I ON BEING A PHYSICIAN

David C. Dale, m.d., f.a.c.p.
Daniel D. Federman, m.d., m.a.c.p.

Medical practice is constantly changing. Almost daily there is important new information regarding basic disease mechanisms and new therapies. There is a constant need to reconsider how we diagnose and treat both common and rare diseases. The way that hospitals and clinics are organized, how we pay for health care, and how our services are evaluated are also changing. The future promises more changes, some of which will undoubtedly increase the burden of illness and the work for physicians. Population growth, poverty, and emerging infectious diseases, as well as inactivity, dietary changes, and obesity, are worldwide problems that have immense medical implications. Other changes are coming rapidly that should improve health care and be helpful to physicians—new information technologies to aid in obtaining current medical information and record keeping, applications of discoveries from basic sciences and the human genome project to make diagnosis more precise and treatments more specific, and the development of new drugs and vaccines on the basis of increasing understanding of normal physiology and disease processes.

Although change is a watchword for medicine, many aspects of medicine are not changing rapidly; some of these are the most important and most satisfying aspects of being a physician. In the community and in the patient-doctor relationship, physicians are still seen as persons skilled in the art of healing and in teaching others about health and disease. Physicians are still the ones who receive the extensive training, the licensure by the state, and the approval of society to provide all levels of care: to give advice for a healthy life, to examine and diagnose illness, to prescribe drugs to relieve suffering, and to care for those who are seriously ill and dying. Although physicians now share the many responsibilities involved in patient care and work closely with nurses, physician assistants, pharmacists, technicians, therapists, and family members of patients, it is still the physician who bears most of the responsibility for the care of the patient.

Being a patient's physician carries many responsibilities and requires at least three attributes. First, knowledge of the applicable biomedical science and clinical medicine is necessary to understand a patient's problem. There is no limit to the knowledge that may be needed, but it is important to be able to answer correctly the patient's questions, such as "How did this happen to me?" and "Will I be better soon?" The physician needs to understand disease processes well enough to identify and categorize a patient's problem quickly. It is important, and sometimes critical, to know whether the problem will resolve spontaneously or whether detailed investigations, consultations, or hospitalization is needed. A thorough and up-to-date understanding of diagnosis and treatment is essential for the day-to-day exchange of information that occurs between physicians as they solve the problems of individual patients and work together to organize systems to improve patient care.

Second, some specific skills are necessary to diagnose and treat a patient. The ability to communicate—both to speak and to listen—is essential, especially for physicians providing primary care. Effective and sensitive communication can be challenging in communities characterized by diverse cultures and languages. At times, the physician must be, in part, an anthropologist to grasp the patient's understanding of illness and of the roles of patient and doctor. Knowing how to communicate empathically is also invaluable: It is important to welcome each patient at every visit, to reach out and hold the hand of a troubled person, and to express understanding and concern. The ability to balance the time spent with the patient and the time required for organizing services for the patient in a busy practice is an increasingly important skill.

The physical examination remains a fundamental skill; the ability to recognize the difference between normal and abnormal findings, adjusting for age, sex, ethnicity, and other factors, is crucial. Good record keeping is essential—with regard to both a written record and a mental record—so that the circumstances of visits are remembered and changes in a patient's appearance or other characteristics that may not have been recorded can be recognized. With practice and attention, these skills—history taking, physical examination, and record keeping—can grow throughout a professional lifetime. Other aspects of care, such as selecting and performing diagnostic tests, procedures, and treatments, require evolving expertise. For all physicians, it is necessary both to practice medicine and to study regularly to maintain all of these essential skills.

The third, but by no means least important, attribute is the physician's responsibility to the patient and the medical community to conform to appropriate professional and ethical conduct. The first principle of the doctor-patient relationship is that the patient's welfare is paramount. Putting the patient first necessitates understanding the patient and the patient's values. It often means spending precious personal time explaining illness, determining the best method of treatment, or dealing with emergencies. It places the physician in service to the patient. Ethical conduct includes seeing clearly and acknowledging situations in which the physician's interest may conflict with the interest of the patient. Ethical conduct also requires recognizing and acknowledging conflicts of interest in profiting from the prescribing of services and treatments, ownerships of equities and properties, and personal and business relationships. Finally, personal exploitation of the intimacy and privacy of the doctor-patient relationship is never allowed.

Thus, the work of the physician—recognizing illness, providing advice and comfort, relieving pain and suffering, and dealing with illness and death—has not changed much even since ancient times. On another level, however, the work has changed greatly. Better medical record keeping, quantitative observation, meticulous experimentation, and carefully conducted clinical trials have contributed to the rapid evolution of medical practice in this century. Simultaneously, medical education at the undergraduate, graduate, and postgraduate levels has been dedicated to the organization of a truly scientific knowledge base and its translation into intellectually cohesive approaches to understanding disease. Extraordinary advances in the biologic sciences, the development of medical and surgical specialties, and the explosion of medical information have brought with them great benefits. They have also added to the costs and the potential costs of almost every aspect of health care.

Efficiency and cost containment are now watchwords of the payers for health service. Practice guidelines, hospital care pathways, and other efforts to codify the practice of medicine are receiving much attention. When based on good evidence, these efforts are beneficial; they save precious resources—time and money—for both patients and physicians. The development of managed care in the United States has created a new challenge for physicians: to serve as advocates for their patients. In this role, physicians are responsible for overcoming organizational, geographic, and financial barriers to the provision of services that are important for their patients. In organizations in which guidelines for care have been established, it may be necessary for a physician to explain to administrators the specific needs and problems of individual patients—sometimes over and over again, because laypersons may be less apt to recognize that guidelines for clinical practice must remain just guidelines. Because more and more physicians are salaried and thus bound to the needs of populations of patients, physicians face the problem of balancing the needs of individual patients with the expectations of the employer. This is a delicate and, in some places, even fragile balance. To serve both patients and the employer well, a physician must develop good judgment in managing patient care under conditions in which the allocation of resources is conservative.

The increasing organization of health care on a for-profit basis has raised new issues. The physician's obligation to put the patient first, the thoroughness inculcated in physicians throughout their training, and the increasing costs of diagnostic tests and therapies can collide head-on with health care management's attempts to protect earnings for investors. Professional responsibility to patients and the public good is clear and at times poses difficult challenges for the physician.

A profession is defined by a specialized body of knowledge requiring advanced training and by the dedication of its practitioners to the public good over their own enrichment. In exchange, professionals are granted considerable autonomy in setting standards and in the conduct of their work. Circumstances within the medical profession have changed. The public in general and patients in particular have much more knowledge of medicine than at any time in the past, and the modern organization of medicine has severely restricted the autonomy of physicians. But delivery of expert medical care and the welfare of the patient remain central to the physician's professional responsibility. Maintaining professionalism as the ground moves under us is more important than ever.

The weight of all these responsibilities may suggest that it is impossible, or nearly impossible, to be a good physician. Quite the contrary, persons with vastly different personalities, interests, and intellects have become and are becoming good physicians and are deeply satisfied in this role. The information necessary for practicing medicine is now more accessible than ever before. The skills the physician needs can be learned through experience, sharpened through practice, and focused through specialization. The ethical requirements of physicians are not onerous. They are, in fact, expectations of all good citizens, regardless of their careers. Being a physician is both exciting and satisfying; it provides a unique opportunity to combine modern scientific knowledge with the traditions of an ancient and honored profession in serving and helping one's fellow man.

ACP Medicine is written and edited by physicians to help other physicians meet the ideals enunciated in this introduction. A principal goal of *ACP Medicine* is to be the most up-to-date textbook of medicine available. The section Clinical Essentials presents the contemporary skills and knowledge needed by all physicians to encourage and maintain good health, to analyze medical information, to deal compassionately with the end of life, and to understand issues of medical ethics. The other sections organize and summarize the most important information on pathophysiology, diagnosis, and treatment for most problems encountered in practicing medicine for adults from general and specialty journals, as interpreted by experienced clinicians. The material is evidence-based, with extensive bibliographic citations that are updated regularly. Authors are selected who understand both the constraints of managed care and the quality of care that is possible with scientific advances. In short, *ACP Medicine* is committed to conveying the information necessary for physicians to provide excellent care to their patients.

II CONTEMPORARY ETHICAL AND SOCIAL ISSUES IN MEDICINE

Christine K. Cassel, m.d., m.a.c.p.
Ruth B. Purtilo, ph.d.

In the past, medical ethics was thought to refer solely to proscriptions against physicians advertising their services and fees or engaging in questionable economic arrangements such as fee-splitting. Within the past 20 years, however, medical ethics has evolved into a discipline in which clinicians (physicians, nurses, and other health professionals), philosophers, theologians, and social scientists speak knowledgeably about value conflicts that arise in the practice of medicine.[1,2] Physicians have come to recognize the need to be knowledgeable about complex and wide-ranging moral issues as the result of advances in biomedical science and technology; changes in the delivery of health care; changing worldwide demographic trends; epidemics (e.g., the AIDS and severe acute respiratory syndrome epidemics) and new or reemerging infectious illnesses (e.g., avian influenza and Marburg viruses); and a growing understanding of the interconnectedness of individual and public health concerns. The AIDS pandemic has brought awareness that global health threats and cross-cultural contacts can present clinical, epidemiologic, and ethical challenges. For example, what responsibility does the international community have to provide assistance to severely underprivileged, impoverished countries experiencing an AIDS epidemic when treatment of the disease is readily available but out of reach of millions of people who suffer from its ravages?

Ethical issues in the clinical setting persist, and physicians need to be aware of legal decisions and new technologies that affect clinical practice. The rapid, continuing advances of medical technology have raised a host of moral issues around such fundamental questions as when does life begin, when and how does life end, which services can patients require of physicians, and which requests can physicians legitimately refuse. These questions become even more complex in a society as diverse and multicultural as our own, where moral norms may conflict. Respect for the personal values of our patients requires physicians to examine ethical dilemmas carefully and analytically. Consider the following ethical dilemmas and the questions that each one raises for physicians today:

- A 90-year-old woman, totally disabled from several strokes, lives at home, where she receives 24-hour care. Her strokes have left her cognitively impaired and unable to communicate. She signed a living will 15 years ago, and her husband is her designated health care proxy. She was hospitalized because she had stopped eating, and while in the hospital, she developed aspiration pneumonia. Four days into her hospitalization, she developed a bleeding ulcer and hemorrhaged several units of blood. She had a cardiac arrest, was resuscitated after 45 minutes of asystole, and is now unresponsive and ventilator dependent. Her husband insists that she be kept alive by whatever means possible. The hospital team is strongly divided about the morally appropriate course of action. Some agree with the patient's husband and argue that the patient should receive life-sustaining treatment, even though she has virtually no chance of recovery. Others argue that it would be more respectful to discontinue intrusive medical care—an action consistent with her advance directive—and allow her to die. What clinical and moral value considerations should govern their final decision?
- A 58-year-old man living in Oregon is suffering from end-stage AIDS with lymphopenia, multiple refractory fungal infections, and Kaposi sarcoma. He has significant pain from mucosal lesions and skin breakdown and has sustained fractures, including one from a spinal metastasis that has led to paraplegia and urinary and fecal incontinence. He is cognitively intact and has given oral and written directives indicating that he does not want to be kept alive any longer. He has repeatedly asked his physician to give him an overdose of sedative so that he will die and be released from his intractable suffering. The physician is convinced that this patient is competent, that he is well-informed about his condition, and that his wish to die is made in good faith. The patient's companion of 15 years agrees with the patient's decision. Both have known the physician for a long time and trust her judgment. Physician-assisted suicide is currently legal in Oregon. Should the physician comply with this patient's wishes? If she cannot do so in good conscience, must she refer her patient to a physician who can? Why or why not?
- Science allows physicians to transplant hearts, livers, kidneys, and other living organs, tissues, and cells. Overall, there are drastic shortages of donors. Hundreds, sometimes thousands, of people die each year before an organ match becomes available. Currently in the United States, people who wish to donate organs are encouraged to indicate that wish on their driver's licenses. In the absence of such clear evidence of consent, physicians and other hospital staff are often reluctant to ask bereaved family members for donations because many people, understandably, cannot deal with such a request in a time of crisis. Should the United States adopt a policy—already practiced in other countries—of allowing hospitals to harvest organs upon the death of a patient unless that person has specified otherwise? Could one policy ever work to everyone's benefit in a diverse society in which there may be differing attitudes about treatment of the dead, the moral use of animals, and other culturally derived considerations? Would therapeutic cloning or xenotransplantation provide ethically preferable alternatives?
- A woman whose family has a strong history of breast and ovarian cancer wants to be tested to determine whether she is a carrier of the *BRCA* family of genes, which confer high risk for these malignancies. She is between jobs and is about to apply for a position with a small, innovative firm that has a self-insured health care plan. She knows that the disclosure of this information would dramatically skew the insurance risk and insurance costs for this company, which is largely composed of young people who have relatively low health care costs. She might be denied the job for these reasons. The patient wants to undergo *BRCA* screening but asks you not to note the results in her medical record. You know that her fears are well founded. What should you do?

These examples highlight the complexity of ethical dilemmas and the need for a common language by which clinicians and society can openly deliberate about ethical issues. Often, there is not a single right answer to an ethical dilemma; in almost all cases, there are competing values that need to be weighed against each other before a decision is made that most fully upholds the moral values by which physicians must guide their practice. As in many other areas of medicine, there may be a high degree of uncertainty. For that reason alone, it is useful to have a framework for ethical decision making.

A Context and Process for Ethical Decision Making

A conflict of values lies at the center of each ethical dilemma. Most medical ethicists agree that several fundamental ethical norms can be drawn from the overarching principle that patients should be treated with respect. These ethical norms include the responsibility to act in a way that benefits the patient (beneficence); the responsibility, whenever possible, to do no harm (nonmaleficence); the responsibility to acknowledge the autonomy of the patient and his or her right to self-determination; and the responsibility to treat people fairly and equitably. Although it would be hard to argue against any of these values taken individually, they come into conflict with one another every day in medical practice. Three steps are useful for making decisions when ethical conflicts arise.

First, the clinician needs to gather all available relevant information regarding the patient. Inadequate information can result in decisions that do not reflect the interests and desires of the patient. However, the clinician must be aware that cultural differences and language barriers may limit a patient's understanding of the choices that need to be addressed.[3] Key information includes not only information about the medical condition of the patient but also information about the patient's values and preferences, the family and social situation, and the realities of the options open to the patient.

Second, ethical dilemmas must be clarified and presented clearly to all those involved in the decision-making process. For example, a spouse of an incompetent patient who argues for aggressive, clinically futile treatment in the face of an imminently terminal and untreatable illness can present the physician with a conflict between respecting the considered wishes of family members and doing what the physician judges is best for the patient.[4] Sometimes, enhanced communication between physician, patient, and family helps bring the matter to resolution.[5,6] For example, having a discussion with the family that is focused on the likelihood that aggressive measures would only prolong the suffering of the patient may convince them to end life-prolonging interventions. In other circumstances, however, the patient's and family's beliefs may necessitate that the physician take aggressive measures to preserve life at all costs.[7,8] It may be important to discuss the spiritual and moral dimensions of the impending decision explicitly. It is often helpful to involve other physicians or nonphysician mediators, such as the hospital ethicists, patient advocates, social workers, and clergy members, in the decision-making process.[9] Once values are explicitly discussed and differences clarified, a plan may be agreed upon by which all parties can abide.

Third, once a decision has been made, it is essential that the decision be carried out effectively, compassionately, and with continuing respect for the patient's needs and wishes. For example, if genetic testing is indicated and there are potential consequences regarding the patient's future eligibility for health insurance, the physician must ensure the confidentiality of information about the tests.[10,11] If complete confidentiality is not possible, the physician should be sure that the patient understands and accepts the risks.[12,13] Whatever the topic at hand, the physician must employ the clinical and interpersonal skills necessary to carry out the patient's wishes respectfully and compassionately.

Areas of Current Ethical Debate

Three broad societal concerns that have important implications for clinical practice lie at the center of many current ethical dilemmas.

DIFFERENCES OF OPINION ABOUT THE MORAL LIMITS OF MEDICAL INTERVENTION IN AN ERA OF TECHNOLOGICAL IMPERATIVES

Modern medicine has been criticized for generating an ethos in which clinicians assume that if an intervention is available, it should always be used. A physician might offer a new intervention as a way of either sustaining hope for the patient and family or avoiding the reality of a poor prognostic situation. In these circumstances, the chances of success can sometimes be overestimated. There are times when the better course is to help patients and families deal realistically with their losses. Physicians' ethics should allow them to consider each medical intervention in the light of their patient's values and wishes and with due regard for the appropriateness of the treatment in that particular setting.[14,15] Several questions frame the current debate about the appropriate use of medical technology—among them, questions as to when life begins and ends, what constitutes quality of life, and is it appropriate to withhold interventions in the face of medical futility.

THE ENIGMA OF WHAT CONSTITUTES A PERSON AND WHEN LIFE BEGINS AND ENDS

Physicians sometimes face extreme and unfamiliar situations in discharging their duty to respect a patient's autonomy. Current research in genetics, for example, challenges traditional assumptions of the uniqueness of individual identity and the acceptability of genetic interventions.[16] Germline interventions were considered completely ethically unacceptable just a few years ago because of the reluctance on the part of geneticists to create changes that would persist through subsequent generations. However, research has now progressed to the point of growing human stem cells under laboratory conditions, and stem cell research is thought to be one of the most promising new areas for clinical interventions.[17] Although the debate has become intensely political and national funding of stem cell research by the National Institutes of Health is strictly proscribed, many in the scientific community are actively supporting stem cell research and are taking steps to address some of the ethical concerns raised by the use of these cells. This shift has occurred in part because stem cell techniques do not create permanent germline changes. Scientific research is ongoing, especially in other countries[18]; in the United States, interest in the clinical promise of stem cell technology continues to grow. Legislators in California and Massachusetts have created and funded state-level research centers, and other states are considering whether to undertake similar initiatives.

Attempts to promulgate practice guidelines for governing the conduct of stem cell research engender extensive debate; such debate generates rich ethical discourse that addresses the very essence of personhood. Reproductive technologies, including the potential for cloning, have an impact on this issue and have spurred new questions about the ethics of medical intervention in human reproduction. The debate about abortion in the United States continues to encompass many points of controversy that directly affect the practice of medicine, sometimes violently.

At the other end of the continuum of care is the question of when life ends. This question is brought into sharp focus by dramatic life-extending technologies. For example, although rational criteria for brain death have served to guide organ transplantation, the extreme shortage of donor organs and evolving technological capabilities have prompted new ethical considerations regarding organ recovery. As utilization of organ donations from non–brain-dead but irreversibly comatose persons has become an increasingly common practice, commitment to clarifying and addressing the ethical dilemmas associated with the use of such donors remains warranted.[19,20] Finally, the debate about assisted suicide raises profound questions of quality of life and the limits of personal choice [*see 11:VII Anoxic, Metabolic, and Toxic Encephalopathies and 13:IV Care and Management of the Patient at the End of Life*].[21-24]

APPROPRIATE APPROACHES TO ASSESSING QUALITY OF LIFE

Discussions of quality of life gain broader clinical relevance as technical advances make it easier to extend life beyond a point where many people would consider it meaningful. When a patient or family member raises the issue, it is important for the physician to learn more about what that person means by "quality of life." Physicians, family members, and patients may disagree about what constitutes an acceptable quality of life. Often, the phrase is used in the context of how long clinicians should continue attempts to extend life. The ideal setting for gathering this key information is in an ongoing caregiving relationship that allows the patient time to think about the issues, discuss them with those close to him or her, and come back to the physician for a fuller discussion.[25] Unfortunately, this ideal relationship is becoming increasingly rare. Crucial decisions must often be made among relative strangers in times of great stress (for example, in an intensive care unit or when the patient is on the brink of having a cardiac arrest precipitated by critical illness).[26]

For that reason, physicians should try to open the door to these discussions with patients ahead of time whenever appropriate.[27] Increasingly frequent discussions of death and dying in the popular media have set the stage for patients and families to be receptive to such discussions and to be better informed about the facts and issues involved.

In general, questions related to acceptable quality of life should be answered by the patient.[28] Often, however, the patient is unable to speak for himself or herself when the answer is needed.[29,30] For example, patients with advanced dementia from Alzheimer disease or with irreversible coma cannot make these decisions, and few such patients have written detailed and specific advance directives. A proxy decision maker, usually a family member or a friend, should be asked about the patient's likely wishes in such a situation. It is crucial to emphasize to a proxy that it is the patient's values, not the proxy's, that should be conveyed in these situations. In addition to providing clear information about prognosis and likely outcomes, it is important for the clinician to recognize that a proxy is in a very difficult position—often in the midst of acute grief or anxiety—and should be provided a comforting context in which to make a decision. A proxy should not be made to feel that he or she is alone in making this decision, especially in the common situation in which the patient is likely to die in any case. Written advance directives—so-called living wills—can be helpful in this regard, mostly as adjuncts to discussions between patient and physician. Assigning a trusted proxy is still recommended, however, because situations are often more complex than can be adequately addressed in a written document.

Traditional Medical Ethics and the Changing World of Medicine

One of the reasons the medical profession has been able to maintain a strong ethical standard for more than 2,000 years is that the standard has been so simple. From the Hippocratic oath to the prayer of Moses Maimonides, statements of medical ethics have required the physician to do what is best for the patient, putting the patient's interest before the physician's own. Admittedly, there have been breaches of the standard. Many physicians became rich selling unproven patent medicines before the advent of scientific medicine. More recently, some have overcharged patients or have ordered unnecessary tests, medications, and procedures to further their own financial interests. Overtreatment can be as unethical as undertreatment, for two reasons: (1) all treatments carry some degree of risk to the patient and (2) rising health care costs contribute to the difficulty our nation faces in extending health care access to the uninsured and underinsured. Generating costs to enhance one's own income, with no benefit to the patient, adds to the barriers facing populations who are underserved by the health care system. Physicians have a responsibility for societal health, as well as the health of their individual patients.[31]

In the past, the accountability structure for health care was clearly delineated between physician and patient. Today, changes in the economics and delivery systems of managed care have so affected this classic ethical construct of undivided loyalty to the patient that even previously inviolable ethical relationships are being challenged.[32,33]

Managed care has been criticized for withholding care from patients, and physicians have been seen as the agents of rich insurance companies rather than as advocates for the best health care for their patients. Good managed care, however, allows physicians to limit risk to patients and reduce waste and cost. Physicians are challenged to examine whether their role in such managed care programs is truly in the patients' best interest; to do so could ameliorate the loss of public trust in the profession.[34] Studies by the Institute of Medicine and the RAND Corporation have spurred new approaches in the provision of quality care.[35,36] Physicians are being asked to measure the quality of their care and make this information available to payers, consumers, certifying/accrediting organizations, and others [*see CE:XIII Performance Measurement in Clinical Practice*]. Such measures are imperfect but will reveal to patients and others standards for ideal care. How should physicians respond to these new demands for transparency? Understanding the fundamental responsibility of the profession to the welfare of the patient is an important starting point for dealing with any ethical problems arising from social change, technological innova-

tion, and changes in the delivery and financing of health care. The Physician Charter on Medical Professionalism identifies a modern framework that may be used to clarify professional standards in this more complex world.[37,38]

POPULATION-BASED MEDICINE AND THE RIGHTS OF THE INDIVIDUAL

Although simple in the abstract, the physician's responsibility to the patient is not always clear in actual practice. For example, the traditional standard requires a physician to do everything possible for patients directly in his or her care. Arguing that a more utilitarian standard is needed, some theorists have suggested changes to meet the requirements of population-based medicine, in which some treatments that are potentially beneficial to the individual patient are forgone to benefit larger numbers of patients with the available resources.

Utilitarian considerations are sometimes discussed in the context of a communitarian philosophy, which holds that all members of the society are better off if standards are based on the benefit to communities rather than to individuals exclusively.[39] Many European governments base policies on communitarian premises, whereas in the United States, policy makers have traditionally focused more sharply on the rights of the individual. However, it may be that the rights of a far greater number of individuals would be better served with a health care structure that emphasizes more collective responsibility and resolution.

One area where this tension can be seen is in end-of-life care. In recent decades, there has been a presumption and a legal standard in the United States that patients may make their own decisions about the care they receive at the end of life and, in particular, that every person has the right to refuse life-sustaining treatment. This freedom of choice is the thrust of the Patient Self-Determination Act of 1990, which requires hospitals and nursing homes to inform patients of local laws regarding advance directives and to help them prepare advance directives if they choose to do so. In several well-publicized cases (e.g., the Quinlan, Cruzan, and Schiavo cases),[40,41] courts supported families or patients who wished to end life-sustaining treatment. However, attention is now being drawn to instances in which patients or their proxies ask for life-sustaining treatment over objections from health care payers and, sometimes, providers. In the relatively few cases in which such conflicts have been brought to litigation, courts, again, have been generally supportive of patients' and families' desires. Interestingly, these cases conflict with the recent judgments that financial incentives to restrict care are acceptable in the context of insurance law.

In recent years, some ethicists have worked to define a standard of medical futility that would give physicians the right to withhold treatment in specific cases.[42,43] There is profound disagreement, however, about the definition of futility and its statistical basis. For example, the chances of success with cardiopulmonary resuscitation (CPR) are remarkably small in patients of very advanced age who have debilitating illness and poor functional status, particularly in cases of an unwitnessed cardiac arrest; however, many physicians would be uncomfortable making the decision to withhold CPR without consulting the patient's family.[44] From one perspective, this inclination to involve and communicate with patient and family is a sound one, motivated by respect and caring.[45] In other cases, however, an insistence on family permission in the context of medical futility is a misguided gesture, perhaps driven by liability con-

Biomedical Ethics Information on the Internet

Federal Government

Bioethicsline
http://www.nih.gov/sigs/bioethics
The National Library of Medicine's database of peer-reviewed bioethics literature.

National Bioethics Advisory Commission
http://bioethics.gov
Agendas and transcripts of meetings, online publications, and other information primarily regarding genetics research and research involving humans.

Ethical, Legal and Social Implications Program, National Human Genome Research Institute
http://www.ornl.gov/sci/techresources/Human_Genome/elsi/elsi.shtml
Information on policy and legislation, research opportunities, grant products and publications, education and training activities.

Professional Societies

American College of Physicians Center for Ethics and Professionalism
http://www.acponline.org/ethics
Position papers, educational programs, and other resources on end-of-life care, managed care, and other issues related to medical ethics.

American Medical Association Institute for Ethics
http://www.ama-assn.org/ama/pub/category/2416.html
Educational and outreach programs for physicians, including the Education for Physicians on End-of-life Care Project.

American Society for Bioethics and Humanities
http://www.asbh.org
Consolidation of the Society for Health and Human Values, the Society for Bioethics Consultation, and the American Association of Bioethics; meeting agendas, position papers.

American Society of Law, Medicine & Ethics
http://www.aslme.org
Conference agendas, publications, online forum.

Bioethics Council
http://www.bioethics.org.nz/about-bioethics/international-links.html
Comprehensive guide to international resources in bioethics.

Institutes and Centers

Case Western Reserve University Center for Biomedical Ethics
http://www.cwru.edu/med/bioethics/bioethics.html
Program news, events, online newsletter.

Georgetown University Kennedy Institute of Ethics
http://www.georgetown.edu/kie
Information on symposia, publications, and services, including the National Reference Center for Bioethics Literature. (**http://www.georgetown.edu/nrcbl**)

The Hastings Center
http://www.thehastingscenter.org
Research and educational programs on ethical issues in medicine, the life sciences, and the environment.

University of Chicago MacLean Center for Clinical Medical Ethics
http://ethics.bsd.uchicago.edu/resources.html
Comprehensive guide to online resources in biomedical ethics; online newsletter.

University of Pennsylvania Center for Bioethics
http://www.med.upenn.edu/bioethic
Online bioethics tutorial, publications, discussion groups; special sections on genetics, cloning, and physician-assisted suicide.

cerns. Ethicists have asserted the physician's duty to regain the responsibility of prognostication and decision making inherent in the older paternalistic model of medical practice.[46,47] This belief can be supported by two arguments: (1) there is a responsibility to avoid wasteful use of scarce resources and (2) the attitude of caring means to avoid inflicting unrealistic choices on grieving families and to offer reassurances of aggressive palliative care and relief from suffering for patients who are dying.

Outside of the context of life and death, the allocation of medical resources is an area in which the tensions between wasteful expenditure and appropriate care are regularly played out. For example, the high cost of brand-name medications would lead a physician to prescribe equally effective generic agents whenever possible; however, direct-to-consumer advertising and drug detailing to physicians have created a demand for brand-name medications, even when there is no evidence that they are better than older formulations. On the basis of biomedical and clinical evidence, government agencies (e.g., the Centers for Medicare and Medicaid Services, the Veterans Administration, and certain state agencies) and health care plans have created formularies that determine the most cost-effective medications. These formularies are regularly used in the filling of prescriptions covered by health care payers. When these formularies are used, patients may not be given the brand of medication they request; however, they receive a formulation that is equally effective. The savings resulting from the adjustment in prescription allows health care plans to cover the health care costs of larger numbers of people. Patients who prefer the more expensive brand-name medication may have it if they pay for it. In these instances, does the physician's responsibility to avoid waste override the responsibility to respect the patient's values? The Physician Charter calls on physicians to reduce waste and improve quality. The achievement of this dual goal may require concerted effort to educate patients—and possibly policy makers—about the best uses of their health care resources.[31,37,38]

A BROADER CONTEXT FOR CLINICAL DECISION MAKING

The role of the physician and the nature of the doctor-patient relationship may be challenged, not only by changes in the practice of medicine but also by the increasing interconnectedness of communities and societies and the emergence of public health as a global concern. Regional and national health care systems are commonplace; epidemics can occur worldwide because of widespread international travel, immigration, and dislocation caused by war and civil strife.

The global, multicultural aspect of modern medicine will have increasingly significant implications for ethical decision making in clinical practice in coming years. For example, in seeking to honor a patient's right to autonomy, a physician may have to balance the traditional standard of care with a patient's desire to choose an alternative or complementary therapy; or following the traditional Hippocratic model, a physician may feel justified in using the most powerful antibiotic available to treat a patient's infection, despite the fact that the widespread use of powerful antibiotics leads to the emergence of new and more resistant organisms throughout the world.

Caring for patients in this new environment raises challenges for modern physicians that their predecessors never faced. Physicians must now analyze ethical issues systematically, understand the conflicts modern medicine poses for some traditional Hippocratic precepts, and come to terms with the conflict between their responsibility to their patients and the consequences of individual clinical decisions for the broader population. Even as electronic communication systems evolve to keep physicians abreast of new global realities, the moral and ethical framework of clinical decision making must begin to encompass those realities [*see Sidebar* Biomedical Ethics Information on the Internet]. It is critical that physicians learn the language of medical ethics and follow its literature closely so that their voices will help shape basic medical values in the future, even as they cope with complex ethical challenges in their daily practice.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Pellegrino ED: The metamorphosis of medical ethics: a 30-year retrospective. JAMA 269:1158, 1993

2. Pellegrino ED: Ethics. JAMA 275:1807, 1996

3. Davis TC, Williams MV, Marin E, et al: Health literacy and cancer communication. CA Cancer J Clin 52:134, 2002

4. Hines SC, Glover JJ, Babrow AS, et al: Improving advance care planning by accommodating family preferences. J Palliat Med 4:481, 2001

5. Quill TE, Brody H: Physician recommendations and patient autonomy: finding a balance between physician power and patient choice. Ann Intern Med 125:763, 1996

6. Medical futility in end-of-life care: report of the Council on Ethical and Judicial Affairs. JAMA 281:937, 1999

7. Asch DA, Hansen-Flaschen J, Lanken PN: Decisions to limit or continue life-sustaining treatment by critical care physicians in the United States: conflicts between physicians' practices and patients' wishes. Am J Respir Crit Care Med 151:288, 1995

8. Cantor NL: Can healthcare providers obtain judicial intervention against surrogates who demand "medically inappropriate" life support for incompetent patients? Crit Care Med 24:883, 1996

9. DuVal G, Clarridge B, Gensler G, et al: Experiences with ethical dilemmas and ethics consultation. J Gen Intern Med 19:251, 2004

10. Dressler L: Genetic testing for the *BRCA1* gene and the need for protection from discrimination: an evolving legislative and social issue. Breast Dis 10:127, 1998

11. Gostin LO: National health information privacy: regulations under the Health Insurance Portability and Accountability Act. JAMA 285:3015, 2001

12. Geller G, Botkin JR, Green MJ, et al: Genetic testing for susceptibility to adult-onset cancer: the process and content of informed consent. JAMA 277:1467, 1997

13. Ensenauer RE, Michels VV, Reinke SS: Genetic testing: practical, ethical, and counseling considerations. Mayo Clin Proc 80:63, 2005

14. Sharpe VA, Faden AI: Appropriateness in patient care: a new conceptual framework. Milbank Q 74:115, 1996

15. Brett AS, Jersild P: "Inappropriate" treatment near the end of life: conflict between religious convictions and clinical judgment. Arch Intern Med 163:1645, 2003

16. Kaji E, Leiden J: Gene and stem cell therapies. JAMA 285:545, 2001

17. Daar AS, Bhatt A, Court E, et al: Stem cell research and transplantation: science leading ethics. Transplant Proc 36:2504, 2004

18. Ethics can boost science. Nature 408:275, 2000

19. Herdman R, Beauchamp TL, Potts JT: The Institute of Medicine's report on non–heart-beating organ transplantation. Ken Inst Ethics J 8:83, 1998

20. DeVita MA, Snyder JV, Arnold RM, et al: Observations of withdrawal of life-sustaining treatment from patients who became non–heart-beating organ donors. Crit Care Med 28:1709, 2000

21. Physician-assisted suicide: toward a comprehensive understanding: report of the Task Force on Physician-Assisted Suicide of the Society for Health and Human Values. Acad Med 70:583, 1995

22. Drickamer MA, Lee MA, Ganzini L: Practical issues in physician-assisted suicide. Ann Intern Med 126:146, 1997

23. Muskin PR: The request to die: role for a psychodynamic perspective on physician-assisted suicide. JAMA 279:323, 1998

24. Kohlwes RJ, Koepsell TD, Rhodes LA, et al: Physicians' responses to patients' requests for physician-assisted suicide. Arch Intern Med 161:657, 2001

25. Meier DE, Morrison RS, Cassel CK: Improving palliative care. Ann Intern Med 127:225, 1997

26. Schneiderman LJ, Gilmer T, Teetzel HD, et al: Effect of ethics consultations on nonbeneficial life-sustaining treatments in the intensive care setting: a randomized controlled trial. JAMA 290:1166, 2003

27. Searight HR, Gafford J: Cultural diversity at the end of life: issues and guidelines for family physicians. Am Fam Physician 71:515, 2005

28. Garrett JM, Harris RP, Norburn JK, et al: Life-sustaining treatments during terminal

illness: who wants what? J Gen Intern Med 8:361, 1993

29. Cranford RE: The vegetative and minimally conscious states: ethical implications. Geriatrics 53(suppl 1):S70, 1998

30. Rubenfeld GD: Principles and practice of withdrawing life-sustaining treatments. Crit Care Med 20:435, 2004

31. Gruen R, Pearson S, Brennan T: Physician-citizens: public roles and professional obligations. JAMA 291:94, 2004

32. Quill TE, Cassel CK: Nonabandonment: a central obligation for physicians. Ann Intern Med 122:368, 1995

33. Mechanic D, Schlesinger M: The impact of managed care on patients' trust in medical care and their physicians. JAMA 275:1693, 1996

34. Mechanic D: Changing medical organization and the erosion of trust. Milbank Q 74:171, 1996

35. Crossing the quality chasm: a new health system for the 21st century/Committee on Quality Health Care in America. Institute of Medicine. National Academy Press, Washington, DC, 2001

36. McGlynn EA, Asch SM, Adams J, et al: The quality of health care delivered to adults in the United States. N Engl J Med 348:2635, 2003

37. Medical professionalism in the new millennium: a physician's charter. ABIM Foundation, American Board of Internal Medicine; ACP-ASIM Foundation, American College of Physicians-American Society of Internal Medicine; European Federation of Internal Medicine. Ann Intern Med 136:243, 2002

38. Brennan TA: Physicians' professional responsibility to improve the quality of care. Acad Med 77:973, 2002

39. Luce JM: The changing physician-patient relationship in critical care medicine under health care reform. Am J Respir Crit Care Med 150:266, 1994

40. Gostin LO: Deciding life and death in the courtroom: from Quinlan to Cruzan, Glucksberg, and Vacco: a brief history and analysis of constitutional protection of the 'right to die.' JAMA 278:1523, 1997

41. Annas G: "Culture of life" politics at the bedside: the case of Terri Schiavo. N Engl J Med 352:1710, 2005

42. Schneiderman LJ, Jecker NS, Jonsen AR: Medical futility: response to critiques. Ann Intern Med 125:669, 1996

43. Marco CA, Larkin GL, Moskip JC, et al: Determination of "futility" in emergency medicine. Ann Emerg Med 35:604, 2000

44. Hofmann JC, Wenger NS, Davis RB, et al: Patient preferences for communication with physicians about end-of-life decisions. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT) investigators. Ann Intern Med 127:1, 1997

45. Doukas DJ, McCullough LB: A preventive ethics approach to counseling patients about clinical futility in the primary care setting. Arch Fam Med 5:589, 1996

46. Helft PR, Siegler M, Lantos J: The rise and fall of the futility movement. N Engl J Med 343:293, 2000

47. Schneiderman LJ, Capron AM: How can hospital futility policies contribute to establishing standards of practice? Camb Q Healthc Ethics 9:524, 2000

Practice Guidelines and Consensus Statements

American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility. J Clin Oncol 21:2397, 2003

Consensus statement of the Society of Critical Care Medicine's Ethics Committee regarding futile and other possibly inadvisable treatments. Crit Care Med 25:887, 1997

Emanuel LL: A professional response to demands for accountability: practical recommendations regarding ethical aspects of patient care. Working Group on Accountability. Ann Intern Med 124:240, 1996

Ethical considerations in the allocation of organs and other scarce medical resources among patients. Council on Ethical and Judicial Affairs, American Medical Association. Arch Intern Med 155:29, 1995

Fletcher JC, Siegler M: What are the goals of ethics consultation? A consensus statement. J Clin Ethics 7:122, 1996

Karlawish JH, Quill T, Meier DE: A consensus-based approach to providing palliative care to patients who lack decision-making capacity. ACP-ASIM End-of-Life Care Consensus Panel. American College of Physicians–American Society of Internal Medicine. Ann Intern Med 130:835, 1999

Lo B, Quill T, Tulsky J: Discussing palliative care with patients: ACP-ASIM End-of-Life Care Consensus Panel. American College of Physicians–American Society of Internal Medicine. Ann Intern Med 130:744, 1999

Position paper of the Ethics Committee of the International Xenotransplantation Association. Sykdes M, d'Apice A, Sandrin M, et al: Transplantation 78:1101, 2004

Practice parameters: assessment and management of patients in the persistent vegetative state (summary statement). The Quality Standards Subcommittee of the American Academy of Neurology. Neurology 45:1015, 1995

Snyder L, Leffler C: Ethics manual: fifth edition. Human Rights Committee of the American College of Physicians. Ann Intern Med 142:560, 2005
http://www.acponline.org/ethics/ethicman5th.htm

III REDUCING RISK OF INJURY AND DISEASE

HAROLD C. SOX, JR., M.D.

Prevention: A Brief Overview

During the past 2 decades, disease and injury prevention has occupied an expanding share of medical practice. Public interest in prevention is very high, driven by a steady accumulation of high-quality evidence that preventive interventions do reduce cause-specific death rates. The purpose of these interventions is to eliminate the root causes of diseases that precede death (e.g., heart disease, cancer, and stroke), which in the United States in 1990 were tobacco use (400,000 deaths), poor diet and inadequate physical activity (300,000 deaths), alcohol consumption (100,000 deaths), microbial agents (90,000 deaths), toxic agents (60,000 deaths), firearms (35,000 deaths), unprotected sexual intercourse (30,000 deaths), motor vehicle accidents (25,000 deaths), and use of illicit drugs (20,000 deaths).[1] These causes of death are the targets of disease and injury prevention. They contribute to 50% of the deaths in the United States. Most are simply bad habits, and changing those habits reduces the risk of dying.

Physicians have two principal roles in prevention: they identify risk factors for disease and injury, and they act as teachers and counselors. Physicians must expand their routine questioning beyond diet, exercise, and substance abuse to include recreational activities that increase the risk of death (e.g., boating, bicycling, and riding motorcycles), gun ownership, use of swimming pools, smoke detectors in the home, and domestic violence. In counseling patients about a healthy diet (including vitamin supplements), exercise, and other elements of a healthy lifestyle, physicians must often help patients adopt healthy living habits. Some patients simply require reinforcement of a chosen lifestyle. Others need help in changing harmful habitual behaviors to healthy behaviors.

The report of the second United States Preventive Services Task Force (USPSTF)[2] contains evidence-based guidelines on 70 prevention topics [*see Table 1*]. As they become available, the third USPSTF reports will appear on its Web site (http://www.ahcpr.gov/clinic/cps3dix.htm). Other literature is also helpful [*see CE:IV Diet and Exercise, CE:V Adult Preventive Health Care, 7:XXXIII HIV and AIDS, 13:III Alcohol Abuse and Dependency, and 13:VI Drug Abuse and Addiction*].

CAVEATS IN DISEASE AND INJURY PREVENTION

Although disease and injury prevention can have a significant effect on the health of the public, physicians should observe several caveats. First, the baseline risk of most diseases is very low in the average person. Each year, colon cancer occurs in 165 per 100,000 men 60 to 64 years of age. The low baseline risk means that the number needed to screen or treat to prevent one death is often very high. Annual fecal occult blood testing must be performed on more than 300 people for 12 years to prevent one death from colon cancer. Whether this inefficiency is important depends partly on the cost of the intervention. Fecal occult blood testing can be costly because it must be performed annually and because abnormal results trigger costly diagnostic tests. Seat belts and smoke alarms are very cost-effective because they incur a onetime cost.

Second, disease prevention does not prevent death. At best, it postpones death by shifting the cause of death from the targeted disease to another disease that strikes later in life. In the Minnesota Colon Cancer Control Study, a randomized trial of fecal occult blood testing, annual testing reduced deaths from colon cancer during 18 years of surveillance. However, the total mortality was the same in the control group and the intervention groups. Our preventive efforts may reduce the likelihood of death from the target disease, whose identity and natural his-

Table 1 Recommendations of the United States Preventive Services Task Force[2]

Tobacco Use

Provide tobacco cessation counseling to patients who use tobacco products. Counsel pregnant women and parents about the potentially harmful effects of smoking on fetal and child health. Prescribe nicotine patches or gum to selected patients as an adjunct to counseling. Give antitobacco messages to young people as part of health promotion counseling.

Alcohol Abuse

Screen all adults and adolescents for problem drinking, using a careful history or a standardized screening questionnaire. Advise pregnant women to limit or abstain from drinking. Counsel all persons who use alcohol about the dangers of operating a motor vehicle or engaging in other potentially dangerous activities while drinking.

*Drug Abuse**

Although the evidence is insufficient for a strong recommendation to be made, it is reasonable to ask adolescents and adults about drug use and drug-related problems. All pregnant women should be advised of the potential adverse effects of drug use on fetal development. Refer drug-abusing patients to specialized treatment facilities where available.

Preventing Motor Vehicle Injuries†

Counsel all patients and the parents of young people to use occupant restraints (lab/shoulder safety belts and child safety seats), to wear helmets when riding motorcycles, and to refrain from driving while under the influence of alcohol or other drugs.

Preventing Falls‡

Counsel elderly patients on specific measures to reduce the risk of falls. Effective measures include exercise, balance training, environmental hazard reduction, and monitoring and adjusting medications. Provide multifactorial individualized interventions to elderly patients at especially high risk for falls.

Fires

Advise homeowners to install smoke detection devices and test them periodically. Infants and children should wear flame-resistant nightclothes. Smokers should cease or reduce smoking.

Drowning

Families with swimming pools should install four-sided 4-ft isolation fences with self-latching gates.

Firearm Injuries

Remove firearms from the home or store them unloaded in a locked compartment.

*There is insufficient evidence to recommend routine screening for drug abuse with standardized questionnaires or biologic assays.
†There is insufficient evidence to recommend for or against counseling patients to avoid pedestrian injuries.
‡There is insufficient evidence of effectiveness of external hip protectors.

tory we know. Inevitably, we raise the lifetime probability of dying from another disease.

Third, we know little about the age at which we should stop our efforts to prevent disease. Our studies provide good information on effectiveness in the study population, which is usually in middle age, but we do not know how the results apply to older people, whose care will occupy an increasing amount of the primary care physician's time.[3] Interventions that take years to show their impact may be ill suited to people whose life expectancy is measured in years rather than decades.

The decision to do a screening test on an older person should depend on the person's general health, which may be quantified as the person's physiologic age. The physician can determine a patient's physiologic age by asking the patient to rate his or her health as excellent, good, fair, or poor.[3] The most likely age at death is the sum of the patient's actual age and the life expectancy that corresponds to the patient's physiologic age. For example, a 75-year-old man in excellent health has a physiologic age of 67 years, which corresponds to a 14.7-year life expectancy [*see Figure 1*]. The most likely age at death is 75 years plus 14.7 years, or 90 years. This information can be very helpful in deciding how hard to press preventive efforts. With a life expectancy of almost 15 years, a 75-year-old man has plenty of time in which to experience gains from preventive efforts.

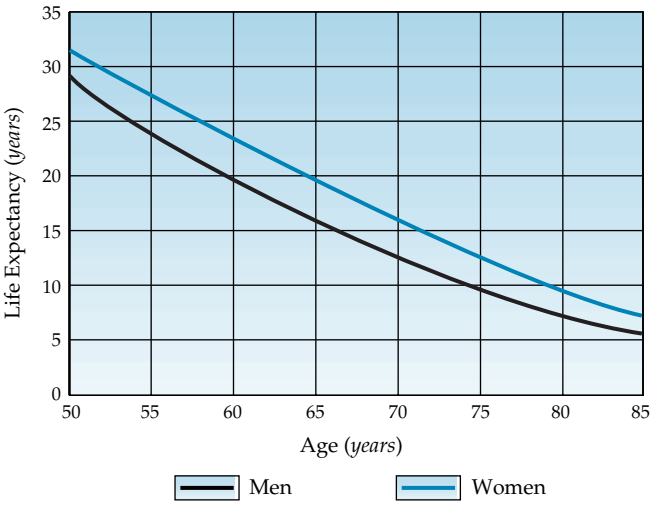


Figure 1 **Life expectancy of men and women in the United States.**

CHANGING BEHAVIOR

Many interventions of proven efficacy are not completely effective because patients are reluctant to change long-established risky behaviors. The USPSTF recommends the following steps for helping patients use their ability to change (self-efficacy)[4]:

1. Match teaching to the patient. Identify a patient's beliefs about a behavior, and adjust advice to the patient's lifestyle. Building the patient's confidence in his or her ability to change requires recognizable successes; define success in terms of goals the patient can achieve.
2. Tell why, what, and when. Patients need to know the reason for a recommendation and the results of following the recommendation. They must also know the time scale for the results so that they do not become discouraged when results do not occur immediately.
3. Small changes succeed. As the patient achieves small successes, propose larger but achievable goals.
4. Be specific. Couch suggestions in terms of current behavior, and give precise instructions in writing.
5. Add new behaviors. Adopting good habits is often easier than discarding bad ones.
6. Link positive behaviors with the daily routine. For example, patients can be encouraged to exercise before lunch or to take medication immediately after brushing teeth.
7. Do not mince words. Tell the patient directly, simply, and specifically what you want and why.
8. Extract promises. Get explicit commitments from the patient. Have the patient tell you exactly how he or she will achieve a goal. Assess the patient's self-confidence and address concerns about succeeding.
9. Use combination strategies. An approach that combines several strategies is more likely to succeed than a single strategy.
10. Involve others. Members of the physician's office staff can become educators. Anyone can offer encouragement to patients.
11. Refer. Subspecialists in many chronic diseases have trained teams that can educate patients far more effectively than individual physicians can. Another form of referral is sending novice patients to talk with successful patients.
12. Stay interested. According to research findings, a call from a health professional to inquire about progress is very effective in changing a behavior. A well-organized office will have a protocol for making these calls a matter of routine.

Health Risks from Substance Abuse

Substance abuse exacts a large toll on the health of the American people, accounting for at least 520,000 deaths in 1990, or 24% of all deaths.[1] Unfortunately, many physicians do not place diseases resulting from self-abuse in the same category as diseases that strike seemingly by chance. Two factors make helping patients shed their habitual use of tobacco, alcohol, and illicit drugs a very efficient way for physicians to add healthy years to patients' lives. First, there is a high baseline risk that substance abuse will lead to serious disease. Therefore, the absolute reduction in risk from a successful intervention is high. This principle is especially true of substance abuse during pregnancy. Second, substance abuse is primarily a problem of youth and middle age, so a successful intervention can add many years of healthy life.

TOBACCO USE

Tobacco contains an addictive drug, nicotine, as well as other substances that contribute to death from cardiovascular disease, cancer, and chronic lung disease. Smoking also contributes to 10% of infant deaths and 20% to 30% of low-birth-weight infants.[5] Tobacco use contributed to one in every five deaths in the United States in 1990 (420,000 deaths a year). In one 40-year cohort study of male physicians in the United Kingdom, half of the deaths after age 35 were smoking related, and smoking caused 25% of the residential fires that resulted in death.[5] Tobacco use is less common than it was several decades ago, but 25% of adults smoke, and an increasing number of smokers are women. Women are starting to smoke earlier—many as high-school students—and they are heavier smokers; nearly twice as many women smoked at least 25 cigarettes a day in 1985 as in 1965.[6]

Cigarettes are highly addictive. Fewer than 10% of people who quit smoking for a day are still abstinent 1 year later. Nicotine, like other highly addictive substances, acts on the dopaminergic mesolimbic pathway, the brain reward pathway that controls motivated behaviors [*see 13:VI Drug Abuse and Addiction*]. The use of nicotine is self-reinforcing, leading to compulsive use. Nicotine produces a withdrawal syndrome that begins within a few hours of abstinence, peaks within the first week, and continues for several weeks. The withdrawal syndrome includes dysphoria, insomnia, irritability, anxiety, difficulty in concentrating, restlessness, slowed heart rate, and increased appetite.[7]

Detection of cigarette smoking is easy; most smokers are truthful when asked about their habit and its extent, and the odor of tobacco is an additional diagnostic clue.

Smoking cessation reduces mortality dramatically. The risk of some diseases (e.g., myocardial infarction and stroke) declines rapidly within a few years after quitting [*see Table 2*].[8] This information is important when one is trying to convince long-term smokers to quit.

Research has shown that a strong message from a personal physician is the most important factor in successful quitting. The elements of successful quitting are consistent, repeated, and strong advice to stop smoking; setting a specific quit date; and follow-up visits to reinforce behavior [*see Table 3*]. However, not all physicians counsel smokers to quit. In one study, only 78% of cigarette smokers reported that their physician had advised them to quit.[9]

A current theory in behavioral psychology suggests that changes in behavior reflect predictable stages in the readiness to change, ranging from no intention to change, to definite plans to change in the near future, to active attempts at change. A stage of readiness to change is predictive of quitting,[10] and this behavioral model may help clinicians shape smoking-cessation efforts to the patient's state of mind.

School-based prevention has received extensive study, and it is effective for at least 2 to 4 years. Clinicians, especially those caring for adolescents, must reinforce the messages of school-based programs.

Nicotine products are an important adjunct to counseling.[11] Drugs are most reinforcing when the level in the brain rises very rapidly, as with inhaled nicotine. Nicotine in medication form, especially transdermal products, appears in the blood much more slowly than inhaled nicotine and produces much less of the reinforcing effect that leads to craving for cigarettes. Plasma nicotine concentrations after transdermal administration reach stable levels in 2 to 3 days. Nicotine medications reduce the symptoms of withdrawal, so that symptoms in the first week are reduced to the level of symptoms at 5 to 10 weeks. Nicotine medications may also provide some nicotine-like effects, such as helping patients to sustain concentration and deal with stress.

Nicotine medications improve abstinence rates, but abstinence at 1 year is still the exception. A meta-analysis summarized the results of 46 trials of nicotine gum and 20 trials of nicotine patches.[12] At 12 months, 19% of patients who received nicotine gum were abstinent, as compared with 11% of patients who did not receive gum. The number needed to treat to achieve one success at 1 year (NNT) was 17 ($P < 0.001$). Transdermal nicotine led to similar rates. At 12 months, 16% of patients who received a transdermal nicotine patch had quit smoking, as compared with 9% who did not receive the patch (NNT, 16). Clonidine also increases the rate of abstinence at 12 months. Weight gain is a common occurrence in patients who have stopped smoking. The average weight gain in a national sample of adults who had stopped smoking was 4.4 kg for men and 5.0 kg for women.[13]

Table 2 Years of Smoking Abstinence Needed to Reduce Risk of Disease[2,8]

Disease	*Years until Risk Is Half of a Current Smoker's Risk*	*Years until Risk Is Equal to a Never-smoker's Risk*
Recurrent myocardial infarction or death from coronary artery disease	1	15
Stroke	2–4	5–15
Oral and esophageal cancer	5	—
Lung cancer	10	20

The dose of nicotine medications should depend on the degree of nicotine dependence.[11] The score on the Fagerstrom questionnaire[14] and the number of cigarettes a day are measures of dependence. Follow-up calls at prearranged times will help the patient to maintain abstinence. It is best to designate a specific member of the physician's office staff to be the smoking-cessation coordinator.

The starting dose of nicotine gum (nicotine polacrilex) is 2 mg per two cigarettes; in patients who smoke more than 20 cigarettes a day, the dose should be 4 mg per three or four cigarettes. Patients may take additional unit doses if their withdrawal symptoms are unpleasant. The medication should be taken at regular intervals throughout the waking hours. The patient should compress the gum a few times with the teeth and then hold it in the mouth, repeating the cycle every minute or so for 15 to 30 minutes for each dose. After 1 to 2 months, weaning can begin with a reduction of 1 unit dose a week.

With transdermal nicotine, patients who smoke more than 10 cigarettes a day should use the largest patch (21 mg). After 1 to 2 months, weaning can begin with each of the lower doses (usually 14 mg and 7 mg, respectively), prescribed for 2 to 4 weeks. Patients who smoke fewer than 10 cigarettes a day can start with the 14 mg dose. A hairless site allows the best absorption. The patient should rotate sites to avoid skin irritation.

Nicotine medications are quite safe[11]—certainly safer than cigarette smoking—even for patients with cardiovascular disease. The only contraindication is hypersensitivity to nicotine or to a component of the delivery system. Twenty-four-hour application of transdermal medication can result in sleep distur-

Table 3 Elements of a Successful Smoking Cessation Strategy

- Direct, face-to-face advice and suggestions
- Reinforcement, especially in first 2 weeks
- Office reminders: a sticker on chart of smokers may stimulate physician to deliver antitobacco message at each visit
- Self-help materials
- Community programs for additional help
- Drug therapy

bance, which subsides if the medication is removed before sleep. Nicotine medication during pregnancy is of concern but probably of less concern than heavy smoking during pregnancy. Medication during pregnancy should be reserved for women who have failed to quit without medication and who smoke more than 10 to 15 cigarettes a day. Dependence on nicotine medications is most likely with delivery systems such as a nasal spray, which causes a rapid rise in the plasma nicotine concentration, and a small number of patients will still be using nicotine medication 1 year after starting treatment.

Nicotine therapy is not the only pharmacologic approach to smoking cessation. Another approach focuses on dopamine. Nicotine releases norepinephrine in the brain and increases dopamine in areas of the brain associated with reinforcing the effects of addictive substances, such as opioids. Bupropion potentiates the effect of norepinephrine and dopamine by acting as a weak inhibitor of their neuronal uptake. Thus, bupropion can mimic some of the central nervous system effects of nicotine and act as a substitute for nicotine in people who are trying to quit cigarettes. Randomized clinical trials have shown that abstinence rates are approximately twice as high with bupropion as with placebo.[15] In one trial, for example, the rates of abstinence from tobacco at 1 year were 12% for those taking placebo and 23% for those taking 150 or 300 mg of bupropion.[16] Side effects of bupropion include agitation and insomnia. Seizures are very uncommon when the daily dose of bupropion is 300 mg or less.

The recommended dose of bupropion is 150 mg/day for the first 3 days and then 150 mg twice daily. The patient should wait to stop smoking until he or she has been on bupropion for 1 week. There are no peer-reviewed reports comparing nicotine-replacement products with bupropion or detailing possible synergy between the two drugs.

ALCOHOL ABUSE

Habitual excessive alcohol consumption causes 100,000 deaths annually in the United States.[1] Although more than one million adults are under treatment for alcoholism, a far greater number engage in drinking that injures their health or has social consequences. The Institute of Medicine estimates that 20% of the population of the United States are problem drinkers, but only 5% are alcohol dependent.[17] Therefore, it is important to distinguish alcohol dependence from problem drinking. Alcohol dependence is associated with major withdrawal symptoms, tolerance, complete loss of self-control, and preoccupation with drinking. Problem drinking, on the other hand, is a less severe condition. Problem drinkers are younger, have a shorter drinking history, have fewer alcohol-related job problems, and have better social resources. In community surveys, the prevalence of problem drinking has been shown to be highest in young men (17% to 24% in 18- to 29-year-olds) and lowest in men and women older than 65 years (1% to 3% and less than 1%, respectively). Women are more frequently problem drinkers than alcohol dependent.

Problem drinking has consequences that affect others, such as motor vehicle accidents, fetal-alcohol syndrome, unsafe sex, domestic violence, and psychological damage to children of problem drinkers. Binge drinking, which is especially prevalent in young adults, leads to violence, unsafe sex, and drunk driving.

Screening for problem drinking and alcohol dependence can be time consuming, and most methods are inaccurate. The gold-standard test for alcoholism is the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria, which require a detailed interview and do not constitute a suitable screening instrument. Results of physical examination and laboratory tests are often normal in problem drinkers. Screening questionnaires such as the modified Michigan Alcoholism Screening Test (MAST), the Alcohol Use Disorders Identification Test (AUDIT), and the CAGE test are the most accurate instruments for detecting problem drinking[2] [*see Tables 4 and 5 and 13:III Alcohol Abuse and Dependency*]. The questions in the MAST and CAGE instruments focus on alcohol dependence and are much less sensitive or specific for binge drinking. The AUDIT screening instrument[2] may be more generally useful because it also asks about quantity of alcohol imbibed, frequency of drinking, and binge behavior. The CAGE and MAST questionnaires may also fail to detect a level of alcohol use that is dangerous during pregnancy.

The treatment of problem drinking depends on the severity of the problem. Problem drinkers often respond to brief office interventions (as short as 10 or 15 minutes), which use motivational techniques such as goal setting, contracts, and enhancing self-efficacy. In most instances, the goal is usually controlled moderate drinking rather than abstinence. The first step toward successful counseling is to get the patient to recognize that there is a problem. It is important to help the patient see a relationship between drinking and current medical or psychosocial problems. Strong advice to reduce consumption is also important. Regular follow-up visits to monitor progress are just as important for problem drinkers as they are for patients with high blood pressure.[18] One meta-analysis of brief intervention trials for nondependent problem drinkers showed a reduction of 24% in average alcohol consumption.[19] Another meta-analysis of nine studies showed somewhat smaller effects. In five of nine studies in men, the number of drinks a week decreased (range of decrease, five to 20 a week). The effects in women were smaller.[20] A third meta-analysis of nine randomized trials found that brief interventions

Table 4 Test Performance of Screening Questionnaires for Alcohol Abuse[2]

Screening Instrument	*Number of Items*	*Sensitivity (%)*	*Specificity (%)*	*Likelihood Ratio Positive*	*Likelihood Ratio Negative*
MAST	25	84–100	87–95	10.2	0.10
CAGE	4	74–89	79–95	6.3	0.21
AUDIT*	10	96	96	24	0.04
AUDIT†	10	61	90	6.1	0.43

*In inner-city clinic population.
†In rural clinic.
AUDIT—Alcoholism Use Disorders Identification Test CAGE—see Table 5 MAST—Michigan Alcoholism Screening Test

Table 5 The CAGE Questionnaire

C: Have you ever felt you ought to **C**ut down on drinking?

A: Have people **A**nnoyed you by criticizing your drinking?

G: Have you ever felt bad or **G**uilty about your drinking?

E: Have you ever had a drink in the morning to steady your nerves or get rid of a hangover (**E**ye-opener)?

were associated with alcohol moderation 6 to 12 months later (pooled odds ratio, 1.91; 95% confidence interval = 1.61 to 2.27).[21] An excellent guide to managing problem drinking is available from the National Institute on Alcohol Abuse and Alcoholism (http://www.niaaa.nih.gov/publications/brochures.htm).

In contrast to the success of brief office-based interventions for problem drinking, successful treatment of alcohol dependence requires intensive therapy from specialists in substance abuse. A randomized trial of employees with alcohol dependence showed the importance of intervening intensively. Participants were randomly allocated to compulsory 3-week hospitalization followed by 1 year of attendance at Alcoholics Anonymous (AA) meetings, mandatory attendance at AA meetings at least three times a week, or a choice of treatment. Rates of being fired from work were similar in all groups, but rates for hospitalization for additional alcohol treatment were much lower in the mandatory-hospitalization group.[22]

The personal physician does have an important role to play in the management of patients with alcohol dependence. In addition to managing medical complications of alcoholism, physicians should be able to use adjunctive therapy for alcohol dependence, such as naltrexone. Another key role for the personal physician is encouragement. Patients need to know that the abstinence rate can be as high as 60% at 10 years after intensive treatment. Finally, the personal physician can lead efforts to help patients solve life problems that are contributing to alcohol dependence.

DRUG ABUSE

The abuse of illicit and legal drugs is a large problem in the United States. A 2001 household survey showed that use of illicit drugs within the previous month peaked among 18 to 20 year olds, at 22.4%, and declined steadily with increasing age. Casual use of marijuana accounts for most of these reports, but as many as 1.7 million Americans use cocaine weekly and 130,000 use heroin. The drug abuser is at risk for many medical complications, but the social cost of drug abuse far outweighs the personal costs. Illicit drug use plays a major role in spreading HIV infection and in homicide, suicide, and motor vehicle accidents. The health care costs of drug abuse are estimated to be $3.2 billion annually, and the cost of federal and state government efforts to stem the flow of illicit drugs is several times higher [*see 13:VI Drug Abuse and Addiction*].

Many professional organizations recommend that physicians ask about drug abuse as part of a periodic health examination of a well person. However, learning about drug abuse may be difficult in the office setting. Patients may be unwilling to acknowledge drug abuse until presented with incontrovertible evidence or after persistent questioning by an alerted physician. There is little information about the accuracy of the history or questionnaires in detecting drug abuse.

Toxicologic testing is the best way to detect illicit drug use. Compared with reference tests, current tests can detect drugs in the urine with 99% sensitivity. However, detection depends on when the patient supplies the specimen relative to the last drug exposure. Marijuana is detectable up to 14 days after use, whereas cocaine, opiates, amphetamines, and barbiturates are present for only 2 to 4 days after use.

Although physicians sometimes test for illicit drugs without obtaining the patient's consent, they do so in the context of trying to determine the cause of a clinical problem that could be caused by an illicit drug. Whether it is ethical to test for illicit drugs in an apparently healthy person who is at high risk for drug abuse is an open question. Regardless of the circumstances leading to testing, abnormal results deserve the physician's best efforts to maintain confidentiality, because they may affect the patient's employability, insurability, and personal relationships. At present, no professional organization recommends drug testing in apparently healthy people.

Physicians must learn to think of drug abuse as a chronic disease. Recidivism after intensive treatment programs is very common, no doubt in part because psychiatric disorders, unemployment, and homelessness often coexist with drug abuse. On the other hand, treating heroin abuse with maintenance methadone, an opioid agonist, can dramatically reduce the social effects of abusing the drug. Heroin addicts in methadone maintenance programs report less use of heroin and reduced rates of HIV infection, criminal behavior, and unemployment. There is no similarly effective treatment for cocaine addiction. Changes in the law now encourage methadone maintenance in office practice. Physicians who wish to treat their regular patients will need to learn about the new regulations.[23]

Health Risks from Accidents and Violence

A person's environment contains many threats to health: motor vehicle accidents, accidents in the home, recreation-related accidents, and domestic violence. Passive strategies, which change the environment in which accidents can occur, are generally more successful at accident prevention than active strategies, which require people to change their behavior. Improving roads saves more lives than exhorting people to drive carefully. For a fuller account, refer to the report of the USPSTF[2] and to a comprehensive review published in 1997.[24]

MOTOR VEHICLE INJURIES

Motor vehicle accidents are the leading cause of loss of potential years of life before age 65. Alcohol-related accidents account for 44% of all motor vehicle deaths. One can experience a motor vehicle accident as an occupant, as a pedestrian, or as a bicycle or motorcycle rider.

Injuries to Motor Vehicle Occupants

In 2000, 37,409 people died of injuries sustained in motor vehicle accidents in the United States.[25] The two greatest risk factors for death while one is driving a motor vehicle are driving while intoxicated and failing to use a seat belt. The physician's role is to identify patients with alcoholism [*see* Alcohol Abuse, *above*], to inquire about seat-belt use, and to counsel people to use seat belts and child car seats routinely. In one study, 53.5% of patients in a university internal medicine practice did not use seat belts. Problem drinking, physical inactivity, obesity, and low income were indicators of nonuse. The prevalence of nonuse was 91% in people with all four indicators and only

25% in those with no indicators.[25] Seat belts confer considerable protection, yet in one survey, only 3.9% of university clinic patients reported that a physician had counseled them about using seat belts.[25]

Three-point restraints reduce the risk of death or serious injury by 45%.[26] Air bags reduce the risk of death by an additional 6% in drivers using seat belts.[27] Because air bags reduce the risk of death by only 14% in unbelted drivers,[27] physicians must tell their patients not to rely on air bags.

Injuries to Motorcyclists

Motorcycle deaths in the United States have been rising since 1997, reaching 3,181 in 2001.[28] The chance of death per mile when one is riding a motorcycle is 35 times higher than when one is riding in an automobile. Most deaths are caused by head injuries. Helmets reduce the risk of a fatal head injury by 27%, but only 50% of riders use helmets. Laws mandating helmet use are quite effective; the rate of helmet use rose to 95% in California after passage of a law, and the rate of head injuries dropped by 34%. Substance abuse is very common among injured motorcyclists.

Physicians should inquire about motorcycle use. They should redouble their efforts to screen for substance abuse in motorcyclists and should recommend using helmets.

Injuries to Pedestrians

Pedestrian injuries caused by motor vehicles accounted for 4,739 deaths in 2000, a 27% decrease from 1990. Children are at greatest risk for injury. Among adults, the elderly are at greatest risk, principally because of sensory deficits, locomotor disability, and inability to process simultaneous stimuli.

Injuries to Cyclists

Each year, there are approximately 900 deaths from bicycle injuries in the United States.[29] Children are at greatest risk. Head injuries account for two thirds of hospitalizations and three quarters of deaths related to bicycling. A meta-analysis of case-control studies showed that use of safety helmets reduced the risk of head injuries by 63% to 88%.[29] Helmets are effective for all ages and provide protection even in collisions with motor vehicles. Community-based education efforts have raised the rate of helmet use to 50%. Physicians should ask about bicycle use and counsel riders to use safety helmets.

INJURIES FROM FALLING

Falling is a serious health risk for older persons [*see Table 6*]. The lifetime risk of hip fracture, perhaps the most important consequence of a fall, is 40% for a 50-year-old woman. One approach to reducing the risk of hip fracture is to prevent osteoporosis. Strategies for prevention of osteoporosis include vitamin D and supplementation of dietary calcium intake,[30] drugs that increase bone mass, such as etidronate and alendronate, and estrogen replacement after menopause. In a cohort study, weight-bearing exercise, such as walking, was associated with a 40% lower risk of hip fracture in women and a 50% lower risk in men.[31] Exercise works in part by increasing bone mass and in part by reducing the likelihood of a fall. Combined interventions that included home visits, modifying home hazards, and exercise and gait programs reduced the risk of falls by 31% in a randomized clinical trial.[32] Physicians should identify patients who are at greatest risk for falling [*see Table 6*], treat osteoporosis, and link the patient to community-based programs for improving mobility and reducing hazards in the home. Hip pads are also effective; in a randomized trial, a pad worn over each hip reduced the risk of hip fracture from 46 per 1,000 patient-years to 21.3 per 1,000 patient-years ($P = 0.008$).[33]

Table 6 Risk Factors for Falls among the Elderly[32]

Prior falls	Low body mass index
Cognitive impairment	Female sex
Chronic illness	General frailty
Balance and gait impairment	Hazards in the home

INJURIES FROM FIRE

Prevention of death from fires is an example of a successful passive strategy. Smoke detectors prevent fire injury. A study in Oklahoma City measured the effects of door-to-door distribution of smoke detectors to residents of an area that had much higher rates of burn injuries than the rest of the city. The fire-injury rate declined 80%, to the same level as that in the rest of the city. The injury rate per fire also declined dramatically.[34] Physicians should inquire about smoke detectors in the home and recommend them to people who don't have them. Persons who are alcohol dependent or who smoke in bed are at high risk and need special effort.

DROWNING

In most instances of witnessed drowning, bystanders report that the victim becomes motionless while swimming or simply fails to surface after a dive. Struggle is unusual. This observation raises the possibility that many cases of drowning occur when something such as a seizure, an arrhythmia, or an injury occurs.[35]

All victims of immersion have hypoxemia. Aspirated freshwater is hypotonic and therefore rapidly absorbed by the pulmonary circulation and distributed throughout the body water compartment. Freshwater alters pulmonary surfactant and causes alveolar collapse and atelectasis. Saltwater is hypertonic and draws water into the alveoli, causing perfused but poorly ventilated alveoli and hypovolemia with concentration of electrolytes. The end result with both types of water is venous admixture and hypoxemia, often resulting in metabolic acidosis. Saltwater drowning often leads to hypovolemia as well.

The main goal of treatment is to prevent brain injury.[35] The first step is to initiate cardiopulmonary resuscitation if the victim is apneic and pulseless. The American Heart Association recommends abdominal thrusts only to clear the airway in case of suspected foreign-body aspiration or failure to respond to artificial ventilation. Supplemental oxygen is indicated as long as the patient is hypoxemic. The most effective single treatment of hypoxemia is continuous positive airway pressure (CPAP), using mechanical ventilation to expand collapsed alveoli caused by freshwater immersion.[35] Hypothermia, which often accompanies near-drowning, can protect the brain from injury by reducing its metabolic requirements when the patient is hypoxemic.

Most efforts to prevent drowning focus on children, for whom the passive strategy of requiring fencing around swimming pools is associated with reduced drowning rates. In adults, alcohol ingestion is a risk factor for drowning. The efficacy of personal flotation devices is not known. Relatively few boaters (14%) wear personal flotation devices, but this rate is

Domestic Violence Information on the Internet

Medical Resources

Family Peace Project
http://www.family.mcw.edu/d_FamilyPeace.htm

Domestic Violence: A Practical Approach for Clinicians
http://www.sfms.org/domestic.html

Legal Resources

Women's Law Initiative
http://www.womenslaw.org

American Bar Association Commission on Domestic Violence
http://www.abanet.org/domviol/home.html

similar to that of drowning victims. Physicians should ask patients whether they use boats recreationally and advise avoiding alcohol and using a personal flotation device while boating.

DOMESTIC VIOLENCE

For women especially, the home is the most dangerous place. In one large study of women in a primary care clinic, one in 20 had experienced domestic violence in the previous year, one in four had experienced it as adults, and one in three had experienced it in their lifetime.[36,37] A condition so prevalent in primary care practice demands the attention of the physician.

Among those abused in the previous year, approximately equal numbers had been abused once, two or three times, or four or more times. In this study, the definition of domestic violence was an affirmative answer to the question "Have you been hit, slapped, kicked, or otherwise physically hurt by someone?" or "Has anyone forced you to have sexual activities?" Generally, a husband, ex-husband, boyfriend, or relative is the abuser in domestic violence.

Most of the rapidly developing literature on domestic violence focuses on screening, diagnosis, and management, and there is little on preventing the first episode. Screening for domestic violence typically occurs in the office setting. Many authorities recommend that physicians routinely ask about domestic violence as part of the screening history.[37,38] Domestic violence occurs in homosexual relationships as well, so it is best to ask both men and women. Some physicians introduce the question by saying that they are now asking all their patients about domestic abuse, in view of the growing awareness of the problem. Then they ask, "At any time [or since I last saw you] has your husband [lover, partner, boyfriend] hit, kicked, threatened, or otherwise frightened you?" If the patient replies in the affirmative, the physician should gather more information, including the name of the abuser, and record it. Because many people are ashamed of their situation and their inability to break out of it, the physician should avoid any judgmental statements other than to confirm that what is being done to the patient is wrong.

In many cases, patients will not disclose an abusive relationship but their medical and social history contains clues to the true situation. Somatic symptoms that are particularly indicative (prevalence ratio > 2.5) of an abusive relationship include multiple symptoms (especially with no apparent physical cause), poor appetite, nightmares, eating binges, pain in the pelvic region, vaginal discharge, musculoskeletal injuries, and diarrhea.[36] Even more indicative are emotional symptoms such as high anxiety, severe depression, a high level of somatization, and low self-esteem; current or past use of street drugs; positive items on the CAGE questionnaire for alcohol abuse; a current or past drinking problem; a husband or partner who abuses alcohol or uses street drugs; a history of suicide attempt; and abuse as a child (prevalence ratio > 10.0 for all of these).[36,39] Pregnancy is often associated with an escalation of violence.

Some abused patients will disclose a history of abuse, but many will not. Therefore, during the physical examination, the physician should be alert to signs of injury. One expert states that a woman who presents with any injury should be considered a victim of domestic abuse until proved otherwise.[38] Trauma to the face, abdomen, breasts, or genitals is especially likely to be from domestic abuse, as are bilateral or multiple injuries, injuries in different stages of healing, and injuries that occurred well before the patient sought help. Injuries to the ulnar aspect of the elbows may occur as a woman raises her hands to protect herself during an assault.

Older persons are subject to several forms of abuse: self-neglect or caregiver neglect, emotional and psychological abuse, fiduciary exploitation, and physical abuse. Signs of elder abuse include bruising and other signs of trauma, malnutrition, volume depletion, and poor hygiene.

The physician should communicate concern and validate the patient's belief that domestic abuse is wrong. The physician should not only provide medical treatment of injuries but also talk with the patient about how to avoid serious injury during an assault, review the patient's options, and facilitate referral to community and other resources for abused partners [*see Sidebar* Domestic Violence Information on the Internet].

INJURIES FROM FIREARMS

The rate of death by firearms in the United States peaked at 39,595 in 1993, and it has since declined, reaching 28,663 in 2000.[40,41] Firearms accounted for 992 accidental deaths, 13,677 homicides, and 17,767 suicides in 1997. In two large communities, 58% of suicide victims used a firearm. Seventy percent of the suicides occurred at home. Firearms kill more teenagers than all natural causes of death combined. The Bureau of Alcohol, Tobacco and Firearms estimates that there are 192 million firearms in private hands. Firearms, often bought for protection in the home, are far more dangerous to the occupants than to an intruder. After controlling for other suicide risk factors, the odds of suicide are 1.9 times greater in homes in which there is at least one gun.[42,43] The odds of homicide are 2.2 times higher in homes in which there is a firearm.[44]

Physicians strongly support regulation of firearms and community efforts to restrict ownership.[45] Physicians also have a role to play in preventing injury from firearms.[46] They should inquire about firearms in the home and counsel owners about storing their firearms in a safe place. With the increased number of teenagers who own guns and commit homicide with guns, the need to educate parents is urgent. The American Academy of Pediatrics has developed an information kit for physicians to use in counseling parents and children. The kits, called Steps To Prevent Firearm Injury In The Home, are available without charge from the Brady Center to Prevent Gun Violence, 1225 Eye Street NW, Suite 1100, Washington, DC 20005, or they can be obtained on the Internet, at http://www.bradycenter.com/stop2/.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. McGinnis JM, Foege WH: Actual causes of death in the United States. JAMA 270: 2207, 1993

2. United States Preventive Services Task Force: Guide to Clinical Preventive Services. Williams & Wilkins, Baltimore, 1996

3. Welch HG, Albertsen PC, Nease RF, et al: Estimating treatment benefits for the elderly: the effect of competing risks. Ann Intern Med 124:577, 1996

4. Lorig K: Patient Education and Counseling. United States Preventive Services Task Force. Williams & Wilkins, Baltimore, 1996.

5. American Academy of Pediatrics: Tobacco's toll: implications for the pediatrician. Committee on Substance Abuse. Pediatrics 107:794, 2001

6. Bartecchi CE, MacKenzie TD, Schrier RW: The human costs of tobacco use. N Engl J Med 330:907, 1994

7. Diagnostic and Statistical Manual of Mental Disorders, 4th ed.: DSM-IV. American Psychiatric Association, Washington, DC, 1994

8. Kawachi I, Colditz GA, Stampfer MJ, et al: Smoking cessation and decreased risk of stroke in women. JAMA 269:232, 1993

9. Dietrich AJ, O'Connor GT, Keller A, et al: Cancer: improving early detection and prevention: a community practice randomised trial. BMJ 304:687, 1992

10. Djikstra A, Roijackers J, De Vries H: Smokers in four stages of readiness to change. Addict Behav 23:339, 1998

11. Henningfield JE: Nicotine medications for smoking cessation. N Engl J Med 333:1196, 1995

12. Silagy C, Lancaster T, Stead L, et al: Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev (4):CD000146, 2002

13. Flegal KM, Troiano RP, Pamuk ER, et al: The influence of smoking cessation on the prevalence of overweight in the United States. N Engl J Med 333:1165, 1995

14. Heatherton TF, Kozlowski LT, Frecker RC, et al: The Fagerstrom test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. Br J Addict 86:1119, 1991

15. Holm KJ, Spencer CM: Bupropion: a review of its use in the management of smoking cessation. Drugs 59:1007, 2000

16. Hurt RD, Sachs DPL, Glover ED, et al: A comparison of sustained-release bupropion and placebo for smoking cessation. N Engl J Med 337:1195, 1997

17. Institute of Medicine. Broadening the base of treatment for alcohol problems. National Academy Press, Washington, DC, 1990

18. Bien TH, Miller WR, Tonigan JS: Brief interventions for alcohol problems: a review. Addiction 88:315, 1993

19. Brief interventions and alcohol use. Effective Health Care, Bulletin No. 7. Nuffield Institute for Health Care, University of Leeds, Leeds, England, 1993

20. Kahan M, Wilson L, Becker L: Effectiveness of physician-based interventions with problem drinkers: a review. Can Med Assoc J 152:851, 1995

21. Wilk AI, Jensen NM, Havighurst TC: Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J Gen Intern Med 12:274, 1997

22. Walsh DC, Hingson RW, Merrigan DM, et al: A randomized trial of treatment options for alcohol-abusing workers. N Engl J Med 325:775, 1991

23. Fiellin DA, O'Connor PG: Clinical practice: office-based treatment of opioid-dependent patients. N Engl J Med 347:817, 2002

23. Rivara FP, Grossman DC, Cummins P: Medical progress: injury prevention. N Engl J Med 337: 536, 1997

24. Traffic safety facts 2000. A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system (DOT HS 809 337). National Highway Traffic Safety Administration, US Department of Transportation. December, 2001
http://www-fars.nhtsa.dot.gov/pubs/1.pdf

25. Hunt DR, Lowenstein SR, Badgett RG, et al: Safety belt nonuse by internal medicine patients: a missed opportunity in preventive medicine. Am J Med 98:343, 1995

26. Effectiveness of occupant protection systems and their use: third report to Congress. National Highway Traffic Safety Administration, US Department of Transportation. Washington, DC, December 1996
http://www.nhtsa.dot.gov/people/injury/airbags/208con2e.html

27. Kahane CJ: Fatality reduction by airbags: analyses of accident data through early 1996 (DOT HS 808 470). National Highway Traffic Safety Administration, US Department of Transportation. Washington, DC, August 1996
www.nhtsa.dot.gov/cars/rules/regrev/evaluate/808470.html

28. The National Highway Traffic Safety Administration Motorcycle Safety Program. National Highway Traffic Safety Administration, US Department of Transportation, January 2003
http://www.nhtsa.dot.gov/people/injury/pedbimot/motorcycle/motorcycle03

29. Thompson DC, Rivara FP, Thompson R: Helmets for preventing head and facial injuries in bicyclists. Cochrane Database Syst Rev (2):CD001855, 2000

30. Bischoff HA, Stahelin HB, Dick W, et al: Effects of vitamin D and calcium supplementation on falls: a randomized controlled trial. J Bone Miner Res 18:343, 2003

31. Paganini-Hill A, Chao A, Ross RK, et al: Exercise and other risk factors in the prevention of hip fracture: the Leisure World Study. Epidemiology 2:16, 1991

32. Tinetti ME, Baker DI, McAvay G, et al: A multi-factorial intervention to reduce the risk of falling among elderly people living in the community. N Engl J Med 331:821, 1994

33. Kannus P, Parkkari J, Niemi S, et al: Prevention of hip fracture in elderly people with use of a hip protector. N Engl J Med 343:1506, 2000

34. Mallonee S, Istre GR, Rosenberg M, et al: Surveillance and prevention of residential-fire injuries. N Engl J Med 335:27, 1996

35. Bierens JJ, Knape JT, Gelissen HP: Drowning. Curr Opin Crit Care 8:578, 2002

36. McCauley J, Kern DE, Kolodner K, et al: The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. Ann Intern Med 123:737, 1995

37. Barrier PA: Domestic violence. Mayo Clin Proc 73:271, 1998

38. Alpert EJ: Violence in intimate relationships and the practicing internist: new "disease" or new agenda? Ann Intern Med 123:774, 1995

39. Kyriacou DN, McCabe F, Anglin D, et al: Emergency department–based study of risk factors for acute injury from domestic violence against women. Ann Emerg Med 31:502, 1998

40. Nonfatal and fatal firearm-related injuries—United States, 1993–1997. MMWR Morb Mortal Wkly Rep 48:1029, 1999

41. Deaths: Final Data for 2000. National vital statistics reports, vol 50, no 15. National Center for Health Statistics, Hyattsville, Maryland, 2002

42. Miller M, Azrael D, Hemenway D: Household firearm ownership and suicide rates in the United States. Epidemiology 13:517, 2002

43. Cummings P, Koepsell TD: Does owning a firearm increase or decrease risk of death? JAMA 280:471, 1998

44. Kellerman AL, River FP, Rushforth NB, et al: Gun ownership as a risk factor for homicide in the home. N Engl J Med 329:1084, 1993

45. Cassel CK, Nelson EA, Smith TW, et al: Internists' and surgeons' attitudes toward guns and firearm injury prevention. Ann Intern Med 128:224, 1998

46. American College of Physicians: Firearm injury prevention. Ann Intern Med 128:236, 1998

Acknowledgment

Figure 1 Marcia Kammerer. Adapted from Life table from the Vital Statistics of the United States, 1999.

IV DIET AND EXERCISE

Harvey B. Simon, m.d.

Many chronic diseases result from unhealthful eating and a sedentary lifestyle. Poor nutrition and inadequate exercise substantially increase the risk of such maladies as coronary artery disease, hypertension, stroke, diabetes, obesity, osteoporosis, and certain cancers and account for about 300,000 deaths in the United States each year.[1] Dietary factors also contribute to cholelithiasis, hemorrhoids, hernias, constipation, irritable bowel syndrome, and diverticulosis. A rigorous program that combines a low-fat, high-fiber diet with daily exercise can produce dramatic improvement in cardiovascular risk factors in as little as 3 weeks' time.[2]

Diet

In the 20th century, the average American diet shifted from one based on fresh, minimally processed vegetable foods to one based on animal products and highly refined, processed foods. As a result, Americans now consume far more calories, fat, cholesterol, refined sugar, animal protein, sodium, and alcohol and far less fiber and starch than is healthful.

In the United States, two out of every three adults are overweight (body mass index [BMI] of 25 to 30) or obese (BMI > 30), compared with fewer than one in four in the early 1960s.[3] The consequences include a substantial decrease in life expectancy and an increase in morbidity similar in magnitude to the burden imposed by smoking.[4]

Obesity is a complex, multifactorial disorder, but an element common to all cases is a positive energy balance in which more calories are consumed than expended. Excess calories are stored in body fat; each pound of adipose tissue contains 3,500 calories. Weight loss is accomplished only by achieving a negative energy balance.

ENERGY

Genetic, metabolic, and behavioral variables make it difficult to predict an individual's caloric requirements with precision. However, physicians can provide estimates: sedentary adults require about 30 cal/kg/day to maintain body weight; moderately active adults require 35 cal/kg/day; and very active adults require 40 cal/kg/day. On average, therefore, a 70 kg (154 lb) person can expect to maintain body weight by consuming 2,100 to 2,800 calories daily.

Although any source of dietary energy, including carbohydrate, protein, and alcohol, can be converted in the body to fatty acids and cholesterol, the caloric value of foods varies considerably; for example, fat provides 9 cal/g and alcohol provides 7 cal/g, but protein and carbohydrates each provide only 4 cal/g. Patients with excess body fat should be encouraged to shift from high-fat, calorie-dense foods to low-fat, less-caloric foods. Even in a well-balanced diet, it is important to control food-portion size, which has increased dramatically[5] in concert with the obesity epidemic in the United States. As an example, to lose 1 lb a week, patients must consume 500 fewer calories than they expend each day; in almost all cases, sustained weight loss requires both an energy-restricted diet and regular vigorous exercise.

FAT AND CHOLESTEROL

Structure

Most dietary lipids are triglycerides, in which three fatty acids are joined to one glycerol molecule. At the core of every fatty acid is a chain of carbon atoms with a methyl group at one end and a carboxyl group at the other [*see Figure 1*]. The biologic properties of fatty acids are determined by the presence or absence of double bonds between carbon atoms, the number and location of the double bonds, and the configuration of the molecules.

Most of the fatty acids in foods are composed of an even number of carbon atoms, generally in chains of 12 to 22 atoms. The number of double bonds between carbon atoms determines the saturation of fats. Fatty acids with no double bonds are fully saturated; they have no room for additional hydrogen atoms. Fatty acids with one double bond are monounsaturated, and those with two or more double bonds are polyunsaturated.

Fatty acids contain zero to six double bonds, where additional hydrogen atoms can be attached. The location of the double bonds is of great physiologic importance; an unsaturated fatty acid's group (i.e., omega-3, omega-6, or omega-9) is determined by the position of the double bond closest to the methyl group. In omega-3 fatty acids, for example, three carbon atoms lie between the methyl end of the chain and the first double bond.

Most of the fatty acids in natural foods are in the curved, or *cis*, configuration. When hydrogen is added back to unsaturated fats during food manufacturing, however, the molecules assume a straightened, or *trans*, configuration [*see Figure 1*].

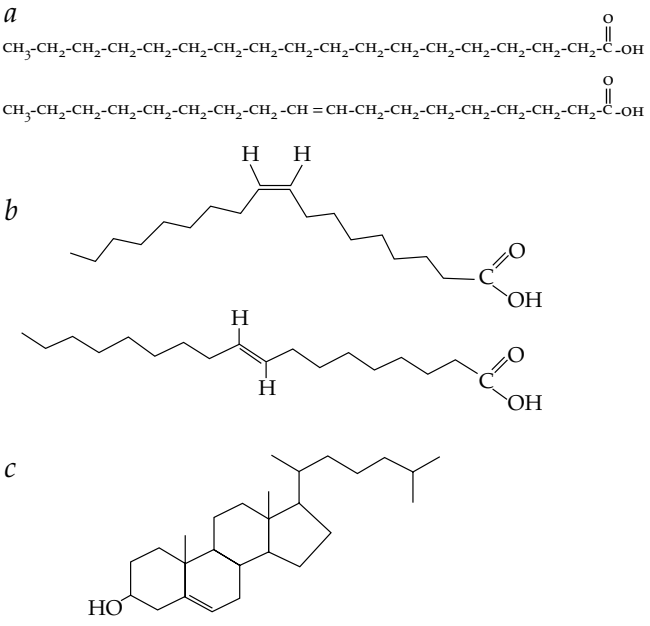


Figure 1 **The structure of fat and cholesterol is shown. (*a*) Stearic acid (top) is a saturated fatty acid. Oleic acid (bottom) is a monounsaturated omega-9 fatty acid. (*b*) Oleic acid (top) displays a *cis* double bond. Elaidic acid (bottom) displays a *trans* double bond. (*c*) Cholesterol has a structure similar to that of fatty acids.**

Cholesterol is a waxy, fatlike molecule that is present in the membranes of all animal cells but is absent from plant cells. Although cholesterol is a sterol rather than a true fat, its metabolism is intimately linked to the dietary intake of fatty acids.

Effects on Blood Lipids and Cardiovascular Risk

Although all fats have the same caloric value (9 cal/g), their effects on human health vary greatly, largely because of their disparate effects on blood cholesterol levels. Saturated fats stimulate hepatic cholesterol production, thus increasing blood cholesterol levels. Of the four saturated fatty acids that predominate in the American diet, myristic acid (14 carbons) has the most potent hypercholesterolemic effect, followed by palmitic acid (16 carbons) and lauric acid (12 carbons). Stearic acid (18 carbons) has little effect on blood cholesterol levels. Evidence strongly suggests that the degree to which saturated fat and cholesterol intake increase the risk of coronary artery disease depends on their effects on blood cholesterol concentration.[6]

Unsaturated fatty acids are generally derived from vegetable and marine sources; they are often called oils rather than fats because they are liquid at room temperature. When monounsaturated or polyunsaturated fatty acids are substituted for saturated fats, blood cholesterol levels fall. Neither type of unsaturated fat, however, has a direct ability to lower low-density lipoprotein (LDL) cholesterol or raise high-density lipoprotein (HDL) cholesterol levels. Although monounsaturated and polyunsaturated fats have a similar, generally neutral, effect on blood cholesterol levels, monounsaturated fats are less susceptible to oxidation and may therefore be less atherogenic. Omega-3 polyunsaturated fatty acids in particular have been shown to have a cardioprotective effect.

Consumption of omega-3 fatty acids is inversely related to the incidence of atherosclerosis and the risk of sudden death[7,8] and stroke.[9] In high doses, omega-3 fatty acids may reduce blood triglyceride levels, but in dietary amounts, they have little effect on blood lipids. Even in modest amounts, however, omega-3 fatty acids reduce platelet aggregation, impairing thrombogenesis. They may also have antiarrhythmic[10] and plaque-stabilizing properties.[11] Diets high in α-linolenic acid appear to reduce the risk of coronary artery diseases[12-14] and stroke.

Like saturated fats, *trans*-fatty acids increase blood LDL cholesterol levels; unlike saturated fats, *trans*-fatty acids reduce HDL cholesterol levels, making *trans*-fatty acids even more detrimental.[15] Diets high in *trans*-fatty acids have been associated with an increased risk of atherosclerosis and coronary events.

Dietary cholesterol increases blood LDL cholesterol levels but has a less potent hypercholesterolemic effect than saturated fat. Diets high in cholesterol are associated with an increased risk of coronary artery disease independent of their effects on blood cholesterol levels,[12] reinforcing the importance of reducing cholesterol intake.

Fat and Health

A high intake of saturated fat from animal sources appears to increase the risk of colon cancer[16] and prostate cancer[17] but not breast cancer.[18] However, some dietary fat is essential. For example, omega-3 and omega-6 fatty acids cannot be synthesized endogenously and therefore must be obtained from food. Dietary fat is required for the absorption of fat-soluble vitamins. Lipids are essential components of cell membranes and steroid hormones; adipose tissue is the body's major energy depot, and it provides insulation against heat loss. As little as 15 to 25 g of dietary fat a day can provide essential physiologic functions.

Table 1 Recommended Daily Intake of Fat and Other Nutrients*

Nutrient	*Recommended Intake*
Total fat	20%–35% of total calories
Saturated fat†	< 7% of total calories
Polyunsaturated fat	≤ 10% of total calories
Monounsaturated fat	≤ 20% of total calories
Cholesterol	< 300 mg/day 50%–60% of total calories
Carbohydrate††	≥ 25 g/day
Fiber	15% of total calories
Protein	—
Total calories§	Balance energy intake and expenditure to maintain desirable body weight and prevent weight gain

*See reference 24
†*Trans*-fatty acids, which raise low-density lipoprotein (LDL) and lower high-density lipoprotein (HDL) cholesterol, should also be kept at low levels.
††Carbohydrates should be derived predominantly from foods rich in complex carbohydrates, including grains, especially whole grains, fruits, and vegetables. Simple sugars should contribute no more than 25% of total calories.
§Daily energy expenditure should include at least moderate physical activity (consuming 200 kcal/day).

Dietary Recommendations

The American Heart Association (AHA) dietary guidelines[19] for healthy adults suggest that no more than 30% of calories should come from fat, with less than 10% coming from saturated fat and the remainder coming from unsaturated fat in vegetables, fish, legumes, and nuts. The AHA guidelines also specify consumption of less than 300 mg of cholesterol a day. Patients with atherosclerosis or diabetes and persons who are hyperlipidemic or obese should follow more stringent limits, such as a saturated-fat intake of no more than 7% of daily calories, with a corresponding decrease in cholesterol consumption to less than 200 mg a day. In some persons, very low fat diets providing 15% to 22% of calories from fat can reduce blood HDL levels and produce other adverse effects,[20,21] but in carefully monitored high-risk persons, diets with about 10% fat and virtually no cholesterol have been beneficial.[22] Although reductions in total fat intake can help reduce body fat and serum cholesterol levels, the risk of coronary artery disease may depend more on the type of fat in the diet; saturated fats and *trans*-fatty acids are the most atherogenic, whereas monounsaturated and omega-3 fatty acids are the most desirable [*see Table 1*].[7,12-15,23,24]

Food labels list the fat, saturated fat, and cholesterol contents of packaged foods. They will soon be required to list *trans*-fatty acids; until then, patients should be advised to check the ingredients list at the bottom of the label for the presence of partially hydrogenated vegetable oils.

CARBOHYDRATES

Carbohydrates are a vital source of energy for metabolic processes. They are also vital constituents of nucleic acids, glycoproteins, and cell membranes.

Plants are the principal dietary sources of carbohydrates. The only important carbohydrates that originate from animal sourc-

Table 2 Types of Dietary Fiber and Representative Food Sources

Fiber Type	*Food Sources*
Gums*	Oats, beans, legumes, guar
Pectin*	Apples, citrus fruits, soybeans, cauliflower, squash, cabbage, carrots, green beans, potatoes
Mucilage*	Psyllium
Hemicellulose*†	Barley, wheat bran and whole grains, brussels sprouts, beet roots
Lignin†	Green beans, strawberries, peaches, pears, radishes
Cellulose†	Root vegetables, cabbage, wheat and corn, peas, beans, broccoli, peppers, apples

*Soluble fiber †Insoluble fiber

es are the lactose in milk and the glycogen in muscle and liver. Carbohydrate-rich foods contain varying amounts of simple and complex carbohydrates. Simple carbohydrates include monosaccharides such as glucose, fructose, and galactose and disaccharides such as sucrose (table sugar), maltose, and lactose. Complex carbohydrates include polysaccharides (e.g., starch and glycogen that can be digested into sugars by intestinal enzymes) and fiber (i.e., high-molecular-weight carbohydrates that cannot be split into sugars by human intestinal enzymes). Sugars, starches, and glycogen provide 4 cal/g; because fiber is indigestible, it has no caloric value.

Carbohydrates contribute about 50% of the calories in the average American diet—half from sugar and half from complex carbohydrates. Because sugars are more rapidly absorbed, they have a higher glycemic index than starches. In addition to provoking higher insulin levels, carbohydrates with a high glycemic index appear to reduce HDL cholesterol levels and may increase the risk of coronary artery disease.[25] Processed foods containing simple sugars are often calorie dense, whereas foods that are rich in complex carbohydrates provide vitamins, trace minerals, and other valuable nutrients. A healthful diet should provide 55% to 65% of calories from complex carbohydrates found in fresh fruits and vegetables, legumes, and whole grains.[19,23]

DIETARY FIBER

Dietary fiber is a heterogeneous mix of very long chain branched carbohydrates that resist digestion by human intestinal enzymes because of the ways their monosaccharide components are linked to one another. Fiber is found only in plants, particularly in the bran of whole grains, in the stems and leaves of vegetables, and in fruits, seeds, and nuts. The two general categories of dietary fiber are soluble and insoluble.

Soluble fiber delays gastric emptying, which produces a sensation of satiety, and slows the absorption of digestible carbohydrates, which reduces insulin levels. Soluble fiber also lowers blood cholesterol levels, probably by inhibiting bile acid and nutrient absorption in the small intestine and by promoting bile acid sequestration by colonic bacteria.[26] Because soluble fiber is metabolized by these bacteria, it has little effect on fecal bulk. In contrast, insoluble fiber increases the water content and bulk of feces and shortens intestinal transit time [*see Table 2*].

Diets that are high in fiber also tend to be low in fat. Such diets have been associated with a reduced risk of intestinal disorders, including constipation, irritable bowel syndrome, cholelithiasis, hemorrhoids, and diverticulosis. Although earlier data are mixed, new studies suggest that a high intake of fiber can substantially reduce the risk of colorectal cancer.[27,28] In addition, a dietary pattern that includes a high intake of fruits, vegetables, legumes, fish, poultry, and whole grains but little red meat, processed meats, sweets, and refined grains appears protective.[29,30] A high intake of fiber is associated with a reduced risk of diabetes and, in patients with diabetes,[31] improved glycemic control and decreased blood lipids[32]; it is also associated with a reduced risk of obesity[33] and coronary artery disease[34,35] and a lower all-cause mortality. A healthful diet should contain at least 25 to 30 g of fiber a day, including substantial amounts of soluble fiber.

PROTEINS

Unlike reserves of fat (which is stored in large amounts as triglyceride in adipose tissue) and reserves of carbohydrate (which is stored in small amounts as glycogen in liver and muscle), there are no endogenous reserves of amino acids or protein; all the proteins in the body are serving a structural or metabolic function. As a result, bodily function can be impaired if proteins are catabolized because of energy deficiency, wasting diseases, or dietary protein intake that is not sufficient to replace protein losses.

All proteins in human cells are continuously catabolized and resynthesized. In a healthy 70 kg adult, about 280 g of protein is degraded and replaced daily. In addition, about 30 g of protein is lost externally through the urine (urea), feces, and skin.

In healthy adults, daily protein losses can be fully replaced by as little as 0.4 g/kg. Because not all dietary proteins are fully digestible, the recommended dietary allowance (RDA) of protein for healthy adults is 0.8 g/kg. People who exercise strenuously on a regular basis may benefit from extra protein to maintain muscle mass; a daily intake of about 1 g/kg has been recommended for athletes. Women who are pregnant or lactating require up to 30 g/day in addition to their basal requirements. To support growth, children should consume 2 g/kg/day.

A healthful diet should provide 10% to 15% of its calories from protein.[19] For healthy, nonpregnant women, an intake of 44 to 50 g/day of protein is required, and for men, an intake of 45 to 63 g/day of protein is needed. Although excessive protein intake has not been proved to be harmful, there are several potential disadvantages to a very high protein intake. The protein in foods derived from animals is often accompanied by large amounts of fat. In the body, excessive protein can be transaminated to carbohydrate, adding to the energy surplus responsible for obesity. When excess protein is eliminated from the body as urinary nitrogen, it is often accompanied by increased urinary calcium, perhaps increasing the risk of nephrolithiasis and osteoporosis. Because nitrogen is excreted in the urine, an increased protein intake is associated with an increase in renal plasma flow and glomerular filtration rates and, eventually, with increased renal size. In some animal models, increased dietary protein is associated with accelerated renal aging; and in humans with kidney disease, high dietary protein intake is associated with more rapid disease progression.[36] On the other hand, high dietary protein intake appears linked to somewhat reduced blood pressure readings,[37] possibly because of increased urinary sodium losses, and protein supplements may be beneficial for patients with acute or chronic illnesses.[38]

The thousands of proteins in the human body are synthesized from just 21 amino acids. Most amino acids can be synthe-

Table 3 The Vitamins

Vitamin	*Functions*	*Deficiency Effects*	*Toxic Effects*	*Sources*	*RDA for Adults*
A (retinol, retinoic acid)	Vision, epithelial integrity; possible protection against epithelial cancers and atherosclerosis	Night blindness; increased susceptibility to infection	Teratogenicity, hepatotoxicity, cerebral edema, desquamation; yellowish skin discoloration by carotenoids	Liver, dairy products, eggs; dark-green and yellow-orange vegetables (carotenoids)	Men, 5,000 IU or 1,000 RE; women, 4,000 IU or 800 RE
B_1 (thiamine)	Metabolism of carbohydrates, alcohol, and branched-chain amino acids	Beriberi, Wernicke-Korsakoff syndrome	None	Grains, legumes, nuts, poultry, meat	Men 19–50 yr: 1.5 mg; men > 50 yr: 1.2 mg; women 19–50 yr: 1.1 mg; women > 50 yr: 1.0 mg
B_2 (riboflavin)	Cellular oxidation-reduction reactions	Stomatitis, dermatitis, anemia	None	Grains, dairy products, meat, eggs, dark-green vegetables	Men 19–50 yr: 1.7 mg; men > 50 yr: 1.4 mg; women 19–50 yr: 1.3 mg; women > 50 yr: 1.2 mg
B_3 (niacin, nicotinic acid)	Oxidative metabolism; reduces LDL cholesterol; increases HDL cholesterol	Pellagra	Flushing, headaches, pruritus, hyperglycemia, hyperuricemia, hepatotoxicity	Meat, poultry, fish, grains, peanuts; synthesized from tryptophan in foods	Men 19–50 yr: 19 mg; men > 50 yr: 15 mg; women 19–50 yr: 15 mg; women > 50 yr: 13 mg
B_6 (pyridoxine)	Amino acid metabolism and heme synthesis; neuronal excitability; reduces blood homocysteine levels	Anemia, cheilosis, dermatitis	Neurotoxicity	Meat, poultry, fish, grains, soybeans, bananas, nuts	2 mg
B_{12} (cobalamin)	DNA synthesis (with folate); myelin synthesis (without folate); reduces blood homocysteine levels	Megaloblastic anemia, neuropathies	None	Meat (especially liver), poultry, fish, dairy products	2–4 μg
Folic acid	DNA synthesis (with B_{12}); reduces blood homocysteine levels	Megaloblastic anemia, birth defects	None	Vegetables, legumes, grains, fruit, poultry, meat	400 μg
Biotin	Metabolic processes	Rare	None	Many foods	30–100 μg
Pantothenic acid	Metabolic processes	Rare	None	Many foods	4–7 mg
C (ascorbic acid)	Collagen synthesis; possible protection against certain neoplasms	Scurvy	Nephrolithiasis, diarrhea	Fruits, green vegetables, potatoes, cereals	Men, 90 mg; women, 75 mg
D (calciferol)	Intestinal calcium absorption	Osteomalacia and rickets	Hypercalcemia	Fortified dairy products, fatty fish, egg yolks, liver	< 50 yr, 200 IU; 50–70 yr, 400 IU; > 70 yr, 600 IU
E (α-tocopherol)	Reduces peroxidation of fatty acids: possible protection against atherosclerosis	Rare	Antagonism of vitamin K, possible headaches	Vegetable oils, wheat germ, nuts, broccoli	15 mg
K	Synthesis of clotting factors VII, IX, X, and possibly V	Hemorrhagic diathesis	None	Leafy green vegetables (K_1), intestinal bacteria (K_2)	Men, 120 μg; women, 90 μg

HDL—high-density lipoprotein IU—international units LDL—low-density lipoprotein RDA—recommended dietary allowance RE—retinol equivalents

sized endogenously, but nine cannot. Not all dietary proteins contain all nine essential amino acids; in particular, vegetable proteins may be incomplete. However, by eating a varied diet with foods that contain a mix of proteins, even strict vegetarians can obtain all the amino acids they need.

VITAMIN AND MINERAL CONSUMPTION

Vitamins

Vitamins are either fat soluble or water soluble. Vitamins A, D, E, and K are fat soluble. They are found in fatty foods and are absorbed, transported, and stored with fat. Because excretion is minimal and storage in fat is abundant, deficiencies of fat-soluble vitamins are rare, but toxic amounts can accumulate if intake is excessive. Vitamin C and the B-complex group are water soluble; they are absorbed in the intestine, bound to transport proteins, and excreted in the urine. Because storage is minimal, water-soluble vitamins should be ingested regularly, and except for large doses of B_3 and B_6, toxicity is rare [*see Table 3*].

Although there is great disparity between popular beliefs about vitamins and their known physiologic effects, new medical information may narrow the gap. First, it is becoming clear that many persons in the United States, particularly the elderly and the poor, do not consume adequate amounts of vitamin-rich foods. Second, laboratory and animal experiments demonstrate that antioxidant vitamins can retard atherogenesis and suggest that antioxidants may lower the risk of carcinogenesis. Indeed, many epidemiologic and observational studies have demonstrated an association between a low dietary intake or low plasma levels of antioxidants and an increased risk of ath-

erosclerosis and certain cancers. Similarly, studies have linked low levels of folic acid, vitamin B_6, and vitamin B_{12} with elevated blood homocysteine levels and increased cardiovascular risks.[39] People who consume multivitamins appear to have a reduced risk of coronary artery disease[40] and colon cancer[41]; protection may be attributed principally to folic acid. However, with the exception of the Cambridge Heart Antioxidant Study, which demonstrated the efficacy of vitamin E in reducing the risk of myocardial infarction in patients with coronary artery disease, randomized trials have not demonstrated benefit from vitamin supplements.[42] Moreover, β-carotene supplements actually appear to increase the risk of lung cancer in smokers,[43] and hypervitaminosis A is linked to an increased risk of fractures.[44] It is clear that additional studies are required to clarify the impact of vitamins on health.

Women of childbearing age, the elderly, and people with suboptimal nutrition should take a single multivitamin tablet daily; others may benefit as well.[45] Strict vegetarians should take vitamin B_{12} in the recommended daily amount (2–4 μg); because many people older than 60 years have atrophic gastritis and cannot absorb B_{12} bound to food protein, they may also benefit from supplementary B_{12}. Multivitamin supplements may also be necessary to avert vitamin D deficiencies, particularly in the elderly.[46] A supplement that combines antioxidants with zinc can slow the progression of age-related macular degeneration.[47] Use of so-called megadose vitamins should be discouraged. Expensive brand-name and so-called all-natural preparations are no more effective than reputable generic preparations. In any case, vitamin supplements should never be used as a substitute for a balanced healthful diet that provides abundant amounts of vitamin-rich foods.

Minerals

Although minerals are chemically the simplest of nutrients, their roles in metabolism and health are complex. At least 16 minerals are essential for health [*see Table 4*]; 10 are classified as trace elements because only small amounts are required. Other minerals, such as boron, nickel, vanadium, and silicon, have been shown to be essential in various animal studies but have not been found to be necessary for humans. Many persons in the United States consume too little of some minerals (e.g., calcium and iron) or too much of others (e.g., sodium).

Table 4 Essential Minerals and Trace Elements

Minerals and Elements	*RDA/ESADDI for Healthy Individuals*
Macrominerals	
Calcium	1,000 mg before age 50; 1,200 mg after age 50
Phosphorus	800 mg
Magnesium	Men, 350 mg; women, 280 mg
Sodium	1,100–3,300 mg
Potassium	1,875–5,625 mg
Chloride	1,700–5,100 mg
Trace elements	
Iron	Men and postmenopausal women, 8 mg; premenopausal women, 18 mg; pregnant women, 27 mg
Zinc	Men, 15 mg; women, 12 mg
Iodine	150 μg
Copper	1.5–3.0 mg
Manganese	2–5 mg
Fluoride	1.5–4.0 mg
Chromium	Men, 35 μg; women, 25 μg
Molybdenum	75–250 μg
Selenium	55 μg
Cobalt	Required in small amount as a component of vitamin B_{12}

ESADDI—estimated safe and adequate daily dietary intake RDA—recommended dietary allowance

Sodium The body can conserve sodium so effectively that only small amounts are required in the diet. The Food and Nutrition Board of the National Academy of Science estimates that an intake of no more than 500 mg of sodium a day is needed for health; the average American diet contains more than 4,000 mg a day.

Population studies have demonstrated conclusively that a high sodium intake increases blood pressure, especially in older people.[48] The Dietary Approaches to Stop Hypertension (DASH) trial demonstrated that reduction of sodium intake from high amounts to moderate amounts will result in lower blood pressures and that further reductions in sodium intake will produce additional benefits.[49] When combined with other elements of the DASH diet (increased consumption of fruits, vegetables, whole grains, and low-fat dairy products, along with decreased consumption of saturated fat and sugar), sodium restriction can lower systolic blood pressure by an average of 7.1 mm Hg in normotensive persons and 11.5 mm Hg in patients with hypertension. Hence, reductions in dietary sodium could substantially reduce the risk of stroke and coronary artery disease. A high sodium intake also increases urinary calcium excretion, which increases the risk of osteoporosis.

There is no RDA for sodium, and additional controlled clinical trials will be needed to provide conclusive evidence that sodium restriction is beneficial to normotensive persons. Pending such information, the AHA recommends that daily consumption of sodium not exceed 2,400 mg,[19] and the National Academy of Science proposes a 2,000 mg maximum. Patients with illnesses such as hypertension, congestive heart failure, cirrhosis, and nephrotic syndrome may benefit from substantially lower sodium intakes.

About 80% of dietary sodium comes from processed foods. Physicians should review these hidden sources of salt with patients who would benefit from sodium restriction.

Calcium A high intake of calcium, either from dairy products or supplements,[50] improves bone density. Dietary calcium intake is inversely related to blood pressure[51] and to the risk of stroke[51,52]; however, calcium supplements produce only small reductions in systolic blood pressure.[53] Calcium supplements appear to reduce the risk of colorectal adenomas[54] but may increase the risk of prostate cancer.[55,56]

At present, fewer than 50% of persons in the United States consume the RDA of calcium [*see Table 4*]. Persons who do not consume enough calcium from foods should consider a supplement such as calcium carbonate or calcium citrate. High-calcium diets do not increase the risk of nephrolithiasis,[57] but prolonged overdoses of supplements may produce hypercalcemia (milk-alkali syndrome) or nephrolithiasis.

Iron Iron deficiency is the most common cause of anemia. In the United States, 9% to 11% of women of childbearing age are iron deficient and 2% to 5% have iron deficiency anemia, but only 1% of men are iron deficient. Routine administration of iron

supplements is recommended only for infants and pregnant women[58]; dietary sources should provide adequate amounts of iron for other healthy people.

A high intake of iron is harmful for patients with hemochromatosis and for others at risk for iron overload. A Finnish study linked high iron levels to cardiac risk.[59] However, studies in the United States have not confirmed these observations, and one study indicated a possible inverse association between iron stores and mortality from cardiovascular disease and other causes.[59]

Potassium Dietary potassium is inversely related to blood pressure and to stroke mortality in hypertensive men.[60] Although potassium supplements may assist in the treatment of hypertension,[61] current data do not justify the routine use of potassium supplements. Physicians should encourage a high dietary potassium intake in most individuals,[19,49] but low-potassium diets may be necessary for patients with renal disease or other conditions that cause hyperkalemia.

Selenium Selenium is a cofactor of the free radical scavenger enzyme glutathione peroxidase. A randomized clinical trial reported that selenium supplements of 200 μg/day appear to reduce mortality from various cancers.[62] Selenium levels have been inversely associated with mortality from prostate cancer[63] and gastroesophageal malignancies.[64] These data, however, do not yet support the routine use of selenium supplements, which can be toxic in high doses. Selenium is present in many foods, including tomatoes, poultry, shellfish, garlic, meat, egg yolks, and grains grown in selenium-rich soil.

Chromium Although chromium plays a role in glucose metabolism, there is no scientific basis for the claims that chromium supplements contribute to weight loss or increased energy. Chromium supplements may be beneficial for persons with low HDL cholesterol levels, but more study is needed. Dietary sources of chromium include brewer's yeast, whole grains, legumes, peanuts, and meats.

Magnesium Magnesium deficiency is common in diabetics, alcoholics, patients who take diuretics, and hospitalized patients. Persons with hypomagnesemia may require magnesium supplements, but others can rely on foods such as green vegetables, whole grains, bananas, apricots, legumes, nuts, soybeans, and seafood to provide magnesium.

WATER AND FOOD CONSUMPTION

Water

On average, adults consume about 2 L/day of water, with two thirds coming from beverages and the remainder coming from food. Healthy people have no need to track their water intake. Patients with conditions such as nephrolithiasis and urinary tract infections may benefit from consciously increasing their fluid intake; patients who are at risk for hyponatremia should restrict their water consumption.

Foods

Fruits and vegetables Fruits and vegetables provide many desirable nutrients, including complex carbohydrates, fiber, vitamins, and minerals. Deep-green and yellow-orange vegetables may be particularly beneficial because of their carotenoids, and citrus fruits may be valuable because of their vitamin C, soluble fiber, and potassium. Cruciferous vegetables, such as cabbage, may reduce the risk of certain cancers. Vegetables and fruits are low in sodium and calories; none contain cholesterol, and only coconut, palm oil, and cocoa butter contain saturated fat.

The findings of case-control and cohort studies strongly suggest that the consumption of fruits and vegetables is inversely related to the risk of coronary artery disease,[65] stroke,[66] malignancies of the respiratory and digestive tracts,[29] chronic obstructive lung disease,[67] and all-cause mortality.[68] A dietary-intervention trial demonstrated that a diet rich in vegetables, fruits, and low-fat dairy products can substantially reduce blood pressure.[49] The United States Department of Agriculture's Dietary Guidelines for Americans recommends eating two to four servings of fruit and three to five servings of vegetables a day; at present, only 35% of women and 19% of men meet these standards.[69]

Legumes Often neglected in the Western diet, legumes (beans, peas, and lentils) are rich in complex carbohydrates with low glycemic indices, iron, and B vitamins. Legumes are an excellent source of dietary fiber, including soluble fiber that can reduce blood cholesterol levels. Because of their high protein content, legumes are an excellent meat substitute. Soy protein can reduce blood cholesterol levels, and soy intake is inversely related to the risk of prostate and breast cancers.

Legumes can increase intestinal gas, causing bloating, flatulence, and cramps. Distress can be minimized by use of the nonprescription α-galactosidase preparation Beano.

Grains The seed-bearing fruits of grains, called kernels, consist of three layers: the inner germ, which contains vitamins and polyunsaturated fats; the middle endosperm, which contains complex carbohydrates; and the outer bran, which contains dietary fiber. Because milling removes the bran and endosperm, whole grains are nutritionally superior to refined grain; whole-grain consumption is inversely related to the risk of coronary artery disease[70] and stroke.[71] Whole-grain flour can be used to make cereals, baked goods, and even pasta. Whole grains such as brown rice, couscous, and yellow cornmeal (polenta) are easily prepared and healthful side dishes. Oats and barley contain soluble fiber that can lower blood cholesterol levels.

Meat and poultry Although meat is a source of protein, vitamins, and iron and other minerals, its high content of saturated fat, cholesterol, and calories makes it a potentially unhealthy food. Patients who eat meat should be encouraged to select lean cuts, trim away visible fat, and use cooking methods that remove, rather than add, fat. It is even more beneficial to reduce the amount of meat consumed by reducing portion size and frequency; a reasonable goal is to eat about 4 oz one to three times a week.

Poultry is a more healthful source of protein and other nutrients. Chicken and turkey are best, but the skin should be removed before cooking to reduce the fat content.

Dairy products and eggs To reduce intake of saturated fat and cholesterol, nonfat or low-fat dairy products can be substituted for whole-milk products. The use of nondairy creamers, imitation cheese, margarine, and other products that contain *trans*-fatty acids in partially hydrogenated vegetable oils should be limited. The consumption of up to one egg a day does not appear to increase the risk of cardiovascular disease in healthy, nondiabetic

Table 5 Dietary Guidelines for Healthy People

Eat more vegetable products than animal products
Eat more fresh and homemade foods than processed foods
Less than 30% of calories should come from fat
Limit cholesterol to less than 300 mg a day
Eat at least 25 g of fiber a day
55%–65% of calories should come from complex carbohydrates
10%–15% of calories should come from protein
Limit sodium to less than 2,400 mg a day
Obtain 1,200–1,500 mg of calcium a day from food or supplement
Eat 6 or more servings of grain products a day
Eat 3–5 servings of vegetables and legumes a day
Eat 2–4 servings of fruit a day
Eat two 4 oz servings of fish a week
Eat no more than two 4 oz servings of red meat a week
Chicken and turkey should be eaten in moderation with skin removed
Eat no more than one egg yolk a day, including those used in cooking and baking
Use vegetable oils, preferably olive and canola oils, in moderation
Have no more than two alcoholic drinks a day
Adjust caloric intake and exercise level to maintain a desirable body weight
Avoid fad diets and extreme or unconventional nutrition schemes
Avoid untested nutritional supplements, including megavitamins, herbs, food extracts, and amino acids

people,[72] but additional egg-yolk consumption should be limited. One egg yolk contains about two thirds of the total amount of cholesterol that is recommended for an entire day. Egg whites and egg substitutes are good alternatives to egg yolks.

Fish A 1997 study reported that participants who consumed 245 g or more of fish a week enjoyed a 38% reduction in fatal myocardial infarctions over a 30-year period.[73] Although at least three other observational studies did not find that fish consumption was protective, a 1989 intervention trial that randomized 2,033 myocardial infarction survivors to usual care or usual care plus fish consumption found that eating two or three fish meals a week reduced 7-year mortality by 29%.[74] Fish consumption has also been associated with a reduced risk of primary cardiac arrest,[75] hypertension, stroke,[9] and prostate cancer.[76,77] As little as 4 oz of fish twice a week may provide protection.[19] Fish should be baked, broiled, grilled, steamed, or poached rather than fried, and high-fat sauces should be avoided. Because of their higher content of omega-3 fatty acids, oily, deep-water fish may be best. People who are reluctant to eat fish may benefit from fish oil supplements in the modest dose of about 1 g/day.[78] More study is required to confirm the value of fish and fish oil and to define the optimal types and amounts of fish.

Cooking oils Canola oil contains an omega-3 fatty acid, α-linolenic acid. High serum levels of α-linolenic acid have been associated with a decreased risk of stroke, and consumption of canola oil is inversely related to the risk of myocardial infarction.[12] Canola oil and olive oil have a high content of oxidation-resistant monounsaturated fatty acids. Olive oil may be a cardioprotective element in the Mediterranean diet and may also reduce the risk of breast cancer. Although more study is needed, canola and olive oils appear to be the most beneficial oils for food preparation.

Nuts Nuts are high in monounsaturated and polyunsaturated fatty acids and fiber. Nut consumption appears to be inversely related to the risk of coronary artery disease[79] and diabetes.[80]

Garlic Medical studies of garlic have shown mixed results. Some meta-analyses suggest that garlic extracts can improve blood cholesterol levels, but others do not.[81] The putative benefits of garlic on blood pressure and coagulation are even less clear.

Flavonoid-rich foods Flavonoids are polyphenolic antioxidants that are found in a variety of vegetable foods, including apples, onions, tea, and red wine. Although not all studies agree, consumption of these foods has been inversely related to the risk of coronary artery disease and stroke.[82]

Alcohol Rarely regarded as a nutrient, alcohol should be considered when dietary recommendations are formulated. Containing 7 cal/g, alcohol is a calorie-dense food. Numerous studies demonstrate that low to moderate alcohol consumption substantially reduces the risk of coronary artery disease, peripheral vascular disease, and all-cause mortality.[83] The major mechanism of protection is alcohol's ability to increase HDL cholesterol levels; favorable effects on blood coagulation mechanisms may also contribute. Protective doses of alcohol can be obtained from one to two drinks a day; 5 oz of wine, 12 oz of beer, or 1.5 oz of spirits is counted as one drink. Despite its antioxidant content, red wine is no more protective than other alcoholic beverages.[84]

Caffeine Studies have failed to confirm putative links between caffeine and peptic ulcers, hypertension, coronary artery disease, breast disease, or cancer. Caffeine can trigger migraines in sensitive individuals, and caffeine withdrawal can precipitate headaches or depression in habitual consumers. Caffeine can cause anxiety, insomnia, and gastroesophageal reflux. Brewed coffee can increase blood cholesterol levels, but filtered coffee does not. The effects of caffeine on pregnancy are not fully understood, but it is reasonable to discourage consumption of large amounts.[85] Caffeine restriction does not reduce palpitations in patients with idiopathic premature ventricular contractions.[86]

DIET AND HEALTH

Much remains to be learned about the complex relation between nutrition, health, and disease. Dietary preferences are no less complex and individual. Despite these uncertainties, a dietary pattern characterized by a high intake of vegetables, fruits, legumes, whole grains, fish, and poultry is associated with major health benefits for men[87] and women.[30,88] Physicians have an important role in educating patients about healthful nutrition and providing dietary guidelines [*see Table 5*].

Exercise

Numerous observational studies have demonstrated a dose-related inverse relation between habitual physical activity and the risk for many of the chronic illnesses that afflict people in industrialized societies.[89] The protective effect of exercise is strongest against coronary artery disease but is also significant against hypertension, stroke, type 2 (non–insulin-dependent) diabetes mellitus, obesity, anxiety, depression, osteoporosis, and cancers of the colon, breast, and female reproductive tract. Despite these

proven benefits, only 25% of adults in the United States exercise at recommended levels.[90] Of all deaths in the United States, as many as 12%, or about 250,000 annually, can be attributed to a sedentary lifestyle.[91]

EXERCISE PHYSIOLOGY

The physiologic effects of exercise depend on the type of exercise, its intensity, its duration, and its frequency.[92] Exercise is either isometric or isotonic. Isometric contraction of muscle is characterized by an increase in muscle tension without a significant change in fiber length. No external work is accomplished, but substantial energy is expended. Examples of isometric work include handgrip exercises, pushing or pulling against a fixed resistance, and holding a heavy weight. In contrast, isotonic work involves a shortening of muscle fibers with little increase in tension; examples include swimming, bicycling, and running. Most exercise includes both isometric elements and isotonic elements.

Isometric and isotonic exercises differ substantially in their physiologic effects. Isometric work increases total peripheral resistance; both systolic blood pressure and diastolic blood pressure rise substantially, with relatively little increase in stroke volume or cardiac output. Isotonic work lowers total peripheral resistance, but heart rate and cardiac output rise. Systolic blood pressure rises substantially, but diastolic pressure changes little, resulting in a small increase in mean arterial pressure. Isometric work places a pressure load on the heart, whereas isotonic work imposes a volume load.

Isometric exercise increases muscle strength and bulk, which is desirable for competitive athletes, for patients recovering from musculoskeletal injuries, and for individuals who wish to attenuate the loss of muscle mass and bone strength that accompanies sedentary aging and certain chronic illnesses.[93] However, static exercises produce minimal cardiovascular conditioning, and the circulatory demands of intense isometric work can be hazardous to patients with heart disease. In contrast, dynamic exercises enhance endurance and can produce adaptive cardiovascular changes in healthy individuals and cardiac patients.

Cardiovascular Response to Dynamic Exercise

The acute circulatory response to maximal dynamic exercise is a dramatic rise in cardiac output, from about 5 L/min to 20 L/min in healthy young men. The increased cardiac output results from a 300% increase in heart rate. This increased transport of oxygen is matched by a threefold increase in peripheral oxygen extraction. Total peripheral resistance falls, and blood is shunted away from nonworking muscles and the viscera toward exercising muscles and the coronary circulation, where blood flow increases fourfold.

The physiologic adaptations produced by repetitive dynamic exercise are known collectively as the training effect. The magnitude of the training effect depends on the intensity, duration, and frequency of exercise. Training requires rhythmic, repetitive use of large muscle groups for prolonged periods. Fitness can be developed and maintained in healthy adults with three to five exercise sessions a week. Each day's exercise should involve isotonic work at 60% to 90% of maximal heart rate for 20 to 60 minutes, either continuously or in increments of 10 minutes or longer.[94] Obviously, sedentary persons and patients with cardiopulmonary disease must initiate training at lower intensities and shorter durations and build up gradually.

Perhaps the most obvious training effect is resting bradycardia; heart rates of 40 to 50 beats/min are common in highly trained endurance athletes. The mechanisms responsible are not fully understood but probably involve increased vagal tone, decreased sympathetic activity, and increased stroke volume. The best overall measurement of the training effect and of physical fitness is the maximal oxygen uptake ($\dot{V}O_{2MAX}$).

Oxygen consumption relates directly to the amount of muscular work; maximal oxygen uptake therefore reflects maximal work capacity. Many factors determine an individual's $\dot{V}O_{2MAX}$, including age, gender, lean body mass, genetics, and, most important, the level of habitual exercise. Just 3 weeks of bed rest will cause a 20% to 25% decline in $\dot{V}O_{2MAX}$. It is no wonder that patients are debilitated after being confined to bed by illness or treatment regimens. In contrast, regular training lasting weeks or months will increase $\dot{V}O_{2MAX}$, typically by 30% to 40%.

Both central (cardiac) and peripheral (muscular) adaptations are involved in the training effect. In healthy individuals, training produces dramatic changes in cardiac structure. The dimensions of all cardiac chambers increase by up to 20%, and myocardial mass may increase as much as 70%.[95] Although increased coronary blood flow and collateralization have not been demonstrated directly in humans, echocardiographic studies show that elite athletes have increased proximal coronary artery size, which is proportional to their increased left ventricular mass. Cardiac function is also enhanced by training; left ventricular contractility and stroke volume increase, and angiographic studies have demonstrated increased dilating capacity in the coronary arteries of endurance athletes. Exercise training also improves endothelial function in patients with coronary artery disease and in elderly persons.[96,97]

In addition to these cardiac changes that allow enhanced oxygen delivery, there is improvement in peripheral oxygen extraction caused by enhanced O_2 extraction by the skeletal muscles themselves. This effect on skeletal muscle is specific for the muscles that have been trained; if only leg muscles are trained, the circulatory response to strenuous leg exercise will improve, but the response to vigorous arm exercise will not change.

Exercise training decreases the risk of hypertension. A meta-analysis of 54 controlled intervention studies concluded that isotonic exercise training lowers both systolic blood pressure by about 4 mm Hg and diastolic blood pressure by about 3 mm Hg.[98] Regular exercise can even reduce left ventricular hypertrophy and blood pressure in patients with severe hypertension.[99] Regular exercise also lowers catecholamine levels, protecting against arrhythmias, and it reduces myocardial oxygen demands.[89]

Although isotonic exercise reduces resting blood pressure, isometric exercise increases total peripheral resistance and acutely elevates blood pressure. However, sustained hypertension is not a complication of resistance training, which may even reduce resting blood pressure.[100] Unsupervised isometric exercising should be avoided by patients with cardiovascular disease; with appropriate precautions, however, it can be safe for selected cardiac patients and can produce favorable effects on muscular function.[101]

Pulmonary Response

Except in people with intrinsic lung disease, the pulmonary diffusion capacity does not limit exercise. At heavy work loads, however, skeletal muscle oxygen demands exceed oxygen delivery. As a result, muscle metabolism becomes anaerobic; the lactic acid that accumulates is buffered by bicarbonate, so that

the pH remains nearly normal. The CO_2 that is liberated by the buffering reaction produces an increased ventilatory drive and tachypnea. Athletes know when they have crossed the anaerobic threshold by a markedly increased respiratory rate and a sensation of dyspnea. Habitual exercise does not improve pulmonary function in healthy people, but exercise training may be helpful in patients with chronic lung disease as a result of adaptations in muscles rather than in the lungs.

Musculoskeletal Response

Isotonic exercises increase muscle endurance. Training increases capillary density, and it can increase muscle mitochondria and oxidative capacity more than twofold. These changes account for the greater oxygen extraction that is an important element of the training effect. Isometric training builds muscle mass, which improves performance and may decrease injuries. Isometric exercises involving slow repetitions of work against high resistance produce fiber hypertrophy and strength but do not alter muscle enzyme content.

Exercise training affects tissues in addition to muscles. Of great importance, weight-bearing exercises increase bone mineral density, reducing the risk of osteoporosis. Repetitive performance of athletic tasks improves coordination and efficiency; changes in neuromuscular recruitment may be partially responsible. Tendon strength and bone density increase as a result of repetitive use. Joint wear and tear remains a concern, but as long as there is no trauma, habitual exercise probably does not produce degenerative joint disease.[102] In fact, aerobic and resistance exercise may help reduce disability in patients with osteoarthritis and fibromyalgia.[103,104]

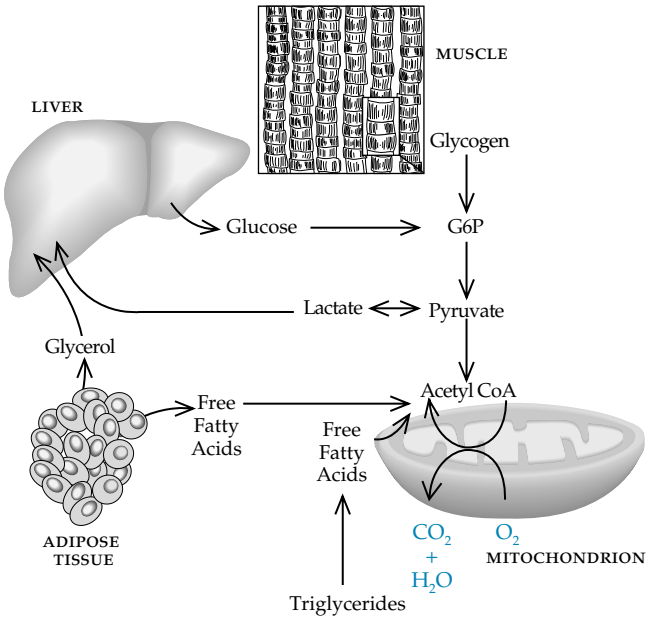


Figure 2 **During exercise, catecholamine stimulation of adipose tissue rapidly mobilizes free fatty acids to achieve blood levels that are six times the normal level, which are far higher than the muscle can use. Glucose derived from the liver and muscle glycogen are initially phosphorylated to yield glucose-6-phosphate (G6P). The G6P, the free fatty acids from adipose tissue, and the muscle's own triglycerides are metabolized to acetyl coenzyme A (acetyl CoA). This compound then undergoes oxidative metabolism in the mitochondrial Krebs cycle (blue), thus providing energy for exercising muscle.**

Metabolic Effects

Skeletal muscle contains only very limited energy stores; preformed adenosine triphosphate (ATP) and creatine phosphate (CP) can supply less energy than that which is consumed in a 100-yard dash. Clearly, ATP and CP must be generated during exercise. Only three sources of fuel are available to skeletal muscle for this purpose: endogenous muscle glycogen, blood glucose, and free fatty acids (FFAs) derived either from muscle triglyceride or from adipose tissue. Normally, the body's skeletal muscle contains only 120 g of glycogen and the liver only 70 g. The 600 kcal of energy available from these two sources could sustain running for only 6 miles. The blood glucose provides only 40 kcal more. In contrast, the average person's 15,000 g of adipose tissue provides 100,000 kcal of energy, theoretically enough to fuel a run from Boston to Atlanta.

At rest and during low-intensity exercise, both FFAs and muscle glycogen provide energy. As exercise begins, catecholamines stimulate adipose lipase, which cleaves triglyceride into glycerol and three FFA molecules [*see Figure 2*]. In muscle cells, FFAs are metabolized to acetyl coenzyme A (acetyl CoA); in the presence of oxygen, acetyl CoA undergoes oxidative metabolism by enzymes of the citric acid (Krebs) cycle in mitochondria.

As the intensity of exercise increases, the relative contribution of FFAs decreases and glycogen becomes more important, and at maximum work, muscle depends entirely on glycogen. When oxygen is available, glycogen is metabolized in the cytoplasm to pyruvate, which then undergoes oxidation in the mitochondria via the citric acid cycle to water and CO_2. However, when the demands of muscle outstrip the availability of oxygen, energy can be generated only anaerobically via glycolysis. Anaerobic metabolism is much less efficient: from a gram of glycogen, anaerobic metabolism generates only 5% of the energy that aerobic metabolism generates. In addition, pyruvate cannot be converted to acetyl CoA. Instead, pyruvate is reduced to lactate. Acidosis limits muscular performance, and buffering by the bicarbonate system generates CO_2, causing tachypnea.

Although the blood glucose itself constitutes only a modest caloric reserve, glucose turnover is greatly accelerated by exercise. During exercise, the liver releases glucose by both glycogenolysis and gluconeogenesis. Simultaneously, peripheral glucose uptake is enhanced. As a result of these metabolic events, blood glucose can account for 10% to 30% of exercising muscle's metabolic needs. The blood glucose level remains normal and may even rise during modest exertion. However, hypoglycemia can occur if hepatic glycogen stores are depleted and high-intensity exercise continues to consume blood glucose and muscle glycogen.

These changes in glucose metabolism are moderated by a number of hormonal alterations. Circulating catecholamines, growth hormone, cortisol, and glucagon levels rise. Insulin levels fall. All of these factors tend to elevate blood glucose levels. Glucose that is ingested during exercise will also tend to maintain blood glucose levels, but ingestion of glucose before exercise may actually raise insulin levels, thus impeding energy mobilization. Contrary to popular so-called instant-energy theories, preexercise meals should not contain concentrated sweets. Indeed, preexercise meals should be sparse, and people should probably ingest little other than water during the 2 hours before exercise.

Exercise increases the insulin sensitivity of muscle, thereby increasing glucose transport and muscle glycogen synthesis.

Even moderate physical activity such as walking can help prevent the development of type 2 diabetes mellitus[105] and the metabolic syndrome.[106] Because exercise improves glucose tolerance in diabetic patients, patients taking insulin may require special precautions to exercise safely [*see* Medical Complications of Exercise, *below*].

During exercise, the rate of protein synthesis is depressed. As a result, amino acids are available for anabolic processes, including hepatic gluconeogenesis. Amino acids also may directly provide a small fraction of the energy needed for muscle contraction. It is not clear whether athletes have higher nutritional protein requirements than sedentary persons; the ingestion of protein and amino acid supplements does not enhance athletic performance.

Regular aerobic exercise also alters body weight and body composition. If dietary caloric intake remains constant, exercise will produce slow weight loss.[107] It takes 35 miles of walking or jogging to consume the calories present in 1 lb of adipose tissue. Intense exercise also stimulates both energy expenditure and lipid oxidation for up to 17 hours after exercise itself, thus further contributing to a reduction in body fat. Even as body fat declines, muscle mass increases; because muscle is denser than fat, net weight loss may be slight. Swimming appears to be less effective than land exercise for reducing body fat and increasing bone mineral content.

Effects on Blood Lipids

Exercise increases serum levels of HDL-associated cholesterol (HDL-C), probably by delaying hepatic HDL-C catabolism. The amount of exercise appears to be the major determinant of the magnitude of the increase in HDL-C. As little as 5 to 10 miles of jogging a week will elevate HDL-C levels, which rise with increasing exercise in a dose-response fashion; beyond about 35 miles a week, however, additional training does not produce a further increase in HDL-C levels.[108] Similar changes in HDL-C levels have also been demonstrated in walkers, cross-country skiers, tennis players, bicyclists, and other endurance athletes. The effects of exercise are independent of other factors known to alter HDL-C levels, such as diet, body weight, smoking, and alcohol consumption. Exercise must be sustained to maintain high HDL-C levels.

The effects of exercise on HDL-C levels are observed consistently, but changes in the other blood lipid levels have varied. In general, exercise produces a fall in triglyceride and chylomicron levels. Total cholesterol and LDL cholesterol levels also tend to decline. Heritable factors, in part, determine lipid profile responses to exercise.[109]

Hematologic Effects

A mild decrease in hematocrit is commonly observed in endurance athletes. This so-called sports anemia is usually a pseudoanemia, because red blood cell mass is normal but plasma volume is increased; decreased viscosity has also been observed. Exercise-related hemolysis or gastrointestinal blood loss may be an additional factor in some cases of anemia in athletes. No consistent long-term changes in polymorphonuclear leukocytes, lymphocytes, or immunoglobulins have been noted.

Hemostatic mechanisms are influenced by exercise. Endurance exercise acutely increases fibrinolytic activity, and repetitive exercise is associated with reduced fibrinogen levels. In contrast, intense exercise can activate platelets,[110] perhaps contributing to a prothrombotic state that may contribute to exertion-induced cardiac events[111] [*see* Medical Complications of Exercise, *below*]. The effects of exercise on platelet function require further study.

Effects on Body Fluids

During exercise, skeletal muscle generates a tremendous amount of heat. Sweating is necessary to dissipate this heat. During strenuous exercise in a warm environment, up to 2 L can be lost each hour. Because sweat is hypotonic, the serum sodium concentration rises. Even in the absence of systemic acidosis, serum potassium levels may rise because of an efflux of potassium from muscle cells, but potassium levels normalize within minutes after exertion ceases.

The decline in blood volume, together with a shift in blood flow from the kidneys to skeletal muscle, produces a sharp decline in urine volume during exercise. The rise in plasma osmolarity increases thirst. However, thirst lags behind volume requirements, and fluid intake is often inadequate during athletic events. Volume depletion impairs athletic performance and can contribute to renal dysfunction or heatstroke. Unfortunately, coaching lore often limits fluid intake for fear of cramps, when, in fact, athletes can tolerate large volumes of fluids during brief pauses in exercise. Although water is an excellent fluid replacement, excessive amounts during prolonged exercise can produce severe, even fatal, hyponatremia.[112] Athletes do not require supplemental potassium or salt, so popular glucose-sodium-potassium solutions make little sense physiologically.

Psychological Effects

Endurance exercise produces improvements in mood, self-esteem, and work behavior both in healthy people and in patients undertaking cardiac rehabilitation; exercise training can help treat depression.[113] Several mechanisms have been suggested to explain the psychological effects of exercise. Purely psychological factors, such as distraction, may be involved. The serum levels of β-endorphin, monoamines, and other neuropeptides are affected by exercise in direct relation to the intensity and duration of exercise. Changes in endogenous opioid peptides may mediate the subjective effects of exercise (so-called runner's high).

EXERCISE AND THE ELDERLY

Many physiologic changes attributed to aging closely resemble those that result from inactivity.[114] In both circumstances, bone calcium wastage occurs, and there are decreases in $\dot{V}O_{2MAX}$, cardiac output, red blood cell mass, glucose tolerance, and muscle mass; total peripheral resistance and systolic blood pressure are increased, as are body fat and serum cholesterol levels. Regular exercise appears to retard these age-related maladies. Exercise training improves left ventricular systolic function and increases stroke volume to maintain exercise cardiac output in healthy, older people.[115] The age-related decline in $\dot{V}O_{2MAX}$ has been found to be twice as great for sedentary men as for active men, and even low-intensity training can improve $\dot{V}O_{2MAX}$ in the elderly. Exercise training also helps blunt the age-related decline in peripheral vascular function experienced by sedentary people. Endurance training improves glucose tolerance and serum lipid levels in older men and women, and regular exercise appears to blunt the age-related decline in resting metabolic rate. Physical activity in the elderly is associated with increased functional status and decreased mortality. Exercise is safe in the elderly if simple precautions are observed [*see* Prescribing Exer-

cise, *below*]. Walking programs increase aerobic capacity in persons 70 to 79 years of age, with few injuries; healthy elderly persons who are randomly assigned to aerobic exercise acquire fewer new cardiovascular disorders than control subjects. Appropriate resistance weight programs are not hemodynamically stressful in the elderly and produce increases in muscle strength, functional mobility, and walking endurance. Even frail nursing home residents (mean age, 87 years) responded to resistance training with an increase in muscle mass and strength, as well as improved gait velocity, stair-climbing power, and spontaneous activity. Although more studies are needed to clarify correlations between aging, inactivity, and exercise, enough information is available to warrant a recommendation of carefully planned exercise programs for the elderly.

EXERCISE AND LONGEVITY

Primary Prevention of Atherosclerosis

Exercise training can favorably modify many of the conditions associated with an increased risk of coronary artery disease, including hypercholesterolemia, elevated blood pressure, glucose intolerance, obesity, and the less firmly incriminated traits of hypertriglyceridemia, hyperinsulinemia, hyperfibrinogenemia, and psychological stress. Studies conducted in men, women, and children demonstrated a consistent inverse relation between physical fitness and body weight, percent body fat, systolic blood pressure, and serum levels of cholesterol, triglycerides, and glucose.[116] In addition, exercise appears to reduce levels of C-reactive protein and other inflammatory markers of cardiovascular risk.[117]

Is a sedentary way of life itself a risk factor independent of these other traits? Investigators at the Centers for Disease Control and Prevention (CDC) reviewed 43 methodologically sound studies of exercise and coronary artery disease.[118] Collectively, these studies showed that sedentary living increases coronary risk by 1.9 times. An independent meta-analysis derived the same relative risk.[119] The magnitude of this excess risk is similar to that conferred by other risk factors: hypertension, 2.1 times; hypercholesterolemia, 2.4 times; and cigarette smoking, 2.5 times.[118] Because sedentary living is at least two to three times more prevalent than any of these other risk factors, it can be argued that physical inactivity makes the most significant contribution to the epidemic of coronary artery disease in the United States. Maintaining a physically active way of life can be expected to reduce the risk of myocardial infarction by 35% to 70%.

Although reductions in coronary artery disease account for the great majority of the improvements in survival conferred by exercise, other factors may play a role. Physical activity protects against stroke[120] and reduces the risk of colon cancer.[121] Exercise also reduces the risk of breast cancer[122] and cancer of the reproductive organs in women, and very intensive exercise may reduce the risk of prostate cancer. Physical activity is associated with a reduced risk of age-related cognitive dysfunction[123] and hip fractures.[124] Current studies continue to confirm observations that have occurred over the past 50 years, demonstrating that there is a graded, inverse association between activity and mortality.[125,126]

Secondary Prevention of Ischemic Heart Disease

Since the 1970s, interest in the potential role of exercise in the rehabilitation of patients after myocardial infarction and in the prevention of recurrent cardiac events has grown.[127] Certain benefits of supervised exercise programs have been clearly established, including physiologic and symptomatic improvements and the reduction of risk factors. Patients completing exercise programs demonstrate the training effect, including a lower heart rate at rest and both a lower heart rate and a lower systolic blood pressure at submaximal work loads. These changes reduce myocardial oxygen demands, thereby increasing the angina threshold. Significant improvements in maximal oxygen uptake and work capacity can also be demonstrated. Exercise training can improve walking distances in patients with claudication.[128] Exercise can be useful even for patients with severe ischemic left ventricular dysfunction and chronic congestive heart failure, although extra precautions should be taken in these patients. In addition, cardiac exercise programs are safe.[129] Most centers note a substantial improvement in mental attitudes, a decrease in the use of medications, and widespread patient satisfaction. Most important, randomized trials demonstrate that cardiac exercise programs reduce mortality by 20% to 25%.[129,130] Unsupervised moderate exercise, such as walking or gardening, also appears to reduce mortality in older patients with coronary artery disease.[131]

PRESCRIBING EXERCISE

Physicians can provide important incentives for their patients by educating them about the benefits, as well as the risks, of habitual exercise. Healthy, sedentary individuals are the largest group in need of such advice. In addition, physicians may be responsible for the medical screening of competitive athletes or for prescribing exercise for patients with chronic illnesses.

A careful history and physical examination are central to the medical evaluation of all potential exercisers. Particular attention should be given to a family history of coronary disease, hypertension, stroke, or sudden death and to symptoms suggestive of cardiovascular disease. Cigarette smoking, sedentary living, hypertension, diabetes, and obesity all increase the risks of exercise and may indicate the need for further testing. Physical findings suggestive of pulmonary, cardiac, or peripheral vascular disease are obvious causes for concern. A musculoskeletal evaluation is also important.

The choice of screening tests for apparently healthy individuals is controversial. A complete blood count and urinalysis are reasonable in all cases. Determination of blood glucose, serum cholesterol, and creatinine levels may also be useful in screening for risk factors or occult disease. The Valsalva maneuver and the isometric handgrip may be useful additions to the workup.

Young adults who are free of risk factors, symptoms, and abnormal physical findings do not require further evaluation. It is not at all clear that more aggressive medical screening can prevent sudden cardiac death. Although echocardiography and electrocardiography might reveal asymptomatic hypertrophic cardiomyopathy in some patients, the infrequency of this problem makes routine screening impractical.

The role of exercise electrocardiography as a screening test before an individual begins an exercise program is controversial. The AHA no longer recommends routine exercise testing for asymptomatic individuals.[89] In fact, a study of 3,617 asymptomatic men 35 to 59 years of age casts doubt on the value of exercise electrocardiography for routine preexercise screening.[132] None of the men had known coronary artery disease on entry into the study, but all were at increased risk because of hypercholesterolemia. Each individual had an exercise test on entry;

Table 6 Exercise Time Required to Consume 2,000 kcal

Activity	*Time (hr)*
Strolling	10
Bowling	8.5
Golf	8
Raking leaves	7
Doubles tennis	6
Brisk walking	5.5
Biking (leisurely)	5.5
Ballet	4.5
Singles tennis	4.5
Racquetball, squash	4
Biking (hard)	4
Jogging	4
Downhill skiing	4
Calisthenics, brisk aerobics	3.3
Running	3
Cross-country skiing	3

the tests were repeated annually over a mean follow-up period of 7.4 years. Exercise proved safe in this group, with approximately 2% experiencing exercise-related cardiac events. Only 11 of the 62 men who experienced such events had abnormal exercise tests on entry, a sensitivity of only 18%. The cumulative sensitivity of annual tests was also low (24%). Even in elderly people, routine exercise testing before starting a moderate exercise program may not be necessary.[133]

Despite its limitations as a screening test for silent coronary artery disease, exercise testing can be useful for detecting exercise-induced arrhythmias, establishing a maximal heart rate for the exercise prescription, and determining work capacity. Serial testing may help motivate a patient by demonstrating increased work capacity. Specialized tests such as pulmonary function tests and exercise ergometry, Holter or telemetric monitoring during exercise, and echocardiography may be very useful in the evaluation of patients who have known or suspected cardiovascular abnormalities.

Screened patients will fall into one of three groups:

1. Healthy persons who can exercise without supervision. (Medical guidelines [see below] may still be helpful.)
2. Patients with ischemic heart disease or other significant cardiovascular abnormalities who should have medically supervised, graded exercise programs. (If structured programs are not available, such patients should engage in milder forms of exercise, such as walking or bicycling, with appropriate precautions.)
3. Patients for whom physical exertion is contraindicated because of decompensated congestive heart failure, complex ventricular irritability, unstable angina, significant aortic valve disease, aortic aneurysm, uncontrolled diabetes, or uncontrolled seizure disorders.

People can exercise in the course of daily life or in formal exercise programs. Although most physicians have recommended structured exercise, recent studies demonstrate that even modest levels of physical activity are beneficial.[131,134] Walking and gardening are good examples[131,134,135]; such activities are protective even if they are not started until midlife or late in life.[136] In one study, for example, elderly men who walked less than 1 mile a day had nearly twice the mortality of men who walked more than 2 miles a day.[137] Compliance with walking is good,[138] and lifestyle interventions appear to be as effective as formal exercise programs of similar intensity in improving cardiopulmonary fitness, blood pressure, and body composition.[139,140]

People should be encouraged to exercise nearly every day. Formal, intense exercise is not necessary; even moderate exercise that consumes about 150 kcal/day or 1,000 kcal/wk is very beneficial to health. Warm-ups, stretches, and a graded increase in exercise intensity can help prevent musculoskeletal problems.

Whereas all people can benefit from moderate daily activity, additional benefit can be obtained from more intense exercise; people who consume about 2,000 kcal in exercise a week obtain the greatest reduction of cardiovascular risk and mortality[125,141] [*see Table 6*]. On average, people can obtain optimal health benefits from about 30 minutes of intense exercise or 45 to 60 minutes of mild to moderate exercise a day.

Physicians who provide specific practical advice are most likely to motivate their patients to adopt better health habits, including diet and exercise [*see Table 7*].

The success of a structured fitness program depends on the frequency, duration, and intensity of exercise. At least three sessions a week are needed. An alternate-day schedule will help prevent muscle soreness, but as fitness improves, individuals should be encouraged to increase exercise sessions to five or even seven times a week. Each session consists of 15 to 60 minutes of continuous aerobic activity. Untrained individuals may not be able to sustain even 15 minutes at first, but they should be encouraged to progress slowly as they improve. Each exercise session should be preceded by a 5- to 10-minute warm-up period and followed by a 5- to 10-minute cool-down period; stretching, gentle calisthenics, and walking are ideally suited for this purpose. These same exercises are excellent for a 5- to 10-minute cool-down period.

The intensity of exercise can best be judged by the target heart rate. A heart rate of 60% to 85% of maximum is considered optimal for training. If an exercise test has not been performed, a maximal heart rate can be calculated by subtracting

Table 7 Exercise Advice for Patients

- Get a medical checkup before beginning a formal exercise program
- Warm up before each exercise session, and cool down afterward with 10 minutes of stretching and light calisthenics
- Start slowly and build up to 30 minutes of moderate to intense exercise or 45–60 minutes of mild to moderate exercise
- Begin with aerobic-type exercise, and later add stretching exercises for flexibility and low-resistance weight training for strength
- Exercise daily if possible, and alternate harder workouts with easier ones
- Dress comfortably
- Use good equipment, especially good shoes
- Do not eat during the 2 hr before you exercise, but drink plenty of water before, during, and after exercise, particularly in warmer weather
- Do not ignore aches and pains that may signify injury
- Do not exercise if you are feverish or ill
- Learn warning signals of heart disease, including chest pain or pressure, disproportionate shortness of breath, fatigue, sweating, erratic pulse, light-headedness, or even indigestion
- Consider getting instruction or joining a health club

the patient's age from 220. Unfit people should start at the lower end of the target heart rate range. Healthy people need not monitor pulse rate. Instead, they can adjust the intensity of effort to a talking pace: they are working hard but still able to talk to a companion without a sensation of dyspnea.

Many kinds of exercise can be used to attain fitness. Dynamic (i.e., isotonic or aerobic) exercises in which large muscle groups are used continuously in a rhythmic, repetitive fashion for prolonged periods are ideal. The energy requirements of various activities have been measured. An energy expenditure of 5 to 6 METs (metabolic equivalents) or more is desirable for exercise training (1 MET is equal to the energy expenditure at rest or equivalent to approximately 3.5 ml O_2/kg body weight/min). Brisk walking, jogging, swimming, cross-country skiing, skating, bicycling, and vigorous singles racket sports all provide good conditioning. Sports that allow prolonged periods of inactivity, such as doubles tennis, golf, bowling, and baseball, are much less desirable for fitness.[125] Activities requiring sudden bursts of intense isometric activity, such as weight lifting, provide little cardiovascular conditioning and are contraindicated for patients with hypertension or heart disease. Contact sports cannot be recommended for health.

Although physicians should encourage patients to choose the sports that appeal most to them, medical considerations may also be important. For example, swimming is particularly desirable for individuals who have various musculoskeletal problems, and it is also ideal for people who experience exercise-induced asthma (EIA). Walking and bicycling are ideal for older individuals or for anyone who is starting from a low level of fitness. Jogging can be recommended because it is convenient and because the participants can easily adjust intensity and duration upward as fitness develops. Most desirable of all is a balanced program containing a variety of activities that exercise different muscle groups. People who have several activities at their command find it easier to remain active despite constraints of climate, schedules, and minor injuries. Although aerobic exercise is most important for metabolic improvement and cardiovascular health, exercises for flexibility and strength should be part of a balanced fitness program.[94] Stretching exercises promote flexibility and help prevent injuries. A stretching routine should be performed at least two to three times a week, but it is best when incorporated in the warm-up and cool-down periods that should surround aerobic exercise. Low-resistance strength training is important to preserve muscle mass and power in the face of the aging process; two to three sessions a week are ideal.

COMPLICATIONS OF EXERCISE

Reducing Risk of Injury and Complications

Physicians can minimize injuries and medical complications associated with exercise by educating patients about potential problems. Physicians should stress the need for such safety devices as helmets for biking, eye guards for squash and racquetball, and elbow and knee pads for roller-skating. Diet, weight control, stress management, smoking cessation, and other preventive health measures should be discussed [*see CE:III Reducing Risk of Injury and Disease*], as should the warning signs of cardiac disease and the precautions for exercising in cold or hot climates.

Medical Complications of Exercise

Exercise promotes health, but it can also have adverse consequences. In some cases, the physiologic adaptations to exercise produce changes that may be misinterpreted as pathologic; athlete's heart is one example. In other cases, however, exercise can precipitate clinically important problems.

The cardiac complications of exercise include ischemia, infarction, and sudden death, often caused by rupture of an atherosclerotic plaque.[142] These dire events are infrequent and can be minimized by proper patient screening and instruction [*see* Prescribing Exercise, *above*]. Exercise-induced cardiac events are less common in people who exercise regularly than in sedentary individuals.[143] On balance, exercise is clearly beneficial for the heart.

The most common pulmonary complication of exercise is EIA, which usually responds well to treatment.[144] A much less common problem that can mimic hypersensitivity disorders is exercise-induced anaphylaxis.

The gastrointestinal response to exercise may produce reflux, diarrhea, or bleeding, which is usually occult and transient. Women who exercise very strenuously may experience oligomenorrhea or amenorrhea; the menstrual dysfunction is reversible but may be accompanied by osteoporosis. With appropriate precautions, exercise is safe during pregnancy. Precautions are also in order for prevention of hypoglycemia in diabetics who exercise.

People who exercise regularly can experience increased plasma volume that produces hemodilution or pseudoanemia. True anemia is less common but may result from shortened red cell life span caused by vascular trauma or iron deficiency. Exercise can produce proteinuria or hematuria; both are benign but are indications for studies to rule out renal disease. In warm, humid weather, exercise can produce heat cramps, hyperthermia, or heatstroke, all of which are preventable.

Exercise does not appear to cause or accelerate osteoarthritis.[102] Acute muscle injury, manifested by transient elevation of creatine phosphokinase levels, is common, but exertional rhabdomyolysis is rare. Musculoskeletal problems, however, are the most frequent side effects of exercise.[145] Overstress, overuse, or trauma is usually responsible. Poor technique, faulty equipment, or fatigue often contributes to injury. Soft tissue injuries such as sprains, strains, and tendinitis usually respond well to simple treatment regimens. The same is true of stress fractures. Primary care physicians can manage many of these problems, but more serious injuries may merit referral to a sports medicine facility.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. McGinnis JM, Foege WH: Actual causes of death in the United States. JAMA 270:2207, 1993

2. Roberts CK, Vaziri ND, Barnard RJ: Effect of diet and exercise intervention on blood pressure, insulin, oxidative stress, and nitric oxide availability. Circulation 106:2530, 2002

3. Manson JE, Bassuk SS: Obesity in the United States. A fresh look at its high toll. JAMA 289:229, 2003

4. Peeters A, Barendregt JJ, Willekens F, et al: Obesity in adulthood and its consequences for life expectancy: a life-table analysis. Ann Intern Med 138:24, 2003

5. Nielsen SJ, Popkin BM: Patterns and trends in food portion sizes, 1977-1998. JAMA 289:450, 2003

6. Ascherio A, Rimm EB, Giovannucci EL, et al: Dietary fat and risk of coronary heart disease in men: cohort follow up study in the United States. BMJ 313:84, 1996

7. Rissanen T, Voutilainen S, Nyyssonen K, et al: Fish oil–derived fatty acids, docosahexaenoic acid and docosapentaenoic acid, and the risk of acute coronary events. The Kuopio Ischaemic Heart Disease Risk Factor Study. Circulation 102:2677, 2000

8. Hu FB, Bronner L, Willett WC, et al: Fish and omega-3 fatty acid intake and risk of coronary heart disease in women. JAMA 287:1815, 2002

9. He K, Rimm EB, Merchant A, et al: Fish consumption and risk of stroke in men. JAMA 288:3130, 2002

10. Kang JX, Leaf A: The cardiac antiarrhythmic effects of polyunsaturated fatty acid. Lipids 31:S41, 1996

11. Thies F, Garry JM, Yaqood P, et al: Association of n-3 polyunsaturated fatty acids with stability of atherosclerotic plaques: a randomized controlled trial. Lancet 361:477, 2003

12. de Lorgeril M, Salen P, Martin JL, et al: Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. Circulation 99:779, 1999

13. Hu FB, Stampfer MJ, Manson JE, et al: Dietary intake of α-linolenic acid and risk of fatal ischemic heart disease among women. Am J Clin Nutr 69:890, 1999

14. Singh RB, Dubnov G, Niaz MA, et al: Effect of an Indo-Mediterranean diet on progression of coronary artery disease in high risk patients (Indo-Mediterranean Diet Heart Study): a randomized single-blind trial. Lancet 360:1455, 2002

15. Lichtenstein AH, Ausman LM, Jalbert SM, et al: Effects of different forms of dietary hydrogenated fats on serum lipoprotein cholesterol levels. N Engl J Med 340:1933, 1999

16. Cummings JH, Bingham SA: Diet and the prevention of cancer. BMJ 317:1636, 1998

17. Kolonel LN, Nomura AMY, Cooney RV: Dietary fat and prostate cancer: current status. J Natl Cancer Inst 91:414, 1999

18. Velie E, Kulldorff M, Schairer C, et al: Dietary fat, fat subtypes, and breast cancer in postmenopausal women: a prospective cohort study. J Natl Cancer Inst 92:833, 2000

19. Krauss RM, Eckel RH, Howard B, et al: AHA dietary guidelines: revisions 2000: a statement for healthcare professionals from the nutritional committee of the American Heart Association. Circulation 102:2284, 2000

20. Knopp RH, Walden CE, Retzlaff BM, et al: Long-term cholesterol-lowering effects of 4 fat-restricted diets in hypercholesterolemic and combined hyperlipidemic men. The Dietary Alternatives Study. JAMA 278:1509, 1997

21. Barnard ND, Scialli AR, Berton R, et al: Effectiveness of a low-fat vegetarian diet in altering serum lipids in healthy premenopausal women. Am J Cardiol 85:969, 2000

22. Ornish D, Scherwitz LW, Billings JH, et al: Intensive lifestyle changes for reversal of coronary heart disease. JAMA 280:2001, 1998

23. Hu FB, Willett WC: Optimal diets for prevention of coronary heart disease. JAMA 288:2569, 2002

24. Third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III): Final report. US Department of Health and Human Services; Public Health Service; National Institutes of Health; National Heart, Lung, and Blood Institute. Circulation 106:3143, 2002

25. Frost G, Leeds AA, Dore CJ, et al: Glycaemic index as a determinant of serum HDL-cholesterol concentration. Lancet 353:1045, 1999

26. Spiller RC: Cholesterol, fibre, and bile acids. Lancet 347:415, 1996

27. Bingham SA, Day NE, Luben R, et al: Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. Lancet 361:1496, 2003

28. Peters U, Sinha R, Chatterjee N, et al: Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. Lancet 361:1491, 2003

29. Terry P, Giovannucci E, Michels KB, et al: Fruit, vegetables, dietary fiber, and risk of colorectal cancer. J Natl Cancer Inst 93:525, 2001

30. Fung T, Hu FB, Fuchs C, et al: Major dietary patterns and the risk of colorectal cancer in women. Arch Intern Med 163:309, 2003

31. Salmeròn J, Manson JE, Stampfer MJ, et al: Dietary fiber, glycemic load, and risk of non-insulin-dependent diabetes mellitus in women. JAMA 277:472, 1997

32. Chandalia M, Garg A, Lutjohann D, et al: Beneficial effects of high dietary fiber intake in patients with type 2 diabetes mellitus. N Engl J Med 342:1391, 2000

33. Ludwig DS, Pereira MA, Kroenke CH, et al: Dietary fiber, weight gain, and cardiovascular disease risk factors in young adults. JAMA 282:1539, 1999

34. Rimm EB, Ascherio A, Giovannucci E, et al: Vegetable, fruit, and cereal fiber intake and risk of coronary heart disease among men. JAMA 275:447, 1996

35. Wolk A, Manson JE, Stampfer ME, et al: Long-term intake of dietary fiber and decreased risk of coronary heart disease among women. JAMA 281:1998, 1999

36. Pedrini MT, Levey AS, Lau J, et al: The effect of dietary protein restriction on the progression of diabetic and nondiabetic renal diseases: a meta-analysis. Ann Intern Med 124:627, 1996

37. Stamler J, Elliott P, Kesteloot H, et al: Inverse relation of dietary protein markers with blood pressure: findings for 10,020 men and women in the INTERSALT study. Circulation 94:1629, 1996

38. Potter J, Langhorne P, Roberts M: Routine protein energy supplementation in adults: systematic review. BMJ 317:495, 1998

39. Wald DS, Law M, Morris JK: Homocysteine and cardiovascular disease: evidence on causality from a meta-analysis. BMJ 325:1202, 2002

40. Rimm EB, Willett WC, Hu FB, et al: Folate and vitamin B6 from diet and supplements in relation to risk of coronary heart disease among women. JAMA 279:359, 1998

41. Giovannucci E, Stampfer MJ, Colditz GA, et al: Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. Ann Intern Med 129:517, 1998

42. Asplund K: Antioxidant vitamins in the prevention of cardiovascular disease: a systematic review. J Intern Med 251:372, 2002

43. Omenn GS, Goodman GE, Thornquist MD, et al: Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease. N Engl J Med 334:1150, 1996

44. Michaelsson K, Lithell H, Vessby B, et al: Serum retinol levels and the risk of fracture. N Engl J Med 348:287, 2003

45. Vitamin supplements. Med Lett Drugs Ther 40:75, 1998

46. Utiger RD: The need for more vitamin D. N Engl J Med 338:828, 1998

47. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss: AREDS Report No. 8. Arch Ophthalmol 119:1417, 2001

48. Elliott P, Stamler J, Nichols R, et al: INTERSALT revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations. BMJ 312:1249, 1996

49. Sacks FM, Svetkey LP, Vollmer WM, et al: Effects on blood pressure of reduced dietary sodium and the dietary approaches to stop hypertension (DASH) diet. N Engl J Med 344:3, 2001

50. Dawson-Hughes B, Harris SS, Krall EA, et al: Effect of calcium and vitamin D supplementation on bone density in men and women 65 years of age or older. N Engl J Med 337:670, 1997

51. Iso H, Stampfer MJ, Manson JE, et al: Prospective study of calcium, potassium, and magnesium intake and risk of stroke in women. Stroke 30:1772, 1999

52. Abbott RD, Curb JD, Rodriguez BL, et al: Effect of dietary calcium and milk consumption on risk of thromboembolic stroke in older middle-aged men: the Honolulu Heart Program. Stroke 27:813, 1996

53. Allender PS, Cutler JA, Follmann D, et al: Dietary calcium and blood pressure: a meta-analysis of randomized clinical trials. Ann Intern Med 124:825, 1996

54. Baron JA, Beach M, Mandel JS, et al: Calcium supplements for the prevention of colorectal adenomas. Calcium Polyp Prevention Study Group. N Engl J Med 340:101, 1999

55. Kristal AR, Cohen JH, Qu P, et al: Associations of energy, fat, calcium, and vitamin D with prostate cancer risk. Cancer Epidemiol Biomarkers Prev 11:719, 2002

56. Giovannucci E, Rimm EB, Wolk A, et al: Calcium and fructose intake in relation to risk of prostate cancer. Cancer Res 58:442, 1998

57. Curhan GC, Willett WC, Speizer FE, et al: Comparison of dietary calcium with supplemental calcium and other nutrients as factors affecting the risk for kidney stones in women. Ann Intern Med 126:497, 1997

58. U.S. Department of Health and Human Services: Recommendations to prevent and control iron deficiency in the United States. MMWR Morb Mortal Wkly Rep 47:1, 1998

59. Danesh J, Appleby P: Coronary heart disease and iron status: meta-analyses of prospective studies. Circulation 99:852, 1999

60. Fang J, Madhavan S, Alderman MH: Dietary potassium intake and stroke mortality. Stroke 31:1532, 2000

61. Whelton PK, He J, Cutler JA, et al: Effects of oral potassium on blood pressure: meta-analysis of randomized controlled clinical trials. JAMA 277:1624, 1997

62. Clark LC, Combs GF Jr, Turnbull BW, et al: Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: a randomized controlled trial. Nutritional Prevention of Cancer Study Group. JAMA 276:1957, 1996

63. Yoshizawa K, Willett WC, Morris SJ, et al: Study of prediagnostic selenium level in toenails and the risk of advanced prostate cancer. J Natl Cancer Inst 90:1219, 1998

64. Mark SD, Qiao YL, Dawsey SM, et al: Prospective study of serum selenium levels and incident esophageal and gastric cancers. J Natl Cancer Inst 92:1753, 2000

65. Joshipura KJ, Hu FB, Manson JE, et al: The effect of fruit and vegetable intake on risk for coronary heart disease. Ann Intern Med 134:1106, 2001

66. Joshipura KJ, Ascherio A, Manson JE, et al: Fruit and vegetable intake in relation to risk of ischemic stroke. JAMA 282:1233, 1999

67. Romieu I, Trenga C: Diet and obstructive lung disease. Epidemiol Rev 23:268, 2001

68. Gillman MW: Enjoy your fruits and vegetables: eating fruits and vegetables protects against the common chronic diseases of adulthood (editorial). BMJ 313:765, 1996

69. DeBoer SW, Thomas RJ, Brekke MJ, et al: Dietary intake of fruits, vegetables, and fat in Olmstead County, Minnesota. Mayo Clin Proc 78:161, 2003

70. Liu S, Stampfer MJ, Hu FB, et al: Whole-grain consumption and risk of coronary heart disease: results from the Nurses' Health Study. Am J Clin Nutr 70:412, 1999

71. Liu S, Manson JE, Stampfer MJ, et al: Whole grain consumption and risk of ischemic stroke in women: a prospective study. JAMA 284:1534, 2000

72. Hu FB, Stampfer MJ, Rimm EB, et al: A prospective study of egg consumption and risk of cardiovascular disease in men and women. JAMA 281:1387, 1999

73. Daviglus ML, Stamler J, Orencia AJ, et al: Fish consumption and the 30-year risk of fatal myocardial infarction. N Engl J Med 336:1046, 1997

74. Burr ML, Fehily AM, Gilbert JF, et al: Effects of changes in fat, fish, and fibre intakes on death and myocardial reinfarctions: diet and reinfarction trial (DART). Lancet 2:757, 1989

75. Albert CM, Hennekens CH, O'Donnell CJ, et al: Fish consumption and risk of sudden cardiac death. JAMA 279:23, 1998

76. Terry P, Lichtenstein P, Feychting M, et al: Fatty fish consumption and risk of prostate cancer. Lancet 357:1764, 2001

77. Augustsson K, Michaud DS, Rimm EB, et al: A prospective study of intake of fish and marine fatty acids and prostate cancer. Cancer Epidemiol Biomarkers Prev 12:64, 2003

78. Kris-Etherton PM, Harris WS, Appel LJ: Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. Circulation 106:2747, 2002

79. Tunstall-Pedoe H: Nuts to you (...and you, and you): eating nuts may be beneficial—though it is unclear why. BMJ 317:1332, 1998

80. Jiang R, Manson JE, Stampfer MJ, et al: Nut and peanut butter consumption and risk

of type 2 diabetes in women. JAMA 288:2554, 2002

81. Neil HA, Silagy CA, Lancaster T, et al: Garlic powder in the treatment of moderate hyperlipidaemia: a controlled trial and meta-analysis. J R Coll Physicians Lond 30:329,1996

82. Keli SO, Hertog MGL, Feskens EJ, et al: Dietary flavonoids, antioxidant vitamins, and incidence of stroke: the Zutphen Study. Arch Intern Med 156:637, 1996

83. Mukamal KJ, Conigrave KM, Mittleman MA, et al: Roles of drinking pattern and type of alcohol consumed in coronary heart disease in men. N Engl J Med 348:109, 2003

84. Rimm EB, Stampfer MJ: Wine, beer, and spirits. Are they really horses of a different color? Circ 105:2806, 2002

85. Hinds TS, West WL, Knight EM, et al: The effect of caffeine on pregnancy outcome variables. Nutr Rev 54:203,1996

86. Newby DE, Neilson JM, Jarvie DR, et al: Caffeine restriction has no role in the management of patients with symptomatic idiopathic ventricular premature beats. Heart 76:355, 1996

87. Hu FB, Rimm EB, Stampfer MJ, et al: Prospective study of major dietary patterns and risk of coronary heart disease in men. Am J Clin Nutr 72:912, 2000

88. Kant AK, Schatzkin A, Graubard BI, et al: A prospective study of diet quality and mortality in women. JAMA 283:2109, 2000

89. Fletcher GF, Balady GJ, Amsterdam EA, et al: Exercise standards for testing and training: a statement for healthcare professionals from the American Heart Association. Circulation 104:1694, 2001

90. Physical Activity Trends, United States 1990-1998. MMWR Morb Mortal Wkly Rep 50:166, 2001

91. Myers J: Exercise and cardiovascular health. Circ 107:e2, 2003

92. Jones NL, Killian KJ: Exercise limitations in health and disease. N Engl J Med 343:632, 2000

93. Castaneda C, Gordon PL, Uhlin KL, et al: Resistance training to counteract the catabolism of a low-protein diet in patients with chronic renal insufficiency: a randomized, controlled trial. Ann Intern Med 135:965, 2001

94. The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults. Med Sci Sports Exerc 30:975, 1998

95. Huonker M, Halle M, Keul J: Structural and functional adaptations of the cardiovascular system by training. Int J Sports Med 17:S164, 1996

96. Hambrecht R, Wolf A, Gielen S, et al: Effect of exercise on coronary endothelial function in patients with coronary artery disease. N Engl J Med 342:454, 2000

97. Gielen S, Schuler G, Hambrecht R: Exercise training in coronary artery disease and coronary vasomotion. Circulation 103:e1, 2001

98. Whelton SP, Chin A, Xin X, et al: Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. Ann Intern Med 136:493, 2002

99. Kokkinos PF, Narayan P, Colleran JA, et al: Effects of regular exercise on blood pressure and left ventricular hypertrophy in African-American men with severe hypertension. N Engl J Med 333:1462, 1995

100. Kelley GA, Kelley KS: Progressive resistance exercise and resting blood pressure: a meta-analysis of randomized controlled trials. Hypertension 35:838, 2000

101. Pollock ML, Franklin BA, Balady GJ, et al: Resistance exercise in individuals with and without cardiovascular disease: an advisory from the committee on exercise, rehabilitation, and prevention, Council on Cardiology, American Heart Association. Circulation 101:828, 2000

102. Sutton AJ, Muir KR, Mockett S, et al: A case-control study to investigate the relation between low and moderate levels of physical activity and osteoarthritis of the knee using data collected as part of the Allied Dunbar National Fitness Survey. Ann Rheum Dis 60:756, 2001

103. Penninx BW, Messier SP, Rejeski WJ, et al: Physical exercise and the prevention of disability in activities of daily living in older persons with osteoarthritis. Arch Intern Med 161:2309, 2001

104. Richards SC, Scott DL: Prescribed exercise in people with fibromyalgia: parallel group randomized controlled trial. BMJ 325:185, 2002

105. Hu FB, Sigal RJ, Rich-Edwards JW, et al: Walking compared with vigorous physical activity and risk of type 2 diabetes in women: a prospective study. JAMA 282:1433, 1999

106. Laaksonen DE, Lakka HM, Salonen JT, et al: Low levels of leisure-time physical activity and cardiorespiratory fitness predict development of the metabolic syndrome. Diabetes Care 25:1612, 2002

107. Irwin ML, Yasui Y, Ulrich CM, et al: Effect of exercise on total and intra-abdominal body fat in postmenopausal women: a randomized controlled trial. JAMA 289:323, 2003

108. Williams PT: Relationship of distance run per week to coronary heart disease risk factors in 8,283 male runners. The National Runners' Health Study. Arch Intern Med 157:191, 1997

109. Rice T, Despres JP, Perusse L, et al: Familial aggregation of blood lipid response to exercise training in the health, risk factors, exercise training, and genetics (HERITAGE) Family Study. Circulation 105:1904, 2002

110. Li N, Wallen NH, Hjemdahl P: Evidence for prothrombotic effects of exercise and limited protection by aspirin. Circulation 100:1374, 1999

111. Bartsch P: Platelet activation with exercise and risk of cardiac events. Lancet 354:1747, 1999

112. Montain SJ, Sawka MN, Wenger CB: Hyponatremia associated with exercise: risk factors and pathogenesis. Exerc Sport Sci Rev 29:113, 2001

113. Blumenthal JA, Babyak MA, Moore KA, et al: Effects of exercise training on older patients with major depression. Arch Intern Med 159:2349, 1999

114. Cassel CK: Use it or lose it: activity may be the best treatment for aging. JAMA 288:2333, 2002

115. McGuire DK, Levine BD, Williamson JW, et al: A 30-year follow-up of the Dallas bed rest and training study: II. Effect of age on cardiovascular adaptation to exercise training. Circulation 104:1358, 2001

116. LaMonte MJ, Eisenman PA, Adams TD, et al: Cardiorespiratory fitness and coronary heart disease risk factors: the LDS Hospital Fitness Institute Cohort. Circulation 102:1623, 2000

117. Ford ES: Does exercise reduce inflammation? Physical activity and C-reactive protein among U.S. adults. Epidemiology 13:561, 2002

118. Protective effect of physical activity on coronary heart disease. MMWR Morb Mortal Wkly Rep 36:426, 1987

119. Berlin JA, Colditz GA: A meta-analysis of physical activity in the prevention of coronary heart disease. Am J Epidemiol 132:612, 1990

120. Lee CD, Blair SN: Cardiorespiratory fitness and stroke mortality in men. Med Sci Sports Exerc 34:592, 2002

121. Batty D, Thune I: Does physical activity prevent cancer? Evidence suggests protection against colon cancer and probably breast cancer. BMJ 321:1424, 2000

122. Friedenreich CM, Bryant HE, Courneya KS: Case-control study of lifetime physical activity and breast cancer risk. Am J Epidemiol 154:336, 2001

123. Yaffe K, Barnes D, Nevitt M, et al: A prospective study of physical activity and cognitive decline in elderly women: women who walk. Arch Intern Med 161:1703, 2001

124. Feskanich D, Willett W, Colditz G: Walking and leisure-time activity and risk of hip fracture in postmenopausal women. JAMA 288:2300, 2002

125. Lee IM, Paffenbarger RS Jr: Associations of light, moderate and vigorous intensity physical activity with longevity. The Harvard Alumni Health Study. Am J Epidemiol 151:293, 2000

126. Andersen LB, Schnohr P, Schroll M, et al: All-cause mortality associated with physical activity during leisure time, work, sports, and cycling to work. Arch Intern Med 160:1621, 2000

127. Ades PA: Cardiac rehabilitation and secondary prevention of coronary heart disease. N Engl J Med 345:892, 2001

128. Stewart KJ, Hiatt WR, Regensteiner J: Exercise training for claudication. N Engl J Med 347:1941, 2002

129. Frankin BA, Bonzheim K, Gordon S, et al: Safety of medically supervised outpatient cardiac rehabilitation exercise therapy: a 16-year follow-up. Chest 114:902, 1998

130. Ades PA, Coello CE: Effects of exercise and cardiac rehabilitation on cardiovascular outcomes. Med Clin North Am 84:251, 2000

131. Wannamethee SG, Shaper G, Walker M: Physical activity and mortality in older men with diagnosed coronary heart disease. Circulation 102:1358, 2000

132. Siscovick DS, Ekelund LG, Johnson JL, et al: Sensitivity of exercise electrocardiography for acute cardiac events during moderate and strenuous physical activity. Arch Intern Med 151:325, 1991

133. Gill TM, DiPietro L, Krumholz HM: Role of exercise stress testing and safety monitoring for older persons starting an exercise program. JAMA 284:342, 2000

134. Lemaitre RN, Siscovic DS, Raghunathan TE, et al: Leisure-time physical activity and the risk of primary cardiac arrest. Arch Intern Med 159:686, 1999

135. Manson JE, Greenland P, LaCroix AZ, et al: Walking compared with vigorous exercise for the prevention of cardiovascular events in women. N Engl J Med 347:716, 2002

136. Wannamethee SG, Shaper AG, Walker M: Changes in physical activity, mortality, and incidence of coronary heart disease in older men. Lancet 351:1603, 1998

137. Hakim AA, Petrovitch H, Burchfiel CM, et al: Effects of walking on mortality among nonsmoking retired men. N Engl J Med 338:94, 1998

138. Pereira MA, Kriska AM, Day RD, et al: A randomized walking trial in postmenopausal women: effects on physical activity and health 10 years later. Arch Intern Med 158:1695, 1998

139. Andersen RE, Wadden TA, Bartlett SJ, et al: Effects of lifestyle activity vs structured aerobic exercise in obese women: a randomized trial. JAMA 281:335, 1999

140. Dunn AL, Marcus BH, Kampert JB, et al: Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: a randomized trial. JAMA 281:327, 1999

141. Sesso HD, Paffenbarger RS Jr, Lee IM: Physical activity and coronary heart disease in men. The Harvard Alumni Health Study. Circulation 102:975, 2000

142. Burke AP, Farb A, Malcom GT, et al: Plaque rupture and sudden death related to exertion in men with coronary artery disease. JAMA 281:921, 1999

143. Albert CM, Mittleman MA, Chae CU, et al: Triggering of sudden death from cardiac causes by vigorous exertion. N Engl J Med 343:1355, 2000

144. Leff JA, Busse WW, Pearlman D, et al: Montelukast, a leukotriene receptor antagonist, for the treatment of mild asthma and exercise-induced bronchoconstriction. N Engl J Med 339:147, 1998

145. Hootman JM, Macera CA, Ainsworth B, et al: Epidemiology of musculoskeletal injuries among sedentary and physically active adults. Med Sci Sports Exer 34:838, 2002

Acknowledgments

Figure 1 Marcia Kammerer.

Figure 2 Talar Agasyan.

V ADULT PREVENTIVE HEALTH CARE

MARK HELFAND, M.D.

Over the past 20 years, prevention has become a major activity in primary care. During a typical day, primary care clinicians spend much of their time managing asymptomatic conditions in which the main goal is to prevent death or complications (e.g., hypertension, hyperlipidemia, osteoporosis). Many chapters in *ACP Medicine* include information on screening or prevention of specific disorders in asymptomatic patients or those at increased risk [*see Table 1*]. This chapter focuses primarily on preventive screening recommendations from the United States Preventive Services Task Force (USPSTF).

Rationale and Evolution of Preventive Care Guidelines

The rationale for delivering preventive care during an office visit is strong. In 2002, life expectancy in the United States was 77.4 years, an all-time high.[1] Behavioral risk factors, including tobacco use, diet, and alcohol use, as well as factors such as hyperlipidemia and hypertension, contributed to the most frequent causes of death [*see Table 2*]. From the viewpoint of clinical preventive services, modifiable risk factors such as these, rather than the diseases they affect, are the true causes of death.[2]

Primary care visits provide an opportunity to assess risk, discuss options, and recommend behaviors and treatments that have been proved to reduce the risks of diseases and death. During 2002, an estimated 558 million visits were made to primary care physicians in the United States, an overall rate of about two visits per person per year.[3] On average, these physicians spent 20 minutes with the patient at each visit.

In 1975, Frame and Carlson published a series of articles that examined the quality of evidence for periodic screening conducted in the routine physical examination.[4] These authors argued that any preventive strategy should meet certain criteria of accuracy and usefulness [*see Table 3*]. The criteria are helpful in understanding the controversy about screening proposals. Several scholars have pointed out that clinical intuition about screening is often wrong, leading to errors in inference about the effects of screening. Some of these logical fallacies and hidden assumptions are now well recognized and even find their way into board examinations [*see Table 4*].

The work of Frame and Carlson gave rise to evidence-based decision making in prevention. The Canadian Task Force on the Periodic Health Exam used independent reviews of the scientific literature and a set of rules to grade the strength of evidence supporting a clinical service.

The USPSTF, founded in 1984, was modeled on the Canadian Task Force. It published its first set of guidelines for clinical preventive services in 1989.[5] The current USPSTF has experts from the specialties of family medicine, pediatrics, internal medicine, obstetrics and gynecology, geriatrics, preventive medicine, public health, behavioral medicine, and nursing.

Other expert panels also make recommendations about prevention [*see Table 5*]. Despite general agreement that recommendations should be evidence based, opinions about the effectiveness of specific preventive services differ. These differences arise because interpretation of the evidence is ultimately a subjective process, especially regarding the balancing of benefits and risks—an equation that includes such disparate factors as mortality reduction, costs or burden of illness, and patient discomfort.

To avoid errors in judging the evidence and weighing benefits and harms, expert panels, as well as individual clinicians, should do the following: (1) use an independent systematic review to distinguish assertions based on evidence from those based on other grounds, (2) make the rationale for a recommendation explicit, and (3) be free from financial and political conflicts of interest. Although the use of these measures does not guarantee a correct decision, they represent the best safeguards against bias.

USPSTF Evidence Ratings

The USPSTF assigns an overall grade of A, B, C, D, or I to each prevention service. The grades reflect the overall strength of evidence and the magnitude of benefit, defined as benefits minus harms [*see Table 6*].[6]

A grade of A indicates services that have solid supporting evidence and at least a moderate net benefit. A grade of B suggests that there are information gaps (so-called fair evidence) or that the benefits are only moderately greater than the harms for all patients. A grade of C denotes a toss-up, whereas a D grade indicates a service that is either proven ineffective or unlikely to provide benefits that outweigh the harms.

When there is too little evidence to determine whether or not a service works, the USPSTF assigns a grade of I for insufficient evidence. Some of the services with an I grade make good clinical sense and some are very promising, but without better research, it is not possible to say with confidence that they improve outcomes. Other grade I services have uncertain benefits but definite harms.

Noncancer Prevention Imperatives

Several preventive measures have earned an A grade on the strength of their good supportive evidence, substantially greater benefits than harms, and broadest applicability to primary care practice [*see Table 7*]. Implementation of these measures is described in detail in other *ACP Medicine* chapters [*see Table 1*].

IMMUNIZATION

The USPSTF has not issued recommendations about immunization since 1996, and those recommendations are now out of date. The Advisory Committee on Immunization Practices (ACIP), which consists of 15 experts in fields associated with immunization, is currently the only entity in the United States federal government that makes recommendations about immunizations. In contrast to the USPSTF, the ACIP does not use systematic reviews and does not usually describe the quality of evidence supporting a recommendation.

The ACIP publishes schedules for vaccination against certain infectious diseases in adults, depending on age and risk factors; these recommendations are discussed in individual *ACP Medicine* chapters and are available on the Internet (http://www.cdc.gov/nip/recs/adult-schedule.pdf). For example, general recommendations include a tetanus-diphtheria booster every 10

Table 1 Selected Prevention-Related Content in *ACP Medicine*

Chapter	*Relevant Content*
CE:III Reducing Risk of Injury and Disease	Alcohol, tobacco, and other drug abuse; injury; violence; cites USPSTF
CE:IV Diet and Exercise	AHA recommendations
CE:VII Health Advice for International Travelers	CDC-recommended pretravel immunizations, other prophylactic measures
1:III Hypertension	Prevention
1:VIII Acute Myocardial Infarction	Secondary prevention; drugs and risk-factor modification
1:XI Valvular Heart Disease	Prophylactic drug therapy for endocarditis; drugs and surgery for valvular disease
1:XVIII Venous Thromboembolism	Primary and secondary prophylaxis
2:V Contact Dermatitis and Related Disorders	Prevention
2:X Malignant Cutaneous Tumors	Prevention
3:VI Diseases of Calcium Metabolism and Metabolic Bone Disease	Osteoporosis prevention
4:VII Acute Viral Hepatitis	Immunization
6:XIII Urticaria, Angioedema, and Anaphylaxis	Prevention of anaphylaxis
6:XV Allergic Reactions to Hymenoptera	Prevention
7:I Infections Due to Gram-Positive Cocci	Prevention of spread of staphylococcal infection
7:II Tuberculosis	Tuberculosis prevention
7:III Infections Due to Neisseria	Prophylaxis for meningococcal disease
7:IV Infections Due to Gram-Positive Bacilli	Prevention of diphtheria, listeria, anthrax
7:V Anaerobic Infections	Tetanus prevention
7:VII Lyme Disease and Other Spirochetal Zoonoses	Prevention of Lyme disease
7:X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas	*H. influenzae* immunization and secondary prevention; pertussis immunization
7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella	Prevention of Rocky Mountain spotted fever, typhus
7:XVIII Infective Endocarditis	Prevention for high-risk patients/procedures
7:XXII Vaginitis and Sexually Transmitted Diseases	Screening recommendations for sexually transmitted diseases
7:XXV Respiratory Viral Infections	Prevention of influenza and other respiratory viral infections
7:XXVI Herpesvirus Infections	Prevention of herpes simplex, varicella-zoster, and cytomegalovirus infections
7:XXVIII Enteric Viral Infections	Polio prevention
7:XXIX Measles, Mumps, Rubella, Parvovirus, and Poxvirus	MMR, smallpox vaccination

years in all adults, influenza vaccination every year in adults 50 years of age and older, and pneumococcal vaccination once in adults 65 years and older. The ACIP has made specific recommendations for vaccination of health care workers [*see Table 8*].[7]

Cancer Prevention

Only two cancer screening tests meet the USPSTF criteria for a strong recommendation: (1) Papanicolaou (Pap) smears for cervical cancer and (2) fecal occult blood testing or endoscopic procedures for colorectal cancer [*see Table 9*]. With both of these conditions, the aim of screening is to remove precancerous lesions, which prevents invasive cancer, saves the involved organ, and reduces disease-specific mortality. By contrast, the more controversial cancer screening tests, such as prostate-specific antigen (PSA) and mammography, detect invasive cancers and lead to aggressive treatments (prostatectomy and mastectomy) that often destroy the involved organ and that have more substantial morbidity than cone biopsy for cervical cancer and polypectomy for colorectal cancer.

CERVICAL CANCER

Although no data from randomized, controlled trials support the value of the Pap smear in reducing mortality from cervical cancer, indirect evidence suggests that it is among the most effective cancer screening techniques.[8] By current standards, the sensitivity of traditional Pap testing is low (51%).[9,10] Cervical dysplasia is slow to progress to invasive carcinoma, however, so periodic screening can make up for the low sensitivity of a single exam.

The USPSTF recommends screening with Pap smears at least

Table 1 (*continued*)

Chapter	*Relevant Content*
7:XXXI Viral Zoonoses	Vaccination for yellow fever, Japanese encephalitis, and rabies
7:XXXIII HIV and AIDS	Prevention of HIV infection
7:XXXIV Protozoan Infections	Prevention of malaria, toxoplasmosis, giardiasis, and amebiasis
7:XXXVIII Mycotic Infections in the Compromised Host	Prevention of several opportunistic fungal infections
8:II Bites and Stings	Prophylactic antibiotics for bites
8:IV Preoperative Assessment and Care of the Surgical Patient	Assessing operative risk and preventing complications
8:V Bioterrorism	Vaccination and postexposure prophylaxis
8:VIII Assessment of the Geriatric Patient	Evidence-based preventive services, assessment of in-home hazards
8:IX Management of Common Clinical Disorders in Geriatric Patients	Prevention of falls, incontinence, pressure ulcers, malnutrition, and iatrogenic illness
8:X Rehabilitation of Geriatric Patients	Prevention of stroke complications
9:II Diagnosis and Treatment of Dyslipidemia	Primary and secondary prevention
9:VI Diabetes Mellitus	ADA screening recommendations, prevention of type 2 diabetes, prevention of diabetic complications
10:VI Acute Renal Failure	Prevention
10:XII Nephrolithiasis	Prevention of recurrent kidney stones
11:IV Cerebrovascular Disorders	Risk reduction for stroke
11:VIII Headache	Migraine prophylaxis
12:I Cancer Epidemiology and Prevention	Screening of asymptomatic patients for prevention and early detection; ACS recommendations
12:V Colorectal Cancer	Risk reduction, screening tests; ACS recommendations
12:VII Breast Cancer	Screening and prophylaxis
12:VIII Lung Cancer	Prevention
12:IX Prostate Cancer	Risk reduction, screening; ACS recommendations
13:III Alcohol Abuse and Dependency	Screening for alcoholism and treatment to prevent relapse
16:IX Medical Problems in Pregnancy	Limited discussion of screening
16:XI Menopause	Prevention and screening per USPSTF
16:XVI Approach to the Patient with an Abnormal Pap Smear	Prevention and screening per USPSTF

ACS—American Cancer Society ADA—American Diabetes Association AHA—American Heart Association CDC—Centers for Disease Control and Prevention MMR—measles, mumps, rubella vaccine USPSTF—United States Preventive Services Task Force

every 3 years, beginning within 3 years after the start of sexual activity or age 21 (whichever comes first). The Task Force recommends against annual screening. In women who have had consistently negative Pap smear results, continuing screening past age 65 is unnecessary because of the declining incidence of high-grade cervical lesions and an increased risk for potential harms, including false positive results and invasive procedures. The USPSTF also recommends against routine Pap smear screening in women who have had a total hysterectomy for benign disease.

The specificity of Pap smears for detection of dysplasia and cancer is 98%. False positive results occur infrequently, but Pap smears may correctly detect a large number of low-grade lesions that, without treatment, would remain stable or regress.[11] As a consequence, many women who would never develop invasive cervical cancer are subjected to anxiety and to colposcopy and biopsy.

In a systematic review, the effectiveness of liquid-based cytology, computerized rescreening, and algorithm-based screening have been compared with that of conventional Pap smear screening in reducing the incidence and mortality of invasive cervical cancer. The review concluded that the liquid-based monolayer preparation (ThinPrep) appears to offer higher sensitivity but lower specificity than conventional Pap smears.[10] However, the USPSTF could not determine whether the potential benefits of the three new screening approaches relative to conventional Pap smears are sufficient to justify a possible increase in potential harm or cost. They also found insufficient evidence to recommend for or against the routine use of human papillomavirus testing as a primary screening test for cervical cancer.

COLORECTAL CANCER

Screening modalities for colorectal cancer include fecal occult blood testing (FOBT), sigmoidoscopy, double-contrast barium enema, colonoscopy, and computed tomographic colonogra-

Table 2 Major Causes of Death in the United States[1]*

Cause of Death	*Number of Deaths*	*Age-Adjusted Death Rate (per 100,000 population)*
Diseases of the heart	695,754	204.4
Malignant neoplasms	558,847	194
Cerebrovascular diseases	163,010	56.3
Chronic lower respiratory diseases	125,500	43.7
Accidents (unintentional injuries)	102,303	35.3
Diabetes mellitus	73,119	25.4
Influenza and pneumonia	65,984	22.7
Alzheimer disease	58,785	20.2

*Preliminary data for 2002; these causes account for three quarters of all deaths.

phy [*see 12:V Colorectal Cancer*]. FOBT is the only screening modality that has been shown in randomized controlled trials to reduce colorectal cancer mortality. In the Minnesota Colon Cancer Control Study, 33 volunteers 50 to 80 years of age were randomized to annual FOBT, biennial FOBT, or a control group. After 18 years of follow-up, colorectal cancer mortality was 33% lower in the annual FOBT group and 21% lower in the biennial group than in the control group.[12] In this study, the slides were rehydrated, a technique that increases sensitivity but reduces specificity; during the trial, 38% of patients in the annual FOBT group underwent colonoscopy because of a positive test result. Two randomized, controlled trials from Europe have demonstrated 16% and 18% reductions in colorectal cancer mortality using FOBT.[13,14] In the European trials, unlike in the Minnesota study, patients were drawn from the general population, the slides were not rehydrated, and all testing was biennial.

In the screening trials, FOBT reduced mortality from colon cancer but did not reduce all-cause mortality. For example, the Minnesota trial findings indicate that 10 years of screening would result in 12 (95% confidence interval, 1 to 24) fewer colon cancer deaths per 10,000 persons screened. In that trial, however, the 95% confidence interval for all-cause mortality was 334 to 350 with annual screening, 333 to 348 with biennial screening, and 336 to 351 in control subjects.[12]

Evidence for the efficacy of sigmoidoscopy comes from case-control studies, which suggest that the protective effect of a single sigmoidoscopy lasts at least 6 years. The results of a large United Kingdom trial of screening with flexible sigmoidoscopy are not yet complete. Preliminary results suggest that flexible sigmoidoscopy is safe and that about 5% of persons 55 to 64 years of age have high-risk polyps (three or more adenomas; size 1 cm or greater; villous, severely dysplastic, or malignant).[15]

Because of the imperfect sensitivity of FOBT and sigmoidoscopy and because many patients who undergo these procedures end up requiring colonoscopy anyway, many clinicians are advising their average-risk patients to undergo colonoscopy as a screening test, either as a one-time procedure or periodically (e.g., every 10 years) beginning at age 50. Colonoscopy is the most sensitive test for detecting polyps; however, as for other slow-growing lesions, such as cervical dysplasia, it is not clear whether improved sensitivity for polyps at a single point in time will translate into fewer invasive cancers in the long run.

In 2002, for the first time, the USPSTF included screening colonoscopy as an option, but with the qualification that the potential added benefits of colonoscopy may not always be great enough to justify the increased risks and inconvenience.[16] All colon cancer screening tests have a low yield—over 500 patients must be screened to prevent one invasive cancer[17]—so even a slightly increased rate of serious complications with colonoscopy might negate the benefit. Several gaps in the evidence base for colonoscopy can also be mentioned. First, the frequency of one procedure every 10 years was arrived at by means of mathematical models; in fact, no one knows how many patients will develop invasive cancer less than 10 years after a negative colonoscopy. Second, surveys suggest that gastroenterologists overuse colonoscopy for surveillance in patients who have clinically insignificant hyperplastic polyps or low-risk lesions, such as small adenomas. As a result, colonoscopic screening may lead to the use of a scarce, expensive resource, primarily in patients who have little chance of benefit. Third, the accuracy of colonoscopy when performed by the so-called average colonoscopist is not known. The primary advantage of colonoscopy, visualization of the entire colon, is negated if the operator cannot reach the cecum consistently or does not view the entire circumference of the lumen during the procedure.

No direct evidence supports the use of double-contrast barium enema for screening, and patients find it more uncomfortable than other alternatives. CT colonography may prove to be more sensitive and better tolerated than double-contrast barium enema and safer than colonoscopy; as of yet, however, there are insufficient data to determine whether it would result in better outcomes.[18]

BREAST CANCER

It was predicted that in the United States in 2004, invasive breast cancer would be diagnosed in an estimated 215,990 women; in situ disease would be diagnosed in 55,700 women; and 40,110 women would die of the disease.[19] A 40-year-old woman has a 13.2% (approximately one in eight) chance of developing invasive breast cancer during her life, but her risk of developing breast cancer within 10 years is only 1.47% (approximately one in 68). Modalities for breast cancer screening include mammography, clinical breast examination, and breast self-examination.

Table 3 Criteria for Evaluating a Screening Program

1. Does the program target a disease that causes serious morbidity and mortality that might be prevented by the service?
2. Can the screening test accurately identify healthy people who are at high risk for developing advanced disease?
3. Is the screening test feasible to use in primary care?
4. Does treatment given before symptoms occur result in better outcomes than treatment given later?
5. Do the overall benefits outweigh the harms of screening and treatment?

Table 4 Sample Board Examination Questions About Screening

Question	*Answer and Explanation*
A screening test correctly identifies 95% of patients who have prexerostosis and 95% of patients who are well. If 1 of every 500 patients has prexerostosis, what is the likelihood that a patient who has a positive test has the disease?	The correct answer is 3%; the positive predictive value is commonly overestimated because of neglecting Bayes theorem.
The 5-year survival of stage 0 lung cancer is 95%, versus 10% for more advanced stages. In usual care, 70% of patients present in advanced stages. When screening with a CT scan, 90% of patients have stage 0 disease. By how much will screening reduce mortality?	The correct answer is that the effect of screening on survival cannot be determined; increasing detection of disease in a "curable stage" may improve 5-yr survival but does not necessarily reduce mortality because of overdiagnosis bias, length bias, and lead-time bias—for example, screening may detect slower-growing cancers that would never have become lethal.
With improvements in treatment, mortality from advanced HIV infection has dropped by 63%. Because effective treatment is now available, screening and early treatment should result in even greater mortality reductions. True or false?	The conclusion may be, but is not necessarily, true. If treatment of advanced disease is very effective, screening may not confer any additional advantage. Screening is most likely to improve outcomes when advanced disease is untreatable but treatment of earlier, asymptomatic disease can result in cure.

Mammography

In 2000, a Danish meta-analysis of the major randomized trials of mammography concluded that there was no evidence that mammography reduced mortality from breast cancer.[20] However, another analysis of the same trials conducted for the USPSTF concluded that mammography reduced breast cancer mortality in women 40 to 70 years of age.[21] The controversy centered on disagreement about the quality of the randomized trials of mammography: the Danish investigators excluded five of the eight trials that showed mammography to be beneficial, whereas the United States investigators excluded only two of those eight trials on grounds of quality.

The USPSTF demoted mammography from grade A to grade B to reflect their view that the quality of the evidence was fair and that the net benefit (benefits minus harms) was moderate. Coming after the widely publicized Danish study, the USPSTF recommendation of grade B for mammography received a mixed reception. One independent review, published in 2003, confirmed the USPSTF view that although the trials were flawed, the balance of the evidence still favored screening mammography in women 40 years of age and older at least every 2 years.[22] Conversely, the National Cancer Institute's Physician Data Query program largely endorsed the idea that most of the mammography trials were seriously flawed.

The USPSTF's most controversial decision regarding mammography was to promote screening in women 40 to 50 years of age from a grade C to a grade B. This was done because with several additional years of follow-up since the previous recommendations, in 1996, the pooled risk reduction for women who began screening at this age had become statistically significant. Nevertheless, the number needed to screen is higher, and the balance of benefits and harms narrower, in women 40 to 50 years of age than in older women.

For clinicians, the most difficult question is how to present information about the risks and benefits clearly and fairly to patients. At the time of an earlier controversy over the effectiveness of mammography in women 40 to 49 years of age, a survey of 509 women in the United States found that most believed the

Table 5 Government-Sponsored Preventive Guidelines Programs

Organization	*Sponsorship*	*Focus*	*Web Site*
U.S. Preventive Services Task Force	Agency for Healthcare Research and Quality	Clinical preventive services	http://www.ahrq.gov/clinic/uspstfix.htm
Canadian Task Force on Preventive Health Care	Health Canada (Canadian Federal Government)	Clinical prevention, periodic health examination	http://www.ctfphc.org
Physician Data Query Program	National Cancer Institute	Cancer prevention	http://cancernet.nci.nih.gov/cancertopics/pdq/screening
Task Force on Community Preventive Services	CDC	Community, population, and health care system strategies	http://www.thecommunityguide.org
Advisory Committee on Immunization Practices	DHHS and CDC	Immunizations, bioterrorism response	http://www.cdc.gov/nip/recs/adult-schedule.pdf
National Heart, Lung, and Blood Institute	National Institutes of Health	Asthma, cholesterol, hypertension, obesity	http://www.nhlbi.nih.gov/guidelines
Board on Health Promotion and Disease Prevention	Institute of Medicine	Population-based public health measures and the public health infrastructure	http://www.iom.edu/board.asp?id=3793

CDC—Centers for Disease Control and Prevention DHHS—Department of Health and Human Services

Table 6 United States Preventive Services Task Force Grading System[6]

*Grade**	*Strength of Evidence*	*Magnitude of Benefit*
A	Good	Large
B	Good Fair	Moderate Moderate to large
C	Fair to good	Small
D	Fair to good	None
I	Poor	None to large

*A—Service strongly recommended B—Service recommended
C—No recommendation for or against D—Service not recommended
I—Insufficient evidence

controversy was really about cost.[23] Women may interpret the lifetime risk of one in eight to be their immediate risk of developing breast cancer if they defer or miss their next mammogram.[24] In deciding how to inform patients, clinicians should carefully consider the major criticisms of the USPSTF recommendation. These criticisms represent differences in values rather than disagreements over the facts. There are four principal issues:

1. Is reducing breast cancer mortality important? In the trials, which involved nearly half a million women, mammography clearly had no effect on all-cause mortality. The USPSTF, although fully aware of this fact, chose to base their assessment of the benefits on the narrower grounds of breast cancer mortality. They chose to let women decide for themselves whether reducing the risk of dying of breast cancer was important to them.
2. How large is the reduction in breast cancer mortality? Judging from the trials, about 1,200 women 40 to 70 years of age must be invited to be screened four to five times over 10 years to prevent one death from breast cancer. Of women 40 to 49 years of age, 1,792 (95% CI, 764 to 10,540) must be invited to be screened to prevent one death from breast cancer, a death that would not have occurred until about 20 years after screening began. The specification "invited to" is important: it is likely that the trials underestimated the true benefit because they are diluted by a large number of subjects who were assigned to have mammography but did not.[25] Nevertheless, to benefit even one woman, a large number of women must have a large number of mammograms over many years.
3. Are these estimates from the randomized trials still valid? Evidence from the trials may be out of date**.** The first trial began in 1963, and the others began between 1976 and 1982. Improvements in mammography since then might translate into better outcomes than were seen in the trials. On the other hand, improved systemic treatment for clinically detected breast cancer may have eliminated the advantage that earlier detection conferred in the era of the trials.
4. How large are the harms? The USPSTF was criticized for ignoring or underestimating harms. In fact, the USPSTF considered the harms, but it was also influenced by evidence that many healthy women stated that they would be willing to take on these risks, as well as the morbidity associated with treatments, to avoid a breast cancer death.[26]

What are the harms? Women who get 10 annual mammograms have about a 50% chance that at least one of them is a false positive result; many of these false positive results necessitate a biopsy. Of women who are found to have invasive cancer, about 30 must undergo major surgery, or surgery plus radiation or tamoxifen, to prevent one death from breast cancer. In addition, screening identifies many women with ductal carcinoma in situ, and many of these women also undergo surgery, with uncertain benefit. In sum, many women experience immediate morbidity from treatment; without screening, most of them would not have had consequences of their breast cancer (and no morbidity from mastectomy) for many years, if ever.

Table 7 Strongly Recommended Noncancer Preventive Services in Adults*

Service	*Candidates*	*Established Benefits*
Aspirin for primary prevention of cardiovascular events	Adults at high cardiovascular risk	Reduces the risk of stroke
Blood pressure screening	All adults	Reduces the risk of stroke
Screening for lipid disorders	Men 35 yr of age and older; women 45 yr of age and older; and younger adults at increased risk for coronary artery disease	Reduces overall mortality, as well as mortality from cardiovascular disease
Chlamydial infection screening	Sexually active women 25 yr of age and younger; other asymptomatic women at increased risk for infection	Reduced the risk of pelvic inflammatory disease in one randomized trial[36]
Hepatitis B virus (HBV) infection screening	Pregnant women	Reduces prenatal transmission of HBV
Syphilis screening	Persons at increased risk for infection; all pregnant women	Penicillin treatment during pregnancy reduces the risk to the fetus of acquiring congenital syphilis
HIV screening†	Pregnant women High-risk men and women	Reduces prenatal transmission of HIV Delays mortality from HIV disease and permits counseling to reduce transmission
Screening for asymptomatic bacteriuria	Pregnant women (urine culture at 12–16 weeks' gestation)	Prevents symptomatic urinary tract infections, low birth weight, and preterm delivery

*As per the United States Preventive Services Task Force. †As per the CDC.[37,38]

Table 8 Recommended Vaccination Schedule for Health Care Workers[7]*

Vaccine	*Schedule*
Tetanus-diphtheria	Every 10 yr after complete primary series or for persons lacking documentation of vaccination
Influenza	Annual
Pneumococcal (polysaccharide)	For persons with medical indications† or at risk for exposure
Hepatitis B	For persons lacking documentation of vaccination or evidence of disease
Hepatitis A	No data to support a recommendation
Measles, mumps, rubella (MMR)	For persons lacking documentation of vaccination or history of disease‡
Varicella	For persons lacking documentation of vaccination or history of disease

*As per the Advisory Committee on Immunization Practices.

†*Medical indications:* Chronic pulmonary disorders, excluding asthma; cardiovascular disease; diabetes mellitus; chronic liver disease, including liver disease as a result of alcohol abuse (i.e., cirrhosis); chronic renal failure; chronic renal failure or nephrotic syndrome; functional or anatomic asplenia (e.g., sickle cell disease or splenectomy); lymphoma, multiple myeloma, generalized malignancy, or organ or bone marrow transplantation; chemotherapy with alkylating agents, antimetabolites, or long-term systemic corticosteroids; or cochlear implants.

‡*Measles component:* Adults born before 1957 can be considered immune to measles. Health care workers born during or after 1957 should receive two doses of MMR vaccine unless they have a medical contraindication, documentation of one or more dose, or other acceptable evidence of immunity. *Mumps component:* One dose of MMR vaccine should be adequate for protection. *Rubella component:* Administer one dose of MMR vaccine to women whose rubella vaccination history is unreliable, and counsel women to avoid becoming pregnant for 4 wk after vaccination. For women of childbearing age, routinely determine rubella immunity and counsel regarding congenital rubella syndrome. Do not vaccinate pregnant women or those planning to become pregnant during the next 4 wk. For women who are pregnant and susceptible, vaccinate as early as possible in the postpartum period.

Clinical Breast Examination

The USPSTF could not determine the benefits of clinical breast examination (CBE) alone or the incremental benefit of adding CBE to mammography (grade I recommendation). No screening trial has examined the benefits of CBE alone (without accompanying mammography). Four of the eight trials of screening used mammography alone, and four used mammography plus CBE. In the trials that used both methods, CBE detected 40% to 69% of breast cancers. It is not clear from the trials whether CBE contributed to the reduction in breast cancer mortality observed in some of the trials.

Breast Self-examination

A randomized trial from China failed to show a reduction in breast cancer mortality or an improvement in tumor stage at presentation in women receiving instruction in breast self-examination.[27] Results from a Russian trial were similar.[28] In both trials, women who had been instructed in breast self-examination were more likely to seek medical advice for benign breast lesions.

Genetic Risk Assessment

In women whose family history suggests an increased risk of deleterious *BRCA1* or *BRCA2* mutations, the USPSTF recommends referral for genetic counseling and evaluation for *BRCA* testing (grade B recommendation). However, the USPSTF recommends against routine testing for breast cancer susceptibility genes (i.e., *BRCA1* or *BRCA2*) or routine referral for genetic counseling in women whose family history does not suggest an increased risk of deleterious mutations in these genes (grade D recommendation). Such screening and counseling have few or no benefits and could have important adverse ethical, legal, social, and medical consequences.

Cancer Screening Measures That Are Not Recommended

The USPSTF recommended against screening for bladder, ovarian, pancreatic, and testicular cancers. In each case, the deciding factor was that screening and treatment caused serious, immediate harms, whereas evidence of a benefit was inconclusive. As with mammography for breast cancer, screening for

Table 9 Recommended and Strongly Recommended Measures for Cancer Prevention*

Service	*Recommendation Grade*	*Candidates*	*Comment*
Cervical cancer screening	A	Women who have been sexually active and have a cervix; begin screening within 3 yr of onset of sexual activity or at age 21 (whichever comes first) and screen at least every 3 yr, stopping at age 65	Reduces the risk of invasive cervical cancer and mortality from cervical cancer
Colorectal cancer screening	A	Adults 50 yr and older (earlier in patients with a strong family history)	Reduces the risk of invasive colon cancer and mortality from colon cancer
Breast cancer screening	B	Women 40 yr and older	Reduces mortality from breast cancer
Breast cancer chemoprophylaxis	B	Women at high risk for developing breast cancer	Reduces the incidence of invasive breast cancer

*As per the United States Preventive Services Task Force [*see Table 6*].

Table 10 Recommended Preventive Noncancer Screening*

Condition	*Screening Measure*	*Comments*
Abdominal aortic aneurysm	Abdominal palpation, ultrasonography	Men 65 to 70 yr of age who have ever smoked should be screened one time by ultrasonography
Depression	Standardized questionnaire	In most trials, screening alone had nonsignificant effects on treatment rates and on clinical outcome; however, larger benefits were observed in studies in which the communication of screening results was coordinated with effective follow-up and treatment; in such settings, 110 patients would need to be screened to produce one additional remission after 6 mo of treatment
Obesity	Measurement of body mass index (BMI)	Screening can identify obesity (BMI ≥ 30 kg/m^2); programs that combined diet and physical activity produced modest weight loss (6.4 lb on average for 1 yr or more); most trials did not report the proportion of subjects who lost weight
Osteoporosis	Dual–energy x-ray absorptiometry	Women older than 65 yr and high-risk women 50 yr of age and older should be screened; alendronate reduces the risk of fracture over 3–5 yr, but the longer-term benefit of treatment is unclear

*"B" recommendations, United States Preventive Services Task Force.

these cancers is aimed at detection of early invasive disease, and treatment has substantial morbidity. This degree of morbidity is in contrast to that associated with screening for colonic polyps or cervical dysplasia, for which treatment is relatively safe and is aimed at preserving, rather than removing, the involved organ.

Prostate Cancer Screening

The USPSTF concluded that evidence was insufficient to recommend for or against prostate cancer screening. This conclusion was based on the following considerations: (1) there are no completed randomized, controlled trials of screening, although studies are ongoing in the United States[29] and in Europe[30]; (2) although prostate cancer is a major cause of cancer death in men, many cases are clinically indolent (in autopsy studies, the prevalence of histologic prostate cancer in men older than 50 years is about 30%, but only 3% of men die of prostate cancer)[31]; (3) the value of treatment for the localized cancers targeted by screening is unknown; the one randomized, controlled trial of radical prostatectomy, which found no improvement in the 15-year survival rates of patients undergoing surgery, has been criticized for methodological problems[32] (another randomized, controlled trial comparing expectant management with radical prostatectomy for the treatment of localized cancer is under way)[33]; (4) aggressive treatments for localized disease are associated with significant morbidity; and (5) mortality from prostate cancer has not declined in the United States despite 15 years of widespread use of PSA testing.

Lung Cancer Screening

The USPSTF concluded that evidence was insufficient to recommend for or against screening asymptomatic patients for lung cancer with low-dose CT, chest x-ray, sputum cytology, or a combination of these tests. Although there is fair evidence that screening with these measures can result in detection of lung cancer at an earlier stage, there is poor evidence that any screening strategy for lung cancer decreases mortality. Moreover, the invasive nature of diagnostic testing and the possibility of a high number of false positive tests in certain populations raises the potential for significant harms from screening.

Noncancer Screening

Selected screening tests for diseases other than cancer are recommended for all adults, or for groups defined by age and sex. These diseases include abdominal aortic aneurysm in older men, depression, obesity, and osteoporosis [*see Table 10*].

Behavioral-Counseling Interventions

Unhealthy behaviors have a huge impact on mortality and morbidity. Tobacco use remains the leading preventable cause of death in the United States, contributing to more than 440,000 deaths each year. Misuse of alcohol is responsible for 100,000 more deaths. Although tobacco use has decreased, alcohol abuse, obesity, and diabetes have increased in recent years, bringing new attention to the need to eat, drink, and exercise sensibly.

The evidence base supporting brief counseling by primary care physicians has grown substantially in the past 10 years. To date, however, efficacy has been proved only for counseling on tobacco cessation and alcohol use [*see Table 11*]. Evidence to support counseling on diet, exercise, and other behaviors (e.g., use of sunscreens, seat-belt use) is limited. In many instances, follow-up in the available studies was too short to confirm that behavior change is sustained long enough to reduce the risk of developing disease or injury.

Table 11 Selected Recommendations for Counseling and Patient Education

Counseling Topic	*USPSTF Grade**
Tobacco use	B
Alcohol use/driving after drinking	B
Healthy diet	I
Physical activity	I
Seat-belt use	I
Regular dental care	I
Avoidance of sun exposure/use of protective clothing	I
Adequate calcium intake (women)	I
Use of sunscreens	I

*See Table 6.
USPSTF—United States Preventive Services Task Force

SMOKING CESSATION

There is strong evidence that smoking bans, increasing the price of tobacco products, and public-information campaigns can discourage people from starting to smoke and encourage them to stop. Smoking cessation rapidly decreases the risk of stroke and heart disease and slowly decreases the risk of lung cancer [*see CE:III Reducing Risk of Injury and Disease*]. In patients with peripheral vascular disease, smoking cessation reduces the risk of limb amputation and recurrent stroke.

Brief counseling by clinicians can help smokers take action. Counseling by physicians becomes increasingly important as more patients become motivated to quit. Because many patients have tried and failed before, brief messages should emphasize that repeated efforts often bring success.

ALCOHOL USE

Screening and counseling of alcohol use in primary care is aimed at drinkers who are at risk for harm from alcohol consumption that exceeds daily, weekly, or per-occasion norms (i.e., risky or hazardous drinking) [*see 13:III Alcohol Abuse and Dependency*]. Unlike harmful drinking and alcohol abuse or dependence, risky drinking behavior has not yet resulted in physical, social, or psychological harm to the drinker, and such drinkers do not meet diagnostic criteria for alcohol dependence.[34] In contrast to persons who engage in risky drinking, alcohol-abusing and alcohol-dependent drinkers may require intense addiction treatment and are unlikely to respond to brief advice from a physician.

Self-administered questionnaires or brief interviews can be used to assess average quantity or frequency and binge use. In the United States, about 8% to 18% of patients screen positive for binge drinking. CAGE is a four-item screening questionnaire to detect alcohol abuse and dependence. Its name derives from the topics of the four questions: Have you ever felt you ought to **C**ut down on drinking? Have people **A**nnoyed you by criticizing your drinking? Have you ever felt bad or **G**uilty about your drinking? Have you ever had a drink in the morning to steady your nerves or get rid of a hangover (**E**ye-opener)? In contrast, the Alcohol Use Disorders Identification Test (AUDIT), a 10-item instrument, is designed to identify risky and harmful use. In several controlled trials conducted in primary care settings, it was found that brief, multicontact behavioral-counseling interventions reduced risky and harmful alcohol use. About one in 10 risky drinkers reduced their alcohol use to sensible levels for up to 1 year.[35]

Reminder Systems

The USPSTF has created patient pocket guides that are based on its guidelines and that clinicians can use as reminder systems to promote patients' involvement in their own preventive care. These pocket guides are available on the Internet. There is one for all adults (http://www.ahrq.gov/ppip/adguide), one for adults older than 50 years (http://www.ahrq.gov/ppip/50plus/index.html), and one for women (http://www.ahrq.gov/ppip/healthywom.htm).

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Kochanek KD, Smith BD: Deaths: Preliminary Data for 2002. National Center for Health Statistics, Hyattsville, Maryland, 2004
http://www.cdc.gov/nchs/data/nvsr/nvsr52/nvsr52_13.pdf

2. McGinnis M: Actual causes of death in the United States. JAMA 270:2207, 1993

3. Woodwell DA, Cherry DK: National Ambulatory Medical Care Survey: 2002 Summary. Advance data from vital and health statistics; no. 346. National Center for Health Statistics, Hyattsville, Maryland, 2004
http://www.cdc.gov/nchs/data/ad/ad346.pdf

4. Frame PS, Carlson SJ: A critical review of periodic health screening using specific criteria. J Fam Pract 2:29, 1975

5. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions, 1st ed. US Preventive Services Task Force. Williams & Wilkins, Baltimore, 1989

6. Harris R, Helfand M, Woolf SH, et al: Current methods of the third US Preventive Services Task Force: a review of the process. Am J Prev Med 20:21, 2001

7. Recommended adult immunization schedule—United States, October 2004–September 2005. MMWR Morb Mortal Wkly Rep 53:Q1, 2004
http://www.cdc.gov/nip/recs/adult-schedule.pdf

8. Guzick D: Efficacy of screening for cervical cancer: a review. Am J Public Health 68:125, 1978

9. McCrory DC, Matcher DB, Bastian L: Evaluation of Cervical Cytology: Evidence Report Number 5, Summary Agency for Health Care Policy and Research. Agency for Health Care Policy and Research, Rockville, Maryland, 1999

10. Nanda K, McCrory DC, Myers E, et al: Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 132:810, 2000

11. Holowaty P, Miller AB, Rohan T, et al: Natural history of dysplasia of the uterine cervix. J Natl Cancer Inst 91:252, 1999

12. Mandel JS, Church TR, Ederer F, et al: Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst 91:434, 1999

13. Kronborg O, Fenger C, Olsen J, et al: Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet 348:1467, 1996

14. Hardcastle J, Chamberlain J, Robinson M, et al: Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet 348:1472, 1996

15. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. UK Flexible Sigmoidoscopy Screening Trial Investigators. Lancet 359:1291, 2002

16. Screening for colorectal cancer: recommendation and rationale. US Preventive Services Task Force. Ann Intern Med 137:129, 2002

17. Pignone M, Rich M, Teutsch SM, et al: Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 137:132, 2002

18. van Dam J, Cotton P, Johnson CD, et al: AGA future trends report: CT colonography. Gastroenterology 127:970, 2004

19. Cancer Facts and Figures. American Cancer Society. American Cancer Society, Atlanta, 2004

20. Gøtzsche PC, Olsen O: Is screening for breast cancer with mammography justifiable? Lancet 355:129, 2000

21. Humphrey LL, Helfand M, Chan BK, et al: Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 137:347, 2002

22. Green BB, Taplin SH: Breast cancer screening controversies. J Am Board Fam Pract 16:233, 2003

23. Woloshin S, Schwartz LM, Byram SJ, et al: Women's understanding of the mammography screening debate. Arch Intern Med 160:1434, 2000

24. Baines CJ: Mammography screening: are women really giving informed consent? J Natl Cancer Inst 95:1508, 2003

25. Berry D: Commentary: screening mammography: a decision analysis. Int J Epidemiol 33:68, 2004

26. Schwartz LM, Woloshin S, Sox HC, et al: US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. BMJ 320:1635, 2000

27. Thomas DB, Gao DL, Ray RM, et al: Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst 94:1445, 2002

28. Semiglazov VF, Manikhas AG, Moiseenko VM, et al: [Results of a prospective randomized investigation [Russia (St. Petersburg)/WHO] to evaluate the significance of self-examination for the early detection of breast cancer]. Voprosy Onkologii 49:434, 2003

29. Gohagan JK, Prorok PC, Kramer BS, et al: Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial of the National Cancer Institute. J Urol 151:1283, 1994

30. Schroder FH, Kranse R, Rietbergen J, et al: The European randomized study of screening for prostate cancer (ERSPC): an update. Eur Urol 35:539, 1999

31. Coley CM, Barry MJ, Fleming C, et al: Early detection of prostate cancer: part I: prior probability and effectiveness of tests. Ann Intern Med 126:394, 1997

32. Graversen PH, Corle DK, Nielsen KT, et al: Radical prostatectomy versus expectant primary treatment in stages I and II prostatic cancer: a fifteen-year follow-up. Urology 36:493, 1990

33. Wilt TJ, Brawer MK: The prostate cancer intervention versus observation trial (PIVOT). Oncology 8:1133, 1997

34. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC, 1994

35. Whitlock EP, Polen MR, Green CA, et al: Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 140:557, 2004

36. Scholes D, Stergachis A, Heidrich FE, et al: Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. N Engl J Med 334:1362, 1996

VI OCCUPATIONAL MEDICINE

Linda Rosenstock, m.d., m.p.h., f.a.c.p.
Mark R. Cullen, m.d.

Awareness of the impact of the work environment on health has increased dramatically in the past few decades. Common clinical problems, such as carpal tunnel syndrome and respiratory irritation and allergy, are increasingly being related to physical, chemical, and biologic hazards at work.[1] In this chapter, we cover some of the most common occupational disorders diagnosed in industrialized countries, and we present examples of known or suspected causes [*see Table 1*]. More extensive descriptions of specific disorders are presented in other chapters of *ACP Medicine* and in textbooks of occupational medicine.[2-4] An increasing amount of information about occupational medicine is available on the Internet from the National Institute for Occupational Safety and Health (http://www.cdc.gov) and the Occupational Safety and Health Administration (http://www.osha.gov).

Data on the frequency of occurrence of most occupational disorders are limited; however, data demonstrating the extent of the problem are available. Recent estimates are that each year, approximately 55,000 deaths result from occupational illness, and 3.8 million disabling occupation-related injuries occur.[5] Costs of occupational deaths and related injuries have been estimated to be $125 billion to $155 billion a year.[5,6] Occupational illness is common and has substantial clinical ramifications.

Basic Principles of Occupational Disease

It is important to debunk the widespread and erroneous perception that most occupational disorders are pathologically unique. Although some disorders, such as silicosis, do have distinguishing pathologic characteristics, the majority do not. Most occupational diseases—such as lung cancer induced by ionizing radiation, bladder cancer caused by fumes from coke ovens, asthma triggered by the inhalation of platinum salts, and fatty liver resulting from the absorption of the solvent dimethylformamide through the skin—are pathologically indistinguishable from disorders with more familiar causes. However, it is virtually always possible to differentiate occupational diseases from their nonoccupational counterparts. Laboratory testing and data gathering provide the best clues for the diagnosis of occupational disease, but to recognize these disorders, it is critical to ask appropriate questions when taking the medical history [*see* Clinical Evaluation, *below*].

Workplace toxins and hazards, when adequately studied, have predictable and discrete pathologic consequences. Although other diseases share common final pathways, the initial mechanisms of injury are generally highly specific for each agent. Aside from the possibility of idiosyncratic responses, as occur with pharmacologic agents, the actual potential effects of most toxins are few. For example, beryllium may cause an acute inflammatory pneumonia (acute beryllium disease) within hours after intense exposure, or it may cause a delayed hypersensitivity response with granulomatous lung disease (chronic beryllium disease [CBD]) in persons with recurrent or long-term exposures; no other form of nonmalignant lung disease is known to be caused by this metal or its salts.

Both the likelihood that workplace hazards will produce effects and the severity of those effects are determined by the amount of toxin to which the patient is exposed (hereafter referred to as dose). The nature of the relation between dose and response depends on the mechanism of action of the agent. For direct-acting toxins, which cause effects by directly disrupting cellular function or cell death at the target-organ level, there is usually a dose beneath which no biologic effects are observed—a so-called threshold level. Above this level, there is typically a sigma-shaped dose-response correlation as dose rises, until a lethal dose is reached. Similarly, an increasing percentage of the exposed population is affected as dose rises; eventually, everyone is affected. This is characteristic of heavy metals, organic solvents, and pesticides. For agents that cause allergic-type or idiosyncratic responses, such as latex and epoxy resins, which affect only susceptible people, dose contributes to the likelihood of sensitization, though not necessarily to the severity of the reaction. Further, once a worker has become sensitized, a very low dose may be sufficient to induce a full-blown clinical response. For mutagens and carcinogens, current knowledge presumes a linear dose-response model, with each increment in cumulative dose resulting in a proportional increase in the risk of cancer. The severity of the resultant cancer bears no predictable relation to the induction dose, though the time from exposure to onset generally is shorter when doses are higher.

The temporal relation between exposure and effect is highly predictable for each agent and each effect. For many direct toxins, effects occur within minutes or hours after exposure to an appropriate dose, such as the syndrome of cholinergic storm after organophosphate pesticide poisoning. Similarly, immunologically mediated responses, such as asthma and dermatitis, will occur within minutes or hours after exposure. Conversely, other effects are predictably delayed. Asbestos and silica rarely cause pneumoconiosis in less than 10 years after first exposure, except after very high exposure levels. Solid tumors, such as lung cancer associated with these same dusts, emerge, on average, 20 to 30 years after first exposure. Other effects occur in an intermediate time frame: some organophosphates cause a paralysis whose onset is delayed by weeks to months after an intense overexposure. The presentations of acute lead, mercury, or arsenic poisoning are insidious, coming after the poison accumulates to a dangerous level, usually after weeks or months of exposure.

When the clinician is approaching patients with new medical problems, consideration should be given to occupational causes. If the problem is acute, such as the relatively sudden onset of a rash or of liver function abnormalities or hemolysis, the search for a possible occupational cause should focus on recent events: Has there been a new or increased exposure to an agent that can cause such toxicity in the hours, days, or, at most, weeks before onset? On the other hand, for chronic disorders, such as pulmonary fibrosis and cancer, the search for causes should begin with a work history that goes back years.

With regard to work histories, it is important to note that host factors may modify temporal and dose-response correlations; all workers do not react alike to comparable exposures.

Table 1 Common Occupational Disorders

	Disorders	*Examples of Causal Factors*
Respiratory tract	Pneumoconiosis	Coal, silica, asbestos
	Asthma	Latex, polyurethane
	Allergic alveolitis	Vegetable matter, machining fluids
	Metal fume fever	Metal fumes
Skin	Contact dermatitis	Oils, rubber, metals
	Acne	Herbicides, oils, friction
	Urticaria	Latex
Urinary tract	Glomerular disease	Organic solvents, mercury
	Tubulointerstitial disease	Cadmium, lead
Liver	Acute or subacute necrosis	Organic solvents, TNT, 2-nitropropane
	Cholestatic hepatitis	Methylene dianiline
	Acute and chronic hepatitis	Viruses (hepatitis B, C)
	Steatosis	Organic solvents
	Hepatoportal sclerosis	Vinyl chloride, arsenical compounds
	Hepatocellular injury	Lead, arsenic, phosphorus, dioxin
Musculoskeletal	Carpal tunnel syndrome	Repetitive trauma
	Raynaud phenomenon	Repetitive vibrations, vinyl chloride
	Scleroderma	Coal mining
Nervous system	Parkinsonism	Manganese
	Peripheral neuropathy	Solvents, lead, acrylamide, arsenic
	Acute encephalopathy	Organic solvents, asphyxiants
	Acute or subacute cholinergic crisis	Organophosphate and carbamate pesticides
	Subacute encephalopathy	Mercury, lead, arsenic, manganese, carbon disulfide
	Subacute peripheral neuropathy	Organophosphates
	Chronic basal gangliar disorder	Manganese, carbon monoxide (postasphyxiation)
	Chronic encephalopathy	Recurrent organic solvent exposures
Hematologic conditions	Hemolysis	Lead, organic nitrites
	Accelerated red cell destruction	
	Acute hemolysis	Nitro and amine compounds
	Subacute hemolysis	Lead
	Disorders of oxygen transport	
	Methemoglobinemia	Nitro and amine compounds
	Carboxyhemoglobinemia	Carbon monoxide
	Disorders of red cell production	
	Hyperplastic anemia	Lead
	Aplastic anemia, hypoplastic anemia	Ethylene glycol ethers, benzene, arsenic, ionizing radiation
	Myelodysplasia	Benzene, ionizing radiation
	Polycythemia	Cobalt
Infectious	Hepatitis B, C	Health care
	Influenza A (H5N1)	Poultry workers
	SARS	Health care
	Anthrax	Animal handling
Endocrine and reproductive	Hypogonadism	Lead
	Azoospermia, oligospermia	DBCP, ionizing radiation
	Teratogenesis	Organic mercury, PCBs

DBCP—1,2-dibromo-3-chloropropane (pesticide) PCBs—polychlorinated biphenyls SARS—severe acute respiratory syndrome

In every workplace, some people appear to be immune to the effects of even the most toxic agents, and others seem to react to low doses, often lower than the threshold deemed toxic by regulatory authorities. These differences may be caused by genetic, dietary, or constitutional factors or by the preexistence of other illnesses.

In addition, many workplace hazards and toxins interact with one another and with nonoccupational factors to cause disease. Dose-response correlations for industrial hazards may be markedly shifted in the presence of other hazards, habits, or medications. An important example is the likelihood of disease resulting from thermal stress (i.e., heat or cold) in the presence of hemodynamically active agents, such as calcium channel blockers, autonomic agents, and diuretics.[7] Likewise, the effects of vibration trauma on wrists and digits may be amplified by nicotine.[8] The effects of one hazard may be significantly altered in the presence of another; for example, the combined effect of noise and solvents on hearing loss[9] and of asbestos and smoking on lung cancer[10] are greater than the effect of exposure to each hazard alone.

Defining the Pathophysiologic Basis of the Patient's Complaints

When searching for the pathophysiologic basis of a patient's complaints, it is important to ascertain the following: Is the process an acute or relapsing process, with precipitous changes in physiologic status, reflecting a recent or ongoing exposure? Or is it a chronic process, more likely the result of noxious exposure in the distant past? Dysfunction of what organ or organs best explains symptoms? Is there evidence of physiologic disruption, or is the disorder predominantly one of subjective difficulties?

Taking the Occupational History

Every patient should be questioned regarding the essentials of occupation, including current and past workplaces, job type, and materials used. Open-ended questions are always appropriate (e.g., "Are there dangerous materials or hazards in your workplace?" and "Do you believe that your work is causing you any health problems?").[11] The exploration of work as the basis for a complaint or medical problem entails an incisive approach and depends on the nature of the clinical problem being investigated. Evidence suggests that physicians need to become more adept at assessing a patient's occupational history.[12]

Approach to the Patient with an Acute Disorder

The emphasis should be on new exposures, increased exposures, and accidental exposures. Has the patient recently begun a new job or task involving hazards? Were new materials recently introduced at work? Has there been a change in working conditions, such as a failure of the ventilation system? Has there been a leak, spill, or accident? If the answer to all of these questions is no, the likelihood is low that the acute illness is related to work processes or chemicals.

Other than acute effects that are immunologically mediated, most effects are not idiosyncratic and will follow a sigma-shaped dose-response correlation like that discussed for direct-acting toxins (see above). In such circumstances, it would be expected that a high proportion of exposed persons would be affected, although individual thresholds and dose responses may differ. Questions probing effects in other exposed persons are extremely helpful, as in the investigation of food poisoning or respiratory infections. Although a negative answer does not exclude a work-related effect, the suggestion of an outbreak or a cluster makes the probability of an association high and increases the urgency of a prompt, correct diagnosis.

Approach to the Patient with Recurrent Manifestations

A patient may have repeated or recurring manifestations, such as intermittent cough, rash, or nausea. Although the cause may be difficult to establish in some situations, especially when symptoms have been very persistent or chronic, the time course, particularly at the onset of recurring manifestations, is often extremely revealing. For example, a new asthma patient whose symptoms occur on vacations and weekends is unlikely to have an occupationally related disorder.

Approach to the Patient with Chronic Disease

When patients present with evidence of irreversible organ damage or malignancy, the approach is altogether different. Although the longer latency between initial exposure and disease onset is useful in determining whether occupational exposures have played an important role, questions directed at temporal associations between symptoms and exposures are not helpful. Rather, the first step is to establish a clear pathophysiologic picture of the disease process itself. Sometimes, knowledge of past exposures may assist in directing this evaluation. For example, a worker who has been exposed to asbestos and who presents with a malignant pleural effusion should be carefully evaluated for mesothelioma, which is otherwise an uncommon disorder.

Once the disease process is characterized, a role for occupational factors can be more seriously considered by obtaining a more detailed history of exposures. Because only a handful of agents are suspected of causing or have been proved to cause any single chronic disease, the goal of this history is to determine whether exposure to any of those agents has occurred and whether the exposure occurred at a time and dose that suggest a causal connection to the disease.

Approach to Subacute and Insidious Disease

The greatest diagnostic challenge in clinical occupational medicine is the clinical disorder of gradual onset over days to weeks for which none of the above approaches are effective. Examples include peripheral neuropathies, anemia, and a change in bowel habit in the absence of evidence of malignant or irreversible organ system damage. Often, in such cases, the search for the underlying pathophysiologic process and the search for its cause seem intricately related and must proceed simultaneously. Lessons from these paradigms may be helpful. If indeed the subacute process is toxic, it most likely reflects the effects of a recent exposure, typically of an agent that is accumulating slowly. Heavy metals, pesticides, and various toxic organic chemicals often accumulate in this fashion; under typical conditions of exposure, it may take weeks or months for these agents to accrue to levels of pathogenic significance. Although it is unnecessary to identify an accidental leak or spill to make a diagnosis in such cases, it is essential to note any enhanced opportunity for exposure or any novel exposure that may have occurred relatively recently. The distant exposure history is not likely to be helpful, because the subacute disorders almost always present at the point of maximal accumulation; once the worker is removed from the site of the exposure, latency or delay in onset is unusual.

CONFIRMING AND QUANTIFYING EXPOSURE

There are two basic approaches to obtaining additional exposure information. The first involves the collection of independent information about present or former work (depending on which is relevant). After the physician obtains consent from the patient (to ensure that the patient is protected from unwanted consequences), information about exposures is requested from the employer, a trade union, or a regulatory agency. Such information is usually reported through the use of a material safety data sheet (MSDS). The MSDS provides generic chemical names, compositions, and basic toxicity information of all materials used. In addition, employers may be able to provide evidence of objective sampling that may have been done to test air levels of hazardous substances. Job descriptions, results of medical tests performed at work, information about other workers with health problems, and the use of protective equipment or other methods to limit exposure may all be of value in assessing workplace exposures.

The second potential source of dose information is biologic testing. For a few hazards, testing of urine, blood, or hair may enable the physician to determine the body burden of the agent;

Table 2 Common Occupational Hazards for Which There Are Widely Available Biologic Tests of Exposure

Hazard	*Comments*
Metals	
Arsenic	Hair sampling can detect historic exposures
Cadmium	Detectable in urine for many years if there is renal injury
Fluorides	Transient in urine
Lead	Half-life 40 days in blood
Mercury	Detectable in urine for days to weeks
Asphyxiants	
Carbon monoxide	Half-life 4 hr in blood
Pesticides	
Organophosphates	Detectable indirectly, by measurement of cholinesterase, which may be depressed for days to months
Organochlorines (e.g., DDT, chlordane, dieldrin)	Persists in blood
Organic solvents	
Benzene and toluene	Metabolites transiently in urine
Antigens	IgE antibodies measurable by RAST
Miscellaneous	
PCBs	Persists in blood

PCBs—polychlorinated biphenyls RAST—radioallergosorbent test

the results of such testing correlate with current or recent levels and, less commonly, with remote exposures [*see Table 2*]. Most of these tests cannot detect chemicals that have been cleared from the body or deposited in bodily organs; this substantially limits their usefulness for diagnostic decision making. Of course, there are no simple tests for chemicals that cause topical injury to skin or respiratory mucosa but are not absorbed. For agents that act by immune sensitization, radioallergosorbent testing or skin-patch testing may be useful both for documenting exposure and for subsequent elicitation of an immune response.

Most important of all is to remember that a test for exposure can be interpreted only in the context of the history and the clinical problem. It should not be directly interpreted as a test for disease, regardless of how the laboratory reports the data. For example, a whole blood lead level of 25 mg/dl is clear evidence of excess lead exposure. If the history indicated that the patient had recently been exposed for the first time, this level would suggest a modest, generally subtoxic dose of lead. If, however, the patient had worked around lead for many years and quit a year before the test was performed, this same value would suggest a very high previous exposure and might well be associated with health effects caused by high long-term exposure. Similarly, a large proportion of bakers working around flour dust may have IgE antibodies to wheat, rye, or other grain antigens, even though the vast majority of those bakers are symptom free and will most likely remain so. Given all these limitations, biologic testing plays only a limited role in occupational medicine and can never be a substitute for the occupational history.

DIAGNOSTIC DECISION MAKING

The determination that a patient's symptoms are work related often entails extensive ramifications for the patient's employer, as well as potentially serious public health and medicolegal implications. These may present a significant challenge to the clinician, because for many occupational disorders, there is no gold standard for diagnosis.

The decision-making process should address the following questions:

1. Is the clinical illness—including the history, physical examination, and laboratory findings—consistent with other case descriptions?
2. Is the timing between exposure and clinical onset compatible with the known biologic facts about the hazard?
3 Is the exposure dose within the range of doses believed to cause such effects?
4. Are there special attributes of the particular patient that make it more or less likely that he or she would be so affected?
5. Are there alternative ways of constructing the case that better fit the available facts?
6. Where there remains significant uncertainty about the cause, how important is it to be certain?

Regarding the certainty of identifying the cause, the general legal standard for workers' compensation purposes is "more likely than not," which is a relatively low hurdle of certainty (i.e., at least 50% certain). However, there may be other situations that demand a higher level of confidence, irrespective of the standard for obtaining compensation benefits. In general, problems involving current working conditions demand a far greater level of certainty than historical ones. For example, a diagnosis of occupational asthma in a spray painter would likely dictate removing the patient completely from exposure to the offending paint or constituent; correct identification of that agent might be crucial to saving his or her career. Similarly, if a surgeon presented with recurrent anaphylactic reactions, it would be very important to determine whether the reactions were to latex, an anesthetic agent, or some extrinsic factor.

In situations where a high level of certainty is needed, it is often worth the effort to refine the diagnostic impression by serial observations, usually while the patient remains exposed, or by utilizing diagnostic challenges of removal followed by reexposure. Using serial functional measurements, such as peak expiratory flow records or serial blood tests, a more certain judgment can be made. This may also be an appropriate circumstance for referral to occupational physicians who specialize in evaluating challenging cases.

Major Occupational Disorders in Developed Countries

The spectrum of occupational disorders of clinical importance is rapidly shifting as a result of several factors: these include changes in the economy, which have brought about a decline in traditional manufacturing and a rise in service-sector activities; better control of many hazards, such as mineral dusts (e.g., asbestos, silica, coal), heavy metals (e.g., lead, arsenic, mercury), and the most toxic solvents (e.g., benzene); rapid introduction of many new technologies whose health risks remain inadequately characterized; and changing demographics in the workplace, in which the proportion of women, minority, and older workers is increasing. In the sections that follow, the disorders that are most important in clinical practice in developed countries are briefly discussed by organ system.

Only a small fraction of known chemical agents and a handful of physical and biologic hazards appear capable of inducing neoplastic change in mammalian tissues. In general, the risk of cancer being induced increases in direct proportion to total dose of toxin to which the person is exposed. Typically, the target organ is relatively specific and is determined by metabolism and transport of the agent. However, a few agents, including ionizing radiation and asbestos, appear to have potential to cause malignancy at more than one human site. There is invariably a long lag time between initial exposure and onset of clinical disease. Only a small number of hazards found in the workplace have been clearly established as carrying substantive cancer risk for workers. An additional group of hazards are suspected, but additional studies are needed. The list of potential carcinogens is expanding; for example, evidence suggests that exposure to cadmium may play a role in the development of prostate cancer.[13] Studies provide some indication that workers in print shops, service-station employees, farm-product vendors, horticulturists, farmers, and aircraft mechanics are at increased risk for renal cell carcinoma[14,15] [*see Table 3*].

Table 3 Established Occupational Carcinogens

Cancer Site	*Hazard*	*Setting*
Lung	Asbestos	Insulation, textiles
	Ionizing radiation	Uranium mining
	Arsenic	Refining
	Polyaromatic hydrocarbons	Coke ovens
	Nickel	Nickel refining
	Chromium	Tanning, pigments
	Alkylating agents	Chemical industry
	Silica	Mining, stonecutting
	Ceramic fibers	Insulation
	Formaldehyde	Chemicals, plastics
	Beryllium	Nuclear weapons, aerospace industry
	Cadmium	Batteries
	Acrylonitrile	Plastics
	1,3-Butadiene	Rubber, plastics
Pleura and peritoneum	Asbestos	Construction materials
Upper respiratory tract	Wood dust	Carpentry
	Nickel	Refining
	Chromium	Plating
	Asbestos	Friction products
	Formaldehyde	Chemicals, plastics
Urinary bladder	Benzidine and related amines	Dyes, chemicals
	Polyaromatic hydrocarbons	Aluminum reduction
Liver	Vinyl chloride monomer	Plastics
	Arsenic	Pesticides
Upper GI tract	Asbestos	Shipbuilding
	Coal dust	Mining
	Acrylonitrile	Plastics
Hematologic system	Benzene	Chemicals, rubber
	Ionizing radiation	Defense industry
	Ethylene oxide	Chemicals, sterilizers
Soft tissue	Dioxin	Chemical industry
Brain	Vinyl chloride	Chemical industry
	Formaldehyde	Chemical industry

The respiratory tract is a frequent target of toxic effects. Complaints referable to the lungs or upper respiratory tract often require a careful evaluation for occupational causes. The presence of other possible causal factors, such as common allergy and smoking, does not exclude the possibility of an occupational cause and may, in fact, increase the likelihood of one.

Acute Disorders and Recurrent Disorders

The most prevalent acute effects—inflammatory reactions of the mucosae of the upper or lower airway system—are caused by environmental irritants.[16] An extraordinary array of agents are irritating, including simple inorganic gases (e.g., ammonia and chlorine), organic solvents, acid and alkaline mists, metal fumes (i.e., tiny particles of metal and metal oxide that occur when vaporized metals hit cool air), mineral dusts (e.g., fibrous glass and coal), and almost all the pyrolytic products of combustion. The anatomic site of irritation for dusts, mists, and fumes depends on the deposition of particles; for gases, it depends on water solubility (i.e., the more water soluble the gas, the more it will dissolve in the upper respiratory tract). Expression of symptoms, from mild burning of the eyes, nose, and throat to small airway and alveolar injury associated with the acute respiratory distress syndrome, depends on dose, duration of exposure, and the potency and composition of the irritant; there is also substantial host variability. The period from the time of exposure to the onset of symptoms is very brief for the upper respiratory structures and can be from minutes to hours for lower structures.

Most of the consequences of acute irritation are self-limited; the upper respiratory tract is particularly resilient, although patients who work in areas of poor air quality will experience frequent recurrences, punctuated by commonplace complications such as sinusitis. Such cases require steps to modify exposure. More severe insults may result in fixed scarring of airways or lung parenchyma; late inflammatory sequelae such as bronchiolitis obliterans are occasionally reported. A newly recognized and probably common outcome of significant lower airway injury is the occurrence of persistent mucosal irritation and bronchospasm, a variant of asthma induced by a single exposure or repeated exposures to irritants. Initially dubbed reactive airways dysfunction syndrome,[17] this disorder is best classified as nonimmune occupational asthma or simply asthma without latency. Unfortunately, the condition tends to be highly resistant to therapy, and patients derive only modest benefit from inhaled steroids or other bronchodilators. Typically, cough with some phlegm, chest discomfort, and occasionally even dyspnea persist despite early and intensive therapy. Reassurance and reduction of further exposures to irritants are the mainstays of treatment.

Occupational asthma, including the nonimmune- and the immune-mediated varieties, is prevalent.[18] There are now over 200 established causes of presumed immune-mediated asthma[19]; these are usually categorized as proteins and other high-molecular-weight antigens (e.g., animal danders, latex antigens, and grains) and small molecules such as the isocyanates—the ubiquitous chemicals used in polyurethane products. Typically, the classic antigens differentially affect those with atopy and are associated with identifiable IgE antibody responses to the sensitizer.[20] In such cases, the greatest diagnostic dilemma is distinguishing occupational sources from other causes of asthma, though the periodicity as documented by history or peak expiratory flow records (PEFR) aid in identifying a relation to work. Latex has become a particularly important cause, especially when ren-

dered airborne in association with the use of powdered gloves.[21,22] More troublesome are the low-molecular-weight agents such as toluene-2,4-diisocyanate (TDI) and other isocyanates, for which atopy is not a risk factor.[19,23] Onset is often insidious, with cough and chest discomfort relatively more common than in asthma of other causes. Far more often than with the IgE-mediated agents, symptoms may be delayed some hours after exposure, so patterns may include nocturnal complaints. Once the physiologic hallmarks of asthma are established, the history and PEFR are the keys to specific diagnosis. Studies have shown that detailed histories can be inconclusive; in some cases, objective measurements can establish the diagnosis of occupational asthma.[24] Specific inhalation tests may be valuable, but they should be performed only under medical supervision.

Current evidence suggests that correct diagnosis of occupational asthma makes a difference. People who are removed early from further contact have a better likelihood of reducing their dependence on medication; many will become nonasthmatic over time.[19,20] Most who remain exposed will develop persistent nonspecific bronchial hyperreactivity, as well as possible fixed obstructive changes. These patients will typically fail to recover after they are removed from contact with the agent, and their conditions may even worsen; this is the basis for an aggressive posture toward early evaluation and management.

Acute infectious diseases occur in an extraordinarily wide variety of workplaces, from health care to industrial and agricultural settings. Anthrax and other agents of bioterrorism, as well as emerging infectious diseases such as severe acute respiratory syndrome and influenza A (H5N1) are of particular concern to workers.[25,26]

Allergic alveolitis, with its more benign variants, such as humidifier fever, continues to occur sporadically in a wide range of settings. This disorder was traditionally associated with agricultural exposures to molds and bacilli. Cases are now reported to occur in manufacturing and other industrial settings because of the appearance of a few chemicals that appear capable of inducing the immune response (e.g., plastic resin constituents) and because of the contamination of many industrial processes with microorganisms.[27] The office environment continues to be an occasional source of this condition as well, though the reservoir of causal microbes may be obscure; such organisms may potentially reside in heating and air-conditioning systems remote from the patient's work area.[28]

Chronic Conditions

The pneumoconioses continue to occur, in part because of their very long latency from first exposure and because pockets of very poor industrial conditions continue to exist even in developed countries. Construction activities have been particularly problematic. In general, asbestosis, silicosis, and coal workers' pneumoconiosis are diseases that occur after extensive work exposures. The diagnosis can usually be made on the basis of clinical findings and the history of exposure, once the patient's lifetime job history is obtained.

The granulomatous diseases, including CBD and so-called hard metal disease, are less common but important and increasingly recognized disorders of sensitization. CBD is clinically almost identical to idiopathic sarcoidosis except that all cases involve the lung and that the prognosis—even after the patient is removed from exposure to beryllium metal, compound, or fumes—is generally unfavorable. All patients with sarcoidosis should be asked if they work with metals, and the least suspicion should prompt specific testing; there is a highly sensitive test that can distinguish sarcoidosis from CBD on blood or bronchoalveolar lavage (BAL) fluid.[29] Hard metal disease is a giant cell alveolitis induced through an idiosyncratic reaction in workers exposed to the metal cobalt.[30] Most often, it occurs in workers making or using tungsten carbide, the very hard metal used for machine tools. Onset may be insidious and may include asthmatic symptoms, because cobalt is asthmogenic as well. Recognition of the parenchymal process by BAL or biopsy is crucial because hard metal disease is progressive, often refractory to treatment with steroids, and often lethal; there is anecdotal evidence favoring the use of cytotoxic drugs. Once hard metal disease is diagnosed, the patient should be promptly removed from any further exposure.

In 1998, a novel form of interstitial fibrosis related to an industrial exposure was reported: flock worker's lung, named after the nylon flocking used for making feltlike textiles.[31] Cases of flock worker's lung are distinctive, with pathologic evidence of both parenchymal fibrosis and lymphocytic bronchiolitis. The reporting of flock worker's lung underscores a key principle of occupational medicine: that new occupational diseases and other clinical consequences of work continue to be uncovered.[32]

DERMATOLOGIC DISORDERS

Despite increased recognition of the need to reduce contact between the skin and the chemical and physical environment, dermal conditions remain responsible for significant morbidity in the workplace. Most disorders are caused by direct exposure of the skin to workplace irritants, sensitizers, pigments, carcinogens, and materials that interfere with normal dermal function by disrupting sebaceous and follicular secretions (e.g., oils that cause acne) or solvents that erode protective lipids. Trauma, foreign bodies, ionizing and nonionizing radiation, and extremes of temperature may modify or disrupt skin growth, vascular integrity, or both. On occasion, systemic exposure may have a dermal consequence, as in urticarial responses to inhaled antigens, pigmentary alterations from the deposition of metals (e.g., silver), and the much-described though rarely seen chloracne, a variant of acne induced by dioxins and related chemicals. Workers who are at increased risk for allergic contact dermatitis include tanners, cast-concrete product workers, leather-goods workers, footwear workers, machine and metal product assemblers, electrical and telecommunications equipment assemblers, print-shop workers, and machine and engine mechanics.[33] Several excellent texts of occupational skin diseases are available.[34-36]

Overwhelmingly, the major skin problem in the workplace remains dermatitis, either irritant induced or caused by allergy. Many agents may be responsible, including organic and inorganic chemicals, plastics and rubber, oils and lubricants, metals and construction materials, paints, and coatings.[37] Both allergic dermatitis and irritant-induced dermatitis are more likely to affect persons with atopic conditions, dry skin, or other dermal risk factors. Distinguishing between the two is less important than recognizing occupational precipitants in the first place; both are difficult to differentiate from other commonplace skin disorders, such as eczema. The key to correct diagnosis is the history of skin contact and the temporal relation between contact and manifestations. Unfortunately, there is seldom a perfect or obvious correlation between the two, and some sleuthing is necessary, especially to discern the extent to which chemical contact

may spread to places like the groin or areas where hand contact occurs. Airborne exposure may cause lesions in apparently untouched areas, such as the face; such occurrences are signs of likely hypersensitivity. Vexingly, symptoms do not always abate dramatically over weekends or during short periods in which exposure is avoided; removing the patient from the toxin for a week or two may be necessary to observe response. This, combined with observation of the patient during reexposure, is often the most valuable diagnostic test. Patch testing, performed by an experienced clinician aware of the exposures of concern, may be useful in difficult cases, though the clinician should keep in mind that irritants may yield false negative results and that even many healthy atopic persons will experience reactions to common contactants, such as nickel.[38] Often, complete isolation from offending agents is economically infeasible, and materials that previously were well tolerated become sources of irritation and exacerbation. Combinations of work modification, aggressive treatment of flares and complications, and careful attention to routine skin care are necessary to control disease.

DISORDERS OF THE URINARY TRACT

Although innumerable toxins are known to cause acute injury to the kidney, exposures to chemical and physical agents at concentrations found in the workplace rarely cause such effects (exceptions include cases involving overwhelming accidental overexposure or ingestion). Of far greater concern are recurring exposures to agents at more typical workplace exposure levels that have subclinical effects but can lead to late nephropathy. Although there remains a vast burden of unexplained nephropathology in the population and despite epidemiologic data suggesting an occupational cause,[39,40] chronic renal injury resulting from workplace exposures remains poorly characterized.

The best-established effects on the urinary tract are those caused by exposure to heavy metals, especially lead, mercury, and cadmium; each of these metals is associated with a unique pattern of effects. Workers whose jobs entail exposure to lead include traffic police, hazardous-waste incineration workers, industrial workers, and furniture strippers; workers at risk for exposure to mercury include gold-mine workers, workers at chloralkali plants, workers exposed to hazardous waste, and construction workers; workers at risk for exposure to cadmium include those involved in the manufacture of batteries. Long-standing heavy-lead exposure results in a pattern of injury difficult to distinguish pathologically and clinically from the effects of hypertension; signs and symptoms include nephrosclerosis and evidence of both glomerular and tubular defects. The ability to clear urate is impaired early in the course and may be a clue; saturnine gout may occur a decade later. There is debate about the possibility of low-level or brief exposures to lead predisposing to hypertension or enhancing the degree of renal injury associated with essential hypertension or gout.[41,42] Proponents of this view stress the importance of assessment of lead exposure in patients with mild chronic renal insufficiency.[43]

Long-term occupational exposure to inorganic mercury—principally through exposure to mercury vapor—may result in renal alterations involving the tubules and glomeruli. The monitoring of urinary mercury is useful for controlling such risk.[44]

Cadmium exposure in jewelry making, battery production, and other metal-processing operations leads to bioaccumulation of cadmium in the kidney, which results in proximal tubular injury with excessive excretion of β_2-microglobulin and other tubular proteins. Later, a pattern of renal tubular acidosis may occur, which subsequently may lead to the development of renal insufficiency. Because the tubular dysfunction is only partially reversible,[45] it is important to carefully monitor cadmium exposure, which is best done with regular blood and urine cadmium testing.[46] Renal damage can occur at relatively low levels of cadmium exposure.[47]

Organic solvents have been implicated in renal tubular and renal parenchymal injury[48]; despite uncertainty of their role in renal toxicity, growing evidence suggests the need for evaluation of these substances in all new cases of unexplained nephropathy.

LIVER DISEASE

The liver is highly sensitive to effects of numerous organic and inorganic substances used in the workplace [*see Table 1*]. Despite the impressive potential for harm, often at exposure levels not uncommon in the workplace, occupational liver diseases are rarely recognized except during outbreaks.[49] This is almost certainly because the clinical presentation is nonspecific, most often consisting of unsuspected elevations of hepatocellular enzymes occasionally associated with mild gastrointestinal symptoms. The single exception to this is the now extremely rare vascular disorder resembling veno-occlusive disease that is caused by vinyl chloride.

The more common hepatic effects of occupational hazards—steatosis and nonspecific hepatocellular injury—have numerous causes and are prevalent in the general population; a given case may be readily attributed to infection, alcohol use, drug toxicity, biliary tract disease, diabetes, obesity, or weight change. When persistent elevations of hepatic enzymes prompt more extensive workup with radiographic studies and biopsy, results rarely provide specific evidence of an occupational cause. Only high suspicion of a workplace culprit, combined with evidence of exposure to a suspect agent, serves to distinguish etiology.

CENTRAL AND PERIPHERAL NERVOUS SYSTEMS

Most pesticides,[50] organic solvents,[51] and many metals[52] are neurotoxic at doses that may be seen in the workplace [*see Table 1*]. A handful of other chemicals used in plastics, lubricating fluids, and chemical operations are also neurotoxic; most cases occur after accidental or unusual exposures. In addition, persons exposed to asphyxiants, such as carbon monoxide and cyanide, may present with acute or recurring central nervous system symptoms. Both acute and late effects may occur—the former typically occurring immediately after an intense exposure, the latter often after prolonged periods of exposure. Importantly, the late or chronic effects usually result from prolonged periods of bioaccumulation or recurrent mild or subclinical acute exposures or as sequelae of acute intoxication. A direct consequence of this toxicologic fact is that neurotoxicity almost invariably presents during the time of occupational exposure to the offending agent and not long afterward, as may occur with carcinogenic substances or dusts causing pneumoconiosis.

Because of the extraordinarily diverse range of clinical symptoms that may herald CNS toxicity, including subtle changes in cognitive and affective function, the evaluation of suspected cases follows the general principles for all occupational disease, with increased attention given to recent exposures. The acute disorders usually occur as mild alterations of CNS function,[53] often with associated GI or other systemic effects; they are often recurrent, cycling with periods of work exposure, as might be seen in a painter (through exposure to solvents) or a pest-control worker. The key to recognition is the temporal pattern, with re-

mission of symptoms occurring over a course of time consistent with the metabolism of the toxin. There may also be evidence of symptoms associated with withdrawal, similar to the effects associated with ethyl alcohol. For the subacute and chronic effects, the key to diagnosis is identification of evidence of substantial exposure occurring over a course of time consistent with the evolving neurotoxic picture. None of the neurologic disorders appear to involve allergy or idiosyncrasy; thus, the doses of exposure involved must be substantive.

In many cases, the exposure to the agent can be biologically confirmed with measurement of the levels of metal in the urine or blood, measurement of cholinesterase levels, or identification of a metabolite of an organic chemical in urine. There may also be some clinical or pathophysiologic clues. For example, the constellation of cerebellar ataxia, personality change, and salivary gland hypersecretion should prompt consideration of inorganic mercurialism, possibly with associated renal effects. An asymmetrical motor neuropathy should always raise the specter of lead poisoning. Insidious symmetrical distal sensory neuropathies, on the other hand, are far more common with solvents and acrylamide; electrophysiologic or pathologic evaluation reveals almost pure axonal degeneration, with minimal secondary demyelination—an important differential feature. Highly localized neuropathies, either unilateral or bilateral, should raise the possibility of a compressive etiology, not uncommon with repetitive work activities [*see* Musculoskeletal Disorders, *below*].[54]

Although diagnosis may be straightforward once the possibility of a workplace agent is considered, management remains challenging. Treatment of acute disorders involves ending the exposure and providing support where clinically necessary. Several hazards, such as certain cholinesterase inhibitors and cyanide, have specific antidotes that should be administered under medical supervision. The subacute and chronic conditions all require removal from further exposure. In addition, patients with heavy-metal exposure may be given chelation therapy when signs and symptoms of severe intoxication are evident; this, too, must be done under very close supervision in view of the risk of enhancing CNS effects early in treatment. Moreover, the possibility of rebound effects from reequilibration of metal into the nervous system must be anticipated when chelation is stopped. Most important, whatever strategy is chosen, physician and patient must be aware that the prognosis for full recovery from all but the most acute effects is somewhat guarded. Axons regrow very slowly, and higher integrative functions, such as affective or cognitive functions of the CNS, resolve even more slowly or not completely. Early efforts at functional rehabilitation, as may be used for trauma or stroke patients, are indicated when impairments limit work or other major life activities.

Possibly the most challenging diagnostic situation in occupational neurology is the worker who presents with CNS-related complaints that exhibit a temporal pattern consistent with a workplace origin but who does not have substantial exposure to neurotoxic agents. Such symptoms are a common part of the so-called sick-building syndrome, now referred to as nonspecific building-related illness, and are universal among persons who have acquired intolerance to low levels of chemicals (multiple chemical sensitivities).[55] It is important to recognize early that these syndromes are different from the neurotoxic disorders discussed here with regard to evaluation, prognosis, and treatment. They are discussed more fully later in this chapter [*see* Clinical Problems Associated with Low-Level Environmental Exposures, *below*].

MUSCULOSKELETAL DISORDERS

There has been a marked increase in the awareness of the role that work factors play in musculoskeletal disorders, ranging from such well-defined clinical problems as arthropathies and nerve compression syndromes to the less well characterized ailments causing pain of the trunk and extremities.[54,56] In developed countries, such disorders account for billions of dollars of costs in medical care and lost productivity. The overwhelming bulk of this epidemic relates to suspected consequences of physical stressors and trauma that occur at work. A number of systemic occupational disorders may also have expression in the muscles, bones, joints, and connective tissues; important examples of such disorders are the arthralgias and gouty consequences of lead intoxication, bony pain in association with systemic fluorosis, and the apparent increased risk of scleroderma in miners.[57]

It is clinically useful to divide potential occupational musculoskeletal disorders into those that have a well-defined anatomic structure of involvement, such as carpal tunnel syndrome, and those that lack such a clear-cut pattern, such as low back pain.[58,59] Although extensive data suggest that physical aspects of work, such as overall force, repetition, awkward posture, and vibration, contribute in a cumulative fashion to the development of both localizable and nonspecific symptoms, the approach to diagnosis and treatment is somewhat different for each. There is also evidence that factors other than physical strain, such as work stress, work fatigue, and adverse relationships in the workplace, may be important contributory factors, partially explaining high rates of musculoskeletal disorders among certain white-collar workers.[60,61]

For disorders of new onset involving the trunk or extremities or for clinically mild disorders, the initial approach should be short-term palliation with minimal workup. Rest from physically demanding tasks, use of nonsteroidal anti-inflammatory drugs or other nonnarcotic pain relievers, reassurance, and follow-up after a few days of treatment are suggested; further evaluation is indicated only if suggested by physical findings. If conservative steps fail to alleviate symptoms rapidly, additional examination and laboratory evaluation may be appropriate to rule out an anatomically discrete lesion that could be amenable to treatment. Where specific lesions are identified, such as compression of a nerve or disk or tenosynovial inflammation, longer-term efforts at elimination of strain in the affected region combined with anti-inflammatory drugs or other therapies are appropriate, followed by surgical intervention should these fail. In such cases, it is crucial to remember that the work-related stressors that caused the problem will complicate recovery unless they are modified.[62-64]

The most perplexing problem is the management of patients whose complaints cannot be specifically localized by physical examination or, when necessary, electrophysiologic or radiologic evaluation. Such complaints are no less real than those that are more readily understood and treated. Modification of work activities is often necessary but is rarely sufficient to resolve the problem. Pain may be persistent and refractory to treatment, and the value of physical therapy or pain medications is questionable. Rather, it is important for the treating physician to establish early that the symptoms are troublesome but not the result of a progressive process and that the patient may have to adapt to them despite discomfort. Expectation of cure often leads to unnecessary treatment, prolonged (and clinically unhelpful) loss of work time, and, ultimately, frustration on the part of the employer, the insurance company, the patient, and the physician.

HEMATOLOGIC DISORDERS

A host of disturbances of red cell function, survival, and production have been attributed to workplace exposures, including acute, subacute, and chronic processes [*see Table 1*]. Effects involving other cell lines have seldom been reported and will not be discussed. In clinical practice, the biggest concerns are the risk of acute hemolysis in workers exposed to nitrogen-containing oxidant chemicals in pharmaceutical, chemical, and explosives manufacturing; the effects of lead, which remains ubiquitous in the work environment; and the potential for solvent-induced marrow injury. The problem of oxidant stressors is somewhat difficult. Although workers with marked deficiency of glucose-6-phosphate dehydrogenase (G6PD) should probably avoid significant contact with such chemicals, there is not a clear relation between any of the measurable enzyme levels and risk. It is prudent to periodically screen all exposed workers for subclinical evidence of hemolysis, as well as for subclinical accumulation of methemoglobin, which is often induced by the same agents; workers who show evidence of early effects should probably be removed from harm's way, irrespective of identifiable factors.[65]

The hematologic effects of lead are widely misunderstood.[66,67] Although there is a dose-related inhibition of heme synthetase by lead that can be readily quantified by determining the accumulation of the precursor protoporphyrin (usually measured as whole blood zinc protoporphyrin), this biochemical effect of lead on blood hemoglobin or hematocrit is minimal until very high levels are reached, and there is almost no impact on red cell volume. In other words, anemia associated with hypoproliferation of red cells is very rare, and the absence of anemia should never be used to exclude a role for lead in causing toxicity to organs and systems that are far more sensitive, such as the nervous system and renal tubules. Furthermore, microcytosis can only occasionally be attributed to lead alone; when it is seen, especially in children, it most often signifies coincident iron deficiency. On the other hand, rapid accumulation of lead in acute lead poisoning, typically heralded clinically by the onset of abdominal pain, is almost always associated with evidence of rapid hemolysis; reticulocyte counts are in the range of 5% to 20%. In this setting, the notorious basophilic stipples are frequently seen as well, though they are by no means pathognomonic for lead toxicity. In general, this syndrome will occur only after lead levels have exceeded 60 mg/dl in whole blood. The hemolysis tends to abruptly stop after effective chelation therapy, which is usually indicated in this acute symptomatic form of lead poisoning.

The bone marrow effects of workplace chemicals are only slowly being unraveled, but certain conclusions seem warranted. Benzene, the aromatic constituent of petroleum products, was once widely prevalent in the work environment as a solvent and a component of gasoline. It can cause hypoplastic injury to the marrow, which may directly progress to a chronic blood dyscrasia (i.e., myelodysplasia or leukemia), or dyscrasia may occur after apparent recovery.[68] In other words, an exposed worker may show depressed cell counts, be removed from the source of toxicity, improve, and years later (possibly long after exposure ceases) develop myelodysplasia or a myeloproliferative syndrome. It is likely that some workers will develop the obviously more serious dyscrasias without direct marrow injury having been recognized while exposure was ongoing. There are no hallmark features of either the hypoplastic state (occurring during ongoing exposure) or the myelodysplastic state (occurring later) that distinguish benzene toxicity from other causes of such disorders; this differentiation depends on the history of substantial benzene exposure, because the disorders are not believed to be idiosyncratic but dose related. Although there is some evidence that a few other solvents, such as the glycol ethers that are widely used in paints and coatings,[69] may cause such injury, the vast majority of solvents, including many benzene congeners such as toluene and xylene, do not appear to have potential for marrow injury. For this reason, most products that formerly contained benzene that are used in developed countries have been modified, and benzene is not used directly except for specific purposes in the manufacture of chemicals and pharmaceuticals. Obviously, exposed persons should be carefully monitored for hematologic effects, the presence of which would be clear evidence of overexposure.

ENDOCRINE AND REPRODUCTIVE EFFECTS

Despite an exceptional upsurge in interest in the endocrine-disrupting effects of environmental contaminants, there is little evidence that occupational exposures to chemical hazards cause clinically relevant endocrinopathies in adults.[70] Lead has been shown to impair hypothalamic–pituitary axis secretions and probably testosterone regulation in men heavily exposed, but the clinical relevance of these observations is unclear. Several compounds used in the pharmaceutical industry and other industries have been shown to have estrogenic activity, with predictable clinical consequences in both men and women.

The effects of work on male and female reproduction are a more formidable concern.[71] Although data are far from complete because many chemicals have never been studied adequately, several substances at occupational levels of exposure have been proved to cause infertility and decreases in sperm counts; such substances include lead, the pesticides 1,2-dibromo-3-chloropropane (DBCP) and ethylene dibromide (EDB), ethylene glycol ethers, and carbon disulfide. Heat and ionizing radiation have also been associated with infertility and decreased sperm counts. In addition, a host of other metals, anesthetic agents, and plastic reagents have been shown to cause worrisome gonadal effects in toxicologic experiments on male animals. For this reason, infertile men should be carefully questioned about work exposures; they should be observed for signs of improvement for about 9 months (which equals four cycles of spermatogenesis) should suspicion of an occupational cause be entertained.

Female reproduction is harder to study for lack of a single body fluid to analyze and because of the absence of a simple animal model. There is evidence that several common exposures, including waste anesthetic gases, lead, glycol ethers, ethylene oxide, and antineoplastic drugs, have the potential to increase the risk of miscarriage. Lead, organic mercury, polychlorinated biphenyls (PCBs), heat, and ionizing radiation are established teratogens; organic solvents are also suspect on the basis of animal studies and new epidemiologic reports.[72] Most of the agents that cause human cancer [*see Table 3*] are considered likely fetal hazards as well. In most cases, there is risk of adverse effects at doses considered acceptable in the workplace, because regulations have not traditionally been developed on the basis of reproductive concern. To a disturbing degree, knowledge of the reproductive effects of thousands of additional chemicals is unknown. Even the effects of hard physical work during pregnancy remain unclear, though there is evidence that excessive lifting and standing late in the third trimester may induce prematurity.

With the majority of women of reproductive age now in the workforce, many are questioning the safety of work during pregnancy, and clinicians are being confronted with trade-offs between fetal risks and the worker's economic security. Although each case must be studied individually, a reasonable guideline is to rigorously protect patients from the established teratogens or ensure the levels of exposure below those established for pregnancy. For others, reasonable steps can be taken to minimize exposure, including job transfer if the patient prefers and the employer has alternative work. For the patient for whom any risk represents an unacceptable psychological impediment, transfer or removal is probably in the best interest of all parties.

CLINICAL PROBLEMS ASSOCIATED WITH LOW-LEVEL ENVIRONMENTAL EXPOSURES

One of the most common problems emerging in developed countries is the constellation of respiratory and systemic complaints that are appearing with increasing frequency in office workers and others in what are traditionally considered safe jobs.[73,74] Typical symptoms of sick-building syndrome, or nonspecific building-related illness, include upper and lower respiratory symptoms, often combined with neurologic problems, such as fatigue, headache, and cognitive deficits, as well as rashes and other nonspecific complaints.[74] Usually, the patient will relate that others in the environment are experiencing similar difficulties and that the symptoms improve when the patient is away from work and return upon reexposure. Although in a minority of cases, investigation may reveal a specific allergy (e.g., in patients with asthma, rhinitis, or allergic alveolitis) or a specific hazard (e.g., fibrous glass released during a renovation or from a ventilation duct, causing pruritus), in the majority of cases, the environment is usually best described as poorly ventilated.[74,75] At present, there is no specific treatment of this syndrome other than palliative care and reassurance that it is neither progressive nor life threatening.[76,77] Expensive testing of either the patient or the work environment is rarely necessary or beneficial.[74] Ideally, remediation of both should be undertaken as soon as more dire possibilities are excluded by history and a walk-through of the workplace by an industrial hygienist or comparable environmental professional. In the vast majority of cases, improvement of ventilation will result in symptomatic improvement for most workers.[74]

On occasion, a patient in an affected building will start to experience similar discomfort in other situations, such as driving behind a bus, being in a store, or using a perfume or detergent.[55] The net impression is that the patient has become reactive to everything that has an odor. Many also have fatigue or other asthenic symptoms between exposures. Symptoms reminiscent of those in panic disorder may also occur. Dubbed multiple chemical sensitivities (MCS), this disorder is not associated with measurable abnormalities of organ system function but may be highly disabling.[55] Although there are many physical and psychological theories regarding the origin of MCS, present knowledge is limited. Patients do not easily tolerate pharmacologic agents and usually do not respond to treatment for anxiety or depression. Avoidance is equally fruitless, with shorter and more trivial exposures causing problems in those who quit work and minimize human contact. At present, the recommended treatment is supportive care coupled with moderate life modifications to avoid the most provocative exposures while preserving everyday functioning, including work if possible. Unrealistic expectations of cure or remission are as harmful as unwarranted fears of deterioration; neither outcome appears common among patients followed for many years.

Linda Rosenstock, M.D., M.P.H., F.A.C.P., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Mark R. Cullen, M.D., is a consultant for Alcoa, Inc.

References

1. Palmer K, Coggon D: ABC of work related disorders: investigating suspected occupational illness and evaluating the workplace. BMJ 313:809, 1996

2. Encyclopedia of Occupational Health and Safety, 4th ed. Stellman JM, Ed. International Labor Office, Geneva, Switzerland, 1999

3. Maxcy Rosenau Textbook of Public Health and Preventive Medicine, 14th ed. Last WB, Ed. Appleton & Lange, Norwalk, Connecticut, 1998

4. Textbook of Clinical Occupational and Environmental Medicine, 2nd ed. Rosenstock L, Cullen M, Brodkin CA, et al, Eds. Elsevier Saunders, 2005

5. Schulte P: Characterizing the burden of occupational injury and disease. J Occup Environ Med 47:607, 2005

6. Injury Facts, 1998. National Safety Council. Itasca, Illinois, 1999

7. Epstein Y, Albukrek D, Kalmovitc B, et al: Heat intolerance induced by antidepressants. Ann NY Acad Sci 813:553, 1997

8. Tanaka S, Wild DK, Cameron LL, et al: Association of occupational and non-occupational risk factors with the prevalence of self-reported carpal tunnel syndrome in a national survey of the working population. Am J Ind Med 32:550, 1997

9. Lataye R, Campo P: Combined effects of a simultaneous exposure to noise and toluene on hearing function. Neurotoxicol Teratol 19:373, 1997

10. Erren TC, Jacobssen M, Piekarski C: Synergy between asbestos and smoking on lung cancer. Epidemiology 10:405, 1999

11. Lax MB, Grant WD, Manetti FA, et al: Recognizing occupational disease: taking an effective occupational history. Am Fam Physician 58:935, 1998

12. Politi BJ, Arena VC, Schwerha J, et al: Occupational medical history taking: how are today's physicians doing? A cross-sectional investigation of the frequency of occupational history taking by physicians in a major US teaching center. J Occup Environ Med 46:550, 2004

13. Achanzar WE, Diwan BA, Liu J, et al: Cadmium-induced malignant transformation of human prostate epithelial cells. Cancer Res 61:455, 2001

14. Parent ME, Hua Y, Siemiatycki J: Occupational risk factors for renal cell carcinoma in Montreal. Am J Ind Med 38:609, 2000

15. Zhang Y, Cantor KP, Lynch CF, et al: A population-based case-control study of occupation and renal cell carcinoma risk in Iowa. J Occup Environ Med 46:235, 2004

16. Baur X, Chen Z, Liebers V: Exposure-response relationships of occupational inhalative allergens. Clin Exp Allergy 28:537, 1998

17. Brooks SM, Bernstein IL: Reactive airways dysfunction syndrome or irritant induced asthma. Asthma in the Workplace. Bernstein IL, Chan-Yeung M, Malo JL, et al, Eds. Marcel Dekker, New York, 1993, p 533

18. Rabatin JT, Cowl CT: A guide to the diagnosis and treatment of occupational asthma. Mayo Clin Proc 76:633, 2001

19. Chan-Yeung M, Malo JL: Occupational asthma. N Engl J Med 333:107, 1995

20. Wild LG, Lopez M: Occupational asthma caused by high-molecular-weight substances. Immunol Allergy Clin North Am 23:235, 2003

21. Liss GM, Sussman GL: Latex sensitization: occupational versus general population prevalence rates. Am J Ind Med 35:196, 1999

22. Zeiss CR, Gomaa A, Murphy FM, et al: Latex hypersensitivity in Department of Veterans Affairs health care workers: glove use, symptoms, and sensitization. Ann Allergy Asthma Immunol 91:510, 2003

23. Meredith SK, Bugler J, Clark RL: Isocyanate exposure and occupational asthma: a case-referent study. Occup Environ Med 57:830, 2000

24. Malo JL, Chan-Yeung M: Occupational asthma. J Allergy Clin Immunol 108:317, 2001

25. Drazen J: SARS: looking back over the first 100 days. N Engl J Med 349:319, 2003

26. Bridges CB, Lim W, Hu-Primmer J, et al: Risk of influenza A (H5N1) infection among poultry workers. J Infect Dis 185:1005, 2002

27. Hodgson MJ, Bracker A, Yang C, et al: Hypersensitivity pneumonitis in a metalworking environment. Am J Ind Med 39:616, 2001

28. Hoffman RE, Wood RC, Kreiss K: Building-related asthma in Denver office workers. Am J Public Health 83:89, 1993

29. Middleton DC: Chronic beryllium disease: uncommon disease, less common diagnosis. Environ Health Perspect 106:765, 1998

30. Cugell DW: The hard metal diseases. Clin Chest Med 13:269, 1992

31. Kern DG, Crausman RS, Durand KT, et al: Flock worker's lung: chronic interstitial lung disease in the nylon flocking industry. Ann Intern Med 129:261, 1998

32. Kern DG, Kuhn C 3rd, Pransky GS, et al: Flock worker's lung: broadening the spectrum of clinicopathology, narrowing the spectrum of suspected etiologies. Chest 117:251, 2000

33. Kanerva L, Jolanki R, Estlander T, et al: Incidence rates of occupational allergic contact dermatitis caused by metals. Am J Contact Dermat 11:155, 2000

34. Hogan DJ: Occupational Skin Disorders. Igaku-Shoin, New York, 1994

35. Adams R: Occupational Skin Disease, 3rd ed. WB Saunders Co, Philadelphia, 1999

36. Marks JG, DeLeo VA: Contact and Occupational Dermatology, 3rd ed. Mosby, St. Louis, 2001

37. Pratt MD, Belsito DV, DeLeo VA, et al: North American Contact Dermatitis Group patch-test results, 2001–2002 study period. Dermatitis 15:176, 2004

38. Krob HA, Fleischer AB Jr, D'Agostino R Jr, et al: Prevalence and relevance of contact dermatitis allergens: a meta-analysis of 15 years of published T.R.U.E. test data. J Am Acad Dermatol 51:349, 2004

39. de Broe ME, D'Haese PC, Nuyts GD, et al: Occupational renal diseases. Curr Opin Nephrol Hypertens 5:114, 1996

40. Wedeen RP: Occupational and environmental renal disease. Semin Nephrol 17:46, 1997

41. Batuman V, Landy E, Maesaka JK, et al: Contribution of lead to hypertension with renal impairment. N Engl J Med 309:17, 1983

42. Lin JL, Lin-Tan DT, Hsu KH, et al: Environmental lead exposure and progression of chronic renal diseases in patients without diabetes. N Engl J Med 23:277, 2003

43. Erhlich R, Robins T, Jordaan E, et al: Lead absorption and renal dysfunction in a South African battery factory. Occup Environ Med 55:453, 1998

44. Roels HA, Hoet P, Lison D: Usefulness of biomarkers of exposure to inorganic mercury, lead, or cadmium in controlling occupational and environmental risks of nephrotoxicity. Ren Fail 21:251, 1999

45. Roels HA, Van Assche FJ, Overstseyns M, et al: Reversibility of microproteinuria in cadmium workers with incipient tubular dysfunction after reduction of exposure. Am J Ind Med 31:645, 1997

46. Jarup L, Persson B, Elinder CG: Blood cadmium as an indicator of dose in a long-term follow-up of workers previously exposed to cadmium. Scand J Work Environ Health 23:31, 1997

47. Jarup L, Hellstrom L, Alfven T, et al: Low level exposure to cadmium and early kidney damage: the OSCAR study. Occup Environ Med 57:668, 2000

48. Ravnskow U: Hydrocarbons may worsen renal function in glomerulonephritis. Am J Ind Med 37:599, 2000

49. Hoet P, Graf MI, Bourdi M, et al: Epidemic of liver disease caused by hydrochlorofluorocarbons used as ozone-sparing substitutes of chlorofluorocarbons. Lancet 350:556, 1997

50. Keifer MC, Mahurin RK: Chronic neurologic effects of pesticide overexposure. Occup Med 12:291, 1997

51. Rutchik JS, Whittman RI: Neurologic issues with solvents. Clin Occup Environ Med 4:621, 2004

52. Manzo L, Artigas F, Martinez E, et al: Biochemical markers of neurotoxicity: a review of mechanistic studies and applications. Hum Exp Toxicol 15(suppl):S20, 1996

53. Stallones L, Beseler C: Pesticide illness, farm practices, and neurological symptoms among farm residents in Colorado. Environ Res 90:89, 2002

54. Musculoskeletal Disorders and Workplace Factors: A Critical Review of Epidemiologic Evidence for Work-Related Musculoskeletal Disorders of the Neck, Upper Extremity, and Low Back. Publication no. 97-141, U.S. Department of Health and Human Services. National Institute for Occupational Health and Safety, Cincinnati, Ohio, 1997 http://www.cdc.gov/niosh/ergosci1.html

55. Cullen MR: Low-level environmental exposures. Textbook of Clinical Occupational and Environmental Medicine, 2nd ed. Rosenstock L, Cullen MR, Brodkin CA, et al, Eds. Elsevier Saunders 2005, p 1127

56. Tanaka S, Petersen M, Cameron L: Prevalence and risk factors of tendinitis and related disorders of the distal upper extremity among U.S. workers: comparison to carpal tunnel syndrome. Am J Ind Med 39:328, 2001

57. Katz JN, Brissot R, Liang MH: Systemic rheumatologic disorders. Textbook of Clinical Occupational and Environmental Medicine, 2nd ed. Rosenstock L, Cullen MR, Brodkin CA, Eds. WB Saunders Co, Philadelphia, 2005, p 533

58. Pascarelli EF, Hsu YP: Understanding work-related upper extremity disorders: clinical findings in 485 computer users, musicians, and others. J Occup Rehabil 11:1, 2001

59. Hoogendoorn WE, Bongers PM, de Vet HC, et al: Flexion and rotation of the trunk and lifting at work are risk factors for low back pain: results of a prospective cohort study. Spine 25:3087, 2000

60. Macfarlane GJ, Hunt IM, Silman AJ: Role of mechanical and psychosocial factors in the onset of forearm pain: prospective population based study. BMJ 321:676, 2000

61. Haufler AJ, Feuerstein M, Huang GD: Job stress, upper extremity pain and functional limitations in symptomatic computer users. Am J Ind Med 38:507, 2000

62. Jayson MI: ABC of work related disorders: back pain. BMJ 313:355, 1996

63. Hagberg M: ABC of work related disorders: neck and arm disorders. BMJ 313:419, 1996

64. Buckle PW: Work factors and upper limb disorders. BMJ 315:1360, 1997

65. Luster MI, Wierda D, Rosenthal GJ: Environmentally related disorders of the hematologic and immune systems. Med Clin North Am 74:425, 1990

66. Nelson JC, Westwood M, Allen KR, et al: The ratio of erythrocyte zinc-protoporphyrin to protoporphyrin IX in disease and its significance in the mechanism of lead toxicity on haem synthesis (pt 3). Ann Clin Biochem 35:422, 1998

67. Solliway BM, Schaffer A, Pratt H, et al: Effects of exposure to lead on selected biochemical and haematological variables. Pharmacol Toxicol 78:18, 1996

68. Smith MT: Overview of benzene-induced aplastic anaemia. Eur J Haematol 57:107, 1996

69. Cullen MR, Solomon L, Pace PE, et al: Morphologic, biochemical and cytogenetic studies of bone marrow and circulating blood cells in painters exposed to ethylene glycol ethers. Environ Res 59:250, 1992

70. Cullen MR: Endocrine disorders. Textbook of Clinical Occupational and Environmental Medicine, 2nd ed. Rosenstock L, Cullen M, Brodkin CA, et al, Eds. WB Saunders Co, Philadelphia, 2005, p 609

71. Reproductive Hazards. U.S. Department of Labor. Occupational Health and Safety Administraiton, Washington, DC, 2005
http://www.osha.gov/SLTC/reproductivehazards

72. Khattak S, Moghtader GK, McMartin K, et al: Pregnancy outcome following gestational exposure to organic solvents. JAMA 281:12, 1999

73. Seltzer JM: Effects of the Indoor Environment on Health. Occupational Medicine: State of the Art Reviews, Vol 10, No 1. OEM Press, 1995

74. Redlich CA, Sparer JS, Cullen MR: Sick building syndrome. Lancet 349:1013, 1997

75. Mitchell LS, Donnay A, Hoover DA, et al: Immunologic parameters of multiple chemical sensitivity. Occup Med State Art Rev 15:539, 2000

76. Husman T: Health effects of indoor-air microorganisms. Scand J Work Environ Health 22:5, 1996

77. Bascom R, Kesavanathan J, Swift DL: Human susceptibility to indoor contaminants. Occup Med 10:119, 1995

VII HEALTH ADVICE FOR INTERNATIONAL TRAVELERS

PETER F. WELLER, M.D.

The provision of health advice and the administration of prophylactic measures can help reduce the morbid and, at times, mortal risks of infectious illnesses that may be acquired during international travel. The Centers for Disease Control and Prevention (CDC) publishes *Health Information for International Travel*, which provides information on required and recommended vaccinations and malaria prophylaxis, as well as general advice.[1] *Health Information for International Travel 2003–2004* is available for purchase on the Internet, at http://bookstore.phf.org/cat24.htm, or by phone, at 1-877-252-1200. Other information regarding international travel is readily available on the CDC Web site (http://www.cdc.gov), such as the Green Sheet, which provides reports of cruise-ship sanitation inspections; detailed guidelines on the need for yellow fever immunizations; guidelines on international health issues for travelers by country; recently recognized disease outbreaks; and general guidelines on immunizations and other medical issues for travelers. Information may also be obtained from state public health departments, local physicians or clinics catering to travelers, the embassies of individual countries, and Internet-based advisory services. Even the most up-to-date information sources, however, may not be able to provide precise information on specific diseases prevalent in specific locales, because mechanisms for recognizing and reporting diseases are often lacking in developing areas.

Pretravel Evaluation and Immunizations

Medical consultation should be obtained at least 1 month before international travel to allow time for immunizations [*see Table 1*]. A general patient medical history should be obtained to define pertinent underlying medical conditions. For instance, splenectomy predisposes a person to more severe malaria, babesiosis, and infections with encapsulated bacteria, including meningococcal infections. A history of allergies to antimicrobial agents or to other components of vaccines should be determined. Knowledge of the duration and purpose of a trip, as well as of the countries and locales to be visited, can help in estimating the risks of exposure to endemic diseases. In addition, specific groups of travelers—including pregnant women; persons with HIV; persons with chronic diseases such as chronic obstructive pulmonary disease, diabetes mellitus, hypercoagulable states, and cardiovascular disease; and health care workers—may require more time before travel to address their potentially altered needs for immunization and prophylaxis.

Table 1 Guidelines for Immunizations for Travelers

	Asia	*Eastern Mediterranean, North Africa*	*Middle East*	*Sub-Saharan Africa*	*Pacific Islands*	*Caribbean, Mexico, Central and South America*	*North America, Europe, Japan, Australia, New Zealand*
Yellow fever				X (some countries)		X (some South American countries)	
Cholera*							
Polio	X (some countries			X (some countries			
Tetanus/diphtheria (booster every 10 yr)	X	X	X	X	X	X	X
Measles (if born after 1957 and not recipient of 2 doses of vaccine)	X	X	X	X	X	X	X
Typhoid	X	X	X	X	X	X	
Rabies (for prolonged visits)	X	X	X	X		X	
Hepatitis A	X	X	X	X	X	X	
Hepatitis B (especially for prolonged visits)	X	X	X	X	X	X	
Meningococci	X (during outbreaks)	X (during outbreaks)	X (especially Mecca, during Hajj)	X (especially in meningococcal "belt" countries)		X (during outbreaks)	
Japanese encephalitis	X†						

*Only if required by a country.
†Prolonged visits to some regions.

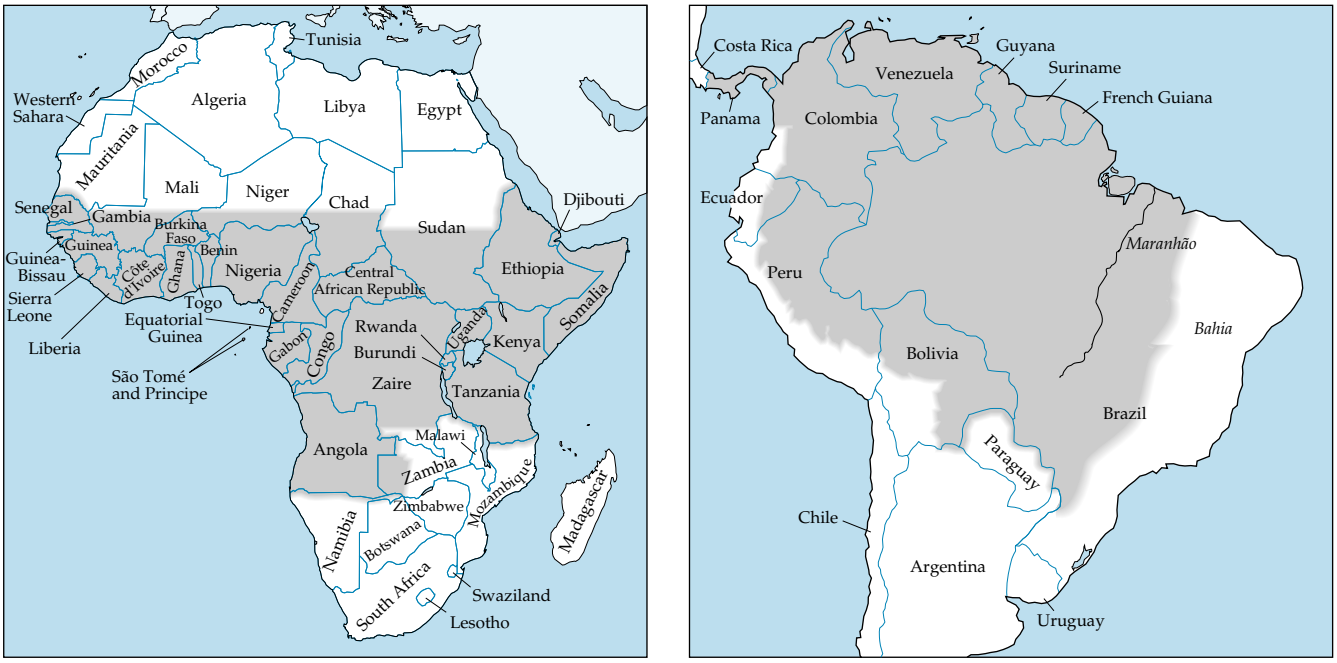


Figure 1 **Yellow fever (gray areas) is endemic in parts of Africa (left) and South America (right). Several countries consider these zones infected areas and require an International Certificate of Vaccination against yellow fever from travelers from these zones.**

REQUIRED IMMUNIZATION

The only immunization legally required for entrance into specific countries is that for yellow fever.

Yellow Fever

Yellow fever is a mosquito-transmitted viral infection, whose severity may range from an influenzalike illness to potentially fatal hepatitis and hemorrhagic fever. Yellow fever occurs only in equatorial Africa and in areas of tropical South America [*see Figure 1*].[1] Persons older than 6 months visiting countries where yellow fever is known to exist should be immunized. In addition, those traveling outside of urban areas in countries that are in the yellow fever endemic zones but are not officially reporting the infection should be immunized because the disease may be underrecognized. Some countries, especially in Asia, may require yellow fever immunizations for entry, especially for persons who have traveled in potentially endemic countries.[1] Yellow fever vaccine, a live virus vaccine grown in chick embryos, is effective.

Although generally safe, yellow fever vaccines are uncommonly associated with encephalitis (referred to as yellow fever vaccine–associated neurotropic disease [YEL-AND]) and a potentially fatal multiorgan system failure (referred to as yellow fever vaccine–associated viscerotropic disease [YEL-AVD]). In the United States, estimated rates of these two complications are four to six cases per million doses for encephalitis and three to five cases per million doses for multiorgan failure.[1] However, the risks of illness and death due to yellow fever in an unvaccinated traveler are greater, estimated to be one per 1,000 and one per 5,000 a month, respectively.[1] Thus, for those entering yellow fever endemic areas, yellow fever vaccination is indicated; however, vaccine use should be limited to those who are truly at risk. This is especially true for travelers older than 60 years, because they appear to be at greater risk for developing YEL-AVD.[2] For these travelers, vaccine may be indicated only if the risk of potential exposure is high (e.g., when travel in endemic areas outside urban centers is anticipated). A history of thymus disorders appears to be a contraindication for vaccine. In an analysis of 23 vaccinated persons who developed YEL-AVD, four (17%) had a history of thymus disease.[2] Travelers with a history of thymus disorders or dysfunction, including myasthenia gravis, thymoma, thymectomy, or DiGeorge syndrome, should not receive yellow fever vaccine. If travel to yellow fever–endemic regions cannot be deferred, persons with thymus disorders should be advised to use N,N-diethyl-m-toluamide (DEET) and permethrin to reduce mosquito bites [*see* Insect Repellents and Avoidance, *below*].[2]

Initially, a single subcutaneous dose of 0.5 ml is given. A booster dose is required every 10 years. Immunizations, which are recorded on the International Certificate of Vaccination, are available only from designated physicians and centers, an updated listing of which can be found on the Internet at http://www2.ncid.cdc.gov/travel/yellowfever. Yellow fever vaccine, which contains both egg proteins and gelatin, rarely causes anaphylaxis.[3] For those persons allergic to egg proteins or gelatin, skin testing with yellow fever vaccine (per directions included on the package insert) may help determine whether vaccine can be given safely.

Because yellow fever vaccine is a live virus vaccine, cautions and potential contraindications to its use apply to those who are pregnant, lactating, or immunocompromised. In these persons, if the sole indication for administration of yellow fever vaccine is to satisfy legal requirements for entry, a physician's letter documenting the contraindications to vaccination can be provided to the traveler; in addition, advice should be sought from the embassy or consulate of the country or countries to be visited.

Influenza

Travelers are at increased risk for influenza infection.[1] Influenza, like hepatitis A, has become one of the more common infections in travelers that are preventable by vaccine. In temperate countries, influenza is prevalent in the winter months; whereas in the tropics, influenza transmission occurs year-round. For travelers to the Northern Hemisphere, the risk is greatest during December through February; and for travelers to the Southern Hemisphere, the risk is greatest from April through September. Summertime outbreaks of influenza have occurred on cruise ships in the Northern and Southern hemispheres. For travelers to tropical countries, the risk of influenza exists throughout the year.[1]

Persons at high risk for influenza, including those older than 50 years, should receive influenza vaccine (1) if influenza vaccine was not received during the preceding fall or winter, (2) if travel is planned to the tropics, (3) if travel is planned with large groups of tourists (e.g., on cruise ships), or (4) if travel is planned during seasons in which influenza is prevalent.[1,4] In North America, travel-related influenza vaccination should be administered in the spring, if possible, because vaccine may be unavailable in the summer.

Cholera

Cholera is caused by toxigenic *Vibrio cholerae* groups 01 and 0139. Although the number of cases of cholera seen in the United States has increased in recent years, many of the cases have been the result of the illness being imported into the United States by travelers from other countries; tourists from the United States visiting other countries have only rarely been infected. Cholera is acquired by ingestion of contaminated water, ice, or food, including raw or undercooked fish and shellfish. Travelers in endemic regions should be advised of the precautions to be followed to minimize risks of acquiring cholera and other enteric infections [*see* Travel-Related Illness, *below*] and of the importance of rehydration in the treatment of cholera. Dietary precautions include consuming only boiled or treated water, eating thoroughly cooked food, avoiding all fruit not peeled by onself, and avoiding undercooked or raw fish or shellfish, including seviche.

Routine immunization is not recommended for travelers.[1] In the unlikely event that a locale requires immunization for cholera, immunization would need to be obtained outside the United States in countries in which current cholera vaccine is available. Currently, no country requires proof of cholera immunization as a condition for entry, and the World Health Organization recommends against such a requirement. Some local authorities, however, may require immunization (to determine local requirements, travelers may consult the embassies of the countries to which they will be traveling). The only cholera vaccine licensed for use in the United States is no longer manufactured. In other countries, two cholera vaccines (Dukoral, from Biotec AB, and Mutacol, from Berna) have been licensed for use, but neither of these vaccines is indicated for most travelers.[1] Travelers who are at risk for cholera and who expect to travel to areas remote from medical care should take with them packets of oral rehydration salts. Antimicrobial agents often employed for therapy for traveler's diarrhea, such as ciprofloxacin, are usually very effective in helping terminate cholera infections.[5]

Poliomyelitis

Travelers to countries in which polio is endemic or in which there is a current epidemic are at risk for the disease and should be immunized. Countries considered to be free of wild poliovirus are all countries in the Western Hemisphere, the Western Pacific Region (which includes China), and the European region.[1] Polio transmission continues in some developing countries, including Afghanistan, India, Pakistan, Nigeria, and Niger, although efforts to achieve global eradication of polio are ongoing.

Travelers who were immunized previously should receive one booster dose of polio vaccine. Oral live virus vaccine is no longer recommended for immunizations in the United States.[6] The inactivated vaccine is preferred to avoid the small risk of paralytic disease from the oral vaccine. Patients with an altered immune status should receive inactivated vaccine. Children who have not been immunized should receive a full series of immunizations with inactivated polio vaccine. Adults who have not been immunized should receive a series of three doses of enhanced-potency inactivated vaccine.[1] If there is insufficient time before travel for at least three doses of inactivated vaccine to be given at intervals of 1 to 2 months, the following alternatives are recommended[1]: if less than 1 month is available before travel, a single dose of inactivated vaccine is given; if between 1 and 2 months is available before travel, two doses of inactivated vaccine are administered 4 weeks apart. Travelers who were incompletely immunized previously should receive the remaining required doses of vaccine.

Tetanus and Diphtheria

A tetanus-diphtheria booster should be administered every 10 years [*see 7:V Anaerobic Infections*].[1] Older persons and women are more likely to lack the tetanus and diphtheria antibodies and thus are more likely to require boosters.[7,8]

Pneumococcal Infections

There are no data on the risk to travelers of acquiring pneumococcal infections; however, those at risk, including those older than 65 years [*see 7:I Infections Due to Gram-Positive Cocci*], are recommended as candidates to receive pneumococcal vaccinations.

Measles

Because of the declining prevalence of measles in the United States, disease imported by immigrants and by returning residents accounts for an increasing proportion of cases in this country. Measles may be acquired during travel in developed countries, including those in Europe and Asia, as well as in less developed countries.[1] Most persons who were born before 1957 are immune because of natural exposure and do not require vaccination. Persons who were born after 1956 who either have not been immunized or were immunized before 1980 and who have neither serologic evidence of infection nor a history of physician-diagnosed measles should be immunized with a single subcutaneous dose of measles vaccine before travel. Measles vaccine is contraindicated for both pregnant and immunodeficient patients. HIV-infected patients, unless they are severely immunocompromised, should be immunized before travel because measles can be severe and even fatal in persons with HIV infection.[1,9]

Typhoid

Salmonella typhi infection is prevalent in many areas of Asia, Africa, and Latin America. Typhoid is acquired from contaminated food or water. Although the overall risk of acquiring typhoid during travel remains low (2.3 million cases per million

travelers),[10] foreign travel accounted for 74% of 1,393 cases of typhoid reported to the CDC between 1994 and 1999.[11] The risk was greatest for those traveling to the Indian subcontinent (India, Pakistan, and Bangladesh) and Haiti. Of note, even those traveling for no more than a couple of weeks were at risk of acquiring typhoid. Given the safety of current typhoid vaccines, typhoid vaccination should be considered for short-term travel in high-risk areas, as well as for any travel to areas off the usual tourist itinerary.[11]

Two typhoid vaccines are available for use in the United States. One typhoid vaccine is an oral vaccine (Vivotif Berna, from Berna) that uses the attenuated Ty21a strain of *S. typhi;* this vaccine does not cause the local and systemic side effects frequently produced by the older, parenteral vaccine. The oral vaccine is supplied as a packet of four enteric-coated capsules that must be refrigerated. Patients need explicit guidance on refrigerating the vaccine because failure to do so might compromise its efficacy.[12] At least 2 weeks before departure, the traveler takes one capsule every other day until all four capsules have been taken. Because mefloquine and antibiotics inhibit the growth in vitro of *S. typhi* strains, including Ty21a, it is prudent to separate the oral administrations of mefloquine and antibiotics and of Ty21a vaccine by 24 hours.[1] It is recommended that a booster dose of Ty21a vaccine, consisting of four capsules taken on alternate days, be given every 5 years to persons who continue to be at risk for exposure to typhoid. The safety of the oral vaccine has not been established for patients with deficient humoral or cell-mediated immunity, and thus, patients with congenital or acquired immunodeficiencies should not receive it. It may be given to children 6 years of age or older.

The second typhoid vaccine is a capsular polysaccharide vaccine for parenteral use (Typhim Vi, from Aventis Pasteur). Primary vaccination consists of one I.M. dose of 0.5 ml; the same dose is administered as a booster every 2 years. The vaccine is well tolerated but, like the oral vaccine, protects only 50% to 80% of recipients.[1,13] This vaccine is safe for immunocompromised persons, including HIV-infected patients.[1] The only contraindication to its use is a history of serious reactions to the vaccine. It may be given to children 2 years of age and older.

Rabies

Rabies vaccine—either human diploid cell rabies vaccine (HDCV), purified chick embryo cell vaccine (PCEC), or rabies vaccine adsorbed (RVA)—is an inactivated viral preparation. Immunizations with either of the three vaccine preparations consists of three I.M. doses, 1 ml each, administered on days 0, 7, and 21 or 28.[1] Preexposure immunization with rabies vaccine is not indicated for most travelers but should be strongly considered for persons who anticipate contact with wild animals or who are living for a month or more where rabies is endemic. Dog rabies is present in most countries of Asia, Africa, and Central and South America and is prevalent in parts of Brazil, Bolivia, Mexico, El Salvador, Guatemala, Colombia, Ecuador, Peru, India, Nepal, the Philippines, Sri Lanka, Thailand, and Vietnam.[1] Preexposure immunization does not eliminate the need for postexposure immunization but abbreviates its course and eliminates the need to administer rabies immune globulin. If left untreated, rabies is fatal, and postexposure rabies immune globulin and postexposure vaccine are frequently unavailable in many areas of the world.

Plague

Plague vaccine is no longer commercially available. Vaccination against plague is not indicated for most travelers.[1] However, prophylaxis should be considered for travelers to areas in which plague is epidemic or actively epizootic. For adults, tetracycline or doxycycline is appropriate prophylactic therapy; for children younger than 8 years, trimethoprim-sulfamethoxazole is recommended.[1]

Hepatitis A

Hepatitis A is prevalent in many less-developed countries [*see Figure 2*] and is the most common infection acquired by travelers that is preventable by vaccine. In visitors to developing countries, even those staying in luxury hotels, the incidence of hepatitis A in unprotected travelers is about 3 per 1,000 travelers per month of stay, and this rate rises to 20 per 1,000 travelers per month for those eating or drinking under poor hygienic conditions.[14]

Immunization for hepatitis A is recommended for persons who will be traveling or working in countries with intermediate or high endemicity for hepatitis A infection.[1] Although hepatitis A previously was prevented solely by the administration of immune globulin, two monovalent inactivated hepatitis A vaccines and a combined hepatitis A and hepatitis B vaccine are now available. The monovalent vaccines, HAVRIX (GlaxoSmithKline) and VAQTA (Merck), have proved safe and highly effective.[15] For adults, two I.M. 1.0 ml doses should be administered in the deltoid muscle at 0 and 6 to 12 months. For persons between 2 and 18 years of age, two doses, 0.5 ml each, should be administered at 0 and 6 to 12 months. VAQTA contains no preservative, whereas HAVRIX contains 2-phenoxyethanol. The bivalent hepatitis A and hepatitis B vaccine, TWINRIX (GlaxoSmithKline), is likewise safe and effective and is administered to those 18 years of age or older in three I.M. 1.0 ml doses at 0, 1, and 6 months. TWINRIX contains 2-phenoxyethanol.

With the monovalent vaccines, many persons who have been vaccinated will have detectable antibody responses within 2 weeks after the first dose; 94% to 100% of persons treated will have protective levels of antibody by 1 month after the first dose. The second dose of vaccine provides longer-term protection. If the immunization schedule is unduly interrupted, it is not necessary to restart the full regimen; the second dose may simply be administered. A vaccination series started with one brand of vaccine may be completed with the same or the other brand of hepatitis A vaccine. Travelers who receive vaccine less than 2 weeks before travel are at risk for acquiring hepatitis and should also receive immune globulin, given at an injection site different from the one for vaccine.

For travelers who are allergic to vaccine components or who opt not to receive the vaccine, immune globulin should be administered. Administration of immune globulin should begin shortly before departure in a dose of 2.0 ml I.M. for adults (1.0 ml for patients, including children, weighing 23 to 45 kg; 0.5 ml for those weighing less than 23 kg). If the stay is to be longer than 3 months, the adult dose is 5.0 ml (2.5 ml for patients weighing 23 to 45 kg; 1.0 ml for those weighing less than 23 kg). If the duration of stay is prolonged, the latter dosage schedule should be repeated every 4 to 6 months. Immune globulin should be given at least 2 weeks after measles, mumps, or rubella live virus vaccines. Conversely, these vaccines should be given at least 3 months after immune globulin. Immune globulin does not interfere with the immune response to killed virus vaccines or to yellow fever or polio vaccines.

Because immune globulin has been in limited supply and because the hepatitis A vaccines have proved to be highly effective against hepatitis A infection, which is frequent in travelers, immunization with hepatitis A vaccine has become the principal approach for preventing hepatitis A infection in travelers. The hepatitis A vaccines are safe in pregnancy and immunosuppression.

Hepatitis B

The risk to travelers of acquiring hepatitis B is generally low, compared with the risk of acquiring hepatitis A. The risk increases, however, in regions where hepatitis B is highly prevalent [*see Figure 3*], if there is contact with blood or bodily secretions, if sexual contact with infected persons occurs, or if travel is prolonged.[1] Immunization for hepatitis B, which is recommended for all persons who work in health care fields with potential exposure to human blood, is especially important for medical workers traveling in countries with high or intermediate hepatitis B endemicity. Hepatitis B immunization should be considered for persons residing for more than 6 months in regions where hepatitis B is endemic and for persons with potential contact with blood (including those receiving tattoos or body piercing), potential sexual contact, or potential need for medical or dental procedures.

Two monovalent hepatitis B vaccines are available, both of which include recombinant HBsAg (hepatitis B surface antigen) protein produced in yeast. Except for rare hypersensitivity reactions to vaccine components, including yeast proteins, the two recombinant vaccines are safe and efficacious; there are no other medical contraindications, including pregnancy and immunosuppression, for administration of these vaccines. The two vaccines, Recombivax HB (Merck) and Engerix-B (GlaxoSmithKline), are administered in three I.M. doses: at 0, 1, and 6 months. Engerix-B may also be given in four doses: at 0, 1, 2, and 12 months. Immunization should start 6 months before travel, but if this schedule is not feasible, some protection is afforded by one or two doses administered before travel. Full protection will be achieved in most cases by completion of the three-dose or four-dose schedule.[1] For travelers who will depart before the recommended series can be completed, an accelerated regimen, involving doses given on days 0, 7, and 14, can be administered; the accelerated regimen is not approved by the Food and Drug Administration. Travelers receiving the accelerated course should receive a booster at least 6 months later to provide long-term immunity.[1]

An additional vaccine for hepatitis B is TWINRIX, a combined hepatitis A and hepatitis B vaccine. Primary immunization with TWINRIX consists of three doses administered at 0, 1, and 6 months. The bivalent vaccine can be used to complete immunization series started with monovalent hepatitis A and B vaccines.

Meningococcal Disease

Although acquisition of meningococcal disease is uncommon in travelers from the United States, immunization should be considered for travelers to areas with recognized epidemics or to regions where such disease is hyperendemic, especially if prolonged contact with the local populace is anticipated. Epidemics of meningococcal disease are frequent in the area of sub-Saharan Africa extending from Guinea in the west to Ethiopia in the east [*see Figure 4*]. Vaccination against meningococcal disease is legally required only for pilgrims who make the Hajj pilgrimage to Mecca, Saudi Arabia. Routine immunization is also indicated for persons who have either deficiencies of terminal complement components or functional or anatomic asplenia. The currently available quadrivalent vaccine is composed of meningococcal

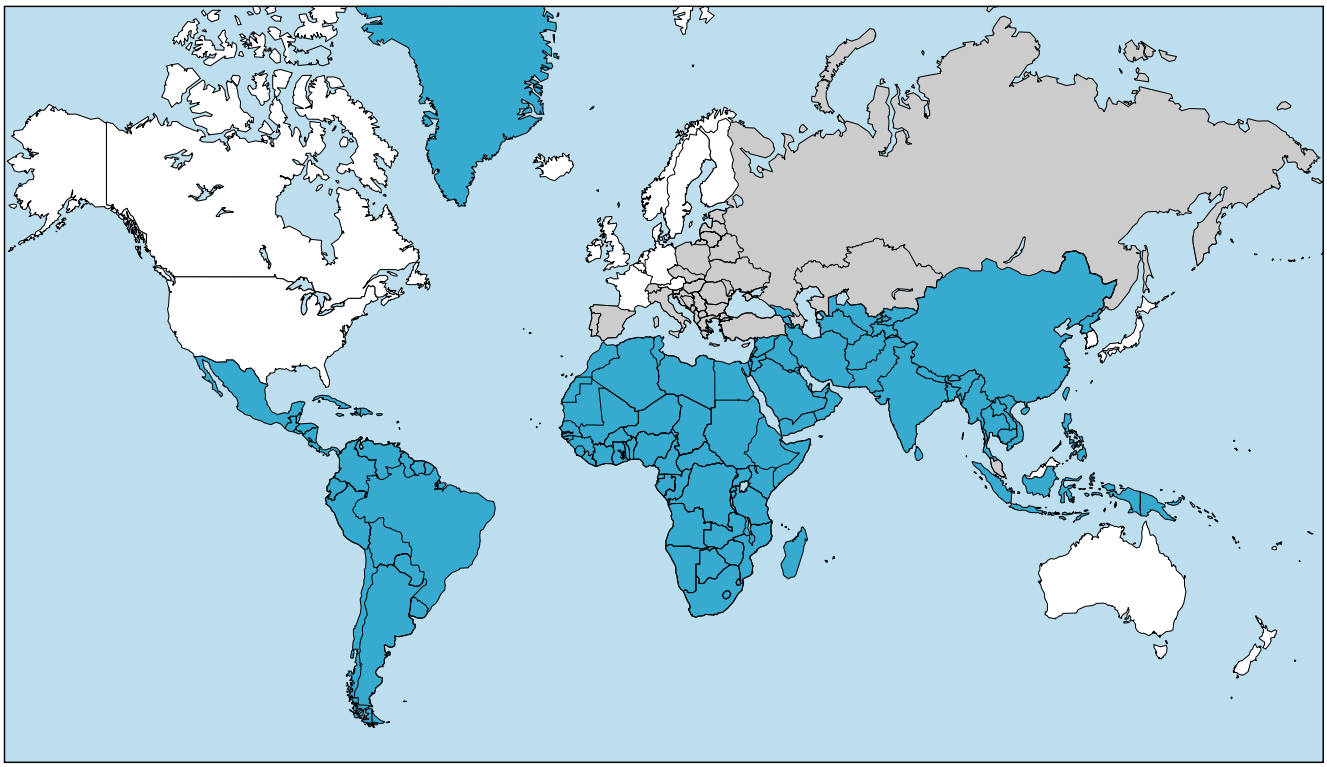


Figure 2 **The prevalence of hepatitis A is high in those countries shaded blue, intermediate in those shaded gray, and low in the white areas of the map.**

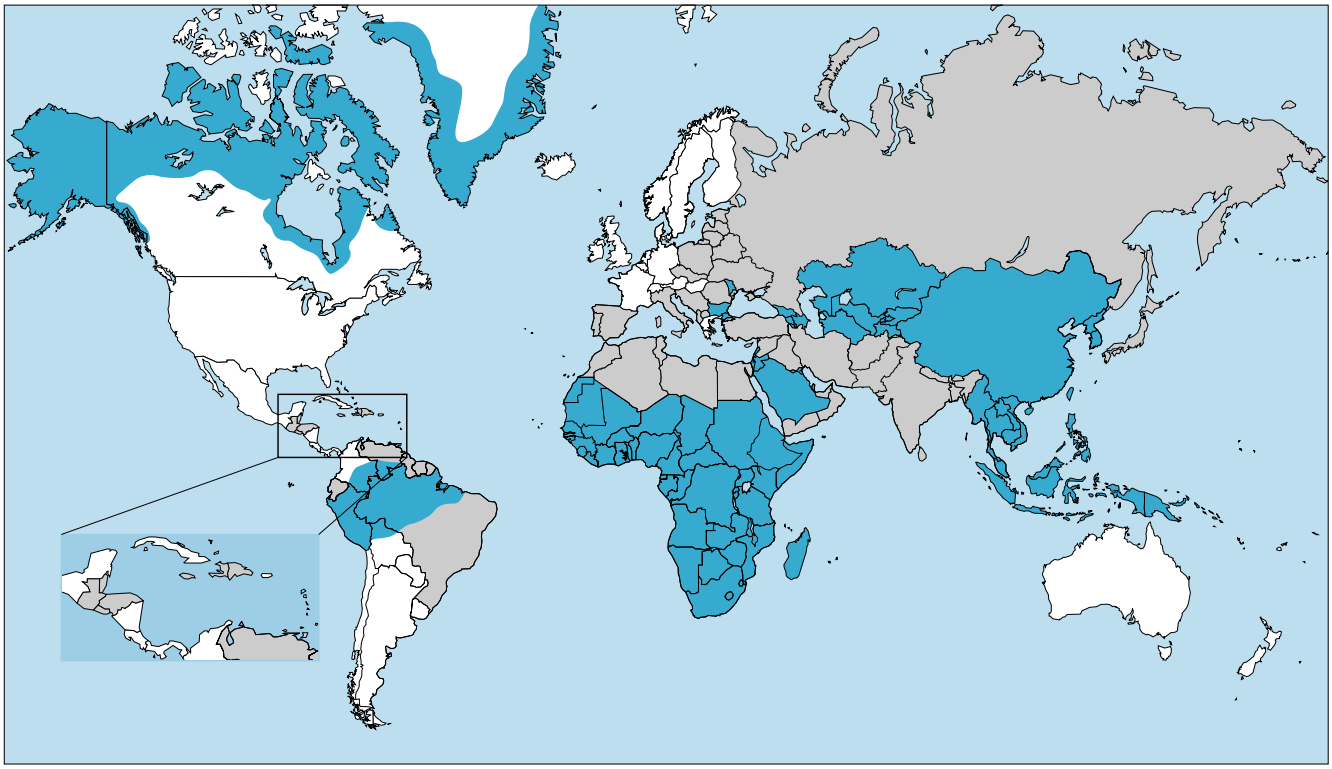


Figure 3 **Hepatitis B is highly endemic in those countries shaded blue (prevalence > 8%). Those regions shaded gray, where the prevalence is 2% to 7%, are considered to be of intermediate endemicity. The prevalence of hepatitis is less than 2% in the white areas of the map.**

polysaccharides from *Neisseria meningitidis* serogroups A, C, Y, and W-135. A single 0.5 ml subcutaneous dose of vaccine is administered to both adults and children and will induce an antibody response in 10 to 14 days.[1] Duration of immunity is at least 3 years.

Japanese Encephalitis

Japanese encephalitis, an arboviral infection transmitted by mosquitoes, may occur in epidemics during the late summer and autumn in northern tropical areas and temperate regions of some countries. The risk of acquiring Japanese encephalitis infection varies by season and geographic area [*see Table 2*].[1] The disease rarely occurs in Hong Kong or Japan. Persons at highest risk are those who live for extended periods in endemic or epidemic areas. The risk for short-term travelers to urban centers is low, and in temperate countries, the risk for travelers to either an urban or a rural area is negligible during the winter.

Although Japanese encephalitis is highly uncommon, prevention is important for those traveling specifically to epidemic or endemic areas [*see Table 2*], because the risk of serious neurologic sequelae is high. Exposure to mosquitoes should be minimized by the use of insect repellents, protective clothing, and mosquito screens. Also, vaccination should be considered for persons traveling during summer monsoon months, for those visiting rural areas, and for those planning to stay more than 1 month in urban or rural areas.

Vaccination is not usually recommended for travelers to Singapore or Hong Kong, urban Japan or China, or high-altitude regions in Nepal. An effective formalin-inactivated, mouse-derived vaccine (JE-Vax, Aventis Pasteur) has been licensed by the FDA. Primary immunization for persons older than 3 years consists of three doses, 1 ml each, administered subcutaneously on days 0, 7, and 30. An abbreviated schedule of 0, 7, and 14 days can be used if there is insufficient time before travel to administer the standard immunization. A booster dose of 1 ml may be administered after 2 years. About 20% of recipients of JE-Vax vaccine experience local reactions and mild systemic side effects (e.g., fever, headache, myalgias, and malaise).[1]

Allergic reactions, including generalized urticaria, angioedema, respiratory distress, and anaphylaxis, have developed in about six per 1,000 recipients; at times, the onset of allergic reaction is delayed for hours or even a week after vaccine administration. Those with a history of urticaria and allergies (including hay fever and reactions to hymenoptera venom) appear to have a greater risk of developing allergic reactions to the vaccine. Reactions have been responsive to epinephrine, antihistamines, steroids, or a combination of these agents.[1,16] Because of late-developing allergic reactions, immunizations should be completed 10 days before travel; vaccine recipients need to be advised to remain accessible to emergency medical care.

Tick-Borne Encephalitis

Tick-borne encephalitis is a viral infection of the central nervous system that is transmitted by ticks. The disease occurs in Scandinavia, western and central Europe, and countries of the former Soviet Union. The disease is transmitted principally from April through August, when the tick vector, *Ixodes ricinus,* is most active. Infections may also be acquired by consumption of unpasteurized dairy products from infected cows, goats, or sheep. Effective vaccines are available in Europe and in many travel clinics in Canada; vaccines are not available in the United States. Vaccination should be considered for travelers who anticipate extensive outdoor exposure (e.g., camping or related activities) in the endemic regions during the spring and summer months.[1]

VACCINE CONTRAINDICATIONS

Vaccines that contain live attenuated viruses (i.e., oral polio, measles, mumps, rubella, and yellow fever vaccines) should not be given to pregnant women or to persons who have known or potential immunodeficiencies (e.g., leukemia, lymphoma, or a generalized malignant disorder) or who are receiving corticosteroids, alkylating agents, antimetabolites, or irradiation. Oral polio vaccine, which is no longer recommended in the United States, should not be given to a patient if an immunodeficient person resides in the same household. If a pregnant woman cannot defer travel to areas of high risk for yellow fever, yellow fever vaccine may be given.[1] For travelers infected with HIV, immunization with live oral polio and attenuated oral typhoid vaccines should be avoided in favor of killed parenteral vaccines. The risks of live yellow fever vaccine have not been defined for HIV-infected persons, but persons with asymptomatic HIV infection who cannot avoid exposure in areas endemic for yellow fever should be offered the choice of immunization.[1] Because measles can be severe in patients with HIV, measles immunization should be provided, unless the patient is severely immunocompromised (i.e., total $CD4^+$ T cell count < 200 µl).[9]

Contraindications to vaccination also include hypersensitivity to components of the vaccine. Neomycin and gelatin are present in some vaccines. Persons who have immediate hypersensitivity reactions to neomycin, gelatin, or preservative agents should avoid vaccines containing these substances. Yellow fever vaccine, which contains egg proteins and gelatin, may be contraindicated in patients who have allergic reactions to these proteins. In general, there is a poor correlation between a history of egg sensitivity and skin-test reactivity to egg antigen. The most reliable predictor of reactions to egg-containing vaccines is skin testing with the vaccine itself.[1] If travel plans cannot be changed, persons who have positive skin tests or known egg hypersensitivity (i.e., urticaria, oropharyngeal swelling, bronchospasm, or hypotension) should be given a letter documenting the contraindication to immunization and obtain a waiver before travel from the embassy of any country requiring yellow fever immunization.

INSECT REPELLENTS AND AVOIDANCE

To reduce the risk of all mosquito-borne infections (e.g., malaria, yellow fever, and dengue fever), travelers should be instructed about the importance of minimizing the potential for insect bites. The most effective insect repellents contain DEET.[17,18] DEET is available in many products in concentrations ranging from 25% to more than 75% and repels mosquitoes, ticks, fleas, and biting flies. Protection lasts for several hours but is shortened by losses from swimming, washing, rainfall, sweating, and wiping. A long-acting formulation, which contains polymer to limit the losses of DEET that result from dermal absorption and evaporation, has been developed by the military and is available in the United States as Ultrathon (3M).

The absorption of DEET through the skin can cause such adverse reactions as dermatitis, allergic reactions, and neurotoxicity. Potential toxicity can be avoided by using solutions of 30% to 35% DEET and following instructions for its use. The repellent should be applied sparingly to clothing and exposed skin only. The product should be applied carefully to avoid introducing it into the eyes, to avoid contact with wounds and sensitive skin, and to prevent inhalation or ingestion. Clothing and bed netting can also be treated with permethrin for protection against mosquitoes and ticks.[19] Treated clothing will effectively repel mosquitoes for more than 1 week even with washing and field use. Permethrin is available, often in outdoor supply stores, as a nonstaining aerosol clothing spray (e.g., Permanone Tick Repellent).

Figure 4 **Epidemics of meningococcal disease are frequent in the area of sub-Saharan Africa that extends from Guinea in the west to Ethiopia in the east.**

Table 2 Risk of Japanese Encephalitis by Country, Region, and Season[1]

Country	*Affected Areas*	*Transmission Season*
Australia	Islands of Torres Strait	Probably year-round transmission
Bangladesh	Few data, probably widespread	Possibly July through December
Bhutan	No data	No data
Brunei	Presumed to be sporadic-endemic, as in Malaysia	Presumed year-round transmission
Myanmar (Burma)	Presumed to be endemic-hyperendemic countrywide	Presumed to be May through October
Cambodia	Presumed to be endemic-hyperendemic countrywide	Presumed to be May through October
India	Reported cases from many states	South India: May through October in Goa, October through January in Tamil Nadu, and August through December in Karnataka Andhra Pradesh: September through December North India: July through December
Indonesia	Kalimantan, Bali, Nusa, Tenggara, Sulawesi, Mollucas, Irian, Jaya, and Lombok	Probably year-round risk (varies by island); peak risks associated with rainfall, rice cultivation, and presence of pigs Peak periods of risk are November through March and, in some years, June through July
Japan	Rare, sporadic cases on all islands, except Hokkaido	June through September; Ryukyu Islands (Okinawa), April through October
Korea	Sporadic in South Korea; endemic with occasional outbreaks	July through October
Laos	Presumed to be endemic-hyperendemic countrywide	Presumed to be May through October
Malaysia	Sporadic-endemic in all states of Malay Peninsula, Sarawak, and probably Sabah	No seasonal pattern; year-round transmission
Nepal	Hyperendemic in southern lowlands (Terai)	July through November
People's Republic of China	Hyperendemic in southern China; periodically epidemic in temperate areas	Northern China: May through September Hong Kong and southern China: April through October
Pakistan	May be transmitted in central deltas	Presumed to be June through January
Philippines	Presumed to be endemic on all islands	Uncertain
Russia	Far eastern maritime areas south of Khabarovsk	Peak period July through September
Singapore	Rare cases	Year-round transmission; April peak
Sri Lanka	Endemic in all but mountainous areas; periodically epidemic in northern and central provinces	October through January; secondary peak of enzootic transmission May through June
Taiwan	Endemic-sporadic cases island-wide	April through October; June peak
Thailand	Hyperendemic in north; sporadic-endemic in south	May through October
Vietnam	Endemic-hyperendemic in all provinces	May through October

Malaria Chemoprophylaxis

The provision of appropriate malaria chemoprophylaxis is the most important preventive measure for travelers to malarious areas. Several hundred United States civilians contract malaria each year,[1] and infections from *Plasmodium falciparum* are potentially lethal and do cause deaths in travelers [*see 7:XXXIV Protozoan Infections*].[20] Morbidity and mortality are largely avoidable with chemoprophylaxis. Malaria is prevalent in parts of Mexico, Haiti, Central and South America, Africa, the Middle East, Turkey, the Indian subcontinent, Southeast Asia, China, the Malay archipelago, and Oceania. Chloroquine-resistant *P. falciparum* (CRPF) malaria occurs in most areas [*see Figure 5*]. Details on the prevalence by country and regions within countries of both malaria and CRPF malaria are reported annually by the CDC and may be accessed online (www.cdc.gov/travel/yb/index.htm).[1] Because even brief exposures to infected mosquitoes can transmit malaria infections, travel in malarious regions, no matter how brief, mandates the use of chemoprophylaxis. When uncertainty exists over the need for chemoprophylaxis, it should be initiated. If a traveler can ascertain that malaria is not a risk after arriving in an area, prophylaxis can be terminated as long as further travel into malarious areas is not planned.

Travelers should be advised that it is possible to acquire malaria despite prophylaxis and regardless of the prophylactic regimen used. Symptoms can begin as early as 8 days after infection or as late as several months after departure from a malarious area. Travelers should be cautioned to seek medical attention promptly for any febrile illness and to inform the physician of their prior itinerary. The wisdom of general protective measures against mosquito bites should also be stressed for all travelers to malarious areas. Because the vector mosquitoes usually feed at night, it is advisable to diminish exposure between dusk and dawn by remaining in screened areas; using mosquito net-

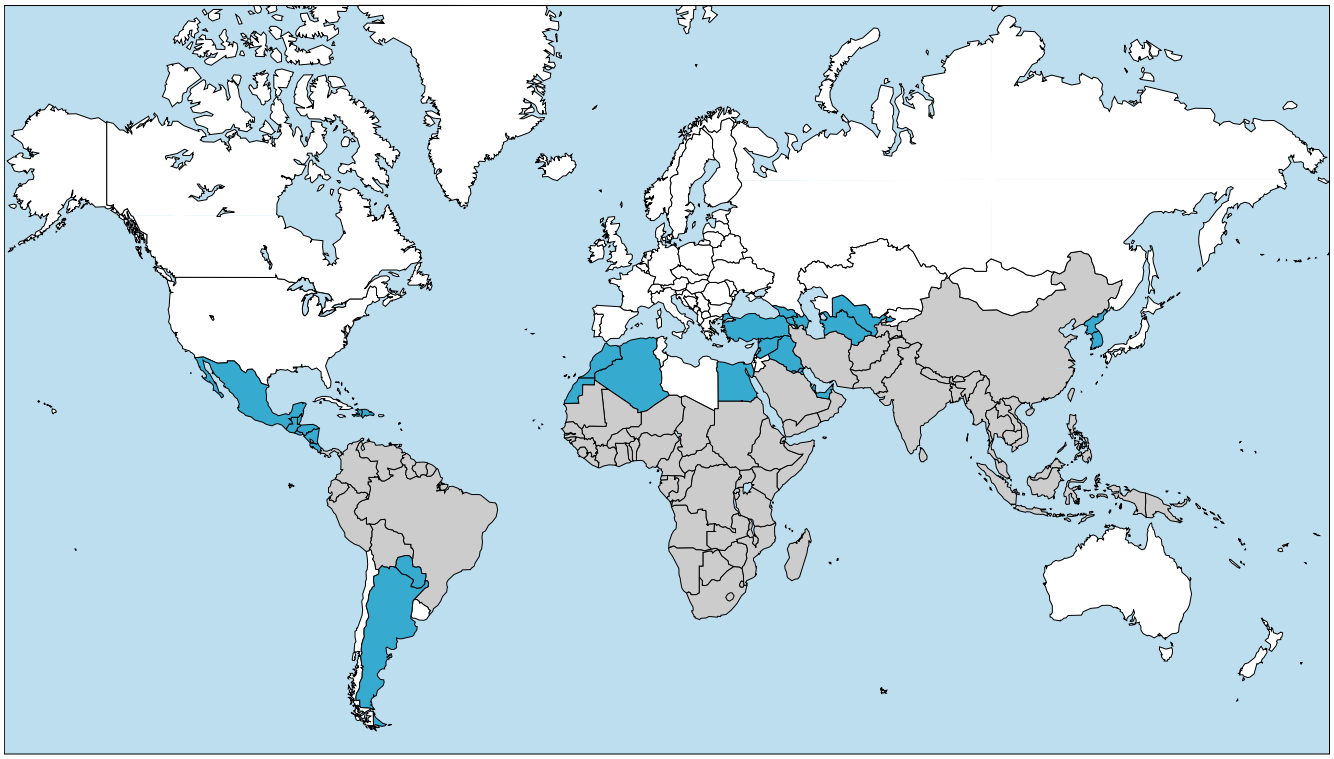


Figure 5 **This map displays the distribution of the chloroquine-resistant malaria (gray areas) and chloroquine-sensitive malaria (blue areas) in the Americas and in Asia, Europe, and Africa.**

ting, ideally treated with permethrin; covering exposed skin with clothing; and using insect repellent.

The choice of appropriate chemoprophylactic agents against malaria depends on the geographic areas to be visited and, importantly, whether these areas are endemic for CRPF [*see Figure 5*]. If travel is not to include areas where CRPF has been reported (e.g., Central America and the Caribbean), chloroquine remains the chemoprophylactic agent of choice. For most of the world, however, alternatives to chloroquine are required. Mainline alternatives to chloroquine include mefloquine and atovaquone-proguanil (Malarone).

Chloroquine

For those limited geographic regions not yet experiencing CRPF [*see Figure 5*], the chemoprophylactic agent of choice is chloroquine, given as either chloroquine phosphate (Aralen) or hydroxychloroquine sulfate (Plaquenil).[1] Chloroquine phosphate, 500 mg (300 mg of chloroquine base), or hydroxychloroquine sulfate, 400 mg (310 mg of hydroxychloroquine base), should be taken once weekly beginning 1 to 2 weeks before travel and continuing during the stay and for 4 weeks after departure from malarious areas. Minor side effects, including gastrointestinal disturbances, dizziness, blurred vision, and headache, may be alleviated by taking the drug after meals. Serious side effects are rare. Specifically, retinal injury, which can occur when high doses of chloroquine are used to treat rheumatoid arthritis, does not occur with the weekly dosages used for malaria prevention, even when such a regimen is continued for 5 years. However, deaths from malaria have occurred among tourists from the United States who avoided chloroquine prophylaxis out of a misguided concern for ocular toxicity.

Mefloquine

Mefloquine (Lariam) is active against CRPF and against *P. falciparum* that is resistant to sulfadoxine with pyrimethamine (Fansidar). With the now-widespread geographic prevalence of CRPF [*see Figure 5*], either mefloquine or atovaquone-proguanil is for many travelers the mainstay of malarial chemoprophylaxis. Strains of *P. falciparum* that are resistant to mefloquine, however, have been recognized in Africa and along the border between Thailand and Cambodia. Mefloquine, 250 mg, is taken once a week, beginning 1 to 2 weeks before travel and continuing during the stay and for 4 weeks after departure from a malarious area.[1] (This schedule is similar to that for chloroquine.) For travelers who will be immediately arriving in malarious areas, a loading dose of mefloquine (250 mg daily for the first 3 days) is advisable.

Despite the benefits of mefloquine to travelers in regions with CRPF malaria, mefloquine has acquired an unsalutary reputation. Mefloquine causes side effects, including nausea, dizziness, vertigo, light-headedness (described as an inability to concentrate), bad dreams, seizures, and psychosis. These reactions occur principally when the drug is given at therapeutic doses, which are higher than those given for prophylaxis. The incidence of psychosis or seizures has been about one per 10,000 travelers treated with chemoprophylactic mefloquine, which is comparable to the incidence associated with chloroquine use.[21] Other controlled trials have demonstrated that mefloquine is reasonably well tolerated in groups receiving this agent.[22-24] Thus, the uncommon and self-limited, but bothersome, side effects of mefloquine are to be weighed against the very real risks of serious and fatal malaria in many nonimmune travelers.

Mefloquine use has also been associated with sinus bradycardia and prolongation of the QT interval. Therefore, mefloquine

probably should not be used by persons with cardiac conduction abnormalities but may be used by patients without arrhythmias who are taking beta blockers.[1] Other contraindications to mefloquine include a history of serious neuropsychiatric disorders or seizures. Mefloquine appears to be safe and effective for young children.[25] Studies indicate that use of mefloquine in pregnancy during the second and third trimesters is not associated with adverse fetal or pregnancy outcomes; more limited data suggest that mefloquine is probably safe during the first trimester.[26-28] Mefloquine has no deleterious effects on fine motor skills, such as those required by airplane pilots.[29]

Atovaquone-Proguanil

Atovaquone-proguanil (Malarone) is available in many countries, including the United States, for the chemoprophylaxis of malaria. Atovaquone-proguanil is formulated as a fixed-dose tablet in adult strength (250 mg atovaquone/100 mg proguanil) and in pediatric strength (62.5 mg atovaquone/25 mg proguanil). For prophylaxis, one tablet is taken daily, beginning 1 to 2 days before travel and continuing for the duration of travel and for 1 week after departure from malarious areas. One, two, or three pediatric-strength tablets are taken by children weighing 11 to 20 kg, 21 to 30 kg, or 31 to 40 kg, respectively.

Atovaquone-proguanil is well tolerated; side effects, which are uncommon, are abdominal pain, nausea, vomiting, headache, and rash. Atovaquone-proguanil is safe and efficacious for prophylaxis of *P. vivax* and *P. falciparum* malaria, including CRPF. For *P. vivax* and *P. ovale* malaria, atovaquone-proguanil, like mefloquine and chloroquine, does not prevent development of hepatic hypnozoite stages, so treatment with primaquine (so-called terminal prophylaxis) may be necessary to prevent relapses with these species [*see* Primaquine, *below*]. Atovaquone-proguanil, therefore, is an alternative to mefloquine for malaria chemoprophylaxis[1] in regions of Thailand, Myanmar (Burma), and Cambodia where mefloquine-resistant *P. falciparum* malaria is present.

Doxycycline

Doxycycline, taken alone, is an alternative chemoprophylactic agent.[1] It should be taken in a dosage of 100 mg daily, beginning 1 to 2 days before travel and continuing for 4 weeks after departure from malarious areas. The use of doxycycline is appropriate for persons who are intolerant of sulfonamides, pyrimethamine, chloroquine, or mefloquine and for persons who are planning short-term visits in forested areas of Thailand, Myanmar (Burma), or Cambodia, where strains of malaria that are resistant to chloroquine, mefloquine, and sulfadoxine with pyrimethamine (Fansidar) are present.[1] Doxycycline may cause photosensitivity skin reactions and is contraindicated in pregnant women and in children younger than 8 years.

Proguanil

Proguanil (Paludrine) is not available in the United States but is available in Canada, Europe, and much of Africa. This agent, like pyrimethamine, is a dehydrofolate reductase inhibitor, and some strains of malaria are resistant to it. Proguanil (200 mg) is taken daily in combination with a weekly dose of chloroquine. The combination of proguanil and chloroquine, however, is much less effective than mefloquine or atovaquone-proguanil against chloroquine-resistant *P. falciparum* malaria and hence is not recommended.[1,30]

Primaquine

Primaquine may be used either as a single agent taken daily for chemoprophylaxis against all species of malaria or as an agent to eradicate residual intrahepatic stages of *P. vivax* and *P. ovale*. For the latter purpose, primaquine is administered during the last weeks of or just after a course of prophylaxis with either chloroquine or mefloquine. When intended as terminal prophylaxis, primaquine may be administered as 30 mg of the base daily for 14 days. Such terminal prophylaxis is generally reserved for persons who have had more than a casual potential exposure to *P. vivax* or *P. ovale*; other persons may be followed clinically and evaluated if they become symptomatic. For use as a primary chemoprophylactic agent, 30 mg of primaquine base is taken daily starting 1 day before travels and continuing for 2 days after departure from a malarious area.[31]

Because primaquine can cause severe hemolysis in patients who have glucose-6-phosphate dehydrogenase (G6PD) deficiency, this disorder must be excluded before the drug is administered. As a chemoprophylactic agent, primaquine is reserved for the rare individual who is unable to take other recommended chemoprophylactic regimens. CDC suggests primaquine be used only after consultation with malaria experts, including those at the CDC Malaria Hotline (1-770-488-7788).[1]

Sulfadoxine with Pyrimethamine

For chemoprophylaxis in areas where CRPF malaria occurs, it was formerly recommended that a single tablet of Fansidar, which contains 500 mg of long-acting sulfadoxine and 25 mg of pyrimethamine, be taken once a week along with chloroquine beginning 1 to 2 weeks before arrival in an endemic area and continuing for 4 weeks after departure from such an area. However, severe mucocutaneous reactions, including erythema multiforme, Stevens-Johnson syndrome, and toxic epidermal necrolysis, have developed after the use of two or more doses of Fansidar. These reactions produced fatalities with an incidence of about one per 11,000 to 20,000 travelers from the United States. Moreover, *P. falciparum* malaria is increasingly resistant to antifolate agents. Consequently, Fansidar is not recommended for chemoprophylactic use.

PROPHYLAXIS DURING PREGNANCY

Malaria infections represent a major health hazard to the mother and fetus.[32,33] Infections are potentially more serious during pregnancy and increase the risks of stillbirths, abortions, and other adverse pregnancy outcomes. For pregnant women who cannot defer travel or residence in malarious areas, chloroquine, which is without established teratogenicity, may be used.[1] Mefloquine appears to be safe in pregnancy.[26-28] For the pregnant traveler in regions with CRPF malaria, the benefits of effective mefloquine chemoprophylaxis need to be balanced with any potential, but as yet not recognized, adverse effects of mefloquine in pregnancy. Sulfadoxine should be avoided before delivery because of the risk of neonatal jaundice. Pyrimethamine, which is teratogenic in animals because it interferes with folate metabolism, is generally avoided but probably could be used. Doxycycline should not be used during pregnancy because of the effects of tetracyclines on the fetus, which include dental discoloration and dysplasia and inhibition of bone growth. To avoid the risk of inducing hemolytic anemia in utero in a G6PD-deficient fetus, primaquine should not be taken during pregnancy. The safety of malarone in pregnancy has not been established.

Travel-Related Illness

In a study of more than 10,000 Swiss who had traveled in developing countries for less than 3 months, 15% experienced health problems, and 3% were unable to work for an average of 15 days.[34] Infections with the greatest incidence per month abroad included giardiasis (seven cases per 1,000 months abroad), amebiasis (four cases per 1,000), hepatitis (four cases per 1,000), and gonorrhea (three cases per 1,000). Malaria, syphilis, and helminthic infections occurred at a lower incidence (fewer than one case per 1,000). No cases of typhoid fever or cholera were reported. The most common modes of acquisition of infection were enteral and sexual. Travelers should be cautioned about sexual contacts, especially in areas where hepatitis B or HIV is prevalent, and be advised to use condoms and barrier protection during sexual encounters.

Because of the global prevalence of HIV, postexposure antiretroviral prophylaxis may be germane for travelers who may have occupational exposures (e.g., health care workers) and for students and workers who are traveling and may be at risk for HIV exposure. The availability of local postexposure prophylactic medications should be ascertained at overseas work or study sites. Options for two- or three-drug regimens of postexposure antiretroviral therapy are discussed elsewhere [*see 7:XXXIII HIV and AIDS*]. If selected antiretroviral therapy is not assuredly available at work or study sites, sufficient medication should be carried by the traveler to ensure that a 28-day course of antiretroviral therapy is available.

Stays at major resorts and first-class hotels are associated with less risk than stays in less frequented locales or rural dwellings or encampments. In areas where sanitation and personal hygiene may be poor, it is prudent to be careful of food and water, although such care does not necessarily diminish the risk of diarrheal disease. Fruit that is peeled by the traveler is safe, whereas vegetables may be contaminated with fecally passed organisms in the soil and should not be consumed raw. Unpasteurized dairy products should be avoided, as should inadequately cooked fish or meat. If water is of uncertain quality, travelers should avoid drinking it or using ice made from it. Boiling will render water safe. Chlorination will kill most bacterial and viral pathogens, but protozoal cysts of *Giardia lamblia* and *Entamoeba histolytica* may survive. Carbonated beverages, beer, wine, and drinks made from boiled water are safe.

In areas where schistosomiasis is prevalent, swimming in freshwater should be avoided, although swimming in chlorinated or saltwater is safe. Even short exposures to infested water during rafting or swimming have caused the onset of acute schistosomiasis.

Most infections acquired during travels will present within weeks of travel, but some may not manifest themselves until much later; hence, knowledge of a patient's travel history is important.

ALTITUDE ILLNESS

Altitude illnesses may develop in travelers who arrive at heights between 6,000 and 8,000 ft (1,829 and 2,438 m) above sea level.[1] Travelers may arrive at these altitudes rapidly by flying into an airport at these elevations or more slowly by driving or climbing. Altitude illness includes three syndromes: acute mountain sickness (AMS), high-altitude pulmonary edema (HAPE), and high-altitude cerebral edema (HACE). AMS, the most common form of altitude illness, may occur at altitudes between 4,000 and 6,000 ft. Symptoms include headache, fatigue, loss of appetite, nausea, and, sometimes, vomiting. AMS usually develops 6 to 12 hours after arrival at the higher altitude. HACE is a progression of AMS characterized by extreme lethargy, confusion, and an ataxic gait during a tandem gait test.

HAPE may develop alone or in conjunction with HACE. Symptoms include increasing breathlessness. HAPE is more likely than HACE to be fatal. Travelers who develop HACE or HAPE must immediately descend to lower altitudes. Travelers to elevated altitudes need to be cautioned about the symptoms of these syndromes [*see 14:X Pulmonary Edema*], advised about the gravity of HACE and HAPE, and admonished not to delay descent to lower altitudes if these potentially lethal syndromes develop.

Three medications can be used to prevent and treat altitude illnesses. Acetazolamide can prevent AMS if taken before ascent; it also can hasten recovery. Dosing is 125 mg every 12 hours beginning the day of ascent. Dexamethasone (4 mg every 6 hours) can be used to prevent and treat AMS and HACE. Some investigators recommend relying on acetazolamide for prophylaxis and reserving dexamethasone for treatment of symptoms.[1] Persons who have experienced HAPE are at increased risk of its recurrence. If travel to high altitudes is unavoidable, nifedipine (10 to 20 mg every 8 hours) can prevent and ameliorate HAPE in those prone to experience this syndrome.

TRAVELER'S DIARRHEA

Diarrhea is the most common illness of travelers.[35] Infectious agents, primarily bacterial but also viral and parasitic pathogens, are responsible for traveler's diarrhea. Over 75% of cases of traveler's diarrhea are caused by bacteria, with enterotoxigenic *Escherichia coli* being the most frequent cause. Other common bacterial causes of traveler's diarrhea include *Shigella* species, *Campylobacter jejuni*, *Aeromonas* species, *Plesiomonas shigelloides*, *Salmonella* species, and noncholera *Vibrio* species.[35] Rotavirus and Norwalk agent are the most common viral causes; *Giardia, Cryptosporidium, Cyclospora,* and, less commonly, *Dientamoeba fragilis, Isospora belli, Balantidium coli, Strongyloides stercoralis,* and *E. histolytica* are parasitic causes.

In addition to exercising caution about food and water,[36] travelers may take either of two approaches: chemoprophylaxis and postonset treatment.

Chemoprophylaxis

The benefits of chemoprophylaxis may be offset by the risks of taking chemoprophylactic agents. Side effects of short-term prophylactic doses of bismuth subsalicylate may include tinnitus, blackening of the stool and tongue, and impaired absorption of doxycycline, which is important if doxycycline is used as daily antimalarial chemoprophylaxis. Side effects of antibiotics may include skin rashes and vaginal candidiasis, photosensitivity skin eruptions (especially with doxycycline), and, in rare instances, potentially life-threatening bone marrow suppression, mucocutaneous reactions, or anaphylaxis. Although these potential side effects temper the routine use of chemoprophylaxis, specific needs or wishes of travelers may dictate its use. Patients with underlying medical conditions that may be aggravated by a serious diarrheal illness, including active inflammatory bowel disease, type 1 (insulin-dependent) diabetes mellitus, and heart disease in the elderly, as well as patients whose activities during travel cannot tolerate interruption by an episode of diarrheal illness, should consider chemoprophylaxis. Several regimens are available [*see Table 3*]. Bismuth subsalicylate, which should not

Table 3 Chemoprophylaxis and Treatment of Traveler's Diarrhea

Drug	*Dose*
Prophylaxis	
Bismuth subsalicylate	Two 262 mg tablets chewed q.i.d. with meals and at bedtime
Quinolone antibiotics	
Norfloxacin	400 mg/day
Ciprofloxacin	500 mg/day
Ofloxacin	300 mg/day
Levofloxacin	500 mg/day
Doxycycline	100 mg/day
Treatment	
Loperamide	4 mg loading dose, then 2 mg after each loose stool, to a maximum of 16 mg/day
Quinolone antibiotics	
Norfloxacin	400 mg b.i.d. for up to 3 days
Ciprofloxacin	500 mg b.i.d. for up to 3 days
Ofloxacin	300 mg b.i.d. for up to 3 days
Levofloxacin	500 mg/day for up to 3 days
Azithromycin	1,000 mg single dose or 500 mg/day for 3 days
Rifaximin	200 mg t.i.d. for 3 days

be taken by persons with peptic ulcer disease, coagulopathies, or allergies to salicylates, is not as completely effective as quinolone antibiotics but has fewer side effects and enables the use of quinolone antibiotics, if they are needed for therapy. Resistance among bacterial causes of traveler's diarrhea is not common at present for the quinolone antibiotics (except for quinolone-resistant *Campylobacter* infection prevalent in Thailand) but is quite common for trimethoprim-sulfamethoxazole and doxycycline, limiting their efficacy. Chemoprophylactic medications should be started on the first day of arrival and continued for 1 to 2 days after returning home but not for more than 3 weeks.

Postonset Treatment

A generally preferable alternative to chemoprophylaxis is early therapy for traveler's diarrhea [*see Table 3*]. Because of the likelihood of bacterial resistance, trimethoprim-sulfamethoxazole is less effective than regimens employing quinolone antibiotics. Antibiotics will shorten the duration of traveler's diarrhea to a range of 16 to 30 hours, compared with a range of 59 to 93 hours in those not receiving antibiotics. The use of loperamide, which diminishes intestinal motility and fluid and electrolyte losses, together with antibiotics can further abbreviate symptoms. In a study of patients with dysentery caused by *Shigella* or enteroinvasive *E. coli*, the use of loperamide with ciprofloxacin, in comparison with ciprofloxacin alone, led to briefer (median, 19 hours versus 42 hours) and milder (median, two stools versus 6.5 stools) diarrheal illness, without untoward effects.[37] Loperamide has not been studied in children, and adults with prolonged fever or bloody stools should be advised to cease loperamide use and seek medical attention. Azithromycin is an alternative to quinolone antibiotics that can be used by pregnant patients; it is the agent of choice where quinolone-resistant *Campylobacter* infection is prevalent. Rifaximin (Xifaxan, from Salix), a nonabsorbable agent, is approved by the FDA for traveler's diarrhea caused by noninvasive strains of *Escherichia coli*.[38] Rifaximin should not be used if dysentery is suspected (i.e., if symptoms include fever and bloody stools) or if other causes of diarrhea (e.g., *Campylobacter*, *Shigella*, or invasive *E. coli*) are possible or isolated.

For any diarrheal illness, maintenance of hydration is of cardinal importance and can often be achieved by oral replacement of lost fluid and electrolytes. Convenient and inexpensive packets of oral rehydration salts formulated according to World Health Organization recommendations (i.e., 3.5 g of sodium chloride, 1.5 g of potassium chloride, 20 g of glucose, and 2.9 g of trisodium citrate in each packet) are available in both developed and developing countries. Each packet of oral rehydration salts is added to a liter of boiled or treated water and should be consumed or discarded within 12 hours if kept at ambient temperature or within 24 hours if kept refrigerated.

MEDICAL ISSUES DURING TRANSIT

Cruise ships that dock at ports in the United States are inspected for sanitation by officials from the CDC. Inspections are aimed at minimizing the potential for outbreaks of gastrointestinal disease on board. Travelers may obtain information on whether specific cruise ships meet sanitation standards from travel agents, state health departments, or the CDC.[1] Outbreaks of influenza have occurred aboard cruise ships in the past 10 years in various regions, including Alaska and the Yukon Territory. Travelers older than 50 years should consider influenza vaccination.

Because jet aircraft are not pressurized to sea level, passengers will be exposed to high-altitude environments. The atmospheric pressure maintained within the cabin of an airplane flying at 27,000 to 42,000 ft is equivalent to the pressure at an altitude of 3,000 to 8,000 ft, so that at a cruising altitude of 35,000 ft, the cabin pressure is about 600 mm Hg. Because of the decreased pressure, the arterial oxygen tension (P_aO_2) of normal persons will fall to about 68 mm Hg. In patients with chronic obstructive lung disease, the P_aO_2 will fall even lower. However, despite a fall in P_aO_2, patients may not show symptoms of hypoxia. Although hypoxia occurs in pregnant women, jet air travel has no deleterious effects on them or their fetuses. It is difficult to establish precise criteria for the use of supplemental oxygen for air travelers. Caution is indicated, however, for patients with impaired cardiopulmonary function: supplemental oxygen may be administered during flights at altitudes higher than 22,500 ft.

Scuba divers should wait 12 to 48 hours, depending on the length of their diving exposures, before boarding a commercial aircraft. This measure is important for avoiding the occurrence of aeroembolism, commonly known as the bends, which could develop in an underpressurized cabin if nitrogen gas dissolved in the person's fat cells is mobilized.

In patients with upper respiratory tract infections, differential air pressures between blocked eustachian tubes or sinuses and the cabin may develop on ascent or descent and impair hearing or cause pain in the ears or sinuses; symptoms can be relieved by the use of decongestants. Persons prone to motion sickness should take a prophylactic medication. Prolonged immobilization during flight may cause venous thrombosis in persons with preexisting thrombotic or venous disease [*see 1:XVIII Venous Thromboembolism*]. The exact risks and rates for developing venous thromboembolism during air travel are not yet defined.[39] Leg exercise and walking during the flight and use of below-the-knee stockings have been suggested to be beneficial, but evidence is lacking.

The author has received grant or research support from and is an advisor to GlaxoSmithKline.

References

1. Centers for Disease Control and Prevention: Health information for international travel 2003-2004. US Department of Health and Human Services, Atlanta, 2003 (http://www.cdc.gov)

2. Rachel Barwick: History of thymoma and yellow fever vaccination (correspondence). Eidex for the Yellow Fever Vaccine Safety Working Group. Lancet 364:936, 2004

3. Kelso JM, Mootrey GT, Tsai TF: Anaphylaxis from yellow fever vaccine. J Allergy Clin Immunol 103:698, 1999

4. Prevention and control of influenza. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 53(RR-6):1, 2004 (http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5306a1.htm)

5. Khan WA, Bennish ML, Seas C, et al: Randomised controlled comparison of single-dose ciprofloxacin and doxycycline for cholera caused by *Vibrio cholerae* 01 or 0139. Lancet 348:296, 1996

6. Poliomyelitis prevention in the United States. Updated recommendation of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 49(RR-5):1, 2000 (http://www.cdc.gov/mmwr/PDF/RR/RR4905.pdf)

7. Maple PA, Jones CS, Wall EC, et al: Immunity to diphtheria and tetanus in England and Wales. Vaccine 19:167, 2000

8. Leder K, Weller PF, Wilson ME: Travel vaccines and elderly persons: review of vaccines available in the United States. Clin Infect Dis 33:1553, 2001

9. Measles immunization in HIV-infected children. American Academy of Pediatrics. Committee on Infectious Diseases and Committee on Pediatric AIDS. Pediatrics 103:1057, 1999

10. Mermin JH, Townes JM, Gerber M, et al: Typhoid fever in the United States, 1985-1995: changing risks of international travel and antimicrobial resistance. Arch Intern Med 158:633, 1998

11. Steinberg EB, Bishop R, Haber P, et al: Typhoid fever in travelers: who should be targeted for prevention? Clin Infect Dis 39:186, 2004

12. Stubi CL, Landry PR, Petignat C, et al: Compliance to live oral Ty21a typhoid vaccine, and its effect on viability. J Travel Med 7:133, 2000

13. Plotkin SA, Bouveret-Le Cam N: A new typhoid vaccine composed of the Vi capsular polysaccharide. Arch Intern Med 155:2293, 1995

14. Steffen R, Kane MA, Shapiro CN, et al: Epidemiology and prevention of hepatitis A in travelers. JAMA 272:885, 1994

15. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 48(RR-12):1, 1999

16. Plesner A, Ronne T, Wachmann H: Case-control study of allergic reactions to Japanese encephalitis vaccine. Vaccine 18:1830, 2000

17. Fradin MS, Day JF: Comparative efficacy of insect repellents against mosquito bites. N Engl J Med 347:13, 2002

18. Insect repellents. Med Lett Drugs Ther 24:41, 2003

19. D'Allessandro V: Insecticide treated bed nets to prevent malaria. BMJ 322:249, 2001

20. Shah S, Filler S, Causer LM, et al: Malaria surveillance—United States, 2002. MMWR Surveill Summ 53(1):21, 2004

21. Steffen R, Fuchs E, Schildknecht J, et al: Mefloquine compared with other chemoprophylactic regimens in tourists visiting East Africa. Lancet 341:1299, 1993

22. Jaspers CA, Hopperus Buma AP, van Thiel PP, et al: Tolerance of mefloquine chemoprophylaxis in Dutch military personnel. Am J Trop Med Hyg 55:230, 1996

23. Davis TM, Dembo LG, Kaye-Eddie SA, et al: Neurological, cardiovascular and metabolic effects of mefloquine in healthy volunteers: a double-blind, placebo-controlled trial. Br J Clin Pharmacol 42:415, 1996

24. Croft AM, Clayton TC, World MJ: Side effects of mefloquine prophylaxis for malaria: an independent randomized controlled trial. Trans R Soc Trop Med Hyg 91:199, 1997

25. Luxemburger C, Price RN, Nosten F, et al: Mefloquine in infants and young children. Ann Trop Paediatr 16:281, 1996

26. Smoak BL, Writer JV, Keep LW, et al: The effects of inadvertent exposure of mefloquine chemoprophylaxis on pregnancy outcomes and infants of US Army servicewomen. J Infect Dis 176:831, 1997

27. Steketee RW, Wirima JJ, Slutsker L, et al: Malaria treatment and prevention in pregnancy: indications for use and adverse events associated with use of chloroquine or mefloquine. Am J Trop Med Hyg 55(suppl 1):50, 1996

28. Schlagenhauf P: Mefloquine for malaria chemoprophylaxis 1992-1998: a review. J Travel Med 6:122, 1999

29. Schlagenhauf P, Lobel H, Steffen R, et al: Tolerance of mefloquine by SwissAir trainee pilots. Am J Trop Med Hyg 56:235, 1997

30. Bradley DJ, Bannister B: Guidelines for malaria prevention in travelers from the United Kingdom for 2003. Health Protection Agency Advisory Committee on Malaria Prevention for UK Travellers. Commun Dis Public Health 6:180, 2003

31. Schwartz E, Rgev-Yochay G: Primaquine as prophylaxis for malaria for nonimmune travelers: a comparison with mefloquine and doxycycline. Clin Infect Dis 29:1502, 1999

32. Steketee RW, Wirima JJ, Hightower AW, et al: The effect of malaria and malaria prevention in pregnancy on offspring birthweight, prematurity, and intrauterine growth retardation in rural Malawi. Am J Trop Med Hyg 55(suppl):33, 1996

33. Steketee RW, Wirima JJ, Slutsker L, et al: Malaria parasite infection during pregnancy and at delivery in mother, placenta, and newborn: efficacy of chloroquine and mefloquine in rural Malawi. Am J Trop Med Hyg 55(suppl 1):24, 1996

34. Steffen R, Rickenbach M, Wilhelm U, et al: Health problems after travel to developing countries. J Infect Dis 156:84, 1987

35. Ryan ET, Kain KC: Health advice and immunizations for travelers. N Engl J Med 342:1716, 2000

36. Herwaldt BL, de Arroyave KR, Roberts JM, et al: A multiyear prospective study of the risk factors for and incidence of diarrheal illness in a cohort of Peace Corps volunteers in Guatemala. Ann Intern Med 132:982, 2000

37. Murphy GS, Bodhidatta L, Echeverria P, et al: Ciprofloxacin and loperamide in the treatment of bacillary dysentery. Ann Intern Med 118:582, 1993

38. Infante RM, Ericsson CD, Jiang ZD, et al: Enteroaggregative Escherichia coli diarrhea in travelers: response to rifaximin therapy. Clin Gastroenterol Hepatol 2:135, 2004

39. Mendis S, Yach D, Alwan A: Air travel and venous thromboembolism. Bull World Health Organ 80:403, 2002

Acknowledgment

Figures 1 through 5 Tom Moore.

VIII QUANTITATIVE ASPECTS OF CLINICAL DECISION MAKING

Brian Haynes, m.d., ph.d.
Harold C. Sox, m.d.

An increasing amount of very useful quantitative evidence from health care research is available to practitioners. New research findings continually expand the knowledge base of what does more good than harm for patients, and institutional forces, both professional and financial, are accelerating the adoption of research findings. More and more information is available on issues related to such important clinical topics as screening and diagnostic tests, preventive and therapeutic interventions, prognosis and clinical prediction, risk of adverse outcomes, improvement in quality of care, and cost-effectiveness of tests and treatments. Clinical application of this evidence has lagged, however, for a number of reasons.[1] First, evidence from research is often not definitive or covers only some aspects of practice. Second, clinicians are often slow to adopt research findings, even those that are well validated. Third, resources may be inadequate or too poorly organized in the local setting to permit implementation. Fourth, clinicians may be unfamiliar with the concepts that lie behind the application of quantitative reasoning to clinical care. This chapter addresses the last of these barriers: principles and methods for quantitative reasoning.

Lack of precision in clinical thinking is beginning to yield to several encouraging developments—in particular, clinicians increasingly applying principles of critical appraisal to evidence in the medical literature; formulation of methods for medical decision analysis; increasing clinical comfort with terms such as sensitivity, specificity, likelihood ratio, number needed to treat, and confidence interval (CI); and creation of print and electronic resources that minimize the effort that clinicians must make to find and interpret valid quantitative evidence when it is needed. These developments notwithstanding, the possibility of miscommunication is still considerable. A 2003 study of primary care physicians reported that just over 50% of respondents were able to answer questions about critical appraisal of methods and interpretation of results of studies focusing on treatments and diagnostic tests.[2] Patients are entitled to expect clearer thinking from their physicians, especially because many patients have difficulty themselves interpreting information about risks, benefits, and prognoses provided by their doctors.[3] Moreover, the current health care environment increasingly demands that physicians be able to justify clinical policies and decisions with an evidence-based, quantitative approach.

We have two principal goals in this chapter. The first is to provide a basic explanation of the measurements used in critical appraisal of the literature and the ways in which physicians interpret these measurements in evidence-based clinical decision making. With the advent of electronic access to MEDLINE and its clinical subsets, specialized compendia of studies (e.g., Clinical Evidence[4] and Physicians' Information and Education Resource [PIER][5]), systematic reviews of studies (e.g., the Cochrane Library[6]), and alerting services for new, clinically relevant evidence (e.g., bmjupdates+[7] and MEDSCAPE Best Evidence alerts[8]), the current best evidence for clinical practice is becoming more and more accessible to clinicians.

The second goal is to introduce the topic of medical decision analysis. Clinicians use decision analysis in two ways. One way is essentially indirect: reliance on products of decision analyses conducted by others. For example, practice guidelines increasingly influence many of the quick, straightforward decisions that occur in daily practice. Many of these guidelines are based on formal decision analyses. The second way of employing decision analysis is more direct: using the tools of decision analysis to assist in making major decisions about the care of an individual patient. Although few physicians spend the hours required to conduct a formal decision analysis from scratch, some tools of decision analysis (e.g., likelihood ratios of test results) are easy to apply; moreover, some decision analyses are accessible on a desktop or palmtop computer and only require the clinician to enter the clinical findings required by the decision tree.

It is important to understand the intent of this effort to achieve precision and quantitation in measurement and decision making: to enhance the quality of care by making it more tailored to the individual patient. Anything that can be measured, even if only qualitatively, can be counted and turned into a clinically useful quantitative measure. For example, a study might classify clinical outcomes only qualitatively (e.g., as satisfactory or unsatisfactory), but if the numbers of participants in the study who fall into one or the other of the two outcome states are counted, the result then becomes quantitative. If physicians can define individual states and measure them quantitatively (e.g., by using a continuous scale to assess functional status), they can describe individual patient status more precisely and therefore can make finer distinctions between groups of patients. By placing patients in distinctive groups, physicians can achieve one of the great goals of patient care: to inform patients of the choices between alternative treatments by the known predictors of response to those treatments.

What is the role of the individual practitioner in retrieving and evaluating evidence from research and incorporating it into individual clinical decisions? The answer to this innocuous question distills the angst of contemporary health care. In some settings, the practitioner has the freedom to act as circumstances dictate, whereas in others (e.g., certain managed care settings), someone else tries to dictate how to translate research results into patient care. We believe that practitioners cede their responsibility for clinical decision making to others at great risk to their patients and themselves, because any clinical decisions must take into account not only the evidence available and the guidelines in force but also the patient's unique circumstances and individual wishes. In today's world, the freedom to determine the content of one's practice is increasingly precious. To use this freedom responsibly, practitioners must have ready access to information that is based on current best evidence, must understand the basic principles of quantitative decision making and decision analysis, must be able to determine whether others have applied these principles appropriately in published works or in practice,

Table 1 Abbreviated Users' Guides for Appraisal of Medical Journal Articles

Purpose of Study	*Source of Data*	*Method of Arriving at Findings*	*Method of Reducing Bias of Findings*
Diagnosis	Clearly identified comparison groups, all suspected of having the disorder, but one of which is free of the disorder	Objective or reproducible diagnostic standard applied to all participants	Blinded assessment of test and diagnostic standard
Therapy	Random allocation of patients to comparison groups	Outcome measure of known or probable clinical importance	Follow-up of ≥ 80% of subjects
Prognosis	Inception cohort, early in the course of the disorder and initially free of the outcome of interest	Objective or reproducible assessment of clinically important outcomes	Follow-up of ≥ 80% of subjects
Causation	Clearly identified comparison group for those who are at risk for, or for those having, the outcome of interest	Blinding of observers of outcome to exposure; blinding of observers of exposure to outcome	—
Review	Comprehensive search for relevant articles	Explicit criteria for rating relevance and merit of studies	Inclusion of all relevant studies

and must be able to understand how to use evidence from research to make decisions in clinical practice.

How to Critically Evaluate Research Reports

To use numbers wisely in making decisions about patients, the physician must have some way of determining whether the numbers are derived from sound research. Detailed users' guides for interpreting the medical literature are available[9]; in an effort to simplify this issue, we have provided an abbreviated set of such guides [*see Table 1*].[10] Physicians may find these guides especially useful when reading research reports in the primary literature. However, when physicians are not getting and interpreting evidence themselves, they should look to evidence-based publications, such as *Clinical Evidence* and *PIER*; systematic review articles, such as those from the Cochrane Collaboration and clinical journals; and practice guidelines that use explicit criteria for evaluating evidence [*see Table 1*].

How to Apply Research Results to Patient Care

Once a physician is satisfied that the quantitative results from the relevant research were derived through sound methods, he or she can interpret them in light of the patient's circumstances and use them to help determine the best way to proceed with management. The interpretation of research results takes five main forms: (1) measures of disease frequency, (2) measures of diagnostic certainty, (3) measures of diagnostic test performance and interpretation, (4) measures of the effects of treatment, and (5) measures of treatment outcomes adjusted for quality of life.

Measures of Disease Frequency

Clinically useful measures of disease frequency include incidence, prevalence, the case-fatality rate, the *P* value, and the CI [*see Table 2*]. The use of such terms is illustrated in more detail elsewhere (see below).

Measures of Diagnostic Certainty: Use of Probabilities

When asked how sure they are of their diagnoses, most physicians express their degree of certainty in words rather than numbers. A classic study illustrates the difficulty of this approach.[11] The authors examined pathology and radiology reports and recorded various terms expressive of the probability of a disorder, such as "compatible with," "consistent with," "likely," "probably," and "pathognomonic." They then asked a group of clinicians to assign numerical probabilities to all of these terms. For each term (even "pathognomonic"), the range of probabilities stretched over half the scale. For example, to one physician, "likely" meant there was a 45% chance that the disease in question was present, whereas to another, "likely" meant the probability was higher than 90%. When diagnostic-test specialists were asked on two different occasions what they meant by these terms, the earlier and later answers were highly consistent for each individual specialist but highly inconsistent from one specialist to the next.

An alternative to using words to express the degree of diagnostic certainty is to use a number—namely, the probability that the diagnosis is present. A probability is a number between 0 and 1 that expresses the likelihood that an event will occur; 0 represents certainty that it will not occur, and 1 represents certainty that it will. Using probability to express diagnostic certainty has two key advantages. First, it facilitates precise communication. Comparison of probability estimates is a far more precise method of comparing degrees of diagnostic certainty than ex-

Table 2 Clinically Useful Measures of Disease Frequency

- Incidence: the proportion of new cases of a disorder occurring in a defined population during a specified period of time, typically 1 year.
- Prevalence: the proportion of cases of a disorder at a designated point in time in a specified population.
- Case-fatality rate: the proportion of cases of a specified disorder that are fatal during a specified period of follow-up (typically 1 yr) from the onset of the disorder.
- Quality-adjusted life year (QALY): a measure of survival in which each year of a patient's survival is discounted according to a measure (usually an index) of the patient's quality of life.
- *P* value: the probability of obtaining the observed data, or more unlikely data, when the null hypothesis is true. The *P* value does not indicate the magnitude of the effect of interest, or even its direction, nor does it indicate how much uncertainty is associated with the results.
- Confidence interval (CI): the range of values of a true effect that is consistent with the data observed in a study. A common (although not entirely correct) interpretation of a 95% confidence interval is that 95% of the time, the true value lies within the stated range of values.

changing verbal assessments. Second, there exists an accurate method of calculating changes in the likelihood of disease as new information (e.g., a test result) becomes available. This method, Bayes' theorem, should be one of the central principles that underlie medical practice. This claim may seem audacious to some readers, but we all recognize that the interpretation of new information about the patient moves us either away from or closer to a diagnosis and, therefore, away from or closer to the decision to use a specific treatment.

The probability of an event is not precisely the same thing as the odds of an event occurring, even though the two are mathematically equivalent ways of expressing diagnostic uncertainty. Habitués of the racetrack are reputed to use odds directly, but most clinicians are likely to find probabilities easier to use. Each of these measures can be readily converted to the other, as follows:

$$\text{Odds} = \frac{\text{probability}}{1 - \text{probability}}$$

$$\text{Probability} = \frac{\text{odds}}{1 + \text{odds}}$$

To use a test result quantitatively, a physician must first estimate the pretest probability of the disease. Unaided, physicians are not particularly good at this task. In a 1982 study, when primary care physicians were given clinical scenarios and asked for their estimates of the probabilities of given disorders, they provided estimates—quite confident ones—but their estimates did not agree with those of their fellow clinicians.[12] Indeed, when individual physicians were tested subsequently with the same scenarios, their later estimates did not agree with their initial ones.

How does a physician estimate the probability that a patient's chief complaint is a manifestation of a particular disease? The first step is to take a careful history and do a physical examination. From this point, the physician may take any of three basic approaches to estimating the probability of a disease[13]: (1) subjective estimation, (2) estimation based on the prevalence of disease in other patients with the same syndrome, or (3) application of clinical prediction rules.

Subjective Estimation

In principle, the physician can draw on personal experience with similar patients and use the estimated frequency of the disease in those patients. In practice, this approach is little more than a semiquantitative guess and is prone to error because of defective recall, as well as to bias in the application of the heuristics (i.e., the rules of thumb) for estimating probability. Examples of such heuristics are representativeness, by which one estimates a probability on the basis of the similarity of the patient's signs and symptoms to the features of the classic description of the disease, and availability, by which one estimates a probability partly on the basis of how easy it is to recall similar cases. One very useful heuristic is anchoring and adjustment, by which one establishes an initial estimate (e.g., the prevalence of pulmonary embolism in 100 patients presenting to the emergency department with pleuritic chest pain) and then adjusts the estimate upward or downward by taking into account the patient's findings (e.g., hypoxemia, unilateral leg swelling, or a history of cancer). Physicians can, in principle, calculate the extent of such adjustments by using Bayes' theorem (see below).

Estimation Based on the Prevalence of Disease in Other Patients with the Same Syndrome

One antidote to the failures of subjective probability estimation is to base the estimate on accurate diagnoses established in a series of patients with the same clinical syndrome as the patient under consideration. The best example is the diagnosis of suspected coronary artery disease in patients with chronic chest pain. On the basis of the clinical history, the physician can place the patient in one of three categories: typical angina pectoris, atypical angina, or nonanginal chest pain. Many published studies have reported the frequency of angiographically proven coronary disease in patients with these syndromes. These studies have shown, for example, that in a man with atypical angina, the probability of significant coronary artery disease is approximately 0.70 (see below).

Application of Clinical Prediction Rules

Clinical prediction rules describe the key clinical findings that predict a disease and show how to use these findings to estimate the probability of disease in a patient. Such rules are based on analysis of a standardized set of data, including clinical findings and the final diagnosis, for each of many patients with a diagnostic problem. One type of clinical prediction rule uses regression analysis to identify the best clinical predictors and their diagnostic weights. The sum of the diagnostic weights corresponding to a patient's findings is a score, and the probability of disease for each patient is equivalent to the prevalence of disease among patients with a similar score. A well-known example of this approach is the rule for estimating the probability of cardiac complications from noncardiac surgery.[14] Another interesting example showed that the prevalence of coronary artery disease in patients with similar chest pain scores varied systematically according to the overall prevalence of coronary artery disease in several study populations.[15] This study suggested that the probability of disease corresponding to a patient's clinical history varies depending on whether the setting of care is a primary care practice or a referral practice. *Diagnostic Strategies for Common Medical Problems*[16] is an excellent source of pretest probabilities, as is *Evidence-Based Physical Diagnosis*.[17]

MEASURES OF DIAGNOSTIC TEST PERFORMANCE AND INTERPRETATION

Clinically useful measures of diagnostic test performance include sensitivity, specificity, and the likelihood ratio; clinically useful measures of test interpretation include pretest odds, pretest probability, probability after a positive test result, and probability after a negative test result [*see Table 3*]. Physicians should memorize and internalize the definitions of these terms to avoid becoming muddled when attempting to use information from diagnostic tests in decision making.

In the past, articles usually described the performance of a diagnostic test only in terms of sensitivity and specificity. These familiar terms do not directly describe the effect of a test result on the probability of disease. To correct this shortcoming, many articles now use the likelihood ratio (LR), which is the amount by which the odds of a disease change with new information. This value is calculated as follows:

$$\text{LR} = \frac{P\,[\text{test result if disease present}]}{P\,[\text{test result if disease absent}]}$$

Because physicians often express test results as either positive or negative, there is a likelihood ratio for a positive test re-

Table 3 Definitions of Clinically Useful Measures of Diagnostic Test Performance and Interpretation

The typical approach to evaluation of most diagnostic tests, particularly those with so-called binary outcomes (e.g., a positive or a negative test result, with no other categories), makes use of a 2 × 2 table, as follows:

Diagnostic Test Result	*Presence or Absence of Disease on a Reference Test (Gold Standard)*		*No. of Patients with Given Test Result*
	Present	*Absent*	
Positive	*a*	*b*	*a* + *b*
Negative	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	

Measures of diagnostic test performance, defined below, are calculated from this table.

- Sensitivity: the proportion of people with a disease of interest who are detected by a diagnostic test; calculated as $a/(a+c)$.
- Specificity: the proportion of people who do not have a disease who are correctly identified by a negative result on a diagnostic test; calculated as $d/(b+d)$.
- Likelihood ratio: the amount by which the odds of having a disease change after a test result; calculated as $[a/(a+c)]/[b/(b+d)]$ for a positive test result and as $[c/(a+c)]/[d/(b+d)]$ for a negative test result.
- Pretest probability: the proportion of people with the disorder of interest in a group suspected of having the disorder; calculated as $(a+c)/(a+b+c+d)$.
- Odds: calculated as probability/(1 – probability).
- Probability: calculated as odds/(1+ odds).
- Posttest odds: calculated as pretest odds × likelihood ratio.
- Probability after a positive test: the proportion of people with a positive test result who have the disease of interest; calculated as $a/(a+b)$.
- Probability after a negative test: the proportion of people with a negative test result who have the disease of interest; calculated as $c/(c+d)$.

sult (LR^+) and a likelihood ratio for a negative test result (LR^-). The formula for the likelihood ratio for a positive test result is as follows:

$$LR^+ = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

The formula for the LR for a negative test is as follows:

$$LR^- = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

The likelihood ratio is generally a better descriptor than sensitivity or specificity because it more directly describes the effect of a test result on the odds of disease. The probability after obtaining new information is an application of Bayes' theorem. The most useful form of Bayes' theorem for this purpose is the odds ratio format:

$$\text{Posttest odds} = \text{pretest odds} \times \text{likelihood ratio}$$

This form of Bayes' theorem illustrates a very powerful concept that clinicians often overlook: new information has meaning only in context. Operationally, the statement means that a physician should never interpret a test result in isolation but should always take into account the individual patient's pretest probability. Simply stated, the posttest probability after a positive test result will be greater if the pretest index of suspicion was high than if the pretest index of suspicion was low. The most important practical application of this reasoning is to be suspicious when a test result is negative in a patient whose clinical findings strongly point toward a disease—that is, the probability of the disease may still be high, even after the negative test. One should also be suspicious when a test is positive in a patient for whom the likelihood of disease is very low.

The evaluation of suspected pulmonary embolism (PE) is a good example of the practical use of these statistical terms and methods. A 37-year-old woman presents to the emergency department (ED) with pleuritic chest pain and new dyspnea. She has a low-grade fever and has no cough or hemoptysis, but the ED physician believes it necessary to rule out PE. The patient has none of the other known risk factors for PE (e.g., recent surgery or prolonged bed rest, previous deep vein thrombosis [DVT], coagulopathy, malignancy, pregnancy, and use of oral contraceptives), and physical examination reveals no evidence of DVT. The arterial oxygen tension (P_aO_2) is 92 mm Hg on room air. The patient is quite distressed. The ED physician orders a chest x-ray and a helical CT scan. The CT scan is interpreted as negative for PE. The resident wishes to explain this result to the patient and then to take the appropriate next steps.

A useful flowchart for working up patients with suspected PE is provided elsewhere [*see 1:XVIII Venous Thromboembolism*]; however, this chart provides no guidance on how to estimate the clinical probability of PE. It is instructive to examine how the results of a quantitative, evidence-based approach to this patient's case relate to the recommendations outlined in the flowchart.

The initial step is to estimate the pretest probability of PE by one of two approaches. The first is to use the anchoring and adjustment heuristic. The anchor, or starting point, is the prevalence of PE in adults who present to the ED with pleuritic chest pain. One very careful study found that 21% of such patients (36/173) had a positive pulmonary angiogram.[18] The physician should use this 21% initial probability as the starting point (the anchor) for the patient under discussion and adjust it on the basis of the history and the physical examination. As noted, this patient has no predisposing factors for PE and no evidence of DVT, and her P_aO_2 is greater than 90 mm Hg. Using this approach, the ED physician concludes that the probability of PE before helical CT is quite low, perhaps 10%.[19]

The second approach is to use a clinical prediction rule.[20] This model places patients into three categories on the basis of clinical findings (typical for PE, atypical for PE, severe PE), the likelihood of alternative diagnoses, and the presence of risk factors for DVT. The prevalence rates of PE in the three categories are 3.4%, 27.8%, and 78.4%, respectively. The algorithm for placing patients into one of the three categories is somewhat complex but is easy to use when represented on the screen of a palmtop computer. Assuming that the ED physician did not identify an alternative diagnosis that seemed more likely than PE, the patient's pretest probability of PE was 28%, considerably higher than the ED physician's subjective probability.

With an estimate for the pretest probability of PE, the next step is to obtain the likelihood ratio for a negative helical CT scan. The sensitivity and specificity of the helical CT scan have varied considerably among studies. A recent meta-analysis of studies of diagnostic tests for pulmonary embolism found the likelihood ratio for a positive chest CT scan to be 24.1 (95% CI,

12.4 to 46.7). The likelihood ratio for a negative scan was 0.04 (95% CI, 0.03 to 0.06).[21]

To calculate the posttest odds of PE, the ED physician must combine the patient's pretest odds with the test's likelihood ratio by means of the odds ratio format of Bayes' theorem mentioned earlier (posttest odds = pretest odds × likelihood ratio). An alternative to converting the pretest probability to odds and doing the calculation of posttest odds is to use a nomogram [*see Figure 1*]. To estimate posttest probability, anchor a straightedge at a pretest probability of 28% (corresponding to the clinical predictive rule's estimate of pretest probability) in the left-hand column; then pass the straightedge through a likelihood ratio for a negative helical CT scan, 0.04, in the middle column. Read the posttest probability from the right-hand column: about 1.5%. The math for this estimate is as follows, with the 0.28 pretest probability of PE first needing to be converted to pretest odds:

$$\begin{aligned}\text{Pretest odds} &= \text{pretest probability}/(1 - \text{pretest probability})\\ &= 0.28/(1 - 0.28)\\ &= 0.28/0.72\\ &= 0.39\end{aligned}$$

Now, the post–helical CT scan odds of PE for this patient must be determined by multiplying the pretest odds of PE, 0.39, by the likelihood ratio for a negative helical CT, 0.04:

$$\begin{aligned}\text{Posttest odds} &= \text{pretest odds} \times \text{likelihood ratio}\\ &= 0.39 \times 0.04\\ &= 0.0156\end{aligned}$$

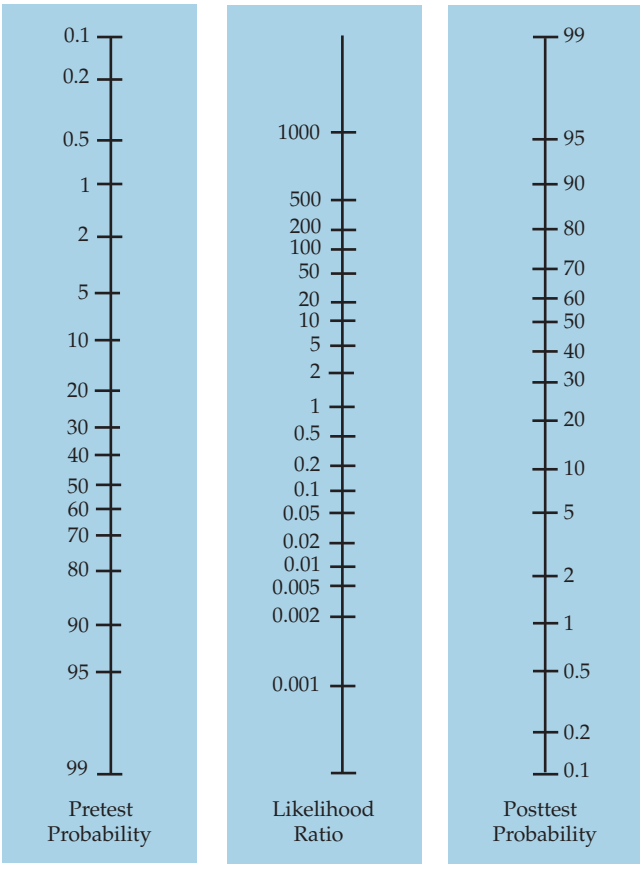


Figure 1 **Nomogram for converting pretest probabilities to posttest probabilities when test results are presented as likelihood ratios.**

Table 4 Definitions of Clinically Useful Measures of Treatment Effects from Clinical Trials

Like evaluation of diagnostic tests, evaluation of treatment effects often makes use of a 2 × 2 table, as follows:

Treatment Group	*Treatment Outcome*		*No. of Patients in Treatment Group*
	Bad	*Good*	
Experimental	*a*	*b*	*a* + *b*
Control	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	

Measures of treatment effects when treatment reduces the risk of bad outcomes are calculated from this table.

- Experimental event rate (EER): the rate of an adverse clinical outcome in the experimental group; calculated as $a/(a+b)$.
- Control event rate (CER): the rate of an adverse clinical outcome in the control group; calculated as $c/(c+d)$.
- Absolute risk reduction (ARR): the absolute arithmetic difference in outcome rates between control and experimental groups in a trial; calculated as CER – EER, or $[c/(c+d)] - [a/(a+b)]$.
- Relative risk reduction (RRR): the proportional reduction in the rate of an adverse clinical outcome in the experimental group in comparison with the control group in a trial; calculated as ARR/CER, or (CER – EER)/CER, or $\{[c/(c+d)] - [a/(a+b)]\}/[c/(c+d)]$.
- Number needed to treat (NNT): the number of patients to whom one would have to give the experimental treatment to prevent one adverse clinical outcome; calculated as 1/ARR, or $1/\{[c/(c+d)] - [a/(a+b)]\}$, and reported as a whole number rounded to the next highest integer.
- Odds ratio: the odds that an experimental patient will experience an adverse event relative to the odds that a control subject will experience such an event; calculated as $(a/b)/(c/d)$.

Convert the posttest odds to the posttest probability, as follows:

$$\begin{aligned}\text{Posttest probability} &= \text{posttest odds}/(1 + \text{posttest odds})\\ &= 0.0156/(1 + 0.0156)\\ &= 0.0154\end{aligned}$$

At a posttest probability of PE of 1.5%, only 15 patients per 1,000 would have PE. Anticoagulating patients with a 1.54% probability of PE would mean exposing 65 patients (i.e., 1/0.0154) to the harms of anticoagulation to benefit one patient with a PE. Most physicians would follow this patient closely without giving specific treatment for PE. This same logic can be applied to all screening and diagnostic tests for PE, including D-dimer testing (high sensitivity and low specificity), which is therefore more useful for ruling out PE (when it is negative) than ruling it in (when it is positive).[22] D-dimer tests can also be used for calibrating clinical observations to enhance the quantitation of pretest probabilities.[23]

MEASURES OF TREATMENT EFFECTS

One of the most important tasks of clinicians is to advise patients about the current best treatment for their condition. Such advice should be based on the best evidence available. Clinically useful measures of treatment effects reported in clinical trials include the experimental event rate (EER), the control event rate (CER), relative risk reduction (RRR), absolute risk reduction (ARR), the number needed to treat (NNT), and the number needed to harm [*see Table 4*]. These measures can be effective tools for quantifying the magnitude of treatment benefits and

risks, provided that there is a statistically significant difference in the clinical event rate between experimental subjects and control subjects (i.e., between the EER and the CER).

Again, we illustrate the practical application of these terms by a specific example. A 69-year-old hypertensive male smoker has experienced a partial left hemispheric stroke, with good recovery of function. He has a 75% ipsilateral internal carotid artery stenosis. One option would be to give this patient aspirin or clopidogrel and manage his risk factors for cerebrovascular disease; another would be to offer him carotid endarterectomy in addition to medical treatment. The question is, how and on what evidentiary basis does the clinician choose one treatment over another? It is tempting to think of treatments in black-and-white terms, as either working or not working, but the reality is rarely so absolute; often, the choice is between two or more treatments, each of which works after a fashion in certain situations. To apply the available evidence to the decision-making process in the most effective manner, the clinician must interpret it quantitatively, offering accurate, relevant figures instead of gut feelings when the patient asks what his chances are with each therapeutic approach.

Three randomized, controlled trials of carotid endarterectomy for symptomatic carotid artery stenosis[24-26] can inform our choice of treatment in this hypothetical patient. Examination of the North American Symptomatic Carotid Endarterectomy Trial (NASCET)[24] in the light of the users' guides discussed earlier [*see Table 1*] reveals that this study meets the three criteria for a study focusing on therapy. First, patients with symptomatic hemispheric transient ischemic attacks or partial strokes and ipsilateral carotid stenoses of 70% to 99% were randomly assigned to either an experimental group that underwent carotid endarterectomy or a control group that did not. All patients received continuing medical care, with special attention given to risk factors for cerebrovascular disease. Second, the study assessed the effect of carotid endarterectomy on important clinical events—namely, recurrence of stroke or perioperative stroke or death. Third, none of the patients were lost to follow-up. Consequently, the data from the study are likely to be valid guides in determining which treatment is best for this patient.

In the NASCET report, the risk of major or fatal ipsilateral stroke within a 2-year follow-up period was 2.5% in the group that underwent carotid endarterectomy and 13.1% in the control group. The absolute risk reduction, therefore, was 13.1% – 2.5%, or 10.6% ($P < 0.001$; CI, 5.5% to 15.7%), and the relative risk reduction was 10.6%/13.1%, or 81%. The number needed to treat was 10 (1/0.106); that is, 10 patients (CI, 7 to 18) would have to be treated with carotid endarterectomy (rather than medical treatment alone) to ensure that one major or fatal ipsilateral stroke would be prevented. The NASCET report indicates that this benefit is somewhat lower for patients with less severe stenosis (70% to 79%) and somewhat higher for patients with multiple risk factors for cerebrovascular disease—circumstances that offset one another in the case of the patient under consideration here.

Having determined the NNT, the next question is whether an NNT of 10 for major or fatal stroke over a 2-year period is a small benefit or a large one. By contrast, treatment of elevated diastolic blood pressures that do not exceed 115 mm Hg is associated with an NNT of 167 to prevent one stroke over a 5-year period.[25] Thus, for patients who have symptomatic, severe carotid artery stenosis, carotid endarterectomy is highly beneficial.

Given this conclusion, the next question is, do these research results apply to a specific patient, hospital, and surgeon? For example, the NASCET data reflect operative procedures performed by highly competent surgeons in specialized centers. One would have to know the perioperative complication rates for local surgeons to be able to assess a patient's level of risk if referred to any of those surgeons. If the local surgeons' perioperative complication rates for carotid endarterectomy are lower than 7%, the results are comparable to the NASCET results. On the other hand, patients with a stenosis of less than 70% are at substantially less risk for subsequent stroke to begin with. Potential benefit is similar to potential harm for patients with stenoses of 50% to 70%; for patients with stenoses of less than 50%, current evidence indicates that carotid endarterectomy would not yield any net reduction of this risk, even when the procedure is done by a highly skilled surgeon.[26,27]

MEASURES OF TREATMENT OUTCOME, ADJUSTED FOR QUALITY OF LIFE

Measures of treatment outcome, such as reduction in mortality, are important in deciding whether to start a medication or perform an operation, but they do not answer a question that is important to many patients: How much longer can they expect to live if treatment is started? One way of responding is to frame the answer in terms of life expectancy, the average length of life after starting treatment, which has a simple relation to the annual mortality in patients undergoing treatment.[13]

Although life expectancy is a useful measure of treatment outcome, it has one shortcoming: it places the same value on years in perfect health as on years in poor health. Arguably, a year with partially treated chronic disease is not equivalent to a year in perfect health. A solution to this problem is to adjust life expectancy for the quality of life that the patient experiences during a year of poor health by multiplying life expectancy by a number, expressed on a scale of 0 to 1, that reflects how the patient feels about the quality of life experienced during an illness. This number is usually called a utility. When life expectancy, expressed in years, is multiplied by a utility, the result is a quality-adjusted life year (QALY). One QALY is equivalent to a year in perfect health.

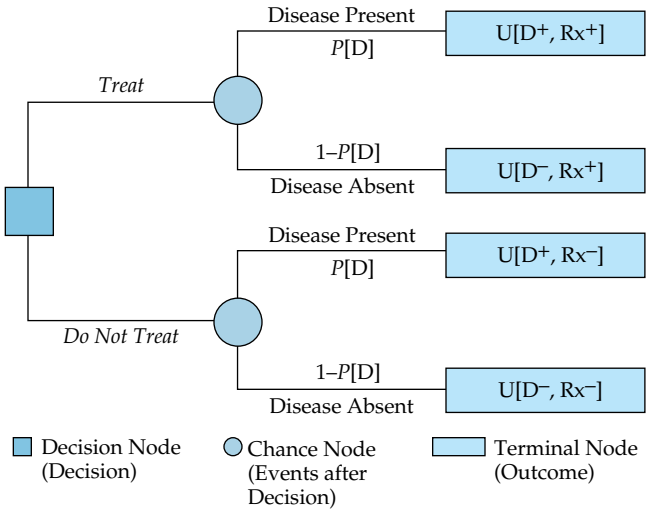


Figure 2 **Shown is a decision tree for calculating the treatment threshold probability in a patient who is a possible candidate for carotid endarterectomy. (D—disease; U—utility)**

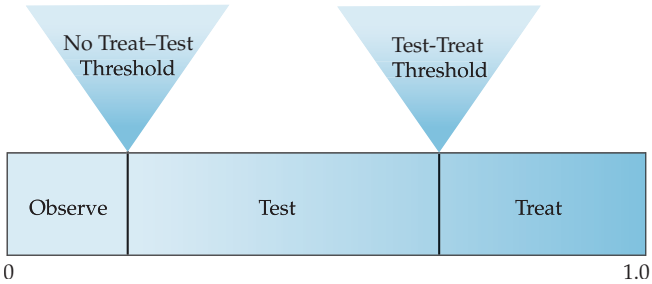


Figure 3 **Probability scale showing the ranges of probability corresponding to different actions following the initial history and physical examination.**

Medical Decision Analysis

Clearly, there is more to clinical decision making than simply collecting numbers that measure treatment effects. Reports of treatment effects in randomized, controlled trials are important starting points that help determine whether a treatment has merit in its own right, but the actual decision whether to offer a given patient a particular treatment is complex and must take into account each patient's specific clinical circumstances and individual wishes. For example, if the patient has significant comorbidity that would result in an especially high risk of perioperative complications, surgical therapy might not be the best choice. Even if the patient is well enough to undergo operation, individual preferences and values must be taken into account: the patient might be strongly averse to the immediate risks posed by surgery or might lack the resources to pay for the procedure.

THE THRESHOLD MODEL OF DECISION MAKING

At the conclusion of every history and physical examination, the clinician must choose one of three options: to treat, to observe, or to obtain more information. The optimal approach to making this choice starts with the assumption that the physician will seek more information (i.e., order diagnostic tests) only if the results may alter the treatment decision. Although occasional exceptions are easily justified, this rule is a good guiding principle for a lean style of practice. It is also the central assumption behind the threshold model of decision making.

When a diagnosis is uncertain, the decision whether to start treatment depends on the probability of the diagnosis. If the probability is 0, no one would start treatment; if the probability is 1, everyone would start treatment. Therefore, there must be a probability between 0 and 1 at which a physician would have no preference between treating and not treating. This probability is called the treatment threshold probability.

The treatment threshold probability is a key to solving the important decision-making problem of whether to treat, to observe, or to obtain more information. The most elegant way of obtaining the treatment threshold probability is to construct a decision tree that represents the choice between starting treatment and withholding treatment [*see Figure 2*]. In a decision tree, decisions are represented by squares (decision nodes), and the chance events that follow a decision are represented by circles (chance nodes). The probabilities of the events after a chance node must total 1.0. A terminal node (represented by a rectangle enclosing the name of the state) represents a state in which there are no subsequent chance events. Each terminal node has a value, which is a measure of the outcome associated with the event.

In a decision tree for starting or withholding treatment, each branch of the two chance nodes ends in a terminal node whose value is the utility (U) for being in the state specified. For example, $U[D^+, Rx^+]$ is the utility for having the disease (D) and being treated for it, which one could calculate by representing that state as a tree with chance nodes and terminal nodes. To obtain the treatment threshold probability, one sets the expected utility of treatment at a value equal to the value for the expected utility of no treatment and then solves for the probability of disease. The general solution to the equation is as follows:

$$\text{Treatment threshold probability} = \frac{\text{harm}}{\text{harm} + \text{benefit}}$$

where harm is the net utility of being treated when disease is absent ($U[D^-, Rx^+] - U[D^-, Rx^-]$) and benefit is the net utility of being treated when disease is present ($U[D^+, Rx^+] - U[D^+, Rx^-]$). This relationship between harms and benefits of treatment is fundamental to solving the common decision problem of deciding about treatment when the diagnosis is not known with certainty. Because the treatment threshold depends on the benefits and harms of the treatment, it will vary from treatment to treatment. When the benefits of a treatment exceed harms, which is usually the case, the treatment threshold probability must be less than 0.50.

To make the choice between treating, not treating, and ordering tests to obtain additional information, the physician needs to know the range of probabilities of disease within which testing is the preferred action. The probability scale can be divided into three ranges [*see Figure 3*], one of which is the test range. The first step in defining the test range is to establish the treatment threshold probability. For the next step, we must invoke the principle that the physician should seek more information only if the results might alter the treatment decision. Translated to the threshold model, this principle takes the following form: testing is indicated only if the result of the test might move the probability of disease from one side of the treatment threshold (the do-not-treat side) to the other (the treat side). A physician can use this principle to decide whether to obtain a test in an individual patient. If the patient's pretest probability is below the treatment threshold and therefore in the do-not-treat zone, the physician should order the test only if the posttest probability of disease after a positive test result would be higher than the treatment threshold probability.

To obtain the test range, we must extend this example to a more general solution, which is to use the test's likelihood ratio and Bayes' theorem to calculate the pretest probability at which the posttest probability is exactly equal to the treatment threshold probability [*see Figure 3*]. This probability is called the no treat–test threshold probability. Clearly, if the pretest probability is lower than the no treat–test threshold probability, the test should not be done, because the posttest probability will be lower than the treatment threshold probability (i.e., a positive result will not change the management decision); conversely, if the pretest probability is higher than the no treat–test threshold probability, the test should be done, because the posttest probability will be higher than the treatment threshold probability (i.e., a positive test result would change the management decision from do not treat to treat).

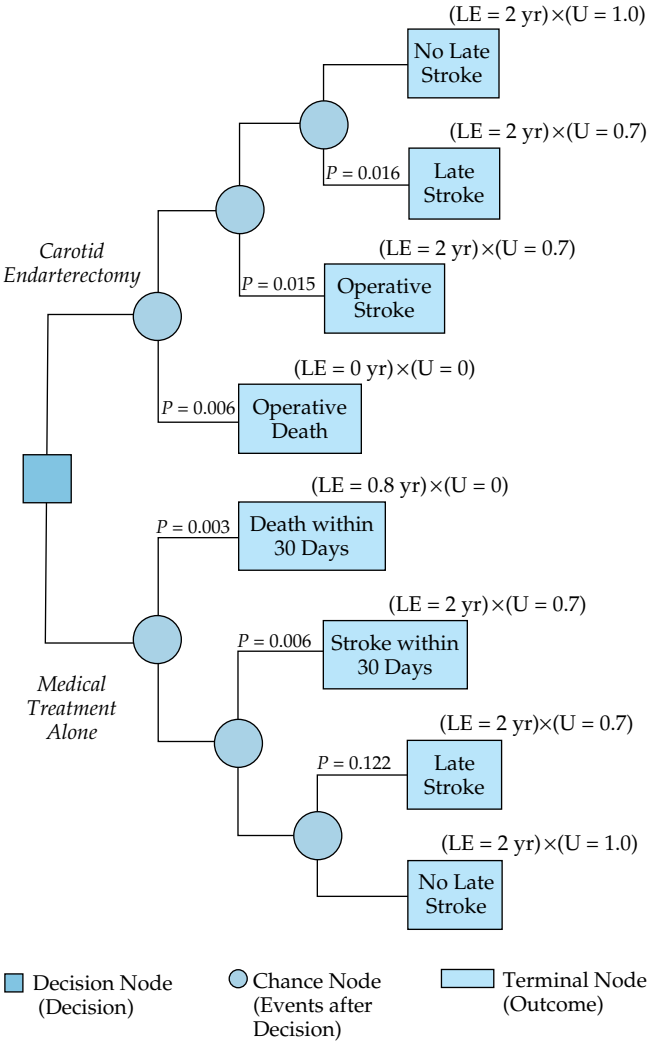


Figure 4 **Shown is a decision tree depicting the application of expected-outcome decision analysis to the same patient referred to in Figure 2. (LE—life expectancy; U—utility)**

The size of the test range depends on the likelihood ratios reported for the test. If LR^- is close to zero and LR^+ is much greater than 1.0, the test range will be very wide. In general, the better the test, the larger the test range. If the posttest probability falls within the treat zone, the physician must then decide which treatment to offer. The choice among treatments offers a good opportunity to explore the principles of decision making under conditions of uncertainty.

MEASURES OF EXPECTED-OUTCOME DECISION MAKING: THE TREATMENT DECISION

The purpose of decision analysis is to help with those decisions for which the outcome cannot be foretold (e.g., the decision whether to treat carotid artery stenosis surgically). Even when randomized trial results indicate that one treatment generally gives better results than another, some degree of uncertainty remains: individual patients may still exhibit idiosyncratic outcomes or may experience unusual but serious side effects of treatment. Faced with this uncertainty, most physicians choose the treatment that gives the best results averaged over a large number of patients. In so doing, they become, perhaps unwittingly, what are known as "expected-value decision makers." Expected value is the value of an intervention when the outcomes of that intervention are averaged over many patients. A more general term might be "expected-outcome decision maker," which would denote a physician who chooses the treatment that gives the best outcome when averaged over many patients. This concept is the basis of expected-outcome decision analysis, which is a method of framing a decision problem in terms of the expected outcome of each decision alternative. Thus, in a patient with stable angina, the physician would decide between medical management, coronary angioplasty, and coronary artery bypass surgery by first calculating a patient's life expectancy, expressed in years in good health, after undergoing each of these treatment options; then, the physicial would choose the treatment with the highest life expectancy.

We can illustrate the application of expected-outcome decision making by returning to the example of the 69-year-old man who has recovered from a hemispheric stroke and has a 75% carotid stenosis. The question to be answered is the same: Should the patient be offered carotid endarterectomy in addition to best medical treatment? The first step is to represent the problem by a decision tree [*see Figure 4*]. Each of the terminal nodes in this decision tree is associated with a life expectancy, as well as a utility representing the value of life in the outcome state, represented by the terminal node. As noted earlier [*see* How to Apply Research Results to Patient Care, Measures of Treatment Outcome Adjusted for Quality of Life, *above*], life expectancy by itself is not a sufficiently precise measure: clearly, 2 years of life after a major stroke is not equivalent to 2 years in perfect health. The decision maker needs a quantitative measure of the patient's feelings about being in an outcome state. The physician can obtain the patient's utility for that state by asking the patient to indicate the length of time in perfect health that he would consider equivalent to his life expectancy in a disabled state (e.g., after a major stroke). This technique is called time trade-off. Other techniques used to obtain this utility include linear scaling and the standard reference gamble.[13]

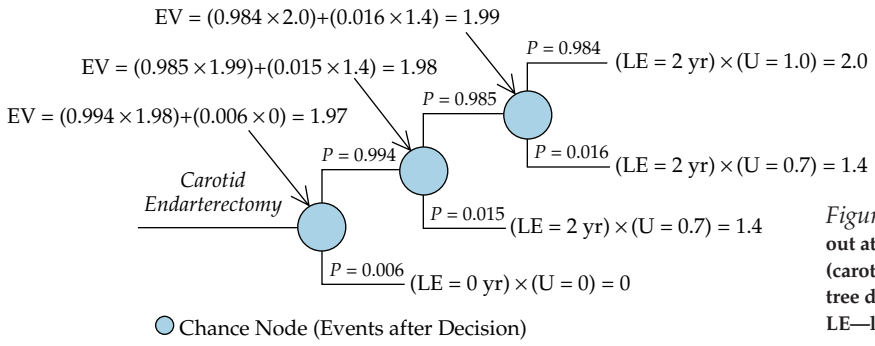


Figure 5 **Illustrated is the process of averaging out at a chance node, as applied to the upper (carotid endarterectomy) portion of the decision tree depicted in Figure 2. (EV—expected value; LE—life expectancy; U—utility)**

To calculate the expected value of surgical management, the decision maker starts at the chance nodes that are farthest from the decision node (the tips of the branches of the decision tree), multiplies the probability of each event at each chance node by the value of the event, and sums these products over all the events at the chance node. This calculation is known as averaging out at a chance node [*see Figure 5*]. The value obtained for each chance node by means of this process becomes the outcome measure for the next step, which is to repeat the averaging-out process at the next chance node to the left.

With either therapeutic option—aspirin combined with carotid endarterectomy or continued management with aspirin alone—there is a chance of death within 30 days, stroke within 30 days, or stroke within 2 years [*see Figure 4*]. As noted [*see* How to Apply Research Results to Patient Care, Measures of Treatment Effects, *above*], reliable data on the probabilities of these adverse events are available in the NASCET report.[24] To simplify the presentation of the decision analysis, we measure survival only within the 2-year time frame addressed in the NASCET report, and we assume that all late strokes occur at the start of this 2-year period. Further, we assume that a patient would value 2 years of disability resulting from a stroke as equivalent to 17.5 months of healthy life, which means that the utility representing the state of having experienced a major stroke is 0.70.

The decision analysis indicates that the decision maker should prefer surgical treatment to medical treatment. The expected value of carotid endarterectomy for this patient is 1.96 QALY, whereas the expected value of medical treatment is 1.91 QALY. Admittedly, this difference is not very large, indicating a close call, and it is reasonable to ask how high the operative mortality would have to be to make medical treatment the favored approach. Sensitivity analysis, one of the most powerful features of decision analysis, shows that the operative mortality would have to increase considerably before medical treatment would become preferable. The baseline figure for operative mortality in the NASCET report was 0.6%. The sensitivity analysis indicates that medical treatment would have a higher expected value than surgical treatment only if the operative mortality were 3.2% or higher, which might be the case if considerable comorbidity were present or if the surgeon seldom performed carotid endarterectomy. Although most physicians would not have the time or expertise to carry out this decision analysis, storing the appropriate decision tree in a palmtop computer would make it possible to do the decision analysis easily in the office setting, using values specific to the clinical setting and the patient.

cost-effectiveness analysis

Cost-effectiveness analysis is a method for comparing the impact of expenditures on different health outcomes. Cost-effectiveness analysis assesses the trade-off between added benefit and added cost by examining costs and benefits at the margin (i.e., comparing one intervention with another or with no intervention). The cost-effectiveness of one intervention (A) versus another (B) is calculated as follows:

$$\text{Cost-effectiveness (A vs. B)} = \frac{\text{cost A} - \text{cost B}}{\text{effectiveness A} - \text{effectiveness B}}$$

In the carotid endarterectomy example, the costs would include all costs associated with a subsequent stroke. If we assume that the average lifetime cost associated with carotid endarterectomy is $10,000 and the average lifetime cost associated with medical treatment is $8,000, then the cost-effectiveness of surgical treatment, as compared with medical treatment, would be calculated as follows:

Cost-effectiveness (surgery vs. no surgery)
= ($10,000 – $8,000)/(1.96 –1.91 QALY)
= $2,000/0.05 QALY
= $40,000/QALY

One may then ask, is a treatment choice that costs $40,000 for each extra QALY cost-effective? There is no absolute answer to this question. In practice, a physician compares the cost-effectiveness of carotid endarterectomy with that of other interventions. How this information should affect the decision whether to offer surgical treatment to any given patient is an even more difficult question. Indeed, most experts would say that cost-effectiveness is a technique for deciding policies that would apply to many patients. An organization with limited resources would choose policies that prescribe interventions with the lowest cost per added QALY. The organization would not offer interventions that have a high cost relative to the magnitude of the anticipated benefit.

Conclusion

Quantitative approaches to clinical reasoning are still evolving. By combining better evidence from health care research with today's burgeoning information technology, physicians can apply evidence effectively to individual patient care. As requirements for efficiency and accountability continue to increase, physicians are under more and more pressure to adopt a quantitative, evidence-based approach to patient care. Physicians who can back up their decisions with sound research and sound reasoning will be in a better position to provide their patients with optimal care.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Antman EM, Lau J, Kupelnick B, et al: A comparison of results of meta-analyses of randomized control trials and recommendations of experts. JAMA 268:240, 1992

2. Godwin M, Seguin R: Critical appraisal skills of family physicians in Ontario, Canada. BMC Med Educ, 2003
http://www.biomedcentral.com/1472-6920/3/10

3. Lobb EA, Butow PN, Kenny DT, et al: Communicating prognosis in early breast cancer: do women understand the language used? Med J Aust 171:290, 1999

4. Clinical Evidence: A Compendium of the Best Available Evidence for Effective Health Care (Online). BMJ Publishing Group, Serial Electronic Publication, London, 1999
http://www.clinicalevidence.org

5. Physicians' Information and Education Resource (Online). American College of Physicians, Philadelphia, 2005
http://pier.acponline.org/info/?hp

6. The Cochrane Library (Online). Serial Electronic Publication, Oxford
http://www.update-software.com/cochrane

7. BMJupdates^{+}(Online). BMJ Publishing Group and McMaster University Health Information Research Center, Hamilton, Ontario, Canada, 2005
http://www.bmjupdates.com

8. MEDSCAPE Best Evidence alerts (Online). WebMD, New York, 2005
https://profreg.medscape.com/px/newsletter.do

9. Sackett DL, Straus S, Richardson SR, et al: Evidence-Based Medicine: How to Practice and Teach EBM, 2nd ed. Churchill Livingstone, London, 2000

10. Haynes RB, Sackett DL, Cook DJ, et al: Transferring evidence from health care research into medical practice: 2. Getting the evidence straight. ACP J Club 126:A14, 1997

11. Bryant G, Norman G: Expression of probability: words and numbers (letter). N Engl J Med 302:411, 1980

12. Feightner JW, Norman GR, Haynes RB: The reliability of likelihood estimates by physicians. Clin Res 30:298A, 1982

13. Sox HC, Blatt M, Marton KI, et al: Medical Decision Making. Butterworths, Stoneham, Massachusetts, 1988

14. Goldman L, Caldera DL, Nussbaum SR, et al: Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 297:845, 1977

15. Sox HC, Hickam DH, Marton KI, et al: Using the patient's history to estimate the probability of coronary artery disease: a comparison of referral and primary care practice. Am J Med 89:7, 1990

16. Black E, Bordley DR, Tape TG, et al: Diagnostic Strategies for Common Medical Problems, 2nd ed. ACP Publications, Philadelphia, 1999

17. McGee S: Evidence-Based Physical Diagnosis. WB Saunders Co, Philadelphia, 2001

18. Hull RD, Raskob GE, Carter CJ, et al: Pulmonary embolism in outpatients with pleuritic chest pain. Arch Intern Med 148:838, 1988

19. Mayewski RJ: Respiratory problems: pulmonary embolism. Diagnostic Strategies for Common Medical Problems. Panzer R, Black E, Griner P, Eds. ACP Library on Disk (CD-ROM publication). American College of Physicians, Philadelphia, 1996

20. Wells PS, Ginsberg JS, Anderson DR, et al: Use of a clinical model for safe management of patients with suspected pulmonary embolism. Ann Intern Med 129:997, 1998

21. Roy PM, Colombet I, Durieux P, et al: Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. BMJ 331:259, 2005

22. Stein PD, Hull RD, Patel KC, et al: D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. Ann Intern Med 140:589, 2004

23. Chunilal SD, Eikelboom JW, Attia J, et al: Does this patient have pulmonary embolism? JAMA 290:2849, 2003

24. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. North American Symptomatic Carotid Endarterectomy Trial collaborators. N Engl J Med 325:445, 1991

25. Cook RJ, Sackett DL: The number needed to treat: a clinically useful measure of treatment effect. BMJ 310:452, 1995

26. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. Carotid Endarterectomy Trialists' Collaboration. Lancet 361:107, 2003

27. Cina CS, Clase CM, Haynes RB: Carotid endarterectomy for symptomatic carotid stenosis. Cochrane Database Syst Rev (2):CD001081, 2000

IX PALLIATIVE MEDICINE

Cynthia X. Pan, m.d.

One unanticipated result of the advances in health care during the past century has been the emergence of chronic illness as the leading cause of death [*see Table 1*]. At the same time, the enhanced ability to significantly extend life for patients with chronic diseases has blurred the boundary between curable illnesses and illnesses that inevitably result in death. As a result, over the course of the 20th century, Western society increasingly attributed near-miraculous powers to medical science—and increasingly avoided the subject of death. Many patients and physicians came to regard the prolongation of life and the cure of disease as the fundamental and exclusive goals of modern medicine. Viewed from this perspective, death is a medical failure.

Recent decades, however, have seen a growing recognition that this view is unrealistic and potentially harmful. This recognition has supported the emergence of the field of palliative care. Unlike curative care, which focuses on the disease process, palliative care focuses on the patient, striving to minimize the patient's burden and maximize the patient's quality of life. A distinguishing feature of palliative care is that it openly acknowledges dying as part of living and does not consider death an enemy.[1]

This chapter describes the general concepts of palliative care, reviews the clinical skills needed to provide competent palliative care to patients who are chronically ill or near the end of life, and discusses some of the challenging legal and ethical issues often encountered in palliative and terminal care.

History and Rationale

Palliative medicine was first recognized as a medical specialty in Great Britain in 1987. This discipline grew out of the hospice care movement, a special interdisciplinary system of comprehensive care for the dying and for their families.[1] Over time, the palliative care model has been extended. It now applies not only to patients who are clearly at the end of life but also to those with chronic illnesses that, although not imminently fatal, cause significant impairment in function, quality of life, and independence. Palliative medicine for patients with serious illness thus should no longer be seen as the alternative to traditional life-prolonging care. Instead, it should be viewed as part of an integrated approach to medical care. Palliative care is not characterized by less care or by withdrawal of care. On the contrary, palliative care may involve intensive and highly sophisticated medical interventions, albeit ones intended to relieve suffering or improve quality of life.

Table 1 Leading Causes of Death in the United States: 2000[46]

Rank	*Condition*	*Percent of Total Deaths*
1	Heart disease	29.6
2	Malignant neoplasm	23.0
3	Cerebrovascular diseases	7.0
4	Chronic lower respiratory tract diseases	5.1
5	Accidents	4.1
6	Diabetes mellitus	2.9
7	Influenza and pneumonia	2.7

Note: These conditions account for approximately 75% of all deaths.

Settings for Delivery of Palliative Care

Palliative care may be delivered in a variety of settings, including a hospital, nursing home, hospice, or private home. In some cases, the level of care required will dictate the choice of setting. For the most part, however, palliative care depends more on the attitude of the clinician than on the setting.

HOSPITALS

Increasing numbers of hospital-based palliative care programs have been developed in recent years to meet the needs of people who are chronically and critically ill and eventually die in hospitals.[2] A national Center to Advance Palliative Care has been created to provide technical support and resources for hospitals that want to establish such programs [*see Sidebar* Palliative Care Information and Resources on the Internet].

HOSPICE

Hospice is one way to deliver palliative care [*see Table 2*]. Hospice care traditionally has been characterized as low tech, high touch. Hospice provides home nursing, support for the family, spiritual counseling, pain treatment, medications for the illness that prompted the referral, medical care, and some inpatient care. The National Hospice and Palliative Care Organization (NHPCO) estimates that in the United States, hospices admitted 775,000 patients in 2001 (compared with approximately 340,000 persons in 1994) and that, in 2000, one in four persons who died of all causes were receiving hospice care at the time of death.[3]

Palliative care and hospice share similar philosophies, and both are delivered by an interdisciplinary team of health care professionals. Palliative care differs from hospice care in that palliative care can be provided at any time during an illness and in a variety of settings, may be combined with curative treatments, and is independent of the third-party payer. In the United States, hospice is paid through the Medicare Hospice Benefit. Medicare requires that recipients spend 80% of hospice care days at home, which means that to qualify for hospice, the patient must have a home and have caregivers (e.g., family members) capable of providing care. In addition, primarily for financial reasons, Medicare requires that recipients have an estimated survival of 6 months or less and that their care be focused on comfort rather than cure.[4] These eligibility rules were created at a time when hospice programs principally served patients with cancer or AIDS, in which the trajectory of dying is relatively predictable; in 1994, for example, 80% of hospice patients had cancer, and the average patient enrolled about 1 month before death.[5] Because hospice increasingly serves pa-

Palliative Care Information and Resources on the Internet

American Academy of Hospice and Palliative Medicine (AAHPM)
http://www.aahpm.org
Organization for physicians dedicated to the advancement of hospice/palliative medicine.

Americans for Better Care of the Dying (ABCD)
http://www.abcd-caring.org
Nonprofit public charity dedicated to social, professional, and policy reform aimed to improve the care system for patients with serious illness and their families.

Center to Advance Palliative Care (CAPC)
http://www.capcmssm.org
For hospitals and health systems interested in developing palliative care programs.

Death and Dying: MEDLINEplus
http://www.nlm.nih.gov/medlineplus/deathanddying.html
Links from the U.S. National Library of Medicine and the National Institutes of Health.

Education for Physicians on End-of-life Care (EPEC)
http://www.epec.net
Provides a core curriculum for physicians on the basic knowledge and skills needed to appropriately care for dying patients.

End of Life/Palliative Education Resource Center (EPERC)
http://www.eperc.mcw.edu
Identifies and disseminates information on end-of-life care education and training materials, publications, conferences, and other resources.

End-of-life Nursing Education Consortium (ELNEC) Project
http://www.aacn.nche.edu/elnec
Provides a comprehensive national education program to develop a core of expert nursing educators and to coordinate national nursing efforts in end-of-life care.

Growth House
http://www.growthhouse.org
Gateway to international resources for life-threatening illness and end-of-life care; intended to improve the quality of care for dying people through public education and global professional collaboration; includes links and search engine on reviewed resources for end-of-life care.

National Hospice and Palliative Care Organization (NHPCO)
http://www.nhpco.org
Nonprofit membership organization representing hospice and palliative care programs and professionals in the United States. Offers information in Spanish and English on local hospice and palliative care programs across the country. Toll-free telephone number: 800-658-8898.

Project on Death in America (PDIA)
http://www.soros.org/death
Strives to increase understanding and transform the culture and experience of dying and bereavement through initiatives in research, scholarship, the humanities, and the arts, and to foster innovations in the provision of care, public education, professional education, and public policy.

tients with chronic conditions in which prognostication remains inaccurate, these eligibility rules now limit access to care.[6,7] Asking patients and families to choose between curative care and palliative care is difficult for all concerned and is inconsistent with the current model of care, which views palliative care on a continuum with life-prolonging therapy [*see Figure 1*].[8] Also, this either/or situation contributes to late referrals and underutilization of hospice services.[9]

Palliative care can be provided in nursing homes, and increasing numbers of nursing homes strive to do so. Policies governing the coordination of palliative care in nursing homes vary according to reimbursement venues and availability of trained staff. Many nursing homes coordinate palliative care through local hospices, taking advantage of the skilled hospice nurses and other health care professionals.

Demographics of Death and Dying in the United States

Most people in the United States can now expect to die in old age. Of the over 2 million deaths per year in the United States, 73% occur in persons 65 years of age or older: 49% in persons 65 to 84 years of age, and 24% in persons 85 years of age or older. In 2001, the estimated life expectancy at birth reached 77.2 years, compared with less than 50 years in 1900.[10]

The median age of death in the United States is 77 years of age; of persons who survive to 65 years of age, median age at death is 84 years for women and 80 years for men.[11] Persons 65 years of age or older constitute an increasingly large number and proportion of the United States population, and those persons 85 years of age or older constitute the most rapidly growing segment. In 1999, persons 65 years of age or older accounted for about 13% of the population; this proportion is projected to rise to 20% by the year 2030.[12]

The elderly population is extremely heterogeneous, varying in socioeconomic status, educational level, and cultural and ethnic background. This diversity is likely to increase in the coming years. For example, African Americans 65 years of age and older numbered 2.5 million in 1990 (constituting 8% of the population of persons older than 65 years), and their number is expected to more than triple, to 8.4 million (or 10.5% of that population) by 2030. Similarly, there were approximately 1.1 million Hispanic elderly persons in 1990 (3.5% of the population of persons older than 65 years), but by 2030 this number will skyrocket to 12.5 million (15.6% of that group).[12]

Compared with the current elderly population, elderly baby boomers will be far more knowledgeable about health care and far more demanding of health care providers. Their expectations are likely to lead them to challenge the health care profession to deliver high quality end-of-life care tailored to patients' individual need and to provide that care in a culturally sensitive manner.

Although most deaths occur in the elderly, people can become critically ill at any point in their lives and can die at any age. In fact, the persons whose cases were the basis for establishing important precedents for ethical and legal decisions related to death and dying were young adults: 26-year-old Nancy Cruzan,[13] whose case involved the issue of artificial feeding of patients in a persistent vegetative state; and 21-year-old Karen Ann Quinlan,[14] whose case involved the withdrawal of artificial ventilation from patients in a persistent vegetative state.

LEADING CAUSES AND SETTINGS OF DEATH

The three leading causes of death in adults in the United States in 2001 were heart disease, malignant neoplasm, and stroke.[10] Chronic obstructive pulmonary disease (COPD), pneumonia, and accidents each accounted for less than 10% of all deaths. Most adult deaths in the United States occur in hospitals (56%), followed sequentially by deaths occurring at home (21%), in nursing homes (19%), and in other settings (4%).[15] These statistics vary substantially according to geographic site, primarily because of regional variations in hospital, hospice, and nursing home bed supply.

Table 2 Comparison of Hospice and Palliative Care

Feature	*Hospice*	*Palliative Care*
Initiation	Prognosis of < 6 mo survival	Any time during illness
Clinical focus	Comfort care; no curative care	May involve both comfort care and curative care
Third-party payment	Medicare hospice benefit	Independent of payer
Personnel	Volunteers integral and required	Health care professionals
Setting	> 90% provided at home; inpatient hospice for acute deterioration, very short life expectancy (e.g., < 2 wk), or high symptom burden	Anywhere (hospital, home, nursing home)
Do not resuscitate (DNR) orders	Not required	Not required

Prognosis and Palliative Care

Traditionally, palliative care has been narrowly conceptualized as an alternative to standard life-prolonging therapy and has been provided to patients whose disease no longer responds to curative treatment. Although this model may be appropriate for patients dying of metastatic cancer, in which prognosis is relatively predictable and response to treatment is typically well defined, fewer than a quarter of persons in the United States die of cancer; the majority die of chronic diseases (e.g., heart disease) in which the prognosis is often uncertain, functional decline is nonlinear, and life-prolonging therapies coexist with or are identical to therapies directed at palliation and comfort (e.g., diuresis for fluid overload in heart failure).

One of the barriers to initiating palliative care is the uncertainty of predicting prognosis in these complex, chronic medical illnesses. For example, timing of death in heart failure is far less predictable than in many other fatal disorders. A patient dying of colon cancer usually has a long period of functional stability, then several months of progressive functional decline and weight loss just before death. In contrast, most heart failure patients experience a lengthy decline in daily function, with periodic bouts of severe symptoms and disability and multiple hospital admissions for exacerbation and for adjustment of therapy. Death may occur during a severe exacerbation but often occurs suddenly and relatively unpredictably from cardiac arrhythmia [*see Figure 2*]. In the SUPPORT project (Study to Un-

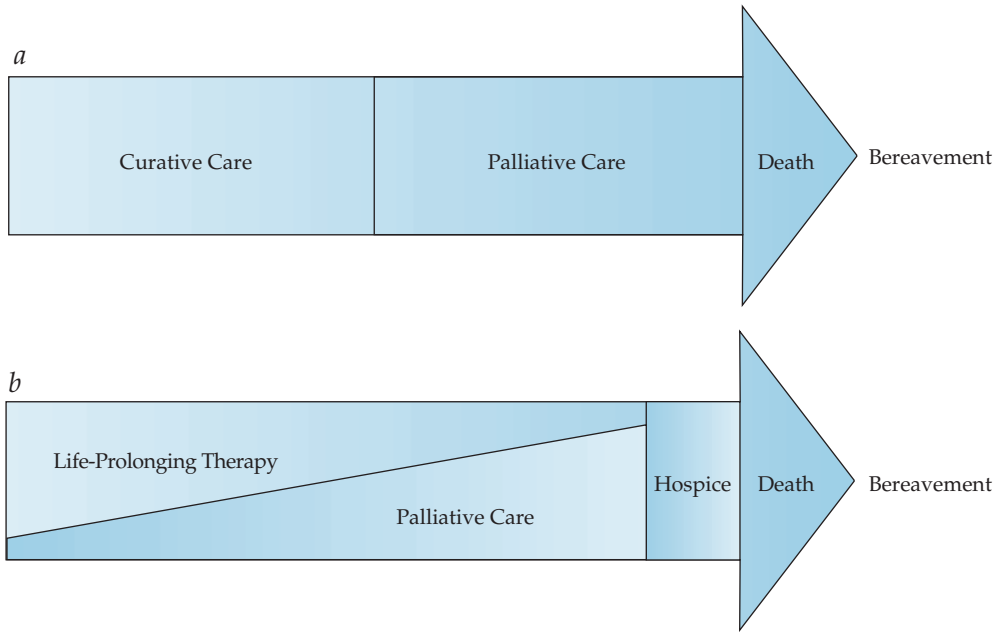


Figure 1 **(*a*) Formerly, curative care and palliative care were viewed as mutually exclusive; when death became inevitable, curative care was abandoned and palliative care begun. This model of care is now outdated, because prognosis is so difficult to determine. (*b*) The current model views palliative care on a continuum with life-prolonging therapy, with palliative care assuming increasing importance as the patient's illness progresses and curative options are exhausted. Also, many chronic and life-threatening illnesses have no cures; the goals of treatment are to contain the illness and maintain an acceptable level of function and quality of life.**

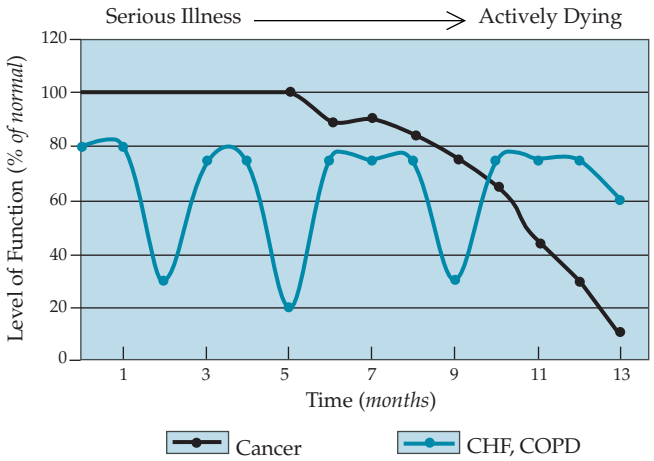


Figure 2 **Prognosis is relatively predictable in metastatic cancer; these patients typically have a long period of functional stability, then several months of progressive functional decline and weight loss just before death. In contrast, prognosis can be difficult to predict in diseases such as chronic heart failure (CHF), chronic obstructive pulmonary disease (COPD), and Alzheimer disease; these patients typically experience a lengthy decline in daily function, with periodic bouts of severe symptoms and disability and multiple hospital admissions for exacerbation and for adjustment of therapy. Death may occur during a severe exacerbation, but—especially in CHF—often occurs suddenly and relatively unpredictably from cardiac arrhythmia.**

derstand Prognoses and Preferences for Outcomes and Risks of Treatments),[16] heart failure patients were given surprisingly long prognoses even up to the day before death. The median prognosis on the day before death was a 50% chance of living 2 months.[17] Dementia is another condition that often progresses over years rather than months.

To help clinicians assess prognosis in various nonneoplastic conditions, the NHPCO has compiled guidelines describing factors associated with poor prognosis; these guidelines can promote discussion about preferences for care and advance care planning.[18] However, SUPPORT data indicate that for seriously ill, hospitalized patients with advanced COPD, heart failure, or end-stage liver disease, recommended clinical-prediction criteria cannot reliably identify those patients whose survival prognosis is 6 months or less.[19] Because it is not possible to consistently and accurately predict the timing of death, palliative care interventions should be incorporated early in a patient's course of illness, even in the face of substantial uncertainty about prognosis. As disease progresses, the goals of care should change accordingly, with the balance shifting from curative to palliative.

Clinical Skills

Caring for dying patients and for those who suffer from chronic and severe illnesses with uncertain prognoses requires an interdisciplinary approach and specific clinical skills. In particular, the clinician who provides palliative care must be competent in clinical communication, management of symptoms (physical, emotional, and psychological), and planning for continuity of care.

COMMUNICATION

The ability to communicate well with both patient and family is paramount in palliative care. In the beginning, it enables the physician to deliver bad news, assess the patient's and the family's knowledge and understanding of the disease process, determine the factors that they consider important to quality of life, and discuss goals and preferences for future care. As the illness progresses, regular communication about the course of illness and the patient's needs and expectations enables the physician to provide the most appropriate care for the patient and support for the family. Communication continues to be important after the patient's death, because the period of bereavement poses major challenges and increased risks of medical and psychiatric illness for family members.[1]

Patients whose cultural background and language differ from that of the physician present special challenges and rewards and need to be approached in a culturally sensitive manner [*see* Accounting for Cultural Differences, *below*]. Physicians also need to communicate effectively with colleagues and interdisciplinary team members to achieve optimal care for their patients. Communication in palliative care is discussed in detail elsewhere [*see CE:XI Management of Psychosocial Issues in Terminal Illness*].

SYMPTOM MANAGEMENT

Symptom management in palliative care encompasses the assessment and treatment not only of physical symptoms but of emotional and psychological symptoms as well. Physical symptoms that can contribute to discomfort, disability, and dependence include pain, dyspnea, constipation, nausea and vomiting, delirium, fatigue, and anorexia. Emotional and psychological symptoms include depression, anxiety, delirium, cognitive impairment, fear, and agitation or sedation, as well as spiritual and existential angst.

Pain is the most common symptom of terminal illness, reported by 84% of patients with cancer and 67% of patients dying of other causes.[11] Surprisingly, the leading cause of physical distress in patients dying of heart failure is also pain, followed by fatigue and shortness of breath. Other common symptoms reported by dying patients include trouble with breathing (49%), nausea and vomiting (33%), sleeplessness (40%), depression (36%), loss of appetite (47%), and constipation (36%). Apart from illness, symptoms that become more prevalent with increasing age include mental confusion, loss of bladder and bowel control, difficulty seeing and hearing, and dizziness.

At present, the identification and management of many symptoms, including pain, remain suboptimal. Undertreatment of symptoms is common in elderly patients whether they have cancer[20] or other chronic conditions, whether they reside in long-term care settings (45% to 80% prevalence)[21,22] or in the community, and whether they are white or are members of minority groups.[20,23] Undertreatment of pain in the elderly may be more common in patients who are women, are members of minority groups, or have mild to moderate cognitive impairment.[24] Clinicians may contribute to undertreatment of pain through lack of proper pain-assessment procedures, misconceptions regarding both the efficacy of nonpharmacologic pain-management strategies and the attitudes of the elderly toward such treatments,[25] and legitimate concerns about drug interactions and side effects.

Education and involvement of the patient and family as partners in care are key to the successful management of symptoms. Specific strategies for symptom management include both pharmacologic and nonpharmacologic measures. These

strategies are discussed in detail elsewhere [*see CE:IX Symptom Management in Palliative Medicine and 11:XIV Pain*].

ACCOUNTING FOR CULTURAL DIFFERENCES

The United States is a culturally heterogeneous country. Culture can broadly include race and ethnicity, as well as country of origin, religion, spirituality, and profession.[26] Medicine has its culture as well, with its attendant values, beliefs, behaviors, and language. Thus, in some cases, patients may be encountering two unfamiliar cultures: that of the United States mainstream and that of Western medicine.

Cultural traits may have a far-reaching impact on a patient's views on illness, preferences, and ultimate decisions.[27] Compared with patients from mainstream United States culture (most of whom are whites of European descent), people from other cultural backgrounds may be less willing to discuss resuscitation status,[27] less likely to forgo life-sustaining treatment,[27-29] and more reluctant to complete advance directives.[30]

Although many individual variations exist, some frequently encountered examples of cultural differences include the following:

(1) Hindus traditionally respect the doctor's medical opinion and may request the physician, rather than a family member, be appointed as health care proxy. They may prefer to die at home, preferably on the floor near the earth. After death, the relatives may want to wash the body themselves and dress it in new clothes. Autopsy is not forbidden but is considered distasteful, and cremation is usual.

(2) Traditional Chinese (and some other Asian) families usually will ask the clinician not to inform the patient about a terminal diagnosis (especially cancer) for fear that the patient will lose hope and die. In these cultures, the patient ideally will be informed after a period of adjustment. Decision-making is often entrusted to the eldest child, usually a son. Patients may seek traditional therapies, such as acupuncture and herbal medicine, often in conjunction with allopathic care.

(3) African Americans may decline participation in research studies, because of their long history of abuse as experimental subjects in research. Because of their history of receiving inappropriate undertreatment, they may continue to request aggressive care, even in terminal illness.[31]

With patients who do not speak English, it is extremely helpful to have access to a trained interpreter who can provide an objective translation and shift the translation burden from family members. This can prevent awkward and inappropriate situations, such as having to ask a male teenager to interpret for his mother who has cervical cancer. Translators may also be able to provide valuable information about patients' cultural attitudes and expectations.

Although it is important to learn about and respect different cultural practices, it is even more important not to stereotype or assume that membership in a group determines preferences. Instead, the physician should treat each patient as an individual. When in doubt, ask: "I have had patients from your cultural group who told me…. Does this apply to you?" Or, "I don't know much about medical practices or beliefs in your culture; can you tell me more about this?"

Advance Directives

Public opinion polls have revealed that close to 90% of adults in the United States would not want to be maintained on life-support systems without prospect of recovery. Yet a survey of emergency department patients found that 77% did not have advance directives, and of those patients who had one, only 5% had discussed their advance directive with their primary care physician.[32] A survey of community-dwelling older adults found that only about 16% had written advance directives.[33] In a survey of adult outpatients, most felt that discussions about advance directives should take place at an earlier age, earlier in the course of the disease, and earlier in the patient-physician relationship; most subjects also agreed that it was the physician's responsibility to initiate the discussion.[34]

Primary care physicians are in an excellent position to speak with patients about their care preferences because of the therapeutic relationship that already exists between patient and doctor. Conversations about preferences of care should be a routine aspect of care, even in healthy older patients. Determination of the patient's preferences can be made over two or three visits and then updated on a regular basis (e.g., annually). Reevaluation is indicated if the patient's condition changes acutely. In general, it is preferable that a close family member or friend accompany the patient during these discussions, so that these care preferences can be witnessed and any potential surprises or conflicts can be explored with the family.

Such discussions have particular urgency in patients who are showing early signs of cognitive impairment, because advancing impairment may render these discussions impossible. In older persons with existing cognitive impairment, it is important to assess both their current degree of decision-making capacity and any evidence, written or verbal, of previously stated preferences.

Decision-making capacity refers to the capacity to provide informed consent to treatment. This is different from competence, which is a legal term; competence is determined by a court. Any physician who has adequate training can determine capacity. It does not need to be determined by a psychiatrist. Primary care physicians often have more insight and knowledge about their patients than a psychiatrist, who might be seeing the patient for the first time.[35] The more complicated and serious the decision, the more stringent the requirements for understanding. For instance, a demented patient may have the capacity to appoint a trusted family member to serve as health care proxy but may not have the capacity to decide whether to have a permanent feeding tube placed.

A patient must meet three key criteria to demonstrate decision-making capacity: (1) the ability to understand information about diagnosis and treatment; (2) the ability to evaluate, deliberate, weigh alternatives, and compare risks and benefits; and (3) the ability to communicate a choice, whether verbally, in writing, or with a nod or gesture.

In eliciting patient preferences, the clinician should explore the patient's values—what is important to the patient and what makes life worth living or what makes life intolerable. The clinician should help the patient identify and set realistic goals, then direct treatment decisions according to these goals. More specifically, it is important to evaluate whether the patient would prefer to focus on length of life or quality of life if faced with a serious illness. In older persons who have chronic conditions that are not immediately life threatening, there is more time to explore these issues and to modify decisions over time. Outlining the available treatment options (e.g., probability and extent of response to treatment, duration and quality of extended life, anticipated side effects), identifying patients' short- and long-term

goals and needs, uncovering their expectations about therapy, evaluating their coping strategies, and identifying their support networks are critical components of this discussion.

Discussions of care preferences should cover specific life-sustaining treatments such as cardiopulmonary resuscitation (CPR), artificial nutrition and hydration, and mechanical intubation and ventilation. Physicians should review with the patient the potential indications for such therapies and, if possible, offer an explicit appraisal of the outcome in their situation. A helpful strategy is to ask patients how long they think they would want a particular treatment to be continued if it did not seem to be helping. For example, the physician might ask, "If you had a brain injury that left you in a coma and the neurologists determined that only a miracle would restore your brain function, how long would you want to stay on treatments that were keeping you alive?" Some patients may specify a week, some a month, and still others, a year. Such discussions help clarify the patient's preferences and tolerance for uncertainty.

In eliciting patient preferences for care, it is critical to consider the person's cultural, ethnic, and religious background.[26] For example, it is fairly well known that Jehovah's Witnesses will not accept blood transfusions, even in the face of life-threatening conditions, but may want all other invasive treatments. Such differences can make a patient unwilling to accept a physician's recommendations and can make a physician angry at the patient's resistance to those recommendations. With patience and training, however, it is usually possible to uncover these beliefs and negotiate treatment plans that are acceptable to all concerned.

TYPES OF ADVANCE DIRECTIVES

There are three types of advance directives: (1) do not resuscitate (DNR) orders, (2) directives involving health care proxies, and (3) living wills. All are legal instruments. The federal Patient Self-Determination Act of 1990 requires hospitals, skilled nursing facilities, home health agencies, hospice programs, and health maintenance organizations to maintain written policies and procedures guaranteeing that every adult receiving medical care be given written information concerning advance directives. Although forms for designating health care proxies and living wills are completed by patients themselves, physicians may wish to secure copies of the forms used in their state and assist their patients in completing these forms.

A health care proxy is a person appointed by the patient to make health care decisions in the patient's stead, in the event that the patient loses the capacity to make those decisions. In general, it is preferable for the physician to speak to the patient first, ask the patient to think about appointing a health care proxy, and then ask the patient to bring the potential proxy to a follow-up meeting. The proxy should be aware of and advised about the patient's goals of care and preferences and should be able and willing to assume the responsibilities of serving as proxy. Typically, an alternative proxy is also appointed.

A living will is a document that directs health care personnel to withhold or withdraw life-sustaining treatment in the event that the patient is in an incurable or irreversible condition with no reasonable expectation of recovery. Not all states have statutes recognizing living wills. However, courts have recognized and upheld the use of living wills as long as these documents provide "clear and convincing evidence" of a competent patient's wishes.

Ethical Issues in Palliative Care

Chronic illness and end-of-life care bring into focus some compelling ethical issues. These include limiting life-sustaining treatments, physician-assisted suicide, and euthanasia. Guidelines and principles on these issues have been established to enable patients, families, and physicians to reach medically sound, ethical treatment decisions in cases of irreversible illness. As a result, and despite widespread physician feeling to the contrary, these treatment decisions are almost devoid of litigation danger. Nevertheless, physicians should work with their hospital attorneys to clarify the status of legislation and case law on these issues in their particular jurisdiction.

LIMITING TREATMENT (REFOCUSING GOALS OF CARE)

In the discussion of treatment goals and plans with patients or family members, the language a physician uses can make a tremendous difference. If the physician says, "It is time to stop [or limit] the treatments," the patient or family will likely feel abandoned and hopeless and therefore ask for more interventions that may not be appropriate or useful. However, if the physician says, "It is time to refocus our efforts; let's strive to maximize comfort and dignity rather than prolong the dying process," the patient and family are more likely to feel validated and reassured. Similarly, if the physician refers to mechanical ventilation, dialysis, or artificial nutrition as "life-sustaining" treatments, it is a rare individual who will elect to forgo them. Rather, the physician should refer to them as medical interventions used to achieve specific goals. For example, one might speak of instituting mechanical ventilation to support breathing, in the hope that the patient will regain spontaneous breathing; if this hope is not realized, it is then time to discuss what the goals of care are and whether they need to be modified.

Ordinarily, discussions of goals of care involving limitation of life-sustaining treatment occur in three categories of patients. The first category includes patients whose illness is judged irreversible and who are moribund; these patients usually do not benefit from aggressive medical interventions, which can become invasive and burdensome. For patients who will die with or without treatment, such as a patient with advanced metastatic cancer or a patient with end-stage cardiomyopathy for whom a transplant is not possible, interventions often pose more burdens and risks than benefit. The second category consists of patients with capacity who are not moribund but have an irreversible illness, such as amyotrophic lateral sclerosis or multiple sclerosis. These patients often wish to discuss their ultimate goals of care and their right to refuse medical treatments so as to retain control over their health care as their disease progresses. The third category includes patients with capacity who have a reversible illness. As with any patient with capacity, the principle of autonomy guarantees these patients the right to refuse any treatment, even a lifesaving one, although physicians obviously will question these refusals much more vigorously than refusals in cases of irreversible and progressive illnesses.

An important caveat here is that although supreme autonomy of the patient is valued by mainstream culture in the United States, it is not the guiding value of many other cultures. In fact, most ethnic groups in the United States (e.g., Hispanic Americans, Asian Americans, Orthodox Jews, African Americans) favor a family-based decision-making process. Furthermore, autonomy does not always mean that the patient must be informed or must participate in decision making. Autonomy

means that patients should be asked whether they wish to be informed or participate in decision making; they may refuse to do either.

In some cases, a limited trial of life-prolonging treatment may clarify the patient's chances of recovery. The treatment can be stopped if it becomes clear that health (or the extent of recovery acceptable to the patient) cannot be restored. However, sometimes it is psychologically more difficult to stop such a treatment once it has been started, even if its original justification no longer applies.

When the patient does not have capacity, there are several ways to resolve treatment decisions. Advance directives are the most helpful. Otherwise, common sense should be followed, and the patient's next of kin should be asked to provide a substituted judgment about what the patient would have wanted or what decision would be in the patient's best interest.

Ethically and legally, there is no difference between forgoing or withholding a medical treatment (such as mechanical ventilation) and stopping or withdrawing it. However, family members and health care providers may feel that withholding and withdrawing interventions are emotionally different. It is therefore critical to counsel families and health care professionals that the decisions about any medical treatment should be guided by overall goals of care. Consultation with the hospital's palliative care service or ethics committee may be valuable for resolving conflicts over life-prolonging treatments.

FUTILE TREATMENT

Conflicts that require arbitration often center on treatments that either the family or the treatment team regard as futile or ineffective. Futility is a narrowly applied term that is used in the setting of CPR to describe a resuscitation attempt that would not succeed in resuscitating the patient or that, if successful, would likely be followed shortly afterward by another arrest. In many cases, when a patient is irreversibly ill and dying, CPR would be futile. Application of it is contrary to the standards of medical practice; it is unethical and inhumane. In such a case, the physician does not have a duty to consult anyone before writing a DNR order but should inform the family that a DNR order is being implemented. This is an opportunity for the physician to remind the family just how severe the illness is and to refocus attention to meeting other needs of the patient and the family.

Defining futility is currently a major goal of medical ethics.[36] The negative right of refusal has become transformed by some into a positive right to demand of physicians any life-sustaining treatment. Others argue that physicians have a duty not to offer or provide treatments that are ineffective.[37] Because most risk-versus-benefit considerations of life-sustaining treatment involve value judgments and because the principle of autonomy requires that the patient's values come first, some argue that so-called objective standards for futility are impossible to formulate and that physicians should make no such judgments.[38] However, for patients in an irreversible coma and, increasingly, for those in a persistent vegetative state, life-sustaining treatments are seen to be futile.

Controversial questions about defining treatments as futile will most likely be resolved city by city by a panel of experts set up to judge whether a treatment is futile after hearing all evidence presented by the family, the medical team, and others. This was the approach used by the Houston citywide consortium of hospitals.[39]

THE REQUEST FOR ASSISTED SUICIDE OR EUTHANASIA

In the United States, the public increasingly accepts physician-assisted suicide and euthanasia as moral practices and believes that these practices should be legal.[40] These views can be seen as the public's condemnation of at least two things: the way hospitals and physicians overtreat sick patients in their last days, making death a painful journey; and medicine's inadequate and ineffective treatment of suffering. These views also reflect a demand for more control in decisions about the end of life. Some people equate physician-assisted suicide with euthanasia, but they are different concepts. In physician-assisted suicide, the patient requests the physician's help in dying, usually in the form of a prescription of a lethal dose of medication to be taken at home. Euthanasia occurs when there is no patient request but the physician (or other health care professional) decides to hasten the patient's dying process in order to relieve suffering (the patient's or the physician's).

In June 1997, the United States Supreme Court ruled that there is no constitutional right to physician-assisted suicide.[41] This opinion did not remove the authority of individual states to outlaw or decriminalize physician-assisted suicide, however; and in November 1997, Oregon voters confirmed their acceptance of the Death with Dignity Act, which allows terminally ill Oregon residents to obtain from their physicians and to use prescriptions for self-administered, lethal medications. The act states that ending one's life in accordance with the law does not constitute suicide, and it specifically prohibits euthanasia (i.e., direct administration of a medication to end another person's life).

Many, if not most, of those patients who want physician-assisted suicide want it not to relieve suffering as ordinarily understood but to maintain control over their dying.[42] As of 1999 (2 years after legalization of physician-assisted suicide in Oregon), a survey of Oregon physicians found that they granted about one in six requests for a prescription for a lethal medication and that one in 10 requests actually resulted in suicide. Substantive palliative interventions led some—but not all—patients to change their minds about assisted suicide.[43,44] As of 2002, a total of 129 people had committed physician-assisted suicide in Oregon, corresponding to a rate of 8.8 per 10,000 deaths from any cause in the state.[45] Compared with Oregon residents who died of similar underlying causes, rates of physician-assisted suicide decreased with age and were higher among those who had been divorced and among those with higher levels of education. The rate of physician-assisted suicide was also higher among those afflicted with amyotrophic lateral sclerosis and cancer. The majority of patients using physician-assisted suicide were non-Hispanic whites, but a significant minority were Asian Americans. Overall, the number of Oregon residents using physician-assisted suicide has increased over the years, but it remains a very small minority relative to overall deaths.

Regardless of legal issues, however, when a patient requests a prescription for enough medication to commit suicide or to hasten death, the physician has the ethical responsibility to try to learn why. What is it that now makes death seem a better option than life? What is it that the patient feels must be avoided? From what is the patient trying to escape? Is the patient depressed? Why does the patient feel that that he or she can no longer be someone who matters? Are there financial considerations—that is, does the patient fear becoming a financial burden, a burden to care for, or both? Has any of this been discussed with the fam-

ily? How would the family understand the patient's requests and be affected by them? If the patient considers life to be devoid of value and meaning, does the patient's life still have meaning for other persons? Does this affect the patient? Has the patient made any effort to achieve consensus so that his or her death can be a meaningful, shared family experience?

FEAR OF LEGAL REPRISAL

When a physician makes a reasonable clinical judgment of irreversible illness and decides to forgo or stop life-sustaining treatment—whether on the wish of the patient or, if the patient is incompetent, with the agreement of the patient's proxy or surrogates—fear of litigation is neither a reasonable nor a legitimate excuse not to proceed. The courts have made it clear that these decisions are valid and that the persons involved in those decisions should not be brought to trial. It is irrational to demand guarantees that no litigation will follow, however. It is hoped that physicians' energies will be spent doing the best they can for the patient, in accord with the patient's wishes. Should litigation follow an action taken in accord with the above guidelines, the physician will be well prepared to defend the decisions in court.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Billings J: Recent advances: palliative care. BMJ 321:555, 2000

2. Pan CX, Morrison RS, Meier DE, et al: How prevalent are hospital-based palliative care programs? Status report and future directions. J Palliat Med 4:315, 2001

3. NHPCO facts and figures. National Hospice and Palliative Care Organization, 2003
http://www.nhpco.org

4. Health Care Financing Administration: Medicare hospice benefits. U.S. Dept. of Health and Human Services, Baltimore, 1999

5. Lynn J: Caring at the end of our lives. N Engl J Med 335:201, 1996

6. Davis MP, Walsh D, Nelson KA, et al: The business of palliative medicine. Part 2: The economics of acute inpatient palliative medicine. Am J Hosp Palliat Care 19:89, 2002

7. Covinsky KE, Eng C, Lui LY, et al: The last 2 years of life: functional trajectories of frail older people. J Am Geriatr Soc 51:492, 2003

8. Good care of the dying patient. Council on Scientific Affairs, American Medical Association. JAMA 275:474, 1996

9. Christakis NA: Timing of referral of terminally ill patients to an outpatient hospice. J Gen Intern Med 9:314, 1994

10. Deaths: Preliminary Data for 2001. National Vital Statistics Reports 51:1, 2003
http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_05.pdf

11. Seale C, Cartwright A: The Year Before Death. Ashgale Publishing Co, Brookfield, Vermont, 1994

12. National Population Projections. United States Census Bureau, 2002
http://www.census.gov/population/www/projections/natproj.html

13. Fairman RP: Withdrawing life-sustaining treatment: lessons from Nancy Cruzan. Arch Intern Med 152:25, 1992

14. Healey JM: Decisions at the end of life: the legacy of Karen Ann Quinlan. Conn Med 49:549, 1985

15. National Mortality Followback Survey. National Center for Health Statistics, 1998
http://www.cdc.gov/nchs/about/major/nmfs/nmfs.htm

16. A controlled trial to improve care for seriously ill hospitalized patients: the study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. JAMA 274:1591, 1995

17. Lynn J, Harrell F Jr, Cohn F, et al: Prognoses of seriously ill hospitalized patients on the days before death: implications for patient care and public policy. New Horiz 5:56, 1997

18. Koretz B, Whiteman E: Hospice eligibility. University of California Office of the President Academic Geriatric Resource Plan, 2002
http://www.ucop.edu/agrp/docs/la_hospice.pdf

19. Evaluation of prognostic criteria for determining hospice eligibility in patients with advanced lung, heart, or liver disease. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. JAMA 282:1638, 1999

20. Management of pain in elderly patients with cancer. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology. JAMA 279:1877, 1998

21. Ferrell BA, Ferrell BR, Rivera L: Pain in cognitively impaired nursing home patients. J Pain Symptom Manage 10:591, 1995

22. Sengstaken EA, King SA: The problems of pain and its detection among geriatric nursing home residents. J Am Geriatr Soc 41:541, 1993

23. Pain and treatment of pain in minority patients with cancer. The Eastern Cooperative Oncology Group Minority Outpatient Pain Study. Ann Intern Med 127:813, 1997

24. Feldt KS, Warne MA, Ryden MB: Examining pain in aggressive cognitively impaired older adults. J Gerontol Nurs 24:14, 1998

25. Gagliese L, Melzack R: Chronic pain in elderly people. Pain 70:3, 1997

26. Culture and Nursing Care: A Pocket Guide. Lipson JDS, Minarik PA, Eds. UCSF Nursing Press, San Francisco, 1996

27. Caralis PV, Davis B, Wright K, et al: The influence of ethnicity and race on attitudes toward advance directives, life-prolonging treatments, and euthanasia. J Clin Ethics 4:155, 1993

28. Leggat JE Jr, Bloembergen WE, Levine G, et al: An analysis of risk factors for withdrawal from dialysis before death. J Am Soc Nephrol 8:1755, 1997

29. Blackhall LJ, Frank G, Murphy ST, et al: Ethnicity and attitudes towards life sustaining technology. Soc Sci Med 48:1779, 1999

30. Garrett JM, Harris RP, Norburn JK, et al: Life-sustaining treatments during terminal illness: who wants what? J Gen Intern Med 8:361, 1993

31. Crawley L, Payne R, Bolden J, et al: Palliative and end-of-life care in the African American community. JAMA 284:2518, 2000

32. Llovera I, Ward MF, Ryan JG, et al: Why don't emergency department patients have advance directives? Acad Emerg Med 6:1054, 1999

33. Kvale JN, Melloh JR, Oprandi A, et al: Use of advance directives by community-dwelling older adults. J Clin Outcomes Meas 6:39, 1999

34. The discussion about advance directives: patient and physician opinions regarding when and how it should be conducted. End of Life Study Group. Arch Intern Med 155:1025, 1995

35. Carney MT, Neugroschl J, Morrison RS, et al: The development and piloting of a capacity assessment tool. J Clin Ethics 12:17, 2001

36. Pellegrino ED: Ethics. JAMA 270:202, 1993

37. Jecker NS, Schneiderman LJ: Medical futility: the duty not to treat. Camb Q Health Ethics 2:151, 1993

38. Truog RD, Brett AS, Frader J: The problem with futility. N Engl J Med 326:1560, 1992

39. Halevy A, Brody BA: Medical futility in end-of-life care. JAMA 282:1331, 1999

40. Paul P: Euthanasia and assisted suicide. Am Demogr 24:20, 2002

41. Vacco v. Quill. 521 US 793, 1997

42. Ganzini L, Harvath TA, Jackson A, et al: Experiences of Oregon nurses and social workers with hospice patients who requested assistance with suicide. N Engl J Med 347:582, 2002

43. Ganzini L, Nelson HD, Schmidt TA, et al: Physicians' experiences with the Oregon Death with Dignity Act. N Engl J Med 342:557, 2000

44. Ganzini L, Nelson HD, Lee MA, et al: Oregon physicians' attitudes about and experiences with end-of-life care since passage of the Oregon Death with Dignity Act. JAMA 285:2363, 2001

45. Hedberg K, Hopkins D, Kohn M: Five years of legal physician-assisted suicide in Oregon. N Engl J Med 348:961, 2003

46. Deaths: Final Data for 2000. National Vital Statistics Reports 50:8, 2002
http://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_15.pdf

X SYMPTOM MANAGEMENT IN PALLIATIVE MEDICINE

Maria Torroella Carney, m.d.
Jennifer Rhodes-Kropf, m.d.

The goal of palliative care is to provide comfort and support for both patient and family through the course of a life-threatening illness. Symptom control is essential to meeting that goal. This chapter discusses symptoms that commonly contribute to patients' suffering in terminal illness: pain; respiratory, gastrointestinal, mouth, and skin problems; and delirium.

Although this chapter focuses on physical and psychological symptoms, achieving symptom control requires the physician to address the patient's suffering in all its aspects: physical, psychological, social, and spiritual. Physical distress cannot be effectively treated in isolation from the emotional and spiritual components that contribute to it, nor can these sources of suffering be addressed adequately when patients are in physical distress. The various components of suffering must be addressed simultaneously [*see Clinical Essentials: XI Management of Psychosocial Issues in Terminal Illness*].

Symptom Assessment

A full and formal symptom assessment is necessary before effective treatment can be instituted.[1] Symptoms are inherently subjective[2]; therefore, patient self-reporting must be the primary source of information, and the clinician must believe what the patient says. If the patient is unable to report, a family member or professional can provide a surrogate assessment. However, several studies have demonstrated that observer and patient assessments are not well correlated.[3,4]

To compensate for this inherent subjectivity, researchers have developed symptom measurement systems that are intended to quantify patients' perceptions in a manner that is valid and reliable. Often, these measurement systems have taken the form of symptom checklists.[5,6] For example, the Edmonton Symptom Assessment Scale[5] comprises 14 questions that evaluate eight physical and psychological symptoms [*see Table 1*]. This scale has been extensively employed in palliative care research, in part because of its ease of use. Although the scale yields a numeric score (the higher the score, the more severe the patient's condition), the formal scoring mechanism is used only in research. In clinical practice, the scale can be used informally to evaluate a patient's status and follow it over time.

The Memorial Symptom Assessment Scale[7] characterizes 32 physical and psychological symptoms in terms of intensity and frequency, as well as the level of distress from the symptoms [*see Table 2*]. Although the Memorial Symptom Assessment Scale provides a greater range of information than the Edmonton Symptom Assessment Scale, the former is correspondingly more time consuming to use.

Physical Symptoms

PAIN

Diagnosis

Management of pain begins with a careful and detailed assessment [*see 11:XIV Pain*]. The goal of this assessment is to determine the location and character of the pain, define its cause (or causes), and develop a plan of care.

Pain cannot be measured objectively, and several studies have shown that medical care providers' estimates of patients' pain severity are significantly lower than the patients' self-reports.[8,9] Pain is independent of age, gender, marital status, physical function, and cognitive function.[10] Therefore, the central guiding principle of pain assessment is to ask the patient and believe the patient's description of pain.

Pain assessment in the elderly is often complicated by coexistent cognitive impairment. The cognitively impaired patient may be unable to express pain adequately or request analgesics and, therefore, is at increased risk for undertreatment of pain.[11,12] As with cognitively intact patients, the first step in the assessment of pain in demented patients is to ask them about their pain. Although patients with severe dementia may be incapable of communicating, many patients with mild or moderate impairment can accurately localize and grade the severity of their pain,[13] and these self-reports should be regarded as valid.

Untreated pain can result in agitation and disruptive behavior, and it may worsen or precipitate delirium, particularly in cognitively impaired patients.[14,15] When delirium prevents communication with the patient, the physician may have to infer that pain is present and proceed with treatment.

Treatment

Opioids are the standard choice for treating pain in terminally ill patients. The physician who provides palliative care needs to have the confidence and competence to prescribe opioids at whatever dose is needed to control pain, as well as the skill to determine when adjuvant analgesics (e.g., antidepressant or antiseizure medication) are needed to manage certain types of pain.[16,17] Terminally ill patients are a special population, often suffering chronic pain and taking pain medications over longer periods of time and at higher dosages.[18] Indeed, tolerance to opioids may require that they be used in amounts that would be fatal to the opioid-naive patient.

In a multisite study of terminally ill patients in the United States, Weiss and colleagues[19] found that half of terminally ill patients experienced moderate to severe pain but that less than one third wanted additional pain treatment from their primary care physician. Reasons for not wanting additional therapy included dislike of analgesic side effects and not wanting to take more pills or injections. Some patients, however, mentioned fear of addiction. This is a common—and unwarranted—concern not only of patients but of some medical personnel, as well.

As the goals of care change in the course of a life-threatening illness, higher dosages of pain medication may be needed to achieve comfort. In the last days of life, relief of suffering may require sedation to the point of unconsciousness, a technique referred to as palliative sedation (see below).

RESPIRATORY SYMPTOMS

Dyspnea

Shortness of breath has been described in 70% of cancer patients during the last 6 weeks of life[20] and in 50% to 70% of patients dying

Table 1 Modified Edmonton Symptom Assessment Scale[5]

1a. Please rate your *pain* now.
- 1. ☐ No pain
- 2. ☐ Mild pain
- 3. ☐ Moderate pain
- 4. ☐ Severe pain

1b. Please rate your *pain* over the past 3 days.
- 1. ☐ No pain
- 2. ☐ Mild pain
- 3. ☐ Moderate pain
- 4. ☐ Severe pain

1c. Is your *pain control* acceptable to you?
- 1. ☐ Very acceptable
- 2. ☐ Acceptable
- 3. ☐ Not acceptable

2. How would you describe your *activity level* over the past 3 days?
- 1. ☐ Very active
- 2. ☐ Somewhat active
- 3. ☐ Minimally active
- 4. ☐ Not active

3. How would you describe your amount of *nausea* over the past 3 days?
- 1. ☐ Not nauseated
- 2. ☐ Mildly nauseated
- 3. ☐ Moderately nauseated
- 4. ☐ Very nauseated

4a. How would you describe your level of *constipation* over the past 3 days?
- 1. ☐ No constipation
- 2. ☐ Mild constipation
- 3. ☐ Moderate constipation
- 4. ☐ Severe constipation

4b. When was your *last bowel movement*?
- 1. ☐ Today
- 2. ☐ Yesterday
- 3. ☐ 2–3 days ago
- 4. ☐ More than 4 days ago

5. How would you describe your feelings of *depression* over the past 3 days?
- 1. ☐ Not depressed
- 2. ☐ Mildly depressed
- 3. ☐ Moderately depressed
- 4. ☐ Very depressed

6. How would you describe your feelings of *anxiety* over the past 3 days?
- 1. ☐ Not anxious
- 2. ☐ Mildly anxious
- 3. ☐ Moderately anxious
- 4. ☐ Very anxious

7. How would you describe your level of *fatigue* over the past 3 days?
- 1. ☐ Not fatigued
- 2. ☐ Mildly fatigued
- 3. ☐ Moderately fatigued
- 4. ☐ Very fatigued

8. How has your *appetite* been over the past 3 days?
- 1. ☐ Very good appetite
- 2. ☐ Moderate appetite
- 3. ☐ Poor appetite
- 4. ☐ No appetite

9. How would you describe your sensation of *well-being* over the past 3 days?
- 1. ☐ Very good sensation of well-being
- 2. ☐ Moderately good sensation of well-being
- 3. ☐ Not very good sensation of well-being
- 4. ☐ Poor sensation of well-being

10. How *short of breath* have you been over the past 3 days?
- 1. ☐ No shortness of breath
- 2. ☐ Mild shortness of breath
- 3. ☐ Moderate shortness of breath
- 4. ☐ Very short of breath

11. How has your *physical discomfort* been over the past 3 days?
- 1. ☐ No physical discomfort
- 2. ☐ Mild physical discomfort
- 3. ☐ Moderate physical discomfort
- 4. ☐ Severe physical discomfort

of other illnesses.[21] Ventafridda and colleagues[22] observed "horrible and unpleasant" dyspnea in 10% of cancer patients dying in a palliative care unit. Like pain, dyspnea is a subjective symptom that may not correlate with any objective signs of respiratory compromise,[23] and hence, its management can be challenging.

It is important to diagnose and treat any underlying reversible causes of dyspnea. For example, dyspnea caused by congestive heart failure will require diuretics and inotropic support [*see 1:II Congestive Heart Failure*].

When therapy specific to the underlying cause is unavailable or ineffective, several techniques may alleviate breathlessness. Simple measures include pursed-lip breathing and diaphragmatic breathing, leaning forward with arms on a table, cool-air ventilation (from a fan or an open window), and nasal oxygen.

Opioids have been shown in numerous studies to be highly effective in the amelioration of dyspnea.[24,25] In one study,[24] morphine in doses sufficient to relieve dyspnea had no measurable adverse effect on respiratory rate or effort, oxygen saturation, and carbon dioxide concentration. Therefore, morphine is the drug of choice for treating otherwise refractory dyspnea in terminal illness.

Lorazepam and other benzodiazepines are also widely used, especially in terminally ill patients whose dyspnea has an anxiety component, although evidence to support this practice is limited.[26] In addition, steroids and oxygen therapy may be of benefit [*see Table 3*].

Cough

Cough can be an annoyance or can develop into a major source of suffering by causing muscle strain, increasing fatigue, and interrupting sleep. In one study of lung cancer patients, cough was the most common symptom, affecting 80% of patients until just before death.[27] Because the causes of cough are varied, the optimal treatment is treatment of the underlying problem, if possible. When such treatment is not possible, management depends on whether the cough is productive [*see Figure 1*].[28] A productive cough may improve with chest physiotherapy, oxygen, humidity, and suctioning. Antibiotics for infection, *N*-acetylcysteine, bronchodilators, and guaifenesin are also effective.[29,30] Opioids, antihistamines, and anticholinergics decrease mucus production, which can decrease the stimulus for cough. Cough suppressants can be harmful if used in patients with productive coughs by causing mucus

retention,[29,30] which may lead to the formation of mucous plugs and airway obstruction. A patient with a nonproductive cough may benefit from a cough-suppressing agent such as a local anesthetic (e.g., nebulized bupivacaine), bronchodilators, opioids, or a soothing agent such as a lozenge. Benzonatate, steroids, and opiates are effective treatments. Opioids act centrally and are one of the most effective agents against cough. Nonopioid antitussives, such as dextromethorphan, may work synergistically with opiates.[30]

GASTROINTESTINAL SYMPTOMS

Anorexia, nausea and vomiting, constipation, bowel obstruction, and diarrhea are common and potentially devastating in terminal illness.

Anorexia

Anorexia is nearly universal in patients with a terminal illness.[31] Evaluation of anorexia should be concentrated on finding a reversible or treatable cause. It is important to note that cognitive impairment, which is also highly prevalent in advanced disease, may cause a person to be misdiagnosed as anorexic, because the person may be unable to obtain, prepare, or eat meals.[32] Often in terminal disease, however, the patient simply loses the desire to eat.

Patients themselves may complain of anorexia, in some cases because they find the resulting cachexia unacceptable. In those cases, the decision to treat is straightforward. However, anorexia can often be of more concern to family, friends, and medical staff than to patients themselves. The family may be concerned because loss of appetite is seen as a certain sign of impending death.[33] Concern about anorexia may also be rooted in the emotional and psychological meanings that surround food and its consumption: not feeding the patient may be considered equivalent to not caring about the patient. The family should be reassured that anorexia in terminal disease is usually not associated with suffering; especially at the end of life, patients rarely feel hunger or thirst, and many patients who stop eating experience analgesia and even euphoria. Excessive proteins and lipids can induce nausea and vomiting in such cases, and excessive hydration can result in edema and dyspnea.[34]

In the early stages of terminal illness, however, studies have shown that the treatment of anorexia with appetite stimulants may improve patients' quality of life.[35,36] Treatment can begin with simple measures. The patient should be encouraged to eat without any restrictions on sugar, salt, or fats, when possible. Alcohol has appetite-stimulating properties, so patients may wish to consider a cocktail or glass of wine before the evening meal.[37]

Appetite stimulants with proven efficacy in palliative care include dexamethasone, in dosages of 2 to 20 mg/day (recommended because its long half-life permits once-daily dosing and because it has minimal mineralocorticoid effects); megestrol acetate (beginning with 200 mg every 8 hours and titrating to 800 mg/day); and cannabinoids (e.g., tetrahydrocannabinol [THC]), starting with a small dose and titrating to effect and tolerability. Dexamethasone and megestrol tend to be used more often than cannabinoids because of the restricted availability of cannabinoids.

Anorexia in patients with dementia Because Alzheimer disease destroys higher brain function while sparing the other major organ systems, many patients with Alzheimer disease progress to a stage at which they are unable to eat on their own or even chew and swallow reliably but may survive for years if artificial hydration and nutrition are provided. Deciding whether to insert a gastrostomy tube in such patients can be challenging. Complications of tube feeding are common and include repeated infections, whose treatment may require needle sticks, transfer to a hospital, and restraints; these are especially burdensome for a confused patient who cannot understand the reason for such interventions.[38] In addition, patients with advanced neurologic impairment are at high risk for pneumonia from a variety of causes, including but not limited to aspiration. There is no evidence that tube feeding reduces the risk of pneumonia in such patients; it may even increase the risk.[39] One may ask what is to be gained with artificial nutrition and hydration in such cases.

Because of the terminal and irreversible nature of end-stage dementia and the substantial burden that continued life-prolonging care may pose for these patients, they may be better served by palliative care that focuses predominantly on their comfort. Comfort care is viewed as preferable to life-prolonging

Table 2 Memorial Symptom Assessment Scale[14]

For physical symptoms, patients are instructed to check off all symptoms experienced during the past week and the degree to which the symptom bothered or distressed them. Categories and scores are as follows: Not at all (0), A little bit (1), Somewhat (2), Quite a bit (3), and Very much (4). Patients may also add symptoms not listed and rate them on the same scale. For psychological symptoms, patients are instructed to check off all symptoms experienced during the past week and how often each occurred. Categories and scores are as follows: Rarely (1), Occasionally (2), Frequently (3), and Almost constantly (4). Patients may also add symptoms not listed and rate them on the same scale.

Physical Symptom	*Severity*				
	0	*1*	*2*	*3*	*4*
Difficulty concentrating					
Pain					
Lack of energy					
Cough					
Changes in skin					
Dry mouth					
Nausea					
Feeling drowsy					
Numbness or tingling in hands and feet					
Hair loss					
Constipation					
Swelling of arms or legs					
Psychological Symptom	*Frequency*				
	0	*1*	*2*	*3*	*4*
Feeling sad					
Worrying					
Feeling irritable					
Feeling nervous					

Table 3 Drug Treatment for Dyspnea[33]

Drug (Trade Name)	*Dosage*	*Comment*
Oral morphine	2.5–5 mg p.o., q. 4 hr while awake	Doses for opiate-naive patients
I.V. morphine	0.5 mg/hr; titrate to relief	Once dose requirement established, switch to long-acting oral opiate or fentanyl patch
Nebulized morphine	2.5–10 mg injectable in 2 ml NS	—
Nebulized hydromorphone	0.25–1 mg injectable in 2 ml NS	—
Nebulized albuterol	0.083% (3 ml)	Possible adjunct to opioid
Nebulized methylprednisolone (Solu-Medrol)	10 mg	Possible adjunct to opioid
Dexamethasone	Day 1: 16 mg p.o.; days 2–3: 8 mg b.i.d.; days 3–4: 4 mg b.i.d.; subsequent: 2 mg b.i.d.	Possible adjunct to opioid
Prednisone	40 mg b.i.d. for 5–7 days	Possible adjunct to opioid
Lorazepam (Ativan)	1–10 mg/day in two or three divided doses; usual dose, 2–6 mg/day in divided doses. Elderly: 0.5–4 mg/day	For patients whose dyspnea has an anxiety component
Oxygen	2 L/min by nasal cannula; titrate to relief	—

NS—normal saline

measures by a substantial proportion of nursing home patients and family members.[40] Families should be reassured that it is never unethical to withhold nutrition and hydration if they are not helping the patient.

Nausea and Vomiting

Nausea and vomiting occur in up to 62% of patients with terminal cancer[41] and 27% of patients dying of other causes. There are multiple potential causes for both nausea and vomiting [*see Table 4*].[37] Once the cause has been determined, symptomatic relief is relatively easy to achieve with the appropriate medications [*see Table 5*].[28] Without an understanding of the underlying etiology, it may be impossible to find the most beneficial form of treatment.

The central nervous system and the gastrointestinal tract are particularly important in nausea and vomiting.[42] The gastric lining, the chemoreceptor trigger zone in the base of the fourth ventricle, the vestibular apparatus, and the cortex are all involved in the physiology of nausea. Stimulation of the vomiting center in the brain from one or more of these areas is mediated through the neurotransmitters serotonin, dopamine, acetylcholine, and histamine. Serotonin seems to be important in the gastric lining and the CNS, whereas acetylcholine and histamine are important in the vestibular apparatus. Cortical responses are mediated via both neurotransmitters and learned responses (e.g., nausea related to anxiety or anticipatory nausea with chemotherapy).

The major causes of nausea and vomiting can be classified by the mechanisms' principal site of action. Dopamine-mediated nausea is probably the most common form of nausea and the most frequently targeted one for initial symptom management. Antidopamine medications are phenothiazines and butyrophenone neuroleptics (metoclopramide and prochlorperazine). They may cause drowsiness and extrapyramidal symptoms. Haloperidol is a highly effective antinausea agent and may be less sedating. Antihistamines such as diphenhydramine can be used to control nausea, but they may cause sedation. Antihistamines also have anticholinergic properties. Serotonin has been implicated in chemotherapy-associated nausea. Antiserotonin medications (e.g., odansetron) can be effective, but they are expensive.

Nausea can also result from slow gastric and intestinal motility, so-called squashed stomach syndrome from mechanical compression of the stomach, and constipation. Hence, prokinetic agents (e.g., metoclopramide) and aggressive fecal disimpaction and institution of a bowel regimen (see below) should be considered as therapeutic modalities. In some patients, hyperacidity and mucosal erosion may also be associated with significant nausea. In these patients, consider the use of antacids, histamine$_2$ blockers, proton-pump inhibitors, and misoprostol [*see 4:XIV Gastrointestinal Motility Disorders*].

Constipation

Constipation can lead to serious complications, such as bowel obstruction, ulceration, or perforation, as well as delirium. Because constipation is so common in terminal illness, appropriate management includes the institution of preventive measures in patients at high risk for this complication.

Diagnosis Assessment of constipation begins with inquiry about the frequency and consistency of stools; possible contributing factors, such as medications, reduced mobility, and a low-fiber diet; and any accompanying symptoms that suggest complications, such as nausea, vomiting, abdominal pain, distention, and discomfort.[43] As with any symptom, the search for a reversible cause is primary. A plain x-ray can be useful to evaluate for ileus or bowel obstruction. Invasive evaluation with colonoscopy should be considered in difficult, refractory, or complicated cases.

Many medications can contribute to constipation. These include beta blockers, calcium channel blockers, anticholinergic agents, and diuretics.[43,44] First and foremost, however, are opioid analgesics: constipation is a universal side effect of opioid therapy, especially in the terminally ill. For that reason, every terminal-

ly ill patient who is placed on opioids should also be started on a preventive regimen for constipation. The bowel regimen in these patients starts with stool softeners and stimulant laxatives and progresses through hyperosmotic agents and enemas, as necessary [*see Table 6*].[45] This regimen can also be utilized for treatment of constipation from other causes, once intestinal obstruction is ruled out.

Treatment Treatment of constipation is with oral agents, rectal suppositories, or enemas and can focus on softening the stool, enlarging stool volume, or promoting bowel peristalsis. Laxative categories include detergents, stimulants, osmotic agents, prokinetic agents, lubricant stimulants, and large-volume enemas [*see Table 7*]. Polyethylene glycol solution (GoLYTELY) or powder (MiraLax) is an osmotic agent that is marketed as a bowel cleanser to prepare patients for colonoscopy, but it is often effective in relieving constipation and may cause less cramping than other laxatives. Whichever laxative is chosen, the clinician should prescribe the maximum therapeutic dose of the agent before switching to another one.

Fecal impaction Although impaction of stool in the rectum is a complication of constipation, the typical clinical manifestation is so-called overflow diarrhea from leakage of unformed stool around the obstruction. A digital rectal examination may confirm fecal impaction in the distal rectum, but abdominal x-rays may be required for the diagnosis of more proximal impaction. Treatment of fecal impaction is from below, utilizing digital disimpaction and rectal laxatives (suppositories, enemas, or both); only if those fail should oral treatment be attempted.[43]

Bowel Obstruction

The prevalence of bowel obstruction is as high as 40% in bowel and pelvic cancers.[46] Constipation and fecal impaction are the most common causes of bowel obstruction in terminal illness. Symptoms of bowel obstruction include anorexia, confusion, nausea and vomiting, constipation, and pain. Diagnosis is made on the basis of the clinical presentation and abdominal x-rays.

Consultation with a surgeon is advisable to establish a treatment plan. In addition to aggressive measures to prevent or treat constipation and fecal impaction (see above), treatment of bowel obstruction may involve surgical relief of obstruction, nasogastric suction, and pharmacologic measures. Colicky or cramping pain may respond to dicyclomine, opioids (parenteral or rectal), and warm soaks to the abdomen. The obstruction and associated nausea and vomiting may respond to metoclopramide, haloperidol, or dexamethasone. Parenteral octreotide is also useful in this setting to decrease the volume of bowel secretions.

Diarrhea

Diarrhea, which is often secondary to fecal impaction or antibiotic-associated colitis, is a particularly distressing and exhausting symptom in the terminally ill patient.[43] Once impaction, overgrowth, and other causes (e.g., gastrointestinal bleeding, malabsorption, and medications) have been ruled out, kaolin-pectin, psyllium, loperamide, or tincture of opium may be tried.

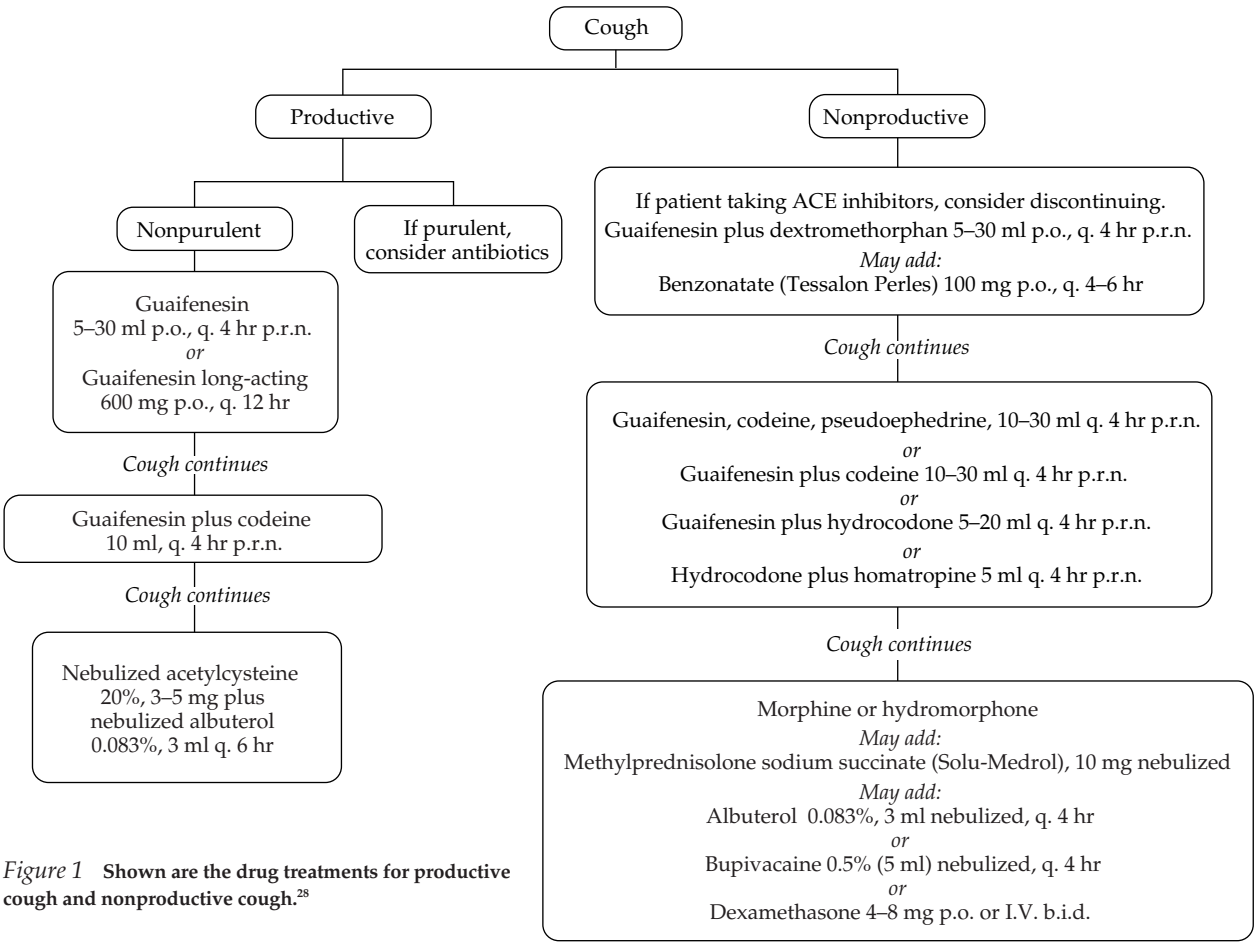


Figure 1 **Shown are the drug treatments for productive cough and nonproductive cough.[28]**

Table 4 Management of Nausea and Vomiting[37]

Etiology	*Pathophysiology*	*Therapy*
Mechanical obstruction—intraluminal	Constipation, obstipation	Laxatives; disimpaction
Mechanical obstruction—extraluminal	Tumor, fibrotic stricture	Surgery, fluid management, steroids, octreotide, scopolamine
Medications—chemotherapy	Chemoreceptor trigger zone, GI tract	Antiserotonin, antidopamine, steroids
Medications—NSAIDs	GI tract irritation	Cytoprotective agents, antacids
Medications—opioids	Chemoreceptor trigger zone, vestibular effect, GI tract	Antidopamine, anticholinergic, prokinetic agents, stimulant cathartics
Medications—other	Chemoreceptors	Antidopamine, antihistamine
Meningeal irritation	Increased intracranial pressure	Steroids
Mentation (e.g., anxiety)	Cortical	Anxiolytics
Metabolic—hypercalcemia	Chemoreceptor trigger zone	Antidopamine, antihistamine
Metabolic—hyponatremia	Chemoreceptor trigger zone	Antidopamine, antihistamine
Metabolic—hepatic/renal failure	Chemoreceptor trigger zone	Rehydration, steroids
Metastases—cerebral	Increased intracranial pressure	Steroids, mannitol
	Chemoreceptor trigger zone	Antidopamine, antihistamine
Metastases—liver	Toxin buildup	Antidopamine, antihistamine
Microbes—gastroenteritis	GI tract	Anti-infectives, antacids
Microbes—sepsis	Chemoreceptor trigger zone	Antidopamine, antihistamine, anti-infectives
Movement	Vestibular stimulation	Anticholinergic
Mucosal irritation	GI hyperacidity, GERD	Cytoprotective agents, antacids
Myocardial—ischemia, CHF	Vagal stimulation, cortical, chemoreceptor trigger zone	Oxygen, opioids, antidopamine, antihistamine, anxiolytics

CHF—congestive heart failure GERD—gastroesophageal reflux disease NSAIDs—nonsteroidal anti-inflammatory drugs

Octreotide (see above) is an effective means of reducing gastrointestinal secretions.

MOUTH SYMPTOMS

Oral problems can cause altered taste, pain, and difficulty swallowing, which may lead to reduced food and fluid intake. Good hydration, hygiene, and regular observation can keep oral problems to a minimum. The patient's teeth should be brushed twice daily with toothpaste. Daily observation of the oral mucosa is recommended.

Dentures also require regular cleansing. Dentures may cease to fit properly in patients who lose a significant amount of weight. Some of those patients may wish to have their dentures refitted; others (especially those nearing death) will choose to forgo this arduous process.

Key questions to ask regarding the mouth include the following: Is the mouth dry? Is infection present? Is the mouth dirty? Is the mouth painful? Are oral ulcerations present? [*see Table 8*][47]

Dry Mouth

The presence of saliva is usually taken for granted, but the lack of it can seriously damage the quality of life. Xerostomia (the subjective sensation of dry mouth) may result from salivary gland disease or systemic conditions such as Sjögren syndrome, Parkinson disease, AIDS, or diabetes[48]; it may also be a side effect of medications, including those with anticholinergic action, benzodiazepines, diuretics, and interleukin-2.[49] Regardless of the cause, xerostomia almost always requires symptomatic treatment. The goal of therapy is to moisten the oral mucosa, and the best, simplest way is for the patient to sip water frequently. However, mouth moisteners and artificial salivas exist and may be preferred by some patients.[47,49] Pilocarpine tablets may be used, at a dosage of 5 to 10 mg every 8 hours, if the above measures fail. Side effects may include nausea, diarrhea, urinary frequency, and dizziness. Other nonpharmacologic treatments include eating ice chips and sucking on hard candy.

Oral Ulcers/Mucositis

Oral infection can have multiple causes. Aphthous ulcers are common and can be eased by topical corticosteroids, tetracycline mouthwash, or thalidomide. Oral candidiasis usually presents as adherent white plaques but can also present as erythema or angular cheilitis. Nystatin suspension is the usual treatment, but a 5-day course of oral ketoconazole, 200 mg daily, can also be used. Severe viral infection (herpes simplex or zoster) requires treatment with acyclovir, 200 mg every 4 hours for 5 days. Malignant ulcers are often associated with anaerobic bacteria and may respond to metronidazole, 400 to 500 mg orally or rectally every 12 hours or as a topical gel.[47]

Pressure Ulcers

Pressure ulcers typically result from both intrinsic and extrinsic factors [*see Table 9*]. Major sites of pressure ulcers in terminally ill patients include the ear and the skin overlying the spine (apex of kyphosis), sacrum, greater femoral trochanter, head of the fibula, and malleolus. Prevention should emphasize these sites and should include daily visual inspection of them in patients at risk for pressure sores.

Prevention and treatment of pressure sores require targeting risk factors and minimizing them. Caregivers need to minimize pressure by turning and repositioning the patient frequently and avoiding shear (sliding movement) and friction. They should be

Table 5 Medications for Nausea and Vomiting[28]

Administration	*Category*	*Drug (Trade Name)*	*Dosage*
Oral	Corticosteroid	Dexamethasone	2–8 mg q. 6–12 hr
	Antidopamine	Haloperidol (Haldol)	0.5–5 mg q. 6–8 hr
		Prochlorperazine (Compazine)	5–10 mg q. 4–6 hr
		Prochlorperazine SR	10–15 mg b.i.d.
	Antihistamine	Diphenhydramine (Benadryl)	25–50 mg q. 4–6 hr
		Hydroxyzine (Atarax)	25–50 mg t.i.d.–q.i.d.
		Promethazine (Phenergan)	12.5–25 mg t.i.d.–q.i.d.
	Anticholinergic	Hyoscyamine (Levsin)	0.125–0.25 S.L. q. 4 hr
		Meclizine (Antivert)	12.5–25 mg b.i.d.–q.i.d.
	Anxiolytic	Lorazepam (Ativan)	1–2 mg q. 2–4 hr
	Prokinetic	Metoclopramide (Reglan)	10–40 mg q.i.d.
	Antiserotonin	Ondansetron (Zofran)	8 mg p.o., t.i.d.–q.i.d.
	Other	Dronabinol (Marinol)	2.5–10 mg b.i.d., t.i.d.
		Thiethylperazine (Torecan)	10 mg q.d.–t.i.d.
		Trimethobenzamide (Tigan)	250 mg t.i.d.–q.i.d.
Rectal suppositories	Antidopamine	Prochlorperazine (Compazine)	25 mg q. 6 hr
	Antihistamine	Promethazine (Phenergan)	12.5, 25, 50 mg t.i.d.–q.i.d.
	Other	Trimethobenzamide (Tigan)	200 mg t.i.d.–q.i.d.
Continuous intravenous infusion	Corticosteroids	Dexamethasone	8–100 mg/24 hr
	Antidopamine	Haloperidol (Haldol)	2.5–10 mg/24 hr
	Anticholinergic	Hyoscyamine (Levsin)	1–2 mg/24 hr
		Scopolamine	0.8–20 mg/24 hr
	Antiserotonin	Odansetron (Zofran)	0.45 mg/kg/24 hr
	Prokinetic	Metoclopramide (Reglan)	20–80 mg/24 hr
Intermittent intravenous infusion	Corticosteroids	Dexamethasone	2–8 mg q. 4–6 hr
	Antidopamine	Haloperidol (Haldol)	0.5–2 mg q. 4–6 hr
		Prochlorperazine (Compazine)	5–10 mg q. 4–6 hr
	Antihistamine	Diphenydramine (Benadryl)	25–50 mg q. 6 hr
	Anxiolytic	Lorazepam (Ativan)	1–2 mg q. 6–8 hr
	Prokinetic	Metoclopramide (Reglan)	10–20 mg q. 6 hr
	Antiserotonin	Ondansetron (Zofran)	4–8 mg q. 8 hr
		Granisetron (Kytril)	10 µg/kg q.d.
	Other	Dronabinol (Marinol)	5 mg/m^2 q. 4 hr (maximum, six doses/day)

aware that even crumpled bedclothes can impair circulation. How a patient moves or is moved by caregivers needs to be assessed and monitored. Even with regular turning and careful lifting and positioning, special pressure surfaces or mattresses are sometimes needed.[47] Fragile skin that is at risk for breakdown should be covered with clear, occlusive dressings; pressure points should be covered with thin, hydrocolloid dressings.

Caregivers must keep the patient's skin clean and dry. Absorbent surfaces, urinary catheters, and rectal tubes may be helpful, but they must be used carefully because of their attendant complications.[37]

Nutrition is an important factor in both prevention and treatment. Good hydration, a diet that is high in protein and carbohydrates, and vitamin C supplements help maintain skin integrity and encourage healing.

If pressure ulcers develop, they should be covered with gel or colloid dressings, which keep the area moist, reduce pain, and can be left in place for several days. The pain of dressing changes can be eased by extra analgesia before each change.[47] The clinician should instruct caretakers to give oral pain medication one-half hour before the dressing change. The dose is determined by whether or not the patient is on regular opioid medications. If the patient is not on regular pain medications, start with 15 mg of immediate-release morphine. If the patient is on a regular opioid regimen, the predressing dose should be the same as the rescue dose.

Pressure ulcer management needs to be consistent with the overall goals of care. If maintenance or improvement of function is the goal and the patient's life expectancy is weeks to months, the ulcer should be treated according to the usual management guidelines. If life expectancy is very limited (e.g., days), the intent should be to optimize quality of life and minimize pain and discomfort (such as from excessive dressing changes or debridement).

Malignant Ulcers

For uncomplicated malignant ulcers, pain relief and wound care are managed in the same way as pressure ulcers. Malignant wounds can present special problems, however, which may include bleeding, exudate, infection, odor, and disfigurement. A bleeding malignant ulcer should be treated with radiation therapy, topical sucralfate, or topical tranexamic acid. Dirty ulcers should be debrided, which can be accomplished chemically. Altered body image from disfiguring wounds can be lessened with cavity foam dressings. Furthermore, empathetic listening is often therapeutic in itself. Anxiety, anger, or depression needs specific support, however[47] [*see Clinical Essentials: XI Management of Psychosocial Issues in Terminal Illness*].

Foul-Smelling Wounds

Odors may be very distressing to patients, families, and caregivers and may lead to poor quality of care, as even professional caregivers tend to avoid sickening smells. Odors are usually due to anaerobic infections or poor hygiene. Treat superficial infections with topical metronidazole or silver sulfadiazine. These agents are expensive, however; and if a less costly alternative is required, a diluted hydrogen peroxide solution can be used.[37] For soft tissue infections, add systemic metronidazole to topical management.

To control odors, place a pan containing kitty litter or activated charcoal under the patient's bed, provide adequate room ventilation, place an open cup of vinegar in the room, or burn a candle. Special charcoal-impregnated dressings placed over the odorous wound may also be helpful.[37]

Psychiatric Symptoms

Adjustment disorders, depression, anxiety, dementia, and delirium are the most common psychiatric problems encountered in dying patients.[50,51] Depending on their severity, management of these psychiatric problems may be within the capacity of the primary care physician or may require referral [*see Clinical Essentials: XI Management of Psychosocial Issues in Terminal Illness*].

DELIRIUM

Delirium occurs in roughly 75% of terminally ill patients.[52,53] Symptoms of delirium include inability to maintain attention, waxing and waning of consciousness, psychomotor changes, disturbance of sleep-wake cycle, disorientation, visual or auditory hallucinations, and problems with memory and language.[54] Other terms often used synonymously with delirium include acute confusional state, metabolic encephalopathy, and sundowning. In contrast to dementia, delirium is more rapid in onset (developing over hours to days), fluctuates in severity, is potentially reversible, and is associated with a lesser degree of memory impairment.

Table 6 A Progressive Bowel Regimen for Patients Receiving Opioid Therapy[45*]

Step 1
Docusate, 100 mg b.i.d.
Senna, 1 tablet q.d. or b.i.d.

Step 2
Docusate, 100 mg b.i.d.
Senna, 2 tablets b.i.d.
Bisacodyl rectal suppositories, 1–2 after breakfast

Step 3
Docusate, 100 mg b.i.d.
Senna, 3 tablets b.i.d.
Bisacodyl rectal suppositories, 3–4 after breakfast

Step 4
Docusate, 100 mg b.i.d.
Senna, 4 tablets b.i.d.
Lactulose or sorbitol, 15 ml b.i.d.
Bisacodyl suppositories, 3–4 after breakfast

Step 5
Sodium phosphate or oil-retention enema; if no results, add a high-colonic tap-water enema

Step 6
Docusate, 100 mg b.i.d.
Senna, 4 tablets b.i.d.
Lactulose or sorbitol, 30 ml b.i.d.
Bisacodyl rectal suppositories, 3–4 after breakfast

Step 7
Docusate, 100 mg b.i.d.
Senna, 4 tablets b.i.d.
Lactulose or sorbitol, 30 ml q.i.d.
Bisacodyl rectal suppositories, 3-4 after breakfast

*The bowel regimen is started at the time of or before the initiation of opioid therapy, and it should be continued for the duration of opioid therapy. The clinician should start with step 1 and progress through higher steps until an effective regimen is found.

Table 7 Treatments for Constipation

Laxative type	*Mechanism*	*Agent*	*Dosage*	*Comment*
Stimulant	Irritate the bowel and increase peristaltic activity	Prune juice	120-240 ml q.d. or b.i.d.	
		Senna	1–2 tablets p.o., q.h.s.	Titrate to effect; ≤ 8 tablets b.i.d.
		Bisacodyl	10–15 mg p.o., h.s.; or 10 mg p.r., after breakfast	Titrate to effect
Osmotic	Draw water into the bowel lumen, increase overall stool volume	Lactulose	30 ml p.o., q. 4–6 hr	Titrate to effect
		Sorbitol, 70% solution	2 ml/kg, up to 50 ml p.o., q.d.–t.i.d.	
		Milk of magnesia	1–2 tbsp, q.d.–t.i.d.	
		Magnesium citrate	1–2 bottles p.r.n.	
		Polyethylene glycol solution	1–4 L p.o.	Drink 8 oz q. 10 min until consumed
		Polyethylene glycol powder	17 g (1 tbsp) powder in 8 oz water, q.d.	2–4 days may be required to produce a bowel movement; increase dose as needed
Detergent (stool softeners)	Increase water content in stool by facilitating the dissolution of fat	Docusate sodium	1–2 capsules p.o., q.d.–b.i.d.	Titrate to effect
		Docusate calcium*	1–2 capsules p.o., q.d.–b.i.d.	Titrate to effect
Prokinetic agents	Stimulate the bowel's myenteric plexus, and increase peristaltic activity and stool movement	Metoclopramide	10–20 mg p.o., q. 6 hr	
Lubricant stimulants	Lubricate the stool and irritate the bowel, increasing peristaltic activity and stool movement	Glycerin suppositories	Daily	
		Mineral oil or peanut oil enema	Daily	
Large-volume enemas	Soften stool by increasing its water content; distend the colon and induce peristalsis	Warm-water enema	Daily	Addition of soapsuds irritates bowel wall to induce peristalsis
High-colonic enemas	Utilize gravity to bring fluid to more proximal parts of bowel	2 L of water or saline warmed to body temperature, hung on I.V. pole at ceiling level	Run in over 30 min, repeat q. 1 hr	

*Not available in the United States.

Delirium is a multifactorial syndrome, involving preexisting risk factors and precipitating factors that occur during hospitalization. Factors that predispose a patient to delirium include vision impairment, severe illness, cognitive impairment, and dehydration.[55] In older patients, cognitive impairment that is so mild as to be inapparent when they are well may nevertheless increase the risk of delirium. Precipitating factors include the use of physical restraints, malnutrition, taking more than three drugs, bladder catheter use, and any iatrogenic event.[55] Prevention of delirium can be accomplished by targeting risk factors.[55]

Management of delirium in the terminally ill patient includes correction of the cause and provision of symptomatic relief. Identification and treatment of underlying diseases or conditions is paramount—for example, give antibiotics for sepsis or oxygen for shortness of breath. In patients with underlying dementia, the possibility of untreated pain deserves special consideration. In the past, physicians were taught that the use of narcotic analgesics is dangerous in patients with dementia because those agents cause delirium. That is not true of a demented patient who becomes agitated or belligerent because of pain, however; in those cases, a dose of a narcotic analgesic may calm the patient within an hour or so. The risk of undertreating severe pain should be of greater concern, both medically and ethically, than the risk of worsening delirium with analgesic medications.

Additional means of treating delirium include minimizing any sensory impairments by providing appropriate eyeglasses or hearing aids and maintaining a quiet, familiar, and reassuring

Table 8 Local Measures for Oral Problems[47]

Dry mouth
- Semifrozen fruit juice
- Frequent sips of cold water or water sprays
- Petroleum jelly rubbed on lips

Dirty mouth
- Regular brushing with soft toothbrush and toothpaste
- Pineapple chunks
- Cider and soda mouthwash

Infected mouth
- Tetracycline mouthwash, 250 mg every 8 hr (one capsule dissolved in 5 ml water)

Painful mouth
- Topical corticosteroids: betamethasone, 0.5 mg in 5 ml water, as mouthwash; or triamcinolone in carmellose paste
- Coating agents: sucralfate suspension as mouthwash, carmellose paste, carbenoxolone
- Topical anesthesia: benzocaine or lozenges containing local anesthetics

setting. It is important to maintain communication with the patient, using frequent reorientation; familiar objects, places, and people; and avoidance of stimulus overload or deprivation.[56]

Pharmacologic symptom relief is best achieved with the use of an antipsychotic agent such as haloperidol or risperidone [*see Table 10*]. Benzodiazepines or sedatives should be used only if antipsychotic agents fail.[57]

Terminal Delirium

Delirium may be an irreversible part of the dying process. Many terminally ill patients have escalating restlessness, agitation, or hallucinations that can be relieved only with sedation.[58] When death is imminent, reversing the underlying causes of delirium is not possible. Instead, the clinician should focus on the management of the symptoms associated with the terminal delirium and bring comfort to the patient and family.

Benzodiazepines are widely used in the management of terminal delirium because they are anxiolytics, amnestics, skeletal muscle relaxants, and antiepileptics. Oral lorazepam (1 to 2 mg as an elixir, or the tablet predissolved in 0.5 to 1.0 ml of water and administered against the buccal mucosa) should be given every hour as needed; it will settle most patients at a daily dose of 2 to 10 mg. The lorazepam can then be given in divided doses, every 3 to 4 hours, to keep the patient settled. For a few extremely agitated patients, high doses of lorazepam—20 to 50 mg or more per 24 hours—may be required. A midazolam infusion (1 to 5 mg S.C. or I.V. every 1 hour, preceded by repeated loading boluses of 0.5 mg every 15 minutes to effect) may be a rapidly effective alternative.[37]

Palliative sedation When terminal delirium cannot be adequately controlled despite aggressive efforts to identify a tolerable therapy that does not compromise consciousness, it may be necessary to resort to palliative sedation. Most physicians define palliative sedation as the act of purposely inducing and maintaining a pharmacologically sedated and unconscious state, without the intent to cause death.

Once palliative sedation is initiated, the dosage of the sedative agent should not be increased unless the patient awakens or becomes restless, tachypneic, or tachycardic. Increasing the level of sedation in the absence of a clinical indication might imply that the physician is intending to hasten death, which if true would cross the line between palliative sedation and physician-assisted suicide or euthanasia [*see Clinical Essentials: IX Palliative Care*].[59]

Table 9 Risk Factors for Pressure Ulcers

Intrinsic	*Extrinsic*
Malnutrition	Pressure
Protein	Shear
Vitamin C	Trauma
Zinc	Friction
Diminished mobility	Crumpled bedclothes
Tissue fragility	Restraints
Anemia	Bed rails
Dehydration	Poor hygiene
Hypotension	Hospital equipment
Poor peripheral perfusion	Oxygen tubing
Incontinence	Heart monitor wires
Neurologic deficit	
Sensory	
Motor	
Older age	
Coma	
Moribund state	

Table 10 Drug Treatment for Agitation or Delirium[28]

Acute	Haloperidol, 0.5–5 mg p.o., p.r., I.M., I.V., or S.C.; titrate until calm
	Chlorpromazine, 1 mg I.V. q. 2 min until calm
Chronic	Haloperidol, 0.5–5 mg p.o. or p.r., b.i.d. (maximum dose, 100 mg/day)
	Thioridazine, 10–25 mg p.o., b.i.d. (maximum dose, 800 mg/day)
	Risperidone, 0.5 mg p.o., b.i.d.; increase by 0.5 mg b.i.d. q. 24 hr (maximum dose, 6 mg/day)
	Chlorpromazine, 10–50 mg p.o. or p.r., b.i.d. (maximum dose, 500 mg/day)
	Olanzepine, 2.5–15 mg p.o., q.d.

Terminal Wean

Mechanical ventilation is often tried in patients with respiratory distress, when there is hope that their condition will improve. This is best referred to as a time-limited trial. If reversal of the acute medical condition proves unsuccessful, the physician needs to discuss discontinuance of ventilation with the family.

Terminal ventilation withdrawal should be approached with attention to ensuring the patient's comfort and to enhancing the family's access to the bedside. Miles[60] recommends a 10-step protocol, which applies to unconscious patients dependent on a ventilator:

1. Shut off and remove all monitors and alarms from the patient's room.
2. Remove equipment that impedes access to the patient's hands (e.g., intravenous lines, pulse oximeter, restraints). Hands are for holding.
3. Remove encumbering or disfiguring devices from the bedside.
4. Invite the family to be with the patient.
5. Quietly and personally request that pressors be turned off and that intravenous infusions be set to keep veins open.
6. Watch for distressing symptoms, such as agitation, tachypnea, or seizures; treat appropriately (e.g., with diazepam) if they appear.
7. Turn the fraction of inspired oxygen (F_IO_2) down to 20% and observe the patient for respiratory distress.
8. If the patient appears comfortable, remove the endotracheal tube with a clean towel in hand.
9. Educate and debrief the house staff and nursing staff about the process.
10. Consider contacting the family during the bereavement period, whether by letter or visit.

The goal is for a peaceful, pain-free death for the patient and a supportive, comfortable environment for the family and friends. It is important to warn family that a patient removed from the ventilator may live for hours to days afterward and to reassure them that all measures necessary to ensure comfort during the dying process will be used.

Symptom Management in the Last Hours of Life

The final hours of living can be some of the most important ones for the patient and for family. Managed well, they can lead to a peaceful death and healthy grief and bereavement.[37]

During the final hours, patients usually need skilled care around the clock. Ideally, the environment will allow family and friends both easy access to their loved one and privacy. All who are present should presume that the unconscious patient hears everything.[37]

It is important to be knowledgeable about the normal physiologic changes that occur in the last hours and to educate the patient's family about them. Reassure the family that dehydration in the final hours of living does not cause distress and may stimulate endorphin release that adds to the patient's sense of well-being. Moaning and groaning, although frequently misinterpreted as pain, is often terminal delirium (see above). Decreased hepatic and renal function lead to the accumulation of metabolites, which may cause terminal delirium. Use only essential medications and dose them accordingly.[37]

In the final hours of life, many persons in semiconscious or unconscious states are unable to swallow saliva reflexively or to cough up mucus. This inability to clear secretions from the oropharynx and trachea results in the so-called death rattle—noisy respiration as the secretions move up and down with expiration and inspiration. Explain the reason for the death rattle to the family and administer an anticholinergic drug to reduce pharyngeal secretions (e.g., hyoscine, as a single parenteral dose or by continuous infusion, or scopolamine by patch).[61] At times, it may be necessary to reposition the patient or to suction the airway gently with a soft catheter. Reassure the family that despite the way the breathing sounds, the patient is not uncomfortable.

The removal of the body too soon after death can be even more upsetting to the family than the moment of death, so give the family time with the body.[37] After the patient has died, follow-up with the family is important to ensure that grief and bereavement are progressing normally [*see CE: X Management of Psychosocial Issues in Terminal Illness*].

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. O'Neill B, Fallon M: Principles of palliative care. BMJ 315:801, 1997

2. Ingham J, Portenoy RK: The measurement of pain and other symptoms. Oxford Textbook of Palliative Medicine. Doyle D, Hanks GW, MacDonald N, Eds. Oxford University Press, Oxford, England, 1993, p 202

3. Grossman SA, Sheidler VR, Swedeen K, et al: Correlation of patient and caregiver ratings of cancer pain. J Pain Symptom Manage 6:53, 1991

4. Clipp EC, George LK: Patients with cancer and their spouse caregivers: perceptions of the illness experience. Cancer 69:1074, 1992

5. Bruera E, Kuehn N, Miller M, et al: Symptom Assessment System: a simple method for the assessment of palliative care patients. J Palliative Care 7:6, 1991

6. Donnelly S, Walsh D: The symptoms of advanced cancer. Semin Oncol 22 (suppl 3):67, 1995

7. Portenoy RK , Thaler HT, Kornblith AB, et al: The Memorial Symptom Assessment Scale: an instrument for the evaluation of symptom prevalence, characteristics and distress. Eur J Cancer 30A:1326, 1994

8. Camp L: A comparison of nurses' record assessment of pain with perceptions of pain as described by cancer patients. Cancer Nurs 11:237, 1988

9. Teske K, Daut R, Cleeland C: Relationships between nurses' observations and patients' self-reports of pain. Pain 16:289, 1983

10. Bernabei R, Gambassi G, Lapane K, et al: Management of pain in elderly patients with cancer. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology. JAMA 279:1877, 1998

11. Feldts KS, Ryder MB, Miles S: Treatment of pain in cognitively impaired compared with cognitively intact older patients with hip fracture. J Am Geriatr Soc 46:1069, 1998

12. Morrison RS, Siu AL: A comparison of pain and its treatment in advanced dementia and cognitively intact patients with hip fracture. J Pain Symptom Manage 19:240, 2000

13. Ferrell B, Ferrell B, Rivera L: Pain in cognitively impaired nursing home patients. J Pain Symptom Manage 10:591, 1995

14. Duggleby W, Lander J. Cognitive status and postoperative pain: older adults. J Pain Symptom Manage 9:19, 1994

15. Lynch EP, Lazor MA, Gellis JE, et al: The impact of postoperative pain on the development of postoperative delirium. Anesth Analg 86:781, 1998

16. Super A: Going one step further: skilled pain assessment and the art of adjuvant analgesia. Am J Hosp Palliat Care 14:279, 1997

17. Ahmedzai S: Current Strategies for Pain Control. Ann Oncol 8 (suppl 3):521, 1997

18. Sallerin-Caute B, Lazorthes Y, Deguine O, et al: Does intrathecal morphine in the treatment of cancer pain induce the development of tolerance? Neurosurgery 42:44, 1998

19. Weiss SC, Emanuel LL, Fairclough DL, et al: Understanding the experience of pain in terminally ill patients. Lancet 357:1311, 2001

20. Reuben DB, Mor V: Dyspnea in terminally ill cancer patients. Chest 89:234, 1986

21. Hockely JM, Dunlop R, Davies RJ: Survey of distressing symptoms in dying patients and their families in hospital and their response to a symptom control team. BMJ 296:1715, 1988

22. Ventafridda V, De Conno F, Ripamonti C, et al: Quality-of-life assessment during a palliative care programme. Ann Oncol 1: 415, 1990

23. Carrieri VK, Janson-Bjerklie S: The sensation of dyspnea: a review. Heart Lung 13:436, 1984

24. Bruera E, Macmillan K, Pither J, et al: Effects of morphine on the dyspnea of terminal cancer patients. J Pain Symptom Manage 5:341, 1990

25. Cohen MH, Anderson AJ, Krasnow SH, et al: Continuous intravenous infusion of morphine for severe dyspnea. South Med J 84:229, 1991

26. Bruera E, Neumann CM: Management of specific symptom complexes in patients receiving palliative care. CMAJ 159:1242, 1998

27. Muers MF, Round CE: Palliation of symptoms in non-small cell lung cancer: a study by the Yorkshire Regional Cancer Organization Thoracic Group. Thorax 48:339, 1993

28. Stegman MB: Non-pain symptoms. Pain and Symptom Control in Palliative Medicine. Stegman MB, Ed. Hospice Resources, Fort Myers, Florida, 1997, p 6.1

29. Fuller RW, Jackson DM: Physiology and treatment of cough. Thorax 45:425, 1990

30. Hagen NA: An approach to cough in cancer patients. J Pain Symptom Manage 6:257, 1991

31. Bruera E: ABC of palliative care: anorexia, cachexia, and nutrition. BMJ 315:1219, 1997

32. Bruera E, Miller L, MacCallion J, et al: Cognitive failure in patients with terminal cancer: a prospective study. Proc Am Soc Clin Oncol 9:308, 1990

33. Holden CM: Anorexia in the terminally ill cancer patient: the emotional impact on the patient and the family. Hosp J 7:73, 1991

34. Strang P: Quality of life is the most important goal in nutritional support of the dying. Lakartidningen 97:1141, 2000

35. Bruera E, Macmillan K, Hanson J, et al: A controlled trial of megestrol acetate on appetite, caloric intake, nutritional status, and other symptoms in patients with advanced cancer. Cancer 66:1279, 1990

36. Tchekmedyian NS, Hariri L, Siau J, et al: Megestrol acetate in cancer anorexia and weight loss. Proc Am Soc Clin Oncol 9:336, 1990

37. Emanuel LL, von Gunten CF, Ferris FD: The Education for Physicians on End-of-Life Care (EPEC) Curriculum, Institute for Ethics at the American Medical Association, 1999, p 14

38. Volicer L, et al: Ethical issues in the treatment of advanced Alzheimer dementia: hospice approach. Clinical Management of Alzheimer Disease. Aspen Publishers, Rockville, Maryland, 1988, p 167

39. Finucane TE, Christmas C, Travis K: Tube feedings in patients with advanced dementia: a review of the evidence. JAMA 284:1365, 1999

40. Luchins DJ, Hanrahan P: What is appropriate health care for end-stage dementia? J Am Geriatr Soc 41:25, 1993

41. Reuben DB, Mor V: Nausea and vomiting in terminally ill cancer patients. Arch Intern Med 146:2021,1986

42. Baines MJ: ABC of palliative care: Nausea, vomiting and intestinal obstruction. BMJ 315:1148, 1997

43. Fallon M, O'Neill B: ABC of palliative care: constipation and diarrhea. BMJ 315:1293, 1997

44. Meiring PJ, Joubert G: Constipation in elderly patients attending a polyclinic. S Afr Med J 88:888, 1998

45. Carney MT, Meier DE: Palliative care and end-of-life issues. Anaesthesiol Clin North America 18:183, 2000

46. Ripamonti C: Malignant bowel obstruction in advanced and terminal cancer patients. European Journal of Palliative Care 1:23, 1994

47. Regnard C, Allport S, Stephenson L: ABC of palliative care: mouth care, skin care, and lymphoedema. BMJ 315:1002, 1997

48. Fox PC, Van der Ven PF, Sonies BC, et al: Xerostomia: evaluation of a symptom with increasing significance. J Am Dent Assoc 110:519, 1985

49. Narhi TO, Meurman JH, Ainamo A: Xerostomia and hyposalivation: causes, consequences and treatment in the elderly. Drugs and Aging 15:103,1999

50. Breitbart W, Passik SD: Psychiatric aspects of palliative care. Oxford Textbook of Palliative Medicine. Doyle D, Hanks GW, MacDonald N, Eds. Oxford University Press, Oxford, England, 1993, p 609

51. Barraclough J: ABC of palliative care: depression, anxiety, and confusion. BMJ 315:1365, 1997

52. Massie MJ, Holland J, Glass E: Delirium in terminally ill cancer patients. Am J Psychiatry 140:1048, 1983

53. Leipzig RM, Goodman H, Gray G, et al: Reversible, narcotic-associated mental status impairment in patients with metastatic cancer. Pharmacology 35:47, 1987

54. Lipowski ZJ: Delirium. Acute Confusional States. Oxford University Press, New York, 1990

55. Inouye SK: Prevention of delirium in hospitalized older patients: risk factors and targeted intervention strategies. Ann Med 32:257, 2000

56. Rummans TA, Evans JM, Krahn LE, et al: Delirium in elderly patients: evaluation and management. Mayo Clin Proc 70:989, 1995

57. Pan CX, Meier DE: Clinical aspects of end-of-life care. Annual Review of Gerontology and Geriatrics, Year 2000: Focus on the End-of-life: Scientific and Social Issues, Lawton P, Ed. Springer-Verlag, New York, 2000, p 273

58. Fainsinger RL, Waller A, Bercovici M, et al: A multicentre international study of sedation for uncontrolled symptoms in terminally ill patients. Palliat Med 14:257, 2000

59. Rousseau P: Existential suffering and palliative sedation: a brief commentary with a proposal for clinical guidelines. Am J Hospice Palliat Care 18:299, 2001

60. Miles S: Protocol for rapid withdrawal of ventilator support in anticipation of death. Ethical Currents 45(Spring), 1996

61. Doyle D: Domiciliary Terminal Care. Churchill Livingstone, Edinburgh, Scotland, 1987

XI MANAGEMENT OF PSYCHOSOCIAL ISSUES IN TERMINAL ILLNESS

Jennifer Rhodes-Kropf, m.d.
Ned H. Cassem, m.d.

Like all good medical care, palliative care addresses patients' needs at many levels. The physical deterioration as death approaches can challenge the ingenuity and equanimity of health care professionals [*see CE:IX Palliative Medicine* and *CE:X Symptom Management in Palliative Medicine*]. Yet symptom management is only one aspect of the care that these patients need; the psychosocial and spiritual problems that arise at the end of life also require attention.

A fatal illness—such as untreatable heart disease, terminal cancer, or AIDS—brings with it not only physical pain but also emotional suffering. Fear is prominent. The body, once regarded as a friend, may seem more like a dormant adversary programmed for betrayal. Even innocuous bodily changes may be interpreted as ominous. Patients become fearful before disease symptoms occur. Cancer patients, for example, fear pain, shortness of breath, nausea and vomiting, anorexia, dyspnea, and isolation.[1] Well before the terminal stages of an illness, patients fear the unknown, losing autonomy, disfigurement, dementia, and, last but not least, becoming a burden to their families.

The stress of terminal illness may manifest itself in many ways. Patients may deny or be unable to accept diagnosis or treatment; they may have unrealistic hopes of being cured, and they may persistently ask why there is no improvement. They may express anxiety, often extreme, with near panic and unspecified fears about dying. They may experience intense feelings of ambivalence and guilt regarding their personal relationships.

Considering the severe distress that is frequently involved, it is remarkable how well most patients do cope with a terminal illness and its treatments. The unique set of coping mechanisms that they have used to maintain self-esteem and stability in the past plays a vital role in this process. Religion and spirituality may be most helpful to patients and families at the end of a patient's life and should therefore be recognized and encouraged.

When a patient is dying, the entire family is the appropriate focus of treatment. A family member may be the first to notice a symptom, such as a personality change, in the patient or in another family member and can thus serve as an indispensable historian. The physician must also contend with the psychosocial forces that can lead to family fragmentation and interfere with care. A family member may have more difficulty than the patient in coping with the illness; this may irritate and distract caregivers and ultimately disrupt the relationship between the physician and the family. For example, the physician may be inclined to schedule visits to the patient so as to avoid encounters with the family. This could seriously jeopardize the patient's treatment and the chance to make the patient's death meaningful and dignified. After the patient's death, the survivors may experience abnormal grieving patterns that deserve medical attention.

Most of the psychosocial aspects of palliative care are within the capability of the primary care physician. In some cases, however, psychiatric consultation may be necessary to help the dying patient cope with major depression, personality disorders, continuous treatment-resistant pain, substance abuse, or grieving.

Preliminary Considerations

BREAKING BAD NEWS

Psychosocial care of the patient with terminal illness begins with delivery of the diagnosis. Because there are so many possible reactions a patient may have when informed of the diagnosis, it is helpful to have some plan of action in mind that will permit the greatest range and freedom of response by the patient. Guidelines have been developed for communicating bad news.[2] When the diagnosis is made and it is time to inform the patient, the physician should meet with the patient in a private place. The spouse (and sometimes the family) should be included in the discussion, unless there is a good reason not to do so. If possible, the patient should be informed ahead of time that after all the tests are completed the physician will review the results and discuss treatment plans in detail. With inpatients, the physician should sit down at the bedside to deliver bad news. Standing while conveying bad news may be regarded by a patient as unkind and expressive of wanting to leave as quickly as possible. If the patient is tested as an outpatient and returns home before the results are known, he or she should be told that the diagnostic information is too important to convey by phone, and a meeting to discuss the results should be arranged. Relaying bad news by phone may be perceived by patients as thoughtless, even though they may have asked for information. A physician must also be prepared to respond to a patient who wishes no or minimal information about the diagnosis.

When the findings are life threatening (e.g., a biopsy positive for malignancy), how can the news best be conveyed? A good opening statement is one that is (1) rehearsed so that it can be delivered calmly, (2) brief (three sentences or less), (3) designed to encourage further dialogue, and (4) reassuring of continued attention and care. A typical delivery might go as follows: "The tests confirmed that your tumor is malignant. I have therefore asked the surgeon (or radiotherapist or oncologist) to speak with you, examine you, and make recommendations for treatment. After this, we can discuss how we should proceed." Silence and quiet observation at this point will yield more valuable information about the patient than any other part of the exchange. What are the emotional reactions? What sort of coping is seen at the very start? While observing, one can decide how best to continue with the discussion. Just sitting with the patient for a period of time, however, is the most important part of this initial encounter with a grim reality that both patient and physician will continue to confront together, possibly for a long time.

TELLING THE TRUTH

Given the difficulties that can follow the disclosure of a life-threatening illness, it may be tempting to avoid telling the patient the diagnosis. This tactic has ancient roots: Hippocrates himself recommended concealing bad news from patients, lest they become discouraged.[3] Nevertheless, most empirical studies in which patients were asked whether they wanted to be told the truth about malignancy have indicated an overwhelming desire in the affirmative. Of 740 patients in a cancer-detection clinic who were asked before diagnosis whether they wanted to be told their diagnosis, 99% said that they did.[4] Another group of

patients in this clinic were asked the same question after the diagnosis was established, and 89% of them replied affirmatively, as did 82% of patients in still another group who had been examined and found to be free of malignancy.

Truth telling entails eliciting the patient's concerns. Studies indicate the clinician's use of specific communication skills enhances a patient's disclosure of his or her concerns.[5] Such communications skills include making eye contact with the patient, asking open-ended questions, responding to the patient's affect, and demonstrating empathy. The desire for truth telling may vary among different ethnic groups.[6] In a study of elderly persons in the United States, Korean Americans and Mexican Americans were less likely than African Americans and European Americans to believe that a patient should be told the diagnosis of metastatic cancer.[7] However, a population study in Hong Kong reported that the majority of persons canvassed would want to know if they had terminal cancer; this finding is at odds with the cultural preference of many Chinese, who usually prefer to withhold diagnostic information from terminally ill family members.[8] This study emphasizes that cultural preferences give only general indications of a patient's readiness to hear bad news; truth telling ultimately depends on the physician's assessment of what the patient wants to know and is prepared to know about the diagnosis. Socioeconomic factors may also be involved in a patient's willingness to hear bad news: younger age and higher income and education make patients more likely to want detailed diagnostic information.[9]

Is the truth harmful? Gerle and colleagues[10] studied 101 patients who were divided into two groups, with one group, along with their families, being told the frank truth of their diagnoses and the other group being excluded from discussion of the diagnosis (although the patients' families were informed). Initially, there appeared to be greater emotional upset in the group of patients and families who were informed together. The investigators observed in follow-up, however, that the emotional difficulties of the families of the patients who were shielded from the truth far outweighed those of the patients and families that were told the diagnosis simultaneously. In general, empirical studies support the idea that the truth about the diagnosis is desired by terminally ill patients and does not harm those to whom it is told. Honesty sustains the relationship with a dying person rather than jeopardizing it.[11] Individual variations in willingness to hear the initial diagnosis are extreme, however, and diagnosis is entirely different from prognosis. Many patients have said that they were grateful to their physician for telling them they had a malignancy. Very few, however, reacted positively to being told that they were dying. In our experience, "Do I have cancer?" is a common question, whereas "Am I dying?" is a rare one. The question about dying is more commonly heard from patients who are dying rapidly, such as those in cardiogenic shock.

Honest communication of the diagnosis by no means precludes later avoidance or even denial of the truth of the diagnosis. In two studies, patients who had been explicitly told their diagnosis (using the words cancer or malignancy) were asked 3 weeks later what they had been told: in both studies, about 20% of the patients sampled denied that their condition was cancerous or malignant.[12,13] Croog and colleagues[14] interviewed 345 men 3 weeks after myocardial infarction; 20% of those patients said they had not had a heart attack. All had been explicitly told their diagnosis. For a person to function effectively, truth's piercing voice may be muted or even excluded from awareness. Denial can reduce psychological distress, and preliminary evidence suggests that in women with nonmetastatic breast cancer, it may be associated with prolonged survival.[15] However, information-seeking behavior and fighting spirit have been more consistently associated with higher rates of recurrence-free survival 5 and 10 years after diagnosis of breast cancer.[16,17] Communicating a diagnosis honestly, though difficult, is easier than the labors that lie ahead. Telling the truth is merely a way to begin, but it provides a firm basis on which to build a relationship of trust.

COMMUNICATING WITH THE PATIENT

The most important component of communication is listening. The real issue is not what you tell your patients but, rather, what you let your patients tell you.[1] Most people are afraid to let dying patients say what is on their minds. If a patient who is presumed to be 3 months from death says, "My plan was to buy a new car in 6 months, but I guess I won't have to worry about that now," a poor listener might say nothing or, "Right. Don't worry about it." A better listener might ask, "What kind of car were you thinking about?" In a study of 126 patients with incurable cancer, 98% wanted their physicians to be realistic about their health status, to provide an opportunity to ask questions, and to acknowledge them as an individual when discussing prognosis.[18]

It is essential to get to know the patient as a person. The best way to recognize and acknowledge the person's worth is to learn those features of his or her history and nature that make him or her unique. Encourage dying persons to tell their stories. Learn about significant areas of the patient's life—such as family, work, or school—and chat about common interests. This is the most natural way to give the patient the sense that she or he is known and appreciated.

Patients occasionally complain about professionals and visitors who regard them as "the dying patient," not as a unique person. The physician can help dissolve communication barriers for staff members by showing them the remarkable qualities of each patient. Comments such as, "She has 34 grandchildren," or, "This woman was an Olympic sprinter," convey information that helps other members of the health care team appreciate their patient and to find something to talk about with them. Listen for the patient's own conversational cues whenever possible. Awkwardness subsides when a patient is appreciated as a real person and not merely "a breast cancer patient." This rescue from anonymity is essential to prevent a sense of isolation. The most important communication is often not verbal. A pat on the arm, a wave, a wink, or a grin communicates important reassurances. Back rubs and physical examinations can also be an opportunity to convey reassurance.

Psychosocial Support during Terminal Illness

The diagnosis of a terminal illness impacts the patient's relationships with family, friends, and coworkers and can thus undermine the patient's sense of self. Although death is a natural part of life, the adjustment to diminished function and role in relationships can be stressful, both for the patient and the patient's family.

FAMILY AND FRIENDS

Family members and friends must be helped to support the patient and one another. To provide this help, the physician must get to know both the patient and the family members. When patients are permitted to give support to their families, they often feel they are less of a burden.

One must appreciate the fact that for family and close friends, a fatal illness of a loved one may be the only event important enough to resolve long-standing conflicts. Peacemaking should be a priority. Specific plans for the family are important. The writing of wills, the clarification of family history, the review of memorable family gatherings and achievements, the carrying out of such family projects as trips or photo-album reviews, and planning a funeral or memorial service are all important activities.

The care of a dying person can be a process of mutual growth for the patient and the family. Just as the deterioration of a person with a fatal disease can be threatening (family members may feel both horrified at the prospect of the same thing happening to them and helpless to assist), the response of the dying person to the challenge may be not only edifying but also an invaluable lesson in coping. Indeed, family members who act as caregivers report strong positive emotions regarding the opportunity to express their love through care. Those caregivers may experience extreme grief after the patient's death, however, and may require special support and attention from health care providers at that time.[19]

Near the end of life, a patient may be too weak to communicate by speech, and sometimes, consciousness itself may be difficult to assess. Most patients who have lost the ability to communicate have a period when they can still hear or perceive those in attendance. Family feelings of helplessness can be minimized by reading especially meaningful passages to the dying person (e.g., the daily headlines, articles by favorite authors or columnists, poetry, passages from the Bible, the Dow-Jones average, sports scores, and letters new and old). Conversations should make natural reference to the person as though hearing and understanding were intact. Singing favorite songs, playing favorite music, or praying aloud may increase the sense of unity and purpose for the family. Although the patient may never be able to tell us how important that time is, an occasional incident will do so dramatically, as when a supposedly unconscious person suddenly smiles appropriately, gestures, or even speaks. Often, this conveys gratitude for the attention given to him or her, which is very rewarding for the loved ones in attendance.

The end of life is an opportunity to educate the younger generation. Whenever possible, children should be included in all the planning, meetings, discussions, activities, and care, as well as the final attendance at death. Children can learn that death need not be violent or terrifying and that we face our losses best when we face them together.

Investigators have consistently learned that the visits of children are as likely as any other intervention to bring consolation and relief to the terminally ill patient. How can one determine whether a particular child should visit a dying patient? No better approach has been found than asking the child directly whether he or she wants to visit.

Ideally, this mutual work at the end of life will confirm the dying person's sense of self. It also can give family, friends, and caregivers the wonderful feeling that they have provided good care and safe passage.

OCCUPATION AND WORK

Work is critical for the self-esteem of many people. Many people begin to feel less valuable when work ceases or they retire, and the approaching end of life may intensify a sense of failure. The continuation of work for as long as feasible, as well as continued contact with work colleagues, can remind the dying of who they are and what they have accomplished. It encourages the belief that they are remembered and respected, regardless of their illness. Similarly, continued involvement in recreational activities can be very satisfying.

Near the end of life, a person is often too disabled to get around or to contact colleagues and friends. Wherever possible, such contact should be arranged for and encouraged.

RELIGION AND SPIRITUALITY

Studies find that people who have a strong internalized faith possess a resource that helps significantly in coping with a fatal illness.[1,20-23] It is a well-documented finding that religious persons usually belong to a community that can be unusually thoughtful and generous in providing support. However, the community may not know of the patient's plight and may need to be contacted. Thus, the appreciation of a person's religion or spirituality is extremely important.

The Physician's Psychological Role in Patient Care

The first responsibility of the caregiver, as Saunders[24] points out, is "above all to listen." A suffering person often wants to communicate just how awful a fatal illness is. Words from the caregiver may be irrelevant: "When no answers exist," says Saunders, "one can offer silent attention."

It is important to be aware of the impact that patients' feelings can have on one's own mood and the amount of time spent with the patient. The relentless approach of death from cancer or AIDS may leave a patient with feelings of terror, hopelessness, and despair. Those feelings tend to be contagious, intensifying our feelings of impotence. A caregiver's own helplessness and despair may result in neglect or avoidance of the patient or feelings that the patient would be better off dead. Sensing that a patient is burdensome to caregivers can be devastating to the patient who looks to a doctor or nurse for some sense of hope.[25] In one study of terminal cancer patients, the majority of patients (87%) indicated that seeing their physician nervous or uncomfortable in their presence did not promote hope.[16] Thus, it is of the utmost importance that caregivers remain empathetic and reassure patients that they will continue to be there for them at all stages of their illness and that caregivers learn to live with negative feelings and resist the urge to avoid certain patients—attitudes that could convey that care of a patient is difficult for us or that the patient no longer matters to us.

COMPASSION

Of all attributes in physicians and nurses, none is more highly valued by terminally ill patients than compassion. Although they may never convey it precisely by words, some physicians and nurses are able to tell the patient that they are genuinely touched by his or her predicament. Although universally praised as a quality for a health professional, compassion exacts a cost that is usually overlooked in professional training. This cost is conveyed by its two Latin roots: *com*, meaning with, and *passio*, from *pati*, meaning to suffer—that is, to suffer with another person. It is important for caregivers to have a source of support for themselves—such as colleagues, friends, and family—so that they can continue to be there for their patients.[26]

CHEERFULNESS

The possessor of a gentle and appropriate sense of humor can bring relief to all parties involved. Often, patients provide this, and their wit may soften many a difficult incident. Humor needs

to be used sensitively, however: forced or inappropriate mirth with a sick person can increase feelings of distance and isolation.

CONSISTENCY AND PERSEVERANCE

Dying patients have a realistic fear of progressive isolation. The physician or nurse who regularly visits the sickroom provides tangible proof of continued support and concern. A brief visit is far better than no visit at all; the patient may not even be able to tolerate a prolonged visit. Do continue to visit: patients are quick to identify those who show interest at first but gradually disappear from the scene. Stay the course even if this means that you must listen to repeated or irrelevant complaints.

The Patient's Psychological Response to Terminal Illness

Any serious illness inflicts some loss on the patient. A diagnosis itself, with no change in subjective symptoms, can cause a feeling of loss, as concepts of self and plans for the future are swept away.

The emotional reactions to a myocardial infarction serve as a model for the reactions of a person who has experienced physical loss. In a series of 149 coronary patients whose emotional difficulties were severe enough to warrant psychiatric consultation, the majority of problems during the first 2 days stemmed from fear and anxiety. These patients generally showed a sequence of emotions beginning with anxiety, followed shortly thereafter by denial (at this stage, a few wanted to sign out of the hospital) and then by despondency, which sometimes persisted. A final group of management problems, related mostly to dependency or personality disorders, rounded out the sequence.[27]

In essence, this reaction pattern suggests that the most common difficulty for a patient immediately after admission is fear. The patient fears imminent death, the presence or return of pain or breathlessness, or some vague but ominous threat to well-being. As symptoms stabilize or subside, the patient is likely to imagine that admission symptoms were false alarms and, in some cases, to insist on signing out of the hospital. When the diagnostic tests confirm the presence of myocardial infarction, however, the patient is confronted with the reality of the illness and feels demoralized. As hospitalization continues, any personality flaws (e.g., passive aggression) further complicate interactions between the patient and the hospital personnel. The sequence of acute onset of illness, fear, stabilization, denial, confirmation of illness, and depression provides a convenient framework for assessing the mental state of an individual hospitalized for any serious illness.

FEAR

Anxiety and despondency are the most common emotional reactions to illness. Panic distorts personality as nothing else does. Yet fear assumes many guises. If a patient seems impossible to deal with on the first day of hospitalization, the reason very likely is underlying fright. However, if difficult behavior continues after 4 or 5 days in the absence of new events that are frightening, it is probably because of the patient's personal style. Excessive talkativeness or mute withdrawal is a typical sign of fear in the acute phase.

Medication, quiet reassurance, or both can relieve a patient's fear and anxiety. Minor tranquilizers are the agents most commonly employed, but explanation and reassurance can be even more effective than medication.

When the physician senses that the patient is afraid, it is safe to assume that the patient regards the illness as an overwhelming threat to well-being. This threat is based on what the patient already knows or presumes about the disease. The physician, therefore, may ask questions designed to uncover erroneous concepts about the patient's condition, such as, "Have you ever known anyone with this disease?" or, "What is your notion of this disease?" If any family member has died of the disease, his or her age at death may also contribute heavily to the patient's fear of the same fate.

After false notions have been corrected, it is important to emphasize the positive aspects of the treatment plan. Even when the prognosis is grave, a calm statement of the treatments planned to counteract and contain the disorder is of value to the anxious patient. The more ominous the prognosis, the more important it is to encourage the patient to specify the fear, so that correspondingly true reassurances (e.g., "the medication can control pain") can be given. False comfort is not recommended. It robs the physician of credibility and, therefore, of the ability to reassure the patient as the illness progresses. An empathetic yet silent presence can sometimes be more helpful than well-meant counsel.[28]

DENIAL AND PANIC

Denial is a common defense mechanism in the initial stage of life-threatening illness. The ability to minimize or to completely deny the threatening implications of the disease ("There's nothing to worry about; I'll be all right.") is essential for controlling panic. When panic sets in, denial fails and people want to flee. Panic is the most common reason why acutely sick patients sign out of a hospital. Although it may simply mean that the patient does not take his or her illness seriously enough, the threat to sign out should be considered a panic reaction. The patient's panic conviction is, "I'll die if I don't get out of here."

Because patients who are experiencing a panic reaction are feeling desperate, they may become antagonistic to efforts to detain them. A gentle approach is essential. For example, the doctor, seated if possible, may begin with "Mr. Jordan, I'm not here to force you to do anything; I just ask that you hear me out." Then the patient needs to hear the truth—that he is seriously ill—expressed in direct but reassuring terms. To quiet the panic, it is most important to explain that the illness is manageable. As the patient calms down, other questions designed to reduce fear can be asked. Even if calmed, however, an anxious patient will not remain calm for long and should be promptly medicated. Family members should also be mobilized and informed.

ANXIETY

Anxiety disorders may or may not intensify during a terminal illness, but they clearly require psychiatric attention when they do. The four most common anxiety-provoking fears associated with death are (1) helplessness or loss of control, (2) being considered bad (guilt and punishment), (3) physical injury or symbolic injury (castration), and (4) abandonment.[29]

In the clinical examination, a severely anxious patient usually does not know what it is about death that is so frightening. Increased anxiety may result from specific memories and associations related to the death of parents or others with whom one identifies; patients may picture the same fate for themselves (e.g., agonizing pain or excessive use of technology). Memories of someone who died of the same illness (e.g., a woman with breast cancer who had relatives who died of breast cancer or a patient

with AIDS who tended to a lover dying of AIDS) or particular associations with the illness may produce specific reasons for anxiety (e.g., the disease will be disfiguring). For the sake of the patient's mental health, it is important to explore these issues.

Pharmacologic therapy with an antidepressant such as a selective serotonin reuptake inhibitor (SSRI) or benzodiazepine may be warranted for patients with moderate levels of anxiety. The effect of benzodiazepine is immediate, whereas the therapeutic effect of an SSRI is delayed, occurring 2 weeks after initiation of treatment. Benzodiazepines (e.g., lorazepam and clonazepam) are therefore the preferred class of drug. Lorazepam (0.5 to 2 mg, given orally two to four times a day) peaks in 1 to 6 hours and is available in tablet, elixir, sublingual, rectal, or I.V. formulations. Clonazepam (0.25 to 0.5 mg, given orally two to four times a day) has a slightly longer half-life; it is available in tablet formulation. Patients who have anxiety in the setting of dyspnea respond well to morphine [*see CE:X Symptom Management in Palliative Medicine*].[30] If psychotic symptoms accompany anxiety, a neuroleptic agent such as haloperidol (0.5 to 5 mg, given orally, I.V., or S.C. every 2 to 12 hours) is helpful.[31]

DESPONDENCY

Despondency—a mixture of dread, bitterness, and despair—is the result of an attack on the patient's self-image. The patient feels broken, scarred, and ruined. Work and personal relationships appear jeopardized. It may seem too late to realize cherished goals. The patient is haunted by disappointment with both what has been done and what has been missed. He or she may feel old and that life has been a failure.

Despondency is a contagious feeling, and in most cases, the physician can sense that the patient is depressed. Simply asking about the depression is helpful: "You look a bit blue today. What's on your mind?" The patient is likely to respond with the feelings already described. The patient should be told that such feelings are a normal part of any serious illness. It is important to remind even those who deny despondency that there is nothing unusual about feeling low from time to time in the struggle with any illness and that these feelings are time-limited. When the patient has acknowledged feelings of depression, even in the first few days of illness, it is very helpful if the physician describes future plans for medical treatment.

DEPRESSION

The more seriously ill a patient becomes, the more likely it is that a major depression will develop.[32] In a review of the literature, major depression was reported to affect as many as 29% of palliative care patients[33]; however, this figure may be low. Researchers identified depression in 62% of patients in a palliative care unit in Winnipeg, Canada.[34] Standard depression inventories (e.g., Beck) are not as useful for diagnosing depression in terminal patients, because some of the physical symptoms of depression that these inventories target can occur in terminal illness without depression. At present, there is no validated instrument to assess depression in patients with terminal illness, although research is under way. Emotional symptoms remain helpful, however.[35] These include anhedonia, depressed mood, suicidal thoughts, and guilt.

Patients in pain have a significantly higher rate of depression than comparable patients without pain.[36] Extreme depression and hopelessness are the strongest predictors that patients may develop a desire for hastened death.[29] Ganzini and colleagues[37] documented that severely depressed patients make more restricted advance directives when depressed and change them when the depression is in remission.

Dignity therapy, a novel psychotherapeutic intervention, may hold promise as a treatment for depression in palliative care patients. As part of the intervention, patients are asked to discuss issues that matter most to them or that they most want remembered by their families; these discussions are recorded and transcribed for the patients and their families. In a multicenter study, the majority of patients who received dignity therapy reported a heightened sense of dignity (76%) and an increased sense of purpose (68%) after the intervention.[38] These results suggest dignity therapy may be useful in treating depression and distress common in palliative care patients.

Pharmacologic options for depression in palliative care extend beyond the traditional agents [*see Table 1*]. Because standard antidepressant medications typically require several weeks to take effect, psychostimulants such as methylphenidate (Ritalin) and pemolin (Cylert) are increasingly being used for short-term treatment of depression for terminally ill patients in pain. They may be used instead of traditional antidepressants, in patients whose life expectancy is less than 3 weeks, or as an interim measure until traditional antidepressants take effect.[39] They

Table 1 Antidepressant Medications Used in Patients with Advanced Disease[37,56]

Class	*Agent (Trade Name)*	*Dosage*
Tricyclic antidepressants	Amitriptyline (Elavil)	10–150 mg p.o./I.M./p.r., q.d.
	Clomipramine (Anafranil)	10–150 mg p.o., q.d.
	Desipramine (Norpramin)	12.5–150 mg p.o./I.M. q.d.
	Doxepin (Sinequan)	12.5–150 mg p.o./I.M. q.d.
	Imipramine (Tofranil)	12.5–150 mg p.o./I.M. q.d.
	Nortriptyline (Pamelor)	10–125 mg p.o., q.d.
Second-generation antidepressants	Bupropion (Wellbutrin)	200–450 mg p.o., q.d.
	Citalopram (Celexa)	10–60 mg p.o., q.d.
	Fluoxetine (Prozac)	10–60 mg p.o., q.d.
	Fluvoxamine (Luvox)	50–300 mg p.o., q.d.
	Mirtazepine (Remeron)	15–45 mg p.o., q.d.
	Paroxetine (Paxil)	10–60 mg p.o., q.d.
	Sertraline (Zoloft)	25–200 mg p.o., q.d.
	Trazodone (Desyrel)	25–300 mg p.o., q.d.
	Venlafaxine (Effexor)	37.5–225 mg p.o., q.d.
Psychostimulants	Dextroamphetamine (Dexedrine)	2.5–20 mg p.o. in the morning and at noon*
	Lithium carbonate	600–1,200 mg p.o., q.d.
	Methylphenidate (Ritalin)	2.5–20 mg p.o. in the morning and at noon*
	Pemoline (Cylert)	37.5–75 mg p.o. in the morning and at noon*

*Give last dose at noon to avoid insomnia at night.

are also useful to counteract opiate-induced sedation and may potentiate opiate analgesia.

Of the antidepressant agents, the selective serotonin reuptake inhibitors (SSRIs) are associated with fewer side effects than traditional tricyclic agents, which are associated with a high incidence of anticholinergic toxicity, including constipation, urinary retention, confusion, and altered cardiac conduction. The SSRIs (fluoxetine, sertraline, and paroxetine) are effective antidepressants and are generally well tolerated. Major side effects include anorexia, nausea, restlessness, and insomnia. Antidepressants with demonstrated efficacy as adjuvant therapy for treatment of pain include the tricyclic antidepressants and paroxetine.

PERSONALITY DISORDERS

Seriously ill people share common objectives with their physician: the relief of suffering and, as far as possible, the restoration of health. Dysfunctional personality traits (e.g., passive, hysterical, obsessive, dependent) that are the residue of past problems, such as parental conflicts, can distract both patient and doctor from those shared objectives. The doctor has enough to do to care for the physical illness and its normal emotional consequences to the patient (e.g., fear, anger, or despondency) without trying to alter personality traits. If reasonable efforts do not suffice, further intervention is best left to a consulting psychiatrist.

Preparation for the End of Life

THE CHOICE OF WHERE TO DIE

Where a person wishes to spend the end of his or her life is a very personal decision. The options are to remain at home, to move to an inpatient hospice [*see CE:IX Palliative Medicine*], or to die in the hospital. Factors that influence this decision include the degree of support at home to care for the patient (emotional and physical), how comfortable the caretakers are with the care of a person who is dying, financial resources, and the technical support needed to keep the patient comfortable. In most cases, special equipment and services can be set up in the home, but this can be prohibitively expensive.

If it is anticipated that the patient has less than 6 months before death, this is an appropriate time to discuss hospice, whether inpatient or at home [*see CE:IX Palliative Medicine*].

Health care providers frequently overlook the financial burden for patients and families resulting from terminal illness. Financial costs can be devastating. It is important to address this issue with patients and families and to refer them to appropriate financial counseling. Social workers can provide invaluable assistance in facilitating the provision of the home services to which patients are entitled under Medicare or Medicaid, and they can usually tell the patient and family what services will have to be paid for out of pocket.

Remember that while patients are in the hospital, caregivers surround them. When they return home, they often feel isolated and abandoned. Every effort should be made to maintain channels of communication among patients, family, and home health care workers.

ADVANCE DIRECTIVES

It is a mistake to delay the discussion of advance directives until the patient is in the terminal stages of illness. Rather, this issue should be dealt with soon after the diagnosis of terminal illness [*see CE:IX Palliative Medicine*].

FINAL CLOSURE

The end of life is the opportunity for closure in relationships with loved ones. Relationship completion comprises five types of communications: I forgive you; forgive me; thank you; I love you; and good-bye.[40] These messages are vital to the peace of mind of the patient and the patient's family and should be encouraged by the physician as an aspect of standard palliative care.

Other actions that help with life's closure are a discussion of personal preferences for a memorial service, the settlement of financial affairs, and, if applicable, the completion of a plan for care of the children.

The physician should instruct the family in practical considerations concerning their loved one's death. For instance, the family should be told that there is no need to call 911 when the patient dies; instead, they should contact the funeral director. If a patient is dying at home and the family panics and calls 911, it is important that they have a "Do Not Resuscitate (DNR)" form in the home. Otherwise, the emergency medical services in some states are required to automatically intubate the patient.

Grief and Bereavement

In one respect, life can be described as one loss after another. The degree of recovery from each loss determines whether an individual regains a stable life or remains disabled. When losses occur, the resulting sadness can eventually give way to a process of reorganization that restores the person's ability to function normally. For example, the death of a parent can cause a child to become self-reliant. Some persons maintain a satisfying, productive life despite seemingly overwhelming losses, whereas others never recover from less severe losses. What makes the difference?

NORMAL GRIEVING

Grief is the psychological process by which an individual copes with loss, struggles to understand it, regains perspective, and goes on with life. Causes of grief include not only the loss of a loved one, of valued possessions, or of employment but also the loss of good health that occurs with major illness or injury. Serious illness or injury challenges personal integrity; it could be said, for example, that every myocardial infarction causes an ego infarction. Therefore, recovery from major illness is not complete until the patient has also recovered from the accompanying emotional damage to the self.

Surrounded daily by the sick and injured, physicians see grief-work in process. It is important for the physician to realize that grief is a normal reaction serving an important restitutive function, that it follows a typical pattern, and that marked deviation from this pattern may be a sign that psychological intervention is required.

The normal grieving process follows a similar course in individuals suffering from any serious loss. Several prominent features of normal grieving have been identified.[41,42] Because these features are often mistakenly labeled as pathologic, familiarity with their correlation to grief can prevent well-meaning but misguided efforts to intervene in a necessary process.

Somatic symptoms of grieving may be prominent, including sighing respirations, exhaustion, gastrointestinal symptoms of all kinds, restlessness, yawning, and choking. Feelings of guilt, especially early in the wake of loss, seem to be universal. "What more could I have done?" or other references to unresolved emotional conflicts are common expressions of these feelings.

Preoccupation with the image of the deceased person, often seeming bizarre even to the griever, is a sure sign that normal mourning is under way. The intense focus on the deceased may be manifested in several ways: by continual mental conversations with the dead person; by a sense of the dead person's presence so vivid, especially at night, that the griever hears, sees, or is touched by the person; or by the simultaneous feeling that all other persons are emotionally distant.

Hostile reactions and irritability also seem to be the rule, combined with a disconcerting loss of warm feelings toward others. Some disruption of normal patterns of conduct is present, such as a desire to be alone, uncharacteristic procrastination, and indecisiveness toward others. The style, traits, mannerisms, or even the physical symptoms of the dead person may alarmingly appear in the mourner; such identification phenomena signify only that grief is in process. Finally, it is routine for the griever to feel that part of the self has been destroyed or mutilated.

How long will it take for the acute symptomatology of grieving to subside? Although the usual estimate is 1 to 3 months, many factors affect the actual time required. They include the number of strong remaining relationships, the intensity and duration of the bond with the lost person, the number and severity of any unresolved conflicts, the degree of dependence on the lost person, and how much of the survivor's mental life habitually assumed the dead person's physical or emotional presence. The main signs of resolution of acute grief are the reappearance of normal functioning, the capacity to experience pleasure, and the ability to enter new relationships.

The acute phase is followed by the disorganization phase. In this phase, the pain of the experience becomes foremost in the person's consciousness. Turmoil, emptiness, despair, and thoughts about the pointlessness of life and the reasonableness of suicide are common. Social interaction seems impossible and is avoided, even though solitude itself is dreaded and intolerable.

Finally, there is reorganization, characterized by a return of normal functioning and behavior. Reversals during this time are the rule, and reappearance of the earlier two phases should be expected. The bereaved person is caught off guard by sudden reminders of the lost person (e.g., a special coat discovered in storage) or by new and painful realizations (e.g., no more shared holidays) that reopen the wound of loss.

The grieving process is often delayed when death follows a prolonged and difficult illness. In such circumstances, death is entirely acceptable, even welcomed as the end of suffering. Later, especially when returning to a scene that sharply evokes the memory of the dead person when healthy, death becomes unacceptable, and feelings of protest or resentment spontaneously emerge.

ABNORMAL GRIEVING

Preexisting personality traits in survivors can interfere with the normal grief process. Additionally, survivors are at heightened risk of abnormal or complicated bereavement if the loved one died suddenly or unexpectedly, if the death was violent, or if no bodily remains were found. Because grief serves an important restitutive function, failure to grieve normally may result in serious psychological symptoms.

Some markers of abnormal grief are evident immediately; others do not appear for 3 months or longer after the loss. An inability to grieve immediately after the loss, typically manifested by absence of weeping, is the best predictor of later problems. Prolonged hysterical grieving that is defined as excessive by the individual's own subcultural norms (not those of the physician) is an equally ominous prognostic sign. Overactivity without a sense of loss is an early sign of distorted grieving. Furious hostility against specific persons—for example, the doctor or hospital staff—which may assume true paranoid proportions, can be regarded as a sign of abnormality when the individual dwells on it to the exclusion of the other concerns of normal grief. A suppression of hostility to the degree that the person's affect and conduct appear frozen (masklike appearance, stilted robotlike movements, and no emotional expressiveness) and self-destructive behavior (giving away belongings, foolish business deals, or other self-punitive actions with no attendant guilt feelings) are also early indicators of abnormal grieving.

Ultimately, it may become apparent that social isolation has become progressive, with a lasting loss of interpersonal initiative. When symptoms of the deceased person appear in the survivor as conversion symptoms or have become the focus of hypochondriacal complaints overshadowing all other manifestations of grief, pathologic grief is likely. Unresolved grief can also be suspected when the dead person is portrayed either as a saint who had no shortcomings or as one who never occasioned the least feeling of anger, burden, or disagreement in the survivor. In such cases, the mourner usually harbors intense feelings that are in conflict with those feelings outwardly expressed, and fear that these feelings will be discovered immobilizes the grieving process.

The result of prolonged grieving may be prolonged sadness, social isolation, somatic complaints, or loss of ability to function. A few sessions with a psychiatrist, aimed at helping the patient bring his or her own feelings into the open so that the process of grieving can be completed, often provide great relief.

Helping the Bereaved

Mourners tend to be outcasts from society. Their presence is painful to many around them, and efforts to silence, impede, or stop the manifestations of their grief are common. Allowing the grieving person to express feelings is essential, however. Most important is avoidance of maneuvers that negate grieving, such as clichés ("It's God's will"), efforts to distract ("After all, you've got three other children"), and outright exhortations to stop grieving ("Cheer up, life must go on").

Seeing the body of the deceased facilitates grieving, probably by establishing the irrevocable fact of death.[39] Permitting survivors to express their feelings and reminders that grief is a normal process are helpful. Gentle review of the deceased person's last days of life, last conversations, and final exchange of words, as well as talking about the deceased's general lifestyle, help initiate grieving. The memories most obstructive of grieving are those of hostile interactions with the deceased and any other interactions that leave the survivor feeling guilty. The more negative these interactions were, the longer it takes to begin recalling and discussing them.

In helping the bereaved, presence means more than words. Someone who can remain calm and accepting in the presence of a weeping, angry, or bitter mourner is highly valued. A hand on the shoulder can be just what is needed. Over time, helping the griever complete memories of the deceased also facilitates mourning. Old photograph albums and letters can be helpful in this regard. Anniversaries are key points in the grieving process, and special attention to the bereaved on these days is a basic element in the care of mourners.

A return to a job is an essential feature of the recovery process because it brings the mourner back into contact with concerned fellow workers. In addition, the therapeutic effects of work on self-esteem play an important part in alleviating the narcissistic component of the response to the loss. Most bereaved persons benefit from returning to work within 2 to 4 weeks after the death of a loved one.

Self-help groups can be extremely effective for permitting expression of emotion, showing that grief is universal, and supplying the compassion and respect necessary for rebuilding self-esteem. Books that recount events such as losing a spouse or that give instructions for the surviving spouse and children may be helpful.

SPECIFIC TYPES OF LOSS

Each type of loss carries specific challenges to mourners, and each type has a specific literature that can be helpful.[43] Loss of a parent by an adult, although a nearly universal occurrence, is not trivial, and loss of the second parent may leave the bereaved feeling particularly alone and vulnerable. Loss of a parent by a child invariably worries the adult survivors responsible for the child's care because successful mourning in a child is a more complex process than in an adult.[44,45] For example, the child may face adjustment to parental surrogates, to a parent stressed by the responsibility of raising the child alone, to the loss of a gender role model, or, eventually, to the replacement of the deceased parent by remarriage and competition for the affection of the surviving parent. However, studies of bereaved children from stable families have shown optimistic results: 8 weeks after the death of a parent, children 5 to 12 years of age were similar to nonbereaved children in school behavior, interest in school, peer involvement, peer enjoyment, and self-esteem.[46]

Research on loss of a sibling appears to be lacking, but the available data indicate that death of a sibling forces surviving siblings to reorganize their roles and relationships with their parents and with one another.[43]

Loss of a spouse, ranked on life-event scales as the most stressful of all possible losses,[47] is more detrimental for men than for women and leads to increased morbidity and mortality in elderly men.[48] The bereaved spouse is left with sole responsibility for children, finances, management, and planning; faced with possible loss of income; and forced to cope with a changed social role in the community.

Each year, about 800,000 parents lose a child younger than 25 years. This loss is particularly traumatic because it is so contrary to life-cycle expectancies.[49]

Sudden death, such as death in an emergency ward, stillbirth, sudden infant death, accidental or traumatic death, cardiac arrest, or death during or after surgery, inflicts a uniquely intense trauma on the survivors. Shock is dramatically intensified. Guilt is likely to be a much more serious problem than it is with nonsudden death because of the total absence of preparation. Violence or disfigurement further intensifies the survivor's feelings.

General rules for dealing with the bereaved also apply here, with certain specific emphases. The chance to view the body, even when mutilated, should be offered to the family members. If there is severe mutilation, the family should be warned. The need to view the body, an aid to normal mourning, is greater when death is sudden.

Suicide is an especially difficult way to lose a loved one. Feeling abandoned and rejected, the survivor often experiences unsettling anger or, if the relationship had been hostile and stormy, equally unsettling relief. The bereaved scours through memories for an action that might have caused or prevented the suicide. Guilt is such an inevitable consequence of suicide that even casual acquaintances wonder what they might have done that contributed to the death. Shame can cause avoidance of others, falsification of the event as an accident, or unwillingness to let others know that a family member has died. A scapegoat may be sought, such as the deceased's therapist, spouse, or boss or the medical examiner who labeled the death a suicide.[50]

Loss by homicide also produces especially intense grief reactions.[51] Flashbacks of the violent death are unavoidable. Survivors tend to avoid locations associated with the death and to stop watching television news because of possible reports of violence. Rage and desire for proportional revenge may cause intense discomfort for the bereaved, if suppressed, or for those around the bereaved, if excessively expressed. Children who witness the murder of one parent by another are afflicted with traumatic intrusive memories of the parents, massive conflicts of loyalty, and the intense need for secrecy because of the stigmatizing nature of their loss. They may inadvertently become so-called neglected victims and are at risk for perpetuating an intergenerational cycle of violence.[52,53]

Patience and gentleness with the family's prolonged numbness and shock are essential features of caring for bereaved family members. Physical acts of kindness may be the only avenue of communication at first. Leading the family to a quiet room, providing comfortable seats, bringing beverages, and making sure that all possible members are included are all helpful and may lay the groundwork for dialogue.

Immediately after imparting the news of death, the physician may be able to bring the family together and start a dialogue by offering to give them as detailed an account as possible. Teamwork is usually required to get everyone present and seated with beverages, ashtrays, and any other comforts that seem appropriate. Survivors may benefit from very gentle questions about the last hours of the deceased: Were there any prodromal syndromes? Any premonitions? Who saw the deceased last? Families who do not wish to explore these crucial questions at this time should not be pushed, however.

A chaplain, nurse, or other team member with counseling skills, present from the time the physician begins communicating the bad news, may be able to address sensitive issues that arise. Family members or other supportive figures (e.g., family doctor or clergyman) who are absent should be notified and asked to come to the hospital when appropriate. When the family members are too shaken to sit down or participate in any dialogue, it is important to leave them a telephone contact at the hospital should any questions arise.

MEDICATIONS AND BEREAVEMENT

Treatment of bereavement-related major depressive episodes has recently been shown to be beneficial. In one trial, persons who had lost their spouses within 6 to 8 weeks and met the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV) criteria for a major depressive episode were treated with sustained-release bupropion. Improvement was noted in both depression and grief intensity.[54] In another study, persons with major depressive episodes that began within 6 months before or 12 months after the loss of a spouse were randomly assigned to a 16-week double-blind trial of one of four treatments: nortriptyline plus interpersonal psychotherapy, nortriptyline alone in a medication clinic, placebo plus interpersonal psy-

chotherapy, or placebo alone in a medication clinic. Nortriptyline proved superior to placebo in achieving remission of bereavement-related major depressive episodes, but the combination of medication and psychotherapy was associated with the highest rate of treatment completion. The investigators concluded that the results support the use of pharmacologic treatment of major depressive episodes in the wake of a serious life stressor such as bereavement.[55]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Magni KG: The fear of death. Death and Presence: The Psychology of Death and the After-Life. Godin A, Ed. Lumen Vitae, Brussels, Belgium, 1972, p 125

2. The EPEC Project: Education on Palliative Care and End-of-Life Care. Feinberg School of Medicine, Northwestern University, Chicago
http://www.epec.net

3. Hippocrates: Decorum, XVI. Hippocrates with an English Translation, Vol 2. Jones WH, Ed. Heinemann, London, England, 1923

4. Kelly WD, Friesen SR: Do cancer patients want to be told? Surgery 27:822, 1950

5. Morrison RS, Meier DE: Palliative Care. N Engl J Med 350:2582, 2004

6. Searight HR, Gafford J: Cultural diversity at the end of life: issues and guidelines for family physicians. Am Fam Physician 71:515, 2005

7. Blackhall LJ, Murphy ST, Frank G, et al: Ethnicity and attitudes toward patient autonomy. JAMA 274:820, 1995

8. Tse CY, Chong A, Fok SY: Breaking bad news: a Chinese perspective. Palliat Med 17:339, 2003

9. Sullivan RJ, Menapace LW, White RM: Truth-telling and patient diagnoses. J Med Ethics 27:192, 2001

10. Gerle B, Lunden G, Sandblom P: The patient with inoperable cancer from the psychiatric and social standpoints. Cancer 13:1206, 1960

11. Emanuel LL, Alpert HR, Baldwin DC, et al: What terminally ill patients care about: toward a validated construct of patients' perspectives. J Palliat Med 3:419, 2000

12. Aitken-Swan J, Easson EC: Reactions of cancer patients on being told their diagnosis. Br Med J 1:779, 1959

13. Gilbertsen VA, Wangensteen OH: Should the doctor tell the patient that the disease is cancer? Surgeon's recommendation. The Physician and the Total Care of the Cancer Patient. American Cancer Society, New York, 1962, p 80

14. Croog SH, Shapiro SD, Levine S: Denial among male heart patients. Psychosom Med 33:385, 1971

15. Greer S: The management of denial in cancer patients. Oncology (Williston Park) 6:33, 1992

16. Watson M, Homewood J, Haviland J, et al: Influence of psychological response on breast cancer survival: 10-year follow-up of a population-based cohort. Eur J Cancer 41:1710, 2005

17. Nelson DV, Friedman LC, Baer PE, et al: Attitudes of cancer: psychometric properties of fighting spirit and denial. J Behav Med 12:341, 1989

18. Hagerty RG, Butow PH, Ellis PM, et al: Communicating with realism and hope: incurable cancer patients' views on the disclosure of prognosis. J Clin Oncol 23:1278, 2005

19. Grbich C, Parker D, Maddocks I: The emotions and coping strategies of caregivers of family members with a terminal cancer. J Palliat Care 17:30, 2001

20. Koenig HG, George LK, Titus P: Religion, spirituality, and health in medically ill hospitalized older patients. J Am Geriatr Soc 52:554, 2004

21. Larson DB, Swyers JP, McCullough ME: Scientific research on spirituality and health: a consensus report. National Institute for Healthcare Research, Rockville, Maryland, 1998

22. Mueller PS, Plevak DJ, Rummans TA: Religious involvement, spirituality and medicine: implications for clinical practice. Mayo Clin Proc 76:1225, 2001

23. McClain CS, Rosenfeld B, Breibart W: Effect of spiritual well-being on end-of-life despair in terminally-ill patients. Lancet 362:1603, 2003

24. Saunders CM: Foreword. Mortally Wounded: Stories of Soul, Pain, Death, and Healing. Kearney M, Ed. Simon & Schuster, New York, 1997, p 11

25. Wilson KG, Curran D, McPherson CJ: A burden to others: a common source of distress for the terminally ill. Cogn Behav Ther 34:115, 2005

26. Kash KM, Holland JC, Breitbart W, et al: Stress and burnout in oncology. Oncology (Williston Park) 14:1621, 2000

27. Cassem NH, Hackett TP: Psychiatric consultation in a coronary care unit. Ann Intern Med 75:9, 1971

28. Penson RT, Partridge RA, Shah MA, et al: Fear of death. Oncologist 10:160, 2005

29. Breitbart W, Rosenfeld B, Pessin H, et al: Depression, hopelessness, and desire for hastened death in terminally ill patients with cancer. JAMA 284:2907, 2000

30. Brurera E, MacEachern T, Ripamonti C, et al: Subcutaneous morphine for dyspnea in cancer patients. Ann Intern Med 119:906, 1993

31. Massie MJ: Anxiety, panic and phobias. Handbook of Psycho-oncology: Psychological Care of the Patient with Cancer. Holland JC, Rowland JH, Eds. Oxford University Press, New York, 1989, p 300

32. Cassem NH: Depression and anxiety secondary to medical illness. Psychiatry Clin North Am 13:597, 1990

33. Hotopf M, Chidgey J, Addington-Hall J, et al: Depression in advanced disease: a systematic review. Part 1. Prevalence and case finding. Palliat Med 16: 81, 2002

34. Chochinov HM: Management of grief in the cancer setting. Psychiatric Aspects of Symptom Management in Cancer Patients. Breitbart W, Holland JC, Eds. American Psychiatric Press, Washington, DC, 1993, p 231

35. Block SD: Assessing and managing depression in the terminally ill patient. ACP-ASIM End of Life Care Consensus Panel. American College of Physicians-American Society of Internal Medicine. Ann Intern Med 132:209, 2000

36. Plumb MM, Holland JC: Comparative studies of psychological function in patients with advanced cancer. Psychosom Med 39:264, 1977

37. Ganzini L, Lee MA, Heintz RT, et al: The effect of depression treatment on elderly patients' preferences for life-sustaining medical therapy. Am J Psychiatry 151:1631, 1994

38. Chochinov HM, Hack T, Hassard T, et al: Dignity therapy: a novel psychotherapeutic intervention for patients near the end of life. J Clin Oncol 23:5520, 2005

39. Clinical practice guidelines for quality palliative care, executive summary. National Consensus Project for Quality Palliative Care. J Palliat Med 7:611, 2004

40. Byock I: Dying Well: Peace and Possibilities at the End of Life. Riverhead Books, Berkley Publishing Group, New York, 1997, p 140

41. Parkes CM, Weiss RS: Recovery from Bereavement. Basic Books Inc, New York, 1983

42. Lindemann E: Symptomatology and management of acute grief. Am J Psychiatry 101:141, 1944

43. Bereavement: Reactions, Consequences, and Care. Osterweis M, Solomon F, Green M, Eds. National Academy Press, Washington, DC, 1984

44. Furman E: Child's Parent Dies: Studies in Childhood Bereavement. Yale University Press, New Haven, Connecticut, 1988

45. Geis HK, Whittlesey SW, McDonald NB, et al: Bereavement and loss in childhood. Child Adolesc Psychiatr Clin N Am 7:73, 1998

46. Fristad MA, Jedel R, Weller RA, et al: Psychosocial functioning in children after the death of a parent. Am J Psychiatry 150:511, 1993

47. Holmes TH, Rahe RH: The social readjustment rating scale. J Psychosom Res 11:213, 1967

48. Schaefer C, Quesenberry CP Jr, Wi S: Mortality following conjugal bereavement and the effects of a shared environment. Am J Epidemiol 141:1142, 1995

49. Cassem NH: The person confronting death. The New Harvard Guide to Psychiatry. Harvard University Press, Cambridge, Massachusetts, 1988, p 728

50. Ness DE, Pfeffer CR: Sequelae of bereavement resulting from suicide. Am J Psychiatry 147:279, 1990

51. Rynearson EK, McCreery JM: Bereavement after homicide: a synergism of trauma and loss. Am J Psychiatry 150:258, 1993

52. Black D, Kaplan T: Father kills mother: issues and problems encountered by a child psychiatric team. Br J Psychiatry 153:624, 1989

53. McCune N, Donnelly P: Child surviving parental murder (letter). Br J Psychiatry 154:889, 1989

54. Zisook S, Shucter SR, Pedrelli P, et al: Bupropion sustained release for bereavement: results of an open trial. J Clin Psychiatry 62:227, 2001

55. Reynolds CF 3rd, Miller MD, Pasternak RE, et al: Treatment of bereavement-related major depressive episodes in later life: a controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry 156: 202, 1999

56. Massie MJ: Depression: Handbook of Psycho-oncology. Holland JC, Rowland JH, Eds. Oxford University Press, New York, 1989, p 283

XII COMPLEMENTARY AND ALTERNATIVE MEDICINE

Bimal H. Ashar, m.d., f.a.c.p.
Adrian S. Dobs, m.d., m.h.s.

Definitions

The term alternative medicine encompasses a spectrum of approaches to medical conditions not routinely used by conventional practitioners. Historically, the term has been associated with negative conceptions about medical practices that did not conform to accepted standards of care. The term complementary medicine has since evolved to describe a more positive, symbiotic relationship between unconventional medicine and conventional medicine. The field of complementary and alternative medicine (CAM) now encompasses a multitude of different approaches and beliefs that are generally linked by their emphasis on so-called natural modalities of healing and wellness. More and more, the term integrative medicine is being used, suggesting that CAM should be integrated into conventional care. This chapter describes modalities that are complementary to conventional medicine in the United States; in other countries, many of these modalities are part of mainstream medical practice.

Classification

Patient demand, media attention, and the growth of an approximately $40 billion industry[1] have stimulated leaders in governmental agencies and academic medicine to recognize and categorize CAM and to direct research initiatives on the subject. Although there is currently no universally accepted classification of CAM modalities, the National Center for Complementary and Alternative Medicine (NCCAM) has grouped CAM practices into five domains [*see Table 1*]. It should be recognized that these categories are not mutually exclusive. Certain practices will overlap (e.g., qigong is considered an energy therapy but is part of Chinese medicine, which is an alternative medical system). Also, as evidence emerges regarding mechanisms of action, safety, and efficacy, certain modalities will naturally move beyond the CAM label and become part of mainstream medicine.

Use of CAM

Prevalence and demographics

The widespread use of CAM by the public has been well documented. In 2002, about 62% of adults in the United States reported using at least one form of alternative medicine within the previous year.[1] It has been estimated that 75% of people in the United States have used at least one CAM therapy over their lifetime.[1] Public-opinion surveys have suggested similar overall patterns of use in European countries, although the popularity of specific CAM modalities varies greatly from country to country.[2] Patients across all demographic groups use alternative medicine. However, some surveys have noted that predictors of CAM use may include female gender, white race (as opposed to African-American or Hispanic), higher socioeconomic status, and higher levels of education.[3,4] Many CAM users have chronic, non–life-threatening medical conditions,[3,5] and they may have an interest in spirituality.[6] A number of diagnosis-based surveys suggested exceptionally high usage of alternative medicine in patients with cancer,[7] HIV infection,[8] fibromyalgia,[9] and inflammatory bowel disease.[10]

Public perception

The alternative-medicine movement has clearly been a public-driven process that has spanned decades. It was initially thought that this movement was primarily the result of dissatisfaction with conventional medicine.[11] Subsequent studies have shown that this is not the case[6,12] and that patients continue to see their conventional health care practitioners while using CAM therapies; however, about 27% of CAM users believe that conventional medicine will not help their health care problem.[1] Two disturbing observations are that most patients do not disclose their use of alternative therapies to their physicians and that such patients are never asked about CAM use by their physicians.[13] Furthermore, many patients feel no need to communicate their CAM use to their physicians because they believe that their physicians would be unable to understand and incorporate that information into their treatment plan.[12,14] On the other hand, current data suggest that about one quarter of patients who use CAM do so on the advice of a conventional medical professional.[1]

A number of other factors have stimulated public use of alternative medical therapies. The fact that many CAM modalities emphasize natural forms of healing seems to form the fundamental basis for its use. Many patients desire a more holistic approach to their medical care.[6] They may feel that conventional medicine focuses excessively on suppression of symptoms (e.g., pharmacologic lowering of elevated blood pressure) rather than addressing the root cause of symptoms. They believe that so-called natural products are better and safer than synthetic medications. In many cases, they may turn to alternative medical practices to get relief from chronic conditions that have not responded to conventional symptomatic therapy. Additionally, media hype, direct-to-consumer advertising, and the wide-

Table 1 NIH/NCCAM Classification of Complementary and Alternative Medicine Practices

Category	*Examples*
Alternative medical systems	Ayurveda (traditional Indian medicine), traditional Chinese medicine, homeopathy
Mind-body interventions	Biofeedback, hypnosis, meditation, prayer
Biologic-based therapies	Dietary therapy, herbal medicine, megavitamins, shark cartilage
Manipulative and body-based methods	Chiropractic, massage therapy, craniosacral therapy
Energy therapies	Therapeutic touch, qigong, bioelectric field manipulation, reiki

NIH/NCCAM—National Institutes of Health/National Center for Complementary and Alternative Medicine

spread availability of information over the Internet have all played a role in the popularity of CAM and have served to expand the public's health care choices. Of concern to many physicians is that these choices are frequently based on insufficient basic science or clinical evidence.

Research Concerns

SCIENTIFIC ISSUES

One of the defining characteristics of alternative medicine is the paucity of definitive evidence supporting mechanism of action, efficacy, and safety. Although a number of clinical trials on CAM have been published, the overall quality of those trials is quite poor, primarily because of inadequate sample size, randomization, and blinding.[15,16] Additionally, publication bias may be common in the international literature. Critical reviews of published studies on CAM therapies from a number of countries have shown that the studies almost universally report positive findings pertaining to CAM. This suggests that studies reporting negative findings may never make it to press.[17,18]

There are a number of barriers to the proper evaluation of CAM studies. First, the establishment of adequate control groups is frequently very difficult. Studies on acupuncture, for example, have attempted to incorporate a placebo control by stimulating nonacupuncture points, stimulating actual points unrelated to the treated condition, or applying pressure instead of inserting needles. Some critics argue that so-called sham acupuncture is an inadequate placebo that does not preserve subject blinding. Proponents of acupuncture may argue that such control methods are still potentially therapeutic because of their possible positive effect on the flow of subtle energy through the body. Similar pitfalls are inherent in mind-body research. In the study of personal prayer, of prayer groups, or of intercessory prayer that occurs in the presence of the patient, the intervention group can be compared with those who do not partake in organized prayer. Such a design clearly does not lend itself to adequate blinding. Additionally, any positive results could reflect aspects of prayer that are unrelated to its spiritual qualities (e.g., relaxation), making definitive conclusions difficult.

Another major problem with interpreting CAM research stems from inconsistencies in the intervention groups. Drawing meaningful conclusions from herbal-medication studies is difficult because extracts are not standardized. For example, although positive effects have been seen in a number of published clinical trials with the plant genus *Echinacea* for treatment of upper respiratory tract infections, definitive conclusions cannot be drawn because of variation in the species of plant studied, the part of the plant utilized (root, leaf, or flower), and extraction methods.[19] In addition, the manufacturing processes within and between companies vary widely, so that the concentration of active product in an over-the-counter preparation is rarely known.

Lack of standardization is also a flaw in acupuncture research. Many different types of acupuncture are practiced around the world, and each type may utilize a completely different set of points for the same condition. Even among providers who practice the same type of acupuncture, variation in point selection is common because approaches differ on the basis of the patient's history and physical examination and on the acupuncturist's personal style. This individualization of therapy is alluring to patients, but the unwillingness of practitioners to agree on what constitutes acceptable technique challenges conventional study methodology. CAM practitioners often criticize the typical scientific model that employs randomized clinical trials because in clinical practice, there are multiple interventions, such as mind-body and herbal treatments, that occur simultaneously. Thus, any research in the area may have to be multidimensional.

FINANCIAL ISSUES

Unlike conventional pharmaceutical and medical-device research, large-scale studies in CAM derive their funding almost exclusively from government sources. Modalities such as prayer, acupuncture, and massage therapy are not lucrative enough endeavors to support large, privately funded trials. Dietary supplements, such as herbs, may have a significant profit potential, but the incentive for research is weakened by the fact that herbs, like other natural substances, cannot be patented. In addition, the rules and regulations under which foods and natural products are regulated differ from those for pharmaceuticals, which must meet stringent standards of efficacy and safety.

In an effort to boost CAM research, the United States Government set up NCCAM (http://nccam.nih.gov), under the National Institutes of Health (NIH). With an annual working budget of about $120 million, NCCAM has funded a number of individual projects, as well as specialty centers around the country [*see Table 2*].

Specific CAM Modalities

ALTERNATIVE MEDICAL SYSTEMS

Traditional Chinese Medicine

Acupuncture Acupuncture has been used for centuries as a component of traditional Chinese medicine (TCM). It involves the insertion of thin needles into specific points on the skin to facilitate the movement of energy (qi). Chinese medicine posits that qi (pronounced *chee*) flows along distinct channels (called meridians) in the body and that balanced circulation of qi is a prerequisite for good health. A block in the flow of qi can result in either a deficiency or an excess of qi along a meridian; those imbalances can be corrected by accurate needle placement (or pressure, in the case of acupressure) at specific points on the body. Acupuncture practitioners often enhance the effect of the needles by electrical stimulation; manual manipulation (e.g., twirling); or moxibustion, which involves burning mugwort (*Artemisia vulgaris*) on the acupuncture point or the end of the needle. Practitioners of TCM frequently combine acupuncture with other modalities, including herbal remedies, to achieve the desired physiologic response. Each treatment is individualized on the basis of the patient's history and physical examination, including pulse and tongue examinations. Many types of acupuncture are practiced today; a few examples are traditional Chinese acupuncture, five-elements acupuncture, and auricular acupuncture.

To date, no clear physical mechanism of action has emerged to explain the potential therapeutic response to acupuncture. Changes in blood flow and biologic mediators (e.g., hormones, neurotransmitters, and endorphins) have been shown to occur with needle manipulation.[20-23] There are numerous published clinical studies on acupuncture treatment for a variety of ailments. Most are small in size and have methodologic flaws that make consensus difficult. Nevertheless, in 1997, an NIH consensus panel concluded that there is clear evidence to support the use of acupuncture for postoperative, chemotherapy-induced, and prob-

ably pregnancy-associated nausea and vomiting.[23] Although the data are less compelling, evidence also suggests a positive effect of acupuncture on idiopathic headache,[24] fibromyalgia,[25,26] and osteoarthritis of the knee.[27] Current evidence does not support its use for smoking cessation,[28] asthma,[29] or low back pain.[30]

If done correctly, acupuncture is quite safe.[31] Rare case reports of serious adverse events, including skin infections, hepatitis, pneumothorax, and cardiac tamponade, seem to stem from inadequate sterilization of needles and practitioner negligence.[32,33] To prevent transmission of infection, most practitioners now use disposable needles. Minor side effects, including insertion-site pain or bleeding, fatigue, and vasovagal syncope, are probably more common.[34]

Table 2 Government-Funded Specialty Centers for Research into Complementary and Alternative Medicine

Specialty	*Center and Location*
Acupuncture—addiction	Minneapolis Medical Research Foundation, Minneapolis; New England School of Acupuncture, Watertown, Massachusetts
Acupuncture—neuroimaging	Massachusetts General Hospital, Charlestown
Aging and women's health	Columbia University, New York
Antioxidants	Oregon State University, Corvallis
Arthritis	University of Maryland, Baltimore
Botanical treatment of age-related diseases	Purdue University, West Lafayette, Indiana
Botanical dietary supplements for women's health	University of Illinois, Chicago
Botanical dietary supplements	UCLA, Los Angeles
Cancer	Johns Hopkins University, Baltimore
Cancer and hyperbaric oxygen	University of Pennsylvania, Philadelphia
Cardiovascular diseases	University of Michigan, Ann Arbor
Cardiovascular disease and aging in African Americans	Maharishi University of Management, Fairfield, Iowa
Chiropractic	Palmer Center for Chiropractic Research, Davenport, Iowa
Craniofacial disorders	Kaiser Foundation Hospitals, Portland, Oregon
Frontier medicine (biofield science)	University of Arizona, Tucson
Frontier medicine (therapeutic touch)	University of Connecticut, Farmington
Neurodegenerative diseases	Emory University School of Medicine, Atlanta
Neurologic disorders	Oregon Health Sciences University, Portland
Pediatrics	University of Arizona Health Sciences Center, Tucson
Phytomedicine	University of Arizona, Tucson
Phytonutrient and phytochemical studies	University of Missouri, Columbia

Homeopathy

Homeopathy is one of the most controversial modalities in CAM, primarily because of its theoretical implausibility. The roots of homeopathy trace back to the 1700s, when it was first described by Samuel Hahnemann. Homeopathic principles revolve around two basic tenets: the law of similars and the principle of serial dilutions. According to the law of similars, substances that cause symptoms in healthy people can cure those same symptoms in people who are sick. A number of examples of this principle exist in conventional medicine. Digoxin is used to treat the same arrhythmias that it is capable of inducing. Similarly, methylphenidate, which is a stimulant, has been used to treat attention-deficit/hyperactivity disorder.[35]

The principle of serial dilutions (or the minute dose) is another controversial aspect of homeopathy. According to this principle, medications can have a biologic effect even if diluted to levels at which the original substance is undetectable (a so-called homeopathic dose).

A homeopath's approach to patients differs from that of the conventional physician. Homeopaths concentrate almost exclusively on subjective symptoms and sensations. They choose medications on the basis of the patient's symptomatology, rather than on the objective medical diagnosis. This results in the use of a wide array of different medications for any one conventionally diagnosed condition. A number of homeopathic encyclopedias (*materia medica*) are available that describe symptoms induced by different remedies when given to healthy individuals (provings). These provings are matched to a patient's symptoms to determine the therapeutic regimen. Patients are typically followed closely so that the homeopath can titrate dosing schedules.

Meta-analyses of a number of trials of homeopathic remedies have suggested an effect superior to placebo.[36,37] However, many studies and reviews on specific medications have shown negative or inconclusive results.[37-39] Given the conflicting clinical data and the lack of evidence regarding mechanism of action, it is difficult to support the general use of homeopathy until more high-quality research is available.

Because homeopathic remedies typically contain little or no detectable active ingredients, serious side effects are rare. Homeopathic preparations are generally marketed as over-the-counter remedies and are usually exempt from government requirements for finished product testing or expiration dating. In the United States, the Food and Drug Administration requires that all homeopathic remedies list the indications for their use, the ingredients, instructions for safe use, and dilutions. Dilutions in a ratio of 1:10 are labeled with an X, and dilutions in a ratio of 1:100 are labeled with a C. For example, a 3X product has been diluted 1:10 three times; a 3C product has been diluted 1:100 three times. It should also be noted that these remedies are not restricted to the 10% alcohol limit of conventional drugs.[40]

MIND-BODY INTERVENTIONS

The relationship between psychological stress and physical health has been studied extensively over the past 30 years. Despite positive results from some trials, interventions designed to alter the stress response have not become part of mainstream medical practice. The reluctance of physicians to incorporate

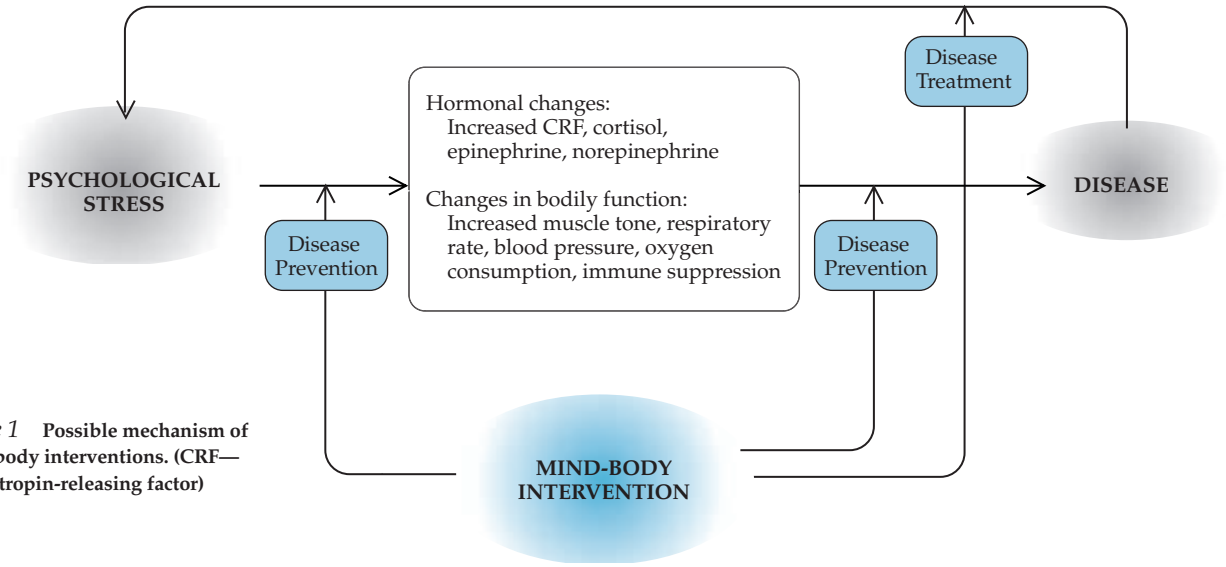


Figure 1 **Possible mechanism of mind-body interventions. (CRF—corticotropin-releasing factor)**

mind-body strategies into their therapeutic armamentarium likely stems from their unfamiliarity with such interventions; time constraints; and the lack of a clear mechanistic pathway to disease. Furthermore, many physicians may feel that these therapies need to be patient driven rather than physician driven, because they require significant changes in self-care.

The proposed theory of mind-body medicine stems from work done in the early 1900s. The fight-or-flight response was described as physiologic preparation for combating or fleeing an external threat.[41] Stimulation of the hypothalamus and increased sympathetic nervous system activity lead to neurohormonal stimulation and increases in blood pressure, heart rate, respiratory rate, and muscle tension. This response has historically been protective, ensuring survival in the face of physical danger. In today's society, however, we are continually faced with innumerable stressors that can elicit the fight-or-flight response, yet fighting or running away is inappropriate or impossible. The body is primed for action but can take none. This chronic physiologic stimulation is thought to increase the likelihood of disease. Furthermore, the development of a chronic disease may stimulate the response through a feedback mechanism, potentially worsening the condition. The effect of the chronic fight-or-flight response on immunosuppression and cytokine and hormone production needs greater elucidation.

Mind-body interventions can elicit a relaxation response that may prevent or aid in the treatment of a number of medical ailments[42] [*see Figure 1*]. A number of modalities can be used for this purpose. Many people have incorporated yoga, meditation, or self-hypnosis into their daily self-care regimen. Several clinical studies have suggested that there are positive results from mind-body modalities for many conditions [*see Table 3*]. As with other CAM interventions, however, limitations in study methodology and sample size, as well as lack of an adequate control, make definitive conclusions difficult.

The mind-body category also encompasses techniques for which a mechanism is not even remotely understood. No physical explanation for distant healing modalities—such as intercessory prayer, spiritual healing, and mental healing—is currently accepted, despite some evidence for positive treatment effects.[43] No harmful effects are seen when most mind-body interventions are used as an adjunct to conventional care. However, there is concern that patients might choose exclusive use of one or more of these methods in lieu of appropriate diagnosis and therapy.

BIOLOGIC-BASED THERAPIES

Biologic therapy is the most popular of all fields of CAM. Its popularity stems from its similarity to the process of using conventional medications. Some people consider biologics to be a possible quick fix for their ailments, without the need for physician visits or potentially harmful prescription medications. Others turn to biologics in the hope of preventing potentially serious diseases through the use of so-called natural substances.

Dietary supplements, including herbal and nonherbal products, make up the preponderance of medications in this category. The supplement industry has become a billion-dollar business, largely as a result of the loosening of federal regulations. The Dietary Supplement Health and Education Act (DSHEA) of 1994 expanded the definition of dietary supplements to include vitamins, amino acids, herbs, and other botanicals. Furthermore, under DSHEA, supplements no longer require premarket testing for safety and efficacy. Supplements are assumed to be safe unless proved otherwise by the FDA. Given the number and variety of products currently available, the FDA's ability to effectively regulate all products after they have been marketed is limited. An example of the regulatory process was the banning of products containing ephedra. The FDA first expressed concern over the herb in 1997, when it proposed limitations on its use. The General Accounting Office viewed these limitations as inappropriate, because of insufficient evidence proving harm. It took 7 years (and a few high-profile deaths) for the FDA to be able to effect a ban on ephedra.[44] Although most dietary supplements are well tolerated and are associated with few adverse effects, the potential for harm from the lack of regulation can be seen from examples of misidentification of plant species,[45] contamination with heavy metals, and addition of pharmaceutical agents.[46,47]

Overall, there is only limited evidence supporting the use of most dietary supplements. Most clinical trials have been small, nonrandomized, or unblinded. In general, physicians and patients should view herbs as medications. Physicians should advise patients to be wary of products for which grandiose claims are made, because misleading advertising is common.[48] The potential for significant toxicity and drug interactions does exist.

The list of currently used supplements is immense, and this chapter can touch on only the most popular [*see Tables 4 through 6*]. More comprehensive resources for dietary supplements (and other CAM modalities) are now available [*see Table 7*].

MANIPULATIVE AND BODY-BASED THERAPIES

Chiropractic

Many would argue that chiropractic medicine should not be considered alternative therapy. Patients, physicians, and insurance companies have all shown some degree of support for chiropractic care in recent years. Between 10% and 20% of the population have used chiropractors.[6] Health care insurance plans, including Medicare, cover many of the services performed during chiropractic visits. Most chiropractor visits are for musculoskeletal problems, including low back pain, neck pain, and extremity pain. However, a small proportion of patients currently seek out chiropractic care for a variety of other conditions, as well as general health concerns.[49] The tenets of chiropractic medicine place the spinal cord and nervous system at the center of a person's well-being. The nervous system is thought to control and influence all other bodily systems. Malalignments (subluxations) of the vertebrae are thought to cause or perpetuate disease. Once these subluxations are identified and corrected (via manipulation), the body uses its natural healing abilities to restore physiologic balance and health. Chiropractors typically look for spinal pain, asymmetry, impaired range of motion, or abnormalities in tone, texture, and temperature when evaluating patients.[50] Laboratory testing, including x-rays, electromyography (EMG), and ultrasonography, may be used to aid in diagnosis. Actual spinal manipulation is performed by direct or indirect delivery of thrusts to the spine. Frequently, the patient will experience a cracking noise. Some chiropractors may use adjunctive therapies, including massage, heat, and trigger-point injections.[50]

Chiropractic manipulation has been touted as treatment for a number of conditions, including hypertension, asthma, pelvic pain, and fibromyalgia. Very little data exist to support its use for these conditions.[51,52] Use of chiropractic therapy for neck pain and headaches is also weakly supported.[53,54] Much of the current use of chiropractic care stems from its utility in cases of low back pain. A number of controlled trials on chiropractic treatment for low back pain have been done, with conflicting results. A meta-analysis concluded that spinal manipulation appears to be more effective than sham therapy or treatments previously judged to be ineffective, but not to be superior to other standard treatments for acute or chronic low back pain, such as analgesics, physical therapy, or exercises.[55] Patient satisfaction also seems to be high with such therapy.[56]

Serious complications from lumbar spinal manipulation seem to be uncommon, although there are reports of cauda equina syndrome.[57] Many patients, however, experience mild to moderate side effects, including localized discomfort, headache, or tiredness. These reactions usually disappear within 24 hours.[58] Brain stem or cerebellar infarction, vertebral fracture, tracheal

Table 3 Selected Mind-Body Interventions

Modality	*Description*	*Potential Applications*	*Comments*
Aromatherapy	The use of essential oils (e.g., jasmine, chamomile, lavender) to enhance physical or psychological well-being; often combined with massage	Anxiety, agitation	Long-term efficacy data (independent of massage) are lacking[71]
Biofeedback	Voluntary control of physiologic processes—e.g., brainwaves, smooth muscle contraction, vasodilation—learned and reinforced with the aid of instrumentation (EEG, EMG, skin temperature/sweat monitors)	Asthma, ADHD, back pain, fibromyalgia, headache, hypertension, incontinence, neuromuscular disorders, Raynaud disease	Effective for fecal incontinence[72]; techniques utilized may vary between patients and practitioners; learning process can be slow, requiring multiple sessions with therapist and regular practice by patient
Guided imagery	Use of the imagination to positively stimulate the senses to bring about emotional and physiologic change	Chronic pain, perioperative management, headaches, nausea, posttraumatic stress disorder	Small studies suggest a positive impact on surgical and cancer treatment outcomes[73]
Hypnotherapy	The induction of a trancelike state to induce relaxation and susceptibility to positive suggestion; used as a diagnostic and therapeutic tool	Anesthesia, headache, irritable bowel syndrome, smoking cessation	Success of therapy may depend on patient susceptibility and attitude toward hypnosis; no conclusive data on most conditions
Intercessory prayer	Request to God (or other spiritual beings) for the benefit of others; can take place in the presence of the patient or at a distance	Cardiac disease, HIV infection, RA	Studies are conflicting and inconclusive[74]; mechanism is unclear
Meditation	Release of the mind from attachment to discursive thought, typically aided by focusing on the breath or a mantra	Anxiety, chronic pain, hypertension, substance abuse	Many types of meditation exist; large-scale studies are needed to prove the absolute impact of this simple intervention on health
Music therapy	Use of music to improve psychological, physical, cognitive, or social functioning	Anxiety, dementia, chronic pain, Parkinson disease	Treatment is guided by a trained music therapist; may have a short-term effect on anxiety and mood during treatments or procedures[75]
Writing therapy	Creative writing exercise about an emotionally traumatic event	General emotional health, asthma, RA, HIV infection	Short-term studies show positive response in asthma and RA[76,77]

ADHD—attention-deficit/hyperactivity disorder EEG—electroencephalography EMG—electromyography RA—rheumatoid arthritis

Table 4 Commonly Used Herbal Dietary Supplements

Herb	*Suggested Uses*	*Potential Toxicity*	*Potential Drug Interactions*	*Comments*
Black cohosh (*Cimicifuga racemosa*)	Menopausal symptoms	Gastrointestinal discomfort	None known	Scant efficacy data, long-term safety unknown[78]
Chaste tree berries (*Vitex agnus-castus*)	Premenstrual syndrome, mastodynia	Pruritus	May have dopaminergic activity; therefore, avoid with use of dopamine-receptor antagonists (e.g., neuroleptics)	Small, short-term studies suggest efficacy[79]
Cranberry (*Vaccinium macrocarpon*)	Urinary tract infections	Nephrolithiasis (with cranberry concentrate tablets)[80]	Possible interaction with warfarin[81]	No clear role in treatment of UTIs; may be effective for prophylaxis[82]
Dong quai (*Angelica sinensis*)	Menopausal symptoms	Rash	Increased INR in patients taking warfarin	No clinical evidence of efficacy[78]
Echinacea (*E. purpurea, E. pallida, E. angustifolia*)	Upper respiratory infections	Hypersensitivity reactions	Theoretically, may antagonize the effect of immunosuppressive medications	Variations in plant species studied, part of plant used, and extraction methods make conclusions regarding efficacy difficult[83]
Ephedra (*E. sinica,* mahuang)	Asthma, congestion, weight loss	Hypertension, arrhythmia, myocardial infarction, stroke	Interaction with monoamine oxidase inhibitors and cardiac glycosides; potential for serious toxicity when combined with other stimulants	Banned by the FDA, effective April 2004, but still available internationally over the Internet
Evening primrose (*Oenothera biennis*)	Eczema, irritable bowel syndrome, mastalgia, premenstrual syndrome, rheumatoid arthritis	Nausea, vomiting, diarrhea, flatulence	Possible lowering of seizure threshold in patients taking antiepileptic medications[84]	Conflicting efficacy data for a number of conditions
Feverfew (*Tanacetum parthenium*)	Migraine prophylaxis	Hypersensitivity reactions	Theoretical risk of increased bleeding when combined with anticoagulants	No clear evidence to support efficacy[85]
Garlic (*Allium sativum*)	Cardiovascular protection	Gastrointestinal upset, bleeding	Theoretical risk of increased bleeding when combined with anticoagulants	Possible short-term improvement in cardiovascular risk factors, but impact on disease unknown; active ingredient unclear[86]
Ginger (*Zingiberis rhizoma*)	Nausea, motion sickness, dyspepsia	None known	Theoretical risk of increased bleeding when combined with anticoagulants	May be useful for nausea and vomiting of pregnancy[87]

rupture, internal carotid artery dissection, and diaphragmatic paralysis are rare but have all been reported with cervical manipulation.[59] Given the lack of efficacy data and the risk (although small) of catastrophic adverse events, it is difficult to advocate routine use of this technique for treatment of neck or headache disorders. Physicians should also recognize potential contraindications to chiropractic therapy. Patients with coagulopathy, osteoporosis, rheumatoid arthritis, spinal neoplasms, or spinal infections should be advised against such treatments.[59]

Massage Therapy

A number of different types of massage are in practice today. Many therapists combine aspects of Swedish massage (stroking and kneading), shiatsu (pressure-point manipulation), and neuromuscular massage (total body, deeper therapy) to relieve stress, anxiety, and muscle tension, as well as improve circulation. Frequently, aromatic oils are employed to enhance the relaxation response. A number of small studies have suggested a potential beneficial effect of massage on fibromyalgia, headaches, and anxiety,[60] although the paucity of data precludes definitive conclusions. Massage therapy does seem to be effective for subacute and chronic back pain.[30] No significant adverse effects are seen with properly performed massage, although caution must be advised for patients with coagulation disorders.

Structural integration (rolfing) is a system of deep-tissue manipulation that involves stretching of the fascial planes. In this system, the fascia is thought to be the key supporting structure for bones and muscles. When injury or stress occurs, the fascia tends to become shorter and thicker. Manipulation of the fascia with fingers, thumbs, and elbows is supposed to relieve tension, restore structural integrity, and improve physiologic and psychological function. Limited data exist to support the efficacy of rolfing for any particular condition.

ENERGY THERAPIES

Many traditional cultures describe the physical body as existing within a field of energy. Such energy is called prana by Indians and qi by the Chinese; English terms include subtle energy, vital energy, and life energy. Many ancient and modern CAM techniques involve the manipulation of this energy or the transfer of additional energy into the patient's field in an effort to restore or maintain balance. Because the field extends beyond the body, energy therapies do not always involve physical contact between practitioner and patient. Further, the presumed connec-

Table 4 (*continued*)

Herb	*Suggested Uses*	*Potential Toxicity*	*Potential Drug Interactions*	*Comments*
Ginkgo biloba	Dementia, claudication, tinnitus	Gastrointestinal upset, headache, dizziness, bleeding, seizure	Theoretical risk of increased bleeding when combined with anticoagulants	May have modest effects on cognitive performance and functioning in patients with Alzheimer disease or multi-infarct dementia[88]; no evidence to support prevention of memory loss or dementia
Ginseng (*Panax* species; Asian ginseng, Korean ginseng, American ginseng)	Fatigue, diabetes	Generally considered safe; rare reports of hypertension, insomnia, headache, and mastalgia	May interact with monoamine oxidase inhibitors and warfarin (decreased prothrombin time)	Currently, little data to support its use[89]
Kava kava (*Piper methysticum*)	Anxiety	Rash, sedation, liver toxicity	May potentiate effects of benzodiazepines; best to avoid with other anxiolytics or alcohol because of risk of excess sedation	Studies suggest efficacy for short-term treatment of anxiety[90]; no data on addiction potential; banned in many European countries because of cases of hepatic failure
Kola nut (*Cola nitida*)	Fatigue	Irritability, insomnia	Caution when used with other stimulants	Contains caffeine
Milk thistle (*Silybum marianum*)	Chronic liver disease	Rare mild laxative effect, gastrointestinal upset	None known	Appears to be safe and well tolerated; efficacy data too limited to exclude a substantial benefit or harm[91]
Saw palmetto (*Serenoa repens*)	BPH	Mild gastrointestinal effects	None known	Short-term studies show improvement in symptoms[89,92]; no evidence for prevention of BPH or its complications, or prevention of prostate cancer
St. John's wort (*Hypericum perforatum*)	Depression, anxiety	Headache, insomnia, dizziness, gastrointestinal irritation	Can decrease levels of cyclosporine, digoxin, oral contraceptives, theophylline, and indinavir; serotonin syndrome can occur when combined with prescription SSRIs	May be effective for mild to moderate depression[89,93]
Valerian (*Valeriana officinalis*)	Insomnia	Headaches	Avoid use with benzodiazepines because of sedation	Efficacy data inconclusive[94]; theoretical risk of addiction with prolonged use

BPH—benign prostatic hyperplasia FDA—Food and Drug Administration INR—international normalized ratio SSRI—selective serotonin reuptake inhibitor UTI—urinary tract infection

tion of these individual fields with a universal field is believed to permit the use of some of these therapies at a distance.

Qigong

Qigong is a branch of traditional Chinese medicine designed to affect the flow of energy (qi) to preserve health. This system combines relaxation techniques with movement to achieve a meditative state designed to ensure mental and physical health. Tai chi (tai chi chuan) is a type of movement-oriented qigong that utilizes a sequence of slow, dancelike maneuvers to enhance the flow of qi through the body. In the course of a tai chi session, the person shifts body weight constantly from one foot to the other. Studies of tai chi in elderly persons have shown that long-term regular practice may improve balance, flexibility, and cardiovascular fitness and, possibly, decrease the risk of falls in older individuals.[61,62] Meditative qigong is accomplished without movement and is intended to establish inner harmony. Breathing exercises can also be part of qigong. They are designed to enhance circulation of qi and expel negative energy. Qigong has been used extensively in China for a number of conditions, including hypertension, anxiety, asthma, and nausea and vomiting.[63] Data to support use for any individual condition are lacking, despite historical successes. Although the principles of qigong seem simple, it involves a complex set of processes that are not clearly understood. Inappropriate training has reportedly been associated with physical and mental disturbances.[64]

Yoga

Yoga is an ancient Indian philosophical practice that uses postures or stretching exercises (asanas), breathing exercises (pranayama), and meditation to help unite the body and the mind. It was developed as a means of enlightenment through self-realization and self-mastery. Only recently, with its migration to the West, has yoga come to be seen as a means to heal illness or reduce anxiety. As with most CAM modalities, there are limited data for or against the use of yoga for particular conditions. Studies on the use of yoga in patients with carpal tunnel syndrome seem promising.[65] Yogic breathing exercises may have

Table 5 Commonly Used Nonherbal Dietary Supplements

Supplement	*Common Uses*	*Potential Toxicity*	*Potential Drug Interactions*	*Comments*
Coenzyme Q10	Heart failure, hypertension, angina, Parkinson disease	Nausea, heartburn, diarrhea	Decreased INR in patients on warfarin	Data inconclusive for treatment or prevention of cardiovascular disease[95]; early data promising for slowing the progression of Parkinson disease[96]
Glucosamine and chondroitin	OA	Hyperglycemia in diabetic patients	Theoretical risk of increased bleeding in patients taking chondroitin and anticoagulants	Current data suggest symptomatic improvement for OA of the hips and knees,[97] with a slowed progression of joint space narrowing[98]
Melatonin	Jet lag, insomnia	Fatigue, drowsiness, headache	Theoretical risk of bleeding in patients taking anticoagulants	Data suggest efficacy for jet lag[99]; no data on long-term use
SAMe (S-adenosylmethionine)	OA, depression, liver disease	Nausea, abdominal discomfort	Can lead to serotonin syndrome when used with tricyclic antidepressants	Early data promising[100]; poor oral bioavailability; very expensive; marketed doses are much lower than studied doses

INR—international normalized ratio OA—osteoarthritis

some beneficial effect on the symptoms of asthma and may reduce bronchodilator use, but they do not decrease airway reactivity or improve lung function.[66]

Therapeutic Touch

Therapeutic touch is the use of the hands, without actual physical touching, to influence or direct life energy throughout the body in an effort to promote healing. Therapeutic touch was codeveloped by a nurse, Dolores Krieger,[67] and many of its practitioners are nurses who use the technique for hospital inpatients.

In a therapeutic-touch session, which generally lasts 20 to 30 minutes, the practitioner enters a meditative state (centering) and then assesses the patient's energy field. To do so, the practitioner holds his or her hands a few inches from the patient's body and moves from head to foot. Downward sweeping movements are then used to remove any blockages of energy and correct any energy-field imbalances. The practitioner then transfers energy to the patient's field and finishes the session by smoothing the field.

A number of small trials have suggested a positive effect of therapeutic touch on conditions such as osteoarthritis, tension headache, and anxiety.[43] However, most of these trials are quite small and suffer from methodologic weaknesses that make definitive conclusions difficult.[68] A critical evaluation of relevant trials concluded that the data did not support the hypothesis that therapeutic touch promoted wound healing.[69] More vigorous trials need to be performed to determine the true efficacy of this technique.

CAM and the Practicing Physician

The field of research in CAM is in its infancy. Current levels of evidence are insufficient to support or disprove a majority of CAM modalities. Despite these limitations, the public continues to embrace CAM therapies as alternatives or adjuncts to conventional care. Given that many patients currently do not inform their physician of their use of CAM, it is imperative that physicians take the lead in inquiring about such therapies. Open dialogue needs to be established to uncover the types of modalities being utilized, reasons for pursuing such therapy, and patient experiences. From there, a discussion of the current data on level of efficacy and toxicity can follow. Ultimately, primary care physicians may need to develop referral networks of trusted CAM practitioners who are open to reciprocal communication.

Table 6 Popular Uses for Common Dietary Supplements*

Use	*Supplement*
Anxiety	Kava kava, St. John's wort, valerian
Benign prostatic hyperplasia	Saw palmetto, *Pygeum africanum*
Claudication	*Ginkgo biloba*
Dementia	*Ginkgo biloba*, pyridoxine
Depression	SAMe, St. John's wort
Diabetes	Aloe vera, chromium picolinate, ginseng, *Gymnema sylvestre*
Fatigue	Ginseng, licorice root
Heart failure	Coenzyme Q10, hawthorn extract
Hypercholesterolemia	Fenugreek, garlic, guggulipid, red yeast rice
Hypertension	Coenzyme Q10, garlic
Insomnia	Melatonin, valerian
Irritable bowel syndrome	Acidophilus, aloe, evening primose, peppermint oil
Jet lag	Melatonin
Liver disease	Milk thistle, SAMe
Menopausal symptoms	Black cohosh, dong quai, red clover, soy
Migraine prophylaxis	Butterbur root, feverfew, riboflavin
Nausea and vomiting	Ginger, pyridoxine
Osteoarthritis	Ginger, glucosamine and chondroitin, SAMe
Premenstrual syndrome	Calcium, chaste tree berries, evening primrose, pyridoxine
Tinnitus	*Ginkgo biloba*
Upper respiratory infections	Ascorbic acid, *Echinacea*
Urinary tract infections	Ascorbic acid, cranberry
Weight loss	Chitosan, chromium picolinate, *Garcinia cambogia*

*See Tables 4 and 5 for potential toxicity, drug interactions, and comments.

Table 7 Sources of Information on Complementary and Alternative Medicine

Medium	*Listing*	*Description (Price)*
Internet	Natural Medicine Comprehensive Database www.naturaldatabase.com	Monographs on a multitude of natural products ($92/year)
	Natural Standard www.naturalstandard.com	Monographs on a multitude of natural products and CAM modalities; grades the level of evidence currently available for each potential application ($99/year)
	Memorial Sloan Kettering Cancer Center Integrative Medicine www.mskcc.org/aboutherbs	Brief descriptions of popular herbs/supplements (free)
	National Center for Complementary and Alternative Medicine http://nccam.nih.gov/index.htm	Describes ongoing CAM research; contains links to governmental resources on popular modalities and supplements (free)
	United States Pharmacopeia (USP) www.usp.org	Dietary supplement verification program identifies ingredients and absence of contaminants of specific brands that undergo testing (free)
	ConsumerLab.com www.consumerlab.com	Tests individual supplements and reports on purity and accuracy of labeling ($24/year)
	National Cancer Institute PDQ Cancer Information Summaries: Complementary and Alternative Medicine www.cancer.gov/cancertopics/pdq/cam	Overviews of selected CAM products used in cancer treatment (free)
Print	Essentials of Complementary and Alternative Medicine. Jonas W, Levin J, Eds. Lippincott Williams & Wilkins, Philadelphia, 1999	Concise overview of popular CAM therapies
	Herbal Medicine: Expanded Commission E Monographs. Blumenthal M, Goldberg A, Brinckmann J, Eds. American Botanical Council, Austin, Texas, 2000	Expanded English translation of the German compendium on herbs
	Evidence-Based Herbal Medicine. Rotblatt R, Ziment I, Eds. Hanley & Belfus, Philadelphia, 2001	Pocket-sized handbook that contains a summary of clinical trials as well as a rating of efficacy, evidence, and safety levels of common herbs

These steps should serve to strengthen the physician-patient relationship while limiting the potential for adverse outcomes.

Specific emphasis should be placed on the role of dietary supplements, which pose a risk of significant toxicity and drug interactions. To ensure patient safety, the medication history should include specific questioning about what vitamins, herbs, or other supplements the person is taking. Unfortunately, supplements are often sold as combination products that are identified only by their catchy trade names. Patients should be encouraged to bring in all new medications and supplements at each visit. Depending on their side-effect profile or potential for drug interactions, certain supplements should be discontinued in the perioperative period.[70] Finally, any suspected adverse reactions or drug-supplement reactions should be reported to the FDA's MedWatch program at their web site (http://www.fda.gov/medwatch) or by calling them at 1-800-FDA-1088.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Barnes PM, Powell-Griner E, McFann K, et al: Complementary and alternative medicine use among adults: United States, 2002. Advance data from vital and health statistics; no 343. Hyattsville, Maryland: National Center for Health Statistics, 2004 http://www.cdc.gov/nchs/data/ad/ad343.pdf

2. Fisher P, Ward A: Medicine in Europe: complementary medicine in Europe. BMJ 309:107, 1994

3. Bausell RB, Lee WL, Berman BM: Demographic and health-related correlates of visits to complementary and alternative medical providers. Med Care 39:190, 2001

4. Population-based survey of complementary and alternative medicine usage, patient satisfaction, and physician involvement. South Carolina Complementary Medicine Program Baseline Research Team. South Med J 93:375, 2000

5. Eisenberg DM, Kessler RC, Foster C, et al: Unconventional medicine in the United States. N Engl J Med 328:246, 1993

6. Astin JA: Why patients use alternative medicine. JAMA 279:1548, 1998

7. Bernstein BJ, Grasso J: Prevalence of complementary and alternative medicine use in cancer patients. Oncology 15:1267, 2001

8. Sparber A, Wootton JC, Bauer L, et al: Use of complementary medicine by adult patients participating in HIV/AIDS clinical trials. J Altern Complement Med 6:415, 2000

9. Pioro-Boisset M, Esdaile JM, Fitzcharles MA: Alternative medicine use in fibromyalgia syndrome. Arthritis Care Res 9:13, 1996

10. Rawsthrone P, Shanahan F, Cronin NC, et al: An international survey of the use and attitudes regarding alternative medicine by patients with inflammatory bowel disease. Am J Gastroenterol 94:1298, 1999

11. Campion EW: Why unconventional medicine? N Engl J Med 328:282, 1993

12. Eisenberg DM, Kessler RC, Van Rompay MI, et al: Perceptions about complementary therapies relative to conventional therapies among adults who use both: results from a national survey. Ann Intern Med 135:344, 2001

13. Eisenberg DM, Davis RB, Ettner SL, et al: Trends in alternative medicine use in the United States, 1990–1997. JAMA 280:1569, 1998

14. Blendon RJ, DesRoches CM, Benson JM, et al: Americans' views on the use and regulation of dietary supplements. Arch Intern Med 161:805, 2001

15. Bloom BS, Retbi A, Dahan S, et al: Evaluation of randomized controlled trials on complementary and alternative medicine. Int J Technol Assess Health Care 16:13, 2000

16. Linde K, Jonas WB, Melchart D, et al: The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. Int J Epidemiol 30:526, 2001

17. Vickers A, Goyal N, Harland R, et al: Do certain countries produce only positive results? A systematic review of controlled trials. Control Clin Trials 19:159, 1998

18. Tang JL, Zhan SY, Ernst E: Review of randomized controlled trials of traditional Chinese medicine. BMJ 319:160, 1999

19. Melchart D, Linde K, Fischer P, et al: *Echinacea* for preventing and treating the common cold (review). Cochrane Database Syst Rev (2):CD000530, 2000

20. Lee JD, Chon JS, Jeong HK, et al: The cerebrovascular response to traditional acupuncture after stroke. Neuroradiology 45:780, 2003

21. Sandberg M, Lindberg LG, Gerdle B: Peripheral effects of needle stimulation (acupuncture) on skin and muscle blood flow in fibromyalgia. Eur J Pain 8:163, 2004

22. Han JS: Acupuncture and endorphins. Neurosci Lett 361:258, 2004

23. Acupuncture. NIH Consensus Statement 15(5):1, 1997

24. Linde K, Melchart D, Fischer P, et al: Acupuncture for idiopathic headache (review). Cochrane Database Syst Rev (1):CD001218, 2001

25. Berman BM, Ezzo J, Hadhazy V, et al: Is acupuncture effective in the treatment of fibromyalgia? J Fam Pract 48:213, 1999

26. Holdcraft LC, Assefi N, Buchwald D: Complementary and alternative medicine in fibromyalgia and related syndromes. Best Pract Res Clin Rheumatol 17:667, 2003

27. Ezzo J, Hadhazy V, Birch S, et al: Acupuncture for osteoarthritis of the knee: a systematic review. Arthritis Rheum 44:819, 2001

28. White AR, Rampes H, Ernst E: Acupuncture for smoking cessation (review). Cochrane Database Syst Rev (2):CD000009, 2000

29. McCarney RW, Brinkhaus B, Lasserson TJ, et al: Acupuncture for chronic asthma. Cochrane Database Syst Rev (1):CD000008, 2004

30. Cherkin DC, Sherman KJ, Deyo RA, et al: A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. Ann Intern Med 138:898, 2003

31. Lao L, Hamilton GR, Fu J, et al: Is acupuncture safe? A systematic review of case reports. Altern Ther Health Med 9:72, 2003

32. Ernst E, White A: Life-threatening adverse reactions after acupuncture? A systematic review. Pain 71:123, 1997

33. Yamashita H, Tsukayama H, Tanno Y, et al: Adverse events related to acupuncture. JAMA 280:1563, 1998

34. Ernst E, White AR: Prospective studies of the safety of acupuncture: a systematic review. Am J Med 110:481, 2001

35. Chapman EH: Homeopathy. Essentials of Complementary and Alternative Medicine. Jonas W, Levin JS, Eds. Lippincott Williams & Wilkins, Philadelphia, 1999, p 472

36. Kleijnen J, Knipschild P, ter Riet G: Clinical trials of homeopathy. BMJ 302:316, 1991

37. Linde K, Clausius N, Ramirez G, et al: Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. Lancet 350:834, 1997

38. Jonas WB, Kaptchuk TJ, Linde K: A critical overview of homeopathy. Ann Intern Med 139:393, 2003

39. Ernst E: A systematic review of systematic reviews of homeopathy. Br J Clin Pharmacol 54:577, 2002

40. Stehlin I: Homeopathy: real medicine or empty promises? FDA Consumer 30, 1996 http://www.fda.gov/fdac/096_toc.html

41. Cannon WB: The emergency function of the adrenal medulla in pain and the major emotions. Am J Physiol 33:356, 1914

42. Benson H: The Relaxation Response. William Morrow, New York, 1975

43. Astin JA, Harkness E, Ernst E: The efficacy of "distant healing": a systematic review of randomized trials. Ann Intern Med 132:903, 2000

44. Rados C: Ephedra ban: no shortage of reasons. FDA Consum 38:6, 2004

45. Nortier JL, Martinez MC, Schmeiser HH, et al: Urothelial carcinoma associated with the use of a Chinese herb (*Aristolochia fangchi*). N Engl J Med 342:1686, 2000

46. Ko RJ: Adulterants in Asian patent medicines. N Engl J Med 339:847, 1998

47. Fugh-Berman A: Herb-drug interactions. Lancet 355:134, 2000

48. Ashar BH, Miller RG, Getz KJ, et al: A critical evaluation of Internet marketing of products containing ephedra. Mayo Clin Proc 78:944, 2003

49. Meeker WC, Haldeman S: Chiropractic: a profession at the crossroads of mainstream and alternative medicine. Ann Intern Med 136:216, 2002

50. Lawrence DJ: Chiropractic medicine. Essentials of Alternative and Complementary Medicine. Jonas WB, Levin JS, Eds. Lippincott Williams & Wilkins, Philadelphia, 1999, p 275

51. Hondras MA, Linde K, Jones AP: Manual therapy for asthma. Cochrane Database Syst Rev (4):CD001002, 2002

52. Ernst E: Chiropractic manipulation for non-spinal pain: a systematic review. N Z Med J 116:U539, 2003

53. Ernst E: Chiropractic spinal manipulation for neck pain: a systematic review. J Pain 4:417, 2003

54. Astin JA, Ernst E: The effectiveness of spinal manipulation for the treatment of headache disorders: a systematic review of randomized clinical trials. Cephalgia 22:617, 2002

55. Assendelft WJ, Morton SC, Yu EI, et al: Spinal manipulative therapy for low back pain: a meta-analysis of effectiveness relative to other therapies. Ann Intern Med 138:871, 2003

56. Cherkin DC, Deyo RA, Battie M, et al: A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. N Engl J Med 339:1021, 1998

57. Kaptchuk TJ, Eisenberg DM: Chiropractic: origins, controversies, and contributions. Arch Intern Med 158:2215, 1998

58. Senstad O, Leboeuf-Yde C, Borchgrevink C: Frequency and characteristics of side effects of spinal manipulative therapy. Spine 22:435, 1997

59. Ernst E: Adverse effects of spinal manipulation. Essentials of Complementary and Alternative Medicine. Jonas WB, Levin JS, Eds. Lippincott Williams & Wilkins, Philadelphia, 1999, p 176

60. Field T: Massage therapy. Med Clin North Am 86:163, 2002

61. Lee CT, Lei T: Qigong. Essentials of Complementary and Alternative Medicine. Jonas W, Levin JS, Eds. Lippincott Williams & Wilkins, Philadelphia, 1999, p 392

62. Ng BY: Qigong-induced mental disorders. Aust N Z J Psychiatry 33:197, 1999

63. O'Connor D. Marshall S, Massy-Westropp N: Non-surgical treatment (other than steroid injection) for carpal tunnel syndrome. Cochrane Database Syst Rev (1):CD003219, 2003

64. Cooper S, Oborne J, Newton S, et al: Effect of two breathing exercises (Buteyko and pranayama) in asthma: a randomised controlled trial. Thorax 68:674, 2003

65. Wang C, Collet JP, Lau J: The effect of Tai Chi on health outcomes in patients with chronic conditions: a systematic review. Arch Intern Med 164:493, 2004

66. Wolf SL, Sattin RW, Kutner M, et al: Intense Tai Chi training and fall occurrences in older, transitionally frail adults: a randomized, controlled trial. J Am Geriatr Soc 51:1693, 2003

67. Krieger D: Accepting Your Power to Heal: The Personal Practice of Therapeutic Touch. Bear & Company, Rochester, Vermont, 1993

68. O'Mathuna: Evidence-based practice and reviews of therapeutic touch. J Nurs Scholarsh 32:279, 2000

69. O'Mathuna DP, Ashford RL: Therapeutic touch for healing acute wounds. Cochrane Database Syst Rev (4):CD002766, 2003

70. Ang-Lee MK, Moss J, Yuan CS: Herbal medicines and perioperative care. JAMA 286:208, 2001

71. Cooke B, Ernst E: Aromatherapy: a systematic review. Br J Gen Pract 50:493, 2000

72. Hinninghofen H, Enck P: Fecal incontinence: evaluation and treatment. Gastroenterol Clin North Am 32:685, 2003

73. Barrows KA, Jacobs BP: Mind-body medicine: an introduction and review of the literature. Med Clin North Am 86:11, 2002

74. Roberts L, Ahmed I, Hall S: Intercessory prayer for the alleviation of ill health. Cochrane Database Syst Rev (2):CD000368, 2000

75. Cassileth BR, Vickers AJ, Magill LA: Music therapy for mood disturbance during hospitalization for autologous stem cell transplantation: a randomized controlled trial. Cancer 98:2723, 2003

76. Smyth JM, Stone AA, Hurewitz A, et al: Effects of writing about stressful experiences on symptom reduction in patients with asthma or rheumatoid arthritis. JAMA 281:1304, 1999

77. Petrie KJ, Fontanilla I, Thomas MG, et al: Effect of written emotional expression on immune function in patients with human immunodeficiency virus infection: a randomized trial. Psychosom Med 66:272, 2004

78. Kronenberg F, Fugh-Berman A: Complementary and alternative medicine for menopausal symptoms: a review of randomized, controlled trials. Ann Intern Med 137:805, 2002

79. Fugh-Berman A, Kronenberg F: Complementary and alternative medicine in reproductive-age women: a review of randomized controlled trials. Reprod Toxicol 17:137, 2003

80. Terris MK, Issa MM, Tacker JR: Dietary supplementation with cranberry concentrate tablets may increase the risk of nephrolithiasis. Urology 57:26, 2001

81. Suvarna R, Pirmohamed M, Henderson L: Possible interaction between warfarin and cranberry juice. BMJ 327:1454, 2003

82. Jepson RG, Mihaljevic L, Craig J: Cranberries for preventing urinary tract infections (review.) Cochrane Database Syst Rev (4):CD001321, 2004

83. Melchart D, Linde K, Fischer P, et al: Echinacea for preventing and treating the common cold. Cochrane Database Syst Rev (2):CD000530, 2000

84. Miller LG: Herbal medicinals: selected clinical considerations focusing on known or potential drug-herb interactions. Arch Intern Med 158:2200, 1998

85. Pittler MH, Ernst E: Feverfew for preventing migraine. Cochrane Database Syst Rev (1):CD002286, 2004

86. Ackermann RT, Mulrow CD, Ramirez G, et al: Garlic shows promise for improving some cardiovascular risk factors. Arch Intern Med 161:813, 2001

87. Vutyavanich T, Draisarin T, Ruangsri R: Ginger for nausea and vomiting in pregnancy: randomized, double-masked, placebo-controlled trial. Obstet Gynecol 97:577, 2001

88. LeBars PL, Katz MM, Berman N, et al: A placebo-controlled, double-blind, randomized trial of an extract of Ginkgo biloba for dementia. JAMA 278:1327, 1997

89. Ernst E: The risk-benefit profile of commonly used herbal therapies: ginkgo, St. John's wort, ginseng, Echinacea, saw palmetto, and kava. Ann Intern Med 136:42, 2002

90. Pittler MH, Ernst E: Kava extract for treating anxiety. Cochrane Database Syst Rev (1):CD003383, 2003

91. Jacobs BP, Dennehy C, Ramirez G, et al: Milk thistle for the treatment of liver disease: a systematic review and meta-analysis. Am J Med 113:506, 2002

92. Wilt TJ, Ishani A, Stark G, et al: Saw palmetto extracts for the treatment of benign prostatic hyperplasia: a systematic review. JAMA 280:1604, 1999

93. Gaster B, Holroyd J: St John's wort for depression: a systematic review. Arch Intern Med 160:152, 2000

94. Stevinson C, Ernst E: Valerian for insomnia: a systematic review of randomized clinical trials. Sleep Med 1:91, 2000

95. Shekelle P, Morton SC, Hardy M, et al: Effect of supplemental antioxidants vitamin C, vitamin E, and coenzyme Q10 for the prevention and treatment of cardiovascular disease. Evid Rep Technol Assess June (83):1, 2003

96. Shults CW, Oakes D, Kieburtz K, et al: Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Arch Neurol 59:1541, 2002

97. McAlindon TE, LaValley MP, Gulin JP, et al: Glucosamine and chondroitin for treatment of osteoarthritis. JAMA 283:1469, 2000

98. Richy F, Bruyere O, Ethgen O, et al: Structural and symptomatic efficacy of glucosamine and chondroitin in knee osteoarthritis: a comprehensive meta-analysis. Arch Intern Med 163:1514, 2003

99. Herxheimer A, Petrie KJ: Melatonin for the prevention and treatment of jet lag. Cochrane Database Syst Rev (2):CD001520, 2002

100. Hardy M, Coulter I, Morton SC, et al: S-adenosyl-L-methionine for treatment of depression, osteoarthritis, and liver disease. Evidence Report/Technology Assessment no. 64. AHRQ publication no. 02-E034, 2002

XIII PERFORMANCE MEASUREMENT IN CLINICAL PRACTICE

Stephen D. Persell, m.d., m.p.h.
David W. Baker, m.d., m.p.h., f.a.c.p.
Kevin B. Weiss, m.d., m.p.h., f.a.c.p.

Performance in health care is the degree to which desirable objectives are accomplished. Performance measurement can inform quality-improvement activities and allow health care consumers and commercial health care purchasers to hold physicians and health care organizations accountable for the services they provide. Over the past decade, the methodology supporting performance measurement has matured. With this maturity has come an increasing array of performance measures covering a range of care settings and specialties; the increasing information has given rise to an expanded interest in the application of performance measurement. The Centers for Medicare and Medicaid Services (CMS) and large commercial health care purchasers now pay close attention to publicly released performance data. At present, the vast majority of patients in the United States who are enrolled in managed care organizations or Medicare receive health care from health plans, hospitals, nursing homes, or ambulatory care centers for which publicly reported performance data are available; the trend toward the public release of performance information is likely to accelerate in coming years. Physicians may increasingly find themselves the targets of efforts to profile the care provided by group practices or individual physicians. It is likely that the trend toward performance measurement will have increasingly noticeable effects on the ways in which health care is delivered and physicians are compensated.[1]

Why Measure Performance?

Ideally, individual health care professionals would be able to deliver high-quality care as a matter of course, and performance measurement would be unnecessary. Unfortunately, the best practices often are not followed, and patients frequently do not receive indicated services for acute problems, chronic illnesses, and preventive health care.[2] Measurement of clinical performance is needed to assess the quality of care and to compare what is achieved with what is desired. This information can then serve the related but distinct purposes of internal and external quality review. First, measurement can enable internal quality improvement—in other words, deficiencies in health care can be identified and physicians and health care organizations can implement changes to improve quality and track their progress.[3] Second, the public release of performance information can increase accountability of health care providers by (1) allowing health care purchasers and consumers to make informed health care choices and (2) permitting regulators responsible for licensure or accreditation to evaluate the quality of care offered by specific health care organizations. Both purposes figure prominently in the Institute of Medicine's Strategic Framework for a national quality measurement and reporting system.[4]

Performance Measures

In the past, performance measurement was often conducted through the use of implicit review to assess the quality of care provided. In this method, a physician reviewed patient records and judged whether appropriate care had been provided. This approach to performance measurement is poorly reproducible because implicit review involves a level of subjectivity, and physician reviewers frequently do not agree.[5] Currently, the most widely used quality measures define explicit criteria against which performance is judged. Although no performance measure is ideal, well-developed measures share three characteristics[6]: (1) they are based on strong clinical evidence; (2) they depict uncontroversial clinical practices that have broad consensus among physicians; and (3) they incorporate agreed-upon standards for determining satisfactory performance.

For performance measurement to be undertaken, it must be feasible to collect the necessary data. Methods of data collection that are too burdensome limit a measure's utility. When used for comparisons of health care providers, performance measures should be attributable to the physician or organization being assessed. Furthermore, sufficient numbers of patients should be assessed to allow for meaningful statistical comparisons; if required, statistical adjustments should be made to account for confounding variables in the collected data.[6] Several measures are used to assess the quality of health care; these include measures of clinical performance, measures of patient experience, and measures of efficiency.

Measures of Clinical Performance

Measures of clinical performance can assess health care structures, processes, and outcomes.[7] Structural measures are not measures of clinical performance per se; rather, they describe characteristics of physicians, hospitals, or other health care organizations. Structural characteristics are selected for assessment because they are perceived to be associated with favorable clinical outcomes. Examples of structural measures include the ratio of nurses to patients in the hospital,[8] whether a board-certified critical care physician is available in the intensive care unit of a hospital 24 hours a day,[9] whether a hemodialysis center is operated as a not-for-profit facility,[10] and whether a hospital has a computerized drug-order entry system.[11,12]

Process measures assess specific components of the encounters between physicians and patients: Was a screening test for colon cancer obtained? Was a beta blocker prescribed to a patient with a myocardial infarction? Did a patient who was hospitalized for a mental illness receive an outpatient appointment promptly after discharge?[13]

Outcome measures are direct assessments of patients' health status. Examples include whether a nursing-home patient has a pressure ulcer and whether death occurs within 30 days after coronary artery bypass graft surgery.[14,15]

Structure and process measures have some advantages over outcome measures. Because they are to a greater degree under the control of a single health care organization or individual physician, structure and process measures are less likely than outcome measures to require adjustment for confounding differences between patient groups—a statistical correction referred to as case-mix adjustment, or risk adjustment [*see* Case-Mix Adjustment, *below*]. In addition, structural measures often require very little data collection, as compared with other measures.

Conversely, a limitation of both structure and process measures is that they need to be causally related to desirable health outcomes to be valid measures of health care quality. If structure or process measures are not directly related to desirable health outcomes, efforts made to improve these measures may merely increase the costs of care without improving patients' health status. An advantage of outcome measures is that they directly measure the ultimate objectives of health care—clinical outcomes; however, they are often dependent on the characteristics and actions of individual patients. Observed differences in outcomes may be largely driven by factors that are not under the control of physicians or health care organizations. Patients who receive excellent care may still have bad outcomes, and physicians caring for patients of lower socioeconomic status or educational attainment, who are known to be at higher risk for disease severity and who lack access to care because of mechanisms such as cost sharing of copayments and high deductibles, may falsely appear to be providing inferior care.[16-19]

PATIENT EXPERIENCE AND SATISFACTION MEASURES

Although expert-derived clinical measures may best assess the technical aspects of care, consumers of health care are in the best position to evaluate their own experience and level of satisfaction with the services they receive. Furthermore, care that is timely and well received by patients may lead to better health outcomes. Examples of patient experience measures include the ease with which medical advice can be obtained by phone, the number of times a patient must wait more than 30 minutes past an appointment time to see a physician, and the rating of a personal physician on a scale of 0 to 10.[20] The Consumer Assessment of Health Plans Survey (CAHPS) is a widely used series of surveys to assess patients' experience of health care.[21] CAHPS captures patient satisfaction with the delivery of care occurring during office visits, the level of assistance obtained from health-plan customer service, the perceived accessibility and timeliness of medical and reimbursement services, and the clarity and timeliness of health advice provided by the physician. CAHPS is one form of evaluation used by the National Committee for Quality Assurance (NCQA) to evaluate health plans.[14]

EFFICIENCY MEASURES

Efficiency measures focus on the costs of delivering health care. When combined with measures of clinical performance, efficiency measures assess the value of health care (i.e., the quality of care delivered per unit cost). The methodology for assessing efficiency is less well developed than that of quality measurement.[22] Common units of efficiency measurement (e.g., cost of health plan per member per month and cost per episode of care) do not capture the quality of the care received for that cost.

From Quality Improvement to Accountability

Performance measurement can support quality improvement in several ways. Measuring performance and relaying this information to the physicians who were assessed can produce beneficial, albeit modest, improvements in quality.[23] Within health care organizations, reliable measurements of quality are necessary to assess the impact of quality improvement initiatives such as clinical-reminder systems, disease-management programs, clinical-decision support systems, patient-directed programs, and multimodality interventions. Because they are in a position to implement system-level changes, health care organizations may be in a better position than individual physicians to act on performance data and institute quality improvement.[24,25]

Using performance measurement to hold physicians and organizations accountable for the care they provide is a very different undertaking from that of quality improvement. Although publicly released reports on quality performance may prompt physicians to improve performance, they are intended for an audience external to the health care team.[26] The goal of measurement for accountability is to enable comparisons of different health care organizations. For this reason, the measures used in assessing accountability must be standardized across sites. In the United States, private organizations such as the NCQA, the Quality Consortium of the American Medical Association, the Joint Commission on Accreditation of Healthcare Organizations, and the National Quality Forum (NQF) have taken leading roles in the development of standardized performance measures and the dissemination of performance results. Government agencies, such as CMS, are also a major source of performance data. Consumers and commercial health care purchasers can find comparative information on health plans, hospitals, nursing homes, and hemodialysis centers on the Internet or through health insurers.[13-15]

The use of publicly released performance data is not limited to consumers or purchasers of health care. Increasingly, physicians may find their remuneration influenced by their measured performance. The Institute of Medicine has issued a call for government health care payers to reward high-quality care with increased payments.[1] Nongovernmental organizations also are adopting pay-for-performance (PFP or P4P) plans. For example, the Integrated Healthcare Association, a California consortium of medical groups and health plans, distributed $50 million in bonuses to health care providers on the basis of performance in the first year of the consortium's pay-for-performance initiative.[1,27] In the United Kingdom, the government has instituted an ambitious pay-for-performance initiative in which family practitioners will receive from the government an additional £1 billion ($1.8 billion) more than they ordinarily would have received; this represents a 20% increase over the previous year's family-practice budget. The additional funding will be distributed to family practitioners on the basis of a combination of three factors: their performance on a variety of clinical measures, organizational indicators, and patient surveys.[28] Both positive and negative consequences of pay-for-performance have been predicted,[28] but the full impact of the program remains to be seen.

Aside from influencing payments, performance measurement can also be incorporated into the accreditation process. For example, the American Board of Internal Medicine has incorporated a module on self-assessment of the quality of care provided by the physician as an option by which to meet the 2006 practice-assessment requirement for recertification in internal medicine.[29]

Methodological Issues

As the use of performance measurement increases, continued awareness of its limitations is vital. There remain both methodological and practical limitations to successful performance measurement.

FACTORS AFFECTING DATA UTILITY

Data Collection

Collecting the data required for performance measurement can be labor-intensive and expensive. Measures that rely on data

collected for other purposes (e.g., administrative data) are not burdensome to implement but may not be as reliable as data collected explicitly for quality measurement. Expanded use of clinical computer systems and electronic health records is one way to facilitate the collection of clinical data of high quality,[28] but the implementation of new clinical computer systems is costly.

Use of publicly released data to inform consumer choice or to increase provider accountability also has its drawbacks. One significant problem is that publicly available data may include falsified results. To help maintain the accuracy of data reported to such organizations as the NCQA, auditing is required. Some falsification, however, may be undetectable by audits (e.g., a physician's falsely recording a blood pressure measurement that is below a quality goal). The potential exists for physicians to adjust their performance measures by the selection or dismissal of patients. For example, physicians may dismiss nonadherent or outlier patients (i.e., patients whose data lie far outside the central statistical mass) to improve their measured performance.[30] Physicians who will be judged by clinical outcomes such as mortality following coronary artery bypass graft surgery may avoid taking on patients at high risk for bad outcomes. In one instance, following the release of publicly reported cardiac surgery mortality, cardiologists reported that it became harder to find surgeons willing to operate on high-risk patients who needed surgery.[31]

Sample Size

Another potential problem in performance measurement is insufficient sample size. For any single measure, there may be too few patients for statistically meaningful comparisons, either between providers or between the performances of a single provider over time. An insufficient sample size is frequently a factor when individual physicians are profiled. When statistically meaningful differences cannot be detected because of limited sample size, those who report data should not suggest that providers with statistically insignificant differences in quality scores differ from one another. Combining several years of data is one way to compensate for insufficient sample size; however, this reporting method makes the reported results less timely and may obscure improvements in quality that occur in the short term.

Case-Mix Adjustment

The differences in patient characteristics (referred to as case mix) can introduce systematic biases into quality measurements; case-mix adjustment to correct these biases is an important methodological consideration in the utilization of performance data. Apparent differences in the quality of care received by patients cared for by different physicians may disappear when differences in patients' socioeconomic status or education are taken into account.[16-19] Case-mix adjustment is especially important for measures of clinical outcome or cost efficiency, but it may alter the interpretation of data pertaining to process measures.

Statistical methods can be used to improve the validity of statistical inference when sample sizes are small or when the case mix differs across providers; however, it is not clear whether the methodology to correct for these variables will be generally adopted or, if adopted, applied in uniform ways.[32]

Optional Reporting

Public reporting of performance data may be of limited use to health care purchasers and consumers if poor performers can choose not to report their data. It has been noted that some health plans that participated in the NCQA's Health Plan Employer Data and Information Set (HEDIS) withdrew from HEDIS when they performed poorly.[33]

Internet Resources for Performance Measures

General Information

Agency for Healthcare Research and Quality (AHRQ)
http://www.ahrq.gov/clinic/epcix.htm
A governmental agency and information clearinghouse on issues of quality, safety, efficiency, and effectiveness of health care

Ambulatory Care Quality Alliance (AQA)
http://www.ahrq.gov/qual/aqaback.htm
Initially convened by the American Academy of Family Physicians, American College of Physicians, America's Health Insurance Plans, and AHRQ, the Alliance consists of a large body of stakeholders, including physicians, consumers, health care purchasers, and health plans

Centers for Medicare and Medicaid Services (CMS)
http://www.cms.hhs.gov/quality/hospital
Provides hospital quality information to consumers and others in initiatives designed to improve hospital care in the United States

National Committee for Quality Assurance (NCQA)
http://www.ncqa.org
An independent nonprofit organization that provides information on the quality of managed care plans in the United States

National Quality Forum (NQF)
http://www.qualityforum.org
A private, not-for-profit membership organization created to develop and implement a national strategy for measuring and reporting health care quality

Measures and Tools

National Quality Measures Clearinghouse
http://www.qualitymeasures.ahrq.gov
Public repository for evidence-based quality measures and measure sets

Measuring Healthcare Quality
http://www.ahrq.gov/qual/measurix.htm
Includes National Health Care Quality reports, AHRQ quality indicators, and ambulatory care clinical-performance measures

National Voluntary Consensus Standards for Ambulatory Care
http://www.qualityforum.org/txWEBambreport04-29-05.pdf
Appendix A contains physician standards for ambulatory care pertaining to asthma and respiratory conditions, heart disease, hypertension, prenatal care, and prevention

Quality Profiles
http://www.qualityprofiles.org/quality_profiles/index.asp
Case studies that examine performance measures in such topics as chronic illness, women's health, preventive care, and behavioral health

Quality Improvement Initiative Tools
http://www.qualityprofiles.org/qia_Tools/index.asp
Templates of the tools used by health plans to help implement their health care improvement initiatives

Physician Quality Reports
http://www.ncqa.org/PhysicianQualityReports.htm
Includes measures used to qualify physicians for the Diabetes Physician Recognition Program, the Heart/Stroke Physician Recognition Program, and Physician Practice Connections

Hospital Quality Measures
http://www.qualityindicators.ahrq.gov
Presents hospital quality measures, inpatient quality indicators, and prevention quality indicators

Inadequate Use of Data

Even when public reports of health care quality are available, purchasers and consumers may not put them to good use.[34] Consumers often do not access the performance data that are available and may have difficulty making sense of the information when they find it.[25,35,36]

METHODS TO IMPROVE PERFORMANCE MEASURES

Although all these methodological issues and limitations are formidable, the organizations developing and promoting such measurement seek to identify and promote only those measures meeting clear criteria. An example of this move toward standardization is the recent development by the NQF of a set of voluntary consensus criteria by which to endorse physician-focused ambulatory care performance measures. The NQF's consensus standards include criteria to judge the importance, scientific acceptability, usability, and feasibility of proposed performance measures.[37]

Issues for the Physician

Although encumbered by methodological and practical issues, performance measurement is a reality of modern medical care. Experts caution physicians that they must be leaders in quality measurement and performance improvement or be vulnerable to challenge by economic or political stakeholders and to the potential loss of patient confidence.[38] Because it is likely that we physicians will increasingly be held accountable for the quality of care we provide, it is to our advantage to have a working knowledge of current performance measurement activities in our specialty [*see Sidebar* Internet Resources for Performance Measures]. By understanding the types of measures by which we will be judged and knowing how the data are collected, we may be able to improve our performance—as well as the documentation of the data needed for proper measurement of that performance—without greatly increasing our work.

When possible, individual physicians should assess the quality of the care they provide and correct their own deficiencies. They should consider making systematic changes that facilitate quality measurement and internal quality improvement, such as making use of disease registries, flow sheets, or computer systems, to track the quality of chronic disease management and preventive care. It is unlikely that the costs of implementing these changes will be recovered in the short term; however, if pay-for-performance becomes widespread, physicians and health care organizations that have not prepared for it by routinely measuring their own quality of care will find themselves at a distinct disadvantage in comparison with their competitors who have.

References

1. Epstein AM, Lee TH, Hamel MB: Paying physicians for high-quality care. N Engl J Med 350:406, 2004

2. McGlynn EA, Asch SM, Adams J, et al: The quality of health care delivered to adults in the United States. N Engl J Med 348:2635, 2003

3. Kiefe CI, Allison JJ, Williams OD, et al: Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA 285:2871, 2001

4. Crossing the Quality Chasm: A New Health System for the 21st Century. Committee on Quality Health Care in America, Institute of Medicine. National Academy Press, Washington, D.C., 2001

5. Hayward RA, McMahon LF Jr, Bernard AM: Evaluating the care of general medicine inpatients: how good is implicit review? Ann Intern Med 118:550, 1993

6. Landon BE, Normand SL, Blumenthal D, et al: Physician clinical performance assessment: prospects and barriers. JAMA 290:1183, 2003

7. Brook RH, McGlynn EA, Cleary PD: Quality of health care. Part 2: Measuring quality of care. N Engl J Med 335:966, 1996

8. Needleman J, Buerhaus P, Mattke S, et al: Nurse-staffing levels and the quality of care in hospitals. N Engl J Med 346:1715, 2002

9. Pronovost PJ, Angus DC, Dorman T, et al: Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. JAMA 288:2151, 2002

10. Garg PP, Frick KD, Diener-West M, et al: Effect of the ownership of dialysis facilities on patients' survival and referral for transplantation. N Engl J Med 341:1653, 1999

11. Bates DW, Leape LL, Cullen DJ, et al: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. JAMA 280:1311, 1998

12. Computer Physician Order Entry Fact Sheet: The Leapfrog Group for Patient Safety, Washington, D.C., 2004
http://www.leapfroggroup.org/media/file/Leapfrog-Computer_Physician_Order_Entry_Fact_Sheet.pdf

13. Follow-up after hospitalization for mental illness. State of Health Care Quality Report, 2003. National Committee for Quality Assurance, Washington, D.C., 2003
http://www.ncqa.org/sohc2003/follow_up_after_hospitalization.htm

14. Nursing Home Compare: Medicare, the official U.S. government site for people with Medicare, 2005
http://www.medicare.gov/NHCompare

15. Hospital report cards: mortality and complication based outcomes, 2005. Hospital methodologies. Healthgrades, Golden, Colorado
http://www.healthgrades.com

16. Zaslavsky AM, Hochheimer JN, Schneider EC, et al: Impact of sociodemographic case mix on the HEDIS measures of health plan quality. Med Care 38:981, 2000

17. Greenfield S, Kaplan SH, Kahn R, et al: Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. Ann Intern Med 136:111, 2002

18. Fiscella K, Franks P: Influence of patient education on profiles of physician practices. Ann Intern Med 131:745, 1999

19. Franks P, Fiscella K: Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. Med Care 40:717, 2002

20. Zaslavsky AM, Beaulieu ND, Landon BE, et al: Dimensions of consumer-assessed quality of Medicare managed-care health plans. Med Care 38:162, 2000

21. Crofton C, Lubalin JS, Darby C: Consumer Assessment of Health Plans Study (CAHPS). Foreword. Med Care 37(3 suppl):MS1-9, 1999

22. Provider Efficiency White Paper: Measuring Provider Efficiency. Version 1.0. Bridges to Excellence and the Leapfrog Group for Patient Safety, Washington, D.C., 2004
http://www.bridgestoexcellence.org/bte/white_paper_release.htm

23. Thomson O'Brien MA, Oxman AD, Davis DA, et al: Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev (2):CD000259, 2000

24. Berwick DM, James B, Coye MJ: Connections between quality measurement and improvement. Med Care 41(1 suppl):I30, 2003

25. Marshall MN, Shekelle PG, Leatherman S, et al: The public release of performance data: what do we expect to gain? A review of the evidence. JAMA 283:1866, 2000

26. Solberg LI, Mosser G, McDonald S: The three faces of performance measurement: improvement, accountability, and research. Jt Comm J Qual Improv 23:135, 1997

27. Pay-for-performance takes off in California. ACP Observer Online. January-February 2005
http://www.acponline.org/journals/news/jan05/pfp.htm

28. Roland M: Linking physicians' pay to the quality of care: a major experiment in the United Kingdom. N Engl J Med 351:1448, 2004

29. Self-evaluation of practice performance. Maintenance of Certification. American Board of Internal Medicine, Philadelphia, 2005
http://www.abim.org/moc/sempbpi.shtm

30. Hofer TP, Hayward RA, Greenfield S, et al: The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. JAMA 281:2098, 1999

31. Schneider EC, Epstein AM: Influence of cardiac-surgery performance reports on referral practices and access to care: a survey of cardiovascular specialists. N Engl J Med 335:251, 1996

32. Zaslavsky AM: Statistical issues in reporting quality data: small samples and casemix variation. Int J Qual Health Care 13:481, 2001

33. McCormick D, Himmelstein DU, Woolhandler S, et al: Relationship between low quality-of-care scores and HMOs' subsequent public disclosure of quality-of-care scores. JAMA 288:1484, 2002

34. Hibbard JH, Jewett JJ, Legnini MW, et al: Choosing a health plan: do large employers use the data? Health Aff (Millwood) 16:172, 1997

35. Schneider EC, Epstein AM: Use of public performance reports: a survey of patients undergoing cardiac surgery. JAMA 279:1638, 1998

36. Hibbard JH, Jewett JJ: Will quality report cards help consumers? Health Aff (Millwood) 16:218, 1997

37. Pre-voting review for "National Voluntary Consensus Standards for Ambulatory Care: An Initial Physician-Focused Measure Set." The National Quality Forum. Washington, D.C., 2005
http://www.qualityforum.org/txWEBambreport04-29-05.pdf

38. Blumenthal D: Quality of care: what is it? Part 1. N Engl J Med 335:891, 1996

I APPROACH TO THE CARDIOVASCULAR PATIENT

Catherine M. Otto, m.d.
David M. Shavelle, m.d.

The complete evaluation of the cardiovascular patient begins with a thorough history and a detailed physical examination. These two initial steps will often lead to the correct diagnosis and assist in excluding life-threatening conditions. The history and physical examination findings should be assessed in the context of the overall clinical status of the patient, including lifestyle, comorbidities, and expectations. Cardiovascular conditions that frequently require evaluation include chest pain, dyspnea, palpitations, syncope, claudication, and cardiac murmurs. Each of these conditions will be discussed separately, with an emphasis on a diagnostic algorithm and the appropriate use of invasive and noninvasive cardiac testing.

Chest Pain

BACKGROUND

Chest pain is perhaps the most common cardiovascular symptom encountered in clinical practice. Establishing a cardiac origin of chest pain in a patient with multiple cardiovascular risk factors is essential because it allows initiation of appropriate therapy, thereby reducing the risk of myocardial infarction and death. Similarly, excluding a cardiac origin of chest pain in a low-risk patient is no less essential to avoid costly and potentially risky diagnostic testing that will neither add to the care of the patient nor relieve the patient's discomfort.[1] Cardiac disorders that result in chest pain include myocardial ischemia, myocardial infarction, acute pericarditis, aortic stenosis, hypertrophic cardiomyopathy, and aortic dissection. Noncardiac disorders that may result in chest pain include pulmonary embolism, pneumonia, pleural effusion, reactive airway disease, gastrointestinal and biliary disease, anxiety, and musculoskeletal disorders.

Angina most frequently is caused by atherosclerosis of the coronary arteries. Less common causes of angina include coronary artery spasm (e.g., Prinzmetal angina or spasm secondary to drug use, as with cocaine), coronary artery embolism (from aortic valve endocarditis), congenital coronary anomalies, spontaneous coronary artery dissection, coronary arteritis, and aortic dissection when the right coronary artery is involved. Angina may also occur in the presence of angiographically normal coronary arteries and is referred to as syndrome X. The underlying pathophysiology is thought to be related to microvascular dysfunction; the prognosis is generally good despite frequent episodes of chest pain.[2]

HISTORY AND PHYSICAL EXAMINATION

Essential features of the history include an accurate description of the chest pain, including the severity, frequency, location, radiation, quality, alleviating and aggravating factors, and duration of symptoms [*see Table 1*]. Anginal chest pain is often described as pressure or a heavy sensation. Symptoms may be difficult for the patient to describe and may be better characterized as discomfort, not pain. Angina typically is described as substernal with radiation to the left neck, jaw, or arm; is mild to moderate in severity; and lasts for 5 to 15 minutes. Classically, angina occurs with exercise, stress, or exposure to cold weather and is relieved with rest or use of nitroglycerin. Some of the most useful features of the patient history that help establish that chest pain is angina are (1) reproducibility of the pain with a given degree of activity, (2) brief duration, and (3) alleviation of the pain with rest or use of nitroglycerin. In patients with a history of coronary artery disease (CAD), an accurate characterization of the quality and frequency of the pain is essential to determine whether a change in the anginal pattern has occurred (i.e., a patient with chronic stable angina now has unstable angina) or if a noncardiac origin of pain is now present (e.g., a patient with chronic stable angina now has musculoskeletal pain). Elderly patients, diabetic patients, and women experiencing angina often present with atypical symptoms that may appear to be noncardiac in nature.

Anginal chest pain may also be seen in patients with aortic stenosis or hypertrophic cardiomyopathy secondary to the supply-demand imbalance caused by excessive myocardial hypertrophy. Pericarditis commonly results in a sharp type of chest pain that occurs in the substernal region and worsens on inspiration (pleuritic) when the patient is in a supine position and improves when the patient is in an upright position. The pain of aortic dissection is also substernal, but typically, it is described as a tearing or ripping sensation, radiates to the back or interscapular area, begins abruptly, and fails to improve with rest or use of nitroglycerin. Musculoskeletal pain may be located anywhere on the chest wall, is often reproducible with palpation, and frequently worsens with rotation of the thorax. If the pain is musculoskeletal in origin, recent episodes of excessive lifting or activity may be elicited in the history. Esophageal spasm and gastroesophageal reflux disease are frequent causes of noncardiac chest pain.[3]

Cardiovascular risk factors should be reviewed in all patients presenting with chest pain. These risk factors include (1) a history of hypertension, hyperlipidemia, diabetes mellitus, or cigarette smoking,[4] and (2) a family history of CAD (i.e., a first-degree male relative with myocardial infarction or sudden death occurring before 55 years of age or a first-degree female relative with these events occurring before 65 years of age). Relatively uncommon factors that may also result in angina include prior radiation therapy, drug use (e.g., cocaine and amphetamines), and the presence of a systemic disease (e.g., lupus erythematosus, polyarteritis nodosum, or rheumatoid arthritis) that is associated with coronary arteritis.

The physical examination is usually unremarkable in patients presenting with anginal chest pain. However, certain physical findings can be very helpful in supporting the diagnosis of CAD. Elevated blood pressure by cuff sphygmomanometry and retinal abnormalities on fundoscopic examination (e.g., arteriovenous nicking, microaneurysms, arteriolar narrowing, or hemorrhages) may indicate previously undiagnosed hypertension. Xanthomas (cholesterol-filled nodules that occur subcutaneously or over tendons) indicate severe elevations in serum choles-

terol levels. Femoral, carotid, or renal artery bruits and diminished peripheral pulses signify peripheral vascular disease and markedly increase the probability of CAD.[5] Tenderness to palpation of the chest wall, especially at the costochondral and chondrosternal articulations, suggests a musculoskeletal etiology of chest pain. Occasionally, patients with anginal chest pain also have a component of reproducible pain with palpation. A third heart sound and a holosystolic murmur of mitral regurgitation (secondary to ischemia of a papillary muscle) may be present if a patient with CAD is examined during an episode of anginal pain.

Physical examination also is directed toward findings that suggest an alternative cause of chest pain. Asymmetrical peripheral pulses, an early diastolic murmur, and the appropriate clinical history (tearing chest pain with radiation to the back) indicate an aortic dissection. A systolic murmur that radiates to the base of the neck (aortic stenosis) or a systolic murmur that increases in intensity with the strain phase of the Valsalva maneuver (hypertrophic cardiomyopathy) are uncommon but useful findings. A so-called leatherlike or scratchy series of sounds indicates a pericardial rub and supports a diagnosis of pericarditis. The intensity of the rub may increase with inspiration, indicating associated inflammation of the pleura, or pleuritis. Examination of the lung fields may disclose diminished breath sounds associated with dullness to percussion (pleural effusion), rhonchi, and egophony (pneumonia) or expiratory wheezes (asthma).

DIAGNOSTIC TESTS

On the basis of the history, chest pain is characterized as anginal, atypical anginal (some features of angina combined with some noncharacteristic features), or nonanginal. Estimates of the pretest probability of CAD can be accurately derived from a description of the chest pain syndrome and the presence or absence of cardiovascular risk factors.[6,7] The most widely used method for determining pretest likelihood of CAD is the Duke University Database formula, which considers the patient's age, sex, cardiovascular risk factor profile, description of chest pain, and information from the resting electrocardiogram.[8]

Although the diagnostic yield from the baseline ECG is low, it provides useful information on the advisability of pursuing additional diagnostic testing [*see Figure 1*]. Notable findings include Q waves consistent with a prior myocardial infarction and left ventricular hypertrophy that may be secondary to aortic stenosis, hypertrophic cardiomyopathy, or long-standing hypertension. ST segment depression, T wave abnormalities, and arrhythmias may be present if the ECG is obtained during an episode of anginal chest pain. A normal resting ECG predicts normal left ventricular function with a high degree of certainty (i.e., > 95%).

As with the ECG, a routine chest roentgenogram is usually normal. However, the presence of cardiomegaly, a left ventricular aneurysm, significant coronary or aortic calcification, or pulmonary venous congestion would be useful information and may warrant additional diagnostic testing.

Some physicians have started to use portable or handheld echocardiographic devices to evaluate patients with chest pain. Pertinent findings by echocardiography that would assist in establishing the etiology of chest pain include a pericardial effusion (pericarditis), hypokinesis or akinesis of a left ventricular wall segment (acute coronary ischemia), a dilated right ventricle (pulmonary embolism), and calcification and impaired excursion of the aortic valve leaflets (aortic stenosis).

Noninvasive stress testing is most likely to influence clinical decision making when the pretest probability of CAD is in the intermediate range. Patients with a low risk of CAD should not undergo noninvasive cardiac stress testing, because an abnormal test result would likely be a false positive one, and a negative test result would simply confirm the low probability of CAD. However, if patient reassurance is a consideration, a normal test result may be very useful. In addition, exercise stress testing provides information regarding symptom status, exercise capacity, and the hemodynamic response to exercise if the history is unclear (e.g., the patient denies symptoms but has decreased exercise capacity for "other reasons"). Absolute and relative contraindications to exercise testing should be reviewed in all patients before testing is begun [*see Table 2*].[8]

Similarly, patients with a high risk of CAD in general should not undergo noninvasive cardiac stress testing for the purpose of diagnosing CAD, because a negative test result would likely be a false negative one, and a positive result would simply confirm the high probability of CAD. In such patients, coronary angiography should be used to establish a diagnosis of CAD. However, noninvasive cardiac stress testing in certain patients at high risk for CAD may be useful. Indications for noninvasive stress testing in these patients include (1) assessment of the ef-

Table 1 Differentiating Features in the Patient's History of Chest Pain

Condition	*Location*	*Radiation*	*Quality*	*Alleviating Factors*	*Aggravating Factors*	*Duration*
Angina pectoris	Substernal	Jaw, arm	Pressure	Rest, nitroglycerin	Exercise, cold weather	5–15 minutes
Pericarditis	Left-sided, substernal	Neck, trapezius ridge	Sharp	Sitting up and leaning forward	Inspiration, supine position	Hours
Musculoskeletal	Variable over entire chest wall	None	Sharp or aching	Rest, anti-inflammatory or analgesic medications	Movement, palpation	Variable, but usually constant
Aortic stenosis	Substernal	Occasionally to jaw, arm	Pressure	Rest, nitroglycerin	Exercise, cold weather	Minutes
Hypertrophic cardiomyopathy	Substernal	Occasionally to jaw, arm	Pressure	Rest, nitroglycerin	Exercise, cold weather	Minutes
Aortic dissection	Substernal	Back	Tearing	None	None	Minutes to hours

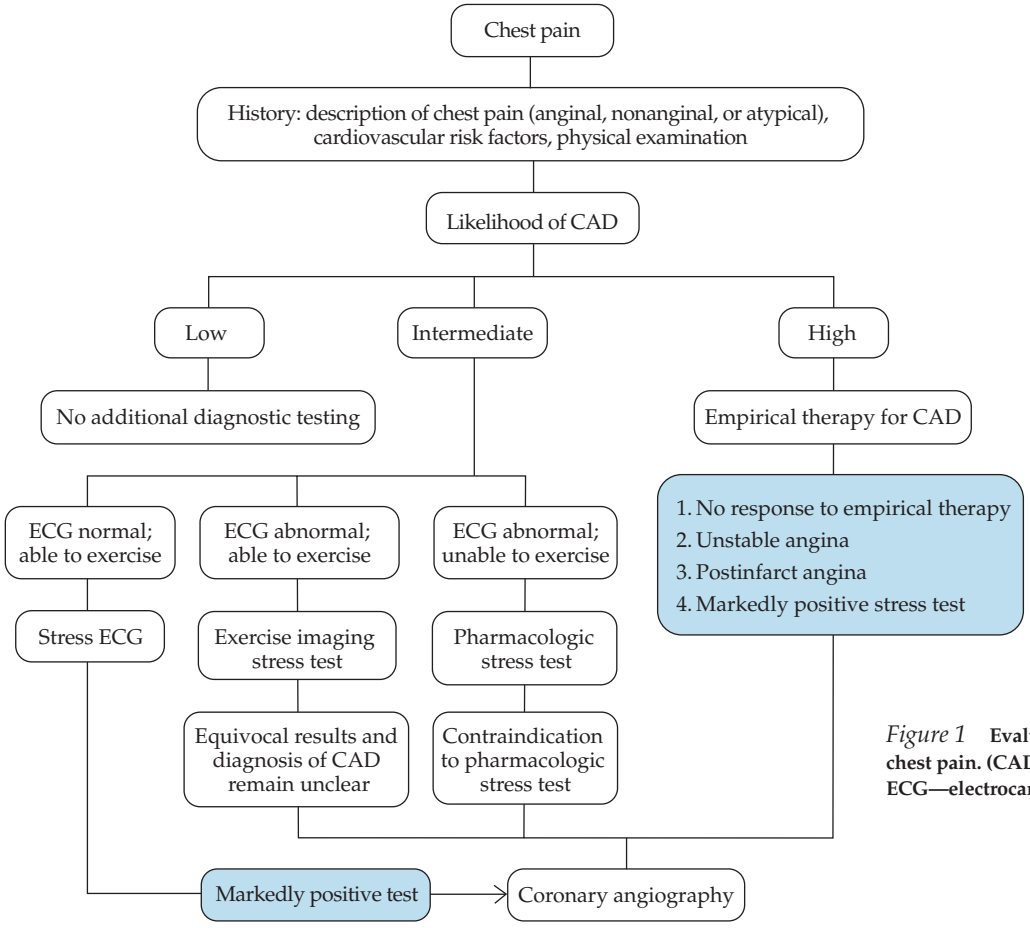


Figure 1 **Evaluation of patients with chest pain. (CAD—coronary artery disease; ECG—electrocardiogram)**

fectiveness of current medical therapy, (2) objective measurement of exercise capacity, (3) evaluation of the extent and location of ischemia or infarction with nuclear or echocardiographic imaging, (4) preoperative risk assessment in patients with known CAD who are undergoing noncardiac surgery, and (5) assessment of prognosis in patients with symptoms consistent with CAD or in patients with known CAD.

To establish the diagnosis of CAD in intermediate-risk patients, a number of noninvasive testing methods are available.[9] The decision whether to perform a specific test is based on various patient characteristics (e.g., body size, associated medical conditions, and ability to exercise), findings on the baseline ECG, and institutional experience with specific testing methods [*see Table 3*].[10-16] The most appropriate noninvasive stress test is chosen on the basis of each of these factors, as indicated in the chest pain algorithm [*see Figure 1*]. For most patients who are able to exercise with a normal baseline ECG, treadmill-ECG stress testing is indicated [*see Table 2*].[14,15,17] Women have a higher incidence of false positive results; therefore, many physicians recommend that, for all women, exercise be combined with an imaging method (e.g., echocardiography or nuclear imaging).[11] In general, to establish the diagnosis of CAD, exercise is preferred over pharmacologic stress agents. For patients who are unable to exercise because of physical limitations (e.g., arthritis or orthopedic problems), severe coexisting pulmonary disease, or general disability, pharmacologic stress agents such as dobutamine, adenosine, or dipyridamole can be employed. Each of these agents has specific contraindications [*see Table 4*].

Coronary angiography is considered the gold standard for the diagnosis of CAD. Although the incidence of major complications is low (< 2%), coronary angiography is costly and has some risk; thus, it is reserved for (1) patients with markedly positive noninvasive tests (i.e., hypotension and significant ST segment depression on ECG stress testing on a treadmill), (2) patients at high risk for CAD in whom a course of empirical antianginal therapy has failed, (3) patients with unstable or postinfarction angina, (4) patients with a contraindication to ex-

Table 2 Absolute and Relative Contraindications to Exercise Testing[8]

Absolute	*Relative*
Recent myocardial infarction (within 48 hr)	Left main coronary stenosis
Unstable angina not previously stabilized with medical therapy	Moderate stenotic valvular heart disease
Uncontrolled cardiac arrhythmias causing symptoms or hemodynamic compromise	Electrolyte abnormalities
Symptomatic severe aortic stenosis	Severe arterial hypertension
Uncontrolled symptomatic heart failure	Tachycardia or bradyarrhythmias
Acute pulmonary embolism or pulmonary infarction	Hypertrophic cardiomyopathy and other forms of outflow tract obstruction
Acute myocarditis or pericarditis	Mental or physical impairment leading to inability to exercise adequately
Acute aortic dissection	High degree of atrioventricular block

Table 3 Diagnostic Testing Methods Available for Evaluating Chest Pain[10-17]

Diagnostic Test	*Indications*	*Information Obtained*	*Limitations*	*Sensitivity*	*Specificity*
Exercise electrocardiographic stress test (stress ECG)	Initial test for most males with chest pain to establish diagnosis of CAD; females have higher rate of false positive test results Assess prognosis and functional capacity in patients with prior MI or known CAD Assess efficacy of current medical therapy in patients with known CAD	Exercise duration and functional aerobic capacity Amount of ST segment depression as indication of extent of ischemia Hemodynamic response to exercise	Normal baseline ECG Ability to exercise (patients who cannot attain adequate cardiopulmonary stress because of respiratory or musculoskeletal problems should receive a pharmacologic stress agent) Contraindications [*see Table 2*] False positives occur with left ventricular hypertrophy, bundle branch block, preexcitation syndromes, electrolyte abnormalities, and digoxin use	68%[1] (females, 61%[2])	77%[1] (females, 70%[2])
Thallium-201 perfusion scintigraphy	Often used when increased diagnostic accuracy for CAD required Can be combined with pharmacologic stress agents such as dobutamine, adenosine, or dipyridamole	Diagnosis of CAD with higher sensitivity and specificity than stress ECG Extent of ischemia Extent of infarction Left ventricular cavity size	Higher cost and longer testing time than stress ECG Imaging artifacts (attenuation) from diaphragm, breast, and intestine Contraindications [*see Table 2 if exercise; see Table 4 if pharmacologic stress agent*]	Ex thall 89%[5] Ph thall 90%[5] Dob thall 88%[4]	Ex thall 76%[5] Ph thall 70%[5] Dob thall 74%[4]
Technetium-99m perfusion scintigraphy	Often used when increased diagnostic accuracy for CAD required Can be combined with pharmacologic stress agents such as dobutamine, adenosine, or dipyridamole	Higher sensitivity and specificity for diagnosis of CAD than stress-ECG Extent of ischemia Extent of infarction Left ventricular cavity size ECG-gated SPECT allows calculation of left ventricular ejection fraction and evaluation of wall motion; evaluation of wall motion reduces false positive scans caused by imaging artifacts (attenuation) Used when excessive body weight precludes thallium imaging	Higher cost and longer testing time than stress ECG Imaging artifacts (attenuation) from diaphragm, breast, and intestine Contraindications [*see Table 2 if exercise; see Table 4 if pharmacologic stress agent*]	Ex tech 89%[5] Ph tech 90%[5] Dob tech 88%[4]	Ex tech 76%[5] Ph tech 70%[5] Dob tech 74%[4]
Exercise or dobutamine echocardiography	Exercise echocardiography often used when patient can exercise and has good-quality echocardiographic images Dobutamine used when exercise not possible	Higher sensitivity and specificity for diagnosis of CAD than stress ECG Left and right ventricular chamber size and function, presence of valve disease and pulmonary arterial pressures	Inadequate image quality may occur in patients with obesity, chronic obstructive pulmonary disease, and chest wall deformities Contraindications [*see Table 2 if exercise; see Table 4 if pharmacologic stress agent*]	Ex echo 85%[6] Dob echo 82%[6]	Ex echo 86%[6] Dob echo 82%[6]
Holter monitoring	Prinzmetal angina	Transient ST segment elevation in presence or absence of chest pain	Difficult to interpret because of baseline abnormalities		
Coronary angiography	Chest pain of unclear etiology despite noninvasive testing Angina not responsive to medical therapy Unstable and postinfarction angina Unclear diagnosis of CAD despite noninvasive stress testing	Anatomic severity of CAD Completely exclude cardiac origin of chest pain—gold standard of diagnostic tests Left ventricular function if left ventricular angiography also performed	Invasive procedure with low (< 2%) but inherent risk of MI, stroke, and death Represents a luminogram; does not evaluate functional significance of arterial narrowing	100%	100%

CAD—coronary artery disease Dob tech—dobutamine technetium Dob thall—dobutamine thallium ECG—electrocardiogram Ex echo—exercise echocardiography Ex tech—exercise technetium Ex thall—exercise thallium MI—myocardial infarction Ph stress—pharmacologic stress Ph tech—pharmacologic (adenosine or dipyridamole) stress combined with technetium Ph thall—pharmacologic (adenosine or dipyridamole) stress combined with thallium SPECT—single-photon emission computed tomography

Table 4 Mechanism of Action, Side Effects, and Contraindications of Pharmacologic Stress Agents

Pharmacologic Stress Agent	*Mechanism of Action*	*Side Effects*	*Contraindications*
Dobutamine	Increase myocardial oxygen demand by increasing heart rate, blood pressure, and myocardial contractility	78% of patients experience side effects: chest pain, palpitations, headache, flushing, malaise, and dyspnea; ventricular and atrial arrhythmias may occur	Severe hypertension at baseline, recent history of ventricular and/or atrial arrhythmias, and current beta-blocker use
Dipyridamole	Coronary artery vasodilatation—indirect response by blocking adenosine uptake and degradation	Increase in heart rate (average, 5–10 beats a minute), decrease in systolic blood pressure (average, 10–15 mm Hg); approximately 50% of patients experience side effects: chest pain, flushing, dizziness, headaches, or nausea; may provoke bronchospasm	Severe reactive airway disease (not contraindicated with chronic obstructive pulmonary disease unless a significant component of reactive airway disease is present), current theophylline use; avoid caffeine use 1 day before testing
Adenosine	Coronary artery vasodilatation—direct response	79% of patients experience side effects (more than with dipyridamole); side effects are chest, throat or jaw pain, headache, flushing, malaise, nausea, and bradyarrhythmias	Similar to dipyridamole; avoid caffeine use 1 day before testing; may cause bradyarrhythmias and is therefore contraindicated with baseline second- or third-degree heart block

ercise or pharmacologic stress testing, and (5) patients with equivocal results on noninvasive stress testing when the diagnosis of CAD remains unclear. Coronary angiography has certain limitations, including the inability to determine (1) the functional significance of a coronary artery stenosis and (2) which coronary plaque is likely to rupture (i.e., the so-called vulnerable plaque) and result in an acute coronary syndrome. Intravascular ultrasound studies have shown that coronary angiography may occasionally underestimate the severity of an area of narrowing, because it represents a so-called luminogram (shadow image) and not the size of the atherosclerotic plaque.[18] Despite these shortcomings, the extent and severity of CAD and measurement of left ventricular function by left heart catheterization are powerful predictors of clinical outcome.[14]

Dyspnea

BACKGROUND

Dyspnea refers to difficulty with breathing and can occur with a wide variety of cardiac, pulmonary, and systemic conditions [*see Table 5*]. Dyspnea can be classified as occurring (1) at rest, (2) with exertion, (3) during the night, awakening a patient from sleep (paroxysmal nocturnal dyspnea), or (4) during episodes of recumbency (orthopnea). Paroxysmal nocturnal dyspnea and orthopnea result from similar mechanisms. Specifically, the recumbent position augments venous return to the right heart. This increase in cardiac filling further increases the pulmonary capillary pressure and results in interstitial (and possibly intra-alveolar) pulmonary edema. Patients find relief by sitting upright, which reduces venous filling and transiently decreases the pulmonary interstitial pressure.

Dyspnea may be acute or chronic. An acute presentation suggests a pulmonary embolism, acute asthma exacerbation, pneumothorax, or rapidly developing pulmonary edema, as occurs with ischemic mitral regurgitation. Chronic dyspnea suggests heart failure resulting from systolic or diastolic dysfunction.

HISTORY AND PHYSICAL EXAMINATION

The history will often exclude less likely conditions and establish the etiology of dyspnea. A history of reactive airway disease, bronchodilator use, or corticosteroid use suggests asthma. Reactive airway disease tends to occur in children and young adults; therefore, in older patients given this diagnosis, a cardiac cause for dyspnea (e.g., new onset of congestive heart failure) should be considered. A significant history of tobacco use, wheezing, chronic cough, and sputum production suggests obstructive airway disease.[19] A recent history of fever, chills, and

Table 5 Causes of Dyspnea

- Cardiac
  - Valve disease
    - Aortic stenosis
    - Aortic regurgitaion
    - Mitral stenosis
    - Mitral regurgitation
  - Myocardial disease
    - Hypertensive heart disease
    - Dilated cardiomyopathy
    - Restrictive cardiomyopathy
    - Hypertrophic cardiomyopathy
  - Pericardial disease
    - Constrictive pericarditis
    - Pericardial tamponade
    - Pericardial effusion
  - Coronary disease
    - Myocardial infarction and ischemia
  - Arrhythmia
    - Ventricular and supraventricular arrhythmias
  - Congenital heart disease
- Pulmonary
  - Reactive airway disease
  - Chronic obstructive lung disease (chronic bronchitis and emphysema)
  - Interstitial lung disease
  - Infection (acute bronchitis and pneumonia)
  - Pulmonary embolism
  - Chest wall disease
  - Pleural effusion
- Deconditioning
- Obesity
- Malingering
- Psychogenic
  - Anxiety and panic disorders
- Anemia

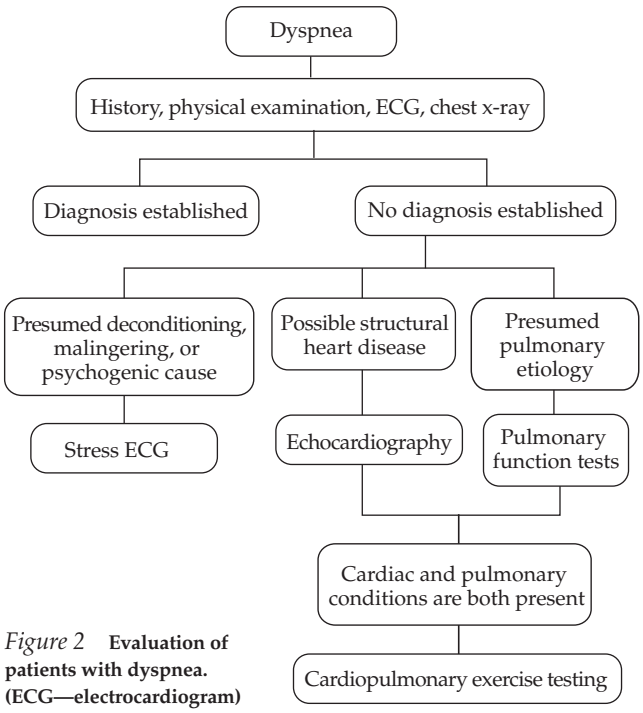


Figure 2 **Evaluation of patients with dyspnea. (ECG—electrocardiogram)**

productive cough may indicate bronchitis or pneumonia. The acute onset of dyspnea associated with pleuritic chest pain after a period of immobilization suggests pulmonary embolism. Paroxysmal nocturnal dyspnea, orthopnea, nocturia, recent weight gain, and lower extremity edema suggest a cardiac cause for dyspnea. Patients with chronic obstructive pulmonary disease may also awaken at night with dyspnea, but they usually have a history of sputum production and expectoration that improves with the patient in the upright position. Occasionally, on the basis of the history alone, it may not be possible to determine whether a cardiac or pulmonary cause of dyspnea is present.[20] In up to one third of patients being evaluated, dyspnea may have more than one cause.[21] In elderly patients, dyspnea may be the only symptom of a myocardial infarction. Hemoptysis may indicate the presence of severe underlying pulmonary disease (e.g., pulmonary embolism or lung cancer) but must be differentiated from hematemesis and nasopharyngeal bleeding.

Several findings on physical examination can assist in excluding a cardiac cause for dyspnea. These findings include a normal level of the jugular venous pressure, a normal point of maximal cardiac impulse, the lack of a third heart sound or cardiac murmurs, the absence of rales on lung examination, and the absence of peripheral edema. Alternatively, elevated jugular venous pressure, a displaced point of maximal cardiac impulse, a third heart sound, a holosystolic murmur of mitral regurgitation, basilar rales, and peripheral edema suggest congestive heart failure. A positive abdominojugular reflux maneuver may also identify dyspnea of cardiac origin.[22]

Obese patients and those with chest wall deformities may experience dyspnea secondary to the increased workload of breathing from the mechanical limitation imposed on the chest wall. Patients with emphysema frequently have an increased anteroposterior chest diameter, prolonged expiratory phase, expiratory wheezes, and diminished breath sounds. Central cyanosis, a normal anteroposterior chest diameter, and expiratory wheezes or rhonchi on lung examination suggest chronic bronchitis. Expiratory wheezing can occur in both cardiac and pulmonary conditions and is therefore not helpful in establishing an etiology. Stridor may result from an upper airway obstruction or vocal cord paralysis and at times may resemble wheezing. Tachypnea, a loud pulmonic component of the second heart sound, and calf tenderness suggest a pulmonary embolism.

DIAGNOSTIC TESTS

An ECG and a chest roentgenogram should be the initial tests in the evaluation of dyspnea. Pertinent ECG findings include Q waves (prior myocardial infarction), a bundle branch block (structural heart disease), left ventricular hypertrophy (aortic stenosis, hypertension), and evidence of atrial chamber enlargement (valvular heart disease). Notable chest roentgenogram findings include an enlarged cardiac silhouette; interstitial or alveolar edema (congestive heart failure); aortic valve calcification (valvular heart disease); lung mass (lung cancer); focal infiltrate (pneumonia); pleural effusion (congestive heart failure, infectious process); and hyperinflation, bullae, and flattened hemidiaphragms (emphysema). Screening laboratory tests may be useful to exclude anemia as a potential cause of dyspnea.

If the diagnosis of dyspnea remains unclear, additional testing can be pursued [*see Figure 2*]. For patients with cardiovascular risk factors, with findings on physical examination that suggest structural heart disease, or with abnormal ECGs, echocardiography is indicated to exclude valvular heart disease and assess systolic and diastolic ventricular function. Patients with a presumed pulmonary etiology for dyspnea that remains undiagnosed should undergo pulmonary function testing to exclude reactive airway and restrictive and chronic obstructive pulmonary disease. Stress-ECG may be useful to objectively evaluate the degree of limitation and may be particularly helpful for patients with presumed deconditioning, malingering, or a psychogenic cause for dyspnea. For patients who may have a component of dyspnea from both a cardiac and a pulmonary source, cardiopulmonary exercise testing can be considered. Serum brain natriuretic peptide (BNP) levels are useful in distinguishing cardiac from noncardiac causes of dyspnea; a BNP greater than 100 pg/ml has a sensitivity of 90% but a specificity of only 73% for establishing the diagnosis of heart failure.[23] Other factors that affect BNP levels include renal failure, acute coronary syndrome, and female gender.

Palpitations

BACKGROUND

Palpitations are a nonspecific symptom associated with severity ranging from an increased awareness of the normal heartbeat to life-threatening ventricular arrhythmias [*see Table 6*]. Although palpitations represent one of the most common complaints requiring evaluation in the outpatient setting,[24] consensus guidelines describing the appropriate evaluation have not yet been established.

For patients with an underlying cardiac disease associated with palpitations, long-term outcome is poor. In contrast, clinical outcome is excellent for those with a noncardiac origin for palpitations, despite a high rate of recurrent episodes.[25] The key, then, in the evaluation of palpitations is to establish or exclude the presence of underlying structural heart disease. This determination can often be made by use of information from the his-

tory, physical examination, and ECG, but it may require additional evaluation with ambulatory ECG monitoring and possibly electrophysiologic testing. Psychiatric illnesses (anxiety, panic, and somatization disorders) account for a certain number of patients who seek medical attention for palpitations[25]; these disorders can initially be screened by use of simple and rapid patient-administered questionnaires. Although an underlying psychiatric illness should be considered in appropriate patients, it does not obviate the need for a complete evaluation to exclude a cardiac origin.[26] A diagnostic algorithm is presented that utilizes a rational approach to diagnostic testing [*see Figure 3*].

Table 6 Causes of Palpitations

General Category	*Prognosis*
Hyperdynamic state	Anemia, thyrotoxicosis, and exercise—all leading to sinus tachycardia
Increase in cardiac stroke volume	Aortic regurgitation, patent ductus arteriosus
Arrhythmia	
Ventricular	Frequent ventricular premature beats, ventricular tachycardia
Supraventricular	Frequent atrial premature beats, atrial fibrillation, atrial flutter, multifocal atrial tachycardia, atrial tachycardia, atrioventricular nodal reentry tachycardia, atrioventricular reentry tachycardia
Psychiatric	Anxiety, panic, or somatization disorder

HISTORY AND PHYSICAL EXAMINATION

Palpitations are often described as a fluttering, a pounding, or an uncomfortable sensation in the chest. Occasionally, patients may complain only of a sensation of awareness of the heart rhythm. Patients may be able to discern whether the episodes are rapid and regular or rapid and irregular. Tapping a finger on the patient's chest in either a regular or an irregular manner may occasionally lead to an accurate description of the events.

A history of palpitations since childhood suggests a supraventricular arrhythmia and possibly an atrioventricular bypass tract, such as in the Wolff-Parkinson-White syndrome. Patients with congenital long QT syndrome typically begin to manifest symptoms in adolescence. A family history of sudden cardiac death, congestive heart failure, or syncope may suggest an inherited dilated or hypertrophic cardiomyopathy.

Knowing the circumstances in which palpitations occur may be useful in determining their origin. Palpitations associated only with strenuous physical activity are normal, whereas episodes occurring at rest or with minimal activity suggest underlying pathology. Episodes associated with a lack of food intake suggest hypoglycemia, and episodes after excessive alcohol intake suggest the toxic effects of alcohol. The resolution of symptoms with vagal maneuvers (breath-holding or the Valsalva maneuver) suggests paroxysmal supraventricular tachycardia. The onset of an episode of palpitations on assuming an upright position after bending over suggests atrioventricular nodal tachycardia.[27] Emotional stress and strenuous exercise may precipitate episodes in patients with long QT syndrome. Palpitations associated with anxiety or a sense of doom or panic suggest, but do not confirm, an underlying psychiatric disorder. An odds-ratio analysis found that regular palpitations, palpitations

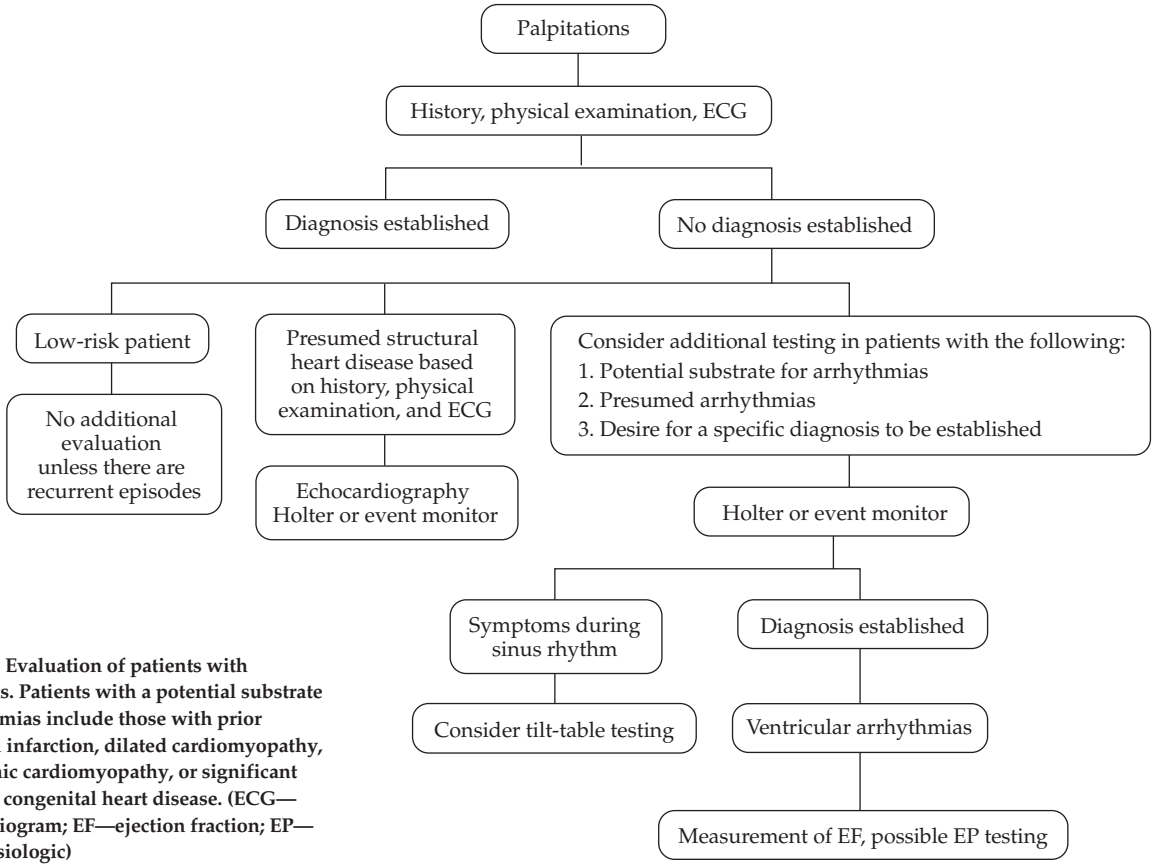


Figure 3 **Evaluation of patients with palpitations. Patients with a potential substrate for arrhythmias include those with prior myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, or significant valvular or congenital heart disease. (ECG—electrocardiogram; EF—ejection fraction; EP—electrophysiologic)**

Table 7 Medications Associated with Prolongation of the QT Interval

- Antibiotics
  - Tetracycline
  - Erythromycin
  - Trimethoprim and sulfamethoxazole
  - Pentamidine
- Antihistamines
  - Terfenadine
  - Astemizole
  - Diphenhydramine
- Antiarrhythmic agents
  - Quinidine
  - Procainamide
  - Disopyramide
  - Sotalol
  - Amiodarone
  - Dofetilide
- Other cardiac drugs
  - Bepridil
- Gastrointestinal
  - Cisapride
- Antifungal drugs
  - Ketoconazole
  - Fluconazole
  - Itraconazole
- Psychotropic drugs
  - Tricyclic antidepressants
  - Phenothiazines
  - Haloperidol
  - Resperidone
- Diuretics
  - Indapamide

experienced at work, and those affected by sleeping were more likely to indicate cardiac origin.[28]

Symptoms associated with an episode of palpitations should also be explored. Syncope or presyncope after an episode suggests ventricular arrhythmias. However, patients with structural heart disease (e.g., severe left ventricular systolic dysfunction) may also experience these symptoms after supraventricular arrhythmias because of dependence on atrial filling. Additional mechanisms of syncope in patients with supraventricular arrhythmias have also been reported.[29] Regardless of the mechanism, syncope and presyncope are worrisome symptoms and merit a complete cardiovascular evaluation. Occasionally, patients may experience an episode of polyuria that follows the palpitations. This condition may suggest supraventricular arrhythmias as the cause of palpitations, although studies have found this to be uncommon.[30]

The physical examination should focus on establishing whether underlying structural heart disease is present. Evidence of cardiac enlargement, third heart sound, and holosystolic murmur of mitral regurgitation suggest an underlying dilated cardiomyopathy. A midsystolic click, often followed by a systolic murmur, indicates mitral valve prolapse, which may be associated with both ventricular and supraventricular arrhythmias. A midsystolic murmur along the left sternal border that varies in intensity with alterations in left ventricular filling (e.g., Valsalva maneuver or changes in body position) is consistent with hypertrophic cardiomyopathy. Although atrial fibrillation is common in hypertrophic cardiomyopathy, ventricular arrhythmias may also occur.

DIAGNOSTIC TESTS

The ECG is the first step in the diagnostic evaluation of a patient with palpitations [*see Figure 3*]. A short PR interval and delta wave (Wolff-Parkinson-White syndrome), prolonged QT interval (long QT syndrome), and left bundle branch block (structural heart disease) are notable findings. Certain medications [*see Table 7*] may result in prolongation of the QT interval (i.e., acquired prolonged QT) and increase the risk of arrhythmias. Extreme voltage amplitudes and Q waves in leads I, aVL, and V4 through V6 are seen with hypertrophic cardiomyopathy. Pathologic Q waves indicate prior myocardial infarction and therefore a substrate for ventricular arrhythmias. Left ventricular hypertrophy or atrial abnormalities are nonspecific findings but suggest underlying structural heart disease. Many pertinent findings for various causes of palpitations can be obtained from the history, physical examination, and ECG [*see Table 8*].

If the cause of palpitations is not apparent after the initial evaluation (history, physical examination, and ECG), additional

Table 8 Diagnosis of the Underlying Etiology of Palpitations

Condition	*History*	*Physical Examination*	*ECG*	*Underlying Etiology of Palpitations*
Congenital long QT syndrome	Symptom onset in adolescence; episodes may be triggered by emotional stress and strenuous exercise	Normal	Prolonged QT interval	Ventricular arrhythmias
Atrioventricular bypass tract (e.g., Wolf-Parkinson-White syndrome)	Childhood episodes of palpitations	Normal	Short PR interval, delta wave	Supraventricular arrhythmias
Inherited dilated cardiomyopathy	Family history of cardiomyopathy, syncope, or sudden cardiac death	Abnormal cardiac impulse, systolic murmur (MR), third heart sound	Atrial enlargement, IVCD, LBBB, ventricular ectopic beats, or Q waves	Supraventricular or ventricular arrhythmias
Hypertrophic cardiomyopathy	Family history of cardiomyopathy, syncope, or sudden cardiac death	Systolic murmur	Increased voltage amplitude (LVH), Q waves in V4-V6, I, aVL	Supraventricular or ventricular arrhythmias
Anxiety, panic, or somatization disorder	Sense of doom, panic, or anxiety associated with episodes; coexisting psychiatric illness	Normal	Normal	Psychiatric
Mitral valve prolapse	Associated fatigue, dyspnea	Midsystolic click, systolic murmur (MR)	Normal or left atrial enlargement	Supraventricular arrhythmias

IVCD—interventricular conduction defect LBBB—left bundle branch block LVH—left ventricular hypertrophy MR—mitral regurgitation

diagnostic testing is indicated for certain patients [*see Figure 3*].[27] Such patients include those with presumed arrhythmias that remain undiagnosed and those with prior myocardial infarction, dilated cardiomyopathy, hypertrophic cardiomyopathy, or significant valvular or congenital heart disease. In addition, patients who desire a specific diagnosis should be considered for additional testing.

Ambulatory ECG devices include Holter monitoring and continuous-loop event recorders. Holter monitors continuously record the heart rhythm for 24 or 48 hours. Patients are asked to maintain a diary documenting the time and describing the symptoms during the monitoring period. The key is to correlate patient symptoms with documented rhythm abnormalities. Patients with significant complaints of palpitations that correlate with periods of normal sinus rhythm should be further evaluated for underlying psychiatric disorders. Event monitors also continuously record the heart rhythm but require the patient to trigger the device to save the information. These devices can be kept by patients for several weeks and are especially useful when symptoms are infrequent. Event monitors are more cost-effective than Holter monitors for evaluating palpitations.[31,32] For patients with underlying structural heart disease and documented ventricular arrhythmias on ambulatory ECG monitoring, additional evaluation is warranted, including determination of left ventricular function and, occasionally, electrophysiologic testing.

Syncope

BACKGROUND

Syncope refers to a transient loss of consciousness accompanied by loss of postural tone. Roughly one third of all persons have an episode of syncope during their lifetime. It is a particularly common problem encountered in emergency departments and accounts for approximately 6% of all hospital admissions.[33] Determining which patients require hospital admission is difficult, given the large number of potential causes of syncope. Although many conditions that result in syncope are life threatening, other common etiologies, such as medication side effects, orthostatic hypotension, and psychiatric disorders, are benign.

Syncope is classified on the basis of the underlying etiology [*see Table 9*]. In elderly patients, the etiology may be multifactorial and related to medication side effects (particularly antihypertensives and antidepressants),[34] orthostatic hypotension, and bradyarrhythmias. Various medications are associated with prolongation of the QT interval and the development of ventricular arrhythmias and resulting syncope [*see Table 7*]. Vasovagal syncope is particularly common in otherwise healthy patients and has a benign prognosis. Episodes often occur in response to injury and are characterized by a sudden decline in blood pressure with or without associated bradycardia.

Establishing the presence of structural heart disease in the evaluation of patients with syncope is essential because such patients may have a 1-year mortality as high as 30%.[35,36] Structural heart disease is usually apparent on the basis of history, physical examination, and information from the baseline ECG. Occasionally, additional diagnostic testing with echocardiography, tilt-table testing, or electrophysiologic testing may be required.

HISTORY AND PHYSICAL EXAMINATION

The first step in establishing the presence of structural heart disease is to obtain an accurate description of the episode of syncope. Key elements of the history include the presence of postural or exertional symptoms; associated chest pain, shortness of breath, or palpitations; and the situation in which the episode occurred (e.g., during micturition). Neurologic symptoms such as focal motor weakness, arm or leg movement, tongue biting, or a postictal state suggest a neurologic rather than cardiac event. However, seizures can occur from cardiac causes if a patient is kept upright during an episode (usually the result of a well-meaning bystander) because of cerebral hypoperfusion. A witness to the episode of syncope may provide a clear description of the event and should be questioned if possible. Medications associated with QT prolongation [*see Table 7*], blood pressure lowering (antihypertensives), and volume depletion (diuretics) should be reviewed. A family history of sudden cardiac death, syncope, or heart failure suggests hypertrophic cardiomyopathy, an inherited dilated cardiomyopathy, or long QT syndrome. A history of myocardial infarction or congestive heart failure raises the possibility of ventricular arrhythmias.

The physical examination focuses on determining whether structural heart disease is present and excluding common causes of syncope. Orthostatic vital signs should be obtained in all patients. Focal neurologic findings such as a motor deficit or a visual-field defect may indicate a neurologic cause for syncope. Pertinent findings on cardiovascular examination include a delayed carotid upstroke (aortic stenosis), an abnormal point of maximal cardiac impulse (cardiomyopathy), an irregular or

Table 9 Classification of Syncope Based on Etiology

- Cardiac
  - Blood flow obstruction
    - Aortic stenosis
    - Pulmonic stenosis
    - Left atrial myxoma
    - Hypertrophic cardiomyopathy
    - Massive pulmonary embolism
  - Reduction in forward cardiac output
    - Pericardial tamponade
    - Severe pump failure
  - Arrhythmia
    - Tachyarrhythmias
      - Ventricular tachycardia
      - Supraventricular tachycardia
    - Bradyarrhythmias
      - Sinus bradycardia
      - Sick sinus syndrome
      - Atrioventricular block
      - Carotid sinus hypersensitivity (can also be considered neurologic cause)
- Neurologic
  - Vasovagal
  - Situational (micturition)
  - Seizures
  - Cerebrovascular accident
  - Cerebrovascular insufficiency
  - Orthostatic hypotension—autonomic dysfunction
- Other
  - Volume depletion
  - Drugs
  - Hypoglycemia
  - Anxiety attack
  - Psychogenic

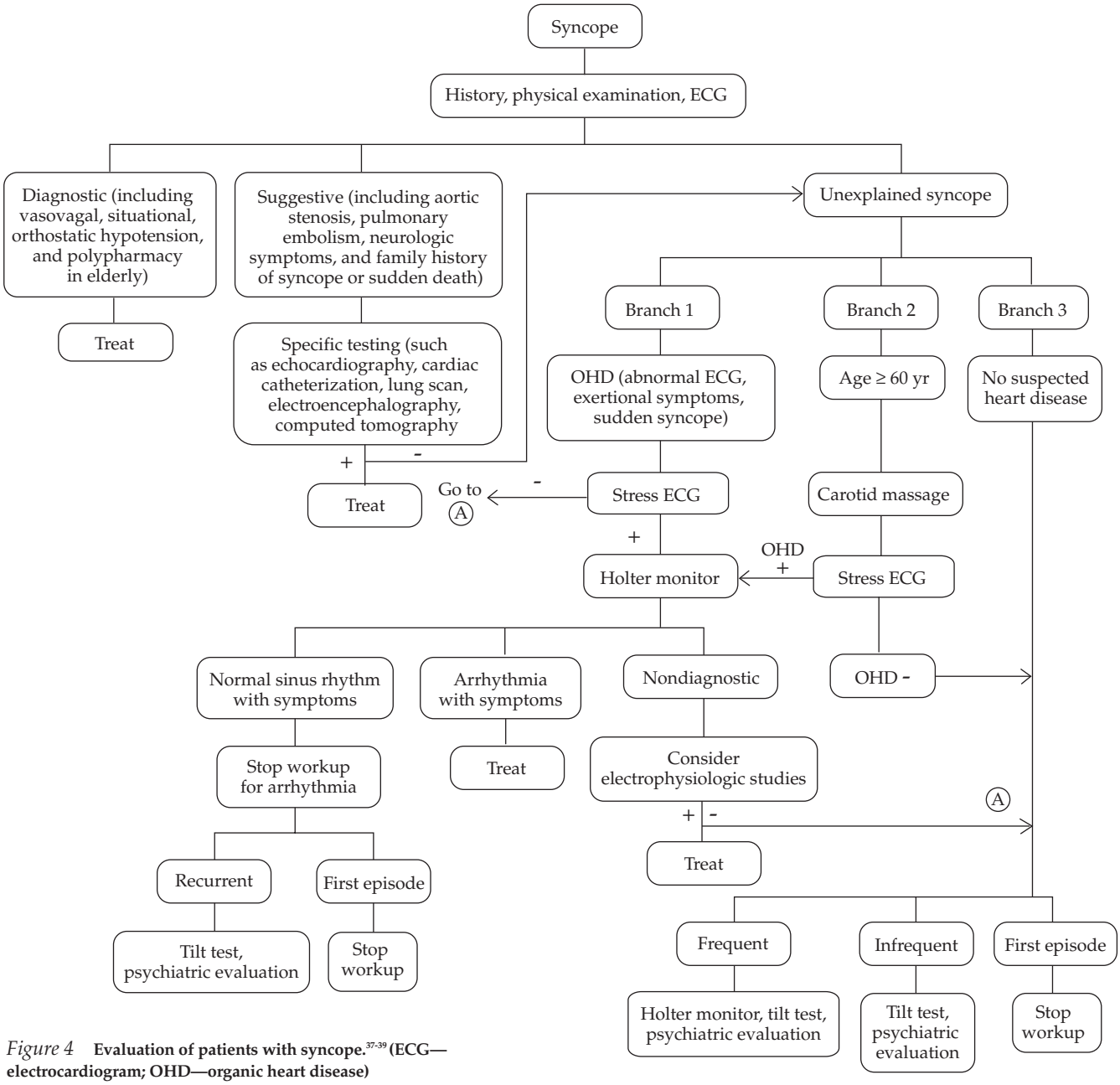


Figure 4 **Evaluation of patients with syncope.[37-39] (ECG—electrocardiogram; OHD—organic heart disease)**

bradycardiac rhythm (arrhythmias), a third heart sound (cardiomyopathy), a midsystolic murmur (aortic stenosis, hypertrophic cardiomyopathy), and a holosystolic murmur (mitral regurgitation secondary to left ventricular dilatation). Less common findings include an early diastolic sound (so-called tumor plop, indicating a left atrial myxoma), asymmetrical peripheral pulses (aortic dissection), and a loud second heart sound (pulmonary hypertension secondary to pulmonary embolism). Information from the history and physical examination yields a cause for syncope in approximately 45% of patients.[37]

DIAGNOSTIC TESTS

An ECG is the initial diagnostic test for all patients with syncope. Although the yield of the baseline ECG is low (approximately 5%), a number of potential findings are useful,[37] including bundle branch block, Q waves indicating prior myocardial infarction, left ventricular hypertrophy, prolonged QT interval, or evidence of atrioventricular block. The presence of sinus bradycardia, first-degree atrioventricular block, and bundle branch block suggests bradyarrhythmias as the cause of syncope. Extreme voltage amplitudes and Q waves in leads I, aVL, and V4 through V6 suggest hypertrophic cardiomyopathy and therefore the possibility of ventricular arrhythmias. An uncommon but unique ECG abnormality is the combination of a right bundle branch block, T wave inversions in leads V1 through V3, and an epsilon wave (a positive wave on the terminal portion of the QRS complex)—findings that indicate right ventricular dysplasia, which is associated with ventricular arrhythmias. Ventricular arrhythmias are also seen in the Brugada syndrome, which can be identified on the ECG by an incomplete right bundle branch block and ST segment elevation in leads V1 through V3.[38] A short PR interval and slurring of the initial portion of the

Table 10 Differential Diagnosis of Claudication

Condition	*History*	*Physical Examination*	*Diagnostic Tests*	*Comments*
Peripheral vascular disease	Symptoms occur with exercise and are relieved by rest	Diminished or absent peripheral pulses	ABI, arterial duplex ultrasound	Angiography reserved for those with severe disease who are considering surgical or percutaneous revascularization
Lumbar spinal stenosis	Paresthesias occur with standing and walking Symptoms are relieved by sitting and/or leaning forward History may include chronic low back pain and prior lumbar surgery	Normal peripheral pulses	Computed tomography or magnetic resonance imaging of the lumbar spine	Referred to as pseudoclaudication
Arthritis	Pain localized to the joint area as opposed to adjacent muscles	Normal peripheral pulses	Radiograph of affected joint	
Myalgia	Pain within a muscle group at rest and with exertion No relief with rest	Tenderness to palpation of the affected muscle group; reduced muscle strength	Laboratory evaluation of muscle inflammation with CPK, aldolase	Associated with hypothyroidism and end-stage renal disease; may be related to drug side effect (e.g., HMG-CoA reductase inhibitors)

ABI—ankle-brachial index CPK—creatinine phosphokinase

QRS complex (the delta wave) suggests preexcitation (i.e., Wolff-Parkinson-White syndrome), with the possibility of rapid antegrade conduction via the accessory pathway.

If the etiology of syncope remains unclear after reviewing the history, physical examination, and ECG, additional diagnostic testing should be pursued. For patients with findings suggestive of an underlying cardiac cause, echocardiography and coronary angiography can be performed; for those with a possible neurologic cause, brain imaging (computed tomography or magnetic resonance imaging), neurovascular studies (carotid and transcranial Doppler ultrasound studies), and electroencephalography can be performed; and for those with a presumed pulmonary cause, lung scanning can be considered [*see Figure 4*]. If the diagnosis remains uncertain despite these tests, one of three pathways can be followed.[39]

The first pathway is for patients with structural heart disease or an abnormal ECG, who therefore have an increased likelihood for underlying arrhythmias or valve disease as a cause for syncope. Echocardiography, noninvasive stress testing, and ambulatory ECG monitoring using either a Holter monitor or continuous-loop event recorder should be considered for these patients. Event recorders have been found to be more accurate than Holter monitors in the diagnosis of syncope and presyncope; however, some patients find event recorders difficult to operate correctly.[40] If ambulatory ECG monitoring documents normal sinus rhythm in the setting of reported syncope, psychiatric evaluation and possibly tilt-table testing are warranted.

The second pathway is for patients older than 60 years, who are more likely to have valve disease (aortic stenosis), ischemic heart disease, carotid sinus syncope, cerebrovascular disease (transient ischemic attacks), and situational events (micturition, defecation, postural) as a basis for syncope. Carotid sinus massage (in the absence of carotid bruits, recent myocardial infarction, or stroke) should be the initial diagnostic test for these patients.[41,42] A positive test is defined as asystolic arrest lasting 3 seconds or longer and may identify those with cardioinhibitory hypersensitivity of the carotid sinus who will benefit from pacemaker placement. For those with a negative test result, echocardiography, noninvasive stress testing, and ambulatory ECG monitoring can be performed.

The third pathway is for patients with unexplained syncope and no suspected structural heart disease. For those who have had a single episode, additional evaluation can be deferred until a second episode occurs. In patients with more than one episode, ambulatory ECG monitoring or tilt-table testing and, possibly, psychiatric evaluation should be considered.

Tilt-table testing was initially developed in the 1980s to evaluate patients with presumed vasovagal syncope. The passive portion of the test involves quickly raising a patient from the supine position to an angle of 60° (the tilt angle) for approximately 45 minutes, which causes pooling of venous blood in the lower extremities, a decrease in venous return, compensatory tachycardia, and enhanced ventricular contraction. For individuals with vasovagal syncope, augmented ventricular contraction causes activation of vasodepressor reflexes that result in hypotension, bradycardia, or both. Approximately 49% of patients referred for evaluation of

Table 11 Ankle-Brachial Index (ABI) Values and Accompanying Findings in Peripheral Vascular Disease (PVD)

Condition	*Symptoms*	*Physical Findings*	*ABI*
Normal	None	None	> 1.0
Mild PVD	Mild claudication on exertion	Diminished pulses	0.8–0.9
Moderate PVD	Moderate or severe claudication on exertion	Diminished or absent pulses; nonhealing ulcers or skin wounds	0.5–0.8
Severe PVD	Severe claudication; symptoms may occur at rest	Absent pulses; nonhealing ulcers or skin wounds	< 0.5

vasovagal syncope have positive responses, compared with 9% of control patients.[43] The active portion of tilt-table testing uses an isoproterenol infusion to enhance the vasodepressor reflex.

Claudication

BACKGROUND

Claudication is a condition of muscle pain or weakness associated with compromised blood flow to the extremities. It is a common complaint of patients who have peripheral vascular disease (PVD). Claudication is also a common symptom of CAD, a disease that shares risk factors with PVD. PVD is associated with an increased risk of stroke, cardiovascular death, and all-cause mortality[44]; it is most frequently caused by atherosclerosis. Studies suggest that PVD is infrequently diagnosed and often undertreated in the primary care setting.[45]

Intermittent claudication associated with PVD is a reproducible discomfort of a muscle group that is induced by exercise and relieved by rest. It is frequently manifested during ambulation as pain in the buttocks, upper thighs, and calves. The differential diagnosis of claudication [*see Table 10*] can be challenging and requires a detailed history and physical examination supplemented with diagnostic testing.

HISTORY AND PHYSICAL EXAMINATION

Claudication is described by patients as pain or cramping in the buttocks, thighs, and calf muscles. Symptoms typically occur during ambulation and are relieved by rest. Patients may also describe generalized weakness or a tired sensation within the legs. Depending on the extent of disease, men may experience impotence. Patients with severe disease may experience pain at rest. Patients with pain originating from compression of a nerve root describe an electric shock–like discomfort that frequently involves both legs and may be relieved by sitting down and leaning forward. More than one cause for leg symptoms may exist in a given patient, necessitating careful and clear delineation of each individual patient. Lumbar stenosis is another common cause of these symptoms (i.e., pseudoclaudication). The pain of pseudoclaudication is often poorly localized and may affect the leg from thigh to calf. Additional information that should be elicited from the history include the presence of a nonhealing ulcer or wound and previous manifestations of PVD, such as prior carotid endarterectomy. Risk factors associated with PVD are similar to those associated with CAD and include older age, cigarette smoking, diabetes mellitus, hypertension, and hyperlipidemia.

Physical examination should focus on assessment of all peripheral pulses, including the carotid, femoral, popliteal, dorsal pedis, and posterior tibial. Peripheral pulses should be described as normal, diminished, or absent, and the presence or absence of a bruit should be noted. For patients with severe PVD, the skin distal to the area of occlusion may be cold, and elevation of the legs may result in pallor of the soles of the feet. The presence and location of ulcers, skin wounds, and gangrene should be noted. Auscultation over the neck and abdomen may reveal bruits suggesting carotid artery stenosis and renal artery stenosis, repectively.

DIAGNOSTIC TESTS

The key diagnostic test in the evaluation of claudication is the ankle-to-brachial systolic pressure index (ABI). This involves measuring the systolic blood pressure in the ankle and the upper arm (brachial artery) in the supine position. The ankle systolic pressure is measured with a standard blood pressure cuff placed around the ankle, with the lower edge of the cuff situated above the malleoli. The blood pressure cuff is inflated to approximately 30 mm Hg above the systolic pressure to temporarily occlude flow. As the cuff is slowly deflated, a Doppler probe is used to monitor the signal; the pressure at which the Doppler flow signal is heard is recorded as the systolic pressure. A normal ABI is greater than 1.0; a value of less than 1.0 indicates the presence of PVD [*see Table 11*]. In addition to establishing the diagnosis of PVD, the ABI gives an assessment of disease severity and has prognostic implications for future cardiovascular and cerebrovascular events.[46,47] Additional noninvasive imaging to evaluate the extent and severity of PVD may include an arterial duplex ultrasound. The vasculature from the abdominal aorta to the distal tibial arteries can be imaged to localize the area of stenosis and assess hemodynamic significance using the Doppler flow signals or Doppler velocity spectra.

Diagnostic angiography involves the use of cineangiographic imaging and radiographic contrast to image the peripheral vessels. Because angiography is an invasive procedure that carries a 1% risk of vascular complications, it is usually reserved for patients being considered for surgical or percutaneous (angioplasty and stenting) revascularization procedures; such procedures are undertaken to relieve limiting claudication, nonhealing ulcers, and severe ischemia [*see Figure 5*]. Magnetic resonance angiography is potentially useful in the initial evaluation of patients with renal insufficiency.

A normal ABI excludes PVD as the cause of claudication and prompts investigation of alternative conditions as the underlying disorder [*see Figure 6*].

For patients with an abnormal ABI and mild symptoms, medical therapy can be initiated and risk factors modified.[48] Patients with moderate to severe symptoms should undergo additional noninvasive testing using arterial Duplex ultrasound. Patients with limiting symptoms or threatened limb loss should undergo angiography in anticipation of surgical or percutaneous revascularization.

Cardiac Murmurs

BACKGROUND

The increased access to health care and the widespread use and availability of echocardiography have resulted in a large number of patients being diagnosed and evaluated for various cardiac murmurs. A cardiac murmur may indicate underlying valvular, congenital, or myocardial disease, but it may also be caused by systemic illnesses and occur in the setting of a structurally normal heart.

Cardiac murmurs result from disturbed or turbulent blood flow, often through diseased cardiac valves or intracardiac structures. The presence of a cardiac murmur, however, does not always indicate underlying cardiac pathology. Hyperthyroidism, anemia, and a febrile illness may all result in increased blood flow through the aortic valve and produce a soft, crescendo-decrescendo, systolic murmur over the aortic area. In this setting, the aortic valve is structurally normal, and the murmur is the result of augmented blood flow (i.e., a flow murmur) caused by the systemic illness. Another common cause of a systolic murmur is calcification of the aortic valve, which is referred

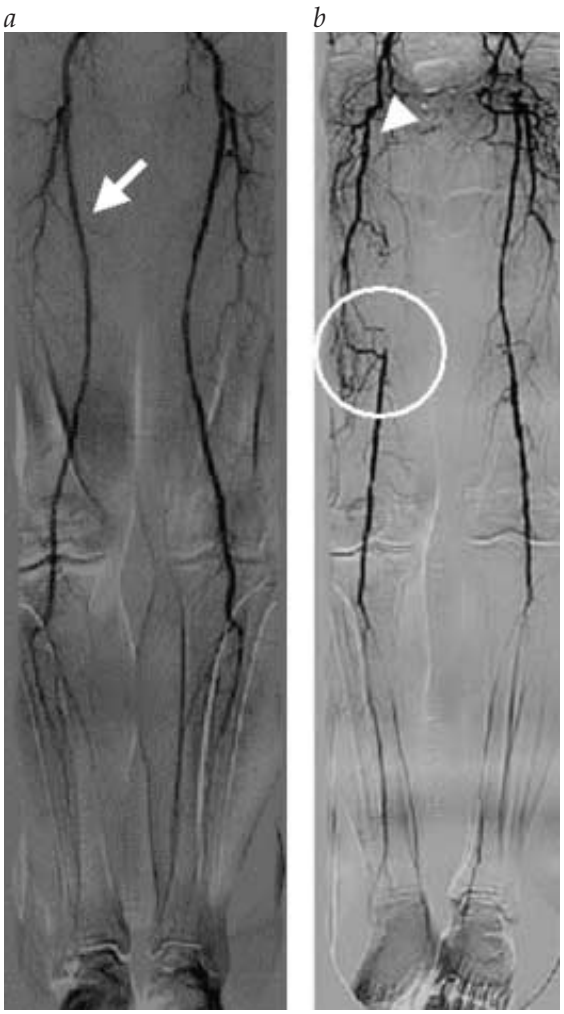


Figure 5 **Abdominal aortogram showing peripheral runoff. (*a*) Mild irregularities of the peripheral vessels are present, but there is no evidence of severe disease. The arrow indicates mild irregularities of the right superficial femoral artery. (*b*) Severe peripheral vascular disease with occlusion of proximal right superficial femoral artery (arrowhead) is evident. The midportion of the right superficial femoral artery is reconstituted from collateral vessels supplied by the right profunda femoralis artery (circle).**

to as aortic sclerosis when there is no obstruction to left ventricular outflow. Aortic sclerosis is a common finding in elderly patients; 25% of those older than 65 years are affected.[49] This condition is often diagnosed when a systolic murmur is detected in an otherwise asymptomatic patient during a routine physical examination. In addition to being caused by diseases of the cardiac valves, murmurs may result from intracardiac communications (atrial and ventricular septal defects), congenital abnormalities (patent ductus arteriosus), and disease of the myocardium (hypertrophic cardiomyopathy).

A thorough history and physical examination can often provide the etiology of a murmur. Additional diagnostic tests, such as the ECG, chest roentgenogram, and echocardiogram, are used to confirm the diagnosis and establish the severity of the abnormality.

HISTORY AND PHYSICAL EXAMINATION

A history of a childhood murmur may indicate a congenital abnormality of a cardiac valve, such as a bicuspid or unicuspid

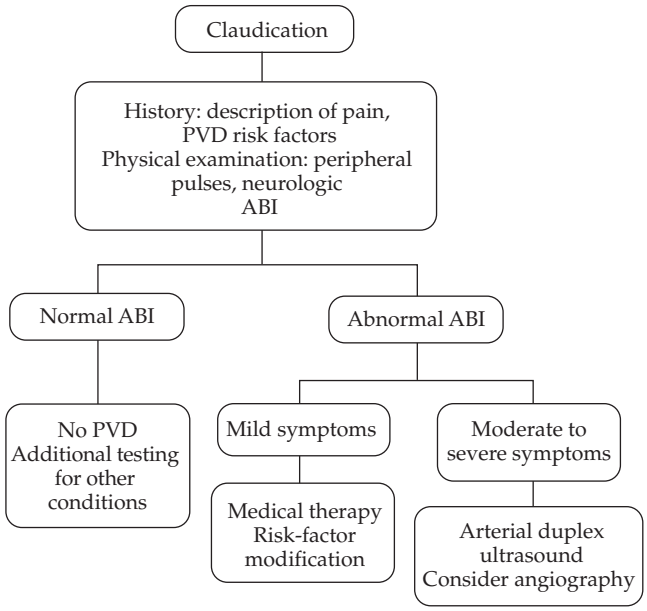


Figure 6 **Evaluation of patients with claudication.**

aortic valve. A febrile illness occurring in childhood should raise the suspicion of rheumatic fever, possibly resulting in rheumatic mitral stenosis. Although rheumatic fever is uncommon in the United States, it may still be seen in immigrants from Asia, Latin America, and the Caribbean.

Establishing the presence or absence of cardiovascular symptoms is essential in the evaluation of a cardiac murmur. Otherwise healthy young adults without cardiac symptoms, with a systolic flow murmur and no other cardiac findings on examina-

Table 12 Differential Diagnosis of a Cardiac Murmur Based on Timing of Cardiac Cycle

- Systolic
  - Midsystolic
    - Innocent flow murmur
    - Aortic stenosis
    - Pulmonic stenosis
    - Atrial septal defect
  - Holosystolic
    - Ventricular septal defect
    - Tricuspid regurgitation
    - Hypertrophic cardiomyopathy
    - Mitral regurgitation
- Diastolic
  - Early diastolic
    - Aortic regurgitation
    - Pulmonic regurgitation
  - Middiastolic
    - Mitral stenosis
    - Tricuspid stenosis
    - Austin Flint murmur associated with chronic aortic regurgitation
    - Severe mitral regurgitation (augmented antegrade mitral valve flow)
- Continuous
  - Patent ductus arteriosus

Table 13 Physical Findings Useful for Evaluating a Cardiac Murmur

Condition	*Timing*	*Location*	*Radiation*	*Characteristics*	*Effects of Maneuvers*	*Associated Findings*
Innocent flow murmur	Midsystolic	Base	Variable or none	Soft, ejection	No change	None
Aortic stenosis	Systolic	Base (right second ICS)	Carotid arteries	Crescendo-decrescendo	Decrease with handgrip or standing	Single S_2, delayed and decreased carotid upstroke, ES if mobile valve leaflets
Mitral regurgitation	Systolic	Apex	Axilla (sometimes back)	Holosystolic	Increase with handgrip	Hyperdynamic apical impulse
Ventricular septal defect	Systolic	Left sternal border	None	Holosystolic	No change	Palpable thrill
Atrial septal defect	Systolic	Left second ICS	None	Crescendo-decrescendo	Possible increase with inspiration	Fixed split S_2
Hypertrophic cardiomyopathy	Systolic	Base	Carotid arteries	Late-peaking crescendo	Increase with standing and strain phase of Valsalva maneuver	Brisk carotid upstroke
Tricuspid regurgitation	Systolic	Left lower sternal border	Right lower sternal border	Holosystolic	Increase with inspiration	Prominent v waves in JVP, pulsatile liver
Pulmonic stenosis	Systolic	Left second ICS	None	Crescendo-decrescendo	No change	ES if mobile valve leaflets
Aortic regurgitation	Diastolic	Left sternal border	None	Decrescendo, high-pitched	Increase with handgrip	Wide pulse pressure, displaced and enlarged apical impulse
Mitral stenosis	Diastolic	Apex	None	Low-pitched rumble, presystolic accentuation	Best heard in left lateral decubitus position	Loud S_1, opening snap
Pulmonic regurgitation	Diastolic	Left second ICS	Left sternal border	Decrescendo	May increase with inspiration	—
Tricuspid stenosis	Diastolic	Right lower sternal border	Right upper abdomen	Low-pitched rumble	Increase with inspiration	Right ventricular heave
Patent ductus arteriosus	Continuous	Left second ICS	Back	Machinery-like	None	Wide pulse pressure, bounding pulses

ES—ejection sound ICS—intercostal space JVP—jugular venous pulse S_1, S_2—first, second heart sounds

tion, often require no additional evaluation.[50] In contrast, the finding of a cardiac murmur in patients with cardiovascular symptoms must be further explored and a diagnosis established.

Aortic stenosis may result in the triad of angina, syncope, and impaired exercise tolerance or dyspnea on exertion. Patients with hypertrophic cardiomyopathy experience similar symptoms but may also complain of palpitations from associated atrial or ventricular arrhythmias. Hypertrophic cardiomyopathy is most commonly familial, with an autosomal dominant inheritance pattern; therefore, the patient should be questioned about a family history of sudden cardiac death, heart failure, and syncope. Symptoms of mitral stenosis include shortness of breath, impaired exercise tolerance, palpitations (from associated atrial fibrillation), and hemoptysis. These symptoms may occur during episodes of tachycardia, volume overload, or both as mitral valve flow is increased and the stenotic mitral valve impairs filling of the left ventricle. Asymptomatic women with mitral stenosis may develop symptoms during pregnancy. Mitral and aortic regurgitation cause a volume overload to the left atrium and left ventricle, respectively, and may result in shortness of breath, orthopnea, paroxysmal nocturnal dyspnea, lower extremity edema, and impaired exercise capacity. A ventricular septal defect is either congenital or ischemic (e.g., occurring after a myocardial infarction). The congenital form often becomes apparent during adolescence; the ischemic form presents several days after a myocardial infarction as a new holosystolic murmur associated with significant respiratory distress.

The physical examination begins with determining the timing of the murmur in the cardiac cycle—systolic, diastolic, or continuous [*see Table 12*]. The grade, quality, location, area of radiation, and change in intensity with maneuvers should then be described. Murmurs are graded on a scale of 1 to 6. Grade 1 is a soft intermittent murmur, grade 4 is a palpable murmur, and grade 6 is a murmur that can be appreciated without a stethoscope. Thus, most murmurs are classified as grade 2 or 3. Midsystolic murmurs are derived from the aortic or pulmonic valves or occur in association with hypertrophic cardiomyopathy. In contrast, holosystolic murmurs are the result of regurgitant blood flow through either the mitral or the tricuspid valves or a ventricular septal defect. The murmur of a ventricular septal

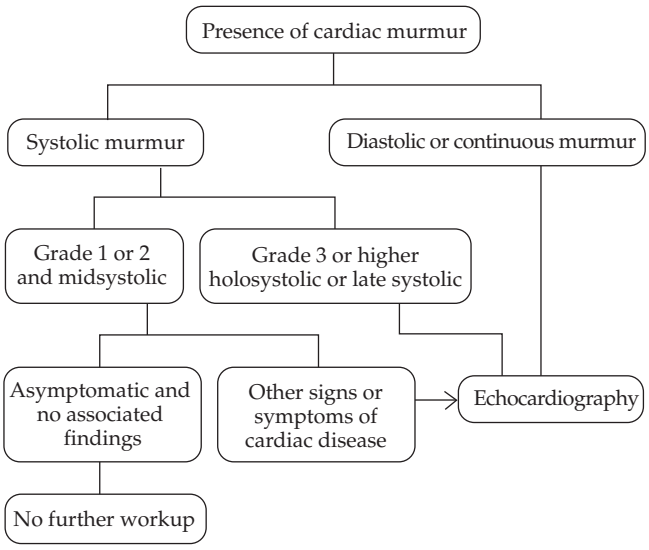


Figure 7 **Evaluation of patients with cardiac murmurs.**[45]

defect is usually well localized to the fourth left intercostal space, does not radiate significantly, and is often associated with a thrill (i.e., grade 4 or higher). Late systolic murmurs occur from mitral regurgitation that is secondary to (1) ischemia or infarction to the papillary muscles (ischemic mitral regurgitation), (2) left ventricular dilatation with functional mitral regurgitation, or (3) mitral valve prolapse. Additional findings on cardiac auscultation, such as a fixed, split second heart sound or an ejection sound, may be helpful in determining the etiology of a systolic murmur [*see Table 13*]. Electronic stethoscopes and handheld ultrasound devices have begun to supplement the bedside evaluation of cardiac murmurs.

Diastolic murmurs always indicate underlying cardiac pathology and commonly occur in either early diastole or middiastole. Early diastolic murmurs begin at the onset of diastole (i.e., with the second heart sound) and originate from regurgitant flow across the pulmonic and aortic valves. Aortic regurgitation occurs because of failure of the aortic valve leaflets to adequately coadapt during diastole and may be the result of disease processes affecting the aortic valve (e.g., endocarditis) or the aortic root (e.g., aortic dissection). Pulmonary regurgitation is most commonly seen in patients with pulmonary hypertension and is therefore associated with a loud second heart sound. Middiastolic murmurs occur from either mitral or tricuspid stenosis; the Austin Flint murmur associated with chronic aortic regurgitation or occurring in the setting of severe mitral regurgitation arises from augmented antegrade flow across the mitral valve in diastole.

In adults, continuous murmurs are usually from a previously undiagnosed patent ductus arteriosus. Occasionally, a patient with chronic aortic regurgitation may have a prominent systolic murmur in addition to the early diastolic murmur, thus simulating a continuous murmur. The systolic murmur in this case is the result of enhanced stroke volume from increased diastolic filling of the left ventricle. Whereas both conditions are associated with a widened pulse pressure and murmurs that occur during both systole and diastole, the murmur of a patent ductus arteriosus is continuous and peaks on the second heart sound; with chronic aortic regurgitation, there is a so-called silent period at the end of systole as the systolic murmur fades, before the beginning of the diastolic murmur.

Additional findings on physical examination can assist in determining the severity of the valve lesion and in excluding other conditions that result in similar murmurs. For patients with a midsystolic murmur presumed to be aortic stenosis, the carotid upstroke and splitting of the second heart sound should be carefully evaluated. A delayed carotid upstroke and single splitting of the second heart sound indicate hemodynamically severe aortic stenosis. However, physical examination has a low sensitivity for diagnosis of severe aortic stenosis, and overreliance on examination findings can lead to serious errors. The threshold for diagnostic imaging should be low in a patient with possible aortic valve stenosis. In contrast, hypertrophic cardiomyopathy results in a brisk carotid upstroke (the so-called spike-and-dome configuration) and normal splitting of the second heart sound. Severe mitral regurgitation can be identified by a holosystolic murmur associated with a third heart sound and a middiastolic murmur that results from the increased blood flow crossing antegrade across the mitral valve in diastole.

Several bedside maneuvers may also be useful in the evaluation of cardiac murmurs.[51] Right-sided murmurs (e.g., tricuspid regurgitation) increase in intensity during inspiration because of augmented right heart filling. The murmur of hypertrophic cardiomyopathy is extremely dependent on left ventricular filling, such that both the strain phase of the Valsalva maneuver and moving from squatting to the standing position augment the intensity of the murmur.

DIAGNOSTIC EVALUATION

An ECG should be obtained to evaluate for the presence of cardiac chamber enlargement and hypertrophy. Aortic stenosis imposes a pressure overload to the left ventricle, resulting in left ventricular hypertrophy by ECG in approximately 50% of patients. Hypertrophic cardiomyopathy is characterized by increased ventricular muscle mass, which is usually apparent on the ECG with extreme voltage amplitudes and small Q waves in leads I, aVL, and V4 through V6, referred to as septal Q waves. Mitral stenosis results in left atrial enlargement and occasionally right axis deviation and right ventricular hypertrophy.

Pertinent Web Sites

http://www.aha.org
The American Heart Association maintains this site with information on recent cardiovascular trials, local and national meetings, information for patients, and clinical guidelines.

http://www.acc.org
The American College of Cardiology maintains this site with information on recent cardiovascular trials, local and national meetings, and clinical guidelines.

http://www.theheart.org
This excellent site has current information on clinical trials, pertinent articles, clinical cases, discussion forums, cybersessions, and links to other cardiovascular sites. It frequently has summaries of clinical trials recently reported at the major cardiology meetings.

http://www.vssgbi.org
The Vascular Surgical Society of Great Britain and Ireland maintains this site, which provides patient information.

http://www.tasc-pad.org
The Trans-Atlantic Inter-Society Consensus on the Management of Peripheral Arterial Disease offers management recommendations for intermittent claudication, acute limb ischemia, and critical limb ischemia.

A chest roentgenogram should be reviewed for chamber enlargement and the presence of calcification. Chronic aortic and mitral insufficiency cause a volume overload to the left ventricle and left atrium, respectively. Left atrial enlargement, without enlargement of the left ventricle, and mitral valve calcification suggest mitral stenosis. Calcification of the aortic valve frequently occurs with valvular aortic stenosis but is rarely apparent on the chest roentgenogram.

In the absence of cardiovascular symptoms and other physical findings, a grade 1 or grade 2 midsystolic murmur does not require additional evaluation [*see Figure 7*].[50] Midsystolic murmurs of grade 3 and higher, holosystolic murmurs, or late systolic murmurs should be further evaluated by echocardiography. All patients with a diastolic or continuous murmur should be referred for echocardiography because these murmurs always indicate underlying cardiac pathology. In addition to confirming the etiology of a cardiac murmur, echocardiography provides evaluation of left ventricular systolic and diastolic function, wall motion abnormalities (that may indicate associated CAD), and estimation of pulmonary arterial pressures. For patients with valvular, congenital, or myocardial diseases, echocardiography provides a baseline from which additional studies can be obtained and used to follow disease progression over time.

Cardiovascular Information on the Internet

There are numerous sources of cardiovascular information on the Internet. The most useful general information sites are listed [*see Sidebar* Pertinent Web Sites].

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Lee TH, Goldman L: Evaluation of the patient with acute chest pain. N Engl J Med 341:1187, 2000

2. Ammann P, Sabine M, Kraus M, et al: Characteristics and prognosis of myocardial infarction in patients with normal coronary arteries. Chest 117:333, 2000

3. Goyal RK: Changing focus on unexplained esophageal chest pain. Ann Intern Med 124:1008, 1996

4. Khot UN, Khot MB, Bajzer CT, et al: Prevalence of conventional risk factors in patients with coronary heart disease. JAMA 290:898, 2003

5. Hertzer NR, Beven EG, Young JR, et al: Coronary artery disease in peripheral vascular patients: a classification of 1,000 coronary angiograms and results of surgical management. Ann Surg 199:223, 1984

6. Diamond GA, Forrester JS: Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Engl J Med 300:1350, 1979

7. Pryor DB, Shaw L, McCants CB, et al: Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 118:81, 1993

8. Gibbons RJ, Balady GJ, Bricker JT, et al: ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). J Am Coll Cardiol 40:1531, 2002

9. Chou TM, Amidon TM: Evaluating coronary artery disease noninvasively: which test for whom? West J Med 161:173, 1994

10. Gianrossi R, Detrano R, Mulvihil D, et al: Exercise-induced ST depression in the diagnosis of coronary artery disease: a meta-analysis. Circulation 80:87, 1989

11. Kwok Y, Kim C, Grady D, et al: Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 83:660, 1999

12. Garber AM, Solomon NA: Cost effectiveness of alternative test strategies for the diagnosis of coronary artery disease. Ann Intern Med 130:719, 1999

13. Kim C, Kwok YS, Heagerty P, et al: Pharmacologic stress testing for coronary disease diagnosis: a meta-analysis. Am Heart J 142:934, 2001

14. Gibbons RJ, Chatterjee K, Daley J, et al: ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients with Chronic Stable Angina). J Am Coll Cardiol 33:2092, 1999

15. Gibbons RJ, Abrams J, Chatterjee K, et al: ACC/AHA 2002 guideline update for the management of patients with chronic stable angina: summary article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on the Management of Patients With Chronic Stable Angina). J Am Coll Cardiol 41:159, 2003

16. Cheitlin MD, Alpert JS, Armstrong WF, et al: ACC/AHA Guidelines for the clinical application of echocardiography. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Clinical Application of Echocardiography). Developed in collaboration with the American Society of Echocardiography. Circulation 95:1686, 1997

17. Morise AP, Olson MB, Merz CN, et al: Validation of the accuracy of pretest and exercise test scores in women with a low prevalence of coronary disease: the NHLBI-sponsored Women's Ischemia Syndrome Evaluation (WISE) study. Am Heart J 147:1085, 2004

18. Nissen SE, Grines CL, Gurley JC, et al: Application of a new phased-array ultrasound catheter in the assessment of vascular dimensions: in-vivo comparison to cineangiography. Circulation 82:660, 1990

19. Holleman DR Jr, Simel DL, Goldberg JS: Diagnosis of obstructive airways disease from the clinical examination. J Gen Intern Med 8:63, 1993

20. Manning HL, Schwartzstein RM: Mechanism of disease: pathophysiology of dyspnea. N Engl J Med 333:1547, 1995

21. Schmitt BP, Kushner MS, Wiener SL: The diagnostic usefulness of the history of the patient with dyspnea. J Gen Intern Med 1:386, 1986

22. Mulrow CD, Lucey CR, Farnett LE: Discriminating causes of dyspnea through clinical examination. J Gen Intern Med 8:383, 1993

23. Mueller C, Scholer A, Laule-Kilian K, et al: Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea. N Engl J Med 350:647, 2004

24. Kroenke K, Arrington ME, Mangelsdorff AD: The prevalence of symptoms in medical outpatients and the adequacy of therapy. Arch Intern Med 150:1685, 1990

25. Weber BE, Kapoor WN: Evaluation and outcomes of patients with palpitations. Am J Med 100:138, 1996

26. Lessmeier TJ, Gamperling D, Johnson-Liddon V, et al: Unrecognized paroxysmal supraventricular tachycardia: potential for misdiagnosis as panic disorder. Arch Intern Med 157:537, 1997

27. Zimetbaum P, Josephson ME: Current concepts: evaluation of patients with palpitations. N Engl J Med 338:1369, 1998

28. Summerton N, Mann S, Rigby A, et al: New-onset palpitations in general practice: assessing the discriminant value of items within the clinical history. Fam Pract 18:383, 2001

29. Leitch JW, Klein GJ, Yee R, et al: Syncope associated with supraventricular tachycardia: an expression of tachycardia rate or vasomotor response? Circulation 85:1064, 1992

30. Brugada P, Gursoy S, Brugada J, et al: Investigation of palpitations. Lancet 341:1254, 1993

31. Kinlay S, Leitch JW, Neil A, et al: Cardiac event recorders yield more diagnoses and are more cost-effective than 48-hour Holter monitoring in patients with palpitations: a controlled clinical trial. Ann Intern Med 124:16, 1996

32. Fogel RI, Evans JJ, Prystowsky EN: Utility and cost of event recorders in the diagnosis of palpitations, presyncope and syncope. Am J Cardiol 79:207, 1997

33. Hayes OW: Evaluation of syncope in the emergency department. Emerg Med Clin North Am 16:601, 1998

34. Hanlon JT, Linzer M, MacMillan JP, et al: Syncope and presyncope associated with probable adverse drug reactions. Arch Intern Med 150:230, 1990

35. Kapoor WN: Evaluation and outcome of patients with syncope. Medicine (Baltimore) 69:160, 1990

36. Brieger D, Eagle KA, Goodman SG, et al: Acute coronary syndromes without chest pain, an underdiagnosed and undertreated high-risk group: insights from the Global Registry of Acute Coronary Events. Chest 126:461, 2004

37. Linzer M, Yang EH, Estes M III, et al: Diagnosing syncope: value of history, physical examination and electrocardiography (pt 1). Ann Intern Med 126:989, 1997

38. Littmann L, Monroe MH, Kerns WP 2nd, et al: Brugada syndrome and "Brugada sign": clinical spectrum with a guide for the clinician. Am Heart J 145:768, 2003

39. Linzer M, Yang EH, Estes M III, et al: Diagnosing syncope: unexplained syncope (pt 2). Ann Intern Med 127:76, 1997

40. Sivakumaran S, Krahn AD, Klein GJ, et al: A prospective randomized comparison of loop recorders versus Holter monitors in patients with syncope or presyncope. Am J Med 115:1, 2003

41. Brignole M, Menozzi C, Gianfranchi L, et al: Neurally mediated syncope detected by carotid sinus massage and head-up tilt table test in sick sinus syndrome. Am J Cardiol 68:1032, 1991

42. McIntosh SJ, Lawson J, Kenny RA: Clinical characteristics of vasodepressor, cardioinhibitory, and mixed carotid sinus syndrome in the elderly. Am J Med 95:203, 1993

43. Kapoor WN, Smith MA, Miller NL: Upright tilt testing in evaluating syncope: a comprehensive literature review. Am J Med 97:78, 1994

44. Newman A, Shemanski L, Manolio T, et al: Ankle-arm index as a predictor of cardiovascular disease and mortality in the Cardiovascular Health Study. Arterioscler Thromb Vasc Biol 19:538, 1999

45. Hirsch AT, Criqui M, Treat-Jacobson D, et al: Peripheral arterial disease detection, awareness, and treatment in primary care. JAMA 286:1317, 2001

46. Leng GC, Fowkes FGR, Lee AJ, et al: Use of the ankle brachial pressure index to predict cardiovascular events and death: a cohort study. BMJ 313:1440, 1996

47. Lee AJ, Price JF, Russell MJ, et al: Improved prediction of fatal myocardial infarction using the ankle brachial index in addition to conventional risk factors: the Edinburgh

Artery Study. Circulation 110:3075, 2004

48. Hiatt WR: Medical treatment of peripheral arterial disease and claudication. N Engl J Med 344:1608, 2001

49. Stewart BF, Siscovick D, Lind BK, et al: Clinical factors associated with calcific aortic valve disease. J Am Coll Cardiol 29:630, 1997

50. Bonow RO, Carabello B, de Leon AC Jr, et al: ACC/AHA guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Valvular Heart Disease). J Am Coll Cardiol 32:1486, 1998

51. Lembo NJ, Dell'Italia LJ, Crawford MH, et al: Bedside diagnosis of systolic murmurs. N Engl J Med 318:1572, 1988

II HEART FAILURE

Shannon Winakur, m.d.
Mariell Jessup, m.d.

Definition

Heart failure is a clinical syndrome resulting from a structural or functional cardiac disorder that impairs the ability of the ventricle to fill with or eject blood to meet the needs of the body. This syndrome, which is a constellation of signs and symptoms, is primarily manifested by dyspnea, fatigue, fluid retention, and decreased exercise tolerance. Heart failure may result from disorders of the pericardium, the myocardium, the endocardium, valvular structures, or the great vessels of the heart or from rhythm disturbances. It is important to emphasize that not all patients with heart failure symptoms have similar cardiac structural abnormalities. Indeed, the major thrust of an initial evaluation of a patient with heart failure is to define the cardiac abnormalities responsible for the symptoms.

Classification

Heart failure has been classified in many ways. One useful framework involves describing the underlying cardiomyopathy, which frequently will suggest the etiology [*see Table 1 and Figure 1*].[1-4] Some examples of the World Health Organization (WHO) classification include ischemic cardiomyopathy, hypertrophic or restrictive cardiomyopathy, and idiopathic dilated cardiomyopathy. In the United States, the most common cause of heart failure is ischemic cardiomyopathy from coronary artery disease (CAD).[5,6]

Another practical approach for classification is to divide patients with heart failure into those with primarily systolic dysfunction and those with diastolic dysfunction. For the clinician, this usually means assessing the patient's left ventricular ejection fraction (LVEF), most commonly with echocardiography.[7,8] Patients with systolic heart failure typically have a low LVEF (usually less than 40% to 45%), a dilated left ventricular cavity, and a reduced cardiac output because of diminished contractility of the myocardium. In contrast, patients with diastolic heart failure have a normal LVEF and normal contractility, but there is impaired filling of the heart secondary to a variety of pathophysiologic abnormalities.[9-11]

Despite an increased understanding of the etiologies and pathophysiology of heart failure and advancements in treatment, morbidity and mortality remain unacceptably high for the majority of patients stricken with this disorder.[12,13] Most experts agree that earlier recognition of the syndrome or better identification of patients at risk for heart failure may be our best hope for the future reduction of heart failure's death toll. This is analogous to the concerted efforts to screen for cancer at its earliest stages, before the disease can defy therapy. Consequently, the committee charged with revising the American College of Cardiology/American Heart Association (ACC/AHA) Guidelines for the Evaluation and Management of Heart Failure took the bold step of developing a new classification for patients with heart failure.[14] These guidelines can be obtained from the ACC or AHA Web sites: http://www.acc.org/clinical/statements.htm and http://my.americanheart.org/portal/professional/guidelines.

The ACC/AHA classification emphasizes the evolution and progression of heart failure; it defines four stages of the disorder [*see Table 2 and Figure 2*]. Stage A identifies patients who are at high risk for developing heart failure but who have no apparent structural abnormality of the heart. This includes patients with hypertension, diabetes, or CAD; patients with a history of rheumatic fever, alcohol abuse, or exposure to cardiotoxic drugs; and patients with a family history of cardiomyopathy. Stage B denotes patients with a structural abnormality of the heart but in whom symptoms of heart failure have not yet developed. This includes patients found to have left ventricular hypertrophy or dilatation, a decreased LVEF, or valvular disease, as well as patients with prior myocardial infarction. Stage C refers to patients with a structural abnormality of the heart and symptoms of heart failure. This would include patients with dyspnea, fatigue, or fluid overload, as well as patients with a prior diagnosis of heart failure who are receiving treatment that has relieved their symptoms. Importantly, once patients have had symptoms of heart failure, they remain in stage C even if they subsequently experience clinical improvement. Stage D includes the patient with end-stage heart failure that is refractory to standard treatment. Typical patients include those who require frequent hospital admissions for heart failure, are awaiting a heart transplant, are being supported with intravenous agents or mechanical assist devices, or are receiving hospice care for end-stage heart failure.

The ACC/AHA classification is a departure from the traditional New York Heart Association (NYHA) classification, which characterizes patients by symptom severity.[15] Patients with heart failure may progress from stage A to stage D, but never the reverse. In contrast, many patients with NYHA class IV symptoms can be restored to class II with appropriate therapy. The ACC/AHA classification highlights the importance of known risk factors and structural abnormalities in the development of heart failure. Additionally, it reinforces the concept that heart failure is a progressive disease whose onset can be prevented, or its progression halted, by early identification and intervention.

Epidemiology

Heart failure is one of the major public health problems in the United States today, both in terms of the number of patients affected and health care dollars spent. Nearly five million patients have heart failure, and almost 500,000 patients are diagnosed with the disease each year. Estimated direct and indirect costs for heart failure came to $21 billion in 2001, more than 5% of that year's health care budget[16]; annual spending on drugs for heart failure treatment is about $500 million.[14] Hospitalizations for heart failure increased by 159% from 1979 to 1998,[16] and this trend will likely continue as the United States population ages.

Heart failure is primarily a disease of the elderly.[17] Approximately 6% to 10% of people older than 65 years have heart failure,[18] and roughly 80% of patients hospitalized with heart failure are older than 65 years.[19] More Medicare dollars are spent on heart failure than on any other disease, and heart failure is the most common Medicare diagnosis-related group.[5]

Table 1 Examples of Descriptive and Etiologic Classifications of Heart Failure

Classification Scheme	*Disorder or Disease Process*	*Comments*
By disorder	Dilated cardiomyopathy	Dilatation and impaired function of left ventricle or both ventricles; multiple etiologies: ischemia, valvular disease, infectious process, inflammatory process, toxins, familial/genetic cause, idiopathic
	Hypertrophic cardiomyopathy	Hypertrophy of left ventricle or both ventricles, often asymmetrical and involving the interventricular septum; often associated with mutations in sarcoplasmic proteins; associated with arrhythmias and sudden death
	Restrictive cardiomyopathy	Usually associated with normal systolic function and impaired diastolic function; can be idiopathic or associated with infiltrative diseases, such as amyloidosis, sarcoidosis, and endomyocardial fibrosis
	Arrhythmogenic right ventricular cardiomyopathy	Replacement of myocardium with fatty tissue; can involve left ventricle as well; associated with ventricular arrhythmias; may have a genetic component
By underlying disease process	Ischemic heart disease	Secondary to coronary artery disease
	Valvular disease	Caused by primary valvular disease
	Hypertension	Usually associated with left ventricular hypertrophy; can involve systolic and/or diastolic dysfunction
	Diabetes mellitus	Associated with systolic and/or diastolic dysfunction and left ventricular hypertrophy, even independent of coexisting hypertension or coronary artery disease
	Inflammatory/infectious disease	Systolic dysfunction from myocarditis; multiple infectious etiologies, both viral (e.g., coxsackievirus, echovirus, HIV) and bacterial (rheumatic fever)
	Metabolic disorders	Associated with endocrine abnormalities (e.g., hyperthyroidism, hypothyroidism), electrolyte deficiencies (potassium, magnesium), nutritional deficiencies (e.g., beriberi), and glycogen storage disease (e.g., Pompe disease, Gaucher disease)
	General systemic disease	Associated with connective tissue diseases (e.g., systemic lupus erythematosus, rheumatoid arthritis) and infiltrative diseases (e.g., sarcoidosis, amyloidosis)
	Muscular dystrophies	Includes Duchenne, Becker, and myotonic muscular dystrophies
	Neuromuscular disease	Includes Friedreich ataxia and Noonan syndrome
	Toxins	Associated with alcohol and cocaine abuse, treatment with cardiotoxic chemotherapeutic agents (e.g., anthracyclines), and radiation therapy
	Tachycardia	Associated with uncontrolled tachycardias (e.g., atrial fibrillation and other supraventricular tachycardias)
	Genetic/familial disorders	Associated with a family history of cardiomyopathy and/or sudden death; many cardiomyopathies previously designated as idiopathic may fall into this category
	Pregnancy	Manifests in peripartum period

It is important to recognize that heart failure has diverse causes and affects diverse populations. Until recently, this diversity was not reflected in the composition of heart failure trials in the United States, which typically enrolled middle-aged white men with ischemic cardiomyopathy. In fact, the heart failure population in the United States includes significant numbers of women, elderly persons, and members of racial minorities—and these patients tend to have different forms of heart failure. For example, in an estimated 20% to 50% of patients with heart failure, ventricular systolic function is preserved (i.e., the patients have diastolic heart failure), and these patients are more likely to be elderly women.[20-23]

Etiology

CAD is responsible for roughly two thirds of cases of heart failure in the United States.[24] Coronary ischemia or infarction can lead to heart failure through a variety of mechanisms: acute coronary syndromes or infarction can cause acute heart failure in an otherwise normal heart; likewise, repeated insults of ischemia or infarction can cause a chronic cardiomyopathy. Moreover, many patients with diastolic heart failure have underlying CAD.

Ventricular dysfunction can result from a multitude of nonischemic causes [*see Table 1*]. These include hypertension; diabetes; valvular disease; arrhythmias; myocardial toxins; myocarditis from a variety of infectious agents (including HIV); and hypothyroidism. Infiltrative causes of ventricular dysfunction, which are usually associated with restrictive cardiomyopathy, include amyloidosis, hemochromatosis, and sarcoidosis. Myocardial systolic dysfunction for which there is no apparent cause is labeled idiopathic cardiomyopathy. Over the past several years, there has been increased recognition that many of these so-called idiopathic dilated cardiomyopathies are familial; a number of centers are actively focusing on the identification of the genetic irregularities responsible for the abnormal phenotypes.[25]

Pathophysiology

There is no single, simple model that effectively explains the syndrome of heart failure; currently, the consensus view integrates multiple pathophysiologic models to explain the complex cascade of events leading to this clinical syndrome.[26] The different structural, functional, and biologic changes that culminate in heart failure have led to a variety of treatment

modalities to target this array of causative factors.[27,28] For example, for many years, beta blockers were contraindicated in patients with heart failure because the disorder was thought to be primarily a result of decreased myocardial contractility that would worsen with negative inotropic therapy. However, we have come to realize the central role of pathologic sympathetic activation in heart failure—the maladaptive mechanisms that lead to vasoconstriction, arrhythmias, and ventricular remodeling (see below). This model explains the therapeutic benefits of beta blockade.

The hemodynamic model of heart failure concentrated on the role of increased load on a failing ventricle; this conceptual approach led to the successful use of vasodilators and inotropes. Later, the neurohumoral model of heart failure identified the critical importance of the renin-angiotensin-aldosterone axis and the sympathetic nervous system in the progression of car-

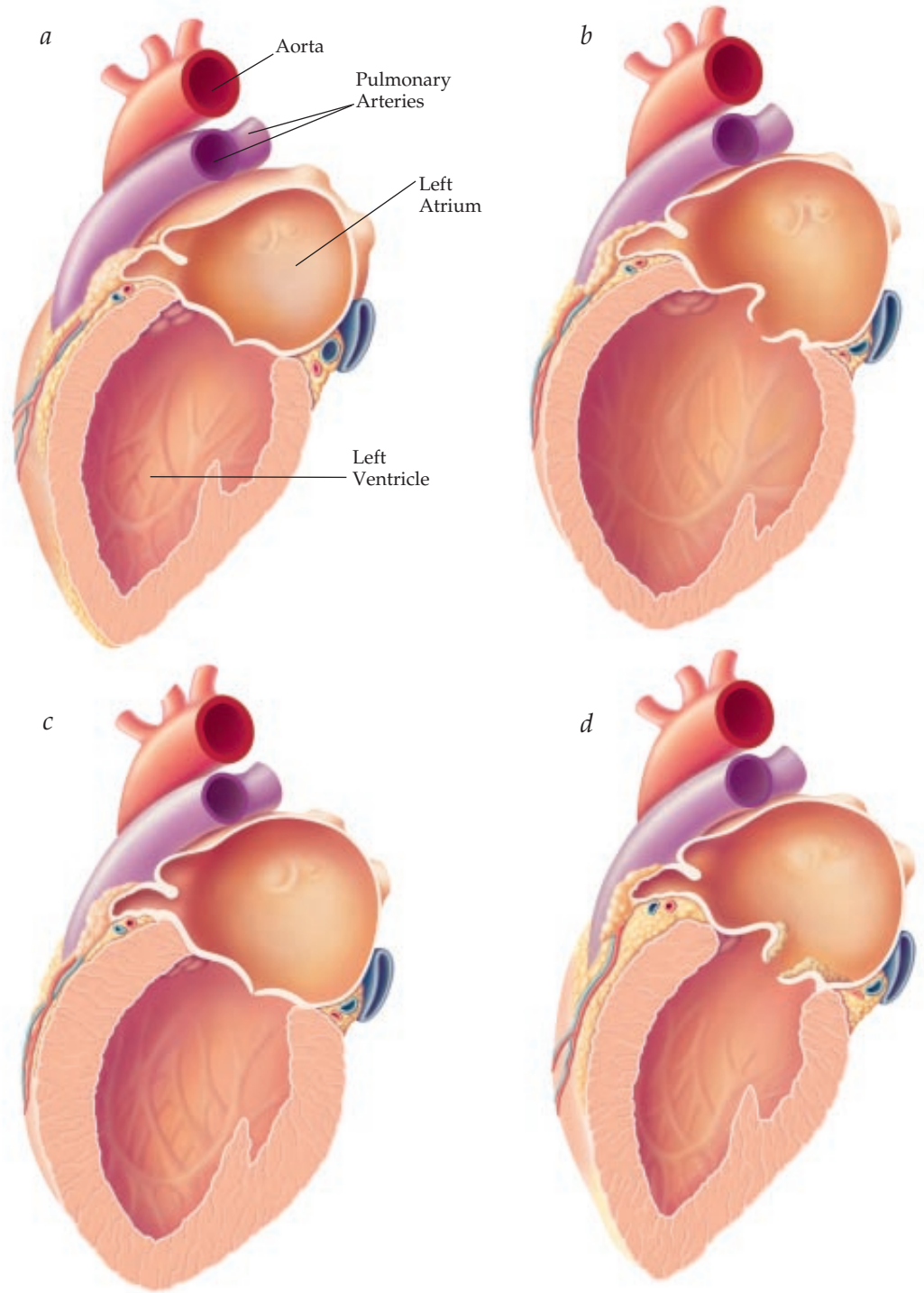


Figure 1 **The different cardiac morphologies in heart failure. The heart is viewed from the left side, with the mitral valve partially cut away; the aortic valve is visible in the upper portion of the left ventricle. (*a*) Normal; (*b*) dilated cardiomyopathy; (*c*) hypertrophic cardiomyopathy; (*d*) diastolic dysfunction.**

Table 2 Stages of Heart Failure[14]

Stage	*Description*	*Examples*
A	Patients at high risk for heart failure because of the presence of conditions strongly associated with the development of heart failure; no identified structural or functional abnormalities of the pericardium, myocardium, or cardiac valves; no current or previous history of signs or symptoms of heart failure	Patients with systemic hypertension, coronary artery disease, diabetes mellitus, history of cardiotoxic drug therapy or alcohol abuse, history of rheumatic fever, family history of cardiomyopathy
B	Patients with structural heart disease that is strongly associated with the development of heart failure but who have no current or previous history of signs or symptoms of heart failure	Patients with left ventricular hypertrophy or fibrosis, left ventricular dilatation or hypocontractility, asymptomatic valvular heart disease, previous myocardial infarction
C	Patients who currently have or who in the past have had symptoms of heart failure associated with underlying structural heart disease	Patients with dyspnea or fatigue due to left ventricular systolic dysfunction; asymptomatic patients undergoing treatment for prior symptoms of heart failure
D	Patients with advanced structural heart disease and marked symptoms of heart failure at rest despite maximal medical therapy; need for specialized interventions	Patients who are frequently hospitalized for heart failure and cannot be safely discharged from the hospital; patients in hospital awaiting heart transplantation; patients at home receiving continuous intravenous support for symptom relief or support with a mechanical circulatory assist device; patients in a hospice setting for the management of heart failure

diac dysfunction, leading to widespread use of angiotensin-converting enzyme (ACE) inhibitors and beta blockers.

The recognition that progressive ventricular dilatation serves as a marker for disease progression has focused attention on the myocyte and on the role of the cardiac interstitium. Both medical and surgical therapies have been directed at this mechanism.

Left ventricular dysfunction begins with an injury to the myocardium. The unanswered question is why ventricular systolic dysfunction continues to worsen in the absence of recurrent insults. This pathologic process, which has been termed remodeling, is the structural response to the initial injury. Mechanical, neurohormonal, and possibly genetic factors alter ventricular size, shape, and function to decrease wall stress and compensate for the initial injury. Remodeling involves hypertrophy, loss of myocytes, and increased fibrosis and is secondary to both neurohormonal activation and other mechanical factors.[29,30] Ultimately, the changes in ventricular shape lead to a less efficient cardiac pump. Functional mitral regurgitation often occurs as the left ventricle dilates and becomes more globular, increasing volume overload. Remodeling seems to beget more remodeling.

Arrhythmias often contribute to myocardial dysfunction and are an unwelcome side effect of heart failure. Supraventricular arrhythmias, particularly atrial fibrillation, often unmask systolic or diastolic dysfunction in a previously asymptomatic patient.[31] In addition, intraventricular conduction delays and bundle branch block are often present in patients with heart failure. Abnormal ventricular conduction, particularly left bundle branch block, has significant detrimental hemodynamic effects.[32-35] In addition to contributing to worsening heart failure, ventricular arrhythmias are likely a direct cause of death in many of these patients; the rate of sudden cardiac death in persons with heart failure is six to nine times that seen in the general population.[36]

These pathophysiologic models do not easily explain diastolic heart failure.[37] In the 20% to 50% of patients who have heart failure despite normal systolic function, cardiac output is limited by abnormal filling and disordered relaxation of the ventricles, especially during exercise. Ventricular pressures are elevated for a given ventricular volume, leading to pulmonary congestion, dyspnea, and peripheral edema identical to that seen in patients with a dilated, poorly contracting heart.[9,11,38,39] CAD or ischemia frequently compounds the impairment of ventricular performance in patients with diastolic heart failure, who typically are elderly women[22] with hypertension, diabetes, and obesity.

Diagnosis

STAGE A

The first step in the diagnosis of heart failure is to identify patients who are at risk for developing the syndrome; this concept was part of the reasoning behind the new ACC/AHA staging system.[14] Patients in stage A are those with CAD, hypertension, diabetes, a history of alcohol abuse or exposure to cardiotoxic drugs (e.g., certain chemotherapeutic agents, cocaine), a history of rheumatic fever, or a family history of cardiomyopathy or sudden death. In these high-risk patients, reversible risk factors should be aggressively treated to prevent heart failure from developing.[40,41]

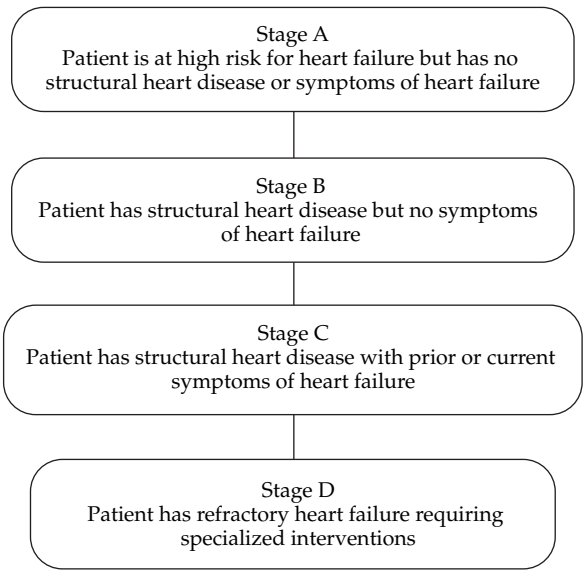


Figure 2 **The evolution of heart failure by stage.[14]**

STAGE B

Stage B patients have asymptomatic, structural heart disease. Echocardiography is easily the best diagnostic tool to uncover left ventricular hypertrophy or dilatation, valvular disease, or wall motion abnormalities indicative of previous myocardial infarction. Patients in stage B represent a significant portion of the heart failure population and constitute a key opportunity for intervention. In a community-based survey, less than half of patients with moderate or severe systolic or diastolic dysfunction, as defined by echocardiographic parameters, had recognized heart failure.[42] At this time, the ACC/AHA guidelines do not recommend routine screening echocardiography for the large number of patients at risk for the development of heart failure. Nonetheless, a class IIa indication is given for noninvasive evaluation of left ventricular function in patients with a strong family history of cardiomyopathy or in those exposed to cardiotoxic therapies.[14]

STAGES C AND D

Stages C and D fit the traditional definition of heart failure. Patients in stage C or D usually present with decreased exercise tolerance, fluid retention, or both. Initial assessment of these patients should focus on the structural abnormality leading to heart failure, as well as evaluation of its etiology. Initial testing should include a 2-D echocardiogram with Doppler flow studies, a chest x-ray, electrocardiography, and laboratory studies, including urinalysis, complete blood count, serum chemistries, liver function studies, and thyroid-stimulating hormone measurement. These tests serve primarily to exclude other potential causes of dyspnea or fatigue.[14] In patients with dyspnea, measurement of serum brain natriuretic peptide (BNP) may aid in the diagnosis; marked elevation of BNP levels suggests that the dyspnea is cardiac rather than pulmonary in origin.[43] Strong consideration should be given to excluding significant CAD, because CAD is the leading cause of left ventricular dysfunction.[24] The ACC/AHA guidelines strongly encourage that patients with heart failure be evaluated with coronary angiography rather than noninvasive testing, even if they do not have a known history of CAD; the guidelines cite the fact that noninvasive testing can often lead to inaccurate results in patients with cardiomyopathies (e.g., perfusion defects or wall motion abnormalities in patients with a nonischemic cardiomyopathy).[14] Nonetheless, some argue that there is little evidence that revascularization changes outcome or prognosis in patients with left ventricular dysfunction and that it should therefore be used only to relieve angina.[44]

Several clinical parameters are useful for the subsequent evaluation and management of heart failure. Patients' weights should be measured in the office, and patients should be taught to follow their weights at home to assess for fluid retention. Office evaluation of jugular venous pressure, hepatojugular reflux, the presence of a gallop rhythm, and peripheral edema can aid in initial diagnosis and can guide the need for diuresis. In addition, these signs of heart failure may be prognostically important.[45]

DIASTOLIC HEART FAILURE

There is no precise definition of diastolic heart failure[11]; the diagnosis is usually made by a clinician who recognizes the typical signs and symptoms of heart failure despite the finding of normal systolic function (i.e., a normal LVEF) on an echocardiogram. Doppler echocardiographic techniques can also aid in establishing the diagnosis of diastolic dysfunction.[46,47]

Table 3 Treatment of Heart Failure[14]

Stage A	Treat hypertension Encourage smoking cessation Treat lipid disorders Encourage regular exercise Discourage alcohol intake, illicit drug use Prescribe ACE inhibitors if appropriate
Stage B	All measures used for stage A ACE inhibitors if appropriate (see text) Beta blockers if appropriate
Stage C	All measures used for stage A Drugs for routine use: ACE inhibitors Beta blockers Digitalis Dietary salt restriction
Stage D	All measures used for stages A, B, and C Mechanical assist devices Heart transplantation Continuous (not intermittent) I.V. inotropic infusions for palliation Hospice care

ACE—angiotensin-converting enzyme

Treatment

Treatment for heart failure is keyed to the stage of the syndrome as defined by the recent ACC/AHA guidelines [*see Table 3*]. Treatment in all stages is aimed at preventing or palliating the remodeling process [*see* Pathophysiology, *above*]. In addition, therapy in stages C and D is intended to relieve the disabling symptoms of heart failure.

STAGE A

The goal of treatment in stage A is to prevent structural heart disease. This is achieved by controlling risk factors (e.g., hypertension, CAD, diabetes mellitus, hyperlipidemia, smoking, alcohol ingestion, and use of cardiotoxic drugs), which lowers the incidence of later cardiovascular events. For example, effective treatment of hypertension decreases left ventricular hypertrophy and cardiovascular mortality; it can also reduce the incidence of heart failure by 30% to 50%.[40,41]

Diabetes deserves particular attention because diabetes patients have a high incidence both of CAD and of heart failure in the absence of CAD; diabetes causes many detrimental biochemical and functional cardiac changes independent of ischemia.[48] ACE inhibitors and angiotensin receptor blockers (ARBs) have assumed a major role in risk reduction for diabetic patients (see below). Studies have shown that in asymptomatic high-risk patients with diabetes or vascular disease who have no history of heart failure or left ventricular dysfunction, treatment with these agents yields significant reductions in death, myocardial infarction, and stroke[49-51] or delays the first hospitalization for heart failure.[52]

STAGES B, C, AND D

The goals of therapy for patients with heart failure and a low LVEF are to decrease the progression of disease and the number of hospitalizations, improve symptoms and survival, and

Table 4 Pharmacotherapy of Heart Failure

Category	*Drug (Trade Name)*	*Dosage*		*Comment*
		Initial Daily Dose	*Maximum Daily Dose*	
Loop diuretics	Bumetanide (Bumex) Furosemide (Lasix) Torsemide (Demadex)	0.5–1 mg q.d. or b.i.d. 20–40 mg q.d. or b.i.d. 10–20 mg q.d. or b.i.d.	Up to 10 mg Up to 400 mg Up to 200 mg	Titrate to achieve dry weight; carefully monitor serum potassium and creatinine levels
ACE inhibitors	Captopril (Capoten) Enalapril (Vasotec) Fosinopril (Monopril) Lisinopril (Prinivil, Zestril) Quinapril (Accupril) Ramipril (Altace)	6.25 mg t.i.d. 2.5 mg b.i.d. 5–10 mg 2.5–5 mg 10 mg b.i.d. 1.25–2.5 mg	50 mg t.i.d. 10–20 mg b.i.d. 40 mg 20–40 mg 40 mg b.i.d. 10 mg	Carefully monitor serum potassium and creatinine levels
Beta blockers	Bisoprolol (Zebeta) Carvedilol (Coreg) Metoprolol tartrate (Lopressor) Metoprolol succinate extended release (Toprol-XL)	1.25 mg 3.125 mg b.i.d. 6.25 mg b.i.d. 12.5–25 mg	10 mg 25 mg b.i.d. (50 mg b.i.d. for patients > 85 kg) 75 mg b.i.d. 200 mg	Titrate dosage up over 2- to 4-week intervals, carefully monitoring for signs and symptoms of fluid overload
Digitalis glycosides	Digoxin (Lanoxin)	0.125–0.25 mg	0.125–0.25 mg	Narrow therapeutic window; monitor levels carefully in older patients and those with renal insufficiency
Aldosterone inhibitors	Spironolactone (Aldactone)	25 mg	50 mg	50 mg q.d. was maximum dosage used in RALES trial[78]; use carefully with concurrent ACE inhibitor or ARB; carefully monitor serum potassium and creatinine levels; use if potassium < 5.0 mmol/L, creatinine < 2.5 mg/dl
Angiotensin receptor blockers	Candesartan (Atacand) Irbesartan (Avapro) Losartan (Cozaar) Valsartan (Diovan)	8 mg 75 mg 25 mg 80 mg	32 mg 300 mg 100 mg 320 mg	Use if patients have cough or angioedema on ACE inhibitor

ACE—angiotensin-converting enzyme ARB—angiotensin receptor blocker RALES—Randomized Aldosterone Evaluation Study

minimize risk factors. Simple interventions can help patients control their disease. For example, basic habits of moderate sodium restriction, weight monitoring, and adherence to medication schedules serve to prevent hospitalizations for rapid fluid overload. Other frequent causes of decompensation in heart failure include anemia, arrhythmias (especially atrial fibrillation), noncompliance with medications and diet, or the use of nonsteroidal anti-inflammatory drugs (NSAIDs).[52-55]

Medical Therapy

Pharmacologic treatment of heart failure routinely includes diuretics, angiotensin antagonists, beta blockers, and digoxin; spironolactone or inotropes may be beneficial in some cases [*see Table 4*].

Diuretics In symptomatic patients in stage C and stage D, diuretics are often the first drugs prescribed to decrease fluid overload and congestive symptoms. Loop diuretics are most often given to these patients, either as maintenance therapy or on an as-needed basis. Loop diuretics can be combined with thiazides to optimize diuresis.[56,57]

ACE inhibitors ACE inhibitors are recommended for all patients in stages B, C, and D. By decreasing the conversion of angiotensin I to angiotensin II, ACE inhibitors minimize the multiple pathophysiologic effects of angiotensin II, such as vasoconstriction and fibrosis. ACE inhibitors (but not ARBs) also decrease the degradation of bradykinin, a substance that causes vasodilation and natriuresis. In patients with heart failure, ACE inhibitors have been shown to improve survival and cardiac performance, to decrease symptoms and hospitalizations, and to decrease or slow the remodeling process.[58-60]

Currently, it is not clear whether all ACE inhibitors are equally effective in all forms of heart failure. There are few data from controlled trials, for example, about the efficacy of ACE inhibitors in diastolic heart failure. Moreover, although several guidelines have emphasized the need to maximize the dose of ACE inhibitor to target levels (rather than using blood pressure alone to guide dose titration), current recommendations underscore the need to add beta blockers to the regimen of patients in stage C early in the course of treatment, even if target ACE inhibitor doses have not been achieved.

Angiotensin receptor blockers What is the role of ARBs in heart failure? These agents block the effects of angiotensin II at the angiotensin II type 1 receptor site. ACC/AHA guidelines recommend the use of ARBs only in patients who cannot tolerate ACE inhibitors because of cough or angioedema[14]; the guidelines stress that ARBs are comparable to ACE inhibitors but are not superior.[61-63] Since publication of the guidelines,

however, several key trials have reported successful intervention with ARBs in stage B and C patients.[64,65] The role of ARBs in patients already on beta blockers, with or without an ACE inhibitor, remains to be elucidated. Symptomatic patients who cannot tolerate ACE inhibitors or ARBs, usually because of renal insufficiency, may benefit from a combination of hydralazine and isosorbide dinitrate for afterload reduction.[66]

Beta blockers Although it was once taught that beta blockers were contraindicated in heart failure secondary to systolic dysfunction, multiple studies have now shown an impressive effect of these drugs on many aspects of heart failure and at all stages of the syndrome. The primary action of these agents is to counteract the harmful effects of the increased sympathetic nervous system activity in heart failure. Beta blockers improve survival, ejection fraction, and quality of life; they also decrease morbidity, hospitalizations, sudden death, and the maladaptive effects of remodeling.[67,68] Long-term placebo-controlled trials have shown improvement in systolic function and reversal of remodeling after 3 to 4 months of treatment with beta blockers.[69-71] A recent analysis showed that even in the sickest of heart failure patients, beta-blocker therapy was well tolerated and led to a decrease in mortality and hospitalizations as early as 14 to 21 days after initiation of therapy.[72] However, clinicians should be extremely cautious about starting beta blockers in patients with significant reactive airway disease, in diabetic patients with frequent episodes of hypoglycemia, or in patients with bradyarrhythmias or heart block who do not have a pacemaker implanted.

In the United States, two beta blockers are specifically approved for treatment of heart failure: carvedilol and long-acting metoprolol. Beta blockers should be started at the lowest possible dose and titrated up slowly at 2- to 4-week intervals. Patients should be closely monitored for worsening of symptoms or fluid retention, which can sometimes occur early in therapy with these agents. If patients do have exacerbations during initiation of beta blockade, diuretic therapy can be increased, and titration of the beta blocker can proceed more slowly.

Digoxin Digoxin has long been a mainstay in the treatment of symptomatic patients with left ventricular dysfunction, despite a lack of data from clinical trials showing benefit. A large randomized study demonstrated that digoxin was successful in decreasing hospitalization for heart failure—an important clinical end point—but did not decrease mortality.[73] Recent post hoc analysis of data from this trial showed that in the patients randomized to receive digoxin therapy, mortality may have been higher in women than in men.[74] It is hypothesized that the therapeutic window for digoxin may be different in men and women, with women perhaps needing a lower dose of the drug.[75] Indeed, data suggest that digoxin improves morbidity as effectively at low serum concentrations (< 0.09 ng/ml) as at higher levels, and with less toxicity.[76] Clinicians should carefully monitor all patients for signs and symptoms of digoxin toxicity, especially those patients who are elderly or have renal dysfunction. Physicians and patients should also keep in mind that digoxin interacts with numerous other drugs.

Spironolactone Another relatively old drug with new data to support its use in heart failure is the aldosterone antagonist spironolactone. Because of the activation of the renin-angiotensin-aldosterone axis, which is incompletely suppressed by ACE inhibitors, patients with heart failure have increased circulating levels of aldosterone. This leads to sodium retention and potassium loss. Aldosterone also works locally within the myocardium, contributing to hypertrophy and fibrosis in the failing heart.[77] A large randomized trial has shown that the addition of low-dose spironolactone (25 mg daily) to standard treatment reduces morbidity and mortality in patients with NYHA class III and IV heart failure (stage C and D patients).[78]

Intravenous inotropes Patients with refractory heart failure (stage D patients) often require intermittent intravenous inotropic therapy to aid in diuresis and to improve symptoms. No survival benefit has been demonstrated with inotropic treatment. These agents have received a class IIb indication in the ACC/AHA guidelines[14]—that is, they are regarded as palliative.

DIASTOLIC HEART FAILURE

Despite the large number of patients with primarily diastolic heart failure, few clinical trials have addressed the management of these cases. Physiologic principles used to guide treatment in these patients include control of blood pressure, heart rate, myocardial ischemia, and blood volume.[14]

REVASCULARIZATION AND SURGICAL THERAPY

Patients in all stages of heart failure must be evaluated for CAD. Angioplasty and surgical revascularization improve ischemic symptoms and can lead to improved ejection fraction and decreased incidence of sudden death.[79]

Clinical trials to investigate the role of surgical interventions in halting or reversing the remodeling process are now under way. Such interventions include mitral valve repair or replacement, mechanical devices to reduce wall stress, and surgical excision of infarcted tissue.[80-83]

Cardiac transplantation remains the only definitive treatment for stage D patients, but it is available only to roughly 2,500 patients a year in the United States.[84] Left ventricular assist devices are available to support patients waiting for heart transplant. There is growing evidence supporting the use of these devices as destination therapy for stage D patients, many of whom are not eligible for cardiac transplantation.[85]

IMPLANTED DEVICES

Biventricular Pacing Systems

Many heart failure patients have intraventricular conduction delays that may contribute to altered myocardial contractility or dyssynchrony. Biventricular pacing is a novel therapy for patients with left ventricular systolic dysfunction, particularly those with a left bundle branch block. The goal of this therapy is to restore the usual pattern of electrical activation of the left ventricle and thereby restore ventricular synchrony. Pacing leads are placed in the right atrium and the right ventricle and into a cardiac vein in the lateral wall of the left ventricle via the coronary sinus. There is evidence that with restored ventricular synchrony from a biventricular pacing system, the remodeling process is halted and reversed. Trials have shown that implantation of a biventricular pacer results in decreased ventricular size and volumes, improved ventricular function, and less mitral regurgitation. This has led to improved exercise tolerance, decreased hospitalizations, and improved quality of life.[86-88] Although individual randomized trials have not shown a mortality benefit for biventricular pacing, a recent meta-analysis of four of the largest trials to date showed a 51% decrease in death

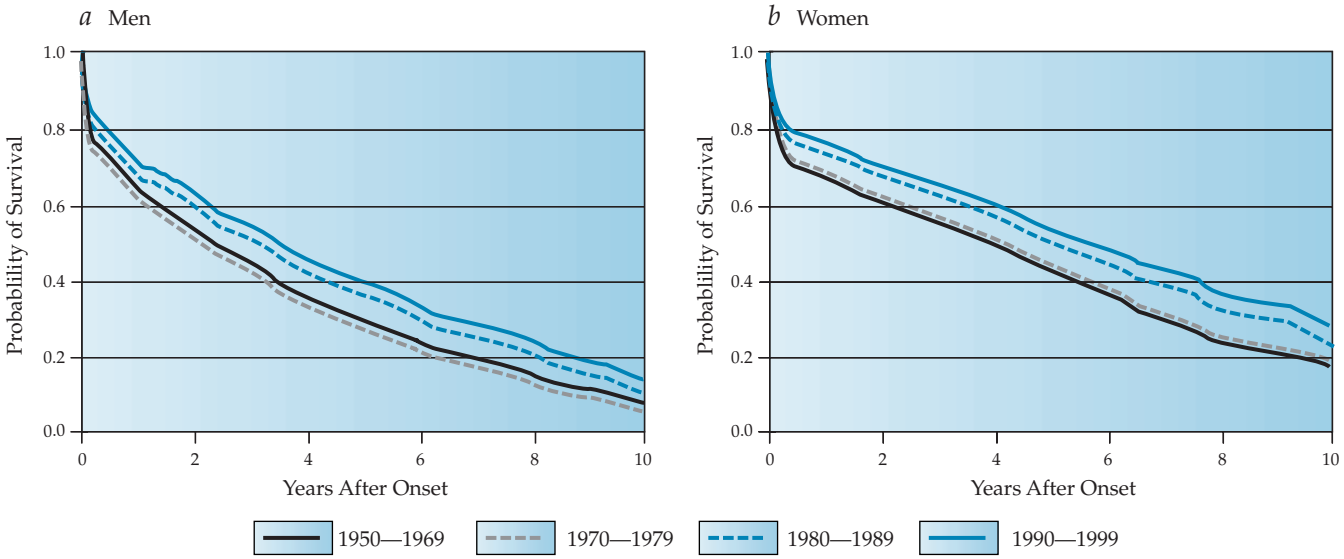


Figure 3 **Data from the Framingham Heart Study indicate a steady upward trend since the 1950s in age-adjusted survival after the onset of heart failure.[96] Estimates shown are for patients 65 to 74 years of age.**

from progressive heart failure.[89] In addition, a large clinical trial of biventricular pacing in patients with heart failure was stopped early because resynchronization therapy was found to confer a statistically significant benefit regarding the combined end point of mortality and hospitalization.[90]

Cardioverter-Defibrillators

The use of implantable cardioverter-defibrillators (ICDs) for the primary prevention of sudden death in patients with left ventricular dysfunction has grown enormously in recent years. There is increasing evidence that ICD placement reduces mortality in patients with ischemic cardiomyopathy, regardless of whether they have nonsustained ventricular arrhythmias.[91] The role of these devices in patients with heart failure of a nonischemic cause has yet to be elucidated and is the subject of several ongoing trials.

Prognosis

Despite many advances in the management of heart failure, this disorder remains life-threatening. Symptomatic heart failure continues to confer a worse prognosis than the majority of cancers in the United States, with 1-year mortality averaging 45%.[12,13] Nonetheless, it is difficult to discuss the prognosis of heart failure as a whole, because an individual patient's likelihood of survival is related to the cause of the heart failure, as well as multiple other clinical factors. For example, given the same severity of heart failure symptoms, an 85-year-old woman with ischemic cardiomyopathy would have a lower likelihood of survival than a 45-year-old man with idiopathic cardiomyopathy. One study of 1,230 patients with cardiomyopathy found that survival was significantly worse in patients with cardiomyopathy from ischemia, infiltrative disease, cardiotoxic chemotherapy, HIV infection, or connective tissue disease than in patients with idiopathic cardiomyopathy.[92]

There are conflicting data about the prognosis of diastolic heart failure. However, recent studies have shown that mortality in these cases may be as high as in systolic heart failure, and hospitalization rates are equal.[42,93]

It is also important for clinicians to remember that a low LVEF is not universally predictive of poor outcome. In patients referred for transplantation, survival has correlated more closely with other variables—notably, peak exercise oxygen consumption.[94] One prospectively validated model for predicting survival in patients with severe heart failure incorporates LVEF with six other clinical factors: presence of coronary disease, resting heart rate, mean arterial blood pressure, presence of intraventricular conduction delays, serum sodium concentration, and peak exercise oxygen consumption.[95] These tools can be used to stratify patients according to risk and to make the most appropriate use of modern therapies and treatment modalities.

How can we improve the prognosis of patients with heart failure? A recent report from the Framingham Heart Study has shown promising evidence of increasing survival after the diagnosis of heart failure [*see Figure 3*].[96] To further this trend, we must work toward widespread implementation of the therapies known to decrease morbidity and mortality in heart failure. We must also investigate more completely the impact of medical therapy on the survival of patients with diastolic heart failure. There should be continued efforts to increase the number of traditionally underrepresented patients (e.g., women and minorities) enrolled in heart failure trials. Finally, in keeping with the emphasis of the ACC/AHA guidelines, we must concentrate on identifying and treating those patients at greatest risk for heart failure to prevent it from occurring.

Shannon Winakur, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection. Mariell Jessup, M.D., during the past 12 months has received grants for clinical research from, and served as an advisor or consultant to, Acorn Cardiovascular, Inc.; Medtronic, Inc.; Guidant Corporation; GlaxoSmithKline; AstraZeneca Pharmaceuticals LP; and Scios Inc.

References

1. Boffa GM, Thiene G, Nava A, et al: Cardiomyopathy: a necessary revision of the WHO classification. Int J Cardiol 30:1, 1991

2. Goodwin JF: Overview and classification of the cardiomyopathies. Cardiovasc Clin 19:3, 1988

3. Keren A, Popp RL: Assignment of patients into the classification of cardiomy-

opathies. Circulation 86:1622, 1992

4. Pisani B, Taylor DO, Mason JW: Inflammatory myocardial diseases and cardiomyopathies. Am J Med 102:459, 1997

5. Massie BM, Shah NB: Evolving trends in the epidemiologic factors of heart failure: rationale for preventive strategies and comprehensive disease management. Am Heart J 133:701, 1997

6. Wilhelmsen L, Rosengren A, Eriksson H, et al: Heart failure in the general population of men: morbidity, risk factors and prognosis. J Intern Med 249:253, 2001

7. Gadsboll N, Hoilund-Carlsen PF, Neilsen GG, et al: Interobserver agreement and accuracy of bedside estimation of right and left ventricular ejection fraction in acute myocardial infarction. Am J Cardiol 63:1301, 1989

8. Ghali JK, Kadakia S, Cooper RS, et al: Bedside diagnosis of preserved versus impaired left ventricular systolic function in heart failure. Am J Cardiol 67:1002, 1991

9. Zile MR, Brutsaert DL: New concepts in diastolic dysfunction and diastolic heart failure. Part II: causal mechanisms and treatment. Circulation 105:1503, 2002

10. Zile MR, Brutsaert DL: New concepts in diastolic dysfunction and diastolic heart failure. Part I: diagnosis, prognosis, and measurements of diastolic function. Circulation 105:1387, 2002

11. Vasan RS, Levy D: Defining diastolic heart failure: a call for standardized diagnostic criteria. Circulation 101:2118, 2000

12. Khand A, Gemmel I, Clark AL: Is the prognosis of heart failure improving? J Am Coll Cardiol 36:2284, 2000

13. Konstam MA: Progress in heart failure management? Lessons from the real world. Circulation 102:1076, 2000

14. Hunt SA, Baker DW, Chin MH, et al: ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to revise the 1995 Guidelines for the Evaluation and Management of Heart Failure). J Am Coll Cardiol 38:2101, 2001

15. Gibelin P: An evaluation of symptom classification systems used for the assessment of patients with heart failure in France. Eur J Heart Failure 3:739, 2001

16. 2001 Heart and Stroke Statistical Update. American Heart Association, Dallas, Texas, 2000

17. Kannel WB, Belanger AJ: Epidemiology of heart failure. Am Heart J 121:951, 1991

18. Kannel WB: Epidemiology and prevention of cardiac failure: Framingham Study insights. Eur Heart J 8:23, 1987

19. Haldeman GA, Croft JB, Giles WH, et al: Hospitalization of patients with heart failure: National Hospital Discharge Survey, 1985 to 1995. Am Heart J 137:352, 1999

20. Senni M, Tribouilloy CM, Rodeheffer RJ, et al: Congestive heart failure in the community: a study of all incident cases in Olmsted County, Minnesota, in 1991. Circulation 98:2282, 1998

21. Vasan RS, Larson MG, Benjamin EJ, et al: Congestive heart failure in subjects with normal versus reduced left ventricular ejection fraction. J Am Coll Cardiol 33:1948, 1999

22. Masoudi FA, Havranek EP, Smith G, et al: Gender, age, and heart failure with preserved left ventricular systolic function. J Am Coll Cardiol 41:217, 2003

23. Importance of heart failure with preserved systolic function in patients ≥ 65 years of age. Cardiovascular Health Study (CHS) Research Group. Am J Cardiol 87:413, 2001

24. Gheorghiade M, Bonow RO: Chronic heart failure in the United States: a manifestation of coronary artery disease. Circulation 97:282, 1998

25. Schönberger J, Seidman CE: Many roads lead to a broken heart: the genetics of dilated cardiomyopathy. Am J Hum Genet 69:249, 2001

26. Mann DL: Mechanisms and models in heart failure: a combinatorial approach. Circulation 100:999, 1999

27. McMurray J, Pfeffer MA: New therapeutic options in congestive heart failure: part I. Circulation 105:2099, 2002

28. McMurray J, Pfeffer MA: New therapeutic options in congestive heart failure: part II. Circulation 105:2223, 2002

29. Sutton MGSJ, Sharpe N: Left ventricular remodeling after myocardial infarction: pathophysiology and therapy. Circulation 101:2981, 2000

30. Eichhorn EJ, Bristow MR: Medical therapy can improve the biological properties of the chronically failing heart: a new era in the treatment of heart failure. Circulation 94:2285, 1996

31. Benjamin EJ, Wolf PA, D'Agostino RB, et al: Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation 98:946, 1998

32. Gerber TC, Nishimura RA, Holmes DR Jr, et al: Left ventricular and biventricular pacing in congestive heart failure. Mayo Clin Proc 76:803, 2001

33. Hultgren HN, Craige E, Fujii J, et al: Left bundle branch block and mechanical events of the cardiac cycle. Am J Cardiol 52:755, 1983

34. Sadaniantz A, Saint Laurent L: Left ventricular Doppler diastolic filling patterns in patients with isolated left bundle branch block. Am J Cardiol 81:643, 1998

35. Xiao HB, Lee CH, Gibson DG: Effect of left bundle branch block on diastolic function in dilated cardiomyopathy. Br Heart J 66:443, 1991

36. Stevenson WG, Stevenson LW: Prevention of sudden death in heart failure. J Cardiovasc Electrophysiol 12:112, 2001

37. Burkhoff D, Maurer MS, Packer M: Heart failure with a normal ejection fraction: is it really a disorder of diastolic function? Circulation 107:656, 2003

38. Banerjee P, Banerjee T, Khand A, et al: Diastolic heart failure: neglected or misdiagnosed? J Am Coll Cardiol 39:138, 2002

39. Brutsaert DL, Sys SU: Diastolic dysfunction in heart failure. J Cardiac Failure 3:225, 1997

40. Mosterd A, D'Agostino RB, Silbershatz H, et al: Trends in the prevalence of hypertension, antihypertensive therapy, and left ventricular hypertrophy from 1950 to 1989. N Engl J Med 340:1221, 1999

41. Deedwania PC: Hypertension and diabetes: new therapeutic options. Arch Intern Med 160:1585, 2000

42. Redfield MM, Jacobsen SJ, Burnett JC Jr, et al: Burden of systolic and diastolic ventricular dysfunction in the community. JAMA 289:194, 2003

43. Morrison LK, Harrison A, Krishnaswamy P, et al: Utility of a rapid B-natriuretic peptide assay in differentiating congestive heart failure from lung disease in patients presenting with dyspnea. J Am Coll Cardiol 39:202, 2002

44. Cleland JGF, Alamgir F, Nikitin NP, et al: What is the optimal medical management of ischemic heart failure? Prog Cardiovasc Dis 43:433, 2001

45. Drazner MH, Rame JE, Stevenson LW, et al: Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure. N Engl J Med 345:574, 2001

46. Nishimura RA, Tajik AJ: Evaluation of diastolic filling of left ventricle in health and disease: Doppler echocardiography is the clinician's Rosetta Stone. J Am Coll Cardiol 30:8, 1997

47. Ommen SR, Nishimura RA, Appleton CP, et al: Clinical utility of Doppler echocardiography and tissue Doppler imaging in estimation of left ventricular filling pressures: a comparative simultaneous Doppler-catheterization study. Circulation 102:1788, 2000

48. Taegtmeyer H, McNulty P, Young ME: Adaptation and maladaptation of the heart in diabetes: part I. Circulation 105:1727, 2002

49. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. UK Prospective Diabetes Study Group. BMJ 317:703, 1998

50. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Lancet 355:253, 2000

51. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. N Engl J Med 342:145, 2000

52. Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. RENAAL Study Investigators. N Engl J Med 345:861, 2001

53. Page J, Henry D: Consumption of NSAIDs and the development of congestive heart failure in elderly patients: an underrecognized public health problem. Arch Intern Med 160:777, 2000

54. Hambrecht R, Gielen S, Linke A, et al: Effects of exercise training on left ventricular function and peripheral resistance in patients with chronic heart failure: a randomized trial. JAMA 283:3095, 2000

55. Coats AJS: Exercise training for heart failure: coming of age. Circulation 99:1138, 1999

56. Ellison DH: Diuretic drugs and the treatment of edema: from clinic to bench and back again. Am J Kidney Dis 23:623, 1994

57. Brater DC: Diuretic therapy. N Engl J Med 339:387, 1998

58. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group on ACE Inhibitor Trials. JAMA 273:1450, 1995

59. Khalil ME, Basher AW, Brown EJ Jr, et al: A remarkable medical story: benefits of angiotensin-converting enzyme inhibitors in cardiac patients. J Am Coll Cardiol 37:1757, 2001

60. Münzel T, Keaney JF Jr: Are ACE inhibitors a "magic bullet" against oxidative stress? Circulation 104:1571, 2001

61. Havranek EP, Thomas I, Smith WB, et al: Dose-related beneficial long-term hemodynamic and clinical efficacy of irbesartan in heart failure. J Am Coll Cardiol 33:1174, 1999

62. Pitt B, Poole-Wilson PA, Segal R, et al: Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial—the Losartan Heart Failure Survival Study ELITE II. Lancet 355:1582, 2000

63. Pitt B, Segal R, Martinez FA, et al: Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). Lancet 349:747, 1997

64. Cohn JN, Tognoni G: A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. N Engl J Med 345:1667, 2001

65. Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. LIFE Study Group. Lancet 359:995, 2002

66. Gomberg-Maitland M, Baran DA, Fuster V: Treatment of congestive heart failure: guidelines for the primary care physician and the heart failure specialist. Arch Intern Med 161:342, 2001

67. Foody JM, Farrell MH, Krumholz HM: β-Blocker therapy in heart failure: scientific review. JAMA 287:883, 2002

68. Farrell MH, Foody JM, Krumholz HM: β-Blockers in heart failure: clinical applications. JAMA 287:890, 2002

69. Bristow MR: β-Adrenergic receptor blockade in chronic heart failure. Circulation 101:558, 2000

70. Groenning BA, Nilsson JC, Sondergaard L, et al: Antiremodeling effects on the left ventricle during beta-blockade with metoprolol in the treatment of chronic heart failure. J Am Coll Cardiol 36:2072, 2000

71. Hall SA, Cigarroa CG, Marcoux L, et al: Time course of improvement in left ventricular function, mass and geometry in patients with congestive heart failure treated with beta-adrenergic blockade. J Am Coll Cardiol 25:1154, 1995

72. Effects of initiating carvedilol in patients with severe chronic heart failure: results from the COPERNICUS study. Carvedilol Prospective Randomized Cumulative Survival (COPERNICUS) Study Group. JAMA 289:712, 2003

73. The effect of digoxin on mortality and morbidity in patients with heart failure. The Digitalis Investigation Group. N Engl J Med 336:525, 1997

74. Rathore SS, Wang Y, Krumholz HM: Sex-based differences in the effect of digoxin for the treatment of heart failure. N Engl J Med 347:1403, 2002

75. Eichhorn EJ, Gheorghiade M: Digoxin: new perspective on an old drug. N Engl J Med 347:1394, 2002

76. Adams KF Jr, Gheorghiade M, Uretsky BF, et al: Clinical benefits of low serum digoxin concentrations in heart failure. J Am Coll Cardiol 39:946, 2002

77. Weber KT: Aldosterone in congestive heart failure. N Engl J Med 345:1689, 2001

78. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. N Engl J Med 341:709, 1999

79. Baumgartner WA: What's new in cardiac surgery. J Am Coll Surg 192:345, 2001

80. Bitran D, Merin O, Klutstein MW, et al: Mitral valve repair in severe ischemic cardiomyopathy. J Card Surg 16:79, 2001

81. Bishay ES, McCarthy PM, Cosgrove DM, et al: Mitral valve surgery in patients with severe left ventricular dysfunction. Eur J Cardiothorac Surg 17:213, 2000

82. Raman JS, Hata M, Storer M, et al: The mid-term results of ventricular containment (ACORN WRAP) for end-stage ischemic cardiomyopathy. Ann Thorac Cardiovasc Surg 7:278, 2001

83. Starling RC, McCarthy PM, Buda T, et al: Results of partial left ventriculectomy for dilated cardiomyopathy: hemodynamic, clinical and echocardiographic observations. J Am Coll Cardiol 36:2098, 2000

84. Hosenpud JD, Bennett LE, Keck BM, et al: The Registry of the International Society for Heart and Lung Transplantation: eighteenth Official Report—2001. J Heart Lung Transplant 20:805, 2001

85. Jessup M: Mechanical cardiac-support devices: dreams and devilish details. N Engl J Med 345:1490, 2001

86. Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. Multisite Stimulation in Cardiomyopathies (MUSTIC) Study Investigators. N Engl J Med 344:873, 2001

87. Touiza A, Etienne Y, Gilard M, et al: Long-term left ventricular pacing: assessment and comparison with biventricular pacing in patients with severe congestive heart failure. J Am Coll Cardiol 38:1971, 2001

88. Abraham WT, Fisher WG, Smith AL, et al: Cardiac resynchronization in chronic heart failure. Multicenter InSync Randomized Clinical Evaluation (MIRACLE) Study Group. N Engl J Med 346:1845, 2002

89. Bradley DJ, Bradley EA, Baughman KL, et al: Cardiac resynchronization and death from progressive heart failure: a meta-analysis of randomized controlled trials. JAMA 289:730, 2003

90. Pinski SL: Continuing progress in the treatment of severe congestive heart failure. JAMA 289:754, 2003

91. Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. The Multicenter Automatic Defibrillator Implantation Trial II Investigators. N Engl J Med 346:877, 2002

92. Felker GM, Thompson RE, Hare JM, et al: Underlying causes and long-term survival in patients with initially unexplained cardiomyopathy. N Engl J Med 342:1077, 2000

93. Senni M, Redfield MM: Heart failure with preserved systolic function: a different natural history? J Am Coll Cardiol 38:1277, 2001

94. Mancini DM, Eisen H, Kussmaul W, et al: Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation 83:778, 1991

95. Aaronson KD, Schwartz JS, Chen TM, et al: Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation 95:2660, 1997

96. Levy D, Kenchaiah S, Larson MG, et al: Long-term trends in the incidence of and survival with heart failure. N Engl J Med 347:1397, 2002

Acknowledgment

Figure 1 Alice Y. Chen.

III HYPERTENSION

Gary L. Schwartz, m.d.
Sheldon G. Sheps, m.d.

Hypertension is the most common chronic disorder in the United States, affecting 29% of the adult population.[1] It is the most common reason adults visit the doctor's office. In the year 2000, hypertension accounted for more than 1 million office visits to health care providers. The prevalence increases with age: for a normotensive middle-aged person in the United States, the lifetime risk of developing hypertension approaches 90%.[2] With the increasing age of the population in most developed and developing societies, it seems safe to assume that hypertension will become steadily more widespread in the coming years.

Hypertension is a major risk factor for stroke, myocardial infarction, heart failure, chronic kidney disease, progressive atherosclerosis, and dementia.[3-5] The treatment of hypertension is highly effective in reducing cardiovascular (CV) morbidity and mortality.[6,7] However, despite widespread public and professional education regarding the risks of hypertension and the benefits of treatment and despite the ready availability of effective therapies, only 58% of adults with hypertension are receiving treatment, and in only 31% is hypertension controlled.[1,8]

Improving control rates for hypertension depends on setting appropriate, patient-specific, evidence-based therapeutic goals, understanding and using available treatment options in an efficient and cost-effective manner, involving the patient in goal setting and the care process in an empathetic manner, and employing timely follow-up to monitor and adjust therapy as necessary. It is important to recognize that multiple classes of drugs are often needed for control; that patient education, communication, and involvement in the process is vital to long-term compliance; and that systematic approaches to therapy are required to achieve and maintain control over time.

Definition

Blood pressure (BP) is a quantitative trait that is continuously distributed in the population. Essential hypertension represents the upper end of the distribution of this trait and is defined by the BP level associated with a threshold value of increased CV risk. Any definition of hypertension is arbitrary because the risk of CV disease related to BP level increases steadily across the spectrum of BP values. Based on a meta-analysis of studies relating BP level to vascular mortality, optimum BP is defined as less than 115/75 mm Hg.[9] According to the Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7), normal BP (the level associated with minimal risk) for adults 18 years of age or older is a systolic BP of less than 120 mm Hg and a diastolic BP of less than 80 mm Hg [*see Table 1*].[8] Blood pressures ranging from 120 to 139 mm Hg systolic or 80 to 89 mm Hg diastolic are considered prehypertensive. Patients with BP in this range are at increased risk for the development of target organ injury and for progression to definite hypertension over time.[9,10] Therefore, these patients should have annual BP checks and be educated in strategies to lower BP and CV risk and to prevent the development of hypertension [*see* Prevention of Hypertension, *below*]. For patients with diabetes or renal disease, BP in the prehypertensive range poses a significantly higher risk than for healthy persons, and a lower threshold for intervention is indicated for these patients: above 130 mm Hg systolic or 80 mm Hg diastolic.

For the general population, hypertension is defined as a systolic BP of 140 mm Hg or higher or a diastolic BP of 90 mm Hg or higher. Hypertension is further divided into two stages, based on the highest level of either the systolic or diastolic BP. Prospective drug intervention trials have demonstrated the benefit of treatment for patients with a diastolic BP of 90 mm Hg or higher. Isolated systolic hypertension (ISH), which occurs mainly in persons older than 55 years, is defined as a systolic BP of 140 mm Hg or higher and a diastolic BP of less than 90 mm Hg. ISH is the most common hypertension subtype in older adults, who are the most rapidly growing segment of the population.[11] Epidemiologic data clearly demonstrate elevated and graded risk associated with systolic BP higher than 115 mm Hg.[9] However, those drug intervention trials that showed a benefit enrolled only subjects whose systolic BP was 160 mm Hg or higher[7]; the benefits of drug intervention for patients with ISH whose pretreatment systolic BP is below 160 mm Hg is inferred.

Epidemiology

Currently, it is estimated that 58 million adults in the United States have hypertension or are taking antihypertensive medications.[1] In addition to definitive hypertension, an additional 45 million adults in the United States have prehypertension.

In developed societies, BP increases with age. Diastolic BP plateaus in the fifth decade and may decline thereafter, but systolic BP continues to rise through the seventh decade. In persons younger than 50 years, diastolic BP level is the major predictor of CV risk, whereas systolic BP is the major predictor in those older than 60 years.[12] Individual risk is related to the level and duration of BP, as well as to the presence of other CV risk factors and of injury to so-called target organs—brain, heart, kidneys, peripheral arteries, and retina.[3]

The relationship between BP and CV morbidity and mortality begins in patients whose BP is higher than optimal levels (115/75 mm Hg) and is strong, continuous, graded, consistent, and independent. The relationship between BP and CV risk has largely been determined in middle-aged and older people, but above-normal BP in young adulthood is also related to increased long-term CV and all-cause mortality.[13] In young adult-

Table 1 Classification of Blood Pressure for Adults 18 Years of Age and Older[8]

Category	*Blood Pressure Level (mm Hg)*
Normal	Systolic < 120 and diastolic < 80
Prehypertension	Systolic 120–139 or diastolic 80–89
Hypertension	
Stage 1	Systolic 140–159 or diastolic 90–99
Stage 2	Systolic ≥ 160 or diastolic ≥ 100

Note: These categories apply to patients who are not taking antihypertensive drugs and are not acutely ill. When systolic and diastolic blood pressures fall into different categories, the higher category should be selected to classify the person's blood pressure status.

hood and early middle age, hypertension is more common in men than in women, but the opposite is the case in persons 60 years of age and older.[1] At all ages, hypertension is more common in African Americans than in whites; in all ethnic and racial groups, it is more common in the economically disadvantaged. At any given level of BP, CV risk is greater in men than women, in African Americans than whites or members of other racial or ethnic groups, in older persons than younger ones, and in patients with target organ disease and longer duration of hypertension.[3]

Etiology and Genetics

Essential hypertension develops as the consequence of a complex interplay over time between susceptibility genes and environmental factors. Numerous family and population studies suggest a significant role for genetic factors. Hypertension in persons younger than 55 years is four times more common in individuals with a family history of hypertension than in those with no family history of it. Estimates of the genetic contribution to BP variation range from 30% to 50%. However, the genetic contribution to essential hypertension is complex. Multiple genes are likely involved, and although the effects of some genes may affect BP independently, most genetic effects involve both gene-gene interaction (epistasis) and gene-environment interaction. Important interactions between the effects of specific genes and environments may occur at a particular time (perinatal life) or over the lifetime of an individual. Thus, sorting out the genetic contribution to essential hypertension is complex and challenging.

Some insight into the genetic contribution to essential hypertension has been gained from the identification of rare monogenic forms of hypertension.[14] Interestingly, most of these forms of hypertension arise from gene mutations that result in impairment of renal sodium excretion; renal impairment occurs either through the disruption of the renal sodium transport systems or through interference with mineralocorticoid receptor activity. Additional insight has come from studies that employ a candidate gene approach, in which genes are chosen on the basis of animal studies or previous knowledge of genes that encode proteins involved in BP regulatory pathways. Polymorphisms of candidate genes have been studied in association and linkage studies to assess their potential role in essential hypertension in humans. Most extensively studied have been genes encoding components of the renin-angiotensin-aldosterone system. Results of this line of investigation have implicated polymorphisms of the angiotensinogen *(AGT)* gene and the angiotensin-converting enzyme *(ACE)* gene in human essential hypertension.[15,16] The M235T variant of the *AGT* gene has been associated with higher circulating levels of AGT and is found more often in hypertensive than in normotensive persons.[16] An insertion/deletion polymorphism of the *ACE* gene has been associated with differences in ACE activity, with higher levels associated with the deletion allele.[17] The deletion allele has also been associated with several cardiovascular phenotypes, including higher BP levels and greater risk of target organ complications in hypertensive individuals.[15,18] In large samples, the observed associations have frequently been gender specific, suggesting interaction between the effects of the insertion/deletion polymorphism and gender.[15,19]

Variants of other genes have been implicated in essential hypertension. Adducin is a membrane skeleton protein consisting of α and β subunits that may influence ion transport across membranes. A variant of the α-adducin gene (Gly 460 Trp) has been associated with essential hypertension in case-control studies and may play a role in salt-sensitive hypertension.[20] Linkage studies have also found evidence indicating a role for variants of adrenergic and dopamine receptor genes.[21]

Experts expect that the number of gene polymorphisms associated with essential hypertension will continue to grow. However, because of the limitations of candidate gene studies, as well as linkage and association studies, there is a need for new approaches that can assess the effects of multiple genes and environments. These types of studies will be difficult to perform but will be necessary if we are to someday be able to determine the specific genotypes and environments present in the majority of persons destined to develop hypertension.[22]

Regarding the role of genetics in hypertension, it is important to note that hypertension is virtually nonexistent in primitive peoples who follow a preagricultural hunter-gatherer lifestyle. This lifestyle involves significant daily physical activity and a diet rich in potassium and low in fat and sodium. Obesity is uncommon. Dietary patterns involve periods of feasting interspersed with long periods with minimal food. Given that the human gene pool has changed little over the past 30,000 years, some suggest that hypertension is the consequence of a human genome selected for a hunter-gatherer lifestyle but now interacting with a modern one. In contrast to primitive societies, modern societies are characterized by a low level of physical activity and constant availability of abundant food that is rich in sodium and fat and low in potassium; the result is an increase in body weight with aging and a high incidence of obesity. Genetic adaptation to the hunter-gatherer lifestyle provided survival advantages in that environment but may now be contributing to many modern diseases such as obesity and hypertension.

Pathophysiology and Pathogenesis

Simplistically, BP is the product of cardiac output and peripheral vascular resistance (BP = cardiac output × peripheral vascular resistance). Thus, variation in extracellular fluid volume, the contractile state of the heart, and vascular tone determine variation in BP level. The hemodynamic hallmark of established essential hypertension is elevated peripheral vascular resistance. An increase in cardiac output is occasionally noted early but is not a persistent finding. Hypertension can be viewed as the final outcome of a complex interaction between genetic and environmental factors that act on intermediate physiologic systems involved in BP regulation (i.e., those that influence fluid volume, heart contractility, and vascular tone).

A central hypothesis for the pathogenesis of essential hypertension involves an interaction between the high dietary sodium intake typical of industrial society and defects in renal sodium excretion. Evidence of a role for dietary sodium comes from animal studies and from epidemiologic and experimental studies in humans.[23,24] Guyton hypothesized that hypertension develops when the kidneys require a higher BP to maintain extracellular volume within normal limits.[25] This would occur in persons with impaired renal sodium excretion. Studies support the possibility of an inherited defect in renal sodium excretion as the basis of human essential hypertension. Most monogenic forms of hypertension discovered so far involve mutations that impair renal sodium excretion by increasing mineralocorticoid activity or by influencing tubular sodium transport systems.[14]

Moreover, renal sodium excretion can be influenced by variation in activity of both the renin-angiotensin-aldosterone system and the sympathetic nervous system. Angiotensin II enhances renal tubular sodium reabsorption directly and indirectly through stimulation of aldosterone release and the sympathetic nervous system. Additional mechanisms that may explain defective renal sodium excretion include an inherited reduction in the number of nephrons, as well as the presence of a subpopulation of so-called ischemic nephrons, which occur as a result of increased afferent renal artery tone and that lead to increased renin activity.[26,27] Extracellular volume expansion could lead to chronic increases in vascular resistance through mechanisms of organ autoregulation of blood flow (i.e., variation in the tone of vessels that occurs so as to regulate organ blood flow to meet metabolic needs). Some studies have suggested that volume expansion stimulates the release of a sodium-potassium-adenosine triphosphatase (Na^+,K^+-ATPase) inhibitor (i.e., an ouabain-like substance) that facilitates renal sodium excretion but increases vascular tone by interfering with sodium-calcium exchange in vascular smooth muscle cells.[28]

Other evidence suggests that increased sympathetic nervous system activity has a role in causing hypertension in some persons.[29] These cases could be the result of a genetic tendency toward increased sympathetic activity interacting with repetitive psychogenic stress, obesity, or high sodium intake. Hypertension could also arise or be sustained by defects in baroreceptor function.[30]

Weight gain and obesity (especially abdominal fat accumulation) are associated with an increased risk of hypertension. A number of humoral factors may be responsible, including increased activity of the sympathetic nervous system and the renin-angiotensin system.[31] In addition, obesity is associated with insulin resistance and hyperinsulinemia. Hyperinsulinemia may directly stimulate sympathetic activity, in addition to promoting vascular hypertrophy (increased vascular tone) and renal sodium retention.[32] In addition, leptin levels are increased in obese individuals. Leptin may also increase BP by stimulation of the sympathetic nervous system.[33]

More general abnormalities of cell membranes or multiple ion transport systems acting across cell membranes could contribute to the development of hypertension. In addition to impairing sodium excretion in the kidneys, these defects could act in a variety of ways to influence vascular structure and tone.[34] Vascular tone could also be influenced by variation in vascular endothelial function through an imbalance in the production of substances that cause vasodilation (e.g., nitric oxide) and those that cause vasoconstriction (e.g., endothelin).[35]

Table 2 Proper Blood Pressure Measurement Technique

Patient should refrain from smoking or caffeine ingestion for 30 min before measurement
Patient should be at rest, seated in a chair with back and feet supported, for at least 5 min before measurement is taken
Patient should not speak while blood pressure is being measured
Patient's arm should be bare, with no tight clothing constricting the upper arm
Select a proper cuff size for the arm: bladder should encircle at least 80% of arm (many adults will require a large cuff)
Position patient's arm so cuff is at the level of the heart
Place stethoscope bell over brachial artery
Inflate cuff to occlude the pulse
Deflate at rate of 2–3 mm/sec
Measure systolic (first sound) and diastolic (last sound) to nearest 2 mm Hg
Repeat measurement after 2 min
Under special circumstances, measure blood pressure with patient in standing position

Diagnosis

The diagnosis of hypertension relies on multiple office measurements of BP performed in a rigorous manner with a validated and well-maintained mercury or aneroid sphygmomanometer and a cuff of appropriate size. Several expert groups have published guidelines for proper BP measurement; unfortunately, these guidelines are rarely complied with in most clinics [*see Table 2*].[36] The diagnosis of hypertension requires findings of an elevated average BP on at least two office visits, with at least two standardized measures of BP made at each visit. For most patients, confirmation can occur over a 1- to 2-month period. If an initial BP is severely elevated, confirmation should occur over a shorter period of time. Self-measurements of BP outside the office setting can be used to distinguish sustained hypertension from isolated clinic hypertension; self-measurement has the further advantages of involving patients in the process of care (which often improves compliance) and aiding in the assessment of response to therapy. On average, home readings are lower than office readings; therefore, values above 135 mm Hg systolic or 85 mm Hg diastolic are considered elevated.[8] BP devices for home use (aneroid or oscillometric) need to be validated twice yearly by the health care provider, and patients need to be educated in the technique of proper BP measurement [*see Table 2*].

AMBULATORY BP MONITORING

Cross-sectional studies show that BP averages from ambulatory BP monitoring (ABPM) correlate better with the presence of target organ injury (especially left ventricular hypertrophy [LVH]) than office BP measurements.[37] Also, prospective studies and population-based observational studies have shown that average BP derived from ABMP predicts additional risk for CV events after adjustment for clinic or office BP.[38] This is true for both untreated as well as treated patients.[39] ABPM is the best method to establish the presence of isolated clinic hypertension (so-called white-coat hypertension), which is defined as an elevation in BP that occurs only in the clinic setting, with normal BP in all other settings, in the absence of evidence of target organ injury.[8] Screening for white-coat hypertension is currently a reimbursable indication for ABPM by Medicare.[40] The possibility of a white-coat effect should be considered in selected patients with resistant hypertension, in elderly patients with significant office systolic hypertension, and in some pregnant women. Other uses for ABPM include assessment of hypotensive symptoms, episodic hypertension, and suspected autonomic dysfunction in patients with postural hypotension.[8] ABPM is also useful in the evaluation of the occasional patient with hypertensive target organ injury (LVH, stroke) whose office BP is normal. In addition, it is now recognized that some patients have so-called white-coat normotension, or masked hypertension; for these patients, BP is normal in the office but is elevated outside the office setting.[41] This important group is often missed in routine practice.

Table 3 Classic Features of Essential Hypertension

- Onset of hypertension in the fourth or fifth decade of life
- Family history of hypertension
- BP < 180/< 110 mm Hg at diagnosis
- Asymptomatic
- History, physical examination, and routine laboratory studies are normal (no target-organ damage at time of diagnosis)
- BP control achieved with lifestyle changes and one or two drugs
- BP control is maintained once achieved

INITIAL EVALUATION

The initial evaluation of patients with elevated BP has four major objectives: (1) to identify lifestyle factors contributing to elevated BP and higher CV disease risk, (2) to identify associated modifiable CV risk factors, (3) to assess for target organ injury or clinical CV disease, and (4) to identify any secondary causes of hypertension.[8] The second and third objectives are important for risk stratification, which defines the BP threshold for initiation of drug therapy and establishes the BP goal to achieve with therapy.[42,43]

The overall frequency of secondary hypertension is 5% to 10% in primary care practices. The classic picture of essential hypertension should be compared to the individual patient's presentation [*see Table 3*]. Secondary hypertension should be suspected on finding features that are not consistent with essential hypertension. Such features include age at onset younger than 30 or older than 50 years; BP higher than 180/110 mm Hg at diagnosis; significant target organ injury at diagnosis; hemorrhages and exudates on fundus examination; renal insufficiency; LVH; poor response to appropriate three-drug therapy; and accelerated or malignant hypertension. Specific features that suggest secondary causes of hypertension vary with the individual condition [*see Table 4*].

HISTORY

The clinician should inquire about a family history of hypertension, premature CV disease, and disorders that would increase the possibility of secondary hypertension (e.g., polycystic kidney disease or other renal disease, medullary cancer of the thyroid, hyperparathyroidism, or pheochromocytoma). The patient should be questioned about lifestyle habits that influence BP (e.g., level of physical activity, sodium intake, use of caffeine and alcohol, history of weight gain), and CV risk (tobacco use); in addition, the patient should be asked about symptoms suggesting target organ disease (angina, symptoms of heart failure, transient cerebral ischemia, or renal disease) or secondary hypertension (spells suggesting pheochromocytoma). A known history of dyslipidemia, diabetes, or cerebrovascular, heart, or renal disease also should be documented. In addition, a thorough medication review (including prescription and over-the-counter drugs, herbs and herbal compounds, and street drugs) is important to identify drugs that can raise BP or interfere with the antihypertensive effect of planned drug therapy [*see Table 5*].[44] In patients with a history of hypertension, the duration of hypertension, the previous BP levels, and the specific drugs used for treatment, together with the efficacy of those drugs and the reasons for discontinuation, should be ascertained. Other comorbid conditions and their treatments need to be documented, because they may influence antihypertensive drug selection.

PHYSICAL EXAMINATION

The examination should include at least two standardized measurements of BP with the subject in the seated position. Initially, BP should also be measured in the opposite arm (to identify arterial narrowing, which can cause an inaccurately low reading in one arm) and in the standing position, especially in diabetic patients and older patients (to identify orthostatic declines). Height and weight should be determined, to permit calculation of body mass index, and waist circumference (a potential CV risk factor) should be recorded.

The physical examination is directed toward identifying target organ injury or features suggesting secondary hypertension [*see Table 4*]. Retinal examination should be performed, primarily to identify retinal changes of diabetes or severe hypertension (i.e., hemorrhages, exudates, papilledema). Arteriolar narrowing, focal constrictions, and arteriovenous nicking on retinal examination are more closely associated with atherosclerosis and are of limited value for predicting the severity of hypertension or assessing overall CV risk.[43,45]

Table 4 Features Suggesting Specific Causes of Secondary Hypertension

Condition	*Features*
Primary aldosteronism	Unprovoked hypokalemia
Pheochromocytoma	Labile BP with episodic headache, sweats, tachycardia, pallor, abdominal pain, weight loss Neurofibromas, café-au-lait spots Orofacial neuromas (multiple endocrine neoplasia type II) Retinal angiomas (von Hippel–Lindau syndrome)
Renovascular hypertension	Abdominal or flank bruits Peripheral bruits/diminished pulses from atherosclerosis Elevated serum creatinine level, hypokalemia Flash pulmonary edema Hypertension in a patient younger than 30 yr (fibromuscular dysplasia) Sudden onset or worsening of systolic-diastolic hypertension after age 50 yr (atheromatous disease) Accelerated-malignant hypertension Treatment-resistant hypertension Unexplained subacute decline in renal function
Cushing syndrome	Truncal obesity, proximal muscle weakness and atrophy Stria, acne, thin skin, bruises, hyperpigmentation Elevated plasma glucose, hypokalemia
Coarctation of the aorta	Headaches Cold feet, claudication Delay of femoral pulse compared to radial pulse Weak or absent femoral pulses High BP in arms/low BP in legs Murmurs front/back chest
Polycystic renal disease	Abdominal/flank masses, family history of renal disease

Table 5 Drugs That Can Increase Blood Pressure or Interfere with Antihypertensive Drug Efficacy

Drug	*Mechanism*
Oral contraceptives	Sodium retention, increase level of angiotensinogen, facilitate action of catecholamines
Alcohol (moderate or heavy intake)	Activate sympathetic nervous system, increase cortisol secretion and intracellular calcium levels
Sympathomimetics and amphetamine-like substances (over-the-counter cold or allergy formulas, diet pills)	Increase peripheral vascular resistance
Nonsteroidal anti-inflammatory drugs	Sodium retention, renal vasoconstriction; interfere with efficacy of all antihypertensive drugs, especially diuretics, beta blockers, angiotensin-converting enzyme inhibitors, angiotensin receptor blockers
Corticosteroids	Iatrogenic Cushing disease
Tricyclic antidepressants	Inhibit action of centrally acting sympatholytics (clonidine, guanfacine)
Serotoninergics (antidepressants)	Systemic vasoconstriction (increase peripheral vascular resistance)
Cyclosporine	Renal and systemic vasoconstriction (sodium retention, increased peripheral vascular resistance)
Erythropoietin	Systemic vasoconstriction (increased peripheral vascular resistance)
Monoamine oxidase inhibitors + tyramine-containing foods (aged cheeses, red wine)	Prevents degradation of norepinephrine released by tyramine-containing foods; increase in BP with reserpine
Cocaine	Systemic vasoconstriction (increased peripheral vascular resistance)
Marijuana	Increases systolic blood pressure
Glycyrrhizinic acid (chewing tobacco, imported licorice, health food products)	Inhibits renal metabolism of cortisol to cortisone (increased mineralocorticoid activity causing sodium retention, loss of potassium)
Grapefruit products	Inhibit cytochrome P-450 metabolism of some drugs
Herbs	
Bloodroot	CNS stimulant
Blue cohosh	Action of methylcytisine
Broad bean	Unknown
Scotch broom	Sympathomimetic
Cola nut	Sympathomimetic
Ephedra	Sympathomimetic, CNS stimulant
Foxglove	Cardiac inotrope
Gentian	Unknown
Ginseng	CNS stimulant, glucocorticoid effect
Goldenseal	Systemic vasoconstriction
Grindelia	CNS stimulant
Jimson weed	Anticholinergic effect
Juniper	Aquaretic
Kava	Unknown
Yohimbe	Central alpha blocker

Table 6 Laboratory Evaluation of Newly Diagnosed Hypertensive Patients

Purpose	*Tests*
Identify cardiovascular risk factors	Cholesterol (total, HDL), triglycerides, fasting blood glucose
Identify target-organ injury	Chest x-ray, ECG, urinalysis, serum creatinine or BUN, uric acid
Screen for secondary hypertension	Serum creatinine, potassium, calcium; urinalysis
Calculate kidney function	Cockcroft and Gault equation: GFR = (140 – age in yr) × (weight in kg) × 0.85 (if patient is female)/72 × S_{Cr} Modification of Diet in Renal Disease (MDRD) equation: GFR = 170 × $(S_{Cr})^{-0.999}$ × (age in yr)$^{-0.176}$ × 0.762 (if patient is female) × 1.18 (if patient is black) × (BUN)$^{-0.17}$ × (alb)$^{0.318}$

alb—serum albumin concentration (g/dl) BUN—blood urea nitrogen (mg/dl)
ECG—electrocardiogram GFR—glomerular filtration rate (ml/min)
HDL—high-density lipoprotein S_{Cr}—serum creatinine (mg/dl)

LABORATORY TESTS

Laboratory studies are performed to support the general goals of the initial evaluation [*see Table 6*]. In addition, they provide baseline information for monitoring in patients who are subsequently treated with antihypertensive drugs that can influence laboratory values (i.e., diuretics, beta blockers, ACE inhibitors, and angiotensin receptor blockers [ARBs]). Additional studies are not advised unless the history, physical examination, or initial laboratory studies are inconsistent with essential hypertension or suggest a specific secondary etiology.

If the initial assessment suggests renal dysfunction, the patient should be evaluated for chronic kidney disease by measuring 24-hour urinary protein excretion and estimating glomerular filtration rate (GFR). Equations are available to estimate GFR [*see Table 6*].[46,47] The Modification of Diet in Renal Disease (MDRD) equation requires measurement of blood urea nitrogen and serum albumin concentrations in addition to serum creatinine concentration. The estimate of GFR can also be calculated with an online tool (available at www.hdcn.com/calcf/gfr.htm).

RISK STRATIFICATION

At any given level of BP, specific factors in an individual patient may result in deviations above or below the average CV risk observed in population studies. These factors are used to determine the BP threshold and timing of drug therapy and the BP goal for the individual patient. Individual specific factors that determine risk include the presence of other CV risk factors and the presence of injury to the target organs of hypertension or clinical CV disease.[3] A simple and clinically useful scheme modified from the JNC VI report separates patients into three levels of risk [*see Table 7*].[42] This scheme suggests aggressive treatment and lower BP goals for patients at the highest level of risk and more conservative treatment and BP goals for patients at the lowest level of risk. For example, in a patient with diabetes, drug therapy is indicated initially (along with lifestyle changes) when BP exceeds 130/80 mm Hg. In contrast, in a young patient who has no other CV risk factors or evidence of target organ injury or CV disease, a 6- to 12-month trial of lifestyle changes rather than drugs is indicated as initial therapy

Table 7 Risk Stratification and Treatment in Hypertensive Patients[8,42,43]

Blood Pressure Stage (mm Hg)	*Risk Group A (no risk factors, no TOD/CCD*)*	*Risk Group B (≥ 1 risk factor, not including diabetes; no TOD/CCD)*	*Risk Group C (TOD/CCD and/or diabetes ± other risk factors)*
Prehypertension 120–139/80–89)	Lifestyle modification	Lifestyle modification	Lifestyle modification, drug therapy‡
Stage 1 (140–159/90–99)	Lifestyle modification (up to 12 mo)	Lifestyle modification (up to 6 mo)†	Lifestyle modification, drug therapy
Stage 2 (≥ 160/≥ 100)	Lifestyle modification, drug therapy	Lifestyle modification, drug therapy	Lifestyle modification, drug therapy

*Risk factors are cigarette smoking, dyslipidemia, diabetes, age > 55 yr in men and > 65 yr in women, male sex, postmenopausal status in women, family history of premature cardiovascular disease (women < 65 yr, men < 55 yr), nephropathy (microalbuminuria or glomerular filtration rate < 60 ml/min), obesity (body mass index ≥ 30 kg/m^2; waist circumference ≥ 102 cm in men and ≥ 88 cm in women), C-reactive protein level ≥ 1 mg/dl, physical inactivity. TOD/CCD includes left ventricular hypertrophy, angina, prior myocardial infarction, heart failure, previous coronary revascularization procedure, stroke, transient ischemic attack, nephropathy, peripheral arterial disease, retinopathy.
†For patients with multiple risk factors, consider drugs initially in addition to lifestyle modifications.
‡Use drugs if BP > 130 mm Hg systolic or > 80 mm Hg diastolic and patient has heart failure, chronic kidney disease, or diabetes.
TOD/CCD—target-organ disease/clinical cardiovascular disease

unless BP is of stage 2 (≥ 160 mm Hg systolic or ≥ 100 mm Hg diastolic). In these low-risk cases, the goal BP is less than 140/90 mm Hg. Guidelines from Europe provide an even more detailed approach to risk stratification.[43]

Prevention of Hypertension

In many cases, the assessment will show BP in the prehypertensive range (i.e., 120 to 139/80 to 89 mm Hg); in the United States, 22% of adults, or approximately 45 million persons, fit this category. Preventive care is indicated in these patients.

Multiple studies support the effectiveness of environmental manipulation in preventing or delaying the onset of hypertension.[48-51] Prevention of hypertension is important, given that treatment of established hypertension is only partly effective in reducing the associated morbidity and mortality.[52] Furthermore, the relationship between BP level and CV morbidity and mortality is continuous and extends into nonhypertensive levels; approximately one third of the coronary artery disease deaths attributable to BP occur in persons whose BP is in the prehypertensive range. Prevention strategies that lower BP in prehypertensive patients extend the benefits of BP reduction to this large group.

The risk of developing hypertension is increased in African Americans and in all persons with prehypertension or a family history of hypertension. Reversible patient characteristics associated with an increased risk of developing hypertension include being overweight or obese; having a sedentary lifestyle; ingesting a high-sodium, low-potassium diet; using excessive amounts of alcohol; and manifesting the so-called metabolic syndrome. The metabolic syndrome is defined as three or more of the following conditions: abdominal obesity (waist circumference > 40 inches in men or > 35 inches in women), glucose intolerance (fasting blood glucose ≥ 110 mg/dl), BP of 130/85 mm Hg or higher, elevated triglycerides (≥ 150 mg/dl) or low high-density lipoprotein (HDL) cholesterol (< 40 mg/dl in men or < 50 mg/dl in women).[53] Clinical trials support the efficacy of six interventions in such people for the primary prevention of hypertension [*see Table 8*].[48,49,51] Combining interventions is beneficial.[54,55] For patients with the metabolic syndrome, in addition to intensive lifestyle modifications, drug therapy is recommended for management of each of its components when appropriate.

Treatment

The overall goal of treatment in hypertensive patients is to reduce the risk of CV morbidity and mortality by lowering BP and treating other modifiable risk factors. In general, the goal is to lower BP to below 140/90 mm Hg. In patients with heart failure, diabetes, or renal disease, the goal is to lower BP to below 130/80 mm Hg. In older patients with ISH, the goal is to lower systolic BP to below 140 mm Hg.

These goals are achieved through lifestyle modification and, in most cases, drug therapy. In addition, comorbid conditions such as dyslipidemia or diabetes should be addressed.[53] Low-dose aspirin should be considered once BP is controlled.[56] Self-measurement of BP should be encouraged.

LIFESTYLE FACTORS

Observational studies have identified several environmental factors associated with hypertension, and prospective studies have demonstrated BP lowering with manipulation of these factors [*see Table 8*].[48,49,51,54,55,57-60] In addition to lowering BP, lifestyle recommendations are designed to reduce overall CV risk. These measures should be advised for all patients with BP above the normal level. Tobacco use should be discouraged because, in addition to being a powerful CV risk factor, each cigarette smoked elevates BP for 15 to 30 minutes, and multiple cigarettes can raise BP for most of the day. A new device that facilitates deep-breathing exercises (RESPeRATE) has been shown to low-

Table 8 Lifestyle Modifications for Hypertension Prevention and Management

- Lose weight if overweight
- Reduce sodium intake to ≤ 100 mmol/day (2.4 g sodium, 6 g salt)
- Increase aerobic exercise (30–45 min/day)
- Limit alcohol intake to no more than 1 oz (30 ml; e.g., 24 oz of beer, 10 oz of wine, 2 oz of 100-proof whiskey) or to 0.5 oz for women and lighter-weight people
- Maintain adequate intake of potassium (90 mmol/day)
- Ingest a diet rich in fruits and vegetables and low-fat dairy products but reduced in saturated and total fat (e.g., Dietary Approaches to Stop Hypertension [DASH] diet)
- Discontinue tobacco use

er BP and can be considered as an adjunct to lifestyle and drug treatments.[61]

PHARMACOLOGIC TREATMENT

The JNC 7 report recommends thiazide diuretics as initial drugs of choice for most patients; this recommendation is based on the totality of data from randomized trials, including the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT).[8,62-64] Critics of diuretics have cited evidence suggesting that diuretic-based treatment does not provide protection from coronary artery disease events to the degree predicted from epidemiologic studies. The ALLHAT was designed to determine whether treatment with a diuretic would be inferior to treatment with an alpha blocker, a calcium antagonist, or an ACE inhibitor in preventing fatal and nonfatal coronary artery disease events in a high-risk group of adults with essential hypertension. The study showed no difference among the drugs for the outcome of fatal and nonfatal coronary artery disease or total mortality. Moreover, diuretic treatment was superior to alpha blocker, calcium antagonist, or ACE inhibitor treatment with some CV disease outcomes. The alpha-blocker arm of the trial was terminated early because of an almost twofold increase in the risk of heart failure compared with the diuretic group. On the basis of these results, alpha blockers are no longer considered an appropriate initial therapy for hypertension. Compared with the diuretic group, the calcium antagonist group also had a higher risk of heart failure. Compared with the diuretic group, the ACE inhibitor group had an increased risk of stroke and combined CV disease, but much of the increased risk occurred in blacks, in whom BP control with the ACE inhibitor was inferior to the control achieved with the diuretic.

Alternative medications should be considered if diuretics are contraindicated or are poorly tolerated or there is a compelling indication for a drug from a different class. Alternative drug choices are beta blockers, ACE inhibitors, ARBs, and calcium antagonists.

A subsequent study contradicted the results of ALLHAT and suggested that ACE inhibitors are superior to diuretics in older men.[65] In truth, differences in outcomes by drug choice likely reflect differences in achieved BP rather than unique effects of specific agents.[66] Therefore, achieving the BP goal is more important than the specific agents used to achieve it.

Randomized clinical trials suggest that the presence of certain comorbid conditions constitutes a so-called compelling indication for selection of specific drugs [*see Table 9*]. Other considerations that should influence drug selection include concomitant conditions for which some agents may be beneficial and others contraindicated [*see Tables 9 and 10*], potential drug-drug interactions, concerns about quality of life, cost (generic formulations are available for diuretics, beta blockers, calcium antagonists, and ACE inhibitors), and, finally, demographics (in general, older patients and blacks respond better to diuretics and calcium antagonists, whereas younger patients and whites respond better to beta blockers, ACE inhibitors, and ARBs). In general, the drug chosen should have a long half-life (once-daily dosing is preferable). It should be continued only if the patient tolerates it and is comfortable with its cost, because these are important factors in long-term compliance. To achieve currently recommended goal BP levels, many patients will require more than one drug; this possibility should be discussed at the outset with the patient. Regardless of the agent chosen, BP should be reassessed after 2 to 4 weeks of treatment [*see Figure I*].

Combination Therapy

The JNC 7 report suggests initiation of therapy with two drugs (combination therapy) rather than a single agent if BP is more than 20 mm Hg systolic or 10 mm Hg diastolic above the treatment goal.[8] Generally, a two-drug regimen should include a diuretic appropriate for the level of renal function. An increasing number of antihypertensive combination products are available in a number of dosing options.[8] Although combination products may be more convenient, it is often less expensive to use individual agents, because generic drugs are frequently available. In addition, titration of doses of the two agents may be easier when the two drugs are prescribed separately. Once BP control is achieved with given doses of two agents, switching to the same therapy in combination form can be considered.

The advantages and disadvantages of using combination products have been reviewed.[67] Caution is advised when using combination therapy in older persons and diabetic patients, because of the increased risk of precipitous declines in BP or aggravation of orthostatic hypotension.

Improving Control Rates

In general, significant progress has been made in lowering BP in patients with hypertension. Although the proportion of patients with BP lower than 160/95 mm Hg has increased significantly since the 1970s, the percentage of patients with controlled hypertension (defined as systolic BP maintained below 140 mm Hg and diastolic BP, below 90 mm Hg) remains low. It is estimated that control of hypertension was accomplished in 31% of patients for the period from 1999 to 2000.[1] This is well below the Healthy People 2010 goal of at least 50% of patients achieving control. It is commonly believed that the major factors responsible for lower control rates are lack of access to health care and patient noncompliance and believed that the population of patients with uncontrolled hypertension comprises disproportionately large numbers of ethnic and racial minorities. However, studies suggest that other factors are also important. Analyses of the Third National Health and Nutrition Examination Survey (NHANES III) identified factors associated with the likelihood both of attaining control of hypertension and of failing to attain control.[68] Factors associated with an increased likelihood of controlling hypertension included being married (greater social support), having private health insurance, visiting the same health care facility or having the same provider over time, having had BP measured within the previous 6 to 11 months, and using lifestyle modifications in the treatment program. On the other hand, factors associated with an increased likelihood of uncontrolled hypertension included being 65 years or age or older, being male, being black, and failing to see a physician in the preceding year. Interestingly, not having health insurance or not having a source of health care was not predictive of uncontrolled hypertension.

Most cases of uncontrolled hypertension occur in older persons and represent mild ISH (systolic BP, 140 to 160 mm Hg).[69] In a study of self-reported treatment practices among primary care physicians, 43% of physicians would neither start drug therapy for a patient whose systolic BP is between 140 and 160 mm Hg nor intensify treatment for a patient whose systolic BP is 158 mm Hg.[70] In this same study, 41% of the care givers were unfamiliar with national hypertension guidelines. In a further analysis, familiarity with the guidelines lowered the provider's BP treatment threshold. Other studies of physician practices

Table 9 Patient Condition and Choice of Antihypertensive Drugs

Conditions	*Drug Choice*
No comorbid conditions	Diuretics
Isolated systolic hypertension (elderly patients)	Diuretics (preferred), calcium antagonists (DHP)*
Angina	Beta blockers,* calcium antagonists (non–short-acting DHP)
Angina (with diabetes or LV dysfunction)	ACE inhibitors† (in addition to beta blockers and calcium antagonists)
Atrial fibrillation	Beta blockers,* calcium antagonists (rate limiting)*†
Cough with ACE inhibitors	ARBs*
Diabetes mellitus type 1 with proteinuria	ACE inhibitors*; calcium antagonists (non-DHP); diuretics, beta blockers†
Diabetes mellitus type 2 with proteinuria	ARBs*†; calcium antagonists (non-DHP)†; diuretics, beta blockers†
High risk of type 2 diabetes	ACE inhibitors†
Essential tremor	Beta blockers (noncardioselective)†
Heart failure, LV dysfunction	ACE inhibitors, beta blockers, diuretics, aldosterone antagonists*; ARBs†; generally, an ACE inhibitor is first choice, ± a beta blocker in asymptomatic patients; diuretic used to treat congestion; aldosterone antagonist used only in advanced disease in combination with other agents; ARB should not be used in patients on an ACE inhibitor and beta blocker [*see 1:II Heart Failure*]
High risk of cardiovascular disease or type 2 diabetes	ACE inhibitor†
Hyperlipidemia	Alpha blockers (not considered first-line therapy)†
Intolerance to other antihypertensive drugs	ARBs†
Left ventricular hypertrophy (by ECG)	ARBs†
Migraine	Beta blockers (noncardioselective), calcium antagonists (non-DHP)†
Myocardial infarction	Beta blocker (non-ISA) most often drug of choice, with ACE inhibitor added if LV function impaired*; aldosterone antagonist can be added to standard therapy in patients with LV dysfunction*; diltiazem (non–Q wave infarction)†; verapamil†
Osteoporosis	Thiazide diuretics†
Peripheral vascular disease	Calcium antagonists†
Preoperative hypertension if at increased cardiovascular risk	Beta blockers†
Previous stroke	Diuretic + ACE inhibitor*; ACE inhibitor as monotherapy had no effect on BP or outcome; benefit noted only with combination that lowered BP
Prostatism	Alpha blockers (not considered first-line therapy)†
Renal insufficiency with proteinuria from any cause	ACE inhibitors, ARBs, calcium antagonists (non-DHP)†

*Compelling indication.
†Specific indication.
ACE—angiotensin-converting enzyme ARB—angiotension II receptor blocker DHP—dihydropyridine ISA—intrinsic sympathomimetic activity LV—left ventricle

have shown similar results. Emerging from these studies is the realization that a major factor in continued poor control rates for hypertension is a tolerance by the health care provider of elevated systolic BP, especially in older patients. On the basis of these study results, health care providers should consider steps to improve control rates in their practice [*see Table 11*].

REFRACTORY/RESISTANT HYPERTENSION

Studies conducted to determine what causes resistant hypertension have used different definitions of the term. In most studies, hypertension was considered resistant or refractory if control was not achieved with a combination of lifestyle modifications and the rational use of full therapeutic doses of two or three antihypertensive medications, one of which was a diuretic appropriate for the level of renal function. Studies suggest five issues to consider when evaluating patients with resistant hypertension[42]: noncompliance with therapy, interfering substances, an inappropriate drug regimen, office hypertension or pseudohypertension, and secondary hypertension. In most cases, causative factors will be identified if these five issues are given careful attention.

Noncompliance

Lack of BP control often results from noncompliance with the drug regimen or diet. Common reasons for noncompliance with drug therapy include drug costs, side effects, complex dosing schedules, and inadequate follow-up. Patients are reluctant to admit noncompliance with drug treatment, so a high degree of vigilance is required. Asking an open-ended question

Table 10 Contraindications to Antihypertensive Drugs

Class of Drug	*Possible Contraindications*	*Compelling Contraindications*
Diuretics	Dyslipidemia (high doses), allergy to sulfa-based antibiotics, patient is sexually active man, diabetes mellitus (high doses)	Gout, allergy to sulfa-based diuretics
Beta blockers	Bronchospastic disease (asthma, COPD, noncardioselective agents), dyslipidemia (non-ISA agents), severe peripheral vascular disease, athletes	Bronchospastic disease (noncardioselective agents) second- or third-degree heart block
ACE inhibitors, ARBs	Renovascular disease (bilateral renal artery stenosis), renal insufficiency	Pregnancy, hyperkalemia
Calcium antagonists	—	Second- or third-degree heart block (non-DHP agents); heart failure (except amlodipine, felodipine)
Alpha blockers	Postural hypotension	Urinary incontinence
Reserpine	Peptic ulcer, nasal allergy	Depression
Methlydopa	Liver disease	—
Labetalol	Liver disease	—
Central alpha agonists	Depression, sleep disorders	—

ACE—angiotensin-converting enzyme ARBs—angiotensin receptor blockers COPD—chronic obstructive pulmonary disease DHP—dihydropyridine ISA—intrinsic sympathomimetic activity

such as, "Many people have problems remembering their drug schedule; do you?" is occasionally effective. Clues to noncompliance include failure to keep follow-up appointments or renew prescriptions, or complaints about the cost of drugs or side effects. Certain drugs are expected to cause findings on the physical examination or laboratory evaluation. An absence of these findings may indicate noncompliance. Examples are slowing of the heart rate with beta blockers, electrolyte changes with diuretics, or dry mouth with clonidine. Noncompliance with a low-salt diet can also be important. A high-salt diet can interfere with the effectiveness of almost all of the currently used antihypertensive drugs.

Interfering Substances

Certain prescription drugs, over-the-counter medications, herbals, and street drugs can raise BP or interfere with the BP-lowering effect of antihypertensive drugs [*see Table 5*]. Taking a complete medication history and asking patients to bring in all their medication bottles is essential for identifying interfering substances. Alcohol abuse should also be considered, because in addition to its physiologic effects, alcohol abuse is often associated with poor compliance and lack of BP control.

Inappropriate Drug Regimens

The drug regimen should be carefully reviewed. Full therapeutic doses of drugs should be employed. In general, it is preferable to use drugs that have complementary actions and that work by interfering with different BP regulatory pathways. In compliant patients, inadequate control of extracellular volume is the most common cause of resistant hypertension.[71] Extracellular volume expansion tends to occur as BP is lowered and is a secondary effect of some drugs (e.g., centrally acting sympatholytics in modest doses and some direct vasodilators). In patients with renal dysfunction, impaired renal excretion of sodium often is an important factor in raising BP. Thiazide diuretics are often ineffective when serum creatinine is higher than 2.0 mg/dl or creatinine clearance is less than 30 ml/min. In

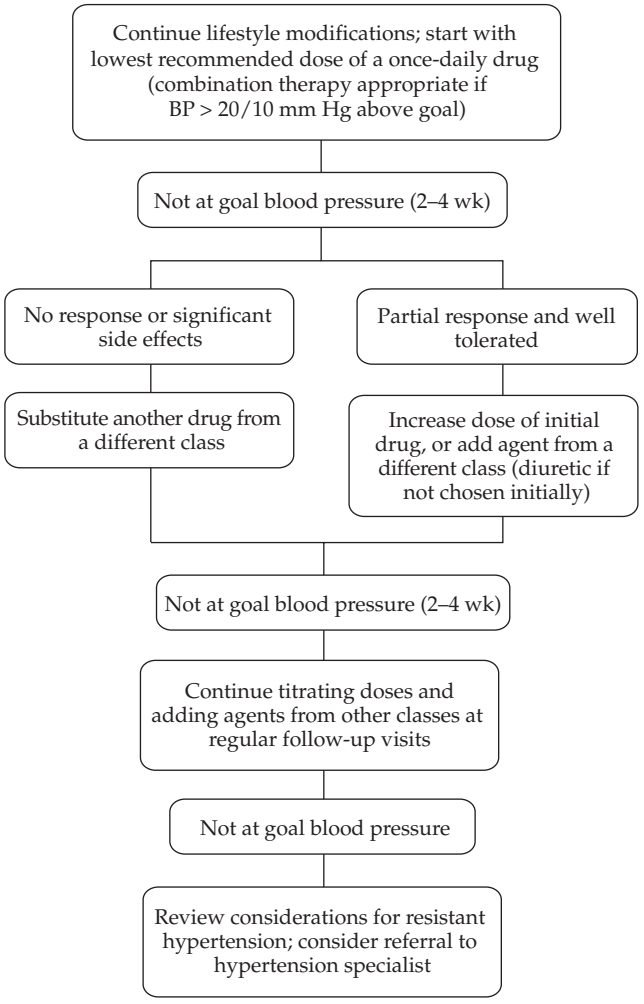


Figure 1 **Overview of drug treatment for hypertension.**

Table 11 Considerations for Improving Blood Pressure Control Rates

Become familiar with national guidelines (set a BP goal with the patient)
Schedule regular follow-up visits
Recommend self-monitoring of BP (involve patients in the treatment process)
Measure BP at every follow-up office visit and articulate a treatment recommendation if BP is above goal (be more aggressive, especially with systolic hypertension in older patients)
Emphasize lifestyle factors as part of the treatment program (involve patients in the treatment process); review progress and barriers at each visit
Use adequate doses of antihypertensive drugs; be willing to use multiple drugs
Encourage communication regarding medication costs and side effects
Be aware of poor control rates in men and African Americans

such patients, loop diuretics are required. For patients on multidrug regimens, the lack of a diuretic or the use of low doses of short-acting loop diuretics given only once daily may explain the resistant state. In some patients with renal disease, combinations of loop agents and thiazide diuretics are required to control volume.

Office Hypertension/Pseudohypertension

Office measures of BP may overestimate the usual or average level [*see* Ambulatory BP Monitoring, *above*]. Before embarking on further evaluation, the clinician should consider using out-of-office BP readings or ABPM to exclude a white-coat effect. Patients with white-coat hypertension are more likely to be younger (although cases of white-coat ISH do occur in elderly patients), female, and of normal weight. They often have no target organ injury and complain of fatigue and weakness (which are symptoms of hypotension) when drug doses are increased. One study suggested that up to 50% of hypertensive patients deemed resistant by office determinations of BP in fact had controlled hypertension.[72]

Some elderly patients may have pseudohypertension—falsely elevated systolic and diastolic BP as determined by cuff measurement that results from atherosclerosis of the brachial artery. Because of the excessive stiffness of the vessel wall, higher cuff pressure must be applied to produce vascular occlusion. In addition, the accuracy of oscillometric devices is impaired under these circumstances. Such patients often have evidence of severe generalized atherosclerosis and remarkable elevations of systolic BP without concomitant symptoms. They may complain of weakness and fatigue with increases in drug doses. The ability to palpate the pulseless radial artery after cuff inflation (i.e., a positive Osler sign) increases the likelihood of pseudohypertension, but this is not a sensitive test.[73] Confirmation of pseudohypertension requires intra-arterial measures of BP.

Secondary Hypertension

Secondary forms of hypertension are relatively uncommon in the general hypertensive population but may account for a significant proportion of cases of resistant hypertension. Once the other considerations have been eliminated, patients with resistant hypertension should be considered for further evaluation of secondary causes [*see* Secondary Hypertension, *below*].

HYPERTENSIVE CRISIS

An acute and severe rise in BP is a serious medical concern. Prompt therapy may be lifesaving. Clinically, acute and severe increases in BP can be classified as either hypertensive urgencies or emergencies (crises).[42]

The term hypertensive emergency or hypertensive crisis is defined as severely elevated BP associated with acute injury to target organs (i.e., brain, heart, kidneys, vasculature, and retina). Prompt hospitalization and reduction of BP with parenteral therapy is required. Examples of hypertensive emergencies include malignant hypertension, hypertensive encephalopathy, aortic dissection, eclampsia, unstable angina or acute myocardial infarction, pulmonary edema, and acute renal failure.

Malignant hypertension is an old term that describes a clinical syndrome associated with acute severe elevation of BP that may be fatal if not promptly treated. It is associated with a marked increase in peripheral vascular resistance caused by systemic (angiotensin II) or locally generated (endothelin) vasoconstrictors. Any form of hypertension can progress to the malignant phase. Clinical characteristics include severe hypertension (diastolic BP > 130 mm Hg); hemorrhages, exudates, and papilledema on retinal examination; encephalopathy (i.e., headache, confusion, somnolence, stupor, visual loss, focal neurologic deficits, seizure, or coma); oliguria and azotemia; nausea, vomiting, and dyspnea; and physical findings of heart failure (rales, an S_3 heart sound). Encephalopathy arises from the failure of cerebral autoregulation of blood flow at critically high pressures, which results in cerebral vasodilation, hyperperfusion, vascular leakage, and cerebral edema. The hallmark vascular lesion of malignant hypertension is fibrinoid necrosis of arterioles that, in turn, increases both ischemic injury and further vasoactive substance release, setting up a vicious cycle. Microangiopathic hemolytic anemia with fragmentation of red cells and intravascular coagulation may occur in the setting of fibrinoid necrosis.

Hypertensive urgency is defined as severe hypertension without evidence of acute target organ injury that requires BP reduction over 24 to 48 hours. Oral therapy in the outpatient setting is often adequate. Examples include severe hypertension in a patient with known coronary artery disease, an aortic aneurysm (or aneurysm at another site), or a history of heart failure. The term accelerated hypertension is often used to describe a state of acute, severe hypertension with hemorrhages and exudates on retinal examination (but not papilledema) but without other findings of acute organ injury. This condition can be managed with oral therapy but may progress to malignant hypertension if left untreated.

The causes of hypertensive urgencies and emergencies include neglected essential hypertension (approximately 7% of untreated hypertension can progress to the malignant phase), sudden discontinuance of drug therapy (especially multiple drug regimens or regimens containing clonidine or beta blockers), renovascular disease, collagen vascular disease (especially scleroderma), eclampsia, acute glomerulonephritis, and pheochromocytoma. Approximately 50% of hypertensive crises occur in patients with preexisting hypertension.

The goals of the initial evaluation are to assess for target organ injury and to define potential causes. The evaluation begins with a focused medical history and physical examination. In taking the history, the clinician must ask about compliance with prescribed antihypertensive medications and the use of drugs that can raise BP [*see Table 5*]. Retinal examination is a mandatory aspect of the physical examination. Immediate laboratory

Table 12 Parenteral Therapy for Hypertensive Crisis

Drug	*Dosage*	*Comments*
Sodium nitroprusside	0.25–10.0 μg/kg/min I.V. infusion	General drug of choice; produces direct arteriolar and venous dilation; immediate onset and offset; side effects include metabolic acidosis, nausea, vomiting, agitation, psychosis, tremor (monitor thiocyanate levels)
Labetalol	Repetitive I.V. boluses of 20–80 mg q. 10 min or constant infusion of 0.5–2.0 mg/min	Combination alpha/beta blocker; onset 5–10 min, offset 3–6 hr; useful in most settings, especially postoperative state, hypertensive crisis of pregnancy; avoid in acute heart failure; take beta-blocker precautions; side effects include scalp tingling, vomiting, heart block, orthostatic hypotension
Glyceryl trinitrate	5–100 μg/min I.V. infusion	Produces direct arteriolar and venous dilation; onset 5–10 min, offset 3–5 min; especially useful in acute coronary ischemia, CHF; tolerance with prolonged infusion; side effects include headache, flushing, nausea, methemoglobinemia
Esmolol	50–300 μg/kg/min I.V.	Cardioselective beta blocker, onset 1–2 min, offset 10–20 min; especially useful in postoperative state, aortic dissection, ischemic heart disease; take beta-blocker precautions; side effects include bradycardia, nausea
Hydralazine	10–20 mg I.V. bolus	Causes direct arteriolar vasodilation; onset 10–20 min, offset 3–8 hr; used primarily for hypertensive crisis of pregnancy; avoid in acute MI, angina, aortic dissection; side effects include headache, flushing, nausea, vomiting, tachycardia, angina
Enalapril	1.25–5 mg I.V. bolus, q. 6 hr	ACE inhibitor; onset 15 min, offset 6 hr; especially useful in acute heart failure in postoperative state; lower doses in renal disease; side effects include precipitous decline in BP (high-renin states), acute renal failure (presence of renal vascular disease)
Nicardipine	5–15 mg/hr I.V. infusion	Dihydropyridine calcium antagonist; onset 5–10 min, offset 1–4 hr; especially useful in postoperative state; avoid in acute heart failure; side effects include headache, nausea, flushing, phlebitis
Fenoldopam	0.1–1.6 μg/kg/min I.V. infusion	Dopamine (DA1) agonist; onset 5 min, offset 30–60 min; especially useful in patients with impaired renal function because it increases renal blood flow and sodium excretion; side effects include nausea, vomiting, headache, flushing
Phentolamine	5–15 mg I.V. bolus	Alpha blocker; onset instantaneous, offset 3–10 min; drug of choice for pheochromocytoma crisis; side effects include flushing, tachycardia
Trimethaphan	0.5–15 mg/min I.V. infusion	Ganglionic blocker; onset 1–5 min, offset 10 min; tachyphylaxis common with prolonged infusion; side effects include urinary retention, paralytic ileus, dry mouth, blurred vision, orthostatic hypotension
Diazoxide	1–150 mg/kg I.V. bolus over 10 min; repeat at 10–15 min intervals if needed	Considered obsolete; direct arteriolar dilator; onset 1 min, offset 3–18 hr; avoid in acute MI, aortic dissection; side effects include hyperglycemia, hyperuricemia, fluid retention

ACE—angiotensin-converting enzyme CHF—chronic heart failure MI—myocardial infarction

studies include a complete blood count (to check for anemia); blood smear (to look for fragmented red blood cells); serum creatinine and blood urea nitrogen assays; urinalysis; serum sodium, potassium, and glucose assays; a chest x-ray; and an electrocardiogram. In hypertensive crisis, evaluation for secondary hypertension should be deferred until the patient is stable. If a cause of the crisis is not apparent, such patients should eventually have an evaluation to exclude renal vascular disease, pheochromocytoma, scleroderma, and primary aldosteronism.

Patients with hypertensive crisis should be hospitalized in an intensive care unit. The challenge of treatment is to lower BP without aggravating ischemia to vital organs. Parenteral therapy should be used [*see Table 12*]. Sodium nitroprusside is generally the drug of choice. Diazoxide is considered obsolete, because of the availability of newer and safer drugs. Mean BP should be lowered by 20% in the first hour (diastolic BP should be reduced to 100 to 110 mm Hg). As BP is lowered, the patient should be monitored for evidence of worsening cerebral, renal, or cardiac function. If the patient is stable, BP should be further lowered over the next 24 hours. Oral therapy can be started, and parenteral therapy gradually discontinued.

TREATMENT FOR SPECIFIC PATIENT GROUPS

The Elderly

Approximately 60% to 70% of persons 60 years of age or older have hypertension.[1] In this age group, systolic BP is the dominant predictor of adverse events, and ISH is the most common type of blood pressure disturbance.[11,12] Treatment of hypertension in the elderly reduces CV disease event rates and lessens the risk of development and progression of cognitive dysfunction and dementia.[7,74] The benefits of treatment have been shown for persons with either systolic-diastolic hypertension or ISH and for those older than 80 years. Although most elderly persons have primary hypertension, secondary forms of hypertension should be considered if the onset is recent or the hypertension is resistant.

There are special concerns regarding BP measurement in the elderly. Systolic BP is often quite variable, and the phenomenon of white-coat hypertension may be common in the elderly, especially in older women. Thus, readings of BP outside the office should be encouraged, as should selective use of ambulatory monitoring, especially if the patient has no target organ changes

related to hypertension or complains of side effects that suggest hypotension with treatment. As noted, white-coat hypertension is an indication for ambulatory monitoring that is covered under Medicare.[40] Orthostatic hypotension and postprandial hypotension are more common in the elderly, in most cases because of dysautonomia of aging. Systolic hypertension is a predictor of orthostatic hypotension, and diabetic patients are at greater risk because of autonomic neuropathy. Thus, BP measurement in the standing position is required in all elderly patients at all office visits. Pseudohypertension should be considered in elderly patients who have palpably stiff vessels, who lack significant target organ changes despite very high BP readings, and who complain of hypotensive symptoms with treatment. Such patients may require a direct intra-arterial measure of BP for clarification.

In general, treatment of hypertension in the elderly follows the same principles as treatment in younger patients. The BP goals are the same as for the general hypertensive population. However, because the benefit of treatment on longevity is less in most elderly patients, the costs of drugs, side effects, and quality of life are important considerations. Goal BP may be difficult to achieve in some patients with systolic hypertension, but any reduction is beneficial. Thus, in some patients, a higher systolic goal may be reasonable.

Modification of adverse lifestyle factors is beneficial in the elderly and should be encouraged.[75] Salt sensitivity increases with age and with the reduction in renal function that is common in the elderly.[76] In patients who require drugs, lower initial doses should be considered, especially in the presence of orthostatism or comorbid vascular diseases. However, many elderly patients ultimately require multiple drugs for BP control.

In the elderly with systolic-diastolic hypertension, placebo-controlled studies have shown that initial therapy with a diuretic or a beta blocker is beneficial. In one trial, treatment using newer drugs (calcium antagonists or ACE inhibitors) was not superior to treatment using diuretics and beta blockers.[77] In another study, however, starting treatment with an ACE inhibitor rather than a diuretic was associated with better outcomes, particularly in men.[65] Studies in patients with ISH have shown efficacy of thiazide diuretics and long-acting dihydropyridine calcium antagonists.[7] In elderly patients with LVH, the LIFE (Losartan Intervention For Endpoint reduction in hypertension) trial demonstrated that, compared with therapy using a beta blocker (atenolol), use of an ARB (losartan) was associated with fewer CV events, including strokes.[78] This observation was noted overall and in the subset of elderly patients with ISH. In elderly patients with a history of stroke or transient ischemic attack, the combination of indapamide and perindopril reduced the risk for subsequent stroke and progression to dementia.[79] In many elderly patients, comorbid conditions will determine the use of specific drugs. Because of the problem of polypharmacy in the elderly, a goal should always be to keep the program as simple as possible.

Diabetic Patients

Patients who have both hypertension and diabetes have twice the risk of CV disease as nondiabetic hypertensive patients. In addition, hypertension increases the risk of diabetic retinopathy and nephropathy.[80] Epidemiologic and observational studies have shown that the risk of BP-related CV disease and mortality in diabetic patients begins to rise when BP exceeds 120/70 mm Hg.[80,81] There does not appear to be a threshold value for risk associated with systolic BP in diabetic patients. In the Hypertension Optimal Treatment Trial (HOT), diabetic patients randomized to the lowest diastolic BP goal (≤ 80 mm Hg; the achieved diastolic BP was 82.6 mm Hg) had the best outcomes.[56] In the United Kingdom Prospective Diabetes Study (UKPDS), a mean achieved diastolic BP of 82 mm Hg was beneficial, as compared with less aggressive BP reduction.[82] On the basis of these data, the American Diabetes Association, the National Kidney Foundation, and the JNC 7 report recommend a goal BP of less than 130/80 mm Hg in hypertensive diabetic patients.[8,81,83]

All patients with diabetes should be encouraged to adopt lifestyle modifications [*see Table 8*]. Weight loss (if the patient is overweight or obese) and moderate exercise are especially beneficial in diabetic patients because in addition to lowering BP, these interventions improve insulin sensitivity and blood lipid levels. Many patients will require lifestyle modifications and three or more drugs to achieve the BP goals. Meeting these goals may be difficult in some patients. The clinician must balance benefit from lower BP with cost of medication, side effects, and risks associated with the lower goals in some patients. The American Diabetes Association recommends a trial of lifestyle modifications alone for up to 3 months if the initial systolic BP is 130 to 139 mm Hg or the diastolic BP is 80 to 89 mm Hg. Drug monotherapy should be considered initially along with lifestyle modifications if the initial systolic BP is 140 mm Hg or higher or if the diastolic BP is 90 mm Hg or higher.[81] The JNC 7 report suggests that if the initial systolic BP is 150 mm Hg or higher or the initial diastolic BP is 90 mm Hg or higher, consideration should be given to starting therapy with a combination of two drugs, one of them a thiazide diuretic.[8] Before initiating drug therapy, it is important to measure BP in the standing position to detect orthostatism, the presence of which may be a clue to autonomic neuropathy and would necessitate a modification to the treatment approach.

Placebo-controlled trials in diabetic patients have shown the efficacy of ACE inhibitors, ARBs, diuretics, and beta blockers as initial therapy. Numerous studies have shown the effectiveness of ACE inhibitors and ARBs in retarding progression of diabetic nephropathy.[84,85] For diabetic patients with nephropathy, the American Diabetes Association guidelines recommend ACE inhibitors as initial drugs of choice in type 1 diabetes but ARBs in type 2 diabetes.[81] It is unclear whether ARBs are as cardioprotective in diabetic patients as ACE inhibitors have been shown to be. In some studies, the incidence of cardiac events has been higher in diabetic patients treated with dihydropyridine calcium antagonists, as compared with ACE inhibitors.[86] Beta blockers should be considered in the setting of coronary artery disease, a common comorbidity in patients with diabetes. On balance, treatment data suggest that reaching the goal BP in diabetic patients is probably more important than the choice of drugs used to achieve it.

Patients with Heart Disease

Ischemic heart disease is the most common cause of death in patients with hypertension. Poorly controlled hypertension also results in the development of LVH. Both LVH and ischemic injury lead to the development of heart failure from either systolic or diastolic dysfunction.[87] Hypertension is the most common antecedent of heart failure.[88] Hypertensive effects on the heart also increase the risk for atrial fibrillation.

For asymptomatic patients with known coronary artery dis-

ease, an ACE inhibitor should be considered initially because some studies (but not all) suggest that their use may be associated with a reduced risk of cardiovascular events.[89] An ACE inhibitor would also be the initial drug of choice for patients with concomitant reduced systolic function or concomitant diabetes with renal involvement.[81,84,90] If there is a history of myocardial infarction, the first drug should be a beta blocker.[91] For hypertensive patients with previous myocardial infarction and reduced left ventricular function, combination therapy with a beta blocker and an ACE inhibitor should be considered.[92] In addition, the aldosterone antagonist eplerenone has been shown to be effective.[93] If eplerenone is used, serum potassium levels should be monitored carefully, especially in patients with renal dysfunction or if ACE inhibitors or ARBs are used.

The drug of choice in hypertensive patients with stable angina, both to lower BP and to relieve symptoms and ischemia, is a beta blocker. Long-acting dihydropyridine or nondihydropyridine calcium antagonists have been shown to relieve symptoms and are alternative agents if beta blockers are contraindicated; these alternative agents are also suitable as additional therapy for BP or symptom control. Newer vasoselective, long-acting dihydropyridine calcium antagonists such as amlodipine or felodipine can be used safely to lower BP in patients with impaired left ventricular function. Nitrates can be used in combination with either beta blockers or calcium antagonists for symptomatic relief and may lower systolic BP. Beta blockers should be avoided in pure vasospastic angina, a disorder best managed with long-acting calcium antagonists or nitrates. Diuretics are safe antihypertensives for patients with coronary artery disease; they work well with other agents to lower BP. Hypokalemia should be avoided.

LVH is associated with a doubling of the risk of myocardial infarction and death in hypertensive patients.[94] Effective BP control causes regression of LVH and improves prognosis. Weight loss and the use of antihypertensive drugs of all major classes have been shown to induce regression of LVH; however, increasing evidence suggests that ACE inhibitors and ARBs may be more effective than other agents.[95]

The goal of treating hypertensive patients with heart failure is a BP of less than 130/80 mm Hg. The American College of Cardiology/American Heart Association have developed guidelines for the evaluation and management of heart failure in adults that encompass a staging system and evidence-based treatment recommendations for patients with heart failure [*see I:II Heart Failure*].

Patients with Chronic Kidney Disease

Kidney disease is both a cause and a consequence of hypertension. Hypertension is the second most common cause of the development of end-stage kidney disease, and most people with kidney disease have hypertension. Aggressive control of elevated BP can slow progression of renal damage and delay or prevent the development of end-stage disease.[83-85,96] The currently recommended goal BP for patients with kidney disease is a level below 130/80 mm Hg. Also, patients with chronic kidney disease are at high risk for CV morbidity and mortality. Therefore, in addition to elevated BP, other modifiable CV risk factors require management.

Chronic kidney disease is defined as either a GFR of less than 60 ml/min/1.73 m^2 or the presence of albuminuria (> 300 mg/day or > 200 mg albumin per gram of creatinine).[83] The GFR can be estimated using the Cockcroft-Gault or MDRD equation [*see Table 6*]. Determination of creatinine clearance using timed urine collections generally does not improve upon the estimates of GFR obtained using these equations.

ACE inhibitors and ARBs may be more effective than other drugs in slowing progression of proteinuric kidney disease. Whether these agents provide a specific advantage in the absence of proteinuria is less certain.[83] Serum creatinine concentrations often increase acutely when these drugs are used, so serum creatinine and potassium should be measured within several days of initiating treatment. An increase in creatinine is not a reason to stop the drug unless it is excessive or associated with severe hyperkalemia. Concomitant use of potassium-sparing diuretics, potassium supplements, or nonsteroidal anti-inflammatory drugs should be avoided. A persistent increase in creatinine with treatment raises the possibility of renal artery stenosis. Most patients with kidney disease will require a diuretic as part of the treatment regimen. If GFR is estimated to be less than 30 ml/min, thiazide diuretics are usually ineffective, and loop diuretics are required. Often, three or more drugs are required to control BP.

Patients with Acute Stroke

The majority of patients presenting with either acute ischemic or hemorrhagic stroke have hypertension.[97] The temporal profile is that of an initial acute rise in BP in the first 24 hours, followed by a slow decline over the next several days. On the whole, observational studies show that high BP at stroke onset is associated with an increased risk of death or dependency.[98] However, this association is not evident in some studies, especially studies in patients with ischemic stroke.[99]

Unfortunately, at present there is little evidence from clinical trials to provide clear recommendations for the appropriate management of BP during acute stroke. Currently, there is consensus that in patients with acute intracranial hemorrhage, BP should be lowered if it exceeds 200/120 mm Hg, to prevent growth of the hematoma or rebleeding. Lowering of BP by less than 20% is suggested in this setting.[100] Guidelines for BP management in acute ischemic stroke from the Stroke Council of the American Heart Association suggest that in patients who are not candidates for thrombolytic therapy, hypertension should be managed with observation alone if BP is less than 220 mm Hg systolic and 120 mm Hg diastolic, unless there is evidence of other acute target-organ injury (e.g., aortic dissection, acute myocardial infarction, pulmonary edema, hypertensive encephalopathy).[101] For patients with systolic BP higher than 220 mm Hg or diastolic BP of 121 to 140 mm Hg, treatment with intravenous labetalol or nicardipine is recommended. Labetalol is given in a dosage of 10 to 20 mg over 1 to 2 minutes; the dose is repeated or doubled as needed every 10 minutes to a maximum dose of 300 mg. Nicardipine is given in an initial 5 mg/hr infusion and titrated to desired effect by increasing the dosage by 2.5 mg/hr every 5 minutes, to a maximum rate of 15 mg/hr. It is suggested that BP be lowered by 10% to 15%. Nitroprusside is recommended if diastolic BP is higher than 140 mm Hg; the dose should be titrated to lower BP by 10% to 15%.

In patients who are eligible for thrombolytic therapy, the Stroke Council suggests lowering BP before initiating thrombolysis if the BP is higher than 185 mm Hg systolic or 110 mm Hg diastolic. Treatment with labetalol, 10 to 20 mg intravenously over 1 to 2 minutes, is advised. If needed, this dose can be repeated once; or nitroglycerin paste, 1 to 3 inches, can be applied.

Table 13 Screening Options for Secondary Hypertension

Disorder	*Screening Tests (Sensitivity/specificity)*	*Comments*
Renovascular hypertension[112-114]	Captopril radionuclide renal scan (75%/85%)	Advantage: no contrast allergy Disadvantages: renal dysfunction impairs interpretation, may miss accessory- or branch-vessel disease
	Duplex ultrasound (80%–90%/90%)	Advantages: no contrast allergy, can be used in patients with renal dysfunction; calculation of resistive index identifies patients with renal dysfunction likely to benefit from intervention Disadvantages: failure to visualize both renal arteries; may miss accessory- or branch-vessel disease
	Spiral CT angiography	Advantages: excellent images of renal arteries; can identify dissection, accessory vessels, and fibromuscular disease Disadvantage: considerable contrast load precludes use in presence of renal dysfunction
	Magnetic resonance angiography (85%–100%/79%–98%)	Advantages: contrast allergy and renal dysfunction; no radiation exposure Disadvantages: cost, may overstate degree of stenosis, claustrophobic patients may not tolerate test
	Renal angiography	Gold standard Advantages: identifies accessory- and branch-vessel disease; percutaneous interventions can be performed as part of study Disadvantages: cost, contrast exposure, invasive (atheroemboli)
Primary aldosteronism[105,115]	Measurement of serum sodium, potassium, PRA, and PAC; 24-hr urinary aldosterone, sodium, and PRA after 3 days of 200 mEq sodium diet	Diagnosis confirmed if U_{Na} > 200 mEq, U_{aldo} > 12, and PRA < 1.0; 30% of patients with primary aldosteronism will be normokalemic at presentation Ratio of PAC/PRA > 20 (PAC >15 ng/dl and PRA < 2.0 ng/ml) Advantage: simple Disadvantages: many antihypertensive drugs can influence values of PRA and PAC; sensitive screen but not specific
Pheochromocytoma[107]	Plasma free metanephrine (highly sensitive); 24-hr fractionated urinary metanephrines	—
Cushing syndrome	24-hr urinary free cortisol (95%–100%/97%–100%)	Diagnosis certain if 24-hr urinary free cortisol level > 3× normal; diagnosis excluded if level normal; use low-dose dexamethasone suppression test if elevation < 3× normal
Coarctation of the aorta	Chest x-ray; transesophageal echocardiogram; CT or MRI of the aorta	Diagnostic findings on chest x-ray: "3" sign from dilation of aorta above and below the coarctation, rib notching from collateral vessels

PAC— plasma aldosterone concentration PRA—plasma renin activity

If antihypertensive treatment does not reduce BP to below 185/110 mm Hg, thrombolytic therapy is not advised. During and after thrombolytic treatment, BP should be monitored frequently (every 15 minutes for 2 hours, then every 30 minutes for 6 hours, and then every hour for 16 hours). During this period, treatment with nitroprusside is advised for diastolic BP higher than 140 mm Hg. For systolic BP higher than 180 mm Hg or diastolic BP of 105 to 140 mm Hg, intravenous labetalol in a dosage of 10 mg administered over 1 to 2 minutes is recommended; the dosage should be repeated or doubled every 10 minutes to a maximum of 300 mg, or a drip at a rate of 2 to 8 mg/min should be started.

Secondary Hypertension

Detection of secondary hypertension is important because, depending on the cause, it may be possible to cure the underlying condition or tailor therapy to achieve optimal BP control. Certain features suggest the presence of specific secondary forms of hypertension [*see Table 4*], which should then direct further testing [*see Table 13*].

Common reversible causes of hypertension include obesity, the use of drugs that raise BP [*see Table 5*], obstructive sleep apnea, and renal disease. Obstructive sleep apnea is prevalent in the population and is often associated with hypertension. Renal insufficiency from any etiology causes BP to rise. Elevated BP in turn accelerates loss of renal function, and a vicious cycle ensues. Traditional secondary causes of hypertension include renal vascular disease, coarctation of the aorta, the adrenal causes of primary aldosteronism, pheochromocytoma, and Cushing syndrome.

RENOVASCULAR HYPERTENSION

Renovascular hypertension is the most common form of potentially curable secondary hypertension. It probably occurs in 1% to 2% of the overall hypertensive population. The prevalence may be as high as 10% in patients with resistant hypertension, and even higher in patients with accelerated or malignant hypertension.

Stenosing lesions of the renal circulation cause hypertension through ischemia-mediated stimulation of the renin-angiotensin-aldosterone axis. Correcting renal ischemia eliminates excess renin production and improves or cures the hypertension. In unilateral disease, prolonged hypertension can cause nephrosclerosis in the nonischemic kidney; nephrosclerosis lessens the likelihood of benefit from correction of the renal vascular lesion.

Fibromuscular disease is the most common cause of renovascular hypertension in younger patients, especially women between 15 and 50 years of age; it accounts for approximately 10% of cases of renovascular hypertension.[102] Vascular lesions typically affect the middle and distal portions of the renal artery and often extend into branches. Three subtypes are defined on the basis of the layer of the vascular wall affected: (1) intimal hyperplasia (1% to 2% of cases), (2) medial fibromuscular dysplasia (95% of cases), and (3) periadventitial fibrosis (1% to 2% of cases). The most common subtype, medial fibromuscular dysplasia, presents as a classic string-of-beads (aneurysmal dilatations) on angiography; it progresses in 30% of cases. It is rarely associated with dissection or thrombosis. In contrast, the rarer forms can progress rapidly, and dissection and thrombosis are common. Fibromuscular dysplasia is a rare cause of renal artery occlusion.

Atheromatous disease is the most common cause of renovascular hypertension in middle-aged and older patients and accounts for approximately 90% of renovascular hypertension.[102] Vascular lesions are usually in the proximal third of the renal arteries, often near or at the orifice. The prevalence of atheromatous renal artery disease increases with age and is common in older hypertensive patients, especially in those with diabetes or with atherosclerosis in other vascular beds. Most patients with atheromatous renal vascular disease and hypertension have essential hypertension. The disease is frequently bilateral (30%) and is often progressive. The likelihood of progression can be decreased by aggressive control of risk factors (e.g., dyslipidemia, cigarette smoking, and hypertension).

The presentations of hemodynamically significant bilateral renal artery disease (ischemic nephropathy) include the following: an acute decline in renal function with use of an ACE inhibitor or ARB or with a sudden decrease in blood pressure; acute hypertension and pulmonary edema (flash pulmonary edema); or an unexplained subacute decline in renal function with or without worsening of hypertension.[103] Bilateral atherosclerotic renal artery disease accounts for a small but increasing number of cases of end-stage renal disease in older persons.[104]

Atheroembolic renal disease can mimic renovascular hypertension and ischemic nephropathy, in that it may present as hypertension of acute onset or as a worsening of hypertension in conjunction with a subacute decline in renal function. Atheroembolic renal disease often occurs after angiography or vascular surgery. Physical findings include the presence of distal livido reticularis and peripheral emboli. Laboratory findings include an elevated erythrocyte sedimentation rate, anemia, hematuria, eosinophilia, and eosinophiluria.

PRIMARY ALDOSTERONISM

The classic syndrome of primary aldosteronism consists of hypertension, hypokalemia from excessive renal excretion, alkalosis, suppressed plasma renin activity, and increased aldosterone secretion.[105] Hypokalemia is the abnormality that most often raises suspicion of this disorder, but approximately 30% of patients with primary aldosteronism present with normal serum potassium levels.

Although several subtypes of primary aldosteronism have been identified, the most common are unilateral aldosterone-producing adenoma, which comprises 30% to 40% of cases; and bilateral adrenal zona glomerulosa hyperplasia (also known as idiopathic hyperaldosteronism [IHA]), which comprises 60% to 70% of cases. Rare subtypes include glucocorticoid-suppressible hyperplasia, unilateral hyperplasia, and aldosterone-producing cortical carcinoma. The prevalence of primary aldosteronism is probably around 2%, but studies have suggested the prevalence to be as high as 15% of the hypertensive population. The higher prevalence estimates reflect an increase in the number of patients being diagnosed with IHA, a condition that may be part of the spectrum of essential hypertension.

Patients for whom the diagnosis of primary aldosteronism should be considered include the following: all hypertensive patients with spontaneous hypokalemia of renal origin (for a hypokalemic patient, a 24-hour urinary potassium level higher than 30 mEq/L is consistent with renal potassium wasting); most patients with excessive hypokalemia who are receiving usual doses of diuretics (serum potassium < 3.0 mEq/L); most patients with resistant hypertension, even if normokalemic; and all patients with hypertension and an adrenal mass.

PHEOCHROMOCYTOMA

Pheochromocytomas are rare tumors of chromaffin cell origin that produce excess amounts of catecholamines, which leads to paroxysmal or sustained hypertension. The incidence in the general population is 2 to 8 cases per million persons per year. The prevalence is about 0.5% in patients with hypertension who have suggestive symptoms, and approximately 4% in patients with adrenal incidentalomas. Most tumors are benign, but approximately 10% are malignant. Symptomatic paroxysms occur in less than 50% of patients. Episodes are characterized by symptoms of headache, diaphoresis, palpitations, and pallor associated with increases in blood pressure.[106] Such paroxysms are usually rapid in onset, and offset and can be precipitated by a variety of activities (e.g., exercise, bending over, urination, defecation, induction of anesthesia, infusion of intravenous contrast media, smoking). A history of unintended weight loss is not uncommon. The hypertension may be associated with marked BP lability and orthostatic hypotension. Rarely, patients may present with catecholamine-induced cardiomyopathy, fever, or peripheral vasospasm. The hypertension can be severe and resistant to control.

Most pheochromocytomas are sporadic, but 10% are familial. Familial syndromes include a simple autosomal dominant form not associated with other abnormalities, the multiple endocrine neoplasias (type IIA [medullary thyroid carcinoma, hyperparathyroidism] and type IIB [medullary thyroid carcinoma, mucosal neuromas, marfanoid habitus, thickened corneal nerves, intestinal gangliomatosis]), neurofibromatosis, and the von Hippel-Lindau syndrome (retinal hemangiomatosis, cerebellar hemangioblastomas, renal cell carcinoma). Familial pheochromocytomas can be bilateral.

Most pheochromocytomas (90%) are located in one or both adrenal glands. Extra-adrenal pheochromocytomas can occur anywhere along the sympathetic chain and, rarely, in other sites (i.e., the superior para-aortic region, the glomus jugulare, the inferior para-aortic region, the bladder, or the thorax). About 98% of pheochromocytomas are located in the abdomen.

Screening for pheochromocytoma should be selective and based on suggestive clinical features. Screening tests include measurement of catecholamines (i.e., epinephrine, norepinephrine, dopamine) and their metabolites (i.e., metanephrine, normetanephrine, and vanillylmandelic acid [VMA]) in the plasma and urine. Traditionally, most experts have considered measurement of 24-hour urinary catecholamines or catecholamine metabolites to be the screening tests of choice.[106] However, stud-

Table 14 Causes for False Positive Screening Results for Plasma Free Metanephrines and 24-Hour Urinary Metanephrines

Category	*Sources*	*Tests Affected*
Diet	Coffee (including decaffeinated)	Plasma metanephrines/urinary HPLC electrophoresis
	Acetaminophen (direct effect)	Plasma metanephrines/urinary HPLC electrophoresis
	Caffeine (increases plasma catecholamines)	All
	Unknown diet sources	Plasma metanephrines/urinary HPLC electrophoresis
Drugs	Nicotine (increases plasma catecholamines)	All
	Tricyclic antidepressants (norepinephrine and its metabolites)	All
	Dibenzyline (norepinephrine and its metabolites)	All
	Drugs containing catecholamines (decongestants)	All
	Labetalol	Urinary HPLC electrophoresis
	Withdrawal from clonidine	All
	Withdrawal from alcohol	All
	Withdrawal from benzodiazepines	All
	Levodopa	All
	Cyclobenzaprine	All
	Amphetamines	All
	Phenothiazines	Unknown
	Benzodiazepines	Unknown
Physiologic stress	Obstructive sleep apnea, heart failure	All

HPLC—high-pressure liquid chromatography

ies now suggest that measurement of plasma free metanephrines is a much more sensitive screening test (for hereditary tumors, sensitivity is 97%, versus 60% for urinary metanephrines; for sporadic tumors, sensitivity is 99%, versus 88% for urinary metanephrines).[107] Also, this screening test obviates the concerns associated with obtaining an adequate 24-hour urine collection. Although the sensitivity of the plasma screen is higher than that of urinary tests, its specificity is lower with regard to screening for sporadic tumors (for sporadic tumors, specificity is 82% with the plasma test versus 89% with the urinary test; in hereditary tumors, specificity is 96% with the plasma test versus 97% with the urinary test). Plasma metanephrine assay should be strongly considered as the screening test of choice if a hereditary form of pheochromocytoma is suspected; it should also be considered the test of choice for patients with a history of pheochromocytoma and for patients in whom the clinical suspicion is high. A negative result on either a plasma or urinary metanephrine test excludes the diagnosis in most cases.

Because of the low prevalence of pheochromocytomas in patients screened for this disorder, false positive results outnumber true positive results. This is of major concern because positive results from screening tests often lead to additional tests and anxiety on the part of the both the physician and the patient. There are three main factors associated with false positive results: diet, drugs, and physiologic stressors [*see Table 14*]. Specific dietary factors and drugs affect plasma screens and urinary metanephrine screens differently. Moreover, for urinary metanephrine screens, different drugs affect the results differently depending on the method of analysis used (i.e., high-pressure liquid chromatography [HPLC] versus mass spectrophotometry). Anticipation of these potential problems and proper preparation of the patient can prevent many false positive results.

A positive screening test should prompt a search for the tumor if sources of a false positive result have been excluded. Abdominal imaging with computed tomography or magnetic resonance imaging is the initial test of choice, given that 90% of pheochromocytomas are on the adrenal glands and 98% are in the abdomen. Additional studies may be required if a tumor is not found with initial imaging.[108] Medical treatment is required before surgical intervention. The mainstay of treatment is alpha blockade with phenoxybenzamine. Beta blockers can be used to control the tachycardia that occasionally follows adequate alpha blockade. Because pheochromocytomas can recur in 10% of patients, long-term biochemical follow-up is required.

CUSHING SYNDROME

Cushing syndrome arises from excess production of glucocorticoids. It is rare: the incidence of the ectopic adrenocorticotropic hormone (ACTH) syndrome is about 660 cases per million population; in 50% of these cases, the underlying cause is small cell lung cancer. The incidence of adrenal tumors is one to five cases per million population per year. The incidence of pituitary ACTH-dependent disease is estimated to be five to 25 cases per million population per year. The signs and symptoms of Cushing syndrome arise from long-term exposure to excess glucocorticoids. They include central obesity, skin atrophy, striae, acne, slow wound healing, proximal muscle wasting and weakness, osteoporosis, menstrual irregularity, hyperpigmentation (ACTH dependent), glucose intolerance, hypokalemia, and hypertension. Clinical manifestations vary on the basis of degree and duration of glucocorticoid excess, the presence or absence of androgen excess (in women, androgen excess produces hirsutism, decreased libido, virilization, and oily skin), and the cause of hypercortisolism (hyperpigmentation results from excessive ACTH; androgen excess is more common in adrenal carcinomas). States of pseudo–Cushing syndrome can result from significant stress, severe obesity, depression, and chronic alcoholism.[109]

Patients suspected of having Cushing syndrome should undergo measurement of 24-hour urinary free cortisol. Normal levels exclude the diagnosis, and levels higher than threefold normal confirm it. In patients with equivocal results, a low-dose

dexamethasone suppression test can be used. For this test, the patient is given a 1 mg tablet at 11 P.M. Serum cortisol is measured on a specimen drawn the next morning at 8 A.M. A normal response is a serum cortisol level of less than 5 μg/dl. An alternative method is to give a 0.5 mg tablet every 6 hours for eight doses and to measure 24-hour urinary cortisol excretion on the second day. A normal response is a urinary cortisol excretion of less than 10 μg/24 hr and a serum cortisol level of less than 5 μg/dl.

COARCTATION OF THE AORTA

Congenital constriction of the aorta accounts for approximately 7% of congenital cardiovascular diseases. Coarctation can occur anywhere along the aorta but most often occurs just distal to the takeoff of the left subclavian artery. The disorder is usually detected in childhood, but occasionally it escapes detection until adulthood. Symptoms include headache, cold feet, and claudication. The classic feature of coarctation is elevated blood pressure in the arms and low or unobtainable blood pressure in the legs. This finding can be identified by direct measurement. The presence of weak femoral pulses or a delay in sensing the femoral pulse when simultaneously palpating the radial pulse is cause to suspect coarctation. Other findings include visible pulsations in the neck or chest wall and murmurs in the front and back of the chest from collateral vessels. Physical findings may be subtle. If the diagnosis is suspected, screening tests include transesophageal echocardiography or MRI or CT imaging of the aorta. Treatment is surgical in most cases [*see 1:XII Diseases of the Aorta*].

Complications

Left untreated, hypertension leads to premature death or disability from complications of CV diseases, especially atherosclerosis.[3-5] Hypertension affects blood vessels directly, inducing endothelial dysfunction, and acts in concert with other factors (e.g., smoking, hyperlipidemia, and diabetes) to promote the atherosclerotic process. Although the effect of hypertension on blood vessels is systemic, it expresses itself by characteristic effects on target organs—the heart, brain, kidneys, and eyes.

Hypertension increases the risk of myocardial infarction and sudden cardiac death twofold.[3] It contributes to the risk of atrial fibrillation and is the single most important antecedent to the development of heart failure.[87,88] These adverse effects of hypertension reflect both acceleration of atherosclerosis and the development of structural adaptation of the heart (LVH and left atrial enlargement) to increased afterload. The structural changes limit coronary reserve.

Heart failure can be the result of either systolic or diastolic dysfunction. Hypertension is commonly associated with abnormal diastolic relaxation, which can be demonstrated by echocardiography. Progression of these effects on the heart can lead to symptoms of heart failure with preserved systolic function, a condition known as diastolic heart failure [*see 1:II Heart Failure*]. In addition, long-standing hypertension leads to LVH and ventricular remodeling, which progresses to systolic dysfunction. This process is aggravated by myocardial infarction.

Hypertension is the single most important cause of stroke, which itself is the third leading cause of death in the United States.[110] Hypertension increases the risk of stroke by aggravating atherosclerosis in the aortic arch and carotid and cerebral arteries (causing thrombotic or embolic ischemic strokes) and by inducing arteriosclerosis in small, penetrating subcortical cerebral vessels, leading to leukoaraiosis (periventricular leukoenceophalopathy) and lacunar strokes. Severe hypertension is also associated with intraparenchymal and subarachnoid hemorrhage.

Hypertension in midlife is associated with an increased risk of cognitive dysfunction and dementia in later life.[5] This may be a complication of multiple cerebral infarctions (multi-infarct dementia), but it also occurs in the absence of previous strokes. In some persons, cognitive dysfunction may arise from the effects of elevated BP on the small penetrating subcortical arterioles, leading to ischemic injury to white matter (visible as leukoaraiosis on brain imaging studies). Although vascular dementia in the elderly is strongly related to hypertension, the relationship between BP and cognition is less clear in persons older than 75 years. In some cases, an inverse relationship has been noted. This may reflect a shift in cerebral autoregulation to a higher range in patients with hypertension-induced small vessel disease, making these patients more vulnerable to further ischemic brain injury when BP is lowered.

Hypertension is a risk factor for abdominal aortic aneurysm. In addition, the majority of patients with aortic dissection have hypertension. Aortic dissection arises from the combined effects of accelerated aortic atherosclerosis and increased pulsatile stress on the aortic wall. Hypertension increases the risk of peripheral vascular disease, especially in cigarette smokers and diabetic patients.

Hypertension is the second leading cause of end-stage renal disease.[104] Arteriosclerotic changes lead to ischemic injury and loss of glomeruli and tubular elements, ultimately leading to the shrunken kidney of nephrosclerosis. End-stage kidney disease from hypertension is much more common in blacks. Malignant hypertension induces fibrinoid necrosis of renal arterioles and can lead to acute renal failure.

Hypertension-related vascular disease causes loss of vision through a variety of mechanisms.[111] Chronic hypertension causes arteriosclerosis of retinal vessels. These changes at the site of arterial-venous crossings can lead to branch retinal vein occlusion. Central retinal vein occlusion can also occur. Ischemic optic neuropathy can be a complication of chronic hypertension or acute severe hypertension. Acute, severe elevations in BP can also cause retinal hemorrhages, exudates, and papilledema. Hypertension accelerates atherosclerosis. Atherosclerotic emboli can occlude central or branch retinal arteries, with sudden and irreversible visual loss. Reduced blood flow in the carotid or ophthalmic artery because of severe atherosclerosis can cause venous stasis retinopathy. Occlusive disease of retinal vessels can lead to cystoid macular edema, epiretinal membrane formation, and collateral vessel formation.

Prognosis

Effective treatment has a dramatic effect on the prognosis of patients with hypertension. Prospective treatment trials have established that BP reduction with drug therapy markedly reduces CV morbidity and mortality. Active treatment of hypertension lessens the tendency for BP to increase over time. For patients whose diastolic BP is 90 mm Hg more or whose systolic BP is 160 mm Hg or more, drug intervention has been shown to reduce the risk of stroke by 35% to 40%; the risk of myocardial infarction is reduced by 20% to 25%; and the risk of heart failure is reduced by over 50%. In hypertensive patients with chronic

kidney disease, drug intervention reduces the risk of progression to dialysis, transplantation, and death. However, even when BP is brought down to current recommended levels, hypertensive individuals remain at higher risk for CV disease events compared with normotensive individuals. Patients with target-organ disease remain at even higher risk, despite good BP control. These observations argue for application of public health and individual patient strategies to prevent the development of hypertension and for early detection and effective treatment of high BP.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hajjar I, Kotchen TA: Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988–2000. JAMA 290:199, 2003

2. Vasan RS, Beiser A, Seshadri S, et al: Residual lifetime risk for developing hypertension in middle-aged women and men: the Framingham Heart Study. JAMA 287:1003, 2002

3. Kannel WB: Blood pressure as a cardiovascular risk factor: prevention and treatment. JAMA 275:1571, 1996

4. Whelton PK, He J, Perneger TV, et al: Kidney damage in "benign" essential hypertension. Curr Opin Nephrol Hypertens 6:177, 1997

5. Launer LJ, Masaki K, Petrovitch H, et al: The association between midlife blood pressure levels and late-life cognitive function: the Honolulu-Asia Aging Study. JAMA 274:1846, 1995

6. Staessen JA, Wang JG, Thijs L: Cardiovascular protection and blood pressure reduction: a meta-analysis. Lancet 358:1305, 2001

7. Staessen JA, Gasowski J, Wang JG, et al: Risks of untreated and treated isolated systolic hypertension in the elderly: meta-analysis of outcome trials. Lancet 355:865, 2000

8. Chobanian AV, Bakris GL, Black HR, et al: The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 Report. JAMA 289:2560, 2003

9. Lewington S, Clarke R, Qizilbash N, et al: Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data from one million adults in 61 prospective studies. Lancet 360:1903, 2002

10. Vasan RS, Larson MG, Leip EP, et al: Assessment of frequency of progression to hypertension in non-hypertensive participants in the Framingham Heart Study: a cohort study. Lancet 358:1682, 2001

11. Franklin SS, Jacobs MJ, Wong ND, et al: Predominance of isolated systolic hypertension among middle-aged and elderly US hypertensives: analysis based on National Health and Nutrition Examination Survey (NHANES) III. Hypertension 37:869, 2001

12. Franklin SS, Larson MG, Khan SA, et al: Does the relation of blood pressure to coronary heart disease risk change with aging? The Framingham Heart Study. Circulation 103:1245, 2001

13. Miura K, Daviglus ML, Dyer, et al: Relationship of blood pressure to 25-year mortality due to coronary heart disease, cardiovascular diseases, and all causes in young adult men: the Chicago Heart Association Detection Project in Industry. Arch Intern Med 161:1501, 2001

14. Lifton RP, Gharavi AG, Geller DS: Molecular mechanisms of human hypertension. Cell 104:545, 2001

15. O'Donnell CJ, Lindpainter K, Larson MG, et al: Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not in women in the Framingham Heart Study. Circulation 97:1766, 1998

16. Caulfield M, Lavender P, Farrall M, et al: Linkage of the angiotensinogen gene to essential hypertension. N Engl J Med 330:1629, 1994

17. Rigat B, Hubert C, Alhenc-Gelas F, et al: An insertion/deletion polymorphism in the angiotensin I–converting enzyme gene accounting for half the variance of serum enzyme levels. J Clin Invest 86:1343, 1990

18. Cambien F, Poirier O, Lecerf L, et al: Deletion polymorphism in the gene for angiotensin-converting-enzyme is a potent risk factor for myocardial infarction. Nature 359:641, 1992

19. Turner ST, Boerwinkle E, Sing CF: Context-dependent associations of the ACE I/D polymorphism with blood pressure. Hypertension 34(4 pt 2):773, 1999

20. Cusi D, Barlassina C, Azzani T, et al: Polymorphisms of α-adducin and salt sensitivity in patients with essential hypertension. Lancet 349:1353, 1997

21. Krushkal J, Xiong M, Ferrell R, et al: Linkage and association of adrenergic and dopamine receptor genes in the distal portion of the long arm of chromosome 5 with systolic blood pressure variation. Hum Mol Genet 7:1379, 1998

22. Sing CF, Stengard JH, Kardia SLR: Genes, environment, and cardiovascular disease. Arterioscler Thromb Vasc Biol 23:1190, 2003

23. Intersalt: an international study of electrolyte excretion and blood pressure: results for 24 hour urinary sodium and potassium excretion. Intersalt Cooperative Research Group. BMJ 297:319, 1988

24. Cutler JA, Follmann D, Allender PS: Randomized trials of sodium reduction: an overview. Am J Clin Nutr 65(2 suppl):643S, 1997

25. Guyton AC: Kidneys and fluids in pressure regulation: small volume but large pressure changes. Hypertension 19(suppl):I2, 1992

26. Brenner BM, Chertow GM: Congenital oligonephropathy and the etiology of adult hypertension and progressive renal injury. Am J Kidney Dis 23:171, 1994

27. Sealey JE, Blumenfeld JD, Bell GM, et al: On the renal basis for essential hypertension: nephron heterogeneity with discordant renin secretion and sodium excretion causing a hypertensive vasoconstriction-volume relationship. J Hypertens 6:763, 1988

28. Blaustein MP: Endogenous ouabain: role in the pathogenesis of hypertension. Kidney Int 49:1748, 1996

29. Esler M, Rumantir M, Kaye D, et al: The sympathetic neurobiology of essential hypertension: disparate influences of obesity, stress, and noradrenaline transporter dysfunction? Am J Hypertens 14(6 pt 2):139S, 2001

30. Izzo JL, Taylor AA: The sympathetic nervous system and baroreflexes in hypertension and hypotension. Curr Hypertens Rep 1:254, 1999

31. Grassi G, Seravalle G, Dell'Oro R, et al: Adrenergic and reflex abnormalities in obesity-related hypertension. Hypertension 36:538, 2000

32. Landsberg L: Insulin-mediated sympathetic stimulation: role in the pathogenesis of obesity-related hypertension (or, how insulin affects blood pressure, and why). J Hypertens 19(3 pt 2):523, 2001

33. Haynes WG, Morgan DA, Walsh SA, et al: Cardiovascular consequences of obesity: role of leptin. Clin Exp Pharmacol Physiol 25:65, 1998

34. Jackson WF: Ion channels and vascular tone. Hypertension 35(1 pt 2):173, 2000

35. Cosentino F, Lüscher TF: Effects of blood pressure and glucose on endothelial function. Curr Hypertens Rep 3:79, 2001

36. Recommendations for routine blood pressure measurement by indirect cuff sphygmomanometry. American Society of Hypertension. Am J Hypertens 5(4 pt 1):207, 1992

37. Gosse P, Ansoborlo P, Jullien V, et al: Ambulatory blood pressure and left ventricular hypertrophy. Blood Press Monit 2:70, 1997

38. Verdecchia P, Porcellati C, Schillaci G, et al: Ambulatory blood pressure: an independent predictor of prognosis in essential hypertension. Hypertension 24:793, 1994

39. Clement DL, De Buyzere ML, De Bacquer DA, et al: Prognostic value of ambulatory blood pressure recordings in patients with treated hypertension. N Engl J Med 348:2407, 2003

40. Coverage and billing of ambulatory blood pressure monitoring (ABPM). Centers for Medicare & Medicaid Services, 2001
http://cms.hhs.gov/manuals/pm_trans/AB01188.pdf

41. Pickering T, Davidson K, Gerin W, et al: Masked hypertension. Hypertension 40:795, 2002

42. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med 157:2413, 1997

43. 2003 European Society of Hypertension–European Society of Cardiology guidelines for the management of arterial hypertension. J Hypertens 21:1011, 2003

44. Kaplan NM: Other secondary forms of hypertension. Kaplan's Clinical Hypertension, 8th ed. Kaplan NM, Ed. Lippincott Williams & Wilkins, Philadelphia, 2002, p 495

45. Fuchs FD, Maestri MK, Bredemeier M, et al: Study of the usefulness of optic fundi examination of patients with hypertension in a clinical setting. J Hum Hypertens 9:547, 1995

46. Cockcroft DW, Gault MH: Prediction of creatinine clearance from serum creatinine. Nephron 16:31, 1976

47. Levey AS, Bosch JP, Lewis JB, et al: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461, 1999

48. The effects of nonpharmacologic interventions on blood pressure of persons with high normal levels: results of the Trials of Hypertension Prevention, phase I. Trials of Hypertension Prevention Collaborative Research Group. JAMA 267:1213, 1992

49. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in overweight people with high-normal blood pressure: the Trials of Hypertension Prevention, phase II. Trials of Hypertension Prevention Collaborative Research Group. Arch Intern Med 157:657, 1997

50. Xin X, He J, Frontini MG, et al: Effects of alcohol reduction on blood pressure: a meta-analysis of randomized controlled trials. Hypertension 38:1112, 2001

51. Appel LJ, Moore TJ, Obarzanek E, et al: A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N Engl J Med 336:1117, 1997

52. Andersson OK, Almgren T, Persson B, et al: Survival in treated hypertension: follow-up study after two decades. BMJ 317:167, 1998

53. Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 285:2486, 2001

54. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N Engl J Med 344:3, 2001

55. Appel LJ, Champagne CM, Harsha DW, et al: Effects of comprehensive lifestyle modification on blood pressure control: main results of the PREMIER clinical trial. Writing Group of the PREMIER Collaborative Research Group. JAMA 289:2083, 2003

56. Hansson L, Zanchetti A, Carruthers SG, et al: Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomized trial. Lancet 351:1755, 1998

57. He J, Whelton PK, Appel LJ, et al: Long-term effects of weight loss and dietary sodium reduction on incidence of hypertension. Hypertension 35:544, 2000

58. Chobanian AV, Hill M: National Heart, Lung, and Blood Institute Workshop on Sodium and Blood Pressure: a critical review of current scientific evidence. Hypertension 35:858, 2000

59. Kelley GA, Kelley KS: Progressive resistance exercise and resting blood pressure: a meta-analysis of randomized controlled trials. Hypertension 35:838, 2000

60. Whelton SP, Chin A, Xin X, He J: Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. Ann Intern Med 136:493, 2002

61. Schein MH, Gavish B, Herz M, et al: Treating hypertension with a device that slows and regularises breathing: a randomized double-blind controlled study. J Hum Hypertens 15:271, 2001

62. Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. JAMA 283:1967, 2000

63. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. JAMA 288:2981, 2002

64. Psaty BM, Lumley T, Furberg CD, et al: Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. JAMA 289:2534, 2003

65. Wing LM, Reid CM, Ryan P, et al: A comparison of outcomes with angiotensin-converting-enzyme inhibitors and diuretics for hypertension in the elderly. N Engl J Med 348:583, 2003

66. Wang J, Staessen JA: Benefits of antihypertensive pharmacologic therapy and blood pressure reduction in outcome trials. J Clin Hypertens 5:66, 2003

67. Moser M, Pickering T, Sowers JR: Combination drug therapy in the management of hypertension: when, with what, and how? J Clin Hypertens 2:94, 2000

68. He J, Muntner P, Chen J, et al: Factors associated with hypertension control in the general population of the United States. Arch Int Med 162:1051, 2002

69. Hyman DJ, Pavlik VN: Characteristics of patients with uncontrolled hypertension in the United States. N Engl J Med 345:479, 2001

70. Hyman DJ, Pavlik VN: Self-reported hypertension treatment practices among primary care physicians: blood pressure thresholds, drug choices, and the role of guidelines and evidence-based medicine. Arch Intern Med 160:2281, 2000

71. Taler SJ, Textor SC, Augustine JE: Resistant hypertension: comparing hemodynamic management to specialist care. Hypertension 39:982, 2002

72. Redon J, Campos C, Narciso ML, et al: Prognostic value of ambulatory blood pressure monitoring in refractory hypertension: a prospective study. Hypertension 31:712, 1998

73. Messerli FH: Osler's maneuver, pseudohypertension, and true hypertension in the elderly. Am J Med 80:906, 1986

74. Di Bari M, Pahor M, Franse LV, et al: Dementia and disability outcomes in large hypertension trials: lessons learned from the systolic hypertension in the elderly program (SHEP) trial. Am J Epidemiol 153:72, 2001

75. Whelton PK, Appel LJ, Espeland MA, et al: Sodium reduction and weight loss in the treatment of hypertension in older persons: a randomized controlled trial of nonpharmacologic interventions in the elderly (TONE). TONE Collaborative Research Group. JAMA 279:839, 1998

76. Midgley JP, Matthew AG, Greenwood CM, et al: Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. JAMA 275:1590, 1996

77. Hansson L, Lindholm LH, Ekbom T, et al: Randomised trial of old and new antihypertensive drugs in elderly patients: cardiovascular mortality and morbidity in the Swedish Trial in Old Patients with Hypertension–2 Study. Lancet 354:1751, 1999

78. Dahlof B, Devereux RB, Kjeldsen SE, et al: Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. Lancet 359:995, 2002

79. Randomized trial of a perindopril-based blood-pressure lowering regimen among 6,105 individuals with previous stroke or transient ischemic attack. PROGRESS Collaborative Group. Lancet 358:1033, 2001

80. Adler AI, Stratton IM, Neil HA, et al: Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36): prospective observational study. BMJ 321:412, 2000

81. Treatment of hypertension in adults with diabetes. American Diabetes Association. Diabetes Care 26(suppl 1):S80, 2003

82. Efficacy of atenolol and captopril in reducing risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. UK Prospective Diabetes Study Group. BMJ 317:713, 1998

83. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Kidney Disease Outcome Quality Initiative. National Kidney Foundation (NKF) Disease Outcome Quality Initiative (K/DOQI) Advisory Board. Am J Kid Dis 39(2 suppl 2):S1, 2002

84. Lewis EJ, Hunsicker LG, Bain RP, et al: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. N Engl J Med 329:1456, 1993

85. Brenner BM, Cooper ME, de Zeeuw D, et al: Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med 345:861, 2001

86. Estacio RO, Jeffers BW, Hiatt WR, et al: The effect of nisoldipine as compared with enalapril on cardiovascular outcomes in patients with non-insulin-dependent diabetes and hypertension. N Engl J Med 338:645, 1998

87. Vasan RS, Levy D: The role of hypertension in the pathogenesis of heart failure: a clinical mechanistic overview. Arch Intern Med 156:1789, 1996

88. Levy D, Larson MG, Vasan RS, et al: The progression from hypertension to heart failure. JAMA 275:1557, 1996

89. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000

90. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. The SOLVD investigators. N Engl J Med 325:293, 1991

91. Freemantle N, Cleland J, Young P, et al: β Blockade after myocardial infarction: systematic review and meta regression analysis. BMJ 318:1730, 1999

92. Tepper D: Frontiers in congestive heart failure: effect of metoprolol CR/XL in chronic heart failure: Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF). Congest Heart Fail 5:184, 1999

93. Pitt B, Remme W, Zannad F, et al: Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. Eplerenone Post-Aute Myocardial Infarction Heart Failure Efficacy and Survival Study Investigators. N Engl J Med 348:1309, 2003

94. Vakili BA, Okin PM, Devereaux RB: Prognostic implications of left ventricular hypertrophy. Am Heart J 141:334, 2001

95. van Zwieten PA: The influence of antihypertensive drug treatment on the prevention and regression of left ventricular hypertrophy. Cardiovasc Res 45:82, 2000

96. Effect of ramipril vs amlodipine on renal outcomes in hypertensive nephrosclerosis: a randomized controlled trial. African American Study of Kidney Disease and Hypertension (AASK) Study Group. JAMA 285:2719, 2001

97. Morfis L, Schwartz RS, Poulos R, et al: Blood pressure changes in acute cerebral infarction and hemorrhage. Stroke 28:1401, 1997

98. Wilmot M, Leonardi-Bee J, Bath PMW: High blood pressure in acute stroke and subsequent outcome: a systematic review [abstract]. Cerebrovasc Dis 13(suppl 3):95, 2002

99. Semplicini A, Maresca A, Boscolo G, et al: Hypertension in acute ischemic stroke: a compensatory mechanism or an additional damaging factor? Arch Intern Med 163:211, 2003

100. Bath P, Chalmers J, Powers W, et al: International Society of Hypertension (ISH): statement on the management of blood pressure in acute stroke. International Society of Hypertension Writing Group. J Hypertens 21:665, 2003

101. Adams HP Jr, Adams RJ, Brott T, et al: Guidelines for the early management of patients with ischemic stroke: a scientific statement from the Stroke Council of the American Heart Association. Stroke 4:1056, 2003

102. Safian RD, Textor SC: Renal-artery stenosis. N Engl J Med 344:431, 2001

103. Alcazar JM, Rodicio JL: Ischemic nephropathy: clinical characteristics and treatment. Am J Kid Dis 36:883, 2000

104. Fatica RA, Port FK, Young EW: Incidence trends and mortality in end-stage renal disease attributed to renovascular disease in the United States. Am J Kid Dis 37:1184, 2001

105. Young WF Jr: Primary aldosteronism: a common and curable form of hypertension. Cardiol Rev 7:207, 1999

106. Young WF Jr: Pheochromocytoma: issues in diagnosis & treatment. Compr Ther 23:319, 1997

107. Lenders JW, Pacak K, Walther MM, et al: Biochemical diagnosis of pheochromocytoma: which test is best? JAMA 287:1427, 2002

108. Pacak K, Linehan WM, Eisenhofer G, et al: Recent advances in genetics, diagnosis, localization, and treatment of pheochromocytoma. Ann Intern Med 134:315, 2001

109. Newell-Price J, Trainer P, Besser M, et al: The diagnosis and differential diagnosis of Cushing's syndrome and pseudo-Cushing's states. Endocr Rev 19:647, 1998

110. Deaths: Preliminary Data for 2001. National Vital Statistics Reports 51:1, 2003 http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_05.pdf

111. Lock LC: Ocular manifestations of hypertension. Optometry Clin 2:47, 1992

112. Radermacher J, Chavan A, Bleck J, et al: Use of doppler ultrasonography to predict the outcome of therapy for renal-artery stenosis. N Engl J Med 344:410, 2001

113. Vasbinder GB, Nelemans PJ, Kessels, AG, et al: Diagnostic tests for renal artery stenosis in patients suspected of having renovascular hypertension: a meta-analysis. Ann Intern Med 135:401, 2001

114. Carman TL, Olin JW, Czum J: Noninvasive imaging of the renal arteries. Urologic Clin North Am 28:815, 2001

115. Mulatero P, Rabbia F, Milan A, et al: Drug effects on aldosterone/plasma renin activity ratio in primary aldosteronism. Hypertension 40:897, 2002

IV ATRIAL FIBRILLATION

Anthony Aizer, m.d.
Valentin Fuster, m.d., ph.d.

Atrial fibrillation (AF) is a supraventricular tachyarrhythmia defined by rapid, irregular atrial activation. This disordered atrial activation results in loss of coordinated atrial contraction; irregular electrical input to the atrioventricular (AV) node typically leads to sporadic ventricular contractions. On an electrocardiogram, AF is characterized by the absence of visible discrete P waves, the presence of irregular fibrillatory waves, or both, and an irregularly irregular ventricular response [*see Figure 1*].

AF may occur by itself or with other arrhythmias, notably, atrial flutter. Atrial flutter is more organized than AF, involving regular atrial activation that often produces a characteristic sawtooth pattern on ECG. Cardiac rhythm may alternate between AF and atrial flutter, AF may trigger atrial flutter, or atrial flutter may degenerate into AF.

Classification

Numerous classification schemes have been used to characterize AF patients, and the lack of a consistent classification scheme across studies has led to difficulties in comparison of analyses and an inability to extrapolate results to all patients. Consequently, the American College of Cardiology (ACC), the American Heart Association (AHA), and the European Society of Cardiology (ESC), in collaboration with the North American Society of Pacing and Electrophysiology, have established guidelines for the classification of AF.[1] The ACC/AHA/ESC guidelines include the following categories:

- Recurrent—AF occurring in a patient who has experienced an episode of AF in the past.
- Lone—AF occurring in a patient younger than 60 years who has no clinical or echocardiographic evidence of cardiopulmonary disease.
- Valvular or nonvalvular—Valvular AF is AF that occurs in a patient who has evidence or history of rheumatic mitral valve disease or who has a prosthetic heart valve; all other forms of AF are classified as nonvalvular.
- Paroxysmal—AF that typically lasts 7 days or less and that converts spontaneously to sinus rhythm.
- Persistent—AF that typically lasts longer than 7 days or requires pharmacologic or direct current (DC) cardioversion.
- Permanent—AF that is refractory to cardioversion or that has persisted for longer than 1 year.

Paroxysmal, persistent, and permanent AF categories do not apply to episodes of AF lasting 30 seconds or less or to episodes precipitated by a reversible medical condition. Reversible conditions include acute myocardial infarction, cardiac surgery, pericarditis, myocarditis, hyperthyroidism, pulmonary embolism, and acute pulmonary disease.

Epidemiology

AF is the most common sustained arrhythmia, currently affecting more than 2.2 million persons in the United States.[2] The incidence is approximately 0.1% per year for the entire population; however, the incidence of AF increases steadily with age. As a result, one out of 11 Americans older than 80 years has AF.[3-5]

AF is associated with significant morbidity and mortality. The annual incidence of ischemic stroke in patients with AF is 5%, which is two to seven times higher than the incidence in the general population. In addition, the mortality in patients with AF is approximately twice that of patients without AF.[3,6,7] AF frequently leads to reduced functional capacity, dyspnea, palpitations, fatigue, tachycardia-induced cardiomyopathy, heart failure, and angina, significantly impairing quality of life.[8]

Finally, AF results in tremendous health care expenditures. There are more than 370,000 hospital admissions for AF annually.[9] After the first diagnosis of AF, hospitalization costs are typically 35% higher for patients with AF than for age-matched control subjects.[10]

Pathophysiology

Central to the pathophysiology of AF are two factors: the electrical trigger that initiates the arrhythmia and the abnormal myocardial substrate that allows AF to be maintained. A spectrum of triggers is thought to initiate AF, ranging from premature atrial contractions to atrial tachycardias; ultimately, AF may be self triggering.[11-13] Ectopic atrial foci, frequently located in the pulmonary veins, have been shown to trigger AF.

For AF to persist, the atrial tissue must be primed to allow the propagation of multiple wavelets of electrical depolarization throughout the atria.[14] If a wavelet encounters refractory tissue, the wavelet can extinguish, divide into additional wavelets, or change direction. If the underlying atrial substrate leads to the extinction of the wavelets, then AF will not persist. In contrast, if the underlying atrial substrate promotes the generation of additional wavelets or the maintenance of the existing wavelets, then AF will continue. Fibrosis, hypertrophy, and fatty infiltration of atrial tissue likely allow for abnormal atrial electrical conduction and the maintenance of AF wavelets.

Diagnosis

clinical manifestations

AF can result in a wide variety of signs and symptoms. Some patients are asymptomatic, although they may have an irregularly irregular pulse. Other patients experience strokes, palpitations, fatigue, dyspnea, reduced exercise capacity, heart failure, angina, presyncope, or syncope. Additional complications include

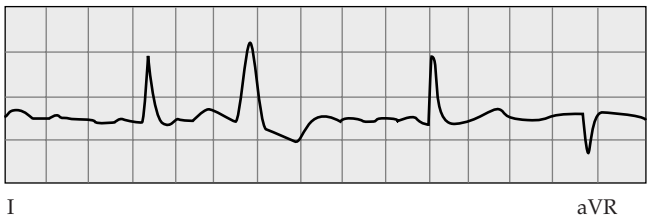


Figure 1 **An electrocardiographic tracing shows characteristic features of atrial fibrillation, with absent P waves, irregular fibrillatory waves, and an irregularly irregular ventricular response.**

Table 1 Initial Clinical Evaluation of Atrial Fibrillation[10]

Evaluation	*Features to Assess*
History and physical examination	Presence, frequency, onset, duration, termination, exacerbating and alleviating factors of AF; date of AF onset; AF classification; associated symptoms; reversible and irreversible contributing conditions; thromboembolic and hemorrhagic risk factors; response to pharmacologic or mechanical interventions
Laboratory studies	Thyroid function,* serum electrolytes, hemoglobin or hematocrit
Chest radiography	Lung parenchyma for intrinsic lung disease; abnormal pulmonary vasculature for pulmonary hypertension; cardiac size and shape for heart failure and pericardial disease
ECG	AF verification; P wave morphology for atrial flutter; preexcitation; atrial arrhythmias besides AF, as possible AF triggers; LVH, for hypertension and hypertrophic cardiomyopathy; bundle branch block and previous MI as markers for CAD, left ventricular dysfunction, and conduction system disease; RR, QRS, and QT intervals to guide antiarrhythmic drug therapy
Transthoracic echocardiography	Left and right atrial size and function; left ventricular size, function, and hypertrophy; valvular heart disease, including rheumatic heart disease; right ventricular systolic pressure for pulmonary hypertension; left atrial thrombus; spontaneous echocardiographic contrast (low sensitivity); pericardial disease; aortic plaque (low sensitivity)

*Reassessment of thyroid function should be considered if ventricular rate becomes difficult to control or atrial fibrillation recurs unexpectedly after conversion to sinus rhythm.
AF—atrial fibrillation CAD—coronary artery disease ECG—electrocardiogram LVH—left ventricular hypertrophy MI—myocardial infarction

thromboembolism and tachycardia-induced cardiomyopathy.[15] The effect of AF on the patient's quality of life is often a critical component that guides decisions regarding AF management.

CLINICAL EVALUATION

The initial evaluation of a patient with AF focuses on the following tasks: (1) confirming the diagnosis of AF, (2) classifying the type of AF, (3) identifying factors (both reversible and irreversible) that contribute to or cause AF, (4) establishing the risk of thromboembolism and additional adverse outcomes, and (5) defining the most effective treatment strategy. In taking the history, the clinician should try to determine whether this is the first episode of AF. If more than one episode of AF has occurred, the AF is defined as recurrent. If no reversible condition is detected in recurrent AF, the clinician may be able to classify the AF as paroxysmal, persistent, or permanent [*see* Classification, *above*].

LABORATORY STUDIES

The standard blood tests that are recommended by the ACC/AHA/ESC are thyroid function tests and measurement of serum electrolytes and hemoglobin or hematocrit. Other recommended laboratory studies include chest radiography, ECG, and transthoracic echocardiography [*see Table 1*]. Additional tests that may be indicated in specific situations are event and Holter monitoring, exercise testing, transesophageal echocardiography (TEE), and electrophysiologic study (EPS).

Event and Holter Monitors

Event monitors are of particular use for documenting infrequent symptomatic episodes in patients in whom AF has not been confirmed previously. In addition to their diagnostic utility for documenting AF, Holter monitors may be used for therapeutic follow-up to evaluate rate control.[16]

Exercise Testing

Exercise testing can confirm the presence of ischemic heart disease and may unmask exercise-mediated AF. In addition, exercise testing can be used to explore the safety of using specific antiarrhythmic medications and to assess rate control.

Transesophageal Echocardiography

TEE is of greatest use in establishing the risk for embolic stroke, most notably in association with cardioversion to sinus rhythm. Risk factors for cardiogenic embolism that are best identified with TEE include the following: left atrial and left atrial appendage thrombus, left atrial and left atrial appendage spontaneous echo contrast (smoke), left atrial appendage flow velocity, and aortic plaque.[17]

Electrophysiologic Study

EPS can define specific forms of AF that are amenable to catheter-based intervention (i.e., radiofrequency ablation). In addition, EPS allows for assessment of the underlying conduction system to determine the etiology of wide-complex tachycardias, whether supraventricular or ventricular in origin.

Management

Treatment of AF includes either restoration and maintenance of sinus rhythm or control of ventricular rate if AF is persistent or if future paroxysmal events are likely to occur. In ad-

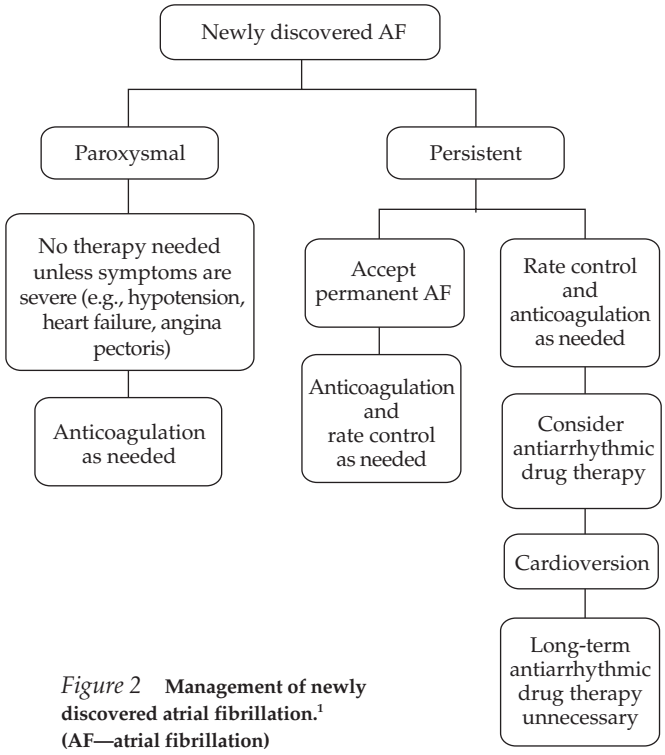


Figure 2 **Management of newly discovered atrial fibrillation.[1] (AF—atrial fibrillation)**

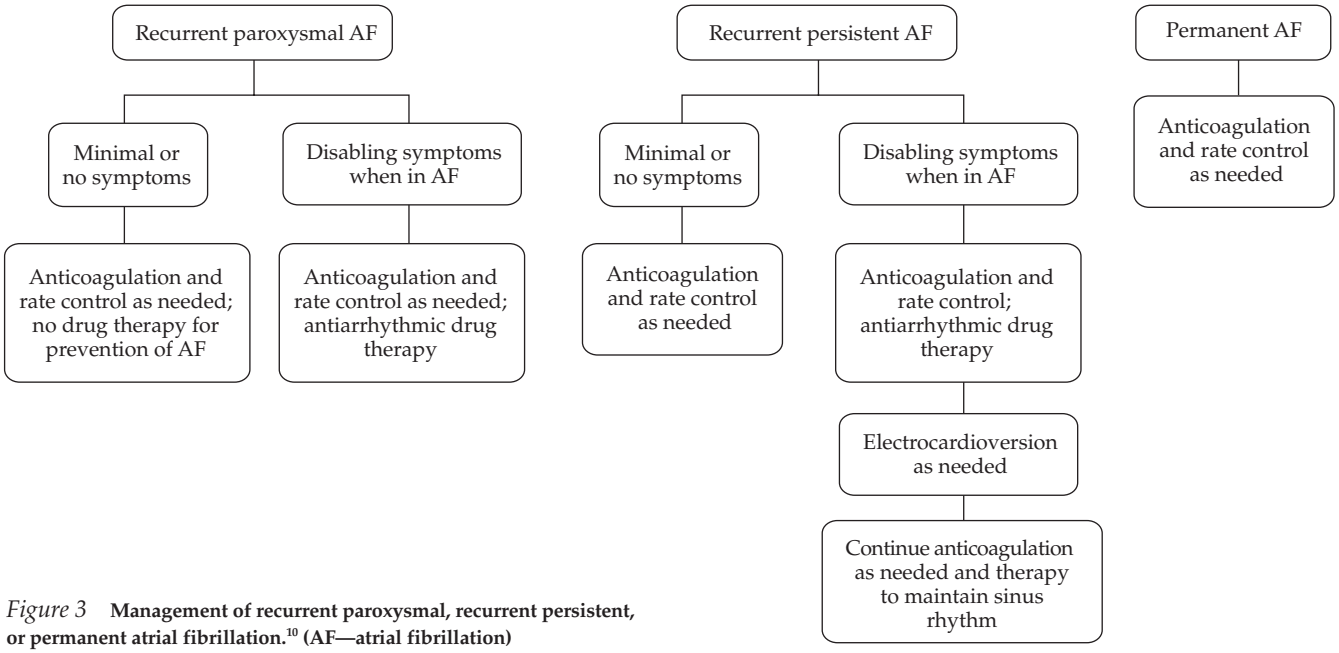


Figure 3 **Management of recurrent paroxysmal, recurrent persistent, or permanent atrial fibrillation.[10] (AF—atrial fibrillation)**

dition, antithrombotics are used to reduce embolic risk [*see Figures 2 through 4*].[10] Treatment decisions involve a synthesis of research results with the characteristics of the individual patient.

Several trials have compared restoration of sinus rhythm with control of ventricular rate in patients with AF. Outcomes evaluated have included overall mortality, stroke, symptoms, and quality of life. Contrary to the expectations of many experts, maintenance of sinus rhythm provided no survival advantage and possibly a higher mortality when compared with ventricular rate control.[18,19] Maintenance of sinus rhythm frequently requires the use of antiarrhythmic medications that may precipitate ventricular arrhythmias, bradycardia, and depression of left ventricular function. It was theorized that maintenance of sinus rhythm would reduce rates of thromboembolism and the need for anticoagulation; however, trial results demonstrated no significant reduction in thromboembolic risk. Peak exercise capacity may improve with maintenance of sinus rhythm, but both treatment strategies result in a similar degree of perceived symptomatic impairment.[8,20,21]

Nevertheless, ventricular rate control frequently is not feasible because of the complications that patients experience while in AF. Clinical scenarios in which AF often is not tolerated include unstable angina, acute myocardial infarction, heart failure, and pulmonary edema. In addition, patients in whom atrial contraction provides a significant proportion of ventricular filling because of impaired ventricular relaxation often need to be maintained in sinus rhythm.

RESTORATION AND MAINTENANCE OF SINUS RHYTHM

Sinus rhythm can be restored with medication, electrical shocks, or a combination of both. Electrical shocks typically are more effective than medication for cardioversion and pose a lower risk of life-threatening ventricular arrhythmias. However, shocks require conscious sedation. In a proportion of patients refractory to medication or electrical shocks, the combination of both therapies results in return of sinus rhythm.

Pharmacologic Cardioversion

Antiarrhythmic medications typically alter the conduction properties of both diseased and normal atrial tissue, suppressing AF triggers or inhibiting the propagation of AF electrical wavelets. Although pharmacologic cardioversion might seem simpler than electrical cardioversion, it has a lower success rate and it poses a risk of life-threatening arrhythmias; the latter risk often precludes use of this strategy. The efficacy of medications for cardioversion of AF typically declines as the duration of AF increases.[22]

A number of medications can be used for cardioversion or for maintenance of sinus rhythm [*see Tables 2 and 3*]. Some medications can be used for both purposes, but others should be used for cardioversion only or for maintenance of sinus rhythm only.

Medication selection for pharmacologic cardioversion must be based on individual patient characteristics. Amiodarone, dofetilide, and ibutilide (agents with potassium channel blocking effects) can be given safely to patients with heart failure or reduced left ventricular systolic function. In contrast, flecainide and propafenone may exacerbate heart failure and should be avoided in such patients. Dofetilide and ibutilide have higher success rates for conversion of atrial flutter than of AF, whereas flecainide and propafenone have higher success rates with conversion of AF than of atrial flutter. Flecainide, propafenone, disopyramide, procainamide, and quinidine also may increase ventricular rate response, especially if patients convert from AF to atrial flutter. Before receiving one of these medications, the patient should be pretreated with an AV nodal blocking agent (typically, diltiazem or verapamil, or possibly digoxin).

Disopyramide, procainamide, and quinidine have either limited efficacy for cardioversion of AF or are associated with significant adverse effects that preclude their use except in rare circumstances. Sotalol effectively maintains sinus rhythm and controls ventricular rate in patients who have undergone cardioconversion from AF, but it has not been shown to effectively convert AF to sinus rhythm. Similarly, beta blockers, verapamil, diltiazem, and digoxin are effective for control of ven-

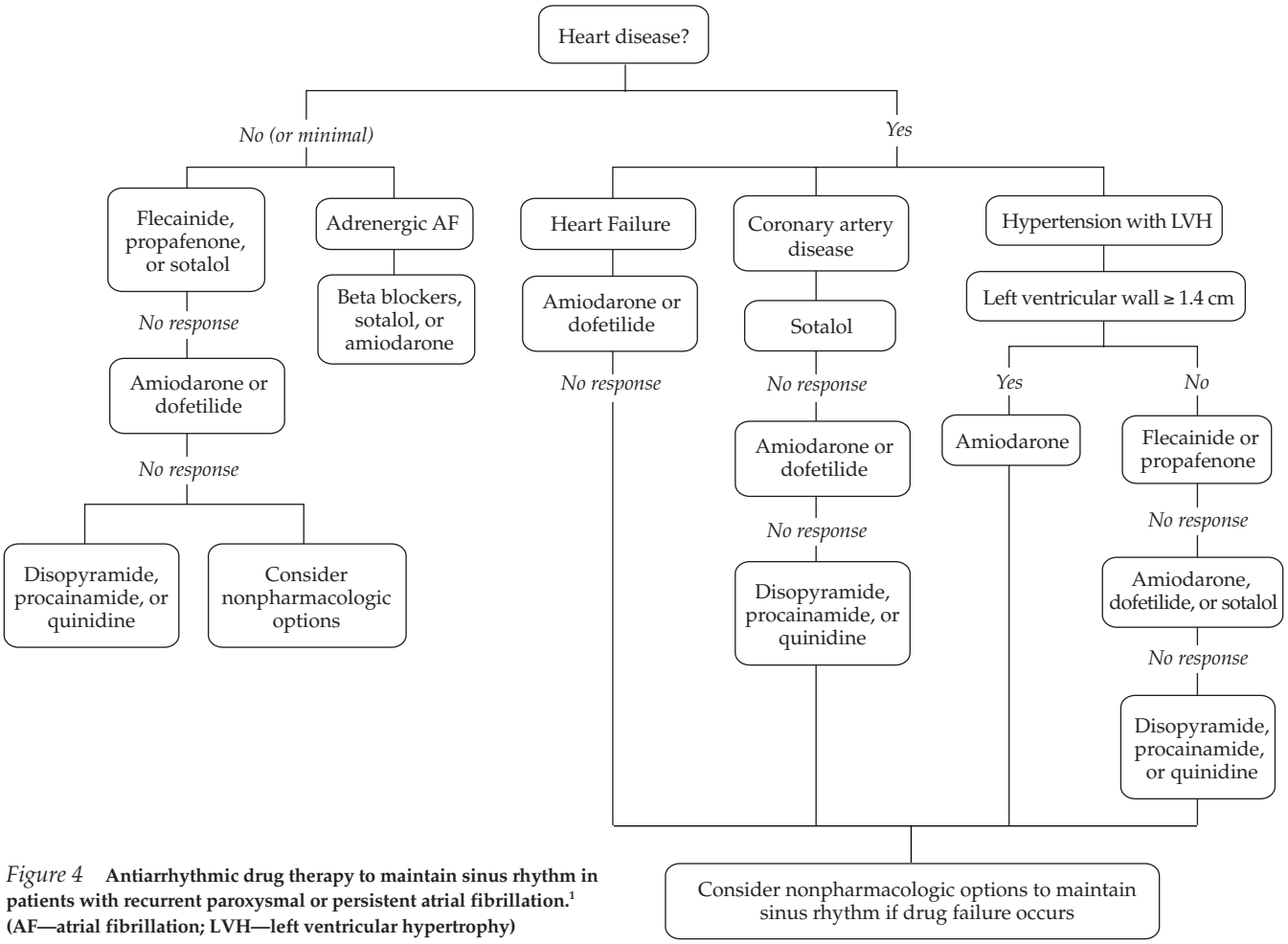


Figure 4 **Antiarrhythmic drug therapy to maintain sinus rhythm in patients with recurrent paroxysmal or persistent atrial fibrillation.[1] (AF—atrial fibrillation; LVH—left ventricular hypertrophy)**

tricular rate in patients with AF, but these medications have little role in AF cardioversion.

Electrical Cardioversion

DC cardioversion is the most effective mechanism for achieving sinus rhythm, with success rates of approximately 70% to 90%.[23,24] DC cardioversion has an even greater rate of success with atrial flutter, approximating 95%.[25] The efficacy of DC cardioversion can be optimized by enhancing delivery of energy to the atrial myocardium. This is achieved through a number of maneuvers:

- Electrode paddle positioning. Anteroposterior positioning is more effective than anterolateral positioning.[26] In addition, applying pressure to the paddles during conversion reduces transthoracic impedance, improving energy conduction.
- Timing of cardioversion. Application of the energy when the patient has fully exhaled reduces pulmonary resistance to the current.[27]
- Use of rectilinear biphasic energy. Traditional energy sources supply monophasic energy. Biphasic energy transfers more efficiently to atrial tissue, leading to higher cardioversion success rates and lower cumulative energy discharge.[28]

Although numerous protocols have been validated, a reasonable protocol that uses monophasic energy to convert AF is to start at 200 joules (J), followed by 300 J, then by 360 J or 400 J.[29] For patients with atrial flutter, cardioversion is frequently achieved with 50 J of monophasic energy; therefore, the monophasic AF protocol can be modified for AF by starting with 50 J, followed by 100 J. If biphasic energy is utilized for AF, a protocol of 70 J or 100 J followed by 150 J and then by 200 J may be utilized.[28,30]

Although success rates are high with DC cardioversion, a number of risk factors for cardioversion failure have been identified. These include longer duration of AF (notably, greater than 1 year), older age, left atrial enlargement, cardiomegaly, rheumatic heart disease, and transthoracic impedance.[25,28] Pretreatment with amiodarone, ibutilide, sotalol, flecainide, propafenone, disopyramide, and quinidine have been shown to increase DC cardioversion success rates.[1] Transvenous cardioversion also may be successfully used for cardioversion for patients in whom transthoracic cardioversion fails.[31,32]

DC cardioversion of AF is extremely safe, typically resulting in no significant myocardial damage if cardioversion attempts are separated by at least 1 minute. Nevertheless, clinicians must give consideration to two types of adverse events[33,34]:

- Reprogramming or malfunction of permanent pacemakers or implantable cardioverter-defibrillators (ICDs). Electricity transmitted from endocardial wires to myocardium can lead to tissue scarring and an increased threshold for tissue capture.[35] In addition, cardioversion energy can erase or alter the programming of permanent pacemakers or ICDs. For that

Table 2 Drugs for Cardioversion of Atrial Fibrillation and Maintenance of Sinus Rhythm[10]

Medication	*Route*	*Time to Conversion*	*Precautions*	*Drug Interactions*	*Side Effects*	*Comments*
Amiodarone	Oral/ I.V.	Hours to weeks	—	Increases digoxin, procainamide, quinidine, and warfarin levels	Bradycardia, visual disturbances, nausea, constipation, phlebitis (I.V. form); hepatic, ocular, pulmonary, thyroid, neurologic toxicity	Safe for use in patients with left ventricular dysfunction; TdP/VT less common than with dofetilide, ibutilide, or sotalol
Dofetilide	Oral	Days to weeks	—	Levels increased by cimetidine and verapamil	—	Safe for use in patients with left ventricular dysfunction; associated with TdP
Ibutilide	I.V.	< 1 hr	Check serum potassium, magnesium levels; requires 4 hr of monitoring for TdP	—	—	Safe for use in patients with left ventricular dysfunction; associated with TdP; not used for maintenance of sinus rhythm
Sotalol	Oral	Incompletely studied; reduced efficacy or no proven efficacy for cardioversion of AF	May exacerbate CHF and/or COPD	—	Bradycardia	Use with caution in patients with reduced left ventricular function; associated with TdP
Flecainide	Oral	3 hr	Pretreat with AV nodal blocking agents* to avoid accelerated ventricular response; avoid in patients with heart failure, left ventricular dysfunction, or CAD	Levels increased by amiodarone	—	
Propafenone	Oral/I.V.	< 6 hr	Pretreat with AV nodal blocking agents* to avoid accelerated ventricular response; avoid in patients with heart failure, left ventricular dysfunction, or CAD; may exacerbate COPD	Increases digoxin and warfarin levels	Blurred vision, hypotension	Efficacy reduced in patients with structural heart disease
Quinidine	Oral/I.V.	2–6 hr	Pretreat with AV nodal blocking agents* to avoid accelerated ventricular response; avoid in patients with heart failure or left ventricular dysfunction	Increases digoxin levels; levels increased by verapamil	Hypotension, nausea, diarrhea, fever, hepatic dysfunction, thrombocytopenia, hemolytic anemia	Safety limits use in cardioversion; side effects limit use; associated with TdP
Disopyramide	Oral/I.V.	< 12 hr	Incompletely studied, reduced efficacy or no proven efficacy for cardioversion of AF; pretreat with AV nodal blocking agents* to avoid accelerated ventricular response; avoid in patients with heart failure or left ventricular dysfunction	—	Dry mucous membranes, constipation, urinary retention; significant reduction of left ventricular function	Side effects limit use; associated with TdP
Procainamide	I.V.	< 24 hr	Incompletely studied, reduced efficacy or no proven efficacy for cardioversion of AF; pretreat with AV nodal blocking agents* to avoid accelerated ventricular response; avoid in patients with heart failure or left ventricular dysfunction	—	Drug-induced lupus, vasculitides, blood dyscrasias, central nervous system disturbances	Reduced efficacy, side effects limit use; associated with TdP

*AV nodal blocking agents typically used are verapamil or diltiazem, and possibly digoxin.
AF—atrial fibrillation CAD—coronary artery disease CHF—chronic heart failure COPD—chronic obstructive pulmonary disease TdP—torsade de pointes VT—ventricular tachycardia

Table 3 Dosages of Drugs for Pharmacologic Cardioversion of Atrial Fibrillation and Maintenance of Sinus Rhythm[10,80]

Drug	*Dosage for Cardioversion*	*Daily Dosage for Maintenance of Sinus Rhythm*
Amiodarone	Oral, inpatient 1.2–1.8 g/day in divided doses until 10 g total, then 200–400 mg/day maintenance; or 30 mg/kg as single dose Oral, outpatient 600–800 mg/day in divided doses until 10 g total Intravenous/oral 5–7 mg/kg over 30–60 min, then 1.2–1.8 g/day continuous I.V. or in divided oral doses until 10 g total	100–400 mg
Dofetilide	Oral dosages for specified C_{Cr} values 500 μg b.i.d. for C_{Cr} > 60 ml/min 250 μg b.i.d. for C_{Cr} 40 to 60 ml/min 125 μg b.i.d. for C_{Cr} 20 to 40 ml/min Contraindicated for C_{Cr} < 20 ml/min	500–1,000 μg; dosage adjustment based on QTc
Ibutilide	I.V.: 1 mg over 10 min; repeat once, if necessary	Not available
Sotalol	Not effective for cardioversion	240–320 mg; dosage adjustment based on QTc; reduced dosing with renal insufficiency
Flecainide	Oral: 200–300 mg	200–300 mg; reduced dosing with renal insufficiency
Propafenone	Oral: 450–600 mg I.V.: 1.5–2.0 mg/kg over 10–20 min; reduced dosing with renal insufficiency	450–900 mg; reduced dosing with hepatic dysfunction
Quinidine	Oral: 0.75–1.5 g in divided doses over 6–12 hr I.V.: 1.5–2.0 mg/kg over 10–20 min	600–1,500 mg
Disopyramide	Oral: 200 mg q. 4 hr, up to 800 mg	400–750 mg; reduced dosing with renal insufficiency
Procainamide	I.V.: 100 mg q. 5 min, up to 1,000 mg	1,000–4,000 mg; reduced dosing with renal insufficiency or hepatic dysfunction

Note: Dosages given may differ from those recommended by the manufacturer; see Table 2 for guidance regarding medication selection and dosing adjustments.
C_{Cr}—creatinine clearance QTc—corrected QT interval

reason, all such devices should be interrogated before and after DC cardioversion. Distancing of paddles from implanted devices may limit these adverse events.

- Arrhythmias. Life-threatening arrhythmias are more common with pharmacologic conversion but can occur with DC cardioversion. Ventricular tachycardia and ventricular fibrillation can result from cardioversion in patients with hypokalemia or digoxin toxicity. Failure to synchronize DC energy with ventricular rhythm can lead to ventricular fibrillation if energy is applied during ventricular repolarization.

Finally, many patients with AF have underlying sinus node dysfunction that may require permanent pacing once cardioversion is completed.[36]

Pharmacologic Approaches to Maintaining Sinus Rhythm

Except for patients in whom the cause of AF is reversible, pharmacologic therapy likely will be required to maintain sinus rhythm after cardioversion. In approximately 50% of AF patients who undergo cardioversion to sinus rhythm, AF will return within 1 year if prophylactic drug therapy is not employed; AF will recur in approximately 75% of patients within 4 years.[24] Before prescribing medication to maintain sinus rhythm, the clinician must assess the patient for underlying cardiovascular disease [*see Table 1*]. The presence of heart failure, coronary artery disease (CAD), or hypertension with left ventricular hypertrophy has a critical impact on the selection of antiarrhythmic medications [*see Figure 4*].

Class I antiarrhythmics frequently suppress left ventricular function. Randomized clinical trials have demonstrated that amiodarone and dofetilide maintain sinus rhythm without reducing survival in AF patients with heart failure.[37-39] As a result, these two drugs have become first-line therapy in this patient subgroup. In patients with ICDs, sotalol may be used safely.[40,41]

Agents with beta-blocking properties are preferred for patients with CAD. Sotalol has the advantage of blocking both beta-adrenergic receptors and potassium channels. In addition, sotalol has been shown to reduce reinfarction rates after a myocardial infarction, and its use has been associated with a trend toward reduced mortality.[42] However, in patients with concomitant heart failure or reduced ventricular function, amiodarone or dofetilide is preferable.

Hypertension and left ventricular hypertrophy may affect drug selection. If the left ventricular wall thickness is 14 mm or greater, amiodarone is recommended.

Although these recommendations can be applied to the majority of patients with AF, a number of distinct clinical scenarios require a tailored approach. In patients who do not have structural heart disease but who experience AF during exercise or under adrenergic stimulation, beta blockers are the treatment of choice, followed by sotalol or amiodarone. Vagally mediated AF that is not associated with structural heart disease often re-

sponds to disopyramide, a vagolytic medication. Second-line therapy includes flecainide and amiodarone.

Combination therapy may be used when a single medication fails to maintain sinus rhythm. With the combination of medications comes the increased risk of drug-induced side effects, notably, torsade de pointes and heart failure. Monitoring of symptoms and the width of the QTc and QRS intervals is critical.

Monitoring of antiarrhythmic therapy ECG monitoring is necessary in all patients receiving antiarrhythmic medications for maintenance of sinus rhythm. If flecainide or propafenone is used, QRS widening should not exceed 150% of pretreatment QRS width. QRS width should be assessed during exercise ECG testing, typically within 3 days after starting the medication. With all antiarrhythmics except amiodarone, QTc width should not exceed 520 msec. In addition, renal function and levels of serum potassium and serum magnesium should be monitored periodically, because abnormalities in these levels may predispose to arrhythmias.

Outpatient Initiation of Antiarrhythmic Drugs

In a subset of patients with AF, drugs for restoration and maintenance of sinus rhythm can be started safely in the outpatient setting. Advantages of this approach are elimination of the need for DC cardioversion, reduction of hospitalization time, and a decrease in early recurrences of AF after conversion to sinus rhythm. Although outpatient pharmacologic therapy to restore sinus rhythm is appealing, the concern for induction of life-threatening arrhythmias often precludes use of this approach.

Flecainide and propafenone may be initiated on an outpatient basis if the patient has no history of heart failure; if there is no left ventricular dysfunction; if the QRS width is normal; and if the QTc interval is not prolonged. Patients should have both a normal ECG (without any evidence of bradycardia, sinus node disease, or AV nodal disease) and a documented history of at least one episode of inpatient cardioversion with these medications during which no conduction abnormality was unmasked. Amiodarone and sotalol may be started in the ambulatory setting, provided there is no history of structural heart disease, left ventricular hypertrophy, reduced left ventricular function, bradycardia, sinus node or AV nodal conduction disease, hypokalemia, hypomagnesemia, or previous arrhythmias other than AF or atrial flutter. Flecainide, propafenone, amiodarone, or sotalol should not be started if the patient is also taking other medications that may prolong the QTc interval or predispose to electrolyte abnormalities. Dofetilide, disopyramide, procainamide, and quinidine typically should not be started in the ambulatory setting.[1]

Nonpharmacologic Approaches to Maintaining Sinus Rhythm

Several mechanical techniques offer the benefit of reducing the use of antiarrhythmics. The need for anticoagulation with these techniques remains uncertain, however.

Catheter-based ablation Radiofrequency energy emitted from intravascular catheters promotes the generation of endocardial scars to eliminate AF. These procedures focus primarily on elimination or isolation of ectopic foci, many of which are located in the pulmonary veins. Although these procedures have the potential to cure AF, many patients experience recurrence of AF. The risks of catheter-based ablation include thromboembolism, pulmonary vein stenosis, and cardiac perforation.[43]

Endovascular radiofrequency ablation is less suited to AF than to atrial flutter, which it can cure with minimal risks and a high rate of success. Ablation of atrial flutter typically involves creating a scar within the right atrium and therefore has a lower risk of complications than AF ablation of the pulmonary veins. Radiofrequency ablation for atrial flutter is curative in more than 90% of cases and should be considered primary therapy for these patients.[44]

Surgical ablation Surgical ablation of AF is similar in concept to catheter-based ablation. During open thoracotomy, linear lesions are created across atrial tissue to generate scars that will act as electromechanical obstacles, extinguishing the reentrant circuits needed for the maintenance of AF. There is a greater than 90% rate of success in eliminating AF with this procedure; however, approximately 25% of patients require a permanent pacemaker for sinus node dysfunction postoperatively.[45-47] This approach has an operative mortality of less than 1% but involves the morbidity of an invasive surgical procedure. The procedure is most often utilized when patients are undergoing cardiac surgery for other indications. The techniques utilized to generate the scars, as well as the location and number of scars created, continue to be modified to reduce surgical time while maintaining efficacy.

Atrial pacing In patients requiring ventricular pacing, the addition of atrial pacing reduces the risk of AF. However, the use of atrial pacing as the primary treatment to prevent AF has not been validated.[48-50]

Atrial defibrillators Implantable devices to detect and provide DC cardioversion for AF have been shown to successfully terminate AF in more than 95% of episodes.[51] Although promising, the use of atrial defibrillators is limited by the generation of pain associated with the release of the electrical shock, as well as the risks associated with device implantation (typically, bleeding and infection). As a result, atrial defibrillators have been used in patients who are unable to tolerate a strategy of ventricular rate control and whose condition is refractory to pharmacologic and ablative therapies.

CONTROL OF VENTRICULAR RATE

Ventricular rate control must be addressed both in the acute and the chronic setting. Medication selection in these scenarios is influenced by the rate of onset of the medication, its potential side effects, and its convenience of use.

Hemodynamically unstable patients with angina, myocardial infarction, heart failure, or symptomatic hypotension should be considered for acute conversion to sinus rhythm rather than rate control. In contrast, acute rate control can often be achieved rapidly in hemodynamically stable patients through the use of intravenous beta blockers, diltiazem, verapamil, or digoxin. Oral formulations of these medications are utilized for transition to long-term rate control [*see Table 4*]. More than one medication is often required to achieve ventricular rate control. Although digoxin is available orally and intravenously, its onset of action is at least 1 hour after infusion, so it is rarely sufficient for stand-alone therapy in the acute clinical setting.

Depending on the clinical scenario, specific agents may be more or less preferable for rate control. This is true of patients with reduced ventricular function, CAD, high sympathetic tone, pulmonary disease, and atrial flutter.

Table 4 Drugs for Ventricular Rate Control in Atrial Fibrillation[10]

Drug	*I.V. Loading Dose*	*I.V. Onset*	*I.V. Maintenance Dose*	*Oral Loading Dose*	*Oral Onset*	*Oral Maintenance Dose*	*Drug Interactions and Precautions*
Esmolol*	0.5 mg/kg over 1 min	5 min	5–20 μg/kg/min	Available in I.V. form only	—	—	—
Metoprolol*	2.5–5 mg over 2 min, up to 15 mg	5 min	Bolus every 4–6 hr	Not applicable	4–6 hr	50–200 mg daily in divided doses	—
Propranolol*	0.15 mg/kg over 1 min, repeat once	5 min	Bolus every 4 hr	Not applicable	1–1.5 hr	80–240 mg daily in divided doses	—
Diltiazem	0.25 mg/kg over 2 min	2–7 min	5–15 mg/hr	Not applicable	2–4 hr	120–360 mg daily in divided doses	Increases levels of digoxin, quinidine, simvastatin
Verapamil	75–150 μg/kg over 2 min	3–5 min	Bolus q. 3–6 hr	Not applicable	1–2 hr	120–360 mg daily in divided doses	Increases levels of digoxin, dofetilide, quinidine, simvastatin
Digoxin	0.25 mg q. 2 hr, up to 1.5 mg	2 hr	0.125–0.25 mg daily	0.25 mg q. 2 hr, up to 1.5 mg	2 hr	0.125–0.250 mg/day	Reduce dosing with renal insufficiency; levels increased by amiodarone, propafenone, quinidine, diltiazem, verapamil, spironolactone
Amiodarone	1.2–1.8 g/day until 10 g total	1–3 wk	720 mg/day up to 3 wk; limited data on continuous infusion beyond 3 wk	800 mg/day × 1 wk, 600 mg/day × 1 wk, 400 mg/day × 4–6 wk	1–3 wk	200 mg/day	Increases levels of digoxin, procainamide, quinidine, and warfarin

Note: Typical dosing regimens are provided; however, adjustments are necessary based on individual patient characteristics.
*Other beta-blocking medications may also be used.

Reduced Ventricular Function

Diltiazem and verapamil can significantly exacerbate left ventricular dysfunction and associated heart failure and so should be avoided in the acute setting. Beta blockers can also have this effect but are preferable for acute rate control. Intravenous esmolol has the advantage of rapid onset and clearance and so may be used to determine whether a patient with left ventricular dysfunction tolerates intravenous beta blockade. However, the large infusion of saline given with esmolol makes long-term intravenous use unattractive for patients with heart failure. If the patient tolerates intravenous esmolol, the clinician should consider changing to another intravenous beta blocker or to oral beta blockade. In addition, digoxin can be utilized in patients with left ventricular dysfunction without concern for exacerbating heart failure. Intravenous amiodarone may also be used in the subacute setting for rate control of patients with AF and reduced ventricular function.

Chronic rate control can be achieved through the oral administration of beta blockers. Bisoprolol, extended-release metoprolol, and carvedilol improve symptoms and survival in patients with systolic dysfunction and heart failure independent of atrial rhythm.[52-54] These medications should be first-line therapy for long-term rate control in these patients. If these medications are not tolerated, oral amiodarone should be considered. In addition, digoxin is effective and well tolerated in heart failure patients with AF.

Coronary Artery Disease

Beta blockers have been shown to reduce mortality in patients with CAD. Because of this additive benefit, beta blockers typically should be selected for CAD patients.

High Sympathetic Tone

The effects of digoxin are attenuated in patients with high sympathetic tone, so this agent rarely provides significant control of heart rate in acute, high sympathetic tone states.

Pulmonary Disease

Patients with asthma can experience significant exacerbation of their lung disease with the use of beta blockers. In these patients, diltiazem and verapamil should be used. Patients with chronic obstructive pulmonary disease without reactive airway disease may or may not tolerate beta blockers. Use of beta blockers in this population should be carefully monitored.

Atrial Flutter

It is often more difficult to achieve ventricular rate control in patients with atrial flutter than in those with AF. If rate control cannot be achieved easily in patients with atrial flutter, radiofrequency ablation should be reconsidered.

Monitoring Rate Control

Adequacy of rate control should be assessed both with the patient at rest and under stress. The history, physical examination, and ECG provide significant data for this assessment, but Holter monitoring and exercise stress testing also can be used. The ventricular rate should be maintained between 60 and 80 beats/min during rest and 90 to 115 beats/min during moderate exercise.[55,56] If rate control cannot be achieved with pharmacologic therapy, AV nodal ablation, combined with permanent pacemaker insertion, should be considered. In addition, permanent pacemaker insertion may be necessary for patients with AF who have labile responses to pharmacologic therapy to avoid episodes of symptomatic bradycardia.

Table 5 Data Collection for Assessment of Thromboembolic Risks and Need for Antithrombotic Therapy in Atrial Fibrillation

Characteristic	*Comments*
Age	
Sex	
History of hypertension	Patients with medically treated hypertension are considered hypertensive for risk-stratification guidelines
Diabetes mellitus	Irrespective of control with insulin or oral medications
Coronary artery disease	
Heart failure	Past or current
Hyperthyroidism	Treatment varies depending on whether currently euthyroid
Rheumatic heart disease	Defined as involving the mitral valve
Previous thromboembolism	Includes strokes, transient ischemic attacks, and other emboli
Prosthetic heart valves	
LVEF less than 35%	

LVEF–left ventricular ejection fraction

ANTITHROMBOTIC THERAPY

AF (including paroxysmal, permanent, and chronic forms) is associated with an increased risk of stroke and other embolic phenomena. The risk of stroke for an individual AF patient varies according to the presence or absence of a number of thromboembolic risk factors. These factors can be garnered from the baseline history, physical examination, laboratory evaluation, ECG, and transthoracic echocardiogram; assessment of these thromboembolic risk factors can serve to guide antithrombotic therapy [*see Table 5*].

Current ACC/AHA/ESC guidelines for anticoagulation recommend the use of aspirin or warfarin [*see Table 6*]. Clinical trials have shown that both aspirin and warfarin significantly reduce AF-related strokes in high-risk patients.[57-59] Warfarin reduces the risk of stroke by greater than 60%, whereas aspirin reduces stroke risk by 19%. However, the increased benefits of warfarin must be counterbalanced by the increased risk of hemorrhage.[60] Use of lower-intensity warfarin in combination with aspirin provides no additional stroke prevention over aspirin alone, and the combination of full-dose warfarin with aspirin further increases the risk of intracranial hemorrhage.[61,62] After warfarin therapy is started, the international normalized ratio (INR) of prothrombin time should be measured at least weekly until stable dosing is reached, and monthly thereafter [*see 1:XVIII Venous Thromboembolism*].

Atrial Flutter

Although clinical trial data are limited, epidemiologic studies demonstrate that the risk of stroke with atrial flutter, although less than that with AF, remains elevated.[63] As a result, use of warfarin and aspirin in atrial flutter should be based on the current AF guidelines.

Elderly Patients

Patients who are 75 years of age or older are at increased risk for both stroke with AF and bleeding with AF anticoagulation.[64] As a result of these increased risks, anticoagulation must be tightly monitored in elderly patients, with a goal of maintaining the INR at 2.

Surgical Procedures

Anticoagulation may need to be discontinued in patients scheduled for elective surgical procedures. AF anticoagulation can be discontinued for up to 1 week for surgical procedures in patients without mechanical heart valves. In patients with mechanical valves, the practice has been to discontinue warfarin 1 week before surgery but to maintain anticoagulation with either unfractionated or low-molecular-weight heparin (LMWH). However, current case reports suggest that LMWH may not provide sufficient anticoagulation for patients with mechanical valves, irrespective of concomitant AF.[65] Until further data become available, intravenous unfractionated heparin should be utilized.[66]

Anticoagulation and Cardioversion

Cardioversion from AF or atrial flutter to sinus rhythm—whether it occurs spontaneously or is accomplished with drugs or electricity—is associated with a 1% to 5% risk of thromboembolism. Therefore, strategies for cardioversion of AF should include consideration of anticoagulation; the anticoagulation may start before cardioversion, extend after it, or both [*see Figure 5*].

If warfarin anticoagulation (to an INR of 2 to 3) is used for 3 to 4 weeks before and after cardioversion, the risk of stroke is reduced to 0.5% in the immediate follow-up period.[36,67,68] For that reason, anticoagulation before cardioversion has been strongly advocated.

TEE has been validated as an alternative mechanism to gauge the risk of thromboembolism at the time of cardioversion and immediately afterward. If TEE reveals no evidence of thrombus in the left atrium or left atrial appendage, cardioversion can be performed immediately, with a risk of thromboembolism comparable to that in patients pretreated with 3 to 4

Table 6 ACC/AHA/ESC Recommendations for Antithrombotic Therapy in Atrial Fibrillation Based on Underlying Risk Factors[10]

Patient Characteristics	*Antithrombotic Therapy*
Age < 60 yr, no heart disease (lone atrial fibrillation)	Aspirin, 325 mg daily, or no therapy
Age < 60 yr, heart disease but no risk factors	Aspirin, 325 mg daily
Age ≥ 60 yr but no risk factors	Aspirin, 325 mg daily
Age ≥ 60 yr with DM or CAD	Warfarin (INR, 2.0–3.0); consider addition of aspirin, 81–162 mg daily
Age ≥ 75 yr, especially in women	Warfarin (INR, 2.0)
Heart failure	Warfarin (INR, 2.0)
LVEF ≤ 0.35	Warfarin (INR, 2.0–3.0)
Thyrotoxicosis	Warfarin (INR, 2.0–3.0)
Hypertension	Warfarin (INR, 2.0–3.0)
Rheumatic heart disease (mitral stenosis)	Warfarin (INR, 2.5–3.5 or possibly higher)
Prosthetic heart valves	Warfarin (INR, 2.5–3.5 or possibly higher)
Prior thromboembolism	Warfarin (INR, 2.5–3.5 or possibly higher)
Persistent atrial thrombus on TEE	Warfarin (INR, 2.5–3.5 or possibly higher)

ACC/AHA/ESC—American College of Cardiology/American Heart Association/European Society of Cardiology CAD–coronary artery disease DM—diabetes mellitus INR—international normalized ratio LVEF–left ventricular ejection fraction TEE–transesophageal echocardiography

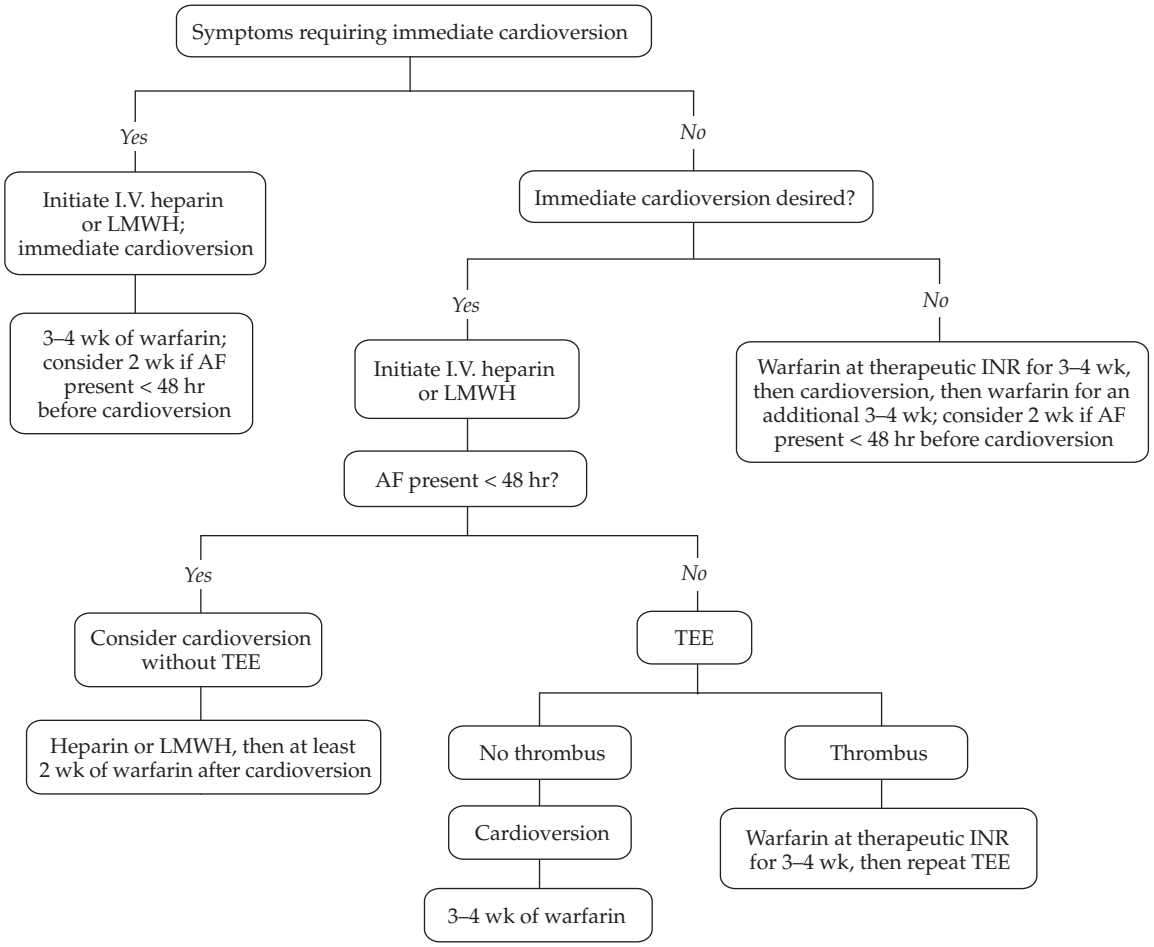


Figure 5 **Cardioversion and anticoagulation strategy for atrial fibrillation. Symptoms that frequently require cardioversion include hypotension, altered mental status, heart failure, pulmonary edema, angina, and myocardial infarction. Adjustment of warfarin intensity and therapy duration is based on individual patient characteristics; the anticoagulation goal is typically an INR of 2–3. (AF—atrial fibrillation; INR—international normalized ratio; LMWH—low-molecular-weight heparin; TEE—transesophageal echocardiography)**

weeks of warfarin therapy.[69] This approach allows for immediate cardioversion; however, because cardioversion frequently results in so-called stunning of left atrial and left atrial appendage tissue (a condition that may predispose to thrombus formation), warfarin anticoagulation is required for 3 to 4 weeks after cardioversion, even when TEE performed before cardioversion showed no thrombus. If TEE does identify thrombus, cardioversion should be postponed for 3 to 4 weeks of anticoagulation therapy with warfarin, after which TEE should be repeated.

Cardioversion without 3 to 4 weeks of warfarin pretreatment and without TEE assessment can be considered if the cardioversion can be done within 48 hours of the onset of AF or if the patient is started on heparin within 48 hours of AF initiation. Limited data suggest that LMWH may be used instead of intravenous unfractionated heparin, allowing both simplified dosing and transition to warfarin therapy on an outpatient basis.[70] This strategy should be most strongly considered in AF patients with significant symptoms of cardiac compromise, including hemodynamic instability, angina, myocardial infarction, heart failure, and shock. The need for anticoagulation after cardioversion in this scenario is unclear, but considering that more than 95% of postcardioversion thromboemboli occur within 10 days after cardioversion, at least 2 weeks of warfarin therapy should be strongly considered if the patient has no contraindications.[71]

Even if heparin was not started until more than 48 hours after the onset of AF, immediate cardioversion also may be necessary if the patient has symptoms of cardiac compromise. Unlike patients who present less than 48 hours after onset of AF, patients with AF of longer duration should receive 3 to 4 weeks of warfarin therapy after cardioversion.

Prolonged anticoagulation after cardioversion should be considered in patients at high risk for both AF recurrence and thromboembolic complications. Atrial flutter is associated with a risk of thromboembolism in the setting of elective cardioversion and should be treated in the same manner as AF.[67]

Ximelagatran

Ximelagatran is a direct thrombin inhibitor that can be administered orally. In the Sport Prevention Using Oral Thrombin Inhibitor in Atrial Fibrillation–III (SPORTIF III) trial, which was an open-label comparison of adjusted-dose warfarin and fixed-dose ximelagatran, there were no significant differences in rates of stroke, systemic thromboembolism, bleeding, or death with

the two drugs.[72] The SPORTIF V trial will compare the two drugs in a double-blinded format. Ximelagatran offers a wider therapeutic window than warfarin and requires no monitoring with coagulation studies. If the equivalence of ximelagatran with warfarin is confirmed, the ease of use of this medication will likely lead to its replacing warfarin for many patients with AF. Future studies will be required to determine whether ximelagatran can be applied to specific settings in which warfarin therapy has been validated.

TREATMENT IN SPECIFIC CLINICAL SCENARIOS

Cardiac Surgery

AF occurs after 25% of all coronary artery bypass surgeries and after more than 60% of combined coronary artery bypass and mitral valve surgeries.[73] Additional risk factors in these cases included advanced age, male sex, preoperative atrial arrhythmias, left atrial enlargement, chronic lung disease, and previous cardiac surgery.[74] AF after cardiac surgery leads to a significant increase in length of hospital stay and cost.[75] A number of prophylactic therapies to prevent postoperative AF have been examined and validated, including use of beta blockers, sotalol, amiodarone, and postoperative temporary atrial pacing.[76] The incremental cost of prophylactic therapy must be balanced against the potential savings achieved by reducing length of stay if AF is prevented. Unless contraindicated, beta blockers should be given to all patients scheduled for cardiac surgery. Sotalol, amiodarone, and biatrial pacing should be considered if patients are at high risk for postoperative AF because of additional risk factors.

Anticoagulation should be given if AF occurs after cardiac surgery and lasts longer than 48 hours. Although sinus rhythm returns spontaneously within 6 weeks in 95% of patients with postoperative AF, pharmacologic or DC cardioversion is often performed, particularly in patients who are symptomatic or hemodynamically unstable.[77] Medications to maintain sinus rhythm or to achieve ventricular rate control can be selected on the basis of patient characteristics.

Acute Myocardial Infarction

In patients with acute myocardial infarction, AF is an independent predictor of mortality and stroke. Immediate DC cardioversion should be performed in patients with severe hemodynamic compromise or persistent ischemia. If rate control is possible, digoxin can be combined with a beta blocker if left ventricular function is preserved. Because of the thromboembolic risk, heparin should be given acutely and followed with warfarin if AF persists or significant left ventricular dysfunction develops.

Wolff-Parkinson-White Syndrome

Wolff-Parkinson-White syndrome (WPW) in association with AF can be a life-threatening condition. The bypass tract of WPW may allow rapid conduction of atrial activity to the ventricles, precipitating hemodynamic compromise or ventricular fibrillation. In the acute setting, DC cardioversion should be pursued if hemodynamic compromise is present. If the patient is hemodynamically stable, the clinician may consider pharmacologic cardioversion to sinus rhythm with intravenous procainamide or ibutilide.[78] Agents that slow AV conduction are contraindicated, including digoxin, diltiazem, and verapamil. Beta blockers should be used rarely and with extreme caution. Once stabilization is achieved, catheter ablation of the WPW bypass tract should be pursued in all symptomatic WPW patients with AF.

Hyperthyroidism

Hyperthyroidism may cause AF and is associated with an increased risk of stroke. Hence, these patients require anticoagulation. Rate control should be attempted with beta blockers, supplemented with diltiazem, verapamil, or digoxin as needed. Warfarin should be given while the patient is thyrotoxic. Once the euthyroid state has returned, use of aspirin or warfarin should be based on underlying risk factors.

Hypertrophic Cardiomyopathy

AF in patients with hypertrophic cardiomyopathy is associated with a high risk of death and stroke.[79] Warfarin therapy is recommended (INR, 2 to 3).

Pulmonary Disease

In patients with pulmonary disease, hypoxia and other metabolic disturbances frequently initiate AF. Initial therapy focuses on treating the underlying lung disease. Beta blockers, propafenone, sotalol, and adenosine are contraindicated in patients with reactive airway disease. Diltiazem or verapamil, with or without digoxin, should be utilized for rate control in these patients.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Fuster V, Ryden LE, Asinger RW, et al: ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients With Atrial Fibrillation): developed in collaboration with the North American Society of Pacing and Electrophysiology. J Am Coll Cardiol 38:1231, 2001

2. Atrial fibrillation as a contributing cause of death and Medicare hositalization—United States, 1999. MMWR Morb Mortal Wkly Rep 52:128, 2003

3. Krahn AD, Manfreda J, Tate RB, et al: The natural history of atrial fibrillation: incidence, risk factors, and prognosis in the Manitoba Follow-up Study. Am J Med 98:476, 1995

4. Psaty BM, Manolio TA, Kuller LH, et al: Incidence of and risk factors for atrial fibrillation in older adults. Circulation 96:2455, 1997

5. Wolf PA, Abbott RD, Kannel WB: Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 22:983, 1991

6. Kannel WB, Abbott RD, Savage DD, et al: Coronary heart disease and atrial fibrillation: the Framingham Study. Am Heart J 106:389, 1983

7. Flegel KM, Shipley MJ, Rose G: Risk of stroke in non-rheumatic atrial fibrillation. Lancet 1:526, 1987

8. Gosselink AT, Crijns HJ, van den Berg MP, et al: Functional capacity before and after cardioversion of atrial fibrillation: a controlled study. Br Heart J 72:161, 1994

9. Wattigney WA, Mensah GA, Croft JB: Increasing trends in hospitalization for atrial fibrillation in the United States, 1985 through 1999: implications for primary prevention. Circulation 108:711, 2003

10. Wolf PA, Mitchell JB, Baker CS, et al: Impact of atrial fibrillation on mortality, stroke, and medical costs. Arch Intern Med 158:229, 1998

11. Haissaguerre M, Jais P, Shah DC, et al: Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. N Engl J Med 339:659, 1998

12. Franz MR, Karasik PL, Li C, et al: Electrical remodeling of the human atrium: similar effects in patients with chronic atrial fibrillation and atrial flutter. J Am Coll Cardiol 30:1785, 1997

13. Bennett MA, Pentecost BL: The pattern of onset and spontaneous cessation of atrial fibrillation in man. Circulation 41:981, 1970

14. Moe G: On the multiple wavelet hypothesis of atrial fibrillation. Arch Int Pharmacodyn Ther 140:183, 1962

15. Grogan M, Smith HC, Gersh BJ, et al: Left ventricular dysfunction due to atrial fibrillation in patients initially believed to have idiopathic dilated cardiomyopathy. Am J Cardiol 69:1570, 1992

16. Gillis AM, Klein GJ, MacDonald RG: Investigation of the patient with atrial fibrillation. Can J Cardiol 12(suppl A):12A, 1996

17. Zabalgoitia M, Halperin JL, Pearce LA, et al: Transesophageal echocardiographic correlates of clinical risk of thromboembolism in nonvalvular atrial fibrillation. Stroke Prevention in Atrial Fibrillation III Investigators. J Am Coll Cardiol 31:1622, 1998

18. Wyse DG, Waldo AL, DiMarco JP, et al: A comparison of rate control and rhythm control in patients with atrial fibrillation. N Engl J Med 347:1825, 2002

19. Van Gelder IC, Hagens VE, Bosker HA, et al: A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. N Engl J Med 347:1834, 2002

20. Atwood JE, Myers J, Sullivan M, et al: The effect of cardioversion on maximal exercise capacity in patients with chronic atrial fibrillation. Am Heart J 118:913, 1989

21. Gosselink AT, Bijlsma EB, Landsman ML, et al: Long-term effect of cardioversion on peak oxygen consumption in chronic atrial fibrillation: a 2-year follow-up. Eur Heart J 15:1368, 1994

22. Suttorp MJ, Kingma JH, Lie AHL, et al: Intravenous flecainide versus verapamil for acute conversion of paroxysmal atrial fibrillation or flutter to sinus rhythm. Am J Cardiol 63:693, 1989

23. Lundstrom T, Ryden L: Chronic atrial fibrillation: long-term results of direct current conversion. Acta Med Scand 223:53, 1988

24. Van Gelder IC, Crijns HJ, Tieleman RG, et al: Chronic atrial fibrillation: success of serial cardioversion therapy and safety of oral anticoagulation. Arch Intern Med 156:2585, 1996

25. Van Gelder IC, Crijns HJ, Van Gilst WH, et al: Prediction of uneventful cardioversion and maintenance of sinus rhythm from direct-current electrical cardioversion of chronic atrial fibrillation and flutter. Am J Cardiol 68:41, 1991

26. Botto GL, Politi A, Bonini W, et al: External cardioversion of atrial fibrillation: role of paddle position on technical efficacy and energy requirements. Heart 82:726, 1999

27. Ewy GA, Hellman DA, McClung S, et al: Influence of ventilation phase on transthoracic impedance and defibrillation effectiveness. Crit Care Med 8:164, 1980

28. Mittal S, Ayati S, Stein KM, et al: Transthoracic cardioversion of atrial fibrillation: comparison of rectilinear biphasic versus damped sine wave monophasic shocks. Circulation 101:1282, 2000

29. Joglar JA, Hamdan MH, Ramaswamy K, et al: Initial energy for elective external cardioversion of persistent atrial fibrillation. Am J Cardiol 86:348, 2000

30. Page RL, Kerber RE, Russell JK, et al: Biphasic versus monophasic shock waveform for conversion of atrial fibrillation: the results of an international randomized, double-blind multicenter trial. J Am Coll Cardiol 39:1956, 2002

31. Schmitt C, Alt E, Plewan A, et al: Low energy intracardiac cardioversion after failed conventional external cardioversion of atrial fibrillation. J Am Coll Cardiol 28:994, 1996

32. Levy S, Lauribe P, Dolla E, et al: A randomized comparison of external and internal cardioversion of chronic atrial fibrillation. Circulation 86:1415, 1992

33. Lund M, French JK, Johnson RN, et al: Serum troponins T and I after elective cardioversion. Eur Heart J 21:245, 2000

34. Jakobsson J, Odmansson I, Nordlander R: Enzyme release after elective cardioversion. Eur Heart J 11:749, 1990

35. Levine PA, Barold SS, Fletcher RD, et al: Adverse acute and chronic effects of electrical defibrillation and cardioversion on implanted unipolar cardiac pacing systems. J Am Coll Cardiol 1:1413, 1983

36. Mancini GB, Goldberger AL: Cardioversion of atrial fibrillation: consideration of embolization, anticoagulation, prophylactic pacemaker, and long-term success. Am Heart J 104:617, 1982

37. Roy D, Talajic M, Dorian P, et al: Amiodarone to prevent recurrence of atrial fibrillation. Canadian Trial of Atrial Fibrillation Investigators. N Engl J Med 342:913, 2000

38. Singh S, Zoble RG, Yellen L, et al: Efficacy and safety of oral dofetilide in converting to and maintaining sinus rhythm in patients with chronic atrial fibrillation or atrial flutter: the symptomatic atrial fibrillation investigative research on dofetilide (SAFIRE-D) study. Circulation 102:2385, 2000

39. Torp-Pedersen C, Moller M, Bloch-Thomsen PE, et al: Dofetilide in patients with congestive heart failure and left ventricular dysfunction. Danish Investigations of Arrhythmia and Mortality on Dofetilide Study Group. N Engl J Med 341:857, 1999

40. Pinto JV Jr, Ramani K, Neelagaru S, et al: Amiodarone therapy in chronic heart failure and myocardial infarction: a review of the mortality trials with special attention to STAT-CHF and the GESICA trials. Grupo de Estudio de la Sobrevida en la Insuficiencia Cardiaca en Argentina. Prog Cardiovasc Dis 40:85, 1997

41. Pacifico A, Hohnloser SH, Williams JH, et al: Prevention of implantable-defibrillator shocks by treatment with sotalol. D,L-Sotalol Implantable Cardioverter-Defibrillator Study Group. N Engl J Med 340:1855, 1999

42. Julian DG, Prescott RJ, Jackson FS, et al: Controlled trial of sotalol for one year after myocardial infarction. Lancet 1:1142, 1982

43. Wellens HJ: Pulmonary vein ablation in atrial fibrillation: hype or hope? Circulation 102:2562, 2000

44. Natale A, Newby KH, Pisano E, et al: Prospective randomized comparison of antiarrhythmic therapy versus first-line radiofrequency ablation in patients with atrial flutter. J Am Coll Cardiol 35:1898, 2000

45. Cox JL, Jaquiss RD, Schuessler RB, et al: Modification of the maze procedure for atrial flutter and atrial fibrillation. II. Surgical technique of the maze III procedure. J Thorac Cardiovasc Surg 110:485, 1995

46. Cox JL, Boineau JP, Schuessler RB, et al: Modification of the maze procedure for atrial flutter and atrial fibrillation. I. Rationale and surgical results. J Thorac Cardiovasc Surg 110:473, 1995

47. Cox JL, Schuessler RB, D'Agostino HJ Jr, et al: The surgical treatment of atrial fibrillation. III. Development of a definitive surgical procedure. J Thorac Cardiovasc Surg 101:569, 1991

48. Gillis AM, Wyse DG, Connolly SJ, et al: Atrial pacing periablation for prevention of paroxysmal atrial fibrillation. Circulation 99:2553, 1999

49. Andersen HR, Nielsen JC, Thomsen PE, et al: Long-term follow-up of patients from a randomised trial of atrial versus ventricular pacing for sick-sinus syndrome. Lancet 350:1210, 1997

50. Delfaut P, Saksena S, Prakash A, et al: Long-term outcome of patients with drug-refractory atrial flutter and fibrillation after single- and dual-site right atrial pacing for arrhythmia prevention. J Am Coll Cardiol 32:1900, 1998

51. Wellens HJ, Lau CP, Luderitz B, et al: Atrioverter: an implantable device for the treatment of atrial fibrillation. Circulation 98:1651, 1998

52. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. Lancet 353:9, 1999

53. Packer M, Bristow MR, Cohn JN, et al: The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. U.S. Carvedilol Heart Failure Study Group. N Engl J Med 334:1349, 1996

54. Hjalmarson A, Goldstein S, Fagerberg B, et al: Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT-HF). MERIT-HF Study Group. JAMA 283:1295, 2000

55. Rawles JM: What is meant by a "controlled" ventricular rate in atrial fibrillation? Br Heart J 63:157, 1990

56. Resnekov L, McDonald L: Electroversion of lone atrial fibrillation and flutter including haemodynamic studies at rest and on exercise. Br Heart J 33:339, 1971

57. Hart RG, Benavente O, McBride R, et al: Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 131:492, 1999

58. Hart RG, Halperin JL: Atrial fibrillation and thromboembolism: a decade of progress in stroke prevention. Ann Intern Med 131:688, 1999

59. Hylek EM, Skates SJ, Sheehan MA, et al: An analysis of the lowest effective intensity of prophylactic anticoagulation for patients with nonrheumatic atrial fibrillation. N Engl J Med 335:540, 1996

60. Hylek EM, Singer DE: Risk factors for intracranial hemorrhage in outpatients taking warfarin. Ann Intern Med 120:897, 1994

61. Adjusted-dose warfarin versus low-intensity, fixed-dose warfarin plus aspirin for high-risk patients with atrial fibrillation. Stroke Prevention in Atrial Fibrillation III randomised clinical trial. Lancet 348:633, 1996

62. Hart RG, Benavente O, Pearce LA: Increased risk of intracranial hemorrhage when aspirin is combined with warfarin: a meta-analysis and hypothesis. Cerebrovasc Dis 9:215, 1999

63. Biblo LA, Yuan Z, Quan KJ, et al: Risk of stroke in patients with atrial flutter. Am J Cardiol 87:346, 2001

64. Bleeding during antithrombotic therapy in patients with atrial fibrillation. The Stroke Prevention in Atrial Fibrillation Investigators. Arch Intern Med 156:409, 1996

65. Ginsberg JS, Chan WS, Bates SM, et al: Anticoagulation of pregnant women with mechanical heart valves. Arch Intern Med 163:694, 2003

66. Hirsh J, Fuster V, Ansell J, et al: American Heart Association/American College of Cardiology Foundation guide to warfarin therapy. J Am Coll Cardiol 41:1633, 2003

67. Gallagher MM, Hennessy BJ, Edvardsson N, et al: Embolic complications of direct current cardioversion of atrial arrhythmias: association with low intensity of anticoagulation at the time of cardioversion. J Am Coll Cardiol 40:926, 2002

68. Prystowsky EN, Benson DW Jr, Fuster V, et al: Management of patients with atrial fibrillation: a statement for healthcare professionals. Subcommittee on Electrocardiography and Electrophysiology, American Heart Association. Circulation 93:1262, 1996

69. Klein AL, Grimm RA, Murray RD, et al: Use of transesophageal echocardiography to guide cardioversion in patients with atrial fibrillation. N Engl J Med 344:1411, 2001

70. Wodlinger AM, Pieper JA: Low-molecular-weight heparin in transesophageal echocardiography–guided cardioversion of atrial fibrillation. Pharmacotherapy 23:57, 2003

71. Berger M, Schweitzer P: Timing of thromboembolic events after electrical cardioversion of atrial fibrillation or flutter: a retrospective analysis. Am J Cardiol 82:1545, 1998

72. Maisel WH, Rawn JD, Stevenson WG: Atrial fibrillation after cardiac surgery. Ann Intern Med 135:1061, 2001

73. Creswell LL, Schuessler RB, Rosenbloom M, et al: Hazards of postoperative atrial arrhythmias. Ann Thorac Surg 56:539, 1993

74. Zimmer J, Pezzullo J, Choucair W, et al: Meta-analysis of antiarrhythmic therapy in the prevention of postoperative atrial fibrillation and the effect on hospital length of stay, costs, cerebrovascular accidents, and mortality in patients undergoing cardiac surgery. Am J Cardiol 91:1137, 2003

75. Crystal E, Connolly SJ, Sleik K, et al: Interventions on prevention of postoperative atrial fibrillation in patients undergoing heart surgery: a meta-analysis. Circulation 106:75, 2002

76. Kowey PR, Stebbins D, Igidbashian L, et al: Clinical outcome of patients who develop PAF after CABG surgery. Pacing Clin Electrophysiol 24:191, 2001

77. Glatter KA, Dorostkar PC, Yang Y, et al: Electrophysiological effects of ibutilide in patients with accessory pathways. Circulation 104:1933, 2001

78. Maron BJ, Olivotto I, Bellone P, et al: Clinical profile of stroke in 900 patients with hypertrophic cardiomyopathy. J Am Coll Cardiol 39:301, 2002

79. Halperin JL: SPORTIF III: a long-term randomized trial comparing ximelagatran with warfarin for prevention of stroke and systemic embolism in patients with nonvalvular atrial fibrillation. American College of Cardiology Annual Scientific Session, Chicago, 2003

80. Falk RH: Atrial fibrillation. N Engl J Med 344:1067, 2001

V SUPRAVENTRICULAR TACHYCARDIA

Melvin M. Scheinman, m.d.

Over the past decade, enormous strides have been made in the treatment of patients with supraventricular tachycardia (SVT). Although acute therapy for SVT continues to require drugs or cardioversion, advances in the understanding of the mechanisms of SVT have led to the development of catheter ablation procedures for most forms of SVT.[1] These procedures often cure the condition, freeing the patient from the need for lifelong drug therapy. This chapter focuses on the most common forms of SVT—excluding atrial fibrillation, which is discussed in detail elsewhere [*see 1:IV Atrial Fibrillation*].

Classification

SVT is often paroxysmal (PSVT). Clinically, PSVT is marked by palpitations, occurring in episodes that start and end abruptly. During these episodes, the 12-lead ECG shows a heart rate greater than 100 beats/min and, typically, narrow QRS complexes. For almost all patients with PSVT, the underlying mechanism of the tachycardia is atrioventricular node reentry (AVNRT), reentry involving an accessory pathway (AVRT), or atrial tachycardia. AVNRT and AVRT are the most common and the second most common causes of PSVT, respectively. Atrial flutter also presents as a rapid regular tachycardia, but this arrhythmia usually does not begin and end abruptly.

The clinician has a variety of tools to distinguish the various mechanisms of SVT [*see Figure 1*]. The use of carotid massage[2] or intravenous adenosine[3] [*see Figure 2*] may be diagnostic, therapeutic, or both. If vagal maneuvers terminate the arrhythmia acutely or produce no effect, the patient probably has AVNRT or AVRT. In patients with atrial tachycardia, these maneuvers will frequently result in transient AV block. Perpetuation of the arrhythmia in the face of AV block strongly suggests atrial tachycardia or atrial flutter.[3] Intravenous adenosine will almost always terminate tachycardia from AVNRT or AVRT, but focal atrial tachycardia may also terminate abruptly after adenosine. Hence, the use of adenosine does not reliably distinguish those disorders from atrial tachycardia unless it produces AV block.[3]

Paying careful attention to the relationship between the P wave and the QRS complex during tachycardia is also very helpful in distinguishing tachycardia mechanisms[4] [*see Figure 1*]. If the retrograde P wave falls within or just after the QRS, the most likely diagnosis is AVNRT [*see Figure 3*]. If the tachycardia shows a retrograde P wave in the ST segment [*see Figure 4*], AVRT is

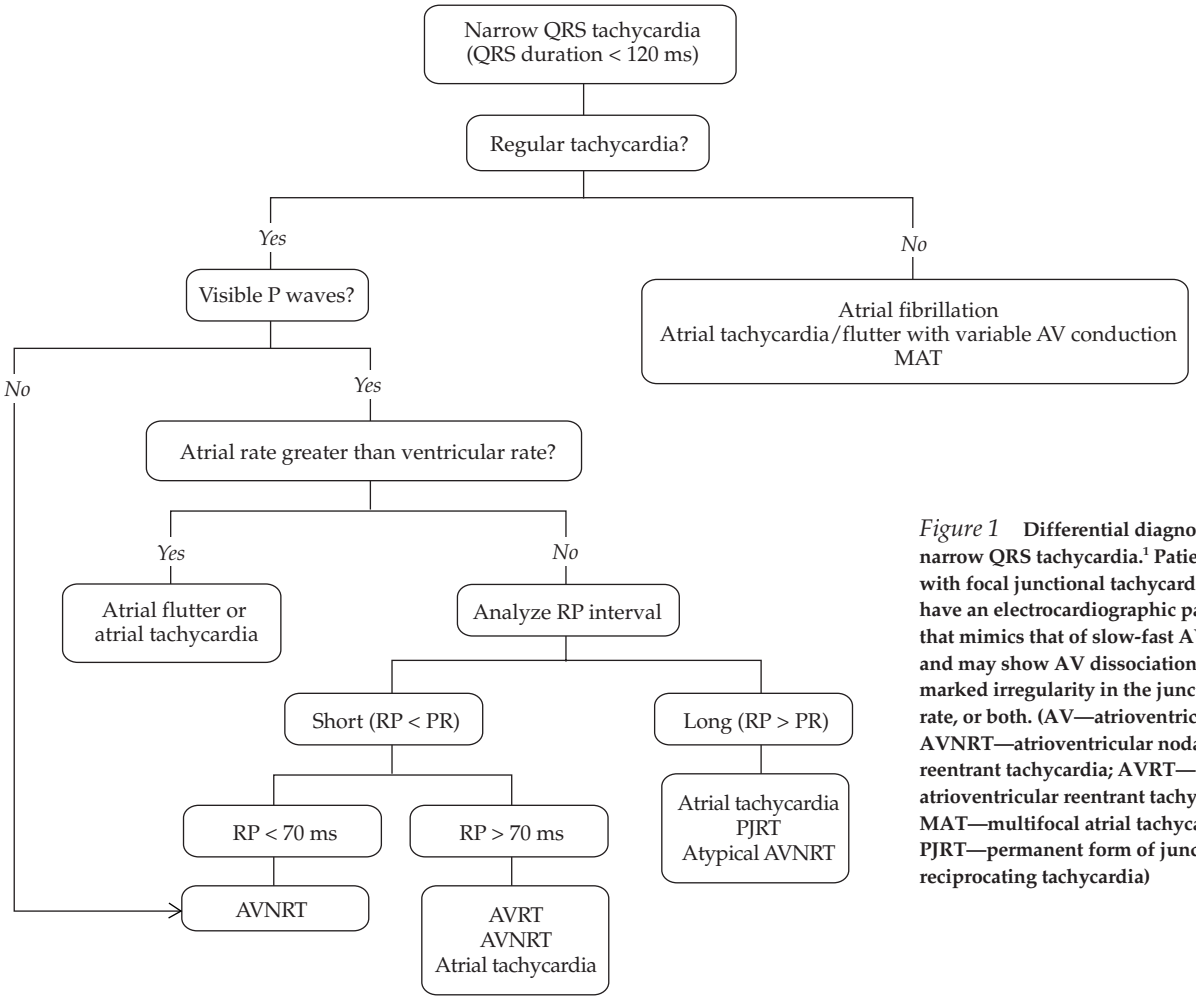


Figure 1 **Differential diagnosis for narrow QRS tachycardia.[1] Patients with focal junctional tachycardia may have an electrocardiographic pattern that mimics that of slow-fast AVNRT and may show AV dissociation, marked irregularity in the junctional rate, or both. (AV—atrioventricular; AVNRT—atrioventricular nodal reentrant tachycardia; AVRT—atrioventricular reentrant tachycardia; MAT—multifocal atrial tachycardia; PJRT—permanent form of junctional reciprocating tachycardia)**

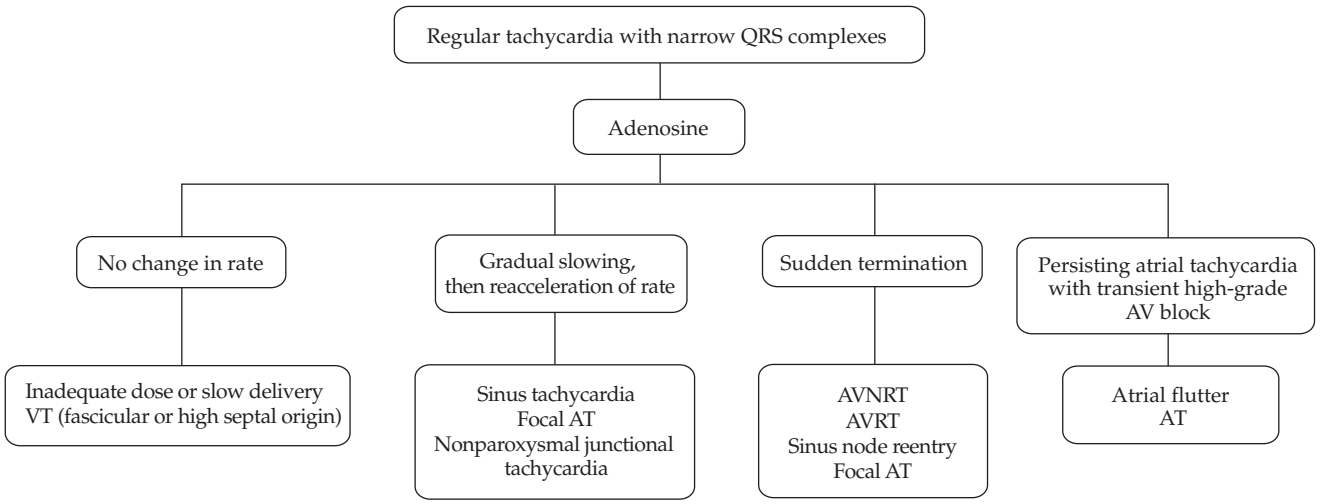


Figure 2 **The response to intravenous adenosine can be useful in determining the cause of tachycardia.[1] (AT—atrial tachycardia; AV—atrioventricular; AVNRT—atrioventricular nodal reentrant tachycardia; AVRT—atrioventricular reentrant tachycardia; VT—ventricular tachycardia)**

most likely. Finally, atrial tachycardia is characterized by the presence of P waves immediately in front of the QRS (long RP tachycardia) [*see Figure 5*].

Although the QRS complex is usually narrow in SVT, it may be broad (> 120 ms) in patients who have either bundle branch block or aberrant conduction. A number of ECG findings have been found very helpful in distinguishing SVT with a broad QRS complex from ventricular tachycardia (VT).[5,6] For example, AV dissociation (i.e., independent atrial activity during tachycardia), fusion beats, or capture beats prove the presence of VT. Unfortunately, AV dissociation is not apparent in 80% to 85% of patients with rapid VT, because the P wave is obscured by the QRS complex and T waves.[6] In this setting, morphologic criteria may be very helpful in distinguishing SVT from VT.

Use of morphologic criteria begins with careful attention to the precordial leads [*see Figure 6*]. Any of the following features in the precordial tracings will favor the diagnosis of VT: (1) Concordance of all the precordial leads (i.e., all are positive or all are negative); (2) absence of an initial positive deflection (r wave) followed by a negative deflection (s wave; recall that in ECG nomenclature, upper-case letters denote dominance; small waves are designated by lowercase letters); (3) an r/s pattern is present but the time from the initial r to the nadir of the s wave is greater than 60 ms; (4) presence of a right bundle branch pattern in lead V1, with an r greater than s or a qr pattern, where q indicates the initial negative deflection; (5) presence of a left bundle branch pattern in V1, with a broad r wave (> 30 ms) or an interval of greater than 60 ms from the onset of the r wave to the nadir of the s wave; (6) extreme left axis deviation; or (7) very broad QRS complexes (> 160 ms).

Atrioventricular Nodal Reentry Tachycardia

PATHOGENESIS

Normally, sinus impulses are discharged into the surrounding atria and directed to the region of the node that resides in the atrial septum. The AV nodal impulses then propagate through the ventricles over the His-Purkinje system. The normal AV node has a single transmission pathway. In two to three persons

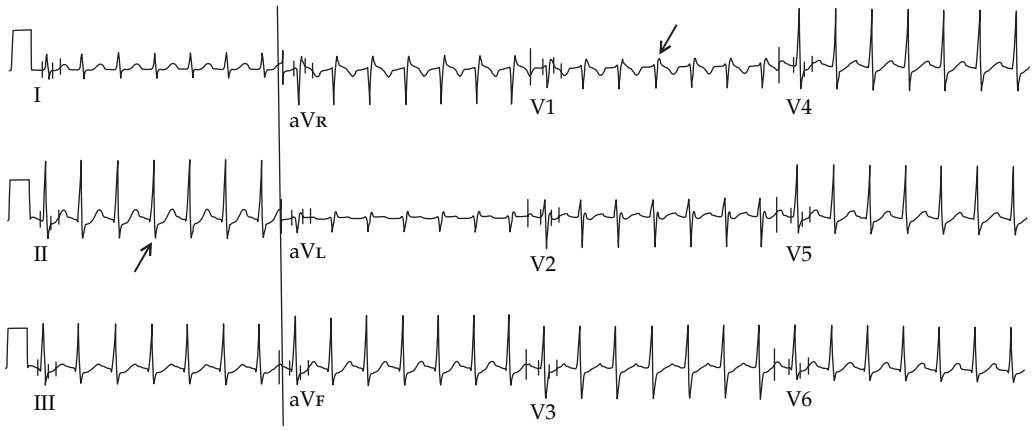


Figure 3 **A 12-lead ECG shows paroxysmal supraventricular tachycardia from AV nodal reentry (AVNRT). The arrows point to a pseudo r[1] in lead V1 and S waves in the inferior leads (II, III, and aVF), which disappeared with conversion to sinus rhythm.**

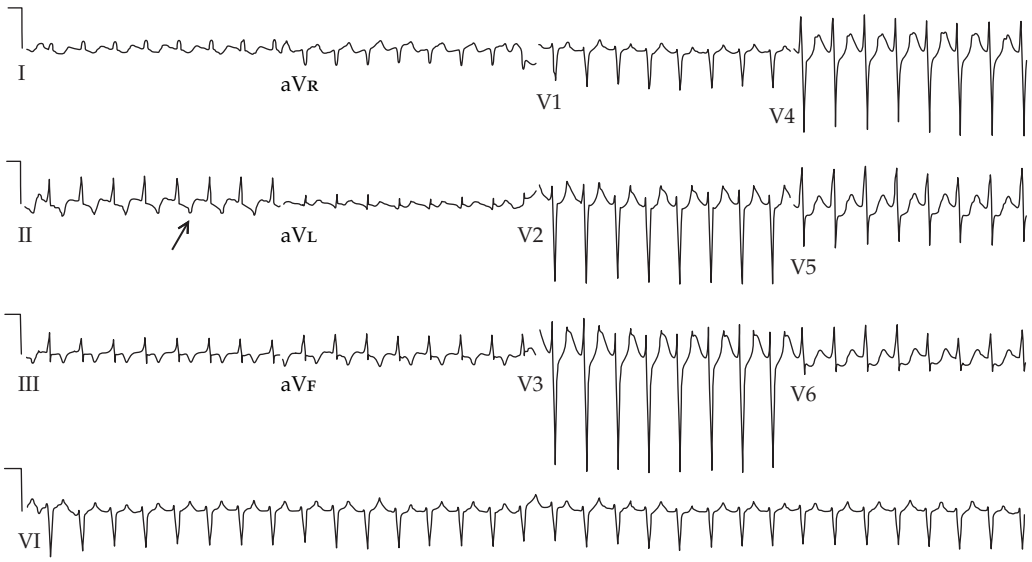


Figure 4 **A 12-lead ECG showing narrow complex tachycardia with P waves (arrow) inscribed well after the QRS, taken from a patient who had paroxysmal supraventricular tachycardia supported by an accessory pathway.**

per 1,000 population, however, the AV node has both a normal (fast) pathway and a second, slow pathway.[7,8] In such persons, the sinus impulse is ordinarily transmitted over the fast pathway to the ventricle, and slow pathway conduction is preempted. However, if an atrial premature complex (APC) occurs at a critical point in the conduction cycle, the impulse can block in the fast pathway, thus allowing for anterograde (forward) conduction over the slow pathway and retrograde (backward) conduction over the fast pathway [*see Figure 7*]. The latter situation may produce a single echo beat (a beat that returns to the chamber of origin) or stabilize into a circus-movement tachycardia.

DIAGNOSIS

The diagnosis of AVNRT can usually be made by careful analysis of the 12-lead ECG.[4] Because retrograde conduction over the AV node is occurring more or less simultaneously with anterograde conduction to the ventricles, the P wave is either buried within the QRS complex or inscribed just after the QRS. The P wave inscribed by retroconduction over the AV node will be negative in the inferior leads and positive in lead V1; therefore, PSVT from AVNRT may manifest as small negative deflections in the inferior leads and a small positive deflection in V1—the so-called pseudo r[1] pattern[5] [*see Figure 3*].

MANAGEMENT

Acute Therapy

AVNRT may respond to carotid sinus massage[2] but is highly responsive to intravenous adenosine,[3] beta blockers,[9] or calcium channel blockers[10] [*see Table 1*].

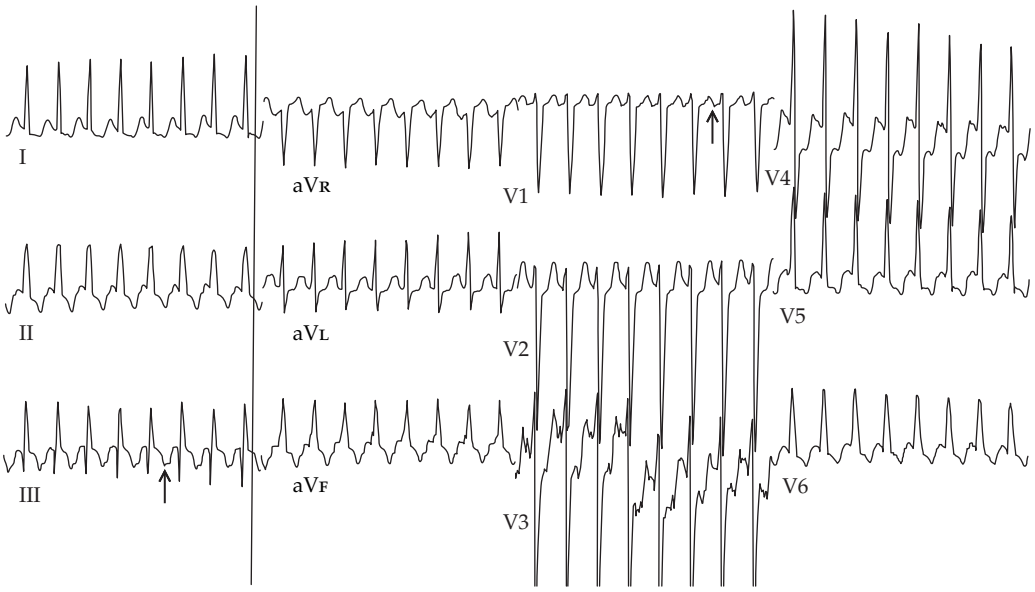


Figure 5 **A 12-lead ECG shows tachycardia with P waves (arrows) just preceding the QRS complex. The patient in this case had a focal atrial tachycardia emanating from the lateral tricuspid annulus.**

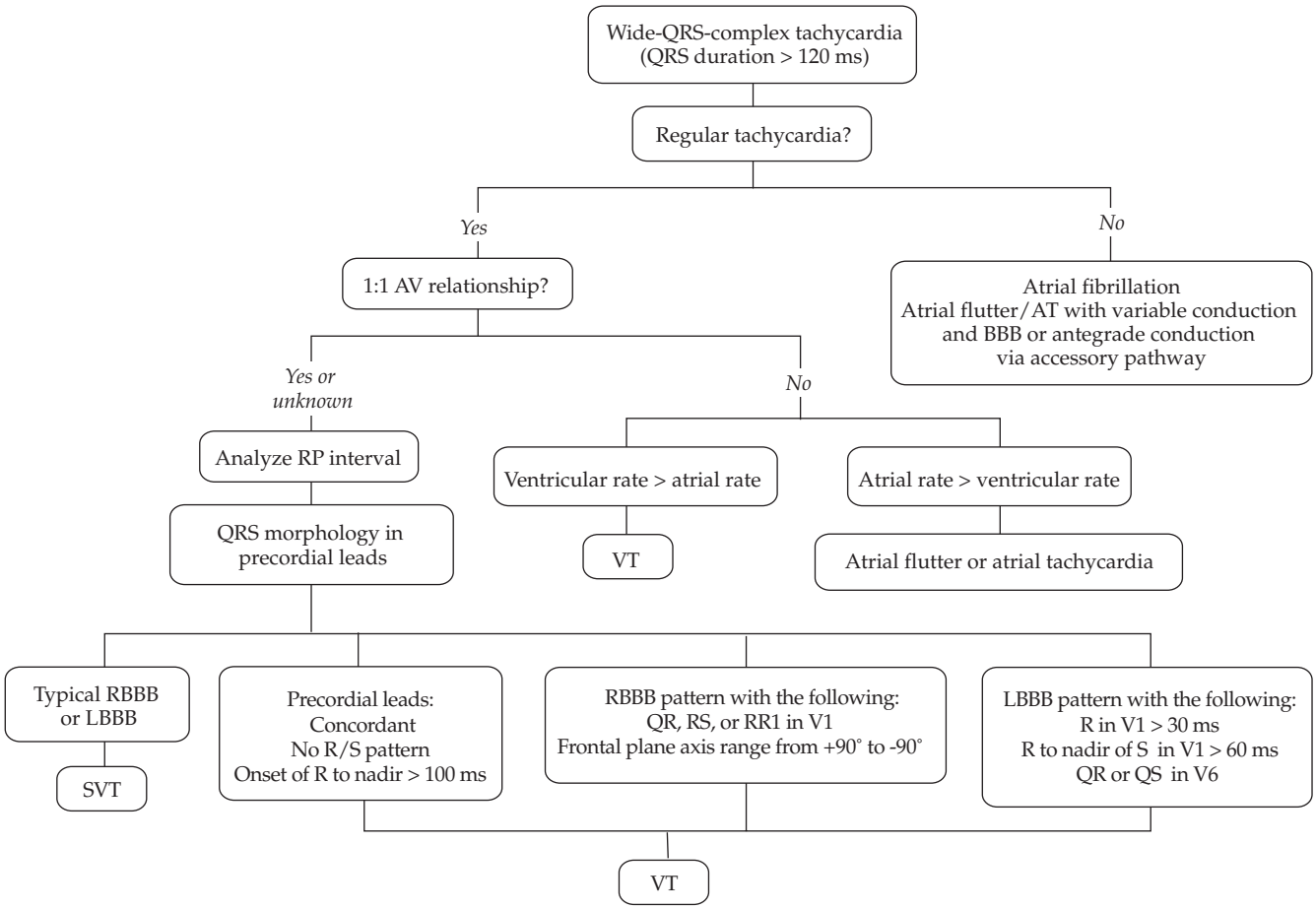


Figure 6 **Differential diagnosis for wide (> 120 ms) QRS complex tachycardia.[1] If the tachycardia is regular and comparison with a baseline electrocardiogram shows that the QRS complex is identical to that during sinus rhythm, the patient may have supraventricular tachycardia (SVT) with bundle branch block (BBB) or antidromic atrioventricular reentrant tachycardia (AVRT). If the patient has a history of myocardial infarction or has structural heart disease, ventricular tachycardia (VT) is likely. Vagal maneuvers or adenosine may convert regular tachycardia, although adenosine should be used with caution when the diagnosis is unclear, because this drug may produce ventricular fibrillation (VF) in patients with coronary artery disease and patients with alternative pathways who have atrial fibrillation with a rapid ventricular rate. Precordial leads are concordant when all show either positive or negative deflections. Fusion complexes are diagnostic of VT. In preexcited tachycardias, the QRS is generally wider (i.e., more preexcited) than during sinus rhythm. (AT—atrial tachycardia; AV—atrioventricular; LBBB—left bundle branch block; RBBB—right bundle branch block)**

Adenosine If carotid massage fails to convert SVT, the drug of choice is intravenous adenosine, which is effective in 95% of cases.[10,11] The initial dose is given as a rapid bolus infusion of 6 mg, followed by 12 mg and finally 18 mg if necessary. The bolus must be given rapidly and then followed by a saline flush. If administration is too slow, the adenosine may be metabolized before it reaches the AV node. Possible adverse effects include headache, wheezing, and flushing. These effects disappear within 45 to 60 seconds. It is important to note that atrial, ventricular, and junctional premature beats are commonly observed after adenosine. In 3% to 5% of cases, the APCs trigger atrial fibrillation,[3] which may result in serious problems for patients with accessory pathways (see below). If possible, an external defibrillator should be readily available when adenosine is administered.

The most common reason for failure to respond to adenosine is that multiple premature beats are retriggering the tachycardia. In this setting, a longer-acting intravenous preparation (i.e., 5 mg of metoprolol or 0.1 mg/kg of verapamil) is indicated. Agents that more selectively block purogenic receptors have been

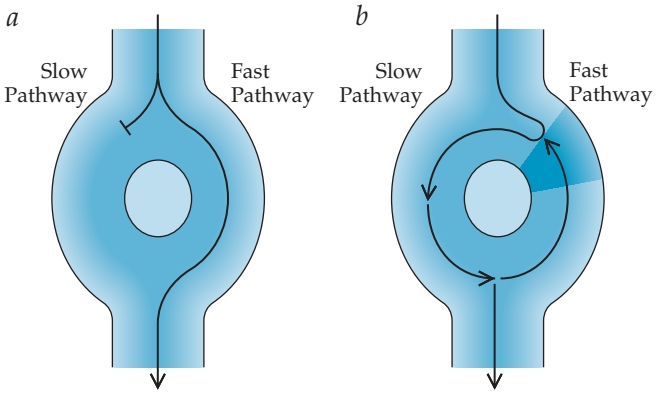


Figure 7 **In persons with dual pathways in the AV node, the sinus impulse is normally transmitted over the fast pathway to the ventricle and slow pathway conduction is preempted (*a*). However, if an atrial premature complex occurs during the fast pathway's refractory period, the impulse can block in the fast pathway. This may allow for anterograde (forward) conduction over the slow pathway and retrograde (backward) conduction over the fast pathway (*b*).**

Table 1 Drugs Used to Maintain Sinus Rhythm in Patients with Supraventricular Tachycardia[55]

Drug	*Typical Daily Dose*	*Potential Adverse Effects*
Amiodarone	100–400 mg*	Photosensitivity, pulmonary toxicity, polyneuropathy, GI upset, bradycardia, torsade de pointes (rare), hepatic toxicity, thyroid dysfunction
Disopyramide	400–750 mg	Torsade de pointes, heart failure, glaucoma, urinary retention, dry mouth
Dofetilide	500–1,000 μg	Torsade de pointes
Flecainide	200–300 mg	Ventricular tachycardia, heart failure, enchanced AV nodal conduction (conversion to atrial flutter)
Procainamide	1,000–4,000 mg	Torsade de pointes, lupuslike syndrome, GI symptoms
Propafenone	450–900 mg	Ventricular tachycardia, heart failure, enhanced AV nodal conduction (conversion to atrial flutter)
Quinidine	600–1,500 mg	Torsade de pointes, GI upset, enhanced AV nodal conduction
Sotalol	240–320 mg†	Torsade de pointes, heart failure, bradycardia, exacerbation of chronic obstructive or bronchospastic lung disease

*A loading dose of 600 mg/day is usually given for 1 month, or a dose of 1,000 mg/day is given for 1 week.
†Adjust dose for renal function and QT-interval response during in-hospital initiation phase
AV—atrioventricular GI—gastrointestinal

shown to be very effective and associated with fewer side effects than older agents. Selective purogenic blockers are currently under investigation.

Long-term Therapy

A wide variety of drugs have proved effective for controlling episodes of AVNRT, including beta blockers,[9] calcium channel blockers,[12] and digoxin[13] [*see Table 1*]. Long-term drug therapy is associated with frequent recurrences and adverse effects, however. In patients without structural cardiac disease, class IC antiarrhythmic agents (e.g., flecainide, propafenone) are more effective than drugs that act by blocking AV nodal conduction, but recurrence rates nevertheless range from 25% to 35%.[14-16] For patients who have episodes infrequently and tolerate them well, some cardiologists will prescribe medication for use as needed—the "pill in the pocket" approach. For example, single-dose diltiazem (120 mg) and propranolol (80 mg) have been shown to be more effective than placebo or flecainide in patients with PSVT.[17]

Catheter Ablation

Current catheter ablative techniques involve placement of an electrode catheter between the tricuspid annulus and coronary sinus in the so-called slow pathway region.[18] One or more applications of radiofrequency energy are delivered through the catheter to destroy or attenuate the slow pathway. The success rate of ablation is over 96%, and the only significant complication is AV block, which occurs in approximately 1% of patients.[19]

Catheter ablation for AVNRT has proved so safe and effective that it is clearly the procedure of choice for patients in whom drug therapy fails. Moreover, it can be offered to those with milder symptoms who prefer to avoid long-term drug therapy. Precise recommendations for drug therapy versus ablative therapy are provided in the American College of Cardiology/American Heart Association/European Society of Cardiology guidelines.[1]

Atrioventricular Reentry Tachycardia

PATHOGENESIS

The normal conduction system of the heart limits the propagation of electrical impulses from the atria to a single pathway through the AV node and the His-Purkinje system. This limitation delays ventricular activation and thus optimizes mechanical function. The presence of an alternative pathway of atrioventricular conduction creates the potential for reentrant tachycardia.

The most prominent manifestation of accessory atrioventricular pathways is the Wolff-Parkinson-White (WPW) syndrome. In this syndrome, the accessory pathway can be located at various regions around the tricuspid and the mitral atrioventricular rings, but it is most commonly sited at the left free wall of the mitral annulus. The next most common pathway sites are the posteroseptal and right free wall areas. Pathways in the anteroseptal and the midseptal regions are relatively rare. Occasionally, posteroseptal pathways can be associated with a branching vein from the coronary sinus. On occasion, a patient will have more than one accessory pathway.

The basic mechanism of tachycardia in AVRT is similar to that of AVNRT. Electrical impulses can travel down both the AV node and the accessory pathway to activate the ventricles, with ventricular activation occurring earlier at sites near the accessory pathway than at sites activated normally (i.e., ventricular preexcitation). An APC may block in the accessory pathway but conduct over the normal pathway to activate the ventricle. After ventricular depolarization, the impulse may return to the atrium via retrograde conduction over the accessory pathway, leading to a sustained tachycardia.[20]

The most feared arrhythmia in the WPW syndrome involves atrial fibrillation with dominant conduction over an accessory pathway that has rapid conduction properties[21,22] [*see Figure 8*]. These patients may experience extraordinarily rapid ventricular rates and are at risk for sudden cardiac death from ventricular fibrillation.[23] In one large series, atrial fibrillation developed in 30% of patients with the WPW syndrome.[24]

DIAGNOSIS

Clinical Presentation

Symptomatic tachyarrhythmias associated with the WPW syndrome generally begin in the teenage years or during early adulthood. Pregnancy may produce an initial attack in some women. Pregnancy can also be associated with an increasing frequency of attacks and more symptomatic episodes. Symptoms are generally paroxysmal palpitations with or without dizziness, syncope, shortness of breath, weakness, or chest pain. Diuresis is another frequently described symptom; it occurs 30 minutes to an hour after onset of tachycardia and may be related to production of atrial natriuretic factor during the arrhythmia.

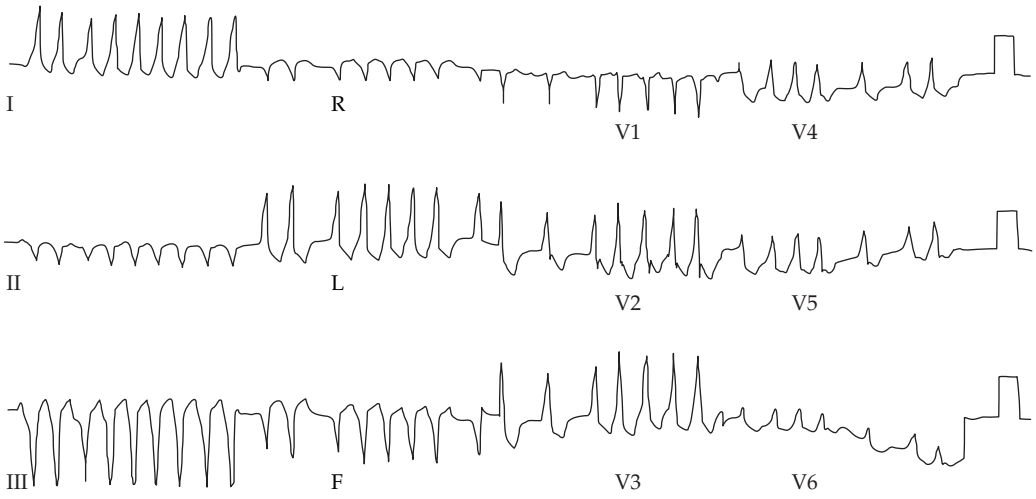


Figure 8 **A 12-lead ECG in a patient with Wolff-Parkinson-White syndrome shows the rapid, irregularly irregular ventricular rate and wide QRS complexes of atrial fibrillation with a very short refractory period. This is an especially dangerous arrhythmia.**

Electrocardiographic Findings

Ventricular preexcitation may be evident on a baseline ECG as fusion complexes (WPW pattern). The WPW pattern comprises a short PR interval and an earlier than normal deflection on the QRS complex (delta wave).[25] The ECG during AVRT will usually show a narrow complex with the retrograde P wave falling in the ST segment because atrial activation occurs well after ventricular depolarization[4] [*see Figure 4*]. Of interest is that a subset of patients with AVRT never show manifest anterograde conduction over the accessory pathway yet still have this form of tachycardia.[20] The only evidence that the tachycardia is supported by an accessory pathway is that the retrograde P wave clearly occurs after the QRS during tachycardia. On rare occasions, patients may have slowly conducting retrograde pathways[26]; their ECG will show a long RP–short PR relationship. These patients tend to have persistent tachycardias that have been referred to as the permanent form of junctional tachycardia (PJRT). In addition, approximately 5% of patients with the WPW syndrome (WPW pattern and arrhythmias) will show anterograde conduction over the accessory pathway with retrograde conduction through the AV node or over a separate accessory pathway. The ECG in these patients will show a wide-complex tachycardia with retrograde P waves preceding the QRS complex.

MANAGEMENT

Acute Therapy

Acute management of AVRT is similar to that for AVNRT: adenosine is the drug of choice,[11] but calcium channel blockers[13] or beta blockers[9] are also effective. Again, because adenosine usually provokes APCs and thus may in rare instances precipitate atrial fibrillation,[3] it is advisable to have ready access to an external defibrillator when using this agent.

Long-term Therapy

Long-term therapy for AVRT may be directed at interfering with conduction either through the AV node (i.e., with beta blockers or calcium channel blockers[27]) or through the accessory pathway (i.e., with class IC or class III antiarrhythmic agents[28-32]). Oral digitalis therapy is contraindicated because very rapid ventricular rates may occur if atrial fibrillation develops. Class IC agents appear to be more effective than AV nodal blockers, but their use is restricted to patients who do not have significant cardiac disease. Class III agents (particularly amiodarone) are limited by long-term systemic toxicity and modest efficacy for patients with WPW and atrial fibrillation.[30] In general, drug therapy is attended by a significant risk of arrhythmic recurrence and adverse drug effects.

Treatment of WPW and Atrial Fibrillation

The treatment of WPW and atrial fibrillation is different from the treatment of AVRT. Because atrial fibrillation may precipitate a life-threatening arrhythmia, urgent therapy is required. If the patient presents with hemodynamic collapse, emergency direct current (DC) cardioversion is the first step. If the patient is less ill, trials of intravenous drug therapy are in order.[33] The drug of choice is procainamide, 50 mg/min to a total of 1 g, or ibutilide, 2 mg infused over 15 minutes. Ibutilide is very effective but should be used only in patients without significant structural cardiac disease. Intravenous digoxin or calcium channel blockers may result in an inordinate increase in heart rate and so should be avoided. Beta blockers, lidocaine, and adenosine are not likely to be effective, and their use will tend only to delay effective therapy.

Catheter Ablation

Reports from both single centers[34,35] and multicenter prospective registries[36,37] have documented the efficacy and possible adverse effects of ablative therapy in AVRT. Current techniques allow for successful ablation of accessory pathways that traverse the AV annulus or the anterior or posteroseptal spaces. For pathways over the left AV groove, current ablation techniques involve use of either transseptal or retrograde aortic approaches.[38] The overall success rate for ablation is approximately 95%.

Complications of ablation are primarily related to the site of the accessory pathway. For example, patients with an anteroseptal accessory pathway are at risk for injury to the AV node (5%), whereas ablations of left-sided accessory pathways carry a risk of cerebrovascular accident, myocardial perforation, or coronary artery occlusion.[36,37] The overall incidence of significant adverse

effects varies from 2% to 4%. Death associated with ablative procedures is quite rare, occurring in 0.13% to 0.2% of cases.[36,37,39]

Treatment Selection

The remarkable efficacy and safety of ablation make this mode of therapy more attractive than long-term drug therapy for symptomatic patients. Drug therapy carries the possibility of recurrent arrhythmias, including atrial fibrillation. Hence, ablation is currently recommended for all patients with symptomatic WPW. Patients with mild symptoms and without manifest preexcitation can be managed with drug therapy, but even in these cases ablation would appear to be a favored approach. Some of these patients decline long-term drug therapy, leaving ablation as the only alternative.

Asymptomatic preexcitation The management of asymptomatic preexcitation remains controversial. The vast majority of these patients have an overall good prognosis; sudden cardiac death is a rare initial manifestation. Leitch and colleagues followed asymptomatic WPW subjects in whom atrial fibrillation was induced during invasive electrophysiologic study; although approximately 20% demonstrated the capacity for rapid ventricular conduction, on follow-up few became symptomatic and none died suddenly.[39] A later study, however, emphasized findings on electrophysiologic testing (e.g., inducible AVRT, atrial fibrillation, and multiple pathways) that indicated increased risk for subsequent spontaneous development of atrial fibrillation or even sudden death.[40] Whether to treat an asymptomatic patient can also be decided on an individual basis[1]; for example, patients judged to be in high-risk occupations (e.g., airplane pilots, bus drivers) might well be considered for ablative therapy.

Focal Atrial Tachycardia

Regular tachycardias emanating in an atrial area and showing a centripetal pattern of spread are designated as focal atrial tachycardias (FATs). These arrhythmias are the least common cause of PSVT but nevertheless can cause significant morbidity. This is particularly true if the arrhythmia is incessant, which can result in the development of so-called tachycardia myopathy.

Atrial tachycardia may arise from sites in either the right or left atrium. The most common site of FAT is in the right atrium, with predilection for sites over the crista terminalis, tricuspid annulus, or coronary sinus.[41,42] In the left atrium, FAT is more apt to develop at the ostium of the pulmonary veins or over the mitral annulus.[42]

ELECTROCARDIOGRAPHIC DIAGNOSIS

In patients with FAT, the P wave may appear anywhere in the diastolic cycle but most often appears in front of the QRS (long RP tachycardia) [*see Figure 5*]. The ectopic P wave has a different shape than the sinus P wave unless the tachycardia originates from the high cristal or right pulmonary vein area. The P wave morphology gives excellent clues to tachycardia localization.[43,44] For example, P waves from left atrial foci will show negative deflection in leads I or aVL and positive deflections in the precordial leads. Right atrial foci tend to show negative P waves in lead V1 but positive or biphasic deflection in aVL. As a rule, foci from superior atrial sites generally produce strongly positive P waves in the inferior leads, whereas those arising from the inferior atrium produce negative P waves.

MANAGEMENT

Acute Therapy

Acute treatment of FAT attempts either to convert the arrhythmia or to slow the heart rate. Drugs used to slow the rate are the AV nodal blockers (i.e., digoxin, beta blockers, or calcium channel blockers). In contrast, class IC antiarrhythmic agents (e.g., flecainide, propafenone) or class III agents (e.g., amiodarone or sotalol) may terminate the tachycardia. Intravenous adenosine may be effective in terminating FAT and should be tried early. DC cardioversion may not be effective, particularly if the tachycardia results from an automatic mechanism.

Long-term Therapy

Long-term oral therapy for FAT is not well defined.[1] The general approach is empirical, with initial use of AV nodal blockers followed by class IC or III antiarrhythmic agents if the AV nodal blockers are ineffective.

Catheter Ablation

Catheter ablative procedures have been successfully applied to patients with FAT. Ablation has proved more effective for patients with right atrial foci (in whom the success rate is approximately 90%) than for those with left atrial foci (in whom the success rate is approximately 70%). A study of pooled data that included 514 patients showed an overall success rate of 86%, with an incidence of significant complications from 1% to 2%.[45]

Multifocal Atrial Tachycardia

Multifocal atrial tachycardia, generally regarded as automatic in origin, is characterized by atrial rates of 100 to 130 beats/min, three or more morphologically distinct (nonsinus) P waves, and variable AV conduction. It is commonly associated with respiratory disease and heart failure. Hypoxemia is a frequent finding. The arrhythmia may be exacerbated by digitalis excess, theophylline toxicity, or hypokalemia.

Treatment of multifocal atrial tachycardia is usually directed at the underlying precipitants. Metoprolol (used cautiously in patients with bronchospasm) or verapamil may slow atrial and ventricular rates and, occasionally, may restore sinus rhythm. Potassium and magnesium supplements may help suppress the arrhythmia. Amiodarone has also been useful in restoring sinus rhythm.

Atrial Flutter

Rapid reentrant atrial arrhythmias are referred to as atrial flutter. The most common circuit involves reentry around the tricuspid annulus. The reentrant circuit is usually counterclockwise (in the left anterior oblique projection) but may be clockwise.[46] Other circuits may involve the upper portion of the right atrium.[47] Less commonly, left atrial (LA) circuits are operative. LA circuits may involve the mitral annulus or scars around the posterior LA wall, pulmonary veins, or the foramen ovale.[47]

DIAGNOSIS

Clinical Presentation

Atrial flutter generally occurs in older patients who have associated cardiopulmonary disease. Atrial flutter may appear acutely during acute myocardial infarction, after cardiac surgery, or

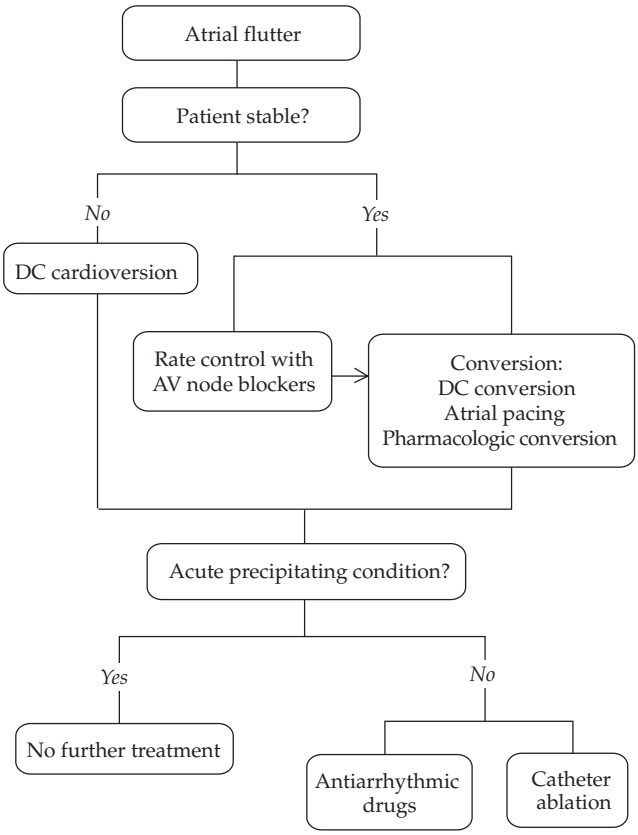


Figure 9 **Management of atrial flutter. With patients whose condition is unstable (e.g., because of heart failure, shock, or acute myocardial infarction), atrial flutter typically does not recur once the underlying disorder is resolved. Anticoagulant precautions, as per atrial fibrillation, should be taken in patients undergoing elective attempts to convert atrial flutter to sinus rhythm. (AV—atrioventricular; DC—direct current)**

with acute pulmonary insufficiency; in such cases, the arrhythmia usually does not recur once the inciting event has resolved. In contrast, atrial flutter in patients without concomitant acute illness tends to recur; like atrial fibrillation, it is usually a relapsing and remitting disease.

Electrocardiographic Findings

The ECG in patients with atrial flutter usually shows a flutter rate of 300 beats/min with 2:1 AV block. The most common atrial flutter pattern—a counterclockwise loop around the annulus—manifests as negative flutter waves in the inferior leads and positive waves in V1.[48] The ECG shows a continuous or so-called picket-fence appearance. In contrast, patients with a clockwise pattern will have positive flutter waves in the inferior leads and negative waves in V1. The ECG is much more variable for nonannular types of flutter circuits.[47]

MANAGEMENT

Acute Therapy

Treatment of atrial flutter is directed at attempts to convert the arrhythmia or use of AV nodal blockers to slow the ventricular response [*see Figure 9*]. Acute conversion of atrial flutter can be accomplished electrically by use of external DC shocks[49] or by pacing.[50,51] Atrial flutter is usually exquisitely responsive to a small "dose" of DC shock (i.e., 25 to 50 joules).[49] Atrial overdrive pacing is also quite effective for terminating flutter, especially when the patient has been pretreated with drugs (i.e., ibutilide or procainamide).[51] Overdrive pacing is particularly appropriate for atrial flutter that occurs after cardiac surgery, because atrial wires are routinely left in place postoperatively in such patients. Transesophageal pacing has also been used to terminate flutter,[52] but its popularity has been limited by the need for analgesics to alleviate the associated chest pain.

Ibutilide (a class III antiarrhythmic agent) may be used to convert atrial flutter to sinus rhythm. Randomized prospective studies have shown that ibutilide is approximately 70% effective for this purpose.[53] In addition, ibutilide has been shown to be far more effective than intravenous procainamide.[54] Ibutilide is given in 1 mg aliquots over 10 minutes separated by a 10-minute rest period. A total of 2 mg of the drug is used, and the patient must remain under telemetry monitoring for approximately 4 hours after drug delivery. Ibutilide should not be given to patients with severe structural cardiac disease (i.e., those with a left ventricular ejection fraction less than 30%) because the risk of torsade de pointes becomes significant in this setting.

Patients with atrial flutter are at risk for thromboembolism. The current recommendations for anticoagulant therapy are the same as those for patients with atrial fibrillation.[55] For example, if the flutter duration is less than 48 hours, the risk of left atrial clot is small, and one may proceed with chemical or electrical cardioversion without full anticoagulation. Anticoagulant therapy is still required for 4 to 6 weeks after conversion because of the increased risk of thromboembolism secondary to decreased left atrial flow velocity after conversion. If the flutter duration is greater than 48 hours, a transesophageal echocardiogram to exclude clot is recommended before cardioversion. Complete guidelines for antithrombotic therapy in patients with atrial flutter are described elsewhere [*see 1:IV Atrial Fibrillation*].

As an alternative to cardioversion, AV nodal blocking agents can be used to decrease the ventricular response in patients with flutter. Controlled trials have demonstrated the efficacy of intravenous calcium channel blockers (verapamil or diltiazem) in producing prompt decreases in heart rate.[56] Calcium channel blockers have been shown to reduce the heart rate below 100 beats/min more promptly than digoxin or amiodarone.

Long-term Therapy

Drug therapy for chronic atrial flutter is notoriously unreliable, and long-term rate control alone usually requires large doses of AV nodal blocking agents. A more effective intervention involves an ablative procedure in which radiofrequency lesions are applied in a line from the tricuspid annulus to the inferior vena cava.[57] This area is the critical isthmus for the usual type of atrial flutter circuit. Ablation of this area resulting in total conduction block of the isthmus is associated with a 90% to 100% cure rate in flutter.[58] Non–isthmus-dependent flutter circuits may involve either the right or the left atrium and usually require sophisticated mapping tools to determine the tachycardia circuit and the critical isthmus needed for curative ablation. These patients should be referred to experienced centers for evaluation.

Many patients have both atrial flutter and atrial fibrillation. For example, atrial flutter may deteriorate into atrial fibrillation, or bursts of atrial fibrillation may trigger atrial flutter. In addition, approximately 15% to 30% of patients treated with

class IC antiarrhythmics or amiodarone for atrial fibrillation will develop stable atrial flutter.[59] In these patients, radiofrequency ablation of the flutter circuit together with continuance of drug therapy is usually quite effective in controlling both atrial fibrillation and flutter. Ablation of the flutter usually does not cure the atrial fibrillation.[60]

Sinus Tachycardia

Sinus tachycardia is usually a normal reflex response to changes in physiologic, pharmacologic, or pathophysiologic stimuli, such as exercise, emotion (e.g., anxiety, anger), fever, hemodynamic or respiratory compromise, anemia, thyrotoxicosis, poor physical condition, sympathomimetic or vagolytic agents, and abnormal hemoglobins. Heart rate during sinus tachycardia generally does not exceed 180 beats/min, except perhaps in young persons, who may achieve sinus rates greater than 200 beats/min during vigorous exercise.

When sinus tachycardia is a reflex response to altered physiology, the resulting increase in cardiac output is usually beneficial. Tachycardia resolves when conditions return to normal.

Inappropriate Sinus Tachycardia

An infrequent but troublesome problem, inappropriate sinus tachycardia (IST) appears to be a true syndrome with cardiac, neurologic, and psychiatric components. It affects women more often than men. Structural heart disease is generally absent. In one series of 475 patients, IST was the indication for catheter ablation in 2.3%.[61]

DIAGNOSIS

Clinical Presentation

IST may be persistent or episodic. It is often precipitated by arising from a reclining or sitting position (postural orthostatic tachycardia).[62] Very rapid rates (> 170 beats/min) may be triggered by minimal exertion.

The tachycardia is frequently accompanied by symptoms of dizziness, near-syncope, or syncope. Fatigue and atypical chest pain may also accompany IST. Peculiar but inconsistent autonomic and hemodynamic findings may be seen in these patients. This suggests that the syndrome is not uniform in etiology.

Electrocardiographic Findings

Because tachycardia rates may arise from higher foci, the P waves seen during IST may differ slightly from those seen at rest.

MANAGEMENT

Drug Therapy

Beta blockers and calcium channel blockers (i.e., verapamil or diltiazem) may be used to alleviate tachycardia in IST. Unfortunately, these drugs are often not effective and tend to exacerbate the nonspecific symptoms that accompany this syndrome. Agents that alter sinus node automaticity, autonomic tone, or both, such as flecainide, propafenone, and amiodarone, may be tried in selected patients.[63]

Catheter Ablation

Radiofrequency ablation has been employed to ablate or modify the sinus node in IST. Large-tipped (8 to 10 mm) catheters are often required to create more sizable lesions. Successful modification or ablation has been achieved in 70% to 100% of patients.[64] Sinus nodal modification is associated with a 10% to 27% risk of sinus node damage necessitating permanent pacing.

Both intracardiac electrograms and intracardiac ultrasonography have been used to target lesion delivery. Intracardiac ultrasonography targets the fastest portions of the sinus node by ablating the uppermost portion of the crista terminalis. This approach seems to require fewer radiofrequency applications than do electrogram-guided approaches. It may also reduce the need for permanent pacing.

Long-term follow-up after radiofrequency modification has been less encouraging. Recurrence rates are high. Ablation to the extent that permanent pacing is required may be necessary for sustained success. Thus, patients require careful follow-up for recurrent tachycardia or progressive sinus node dysfunction. Surgical isolation of the sinus node for IST has also been followed by recurrent tachycardia at new foci.

The author has received a grant for educational activities from the National Association for Sport and Physical Education and has during the past 12 months owned stock in or served as a consultant or member of the speakers' bureaus for AstraZeneca Pharmaceuticals LP; Bristol-Myers Squibb Company; Guidant Corporation; Medtronic, Inc.; The Procter & Gamble Company; Solvay Pharmaceuticals, Inc.; 3M; and Wyeth-Ayerst Laboratories, Inc.

References

1. Blomström-Lundqvist C, Scheinman MM, Aliot EM, et al: ACC/AHA/ESC Guidelines for the management of patients with supraventricular arrhythmias—executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Develop Guidelines for the Management of Patients with Supraventricular Arrhythmias). Circulation 108:1871, 2003

2. Mehta D, Wafa S, Ward DE, et al: Relative efficacy of various physical manoeuvres in the termination of junctional tachycardia. Lancet 1:1181, 1988

3. Glatter KA, Cheng J, Dorostkar P, et al: Electrophysiologic effects of adenosine in patients with supraventricular tachycardia. Circulation 99:1034, 1999

4. Kay GN, Pressley JC, Packer DL, et al: Value of the 12-lead electrocardiogram in discriminating atrioventricular nodal reciprocating tachycardia from circus movement atrioventricular tachycardia utilizing a retrograde accessory pathway. Am J Cardiol 59:296, 1987

5. Brugada P, Brugada J, Mont L, et al: A new approach to the differential diagnosis of a regular tachycardia with a wide QRS complex. Circulation 83:1649, 1991

6. Wellens HJ, Bar FW, Lie KI: The value of the electrocardiogram in the differential diagnosis of a tachycardia with a widened QRS complex. Am J Med 64:27, 1978

7. Akhtar M, Jazayeri MR, Sra J, et al: Atrioventricular nodal reentry: clinical, electrophysiological, and therapeutic considerations. Circulation 88:282, 1993

8. Sung RJ, Waxman HL, Saksena S, et al: Sequence of retrograde atrial activation in patients with dual atrioventricular nodal pathways. Circulation 64:1059, 1981

9. Amsterdam EA, Kulcyski J, Ridgeway MG: Efficacy of cardioselective beta-adrenergic blockade with intravenously administered metoprolol in the treatment of supraventricular tachyarrhythmias. J Clin Pharmacol 31:714, 1991

10. Lee KL, Chun HM, Liem LB, et al: Effect of adenosine and verapamil in catecholamine-induced accelerated atrioventricular junctional rhythm: insights into the underlying mechanism. Pacing Clin Electrophysiol 22:866, 1999

11. Overholt ED, Rheuban KS, Gutgesell HP, et al: Usefulness of adenosine for arrhythmias in infants and children. Am J Cardiol 61:336, 1988

12. Mauritson DR, Winniford MD, Walker WS, et al: Oral verapamil for paroxysmal supraventricular tachycardia: a long-term, double-blind randomized trial. Ann Intern Med 96:409, 1982

13. Winniford MD, Fulton KL, Hillis LD: Long-term therapy of paroxysmal supraventricular tachycardia: a randomized, double-blind comparison of digoxin, propranolol and verapamil. Am J Cardiol 54:1138, 1984

14. Flecainide acetate prevents recurrence of symptomatic paroxysmal supraventricular tachycardia. The Flecainide Supraventricular Tachycardia Study Group. Circulation 83:119, 1991

15. Long-term safety and efficacy of flecainide in the treatment of supraventricular tachyarrhythmias: the United States experience. The Flecainide Supraventricular Tachyarrhythmia Investigators. Am J Cardiol 70:11A, 1992

16. A randomized, placebo-controlled trial of propafenone in the prophylaxis of paroxysmal supraventricular tachycardia and paroxysmal atrial fibrillation. UK Propafenone PSVT Study Group. Circulation 92:2550, 1995

17. Yeh SJ, Lin FC, Chou YY, et al: Termination of paroxysmal supraventricular tachycardia with a single oral dose of diltiazem and propranolol. Circulation 71:104, 1985

18. Clague JR, Dagres N, Kottkamp H, et al: Targeting the slow pathway for atrioventricular nodal reentrant tachycardia: initial results and long-term follow-up in 379 consecutive patients. Eur Heart J 22:82, 2001

19. Scheinman MM, Huang S: The 1998 NASPE prospective catheter ablation registry. Pacing Clin Electrophysiol 23:1020, 2000

20. Ross DL, Uther JB: Diagnosis of concealed accessory pathways in supraventricular tachycardia. Pacing Clin Electrophysiol 7:1069, 1984

21. Dreifus LS, Haiat R, Watanabe Y, et al: Ventricular fibrillation: a possible mechanism of sudden death in patients and Wolff-Parkinson-White syndrome. Circulation 43:520, 1971

22. Klein GJ, Bashore TM, Sellers TD, et al: Ventricular fibrillation in the Wolff-Parkinson-White syndrome. N Engl J Med 301:1080, 1979

23. Timmermans C, Smeets JL, Rodriguez LM, et al: Aborted sudden death in the Wolff-Parkinson-White syndrome. Am J Cardiol 76:492,1995

24. Campbell RW, Smith RA, Gallagher JJ, et al: Atrial fibrillation in the preexcitation syndrome. Am J Cardiol 40:514, 1977

25. Krahn AD, Manfreda J, Tate RB, et al: The natural history of electrocardiographic preexcitation in men. The Manitoba Follow-up Study. Ann Intern Med 116:456, 1992

26. Murdock CJ, Leitch JW, Teo WS, et al: Characteristics of accessory pathways exhibiting decremental conduction. Am J Cardiol 67:506, 1991

27. Lai WT, Voon WC, Yen HW, et al: Comparison of the electrophysiologic effects of oral sustained-release and intravenous verapamil in patients with paroxysmal supraventricular tachycardia. Am J Cardiol 71:405, 1993

28. Flecainide acetate treatment of paroxysmal supraventricular tachycardia and paroxysmal atrial fibrillation: dose-response studies. The Flecainide Supraventricular Tachycardia Study Group. J Am Coll Cardiol 17:297, 1991

29. Cockrell JL, Scheinman MM, Titus C, et al: Safety and efficacy of oral flecainide therapy in patients with atrioventricular re-entrant tachycardia. Ann Intern Med 114:189, 1991

30. Kappenberger LJ, Fromer MA, Steinbrunn W, et al: Efficacy of amiodarone in the Wolff-Parkinson-White syndrome with rapid ventricular response via accessory pathway during atrial fibrillation. Am J Cardiol 54:330, 1984

31. Kunze KP, Schluter M, Kuck KH: Sotalol in patients with Wolff-Parkinson-White syndrome. Circulation 75:1050, 1987

32. Rosenbaum MB, Chiale PA, Ryba D, et al: Control of tachyarrhythmias associated with Wolff-Parkinson-White syndrome by amiodarone hydrochloride. Am J Cardiol 34:215, 1974

33. Glatter KA, Dorostkar PC, Yang Y, et al: Electrophysiological effects of ibutilide in patients with accessory pathways. Circulation 104:1933, 2001

34. Jackman WM, Wang XZ, Friday KJ, et al: Catheter ablation of accessory atrioventricular pathways (Wolff-Parkinson-White syndrome) by radiofrequency current. N Engl J Med 324:1605, 1991

35. Kuck KH, Schluter M, Geiger M, et al: Radiofrequency current catheter ablation of accessory atrioventricular pathways. Lancet 337:1557, 1991

36. Scheinman MM: NASPE Survey on Catheter Ablation. Pacing Clin Electrophysiol 18:1474, 1995

37. The Multicentre European Radiofrequency Survey (MERFS): complications of radiofrequency catheter ablation of arrhythmias. Multicentre European Radiofrequency Survey (MERFS) investigators of the Working Group on Arrhythmias of the European Society of Cardiology. Eur Heart J 14:1644, 1993

38. Lesh MD, Van Hare GF, Scheinman MM, et al: Comparison of the retrograde and transseptal methods for ablation of left free wall accessory pathways. J Am Coll Cardiol 22:542, 1993

39. Leitch JW, Klein GJ, Yee R, et al: Prognostic value of electrophysiology testing in asymptomatic patients with Wolff-Parkinson-White pattern. Circulation 82:1718, 1990

40. Pappone C, Santinelli V, Rosanio S, et al: Usefulness of invasive electrophysiologic testing to stratify the risk of arrhythmic events in asymptomatic patients with Wolff-Parkinson-White pattern: results from a large prospective long-term follow-up study. J Am Coll Cardiol 41:239, 2003

41. Kalman JM, Olgin JE, Karch MR, et al: "Cristal tachycardias": origin of right atrial tachycardias from the crista terminalis identified by intracardiac echocardiography. J Am Coll Cardiol 31:451, 1998

42. Natale A, Breeding L, Tomassoni G, et al: Ablation of right and left ectopic atrial tachycardias using a three-dimensional nonfluoroscopic mapping system. Am J Cardiol 82:989, 1998

43. Tang CW, Scheinman MM, Van Hare GF, et al: Use of P wave configuration during atrial tachycardia to predict site of origin. J Am Coll Cardiol 26:1315, 1995

44. Tada H, Nogami A, Naito S, et al: Simple electrocardiographic criteria for identifying the site of origin of focal right atrial tachycardia. Pacing Clin Electrophysiol 21:2431, 1998

45. Hsieh MH, Chen SA: Catheter ablation of focal AT. Catheter Ablation of Arrhythmias. Zipes DP, Haissaguerre M, Eds. Futura Publishing Co, Armonk, New York, 2002

46. Waldo AL: Pathogenesis of atrial flutter. J Cardiovasc Electrophysiol 9:S18, 1998

47. Yang Y, Cheng J, Bochoeyer A, et al: Atypical right atrial flutter patterns. Circulation 103:3092, 2001

48. Kalman JM, Olgin JE, Saxon LA, et al: Electrocardiographic and electrophysiologic characterization of atypical atrial flutter in man: use of activation and entrainment mapping and implications for catheter ablation. J Cardiovasc Electrophysiol 8:121, 1997

49. Lown B: Electrical reversion of cardiac arrhythmias. Br Heart J 29:469, 1967

50. Zeft HJ, Cobb FR, Waxman MB, et al: Right atrial stimulation in the treatment of atrial flutter. Ann Intern Med 70:447, 1969

51. Stambler BS, Wood MA, Ellenbogen KA: Comparative efficacy of intravenous ibutilide versus procainamide for enhancing termination of atrial flutter by atrial overdrive pacing. Am J Cardiol 77:960, 1996

52. Doni F, Manfredi M, Piemonti C, et al: New onset atrial flutter termination by overdrive transoesophageal pacing: effects of different protocols of stimulation. Europace 2:292, 2000

53. Ellenbogen KA, Stambler BS, Wood MA, et al: Efficacy of intravenous ibutilide for rapid termination of atrial fibrillation and atrial flutter: a dose-response study. J Am Coll Cardiol 28:130, 1996

54. Volgman AS, Carberry PA, Stambler B, et al: Conversion efficacy and safety of intravenous ibutilide compared with intravenous procainamide in patients with atrial flutter or fibrillation. J Am Coll Cardiol 31:1414, 1998

55. Fuster V, Ryden LE, Asinger RW, et al: ACC/AHA/ESC guidelines for the management of patients with atrial fibrillation: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients with Atrial Fibrillation) developed in collaboration with the North American Society of Pacing and Electrophysiology. Circulation 104:2118, 2001

56. Ellenbogen KA, Dias VC, Plumb VJ, et al: A placebo-controlled trial of continuous intravenous diltiazem infusion for 24-hour heart rate control during atrial fibrillation and atrial flutter: a multicenter study. J Am Coll Cardiol 18:891, 1991

57. Cosio FG, Goicolea A, Lopez-Gil M, et al: Catheter ablation of atrial flutter circuits. Pacing Clin Electrophysiol 16:637, 1993

58. Chen SA, Chiang CE, Wu TJ, et al: Radiofrequency catheter ablation of common atrial flutter: comparison of electrophysiologically guided focal ablation technique and linear ablation technique. J Am Coll Cardiol 27:860, 1996

59. Nabar A, Rodriguez LM, Timmermans C, et al: Radiofrequency ablation of "class IC atrial flutter" in patients with resistant atrial fibrillation. Am J Cardiol 83:785, 1999

60. Natale A, Newby KH, Pisano E, et al: Prospective randomized comparison of antiarrhythmic therapy versus first-line radiofrequency ablation in patients with atrial flutter. J Am Coll Cardiol 35:1898, 2000

61. McKenzie JP, Frazier DW, Smith JM, et al: Successful radio frequency ablation of inappropriate sinus tachycardia (abstr). Circulation 92:I, 1995

62. Low PA, Opfer-Gehrking TL, Textor SC, et al: Postural tachycardia syndrome (POTS). Neurology 45(4 suppl 5):S19, 1995

63. Kanjwal MY, Kosinski DJ, Grubb BP: Treatment of postural orthostatic tachycardia syndrome and inappropriate sinus tachycardia. Curr Cardiol Rep 5:402, 2003

64. Shen WK: Modification and ablation for inappropriate sinus tachycardia: current status. Card Electrophysiol Rev 6:349, 2002

VI VENTRICULAR ARRHYTHMIAS

Jonathan J. Langberg, m.d.
David B. DeLurgio, m.d.

Ventricular tachyarrhythmias characteristically are sudden in onset, unpredictable, and transitory. Consequently, their assessment and treatment present extraordinary challenges to the clinician. Moreover, the prognosis for patients with these arrhythmias is quite variable. In some patients, ventricular ectopic activity may be benign and without sequelae, but in other patients, comparable ectopy is a harbinger of ventricular fibrillation and sudden cardiac death.[1] This chapter summarizes the practical aspects of evaluation and treatment of patients with ventricular arrhythmias.

Pathophysiology

Ventricular tachyarrhythmias are mediated by one of three basic mechanisms: reentry, abnormal automaticity, and triggering. Although causation cannot be directly determined in individual patients, experimental and clinical observations make it possible to infer the mechanism underlying many of the ventricular arrhythmia syndromes encountered in practice.

Ventricular tachycardia caused by reentry

Reentrant arrhythmias (also called circus-movement tachycardias) are produced by a continuous circular or looping pattern of myocardial activation. Reentry can occur around lines of anatomic or functional block or occur as spinning wavefronts or rotors that lack a fixed anatomic path. When reentry occurs around lines of block, two features must be present for reentry to occur: (1) a barrier around which the wavefront circulates, either a fixed region of inexcitability caused by scarring or a dysfunctional region resulting from local refractoriness, and (2) unidirectional block at the entrance of the circuit. If activation spreads down both sides of the barrier, the impulses will collide distally and reentry will not occur; however, if propagation is blocked in one limb and proceeds in an anterograde direction over the other, the activation wavefront may be capable of retrograde invasion of the initially blocked pathway, thereby initiating sustained reentry.

In patients with structural heart disease, most symptomatic ventricular arrhythmias are mediated by reentry.[2,3] Sustained monomorphic ventricular tachycardia often occurs after transmural myocardial infarction (MI). The arrhythmia usually arises in the border zone of the scar [*see Figure 1*]. The larger the extent of this heterogeneous border zone, the greater the probability of a circuit capable of mediating reentrant ventricular tachycardia. This is consistent with the observation that the risk of malignant ventricular arrhythmias is proportional to the volume of the scar and the severity of left ventricular dysfunction after MI.[4]

Ventricular fibrillation is also a reentrant phenomenon.[5] Unlike ventricular tachycardia, during which a single activation wavefront circulates around a fixed barrier, ventricular fibrilla-

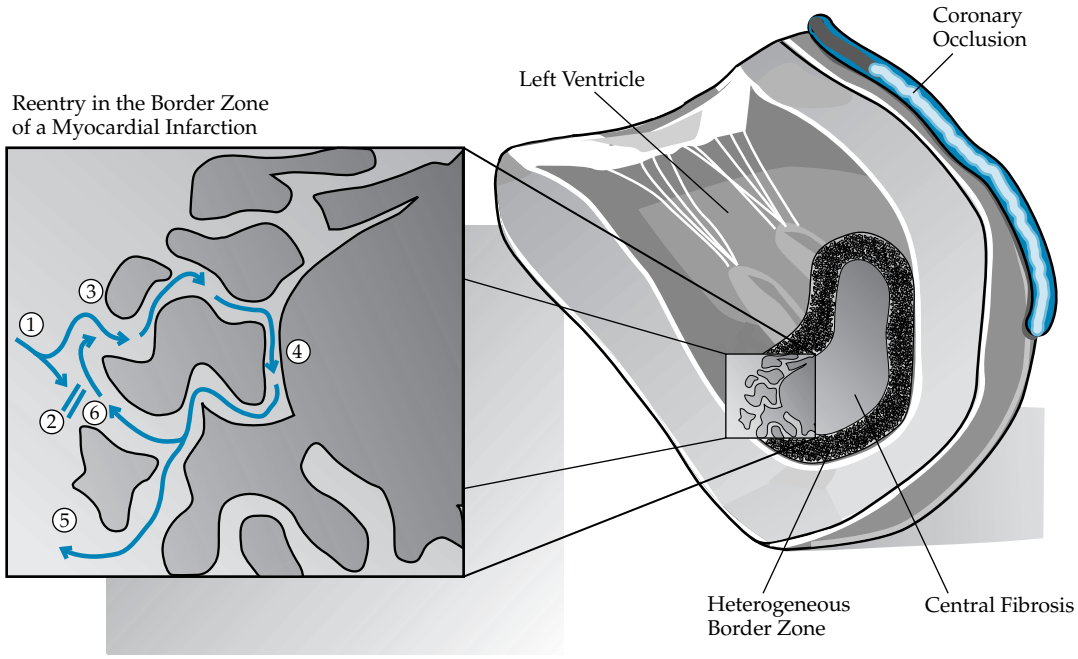


Figure 1 **Reentrant ventricular tachycardia usually arises as the result of reentry within the border zone of a myocardial infarction. This region consists of strands of viable myocytes interspersed with inexcitable fibrous tissue. Reentry begins when a wavefront of activation (1) encounters a bifurcation and blocks in one of the two pathways around an obstacle (2). The activation wavefront then conducts exclusively through the orthodromic pathway (3) and encounters a region of relatively slow conduction within the tachycardia circuit (4). The activation wavefront may exit from the tachycardia circuit at a site quite different from the entrance point (5). Although the anterograde limb of the circuit is initially refractory, it recovers excitability by the time it is depolarized by the reentrant wavefront (6). The activation wavefront reenters the orthodromic limb of the circuit, and the circus movement is established.**

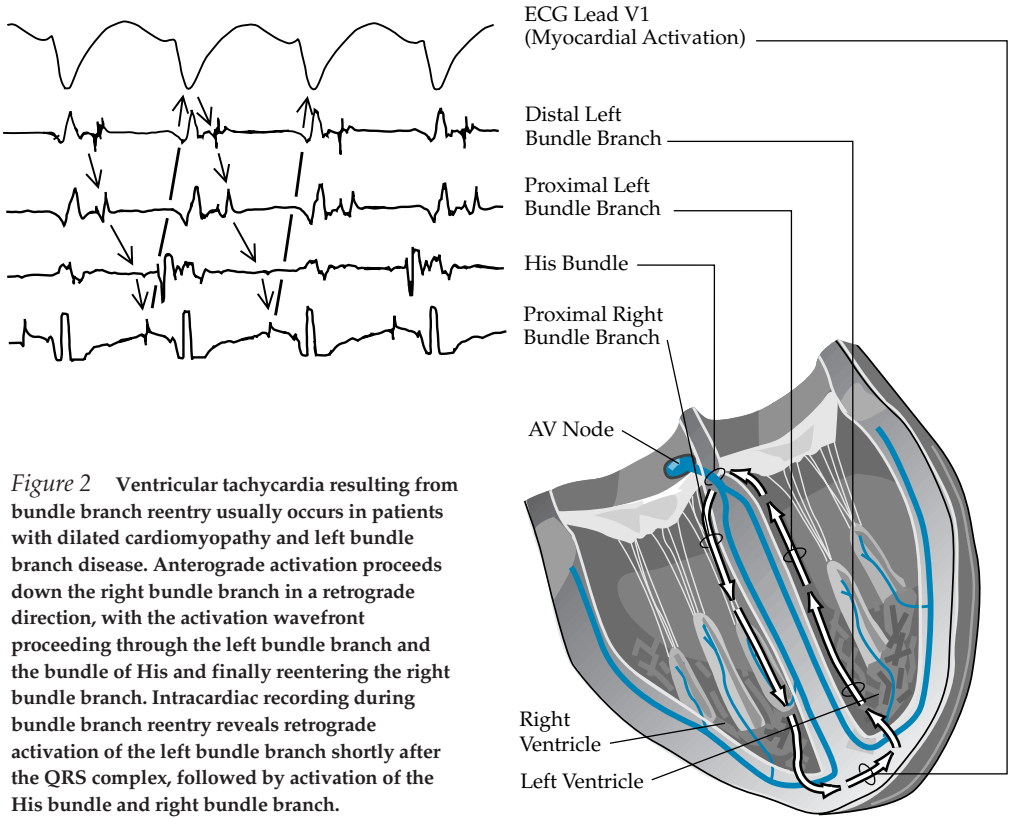


Figure 2 **Ventricular tachycardia resulting from bundle branch reentry usually occurs in patients with dilated cardiomyopathy and left bundle branch disease. Anterograde activation proceeds down the right bundle branch in a retrograde direction, with the activation wavefront proceeding through the left bundle branch and the bundle of His and finally reentering the right bundle branch. Intracardiac recording during bundle branch reentry reveals retrograde activation of the left bundle branch shortly after the QRS complex, followed by activation of the His bundle and right bundle branch.**

tion is caused by multiple simultaneous impulses that travel around functional barriers of refractory tissue, moving continuously throughout the myocardium to create very rapid, irregular, and ineffective activation. Alternatively, in some patients, ventricular fibrillation may be initiated by very early ectopic beats in the specialized conduction system.[6]

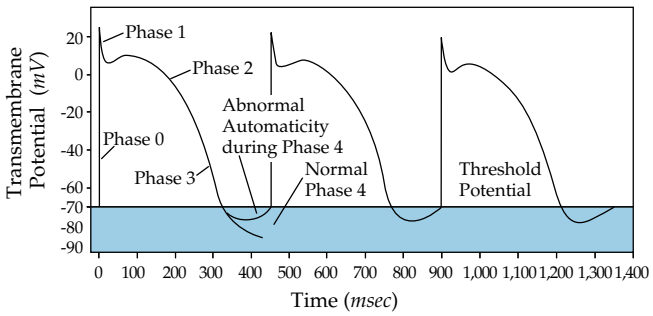


Figure 3 **The resting transmembrane potential of the myocardial cell is created by active maintenance of sodium and potassium gradients. The cell is depolarized (phase 0) by an electrical stimulus that allows a sudden influx of sodium (Na^+). Repolarization, phases 1 through 3, requires an early rapid chloride influx, a plateau phase mediated by calcium currents, and reestablishment of the resting transmembrane potential via potassium (K^+) efflux. Between action potentials, the resting potential is designated as phase 4. In cells with automaticity, depolarization mediated by calcium (Ca^{2+}) and Na^+ currents may occur during phase 4, resulting in spontaneous generation of the next action potential. In normal ventricular myocytes, the resting potential during electrical diastole (phase 4) remains in the region of –80 to –90 mV. The rate of automatic firing is determined by the resting potential, the slope of phase 4, and the threshold potential.**

Like postinfarction arrhythmias, the ventricular tachycardia in patients with nonischemic cardiomyopathy is often the result of reentry in a zone of patchy fibrosis. However, in patients with left ventricular dilatation and slowed conduction in the specialized conduction system, the tachycardia may be mediated by bundle branch reentry: anterograde conduction over the right bundle branch, activation of the septum, and retrograde conduction over the left bundle branch [*see Figure 2*].[7] Although an infrequent cause of ventricular tachycardia, bundle branch reentry is of interest to cardiac electrophysiologists because it can be cured by selective destruction of either the right or the left bundle branch by use of radiofrequency catheter ablation [*see 1:VII Pacemaker Therapy*].

VENTRICULAR TACHYCARDIA MEDIATED BY ABNORMAL AUTOMATICITY

Normal ventricular myocytes maintain a steady transmembrane resting potential of –80 to –90 mV, depolarizing only when stimulated by an activation wavefront. Extrinsic factors, such as electrolyte imbalance and ischemia, or intrinsic disease may reduce the resting potential and produce simultaneous diastolic (phase 4) depolarization [*see Figure 3*].

Unlike reentry, which can usually be induced and terminated by premature beats, automatic rhythms tend not to be influenced by pacing. Changes in heart rate at the onset of ventricular tachycardia may also provide insight into the arrhythmia mechanism. Reentrant tachycardias are usually stable because of a fixed conduction time around the circuit. In contrast, automaticity often shows warm-up, with progressive acceleration during the first few seconds of the tachycardia.

Abnormal automaticity may play a role in a number of clinical arrhythmia syndromes. An accelerated idioventricular rhythm

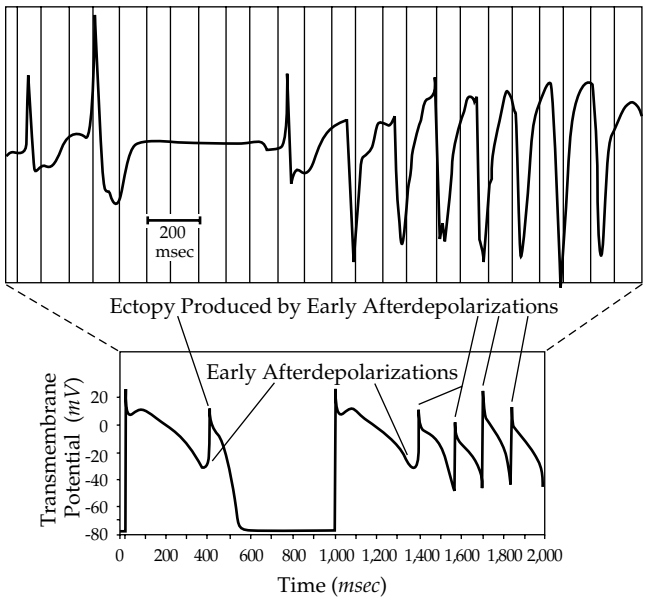


Figure 4 **In ventricular tachycardia caused by triggering, prolongation of the action potential (and the QT interval) results in depolarization during phase 3. Such early afterdepolarizations are manifested as positive deflections at the end of the phase 2 plateau or during the phase 3 rapid repolarization of the action potential. If this deflection exceeds the threshold potential, one or more triggered beats will occur. Bradycardia-dependent torsade de pointes is an example of an arrhythmia caused by early afterdepolarizations. The electrocardiogram of a patient with quinidine intoxication reveals an extrasystole and polymorphic ventricular tachycardia.**

(60 to 100 beats/min) or episodes of slow ventricular tachycardia (100 to 140 beats/min) occur in approximately 20% of patients who are monitored after transmural MI.[8] These slow-fast rhythms are probably the result of abnormal automaticity in ischemic Purkinje fibers.

More rapid ventricular tachycardia is also a frequent complication of acute ischemia, reperfusion, or both. These arrhythmias are often polymorphic, characterized by QRS complexes that change in amplitude and cycle length, with heart rates that may approach 300 beats/min. Abnormal automaticity in ischemic myocardium probably causes many of these episodes.

Ventricular tachycardia occasionally occurs in patients without apparent structural heart disease.[9] This idiopathic arrhythmia generally originates in the right ventricular outflow tract, just beneath the pulmonary valve. A number of observations suggest that it, too, is sometimes mediated by abnormal automaticity. It can develop spontaneously in response to increased adrenergic tone and, as a rule, cannot be induced or terminated by pacing. It may occur as a pattern of recurrent short bursts of tachycardia interspersed with equally short interludes of sinus rhythm, a pattern more consistent with automaticity than reentry.[10]

VENTRICULAR TACHYCARDIA CAUSED BY TRIGGERING

Early Afterdepolarization

Triggered activity, defined as premature activation caused by one or more preceding impulses, is the result of afterdepolarizations that occur either during (early afterdepolarization) or just after (delayed afterdepolarization) completion of the repolarization process [*see Figure 4*]. Factors that slow the heart rate tend to prolong the duration of depolarization, which is identified by a lengthened QT interval on the electrocardiogram, often sufficiently to bring early afterdepolarizations to threshold. Thus, triggered ventricular tachycardia that results from early afterdepolarizations is characteristically bradycardia dependent or pause dependent.

Early afterdepolarizations have been produced experimentally under a variety of conditions, including ischemia, hypokalemia, and antiarrhythmic drug toxicity. The arrhythmias seen in these studies are bradycardia dependent and, typically, are both rapid and polymorphic. Slowing of the tachycardia rate just before spontaneous termination is another characteristic feature of early afterdepolarization–mediated ventricular tachycardia.

Although it is difficult to prove, it seems likely that early afterdepolarizations mediate a variety of clinical arrhythmias. Prolongation of the QT interval—whether congenital or acquired as a result of drugs (class IA antiarrhythmic agents or, more commonly, other drugs such as haloperidol or erythromycin) or electrolyte depletion—increases the risk of a polymorphic ventricular tachycardia. As in the experimental situation, patients with QT prolongation tend to develop polymorphic ventricular tachycardia as a result of slowing of the heart rate, heart rate pauses, or sudden surges in adrenergic tone. Unlike rhythms mediated by automaticity or reentry, ventricular tachycardia in the setting of QT prolongation is almost always polymorphic, sometimes with the twisting pattern that characterizes torsade de pointes.

Delayed Afterdepolarization

Arrhythmias mediated by delayed afterdepolarization are distinctly different from those associated with early afterdepolarization and appear to be caused by abnormal accumulation and oscillation of cytosolic calcium concentration. The amplitude of these arrhythmias is augmented by acceleration rather than slowing of the heart rate. Delayed afterdepolarizations have been implicated in the genesis of ventricular tachycardia in patients with digitalis toxicity, and in some patients with ventricular tachycardia who have no apparent structural heart disease. Verapamil may be therapeutic in this subset of patients.[11] Although these arrhythmias have been recorded from surviving Purkinje fibers and infarcted canine myocardium, their role in clinical arrhythmias during and after MI is less well established.

Delayed afterdepolarizations are induced at a critical heart rate range, which is patient specific, either spontaneously or during atrial or ventricular pacing. As with reentrant arrhythmias, tachycardia resulting from delayed afterdepolarizations is often terminated by overdrive pacing, although it will frequently persist for several cycles after cessation of pacing.

Asymptomatic Ventricular Ectopy

Ventricular ectopy is recorded in more than half of normal persons undergoing ambulatory electrocardiographic monitoring. Complex ectopy (multifocal premature ventricular complexes and nonsustained ventricular tachycardia) is less frequent but is still observed in 5% to 10% of healthy persons with no apparent heart disease.[12]

The prognostic significance of ventricular ectopy depends on the severity of left ventricular dysfunction. In the absence of structural heart disease, asymptomatic ventricular ectopic activity is benign, with no demonstrable risk of sudden death, even in the presence of ventricular tachycardia. In patients with structural heart disease, however, ventricular ectopic activity is associat-

ed with an increased risk of sudden cardiac death. This risk is markedly increased with progressive left ventricular dysfunction.[13] For example, post-MI patients with a left ventricular ejection fraction (LVEF) greater than 40% who experience fewer than 10 ventricular premature complexes (VPCs) an hour after MI have a mortality of 5% to 7% a year. Those patients who experience more than 10 VPCs an hour, however, have a mortality of 12% to 18%. The combination of an LVEF of less than 40% and more than 10 VPCs an hour raises the annual mortality to between 27% and 40%.

The presence of frequent ventricular premature beats 7 to 10 days after MI is associated with a fivefold increase in the risk of symptomatic or fatal arrhythmias during follow-up.[4] Because many patients with frequent ectopy do not develop malignant ventricular arrhythmias, the positive predictive accuracy of this finding is only 16%. Conversely, because the majority of patients without frequent ectopy remain free of fatal arrhythmias, its absence is associated with a negative predictive accuracy of 82%. The occurrence of nonsustained ventricular tachycardia (fewer than three consecutive rapid beats over a period of less than 30 seconds) during monitoring appears to confer an even greater risk than does the presence of frequent isolated ventricular premature beats.[4,12,13]

The association between ambient ventricular ectopy and the risk of arrhythmic death is less well established in patients with nonischemic (i.e., valvular, hypertensive, or idiopathic) cardiomyopathy. However, most reports in the literature do suggest that the presence of high-grade ventricular arrhythmias, defined as multifocal VPCs or nonsustained ventricular tachycardia, confers an increased risk of sudden death that is independent of the severity of left ventricular dysfunction.[14,15]

Because the significance of ventricular ectopy depends on the degree of ventricular function impairment, cardiac imaging should be part of the initial evaluation. Echocardiography is the most versatile test; it provides information regarding regional wall motion abnormalities and valvular lesions as well as the LVEF. Radionuclide ventriculography also gives precise information regarding ejection fraction and may be of value in patients whose heart disease is already well characterized. If ventricular function is normal or close to normal, reassurance or treatment with beta blockers or calcium channel blockers to suppress bothersome symptoms is appropriate. In contrast, if patients have evidence of significant ventricular dysfunction or other significant structural heart disease such as hypertrophic cardiomyopathy, further evaluation and therapy may be appropriate. This evaluation may include additional tests that can help define the risk of a sustained arrhythmic event and the need for a prophylactic implantable cardioverter-defibrillator (ICD).

SIGNAL-AVERAGED ELECTROCARDIOGRAPHY

Signal-averaged electrocardiography may be useful for estimating risk in patients with heart disease and ventricular ectopy. This noninvasive test detects signals from areas of slow conduction in the arrhythmogenic regions on the periphery of an MI. The surface ECG is recorded for approximately 250 beats, and the signal is averaged by a computer and filtered, resulting in dramatic reduction of the signal-to-noise ratio. This allows detection of low-amplitude, high-frequency late potentials that result from the activation of zones of slow conduction just after the offset of the QRS complex.

Low-amplitude, high-frequency late potentials are recorded in about one third of patients after MI. These patients have a 20% incidence of life-threatening ventricular arrhythmias during the first year after infarction, compared with a 3% incidence in patients without late potentials.[16] Signal-averaged ECG findings are independently predictive of adverse events after MI and provide additional information regarding risks in patients with frequent ventricular premature contractions and impaired left ventricular function. A limitation of signal-averaged ECG is that it cannot be used in patients with bundle branch block or atrial fibrillation.

MICROVOLT T WAVE ALTERNANS

Another screening test that may be useful for assessing risk of sudden cardiac death in patients with left ventricular dysfunction is microvolt T wave alternans.[17] In this technique, signal processing is used to detect minute beat-to-beat variation in T wave amplitude that takes place during low-level exercise. Like signal-averaged ECG, microvolt T wave alternans has been approved by the Food and Drug Administration,[18] although currently it has little role in the selection of patients for ICDs. Unlike signal-averaged ECG, microvolt T wave alternans appears to have prognostic value in nonischemic as well as ischemic cardiomyopathy. A limitation of the test is that patients must be able to exercise for 5 minutes to a heart rate of at least 110 beats a minute.

ELECTROPHYSIOLOGIC TESTING

Electrophysiologic study can be used to assess the inducibility of sustained ventricular arrhythmias in patients with structural heart disease.[4] Electrode catheters are introduced percutaneously into the venous system, usually via the femoral vein, and advanced under fluoroscopic guidance into the right ventricle. Programmed electrical stimulation is performed in an attempt to elicit ventricular tachycardia or fibrillation. This usually consists of a drive train at a constant paced cycle length followed by one, two, or three extra stimuli (premature beats). The stimuli are introduced at progressively more premature coupling intervals until tachycardia is induced or the stimuli fail to capture as the result of local refractoriness [*see 1:VII Pacemaker Therapy*].

Programmed stimulation can induce sustained monomorphic ventricular tachycardia in about 20% of patients with reduced left ventricular function after MI and can induce ventricular fibrillation in an additional 10% to 15% of such patients. During follow-up, arrhythmic events occur in 5% of the noninducible patients, in 10% of patients with inducible ventricular fibrillation, and in 50% of patients with inducible ventricular tachycardia.

Although electrophysiologic study has reasonable sensitivity for prediction of subsequent arrhythmic events, the positive predictive value of the test is probably no better than that of the signal-averaged ECG, T wave alternans testing, or both, especially when such tests are combined with measurements of left ventricular systolic function and quantification of ambient ectopy. Electrophysiologic study is invasive and relatively expensive. Moreover, there is no evidence to suggest that treatment of this group of patients with antiarrhythmic drugs improves survival. Thus, it is difficult to justify routine electrophysiologic testing in asymptomatic patients after MI. The role of invasive electrophysiologic study for risk stratification in asymptomatic patients after MI has diminished with the increasing use of ICDs for primary prevention of sudden cardiac death.

Electrophysiologic testing is of uncertain value for stratification of risk in patients with nonischemic cardiomyopathy and asymptomatic ventricular ectopy. In this population, induction of sustained monomorphic ventricular tachycardia is infrequent

and does not appear to be predictive of subsequent sudden cardiac death.

Currently, electrophysiologic studies are perhaps most useful in patients who have an LVEF of 30% to 40% and evidence of nonsustained ventricular tachycardia. In this select group, electrophysiologic testing can help determine whether the risk of sudden cardiac death is sufficiently high to merit prophylactic implantation of an ICD [*see* The Implantable Cardioverter-Defibrillator, *below*].

Syncope and Ventricular Arrhythmias

Syncope, defined as transient loss of consciousness, is a common phenomenon, accounting for about 3% of all emergency room visits.[19] Because the spells usually resolve by the time the patient is initially evaluated, determination of the cause of loss of consciousness is difficult but extremely important, because prognosis depends on the nature of the episode. If ventricular arrhythmias are detected during subsequent monitoring, additional evaluation should be undertaken to determine whether the syncope was produced by a paroxysm of ventricular tachycardia.

HISTORY AND PHYSICAL EXAMINATION

A thorough history may provide important clues to the diagnosis of ventricular tachycardia. The onset of syncope mediated by ventricular tachycardia is usually abrupt, with only a brief prodrome of light-headedness or no premonitory symptoms at all. The absence of rapid heartbeat does not exclude the diagnosis, because only about one half of patients with documented sustained ventricular tachycardia experience this symptom. The duration of unconsciousness is brief, rarely lasting longer than several minutes. Because of the abrupt onset, traumatic injury is common.

Spontaneous movements during syncope often cause confusion and misdiagnosis. Cerebral hypoperfusion from any cause, including ventricular tachycardia, may produce one or more clonic jerks of the extremities. However, syncopal episodes differ from seizure activity in three respects: (1) the movements in syncopal episodes are not reciprocating (tonic-clonic), (2) they are much briefer in duration, and (3) bladder or bowel incontinence rarely occurs.

Historical information regarding the patient's condition after awakening is frequently overlooked but may be very helpful. Patients typically recover quickly from ventricular tachycardia–mediated syncope. Postictal confusion lasting longer than 5 minutes suggests a grand mal event rather than an arrhythmic one. Similarly, persistent residual malaise, nausea, and weakness are characteristic of a faint produced by the vasodepressor syndrome rather than arrhythmic syncope.

Ventricular tachycardia of sufficient rate or duration to produce loss of consciousness is rare in patients with normal ventricular function. Thus, patients in whom ventricular arrhythmias are identified after a syncopal episode must be thoroughly evaluated for structural heart disease. The presence of severe left ventricular dysfunction in these patients is associated with an ominous prognosis.

Patients with coronary artery disease, syncope, or ventricular arrhythmias require evaluation of myocardial ischemia with a functional study (e.g., thallium scintigraphy), coronary angiography, or both, in addition to quantification of ventricular function. Acute ischemia may precipitate rapid ventricular tachycardia that is sufficient to cause loss of consciousness. In such cases, exercise treadmill testing may induce ventricular ectopy, thereby suggesting the diagnosis, especially if premonitory symptoms are reproduced.

ELECTROCARDIOGRAPHY

On occasion, findings on a 12-lead ECG will suggest the cause of the loss of consciousness in a patient with unexplained syncope. A prolonged QT interval can indicate congenital long QT syndrome; ST segment elevation in lead V1 can indicate the Brugada syndrome; and a short QT interval may indicate short QT syndrome [*see* Heritable Ventricular Arrhythmias, *below*].

Signal-averaged electrocardiography plays a limited but important role in the evaluation of patients with syncope and ventricular arrhythmias. The positive predictive accuracy of this test is inadequate to confirm the diagnosis of an arrhythmic event. However, a negative result makes the possibility of sustained ventricular tachycardia unlikely enough that additional, more invasive studies are probably not justified.

Ambulatory electrocardiography is useful in selected patients with a history of syncope and ventricular arrhythmias. The yield of 24-hour or 48-hour Holter monitoring is low in patients with infrequent arrhythmic episodes, however. In such patients, a transtelephonic loop recorder is more likely to provide diagnostic information. This device is worn by the patient for 4 to 6 weeks, continuously recording and storing several minutes of the ECG in an endless loop. Immediately after presyncope or a syncopal spell, the patient presses the event button on the device to stop the recording and store the preceding ECG in memory. The output of the device is then transmitted over the telephone to a receiving station. This system has been shown to be more cost-effective than Holter monitoring and is preferable unless symptoms are present on a daily basis.

ELECTROPHYSIOLOGIC TESTS

Electrophysiologic testing can be useful in determining whether an episode of loss of consciousness was produced by ventricular tachycardia.[20] Assessment of sinus node function and atrioventricular conduction should be performed during electrophysiologic testing even when ventricular tachycardia is suspected, because episodic bradyarrhythmias may produce spells with very similar symptoms.

The induction of sustained monomorphic ventricular tachycardia during programmed stimulation increases the probability that the patient's spontaneous episode was mediated by ventricular tachycardia and increases the likelihood that therapy will be effective. Several studies have shown a lower rate of recurrent syncope in patients whose therapy is based on results of electrophysiologic testing, compared with those in whom the study was unrevealing or for whom no effective treatment could be found.[20,21]

Evaluation of the Patient Rescued from Cardiac Arrest

In 80% to 90% of patients who develop out-of-hospital cardiac arrest, the precipitating event is either primary ventricular fibrillation or a rapid ventricular tachycardia that degenerates into ventricular fibrillation. Bradyarrhythmic events occur occasionally, but when asystole is recorded as the initial rhythm, it is usually indicative of a prolonged downtime interval and is associated with a very poor prognosis.

The majority of patients who sustain cardiac arrest have structural heart disease. In industrialized societies, this is most often the result of coronary atherosclerosis. Studies of both victims

and survivors of cardiac arrest show significant coronary obstruction in 75% to 80% of patients. Unfortunately, sudden cardiac death is the initial manifestation of coronary artery disease in 10% to 20% of patients, making it the most common cause of mortality in adults younger than 65 years.[22]

Despite the close association between coronary artery disease and sudden cardiac death, acute MI is an infrequent cause of cardiac arrest. Only about 20% of patients rescued from an episode of ventricular fibrillation have evidence of an evolving MI during their subsequent hospitalization.[23] The prognosis is favorable for cardiac arrest survivors in whom the event can be clearly linked to acute myocardial ischemia, with a recurrence rate of only 2% during the subsequent year. In contrast, patients with ventricular fibrillation not related to an ischemic event have an annual recurrence rate of greater than 20%, presumably because they have a chronic substrate capable of mediating malignant ventricular arrhythmias.[22,23]

All patients rescued from cardiac arrest require serial ECGs and enzyme measurements to determine whether the event was a consequence of acute MI. Coronary angiography should be performed in all patients as well, except those in whom the precipitating factor has already been unequivocally identified.

ELECTROCARDIOGRAPHY

Laboratory evaluation of patients rescued from cardiac arrest should be directed at the identification of specific reversible causative factors. As in patients who have experienced syncope, the postresuscitation ECG may provide important information. A prolonged QT interval suggests the possibility of drug-induced torsade de pointes or the congenital long QT syndrome. A short PR interval and slurring of the QRS onset (a delta wave) are manifestations of the Wolff-Parkinson-White (WPW) syndrome [*see 1:V Supraventricular Tachycardia*]. Patients with WPW syndrome have an accessory connection linking the atrium and ventricle across either the mitral or the tricuspid annulus. A subset of patients with the WPW syndrome are capable of very rapid anterograde conduction over the accessory connection. If these patients develop atrial fibrillation, the ventricular response may be in excess of 300 beats/min and can degenerate into ventricular fibrillation.

LABORATORY TESTS

The initial evaluation of serum electrolytes is sometimes revealing, because severe depletion of serum potassium, serum magnesium, or both may precipitate ventricular arrhythmias. Such depletions are characteristic of patients with chronic heart failure who are maintained on long-term diuretic therapy with inadequate electrolyte supplementation.

ELECTROPHYSIOLOGIC TESTS

Electrophysiologic study was once an important part of the evaluation of cardiac arrest survivors in whom a reversible cause cannot be identified. With ICD therapy becoming commonplace for such patients, however, the usefulness of electrophysiologic testing has become limited to patients in whom the exact nature of the arrhythmia that precipitated the arrest remains uncertain. In a study of electrophysiology testing in 572 patients with ventricular fibrillation, ventricular tachycardia with syncope, or sustained ventricular tachycardia in the setting of left ventricular dysfunction, 67% of patients had inducible sustained ventricular tachycardia or ventricular fibrillation, but inducibility of these arrhythmias did not predict death or arrhythmia recurrence. These investigators concluded that electrophysiologic testing may not be worth the risks and costs of the procedure in this patient population, particularly in those patients likely to receive an ICD.[24]

Heritable Ventricular Arrhythmias

Alterations in the duration of the QT interval are most often acquired, typically from drugs.[25] In rare cases, however, ventricular arrhythmias result from genetic disorders that alter ventricular repolarization. These disorders include long QT syndrome, Brugada syndrome, and short QT syndrome.

LONG QT SYNDROME

A familial disorder with distinct clinical features, the congenital long QT syndrome usually presents as syncope (or, in rare instances, as cardiac arrest) during childhood or the teenage years, mediated by recurrent bouts of rapid, polymorphic ventricular tachycardia. Many patients are incorrectly diagnosed with a grand mal seizure disorder. Loss of consciousness characteristically occurs with a sudden surge in adrenergic tone caused by abrupt physical, emotional, or auditory stimulation. There is often a family history of unexplained syncope or premature sudden cardiac death.

The hallmark of this disorder is abnormal prolongation of the QT interval on the ECG. Prolongation is present if the heart rate–corrected QT interval (QT/RR interval) exceeds 0.47 in children, 0.46 in men, or 0.48 in women. Other depolarization abnormalities are often present in the long QT syndrome. The T wave is flattened and may have a bifid, or double-hump, appearance. In addition, a prominent U wave may be seen. About one third of patients will have a resting heart rate of less than 60 beats/min.

Congenital long QT syndrome has two principal phenotypes. The originally described Jervell and Lange-Nielsen syndrome, an autosomal recessive disorder with associated deafness, has proved to be quite rare.[26] The more common Romano-Ward syndrome is an autosomal dominant disorder and is not associated with hearing loss. Genomic studies in families with congenital long QT syndrome have shown that the disorder is produced by mutations of membrane ion channel proteins. To date, more than 300 of these mutations have been identified in seven genes, accounting for approximately 70% of affected patients.[27] Interestingly, the different mutations seem to produce slightly different ECG appearances.

Evaluation of a patient for the long QT syndrome should include screening of all first-degree relatives. A careful history regarding unexplained syncope and a 12-lead ECG should be obtained. A point system has been developed that combines ECG findings, clinical history, and family history (e.g., unexplained sudden death at a young age in an immediate family member) into a score that indicates the likelihood of disease.[28] Genetic testing for long QT syndrome is now commercially available and includes analysis of five major cardiac ion channel genes. The sensitivity of this test is approximately 70%, and its role in the management of affected patients and their families has not been established. Treatment of long QT syndrome is with beta blockers, ICD placement in high-risk patients, and left thoracic sympathectomy in selected cases.[29] Treatment of asymptomatic family members should be considered if screening uncovers a prolonged QT interval.

BRUGADA SYNDROME

Brugada syndrome is an inherited disorder that is manifested by syncope or sudden cardiac death. It is characterized by an

ECG that shows an incomplete right bundle branch block and ST segment elevation in leads V1 through V3.[30] However, these ECG findings also occur in many patients who do not have Brugada syndrome. Because of the low specificity of the ECG characteristics, the diagnosis should not be made on the basis of the ECG alone.[31] A set of diagnostic criteria that includes history, ECG and electrophysiologic test results, and family history has been proposed.[32] As with the long QT syndrome, genetic testing for Brugada syndrome is commercially available, but it is of uncertain utility. ICD placement is recommended for patients with Brugada syndrome who have experienced symptoms; recommendations for ICD placement are less well established for patients with Brugada syndrome who are asymptomatic or have inducible arrhythmias.

SHORT QT SYNDROME

Short QT syndrome is an inherited disorder characterized by a family history of sudden death (perhaps including sudden infant death syndrome), an abnormally short QT (QTc < 300 msec), and inducible ventricular fibrillation.[33,34] The syndrome has been traced to mutations in the cardiac ion channel genes.[35] Because short QT syndrome has only recently been recognized, its incidence remains uncertain. It is important to remember that electrolyte and drug effects that cause QT shortening (e.g., hypercalcemia and digitalis) need to be excluded before this diagnosis is entertained.

Pharmacologic Therapy

As a result of changes in the medical care system, more primary care practitioners bear direct responsibility for treatment decisions in patients with cardiac arrhythmias. The use of antiarrhythmic drugs in patients with ventricular arrhythmias presents a growing challenge, especially given that the medical literature contains reports of real and potential harm associated with the use of antiarrhythmic drugs.

CLASSIFICATION AND MECHANISMS OF ANTIARRHYTHMIC DRUGS

Antiarrhythmic drugs directly alter the electrophysiologic properties of myocardial cells. Therefore, an understanding of basic cellular electrophysiology is critical for an informed use of these compounds [*see Figure 5*].[36]

The most widely accepted classification of antiarrhythmic drugs, originally proposed by Vaughan Williams in 1970, involves four main classes of drugs, with the first class further divided into three subgroups [*see Table 1*].[37] This classification is based primarily on the ability of the drug to control arrhythmias by blocking ionic channels and currents. Few drugs demonstrate pure class effects, however, and other characteristics, such as influence of the drug on autonomic tone, contractility, and adverse effects, may be more important clinically and will be discussed as they pertain to individual drugs.

Class I agents inhibit the fast Na^+ channel during depolarization (phase 0) of the action potential, with resultant decreases in depolarization rate and conduction velocity [*see Figure 5*]. Agents in class IA (quinidine, procainamide, disopyramide, and moricizine) significantly lengthen both the action potential duration and the effective refractory period (and therefore the QT interval) through a combination of the class I effect of Na^+ channel inhibition and the lengthening of repolarization by K^+ channel blockade, a class III effect.

Class IB drugs (lidocaine, mexiletine, and phenytoin) are less powerful Na^+ channel blockers and, unlike class IA agents, shorten the action potential duration and refractory period in normal ventricular tissue, probably by inhibition of a background Na^+ current during phase 3 of the action potential.[38,39] In ischemic tissue, lidocaine may also block an adenosine triphosphate (ATP)–dependent K^+ channel, thus preventing ischemically mediated shortening of depolarization.[40]

Class IC drugs (flecainide and propafenone), the most potent Na^+ channel blockers, markedly decrease phase 0 depolarization rate and conduction velocity. Unlike other class I agents, they have little effect on the action potential duration and the effective refractory period in ventricular myocardial cells, but they do shorten the action potential of the Purkinje fibers.[41,42] This inhomogeneity of depolarization combined with marked slowing of conduction may contribute to the proarrhythmic effects of this class of drugs.

Class II agents are the beta-adrenergic antagonists. The efficacy of these drugs in the reduction of arrhythmia-related morbidity and mortality has become more evident in recent years, but the precise ionic bases for their salutary effects have not been fully elucidated. Beta-adrenergic antagonism has been shown to decrease spontaneous phase 4 depolarization and, therefore, to decrease adrenergically mediated automaticity, an effect that may be of particular importance in the prevention of ventricular arrhythmias during ischemia and reperfusion. Beta blockade also results in the slowing of heart rate and decreased oxygen consumption, effects long recognized as desirable in MI patients.[43] Effects on the cardiac action potential differ in atrial, ventricular, and specialized conduction tissues. For example, conduction velocity is slowed most profoundly in specialized conduction tissue, resulting in prolongation of the PR interval, whereas action potential duration in ventricular myocardium is generally not affected.

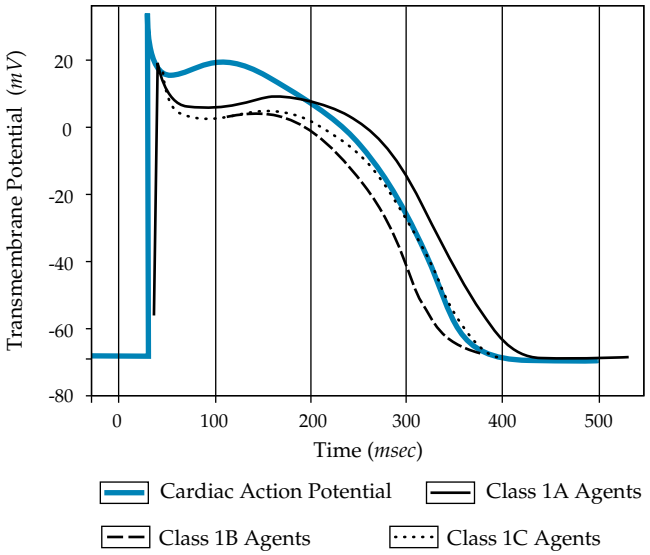


Figure 5 **The electrophysiologic hallmark of class I antiarrhythmic drugs is inhibition of the fast Na^+ channel, which results in a decrease in the slope and amplitude of phase 0 of the cardiac action potential. Class IA agents (quinidine, procainamide, and disopyramide) also prolong the action potential duration, whereas class IB agents (lidocaine and mexiletine) may shorten the action potential duration, particularly in ischemic tissue. Class IC agents (flecainide and propafenone) have little effect on action potential duration.**

Table 1 Classification of Antiarrhythmic Drugs

Class (Agents)	*Action*	*I.V. Dosage*	*Oral Dosage*	*Route of Elimination*	*Side Effects*
I	Inhibit membrane sodium channels; affect Purkinje fiber action potential during depolarization (phase 0)				
IA					
Quinidine	Slow the rate of rise of the action potential and prolong its duration; slow conduction; increase refractoriness	6–10 mg/kg (I.M. or I.V.) over 20 min	200–400 mg every 4–6 hr or every 8 hr (long-acting)	Hepatic	GI, ↓LVF, ↑Dig, torsade de pointes
Procainamide		100 mg every 1–3 min to 500–1,000 mg; maintain at 2–6 mg/min	50 mg/kg/day in divided doses every 3–4 hr or every 6 hr (long-acting)	Renal	SLE, hypersensitivity, ↓LVF, torsade de pointes
Disopyramide			100–200 mg every 6–8 hr	Renal	Urinary retention, dry mouth, markedly ↓LVF
Moricizine			200–300 mg every 8 hr	Hepatic	Dizziness, nausea, headache, ↓theophylline level, ↓LVF
IB					
Lidocaine	Shorten action potential duration; do not affect conduction or refractoriness	1–2 mg/kg at 50 mg/min; maintain at 1–4 mg/min		Hepatic	CNS, GI
Mexiletine			100–300 mg every 6–12 hr; maximum, 1,200 mg/day	Hepatic	CNS, GI, leukopenia
IC					
Flecainide	Slow the rate of rise of the action potential and slow repolarization (phase 4); slow conduction; increase refractoriness		100–200 mg twice daily	Hepatic	CNS, GI, ↓↓LVF, incessant VT, sudden death
Propafenone			150–300 mg every 8–12 hr	Hepatic	CNS, GI, ↓↓LVF, ↑Dig
II					
Beta blockers					
Esmolol	Inhibit sympathetic activity; decrease automaticity; prolong atrioventricular conduction and refractoriness	500 μg/kg over 1–2 min; maintain at 25–200 μg/kg/min	Other beta blockers may be used	Hepatic	↓LVF, bronchospasm
Propranolol		1–5 mg at 1 mg/min	40–320 mg in 1–4 doses (depending on preparation)	Hepatic	↓LVF, bradycardia, AV block, bronchospasm
Acebutolol			200–600 mg twice daily	Hepatic	↓LVF, bradycardia, positive ANA, lupuslike syndrome
III					
Amiodarone	Block potassium channels; predominantly prolong action potential duration, prolong repolarization, widen QRS complex, prolong QT interval, decrease automaticity and conduction, and prolong refractoriness	150 mg I.V. over 10 min, then 1 mg/min for 6 hr; maintain at 0.5 mg/min; overlap with initiation of oral treatment	800–1,600 mg/day for 7–21 days; maintain at 100–400 mg/day (higher doses may be needed)	Hepatic	Pulmonary fibrosis, hypothyroidism, hyperthyroidism, corneal and skin deposits, hepatitis, ↑Dig, neurotoxicity, GI
Sotalol			80–160 mg every 12 hr (higher doses may be used for life-threatening arrhythmias)	Renal (dosing interval should be extended if creatinine clearance < 60 ml/min)	↓LVF, bradycardia, fatigue and other side effects associated with beta blockers
Dofetilide			125–500 μg b.i.d.	Renal (dosing based on creatinine clearance)	Infrequent (rare CNS)
IV					
Verapamil	Slow calcium channel blockers; block the slow inward current; decrease automaticity and atrioventricular conduction	10–20 mg over 2–20 min; maintain at 5 μg/kg/min	80–120 mg every 6–8 hr; 240–360 mg once daily with sustained-release preparation (not approved for arrhythmia)	Hepatic	↓LVF, constipation, ↑Dig
Diltiazem		0.25 mg/kg over 2 min; second 0.35 mg/kg bolus after 15 min if response is inadequate; infusion rate, 5–15 mg/hr	180–360 mg daily in 1–3 doses, depending on preparation (oral forms not approved for arrhythmias)	Hepatic metabolism, renal excretion	Hypotension, ↓LVF

ANA—antinuclear antibodies AV—atrioventricular CNS—central nervous system ↑Dig—elevation of serum digoxin level GI—gastrointestinal (nausea, vomiting, diarrhea) ↓LVF—reduced left ventricular function SLE—systemic lupus erythematosus VT—ventricular tachycardia

The primary actions of class III agents (amiodarone, sotalol, and dofetilide) are prolongation of depolarization, the action potential duration, and the effective refractory period by K^+ channel blockade. These effects may prevent arrhythmias by decreasing the relative proportion of the cardiac cycle during which the myocardial cell is excitable and therefore susceptible to a triggering event. Reentrant tachycardias may be suppressed if the action potential duration becomes longer than the cycle length of the tachycardia circuit and if the leading edge of the wavefront suddenly impinges on inexcitable tissue. Class III agents have proven efficacy and an incidence of proarrhythmia lower than that seen with class IA agents.

Class IV agents act by inhibiting the inward slow Ca^{2+} current, which may contribute to late afterdepolarizations and therefore to ventricular tachycardia. These Ca^{2+} channel blockers reduce afterdepolarizations and are useful in the treatment of idiopathic ventricular tachycardia.[11,44,45] They have no appreciable effect on conduction velocity or repolarization and tend to evoke sympathetic activation. Thus, their role in the treatment of ventricular tachycardia in the setting of structural heart disease is limited.

Antiarrhythmic drugs in clinical use today have activity in multiple classes. For example, in addition to its class III effects, amiodarone also exhibits prominent Na^+ channel blockade (class I), beta blockade (class II), and Ca^{2+} channel blockade (class IV). Sotalol is a racemic mixture of d and l isomers, which have similar class III effects, whereas the l-isomer is essentially a beta blocker. d-Sotalol has been shown to increase mortality in patients with left ventricular dysfunction and recent MI.[46] The lower incidence of proarrhythmia seen with amiodarone or racemic sotalol therapy may be related to beneficial class II effects.

PROARRHYTHMIA

Proarrhythmia refers to the worsening of an existing arrhythmia or the induction of a new one by an antiarrhythmic drug. Three types of proarrhythmia have been described: torsade de pointes (the most common), incessant ventricular tachycardia, and extremely wide complex ventricular rhythm.

Torsade de Pointes

Torsade de pointes is triggered by early afterdepolarizations in a setting of delayed repolarization and increased dispersion of refractoriness. Class IA and class III drugs, which prolong refractoriness (and thus the QT interval) by K^+ channel blockade, provide the milieu for torsade de pointes. Drug-induced torsade de pointes is often pause dependent or bradycardia dependent, because the QT interval is longer at slower heart rates and after pauses. Exacerbating factors, such as hypokalemia, hypomagnesemia, and the concomitant use of other QT-prolonging drugs, are particularly important in this type of proarrhythmia.

Incessant Ventricular Tachycardia

Incessant ventricular tachycardia may be induced by drugs that markedly slow conduction (class IA and class IC) sufficiently to make the patient's own ventricular tachycardia continuous.[47,48] The arrhythmia is generally slower because of the drug effect, but it may become resistant to drugs or cardioversion, with potentially disastrous consequences in the presence of hemodynamic instability. This proarrhythmia is rarely associated with class IB drugs, which affect weaker Na^+ channel blockades.

Extremely Wide Complex Ventricular Rhythm

Extremely wide complex ventricular rhythm is usually associated with class IC agents, also in the setting of structural heart disease, and has been linked to excessive plasma drug levels or a sudden change in dose. The arrhythmia is not thought to represent a preexisting reentrant tachycardia and easily degenerates to ventricular fibrillation.

EFFICACY AND OUTCOMES OF ANTIARRHYTHMIC DRUG USE

Suppression of ambient ventricular ectopy by an antiarrhythmic agent does not prevent future life-threatening arrhythmias. In fact, patients effectively treated with class IC agents in the Cardiac Arrhythmia Suppression Trial (CAST) had a greater risk of sudden cardiac death than those who received placebo, a finding that underlines the proarrhythmic potential of these agents.[49] Conversely, beta blockers, which typically do not suppress ambient ectopy, appear to reduce the risk of malignant ventricular arrhythmias. A retrospective analysis of the CAST data showed that mortality related to arrhythmias, as well as from all causes, was reduced in patients who received beta blockers. The Electrophysiologic Study versus Electrocardiographic Monitoring (ESVEM) trial compared seven antiarrhythmic drugs and found that the risk of arrhythmia recurrence and cardiac mortality was greater with the class I agents than with sotalol.

As mentioned, patients with a history of MI and ventricular arrhythmias have an increased risk of fatal arrhythmias during follow-up. Meta-analysis of 138 trials involving 98,000 patients showed increased mortality with class I drugs.[50] Beta blockers have been conclusively associated with short-term and long-term survival in this population.[51] Therefore, all such patients should receive a beta blocker unless it is specifically contraindicated. In contrast, evidence that class IA and class IC agents increase mortality suggests that these drugs should be avoided in MI patients. Class IV agents have shown neither benefit nor harm.

Amiodarone is not associated with a significant survival benefit in MI patients, nor does it seem to be associated with an increased risk of sudden death. For example, the randomized, double-blind, placebo-controlled Canadian Amiodarone Myocardial Infarction Arrhythmia Trial (CAMIAT) was conducted in 1,202 MI patients with frequent or repetitive ventricular premature depolarizations. Resuscitated ventricular fibrillation or arrhythmic death occurred in 6.9% of patients in the placebo group and in 4.5% of those in the amiodarone group.[52]

Treatment of ventricular arrhythmias in patients with chronic heart failure is particularly challenging. The presence of a reduced ejection fraction and ventricular ectopy significantly increases the risk of sudden death. No antiarrhythmic drug has been shown to produce a significant survival benefit in this population. The proarrhythmic and negative inotropic effects of class IA and class IC drugs preclude their use in these patients. Amiodarone, overall, appears to be neutral in its effects. The Survival Trial of Antiarrhythmic Therapy in Congestive Heart Failure did not show significantly greater improvement in survival in patients treated with amiodarone than in those who received placebo, despite an antiarrhythmic effect.[15] Similarly, in the European Myocardial Infarct Amiodarone Trial (EMIAT), a randomized, double-blind, placebo-controlled trial conducted in 1,486 MI survivors with an LVEF of 40% or less, neither all-cause mortality nor cardiac mortality differed between the amiodarone and the placebo groups. The investigators noted, however, that the 35% risk reduction in arrhythmic deaths in the amiodarone group support the use of amiodarone in patients for whom antiarrhythmic therapy is indicated.[53] Therefore, the only indication for the use of amiodarone in patients with chronic heart failure appears

to be to suppress symptoms from frequent ectopy and nonsustained ventricular tachycardia. Improvement in survival requires ICD therapy.

Nonpharmacologic Therapy

SURGERY AND CATHETER ABLATION OF VENTRICULAR TACHYCARDIA

Surgical techniques for the treatment of ventricular tachycardia after MI were introduced in the late 1970s. However, these procedures are associated with relatively high perioperative mortality and require ventriculotomy, which can further compromise an already damaged ventricle. For these reasons, along with the increased simplicity of ICD implantation, surgical treatment is now rarely performed.

Radiofrequency catheter ablation has a role in selected patients with idiopathic ventricular tachycardia. Ablation is also useful for palliation in patients who have had an ICD implanted and are experiencing frequent shocks.

THE IMPLANTABLE CARDIOVERTER-DEFIBRILLATOR

The ICD automatically detects ventricular tachycardia or fibrillation and terminates the arrhythmia by overdrive pacing, high-energy shocks, or both. Since the first implantation of an ICD in a human, in 1980, the device has been utilized in hundreds of thousands of patients worldwide, and its use is growing exponentially.

All ICD systems contain three elements: the generator, rate-sensing leads, and electrodes to deliver high-energy shocks. In the early ICDs, defibrillating shocks were delivered via wire-mesh patch electrodes applied directly to the epicardial surface, and the generator was implanted subcutaneously in the abdomen. The implantation procedure required a thoracotomy and was associated with considerable morbidity and a perioperative mortality of 3% to 5%.[54] Current ICD models use transvenous leads, and the generator is implanted in a subcutaneous pocket in the anterior chest wall [*see Figure 6*].[55] ICD implantation is simple and safe, with a median duration of less than an hour and a median postoperative stay of 24 hours or less. The incidence of surgical complications is less than 2%—similar to that with routine pacemaker implantation.[55] As with modern pacemakers, the current generation of ICDs are multiprogrammable, microprocessor-based devices capable of automatically detecting ventricular tachycardia or fibrillation on the basis of timing information. The heart rate and duration of a tachycardia episode that will trigger overdrive pacing or shock therapy can be programmed. Additional detection enhancements can be used to reduce the probability that inappropriate pacing or shock will be delivered during episodes of sinus tachycardia or atrial fibrillation that exceed the programmed rate cutoff. The device can also be programmed to initiate therapy only if the heart rate increases abruptly during one cycle and only if the rate variability during the episode is less than a specified amount.

The ICD's output can also be tailored to suit patients' individual needs. For patients with a history of primary ventricular fibrillation, the ICD is programmed to deliver high-energy shocks when it detects tachycardia. Patients with a history of stable monomorphic ventricular tachycardia may benefit from overdrive pace termination. Cardioverting shocks will be delivered only if the specified number of pacing trains fails to terminate or if pacing accelerates the arrhythmia. Because overdrive pacing is associated with little or no discomfort, the device may be considered in patients with recurrent episodes of tolerated ventricular tachycardia.

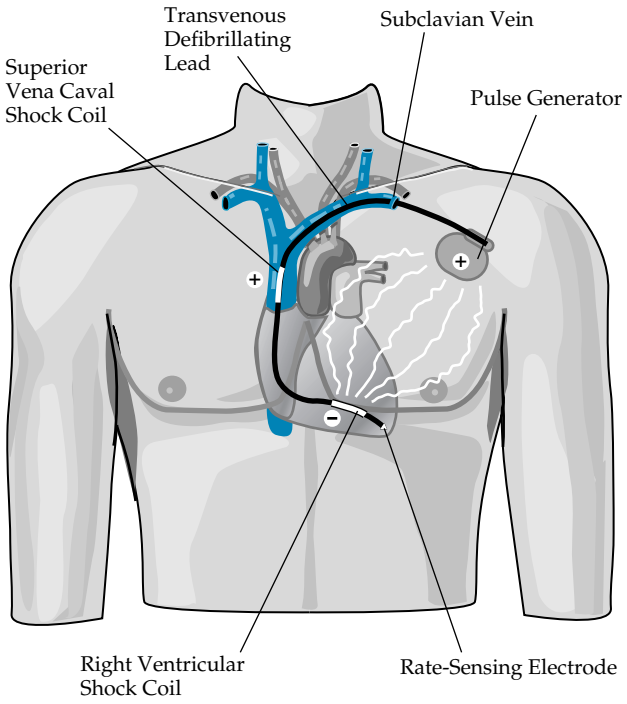


Figure 6 **An implantable cardioverter-defibrillator (ICD) consists of a pulse generator and one or more leads for cardioversion and defibrillation. The pulse generator is usually installed in a subcutaneous pocket in the pectoral region. It comprises a battery, capacitors, memory chips, integrated circuits and microprocessors, and a telemetry module, which are sealed within a titanium casing. A transvenous defibrillating lead from the pulse generator is inserted into the subclavian vein and advanced into the apex of the right ventricle. When a persistent ventricular tachyarrhythmia with a rate faster than the programmed rate cutoff is detected by the rate-sensing electrode in the lead's tip, the device charges and delivers a high-voltage defibrillating shock. For this purpose, the shock coil in the right ventricle serves as the cathode, whereas the proximal shock coil in the superior vena cava portion of the lead, plus the metal casing of the generator, serve as the anode. In older ICD models, the metal casing alone serves as the anode.**

The ICD also functions as a ventricular demand pacemaker, obviating a second device in patients with symptomatic bradyarrhythmias. This feature is also useful for prevention of the transitory bradycardia that sometimes occurs after delivery of a defibrillating shock.

The ICD has the capability of recording individual arrhythmia episodes. When tachycardia is detected, the device stores the electrograms in memory that can then be played back through the programmer at the time of a follow-up visit. This Holter function provides valuable diagnostic information regarding arrhythmia frequency, duration, rate, and response to therapy.

ICD Trials

ICDs were initially used for secondary prevention in survivors of cardiac arrest and in patients with documented life-threatening ventricular arrhythmias. Three large randomized trials have compared ICD therapy with pharmacologic treatment for the prevention of death in survivors of ventricular fibrillation or sustained ventricular tachycardia: the Antiarrhythmics vs Implantable Defibrillator (AVID) study,[56] the Cardiac Arrest Study

Table 2 Recommendations for Implantable Cardioverter-Defibrillator Therapy[66]

Recommendation Class	*Indication*	*Level of Evidence*
I	Cardiac arrest due to VF or VT not due to a transient or reversible cause	A
	Spontaneous sustained VT in association with structural heart disease	B
	Syncope of undetermined origin with clinically relevant, hemodynamically significant sustained VT or VF induced at electrophysiologic study when drug therapy is ineffective, not tolerated, or not preferred	B
	Nonsustained VT in patients with coronary disease, prior MI, LV dysfunction, and inducible VF or sustained VT at electrophysiologic study that is not suppressible by a class I antiarrhythmic drug	BA
	Spontaneous sustained VT in patients without structural heart disease not amenable to other treatments	C
IIa	LVEF ≤ 30% at least 1 mo after MI and 3 mo after coronary artery revascularization surgery	B
IIb	Cardiac arrest presumed to be due to VF when electrophysiologic testing is precluded by other medical conditions	C
	Severe symptoms (e.g., syncope) attributable to ventricular tachyarrhythmias in patients awaiting cardiac transplantation	C
	Familial or inherited conditions with a high risk for life-threatening ventricular tachyarrhythmias such as long QT syndrome or hypertrophic cardiomyopathy	B
	Nonsustained VT with coronary artery disease, prior MI, LV dysfunction, and inducible sustained VT or VF at electrophysiologic study	B
	Recurrent syncope of undetermined origin in the presence of ventricular dysfunction and inducible ventricular arrhythmias at electrophysiologic study when other causes of syncope have been excluded	C
	Syncope of unexplained origin or family history of unexplained sudden cardiac death in association with typical or atypical right bundle branch block and ST segment elevations (Brugada syndrome)	C
	Syncope in patients with advanced structural heart disease in whom thorough invasive and noninvasive investigations have failed to define a cause	C
III	Syncope of undetermined cause in a patient without inducible ventricular tachyarrhythmias and without structural heart disease	C
	Incessant VT or VF	C
	VF or VT resulting from arrhythmias amenable to surgical or catheter ablation (e.g., atrial arrhythmias associated with the Wolff-Parkinson-White syndrome, right ventricular outflow tract VT, idiopathic left ventricular tachycardia, or fascicular VT)	C
	Ventricular tachyarrhythmias due to a transient or reversible disorder (e.g., acute MI, electrolyte imbalance, drugs, or trauma) when correction of the disorder is considered feasible and likely to substantially reduce the risk of recurrent arrhythmia	CB
	Significant psychiatric illnesses that may be aggravated by device implantation or may preclude systematic follow-up	C
	Terminal illness with projected life expectancy less than 6 mo	C
	Patients with coronary artery disease with LV dysfunction and prolonged QRS duration in the absence of spontaneous or inducible sustained or nonsustained VT who are undergoing coronary bypass surgery	B
	NYHA class IV drug-refractory congestive heart failure in patients who are not candidates for cardiac transplantation	C

LV—left ventricular LVEF—left ventricular ejection fraction MI—myocardial infarction NYHA—New York Heart Association VF—ventricular fibrillation VT—ventricular tachycardia

Hamburg (CASH),[57] and the Canadian Implantable Defibrillator Study (CIDS).[58] A meta-analysis of the three trials showed consistent benefit from ICDs: patients who received ICDs had a significant reduction in death from any cause, with a summary hazard ratio (ICD:amiodarone) of 0.72; for arrhythmic death, the hazard ratio was 0.50.[59] Furthermore, 11-year follow-up of a subset of CIDS patients found that the benefit of the ICD over amiodarone increases with time; eventually, most amiodarone-treated patients develop side effects, experience recurrences of arrhythmia, or die.[60]

More recently, ICDs have been used for the primary prevention of sudden death. The first Multicenter Automatic Defibrillator Implantation Trial (MADIT I) compared ICD therapy with conventional medical therapy in MI patients with reduced ejection fraction, nonsustained ventricular tachycardia, and inducible nonsuppressible ventricular tachycardia on electrophysiologic testing. MADIT I showed that compared with conventional therapy, ICD therapy saved lives, with an ICD to non-ICD hazard ratio of 0.46.[61] The magnitude of the survival benefit increased with the severity of cardiac dysfunction.[62]

To study the role of ICDs in primary prevention, MADIT II enrolled MI patients with advanced left ventricular dysfunction (LVEF, 30% or less) who did not necessarily have manifest or inducible ventricular tachycardia. ICD implantation also increased survival in this population: over 20 months of follow-up, the ICD to non-ICD hazard ratio for death from any cause was 0.69.[63]

Additional trials have been performed to determine whether prophylactic ICD implantation is beneficial in all patients with chronic heart failure of any cause, ischemic or nonischemic. In the Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT), patients were evenly divided between those with ischemic and those with nonischemic cardiomyopathy. Patients who were receiving conventional treatment for heart failure were randomly assigned to supplemental therapy with amiodarone, placebo, or an ICD. Amiodarone did not increase survival, but simple, shock-only ICDs decreased mortality by 23%. The protective effect of the device was independent of the cause of the heart failure.[64] The Comparison of Medical Therapy, Pacing and Defibrillation in Heart Failure (COMPANION) trial was designed to compare medical therapy with a biventricular pacemaker and with a ventricular ICD in patients with advanced chronic heart failure and a wide QRS. Cardiac resynchronization with a biventricular pacemaker improved outcome, compared with medical therapy. Compared with biventricular pacing, biventricular ICD had an additional 21% survival benefit.[65]

Indications for ICD Implantation

Guidelines for the selection of patients for ICD implantation have been developed by the American College of Cardiology, the American Heart Association, and the National Association for Sport & Physical Education [*see Table 2*].[66] Since the 2002 revision of these guidelines, the indications for ICD implantation have expanded to include patients with nonischemic cardiomyopathy and an LVEF of less than 30%, even in the absence of symptomatic arrhythmias.[64]

AUTOMATED EXTERNAL DEFIBRILLATORS

An automated external defibrillator (AED) is a compact, easily portable device that can automatically analyze a patient's cardiac rhythm and, if it detects ventricular fibrillation, direct the rescuer to apply a shock. AEDs require minimal training to operate and are achieving widespread distribution.

Public-Access AEDs

AEDs can now be found in many public places, such as airports, stadiums, casinos, and large office buildings. Preliminary data suggest that these devices may confer a survival benefit,[67] although cost-effectiveness is difficult to calculate. It seems safe to say that the availability of public-access AEDs will result in increased numbers of patients successfully resuscitated from cardiac arrest, who will then require follow-up treatment.

Home AEDs

The FDA has approved several AED models for consumer use in the home, without a prescription, and these devices are now being marketed directly to the public for this purpose. The utility of home AEDs is uncertain, but the patients for whom these devices should be considered are those who meet the criteria for prophylactic ICD therapy but either have declined the implantation procedure or have comorbidities that make the implantation procedure inadvisable. The cost-effectiveness of these devices will be difficult to measure, and the potential medicolegal liability issues involved may be complex.

Wearable Automatic Defibrillators

An automatic defibrillator that is worn as a vest has been approved by the FDA.[68] This device is typically worn by patients who are awaiting heart transplants or who recently experienced an MI or underwent coronary revascularization. At our institution, we have used the device to provide temporary prophylaxis for a patient who required removal of an ICD because of site infection.

The authors serve as clinical investigators for Medtronic, Inc, Guidant Corporation, and St. Jude Medical, Inc.

References

1. Zipes DP: An overview of arrhythmias and antiarrhythmic approaches. J Cardiovasc Electrophysiol 10:267, 1999

2. de Bakker JM, van Capelle FLJ, Janse MJ, et al: Reentry as a cause of ventricular tachycardia in patients with chronic ischemic heart disease: electrophysiological and anatomic correlation. Circulation 77:589, 1988

3. de Bakker JM, Coronel R, Tasseron S, et al: Ventricular tachycardia in the infarcted, Langendorff-perfused human heart: role of the arrangement of surviving cardiac fibers. J Am Coll Cardiol 15:1594, 1990

4. Callans DJ, Josephson ME: Ventricular tachycardias in the setting of coronary artery disease. Cardiac Electrophysiology: From Cell to Bedside. Zipes DP, Jalife J, Eds. WB Saunders Co, Philadelphia, 1995, p 732

5. Kenknight BH, Bayly PV, Gerstle RJ, et al: Regional capture of fibrillating ventricular myocardium: evidence of an excitable gap. Circ Res 77:849, 1995

6. Haissaguerre M, Shoda M, Jais P, et al: Mapping and ablation of idiopathic ventricular fibrillation. Circulation 106:962, 2002

7. Caceres J, Jazayeri M, McKinnie J, et al: Sustained bundle branch reentry as a mechanism of clinical tachycardia. Circulation 79:256, 1989

8. Kaplinsky E, Ogawa S, Michelson EL, et al: Instantaneous and delayed ventricular arrhythmias after reperfusion of acutely ischemic myocardium: evidence for multiple mechanisms. Circulation 63:333, 1981

9. Wall TS, Freedman RA: Ventricular tachycardia in structurally normal hearts. Curr Cardiol Rep 4:388, 2002

10. Nibley C, Wharton JM: Ventricular tachycardias with left bundle branch block morphology. Pacing Clin Electrophysiol 18:334, 1995

11. Ohe T, Shimomura K, Aihara N, et al: Idiopathic left ventricular tachycardia: clinical and electrophysiological characteristics. Circulation 77:560, 1988

12. Gordon T, Kannel WB: Premature mortality from coronary heart disease: the Framingham Study. JAMA 215:1617, 1971

13. Myerburg RJ, Kessler KM, Castellanos A: Sudden cardiac death: epidemiology, transient risk and intervention assessment. Ann Intern Med 119:1187, 1993

14. Doval HC, Nul DR, Grancelli HO, et al: Randomized trial of low-dose amiodarone in severe congestive heart failure. Lancet 344:493, 1994

15. Singh SN, Fletcher RD, Fisher SG, et al: Amiodarone in patients with congestive heart failure and asymptomatic ventricular arrhythmia. Survival Trial of Antiarrhythmic Therapy in Congestive Heart Failure. N Engl J Med 333:77, 1995

16. Breithardt G, Schwarzmaier M, Borggrefe M, et al: Prognostic significance of late ventricular potentials after acute myocardial infarction. Eur Heart J 4:487, 1983

17. Hohnloser SH, Klingenheben T, Li YG, et al: T wave alternans as a predictor of recurrent ventricular tachyarrhythmias in ICD recipients: prospective comparison with conventional risk markers. J Cardiovasc Electrophysiol 9:1258, 1998

18. Kunavarapu C, Bloomfield DM: Role of noninvasive studies in risk stratification for sudden cardiac death. Clin Cardiol 27:192, 2004

19. Schnipper JL, Kapoor WK: Diagnostic evaluation and management of patients with syncope. Med Clin North Am 85:423, 2001

20. Krol RB, Morady FF, Flaker GC, et al: Electrophysiological testing in patients with unexplained syncope: clinical and noninvasive predictors of outcome. J Am Coll Cardiol 10:358, 1987

21. Denes P, Ezri MD: The role of electrophysiological studies in the management of patients with unexplained syncope. Pacing Clin Electrophysiol 8:424, 1985

22. Zheng ZJ, Croft JB, Giles WH, et al: Sudden cardiac death in the United States, 1989 to 1998. Circulation 104:2158, 2001

23. Bardy GH, Olsen WH: Clinical characteristics of spontaneous-onset sustained VT and VF in survivors of cardiac arrest. Cardiac Electrophysiology: From Cell to Bedside. Zipes DP, Jalife J, Eds. WB Saunders Co, Philadelphia, 1995, p 778

24. Brodsky MA, Mitchell LB, Halperin BD, et al: Prognostic value of baseline electrophysiology studies in patients with sustained ventricular tachyarrhythmia: the Antiarrhythmics Versus Implantable Defibrillators (AVID) trial. Am Heart J 144:478, 2002

25. Roden DM: Drug-induced prolongation of the QT interval. N Engl J Med 350:1013, 2004

26. Wehrens XH, Vos MA, Doevendans PA, et al: Novel insights in the congenital long QT syndrome. Ann Intern Med 137:981, 2002

27. Chiang CE: Congenital and acquired long QT syndrome: current concepts and management. Cardiol Rev 12:222, 2004

28. Schwartz PJ: The long QT syndrome. Curr Probl Cardiol 22:297, 1997

29. Khan IA: Long QT syndrome: diagnosis and management. Am Heart J 143:7, 2002

30. Brugada P, Brugada J: Right bundle branch block, persistent ST segment elevation and sudden cardiac death: a distinct clinical and electrocardiographic syndrome: a multicenter report. J Am Coll Cardiol 20:1391, 1992

31. Hermida JS, Lemoine JL, Aoun FB, et al: Prevalence of the Brugada syndrome in an apparently healthy population. Am J Cardiol 86:91, 2000

32. Wilde AA, Antzelevitch C, Borggrefe M, et al: Proposed diagnostic criteria for the Brugada syndrome. Eur Heart J 23:1648, 2000

33. Brugada R, Hong K, Dumaine R, et al: Idiopathic short QT interval: a new clinical syndrome? Cardiology 94:99, 2000

34. Faita F, Giustetto C, Bianchi F, et al: Short QT syndrome: a familial cause of sudden death. Circulation 108:965, 2003

35. Brugada R, Hong K, Dumaine R, et al: Sudden death associated with short-QT syndrome linked to mutations in HERG. Circulation 109:30, 2004

36. Singh SN, Patrick J, Patrick J: Antiarrhythmic drugs. Curr Treat Options Cardiovasc Med 6:357, 2004

37. Vaughan Williams EM: Cardiac Arrhythmias. Sandoe E, Fiensted-Jensen E, Olson KH, Eds. Astra, Sodertalje, Sweden, 1970, p 449

38. Singh BN, Opie LH, Marcus FI: Antiarrhythmic agents. Drugs for the Heart, Third Edition. Opie LH, Ed. WB Saunders Co, Philadelphia, 1991, p 180

39. Opie LH: The Heart: Physiology, Metabolism, Pharmacology and Therapy. Grune & Stratton, Orlando, 1984

40. Olschewski A, Brau ME, Olschewski H, et al: ATP-dependent potassium channel in rat cardiomyocytes is blocked by lidocaine. Circulation 93:656, 1996

41. Cowan JC, Vaughan Williams EM: Characterization of a new oral antiarrhythmic drug, flecainide (R818). Eur J Pharmacol 73:333, 1981

42. Ikeda N, Singh BN, Davis LD, et al: Effects of flecainide on the electrophysiological properties of isolated canine and rabbit myocardial fibers. J Am Coll Cardiol 5:303, 1985

43. Expert consensus document of beta-adrenergic receptor blockers. Task Force on Beta-Blockers of the European Society of Cardiology. Eur Heart J 25:1341, 2004

44. Takanaka C, Singh BN: Barium induced nondriver action potential as a model of triggered potentials from early afterdepolarizations: significance of slow channel activity and differing effects of quinidine and amiodarone. J Am Coll Cardiol 15:213, 1990

45. Gaita F, Giustetto C, Leclercq JF, et al: Idiopathic verapamil-responsive left ventricular tachycardia: clinical characteristics and long-term follow-up of 33 patients. Eur Heart J 15:1252, 1994

46. Waldo AL, Camm AJ, deRuyter H, et al: Survival with oral d-sotalol in patients with left ventricular dysfunction after myocardial infarction: rationale, design, and methods (the SWORD trial). Am J Cardiol 75:1023, 1995

47. Levine JH, Morganroth J, Kadish AH: Mechanisms and risk factors for proarrhythmia with type Ia compared with Ic antiarrhythmic drug therapy. Circulation 80:1063, 1989

48. Chaudhry GM, Haffajee CI : Antiarrhythmic agents and proarrhythmia. Crit Care Med 28(10 suppl):N158, 2000

49. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. The Cardiac Arrhythmia Suppression Trial (CAST) investigators. N Engl J Med 321:406, 1989

50. Teo KK, Yusuf S, Furburg CD: Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction. JAMA 270:1589, 1993

51. Kennedy HL, Brooks MM, Barker AH, et al: Beta blocker therapy in the Cardiac Arrhythmia Suppression Trial. CAST Investigators. Am J Cardiol 74:674, 1994

52. Cairns JA, Connolly SJ, Roberts R, et al: Randomised trial of outcome after myocardial infarction in patients with frequent or repetitive ventricular premature depolarisations: CAMIAT. Canadian Amiodarone Myocardial Infarction Arrhythmia Trial Investigators. Lancet 349:675, 1997

53. Julian DG, Camm AJ, Frangin G, et al: Randomised trial of effect of amiodarone on mortality in patients with left-ventricular dysfunction after recent myocardial infarction: EMIAT. European Myocardial Infarct Amiodarone Trial Investigators. Lancet 349:667, 1997

54. Bardy GH, Hofer B, Johnson G, et al: Implantable transvenous cardioverter-defibrillators. Circulation 87:1152, 1993

55. DiMarco JP: Implantable cardioverter-defibrillators. N Engl J Med 349:1836, 2003

56. A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. The Antiarrhythmics versus Implantable Defibrillators (AVID) Investigators. N Engl J Med 337:1576, 1997

57. Kuck KH, Cappato R, Siebels J, et al: Randomized comparison of antiarrhythmic drug therapy with implantable defibrillators in patients resuscitated from cardiac arrest: the Cardiac Arrest Study Hamburg (CASH). Circulation 102:748, 2000

58. Connolly SJ, Gent M, Roberts RS, et al: Canadian implantable defibrillator study (CIDS): a randomized trial of the implantable cardioverter defibrillator against amiodarone. Circulation 101:1297, 2000

59. Connolly SJ, Hallstrom AP, Cappato R, et al: Meta-analysis of the implantable cardioverter defibrillator secondary prevention trials: AVID, CASH and CIDS studies. Eur Heart J 21:2071, 2000

60. Bokhari F, Newman D, Greene M, et al: Long-term comparison of the implantable cardioverter defibrillator versus amiodarone: eleven-year follow-up of a subset of patients in the Canadian Implantable Defibrillator Study (CIDS). Circulation 110:112, 2004

61. Moss AJ, Hall WJ, Cannom DS, et al: Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter Automatic Defibrillator Implantation Trial Investigators. N Engl J Med 335:1933, 1996

62. Moss AJ, Fadl Y, Zareba W, et al: Survival benefit with an implanted defibrillator in relation to mortality risk in chronic coronary heart disease. Am J Cardiol 88:516, 2001

63. Moss AJ, Zareba W, Hall WJ, et al: Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. Multicenter Automatic Defibrillator Implantation Trial II Investigators. N Engl J Med 346:877, 2002

64. Bardy GH, Lee KL, Mark DB, et al: Amiodarone or an implantable cardioverter-defibrillator for congestive heart failure. N Engl J Med 352:225, 2005

65. Bristow MR, Saxon LA, Boehmer J, et al: Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) Investigators. N Engl J Med 350:2140, 2004

66. ACC/AHA/NSAPE 2002 Guideline Update for Implantation of Cardiac Pacemakers and Antiarrhythmia Devices: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/NASPE Committee on Implantation). American College of Cardiology, 2002
http://www.acc.org/clinical/guidelines/pacemaker/II_implantable.htm

67. Caffrey SL, Willoughby PJ, Pepe PE, et al: Public use of automated external defibrillators. N Engl J Med 347:1242, 2002

68. Auricchio A, Klein H, Geller CJ, et al: Clinical efficacy of the wearable cardioverter-defibrillator in acutely terminating episodes of ventricular fibrillation. Am J Cardiol 81:1253, 1998

Acknowledgments

Figures 1, 2, and 7 Joseph Bloch, CMI.

Figures 3 through 6 Marcia Kammerer.

VII PACEMAKER THERAPY

Jonathan Lowy, m.d.
Roger A. Freedman, m.d.

Worldwide, more than 250,000 permanent cardiac pacemakers are implanted each year. As the population ages and as indications for pacemakers expand, the number of implants continues to increase. Advances in technology have played an important role in the evolution of pacemaker therapy: currently available pacemakers are smaller and more reliable than older models and contain a multitude of sophisticated programmable features.

Normal Cardiac Electrical System

The primary role of cardiac pacing is to augment or replace the heart's intrinsic electrical system. This specialized system consists of structures capable of automaticity and conduction and provides the timing and synchrony needed to maintain appropriate cardiac output.

SINOATRIAL NODE

In normal circumstances, the sinoatrial (SA) node (also referred to as the sinus node) is the origin of impulse generation and dictates the intrinsic heart rate. The SA node is located in the superior aspect of the right atrium. It is composed of specialized tissue that demonstrates the fastest rate of spontaneous depolarization (automaticity) of any of the cardiac tissues.

ATRIOVENTRICULAR NODE

The atrioventricular (AV) node is the junction between the atria and the ventricular conduction system. This node is a dense and complex structure that plays three important roles. First, it demonstrates spontaneous depolarization and is capable of acting as an auxiliary pacemaker. Second, it delays propagation of the impulse between the atria and the ventricles, thereby allowing normal atrioventricular synchrony. Third, it acts as a filter, limiting the number of impulses that can be propagated from the atria to the ventricles and protecting the heart from rapid ventricular rates.

HIS-PURKINJE SYSTEM

The His-Purkinje system originates at the inferior border of the AV node. From this point, the bundle of His courses down the interventricular septum, where it diverges into the left and right bundle branches and terminates in the Purkinje fiber network. The bundle of His and the bundle branches provide rapid and synchronous depolarization of the ventricles. The Purkinje fibers serve as the interface between the specialized conduction system and the local ventricular myocardium.

MODULATION OF HEART RATE

The basal heart rate is maintained by the balance between sympathetic and parasympathetic tone. Changes in the heart rate are mediated by the autonomic nervous system and circulating catecholamines. There is a normal physiologic acceleration of the heart rate that results from increased demand for cardiac output. This acceleration is mediated by both increased sympathetic tone and reduced parasympathetic tone. Inability to increase the heart rate in response to increased demand for cardiac output can result in a number of symptoms, including fatigue, poor exercise tolerance, and exertional dyspnea.

Disruption or imbalance of sympathetic and parasympathetic inputs to the SA node or the AV node can cause profound abnormalities in the heart rate, resulting in inappropriate increases or decreases that give rise to significant symptoms. SA node dysfunction may be caused by intrinsic abnormalities of the conduction system or by imbalances in autonomic tone.

Indications for Permanent Pacing

GENERAL CONSIDERATIONS

The cardiac conduction system can be affected by any of a wide variety of pathologic states, ranging from benign abnormalities to conditions that can lead to severe symptoms and substantial morbidity and mortality.

Guidelines for permanent pacemaker implantation were established by a joint task force of the American College of Cardiology and the American Heart Association and were first published in 1984.[1] These guidelines were subsequently revised in 1991,[2] 1998,[3] and 2002.[4] The North American Society of Pacing and Electrophysiology (NASPE) was also involved in the 2002 revision [*see Table 1*]. Current recommendations are divided into the following three broad categories on the basis of (1) the strength of the available data and (2) the consensus of experts in the field:

- Class I: conditions for which there is evidence or general agreement that a given procedure or treatment is beneficial, useful, and effective.
- Class II: conditions for which there is conflicting evidence or a divergence of opinion about the usefulness or efficacy of a procedure or treatment.
  IIa: conditions for which the weight of the evidence or expert opinion is in favor of usefulness/efficacy.
  IIb: conditions for which usefulness or efficacy is less well established by evidence or opinion.
- Class III: conditions for which there is evidence or general agreement that a procedure or treatment is not useful or effective and, in some cases, may be harmful.

GUIDELINES FOR SPECIFIC PACEMAKER INDICATIONS

Acquired Atrioventricular Block

AV block is defined as delayed or failed conduction from the atria to the ventricles.[5-9] It is usually categorized as occurring either at or below the level of the AV node. First-degree AV block describes conduction delay from the sinus impulse to the ventricles and is defined as prolongation of the PR interval without a dropped QRS complex. Usually, first-degree AV block occurs at the level of the AV node, though it may also occur in the His-Purkinje system.

Second-degree AV block is present when some, but not all, P waves are conducted to the ventricles. It can be further subdivided into Mobitz type I (Wenckebach) and Mobitz type II. In type I second-degree AV block, there is a progressive prolongation of the PR interval preceding a nonconducted P wave. The anatom-

Table 1 Guidelines for Permanent Pacemaker Implantation

Condition	*Indications for Pacing*			
	Class I	*Class IIa*	*Class IIb*	*Class III*
Acquired AV block	Third-degree AV block or advanced second-degree AV block associated with any of the following: Symptomatic bradycardia Medical conditions requiring medications that result in symptomatic bradycardia (e.g., beta blockers, calcium channel blockers, antiarrhythmic agents) Asymptomatic asystole ≥ 3 sec or escape rate < 40 beats/min in awake patient Ablation of AV junction Postoperative AV block not expected to resolve after cardiac surgery Neuromuscular disease, including myotonic muscular dystrophy, Kearns-Sayre syndrome, Erb dystrophy, and peroneal muscular atrophy, with or without symptoms of bradycardia Second-degree AV block, regardless of type, with documented associated bradycardia	Asymptomatic third-degree AV block with average ventricular rate ≥ 40 beats/min when awake Asymptomatic type II second-degree AV block with narrow QRS complex Asymptomatic type I second-degree AV block found during electrophysiologic study performed for another reason First- or second-degree AV block with symptoms similar to those of pacemaker syndrome	Marked first-degree AV delay > 30 msec in patients with left ventricular dysfunction and congestive symptoms of heart failure Neuromuscular disease, including myotonic muscular dystrophy, Kearns-Sayre syndrome, Erb dystrophy, and peroneal muscular atrophy, with or without symptoms of bradycardia, with any degree of AV block, with or without symptoms	Asymptomatic first-degree AV block Asymptomatic type I (Wenckebach) second-degree AV block Any AV block that is expected to resolve and does not
Chronic bifascicular and trifascicular block	Intermittent third-degree AV block Type II second-degree AV block Alternating bundle-branch block	Syncope in which other causes (specifically, ventricular tachycardia) have been excluded but that has not been demonstrated to be due to AV block Asymptomatic patients in whom electrophysiologic study reveals prolonged HV interval Electrophysiologic study finding of nonphysiologic block below His bundle	Neuromuscular diseases with any degree of fascicular block, with or without symptoms, in which there is unpredictable progression	Fascicular block without AV block or symptoms Fascicular block with first-degree AV block without symptoms
Myocardial infarction	Persistent second-degree AV block in His-Purkinje system with bifascicular block or third-degree AV block within or below His-Purkinje system after acute MI Persistent and symptomatic second- or third-degree AV block Transient advanced second- or third-degree infranodal AV block and associated bundle-branch block; electrophysiologic study may be indicated to identify level of block	None	Persistent second- or third-degree block at level of AV node	Transient AV block in absence of intraventricular conduction delay Transient AV block in presence of isolated left anterior fascicular block Acquired left anterior hemiblock in absence of AV block Persistent first-degree AV block in presence of old bundle-branch block
SA node dysfunction	SA node dysfunction with documented symptomatic bradycardia Symptomatic SA node dysfunction resulting in bradycardia that occurs as consequence of essential drug therapy to which there is no acceptable alternative Symptomatic chronotropic incompetence	SA node dysfunction occurring either spontaneously or as a result of drug therapy with heart rates < 40 beats/min where there is clear association between symptoms but where actual presence of bradycardia during symptoms has not been documented Syncope of unexplained origin in which major abnormalities of SA node are elicited during electrophysiologic studies	Patients with minimal symptoms with resting heart rates < 40 beats/min while awake	Asymptomatic patients, including those on drug therapy with resting heart rates < 40 beats/min Patients with symptoms of bradycardia in which SA node dysfunction is clearly not associated with symptoms SA node dysfunction with symptomatic bradycardia caused by unnecessary drug therapy
Neurocardiogenic syncope and hypersensitive carotid sinus syndrome	Recurrent syncope caused by carotid sinus massage that results in ventricular asystole ≥ 3 sec (must occur in absence of any medication that depresses SA node or AV conduction)	Recurrent syncope without another cause and cardioinhibitory response to carotid sinus massage Symptomatic and recurrent neurocardiogenic syncope associated with documented bradycardia	None	Hyperactive cardioinhibitory response to carotid sinus massage in absence of symptoms or in presence of vague symptoms Recurrent symptoms in absence of documented cardioinhibitory response Situational vasovagal syncope in which avoiding behavior or environmental factors is effective

ic site of the block is usually the AV node, and the QRS complex is usually narrow. In type II second-degree AV block, there is a fixed PR interval preceding the dropped QRS complex. Type II block is often accompanied by bundle branch block, and its anatomic location is usually below the AV node in the His-Purkinje system.

When every other P wave is conducted, 2:1 AV block is present; 2:1 block cannot be classified as either type I or type II block, because there are not consecutive PR intervals preceding the nonconducted P wave. When 2:1 block is accompanied by bundle branch block, the site of the block is likely to be below the AV node in the His-Purkinje system. High-degree (or advanced) type II AV block is defined as blockage of two or more consecutive P waves. Complete heart block, or third-degree block, denotes a complete absence of conduction from the atria to the ventricles.

The anatomic location of AV block has important prognostic implications. Typically, a block occurring at the level of the AV node—such as first-degree block, type I second-degree block, and 2:1 block at the level of the AV node—does not typically lead to abrupt complete heart block, though gradual progression is common. A block occurring below the level of the AV node, on the other hand, can often progress quickly to complete heart block. In addition, high-degree or complete heart block at the level of the AV node is often ameliorated by junctional escape rhythms, whereas escape rhythms are much less reliable when the block is at the level of the His-Purkinje system.

Chronic Bifascicular and Trifascicular Block

The conduction system below the AV node is composed of three fascicles: the right bundle branch, the left anterior fascicle, and the left posterior fascicle. The left anterior and left posterior fascicles are divisions of the left bundle branch. Bifascicular block denotes blockage of the right bundle and either the left anterior or the left posterior fascicle; trifascicular block is present when alternating bundle branch block is seen or when right bundle branch block occurs in conjunction with alternating left anterior and left posterior hemiblock.[10] Trifascicular block may also be present when bifascicular block is accompanied by first-degree AV block. More commonly, however, this electrocardiogram pattern is the result of bifascicular block combined with conduction delay at the AV node.

Acute Myocardial Infarction

Conduction abnormalities are common in the setting of acute myocardial infarction.[11-15] Pathophysiologic mechanisms include ischemia, necrosis, autonomic influences, and the neurohumoral response to injury. Temporary transvenous pacing is often required during the acute phase of an infarction. The need for temporary pacing does not, however, predict the need for permanent pacing, given that many of the conduction abnormalities are transient and resolve after revascularization or upon recovery from the acute phase of the infarction.

Patients with acute inferior infarction can manifest a variety of abnormalities, including SA node dysfunction, first-degree AV block, type I second-degree block, and third-degree block at the level of the AV node. It is uncommon for any of these conduction disturbances to persist after the acute phase of the infarction. These patients often require temporary pacing if they manifest hemodynamic instability, but they rarely require permanent pacing.

Patients with anterior infarction can manifest bundle branch block, bifascicular block, trifascicular block, type II second-degree block, or complete heart block. These patients are much more likely to require permanent pacing than those with inferior infarction are. Although conduction abnormalities are associated with higher mortality in the setting of anterior infarction, the increased mortality is a consequence of the larger infarct size and is not directly related to the conduction abnormality.

SA Node Dysfunction

SA node dysfunction is a loose term that includes a number of different arrhythmias, including sinus bradycardia, sinus arrest, sinoatrial block, and the bradycardia-tachycardia syndrome.[16-20] The bradycardia-tachycardia syndrome is characterized by atrial tachyarrhythmias (usually atrial fibrillation) alternating with periods of bradycardia or sinus pauses. SA node dysfunction must be differentiated from the physiologic sinus bradycardia seen in trained athletes. During sleep, sinus rates as low as 30 beats/min and type I second-degree AV block are commonly seen in normal persons.

Pacing for Neurocardiogenic Syncope and Hypersensitive Carotid Syndrome

Neurocardiogenic syncope is syncope secondary to vasodilatation or bradyarrhythmias resulting from abrupt imbalance of autonomic input to the heart and the vascular system.[21-25] Classic neurocardiogenic syncope involves sinus tachycardia followed by bradycardia, vasodilatation, and syncope. Some patients have primarily a vasodepressive (vasodilatation) syndrome, whereas others have a syndrome with a significant cardioinhibitory component (bradycardia). Thus, bradycardia is not always a contributing component in neurocardiogenic syncope. Head-up tilt testing is often useful for diagnosing the presence and type of neurocardiogenic syncope.

The hypersensitive carotid syndrome is characterized by a similar abnormal response of the autonomic nervous system, in which baroreceptors in the carotid sinus trigger a vasodepressive or cardioinhibitory response. A hyperactive carotid sinus response is defined as a sinus pause longer than 3 seconds or a substantial symptomatic decrease in systolic blood pressure.

Other Pacemaker Indications

Besides those already mentioned, there are several indications for which pacemakers are implanted that warrant mention, including treatment of hypertrophic cardiomyopathy, prevention or suppression of tachyarrhythmias, and resynchronization therapy for congestive heart failure. Cardiac resynchronization therapy is an exciting new development in the treatment of heart failure but lies outside the scope of this chapter.

Pacemaker Systems

A basic pacemaker system is made up of three main components: the pulse generator, the pacemaker lead(s), and the programmer.

PULSE GENERATOR

Over the past 30 years, pulse generators have evolved from large, bulky devices into small, sophisticated systems [*see Figure 1*]. All pulse generators contain hardware, software, and a battery; however, the systems currently available can differ from one another with respect to a number of factors, including number of chambers, biventricular pacing capability, presence and type of activity sensor, size, battery life, and cost. All of these fac-

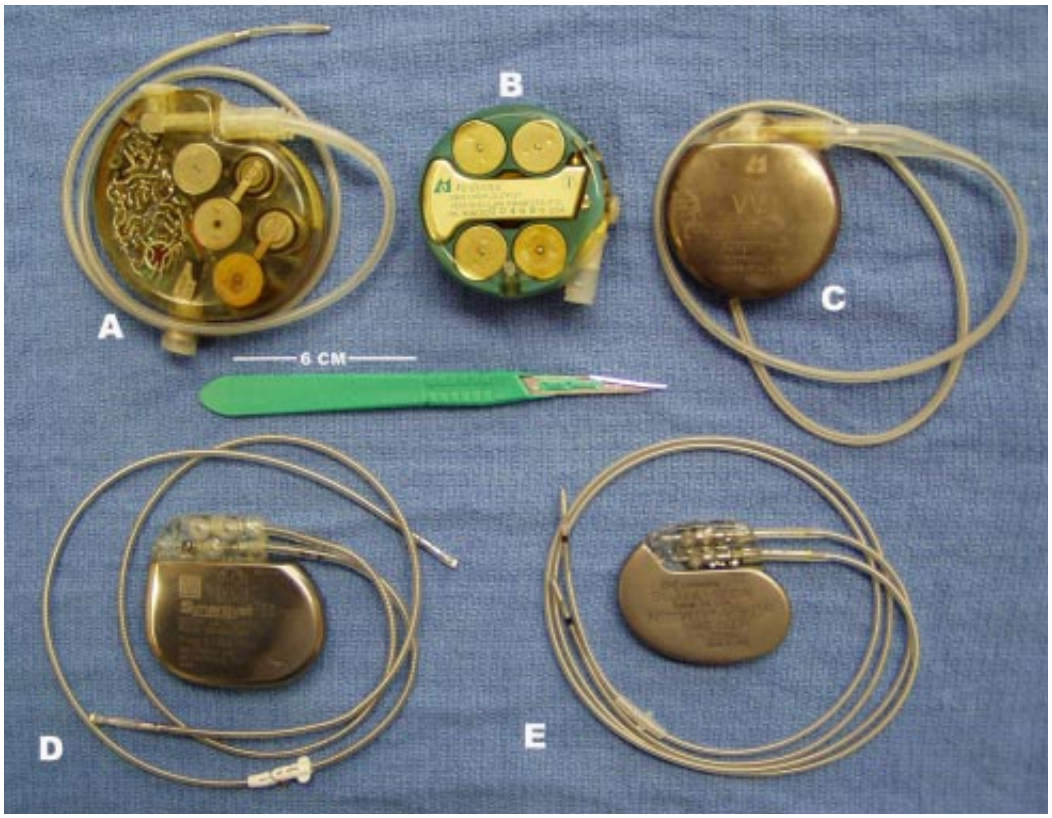


Figure 1 **Shown are five different pacemaker generators. The first three are older single-chamber devices from (*a*) 1972, (*b*) 1977, and (*c*) 1983. The last two are modern dual-chamber devices from (*d*) 1994 and (*e*) 2000.**

tors are taken into account in selecting a specific generator for a specific patient.

Generators are usually described as being either single-chamber or dual-chamber. Single-chamber systems have one lead, which is usually placed in the right ventricle (though it may, on occasion, be placed in the atrium). Dual-chamber systems have two leads, one of which is implanted in the right atrium and the other in the right ventricle. The biventricular pacemaker devices currently used in patients with heart failure have a third lead that is usually placed in a branch of the coronary sinus to provide left ventricular pacing. Dual-chamber systems can be programmed to single-chamber modes of operation.

At present, most generators currently use lithium iodine batteries that have a typical life span of 5 to 10 years. These batteries are not rechargeable or replaceable; accordingly, when the battery reaches the end of its life, a new generator must be implanted.

PACEMAKER LEADS

Pacemaker leads are the conduits from the generator to the myocardium. Most leads are implanted transvenously. There are still occasional applications for epicardial leads, but these are generally limited to patients with mechanical tricuspid valves, certain congenital heart abnormalities, or other conditions that preclude transvenous leads. Like pulse generators, leads have gone through a complex evolution since they were first developed. Various types are currently used [*see Figure 2*]; the major differences among them have to do with type of insulation, fixation mechanism, and polarity.

Most pacemaker leads are insulated with either silicone or polyurethane. In the past, there were significant differences between the two materials with respect to durability and handling. Today, however, the differences are minimal, and the choice of material is usually operator dependent.

Leads can be attached to the myocardium via either passive or active fixation. Passive-fixation leads usually have tines at the distal tip to help maintain stability. Active-fixation leads have a corkscrew helix mechanism at the distal end, which inserts into the myocardium. Both fixation mechanisms are reliable, and lead dislodgment is uncommon with either one.

Finally, leads can be either unipolar or bipolar. Unipolar leads have a single conductor and a single electrode; the unipolar pacing circuit involves the single electrode and the metal housing of the generator. Bipolar leads have two conductors and two electrodes; the pacing circuit is between the two electrodes. Advantages of unipolar leads include decreased diameter and reduced susceptibility to lead fracture. Advantages of bipolar leads include reduced risk of inappropriate sensing of myopotentials, greater resistance to electromagnetic interference (EMI), less likelihood of pectoral muscle stimulation, and better compatibility with implanted defibrillators. At present, bipolar leads are more commonly used, but unipolar leads are still employed on occasion.

Currently available lead systems are very reliable: failure rates at 5 years are typically 5% or lower.

PACEMAKER PROGRAMMER

The programming computer allows telemetric communication with the implanted pulse generator and serves as the interface between the health care provider and the pacemaker. Because there is no standardization among pacemaker manufacturers, each company's device requires its own programmer.

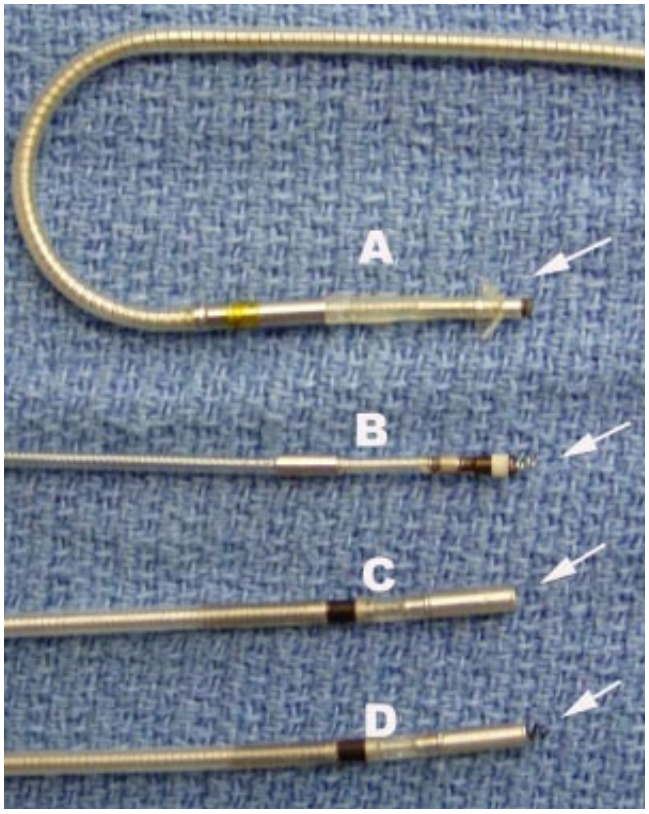


Figure 2 **Shown are four different pacemaker leads. The first (*a*) is a passive-fixation lead with soft tines at the tip (arrow); it is also a preformed J lead used for atrial pacing. The second (*b*) is an active-fixation lead with a fixed helix. The third (*c*) and fourth (*d*) are active-fixation leads with a retractable helix; the fourth has the helix mechanism exposed.**

Programmers are equipped with a wand that provides external telemetry through the skin, thus allowing direct communication with the pacemaker generator and access to the software contained within it. The pacemaker programmer is used to perform a multitude of functions, including assessing battery status, modifying pacemaker settings, and providing access to diagnostic information the pacemaker has stored (e.g., heart rate trends and tachyarrhythmia documentation).

PACEMAKER MAGNETS

Pacemaker generators are designed to respond to the placement of a strong magnet over the device. The response of most pacemakers is to pace at a set "magnet rate" in an asynchronous mode. Magnets also can be used to perform any of a number of functions designated by the manufacturer, including checking battery life, threshold testing, and obtaining event snapshots (in much the same way as an event monitor). Magnets should be available in the hospital and clinic, as well as on code carts for immediate access.

Although such use is beyond the scope of this chapter, it is worth mentioning that magnets can also temporarily turn off defibrillation therapy in implantable cardioverter-defibrillators.

Pacemaker Programming

Detailed description of specific programming techniques and indications is beyond the scope of this chapter; however, familiarity with the basic functions and nomenclature is critical for understanding how pacemakers function.

BASIC FUNCTIONS

A pacemaker has three basic functions: pacing, sensing, and action. Its other, more complicated functions are based on these three. Pacing is the delivery of an electrical impulse to the myocardium to elicit depolarization. Sensing is the ability to "see" intrinsic depolarization (i.e., the local intrinsic electrical signal that passes by the tip of the lead). Action is the response of the pacemaker to a sensed event—namely, either inhibition or triggering of a paced event.

CODES

The basic functions—pacing, sensing, and action—are determined by basic pacemaker programming. In 1974, the American Heart Association and the American College of Cardiology proposed a three-letter code for describing the basic functions of pacemakers. Under the guidance of NASPE and the British Pacing and Electrophysiology Group (BPEG), this code evolved into the five-position code currently in use [*see Table 2*].[26] The first position denotes the chamber or chambers paced; the second denotes the chamber or chambers sensed; the third denotes the action or actions performed; the fourth denotes rate response; and the fifth denotes multiple-site pacing. The simplest mode of pacing is VVI, otherwise known as ventricular demand pacing or ventricular inhibited pacing. The most commonly used mode in dual-chamber pacing is DDD.

TIMING CYCLES

A pacemaker is governed by timing cycles, which are a hierarchy of clocks that regulate how the pacemaker functions. The most basic timing cycle is the lower rate, which reflects how long the pacemaker will wait after a paced or sensed beat before initi-

Table 2 NASPE-BPEG Generic Five-Position Code for Antibradycardia Pacing

	Position				
	I	*II*	*III*	*IV*	*V*
Parameter measured	Chamber(s) paced	Chamber(s) sensed	Response or action	Rate modulation	Multisite pacing
Possible values	O = None A = Atrium V = Ventricle D = Dual (A + V)	O = None A = Atrium V = Ventricle D = Dual (A + V)	O = None I = Inhibited T = Triggered D = Dual (I + T)	O = None R = Rate response on	O = None A = Atrium V = Ventricle D = Dual (A + V)

NASPE—North American Society of Pacing and Electrophysiology BPEG—British Pacing and Electrophysiology Group

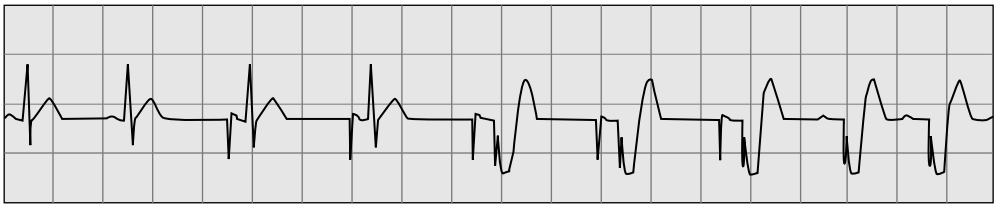


Figure 3 **Illustrated are different forms of DDD pacing. In the first two beats (labeled "Inhibited"), the pacemaker senses both the intrinsic P wave and the QRS complex; the result is inhibition of pacing. In the next two beats (labeled "Atrial"), there is a pacing spike preceding each P wave; the result is atrial pacing. The intrinsic QRS complex is then sensed, and ventricular pacing is inhibited. In the third set of beats (labeled "Sequential"), there are pacing spikes preceding both the P wave and the QRS complex. Both chambers are paced. The paced QRS morphology is noticeably different from the intrinsic complexes seen in the previous examples. In the final set of beats (labeled "Tracking"), an intrinsic P wave is followed by a paced QRS. The intrinsic atrial beat is sensed and triggers ventricular pacing.**

ating pacing. If the pacemaker is set to VVI mode at a lower rate of 60 beats/min, then as long as the interval between intrinsic beats is less then 1,000 msec, the pacemaker will reset the lower rate clock with each sensed QRS complex, and pacing will not occur. If, however, the intrinsic heart rate falls below 60 beats/min, the pacemaker's lower-rate clock will time out before an intrinsic beat is sensed, and pacing will occur. After a paced beat, the lower-rate clock is reset and the cycle repeats. In a modern dual-chamber pacemaker, there are a number of additional timing cycles that regulate how the pacemaker responds to these paced and sensed events [*see Figure 3*].

Patients with chronic atrial fibrillation and slow ventricular response are generally treated with single-chamber ventricular pacemakers. Such devices are also occasionally used in patients with isolated SA node dysfunction.

Pacemaker Implantation

Most pacemakers are implanted by cardiologists, and most implantation procedures are performed in the cardiac catheterization laboratory.[27]

PREPROCEDURAL CONSIDERATIONS

There are several issues that should be considered after the need to implant a pacemaker has been established. In particular, the patient's underlying health must be assessed and any comorbid conditions evaluated.

In select patients, the issue of reversal and reinitiation of oral anticoagulation must be addressed before implantation. In the past, all patients receiving warfarin had their international normalized ratios (INRs) normalized before the procedure. Furthermore, patients with a strong indication for anticoagulation (e.g., a mechanical heart valve) required prolonged hospitalization for reinitiation of oral anticoagulation after the procedure. In the past few years, however, favorable results have been reported with routine pacemaker implantation in patients undergoing therapeutic anticoagulation with warfarin. These results suggest that preprocedural reversal of anticoagulation may not be necessary.[28-29]

Pacemakers can interfere with or preclude certain imaging procedures, such as mammography and magnetic resonance imaging. In the case of elective pacemaker implants, a baseline mammogram should be performed beforehand.[30] Any MRI pro-

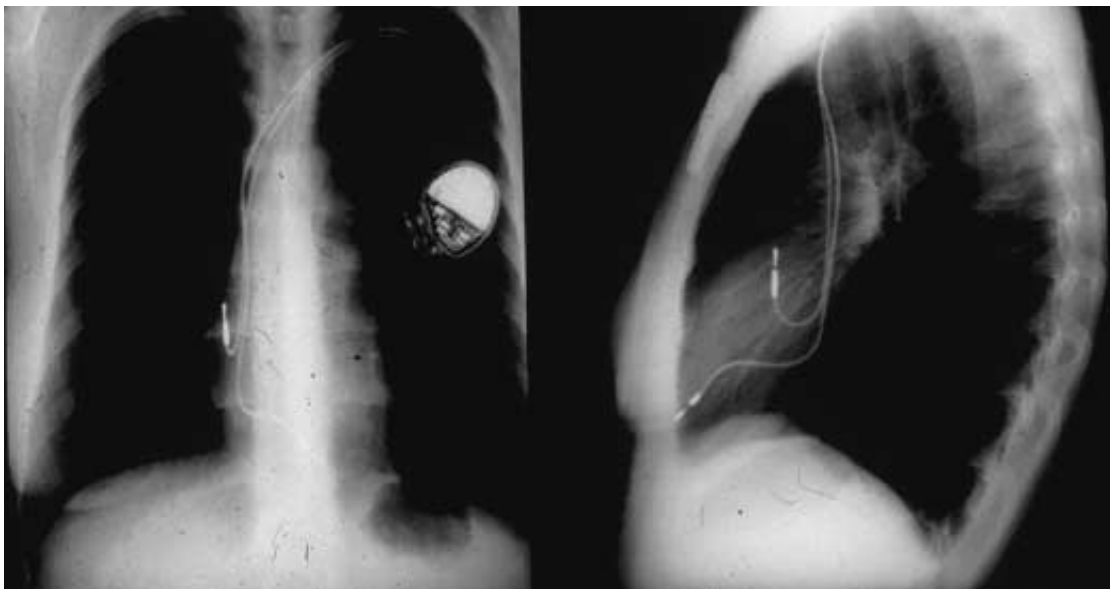


Figure 4 **Shown is the typical appearance of a dual-chamber pacemaker on posteroanterior (left) and lateral (right) chest x-rays. The RV lead is at the apex, and the RA lead is in the right atrial appendage.**

cedures that may be indicated should also be performed before implantation.

Local anesthesia is typically employed in conjunction with parenteral sedation. In certain circumstances (e.g., in pediatric patients or other patients who would tolerate the procedure poorly under local anesthesia), an anesthetist should be involved, but such circumstances are relatively uncommon. Antibiotic prophylaxis is commonly employed, but not in a uniform manner. There are no strict guidelines, and antibiotic regimens vary greatly.[31]

PACEMAKER POCKET PLACEMENT

The pulse generator pocket is usually placed on the upper left aspect of the chest, just medial to the angle of the deltopectoral grove and 2 to 3 cm below the clavicle. In the case of left-handed patients or in certain other specific situations (e.g., when left subclavian vein acclusion is present or the patient has undergone a left mastectomy), the pacemaker may be located on the right side. It is important to locate the generator medially enough that it does not interfere with normal shoulder function. The pocket is formed deep to the subcutaneous tissue and above the plane of the pectoral fascia. Occasionally, if the patient is extremely thin or if cosmetic considerations are a priority, the generator may be placed either below the pectoral muscle or via a retromammary approach.

VASCULAR ACCESS

Vascular access is most frequently gained by means of the Seldinger technique. The subclavian vein remains the most common venous access site; however, the axillary vein is becoming an increasingly popular site. Venous access may also be obtained via the cephalic vein or the internal jugular vein. In addition, leads may be tunneled subcutaneously from a remote entry site (e.g., the internal jugular vein) to the site of the generator pocket. Occasionally, thoracotomy and the use of epicardial lead systems are still necessary.

RISKS

Overall, transvenous pacemaker implantation is both safe and well tolerated. The risk of major adverse events (e.g., death, myocardial infarction, stroke, and the need for emergency thoracotomy) is approximately 0.1%. Other complications sometimes encountered include pneuomothorax, vascular injury, cardiac perforation, tamponade, local bleeding, pocket hematoma, infection, and venous thrombosis. There is also a small risk that one or more leads may become dislodged and have to be repositioned in a second procedure.

POSTPROCEDURAL CARE

At most institutions, it is standard practice to admit patients for overnight observation after routine pacemaker implantation. Routine exchange of the pacemaker generator because of battery depletion is often performed as a same-day outpatient procedure. Longer hospitalizations may be required in certain specific situations, as when anticoagulation must be reversed and reinitiated or when a major comorbid condition must be treated.

After implantation of new devices or leads, the ipsilateral arm is placed in a sling or a soft restraint for 12 to 24 hours. Nonnarcotic analgesics are usually sufficient for pain control, but occasionally, oral narcotics are indicated. Patients are monitored via continuous telemetry. We routinely obtain a portable chest x-ray and a 12-lead ECG immediately after implantation.

The day after the procedure, the pacemaker is interrogated and the final settings confirmed. Posteroanterior and lateral chest x-rays are obtained both to verify the positioning of the leads and to rule out the possibility of a slowly accumulating pneumothorax [*see Figure 4*].

Before discharge, the patient receives instruction about the pacemaker teaching and is given a temporary pacemaker card that lists the manufacturer, the specific generator and lead(s) used, and complete serial-number information. Later, the manufacturer mails the patient a permanent identification card, which the patient is asked to keep on hand at all times.

POSTDISCHARGE INSTRUCTIONS AND RESTRICTIONS

Postoperative care focuses on averting hematoma and preventing lead dislodgment. Patients are prohibited from showering for the first 48 to 72 hours. After this period, they may shower, but for the first week, they are advised to cover the implantation site with plastic wrap to protect it from contamination. When 24 hours have passed after implantation, minimal range-of-motion restrictions are placed on the ipsilateral arm and shoulder. Patients are asked to refrain from raising the arm above shoulder level and to perform only limited heavy lifting for the first few weeks. After this period, patients may return to normal activity levels without having to be concerned about displacing the leads or the generator system.

Usually, a follow-up visit is scheduled 7 to 10 days after implantation. During this visit, a wound check is performed to ensure proper healing and to remove the skin suture if it is nonresorbable. As a rule, the pacemaker pocket heals completely within 2 to 4 weeks.

LONG-TERM FOLLOW-UP

Pacemaker patients need routine follow-up care, including interrogation of the pacemaker. Follow-up care can be provided during office visits, via transtelephonic monitoring (TTM), or both. Guidelines for follow-up have been published by NASPE,[32] as well as by the Canadian Working Group in Cardiac Pacing.[33] We recommend that patients either be seen in the office or undergo TTM every 3 months. As the battery approaches the end of its life, more frequent visits may be required.

Complications

Pacemaker complications are infrequent but can lead to serious situations. To minimize adverse consequences, it is important to identify problems early in their course, initiate appropriate workup and treatment, and refer when necessary [*see Table 3*]. Generally, pacemaker complications can be classified according to whether they primarily affect the pocket, the generator, or the leads.

GENERATOR POCKET COMPLICATIONS

Pocket hematomas can occur in any patient but are especially likely to occur in those receiving anticoagulants. These hematomas are usually self-limited, and intervention is rarely necessary. Acute management includes direct manual compression, sandbag compression, pressure dressings, or a combination thereof. Needle aspiration and opening the pocket to drain the hematoma are discouraged because of the risk of introducing infection. Reoperation is generally limited to situations in which there is impending compromise of the incision, uncontrollable bleeding, uncontrollable pain, or suspected infection. Other possible pocket problems include erosion of the underlying hardware, infection, pocket pain, migration of the pulse generator,

Table 3 Common Findings Related to Pacemaker Problems

Findings	*Potential Causes*	*Treatment/Workup*	*When to Refer*
Ecchymoses Hematoma Oozing at incision site	Local bleeding Anticoagulation	External compression; avoid needle aspiration or surgical drainage if possible Withhold anticoagulation	Impending wound dehiscence, uncontrolled pain Signs of infection
Palpable hardware, including header or leads	Benign unless findings consistent with impending erosion	No treatment Cushion with gauze or dressing to avoid irritation from clothing Pocket revision	Signs of impending erosion Pain requiring consideration of pocket revision
Adhesion of skin Thinning or atrophy of skin Scaling of skin Erythema	Impending generator or lead erosion	Pocket revision	Early If hardware becomes exposed, extraction may be required
Exposed hardware	Erosion with infection	Blood cultures Blood count Antibiotics Hardware extraction	Immediately
Pocket erythema Pocket swelling Purulent discharge	Infection Local inflammatory reaction Local trauma	Blood cultures Chest x-ray Blood counts	Early
Pocket pain	Superficial implant Infection Generator migration Pacemaker allergy Superficial irritation from bra strap or clothing	Chest x-ray Examination of generator pocket for signs of migration or infection	Signs of infection Continued pain despite mild analgesics
Fever, chills, or other signs of systemic infection or bacteremia, even without signs of pocket infection	Systemic infection, including bacteremia, bloodstream infection, or endocarditis	Blood cultures Chest x-ray Blood counts Echocardiography	Immediately
Ipsilateral arm swelling Arm heaviness Superior vena cava syndrome	Venous thrombosis	Doppler ultrasonography Arm elevation Anticoagulation	Early
Pectoral muscle twitching Diaphragmatic stimulation	Lead fracture Unipolar pacing Autocapture feature Phrenic nerve stimulation	Chest x-ray Pacemaker interrogation	Early

and misplacement of the generator (so that it interferes with shoulder movement).

Erosion of the underlying hardware can be quite serious, in that it usually leads to infection of the system. In normal circumstances, the underlying hardware, including the leads, can be felt during palpation of the pacemaker pocket, especially if the patient is thin. In extreme cases, the outlines of the generator and the leads can be clearly seen through the skin. It is important to be able to distinguish between normal palpability or visibility and impending pacemaker pocket erosion. Normally, the skin overlying the pacemaker is freely mobile, without discoloration or tenderness to palpation. Fixation, erythema, thinning, atrophy, and scaling of the skin over the underlying hardware are signs of impending erosion. It is crucial to identify early signs of erosion before the hardware breaks the skin. If the skin is intact, surgical revision of the pocket is often all that is needed to protect the hardware from contamination and infection. Once the hardware has been exposed, however, the device must be assumed to be infected, and treatment usually involves a much more complex procedure that includes removal of all the hardware.[34]

Device migration is unusual but can cause significant discomfort. In some cases, surgical revision of the pocket is required to restore an appropriate position.

Chronic pacemaker pocket pain is also infrequent. There is normally some postoperative discomfort while the site heals and the capsule of scar tissue develops. Chronic pain may indicate that the device is not properly located in relation to the shoulder joint and the clavicle or may be an early sign of subacute infection.

GENERATOR COMPLICATIONS

On the whole, pacemaker generators are highly reliable: normal battery depletion aside, failure is unusual. True allergy to pacemaker materials does occur but is rare.

LEAD COMPLICATIONS

Pacemaker lead complications include dislodgment, fracture,

and infection. Fractures can occur throughout the body of the lead, but the most common location is the area where the lead passes between the first rib and the clavicle; fracture at this site leads to the so-called subclavian crush syndrome. Lead fractures may be asymptomatic or may give rise to symptoms related to failure to pace or sense appropriately. Extracardiac stimulation and changes in measured parameters of lead function may be noted. Some lead fractures may be evident on chest x-ray; however, only the conductors are radiopaque, and thus, simple disruption of the outer insulation will not be visible.

A common lead complication is the so-called twiddler's syndrome, which refers to patients who, whether intentionally or subconsciously, continually manipulate the generator within the pocket, eventually causing lead damage or dislodgment.

PACEMAKER INFECTIONS

Bacterial infections can affect any part of the pacemaker system, and the consequences can be devastating. The most com-

Table 4 Sources of Electromagnetic Interference That Can Affect Pacemakers

	Source	*Safe with Pacemaker*	*Specific Recommendations*
Medical sources	MRI	No	Rarely done; restricted to life-threatening situations with close monitoring
	CT scanning	Yes	Pacemaker may interfere with images of thorax
	Lithotripsy	Yes	Activity sensors should be disabled Pacemaker-dependent patients should be programmed to asynchronous mode Shocks should be synchronized to R wave Contraindicated in patients with abdominal implants
	External direct current cardioversion	Yes	Avoid placing patches or paddles directly over pacemaker Have transcutaneous pacing available Use lowest possible energy and biphasic waveform when possible Interrogate pacemaker after procedure
	Neurostimulation	Yes	Test at highest output for pacemaker inhibition before discharge
	Peripheral nerve stimulation	Yes	Nerve conduction studies below the elbow or knee are safe
	Transcutaneous electric nerve stimulation (TENS)	Yes	May require increasing sensing threshold Avoid placing TENS electrodes parallel to pacing vector
	Radiation therapy	Yes	Avoid direct irradiation; maximize shielding If total dose is expected to exceed 10 Gy, device may have to be relocated out of field Reprogram to asynchronous mode if patient is pacemaker dependent Initiate continuous monitoring if patient is pacemaker dependent Check device function after each session and for first few weeks after therapy
	Diagnostic ultrasonography, including echocardiography	Yes	No precautions needed
	Surgical electrocautery	Yes	[*See Figure 5*]
Household and industrial sources	Microwave ovens, TV remote-control devices, cordless telephones, other household appliances	Yes	All devices considered safe; controlled studies lacking
	Slot machines	Yes	May cause interaction and spurious shocks with ICDs
	Walk-through metal detectors	Yes	Do not dwell in scanner; device will probably set off alarm Patients should be advised to carry pacemaker ID card as proof
	Handheld security wand	Yes	Patient should instruct person conducting search not to put wand directly over pacemaker generator
	Cellular telephones	Yes	Keep phone at least 10 cm from pacemaker; do not keep phone in shirt pocket above pacemaker; try to use contralateral ear when using phone
	Electronic article surveillance devices	Yes	Do not dwell in scanner
	Industrial sources, including large electric motors, magnets, and high-voltage power	Yes/No	Depends on source and proximity of pacemaker; site visit may be needed to determine safety
	Arc-welding equipment	No	Cannot be used because of magnetic field of cable

ICD—implantable cardioverter-defibrillator

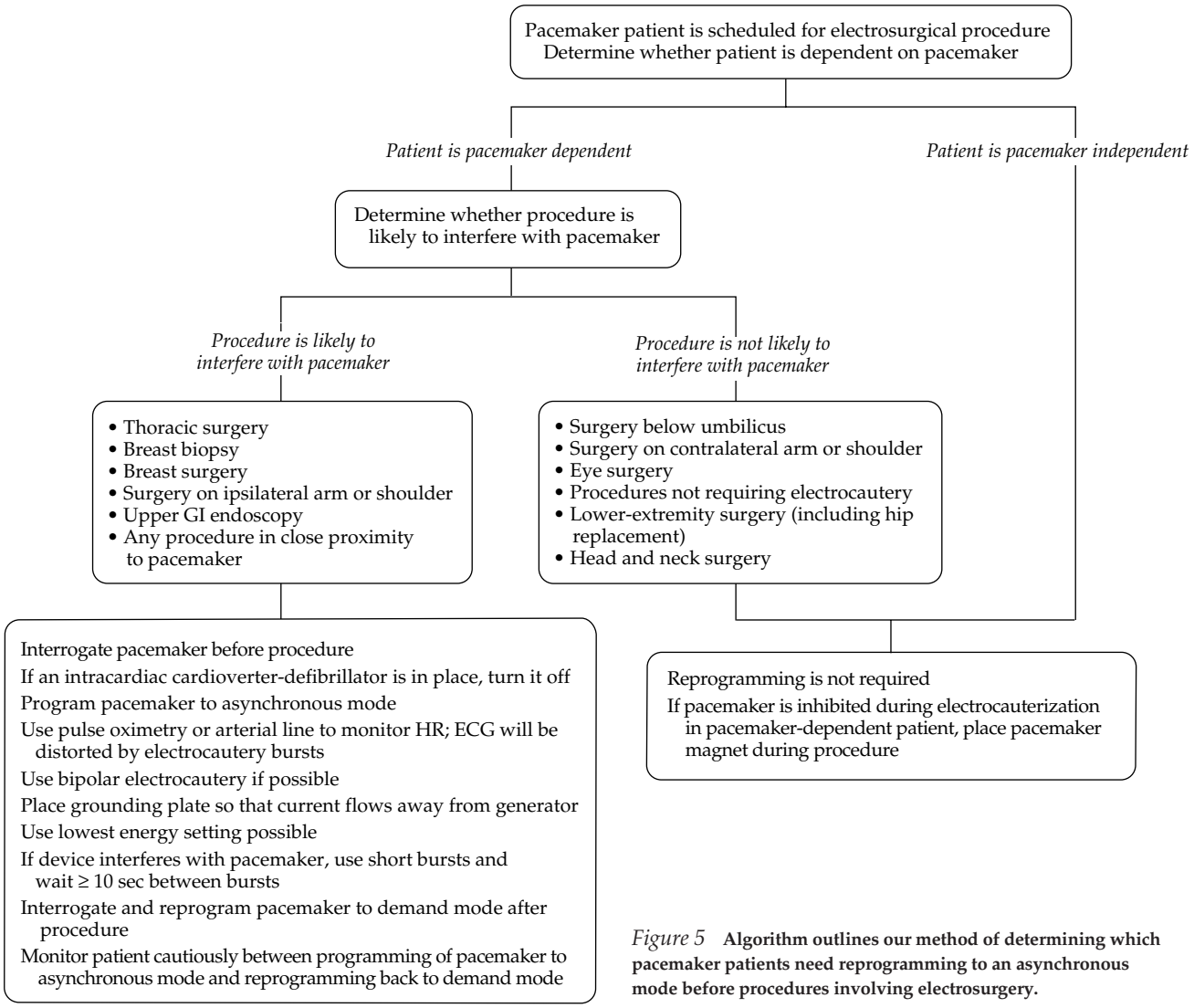


Figure 5 **Algorithm outlines our method of determining which pacemaker patients need reprogramming to an asynchronous mode before procedures involving electrosurgery.**

mon pathogens are staphylococci, especially *Staphylococcus epidermidis*. Once a pacemaker infection is established, it is difficult to eradicate with antibiotics; thus, infected pacemaker systems usually must be removed in their entirety. Patients with pacemakers in place who acquire *S. aureus* bacteremia are at significant risk for a secondary device infection.[35] If infection of an implanted cardiac device is suspected, prompt referral to an experienced center is critical.

External Interference with Pacemaker Function

To function appropriately, pacemakers must be able to sense a clean signal from the myocardium. A number of potential sources can interfere with such signals and thereby affect pacemaker function.[36,37] The most significant of these is EMI, which can have several detrimental effects on pacing systems. The most common detrimental effect of EMI is inhibition of pacing: the pacemaker senses the EMI and interprets it as cardiac activity. In a pacemaker-dependent patient, this misinterpretation can have catastrophic consequences. Other detrimental effects include reversion to an asynchronous pacing mode, reversion to a backup pacing mode, inappropriate activation of other features, and damage to the pacemaker circuitry. Modern pacemakers with bipolar leads are less susceptible to EMI; in addition, they often contain filters and other features designed to protect the patient from device malfunction.

SOURCES OF EMI

Sources of EMI can be divided into household sources, industrial sources, and medical sources [*see Table 4*]. In general, household appliances such as microwave ovens, hairdryers, and television remote controls are safe for pacemaker recipients to use.[38-40] Medical sources of EMI are common in both noninvasive and invasive procedures. MRI scans are generally contraindicated in pacemaker patients; they should be performed only in life-threatening situations and with close monitoring.[41,42] Surgical procedures involving electocauterization are important sources of EMI and often necessitate pacemaker reprogramming before and after the procedure.[43] As a rule, only patients who are pacemaker dependent require reprogramming. The location of the procedure in relation to the pacemaker generator is also an important consideration in deciding whether reprogramming is indicated. On the basis of case reports and our own clinical experience, we have developed an approach we use to determine who needs pacemaker reprogramming before surgery [*see Figure 5*].

The Future

Pacemaker technology is advancing on many fronts.[44,45] Devices are becoming smaller and more sophisticated. Improvements in pacemaker software are allowing closer imitation of normal physiologic cardiac function. New automatic features (e.g., automatic mode switching in response to atrial fibrillation, automatic capture verification, and automatic sensing) are leading to greater reliability and simplified follow-up. New indications for pacing (including cardiac resynchronization therapy for heart failure and treatment of sleep apnea) are evolving. Pacemaker and implantable cardioverter-defibrillator technologies are converging. New information technology is allowing improved collection, storage, and analysis of pacemaker patient data. Internet-based patient management systems are being developed that will include automatic wireless interrogation performed at the patient's home.

Roger A. Freedman, M.D., has received grants for clinical research from Guidant Corporation, Medtronic Inc., and St. Jude Medical, Inc., has served as a consultant to Guidant Corporation and St. Jude Medical, Inc., has received program support from Guidant Corporation and Medtronic Inc., and has served as a speaker for Guidant Corporation.

Jonathan Lowy, M.D., has received grants for clinical research from Guidant Corporation.

References

1. Frye RL, Collins JJ, DeSanctis RW, et al: Guidelines for permanent cardiac pacemaker implantation, May 1984: a report of the Joint American College of Cardiology/American Heart Association Task Force on assessment of cardiovascular procedures (Subcommittee on Pacemaker Implantation). Circulation 70:331A, 1984

2. Dreifus LS, Fisch C, Griffin JC, et al: Guidelines for implantation of cardiac pacemakers and antiarrhythmia devices: a report of the American College of Cardiology/American Heart Association Task Force on assessment of diagnostic and therapeutic cardiovascular procedures. Circulation 84:455, 1991

3. Gregoratos G, Cheitlin MD, Conill A, et al: ACC/AHA Guidelines for implantation of cardiac pacemakers and antiarrhythmia devices: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on Pacemaker Implantation). JACC 31:1175, 1998

4. Gregoratos G, Abrams J, Epstein AE, et al: ACC/AHA/NASPE 2002 guideline update for implantation of cardiac pacemakers and antiarrhythmia devices. Circulation 106:2145, 2002

5. Mymin D, Mathewson FA, Tate RB, et al: The natural history of primary first-degree atrioventricular block. N Engl J Med 315:1183, 1986

6. Shaw DB, Kekwick CA, Veale D, et al: Survival in second degree atrioventricular block. Br Heart J 53:587, 1985

7. Strasberg B, Amat-Y-Leon F, Dhingra RC, et al: Natural history of chronic second-degree atrioventricular nodal block. Circulation 63:1043, 1981

8. Connelly DT, Steinhaus DM: Mobitz type I atrioventricular block: an indication for permanent pacing? Pacing Clin Electrophysiol 19:261, 1996

9. Barold SS: Indications for permanent cardiac pacing in first degree AV block: class I, II, or III? Pacing Clin Electrophysiol 19:747, 1996

10. McAnulty JH, Rahimtoola SH, Murphy E, et al: Natural history of "high risk" bundle-branch block: final report of a prospective study. N Engl J Med 307:137, 1982

11. Ryan TJ, Antman EM, Brooks NH, et al: 1999 update: ACC/AHA guidelines for the management of patients with acute myocardial infarction: executive summary and recommendations: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on Management of Acute Myocardial Infarction). Circulation 100:1016, 1999

12. Nicod P, Gilpin E, Dittrich H, et al: Long-term outcome in patients with inferior myocardial infarction and complete atrioventricular block. J Am Coll Cardiol 12:589, 1988

13. Dubois C, Pierard LA, Smeets JP, et al: Long-term prognostic significance of atrioventricular block in inferior acute myocardial infarction. Eur Heart J 10:816, 1989

14. Goldberg RJ, Zevallos JC, Yarzebski J, et al: Prognosis of acute myocardial infarction complicated by complete heart block (the Worcester Heart Attack Study). Am J Cardiol 69:1135, 1992

15. Spencer FA, Jabbour S, Lessard D, et al: Two-decade-long trends (1975–1997) in the incidence, hospitalization, and long-term death rates associated with complete heart block complicating acute myocardial infarction: a community-wide perspective. Am Heart J 145:500, 2003

16. Kusumoto FM, Goldschlager N: Cardiac pacing. N Engl J Med 334:89 1996

17. Anderson HR, Nielson JC, Thomsen PEB, et al: Long-term follow-up of patients from a randomized trial of atrial versus ventricular pacing for sick sinus syndrome. Lancet 350:1210, 1997

18. De Marneffe M, Gregoire JM, Waterschoot P, et al: The sinus node function: normal and pathological. Eur Heart J 14:649, 1993

19. Santini M, Alexidou G, Ansalone G, et al: Relation of prognosis in sick sinus syndrome to age, conduction defects and modes of permanent cardiac pacing. Am J Cardiol 65:729, 1990

20. Lamas GA, Lee KL, Sweeney MO, et al: Ventricular pacing or dual chamber pacing for sinus node dysfunction. N Engl J Med 346:1854, 2002

21. Grubb BP, Kosinski DJ: Syncope resulting from autonomic insufficiency syndromes associated with orthostatic intolerance. Med Clin North Am 85:457, 2001

22. Connolly SJ, Sheldon R, Roberts RS, et al: The North American vasovagal pacemaker study (VPS): a randomized trial of permanent cardiac pacing for the prevention of vasovagal syncope. J Am Coll Cardiol 33:16, 1999

23. Ammirati F, Colivicchi F, Santini M: Permanent cardiac pacing versus medical treatment for the prevention of recurrent vasovagal syncope: a multicenter, randomized, controlled trial. Circulation 104:52, 2001

24. Connolly SJ, Sheldon R, Thorpe KE, et al: Pacemaker therapy for the prevention of syncope in patients with recurrent severe vasovagal syncope: Second Vasovagal Pacemaker Study (VPS II). JAMA 289:2224, 2003

25. Morillo CA, Camacho ME, Wood MA, et al: Diagnostic utility of mechanical, pharmacological and orthostatic stimulation of the carotid sinus in patients with unexplained syncope. J Am Coll Cardiol 34:1587, 1999

26. Bernstein AD, Daubert JC, Fletcher RD, et al: The Revised NASPE/BPEG generic code for antibradycardia, adaptive-rate, and multisite pacing. Pacing Clin Electrophysiol 25:260, 2002

27. Hayes DL, Naccarelli GV, Furman S, et al: Report of the NASPE Policy Conference training requirements for permanent pacemaker selection, implantation, and follow-up. North American Society of Pacing and Electrophysiology. Pacing Clin Electrophysiol 17:6, 1994

28. al-Khadra AS: Implantation of pacemakers and implantable cardioverter defibrillators in orally anticoagulated patients. Pacing Clin Electrophysiol 26:511, 2003

29. Michaud GF, Pelosi F Jr, Noble MD, et al: A randomized trial comparing heparin initiation 6 h or 24 h after pacemaker or defibrillator implantation. J Am Coll Cardiol 35:1915, 2000

30. Roelke M, Rubinstein VJ, Kamath S, et al: Pacemaker interference with screening mammography. Pacing Clin Electrophysiol 22:1106, 1999

31. Da Costa A, Kirkorian G, Cucherat M, et al: Antibiotic prophylaxis for permanent pacemaker implantation: a meta-analysis. Circulation 97:1796, 1998

32. Bernstein AD, Irwin ME, Parsonnet V, et al: Report of the NASPE policy conference on antibradycardia pacemaker follow-up: effectiveness, needs, and resources. North American Society of Pacing and Electrophysiology. Pacing Clin Electrophysiol 17:1714, 1994

33. Fraser JD, Gillis AM, Irwin ME, et al: Guidelines for pacemaker follow-up in Canada: a consensus statement of the Canadian Working Group on Cardiac Pacing. Can J Cardiol 16:355, 2000

34. Chua JD, Wilkoff BL, Lee I, et al: Diagnosis and management of infections involving implantable electrophysiologic cardiac devices. Ann Intern Med 133:604, 2000

35. Chamis AL, Peterson GE, Cabell CH, et al: *Staphylococcus aureus* bacteremia in patients with permanent pacemakers or implantable cardioverter-defibrillators. Circulation 104:1029, 2001

36. Pinski SL, Trohman RG: Interference in implanted cardiac devices, part I. Pacing Clin Electrophysiol 25:1367, 2002

37. Pinski SL, Trohman RG: Interference in implanted cardiac devices, part II. Pacing Clin Electrophysiol 25:1496, 2002

38. Niehaus M, Tebbenjohanns J: Electromagnetic interference in patients with implanted pacemakers or cardioverter-defibrillators. Heart 86:246, 2001

39. Hayes DL, Wang PJ, Reynolds DW, et al: Interference with cardiac pacemakers by cellular telephones. N Engl J Med 336:1473, 1997

40. McIvor ME, Reddingger J, Floden E, et al: Study of pacemaker and implantable cardioverter defibrillator triggering by electronic article surveillance devices (SPICED TEAS). Pacing Clin Electrophysiol 21:1847, 1998

41. Hayes DL, Holmes DR Jr, Gray JE: Effect of 1.5 Tesla nuclear magnetic resonance imaging scanner on implanted permanent pacemakers. J Am Coll Cardiol 10:782, 1987

42. Lauck G, von Smekal A, Wolke S, et al: Effects of nuclear magnetic resonance imaging on cardiac pacemakers. Pacing Clin Electrophysiol 18:1549, 1995

43. Madigan JD, Choudhri AF, Chen J, et al: Surgical management of the patient with an implanted cardiac device. Ann Surg 230:639, 1999

44. Bryce M, Spielman SR, Greenspan AM, et al: Evolving indications for permanent pacemakers. Ann Intern Med 134:1130, 2001

45. Gold MR: Permanent pacing: new indications. Heart 86:355, 2001

VIII ACUTE MYOCARDIAL INFARCTION

Peter B. Berger, m.d.
James L. Orford, m.b.chb., m.p.h.

In the 1970s, coronary angiography demonstrated that almost all cases of acute myocardial infarction were caused by thrombotic occlusion of a coronary artery. This discovery has led to the development of therapies to restore coronary blood flow in the occluded artery, which has dramatically reduced the morbidity and mortality associated with acute myocardial infarction.

Epidemiology

In the past decade, the number of people who die each year of myocardial infarction has decreased significantly. Both in-hospital mortality and out-of-hospital mortality have declined as a result of substantial increases in the use of aspirin, heparin, thrombolytic therapy, and coronary angioplasty, as well as a reduction in the risk factors for coronary artery disease (e.g., hypertension, hyperlipidemia, smoking, and sedentary lifestyle) [*see* Risk-Factor Modification, *below*]; however, it must be emphasized that there is, unfortunately, persistent discordance between existing guidelines for management of acute coronary artery disease syndromes and current clinical practice.[1,2]

Despite these advances, approximately 1.5 million people in the United States suffer acute myocardial infarction each year, and nearly 500,000 of these patients die of coronary artery disease.[3] Nearly half of these deaths occur before the patients receive medical care either from emergency medical technicians or in a hospital.[3,4]

Pathogenesis

The factors responsible for the sudden thrombotic occlusion of a coronary artery have only recently been elucidated.[5-7] Atherosclerotic plaques rich in foam cells (lipid-laden macrophages) are susceptible to sudden plaque rupture and hemorrhage into the vessel wall, which may result in the sudden partial or total occlusion of the coronary artery. Although severe stenosis of a coronary artery (i.e., stenosis ≥ 70% of the diameter of the artery) is generally required to produce anginal symptoms, such stenoses tend to have dense fibrotic caps and are less prone to rupture than mild to moderate stenoses, which are generally more lipid laden. Studies of patients in whom angiography was performed before and after a myocardial infarction revealed that in most cases, acute coronary artery occlusion occurred at sites in the coronary artery circulation with stenoses of less than 70%, as demonstrated on the preinfarction angiogram.[8] Although patients who have unstable anginal syndromes with increasingly frequent and severe angina are clearly at increased risk for myocardial infarction, the ability of physicians to predict which patients with stable anginal syndromes are likely to experience infarction and which coronary artery stenoses are likely to result in acute thrombotic occlusion is poor.

Diagnosis

According to the World Health Organization, the diagnosis of myocardial infarction requires at least two of the following three criteria: (1) a clinical history of ischemic-type chest discomfort, (2) serial electrocardiographic tracings indicative of myocardial infarction, and (3) a rise and fall in serum cardiac markers.[9] However, the advent and widespread adoption of novel diagnostic tools, including highly sensitive and specific serologic biomarkers and precise imaging techniques, have necessitated reevaluation of this established definition. The Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction has integrated these diagnostic modalities and published updated definitions of acute myocardial infarction, evolving or recent myocardial infarction, and established myocardial infarction that more accurately reflect current clinical practice [*see Table 1*].[10]

Clinical Manifestations

Patients with acute myocardial infarction often describe a heaviness, pressure, squeezing, or tightness in the chest that has persisted for more than 30 minutes. The discomfort may radiate or be located primarily in the arms, neck, or jaw. Chest pain, particularly severe or stabbing chest pain, and pain that causes writhing are unusual for coronary artery ischemia and should lead the clinician to consider causes other than myocardial infarction. Many patients with acute myocardial infarction, particularly those with inferior infarction, are diaphoretic; nausea and emesis are common as well. Dyspnea is also a common associated symptom. Syncope may occur and is more frequent with inferior than anterior infarction, in part because of the more frequent occurrence of bradyarrhythmias, heart block, and tachyarrhythmias with inferior infarction. Elderly patients with

Table 1 Clinical Definitions of Myocardial Infarction as Determined by the Joint European Society of Cardiology/American College of Cardiology Committee[10]

Acute, Evolving, or Recent Myocardial Infarction
Biochemical markers of myocardial necrosis (i.e., typical rise and gradual fall of troponin or more rapid rise and fall of CK-MB) with at least one of the following:
Ischemic symptoms
Development of pathologic Q waves on the ECG
ECG changes indicative of ischemia (ST segment elevation or depression)
Coronary artery intervention (e.g., primary coronary angioplasty)
Pathologic findings of an acute myocardial infarction
Established Myocardial Infarction
Development of new pathologic Q waves on serial ECGs; the patient may or may not remember previous symptoms; biochemical markers of myocardial necrosis may have normalized, depending on the length of time that has passed since the infarct developed
Pathologic findings of a healed or healing myocardial infarction

CK-MB—creatine kinase–myocardial band

infarction often present with symptoms that differ from the symptoms of infarction in younger patients; more than half of elderly patients present with shortness of breath as their main complaint, and many others present with dizziness or symptoms of arrhythmia rather than the classic symptoms of acute myocardial infarction.[11]

Approximately two thirds of patients describe the new onset of angina or a change in their anginal pattern in the month preceding infarction.[12] However, in approximately one fourth of patients, myocardial infarction is associated with only mild symptoms or no symptoms at all.[13]

PHYSICAL EXAMINATION

The patient with acute myocardial infarction often appears anxious and in distress. Vital signs are often normal, but sinus tachycardia is not uncommon. The pulse may be rapid or slow if arrhythmias are present. Either hypotension caused by left or right ventricular dysfunction or arrhythmia or hypertension caused by adrenergic discharge may be present. The respiratory rate may be elevated because of anxiety or pain or because of hypoxia in patients with significant congestive heart failure. The jugular venous pressure may be elevated, reflecting right ventricular dysfunction caused by right ventricular involvement (more common with inferior infarction); arrhythmia in which atrioventricular dissociation is present may produce so-called cannon A waves, which are abnormally high jugular venous waves caused by atrial systole occurring when the atrioventricular valves are closed. The lung examination is typically normal, but moist rales indicative of congestive heart failure resulting from left ventricular dysfunction may be present. The cardiac examination may reveal a dyskinetic apical pulsation on palpation; a fourth and, less commonly, a third heart sound may be audible. The murmur of ischemic mitral regurgitation may be present. If a left bundle branch block is present, abnormal splitting of the second heart sound may be heard.

It must be emphasized that the physical examination in acute myocardial infarction is generally most useful in excluding other potentially serious causes of the patient's chest discomfort, including pulmonary embolism, aortic dissection, spontaneous pneumothorax, pericarditis, and cholecystitis, rather than in confirming a diagnosis of acute myocardial infarction.

ELECTROCARDIOGRAPHY

ECG is a valuable tool both in confirming the diagnosis of acute myocardial infarction and in selecting the most appropriate therapy for the patient. Although rhythm and conduction disturbances may be present, the presence and type of repolarization abnormalities are most useful in identifying myocardial infarction. If ST segment elevation is present in a patient with chest pain typical of acute myocardial infarction, the likelihood that the patient has acute myocardial infarction is greater than 90%.[14] Other findings, such as ST segment depression, T wave inversion, and bundle branch block, are less specific but may also support a diagnosis of acute myocardial infarction, particularly when typical symptoms are present [*see Figure 1*]. Fully 50% of patients with myocardial infarction do not have ST segment elevation on their ECGs, although the ECG is seldom normal even at an early stage.[15] In such patients, the ECG can help predict complications and early mortality.[16] Patients with ST segment depression are at high risk; 30-day mortality in such patients is nearly as high as in patients with anterior ST segment elevation.[17] Patients with other nonspecific ECG abnormalities are at lesser

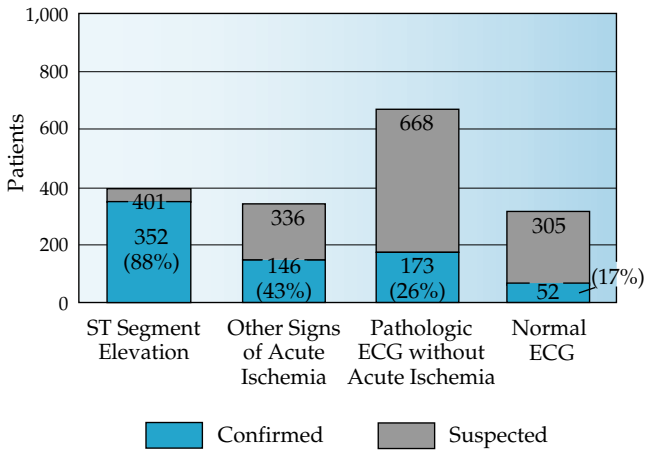


Figure 1 **Relation between the initial electrocardiographic changes and the development of infarction in 1,715 patients strongly suspected of having an acute myocardial infarction. Each column shows the total number of patients and the number of patients later found to have had an infarction. Although infarction is less frequently confirmed in patients without ST segment elevation than in those with ST segment elevation, even patients with normal ECG findings may suffer acute myocardial infarction.[14]**

risk; those with normal ECGs who suffer infarction generally have the best prognosis [*see Figure 2*]. Regardless of the findings on the initial ECG, the most important element in the evaluation of a patient with suspected acute myocardial infarction is the patient's description of symptoms. All patients suspected of having acute myocardial infarction should be admitted to the hospital and receive rapid and appropriate therapy.

LABORATORY FINDINGS

Injury to myocardial cells results in the release of intracellular enzymes into circulating blood, permitting their detection by blood tests. Traditionally, the serum cardiac marker creatine kinase (CK) and an isoenzyme, creatine kinase–myocardial band (CK-MB), which are found in high concentration in myocardial cells, have been used to diagnose myocardial infarction in its earliest stages.[18] Rapid assays of these enzymes have been developed, permitting the determination of the blood levels of these enzymes within 30 to 60 minutes. Drawbacks to the use of CK-MB include its lack of specificity for cardiac muscle and the time required for CK-MB levels to rise during myocardial infarction. CK and CK-MB usually require at least 3 hours of profound ischemia to rise above normal levels; patients who present early in their infarction would not be expected to have elevated CK levels. Furthermore, patients may have only partial obstruction of the infarct-related artery, or there may be extensive collateralization of the infarct-related artery, which further delays the release of these enzymes. In patients suspected of having acute myocardial infarction, it is not appropriate to delay treatment until an elevation of CK or CK-MB is present, because the goal of treatment is to prevent injury to the myocardium. The challenge facing physicians is to identify patients suffering myocardial infarction even before CK becomes elevated, because these patients require emergency therapy and stand to benefit the most from reperfusion therapy.

To overcome these limitations and more accurately and rapidly identify patients in need of emergency reperfusion therapy,

other blood tests have been developed to help identify patients with ischemia. Myoglobin is a low-molecular-weight heme protein found in cardiac muscle. Its advantage for diagnosis is that it is released more rapidly from infarcted myocardium than is CK-MB. However, myoglobin is also found in skeletal muscle, and the lack of specificity is a drawback.[19] Troponin is a cardiac-specific marker for acute myocardial infarction; an increase in serum levels of troponin occurs soon after myocardial cell injury. An elevated cardiac troponin level on admission is a predictor of subsequent cardiac events.[20,21] The recent changes to the definition of acute myocardial infarction reflect the increased emphasis on these specific biomarkers of myocardial injury [*see Table 1*].[10]

IMAGING STUDIES

Echocardiography

Echocardiography may be useful in identifying patients with myocardial infarction in the emergency department. Regional wall motion abnormalities occur within seconds of coronary occlusion and well before myocyte necrosis,[22-26] and most patients with acute myocardial infarction have regional wall motion abnormalities readily seen on echocardiography. However, echocardiographic evidence of myocardial infarction is not required in patients with symptoms and electrocardiographic evidence typical of acute myocardial infarction, and treatment should not be delayed, so that an echocardiogram can be performed. Similarly, wall motion abnormalities are not specific for acute myocardial infarction and may be caused by ischemia or prior infarction. Echocardiography may be useful in patients with left bundle branch block or abnormal ECGs without ST segment elevation whose symptoms are atypical and in whom the diagnosis is uncertain.[27]

Radionuclide Imaging

Perfusion imaging with both thallium and sestamibi in the emergency department has been reported to be both sensitive and specific in the evaluation of patients in whom the diagnosis is uncertain.[27-29] A prospective randomized trial of 2,475 patients found that resting technetium-99m sestamibi imaging reduced unnecessary hospitalization in patients with acute ischemia without reducing admission of patients with acute ischemia.[30]

Emergent Therapy

Treatments have been developed that reduce the morbidity and mortality of acute myocardial infarction, particularly when initiated early; it is therefore important to avoid delay in administering therapy.[4,31] Although the greatest delay in treatment of acute myocardial infarction is usually the time that it takes a patient to seek medical care, much of the emphasis on reducing delay has focused on the time between a patient's presentation to the emergency department and the administration of reperfusion therapy. A patient with symptoms suggestive of myocardial infarction should be evaluated within 10 minutes after arrival in the emergency department. Early steps should include the assessment of hemodynamic stability by measurement of the patient's heart rate and blood pressure; the performance of a 12-lead ECG; and the administration of oxygen by nasal prongs, I.V. analgesia (most commonly morphine sulfate), oral aspirin, and sublingual nitroglycerin if the blood pressure is greater than 90 mm Hg. The challenge facing physicians who work in emergency departments is that more than 90% of patients who present to the emergency department complaining of chest pain are not suffering myocardial infarction; many do not have a cardiac etiology for their chest pain.

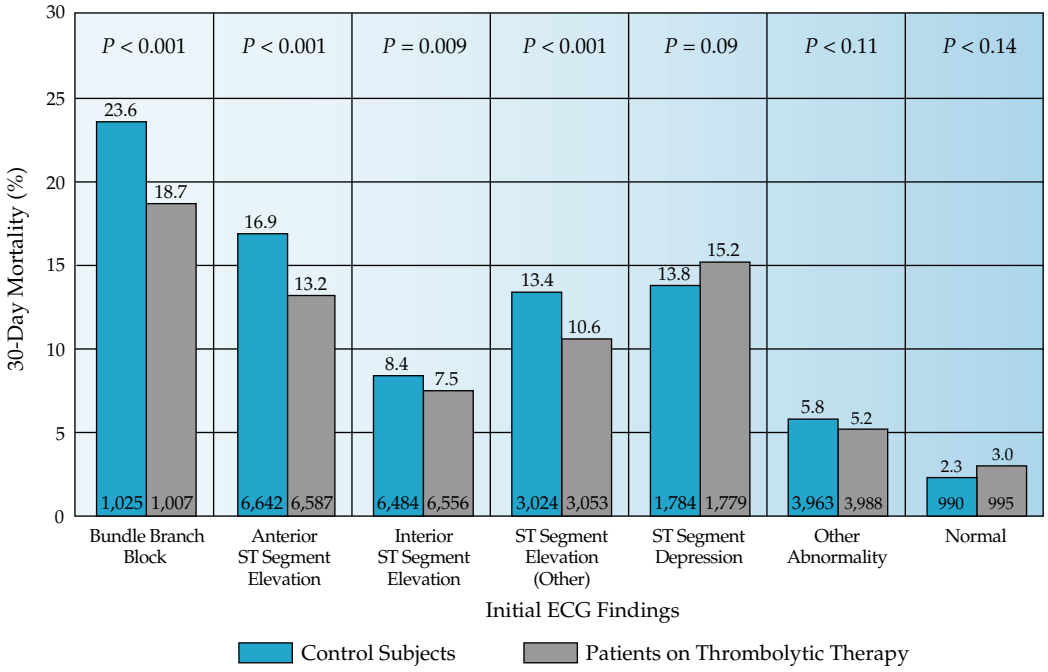


Figure 2 **Thirty-day mortality in patients with suspected acute myocardial infarction from placebo-controlled trials of thrombolytic therapy on the basis of their initial ECGs. Patients with ST segment depression are at high risk, nearly as high as patients with anterior ST segment elevation. The mortality among such patients is not reduced (and may be increased) by thrombolytic therapy. Patients with other nonspecific electrocardiographic abnormalities are at lesser risk, and those with normal ECG findings have the best prognosis.**

All patients with definite or suspected myocardial infarction should be admitted to the hospital, undergo preparation for I.V. access, and be placed on continuous ECG monitoring. High-risk patients should be admitted to a coronary care unit. In many hospitals, patients at low risk for major complications are admitted to a telemetry unit, where emergency medical care can be quickly administered, rather than to a coronary care unit. Tachyarrhythmias and bradyarrhythmias may occur even in low-risk patients, particularly in the first 24 hours. Lidocaine, atropine, an external or internal pacemaker, and a defibrillator should be readily available.

OXYGEN

Oxygen is generally recommended for all patients with acute myocardial infarction for the first several hours after admission and is mandatory for patients with pulmonary congestion or evidence of oxygen desaturation.

ASPIRIN

Aspirin should be given to all patients as soon as a diagnosis of myocardial infarction is made.[17] In the second International Study of Infarct Survival (ISIS-2), aspirin was found to be nearly as effective as streptokinase, reducing 30-day mortality 23% in 17,000 patients with acute myocardial infarction; the benefit was additive in patients receiving both aspirin and streptokinase [*see Figure 3*].[17] Other studies have revealed similar benefit from immediate aspirin therapy.[32] The beneficial effect of aspirin is the result of its antiplatelet effect, which is achieved through the rapid inhibition of thromboxane A_2 production.

Patients should be maintained on aspirin indefinitely. Prolonged administration of aspirin in patients with a history of myocardial infarction is associated with a 25% reduction in death, nonfatal reinfarction, and stroke.[32]

ANALGESIA

Pain relief should be among the initial therapies offered to patients with acute myocardial infarction. Persistent chest discomfort is generally caused by ongoing myocardial ischemia; although the ultimate goal of therapy is to eliminate ischemia, analgesia should be administered without delay. In addition to making patients more comfortable, pain relief may reduce the outpouring of catecholamines characteristic of the early stages of acute myocardial infarction and thereby reduce myocardial oxygen demand. Intravenous morphine sulfate is commonly used for pain relief in this setting.

Reperfusion Therapy

REPERFUSION STRATEGIES AND OUTCOMES

Importance of Time to Reperfusion

Many important predictors of early clinical outcome in myocardial infarction are independent of treatment. Most of the early mortality is explained by factors such as the age of the patient, initial heart rate and blood pressure, initial Killip classification [*see Table 2*], and infarct location. However, the time to administration of reperfusion therapy is a critical determinant of outcome and one of the few determinants of early clinical outcome under the control of the physician. Many studies have revealed that patients with myocardial infarction treated most rapidly have a lower mortality and, among survivors, reduced infarct size [*see Figure 4*].[33] This observation has led to recommendations that the time between a patient's presentation to the emergency department and the administration of thrombolytic therapy not exceed 60 minutes; ideally, this period should not exceed 30 minutes.[34] The most critical interval is the time between symptom onset and the achievement of reperfusion, not the time to the initiation of therapy. Thus, therapy that takes longer to initiate (e.g., primary coronary angioplasty) may actually be superior if it

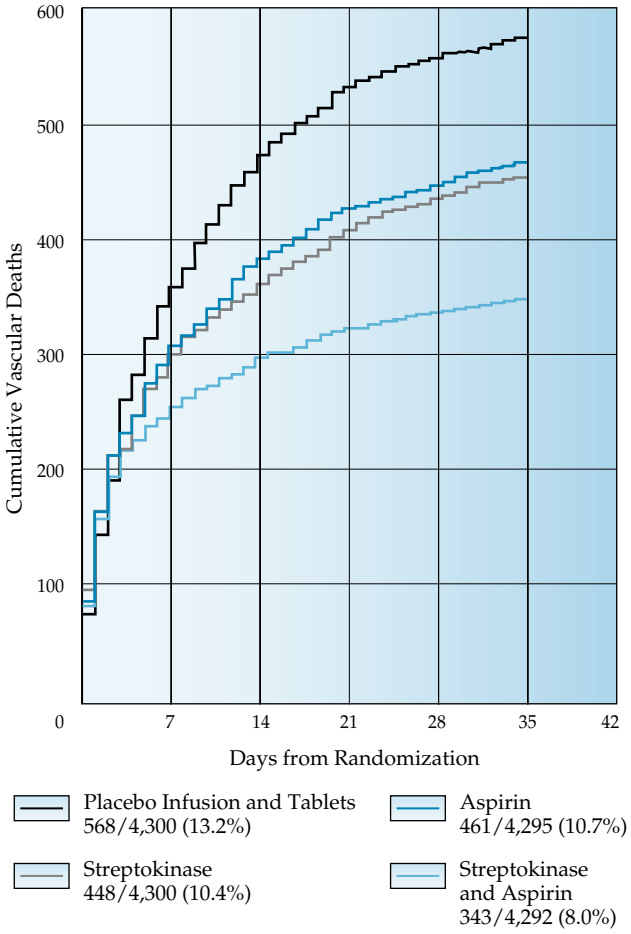


Figure 3 **Mortality at 35 days in 17,187 cases of suspected acute myocardial infarction in the second International Study of Infarct Survival (ISIS-2). In this study, aspirin reduced 30-day mortality by 23% and was nearly as effective as streptokinase; the benefit was additive in patients receiving both aspirin and streptokinase.[17]**

Table 2 Killip Classification of Acute Myocardial Infarction

Class I	No clinical heart failure
Class II	Findings consistent with mild or moderate heart failure (e.g., isolated S_3 gallop, bilateral rales in up to 50% of lung fields)
Class III	Pulmonary edema, rales in all lung fields, acute mitral regurgitation
Class IV	Cardiogenic shock (e.g., stuporous state of consciousness, systolic blood pressure < 90 mm Hg, decreased urine output, pulmonary edema, and cold, clammy skin)

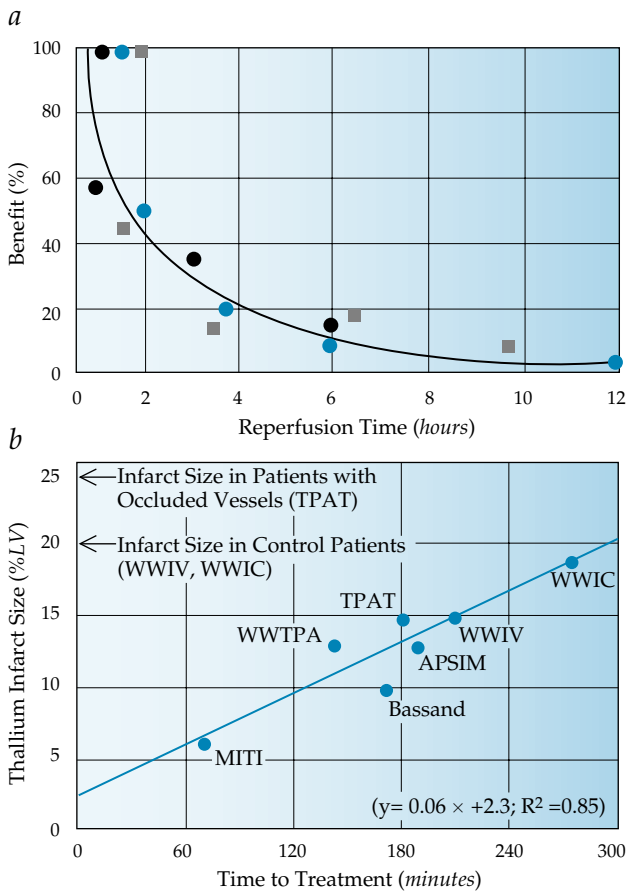


Figure 4 **Many studies have revealed lower mortality (*a*) and reduced infarct size among survivors (*b*) of myocardial infarction treated most rapidly. The equation shows the linear relation between infarct size and time to treatment.[33] (APSIM—APSAC dans l'Infarctus du Myocarde; Bassand—Bassand study; MITI—Myocardial Infarction Triage and Intervention; TPAT—Tissue Plasminogen Activator, Toronto trial; WWIC—Western Washington Intracoronary streptokinase trial; WWIV—Western Washington Intravenous streptokinase trial; WWTPA—Western Washington Tissue Plasminogen Activator trial)**

achieves reperfusion more rapidly than another therapy that can be initiated more rapidly (e.g., thrombolytic therapy).

The importance of avoiding hospital delay in performing primary coronary angioplasty was evident in the Global Use of Strategies to Open Occluded Arteries (GUSTO-IIb) substudy, which compared primary coronary angioplasty with tissue plasminogen activator (t-PA) therapy.[35] There was a clear relation between the length of time until angioplasty was performed after enrollment in the study and 30-day mortality [*see Figure 5*]. Analysis of 27,080 patients in the second National Registry of Myocardial Infarction also revealed a relation between time to treatment with primary percutaneous transluminal coronary angioplasty (PTCA) and survival, even after adjusting for other mortality risk factors.[36] In that study, the volume of patients treated with angioplasty at the hospital was also a predictor of outcome; a lower mortality was seen at hospitals in which a high number of patients with acute myocardial infarction were treated with coronary angioplasty. There have been studies in which unacceptably high mortality was seen at hospitals when primary angioplasty was not performed rapidly; reducing delay led to a reduction in mortality.[37] Therefore, as is the case with thrombolytic therapy, the speed with which reperfusion is achieved appears to be an important determinant of clinical outcome. The best reperfusion therapy (coronary angioplasty or thrombolytic therapy) is not necessarily the one that can be most rapidly initiated but, rather, the one that achieves coronary patency most rapidly. Clinicians should know, at their own institution, whether coronary angioplasty is more rapid or less rapid than thrombolytic therapy at restoring flow to the infarct-related artery; in general, the therapy that restores flow most rapidly should be preferred. At institutions where a skilled catheterization team is on call 24 hours a day and can be rapidly assembled, coronary angioplasty would most likely be able to restore coronary blood flow in more patients more rapidly than thrombolytic therapy. Elsewhere, thrombolytic therapy may be preferable.[38-40]

Transfer for Primary Angioplasty versus Immediate Thrombolytic Therapy

Time to reperfusion is an important modifiable predictor of clinical outcome for both thrombolysis and primary angioplasty, although it has the greatest impact on patients treated with thrombolytic therapy. An alternative treatment strategy for patients with ST segment elevation myocardial infarction (STEMI) initially assessed at a hospital without on-site cardiac surgery facilities is immediate transfer for primary percutaneous coronary intervention (PCI). Although immediate transfer is intuitively an attractive option, reasonable concerns regarding the unavoidable delay before transfer and the risk of transporting such critically ill patients during the acute phase of myocardial infarction have mandated formal evaluation of this strategy.

The PRAGUE-2 investigators randomized 850 patients with acute STEMI presenting within 12 hours to a hospital without a catheterization laboratory to either immediate thrombolysis or transfer for primary PCI.[41] The investigators determined that in the acute phase of STEMI, long-distance transport from a community hospital to a facility with PCI is safe and is associated with decreased mortality in patients presenting more than 3 hours after symptom onset. Similarly, the Danish Trial in Acute Myocardial Infarction (DANAMI)–2 trial investigators concluded that immediate transfer for primary PCI, in preference to immediate thrombolysis, was safe and efficacious.[42]

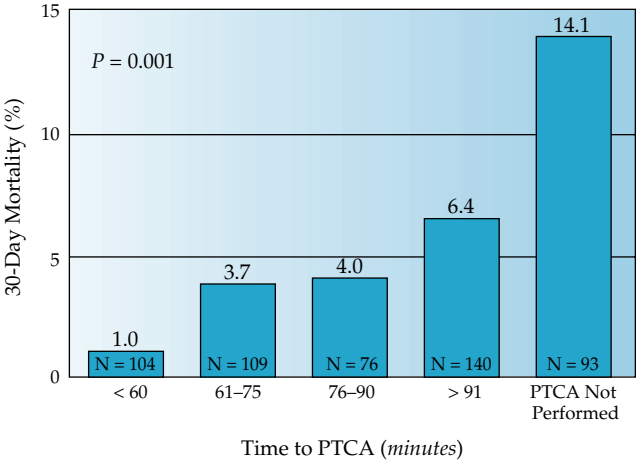


Figure 5 **Relation between the time from study enrollment to the first balloon inflation and 30-day mortality in the GUSTO-IIb substudy. Patients assigned to angioplasty in whom angioplasty was not performed are also shown.[35] (PTCA—percutaneous transluminal coronary angioplasty)**

However, data from these trials need to be understood within the context of the individual trial designs and their actual conduct. The maximum transport distance in the PRAGUE-2 trial was 120 kilometers, and the time from transport to balloon inflation was only 97 ± 27 minutes. In the DANAMI-2 trial, the time from arrival to initiation of treatment in the thrombolysis arm of the study was 51 minutes, whereas the time required to transfer patients for primary PCI was 155 minutes ($P < 0.0001$). In contrast, a report from the National Registry of Myocardial Infarction (NRMI)–4 (http://www.nrmi.org/nrmi_data.html) revealed a median door-to-balloon time of 185 minutes for patients transferred for primary PCI in the United States and a door-to-balloon time of less than 90 minutes in only 3% of patients. It is reasonable to assume that the reported benefit associated with this particular treatment strategy in the aforementioned randomized, controlled clinical trials (i.e., PRAGUE-2 and DANAMI-2) can be realized only if similar transfer times and door-to-balloon times are reproduced in clinical practice; the findings of the NRMI 4 investigators are not reassuring in this regard.

Coronary Angiography after Uncomplicated Myocardial Infarction

The role of coronary angiography after uncomplicated myocardial infarction remains controversial for patients who have received thrombolytic therapy. Coronary angiography in patients initially treated with thrombolytic agents has been studied in the second Thrombolysis in Myocardial Infarction (TIMI II) study, the Should We Intervene Following Thrombolysis? (SWIFT) study, the Treatment of Post-thrombolytic Stenoses (TOPS) study, and, most recently, a German study.[43-46] It is clear from these studies that patients treated with thrombolytic therapy in whom complications do not occur are at low risk for reinfarction and death after discharge and that the routine performance of coronary angiography and coronary angioplasty does not reduce the occurrence of these adverse events. Despite the publication of these well-designed studies, there has been considerable reluctance among physicians to accept their results, and there remains considerable variability throughout the United States and the world in the frequency with which coronary angiography is performed in such patients.

Many cardiologists feel more comfortable caring for patients who have suffered a myocardial infarction if the patient's coronary anatomy is known. Patients at low risk may be discharged from the hospital more rapidly. Patients who have left mainstem or multivessel disease, particularly those who have reduced ventricular function, may be referred for coronary artery bypass surgery or percutaneous revascularization. Patients with persistent occlusion of the infarct-related artery may benefit from revascularization because of favorable effects on remodeling, a reduction in ventricular arrhythmia, and the improved ability of the infarct-related artery to provide collateral blood flow to other coronary arteries in the future. Nonetheless, until the benefits of cardiac catheterization are demonstrated in asymptomatic patients after an uncomplicated myocardial infarction, a conservative strategy is recommended in patients who have been given thrombolytic therapy, and coronary angiography is recommended only for patients with hemodynamic instability or for patients in whom spontaneous or exercise-induced ischemia occurs; such a strategy is safe and is associated with a good clinical outcome.

Patients who are not given thrombolytic therapy are at higher risk for reinfarction and death than those receiving thrombolytic therapy. The role of coronary angiography in patients with acute myocardial infarction not receiving thrombolytic therapy has not been studied. In such patients whose infarctions are complicated by hemodynamic compromise or postinfarction chest pain or in patients in whom multivessel disease or reduced ventricular function is believed to be present, coronary angiography is probably helpful. It remains unclear whether coronary angiography should be performed in patients not treated with thrombolytic therapy who do not have these high-risk characteristics. It is impossible to be definitive about recommendations in the absence of appropriate studies, and not surprisingly, practice patterns vary widely throughout the United States and the world in such patients.

Reperfusion Therapy in Patients without ST Segment Elevation

Primary PTCA has not been appropriately studied in patients without ST segment elevation, and it is not possible to be definitive about its use in this setting. However, regardless of the findings on ECG, PTCA is widely believed to be beneficial in patients with ischemic-type chest discomfort that persists despite medical therapy. Many patients with prolonged chest pain without ST segment elevation are not suffering from myocardial infarction; the likelihood that infarction is present is increased if repolarization abnormalities are present on the ECG and the patient has risk factors for coronary artery disease. In patients with critical coronary artery stenoses, immediate PTCA or bypass surgery may be appropriate. In patients without significant coronary artery disease, immediate angiography can also be extremely useful and can lead to the withdrawal of cardiac med-

Table 3 Class I Recommendations for the Use of an Invasive Strategy in the Management of Patients with Unstable Angina or Non–ST Segment Elevation Myocardial Infarction

An early invasive strategy is recommended for patients who have unstable angina or non–ST segment elevation myocardial infarction (NSTEMI) without serious comorbidity and who have any of the following high-risk indicators (level of evidence: A):

- Recurrent angina/ischemia at rest or with low-level activities despite intensive anti-ischemic therapy
- Elevated levels of troponin T or troponin I
- New or presumably new ST segment depression
- Recurrent angina/ischemia with symptoms of congestive heart failure, an S_3 gallop, pulmonary edema, worsening rales, or new or worsening mitral regurgitation
- High-risk findings on noninvasive stress testing
- Depressed left ventricular systolic function (e.g., ejection fraction < 0.40 on noninvasive study)
- Hemodynamic instability
- Sustained ventricular tachycardia
- Percutaneous coronary intervention within 6 mo
- Prior coronary artery bypass grafting

In the absence of any of these findings, an early conservative strategy or an early invasive strategy may be offered in hospitalized patients without contraindications for revascularization (level of evidence: B)

Note: Class I recommendations pertain to conditions for which there is evidence or general agreement that a given procedure is useful and effective. Level A evidence (highest)—Derived from multiple randomized clinical trials. Level B evidence (intermediate)—Derived from limited number of randomized clinical trials, nonrandomized studies, or observational registries.

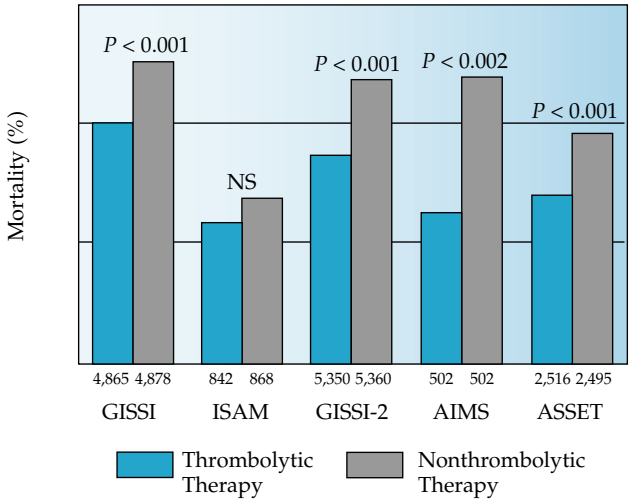


Figure 6 **Data from five controlled megatrials of thrombolytic therapy large enough to detect a mortality difference between the thrombolytic and nonthrombolytic control arms of the trials. Pooled data from these five trials (not shown) reveal a 29% mortality reduction in patients treated within 6 hours of symptom onset.[53] (AIMS—APSAC International Mortality Study; ASSET—Anglo-Scandinavian Study of Early Thrombosis; GISSI—Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico; GISSI-2—Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Myocardico; ISAM—Intravenous Streptokinase in Acute Myocardial Infarction; NS—not significant)**

ications, discharge from the coronary care unit, and appropriate diagnostic evaluation, in many cases as an outpatient. Immediate angiography is recommended in all patients with hypotension, severe congestive heart failure, or cardiogenic shock regardless of the initial ECG results, because immediate revascularization appears to reduce mortality in this setting.[47] In addition, there is a compelling case to be made for routine invasive evaluation of all patients who are admitted with unstable angina or non–ST segment elevation myocardial infarction (NSTEMI).[48]

In the TIMI-IIIb study, an early intervention strategy was compared with a conservative strategy in 3,000 patients with either unstable angina, recent non–Q wave myocardial infarction, or prolonged chest pain without ST segment elevation on ECG.[49] Patients were randomized to receive either early angiography or medical therapy; only those patients who subsequently experienced recurrent chest pain or had an exercise test underwent angiography. Although death and myocardial infarction occurred with similar frequency in the two groups, the study showed that the initial hospitalization was longer and the need for rehospitalization more frequent in the group receiving conservative therapy. More recently, in the second Fragmin and Fast Revascularization during Instability in Coronary Artery Disease (FRISC-II) trial, an early invasive strategy was shown to reduce both mortality and myocardial infarction at 1 year.[50] In the Treat Angina with Aggrastat [tirofiban] and Determine Cost of Therapy with Invasive or Conservative Strategy—Thrombolysis in Myocardial Infarction–18 (TACTICS-TIMI-18) study, an early invasive strategy was found to reduce the combined end point of death or myocardial infarction.[51] In both the FRISC-II and TACTICS-TIMI-18 studies, patients at greatest risk, such as those with positive troponin values and with ST segment depression at study entry, had the highest event rates and derived the greatest benefit from an invasive strategy.

The 2002 ACC/AHA unstable angina guideline update has summarized the data regarding early invasive versus early conservative strategies in patients with unstable angina or NSTEMI and issued an updated set of recommendations to guide clinical decision making in this setting [*see Table 3*].[52]

THROMBOLYTIC THERAPY

Thrombolytic therapy has been widely studied in prospective, randomized, controlled trials involving more than 50,000 patients and has been proved to reduce mortality 29% in patients with ST segment elevation treated within 6 hours after the onset of chest pain [*see Figure 6*].[53] The survival benefit of thrombolytic therapy is maintained for years.[54] The benefit of thrombolytic therapy is achieved through rapid restoration of blood flow in an occluded coronary artery.[55-57]

Thrombolytic therapy is strongly recommended for patients with ST segment elevation in two or more contiguous leads who have had less than 6 hours of chest pain; for patients with classic symptoms of infarction in whom a bundle branch block precludes detection of ST segment elevation[53]; and for patients presenting with 6 to 12 hours of chest pain, although the expected benefits for this last group of patients are fewer. The potential benefits should be weighed against the potential risks in patients with relative contraindications to thrombolytic therapy (see below).[17,58] It is important to calculate the duration of infarction as the time from the last pain-free interval. The infarct-related artery often opens and closes spontaneously during the early stages of infarction, which the patient may experience as alternating pain-free and painful intervals; the window of benefit from thrombolytic therapy may be greater than 12 hours if antegrade flow was even briefly restored.

Contraindications to Thrombolytic Therapy

Contraindications to thrombolytic therapy include all conditions that predispose a patient to significant bleeding. The most feared bleeding complication is intracerebral hemorrhage, which is fatal in over half of cases. Risk factors for intracerebral bleeding include advanced age, low body weight, hypertension, warfarin use, and previous stroke.[53,59] Patients with gastrointestinal bleeding and those who have recently undergone surgery are also at increased risk for bleeding. Even when risk factors for bleeding are present, however, the potential benefits of thrombolytic therapy may still outweigh the risks. For example, although the elderly have a higher risk of intracerebral bleeding than younger patients, elderly patients should certainly be considered candidates for thrombolytic therapy, because their increasing absolute mortality results in a greater reduction in absolute mortality with thrombolytic therapy than is seen in younger patients.[53]

In patients with ECG findings other than ST segment elevation or bundle branch block, thrombolytic therapy has been found to be either of no use or deleterious; its use is not recommended in such patients.[17,53]

Choice of Thrombolytic Agent

Many different thrombolytic regimens have been proved effective for the treatment of acute myocardial infarction, and many more are being studied. In principle, the preferred thrombolytic regimen would restore normal antegrade blood flow to an occluded coronary artery most rapidly and in the greatest number of patients, would have the lowest reocclusion rate, and would be associated with the lowest risk of severe hemorrhagic

complications. The first Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Arteries (GUSTO-I) trial evaluated four thrombolytic regimens to determine which was associated with the greatest overall survival and stroke-free survival at 30 days: (1) a regimen of front-loaded, weight-adjusted t-PA and I.V. heparin, (2) a regimen of streptokinase and I.V. heparin, (3) a regimen of streptokinase and subcutaneous heparin, and (4) a combination of I.V. t-PA and streptokinase given concurrently with I.V. heparin. Front-loaded t-PA was found to be moderately superior to the other thrombolytic regimens [*see Figure 7*].[56] However, because of the approximately 10 times greater cost of t-PA than I.V. streptokinase and the low margin of superiority of t-PA (one life saved per thousand patients treated), some physicians prefer the less expensive streptokinase therapy, particularly for patients at low risk of dying (such as those with uncomplicated inferior infarctions) and the elderly, who are more likely to have hemorrhagic complications with t-PA than with streptokinase; t-PA is associated with a greater frequency of intracerebral hemorrhage than streptokinase.[56] The recommendation of streptokinase in these patient groups is largely driven by its lower cost; if the costs of t-PA and streptokinase were similar, t-PA would most likely be the preferred therapy in all patient subgroups, with the possible exception of those at increased risk for intracerebral hemorrhage, in whom streptokinase might be preferred.

Streptokinase therapy is contraindicated in patients who have recently received a dose of streptokinase because of antibodies that form against the drug; these antibodies limit the efficacy of repeat doses and increase the risk of allergic reactions. It has been suggested that the drug not be readministered for at least 2 years.

New thrombolytic agents are continuously being developed in the hope of finding safer and more effective therapies. One such agent, reteplase, is a recombinant tissue plasminogen activator (rt-PA) that is a mutant of alteplase. Reteplase is easier to administer than alteplase; because of its longer half-life, it can be administered as two 10 mU boluses given 30 minutes apart, with concomitant aspirin and I.V. heparin administration. Several pilot studies suggest that reteplase has an early patency rate that is superior to the patency rates of streptokinase and alteplase. In the International Joint Efficacy Comparison of Thrombolytics (INJECT) trial, 6,010 patients with acute myocardial infarction received either reteplase or streptokinase within 12 hours after the onset of symptoms.[60] Mortality at 35 days, the primary end point of the study, was 9.02% for patients given reteplase, compared with 9.53% for patients given streptokinase, a nonsignificant difference (95% confidence interval, 1.98 to 0.96). This lack of significant difference indicates that reteplase was at least as effective as streptokinase.

In the GUSTO-III trial, reteplase was compared with t-PA in 15,059 patients with acute myocardial infarction who presented within 6 hours of symptom onset.[61] Patients received either reteplase or an accelerated infusion of t-PA. For patients receiving reteplase, the mortality at 30 days was 7.47%, compared with 7.24% for patients receiving t-PA ($P = 0.54$; odds ratio, 1.03; 95%

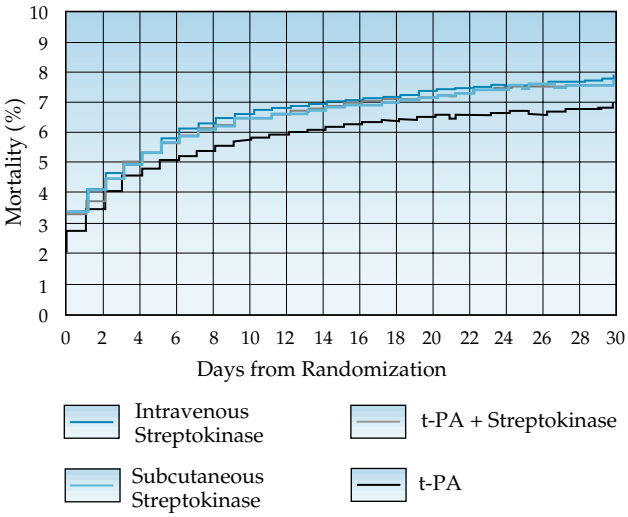
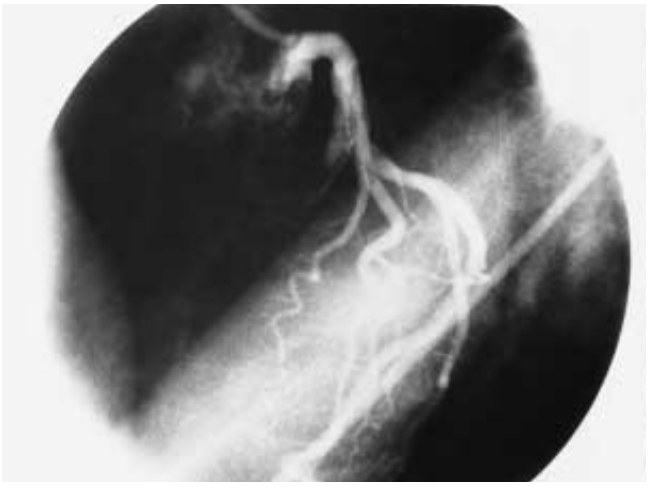


Figure 7 **The frequency of death or disabling stroke in the 30 days after enrollment in 41,021 patients in the Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Arteries (GUSTO-I) trial. Front-loaded t-PA was found to be superior to the other thrombolytic regimens.[56]**

a



b

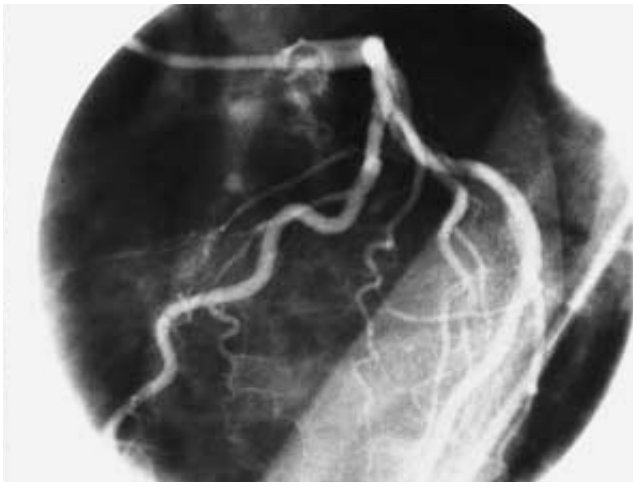


Figure 8 **(*a*) Left anterior oblique view of an occluded left anterior descending artery in a patient suffering an acute anterior myocardial infarction. (*b*) Patency was restored with direct coronary angioplasty 17 minutes after the patient had arrived in the catheterization laboratory, and the patient had immediate resolution of his symptoms.**

confidence interval, 0.91 to 1.18). The mortality rates with the two agents were therefore similar, and the two agents are probably, although not definitely, equivalent in efficacy.

Combination Therapy

Combination therapy, defined as the use of a thrombolytic agent and a glycoprotein IIb/IIIa inhibitor, has been proposed as an alternative to thrombolytic therapy alone for the primary treatment of STEMI. This strategy is supported by data from a number of trials that demonstrated improved rates of TIMI-3 flow after combination therapy, as compared with thrombolytic therapy alone.[62,63] However, the results of two randomized, controlled trials evaluating clinical outcomes after the respective aforementioned reperfusion strategies have been somewhat disappointing.

The GUSTO-V trial randomized 16,588 patients who presented within 6 hours after symptom onset with STEMI to either standard-dose reteplase or half-dose reteplase and full-dose abciximab.[64] At 30 days, the incidence of death in the reteplase arm was 5.9%, compared with 5.6% in the combination-therapy arm ($P = 0.43$), suggesting no mortality benefit associated with combination therapy. However, five of 16 prespecified secondary end points were reduced to a statistically significant degree ($P < 0.05$), suggesting a beneficial impact of combination therapy on the incidence of recurrent ischemic events and the mechanical and electrical complications of acute myocardial infarction (e.g., nonfatal reinfarction, recurrent ischemia, ventricular fibrillation, sustained ventricular tachycardia, and atrioventricular block). However, these clinical benefits were offset by an increased incidence of bleeding of any kind (13.7% with monotherapy versus 24.6% with combination therapy; $P < 0.0001$); severe or moderate, spontaneous, nonintracranial bleeding (1.9% versus 4.3%, $P < 0.0001$); severe bleeding (0.5% versus 1.1%; $P < 0.0001$), and bleeding sufficient to require blood transfusion (4.0% versus 5.7%; $P < 0.0001$). Furthermore, there were subgroups of patients in whom intracranial bleeding was increased by combination therapy; there was a significant ($P = 0.033$) association between age (< 75 or ≥ 75 years) and intracranial hemorrhage in the combination-therapy arm.

The Assessment of the Safety and Efficacy of a New Thrombolytic (ASSENT)–3 trial randomized 6,095 patients to one of three treatment regimens: full-dose tenecteplase and enoxaparin, half-dose tenecteplase and weight-adjusted low-dose unfractionated heparin infusion plus 12-hour infusion of abciximab, or full-dose tenecteplase and weight-adjusted unfractionated heparin infusion for 48 hours.[65] In-hospital mortality was not significantly different between the three groups. The abciximab treatment regimen reduced in-hospital rates of reinfarction and refractory ischemia. Although rates of intracranial hemorrhage were similar, other types of major bleeding were significantly higher in the abciximab arm of the trial. The enoxaparin arm suggested that there is a higher rate of intracranial bleeding in the elderly; this finding was corroborated by the ASSENT-PLUS study, which necessitated a change in dose of enoxaparin when administered with a thrombolytic agent in elderly patients.[66]

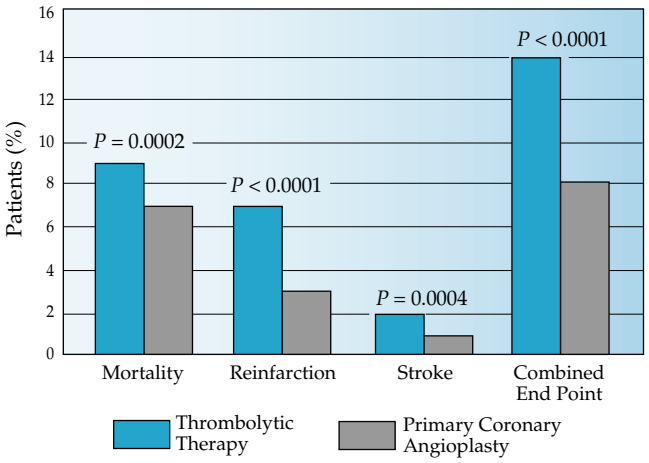


Figure 9 **Results from a quantitative review of 23 randomized trials of primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction.[72] Clinical outcome was improved in patients who received angioplasty.**

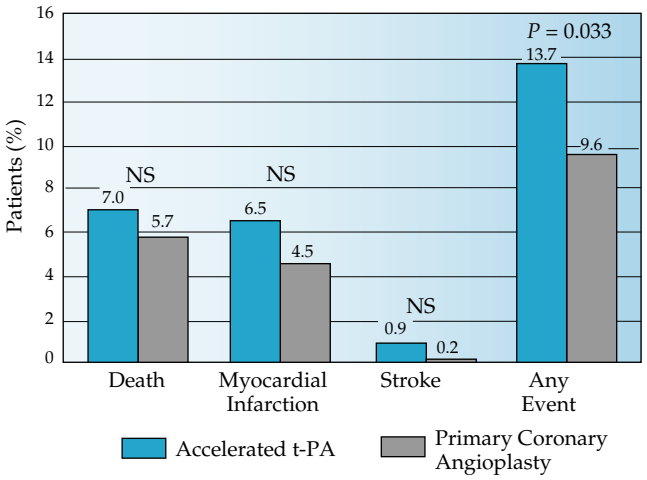


Figure 10 **Results from the GUSTO-IIb substudy trial comparing primary coronary angioplasty and accelerated t-PA indicate that primary angioplasty was associated with a lower mortality, reinfarction rate, and frequency of stroke in the 30 days after enrollment than was accelerated t-PA.[73] (NS—not significant)**

PRIMARY CORONARY ANGIOPLASTY

Coronary angioplasty without antecedent thrombolytic therapy, termed primary or direct coronary angioplasty, has been studied in the treatment of acute myocardial infarction [*see Figure 8*]. In prospective, randomized trials comparing primary coronary angioplasty with different thrombolytic agents, primary coronary angioplasty was associated with a lower morbidity and mortality than thrombolytic therapy.[67-71] Although most of the individual trials were too small to detect statistically significant differences in mortality, pooled data from these trials suggest that primary coronary angioplasty is the preferred therapy for acute myocardial infarction at institutions where it can be performed without delay [*see Figure 9*].[72]

The GUSTO-IIb trial was designed to be large enough to confirm the reduction in mortality found in the smaller randomized trials.[73] The results of the study indicate that compared with thrombolytic therapy, primary coronary angioplasty is associated with a lower mortality, reinfarction rate, and frequency of stroke in the 30 days after enrollment [*see Figure 10*]. However, the degree of benefit associated with primary coronary angioplasty was much smaller than that seen in the earlier randomized studies; this finding was in part related to the lower frequency with which patients assigned to undergo angioplasty in GUSTO-IIb actually underwent the procedure and in part relat-

ed to the lower frequency with which normal antegrade coronary blood flow was achieved in patients who did undergo coronary angioplasty.

The consistency of the results favoring primary coronary angioplasty and the greater speed and frequency with which coronary angioplasty can restore flow to an occluded coronary artery support the conclusion that primary coronary angioplasty is preferable to thrombolytic therapy at institutions where it can be performed quickly with a high success rate.[74] Studies have shown that excessive delay in performing primary coronary angioplasty and operator inexperience lead to a higher mortality than that seen when primary coronary angioplasty is performed rapidly by experienced operators.[36] It has been recommended that primary angioplasty be performed only in hospitals where a high success rate and low complication rate can be demonstrated and where primary angioplasty is performed in at least 80% to 90% of patients in whom acute myocardial infarction is confirmed.[72,75] The need for surgical backup is controversial, as excellent results have been obtained at centers without surgical backup.[76] However, surgical backup is recommended because approximately 5% of patients with acute myocardial infarction who undergo immediate coronary angiography require emergency surgery either for angioplasty that has failed or, more commonly, because lethal coronary anatomy precludes primary angioplasty.

Immediate transfer for primary angioplasty is an alternative treatment strategy for patients with STEMI initially assessed at a hospital without on-site cardiac surgery facilities. The DANAMI-2 investigators randomized STEMI patients to thrombolysis or primary angioplasty.[42] This included a prespecified substudy of patients admitted to a hospital without primary angioplasty facilities; these patients were randomized to immediate thrombolytic therapy or ambulance transfer for primary angioplasty (providing this procedure could be performed within 3 hours of randomization). Among patients who were randomized to hospitals without primary angioplasty facilities, the primary end point of death, myocardial infarction, or disabling stroke at 30 days was reached in 8.5% of the patients in the primary angioplasty group, as compared with 14.2% of those in the fibrinolysis group ($N = 1,129$; $P = 0.002$). Similarly, the PRAGUE-2 investigators concluded that long-distance transport from a community hospital to a facility with angioplasty facilities in the acute phase of STEMI is safe and is associated with decreased mortality in patients who present more than 3 hours after symptom onset.[41]

Glycoprotein IIb/IIIa Inhibitor Therapy and Primary Angioplasty for Acute Myocardial Infarction

The role of the platelet glycoprotein inhibitor abciximab in conjunction with primary coronary angioplasty has been examined in the Randomized, Placebo-Controlled Trial of Abciximab with Primary Angioplasty for Acute Myocardial Infarction (RAPPORT), the Intracoronary Stenting and Antithrombotic Regimen–2 (ISAR-2) study, the Abciximab before Direct Angioplasty and Stenting in Myocardial Infarction Regarding Acute and Long-term Follow-up (ADMIRAL) study, and, most recently, the Controlled Abciximab and Device Investigation to Lower Late Angioplasty Complications (CADILLAC) study.[77-80] The results of the four studies differ, in part because the trials used different end points, in part because of the high frequency of noncompliance with the protocol in some trials, and in part because of differences in the treatments utilized (balloon angioplasty alone versus balloon angioplasty followed by stent placement).

The largest of the four studies, the CADILLAC study, found that abciximab is beneficial at reducing major adverse events, but the benefit appeared to be limited to patients undergoing balloon angioplasty without stent placement. The apparent lack of benefit following treatment with abciximab in patients in the CADILLAC trial who received a coronary stent is contrary to the results of the ISAR-2 and ADMIRAL trials, in which abciximab was found to be beneficial in patients receiving stents. There are data suggesting that stent placement in the setting of acute myocardial infarction slightly reduces the frequency with which normal antegrade blood flow in the infarct-related artery is achieved.[81] This would suggest that glycoprotein IIb/IIIa inhibitors should be beneficial in this setting. Taken together, the results of these studies that are currently available suggest that abciximab is beneficial in patients with acute myocardial infarction but that the benefit in patients undergoing balloon angioplasty alone may differ from that in patients who undergo balloon angioplasty with stent placement. Clearly, however, stents markedly reduce the frequency with which a repeat revascularization procedure is needed in the months after the angioplasty procedure.[81] The combined use of stents and platelet glycoprotein inhibitors may maximize the frequency with which normal antegrade blood flow is achieved while reducing the need for repeat procedures in the following year.

One study compared the outcome of thrombolytic therapy using t-PA with that of primary angioplasty utilizing both stents and abciximab.[82] The reduction in infarct size was far greater in the group undergoing primary angioplasty; the clinical outcome was also better in the patients who underwent angioplasty.

CORONARY ARTERY BYPASS SURGERY

Coronary artery bypass surgery can restore blood flow in an occluded infarct-related artery. However, because of the time required to perform coronary angiography and to transport patients to the operating room, reperfusion is achieved more slowly with bypass surgery than with thrombolytic therapy and primary coronary angioplasty.[83] Emergency coronary artery bypass surgery should generally be reserved for patients in whom immediate angiography reveals coronary anatomy that precludes primary coronary angioplasty; for patients in whom angioplasty has failed; and for patients with a ventricular septal defect, severe mitral regurgitation, or myocardial rupture.

RESCUE CORONARY ANGIOPLASTY

Depending on the regimen used, only 33% to 60% of patients treated with thrombolytic therapy have restoration of normal antegrade flow in the infarct-related artery 90 minutes after the initiation of therapy.[55] Accordingly, immediate coronary angiography has been studied to determine whether patients with persistent occlusion of the infarct-related artery benefit from coronary angioplasty; this procedure has been termed rescue angioplasty. A single small, randomized trial has examined the clinical outcome of patients with anterior infarction and coronary occlusion that persist despite thrombolytic therapy.[84] Patients were randomized to either undergo rescue coronary angioplasty or receive continued medical therapy alone. The results of the trial suggested improved outcome with rescue angioplasty, although the benefits were not compelling. Three additional randomized trials evaluated the role of rescue angioplasty.[85-87] Analyzed together, the four trials suggest that rescue angioplasty offers benefit, although the data are not compelling. Although use of coronary stents and platelet glycoprotein inhibitors improves the results of percutaneous revascularization

procedures and would be expected to further increase the benefit of angioplasty after failed thrombolytic therapy, this has not yet been proved. There are insufficient data to recommend immediate angiography and angioplasty in all patients early after thrombolytic therapy. Immediate angiography is most likely to be beneficial in patients with large myocardial infarctions in whom persistent pain, ST segment elevation, or hemodynamic compromise is present more than 90 minutes after the administration of a thrombolytic agent.

The routine performance of angioplasty immediately after the administration of thrombolytic therapy in all patients with a significant residual stenosis (not just those patients with occluded coronary arteries) has been well studied in three prospective, randomized trials and has been found to be either of no benefit or deleterious.[43,88] Angioplasty should not be routinely performed in such patients.

Stents appear to improve the ability to achieve arterial patency early after thrombolytic therapy, as compared with balloon angioplasty alone[89]; therapy with a glycoprotein IIb/IIIa inhibitor may also do so, although an increase in bleeding has been seen when glycoprotein IIb/IIIa inhibitors are used early after full-dose thrombolytic therapy.[90] Data from several pilot studies suggest that the combination of a fibrin-specific thrombolytic agent, either t-PA or reteplase, combined with the glycoprotein IIb/IIIa inhibitor abciximab, may actually facilitate the performance of angioplasty rather than reduce its safety and efficacy, as was seen when balloon angioplasty was performed after thrombolytic therapy.[62,63] Hence, the term facilitated angioplasty has been coined for the routine performance of angioplasty after the combination of half-dose thrombolytic therapy with a glycoprotein IIb/IIIa inhibitor.

The Plasminogen-activator Angioplasty Compatibility Trial (PACT) investigators randomized 606 patients to a reduced dose of a short-acting fibrinolytic regimen (50 mg bolus of reteplase) or a placebo followed by immediate angiography with angioplasty, if needed.[91] In the group receiving reduced-dose reteplase, there was no increase in the incidence of stroke or major bleeding, and convalescent left-ventricular ejection fraction was higher, as evidenced by a patent infarct–related artery (TIMI-3 flow) on arrival in the catheterization laboratory (62%) or a TIMI-3 flow that was achieved by angioplasty within 1 hour after administration of the drug bolus (58%). However, only 12% of successful angioplasty procedures resulted in a patent infarct–related artery within 1 hour, because of routine delay in transfer to the catheterization laboratory, and there was no difference between the two treatment groups by a traditional intention-to-treat analysis.

The Southwest German Study in Acute Myocardial Infarction III (SIAM III) investigators randomized 163 patients initially treated with thrombolysis at a community hospital with no on-site PCI facilities either to hospital transfer and immediate stenting within 6 hours of thrombolysis or to delayed, elective stenting approximately 2 weeks after acute myocardial infarction.[92] Transfer and immediate stenting were associated with a statistically significant reduction in the incidence of the composite primary end point (i.e., death, reinfarction, ischemic events, and target-lesion revascularization at 6 months), as compared with delayed stenting (25.6% versus 50.6%; $P = 0.001$). The difference in outcome was driven by events occurring during the 2 weeks that patients waited for elective stenting in the deferred PCI group. This trial design does not address whether PCI performed 1 to 2 days after thrombolytic therapy, as is usual in the United States, is as effective as PCI performed immediately after thrombolytic therapy. A randomized trial that analyzed the use of combination therapy (half-dose thrombolytic therapy with full-dose abciximab) before routinely performing primary PCI was recently reported.[93] In the Bavarian Reperfusion Alternatives Evaluation (BRAVE) Trial, patients with STEMI were randomly assigned to receive either half-dose reteplase and full-dose abciximab or abciximab alone, and all patients underwent PCI as rapidly as possible. No advantage was seen with combination therapy; in fact, more bleeding complications occurred in this group. The results argue against the use of a facilitated PCI approach using the combination regimen of reteplase and abciximab.

Adjunctive Medical Therapy

INTRAVENOUS HEPARIN

The need for I.V. heparin after thrombolytic therapy varies with the thrombolytic agents used. A retrospective analysis of the GUSTO-I trial suggested that I.V. heparin with a partial thromboplastin time of 50 to 70 seconds was associated with the best clinical outcome in patients treated with t-PA.[94] Data from GUSTO-I also suggest that I.V. heparin is not required when I.V. streptokinase is used, although heparin is recommended in patients with large anterior infarctions to prevent the development of apical mural thrombus and embolization.[56] In patients in whom I.V. heparin is not administered, subcutaneous heparin should be administered during the period of bed rest to reduce the risk of deep vein thrombosis.[95]

The optimal duration of I.V. heparin therapy is unclear. Standard practice was to administer I.V. heparin for 3 to 5 days, although patients are now often discharged after only 3 days. It is recommended that heparin be discontinued more than 24 hours before patient discharge from the hospital because of the possibility of a rebound effect and recurrent thrombosis within 24 hours after cessation of heparin therapy.[96]

Randomized studies from the prethrombolytic era suggested that administration of I.V. heparin reduces mortality and reinfarction in patients not treated with thrombolytic agents.[95] Aspirin and beta blockers were not routinely administered in those early trials; consequently, the true benefits of heparin when these drugs are administered are unknown. However, on the basis of the early data, I.V. heparin is generally recommended for patients with suspected myocardial infarction who are not treated with thrombolytic therapy.[75]

LOW-MOLECULAR-WEIGHT HEPARIN

Low-molecular-weight heparins (depolymerized unfractionated heparin with a mean molecular weight of approximately 5,000) have a number of potential pharmacokinetic advantages over the parent molecule, including decreased binding to plasma proteins, decreased sensitivity to platelet factor 4, enhanced factor Xa activity, and improved bioavailability. These factors are associated with a predictable dose-response relationship and, combined with the ease of administration (once or twice daily S.C. dosing regimens) and lower rates of heparin-induced thrombocytopenia, have increased investigators' interest in low-molecular-weight heparins in preference to unfractionated heparin for the treatment of myocardial infarction. The results of four large randomized trials comparing three different low-molecular-weight heparins with unfractionated heparin have suggested that low-molecular-weight heparin is at least as effective

in reducing ischemic events in patients with NSTEMI acute coronary syndromes. Following patients with STEMI, the second trial of Heparin and Aspirin Reperfusion Therapy (HART II) and the Acute Myocardial Infarction–Streptokinase (AMI-SK) trial demonstrated evidence of improved rates of reperfusion when the low-molecular-weight heparin enoxaparin was combined with t-PA or streptokinase, respectively.[97,98] The ASSENT-3 trial compared three regimens: full-dose tenecteplase and enoxaparin, full-dose tenecteplase and unfractionated heparin, and half-dose tenecteplase and abciximab for the treatment of thrombolytic-eligible STEMI; the study revealed that enoxaparin fared better than unfractionated heparin in terms of 30-day mortality, in-hospital reinfarction, and in-hospital refractory ischemia (P = 0.0001) at 30 days.[99] However, at 1 year, the benefits had diminished, and the mortality with enoxaparin was identical to that with unfractionated heparin. In addition, when the data from ASSENT-3 and ASSENT-3–PLUS (a study examining the administration of tenecteplase with enoxaparin in the prehospital setting) were pooled, a marked and prohibitive increase in the risk of intracerebral hemorrhage was seen in the elderly.[66,99] As a result, ongoing trials utilizing enoxaparin with a fibrinolytic agent have adjusted the dose of enoxaparin downward in elderly patients. At present, the adjunctive anticoagulant that should be administered with fibrinolytic therapy and its optimal dose are not known.

DIRECT THROMBIN INHIBITORS

Direct thrombin inhibitors are an attractive alternative to indirect thrombin inhibitors, such as heparin or low-molecular-weight heparins, particularly because they block both circulating and clot-bound thrombin. A collaborative meta-analysis of phase-3 trials of direct thrombin inhibitors for the treatment of acute coronary artery syndromes demonstrated superiority over unfractionated heparin for the prevention of the composite end point of death or myocardial infarction.[100] The Hirulog and Early Reperfusion or Occlusion (HERO)–2 trial demonstrated a similar benefit in a comparison of bivalirudin with heparin in patients with STEMI.[101] Bivalirudin was associated with a 30% reduction in the incidence of reinfarction, but mortality did not decrease; however, the bivalirudin arm of the study exhibited a trend toward more bleeding events.

BETA BLOCKERS

Numerous studies of beta-blocker therapy in patients with acute myocardial infarction have documented significant reductions in in-hospital and long-term mortality. Early administration of beta blockers has been promoted because it may reduce infarct size by reducing heart rate, blood pressure, and myocardial contractility, all of which diminish myocardial oxygen demand. Meta-analysis of the effects of early administration of I.V. beta blockers in 27,486 patients with acute myocardial infarction enrolled in 28 randomized trials revealed a 14% reduction in mortality during the first week of therapy; reinfarction was reduced by 18%.[102]

The TIMI-II study compared immediate beta-blocker therapy with deferred beta-blocker therapy in acute myocardial infarction; all patients also received I.V. t-PA.[103] Results indicated that immediate beta-blocker therapy reduced the incidence of nonfatal reinfarction and recurrent ischemia, compared with oral metoprolol therapy begun on the sixth hospital day; as in earlier studies, only about 40% of patients with acute myocardial infarction were eligible for acute beta-blocker therapy.[43] There are also data suggesting that immediate beta-blocker therapy reduces the risk of intracranial hemorrhage after lytic therapy.[104] It is recommended that all patients with acute myocardial infarction without contraindications receive I.V. beta blockers as early as possible, whether or not they receive reperfusion therapy.

In patients in whom contraindications preclude early beta-blocker therapy, reevaluation should take place before discharge. Many patients will no longer have contraindications at the time of discharge. Patients without contraindications should be routinely started on beta-blocker therapy before discharge from the hospital. The optimal duration of benefit remains unclear, but it appears that the benefit of beta-blocker therapy is maintained for years. Patients with the largest infarctions benefit the most from the use of beta blockers. Current recommendations are that beta-blocker therapy be continued indefinitely in the absence of contraindications or side effects.

ANGIOTENSIN-CONVERTING ENZYME INHIBITORS

Several large randomized, controlled clinical trials evaluating the use of angiotensin-converting enzyme (ACE) inhibitors early after acute myocardial infarction have been performed; all but one trial revealed a significant reduction in mortality. Meta-analysis of these large trials and many smaller trials, which together included over 100,000 patients, suggested a 6.5% reduction in deaths, with an absolute reduction in mortality of 4.6 deaths per 1,000 patients among those treated with an ACE inhibitor.[105] All patients with significant ventricular dysfunction (an ejection fraction < 40%) without contraindications should be treated with an ACE inhibitor; treatment should begin within the first 48 hours of infarction and be increased cautiously to avoid hypotension. If hypotension results from the early administration of ACE inhibitors, short-term mortality may be increased.[106]

The benefit of ACE inhibitors is clear in patients with large anterior infarctions and an ejection fraction less than 40%; whether patients with an ejection fraction greater than 40% benefit from ACE inhibitor therapy is less clear. However, the results of two large trials suggest that patients with a normal ejection fraction after myocardial infarction, as well as even patients with coronary artery disease without a previous myocardial infarction, have a reduction in mortality when treated with an ACE inhibitor. In the Heart Outcomes Prevention Evaluation (HOPE) study, 9,297 patients 55 years of age or older with vascular disease (or with diabetes and another cardiovascular risk factor) without a low ejection fraction or congestive heart failure were randomly assigned to receive either the ACE inhibitor ramipril or placebo for a mean of 5 years.[107] The reduction in the combined end point of death from cardiovascular causes, myocardial infarction, or stroke with ramipril was remarkable; it occurred in 17.7% of placebo-treated patients versus 14.1% of patients receiving ramipril (relative risk, 0.78; 95% confidence interval, 0.70 to 0.86; $P < 0.001$). A statistically significant reduction was also present in the individual end points of cardiovascular death, myocardial infarction, and stroke. The study was stopped prematurely by the Data Safety Monitoring Board when clear evidence of a beneficial effect of ramipril was found. These findings are supported by the results of the European Trial on Reduction of Cardiac Events with Perindopril in Patients with Stable Coronary Artery Disease (EUROPA) trial, which randomized 12,218 patients with stable coronary artery disease and no evidence of congestive heart failure to 8 mg of perindopril or conventional therapy.[108] Treatment with perindopril was associated with a highly statistically significant reduction in the

incidence of fatal and nonfatal myocardial infarction at 4 years. Whether these favorable results are unique to tissue-specific ACE inhibitors or represent a class effect is unknown.

INTRAVENOUS NITROGLYCERIN

Randomized studies examining the role of I.V. nitroglycerin in acute myocardial infarction revealed beneficial effects on left ventricular function and a reduction in infarct size and mortality.[109] However, these studies were small and were performed before the reperfusion era. To determine whether nitroglycerin therapy is beneficial in patients treated with reperfusion, 58,050 patients with acute myocardial infarction in the fourth International Study of Infarct Survival (ISIS-4) were randomized to receive either oral controlled-release mononitrate therapy or placebo; thrombolytic therapy was administered to patients in both groups.[105] The results of this study revealed no benefit to the routine administration of oral nitrate therapy in this setting. Similar results were seen among 19,000 patients in the third Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico (GISSI-3) study, in whom I.V. nitroglycerin was administered for the first 24 hours, followed by transdermal nitrates.[110] Whether these disappointing results in the ISIS-4 and GISSI-3 trials were caused by the routes of administration of the nitroglycerin preparation or the administration of thrombolytic therapy is unknown. However, on the basis of existing data, it does not appear that the routine administration of nitroglycerin to patients receiving early thrombolytic therapy is beneficial. I.V. nitroglycerin is probably most likely to be beneficial in patients with persistent or recurrent chest pain after reperfusion therapy and in patients in whom reperfusion therapy is not administered.

PROPHYLACTIC ANTIARRHYTHMIC THERAPY

Previously, routine prophylactic antiarrhythmic therapy with I.V. lidocaine was recommended for all patients in the early stages of acute myocardial infarction. However, studies have revealed that prophylactic therapy with lidocaine does not reduce and may actually increase mortality because of an increase in the occurrence of fatal bradyarrhythmia and asystole.[111] Neither I.V. lidocaine nor other antiarrhythmic agents are recommended as prophylactic therapy for patients without malignant ventricular ectopy.[111,112]

CALCIUM CHANNEL ANTAGONISTS

Calcium channel antagonists should not be routinely administered for acute myocardial infarction. Calcium channel antagonists have been studied in prospective, double-blind, placebo-controlled trials; and neither verapamil,[113,114] nifedipine,[115,116] nor diltiazem[117] appears to reduce postinfarction mortality. Verapamil and diltiazem may be useful in patients with preserved left ventricular function and no heart failure in whom contraindications to beta blockers exist.[118,119] However, the data are insufficient to recommend the routine administration of these agents. On the basis of existing data, treatment with calcium channel blockers should be reserved for patients with ischemia that persists despite use of aspirin, beta blockers, nitrate therapy, and I.V. heparin and for patients with other indications for their administration.

MAGNESIUM

Magnesium has been studied in many prospective, randomized trials of acute myocardial infarction, and the results have been conflicting. Magnesium is involved in hundreds of enzymatic steps and produces systemic and coronary vasodilatation, inhibits platelet function, and reduces reperfusion injury. Meta-analysis of seven prospective, randomized trials revealed a significant reduction in mortality with the use of magnesium (odds ratio, 0.44; confidence interval, 0.27 to 0.71).[120] Subsequently, the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2) revealed a 25% reduction in mortality with the use of magnesium in over 2,000 patients who had acute myocardial infarction.[121] However, in ISIS-4, in which 58,050 patients were randomized to receive either I.V. magnesium or no magnesium, there was no reduction in 30-day mortality.[105] It is possible that the later administration of magnesium in this study, compared with the previous studies, and the concomitant use of thrombolytic therapy in 70% of patients contributed to the lack of efficacy of magnesium in ISIS-4; only one third of patients in the LIMIT-2 study received thrombolytic therapy. A subsequent small randomized trial reexamined the role of magnesium in patients in whom reperfusion therapy was not administered and did find a mortality reduction associated with its use.[122] However, the Magnesium in Coronaries (MAGIC) trial investigators, who randomized 6,213 patients with STEMI to a 2 g intravenous bolus of magnesium followed by a 17 g infusion over 24 hours or matching bolus and 24-hour placebo infusions failed to demonstrate any statistically or clinically significant effect on 30-day mortality.[123] Therefore, on the basis of the existing evidence, current recommendations are that magnesium not be routinely given to patients in whom reperfusion therapy is administered. It is possible that magnesium is of benefit, particularly in patients not receiving reperfusion therapy. Magnesium is clearly indicated in patients with myocardial infarction who have torsade de pointes–type ventricular tachycardia and in patients with magnesium deficiency.

Complications of Acute Myocardial Infarction

VENTRICULAR ARRHYTHMIAS

Ventricular arrhythmias are a frequent cause of death in the earliest stages of acute myocardial infarction. The development of coronary care units, continuous ECG surveillance, and defibrillators in the 1960s led to a reduction in mortality from acute myocardial infarction through the prompt identification and treatment of ventricular arrhythmia; and emergency medical technicians have reduced outpatient mortality in the earliest minutes of myocardial infarction. In cities with well-developed emergency response systems, such as Seattle, Washington, and Rochester, Minnesota, where the average response time is less than 5 minutes, survival of patients with myocardial infarction complicated by cardiac arrest has increased.[4] In fact, long-term survival of patients who have undergone rapid defibrillation after out-of-hospital cardiac arrest is similar to that of age-, sex-, and disease-matched patients who did not have out-of-hospital cardiac arrest; the quality of life of the majority of survivors is similar to that of the general population.[124]

Ventricular Fibrillation

In the setting of acute myocardial infarction, ventricular fibrillation is often described as either primary, when it occurs in the absence of hypotension or heart failure, or secondary, when hypotension or heart failure is present. Primary ventricular fibrillation occurs in approximately 3% to 5% of patients with acute myocardial infarction; the peak incidence is in the first 4 hours of infarction. Primary ventricular fibrillation is infrequent more

than 24 hours after symptom onset. Mortality is increased in patients who suffer this complication.[125,126] In patients who are successfully resuscitated and survive to hospital discharge, however, the long-term prognosis does not appear to be affected.[125] Although lidocaine was shown to reduce the occurrence of primary ventricular fibrillation, mortality in patients receiving lidocaine increased because of an increase in fatal bradycardia and asystole; therefore, prophylactic lidocaine is no longer recommended if defibrillation can rapidly be performed.[111] Beta blockers may reduce the early occurrence of ventricular fibrillation and should be administered to patients who have no contraindications.

Hypokalemia is a risk factor for primary ventricular fibrillation and should be rapidly corrected if present. When ventricular fibrillation occurs, rapid defibrillation with 200 to 300 joules should be attempted, and repeated shocks of 360 joules should be administered. The Advanced Cardiac Life Support (ACLS) guidelines recommend medical therapy, including epinephrine, lidocaine, and bretylium; in addition, I.V. amiodarone should be considered in patients in whom defibrillation is initially unsuccessful.

Secondary ventricular fibrillation is associated with a high mortality, in part because of the underlying hypotension and heart failure. Treatment must be aimed not only at terminating the arrhythmia but also at treating the hemodynamic abnormalities and their causes.

Ventricular Tachycardia

Ventricular tachycardia (three or more consecutive ventricular ectopic beats) is common in patients with acute myocardial infarction; however, short runs of nonsustained ventricular tachycardia are no longer believed to predispose a patient to sustained ventricular tachycardia or ventricular fibrillation. In patients in whom sustained or hemodynamically significant ventricular tachycardia occurs, prompt electrical cardioversion should be performed. If the ventricular tachycardia is monomorphic, synchronic cardioversion with 100 joules should first be attempted. As with ventricular fibrillation, polymorphic ventricular tachycardia should be treated with unsynchronized discharge. Prolonged runs of asymptomatic ventricular tachycardia can be initially treated with I.V. lidocaine, procainamide, or amiodarone. These medications may also be helpful in reducing recurrent ventricular tachycardia.

ATRIAL ARRHYTHMIA

Atrial Fibrillation

Atrial fibrillation is the most common atrial arrhythmia in acute myocardial infarction, occurring in 10% to 16% of patients. Atrial fibrillation may result either from an acute increase in left atrial pressure caused by left ventricular dysfunction or from atrial ischemia as a result of occlusion of a coronary artery (usually the right coronary artery) proximal to the origin of atrial branches. The incidence of atrial fibrillation is decreased in patients given thrombolytic therapy.[56]

The treatment of atrial fibrillation in acute myocardial infarction should be similar to the treatment of atrial fibrillation in other settings. When there is hemodynamic compromise caused by loss of atrial systole or a rapid ventricular response with a reduction in cardiac output, cardioversion should be performed immediately. In patients with preserved left ventricular function in whom the atrial fibrillation is well tolerated, beta-blocker therapy is indicated. Verapamil and diltiazem may also be effective in such patients. In patients with congestive heart failure, digoxin is a reasonable alternative and may slow the ventricular response. If atrial fibrillation recurs, antiarrhythmic agents may be used, although their impact on clinical outcomes is unproved. The Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) investigators randomized 4,060 patients to either rhythm-control or rate-control treatment strategies.[127] The rhythm-control strategy achieved no survival advantage over the rate-control strategy and there were both fewer hospital admissions and fewer adverse drug reactions in the rate-control group.

BRADYARRHYTHMIAS AND HEART BLOCK

Sinus bradycardia is common in acute myocardial infarction, particularly in patients with inferior myocardial infarction. However, treatment with atropine and a temporary pacemaker is required infrequently and, generally, only in patients with significant hemodynamic compromise manifested by increased angina, hypotension, or congestive heart failure.

High-degree (second- or third-degree) heart block occurs in approximately 20% of patients with inferior infarction; it is uncommon with infarction at other sites.[128] About half of the cases of heart block seen with inferior infarction are Wenckebach-type second-degree heart block; the remainder are cases of third-degree heart block. The heart block is often easily treated with atropine, but a temporary pacemaker is required in as many as 50% of cases. The heart block generally lasts for hours to days; placement of a permanent pacemaker is needed in fewer than 1% of cases. However, the development of heart block with inferior infarction is associated with a threefold to fourfold increase in in-hospital mortality over inferior infarction without heart block.[128,129] The increased mortality appears to result from the association between heart block and more severe left and right ventricular infarction rather than from the heart block itself or treatment of the heart block.

Heart block during anterior infarction is uncommon, occurring in fewer than 1% of cases. It is generally associated with extensive left ventricular myocardial infarction involving the conduction system below the atrioventricular node and carries a very poor prognosis.

MITRAL REGURGITATION

Mitral regurgitation may result from injury to any of the components of the mitral valve apparatus, including the papillary muscles and ventricular walls to which they attach. Mild mitral regurgitation is common in acute myocardial infarction and is present in nearly 50% of patients. Severe mitral regurgitation caused by acute myocardial infarction is rare and generally results from partial or complete rupture of a papillary muscle. The characteristic murmur of severe chronic mitral regurgitation may not be present with acute rupture of a papillary muscle. Instead, a decrescendo systolic murmur is often present, extending less throughout systole as systemic arterial pressure falls and left arterial pressure rises. In many cases, the significance of the murmur is not recognized. The blood supply of the anterior papillary muscle arises from branches of both the left anterior descending and the circumflex arteries; therefore, rupture of the anterior papillary muscle is rare. However, the posterior papillary muscle receives blood only from the dominant coronary artery (the right coronary artery in nearly 90% of patients); thrombotic occlusion of this artery may cause rupture of the posterior papillary muscle, resulting in severe mitral regurgitation. Severe mitral regurgitation is 10 times more likely to occur with inferior in-

farction than with anterior infarction. Acute severe mitral regurgitation is poorly tolerated and generally results in pulmonary edema, often with cardiogenic shock. Prompt surgical repair is recommended. Although the mortality associated with mitral valve surgery is high in this setting, approaching 50%, survival appears to be greater than with medical therapy alone. Therapy aimed at reducing left ventricular afterload, such as use of I.V. nitroprusside and an intra-aortic balloon pump, reduces the regurgitant volume and increases forward blood flow and cardiac output and may be helpful as a temporizing measure.

VENTRICULAR SEPTAL DEFECTS

Ventricular septal defects are slightly more frequent in patients with anterior infarction than in patients with inferior infarction. The characteristic holosystolic murmur of ventricular septal defects may be difficult to distinguish from that of severe mitral regurgitation; however, ventricular septal defects are generally better tolerated and less frequently result in severe congestive heart failure. Surgical repair is recommended and results in the best outcome when repaired emergently in the hemodynamically compromised patient. As with acute severe mitral regurgitation, therapy aimed at reducing afterload, including I.V. nitroprusside and an intra-aortic balloon pump, may be beneficial. Repair of the septum is generally more difficult when associated with inferior infarction, because there may not be a viable rim of myocardial tissue beneath the defect to facilitate repair. The surgical mortality associated with repair of a postinfarction ventricular septal defect is approximately 20% but is largely related to the age of the patient, whether cardiogenic shock is present, the infarction site, and the severity of the underlying coronary artery disease.

MYOCARDIAL RUPTURE

As more and more patients survive the acute phase of myocardial infarction because reperfusion therapy reduces myocardial infarct size, myocardial rupture has increased in frequency as a cause of early death. Myocardial rupture has been reported to account for more than 20% of in-hospital deaths in some series in the thrombolytic era. Physicians must have a heightened awareness of the diagnosis if a patient is to survive this catastrophic occurrence, because emergency surgery is required. Symptoms suggestive of rupture include repetitive vomiting, pleuritic chest pain, restlessness, and agitation. ECG evidence of rupture includes a deviation from the normal pattern of ST segment and T wave evolution. Resolution of ST segment elevation and T wave inversion, with maximal T wave negativity in the leads with maximal ST segment elevation, should normally occur; however, in patients with rupture, there is progressive or recurrent ST segment elevation and persistently positive T wave deflections or reversal of initially inverted T waves.[130] Echocardiography can quickly confirm the diagnosis. Even when emergency surgery is performed, fewer than 50% of patients survive to discharge.

RIGHT VENTRICULAR INFARCTION

Right ventricular infarction occurs in approximately one third of patients with acute inferior left ventricular infarction and is hemodynamically significant in approximately 50% of affected patients.[131] Hemodynamically significant right ventricular infarction associated with anterior infarction or isolated right ventricular infarction is rare. The classic findings associated with hemodynamically significant right ventricular infarction are hypotension with clear lung fields and an elevated jugular venous pressure, often with the Kussmaul sign. Although nearly all patients with right ventricular infarction suffer both right and left ventricular infarction, the characteristic hemodynamic findings of right ventricular infarction generally dominate the clinical course and must be the main focus of therapy. Right ventricular involvement during inferior myocardial infarction is associated with a significant increase in mortality, and aggressive attempts at early reperfusion should be pursued.[128,131] Prompt recognition of right ventricular involvement is clinically important because therapy that reduces right ventricular filling, such as use of nitrates or diuretics, should be avoided. Volume therapy should be administered to maintain cardiac output; in patients whose hypotension is refractory to volume therapy, dopamine may be beneficial. Heart block, which occurs in as many as 50% of patients with right ventricular infarction, should be treated rapidly, and maintenance of atrioventricular synchrony with dual atrial and ventricular pacing is often required to maintain filling of the ischemic noncompliant right ventricle and an adequate cardiac output.

Cardiogenic shock resulting from right ventricular infarction is generally reversible with these measures. Improvement in right ventricular function generally occurs over time, particularly in patients in whom reperfusion therapy was successful in achieving vessel patency.[125] In patients who survive the initial hospitalization, left ventricular function is the most potent predictor of long-term outcome.

STROKE

Extensive infarction of the anterior wall and apex of the left ventricle leads to thrombus formation in the apex of the left ventricle in approximately 30% of patients; systemic embolization occurs in about 15% of these patients. Left ventricular thrombus formation is much less common after inferior infarction. The thrombus generally appears within the first several days after infarction; it is more likely to embolize and cause stroke if it is pedunculated, protrudes into the left ventricular cavity, or is mobile. Left ventricular thrombus is an indication for anticoagulation with I.V. heparin, followed by warfarin therapy for 3 to 6 months.

Therapy that reduces infarct size, such as thrombolytic therapy, reduces the frequency of thrombus formation and therefore the risk of systemic embolization and stroke. However, in 0.3% to 1.0% of patients, thrombolytic therapy causes hemorrhagic stroke, most commonly in the 24 hours after its administration, which is fatal in more than 50% of cases. Hemorrhagic stroke is rare in acute myocardial infarction except as a consequence of thrombolytic therapy, although an ischemic stroke may become hemorrhagic because of thrombolytic, antiplatelet, and anticoagulation therapy. Hemorrhagic stroke, the most feared complication of thrombolytic therapy, is more likely in elderly patients; in patients with low body weight, with hypertension, or who have previously had a stroke; and in those on warfarin.[53,59] Although thrombolytic therapy decreases the risk of ischemic stroke, there is a slight net increase in the overall risk of stroke because of the risk of hemorrhagic stroke. Primary coronary angioplasty is believed to reduce the incidence of ischemic stroke without increasing the risk of hemorrhagic stroke.

Predischarge Exercise Testing

In patients with spontaneous postinfarction angina, congestive heart failure, hypotension, or malignant ventricular arrhythmia, exercise testing should generally be deferred and coronary angiography should be performed. However, in patients with-

out these high-risk characteristics, exercise testing is generally recommended before discharge from the hospital to assess a patient's functional capacity and ability to return to activities of daily living and work.[132] Most data indicating that predischarge exercise testing can identify patients at increased risk for cardiac events after discharge are from the prethrombolytic era, when the risk of adverse cardiac events was much higher. In the modern era, in which thrombolytic therapy or primary coronary angioplasty is frequently performed and in which aspirin, beta blockers, ACE inhibitors, and lipid-lowering agents are routinely administered—all of which reduce the frequency of adverse events in the years after discharge—it is difficult to identify patients at risk, because the adverse event rate is so low. Nonetheless, exercise testing is generally recommended to provide a measure of comfort to both the patient and the physician, to help determine the appropriateness of medical therapy, and to facilitate entry of the patient into a cardiac rehabilitation program.

Although predischarge exercise testing has been the standard of care in the United States for some time, only recently has a study examined whether therapy based on the results of a predischarge exercise test improves clinical outcome. The Danish Trial in Acute Myocardial Infarction (DANAMI) was the first study to examine the usefulness of exercise testing in patients treated with thrombolytic agents (a low-risk group) and to provide support for what has been the standard of care in the United States [*see Figure 11*].[133] The results of this study revealed that clinical outcome was improved in patients who received angiography and coronary angioplasty, compared with those who received medical therapy alone. Use of the results of exercise testing to decide whether or not to employ revascularization in patients without spontaneous angina is less common outside of the United States.

Patients with acute myocardial infarction who do not receive thrombolytic therapy or do not undergo primary angioplasty are at greater risk for adverse events after discharge from the hospital, and predischarge exercise testing is of even greater utility in such patients.

Prognostic variables indicating increased risk during exercise testing are exercise-induced angina or ST segment depression, particularly when it occurs during exercise at a low work load, and an abnormal drop in systolic blood pressure. However, electrocardiographic, symptomatic, and scintigraphic risk markers of ischemia (e.g., ST segment depression, angina, or a reversible perfusion defect) are less sensitive for identifying morbid and fatal outcomes than markers of left ventricular dysfunction or heart failure (e.g., exercise duration, impaired systolic blood pressure response, and peak left ventricular ejection fraction).[134] The patients at greatest risk are those unable to exercise; such patients have the highest mortality after discharge.[135]

The type of exercise test that should be performed has been the subject of controversy. It is generally recommended that only simple treadmill testing be performed before discharge; in patients with abnormalities in the baseline ECG, stress testing with perfusion imaging or stress echocardiography may be helpful. In patients without widespread abnormalities on the ECG, perfusion imaging or stress echocardiography is generally deferred until at least 4 weeks after discharge, when a more vigorous exercise test can be performed. Whether the predischarge treadmill test should be a low-level test or a more vigorous symptom-limited test is unclear. It has been shown that a symptom-limited Bruce protocol exercise test detects ischemia more frequently than a submaximal test; however, it is not known which test has

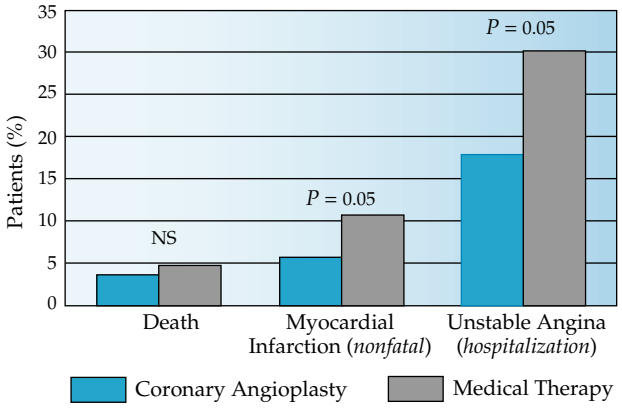


Figure 11 **In the Danish Acute Myocardial Infarction (DANAMI) study, 1,008 patients treated with thrombolytic therapy in whom exercise-induced ischemia was present on a predischarge exercise test were randomized to receive either coronary angioplasty or medical therapy alone. Clinical outcome was improved in patients in the invasive arm of the study. (NS—not significant)**

the greater positive and greater negative predictive value for identifying patients at risk. Currently, a lower-level exercise test is preferred, although a more vigorous test may be appropriate in patients likely to resume a more active and vigorous lifestyle shortly after discharge and in whom a low-level test may not cause the patient to expend the amount of energy he or she will be using during activities of daily living.

There has been concern that the use of beta blockers before the predischarge exercise test may mask the presence of significant coronary artery disease and prevent the identification of high-risk patients. This concern does not appear to be significant enough to outweigh the benefits of early beta-blocker therapy.

Secondary Prevention

PHARMACOTHERAPY

Lipid-Lowering Therapy

Recent studies have demonstrated that in patients with coronary artery disease, lipid-lowering therapy with HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A) reductase inhibitors reduces not only fatal and nonfatal infarction but also mortality from all causes. The Scandinavian Simvastatin Survival Study revealed a 42% reduction in cardiac mortality and a 30% reduction in all-cause mortality in 4,444 men and women with coronary artery disease over the 5.4 years of the study.[136] The reductions in mortality were similar in patients in the lowest and those in the highest quartiles of serum low-density lipoprotein (LDL) cholesterol. It has been demonstrated that postinfarction patients with an LDL cholesterol level at or above 130 mg/dl benefit from lipid-lowering therapy within as little as 2 years after the initiation of such therapy.[137] Initial measurement of cholesterol should be made within 24 hours after myocardial infarction; measurement of lipids 24 hours or more after myocardial infarction can be misleading in that cholesterol levels may be reduced below baseline levels during this period and remain low for up to 1 month. Early initiation of statins may be more beneficial than later initiation.[138] Exercise, weight reduction in overweight patients, avoidance of dietary saturated fat and cholesterol, and

smoking cessation have all been reported to favorably influence blood lipid levels and should be recommended whether or not lipid-lowering medications are prescribed [*see CE:III Reducing Risk of Injury and Diseases and CE:IV Diet and Exercise*].

Anticoagulation Therapy

Several prospective, randomized trials revealed that warfarin therapy reduces mortality after discharge from the hospital in patients with acute myocardial infarction. However, in these studies, in which warfarin therapy was compared with placebo, aspirin was not administered in either arm of the study.[139,140] The Coumadin Aspirin Reinfarction Study (CARS) revealed that the risk of reinfarction in patients treated with aspirin alone was similar to that in patients treated with aspirin and either low-dose (1 mg) or higher-dose (3 mg) warfarin.[141] Warfarin is also ineffective at preventing coronary reocclusion in patients in whom thrombolytic therapy was successful.[142] The routine administration of warfarin is not currently recommended to prevent reinfarction in patients who have survived myocardial infarction.

Antiarrhythmic Therapy

Although Holter monitoring before discharge can help identify patients at increased risk for sudden cardiac death, antiarrhythmic therapy has not been shown to decrease the risk of death in such patients, and in fact, it increased mortality in the Cardiac Arrhythmia Suppression Trial (CAST).[143] Since CAST, several prospective, randomized studies have been performed that have examined the role of amiodarone in patients at increased risk for sudden death. Taken together, the results of those studies do not indicate that amiodarone reduces mortality. Further studies are needed before the routine use of amiodarone can be recommended in high-risk patients, such as those included in these trials.

Automated, Implantable Cardioverter-Defibrillator

Automated implantable cardioverter-defibrillators are of proven benefit in patients with coronary artery disease, reduced left ventricular ejection fraction, nonsustained ventricular tachycardia, and inducible ventricular tachycardia.[144] It has been proposed that patients with a prior myocardial infarction and advanced left ventricular dysfunction may benefit from prophylactic implantation of a defibrillator (in the absence of electrophysiologic testing to induce arrhythmias). The Multicenter Automatic Defibrillator Implantation Trial (MADIT)–II investigators randomized 1,232 patients with a prior myocardial infarction and a left ventricular ejection fraction of 0.30 or less to an implantable defibrillator or conventional medical therapy. During an average follow-up of 20 months, the mortality rates were 19.8% in the conventional-therapy group and 14.2% in the defibrillator group (P = 0.016). Prophylactic implantation of a defibrillator is a recommended therapy in this patient population.[145]

RISK-FACTOR MODIFICATION

An important and often neglected aspect of medical care after a myocardial infarction is the identification and modification of risk factors for atherosclerosis. Hypertension and hypercholesterolemia should be treated. Cessation of smoking [*see CE:III Reducing Risk of Injury and Disease*] has been shown to prolong life in patients who have survived a myocardial infarction; behavior modification and group therapy can increase the likelihood of kicking the habit. Cardiac rehabilitation and the establishment of a healthier lifestyle with an exercise program[146] [*see CE:IV Diet and Exercise*] can further reduce the likelihood of a return to smoking. Hypercholesterolemia should be aggressively treated as described above.

Although there are few data that conclusively indicate that patients who participate in a cardiac rehabilitation program after discharge have increased survival, an exercise rehabilitation program appears to improve a patient's sense of well-being and hasten return to work and leisure activities. A cardiac rehabilitation program can also help improve diet and aid weight reduction in overweight patients, help smokers refrain from smoking, and help establish an exercise program that the patient can maintain long after the formal rehabilitation program has ended. In summary, participation in a cardiac rehabilitation program often leads to the establishment of a healthier lifestyle.

Long-term Prognosis

Long-term prognosis after myocardial infarction is determined primarily by the severity of left ventricular dysfunction, the presence and degree of residual ischemia, and the potential for malignant ventricular arrhythmia. These adverse prognostic factors are related to each other but are also independently associated with death after discharge. Age is also an important determinant of outcome. Most deaths that occur in the first year after discharge occur in the first 3 months, a fact that stresses the importance of assessing risk and optimizing therapy before discharge from the hospital. However, there can be substantial improvement in ventricular function in the weeks and months after acute myocardial infarction, particularly in patients in whom early reperfusion was achieved. Therefore, measurement of ventricular function 2 to 3 months after myocardial infarction is a more accurate predictor of long-term prognosis than measurement of left ventricular function in the acute stages.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Rogers WJ, Canto JG, Lambrew CT, et al: Temporal trends in the treatment of over 1.5 million patients with myocardial infarction in the U.S. from 1990 through 1999: the National Registry of Myocardial Infarction 1, 2 and 3. J Am Coll Cardiol 36:2056, 2000

2. Eagle KA, Goodman SG, Avezum A, et al: Practice variation and missed opportunities for reperfusion in ST-segment-elevation myocardial infarction: Findings from the Global Registry of Acute Coronary Events (GRACE). Lancet 359:373, 2002

3. Mortality from coronary heart disease and acute myocardial infarction—United States, 1998. MMWR Morb Mortal Wkly Rep 50:90, 2001

4. Patient/bystander recognition and action: rapid identification and treatment of acute myocardial infarction (NIH Publication No 93-3303). National Heart Attack Alert Program (NHAAP). National Heart, Lung, and Blood Institute. National Institutes of Health, Bethesda, Maryland, 1993

5. Fuster V: Lewis A. Conner Memorial Lecture. Mechanisms leading to myocardial infarction: insights from studies of vascular biology. Circulation 90:2126, 1994

6. Libby P: Molecular bases of the acute coronary syndromes. Circulation 91:2844, 1995

7. Libby P: Current concepts of the pathogenesis of the acute coronary syndromes. Circulation 104:365, 2001

8. Ambrose JA, Tannenbaum MA, Alexopoulos D, et al: Angiographic progression of coronary artery disease and the development of myocardial infarction. J Am Coll Cardiol 12:56, 1988

9. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, et al: Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Circulation 90:583, 1994

10. Myocardial infarction redefined: a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the Redefinition of Myocardial Infarction. J Am Coll Cardiol 36:959, 2000

11. Aronow WS: Prevalence of presenting symptoms of recognized acute myocardial infarction and of unrecognized healed myocardial infarction in elderly patients. Am J Cardiol 60:1182, 1987

12. Kouvaras G, Bacoulas G: Unstable angina as a warning symptom before acute myocardial infarction. Q J Med 64:679, 1987

13. Kannel WB, Abbott RD: Incidence and prognosis of unrecognized myocardial infarction: an update on the Framingham study. N Engl J Med 311:1144, 1984

14. Karlson BW, Herlitz J, Edvardsson N, et al: Eligibility for intravenous thrombolysis in suspected acute myocardial infarction. Circulation 82:1140, 1990

15. Grijseels EW, Deckers JW, Hoes AW, et al: Prehospital triage of patients with suspected acute myocardial infarction. Evaluation of previously developed algorithms and new proposals. Eur Heart J 16:325, 1995

16. Cragg DR, Friedman HZ, Bonema JD, et al: Outcome of patients with acute myocardial infarction who are ineligible for thrombolytic therapy. Ann Intern Med 115:173, 1991

17. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Second International Study of Infarct Survival Collaborative Group. Lancet 2:349, 1988

18. Roberts R, Kleinman N: Earlier diagnosis and treatment of acute myocardial infarction necessitates the need for a "new diagnostic mind-set." Circulation 89:872, 1994

19. Zabel M, Hohnloser SH, Koster W, et al: Analysis of creatine kinase, CK-MB, myoglobin, and troponin T time-activity curves for early assessment of coronary artery reperfusion after intravenous thrombolysis. Circulation 87:1542, 1993

20. Ohman EM, Armstrong P, Califf RM, et al: Risk stratification in acute ischemic syndromes using serum troponin T. The GUSTO Investigators. J Am Coll Cardiol 25(suppl):148A, 1995

21. Ravkilde J, Nissen H, Horder M, et al: Independent prognostic value of serum creatine kinase isoenzyme MB mass, cardiac troponin T and myosin light chain levels in suspected acute myocardial infarction: analysis of 28 months of follow-up in 196 patients. J Am Coll Cardiol 25:574, 1995

22. Hauser AM, Gangadharan V, Ramos RG, et al: Sequence of mechanical, electrocardiographic and clinical effects of repeated coronary arterial occlusion in human beings: echocardiographic observation during coronary angioplasty. J Am Coll Cardiol 5:193, 1985

23. Sabia P, Abbott RD, Afrookteh A, et al: Importance of two-dimensional echocardiographic assessment of left ventricular systolic function in patients presenting to the emergency room with cardiac-related symptoms. Circulation 84:1615, 1991

24. Fleischmann KE, Lee TH, Come PC, et al: Echocardiographic prediction of complications in patients with chest pain. Am J Cardiol 79:292, 1997

25. Talreja D, Gruver C, Sklenar J, et al: Efficient utilization of echocardiography for the assessment of left ventricular systolic function. Am Heart J 139:394, 2000

26. St John Sutton M, Pfeffer MA, Plappert T, et al: Quantitative two-dimensional echocardiographic measurements are major predictors of adverse cardiovascular events after acute myocardial infarction: the protective effects of captopril. Circulation 89:68, 1994

27. Hilton TC, Thompson RC, Williams HJ, et al: Technetium-99m sestamibi myocardial perfusion imaging in the emergency room evaluation of chest pain. J Am Coll Cardiol 23:1016, 1994

28. Duca MD, Giri S, Wu AH, et al: Comparison of acute rest myocardial perfusion imaging and serum markers of myocardial injury in patients with chest pain syndromes. J Nucl Cardiol 6:570, 1999

29. Radensky PW, Hilton TC, Fulmer H, et al: Potential cost effectiveness of initial myocardial perfusion imaging for assessment of emergency department patients with chest pain. Am J Cardiol 79:595, 1997

30. Udelson JE, Beshansky JR, Ballin DS, et al: Myocardial perfusion imaging for evaluation and triage of patients with suspected acute cardiac ischemia. JAMA 288:2693, 2002

31. Rapid identification and treatment of acute myocardial infarction. National Heart, Lung, and Blood Institute (NIH Publication No 94-3302). National Institutes of Health, Bethesda, Maryland, 1994

32. Secondary prevention of vascular disease by prolonged antiplatelet treatment. Antiplatelet Trialists Collaboration. Br Med J 296:320, 1988

33. Weaver WD: Time to thrombolytic treatment: factors affecting delay and their influence on outcome. J Am Coll Cardiol 25(suppl 7):3S, 1995

34. Emergency department rapid identification and treatment of patients with acute myocardial infarction. National Heart Attack Alert Program Coordinating Committee 60 minutes to Treatment Working Group. Ann Emerg Med 23:311, 1994

35. Berger PB, Ellis SG, Holmes DR Jr, et al: The relationship between delay in performing direct coronary angioplasty and early clinical outcome in patients with acute myocardial infarction: results from the global use of strategies to open occluded arteries in acute coronary syndromes (GUSTO-IIb) trial. Circulation 100:14, 1999

36. Cannon CP, Gibson CM, Lambrew CT, et al: Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. JAMA 283:2941, 2000

37. Caputo RP, Ho KKL, Stoler RC, et al: Effect of continuous quality improvement on the delivery of primary percutaneous transluminal coronary angioplasty for acute myocardial infarction. Am J Cardiol 79:1159, 1997

38. Jhangiana AH, Jorgenson MB, Kotlewski A, et al: Community practice of primary angioplasty for myocardial infarction. Am J Cardiol 80:209, 1997

39. Patel S, Reese C, O'Connor RE, et al: Adverse outcomes accompanying primary angioplasty (PTCA) for acute myocardial infarction (AMI)—dangers of delay. J Am Coll Cardiol 27(suppl A):62A, 1996

40. Rosman HS, Ciolino D, Nerenz D, et al: Primary PTCA and thrombolysis for acute myocardial infarction: observations from the community. J Am Coll Cardiol 27(suppl A):26A, 1996

41. Widimsky P, Budesinsky T, Vorac D, et al: Long distance transport for primary angioplasty vs immediate thrombolysis in acute myocardial infarction. Final results of the randomized national multicentre trial—PRAGUE-2. Eur Heart J 24:94, 2003

42. Andersen HR, Nielsen TT, Rasmussen K, et al: A comparison of coronary angioplasty with fibrinolytic therapy in acute myocardial infarction. N Engl J Med 349:733, 2003

43. Immediate versus delayed catheterization and angioplasty following thrombolytic therapy for acute myocardial infarction. The TIMI Study Group. JAMA 260:2849, 1988

44. SWIFT trial of delayed elective intervention v conservative treatment after thrombolysis with anistreplase in acute myocardial infarction. SWIFT Study Group. BMJ 302:555, 1991

45. Ellis SG, Mooney MR, George BS, et al: Randomized trial of late elective angioplasty versus conservative management for patients with residual stenoses after thrombolytic treatment of myocardial infarction. Treatment of Post-Thrombolytic Stenoses (TOPS) Study Group. Circulation 86:1400, 1992

46. Zeymer U: The ALKK Trial. Oral presentation at the Late Breaking Clinical Trials Session of the 49th Annual Scientific Sessions of the American College of Cardiology, March 2000

47. Berger PB, Holmes DR Jr, Stebbins AL, et al: Impact of an aggressive invasive catheterization and revascularization strategy on mortality in patients with cardiogenic shock in the Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO-I) trial: an observational study. Circulation 96:122, 1997

48. Prasad A, Mathew V, Holmes DR, et al: Current management of non-ST-segment-elevation acute coronary syndrome: reconciling the results of randomized controlled trials. Eur Heart J 24:1544, 2003

49. Effects of tissue plasminogen activator and a comparison of early invasive and conservative strategies in unstable angina and non-Q-wave myocardial infarction: results of the TIMI IIIB Trial. The TIMI IIIB Investigators. Circulation 89:1545, 1995

50. Wallentin L, Lagerqvist B, Husted S, et al: Outcome at 1 year after an invasive compared with a non-invasive strategy in unstable coronary-artery disease: the FRISC II invasive randomised trial. Lancet 356:9, 2000

51. Cannon CP, Weintraub WS, Demopolous LA, et al: Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. N Engl J Med 344:1879, 2001

52. Braunwald E, Antman EM, Beasley JW, et al: ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction – summary article: A report of the American College of Cardiology/American Heart Association task force on practice guidelines. J Am Coll Cardiol 40:1366, 2002

53. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. Fibrinolytic Therapy Trialists' (FTT) Collaborative Group. Lancet 343:311, 1994

54. French JK, Hyde TA, Patel H, et al: Survival 12 years after randomization to streptokinase: the influence of thrombolysis in myocardial infarction flow at three to four weeks. J Am Coll Cardiol 34:62, 1999

55. The effects of tissue plasminogen activator, streptokinase, or both on coronary-artery patency, ventricular function, and survival after acute myocardial infarction. The GUSTO Angiographic Investigators. N Engl J Med 329:1615, 1993

56. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO Investigators. N Engl J Med 329:673, 1993

57. Simes RJ, Topol EJ, Holmes DR Jr, et al: Link between the angiographic substudy and mortality outcomes in a large randomized trial of myocardial reperfusion: importance of early and complete infarct artery reperfusion. Circulation 91:1923, 1995

58. Late Assessment of Thrombolytic Efficacy (LATE) study with alteplase 6 to 24 hours after onset of acute myocardial infarction. Lancet 342:759, 1993

59. de Jaegere PP, Arnold AA, Balk AH, et al: Intracranial hemorrhage in association with thrombolytic therapy: incidence and clinical predictive factors. J Am Coll Cardiol 19:289, 1992

60. Randomised double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate the equivalent. International Joint Efficacy Comparison of Thrombolytics. Lancet 346:329, 1995

61. A comparison of reteplase with alteplase for acute myocardial infarction. The Global Use of Strategies to Open Occluded Coronary Arteries (GUSTO III) Investigators. N Engl J Med 337:1118, 1997

62. Antman EM, Giugliano RD, Gibson CM, et al: Abciximab facilitates the rate and extent of thrombolysis: results of the Thrombolysis in Myocardial Infarction (TIMI 14) Trial. Circulation 99:2720, 1999

63. Ohman EM, Lincoff AM, Bode C, et al: Enhanced early reperfusion at 60 minutes with low-dose reteplase combined with full-dose abciximab in acute myocardial infarction: preliminary results from the GUSTO-IV Pilot (SPEED) Dose-ranging Trial (abstr). Circulation 98(suppl I):I-504, 1998

64. Topol EJ: Reperfusion therapy for acute myocardial infarction with fibrinolytic therapy or combination reduced fibrinolytic therapy and platelet glycoprotein IIb/IIIa inhibition: the GUSTO V randomised trial. Lancet 357:1905, 2001

65. Assessment of the Safety and Efficacy of a New Thrombolytic regimen (ASSENT-3) Investigators. Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: The ASSENT-3 randomized trial of acute myocardial infarction. Lancet 358:605, 2001

66. Wallentin L, Goldstein P, Armstrong PW, et al: Efficacy and safety of tenecteplase in combination with the low-molecular-weight heparin enoxaparin or unfractionated heparin in the prehospital setting: the Assessment of the Safety and Efficacy of a New

Thrombolytic Regimen (ASSENT)-3 PLUS randomized trial in acute myocardial infarction. Circulation 108:135, 2003

67. O'Neill W, Timmis G, Bourdillon P, et al: A prospective randomized clinical trial of intracoronary streptokinase versus coronary angioplasty for acute myocardial infarction. N Engl J Med 314:812, 1986

68. de Boer MJ, Hoorntje JCA, Ottervanger JP, et al: Immediate coronary angioplasty versus intravenous streptokinase in acute myocardial infarction: left ventricular ejection fraction, hospital mortality, and reinfarction. J Am Coll Cardiol 23:1004, 1994

69. de Boer MJ, Suryapranata H, Hoorntje JC, et al: Limitation of infarct size and preservation of left ventricular function after primary coronary angioplasty compared with intravenous streptokinase in acute myocardial infarction. Circulation 90:753, 1994

70. Zijlstra F, de Boer MJ, Hoorntje JCA, et al: A comparison of immediate coronary angioplasty with intravenous streptokinase in acute myocardial infarction. N Engl J Med 328:680, 1993

71. Grines CL, Browne KF, Marco J, et al: A comparison of immediate angioplasty with thrombolytic therapy for acute myocardial infarction. N Engl J Med 328:673, 1993

72. Keeley EC, Boura JA, Grines CL: Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomized trials. Lancet 361:13, 2003

73. A clinical trial comparing primary coronary angioplasty with tissue plasminogen activator for acute myocardial infarction. The Global Use of Strategies to Open Occluded Coronary Arteries in Acute Coronary Syndromes (GUSTO IIb) Angioplasty Substudy Investigators. N Engl J Med 336:1621, 1997

74. Berger PB, Bell MR, Holmes DR Jr, et al: Time to reperfusion with direct coronary angioplasty and thrombolytic therapy in acute myocardial infarction. Am J Cardiol 73:231, 1994

75. Ryan TJ, Anderson JL, Antman EM, et al: ACC/AHA guidelines for the management of patients with acute myocardial infarction: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (Committee of Management of Acute Myocardial Infarction). J Am Coll Cardiol 28:1328, 1996

76. Wharton TP, McNamara SN, Fedele FA, et al: Primary angioplasty for the treatment of acute myocardial infarction: experience at two community hospitals without cardiac surgery. J Am Coll Cardiol 33:1257, 1999

77. Brener SJ, Barr LA, Burchenal JE, et al: Randomized, placebo-controlled trial of platelet glycoprotein IIb/IIIa blockade with primary angioplasty for acute myocardial infarction. ReoPro and Primary PTCA Organization and Randomized Trial (RAPPORT) Investigators. Circulation 98:73, 1998

78. Neumann FJ, Blasini R, Schmitt C, et al: Effect of glycoprotein IIb/IIIa receptor blockade on recovery of coronary flow and left ventricular function after placement of coronary-artery stents in acute myocardial infarction. Circulation 98:2695, 1998

79. Montalescot G, Barragan P, Wittenberg O, et al: Platelet glycoprotein IIb/IIIa inhibition with coronary stenting for acute myocardial infarction. N Engl J Med 344:1895, 2001

80. Stone GW, Grines CL, Cox DA, et al: Comparison of angioplasty with stenting, with or without abciximab, in acute myocardial infarction. N Engl J Med 346:957, 2002

81. Grines CL, Cox DA, Stone GW, et al: Coronary angioplasty with or without stent implantation for acute myocardial infarction. Stent Primary Angioplasty in Myocardial Infarction Study Group. N Engl J Med 341:1949, 1999

82. Schomig A, Kastrati A, Dirschinger J, et al: Coronary stenting plus platelet glycoprotein IIb/IIIa blockade compared with tissue plasminogen activator in acute myocardial infarction. Stent versus Thrombolysis for Occluded Coronary Arteries in Patients with Acute Myocardial Infarction Study Investigators. N Engl J Med 343:385, 2000

83. Berger PB, Stensrud PE, Daly RC, et al: Emergency coronary artery bypass surgery following failed coronary angioplasty: time to reperfusion and other procedural characteristics. Am J Cardiol 76:565, 1995

84. Ellis SG, Ribeiro da Silva E, Heyndrickx G, et al: Randomized comparison of rescue angioplasty with conservative management of patients with early failure of thrombolysis for acute anterior myocardial infarction. Circulation 90:2280, 1994

85. Belenkie I, Traboulsi M, Hall CA, et al: Rescue angioplasty during myocardial infarction has a beneficial effect on mortality: a tenable hypothesis. Can J Cardiol 8:357, 1992

86. Widimsky P, Groh L, Ascherman M, et al: The "PRAGUE" study: a national multicenter randomized study comparing primary angioplasty vs thrombolysis vs both in patients with acute myocardial infarction admitted to the community hospitals: results of the pilot phase (abstr). Eur Heart J 19(suppl):56, 1998

87. Vermeer F, Brunninkhuis L, van de Berg E, et al: Prospective randomized comparison between thrombolysis, rescue PTCA, and primary PTCA in patients with extensive myocardial infarction admitted to a hospital without PTCA facilities: a safety and feasibility study. Heart 82:426, 1999

88. Simoons ML, Arnold AE, Betriu A, et al: Thrombolysis with tissue plasminogen activator in acute myocardial infarction: no additional benefit from immediate percutaneous coronary angioplasty. Lancet 1:197, 1988

89. Dirschinger J, Kastrati A, Neumann FJ, et al: Influence of balloon pressure during stent placement in native coronary arteries on early and late angiographic and clinical outcome: a randomized evaluation of high-pressure inflation. Circulation 100:918, 1999

90. Miller JM, Smalling R, Ohman EM, et al: Effectiveness of early coronary angioplasty and abciximab for failed thrombolysis (reteplase or alteplase) during acute myocardial infarction: results from the GUSTO-III trial. Am J Cardiol 84:779, 1989

91. Ross AM, Coyne KS, Reiner JS, et al: A randomized trial comparing primary angioplasty with a strategy of short-acting thrombolysis and immediate planned rescue angioplasty in acute myocardial infarction: the PACT trial. PACT investigators. Plasminogen-activator Angioplasty Compatibility Trial. J Am Coll Cardiol 34:1954, 1999

92. Scheller B, Hennen B, Hammer B, et al: Beneficial effects of immediate stenting after thrombolysis in acute myocardial infarction. J Am Coll Cardiol 42:634, 2003

93. Kastrati A, Mehilli J, Schlotterbeck K, et al: Early administration of reteplase plus abciximab versus abciximab alone in patients with acute myocardial infarction referred for percutaneous coronary intervention. JAMA 291:947, 2004

94. Granger CB, Hirsh J, Califf RM, et al: Activated partial thromboplastin time and outcome after thrombolytic therapy for acute myocardial infarction: results from the GUSTO-I trial. Circulation 93:870, 1996

95. Chesebro JH, Fuster V: Antithrombotic therapy for acute myocardial infarction: mechanisms and prevention of deep venous, left ventricular, and coronary artery thromboembolism. Circulation 74:1, 1986

96. Granger CB, Miller JM, Bovill EG, et al: Rebound increase in thrombin generation and activity after cessation of intravenous heparin in patients with acute coronary syndromes. Circulation 91:1929, 1995

97. Ross AM, Molhoek P, Lundergan C, et al: Randomized comparison of enoxaparin, a low-molecular-weight heparin, with unfractionated heparin adjunctive to recombinant tissue plasminogen activator thrombolysis and aspirin: second trial of Heparin and Aspirin Reperfusion Therapy (HART II). Circulation 104:648, 2001

98. Simoons M, Krzeminska-Pakula M, Alonso A, et al: Improved reperfusion and clinical outcome with enoxaparin as an adjunct to streptokinase thrombolysis in acute myocardial infarction. The AMI-SK study. Eur Heart J 23:1282, 2002

99. Assessment of the Safety and Efficacy of a New Thrombolytic Regimen (ASSENT)-3 Investigators. Efficacy and safety of tenecteplase in combination with enoxaparin, abciximab, or unfractionated heparin: the ASSENT-3 randomised trial in acute myocardial infarction. Lancet 358:605, 2001

100. Direct thrombin inhibitors in acute coronary syndromes: principal results of a meta-analysis based on individual patients' data. Direct Thrombin Inhibitor Trialists' Collaborative Group. Lancet 359:294, 2002

101. Thrombin-specific anticoagulation with bivalirudin versus heparin in patients receiving fibrinolytic therapy for acute myocardial infarction: the HERO-2 randomised trial. Hirulog and Early Reperfusion or Occlusion (HERO)-2 Trial Investigators. Lancet 358:1855, 2001

102. Gersh BJ, Rahimtoola SH: Acute myocardial infarction. Pharmacological Management of Acute Myocardial Infarction. Warnica JW, Ed. Elsevier, New York, 1991, p 205

103. Comparison of invasive and conservative strategies after treatment with intravenous tissue plasminogen activator in acute myocardial infarction: results of the Thrombolysis in Myocardial Infarction (TIMI) phase II trial. The TIMI Study Group. N Engl J Med 320:618, 1989

104. Barron HV, Rundle AC, Gore JM, et al: Intracranial hemorrhage rates and effect of immediate beta-blocker use in patients with acute myocardial infarction treated with tissue plasminogen activator. Participants in the National Registry of Myocardial Infarction–2. Am J Cardiol 85:294, 2000

105. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. Fourth International Study of Infarct Survival Collaborative Group. Lancet 345:669, 1995

106. Sigurdsson A, Swedberg K: Left ventricular remodelling, neurohormonal activation and early treatment with enalapril (CONSENSUS II) following myocardial infarction. Eur Heart J 15(suppl B):14, 1994

107. Yusuf S, Sleight P, Pogue J, et al: Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000

108. Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). European Trial on Reduction of Cardiac Events with Perindopril in Stable Coronary Artery Disease Investigators. Lancet 362:782, 2003

109. Judgutt BI, Warnica JW: Intravenous nitroglycerin therapy to limit myocardial infarct size, expansion, and complications: effect of timing, dosage, and infarct location. Circulation 78:906, 1988

110. GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto Miocardico. Lancet 343:1115, 1994

111. MacMahon S, Collins R, Peto R, et al: Effects of prophylactic lidocaine in suspected acute myocardial infarction: an overview of results from the randomized, controlled trials. JAMA 260:1910, 1988

112. Teo KK, Yusuf S, Furberg CD: Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction: an overview of results from randomized controlled trials. JAMA 270:1589, 1993

113. Gheorghiade M: Calcium channel blockers in the management of myocardial infarction patients. Henry Ford Hosp Med J 39:210, 1991

114. Held PH, Yusuf S: Effects of beta-blockers and calcium channel blockers in acute myocardial infarction. Eur Heart J 14(suppl F):18, 1993

115. Secondary Prevention Reinfarction Israeli Nifedipine Trial (SPRINT): A randomized intervention trial of nifedipine in patients with acute myocardial infarction. The Israeli SPRINT Study Group. Eur Heart J 9:354, 1988

116. Goldbourt U, Behar S, Reicher-Reiss H, et al: Early administration of nifedipine in suspected acute myocardial infarction. The Secondary Prevention Reinfarction Israel Nifedipine Trial 2 Study. Arch Intern Med 153:345, 1993

117. The effect of diltiazem on mortality and reinfarction after myocardial infarction. The Multicenter Diltiazem Postinfarction Trial Research Group. N Engl J Med 319:385, 1988

118. Gibson RS, Hansen JF, Messerli F, et al: Long-term effects of diltiazem and verapamil on mortality and cardiac events in non-Q-wave acute myocardial infarction without pulmonary congestion: post hoc subset analysis of the multicenter diltiazem postinfarction trial and the second Danish verapamil infarction trial studies. Am J Cardiol 86:275, 2000

119. Boden WE, van Gilst WH, Scheldewaert RG, et al: Diltiazem in acute myocardial infarction treated with thrombolytic agents: a randomised placebo-controlled trial. Incomplete Infarction Trial of European Research Collaborators Evaluating Prognosis post-Thrombolysis (INTERCEPT). Lancet 355:1751, 2000

120. Teo KK, Yusuf S, Collins R, et al: Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. BMJ 303:1499, 1991

121. Woods KL, Fletcher S, Roffe C, et al: Intravenous magnesium sulphate in suspected acute myocardial infarction: results of the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2). Lancet 339:1553, 1992

122. Schechter M, Hod H, Chouraqui P, et al: Magnesium therapy in acute myocardial infarction when patients are not candidates for thrombolytic therapy. Am J Cardiol 75:321, 1995

123. Early administration of intravenous magnesium to high-risk patients with acute myocardial infarction in the Magnesium in Coronaries (MAGIC) trial: a randomized controlled trial. Magnesium in Coronaries (MAGIC) Trial Investigators. Lancet 360:1189, 2002

124. Bunch TJ, White RD, Gersh BJ, et al: Long-term outcomes of out-of-hospital cardiac arrest after successful early defibrillation. N Engl J Med 348:2626, 2003

125. Berger PB, Ruocco NA, Ryan TJ, et al: Incidence and prognostic significance of ventricular tachycardia and ventricular fibrillation in the absence of hypotension or heart failure in acute myocardial infarction treated with recombinant tissue-type plasminogen activator: results from the TIMI II trial. J Am Coll Cardiol 22:1773, 1993

126. Behar S, Goldbourt U, Reicher-Reiss H, et al: Prognosis of acute myocardial infarction complicated by primary ventricular fibrillation. Am J Cardiol 66:1208, 1990

127. Wyse DG, Waldo AL, DiMarco JP, et al: A comparison of rate control and rhythm control in patients with atrial fibrillation. N Engl J Med 347:1825, 2002

128. Berger PB, Ryan TJ: Inferior infarction: high risk subgroups. Circulation 81:401, 1990

129. Berger PB, Ruocco NA, Frederick MM, et al: The incidence and significance of heart block during inferior infarction: results from the TIMI II trial. J Am Coll Cardiol 20:533, 1992

130. Oliva PB, Hammill SC, Edwards WD: Cardiac rupture, a clinically predictable complication of acute myocardial infarction: report of 70 cases with clinicopathologic correlations. J Am Coll Cardiol 22:720, 1993

131. Zehender M, Kasper W, Kauder E, et al: Right ventricular infarction as an independent predictor of prognosis after acute inferior myocardial infarction. N Engl J Med 328:981, 1993

132. Flapan AD: Management of patients after their first myocardial infarction. BMJ 309:1129, 1994

133. Madsen JK, Grande P, Saunamaki K, et al: Danish Multicenter Randomized Study of Invasive Versus Conservative Treatment in Patients with Inducible Ischemia after Thrombolysis in Acute Myocardial Infarction (DANAMI). Danish Trial in Acute Myocardial Infarction. Circulation 96:748, 1997

134. Shaw LJ, Peterson ED, Kesler K, et al: A metaanalysis of predischarge risk stratification after acute myocardial infarction with stress electrocardiographic, myocardial perfusion, and ventricular function imaging. Am J Cardiol 78:1327, 1996

135. Chaitman BR, McMahon RP, Terin M, et al: Impact of treatment strategy on predischarge exercise test in the Thrombolysis in Myocardial Infarction (TIMI) II trial. Am J Cardiol 71:131, 1993

136. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Scandinavian Simvastatin Survival Study Group. Lancet 344:1383, 1994

137. Sacks FM, Pfeffer MA, Moye LA, et al: The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. N Engl J Med 335:1001, 1996

138. Stenestrand U, Wallentin L: Early statin treatment following acute myocardial infarction and 1-year survival. JAMA 285:430, 2001

139. Effect of long-term oral anticoagulant treatment on mortality and cardiovascular morbidity after myocardial infarction. Anticoagulants in the Secondary Prevention of Events in Coronary Thrombosis (ASPECT) Research Group. Lancet 343:499, 1994

140. Smith P, Arnesen H, Holme I: The effect of warfarin on mortality and reinfarction after myocardial infarction. N Engl J Med 323:147, 1990

141. Randomised double-blind trial of fixed, low-dose warfarin with aspirin after myocardial infarction. Coumadin Aspirin Reinfarction Study (CARS) Investigators. Lancet 350:389, 1997

142. Meijer A, Verheugt FWA, Wester CPJP, et al: Aspirin versus coumadin in the prevention of reocclusion and recurrent ischemia after successful thrombolysis: a prospective placebo-controlled angiographic study: results of the APRICOT study. Circulation 87:1524, 1993

143. Epstein AE, Hallstrom AP, Rogers WL, et al: Mortality following ventricular arrhythmia suppression by encainide, flecainide, and moricizine after myocardial infarction: the original design concept of the Cardiac Arrhythmia Suppression Trial (CAST). JAMA 270:2451, 1993

144. Moss AJ, Hall WJ, Cannom DS, et al: Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter Automatic Defibrillator Implantation Trial Investigators. N Engl J Med 335:1933, 1996

145. Moss AJ, Zareba W, Jackson Hall W, et al: Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. N Engl J Med 346:877, 2002

146. Rana JS, Mukamal KJ, Morgan JP, et al: Obesity and the risk of death after acute myocardial infarction. Am Heart J 147:841, 2004

Acknowledgment

Figures 1 through 7, 8, and 11 Marcia Kammerer.

IX CHRONIC STABLE ANGINA

Paul R. Sutton, m.d., ph.d., f.a.c.p.
Stephan D. Fihn, m.d., m.p.h., f.a.c.p.

Definitions

Angina pectoris is the cardinal symptom of myocardial ischemia. Ischemia occurs when the coronary blood supply is inadequate to meet the metabolic demands of the myocardium. This mismatch of coronary blood supply and myocardial metabolic demand usually results from narrowing or occlusion of one or more coronary arteries, but in rare cases, it may be caused by coronary vasospasm or solely by excessive myocardial oxygen demand in the absence of significant coronary atherosclerosis.

Angina is typically a substernal, pressurelike discomfort or pain, but it may also take the form of discomfort in the jaw, shoulder, back, or arm. Usually, it is precipitated by physical exertion or emotional stress, and it is promptly relieved by rest or by taking nitroglycerin. Chronic stable angina refers to a pattern of chest pain or discomfort that does not change appreciably in frequency or severity over 2 months or longer and in which the episodes of pain are provoked by exertions or stresses of similar intensity.[1] Unstable angina, by contrast, is defined as rest angina, severe angina of new onset, or an increase in the severity or frequency of previously stable angina. Certain patients with symptoms of unstable angina are at an increased risk for myocardial infarction (MI) or death [*see* Epidemiology, *below*]. Chronic stable angina precedes MI in about half of cases and is common afterward. Although angina is a cardinal symptom of ischemic heart disease (IHD), MI or sudden death is the initial presentation of IHD in as many as half of patients.

Epidemiology

IHD is the leading cause of mortality in the United States and the rest of the developed world; it is responsible for more than 20% of deaths.[2] In the United States, approximately one million persons suffer an MI, and 500,000 coronary deaths occur each year. IHD is the leading cause of death in the United States for both sexes in both white and black populations.

The prevalence of IHD increases with age and is higher in men than in women in every age group. The American Heart Association (AHA) conservatively estimates that more than six million persons in the United States experience angina.[3]

In addition to posing an increased risk of MI and premature death, chronic stable angina often limits affected persons' capacity for work and other activities, which, in turn, negatively affects their quality of life. The direct and indirect costs of hospitalization, diagnostic procedures, and revascularization related to angina are substantial. Estimates of direct hospital costs for Medicare patients with a history of chronic stable angina exceed $7 billion annually.[1] Of patients with angina who undergo a coronary revascularization procedure, 30% or more never return to work.[4]

The major modifiable risk factors for IHD are dyslipidemias—in particular, elevated levels of low-density lipoprotein (LDL) cholesterol and low levels of high-density lipoprotein (HDL) cholesterol—as well as hypertension, diabetes mellitus, and cigarette smoking.[5,6] Other important, but immutable, risk factors are increasing age, a family history of premature coronary disease, and male sex. Obesity, physical inactivity, and atherogenic dietary habits also contribute to cardiovascular risk, although it is difficult to distinguish the risks conferred by these risk factors independently of the risks conferred by the major cardiovascular risk factors because of the potential interaction of these factors. Patients with combinations of risk factors may be at particular risk for developing IHD. Patients with the metabolic syndrome, which consists of obesity (particularly abdominal adiposity), hypertension, dyslipidemia (i.e., elevated triglyceride levels and low HDL levels), and insulin resistance, are at particularly high risk for IHD.[7] It is estimated that in the United States, the metabolic syndrome affects nearly 25% of all persons and 43.5% of adults 60 years of age or older.[8]

Numerous clinical trials have identified important risk factors and effective therapies for coronary artery disease; however, few of these studies have included sufficient numbers of women to draw meaningful conclusions about coronary disease in women.[9] Thus, much of the evidence that supports contemporary recommendations for testing, prevention, and treatment of coronary disease in women is extrapolated from studies conducted predominantly in middle-aged men.

Although it has been proposed that 50% or more of patients with IHD lack any of the traditional major risk factors, two studies have challenged this notion; these studies indicate that the vast majority of patients who experience cardiac events (either fatal or nonfatal) have one or more major risk factors.[10,11] Nevertheless, interest remains in identifying additional laboratory markers of risk of IHD and, in particular, risk of acute coronary syndromes. In observational studies, several measures of inflammation, including C-reactive protein (CRP) levels, interleukin-6 (IL-6) levels, and levels of soluble cellular adhesion molecules, have been associated with risk of IHD and cardiovascular events.[12,13] Measures of fibrinogen, platelet activator inhibitor, and components of the coagulation/fibrinolysis cascade may ultimately be of use in predicting risk of cardiovascular events.[12] A number of genetic polymorphisms and candidate genes that may increase the risk of MI have been identified in specific populations. Currently, however, it is uncertain whether these or other laboratory measurements represent truly independent risk factors or whether they will prove useful in clinical practice.[14]

Pathophysiology and Pathogenesis

Angina occurs as a result of myocardial ischemia, which occurs when cardiac blood supply is insufficient to meet myocardial oxygen demand. Stable angina commonly occurs in the setting of narrowing or partial occlusion of segments of coronary arteries by atherosclerotic plaque. Significant occlusion is defined as a reduction of the diameter of a major coronary artery by 70%, which corresponds to a 50% reduction in vessel lumen surface area; such occlusion is often sufficient to cause angina. Coronary atherosclerosis and angina often progress over time, reflecting both gradual and more abrupt changes in luminal diameter of coronary vessels. Incremental changes in coronary atherosclerosis reflect the progression of existing lesions and the appearance of new stenoses. Abrupt changes in vessel diameter

may be associated with sudden changes in anginal symptoms, termed unstable angina. Unstable angina is commonly caused by rupture of vulnerable atherosclerotic plaque with associated platelet thrombosis. Unstable angina may also be caused by endothelial injury, thrombosis of severely stenotic coronary arteries, or coronary artery spasm. Although angina is usually associated with coronary artery disease (CAD), it may also occur in persons with normal or near-normal coronary arteries; in these settings, angina may occur as a result of increased myocardial oxygen demand associated with aortic stenosis, hyperthyroidism, or anemia.

Myocardial ischemia, regardless of cause, is associated with intracardiac release of adenosine. Adenosine release slows atrioventricular conduction and reduces contractility, which are adaptive changes in the setting of myocardial ischemia. Stimulation of adenosine receptors in the chest is believed to be responsible for the sensation of angina.

Angina occurs most often in the setting of coronary atherosclerosis. Atherosclerosis occurs as a result of vascular injury and subsequent responses to injury. Vascular injury may result from the mechanical stress of blood flow or from direct endothelial injury from toxins, such as those in cigarette smoke. Traditional cardiovascular risk factors, such as smoking, hypertension, hyperlipidemia, and diabetes, increase coronary risk, at least in part by potentiating vascular injury or altering the subsequent response to injury. Direct endothelial injury produces a sequence of events similar to chronic inflammation: the endothelium elaborates procoagulants, vasoactive molecules, growth factors, and cytokines. Platelets, inflammatory cells (e.g., monocytes and T cells), and smooth muscle cells are attracted to the site of injury.[15] This cascade of events is initially adaptive, resulting in repair of endothelial injury; however, repeated cycles of injury and repair can result in progressive atherosclerosis and luminal narrowing.

Support for the inflammatory hypothesis of atherosclerosis comes from clinical studies that suggest a correlation between the risk of future cardiovascular events and the presence of markers of inflammation, including CRP.[12,13] Inflammation may also play a role in acute coronary syndromes (i.e., unstable angina and acute MI). In some studies, levels of CRP during episodes of unstable angina are associated with a greater risk of MI or death,[16] although it is not clear whether inflammation (and corresponding elevations of CRP) contributes to the etiology of unstable coronary syndromes or is simply a consequence of myocardial ischemia and injury.[17]

Histologically, atherosclerotic plaques are composed of a fibrous cap derived from smooth muscle cells; the cap covers a core of oxidized lipids, inflammatory cells, and cellular debris. Immature plaques consist of a necrotic, lipid-rich core surrounded by a thin, fibrous capsule. These so-called vulnerable plaques may not be visible angiographically, but they are prone to disruption and thrombus formation and, thus, are associated with unstable coronary syndromes. In fact, many of the plaques responsible for myocardial infarction may not be associated with significant coronary stenosis.[18] Mature atherosclerotic plaques, by contrast, have less necrotic cores and thicker, more stable fibrous caps. These plaques tend to cause greater degrees of coronary stenosis and are the lesions typically identified by coronary angiography; they are generally less susceptible to fracture and are associated less with acute coronary syndromes than are immature plaques. Atherosclerotic lesions may be associated with chronic stable angina, either because of luminal narrowing or because of dysfunctional vascular reactivity.

Table 1 Grading of Angina Pectoris by the Canadian Cardiovascular Society Classification System[18]

Class I

Ordinary physical activity, such as walking and climbing stairs, does not cause angina. Angina occurs with strenuous, rapid, or prolonged exertion at work or recreation.

Class II

Slight limitation of ordinary activity. Angina occurs while walking or climbing stairs rapidly; while walking uphill; while walking or climbing stairs after meals in cold or in wind; while under emotional stress; or only during the first several hours after waking. Angina occurs while walking more than two blocks on level grade and climbing more than one flight of ordinary stairs at a normal pace and in normal condition.

Class III

Marked limitations of ordinary physical activity. Angina occurs while walking one or two blocks on level grade and climbing one flight of stairs in normal conditions and at a normal pace.

Class IV

Inability to engage in any physical activity without discomfort. Anginal symptoms may be present at rest

Clinical Manifestations

A cardinal manifestation of IHD, angina is characterized by substernal pain or discomfort that may radiate to the neck, jaw, epigastrium, back, or arms. Words characteristically used to describe the sensation of angina include "squeezing," "vicelike," "heavy," "griplike," and "suffocating." Angina ordinarily lasts only a few minutes. Typical angina has three key characteristics: (1) substernal chest discomfort of characteristic quality and duration that is (2) provoked by exertion or emotional stress and is (3) relieved by rest or nitroglycerin. Atypical angina has two of the three characteristics of typical angina; noncardiac chest pain has one or none of the characteristics of typical angina.

The severity of angina is graded according to the Canadian Cardiovascular Society (CCS) classification [*see Table 1*].[19] Angina grade provides a useful way to evaluate functional limitation, treatment efficacy, and stability of symptoms over time.

Anginal chest pain is further characterized as stable or unstable. Unstable angina presents as prolonged angina at rest; new-onset angina that is severe, prolonged, or frequent; and established angina that has become distinctly more frequent, longer in duration, or more easily provoked. Some patients with unstable angina are at increased risk for acute MI and death [*see Table 2*]; the pathophysiology of unstable angina in these patients is often the result of plaque rupture and thrombosis. Patients with unstable angina and intermediate-risk or high-risk clinical features are best evaluated in the hospital.[20]

Differential Diagnosis

Given the potentially life-threatening sequelae of angina and the availability of effective therapies, it is important to consider angina in all patients presenting with chest pain. One approach to chest pain is to consider the differential diagnosis anatomically [*see Table 3*]. Various diseases of the heart and pericardium cause chest pain. Arrhythmias and valvular heart disease cause typical angina; pericarditis often causes pleuritic pain (i.e., pain that worsens on inspiration), but it may produce angina that is

relieved by sitting up and leaning forward. Dissection of the great vessels can cause a characteristic, sudden, excruciating "tearing" pain in the chest or back. Diseases of the esophagus, such as esophageal spasm and acid reflux, may cause chest pain that is often postprandial or that occurs with recumbence. Esophageal spasm, in particular, can mimic angina and may respond to nitrates or calcium channel blockers. Diseases of lungs and pleura, including pulmonary embolism, pneumonia, pleuritis, and empyema, can cause chest pain that is often pleuritic. Chest wall syndromes, such as costochondritis, can cause substernal chest pain, typically reproduced with palpation. Herpes zoster may cause neuralgia that is localized to the chest; the pain may precede the appearance of the characteristic rash. Patients with panic disorder may describe chest pain or tightness accompanied by shortness of breath, diaphoresis, and other symptoms suggesting cardiac disease.

Diagnosis and Risk Stratification

PRELIMINARY EVALUATION

The evaluation of patients with chest pain should take into account symptom characteristics and cardiovascular risk factors, because these indicate the probability of angina and IHD [*see* Patient History and Its Use in Determining Risk for IHD, *below*]. If the history and physical examination suggest the presence of angina and IHD, patients are further evaluated by noninvasive tests, such as exercise treadmill testing or coronary angiography. Noninvasive testing serves to refine the probability of the diagnosis of IHD and stratify patients according to their risk for near-term cardiovascular events [*see* Noninvasive Testing, *below*].[21]

Patient History and Its Use in Determining Risk for IHD

Determining the pretest probability of significant IHD, which is defined as greater than 70% stenosis of one or more of the major epicardial coronary arteries, is an essential step in the evaluation of patients with suspected IHD. Decisions regarding testing and management are strongly influenced by estimates of the probability of significant IHD [*see Figure 1*].[21]

Estimates of the pretest probability of significant IHD can be accurately derived from a description of the chest pain syndrome and the presence or absence of cardiovascular risk factors.[22] It is important to characterize suspected angina by location, quality, duration, associated symptoms, and factors that exacerbate or relieve the pain [*see Table 4*]. Typical angina is substernal, lasts less than 5 minutes, is dull and aching, and is worse with exertion or emotional stress. Atypical angina has some but not all features of anginal chest pain. For example, a patient with aching substernal chest pain that lasts minutes but is unrelated to exertion is considered to have atypical angina. Similarly, a patient with sharp, exertional chest pain may also have atypical angina. Nonanginal pain is chest pain that does not have any features of angina: it is pleuritic or positional, unrelated to exertion, and is fleeting or lasts for many minutes. A detailed chest pain history allows the clinician to classify a patient's chest pain syndrome as typical angina, atypical angina, stable angina, unstable angina, or nonanginal chest pain. Among patients presenting to a clinician with chest pain, the presence of typical angina substantially increases the probability of significant IHD (likelihood ratio, 5.6), whereas the presence of atypical angina does not substantially alter the probability of significant IHD (likelihood ratio, 1.3).[23]

Once a detailed history of chest pain is obtained, cardiovascular risk factors are assessed; risk factors include increased age, male sex, menopausal status, cigarette smoking, hyperlipidemia, diabetes, hypertension, cerebrovascular disease, peripheral vascular disease, and a family history of premature coronary disease.[24] Cardiovascular risk factors greatly affect the pretest probability of significant IHD, particularly for women, younger patients, and patients with atypical chest pain syndromes [*see Table 5*]. For example, a 55-year-old man with atypical angina and no risk factors has a pretest probability of clinically significant IHD of 45%, whereas a 55-year-old man with typical angina and multiple cardiovascular risk factors has a 95% pretest probability of significant IHD.

Estimating the pretest probability of significant IHD is essential to determine whether further testing is warranted. For example, further diagnostic testing of the patient with a 45% probability of IHD would likely clarify the presence or absence of IHD. On the other hand, further testing of the patient with a 95% probability of IHD would be unlikely to alter the diagnosis of IHD, although further testing might help assess risk of cardiovascular events [*see* Risk Stratification in Patients with Chronic Stable Angina, *below*]. In general, further testing is not recommended for patients with a

Table 2 Short-Term Risk of Death or Nonfatal Myocardial Infarction in Patients with Unstable Angina[19]

High Risk	*Intermediate Risk*	*Low Risk*
At least one of the following features must be present: Prolonged ongoing (> 20 min) rest pain Pulmonary edema, most likely related to ischemia Angina with new or worsening MR murmur Angina with S_3 or new/worsening rales Angina with hypotension	No high-risk features but must have any of the following: Prolonged (> 20 min) rest angina, now resolved, with moderate or high likelihood of CAD Rest angina (> 20 min or relieved with sublingual nitroglycerin) Nocturnal angina New-onset CCSC III or IV angina in the past 3 wk with moderate or high likelihood of CAD Pathologic Q waves or resting ST depression ≤ 1 mm in multiple lead groups (anterior, inferior, lateral) Age > 65 yr	No high- or intermediate-risk feature but may have any of the following: Increased angina frequency, severity, or duration Angina provoked at a lower threshold New-onset angina with onset 2 wk to 2 mo before presentation Normal or unchanged ECG

Note: Estimation of the short-term risks of death and nonfatal myocardial infarction in unstable angina is a complex multivariable problem that cannot be fully specified in a table such as this. Therefore, the table is meant to offer general guidance and illustration rather than rigid algorithms.
CAD—coronary artery disease CCSC—Canadian Cardiovascular Society Classification MR—mitral regurgitation

very low pretest probability of significant IHD, as determined by a clinical assessment of patient history and risk factors.

Estimates of the risk of cardiovascular events can also be determined from clinical variables; the Framingham risk equations are commonly used for this purpose [*see Figure 2*].[25] This multivariate model estimates the 10-year risk of developing IHD on the basis of a patient's age; gender; total cholesterol level; and history of diabetes, hypertension, and smoking. This model is widely used, is readily available, and has been validated across a variety of populations.

The decision to pursue further testing in patients with possible angina appropriately incorporates patient preferences regarding diagnosis or intervention and an assessment of comorbidities [*see* Patients Warranting Noninvasive Testing, *below*].

Physical Examination

The physical examination of patients with chronic stable angina is often normal but may indicate the presence of hypertension (e.g., elevated blood pressure, enlarged or laterally displaced point of maximum impulse, S_4 gallop, or retinal vascular changes) or coexisting peripheral vascular disease (e.g., diminished pulses or bruits). In younger patients with premature coronary disease, there may be stigmata of genetic dyslipidemia syndromes (e.g., xanthelasma associated with familial hypercholesterolemia). For patients with chest pain, the presence of any of these findings increases the likelihood of significant IHD.

It is particularly helpful to examine a patient during an episode of angina. The presence of an S_4 or S_3 gallop, a mitral regurgitation murmur, a paradoxically split S_2 heart sound, bibasilar crackles, or a chest wall heave makes IHD more likely, particularly if the finding disappears when the pain goes away.[26] Physical examination findings that wax and wane with anginal symptoms are of particular significance, because they may indicate significant myocardial dysfunction at low work loads.

Laboratory Tests

A resting 12-lead electrocardiogram should be performed in all patients with suspected angina, although the results are normal in 50% of patients with chronic stable angina.[27] The presence of pathologic Q waves is virtually pathognomonic of

Table 3 Differential Diagnosis of Chest Pain

Diagnosis	*Characteristics*	*Comments*
Ischemic heart disease	Typical or atypical angina	Caused by diminished coronary blood flow and/or increased myocardial oxygen demand
Nonischemic heart disease		
Arrhythmias	Palpitations or typical angina	Tachycardia
Valvular heart disease	Typical angina, often exertional	Heart murmur present
Aortic dissection	"Tearing" pain, often abrupt onset	Widened mediastinum, often with hypertension
Pericarditis	Often pleuritic pain, but may be anginal; relieved by sitting up and leaning forward	Friction rub may be present; diffuse ST segment elevation (PR segment depression) on ECG
Pulmonary disease		
Pulmonary embolus	Pleuritic pain, associated dyspnea	Hypoxia/hypoxemia, pulsus paradoxus, and risk factors for thromboembolic disease
Pneumothorax	Acute onset, pleuritic pain, associated dyspnea	Hyperresonance on examination; tension pneumothorax is associated with distended neck veins, hypotension, and tachycardia
Pneumonia	Pleuritic pain	Associated with fever and productive cough
Gastrointestinal disease		
Esophageal disease	May be indistinguishable from angina	Often diagnosed following a negative evaluation for ischemic heart disease
Acid peptic disease	May be indistinguishable from angina	Pain often related to meals
Biliary disease	Right upper quadrant pain that radiates to the back or scapula	Typically worse following meals; right upper quadrant tenderness may be present
Pancreatitis	"Boring" epigastric pain, may radiate to the back	Chronic pancreatitis pain may occur without signs of systemic illness
Chest wall or dermatologic pain	Characteristically reproduced with palpation or movement	Reproduction of pain does not exclude angina
Costochrondritis		
Rib fracture		
Sternoclavicular arthritis		
Herpes zoster		Pain may precede rash
Fibrositis		Characteristic point tenderness
Psychiatric disorders	May be indistinguishable from angina	Often diagnosed following a negative evaluation for angina
Anxiety disorders		Often associated with palpitations, sweating, and anxiety
Affective disorders (e.g., depression)		
Somatoform disorders		
Thought disorders (e.g., fixed delusions)		
Factitious disorders (e.g., Munchausen syndrome)		

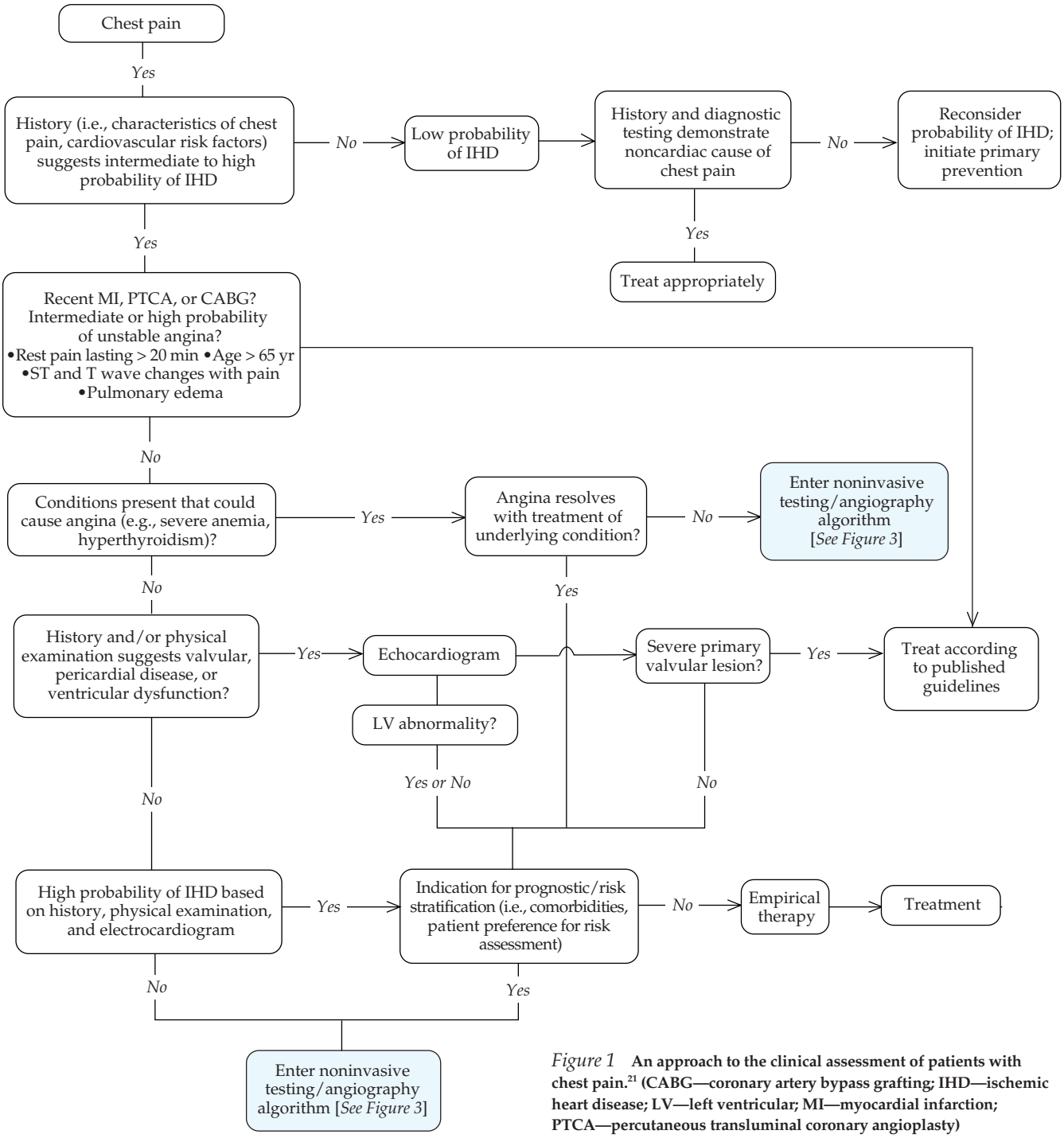


Figure 1 **An approach to the clinical assessment of patients with chest pain.[21] (CABG—coronary artery bypass grafting; IHD—ischemic heart disease; LV—left ventricular; MI—myocardial infarction; PTCA—percutaneous transluminal coronary angioplasty)**

clinically significant IHD, although isolated Q waves in lead III or a QS pattern in leads V1 and V2 is nonspecific. Several other ECG findings increase the clinical probability of IHD; ST segment depression, T wave inversions, and left ventricular hypertrophy (LVH) favor the diagnosis of angina.[28] Arrhythmias, including atrial fibrillation, ventricular tachyarrhythmias, bundle branch blocks, and atrioventricular block, increase the likelihood of IHD somewhat but are nonspecific.

As with the physical examination, an ECG obtained during an episode of pain may be particularly informative. About half of patients with a normal resting ECG will have an abnormality suggestive of ischemia during an episode of chest pain. Suggestive abnormalities include ST segment depression, T wave inversion, or "pseudonormalization" of these abnormalities during pain.[29] Abnormalities on a resting ECG that disappear with resolution of pain may indicate severe IHD, because they suggest ischemia at low work loads or unstable coronary syndromes.

Recommended laboratory tests include measurement of hemoglobin and fasting glucose levels and a fasting lipid panel (including measurement of levels of total cholesterol, HDL cholesterol, triglycerides, and calculated LDL cholesterol).[1] Hyperlipidemia is an important risk factor for IHD; the risk for IHD increases 1% for each 1 mg/dl increase in serum LDL cholesterol.[30] Similarly, patients with impaired glucose tolerance or

Table 4 Features of Typical Angina

Feature	*Typical Angina*	*Comments*
Location	Substernal but may radiate to neck, jaw, shoulder, or arms	The symptom of angina is typically felt below the ears and above the umbilicus
Duration	Typically less than 5 min	Longer-lasting chest pain raises concern for MI but may also be nonanginal
Quality	Dull, aching, pressurelike pain that is difficult to localize precisely	Levine sign, a clenched fist held over the chest well, represents the location and quality of typical angina
Exacerbating or relieving factors	Worsens with exertion or emotional stress and is relieved by rest or nitroglycerin; pain is often precipitated by a reproducible amount of exertion	Typical angina that occurs at lower than the usual amounts of exertion raises concern for unstable angina
Associated symptoms	Diaphoresis, nausea, palpitations, light-headedness, and dyspnea	Palpitations and light-headedness raise concern for arrhythmia; dyspnea suggests left ventricular or valvular dysfunction

frank diabetes are at increased risk for IHD. Normal hemoglobin excludes anemia. Thyroid function tests are indicated in patients who have signs or symptoms compatible with hyperthyroidism.

Imaging Studies

Chest radiography is of limited value in most patients with suspected angina[31]; however, a chest radiograph or other imaging study, such as computed tomography, is indicated in patients with signs or symptoms of congestive heart failure (CHF), valvular heart disease, or pericardial disease or in patients with possible aortic dissection or aneurysm. Echocardiography or multigated equilibrium radionuclide angiography should be obtained in any patient with suspected left ventricular impairment.

Patients Warranting Noninvasive Testing

Noninvasive testing usually has two objectives: to ascertain the probability of clinically important IHD and to estimate the risk of a serious cardiovascular event (e.g., MI or death) in the near future. These two objectives are often pursued concurrently, but it is useful to distinguish between diagnostic testing and testing for purposes of risk stratification.

Noninvasive testing is most likely to influence clinical decision making when the pretest probability of IHD is in the intermediate range. For example, a positive exercise treadmill test in a 55-year-old man with atypical chest pain and no other risk factors (pretest probability of clinically significant IHD, approximately 50%) would significantly increase the suspicion of clinically important IHD (posttest probability, 85%), whereas a negative exercise treadmill test would significantly reduce the suspicion of clinically significant IHD (posttest probability, 15%).

On the other hand, an abnormal exercise treadmill test in a 35-year-old woman with atypical chest pain and no other risk factors (pretest probability of clinically significant IHD, < 5%) would likely be falsely positive and could prompt use of unnecessary medications or potentially invasive diagnostic testing; a negative test would simply support a low clinical suspicion of disease. Therefore, further testing in such a low-risk patient would not be indicated. Similarly, further testing of high-risk patients is not likely to provide information that would alter the diagnosis of IHD. For example, because of the likelihood of significant coronary disease in a 65-year-old man with typical angina (pretest probability, 94%), a positive exercise test would only confirm the high clinical suspicion of IHD; a negative result would only lower the estimate into the moderate range and would not exclude the diagnosis of significant IHD.

Noninvasive testing is commonly obtained in persons with a high clinical probability of having significant IHD. In this setting, however, noninvasive testing is useful to assess risk and establish prognosis but not to establish or refute the diagnosis of coronary disease, as is the case for patients with an intermediate clinical probability of significant IHD.

Table 5 Comparison of Pretest Likelihood of IHD in Low-Risk Symptomatic Patients and High-Risk Symptomatic Patients[49]

Age (yr)	*Nonanginal Chest Pain (%)**		*Atypical Angina (%)**		*Typical Angina (%)**	
	Men	Women	Men	Women	Men	Women
35	3–35	1–19	8–59	2–39	30–88	10–78
45	9–47	2–22	21–70	5–43	51–92	20–79
55	23–59	23–59	45–79	10–47	80–95	38–82
65	49–69	49–69	71–86	20–51	93–97	56–84

*Each value represents the percentage with significant IHD. The first number given in each range (e.g., 3–35) is the percentage for a low-risk, mid-decade patient who does not have diabetes, does not smoke, and does not have hyperlipidemia. The second number in each range is the percentage for a same-age patient who does have diabetes, does smoke, or does have hyperlipidemia. Both high- and low-risk patients have normal resting ECGs. If ST-T wave changes or Q waves are present, the likelihood of IHD is higher in each entry of the table.

IHD—ischemic heart disease

Estimate of 10-Year Risk for Men

Age (yr)	Points
20–34	−9
35–39	−4
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	11
70–74	12
75–79	13

Total Cholesterol (mg/dl)	Points				
	Age 20–39 yr	Age 40–49 yr	Age 50–59 yr	Age 60–69 yr	Age 70–79 yr
< 160	0	0	0	0	0
160–199	4	3	2	1	0
200–239	7	5	3	1	0
240–279	9	6	4	2	1
≥ 280	11	8	5	3	1

	Points				
	Age 20–39 yr	Age 40–49 yr	Age 50–59 yr	Age 60–69 yr	Age 70–79 yr
Nonsmoker	0	0	0	0	0
Smoker	8	5	3	1	1

HDL (mg/dl)	Points
≥ 60	−1
50–59	0
40–49	1
< 40	2

Systolic Blood Pressure (mm Hg)	Untreated	Treated
< 120	0	0
120–129	0	1
130–139	1	2
140–159	1	2
≥ 160	2	3

Total Points	10-Year Risk (%)
< 0	< 1
0–4	1
5–6	2
7	3
8	4
9	5
10	6
11	8
12	10
13	12
14	16
15	20
16	25
≥ 17	≥ 30

Estimate of 10-Year Risk for Women

Age (yr)	Points
20–34	−7
35–39	−3
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	12
70–74	14
75–79	16

Total Cholesterol (mg/dl)	Points				
	Age 20–39 yr	Age 40–49 yr	Age 50–59 yr	Age 60–69 yr	Age 70–79 yr
< 160	0	0	0	0	0
160–199	4	3	2	1	1
200–239	8	6	4	2	1
240–279	11	8	5	3	2
≥ 280	13	10	7	4	2

	Points				
	Age 20–39 yr	Age 40–49 yr	Age 50–59 yr	Age 60–69 yr	Age 70–79 yr
Nonsmoker	0	0	0	0	0
Smoker	9	7	4	2	1

HDL (mg/dl)	Points
≥ 60	−1
50–59	0
40–49	1
< 40	2

Systolic Blood Pressure (mm Hg)	Untreated	Treated
< 120	0	0
120–129	1	3
130–139	2	4
140–159	3	5
≥ 160	4	6

Total Points	10-Year Risk (%)
< 9	< 1
9–12	1
13–14	2
15	3
16	4
17	5
18	6
19	8
20	11
21	14
22	17
23	22
24	27
≥ 25	≥ 30

Figure 2 **Estimates of major cardiovascular risk for men and women, based on the Framingham risk equations.[25]**

For purposes of deciding on a course of noninvasive testing, there is no precise definition of the upper and lower boundaries of intermediate probability of IHD; rather, this is a matter of clinical judgment in individual situations. Relevant issues include the degree of uncertainty acceptable to the physician and patient, the probability of an alternative diagnosis, the costs and risks of additional testing, and the benefits and risks of treatment in the absence of additional testing.[32] It is reasonable to consider a risk of clinically significant IHD of 10% to 20% or lower as low probability and of 80% to 90% risk or greater as high probability.[32]

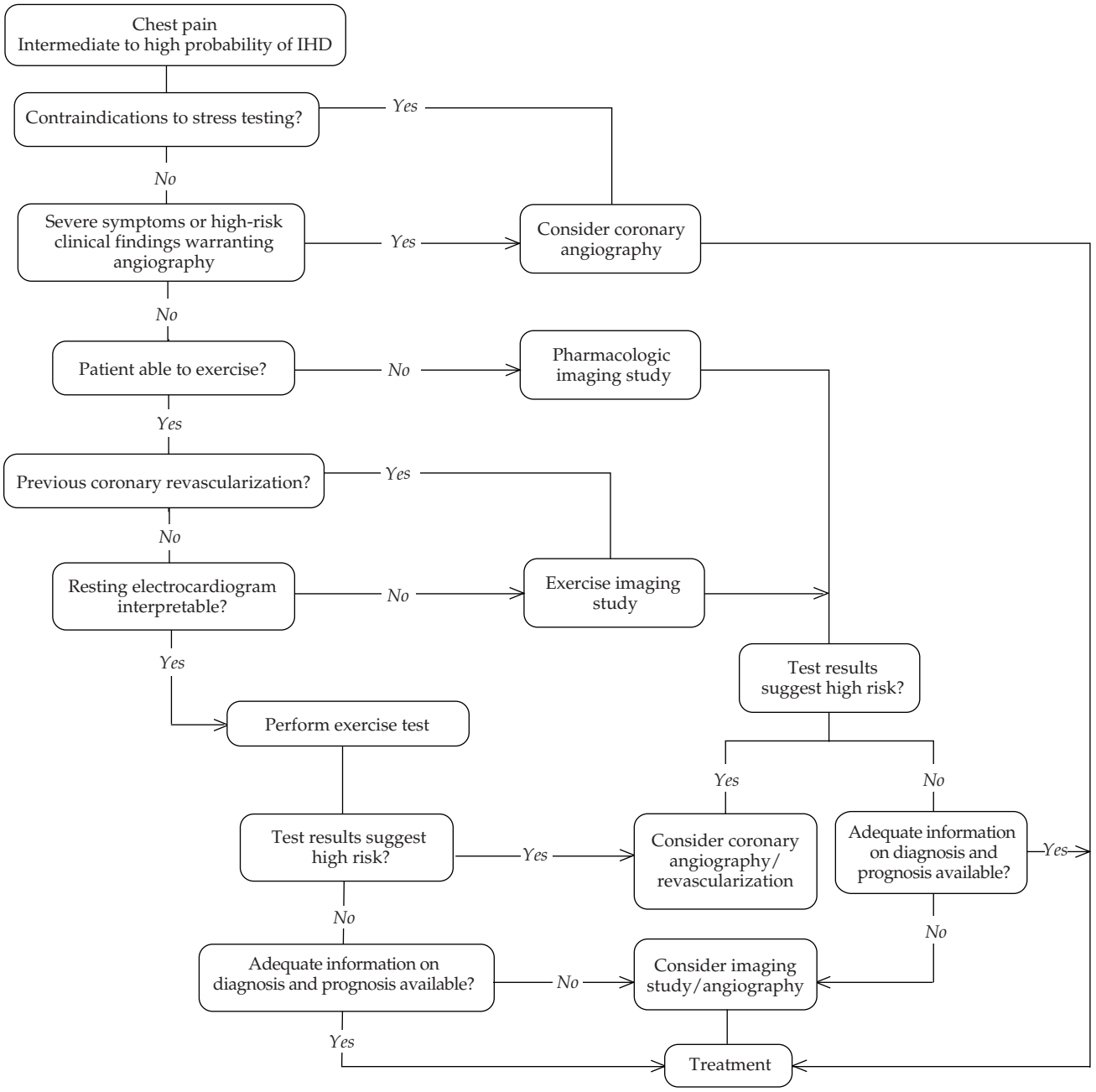


Figure 3 **Noninvasive testing and angiography in the evaluation of patients suspected of having ischemic heart disease.[21]**

NONINVASIVE TESTING

Patients whose history, physical examination, and ECG results indicate an intermediate or high probability of IHD usually should undergo further diagnostic evaluation [*see Figure 3*]. Noninvasive testing in patients with an intermediate probability of IHD provides important information about diagnosis (i.e., the presence or absence of coronary disease). In patients with an intermediate or high probability of IHD, noninvasive testing helps to stratify risk for major cardiovascular events. This stratification helps determine treatment strategies.

Commonly performed noninvasive studies include exercise ECG testing, stress radionuclide myocardial perfusion imaging, and stress echocardiography. The performance characteristics of various noninvasive tests and posttest probabilities of IHD for a range of pretest probabilities are listed [*see Table 6*].

ECG Exercise Testing

Guidelines from the American College of Cardiology/American Heart Association/American College of Physicians (ACC/AHA/ACP) recommend exercise ECG as the first-choice diagnostic test for the average patient with an intermediate pretest probability of IHD and a normal resting ECG.[32,33] Exercise testing has imperfect sensitivity and specificity (68% and 77%, respectively), but it is widely available and inexpensive. In addition, it readily identifies patients at high risk for IHD and provides important prognostic information [*see* Risk Stratification in Patients

with Chronic Stable Angina, *below*]. Exercise testing is generally safe; MI or death occurs with a frequency of less than 1 in 2,500 tests.[34] Symptom-limited exercise testing is safe in patients with unstable angina who lack evidence of MI and CHF and who are free of chest pain at the time of testing.[32] Specific absolute and relative contraindications for exercise testing can be found in guidelines from the ACC/AHA.[1] Other noninvasive tests are preferred in patients who are unable to exercise, in those who previously underwent coronary revascularization, and in those with specific resting ECG abnormalities that would interfere with the interpretation of an exercise ECG (such abnormalities include ST segment depression of greater than 1 mm at rest, preexcitation syndrome, electronically paced rhythm, left bundle branch block, and left ventricular hypertrophy with repolarization abnormalities).[32]

Interpretation of the exercise test depends on the patient's exercise capacity, symptoms, the reasons for stopping the test, the hemodynamic response, any pertinent findings on physical examination (e.g., exercise-induced S_3 heart sound or mitral regurgitation), and any changes in the ECG. Patients who stop the test because of the onset of angina are very likely to have significant IHD. Exercise-induced falls in blood pressure or the development of an exercise-induced S_3 heart sound are strongly suggestive of ischemic left ventricular dysfunction. Specific exercise-induced ECG changes suggestive of IHD include a horizontal or downward-sloping ST segment depression or elevation of greater than 1 mm during or after exercise.[1] Exercise-induced changes in lead V5 are most reliable for the diagnosis of IHD.[1]

Stress Radionuclide Myocardial Perfusion Imaging

Although exercise ECG continues to be recommended as the diagnostic test of choice for most patients with suspected angina, stress myocardial perfusion imaging using single-photon emission computed tomography (SPECT) is the most common noninvasive test performed in the United States. SPECT imaging following exercise or pharmacologic stress (e.g., using dipyridamole, adenosine, or dobutamine) has greater sensitivity for IHD than exercise testing, particularly in patients with an abnormal resting ECG; in addition, SPECT can define vascular regions in which stress-induced coronary flow is limited. Furthermore, SPECT imaging allows an estimation of left ventricular (LV) systolic size and function.

Exercise myocardial perfusion SPECT is preferred for patients who have baseline ECG abnormalities that interfere with the interpretation of an exercise ECG. Such abnormalities include a resting ST segment depression greater than 1 mm and LVH; digoxin therapy and ventricular preexcitation can also cause ECG abnormalities.[35] Myocardial perfusion SPECT with pharmacologic stress is preferred in the setting of left bundle branch block or ventricular pacing.[35] Because of its greater sensitivity, stress myocardial perfusion imaging is an option after a nondiagnostic exercise ECG. Stress myocardial perfusion imaging can also further stratify the risk of IHD in patients with an abnormal exercise ECG.[36]

Stress Echocardiography

Stress echocardiography (i.e., echocardiography performed after exercise or dobutamine administration) is another option for noninvasive testing to establish the diagnosis of IHD. Stress induces regional wall motion abnormalities in the myocardial regions supplied by stenotic coronary vessels; stress echocardiography defines such regions of the left ventricular wall. Stress echocardiography, like myocardial perfusion imaging, is a good choice for patients with ECG abnormalities that might interfere with the interpretation of an exercise ECG.[37] The sensitivity of stress echocardiography, as with that of stress myocardial perfusion imaging, is in the range of 80% to 85%. Exercise echocardiography is marginally more specific than other noninvasive diagnostic tests.[38]

Test Selection

Although exercise ECG remains the recommended initial noninvasive test for most patients with suspected IHD, clinical decision making should also take into account individual patient characteristics, local expertise, and availability. Stress echocardiography and exercise ECG can be performed in a physician's

Table 6 Posttest Probability of Significant IHD Based on Pretest Probabilities of IHD and Normal or Abnormal Results of Noninvasive Studies

Test Result	*Posttest Probability of IHD (%)*					
	20% Pretest Probability		*50% Pretest Probability*		*80% Pretest Probability*	
	Men	Women	Men	Women	Men	Women
Exercise ECG						
Abnormal	71	43	91	75	98	92
Normal	13	16	38	43	71	75
ECG with SPECT						
Abnormal	85	71	96	91		
Normal	3	4	11	13	99	98
Exercise echocardiography					33	36
Abnormal	38	37	71	70		
Normal	15	17	41	44	91	82
Dobutamine echocardiography*					74	76
Abnormal	46	—	77	—		
Normal	3	—	11	—	93	—
					32	—

Note: calculations are based on point estimates of the sensitivities and specificities of noninvasive studies abstracted from Gibbons.[21]
*Available studies did not include women in sufficient numbers to accurately estimate sensitivity and specificity for these tests.
IHD—ischemic heart disease SPECT—single-photon emission computed tomography

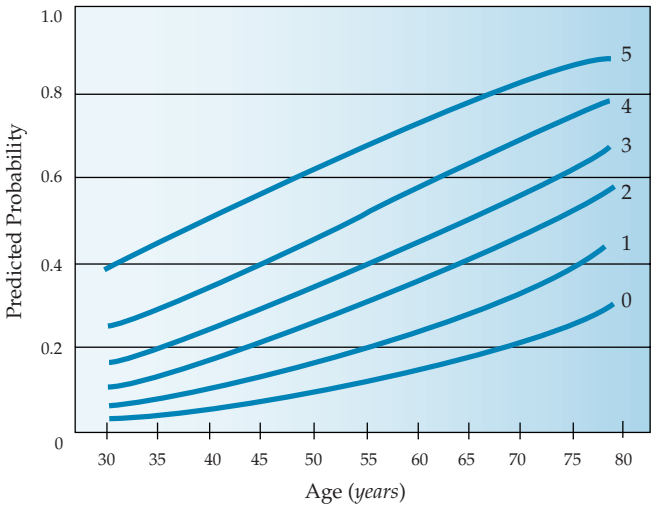


Figure 4 **Nomogram showing the probability of severe coronary disease (i.e., three-vessel disease or left main coronary artery disease) on a five-point score. One point is awarded for each of the following variables: male gender; typical angina; history or electrocardiographic evidence of myocardial infarction; diabetes; and use of insulin. Each curve shows the probability of severe coronary disease as a function of age.[50]**

office, whereas radionuclide myocardial perfusion imaging requires a specialized setting. In addition, myocardial perfusion imaging is considerably more expensive than exercise ECG.

Specific noninvasive tests are preferred in certain patient subsets. Patients unable to exercise should undergo some form of pharmacologic noninvasive test. Pharmacologic stress echocardiography using dipyridamole, adenosine, or, less commonly, dobutamine is an option for patients who are unable to exercise. Nonspecific perfusion image defects may be more common with exercise myocardial perfusion imaging in patients with left bundle branch block; pharmacologic stress imaging or echocardiography is preferred in these patients.[32] Exercise ECG is less accurate in establishing the diagnosis of IHD in women, and some authors have suggested that SPECT imaging and stress echocardiography are more accurate diagnostic tests for women with an intermediate probability of having IHD.[39,40] The ACC/AHA/ACP expert panel concluded, however, that insufficient data are available to recommend stress imaging or echocardiography over standard exercise ECG in women with suspected angina.[32]

Echocardiography is less sensitive in very obese patients, and examination tables used for SPECT imaging often have weight limitations. Planar scintigraphy, positron-emission tomography, and coronary angiography are testing options for obese patients. Planar scintigraphy is less sensitive than SPECT, and positron emission tomography is less well characterized than other modalities. There is some rationale for performing stress imaging or echocardiography in patients with angina whose histories suggest they are at high risk for major cardiovascular events (e.g., patients who experience angina at low work loads or whose angina is progressive); in such patients, determination of LV systolic function and anatomic distribution of IHD are important considerations in anticipation of a possible need for revascularization. For these patients, it may be appropriate to proceed directly to coronary angiography [*see* Invasive Testing, *below*].

Beta blockade reduces the sensitivity of all noninvasive tests, particularly exercise ECG[41]; it is recommended that, whenever possible, beta blockers be withheld for four half-lives (approximately 48 hours) before noninvasive testing is undertaken. Digoxin causes resting ST segment depression and reduces the specificity of exercise ECG.

In general, it is cost-effective to perform exercise ECG as the initial test in most patients, followed by additional imaging for patients in whom further diagnostic or prognostic testing is warranted.[42] Noninvasive testing is not generally recommended as a screening test in asymptomatic persons. Specific exceptions include patients whose occupation requires periodic stress testing (e.g., pilots, firefighters, competitive athletes, and police).

Imaging Studies under Investigation

Several CT and magnetic resonance imaging techniques are currently under study for use in diagnosing patients with suspected angina or in stratifying such patients for risk of cardiac events. Of these techniques, the most widely used is electron-beam computed tomography (EBCT), which detects coronary artery calcification. Although the sensitivity of EBCT for the diagnosis of significant coronary artery stenosis is high, the specificity of EBCT for significant coronary artery stenoses ranges from only 41% to 76%, yielding many false positive results. Some studies suggest that the "calcium score" derived from EBCT predicts the extent of angiographically detected CAD.[43] Studies that correlate calcium scores with risk of future cardiac events have been fraught with methodologic problems[44,45]; it is not clear whether EBCT adds significantly to clinical risk assessment using validated tools such as the Framingham risk score.[46] To date, no prospective, population-based studies have investigated a potential association between calcium score derived from EBCT and risk of future coronary events; likewise, no studies have shown that screening for IHD with EBCT reduces mortality.

Asymptomatic patients identified as being at potentially high risk for cardiac events on the basis of EBCT may suffer anxiety and undergo unnecessary procedures as a result of the study. Estimates of the cost efficacy of EBCT relative to current strategies for diagnosis and risk stratification are varied and are sensitive to the prevalence of disease in a screened population; currently, the clinical benefits of screening asymptomatic patients are uncertain.[47] The ACC/AHA guidelines do not recommend EBCT and other imaging procedures, such as MRI angiography, in asymptomatic patients.[48] Although EBCT testing is currently not recommended as a screening test, it is reasonable to evaluate patients who have undergone EBCT and who have been found to have severe coronary calcification with some form of noninvasive testing.

INVASIVE TESTING

Coronary angiography provides unequaled detail of coronary anatomy, including detail sufficient to enable evaluation for possible revascularization; however, angiography is invasive and expensive compared with noninvasive testing. In addition, although angiography identifies the degree and distribution of coronary stenoses, it provides inconsistent assessment of the functional significance and stability of a particular coronary lesion[32]; for example, plaques that can possibly rupture and thereby cause acute coronary syndromes are commonly angiographically insignificant.[18] Moreover, studies have called into question the reliability of angiographic measurements of coronary stenosis in some settings.[49]

The decision to pursue angiography should be based on preferences of the patient and provider, coexisting illnesses, esti-

mates of the probability of high-risk IHD, and the urgency to confirm or refute a possible diagnosis of IHD.

Direct referral for coronary angiography is recommended for patients who have survived sudden death, because of the high probability of multivessel IHD in these patients.[32] Other patients for whom angiography is recommended as a diagnostic test include those for whom the diagnosis of IHD remains uncertain despite noninvasive testing, provided the benefit of a more certain diagnosis outweighs the risk and cost of angiography; patients who cannot undergo noninvasive testing because of disability, illness, or morbid obesity; patients with suspected nonatherosclerotic angina (e.g., coronary dissection, coronary anomaly, Kawasaki disease, and coronary artery spasm); and patients with a high pretest probability of disease of the left main coronary artery or three-vessel IHD [*see* Prognostic Value of Coronary Angiography, *below*].[32]

DIAGNOSIS OF CHRONIC STABLE ANGINA

The presence of clinically stable, typical angina for a period of 2 or more months is adequate to establish the diagnosis of chronic stable angina. As mentioned [*see* Preliminary Evaluation, *above*], an accurate estimate of the likelihood of significant IHD can be established from simple clinical criteria (i.e., characteristics of chest pain and cardiovascular risk factors). For patients with an uncertain diagnosis after undergoing clinical assessment, particularly patients with some cardiovascular risk factors and an atypical angina syndrome, noninvasive testing is valuable for establishing the diagnosis of IHD. IHD can be confirmed by coronary angiography, which is also helpful in defining the severity of coronary atherosclerosis [*see* Risk Stratification in Patients with Chronic Stable Angina, *below*].

RISK STRATIFICATION IN PATIENTS WITH CHRONIC STABLE ANGINA

Risk stratification that determines the prognosis for MI or death is essential in determining treatment recommendations for patients with chronic stable angina. In general, a patient's coronary risk is determined by the interplay of four factors[32]: (1) left ventricular systolic function; (2) the extent and severity of atherosclerotic occlusion of the coronary tree (i.e., ischemic burden); (3) plaque stability (i.e., risk of plaque rupture); and (4) coexisting medical conditions. Clinical parameters, results of noninvasive testing, and coronary angiography provide important prognostic and diagnostic information.

Clinical Parameters Indicating High Risk

Although noninvasive testing and coronary angiography are the mainstays of risk stratification, clinical parameters alone are sufficient to identify some patients as having a high probability of severe IHD (i.e., three-vessel disease or left main CAD). Hubbard and colleagues developed a simplified algorithm for predicting the probability of severe IHD on the basis of six clinical parameters: age, gender, presence of typical angina, presence of diabetes, insulin use, and prior MI (as indicated by history or ECG) [*see Figure 4*].[50] Older patients with multiple risk factors have a greater than 50% chance of having severe IHD and should be considered for direct referral for coronary angiography. In one study, a previous history of MI and the presence of a carotid bruit (a marker for peripheral vascular disease) more than doubled the probability of severe IHD.[51] Direct referral for coronary angiography is estimated to be cost-effective when the pretest probability of severe IHD is high.[52]

Follow-up Noninvasive Testing

Noninvasive testing is a sensible approach for the majority of patients without high-risk characteristics. Available diagnostic modalities predict death more accurately than cardiovascular events such as MI. Patients with IHD are stratified according to their risk of death into three categories: those at low risk (< 1% mortality a year); those at intermediate risk (1% to 3% mortality a year); and those at high risk (> 3% mortality a year). Persons estimated to be at low risk for death generally may be managed medically without further diagnostic testing unless their condition deteriorates.[32] Persons estimated to be at intermediate or high risk after initial noninvasive testing may need to undergo additional studies for the purpose of further risk stratification.

Left ventricular systolic function Declining left ventricular systolic function is the strongest single predictor of long-term mortality in patients with IHD.[53] Patients with an ejection fraction of less than 35% have a mortality in excess of 3% a year. Left ventricular function can be assessed by echocardiography, stress myocardial perfusion imaging with SPECT, angiographic ventriculography, or gated nuclear medicine studies. As noted, not all patients with angina require evaluation of left ventricular function. Patients with a normal resting ECG and no history of MI or CHF are very likely to have normal left ventricular systolic function (92% to 95% probability).[54] Patients with a history of CHF, MI, or ECG evidence of prior MI should undergo evaluation of left ventricular function.

Exercise treadmill ECG In addition to the diagnostic information it provides, exercise treadmill ECG testing supplies useful prognostic information. Exercise capacity during a treadmill test is one of the strongest predictors of cardiovascular risk. Exercise capacity is influenced, in part, by LV function, both at rest and with exercise. Several measures of exercise capacity are used: exercise duration, maximum heart rate, exercise duration × heart rate, and estimates of work measured in metabolic equivalents (METs). Other important variables include ECG measures of exercise-induced ischemia, as reflected in ST segment depression or elevation, and the duration of ST segment deviation during the recovery phase of the exercise protocol.

The Duke treadmill score (DTS) combines these exercise test variables and is the most widely used prognostic treadmill score.[55] The DTS is calculated as follows:

Exercise time (in min) – (5 × the ST segment deviation [in mm]) – (4 × the angina index)

The angina index equals 2 when angina is the reason for stopping the exercise test; it equals 1 when angina occurs during the test or the recovery period; and it equals 0 if no angina occurs. In patients with a low-risk DTS (i.e., ≥ +5), the 4-year survival is 99% (average annual mortality, 0.25%) [*see Table 7*]. Patients with a high-risk DTS (< –10) have a 4-year survival rate of 79% (average annual mortality, 5%). In one study, more than two thirds of outpatients with suspected IHD had low-risk scores, and only 4% of patients had high-risk scores.[55] Available data suggest that patients with frequent ventricular ectopy or a slow heart rate recovery time following exercise testing are also at increased risk for death during subsequent follow-up,[56] although it remains uncertain how best to incorporate this information into the stratification of risk.

Stress testing Stress myocardial perfusion imaging and stress echocardiography are commonly used in risk stratification

Table 7 Survival According to Risk Groups Based on Duke Treadmill Score (DTS)

Risk Group (DTS)	*Overall Survival (%)*	*4-Year Survival (%)*	*Annual Mortality (%)*
Low risk (≥ +5)	62	0.99	0.25
Moderate risk (–10 to +4)	34	0.95	1.25
High risk (< –10)	4	0.79	5.0

of patients with IHD, although their prognostic value is less known than that of exercise ECG.[57] In patients with an intermediate-risk exercise ECG (DTS < +5 and ≥ –10), annual mortality is between 1% and 3%; these patients are usually referred for additional diagnostic testing, either stress imaging or coronary angiography. In patients with a high-risk exercise ECG result (DTS < –10), annual cardiac mortality is estimated to exceed 3%; these patients are generally referred for coronary angiography. Patients with a low-risk exercise ECG (DTS ≥ +5) require no further testing and may be medically managed. It is also appropriate to obtain a stress imaging study for purposes of risk stratification for patients who are unable to exercise and for patients with uninterpretable rest ECGs. Stress imaging studies are preferred over exercise ECG for evaluating ischemia in symptomatic patients who have previously undergone revascularization.

A normal poststress myocardial perfusion imaging study, as with a low-risk exercise ECG, indicates an excellent prognosis, even among patients with chronic stable angina.[58] The number, size, and location of abnormalities on stress myocardial perfusion studies indicate the distribution and severity of coronary artery stenoses; larger and more numerous perfusion defects are associated with more severe CAD. Stress-induced LV dysfunction—a marker of severe multivessel IHD—is indicated by LV dilatation or lung uptake of radionuclide tracer.[59,60] Patients with two or more moderate to large stress-induced perfusion defects or evidence of stress-induced LV dysfunction are considered to have a high probability of severe IHD; these patients are candidates for referral for coronary angiography.

A negative stress echocardiogram also indicates a good prognosis.[61] In the presence of significant IHD, stress induces regional wall motion abnormalities in the myocardial regions supplied by the stenotic coronary vessel. Stress-induced wall motion abnormalities involving two or more segments at lower levels of stress predict high-risk IHD. As with stress myocardial perfusion imaging, stress-induced LV dilatation suggests multivessel IHD. Abnormal stress echocardiography provides diagnostic and prognostic information that is incremental to that obtained by exercise ECG,[62] although there are relatively fewer follow-up data for this test than for myocardial perfusion studies.

Prognostic Value of Coronary Angiography

The anatomic extent of IHD is a powerful indicator of prognosis. Referral for coronary angiography should be considered for patients with a high-risk exercise ECG (DTS < –10), a high-risk stress myocardial perfusion imaging study, or a high-risk stress echocardiogram, provided the patient is a candidate for revascularization.

In the Coronary Artery Surgery Study database of patients with suspected IHD who were referred for coronary angiography, the 12-year survival rate for patients with normal coronary arteries was 91%, compared with 74% for patients with one-vessel disease, 59% for patients with two-vessel disease, and 40% for patients with three-vessel disease.[63] Proximal lesions are associated with greater risk than more distal lesions.[64] LV function remains crucially important. For example, the 5-year survival of a 65-year-old man with stable angina, three-vessel coronary stenoses, and normal LV function is 93%, compared with 5-year survival of only 58% in a similar patient with an LV ejection fraction of 30%.[64] Although angiography provides detailed information about the extent and severity of stenoses, it does not define which ones are actually responsible for anginal symptoms. Furthermore, the atherosclerotic plaques most likely to rupture and to thereby cause acute coronary syndromes, including MI and death, are often missed by coronary angiography.[18]

In summary, for patients with IHD, the strongest predictor of long-term survival is LV systolic function. A second determinant of survival is the severity and distribution of coronary lesions. A third determinant is the stability of coronary plaques; plaque instability and rupture increase the short-term risk of unstable coronary syndromes and death. Currently, there is no satisfactory means to measure this third determinant.

Treatment

The two overarching goals of the treatment of patients with chronic stable angina are to reduce the likelihood of untoward clinical events (i.e., acute coronary syndromes and sudden death) and to improve quality of life by reducing anginal symptoms and enhancing function. Treatment options include lifestyle modifications, medications, and revascularization. Evidence-based guidelines have been published, and a general approach to the treatment of patients with chronic stable angina is outlined below. However, management should be individualized in accordance with a patient's risk of adverse outcomes, coexisting conditions, and preferences, as well as in consideration of the cost and effectiveness of therapeutic alternatives. An ACC/AHA/ACP expert panel developed an mnemonic for the treatment of patients with chronic stable angina [*see Table 8*].[21]

LIFESTYLE MODIFICATION

Smoking Cessation

Among nonpharmacologic interventions, smoking cessation has the greatest impact on total mortality and cardiovascular risk. A systematic review of prospective cohort studies of smokers with IHD found a striking 29% to 36% relative risk reduction in all-cause mortality for patients who were able to quit smoking.[65] Most patients in these cohort studies had previous MI, angioplasty, or coronary artery bypass grafting (CABG) at the time

Table 8 The "ABCDEs" of Treatment for Patients with Chronic Stable Angina[20]

A = Aspirin and antianginal therapy; ACE inhibitors should be considered for most patients
B = Beta blocker and blood pressure
C = Cigarette smoking and cholesterol
D = Diet and diabetes
E = Education and exercise

ACE—angiotensin–converting enzyme

of entry into the study. The magnitude of the risk reduction for smoking cessation was as great as or greater than that expected to result from use of aspirin, statins, beta blockers, or angiotensin-converting enzyme (ACE) inhibitors. Smoking cessation should be strongly and repeatedly recommended to all patients with known or suspected IHD who smoke. Physicians should become accustomed to applying effective techniques for counseling patients and using effective medications such as nicotine replacement and bupropion. Patients' efforts to quit are often more effective in the setting of a formal smoking-cessation program; therefore, internists should be familiar with local smoking-cessation resources and programs.

Physical Activity and Dietary Modifications

Physical activity Regular physical activity and dietary modifications also reduce cardiovascular risk. Physical fitness during middle age is associated with lower long-term cardiovascular mortality[66]; in addition, self-reported increases in physical activity is associated with reduced all-cause and cardiovascular mortality in elderly men.[67] In the Health Professionals Follow-up Study, half an hour or more of brisk walking each day was associated with an 18% relative risk reduction in cardiovascular events; in addition, greater duration and intensity of exercise were associated with greater reductions in risk.[68] This study also suggested that weight training was associated with a decreased cardiovascular risk. Although these observational studies are potentially subject to bias, the preponderance of evidence strongly supports recommending regular physical activity to patients.

Patients with chronic stable angina should be encouraged to include moderate aerobic activity in their daily lives.[21] Moderate physical activity consists of walking briskly for 30 minutes or more five to seven times a week or the equivalent. Unresolved questions include whether more vigorous physical activity provides greater risk reduction than moderate exercise; whether sustained aerobic exercise of 30 or more minutes is necessary for cardiovascular benefit or whether an accumulation of 30 or more minutes of physical activity during the day is sufficient; and to what extent physical activity provides reductions in risk above and beyond those achieved simply through modification of specific risk factors, such as dyslipidemia, hypertension, and diabetes.

Dietary modification Diet also has the potential to modify multiple coronary risk factors—namely, lipid levels, obesity, insulin resistance, and hypertension. Although trials assessing the effects of dietary modification on stable IHD have not uniformly demonstrated benefit, several trials have shown reductions in cardiac mortality. In the Lyon Diet Heart Study, patients who had had an MI were randomized either to adopt a Mediterranean diet rich in fresh fruits and vegetables, whole grains, olive oil, fish, and relatively little meat or to adopt a prudent Western diet low in saturated fats. At 4 years' follow-up, patients randomized to the Mediterranean diet enjoyed a 2.5% to 3% per year reduction in cardiac death and nonfatal MI.[69] Another trial randomized patients with IHD (about half of whom had a history of MI) either to adopt an Indo-Mediterranean diet rich in fresh fruits, vegetables, legumes, nuts, and whole grains and supplemented with omega-3 fatty acids or to adopt a prudent Indian diet low in saturated fats. Moderate exercise was recommended for all patients. Patients randomized to the Indo-Mediterranean diet had a 7.4% absolute reduction in the risk of MI or cardiac death at 2 years' follow-up.[70] Patients randomized to the intervention diet had significant reductions in daily intake of calories, protein, fat (mostly saturated), cholesterol, and salt; they ingested significantly more complex carbohydrates, fiber, monounsaturated and polyunsaturated fats, fruits, vegetables, legumes, nuts, and omega-3 polyunsaturated oils.

Other studies demonstrate evidence of the protective benefit of fresh fruits and vegetables. The Nurses' Health Study and Health Professionals Follow-up Study found that persons in the highest quintile of fruit and vegetable intake had a 20% relative risk reduction for nonfatal MI or cardiac death compared with the lowest quintile.[71] A review of diet and IHD concluded that three dietary strategies are effective at reducing the risk of IHD: (1) substituting unsaturated fats (particularly polyunsaturated fats) for saturated fats (e.g., animal fats) and *trans*-fatty acids (e.g., stick margarine, vegetable shortenings, many commercially prepared baked goods, and deep-fried foods); (2) increasing consumption of omega-3 fatty acids (e.g., oily fish, canola oil, soybean oil, and flaxseed oils); and (3) consuming a diet high in fruits, vegetables, nuts, and whole grains and low in refined grains.[72]

MEDICAL THERAPY TO REDUCE CARDIOVASCULAR RISK

Antiplatelet Agents

Aspirin Acute coronary events commonly result from rupture of an atherosclerotic plaque and subsequent platelet aggregation and thrombosis. Aspirin inhibits cyclooxygenase and the synthesis of prothrombotic platelet thromboxane A_2. In studies of more than 3,000 patients with chronic stable angina that compared treatment with aspirin to placebo, the use of aspirin reduced the risk of adverse cardiovascular events by 33% over 6 months.[73,74] This reduction in relative risk corresponds to a reduction in absolute risk of approximately 5%. In other words, five cardiovascular events would be prevented for every 100 persons with known cardiovascular disease treated with aspirin for 6 months.[73] In the Swedish Angina Pectoris Aspirin Trial, which involved patients with stable angina, the addition of aspirin (75 mg daily) to a regimen of sotalol resulted in a 34% decrease in MI and sudden death[75]; most of this decrease involved reductions in the incidence of first MI. In the Physician's Health Study, which involved asymptomatic middle-aged men, the use of aspirin (325 mg every other day) was associated with a decrease in the incidence of MI.[76] Doses ranging from 75 to 325 mg daily were found to offer equivalent benefit,[77] although the incidence of gastrointestinal toxicity was dose-dependent. All patients with chronic stable angina should be treated with aspirin unless there is a history of documented aspirin allergy or life-threatening gastrointestinal hemorrhage.

Aspirin alternatives Two thienopyridine derivatives, clopidogrel and ticlopidine, inhibit adenosine diphosphate–mediated activation of platelet glycoprotein IIb/IIIa. These agents, particularly clopidogrel, are reasonable alternatives for aspirin-intolerant patients with chronic stable angina. In a randomized, controlled trial of patients with symptomatic vascular disease (including patients with chronic stable angina), clopidogrel was slightly more effective than aspirin in reducing the risk of MI, vascular death, and ischemic stroke.[78] Clopidogrel is much more expensive than aspirin; in addition, approximately 200 more patients would need to be treated with clopidogrel than with aspirin for 2 years to prevent one major vascular event.[79] There are no studies demonstrating that ticlopidine reduces cardiovascu-

lar events in outpatients with chronic stable angina. Ticlopidine can cause cytopenia, and there is a reported rare association with thrombotic thrombocytopenic purpura.

Lipid-Lowering Agents

There are abundant data showing the beneficial effects of lipid-lowering therapy in patients with chronic stable angina. Each 1% reduction in total cholesterol is associated with an approximately 2% reduction in coronary events.[80] In the Scandinavian Simvastatin Survival Study, patients with documented ischemic heart disease (including chronic stable angina) and elevated total cholesterol levels (i.e., levels of 212 to 308 mg/dl) were randomized to receive either a statin or placebo.[81] A 30% to 35% reduction in mortality and major coronary events was observed in patients receiving a statin. In the Cholesterol and Recurrent Events Trial, patients with prior MI and somewhat lower cholesterol levels (i.e., mean total cholesterol and LDL cholesterol of 209 mg/dl and 139 mg/dl, respectively) had a 25% relative risk reduction in the composite outcome of fatal or nonfatal MI when treated with a statin.[82] In the Heart Protection Study, patients with IHD or conditions that confer a similarly high risk of MI or coronary death (e.g., peripheral vascular disease, cerebrovascular disease, or diabetes) were randomized to receive either simvastatin (40 mg) or placebo, irrespective of baseline LDL level.[83] Statin therapy reduced total mortality; the incidence of first MI, coronary death, and stroke; and the use of revascularization procedures among all groups of patients, including those with LDL cholesterol levels of less than 116 mg/dl at entry (3.0 mmol/L). From this study, it may be concluded that all patients with chronic stable angina should be treated with a statin, barring specific allergy. Detailed recommendations for lipid-lowering therapy are provided by the National Cholesterol Education Program Adult Treatment Program III (NCEP ATP III).[30]

The target for therapy for patients with known IHD, including patients with chronic stable angina, is a serum LDL cholesterol level of less than 100 mg/dl. Patients with diabetes, peripheral vascular disease, and cerebrovascular disease are regarded as having a risk of cardiovascular events equivalent to patients with established IHD; the target for therapy in these patients is an LDL cholesterol level of less than 100 mg/dl, which is the same as in patients with known IHD. For other patients, including patients with possible IHD (e.g., a patient with atypical angina and a nondiagnostic exercise treadmill test) or two or more cardiovascular risk factors, the aggressiveness of lipid-lowering therapy is determined by calculating cardiovascular risk from the Framingham risk calculator[30] and the Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (http://www.nhlbi.nih.gov/guidelines/cholesterol/index.htm) [*see Figure 2*].

Available data suggest a potential benefit for more aggressive lowering of LDL cholesterol. In one study, low-risk patients with stable, mild to moderate angina were randomized to receive either atorvastatin, 80 mg daily, or percutaneous coronary intervention (PCI) followed by usual care (including lipid-lowering treatment). Patients receiving atorvastatin required fewer revascularization procedures or admissions for worsening angina with objective evidence of ischemia (13.4% versus 20.9%).[84] Patients in the atorvastatin-treated group reached an average LDL cholesterol level of 77 mg/dl, as compared with 119 mg/dl in the PCI group. A second study compared the effects of intensive lipid-lowering therapy using atorvastatin (80 mg/day) with moderate lipid-lowering therapy using pravastatin (40 mg/day).[85] Patients treated with atorvastatin (80 mg/day) achieved an average LDL cholesterol level of 79 mg/dl (2.05 mmol/L) and showed less progression of coronary atherosclerosis than patients treated with pravastatin (40 mg/day), who achieved an average LDL cholesterol level of 110 mg/dl (2.85 mmol/L).

Statin therapy is associated with several important potential adverse reactions. Elevations of liver transaminase levels have been described in patients taking statin drugs, although elevations to levels greater than three times the upper limit of normal (necessitating discontinuance) occur in fewer than 0.3% to 0.5% of patients.[86]

Rhabdomyolysis from statins is an uncommon dose-related phenomenon that usually occurs within the first few weeks of therapy, but it may occur at any time; rhabdomyolysis resolves after withdrawal of the offending drug. Clinically significant myopathy with 10-fold elevations of creatine kinase (CK) occurs in about 0.5% of patients treated with statins.[86] Massive rhabdomyolysis usually occurs only with concomitant use of clofibrate, niacin, or gemfibrozil. The incidence appears to be higher with drugs that interfere with the cytochrome P-450 system (e.g., simvastatin, lovastatin, and atorvastatin) and lower with less potent agents (e.g., pravastatin and fluvastatin). Predispositions to this adverse effect include certain P-450 polymorphisms (CYP3A4), renal failure, liver disease, hypothryoidism, concomitant medications (e.g, macrolides, azole antifungals, cyclosporine, protease inhibitors, selective serotonin reuptake inhibitors, nefazodone, verapamil, diltiazem, and amiodarone), and ingestion of grapefruit juice in quantity.[87] Myopathy can also occur without elevation of CK.[88]

Niacin, fibric acid derivatives (e.g., gemfibrozil and clofibrate), and bile acid sequestrants (e.g., cholestyramine and colestipol) reduce cholesterol an average of 6% to 15%. Niacin raises the HDL cholesterol, lowers the LDL cholesterol, and reduces the level of triglycerides. A meta-analysis of 37 studies suggests that the lipid-lowering effect of these agents is associated with reduced coronary mortality and total mortality.[89] In high-risk patients with a low HDL cholesterol level or hypertriglyceridemia, it is reasonable to consider the use of niacin or a fibric acid derivative, alone or in combination with a statin.[90-92] In the Veterans Affairs HDL Intervention Trial (VA-HIT), patients with established IHD and low HDL cholesterol levels who were randomized to receive gemfibrozil experienced a significant reduction in major cardiovascular events and cardiovascular mortality, as compared with patients who received placebo.[90] In another study, patients with established IHD and low levels of HDL cholesterol who were randomized to receive simvastatin plus niacin had fewer first cardiovascular events, as compared with patients who received placebo or simvastatin alone.[91]

Ezetimibe represents a new class of agent that inhibits intestinal absorption of cholesterol and produces moderate reductions in LDL cholesterol. Its principal indication at this time, as with the bile-acid sequestrants, is to augment the lipid-lowering efficacy of a statin.[93] Ezetimibe is associated with fewer gastrointestinal side effects than bile-acid sequestrants. Niacin commonly causes flushing, a side effect that can be mitigated by gradual titration toward a target dose and by pretreatment with aspirin. Extended-release preparations of niacin are associated with less flushing. Niacin modestly increases glucose intolerance and can cause hyperuricemia. As with statins, niacin and fibric acid derivatives are associated with a risk of elevations in transaminase

levels, as well as with infrequent hepatitis and rare myositis. Although the manufacturers of these drugs recommend routine laboratory evaluation of liver function, routine monitoring was shown to have a low yield in a primary care practice.[94]

Reduction of non-LDL cholesterol lipid fractions may also reduce cardiovascular risk, particularly in patients who have the metabolic syndrome. Although definitions vary, the metabolic syndrome is a constellation of cardiovascular risk factors, including insulin resistance, obesity, hypertension, and dyslipidemia.[95] The dyslipidemia characteristic of the metabolic syndrome consists of elevated triglycerides, a low HDL cholesterol level, and a normal (or near-normal) LDL cholesterol level. Retrospective analysis of the VA-HIT, which studied patients with established coronary disease and low HDL cholesterol levels, suggested that the benefits of treatment with gemfibrozil are most pronounced in patients with insulin resistance (whose fasting plasma insulin levels were comparable to those found in patients with the metabolic syndrome).[96] Because patients with the metabolic syndrome have elevated cardiovascular risk despite often having unremarkable LDL cholesterol levels, NCEP ATP III recommends measuring non-HDL cholesterol in patients with elevated triglyceride levels (i.e., triglyceride levels ≥ 200 mg/dl or 2.25 mmol/L) through use of the following formula:

Non-HDL cholesterol = measured total cholesterol – HDL cholesterol

The level of non-HDL cholesterol determined by this formula corresponds to the sum of LDL cholesterol and atherogenic remnant lipoproteins containing apolipoprotein B (very low density lipoprotein cholesterol). Among patients with hypertriglyceridemia, the target level for non-HDL cholesterol is less than 130 mg/dl for patients with IHD or IHD equivalent conditions, as well as for patients with two or more cardiovascular risk factors and a 10-year risk of cardiovascular events greater than 20%, as determined by the Framingham risk estimates [*see Figure 2*].[30]

In summary, numerous studies demonstrate that patients with IHD benefit from treatment with statins; this includes IHD patients with relatively normal LDL cholesterol levels. The vast majority of patients with chronic stable angina should be treated with a statin. Treatment with niacin or fibric acid derivatives should be considered in patients with a low HDL cholesterol level or an elevated triglyceride level. In view of the consistent benefits of statins across many patient subsets, it would be most reasonable to consider adding niacin or fibric acid derivatives to statin therapy.

Antihypertensive Therapy

Hypertension contributes to cardiovascular risk by increasing myocardial wall stress, oxygen demand, and endothelial injury. Treatment of hypertension in patients with IHD reduces the risk of future cardiovascular events.[97] A reduction in systolic blood pressure of 2 mm Hg is associated with a 7% reduction in mortality from IHD.[98] By lowering myocardial oxygen demand, treatment of hypertension may also improve anginal symptoms. The therapeutic target for patients with IHD is to maintain blood pressure at levels below 140/90 mm Hg.[99]

Two groups of patients with hypertension and chronic stable angina warrant particular consideration: patients with specific coexisting chronic conditions (e.g., diabetes, heart failure, or renal insufficiency) and patients with LVH. Guidelines from the Joint National Committee for the Diagnosis, Evaluation, and Treatment of Hypertension (JNC) recommend a lower therapeutic target blood pressure for hypertensive patients who have diabetes, heart failure, or renal insufficiency—namely, a level below 130/85 mm Hg.[99] Although no specific target has been promulgated for therapy for hypertension and LVH, the latter is a marker for the severity and chronicity of hypertension and is a risk factor for MI, CHF, and cardiac sudden death.[100] Treatment of hypertension results in the regression of LVH; ECG evidence of LVH regression is associated with a significantly reduced risk of cardiovascular events.[101] Patients with LVH should therefore be targeted for aggressive antihypertensive therapy.

Two classes of antihypertensives—beta-adrenergic receptor antagonists (i.e., beta blockers) and calcium channel blockers—are also effective antianginal medications. Beta blockers confer mortality benefit in patients after MI[102] and are recommended as first-line therapy for most patients with chronic stable angina.[32] Calcium channel blockers (e.g., diltiazem, verapamil, and long-acting dihydropyridines) are also effective antihypertensive and antianginal medications. Current ACC/AHA/ACP guidelines recommend beta blockers as first-line antihypertensive therapy in patients with chronic stable angina. Long-acting calcium channel blockers are an acceptable alternative, particularly in patients without a history of MI.[102]

ACE Inhibitors

Results of the Heart Outcomes Prevention Evaluation (HOPE) trial demonstrated that ACE inhibitors reduced MI, stroke, and cardiovascular death in patients at high cardiovascular risk. In this study, patients with a history of IHD, diabetes, stroke, or peripheral vascular disease and at least one additional cardiovascular risk factor (e.g., hypertension, dyslipidemia, cigarette smoking, or microalbuminuria) were randomized to receive either ramipril (10 mg daily) or placebo; patients were followed for an average of 4 years. Total mortality was reduced 1.8% in the ramipril-treated group; in terms of numbers needed to treat (NNT), this means that 56 patients would need to be treated for 4 years to prevent one death. The primary outcome of MI, stroke, or cardiovascular death was reduced by 3.8% (NNT of 26). This study indicated that a broad range of patients at high risk for IHD who had normal left ventricular function obtained an impressive survival benefit from the use of ACE inhibitors. The magnitude of benefit was greater than might have been expected for the small decrement in average blood pressure observed in the study, suggesting a mechanism at work other than reduction in blood pressure.[103] In the European Trial on Reduction of Cardiac Events with Perindopril in Stable Coronary Artery Disease (EUROPA), patients with stable coronary disease and no known congestive heart failure or uncontrolled hypertension were randomized to receive either the ACE inhibitor perindopril or placebo.[104] Combined cardiovascular end points were significantly reduced in the perindopril group; in addition, there was a nonsignificant reduction in total mortality. On the basis of the HOPE and EUROPA trials, it can be concluded that most patients with chronic stable angina should be treated with an ACE inhibitor, barring renal insufficiency, hyperkalemia, or an allergy to ACE inhibitors. Angiotensin receptor blockers may offer a similar benefit, but these agents have not been extensively studied in this regard.

MEDICAL THERAPY FOR ANGINAL SYMPTOMS

The major classes of medications for the treatment of angina include beta blockers, calcium channel blockers, and nitroglycerin/nitrates. Randomized trials demonstrate that beta blockers

and calcium channel blockers are equally effective in relieving angina and improving exercise tolerance.[105,106] Current guidelines, however, recommend beta blockers as first-line therapy[32,33] because they improve survival and reduce cardiac events in patients who have had a previous acute MI[107] and in elderly patients with systolic hypertension[108]; no similar benefits have been demonstrated for calcium channel blockers or nitrates. In addition, beta blockers improve survival and reduce the risk of stroke and congestive heart failure in patients with hypertension.[109] Beta blockers should be considered as initial antianginal therapy in patients with chronic stable angina.

Beta Blockers

Beta blockers decrease heart rate, myocardial contractility, blood pressure, and myocardial oxygen demand by inhibiting cardiac and peripheral beta-adrenergic receptors. Beta blockers delay the onset of angina and increase exercise capacity in patients with exertional angina.[110,111] They are titrated to a dose adequate to reduce the resting heart rate to 55 to 60 beats/min. Titration to lower heart rates may be necessary in patients with more severe angina, provided patients do not develop heart block or symptoms of severe bradycardia.

Beta blockers are generally well tolerated by patients with chronic obstructive pulmonary disease; however, they may exacerbate bronchospasm in patients with severe asthma. Beta blockers are well tolerated in patients with diabetes, and in these patients, they can reduce macrovascular events[112]; theoretically, however, beta blockers can mask the adrenergically mediated symptoms of hypoglycemia insulin (e.g., tachycardia).

Beta blockers are contraindicated in the presence of severe bradycardia, high-degree atrioventricular block, sinus node dysfunction, and uncompensated congestive heart failure. Patients with extensive peripheral vascular disease and claudication may experience worsening of their symptoms. Beta blockers are also contraindicated in the small subset of patients with pure variant or vasospastic angina (i.e., angina occurring in the absence of fixed obstruction of the coronary arteries), in whom beta blockade is unlikely to alleviate symptoms. In these patients, beta blockers may actually worsen angina as a result of unopposed alpha-adrenergic effects. Calcium channel blockers are the preferred first-line agent in this patient group [*see* Calcium Channel Blockers, *below*].

Calcium Channel Blockers

Calcium channel blockers reduce smooth muscle tone and cause coronary and peripheral vasodilatation, improving coronary blood flow and reducing peripheral vascular resistance. Calcium channel blockers can be used as monotherapy in the treatment of chronic stable angina, although combinations of beta blockers and calcium channel blockers relieve angina more effectively than either agent alone. Combination therapy with a beta blocker may blunt the reflex tachycardia that can occur with dihydropyridine calcium channel antagonists. All calcium channel blockers exert some negative inotropic effect, although this effect is typically most significant clinically with the nondihydropyridine agents verapamil and diltiazem. Calcium channel blockers are contraindicated in the presence of decompensated congestive heart failure, although the vasoselective dihydropyridine agents amlodipine and felodipine are tolerated in patients with clinically stable LV dysfunction.[113] Verapamil and diltiazem have a pronounced effect on heart rate and conduction; they should be used with caution in combination with beta blockers because of the increased risk of heart block associated with the combined use of these agents. Constipation and peripheral edema are common side effects of calcium channel blockers.

Nitrates and Nitroglycerin

Nitrates and nitroglycerin dilate coronary arteries and their collateral vessels, directly improve myocardial perfusion, diminish afterload, and increase venous capacitance. These agents exert antianginal effects by improving coronary blood flow and by reducing myocardial oxygen demand. Long-acting nitrate preparations, in tablet or patch form, reduce the severity and frequency of angina and improve exercise tolerance; however, they often induce a reflexive increase in sympathetic tone and increase heart rate. Therefore, long-acting nitrates are often used in combination with beta blockers or calcium channel blockers. Short-acting nitroglycerin tablets or spray is appropriate for the immediate relief of exercise-induced or rest angina. They may also prevent angina when taken several minutes before exertion sufficient to cause angina.

Nitroglycerin and nitrates should not be used within 24 hours of taking sildenafil (Viagra) or other phosphodiesterase inhibitors used in the treatment of erectile dysfunction, because of the potential for life-threatening hypotension.[114] It is important to discuss this interaction with patients taking nitrates or sildenafil. Nitroglycerin and nitrates are relatively contraindicated in patients with severe aortic stenosis or hypertrophic obstructive cardiomyopathy because of an increased risk of syncope resulting from diminished cardiac output. Continued use of long-acting nitrates results in tachyphylaxis; the mechanism of this is unclear. An adequate nitrate-free period (8 to 12 hours each day) is necessary to minimize this effect. Headaches are common and often limit nitroglycerin and nitrate therapy; with continued use, headaches will diminish in up to 80% of patients. Hypotension may occur, particularly in hypovolemic patients.

In general, patients who are found to be at low or moderate risk for cardiovascular complications during risk stratification should be treated aggressively with medical therapy. Medical therapy should not be considered to have failed until the patient has been treated with full therapeutic doses of a beta blocker, a calcium channel blocker, and a long-acting nitrate and continues to experience angina or develops unacceptable adverse effects.

PATIENTS WITH DIABETES MELLITUS

IHD therapy for patients with diabetes merits special consideration. Cardiovascular events are the leading cause of death in patients with diabetes, and this patient group is at particularly high risk for MI. Middle-aged persons with diabetes and no history of MI have a risk of MI and cardiac death equivalent to that of nondiabetic patients with a history of MI.[115] A substudy of the Heart Protection Study demonstrated that treatment with statins reduced major coronary and major vascular events, even among diabetic patients without a prior diagnosis of ischemic heart disease and among diabetic patients with LDL cholesterol levels lower than 116 mg/dl (3 mmol/L).[116] Treatment with statins resulted in a relative risk reduction for major coronary events (nonfatal MI and coronary death [27%]), stroke (25%), and first revascularization procedures (17%). A HOPE substudy demonstrated the benefits of ACE inhibitors in patients with diabetes and one or more cardiovascular risk factors.[117] Treatment with an ACE inhibitor resulted in a 25% reduction in MI, stroke, or cardiovascular death. Total mortality was reduced 24%; progression to overt nephropathy was reduced by 24%. In the United Kingdom

Prospective Diabetes Study, beta blockers were found to be equivalent to ACE inhibitors in reducing the risk of macrovascular complications.[112] Another study found that cardiovascular events were more common in patients with type 2 diabetes mellitus and hypertension who were randomized to receive a dihydropyridine calcium channel blocker, as compared with patients who received an ACE inhibitor.[118] It remains uncertain whether to attribute this result to a higher risk of cardiovascular events among the group taking the calcium channel blocker amlodipine or to a reduced risk of cardiovascular events among the group taking the ACE inhibitor fosinopril. Nevertheless, most experts consider calcium channel blockers to be third-line agents for patients with diabetes and hypertension. Results of the Hypertension Optimal Treatment (HOT) study support aggressive blood pressure reduction in patients with diabetes; the target blood pressure for these patients is 135/80 mm Hg or lower.[119]

Although it remains uncertain when to initiate antiplatelet therapy in diabetic patients, it is reasonable to treat diabetic patients who have any additional cardiovascular risk factor with aspirin or an equivalent antiplatelet therapy.[120] The American Diabetes Association and the NCEP recommend a target LDL cholesterol level of 100 mg/dl (2.59 mmol/L) or lower.[31,121]

REVASCULARIZATION

Techniques for revascularization include PCI, using catheter-based methods with or without placement of intracoronary stents, and CABG. For most patients with angina, survival with optimal medical therapy is equivalent to that resulting from revascularization; in addition, CABG results in excellent symptom relief. For a select few patients with chronic stable angina, revascularization is associated with improved survival, as compared with that achieved with medical therapy.[122] Among 2,649 patients with left main coronary stenoses or multivessel coronary artery disease and reduced LV systolic function, those who were treated with CABG in randomized trials had an absolute mortality at 5 years that was more than 5% lower than that of patients assigned to medical management (10.2% versus 15.8%).[123] This benefit persisted at 10 years' follow-up. There is much weaker evidence to suggest that patients with proximal stenoses of the left anterior descending coronary artery and normal LV function also experience lower mortality with CABG. One randomized trial showed no difference in mortality among patients assigned to CABG, PCI, or medical management.[124] In two trials comparing CABG with PCI, survival was equivalent; however, the patients who underwent surgery had fewer symptoms and required fewer antianginal medications and subsequent revascularization procedures.[125,126] Initial costs and short-term (procedure-related) mortality were higher in patients who underwent CABG. Most of the patients enrolled had two-vessel CAD and normal LV systolic function. In one study, survival was improved in patients with diabetes who underwent CABG, as compared with patients who underwent PCI.[125]

Randomized trials comparing PCI with medical management in patients with one- or two-vessel CAD and normal LV systolic function have demonstrated equivalent survival.[127,128] Relief of symptoms was generally greater with PCI than that seen with medical therapy, although PCI was associated with an increased risk of procedure-related MI and death[127,128] and substantially greater cost.[129]

On the basis of a limited number of randomized trials, it appears that only the subgroup of patients with severe coronary disease (defined as two- or three-vessel disease) and impaired LV function can confidently expect improved average survival after revascularization and that improvement is seen only with CABG. Thus, evaluation for revascularization is generally recommended for patients who are at moderate or high risk of death and who are willing to undergo a revascularization procedure.[32] Patients meeting these criteria who are found to have either extensive areas of ischemia on noninvasive testing or reduced LV systolic function should then be considered candidates for angiography. It should be recognized, however, that the results of currently available studies comparing PCI, CABG, and medical therapy do not reflect recent advancements in all three forms of treatment. For example, restenosis rates with PCI using stents, drug-eluting stents, and platelet inhibitors are lower than previously reported rates with angioplasty techniques.

In general, medical therapy is preferred in patients who are determined through risk stratification by noninvasive testing to be at low risk for death. There is no evidence that revascularization improves survival in such patients. For this reason, currently available evidence suggests that many revascularization procedures conducted in the United States may not be warranted.

COMPLICATIONS

Chronic stable angina, as the name suggests, is a long-standing stable or progressive condition caused, in most instances, by atherosclerotic narrowing of the coronary vessels. Coronary atherosclerosis, however, is a risk factor for acute coronary syndromes, including MI and sudden cardiac death. Much of the diagnostic evaluation of patients with chronic stable angina is undertaken to determine the intermediate-term risk of acute coronary events. Much of the treatment of patients with chronic stable angina is intended to reduce the risk of acute coronary complications, principally MI and sudden cardiac death. Another major complication of chronic stable angina is CHF. Major causes of CHF include MI and hypertension—common conditions in patients with IHD.[130] Finally, patients with IHD are at risk for other vascular diseases, namely stroke and peripheral vascular disease.

It is also important to consider complications of conditions that predispose patients to IHD. For example, diabetes is a major risk factor for IHD and is often complicated by peripheral vascular disease, renal insufficiency, neuropathy, and retinopathy. In addition to increasing the risk of IHD, cigarette smoking confers a significant risk of chronic pulmonary disease, cancer, and peripheral vascular disease.

FOLLOW-UP

Patients with chronic stable angina should be regularly followed in a primary care setting. There is little evidence to recommend a particular frequency of follow-up visits, although ACC/AHA/ACP guidelines suggest regular visits at 4- to 12-month intervals for patients with chronic stable angina.[21] The expert panel recommends addressing the following questions at each visit[21]:

- Has there been a change in the level of activity since the last visit?
- Have anginal symptoms increased in severity or frequency?
- Is current therapy well tolerated?
- Has the patient been successful at modifying cardiac risk factors?
- Has the patient developed new or worsening comorbid illnesses that may have an impact on the patient's angina?

It is important to inquire about changes in anginal symptoms or activity levels to identify patients who require increased intensity of antianginal therapy or further risk stratification. For example, it would be reasonable to repeat noninvasive testing in a patient with stable class II angina who developed new class III anginal symptoms since the last clinic visit. Similarly, in a patient with new symptoms of CHF, it would be appropriate to perform echocardiography and consider referral for coronary angiography; multiple-vessel CAD with reduced left ventricular function would be an indication for CABG.

Medical therapy should be reviewed at each visit to assess adherence to recommended therapy, knowledge about doses and indications, and potential side effects. In addition, it is worth considering whether recent evidence for benefit from new therapies or new indications for existing therapies support a modification of IHD management. Patients with chronic stable angina should be encouraged to quit smoking, to eat a prudent diet, and to regularly engage in moderate exercise. Successful adoption of these interventions is challenging, but repeated encouragement from a personal physician enhances success.[131]

Vital signs (i.e., heart rate, blood pressure, and weight) should be regularly followed. The physical examination is focused on the heart, lungs, and vasculature. Findings of particular note include signs of CHF, new or changing heart murmurs, arrhythmias, or evidence of carotid or peripheral vascular disease.

Laboratory assessment should include periodic measurements of fasting lipid levels. Regular measurements of liver transaminase and CK levels are not recommended in the absence of symptoms. Other laboratory testing is indicated by comorbid conditions or changes in the patient's history and physical examination.

There is no evidence showing that regular ECG studies are helpful in the management of patients with chronic stable angina in the absence of changes in history or physical examination. ECGs are indicated when new medications are introduced that may affect cardiac conduction. Changes in anginal or syncopal symptoms and findings suggestive of dysrhythmia or conduction abnormalities should also prompt a repeat ECG.

There is little evidence to guide the use of repeat stress testing in patients with chronic stable angina. Recommendations for follow-up stress testing vary according to initial assessments of a patient's cardiovascular risk.[21] For example, patients with class II angina whose exercise treadmill test places them at low risk have an annual risk of mortality of less than 1%; these patients do not require follow-up stress testing for a period of 3 to 4 years in the absence of new and concerning symptoms or signs. Similarly, patients who underwent PCI more than 6 months earlier and who have minimal residual stenosis are unlikely to benefit from regular stress testing. It is not known whether patients at intermediate or high cardiovascular risk benefit from periodic stress testing.

Follow-up noninvasive tests are selected according to the approach outlined (see above). When possible, the same form of stress (exercise or pharmacologic) and testing (ECG or imaging) should be repeated, because this permits the most valid comparison with the original study.[21]

Patients should be referred to a cardiologist when appropriate for consideration of revascularization. Candidates for referral include patients with valvular disorders that require repair, patients with angina that is refractory to maximal medical therapy, and patients with comorbidities that complicate therapy.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Gibbons RJ, Balady GJ, Beasley JW, et al: ACC/AHA Guidelines for Exercise Testing. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing). J Am Coll Cardiol 30:260, 1997

2. 1997 Biostatistical Fact Sheets. American Heart Association, Dallas, 1997, p 1

3. 1999 Heart and Stroke Statistical Update. American Heart Association, Dallas, 1999, p 1

4. Five-year clinical and functional outcome comparing bypass surgery and angioplasty in patients with multivessel coronary disease: a multicenter randomized trial. Writing Group for the Bypass Angioplasty Revascularization Investigation (BARI) Investigators. JAMA 277:715, 1997

5. Grundy SM, Pasternak R, Greenland P, et al: Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations: a statement for healthcare professionals from the American Heart Association and the American College of Cardiology. Circulation 100:1481, 1999

6. Pasternak RC, Grundy SM, Levy D, et al: 27th Bethesda Conference: matching the intensity of risk factor management with the hazard for coronary disease events. Task Force 3. Spectrum of risk factors for coronary heart disease. J Am Coll Cardiol 27:978, 1996

7. Isomaa B, Almgren P, Tuomi T, et al: Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care 24:683, 2001

8. Scott CL: Diagnosis, prevention, and intervention for the metabolic syndrome. Am J Cardiol 92:35i, 2003

9. Lee PY, Alexander KP, Hammill BG, et al: Representation of elderly persons and women in published randomized trials of acute coronary syndromes. JAMA 286:708, 2001

10. Greenland P, Knoll MD, Stamler J, et al: Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. JAMA 290:891, 2003

11. Khot UN, Khot MB, Bajzer CT, et al: Prevalence of conventional risk factors in patients with coronary heart disease. JAMA 290:898, 2003

12. Naghavi M, Libby P, Falk E, et al: From vulnerable plaque to vulnerable patient: a call for new definitions and risk assessment strategies: Part II. Circulation 108:1772, 2003

13. Koenig W, Sund M, Frohlich M, et al: C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results from the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. Circulation 99:237, 1999

14. Danesh J, Wheeler JG, Hirschfield GM, et al: C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N Engl J Med 350:1387, 2004

15. Ross R: Atherosclerosis: an inflammatory disease. N Engl J Med 340:115, 1999

16. Morrow DA, Rifai N, Antman EM, et al: C-reactive protein is a potent predictor of mortality independently of and in combination with troponin T in acute coronary syndromes: a TIMI 11A substudy. Thrombolysis in Myocardial Infarction. J Am Coll Cardiol 31:1460, 1998

17. Kennon S, Price CP, Mills PG, et al: Cumulative risk assessment in unstable angina: clinical, electrocardiographic, autonomic, and biochemical markers. Heart 89:36, 2003

18. Ambrose JA, Tannenbaum MA, Alexopoulos D, et al: Angiographic progression of coronary artery disease and the development of myocardial infarction. J Am Coll Cardiol 12:56, 1988

19. Campeau L: Letter: grading of angina pectoris. Circulation 54:522, 1976S

20. Braunwald E, Antman EM, Beasley JW, et al: ACC/AHA 2002 guideline update for the management of patients with unstable angina and non-ST-segment elevation myocardial infarction: summary article: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (Committee on the Management of Patients with Unstable Angina). J Am Coll Cardiol 40:1366, 2002

21. Gibbons RJ, Chatterjee K, Daley J, et al: ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: executive summary and recommendations. A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients with Chronic Stable Angina). Circulation 99:2829, 1999

22. Pryor DB, Shaw L, McCants CB, et al: Value of the history and physical in identifying patients at increased risk for coronary artery disease. Ann Intern Med 118:81, 1993

23. Chun AA, McGee SR: Bedside diagnosis of coronary artery disease: a systematic review. Am J Med 117:334, 2004

24. Swan HJ, Gersh BJ, Graboys TB, et al: 27th Bethesda Conference: matching the intensity of risk factor management with the hazard for coronary disease events. Task Force 7. Evaluation and management of risk factors for the individual patient (case management). J Am Coll Cardiol 27:1030, 1996

25. Wilson PW, D'Agostino RB, Levy D, et al: Prediction of coronary heart disease using risk factor categories. Circulation 97:1837, 1998

26. Levine HJ: Difficult problems in the diagnosis of chest pain. Am Heart J 100:108, 1980

27. Connolly DC, Elveback LR, Oxman HA: Coronary heart disease in residents of Rochester, Minnesota: IV. Prognostic value of the resting electrocardiogram at the time of initial diagnosis of angina pectoris. Mayo Clin Proc 59:247, 1984

28. Levy D, Garrison RJ, Savage DD, et al: Prognostic implications of echocardiographically determined left ventricular mass in the Framingham Heart Study. N Engl J Med 322:1561, 1990

29. Fisch C: Electrocardiography and vectorcardiography. Heart Disease. Braunwald E, Ed. WB Saunders Co, Philadelphia, 1992, p 145

30. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood

Cholesterol in Adults (Adult Treatment Panel III). Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. JAMA 285:2486, 2001

31. Tape TG, Mushlin AI: The utility of routine chest radiographs. Ann Intern Med 104:663, 1986

32. Gibbons RJ, Chatterjee K, Daley J, et al: ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients With Chronic Stable Angina). J Am Coll Cardiol 33:2092, 1999 [erratum, J Am Coll Cardiol 34:314, 1999]

33. Gibbons RJ, Abrams J, Chatterjee K, et al: ACC/AHA 2002 guideline update for the management of patients with chronic stable angina: summary article: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on the Management of Patients with Chronic Stable Angina). J Am Coll Cardiol 41:159, 2003

34. Stuart RJ Jr, Ellestad MH: National survey of exercise stress testing facilities. Chest 77:94, 1980

35. Klocke FJ, Baird MG, Lorell BH, et al: ACC/AHA/ASNC guidelines for the clinical use of cardiac radionuclide imaging: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/ASNC Committee to Revise the 1995 Guidelines for the Clinical Use of Cardiac Radionuclide Imaging). J Am Coll Cardiol 42:1318, 2003

36. Shaw LJ, Hachamovitch R, Peterson ED, et al: Using an outcomes-based approach to identify candidates for risk stratification after exercise treadmill testing. J Gen Intern Med 14:1, 1999

37. Cheitlin MD, Alpert JS, Armstrong WF, et al: ACC/AHA guidelines for the clinical application of echocardiography: executive summary. A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (Committee on Clinical Application of Echocardiography). Developed in collaboration with the American Society of Echocardiography. J Am Coll Cardiol 29:862, 1997

38. Fleischmann KE, Hunink MG, Kuntz KM, et al: Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA 280:913, 1998

39. Amanullah AM, Kiat H, Friedman JD, et al: Adenosine technetium-99m sestamibi myocardial perfusion SPECT in women: diagnostic efficacy in detection of coronary artery disease. J Am Coll Cardiol 27:803, 1996

40. Kim C, Kwok YS, Saha S, et al: Diagnosis of suspected coronary artery disease in women: a cost-effectiveness analysis. Am Heart J 137:1019, 1999

41. Gauri AJ, Raxwal VK, Roux L, et al: Effects of chronotropic incompetence and beta-blocker use on the exercise treadmill test in men. Am Heart J 142:136, 2001

42. Ladenheim ML, Kotler TS, Pollock BH, et al: Incremental prognostic power of clinical history, exercise electrocardiography and myocardial perfusion scintigraphy in suspected coronary artery disease. Am J Cardiol 59:270, 1987

43. Bielak LF, Rumberger JA, Sheedy PF 2nd, et al: Probabilistic model for prediction of angiographically defined obstructive coronary artery disease using electron beam computed tomography calcium score strata. Circulation 102:380, 2000

44. Redberg RF: Coronary artery calcium and cardiac events. Circulation 108:E167, 2003

45. O'Malley PG, Taylor AJ: Prognostic value of coronary artery calcification. Circulation 108:E169, 2003

46. Detrano RC, Wong ND, Doherty TM, et al: Coronary calcium does not accurately predict near-term future coronary events in high-risk adults. Circulation 99:2633, 1999

47. Mark DB, Shaw LJ, Lauer MS, et al: 34th Bethesda Conference: task force #5: is atherosclerosis imaging cost effective? J Am Coll Cardiol 41:1906, 2003

48. O'Rourke RA, Brundage BH, Froelicher VF, et al: American College of Cardiology/American Heart Association Expert Consensus Document on electron-beam computed tomography for the diagnosis and prognosis of coronary artery disease. J Am Coll Cardiol 36:326, 2000

49. Kussmaul WG 3rd, Popp RL, Norcini J: Accuracy and reproducibility of visual coronary stenosis estimates using information from multiple observers. Clin Cardiol 15:154, 1992

50. Hubbard BL, Gibbons RJ, Lapeyre AC 3rd, et al: Identification of severe coronary artery disease using simple clinical parameters. Arch Intern Med 152:309, 1992

51. Pryor DB, Shaw L, Harrell FE Jr, et al: Estimating the likelihood of severe coronary artery disease. Am J Med 90:553, 1991

52. Patterson RE, Eisner RL, Horowitz SF: Comparison of cost-effectiveness and utility of exercise ECG, single photon emission computed tomography, positron emission tomography, and coronary angiography for diagnosis of coronary artery disease. Circulation 91:54, 1995

53. Mock MB, Ringqvist I, Fisher LD, et al: Survival of medically treated patients in the coronary artery surgery study (CASS) registry. Circulation 66:562, 1982

54. O'Keefe JH Jr, Zinsmeister AR, Gibbons RJ: Value of normal electrocardiographic findings in predicting resting left ventricular function in patients with chest pain and suspected coronary artery disease. Am J Med 86:658, 1989

55. Mark DB, Shaw L, Harrell FE Jr, et al: Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. N Engl J Med 325:849, 1991

56. Cole CR, Blackstone EH, Pashkow FJ, et al: Heart-rate recovery immediately after exercise as a predictor of mortality. N Engl J Med 341:1351, 1999

57. Gibbons RJ, Balady GJ, Bricker JT, et al: ACC/AHA 2002 guideline update for exercise testing: summary article. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines). J Am Coll Cardiol 40:1531, 2002

58. Brown KA: Prognostic value of thallium-201 myocardial perfusion imaging; a diagnostic tool comes of age. Circulation 83:363, 1991

59. Boucher CA, Zir LM, Beller GA, et al: Increased lung uptake of thallium-201 during exercise myocardial imaging: clinical, hemodynamic and angiographic implications in patients with coronary artery disease. Am J Cardiol 46:189, 1980

60. Cox JL, Wright LM, Burns RJ: Prognostic significance of increased thallium-201 lung uptake during dipyridamole myocardial scintigraphy: comparison with exercise scintigraphy. Can J Cardiol 11:689, 1995

61. McCully RB, Roger VL, Mahoney DW, et al: Outcome after normal exercise echocardiography and predictors of subsequent cardiac events: follow-up of 1,325 patients. J Am Coll Cardiol 31:144, 1998

62. Marwick TH: Use of stress echocardiography for the prognostic assessment of patients with stable chronic coronary artery disease. Eur Heart J 18(suppl D):D97, 1997

63. Emond M, Mock MB, Davis KB, et al: Long-term survival of medically treated patients in the Coronary Artery Surgery Study (CASS) Registry. Circulation 90:2645, 1994

64. Mark DB, Nelson CL, Califf RM, et al: Continuing evolution of therapy for coronary artery disease: initial results from the era of coronary angioplasty. Circulation 89:2015, 1994

65. Critchley J, Capewell S: Smoking cessation for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev (4):CD003041, 2003

66. Sandvik L, Erikssen J, Thaulow E, et al: Physical fitness as a predictor of mortality among healthy, middle-aged Norwegian men. N Engl J Med 328:533, 1993

67. Bijnen FC, Caspersen CJ, Feskens EJ, et al: Physical activity and 10-year mortality from cardiovascular diseases and all causes: the Zutphen Elderly Study. Arch Intern Med 158:1499, 1998

68. Tanasescu M, Leitzmann MF, Rimm EB, et al: Exercise type and intensity in relation to coronary heart disease in men. JAMA 288:1994, 2000

69. de Lorgeril M, Salen P, Martin JL, et al: Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. Circulation 99:779, 1999

70. Singh RB, Dubnov G, Niaz MA, et al: Effect of an Indo-Mediterranean diet on progression of coronary artery disease in high risk patients (Indo-Mediterranean Diet Heart Study): a randomised single-blind trial. Lancet 360:1455, 2002

71. Joshipura KJ, Hu FB, Manson JE, et al: The effect of fruit and vegetable intake on risk for coronary heart disease. Ann Intern Med 134:1106, 2001

72. Hu FB, Willett WC: Optimal diets for prevention of coronary heart disease. JAMA 288:2569, 2002

73. Collaborative overview of randomised trials of antiplatelet therapy: I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. Antiplatelet Trialists' Collaboration. BMJ 308:81, 1994 [erratum BMJ 308:1540, 1994]

74. Ridker PM, Manson JE, Gaziano JM, et al: Low-dose aspirin therapy for chronic stable angina: a randomized, placebo-controlled clinical trial. Ann Intern Med 114:835, 1991

75. Double-blind trial of aspirin in primary prevention of myocardial infarction in patients with stable chronic angina pectoris. The Swedish Angina Pectoris Aspirin Trial (SAPAT) Group. Lancet 340:1421, 1992

76. Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. N Engl J Med 321:129, 1989

77. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. Antithrombotic Trialists' Collaboration. BMJ 324:71, 2002

78. A randomised, blinded trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. Lancet 348:1329, 1996

79. Gaspoz JM, Coxson PG, Goldman PA, et al: Cost effectiveness of aspirin, clopidogrel, or both for secondary prevention of coronary heart disease. N Engl J Med 346:1800, 2002

80. The cholesterol facts. A summary of the evidence relating dietary fats, serum cholesterol, and coronary heart disease. A joint statement by the American Heart Association and the National Heart, Lung, and Blood Institute. The Task Force on Cholesterol Issues, American Heart Association. Circulation 81:1721, 1990

81. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 344:1383, 1994

82. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. N Engl J Med 335:1001, 1996

83. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Heart Protection Study. Lancet 360:7, 2002

84. Aggressive lipid-lowering therapy compared with angioplasty in stable coronary artery disease. Atorvastatin versus Revascularization Treatment Investigators. N Engl J Med 341:70, 1999

85. Nissen SE, Tuzcu M, Schoenhagen P, et al: Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. JAMA 291:1071, 2004

86. Rosenson RS: Lipid lowering with HMG CoA reductase inhibitors (statins). UpToDate Online, 2004
http://www.uptodate.com

87. Thompson PD, Clarkson P, Karas RH: Statin-associated myopathy. JAMA 289:1681, 2003

88. Phillips PS, Haas RH, Bannykh S, et al: Statin-associated myopathy with normal creatine kinase levels. Ann Intern Med 137:581, 2002

89. Gould AL, Rossouw JE, Santanello NC, et al: Cholesterol reduction yields clinical benefit: impact of statin trials. Circulation 97:946, 1998

90. Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial Study Group. N Engl J Med 341:410, 1999

91. Brown BG, Zhao XQ, Chait A, et al: Simvastatin and niacin, antioxidant vitamins, or the combination for the prevention of coronary disease. N Engl J Med 345:1583, 2001

92. Bays HE, McGovern ME: Once-daily niacin extended release/lovastatin combination tablet has more favorable effects on lipoprotein particle size and subclass distribution than atorvastatin and simvastatin. Prev Cardiol 6:179, 2003

93. Ballantyne CM, Houri J, Notarbolo A, et al: Effect of ezetimibe coadministered with atorvastatin in 628 patients with primary hypercholesterolemia: a prospective, randomized, double-blind trial. Circulation 107:2409, 2003

94. Smith CC, Bernstein LI, Davis RB, et al: Screening for statin-related toxicity: the yield of transaminase and creatine kinase measurements in a primary care setting. Arch Intern Med 163:688, 2003

95. Wilson PW, Grundy SM: The metabolic syndrome: a practical guide to origins and treatment: Part II. Circulation 108:1537, 2003

96. Rubins HB, Robins SJ, Collins D, et al: Diabetes, plasma insulin, and cardiovascular disease: subgroup analysis from the Department of Veterans Affairs High-Density Lipoprotein Intervention Trial (VA-HIT). Arch Intern Med 162:2597, 2002

97. Smith SC Jr, Blair SN, Criqui MH, et al: AHA consensus panel statement: preventing heart attack and death in patients with coronary disease. The Secondary Prevention Panel. J Am Coll Cardiol 26:292, 1995

98. Lewington S, Clarke R, Qizilbash N, et al: Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 360:1903, 2002

99. Chobanian AV, Bakris GL, Black HR, et al: The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 289:2560, 2003

100. Kannel WB, Gordon T, Offutt D: Left ventricular hypertrophy by electrocardiogram: prevalence, incidence, and mortality in the Framingham study. Ann Intern Med 71:89, 1969

101. Levy D, Salomon M, D'Agostino RB, et al: Prognostic implications of baseline electrocardiographic features and their serial changes in subjects with left ventricular hypertrophy. Circulation 90:1786, 1994

102. Heidenreich PA, McDonald KM, Hastie T, et al: Meta-analysis of trials comparing beta-blockers, calcium antagonists, and nitrates for stable angina. JAMA 281:1927, 1999

103. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000 [errata, N Engl J Med 342:748, 2000, and N Engl J Med 342:1376, 2000]

104. Fox KM: Efficacy of perindopril in reduction of cardiovascular events among patients with stable coronary artery disease: randomised, double-blind, placebo-controlled, multicentre trial (the EUROPA study). Lancet 362:782, 2003

105. Dargie HJ, Ford I, Fox KM: Total Ischaemic Burden European Trial (TIBET). Effects of ischaemia and treatment with atenolol, nifedipine SR and their combination on outcome in patients with chronic stable angina. The TIBET Study Group. Eur Heart J 17:104, 1996

106. Rehnqvist N, Hjemdahl P, Billing E, et al: Effects of metoprolol vs verapamil in patients with stable angina pectoris. The Angina Prognosis Study in Stockholm (APSIS). Eur Heart J 17:76, 1996

107. The beta-blocker heart attack trial. Beta-Blocker Heart Attack Study Group. JAMA 246:2073, 1981

108. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension: final results of the Systolic Hypertension in the Elderly Program (SHEP). SHEP Cooperative Research Group. JAMA 265:3255, 1991

109. MacMahon S, Peto R, Cutler J, et al: Blood pressure, stroke, and coronary heart disease: Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 335:765, 1990

110. Double-blind comparison of once daily betaxolol versus propranolol four times daily in stable angina pectoris. Betaxolol Investigators Group. Am J Cardiol 65:577, 1990

111. Comparison of celiprolol and propranolol in stable angina pectoris. Celiprolol International Angina Study Group. Am J Cardiol 67:665, 1991

112. Efficacy of atenolol and captopril in reducing risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. UK Prospective Diabetes Study Group. BMJ 317:713, 1998

113. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomized Amlodipine Survival Evaluation Study Group. N Engl J Med 335:1107, 1996

114. Cheitlin MD, Hutter AM Jr, Brindis RG, et al: ACC/AHA expert consensus document: use of sildenafil (Viagra) in patients with cardiovascular disease. American College of Cardiology/American Heart Association. J Am Coll Cardiol 33:273, 1999

115. Haffner SM, Lehto S, Ronnemaa T, et al: Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 339:229, 1998

116. Collins R, Armitage J, Parish S, et al: MRC/BHF Heart Protection Study of cholesterol-lowering with simvastatin in 5963 people with diabetes: a randomised placebo-controlled trial. Lancet 361:2005, 2003

117. Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. Heart Outcomes Prevention Evaluation (HOPE) Study Investigators. Lancet 355:253, 2003

118. Tatti P, Pahor M, Byington RP, et al: Outcome results of the Fosinopril versus Amlodipine Cardiovascular Events Randomized Trial (FACET) in patients with hypertension and NIDDM. Diabetes Care 21:597, 1998

119. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. HOT Study Group. Lancet 351:1755, 1998

120. Colwell JA: Aspirin therapy in diabetes. Diabetes Care 26(suppl 1):S87, 2003

121. Haffner SM: Management of dyslipidemia in adults with diabetes. Diabetes Care (26 suppl 1):S83, 2003

122. Coronary artery surgery study (CASS): a randomized trial of coronary artery bypass surgery: survival data. Circulation 68:939, 1983

123. Yusuf S, Zucker D, Peduzzi P, et al: Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. Lancet 344:563, 1994 [erratum, Lancet 344:1446, 1994]

124. Hueb WA, Soares PR, Almeida De Oliviera S, et al: Five-year follow-up of the medicine, angioplasty, or surgery study (MASS): a prospective, randomized trial of medical therapy, balloon angioplasty, or bypass surgery for single proximal left anterior descending coronary artery stenosis. Circulation 100(19 suppl):II107, 1999

125. Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. The Bypass Angioplasty Revascularization Investigation (BARI) Investigators. N Engl J Med 335:217, 1996 [erratum, N Engl J Med 336:147, 1997]

126. King SB 3rd, Kosinski AS, Guyton RA, et al: Eight-year mortality in the Emory Angioplasty versus Surgery Trial (EAST). J Am Coll Cardiol 35:1116, 2000

127. Coronary angioplasty versus medical therapy for angina: the second Randomised Intervention Treatment of Angina (RITA-2) trial. RITA-2 trial participants. Lancet 350:461, 1997

128. Parisi AF, Folland ED, Hartigan P: A comparison of angioplasty with medical therapy in the treatment of single-vessel coronary artery disease. Veterans Affairs ACME Investigators. N Engl J Med 326:10, 1992

129. Sculpher M, Smith D, Clayton T, et al: Coronary angioplasty versus medical therapy for angina: health service costs based on the second Randomized Intervention Treatment of Angina (RITA-2) trial. Eur Heart J 23:1291, 2002

130. Cowie MR, Wood DA, Coats AJ, et al: Incidence and aetiology of heart failure: a population-based study. Eur Heart J 20:421, 1999

131. Duncan CL, Cummings SR, Hudes ES, et al: Quitting smoking: reasons for quitting and predictors of cessation among medical patients. J Gen Intern Med 7:398, 1992

X UNSTABLE ANGINA AND NON–ST SEGMENT ELEVATION MYOCARDIAL INFARCTION

Joel Kupersmith, m.d., f.a.c.p.
Amish Raval, m.d.

Definition

Acute coronary syndromes are the constellation of symptoms, signs, and electrocardiographic and laboratory findings associated with new-onset or worsening myocardial ischemia. They include the spectrum of acute ST segment elevation myocardial infarction (MI) with or without Q waves, non–ST segment elevation MI (NSTEMI), and unstable angina. The main difference between NSTEMI and unstable angina is that in the latter, the ischemia is not severe enough to cause cardiac enzyme elevation and tissue injury; however, this difference may not be apparent on initial presentation.[1,2]

Pathophysiology

Unstable angina/NSTEMI results from an acute reduction in myocardial oxygen supply caused by rupture or erosion of an atherosclerotic coronary plaque; this plaque disruption is associated with inflammation, thrombosis, vasoconstriction, and microvascular embolization. The plaques implicated in these syndromes usually had previously produced only minor obstruction to blood flow (on angiography, up to 70% of affected vessels have less than 50% stenosis of the lumen); in addition, these plaques are characterized by a large lipid pool, reduced collagen content, a thin fibrous cap, and inflammatory cells.[3-9] Embolization of platelets and clot fragments into the microvasculature results in microcirculatory ischemia, which may account for the slight elevation of cardiac biomarkers. The events leading to unstable angina/NSTEMI also affect atherosclerotic plaques in the rest of the coronary vascular tree and other vascular territories.[10-12]

Less common causes of unstable angina/NSTEMI include intense focal epicardial spasm, cardiac emboli, and severe progressive atherosclerotic narrowing without superimposed thrombus.[1] Rarely, secondary unstable angina can be precipitated by conditions that increase myocardial demand, such as thyrotoxicosis, sepsis, fever, tachycardia, and anemia. Secondary unstable angina usually occurs in patients who also have underlying stable coronary atherosclerosis. Cocaine and amphetamines can also induce the syndrome.

Diagnosis

clinical presentation

Unstable angina/NSTEMI has three principal presentations: (1) prolonged angina at rest, usually lasting less than 20 minutes; (2) new-onset angina that is severe, disabling, and prolonged or frequent; and (3) established angina that has become distinctly more frequent, longer in duration, or more easily provoked.[13]

history and physical examination

Five clinical factors are key to establishing the diagnosis and the prognosis in patients with suspected unstable angina/NSTEMI: the anginal symptoms; any history of coronary artery disease (CAD); patient sex; patient age; and traditional risk factors for CAD [*see Table 1*].[1]

In the initial evaluation of a patient with suspected unstable angina/NSTEMI, the clinician should elicit a full description of the chest pain, including its character, onset, severity, and duration. Jaw pain, neck pain, epigastric pain, and arm pain may be experienced in isolation or in concert with the chest discomfort. In the National Registry of Myocardial Infarction, which included 440,000 patients, one third had atypical symptoms.[14] In the Alabama Unstable Angina Study of Medicare beneficiaries, which included over 4,000 patients, 51.7% of patients with unstable angina had the following atypical symptoms: dyspnea (69.4%), nausea (37.7%), diaphoresis (25.2%), syncope (10.6%), arm pain (11.5%), epigastric pain (8.1%), shoulder pain (7.4%), and neck pain (5.9%).[15] Atypical symptoms were more common in young patients (i.e., those 25 to 40 years of age), the elderly (i.e., those older than 75 years), diabetic patients, and women.[14]

Although it is important to inquire about the traditional risk factors for CAD, both to assess current risk and to guide future risk reduction, these factors are only weakly predictive of unstable angina/NSTEMI.[1] Important secondary or precipitating causes of unstable angina/NSTEMI should be ruled out, including a history of cocaine or amphetamine abuse, severe hypertension, and hyperthyroidism. Myopericarditis, hypothyroidism, and renal failure are among the conditions that may mimic acute coronary syndrome and can be associated with elevated cardiac enzyme levels.

In many cases, the physical examination will be normal. The examination is nevertheless useful to establish the presence of certain prognostic factors for the purpose of risk stratification and also to rule out such potentially devastating conditions as aortic dissection, pulmonary embolus, pneumothorax, and pericarditis.

The pulse is carefully evaluated to assess for significant bradycardia, tachycardia, or irregularity. Blood pressure is measured, to look for uncontrolled hypertension or significant hypotension (suggesting cardiogenic shock); pressure should be

Table 1 Clinical Factors for Determining Diagnosis and Prognosis in Patients with Suspected Unstable Angina or Non–ST Segment Myocardial Infarction

Nature of the anginal symptoms
History of coronary artery disease
Sex
Age ≥ 65 years
Number of traditional risk factors (smoking, diabetes, hypertension, hyperlipidemia, family history)

measured in both arms, to assess for aortic dissection [*see 1:XII Diseases of the Aorta*]. Examination of the thyroid may suggest hypothyroidism or hyperthyroidism. Heart failure should be suspected if there is evidence of pulmonary rales, an elevated jugular venous impulse, an S_3 gallop, or a displaced and diffuse apical impulse. A murmur of acute mitral regurgitation may be a consequence of ischemia. Thrills, bruits, or pulse deficits may indicate coexisting peripheral vascular disease.

LABORATORY TESTS

Biochemical Cardiac Markers

Cardiac biomarkers—specifically, troponins, cardiac creatine kinase, and myoglobin—have important diagnostic, prognostic, and therapeutic implications in unstable angina/NSTEMI and for detection of MI generally. After ischemia-induced myocardial injury, loss of myocyte membrane integrity results in the release of various intracellular molecules into the interstitial space, lymphatics, and, eventually, into the peripheral circulation. Critical to the interpretation of these tests is the precise time of onset of ischemic symptoms.

Troponin Troponin I and T (TnI and TnT) are cardiac-specific subunits of the thin filament-associated troponin-tropomyosin complex, which regulates striated muscle contraction. Troponins have become the primary biomarkers in the evaluation of patients with acute coronary syndromes. These markers are detected in about one third of patients without elevation in the level of creatine kinase–myocardial band (CK-MB). Troponins may be detectable 3 to 4 hours after the onset of ischemic symptoms; they peak at 12 to 48 hours and persist for 4 to 10 days.[16,17] Generally, they are not detectable in the blood of healthy persons. However, both TnI and TnT are exceptionally sensitive to the presence of even minor myocardial necrosis, such as in supraventricular tachycardia, heart failure, and myocarditis, and they may also be elevated in severe renal impairment. Therefore, they should be evaluated in the context of the patient's clinical presentation.

Creatine kinase Before the advent of troponin assays, CK-MB was the primary cardiac biomarker. The CK-MB assay has considerable sensitivity and specificity for detecting myocardial necrosis at 6 to 48 hours after symptom onset or earlier. Abnormal CK-MB levels can be occasionally found in patients with high total CK levels; on clinical grounds, the CK-MB in such cases is thought to originate from skeletal muscle. Elevations in levels of the CK-MB isoforms CK-MB2 and CK-MB1 are very early markers of myocardial necrosis, but these assays are generally not part of the clinical routine.[18]

Myoglobin Myoglobin is a nonspecific biomarker found both in cardiac and skeletal muscle. It is released rapidly in response to muscle injury and is detectable 2 hours after the onset of ischemia.[17] Serial determination of myoglobin is not useful, but because of its high sensitivity in early ischemia, a negative myoglobin assay could potentially rule out myocardial necrosis.

Twelve-lead electrocardiogram ECG results may be important for determining treatment; patients with acute ST segment elevation should be considered for immediate reperfusion therapy.[19] In patients with suspected unstable angina/NSTEMI, ST segment depression of 0.05 mV or more in two or more contiguous leads is highly consistent with myocardial ischemia (this is especially the case if the ST segment depression is present during chest pain and resolves with the easing of pain). Deep, symmetrical T wave inversion is also highly consistent with myocardial ischemia. Other nonspecific abnormalities, such as transient bundle branch block, atrial or ventricular arrhythmias, and QT prolongation, can also occur with unstable angina/NSTEMI but are not useful for diagnosis. Interestingly, a quarter of patients diagnosed with unstable angina/NSTEMI will go on to develop Q waves,[2] and up to 60% may have a normal 12-lead ECG.[20]

RISK STRATIFICATION

Determining whether a patient is at low, medium, or high risk for ischemic complications (e.g., full-blown MI) is important for deciding treatment of unstable angina/NSTEMI. Depending on whether the therapeutic strategy will be invasive or conservative (see below), the degree of risk can be used to determine the level of therapy. Older age, positive cardiac biomarkers, rales, ST segment depression, hypotension and tachycardia,[21] and reduced left ventricular ejection fraction (< 40%) have been associated with increased mortality. Clinical diabetes mellitus is also associated with higher risk. One specific and widely used method of risk stratification, the Antman/Thrombolysis in Myocardial Infarction (TIMI) risk score, is a seven-point scoring system that helps to predict death, reinfarction, or recurrent ischemia requiring revascularization [*see Table 2*].[22] This risk score was developed from the TIMI 11B[23] and Efficacy and Safety of Subcutaneous Enoxaparin in Non–Q-wave Coronary Events (ESSENCE)[24] trial and has been validated in two other large trials: Therapy with an Invasive or Conservative Strategy—Thrombolysis (TACTICS)–TIMI 18 and Platelet Receptor Inhibition in Ischemic Syndrome Management in Patients Limited by Unstable Signs and Symptoms (PRISM-PLUS).[25,26] The risk of adverse outcomes ranges from 5% to 41%, according to the simple sum of the individual variables.

Table 2 Thrombolysis in Myocardial Infarction (TIMI) Risk Score for Unstable Angina or Non–ST Segment Myocardial Infarction*

Variable	*Points*
Age ≥ 65 years	1
More than three coronary risk factors	1
History of coronary artery disease	1
ST segment deviation	1
More than two anginal events within 24 hr	1
Use of aspirin within past 7 days	1
Elevated cardiac markers	1

Total Score	*Risk of a Cardiac Event (%) Within 14 Days†*
0/1	5
2	8
3	13
4	20
5	26
6/7	41

*To determine a patient's level of risk, the clinician determines the total score, based on the presence of specific risk markers.
†Death, recurrent myocardial infarction, or recurrent ischemia requiring urgent revascularization.

Electrocardiographic Changes

Patients with unstable angina/NSTEMI and ECG findings of bundle branch block, ventricular hypertrophy, paced rhythm, or severe ST segment depression (> 0.2 mV) in multiple leads are independently at high risk for subsequent adverse events.[1,20] Considered at low risk are those with isolated T wave abnormalities or a normal ECG pattern. Continuous ECG monitoring may detect transient ischemic episodes, which have been shown in small studies to have prognostic value, but the use of this technique as a risk marker is not widely recommended.[1,27,28]

Blood Tests

Both TnI and TnT are markers of increased risk in that they reflect the presence and level of myocardial necrosis.[1,16] High-sensitivity C-reactive protein (hsCRP), which is produced by the liver in response to inflammation,[29,30] and other acute-phase reactants such as plasma fibrinopeptide, fibrinogen, serum amyloid A, and interleukin-6, have demonstrated similar predictive value for adverse outcomes.[31-33] B-type natriuretic peptide (BNP), released in response to ventricular wall stress, may also independently predict mortality.[34] Measurement of hsCRP and BNP may be of use at times, but more studies are needed to clarify the roles of these markers in routine care.

Stress Testing

Stress testing for assessment of risk may be performed in patients who have been stable and asymptomatic for 24 to 48 hours and in all patients before discharge. Stress testing is a necessary component of an early conservative strategy (see below). Briefly, stress test results that show the patient to be at high risk for significant ischemia are as follows:

1. Treadmill ECG: inability to achieve a workload of greater than two metabolic equivalents (mets); early and persistent ST segment depression of greater than 2 mm; symptom onset at less than 6.5 mets; or hypotension or ST segment elevation during exercise in the absence of Q waves
2. Stress nuclear scintigraphy: evidence of left ventricular dilation or thallium lung uptake during stress, or moderate to large reversible perfusion defects
3. Stress echocardiography: more than two myocardial wall segments demonstrating reversible impairment of thickening with stress or evidence of left ventricular dilation with stress[35]

Treatment

WHERE TO HOSPITALIZE

In most medical centers in the United States, patients with definite features of unstable angina/NSTEMI are admitted to a cardiac care unit that provides continuous ECG monitoring and specialized nursing. Patients with less than definite features of unstable angina or those in whom the chest pain has ceased by the time of arrival may be admitted to a chest pain unit with telemetry (sometimes called a step-down cardiac unit).

Ideally, patients with unstable angina/NSTEMI should be hospitalized in an institution that offers mechanical revascularization, because these procedures are frequently needed in such patients, especially those who have high-risk features. If this is not possible, high-risk patients should receive interim treatment with an intensive pharmacologic regimen until arrangements can be made for transfer to a facility with interventional capability.

INITIAL THERAPY

The initial management of unstable angina/NSTEMI includes resuscitation and supportive measures for patients who present with hemodynamic instability, as well as prompt administration of medication of proven, evidence-based value [*see* Pharmacologic Therapy, *below, and Figure 1*]. Bed rest with continuous ECG monitoring for ischemia and arrhythmia detection is a class I rec-

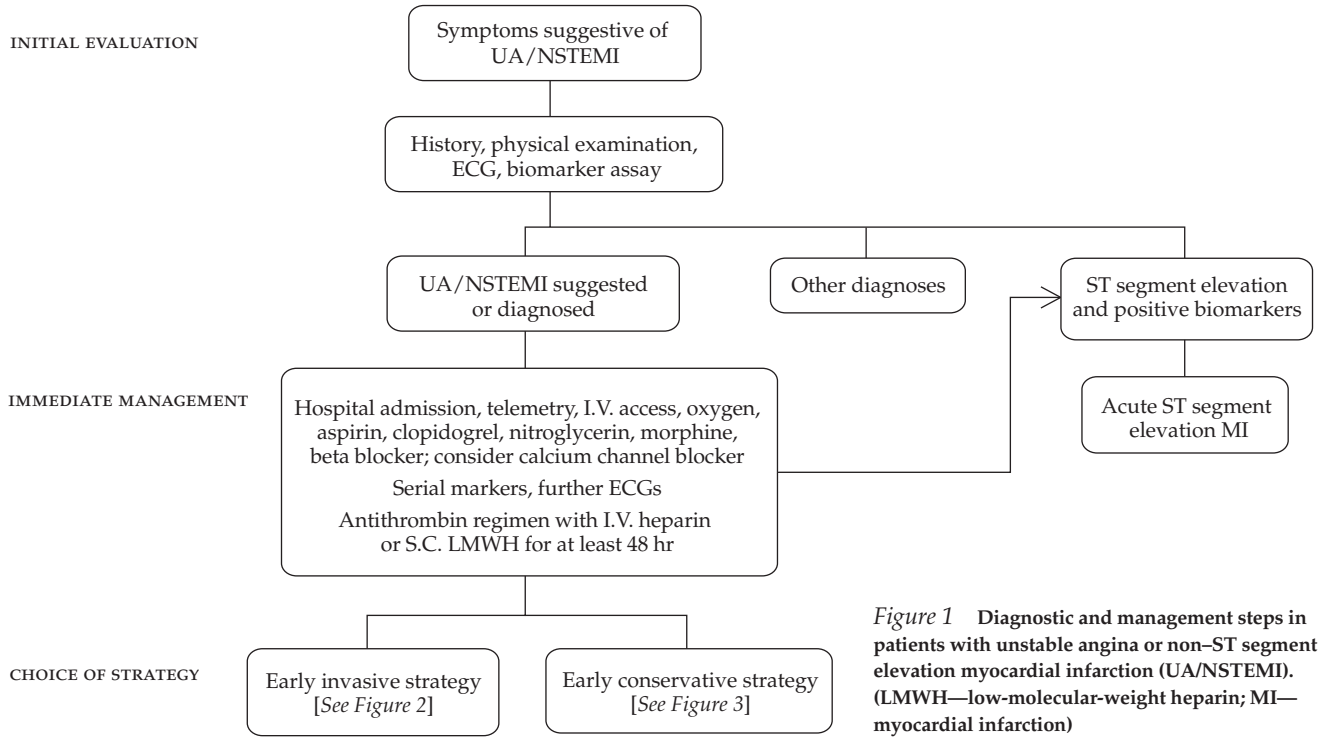


Figure 1 **Diagnostic and management steps in patients with unstable angina or non–ST segment elevation myocardial infarction (UA/NSTEMI). (LMWH—low-molecular-weight heparin; MI—myocardial infarction)**

Table 3 Recommendation Classes and Evidence Levels[1]

Class I: Conditions for which there is evidence or general agreement that a given procedure or treatment is useful and effective

Class II: Conditions for which there is conflicting evidence or a divergence of opinion about the usefulness/efficacy of a procedure or treatment

- Class IIa: Weight of evidence or opinion is in favor of usefulness/efficacy
- Class IIb: Usefulness/efficacy is less well established by evidence or opinion

Class III: Conditions for which there is evidence and/or general agreement that the procedure or treatment is not useful/effective and in some cases may be harmful

Level A: Data are derived from multiple randomized clinical trials that involved large numbers of patients

Level B: Data are derived from a limited number of randomized trials that involved small numbers of patients or from careful analyses of nonrandomized studies or observational registries

Level C: Expert consensus

ommendation for patients who have ongoing anginal pain at rest, on the basis of level C evidence [*see Table 3*]. The strictness of the bed rest requirement can be tailored to the severity of symptoms. For instance, patients can be mobilized to a chair or bedside commode when symptom free.

Oxygen Inhaled oxygen therapy should be reserved for those patients with clear respiratory distress, cyanosis, or arterial hypoxemia. In the absence of these high-risk features, time and resources need not be spent for the sole purpose of oxygen administration. Supplemental oxygen is recommended for patients with cyanosis or respiratory distress. Finger pulse oximetry or arterial blood gas measurements should be done to determine whether the patient has adequate arterial oxygen saturation ($S_aO_2 > 90\%$) or has hypoxemia and requires supplemental oxygen. These class I recommendations are supported by level C evidence.

EARLY INVASIVE VERSUS EARLY CONSERVATIVE STRATEGY

The first decision in management (and one that is often a matter of dispute) is to choose between the two reigning strategies for unstable angina/NSTEMI: early invasive management and early conservative management [*see Figures 2 and 3*]. In the early invasive strategy, early coronary angiography is performed unless contraindicated. In the early conservative strategy, angiography is reserved for those patients who have indications of being at high risk for cardiac events; such indications include evidence of significant ischemia on a noninvasive stress test and recurrent ischemia despite adequate medical therapy. It should be noted that although many patients assigned to the conservative strategy undergo angiography and receive interventions (e.g., in one study, 51% of patients underwent angiography and 36% received subsequent revascularization[25]), in these patients there was an additional indication for the angiography, based on risk, besides the diagnosis of unstable angina/NSTEMI.

Advocates of an early conservative strategy suggest that angiography—and its associated risks—can be avoided in low-risk patients and that costs and resources can be conserved by not performing these procedures in all patients. Advocates of an early invasive strategy suggest that this approach can result in superior clinical outcomes through early identification of patients with high-risk lesions, including those with critical left main coronary artery stenosis or triple-vessel coronary disease. In addition, advocates of an early invasive strategy argue that such intervention results in shorter hospital stays for patients found to have low-risk anatomy; these proponents also note that cardiac

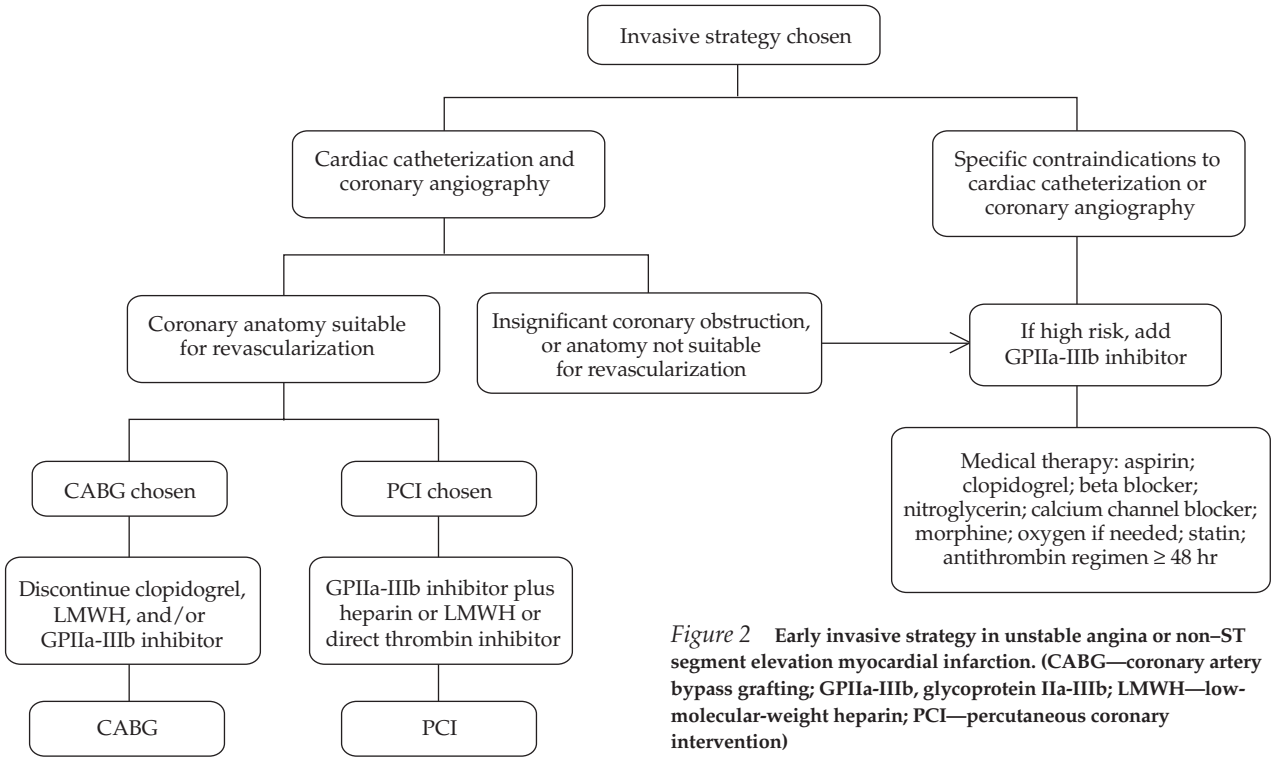


Figure 2 **Early invasive strategy in unstable angina or non–ST segment elevation myocardial infarction. (CABG—coronary artery bypass grafting; GPIIa-IIIb, glycoprotein IIa-IIIb; LMWH—low-molecular-weight heparin; PCI—percutaneous coronary intervention)**

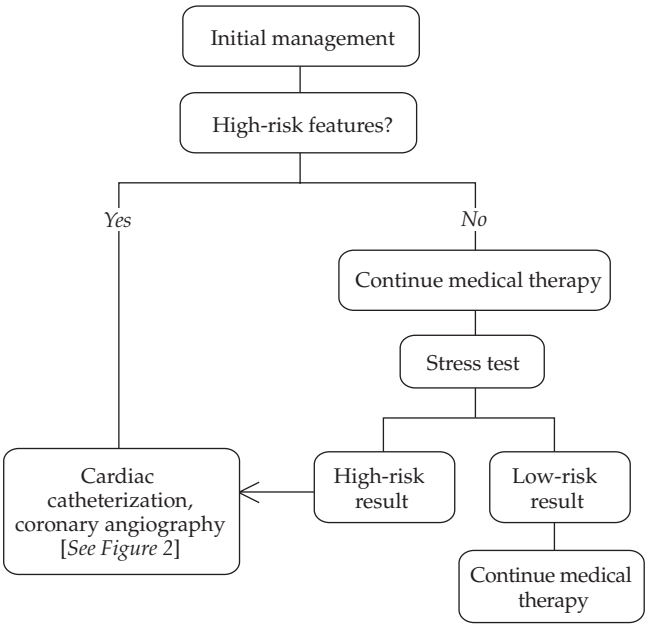


Figure 3 **Early conservative strategy in unstable angina/non–ST segment elevation myocardial infarction.**

catheterization is now available to almost all patients and that catheterization has a very low procedural risk.

Trial Results

Three initial multicenter, randomized trials found that there was no difference in outcome between an early invasive strategy and an early conservative strategy, whereas three subsequent trials showed benefit in favor of the invasive strategy. TIMI IIIB showed no difference in death or MI at 42 days with either strategy in patients with unstable angina/NSTEMI[36]; the Veterans Affairs Non–Q-Wave Myocardial Infarction Strategies In-Hospital (VANQWISH) trial showed no difference in death or recurrent MI in the early invasive versus the conservative group at 2 years' average follow-up (32.9% versus 30.3%; $P = 0.35$), although there were more deaths in the early invasive group at 1 year[37]; and the Medicine versus Angiography in Thrombolytic Exclusion (MATE) trial, conducted in patients with acute coronary syndrome who were ineligible for thrombolysis, showed no difference in clinical outcome between either strategy.[38]

The three subsequent studies have begun to move opinion toward the invasive strategy. The Fragmin in Unstable Coronary Artery Disease II (FRISC II) study was conducted in patients with unstable angina/NSTEMI who were receiving dalteparin; death or MI occurred in 9.4% of patients who received invasive treatment, as compared with 12.1% of those treated conservatively ($P < 0.031$; risk ratio, 0.78; confidence interval, 0.62 to 0.98).[39] However, a substudy of FRISC II showed that the most benefit of the early invasive strategy accrued to those patients with an ST segment shift of more than 2.5 mm or in five or more leads.[40] The Therapy with an Invasive or Conservative Strategy–Thrombolysis—TIMI 18 (TACTICS-TIMI 18) trial involved patients with unstable angina/NSTEMI who had ST or T wave changes, increased cardiac markers, and a history of coronary artery disease and who were taking aspirin, heparin, and tirofiban. This trial showed a combined end point (death, MI, or rehospitalization at 6 months) of 15.9% in the early invasive group, as compared with 19.4% in the early conservative group ($P = 0.025$; relative risk, 0.78; confidence interval, 0.62 to 0.97). Major bleeding was similar in both groups (1.9% versus 1.3%; $P = 0.24$). Subgroup analysis of the trial suggested that early invasive treatment benefited only patients whose TnI or TnT level was initially elevated.[41] Of note, median length of stay was somewhat shorter with the invasive strategy (3.9 days versus 4.3 days; $P < 0.001$). The cost of care in the invasive group was somewhat higher for the initial hospitalization ($15,714 versus $14,047 in year 2000 dollars), though the 6-month average total costs did not differ between the two groups ($21,813 for the invasive group versus $21,227 for the conservative group; P = nonsignificant).[42] In the Randomized Intervention Trial of unstable Angina–3 (RITA-3) in unstable angina/NSTEMI patients treated with enoxaparin, rates of death or MI at 4 months were lower in the early invasive group (9.6% versus 14.5%; $P = 0.001$; relative risk, 0.66; confidence interval, 0.51 to 0.85).[43]

Although the results of these three trials differed, patients who were considered to be at high risk for death or MI consistently seemed to benefit the most from a strategy of early angiography and revascularization. Because of the low event rate in patients considered to be at low risk, it is not clear that the early invasive approach offers a clear advantage over an early conservative strategy for this population. Therefore, until further evidence becomes available, an early conservative strategy may be a reasonable initial approach for the management of low-risk patients with unstable angina/NSTEMI.

Treatment Recommendations

High-risk patients A class I recommendation and level A evidence support the use of an early invasive strategy in patients with unstable angina/NSTEMI who have no serious comorbidity and have any of the following high-risk indicators:

- Recurrent angina at rest or with low-level activity despite intensive medical therapy.
- Elevated TnI or TnT levels.
- New, or presumably new, ST segment depression.
- Recurrent angina or ischemia with symptoms or signs of heart failure.
- High-risk findings on noninvasive stress testing.
- Depressed left ventricular function (ejection fraction < 40%) on noninvasive study.
- Hemodynamic instability.
- Sustained ventricular tachycardia.
- Percutaneous coronary intervention within the past 6 months.
- Previous coronary artery bypass grafting (CABG).

In the absence of any of these findings, either a conservative or an invasive strategy may be offered to hospitalized patients without contraindications for revascularization. This option is supported by a class 1 recommendation and level B evidence.

Women and the elderly There should be no difference in the management of men and women. The elderly should be treated no differently than younger patients, although management should take into consideration general health, comorbid conditions, cognitive status, life expectancy, and altered pharmacokinetics of and sensitivity to hypotensive drugs.

Table 4 Antiplatelet Therapy in Unstable Angina or Non–ST Segment Elevation Myocardial Infarction

Drug (Trade Name)	*Initial Dose*	*Route*	*Duration*	*Adverse Effects*	*Adverse Drug Reactions*	*Contraindications*
Aspirin	81–325 mg q.d.	Oral (first dose chewed)	Indefinite	Bleeding, tinnitus, rash, GI intolerance	None	Active, severe bleeding
Clopidogrel (Plavix)	300 mg loading dose, then 75 mg q.d.	Oral	1 yr	Bleeding, rare TTP	Partial inhibition of effect with statin drugs	Active, severe bleeding; major surgery within < 5 days
Ticlopidine (Ticlid)	250 mg b.i.d.	Oral	1 yr	Bleeding, TTP, rash, neutropenia, diarrhea	None	Active, severe bleeding; major surgery planned
Abciximab (ReoPro)	0.25 mg/kg bolus, then 0.125 µg/kg/min; max 10 µg/min	I.V.	12 hr	Bleeding, thrombocytopenia	None	Active, severe bleeding; major surgery planned
Eptifibatide (Integrilin)	180 µg/kg bolus, then 2 µg/kg/min	I.V.	72–96 hr or 18 hr if PCI performed	Bleeding	None	Active, severe bleeding; major surgery planned
Tirofiban (Aggrastat)	0.4 µg/kg/min bolus (× 30 min), then 0.1 µg/kg/min	I.V.	48 hr	Bleeding	None	Active, severe bleeding; major surgery planned

PCI—percutaneous coronary intervention TTP—thrombotic thrombocytopenic purpura

PHARMACOLOGIC THERAPY

Antiplatelet Drugs

Antiplatelet medications used in unstable angina/NSTEMI include aspirin, thienopyridines (clopidogrel and ticlopidine), abciximab, eptifibatide, and tirofiban [*see Table 4*].

Aspirin Aspirin is considered the benchmark antiplatelet agent for the treatment of unstable angina/NSTEMI. A class I recommendation and level A evidence suggest that aspirin be started immediately in these patients and continued indefinitely.[1,2]

Aspirin's mechanism of action is to decrease the formation of the potent platelet aggregator thromboxane A_2 by irreversibly binding cyclooxygenase-1 in platelets. The effect on platelets is rapid (occurring within 15 to 30 minutes), and it is achieved with an oral dose as low as 81 mg.

Four pivotal randomized trials that evaluated the effectiveness of aspirin in the treatment of acute coronary syndrome showed consistent and durable long-term benefit at doses ranging from 75 to 325 mg daily.[44-47] The Antiplatelet Trialists' Collaboration meta-analysis of more than 100,000 patients in 145 trials showed such benefits in several cardiovascular disorders. For example, in 4,000 patients with unstable angina, rates of so-called vascular events (nonfatal MI, nonfatal stroke, or vascular death) were reduced from 14% to 9% ($P < 0.00001$) after 6 months.[48] Furthermore, the benefit of aspirin in these high-risk patients was sustained for at least 2 years.

Aspirin dosages have varied among several trials; no dosage has been definitively shown to be preferable. For patients with suspected MI in the International Studies of Infarct Survival–2 (ISIS-2) trial, the effective dosage was 160 mg daily.[49] A dose of aspirin between 160 and 325 mg should be administered immediately; the first dose should be chewed, for rapid absorption, and subsequent doses swallowed.

Adverse effects of aspirin include allergy, which may manifest as rash, angioedema, or asthma; a tendency to bleed; gastrointestinal effects, including gastric ulcer; and, rarely, precipitation of acute gout. Contraindications include allergy (especially if the allergic reaction is in the form of asthma), active bleeding, a serious bleeding disorder, severe untreated hypertension, and an active peptic ulcer.

Thienopyridines Thienopyridines irreversibly bind the adenosine diphosphate receptor on platelets, preventing fibrinogen binding and platelet aggregation. The two thienopyridines that have been used clinically are ticlopidine (Ticlid) and clopidogrel (Plavix).

In a single open-label trial in patients with unstable angina, ticlopidine (250 mg twice daily) significantly reduced vascular death and nonfatal MI at 6 months (13.6% versus 7.3%; $P = 0.009$), compared with standard aspirin therapy.[50] Several trials have shown the value of clopidogrel. In 12,562 patients with unstable angina, all of whom were also treated with aspirin, clopidogrel (300 mg followed by 75 mg daily), administered for 3 to 12 months, reduced a combined end point of cardiovascular death, nonfatal MI, and stroke from 11.4% to 9.3% ($P < 0.001$) and also decreased the incidence of ischemia, heart failure, and revascularization procedures. Benefit occurred as early as 24 hours after initiating treatment and persisted for up to 1 year. Clopidogrel increased major bleeding from 2.7% to 3.7% ($P = 0.003$), but there was no difference in life-threatening bleeding or hemorrhagic shock. Importantly, bleeding was increased in patients who underwent CABG within 5 days after stopping clopidogrel.[51] Two other studies found that clopidogrel benefited patients treated with percutaneous coronary intervention.[52,53] The Clopidogrel for the Reduction of Events During Observation (CREDO) study examined the timing of therapy with a combination of clopidogrel and aspirin. In one arm of the study, patients received a bolus load of aspirin and clopidogrel 6 to 24 hours before undergoing percutaneous coronary intervention. In the other arm of the study, patients received a bolus load less than 6 hours before the procedure. Benefit was observed in those patients who received the bolus 6 to 24 hours before the procedure.

The dosage of ticlopidine is 250 mg orally, twice daily; the dosage of clopidogrel is 300 mg orally. Both agents should be started immediately on presentation. The duration of therapy,

which has been better defined for clopidogrel, is up to 1 year at a dose of 75 mg daily.

The principal adverse reaction that has limited the clinical use of ticlopidine is severe neutropenia; rarely, thrombotic thrombocytopenic purpura develops within the first 3 months of therapy. Both conditions are life threatening unless the drug is discontinued promptly. Clopidogrel has been associated with increased bleeding complications when administered with other antithrombotic agents, particularly when arterial puncture is performed for intervention. Bleeding during surgery is an important complication of clopidogrel use; for this reason, surgery should be avoided for 5 and preferably 7 days after the last dose, because of the prolonged duration of action.

Recommendations for the use of clopidogrel are as follows:

1. Patients with unstable angina/NSTEMI in whom a noninterventional approach is planned should receive clopidogrel for at least 1 month (this class I recommendation is supported by level A evidence) and for up to 9 months (class I recommendation, level B evidence).
2. Patients with unstable angina/NSTEMI for whom a percutaneous coronary intervention is planned and who are not at high risk for bleeding should receive clopidogrel for at least 1 month (class I recommendation, level A evidence) and for up to 9 months (class I recommendation, level B evidence).
3. Clopidogrel should be stopped for 5 to 7 days before elective surgical revascularization (class I recommendation, level B evidence).
4. When used to pretreat patients undergoing percutaneous coronary intervention, clopidogrel should be given more than 6 hours before the procedure (level B evidence).[53]

Glycoprotein IIb-IIIa receptor antagonists Abciximab, eptifibatide, and tirofiban act by specifically binding the glycoprotein (GP) IIb-IIIa receptor on platelet surfaces, thereby preventing fibrinogen binding and ultimately preventing platelet aggregation. Abciximab is the Fab fragment of a monoclonal antibody that has a short plasma half-life but irreversibly binds the GPIIb-IIIa receptor for 24 to 48 hours. The half-lives of eptifibatide and tirofiban are 2 to 3 hours, with platelet aggregation returning to normal 4 to 8 hours after drug discontinuance.

In three large clinical trials of patients with unstable angina/NSTEMI who underwent percutaneous coronary intervention, all of these GPIIb-IIIa inhibitors provided significant benefit in the composite outcome of death, MI, or urgent repeat revascularization, with the major benefit seen in recurrent MI and urgent repeat revascularization.[26,54,55] The Do Tirofiban and Reopro Give Similar Efficacy Trial (TARGET), which was conducted in patients with unstable angina/NSTEMI who underwent percutaneous coronary intervention with stenting, found that abciximab conferred greater benefit than tirofiban (although tirofiban was administered at a suboptimal dosage).[56]

For patients not undergoing planned percutaneous intervention, the results with GPIIb-IIIa inhibitors have been less impressive. Abciximab, given for 24 or 48 hours, was found to be no better than placebo in the Global Use of Strategies to Open Occluded Coronary Arteries IV (GUSTO IV) trial in patients with acute coronary syndrome.[57] On the other hand, two trials with tirofiban and eptifibatide demonstrated modest benefit in patients who did not undergo an interventional procedure[26,54]; the benefit was greatest in high-risk patients.

Abciximab, which is recommended for use during percutaneous coronary intervention, is administered as a 0.25 mg/kg intravenous bolus, followed by an infusion at 0.125 μg/kg/min for 12 hours. Eptifibatide is administered as a 180 μg/kg intravenous bolus, followed by a second bolus after 10 minutes. Thereafter, it is infused at a rate of 2 μg/kg/min for 72 to 96 hours if no percutaneous intervention is performed or for 18 hours if such a procedure is performed. Tirofiban is given in an intravenous 0.4 μg/kg/min bolus over 30 minutes, then infused at a rate of 0.1 μg/kg/min for 48 hours.

Bleeding is the most common complication of GPIIb-IIIa inhibitors. Special care should be taken to prevent bleeding in high-risk patients such as the elderly, women, those with low body weight, and those who require arterial puncture for an intervention. Because of their short half-lives, eptifibatide or tirofiban may be better suited for patients who may require surgical revascularization. Abciximab results in serious thrombocytopenia in about 0.3% of patients,[56] so serial platelet measurements are recommended in patients receiving this agent. In a meta-analysis, GPIIb-IIIa inhibitors were found to increase major bleeding from 1.4% to 2.4 % ($P < 0.0001$). There was no increase in the rate of intracranial hemorrhage.[58]

Three evidence-based recommendations can be made for the use of GPIIb-IIIa inhibitors:

1. In patients with unstable angina/NSTEMI who are to undergo a planned percutaneous coronary intervention, a GPIIb-IIIa inhibitor should be given either before or during the procedure. This class I recommendation is supported by level A evidence.
2. Patients with unstable angina/NSTEMI who have high-risk features and who are not to undergo a planned percutaneous intervention should receive eptifibatide or tirofiban (class IIa recommendation, level A evidence).
3. Patients with unstable angina/NSTEMI in whom percutaneous intervention is not planned should not receive abciximab (class III recommendation, level A evidence).

Fibrinolytic Drugs

Fibrinolytic therapy is not indicated for patients with unstable angina/NSTEMI. Level A evidence indicates that intravenous fibrinolytic therapy should not be administered to patients who do not have acute ST segment elevation, unless they have a true posterior MI or a presumed new left bundle branch block (class III recommendation).[59]

Anticoagulants

Heparin, low-molecular-weight heparin (LMWH), bivalirudin, or warfarin can be used for anticoagulant therapy in patients with unstable angina/NSTEMI [*see Table 5*].

Heparin Unfractionated heparin is composed of a number of chains of varying molecular weights that differ with regard to anticoagulant activity. Heparin generally increases the action of circulating antithrombin, which inactivates factor IIa, factor IXa, and factor Xa and prevents thrombus formation.[60]

Three randomized, placebo-controlled trials suggested that early intravenous administration of heparin leads to a modest reduction in the incidence of MI or recurrent ischemia.[61-63] A meta-analysis of six trials showed a relative risk of 0.67 (95% confidence interval, 0.44 to 1.02) in favor of the combination of heparin and aspirin.[64]

Heparin should be given in an initial intravenous bolus of 60 to 70 U/kg (maximum, 5,000 U), followed by an infusion of 12 to

Table 5 Anticoagulant Therapy in Unstable Angina/ Non–ST Segment Elevation Myocardial Infarction

Drug	*Dose*	*Route*	*Duration*	*Adverse Effects*	*Adverse Drug Reactions*	*Contraindications*
Unfractionated heparin	60–70 U/kg bolus (max 5,000 U); 12–15 U/kg/hr (max 1,000 U/hr)	I.V.	48–72 hr	Bleeding, HIT, mild thrombocytopenia	None	Active, severe bleeding; HIT
Enoxaparin	1 mg/kg b.i.d.	S.C.	48–72 hr	Bleeding, HIT	None	Active bleeding; major surgery planned
Bivalirudin	0.75 mg/kg bolus; 1.75 mg/kg/hr	I.V.	Duration of PCI	Bleeding	None	Active, severe bleeding
Warfarin	5–10 mg; dose adjusted to maintain desired INR, typically 2–3	Oral	Varies with indication	Bleeding, warfarin skin necrosis	Dose adjustment with macrolides, cimetidine, digoxin, amiodarone, other drugs	Active, severe bleeding; major surgery planned

HIT—heparin-induced thrombocytopenia INR—international normalized ratio PCI—percutaneous coronary intervention

15 U/kg/hr (maximum, 1,000 U/hr). Doses should be adjusted to maintain an activated partial thromboplastin time (aPTT) of 1.5 to 2.5 times control values.[1]

As with any anticoagulant, heparin is contraindicated in patients with active severe bleeding or who are to undergo imminent surgery. Heparin-induced thrombocytopenia (HIT) is a serious but rare (< 0.2% incidence) antibody-mediated reaction leading to reduced platelet counts and thrombosis [*see 5:XIV Thrombotic Disorders*]. HIT mandates immediate cessation of heparin, including heparinized solutions used for the flushing of intravenous ports. From 10% to 20% of patients receiving heparin may experience mild thrombocytopenia that is not associated with severe thrombosis or excessive bleeding.[1] For serious bleeding, heparin anticoagulation can be immediately reversed with intravenous protamine sulfate.

Low-molecular-weight heparin LMWHs are produced by enzymatic depolymerization of unfractionated heparin, resulting in smaller chain units. Advantages include less protein binding; longer half-life; more stable, dose-dependent clearance; and a greater anti–factor Xa effect that is associated with more thrombin inhibition. The standard aPTT assay cannot be used for LMWH, because this assay is not sensitive to the anti-Xa effects of LMWHs. Rapid factor Xa assays are not widely available, but the stable kinetics of LMWH tends to reduce the need for monitoring.

A number of trials have compared individual LMWHs with unfractionated heparin. Meta-analysis of two trials revealed a modest benefit of enoxaparin over heparin in terms of death and MI at 45 days (7.1% versus 8.6%; $P = 0.02$).[65] In trials with dalteparin and nadroparin, neither showed benefit.[66,67] FRISC showed a significant improvement in death, MI, or urgent revascularization with dalteparin (1.8% versus 4.8%; $P = 0.001$) for up to 43 days.[68] In the Integrilin and Enoxaparin Randomized Assessment of Acute Coronary Syndrome Treatment (INTERACT) trial, enoxaparin was found to be associated with less ischemia, as evidenced on ECG (14.3% versus 25.4%; $P = 0.002$), and to lead to an improvement in the composite end point of death or MI at 30 days (5% versus 9%; $P = 0.031$).[69] Several trials showed evidence of a small increase in minor bleeding with LMWH, as compared with heparin, but no difference in major bleeding was seen.[66-69] One small trial demonstrated that enoxaparin was safe with regard to serious bleeding in unstable angina/NSTEMI patients undergoing percutaneous coronary intervention.[70] Another trial showed significantly lower rates of major CABG–related bleeding with LMWH, as compared with enoxaparin, by 96 hours after percutaneous coronary intervention (1.8% versus 4.6%; $P = 0.03$).[23] When enoxaparin was added to abciximab, no difference in the rate of bleeding events and adverse ischemic outcomes was seen, as compared with historical controls.[71]

In summary, evidence to date suggests that LMWH is at least as effective as unfractionated heparin and is generally safe in patients with unstable angina/NSTEMI in whom surgery or percutaneous coronary intervention is not planned. The dosing of LMWH is simpler than that of unfractionated heparin, and LMWH therapy does not require monitoring with coagulation studies. Abundant data (e.g., from the TIMI 11B, ESSENCE, and INTERACT studies) suggest that enoxaparin is the preferred agent.[23,24,69] In patients for whom percutaneous coronary intervention is planned, LMWH may be started or continued, provided that meticulous attention is given to dosing and the timing of its administration. Enoxaparin has not found wide use in contemporary interventional practice, however, because of practitioners' uneasiness over the inability to monitor the level of anticoagulation. In patients in whom surgery is planned, LMWH should be avoided; if an LMWH has already been started, the patient should be switched to unfractionated heparin, whose effects can be monitored closely.

Enoxaparin is given subcutaneously at a dosage of 1 mg/kg every 12 hours for at least 48 hours. Percutaneous coronary intervention may be performed within 8 hours of starting the drug, without additional anticoagulation therapy. If the percutaneous procedure is performed 8 to 12 hours after starting the drug, an additional intravenous bolus of 0.3 mg/kg may be given. Dose adjustment must be made in patients with moderate to severe renal impairment.

Compared with unfractionated heparin, LMWH has a slightly higher rate of minor bleeding complications and the same rate of major bleeding complications. Although the risk of HIT seems to be lower with LMWH than with unfractionated heparin, LMWH is absolutely contraindicated in patients with a history of HIT because of the danger posed by this reaction.

A class I recommendation and level A evidence support the use of anticoagulation with intravenous heparin or subcutaneous LMWH in patients with unstable angina/NSTEMI; the anticoagulant is given in addition to aspirin, clopidogrel, or both.

Level A evidence indicates that enoxaparin is preferable to unfractionated heparin in patients without renal failure, unless CABG is planned within 24 hours (class IIa recommendation).

Direct thrombin inhibitors These agents directly bind to the fibrinogen-recognition and catalytic sites of thrombin, neutralize clot-bound thrombin, and inhibit thrombin-mediated platelet aggregation; the result is sustained anticoagulation. Hirudin and bivalirudin are the two principal agents that have undergone clinical testing, but hirudin is not used.

In the Randomized Evaluation in Percutaneous Coronary Intervention Linking Angiomax to Reduced Clinical Events–2 (REPLACE-2) trial, which involved patients undergoing percutaneous coronary intervention, bivalirudin in combination with provisional GPIIb-IIIa inhibition was found to be equivalent to unfractionated heparin plus GPIIb-IIIa inhibition in terms of death, MI, urgent repeat revascularization, and in-hospital major bleeding after 30 days of therapy.[72] Rates of in-hospital major bleeding were significantly lower with bivalirudin (2.4% versus 4.1%; $P < 0.001$), although it has been noted that activated clotting time (ACT) values in the patients who received unfractionated heparin and a GpIIb-IIIa inhibitor were higher than was seen in other trials.

Bivalirudin is given as an intravenous bolus of 0.75 mg/kg at the start of percutaneous coronary intervention; for the duration of the procedure, an infusion of 1.75 mg/kg/hr is given. Because of the predictable profile of bivalirudin, measurement of ACT levels is usually not necessary. If the ACT is measured, typical values are 300 to 400 seconds.

Although bleeding seems to occur less commonly with bivalirudin than with unfractionated heparin or GPIIb-IIIa inhibitors, any serious bleeding that does occur could be disastrous because there is no effective way to immediately reverse the anticoagulation (as can be done with protamine sulfate for heparin). However, thrombocytopenia is not a problem, and the offset of bivalirudin's effect is relatively more rapid than that of heparin (1 hour), permitting sheaths to be pulled sooner after the procedure. Mainly because of its better safety profile, some interventional cardiologists have embraced the use of bivalirudin in the catheterization laboratory, but its general use awaits further trials.

If percutaneous coronary intervention is planned, level B evidence indicates that bivalirudin may be used as an alternative to unfractionated heparin and a GPIIb-IIIa inhibitor.

Warfarin Warfarin works by inhibiting the vitamin K–dependent clotting factors II, VII, IX, and X. Achievement of the desired antithrombotic effect takes 4 to 7 days.

In a few small pilot studies, starting warfarin therapy shortly after presentation in patients with unstable angina/NSTEMI showed benefit.[1] In a large trial of patients with unstable angina/NSTEMI who had previously undergone CABG, warfarin provided no advantage over aspirin alone and was associated with excess minor and major bleeding complications.[73]

Warfarin is taken orally once daily. The dose is titrated to maintain the desired international normalized ratio (INR), which is from 2 to 3 for most indications.

Bleeding is the main complication associated with warfarin therapy. Several drugs (e.g., cimetidine, amiodarone) can markedly increase warfarin's effect. Warfarin anticoagulation is reversed by stopping the drug; reversal by administration of vitamin K or of fresh frozen plasma is usually reserved for emergent and life-threatening bleeding.

The available evidence does not support the routine use of warfarin in patients with unstable angina/NSTEMI unless there are other indications for warfarin (e.g., atrial fibrillation, mechanical prosthetic heart valve).

Anti-ischemia Therapy

The agents used for treating ischemia in patients with unstable angina/NSTEMI include nitrates, morphine sulfate, and beta blockers (e.g., metoprolol) [*see Table 6*].

Nitrates Nitroglycerin is an endothelium-independent general arterial and venous dilator. It decreases myocardial oxygen demand through increased venous capacitance and peripheral artery dilation—factors that reduce preload and afterload, respectively, and thereby reduce myocardial wall stress. Epicardial coronary vasodilation and increased collateral flow act to enhance myocardial oxygen delivery.

No large, placebo-controlled clinical trials addressing reductions in major cardiac events or symptoms in unstable angina/NSTEMI have been performed. Multiple small, uncontrolled trials, a well-characterized biologic effect, and decades of experience have made nitrates a standard of care in the early treatment of these patients.[1]

Table 6 Anti-ischemia Therapy

Drug	*Dosage*	*Route*	*Duration*	*Adverse Effects*	*Adverse Drug Reactions*	*Contraindications*
Nitrates	S.L.: 0.4 mg q. 5 min × 3; I.V.: 10 mg/min, titrate up q. 5 min; paste: 2–6 cm; patch: 0.4 mg/hr	S.L., I.V., paste, patch	15–20 min until side effects occur or symptoms resolve	Hypotension, headache, nausea, tolerance	None	Hypotension; sildenafil or vardenafil within 24 hr of nitrate use
Morphine sulfate	1–5 mg every 10–15 min	I.V.	Until side effects or symptoms resolve	Hypotension, respiratory depression, rash, pruritus, nausea	None	Severe ventilatory failure, hypotension
Metoprolol*	5 mg q. 5 min to max 3 doses, then 25–100 mg b.i.d. with titration	I.V., then oral	Symptom resolution (if MI, then indefinite)	Hypotension, bradycardia, bronchospasm, worsened claudication	None	Hypotension, bradycardia, asthma, decompensated heart failure

*Other beta blockers are equally effective, although agents with intrinsic sympathomimetic activity should be avoided. Calcium channel blockers may be used for patients in whom beta blockers are contraindicated.
MI—myocardial infarction

In the emergency department, the initial nitrate dosage in a nonhypotensive patient is typically 0.4 mg of sublingual nitroglycerin (tablet or spray) repeated approximately every 5 minutes if ischemic symptoms do not subside. If this fails to terminate the ischemia, intravenous nitroglycerin at an infusion rate of 10 μg/min is recommended. The dose is titrated upward in increments of 10 to 20 μg/min until symptoms or signs of ischemia subside, hypotension develops, or the recommended maximal dose of 200 μg/min is achieved. After the acute period, topical nitrates, such as a 0.4 mg/hr nitrate patch, can be used for long-term therapy if necessary.

Nitroglycerin can result in significant hypotension, necessitating withdrawal of the agent. It should be avoided in patients who have taken sildenafil (Viagra) or vardenafil (Levitra) in the past 24 hours, because very severe hypotension can occur. Nitrate tolerance develops with prolonged nitrate administration, so patients should have nitrate-free intervals. Other adverse reactions, including intermittent headaches and nausea, are common.

A class I recommendation and level C evidence support the use of nitroglycerin to terminate or prevent ischemic episodes. Nitroglycerin therapy should serve as a bridge to the use of other evidence-based therapies, such as revascularization.

Morphine sulfate Almost a century of experience in acute coronary syndrome (albeit in the absence of clinical trial data) has established morphine sulfate as a useful adjunct to the early management of unstable angina/NSTEMI. Besides its potent analgesic and anxiolytic effect, morphine sulfate causes venodilation and mild arterial dilation, leading to reduced preload and afterload, and it may increase vagal tone to modestly reduce heart rate. These effects make it useful for treating patients with severe pulmonary congestion.

Morphine sulfate is given in a dose of 1 to 5 mg intravenously every 10 to 15 minutes until ischemic symptoms dissipate. At the same time, consideration for intravenous nitroglycerin treatment should be made. Morphine sulfate often causes hypotension, nausea, vomiting, and respiratory depression. Effects are quickly reversible with naloxone. A class I recommendation and level C evidence support the use of morphine sulfate to relieve chest pain that is refractory to sublingual nitroglycerin, to reverse acute pulmonary edema, and to ease severe agitation.

Beta blockers Beta blockers reduce myocardial oxygen demand by reducing heart rate and contractility. Slowing of the heart rate may also permit increased coronary filling during a prolonged diastole.

Much of the evidence in favor of beta-blocker therapy for unstable angina/NSTEMI is extrapolated from the large benefit shown in major clinical trials of acute MI. A meta-analysis of trials of threatened or evolving MI revealed a 13% reduction in progression to acute MI,[74] but there was insufficient power for mortality analysis.

There is no evidence of any difference in efficacy between the various beta blockers available, although agents with intrinsic sympathomimetic activity should be avoided. The dose will vary with the agent selected. In unstable angina/NSTEMI, intravenous loading doses titrated to a target resting heart rate of 50 to 60 beats/min may be used, with rapid conversion to an oral regimen. In patients who may have difficulty tolerating the adverse effects of beta blockers, the initiating dose should be small and titration should proceed more slowly.

Bronchospasm and severe asthma are contraindications to the use of beta blockers. Significant sinus bradycardia, AV nodal block, and hypotension can also occur, typically in patients with preexisting disease of cardiac conductive tissue.

A class I recommendation and level B evidence support the use of beta blockers in patients with ongoing chest pain; in such patients, an initial intravenous dose is followed by oral therapy.

Calcium channel blockers These agents are primarily vasodilators, but they also have effects on atrioventricular nodal conduction and left ventricular contractility. The dihydropyridine calcium channel blockers (e.g., nifedipine and amlodipine) have the most peripheral vasodilatory capability and the least negative inotropic effect.

To date, the trial data generally suggest that calcium channel blockers offer symptom relief in patients with unstable angina, but a meta-analysis found no improvements with regard to death or the occurrence of MI.[75] Nifedipine, compared with a beta blocker (metoprolol), demonstrated a trend toward increased MI.[76] In unstable angina/NSTEMI, the nondihydropyridine agents are used for coronary artery spasm and may be chosen for patients who cannot tolerate beta blockade.

Doses of calcium channel blockers vary with the agent chosen. For diltiazem, the usual immediate dose is a 20 mg/kg intravenous bolus, followed 15 minutes later by a 20 to 25 mg/kg bolus. Thereafter, the drug is administered orally, in a dosage of 30 mg three to four times a day, titrated to a total daily dose of 360 mg if necessary. Long-acting formulations exist as well.

Hypotension occurs with all calcium channel blockers. Bradycardia and negative inotropic effects accompany the nondihydropyridine agents; these agents should be avoided in patients with heart failure. The dihydropyridine agents may cause reflex tachycardia and other sympathomimetic effects.

Recommendations for the use of calcium channel blockers are as follows:

1. Nondihydropyridine calcium channel blockers (i.e., verapamil, diltiazem) may be initiated in patients with continuing or frequently recurring ischemia for whom beta blockers are contraindicated, in the absence of severe left ventricular dysfunction or other contraindications of calcium channel blockade. This class I recommendation is supported by level B evidence.
2. Extended-release forms of nondihydropyridine calcium channel blockers may be used instead of a beta blocker (class IIa recommendation, level B evidence).
3. Immediate-release dihydropyridine calcium channel blockers (i.e., nifedipine) should not be used in the absence of beta-blocker treatment (class III recommendation, level B evidence).
4. Immediate-release dihydropyridine calcium channel blockers may be used, if specifically indicated, in patients who are receiving a beta blocker (class II recommendation, level B evidence).

Lipid-Lowering Agents

There is as yet no evidence from a clinical trial that indicates that the use of lipid-lowering agents in hospital confers a benefit for patients with unstable angina/NSTEMI. However, data from a large Swedish registry showed a reduction in mortality in MI patients given statin therapy before discharge.[77] In addition, pa-

tients given such therapy in the hospital are much more likely to continue it out of hospital, and in-hospital use has therefore been recommended.[78,79] The use of a fibrate or niacin in patients with a high-density lipoprotein cholesterol level of less than 40 mg/dl is supported by a class I recommendation and level B evidence. A class IIa recommendation and level B evidence support treatment with statins and diet for patients whose low-density lipoprotein cholesterol is greater than 100 mg/dl; treatment should begin 24 to 96 hours after admission and continue after hospital discharge.

Angiotensin-Converting Enzyme Inhibitors

Angiotensin-converting enzyme (ACE) inhibitors may block inflammatory processes and encourage plaque stability. In the Heart Outcomes Prevention Evaluation (HOPE) trial, use of the ACE inhibitor ramipril was associated with significant reductions in death, MI, or stroke in moderate- to high-risk patients, most of whom had normal left ventricular function.[80] The American College of Cardiology/American Heart Association (ACC/AHA) recommends ACE inhibitors for patients with unstable angina/NSTEMI and heart failure; left ventricular systolic dysfunction (ejection fraction less than 40%); hypertension; or diabetes. This class I recommendation is supported by level B evidence.

MECHANICAL REVASCULARIZATION

Coronary revascularization with percutaneous procedures or CABG is performed to relieve symptoms and improve prognosis [*see* Early Invasive versus Conservative Strategy, *above*]. Several factors influence the decision to proceed with coronary revascularization, including risk, absence of relevant comorbid conditions, disabling symptoms, viable myocardium at risk, and whether the patient's coronary anatomy is suitable for the procedure.

Percutaneous Coronary Intervention

Advances in percutaneous coronary intervention techniques and devices have improved safety and long-term vessel patency rates. Several changes in the evolution of coronary stent design, including smaller profile, increased flexibility, small strut diameter, and, the newest development, drug-eluting technology, have improved deliverability and reduced the rate of in-stent restenosis. Furthermore, the use of adjunctive antiplatelet and antithrombotic therapies, particularly in the setting of acute coronary syndrome, has improved outcomes.

The ACC/AHA guidelines for the use or avoidance of percutaneous coronary intervention in unstable angina/NSTEMI are as follows[1]:

1. Percutaneous coronary intervention (or CABG) is recommended for patients with single-vessel or two-vessel CAD without significant involvement of the proximal left anterior descending coronary artery (LAD) who have large areas of viable myocardium and high-risk features on noninvasive testing. This class I recommendation is supported by level B evidence.
2. Percutaneous coronary intervention is recommended for patients who have single-vessel or multivessel CAD, have suitable coronary anatomy, have normal left ventricular function, and do not have diabetes (class 1 recommendation, level A evidence).
3. Percutaneous coronary intervention (or CABG) is recommended for patients who have single-vessel or two-vessel CAD without significant proximal LAD involvement but who have a moderate area of viable myocardium and ischemia on noninvasive testing (class IIa recommendation, level B evidence).
4. Percutaneous coronary intervention (or CABG) is recommended for patients who have single-vessel disease with significant proximal LAD involvement (class IIa recommendation, level B evidence).
5. Percutaneous coronary intervention (or CABG) is not indicated for patients who have atypical symptoms or have no evidence of ischemia on noninvasive testing or have not received an adequate trial of medical therapy (class III recommendation, level C evidence).
6. Percutaneous coronary intervention (or CABG) is not indicated for patients with significant left main CAD who are suitable candidates for CABG (class III recommendation, level B evidence).

Surgical Revascularization

Surgical revascularization techniques and perioperative outcomes have improved over the years. Particular advances include use of internal mammary artery conduits, off-pump procedures, minithoracotomy, and, the newest development, robot-assisted procedures.

The ACC/AHA guidelines for CABG in patients with unstable angina/NSTEMI are as follows[1]:

1. CABG is indicated for patients with significant left main CAD. This class I recommendation is supported by level A evidence.
2. CABG is indicated for patients with triple-vessel CAD and abnormal left ventricular function (class 1 recommendation, level A evidence).
3. CABG is indicated for patients who have two-vessel disease with significant proximal LAD disease and abnormal left ventricular function (class 1 recommendation, level A evidence).
4. CABG or percutaneous intervention is indicated for patients who have single-vessel or two-vessel CAD without significant proximal LAD involvement but with large areas of viable myocardium and high-risk features on noninvasive testing (class 1 recommendation, level B evidence).

Diabetes and Revascularization

Overall, patients with diabetes are more likely to require repeat revascularization after percutaneous intervention, because of increased rates of restenosis; in addition, there is a trend toward higher mortality 1 year after both CABG and percutaneous intervention with stents in diabetic patients.[81] In the Bypass Angioplasty Revascularization Investigation (BARI) trial, diabetic patients with multivessel CAD were found to have better survival rates with CABG than with percutaneous intervention.[82] However, analysis of the diabetic subgroup of a randomized trial and registry of percutaneous intervention with bare metal stents versus CABG in unstable angina patients revealed no difference in 3-year survival between the groups.[83] Drug-eluting stents display markedly reduced rates of in-stent restenosis, as compared with traditional bare-metal stents, particularly in diabetic patients.[84,85]

Currently, the available evidence suggests that surgical revascularization should be offered to diabetic patients with CAD in three or more vessels, particularly if they have left ventricular

dysfunction. However, it is common practice to offer percutaneous intervention as the revascularization strategy for diabetic patients with CAD involving one or two vessels. Trials are needed to compare the most advanced drug-eluting stent technology with the most advanced surgical management to define their roles in diabetic patients with CAD.

POSTHOSPITAL CARE

Preparation for the posthospital care of a patient with unstable angina/NSTEMI should begin during the hospitalization, with appropriate education, dietary advice, psychosocial counseling, weight loss advice, exercise prescription, cardiac rehabilitation referral (if appropriate), smoking cessation counseling, and the initiation of drug therapy. Given the importance of aggressive risk modification, the entire medical staff has a responsibility to ensure that all of these therapies and advice are offered, encouraged, and established for the future.[78]

Aspirin therapy should be maintained indefinitely, and clopidogrel should be taken for 9 months. For patients who have had an MI or who have left ventricular dysfunction, beta-blocker therapy is recommended indefinitely, in the absence of contraindications. ACE inhibitor treatment is recommended indefinitely for secondary prevention in moderate- to high-risk patients with atherosclerotic disease, diabetes, a low ejection fraction, or other specific indications. Lipid-lowering therapy (e.g., with a statin drug) should be administered if the patient has a low-density lipoprotein level higher than 100 mg/dl post diet (class I, level B), or a high-density lipoprotein level lower than 40 mg/dl (class IIa, level B).[79] Blood pressure should be kept below 140/90 mm Hg unless the patient has renal disease or diabetes, in which case the target is a pressure lower than 130/80 mm Hg.[86] All patients should be prescribed sublingual nitroglycerin and instructed in its use; in particular, they should understand the importance of returning to the hospital immediately if symptoms persist despite three doses of nitroglycerin. Finally, follow-up should take place 2 to 6 weeks after discharge in low-risk and revascularized patients, or 1 to 2 weeks after discharge in higher-risk patients.[1]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Braunwald E, Antman E, Beasley J, et al: ACC/AHA 2002 guidelines update for the management of patients with unstable angina and non–ST-segment elevation myocardial infarction. J Am Coll Cardiol 40:1366, 2002

2. Braunwald E, Antman EM, Beasley JW, et al: ACC/AHA guidelines for the management of patients with unstable angina and non–ST segment elevation myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force guidelines. J Am Coll Cardiol 36:970, 2000

3. Stary HC, Chandler AB, Dinsmore RB, et al: A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis: a report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation 92:1355, 1995

4. Fuster V, Badimon JJ, Chesebro JH, et al: The pathogenesis of coronary artery disease and the acute coronary syndromes. N Engl J Med 326:242, 1992

5. Falk E, Shah PK, Fuster V: Coronary plaque disruption. Circulation 92:657, 1995

6. Gutstein DE, Fuster V: Pathophysiology and clinical significance of atherosclerotic plaque rupture. Cardiovasc Res 41:323, 1999

7. Moons AH, Levi M, Peters RJ: Tissue factor and coronary artery disease. Cardiovasc Res 53:2:313, 2002

8. Little WC, Constantinescu M, Applegate RJ, et al: Can coronary angiography predict the site of a subsequent myocardial infarction in patients with mild to moderate coronary artery disease? Circulation 78:1157, 1988

9. Levin DC, Fallon JT: Significance of angiographic morphology of localized coronary stenoses: histopathologic correlations. Circulation 66:312, 1982

10. Buffon A, Biasucci LM, Liuzzo G, et al: Widespread coronary inflammation in unstable angina. N Engl J Med 347:1:5, 2002

11. Goldstein A, Demetrou D, Grines CL, et al: Multiple complex coronary plaques in patients with acute myocardial infarction. N Engl J Med 343:915, 2000

12. Rioufol G, Finet G, Ginon I, et al: Multiple atherosclerotic plaque rupture in acute coronary syndrome: a three-vessel intravascular ultrasound study. Circulation 106:804, 2002

13. Braunwald E: Unstable angina classification. Circulation 80:410, 1989

14. Canto JG, Shilpak MG, Rogers WJ, et al: Prevalence, clinical characteristics and mortality among patients with myocardial infarction presenting without chest pain. JAMA 283:3223, 2000

15. Canto JG, Fincher C, Kiefe CI, et al: Atypical presentations among Medicare beneficiaries with unstable angina pectoris. Am J Cardiol 90:248, 2002

16. Bertrand M, Simoons M, Fox K, et al: Management of acute coronary syndromes in patients without persistent ST-segment elevation: Task Force Report. Eur Heart J 23:1809, 2002

17. Panteghini M: Acute coronary syndrome: biochemical strategies in the troponin era. Chest 122:1428, 2002

18. Puleo P, Meyer D, Wathen C, et al: Use of a rapid assay of subforms of creatine kinase MB to diagnose or rule out acute myocardial infarction. N Engl J Med 331:561, 1994

19. ACC/AHA guidelines for the management of patients with acute myocardial infarction: 1999 update. American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 34:3:890, 1999

20. Cannon CP, McCabe CH, Stone PH, et al: The electrocardiogram predicts one-year outcome of patients with unstable angina and non-Q myocardial infarction: results of the TIMI III Registry ECG Ancillary Study. Thrombolysis in Myocardial Ischemia. J Am Coll Cardiol 30:133, 1997

21. Predictors of outcome in patients with acute coronary syndromes without persistent ST segment elevation: results from an international trial of 9146 patients. The PURSUIT Investigators. Circulation 101:2557, 2000

22. Antman E, Cohen M, Bernink P, et al: The TIMI risk score for unstable angina/non–ST elevation MI: a method of prognostication and therapeutic decision making. JAMA 284:835, 2000

23. Antman EM, McCabe CH, Gurfinkel EP, et al: Enoxaparin prevents death and cardiac ischemic events in unstable angina/non–Q wave myocardial infarction: results of the thrombolysis in myocardial infarction (TIMI) 11B trial. Circulation 100:1593, 1999

24. A comparison of low molecular weight heparin with unfractionated heparin for unstable coronary artery disease. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q Wave Coronary Events Study Group. N Engl J Med 337:447, 1997

25. Cannon CP, Weintraub WS, Demopoulos LA, et al: Comparison of early invasive and conservative strategies in patients with unstable coronary syndromes treated with the glycoprotein IIb/IIIa inhibitor tirofiban. N Engl J Med 344:1879, 2001

26. Inhibition of the platelet glycoprotein IIb/IIIa receptor with tirofiban in unstable angina and non–Q wave myocardial infarction. PRISM-PLUS investigators. N Engl J Med 338:1488,1998

27. Patel DJ, Holdright DR, Knight CJ, et al: Early continuous ST segment monitoring in unstable angina: prognostic value additional to the clinical characteristics and the admission electrocardiogram. Heart 75:222, 1996

28. Patel DJ, Knight CJ, Holdright DR, et al: Long-term prognosis in unstable angina: the importance of early risk stratification using continuous ECG monitoring. Eur Heart J 19:240, 1998

29. Haverkate F, Thompson SG, Pyke SD, et al: Production of C-reactive protein and risk of coronary events in stable and unstable angina. Lancet 349:462, 1997

30. Morrow DA, Rifai N, Antman EM, et al: C-reactive protein is a potent predictor of mortality independently of and in combination with troponin T in acute coronary syndromes: a TIMI 11A substudy. J Am Coll Cardiol 31:1460, 1998

31. Becker RC, Cannon CP, Bovill EG, et al: Prognostic value of plasma fibrinogen concentration in patients with unstable angina and non–Q wave myocardial infarction (TIMI IIIB trial). Am J Cardiol 78:142, 1996

32. Morrow DA, Rifai N, Antman EM, et al: Serum amyloid A predicts early mortality in acute coronary syndromes: a TIMI 11A substudy. J Am Coll Cardiol 35:358, 2000

33. Biasucci LM, Vitelli A, Liuzzo G, et al: Elevated levels of interleukin-6 in unstable angina. Circulation 94:874, 1996

34. Morrow DA, de Lemos JA, Sabatine MS, et al: Evaluation of B-type natriuretic peptide for risk assessment in unstable angina/non–ST elevation myocardial infarction. J Am Coll Cardiol 41:1264, 2003

35. Cannon CP, Braunwald E: Unstable angina. Heart Disease: A Textbook of Cardiovascular Medicine, 6th ed. Braunwald E, Zipes DP, Libby P, Eds. WB Saunders Co, Philadelphia, 2001, p 890

36. Effects of tissue plasminogen activator and a comparison of early invasive and conservative strategies in unstable angina and non-Q myocardial infarction: results of the TIMI IIIb trial. TIMI IIIB Trial Investigators. Circulation 89:1545, 1994

37. Boden WE, O'Rourke RA, Crawford MH, et al: Outcomes in patients assigned with acute non–Q-wave myocardial infarction randomly assigned to an invasive as compared with a conservative management strategy. Veterans Affairs Non–Q Wave Infarction Strategies. N Engl J Med 338:1785, 1998

38. McCullough PA, O'Neill WW, Graham M, et al: A prospective randomized trial of triage angiography in acute coronary syndromes ineligible for thrombolytic therapy: results of Medicine versus Angiography in Thrombolytic Exclusion (MATE). J Am Coll Card 32:596, 1998

39. Invasive compared with non-invasive treatment in unstable coronary artery disease: FRISC II prospective randomized multicenter study. Fragmin and Fast Revascularisation during Instability in Coronary Artery Disease Investigators. Lancet 354:708, 1999

40. Holmvang L, Clemmensen P, Lindahl B, et al: Quantitative analysis of the admission electrocardiogram identifies patients with unstable coronary artery disease who benefit the most from early invasive treatment. J Am Coll Card 41:905, 2003

41. Morrow DA, Cannon CP, Rifai N, et al: Ability of minor elevations of troponins I and T to predict benefit from an early invasive strategy in patients with unstable angina and non–ST elevation myocardial infarction: results from a randomized trial. JAMA 286:2405, 2001

42. Mahoney EM, Jurkovitz CT, Chu H, et al: Cost and cost-effectiveness of an early invasive vs conservative strategy for the treatment of unstable angina and non–ST-segment elevation myocardial infarction. JAMA 288:1851, 2002

43. Fox KA, Poole-Wilson PA, Henderson RA, et al: Interventional versus conservative treatment for patients with unstable angina or non–ST segment elevation myocardial infarction: the British Heart Foundation RITA 3 randomised trial. Lancet 360:743, 2002

44. Lewis HD, Davis JW, Archibald DG, et al: Protective effects of aspirin against acute myocardial infarction and death in men with unstable angina: results of a Veterans Administration Cooperative Study. N Engl J Med 309:396, 1983

45. Theroux P, Ouimet H, McCans J, et al: Aspirin, heparin, or both to treat acute unstable angina. N Engl J Med 319:1105, 1988

46. Cairns JA, Gent M, Singer J, et al: Aspirin, sulfinpyrazone, or both in unstable angina. N Engl J Med 313:1369, 1985

47. Risk of myocardial infarction and death during treatment with low dose aspirin and intravenous heparin in men with unstable coronary artery disease. The RISC Group. Lancet 336:827, 1990

48. Collaborative overview of randomized trials of antiplatelet therapy—1: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. Antiplatelet Trialists' Collaboration. BMJ 308:81, 1994

49. Randomised trial of intravenous streptokinase, oral aspirin, both or neither among 17,187 cases of suspected acute myocardial infarction. ISIS-2 Investigators. Lancet 2:349, 1988

50. Balsano F, Rizzon P, Violi F, et al: Antiplatelet treatment with ticlopidine in unstable angina: a controlled multicenter trial. Circulation 82:17, 1990

51. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. The Clopidogrel in Unstable Angina to Prevent Recurrent Events (CURE) Trial Investigators. N Engl J Med 345:494, 2001

52. Mehta SR, Yusuf S, Peters RJG, et al: Effects of pretreatment with clopidogrel and aspirin followed by long-term therapy in patients undergoing percutaneous coronary intervention: the PCI-CURE study. Lancet 358:527, 2001

53. Steinhubl SR, Berger PB, Mann T, et al: Early and sustained dual oral antiplatelet therapy following percutaneous coronary intervention. JAMA 288:2411, 2002

54. Inhibition of platelet glycoprotein IIb/IIIa in unstable angina: receptor suppression using integrilin therapy. PURSUIT trial investigators. N Engl J Med 339:436, 1998

55. Randomised placebo-controlled trial of abciximab before and during coronary intervention in refractory unstable angina. CAPTURE Investigators. Lancet 349:1429, 1997

56. Topol EJ, Moliterno DJ, Hermann HC, et al: Comparison of two platelet glycoprotein IIb/IIIa inhibitors, tirofiban and abciximab, for the prevention of ischemic events with percutaneous coronary revascularization. N Engl J Med 344:1888, 2001

57. Effect of glycoprotein IIb/IIIa receptor blocker abciximab on outcome in patients with acute coronary syndromes without early coronary revascularization. GUSTO-IV ACS investigators. Lancet 357:1915, 2001

58. Boersma E, Harrington RA, Moliterno DJ, et al: Platelet glycoprotein IIb/IIIa inhibitors in acute coronary syndrome: a meta-analysis of all major randomized clinical trials. Lancet 9302:189, 2002

59. Effects of tissue plasminogen activator and a comparison of early invasive and conservative strategies in unstable angina and non–Q wave myocardial infarction. Circulation 97:1195, 1994

60. Hirsh J: Heparin. N Engl J Med 324:1565, 1991

61. Theroux P, Waters D, Qiu S, et al: Aspirin versus heparin to prevent myocardial infarction during the acute phase of unstable angina. Circulation 88:2045, 1993

62. Neri SG, Gensini GF, Poggesi L, et al: Effect of heparin, aspirin or alteplase in reduction of myocardial ischemia in refractory unstable angina. Lancet 335:615, 1990

63. Telford AM, Wilson C: Trial of heparin versus atenolol in prevention of myocardial infarction in intermediate coronary syndrome. Lancet 1:1225, 1981

64. Oler A, Whooley MA, Oler J, et al: Adding heparin to aspirin reduces the incidence of myocardial infarction and death in patients with unstable angina: a meta-analysis. JAMA 276:10:811, 1996

65. Antman EM, Cohen M, Radley D, et al: Assessment of the treatment effect of enoxaparin for unstable angina/non–Q wave myocardial infarction: TIMI 11B – ESSENCE meta-analysis. Circulation 100:1602, 1999

66. Klein W, Buchwald A, Hillis SE, et al: Comparison of low-molecular weight heparin with unfractionated heparin acutely and with placebo for 6 weeks in the management of unstable coronary artery disease: Fragmin in Unstable Coronary Artery Disease Study (FRIC). Circulation 96:61, 1997

67. Comparison of two treatment durations (6 days and 14 days) of a low molecular weight heparin with a 6 day treatment of unfractionated heparin in the initial management of unstable angina or non–Q wave myocardial infarction. The FRAXIS Investigators. Eur Heart J 20:1553, 1999

68. Low molecular weight heparin during instability in coronary artery disease. The FRISC Study Group. Lancet 347:561, 1996

69. Goodman SG, Fitchett D, Armstrong PW, et al: Randomized evaluation of the safety and efficacy of enoxaparin versus unfractionated heparin in high-risk patients with non–ST-segment elevation acute coronary syndromes receiving the glycoprotein IIb/IIIa inhibitor eptifibatide. Circulation 107:238, 2003

70. Collet JP, Montelalescot G, Lison L, et al: Percutaneous coronary intervention after subcutaneous enoxaparin pretreatment in patients with unstable angina pectoris. Circulation 103:658, 2001

71. Kereiakes DJ, Grines C, Fry E, et al: Enoxaparin and abciximab adjunctive pharmacotherapy during percutaneous coronary intervention. J Invasive Cardiol 13:272, 2001

72. Lincoff AM, Bittl JA, Harrington RA, et al: Bivalirudin and provisional glycoprotein IIb/IIIa blockade compared with heparin and planned glycoprotein IIb/IIIa blockade during percutaneous coronary intervention: REPLACE-2 randomized trial. JAMA 289:853, 2003

73. Huynh T, Theroux P, Bogaty P, et al: Aspirin, warfarin or the combination for secondary prevention of coronary events in patients with acute coronary syndromes and prior coronary artery bypass surgery. Circulation 103:3069, 2001

74. Yusuf S, Wittes J, Friedman L: Overview of results of randomized clinical trials in heart disease, II: unstable angina, heart failure, primary prevention with aspirin, and risk factor modification. JAMA 260:2259, 1988

75. Held PYS, Furberg CD: Calcium channel blockers in acute myocardial infarction and unstable angina: an overview. Br Med J 299:1187, 1989

76. Lubson JTJ: Efficacy of nifedipine and metoprolol in the early treatment of unstable angina in the coronary care unit: findings from the Holland Interuniversity Nifedipine/metoprolol Trial (HINT). Am J Cardiol 60:18A, 1988

77. Early statin treatment following acute myocardial infarction and 1-year survival. Swedish Register of Cardiac Intensive Care (RIKS-HIA). JAMA 285:430, 2001

78. Fonarow GC, Gawlinski A, Moughrabi S, et al: Improved treatment of coronary heart disease by implementation of a Cardiac Hospitalization Atherosclerosis Management Program (CHAMP). Am J Cardiol 87:819, 2001

79. Third report of the National Cholesterol Education Program Expert Panel on detection, evaluation and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. NCEP Expert Panel. Circulation 106:3143, 2002

80. Effect of an angiotensin-converting enzyme inhibitor, ramipril, on cardiovascular events in high risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000

81. Lincoff AM: Important triad in cardiovascular medicine: diabetes, coronary intervention, and platelet glycoprotein IIb/IIIa receptor blockade. Circulation 107:1556, 2003

82. Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. Bypass Angioplasty Revascularization Investigation (BARI) Investigators. N Engl J Med 335:217, 1996

83. Sedlis SP, Morrison DA, Lorin JD, et al: Percutaneous coronary intervention versus coronary bypass graft surgery for diabetic patients with unstable angina and risk factors for adverse outcomes with bypass: outcome of diabetic patients in the AWESOME randomized trial and registry. J Am Coll Cardiol 40:1555, 2002

84. Morice MC, Serruys PW, Sousa E, et al: A randomized comparison of a sirolimus-eluting stent with a standard stent for coronary revascularization. N Engl J Med 346:1773, 2002

85. Moses JW, Leon MB, Popma JJ, et al: Sirolimus-eluting stents versus standard stents in patients with stenoses in a native coronary artery. N Engl J Med 349:1315, 2003

86. Chobanian A, Bakris G, Black H, et al: The seventh report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. JAMA 289:2560, 2003

XI VALVULAR HEART DISEASE

Brian P. Griffin, m.d.

Valvular heart disease is an important cause of cardiac morbidity in developed countries despite a decline in the prevalence of rheumatic disease in those countries. Valvular heart disease can give rise to stenosis, regurgitation, or a combination of lesions at one or more valves. The more common significant anomalies currently encountered are mitral regurgitation, caused by mitral valve prolapse; aortic stenosis, caused by a congenital bicuspid valve or by senile valvular calcification; and aortic regurgitation, caused by a bicuspid aortic valve or dilatation of the aorta. Valvular lesions can occur as a result of pathologic changes in the valve leaflets or supporting structures (i.e., the chordae or papillary muscles). Ventricular or aortic enlargement can also produce valvular regurgitation as a result of annular dilatation and inadequate leaflet coaptation in the absence of any specific valve pathology. Valvular heart disease tends to progress over time as degenerative changes are superimposed on the primary pathology. Iatrogenic causes of valvular disease are increasingly recognized. Common causes of major valvular lesions are listed [*see Table 1*].[1]

Etiology

congenital disorders

Anomalies in the development of valve cusps are common at the aortic or pulmonary positions. The normal configuration of these valves is tricuspid. Stenosis is the rule when only one cusp develops, whereas bicuspid valves may be stenotic or regurgitant [*see Figure 1*]. Congenital anomalies of the atrioventricular valves are uncommon; the most common abnormality is congenital cleft mitral valve. Valvular abnormalities can be seen in specific developmental syndromes, such as pulmonary stenosis in rubella syndrome and supravalvular aortic stenosis in Williams syndrome.

myxomatous degeneration

Myxomatous degeneration most often involves the mitral or tricuspid valve. In this condition, leaflet tissue, particularly chordal tissue, is abnormally extensible and weak. The affected valves are therefore more likely to prolapse, leading to significant regurgitation. Chordal rupture is common and may precipitate a rapid clinical deterioration from sudden severe regurgitation. The precise abnormality in valvular tissue is unknown but is thought to involve the structural proteins, such as collagen.[2] A familial tendency is often noted in this disease.[3] Inherited connective tissue diseases such as Marfan syndrome produce valvular abnormalities similar to those found in myxomatous degeneration.

rheumatic heart disease

Rheumatic heart disease remains the most common cause of mitral stenosis and a frequent cause of aortic regurgitation. It is the most common cause of multivalvular heart disease. Isolated outbreaks of rheumatic fever continue to be reported in the United States, even in affluent communities.[4] Rheumatic heart disease remains a significant problem in immigrants, especially those from Latin America and Southeast Asia. Rheumatic fever appears to cause valvular heart disease by an autoimmune phenomenon whereby antibodies against streptococcal antigens cross-react with valvular tissue. Valvular involvement can present acutely as a result of edema of valvular tissue. Progressive fibrosis, superimposed calcification, and scarring with retraction of leaflet tissue lead to valvular stenosis, incompetence, or both. The interval between the occurrence of rheumatic fever and clinical manifestations varies, as does the degree of involvement. Both mitral and aortic valves are usually involved.

degenerative disease

Degenerative calcification is a cause of aortic stenosis in the elderly and in patients with renal dysfunction; it results from calcium deposition on the body of the valvular leaflets rather than on the commissures [*see Figure 1*]. Factors found to promote degenerative valvular changes are increasing age, a low body mass index, hypertension, and hyperlipidemia. Histologic changes that simulate atheroma and involve lipid deposition and inflammatory cell infiltration of the leaflets have been described in patients with early degenerative changes in the aortic leaflets. Even mild degenerative changes in the aortic valve have been reported to be adverse prognostic factors.[5] Calcification of the mitral annulus is common in the elderly; it is more common in women than in men and can produce mitral regurgitation. Occasionally, mitral annular calcification extends onto the valvular leaflets, causing stenosis.

endocarditis

Endocarditis usually occurs on previously abnormal valves, although overwhelming sepsis can infect normal valves. The predominant hemodynamic manifestation of endocarditis is valvular regurgitation. Contributory causes of endocarditis include leaflet prolapse (resulting from a large vegetation), leaflet perforation, and chronic scarring of infected tissue. In rare cases, large vegetations lead to valvular stenosis.

coronary artery disease

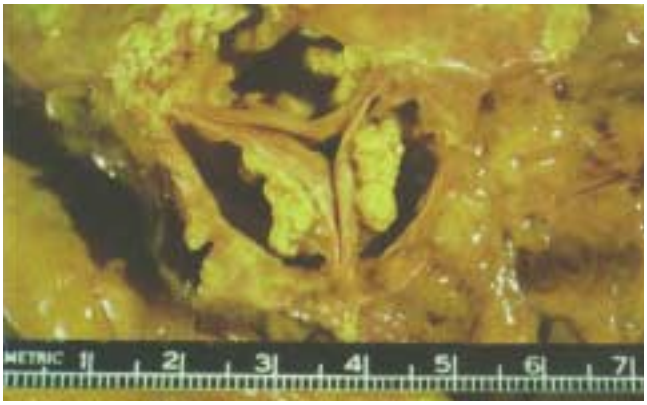
Mitral regurgitation is common in coronary artery disease; it has a number of causal mechanisms. Acute ischemia or infarction

Table 1 Causes of Specific Valvular Lesions

	Mitral	*Aortic*	*Tricuspid*	*Pulmonary*
Stenosis	Rheumatic disease, calcification, SLE	Calcification, congenital disease, rheumatic disease	Rheumatic disease, carcinoid tumor	Congenital disease, carcinoid tumor
Regurgitation	Myxomatous degeneration, ischemia, secondary causes, rheumatic disease, annular calcification, endocarditis, SLE	Congenital disease, secondary causes, rheumatic disease, endocarditis, SLE	Secondary causes, rheumatic disease, endocarditis	Secondary causes

SLE—systemic lupus erythematosus

a



b

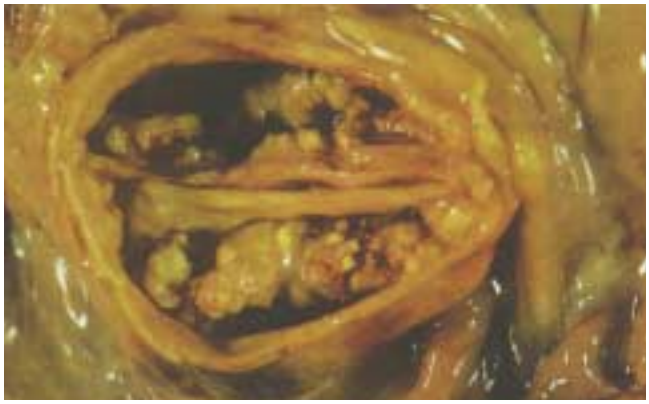


Figure 1 **Pathologic specimens showing degenerative calcification of (*a*) a tricuspid aortic valve and (*b*) a congenital bicuspid valve.[62]**

of a papillary muscle or of the wall to which the papillary muscle is attached leads to impaired leaflet coaptation and mitral regurgitation. Regurgitation can be severe and can vary with the severity of the ischemia. Papillary head rupture or, more rarely, muscle rupture, leads to catastrophic regurgitation that is often fatal.

connective tissue disease

Libman-Sacks endocarditis consists of noninfected warty vegetations involving predominantly the mitral valve; it is characteristic of systemic lupus erythematosus (SLE).[6] Significant regurgitation and stenosis rarely occur acutely but are seen with scarring from chronic disease. Valvular involvement in rheumatoid arthritis is common and leads to valvular thickening but is usually not of hemodynamic significance. Aortitis in ankylosing spondylitis may produce significant aortic regurgitation.

iatrogenic causes of valvular heart disease

Iatrogenic causes include radiation therapy, the use of serotonin agonists such as methysergide, and the use of anorexiants such as fenfluramine and phentermine in combination.[7,8] Radiation leads to scarring and calcification of valvular leaflets many years after the initiating radiation. The effects of both methysergide and anorexiant agents on valvular tissue often simulate rheumatic disease, but regurgitation rather than stenosis predominates as the hemodynamically severe lesion. Serotonin is also thought to play a role in the valvulopathy produced by anorexiants; the precise mechanism by which this occurs remains to be established.

other causes of valvular heart disease

Amyloid disease causes valvular thickening but rarely causes significant stenosis. The carcinoid syndrome most often involves the valves on the right side of the heart and leads to stenosis or incompetence of the tricuspid or pulmonary valve.

secondary involvement

Left ventricular dilatation can cause dilatation of the mitral annulus and, thereby, mitral regurgitation. Common secondary causes of mitral regurgitation include coronary artery disease, aortic valvular disease, and dilated cardiomyopathy. Similarly, tricuspid regurgitation results from right ventricular enlargement secondary to pulmonary hypertension or an atrial septal defect. Dilatation of the ascending aorta, especially involving the annulus of the aortic valve, can lead to aortic regurgitation. This condition is seen in hypertension and in aneurysms of the ascending aorta.

Assessment and Management

Valvular heart disease often remains asymptomatic for many years, but once symptoms develop, survival is reduced if the lesion is not corrected. The assessment of patients with valvular heart disease can be summarized [*see Table 2*]. The evaluation of symptoms can require a stress test or stress echocardiogram in addition to a careful history. Characterization of the lesions and assessment of hemodynamic severity are often possible on physical examination but are aided by additional testing, such as Doppler echocardiography and cardiac catheterization. The auscultatory findings of common valvular lesions also can be summarized [*see Table 3*]. Doppler echocardiography measures the flow velocity across a narrowed valve. By use of the modified Bernoulli equation, the pressure gradient (P), measured in mm Hg, may be estimated from the flow velocity (v), which is measured in m / sec: $P = 4v^2$. Therefore, if the peak velocity recorded across the aortic valve by Doppler echocardiography is 4 m / sec, then the peak pressure gradient will be estimated as $4(4^2)$ or 64 mm Hg. The effects of valvular heart disease on chamber size and function are best assessed serially by echocardiography or, at the time of cardiac catheterization, by ventriculography. In cases of stenotic lesions, intervention is rarely required until symptoms occur. Indications for intervention in regurgitant lesions are more

Table 2 Assessment of Patients with Valvular Heart Disease

Parameters	*Tools*
Symptom severity	History, stress testing
Nature of valve lesion	Auscultation, Doppler echocardiography
Hemodynamic severity of lesion	Physical examination, Doppler echocardiography, cardiac catheterization
Effects of lesion on cardiac chamber size and function	Echocardiography, cardiac catheterization, stress echocardiography
Determination of the optimal time for intervention	Echocardiography, stress echocardiography
Selection of appropriate procedure / prosthesis	Echocardiography

Table 3 Auscultatory Findings Associated with Common Valve Problems

Lesion	*Cardiac Cycle*	*Quality*	*Location*	*Other Sounds*
Aortic stenosis	Systolic, mid-peaking to late peaking	Harsh	Aortic area, left sternal border, apex	Soft S_2, S_4
Aortic regurgitation	Diastolic, early decrescendo	Blowing	Left sternal border, aortic area	—
Mitral stenosis	Diastolic, mid-peaking to late peaking, increases with atrial contraction if rhythm is normal	Rumble	Apex	Opening snap, loud S_1
Mitral regurgitation	Systolic, holosystolic, late systolic with MVP, papillary muscle dysfunction	Blowing	Apex, axilla	Click, soft S_1, S_3
Tricuspid regurgitation	Systolic, increase with inspiration	Blowing	Lower left sternal border, xiphisternum	—
Pulmonary stenosis	Systolic, mid-peaking	Harsh	Pulmonary area, left sternal border	—

MVP—mitral valve prolapse

complex; such indications include significant symptoms or, in the absence of symptoms, increasing ventricular size, overt ventricular contractile dysfunction, or both.

All patients with even mild valvular heart disease require prophylaxis against endocarditis at the time of dental procedures or other procedures that can produce significant bacteremia. The prophylactic regimens recommended by the American Heart Association have been revised [*see Table 4*].[9]

Despite the increase in intravascular volume, pregnancy is usually well tolerated in previously asymptomatic patients with valvular heart disease.[10] During pregnancy, regurgitant lesions are better tolerated than stenosis. Prophylactic intervention to increase the valve area is recommended in patients with hemodynamically severe stenosis before pregnancy.

Patients with hemodynamically significant valvular heart disease should generally avoid participation in competitive sports. Reference should be made to the recommendations of the American College of Cardiology for more information about specific lesions.[11] Valvular heart disease is a chronic disease requiring periodic examination and follow-up, even in asymptomatic patients and in those who have had corrective surgical or other procedures. Patients with prosthetic valves should be seen at least yearly.

Specific Valvular Lesions

mitral stenosis

Normally, the cross-sectional area of the mitral valve is at least 4 cm^2. Mitral stenosis leads to a reduction in valve area and is considered severe when the valve area is less than 1 cm^2. To maintain flow through the valve, left atrial pressure rises, leading to an increase in the pressure gradient across the valve and to increased pulmonary venous and capillary pressures, with resultant dyspnea. Flow through the stenotic valve is dependent on the duration of diastole. Tachycardia shortens diastole disproportionately and causes a further elevation in left atrial pressure and can precipitate symptoms even in patients with relatively mild stenosis. Elevated left atrial pressure contributes to left atrial enlargement, which in turn predisposes the patient to atrial fibrillation, atrial thrombus formation, and thromboembolism, all of which are common complications of mitral stenosis. Severe mitral stenosis is often associated with an increase in pulmonary arterial pressure, leading to right-sided heart failure and secondary tricuspid and pulmonary incompetence. In patients with severe pulmonary hypertension, cardiac output at rest is reduced; this output reduction can cause a relatively low pressure gradient across the mitral valve even in patients with severe stenosis.

Diagnosis

Clinical manifestations Mitral stenosis is often asymptomatic at presentation and for many years thereafter. Symptomatic patients often present with dyspnea, but they can also present with angina, right-sided heart failure, atrial arrhythmia, or embolism. The physical findings in mitral stenosis depend on the severity of the stenosis, the mobility of the valve, and the rhythm. The principal sign is a rumbling diastolic murmur that is best heard at the apex with the stethoscope bell. Such a murmur is accentuated by having the patient lie on the left side and by using provocative maneuvers, such as exercise to increase the heart rate. In sinus rhythm, the murmur increases in intensity with atrial contraction (presystolic accentuation). Increased severity of stenosis is associated with a longer murmur and a thrill. With a pliable valve, an opening sound (the opening snap) is heard, and the sudden closure of the stenotic valve at end diastole gives rise to a loud first heart sound that lends a tapping quality to the apex beat. When the valve calcifies and becomes less mobile, the opening snap and loud first heart sound disappear. A loud pulmonary component of the second heart sound is heard with pulmonary hypertension. The signs and symptoms of mitral stenosis are simulated by left atrial myxoma. In this condition, functional mitral stenosis results from prolapse of a mobile tumor arising from the interatrial septum into the mitral valve opening.

Imaging studies Electrocardiography can reveal left atrial enlargement if the patient is in sinus rhythm. Left atrial enlargement, mitral valve calcification, and signs of pulmonary congestion can all be present on chest x-ray. Doppler echocardiography is the test of choice in confirming the diagnosis, establishing the severity of stenosis, detecting complications, and determining the most appropriate treatment. Echocardiography also allows accurate differentiation of mitral stenosis from a left atrial myxoma.

Typically, the stenotic mitral valve leaflets are thicker and less mobile than normal. The severity of stenosis is determined by measuring the pressure gradient across the valve with Doppler echocardiography and by calculating the valve area. Mitral stenosis should be suspected if the mean gradient exceeds 5 mm Hg; the pressure can exceed 20 mm Hg in severe stenosis. Valve area is measured by tracing the smallest opening of the valve in cross section [*see Figure 2*]. This method is the most accurate way of defining the severity of stenosis, although it is technically demanding and sometimes impossible to perform.[12] The valve area can also be calculated by Doppler echocardiography. Such evaluation is made on the basis of an empirical formula that calculates the time it takes for the pressure gradient to fall to half its initial

Table 4 Summary of American Heart Association Recommendations for Endocarditis Prophylaxis[9]

Procedure	*Patient Condition*	*Drug*	*Regimen**
Dental, oral, respiratory tract, or esophageal†	At risk	Amoxicillin	Adults, 2.0 g; children, 50 mg / kg; orally 1 hr before procedure
	At risk and unable to take oral medications	Ampicillin	Adults, 2.0 g; children, 50 mg / kg. I.M. or I.V. within 30 min before procedure
	At risk and allergic to amoxicillin, ampicillin, and penicillin	Clindamycin *or* Cephalexin‡ or cefadroxil‡ *or* Azithromycin or clarithromycin	Adults, 600 mg; children, 20 mg / kg; orally 1 hr before procedure Adults, 2.0 g; children, 50 mg / kg; orally 1 hr before procedure Adults, 500 mg; children, 15 mg / kg; orally 1 hr before procedure
	At risk and allergic to amoxicillin, ampicillin, and penicillin and unable to take oral medications	Clindamycin *or* Cefazolin	Adults, 600 mg; children, 20 mg / kg. I.V. within 30 min before procedure Adults, 1.0 g; children, 25 mg / kg. I.M. or I.V. within 30 min before procedure
Genitourinary / gastrointestinal	High risk	Ampicillin plus gentamicin	Ampicillin: adults, 2.0 g; children, 50 mg / kg *plus* Gentamicin: 1.5 mg / kg (for both adults and children, not to exceed 120 mg) I.M. or I.V. within 30 min before starting procedure *Then, 6 hr later,* Ampicillin: adults, 1 g; children, 25 mg / kg. I.M. or I.V. *or* Amoxicillin, orally: adults, 1.0 g; children, 25 mg / kg
	High risk and allergic to ampicillin and amoxicillin	Vancomycin plus gentamicin	Vancomycin: adults, 1.0 g; children, 20 mg / kg I.V.; over 1–2 hr *plus* Gentamicin: 1.5 mg / kg (for both adults and children, not to exceed 120 mg) I.M. or I.V. Complete injection / infusion within 30 min before starting procedure
	Moderate risk	Amoxicillin *or* Ampicillin	Adults, 2.0 g; children, 50 mg / kg; orally 1 hr before procedure Adults, 2.0 g; children, 50 mg / kg. I.M. or I.V. within 30 min before starting procedure
	Moderate risk and allergic to ampicillin and amoxicillin	Vancomycin	Adults, 1.0 g; children, 20 mg / kg; over 1–2 hr; complete infusion within 30 min of starting the procedure

*Total children's dose should not exceed adult dose.
†Follow-up dose no longer recommended.
‡Cephalosporins should not be used in patients with immediate-type hypersensitivity reaction to penicillins.
Note: For patients already taking an antibiotic or for other special situations, see reference 9.

value (the pressure half-time). Valve area is estimated as 220 divided by the pressure half-time. Pulmonary arterial (systolic) pressure (PAP) can be determined from the tricuspid regurgitant velocity (TRv) and the estimated right atrial pressure (RAP) (usually estimated as 5 mm Hg) by the following equation: $PAP = 4(TRv)^2 + RAP$. If the tricuspid regurgitant velocity is 3 m / sec, and RAP is estimated to be 5 mm Hg, then the estimated PAP is $4(3^2) + 5 = 41$ mm Hg. The likelihood that the valve may be successfully dilated, either with a balloon or surgically, is estimated by use of a scoring system based on the echocardiographic appearance of the valvular leaflets and supporting structures.

Transesophageal echocardiography is more useful than transthoracic echocardiography in excluding atrial thrombus and determining the severity of mitral regurgitation and is usually performed if balloon valvuloplasty is contemplated. Cardiac catheterization is rarely needed to establish the diagnosis but is used to confirm the severity of stenosis. The valve gradient is the difference between the left atrial pressure or the pulmonary arterial wedge pressure and the left ventricular diastolic pressure. Valve area can be calculated from the pressure gradient and the cardiac output.

Treatment

Once symptoms develop in mitral stenosis, the chance of survival decreases without surgical or balloon dilatation or valve replacement. In the absence of symptoms, management is directed at preventing recurrence of rheumatic fever.[13]

Medical therapy Patients in atrial fibrillation require heart-rate control with a beta blocker (e.g., atenolol, 50 mg q.d.), digoxin (0.125 to 0.25 mg q.d.), or both. Systemic anticoagulation with warfarin is definitely indicated to prevent thromboembolism when (1)

atrial fibrillation is present, (2) there is a history of embolism, or (3) a thrombus is detected in the atrium. Anticoagulation should be considered for patients with paroxysmal atrial fibrillation, a dilated left atrium (> 50 mm in diameter on echocardiography), or severe atrial stasis (as evidenced by swirling echoes or smoke in the left atrium on echocardiography).[14] Regarding symptomatic patients for whom surgical intervention poses a relatively high risk, the judicious use of diuretics and drugs to control heart rate (i.e., digoxin, calcium channel blockers, or beta blockers) may allow symptomatic relief without the need for surgical intervention.

Surgical intervention Intervention to increase valve area is indicated before the onset of symptoms of dyspnea in the following patients: women with severe stenosis who wish to become pregnant but are unlikely to tolerate the volume load of pregnancy, patients who experience recurrent thromboembolic events, and patients who have severe pulmonary hypertension. A number of interventions are currently available to increase the valve area in mitral stenosis. These interventions include percutaneous balloon valvuloplasty, performed in the cardiac catheterization laboratory; surgical commissurotomy; and replacement of the mitral valve with a prosthesis.[15]

Balloon valvuloplasty is performed by inflating a specially designed balloon catheter in the mitral orifice to split the fused commissures. Excellent symptomatic relief is obtained in suitable patients.[16] This intervention is currently the initial choice in mitral stenosis. Typically, the mitral valve area doubles in size from 1.0 to 2.0 cm^2, with a concomitant reduction in the pressure gradient [*see Figure 3*]. Complications of balloon mitral valvuloplasty include severe mitral regurgitation (3%), thromboembolism (3%), and residual atrial septal defect with significant shunting (10% to 20%). Mortality associated with the procedure is less than 1%.[17,18] Contraindications to balloon mitral valvuloplasty include significant mitral regurgitation, which will likely increase after balloon inflation; left atrial thrombus, which can be dislodged at the time of the procedure; and significant subvalvular involvement or leaflet calcification, each of which increases the risk of complications and limits the degree of dilatation produced.[19] In pregnant patients with symptomatically severe mitral stenosis that has not responded to conservative measures such as bed rest and heart-rate control, balloon valvuloplasty is the technique of choice to increase the valve area.[20]

Surgical commissurotomy is now usually performed under direct vision after cardiopulmonary bypass. Surgical commissurotomy may be feasible when balloon valvuloplasty is impossible, such as in patients with significant mitral regurgitation, subvalvular stenosis, or atrial thrombus. A number of studies comparing surgical commissurotomy with balloon commissurotomy have shown equivalent immediate and medium-term (3- to 4-year) results regarding increase in valve area, improvement in symptoms, and freedom from repeat intervention in appropriately selected patients.[19] However, commissurotomy, whether effected by a balloon or surgically, is a palliative procedure, and in most cases further intervention is eventually required. Repeat commissurotomy is sometimes feasible; but most often, mitral valve replacement is also necessary.[21]

A prosthetic replacement is indicated if the valve is heavily scarred or calcified or if severe mitral regurgitation is present. Morbidity and mortality are higher with prosthetic replacement than with either surgical or balloon commissurotomy.

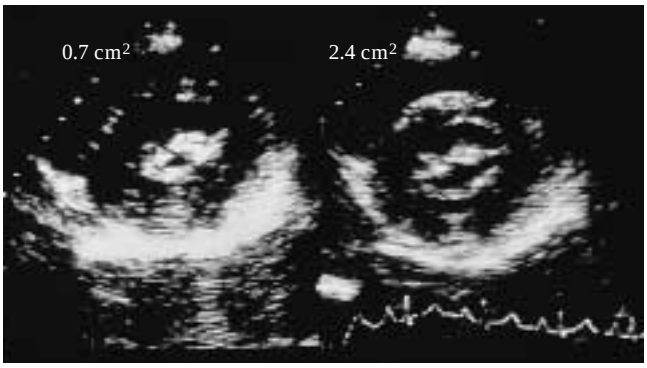


Figure 2 **Two-dimensional echocardiographic parasternal short-axis image of a mitral valve before (left) and after (right) percutaneous balloon mitral valvuloplasty. The valve area is estimated by planimetry and increases from 0.7 cm^2 before valvuloplasty to 2.4 cm^2 after valvuloplasty.**

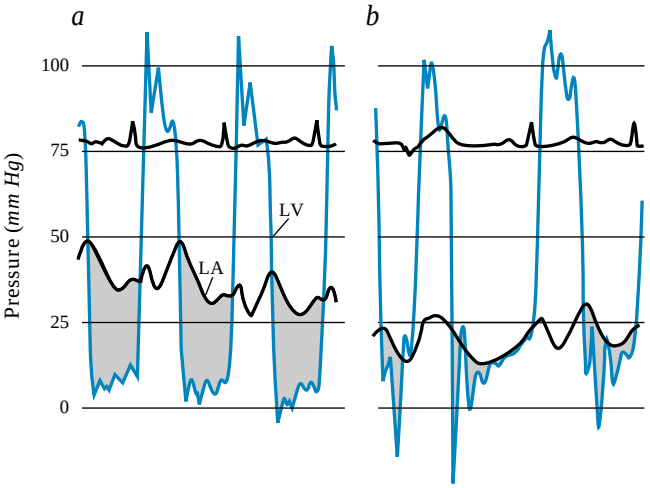


Figure 3 **Simultaneous left atrial pressure (LA, black line) and left ventricular pressure (LV, blue line) are shown (*a*) before and (*b*) after percutaneous mitral valvuloplasty. The shaded area shows the pressure gradient across the mitral valve; the pressure falls after valvuloplasty.**

mitral regurgitation

Mitral regurgitation leads to volume overload of the left ventricle, which must increase in size to achieve a normal stroke output to accommodate the leakage of blood back into the left atrium. Progressive left ventricular dilatation eventually leads to an increase in afterload, contractile impairment, reduction of cardiac output, and heart failure. In acute mitral regurgitation (such as can occur with chordal rupture, ischemia, or endocarditis), left atrial and pulmonary venous and arterial pressures increase quickly, giving rise to dyspnea and, often, acute pulmonary edema. In more chronic forms of mitral regurgitation, an increase in left atrial pressure is often offset by a concomitant increase in atrial compliance; and hence, symptoms appear late in the course of the disease. Left atrial enlargement predisposes the patient to atrial fibrillation and atrial thromboembolism. In long-standing mitral regurgitation, pulmonary hypertension can develop, which in turn leads to tricuspid regurgitation and right-sided heart failure.

Diagnosis

Clinical manifestations In most patients, mitral regurgitation remains asymptomatic for many years. Dyspnea, fatigue

from low cardiac output, and edema occur late in the course of the disease. Mitral regurgitation is recognized clinically by a systolic murmur at the apex, radiating to the axilla and increasing on expiration. In patients with a posteriorly directed jet of mitral regurgitation, the murmur is heard well at the back. In more severe cases, the murmur lasts throughout systole, the first and second heart sounds are soft or difficult to hear, and a third heart sound is present. A midsystolic click can be present in myxomatous disease; in less severe cases, this click can precede the murmur. The murmur can also be confined to late systole, with papillary muscle dysfunction. Mitral regurgitant murmurs caused by ischemia can be variable in duration and intensity, depending on the degree of ischemia and the loading conditions.

Imaging studies Doppler echocardiography is the noninvasive method of choice in confirming the presence of mitral regurgitation. Echocardiography is used to diagnose the mechanism of the regurgitation (e.g., prolapse or annular dilatation); color-flow mapping is used to provide a semiquantitative assessment of severity based on the size and penetration of the left atrium by the regurgitant jet. Additionally, echocardiography can be used to assess the effects of the regurgitation on left ventricular size and function. Quantitative measurements of regurgitation, such as the regurgitant volume, regurgitant fraction ([regurgitant volume plus stroke volume] divided by regurgitant volume), and regurgitant orifice area (the area through which the valve leaks), are now possible with newer Doppler techniques. These techniques are useful in determining the true severity of the lesion and following it over time.[22] Left ventricular size and volume, as well as contractile function assessed by the ejection fraction, are used to determine the need for surgical intervention.

However, asymptomatic mitral regurgitation is more difficult to assess and manage than other valvular lesions because in this condition, the true contractile function of the left ventricle is difficult to determine with conventional measures such as the ejection fraction. These measurements of contractility are confounded by the increase in ventricular preload caused by the extra volume of blood in the left atrium and the variable effect on afterload. Afterload is increased by left ventricular dilatation, but this effect is offset as the ventricle ejects much of its blood into a relatively low pressure system (the left atrium). The left ventricular ejection fraction can appear falsely elevated in mitral regurgitation and usually falls after surgical correction. An ejection fraction of less than 60% should be considered abnormally low in patients with mitral regurgitation.

Transesophageal echocardiography is very sensitive in the detection of mitral regurgitation and is used mainly in those patients who are difficult to evaluate by the transthoracic approach.[23] Contrast ventriculography is used to determine the severity of mitral regurgitation in patients undergoing cardiac catheterization. This procedure involves injecting radiopaque contrast medium into the left ventricle and assessing the extent and duration of opacification of the left atrium. In patients undergoing hemodynamic monitoring, large systolic V waves on the pulmonary arterial wedge tracing raise the suspicion of acute severe mitral regurgitation, as can occur in acute ischemia, but such V waves can occur in the absence of severe regurgitation.

Treatment

Indications for surgery In the management of asymptomatic mitral regurgitation, it should be borne in mind that left ventricular dysfunction is often latent and that, once present, the dysfunction cannot be corrected by operative intervention.[24] Therefore, it is important to refer patients for surgery before the onset of true left ventricular dysfunction even in the absence of symptoms.

Unfortunately, no load-independent measure of contractile function is readily available. Stress echocardiography is useful in detecting latent left ventricular dysfunction not evident on a resting study. Failure of the left ventricular ejection fraction to increase on exercise or of the left ventricular end-systolic volume to decrease on exercise is predictive of incipient left ventricular dysfunction and should be considered an indication for early surgery.[23]

Other echocardiographic indices that have been associated with a less favorable surgical outcome include an absolute end-systolic dimension that is greater than 4.5 cm or that is greater than 2.6 cm/m^2 when indexed for body surface area; an end-systolic volume of greater than 50 ml/m^2; and a resting ejection fraction of less than 60%.[23] Surgical referral should be considered if ventricular size and function approach these indices.[25] Serial echocardiographic evaluation should be performed at least yearly and should be performed more frequently as ventricular dilatation progresses in patients with severe asymptomatic mitral regurgitation. Studies indicate a better long-term survival rate in patients with severe mitral regurgitation when surgery is performed early.[26,27]

Symptomatic severe mitral regurgitation is considered an indication for surgical intervention if the valve is primarily involved. Symptomatic patients with ischemic mitral regurgitation often require mitral valve surgery in addition to revascularization. Mitral regurgitation secondary to left ventricular dilatation often improves with afterload reduction, and surgical intervention is not usually indicated.

Patients with moderately severe or severe left ventricular dysfunction (ejection fraction < 35%) and significant mitral regurgitation were thought in the past to be poor surgical candidates because of high operative risk. However, recent research has shown that there is acceptable risk associated with operations in these patients. Symptoms usually improve, but a survival benefit associated with surgery in this group has not yet been shown.[28,29] Patients who are not considered suitable for surgery because of left ventricular dysfunction often benefit from afterload reduction and diuretics.[30] However, in patients who have primary asymptomatic mitral regurgitation with preserved left ventricular function, afterload reduction has not been shown to delay surgery or improve left ventricular function in the few small studies that have addressed this issue; afterload reduction is not currently recommended to treat such patients.[31] Afterload reduction is beneficial for stabilizing patients with hemodynamically significant acute mitral regurgitation in preparation for surgery.

Surgical intervention Mitral valve repair is currently the technique of choice in the surgical management of mitral regurgitation because the operative mortality is lower, ventricular function is better preserved, and long-term complications such as thromboembolism and infection are lower with repair than with replacement [*see Figure 4*].[32-34] Valve repair is most likely to be feasible in patients with myxomatous disease, especially if such disease involves the posterior leaflet, and is least likely to be successful in patients with rheumatic disease and endocarditis.[34,35] Valve repair is accomplished by use of a variety of techniques, depending on the mechanism and etiology of the regurgitation. Such techniques include partial leaflet resection, chordal shortening or transfer, and insertion of an annuloplasty ring to reduce the size of the annulus. Long-term failure of repair occurs at a rate of 1% to 2% a year but is higher in patients with rheumatic disease. If mitral

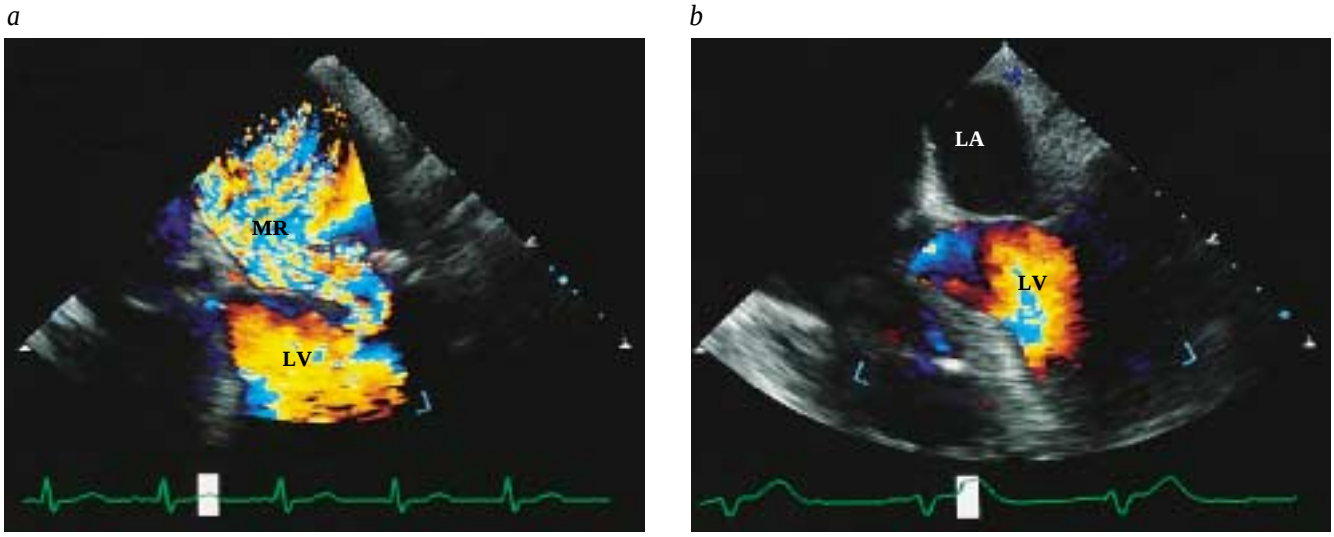


Figure 4 **Transesophageal echocardiogram of a patient with severe myxomatous mitral regurgitation (MR) (*a*) before and (*b*) after mitral valve repair.**

valve repair is not possible, a mitral prosthesis is implanted.[35] Chordal and papillary muscle preservation is increasingly being employed when a mitral prosthesis is inserted, because preserving the muscles has been shown to help conserve left ventricular function after surgery.[36]

mitral valve prolapse

Mitral valve prolapse is a common condition in which the mitral valve leaflets are displaced in systole into the left atrium.[3] It is usually caused by myxomatous degeneration of the valve and can occur in some form in up to 3% of the general population; it is more common in women than in men. In the majority of cases, mitral valve prolapse represents a benign abnormality; in a minority, mainly older men, significant mitral regurgitation results from rupture of a chord or from endocarditis and requires surgical intervention. Mitral valve prolapse is associated with low body weight, low blood pressure, and thoracic skeletal abnormalities such as pectus excavatum. Patients with mitral valve prolapse have a slightly increased risk of stroke, myocardial ischemia, and sudden death. Ventricular extrasystoles are common and can be symptomatically troublesome. Other arrhythmias, such as ventricular or supraventricular tachycardia, are reported but are uncommon.

Diagnosis

Mitral valve prolapse has been associated with multiple nonspecific symptoms, such as atypical chest pain, presyncope, anxiety, and panic attacks. These symptoms are more commonly reported by women than men. A causal relation between these symptoms and mitral valve prolapse has not been established.[3]

A midsystolic click at the mitral area during cardiac auscultation is often the finding that first brings mitral valve prolapse to the attention of the examiner. The click has been attributed to tensing of the redundant valvular tissue with cardiac contraction. A late systolic murmur can follow the click. Maneuvers that reduce intracardiac volume, such as having the patient stand or perform the Valsalva maneuver, cause the click to occur earlier in systole and cause an increase in the duration of the murmur. The typical auscultatory findings and their response to these maneuvers are sufficient to make a diagnosis of mitral valve prolapse. Two-dimensional echocardiography is the method of choice to confirm the diagnosis. Apparent systolic displacement beyond the annular plane is possible with both M-mode and two-dimensional approaches because the annulus is nonplanar and saddle-shaped. The possibility of a false positive diagnosis can be minimized by seeking systolic displacement of the leaflets in a parasternal long-axis view. Myxomatous mitral valve leaflets frequently are thicker, are more redundant, and have longer chordae than normal as seen on echocardiography. Doppler echocardiography is used to detect and quantify associated regurgitation.

Treatment

Asymptomatic mitral valve prolapse requires no specific treatment. Periodic examination is indicated to detect any progression in the severity of mitral regurgitation. Prophylaxis for endocarditis is indicated if both a click and a murmur are present but is not indicated in the absence of mitral regurgitation.[37] Symptomatic ventricular ectopy often responds to beta blockade. Many patients with atypical chest pain and other nonspecific symptoms improve when they are reassured of the relatively benign nature of the condition. Empirical treatment with small doses of beta blockers can also provide symptomatic relief. Mitral regurgitation should be treated as described earlier.

aortic stenosis

The normal aortic valve is 3 to 4 cm^2 in area when fully open. Aortic stenosis is considered severe when the valve area is 1 cm^2 or less and is considered critical when the area is less than 0.75 cm^2. Aortic stenosis causes concentric left ventricular hypertrophy as a compensatory mechanism that maintains cardiac output at rest despite the increased pressure gradient across the valve. Eventually, this compensatory mechanism is overcome, causing the left ventricle to fail and dilate and the resting cardiac output to decline.

Diagnosis

Clinical manifestations There is a variable relation between the severity of stenosis and symptoms. Many patients with critical aortic stenosis are asymptomatic, whereas patients in states of volume overload, such as pregnancy, may have symptoms with

stenosis of lesser severity. Dyspnea is often the presenting feature; it reflects increased left atrial pressure and pulmonary venous hypertension from the increased left ventricular pressure in systole and the diastolic ventricular dysfunction imposed by left ventricular hypertrophy. Angina is common even in the absence of significant obstruction in the epicardial coronary blood vessels because of impaired supply of blood to the subendocardium in the hypertrophied left ventricle. Exertional syncope also occurs with stenosis and can result from the inability to increase cardiac output sufficiently to supply both skeletal muscle and the cerebral vasculature, resulting in impaired cerebral blood supply, or from abnormal baroreceptor reflexes. Serious arrhythmia can also cause syncope and, in severe aortic stenosis, even sudden death. Fatigue is common because of low cardiac output.

In severe aortic stenosis, the carotid pulse typically is reduced in intensity and has a slow delayed upstroke. Aortic stenosis gives rise to a systolic murmur that is heard over the aortic area and that can radiate to the carotid arteries and to the apex. In severe stenosis, the murmur peaks later in systole and can be associated with a thrill. A fourth heart sound is usually present. In mobile congenitally abnormal valves, an ejection click can precede the murmur. Severe calcific aortic stenosis is often associated with a diminished intensity of the aortic component of the second heart sound. Although the physical findings are important in alerting the clinician to the presence of aortic valve disease, the degree of hemodynamic severity is more reliably determined with Doppler echocardiography.

Imaging studies The presence of left ventricular hypertrophy on electrocardiography provides useful supporting evidence for significant aortic stenosis. Doppler echocardiography is used to determine the mechanism and the hemodynamic severity of the stenosis as well as the effects on left ventricular size and function. In aortic stenosis, the opening of the aortic valve is reduced, as seen on the echocardiogram. Continuous wave Doppler echocardiography is used to measure the peak velocity across the valve and thus the aortic pressure gradient; the mean pressure gradient across the valve is often 50 mm Hg or more in patients with severe aortic stenosis. However, the pressure gradient is determined not only by the degree of stenosis but also by flow through the valve and can be relatively low despite severe aortic stenosis if cardiac output is reduced. In most instances, therefore, the valve area as well as the pressure gradient should be calculated. The valve area is estimated readily from the flow through the valve and the pressure gradient across the valve. With cardiac catheterization, the pressure difference across the aortic valve between the left ventricle and the aorta is measured directly. Valve area can be calculated from the cardiac output and the pressure gradient. Because Doppler echocardiography and invasive measurements of aortic valve severity have been shown to agree when both are performed expertly, cardiac catheterization is now used less often as the primary diagnostic tool in assessing aortic stenosis [*see Figure 5*]. Cardiac catheterization is used to confirm the echocardiographic findings in patients being considered for surgery or when there is significant discrepancy between the clinical findings and echocardiographic findings.

Treatment

Indications for surgery Aortic stenosis is a progressive disease, and patients with the disease can remain asymptomatic for many years. The rate of progression varies greatly but increases with age, associated coronary artery disease, and the severity of the stenosis.[38] Progression to symptoms or intervention is likely within 2 years in older patients with severe asymptomatic aortic stenosis.[37] Once symptoms become manifest, the survival rate without surgical treatment is reduced; mean survival is 5 years in patients with angina, 3 years in patients with syncope, and 2 years or less in patients with heart failure.[38] Operative mortality increases with severe symptoms, advanced age, and the presence of left ventricular dysfunction. The onset of symptoms, therefore, is the major indication for surgical intervention. Left ventricular dysfunction attributable to aortic stenosis is another indication for intervention because it demonstrates failure of compensatory mechanisms and incipient symptoms. Sudden death can occur with aortic stenosis, but this is rare in the absence of symptoms. Patients should be instructed to report the onset of any symptoms and should undergo regular follow-up evaluations with physical examination and Doppler echocardiography. Doppler examination should be performed at least yearly and should be performed more frequently in patients with severe stenosis and in older patients. Surgical relief of aortic stenosis usually leads to relief of symptoms and improvement in left ventricular function when such function was abnormal preoperatively.

Aortic valve surgery in the very elderly is associated with an increased mortality but provides excellent palliation of symptoms; surgery should be considered for such patients provided they are otherwise viable candidates.[39] Patients with severe left ventricular dysfunction resulting from aortic stenosis should also be considered for surgery, because significant improvement in ventricular function and symptoms often results, and without surgery the survival rate in these patients is poor.

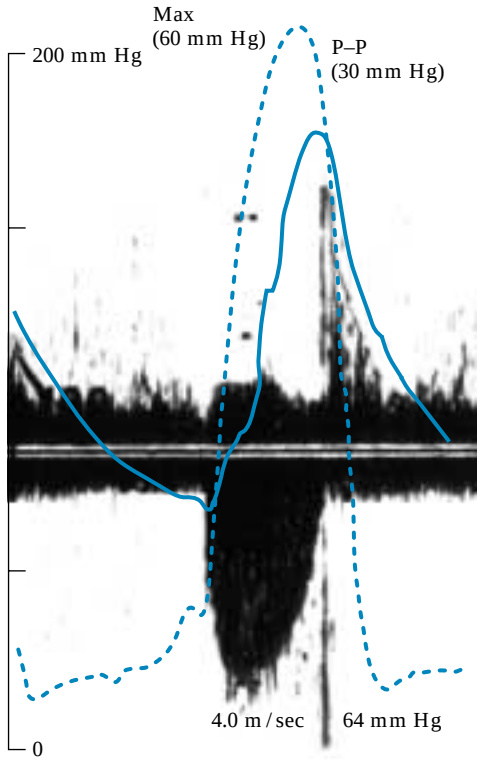


Figure 5 **Simultaneous left ventricular (broken blue line) and aortic (solid blue line) pressure tracings and continuous wave Doppler tracing in a patient with severe aortic stenosis. The pressure gradient (P–P, or 30 mm Hg) is the area between the aortic and LV tracings. Maximal pressure gradient (Max) by cardiac catheterization (60 mm Hg) is similar to that measured by Doppler echocardiography (64 mm Hg).[63]**

Surgical intervention Surgical intervention for patients with aortic stenosis usually involves insertion of a prosthesis or a human valve. In congenital aortic stenosis, valve repair or commissurotomy can be feasible, although significant aortic regurgitation can result. Balloon valvuloplasty has proved disappointing in the long-term treatment of adult calcific aortic stenosis. Valve area typically increases from 0.5 cm^2 to 0.8 cm^2 and is associated with improvement of symptoms in the majority of cases.[18] However, stenosis recurs in as many as 50% of patients within 6 months, and fewer than 25% survive more than 3 years.[40] Balloon valvuloplasty is now indicated in the palliative treatment of adult patients with aortic stenosis who are not surgical candidates because of significant comorbidity; it is also used to stabilize critically ill patients for whom surgery is planned at a later stage. Balloon dilatation is effective in young patients with congenital aortic stenosis and is an alternative to surgery in symptomatic aortic stenosis during pregnancy.

aortic regurgitation

Aortic regurgitation causes a volume overload of the left ventricle. In chronic aortic regurgitation, the volume overload is well tolerated for years. The left ventricle dilates to accommodate the increased volume load and thereby maintains a normal resting cardiac output. Unlike mitral regurgitation, the left ventricle in aortic regurgitation must expel all of the increased volume of blood into the systemic circulation; severe enlargement of the left ventricle is common. Because of a compensatory increase in ventricular compliance, left ventricular diastolic pressure often remains in the normal range despite the increase in ventricular size. The ventricle hypertrophies to maintain normal wall stress. Eventually, compensatory mechanisms fail, and contractile impairment and increased diastolic pressure result in elevated left atrial and pulmonary venous pressures and symptoms. Acute aortic regurgitation can develop as a result of sudden disruption of the valve apparatus with endocarditis or aortic dissection. This condition is poorly tolerated because the left ventricle is unable to dilate fast enough to compensate for the volume load. Left ventricular diastolic pressure rises rapidly and leads to pulmonary congestion and edema. Cardiac output falls, and shock and even death can follow.

Diagnosis

Clinical manifestations In chronic aortic regurgitation, symptomatic presentation occurs late in the course of disease; dyspnea and fatigue are the usual findings. Angina can occur in the absence of coronary artery disease because of the increased demand for oxygen caused by severe left ventricular enlargement and hypertrophy together with the reduced supply of oxygen resulting from the underperfusion of the coronary arteries. Such underperfusion is caused by the low diastolic pressure characteristic of this condition.

The cardinal physical sign of aortic regurgitation is a diastolic murmur that is high pitched and best heard with the diaphragm of the stethoscope with respiration suspended in expiration. The murmur is loudest immediately after aortic valve closure; it progressively diminishes in intensity throughout diastole, paralleling the decline in the pressure gradient between the aorta and the left ventricle. The murmur is best heard on the left of the sternal border in disease of the aortic cusps and on the right of the sternal border in disease of the aortic root. Even in the absence of significant stenosis, an aortic systolic murmur is audible, reflecting the increased flow through the valve. Severe chronic aortic regurgitation is characterized by a wide pulse pressure and an elevated systolic pressure caused by the increased stroke output; also characteristic is a reduction in the diastolic pressure, which occurs as blood leaks back into the left ventricle throughout diastole. If the aortic regurgitant jet hits the mitral valvular leaflet, it can cause partial closure of the valve, creating an apical diastolic murmur that simulates mitral stenosis (Austin Flint murmur). The ejection of a large volume of blood into the systemic circulation and its rapid leak backward into the heart cause many peripheral circulatory manifestations that confirm rather than establish the diagnosis. Acute aortic regurgitation can be more difficult to recognize because the murmur is often short, and the reduced cardiac output leads to reduced intensity of the murmur.

Imaging studies Marked cardiomegaly and prominence of the ascending aorta are often present on chest x-ray in patients with chronic severe aortic regurgitation. Doppler echocardiography confirms the mechanism and severity of aortic regurgitation and its effect on left ventricular size and function. Regurgitant volume and fraction can be quantified by echocardiographic Doppler techniques. More often, the severity of regurgitation is graded on the basis of several qualitative and semiquantitative measures, including the dimensions of the regurgitant jet in the left ventricular outflow tract and of the ventricular cavity, as determined by color flow Doppler mapping, and the presence of diastolic flow reversal in the descending thoracic aorta, as determined with pulsed wave Doppler echocardiography [*see Figure 6*].[41] In severe aortic regurgitation, early closure of the mitral valve and diastolic mitral regurgitation can occur as a result of the increased pressure in the left ventricle in diastole [*see Figure 6*]. Confirmation of the severity of aortic regurgitation is obtained by aortography, a process in which contrast medium is injected into the aortic root and the retrograde filling and clearing of contrast dye from the left ventricle is examined. Aortography is the current gold standard for assessing the severity of aortic regurgitation; it should be performed if there is any discrepancy between the clinical findings and the findings on Doppler echocardiography. Stress ventriculography and echocardiography have both been used to determine the response of the left ventricle to the effects of exercise: a significant fall in left ventricular ejection fraction or an increase in end-systolic volume suggests incipient contractile dysfunction and can be an indication for early surgical intervention.

Treatment

Chronic aortic regurgitation is well tolerated for many years.[42] Operative mortality is increased and long-term survival reduced if the left ventricle is greatly enlarged or if left ventricular dysfunction has been present for more than 1 year. Left ventricular dysfunction that is present for a shorter period is likely to improve and even resolve after surgery. Several studies have shown that asymptomatic patients with normal left ventricular function can be safely followed for a long period (up to 11 years in one study) when serial physical examination and Doppler echocardiographic examination are performed at least yearly and then performed more frequently as left ventricular dilatation progresses.[42]

Surgery is indicated when symptoms develop. In asymptomatic patients, surgery is indicated when resting left ventricular function declines or if severe left ventricular dilatation (end-systolic dimension > 5 cm; end-diastolic dimension = 7 cm) occurs.[43] Evidence suggests that these dimensions should be normalized for body size and that surgery should be considered at an earlier stage, especially in women. Afterload reduction with vasodilators such as hydralazine, captopril, and nifedipine has been shown to

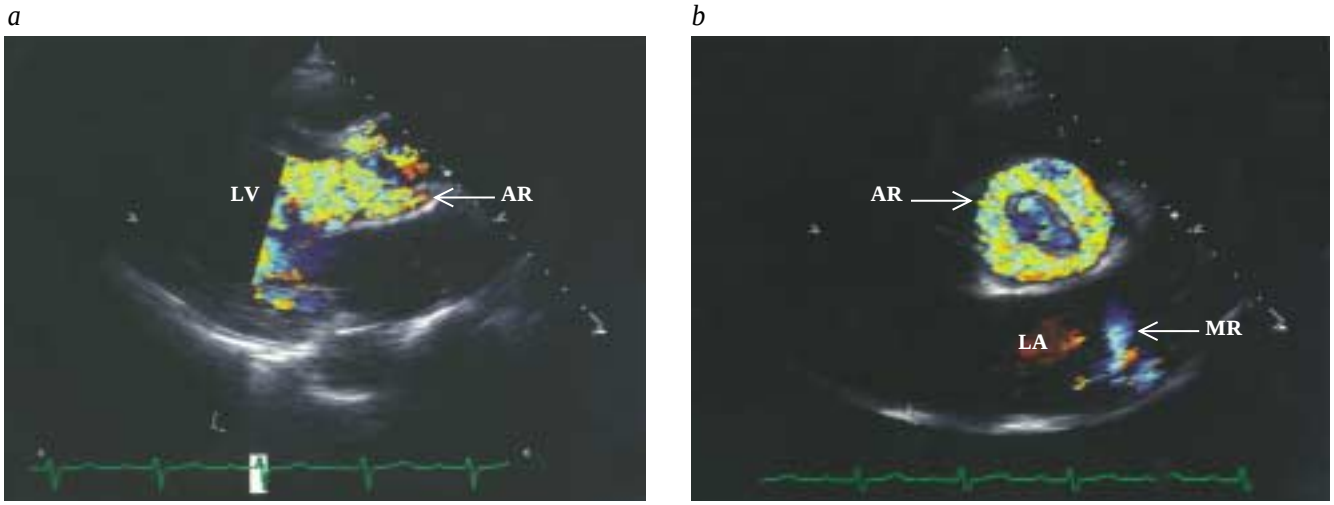


Figure 6 **Parasternal long-axis view (*a*) and short-axis view (*b*) of a severely regurgitant aortic allograft. Aortic regurgitation (AR) is seen circumferentially around the insertion site. Diastolic mitral regurgitation (MR) is also seen.**

reduce regurgitant volume and ventricular size in aortic regurgitation.[31] Treatment with nifedipine, 20 mg twice a day, has been shown to delay the need for surgical intervention in chronic asymptomatic aortic regurgitation but has not been widely used for this indication.[44] Acute severe aortic regurgitation necessitates urgent surgery. Intravenous vasodilatation with sodium nitroprusside or another vasodilator can reduce the regurgitant volume and help stabilize the patient awaiting surgery.

Surgical intervention in aortic regurgitation usually leads to improvement in symptoms and left ventricular size. Although the operative risk is increased when severe left ventricular dilatation or dysfunction is present, significant improvement in symptoms and ventricular function often occurs after surgery; the prognosis without surgery is very poor.[43] Aortic regurgitation usually requires insertion of a prosthesis or a human valve. Occasionally, repair is feasible, especially in prolapsing bicuspid valves or when the aortic ring is dilated.

Tricuspid and Pulmonary Disease

Tricuspid regurgitation is most often secondary to right ventricular dilatation and is the most common valvular problem of the right heart. Tricuspid regurgitation is recognized on physical examination by the characteristic large V waves in the jugular venous pulse and by a systolic murmur heard at the base of the xiphisternum that increases on inspiration. In severe cases, pulsatile hepatomegaly is present. Doppler echocardiography allows rapid detection and assessment of the severity of the regurgitation. Presentation often includes fatigue from reduced forward output and peripheral edema. Severe tricuspid regurgitation is usually treated with surgical repair. If a repair is not possible, a bioprosthesis is usually implanted because of the increased risk of thrombosis of a mechanical prosthesis at this position. Secondary tricuspid regurgitation can improve if the primary condition leading to pulmonary hypertension is treated and leads to a decrease in right heart size.

Tricuspid stenosis occurs in approximately 5% to 10% of patients with severe mitral stenosis. The characteristic physical findings are a large A wave in the jugular venous pressures and a diastolic murmur over the tricuspid area. Doppler echocardiography and right heart catheterization are both used to assess severity. The mean gradient across the tricuspid valve is typically greater than 5 mm Hg. In patients with significant stenosis, either balloon dilatation or surgical repair or replacement is indicated.

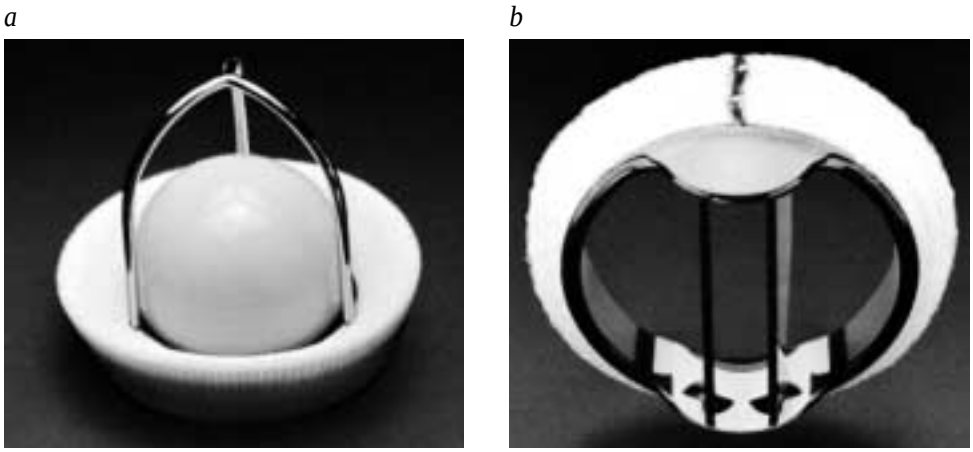


Figure 7 **Two aortic mechanical prostheses. (*a*) Starr-Edwards ball-in-cage prosthesis and (*b*) St. Jude bileaflet tilting-disk prosthesis.**

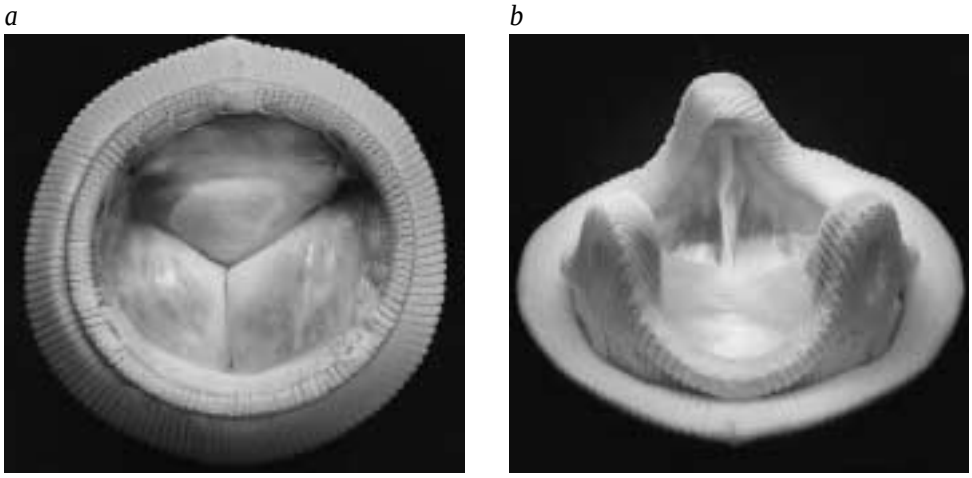


Figure 8 **(*a*) Top view of an aortic porcine xenograft that has been preserved in glutaraldehyde and mounted on a flexible plastic stent. (*b*) Bottom view of the same valve.**

Congenital pulmonary stenosis occurs in isolation or as part of various syndromes and is usually detected before adulthood. Significant pulmonary stenosis is treated with balloon dilatation or surgery. Significant pulmonary insufficiency is rare but can occur with a carcinoid tumor or endocarditis or secondary to pulmonary hypertension. Pulmonary allograft implantation is indicated for severe cases.

prosthetic valves

Prostheses can be classified into two groups—mechanical and biologic—each having different properties, problems, and indications.[45]

Mechanical Prostheses

Mechanical prostheses are of two main types: ball-in-cage and tilting-disk [*see Figure 7*]. The Starr-Edwards valve is the prototypical ball-in-cage valve that has been implanted with various modifications since the 1960s. Tilting-disk valves can consist of one or two leaflets. Single-leaflet models include the Björk-Shiley and Medtronic-Hall valves. The most commonly implanted bileaflet models are the St. Jude valve and the CarboMedics valve. The major advantage of mechanical prostheses is durability. Mechanical prostheses can remain functional for decades and are used especially in young or middle-aged patients to reduce the need for reoperation.[46] Their chief disadvantage is the associated risk of thromboembolism, which necessitates long-term anticoagulation and carries a risk of hemorrhage. An increased incidence of subsequent infection, hemolysis, thrombosis of the valve, and mechanical failure is another problem associated with mechanical prostheses.

Bioprostheses

Three classes of biologic valves are currently available. A xenograft is a prosthesis fashioned from animal tissue [*see Figure 8*]. Most xenografts consist of modified porcine valves that are preserved in glutaraldehyde and mounted on a stent.[47] Prostheses have also been constructed of pericardium and other biologic materials.[48] Stentless bioprostheses are under investigation and are postulated to improve the effective size of the prosthetic valve opening.[49] Allografts (homografts) are human valves that have been harvested postmortem and either cryopreserved or treated with antibiotics.[50] An autograft is a valve from the patient's own body that is moved to a different anatomic site.[51] The most common autograft is the pulmonary valve inserted at the aortic position. A pulmonary allograft is inserted in its place. Patients with biologic valves have a lower risk of thromboembolism than those with mechanical prostheses and do not usually require long-term anticoagulation. Biologic valves are indicated for patients in whom anticoagulation is inappropriate. Xenografts are less durable than mechanical prostheses. Xenograft durability is greatest in patients older than 60 years and improves with age.[52] Xenografts are not usually inserted in patients younger than 60 years because of the poor survival record in this group. Allografts and autografts are alternatives to mechanical prosthetic implantation at the aortic or pulmonary positions in younger patients. Insertion of these valves is technically more demanding and is not widely done. No long-term survival benefit of allografts over xenografts has been demonstrated. Autografts have proved to be durable and have the potential to grow in situ.[51] They are used in the management of pediatric and adolescent patients with aortic valve disease.[53] Both allografts and autografts result in a low reinfection rate when used in the treatment of prosthetic aortic endocarditis and are considered the valve replacement of choice for this condition.[50]

problems and complications of valve prostheses

Thromboembolism

Systemic anticoagulation with warfarin or dicumarol decreases the incidence of, but does not eliminate the occurrence of, thromboembolism with mechanical valves.[54] The incidence of thromboembolic events is lowest in patients younger than 50 years, lower with aortic prostheses than with mitral or multiple prostheses, and lower with bileaflet disk valves than with single-leaflet valves.[54] Hemorrhagic events are more common in older patients. Anticoagulation is generally monitored using the international normalized ratio (INR). Studies have indicated that the level of anticoagulation required to prevent thromboembolism is less than was previously thought. A large study of anticoagulation in patients with mechanical prostheses has suggested that an INR of between 2.5 and 4.0 is desirable in most instances and minimizes hemorrhagic and thromboembolic complications.[54] The appropriate INR for an individual patient will vary depending on the history of embolic or bleeding events; age; and type, position, and number of prostheses. Antiplatelet agents such as aspirin (81 mg

q.d.) or clopidogrel may be added to the anticoagulation regimen in patients who have sustained recurrent thromboembolic events despite adequate anticoagulation. Thromboembolic risk with xenografts is greatest in the first 3 months after surgery.[55] During this period, oral anticoagulation medications are recommended for high-risk patients (e.g., those with mitral prostheses or paroxysmal atrial fibrillation); for patients who are not at high risk, aspirin, 325 mg / day, is recommended.

Valvular Thrombosis

Acute thrombosis of a mechanical valve is more common with a single tilting-disk valve. The incidence is highest at the tricuspid position, followed by the mitral position, and is least common at the aortic position. Thrombosis of left-sided valves can lead to acute pulmonary edema and systemic thromboembolism. Reduced motion of the disk or ball is characteristic of valvular thrombosis and can be demonstrated with transesophageal echocardiography or fluoroscopy.[56,57] There is usually an increased pressure gradient across the valve. Acute thrombosis of a mechanical valve is an indication for emergency surgery to remove the thrombus and to implant another prosthesis. In patients who are not surgical candidates or who are considered at high operative risk, thrombolysis has been used successfully to increase valve opening and motion and reduce the valve gradient. Success rates greater than 70% have been reported in a number of series.[56] Thromboembolism is the most common complication of thrombolysis in this setting and occurs in 12% to 22% of patients. Further episodes of valvular thrombosis after initial successful thrombolysis have been reported.[56]

Valve Failure

In mechanical prostheses, failure of one of the mechanical parts is rare but can have catastrophic consequences. Failure is most common with tilting-disk valves, particularly with the Björk-Shiley single-leaflet tilting-disk valve, which is no longer available commercially in the United States.[58] Failure of the outlet strut in several of these models led to embolization of the disk and acute valve failure, with high morbidity and mortality. Prophylactic repeat surgery has been recommended for certain groups of patients in whom the failure rate is highest.[58] Advanced imaging techniques designed to detect sites of potential strut fractures are currently in development.[59]

Valve failure is expected with bioprostheses and allografts. Fortunately, degeneration is a slow process in biologic prostheses and is usually present for years before significant hemodynamic consequences are seen. Leaflet calcification can give rise to stenosis, whereas cusp degeneration can lead to perforation, with resultant regurgitation. Bioprosthetic degeneration is managed in the same way as stenosis or regurgitation of a native valve. Repeat surgery is indicated for significant symptoms or progressive ventricular enlargement or dysfunction.

Failure of either a mechanical or a biologic prosthesis can occur because of failure of the sutures holding the valve in place. Sutures can fail because of associated infection, but they can also fail spontaneously. A St. Jude valve in which the sewing ring was impregnated with silver nitrate to reduce the likelihood of infection was recalled because of a high incidence of paravalvular leak. The paravalvular leak resulting from suture failure can begin as a relatively mild lesion, but progression is common. In severe instances, partial or complete dehiscence can result in a characteristic rocking motion of the valve, as revealed by echocardiography. Paravalvular leaks are often accompanied by significant hemolysis as red blood cells are destroyed at the site of increased shear stress.[60] Hemodynamically significant paravalvular leaks are considered an indication for reoperation.

Infection

There is an increased risk of endocarditis with mechanical prostheses and xenografts, compared with native valves or allografts. Prosthetic valve endocarditis is often associated with abscess formation. Prosthetic vegetation and abscess formation are best evaluated by using transesophageal echocardiography, which should be performed if a diagnosis of prosthetic valve endocarditis is being considered. Prosthetic valve endocarditis is extremely difficult to eradicate with medical treatment alone; operative intervention is usually required.

Inherent or Acquired Prosthetic Stenosis

All prosthetic valves are inherently stenotic, but in an appropriately selected prosthesis, the degree of stenosis is mild and not of clinical significance. Occasionally, a smaller than desirable prosthesis is implanted because the native valve annulus is small. In such cases, patients can manifest symptoms and signs of valvular stenosis and have severely increased pressure gradients across the valve, especially during exercise. In severe cases, explanation of the prosthesis and annular reconstruction may be necessary to accommodate a prosthesis of sufficient size. In some patients with mechanical prostheses, ingrowth of a fibrous pannus can impede blood flow and lead to hemodynamic stenosis, requiring reoperation.

Problems Associated with Pregnancy

Pregnancy is contraindicated in women with mechanical prostheses because of considerable risk to mother and fetus. The risk to the mother is associated with difficulty in maintaining effective anticoagulation; the risk to the fetus is associated with potential teratogenic effects of warfarin.[61] If possible, valve repair or insertion of an allograft or autograft should be attempted in a woman of childbearing age who wishes to become pregnant. Xenografts are less durable in young patients, especially during pregnancy, and are best avoided.[62] The management of patients with mechanical prostheses who become pregnant or desire pregnancy is controversial. Warfarin is associated with embryopathy and increases the risk of fetal wastage. Optimal anticoagulation is also difficult with heparin, especially when given subcutaneously, and is associated with increased maternal risk for thromboembolism and hemorrhage.[62] Different approaches have been advocated for the management of pregnant patients with mechanical prostheses; self-administration of heparin subcutaneously throughout pregnancy (ideally, from the time of conception) is the preferred approach in the United States.[63] This approach involves administering heparin every 12 hours and keeping the activated partial thromboplastin time at 1.5 to 2.0 times the control value 6 hours after administration, unless low-molecular-weight heparin is used, thereby eliminating the need to monitor the partial thromboplastin time.

anorexiant-induced valvular disorder

Drugs that suppress appetite (anorexiants) have been reported to cause a valvular disorder similar to that caused by ergot derivatives and carcinoid syndrome. This finding was first reported in 1997, and a number of large studies since then have confirmed an increased prevalence of valvular disorders in populations treated with fenfluramine, dexfenfluramine, phentermine, or a combination of these drugs.[7,8] Over 18 million prescriptions were filled for

these drugs in 1996 alone. The precise pathophysiology of the valvular disorder is still unclear. All of these anorexiants affect central serotininergic receptors. A causal relation of serotonin in this disorder is also suggested by the disorder's similarity to carcinoid disease, in which serotonin is also implicated as a causative factor. Initial reports suggested a high prevalence of valvular disease in patients treated with these anorexiants, and they were withdrawn from the market in September 1997. The prevalence of clinically symptomatic valve-related disease in patients receiving these drugs has been reported to be 1 in 1000.[64]

Anorexiant-drug valvulopathy affects mainly the aortic and mitral valve. Leaflet thickening, restricted leaflet motion, chordal thickening, and valve regurgitation without stenosis are the most common abnormalities seen.[65] Although valvular disease severe enough to warrant surgery has been reported, in many instances the valvular lesion appears to be mild or moderate in severity. Factors thought to increase the likelihood of more severe disease are longer duration of treatment with anorexiant therapy, use of drug combinations, and higher dosages of drugs. Patients who received less than 3 months of treatment appear to have a relatively low likelihood of significant valvular disease.[66] Studies also suggest that the valvular lesions may not progress and may even regress after discontinuance of the drug.[67] Patients exposed to anorexiants should undergo a thorough cardiovascular examination for signs of mitral or aortic regurgitation. Echocardiography is indicated if the physical findings suggest valvular disease or if the duration of treatment has been more than 3 months. Patients with evidence of valvular disease on echocardiography should be followed serially and receive prophylactic antibiotics for dental and other procedures associated with significant bacteremia.

References

1. Rose AG: Etiology of valvular heart disease. Curr Opin Cardiol 11:98, 1996

2. Tamura K, Fukuda Y, Ishizaki M, et al: Abnormalities in elastic fibers and other connective-tissue components of floppy mitral valve. Am Heart J 129:1149, 1995

3. Devereux RB: Recent developments in the diagnosis and management of mitral valve prolapse. Curr Opin Cardiol 10:107, 1995

4. Feldman T: Rheumatic heart disease. Curr Opin Cardiol 11:126, 1996

5. Otto CM, Lind BK, Kitzman DN, et al: Association of aortic valve sclerosis with cardiovascular mortality and morbidity in the elderly. N Engl J Med 341:142, 1999

6. Roldan CA, Shively BK, Crawford MH: An echocardiographic study of valvular heart disease associated with systemic lupus erythematosus. N Engl J Med 335:1424, 1996

7. Connolly HM, Crary JL, McGoon MD, et al: Valvular heart disease associated with fenfluramine-phentermine. N Engl J Med 337:581, 1997

8. Gardin JM, Schumacher D, Constantine G, et al: Valvular abnormalities and cardiovascular status following exposure to dexfenfluramine or phentermine / fenfluramine. JAMA 283:1703, 2000

9. Dajani AS, Taubert KA, Wilson W, et al: Prevention of bacterial endocarditis: recommendations by the American Heart Association. JAMA 277:1794, 1997

10. Oakley CM: Valvular disease in pregnancy. Curr Opin Cardiol 11:155, 1996

11. Bonow RO, Carabello BA, de Leon AC, et al: ACC / AHA guidelines for the management of patients with valvular heart disease. Circulation 98:1949, 1998

12. Faletra F, Pezzano A Jr, Fusco R, et al: Measurement of mitral valve area in mitral stenosis: four echocardiographic methods compared with direct measurement of anatomic orifices. J Am Coll Cardiol 28:1190,1996

13. Dajani AS, Taubert K, Ferrieri P, et al: Treatment of streptococcal pharyngitis and prevention of rheumatic fever. Pediatrics 96:758, 1995

14. Gohlke-Burwolf C, Acar J, Oakley C, et al: Guidelines for prevention of thromboembolic events in valvular heart disease: study group of the working group on valvular heart disease of the European Society of Cardiology. Eur Heart J 16:1320, 1995

15. Elliott JM, Tuzcu EM: Recent developments in balloon valvuloplasty techniques. Curr Opin Cardiol 10:128, 1995

16. Palacios IF, Tuzai ME, Weyman AE, et al: Clinical follow-up of patients undergoing percutaneous mitral balloon valvotomy. Circulation 91:671, 1995

17. Dean LS, Mickel M, Bonan R, et al: Four-year follow up of patients undergoing percutaneous balloon mitral commissurotomy: a report from the National Heart, Lung and Blood Institute Balloon Valvuloplasty Registry. J Am Coll Cardiol 28:1452, 1996

18. Orrange SE, Kawanishi DT, Lopez BM, et al: Actuarial outcome after catheter balloon commissurotomy in patients with mitral stenosis. Circulation 95:382, 1997

19. Reyes VP, Raju BS, Wynne K, et al: Percutaneous balloon valvuloplasty compared with open surgical commissurotomy for mitral stenosis. N Engl J Med 331:961, 1994

20. Gupta A, Lokhandwada YY, Satoskar PR, et al: Balloon mitral valvotomy in pregnancy: maternal and fetal outcome. J Am Coll Surg 187:409, 1998

21. Pathan AZ, Mahdi NA, Leon MN, et al: Is redo percutaneous mitral balloon valvuloplasty (PMV) indicated in patients with post-PMV mitral stenosis? J Am Coll Cardiol 34: 49,1999

22. Pu M, Vandervoort PM, Greenberg NL, et al: Impact of wall constraint on velocity distribution in proximal flow convergence zone. J Am Coll Cardiol 27:706, 1996

23. Leung DY, Griffin BP, Stewart WJ, et al: Left ventricular function after valve repair for chronic mitral regurgitation: predictive value of preoperative assessment of contractile reserve by exercise echocardiography. J Am Coll Cardiol 28:1198, 1996

24. Enriquez-Sarano M, Schaff HV, Orszulak TA, et al: Congestive heart failure after surgical correction of mitral regurgitation: a long-term study. Circulation 92:2496, 1995

25. Gaasch WH, John RM, Aurigemma GP: Managing asymptomatic patients with chronic mitral regurgitation. Chest 108:842, 1995

26. Ling LH, Enriquez-Sarano M, Seward JB, et al: Early surgery in patients with mitral regurgitation due to flail leaflets: a long-term outcome study. Circulation 96:1819, 1997

27. Ling LH, Enriquez-Sarano M, Seward JB, et al: Clinical outcome of mitral regurgitation due to flail leaflet. N Engl J Med 335:1417, 1996

28. Bach DS, Bolling SF: Improvement following correction of secondary mitral regurgitation in end stage cardiomyopathy with mitral annuloplasty. Am J Cardiol 78:966, 1996

29. Bolling SF, Pagani FD, Deeb GM, et al: Intermediate term outcome of mitral reconstruction in cardiomyopathy. J Thorac Cardiovasc Surg 115:381, 1998

30. Rosario LB, Stevenson LW, Solomon SD, et al: The mechanism of decrease in dynamic mitral regurgitation during heart failure treatment: importance of reduction of regurgitant orifice size. J Am Coll Cardiol 32:1819, 1998

31. Levine HJ, Gaasch WH: Vasoactive drugs in chronic regurgitant lesions of the mitral and aortic valves. J Am Coll Cardiol 28:1083, 1996

32. Enriquez-Sarano M, Schaff HV, Orszulak TA, et al: Valve repair improves the outcome of surgery for mitral regurgitation: a multivariate analysis. Circulation 91:1022, 1995

33. Grossi EA, Galloway AC, Miller JS, et al: Valve repair versus replacement for mitral insufficiency: when is a mechanical valve still indicated? J Thorac Cardiovasc Surg 115: 389, 1998

34. Stewart WJ: Choosing the "golden moment" for operation in the era of valve repair for mitral regurgitation. American College of Cardiology Heart House Learning Center Highlights 10:2, 1995

35. Gillinov AM, Cosgrove DM, Lytle BW, et al: Reoperation for failure of mitral valve repair. J Thorac Cardiovasc Surg 113:467, 1997

36. Corin WJ, Sutsch G, Murakami T, et al: Left ventricular function in chronic mitral regurgitation: preoperative and postoperative comparison. J Am Coll Cardiol 25:113, 1995

37. Oakley CM: Management of valvular stenosis. Curr Opin Cardiol 10:117, 1995

38. Otto CM, Burwash IG, Legget ME, et al: Prospective study of asymptomatic valvular aortic stenosis: clinical, echocardiographic, and exercise predictors of outcome. Circulation 95:2262, 1997

39. Tseng EE, Lee CA, Cameron DE: Aortic valve replacement in the elderly: risk factors and long-term results. Ann Surg 225:793, 1997

40. Wang A, Harrison JK, Bashore TM: Balloon aortic valvuloplasty. Prog Cardiovasc Dis 40:27, 1997

41. Dolan MS, Castello R, St. Vrain JA, et al: Quantitation of aortic regurgitation by Doppler echocardiography: a practical approach. Am Heart J 129:1014, 1995

42. Bonow RO: Chronic aortic regurgitation: role of medical management and optimal timing of surgery. Cardiol Clin 16:449, 1998

43. Klodas E, Enriquez-Sarano M, Tajik AT, et al: Aortic regurgitation complicated by extreme left ventricular dilation: long-term outcome after surgical correction. J Am Coll Cardiol 27:670, 1996

44. Scognamiglio R, Rahimtoola SH, Fasoli G, et al: Nifedipine in asymptomatic patients with severe aortic regurgitation and normal left ventricular function. N Engl J Med 33:689, 1994

45. Vongpatanasin W, Hillis LD, Lange RA: Prosthetic heart valves. N Engl J Med 335: 407, 1996

46. Zellner JL, Kratz JM, Crumbly AS, et al: Long-term experience with the St Jude Medical valve prosthesis. Ann Thorac Surg 68:1210, 1999

47. Cohn LH, Collins JJ Jr, Rizzo RJ, et al: Twenty-year follow up of the Hancock modified orifice porcine aortic valve. Ann Thorac Surg 66(6 suppl):S30, 1998

48. Banbury MK, Cosgrove DM 3rd, Lytle BW, et al: Long-term results of the Carpentier-Edwards pericardial aortic valve: a 12-year follow up. Ann Thorac Surg 66(6 suppl):S73, 1998

49. Walther T, Falk V, Langebartels G, et al: Prospectively randomized evaluation of stentless versus conventional biological aortic valves: impact on early regression of left ventricular hypertrophy. Circulation 100(19 suppl):II6, 1999

50. Ross DN: Evolution of the homograft valve. Ann Thorac Surg 59:565, 1995

51. Chambers JC, Somerville J, Stone S, et al: Pulmonary autograft procedure for aortic valve disease: long-term results of the pioneer series. Circulation 96:2206, 1997

52. Milano A, Guglielmi C, DeCarlo M, et al: Valve-related complications in elderly patients with biological and mechanical aortic valves. Ann Thorac Surg 66(6 suppl):S82, 1998

53. Lupinetti FM, Warner J, Jones TK, et al: Comparison of human tissues and mechanical prostheses for aortic valve replacement in children. Circulation 96:321, 1997

54. Cannegieter SC, Rosendaal FR, Wintzen AR, et al: Optimal oral anticoagulant therapy in patients with mechanical heart valves. N Engl J Med 333:11, 1995

55. Heras M, Chesebro JH, Fuster V, et al: High risk of thromboemboli early after bioprosthetic cardiac valve replacement. J Am Coll Cardiol 25:1111, 1995

56. Binder T, Baumgartner H, Maurer G: Diagnosis and management of prosthetic valve dysfunction. Curr Opin Cardiol 11:131, 1996

57. Barbetseas J, Nagueh SF, Pitsavos C, et al: Differentiating thrombus from pannus formation in obstructed mechanical prosthetic valves: an evaluation of clinical, transthoracic and transesophageal echocardiographic parameters. J Am Coll Cardiol 32: 1410, 1998

58. Kallewaard M, Algra A, Defauw J, et al: Prophylactic replacement of Bjork-Shiley convexo-concave valves at risk of strut fracture: Bjork-Shiley Study Group. J Thorac Cardiovasc Surg 115:577, 1998

59. O'Neill WW, Chandler JG, Gordon RE, et al: Radiographic detection of strut separations in Björk-Shiley convexo-concave mitral valves. N Engl J Med 333:414, 1995

60. Garcia MJ, Vandervoort PM, Stewart WJ, et al: Mechanism of hemolysis with mitral prosthetic regurgitation: a study using transesophageal echo and fluid dynamic simulation. J Am Coll Cardiol 27:399, 1996

61. Chan WS, Anand S, Ginsberg JS: Anticoagulation of pregnant women with mechanical heart valves: a systemic review of the literature. Arch Intern Med 160:191, 2000

62. Weyman AE: The left ventricular outflow tract. Principles and Practice of Echocardiography, 2nd ed., rev. Weyman AE, Ed. Lea & Febiger, Philadelphia, 1994, p 513

63. Currie PJ, Seward JB, Reeder GS, et al: Continuous wave Doppler echocardiographic assessment of severity of calcific aortic stenosis: a simultaneous Doppler-catheter correlative study in 100 adult patients. Circulation 71:1162, 1985

64. Jick H, Vasilakis C, Weinnarich LA, et al: A population based study of appetite suppressant drugs and the risk of cardiac valve regurgitation. N Engl J Med 339:719, 1998

65. Weissman NJ, Tighe JF Jr, Gottdiener JS, et al: An assessment of heart-valve abnormalities in obese patients taking dexfenfluramine, sustained-release dexfenfluramine, or placebo. N Engl J Med 339:725, 1998

66. Jick H: Heart valve disorders and appetite-suppressant drugs. JAMA 283:1738, 2000

67. Weissman NJ, Tighe JF Jr, Gottdiener JS, et al: Prevalence of valvular-regurgitation associated with dexfenfluramine three to five months after discontinuation of treatment. J Am Coll Cardiol 34:2088, 1999

XII DISEASES OF THE AORTA

Kim A. Eagle, m.d.
William F. Armstrong, m.d.

The Normal Aorta

The normal aorta is composed of three distinct layers: the inner intima, an elastic middle layer called the media, and a thin outer layer called the adventitia. In the media, layers of elastic elements intertwine with collagen and smooth muscle cells, providing the elastic strength that enables the aorta to withstand the pulsatile stress produced by the ejection of blood during ventricular systole. During systole, the aorta is distended by the force of blood ejected into the lumen. The kinetic energy of the ejected blood is transmitted to the wall of the aorta. In diastole, the potential energy stored in the aortic wall is transformed to kinetic energy as it propels the blood forward in the aorta and to its branches. With age, the normal elastic elements of the aorta degenerate, reducing its elasticity and distensibility. The aorta is considered to consist of three anatomic segments: the ascending aorta, the aortic arch, and the descending aorta. The ascending aorta consists of the aortic annulus, the aortic cusps, the sinuses of Valsalva, the sinotubular ridge, and the tubular portion of the ascending aorta. The ascending aorta connects the cardiovascular outflow tract at the aortic valve to the aortic arch, which begins at the brachiocephalic artery. The arch provides branches to the head and neck vessels, coursing just in front of the trachea and then proceeding to the left of the esophagus and the trachea. The descending aorta begins in the posterior mediastinum at the ligamentum arteriosum and courses in front of the vertebral column as it descends to the bifurcation of the leg vessels.

Aortic Aneurysms

Aneurysms may occur at any location in the aorta but are most common in the abdominal segments. Aortic aneurysm is a potentially life threatening entity for which both effective screening and curative therapy are available.

abdominal aortic aneurysms

An aorta is considered aneurysmal when its diameter exceeds 1.5 times the expected normal diameter at any location along its length. Aneurysms are divided into those that affect the abdominal cavity and those that affect the thoracic cavity. More extensive aneurysms (termed thoracoabdominal) involve both aortic areas. In addition, aneurysms are defined as either fusiform or saccular.

Aneurysms of the abdominal aorta are more common than thoracic aortic aneurysms. Among the risk factors for aneurysms, perhaps the most important is age. The incidence of aneurysms increases in men older than 55 years and in women older than 70 years. Overall, men are four to five times more likely to develop aortic aneurysms. Additional risk factors are hypertension, smoking, elevated cholesterol, and a family history suggesting a genetic predisposition to aneurysms.[1] Several reports show that aneurysms develop in as many as 25% of first-degree relatives of patients with abdominal aortic aneurysms.[2] The infrarenal aorta is the most commonly affected region.

screening for abdominal aortic aneurysms

Current recommendations are for noninvasive screening of patients of appropriate age, which is typically defined as older than 65 years but younger if there is a significant family history of or risk factors for aneurysms. Screening may be particularly effective for obese patients, in whom abdominal palpation is of limited value. A large-scale (67,900 patients) randomized trial of ultrasound screening in men 65 to 74 years of age demonstrated a substantial reduction (i.e., 43%; 95% CI, 22 to 58; $P = 0.0002$) in aneurysm-related death with a strategy of routine testing.[3] The cost-effectiveness of various screening strategies has yet to be demonstrated.[4] Careful abdominal palpation is probably cost-effective, particularly in men older than 55 years who are at risk for developing vascular disease. A related issue concerns which patients should undergo noninvasive imaging when the abdominal examination is difficult to perform.

Clinical Presentation

Most abdominal aortic aneurysms produce no symptoms and are discovered during a routine physical examination or as a result of noninvasive screening. The most common symptom is pain, often described as a steady, gnawing discomfort in the lower back or hypogastrium. Generally, the pain is not affected by movement.

In some patients, the abdominal aortic aneurysm is first discovered during a period of rapid expansion or an impending rupture, which is often marked by severe discomfort in the lower abdomen or back, radiating to the buttocks, groin, or legs. Rupture is accompanied by the abrupt onset of back and abdominal pain, abdominal tenderness, the presence of a palpable pulsatile mass, hypotension, and shock. However, only one third of aneurysms present in this fashion. Of note, a ruptured aneurysm may mimic other conditions, including abdominal colic, renal colic, diverticulitis, and gastrointestinal hemorrhage. Not surprisingly, more than 25% of patients presenting with rupture or expansion of an aortic aneurysm are initially misdiagnosed.

Patients with impending or actual rupture must be managed as a surgical emergency in a manner similar to that used for patients with major trauma. Such patients rapidly experience hemorrhagic shock, manifested by peripheral vasoconstriction, hypotension, mottled skin, diaphoresis, oliguria, disorientation, and cardiac arrest. Patients with retroperitoneal rupture may show evidence of hematomas on the flank and in the groin. Although rare, rupture into the duodenum may present as massive upper or lower gastrointestinal hemorrhage.

Diagnostic Evaluation

Physical examination The abdominal aorta is usually detectable on deep palpation, particularly in thin persons. In obese patients, the normal aortic impulse may not be palpable. Obese patients may harbor a large aneurysm without any symptoms or findings on physical examination, unless the aneurysm is exerting pressure on an adjacent structure. Thin patients, in contrast, often feel a pulsatile mass in the abdomen when an abdominal aneurysm has developed.

When palpable, an aneurysm will be identified as a pulsatile mass extending from as high as the xiphoid process to the

suprapubic area. Because of the layers of tissue between the examiner's fingers and the aneurysm, measurements of the transverse diameter of the aneurysm are typically overestimated. Also, it is difficult to differentiate ectatic aorta from aneurysm. Some aneurysms are sensitive to palpation and may be tender if they have recently expanded or are in impending rupture. Thus, palpation should be done with consideration of patient discomfort. Patients with aneurysms often have evidence of other peripheral vascular disease, such as femoral bruits and poor peripheral pulses.

Imaging studies Several diagnostic tools can help identify and measure the size of abdominal aortic aneurysms. For years, aortography was considered the gold standard of diagnostic techniques for evaluating aortic aneurysms. One advantage of aortography is that it can be used to evaluate associated iliofemoral disease and involvement of the renal and mesenteric branches of the aorta. However, aortography is invasive and requires intravascular contrast, which carries a risk of nephrotoxicity. Its use has declined with the development of abdominal ultrasonography, computed tomography, and especially magnetic resonance angiography.

Abdominal ultrasonography is the most frequently used method and the most practical.[5] Ultrasonography has a sensitivity of nearly 100% for diagnosing aneurysms of significant size and can discriminate size to within ± 3 mm. Ultrasonography is inexpensive and noninvasive but may be inadequate for evaluating the most superior or inferior extent of an aneurysm and is generally considered inadequate as a sole diagnostic technique for planning surgical resection.

CT can discriminate aneurysm size to within ± 2 mm. Because CT scanning can determine the inferior and superior extent of the aneurysm and its shape, this method is more useful for planning surgical repair. However, the need for radiographic contrast is a relative disadvantage. When a CT image is compared with an image derived from abdominal ultrasonography, the size of the aneurysm determined by CT is larger by approximately 2.7 mm.[6] New diagnostic techniques such as fast spiral CT have improved the resolution of CT scanning. Magnetic resonance angiography can be successfully used for both screening and surgical planning. It identifies the size and extent of an aneurysm with a high degree of accuracy.

Anatomic landmarks are easily distinguished in the three-dimensional images created with MR angiography, which correctly defines the distal and proximal extent of an aneurysm in more than 75% of the cases examined.[7]

Management to Reduce Risk of Aneurysm Rupture

Current management of abdominal aortic aneurysm is directed at reducing the risk of rupture by intervening with timely surgical resection. Natural history studies show that the likelihood of rupture is greatest in patients with symptomatic, large, or rapidly expanding aneurysms. Aneurysms smaller than 4 cm in diameter have a low (< 2%) risk of rupture. Aneurysms exceeding 10 cm in diameter have a 25% risk of rupture over 2 years. Because aneurysms tend to expand with time, current strategies call for identifying and observing aneurysms that are asymptomatic and are small enough to have a low risk of rupture. The median rate of expansion is slightly less than 0.5 cm a year.[8] However, the tendency for expansion is variable and may not be linear. The more rapidly expanding aneurysms are more likely to rupture than are the stable aneurysms. Aneurysms larger than 6 cm in diameter are generally referred for surgery, whereas aneurysms less than 4 cm in diameter are generally watched.[9-11] Evidence of expansion, particularly if the diameter of the aneurysm has exceeded 5.0 to 5.5 cm, is often taken as an indication to operate.[12] Current data support careful observation and serial noninvasive testing of patients with aneurysms between 4.0 and 5.5 cm in diameter.[13]

Surgical Treatment

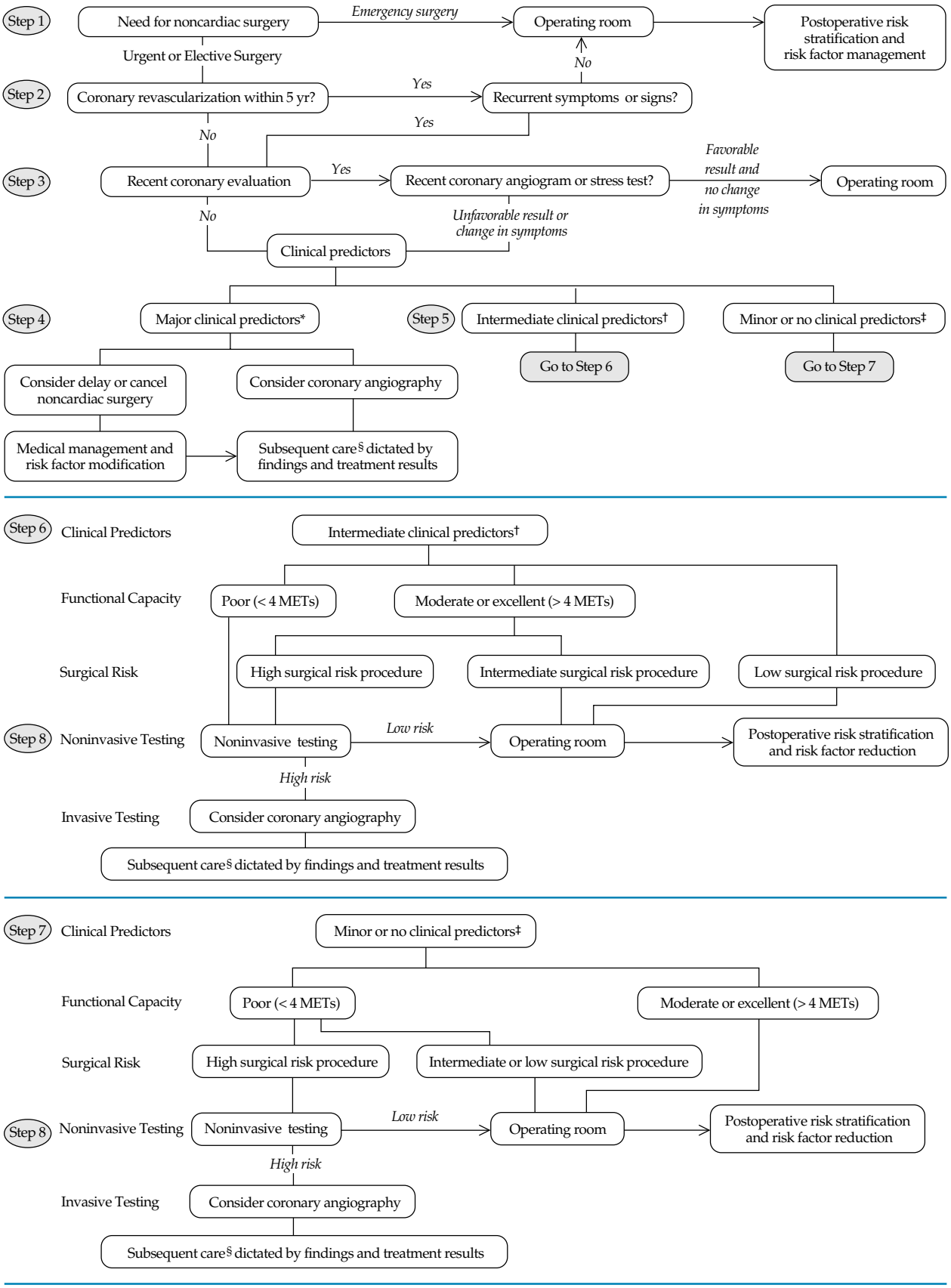
Surgical treatment consists of resection of the aneurysm with insertion of a synthetic (Dacron) graft. Additional distal surgery is often necessary, with resection and interposition of grafts into one or both iliac arteries. For most large aneurysms, the aneurysm wall is left intact, and the Dacron graft is placed inside the aneurysm. The surgical treatment of abdominal aneurysms carries an average operative mortality of 4% to 6%. Surgical mortality is 2% in low-risk patients but may be as high as 20% in patients with impending rupture. For patients in shock with aneurysm rupture who require emergency surgery, operative and perioperative mortality may be as high as 50%.

A therapeutic option is percutaneous placement of implantable endovascular stents, which are similar to those used in patients with coronary artery, renal artery, and peripheral artery stenoses. Some centers are using endovascular stents in nearly 50% of patients referred for treatment of abdominal aortic aneurysms.[14] Larger stents have been used successfully to isolate abdominal aortic aneurysms in patients for whom the risk of surgical resection is unacceptable. However, widespread application of stenting awaits further evaluation of long-term outcomes.[15]

Preoperative evaluation and management Appropriate preoperative evaluation and management of a patient before undergoing elective aortic aneurysm resection are critical. Reports suggest that one third to two thirds of perioperative deaths can be attributed to coronary artery disease. A guideline published by the American College of Cardiology and the American Heart Association reviewed the literature regarding preoperative assessment and presents a simple algorithm to help determine which patients should be considered for preoperative noninvasive testing for coronary disease [*see Figure 1*].[16]

The first consideration is whether the vascular surgery is urgent or emergent. By definition, emergent surgery cannot be delayed, and risk will be higher. In either case, the usual medical approach is to assume the patient may have preexisting coronary disease. Unless contraindicated, beta blockers should be used to treat such patients. Ideally, beta blockers should be started days to several weeks before surgery, titrating the dose to achieve a target heart rate of 50 to 60 beats a minute.[17] The clinical status, electrocardiographic findings, and hemodynamics of these patients should be monitored carefully after surgery.

For determination of perioperative risk, the first issue to address is whether the patient has had a recent coronary revascularization. If the patient has had coronary bypass surgery within the past 6 years and no subsequent coronary symptoms, the risk of perioperative events is relatively low. A second issue is whether the patient has had a recent coronary evaluation. Further preoperative testing is not usually required for patients whose recent stress test or coronary angiogram indicates minimal or no coronary disease, particularly if the evaluation was performed within the previous 2 years and the patient has undergone no change in status.



* Major clinical predictors: unstable coronary syndrome, decompensated CHF, significant arrhythmias, severe valvular disease.
† Intermediate clinical predictors: mild angina pectoris, prior MI, compensated or prior CHF, diabetes mellitus.
‡ Minor clinical predictors: advanced age, abnormal ECG, rhythm other than sinus, low functional capacity, history of stroke, uncontrolled systemic hypertension.
§ Subsequent care of patient may include cancellation or delay of surgery, coronary revascularization followed by surgery, or intensified care.

Figure 1 **Stepwise approach to cardiac assessment. (CHF—congestive heart failure; METs—metabolic equivalents; MI—myocardial infarction)**

Other patients with known prior coronary disease (prior myocardial infarction [MI] or angina), diabetes, or prior congestive heart failure should be more thoroughly evaluated. If such patients have poor functional capacity and have not undergone recent coronary evaluation, they should undergo a preoperative stress test to evaluate the severity of coronary disease and to determine the status of left ventricular function. When possible, exercise is generally the preferred method of stress testing[18] and appears to be safe in most patients. For patients who are unable to exercise, pharmacologic stress testing with either dobutamine echocardiography or adenosine thallium imaging is appropriate. Risk of cardiac events is directly related to the presence and extent of left ventricular (LV) dysfunction and ischemia.

The relative risk of perioperative cardiac morbidity or mortality is low (1% to 5%) in patients with no inducible ischemia and without evidence of fixed perfusion defects or wall motion abnormalities. In patients with extensive areas of ischemia or prior infarction detected during preoperative testing, perioperative event rates (death and MI) may be as high as 20% to 40%. Such patients should probably undergo coronary angiography and possibly coronary revascularization before undergoing major operative procedures.

Although the indications for coronary bypass surgery or percutaneous coronary interventions are generally the same for the preoperative patient and the general population, evaluation for potential heart disease before aneurysm resection may be the patient's first such evaluation. Coronary artery disease must be treated to the fullest extent before undertaking a potentially stressful noncardiac operation on the aorta.

Postoperative modification of risk factors A frequently forgotten issue in the management of patients undergoing abdominal aortic aneurysm resection is long-term modification of cardiovascular risk factors. The preoperative period represents an excellent opportunity to identify and treat hypertension, diabetes, hypercholesterolemia, smoking, obesity, and poor functional status. All patients identified as having vascular disease should take aspirin daily to prevent long-term cardiovascular events. Often, beta blockers are prescribed for patients with coronary artery disease, and the cholesterol profiles of such patients should be routinely assessed. Studies suggest that secondary prevention of vascular disease is enhanced by aggressive treatment of hypercholesterolemia, particularly in persons with a low-density lipoprotein cholesterol level exceeding 100 mg/dl. Currently, the best evidence suggests that the broad class of statin drugs (3-hydroxy-3-methylglutaryl coenzyme A [HMG-CoA] reductase inhibitors) are effective. Beta blockers have been championed as therapy both to reduce risk of MI and to potentially reduce the risk of expansion of aneurysms that may develop or be present elsewhere in patients who previously had significant aneurysms.

THORACIC AORTIC ANEURYSMS

Thoracic aortic aneurysms are less common than abdominal aneurysms. They are classified according to the involvement of the ascending aorta, the descending aorta, or a combination of the two. Aneurysms of the descending aorta are the most common. The etiology of thoracic aneurysms correlates with their location. Aneurysms of the ascending aorta are usually associated with cystic medial necrosis. This association is particularly common in patients with Marfan syndrome, Ehlers-Danlos syndrome, and annuloaortic ectasia, which represents the loss of elastic tensile strength in the aorta. Descending thoracic aortic aneurysms are often seen in hypertensive patients with extensive atherosclerosis. They usually originate beyond the left subclavian artery and may be either fusiform or saccular. Aneurysms of the arch are often contiguous with aneurysms of the ascending or descending thoracic aorta.

Clinical Presentation

More than half of thoracic aortic aneurysms are symptomatic; the rest are discovered only incidentally, often after a routine chest x-ray. Symptoms usually reflect pressure on a contiguous structure or consequences such as concomitant aortic insufficiency. Local mass effects may include a superior vena cava syndrome, caused by obstruction of the superior vena cava; pressure on the trachea, leading to cough or wheezing; and, occasionally, dramatic hemoptysis, resulting from fistula formation between the aneurysm and a major airway. Pressure on the esophagus may produce dysphagia. Pressure on the recurrent laryngeal nerve may result in hoarseness from vocal chord paralysis. Chest pain is usually caused by direct pressure of the aneurysm on an intrathoracic structure or by erosion of a bony structure. Normally, this pain is steady and often severe. Rarely, aortitis may first present as an aortic aneurysm.[19]

A leaking or ruptured aneurysm usually presents with dramatic symptoms. Most such aneurysms leak or rupture into the left pleural space or pericardial space, resulting in hypotension and sudden onset of severe pain. Aortoesophageal fistulas may produce life-threatening gastrointestinal bleeding.

Diagnostic Evaluation

Physical examination The thoracic aorta is generally not palpable unless there is a significant pathologic process. Most often, this pathologic process consists of an ascending aortic arch aneurysm, and the aortic impulse can be palpated just above the sternum or at the right upper sternal border.

Imaging studies The diagnosis of thoracic aortic aneurysms is rarely suspected on physical examination. It is more often initially suspected on chest x-ray and then confirmed with noninvasive or invasive imaging. On chest x-ray, most aneurysms appear as a widening of the mediastinal silhouette. Small aneurysms may not be detected. MRI and spiral CT scanning are the most commonly used methods for delineating the size and extent of thoracic aneurysms. Transthoracic echocardiography (TTE) and transesophageal echocardiography (TEE) are also used to diagnose, measure, and monitor ascending aortic aneurysms. TTE can evaluate only the proximal 3 to 5 cm of the ascending aorta, and neither TTE nor TEE is useful for evaluating aneurysms below the diaphragm.

Management to Reduce Risk of Aneurysm Rupture

The natural history of a thoracic aneurysm can shed light on the disease process that has led to the aneurysm, on the risk factors that may affect the rate of aneurysm expansion, and on the concomitant presence of other vascular disease, including peripheral and coronary disease, that might affect long-term survival. Because size is a critical issue in terms of the risk of rupture, the initial size and potential growth of an aneurysm are important factors in the decision whether to operate on asymptomatic aneurysms. Aneurysms that are invading local structures or creating a marked vascular effect should usually be resected. Careful control of blood pressure is crucial for all pa-

tients and may require medical therapy, particularly with beta blockers, which may also slow the rate of aneurysm growth.[20]

The initial size of a thoracic aneurysm is an important predictor of subsequent growth. In general, small aneurysms tend to grow slowly, whereas large aneurysms have a higher probability of expansion and rupture. On average, thoracic aneurysms grow at 0.4 cm/yr, but the growth rate varies greatly.[21] Small aneurysms (i.e., < 5 cm in diameter) grow at about 0.1 cm/yr. Large aneurysms (i.e., > 5 cm) grow at about 0.5 to 1.0 cm/yr. Although these average growth rates are reassuring, it should be emphasized that rapid expansion can occur and can dramatically affect the natural history and management. In general, thoracic aneurysms smaller than 5 cm in diameter are unlikely to rupture, whereas those larger than 7 cm are at high risk for rupture. Currently, most thoracic centers recommend surgery for aneurysms that exceed 5.5 to 6 cm in an otherwise reasonable surgical candidate.[22,23] Because of their relatively young age, absence of associated disease, and low surgical risk of elective repair, patients with Marfan syndrome should undergo surgery when aneurysms reach 5 cm, particularly if the aneurysm is expanding. Some centers wait until aneurysms reach 6.5 or 7 cm before operating on high-risk surgical candidates. As in the case of treatment of abdominal aneurysm, the use of percutaneously placed aortic stent grafts may emerge as an attractive option in some patients with thoracic aneurysms.[24]

Surgical Treatment

The surgical approach to thoracic aortic aneurysms depends on the site. For ascending aortic aneurysms, the major issue is whether the aortic valve is competent and whether reimplantation of the coronary arteries will be necessary. With the availability of aortic homografts and stentless valves, surgical approaches to thoracic aortic aneurysm are undergoing rapid evolution. Individual patient characteristics and surgical preferences have a great deal to do with a given surgical approach.

Postoperative Complications

Neurologic sequelae are the most serious of potential postoperative complications. Currently, the risk of stroke after thoracic aneurysm resection ranges from 3% to 7%.[25] Efforts to reduce diffuse brain injury caused by prolonged periods of aortic cross clamping include hypothermic arrest and the use of retrograde cerebral perfusion by way of a superior vena cava cannula.[26] Efforts to reduce CNS embolic events focus on meticulous surgical technique to avoid dislodging atheroemboli present in the aortic margins and to avoid air embolism during surgery. The above issues are especially pertinent in aneurysms of the ascending aorta and the arch. Surgery on the posterior thoracic aorta carries a different neurologic risk—namely, postoperative paraplegia as a result of interrupting the supply of arterial blood to the spinal cord—and occurs in more than 5% of patients. Several methods have been devised to deal with this risk, but no definitive solution has yet emerged. Some centers have suggested that reattaching critical intercostal arteries leads to improved outcome,[27] whether or not the spinal cord is treated under epidural cooling during surgery.[28]

Aortic Dissection

The incidence of recognized aortic dissection in the United States is estimated to be 10 to 20 per million population, or about 5,000 cases a year. We stress, however, that the incidence of MI is greater than 500,000 cases annually; that is, MI is at least 100 times more common than aortic dissection. For most patients, dissection entails a tear in the intima, with the subsequent development of a propagating hematoma between the intima and the adventitia. Approximately two thirds of aortic dissections are initiated by a tear in the intima just above the aortic valve. Most of the remaining cases develop in the descending aorta at the attachment of the ligamentum arteriosum. Often, multiple communication sites are present between the true lumen and the false lumen. The dissection often spirals as it courses retrograde or antegrade along the aorta. Approximately 10% to 15% of aortic dissections are caused by intramural hematoma, which is spontaneous rupture of the vaso vasorum within the media, creating a hematoma in the media. This hematoma may extend a variable distance and may eventually rupture into the lumen, resulting in a more typical dissection.

CLASSIFICATION

Aortic dissections are classified as acute or chronic and according to their location. Dissections are termed acute when they are diagnosed within 2 weeks after the onset of symptoms; dissections diagnosed after 2 weeks of symptom onset are termed chronic. A key feature for classification is involvement of the ascending aorta, regardless of where the dissection began. Ascending aortic dissections are also called type A dissections. Dissections not involving the ascending aorta are typically classified as distal, or type B, dissections. Ascending aortic involvement identifies a patient population with high mortality if not treated surgically. A subset of patients with isolated aortic arch dissection has also been described. Normally, the life-threatening condition is caused by communication of the ascending aorta with the pericardial space, creating cardiac tamponade, or by spontaneous rupture or hemorrhage, leading to shock.

The predisposing factors for type A and type B dissections differ somewhat. Disorders of the media that result in cystic medial necrosis are a common precursor of type A dissection. Affected patients may include those with Marfan syndrome or other heritable disorders, such as Ehlers-Danlos syndrome, Noonan syndrome, and Turner syndrome. Another risk factor for ascending aortic dissection is aortic valve disease, such as bicuspid valve disease or prosthetic aortic valve disease. Although these conditions are classically associated with aortic dissection, over 90% of patients with acute aortic dissection do not have any recognized substrate for dissection. Distal, or type B, aortic dissection is most often seen in patients with longstanding hypertension. Patients with type B dissection are older on average than patients with type A dissection. An unexplained relation between aortic dissection and pregnancy also exists, perhaps because of changes in cardiac output, blood pressure, or blood volume or the effects of pregnancy on the aortic wall itself.[29] Aortic dissection after inhalation of crack cocaine has also been reported.

CLINICAL PRESENTATION

The most common distinguishing clinical feature of aortic dissection is the abrupt onset of pain.[30] The abruptness of onset is one of the clinical features reliably distinguishing the pain of aortic dissection from that accompanying other cardiovascular pathology (e.g., myocardial ischemia). This instantaneous pain may begin in the chest or back and may migrate to involve the neck, head, back, and legs as the dissection propagates. The classic combination of abrupt tearing pain, with pulse deficits and

apparent aortic insufficiency, is seldom observed in actual practice.[30,31] Other presentations of type A dissection are sudden syncope or hypotension, resulting from dissection into the pericardial space; stroke, resulting from interruption of the blood supply to one or both internal carotid arteries; and, in rare instances, isolated congestive heart failure, when the dissection involves the ascending aorta and interrupts aortic valve function.

The most typical presentation of type B dissection is onset of severe interscapular pain, which may radiate down the back toward the legs. Type B dissection is frequently accompanied by hypertension, whereas type A dissection more often occurs in the presence of normal or low blood pressure.[30] Spinal cord ischemia, ischemic extremities, and mesenteric ischemia are most frequently encountered in type A dissection that has extended to involve the descending aorta. Whereas aortic insufficiency is noted on auscultation in 35% to 50% of the cases of ascending aortic dissection, it is rather unusual in cases of type B dissection. Pulse deficits are seen in about 25% of patients with type A dissection and in perhaps 5% to 10% of patients with type B dissection.[31]

Acute dissection remains a highly lethal entity. Mortality is commonly quoted as 1% per hour for the first 24 hours. Advanced age, hypotension, and limb and visceral ischemia are all predictors of greater mortality.[32,33] A published review of 500 patients with acute type A dissection identified a number of clinical factors that are predictors of death [*see Figure 2*]. Mortality in the study cohort ranged from 10% to 18%, depending on the number of adverse risk factors.[29]

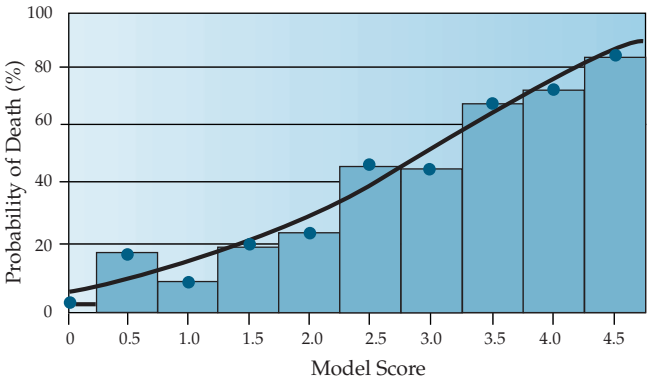


Figure 2 **Graphic demonstration of the increasing mortality in type A dissection when multiple risk factors are present. There is an observed increase in mortality in type A dissection that parallels that of the predictive model. The total risk score was the sum of individual risk factors that was determined from regression analysis to be significantly linked to outcome. The individual factors and their individual scores (in parentheses) were as follows: renal failure (1.6), hypotension/shock/tamponade (1.1), abrupt onset of pain (1.0), pulse deficit (0.7), abnormal ECG (0.6), age ≥ 70 (0.5), female (0.3). (bars—observed findings; line—model probabilities)**

DIAGNOSTIC EVALUATION

Because acute aortic dissection is a life-threatening emergency, rapid and accurate diagnosis is crucial to patient survival. Therefore, sophisticated imaging modalities may be required. Routine ECG in patients with suspected aortic dissection usually reveals only nonspecific abnormalities. Although type A dissection will affect one of the coronary arteries and lead to a transmural MI in 1% to 2% of patients, most patients have nonspecific ST-T wave changes or a finding of left ventricular hypertrophy related to long-standing hypertension.

The typical chest x-ray reveals widening of the mediastinal silhouette and may also demonstrate evidence of a pleural effusion, cardiomegaly, or congestive failure if severe aortic regurgitation is present. A normal-appearing chest x-ray is seen in more than 10% of documented acute aortic dissections.[30] Other laboratory abnormalities are generally nonspecific. An increase of smooth muscle myosin is present in more than 85% of patients presenting within 3 hours after onset of acute aortic dissection.[34] This serum assay, if further developed, may become a useful adjunctive tool to early assessment of suspected aortic dissection.

After a careful history and physical examination, the key to diagnosis is rapid identification of the aortic dissection, ascertainment of whether the ascending aorta is involved, and urgent cardiac surgery if proximal aortic dissection is diagnosed. The importance of rapid diagnosis and institution of definitive therapy for aortic dissection cannot be overemphasized. Given the 1% to 2% mortality per hour in the first 24 hours after presentation, even brief delays to achieve diagnostic imaging are unacceptable.[30,35]

Currently, four diagnostic tools are used to evaluate patients with suspected dissection[36]: CT scanning, echocardiography, MRI, and aortography. In general, the choice of which imaging modality to initially employ will depend on local expertise and availability. In most hospitals, the choice is either CT or TEE.

CT scanning is widely available in most community and tertiary care hospitals. Spiral or ultrafast CT scanning gives even greater resolution than the older scanners and has a reported sensitivity and specificity for aortic dissection exceeding 95%.[37] TEE offers significant advantages in diagnosis [*see Figures 3 and 4*].[38] The primary attractiveness of TEE is its portability, making it suitable for performance in the emergency department, intensive care unit, or operating room. Thus, imaging can be achieved substantially faster with TEE than with other modalities.[39] Second, TEE is highly sensitive for the identification of type A dissection. TEE is also potentially useful when involvement of the aortic valve and the status of the left ventricle, pericardial space, and right and left coronary artery ostia are unknown.[40]

TEE can be very useful in detecting the mechanism of aortic insufficiency and detecting the feasibility of repair.[41,42] Valves in which aortic insufficiency is the result of sinotubular dilatation or extension of the dissection into the sinus are often candidates for repair. Patients with intrinsic disease of the aortic valve leaflets are less optimal candidates for repair.

MRI is less commonly used unless the MRI scanner is part of the emergency department. For most hospitals, however, the delay required in getting a patient into the MRI suite and completing the study makes this technology less efficient than TEE or chest CT.

Finally, although aortography is still used in some hospitals, it is seldom the initial test for aortic dissection. The reported false negative rate for aortography is in the range of 5% to 15%.[43] Aortography frequently misses lesions such as an intramural hematoma. In addition, the time required to get a patient to an angiography suite and complete the study is generally considerably longer than that for TEE. Our medical center and many others follow an algorithmic approach to evaluation and treatment of a suspected aortic dissection [*see Figure 5*].

TREATMENT

The treatment of aortic dissection includes aggressive medical therapy for all patients and definitive surgical therapy in selected patients. The decision to perform surgery depends first and foremost on the site of the aortic dissection [*see Figure 5*].

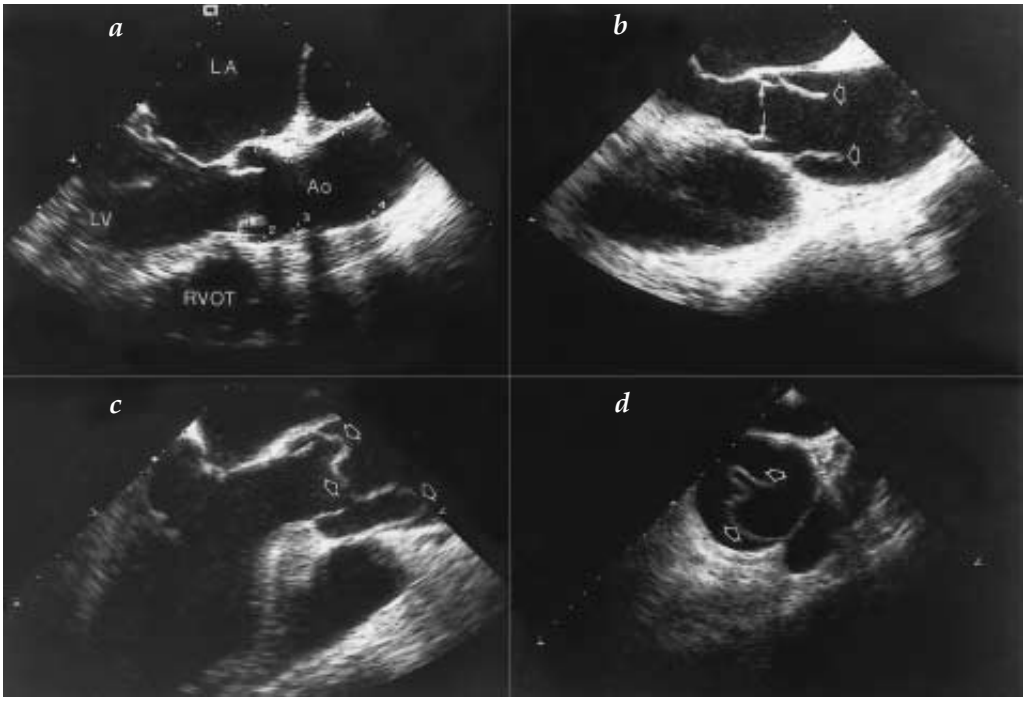


Figure 3 **Transesophageal echocardiograms from a patient with a normal ascending aorta (panel A) and three different ascending aortic dissections (panels B through D). In the normal ascending aorta, the cardiac chambers are noted. The ascending aorta is well visualized, including its annulus (point 1), the coronary sinuses (point 2), the sinotubular junction (point 3), and the true ascending aorta (point 4). Note that the aorta dilates at the level of the sinuses, narrows at the sinotubular junction to a dimension equivalent to that of the annulus, and then slightly dilates further in the ascending aorta. Shown is a normal aortic valve in its open position. Panel B was recorded in a patient with a proximal aortic dissection. The orientation is identical to that in panel A. The solid arrows denote the position of an open aortic valve leaflet. The open arrows represent the margins of a dissection that originated at the sinotubular junction and extended distally. Panel C was recorded in a patient with an ascending aortic dissection (orientation identical to that in panels A and B), and the aortic valve is open. In this instance, a convoluted intimal flap (open arrows) is clearly visualized in the proximal ascending aorta. Panel D was recorded in the short axis of the aorta in a patient with an aortic dissection. In the circular ascending aorta, multiple convolutions of an intimal flap are clearly visualized (open arrows). Note that a communication point (between the downward pointing arrow and the wall of the aorta) allows free communication of flow between the two lumens. (LA—left atrium; Ao—ascending aorta; RVOT—right ventricular outflow tract; LV—left ventricle)**

Surgical Repair

Type A aortic dissection Any involvement of the ascending aorta carries with it a much greater risk of rupture into the pericardial space; development of coronary or cerebral ischemia, aortic regurgitation, and congestive heart failure; or free rupture of the aorta into the thorax. Thus, definitive surgical repair is carried out as quickly as possible for patients with proximal or type A aortic dissection who are appropriate candidates for the procedure.

For patients with type A dissection complicated by malperfusion, medical therapy plus percutaneous reperfusion utilizing aortic stenting or fenestration, or both, and selective branch stenting may allow stabilization and reduce risk associated with the operation. After a period of recovery, repair of the patient's ascending aorta may be undertaken.[44,45]

Definitive aortic repair includes resection of the dissected aorta and insertion of a conduit. The procedure often includes implanting a prosthetic aortic valve. Repair and resuspension of the aortic valve have proved feasible in many patients. For most patients, repair includes reimplantation of the coronary arteries. In some patients, this repair includes resection and placement of a graft to the aortic arch. Even in the best of centers, surgical mortality will range from 10% to 35%, depending on comorbidity.[30,32,46]

Type B aortic dissection Surgery for type B dissection is indicated for patients with life-threatening complications that require a surgical approach. Examples include patients who experience ischemia of both kidneys, leading to reversible renal failure; development of ischemic bowel; ischemia involving one of the legs or arms; development of a progressive aneurysm; impending rupture; and recurrent extension of the dissection. In some centers, percutaneous insertion of aortic stents has been used to stabilize dissections of the descending aorta. This strategy may be preferable to surgery in some candidates.[47-49] In particular, stenting may promote thrombosis of the false channel and thereby reduce the long-term risk of aneurysm formation and aortic rupture. Surgical placement of an endoprosthesis, or so-called elephant trunk, has also been advocated as a preferred strategy for operative type B dissection.[50,51]

Postoperative Complications

In the management of aortic dissection, surgical complica-

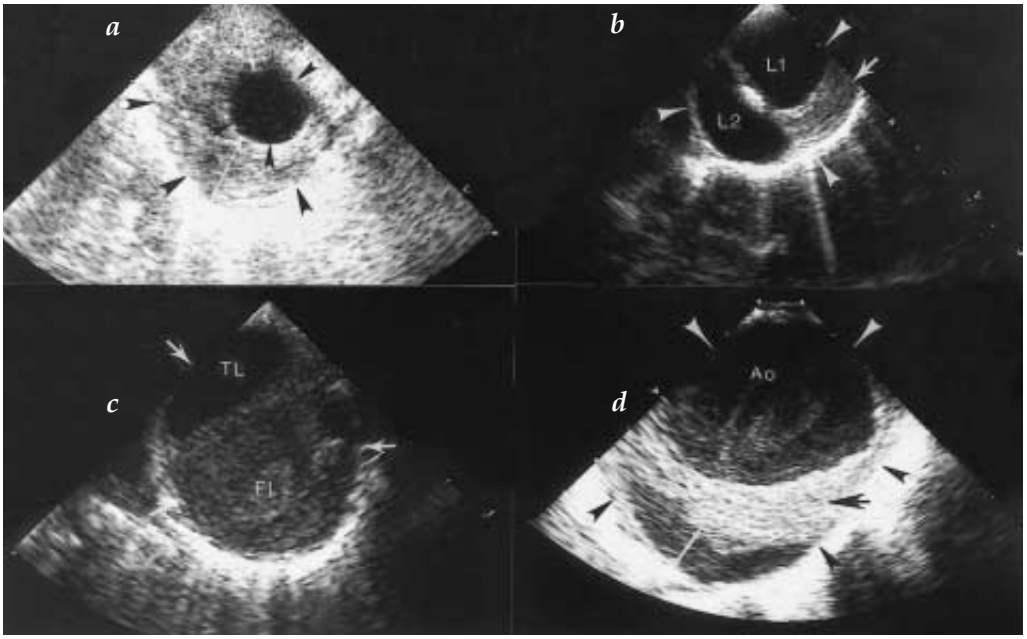


Figure 4 **Panels A through D represent four transesophageal echocardiograms recorded in a short-axis view of the descending thoracic aorta in patients with aortic pathology. Panel A was recorded in a patient with an ascending aortic aneurysm and a large periaortic (adventitial) hematoma extending distally along the thoracic aorta. The smaller black arrows denote the boundaries of the normal-diameter descending thoracic aorta. The larger black arrows pointing inward mark the full dimension of the periaortic hematoma; the full dimension is also noted by the double-headed white arrows. In this instance, the intima of the descending thoracic aorta was not involved in the dissection process. However, a large periadventitial hematoma ruptured along the course of the descending thoracic aorta. Panel B was recorded in a patient with an aortic dissection localized to the descending thoracic aorta. The maximum external dimensions of the aorta are noted by the large white arrowheads. The white arrow notes an area of atherosclerosis and thrombus within the aorta. Two distinct lumens (L1 and L2) can be seen at this level. Panel C was recorded in a patient with an aortic dissection extending from the aortic valve to the bifurcation of the aorta. The large white arrows denote the outer dimension of the aorta. There is an echo-free lumen, or true lumen (TL), and a false lumen (FL) with early thrombus formation. Note the vague echo densities within the false lumen. Panel D was recorded in a patient with a large descending thoracic aortic aneurysm and intramural hematoma. The large arrowheads (black and white) denote the outer dimensions of the aorta. The dilated aortic lumen (Ao) is also noted. The black arrow denotes an area of marked atherosclerosis within the aorta, and the double-headed white arrow denotes an area of intramural hemorrhage, characterized by a lower echo density than the atherosclerotic components. Note also the low-density echoes, which represent stagnant blood flow within the aorta.**

tions can be divided into the sequelae of operations involving the ascending aorta and those of operations involving the arch or descending aorta. Because some period of circulatory arrest is often required to approach the ascending aorta or arch, the most severe complication of surgery in this region is cerebral anoxia, with postoperative neurologic dysfunction. Currently, most aortic centers of excellence perform this operation under conditions of deep hypothermia and circulatory arrest, along with retrograde cerebral perfusion by way of the jugular veins. This technique has dramatically diminished the incidence of severe neurologic injury after aortic surgery.

For surgery on the descending thoracic aorta, the most serious complication is interruption of the blood supply to the spinal cord, with resultant paraplegia. Procedures to reduce this complication include the use of shunts and the careful isolation of ostia of the spinal arteries with reimplantation. This complication remains the one most feared in descending-aorta surgery. Additional risks are acute renal failure, mesenteric ischemia, distal atheroembolic events, and pulmonary complications.

Medical Therapy

All patients with aortic dissection receive aggressive medical therapy. This treatment is first directed at controlling the blood pressure. For hypertensive patients, administration of intravenous beta blockers followed by oral beta blockers, along with the concomitant administration of intravenous or oral vasodilators, is imperative. Patients who are normotensive should maintain a low-normal blood pressure and a low heart rate. The likelihood for propagation of dissection is believed to be in part related to acceleration of flow in the aorta—that is, the force of the aortic jet per unit time (i.e., dp/dt). Accordingly, beta blockers have been the most important therapy for the medical treatment of aortic dissection. Such therapy should maintain heart rates at or below 60 beats/min and keep blood pressure as low as possible while allowing perfusion of the brain, kidneys, and other vital organs. Also important are careful measurements of urine output and filling pressures of the heart.

Long-term management of aortic dissection requires aggressive medical therapy and careful surveillance. Patients who retain patency in the false channel of the aorta after either medical

treatment or surgical repair have a significant risk of aneurysm formation and rupture of the false channel, especially in the first 6 months after initial therapy.[52] Expansion, rupture, or both are more common in patients who are older and have poorly controlled hypertension and chronic obstructive pulmonary disease.[53] Aggressive treatment of blood pressure and heart rate and careful monitoring of the patient's status with physical examination and noninvasive imaging are essential. At many centers, either CT scanning or MRI is performed on a regular basis after initial treatment of the dissection. For instance, the patient might be seen at 2 to 4 weeks after admission for adjustment of dosages of antihypertensive medications and beta blockers. At our center, spiral CT scanning is repeated 3 to 6 months after surgery to screen for the development of aneurysm in the false channel or at the margins of a surgical repair. After this, patients undergo aortic imaging annually, and scrupulous attention is directed to antihypertensive therapy.

Atypical Aortic Dissection

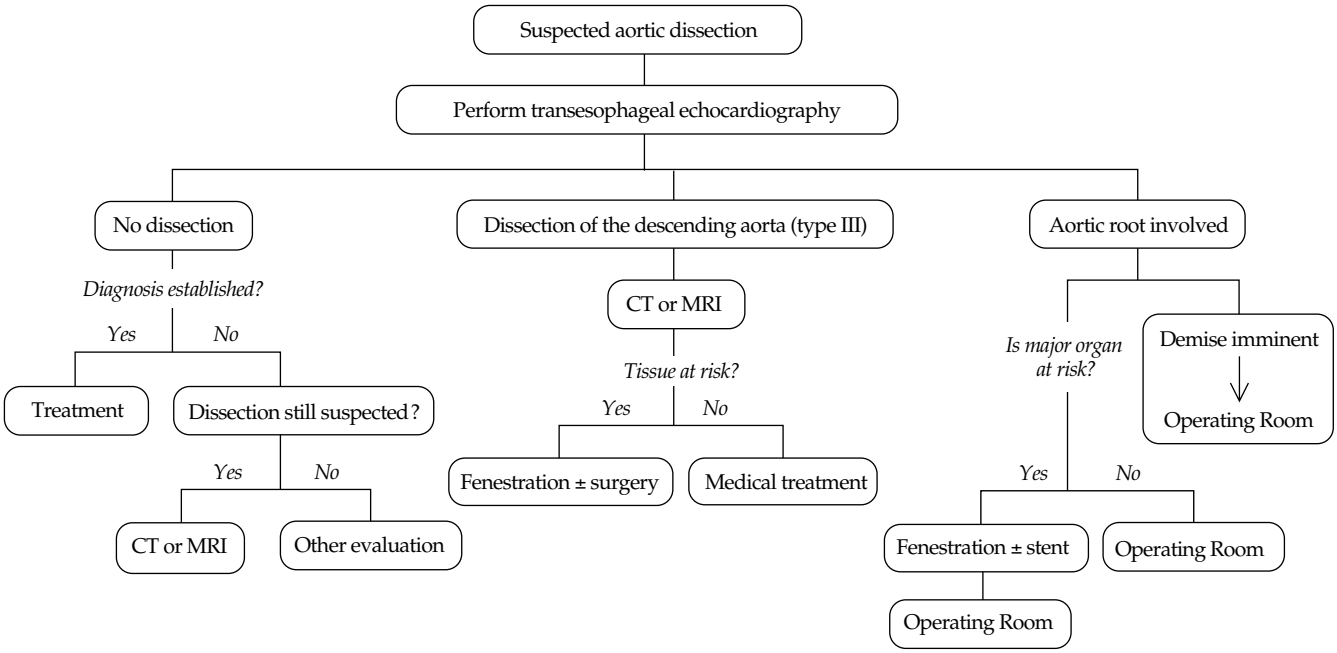
AORTIC DISSECTION WITHOUT INTIMAL TEAR

About 10% to 15% of patients presenting with symptoms suggestive of aortic dissection actually have aortic dissection without an intimal tear (intramural hematoma).[54] This hemorrhage into the medial layer of the aorta may produce a localized or discrete hematoma or may extend for a various distance by dissecting along the plane of the aortic media. Clinically, this hemorrhage mirrors aortic dissection in terms of both its risk factors and its presentation. Intramural hematoma is generally not identified on aortography. It is most easily diagnosed with ultrafast CT scanning. With noncontrast imaging, the hematoma appears as a crescent-shaped high-attenuation area along the aortic wall; moreover, this region cannot be enhanced with contrast imaging. MRI reveals the same crescent-shaped high-intensity area [*see Figure 4*], whereas on TEE, intramural hematoma may appear as a circular or crescentic thickening.[55]

Studies of the natural history of the intramural hematoma suggest that the outcome is similar to that of classic aortic dissection,[56,57] although some recent studies have suggested a more benign prognosis.[58] By 30 days, the rate of aortic expansion or death in patients with medically treated ascending aortic intramural hematoma approaches 50%. Patients with aneurysmal aortas—that is, those with aortas measuring more than 5 cm in diameter—are at particular risk.[59] By contrast, the mortality for intramural hematoma in the descending aorta appears to be between 10% and 15%, a rate similar to that for type B aortic dissection.

PENETRATING ATHEROSCLEROTIC ULCER

A second form of aortic disease that may have an acute pre-



Indications for Surgical Intervention in Aortic Dissection

Ascending Aorta Only	*Ascending and Descending Aorta*	*Descending Aorta Only*
Emergent repair in appropriate candidates	Emergent repair if any of following are present: Aortic insufficiency with CHF Hypotension Pericardial effusion with compromise If ischemia in kidney, CNS, bowel, or major limb, attempt stabilization with stent or fenestration before surgery	Emergent repair if any of following are present: Rupture Impending rupture Uncontrolled pain Major organ at risk (stabilization with stent or fenestration not feasible)

Figure 5 **Decision algorithm for evaluation and treatment of a suspected aortic dissection. Type III dissection originates in the descending aorta and extends distally down the aorta or, in rare instances, retrograde into the aortic arch and ascending aorta. (CHF—congestive heart failure)**

sentation is a penetrating atherosclerotic ulcer.[60] Penetrating ulcers result from erosion of the intima of the aorta, usually because of extensive atherosclerosis. Ulcer formation may produce a hematoma in the media that extends several centimeters from its origin up or down the aorta. Occasionally, pseudoaneurysms are created that may extend into the adventitia and, in rare instances, may rupture. This aortic process develops gradually in elderly patients with extensive atherosclerosis and often is heralded by chest pain or back pain and hypertension. Because it usually presents as a localized process, it is seldom associated with other symptoms of aortic dissection, such as pulse deficit, aortic valve regurgitation, or neurologic defects. Symptomatic penetrating atherosclerotic ulcer rarely requires surgery. Asymptomatic patients who experience progressive enlargement or recurrent atheroemboli may require surgical therapy. For most patients, however, medical therapy suffices and entails aggressive treatment of atherosclerotic risk factors, including cessation of smoking, control of hypertension, lipid-lowering therapy, and careful surveillance. The role of antiplatelet or anticoagulant therapy for this condition is not clear.

Aortic Atheromatous Emboli

Atherosclerosis of the aorta may be so extensive that it leads to overlying thrombosis and subsequent dislodgment of thrombi, cholesterol particles, or fibrinous material into the cerebrovascular or peripheral circulation. Risk factors are hypertension, diabetes, hyperlipidemia, advanced age, and other vascular diseases. Atheromatous disease is most common in the distal aorta but may also occur in the ascending aorta and arch. Evidence of ulceration of atherosclerotic plaques is an independent risk factor for stroke, as is the identification of a mobile, large, protruding aortic atheroma detected with TEE.[60] Plaques more than 4mm in dimension (whereas diameter is used to define the size of the aorta, dimension refers to the maximum size of the atheroma that protrudes into the lumen of the aorta) in the ascending aorta are particularly associated with an increased risk of ischemic stroke.[61] Atheroemboli or cholesterol-particle emboli may also involve the peripheral extremities, leading to ischemic lesions on the feet or toes (so-called blue-toe syndrome). These emboli may present as abdominal pain as a result of ischemic bowel. Acute nonoliguric renal failure is another occasional manifestation, as is gastrointestinal bleeding or pancreatitis. Cutaneous involvement may produce a characteristic skin lesion called livedo reticularis.

Cholesterol embolism syndrome is particularly common after manipulation of the aorta. It is most common in patients undergoing cardiac catheterization or other angiographic procedures in which catheters or wires are manipulated within the aorta. Because the occurrence of atheroemboli may be delayed after aortic manipulation, the relation between the two may not be apparent when the patient is first examined. If cutaneous manifestations are present, a biopsy of the lesions will often identify needle-shaped clefts in the arteriolar lumen characteristic of cholesterol embolization.

Treatment of cholesterol embolism syndrome begins with avoidance of further aortic manipulation (e.g., cardiac catheterization), if this has been a precipitant. Aggressive treatment of hypercholesterolemia is warranted. A search for an aortic aneurysm or protruding mobile atheromas is appropriate in patients for whom the syndrome develops without a concomitant iatrogenic source. Occasionally, recurrent emboli warrant the resection of an aneurysm or of a severely diseased segment of atheromatous aorta.[62] The role of anticoagulant and antiplatelet drugs in this syndrome is uncertain.

Takayasu Arteritis and Giant Cell Arteritis

TAKAYASU ARTERITIS

Takayasu arteritis is a rare inflammatory condition that affects the aorta and its major branches. Other names include aortic arch syndrome, pulseless disease, and young female arteritis.[63] Although Takayasu arteritis is seen throughout the world, most cases occur in Asia and Africa. A specific etiologic agent has yet to be identified, but current evidence favors an autoimmune mechanism. Some studies suggest it may be linked to rheumatic fever, streptococcal infections, certain HLA subtypes, rheumatoid arthritis, and other collagen vascular diseases. Takayasu arteritis is more prevalent in women than in men. By definition, most patients are young, with an average age of 29 years. Takayasu arteritis has been divided into three types.[64] Type I involves the aortic arch and its branches, type II involves the distal aorta and spares the arch, and type III may affect both the ascending aorta and the descending aorta. A suggested fourth category involves the pulmonary arteries.

Pathophysiology

Takayasu arteritis generally involves a granulomatous arteritis of the aorta and its branches, with subsequent involvement of the media and adventitia. Later, the disease may progress to a sclerotic stage in which the intima is hyperplastic, the media degenerates, and the adventitia develops fibrosis. This late fibrotic process may encroach on the lumen of the aorta or its branches. Common areas of involvement are the main aorta and branch points of its major branch vessels. The pulmonary artery may also be involved. The coronary arteries are affected in fewer than 10% of patients. In some patients, involvement of the ascending aorta may lead to aortic valve regurgitation.

Clinical Presentation

The initial symptoms are often typical of an acute or systemic inflammatory process, including fever, loss of appetite, weight loss, night sweats, and arthralgias. Involved vessels may have accompanying localized tenderness over them. By the time the diagnosis is established, most patients have reached a sclerotic phase, in which vascular insufficiency is causing the predominant symptoms. It may involve the upper or lower extremities. Hypertension occurs in more than half of patients. Congestive heart failure occurs in 25% of patients because of hypertension, aortic valve insufficiency, or involvement of the coronary arteries.

Diagnostic Evaluation

Laboratory findings in patients with Takayasu arteritis generally include an elevated erythrocyte sedimentation rate, mild anemia, and a slightly increased white blood cell count. The chest x-ray may demonstrate a rim of calcification around the involved vessels. Aortography often shows an irregular intimal surface with stenoses of the aorta or its branch arteries. Poststenotic dilatation or frank arterial aneurysms may be visible. Similar diagnostic features can also be detected by TEE and MRI.[65] Among the established criteria for the clinical diagnosis of Takayasu arteritis is that patients must be no older than 40 years.

Treatment

The management of Takayasu arteritis begins with high-dose glucocorticoid therapy, which usually leads to abatement of constitutional symptoms and the laboratory signs of inflammation. Serial sedimentation rates are useful for monitoring the benefits of treatment. For patients who fail to respond to steroid therapy, cyclophosphamide at a dosage of 2 mg/kg/day has been used. Alternatively, low-dose methotrexate may enhance the efficacy of steroids or allow steroid tapering. Surgery may be necessary to treat unremitting peripheral ischemia or aortic valve disease or to treat renal artery stenosis that causes severe hypertension. For patients with involvement of the coronary ostia, bypass surgery may be indicated as well.[66] Percutaneously placed arterial stents have successfully treated segmental disease in a variety of vessels in patients with this syndrome.

GIANT CELL ARTERITIS

Giant cell arteritis is another form of aortoarteritis. In contrast to Takayasu arteritis, this illness is more commonly seen in Europe and the United States and in patients older than 50 years (the mean age at onset of disease is 67 years).

Pathophysiology

This form of arteritis often affects the branches of the proximal aorta, particularly the branches supplying the head and neck, the extracranial structures (including the temporal arteries), and the upper extremities. Aortic involvement often coexists with temporal arteritis and polymyalgia rheumatica. Unlike Takayasu arteritis, giant cell arteritis seldom has a sclerotic phase progressing to occlusion of vessels. However, giant cell arteritis may lead to aneurysm formation, aortic regurgitation, or aortic dissection.[67,68]

Clinical Presentation

The classic presentation of giant cell arteritis consists of headache, tenderness over involved arteries in the scalp or the temporal region, jaw claudication, difficulty combing one's hair, and constitutional symptoms. Fever is common, and the blood vessels involved are thick and tender. Pulses may be diminished, and bruits may be present. Occasionally, signs of aortic valve regurgitation are present.

A serious complication of this syndrome is blindness, which results when arteritis affects the ophthalmic artery. The progression to total blindness may be rapid. Visual symptoms of some type occur in as many as 50% of patients. An initial high dose followed by prolonged therapy with corticosteroids remains the treatment of choice.[69] In rare instances, giant cell arteritis may lead to reduced upper extremity pulses and blood pressure along with arm or leg claudication. It also may cause coronary ischemia or abdominal angina in rare cases. Unlike Takayasu arteritis, giant cell arteritis virtually never affects the kidneys. Aortic aneurysms occur in 15% of patients with giant cell arteritis, most commonly involving the ascending aorta. Such aneurysms may develop late in the disease, leading to rupture, aortic dissection, or severe aortic valve regurgitation.

Diagnostic Evaluation

An above normal erythrocyte sedimentation rate is characteristic of this disease, and the diagnosis is confirmed by biopsy of an involved artery, usually the temporal artery. Clinicians need to be aware, however, that temporal artery biopsy may be negative in as many as 15% of patients with confirmed disease; therefore, a second biopsy may be necessary in patients with a high likelihood of temporal arteritis.

Treatment

Standard therapy for giant cell arteritis is high-dose glucocorticoid therapy (e.g., prednisone, 40 to 60 mg/day). Methotrexate may be used to reduce the need for steroids or to treat patients who respond inadequately to steroids. Cyclophosphamide may also be useful for reducing the need for glucocorticoids. Surgery is typically reserved for patients who experience progressive ischemic symptoms or aortic aneurysms.

Traumatic Disease of the Aorta

Finally, a relatively common form of aortic pathology is partial or complete transection as a result of major blunt thoracic trauma, most commonly as a result of a high-speed motor vehicle accident. Most patients with complete aortic transection do not survive long enough for hospital evaluation. If rapidly diagnosed, patients with partial transection may survive long enough to undergo surgical correction. Evidence of aortic trauma is often obscured by other major organ trauma. Patients with aortic transection are typically in shock and may have diminished lower extremity pulses. The transection is usually located at the distal arch, immediately after the origin of the left subclavian artery.

Rapid diagnosis is the key to the survival of patients with aortic transection. The routine chest x-ray typically reveals a widened mediastinum, often with pleural effusions. The gold standard for diagnosis of this disorder remains aortography, but TEE,[60,70] CT,[71] and MRI have also been used successfully.[72] Successful treatment requires vigorous fluid and blood resuscitation and surgical repair of the aortic transsection.

Kim A. Eagle, M.D., has received grant and research support from Pfizer, Inc., and Aventis and is a consultant for COR Therapeutics, Inc. William F. Armstrong, M.D., has served as an advisor or consultant to AstraZeneca Pharmaceuticals LP and St. Jude Medical, Inc.

References

1. Rodin MB, Daviglus ML, Wong GC, et al: Middle age cardiovascular risk factors and abdominal aortic aneurysm in older age. Hypertension 42:61, 2003

2. Crawford ES, Cohen ES: Aortic aneurysm: a multifocal disease. Arch Surg 117:1393, 1982

3. Ashton HA, Buxton MJ, Day NE, et al: The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. Lancet 360:1531, 2002

4. Frame PS, Fryback DG, Patterson C: Screening for abdominal aortic aneurysm in men ages 60 to 80 years: a cost-effectiveness analysis. Ann Intern Med 119:411, 1993

5. Ernst CB: Abdominal aortic aneurysm. N Engl J Med 328:1167, 1993

6. Lederle FA, Wilson SE, Johnson GR, et al: Variability in measurements of abdominal aortic aneurysms. J Vasc Surg 21:945, 1995

7. Petersen MJ, Cambria RP, Kaufman JA, et al: Magnetic resonance angiography in the preoperative evaluation of abdominal aortic aneurysms. J Vasc Surg 21:891, 1995

8. Gadowski GR, Pilcher DB, Ricci MA: Abdominal aortic aneurysm expansion rate: effect of size and beta-adrenergic blockade. J Vasc Surg 19:727, 1994

9. Hollier LD, Taylor LM, Ochsner J: Recommended indications for operative treatment of abdominal aortic aneurysms: report of a subcommittee of the Joint Council of the Society for Vascular Surgery and of the North American Chapter of the International Society for Cardiovascular Surgery. J Vasc Surg 15:1046, 1992

10. Lederle FA, Wilson SE, Johnson GR, et al: Immediate repair compared with surveillance of small abdominal aortic aneurysms. N Engl J Med 346:1437, 2002

11. Lederle FA, Johnson GR, Wilson SE, et al: Rupture rate of large abdominal aortic aneurysms in patients refusing or unfit for elective repair. JAMA 287:2968, 2002

12. Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. The UK Small Aneurysm Trial Participants. Lancet 352:1649, 1998

13. Ballard DJ, Fowkes FG, Powell JT: Surgery for small asymptomatic abdominal aortic aneurysms. Cochrane Database Syst Rev (2):CD001835, 2000

14. Wolf YG, Fogarty TJ, Olcott C, et al: Endovascular repair of abdominal aortic aneurysms: eligibility rate and impact on the rate of open repair. J Vasc Surg 32:519, 2000

15. Faries PL, Brener BJ, Connelly TL, et al: A multicenter experience with the Talent endovascular graft for the treatment of abdominal aortic aneurysms. J Vasc Surg 35:1123, 2002

16. Eagle KA, Brundage BH, Chaitman BR, et al: Guidelines for perioperative cardiovascular evaluation for noncardiac surgery. Report of the American College of Cardiology/ American Heart Association Task Force on Practice Guidelines. Committee on Perioperative Cardiovascular Evaluation for Noncardiac Surgery. Circulation 93:1278, 1996

17. Poldermans D, Boersma E, Bax JJ, et al: The effect of bisoprolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. Dutch Echocardiographic Cardiac Risk Evaluation Applying Stress Echocardiography Study Group. N Engl J Med 341:1789, 1999

18. Best PJ, Tajik AJ, Gibbons RJ, et al: The safety of treadmill exercise stress testing in patients with abdominal aortic aneurysms. Ann Intern Med 129:628, 1998

19. Rojo-Leyva F, Ratliff NB, Cosgrove DM 3rd, et al: Study of 52 patients with idiopathic aortitis from a cohort of 1,204 surgical cases. Arthritis Rheum 43:901, 2000

20. Shores J, Berger KR, Murphy EA, et al: Progression of aortic dilatation and the benefit of long-term beta-adrenergic blockade in Marfan's syndrome. N Engl J Med 330:1335, 1994

21. Dapunt OE, Galla JD, Sadeghi AM, et al: The natural history of thoracic aortic aneurysms. J Thorac Cardiovasc Surg 107:1323, 1994

22. Elefteriades JA: Natural history of thoracic aortic aneurysms: indications for surgery, and surgical versus nonsurgical risks. Ann Thorac Surg 74:S1877, 2002

23. Coady MA, Rizzo JA, Hammond GL, et al: What is the appropriate size criterion for resection of thoracic aortic aneurysm? J Thorac Cardiovasc Surg 113:476, 1997

24. Mitchell RS, Miller DC, Dake MD, et al: Thoracic aortic aneurysm repair with an endovascular stent graft: the "first generation." Ann Thorac Surg 67:1971, 1999

25. Okita Y, Takamoto S, Ando M, et al: Mortality and cerebral outcome in patients who underwent aortic arch operations using deep hypothermic circulatory arrest with retrograde cerebral perfusion: no relation of early death, stroke, and delirium to the duration of circulatory arrest. J Thorac Cardiovasc Surg 115:129, 1998

26. Coselli JS, Buket S, Djukanovic B: Aortic arch operation: current treatment and results. Ann Thorac Surg 59:19, 1995

27. Coselli JA, LeMaire SA, deFigueiredo LP, et al: Paraplegia after thoracoabdominal aortic aneurysm repair: is dissection a risk factor? Ann Thorac Surg 63:28, 1997

28. Cambria RP, Davison JK, Zannetti S, et al: Clinical experience with epidural cooling for spinal cord protection during thoracic and thoracoabdominal aneurysm repair. J Vasc Surg 25:234, 1997

29. Elkayam U, Ostzega E, Shotan A, et al: Cardiovascular problems in pregnant women with the Marfan syndrome. Ann Intern Med 123:177, 1995

30. Hagan PG, Nienaber CA, Isselbacher EM, et al: The international registry of aortic dissection (IRAD): new insights into an old disease. JAMA 283:897, 2000

31. Armstrong WF, Bach DS, Carey LM, et al: Clinical and echocardiographic findings in patients with suspected acute aortic dissection. Am Heart J 136:1051, 1998

32. Mehta RH, Suzuki T, Hagan PG, et al: Predicting death in patients with acute type A aortic dissection. Circulation 105:200, 2002

33. Mehta RH, O'Gara PT, Bossone E, et al: Acute type A aortic dissection in the elderly: Clinical characteristics, management, and outcomes in the current era. J Am Coll Cardiol 40:685, 2002

34. Suzuki T, Katoh H, Tsuchio Y, et al: Diagnostic implications of elevated levels of smooth-muscle myosin heavy-chain protein in acute aortic dissection. The smooth muscle myosin heavy chain study. Ann Intern Med 133:537, 2000

35. Suzuki T, Katoh H, Kurabayashi M, et al: Biochemical diagnosis of aortic dissection by raised concentrations of creatine kinase BB-isozyme. Lancet 350:784, 1997

36. Nienaber CA, vonKodolitsch Y, Nicolas V, et al: The diagnosis of thoracic aortic dissection by noninvasive imaging procedures. N Engl J Med 328:1, 1993

37. Zeman RK, Berman PM, Silverman PM, et al: Diagnosis of aortic dissection: value of helical CT with multiplanar reformation and three-dimensional rendering. AJR Am J Roentgenol 164:1375, 1995

38. Evangelista A, Garcia-del-Castillo H, Gonzales-Alujas T, et al: Diagnosis of ascending aortic dissection by transesophageal echocardiography: utility of M-mode in recognizing artifacts. J Am Coll Cardiol 27:102, 1996

39. Banning AP, Ruttley MST, Musumeci F, et al: Acute dissection of the thoracic aorta: transesophageal echocardiography is the investigation of choice. BMJ 310:72, 1995

40. Armstrong WF, Bach DS, Carey L, et al: Spectrum of acute dissection of the ascending aorta: a transesophageal echocardiographic study. J Am Soc Echocardiogr 9:646, 1996

41. Movsowitz H, Levine RA, Hilgenberg AD, et al: Transesophageal echocardiographic description of the mechanisms of aortic regurgitation in acute type A aortic dissection: implications for aortic valve repair. J Am Coll Cardiol 36:884, 2000

42. Keane MG, Wiegers SE, Yang E, et al: Structural determinants of aortic regurgitation in type A dissection and the role of ventricular resuspension as determined by intraoperative transesophageal echocardiography. Am J Cardiol 85:604, 2000

43. Bansal RC, Chandrasekaran K, Ayala J, et al: Frequency and explanation of false negative diagnosis of aortic dissection by aortography and transesophageal echocardiography. J Am Coll Cardiol 25:1393, 1995

44. Deeb GM, William DM, Bolling SF, et al: Surgical delay for acute type A dissection with malperfusion. Ann Thorac Surg 64:1669, 1997

45. Bavaria JE, Brinster DR, Gorman RC, et al: Advances in the treatment of acute type A dissection: an integrated approach. Ann Thorac Surg 74:S1848, 2002

46. Fann JI, Smith JA, Miller C, et al: Surgical management of aortic dissection during a 30-year period. Circulation 92:II113, 1995

47. Nienaber CA, Fattori R, Lund G, et al: Nonsurgical reconstruction of thoracic aortic dissection by stent-graft placement. N Engl J Med 340:1539, 1999

48. Dake MD, Kato N, Mitchell RS, et al: Endovascular stent-graft placement for the treatment of acute aortic dissection. N Engl J Med 340:1546, 1999

49. Kato M, Matsuda T, Kaneko M, et al: Outcomes of stent-graft treatment of false lumen in aortic dissection. Circulation 98:II305, 1998

50. Palma JH, Almeida DR, Carvalho AC, et al: Surgical treatment of acute type B aortic dissection using an endoprosthesis (elephant trunk). Ann Thorac Surg 63:1081, 1997

51. Umana JP, Miller DC, Mitchell RS: What is the best treatment for patients with acute type B aortic dissections—medical, surgical, or endovascular stent-grafting? Ann Thorac Surg 74:S1840, 2002

52. Yamashita C, Okada M, Ataka K, et al: Cerebral complications and distal false lumen in the repair of aortic dissection with retrograde cerebral perfusion. J Cardiovasc Surg 38:581, 1997

53. Juvonen T, Ergin MA, Galla JD, et al: Risk factors for rupture of chronic type B dissections. J Thorac Cardiovasc Surg 117:776, 1999

54. Nienaber CA, von Kodolitsch Y, Petersen B, et al: Intramural hemorrhage of the thoracic aorta: diagnostic and therapeutic implications. Circulation 92:1465, 1995

55. Vilacosta I, San Roman JA, Ferreiros J, et al: Natural history and serial morphology of aortic intramural hematoma: a novel variant of aortic dissection. Am Heart J 134:495, 1997

56. Maraj R, Rerkpattanapipat P, Jacobs LE, et al: Meta-analysis of 143 reported cases of aortic intramural hematoma. Am J Cardiol 86:664, 2000

57. von Kodolitsch Y, Csosz SK, Koschyk DH, et al: Intramural hematoma of the aorta: predictors of progression to dissection and rupture. Circulation 107:1158, 2003

58. Song JK, Kim HS, Song JM, et al: Outcomes of medically treated patients with aortic intramural hematoma. Am J Med 113:181, 2002

59. Kaji S, Nishigama K, Akasada T, et al: Prediction of progression or regression of type A aortic intramural hematoma by computed tomography. Circulation 100:II281, 1999

60. Vilacosta I, San Román JA, Aragoncillo P, et al: Penetrating atherosclerotic aortic ulcer: documentation by transesophageal echocardiography. J Am Coll Cardiol 32:83, 1998

61. Willens HJ, Kessler KM: Transesophageal echocardiography in the diagnosis of diseases of the thoracic aorta: part II—atherosclerotic and traumatic diseases of the aorta. Chest 117:233, 2000

62. Bojar RM, Payne DD, Murphy RE, et al: Surgical treatment of systemic atheroembolism from the thoracic aorta. Ann Thorac Surg 61:1389, 1996

63. Numano F, Kobayashi Y: Takayasu arteritis—beyond pulselessness. Intern Med 38: 226, 1999

64. Ishikawa K: Diagnostic approach and proposed criteria for the clinical diagnosis of Takayasu's arteriopathy. J Am Coll Cardiol 12:964, 1988

65. Matsunaga N, Hayaski K, Sakamoto I, et al: Takayasu arteritis: MR manifestations and diagnosis of acute and chronic phase. J Magn Reson Imaging 8:406, 1998

66. Endo M, Tomizawa Y, Nishida H, et al: Angiographic findings and surgical treatments of coronary artery involvement in Takayasu arteritis. J Thorac Cardiovasc Surg 125:570, 2003

67. Evans JM, O'Fallon WM, Hunder GG: Increased incidence of aortic aneurysm and dissection in giant cell (temporal) arteritis: a population-based study. Ann Intern Med 122:502, 1995

68. Gravanis MB: Giant cell arteritis and Takayasu aortitis: morphologic, pathogenetic and etiologic factors. Int J Cardiol 75:S21, 2000

69. Hayreh SS, Zimmerman B: Management of giant cell arteritis: our 27-year clinical study: new light on old controversies. Ophthalmologica 217:239, 2003

70. Smith MD, Cassidy JM, Souther S, et al: Transesophageal echocardiography in the diagnosis of traumatic rupture of the aorta. N Engl J Med 332:356, 1995

71. Demetriades D, Gomez H, Velmahos CG, et al: Routine helical computed tomographic evaluation of the mediastinum in high-risk blunt trauma patients. Arch Surg 133:1084, 1998

72. Patel NH, Stephens KE Jr, Mirvis SE, et al: Imaging of acute thoracic aortic injury due to blunt trauma: a review. Radiology 209:335, 1998

XIII DISEASES OF THE PERICARDIUM, CARDIAC TUMORS, AND CARDIAC TRAUMA

E. William Hancock, m.d.

Diseases of the Pericardium

The pericardium provides a protective sac around the heart. The sac contains a thin layer of fluid that permits the heart to move with minimal friction during the cardiac cycle. Neither the sac nor the fluid appears to be necessary for normal function. When one or more of the cardiac chambers dilate acutely, the pericardium restrains the heart. In chronic dilatation of the heart, however, the pericardium stretches and therefore does not exert a restraining effect, except during exercise or other acute stresses.

Pericardial disease results from diverse causes, many of which lead to responses to injury that are pathologically and clinically similar. There are three clinicopathologic responses to injury: acute pericarditis, pericardial effusion, and constrictive pericarditis.

acute pericarditis

Viral infection is usually assumed to be the cause of acute pericarditis that occurs as an apparently primary illness. Because most cases follow a brief and uncomplicated natural course, the syndrome is often termed acute benign pericarditis. Cases resulting from other conditions or treatments, such as rheumatic disease or radiotherapy, often exhibit clinical features similar to those of acute benign pericarditis.

Diagnosis

Clinical manifestations The clinical diagnosis of acute pericarditis rests primarily on the findings of chest pain, pericardial friction rub, and electrocardiographic changes. The chest pain of acute pericarditis typically develops suddenly and is severe and constant over the anterior chest. In acute pericarditis, the pain worsens with inspiration—a response that helps distinguish acute pericarditis from myocardial infarction. Low-grade fever and sinus tachycardia also are usually present.

A pericardial friction rub can be detected in most patients when symptoms are acute. Pericardial friction rubs are typically triphasic: systolic and early diastolic components are followed in later diastole by a third component associated with atrial contraction.

ECG findings Electrocardiographic changes are common in most forms of acute pericarditis, particularly those of an infectious etiology in which the associated inflammation in the superficial layer of myocardium is prominent. The characteristic change is an elevation in the ST segment in diffuse leads. The diffuse distribution and the absence of reciprocal ST segment depression distinguish the characteristic pattern of acute pericarditis from acute myocardial infarction. However, the normal variant pattern of ST segment elevation often complicates the differential diagnosis [*see Figure 1*]. Depression of the PR segment, which reflects superficial injury of the atrial myocardium, is as frequent and specific as ST segment elevation and is often the earliest electrocardiographic manifestation.[1]

Treatment

Analgesic agents, such as codeine (15 to 30 mg taken orally every 4 to 6 hours) or hydrocodone (5 to 10 mg taken orally every 4 to 6 hours), are usually effective in providing symptomatic relief. Salicylates given at an initial dosage of 4 to 6 g a day or a nonsteroidal anti-inflammatory drug (NSAID) such as ibuprofen given at an initial dosage of 800 mg three times daily is often effective in reducing pericardial inflammation.[2] Corticosteroids such as prednisone given at an initial dosage of 40 to 60 mg/day often greatly relieve symptoms; however, steroid therapy should be reserved for severe cases that are unresponsive to other therapy, because symptoms may recur after steroid withdrawal. The corticosteroid dose should be reduced as soon as a clinical response is observed and should be tapered to zero over a period of 2 to 4 weeks.

Other Forms of Acute Pericarditis

Relapsing pericarditis Acute pericarditis of any etiology may follow a recurrent or chronic relapsing course. In many instances, subjective manifestations (e.g., weakness, fatigue, or headache) are present in addition to the chest discomfort. Analgesic agents provide symptomatic relief, and a very slow tapering of the dose of a corticosteroid usually resolves the relapsing course eventually. Treatment with 1 mg of colchicine daily, methylprednisolone in 1 g pulses daily for 3 days, or an immunosuppressant such as prednisone (60 to 100 mg daily) or azathioprine (50 to 100 mg daily) has proved successful in patients with relapsing pericarditis, particularly in patients whose symptoms are mainly related to withdrawal of prednisone.[2,3]

Progression to constriction In a few instances, acute pericarditis progresses to subacute or chronic constrictive pericarditis. In such cases, the pericarditis may be idiopathic or have a bacterial, viral, rheumatoid, radiation-induced, or dialysis-related origin. These patients usually have subacute rather than acute pericarditis initially; pericardial effusion is present at onset, usually with some degree of cardiac tamponade. Acute benign pericarditis unaccompanied by tamponade or substantial pericardial effusion in the acute phase rarely progresses to constrictive pericarditis.

pericardial effusion and cardiac tamponade

Fluid may accumulate in the pericardial cavity in virtually any form of pericardial disease. The fluid may be a transudate or an exudate and is often serosanguineous in neoplastic, idiopathic, dialysis-related, radiation-induced, and tuberculous cases. The fluid is serosanguineous or frankly bloody in cases of coagulopathy, trauma, rupture of acute myocardial infarction, and aortic dissection. Chylopericardium and pneumopericardium also can occur, although rarely.[4] Cardiac tamponade or compression of the heart by effusion is the most important complication of pericardial effusion.

Pathophysiology

The physiologic effect of the accumulation of pericardial fluid depends on whether the fluid is under increased pressure.[5] If ef-

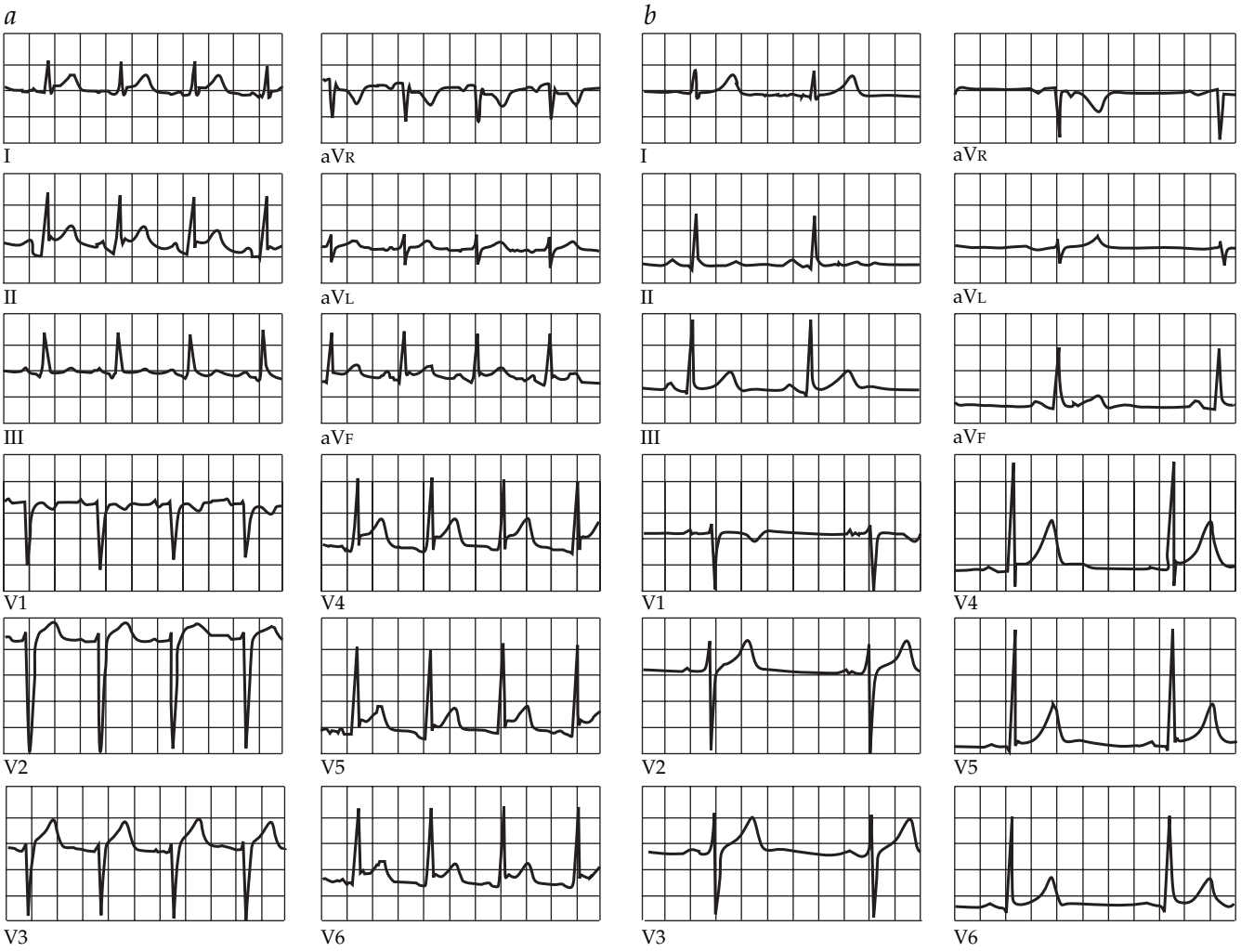


Figure 1 **Electrocardiograms contrast the pattern of ST segment elevation characteristic of acute pericarditis (*a*) with the normal variant (early repolarization) pattern of ST segment elevation (*b*). The normal variant pattern is associated with a normal or slow heart rate and has relatively tall R waves and T waves in V4, V5, and V6. The ST segment elevation is less than 25% of the T wave amplitude. In contrast, the acute pericarditis has PR depression and lower T wave amplitude.**

fusion develops gradually, the pericardium stretches enough to accommodate volumes that may exceed 2,000 ml. However, if the effusion develops acutely, as little as 200 ml of accumulated fluid may raise the intrapericardial pressure and cause cardiac tamponade.

As the pericardial fluid pressure rises, the right atrial and central venous pressures increase correspondingly. Thus, reading the central venous pressure gives an accurate reflection of the intrapericardial pressure.

Cardiac tamponade should be viewed as a spectrum of hemodynamic abnormalities of various severities rather than an all-or-none phenomenon. Depending on the severity of the tamponade, the blood pressure may be lowered or maintained near the normal range; in patients with preexisting hypertension, it may even be increased. The central venous pressure is almost always increased, except in the rare instances of low-pressure cardiac tamponade, which may occur when intravascular volume is depleted.

As a rule, paradoxical pulse—a marked decrease in arterial pressure during inspiration—is present in patients with cardiac tamponade, although it may not be easy to detect on clinical examination [*see Figure 2*]. The arbitrary value of 10 mm Hg is commonly used to indicate the upper limit of the normal decrease in arterial pressure with inspiration.

The inspiratory drop in arterial pressure reflects a selective impairment of diastolic filling of the left ventricle, probably the combined effects of two factors. First, when the filling of the right ventricle is augmented in inspiration, the simultaneous filling of the left ventricle is limited because the entire heart is enclosed in a fixed volume. Second, during inspiration, blood is sequestered in the lungs and pulmonary veins as a result of impaired transmission of changes in intrapleural pressure to the left atrium and to the intrapericardial portions of the pulmonary veins.

Diagnosis

Pericardial effusion Echocardiography is the most accurate and easily applied method for the clinical detection of pericardial effusion [*see Figure 3*]. Echocardiograms detect effusions as small as 20 ml and show characteristic findings with effusions larger than 100 ml. Computed tomography is also a reliable method for detecting both pericardial effusion and pericardial thickening. Magnetic resonance imaging provides information

similar to that provided by CT [*see Figure 4*].[6]

Electrocardiograms usually show low voltage in patients with large pericardial effusions, but this finding is nonspecific. Electrical alternans occurs occasionally when the pericardial effusion is large and permits a beat-to-beat oscillation of the heart from one position to another within the pericardial sac [*see Figure 5*]. Electrical alternans is most common with effusion caused by neoplasm. This type of electrical alternans must be differentiated from other types, such as that occurring in supraventricular tachycardias or alternating intraventricular conduction defects.

Cardiac tamponade The diagnosis of cardiac tamponade is often difficult.[7] The diagnosis is one of the most common important diagnoses made at autopsy but not during life.[8] This diagnosis should be based on a synthesis of various clinical findings, because no single finding is pathognomonic or necessarily present. Echocardiography, although essentially definitive for the demonstration of pericardial effusion, is not as certain a method of assessing tamponade. Several echocardiographic features are helpful, however, particularly the observation of an early diastolic inward motion (so-called collapse) of the right atrial wall or right ventricular wall, indicating similarity of intracavitary and intrapericardial pressures.[9] Another useful sign is an exaggerated respiratory variation in the velocity of flow through the mitral and tricuspid valves or in the left ventricular ejection, as detected in pulsed Doppler recordings [*see Figure 6*]. This phenomenon has the same significance as paradoxical pulse. A third sign suggestive of cardiac tamponade is plethora of the inferior vena cava, which occurs in the more severe cases and is a better guide to the need for pericardiocentesis than the other echocardiographic features.[10]

Loculated pericardial effusions may selectively compress one or more chambers of the heart, producing regional cardiac tamponade. This condition is seen most frequently after cardiac surgery, when bloody fluid accumulates behind the sternum and selectively compresses the right atrium and right ventricle[11]; less often, the left ventricle and left atrium are compressed locally. Similar conditions may occur after closed chest trauma. Fluid accumulations in the mediastinum can compress the heart even when they are not truly within the pericardial space. Transesophageal echocardiography is superior to transthoracic echocardiography in demonstrating such local fluid accumulations, particularly those along the right heart border.

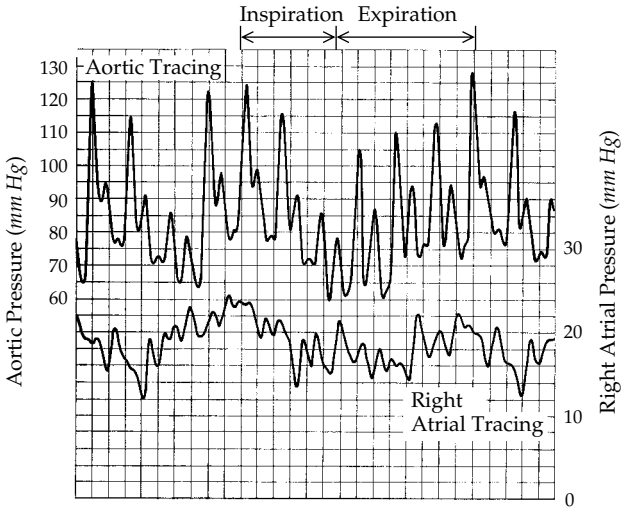


Figure 2 **The aortic pressure and right atrial pressure are recorded during quiet breathing in a patient with cardiac tamponade. The marked fall in arterial pressure that occurs during inspiration is a paradoxical pulse. The decrease in pulse pressure, defined as the difference between systolic and diastolic pressures, that accompanies the fall in systolic pressure indicates that left ventricular stroke volume decreases during inspiration. Central venous pressure, as indicated by the right atrial pressure, also falls during inspiration.**

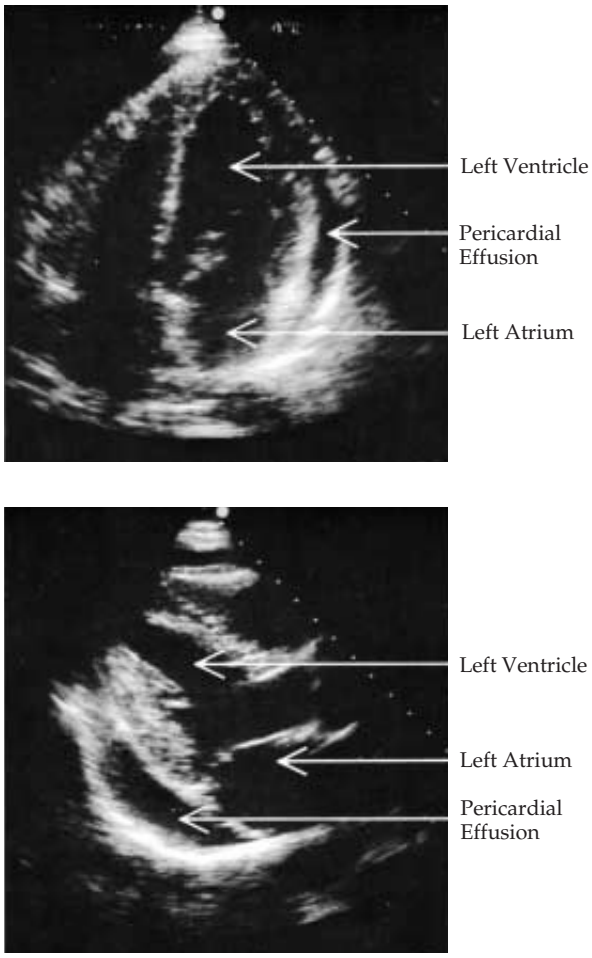


Figure 3 **Pericardial effusion is seen in the two-dimensional echocardiogram as an echo-free space outside the cardiac chambers. Two characteristic sites are lateral to the left ventricle in the apical four-chamber view (top) and posterior to the left ventricle in the parasternal long-axis view (bottom).**

Treatment

Pericardial effusion Occasionally, a syndrome of idiopathic chronic large pericardial effusion is seen, usually without tamponade. Colchicine may be effective in such cases.[12]

In patients with pericardial effusion but no tamponade, pericardiocentesis is rarely performed for the sole purpose of providing diagnostic studies of the fluid, because such specific diagnoses are uncommon in those patients, at least in regions of the world where tuberculous pericarditis has become rare.[13] Pericardial biopsy can be obtained by any of the usual surgical methods. However, pericardiocentesis can be useful in diagnosing infection or neoplastic disease. In addition, pericardioscopy can be

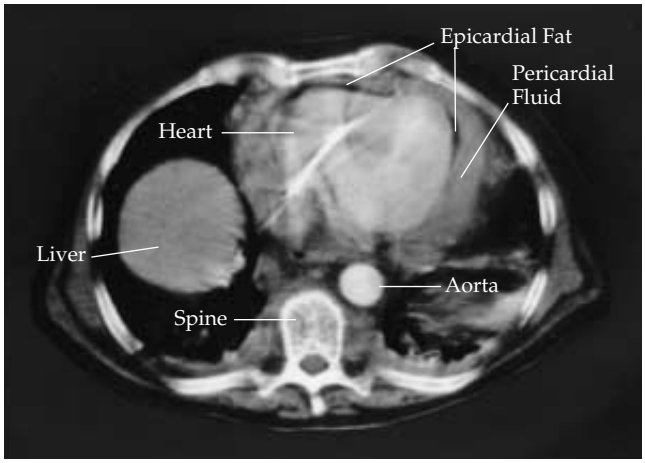


Figure 4 **In a CT scan of the chest of a patient with pericardial effusion, pericardial fluid appears less dense than the heart and is separated from the myocardium by epicardial fat in some areas.**

performed in association with subxiphoid pericardiostomy; biopsy under direct vision may permit the diagnosis of tuberculosis or neoplasm in some instances in which studies of the fluid alone might be inconclusive.[14] Surgery is usually not necessary in chronic idiopathic pericardial effusion in the absence of tamponade and effusive-constrictive disease.[15]

Cardiac tamponade Mild cardiac tamponade may be managed conservatively in some cases, but removal of the fluid is required for definitive treatment and should be carried out in most instances when the central venous pressure is increased. Pericardial fluid may be removed by needle pericardiocentesis or by a surgical technique (subxiphoid pericardiostomy, thoracoscopic pericardiostomy, or thoracotomy).[16,17]

The most acute forms of cardiac tamponade, such as hemopericardium secondary to aortic dissection, penetrating cardiac trauma, or rupture of acute myocardial infarction, require immediate surgery. Tamponade caused by cardiac perforation during invasive intravascular procedures can usually be managed by pericardiocentesis.[18] Pericardiocentesis is effective in most subacute forms of tamponade, such as those associated with idiopathic or viral acute pericarditis, rheumatic diseases, dialysis, or neoplasm.

Thoracoscopy and thoracotomy are usually reserved for patients with recurrent tamponade after an initial pericardiocentesis or subxiphoid pericardiostomy, usually for neoplastic disease. Pericardiostomy by means of a balloon catheter as part of a pericardiocentesis is another alternative for such cases.[19]

SPECIAL ETIOLOGIC FORMS OF ACUTE PERICARDITIS AND PERICARDIAL EFFUSION

Pericarditis Related to Renal Failure and Dialysis

Acute pericarditis with pericardial effusion occurs in patients with end-stage renal disease and in patients who are on chronic dialysis [*see 10:X Chronic Renal Failure and Dialysis*]. In dialysis patients, conservative management with more intensive dialysis and NSAIDs is usually successful. An unexpected decrease in blood pressure during a dialysis session may be the clue to the presence of tamponade. Pericardiocentesis is occasionally necessary for the relief of tamponade, although fluid overload and left ventricular failure are often important factors associated with causing increased central venous pressure. Cardiac catheterization in combination with pericardiocentesis is often useful in assessing the hemodynamic significance of those factors that contribute to an increase of pulmonary and systemic venous pressure in dialysis patients.

Radiation-Induced Pericardial Effusion

Pericardial effusion develops relatively frequently in patients with Hodgkin disease, other lymphomas, or breast carcinoma who survive for long periods after receiving large doses of radiation to the mediastinum. Radiation-induced effusion may evolve into chronic constrictive pericarditis after many years, usually with other forms of myocardial, coronary arterial, valvular, and pulmonary damage.[20]

Neoplastic Pericardial Effusion

Neoplastic pericardial effusion accounts for about one half of the cases of cardiac tamponade in patients who are seen in an internal medicine setting [*see 12:XII Oncologic Emergencies*]. Lung cancer and breast cancer account for the majority of the cases; lymphoma and leukemia account for most of the remainder.[21] In most cases, the primary neoplasm has been previously diagnosed; patients in whom pericardial effusion is the first manifestation of the disease usually have primary cancer of the lung. Cytologic examination of the pericardial fluid is highly accurate in diagnosing common carcinomas but less accurate in diagnosing

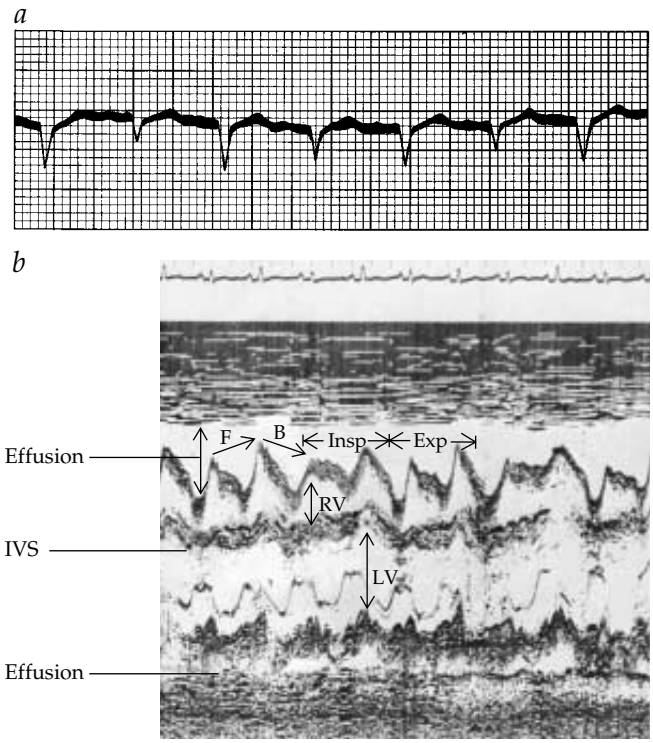


Figure 5 **The electrocardiogram (V2 lead) from a patient with pericardial effusion caused by malignant melanoma reveals a low voltage and electrical alternans (*a*). The echocardiogram (*b*) demonstrates that the heart moves forward (F) and backward (B) within the effusion on alternate beats, thus producing the alternation of the QRS axis characteristic of electrical alternans. The heart also moves with inspiration (Insp) and expiration (Exp), which accounts for a change in anterior wall motion with every two cardiac cycles.**

other neoplasms, especially the lymphomas and leukemias.[22]

Neoplastic pericardial effusion often can be managed conservatively when no symptoms directly related to the pericardial effusion are present. Symptomatic tamponade can be managed palliatively with pericardiocentesis, although recurrent effusion is more likely to form in such cases than in many other types of pericardial effusion. Subxiphoid pericardiostomy is often the preferred procedure, leading to a pericardial reaction that produces adhesion of the parietal and visceral layers of pericardium and thus prevents recurrent effusion. Balloon pericardiostomy is an alternative. Chemotherapy or radiotherapy may be of value, depending on the nature of the primary neoplasm. Intrapericardial instillation of chemotherapeutic agents has often been used with apparent success, but no results from controlled trials are available. Few patients survive longer than a year, and whether pericardiocentesis has a major effect on their longevity is difficult to determine, even when tamponade is relieved.[23]

Purulent Pericarditis

Tamponade is usually present in purulent bacterial pericarditis; after pericardiocentesis, the effusion is highly likely to recur rapidly and progress to constrictive pericarditis. Purulent pericarditis therefore usually requires a surgical drainage procedure; a partial pericardiectomy by a limited left lateral thoracotomy is often the best choice. Surgery may not be required for patients in whom tamponade or constriction does not develop, because antibiotics enter the pericardial cavity in effective concentrations. Active tuberculous pericarditis with effusion is particularly likely to progress to constriction and to require pericardiectomy in addition to antituberculous chemotherapy.[24,25] AIDS is a common cause of large pericardial effusions, the majority of which are not caused by identifiable opportunistic infective agents.[26] However, the incidence of tuberculous and other forms of bacterial pericarditis has increased in the United States as a result of AIDS.

Drug-Induced Pericarditis

Several drugs have been implicated in the etiology of pericardial disease, including procainamide, which leads to a lupuslike syndrome; minoxidil, which has been linked to pericardial effusion; and methysergide, which may lead to constrictive pericarditis.

Pericarditis after Cardiac Surgery

The postcardiotomy syndrome presents primarily as acute pericarditis. Whether it has an infective or autoimmune cause is unclear. A similar condition occurs after blunt or penetrating trauma, hemopericardium from other causes, or epicardial pacemaker implantation. Cardiac tamponade and constrictive pericarditis occur occasionally.

Pericardial Complications of Invasive Procedures

Cardiac tamponade occurs as a complication of various invasive procedures in the cardiac catheterization laboratory and in the intensive care unit. Particularly important, and usually preventable, is the perforation of the heart by central venous catheters that have been allowed to lie in the right atrium rather than in the superior vena cava.[27] The use of newer devices (e.g., stents) and procedures (e.g., rotational atherectomy) has increased the incidence of percutaneous coronary interventions with cardiac tamponade complications.[28] Most of these cases are managed successfully by pericardiocentesis.[18]

CONSTRICTIVE PERICARDITIS

Constrictive pericarditis was formerly widely considered to be primarily a tuberculous lesion and is still so regarded in many areas of the world. Most cases now seen in the United States are idiopathic or are related to previous cardiac surgery or radiotherapy.[29,30] Fewer cases result from purulent pericarditis, rheumatic diseases, dialysis, and various rarer conditions.

In the classic form of chronic constrictive pericarditis, fibrous scarring and adhesion of both pericardial layers obliterate the pericardial cavity. The resulting fibrotic lesion has been likened to a rigid shell around the heart, particularly when there is considerable calcification of the pericardium, a feature seen in longstanding cases. The subacute form of constrictive pericarditis is now more common than the chronic calcific type. In the subacute variant, the constriction is rather fibroelastic and may be produced by fibrous contracture of the visceral pericardial layer (epicardium) alone. The fibroelastic constriction may also exist in combination with persisting loculated or totally free pericardial effusion; this form is termed effusive-constrictive pericarditis, and it can be documented by measuring pericardial and central venous pressures before and after removal of the fluid.

Pathophysiology

The pathophysiology of constrictive pericarditis is similar to that of tamponade in that both conditions impede diastolic filling of the heart and lead to increased venous pressure and ultimately to reduced cardiac output. Differences exist in the diagnostic signs, however. Paradoxical pulse is a regular feature of cardiac tamponade but may be inconspicuous or absent in constrictive pericarditis. The Kussmaul sign (an increase in venous pressure with inspiration) is seen in some patients with constrictive pericarditis but not in patients with pure cardiac tamponade. When tamponade is present, the venous pulse shows a predominant systolic dip, whereas in constrictive pericarditis, the early diastolic dip is the more prominent deflection [*see Figure 6*].

An early diastolic sound (pericardial knock) is often heard in constrictive pericarditis but not in tamponade; this sound is directly related to the extent to which ventricular filling is restricted to early diastole, being abruptly checked at the peak of early filling when the heart reaches the fixed volume imposed by the constricting shell surrounding it.

Diagnosis

ECG and imaging studies Constrictive pericarditis is difficult to diagnose, frequently being misdiagnosed for prolonged periods as liver disease or idiopathic pleural effusion. Clinical diagnosis of constrictive pericarditis depends on the recognition of increased venous pressure in a patient who may not have other obvious signs or symptoms of heart disease. The heart size and lung fields often appear normal in the chest radiograph, and the ECG shows only minor nonspecific abnormalities. Echocardiography is also nondiagnostic in many instances, although the appearance of abnormal septal motion and pericardial thickening often provide clues. Transesophageal echocardiography and chest CT are superior to echocardiography for the demonstration of pericardial thickening [*see Figure 7*]; however, the pericardium is not measurably thicker than normal in noninvasive imaging studies in some patients with constriction.[31]

As in cardiac tamponade, pulsed wave Doppler studies show exaggerated respiratory variation in the mitral and tricuspid diastolic flow velocity in most cases of constrictive pericarditis. Doing the study in the upright position improves the sensitivity of

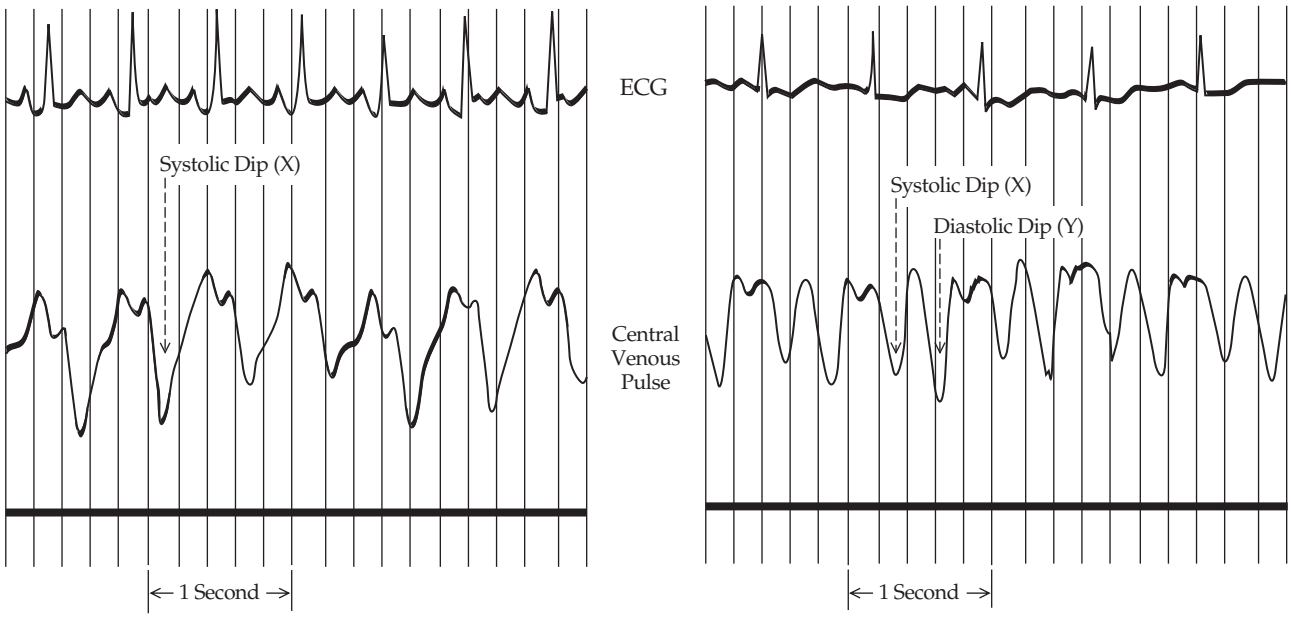


Figure 6 **Differences between the central venous pulse contours characteristic of cardiac tamponade (left) and chronic constrictive pericarditis (right) provide the basis for differential diagnosis. The pressure contour in a patient with pericardial effusion and tamponade has a prominent systolic dip (X) but little or no diastolic deflection. The central venous pulse pattern in a patient with chronic constrictive pericarditis displays an M or W contour consisting of both systolic dip (X) and diastolic dip (Y), with the Y descent being more prominent.**

this test.[32] False positive results occur in patients with chronic obstructive pulmonary disease, but that can be recognized by performing Doppler studies of flow velocity in the superior vena cava; the changes in velocity with respiration are much greater in pulmonary disease than in constrictive pericarditis.[33]

Cardiac catheterization Cardiac catheterization shows characteristic abnormalities, with increased central venous pressure, nondilated and normally contracting right and left ventricles, and near equilibration of the cardiac filling pressures of the right and left sides. These features may also be present in idiopathic restrictive cardiomyopathy or in specific myocardial diseases, especially cardiac amyloidosis; in such cases, the demonstration of pericardial thickness by CT or MRI and the use of endomyocardial biopsy are helpful.[34] Many other clues can assist in this differential diagnosis [*see Table 1*]. The increased interdependence of the two ventricles in constrictive pericarditis causes the right and left ventricular systolic pressures to vary out of phase with each other in respiration; in conditions other than constrictive pericarditis, the two systolic pressures increase and decrease together with respiration. This is perhaps the most useful information yielded by cardiac catheterization in patients with suspected constrictive pericarditis.[35]

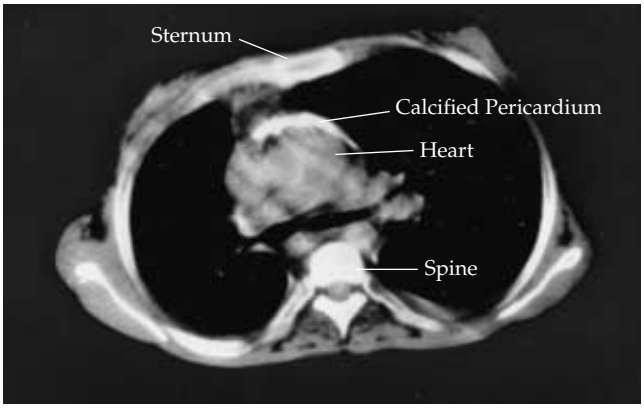


Figure 7 **In this CT scan of the chest of a patient with chronic constrictive pericarditis, the dense layer on the anterior surface of the heart represents thickened and partially calcified pericardium.**

Treatment

Constrictive pericarditis occasionally resolves spontaneously when it develops as a complication of acute pericarditis.[35] In nearly all instances, however, relief of constrictive pericarditis requires surgical stripping and removal of both layers of the adherent, constricting pericardium. This operation is far more difficult to perform than the operation for relief of pericardial effusion. The operation must be thorough, which carries the risk of hemorrhage from perforations in the wall of the heart. Inadequate long-term relief after surgical removal of the pericardium may reflect the presence of associated myocardial disease, particularly in instances of radiation-induced pericardial disease.[36] In most other forms of constrictive pericarditis, however, myocardial function is normal.

Cardiac Tumors

Cardiac tumors may be either primary or secondary, and they may be either benign or malignant. Metastatic cardiac involvement occurs 20 to 40 times more frequently than primary tumors. However, primary tumors are often benign and curable by surgery.

METASTATIC TUMORS

About 10% of patients who die of malignant disease have metastatic cardiac involvement, but the metastases produce symptoms in only 5% to 10% of the affected patients. Neoplasms

Table 1 Clinical Features That Differentiate Constrictive Pericarditis from Amyloidosis and Idiopathic Restrictive Cardiomyopathy

Clinical Feature	*Constrictive Pericarditis*	*Cardiac Amyloidosis*	*Idiopathic Restrictive Cardiomyopathy*
Early diastolic sound (S_3 or pericardial knock)	Frequent	Occasional	Occasional
Late diastolic sound (S_4)	Rare	Frequent	Frequent
Atrial enlargement	Mild or absent	Marked	Marked
Atrioventricular or intraventricular conduction defect	Rare	Frequent	Frequent
QRS voltage	Normal or low	Low	Normal or high
Mitral or tricuspid regurgitation	Rare	Frequent	Frequent
Paradoxical pulse	Frequent but usually mild	Rare	Rare
Exaggerated variation in mitral and tricuspid flow velocity with respiration, out of phase	Usual	Rare	Rare

particularly likely to metastasize to the heart are cancers of the lung or breast, melanoma, leukemia, and lymphoma.[37]

The most frequent clinical manifestation is pericardial effusion with cardiac tamponade. In such cases, the mass of the tumor is often relatively small. Extensive solid tumor in and around the heart is less common but may resemble constrictive pericarditis or effusive-constrictive pericarditis. Invasion of the myocardium most often manifests clinically as arrhythmias; atrial flutter and atrial fibrillation are particularly common.

Usually, the only effective treatment in metastatic involvement of the heart is relief of cardiac tamponade. Otherwise, treatment depends on the nature of the primary tumor.

PRIMARY BENIGN TUMORS

Eighty percent of all primary cardiac tumors are benign; myxomas account for more than half of these in adults, whereas rhabdomyomas and fibromas are the most common benign cardiac tumors in children.[38-40] Cardiac rhabdomyomas in infancy and childhood have a high incidence of spontaneous regression; although they are sometimes responsible for a remarkable syndrome of paroxysmal ventricular tachycardia in infancy, this syndrome can be cured by surgical removal of the tumor. Echocardiography and MRI are both excellent methods for demonstrating intracardiac tumors [*see Figures 8 and 9*].

Myxoma Myxomas consist of scattered stellate cells embedded in a mucinous matrix. They are found in the cavities of the heart, attached to the endocardial wall (or, in rare cases, attached to one of the heart valves) by either a narrow stalk or a broader pedicle. The tumor often shows considerable movement within the cardiac chamber during the cardiac cycle [*see Figure 8*]. About 70% of myxomas are in the left atrium; the rest are mostly in the right atrium. Echocardiography is a reliable method with which to predict tumor size and morphology.[41]

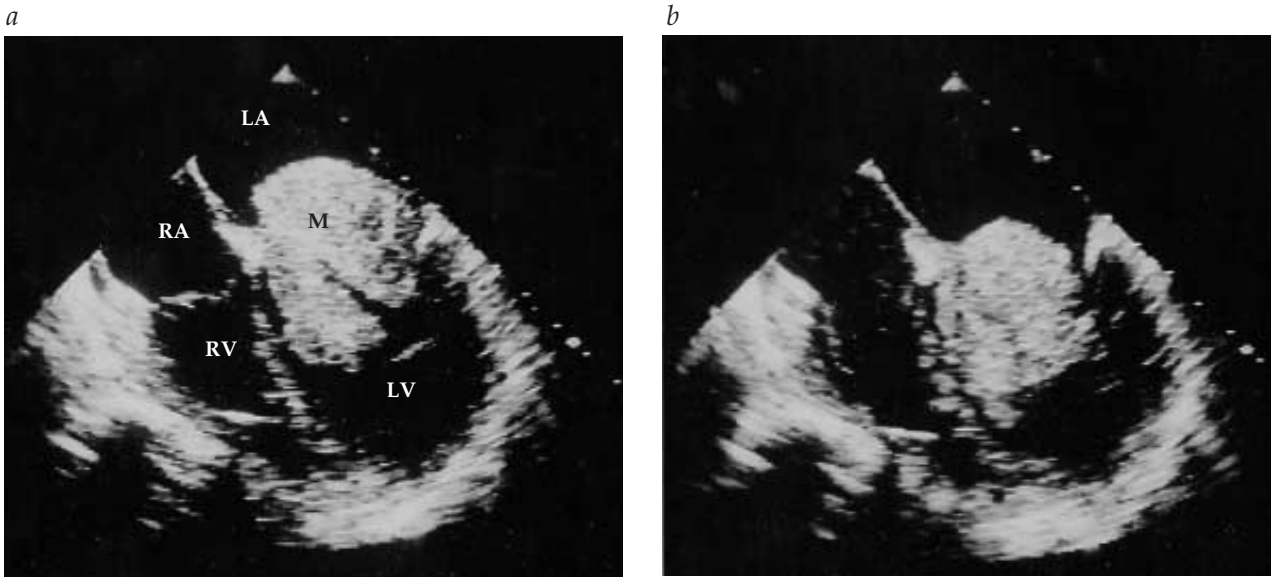


Figure 8 **This transesophageal echocardiogram demonstrates a large myxoma (M) in the left atrium (LA) of a 23-year-old man. The picture on the left (*a*) was taken during early systole, and the picture on the right (*b*) was taken during early diastole. The marked mobility of the tumor is evident as it moves from the left atrium to the left ventricle (LV). The right atrium (RA) and right ventricle (RV) are also visible.**

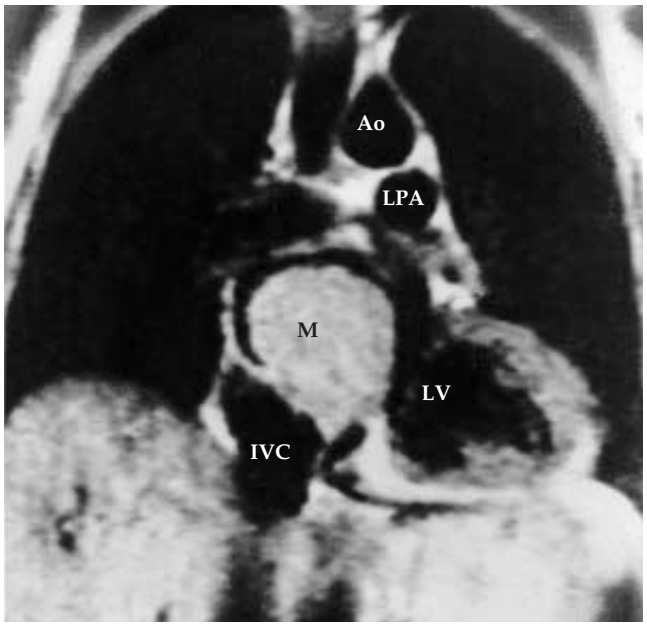


Figure 9 **This magnetic resonance image, coronal view, shows a large left atrial myxoma (M). The left ventricle (LV), aorta (Ao), left pulmonary artery (LPA), and inferior vena cava (IVC) are also visible.**

Myxomas are most often manifested clinically by mechanical hemodynamic effects, which often simulate mitral or tricuspid stenoses when they obstruct the valve orifice. They may simulate mitral or tricuspid regurgitation when they interfere with valve closure or cause a so-called wrecking-ball type of trauma to the mitral or tricuspid valve. Intermittent obstruction of the valve orifice can lead to such dramatic symptoms as syncope or to remarkable changes in physical signs that are sometimes related to changes in body position.

Myxomas also cause thromboembolic complications when portions of the tumor or thrombi from the surface of the tumor are detached. Another manifestation is a constitutional disturbance consisting of fatigue, fever, erythematous rash, myalgias, and weight loss, accompanied by anemia and an increased erythrocyte sedimentation rate. The constitutional symptoms may be caused by production of interleukin-6 by the myxoma.[42]

About 5% of cases of cardiac myxoma are familial, multicentric, or associated with a genetic syndrome that includes cutaneous lentiginosis, cutaneous myxomas, myxoid fibroadenomas of the breast, pituitary adenomas, adrenocortical micronodular hyperplasia with Cushing syndrome, and Sertoli cell tumors of the testis. These cases are referred to as complex myxoma, myxoma syndrome, or the Carney complex. A genetic mutation underlying this syndrome has been identified.[43]

Surgical treatment of cardiac myxomas is usually curative, particularly if the resection includes the portion of the atrial septum or atrial free wall from which the tumor has arisen. Recurrence and distant metastases are rare except in myxoma syndrome.[44]

Papillary fibroelastoma Papillary fibroelastomas are small tumors, usually attached to cardiac valves, that can be a cause of cardioembolic stroke. They have been recognized with increasing frequency since echocardiography has come into more widespread use.[45] Surgery may be indicated, especially if embolism recurs.

PRIMARY MALIGNANT TUMORS

Most malignant tumors of the heart are sarcomas, of either the spindle cell or the round cell type. Spindle cell tumors include fibrosarcomas, hemangiosarcomas, leiomyosarcomas, rhabdomyosarcomas, and fibromyxosarcomas. Round cell tumors include lymphosarcomas or reticulum cell sarcomas. Primary lymphoma of the heart, which is usually seen only in immune-compromised patients, is increasing in incidence.[46]

Malignant tumors are more apt to occur in the right side of the heart than in the left, being about equally frequent in the right atrium and the right ventricle. Signs and symptoms usually stem from intracavitary growth of the tumor, causing obstructive phenomena that simulate congestive heart failure. Pericardial effusion and tamponade are also common.

Malignant pericardial mesothelioma usually presents as pericardial effusion with tamponade or as subacute constrictive pericarditis.

Although surgical excision of malignant cardiac tumors is often attempted, cure is only rarely achieved. The tumors are usually unresponsive to radiation or chemotherapy, and most are fatal within a few months.

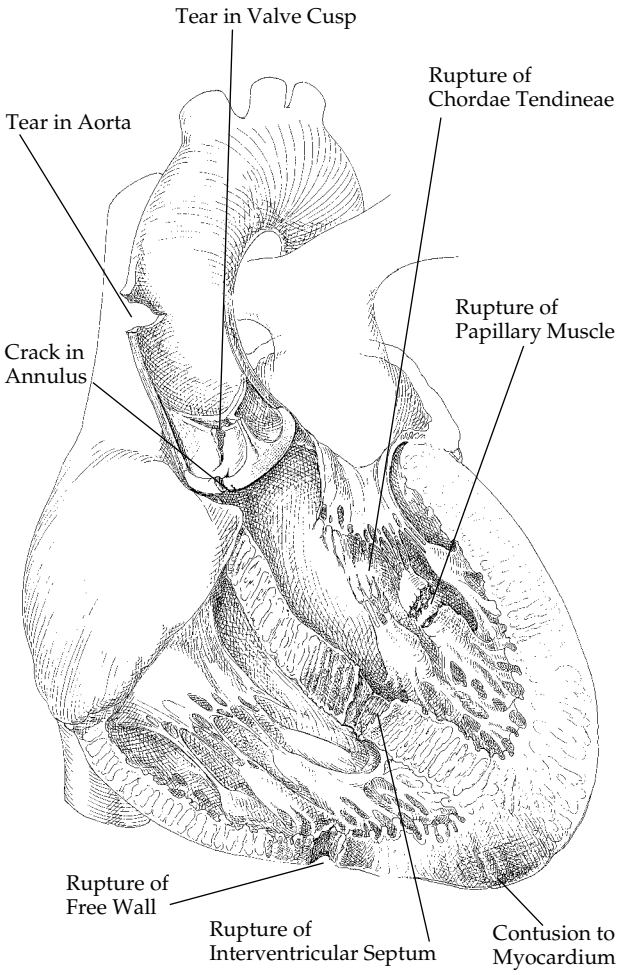


Figure 10 **Blunt trauma, such as that caused by the impact of the chest against the steering wheel in an automobile accident, may injure various cardiac structures. Myocardial contusion is the most frequent injury, but rupture may occur at several sites, including the interventricular septum, the walls of the cardiac chambers, the papillary muscles, and the chordae tendineae. The shearing forces that accompany abrupt deceleration may also cause tearing of the aorta and the valve cusps and cracking of the annulus.**

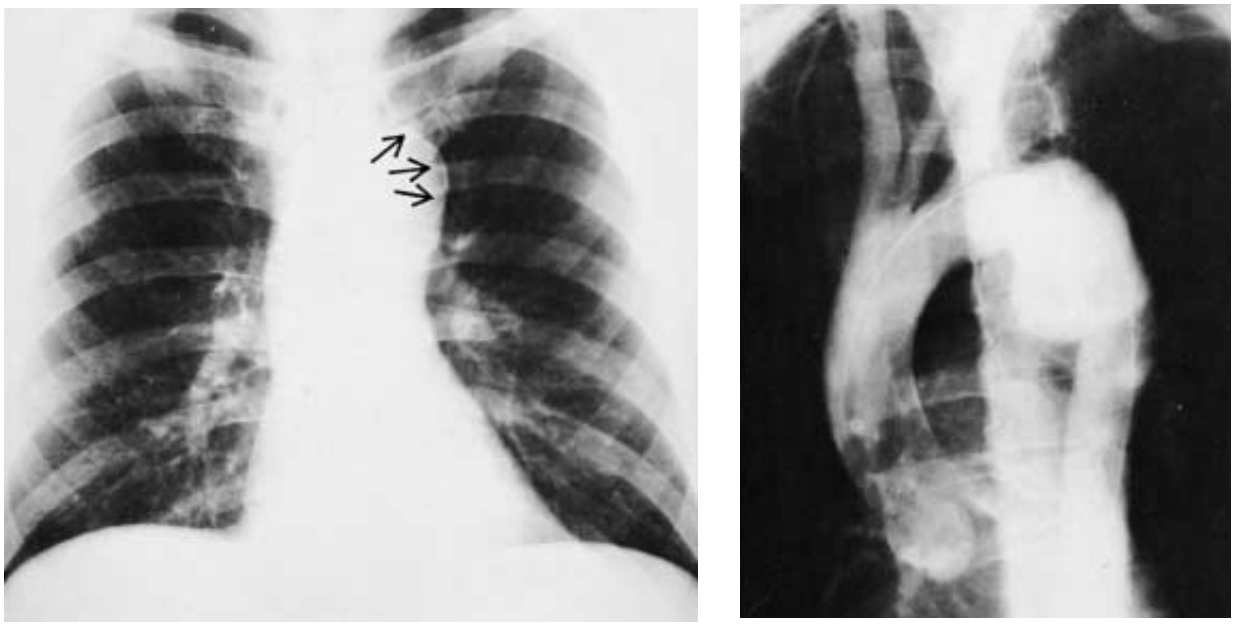


Figure 11 **A posteroanterior chest x-ray (left) reveals a posttraumatic aortic aneurysm at the aortic isthmus. Calcification (arrows) is evident in the wall of the aneurysm, which arose in a 45-year-old policeman who had sustained chest injuries in a motorcycle accident 24 years earlier. The lesion had gradually enlarged during a 10-year period of observation after its discovery. The aneurysm, outlined by angiography (right), was successfully excised.**

Cardiovascular Trauma

Cardiovascular injury may be either blunt (i.e., nonpenetrating) or penetrating.[47] Automobile accidents are the most common cause of blunt cardiovascular trauma; gunshots and stabbings are the most common causes of penetrating trauma. Both types of injury can damage the myocardium, the valves, the coronary arteries, the pericardium, and the great vessels, especially the aorta [*see Figure 10*]. Diagnosis in such instances is often difficult because the associated injuries can mask the cardiovascular trauma; cardiac trauma should therefore be suspected in all patients with chest injuries or severe generalized trauma.

BLUNT CARDIAC TRAUMA

Myocardial Contusion

Myocardial contusion is the most common blunt injury.[48] The right ventricle, because of its immediately substernal location, is the chamber most often involved. The pathologic changes in myocardial contusion consist of myocardial necrosis with hemorrhage, which may range in severity from scattered petechiae to intramural extravasations with associated transmural necrosis. In some instances, coronary arterial occlusion with secondary myocardial infarction is present. Seemingly innocuous blows to the chest by missiles such as baseballs or hockey pucks may cause sudden arrhythmic death, probably when they strike directly over the heart during the vulnerable portion of the T wave and induce ventricular fibrillation.[49]

The most important complication of myocardial contusion is cardiac arrhythmia. Hypotension, intracardiac thrombus, congestive heart failure, and cardiac tamponade occur occasionally.

Myocardial contusion is best recognized clinically by echocardiography, which shows localized areas of impaired wall motion. Transesophageal echocardiography is often superior to the transthoracic evaluation.[50] The abnormalities of wall motion usually resolve within a few days. Increases in the concentrations of creatine kinase (CK) and its MB fraction (CK-MB) in the blood are difficult to interpret because of the release of CK from injured skeletal muscle. Cardiac troponin-I is a more specific marker.[51] Diffuse nonspecific ST-T abnormalities in the electrocardiogram are common in injured patients, even in the absence of echocardiographically detected abnormalities in wall motion. However, localized changes, especially ST segment elevation, are more specific for contusion. Patients with Q wave infarct patterns and irreversible wall motion defects are likely to have a coronary arterial occlusion secondary to trauma with myocardial infarction. Severe contusions may also lead to the formation of traumatic left ventricular aneurysms or pseudoaneurysms that are sometimes detected months or years after the initial trauma. Management of myocardial contusion is conservative unless one or more specific complications (e.g., arrhythmia, tamponade, aneurysm, or perforation) are present.

Valvular Injury

Blunt trauma may injure any of the cardiac valves and lead to valvular regurgitation. Traumatic valvular regurgitation is more likely to be recognized after the patient has recovered from the acute injuries; it is less likely to play a major role in the early postinjury course.[52]

Aortic Injury

Injuries of the aorta result from abrupt deceleration in violent thoracic trauma and are relatively common. The most common injury results from a tear in the wall of the aorta at a point just distal to the left subclavian artery, where the aorta is fixed to the dorsal thoracic cage. Usually, complete transection of the aorta is quickly fatal. Less extensive tears can result in a localized hematoma or a localized false aneurysm. Such aneurysms may be recognized months or years after the initial injury, when they cause symptoms by gradually enlarging, or may be discovered incidentally by chest radiography [*see Figure 11*].

A widened mediastinal shadow in the chest radiograph is often the first clue to the presence of a traumatic aortic rupture. The chest CT and the transesophageal echocardiogram are useful aids in making the diagnosis.

At least 50% of patients with aortic rupture die before they reach a hospital, often of injuries unrelated to the aortic trauma. Surgical therapy should be undertaken as soon as possible, even if the bleeding from the ruptured aorta has stabilized; such therapy results in survival of about 80% of those patients who are still alive when they reach a medical facility. Resection of the injured segment and replacement with a prosthetic graft are usually required.

Usually, false aneurysms that are diagnosed long after the initial injury should be resected electively. Their natural history in most instances is to enlarge gradually and eventually rupture.

PENETRATING TRAUMA

Penetrating injuries of the heart and great vessels are caused either by stab wounds or by gunshot wounds. Any of the cardiac chambers or great vessels may be punctured, and injury of multiple structures is common. The most common sites of involvement, in order of decreasing frequency, are the right ventricle, the left ventricle, the right atrium, and the left atrium.

Stab wounds and, especially, bullet wounds of the heart often are immediately fatal. However, if the penetrating wound is relatively small, cardiac tamponade can occur, and the buildup of pressure in the pericardial sac may help reduce the severity of bleeding and thus increase the chance of survival.

Other sequelae of penetrating trauma include laceration of the aorta or the pulmonary artery; defects in the ventricular or atrial septum; fistulas between the great vessels and between the coronary arteries and the cardiac chambers; coronary arterial fistulas; puncture of any of the heart valves; and atrioventricular block as a result of disruption of the conduction system. Occasionally, a missile that lodges in a cardiac chamber or in one of the great arteries will embolize to a distal site, whereas missiles that initially lodge in distal sites may work their way through the veins to lodge in the chambers of the heart or the pulmonary artery.

The existence of intracardiac shunts, fistulas between the heart and great vessels, coronary arterial fistulas, or valve disruption is usually suggested by the presence of new murmurs. Echocardiography and Doppler studies generally allow precise definition, localization, and quantitation of the lesions.

Penetrating cardiac trauma usually requires prompt surgical intervention, even performance of a thoracotomy in the emergency department. The immediate availability of echocardiography in the emergency department or trauma unit is extremely valuable in management of penetrating wounds of the heart. The survival rate of patients who reach the hospital alive is about 50% for those with knife wounds of the heart and 30% for patients with gunshot wounds.[53]

ELECTRICAL INJURY

Electrical injury, a special type of cardiac trauma, is produced by a direct electrical effect on the tissues, the generation of heat from the passage of current from a high-voltage source through tissue with high electrical resistance, extreme release of catecholamines, or extreme autonomic stimulation. Sudden cardiac arrest occurs with exposure to either household AC current or a lightning strike. Lightning strikes also cause myocardial injury, which may be extensive; in such cases, ECG patterns change, concentrations of cardiac enzymes increase, and wall motion exhibits abnormalities. Pericarditis and pericardial effusion also occur. The abnormalities usually resolve within several weeks.[54]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Baljepally R, Spodick DH: PR-segment deviation as the initial electrocardiographic response in acute pericarditis. Am J Cardiol 81:1505, 1998

2. Spodick DH: Acute pericarditis: current concepts and practice. JAMA 289:1150, 2003

3. Adler Y, Finkelstein J, Guindo A, et al: Colchicine treatment for recurrent pericarditis: a decade of experience. Circulation 97:2183, 1998

4. Akashi H, Tayama K, Ishihara K: Isolated primary chylopericardium. Jpn Circ J 63:59, 1999

5. Spodick DH: Pathophysiology of cardiac tamponade. Chest 113:1372, 1998

6. Smith WH, Beacock DJ, Goddard AJ, et al: Magnetic resonance evaluation of the pericardium. Br J Radiol 74:384, 2002

7. Larose E, Ducharme A, Mercier LA, et al: Prolonged distress and clinical deterioration before pericardial drainage in patients with cardiac tamponade. Can J Cardiol 16:331, 2000

8. Roosen J, Frans E, Wilmer A, et al: Comparison of premortem clinical diagnoses in critically ill patients and subsequent autopsy findings. Mayo Clin Proc 75:562, 2000

9. Mercé J, Sagrista-Sauleda J, Permanyer-Miralda G, et al: Correlation between clinical and Doppler echocardiographic findings in patients with moderate and large pericardial effusion: implications for the diagnosis of cardiac tamponade. Am Heart J 138:759, 1999

10. Shaver JA, Reddy PS, Curtiss EI, et al: Noninvasive/invasive correlates of exaggerated ventricular interdependence in cardiac tamponade. J Cardiol 37:71, 2001

11. Beppu S, Tanaka N, Nakatani S, et al: Pericardial clot after open heart surgery: its specific localization and haemodynamics. Eur Heart J 14:230, 1993

12. Sagristà-Sauleda J, Angel J, Permanyer-Miralda G, et al: Long-term follow-up of idiopathic chronic pericardial effusion. N Engl J Med 341:2054, 1999

13. Soler-Soler J, Sagrista-Sauleda J, Permanyer-Miralda G: Management of pericardial effusion. Heart 86:235, 2001

14. Nugue O, Millaire A, Porte H, et al: Pericardioscopy in the etiologic diagnosis of pericardial effusion in 141 consecutive patients. Circulation 94:1635, 1996

15. Tsang TS, Barnes ME, Gersh BJ, et al: Outcomes of clinically significant idiopathic pericardial effusion requiring intervention. Am J Cardiol 91:704, 2003

16. Allen KB, Faber LP, Warren WH, et al: Pericardial effusion: subxiphoid pericardiostomy versus percutaneous catheter drainage: Ann Thorac Surg 67:437, 1999

17. Flores RM, Jaklitsch MT, DeCamp MM Jr, et al: Video-assisted thoracic surgery pericardial resection for effusive disease. Chest Surg Clin N Am 8:835, 1998

18. Tsang T, Enriquez-Sarano M, Freeman WK, et al: Consecutive 1127 therapeutic echocardiographically guided pericardiocenteses: clinical profile, practice patterns, and outcomes spanning 21 years. Mayo Clin Proc 77:429, 2002

19. Ziskind AA, Pearce AC, Lemmon CC, et al: Percutaneous balloon pericardiotomy for the treatment of cardiac tamponade and large pericardial effusions: description of technique and report of the first 50 cases. J Am Coll Cardiol 21:1, 1993

20. Piovaccari G, Ferretti RM, Prati F, et al: Cardiac disease after chest irradiation for Hodgkin's disease: incidence in 108 patients with long followup. Int J Cardiol 49:39, 1995

21. Wilkes JD, Fidias P, Vaickus L, et al: Malignancy-related pericardial effusion. 127 cases from the Roswell Park Cancer Institute. Cancer 76:1377, 1995

22. Bardales RH, Stanley MW, Schaefer RF, et al: Secondary pericardial malignancies: a critical appraisal of the role of cytology, pericardial biopsy, and DNA ploidy analysis. Am J Clin Pathol 106:29, 1996

23. Laham RJ, Cohen DJ, Kuntz RE, et al: Pericardial effusion in patients with cancer: outcome with contemporary management strategies. Heart 75:67, 1996

24. Trautner BW, Darouiche RO. Tuberculous pericarditis: optimal diagnosis and management. Clin Infect Dis 33:954, 2001

25. Mayosi BM, Ntsekhe M, Volmink JA, et al: Interventions for treating tuberculous pericarditis. Cochrane Database Syst Rev (4):CD000526, 2002

26. Rerkpattanapipat P, Wongpraparut N, Jacobs LE: Cardiac manifestations of acquired immunodeficiency syndrome. Arch Intern Med 160:602, 2000

27. Fletcher SL, Bodenham AR: Safe placement of central venous catheters: where should the tip of the catheter lie? (editorial). Br J Anaesth 85:188, 2000

28. Von Sohsten R, Kopistansky C, Cohen M, et al: Cardiac tamponade in the "new device" era: evaluation of 6999 consecutive percutaneous coronary interventions. Am Heart J 140:279, 2000

29. Dardas P, Tsikaderis D, Ioannides E, et al: Constrictive pericarditis after coronary artery bypass surgery as a cause of unexplained dyspnea: a report of five cases. Clin Cardiol 21:691, 1998

30. Veeragandham RS, Goldin MD: Surgical management of radiation-induced heart disease. Ann Thorac Surg 65:1014, 1998

31. Ling L, Oh J, Tei C, et al: Pericardial thickness measured with transesophageal echocardiography: feasibility and potential clinical usefulness. J Am Coll Cardiol

29:1317, 1997

32. Oh JK, Tajik AJ, Appleton CP, et al: Preload reduction to unmask the characteristic Doppler features of constrictive pericarditis: a new observation. Circulation 95:796, 1997

33. Boonyaratavej S, Oh JK, Tajik AJ, et al: Comparison of mitral inflow and superior vena cava Doppler velocities in chronic obstructive pulmonary disease and constrictive pericarditis. J Am Coll Cardiol 32:2043, 1998

34. Hancock EW: Differential diagnosis of restrictive cardiomyopathy and constrictive pericarditis. Heart 86:343, 2001

35. Nishimura RA. Constrictive pericarditis in the modern era: a diagnostic dilemma. Heart 86:619, 2001

36. Ling LH, Oh JK, Schaff HV, et al: Constrictive pericarditis in the modern era: evolving clinical spectrum and impact on outcome after pericardiectomy. Circulation 100:1380, 1999

37. Silvestri F, Bussani R, Pavletic N, et al: Metastases of the heart and pericardium. G Ital Cardiol 27:1252, 1997

38. Shapiro LM: Cardiac tumours: diagnosis and management. Heart 85:218 2001

39. Meng Q, Lai H, Lima J, et al: Echocardiographic and pathologic characteristics of primary cardiac tumors: a study of 149 cases. Int J Card 84:69, 2002

40. Stiller BB, Hetzer R, Meyer R, et al: Primary cardiac tumours: when is surgery necessary? Eur J Cardiovasc Surg 20:1002, 2001

41. Acebo E, Val-Bernal JF, Gomez-Roman JJ, et al: Clinicopathologic study and DNA analysis of 37 cardiac myxomas: a 28-year experience. Chest 123:1379, 2003

42. Pinede L, Duhaut P, Loire R: Clinical presentation of left atrial cardiac myxoma: a series of 112 consecutive cases. Medicine (Baltimore) 80:159, 2001

43. Matyakhina L, Pack S, Kirschner LS, et al: Chromosome 2 (2p16) abnormalities in Carney complex tumors. J Med Genet 40:268, 20032000

44. Bhan A, Mehrotra R, Choudhary SK, et al: Surgical experience with intracardiac myxomas: long-term follow-up. Ann Thorac Surg 66:810, 1998

45. Sun JP, Asher CR, Yang XS, et al: Clinical and echocardiographic characteristics of papillary fibroelastomas: a retrospective and prospective study in 162 patients. Circulation 103:2687, 2001

46. Rolla G, Bertero MT, Pastena G, et al: Primary lymphoma of the heart. A case report and review of the literature. Leuk Res 26:117, 2002

47. Baum VC: The patient with cardiac trauma. J Cardiothorac Vasc Anesth 14:71, 2000

48. Pretre R, Chilcott M: Blunt trauma to the heart and great vessels. N Engl J Med 336:626, 1997

49. Maron BJ, Gohman TE, Kyle SB, et al: Clinical profile and spectrum of commotio cordis. JAMA 287:1142, 2002

50. Lindstaedt M, Germing A, Lawo T, et al: Acute and long-term clinical significance of myocardial contusion following blunt thoracic trauma: results of a prospective study. J Trauma 52:479, 2002

51. Collins JN, Cole FJ, Weireter LJ, et al: The usefulness of serum troponin levels in evaluating cardiac injury. Am Surg 67:821, 2001

52. Wall MJJ, Soltero ER: Trauma to cardiac valves. Curr Opin Cardiol 17:188, 2002

53. Tyburski JG, Astra L, Wilson RF, et al: Factors affecting prognosis with penetrating wounds of the heart. J Trauma 48:587, 2000

54. Muehlberger T, Voft PM, Munster A: The long-term consequences of lightning injuries. Burns 27:829, 2001

XV CONGENITAL HEART DISEASE

Larry T. Mahoney, m.d.
David J. Skorton, m.d.

Congenital diseases of the heart and vasculature are the most common birth defects, occurring in approximately eight per 1,000 live births. Some patients with congenital heart defects (CHDs) remain asymptomatic for many years; others survive to adulthood, thanks to the impressive progress in medical and surgical management made in recent decades. This relatively high incidence, coupled with improved management, has resulted in a large adult population of these patients: it has been estimated that almost one million adults with CHDs are currently living in the United States.[1] A broad range of clinicians must now become more knowledgeable about the care of these patients,[2] including issues such as endocarditis prophylaxis[3] [*see Sidebar* Selected Internet Resources for Congenital Heart Disease]. This chapter reviews CHDs most likely to be encountered in adult patients.

Acyanotic Disorders—Shunts

Atrial Septal Defects

Atrial septal defects (ASDs) occur in three main locations [*see Figure 1*]: the region of the fossa ovalis (such defects are termed ostium secundum ASDs); the superior portion of the atrial septum near the junction with the superior vena cava (SVC) (sinus venosus ASDs); and the inferior portion of the atrial septum near the tricuspid valve annulus (ostium primum ASDs). The ostium primum ASDs are considered to be part of the spectrum of atrioventricular septal defects (AVSDs) [*see* Atrioventricular Septal Defects, *below*].

Ostium secundum ASDs are the most common variety, accounting for over half of ASDs. A frequent accompanying defect is mitral valve prolapse. Relatively less prevalent is the sinus venosus defect. Anomalous pulmonary venous return is a common associated abnormality. The proximity of the sinoatrial node to the ASD may lead to sinoatrial node dysfunction and atrial arrhythmias.

Pathophysiology

ASDs are associated with left-to-right shunts of varying degrees. The main determinants of the direction and magnitude of shunt flow are the size of the defect and the relative compliances of the left ventricle (LV) and right ventricle (RV).[4]

Clinical Presentation

Most patients with ostium secundum or sinus venosus ASD are asymptomatic through young adulthood. As the patient reaches middle age, compliance of the LV may decrease, increasing the magnitude of left-to-right shunting. Long-standing atrial dilatation may lead to a variety of atrial arrhythmias, including premature atrial contractions, supraventricular tachycardia, and atrial fibrillation. A substantial number of middle-aged patients will report dyspnea, particularly with exertion, even if they do not have pulmonary hypertension. Approximately 10% of patients with ostium secundum ASDs will progress to pulmonary hypertension associated with pulmonary vascular obstructive disease (Eisenmenger syndrome) [*see* Eisenmenger Syndrome, *below*]. As the pulmonary pressure rises, the left-to-right shunt will diminish and eventually be replaced by a right-to-left shunt; cyanosis and pulmonary hypertension will develop.

The hallmark of the physical examination in ASD is the wide and fixed splitting of the second heart sound. A systolic murmur (from increased pulmonary flow) is common, and if a large left-to-right shunt is present, the additional flow across the tricuspid valve may lead to a diastolic rumble reminiscent of tricuspid stenosis.

Laboratory Tests

All patients with suspected ASD should have an electrocardiogram, a chest x-ray, and an echocardiogram.

Electrocardiography The QRS axis usually is normal in ostium secundum ASD but may be slightly rightward, and an rSR′ pattern is common in the right precordial leads. In sinus venosus ASD, the axis may be normal or relatively horizontal (less than 30°). Ectopic atrial rhythms or other evidence of sinoatrial node dysfunction may be seen.

Radiologic studies The chest x-ray reveals enlargement of the right atrium (RA), the RV, and the main pulmonary artery. The pulmonary vessels exhibit diffuse enlargement because of increased pulmonary blood flow. Magnetic resonance imaging, magnetic resonance angiography (MRA), or cardiac catheterization will identify anomalous pulmonary veins; these modalities should be considered when there is suspicion of this associated abnormality in patients with sinus venosus ASD.

The patient with secundum ASD who has pulmonary hypertension may benefit from right-sided heart catheterization to ascertain the level of pulmonary arterial pressure and resistance.

Echocardiography Echocardiography can confirm the presence of an ASD, determine its size, permit calculation of shunt flow through it, and identify any associated anomalies.

Selected Internet Resources for Congenital Heart Disease

International Society for Adult Congenital Cardiac Disease (ISACCD)
http://www.isaccd.org
Professional resources, patient information, and newsletter.

Canadian Adult Congenital Heart Network and the Toronto Congenital Cardiac Centre for Adults at the University of Toronto (CACHNET)
http://www.cachnet.org
Information for physicians and patients.

Grown Up Congenital Heart Patients Association (GUCH)
http://www.guch.demon.co.uk
A United Kingdom site providing information and support for patients and their families.

PediHeart
http://www.pediheart.org
Practitioner and patient information; mailing list.

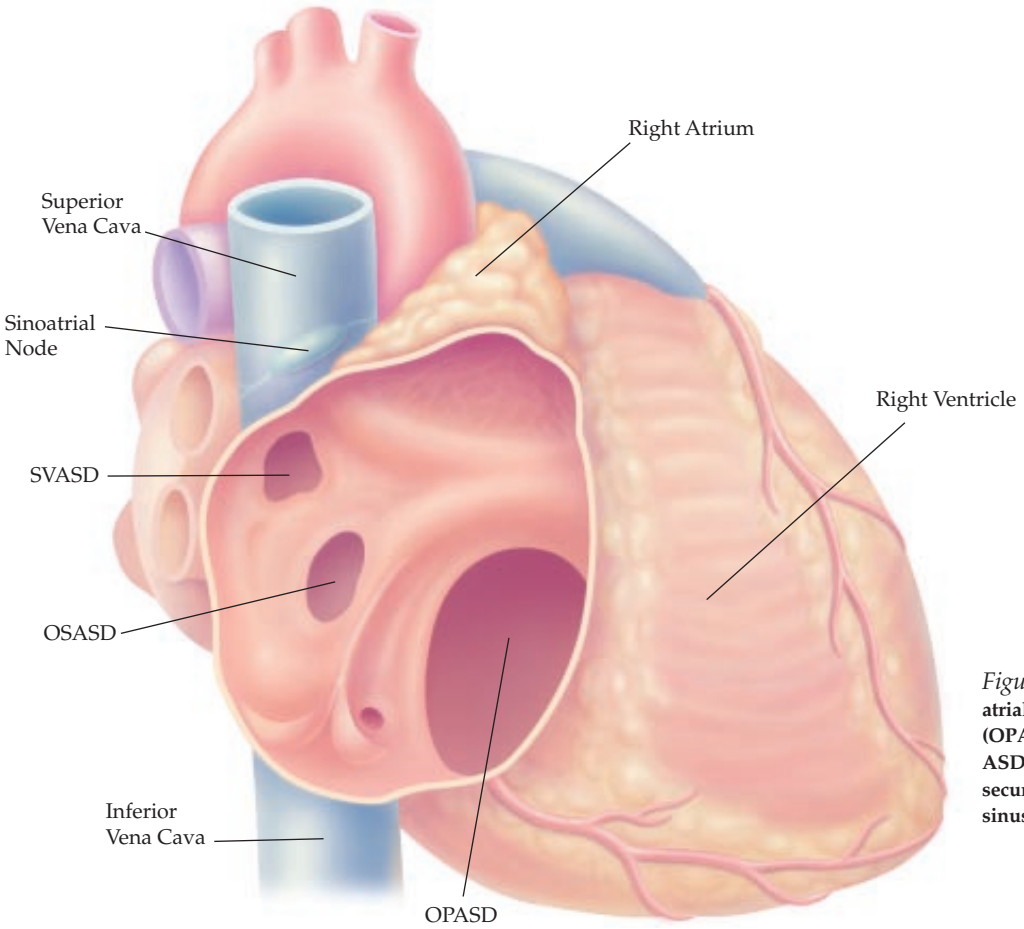


Figure 1 **Anatomy of atrial septal defects (ASDs). (OPASD—ostium primum ASD; OSASD—ostium secundum ASD; SVASD—sinus venosus ASD)**

Management

Large ASDs (defined as those with a pulmonary-to-systemic flow ratio [Qp:Qs] of over 1.5:1) should be closed to prevent the development of pulmonary hypertension and reduce the risk of paradoxical emboli. Direct surgical closure has been the method used, but devices are now available that permit catheterization-based closure of many defects.[5] Postclosure management includes periodic assessment for the development of atrial arrhythmias.[6] The need for endocarditis prophylaxis varies.

ATRIOVENTRICULAR SEPTAL DEFECTS

The septal leaflet of the tricuspid valve normally inserts into the septum slightly closer to the apex than does the septal leaflet of the mitral valve [*see Figure 2*]. Thus, the small portion of septal tissue superior to the tricuspid septal leaflet insertion separates the RA from the LV and so is called the atrioventricular septum. The term AVSD refers to a complex spectrum of disorders involving abnormalities of the atrioventricular septum and, frequently, the atrioventricular valves. Nomenclature for this spectrum of disorders has varied; synonymous terms include atrioventricular canal defect and endocardial cushion defect.

Pathophysiology

The spectrum of AVSDs ranges from a simple ostium primum ASD to a complete AVSD, which allows free communication among all four cardiac chambers. Variations of the anatomy of the anterior leaflet of the mitral valve and the septal leaflet of the tricuspid valve include a cleft or other abnormality in either or both of these leaflets; accessory chordae that attach in anomalous locations and alter function of the valve leaflets; or a common atrioventricular valve leaflet that bridges the septal defect. Physiologic consequences vary according to the extent of the anomaly; for example, the addition of a cleft mitral valve anterior leaflet adds varying degrees of mitral regurgitation (MR). Larger defects that also involve the ventricular septum, as well as complete AVSDs, can be associated with torrential left-to-right shunts or an admixture of venous and arterial blood.

Patients with an unrepaired complete AVSD are at risk for developing pulmonary hypertension. Eisenmenger syndrome is particularly common in AVSD patients who also have Down syndrome (trisomy 21).[7]

Clinical Presentation

Patients with isolated ostium primum ASDs may be asymptomatic until adulthood and then may present with fatigue, dyspnea, or symptoms related to atrial arrhythmias. Severe regurgitation of either atrioventricular valve can produce symptoms of heart failure or arrhythmias. Symptoms related to pulmonary hypertension occur in those patients who develop Eisenmenger syndrome.

Patients with only an ostium primum ASD will have clinical findings similar to those of patients with an ostium secundum ASD. The presence of a cleft in either atrioventricular valve will be associated with a pansystolic murmur. Finally, an additional pansystolic murmur can be found in patients with a complete AVSD.

Laboratory Tests

Electrocardiography Left axis deviation is present in the majority of patients. The combination of physical findings of

ASD along with left axis deviation on the ECG suggests the presence of an AVSD. RV conduction delay may be present as well.

Radiologic studies The chest x-ray shows cardiomegaly and pulmonary vascular engorgement because of the left-to-right shunt.

Echocardiography Echocardiography defines the specific anatomy and functional importance of the defects. Preoperative echocardiographic assessment includes estimation of the severity of atrioventricular valve regurgitation, the Qp:Qs ratio, and pulmonary arterial pressures. Postoperatively, echocardiography is used to identify and assess the significance of residual atrioventricular valve regurgitation or residual shunt.

Management

The rare patient who presents in adulthood with complete AVSD should be evaluated for pulmonary hypertension. If pulmonary pressures are normal or if pulmonary hypertension is not prohibitive (i.e., pulmonary vascular resistance is less than 50% of systemic vascular resistance), then surgical closure of the defect and repair of the atrioventricular valve anomalies should be undertaken. Postoperatively, patients are assessed for the adequacy of atrioventricular valve repair and are monitored for evidence of residual shunt. In patients with residual MR, management focuses on the need for and timing of reoperation, which may involve either repair or replacement of the mitral valve. The patient should also be followed for the development of atrial arrhythmias.

VENTRICULAR SEPTAL DEFECTS

VSDs are among the most common congenital cardiac disorders seen at birth but are less frequently seen as isolated lesions in adulthood. This is because most VSDs in infants either (1) are large and nonrestrictive (i.e., they permit equilibration of pressures between the ventricles) and therefore lead to heart failure, necessitating early surgical closure, or (2) are small and close spontaneously.

Classification systems for VSD vary but usually are referenced to the embryologic divisions of the ventricular septum into inlet, outlet, muscular, and membranous portions [*see Figure 3*]. The most common defects are perimembranous defects. Inlet VSDs, located more posteriorly, may be part of the spectrum of AVSDs (see above). Single or multiple defects may occur in the muscular septum (muscular VSD). Finally, outlet VSDs include subpulmonary defects, which may allow prolapse of an aortic cusp, leading to associated aortic regurgitation (AR).

Pathophysiology

Nonrestrictive VSDs permit equilibration of ventricular pressures between the RV and LV, whereas small defects produce a large pressure gradient across the defect, so right heart pressures remain normal. The magnitude of shunt flow across moderate or large VSDs depends on the relative resistances of the systemic versus the pulmonary vascular bed. Rarely, clinicians may encounter adult patients who have large, nonrestrictive defects in the absence of other lesions. Moderate pulmonic stenosis at either the valve or the subvalvular level may create increased resistance to right ventricular outflow sufficient to reduce the left-to-right shunt; consequently, patients with VSD and mild to moderate pulmonic stenosis may reach adulthood without experiencing symptoms. Adults with long-standing VSD and large shunts may develop Eisenmenger syndrome.

Clinical Presentation

With the exception of those patients who contract infective endocarditis or those with Eisenmenger syndrome, adults with VSD are asymptomatic.

The classic physical finding of a restrictive VSD is a harsh, fre-

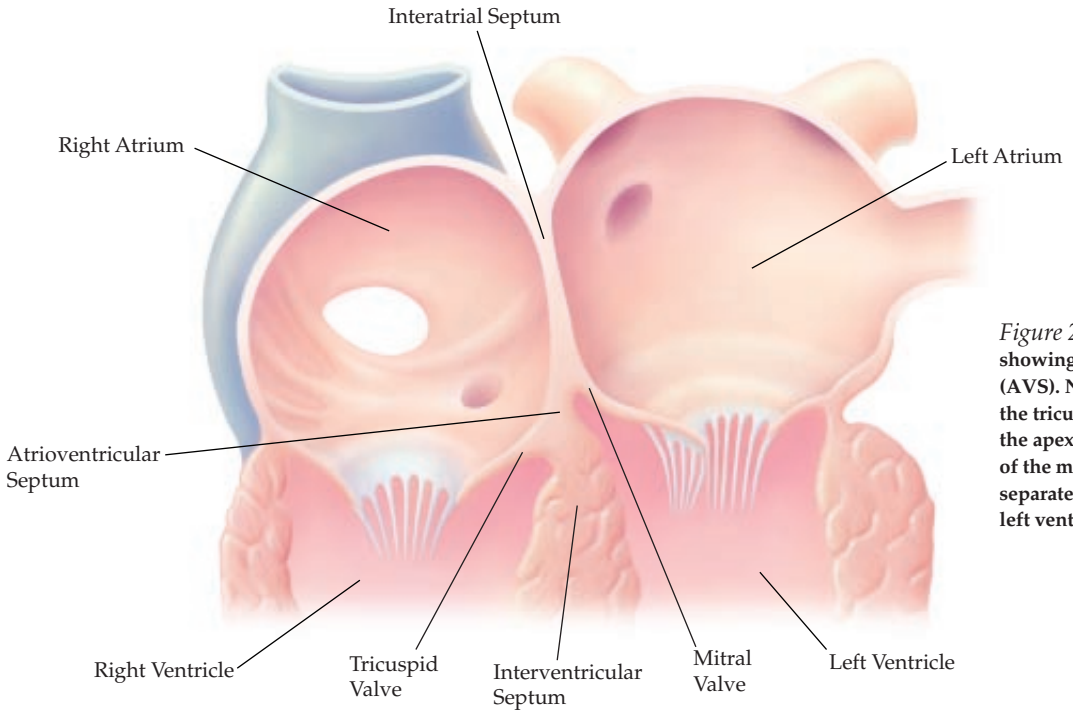


Figure 2 **Anatomic cross-section showing the atrioventricular septum (AVS). Note that the septal leaflet of the tricuspid valve inserts closer to the apex than does the septal leaflet of the mitral valve; thus, the AVS separates the right atrium from the left ventricle.**

quently palpable, pansystolic murmur heard best at the left lower sternal border. Patients who have large defects that allow equilibration of ventricular pressures may present with less impressive murmurs than patients with small defects; the reason is that with small defects, there is a large gradient between the LV and the RV, which results in severe turbulence across the defect. When aortic cusp prolapse occurs, the murmur of AR will be audible.

Laboratory Tests

Electrocardiography The ECG may be normal or show evidence of left ventricular hypertrophy (LVH) and a pattern of so-called diastolic overload, featuring prominent Q waves in left precordial leads V5 and V6 and in leads I and aV_L.

Radiologic studies The chest x-ray may be normal or show left ventricular enlargement and pulmonary arterial engorgement. Patients who have evidence of pulmonary hypertension should undergo right heart catheterization to determine the degree of pulmonary hypertension and the level of pulmonary resistance.

Echocardiography Echocardiography is the procedure of choice for identifying the location, size, and hemodynamic significance of a VSD; the interventricular gradient should be determined (to estimate RV pressure), and an assessment should be made of increased pulmonary blood flow.

Management

Patients with ventricular septal defects in which the Qp:Qs ratio is greater than 1.5:1 should be considered for surgical closure. Patients with pulmonary hypertension may undergo closure if pulmonary resistance is no more than about 50% of systemic resistance. Aortic cusp prolapse with resultant AR may diminish the shunt magnitude, but the presence of a prolapse constitutes an additional potential indication for closure.

Early VSD operative closures were performed through a right ventriculotomy, but now, many defects—particularly those in the perimembranous septum—are closed through a transatrial approach; such an approach leads to fewer problems with RV dysfunction and arrhythmias. Continual progress is being made in the deployment of transcatheter closure devices. Currently, however, surgical closure of VSD is still the most common approach. Postclosure management involves assessment for residual or recurrent VSD and atrial or ventricular arrhythmias, as well as assessment of RV function.

PATENT DUCTUS ARTERIOSUS

During fetal life, the ductus arteriosus connects the pulmonary artery to the aorta. Soon after birth, as a result of changes in circulating prostaglandin levels and arterial oxygen saturation, the ductus constricts; later, it closes permanently. Failure of the ductus to close leads to the condition termed patent ductus arteriosus (PDA).

Pathophysiology

The shunt from aorta to pulmonary artery increases pulmonary blood flow and return to the left heart. The size of the defect and the relative resistances of the pulmonary and systemic vascular beds determine the degree of shunting. Adults who have PDAs commonly present either with a small lesion without a large left-to-right shunt or with larger lesions and Eisenmenger syndrome.

Clinical Presentation

Except for patients with Eisenmenger syndrome, most adults with small to moderate PDAs will be asymptomatic, unless endarteritis supervenes.

The pathognomonic physical finding of PDA is the continuous murmur. A continuous murmur is one that is audible throughout systole and into diastole to any extent. The classic PDA murmur is machinelike and extends through systole and to variable degrees into diastole, peaking in intensity at the time of S_2. The runoff of blood into the pulmonary artery in diastole will produce a wide pulse pressure because of low aortic diastolic pressure.

Laboratory Tests

Electrocardiography The ECG in patients with PDA may be normal or may show evidence of LVH.

Radiologic studies If the shunt is small, the chest x-ray may be normal. Patients with larger shunts will have associated cardiomegaly and increased vascular markings. In adults, calcium may be noted within the wall of the ductus.

Echocardiography Echocardiography will identify the PDA and permit quantification of the Qp:Qs ratio.

Management

With the advent of reliable means of transcatheter closure of PDAs,[8] common practice is to recommend that most PDAs be closed. In rare cases, the ductus may need to be closed surgically if transcatheter closure is not successful. Postoperative management includes assessment for the need of a residual shunt, although this is uncommon. Patients who develop pulmonary hypertension are managed in the same way as those with Eisenmenger syndrome [*see* Eisenmenger Syndrome, *below*].

Acyanotic Disorders—Valvular Lesions

BICUSPID AORTIC VALVE AND OTHER CAUSES OF AORTIC STENOSIS

Abnormalities of the left ventricular outflow tract are common congenital cardiac disorders. In particular, as much as 2% of the population have congenitally bicuspid aortic valves. A bicuspid aortic valve may present as an incidental finding on physical examination or echocardiography done for other reasons; as significant aortic stenosis (AS) or AR; or when it results in infective endocarditis.

Pathophysiology

A stenotic bicuspid aortic valve will produce pressure overload of the LV, which leads to LVH and eventually to heart failure, angina pectoris, or sudden death from tachyarrhythmias. Similarly, the patient with an incompetent bicuspid aortic valve will exhibit LV dilatation, initially with normal systolic function; the condition will later progress to heart failure.

Clinical Presentation

On physical examination, the cardinal sign of a bicuspid aortic valve is an early systolic ejection click. If no significant hemodynamic abnormality is present, either no murmur or a soft ejection murmur may be heard; a very mild murmur of AR is not uncommon, even with hemodynamically insignificant bi-

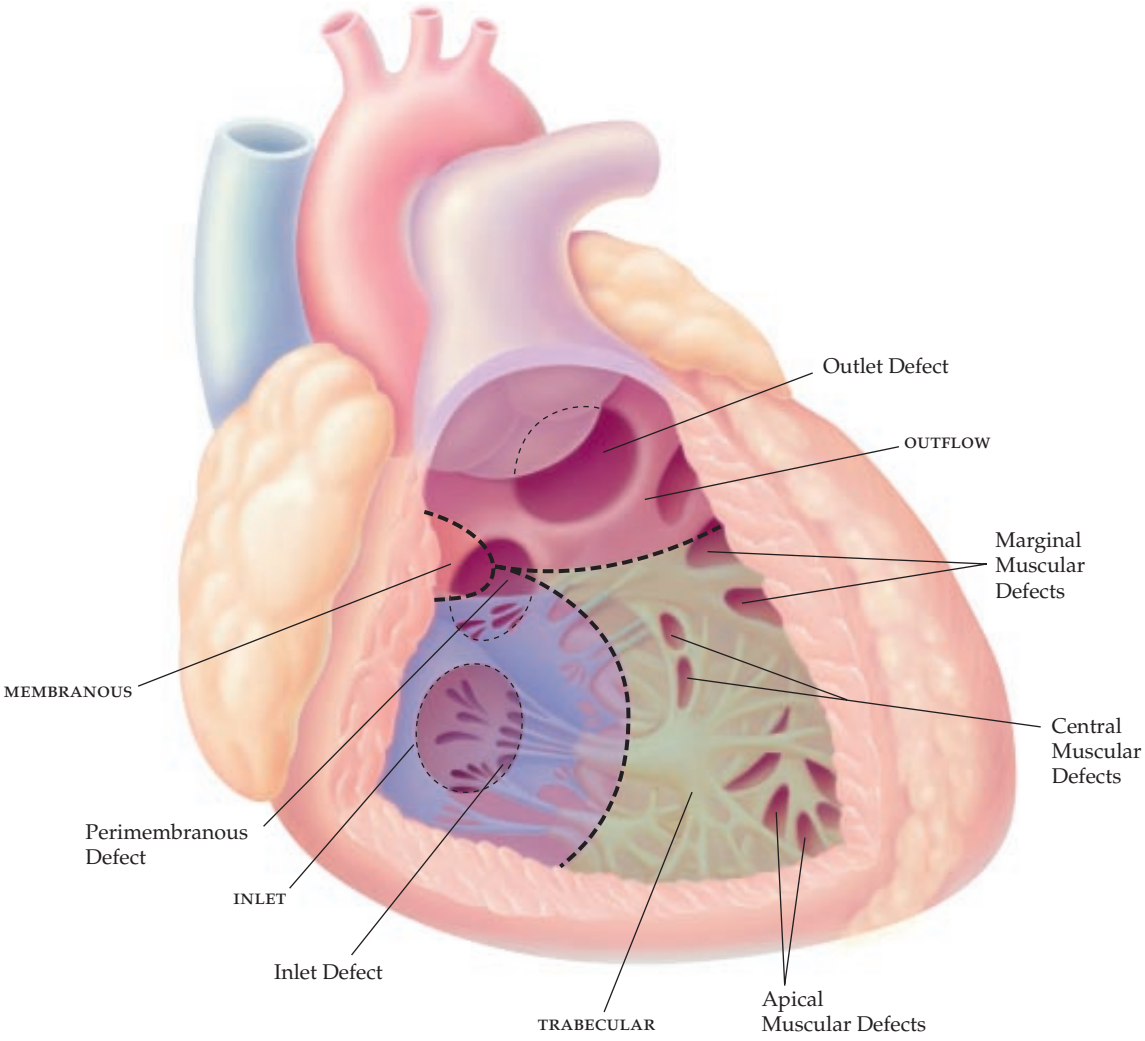


Figure 3 **Anatomic positions of ventricular septal defects. The major anatomic subdivisions of the ventricular septum are the membranous, outflow, inlet, and trabecular portions. Typical VSDs: outlet defect, perimembranous defect, marginal muscular defects, central muscular defects, inlet defect, apical muscular defects.**

cuspid aortic valves. More significant AS or AR will produce findings similar to those in patients with other disorders that cause these lesions.

Laboratory Tests

Electrocardiography The ECG will be normal unless hemodynamically significant stenosis or regurgitation is present, in which case it will show LVH.

Radiologic studies Chest x-ray findings in patients with hemodynamically significant bicuspid aortic valves are similar to those in patients with AS and AR from other etiologies. These may include LV dilatation in AR or in AS progressing to heart failure; the latter will also produce pulmonary congestion.

Echocardiography Both the presence of a bicuspid aortic valve and its hemodynamic significance can be determined by echocardiography. Serial studies are useful in following the progression of the lesion.

Management

All patients with bicuspid aortic valves—even those patients with no significant stenosis or regurgitation—should be given instructions regarding endocarditis prophylaxis. Patients with AR from a bicuspid valve who are asymptomatic and have normal systolic function are followed with echocardiograms and history and physical examinations at regular intervals. If they begin to show evidence of decreasing systolic function, symptoms of heart failure, or progressive dilation of the LV, surgical replacement of the aortic valve is indicated.

Surgical or balloon valvuloplasty (in younger patients) should be considered in a patient with AS who has heart failure, syncope, or chest discomfort. A variety of surgical procedures are available, including direct repair of the valve; replacement with a bioprosthesis or mechanical prosthesis; replacement of the valve and proximal aortic root with a cadaver homograft; and the Ross procedure, in which the abnormal bicuspid valve is removed surgically and replaced with the patient's native pulmonic valve, which in turn is replaced with a cadaver homograft. The Ross procedure eliminates the need for a prosthetic

valve in the aortic position. Postoperative management focuses on assessment for recurrent stenosis or progressive AR.

PULMONIC STENOSIS

Patients with pulmonic stenosis (PS) commonly have a malformed valve, with fusion of one or more of the commissures resulting in a dome-shaped valve. Most of these valves are thin and pliable. However, some patients have thickened valves, termed dysplastic.

Pathophysiology

The stenotic pulmonary valve imposes a pressure load on the ventricle, leading to right ventricular hypertrophy (RVH) and, in a subset of patients, RV failure. In patients with severely hypertrophic right ventricles, imbalance of myocardial oxygen supply and demand may lead to ischemia with attendant anginal chest discomfort and arrhythmias.

Clinical Presentation

Patients with PS of even a moderately severe degree may be asymptomatic for decades. Eventual symptoms may include chest discomfort reminiscent of angina pectoris from coronary artery disease, shortness of breath, fatigability, and symptoms of RV failure. Progression of disease and symptoms beyond adolescence is unusual.

The cardinal physical finding of PS is a systolic crescendo-decrescendo murmur of turbulence through the narrowed valve, preceded by a pulmonic ejection click. The behavior of the pulmonic ejection click during respiration may serve to differentiate it from the click of the bicuspid aortic valve. The pulmonic ejection click will exhibit a selective decrease in intensity with normal inspiration and may even disappear entirely with inspiration; in contrast, the bicuspid aortic valve click will exhibit no such selective decrease.

Laboratory Tests

All patients with suspected PS should have an ECG, a chest x-ray, and an echocardiogram.

Electrocardiography The ECG will be normal in patients with mild to moderate PS. Severe stenosis leads to RVH.

Radiologic studies In patients with mild to moderate degrees of PS, the chest x-ray may show no changes except mild poststenotic dilatation of the proximal pulmonary trunk.

Echocardiography Echocardiography is extremely accurate in identifying and diagnosing the severity of PS. It can also differentiate pliable from dysplastic pulmonary valves. Finally, the echocardiogram can assess RV systolic function.

Management

Early approaches to PS consisted of closed or open surgical valvotomy. In the past 20 years, the advent of reliable methods of balloon pulmonary valvuloplasty has brought a major change in the approach to these cases.[9] Particularly in patients with pliable-dome valves, the initial approach is to perform balloon valvuloplasty in cases of significant stenosis (defined by a right ventricular outflow tract gradient greater than 50 mm Hg). Some dysplastic valves are difficult to treat adequately by balloon valvuloplasty and require surgical repair or replacement. Postoperative management includes surveillance for recurrent stenosis or progressive regurgitation.

Acyanotic Disorders—Aortic Defects

COARCTATION OF THE AORTA

Coarctation of the aorta is a relatively common congenital heart defect that can be seen alone, in association with other defects (especially VSD), and in patients with Turner syndrome. A bicuspid aortic valve is a common associated lesion. Coarctation is a common cause of secondary hypertension and should be sought in all patients presenting with hypertension.

Pathophysiology

The essential pathology in coarctation of the aorta is a narrowing of the aortic lumen, usually in the vicinity of the ligamentum arteriosum, just distal to the take-off of the left subclavian artery. The narrowing of the aorta at the site of the coarctation divides the systemic circulation into a high-pressure zone proximal to the coarctation and a low-pressure zone distal to it. Hypertension may accelerate the development of atherosclerotic coronary artery disease and lead to stroke; stroke is a particular risk when aneurysms of the circle of Willis are present, as occurs with increased incidence in patients with coarctation.

Clinical Presentation

Although lower-extremity claudication may occur, even patients with significant coarctation of the aorta may be entirely asymptomatic. The cardinal feature on physical examination is the difference in pulses and blood pressures above versus below the coarctation. Palpation of the radial and femoral arteries in a normal patient will reveal simultaneous arrival or, perhaps, slightly earlier arrival of the pulse at the femoral artery. In coarctation of the aorta, the femoral pulse will occur later than the radial and is often lower in amplitude. Blood pressure should be evaluated in both arms and either leg when seeking coarctation of the aorta, because of variations in anatomy. When the coarctation is distal to the origin of the left subclavian artery, both arms will be in the high-pressure zone and both legs in the low-pressure zone. However, some coarctations are proximal to the left subclavian. Thus, the left arm and both legs will be in the low-pressure zone, and the diagnosis may be missed if only the left arm is used for measuring blood pressure. More rarely, there may be an anomalous origin of the right subclavian artery; the artery may arise directly from the aorta distal to the left subclavian instead of from the brachiocephalic (innominate) artery. In addition to differential blood pressures, physical examination may also reveal a murmur across the coarctation that can be best heard in the left infrascapular area.

Laboratory Tests

Electrocardiography The ECG in patients with coarctation will show varying degrees of LVH, depending on the severity of the narrowing.

Radiologic studies Dilatation of the aorta proximal and distal to the coarctation site may lead to a so-called 3 sign on chest x-ray. Rib notching is often present; this term refers to apparent effacement, or so-called scalloping, of the lower edges of ribs (usually the third through ninth ribs) because of large, high-flow intercostal collateral vessels that develop as a compensatory mechanism to bypass the narrowing at the coarctation site. Absence of rib notching does not rule out coarctation of the aorta, however.

MRI with MRA can be used effectively to identify coarctation and the collateral circulation. It is also useful for postrepair detection of aneurysms or restenosis at the site of repair.

Echocardiography Echocardiography is extremely helpful in identifying the site of the coarctation by direct visualization, as well as in measuring the pressure gradient across the coarctation site. Echocardiography can also identify bicuspid aortic valves, which frequently accompany coarctation of the aorta.

Management

Coarctation that is sufficient to produce hypertension should always be treated, either surgically or by balloon angioplasty with stent placement. The longest experience is with surgical excision of the coarctation and either end-to-end anastomosis or graft interposition. In recent years, balloon angioplasty has increasingly proved to be a viable alternative for both initial treatment of coarctation and for treatment of restenosis at the coarctation site that develops after repair or angioplasty.[10]

Both before and after correction of coarctation, patients are at risk for infective endarteritis in the vicinity of the coarctation or distal to it and should be treated with prophylactic antibiotics before procedures of risk.

Cyanotic Disorders

DEFINITION AND MECHANISMS

Central cyanosis is caused by an intracardiac shunt or an intrapulmonary right-to-left shunt. Cyanosis becomes evident when reduced (unoxygenated) capillary hemoglobin reaches about 5 g/dl, although this depends on the total hemoglobin concentration: cyanosis is more readily apparent in a patient with polycythemia and is less apparent in a patient with anemia. Mild cyanosis is difficult to detect. Generally, cyanosis does not become clinically apparent until the oxygen saturation falls below 85% (assuming a normal hemoglobin level). Patients with long-standing arterial desaturation will develop clubbing of the fingernails and toenails. Clubbing is characterized by thickening and widening of the nailbeds and loss of the angle between the nail and nail bed, producing a convex nail.

It is helpful to categorize cyanotic CHDs in terms of their effect on pulmonary blood flow. Defects producing decreased pulmonary blood flow include tetralogy of Fallot, tricuspid atresia, Ebstein anomaly, and pulmonary atresia. Defects associated with increased pulmonary blood flow include persistent truncus arteriosus, transposition of the great arteries with or without VSD or PDA, total anomalous venous return, a single or common ventricle, and hypoplastic left heart syndrome. Acyanotic patients with large left-to-right shunts may develop pulmonary vascular occlusive disease (Eisenmenger syndrome).

Adult patients with cyanotic CHD are at increased risk for hyperviscosity secondary to erythrocytosis. The erythrocytosis develops as a compensatory mechanism for red cell oxygen desaturation: a significantly increased red cell mass is necessary to deliver an adequate volume of oxygen to peripheral tissues, given the sometimes severe degree of desaturation. Venous and arterial thrombosis with secondary cerebrovascular accidents have been well documented in cyanotic CHD and have been attributed both to the increased red blood cell mass and to associated iron deficiency anemia, which also increases blood viscosity. This risk is increased in the presence of hypertension or atrial fibrillation and in patients with a history of phlebotomy and microcytosis, suggesting the need for a more conservative approach to phlebotomy and aggressive treatment of iron deficiency.

EISENMENGER SYNDROME

A serious complication of long-standing left-to-right shunts in the atria, ventricles, or great arteries is the development of severe, irreversible pulmonary hypertension, which is termed Eisenmenger syndrome.

Pathophysiology

Normally, the pulmonary vascular resistance is substantially lower than systemic vascular resistance; thus, large intracardiac or great artery communications tend to produce left-to-right shunting. As pulmonary vascular resistance rises, resistance to flow into the pulmonary circulation will eventually exceed that into the systemic circulation, and right-to-left shunting will occur. This will result in varying degrees of cyanosis as well as other physical findings of pulmonary hypertension. Unlike patients with polycythemia vera or polycythemia from chronic obstructive pulmonary disease, patients with Eisenmenger syndrome will often require hematocrits in the 60s, or even low 70s, to deliver sufficient oxygen to tissues to avoid ischemic symptoms.

Clinical Presentation

Patients with Eisenmenger syndrome may be asymptomatic except for cyanosis. Eventually, many patients will note decreased exercise tolerance and chest discomfort, often reminiscent of angina pectoris. If secondary erythrocytosis reaches severe levels, patients may develop symptoms of hyperviscosity, including visual disturbances, headaches, and other complaints.

Physical examination of a patient with Eisenmenger syndrome will reveal manifestations of pulmonary hypertension, including a loud pulmonary component of the second heart sound and the high-pitched diastolic murmur of high-pressure pulmonary regurgitation (the Graham Steell murmur). Additional findings include cyanosis, clubbing, and RV lift or heave.

Laboratory Tests

Electrocardiography The ECG shows right axis deviation and RVH, exhibited as tall R waves and ST-T abnormalities in V1 through V3.

Radiologic studies The chest x-ray will show enlarged central pulmonary arteries with peripheral arterial pruning. Cardiomegaly with specific chamber enlargement will reflect the underlying defect. Right-sided cardiac catheterization often is needed to assess pulmonary arterial pressure and resistance.

Echocardiography Echocardiography can identify and quantify the underlying cardiac shunt and provide an estimate of right heart pressures.

Management

Patients with Eisenmenger syndrome may live for decades after the diagnosis is made.[11] Alternatively, sudden death from ventricular arrhythmias may occur. Because pulmonary resistance is high and fixed in these patients, care needs to be taken to avoid situations that may lead to sudden decreases in systemic vascular resistance, which would exacerbate the right-to-left shunting, sometimes in a life-threatening manner. This would include avoidance of overly hot environments and de-

hydration; in addition, care should be taken during anesthesia or when using vasodilator drugs. Pregnancy is another state in which systemic vascular resistance falls; thus, pregnancy is extremely dangerous for a mother with pulmonary hypertension, as well as for her fetus. Iron deficiency should be treated if present. Only rarely will phlebotomy be required to relieve symptoms of hyperviscosity. A relatively recent therapeutic option for Eisenmenger syndrome is the use of prostacyclin[12] or endothelin-related drugs[13] to lower pulmonary vascular resistance. Heart-lung or lung transplantation has been successfully performed in some patients with Eisenmenger syndrome.[14]

TETRALOGY OF FALLOT

Pathophysiology

Tetralogy of Fallot is the most common form of cyanotic congenital heart disease. Classically, the syndrome includes pulmonary stenosis (subvalvar, valvar, supravalvar, or a combination of all of these), RVH, subaortic VSD, and dextropositioning of the aorta so that it overrides the interventricular septum. Associated anomalies include right aortic arch (25%), atrial septal defect (10%), and coronary artery anomalies (10%).[15] Approximately 15% of patients with tetralogy of Fallot have a deletion of chromosome 22q11 (CATCH 22 syndrome: cardiac anomalies, abnormal facies, thymic hypoplasia, cleft palate, hypocalcemia, and 22q11 deletion).[16]

Surgical Repair in Childhood

Current surgical practice warrants early repair, usually in the first year of life. Without surgery, survival beyond 20 years of age is uncommon.

Surgical repair consists of patch closure of the VSD and alleviation of the RV outflow tract obstruction by one or more of the following methods: infundibular muscle resection, pulmonary valvotomy, outflow tract or transannular patch augmentation, and patch augmentation of the main or proximal branch pulmonary arteries. In some cases, it is necessary to place a conduit from the RV to the pulmonary artery. The conduit may be valved or nonvalved, and it may be bioprosthetic or a homograft.

When pulmonary blood flow is inadequate, surgical repair includes a shunt from the systemic circulation to the pulmonary artery to provide additional pulmonary flow. This may consist of a Blalock-Taussig shunt, a Potts shunt, or a Waterston shunt. The classic Blalock-Taussig shunt connects the subclavian artery to the pulmonary artery; the modified form comprises an interposed tube graft, usually of expanded polytetrafluoroethylene [Gore-Tex]. A Potts shunt connects the descending aorta to the left pulmonary artery. A Waterston shunt connects the ascending aorta to the right pulmonary artery [*see Figure 4*].

Clinical Presentation after Repair

In patients who have undergone surgical repair of tetralogy of Fallot, the examination focuses on residual defects. Not uncommonly, these patients have murmurs related to residual outflow tract obstruction and mild to severe pulmonary regurgitation (PR), which produces a to-and-fro murmur. The severity of RV outflow tract obstruction directly determines the presence and degree of cyanosis. Systolic ejection murmurs are inversely related to the severity of the obstruction: a short, soft murmur suggests severe obstruction with a large right-to-left ventricular level shunt and minimal forward flow in the pulmonary artery, whereas a long, harsh murmur suggests minimal obstruction.

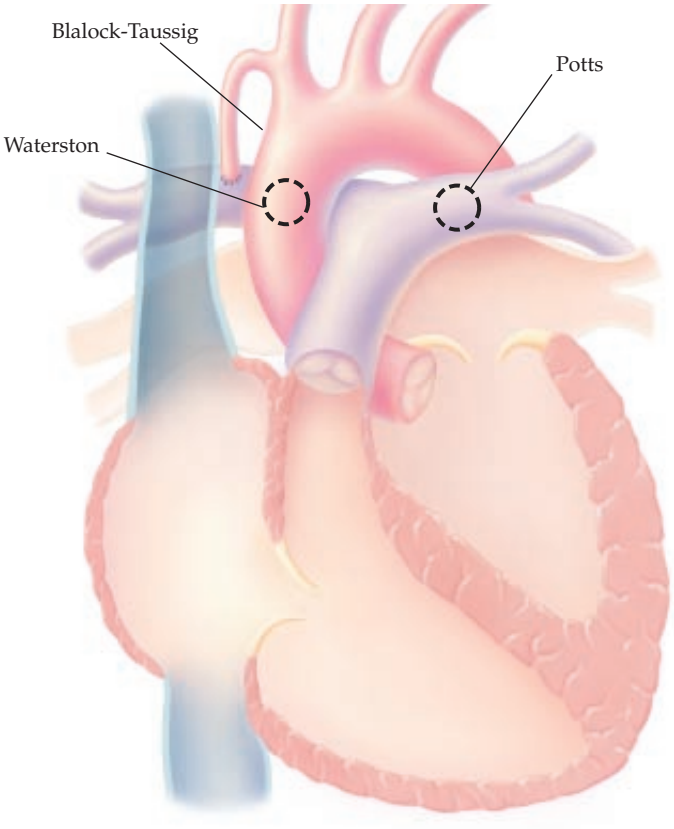


Figure 4 **Systemic artery–to–pulmonary artery shunts. The Blalock-Taussig shunt connects the subclavian artery to a pulmonary artery; the Waterston shunt connects the ascending aorta to the right pulmonary artery; and the Potts shunt connects the descending aorta to the left pulmonary artery.**

Patent shunts will produce a continuous murmur. The degree of cyanosis will depend on the adequacy of pulmonary blood flow provided by the shunt.

A residual VSD may be detected. With increasing RV volume overload, the patient may experience exercise intolerance, right heart failure, and arrhythmias.

Laboratory Data

Electrocardiography In patients who have undergone operative repair of tetralogy of Fallot, the ECG typically shows sinus rhythm, right axis deviation, and RVH; most of these patients also have right bundle branch block. Atrial and ventricular arrhythmias may be detected, especially on a 24-hour monitoring study.

Radiologic studies The findings on chest x-ray vary with the surgical history. A right aortic arch may be noted. The pulmonary artery segment is concave because of the variable degree of pulmonary artery hypoplasia, and the RVH results in an upturned apex; together, these produce the classic finding of a boot-shaped heart. Surgical intervention may result in significant pulmonary regurgitation that eventually will lead to volume overload of the heart, producing cardiomegaly. Over time, patch augmentation of the outflow tract may become aneurysmal, which may be indicated by an enlarged pulmonary artery

segment. Asymmetrical pulmonary blood flow suggests significant branch pulmonary artery obstruction and can be best quantitated by a pulmonary flow study. MRI with MRA is very useful to identify residual defects and assess ventricular function, especially in patients with poor acoustic windows and inadequate echocardiographic studies.

Echocardiography Echocardiography will establish the presence and severity of any residual defects, including progressive enlargement of the RV secondary to pulmonic regurgitation, a residual VSD, and continuous flow in a palliative shunt. Doppler studies will demonstrate the magnitude of the residual outflow tract gradient. The ascending aorta often is enlarged.

Management

Patients who have undergone repair of tetralogy of Fallot must be regularly monitored for progression of residual defects, particularly those with pulmonary regurgitation and conduit obstruction. Branch pulmonary artery stenosis may be approached with balloon angioplasty and stent placement. Repeat surgery should be considered in patients with a significant residual VSD; in patients whose RV pressure is greater than two thirds the systemic pressure because of residual obstruction; in patients with RV enlargement secondary to severe pulmonary regurgitation (which may mandate placement of a bioprosthetic valve, especially if there is associated tricuspid regurgitation [TR]); and in those with reduced exercise tolerance.[17] Reoperation in adults can be performed with low risk. Aortic valve or aortic root replacement is occasionally required because of progressive root dilatation and AR.[18] Ventricular arrhythmias, which are detected in 40% to 50% of patients, have been associated with older age at primary repair, RV volume overload, and QRS prolongation. Marked widening of the QRS to more than 180 msec and LV dysfunction have been identified as risk factors for sudden cardiac death. In such cases, consideration should be given to prophylactic placement of an implantable cardiac defibrillator.[19] Patients should be counseled to follow endocarditis prophylaxis during procedures that place them at risk.

DEXTROTRANSPOSITION OF THE GREAT ARTERIES

Pathophysiology

In the most common form of transposition of the great arteries (TGA), dextro-TGA (D-TGA), the aorta arises in an anterior position from the RV, and the pulmonary artery arises posteriorly from the LV. There is complete separation of the pulmonary and systemic circulations: systemic blood flow traverses the right heart and enters the aorta, whereas pulmonary blood flow traverses the left heart and enters the pulmonary artery. Most surviving patients have a patent ductus arteriosus and foramen ovale, permitting mixing of the two circulations. About one third have associated anomalies, including ASD and VSD. Left ventricular outflow tract obstruction is not uncommon. Unless intracardiac mixing is improved, survival beyond the first year is unusual.

Surgical Repair in Childhood

Initial treatment of D-TGA includes infusion of prostaglandin E to maintain patency of the ductus arteriosus and balloon septostomy (Rashkind procedure) to permit better mixing at the atrial level. Surgery initially consisted of redirecting the systemic venous return to the LV and the pulmonary venous return to the RV. These so-called atrial switch operations (Mustard or Senning procedures), which used a baffle within the atria, restored physiologic circulation but required the RV to function as the systemic ventricle. The arterial switch operation has replaced the atrial switch operation, at least in patients who have normal function of both semilunar valves. In the arterial switch operation, the pulmonary artery and aorta are first transected above the semilunar valves and coronary arteries, and then they are switched. The aorta is connected to the neoaortic valve (formerly the pulmonic valve) arising from the LV, and the pulmonary artery is connected to the neopulmonary valve (formerly the aortic valve) arising from the RV. The coronary arteries are relocated to the neoaorta.

Patients with D-TGA and a large VSD may undergo the Rastelli procedure. The pulmonary artery is divided and oversewn. Flow from the LV must pass through the septal defect and is directed by a baffle to the aortic valve. A conduit from the RV to the pulmonary artery allows egress from the ventricle to the pulmonary circulation.

Clinical Presentation after Repair

Physical findings relate to the presence of associated anomalies (i.e., murmurs of VSD, PS, or PDA). Similarly, the larger the septal defect, the less severe the cyanosis.

Laboratory Tests

Electrocardiography The ECG in patients with the atrial switch shows right axis deviation and RVH. In patients with the arterial switch, the ECG may be normal, provided coronary blood flow is not compromised.

Radiologic studies Patients who have had the atrial switch procedure generally have cardiomegaly from a dilated RV, and the pulmonary artery may show preferential flow to the right lung. Patients with the arterial switch repair are likely to have normal heart size.

Echocardiography Echocardiography is used to assess associated residual defects: depressed RV function, progressive TR, left ventricular outflow tract obstruction, residual VSD, or coronary artery perfusion abnormalities.

Management

The long-term outlook after the atrial switch is quite good, with actuarial survival of 80% at 28 years and 76% of survivors having no symptoms.[20] However, these patients must be monitored for progressive RV enlargement and TR leading to ventricular dysfunction. Although this complication occurs in only 3% of cases, such patients may require cardiac transplantation if medical therapy is ineffective. Atrial arrhythmias, including sick sinus syndrome, are common. The atrial baffle may cause either systemic or pulmonary venous obstruction, which is addressed either by reoperation or by balloon angioplasty and stent placement.

The long-term prognosis of patients with the arterial switch is less well known, but arrhythmias are thought to be less frequent and to occur secondary to imperfections in the operative procedure.[21] Patients should undergo nuclear medicine studies or stress testing to monitor for inadequate coronary perfusion secondary to coronary artery reimplantation abnormalities. Stenosis of the pulmonary artery (the most common complication) or stenosis at aortic anastomosis sites may occur. Complications of the Rastelli procedure include subaortic obstruction

(baffle or VSD obstruction), conduit stenosis (with or without regurgitation), baffle leak, and branch pulmonary artery stenosis. Significant residual defects require reoperation.[17]

THE UNIVENTRICULAR HEART

Pathophysiology

A functional single ventricle may result from hypoplastic left heart syndrome (aortic atresia, mitral atresia, or both), tricuspid atresia, pulmonary atresia with intact ventricular septum, or an unbalanced AVSD resulting in hypoplasia of either the RVor LV.

Surgical Repair in Childhood

The initial presentation of univentricular heart in childhood may include severe cyanosis associated with a marked decrease in pulmonary blood flow, mild cyanosis and heart failure associated with intracardiac admixture of circulations and excessive pulmonary blood flow, or nearly balanced systemic and pulmonary blood flows and mild cyanosis. Patients who survive to adulthood generally have undergone one or more palliative surgical procedures; these include the Norwood, Glenn, and Fontan procedures.

Norwood The Norwood operation establishes a single outlet from the single ventricle by anastomosing the hypoplastic ascending aorta to the main pulmonary artery, producing a so-called neoaorta and connecting the distal pulmonary artery to a systemic shunt, usually a modified Blalock-Taussig shunt [*see Figure 5a*]. Often, an atrial septectomy is required to allow complete mixing at the atrial level.

Glenn The bidirectional Glenn procedure involves anastomosis of the SVC to the pulmonary artery. It includes takedown of a previously placed shunt and repair of any branch pulmonary artery stenosis [*see Figure 5b*]. The term bidirectional refers to the fact that the right pulmonary artery remains in con-

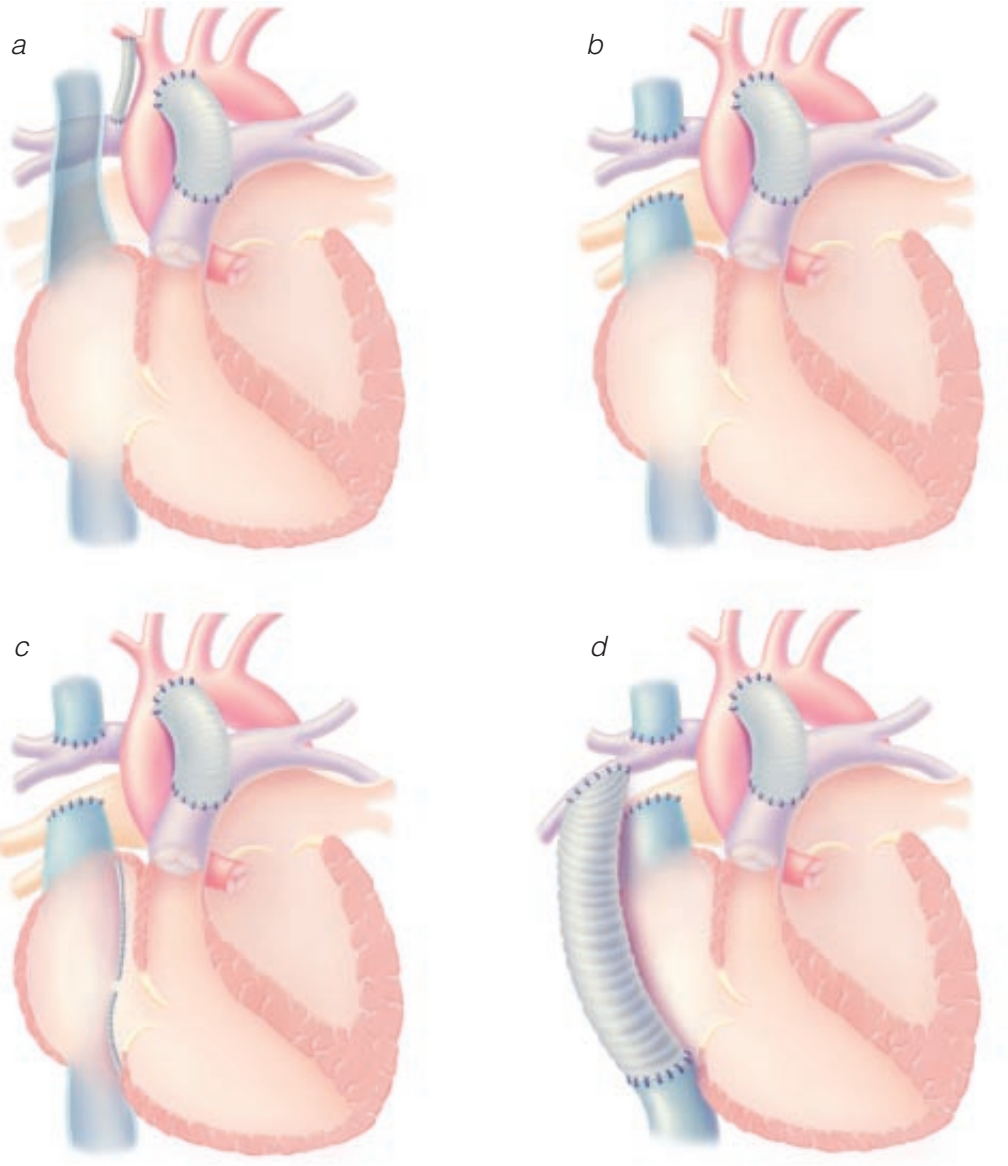


Figure 5 **Stages in the repair of functional single ventricles (see text for details). (*a*) Norwood; (*b*) bidirectional Glenn; (*c*) lateral tunnel; (*d*) extracardiac conduit.**

tinuity with the left pulmonary artery; this contrasts with the classic Glenn procedure, which involves anastomosis of the SVC to a right pulmonary artery that has been disconnected from the main and left pulmonary arteries. The bidirectional Glenn procedure is now done at 4 to 6 months of age.

Fontan The Fontan procedure is the final palliative procedure, providing direct connection of flow from the SVC and inferior vena cava (IVC) to the pulmonary circuit. Initially, this was a one-stage procedure that involved attaching the RA to the pulmonary artery or RV outflow tract and was performed in patients older than 4 years. Current practice is to stage the anastomosis of SVC and IVC to the pulmonary circuit, with the final stage, total cavopulmonary artery anastomosis, occurring at 2 to 3 years of age. The IVC is connected to the pulmonary artery either by a lateral tunnel placed in the RA to direct blood from the IVC to the proximal SVC stump, which is then attached to the pulmonary artery [*see Figure 5c*], or by an extracardiac conduit connecting the IVC to the pulmonary artery directly [*see Figure 5d*]. With any of these routes of flow, a small communication (fenestration) may be made between the caval blood flow conduit and the functional left atrium. Pulmonary blood flow is achieved by passive venous return without assistance of a ventricular pumping chamber. Any mild alteration of pulmonary pressure or resistance will impair adequacy of pulmonary blood flow.

Clinical Presentation after Repair

Clinical features are variable. Some patients may be well palliated, with near-normal oxygen saturation, acceptable activity levels, and negligible findings on cardiac examination. Others will demonstrate progressive heart failure as the single ventricle (especially if it is an anatomic RV) succumbs to the increased pressure and volume overload secondary to progressive atrioventricular valve regurgitation and myocardial dysfunction. Both atrial and ventricular arrhythmias are common. The sluggish pulmonary blood flow may predispose to in situ thrombosis and pulmonary embolism, which in turn will impede pulmonary blood flow by raising pulmonary arterial pressure.

Laboratory Tests

Electrocardiography ECG findings are quite variable and may include atrial or ventricular enlargement, axis deviation, conduction abnormalities, and arrhythmias.

Radiologic studies The chest x-ray may show progressive cardiomegaly. Pulmonary vascular markings may be unequal, indicating stenosis of one or more pulmonary artery branches. MRI with MRA may show areas of branch pulmonary artery stenosis and progressive changes in chamber size and ventricular function.

Echocardiography Echocardiographic studies are aimed at following the progression of atrioventricular valve regurgitation, ventricular enlargement, and dysfunction, as well as detecting so-called smoke or clots in the systemic venous–to–pulmonary artery circuit.

Management

After surgical correction, patients demonstrate significant limitations in exercise tolerance because they rely on passive pulmonary blood flow that does not increase maximally with exertion. Postoperative arrhythmias are common. Arrhythmias may need to be managed medically, because radiofrequency ablation techniques may be limited by access problems secondary to the extracardiac or lateral tunnel connections be-

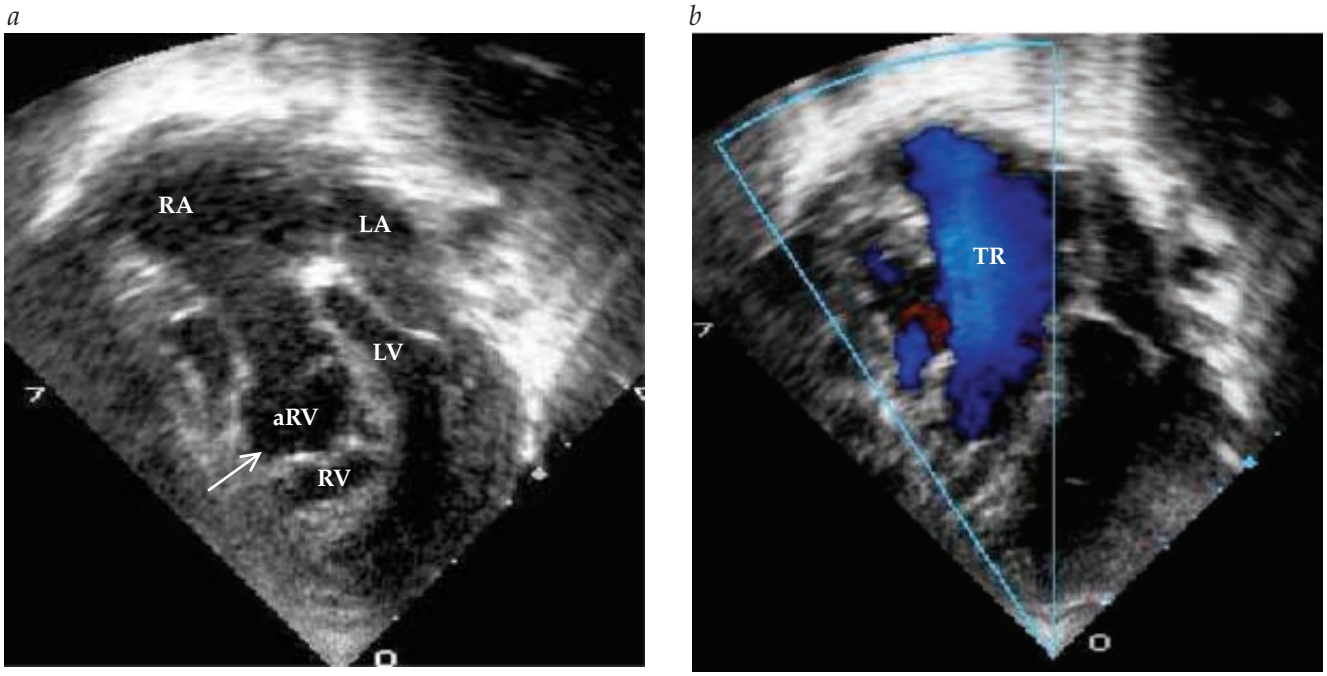


Figure 6 **Echocardiograms of a patient with Ebstein anomaly. (*a*) Apical four-chamber view. The arrow indicates apical displacement of tricuspid leaflet. (*b*) Color Doppler flow image demonstrating severe tricuspid regurgitation (TR), originating deep in the right ventricle from the displaced tricuspid valve leaflet. (ARV—atrialized right ventricle; LA—left atrium; LV—left ventricle; RA—right atrium; RV—right ventricle)**

tween the venous circulation and the pulmonary artery. The need for reoperation after the Fontan procedure is infrequent, with the most common indication being placement of a mechanical pacemaker. Protein-losing enteropathy (PLE) is a serious problem after the Fontan operation. Its cause is not known but probably relates to increased systemic venous and thoracic duct pressures. There may also be a local autoimmune or allergic component in the intestinal wall. PLE is characterized by peripheral edema, malabsorption, and a low serum protein level. Complications have become less frequent with staged surgery and provision of an atrial fenestration. Some older patients may benefit from conversion of classic Fontan to a total cavopulmonary artery anastomosis. Cardiac transplantation may be necessary for systemic ventricular failure or intractable PLE.[17]

EBSTEIN ANOMALY OF THE TRICUSPID VALVE

Pathophysiology

This uncommon anomaly of the tricuspid valve consists of adherence of the posterior and septal leaflets to the myocardium—causing a downward displacement of the functional annulus toward the RV apex—and enlargement of the anterior leaflet. The end result is an atrialization of the RV with resultant TR. In patients who present early in life, Ebstein anomaly is often found in association with other defects, including ASD and PS. Accessory pathways and clinical evidence of preexcitation are not uncommon, and arrhythmias are the most common presenting features in adults. There is an association with maternal lithium administration.

Clinical Presentation

Ebstein anomaly can become clinically evident at any age; the natural history of this lesion ranges from death in early life to adult survival without surgery, depending on the degree of regurgitation and whether significant arrhythmias are present. Cyanosis may occur, in neonates or adults, secondary to right-to-left shunting at the atrial level. Adult patients may complain of fatigue, shortness of breath, palpitations, or syncope. On auscultation, a murmur of TR is apparent and is often associated with a gallop rhythm, multiple systolic ejection sounds, and a widely split second sound.

Laboratory Tests

Electrocardiography ECG findings are quite variable. The PR interval may be normal; short, with preexcitation; or prolonged. The axis may be superior or rightward, with or without a right bundle branch block. There may be evidence of RA enlargement. Arrhythmias are detected in 43% of adolescents and adults.[22]

Radiologic studies The chest x-ray may show cardiomegaly with RA enlargement. Cardiac catheterization is not necessary unless there is concern regarding coronary artery disease or need for electrophysiologic assessment and possible radiofrequency ablation.

Echocardiography Echocardiography can confirm the diagnosis and the degree of the tricuspid valve displacement (which may vary from mild tethering of the septal leaflet to severe apical displacement) and characterize the severity of TR [*see Figure 6*]. The anterior leaflet is large and sail-like and may produce RV outflow obstruction. The atrial septum should be assessed for size of defect and magnitude of shunting.

Management

Surgery is recommended for patients with symptomatic heart failure and cardiomegaly, cyanosis, or arrhythmias; tricuspid valvuloplasty is preferred over valve replacement.[15] Surgery is not recommended for asymptomatic patients,[22] although some authors have advocated surgery if significant cardiomegaly is present, because this may be a better predictor of sudden death than functional status.[23]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Perloff JK, Miner PD: Specialized facilities for the comprehensive care of adults with congenital heart disease. Congenital Heart Disease in Adults, 2nd ed. Perloff JK, Child JS, Eds. Philadelphia, WB Saunders Co, 1998, p 9

2. Driscoll D, Allen HD, Atkins DL, et al: Guidelines for evaluation and management of common congenital cardiac problems in infants, children, and adolescents: a statement for healthcare professionals from the Committee on Congenital Cardiac Defects of the Council on Cardiovascular Disease in the Young, American Heart Association. Circulation 90:2180, 1994

3. Dajani AS, Taubert KA, Wilson W, et al: Prevention of bacterial endocarditis: recommendations by the American Heart Association. J Am Dent Assoc 128:1142, 1997

4. Driscoll DJ: Left-to-right shunt lesions. Pediatr Clin North Am 46:355, 1999

5. Berger F, Ewert P, Bjornstad PG, et al: Transcatheter closure as standard treatment for most interatrial defects: experience in 200 patients treated with the Amplatzer Septal Occluder. Cardiol Young 9:468, 1999

6. Gatzoulis MA, Freeman MA, Siu SC, et al: Atrial arrhythmia after surgical closure of atrial septal defects in adults. N Engl J Med 340:839, 1999

7. Suzuki K, Yamaki S, Mimori S, et al: Pulmonary vascular disease in Down's syndrome with complete atrioventricular septal defect. Am J Cardiol 86:434, 2000

8. Goyal VS, Fulwani MC, Ramakantan R, et al: Follow-up after coil closure of patent ductus arteriosus. Am J Cardiol 83:463, 1999

9. Stanger P, Cassidy SC, Girod DA, et al: Balloon pulmonary valvuloplasty: results of the Valvuloplasty and Angioplasty of Congenital Anomalies Registry. Am J Cardiol 65:775, 1990

10. Magee AG, Brzezinska-Rajszys G, Qureshi SA, et al: Stent implantation for aortic coarctation and recoarctation. Heart 82:600, 1999

11. Cantor WJ, Harrison DA, Moussadji JS, et al: Determinants of survival and length of survival in adults with Eisenmenger syndrome. Am J Cardiol 84:677, 1999

12. Rosenzweig EB, Kerstein D, Barst RJ: Long-term prostacyclin for pulmonary hypertension with associated congenital heart defects. Circulation 99:1858, 1999

13. Prendergast B, Newby DE, Wilson LE, et al: Early therapeutic experience with the endothelin antagonist BQ-123 in pulmonary hypertension after congenital heart surgery. Heart 82:505, 1999

14. Ueno T, Smith JA, Snell GI, et al: Bilateral sequential single lung transplantation for pulmonary hypertension and Eisenmenger's syndrome. Ann Thorac Surg 69:381, 2000

15. Brickner M, Hillis LD, Lange RA: Congenital heart disease in adults: second of two parts. N Engl J Med 342:334, 2000

16. Goldmuntz E, Clark BJ, Mitchell, et al: Frequency of 22q11 deletion in patients with conotruncal defects. J Am Coll Cardiol 32:492, 1998

17. Connelly MS, Webb GD, Somerville J, et al: Canadian consensus conference on adult congenital heart disease—1996. Can J Cardiol 14:399, 1998

18. Niwa K, Siu SC, Webb GD, et al: Progressive aortic root dilation in adults late after repair of tetralogy of Fallot. Circulation 106:1374, 2002

19. Ghai A, Silversides C, Harris L, et al: Left ventricular dysfunction is a risk factor for sudden cardiac death in adults late after repair of tetralogy of Fallot. J Am Coll Cardiol 40:1675, 2002

20. Wilson NJ, Clarkson PM, Barratt-Boyes BG, et al: Long-term outcome after the Mustard repair for simple transposition of the great arteries. J Am Coll Cardiol 32:758, 1998

21. Hunter PA, Kreb DL, Mantel SF: Twenty-five years' experience with the arterial switch operation. J Thorac Cardiovasc Surg 124:790, 2002

22. Celermajer DS, Bull C, Till JA, et al: Ebstein's anomaly: presentation and outcome from fetus to adult. J Am Coll Cardiol 23:170, 1994

23. Gentles TL, Calder L, Clarkson PM: Predictors of long-term survival with Ebstein's anomaly of the tricuspid valve. Am J Cardiol 53:332, 1992

Acknowledgment

Figures 1 through 5 Alice Y. Chen.

XVI PERIPHERAL ARTERIAL DISEASE

Mark A. Creager, m.d.

Peripheral arterial diseases comprise those disorders that compromise blood flow to the limbs. Causes of limb artery obstruction include atherosclerosis, thrombus, embolism, vasculitis, arterial entrapment, adventitial cysts, fibromuscular dysplasia, arterial dissection, trauma, and vasospasm.

Peripheral Atherosclerosis

The most freqently encountered cause of peripheral arterial disease is atherosclerosis. The pathology of atherosclerosis that affects the limbs is similar to that of atherosclerosis of the aorta, coronary arteries, and extracranial cerebral arteries. Of patients who present with symptoms of peripheral atherosclerosis, approximately 80% have femoropopliteal artery stenoses, 30% have lesions in the aorta or iliac arteries, and 40% have tibioperoneal artery stenoses. Most patients have multiple stenoses.

EPIDEMIOLOGY

The prevalence of peripheral atherosclerosis, both asymptomatic and symptomatic, increases with age, ranging from 3% in persons younger than 60 years to greater than 20% in persons 75 years of age and older.[1-3] An epidemiologic study in Germany found that the overall prevalence of peripheral arterial disease in patients older than 65 years was 20% in men and 17% in women.[4] A community-based survey in primary physicians' offices in the United States found that of patients who were between the ages of 50 and 69 years who had diabetes mellitus or smoked cigarettes or who were older than 70 years, 29% had peripheral arterial disease.[5]

The prevalence of claudication ranges from 1% to 5%.[1] The peak incidence of claudication occurs between the sixth and seventh decades and develops later in women than in men. Each year, approximately 2% to 4% of all patients with intermittent claudication develop critical limb ischemia.[6]

Long-term survival is reduced in patients with peripheral atherosclerosis. The risk of death in populations with peripheral atherosclerosis is increased twofold to fourfold. Most patients die as a consequence of myocardial infarction or stroke.[7] Patients with claudication have a 5-year survival rate of approximately 15% to 30%, and patients with critical limb ischemia have a 1-year survival rate of approximately 25%.[1,2,6-8] Overall, there is an inverse relationship between the severity of peripheral arterial disease and survival.

RISK FACTORS

The risk factors associated with the development of peripheral atherosclerosis are similar to those associated with coronary atherosclerosis. These include cigarette smoking, diabetes mellitus, dyslipidemia, hypertension, a family history of premature atherosclerosis, and hyperhomocysteinemia. The risk of developing intermittent claudication is twofold to fivefold higher in smokers than in nonsmokers.[9,10] Moreover, continued cigarette smoking greatly increases the risk of progression from stable claudication to severe limb ischemia and amputation. Diabetes mellitus is associated with a threefold to fourfold increase in the risk of peripheral arterial disease.[9,11] Peripheral atherosclerosis is often more severe and extensive in diabetic patients than in nondiabetic patients with atherosclerosis; in addition, the tibial and peroneal arteries are involved more frequently in diabetic patients than in nondiabetic patients. Prognosis is poor for patients with diabetes who have claudication: 30% to 40% develop critical limb ischemia over a 6-year period. The risk of amputation in diabetic patients is sevenfold to 15-fold higher than in nondiabetic patients with peripheral arterial disease.[2,12] Dyslipidemia, particularly hypercholesterolemia, is present in 40% of patients with peripheral atherosclerosis. The relative risk of peripheral arterial disease is 1.2 to 1.4 for each 40 to 50 mg/dl increase in total cholesterol.[10] Hypertriglyceridemia and an elevated plasma concentration of lipoprotein(a) each increase the risk of developing peripheral arterial disease. Hypertension increases the risk of claudication by at least twofold in men and by fourfold in women.[10] Hyperhomocysteinemia has emerged as an important risk factor for atherosclerosis and increases the risk of peripheral atherosclerosis by twofold to threefold.[13] Elevations in markers of inflammation, including levels of C-reactive protein and soluble intercellular adhesion molecule–1, are also independent predictors of the development of symptomatic peripheral arterial disease in otherwise healthy men.[14,15]

DIAGNOSIS

Clinical Presentation

The two principal symptoms of peripheral atherosclerosis are intermittent claudication and rest pain. However, many patients with peripheral arterial disease are asymptomatic, have symptoms that do not fit the typical pattern of intermittent claudication, or have symptoms from comorbid conditions (e.g., arthritis) that blur the presentation.[16]

Intermittent claudication is described as discomfort, pain, fatigue, or heaviness that is felt in the affected extremity during walking and resolves within a few minutes of resting. Intermittent claudication occurs when the metabolic demand of an exercising muscle exceeds supply. A hemodynamically significant stenosis prevents blood-flow augmentation during exercise. The increased pressure gradient that develops across the stenosis compromises perfusion pressure to the exercising muscle. As ischemia develops, autoregulatory mechanisms cause local vasodilatation and a further reduction in perfusion pressure, and extravascular forces created by the exercising muscle reduce perfusion pressure even further. The location of the symptom depends on the site of stenosis. Thigh, hip, or buttock claudication may develop in cases of proximal arterial occlusive disease involving the aorta or iliac arteries. Involvement of the femoral and popliteal arteries typically causes calf claudication. Tibial and peroneal artery stenoses may cause pedal claudication.

Rest pain occurs when the blood supply does not adequately meet the basic nutritional requirements of the tissues of the affected extremity. Pain typically occurs in the toes or foot. Initially, the pain is worse at night when the patient is lying in bed with the legs in a neutral position. Sitting up and dangling the leg may alleviate the discomfort, because this maneuver increases perfusion pressure via gravitational forces. Conversely, leg elevation worsens the pain. With persistent severe ischemia, skin breakdown occurs, leading to ulceration, necrosis, and gan-

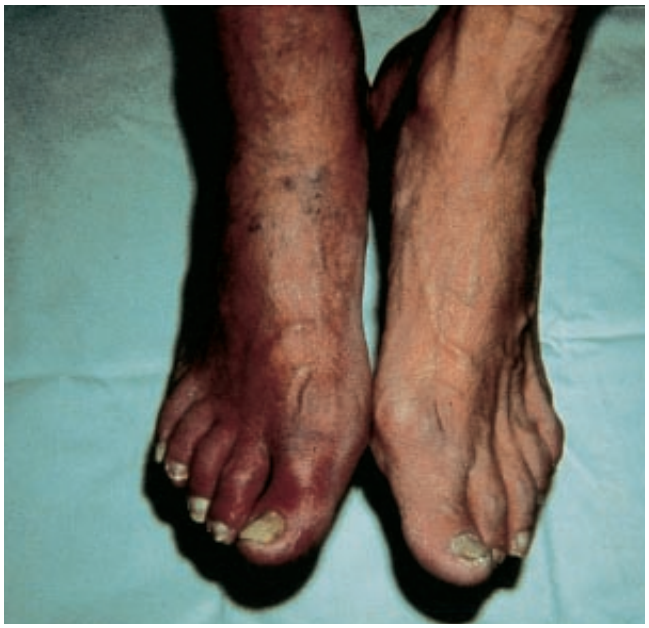


Figure 1 **Photograph shows an ischemic right foot demonstrating dependent rubor.**

grene. Even minor trauma to an ischemic foot may produce a skin lesion that fails to heal.

The most reliable physical finding in patients with peripheral arterial disease is decreased or absent pulses. Examination of femoral, popliteal, posterior tibial, and dorsalis pedis pulses may indicate sites of stenosis. Bruits auscultated in the abdomen, pelvis, and inguinal areas also may indicate the presence of arterial stenosis. Foot pallor may be observed at rest, with leg elevation, or after exercise of the calf muscles. Signs of chronic limb ischemia include subcutaneous atrophy; hair loss; coolness; pallor; and cyanosis, dependent rubor, or both [*see Figure 1*]. Additional signs of critical limb ischemia include petechiae, fissures, ulceration, and gangrene. Ulcers often involve the tips of the toes or the heel of the foot and occur at sites of trauma or pressure caused by poor-fitting footwear. Arterial ulcers have pale bases and irregular borders and are usually quite painful.

Several classifications have been proposed to characterize the severity of limb ischemia in patients with peripheral arterial disease. The most widely recognized classification was developed by René Fontaine [*see Table 1*]. A contemporary classification scheme takes into consideration symptoms, physical findings, perfusion pressure, and exercise capacity; the classification scheme comprises four grades and seven categories [*see Table 2*].[17]

Table 1 Fontaine Classification of Chronic Limb Ischemia

Stage	*Symptoms*
I	Asymptomatic
II	Intermittent claudication
IIa	Pain-free; claudication walking > 200 m
IIb	Pain-free; claudication walking < 200 m
III	Rest pain and nocturnal pain
IV	Necrosis, gangrene

Table 2 Clinical Categories of Chronic Limb Ischemia[75]

Grade	*Category*	*Clinical Description*
I	0	Asymptomatic, not hemodynamically significant
	1	Mild claudication
	2	Moderate claudication
	3	Severe claudication
II	4	Ischemic rest pain
	5	Minor tissue loss; nonhealing ulcer; focal gangrene with diffuse pedal ischemia
III	6	Major tissue loss extending above transmetatarsal level; foot no longer salvageable

Noninvasive Diagnostic Tests

Several noninvasive diagnostic tests can be used to evaluate patients with peripheral arterial disease. Segmental blood pressure measurement of the extremity is a quantitative means to assess the presence and severity of arterial stenoses [*see Table 3*]. Pneumatic cuffs are positioned along the leg and are inflated to suprasystolic pressures. During cuff deflation, the onset of flow (i.e., systolic blood pressure) is assessed by use of a Doppler probe placed over the dorsalis pedis or the posterior tibial arteries. Normally, the systolic blood pressure in the leg is the same as that in the arm. However, because of reflected waves, systolic blood pressure in the leg may be slightly higher than that in the arm. The normal ankle:brachial systolic blood pressure ratio (i.e., the ankle:brachial index) is therefore 1.0 or slightly greater. Taking into consideration the variability in blood pressure measurements, an ankle:brachial index less than 0.95 is considered abnormal. Patients with leg claudication typically have an ankle:brachial index less than 0.8; in patients with ischemia at rest, the ankle:brachial index is frequently less than 0.4 [*see Table 3*].

Measurement of the ankle:brachial index can be performed in a medical office. It is a sensitive indicator of peripheral arterial disease; it is more closely associated with leg function in patients with peripheral arterial disease than is intermittent claudication or other leg symptoms.[18] Because atherosclerosis is a systemic problem, a decreased ankle:brachial index suggests that the burden of disease is increased throughout the body, including the coronary arteries; for that reason, the lower the ankle:brachial index, the higher the risk of a cardiovascular event.[19,20]

In patients with peripheral arterial disease, the pressure gradient across a stenosis increases during exercise, as vascular resistance in the exercising muscle decreases. The exercise-induced increase in systemic pressure (i.e., brachial artery pressure) is not accompanied by a comparable increase in ankle pressure. Thus, the ankle:brachial index will be lower immediately after the patient has exercised than when the patient is at rest.

Plethysmographic devices are used to record the change in limb artery volume that occurs with each pulse (pulse volume recordings). The pulse volume waveform comprises a systolic upstroke with a sharp peak, a dicrotic wave, and a downsloping component. Distal to the site of an arterial stenosis, the amplitude of the pulse volume waveform is diminished, and the dicrotic wave disappears. In the presence of severe ischemia, the waveform may be entirely absent [*see Figure 2*].

Doppler ultrasonography can identify vessels with stenotic lesions. A Doppler probe is positioned at various sites along the limb's arteries. The Doppler waveform has three components,

which correspond to three phases of blood flow: high-velocity antegrade flow during systole, transient flow reversal during early diastole, and low-velocity antegrade flow during late diastole. When stenosis is present, this triphasic waveform is altered distal to the stenosis: the amplitude is decreased, the rate of rise is delayed, and the reverse-flow component disappears. Duplex ultrasound scanning is a direct, noninvasive test that combines B-mode ultrasonography and pulsed Doppler ultrasonography to assess peripheral arterial stenoses. A B-mode scan identifies areas of intimal thickening, plaque formation, and calcification. Color Doppler imaging detects blood-flow abnormalities caused by arterial stenoses. An increase of greater than twofold in the systolic velocity is indicative of a hemodynamically significant stenosis, usually one that exceeds 50% of the artery diameter.

Transcutaneous oximetry, which measures the transcutaneous oxygen tension with oxygen-sensing electrodes placed at various sites on the legs, is used to assess the severity of skin ischemia in patients with peripheral arterial disease. Normally, the transcutaneous oxygen tension of the resting foot is approximately 60 mm Hg; it is often less than 40 mm Hg in patients with ischemia.

Angiography

Magnetic resonance angiography (MRA) and CT angiography can be used to evaluate the location and severity of peripheral atherosclerosis[21]; thus, these modalities can help determine whether a patient is a candidate for an endovascular intervention. In addition, they are free of the risks of conventional angiography. MRA is more widely used, although it is somewhat slower than CT angiography. The current generation of CT angiography machines can provide highly detailed images within minutes; however, CT angiography requires iodinated contrast and radiation exposure, which makes it less suitable for patients with renal insufficiency or contrast allergy.

In most patients, clinical evaluation and noninvasive testing are sufficient for confirming the diagnosis of peripheral arterial disease. Conventional catheter-based angiography is typically performed only when a diagnosis is in doubt or as a prelude to endovascular interventions or surgical reconstruction [*see Figure 3*]. Digital subtraction angiography is a computer-enhancing technique that is used to improve resolution; it is particularly useful in conjunction with the intra-arterial administration of radiographic contrast agent.

Table 3 Leg Segmental Pressure Measurements in Patient with Right Calf Claudication and Right Foot Pain*

	Right	*Left*
Brachial	158	158
Upper thigh	160	162
Lower thigh	94	154
Calf	62	116
Ankle	42	116
Ankle:brachial ratio	0.27	0.68

*Findings are consistent with femoropopliteal and tibioperoneal artery stenoses in the right leg. The right ankle:brachial ratio indicates ischemia. Systolic pressure gradients between the lower thigh and calf and between the calf and ankle in the left leg are consistent with distal femoropopliteal artery and tibioperoneal artery stenoses. The left ankle:brachial ratio is consistent with symptoms of claudication.

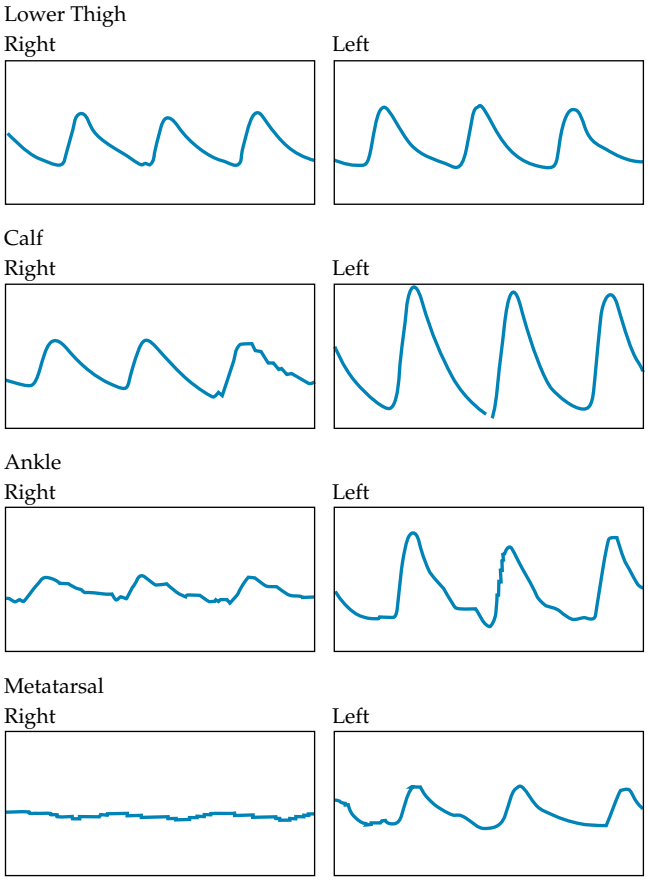


Figure 2 **Pulse volume recordings provide a qualitative assessment of blood flow to the extremity. In this example from a patient with right calf claudication and right foot pain, the pulse volume recordings are abnormal in the right calf, right ankle, and right metatarsal segments. In the right calf and ankle, the amplitude of the pulse is diminished and the rate of rise is delayed. No pulse volume can be recorded in the right metatarsal segment. The pulse volume recordings in the left leg are normal.**

TREATMENT

Risk Factor Modification and Antiplatelet Therapy

Risk factors for atherosclerosis should be identified and treated; this reduces the likelihood of progression of atherosclerosis and also helps to prevent adverse cardiovascular events in patients with peripheral arterial disease.[22,23] Patients who stop smoking cigarettes have a more favorable prognosis than those who continue to smoke. Aggressive lipid-lowering therapy reduces progression of peripheral atherosclerosis, but it has not been established that it prevents progression of symptoms from claudication to critical limb ischemia. Cholesterol-lowering therapy with statin drugs reduces adverse cardiovascular events in patients with atherosclerosis and may improve walking ability.[24-27]

Antihypertensive agents should be tailored to bring blood pressure into a normotensive range to reduce the risk of adverse events such as stroke, congestive heart failure, and renal insufficiency.[2,28] Occasionally, marked reduction of blood pressure may reduce perfusion pressure to an ischemic extremity and potentially aggravate symptoms. Angiotensin-converting enzyme inhibitors are effective antihypertensive drugs that may also reduce the risk of adverse cardiovascular events in patients with atherosclerosis, including those with peripheral arterial disease.[29]

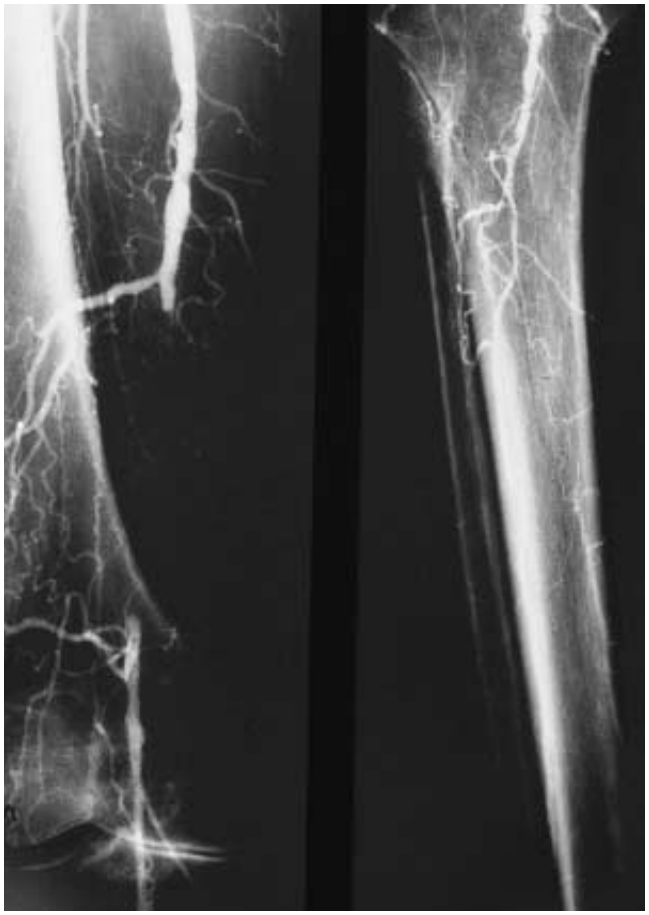


Figure 3 **Arteriogram of a patient with critical ischemia of the right foot. The left panel shows a long, total occlusion of the right superficial femoral artery. The popliteal artery reconstitutes via collaterals. The right panel reveals evidence of anterior tibial, posterior tibial, and peroneal artery occlusions with poor runoff.**

Beta blockers do not worsen intermittent claudication but may cause reflex peripheral cutaneous vasoconstriction and exacerbate critical limb ischemia.[30] Beta blockers are indicated to reduce the risk of myocardial infarction and death in patients with coronary artery disease—a condition that frequently coexists with peripheral arterial disease.

Aggressive treatment of diabetes mellitus reduces microangiopathic complications such as retinopathy and nephropathy.[2,28,31] It is not known whether aggressive treatment of diabetes reduces progression of atherosclerosis or prevents critical limb ischemia or foot ulceration. B-complex vitamins, such as folic acid, cobalamin, and pyridoxine, may lower homocysteine levels, but it is not yet known whether such therapy reduces cardiovascular events or prevents progression of peripheral atherosclerosis.

There is little information regarding the efficacy of platelet inhibition in treating symptoms of peripheral arterial disease. In one study, primary prevention with aspirin was shown to reduce the need for surgical revascularization in patients with peripheral arterial disease.[32] Small angiography trials have suggested that platelet inhibitors reduce the risk of acute peripheral arterial occlusion.[33] These agents may prevent thrombosis after plaque rupture in the peripheral arteries, as they do in coronary arteries. Antiplatelet therapy has been shown to reduce the risk of adverse cardiac events such as nonfatal myocardial infarction and stroke and has been shown to reduce cardiovascular mortality in patients with atherosclerosis.[2,23,34] In one study, clopidogrel was more effective than aspirin in reducing the risk of adverse cardiovascular events, particularly in patients with peripheral arterial disease.[35]

Hygiene and Physical Therapy

Local measures are used to prevent skin ulceration and foot infection, particularly in patients with critical limb ischemia. The feet should be kept clean, and moisturizing cream should be applied to prevent drying and fissuring. The skin of the feet should be inspected frequently, and minor abrasions should be treated promptly. Stockings should be made of natural, absorbent fibers. Elastic hose are contraindicated because they restrict skin blood flow. Shoes should be carefully fitted to reduce the possibility of pressure-induced skin breakdown. In patients with critical limb ischemia, the limbs should be maintained in a dependent position to increase perfusion pressure. This can be achieved by angling the mattress so that the affected limb is below heart level. Cotton wicks placed between the toes absorb moisture and reduce friction. Sheepskin placed beneath the heels of the feet reduces pressure and necrosis. A warm environment is recommended to reduce vasoconstriction. Ulcerations and necrotic areas should be kept dry and covered with dry, nonadhesive material. Infections should be drained. Local antibiotics should be avoided. Pain should be treated with analgesics.

Supervised exercise training programs improve walking capacity in patients with peripheral arterial disease.[2,36] Among the most likely factors that account for the improvement are more efficient skeletal muscle metabolic function and changes in ergonomics.[37] Most studies have not found that exercise training improves blood flow to the exercising extremity, but investigations into the potential angiogenic effects of exercise are ongoing. Training programs should be individualized for each patient. Because supervised settings provide structure and guidance, patients have achieved the most success with supervised training. Programs typically involve treadmill exercise for approximately 1 hour three times a week for at least 3 months. Patients are encouraged to walk independently outside the supervised program.

Pharmacotherapy of Claudication and Critical Limb Ischemia

Drug therapy has generally not been successful in improving symptoms of claudication or reducing the complications of critical limb ischemia.[22] Although arterioles dilate in response to the metabolic demands of exercise, blood-flow augmentation is limited by critical stenoses. Thus, perfusion pressure distal to a stenosis falls further during exercise. Pharmacologic vasodilators may not reduce resistance to blood flow any more than endogenous vasodilators released during exercise. However, vasodilator drugs may increase blood flow to unaffected regions and thereby steal blood away from the ischemic limb.

Two drugs are approved by the Food and Drug Administration for the treatment of intermittent claudication: pentoxifylline and cilostazol. Pentoxifylline is a xanthine derivative with hemorrheologic properties. It has been reported to improve red cell flexibility and decrease blood viscosity. Pentoxifylline improved patients' exercise capacity in several but not all clinical trials.[30,38]

Cilostazol is a quinolinone derivative that inhibits phosphodiesterase III and thereby prevents the degradation of cyclic adenosine monophosphate. It has vasodilatory and platelet in-

hibitory properties, but its precise mechanism of action in patients with peripheral arterial disease is not known. Several trials have found that cilostazol leads to an increase in the distance walked before onset of claudication and also in the maximal walking distance in patients with peripheral arterial disease.[39-42]

Metal-chelating compounds, such as ethylenediaminetetraacetic acid (EDTA), are not useful in the treatment of patients with peripheral arterial disease.[43]

Several classes of drugs are currently undergoing investigation for use in the treatment of claudication, critical limb ischemia, or both. Some drug treatments are designed to increase the efficiency of substrate utilization, which enhances cellular energetics. L-Carnitine and its analogue, propionyl-L-carnitine, may decrease the ratio of acetyl coenzyme A (acetyl CoA) to CoA via the action of CoA:carnitine acetyltransferase and thereby stimulate glucose oxidation and energy production. Small placebo-controlled trials have found that treatment with L-carnitine or propionyl-L-carnitine improves exercise capacity in patients with intermittent claudication.[44,45]

In one study, L-arginine, the precursor of nitric oxide, improved endothelium-dependent vasodilatation and increased claudication distance after 3 weeks of intensive therapy.[46] Initial trials of prostaglandin E_1 (PGE_1) and prostacyclin (PGI_2) or their synthetic analogues suggested that these agents could increase the distance walked before onset of claudication, but subsequent definitive trials failed to show improvement in symptoms.[47,48] Angiogenic growth factors, such as vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF), are undergoing intensive investigation for their potential efficacy in patients with peripheral arterial disease. These angiogenic factors may be delivered parenterally as recombinant proteins or through gene transfer using intra-arterial catheter techniques or intramuscular injection. Both VEGF and bFGF increase collateral blood vessel development and improve blood flow in experimental models of hindlimb ischemia. The efficacy of angiogenic growth factors in patients with intermittent claudication or critical limb ischemia is an active area of investigation. Several placebo-controlled trials in patients with claudication have been reported. In one trial, intra-arterial infusion of recombinant FGF-2 resulted in a significant increase in peak walking time at 90 days.[49] In another study, intramuscular administration of VEGF did not improve exercise performance.[50]

Autologous implantation of bone marrow mononuclear cells is a promising area of study. These cells have the potential to promote angiogenesis, because they can supply endothelial progenitor cells and they secrete angiogenic factors.[51]

Revascularization

Revascularization procedures are indicated for patients with disabling claudication, ischemic rest pain, or impending limb loss. Revascularization can be achieved by catheter-based endovascular interventions [*see Figure 4*] or surgical reconstruction.

Percutaneous transluminal angioplasty (PTA) of iliac arteries has an initial success rate of 90%.[52,53] Patency rates after 4 to 5 years are approximately 60% to 80% and are even higher with implantation of a stent.[54-56] The success rate of PTA of femoral and popliteal arteries is lower than that of PTA of iliac arteries. Patency rates at 1, 3, and 5 years are approximately 60%, 50%, and 45%, respectively.[54] The patency rate is better when PTA is performed for relief of claudication rather than for limb salvage and is also better in patients with good runoff (i.e., in patients with open distal vessels). Stents have not been shown to improve the patency rates of femoral and popliteal arteries over PTA alone. PTA of tibial and peroneal arteries is associated with poorer outcome than PTA of more proximal lesions and is usually performed in patients with critical limb ischemia who are considered at high risk for vascular surgery. Limb salvage rates of 1 to 2 years range from 50% to 75%. Thrombolytic therapy is not used routinely for the treatment of peripheral atherosclerosis but may be effective in restoring patency of native arteries and bypass grafts after acute arterial occlusion [*see* Acute Arterial Occlusion, *below*].

a *b*

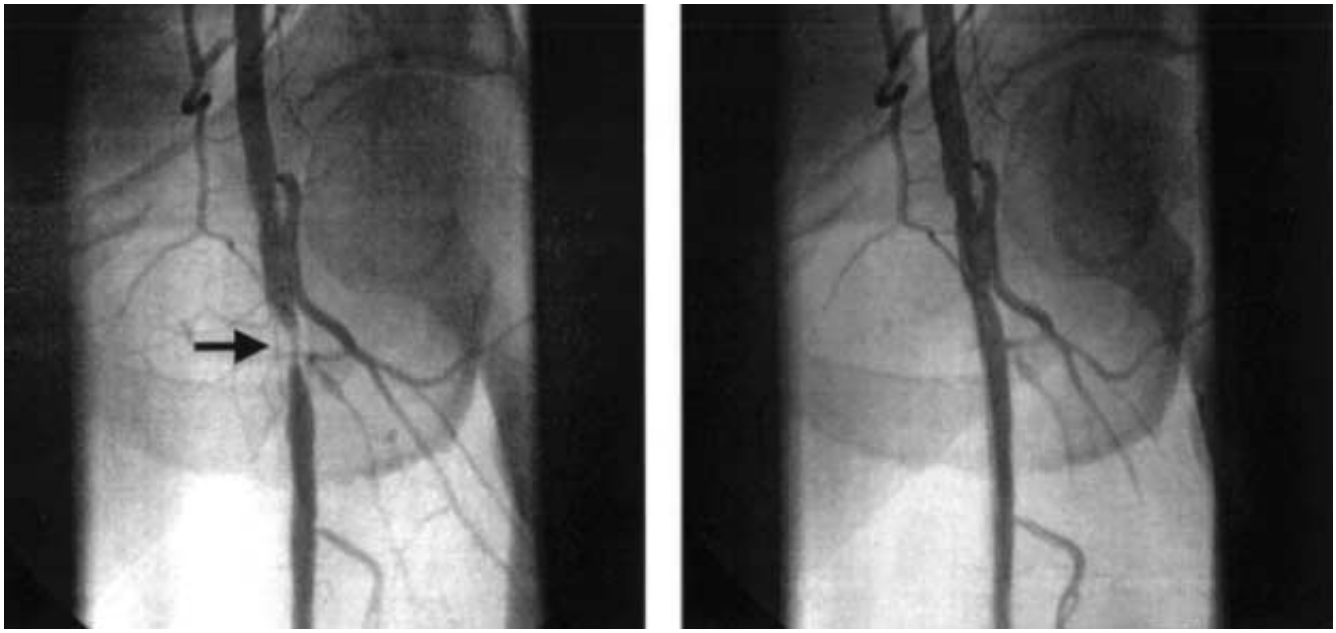


Figure 4 **Arteriograms of a patient with disabling claudication of the left leg. A focal stenosis (arrow) of the superficial femoral artery is apparent (*a*). After percutaneous transluminal angioplasty, patency is restored (*b*).**

The operative procedures used in vascular reconstruction depend on the location and severity of the arterial stenoses. Aortobifemoral bypass with a bifurcated Dacron or polytetrafluoroethylene prosthetic graft is the standard treatment for aortoiliac disease. Operative mortality ranges from 1% to 3% at centers with expertise in this technique. Long-term patency and relief of symptoms exceed 80% over 10 years.[57] Intra-abdominal aortoiliac reconstructive surgery is not feasible in patients whose comorbid conditions pose excessive surgical risk. Axillobifemoral bypass can circumvent the abdominal aorta and achieve revascularization of both legs. Femorofemoral bypass can be performed with the patient under regional anesthesia and is appropriate in cases of unilateral iliac artery obstruction.

Infrainguinal bypass procedures include femoral-popliteal and femoral-tibioperoneal reconstruction. Two techniques are generally used: the in situ saphenous vein bypass graft and the reversed autologous saphenous vein bypass graft. Femoral-popliteal reconstruction is most successful when the distal anastomosis is constructed proximal to the knee. The 5-year patency rate for all saphenous vein infrainguinal bypass grafts, including grafts that have undergone revision, is approximately 75% to 80%.[54,58] Patency rates are higher in claudicants than in patients with critical limb ischemia. Synthetic grafts made of polytetrafluoroethylene are used when veins are not available.[59] The patency of prosthetic grafts is inferior to those composed of veins, particularly because of early thrombotic occlusion. Synthetic grafts inserted below the knee have a very low patency rate and are typically not used for tibioperoneal reconstruction. Operative mortality for infrainguinal vascular reconstruction is 1% to 2%.[58] Antiplatelet agents should be administered to maintain graft patency after bypass grafts.[2,23,33]

Lumbar sympathectomy is used rarely to treat patients with critical limb ischemia. The pathophysiology of limb ischemia suggests that ischemic vessels are maximally vasodilated; thus, lumbar sympathectomy may not increase blood flow.

Amputation is a surgical alternative for patients with advanced limb ischemia in whom revascularization procedures are not possible or have failed. It is a final alternative for patients with unremitting rest pain or gangrene. Selection of the amputation level requires assessment of perfusion. Transphalangeal amputation causes minimal disability. Transmetatarsal amputation of the forefoot may affect balance, but patients are usually able to ambulate after rehabilitation. Patients who undergo below-the-knee amputation and subsequently use a prosthesis expend 10% to 40% more energy to walk on a horizontal surface than a person who has use of both legs. Patients who undergo amputations above the knee and use a prosthetic device expend 65% more energy to walk than a person who has use of both legs. Overall prognosis after major leg amputation is poor, usually because of coexisting coronary and cerebrovascular disease.

Acute Arterial Occlusion

Acute arterial occlusion is to be distinguished from the gradual development of limb artery obstruction caused by peripheral atherosclerosis. The causes of acute arterial occlusion include embolism, thrombosis, dissection, and trauma. The most common cause is arterial embolism. The majority of systemic emboli arise from cardiac sources, including atrial fibrillation, valvular heart disease, congestive heart failure, left ventricular aneurysm, acute myocardial infarction, and cardiac tumors (e.g., left atrial myxomas). Noncardiac sources of embolism include aneurysms of the aorta and aneurysms of the iliac, femoral, and popliteal arteries. A deep vein thrombus may enter the systemic circulation via an intracardiac shunt, resulting in what is termed paradoxical embolism. Thrombosis in situ may develop in peripheral atherosclerotic arteries at a site of plaque rupture and in bypass grafts. Thrombus may also develop in otherwise normal vessels of patients with procoagulant disorders such as hyperhomocysteinemia (including homocysteinuria), antiphospholipid antibody syndrome, and heparin-induced thrombocytopenia. Arterial thrombus formation is uncommon in patients with resistance to activated protein C and in patients deficient in protein C, protein S, or antithrombin. Aortic dissection and trauma may acutely occlude arteries by disrupting the integrity of the vessel lumen.

Acute arterial occlusion may cause severe limb ischemia, resulting in pain, paresthesia, and motor weakness distal to the site of occlusion. There is loss of peripheral pulses, cool skin, and pallor or cyanosis distal to the obstruction site. Noninvasive tests can provide additional evidence of peripheral arterial occlusion and may reveal the severity of ischemia, but definitive treatment should not be delayed. Arteriography is used to define the site of acute arterial occlusion and may distinguish thrombus in an atherosclerotic vessel from an arterial embolism. Once the diagnosis is made, anticoagulation with heparin should be initiated to prevent propagation of the thrombus.

Acute severe limb ischemia requires urgent revascularization. Catheter-directed intra-arterial thrombolysis with agents such as recombinant human tissue plasminogen activator may restore patency in acutely occluded arteries and bypass grafts. An embolectomy catheter can be used to remove arterial emboli. Surgical reconstruction to bypass the occlusion is considered if embolectomy is unsuccessful or not possible. The decision to utilize thrombolysis or surgery for acute arterial occlusion depends in part on the severity of ischemia and urgency of revascularization.[60,61]

Atheroembolism

Atherothrombotic debris from friable plaques in the aorta or other large arteries may dislodge and embolize to small distal limb arteries. Atheroembolism occurs spontaneously, although it occasionally occurs as a complication of arterial catheterization.[62] Violaceous discoloration, petechiae, and livedo reticularis appear when emboli occlude small vessels. Occlusion of digital vessels causes painful cyanotic toes (the blue toe syndrome), despite the presence of palpable pedal arteries [*see Figure 5*]. Embolic occlusion of intramuscular vessels causes pain and tenderness. Abnormal laboratory findings include an elevated eosinophil count and an increased erythrocyte sedimentation rate. Anemia, thrombocytopenia, and hypocomplementemia may also occur. Azotemia may occur if there is concurrent atheroembolism to the kidneys. Sites of shaggy atheroma may be identified by imaging the aorta with transesophageal echocardiography or MRA.[63] Confirmation of the diagnosis is made by skin or muscle biopsy. Tissue examination will reveal elongated needle-shaped clefts in small arteries that are associated with intimal thickening, perivascular fibrosis, inflammatory cells, and lipid-laden giant cells.

The risk of recurrence of atheroembolism is high. Platelet inhibitors have been used in this disorder, although it has not been established that these agents prevent recurrent atheroemboli. The role of warfarin is even less clear. Some investigators have

found that warfarin reduces the likelihood of atheroembolism in patients with mobile atheromas, whereas others have suggested that warfarin may contribute to the development of atheroemboli in persons with a predisposition to the disease.[64-66] Surgical bypass of occluded vessels usually is not possible, because the emboli typically lodge in small distal arteries. If a proximal source, such as an aneurysm, is identified, bypass surgery and removal of the source from the circulation may reduce the risk of recurrence. Risk-factor modification—in particular, lipid-lowering therapy—can serve to stabilize atheromatous plaques and may reduce the risk of cardiovascular events; it is not known whether lipid-lowering therapy can prevent atheroembolism.[66]

Popliteal Artery Entrapment

Popliteal artery entrapment is caused by a congenital anomaly in which the medial head of the gastrocnemius muscle compresses or displaces the popliteal artery. In young patients who present with symptoms of intermittent claudication or rest pain, popliteal artery entrapment should be considered a possible diagnosis. It occurs more frequently in men than in women and is unilateral in two thirds of cases.[67]

The diagnosis is made by measuring ankle pressures before and after exercising the calf muscle, because contraction of the gastrocnemius muscle compresses the popliteal artery. Duplex ultrasonography can demonstrate popliteal artery compression and cessation of blood flow during gastrocnemius contraction. Angiography is used to confirm the diagnosis by delineating the altered course of the popliteal artery and may reveal a popliteal artery thrombus and poststenotic dilatation.

Popliteal artery entrapment should be treated surgically, preferably by relieving compression of the popliteal artery. Occasionally, thrombectomy or bypass grafting is required.

Thromboangiitis Obliterans

Thromboangiitis obliterans is a vasculitis that is also known as Buerger disease.[68] In the United States, the prevalence is approximately 1 per 10,000 population. It occurs throughout the world but is most prevalent in Asia, portions of Eastern Europe, and Israel. Thromboangiitis obliterans affects men primarily but may also occur in women. Onset of the disease usually occurs before 45 years of age. The most important predisposing factor is tobacco use.

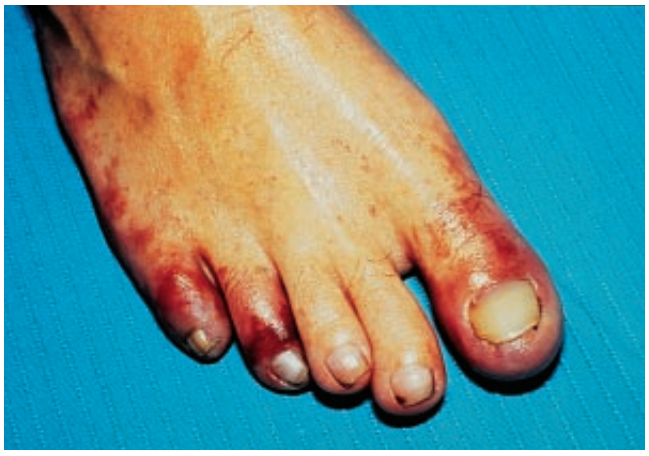


Figure 5 **Ischemia of the toes of the right foot caused by atheroemboli. There is fixed violaceous discoloration of several toes and the lateral aspect of the right foot.**

Thromboangiitis obliterans affects small and middle-sized arteries and veins in the extremities. Inflammatory cells, particularly polymorphonuclear leukocytes, infiltrate the intima, media, and adventitia; thrombi typically occlude the lumen. Leukocytes and multinucleated giant cells may be found within or surrounding the thrombus. The internal elastic lamina and media remain intact.

Involvement of limb arteries causes forearm, calf, or foot claudication. Severe ischemia of the hand and foot causes rest pain, ulcerations, and skin necrosis. Raynaud phenomenon, which is indicative of digital artery obstruction, occurs in approximately 45% of patients. Migratory superficial vein thrombosis develops in approximately 40% of patients.

There are no specific serologic laboratory tests to diagnose thromboangiitis obliterans; however, serologic tests are used to exclude other causes of vasculitis. Serum immunologic markers such as antinuclear antibodies, rheumatoid factor, and antiphospholipid antibodies should not be present, and acute-phase reactants are usually normal. The diagnosis can be supported by arteriography, which reveals interspersed affected and normal segments of blood vessels. Collateral vessels circumventing sites of occlusion are often present. Biopsy of affected vessels should reveal the typical pathologic findings described above but is rarely indicated.

The most effective treatment for patients with thromboangiitis obliterans is smoking cessation. The risk of progression to critical limb ischemia and amputation is greater in patients who continue to smoke. Surgical revascularization is not usually an option because of involvement of small distal vessels. There is no established pharmacologic intervention. The use of vasodilator prostaglandins may be beneficial in some patients. Intramuscular administration of naked plasmid DNA encoding the 165 amino acid isoform of human vascular endothelial growth factor [phVEGF (165)] was reported to heal ulcers and relieve pain in some patients with thromboangiitis obliterans.[69] The efficacy of platelet inhibitors, anticoagulants, and thrombolytic therapy has not been established.

Raynaud Phenomenon

Raynaud phenomenon is episodic vasospastic ischemia of the digits. It is characterized by digital blanching, cyanosis, and rubor after exposure to cold and rewarming and can also be induced by emotional stress. Although many patients describe a triphasic color response, most experience only one or two color changes. The digital discoloration is confined primarily to the fingers or toes. Occasionally, the tongue, tip of the nose, or earlobes are affected. Blanching represents the ischemic phase of the phenomenon, caused by digital vasospasm. Cyanosis results from deoxygenated blood in capillaries and venules. With rewarming and resolution of the digital vasospasm, a hyperemic phase ensues, causing the digits to appear red.[70]

Raynaud phenomenon is categorized as primary or secondary [*see Table 4*]. The primary form of Raynaud phenomenon is also called Raynaud disease. Diagnostic criteria for Raynaud disease include episodic digital ischemia, absence of arterial occlusion, bilateral distribution, absence of symptoms or signs of other diseases that also cause Raynaud phenomenon, and duration of symptoms for 2 years or longer. Most people with Raynaud disease develop symptoms before they reach 40 years of

Table 4 Secondary Causes of Raynaud Phenomenon

- Connective tissue diseases
  - Scleroderma
  - Systemic lupus erythematosus
  - Rheumatoid arthritis
  - Dermatomyositis
  - Mixed connective tissue disease
  - Sjögren syndrome
- Peripheral arterial occlusive diseases
  - Atherosclerosis
  - Thromboangiitis obliterans
  - Thromboembolism
  - Thoracic outlet syndrome
- Neurologic disorders
  - Carpal tunnel syndrome
  - Reflex sympathetic dystrophy
  - Stroke
  - Intervertebral disk disease
  - Spinal cord tumors
  - Syringomyelia
- Blood dyscrasias
  - Hyperviscosity syndromes
  - Myeloproliferative disorders
  - Cold agglutinin disease
  - Cryoglobulinemia
- Trauma
  - Thermal injury
    - Frostbite
  - Percussive injury
    - Exposure to vibrating tools
    - Hypothenar hammer syndrome
- Drugs
  - Antimetabolites
    - Vinblastine
    - Bleomycin
    - Cisplatin
  - Beta-adrenergic blockers
  - Ergot alkaloids
    - Ergotamine
    - Bromocriptine
  - Tricyclic antidepressants
    - Imipramine
  - Amphetamines
- Miscellaneous conditions
  - Primary pulmonary hypertension
  - Hypothyroidism

age. It can occur in young children. Raynaud disease affects women three to five times more frequently than men. The prevalence is lower in warm climates than in cold climates.

ETIOLOGY

The mechanisms postulated to cause Raynaud phenomenon include increased sympathetic nervous system activity, heightened digital vascular reactivity to vasoconstrictive stimuli, circulating vasoactive hormones, and decreased intravascular pressure. The sympathetic nervous system mediates the digital vasoconstrictive response to cold exposure and emotional stress, but sympathetic nervous system activity has been discounted as a primary causal mechanism. Some investigators have suggested that increased sensitivity, increased numbers of postsynaptic alpha$_2$-adrenergic receptors, or both enhance the vasoconstrictive reactivity to sympathetic stimulation.[71] In some cases of Raynaud phenomenon, endogenous vasoactive substances (e.g., angiotensin II, serotonin, and thromboxane A_2) and exogenous vasoconstrictors (e.g., ergot alkaloids and sympathomimetic drugs) may cause digital vasospasm. Many patients with Raynaud phenomenon have low blood pressure. Decreased digital vascular pressure caused by proximal arterial occlusive disease or by digital vascular obstruction may increase the likelihood of digital vasospasm when vasoconstrictive stimuli occur.

DIAGNOSIS

Noninvasive vascular tests that are occasionally used to evaluate patients with Raynaud phenomenon include digital pulse volume recordings and measurement of digital systolic blood pressure and digital blood flow. Nail-fold capillary microscopy is normal in patients with Raynaud disease, whereas deformed capillary loops and avascular areas are present in patients with connective tissue disorders or other conditions that cause digital vascular occlusion.[72] Determinations of the erythrocyte sedimentation rate and titers of antinuclear antibody, rheumatoid factor, cryoglobulins, and cold agglutinins are useful in excluding specific secondary causes of Raynaud phenomenon. Angiography is not necessary to diagnose Raynaud phenomenon but may be indicated in patients with persistent digital ischemia secondary to atherosclerosis, thromboembolism, or thromboangiitis obliterans to identify a cause that may be treated effectively with a revascularization procedure.

TREATMENT

Patients with Raynaud phenomenon should avoid unnecessary exposure to cold and should wear warm clothing. The hands, feet, trunk, and head should be kept warm to avoid reflex vasoconstriction. Pharmacologic intervention is indicated in patients who do not respond satisfactorily to conservative measures. Calcium channel blockers, such as nifedipine, and sympathetic nervous system inhibitors, such as prazosin and its longer-acting analogues, can be used to treat Raynaud phenomenon. Intravenous infusion of vasodilator prostaglandins, including PGE_1, PGI_2, and their analogues, has been reported to facilitate healing of digital ulcers in patients with scleroderma.[73] In patients with persistent severe digital ischemia, selective digital sympathectomy and microarteriolysis may facilitate ulcer healing and improve symptoms. Cervical and limb sympathectomy may also be considered in persons with severe Raynaud phenomenon, but long-term efficacy is not ensured.

Acrocyanosis

Raynaud phenomenon should be distinguished from acrocyanosis, a condition in which there is persistent bluish discoloration of the hands or feet.[74] Like Raynaud phenomenon, cyanotic discoloration intensifies during cold exposure, and rubor may appear with rewarming. Acrocyanosis affects both men and women; the age at onset is usually between 20 and 45 years. The prognosis of patients with idiopathic acrocyanosis is good, and loss of digital tissue is uncommon. Patients should avoid exposure to cold and should dress warmly. Pharmacologic intervention usually is not necessary. Alpha-adrenergic blocking agents and calcium channel blockers may be effective in some patients with acrocyanosis.

The author has received grant or research support from the Bristol-Myers Squibb–Sanofi-Aventis Partnership; he has served as a consultant for ARYx, Bristol-Myers Squibb–Sanofi-Aventis, Genzyme, Sigma Tau, Vasogen, and Wyeth; and he has participated in the speakers' bureau for Bristol-Myers Squibb–Sanofi-Aventis.

References

1. Dormandy JA, Rutherford RB: Management of peripheral arterial disease (PAD). TASC Working Group. J Vasc Surg 31:S1, 2000

2. Hirsch AT, Haskal ZJ, Hertzer NR, et al: ACC/AHA guidelines for the management of patients with peripheral arterial disease (lower extremity, renal, mesenteric, and abdominal aortic). J Am Coll Cardiol (in press)

3. Selvin E, Erlinger TP: Prevalence of and risk factors for peripheral arterial disease in the United States: results from the National Health and Nutrition Examination Survey, 1999–2000. Circulation 110:738, 2004

4. Diehm C, Schuster A, Allenberg JR, et al: High prevalence of peripheral arterial disease and co-morbidity in 6880 primary care patients: cross-sectional study. Atherosclerosis 172:95, 2004

5. Hirsch A, Criqui MH, Treat-Jacobson D, et al: Peripheral arterial disease detection, awareness, and treatment in primary care. JAMA 286:1317, 2001

6. Leng GC, Lee AJ, Fowkes FG, et al: Incidence, natural history and cardiovascular events in symptomatic and asymptomatic peripheral arterial disease in the general population. Int J Epidemiol 25:1172, 1996

7. Criqui M, Langer RD, Fronek A, et al: Mortality over a period of 10 years in patients with peripheral arterial disease. N Engl J Med 326:381, 1992

8. Dormandy J, Heeck L, Vig S: The fate of patients with critical leg ischemia. Semin Vasc Surg 12:142, 1999

9. Fowkes FG, Housley E, Riemersma RA, et al: Smoking, lipids, glucose intolerance, and blood pressure as risk factors for peripheral atherosclerosis compared with ischemic heart disease in the Edinburgh Artery Study. Am J Epidemiol 135:331, 1992

10. Murabito JM, D'Agostino RB, Silbershatz H, et al: Intermittent claudication: a risk profile from the Framingham Heart Study. Circulation 96:44, 1997

11. Hiatt WR, Hoag S, Hamman RF: Effect of diagnostic criteria on the prevalence of peripheral arterial disease. The San Luis Valley Diabetes Study. Circulation 91:1472, 1995

12. Diabetes-related amputations of lower extremities in the Medicare population—Minnesota, 1993–1995. MMWR Morb Mortal Wkly Rep 47:649, 1998

13. Graham IM, Daly LE, Refsum HM, et al: Plasma homocysteine as a risk factor for vascular disease. The European Concerted Action Project. JAMA 277:1775, 1997

14. Ridker PM, Stampfer MJ, Rifai N: Novel risk factors for systemic atherosclerosis: a comparison of C-reactive protein, fibrinogen, homocysteine, lipoprotein(a), and standard cholesterol screening as predictors of peripheral arterial disease. JAMA 285:2481, 2001

15. Pradhan AD, Rifai N, Ridker PM: Soluble intercellular adhesion molecule–1, soluble vascular adhesion molecule–1, and the development of symptomatic peripheral arterial disease in men. Circulation 106:820, 2002

16. McDermott MM, Greenland P, Liu K, et al: Leg symptoms in peripheral arterial disease: associated clinical characteristics and functional impairment. JAMA 286:1599, 2001

17. Rutherford RB, Baker JD, Ernst C, et al: Recommended standards for reports dealing with lower extremity ischemia: revised version. J Vasc Surg 26:517, 1997

18. McDermott MM, Greenland P, Liu K, et al: The ankle brachial index is associated with leg function and physical activity: the Walking and Leg Circulation Study. Ann Intern Med 136:873, 2002

19. Newman AB, Siscovick DS, Manolio TA, et al: Ankle-arm index as a marker of atherosclerosis in the Cardiovascular Health Study. Cardiovascular Heart Study (CHS) Collaborative Research Group. Circulation 88:837, 1993

20. Leng GC, Fowkes FG, Lee AJ, et al: Use of ankle brachial pressure index to predict cardiovascular events and death: a cohort study. BMJ 313:1440, 1996

21. Koelemay MJ, Lijmer JG, Stoker J, et al: Magnetic resonance angiography for the evaluation of lower extremity arterial disease: a meta-analysis. JAMA 285:1338, 2001

22. Hiatt WR: Medical treatment of peripheral arterial disease and claudication. N Engl J Med 344:1608, 2001

23. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. Antiplatelet Trialists' Collaboration. BMJ 324:21, 2002

24. Mondillo S, Ballo P, Barbati R, et al: Effects of simvastatin on walking performance and symptoms of intermittent claudication in hypercholesterolemic patients with peripheral vascular disease. Am J Med 114:359, 2003

25. Mohler ER 3rd, Hiatt WR, Creager MA: Cholesterol reduction with atorvastatin improves walking distance in patients with peripheral arterial disease. Circulation 108:1481, 2003

26. Schillinger M, Exner M, Mlekusch W, et al: Statin therapy improves cardiovascular outcome of patients with peripheral artery disease. Eur Heart J 25:742, 2004

27. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Lancet 360:7, 2002

28. Mehler PS, Coll JR, Estacio R, et al: Intensive blood pressure control reduces the risk of cardiovascular events in patients with peripheral arterial disease and type 2 diabetes. Circulation 107:753, 2003

29. Yusuf S, Sleight P, Pogue J, et al: Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000

30. Radack K, Deck C: Beta-adrenergic blocker therapy does not worsen intermittent claudication in subjects with peripheral arterial disease: a meta-analysis of randomized controlled trials. Arch Intern Med 151:1769, 1991

31. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet 352:837, 1998

32. Goldhaber SZ, Manson JE, Stampfer MJ, et al: Low-dose aspirin and subsequent peripheral arterial surgery in the Physicians' Health Study. Lancet 340:143, 1992

33. Collaborative overview of randomised trials of antiplatelet therapy. II: Maintenance of vascular graft or arterial patency by antiplatelet therapy. Antiplatelet Trialists' Collaboration. BMJ 308:159, 1994

34. Clagett GP, Sobel M, Jackson MR, et al: Antithrombotic therapy in peripheral arterial occlusive disease: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 126(3 suppl):609S, 2004

35. A randomised, blinded trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. Lancet 348:1329, 1996

36. Gardner AW, Poehlman ET: Exercise rehabilitation programs for the treatment of claudication pain: a meta-analysis. JAMA 274:975, 1995

37. Stewart KJ, Hiatt WR, Regensteiner JG, et al: Exercise training for claudication. N Engl J Med 347:1941, 2002

38. Lindgarde F, Labs KH, Rossner M: The pentoxifylline experience: exercise testing reconsidered. Vasc Med 1:145, 1996

39. Beebe HG, Dawson DL, Cutler BS, et al: A new pharmacological treatment for intermittent claudication: results of a randomized, multicenter trial. Arch Intern Med 159:2041, 1999

40. Dawson DL, Cutler BS, Hiatt WR, et al: A comparison of cilostazol and pentoxifylline for treating intermittent claudication. Am J Med 109:523, 2000

41. Thompson PD, Zimet R, Forbes WP, et al: Meta-analysis of results from eight randomized, placebo-controlled trials on the effect of cilostazol on patients with intermittent claudication. Am J Cardiol 90:1314, 2002

42. Regensteiner J, Ware JE Jr, McCarthy WJ, et al: Effect of cilostazol on treadmill walking, community-based walking ability, and health-related quality of life in patients with intermittent claudication due to peripheral arterial disease: meta-analysis of six randomized controlled trials. J Am Geriatr Soc 50:1939, 2002

43. van Rij AM, Solomon C, Packer SG, et al: Chelation therapy for intermittent claudication: a double-blind, randomized, controlled trial. Circulation 90:1194, 1994

44. Brevetti G, Diehm C, Lambert D: European multicenter study on propionyl-L-carnitine in intermittent claudication. J Am Coll Cardiol 34:1618, 1999

45. Hiatt WR, Regensteiner JG, Creager MA, et al: Propionyl-L-carnitine improves exercise performance and functional status in patients with claudication. Am J Med 110:616, 2001

46. Boger RH, Bode-Boger SM, Thiele W, et al: Restoring vascular nitric oxide formation by L-arginine improves the symptoms of intermittent claudication in patients with peripheral arterial occlusive disease. J Am Coll Cardiol 32:1336, 1998

47. Lievre M, Morand S, Besse B, et al: Oral beraprost sodium, a prostaglandin I(2) analogue, for intermittent claudication: a double-blind, randomized, multicenter controlled trial. Beraprost et Claudication Intermittente (BERCI) Research Group. Circulation 102:426, 2000

48. Mohler ER 3rd, Hiatt WR, Olin JW, et al: Treatment of intermittent claudication with beraprost sodium, an orally active prostaglandin I2 analogue: a double-blinded, randomized, controlled trial. J Am Coll Cardiol 41:1679, 2003

49. Lederman RJ, Mendelsohn FO, Anderson RD, et al: Therapeutic angiogenesis with recombinant fibroblast growth factor–2 for intermittent claudication (the TRAFFIC study): a randomised trial. Lancet 359:2053, 2002

50. Rajagopalan S, Mohler ER 3rd, Lederman RJ, et al: Regional angiogenesis with vascular endothelial growth factor in peripheral arterial disease: a phase II randomized, double-blind, controlled study of adenoviral delivery of vascular endothelial growth factor 121 in patients with disabling intermittent claudication. Circulation 108:1933, 2003

51. Tateishi-Yuyama E, Matsubara H, Murohara T, et al: Therapeutic angiogenesis for patients with limb ischaemia by autologous transplantation of bone-marrow cells: a pilot study and a randomised controlled trial. Lancet 360:427, 2002

52. Kandarpa K, Becker GJ, Hunink MG, et al: Transcatheter interventions for the treatment of peripheral atherosclerotic lesions: part I. J Vasc Interv Radiol 12:683, 2001

53. Kanani RS, Garasic JM: Lower extremity arterial occlusive disease: role of percutaneous revascularization. Curr Treat Options Cardiovasc Med 7:99, 2005

54. Hunink MG, Wong JB, Donaldson MC, et al: Revascularization for femoropopliteal disease: a decision and cost-effectiveness analysis. JAMA 274:165, 1995

55. Bosch JL, Hunink MG: Meta-analysis of the results of percutaneous transluminal angioplasty and stent placement for aortoiliac occlusive disease. Radiology 204:87, 1997

56. Tetteroo E, van der Graaf Y, Bosch JL, et al: Randomised comparison of primary stent placement versus primary angioplasty followed by selective stent placement in patients with iliac-artery occlusive disease. Dutch Iliac Stent Trial Study Group. Lancet 351:1153, 1998

57. de Vries SO, Hunink MG: Results of aortic bifurcation grafts for aortoiliac occlusive disease: a meta-analysis. J Vasc Surg 26:558, 1997

58. Belkin M, Knox J, Donaldson MC, et al: Infrainguinal arterial reconstruction with nonreversed greater saphenous vein. J Vasc Surg 24:957, 1996

59. Abbott WM, Green RM, Matsumoto T, et al: Prosthetic above-knee femoropopliteal bypass grafting: results of a multicenter randomized prospective trial. Above-Knee Femoropopliteal Study Group. J Vasc Surg 25:19, 1997

60. Ouriel K, Veith FJ, Sasahara AA: A comparison of recombinant urokinase with vascular surgery as initial treatment for acute arterial occlusion of the legs. Thrombolysis or Peripheral Arterial Surgery (TOPAS) Investigators. N Engl J Med 338:1105, 1998

61. Results of a prospective randomized trial evaluating surgery versus thrombolysis for ischemia of the lower extremity. The STILE trial. Ann Surg 220:251, 1994

62. Fukumoto Y, Tsutsui H, Tsuchihashi M, et al: The incidence and risk factors of cholesterol embolization syndrome, a complication of cardiac catheterization: a prospective study. J Am Coll Cardiol 42:211, 2003

63. Atherosclerotic disease of the aortic arch as a risk factor for recurrent ischemic stroke. The French Study of Aortic Plaques in Stroke Group. N Engl J Med 334:1216, 1996

64. Transesophageal echocardiographic correlates of thromboembolism in high-risk patients with nonvalvular atrial fibrillation. The Stroke Prevention in Atrial Fibrillation Investigators Committee on Echocardiography. Ann Intern Med 128:639, 1998

65. Ferrari E, Vidal R, Chevallier T, et al: Atherosclerosis of the thoracic aorta and aortic debris as a marker of poor prognosis: benefit of oral anticoagulants. J Am Coll Cardiol 33:1317, 1999

66. Tunick PA, Kronzon I: Embolism from the aorta: atheroemboli and thromboemboli. Curr Treat Options Cardiovasc Med 3:181, 2001

67. Persky JM, Kempczinski RF, Fowl RJ: Entrapment of the popliteal artery. Surg Gynecol Obstet 173:84, 1991

68. Olin JW: Thromboangiitis obliterans (Buerger's disease). N Engl J Med 343:864, 2000

69. Isner JM, Baumgartner I, Rauh G, et al: Treatment of thromboangiitis obliterans (Buerger's disease) by intramuscular gene transfer of vascular endothelial growth factor: preliminary clinical results. J Vasc Surg 28:964, 1998

70. Wigley FM: Clinical practice. Raynaud's phenomenon. N Engl J Med 347:1001, 2002

71. Freedman RR, Baer RP, Mayes MD: Blockade of vasospastic attacks by alpha 2-adrenergic but not alpha 1-adrenergic antagonists in idiopathic Raynaud's disease. Circulation 92:1448, 1995

72. Anderson ME, Allen PD, Moore T, et al: Computerized nailfold video capillaroscopy: a new tool for assessment of Raynaud's phenomenon. J Rheumatol 32:841, 2005

73. Wigley FM, Wise RA, Seibold JR, et al: Intravenous iloprost infusion in patients with Raynaud phenomenon secondary to systemic sclerosis: a multicenter, placebo-controlled, double-blind study. Ann Intern Med 120:199, 1994

74. Nousari HC, Kimyai-Asadi A, Anhalt GJ: Chronic idiopathic acrocyanosis. J Am Acad Dermatol 45(6 suppl):S207, 2001

75. Rutherford RB: Standards for evaluating results of interventional therapy for peripheral vascular disease. Circulation 83(suppl I):1-6, 1991

XVIII VENOUS THROMBOEMBOLISM

Clive Kearon, m.b., ph.d.
Jack Hirsh, m.d.

Venous thromboembolism, which involves venous thrombosis and pulmonary embolism, is a leading cause of morbidity and mortality in hospitalized patients and is being seen with increasing frequency in outpatients.[1,2] This increased incidence of venous thromboembolism in outpatients may be attributable to clinicians' heightened awareness of the importance of this condition and the comparatively recent development of reliable noninvasive tests for its diagnosis.

Risk Factors and Etiology

Most patients with venous thromboembolism have one or more clinical risk factors. The most common risk factors are recent surgery, trauma, and immobility, as well as serious illness, including chronic heart failure, stroke, malignancy, and inflammatory bowel disease.[3] The common risk factors in outpatients include hospital admission within the past 6 months, malignancy, presence of antiphospholipid antibody, and familial thrombophilia.[2,4,5] Less common risk factors are paroxysmal nocturnal hemoglobinuria, nephrotic syndrome, and polycythemia vera.

Classification

Although venous thrombosis can occur in any vein in the body, it usually involves superficial or deep veins of the legs. Thrombosis in a superficial vein of the leg is generally benign and self-limiting but can be serious if it extends from the long saphenous vein into the common femoral vein. Superficial thrombophlebitis is easily recognized by the presence of a tender vein surrounded by an area of erythema, heat, and edema. A thrombus can often be palpated in the affected vein. Superficial thrombophlebitis may be associated with deep vein thrombosis (DVT), which typically is clinically silent. Thrombosis involving the deep veins of the leg may be confined to calf veins or may extend into the popliteal or more proximal veins. Thrombi confined to calf veins are usually small and are rarely associated with pulmonary embolism. About 20% of calf vein thrombi, however, extend into the popliteal vein and beyond, where they can cause serious complications.[3,5,6] About 50% of patients with symptomatic proximal vein thrombosis also have clinically silent pulmonary embolism,[5,6] and about 70% of patients with symptomatic pulmonary embolism have DVT, which is usually clinically silent.[3,5,7]

Pulmonary embolism is the most serious and most feared complication of venous thrombosis, but the postthrombotic syndrome is responsible for greater morbidity. The postthrombotic syndrome occurs as a long-term complication in about 25% (and is severe in about 10%) of patients with symptomatic proximal vein thrombosis in the 8 years after the acute event, with most cases developing within 2 years afterward.[5,8,9] Clinically, the postthrombotic syndrome may mimic acute venous thrombosis but typically presents as chronic leg pain that is associated with edema and worsens at the end of the day. Some patients also have stasis pigmentation, induration, and skin ulceration; a smaller number of patients have venous claudication on walking, caused by persistent obstruction of the iliac veins.

Pathophysiology

Venous thrombi are composed predominantly of fibrin and red blood cells. They usually arise at sites of vessel damage or in the large venous sinuses of the calves or the valve cusp pockets in the deep veins of the calves. Thrombosis occurs when blood coagulation overwhelms the natural anticoagulant mechanisms and the fibrinolytic system. Coagulation is usually triggered when blood is exposed to tissue factor on the surface of activated monocytes that are attracted to sites of tissue damage or vascular trauma. Clinical risk factors that activate blood coagulation include extensive surgery, trauma, burns, malignant disease, myocardial infarction, cancer chemotherapy, and local hypoxia produced by venous stasis. Malignant cells contain a cysteine proteinase that activates factor X, which is a key clotting enzyme. Venous stasis and damage to the vessel wall increase the thrombogenic effect of blood coagulation. Venous stasis is produced by immobility, by obstruction or dilatation of veins, by increased venous pressure, and by increased blood viscosity. The critical role of stasis in the pathogenesis of venous thrombosis is exemplified by the observation that thrombosis occurs with equal frequency in the two legs in paraplegic patients but occurs with greater frequency in the paralyzed limb than in the nonparalyzed limb in stroke patients.[3]

Tissue damage also results in impaired fibrinolysis, which occurs through the release of inflammatory cytokines in response to the damage. These cytokines induce endothelial cell synthesis of plasminogen activator inhibitor–1 (PAI-1)[10]; in addition, they reduce the protective effect of the vascular endothelium by downregulating the endothelial-bound anticoagulant thrombomodulin.[3]

Increased central venous pressure, which produces venous stasis in the extremities, may explain the high incidence of venous thrombosis in patients with heart failure. Stasis resulting from venous dilatation occurs in elderly patients, in patients with varicose veins, and in women who are pregnant or using supplemental estrogen, perhaps contributing to the increased incidence of thrombosis in these persons. Venous obstruction contributes to the risk of venous thrombosis in patients with pelvic tumors. Increased blood viscosity, which also causes stasis, may explain the risk of thrombosis in patients with polycythemia vera, hypergammaglobulinemia, or chronic inflammatory disorders. Direct venous damage may lead to venous thrombosis in patients undergoing hip surgery, knee surgery, or varicose vein stripping and in patients with severe burns or trauma to the lower extremities.[3]

Blood coagulation is modulated by circulating inhibitors or by endothelial cell–bound inhibitors. The most important circulating inhibitors of coagulation are antithrombin, protein C, and protein S.[11,12] An inherited deficiency of one of these three proteins is found in about 20% of patients who have a family history of venous thrombosis and whose first episode of venous thrombosis occurs before 41 years of age.[13] Some types of congenital dysfibrinogenemias can also predispose patients to thrombosis, as can a congenital deficiency of plasminogen.[14] An inherited thrombophilic defect known as activated protein C (APC) resistance, or factor V Leiden, has now been established as the most common cause of inherited thrombophilia, occurring in about 5% of whites who do not have a family history of

venous thrombosis and in about 20% of patients with a first episode of venous thrombosis.[15,16] The second most common thrombophilic defect is a mutation (G20210A) in the 3′ untranslated region of the prothrombin gene that results in about a 25% increase in prothrombin levels.[15,17] This mutation is found in about 2% of whites who have no family history of venous thrombosis and in about 5% of patients with a first episode of venous thrombosis.[15] Elevated levels of clotting factors VIII[18] and XI[19] and of homocysteine[20] also predispose patients to thrombosis. The risk of thrombosis in patients with thrombophilic defects is increased through the use of estrogen-containing oral contraceptives.[15,21] Randomized trials have shown that the administration of estrogens in the doses used for postmenopausal hormone replacement therapy increase the risk of a first or recurrent thromboembolism about threefold.[22-24]

Natural History and Prognosis

Most venous thrombi produce no symptoms and are confined to the intramuscular and deep veins of the calf. Many calf vein thrombi undergo spontaneous lysis, but some extend into the popliteal and more proximal veins.[5] Complete lysis of proximal vein thrombosis is less common.[3,5,25] Most symptomatic pulmonary emboli and virtually all fatal emboli arise from thrombi in the proximal veins of the legs. Extensive venous thrombosis causes local valvular damage, which is thought to lead to the postthrombotic syndrome. Patients with a history of venous thrombosis are more likely to experience additional episodes, particularly if they are exposed to high-risk situations.[1]

Untreated or inadequately treated venous thrombosis is associated with a high rate of complications. About 25% of untreated calf vein thrombi extend into the popliteal vein, and about 50% of untreated proximal vein thrombi also undergo extension.[3,5,6,26] Patients with proximal vein thrombosis who are inadequately treated have a recurrence rate of about 40%,[26] and patients with symptomatic calf vein thrombosis who are treated with a 5-day course of intermittent intravenous heparin without continuation of oral anticoagulant therapy have a recurrence rate greater than 20% over the following 3 months.[27]

Complications can be decreased considerably by adequate anticoagulant therapy. Fewer than 3% of patients who have proximal vein thrombosis experience a clinically detectable recurrence during the initial period of treatment with high-dose heparin or low-molecular-weight heparin (LMWH), and fewer than 3% of patients experience recurrence during the subsequent 3 months of moderate-intensity oral anticoagulant therapy or moderate-dose subcutaneous heparin therapy.[28] After 3 months of anticoagulant therapy, patients have an annual recurrence rate of about 3% if their thrombosis developed after a reversible provocation, such as surgery; the recurrence rate is as high as 15% if the thrombosis is idiopathic or associated with ongoing conditions, such as prolonged immobilization or cancer.[16,29-34] The recurrence rate is significantly higher after a 4- or 6-week course of warfarin treatment, compared with a 3- or 6-month course.[29,31,32] Additional risk factors for recurrent venous thrombosis include proximal versus isolated distal thrombosis; hyperhomocysteinemia; elevated levels of factor VIII; an antiphospholipid antibody; homozygous factor V Leiden; and, probably, deficiency of protein C, protein S, or antithrombin.[33,35]

Table 1 Model for Determining Clinical Suspicion of Deep Vein Thrombosis[37]

Variables	*Points**
Active cancer (treatment ongoing or within previous 6 months or palliative)	1
Paralysis, paresis, or recent plaster immobilization of the lower extremities	1
Recently bedridden for more than 3 days, or major surgery within the past 4 weeks	1
Localized tenderness along the distribution of the deep venous system	1
Entire leg swollen	1
Affected calf 3 cm greater than asymptomatic calf (measured 10 cm below tibial tuberosity)	1
Pitting edema confined to the symptomatic leg	1
Dilated superficial veins (nonvaricose)	1
Alternative diagnosis is at least as likely as that of deep vein thrombosis	-2
Total points	

*Pretest probability is calculated as follows: total points ≤ 0, low probability; 1 to 2, moderate probability; ≥ 3, high probability.

Diagnosis

VENOUS THROMBOSIS

Clinical Features

The clinical features of venous thrombosis, such as localized swelling, redness, tenderness, and distal edema, are nonspecific, and the diagnosis should always be confirmed by objective tests.[6,36]

About 75% of ambulatory patients with clinically suspected venous thrombosis have another cause for their symptoms. The conditions that are most likely to simulate venous thrombosis are ruptured Baker cyst, cellulitis, muscle tear, muscle cramp, muscle hematoma, external venous compression, superficial thrombophlebitis, and the postthrombotic syndrome. Of the patients who actually have venous thrombosis, about 85% have proximal vein thrombosis; for the rest, thrombosis is confined to the calf.[5,6]

Although clinical features cannot unequivocally confirm or exclude a diagnosis of venous thrombosis, careful review of the patient's history and of the signs and symptoms at presentation are useful in diagnosis [*see Table 1*].[37] Evidence suggests that patients can be classified as having a high, intermediate, or low probability of having venous thrombosis on the basis of (1) the presence or absence of risk factors (e.g., recent immobilization, hospitalization within the past 6 months, or malignancy), (2) whether the clinical manifestations at presentation are typical or atypical, and (3) whether there is an alternative explanation for the symptoms that is at least as likely as DVT.[35-38]

Laboratory Tests

Three objective tests have been well validated for the diagnosis of venous thrombosis: venography, impedance plethysmography (IPG), and venous ultrasonography.[6] Of these, venography and venous ultrasonography are most widely used. Magnetic resonance imaging appears to be an accurate test for the diagnosis of DVT[39] and may prove to be more accurate than ul-

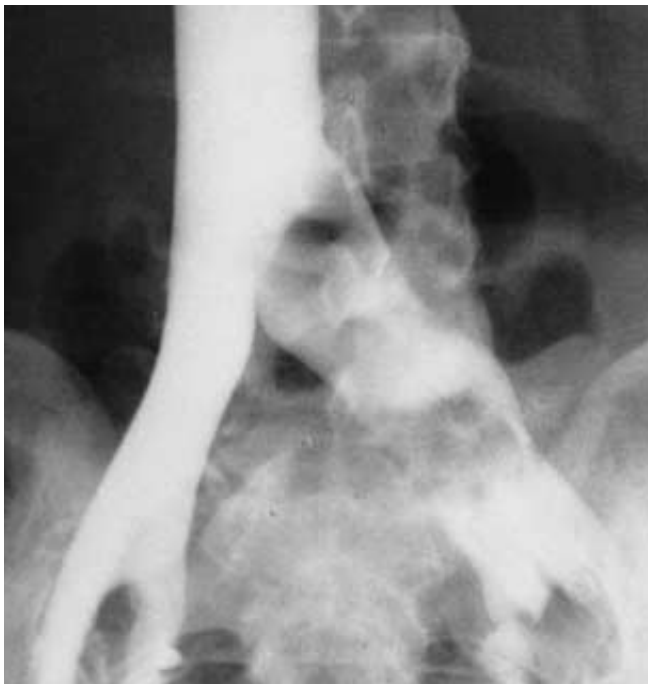


Figure 1 **Filling defects in the left iliac vein, apparent in this venogram, reveal the presence of thrombi.**

trasonography for the diagnosis of isolated calf or pelvic vein thrombosis and of recurrence of DVT.

Venography Venography, which involves the injection of a radiocontrast agent into a distal vein, is the reference standard for the diagnosis of venous thrombosis [*see Figure 1*]. Venography detects both proximal vein thrombosis and calf vein thrombosis. However, it is technically difficult and expensive; can be painful; and requires injection of radiographic contrast, which can cause allergic reactions or renal impairment. For these reasons, venography has been replaced by venous ultrasonography for the diagnosis of most cases of suspected venous thrombosis.[6]

Venous ultrasonography Venous ultrasonography is the noninvasive method of choice for diagnosing venous thrombosis.[6] It is not painful and it is easier to perform than venography. The common femoral vein, superficial femoral vein, popliteal vein, and proximal deep calf veins are imaged in real time and compressed with the transducer probe. Inability to fully compress or obliterate the vein is diagnostic of venous thrombosis.[6] Duplex ultrasonography, which combines real-time imaging with pulsed Doppler and color-coded Doppler technology, facilitates the identification of veins.

Venous ultrasonography is highly accurate for the detection of proximal vein thrombosis in symptomatic patients, with reported sensitivity and specificity approaching 95%. The sensitivity for symptomatic calf vein thrombosis is approximately 70%. In asymptomatic patients who have had elective hip or knee replacement, the sensitivity of real-time or color Doppler ultrasonography for proximal DVT is only about 60%.[6]

Although venous ultrasonography fails to detect a substantial number of calf vein thrombi and small thrombi of the popliteal vein, this limitation is not critical. If the initial test result excludes proximal DVT, the test can be repeated in 7 days to detect the small number of calf vein thrombi that extend after the initial presentation.[6] If the test remains negative after 7 days, the risk that thrombus is present and will subsequently extend to the proximal veins is negligible, and it is safe to withhold treatment.[6,40]

Ultrasonography is accurate when its results are concordant with clinical assessment; its accuracy drops if the results of these two assessments do not agree.[38] Therefore, if the clinical suspicion for DVT is low and the ultrasound is abnormal, or if clinical suspicion is high and the ultrasound is normal, venography should be considered. In about one quarter of such cases, the results of venography differ from those of the ultrasound.[38] Because the prevalence of DVT is only about 2% (most of which is distal), a follow-up test is not necessary when the clinical suspicion of thrombosis is low and the result of an initial proximal venous ultrasound scan is normal [*see Table 2*].[6]

D-dimer blood testing D-dimer is formed when cross-linked fibrin in thrombi is broken down by plasmin; thus, elevated levels of D-dimer can be used to detect DVT and pulmonary embolism. A variety of D-dimer assays are available, and they vary markedly in their accuracy as diagnostic tests for venous thromboembolism.[41]

All D-dimer assays have low positive predictive value for DVT; an abnormal result is nonspecific and cannot be used to diagnose venous thrombosis. However, some D-dimer tests are sensitive for venous thrombosis, and a normal result can be used to exclude venous thromboembolism.[42] Management studies have shown that it is safe to withhold anticoagulant therapy and serial testing in patients who have a normal result on a moderately sensitive and specific D-dimer test in combination with at least one of three other findings: (1) a normal result on IPG (negative predictive value, 98.5%),[43] (2) a normal result on venous ultrasonography of the proximal veins (negative predictive value, 99.8%),[44] or (3) a low clinical suspicion for DVT[45] [*see Table 2*].

RECURRENT VENOUS THROMBOSIS

The diagnosis of acute recurrent DVT can be difficult.[6] A common approach is to perform venous ultrasonography. If the result is normal, the test should be repeated twice over the next 7 to 10 days. If the result is positive and the result of the previous test was negative, a recurrence is diagnosed. This diagnosis can also be made if venous ultrasonography shows convincing evidence of more extensive thrombosis than was seen on a previous examination. If findings on venous ultrasonography are equivocal, as compared with a previous scan, or a previous scan is not available for comparison, venography should be performed. If the venogram shows a new intraluminal filling defect or evidence of thrombus extension since a previous venogram, recurrent venous thrombosis is diagnosed. If no new defect is found, however, the diagnosis must be on the basis of clinical features [*see Table 2*].[6]

PULMONARY EMBOLISM

Clinical Features

Dyspnea is the most common symptom of pulmonary embolism. Chest pain is also common; it is usually pleuritic but can be substernal and compressive. Tachycardia is relatively common; hemoptysis is less frequent. Fewer than 25% of patients with symptomatic pulmonary embolism have clinical features of venous thrombosis.[46,47] However, the clinical features

Table 2 Test Results That Effectively Confirm or Exclude Deep Vein Thrombosis

Purpose	*Test*	*Significant Result*
Diagnostic for first DVT	Venography	Intraluminal filling defect
	Venous ultrasonography	Noncompressible proximal veins at two or more of the common femoral, popliteal, and calf trifurcation sites
Excludes first DVT	Venography	All deep veins seen, and no intraluminal filling defects
	D-dimer	Normal value on a test that has at least a moderately high sensitivity (≥ 85%) and specificity (≥ 70%) and (1) normal results on venous ultrasonography or IPG or (2) low clinical suspicion of DVT at presentation
	Venous ultrasonography or IPG	Normal and (1) low clinical suspicion for DVT at presentation, (2) normal D-dimer test at presentation, or (3) normal serial testing (venous ultrasonography at 7 days; IPG at 2 and 7 days)
Diagnostic for recurrent DVT	Venography	Intraluminal filling defect
	Venous ultrasonography	(1) A new noncompressible common femoral or popliteal vein segment or (2) a ≥ 4.0 mm increase in diameter of the common femoral or popliteal vein since a previous test*
	IPG	(1) Conversion of a normal test to abnormal* (2) An abnormal test 1 year after diagnosis*
Excludes recurrent DVT	Venography	All deep veins seen and no intraluminal filling defects
	Venous ultrasonography or IPG	Normal or ≤ 1 mm increase in diameter of the common femoral or popliteal veins on venous ultrasound since a previous test and continuing normal results (no progression of venous ultrasound) at 2 and 7 days
	D-dimer	Use of D-dimer to exclude recurrent DVT has not been well evaluated

*If other evidence is not consistent with recurrent DVT (e.g., venous ultrasonography or impedance plethysmography, clinical assessment, or D-dimer), venography should be considered.

IPG—impedance plethysmography

of pulmonary embolism, like those of venous thrombosis, are nonspecific, and in fewer than one third of symptomatic patients is the diagnosis confirmed by objective tests.[46,48]

In the past, clinical assessment of the probability of pulmonary embolism was not standardized; physicians made the assessment informally, on the basis of their experience and the results of initial routine tests (e.g., chest x-ray and electrocardiogram). Three groups have published explicit criteria for determining the clinical probability of pulmonary embolism. One group determined probability on the basis of symptoms and electrocardiography in conjunction with perfusion lung scanning,[49] whereas the other two studies used clinical features in conjunction with ventilation-perfusion lung scanning.[46,50] The model created by Wells and colleagues incorporates an assessment of symptoms and signs, the presence of an alternative diagnosis to account for the patient's condition, and the presence of risk factors for venous thromboembolism.[48,50] With this model, a patient's clinical probability of pulmonary embolism can be categorized as low (prevalence of 2%), moderate (prevalence of 19%), or high (prevalence of 60%) [*see Table 3*].[48]

Table 3 Model for Determining a Clinical Suspicion of Pulmonary Embolism[48]

Variables	*Points**
Clinical signs and symptoms of deep vein thrombosis (minimum leg swelling and pain with palpation of the deep veins)	3.0
An alternative diagnosis is less likely than pulmonary embolism	3.0
Heart rate > 100 beats/min	1.5
Immobilization or surgery in the previous 4 weeks	1.5
Previous deep vein thrombosis/pulmonary embolism	1.5
Hemoptysis	1.0
Malignancy (treatment ongoing or within previous 6 months or palliative)	1.0
Total points	

*Pretest probability is calculated as follows: total points < 2, low probability; 2 to 6, moderate probability; > 6, high probability.

Diagnostic Tests

Chest radiography and electrocardiography In patients with pulmonary embolism, chest x-rays show either normal or nonspecific findings. Chest radiography, however, is useful for exclusion of pneumothorax and other conditions that can simulate pulmonary embolism. The electrocardiogram also frequently shows normal or nonspecific findings, but it is valuable for excluding acute myocardial infarction. In the appropriate clinical setting, ECG evidence of right ventricular strain suggests pulmonary embolism.

Ventilation-perfusion lung scanning One of the main diagnostic tests for pulmonary embolism is ventilation-perfusion lung scanning.[7,45] The perfusion component of lung scanning involves the intravenous injection of isotopically labeled microaggregates of human albumin [*see Figure 2*]. These particles become trapped in the pulmonary capillary bed, so the distribution of the particles reflects blood flow in the lung; this pattern of distribution is recorded with an external photoscanner. Perfusion lung scanning is the pivotal test in the diagnostic process because a normal perfusion scan excludes a diagnosis of pulmonary em-

Figure 2 **Posterior, right posterior oblique, and left posterior oblique perfusion scans (top), which were developed by using radiopharmaceutical technetium-99m (^{99m}Tc) microspheres of albumin, show multiple perfusion defects, some relatively large, in both lungs. Three ventilation scans (bottom) made with the patient breathing krypton-81m (^{81m}Kr) were recorded simultaneously with the perfusion scans. The scans were interpreted as showing a marked ventilation-perfusion mismatch, highly suggestive of pulmonary emboli. This diagnosis was confirmed by pulmonary arteriography.**

bolism; an abnormal perfusion scan is nonspecific [*see Table 4*].[7,46]

For ventilation lung scanning, the patient inhales and exhales either radioactive gases or aerosols while a gamma camera records the distribution of radioactivity within the alveolar gas-exchange units. Ventilation imaging improves the specificity of perfusion scanning for the diagnosis of pulmonary embolism, particularly when the ventilation scan shows normal airflow at the site of a large or segmental perfusion defect.[7,46]

Unfortunately, only 40% of lung scans in patients with pulmonary embolism indicate a high probability of the disorder.[7,46] The remaining 60% are classified as non–high probability. Patients with non–high-probability findings on lung scanning require pulmonary angiography or objective tests for venous thrombosis. The latter are useful because approximately 70% of patients with proven pulmonary embolism have DVT of the legs.[46]

Table 4 Test Results that Effectively Confirm or Exclude Pulmonary Embolism[46]

Conclusion	*Test*	*Result*
Diagnostic for PE	Pulmonary angiography	Intraluminal filling defect
	Helical CT	Intraluminal filling defect in a lobar or main pulmonary artery
	Ventilation-perfusion scan	High-probability scan and moderate/high clinical probability
	Tests for DVT*	Evidence of acute DVT with nondiagnostic ventilation-perfusion scan or helical CT scan
	Pulmonary angiography	Normal
Excludes PE	Lung perfusion scan	Normal
	High-sensitivity D-dimer test†	Normal
	Moderate-sensitivity D-dimer test‡	Normal, plus low clinical suspicion of PE and normal alveolar dead space fraction on lung perfusion scan
	Combination of clinical assessment, ventilation-perfusion scan, ultrasonography of proximal leg veins, and D-dimer test‡	Low clinical suspicion, nondiagnostic scan, and negative D-dimer test
	Combination of clinical assessment, helical CT, ultrasonography of proximal leg veins, and D-dimer test‡	Low to moderate clinical suspicion and normal CT, ultrasonography, and D-dimer test results

*See Table 2.
†D-dimer assay with very high sensitivity (i.e., 98%) and at least moderate specificity (i.e., 40%).
‡D-dimer assay with at least moderately high sensitivity (i.e., 85%) and specificity (i.e., 70%).
DVT—deep vein thrombosis PE—pulmonary embolism

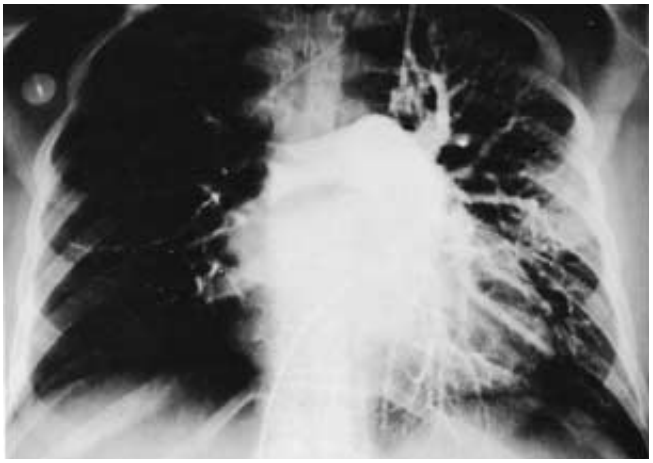


Figure 3 **Pulmonary arteriogram demonstrates nonfilling of pulmonary arterial branches in the right lung, an indication of massive occlusion.**

Pulmonary angiography Pulmonary angiography is the reference standard for establishing the presence or absence of pulmonary embolism [*see Figure 3*].[7,46] Unfortunately, it is invasive, technically difficult, and unavailable in most hospitals. If the test is performed adequately, a normal result excludes the diagnosis of pulmonary embolism; in a patient with a small perfusion defect, however, the diagnosis of a small pulmonary embolism cannot be excluded by pulmonary angiography unless the tertiary pulmonary arteries are visualized.

Pulmonary angiography can be complicated by arrhythmias, cardiac perforation, cardiac arrest, and hypersensitivity to the contrast medium. Complications occur in 3% to 4% of patients undergoing pulmonary angiography.

Computed tomography and magnetic resonance imaging Traditional CT scanning is not suitable for evaluating suspected pulmonary embolism because it is not feasible to opacify the pulmonary arteries with radiographic contrast for the time required to complete imaging (i.e., 3 minutes). Even if the pulmonary arteries could be opacified, motion artifact would interfere with image quality. These problems are overcome by helical CT (also known as spiral or continuous volume CT) because image acquisition can be completed within a single holding of the breath (e.g., 20 seconds).[46,51,52]

Current evidence suggests that helical CT has a sensitivity of about 70% for all pulmonary embolism and a specificity of about 90%. Accuracy varies according to the size of the largest pulmonary artery involved: sensitivity is about 90% for emboli in segmental or larger pulmonary arteries but only about 30% for isolated subsegmental emboli, which account for about 20% of symptomatic pulmonary emboli. Similarly, the positive predictive value of an abnormal helical CT decreases from about 100% with filling defects in main pulmonary arteries to 85% in lobar and about 60% in segmental arteries.[53] These observations suggest the following:

(1) An intraluminal filling defect in a lobar or larger pulmonary artery is likely to be associated with at least a 90% probability of embolism and, therefore, may be interpreted in the same way as a high-probability lung scan.
(2) A normal helical CT scan markedly reduces the probability of pulmonary embolism but does not exclude this diagnosis. It may be reasonable to interpret a normal spiral CT scan in the same way as a low-probability lung scan.
(3) Intraluminal defects that are confined to subsegmental pulmonary arteries are nondiagnostic and require further investigation (e.g., pulmonary angiography or lung scanning).
(4) Intraluminal filling defects that are confined to segmental pulmonary arteries should be considered nondiagnostic if clinical suspicion is low, if other diagnostic tests suggest the absence of pulmonary embolism (e.g., normal D-dimer level), or if helical CT findings are equivocal.

Musset and colleagues published a study evaluating helical CT, ultrasonography of the proximal veins, and clinical assessment for the management of 1,041 patients with suspected pulmonary embolism.[54] The main findings of this study are as follows: (1) a normal helical CT scan plus a normal ultrasound scan excluded pulmonary embolism in patients with a low or moderate clinical probability of the disorder; (2) helical CT was not sensitive enough to exclude pulmonary embolism on its own (16% of pulmonary embolism cases were diagnosed by an abnormal ultrasound scan with a normal helical CT scan); (3) isolated subsegmental filling defects were nondiagnostic for pulmonary embolism; and (4) 8% of helical CT examinations were nondiagnostic. The study provides and validates a helical CT-based algorithm for the diagnosis of pulmonary embolism.

MRI is less well evaluated than helical CT for the diagnosis of pulmonary embolism but appears to have a similar accuracy.[46] Both helical CT and MRI have the advantage of possibly suggesting an alternative pulmonary diagnosis. MRI does not expose the patient to radiation or radiographic contrast media, and the examination may be extended to look for concomitant DVT.

Compression ultrasonography and D-dimer assay Two noninvasive, relatively inexpensive, complementary approaches can be used to simplify the diagnosis of pulmonary embolism in patients with nondiagnostic lung scan findings. These approaches involve the use of compression ultrasonography and the D-dimer test. Because pulmonary emboli usually arise from mostly asymptomatic thrombi in the deep proximal leg veins, a positive compression ultrasound test result can serve as indirect evidence of pulmonary embolism. However, a positive compression ultrasound test result is found in only 5% to 10% of patients with nondiagnostic lung scans.[46,50,55] Furthermore, a negative compression ultrasound test result does not exclude a diagnosis of pulmonary embolism in these patients, possibly because either the original thrombus has embolized or the residual thrombus is too small to be detected by compression ultrasonography. A negative ultrasound test result can, however, eliminate the possibility of an associated large proximal vein thrombosis; in most patients with a negative ultrasound test result, treatment can be withheld while further investigations are performed [*see Table 4*].[46,50,55,56]

Two prospective studies—one utilizing plethysmography[57] and the other utilizing compression ultrasonography[50]—in patients with good cardiopulmonary reserve have reported that patients with non–high-probability lung scan results and a negative noninvasive test result for DVT can be safely managed with serial noninvasive testing for proximal vein thrombosis for 14 days. Both studies reported a very low rate (2%) of confirmed venous thromboembolism over a 6-month follow-up period, provided that the noninvasive test result remained negative.

About 90% of patients with a non–high-probability lung scan have a normal compression ultrasound test; of these, about 80% do not have pulmonary embolism. Therefore, the use of serial

ultrasonography in all patients with a non–high-probability lung scan result and a normal compression ultrasound test result would lead to the testing of a large number of patients to identify the very few at risk for recurrent pulmonary embolism. The noninvasive diagnostic process has been simplified by introducing two additional components: pretest probability [*see Table 3*] and D-dimer testing.

In pulmonary embolism, as in DVT, D-dimer testing is reliable only for excluding the diagnosis. In a prospective study of 1,177 patients with suspected pulmonary embolism, the negative predictive value of D-dimer testing (using the SimpliRED whole blood assay) was 99% in those patients with a low clinical likelihood of pulmonary embolism and a nondiagnostic lung scan.[58] Used alone, a rapid enzyme-linked immunosorbent assay (ELISA) for D-dimer has a high sensitivity (99%) and a low to moderate specificity (45%); it was found to rule out pulmonary embolism in 36% of consecutive patients with suspected embolism [*see Table 4*].[42]

Diagnostic Strategy

Algorithms A number of prospectively validated algorithms have been published that emphasize the use of different initial noninvasive tests in conjunction with ventilation-perfusion lung scanning. The noninvasive tests, as used in these algorithms, are as follows: structured clinical assessment and serial venous ultrasounds[50]; empirical clinical assessment, sensitive D-dimer assay, and venous ultrasound at presentation only[42]; and clinical assessment, moderately sensitive D-dimer assay, and serial venous ultrasounds[57] [*see Figure 4*]. Helical CT can be used rather than ventilation-perfusion scanning; Musset and colleagues have described and validated a diagnostic algorithm for pulmonary embolism that uses routine helical CT, ultrasonography of the proximal veins, and clinical assessment.[54]

Evaluation of patients with nondiagnostic, noninvasive tests Serial venous ultrasonography of the proximal veins (i.e., 1 and 2 weeks after the initial evaluation) is suitable for most patients in whom noninvasive tests are not diagnostic,[50,56] although pulmonary angiography is generally preferred for certain subgroups (see below). An alternative to pulmonary angiography is to perform bilateral venography before serial venous ultrasonography.

Pulmonary angiography is the preferred option in patients with a segmental intraluminal filling defect on helical CT. However, when clinical suspicion is high, this CT finding alone is likely to have a positive predictive value of 85% and could be considered diagnostic for pulmonary embolism. A subsegmental intraluminal filling defect on helical CT and high clinical probability of pulmonary embolism is also an indication for angiography, as is a high-probability ventilation-perfusion scan result and low clinical suspicion. Alternatively, helical CT scanning can be followed with ventilation-perfusion scanning, or vice versa, when the initial scan shows these findings; in such cases, the second test may be diagnostic for pulmonary embolism. If the second test is also nondiagnostic for pulmonary embolism, serial ultrasonography may be reconsidered.

Pulmonary angiography is also preferred when clinical manifestations are severe, posttest probability of pulmonary embolism is moderate, and pulmonary embolism needs to be excluded from the differential diagnosis. Finally, pulmonary angiography is indicated when serial testing is not feasible (e.g., because of impending surgery or geographic inaccessibility).

Prophylaxis and Treatment

PHARMACOLOGY OF ANTITHROMBOTIC AGENTS

Anticoagulants

A less intense anticoagulant effect is required for the prevention of venous thrombosis than is required for its treatment. The anticoagulants in clinical use are heparin, LMWH, and fondaparinux, which are administered subcutaneously or intravenously; and coumarin compounds, which are given orally. Thrombolytic agents are streptokinase, urokinase, and recombinant tissue plasminogen activator (rt-PA).

Heparin and LMWH Heparin is a highly sulfated glycosaminoglycan that produces its anticoagulant effect by binding to antithrombin, markedly accelerating the ability of the naturally occurring anticoagulant to inactivate thrombin, activated factor X (factor Xa), and activated factor IX (factor IXa).[59] At therapeutic concentrations, heparin has a half-life of about 60 minutes. Its clearance is dose dependent. Heparin has decreased bioavailability when administered subcutaneously in low doses but has approximately 90% bioavailability when administered in high therapeutic doses.

Heparin binds to a number of plasma proteins, a phenomenon that reduces the anticoagulant effect of heparin by limiting its accessibility to antithrombin. The concentration of heparin-binding proteins increases during illness, contributing to the variability in anticoagulant response in patients with thromboembolism.[59] Because of this variability, response to heparin should be monitored with the activated partial thromboplastin time (aPTT). The dose should be adjusted as necessary to achieve a therapeutic range, which for many aPTT reagents corresponds to an aPTT ratio of 1.5 to 2.5.[59]

LMWHs are effective in the prevention and treatment of venous thrombosis. They are derived from standard commercial-grade heparin by chemical depolymerization to yield fragments approximately one third the size of heparin.[60] Depolymerization of heparin results in a change in its anticoagulant profile, bioavailability, and pharmacokinetics and in a lower incidence of heparin-induced thrombocytopenia and of osteopenia.[59,60]

The plasma recoveries and pharmacokinetics of LMWHs differ from those of heparin because LMWHs bind much less avidly to heparin-binding proteins than does heparin. This property of LMWHs contributes to their superior bioavailability at low doses and their more predictable anticoagulant response. LMWHs also exhibit less binding to macrophages and endothelial cells than does heparin, a property that accounts for their longer plasma half-life, which is approximately 3 hours, and their dose-independent clearance. These potential advantages over heparin permit once-daily administration of LMWHs without laboratory monitoring. The advantages of LMWHs have been exploited to successfully treat patients with DVT out of hospital[61-63] and to treat patients with acute pulmonary embolism in hospital with once- or twice-daily subcutaneous dosing regimens.[47,63] The published research on LMWHs, which includes over 3,000 patients treated with either once-daily or twice-daily subcutaneous doses, has established this class of anticoagulants as safe, effective, and convenient for treating venous thrombosis and pulmonary embolism.[28]

Fondaparinux Fondaparinux is a new parenteral synthetic anticoagulant composed of the five saccharide units that make up the active site of heparin that binds antithrombin.[64] The fondaparinux-antithrombin complex inhibits factor Xa but has no

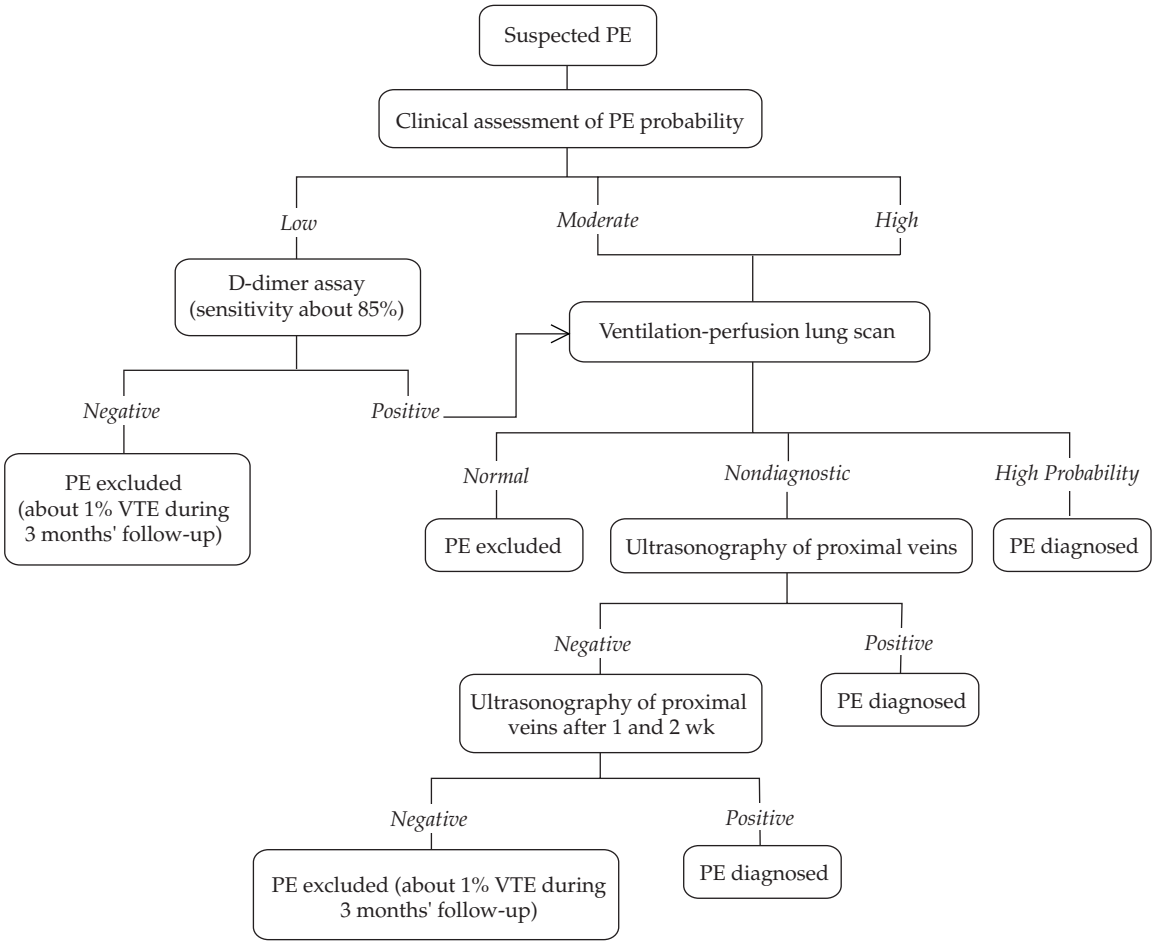


Figure 4 **Diagnostic approach to pulmonary embolism. Use of a very sensitive D-dimer assay can obviate clinical assessment: a negative result excludes pulmonary embolism regardless of the clinical assessment results; a positive test can be followed by a ventilation-perfusion scan. Similarly, a ventilation-perfusion scan can substitute for clinical assessment.**

If the clinical assessment indicates a low probability of pulmonary embolism and, in particular, if a D-dimer test is not done, pulmonary angiography or helical CT may be considered. If helical CT is used instead of ventilation-perfusion lung scanning, intraluminal filling defects in segmental or larger pulmonary arteries are generally diagnostic for pulmonary embolism, but all other findings are nondiagnostic, including normal results and intraluminal filling defects confined to the subsegmental pulmonary arteries. With nondiagnostic results, management is the same as with a nondiagnostic lung scan.

If ultrasonography is negative but overall assessment suggests a high probability of pulmonary embolism, symptoms are severe, or cardiopulmonary reserve is poor, then additional tests (e.g., helical CT or bilateral venography) may be considered. Venography should be considered if there is an increased risk of a false-positive ultrasound result (e.g., previous venous thromboembolism, equivocal ultrasound findings, preceding findings suggesting a low probability of pulmonary embolism). It is reasonable to not repeat ultrasound testing, or to do only one more ultrasound after 1 week, if preceding findings suggest a low probability of pulmonary embolism.

(PE—pulmonary embolism; VTE—venous thromboembolism)

direct activity against thrombin. Fondaparinux is rapidly absorbed and is 100% bioavailable when administered subcutaneously. It is not metabolized, is renally excreted, and has a dose-independent elimination half-life of 15 hours, which makes it suitable for once-daily administration. Fondaparinux is being evaluated for many indications, including prevention and treatment of DVT and pulmonary embolism.

Oral anticoagulants Oral anticoagulants are coumarin compounds, the most common being warfarin, that produce their anticoagulant effect through the production of hemostatically defective, vitamin K–dependent coagulant proteins (prothrombin, factor VII, factor IX, and factor X).[65]

The dose of warfarin must be monitored closely because the anticoagulant response varies widely among individuals. Laboratory monitoring is performed by measuring the prothrombin time (PT), a test responsive to depression of three of the four vitamin K–dependent clotting factors (prothrombin and factors VII and X). Commercial PT reagents vary markedly in their responsiveness to warfarin-induced reduction in clotting factors. This problem of variability in the responsiveness of PT reagents has been overcome by the introduction of the international normalized ratio (INR).[65]

The starting dose of warfarin has been 10 mg, with an average maintenance dose of about 5 mg. However, the dose required varies widely among individuals. Elderly patients, for example, have been shown, on average, to require lower doses. Evidence indicates that it might be safer to use a starting

dose of 5 mg of warfarin because, compared with 10 mg, the 5 mg starting dose does not result in a delay in achieving a therapeutic INR and is associated with a lower incidence of supratherapeutic INR values during the first 5 days of treatment.[66] Warfarin therapy is difficult to manage in some patients because of unexpected fluctuations in dose response, which may reflect changes in diet, inaccuracy in PT testing, undisclosed drug use, poor compliance, or surreptitious self-medication. Certain over-the-counter and prescription drugs can augment or inhibit the anticoagulant effect of coumarin compounds or prolong hemostasis by interfering with platelet function [*see Table 5*].

Patients receiving coumarin compounds are also sensitive to fluctuating levels of dietary vitamin K, which is obtained predominantly from leafy green vegetables. The effect of coumarins can be potentiated in sick patients with poor vitamin K intake, particularly if they are treated with antibiotics and intravenous feeding without vitamin K supplementation, and in states of fat malabsorption.

Direct thrombin inhibitors Direct thrombin inhibitors include hirudin, bivalirudin, argatroban, melagatran, and ximelagatran.[67] Whereas the first four of these compounds must be administered parenterally, ximelagatran is absorbed from the gastrointestinal tract and exhibits no known food or drug interactions. Once absorbed, ximelagatran is converted to melagatran, a partial mimetic of fibrinopeptide A, which blocks the active site of thrombin. Ximelagatran is primarily eliminated by the kidneys and has a half-life of about 3 hours, mandating twice-daily administration. It is being evaluated for many indications, including primary prevention and acute and long-term treatment of venous thromboembolism.

Table 5 Drug and Food Interactions with Warfarin by Level of Supporting Evidence* and Direction of Interaction[92]

	Antibiotics	*Cardiac*	*Anti-inflammatory*	*Central Nervous System*	*Gastrointestinal*	*Miscellaneous*
Potentiation Level 1	Trimethoprim-sulfamethoxazole, erythromycin, isoniazid, fluconazole, metronidazole, miconazole, clarithromycin, amoxicillin	Amiodarone, clofibrate, propafenone, propranolol, sulfinpyrazone†	Phenylbutazone,† piroxicam High-dose intravenous methylprednisolone Acetaminophen	Alcohol (with liver disease)	Cimetidine,‡ omeprazole	—
Level 2	Ciprofloxacin, itraconazole, tetracycline, norfloxacin	Aspirin, quinidine, simvastatin	Aspirin, dextropropoxyphene	Chloral hydrate, disulfiram, phenytoin	—	Anabolic steroids, influenza vaccine, tamoxifen
Level 3	Nalidixic acid, ofloxacin	Disopyramide, lovastatin, metolazone	Topical salicylates, sulindac, tolmetin	—	—	Fluorouracil, ifosfamide
Level 4	Cefamandole, cefazolin, sulfisoxazole, doxycycline	Gemfibrozil, heparin	Indomethacin	—	—	—
Inhibition Level 1	Griseofulvin,* nafcillin, rifampin	Cholestyramine	—	Barbiturates, carbamazepine, chlordiazepoxide	Sucralfate	Foods with a high vitamin K content, enteral nutritional support, large amounts of avocado Ticlopidine
Level 2	Dicloxacillin	—	—	—	—	—
Level 3	—	—	Azathioprine	Trazodone	—	Azathioprine, cyclosporine, etretinate, large amounts of broccoli
No effect Level 1	Enoxacin	Atenolol, bumetanide, felodipine, metoprolol, moricizine	Diflunisal, ketorolac, naproxen	Alcohol, fluoxetine, nitrazepam	Antacids, famotidine, nizatidine, psyllium, ranitidine‡	—
Level 2	Ketoconazole	—	Ibuprofen, ketoprofen	—	—	—
Level 4	Vancomycin	Diltiazem	—	—	—	Tobacco

*Level 1 evidence indicates that the likelihood of an association is very strong; level 2 evidence suggests that a true association is likely; level 3 evidence suggests that a true association is probable; and level 4 evidence suggests that a true association is possible.
†Supporting level 1 evidence was obtained from both patients and volunteers.
‡In a small number of volunteers, an inhibitory drug interaction occurred.

Thrombolytic Agents

Pharmacologic thrombolysis is produced by plasminogen activators—including streptokinase, rt-PA, and urokinase—which convert the proenzyme plasminogen to the fibrinolytic enzyme plasmin.[68]

Streptokinase Streptokinase is a protein produced by β-hemolytic streptococci. In contrast to other plasminogen activators, streptokinase is not an enzyme and does not convert plasminogen directly to plasmin by proteolytic cleavage. Instead, streptokinase binds noncovalently to plasminogen, converting it to a plasminogen-activator complex that acts on other plasminogen molecules to generate plasmin. Streptokinase has a plasma half-life of 30 minutes.

Because streptokinase is a bacterial product, it stimulates antibody production and can prompt allergic reactions. Antistreptococcal antibodies, present in variable titers in most patients before streptokinase treatment, induce an amnestic response that makes repeated treatment with streptokinase difficult or impossible for a period of months or years after an initial course of treatment. Laboratory monitoring of streptokinase can be limited to thrombin time, which is used as a marker for an effective lytic state. If the thrombin time is not prolonged within the first few hours of commencing treatment, resistance to streptokinase resulting from a high titer of antistreptococcal antibodies should be suspected and the dose of streptokinase should be increased.

Urokinase Synthesized by endothelial and mononuclear cells, urokinase is a direct activator of plasminogen. Like streptokinase, urokinase is non–fibrin specific. It has a plasma half-life of 10 minutes.

rt-PA rt-PA, which is fibrin specific, is synthesized by endothelial cells as a single-chain polypeptide. Proteolytic cleavage converts the single-chain form into a two-chain species. Both forms are enzymatically active. rt-PA has a plasma half-life of approximately 5 minutes.

Complications of Antithrombotic Agents

Bleeding is the main complication of antithrombotic therapy.[69] With all antithrombotic agents, the risk of bleeding is influenced by the dose and by patient-related factors, the most important being recent surgery or trauma. Other patient characteristics that increase the risk of bleeding are older age, recent stroke, generalized hemostatic defect, a history of gastrointestinal hemorrhage, and serious comorbid conditions.

Bleeding is more common and more serious with thrombolytic drugs than with anticoagulants. The risk of bleeding with thrombolytic therapy is just as great with rt-PA as with streptokinase and urokinase, which, unlike rt-PA, lack fibrin specificity. With heparin, the incidence of bleeding is influenced by dosage and by means of administration, being higher with intermittent intravenous therapy than with continuous intravenous therapy.[69] Bleeding rates are similar for heparin and LMWH.[28]

Bleeding associated with coumarin anticoagulants is influenced by the intensity of anticoagulant therapy. Such bleeding is reduced to about one third if the targeted INR range is lowered from between 3.0 and 4.5 to between 2.0 and 3.0. Both heparin-induced bleeding and warfarin-induced bleeding are increased by concomitant use of aspirin, which impairs platelet function and produces gastric erosions. When the INR is less than 3.0, coumarin-associated bleeding frequently has an obvious underlying cause or is from an occult gastrointestinal or renal lesion.

Nonhemorrhagic side effects of thrombolytic therapy are limited mainly to allergic reactions to streptokinase. Nonhemorrhagic side effects of heparin include the following: (1) urticaria at sites of subcutaneous injection; (2) thrombocytopenia, which occurs in 2% to 4% of patients treated with high-dose heparin and is complicated by arterial or venous thrombosis in about 0.2% of treated patients [*see 5:XIV Thrombotic Disorders*]; (3) osteoporosis, which occurs with prolonged high-dose heparin use; and, rarely, (4) alopecia, adrenal insufficiency, and skin necrosis. The incidence of thrombocytopenia is lower with LMWHs than with heparin. Similarly, there is evidence that the risk of osteopenia is lower with LMWH than with heparin.[59]

The most important nonhemorrhagic side effect of coumarin anticoagulants is skin necrosis, an uncommon complication usually observed on the third to eighth day of therapy. Skin necrosis is caused by extensive thrombosis of the venules and capillaries within the subcutaneous fat. An association has been reported between coumarin-induced skin necrosis and protein C deficiency—and, less commonly, protein S deficiency—but this complication can occur in patients without these deficiencies.

PRIMARY PROPHYLAXIS

The most effective way of reducing the mortality associated with pulmonary embolism and the morbidity associated with the postthrombotic syndrome is to institute primary prophylaxis in patients at risk for venous thromboembolism. On the basis of well-defined clinical criteria, patients can be classified as being at low, moderate, or high risk for venous thromboembolism, and the choice of prophylaxis should be tailored to the patient's risk [*see Table 6*]. In the absence of prophylaxis, the frequency of fatal postoperative pulmonary embolism ranges from 0.1% to 0.4% in patients undergoing elective general surgery and from 1% to 5% in patients undergoing elective hip or knee surgery, emergency hip surgery, major trauma, or spinal cord injury. Prophylaxis is cost-effective for most high-risk groups.[70]

Prophylaxis is achieved either by modulating activation of blood coagulation or by preventing venous stasis by using the following proven approaches: low-dose subcutaneous heparin, intermittent pneumatic compression of the legs, coumarin anticoagulants, adjusted doses of subcutaneous heparin, graduated compression stockings, LMWHs, or fondaparinux.[64,70] Antiplatelet agents, such as aspirin, also prevent venous thromboembolism but less effectively than the previously stated methods.[70,71]

Low-dose heparin is given subcutaneously at a dose of 5,000 units 2 hours before surgery and 5,000 U every 8 or 12 hours after surgery. In patients undergoing major orthopedic surgical procedures, low-dose heparin is less effective than warfarin, adjusted-dose heparin, or LMWHs. Intermittent pneumatic compression of the legs enhances blood flow in the deep veins and increases blood fibrinolytic activity. This method of prophylaxis is free of clinically important side effects and is particularly useful in patients who have a high risk of serious bleeding. It is the method of choice for preventing venous thrombosis in patients undergoing neurosurgery, it is effective in patients undergoing major knee surgery, and it is as effective as low-dose heparin in patients undergoing abdominal surgery.

Table 6 Risk Categories for Venous Thromboembolism and Recommendations for Prophylaxis

	High Risk	*Moderate Risk*
Calf vein thrombosis	30%–50%	10%–30%
Proximal vein thrombosis	10%–20%	2%–8%
Fatal pulmonary embolism	1%–5%	0.2%–0.7%
Recommended prophylaxis	Low-molecular-weight heparin, oral anticoagulants, adjusted-dose heparin, or fondaparinux	Low-dose heparin, external pneumatic compression, or graduated compression stockings

Graduated compression stockings reduce venous stasis and prevent postoperative venous thrombosis in general surgical patients and in medical or surgical patients with neurologic disorders, including paralysis of the lower limbs.[70] In surgical patients, the combined use of graduated compression stockings and low-dose heparin is significantly more effective than use of low-dose heparin alone. Graduated compression stockings are relatively inexpensive and should be considered in all high-risk surgical patients, even if other forms of prophylaxis are used.

Moderate-dose warfarin (INR = 2.0 to 3.0) is effective for preventing postoperative venous thromboembolism in patients in all risk categories.[70] Warfarin therapy can be started preoperatively, at the time of surgery, or in the early postoperative period. Although the anticoagulant effect is not achieved until the third or fourth postoperative day, warfarin treatment started at the time of surgery or in the early postoperative period is effective in patients at very high risk, including patients with hip fractures and those who undergo joint replacement. Prophylaxis with warfarin is less convenient than that with low-dose heparin or LMWHs, however, because careful laboratory monitoring is necessary.

LMWH is a safe and effective form of prophylaxis in high-risk patients undergoing elective hip surgery, major general surgery, or major knee surgery, as well as in patients who have experienced hip fracture, spinal injury, or stroke. LMWH is more effective than standard low-dose heparin in general surgical patients, patients undergoing elective hip surgery, and patients with stroke or spinal injury.

In patients who undergo hip or major knee surgery, LMWH is more effective than warfarin but is associated with more frequent bleeding; both of these differences may be caused by a more rapid onset of anticoagulation with postoperatively initiated LMWH than with warfarin. It is uncertain whether the superior efficacy of LMWH over warfarin in the prevention of venographically detectable venous thrombosis is mirrored by fewer symptomatic episodes of venous thromboembolism with LMWH.[70,72]

Fondaparinux was shown to reduce the frequency of venographically-detected DVT by 50% but to cause a small increase in bleeding compared with LMWH in a series of large trials in orthopedic surgical patients.[64]

Indications for Prophylaxis

General surgery and medicine Low-dose-heparin prophylaxis is the method of choice for moderate-risk general surgical and medical patients. It reduces the risk of venous thromboembolism by 50% to 70% and is simple, inexpensive, convenient, and safe.[70] If anticoagulants are contraindicated because of an unusually high risk of bleeding, graduated compression stockings, intermittent pneumatic compression of the legs, or both should be used.

Major orthopedic surgery LMWH, fondaparinux, or oral anticoagulants provide effective prophylaxis for venous thrombosis in patients who have undergone hip surgery. Aspirin has also been shown to reduce the frequency of symptomatic venous thromboembolism and fatal pulmonary embolism after hip fracture.[71] The relative efficacy and safety of aspirin versus LMWH, fondaparinux, or oral anticoagulants in patients who have a hip fracture or have undergone hip or knee arthroplasty is uncertain. However, because studies have shown that aspirin is much less effective than LMWH or oral anticoagulants at preventing venographically detectable venous thrombosis, aspirin is not recommended as the sole agent for postoperative prophylaxis.[70]

LMWH, warfarin, fondaparinux, and intermittent pneumatic compression are effective in preventing venous thrombosis in patients undergoing major knee surgery.

Extended prophylaxis with LWMH or warfarin for an additional 3 weeks after hospital discharge should be considered after major orthopedic surgery. Extended prophylaxis is strongly recommended for high-risk patients (e.g., those with previous venous thromboembolism or active cancer).[70,73]

Genitourinary surgery, neurosurgery, and ocular surgery Intermittent pneumatic compression, with or without graduated compression stockings, is effective prophylaxis for venous thrombosis and does not increase the risk of bleeding.

TREATMENT

The objectives of treating patients with venous thromboembolism are to prevent pulmonary embolism, the postthrombotic syndrome, thromboembolic pulmonary hypertension, and recurrent venous thromboembolism and to alleviate the discomfort of the acute event.

Superficial venous thrombosis usually can be treated conservatively with anti-inflammatory drugs. In patients with DVT, anticoagulants can effectively reduce morbidity and mortality from pulmonary embolism.[7] Vena caval interruption, which is usually achieved with an inferior vena caval filter, is also effective but is more complicated, expensive, and invasive and is associated with a doubling of the frequency of recurrent DVT during long-term follow-up.[74] For these reasons, it is in general used only if anticoagulant therapy has failed or is contraindicated because of the risk of serious hemorrhage.[7]

Thrombolytic therapy with streptokinase, urokinase, or rt-PA is more effective than heparin in achieving early lysis of venous thromboembolism and is better than heparin for preventing death in patients with massive pulmonary embolism associated with shock.[75] Thrombolytic therapy is therefore the treatment of choice for patients with life-threatening pulmonary embolism.

Thromboendarterectomy is effective treatment in selected cases of chronic thromboembolic pulmonary hypertension involving proximal pulmonary arterial obstruction.[76] Urgent pulmonary embolectomy is rarely indicated.

In one study, the routine early use of graduated compression

stockings for 2 years reduced the incidence of the postthrombotic syndrome by about 50%.[9] Consequently, use of graduated compression stockings is recommended for 2 years after proximal DVT, particularly if thrombosis was extensive or associated with marked leg swelling, and if the patient finds the stockings comfortable.

Administration and Dosage Guidelines

Anticoagulant therapy Anticoagulants are the mainstay of treatment for most patients with venous thromboembolism. In the past, the treatment of choice was heparin administered by continuous intravenous infusion or subcutaneous injection, in doses sufficient to produce an adequate anticoagulant response. It is now believed that LMWH administered by subcutaneous injection without laboratory monitoring is as effective and safe as heparin.[28]

The anticoagulant effect of intravenous heparin or LMWH is immediate. With subcutaneous injection, the anticoagulant effect of both anticoagulants is delayed for about an hour; peak levels occur at 2 to 3 hours. The anticoagulant effect of subcutaneous heparin is maintained for about 12 hours with therapeutic doses. LMWH is effective when administered subcutaneously once daily.[28]

Heparin therapy is usually monitored by the aPTT and less frequently by heparin assays, which measure the ability of heparin to accelerate the inactivation of factor Xa or thrombin by antithrombin. The anticoagulant effect should be monitored carefully (e.g., every 6 hours until the aPTT is in the therapeutic range, and then daily). The dosage should be adjusted as necessary to achieve an adequate anticoagulant effect because there is a greater risk of recurrent venous thromboembolism if the anticoagulant effect is suboptimal.[7] The therapeutic range of aPTT is equivalent to a heparin level between 0.35 and 7.0 U/ml as measured by an anti–factor Xa assay.[77] For many aPTT reagents, this range is an aPTT ratio of 1.8 to 2.5 times the mean of the normal laboratory control value.[59,77]

LMWH is administered subcutaneously on a weight-adjusted basis at a dosage of either 100 anti-Xa U/kg every 12 hours or 150 to 200 anti-Xa units once daily.[60] Monitoring is not required.

Treatment with heparin or LMWH is usually continued for 5 to 6 days; warfarin therapy is started on the first or second day, overlapping the heparin therapy (or LMWH) for 4 or 5 days, and is continued until an INR of 2.0 is maintained for at least 24 hours.[59] For patients with major pulmonary embolism or extensive DVT, heparin should be given for at least 7 days. A 4- to 5-day period of overlap is necessary because the antithrombotic effects of oral anticoagulants are delayed. The initial course of heparin should be followed by warfarin for at least 3 months.[34] Less intense warfarin therapy (INR = 2.0 to 3.0) is just as effective as the high-intensity regimen (INR = 3.0 to 4.5) and produces significantly less bleeding.[65] Adjusted-dose subcutaneous heparin or intermediate-dose LMWH can also be used in the outpatient setting,[34] but they are more expensive and less convenient than warfarin.

Duration of anticoagulant therapy During the past decade, a series of well-designed studies has helped to define the optimal duration of anticoagulation. The findings of these studies can be summarized as follows:

- Shortening the duration of anticoagulation from 3 or 6 months to 4 or 6 weeks results in a doubling of the frequency of the recurrence of venous thromboembolism during 1 to 2 years of follow-up.[29,31,32]
- Patients with venous thromboembolism that was provoked by a transient risk factor have a risk of recurrence about one third lower than those with unprovoked venous thromboembolism or a persistent risk factor.[8,29,31-34]
- Three months of anticoagulation may not be adequate treatment for an unprovoked (so-called idiopathic) episode of venous thromboembolism; subsequent risk of early recurrence has varied from 5% to 25% per patient-year.[8,29,31-34,78]
- After 6 months of anticoagulation, recurrent DVT is at least as likely to affect the contralateral leg, which suggests that these recurrences result from systemic rather than local factors (including inadequate treatment).[79]
- There is a persistently elevated risk of recurrent venous thromboembolism after a first episode; this risk appears to be 5% to 12% per year after 6 or more months of treatment of an unprovoked episode.[8,30,31,78]
- Oral anticoagulants targeted at an INR of about 2.5 are very effective (risk reduction of 90% or greater) at preventing recurrent unprovoked venous thromboembolism after the first 3 months of treatment.[16,80]
- Time-limited extensions of anticoagulant therapy beyond 3 or 6 months (e.g., to 12 or 24 months) may delay but not ultimately reduce the risk of recurrent venous thromboembolism.[30,34,78]
- Indefinite anticoagulation (i.e., anticoagulation therapy that is not time-limited) is an option for patients who have experienced their first unprovoked venous thromboembolism and who are at low risk for bleeding.[34]
- A second episode of venous thromboembolism does not necessarily indicate a high risk of recurrence or the need for indefinite anticoagulation.[34]
- Risk of bleeding during anticoagulation therapy differs markedly among patients, depending on the prevalence of risk factors (e.g., advanced age, previous bleeding or stroke, renal failure, anemia, antiplatelet therapy, malignancy, or poor anticoagulant control).[34,69]
- Risk of recurrence is lower (about half) after an isolated calf (distal) DVT; in such patients, a shorter duration of treatment is appropriate.[30,31]
- Risk of recurrence is higher with antiphospholipid antibodies (anticardiolipin antibodies or lupus anticoagulants), homozygous factor V Leiden, cancer, and, probably, antithrombin deficiency; the presence of any of these factors argues for a longer duration of treatment.[15,34]
- Heterozygous factor V Leiden and the G20210A prothrombin gene mutations do not appear to be clinically important risk factors for recurrence.[34]
- Other clotting abnormalities (e.g., elevated levels of clotting factors VIII, IX, and XI and homocysteine, and deficiencies of protein C and protein S) may be risk factors for recurrence; they have uncertain implications for duration of treatment.[34]
- Recurrent venous thromboembolism after an initial pulmonary embolism tends to be in the form of another pulmonary embolism (about 60% of cases), whereas recurrent venous thromboembolism after an initial DVT tends to be another DVT (about 80% of cases).[81] Consequently, the chance of dying of recurrent venous thromboembolism after anticoagulants are stopped appears to be at least twice as high after a pulmonary embolism as after a DVT; the difference in prognosis may favor indefinite treatment of selected patients after a first unprovoked pulmonary embolism.[34]

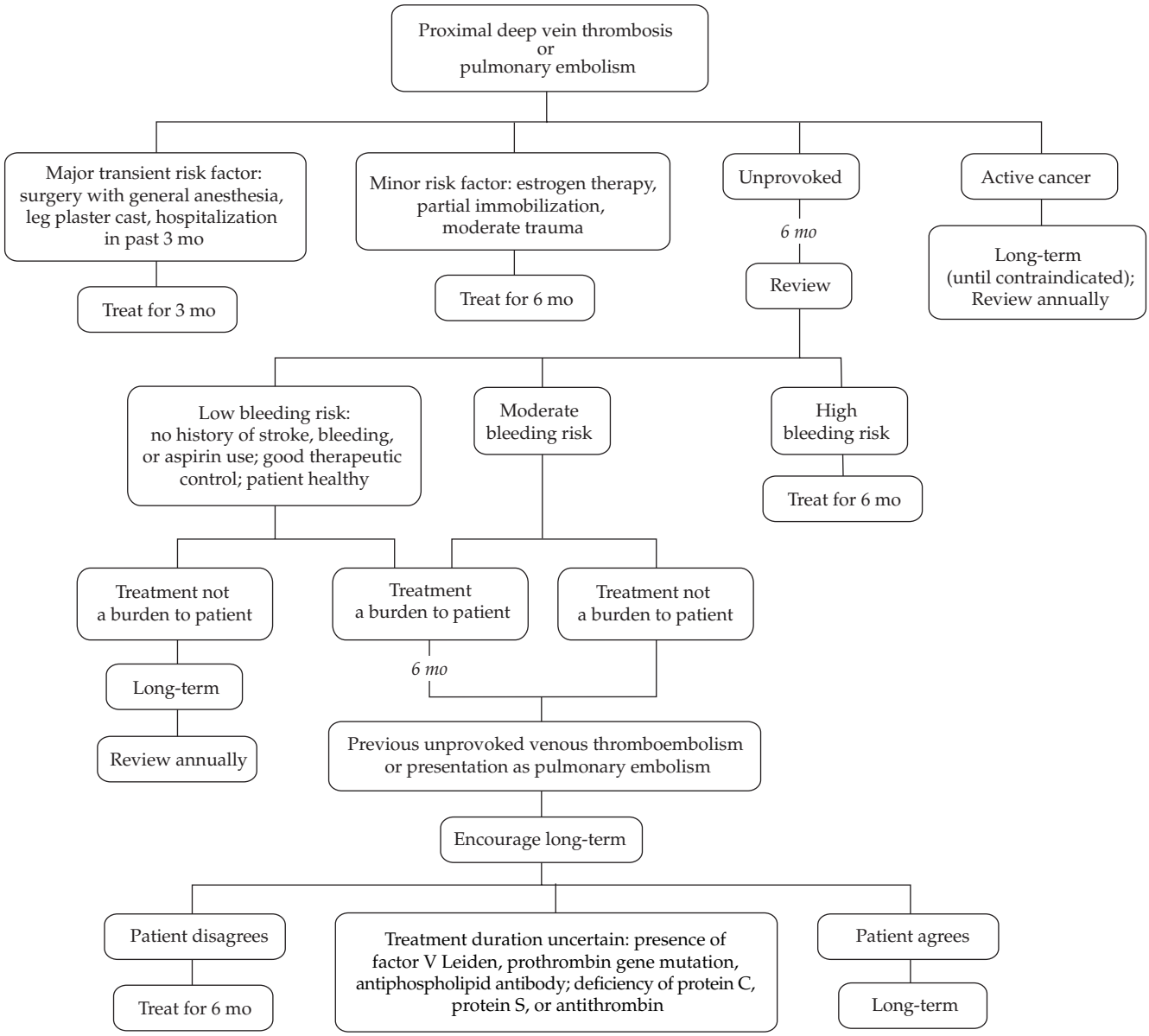


Figure 5 **Algorithm for selecting the duration of anticoagulation for venous thromboembolism.**

These findings permit the construction of an algorithm for selecting duration of anticoagulation for venous thromboembolism [*see Figure 5*]. Whether anticoagulant therapy (INR = 2.0 to 3.0) is recommended for 3 months, 6 months, or an indefinite period (with annual review) depends primarily on the presence of a provoking risk factor for venous thromboembolism (i.e., major or minor transient risk factor, no risk factor, or cancer), risk factors for bleeding, and patient preference (i.e., burden associated with treatment). Secondary factors are whether the patient has had a previous unprovoked venous thromboembolism, whether the venous thromboembolism presented as DVT or as pulmonary embolism, and whether the patient has biochemical risk factors for recurrent venous thromboembolism.

Thrombolytic therapy Thrombolytic therapy produces complete early lysis of acute venous thrombi in 30% to 40% of cases and partial lysis in an additional 30%; in contrast, complete early lysis of venous thrombi occurs in fewer than 10% of patients treated with heparin.[7,25,82] The risk of major bleeding, however, is about three times greater with thrombolytic therapy than with heparin.[25] The risk of hemorrhage increases with the duration of thrombolytic infusion. Hemorrhage often occurs at a site of previous surgery or trauma. Intracranial hemorrhage occurs in 1% to 2% of patients with pulmonary embolism who are treated with thrombolytic agents, which is a rate about five to 10 times higher than that seen in patients with pulmonary embolism who are treated with heparin.[83] Some evidence suggests that thrombolytic therapy with streptokinase reduces the incidence of postthrombotic syndrome,[25,82] but data from properly designed trials are lacking. The potential role of thrombolytic therapy in preventing late sequelae of pulmonary embolism is unknown.

Indications for Treatment

Anticoagulant therapy Most patients with proximal vein thrombosis, calf vein thrombosis, or symptomatic pulmonary

embolism should be treated first with high-dose heparin or LMWH and then with moderate-intensity oral anticoagulant therapy (INR = 2.0 to 3.0) for at least 3 to 6 months (see above).

Long-term anticoagulant therapy should be considered for patients with malignancy; recurrent unprovoked episodes of venous thromboembolism; unprovoked venous thromboembolism and homozygous factor V Leiden; deficiency of protein C, protein S, or antithrombin; an antiphospholipid antibody; or pulmonary embolism.[34] Risk of bleeding and patient preference also strongly influence the decision to use long-term therapy [*see Figure 5*].

Thrombolytic therapy Thrombolytic therapy is indicated in patients who have major pulmonary embolism with hemodynamic compromise (see above). A regimen of 100 mg of rt-PA administered over 2 hours is probably the method of choice because this regimen produces greater lysis at 2 hours than a 24-hour course of conventional urokinase[84] and, compared with heparin, produces more rapid clinical improvement in pulmonary vascular resistance and right ventricular function.[85,86]

The use of thrombolytic therapy in patients who have venous thrombosis is more controversial. Although there is no rigorous supporting evidence for it and it is not part of our clinical practice, regional or systemic thrombolysis may be considered in patients who have large proximal venous thrombi—particularly if the thrombi are confined to the iliac and femoral veins—provided that there are no contraindications.

Absolute contraindications to thrombolytic therapy include active internal bleeding, stroke within the past 3 months, and intracranial disease. Relative contraindications include major surgery within the past 10 days, recent organ biopsy, recent puncture of a noncompressible vessel, recent gastrointestinal bleeding, liver or renal disease, severe arterial hypertension, and severe diabetic retinopathy.

Surgical treatment Pulmonary endarterectomy may be beneficial in selected patients with thromboembolic pulmonary hypertension, which is estimated to occur in about 5% of patients who have undergone treatment of pulmonary embolism.[76] Urgent pulmonary embolectomy is reserved for patients with a saddle embolism lodged in the main pulmonary artery, those with massive embolism whose blood pressure cannot be maintained despite administration of thrombolytic therapy and vasopressor agents, or those in whom there is an absolute contraindication to thrombolytic therapy.[7]

Venous Thromboembolism in Pregnancy

The management of venous thromboembolism during pregnancy is complicated because clinical diagnosis is unreliable, some of the objective diagnostic tests are potentially risky to the fetus, and treatment may cause teratogenicity or fetal bleeding.[87,88]

DIAGNOSIS

In pregnant patients suspected of having venous thrombosis, venous ultrasonography should be used as the initial test.[6,88] If the result is abnormal, a diagnosis of proximal DVT is made and the patient is treated with anticoagulants. If venous ultrasound results are normal, we employ IPG or color Doppler to exclude an isolated iliac vein thrombosis. If both tests are normal, either a limited venogram can be performed to exclude isolated calf vein thrombosis, or serial compression ultrasonography can be performed on two occasions over the next 14 days.[6] MRI may also be considered.

The diagnostic approach to pulmonary embolism in pregnancy is similar to that used in nonpregnant patients. Lung scanning and pulmonary angiography can be performed, but the techniques should be modified to reduce exposure of the fetus to radiation.[88] Although there is little radiation exposure from ventilation-perfusion scanning, exposure can be reduced further without a serious loss of resolution by administering 50% of the standard dose of radioactive particles for perfusion lung scanning and by limiting ventilation scanning to patients with an abnormal perfusion scan. Fetal radiation exposure from pulmonary angiography can be reduced by using the brachial route for contrast injection and by shielding the abdomen with a lead-lined apron.

TREATMENT

The treatment of venous thromboembolism is much more complicated in pregnant patients because oral anticoagulants cross the placenta and, if administered during the first trimester, can cause warfarin embryopathy, which is characterized by nasal hypoplasia and skeletal abnormalities.[87,88] Warfarin administered during the second and third trimesters can cause dorsal midline dysplasia, abnormalities of the ventricular system, and optic atrophy.

Heparin does not cross the placenta and is much safer than oral anticoagulants during pregnancy. Although there have been reports associating heparin therapy during pregnancy with a high incidence of stillbirth or prematurity, most of these complications occurred in mothers receiving heparin for disorders that are known to be associated with a high rate of fetal loss. Other studies have shown that heparin is safe for the fetus but, when used on a long-term basis during pregnancy, can produce osteoporosis in the mother. The incidence of heparin-induced osteopenia diagnosed by dual-photon absorption x-ray or by conventional x-ray may be as high as 30%, but overt fractures are uncommon, occurring in fewer than 5% of patients. Heparin-induced bleeding is not a common problem during pregnancy, provided that heparin therapy is monitored carefully. The anticoagulant response to heparin can be prolonged if the drug is administered in high doses just before parturition, so there is the potential for local bleeding during and immediately after delivery.[87,88]

In pregnant patients with acute venous thromboembolism, continuous intravenous heparin or twice-daily LMWH should be administered for 4 to 7 days, followed by subcutaneous heparin or LMWH given in adjusted therapeutic doses for the remainder of the pregnancy.[89] The injection site should be rotated over the fatty tissue of the lower abdomen and thighs; the site should be compressed for 5 minutes after injection to prevent local bruising. An unwanted anticoagulant effect during delivery can be avoided by discontinuing subcutaneous heparin therapy 24 hours before elective induction of labor.[87,88]

If there is no evidence of excessive postpartum bleeding, heparin therapy can be resumed within 12 hours of delivery and continued until oral anticoagulation is established. The intensity of heparin therapy will depend on the amount of time that has passed since the diagnosis of venous thromboembolism was made: if the diagnosis was made less than 1 month ago, therapeutic doses may be used (with stepwise increases in subcutaneous or intravenous doses over 24 hours); if the diagnosis was made more than 1 month ago, prophylactic or intermediate doses of heparin may be used. Warfarin is started at the

same time as heparin and is continued for a minimum of 6 weeks and preferably until patients have received a minimum of 3 months of anticoagulation. Warfarin does not enter breast milk and therefore can be administered to nursing mothers.[87,88]

Miscellaneous Thromboembolic Disorders

THROMBOSIS IN UNUSUAL SITES

Subclavian or Axillary Veins

Thrombosis of the subclavian or axillary veins may be idiopathic or may occur as a complication of local vascular damage.[90] It is now most frequently seen as a complication of chronic indwelling catheter use, but it also occurs as a complication after mastectomy and local radiotherapy for breast cancer. Idiopathic subclavian or axillary vein thrombosis often occurs in young muscular individuals and may be preceded by repetitive, strenuous activity involving the affected arm. Some of these persons have a fixed stenosis of the subclavian vein that is thought to be caused by external compression of the vein as it courses behind the clavicle. Occasionally, subclavian or axillary vein thrombosis can occur in patients with congenital deficiency of antithrombin, protein C, or protein S or in patients with antiphospholipid antibodies. Thrombosis of the axillary or subclavian vein or the superior vena cava is a rare complication of an implantable perivenous endocardial pacing system.

Subclavian or axillary thrombosis causes pain, edema, and cyanosis of the arm. In rare cases, the thrombosis extends into the superior vena cava and causes edema and cyanosis of the face and neck. Definitive diagnosis is made by venography or venous ultrasonography.[91] Subclavian or axillary vein thrombosis is usually treated with anticoagulants. Regional or systemic thrombolytic therapy may be considered in young patients without contraindications, because a substantial number of these patients experience aching and swelling when they exert the affected arm.

Mesenteric Vein

An uncommon disorder, mesenteric vein thrombosis usually occurs in the sixth or seventh decade of life. It generally involves segments of the small bowel, leading to hemorrhagic infarction. Affected patients often have associated disorders, such as inflammatory bowel disease, malignancy, portal hypertension, or familial thrombophilia or polycythemia vera, or they may have a history of recent abdominal surgery. In about 20% of cases, no underlying cause is found.

The clinical manifestations of mesenteric vein thrombosis include intermittent abdominal pain, abdominal distention, vomiting, diarrhea, and melena. Diagnosis of mesenteric vein thrombosis is often difficult, but the finding of blood-stained ascitic fluid on abdominal paracentesis or peritoneoscopic evidence of hemorrhagic bowel infarction is characteristic of the disorder. Management includes supportive care and surgical resection, followed by anticoagulant therapy. Mortality is about 20%, and up to 20% of patients experience recurrence.

Renal Vein

Renal vein thrombosis can be idiopathic or a complication of the nephrotic syndrome. Patients may be asymptomatic or may present with abdominal, back, or flank pain and tenderness. Pulmonary embolism is a relatively common complication of renal vein thrombosis. Anticoagulant therapy results in a gradual improvement in renal function, but patients may have long-standing proteinuria. Thrombolytic agents have been used, but the data are inadequate for critical appraisal of this form of treatment.

THROMBOPHILIA

The term thrombophilia denotes any increased tendency to thrombosis, whether inherited or acquired. Thrombophilia is discussed in detail elsewhere [*see 5:XIV Thrombotic Disorders*].

Jack Hirsh, M.D., participates in the speakers' bureaus for AstraZeneca Pharmaceuticals LP and Sanofi-Synthelabo Inc. Clive Kearon, M.B., Ph.D., has been an advisor to Sanofi-Synthelabo Inc. during the past 12 months.

References

1. Kearon C: Epidemiology of venous thromboembolism. Semin Vascular Med 1:7, 2001

2. Heit JA, Silverstein MD, Mohr DN, et al: Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. Arch Intern Med 160:809, 2000

3. Kearon C, Salzman EW, Hirsh J: Epidemiology, pathogenesis, and natural history of venous thrombosis. Hemostasis and Thrombosis: Basic Principles and Clinical Practice, 4th ed. Colman RW, Hirsh J, Marder VJ, et al, Eds. JB Lippincott Co, Philadelphia, 2001, p 1153

4. Anderson FA Jr, Wheeler HB, Goldberg RJ, et al: The prevalence of risk factors for venous thromboembolism among hospital patients. Arch Intern Med 152:1660, 1992

5. Kearon C: Natural history of venous thromboembolism. Semin Vascular Med 1:27, 2001

6. Kearon C, Julian JA, Newman TE, et al: Noninvasive diagnosis of deep venous thrombosis. Ann Intern Med 128:663, 1998

7. Ginsberg JS: Management of venous thromboembolism. N Engl J Med 335:1816, 1996

8. Prandoni P, Lensing AWA, Cogo A, et al: The long-term clinical course of acute deep venous thrombosis. Ann Intern Med 125:1, 1996

9. Brandjes DPM, Buller HR, Heijboer H, et al: Randomised trial of effect of compression stockings in patients with symptomatic proximal-vein thrombosis. Lancet 349:759, 1997

10. Prins MH, Hirsh J: A critical review of the evidence supporting a relationship between impaired fibrinolytic activity and venous thromboembolism. Arch Intern Med 151:1721, 1991

11. Lane DA, Mannucci PM, Bauer KA, et al: Inherited thrombophilia (pt 1). Thromb Haemost 76:651, 1996

12. Lane DA, Mannucci PM, Bauer KA, et al: Inherited thrombophilia (pt 2). Thromb Haemost 76:824, 1996

13. Heijboer H, Brandjes DPM, Buller HR, et al: Deficiencies of coagulation-inhibiting and fibrinolytic proteins in outpatients with deep-vein thrombosis. N Engl J Med 323:1512, 1990

14. van den Belt AGM, Prins MH, Huisman MV, et al: Familial thrombophilia: a review analysis. Clin Appl Thromb Hemost 2:227, 1996

15. Kearon C, Crowther M, Hirsh J, et al: Management of patients with hereditary hypercoagulable disorders. Annu Rev Med 51:169, 2000

16. Kearon C, Gent M, Hirsh J, et al: A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med 340:901, 1999

17. Poort SR, Rosendaal FR, Reitsma PH, et al: A common genetic variation in the 3' untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 88:3698, 1996

18. Rosendaal FR: High levels of factor VIII and venous thrombosis. Thromb Haemost 83:1, 2000

19. Meijers JCM, Tekelenburg WLH, Bouma BN, et al: High levels of coagulation factor XI as a risk factor for venous thrombosis. N Engl J Med 342:696, 2000

20. Cattaneo M: Hyperhomocysteinemia, atherosclerosis and thrombosis. Thromb Haemost 81:165, 1999

21. Vandenbrouke JP, Koster E, Briet E, et al: Increased risk of venous thrombosis in oral contraceptive users who are carriers of factor V Leiden mutation. Lancet 344:1453, 1994

22. Grady D, Wenger NK, Herrington D, et al: Postmenopausal hormone therapy increases risk for venous thromboembolic disease. Ann Intern Med 132:689, 2000

23. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

24. Hoibraaten E, Qvigstad E, Arnesen H, et al: Increased risk of recurrent venous thromboembolism during hormone replacement therapy: results of the randomized, double-blind, placebo-controlled Estrogen in Venous Thromboembolism Trial (EVTET). Thromb Haemost 84:961, 2000

25. Hirsh J, Lensing A: Thrombolytic therapy for deep vein thrombosis. Int J Angiol 5:S22, 1996

26. Hull R, Delmore T, Genton E, et al: Warfarin sodium versus low-dose heparin in the long-term treatment of venous thrombosis. N Engl J Med 301:855, 1979

27. Lagerstedt CI, Olsson CG, Fagher BO, et al: Need for long-term anticoagulant treatment in symptomatic calf-vein thrombosis. Lancet 2:515, 1985

28. Dolovich LR, Ginsberg JS, Douketis JD, et al: A meta-analysis comparing low-molecular-weight heparins with unfractionated heparin in the treatment of venous thromboembolism. Arch Intern Med 160:181, 2000

29. Optimum duration of anticoagulation for deep-vein thrombosis and pulmonary embolism. Research Committee of the British Thoracic Society. Lancet 340:873, 1992

30. Pinede L, Ninet J, Duhaut P, et al: Comparison of 3 and 6 months of oral anticoagulant therapy after a first episode of proximal deep vein thrombosis or pulmonary embolism and comparison of 6 and 12 weeks of therapy after isolated calf deep vein thrombosis. Circulation 103:2453, 2001

31. A comparison of six weeks with six months of oral anticoagulant therapy after a first episode of venous thromboembolism. Duration of Anticoagulation Trial Study Group. N Engl J Med 332:1661, 1995

32. Levine MN, Hirsh J, Gent M, et al: Optimal duration of oral anticoagulant therapy: a randomized trial comparing four weeks with three months of wafarin in patients with proximal deep vein thrombosis. Thromb Haemost 74:606, 1995

33. Prandoni P, Lensing AW, Piccioli A, et al: Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. Blood 100:3484, 2002

34. Kearon C: Duration of therapy for acute venous thromboembolism. Clin Chest Med 24:63, 2003

35. Kyrle P, Minar E, Hirschl M, et al: High plasma levels of factor VIII and the risk of recurrent venous thromboembolism. N Engl J Med 343:62, 2000

36. Anand SS, Wells PS, Hunt D, et al: Does this patient have deep vein thrombosis? JAMA 279:1094, 1998

37. Wells PS, Hirsh J, Anderson DR, et al: A simple clinical model for the diagnosis of deep-vein thrombosis combined with impedance plethysmography: potential for an improvement in the diagnostic process. J Intern Med 243:15, 1998

38. Wells PS, Hirsh J, Anderson DR, et al: Accuracy of clinical assessment of deep vein thrombosis. Lancet 345:1326, 1995

39. Fraser DG, Moody AR, Morgan PS, et al: Diagnosis of lower-limb deep venous thrombosis: a prospective blinded study of magnetic resonance direct thrombus imaging. Ann Intern Med 136:89, 2002

40. Heijboer H, Buller HR, Lensing AWA, et al: A comparison of real-time-compression ultrasonography with impedance plethysmography for the diagnosis of deep-vein thrombosis in symptomatic outpatients. N Engl J Med 329:1365, 1993

41. Lee AY, Ginsberg JS: Laboratory diagnosis of venous thromboembolism. Baillieres Clin Haematol 11:2461, 1998

42. Perrier A, Desmarais S, Miron MJ, et al: Non-invasive diagnosis of venous thromboembolism in outpatients. Lancet 353:190, 1999

43. Ginsberg JS, Kearon C, Douketis J, et al: The use of D-dimer testing and impedance plethysmographic examination in patients with clinical indications of deep vein thrombosis. Arch Intern Med 157:1077, 1997

44. Bernardi E, Prandoni P, Lensing AWA, et al: D-dimer testing as an adjunct to ultrasonography in patients with clinically suspected deep vein thrombosis: prospective cohort study. BMJ 317:1037, 1998

45. Kearon C, Ginsberg JS, Douketis J, et al: Management of suspected deep venous thrombosis in outpatients by using clinical assessment and D-dimer testing. Ann Intern Med 135:108, 2001

46. Kearon C: Diagnosis of pulmonary embolism. CMAJ 168:183, 2003

47. Simonneau G, Sors H, Charbonnier B, et al: A comparison of low-molecular-weight heparin with unfractionated heparin for acute pulmonary embolism. N Engl J Med 337:663, 1997

48. Wells PS, Anderson DR, Rodger M, et al: Derivation of a simple clinical model to categorize patients' probability of pulmonary embolism: increasing the model's utility with the SimpliRED D-dimer. Thromb Haemost 83:416, 2000

49. Miniati M, Prediletto R, Formichi B, et al: Accuracy of clinical assessment in the diagnosis of pulmonary embolism. Am J Respir Crit Care Med 159:864, 1999

50. Wells PS, Ginsberg JS, Anderson DR, et al: Use of a clinical model for safe management of patients with suspected pulmonary embolism. Ann Intern Med 129:997, 1998

51. Rathbun SW, Raskob GE, Whitsett TL: Sensitivity and specificity of helical computed tomography in the diagnosis of pulmonary embolism: a systematic review. Ann Intern Med 132:227, 2000

52. Mullins MD, Becker DM, Hagspiel KD, et al: The role of the spiral volumetric computed tomography in the diagnosis of pulmonary embolism. Arch Intern Med 160:293, 2000

53. Perrier A, Howarth N, Didier D, et al: Performance of helical computed tomography in unselected outpatients with suspected pulmonary embolism. Ann Intern Med 135:88, 2001

54. Musset D, Parent F, Meyer G, et al: Diagnostic strategy for patients with suspected pulmonary embolism: a prospective multicentre outcome study. Lancet 360:1914, 2002

55. Kearon C, Ginsberg JS, Hirsh J: The role of venous ultrasonography in the diagnosis of suspected deep venous thrombosis and pulmonary embolism. Ann Intern Med 129:1044, 1998

56. Wells PS, Anderson DR, Rodger M, et al: Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and D-dimer. Ann Intern Med 135:98, 2001

57. Hull RD, Raskob GE, Ginsberg JS, et al: A noninvasive strategy for the treatment of patients with suspected pulmonary embolism. Arch Intern Med 154:289, 1994

58. Ginsberg JS, Wells PS, Kearon C, et al: Sensitivity and specificity of a rapid whole-blood assay for D-dimer in the diagnosis of pulmonary embolism. Ann Intern Med 129:1006, 1998

59. Hirsh J: Heparin. N Engl J Med 324:1565, 1991

60. Weitz JI: Low-molecular-weight heparins. N Engl J Med 337:688, 1997

61. Levine M, Gent M, Hirsh J, et al: A comparison of low-molecular-weight heparin administered primarily at home with unfractionated heparin administered in the hospital for proximal deep-vein thrombosis. N Engl J Med 334:677, 1996

62. Koopman MMW, Prandoni P, Piovella F, et al: Treatment of venous thrombosis with intravenous unfractionated heparin administered in the hospital as compared with subcutaneous low-molecular-weight heparin administered at home. N Engl J Med 334:682, 1996

63. Low-molecular-weight heparin in the treatment of patients with venous thromboembolism. The Columbus Investigators. N Engl J Med 337:657, 1997

64. Turpie AG, Bauer KA, Eriksson BI, et al: Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies. Arch Intern Med 162:1833, 2002

65. Hirsh J, Dalen J, Anderson DR, et al: Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. Chest 119:8S, 2001

66. Harrison L, Johnston M, Massicotte MP, et al: Comparison of 5-mg and 10-mg loading doses in initiation of warfarin therapy. Ann Intern Med 126:133, 1997

67. Weitz JI, Crowther M: Direct thrombin inhibitors. Thromb Res 106:V275, 2002

68. Weitz JI, Stewart RJ, Fredenburgh JC: Mechanism of action of plasminogen activators. Thromb Haemost 82:974, 1999

69. Levine M, Raskob G, Landerfeld S, et al: Hemorrhagic complications of antithrombotic therapy. Chest 119(suppl):1085, 2001

70. Geerts WH, Heit JA, Clagett P, et al: Prevention of venous thromboembolism. Chest 119(suppl):1325, 2001

71. Prevention of pulmonary embolism and deep vein thrombosis with low dose aspirin. Pulmonary Embolism Prevention (PEP) Trial Collaborative Group. Lancet 355:1295, 2000

72. Colwell CW, Collis DK, Paulson R, et al: Comparison of enoxaparin and warfarin for the prevention of venous thromboembolic disease after total hip arthroplasty. J Bone Joint Surg Am 81:932, 1999

73. Eikelboom JW, Quinlan DJ, Douketis JD: Extended-duration prophylaxis against venous thromboembolism after total hip or knee replacement: a meta-analysis of the randomized trials. Lancet 358:9, 2001

74. Decousus H, Leizorovicz A, Parent F, et al: A clinical trial of vena caval filters in the prevention of pulmonary embolism in patients with proximal deep-vein thrombosis. N Engl J Med 338:409, 1998

75. Jerjes-Sanchez C, Ramírez-Rivera A, de Lourdes García M, et al: Streptokinase and heparin versus heparin alone in massive pulmonary embolism: a randomized controlled trial. J Thromb Thrombolysis 2:227, 1995

76. Fedullo PF, Auger WR, Kerr KM, et al: Chronic thromboembolic pulmonary hypertension. N Engl J Med 345:1465, 2001

77. Brill-Edwards P, Ginsberg JS, Johnston M, et al: Establishing a therapeutic range for heparin therapy. Ann Intern Med 119:104, 1993

78. Agnelli G, Prandoni P, Santamaria MG, et al: Three months versus one year of oral anticoagulant therapy for idiopathic deep vein thrombosis. N Engl J Med 345:165, 2001

79. The risk of ipsilateral versus contralateral recurrent deep vein thrombosis in the leg. DURAC Trial Study Group. J Intern Med 247:601, 2000

80. Schulman S, Granqvist S, Holmstrom M, et al: The duration of oral anticoagulant therapy after a second episode of venous thromboembolism. N Engl J Med 336:393, 1997

81. Murin S, Romano PS, White RH: Comparison of outcomes after hospitalization for deep vein thrombosis or pulmonary embolism. Thromb Haemost 88:407, 2002

82. Schweitzer J, Kirch W, Koch R, et al: Short- and long-term results after thrombolytic treatment of deep vein thrombosis. J Am Coll Cardiol 36:1336, 2000

83. Dalen JE, Alpert JS, Hirsh J: Thrombolytic therapy for pulmonary embolism: Is it effective? Is it safe? When is it indicated? Arch Intern Med 157:2550, 1997

84. Goldhaber SZ, Kessler CM, Heit J, et al: Randomised controlled trial of recombinant tissue plasminogen activator versus urokinase in the treatment of acute pulmonary embolism. Lancet 2:293, 1988

85. Konstantinides S, Geibel A, Heusel G, et al: Heparin plus alteplase compared with heparin alone in patients with submassive pulmonary embolism. N Engl J Med 347:1143, 2002

86. Goldhaber SZ, Haire WD, Feldstein ML, et al: Alteplase versus heparin in acute pulmonary embolism: randomised trial assessing right-ventricular function and pulmonary perfusion. Lancet 341:507, 1993

87. Ginsberg JS, Greer I, Hirsh J: Use of antithrombotic agents during pregnancy. Chest 119:122S, 2001

88. Bates SM, Ginsberg JS: How we manage venous thromboembolism during pregnancy. Blood 100:3470, 2002

89. Sanson BJ, Lensing AWA, Prins MH, et al: Safety of low-molecular-weight heparin in pregnancy: a systematic review. Thromb Haemost 81:668, 1999

90. Pradoni P, Polistena P, Bernardi E, et al: Upper-extremity deep vein thrombosis: risk factors, diagnosis, and complications. Arch Intern Med 157:57, 1997

91. Baarslag HJ, van Beek EJ, Koopman MM, et al: Prospective study of color duplex ultrasonography compared with contrast venography in patients suspected of having deep venous thrombosis of the upper extremities. Ann Intern Med 136:865, 2002

92. Wells PS, Holbrook AM, Crowther NR, et al: Interactions of warfarin with drugs and food. Ann Intern Med 121:676, 1994

I CUTANEOUS MANIFESTATIONS OF SYSTEMIC DISEASES

Mark Lebwohl, m.d.

The cutaneous manifestations of systemic diseases are so numerous and varied that a single chapter could not cover them all, even in a cursory way. Instead, this chapter reviews key cutaneous manifestations of systemic diseases that should be recognized by most physicians, and it highlights recent developments in the diagnosis and management of such disorders. For fuller discussions of specific diseases, including their cutaneous manifestations, readers are referred to the chapters devoted to these conditions.

In many of the disorders presented in this chapter, workup and therapy of the underlying systemic condition are essential to a favorable outcome. A finding of cutaneous sarcoidosis, for example, should prompt a search for systemic sarcoidosis. In other conditions—for example, recessive dystrophic epidermolysis bullosa—treatment of the skin disorder is key to the management of the systemic disease.

Cardiopulmonary and Vascular Diseases

SARCOIDOSIS

The cutaneous manifestations of sarcoidosis are as varied as its systemic manifestations [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. Papules around the eyes or nose are most characteristic. The term lupus pernio refers to noncaseating granulomas that result in translucent, violaceous plaques of the ears, cheeks, and nose [*see Figure 1*]. Involvement of underlying bone can occur. Diagnosis is made by skin biopsy. Treatment with intralesional corticosteroids is traditional, and oral antimalarials and methotrexate have been used with success. More recently, infliximab has been successfully used for the treatment of sarcoidosis.[1] Other tumor necrosis factor–α (TNF-α) blockers such as etanercept have been used to successfully treat arthritis and skin lesions associated with sarcoidosis.[2] Infliximab has also been used to treat lupus pernio.[3]

There has been an increase in the use of interferon for the treatment of hepatitis C and multiple sclerosis, and a number of reports of sarcoidosis have been attributed to this treatment. Infliximab therapy has been found to effect a response in these cases.[4]

In some patients with sarcoidosis, erythema nodosum, characterized by deep, tender erythematous nodules, occurs on the lower extremities. Lupus pernio is associated with a more chronic course of sarcoidosis, whereas erythema nodosum indicates a more acute and benign disease.[5]

GRANULOMATOUS VASCULITIS

Wegener Granulomatosis

Wegener granulomatosis is associated with both distinctive and nonspecific mucocutaneous signs. Palpable purpura is one of the most common skin findings, but ulcers, papules, nodules, and bullae have also been described. In addition to upper and lower pulmonary symptoms [*see 14:IV Focal and Multifocal Lung Disease*], saddle-nose deformity, nasal ulcerations, and septal perforation should suggest the diagnosis of Wegener granulomatosis. Definitive diagnosis is made by demonstrating a necrotizing granulomatous vasculitis in a patient with upper and lower respiratory tract disease and glomerulonephritis. Cytoplasmic antineutrophil cytoplasmic autoantibodies (c-ANCA) are often present. Standard therapy is with cyclophosphamide and corticosteroids. TNF-α blockers have proved to be effective for some patients with refractory Wegener granulomatosis.[6]

Lymphomatoid Granulomatosis

Lymphomatoid granulomatosis is a rare, destructive, angiocentric disorder that results from Epstein-Barr virus–associated B cell lymphoproliferative disease.[7] This condition can be associated with skin lesions. Typically, patients develop erythematous papules or nodules that may or may not ulcerate.[8] This disorder is clinically distinguishable from Wegener granulomatosis by the absence of upper respiratory tract involvement. Diagnosis is established by demonstrating a granulomatous necrotizing infiltrate with atypical lymphoid cells around blood vessels. Lymphomatoid granulomatosis is usually fatal; however, rituximab has been used successfully to treat this condition.[9]

Churg-Strauss Syndrome

Churg-Strauss syndrome, or allergic granulomatous angiitis, most commonly presents as asthma and eosinophilia; however, related skin lesions develop in up to 40% of patients. Symmetrical, palpable purpura and petechiae of the lower extremities are the most common findings; these lesions show a leukocytoclastic vasculitis on skin biopsy. Cutaneous nodules caused by extravascular necrotizing granulomas and papules of the elbows also occur.[10] One of the clues to the diagnosis of this disorder is the presence of perinuclear antineutrophil cytoplasmic antibodies (p-ANCA).[11]

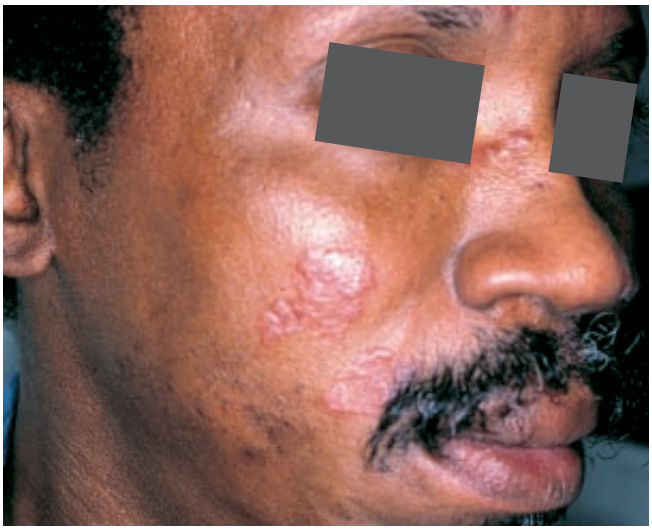


Figure 1 **Characteristic facial lesions of sarcoidosis, called lupus pernio, are shown.**

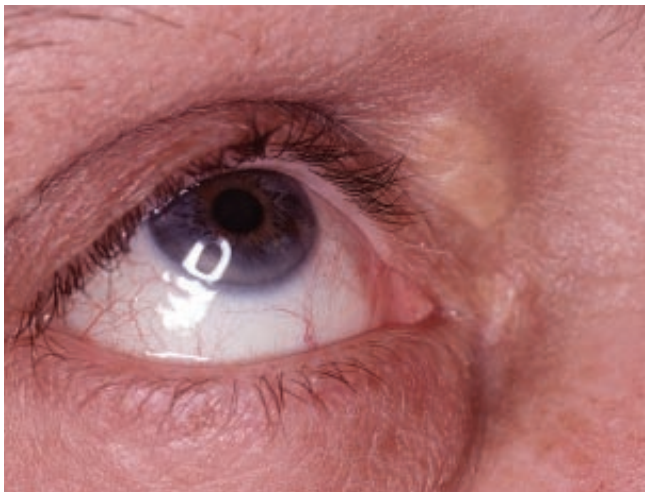


Figure 2 **Xanthelasma and arcus senilis are shown in a patient with hypercholesterolemia.**

HYPERLIPOPROTEINEMIA

Xanthomas are cutaneous manifestations of hyperlipoproteinemias. Several types of xanthomas occur with different lipid abnormalities. Xanthelasmas of the eyelids [*see Figure 2*] are the most common manifestations of familial hypercholesterolemia; however, in at least half the people who have eyelid lesions, plasma lipid levels are normal. Planar xanthomas are flat yellow plaques that can involve the palms, soles, neck, and chest. They can occur in patients with primary biliary cirrhosis or multiple myeloma. Tuberous xanthomas are large yellow or red nodules that appear on the extensor surfaces of joints, such as on the elbows and hands, but are not attached to underlying tendons. They can occur in patients with elevated triglyceride or cholesterol levels. In contrast, tendinous xanthomas, which can appear in patients with familial hypercholesterolemia, are fixed to underlying tendons of the elbows, ankles, knees, and hands. Eruptive xanthomas occur when plasma triglyceride levels suddenly become elevated. Skin lesions consist of small yellow papules that often resolve with lowering of triglyceride levels.

KAWASAKI DISEASE

Kawasaki disease, also called mucocutaneous lymph node syndrome [*see 15:VIII Systemic Vasculitis Syndromes*], is a disorder in children that can be complicated by coronary artery occlusion and myocardial infarction, coronary artery aneurysms, ECG abnormalities, cardiac arrhythmias, or myocarditis.[12] It has been suggested that a toxin secreted by *Staphylococcus aureus* is responsible for this disease, but proof of the precise cause remains elusive.[13] Diagnosis is based on clinical criteria that include fever, conjunctivitis, lymphadenopathy, and rash. In addition to a generalized erythematous eruption, abnormalities of the oral mucosa, as well as swelling and erythema of the hands and feet, may develop. Striking desquamation of the palms and soles ultimately occurs. Perianal and scrotal erythema and scaling are common as well. Thrombocytosis is a late finding, with platelet counts increasing to more than one million over 2 weeks after the onset of the disease. Approximately 15% to 25% of untreated children develop coronary artery aneurysms that may lead to sudden death.[14] Treatment with intravenous immunoglobulin reduces the frequency of coronary artery abnormalities.[15]

PSEUDOXANTHOMA ELASTICUM

Pseudoxanthoma elasticum (PXE) is an autosomal recessively inherited disorder of elastic tissue caused by mutations in the ABCC6 transporter protein.[16] PXE is associated with a wide array of systemic manifestations. Angioid streaks, the ocular hallmark of PXE, are breaks in the Bruch membrane. Retinal bleeding and vision loss commonly occur. Calcification of the internal elastic laminae of arteries can result in bleeding or occlusion of these vessels. As a result, patients develop intermittent claudication on walking and occlusive coronary artery disease at an early age. Cardiac valvular abnormalities have also been described.[17] Skin lesions consist of yellow xanthomalike macules, papules, or redundant folds of skin in flexural areas, particularly the neck and axillae [*see Figure 3*]. Some patients may have systemic manifestations of PXE without clinically apparent skin lesions.[18] Diagnosis is established by biopsy of scar or normal-appearing flexural skin.[19] There is no therapy for the skin lesions associated with PXE.

RHEUMATIC FEVER

The two cutaneous manifestations of rheumatic fever are erythema marginatum and subcutaneous nodules. Erythema marginatum is a transient faint annular erythematous rash that often develops over joints [*see Figure 4*]. The subcutaneous nodules that appear with rheumatic fever are nontender, freely movable nodules measuring approximately 1 cm in diameter; they occur on the extensor surfaces of elbows, hands, or feet.

YELLOW NAIL SYNDROME

Yellow nail syndrome is caused by an abnormality of lymphatics [*see Figure 5*]. Affected patients develop lymphedema, usually of the legs, and pleural effusions. Pulmonary symptoms such as recurrent bronchitis are also common. Diagnosis is made by finding evidence of abnormal lymphatic function associated with yellow nails without other causes of nail pathology. Increased microvascular permeability with leakage of proteins may play a role in the development of the yellow nail syndrome.[20]

Endocrinologic Diseases

DIABETES MELLITUS

There are numerous cutaneous manifestations of diabetes mellitus [*see 9:VI Diabetes Mellitus*]. Acanthosis nigricans can occur in patients with diabetes and other endocrinopathies, such as Cushing syndrome, acromegaly, polycystic ovary syndrome, and thyroid disease. Insulin resistance is an underlying factor in several of the aforementioned endocrinopathies; it also may play a role in the development of acanthosis nigricans. Skin lesions consist of brown velvety patches in intertriginous areas, especially the neck and axillae [*see Figure 6*], and occur more commonly in obese patients with diabetes.[21] Acanthosis nigricans has also been associated with internal malignancies, particularly gastric adenocarcinoma or other gastrointestinal adenocarcinomas.

Necrobiosis lipoidica is a specific cutaneous manifestation of diabetes. Lesions consist of chronic atrophic patches with enlarging erythematous borders. The legs are most commonly affected. The centers of the lesions appear yellow because of subcutaneous fat that is visible through the atrophic dermis and epidermis. Occasionally, the lesions ulcerate. Necrobiosis lipoidica is often associated with diabetic nephropathy or retinopathy.[22]

Scleredema, another manifestation of diabetes, consists of induration of the skin of the back and posterior neck in obese pa-

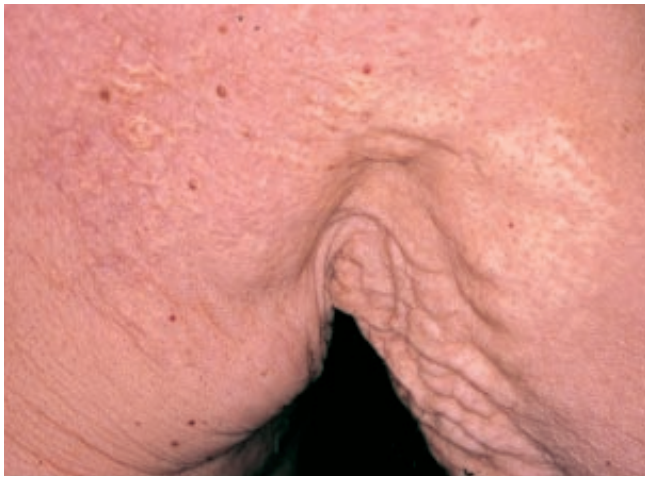


Figure 3 **Xanthomalike papules are characteristic of pseudoxanthoma elasticum. The neck and axillae are the most common sites of involvement.**

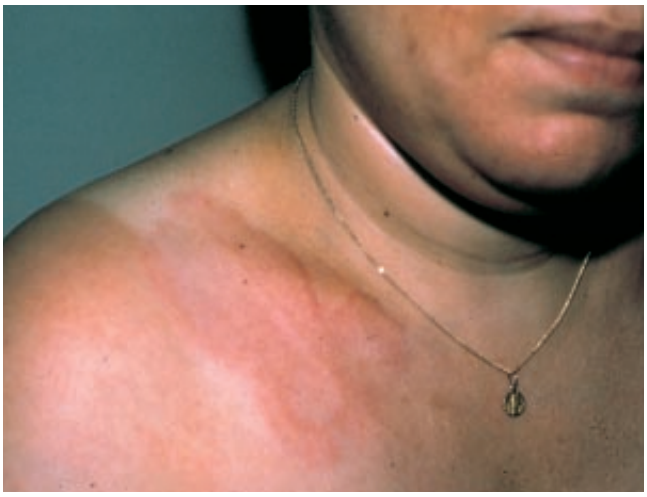


Figure 4 **Transient annular erythematous rashes (erythema marginatum) typically occur in patients with rheumatic fever.**

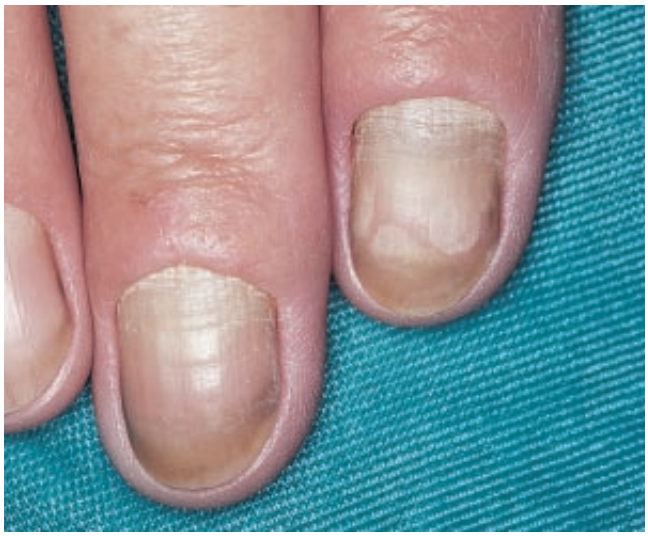


Figure 5 **Yellow nails are a sign of underlying disease of the lymphatics in patients with yellow nail syndrome.**

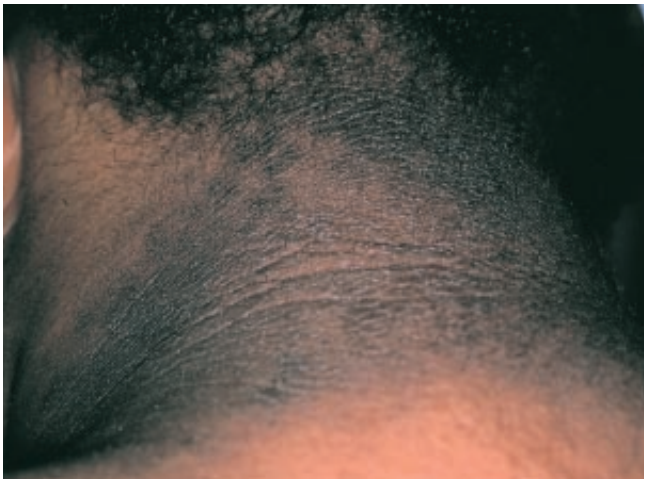


Figure 6 **Acanthosis nigricans, a dark velvety acanthosis that can occur in patients with diabetes mellitus and other endocrine disorders, often appears on the neck.**

tients with type 2 (non–insulin-dependent) diabetes. Scleredema may improve if diabetes is controlled.[23] Less commonly, scleredema occurs in nondiabetic patients after streptococcal pharyngitis; in such patients, the disease is self-limited, resolving within 2 years of onset. High-dose corticosteroids,[24] radiation,[25] and ultraviolet-A1 irradiation (UVA1)[26] have all been used to treat scleredema.

Diabetic bullae, neuropathic ulcers, and so-called waxy skin and stiff joints occur in patients with diabetes. In the last condition mentioned, scleroderma-like induration of the skin over the dorsal aspect of the hands prevents full flexion or extension of the proximal interphalangeal joints.

Diabetic patients are prone to a number of infections, including erythrasma, a corynebacterial infection resulting in asymptomatic reddish-brown patches in intertriginous sites, especially the groin and axillae. Patients are also prone to staphylococcal infections and frequently develop furuncles and carbuncles. Candidal infections are another risk, particularly when blood glucose levels are poorly controlled.

GRAVES DISEASE

Graves disease consists of a triad of exophthalmos, hyperthyroidism, and pretibial myxedema [*see 3:I Thyroid*]. Pretibial myxedema presents as skin-colored nodules and plaques that extend from the pretibial area down to the dorsa of the feet. Lesions often develop after treatment of hyperthyroidism, although they can occur at any stage in the evolution of Graves disease.

Onycholysis, the separation of the nail plate from the nail bed, occurs in many patients with hyperthyroidism. Other autoimmune skin diseases, such as vitiligo and alopecia areata, are increased in patients with Graves disease. Manifestations of thyroid disease include the stigmata of hypothyroidism. Patients can develop alopecia; specifically, they can lose the lateral third of the eyebrows. Edematous thickening of the lips, tongue, and nose occur as well.

Gastrointestinal Diseases

Patients with any of a number of gastrointestinal diseases may present with cutaneous manifestations; similarly, patients

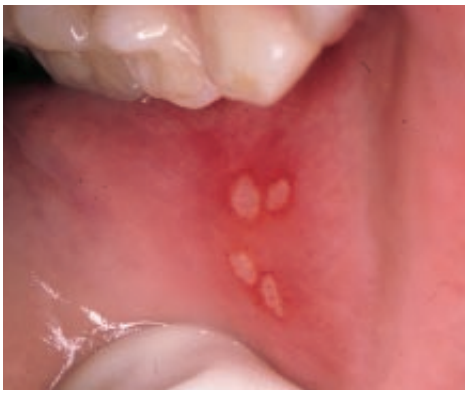


Figure 7 **Aphthous stomatitis is a common finding in patients with ulcerative colitis.**

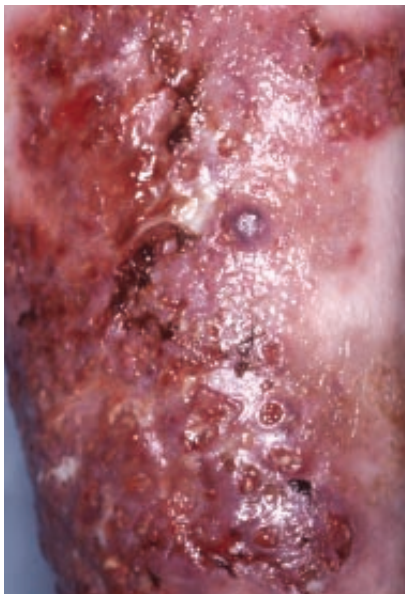


Figure 8 **Pyoderma gangrenosum is characterized by ulcers that begin with craterlike holes draining pus.**

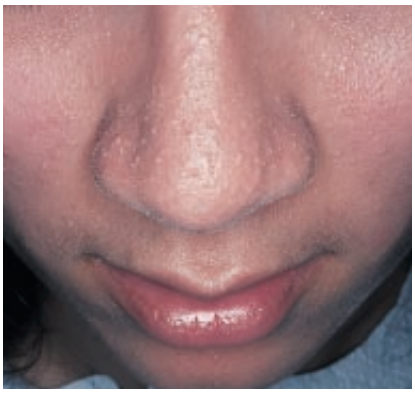


Figure 9 **The patient's nose and cheeks are covered with small papules called trichilemmomas, which represent the cutaneous hallmark of Cowden disease.**

with certain cutaneous diseases can develop gastrointestinal complications.

CARCINOID SYNDROME

The carcinoid syndrome is characterized by episodic flushing that can be associated with abdominal pain, diarrhea, and wheezing. Ninety percent of carcinoid tumors originate in the gastrointestinal tract; however, bronchial carcinoids occur occasionally. Less common cutaneous manifestations of carcinoid tumors include sclerodermatous changes. Cutaneous metastases present as deep nodules; hyperkeratosis may occur; and the patient may experience pigmentation changes similar to those seen in pellagra.

INFLAMMATORY BOWEL DISEASE

There are several specific and nonspecific cutaneous manifestations of inflammatory bowel disease [*see 4:IV Inflammatory Bowel Diseases*]. In both Crohn disease and ulcerative colitis, disease can progress to a hypercoagulable stage, causing venous and arterial thromboses that can lead to loss of digits and limbs. Aphthous stomatitis is another nonspecific manifestation of inflammatory bowel disease [*see Figure 7*]. In patients with Crohn disease, the lesions may appear as noncaseating granulomas, whereas in patients with ulcerative colitis, they may be indistinguishable from canker sores.

Pyoderma gangrenosum occurs in patients with Crohn disease and ulcerative colitis and has also been reported in patients with chronic active hepatitis, rheumatoid arthritis, and a number of myeloproliferative disorders. The lesions are distinguishable from other ulcers by the presence of craterlike holes, pustules, and purulent drainage [*see Figure 8*]. Pyoderma gangrenosum may occur at sites of trauma. Treatment with intralesionally injected or systemic corticosteroids may be required. Immunosuppressive agents such as cyclosporine have proved to be dramatically effective; in refractory cases, thalidomide has been shown to be beneficial.[27] Infliximab has proved to be highly effective in the treatment of refractory pyoderma gangrenosum.[28]

Erythema nodosum is a septal panniculitis that is associated with a number of conditions, including Crohn disease, ulcerative colitis, Behçet syndrome, sarcoidosis, infection, and the ingestion of estrogens and other drugs. Other manifestations of Crohn disease include inguinal abscesses and sinuses and anal fistulas.

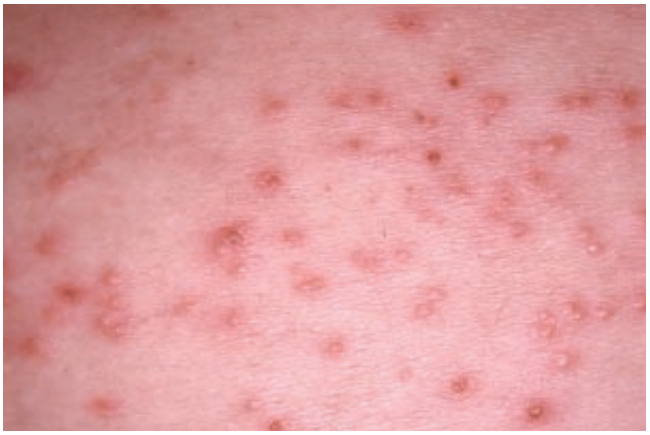


Figure 10 **The primary lesions of herpetiformis are vesicles that quickly break to form crusts and erosions.**

METASTATIC CROHN DISEASE

The term metastatic Crohn disease refers to histologically proven noncaseating granulomas that are remote from the gastrointestinal tract in patients with Crohn disease. The clinical presentation can be quite variable, and the diagnosis of this disorder is frequently missed. In some cases, patients present with marked swelling of the scrotum or vulva.

CUTANEOUS CONDITIONS WITH GASTROINTESTINAL COMPLICATIONS

Cowden Disease

Cowden disease is an autosomal dominant disorder in which gastrointestinal polyps develop along with numerous skin lesions. This disease has been attributed to mutations of the tumor suppressor gene *PTEN*.[29] Wartlike papules known as trichilemmomas occur, particularly around the nose, mouth, and ears but

also on the hands and feet [*see Figure 9*]. Small papules can also develop on the gingival mucosa, creating a cobblestone appearance. Hemangiomas and lipomas can occur.[30] A distinctive nodule of the scalp known as Cowden fibroma has been described. Up to 50% of women with Cowden disease develop breast cancer, a finding that has been associated independently with the *PTEN* mutation.[31] Thyroid carcinomas, thyroid adenomas, and thyroid goiters can occur as well.

Dermatitis Herpetiformis

Dermatitis herpetiformis is an immunobullous disease that is associated with a gluten-sensitive enteropathy [*see 2:IX Vesiculobullous Diseases*]. Skin lesions begin as vesicles that are so pruritic that they are quickly broken by scratching, leaving only excoriations and crusts [*see Figure 10*]. Like patients with celiac disease who are not on a gluten-free diet, patients with dermatitis herpetiformis have an increased risk of gastrointestinal lymphoma.[32]

Peutz-Jegher Syndrome

In Peutz-Jegher syndrome, patients develop hamartomatous polyps of the small intestine that are associated with pigmented macules of the lips and oral mucosa [*see 2:X Malignant Cutaneous Tumors*]. Also, pigmented macules can develop on the palms, fingers, soles, and toes and in areas around the mouth, nose, and rectum. The disease is inherited as an autosomal dominant trait, and a significant proportion of cases are associated with mutations in the serine/threonine protein kinase I1/LKB1 (*STKI1/LKB1*) gene, although mutations in this gene do not account for all cases.[33]

Recessive Dystrophic Epidermolysis Bullosa

Recessive dystrophic epidermolysis bullosa is a congenital bullous disease with recurrent blistering and scarring, particularly on the hands and feet [*see 2:IX Vesiculobullous Diseases*]. The scarring results in pseudosyndactyly, giving rise to mitten-like hands. Ingestion of coarse food can result in mucosal bullae of the esophagus, which heal with scarring and stricture formation. Dysphagia is a frequent complaint. Scarring of the esophagus can lead to squamous cell carcinoma, which is a leading cause of death in this disorder.[34] Gastroenterologists and dermatologists must play key roles in the management of these patients. Liquid and pureed diets and appropriate skin care are essential to the survival of patients with this debilitating disorder. Prenatal diagnosis can be made by sampling DNA from the chorionic villus.[35] All forms of dystrophic epidermolysis bullosa have been attributed to mutations in the type VII collagen gene.[36] Recently, through the use of a self-inactivating minimal lentivirus-based vector, the type VII collagen gene has been delivered and type VII collagen expressed in immunodeficient mice, suggesting the possibility that, in the future, gene therapy may be available to successfully treat this devastating disorder.[37]

Hematologic Diseases

AMYLOIDOSIS

There are several forms of local and systemic amyloidosis [*see 12:XV Chronic Lymphoid Leukemias and Plasma Cell Disorders*]. In a form associated with multiple myeloma, amyloid fibrils consisting of immunoglobulin light chains are deposited in the skin. Shiny translucent papules develop, particularly on the eyelids. Because of amyloid deposits in blood vessels, spontaneous bleeding occurs. Minimal trauma results in petechiae and purpura. Macroglossia also occurs in some patients with myeloma-associated amyloidosis and in some with primary systemic amyloidosis. The systemic manifestations of myeloma-associated and primary systemic amyloidosis are quite varied. Hepatomegaly develops in 50% of patients. Amyloid can affect the heart, resulting in heart failure or myocardial infarction. Survival of patients who undergo heart transplantations for cardiac amyloidosis is lower than survival after cardiac transplantation for other indications.[38] Amyloidosis of the gastrointestinal tract can result in malabsorption and protein-losing enteropathy. Treatment with thalidomide (up to 400 mg daily) and intermittent dexamethasone is rapidly effective in some patients, but side effects are frequent.[39]

MASTOCYTOSIS

Mastocytosis is caused by the infiltration of mast cells into the skin and other organs [*see 4:XI Diseases Producing Malabsorption and Maldigestion*]. Urticaria pigmentosa refers to the skin lesions that occur in most patients with mastocytosis. Reddish-brown macules and papules resembling nevi are characteristic [*see Figure 11*]. Stroking of individual lesions results in urticarial wheals—a phenomenon known as the Darier sign. Pruritus, flushing, abdominal pain, nausea, vomiting, and diarrhea are common complaints.

Most patients with mastocytosis have an indolent form of the disease, even when mast cells have infiltrated the bone marrow.[40] Malignant or aggressive systemic mast cell disease can involve the spleen, liver, and lymph nodes in addition to the skin and bone marrow. Histologically, infiltrates contain atypical nonmetachromatic mast cells that are monoclonal in some patients.[41] Children with urticaria pigmentosa usually have a better prognosis than adults with the disease.[42]

The diagnosis of mastocytosis is made by the demonstration of mast cells on skin biopsy. Because mast cells easily degranulate, making them difficult to identify, biopsies should be performed with a minimum of tissue manipulation.

PORPHYRIAS

The porphyrias result from defective hemoglobin synthesis, leading to excess porphyrins in the blood and in body tissues [*see 9:V The Porphyrias*].

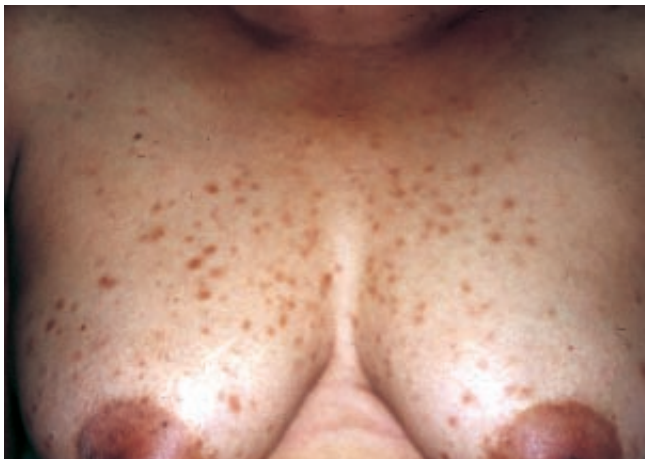


Figure 11 **Multiple brown macules resembling nevi occur in patients with urticaria pigmentosa.**

Congenital Erythropoietic Porphyria

Congenital erythropoietic porphyria is a rare autosomal recessive disorder that has been attributed to mutations in the gene for uroporphyrinogen III synthase.[43] This condition is characterized by severe photosensitivity. Vesicles and bullae develop after sun exposure; these lesions heal with scar formation. Erythrodontia (red-stained teeth) is a characteristic feature [*see Figure 12*]. Digit, ear, and nose loss is common in patients who manage to survive to adulthood [*see Figure 13*]. Hypertrichosis is another frequent complication. Formation of gallstones, splenomegaly, and hemolytic anemia are also associated with this condition.

Porphyria Cutanea Tarda

Porphyria cutanea tarda is characterized by photosensitivity, vesicle formation (especially on the dorsa of the hands) [*see Figure 14*], and hypertrichosis. The condition may be associated with ingestion of alcohol or medications such as estrogens. Diagnosis of the porphyrias can be established by elevated urinary porphyrin levels. Examination of the urine with a Wood lamp will often reveal pink-red fluorescence attributable to the high level of urinary porphyrins. Porphyria cutanea tarda can be associated with hepatitis C. Phlebotomy is effective therapy.

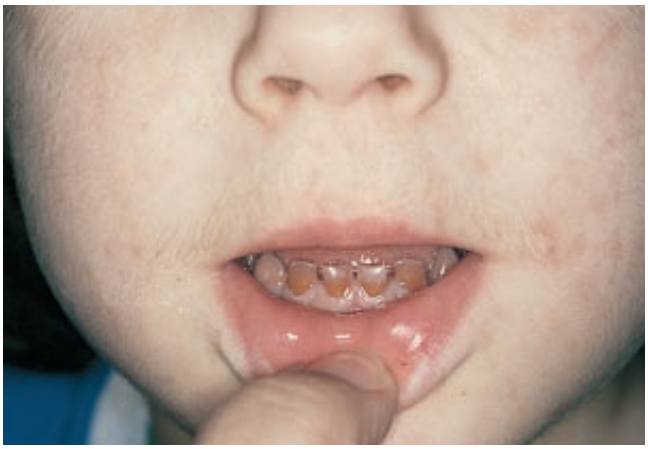


Figure 12 **A reddish pigmentation (erythrodontia) occurs when porphyrins are deposited in the teeth in congenital erythropoietic porphyria.**

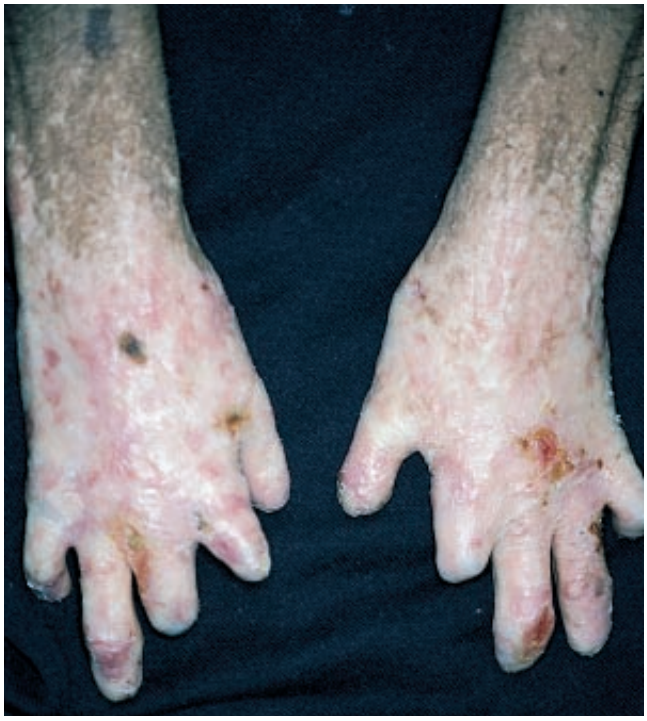


Figure 13 **Skin changes in congenital erythropoietic porphyria can be severe; scarring and loss of digits are common in older patients.**

Immunodeficiency Diseases

AIDS

AIDS may result in cutaneous infections and neoplasms that are often dramatic in their extent and severity. This section focuses on selected cutaneous manifestations of infections and other diseases associated with AIDS. (For a more comprehensive discussion of disorders associated with HIV infection, see *7:XXXIII HIV and AIDS* and other chapters devoted to specific conditions.)

Opportunistic Infections

Viral infections Banal viral infections, such as molluscum contagiosum, that are ordinarily self-limited and easily curable have become widespread, chronic, and enormous problems in patients with AIDS. These umbilicated white papules, ordinarily only a few millimeters in diameter, can reach diameters of 1 to 2 cm in patients with AIDS. Similarly, condyloma acuminatum, caused by human papillomavirus (HPV) infection, is often difficult to treat in patients with AIDS.

Herpes simplex virus infections become chronic and erosive, forming large, nonhealing ulcers [*see 7:XXVI Herpesvirus Infections*]. Acyclovir-resistant strains of herpes simplex virus have been reported in some patients with AIDS[44]; these patients require other antiviral agents, such as foscarnet. Mutations in thymidine kinase and DNA polymerase genes of herpes simplex viruses can render them resistant to acyclovir and foscarnet.[45] Topical cidofovir gel has been reported to be beneficial for herpes infections in patients infected with HIV.[46]

Herpes zoster infections are a common sign of HIV infection. In the non–HIV-infected host, herpes zoster is characterized by grouped vesicles in a dermatomal distribution. The eruption is self-limited, resolving within 1 to 2 weeks. In contrast, herpes zoster infection can develop into a disseminated vesicular eruption in patients with AIDS; and in some AIDS patients, chronic herpetic lesions develop and last for months.

Fungal infections Fungal infections are common in patients with HIV infection. Monilial infections include oral thrush and candidiasis of the groin. Several fungal infections that rarely cause widespread infection in patients with normal immune systems (e.g., cryptococcosis, histoplasmosis, aspergillosis, and sporotrichosis) have emerged as serious pathogens in patients with AIDS.

Bacterial infections Bacterial infections are more frequent and severe in patients with AIDS than in patients with normal immune systems. Bacillary angiomatosis, caused by *Bartonella henselae,* presents as purple papules and nodules that can be mistaken for Kaposi sarcoma (see below). Chronic fever and chills can occur, as can bone lesions. Epidemiologic evidence suggests that cats may be the source of human infection.[47] Diag-

nosis by serologic testing has been commonly used, but in the future, polymerase chain reaction may offer a rapid and convenient way of establishing this diagnosis.[48] The condition resolves upon treatment with oral antibiotics [*see 7:XI Infections Due to Brucella, Francisella, Yersinia pestis, and Bartonella*].

Scabies and other pruritic eruptions Scabies, a severely pruritic eruption, has a predilection for the buttocks, the genitals, the periumbilical area, and the webs between the fingers. Norwegian scabies, a thickly crusted psoriasislike form of the parasitic disease, has been described in patients with Down syndrome and in other immunosuppressed persons. In recent years, Norwegian scabies has been reported most commonly in patients with AIDS. The scales of Norwegian scabies contain thousands of mites that are easily seen with the microscope. Burrows form linear lesions up to 1 cm long. The causative mite, *Sarcoptes scabiei,* can be identified by microscopic examination of scrapings from the burrows.

Eosinophilic pustular folliculitis and papular eruption of AIDS are pruritic rashes that affect patients with HIV infection. It has been suggested that pruritic papular eruption in patients with HIV infection may represent a reaction to arthropod bites.[49]

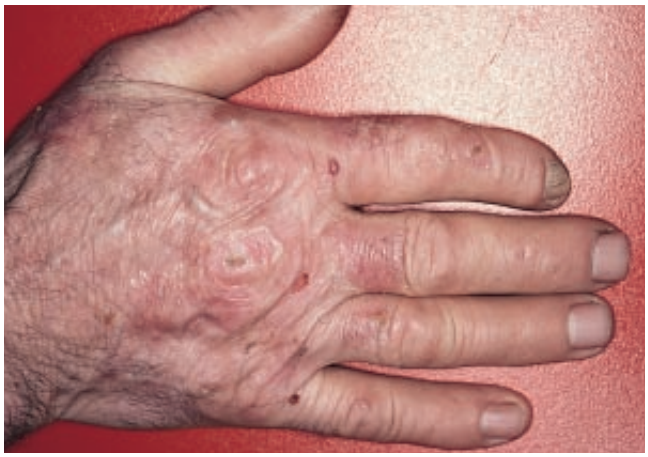


Figure 14 **Crusting and scarring follow the appearance of vesicles and bullae in porphyria cutanea tarda.**

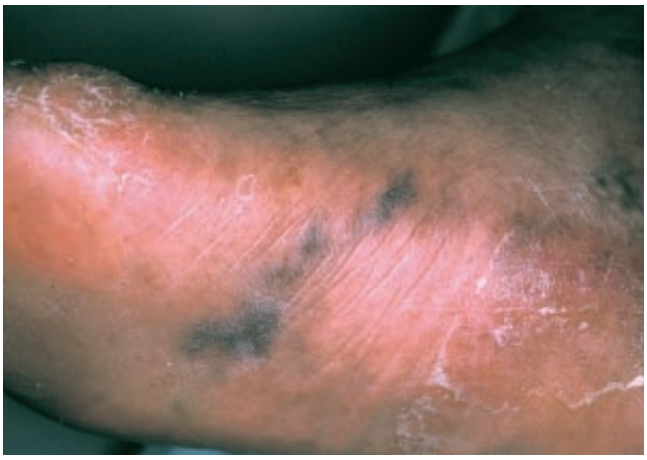


Figure 15 **Kaposi sarcoma is the most common malignancy of AIDS patients. It often presents as purple patches, plaques, or papules. Purple macules on the foot can be seen in patients with classic Kaposi sarcoma but are seen here in a patient with AIDS-related Kaposi sarcoma.**

Both eosinophilic pustular folliculitis and papular eruption of AIDS are characterized by severe itching, and skin-colored papules and excoriations are common in both. Patients with eosinophilic pustular folliculitis can develop pustules and erythematous papules. Both conditions respond to treatment with ultraviolet B.

Kaposi Sarcoma

Kaposi sarcoma, a slowly progressive vascular neoplasm, was originally described in elderly Italian and Jewish men [*see 2:X Malignant Cutaneous Tumors*]. Subsequently, a more rapidly progressive form of the disorder was described in immunosuppressed patients with lymphomas and in kidney transplant patients on immunosuppressive drugs. An aggressive form has been described in patients with AIDS [*see Figure 15*]. Classic Kaposi sarcoma typically affects the lower extremities and only gradually progresses to other sites. In contrast, AIDS-related Kaposi sarcoma can occur on any surface of the body, including mucous membranes. Human herpesvirus type 8 has been implicated in both classic and AIDS-related Kaposi sarcoma.[50] Treatments include radiation therapy, cryotherapy, and intralesional injection with vinblastine; systemic chemotherapy can also be effective. In patients with AIDS, Kaposi sarcoma is best treated with antiretroviral regimens.

Oral Hairy Leukoplakia

Oral hairy leukoplakia, another condition that has been described in HIV-infected patients, consists of linear white papules on the lateral surfaces of the tongue that result in the so-called hairy appearance. Oral hairy leukoplakia can be distinguished from oral thrush in that the lesions cannot be rubbed off, as they can be in thrush.

Thanks to the development of effective antiretroviral therapy, the frequency of opportunistic infections in patients with HIV infection has diminished markedly.

OTHER IMMUNODEFICIENT STATES

Other inherited or acquired immunodeficiency states share a number of clinical features. Susceptibility to monilial infections or bacterial infections is increased in disorders such as chronic granulomatous disease and chemotherapy-induced neutropenia. Oral ulcers similarly occur in cyclic neutropenia and in chemotherapy-induced immunosuppression.

Some immunosuppressive drugs have characteristic cutaneous effects. Corticosteroids, when used long-term, cause vascular fragility, resulting in steroid purpura. They can also cause cutaneous atrophy, formation of striae, and acneiform eruptions. Cyclosporine is associated with hypertrichosis. Aphthous stomatitis is a characteristic effect of numerous immunosuppressive drugs, particularly agents that suppress bone marrow function. Chronic immunosuppression can lead to the development of lymphoma and nonmelanoma skin cancer. Avoidance of excessive exposure to sunlight may prevent development of the latter.

Infectious Diseases

Cutaneous manifestations can be major features of a number of systemic infections; for example, patients with overwhelming septicemia can develop disseminated intravascular coagulation (DIC), which results in cutaneous infarcts and hemorrhage into the skin. Key cutaneous features of selected systemic infections follow.

INFECTIVE ENDOCARDITIS

The cutaneous manifestations of infective endocarditis include petechiae, splinter hemorrhages (linear red streaks under the nail), Osler nodes (tender purpuric nodules on the finger pads and toes), and Janeway lesions (nontender purpuric macules of the palms and soles). Skin lesions are caused by either septic emboli or vasculitis. Treatment of the underlying infection results in resolution of the cutaneous manifestations [*see 7:XVIII Infective Endocarditis*].

STAPHYLOCOCCAL TOXIC-SHOCK SYNDROME

Staphylococcal toxic-shock syndrome was first recognized in menstruating women who used superabsorbent tampons [*see 7:I Infections Due to Gram-Positive Cocci*]. It is caused by an exotoxin produced by certain strains of *S. aureus*.[51] Staphylococcal infections in bone, soft tissue, and other sites have been implicated. Patients develop diffuse sunburnlike erythema, with swelling of the hands and feet, followed by desquamation of the palms and soles. Erythema of mucous membranes, fever, and hypotension also occur. Gastrointestinal symptoms, impaired renal function, elevated liver function values, thrombocytopenia, and myositis can develop.

STAPHYLOCOCCAL SCALDED SKIN SYNDROME

Staphylococcal scalded skin syndrome (SSSS) is caused by a circulating exfoliative toxin produced by *S. aureus* phage group 11. Generalized bulla formation with large areas of desquamation is characteristic of the disorder. Along with tenderness, erythema, and exfoliation of skin, patients have fever. The source of the staphylococcal infection is not always apparent; occasionally, the infection arises in a wound or in an occult abscess. Because the staphylococcal infection is usually remote from the affected skin, culture of the skin does not grow *S. aureus*.

SSSS must be differentiated from toxic epidermal necrolysis. Toxic epidermal necrolysis commonly affects adults and involves mucous membranes; SSSS usually affects children and spares mucous membranes. In addition, toxic epidermal necrolysis can last for several weeks and has a high mortality, whereas SSSS lasts a few days and usually has a good outcome. Histologically, SSSS shows bulla formation in the upper epidermis, and the bulla cavity contains free-floating, normal-appearing, acantholytic cells. In toxic epidermal necrolysis, bulla formation occurs at the basal layer of the epidermis, and the epidermal cells are necrotic. Treatment with antibiotics effective against *S. aureus* eliminates the underlying cause of SSSS.

NECROTIZING FASCIITIS

Necrotizing fasciitis is caused by a mixed anaerobic infection of an ulcer or a surgical or traumatic wound. The affected skin is erythematous, warm, and tender and develops hemorrhagic bullae that rupture to form rapidly enlarging areas of gangrene that extend down to the fascia. Surgical debridement is essential for this life-threatening infection.[52]

MENINGOCOCCEMIA

Acute meningococcemia can occur either in epidemics or in isolated cases [*see 7:III Infections Due to Neisseria*]. Fever, headache, and a hemorrhagic rash develop. If untreated, patients develop DIC, with extensive hemorrhage, hypotension, and ultimately death. The causative organism, *Neisseria meningitidis*, is usually identified in cerebrospinal fluid but can also be identified by smear or cultures of skin lesions or by blood cultures.

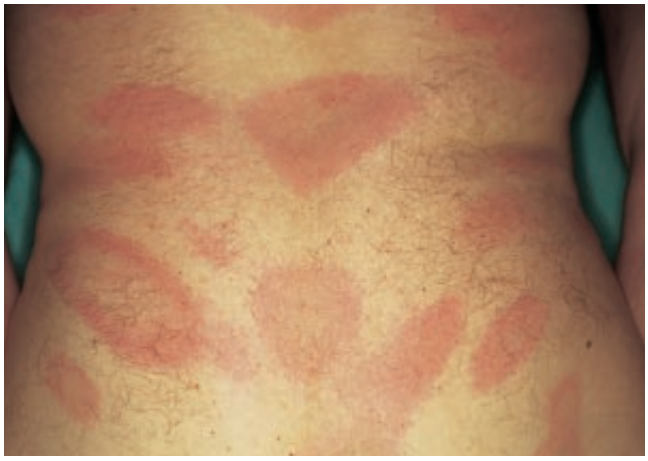


Figure 16 **Several weeks after primary infection with Lyme disease, hematogenous dissemination of spirochetes results in multiple patches of erythema chronicum migrans.**

Treatment with antibiotics and supportive care are essential aspects of therapy.

SCARLET FEVER

Scarlet fever begins with pharyngitis caused by group A *Streptococcus* [*see 7:I Infections Due to Gram-Positive Cocci*]. A generalized rash develops 1 to 2 days after onset of the pharyngitis. The rash is characterized by pinpoint erythematous papules that may be easier to palpate than to see. Other characteristic lesions include a white strawberry tongue and linear petechial macules occurring in body folds (Pastia lines). As the rash fades, desquamation of the palms and soles appears. Treatment with penicillin results in rapid resolution of all symptoms.

VIBRIO INFECTION

Vibrio vulnificus infection arises from minor trauma sustained while swimming in lakes or the ocean or while cleaning seafood. Cellulitis occurs, with lymphangitis and bacteremia. In patients with hepatic cirrhosis, infection can occur after eating raw oysters. These patients develop hemorrhagic bullae, with leukopenia and DIC.[53] Treatment with antibiotics is necessary; management of complications may require intensive supportive care.

LYME DISEASE

Lyme disease is caused by the spirochete *Borrelia burgdorferi* and is transmitted primarily by the tick *Ixodes scapularis* [*see 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*]. The characteristic skin lesion, erythema chronicum migrans, begins as an erythematous macule or papule at the site of the tick bite. Over days and weeks, the erythematous lesion expands to form a red ring, often with central clearing. If left untreated, lesions last weeks or months. Hematogenous dissemination of spirochetes occurs after several weeks, resulting in multiple annular patches of erythema chronicum migrans [*see Figure 16*]. Systemic complications include an acute arthritis involving one or a few large joints a few weeks after the onset of symptoms. A chronic erosive arthritis develops in approximately 10% of patients. Neurologic symptoms, including Bell palsy, can occur, as can cardiac complications, including heart failure and cardiac conduction abnormalities.

Lyme disease can be prevented by the removal of ticks within 18 hours of attachment. Once symptoms have developed, oral

antibiotics are effective at destroying *B. burgdorferi*. A vaccine containing a genetically engineered protein from the surface of the bacteria was found to prevent infection in most vaccinated people[54]; however, for a number of reasons, including lack of demand, the vaccine has been discontinued.[55]

ROCKY MOUNTAIN SPOTTED FEVER

Rocky Mountain spotted fever (RMSF) is a tick-borne illness caused by *Rickettsia rickettsii* [*see 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*]. It is characterized by the sudden onset of fevers, chills, and headache. Approximately 4 days later, a characteristic erythematous rash develops on the wrists and ankles and becomes purpuric. The rash then spreads centrally to involve the extremities, trunk, and face.

Because the mortality of RMSF is high, patients should be treated immediately with intravenous chloramphenicol or tetracycline if RMSF is suspected. Diagnosis can then be established by skin biopsy: immunofluorescence with antibodies against *R. rickettsii* shows the organism in the walls of cutaneous blood vessels. Serologic tests, such as the Weil-Felix reaction, can confirm the diagnosis after the acute phase of the illness.

Neurologic Diseases

BASAL CELL NEVUS SYNDROME

The basal cell nevus syndrome is an autosomal dominant disorder attributed to mutational inactivation of the *PTCH* gene[56] in which patients develop basal cell carcinomas at an early age [*see 2:X Malignant Cutaneous Tumors*]. Multiple skeletal abnormalities are associated with the syndrome, and affected individuals may also develop jaw cysts. Lamellar calcification of the falx cerebri occurs, as well as other neurologic abnormalities, including medulloblastomas.

EPIDERMAL NEVUS SYNDROME

The epidermal nevus syndrome is characterized by systemic manifestations, such as seizures, mental retardation, blindness, and skeletal abnormalities in association with large epidermal nevi. The nevi consist of long pigmented streaks that are linear or whirled and involve large areas of the body [*see Figure 17*].

INCONTINENTIA PIGMENTI

Incontinentia pigmenti is an inherited syndrome that affects the skin and nervous system. Mutations in the *NEMO* gene, an essential component of the nuclear factor–κB signaling cascade, account for 85% of cases.[57] The inheritance pattern is X-linked dominant and is lethal in male fetuses. The first skin manifestations begin within weeks after birth, occasionally occurring in utero, and consist of linear patterns of vesiculobullous lesions. Within weeks, these lesions evolve into verrucous papules and, eventually, into pigmented whirls. Apart from neurologic symptoms, patients may have ocular abnormalities, scarring alopecia, and skeletal malformations.

Hypomelanosis of Ito

Hypomelanosis of Ito, also called incontinentia pigmenti achromians, consists of whirls of hypopigmentation that are associated with neurologic symptoms in 50% of patients. Skin lesions are present at birth or develop in early childhood. In addition to seizures and mental retardation, skeletal and ocular abnormalities occur.

NEUROFIBROMATOSIS

Neurofibromatosis is a common autosomal dominant disorder involving the skin and nervous system [*see 2:XI Benign Cutaneous Tumors*]. Skin lesions include cutaneous neurofibromas, which are soft, skin-colored nodules that are often pedunculated [*see Figure 18*]. Café au lait macules are flat, evenly pigmented patches up to several centimeters in diameter. Six or more café au lait macules greater than 1.5 cm in diameter are found in most patients with neurofibromatosis type 1 (also called von Recklinghausen disease). Plexiform neuromas are larger, deeper tumors that are associated with hypertrophy of bony and soft tissues. In a small proportion of tumors, neurofibrosarcomas will arise. On skin biopsy, café au lait macules are found to contain macromelanosomes—giant granules of pigment in melanocytes and keratinocytes. Axillary and inguinal freckling also appear as pigmented macules that resemble small café au lait spots in intertriginous sites. Lisch nodules—pigmented iris hamartomas—are also found in most patients with neurofibromatosis.

Several variants of neurofibromatosis exist, including segmental neurofibromatosis, in which patients develop a segmental distribution of café au lait spots and cutaneous neurofibromas, and neurofibromatosis type 2, which consists of acoustic neuromas, schwannomas, and meningiomas without Lisch nodules and with fewer café au lait macules than appear in type 1. Patients with neurofibromatosis type 2 may have some cutaneous neurofibromas as well. Neurofibromatosis types 1 and 2 are caused by different genetic defects. Neurofibromatosis type 1 is caused by mutations in the *NF1* gene for neurofibromin on chromosome 17.[58] Neurofibromatosis type 2 has been attributed to inactivating mutations in the *NF2* tumor suppressor gene whose product, merlin, plays a number of roles in tumorigenesis.[59]

SNEDDON SYNDROME

Sneddon syndrome is a disease of the skin and nervous system caused by occlusion of small to medium-sized arteries in persons younger than 45 years. The skin lesions resemble livedo reticularis and have been called livedo racemosa. Transient ischemic attacks or strokes are common. Definitive diagnosis is made by demonstrating characteristic vascular changes on skin biopsy of patients with associated neurologic findings.

TUBEROUS SCLEROSIS

Tuberous sclerosis is an autosomal dominant disease that affects the skin and nervous system. Mutations that inactivate the *TSC1* or *TSC2* tumor suppressor genes affect the respective gene products, hamartin and tuberin, leading to tuberous sclerosis.[60] Affected patients can develop seizures, mental retardation, and brain lesions called tubers, which can be seen on CT scans. Adenoma sebaceum, the most characteristic cutaneous manifestation of tuberous sclerosis, consists of skin-colored papules of the face [*see Figure 19a*]. Other skin lesions are hypopigmented macules referred to as ash-leaf macules [*see Figure 19b*], smaller hypopigmented lesions called confetti macules, periungual and subungual fibromas (skin-colored nodules that arise around the fingers and toenails) [*see Figure 19c*], and the shagreen patch (a skin-colored plaque made of thick dermal connective tissue).

Renal Diseases

FABRY DISEASE

Fabry disease is caused by an abnormality of α-galactosidase

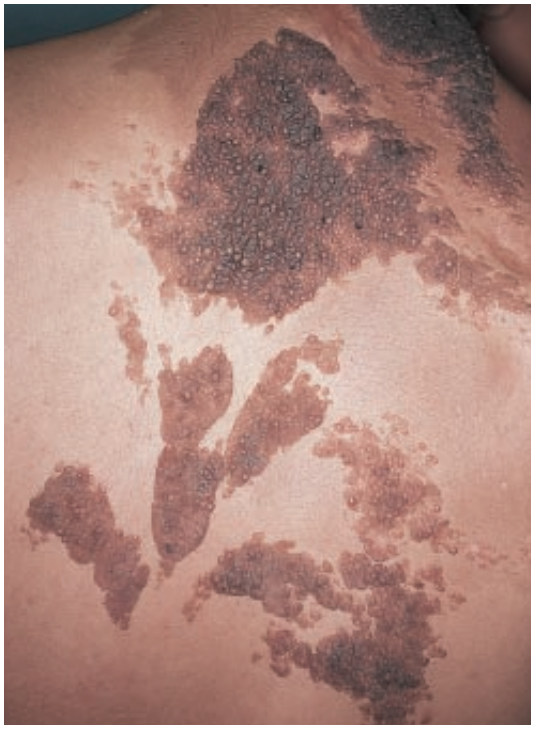


Figure 17 **The epidermal nevus syndrome is characterized by linear or whirled streaks of pigmentation that involve large areas of the body.**

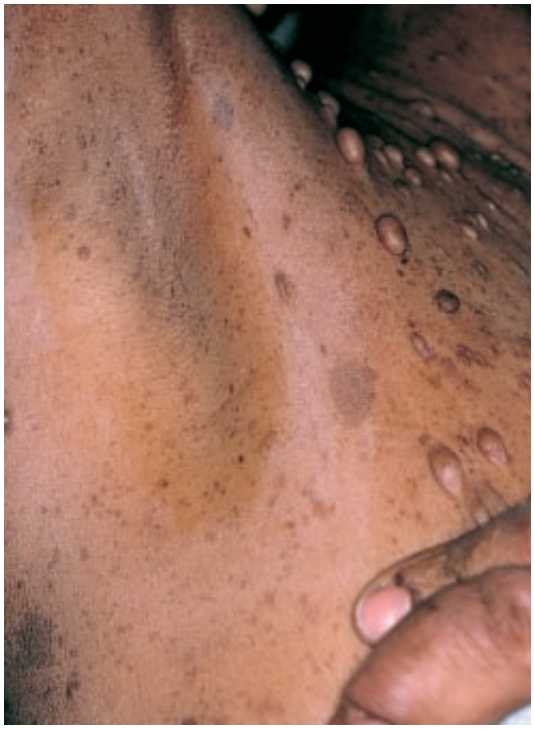


Figure 18 **Axillary freckling, café au lait spots, and neurofibromas are evident in a patient with neurofibromatosis type 1.**

A, resulting in deposition of glycosphingolipids in body tissues. The disorder is inherited as an X-linked recessive trait. A variety of different mutations in the gene for α-galactosidase A have been found in unrelated families with Fabry disease.[61] Affected males often complain of severe pain in the extremities, with burning of the palms and soles. Episodes of pain are transient, but patients complain of persistent paresthesias in the hands and feet.

Skin lesions consist of angiokeratomas, which are pinpoint red or purple papules that resemble cherry hemangiomas [*see Figure 20*]. Angiokeratomas are most commonly found in the periumbilical area but can also occur on the palms, soles, trunk, extremities, and mucous membranes. In adults, glycosphingolipids become deposited in blood vessels and organs, affecting the heart, heart valves, coronary arteries, and kidneys. Replacement therapy with recombinant human α-galactosidase A can improve cutaneous, gastrointestinal, neurologic, and psychiatric symptoms; it has been shown to be safe and can eliminate substrate storage of glycosphingolipids, but questions remain regarding optimal dosing.[62]

POLYARTERITIS NODOSA

Polyarteritis nodosa is an inflammatory condition that affects muscular arteries [*see 15:VIII Systemic Vasculitis Syndromes*]. Aneurysms form in many arteries, including those leading to the kidneys and subcutaneous tissue. Diagnosis of the systemic form of polyarteritis can be made by demonstrating aneurysms of the renal arteries on renal arteriograms.

A localized cutaneous form of polyarteritis nodosa most commonly presents as painful nodules of the lower extremi-

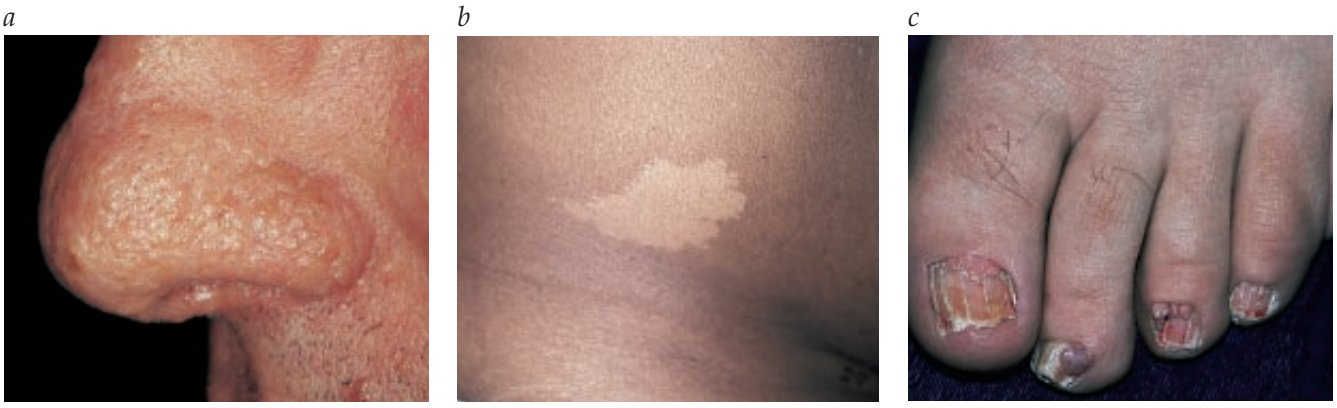


Figure 19 **Several of the characteristic cutaneous findings of tuberous sclerosis are shown: adenoma sebaceum (*a*); ash-leaf macule (*b*); and periungual fibromas (*c*).**

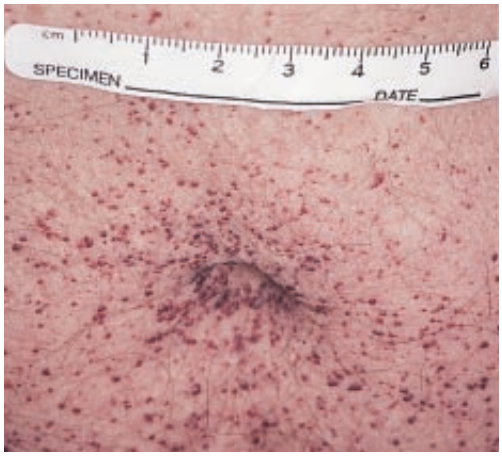


Figure 20 **Angiokeratomas are particularly common in the periumbilical area of patients with Fabry disease.**

ties.[63] In mild cases, patients may only have livedo reticularis; but in severe cases, skin lesions can ulcerate. A polyneuropathy may be associated with the disorder. Patients with classic polyarteritis and microaneurysms have an increased incidence of hepatitis B antigenemia; in contrast to patients with other vasculitides, they usually do not have antineutrophil cytoplasmic antibodies.[64]

PERFORATING DISORDERS

Perforating disorders include several conditions characterized by extrusion of dermal material through the epidermis. These lesions often develop in association with renal failure and diabetes mellitus.[65] Skin lesions are characterized by hyperkeratotic papules with central white craters that histologically can be shown to contain dermal material. Reactive perforating collagenosis, perforating folliculitis, and Kyrle disease are all examples of perforating disorders associated with renal failure.

CALCIPHYLAXIS

Calciphylaxis, also known as calcific uremic arteriolopathy, is a condition of patients with renal failure in which localized areas of skin become necrotic as a result of vascular calcification. Calciphylaxis begins with painful purpuric patches that may be reticulated, resembling livedo reticularis. These patches progress to indurated plaques that may ulcerate, becoming necrotic [*see Figure 21*]. Calciphylaxis often eventuates in amputation or death. Parathyroidectomy may result in healing of affected skin without amputation.[66]

Rheumatologic Diseases

DERMATOMYOSITIS

The best-known cutaneous manifestations of dermatomyositis, an inflammatory disorder of muscle and skin, are Gottron papules and heliotrope erythema. Gottron papules are erythematous scaling macules and papules that occur on the dorsa of the knuckles [*see Figure 22*]. Heliotrope erythema consists of periorbital erythema and edema. Scalp lesions, which can be associated with alopecia, have been described.[67] The lesions are often misdiagnosed as seborrheic dermatitis or psoriasis.

The association between dermatomyositis and malignancy has been established[68,69]; one epidemiologic study indicates patients with dermatomyositis are at particular risk for ovarian and lung cancer.[69]

Classifications of dermatomyositis include a juvenile variant characterized by calcification of skin or muscle. A vasculitic form in children is complicated by cutaneous infarcts and ulceration and by gastrointestinal vasculitis with abdominal pain, bleeding, or perforation. The vasculitic form carries a poor prognosis, with many of the patients dying of this disease.

SCLERODERMA AND SCLERODERMA-LIKE DISEASES

The sclerodermas include a number of distinct syndromes sharing a common feature, induration of the skin [*see 15:V Scleroderma and Related Diseases*].

Progressive Systemic Sclerosis and CREST Syndrome

Progressive systemic sclerosis, also known as systemic scleroderma, is a frequently fatal disease in which patients present with Raynaud phenomenon and sclerodactyly (induration of the skin of the digits) [*see Figure 23*]. Cutaneous induration can become widespread. Involvement of the face can lead to a characteristic appearance with pursed lips and bound-down skin of the nose that creates a beaklike appearance. Patients with antibodies to Scl-70 have a poor prognosis, often succumbing to renal disease and malignant hypertension. Pulmonary fibrosis can

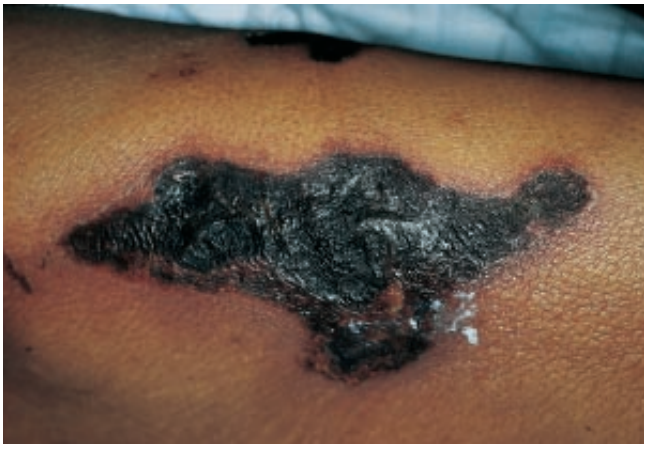


Figure 21 **Calcification of arteries in patients with renal failure results in calciphylaxis. Affected skin forms a black, necrotic eschar.**

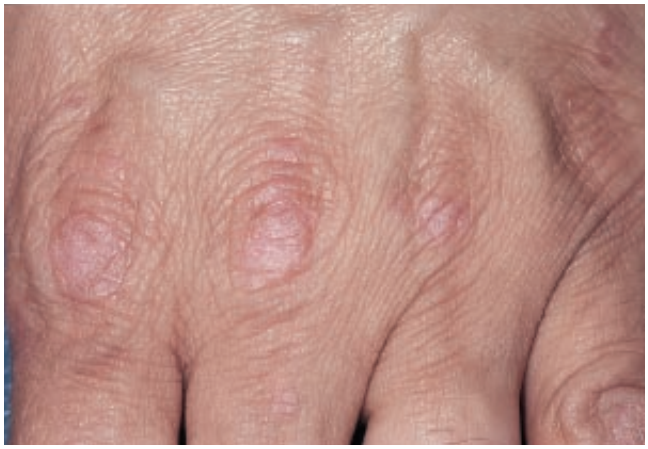


Figure 22 **Erythematous scaling papules on the dorsal aspects of the knuckles (Gottron papules) are a sign of dermatomyositis.**

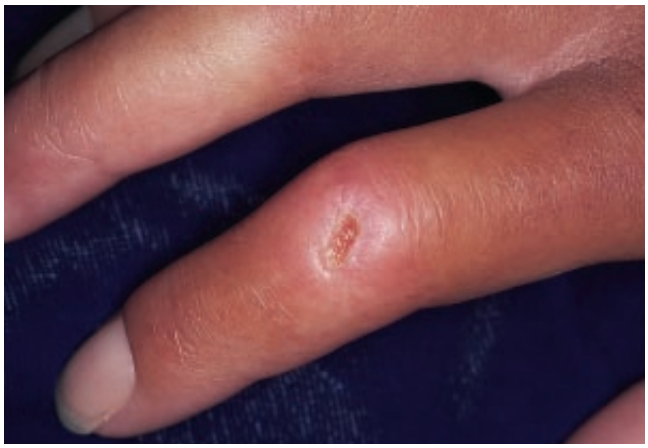


Figure 23 **Sclerodactyly with a nonhealing digital ulcer commonly occurs in progressive systemic sclerosis.**

occur. Patients with anticentromere antibodies have a more slowly progressive variant of scleroderma known as the CREST syndrome, which is characterized by cutaneous calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia. With time, pulmonary hypertension and right-sided heart failure develop.

Morphea

Morphea, also called localized scleroderma, is characterized by sharply demarcated patches of indurated skin that can become generalized. It is distinguished from progressive systemic sclerosis by the absence of Raynaud phenomenon, sclerodactyly, or the systemic complications of scleroderma. There have been innovations in the treatment of both progressive systemic sclerosis and morphea. Exposure to psoralen and longwave ultraviolet light (PUVA) has been reported to improve progressive systemic sclerosis and morphea dramatically,[70] and exposure to UVA1 (the longer UVA spectrum, from 340 to 400 nm) has been reported to benefit patients with localized scleroderma.[71] Anecdotal evidence suggests that topical calcipotriene is an effective treatment for morphea.[72] Further studies must be done to confirm the efficacy of these treatments. Anecdotal reports have indicated that minocycline may benefit patients with progressive systemic sclerosis, but controlled trials are needed.[73]

Graft versus Host Disease

As organ transplantation becomes more common, another scleroderma-like illness, graft versus host disease, increases in frequency, particularly after bone marrow transplantation [*see 5:X Transfusion Therapy and 5:XI Hematopoietic Stem Cell Transplantation*]. There are two stages of graft versus host disease. The first, acute graft versus host disease, develops 10 to 40 days after transplantation and consists of an erythematous macular and papular rash that is often associated with fever, hepatomegaly, lymphadenopathy, or gastrointestinal symptoms. Chronic graft versus host disease usually develops 3 months after transplantation but can occur later; it consists of purple papules resembling lichen planus [*see Figure 24*]. Sclerodermatous skin changes with telangiectasia, reticulated hyperpigmentation, and alopecia are most characteristic. Both cyclosporine and PUVA have proved to be useful in the prevention and treatment of graft versus host disease.[74,75] Infliximab has been used very successfully to treat acute graft versus host disease.[76]

Eosinophilic Fasciitis

Scleroderma-like hardening of the skin also occurs in eosinophilic fasciitis. Puckering of the skin on the extremities typically develops and is associated with pain. In contrast to progressive systemic sclerosis, Raynaud phenomenon does not occur. Definitive diagnosis requires biopsy of skin and fascia overlying the affected muscle. In some cases of eosinophilic fasciitis, hematologic abnormalities develop, including aplastic anemia, thrombocytopenia, Hodgkin disease, and leukemias.[77]

SYSTEMIC LUPUS ERYTHEMATOSUS

There are many cutaneous manifestations of systemic lupus erythematosus (SLE), including nonspecific manifestations such as Raynaud phenomenon, photosensitivity, alopecia, and mucosal ulcers. More specific cutaneous manifestations of SLE include so-called discoid lupus (characterized by round scarred skin lesions with central hypopigmentation and a rim of hyperpigmentation) and malar erythema [*see 15:IV Systemic Lupus Erythematosus*]. As we learn more about lupus, the spectrum of skin diseases associated with this disorder continues to expand. Subacute cutaneous lupus, a variant characterized serologically by anti-Ro and anti-La antibodies, is associated with annular or psoriasiform skin lesions [*see Figure 25*].

Anticardiolipin Antibody Syndrome

The anticardiolipin antibody syndrome, which can occur in patients with SLE, has been described in patients who suffer re-

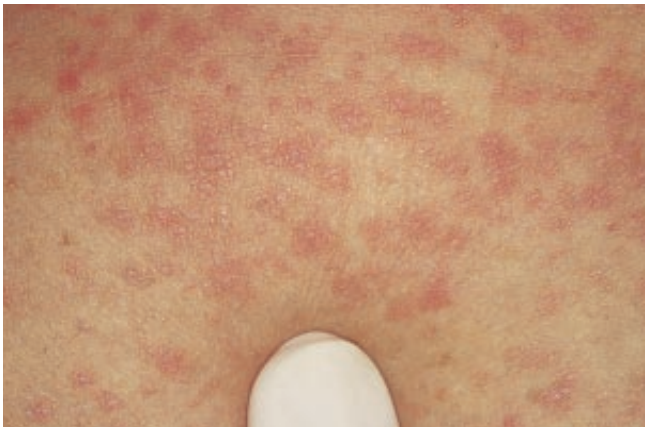


Figure 24 **Flat-topped papules are seen in this chronic lichenoid graft versus host reaction.**

Figure 25 **Annular scaling erythematous patches are characteristic of subacute cutaneous lupus erythematosus.**

peated episodes of phlebitis, arterial thromboses, and repeated miscarriages. Cutaneous infarcts are common manifestations, and livedo reticularis can occur. Patients may have false positive serologies for syphilis and have a circulating lupus anticoagulant. Circulating antiphospholipid antibodies are the serologic hallmark of this syndrome; however, many asymptomatic persons have antiphospholipid antibodies,[78] and antiphospholipid antibody tests can have false negative results. In some patients, a battery of tests may be needed to establish diagnosis; the dilute Russell viper venom time, an assay for circulating lupus anticoagulant, has been found to be among the more sensitive tests.[79]

Livedo Vasculitis

Livedo vasculitis, another disorder that has been associated with lupus, is characterized by painful recurrent ulcers over the lower legs and ankles. The ulcers heal, leaving white sclerotic scars. Affected patients often have livedo reticularis. This condition, also known as atrophie blanche, has been attributed to thrombotic processes rather than immune complex deposition or leukocytoclastic vasculitis.[80]

Neonatal Lupus

Neonatal lupus is a distinct syndrome of annular, erythematous macules and papules occurring on the face of newborn infants. The disorder has been attributed to transplacental passage of anti-Ro and occasionally anti-La antibodies. Mothers are often asymptomatic, but some may have lupus or Sjögren syndrome. Congenital heart block is the most serious complication of this disorder.[81]

The author has served as an investigator, consultant, or speaker for the following companies: Abbott Laboratories, Inc., Allergen, Inc., Amgen, Inc., Biogen, Inc., Centocor, Inc., Connetics Corporation, Fujisawa Healthcare, Inc., Galderma Laboratories, L.P., Genentech, Inc., Leo Pharmaceuticals, and Warner-Chilcott Pharmaceuticals.

The FDA has not approved the following drugs for specific uses described in this chapter: infliximab and TNF-α blockers for the treatment of sarcoidosis and Wegener granulomatosis; rituximab for the treatment of lymphomatoid granulomatosis; and calcipotriene for the treatment of morphea.

References

1. Sorrentino D, Avellini C, Zearo E: Colonic sarcoidosis, infliximab, and tuberculosis: a cautionary tale. Inflamm Bowel Dis 10:438, 2004

2. Khanna D, Liebling MR, Louie JS: Etanercept ameliorates sarcoidosis arthritis and skin disease. J Rheumatol 30:1864, 2003

3. Haley H, Cantrell W, Smith K: Infliximab therapy for sarcoidosis (lupus pernio). Br J Dermatol 150:146, 2004

4. Menon Y, Cucurull E, Reisin EE, et al: Interferon-alpha-associated sarcoidosis responsive to infliximab therapy. Am J Med Sci 328:173, 2004

5. English JC 3rd, Patel PJ, Greer KE: Sarcoidosis. J Am Acad Dermatol 44:725, 2001

6. Kleinert J, Lorenz M, Kostler W, et al: Refractory Wegener's granulomatosis responds to tumor necrosis factor blockage. Wien Klin Wochenschr 116:334, 2004

7. Sebire NJ, Haselden S, Malone M, et al: Isolated EBV lymphoproliferative disease in a child with Wiskott-Aldrich syndrome manifesting as cutaneous lymphomatoid granulomatosis and responsive to anti-CD20 immunotherapy. J Clin Pathol 56:555, 2003

8. Beaty MW, Toro J, Sorbara L, et al: Cutaneous lymphomatoid granulomatosis: correlation of clinical and biologic features. Am J Surg Pathol 25:1111, 2001

9. Zaidi A, Kampalath B, Peltier WL, et al: Successful treatment of systemic and central nervous system lymphomatoid granulomatosis with rituximab. Leuk Lymphoma 45:777, 2004

10. Davis MD, Daoud MS, McEvoy MT, et al: Cutaneous manifestations of Churg-Strauss syndrome: a clinicopathologic correlation. J Am Acad Dermatol 37:199, 1997

11. Seo P, Stone JH: The antineutrophil cytoplasmic antibody–associated vasculitides. Am J Med 117:39, 2004

12. Gong F, Shiraishi H, Momoi MY: Follow-up of coronary artery lesions caused by Kawasaki disease and the value of coronary angiography. Clin Med J (Engl) 115:681, 2002

13. Curtis N: Kawasaki disease and toxic shock syndrome: at last the etiology is clear? Adv Exp Med Biol 549:191, 2004

14. Newburger JW, Takahashi M, Gerber MA, et al: Diagnosis, treatment, and long-term management of Kawasaki disease: a statement for health professionals from the Committee on Rheumatic Fever, Endocarditis and Kawasaki Disease, Council on Cardiovascular Disease in the Young, American Heart Association. Circulation 110:2747, 2004

15. Oates-Whitehead RM, Baumer JH, Haines L, et al: Intravenous immunoglobulin for the treatment of Kawasaki disease in children. Cochrane Database Syst Rev (4):CD004000, 2003

16. Ringpfeil F, Lebwohl MG, Christiano AM, et al: Pseudoxanthoma elasticum: mutations in the MRP6 gene encoding a transmembrane ATP-binding cassette (ABC) transporter. Proc Natl Acad Sci USA 97:6001, 2000

17. Lebwohl MG, DiStefano D, Prioleau PG, et al: Pseudoxanthoma elasticum and mitral valve prolapse. N Engl J Med 307:228, 1982

18. Lebwohl M, Halperin J, Phelps RG: Brief report: occult pseudoxanthoma elasticum in patients with premature cardiovascular disease. N Engl J Med 329:1237, 1993

19. Lebwohl M, Phelps RG, Yannuzzi L, et al: Diagnosis of pseudo-xanthoma elasticum by scar biopsy in patients without characteristic skin lesions. N Engl J Med 317:347, 1987

20. D'Alessandro A, Muzi G, Monaco A, et al: Yellow nail syndrome: does protein leakage play a role? Eur Respir J 17:149, 2001

21. Litonjua P, Pinero-Pilona A, Aviles-Santa L, et al: Prevalence of acanthosis nigricans in newly-diagnosed type 2 diabetes. Endocr Pract 10:101, 2004

22. Verrotti A, Chiarelli F, Amerio P, et al: Necrobiosis lipoidica diabeticorum in children and adolescents: a clue for underlying renal and retinal disease. Pediatr Dermatol 12:220, 1995

23. Rho YW, Suhr KB, Lee JH, et al: A clinical observation of scleredema adultorum and its relationship to diabetes. J Dermatol 25:103, 1998

24. Dogra S, Handa S, Kanwar AJ: Dexamethasone pulse therapy for scleredema. Pediatr Dermatol 21:280, 2004

25. Bowen AR, Smith L, Zone JJ: Scleredema adultorum of Buschke treated with radiation. Arch Dermatol 139:780, 2003

26. Janiga JJ, Ward DH, Lim HW: UVA-1 as a treatment for scleredema. Photodermatol Photoimmunol Photomed 20:210, 2004

27. Hecker MS, Lebwohl MG: Recalcitrant pyoderma gangrenosum: treatment with thalidomide. J Am Acad Dermatol 38:490, 1998

28. Tan MH, Gordon M, Lebwohl O, et al: Improvement of pyoderma gangrenosum and psoriasis associated with Crohn disease with anti–tumor necrosis factor alpha monoclonal antibody. Arch Dermatol 137:930, 2001

29. Kanaseki T, Torigoe T, Hiroshashi Y, et al: Identification of germline mutation of PTEN gene and analysis of apoptosis resistance of the lymphocytes in a patient with Cowden disease. Pathobiology 70:34, 2002

30. Barax CN, Lebwohl M, Phelps RG: Multiple hamartoma syndrome. J Am Acad Dermatol 17:342, 1987

31. Garcia JM, Silva J, Pena C, et al: Promoter methylation of the PTEN gene is a common molecular change in breast cancer. Genes Chromosomes Cancer 41:117, 2004

32. Collin P, Pukkala E, Reunala T: Malignancy and survival in dermatitis herpetiformis: a comparison with coeliac disease. Gut 38:528, 1996

33. Scott RJ, Crooks R, Meldrum CJ, et al: Mutation analysis of the STK11/LKB1 gene and clinical characteristics of an Australian series of Peutz-Jeghers syndrome patients. Clin Genet 62:282, 2002

34. Mallipeddi R: Epidermolysis bullosa and cancer. Clin Exp Dermatol 27:616, 2002

35. Christiano AM, LaForgia S, Paller AS, et al: Prenatal diagnosis for recessive dystrophic epidermolysis bullosa in 10 families by mutation and haplotype analysis in the type VII collagen gene (COL7A1). Mol Med 2:59, 1996

36. Gardella R, Castiglia D, Posteraro P, et al: Genotype-phenotype correlation in Italian patients with dystrophic epidermolysis bullosa. J Invest Dermatol 119:1456, 2002

37. Chen M, Kasahara N, Keene DR, et al: Restoration of type VII collagen expression and function in dystrophic epidermolysis bullosa. Nat Genet 32:670, 2002

38. Dubrey SW, Burke MM, Hawkins PN, et al: Cardiac transplantation for amyloid heart disease: the United Kingdom experience. J Heart Lung Transplant 23:1142, 2004

39. Palladini G, Perfetti V, Perlini S, et al: The association of thalidomide and intermediate-dose dexamethasone is an effective but toxic treatment for patients with AL (primary) amyloidosis. Blood Nov 30, 2004 [epub ahead of print]

40. Topar G, Staudacher C, Geisen F, et al: Urticaria pigmentosa: a clinical, hematopathologic, and serologic study of 30 adults. Am J Clin Pathol 109:279, 1998

41. Horny HP, Ruck P, Krober S, et al: Systemic mast cell disease (mastocytosis): general aspects and histopathological diagnosis. Histol Histopathol 12:1081, 1997

42. Kiszewski AE, Duran-Mckinster C, Orozco-Covarrubias L, et al: Cutaneous mastocytosis in children: a clinical analysis of 71 cases. J Eur Acad Dermatol Venereol 18:285, 2004

43. Shady AA, Colby BR, Cunha LF, et al: Congenital erythropoietic porphyria: identification and expression of eight novel mutations in the uroporphyrinogen III synthase gene. Br J Haematol 117:980, 2002

44. Pottage JC Jr, Kessler HA: Herpes simplex virus resistance to acyclovir: clinical relevance. Infect Agents Dis 4:115, 1995

45. Chibo D, Mijch A, Doherty R, et al: Novel mutations in the thymidine kinase and DNA polymerase genes of acyclovir and foscarnet resistant herpes simplex viruses infecting an immunocompromised patient. J Clin Virol 25:165, 2002

46. Lalezari J, Schacker T, Feinberg J, et al: A randomized, double-blind, placebo-controlled trial of cidofovir gel for the treatment of acyclovir-unresponsive mucocutaneous herpes simplex virus infection in patients with AIDS. J Infect Dis 176:892, 1997

47. Chang CC, Chomel BB, Kasten RW, et al: Molecular epidemiology of *Bartonella henselae* infection in human immunodeficiency virus-infected patients and their cat contacts, using pulsed-field gel electrophoresis and genotyping. J Infect Dis 186:1733, 2002

48. Agan BK, Dolan MJ: Laboratory diagnosis of *Bartonella* infections. Clin Lab Med 22:937, 2002

49. Resnick JS Jr, Van Beek M, Furmanski L, et al: Etiology of pruritic popular eruption with HIV infection in Uganda. JAMA 292:2614, 2004

50. Albrecht D, Meyer T, Lorenzen T, et al: Epidemiology of HHV-8 infection in HIV-positive patients with and without Kaposi sarcoma: diagnostic relevance of serology and PCR. J Clin Virol 30:145, 2004

51. Parsonnet J: Nonmenstrual toxic shock syndrome: new insights into diagnosis, pathogenesis, and treatment. Curr Clin Top Infect Dis 16:1, 1996

52. Childers BJ, Potyondy LD, Nachreiner R, et al: Necrotizing fasciitis: a fourteen-year retrospective study of 163 consecutive patients. Am Surg 68:109, 2002

53. Nakafusa J, Misago N, Miura Y, et al: The importance of serum creatine phosphokinase level in the early diagnosis, and as a prognostic factor, of *Vibrio vulnificus* infection. Br J Dermatol 145:280, 2001

54. Steere AC, Sikand VK, Meurice F, et al: Vaccination against Lyme disease with recombinant *Borrelia burgdorferi* outer-surface lipoprotein A with adjuvant. Lyme Disease Vaccine Study Group. N Engl J Med 339:209, 1998

55. Lymerix: lack of demand kills Lyme disease vaccine. Nursing 32:18, 2002

56. Lam CW, Leung CY, Lee KC, et al: Novel mutations in the PATCHED gene in basal cell nevus syndrome. Mol Genet Metab 76:57, 2002

57. Smahi A, Courtois G, Rabia SH, et al: The NF-kappaB signalling pathway in human diseases: from incontinentia pigmenti to ectodermal dysplasias and immune-deficiency syndromes. Hum Mol Genet 11:2371, 2002

58. Dasgupta B, Dugan LL, Gutmann DH: The neurofibromatosis 1 gene product neurofibromin regulates pituitary adenylate cyclase-activating polypeptide-mediated signaling in astrocytes. J Neurosci 23:8949, 2003

59. Xiao GH, Chernoff J, Testa JR: NF2: the wizardry of Merlin. Genes Chromosomes Cancer 38:389, 2003

60. Nellist M, Sancak O, Goedbloed MA, et al: Distinct effects of single amino-acid changes to tuberin on the function of the tuberin-hamartin complex. Eur J Hum Genet 2004

61. Germain DP, Shabbeer J, Cotigny S, et al: Fabry disease: twenty novel alpha-galactosidase A mutations and genotype-phenotype correlations in classical and variant phenotypes. Mol Med 8:306, 2002

62. Desnick RJ, Brady R, Barranger J, et al: Fabry disease, an under-recognized multisystemic disorder: expert recommendations for diagnosis, management, and enzyme replacement therapy. Ann Intern Med 138:338, 2003

63. Daoud MS, Hutton KP, Gibson LE: Cutaneous periarteritis nodosa: a clinicopathological study of 79 cases. Br J Dermatol 136:706, 1997

64. Guillevin L, Lhote F, Amouroux J, et al: Antineutrophil cytoplasmic antibodies, abnormal angiograms and pathological findings in polyarteritis nodosa and Churg-Strauss syndrome: indications for the classification of vasculitides of the polyarteritis nodosa group. Br J Rheumatol 35:958, 1996

65. Poliak S, Lebwohl MG, Parris A, et al: Reactive perforating collagenosis associated with diabetes mellitus. N Engl J Med 306:81, 1982

66. Arch-Ferrer JE, Beenken SW, Rue LW, et al: Therapy for calciphylaxis: an outcome analysis. Surgery 134:941, 2003

67. Kasteler JS, Callen JP: Scalp involvement in dermatomyositis: often overlooked or misdiagnosed. JAMA 272:1939, 1994

68. Buchbinder R, Forbes A, Hall S, et al: Incidence of malignant disease in biopsy-proven inflammatory myopathy: a population-based cohort study. Ann Intern Med 134:1087, 2001

69. Hill CL, Zhang Y, Sigurgeirsson B, et al: Frequency of specific cancer types in dermatomyositis and polymyositis: a population based study. Lancet 357:96, 2001

70. Kanekura T, Fukumaru S, Matsushita S, et al: Successful treatment of scleroderma with PUVA therapy. J Dermatol 23:455, 1996

71. Kerscher M, Volkenandt M, Gruss C, et al: Low-dose UVA phototherapy for treatment of localized scleroderma. J Am Acad Dermatol 38:21, 1998

72. Tay YK: Topical calcipotriol ointment in the treatment of morphea. J Dermatolog Treat 14:219, 2003

73. Le CH, Morales A, Trentham DE: Minocycline in early diffuse scleroderma. Lancet 352:1755, 1998

74. Zikos P, van Lint MT, Frasoni F, et al: Low transplant mortality in allogeneic bone marrow transplantation for acute myeloid leukemia: a randomized study of low-dose cyclosporin versus low-dose cyclosporin and low-dose methotrexate. Blood 91:3503, 1998

75. Vogelsang GB, Wolff D, Altomonte V, et al: Treatment of chronic graft-versus-host disease with ultraviolet irradiation and psoralen (PUVA). Bone Marrow Transplant 17:1061, 1996

76. Yamane T, Yamamura R, Aoyama Y, et al: Infliximab for the treatment of severe steroid refractory acute graft-versus-host disease in three patients after allogeneic hematopoietic transplantation. Leuk Lymphoma 44:2095, 2003

77. Kim SW, Rice L, Champlin R, et al: Aplastic anemia in eosinophilic fasciitis: responses to immunosuppression and marrow transplantation. Haematologia (Budap) 28:131, 1997

78. Tektonidou MG, Sotsiou F, Nakopoulou L, et al: Antiphospholipid syndrome nephropathy in patients with systemic lupus erythematosus and antiphospholipid antibodies: prevalence, clinical associations, and long-term outcome. Arthritis Rheum 50:2569, 2004

79. Proven A, Bartlett RP, Moder KG, et al: Clinical importance of positive test results for lupus anticoagulant and anticardiolipin antibodies. Mayo Clin Proc 79:467, 2004

80. McCalmont CS, McCalmont TH, Jorizzo JL, et al: Livedo vasculitis: vasculitis or thrombotic vasculopathy? Clin Exp Dermatol 17:4, 1992

81. Brucato A, Franceschini F, Buyon JP: Neonatal lupus: long-term outcomes of mothers and children and recurrence rate. Clin Exp Rheumatol 15:467, 1997

II PAPULOSQUAMOUS DISORDERS

Elizabeth A. Abel, m.d.

Papulosquamous disorders comprise a group of dermatoses that have distinct morphologic features.[1] The characteristic primary lesion of these disorders is a papule, usually erythematous, that has a variable amount of scaling on the surface. Plaques or patches form through coalescence of the primary lesions. Some common papulosquamous dermatoses are pityriasis rosea, lichen planus, seborrheic dermatitis, tinea corporis, pityriasis rubra pilaris, psoriasis [*see 2:III Psoriasis*], and parapsoriasis. Drug eruptions, tinea corporis, and secondary syphilis may also have a papulosquamous morphology. Some papulosquamous disorders may be a cutaneous manifestation of AIDS.[2]

Pityriasis Rosea

Pityriasis rosea is a relatively common, self-limited, exanthematous disease characterized by oval papulosquamous lesions on the trunk and proximal areas of the extremities. Pityriasis rosea typically appears during the spring and fall in temperate climates[3]; its incidence is highest in persons between 10 and 35 years of age.[4]

A population-based 10-year epidemiologic survey identified 939 patients with pityriasis rosea, about one third of whom had antecedent acute infection or atopy.[5] It also showed that peak incidence occurred at 20 to 24 years of age, that the incidence was higher in colder months, and that recurrences were rare. Occurrences among household contacts are uncommon. This study also noted that the incidence of disease had appeared to decline.

etiology

A viral etiology has been suggested for pityriasis rosea on the basis of immunologic and histologic data. The superficial dermis contains aggregates of $CD4^+$ helper T cells in perivascular locations and increased numbers of Langerhans cells. It has been postulated that IgM antibodies to keratinocytes cause the secondary form of the eruption. An association between human herpesvirus type 7 (HHV-7) and pityriasis rosea was initially reported in 1997.[6] Studies using polymerase chain reaction and immunohistochemical analyses of tissue samples to detect HHV-7 DNA sequences and antigens have provided inconclusive evidence of a causal relationship between HHV-7 and pityriasis rosea. In a retrospective study of 13 patients and 14 control subjects, the prevalence of HHV-7 was lower in lesional skin of patients with pityriasis rosea than in control subjects.[7] A subsequent seroepidemiologic study of HHV-6 and HHV-7 was conducted in 44 patients with pityriasis rosea and in 25 patients with other skin eruptions. Although in this study several patients with pityriasis rosea had antibody titers consistent with active infection, the overall prevalence of HHV-6 and HHV-7 was no greater in patients with pityriasis rosea than in control subjects.[8] A meta-analysis reviewed the data from 13 studies and found insufficient evidence to support a causal relationship between HHV-7 infection and pityriasis rosea.[9] A viral etiology of pityriasis rosea thus remains elusive. Certain drugs that cause a pityriasis rosea–like eruption have been implicated in the etiology of this disorder. These drugs include the antihypertensive agent captopril, metronidazole, isotretinoin (13-*cis*-retinoic acid), penicillamine, arsenic, gold, bismuth, barbiturates, and clonidine.[4]

diagnosis

The primary lesion, called a herald patch, appears first as a slightly raised, salmon-colored oval patch with a fine, wrinkled scale resembling cigarette paper. Typically, 7 to 10 days after the appearance of the herald patch, there occurs a bilaterally symmetrical eruption of smaller lesions; this secondary eruption occurs mainly on the trunk and upper extremities [*see Figure 1*]. Secondary lesions tend to follow cleavage lines (Langer lines) in a so-called fir tree distribution. A V-shaped formation on the upper chest and upper back, a circumferential pattern around the shoulders and hips, and a transverse pattern on the lower anterior trunk and lower back are seen in most patients.[10] The lesions are occasionally pruritic. The secondary rash is frequently more helpful in making a diagnosis than the initial herald patch, which is often misdiagnosed.[10] Atypical manifestations occur in 20% of persons affected. Such manifestations include a purpuric form of pityriasis rosea that resembles vasculitis, as well as papular, vesicular, pustular, and urticarial forms. An inverse variant of pityriasis rosea, more common in children than in adults, is characterized by lesions on the face and extremities, with relatively few lesions appearing on the trunk.[4]

differential diagnosis

Because lesions of pityriasis rosea may closely resemble those of secondary syphilis, a serologic test for syphilis may be indicated. Lesions may also resemble tinea corporis or tinea versicolor and should be examined by fungal scrapings and potassium hydroxide (KOH) wet mounts. A careful drug history must be obtained to exclude the possibility of a drug eruption.

treatment

Pityriasis rosea lesions resolve spontaneously after 6 to 8 weeks. The patient should be reassured that the disorder is benign and self-limited; such reassurance, together with educating the pa-

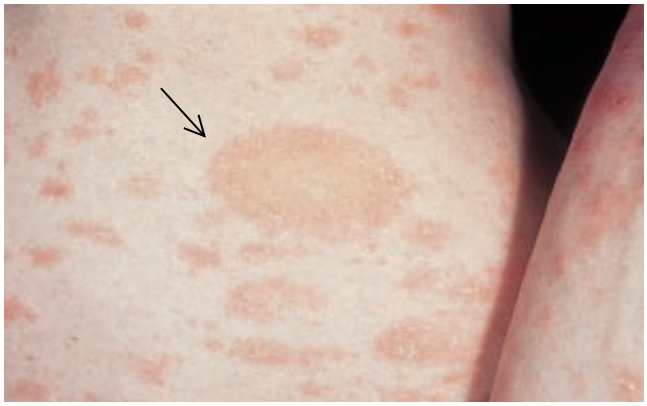


Figure 1 **Pityriasis rosea commonly presents as a single, large salmon-colored plaque called a herald patch (arrow). Appearance of the isolated lesion is followed in a week to 10 days by a bilaterally symmetrical papulosquamous eruption, mainly on the trunk and upper extremities.**

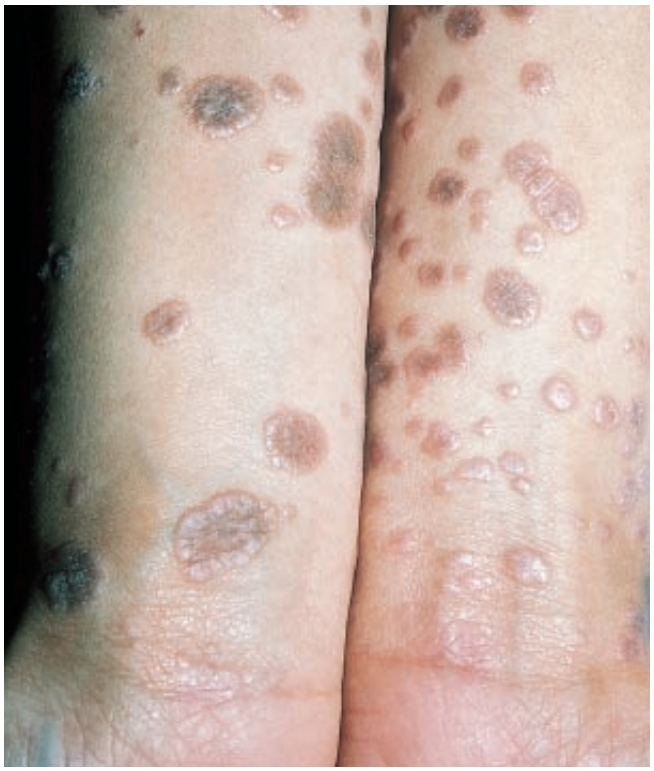


Figure 2 **Violaceous, flat-topped, polygonal papules are typical of lichen planus. A common location is the flexor aspect of the wrists and forearms.**

tient about the disease, is the most important aspect of treatment. Lesions are variably pruritic. Symptoms should be treated with bland emollients or systemic antipruritics. Sun exposure may accelerate clearing. Irradiation with ultraviolet B (UVB) sunlamps is beneficial in decreasing the severity of disease, especially when treatment is initiated within the first week of the eruption. One study found that 10 erythemogenic exposures of UVB substantially decreased the extent of pityriasis rosea, although it neither altered the duration of the disorder nor improved the itching.[11] Other evidence suggests that UVB therapy may hasten resolution of the rash but may cause hyperpigmentation.[12]

In a double-blind, placebo-controlled study in India, oral erythromycin administered in divided doses for 14 days was effective in treating patients with pityriasis rosea.[13] In this cohort, upper respiratory tract infections preceded the skin eruption in 68.8% of the 90 patients. A complete response, with complete resolution of skin lesions occurring within 2 weeks, was reported in 33% of the treatment group, as compared with 0% in the placebo group. The duration of disease was comparable for the two groups of patients. Although not all patients with pityriasis rosea benefit from erythromycin therapy, a trial of erythromycin is a safe treatment approach.

Lichen Planus

Lichen planus is a localized or generalized eruption with violaceous, flat-topped, polygonal papules and little or no observable scaling [*see Figure 2*]. It is often localized to the oral mucosa; 25% of patients with oral lichen planus have skin involvement as well.[14] The incidence is highest in young to middle-aged persons. Lichen planus usually appears in the fifth or sixth decade and affects women more often than men.

ETIOLOGY

The etiology of lichen planus is unknown. An alteration in basal keratinocytes that induces humoral and cell-mediated immune responses has been postulated as a mechanism. Skin and mucous membrane lesions resembling lichen planus have been observed in patients with graft versus host disease (GVHD) [*see 2:VI Cutaneous Adverse Drug Reactions*]. Lichen planus has also been associated with other immune-mediated diseases, including ulcerative colitis, bullous pemphigoid, myasthenia gravis with thymoma, primary biliary cirrhosis, and chronic active hepatitis.[15]

There is an increased prevalence of viral hepatitis, especially hepatitis C, in patients with lichen planus. In a multicenter study of 303 sequential patients with lichen planus, the prevalence of hepatitis C virus (HCV) was 19.1%, compared with 3.2% in control subjects.[16] The role of HCV in the pathogenesis of lichen planus is not clearly understood; some investigators suggest that the cause of lichen planus may be related to the pattern of immune dysregulation induced by HCV.[17] There are a number of reports of lichen planus occurring after administration of different types of hepatitis B vaccine.[18] This is a rare occurrence, considering the widespread use of this vaccine; several cases have been reported from France and Italy, and one case has been reported from the Middle East. An immunologic mechanism has been postulated as the cause. The latency period ranges from several days to 3 months after any one of the three usual injections of vaccine.

A variety of drugs have been reported to cause lichenoid reactions in the skin, usually sparing the mucous membranes. Such drugs include beta blockers, methyldopa, penicillamine, quinidine, and quinine. Other drugs that have been implicated but for which causal evidence is insufficient include angiotensin-converting enzyme inhibitors, sulfonylurea agents, carbamazepine, gold, and lithium.[19] In one study, the administration of penicillamine for primary biliary cirrhosis was followed by the development of lichen planus in 17 of 24 patients[20]; in addition, after treatment with penicillamine, the skin eruption became worse in three of seven patients with biliary cirrhosis and preexisting lichen planus. Nonsteroidal anti-inflammatory drugs have been documented to cause a lichenoid drug eruption; these drugs include naproxen, indomethacin, diflunisal, ibuprofen, acetylsalicylic acid, and salsalate.[21] Although the latency period is highly variable, symptoms usually develop within a few months after drug initiation and resolve within weeks to months after discontinuance of the offending agent.

DIAGNOSIS

Lichen planus appears as flat-topped, shiny, violaceous papules, often with a fine, reticulated scale on the surface. Common sites of involvement include the skin, nails, mucous membranes, vulva, and penis. Wickham striae—white, lacy patterns on the papule surface—are apparent on magnification with a hand lens.[22] The occurrence of papules along a scratch line, as in linear lichen planus, is referred to as the Koebner phenomenon [*see Figure 3*]. In the hypertrophic form of the disease, papules coalesce to form thick plaques or nodules that are often found on the lower extremities. Pruritus may be severe, particularly in the generalized or hypertrophic forms of the disease. Common sites of involvement are the flexor surfaces of the wrists, the sacrum, the

mucous membranes of the mouth, the medial thighs, and the genitalia. Mucous membrane lesions show a white, reticulated mosaic pattern [*see Figure 4*]. A severe erosive form of lichen planus can involve the oral mucous membranes. In rare cases, lesions occur in the esophagus, causing esophageal stricture and dysphagia.[23]

A follicular form known as lichen planopilaris may result in scarring alopecia. Variants of lichen planus with distinct morphologic features include actinic, annular, bullous, hypertrophic, linear, ulcerative, and zosteriform forms. The nails may also be involved [*see 2:XIV Diseases of the Nail*]. The clinical features of some forms of lichen planus may resemble those of lupus erythematosus.[22]

Skin biopsy confirms the clinical diagnosis of lichen planus. Typically, the epidermis shows hyperkeratosis, a prominent granular layer, liquefaction degeneration of the basal cell layer, and an intense upper dermal inflammatory infiltrate. Immunoperoxidase studies using monoclonal antibodies to cell surface antigens have shown that most cells in the infiltrate are of the helper-inducer T cell subset. Colloid bodies (Civatte bodies) coated with immunoglobulin are frequently seen in the dermal papillae. On ultrastructural examination, numerous Langerhans cells can be observed at the dermoepidermal junction.

TREATMENT

Limited data exist for making evidence-based recommendations regarding treatment of lichen planus.[24,25] Definitive clinical trials have not been performed, and information on the efficacy of treatments is derived from small trials and anecdotal evidence.

Body Lesions

Emollients, topical glucocorticoids, a short course of systemic corticosteroids, and systemic antipruritics have been used to treat cutaneous lichen planus. Most experts recommend medium- to high-potency topical corticosteroids as first-line treatment for localized cutaneous lichen planus. Oral corticosteroids may be used for generalized cutaneous lichen planus. As an alternative to systemic corticosteroids, systemic retinoids, such as acitretin, are beneficial in some patients with cutaneous forms of lichen planus.[26] Azathioprine has been used for its steroid-sparing effect in erosive and generalized lichen planus.[27]

Other therapies that have been reported to have efficacy in the treatment of cutaneous lichen planus include phototherapy (psoralen plus ultraviolet A [PUVA]), cyclosporine, and hydroxychloroquine.[24] In a trial of oral psoralen photochemotherapy for widespread recalcitrant lichen planus, clinical remission occurred in six of seven patients and correlated with the disappearance of the upper dermal infiltrate.[28] Oral cyclosporine has also been effective, but potential renal toxicity and hypertension limit its long-term use.[22] Recombinant interferon alfa–2b, administered subcutaneously every other day, was successful in the treatment of generalized lichen planus in three patients with no evidence of hepatitis, further supporting the cell-mediated immunologic etiology of this disease.[29]

Mouth Lesions

For lichen planus that is localized to the oral mucosa, a high-potency corticosteroid such as clobetasol in a vehicle that is adherent to the mucosal surface (Orabase) is helpful.[24,29] Intralesional injections of corticosteroids may be used to treat localized, recalcitrant lesions. Use of miconazole gel in combination with chlorhexidine mouth rinses is effective for prophylaxis against oral candidiasis.[30] Topical isotretinoin gel is an effective alternative to corticosteroids, although relapses often occur after discontinuance of this medication.[31] In a double-blind, placebo-controlled study of 22 patients with biopsy-proven oral lichen planus, an 8-week course of 0.1% isotretinoin gel was found to be effective.[31] Cyclosporine mouth rinses have been helpful for some patients. A 6-month course of hydroxychloroquine, 200 to 400 mg daily, was successful in nine of 10 patients with oral lichen planus; ulcers healed and pain decreased after 1 to 2 months.[32] Topical tacrolimus, a macrolide that suppresses T cell activation, was used to treat erosive mucosal lichen planus in 19 patients whose conditions were resistant to conventional treatment. Therapeutic levels of tacrolimus were demonstrated in eight patients, and areas of ulceration showed a mean decrease of 73.3% over the 8-week study period; however, 13 of 17 patients suffered a relapse after cessation of therapy.[33] Topical pimecrolimus cream is being evaluated as a treatment for oral erosive lichen planus.[34] The role of these agents in the treatment of lichen planus must be further investigated, particularly in view of the alert by the Food and Drug Administration issued in 2005 regarding a possible link between use of topical tacrolimus and pimecrolimus and cases of lymphoma and skin cancer.

Genital and Perianal Lesions

Mild, nonerosive disease can be controlled with topical corticosteroids; erosive disease, although more difficult to treat, may

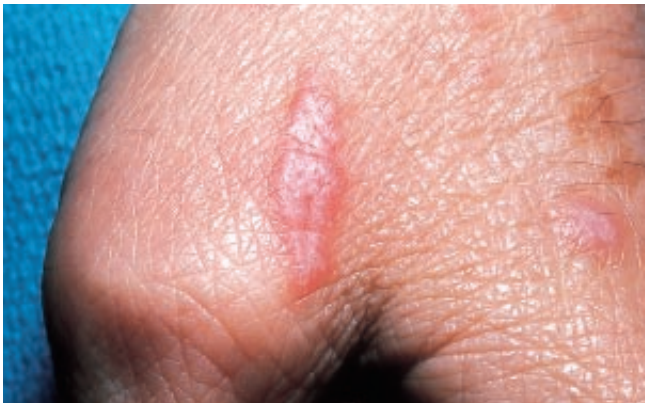


Figure 3 **The Koebner phenomenon—the appearance of lesions along a scratch line—may be seen in patients with lichen planus.**

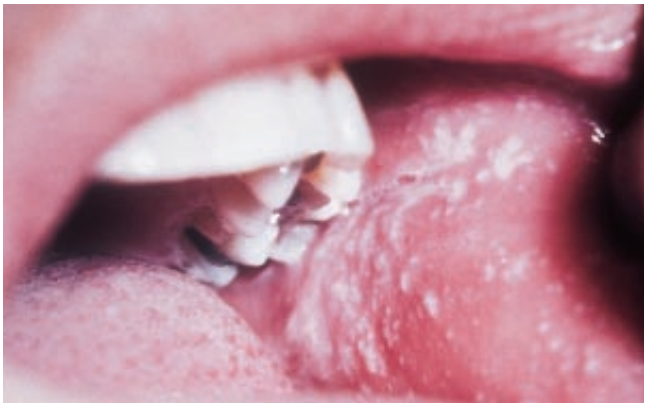


Figure 4 **Lichen planus of the mucous membrane assumes a white, reticulated mosaic pattern, as seen above on the buccal mucosa.**

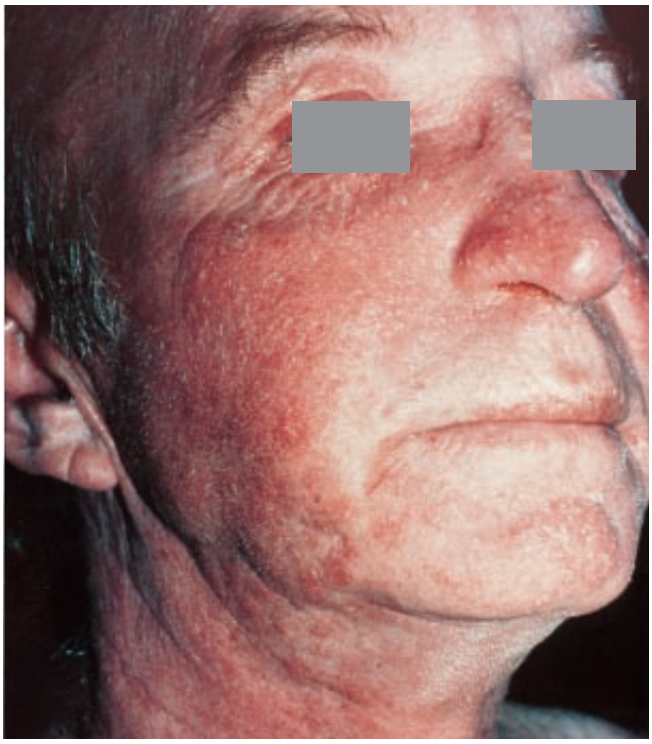


Figure 5 **Seborrheic dermatitis seen on the face of this patient involves sites of sebaceous gland activity.**

be treated with topical corticosteroids in combination with other topical or systemic medications.[35] Topical tacrolimus[36,37] and topical pimecrolimus[38] may be useful in the treatment of recalcitrant cases that have failed to respond to other therapies.

PROGNOSIS

Patients who experience an acute outbreak of lichen planus have a good prognosis; in most cases, the papules clear within several months to a year. The chronic form, however, may last for 10 years or longer. In a study following the long-term course of lichen planus in 214 patients for 8 to 12 years, lichen planus cleared in two thirds of the patients within 1 year. The recurrence rate was 49%, which was higher than recurrence rates reported in previous studies; the authors attributed the high rate of recurrence to treatment with potent topical corticosteroids.[39]

Seborrheic Dermatitis

Seborrheic dermatitis is a papulosquamous condition that is often associated with excessive oiliness or seborrhea, dandruff, and well-defined red, scaly patches on the face, trunk, and intertriginous areas.[40] Some cases may progress to a severe exfoliative erythroderma. Seborrheic dermatitis is a common skin disorder that occurs in otherwise healthy adults. It is increasingly prevalent in middle-aged and elderly persons. Seborrheic dermatitis does not occur before puberty except during infancy (usually between 2 and 12 weeks of age), at which time transplacentally derived maternal hormones are present. The prognosis in adults is one of lifelong recurrence, with each episode lasting weeks to months.

ETIOLOGY

The cause of seborrheic dermatitis is unknown. An occasional association with neurologic abnormalities, especially parkinsonism, has been observed. Genetic predisposition, emotional stress, diet, hormones, and climatic factors may also influence this disorder. It is thought that an association exists between the yeast-like organism *Pityrosporum* and seborrheic dermatitis. Seborrheic dermatitis may, in part, be the result of an abnormal or inflammatory immune response to this yeast, or it may be caused by an epidermal hyperproliferation of this organism.[41]

Patients with classic seborrheic dermatitis may have normal or reduced rates of sebum excretion; therefore, seborrhea is not essential for the development of this disorder.[42] However, seborrhea may play a role in the seborrheic dermatitis present in certain patients, such as those with parkinsonism. Reduction of seborrhea with improvement of the dermatitis has been observed after a favorable neurologic response to levodopa treatment for parkinsonism.

DIAGNOSIS

The scale associated with seborrheic dermatitis may be yellowish and either dry or greasy. Sites of predilection are the areas of sebaceous gland activity [*see Figure 5*], such as the scalp, eyebrows, eyelids, forehead, nasolabial folds, and presternal or interscapular regions. Blepharitis involves granular inflammation of the lid margin, with scaling and shedding of debris into the eye, which may cause conjunctivitis. Seborrheic dermatitis is the most common cause of otitis externa. When the scalp is involved, lesions often extend along the frontal hairline, forming a band of erythema. The postauricular area is a common site of involvement. Lesions of the trunk may consist of erythematous follicular papules covered by greasy scales, which may coalesce to form large plaques or circinate patches. Seborrheic dermatitis can be seen in areas of male pattern baldness, but it is not a cause of hair loss unless there has been a severe intervening secondary infection resulting in a scarring alopecia.

DIFFERENTIAL DIAGNOSIS

Seborrheic dermatitis should be considered in the differential diagnosis of chronic eczematous dermatitis and in that of papulosquamous disorders, particularly psoriasis. The clinical features of seborrheic dermatitis limited to the scalp and face may resemble those associated with psoriasis, giving rise to the term sebopsoriasis. Histologic features range from psoriasiform changes of acanthosis and parakeratosis to the spongiosis of eczema. Seborrheic dermatitis of the face may resemble the facial lesions found in lupus erythematosus or other photosensitivity dermatoses. Lesions on the trunk may be confused with tinea versicolor, but the latter is easily excluded by skin scraping and KOH preparation or Wood light examination. Atopic dermatitis and psoriasis, especially when partially treated, are also included in the differential diagnosis.

SEBORRHEIC DERMATITIS ASSOCIATED WITH AIDS

Severe seborrheic dermatitis can be one of the most common and earliest manifestations of AIDS. From 30% to 80% of patients with AIDS have seborrheic dermatitis, compared with 3% to 5% of HIV-negative young adults.[43] Lesions may be explosive in onset and are often resistant to therapy. Clinical features include a predominantly inflammatory papular eruption on the face, with a tendency to spare the scalp, in contrast to the mild erythema and scaling of the scalp typical of seborrheic dermatitis in persons without AIDS. Truncal involvement in seborrheic areas is common in AIDS patients, and the lesions may resemble psoriasis. Although the cause of the association of seborrheic dermatitis

with AIDS is unknown, immunologic dysfunction may lead to an overgrowth of the yeast *P. orbiculare* in seborrheic areas.

Skin biopsy specimens from AIDS patients with seborrheic dermatitis have distinct histologic features, including keratinocyte necrosis, leukoexocytosis, a superficial perivascular infiltrate of plasma cells, and, frequently, neutrophils.[44]

TREATMENT

The condition on the scalp usually responds well to frequent—as often as daily—shampooing with a preparation containing 3% to 5% sulfur and 2% to 3% salicylic acid. Good response has also been reported with use of ciclopirox 1% shampoo once or twice weekly.[45] For the face and nonhairy areas, a mild cream containing precipitated 3% sulfur and 3% salicylic acid is effective. Involved areas also respond well to low-potency topical glucocorticoids, such as 1% hydrocortisone cream or desonide cream. Caution, however, must be exercised in the use of high-potency fluorinated steroid preparations, especially on the face and in skin folds; prolonged application may lead to chronic skin changes, such as atrophy and telangiectasia. Wet dressings followed by a topical antibiotic preparation are helpful in treating intertriginous areas, in which maceration and superficial secondary infection may occur.

Topical antifungal agents have been used in the treatment of seborrheic dermatitis. In addition to their antifungal properties, certain azoles (e.g., bifonazole, itraconazole, and ketoconazole) have demonstrated anti-inflammatory activity, which may be beneficial in alleviating symptoms.[46] In one study, 575 patients with seborrheic dermatitis underwent twice-weekly treatments with 2% ketoconazole shampoo; an excellent response was seen in 88% of the patients.[47] Continued prophylactic treatment once weekly over 6 months was helpful in preventing relapse of the disorder in a significant number of patients.

In a trial of 38 patients with seborrheic dermatitis, 1% metronidazole gel was found to be effective. Improvement was noted after 2 weeks, and marked improvement or complete clearing was noted at 8 weeks[48]; however, in a randomized, controlled trial, metronidazole 0.75% gel and placebo were found to have similar efficacy.[49] In a small randomized, open-label clinical trial, pimecrolimus 1% cream and betamethasone 1% cream were both found to be effective in reducing symptoms of seborrheic dermatitis, but relapses were observed more frequently with betamethasone.[50] A multicenter, randomized, controlled trial found oral terbinafine (an antimycotic allylamine compound) to be effective in patients with moderate to severe seborrheic dermatitis.[51]

Seborrheic blepharitis may be treated by applying baby shampoo with a cotton-tipped applicator to debride scales. If topical corticosteroids are required, the patient should be referred to an ophthalmologist to monitor potential side effects to the eye, such as increased intraocular pressure, glaucoma, cataracts, and activation of latent herpes infection.[1]

Treatment of HIV-associated seborrheic dermatitis is similar to that of seborrheic dermatitis in general, although HIV-associated seborrheic dermatitis is apt to be recalcitrant, requiring intensive, prolonged therapy. Treatment of the underlying HIV infection may lead to improvement of the associated seborrheic dermatitis.

Pityriasis Rubra Pilaris

Pityriasis rubra pilaris is a relatively uncommon chronic inflammatory dermatosis that is considered to be a disorder of keratinization. The age distribution is bimodal, occurring either in childhood or in the fifth decade; the clinical course is variable. An autosomal dominant inheritance has been postulated for the juvenile form of the disease.[52] Patients with the classic adult form of the disease have the best prognosis; resolution usually occurs over a 3-year period.

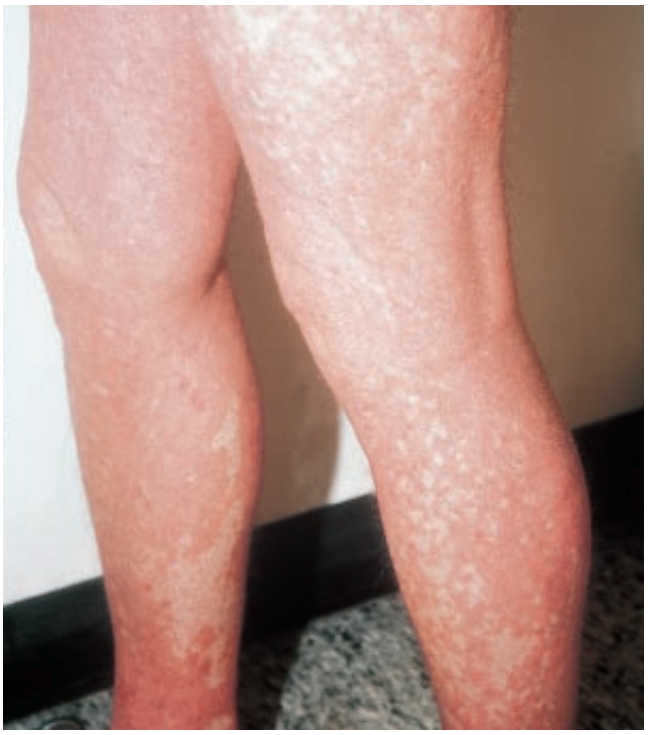


Figure 6 **Islands of spared skin within a background of diffuse erythema are present on the legs of this patient with pityriasis rubra pilaris.**

DIAGNOSIS

Typically, pityriasis rubra pilaris initially manifests itself as a seborrheic dermatitis–like eruption that occurs on sun-exposed areas of the body; this eruption is followed by the development of follicular papules that coalesce into psoriasiform patches on the trunk and extremities, with progression to erythroderma. Generalized involvement is characterized by yellow-orange erythema with desquamation. Diffuse areas of involvement generally show islands of spared skin [*see Figure 6*]. Additional features are palmoplantar hyperkeratosis [*see Figure 7*] and prominent follicular plugging over the dorsal aspects of the fingers. Pruritus is usually mild or absent. A pityriasis rubra pilaris–like eruption with follicular hyperkeratosis is a little known but distinctive cutaneous manifestation of dermatomyositis.[53]

TREATMENT

The response of patients with pityriasis rubra pilaris to conventional antipsoriatic therapies, such as topical corticosteroids, tars, and oral methotrexate, is often unsatisfactory; some patients, however, have shown a favorable response to topical calcipotriene (known outside the United States as calcipotriol).[54] UVB phototherapy may exacerbate the disease.[55] High-dose vitamin A in excess of 200,000 IU daily has been used but can cause liver or central nervous system toxicity. An oral retinoid such as acitretin or isotretinoin is indicated for the treatment of pityriasis rubra pilaris in men and postmenopausal women. In an early study in-

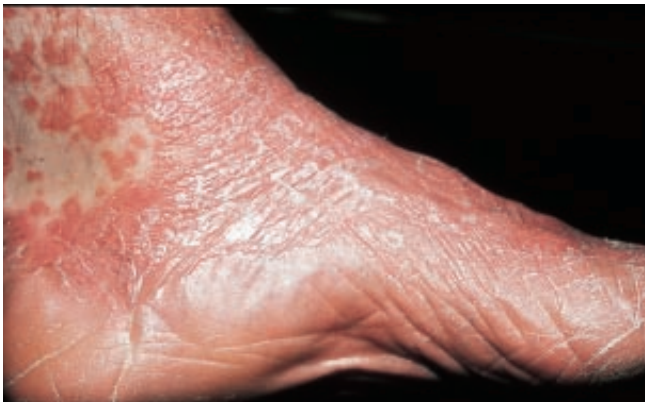


Figure 7 **Plantar hyperkeratosis and confluent erythematous follicular papules typical of pityriasis rubra pilaris are seen on the ankle and foot of this patient.**

volving 45 patients with pityriasis rubra pilaris, isotretinoin produced definite improvement in 50% of the patients after 4 weeks of therapy.[56] Remission of up to 6 months was sustained in some patients after the drug was withdrawn. Long-term use of this drug in patients with keratinizing disorders has been associated with irreversible skeletal toxicity. Because teratogenicity is a concern, women of childbearing age must use effective birth control with either agent.

In a study of patients with pediatric pityriasis rubra pilaris, isotretinoin achieved the best response among a range of therapies including steroids, systemic retinoids, and methotrexate; five of six patients treated with isotretinoin showed 90% to 100% clearing of lesions within 6 months of initiation of treatment.[57] Cyclosporine, 5 mg/kg/day, was effective in the treatment of three adult patients with pityriasis rubra pilaris, with a favorable response being noted within 2 to 4 weeks of initiation of therapy; however, relapse occurred when the dose was decreased to 1.2 mg/kg/day.[58]

Numerous reports have suggested that infliximab, a monoclonal antibody that binds to tumor necrosis factor–α, may be useful in the treatment of cutaneous inflammatory diseases, such as pityriasis rubra pilaris.[59] The drug is currently approved by the FDA for the treatment of rheumatoid arthritis and Crohn disease; further investigation is required to determine its efficacy for cutaneous dermatoses.

Parapsoriasis

Parapsoriasis encompasses a variety of relatively uncommon chronic inflammatory dermatoses of unknown etiology that are resistant to conventional treatment. Despite the designation parapsoriasis, the clinical appearance of the noninfiltrated scaly patches or plaques is distinct from that of psoriatic lesions. Classification of these disorders is controversial and is further complicated by the use of several terms to denote a single entity and by the use of various systems of nomenclature. A proposed standard nomenclature divides parapsoriasis into two distinct subgroups: pityriasis lichenoides, which may be acute or chronic, and small- and large-plaque parapsoriasis.[60]

PITYRIASIS LICHENOIDES

Diagnosis

The acute form of pityriasis lichenoides, also known as pityriasis lichenoides et varioliformis acuta (PLEVA) or Mucha-Habermann disease, is characterized by the abrupt onset of a generalized eruption of reddish-brown maculopapules that evolve during a period of weeks to months. Lesions are typically present at all stages of evolution and may be vesicular, hemorrhagic, crusted, or necrotic [*see Figure 8*]. Healing with varioliform scarring is common. Nonspecific histologic features include intraepidermal lymphocytes and erythrocytes, dermal hemorrhage, and a lymphocytic vasculitis.[61] Skin lesions of PLEVA may resemble those of lymphomatoid papulosis, which has immunohistologic features of a $CD30^+$ cutaneous T cell lymphoma.[62] Lymphomatoid papulosis occurs as a chronic, recurrent, self-healing papulonodular eruption; an association with mycosis fungoides has been observed in some patients. T cell clonality has been documented by PCR in 20 patients with PLEVA; similar findings have been made in patients with lymphomatoid papulosis.[63] Investigators have suggested that PLEVA is a lymphoproliferative process rather than an inflammatory reaction to various trigger factors, such as infectious agents. One case report demonstrated that pityriasis lichenoides lesions evolved concomitantly with a known Epstein-Barr virus (EBV)–mediated disease (i.e., acute infectious mononucleosis), suggesting that pityriasis lichenoides may be caused by EBV infection.[64]

A chronic form of pityriasis lichenoides, pityriasis lichenoides chronica, shows milder skin changes without necrosis. Lesions evolve during a period of weeks and may recur over many years.

Treatment

Treatment of both acute and chronic forms of pityriasis lichenoides is generally unsatisfactory. Topical corticosteroids, tars, and systemically administered methotrexate have all been tried, with variable success. Ultraviolet radiation from sunlamps[65] and oral psoralen photochemotherapy[66] may have a beneficial effect on the course of disease. High-dose tetracycline, 2 g/day for 1 month or more,[67] and minocycline, 100 mg once or twice daily, have also

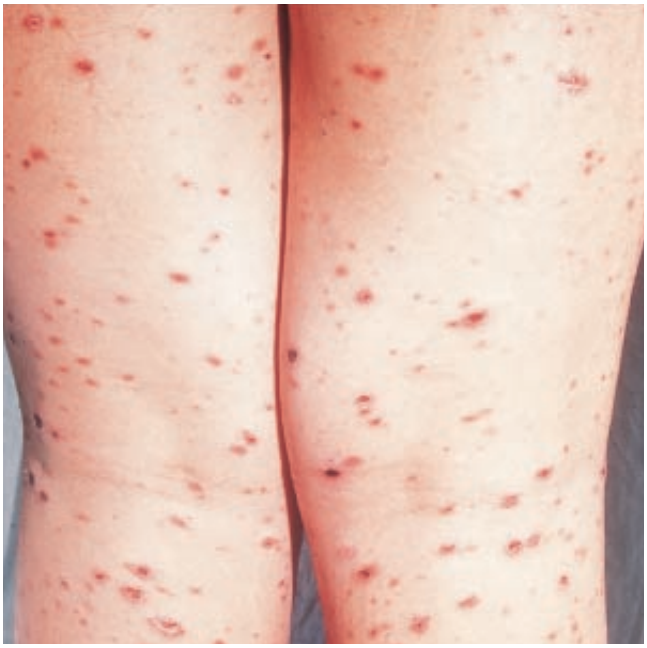


Figure 8 **Hemorrhagic brown-crusted varioliform papules are present on the lower legs of this patient with the acute form of pityriasis lichenoides.**

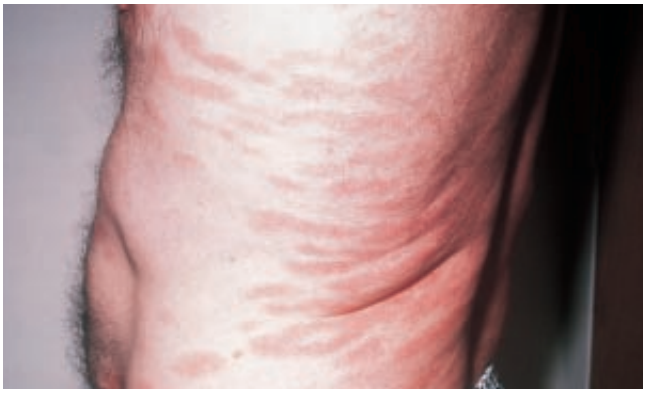


Figure 9 **The digitate variant of small-plaque parapsoriasis is seen in this patient.**

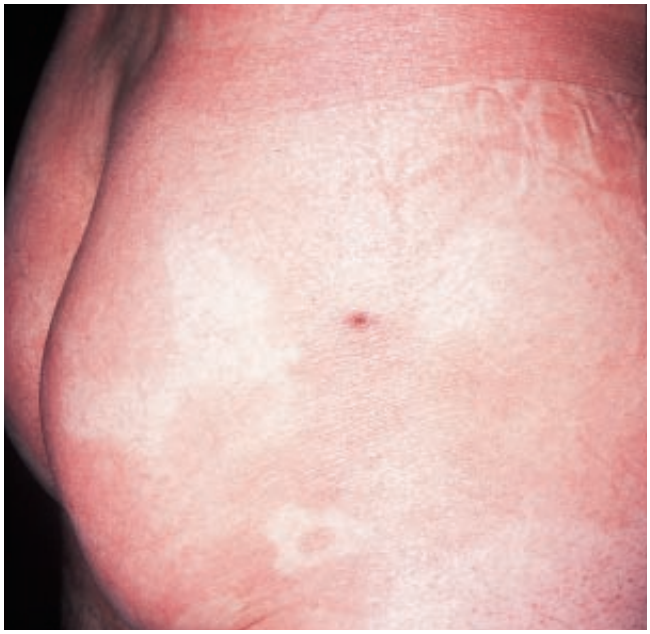


Figure 10 **Large-plaque parapsoriasis as seen on the buttocks of this patient may eventuate in cutaneous T cell lymphoma.**

been shown to be effective treatments. A rare type of Mucha-Habermann disease known as Degos disease (also called malignant atrophic papulosis), characterized by fever and hemorrhagic and papulonecrotic lesions, responds rapidly to the administration of methotrexate [*see 15:VIII Systemic Vasculitis Syndromes*].[68]

SMALL- AND LARGE-PLAQUE PARAPSORIASIS

Diagnosis

Small-plaque parapsoriasis consists of slightly scaly, thin, oval erythematous plaques of less than 5 cm in diameter, commonly located on the trunk and proximal extremities. The variant—digitate dermatosis—shows elongated lesions falling along lines of skin cleavage. The two diseases follow similar chronic, benign courses [*see Figure 9*].

Clinically, large-plaque parapsoriasis consists of slightly thickened, red-brown, scaly plaques that are more than 10 cm in diameter and have ill-defined borders; such lesions are present mainly on the proximal extremities and the buttocks and on the breasts of women [*see Figure 10*]. Frequently, there is a component of poikiloderma, which includes mottled hyperpigmentation and hypopigmentation, atrophy, and telangiectasia. Early lesions may show a nonspecific histology; late lesions show atypical lymphocytes within the epidermis.

It is important to differentiate large-plaque parapsoriasis from the small-plaque form because about 10% of cases of large-plaque parapsoriasis result in a cutaneous T cell lymphoma (mycosis fungoides).[60] Large-plaque lesions may be present for many years before malignant transformation is recognized histologically. The malignant change is suggested clinically by increased pruritus and progressive induration of lesions. The retiform variant may show prominent poikiloderma with atrophy and has a greater potential for malignant transformation.[69] Studies of T cell subsets using monoclonal antibodies to membrane markers have shown a variable predominance of helper T cells in the cutaneous infiltrates in atrophic parapsoriasis; such findings suggest a similarity to lesions of mycosis fungoides, although epidermotropism is absent.[70] Patients with this form of the disease should be evaluated with repeated biopsies of untreated lesions. Once a definitive diagnosis of mycosis fungoides has been established, specific treatment of this disease may be instituted.

Treatment

Treatment of large- and small-plaque parapsoriasis is similar to that of pityriasis lichenoides chronica [*see* Pityriasis Lichenoides, *above*].

Erythroderma

Papulosquamous and psoriasiform eczematous dermatitis may progress to generalized skin involvement with erythema and scaling, known as exfoliative erythroderma. Other causes of eryth-

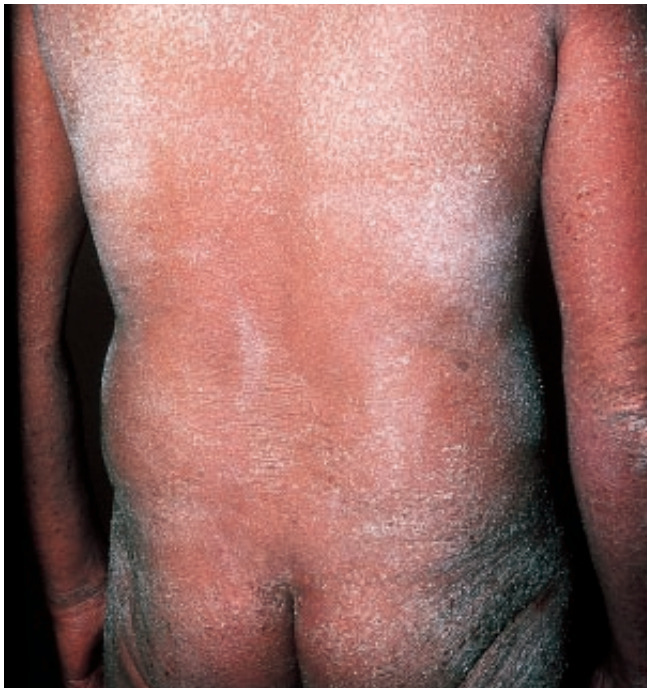


Figure 11 **Erythroderma, which appears as total skin erythema and scaling, can occur as a result of papulosquamous and eczematous disorders caused by a variety of diseases. Cutaneous T cell lymphoma, as seen in this patient with Sézary syndrome, can result in erythroderma.**

roderma include drug eruption, contact dermatitis, and pityriasis rubra pilaris. Eyrthroderma is a rare skin disorder that occurs more often as an exacerbation of a preexisting skin disorder; less commonly, it is idiopathic. There are no accurate studies on the incidence of erythroderma. On the basis of a survey of all dermatologists in the Netherlands, however, the annual incidence was estimated to be one to two patients per 100,000 inhabitants.[71]

DIAGNOSIS

Most cases of exfoliative erythroderma are associated with exacerbation of an underlying dermatosis, such as psoriasis, pityriasis rubra pilaris, seborrheic dermatitis, drug eruptions, atopic dermatitis, or contact dermatitis.[72] Some patients have idiopathic erythroderma, also called red man syndrome.[69] Common associated skin findings include palmoplantar keratoderma, alopecia, and nail dystrophy. Skin biopsy usually shows nonspecific inflammation. Lymph node biopsy may reveal dermatopathic lymphadenopathy. In some patients, idiopathic erythroderma may progress to cutaneous T cell lymphoma (e.g., erythrodermic mycosis fungoides and Sézary syndrome) [*see Figure 11*] [*see 2:X Malignant Cutaneous Tumors*].

Systemic symptoms associated with erythroderma include fever and chills, dehydration from transepidermal water loss, and high-output cardiac failure.

TREATMENT

Nonspecific treatment includes restoration of fluid and electrolyte balance and supportive measures such as administration of antipruritics, application of cool compresses and mild topical corticosteroids, and bed rest. Antibiotics may be required for treatment of secondary bacterial infection. Generally, more aggressive topical and systemic therapies are avoided until the inflammation subsides. More specific treatment depends on the underlying diagnosis and cause of the erythroderma. For example, in patients with erythroderma that is secondary to Sézary syndrome, treatment would be directed toward the underlying cutaneous T cell lymphoma [*see 2:X Malignant Cutaneous Tumors*]. For erythroderma caused by a drug eruption, the offending drug must be discontinued. Systemic agents such as acitretin and methotrexate may be used to treat psoriatic erythroderma [*see 2:III Psoriasis*]. Drug-induced erythroderma may have the best prognosis.[72]

The author has served as advisor or consultant to Abbott Laboratories, Amgen, Inc., Biogen Idec, Inc., and Genentech, Inc.

References

1. Fox BJ, Odom RB: Papulosquamous diseases: a review. J Am Acad Dermatol 12:597, 1985

2. Sadick NS, McNutt NS, Kaplan MH: Papulosquamous dermatoses of AIDS. J Am Acad Dermatol 22:1270, 1990

3. Chuh AA, Molinari N, Sciallis G, et al: Temporal case clustering in pityriasis rosea: a regression analysis on 1379 patients in Minnesota, Kuwait, and Diyarbakir, Turkey. Arch Dermatol 141:767, 2005

4. Chuang T-Y, Ilstrup DM, Perry HO, et al: Pityriasis rosea in Rochester, Minnesota, 1969 to 1978: a 10-year epidemiologic study. J Am Acad Dermatol 7:80, 1982

5. Parsons JM: Pityriasis rosea update: 1986. J Am Acad Dermatol 15:159, 1986

6. Drago F, Ranieri E, Malaguti F, et al: Human herpesvirus 7 in 7 patients with pityriasis rosea. Dermatology 195:374, 1997

7. Kempf W, Adams V, Kleinhans M, et al: Pityriasis rosea is not associated with human herpesvirus 7. Arch Dermatol 135:1070, 1999

8. Kosuge H, Tanaka-Taya K, Miyoshi H, et al: Epidemiological study of human herpesvirus–6 and human herpesvirus–7 in pityriasis rosea. Br J Dermatol 143:795, 2000

9. Chuh AA, Chan HH, Zawar V: Is human herpesvirus 7 the causative agent of pityriasis rosea? Int J Dermatol 43:870, 2004

10. Chuh AA: Rash orientation in pityriasis rosea: a qualitative study. Eur J Dermatol 12:253, 2002

11. Leenutaphong V, Jiamton S: UVB phototherapy for pityriasis rosea: a bilateral comparison study. J Am Acad Dermatol 33:996, 1995

12. Stulberg DL, Wolfrey J: Pityriasis rosea. Am Fam Physician 69:87, 2004

13. Sharma PK, Yadav TP, Gautam RK, et al: Erythromycin in pityriasis rosea: a double-blind placebo-controlled clinical trial. J Am Acad Dermatol 42:241, 2000

14. Eisen D: The clinical features, malignant potential, and systemic associations of oral lichen planus: a study of 723 patients. J Am Acad Dermatol 46:207, 2002

15. Shai A, Halevy S: Lichen planus and lichen planus–like eruptions: pathogenesis and associated diseases. Int J Dermatol 31:379, 1992

16. Lodi G, Giuliani M, Majorana A, et al: Lichen planus and hepatitis C virus: a multicentre study of patients with oral lesions and a systematic review. Br J Dermatol 151:1172, 2004

17. Harden D, Skelton H, Smith KJ: Lichen planus associated with hepatitis C virus: no viral transcripts are found in the lichen planus, and effective therapy for hepatitis C virus does not clear lichen planus. J Am Acad Dermatol 49:847, 2003

18. Callista D, Morri M: Lichen planus induced by hepatitis B vaccination: a new case and a review of the literature. Int J Dermatol 43:562, 2004

19. Thompson DF, Skaehill PA: Drug-induced lichen planus. Pharmacotherapy 14:561, 1994

20. Powell FC, Rogers RS III, Dickson ER: Primary biliary cirrhosis and lichen planus. J Am Acad Dermatol 9:540, 1983

21. Powell ML, Ehrlich A, Belsito DV: Lichenoid drug eruption to salsalate. J Am Acad Dermatol 45:616, 2001

22. Boyd AS, Neldner KH: Lichen planus. J Am Acad Dermatol 25:593, 1991

23. Abraham SC, Ravich WJ, Anhalt GJ, et al: Esophageal lichen planus: case report and review of the literature. Am J Surg Pathol 24:1678, 2000

24. Cribier B, Frances C, Chosidow O: Treatment of lichen planus: an evidence-based medicine analysis of efficacy. Arch Dermatol 134:1521, 1998

25. Zakrzewska JM, Chan ES, Thronhill MH: A systematic review of placebo-controlled randomized clinical trials of treatments used in oral lichen planus. Br J Dermatol 153:336, 2005

26. Laurberg G, Geiger JM, Hjorth N, et al: Treatment of lichen planus with acitretin: a double-blind placebo-controlled study in 65 patients. J Am Acad Dermatol 24:434, 1991

27. Lear JT, English JS: Erosive and generalized lichen planus responsive to azathioprine. Clin Exp Dermatol 21:56, 1996

28. Ortonne JP, Thivolet J, Sannwald C: Oral photochemotherapy in the treatment of lichen planus (LP): clinical results, histological and ultrastructural observations. Br J Dermatol 99:77, 1978

29. Hildebrand A, Kolde G, Luger TA, et al: Successful treatment of generalized lichen planus with recombinant interferon alfa-2b. J Am Acad Dermatol 33:880, 1995

30. Carbone M, Conrotto D, Carrozzo M, et al: Topical corticosteroids in association with miconazole and chlorhexidine in the long-term management of atrophic-erosive oral lichen planus: a placebo-controlled and comparative study between clobetasol and fluocinonide. Oral Dis 5:44, 1999

31. Giustina TA, Stewart JB, Ellis CN, et al: Topical application of oral isotretinoin gel improves oral lichen planus: a double-blind study. Arch Dermatol 122:534, 1986

32. Eisen D: Hydroxychloroquine sulfate (Plaquenil) improves oral lichen planus: an open trial. J Am Acad Dermatol 28:609, 1993

33. Kaliakatsou F, Hodgson TA, Lewsey JD, et al: Management of recalcitrant ulcerative oral lichen planus with topical tacrolimus. J Am Acad Dermatol 46:35, 2002

34. Swift JC, Rees TD, Plemons JM, et al: The effectiveness of 1% pimecrolimus cream in the treatment of oral erosive lichen planus. J Periodontol 76:627, 2005

35. Moyal-Barracco M, Edwards L: Diagnosis and therapy of anogenital lichen planus. Dermatol Ther 17:38, 2004

36. Byrd JA, Davis MD, Rogers RS 3rd: Recalcitrant symptomatic vulvar lichen planus: response to topical tacrolimus. Arch Dermatol 140:715, 2004

37. Watsky KL: Erosive perianal lichen planus responsive to tacrolimus. Int J Dermatol 42:217, 2003

38. Lonsdale-Eccles AA, Velangi S: Topical pimecrolimus in the treatment of genital lichen planus: a prospective case series. Br J Dermatol 153:390, 2005

39. Irvine C, Irvine F, Champion RH: Long-term follow-up of lichen planus. Acta Derm Venereol 71:242, 1991

40. Plewig G: Seborrheic dermatitis. Dermatology in General Medicine, 4th ed, Vol 1. Fitzpatrick TB, Eisen AZ, Wolff K, et al, Eds. McGraw-Hill Book Co, New York, 1993, p 1569

41. Gupta AK, Madzia SE, Batra R: Etiology and management of seborrheic dermatitis. Dermatology 208:89, 2004

42. Burton JL, Pye RJ: Seborrhoea is not a feature of seborrhoeic dermatitis. Br Med J 26:1169, 1983

43. Odom RB: Common superficial fungal infections in immunosuppressed patients. J Am Acad Dermatol 31:S56, 1994

44. Soeprono FF, Schinella RA, Cockerell CJ, et al: Seborrheic-like dermatitis of acquired immunodeficiency syndrome: a clinicopathologic study. J Am Acad Dermatol 14:242, 1986

45. Shuster S, Meynadier J, Kerl H, et al: Treatment and prophylaxis of seborrheic dermatitis of the scalp with antipityrosporal 1% ciclopirox shampoo. Arch Dermatol 41:47, 2005

46. Gupta AK, Nicol K, Batra R: Role of antifungal agents in the treatment of seborrheic dermatitis. Am J Clin Dermatol 5:417, 2004

47. Peter RU, Richarz-Barthauer U: Successful treatment and prophylaxis of scalp seborrhoeic dermatitis and dandruff with 2% ketoconazole shampoo: results of a multicentre, double-blind, placebo-controlled trial. Br J Dermatol 132:441, 1995

48. Parsad D, Pandhi R, Negi KS, et al: Topical metronidazole in seborrheic dermatitis—a double-blind study. Dermatology 202:35, 2001

49. Koca R, Altinyazar HC, Esturk E: Is topical metronidazole effective in seborrheic dermatitis? A double-blind study. Int J Dermatol 42:632, 2003

50. Rigopoulos D, Ioannides D, Kalogeromitros D, et al: Pimecrolimus cream 1% vs. betamethasone 17-valerate 0.1% cream in the treatment of seborrhoeic dermatitis: a randomized open-label clinical trial. Br J Dermatol 151:1071, 2004

51. Scapparo E, Quadri G, Virno G, et al: Evaluation of the efficacy and tolerability of oral terbinafine (Daskil) in patients with seborrhoeic dermatitis: a multicentre, randomized, investigator-blinded, placebo-controlled trial. Br J Dermatol 144:854, 2001

52. Dicken CH: Treatment of classic pityriasis rubra pilaris. J Am Acad Dermatol 31:997, 1994

53. Requena L, Grilli R, Soriano L, et al: Dermatomyositis with a pityriasis rubra pilaris-like eruption: a little-known distinctive cutaneous manifestation of dermatomyositis. Br J Dermatol 136:768, 1997

54. Van de Kerkfho PC, Steijlen PM: Topical treatment of pityriasis rubra pilaris with calcipotriol. Br J Dermatol 130:675, 1994

55. Yaniv R, Barzilai A, Trau H: Pityriasis rubra pilaris exacerbated by ultraviolet B phototherapy (letter). Dermatology 189:313, 1994

56. Goldsmith LA, Weinrich AE, Shupack J: Pityriasis rubra pilaris response to 13-*cis*-retinoic acid (isotretinoin). J Am Acad Dermatol 6:710, 1982

57. Allison DS, El-Azhary RA, Calobrisis SD, et al: Pityriasis rubra pilaris in children. J Am Acad Dermatol 47:386, 2002

58. Usuki K, Sekiyama M, Shimada S, et al: Three cases of pityriasis rubra pilaris successfully treated with cyclosporin A. Dermatology 200:324, 2000

59. Gupta AK, Skinner AR: A review of the use of infliximab to manage cutaneous dermatoses. J Cutan Med Surg 8:77, 2004

60. Lambert WC, Everett MA: The nosology of parapsoriasis. J Am Acad Dermatol 5:373, 1981

61. Hood AF, Mark EJ: Histopathologic diagnosis of pityriasis lichenoides et varioliformis acuta and its clinical correlation. Arch Dermatol 118:478, 1982

62. El Shabrawi-Caelan L, Kerl H, Cerroni L: Lymphomatoid papulosis: reappraisal of clinicopathologic presentation and classification into subtypes A, B, and C. Arch Dermatol 140:441, 2004

63. Klein PA, Jones EC, Nelson JL, et al: Infectious causes of pityriasis lichenoides: a case of fulminant infectious mononucleosis. J Am Acad Dermatol 49:S151, 2003

64. Wenzel J, Gutgemann I, Distelmaier M, et al: The role of cytotoxic skin-homing $CD8^+$ lymphocytes in cutaneous cytotoxic T-cell lymphoma and pityriasis lichenoides. J Am Acad Dermatol 53:422, 2005

65. LeVine MJ: Phototherapy of pityriasis lichenoides. Arch Dermatol 119:378, 1983

66. Satra KH, DeLeo VA: PUVA for photosensitivity and other skin diseases. Photochemotherapy in Dermatology. Abel EA, Ed. Igaku-Shoin Medical Publishers, New York, 1991, p 159

67. Humbert P, Treffel P, Chapuis J-F, et al: The tetracyclines in dermatology. J Am Acad Dermatol 25:691, 1991

68. Fink-Puches R, Soyer HP, Kerl H: Febrile ulceronecrotic pityriasis lichenoides et varioliformis acuta. J Am Acad Dermatol 30:261, 1994

69. Kikuchi A, Naka W, Harada T, et al: Parapsoriasis en plaques: its potential for progression to malignant lymphoma. J Am Acad Dermatol 29:419, 1993

70. Thestrup-Pedersen K, Halkier-Sorensen L, Sogaard H, et al: The red man syndrome: exfoliative dermatitis of unknown etiology: a description and follow-up of 38 patients. J Am Acad Dermatol 18:1307, 1988

71. Sigurdsson V, Steegmans PH, van Vloten WA: The incidence of erythroderma: a survey among all dermatologists in the Netherlands. J Am Acad Dermatol 45:675, 2001

72. Akhyani M, Ghodsi ZS, Toosi S, et al: Erythroderma: a clinical study of 97 cases. BMC Dermatol 5:5, 2005

III PSORIASIS

Elizabeth A. Abel, m.d.
Mark Lebwohl, m.d.

Psoriasis is an immune-mediated inflammatory cutaneous disorder characterized by chronic, scaling, erythematous patches and plaques of skin. It can begin at any age and can vary in severity. Psoriasis can manifest itself in several different forms, including pustular and erythrodermic forms. In addition to involving the skin, psoriasis frequently involves the nails, and some patients may experience inflammation of the joints (psoriatic arthritis). Because of its highly visible nature, psoriasis can compromise both the personal and the working lives of its victims.

Breakthroughs in the treatment of psoriasis have led to a better understanding of its pathogenesis. This chapter reviews current knowledge of the genetics, pathogenesis, and treatment of psoriasis.

Epidemiology

The estimated prevalence of psoriasis ranges from 0.5% to 4.6% worldwide. The reasons for the geographic variation in prevalence are unknown, but climate and genetics may play a role. Psoriasis is uncommon in blacks in tropical zones, but it is more often seen in blacks in temperate zones. It occurs commonly in Japanese persons but rarely in persons native to North and South America. In the United States, studies have variously reported that 4.5 million adults[1] or 7 million adults and children[2] have psoriasis.

Psoriasis can occur at any age, with some cases being reported at birth and others being reported in patients older than 100 years. In Farber and Nall's pioneer study of 5,600 patients, the average age of onset of psoriasis was 27.8 years; in 35% of patients, onset occurred before 20 years of age, and in 10%, onset occurred before 10 years of age.[3] Psoriasis occurs with equal frequency in men and women, but in Farber and Nall's study, onset occurred later in men. In populations in which there is a high prevalence of psoriasis, onset tends to occur at an earlier age. In the Faroe Islands, for example, the prevalence is 3%, and the average age of onset is 12.5 years. The average age of onset is 23 years in the United States. In persons with earlier age of onset, psoriasis is more likely to be severe, with involvement of a large area of skin surface.

Pathogenesis

Psoriasis was once thought to be caused by an abnormality in epidermal cell kinetics; it is now thought that an abnormality in the immune system triggers epidermal proliferation. The role of activated lymphocytes in the development of psoriasis was first proved through investigations of DAB389 interleukin-2 (IL-2), a fusion protein consisting of molecules of IL-2 fused to diphtheria toxin. This fusion protein binds to high-affinity IL-2 receptors on activated T cells, destroying those cells. In a study of DAB389 IL-2 treatment in 10 patients, four patients showed dramatic clinical improvement and four others showed moderate improvement.[4] Unfortunately, the side effects of DAB389 IL-2 have precluded its approval for the treatment of psoriasis.[5]

The skin of patients with lesional psoriasis has higher numbers of antigen-presenting cells that can activate T cells. For T cell activation to occur, antigen-presenting cells must deliver at least two signals to resting T cells. The first signal occurs when major histocompatibility complex (MHC) class II molecules of the antigen-presenting cells present antigens to the T cells. A second costimulatory signal must be delivered through the interaction of ligands on the surface of the antigen-presenting cells with receptors on the surface of T cells. Examples of this process include the interaction of B7 molecules with CD28 on the surface of resting T cells and the interaction of lymphocyte function–associated antigen 3 (LFA-3) with CD2 or intercellular adhesion molecule–1 (ICAM-1) with LFA-1 on the surface of T cells.[6,7] Blockade of any of these steps results in clearing of psoriasis.[8,9] Upon activation, T cells release Th1 (T helper type 1) cytokines, IL-2, and interferon gamma, which together induce proliferation of keratinocytes and further stimulation of T cells. Inflammatory cytokines such as tumor necrosis factor–α (TNF-α) are found in psoriatic skin lesions and joints, and treatment with TNF-α blockers results in clearing of psoriasis and of psoriatic arthritis.[10]

Etiology

Genetic Factors

Several lines of evidence suggest that psoriasis has a genetic etiology. One third of persons affected have a positive family history. Studies have found a higher concordance rate in monozygotic twins than in dizygotic twins or siblings (70% versus 23%).[11]

Current evidence suggests genetic heterogeneity. Both autosomal dominant inheritance with incomplete penetrance and polygenic or multifactorial inheritance have been described. The most important psoriasis susceptibility gene appears to be *PSORS1*, which has been mapped to the region on chromosome 6p21 that codes for the MHC; seven other *PSORS* genes have been found on other chromosomes.[12] Psoriasis is also associated with a single-nucleotide polymorphism on chromosome 17q25 that impairs binding of a runt-related protein (RUNX1).[13]

Contributing Factors

The course and severity of psoriasis can be affected by a number of endogenous and exogenous factors, including stress, climate, concurrent infections, and medications.

Psychological Stress

Many patients believe that anxiety or psychological stress has an adverse effect on the course of their psoriasis. The etiologic significance of stress in psoriasis is difficult to evaluate, however, because of the subjective nature of the evidence used in many of the investigations into this question.[14] In a prospective study, a multivariate statistical method revealed a positive correlation between severity of psoriasis symptoms and psychological stress related to adverse life events.[15] Psoriasis itself can be a source of stress: the effects of psoriasis on physical and mental function have been compared with the effects of cancer, heart disease, diabetes, and depression.[16]

Climate

It has long been known that psoriasis improves when patients are exposed to sunny climates and to regions of lower latitude. In northern latitudes, exacerbation of psoriasis commonly occurs during the fall and winter.

Infection

Viral or bacterial infections, especially streptococcal pharyngitis or tonsillitis, may precipitate the onset or exacerbation of psoriasis.[17] Guttate psoriasis, in particular, is often attributed to a previous streptococcal infection. Attempts to reverse psoriasis by treatment with oral antibiotics have not proved effective in double-blind trials.[18] Nevertheless, some investigators advocate antibiotic therapy for psoriasis.[19]

Infection with HIV has also been associated with psoriasis. In some patients with HIV infection, preexisting psoriasis becomes exacerbated; in other patients, psoriasis develops within a few years after HIV infection. Often, HIV-infected patients present with symptoms similar to those of Reiter syndrome.[20]

Drugs

Numerous drugs can worsen psoriasis.[21] Antimalarial agents such as chloroquine can cause exfoliative erythroderma or pustular psoriasis. Up to 31% of patients experience new onset or worsening of psoriasis as a result of antimalarial therapy. Lithium and beta blockers such as propranolol may precipitate the onset of psoriasis or cause exacerbations of psoriasis.[22] Some nonsteroidal anti-inflammatory drugs (NSAIDs) also exacerbate psoriasis, although this effect is sufficiently minor to allow NSAIDs to be used in the treatment of psoriatic arthritis.[23] Flares of pustular psoriasis may be precipitated by withdrawal from systemic corticosteroids or withdrawal from high-potency topical corticosteroids. Interferon therapy has been associated with development or exacerbation of psoriasis, presumably because of the Th1 effects of this therapy.[24]

Other Factors

Trauma to the clinically uninvolved skin of patients with psoriasis can cause a lesion to appear at the exact site of injury; this phenomenon is known as the Köbner response. Cuts, abrasions, injections, burns resulting from phototherapy, and other forms of trauma can elicit this reaction.

Smoking may be an exacerbating factor in psoriasis.[25] Alcohol use has also been implicated in the exacerbation of psoriasis.[26]

Surveys have suggested that diet plays a role in the development of psoriasis, and attempts have been made to affect the clinical course of psoriasis through modification of diet.[27] Double-blind studies, however, have failed to show that diet has either a beneficial or a detrimental effect on the severity of psoriasis.

Diagnosis

The diagnosis of psoriasis is usually made on clinical grounds. If unusual features are present, biopsy of affected skin can be done to confirm the diagnosis.

CLINICAL VARIANTS

Nearly 90% of patients with psoriasis have plaque type, a form that is characterized by sharply demarcated, erythematous, scaling plaques. The elbows [*see Figure 1*], knees, and scalp [*see Figure 2*] are the most commonly affected sites. The intergluteal cleft [*see Figure 3*], palms [*see Figure 4*], soles [*see Figure 5*], and genitals are also commonly affected, but psoriasis can involve any part of the body. Lesions frequently occur in a symmetrical pattern of distribution.

Many patients have only one or a few lesions that persist for years and that occasionally resolve after exposure to sunlight. Other patients can be covered with plaques that become confluent, affecting nearly 100% of the body surface area. Nail involvement is common, particularly in patients with severe disease.

The second most common form of psoriasis, guttate psoriasis, affects fewer than 10% of patients and is characterized by the development of small, scaling, erythematous papules on the trunk and the extremities [*see Figure 6*]. This form of psoriasis often follows streptococcal infection. Patients with plaque-type psoriasis can develop guttate psoriasis. Conversely, patients with guttate psoriasis frequently develop plaque-type psoriasis. Occasionally, guttate lesions enlarge and become confluent, resulting in the formation of plaques.

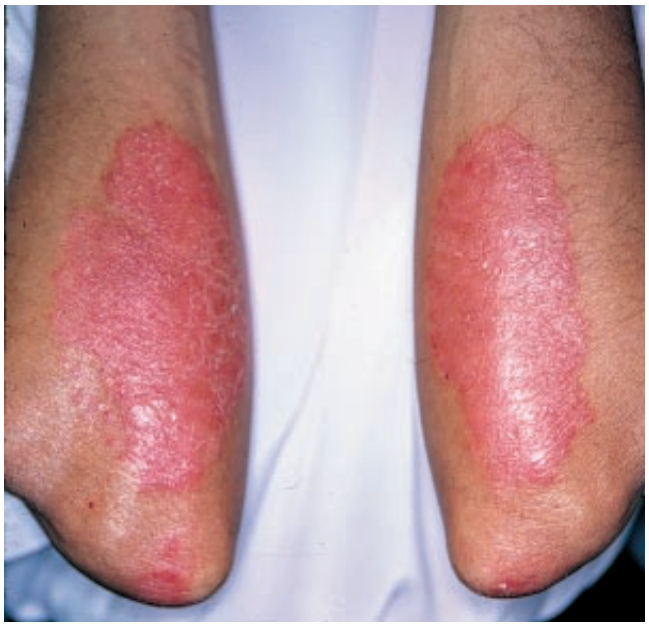


Figure 1 **Involvement of the elbows is characteristic of plaque psoriasis.**

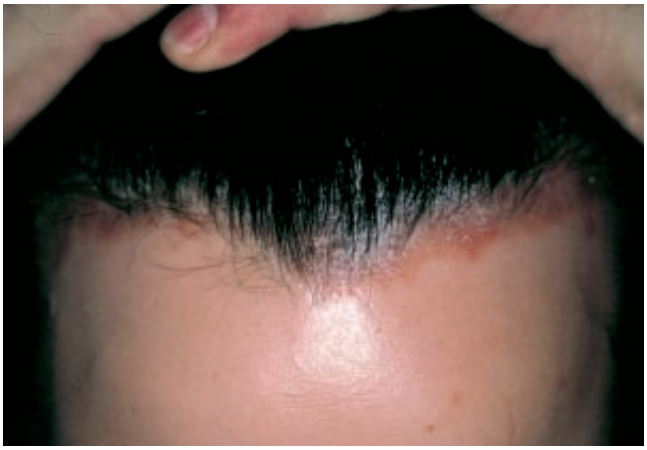


Figure 2 **The scalp is affected in the majority of patients with plaque psoriasis.**

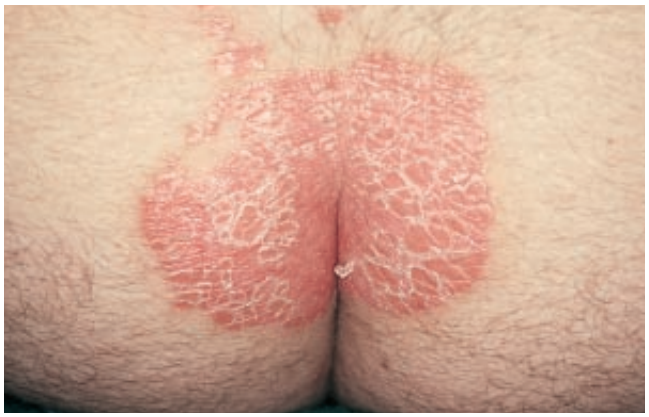


Figure 3 **The intergluteal cleft is a common site of involvement in patients with plaque psoriasis.**

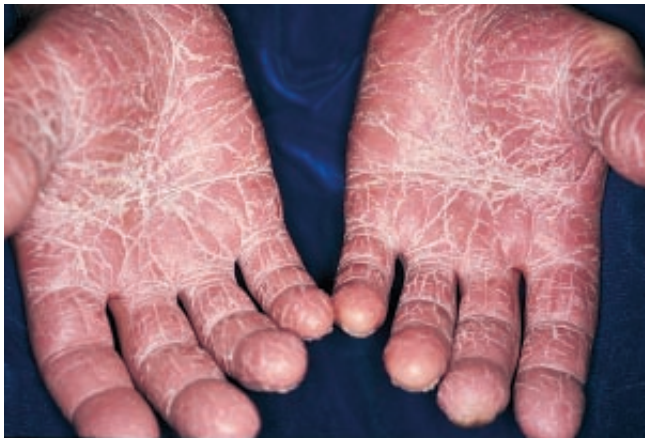


Figure 4 **Psoriasis of the palms is shown in this patient.**

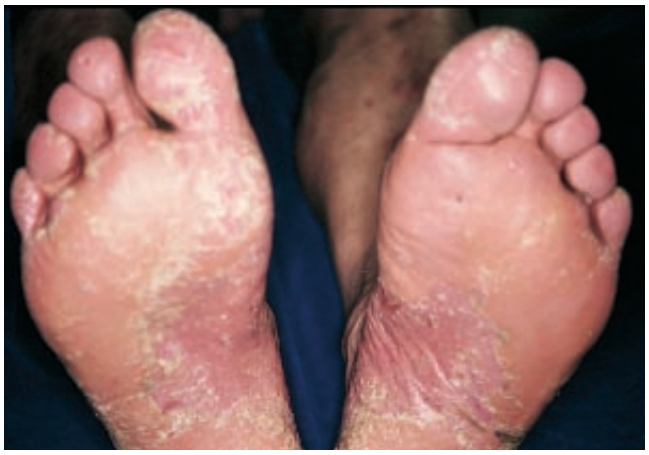


Figure 5 **Sharply demarcated, erythematous, scaling plaques on the feet are apparent in this patient with psoriasis of the soles.**

Erythrodermic psoriasis is a severe form of psoriasis that often affects the entire cutaneous surface. Patients present with an exfoliative erythroderma in which the skin is very red and inflamed and is constantly scaling [*see Figure 7*]. Patients are acutely ill, their skin having lost all protective function. Loss of temperature control, loss of fluids and nutrients through the impaired skin, and susceptibility to infection make this a potentially life-threatening condition.

Erythrodermic psoriasis can develop de novo or evolve from typical plaque-type or guttate psoriasis. Erythrodermic psoriasis can occur after withdrawal of systemic corticosteroids, after phototherapy burns, as a result of antimalarial treatment, as a result of a drug-induced hypersensitivity reaction, or for no apparent reason. Cutaneous T cell lymphoma may also present as erythroderma and needs to be differentiated from erythrodermic psoriasis.

Pustular psoriasis, another severe form of the disease, can occur in patients with preexisting psoriasis or can arise de novo. Pustular psoriasis can be generalized (von Zumbusch type) or localized to the palms and soles [*see Figure 8*]. In either case, the condition is severe and debilitating. In generalized

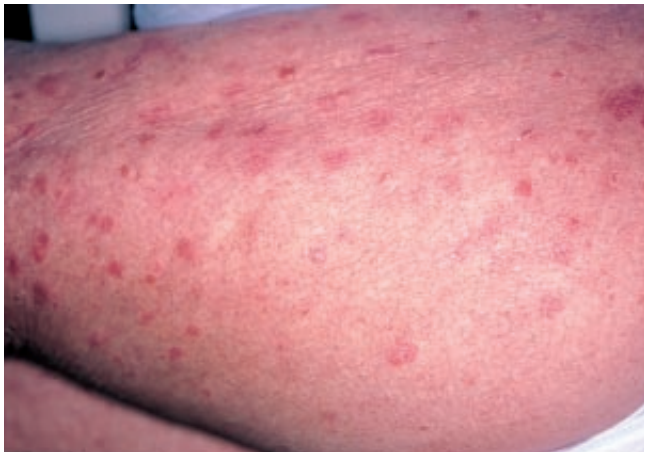


Figure 6 **Guttate psoriasis is characterized by small scaly papules and plaques.**

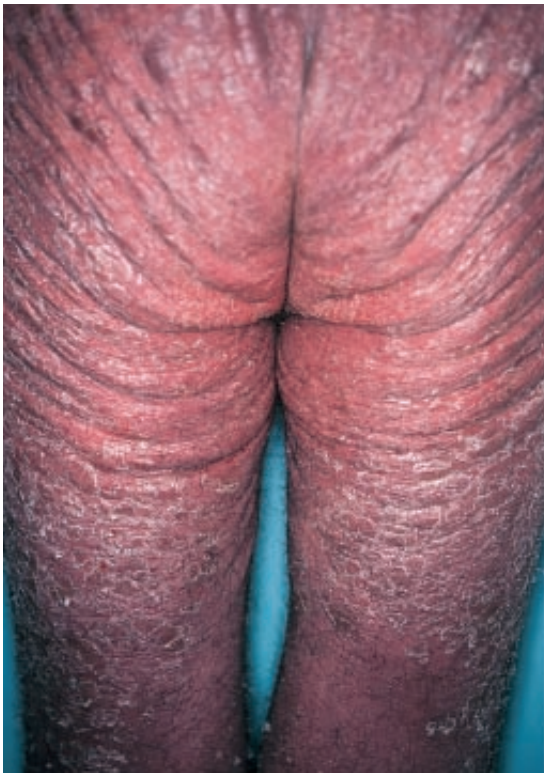


Figure 7 **Erythrodermic psoriasis is characterized by generalized erythema and desquamation.**

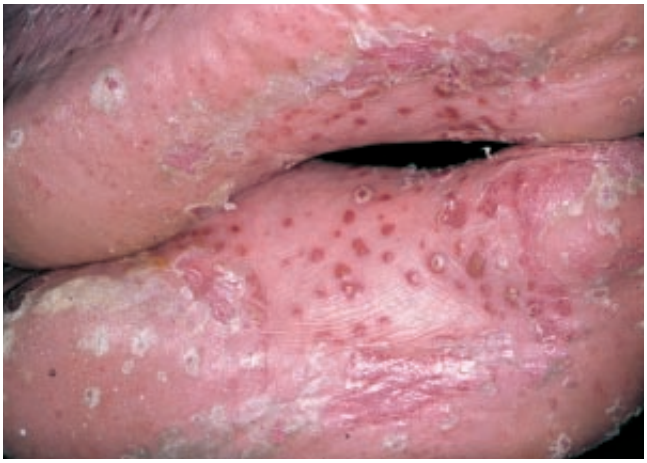


Figure 8 **Pustular psoriasis can be localized to the palms and soles or generalized.**

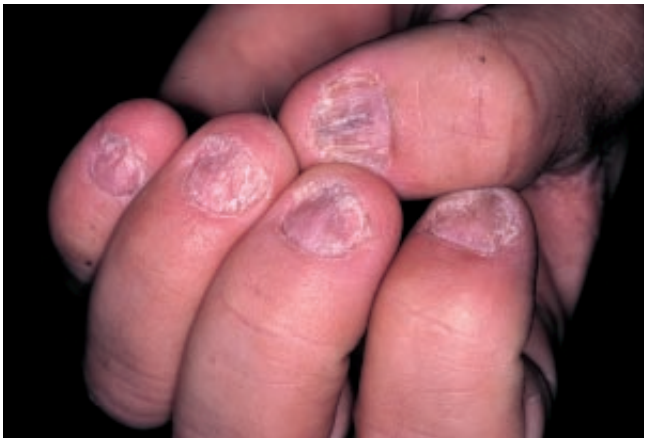


Figure 9 **Involvement of the nails is common in psoriasis.**

pustular psoriasis, the body is covered with sterile pustules. As with erythrodermic psoriasis, the protective functions of the skin are lost, and patients may succumb to infection or hypovolemia and electrolyte imbalance caused by loss of fluid through the skin. Although fever and leukocytosis are common features in pustular psoriasis, the possibility of infection should not be overlooked; patients with pustular psoriasis have died of staphylococcal sepsis.[28]

As with erythrodermic psoriasis, pustular psoriasis is most commonly precipitated by withdrawal of systemic corticosteroids. However, it can also result from therapy with antimalarial drugs or lithium, and it can develop spontaneously.

Nail Psoriasis

Nail changes can be of immeasurable value when the diagnosis is in doubt [*see Figure 9*]. In one study, 55% of patients with psoriasis experienced such changes.[29] The most common change consists of the appearance of tiny pits, as might be made with an ice pick, which often occur in groups. This characteristic pitting of the nails is highly specific for psoriasis, although a few isolated pits may be seen in healthy nails or as a result of past trauma. Yellowish discoloration is common in psoriatic toenails and may appear in fingernails as well. Onycholysis, or distal separation of the nail plate from its bed, frequently occurs.

Other changes include subungual hyperkeratosis—an accumulation of keratinous debris under the nail—as well as transverse and longitudinal ridging. These findings, however, are much less specific because they also occur secondary to dermatitis, fungal infection, vascular insufficiency, and other conditions. Occasionally, a patient shows typical psoriatic nail changes without any other cutaneous signs at initial examination; all such patients are probably psoriatic and may eventually manifest psoriatic lesions.

Psoriatic Arthritis

Psoriatic arthritis has been estimated to occur in 7% to 42% of patients with psoriasis.[30] Joint inflammation in psoriatic arthritis is chronic, with occasional remissions.[31] There are five classic subtypes. The most common presentation is an oligoarthritis in which one or a few joints are affected. This form accounts for approximately 70% of cases of psoriatic arthritis. Skin lesions of psoriasis usually precede articular disease by 5 to 10 years, but joint inflammation develops before skin lesions in some patients. If a diagnosis of psoriatic arthritis is suspected, the physician should carefully examine the scalp, nails, intergluteal cleft, external ear canal, and genital region for psoriasis lesions.

The second most common type of psoriatic arthritis is virtually identical to rheumatoid arthritis. This form is characterized by symmetrical involvement of the joints with ulnar deviation and typical deformities, such as swan-neck deformity and boutonnière deformity. The only distinguishing features are the presence of psoriasis and the absence of circulating rheumatoid factor.

Arthritis mutilans is a rare, severely destructive form of psoriatic arthritis in which the interphalangeal joints of the hands and feet are destroyed, resulting in deformed digits. Ankylosing spondylitis accounts for 5% of cases of psoriatic arthritis. As in other forms of ankylosing spondylitis, the genetic marker HLA-B27 is usually present.

Distal interphalangeal joint involvement is the most characteristic form of psoriatic arthritis. It is usually associated with nail involvement.

HISTOPATHOLOGY

The classic microscopic features of a psoriatic plaque include the following:

- A markedly thickened stratum corneum, with layered zones of parakeratosis (retention of nuclei).
- A moderately to markedly hyperplastic epidermis, with broadening of rete projections and elongation to a uniform depth in the dermis.
- Increased mitotic activity in the lower epidermis.
- Epidermal thinning over the dermal papillae.
- A scant amount of inflammatory infiltrate from mononuclear cells in the superficial dermis.
- Intracorneal or subcorneal collections of polymorphonuclear leukocytes (Munro microabscesses)

Differential Diagnosis

The differential diagnosis of psoriasis includes other scaling dermatoses [*see 2:II Papulosquamous Disorders*]. Such dermatoses include the following:

- Seborrheic dermatitis that involves the scalp, nasolabial folds, and retroauricular folds.
- Pityriasis rosea, which begins with a herald patch and is self-limited.

- Lichen simplex chronicus, which is caused by repeated rubbing or scratching.
- Parapsoriasis, which is characterized by atrophy, telangiectasia, and pigmentary abnormalities.
- Pityriasis rubra pilaris, which is characterized by psoriasiform patches that often begin in sun-exposed areas.
- Other conditions (e.g., discoid eczema or secondary syphilis) that can be differentiated by clinical and pathologic criteria.

Treatment

More treatments are available for psoriasis than perhaps for any other dermatologic disease. New topical therapies, new systemic therapies, and new forms of phototherapy have been introduced, and additional treatments are in development. Biologic therapies that target specific molecules are likely to change the treatment of psoriasis in the future. Topical therapy will continue to be used by most patients, however.

TOPICAL THERAPY

The 1990s saw the development of many new therapies for psoriasis.[32] Topical therapy is the mainstay of treatment for most patients, particularly those with mild disease. Topical corticosteroids are the most commonly prescribed class of medication, but they are now often used together with topical calcipotriene, a vitamin D_3 analogue, or topical tazarotene, a retinoid; both calcipotriene and tazarotene have approval by the Food and Drug Administration for the treatment of psoriasis.[33] Tar and salicylic acid are available by prescription and as over-the-counter products. Use of anthralin has declined as effective nonsteroidal agents have become available.

Emollients are an important part of any topical regimen for psoriasis. Application of petrolatum alone may be sufficient therapy for some patients. More elegant creams and lotions are helpful but are somewhat less effective than greasy ointments. Tar and salicylic acid shampoos are valuable in the treatment of patients with scalp involvement. These preparations are available without prescription.

Corticosteroids

Topical corticosteroids are indicated for limited plaques of psoriasis. Because of their ease of use and their wide availability, topical corticosteroids are the most commonly prescribed medication for treatment of psoriasis. They have anti-inflammatory, antiproliferative, and antipruritic effects. Corticosteroids are more potent when they are applied under occlusion, which increases their percutaneous penetration. Unfortunately, occlusion also increases side effects.

Topical steroids have been ranked in seven categories in decreasing order of potency, with potency determined by a vasoconstriction assay [*see Table 1*]. Superpotent corticosteroids are in group I, and weak over-the-counter topical corticosteroids are in group VII.[34]

Side effects The most commonly encountered side effects of topical corticosteroids are local cutaneous reactions. Development of cutaneous atrophy, telangiectasia, and irreversible striae are the most common side effects. Perioral dermatitis, which is characterized by erythematous papules and pustules on the face, is caused by chronic use of topical corticosteroids. Tachyphylaxis, with habituation to topical corticosteroids and loss of response to them, is noted by most patients. Flare or rebound of psoriasis upon sudden withdrawal of topical corticosteroids can occur. Finally, suppression of the hypothalamic-pituitary-adrenal axis can occur, especially with use of superpotent topical corticosteroids, the widespread application of corticosteroids, occlusion, or chronic use. Because of concern over side effects, the package inserts for some superpotent corticosteroids suggest that use be limited to 2 weeks' duration. A number of regimens have been developed in which, after the initial weeks of continuous treatment with superpotent topical corticosteroids, psoriasis plaques are subsequently treated only on weekends.[35]

Vitamin D Analogues

Calcipotriene The first topical vitamin D analogue to receive FDA approval for use in the United States, calcipotriene has rapidly gained acceptance, despite the fact that it is not as effective as superpotent topical corticosteroids. Calcipotriene is available in ointment and cream form and as a solution. The primary reason for its success is its freedom from any corticosteroid side effects—namely, cutaneous atrophy, telangiectasia, striae, or suppression of the hypothalamic-pituitary-adrenal axis. Calcipotriene is comparable in efficacy to a group II corticosteroid. It is applied twice daily.

Calcipotriene has been used very successfully in combination with several other medications. It is most effective when used in combination with a superpotent topical corticosteroid. A regimen of calcipotriene ointment and halobetasol propionate ointment, each applied once daily, has been found to be more effective than monotherapy with either calcipotriene twice daily or halobetasol propionate twice daily.[36] Up to 90% of patients achieve marked improvement within 2 weeks of combination therapy with once-daily calcipotriene and once-daily halobetasol propionate ointment. For long-term maintenance of remission, a regimen has been developed in which halobetasol propionate is applied only on weekends and calcipotriene is applied on weekdays.[37] Using this regimen, 76% of patients achieved marked improvement for at least 6 months; this level of improvement was achieved in only 40% of patients receiving halobetasol propionate ointment on weekends only. Calcipotriene has also been shown to improve the response to ultraviolet B light (UVB)[38] and to psoralen plus ultraviolet A light (PUVA).[39]

Caution must be used when combining calcipotriene ointment with other medications, because it is easily inactivated. Salicylic acid, for example, completely inactivates calcipotriene on contact. Several other topical medications, including topical corticosteroids, can inactivate calcipotriene. In contrast, halobetasol propionate ointment is compatible with calcipotriene even when one medication is applied on top of the other.[40] UVA has been shown to inactivate calcipotriene,[41] so calcipotriene should be applied after PUVA therapy, not before. Use of calcipotriene should be limited to a maximum of 120 g a week because of isolated reports of hypercalcemia.[42]

A combination product containing calcipotriene and betamethasone dipropionate is now available in Europe and Canada. It appears to be more effective than the individual medications applied separately.[43]

Other vitamin D analogues Several new vitamin D analogues are under investigation in the United States or are in use elsewhere. Tacalcitol and maxacalcitol are promising medications for the treatment of psoriasis. The only common side effect

Table 1 Ranking of Topical Steroids for Psoriasis in Order of High to Low Potency

Group	*Generic Name*	*Trade Name*	*Strength (%)*
I	Betamethasone dipropionate in optimized vehicle	Diprolene ointment	0.05
	Clobetasol propionate	Temovate cream, ointment	0.05
	Diflorasone diacetate	Psorcon ointment	0.05
II	Amcinonide	Cyclocort ointment	0.1
	Betamethasone dipropionate, augmented	Diprolene AF cream	0.05
	Betamethasone dipropionate	Diprosone ointment	0.05
	Mometasone furoate	Elocon ointment	0.1
	Diflorasone diacetate	Florone ointment, Maxiflor ointment	0.05
	Halcinonide	Halog cream	0.1
	Fluocinonide	Lidex cream, ointment; Topsyn gel	0.05
	Desoximetasone	Topicort cream, ointment	0.25
III	Triamcinolone acetonide	Aristocort cream (HP)	0.5
	Betamethasone dipropionate	Diprosone cream	0.05
	Diflorasone diacetate	Florone cream, Maxiflor cream	0.05
	Betamethasone valerate	Valisone ointment	0.1
IV	Triamcinolone acetonide	Aristocort ointment, Kenalog ointment	0.1
	Betamethasone benzoate	Benisone ointment	0.025
	Flurandrenolide	Cordran ointment	0.05
	Mometasone furoate	Elocon cream	0.1
	Fluocinolone acetonide	Synalar-HP cream	0.2
		Synalar ointment	0.025
V	Betamethasone benzoate	Benisone cream	0.025
	Flurandrenolide	Cordran cream	0.05
	Fluticasone propionate	Cutivate cream	0.05
	Betamethasone dipropionate	Diprosone lotion	0.02
	Triamcinolone acetonide	Kenalog cream, lotion	0.1
	Hydrocortisone butyrate	Locoid cream	0.1
	Fluocinolone acetonide	Synalar cream	0.025
	Betamethasone valerate	Valisone cream, lotion	0.1
	Hydrocortisone valerate	Westcort cream	0.2
VI	Alclometasone dipropionate	Aclovate cream	0.05
	Desonide	Tridesilon cream, ointment; DesOwen cream, ointment	0.05
	Flumethasone pivalate	Locorten cream	0.03
	Fluocinolone acetonide	Synalar solution	0.01
VII	Hydrocortisone	Hytone cream, lotion, ointment	2.5
		Hytone, Penecort, Synacort, Cort-Dome, Nutracort	1.0

is irritation, which occurs in up to 20% of patients, most often on the face and in intertriginous areas. Topical calcitriol has FDA approval for the treatment of psoriasis in several countries around the world[44]; it may be less irritating than calcipotriene in intertriginous sites.

Tazarotene

Tazarotene is a retinoid that has been developed for the treatment of psoriasis. It is available in 0.05% and 0.1% gels and in cream formulations. Tazarotene is comparable in efficacy to a group II corticosteroid cream. Patients receiving tazarotene 0.1% gel experience longer periods of remission after discontinuance of therapy than patients receiving corticosteroids.

Tazarotene has several advantages over the corticosteroids. First, it is not associated with cutaneous atrophy, telangiectasia, or the development of striae. In fact, tazarotene, like other retinoids, may actually prevent corticosteroid atrophy. Tazarotene has been shown to enhance the efficacy of UVB phototherapy.[45] It does, however, increase the ability of ultraviolet light to induce erythema.[46] Doses of UVB and UVA should therefore be reduced in patients who are also receiving tazarotene.

Side effects The main side effect of tazarotene is local irritation, which has caused many patients to discontinue its use. The combination of tazarotene and a topical corticosteroid reduces irritation and enhances the efficacy of both agents.

Tars

Tar has been used since the 19th century to treat psoriasis. Crude coal tar, a complex mixture of thousands of hydrocarbon compounds, affects psoriatic epidermal cells through enzyme inhibition and antimitotic action.[47] Crude coal tar is messy to apply, has a strong odor, and stains skin and clothing. It is applied in conjunction with UVB phototherapy in the Goeckerman regimen [see Phototherapy, below]. More refined tar preparations, which are cosmetically acceptable, are available by prescription and over the counter in the form of gels, creams, bath oils, shampoos, and solutions (liquor carbonis detergens). Tar is often used in combination therapies and as maintenance therapy after psoriasis plaques have resolved.

Anthralin

Anthralin (dithranol) has been used to treat psoriasis since

1916.[48] It is an extremely effective topical agent for psoriasis, probably because it inhibits enzyme metabolism and reduces epidermal mitotic turnover.[48]

Indications Because of the staining and irritation associated with the use of anthralin, this agent is usually prescribed for patients who do not respond to other topical therapies.

Formulations and regimens A modified Ingram regimen combines the daily application of anthralin in a stiff paste with tar baths and with exposure to ultraviolet light. This therapy involves application of progressively higher concentrations of anthralin for 6 to 8 hours at a time; it was introduced in the United States for hospitalized psoriatic patients[44] and for ambulatory patients in a psoriasis day care center.[49]

Modified anthralin formulations have been used to minimize the staining from anthralin, to decrease irritation, and to promote home use of the medication. Short-contact therapy consists of the application of anthralin to localized plaques for 30 minutes to 2 hours, after which time the anthralin must be thoroughly removed to minimize irritation of the surrounding skin.[50] Anthralin in a cream base, which can be removed by washing with water, is suitable for home use; it is available in 1% and 0.5% concentrations, for application to localized lesions on the skin and the scalp.

A formulation of 1% anthralin cream, composed of microencapsulated lipid crystals that release anthralin for absorption at skin temperature, is available. When used as short-contact therapy, this preparation carries a low risk of staining and irritation.[51]

Anthralin is most effective therapeutically when it is compounded in the form of a hard paste containing paraffin; this form is most commonly used in ambulatory psoriasis treatment centers. Anthralin ointment is less effective than anthralin paste, and anthralin cream is even less effective. With regard to patient compliance, this order is reversed. The end point of treatment is resolution of plaques to a macular state; this is usually associated with residual postinflammatory hyperpigmentation and temporary staining from anthralin. Resolution of symptoms usually occurs within 2 to 3 weeks after a modified Ingram regimen; remissions last for weeks to months.

Side effects Staining of skin, clothing, and the home is common with anthralin, as is irritation at the site of application.

SUNLIGHT

Ultraviolet radiation has a beneficial effect on psoriasis. Sunbathing for 2 to 4 weeks lessens the morbidity associated with the disorder, and climatotherapy at the Dead Sea is an effective alternative therapy for psoriasis for those who can travel to that part of the world. Because of its unique geographic location, 300 m below sea level, patients are exposed to naturally filtered ultraviolet light, which results in significant improvement or complete resolution of symptoms in 83% of patients over several weeks.[52] The sunlight at the Dead Sea accounts for most of the response, with little additional improvement resulting from bathing in the Dead Sea. Not surprisingly, patients treated at the Dead Sea have higher rates of nonmelanoma skin cancer.[53]

PHOTOTHERAPY

Phototherapy with UVB is an important therapeutic option for patients with extensive psoriasis. UVB irradiation can be used alone, but it has traditionally been combined with topical application of tar. Daily in-hospital application of crude coal tar and exposure to ultraviolet light (the Goeckerman regimen) can lead to a resolution of symptoms in widespread psoriasis within 3 or 4 weeks and can effect remissions that last for weeks to months.

In a reevaluation of the Goeckerman regimen, application of a 1% tar preparation was found to be as effective as a 6% preparation. Furthermore, application of the tar preparation for 2 hours before irradiation was equivalent to longer periods of application.[54] Contraindications to the use of the Goeckerman regimen include the presence of severely excoriated or inflamed psoriasis, erythrodermic and pustular forms of the disease, folliculitis, and a history of photosensitivity.

Newer regimens, which are more convenient and aesthetically acceptable, combine UVB with emollients. The emollient or vehicle decreases reflectance of the psoriatic scale, thereby increasing light transmission. According to a report by Lowe and colleagues, results with emollients are equivalent to those with tar, when used in regimens that utilize UVB in doses sufficient to cause erythema (erythemogenic); however, tar may have an additive effect when combined with a less aggressive regimen of suberythemogenic UVB.[55]

In a comparison study, outpatient UVB phototherapy was administered three times weekly, along with the application of either a tar oil or an emollient twice a day. This approach led to clearing of psoriatic lesions in 78% of patients [*see Figure 10*]. No difference in response was observed between the tar oil and the emollient.[56] Although the Lowe study had shown an additive effect for tar combined with UVB irradiation when patients were evaluated after 3 to 4 weeks (before their lesions had cleared),[57] this comparison study showed no such advantage in patients who were evaluated at the time of lesion clearing. Remission lasted longer in patients who received maintenance UVB phototherapy twice weekly for 1 to 2 months and then once weekly for up to 4 months than in patients who stopped receiving UVB phototherapy after the initial clearing.

Narrow-Band UVB

Narrow-band UVB, which comprises wavelengths of approximately 311 nm (as opposed to the 295 to 320 nm range of broad-band UVB), is a newer approach that is more effective than broad-band UVB.[58] Like other forms of phototherapy, narrow-band UVB works through local effects; therefore, covered areas, such as the scalp, do not respond.[59]

PHOTOCHEMOTHERAPY

Photochemotherapy with PUVA is indicated for patients with extensive, disabling psoriasis that has failed to respond to conventional forms of therapy, including conventional or narrow-band UVB phototherapy. PUVA therapy entails the administration of the photosensitizing drug methoxsalen (8-methoxypsoralen)—in an oral dose or by soaking in a tub containing methoxsalen or applying topical methoxsalen—followed by exposure of the patient to high-intensity longwave ultraviolet light in a walk-in irradiation chamber. The initial UVA dose (in joules/cm^2) is based on the patient's skin type and calculated in accordance with established protocols.[60]

Although its therapeutic effect is local, PUVA is a systemic treatment in which photoactivated methoxsalen binds to epidermal DNA, forming monofunctional and bifunctional adducts. It has been postulated that the resulting interference with epider-

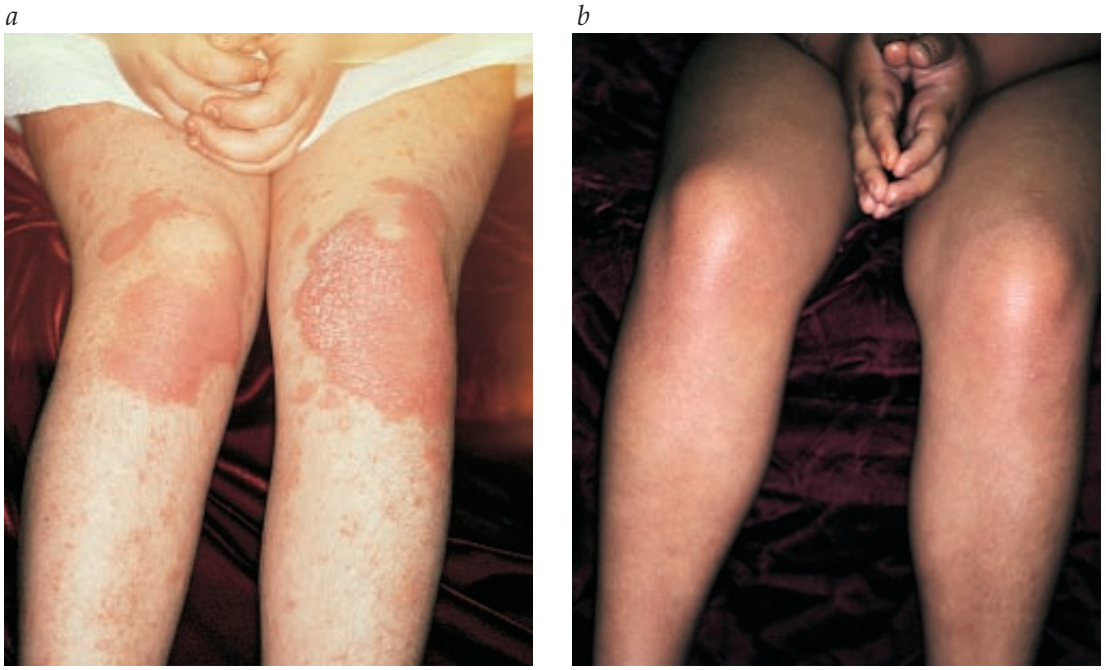


Figure 10 **Psoriasis in a child before (*a*) and after (*b*) phototherapy.**

mal mitosis is one of the mechanisms of action of PUVA therapy for psoriasis, although effects on immune function in the skin play an important role.

The efficacy of oral PUVA therapy has been established by several multicenter clinical trials.[61] A course of PUVA therapy administered two or three times weekly resulted in significant clearing of psoriasis lesions in approximately 90% of patients within a mean of 25 total treatments. After the initial course, a tapering maintenance regimen is instituted, and PUVA therapy is eventually discontinued. In most patients, psoriasis recurs months to years after PUVA is discontinued, indicating that this therapy is palliative rather than curative.

Side Effects

Acute side effects caused by phototoxicity, such as erythema and blistering, are dose related and can therefore be controlled. Pruritus, usually associated with dryness of the skin, is fairly common and can be alleviated by the use of emollients and oral antihistamines. Nausea may follow ingestion of methoxsalen. Of greater concern are the potential long-term side effects, particularly carcinogenicity. Although the FDA has approved the use of PUVA to treat psoriasis, patients must be closely monitored for long-term side effects. A multicenter study of more than 1,300 PUVA-treated patients in the United States who were evaluated after 1 to 3 years of follow-up revealed a significant increase in the number of squamous cell carcinomas (SCCs) in those patients with a history of exposure to ionizing radiation or a history of skin cancer.[62] A higher-than-expected ratio of SCCs to basal cell epitheliomas and an excess of SCCs in areas of the body that were not exposed to the sun were significant findings of the study. A 5.7-year follow-up study of the original cohort group revealed a dose-dependent increase in the risk of SCC.[63] There was only a slight increase in the risk of basal cell carcinoma in these patients. The risk of SCC was almost 13-fold higher in patients who had received high cumulative doses of PUVA than in patients who received low-dose therapy.

A follow-up study of the surviving members of that cohort, at least 15 years after original treatment, again assessed the risk of skin cancers. Of great concern was a small but statistically significant increase in the incidence of malignant melanoma.[64] Because that increase did not become apparent until after a period of at least 15 years, there is great concern that high rates of melanoma will occur in patients who began PUVA therapy years ago. Fortunately, this has not happened thus far.

Studies in animals suggest that PUVA may have ocular side effects. Methoxsalen has been detected in the lenses of rats after they have ingested the drug; subsequent exposure to UVA enhances such ultraviolet-induced changes as cataracts.[65] The risk of ocular toxicity and possible retinal damage is of particular concern in young persons, whose lenses transmit more UVA than the more opaque lenses of older persons, and in aphakic persons, in whom lenses are absent.[66] The use of UVA-opaque goggles during PUVA treatment sessions is extremely important. Glasses that block UVA must be worn from the time that methoxsalen is administered throughout the rest of the day. Some investigators advise protection of the eyes the day after therapy. Thus far, studies of patients treated with PUVA have not revealed an increase in the incidence of cataracts.

SYSTEMIC THERAPY

Methotrexate

Short-term use of the antimetabolite methotrexate can be an extremely effective treatment for psoriasis. Methotrexate is indicated for patients who do not respond adequately to phototherapy and for patients with psoriatic arthritis.

The source of methotrexate's efficacy against psoriasis was once thought to be its antimitotic effect on proliferating keratinocytes. However, tissue culture studies have suggested that activated lymphoid cells in the lymph nodes, blood, and skin are a likely target of methotrexate; proliferating macrophages and T cells are 100 times more sensitive to methotrexate than are prolif-

erating epithelial cells.[67] These findings may be relevant to the mechanism of action of methotrexate in other immunologically based disorders, including psoriatic arthritis, rheumatoid arthritis, and Crohn disease.

Dosage Methotrexate is best given in a single weekly oral dose of up to 30 mg or in three divided doses at 12-hour intervals during a 24-hour period (e.g., at 8:00 A.M., at 8:00 P.M., and again at 8:00 A.M.).

Hepatoxicity and liver biopsy The use of liver biopsy has been advocated for monitoring patients with psoriasis who are receiving methotrexate. This recommendation is controversial, however; critics point out that liver biopsies are not routinely performed in patients with rheumatoid arthritis who are undergoing treatment with methotrexate.[68] Nevertheless, a review of the literature clearly shows that patients with psoriasis who are treated with methotrexate are more likely to develop hepatic fibrosis, possibly because of their underlying disease or because of the concomitant treatments they are given.

Current guidelines call for the use of liver biopsy in patients with psoriasis who have received a cumulative dose of 1 to 1.5 g of methotrexate and who do not have a history of liver disease or alcoholism. Biopsy should be performed early in the course of treatment in patients with a history of hepatitis C, alcoholism, or other liver disease. Other risk factors for hepatotoxicity are obesity, diabetes, and abnormalities on liver function testing.

Pathologic liver changes caused by methotrexate therapy have been graded as follows: grade I, normal liver histology or mild fatty infiltration; grade II, moderate to severe fatty infiltration with portal tract inflammation and necrosis; grade IIIA, mild fibrosis; grade IIIB, moderate to severe fibrosis; and grade IV, cirrhosis. Methotrexate should be discontinued in patients with grade IIIB or IV pathologic liver changes. The importance of strict adherence to current guidelines for the administration of methotrexate is emphasized by the occurrence of methotrexate-induced cirrhosis necessitating liver transplantation in three patients with long-term psoriasis who did not undergo serial liver biopsies.[69]

Other side effects In addition to hepatotoxicity, other side effects of methotrexate therapy include bone marrow suppression, nausea, diarrhea, and stomatitis. Methotrexate is teratogenic and can cause reversible oligospermia. Pneumonitis can occur early in the course of treatment if methotrexate is administered in oncologic doses. Evaluation by tests of liver function, renal function, and blood elements must be made before and throughout the course of methotrexate therapy.

Certain drugs increase the toxicity of methotrexate by reducing renal tubular secretion; these drugs include salicylates, sulfonamides, probenecid, and penicillins. Other drugs increase toxicity by displacing methotrexate from its binding sites on plasma proteins; these drugs include salicylates, probenecid, barbiturates, and phenytoin. Many of the NSAIDs and trimethoprim-sulfamethoxazole enhance methotrexate toxicity.[68] Cases of pancytopenia after low-dose methotrexate therapy underscore the hazards of using this drug in patients with renal insufficiency or in patients who are concomitantly receiving drugs that increase methotrexate toxicity.[70]

Contraindications to treatment with methotrexate and indications for stopping treatment should be heeded. Constant medical supervision is necessary, and therapy must be stopped at once if toxicity develops.

Acitretin

Indications and dosage Acitretin, an oral retinoid, has FDA approval for the treatment of plaque psoriasis. It is highly effective in the treatment of pustular psoriasis and can be very effective as monotherapy for erythrodermic psoriasis. For plaque-type and guttate psoriasis, however, acitretin is most useful in combination with other treatments, particularly UVB and PUVA phototherapy.[71,72] Acitretin is initiated 1 to 2 weeks before UVB or PUVA therapy is started. With combination treatment, symptoms resolve much more quickly. Doses of only 10 to 25 mg daily are effective, thus minimizing retinoid side effects.[71,72] When used as monotherapy, acitretin is prescribed in doses of 25 mg daily, which can be increased to 50 mg a day or higher.

Side effects Acitretin side effects are dose related and are common with doses above 25 mg daily. Hair loss, cheilitis, desquamation of the palms and soles, sun sensitivity, and periungual pyogenic granulomas are among the mucocutaneous side effects. Hyperlipidemia is common but is easily controlled with lipid-lowering agents. Elevations in liver enzyme levels can occur, and enzyme levels must be monitored. Serial liver biopsies have not demonstrated hepatic fibrosis in patients treated with oral retinoids.[73]

Acitretin poses a significant risk of teratogenicity. Characteristic retinoid birth defects occur in a high proportion of fetuses exposed to even small amounts of the drug in utero. Acitretin is eliminated from the body much more quickly than its prodrug etretinate. In the presence of alcohol, however, acitretin is converted back to etretinate,[74] raising concerns that women of childbearing age who take acitretin and who later become pregnant would then be at risk for exposing their fetus to acitretin's teratogenic effects. The FDA therefore requires that acitretin not be given to women planning a pregnancy within 3 years.

Long-term side effects of oral retinoids include calcification of ligaments and tendons and osteoporosis.[75,76] The long-term safety of etretinate, acitretin's prodrug, was examined in a 5-year prospective study of 956 patients with psoriasis. The investigators concluded that with appropriate patient selection and monitoring, there was no substantially increased risk of side effects related to cardiovascular disease, cancer, diabetes, cataracts, and inflammatory bowel disease. Although joint symptoms improved in some patients, more patients had joint problems associated with etretinate. Etretinate also caused short-term changes in liver enzyme levels in some patients and, in rare cases, caused acute hepatitis. The long-term risk of liver disease and cirrhosis with etretinate, however, was less than that associated with comparable periods of methotrexate.[77]

Cyclosporine

Cyclosporine in a microemulsion formulation was approved by the FDA for the treatment of psoriasis after extensive worldwide experience. In dosages of 2.5 to 5 mg/kg/day, cyclosporine is highly effective for psoriasis. Even at such doses, however, it may be associated with significant side effects, which have limited its use in patients with severe or refractory disease.

Indications and dosage Cyclosporine is indicated for patients in whom phototherapy or methotrexate therapy has failed. The microemulsion formulation of cyclosporine is better ab-

sorbed than earlier formulations. It is available in gel capsules of 25 and 100 mg and is most commonly taken in divided doses twice daily. At dosages of 5 mg/kg/day, a response is usually seen within 4 weeks, and some patients respond as quickly as 1 week. It should be noted that in the United States, the package insert for cyclosporine recommends an upper dosage limit of 4 mg/kg/day, although worldwide experience regarding the efficacy and safety of this drug has established an upper limit of 5 mg/kg/day.[78] In the United States, the maximum FDA-approved duration of treatment of cyclosporine is 1 year.

Side effects Cyclosporine is associated with a number of side effects that are easily managed; other side effects are of greater concern. Hypertrichosis, tremors, paresthesias, headache, gingival hyperplasia, joint pain, and fatigue can occur. Elevations in serum lipid levels and minor elevations in liver enzyme levels are also common. Hypomagnesemia may require magnesium supplementation. The most serious common side effects are hypertension and nephrotoxicity. Hypertension can be managed by lowering the dose or by instituting treatment with calcium channel blockers such as amlodipine besylate. There is some evidence that in normotensive patients receiving cyclosporine, amlodipine therapy may prevent some of the nephrotoxicity that has been associated with this potent psoriasis treatment.[79]

Renal interstitial fibrosis and renal tubular atrophy are common in patients on long-term therapy with cyclosporine.[80,81] Consequently, serum creatinine levels must be monitored on a regular basis. If the serum creatinine level rises more than 30% above baseline (or more than 25%, according to the United States package insert), the dosage may have to be reduced.[78]

Organ transplant patients taking cyclosporine, as well as other immunosuppressive drugs, to prevent rejection have experienced an increase in lymphoproliferative diseases and skin cancers.[82,83] It is hoped that the lower doses and intermittent usage of cyclosporine in psoriasis patients will not be associated with an increase in malignancies, but caution must be exercised. In one study, no increase in lymphoproliferative disorders was found in rheumatoid arthritis patients who were treated with cyclosporine for a short period (median, 1.6 years), compared with a parallel group of rheumatoid patients who were not treated with cyclosporine.[84] Nevertheless, caution must be used with this powerful new psoriasis treatment.

Tacrolimus

Although tacrolimus does not have FDA approval for use in psoriasis, it is a potent immunosuppressive agent that may be substituted for cyclosporine in patients who cannot tolerate the hypertrichosis associated with this agent. Tacrolimus has proved to be effective in the treatment of psoriasis. In a double-blind trial, 50 patients with severe recalcitrant psoriasis were given either oral tacrolimus or placebo.[85] In the tacrolimus group, starting dosages were 0.5 mg/kg/day, and the dosages could be increased to 0.10 mg/kg at week 3 or 6 if patient response was judged to be insufficient. After 9 weeks of treatment, patients receiving tacrolimus had an 84% reduction in Psoriasis Area and Severity Index (PASI) scores.

As with cyclosporine, there are concerns about hypertension, nephrotoxicity, and immunosuppressive effects with tacrolimus. This drug is not associated with hypertrichosis or gingival hyperplasia. Tacrolimus has not been studied as extensively as cyclosporine for the treatment of psoriasis, and further investigations are warranted for this very effective antipsoriatic agent.

Hydroxyurea

Hydroxyurea may be considered for the treatment of psoriasis in patients with hepatic disease, because hepatotoxicity is uncommon with this agent.[86] Response is slower and less complete than with methotrexate, however, and resistance to hydroxyurea may develop more frequently. Hydroxyurea is administered orally at a dosage of 1 to 2 g/day. Careful monitoring of blood counts is necessary during therapy.

Sulfasalazine

Sulfasalazine does not have FDA approval for the treatment of psoriasis but is highly effective in selected patients. It is typically given in dosages of 3 to 4 g daily. In one study, over 25% of patients given sulfasalazine stopped the treatment because of side effects (cutaneous eruptions or nausea). In clinical practice, results have been less promising than in studies.[87]

Combination Therapy

Combinations of various psoriasis treatments have proved to be superior in efficacy to monotherapy. Acitretin is routinely used with UVB and PUVA, a combination that allows the use of smaller doses and minimizes toxicities of both retinoid therapy and phototherapy.[71,72] The combination of methotrexate and acitretin has been used successfully despite some concern that both drugs are hepatotoxic.[88] Careful monitoring of liver enzyme levels is essential. Methotrexate and cyclosporine can be used together, and their concurrent administration in small doses can result in greater efficacy and less toxicity than that which can be achieved with higher doses of either agent used alone.[89] Methotrexate has also been used very successfully in combination with UVB[90] and PUVA,[91] although there is some concern that methotrexate may potentiate the carcinogenic effect of PUVA.[92] Because cyclosporine has been associated with skin cancers, it is not routinely used in combination with PUVA. It can be used in combination with retinoids and mycophenolate mofetil.

Other Systemic Therapies

Mycophenolate mofetil, a drug that has FDA approval for the prevention of organ transplant rejection, is highly effective for some patients with psoriasis.[93] Mycophenolate mofetil is the prodrug of mycophenolic acid, a medication that was tested for psoriasis in the 1970s.[94] Although mycophenolic acid was found to be highly effective in the treatment of psoriasis, the manufacturers did not pursue FDA approval for that indication because of its side effects, which included gastrointestinal toxicity and an immunosuppressive effect that resulted in herpes zoster infections in more than 10% of treated patients.

6-Thioguanine is another anticancer chemotherapeutic agent that is highly effective for psoriasis. Unfortunately, it has been associated with bone marrow suppression in approximately 50% of patients.[95] Bone marrow toxicity from 6-thioguanine can be reduced by administering the drug two to three times a week rather than daily.[96]

Biologic Therapies

The ability to create molecules that target specific steps in the pathogenesis of psoriasis has led to the development of biologic agents that can treat psoriasis without the nephrotoxicity associated with cyclosporine and without the bone marrow and liver toxicities associated with methotrexate. Biologic agents are immunosuppressive, and their long-term toxicity is not known. As

with other immunosuppressive agents, there is concern about the potential to predispose patients to infections or malignancies. Several biologic agents have FDA approval for use in psoriasis—namely, alefacept, efalizumab, and etanercept. Others have been approved for use in other diseases but are undergoing clinical trials for use in psoriasis—namely, adalimumab and infliximab. Still other agents, such as onercept, a TNF-α blocking agent, and anti–IL-12, are at earlier stages of development.

Alefacept Alefacept is a fusion protein consisting of LFA-3 fused to the Fc portion of human IgG1. The LFA-3 portion of the molecule attaches to its naturally occurring receptor, CD2, on the surface of a resting T cell, thereby blocking T cell activation. The Fc portion of the molecule attaches to Fc receptors on natural killer cells and macrophages, resulting in apoptosis of the bound T cell.[97]

Alefacept originally received FDA approval as intravenous and intramuscular formulations, but it is now available only in the intramuscular form. Alefacept is administered weekly for 12 weeks in a dose of 15 mg. In one study, by 14 weeks after the start of therapy, 21% of patients achieved PASI 75 (75% reduction in disease from baseline) and 42% of patients achieved PASI 50 (50% reduction in disease from baseline). Improvement typically progresses after the completion of treatment, with maximal disease reduction 8 weeks after a second course of therapy; in one study, 33% of patients achieved PASI 75 and 57% achieved PASI 50 by this point.[98] The most striking benefit of alefacept therapy is the long duration of remission achieved in a subgroup of patients. In patients who achieved PASI 75, the median time to recurrence of psoriasis (as defined by maintenance of PASI 50) was 7 months after a single 12-week course of therapy and more than a year after two courses of therapy.[99]

Drawbacks of alefacept therapy include the high cost of the drug and the need for weekly $CD4^+$ T cell counts because the drug tends to reduce the number of these cells. The onset of action of alefacept is slow, with many patients achieving maximal response weeks after completing the 12-week course. Moreover, only a proportion of patients achieve a satisfactory response.

Efalizumab Efalizumab is a humanized monoclonal antibody directed against the CD11a portion of LFA-1. Efalizumab blocks the interaction between LFA-1 and ICAM-1, an interaction that is responsible for T cell activation and trafficking of T cells into inflamed skin. After a conditioning dose of 0.7 mg/kg the first week, patients self-administer subcutaneous injections of efalizumab at a dose of 1 mg/kg weekly. In double-blind, placebo-controlled trials, 22% to 39% of patients treated with weekly efalizumab for 12 weeks achieved PASI 75,[100-102] and nearly 60% of patients achieved PASI 50. With longer therapy, higher proportions of patients achieve greater degrees of improvement. Like the other biologic agents, efalizumab does not cause the nephrotoxicity associated with cyclosporine or the bone marrow or liver toxicity associated with methotrexate. The drug is fairly expensive, however, and flulike symptoms may develop after the first or second injection; a serious concern is the development of psoriasis rebound (defined as a worsening of psoriasis over baseline), which occurs in up to 15% of patients. To avoid psoriasis rebound, efalizumab should not be stopped abruptly but, rather, slowly converted to alternative therapies.

Etanercept Etanercept is a recombinant fusion protein that includes the p75 TNF receptor that binds to TNF-α, blocking its interaction with cell surface receptors. Etanercept originally received FDA approval for a dosage of 25 mg administered subcutaneously by the patient at home twice weekly for the treatment of psoriatic arthritis. Subsequently, etanercept received approval for the treatment of psoriasis at a dosage of 50 mg administered subcutaneously twice weekly for 3 months and then once weekly. In a double-blind, placebo-controlled, four-arm trial comparing placebo with three dosage regimens, analysis after 12 weeks of treatment showed that PASI 75 was achieved in 14% of patients who received 25 mg once a week, in 34% who received 25 mg twice a week, and in 49% who received 50 mg twice a week. Response rates were even higher at 24 weeks of therapy.[103]

The drawbacks of etanercept include its cost and the need to self-inject the medication on a long-term basis. Injection-site reactions, although common, are almost always minor and seldom require any treatment other than temporarily using a different site for injections. There is evidence that TNF-α blockers can cause an exacerbation of multiple sclerosis, so the drug should be avoided in patients with a personal or family history of demyelinating disease. Some controversy exists as to whether TNF-α blockers exacerbate chronic heart failure, and there is concern that the immunosuppressive effects of TNF-α blockers may contribute to an increase in the development of lymphoproliferative diseases.[104] Antinuclear antibodies also develop in etanercept-treated patients, but they are of questionable physiologic significance.

Infliximab Infliximab is a chimeric monoclonal antibody directed against TNF-α. In the short term (12 weeks), it is the most effective treatment for psoriasis, but it does not yet have FDA approval for this indication. It is administered by slow intravenous infusion at baseline, at weeks 2 and 6, and then every 8 weeks thereafter. In a double-blind, placebo-controlled trial evaluating patients at week 10, after only three infusions, 82% of patients achieved PASI 75.[105] Moreover, 55% of patients maintained PASI 50 or higher during 6 months of follow-up.

Like the other TNF-α blockers, infliximab is associated with worsening of chronic heart failure, multiple sclerosis, and lymphoproliferative diseases. In addition, infusion reactions develop in a significant proportion of patients; these appear to be related to the development of human antichimeric antibodies. Although infusion reactions are mild in the majority of patients, they can be severe, resulting in chest pain and hypotension. Pretreating patients with antibiotics is beneficial. TNF-α blocking plays a significant role in the control of mycobacterial infection, and an increase in reactivation of latent tuberculosis has been observed in patients treated with infliximab. Consequently, patients should undergo tuberculosis testing before starting on this medication.[106]

Adalimumab Adalimumab is a fully human monoclonal antibody against TNF-α. It has FDA approval for the treatment of rheumatoid arthritis and has been successfully tested for psoriasis.[107] Like the other biologics, adalimumab is not toxic to kidneys, liver, or bone marrow; however, also like the other biologic agents, it is quite expensive. The same concerns about heart failure, multiple sclerosis, and lymphoproliferative diseases that exist with etanercept and infliximab are also described in adalimumab's package insert. In a three-arm, placebo-controlled trial, PASI 75 was achieved by 53% of patients who received adalimumab every other week and by 80% of pa-

tients who received it weekly. An even greater number of patients achieved PASI 50. Adalimumab, 40 mg, is given by subcutaneous injection.

Prognosis

Psoriasis is usually lifelong, but the severity of the disease may vary, with periodic exacerbations and relative remissions in some patients. Although pustular psoriasis and erythrodermic psoriasis can be life-threatening, even stable plaque psoriasis can have a negative impact on overall health, possibly because of comorbid conditions such as psoriatic arthritis or obesity or because of complications of therapy.

Severe exacerbation of psoriasis taxes the ingenuity of even the most skilled clinician. Fortunately, because of the wide range of psoriasis therapies now available, clinicians are able to successfully treat almost all patients with psoriasis. The goal of therapy must be to minimize toxicity while achieving satisfactory improvement both in physical signs and symptoms and in patients' quality of life.

Elizabeth A. Abel, M.D., has been an investigator, consultant, or speaker for Abbott Laboratories, Allergan, Inc., Amgen, Inc., Biogen, Inc., Centocor, Inc., Connetics Corp., Genentech, Inc., and 3M.

Mark Lebwohl, M.D., has been an investigator, consultant, or speaker for Abbott Laboratories, Allergan, Inc., Amgen, Inc., Biogen, Inc., Centocor, Inc., Connetics Corp., Fujisawa Healthcare, Inc., Galderma Laboratories, Genentech, Inc., Novartis AG., and Warner Chilcott.

References

1. Stern RS, Nijsten T, Feldman SR, et al: Psoriasis is common, carries a substantial burden even when not extensive, and is associated with widespread treatment dissatisfaction. J Invest Dermatol Symp Proc 9:136, 2004

2. Koo J: Population-based epidemiologic study of psoriasis with emphasis on quality of life assessment. Dermatol Clin 14:485, 1996

3. Farber EM, Nall ML: The natural history of psoriasis in 5,600 patients. Dermatologica 148:1, 1974

4. Gottlieb SL, Gilleaudeau P, Johnson R, et al: Response of psoriasis to a lymphocyte-selective toxin (DAB389IL-2) suggests a primary immune, but not keratinocyte, pathogenic basis. Nat Med 1:442, 1995

5. Bagel J, Garland WT, Breneman D, et al: Administration of DAB389IL-2 to patients with recalcitrant psoriasis: a double-blind, phase II multicenter trial. J Am Acad Dermatol 38:938, 1998

6. Abrams JR, Lebwohl MG, Guzzo CA, et al: CTLA4Ig-mediated blockage of T-cell costimulation in patients with psoriasis vulgaris. J Clin Invest 103:1243, 1999

7. Griffiths CE: T-cell-targeted biologicals for psoriasis. Curr Drug Targets Inflamm Allergy 3:157, 2004

8. Lebwohl M, Tyring SK, Hamilton TK, et al: A novel targeted T-cell modulator, efalizumab, for plaque psoriasis. Efalizumab Study Group. N Engl J Med 349:2004, 2003

9. Lebwohl M, Christophers E, Langley R, et al: An international, randomized, double-blind, placebo-controlled phase 3 trial of intramuscular alefacept in patients with chronic plaque psoriasis. Alefacept Clinical Study Group. Arch Dermatol 139:719, 2003

10. Victor FC, Gottlieb AB, Menter A: Changing paradigms in dermatology: tumor necrosis factor alpha (TNF-alpha) blockade in psoriasis and psoriatic arthritis. Clin Dermatol 21:392, 2003

11. Farber EM, Nall ML, Watson W: Natural history of psoriasis in 61 twin pairs. Arch Dermatol 109:207, 1974

12. Capon F, Trembath RC, Barker JN: An update on the genetics of psoriasis. Dermatol Clin 22:339, 2004

13. Helms C, Cao L, Krueger JG, et al: A putative RUNX1 binding site variant between SLC9A3R1 and NAT9 is associated with susceptibility to psoriasis. Nat Genet 35:349, 2003

14. Lebwohl M, Tan MH: Psoriasis and stress. Lancet 351:82, 1998

15. Gaston L, Lassonde M, Bernier-Buzzanga J, et al: Psoriasis and stress: a prospective study. J Am Acad Dermatol 17:82, 1987

16. Rapp SR, Feldman SR, Exum ML, et al: Psoriasis causes as much disability as other major medical diseases. J Am Acad Dermatol 41:401, 1999

17. Telfer NR, Chalmers RG, Whale K, et al: The role of streptococcal infection in the initiation of guttate psoriasis. Arch Dermatol 128:39, 1992

18. Vincent F, Ross JB, Dalton M, et al: A therapeutic trial of the use of penicillin V or erythromycin with or without rifampin in the treatment of psoriasis. J Am Acad Dermatol 26:458, 1992

19. Rosenberg EW, Noah PW, Zanolli MD, et al: Use of rifampin with penicillin and erythromycin in the treatment of psoriasis: preliminary report. J Am Acad Dermatol 14:761, 1986

20. Obuch ML, Maurer TA, Becker B, et al: Psoriasis and human immunodeficiency virus infection. J Am Acad Dermatol 27:667, 1992

21. Abel EA: Diagnosis of drug-induced psoriasis. Semin Dermatol 11:269, 1992

22. Krueger GG: Psoriasis: current concepts of its etiology and pathogenesis. Yearbook of Dermatology. Dobson RL, Thiers BH, Eds. Year Book Medical Publishers, Chicago, 1982, p 13

23. Katayama H, Kawada A: Exacerbation of psoriasis induced by indomethacin. J Dermatol (Tokyo) 8:323, 1981

24. Magliocco MA, Gottlieb AB: Etanercept therapy for patients with psoriatic arthritis and concurrent hepatitis C virus infection: report of 3 cases. J Am Acad Dermatol 51:580, 2004

25. Gupta MA, Gupta AK, Watteel GN: Cigarette smoking in men may be a risk factor for increased severity of psoriasis of the extremities. Br J Dermatol 135:859, 1996

26. Naldi L, Peli L, Parazzini F: Association of early-stage psoriasis with smoking and male alcohol consumption: evidence from an Italian case-control study. Arch Dermatol 135:1479, 1999

27. Naldi L, Parazzini F, Peli L, et al: Dietary factors and the risk of psoriasis: results of an Italian case control study. Br J Dermatol 134:101, 1996

28. Green MS, Prystowsky JH, Cohen SR, et al: Infectious complications of erythrodermic psoriasis. J Am Acad Dermatol 34:911, 1996

29. Calvert HT, Smith MA, Wells RS: Psoriasis and the nails. Br J Dermatol 75:415, 1963

30. Mease PJ: Recent advances in the management of psoriatic arthritis. Curr Opin Rheumatol 16:366, 2004

31. Gladman DD, Hing EN, Schentag CT, et al: Remission in psoriatic arthritis. J Rheumatol 28:1045, 2001

32. Lebwohl M: Advances in psoriasis therapy. Dermatol Clin 18:13, 2000

33. Feldman SR, Fleischer AB, Cooper JZ: New topical treatments change the pattern of treatment of psoriasis: dermatologists remain the primary providers of this care. Int J Dermatol 39:41, 2000

34. Cornell RC, Stoughton RB: Correlation of the vasoconstriction assay and clinical activity in psoriasis. Arch Dermatol 121:63, 1985

35. Katz HI, Prawer SE, Medansky RS, et al: Intermittent corticosteroid maintenance treatment of psoriasis: a double-blind multicenter trial of augmented betamethasone dipropionate ointment in a pulse dose treatment regimen. Dermatologica 183:269, 1991

36. Lebwohl M, Siskin SB, Epinette W, et al: A multicenter trial of calcipotriene ointment and halobetasol ointment to either agent alone for the treatment of psoriasis. J Am Acad Dermatol 35:268, 1996

37. Lebwohl M, Yoles A, Lombardi K, et al: Calcipotriene ointment and halobetasol ointment in the long-term treatment of psoriasis: effects on the duration of improvement. J Am Acad Dermatol 39:447, 1998

38. Ramsay CA, Schwartz BE, Lowson D, et al: Calcipotriol cream combined with twice weekly broad-band UVB phototherapy: a safe, effective and UVB-sparing antipsoriatric combination treatment. The Canadian Calcipotriol and UVB Study Group. Dermatology 200:17, 2000

39. Speight EL, Farr PM: Calcipotriol improves the response of psoriasis to PUVA. Br J Dermatol 130:79, 1994

40. Patel B, Siskin S, Krazmien BA, et al: Compatibility of calcipotriene with other topical medications. J Am Acad Dermatol 38:1010, 1998

41. Lebwohl M, Hecker D, Martinez J, et al: Interactions between calcipotriene and ultraviolet light. J Am Acad Dermatol 37:93, 1997

42. Georgiou S, Tsambaos D: Hypercalcaemia and hypercalciuria after topical treatment of psoriasis with excessive amounts of calcipotriol. Acta Derm Venereol 79:86, 1999

43. Guenther LC: Fixed-dose combination therapy for psoriasis. Am J Clin Dermatol 5:71, 2004

44. Franssen ME, de Jongh GJ, van Erp PE, et al: A left/right comparison of twice-daily calcipotriol ointment and calcitriol ointment in patients with psoriasis: the effect on keratinocyte subpopulations. Acta Derm Venereol 84:195, 2004

45. Koo JY: Tazarotene in combination with phototherapy. J Am Acad Dermatol 39:S144, 1998

46. Hecker D, Worsley J, Yueh G, et al: Interactions between tazarotene and ultraviolet light. J Am Acad Dermatol 41:927, 1999

47. Lowe NJ, Breeding J, Wortzman MS: The pharmacological variability of crude coal tar. Br J Dermatol 126:608, 1992

48. Fiore M: Practical aspects of anthralin therapy. Cutis 46:351, 1990

49. Abel EA, O'Connell BM, Farber EM: Psoriasis Day Care Center treatment at Stanford: part-time and full-time programs. Int J Dermatol 26:500, 1987

50. Schaefer H, Farber EM, Goldberg L, et al: Limited application period for dithranol in psoriasis: preliminary report on penetration and clinical efficacy. Br J Dermatol 102:571, 1980

51. Volden G, Bjornberg A, Tegner E, et al: Short-contact treatment at home with micanol. Acta Derm Venereol Suppl (Stockh) 172:20, 1992

52. Even-Paz Z, Gumon R, Kipnis V, et al: Dead Sea sun vs. Dead Sea water in the treatment of psoriasis. Dermatol Treat 7:83, 1996

53. Frentz G, Olsen JH, Avrach WW: Malignant tumours and psoriasis: climatotherapy

at the Dead Sea. Br J Dermatol 141:1088, 1999

54. Petrozzi JW, Barton JO, Kaidbey KH, et al: Updating the Goeckerman regimen for psoriasis. Br J Dermatol 98:437, 1978

55. Lowe NJ, Wortzman MS, Breeding J, et al: Coal tar phototherapy for psoriasis reevaluated: erythemogenic versus suberythemogenic ultraviolet with a tar extract in oil and crude coal tar. J Am Acad Dermatol 8:781, 1983

56. Stern RS, Gange RW, Parrish JA, et al: Contribution of topical tar oil to ultra-violet B phototherapy for psoriasis. J Am Acad Dermatol 14:742, 1986

57. Lowe NJ, Stern RS: Contribution of topical tar oil to ultraviolet B phototherapy for psoriasis. J Am Acad Dermatol 15:1053, 1986

58. Barbagallo J, Spann CT, Tutrone WD, et al: Narrowband UVB phototherapy for the treatment of psoriasis: a review and update. Cutis 68:345, 2001

59. Dawe RS, Cameron H, Yule S, et al: UV-B phototherapy clears psoriasis through local effects. Arch Dermatol 138:1071, 2002

60. Abel EA: Administration of PUVA therapy: protocols, indications, and cautions. Photochemotherapy in Dermatology. Abel EA, Ed. Igaku-Shoin Medical Publishers, New York, 1992, p 75

61. Stern RS, Laird N: The carcinogenic risk of treatments for severe psoriasis. Cancer 73:2759, 1994

62. Stern RS, Thibodeau LA, Kleinerman RA, et al: Risk of cutaneous carcinoma in patients treated with oral methoxsalen photochemotherapy for psoriasis. N Engl J Med 300:809, 1979

63. Stern RS, Laird N, Melski J, et al: Cutaneous squamous-cell carcinoma in patients treated with PUVA. N Engl J Med 310:1156, 1984

64. Stern RS, Nichols KT, Vakeva LH: Malignant melanoma in patients treated for psoriasis with methoxsalen (psoralen) and ultraviolet A radiation (PUVA). The PUVA Follow-Up Study.N Engl J Med 336:1041, 1997

65. Lerman S, Borkman RF: A method for detecting 8-methoxypsoralen in the ocular lens. Science 197:1287, 1997

66. Lerman S: Ocular phototoxicity and psoralen plus ultraviolet radiation (320–400 nm) therapy: an experimental and clinical evaluation. J Natl Cancer Inst 69:287, 1982

67. Jeffes EB III, McCullough JL, Pittelkow MR, et al: Methotrexate therapy of psoriasis: differential sensitivity of proliferating lymphoid and epithelial cells to the cytotoxic and growth-inhibitory effects of methotrexate. J Invest Dermatol 104:183, 1995

68. Roenigk HH Jr, Auerbach R, Maibach H, et al: Methotrexate in psoriasis: consensus conference. J Am Acad Dermatol 38:478, 1996

69. Gilbert SC, Lintmalm G, Menter A, et al: Methotrexate-induced cirrhosis requiring liver transplantation in three patients with psoriasis: a word of caution in light of the expanding use of this "steroid-sparing" agent. Arch Intern Med 150:889, 1990

70. Al-Awadhi A, Dale P, McKendry RJ: Pancytopenia associated with low dose methotrexate therapy: a regional survey. J Rheumatol 20:1121, 1993

71. Lebwohl M: Acitretin in combination with UVB or PUVA. J Am Acad Dermatol 4:S22, 1999

72. Lebwohl M, Drake L, Menter A, et al: Consensus conference: acitretin in combination with UVB or PUVA in the treatment of psoriasis. J Am Acad Dermatol 45:544, 2001

73. Roenigk HH Jr, Callen JP, Guzzo CA, et al: Effects of acitretin on the liver. J Am Acad Dermatol 41:585, 1999

74. Larsen FG, Jakobsen P, Knudsen J, et al: Conversion of acitretin to etretinate in psoriatic patients is influenced by ethanol. J Invest Dermatol 100:623, 1993

75. DiGiovanna JJ, Sollitto RB, Abangan DL, et al: Osteoporosis is a toxic effect of long-term etretinate therapy. Arch Dermatol 131:1263, 1995

76. Wilson DJ, Kay V, Charig M, et al: Skeletal hyperostosis and extraosseous calcification in patients receiving long-term etretinate (Tigason). Br J Dermatol 119:597, 1988

77. Stern RS, Fitzgerald E, Ellis CN, et al: The safety of etretinate as long-term therapy for psoriasis: results of the Etretinate Follow-Up Study. J Am Acad Dermatol 33:44, 1995

78. Lebwohl M, Ellis C, Gottlieb A, et al: Cyclosporine consensus conference: with emphasis on the treatment of psoriasis. J Am Acad Dermatol 39:464, 1998

79. Raman GV, Campbell SK, Farrer A, et al: Modifying effects of amlodipine on cyclosporin A–induced changes in renal function in patients with psoriasis. J Hypertens Suppl 16:S39, 1998

80. Zachariae H, Kragballe K, Hansen HE, et al: Renal biopsy findings in long-term cyclosporin treatment of psoriasis. Br J Dermatol 136:531, 1997

81. Lowe NJ, Wieder JM, Rosenbach A, et al: Long-term low-dose cyclosporine therapy for severe psoriasis: effects on renal function and structure. J Am Acad Dermatol 35:710, 1996

82. Jensen P, Hansen S, Moller B, et al: Skin cancer in kidney and heart transplant recipients and different long-term immunosuppressive therapy regimens. J Am Acad Dermatol 40:177, 1999

83. Srivastava T, Zwick DL, Rothberg PG, et al: Posttransplant lymphoproliferative disorder in pediatric renal transplantation. Pediatr Nephrol 13:748, 1999

84. van den Borne BE, Landewe RB, Houkes I, et al: No increased risk of malignancies and mortality in cyclosporin A–treated patients with rheumatoid arthritis. Arthritis Rheum 41:1930, 1998

85. Systemic tacrolimus (FK 506) is effective for the treatment of psoriasis in a double-blind, placebo-controlled study. The European FK 506 Multicenter Psoriasis Study Group. Arch Dermatol 132:419, 1996

86. Smith CH: Use of hydroxyurea in psoriasis. Clin Exp Dermatol 24:2, 1999

87. Gupta AK, Ellis CN, Siegel MT, et al: Sulfasalazine improves psoriasis: a double-blind analysis. Arch Dermatol 126:487, 1990

88. Roenigk HH Jr: Acitretin combination therapy. J Am Acad Dermatol 41:S18, 1999

89. Wong KC, Georgouras K: Low dose cyclosporin A and methotrexate in the treatment of psoriasis. Acta Derm Venereol 79:87, 1999

90. Paul BS, Momtaz K, Stern RS, et al: Combined methotrexate–ultraviolet B therapy in the treatment of psoriasis. J Am Acad Dermatol 7:758, 1982

91. Morison WL, Momtaz K, Parrish JA, et al: Combined methotrexate–PUVA therapy in the treatment of psoriasis. J Am Acad Dermatol 6:46, 1982

92. Stern RS, Laird N: The carcinogenic risk of treatments for severe psoriasis. Photochemotherapy Follow-up Study. Cancer 73:2759, 1994

93. Kirby B, Yates VM: Mycophenolate mofetil for psoriasis. Br J Dermatol 139:357, 1998

94. Epinette WW, Parker CM, Jones EL, et al: Mycophenolic acid for psoriasis: a review of pharmacology, long-term efficacy, and safety. J Am Acad Dermatol 17:962, 1987

95. Zackheim HS, Glogau RG, Fisher DA, et al: 6-Thioguanine treatment of psoriasis: experience in 81 patients. J Am Acad Dermatol 30:452, 1994

96. Silvis NG, Levine N: Pulse dosing of thioguanine in recalcitrant psoriasis. Arch Dermatol 135:433, 1999

97. Krueger GG: Selective targeting of T cell subsets: focus on alefacept: a remittive therapy for psoriasis. Expert Opin Biol Ther 2:431, 2002

98. Hodak E, David M: Alefacept: a review of the literature and practical guidelines for management. Dermatol Ther 17:383, 2004

99. Gordon KB, Langley RG: Remittive effects of intramuscular alefacept in psoriasis. J Drugs Dermatol 2:624, 2003

100. Lebwohl M, Tyring SK, Hamilton TK, et al: A novel targeted T-cell modulator, efalizumab, for plaque psoriasis. Efalizumab Study Group. N Engl J Med 349:2004, 2003

101. Leonardi CL: Efalizumab in the treatment of psoriasis. Dermatol Ther 17:393, 2004

102. Gordon KB, Papp KA, Hamilton TK, et al: Efalizumab for patients with moderate to severe plaque psoriasis: a randomized controlled trial. Efalizumab Study Group. JAMA 290:3073, 2003

103. Leonardi CL, Powers JL, Matheson RT, et al: Etanercept as monotherapy in patients with psoriasis. Etanercept Psoriasis Study Group. N Engl J Med 349:2014, 2003

104. Brown SL, Greene MH, Gershon SK, et al: Tumor necrosis factor antagonist therapy and lymphoma development: twenty-six cases reported to the Food and Drug Administration. Arthritis Rheum 46:3151, 2002

105. Chaudhari U, Romano P, Mulcahy LD, et al: Efficacy and safety of infliximab monotherapy for plaque-type psoriasis: a randomized trial. Lancet 357:1842, 2001

106. Keane J, Gershon S, Wise RP, et al: Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. N Engl J Med 345:1098, 2001

107. Patel T, Gordon KB: Adalimumab: efficacy and safety in psoriasis and rheumatoid arthritis. Dermatol Ther 17:427, 2004

IV ECZEMATOUS DISORDERS, ATOPIC DERMATITIS, AND ICHTHYOSES

Seth R. Stevens, m.d.

Eczematous Disorders

Eczematous dermatitis, or eczema, is a skin disease that is characterized by erythematous vesicular, weeping, and crusting patches. Although the term eczema is often used as a diagnosis, it can in fact be used appropriately to describe lesions seen in several diseases. Itching is a characteristic symptom, and epidermal intercellular edema (spongiosis) is a characteristic histopathologic finding of eczematous conditions. The term eczema is also commonly used to describe atopic dermatitis [*see* Atopic Dermatitis, *below*].

CONTACT DERMATITIS

Contact dermatitis, a paradigmatic example of an eczematous disorder, is common and well studied [*see 2:V Contact Dermatitis and Related Disorders*]. Contact dermatitis can be either allergic or irritant in etiology. Allergic contact dermatitis differs from other eczematous disorders in that determination of the offending contactant is an important part of the evaluation. If the patient's history does not provide the answer, the body site of the lesion may (e.g., head involvement in allergy to paraphenylenediamine in hair dye). Patch testing may be required to confirm the diagnosis.[1]

The manifestations of irritant contact dermatitis are similar to those of allergic contact dermatitis[2]; in the irritant form, however, the mechanism is not immunologic. Given sufficient concentration and duration of contact, offending agents will induce irritation in anyone's skin. Detergents, acids, alkalis, solvents, formaldehyde, and fiberglass are common causes.

SEBORRHEIC DERMATITIS

Seborrheic dermatitis is another common eczematous condition [*see 2:II Papulosquamous Disorders*]. Clinically, seborrheic dermatitis may exist without vesicle formation. Lesional morphology is usually a greasy scale on erythematous patches; however, the scale may be dry and the patches may have an orange hue. Scalp, eyebrows, mustache area, nasolabial folds, and chest are typical areas of involvement. Psoriasis may be part of the differential diagnosis. Treatment is with shampoos containing selenium sulfide, zinc pyrithione, tar, or ketoconazole; emollients; and mild (nonfluorinated) topical steroids. Antimicrobial therapy directed at the commensal yeast *Pityrosporum ovale* can be effective, although a causative role of the organism remains unproved.

OTHER ECZEMATOUS DERMATITIDES

Two other eczematous dermatitides are nummular eczema and dyshidrotic eczema (pompholyx). Nummular eczema describes well-demarcated, coin-shaped eczematous patches that are usually 2 to 4 cm (rarely more than 10 cm) in diameter. The lesions are quite pruritic and require potent topical steroids, antihistamines, and, occasionally, intralesional or systemic corticosteroids for treatment. Dyshidrotic eczema presents as a vesicular eruption of the hands and feet, accompanied on rare occasions by hyperhidrosis. Typically, 1 to 2 mm vesicles appear on the sides of fingers, although more extensive involvement can occur. Treatment is with compresses and soaks, antipruritics, topical steroids, and, in severe recalcitrant cases, systemic corticosteroids. Photochemotherapy with topical psoralen and ultraviolet A irradiation (PUVA) may also be effective.

Atopic Dermatitis

Atopic dermatitis (AD) is a common chronic inflammatory dermatosis that generally begins in infancy. The term atopy was coined in the early 1920s to describe the associated triad of asthma, allergic rhinitis, and dermatitis.[3] Children with AD are at increased high risk of developing asthma and allergic rhinitis, and the risk is further increased for patients with a family history of atopy.[4] The role of reaginic antibodies and allergies in the etiology of AD is controversial; in 80% of patients with AD, however, serum immunoglobulin IgE is elevated, sometimes markedly.

ETIOLOGY AND PATHOGENESIS

The expression of AD is a complex integration of environmental and genetic factors. The lifetime prevalence is estimated to be 30% of the population,[5-7] possibly because of increasing contact with causative agents in the environment. Epidemiologic data suggest a genetic influence—25% of dizygotic twins and 75% of monozygotic twins are concordant for AD.[8] The condition develops in 60% of children who have one affected parent and in 80% of children with two affected parents.[9] The defect is likely carried in the immune system, because both antigen-specific IgE reactivity and AD have been transplanted from an AD-affected bone marrow donor to a previously unaffected recipient.[10] Candidate genes continue to be investigated.[11]

AD can be quickly exacerbated by environmental trigger factors.[12] Wool, lanolin, and harsh detergents are particularly irritating. Emotional stress can also lead to flares, which are characterized by increased itch, erythema, vesiculation, and excoriation, as well as expanded area of involvement. The role of airborne and foodborne allergens is difficult to assess. Although patients with AD frequently have circulating dust mite antigen-specific IgE and Th2 $CD4^+$ T cells,[13] hyposensitization infrequently results in improvement. Contact urticaria to food occurs in AD,[14] but generalized exacerbation after eating is rare. In the absence of a strong supporting history, elimination diets are rarely effective in treating AD. A role has been frequently suggested for cow's milk in inducing AD; however, studies examining the association of AD and early feeding with cow's milk have shown varying results.[15,16] Meta-analyses indicate that exclusive breast-feeding during the first 3 months of life is associated with lower incidence rates of atopic dermatitis during childhood in children with a family history of atopy.[17]

Gut microflora may be a natural source of immune modulation that prevents atopic dermatitis. In a double-blind, randomized, placebo-controlled trial, a probiotic containing a strain of *Lactobacillus* was administered prenatally to mothers who had at least one first-degree relative with atopy and 6 months postnatally to their infants; the frequency of AD in the group receiving *Lactobacillus* was half that of the placebo group.[18] These findings

suggest that impairment of the intestinal mucosal barrier may be involved in the pathogenesis of AD, and strengthening of the mucosal barrier with probiotic bacteria may help prevent AD in high-risk infants. Although intriguing, these results await confirmation.

Mechanisms have been proposed to explain a link between *Staphylococcus aureus* and exacerbations of AD,[19] including effects of cell wall constituents to increase expression of IgE, IgE receptor, and enterotoxin B, a superantigen that activates T cells in an antigen-independent fashion.[20]

The apparent paradox of reduced cell-mediated immunity[21,22] and hyperimmunoglobulinemia E seen in AD is addressed by the so-called Th1/Th2 model of helper T cells. In this model of the murine immune system, $CD4^+$ T cells are divided into two mutually exclusive classes on the basis of cytokine secretion: Th1 cells, which secrete cytokines that promote cell-mediated immunity (e.g., interleukin-2 [IL-2], interferon gamma), and Th2 cells, which secrete cytokines that promote humoral immunity and eosinophil function (e.g., IL-4 and IL-5). Atopy, including AD, has been seen as the paradigmatic condition of a so-called Th1-deficient state. Refinements have shown a heterogeneity of responses within different AD lesions, however. The current model is that blood and acute lesions of AD patients are more often dominated by Th2 cells, whereas chronic lesions are more often dominated by Th1 cells.[23]

Table 1 Diagnostic Criteria for Atopic Dermatitis[14]

- Major criteria
  - Personal or family history of atopy (atopic dermatitis, allergic rhinitis, allergic conjunctivitis, allergic blepharitis, or asthma)
  - Characteristic morphology and distribution of lesions
  - Pruritus
  - Chronic or chronically recurring dermatosis
- Minor features
  - Hyperimmunoglobulinemia E
  - Food intolerance
  - Intolerance to wool and lipid solvents
  - Recurrent skin infections
  - Xerosis
  - Sweat-induced pruritus
  - White (not red) dermatographism
  - Ichthyosis
  - Chronically scaling scalp
  - Accentuation of hair follicles
  - Recurrent conjunctivitis
  - Anterior subcapsular cataracts and keratoconus
  - Morgan line, or Dennie sign (single or double creases in the lower eyelids)
  - Periorbital darkening (allergic shiner)
  - Pityriasis alba (hypopigmented, scaling patches, typically on the cheeks)
  - Cheilitis
  - Anterior neck folds
  - Keratosis pilaris (perifollicular papules with keratotic plugs, typically on the arms and thighs)
  - Nipple eczema
  - Hyperlinear palms (increased folds, typically on the thenar or hypothenar eminence)
  - Recurrent hand and foot dermatitis
  - Exacerbation of symptoms by environmental or emotional factors

Hyperstimulatory dendritic antigen-presenting cells (Langerhans cells) are present in patients with AD.[24] One proposed mechanism for the augmented function of Langerhans cells in AD is the binding of antigen-specific IgE and antigen to the IgE receptors on Langerhans cells as a means of antigen focusing.[25] Another antigen-presenting cell, the monocyte, also manifests altered function in AD. Cyclic adenosine monophosphate (cAMP) phosphodiesterase has increased activity in monocytes of patients with AD—leading to hyperproduction of prostaglandin E_2, among other effects. Increased cAMP phosphodiesterase in AD may explain aberrant adrenergic responses, and the increased prostaglandin E_2 leads to diminished interferon-gamma production. Additionally, monocytes secrete IL-10 in AD, which further augments the so-called Th2 responses.[26] Altered cyclic nucleotide metabolism leads to excessive release of histamine by basophils and, potentially, to mast cell degranulation. High levels of cAMP phosphodiesterase are found in the umbilical cord blood of infants of AD-affected parents.[27] This finding may indicate an early, if not primary, defect in the disease that may become the basis of a diagnostic laboratory test.

Because IL-5 is a critical eosinophil growth factor and activating cytokine, blood eosinophilia may be expected to occur in a Th2 disease such as AD[28]; tissue eosinophilia, however, is variable. Cutaneous endothelial cells are also activated in AD, leading to increased expression of adhesion molecules and recruitment of leukocytes into the skin (i.e., dermatitis).

DIAGNOSIS

AD remains a clinical diagnosis. Major diagnostic criteria are (1) personal or family history of atopy (AD, allergic rhinitis, allergic conjunctivitis, allergic blepharitis, or asthma); (2) characteristic morphology and distribution of lesions; (3) pruritus; and (4) chronic or chronically recurring dermatosis. Several minor features can be added [*see Table 1*].[14] Pruritus is a consistent feature of AD. The lack of itching or of another major diagnostic criterion should prompt consideration of alternative diagnoses [*see* Differential Diagnosis, *below*]. Cutaneous signs can vary, depending on the age of the lesions.

Acute lesions of AD are eczematous—erythematous, scaling, and papulovesicular. Weeping and crusted lesions may develop [*see Figure 1*]. Scratching results acutely in linear excoriations, presenting as erosions or a hemorrhagic crust. In extremely severe cases, exfoliative dermatitis (erythroderma) may occur, with generalized redness, scaling, weeping, and crusting. There may be accompanying systemic toxicity, sepsis, lymphadenopathy, altered thermoregulation (either hyperthermia or hypothermia), and high-output cardiac failure. Erythroderma is a potentially life-threatening condition.

Chronic lesions tend not to be eczematous (thus, atopic eczema is not an ideal synonym for AD). Instead, lichenified plaques [*see Figure 2*] or nodules predominate. Lichenification denotes areas of thickened skin divided by deep linear furrows. Lichenified plaques result from repeated rubbing or scratching and thus often occur in areas of predilection, such as the popliteal and antecubital fossae. As is typical of lesions in AD, lichenification is poorly demarcated. There may be accompanying acute signs. Lichenified lesions are very difficult to treat; once established, they may persist for months even with adequate therapy and avoidance of rubbing or scratching.

Clinical expression of AD also varies with the age of the patient. The infantile stage of AD occurs up to approximately 2 years of age. Of all cases of AD, approximately 90% arise before

the fifth year and 60% in the first year of life; onset before 2 months of age is unusual, however.[8] During infancy, ill-defined, erythematous scaling patches and confluent, edematous papules and vesicles are typical. These lesions may become crusted and exudative. Intense pruritus leads to scratching, which induces linear excoriations and, with time, lichenification. Before the infant begins to crawl, the scalp and face are most often involved [*see Figure 3*], although lesions may be seen anywhere. After the child begins crawling, the extensor surfaces—particularly the knees—become involved. Involvement of fingers can be severe if the child sucks them frequently. Intense pruritus can lead to sleep disturbances of child and parents. Other features may arise [*see Table 1*]. Perifollicular accentuation and papules are commonly seen at any point in the life of an atopic patient, particularly in persons of Asian or African ancestry.

During childhood, the clinical features evolve into those seen in adults. Lesions tend to become less eczematous and drier, with increasing flexural and neck involvement. Scaling, fissured, and crusted hands may become especially troublesome. Infraorbital folds (sometimes called Morgan lines or the Dennie sign) and pityriasis alba can appear. Chronic or chronically relapsing pruritic, erythematous, papulovesicular eruptions that progress to scaling, lichenified dermatitis in a flexural distribution typify adult AD. Extensive areas of skin may be involved, including the face, chest, neck, flanks, and hands. Areas of dyspigmentation may result from repeated skin trauma. Approximately 10% to 15% of childhood AD persists after puberty.[8]

AD that begins after 20 years of age has been termed adult-onset atopic dermatitis.[29] This condition should be considered in patients with characteristic features of AD.

There are many associated features of AD. Asthma and allergic rhinitis, the major and minor criteria, respectively, have already been mentioned. Another important association, cutaneous infection, is related to diminished cutaneous cell-mediated immunity and defective chemotaxis. *S. aureus* is usually found on AD skin, and its density correlates with lesion severity.[30] Although such observations have implicated *S. aureus* as a cause of AD,[19,31] it is also clear that reduction in AD lesions reduces bacterial colonization.[32] Regardless, the high bacterial counts in lesional skin and the relative ease of their reduction suggest the desirability of extra efforts (e.g., use of topical steroids) to reduce the presence of *S. aureus* before elective procedures are performed through involved skin. Frank infection also occurs more commonly in AD, which results in pustules and oozing, crusted lesions.

Cutaneous fungal and viral infections also occur frequently and with increased severity in patients with AD. Eczema herpeticum, an extensive eruption of 2 to 3 mm vesicles, pustules, and punched-out erosions caused by herpes simplex virus, may coalesce into extensive areas of eroded skin. Frequently, the condition is most severe on the face (where it often arises from a herpetic lesion) and diminishes as it progresses to the trunk and extremities. Secondary bacterial infection is common. Lymphadenopathy, fever, and malaise may develop. Antiviral and antibiotic therapy can be lifesaving and should be started empirically upon presentation. Tzanck test, viral culture, and direct fluorescent antibody detection of viral antigens can confirm the diagnosis.

Molluscum contagiosum and common warts are also problematic in patients with AD, as are dermatophyte infections. Because of similar appearance, foot eczema must be distinguished from tinea pedis by potassium hydroxide preparation or fungal culture.

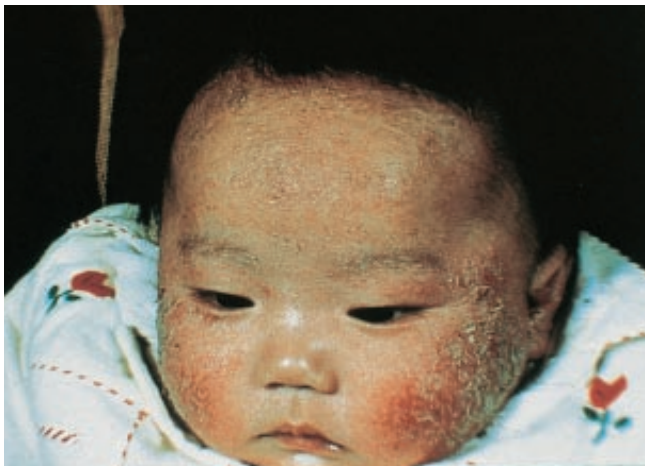


Figure 1 **Extensive, severe, weeping, crusted acute eczematous patches on the face of this infant are characteristic of patients in this age group.**

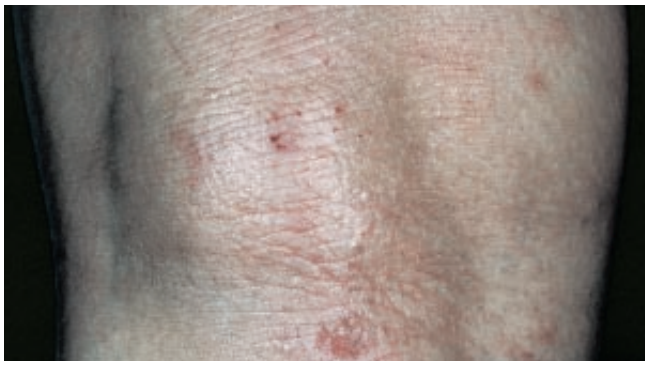


Figure 2 **Lichenified patches appear after chronic rubbing of eczematous patches. These lesions are characteristic of chronic allergic contact dermatitis and atopic dermatitis.**

Numerous ocular complications of AD exist.[33] These include anterior subcapsular cataracts, retinal detachment, keratoconus, blepharitis, conjunctivitis, and iritis.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of AD includes the eczematous conditions and ichthyoses described in this chapter and other immunologic, metabolic, neoplastic, and rheumatologic disorders [*see Table 2*]. Because 80% to 85% of patients with occupational hand dermatitis have AD, the possibility of coexisting AD and contact dermatitis needs to be considered. Another important element of the differential diagnosis is cutaneous T cell lymphoma. Cutaneous T cell lymphoma can arise clinically as scaling, erythematous patches or exfoliative erythroderma. The classic distribution—near axillae, buttocks, and groin—is distinct from that of AD, and patches are frequently well demarcated. There is often sufficient clinical overlap between the two conditions, however, to necessitate further investigation, including histology, immunophenotyping, and gene-rearrangement analysis of T cell receptors. Cutaneous T cell lymphoma can arise in patients with AD, and the lack of conclusive clinical or laboratory tests for either disease can make distinction difficult. Reassessment from time to time in such cases is recommended.[34]

Reduction of Trigger Factors

Reduction of trigger factors (e.g., harsh chemicals, detergents, and wool) and avoidance of occupations that require contact with trigger factors (e.g., hairdressing, nursing, and construction) can be helpful.[35] Appropriate behaviors should be taught to patients and parents early during life, when habits are more easily formed.[36,37]

Bland Emollients

The use of mild, nonalkali soaps and frequent use of emollients are important elements in the long-term management of AD. Because moisture evaporating off the skin can trigger flares, bathing is sometimes discouraged. A better approach is the prompt application of an emollient such as petrolatum (finishing within 3 minutes of the end of the bath), which can serve to seal the moisture from the bath. Lotions and creams containing high amounts of water are usually inadequate, however, and can actually worsen AD. Products containing hydroxy acids, phenol, or urea can reduce dryness and scaling, but these can sting inflamed skin and should therefore be used with caution. Because of a specific reduction of ceramides in AD, a lotion that provides excess ceramides relative to other lipids has been shown to have a therapeutic advantage in AD.[38] Bubble baths and scented salts and oils can be irritating. Scalp care should include a bland shampoo. Topical tar products, such as shampoos and bath solutions, and topical creams and lotions containing 5% to 10% liquor carbonis detergens can help. Baths, soaks, and compresses with Burow solution can ameliorate crusted, infected, eczematous patches. Cotton clothing, washed to remove finishing (which often releases formaldehyde), is preferable to wool or synthetics.

Corticosteroids

Topical corticosteroids are another mainstay of therapy. Application immediately after bathing improves cutaneous penetration. Lowering the risk of side effects with less potent preparations must be balanced against gaining control of a flare quickly with more potent preparations. Long-term use of inadequately potent topical corticosteroids may pose a greater risk of adverse effects than brief use of more potent agents followed by a rapid taper to bland emollients. Because steroid-induced cutaneous atrophy is a greater risk on the face, in intertriginous areas (e.g., groin, axillae, and inframammary folds), and under diapers, less potent steroids (e.g., hydrocortisone and desonide) should be used in these areas, and they should be used with particular caution. For the remainder of the body, midpotency preparations, such as 0.1% triamcinolone acetonide, are helpful. More potent ointments, such as fluocinonide and desoximetasone, are useful for lichenified plaques. Flurandrenolide tape is useful for nodular prurigo (so-called picker's nodules) because it also physically protects the area from manipulation. For the scalp, solutions are preferred.

Systemic corticosteroids (e.g., prednisone, 20 to 80 mg/day orally) may be useful to treat severe, acute flares. Because of the risks of gastrointestinal, endocrine, skeletal, central nervous system, and cardiovascular complications, however, they should not be used more than twice yearly.

Table 2 Differential Diagnosis of Atopic Dermatitis

Type	*Disorders*
Dermatitides	Allergic contact dermatitis Dermatitis herpetiformis Irritant contact dermatitis (may be concomitant with atopic dermatitis) Nummular eczema Seborrheic dermatitis
Ichthyoses	Ichthyosis vulgaris
Immunologic disorders	Graft versus host disease HIV-associated dermatosis Hyperimmunoglobulinemia E syndrome Wiskott-Aldrich syndrome
Infectious diseases	Scabies Dermatophytosis
Metabolic disorders	Zinc deficiency Various inborn errors of metabolism
Neoplastic disorders	Cutaneous T cell lymphoma
Rheumatologic disorders	Dermatomyositis

Calcineurin Inhibitors

The steroid-free topical calcineurin inhibitors, tacrolimus ointment and pimecrolimus cream, are effective alternatives to topical corticosteroids. These agents do not cause the skin atrophy associated with prolonged use of topical corticosteroids and, therefore, are useful for treating skin on the face and neck.

The macrolide antibiotic tacrolimus (formerly FK506) has been found to be effective in treating moderate to severe atopic dermatitis. The efficacy of tacrolimus has been shown in several randomized, controlled trials.[39-41] The most common adverse side effects are skin burning, flulike symptoms, skin erythema, and headache.[41] Topical tacrolimus is available in 0.1% and 0.03% concentrations. In children with moderate and severe AD, treatment with tacrolimus ointment (0.03%) was shown to be superior to conventional 1% hydrocortisone acetate.[42]

The ascomycin derivative pimecrolimus (ASM 981) cream is a cell-selective cytokine inhibitor that was specifically developed for treatment of inflammatory skin diseases. Its mechanism of action is similar to that of topical tacrolimus. Two independent randomized, multicenter studies found pimecrolimus to be effective in infants and children with AD.[43] Another randomized, multicenter study found that pimecrolimus was effective in preventing AD flares, which reduced the need for topical corticosteroids.[44] In adults, pimecrolimus was found to be effective and well tolerated, and it reduced the incidence of AD flare.[45]

A meta-analysis of 16 trials involving more than 5,300 patients showed success rates of tacrolimus and pimecrolimus to be statistically similar; however, tacrolimus success rates were numerically higher than those of pimecrolimus, and tacrolimus was used in patients with more severe disease.[46] The efficacy of these drugs must be balanced against a potential cancer risk. The Food and Drug Administration recently issued a warning that these drugs should be used only as directed and only after other eczema treatments have failed to work.[47]

Other Therapies

Antihistamines can sometimes be helpful in breaking the itch-scratch cycle in AD. Sedating antihistamines, such as hydroxyzine and diphenhydramine, are particularly useful—especially

when itching prevents sleep[48]; however, the sedative properties of antihistamines may limit their use in AD. Cetirizine, a sedating antihistamine, appears to be well tolerated in infants. A multinational, randomized, placebo-controlled trial examined the effects of long-term treatment with cetirizine on infants with AD; the drug proved to be safe, and it reduced the need for topical corticosteroids in patients with more severe disease.[49] Nonsedating antihistamines such as fexofenadine and loratadine are less useful. Doxepin, a tricyclic antidepressant known to have antihistaminic effects, can be beneficial when applied topically in a 5% cream.[50]

Virtually every phototherapy regimen has been reported to ameliorate AD. Some patients cannot tolerate the heat generated by the equipment, however—particularly that used in UVB irradiation. In addition to UVB, the following can be beneficial: UVA, longwave UVA1, narrow-band UVB, UVA-UVB, and PUVA. Extracorporeal photochemotherapy (photopheresis) is reported to be effective therapy for recalcitrant disease.[51] Phototherapies are expensive, and prolonged use of PUVA has been linked to an increased risk of melanoma.[52] Although some patients may benefit from natural sunlight, the risk of sunburn and induction of malignancy by ultraviolet light must be considered.

Antimicrobials are obviously important for patients with infection. Less clear is whether antimicrobial agents can directly treat AD by reducing bacterial products thought to exacerbate the condition. Antistaphylococcal therapy has been advocated for use in patients with AD; however, a double-blind, placebo-controlled study of flucloxacillin did not show improvement in AD despite reduced bacterial counts.[53] Ketoconazole, likewise, has been used; its success, however, may be the result of anti-inflammatory, rather than antifungal, effects.

More advanced therapeutic options exist for severe, recalcitrant AD. The altered expression of cytokines in AD [*see* Etiology and Pathogenesis, *above*] has led investigators to explore the use of interferon gamma. Clinical trials have demonstrated that for some patients, daily subcutaneous administration of interferon gamma is effective in reducing both signs and symptoms of AD[54,55] and that long-term treatment can maintain the benefit.[56] However, moderate results and high costs make interferon gamma less viable as a treatment option.

Oral cyclosporine (2.5 to 5 mg/kg/day orally),[57,58] methotrexate (15 to 25 mg/wk orally), and azathioprine (100 to 200 mg/day orally) can be used in severe,[59] recalcitrant disease provided that patients are monitored for adverse effects specific to those agents.

Traditional Chinese herbal medicine has been found to be effective in the treatment of AD, both in children[60] and in adults,[61] although the efficacy of this treatment remains controversial.[62] The mechanisms of action of these preparations are unclear. A small, randomized, placebo-controlled study found topical treatment with St. John's wort to be significantly superior to placebo in patients with moderate AD.[63] Although evening primrose oil has for many years been proposed to be effective in AD, a well-controlled study failed to show any benefit to patients taking either evening primrose oil or a combination of evening primrose oil and fish oil, as compared with those receiving placebo.[64] Patients should be cautioned that herbal remedies are not risk free and may carry a potential for hepatotoxicity, cardiomyopathy, and other adverse effects; such remedies should be monitored, as should any other treatment. To avoid potential adverse drug reactions, physicians should identify any herbal remedies used by patients.[65]

Topical vitamin B_{12} was found to be significantly superior to placebo in reducing the extent and severity of AD in a randomized, multicenter phase III study[66]; however, larger trials are needed to establish the efficacy of this therapy. The cAMP phosphodiesterase inhibitor cipamfylline in cream form has been shown to be more effective than placebo but significantly less effective than hydrocortisone cream in the treatment of AD.[67] The importance of well-controlled studies to assess efficacy of treatments must be stressed because AD patients on the placebo arms of most controlled studies tend to show benefit, sometimes marked.

Ichthyoses

The ichthyoses are a group of diseases of cornification that are characterized by excessive scaling.[68] Etiologies of the ichthyoses are diverse, including genetic defects of structural proteins and enzymes, as well as acquired forms. Only the major clinical variants will be discussed here.

MAJOR VARIANTS

Ichthyosis Vulgaris

Ichthyosis vulgaris, the most common form of ichthyosis, is found in approximately one in 300 births. This autosomal dominant condition presents as dry skin with fine scaling. The extensor surfaces of extremities are the most commonly affected areas. Ichthyosis vulgaris can occur concomitantly with keratosis pilaris and can also be associated with AD. Age at onset is typically between 3 months and 12 months. Implicated etiologic factors include reduced filaggrin (filament-aggregating protein) and its precursor profilaggrin, whose normal functions are to allow for aggregation of keratin filaments and to serve as sources of compounds that hydrate the skin. The clinical severity of ichthyosis vulgaris correlates with the degree of reduction in filaggrin and profilaggrin. Another possible etiologic factor is the reduced activity of proteases that normally lead to dissociation of keratinocytes.[69]

X-Linked Ichthyosis

Recessive X-linked ichthyosis occurs in approximately one in 2,000 to one in 6,000 male infants. Although collodion membrane may be present at birth, the skin is usually normal, with fine scaling beginning at 1 to 3 weeks of life. Typically, the scales are thick and dark, giving the skin a dirty appearance. Extensor distribution—combined with involvement of the sides of the neck and preauricular skin and sparing the flexural areas—is typical. Steroid sulfatase deficiency is an etiologic factor, causing an increase in cholesterol sulfate and a decrease in cholesterol in the stratum corneum.[70] The accumulated cholesterol sulfate may inhibit proteolysis—a process similar to the inhibition seen in ichthyosis vulgaris. Prenatal diagnosis is available, and gene therapy may be on the horizon.

Lamellar Ichthyosis

Lamellar ichthyosis occurs in one in 300,000 births. It is inherited in an autosomal recessive pattern. Collodion membrane may be present at birth but is then shed, revealing characteristic large, platelike scales. Erythroderma may be present, albeit difficult to discern because of the thickness of the scales. Ectropion is present in most patients and can give rise to ophthalmic complications. Lamellar ichthyosis is often caused by mutations in the gene encoding the enzyme transglutaminase 1.[71]

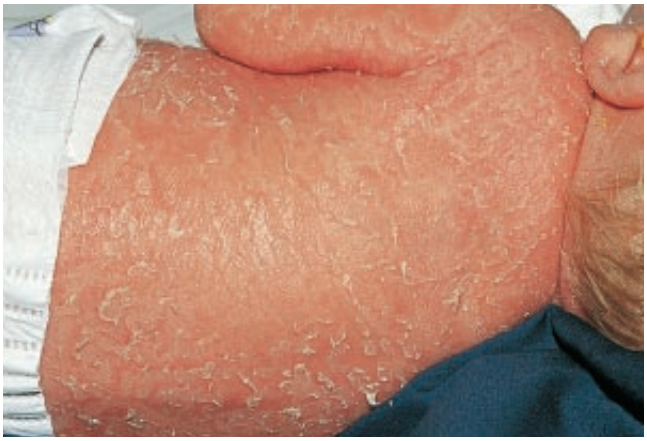


Figure 3 **Erythroderma (total body erythema) and extensive scaling are seen in this infant with congenital ichthyosiform erythroderma.**

Congenital Ichthyosiform Erythroderma

Formerly, congenital ichthyosiform erythroderma [*see Figure 3*] was considered to be a variant of lamellar ichthyosis. Both are inherited as autosomal recessive traits, and collodion membrane may be present at birth in both conditions. Ectropion, eclabion (eversion of the lip), and erythroderma can also occur. Like patients with lamellar ichthyosis, patients with congenital ichthyosiform erythroderma may have platelike scales on the lower extremities, but scales are fine and white on other parts of the body. Also in contrast to lamellar ichthyosis, X-linked ichthyosis, and ichthyosis vulgaris—whose lesions are scaly because of an abnormal ability to desquamate (so-called retention hyperkeratoses)—the lesions of congenital ichthyosiform erythroderma are scaly because of increased production of keratinocytes (so-called hyperproliferative ichthyosis).

Epidermolytic Hyperkeratosis

Epidermolytic hyperkeratosis (formerly called bullous congenital ichthyosiform erythroderma) is autosomal dominant in inheritance. The combinations of large blisters and erythema with denuded skin that appear at birth may be confused with epidermolysis bullosa, staphylococcal scalded skin syndrome, or toxic epidermal necrolysis. Several months to 1 year after birth, the blisters become less prominent, and thick, verrucous plaques comprising rows of hyperkeratotic ridges develop. Flexural skin is usually involved, but the disease can be more extensive. Bacterial colonization leads to a clinically significant foul odor. Abnormal keratin gene expression is the etiologic basis of this condition.[71]

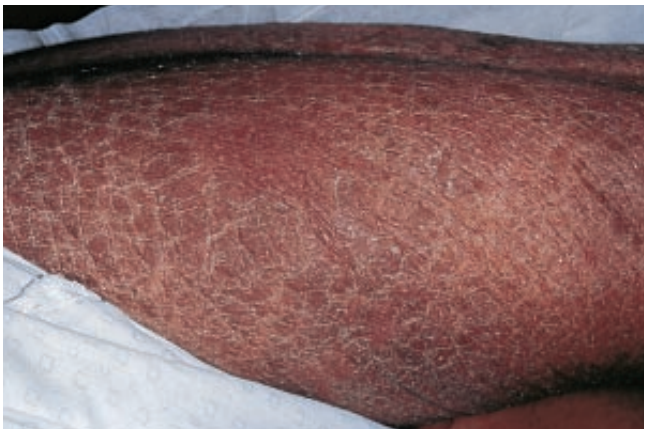


Figure 4 **This patient developed marked scaling (acquired ichthyosis) over a 6-month period. Investigation revealed non-Hodgkin lymphoma.**

Acquired Ichthyosis

Acquired ichthyoses have been associated with numerous systemic diseases and medications. Although the onset of scaling is commonly a manifestation of dryness or ichthyosis vulgaris, patients with unusual manifestations or with severe or recalcitrant disease warrant further investigation. Endocrinopathies (e.g., thyroid disease), autoimmune diseases, infectious diseases (e.g., HIV), and malignancies such as lymphomas [*see Figure 4*] and other carcinomas have been associated with the onset of ichthyosiform dermatosis.

TREATMENT

The standard therapy for the ichthyoses is emollients (e.g., petrolatum) and keratolytics (e.g., lactic acid with or without propylene glycol).[72] Lactic acid should be used cautiously in neonates to avoid causing excess absorption. Oral retinoids (which require lipid monitoring) can be helpful, particularly in the management of X-linked ichthyosis, congenital ichthyosiform erythroderma, and lamellar ichthyosis. Epidermolytic hyperkeratosis is the most difficult of these conditions to treat because of the risk of blistering induced by therapeutic agents. Antimicrobial agents can be useful to reduce the odor caused by bacterial colonization.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Marks JG Jr, Belsito DV, DeLeo VA, et al: North American Contact Dermatitis patch-test results, 1998-2000. Am J Contact Dermat 14:59, 2003

2. Denig NI, Hoke AW, Maibach HI: Irritant contact dermatitis: clues to causes, clinical characteristics, and control. Postgrad Med 103:199, 1998

3. Coca AF, Cooke RA: On the classification of the phenomena of hypersensitiveness. J Immunol 8:163, 1922

4. Gustafsson D, Sjoberg O, Foucard T: Development of allergies and asthma in infants and young people with atopic dermatitis: a prospective follow-up to 7 years of age. Allergy 55:240, 2000

5. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. Lancet 351:1225, 1998

6. Laughter D, Istvan JA, Tofte SJ, et al: The prevalence of atopic dermatitis in Oregon schoolchildren. J Am Acad Dermatol 43:649, 2000

7. Foley P, Zuo Y, Plunkett A, et al: The frequency of common skin conditions in preschool-age children in Australia: atopic dermatitis. Arch Dermatol 137:293, 2001

8. Levy RM, Gelfand JM, Yan AC: The epidemiology of atopic dermatitis. Clin Dermatol 21:109, 2003

9. Uehara M, Kimura C: Descendant family history of atopic dermatitis. Acta Derm Venereol 73:62, 1993

10. Agosti JM, Sprenger JD, Lum LG, et al: Transfer of allergen-specific IgE-mediated hypersensitivity with allogeneic bone marrow transplantation. N Engl J Med 319:1623, 1998

11. Haagerup A, Bjerke T, Schiotz PO, et al: Atopic dermatitis: a total genome-scan for susceptibility genes. Acta Derm Venereol 84:346, 2004

12. Leung DY, Boguniewicz M, Howell MD, et al: New insights into atopic dermatitis. J Clin Invest 113:651, 2004

13. Bos JD, Wierenga EA, Sillevis Smitt JH, et al: Immune dysregulation in atopic eczema. Arch Dermatol 128:1509, 1992

14. Hanifin JM, Rajka G: Diagnostic features of atopic dermatitis. Acta Derm Venereol Suppl (Stockh) 92:44, 1980

15. Saarinen KM, Juntunen-Backman K, Jarvenpaa AL, et al: Breast-feeding and the development of cows' milk protein allergy. Adv Exp Med Biol 478:121, 2000

16. Gustafsson D, Lowhagen T, Andersson K: Risk of developing atopic disease after

early feeding with cows' milk based formula. Arch Dis Child 67:1008, 1992

17. Halken S: Prevention of allergic disease in childhood: clinical and epidemiological aspects of primary and secondary allergy prevention. Pediatr Allergy Immunol 15(suppl 16):4, 2004

18. Kalliomaki M, Salminen S, Poussa T, et al: Probiotics and prevention of atopic disease: 4-year follow-up of randomized placebo-controlled trial. Lancet 361:1869, 2003

19. Zoller TM, Wichelhaus TA, Hartung A, et al: Colonization with superantigens producing *Staphylococcus aureus* is associated with increased severity of atopic dermatitis. Clin Exp Allergy 30:994, 2000

20. Herz U, Bunikowski R, Renz H: Role of T cells in atopic dermatitis: new aspects on the dynamics of cytokine production and the contribution of bacterial superantigens. Int Arch Allergy Immunol 115:170, 1998

21. Rees J, Friedmann PS, Matthews JN: Contact sensitivity to dinitrochlorobenzene is impaired in atopic subjects. Arch Dermatol 126:1173, 1990

22. Akdis CA, Akdis M, Trautmann A, et al: Immune regulation in atopic dermatitis. Curr Opin Immunol 12:641, 2000

23. Grewe M, Bruijnzeel-Koomen CAFM, Schöpf E, et al: A role for Th1 and Th2 cells in the immunopathogenesis of atopic dermatitis. Immunol Today 19:359, 1998

24. Taylor RS, Baadsgaard O, Hammerberg C, et al: Hyperstimulatory CD1a+CD1b+CD36+ Langerhans cells are responsible for increased autologous T lymphocyte reactivity to lesional epidermal cells of patients with atopic dermatitis. J Immunol 147:3794, 1991

25. Stingl G, Maurer D: IgE-mediated allergen presentation via Fc epsilon RI on antigen-presenting cells. Int Arch Allergy Immunol 113:24, 1997

26. Hanifin JM, Chan SC: Monocyte phosphodiesterase abnormalities and dysregulation of lymphocyte function in atopic dermatitis. J Invest Dermatol 105(1 suppl):84S, 1995

27. Heskel NS, Chan SC, Thiel ML, et al: Elevated umbilical cord blood leukocyte cyclic adenosine monophosphate-phosphodiesterase activity in children with atopic parents. J Am Acad Dermatol 11:422, 1984

28. Uehara M, Izukura R, Sawai T: Blood eosinophilia in atopic dermatitis. Clin Exp Dermatol 15:264, 1990

29. Bannister MJ, Freeman S: Adult-onset atopic dermatitis. Australas J Dermatol 41:225, 2000

30. Roll A, Cozzio A, Fisher B, et al: Microbial colonization and atopic dermatitis. Curr Opin Allergy Clin Immunol 4:373, 2004

31. Hofer MF, Lester MR, Schlievert PM, et al: Upregulation of IgE synthesis by staphylococcal toxic shock syndrome toxin-1 in peripheral blood mononuclear cells from patients with atopic dermatitis. Clin Exp Allergy 25:1218, 1995

32. Nilsson EJ, Henning CG, Magnusson J: Topical corticosteroids and *Staphylococcus aureus* in atopic dermatitis. J Am Acad Dermatol 27:29, 1992

33. Rich LF, Hanifin JM: Ocular complications of atopic dermatitis and other eczemas. Int Ophthalmol Clin 25:61, 1985

34. Fletcher CL, Orchard GE, Hubbard V, et al: CD30(+) cutaneous lymphoma in association with atopic eczema. Arch Dermatol 140:449, 2004

35. Ellis C, Luger T, Abeck D, et al: International consensus conference on atopic dermatitis II (ICCAD II): clinical update and current treatment strategies. Br J Dermatol 148(suppl 63):3, 2003

36. McHenry PM, Williams HC, Bingham EA: Management of atopic eczema. Joint Workshop of the British Association of Dermatologists and the Research Unit of the Royal College of Physicians of London. BMJ 310:843, 1995

37. Hanifin JM, Cooper KD, Ho VC, et al: Guidelines of care for atopic dermatitis, developed in accordance with the American Academy of Dermatology (AAD)/American Academy of Dermatology Association "Administrative Regulations for Evidence-Based Clinical Practice Guidelines." J Am Acad Dermatol 50:391, 2004

38. Chamlin SL, Fao J, Frieden IJ, et al: Ceramide-dominant barrier repair lipids alleviate childhood atopic dermatitis: changes in barrier function provide a sensitive indicator of disease activity. J Am Acad Dermatol 47:198, 2002

39. Paller A, Eichenfield LF, Leung DY, et al: A 12-week study of tacrolimus ointment for the treatment of atopic dermatitis in pediatric patients. J Am Acad Dermatol 44(suppl 1):S47, 2001

40. Hanifin JM, Ling MR, Langley R, et al: Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: part I, efficacy. J Am Acad Dermatol 44(suppl 1):S28, 2001

41. Soter NA, Fleisher AB Jr, Webster GF, et al: Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: part II, safety. J Am Acad Dermatol 44(suppl 1):S39, 2001

42. Reitamo S, Harper J, Bos JD, et al: 0.03% Tacrolimus ointment applied once or twice daily is more efficacious than 1% hydrocortisone acetate in children with moderate to severe atopic dermatitis: results of a randomized double-blind controlled trial. Br J Dermatol 150:554, 2004

43. Eichenfield LF, Lucky AW, Boguniewicz M, et al: Safety and efficacy of pimecrolimus (ASM 981) cream 1% in the treatment of mild and moderate atopic dermatitis in children and adolescents. J Am Acad Dermatol 46:495, 2002

44. Wahn U, Bos JD, Goodfield M, et al: Efficacy and safety of pimecrolimus cream in the long-term management of atopic dermatitis in children. Pediatrics 110:e2, 2002

45. Meurer M, Fartasch M, Albrecht G, et al: Long-term efficacy and safety of pimecrolimus cream 1% in adults with moderate atopic dermatitis. Dermatology 208:365, 2004

46. Iskedjian M, Piwko C, Shear NH, et al: Topical calcineurin inhibitors in the treatment of atopic dermatitis: a meta-analysis of current evidence. Am J Clin Dermatol 5:267, 2004

47. FDA issues public health advisory informing health care providers of safety concerns associated with the user of two eczema drugs, Elidel and Protopic. FDA Talk Paper. U.S. Food and Drug Administration. March 10, 2005
http://www.fda.gov/bbs/topics/answers/2005/ANS01343.html

48. Klein PA, Clark RA: An evidence-based review of the efficacy of antihistamines in relieving pruritis in atopic dermatitis. Arch Dermatol 135:1522, 1999

49. Diepgen TL: Long-term treatment with cetirizine of infants with atopic dermatitis: a multi-country, double-blind, randomized, placebo-controlled trial (the ETAC trial) over 18 months. Early Treatment of the Atopic Child Study Group. Pediatr Allergy Immunol 13:278, 2002

50. Groene D, Martus P, Heyer G: Doxepin affects acetylcholine induced cutaneous reactions in atopic eczema. Exp Dermatol 10:110, 2001

51. Richter HI, Billmann-Eberwein C, Grewe M, et al: Successful monotherapy of severe and intractable atopic dermatitis by photopheresis. J Am Acad Dermatol 38:585, 1998

52. Stern RS: The risk of melanoma in association with long-term exposure to PUVA. J Am Acad Dermatol 44:755, 2001

53. Ewing CI, Ashcroft C, Gibbs AC, et al: Flucloxacillin in the treatment of atopic dermatitis. Br J Dermatol 138:1022, 1998

54. Hanifin JM, Schneider LC, Leung DY, et al: Recombinant interferon gamma therapy for atopic dermatitis. J Am Acad Dermatol 28:189, 1993

55. Ellis CN, Stevens SR, Blok BK, et al: Interferon-gamma therapy reduces blood leukocyte levels in patients with atopic dermatitis: correction with clinical improvement. Clin Immunol 92:49, 1999

56. Stevens SR, Hanifin JM, Hamilton T, et al: Long-term effectiveness and safety of recombinant human interferon gamma therapy for atopic dermatitis despite unchanged serum IgE levels. Arch Dermatol 134:799, 1998

57. Berth-Jones J, Graham-Brown RA, Marks R, et al: Long-term efficacy and safety of cyclosporin in severe adult atopic dermatitis. Br J Dermatol 136:76, 1997

58. Harper JI, Ahmed I, Barclay G, et al: Cyclosporin for severe childhood atopic dermatitis: short course versus continuous therapy. Br J Dermatol 142:52, 2000

59. Berth-Jones J, Takwale A, Tan E, et al: Azathioprine in severe adult atopic dermatitis: a double-blind, placebo-controlled, crossover trial. Br J Dermatol 147:324, 2002

60. Sheehan MP, Atherton DJ: A controlled trial of traditional Chinese medicinal plants in widespread non-exudative atopic eczema. Br J Dermatol 126:179, 1992

61. Sheehan MP, Atherton DJ: One-year follow up of children treated with Chinese medicinal herbs for atopic dermatitis. Br J Dermatol 130:488, 1994

62. Fung AY, Look PC, Chong LY, et al: A controlled trial of traditional Chinese herbal medicine in Chinese patients with recalcitrant atopic dermatitis. Int J Dermatol 38:387, 1999

63. Schempp CM, Windeck T, Hezel S, et al: Topical treatment of atopic dermatitis with St. John's wort cream: a randomized, placebo controlled, double blind half-side comparison. Phytomedicine 10(suppl 4):31, 2003

64. Berth Jones J, Graham Brown RA: Placebo-controlled trial of essential fatty acid supplementation in atopic dermatitis. Lancet 341:1557, 1993

65. Simpson EL, Basco M, Hanifin J: A cross-sectional survey of complementary and alternative medicine use in patients with atopic dermatitis. Am J Contact Dermat 14:144, 2003

66. Stucker M, Pieck C, Stoerb C, et al: Topical vitamin B_{12}: a new therapeutic approach in atopic dermatitis: evaluation of efficacy and tolerability in a randomized, placebo-controlled multicentre clinical trial. Br J Dermatol 150:977, 2004

67. Griffiths CE, Van Leent EJ, Gilbert M, et al: Randomized comparison of the type 4 phosphodiesterase inhibitor cipamfylline cream, cream vehicle and hydrocortisone 17-butyrate cream for the treatment of atopic dermatitis. Br J Dermatol 147:299, 2002

68. Shwayder T: Disorders of keratinization: diagnosis and management. Am J Clin Dermatol 5:17, 2004

69. Rabinowitz LG, Esterly NB: Atopic dermatitis and ichthyosis vulgaris. Pediatr Rev 15:220, 1994

70. Paller AS: Laboratory tests for ichthyosis. Dermatol Clin 12:99, 1994

71. DiGiovanna JJ, Robinson-Bostom L: Ichthyosis: etiology, diagnosis, and management. Am J Clin Dermatol 4:81, 2003

72. Fleckman P: Management of the ichthyoses. Skin Therapy Lett 8:3, 2003

V CONTACT DERMATITIS AND RELATED DISORDERS

James S. Taylor, m.d.

Contact dermatitis is an acute or chronic skin inflammation resulting from interaction with a chemical, biologic, or physical agent.[1] It is one of the most common conditions seen by physicians, accounting for more than 6.5 million physician visits a year and 95% of all reported occupational skin diseases.[2] Substances that produce contact dermatitis after a single exposure or multiple exposures may be irritant or allergenic. Direct tissue damage results from contact with irritants. Tissue damage by allergic substances is mediated through immunologic mechanisms. Eczema or dermatitis is the most common clinical expression of this induced inflammation. Of the more than 85,000 chemicals in our environment, most can be irritants, depending on the circumstances of exposure.[1] More than 3,700 substances have been identified as contact allergens.[3] The potential for these substances to cause contact dermatitis varies greatly, and the severity of the dermatitis ranges from a mild, short-lived condition to a severe, persistent, job-threatening, and possibly life-threatening disease.

Major Types of Contact Dermatitis

irritant contact dermatitis

Irritants cause as much as 80% of cases of contact dermatitis, act by direct nonimmunologic chemical or physical action on the skin, and are divided into marginal and acute types. Marginal irritants are the most common. Repeated daily exposures to low-grade irritants such as soap, detergents, surfactants, organic solvents, and oils may not cause clinical changes for days or months. Dryness of the skin with a glazed, parched appearance are often the initial signs; erythema, hyperkeratosis, and fissuring may supervene.

In contrast, acute irritants cause a more immediate reaction. Some irritants, such as strong acids and alkalis, aromatic amines, phosphorus, and metallic salts, produce a marked observable effect within minutes.[4,5] Others, such as hydrofluoric acid, ethylene oxide, podophyllin, and anthralin, produce a reaction within 8 to 24 hours after exposure.[4] Acute irritant contact dermatitis (ICD) is usually easily diagnosed by the patient history and often results from occupational accidents. The clinical appearance varies depending on the irritant and ranges from burns and deep-red ulcerations with sharp circumspection of the dermatitis, sometimes with a gravitational, dripping effect, to a vesicular dermatitis that is indistinguishable from acute allergic contact dermatitis.

Almost any substance can be an irritant, depending on the conditions of exposure [*see Figure 1*]. The nature of the irritant (i.e., its pH, solubility, physical state, and concentration), the duration of contact, and the nature of the vehicle affect disease severity. Host factors that predispose to ICD include preexisting dermatitis, skin dryness, sweating, and decreased thickness or breaks in the stratum corneum; environmental factors include high temperature, low humidity, friction, and pressure.

ICD provoked by work materials is believed to be a frequent cause of occupational skin disease. In one large population-based study, the highest annual incidence rates of ICD were reported for hairdressers (46.9 per 10,000 workers per year), bakers (23.5 per 10,000 workers per year), and pastry cooks (16.9 per 10,000 workers per year).[6] The causative factors of ICD are complex and usually involve exposure to a combination of irritants. The sentinel event for irritant hand eczema in hairdressers is dermatitis developing in moist areas that are difficult to rinse and dry, such as under rings and in the web spaces of the fingers.[7] Dermatitis may spread to the dorsum of the hand, where the skin is thinner and less resistant than on the palms.

No universally accepted test exists for diagnosing ICD, which is often diagnosed by excluding allergic contact dermatitis (ACD). Because of the clinical similarities between allergic and irritant contact dermatitis, it is important that patients who are thought to have either disorder undergo patch testing, the results of which are positive with ACD and negative with ICD.

ICD may become chronic if it is not treated early [*see* Treatment of Irritant and Allergic Contact Dermatitis, *below*]. Even when the skin appears to be healed, its protective capacity remains impaired for weeks or months. Additionally, ICD impairs the barrier function of the skin, allowing penetration of potential contact allergens. Individuals who had childhood atopic eczema are more likely than others to develop ICD of the hands when their jobs involve wet work.

allergic contact dermatitis

Allergic contact dermatitis is a type 4, T cell–mediated, delayed hypersensitivity reaction in the skin. The disorder affects

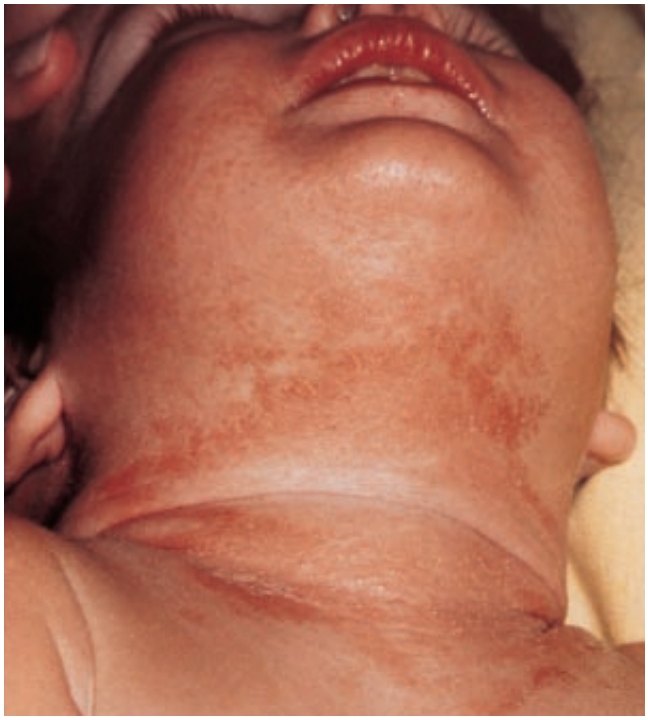


Figure 1 **Wearing a plastic bib resulted in irritant dermatitis in an 18-month-old child.**

Table 1 Body Sites Often Affected by 10 Common Contact Allergens

Allergen	*Common Uses*	*Localization Site*
Nickel	Costume jewelry	Earlobes, neck, fingers, wrists, abdomen
Neomycin	Topical antibiotics (dermatologic; ophthalmologic; ear, nose, throat)	Face, neck, trunk, extremities
Balsam of Peru	Fragrances, cosmetics, medications, flavorings	Face, trunk, extremities, perianal area
Fragrance mix	Toothpaste, fragrances, toiletries, cosmetics	Same as for balsam of Peru
Thimerosal	Topical antiseptic, contact lens solutions, eye cosmetics, nasal sprays	Eyelids, face, neck (relevance hard to prove)
Gold	Jewelry	Eyelids, earlobes, wrists, fingers
Formaldehyde	Cosmetics (preservative), shampoos, nail enamel	Eyelids, face, neck, trunk (especially intertriginous areas)
Quaternium-15	Cosmetics (preservative), shampoos, soaps, lotions	Face, trunk, extremities, hands
Cobalt	Metal-plated objects, jewelry	Earlobes, neck, fingers, wrists
Bacitracin	Topical antibiotics (dermatologic; ophthalmologic; ear, nose, throat)	Face, neck, trunk, extremities

only certain sensitized individuals, typically after two or more exposures, and accounts for about 20% of contact dermatitis cases.

Predisposing Factors

Immunologic status Predisposing factors to ACD include the patient's immunologic status, which in turn is influenced by genetics, age, gender, and the presence of systemic disease. Patients with AIDS, severe combined immunodeficiency, advanced lymphoma or other malignancy, sarcoidosis, lepromatous leprosy, cachexia, and atopic dermatitis may have impaired cell-mediated immunity or anergy.[8] However, contact allergy should not be excluded in these individuals, especially those with atopic eczema. In experimental models, agents that affect the immune system, such as ultraviolet light (ultraviolet B or psoralen and ultraviolet A [PUVA]), glucocorticoids, cyclosporine, and various other drugs, may downregulate ACD.[8] Administration of systemic corticosteroids below certain dosages (e.g., prednisone, 20 mg or less daily), however, does not inhibit strong patch-test reactions.[9]

In patients with occupational dermatitis, a form of natural hyporeactivity termed hardening may occur with diminished but continued exposure to chemical irritants. The process is inducible and is not localized.[10] This acquired state of unresponsiveness, when describing adaptation to allergens, is called tolerance.[8]

Environment The chemical environment in which we live defines opportunities for exposure to various allergens. A patient's age, gender, occupation, avocation, habits, and nationality are among the factors that determine the environment and thus the chemicals to which an individual is exposed. The most common source of contact allergy in the United States is *Toxicodendron*, a plant genus that includes poison ivy, poison oak, and poison sumac. In addition to *Toxicodendron*, 10 sources of contact allergens are commonly encountered in North America [*see Table 1*],[11] and numerous other allergens are known to cause contact reactions.[3]

Other cutaneous disorders Skin that is infected, inflamed, burned, or eczematous predisposes a patient to ACD. Patients with stasis, hand and foot eczema, or chronic actinic dermatitis are at high risk for ACD. ACD occasionally occurs with other skin disorders, including seborrheic dermatitis, psoriasis, prurigo nodularis, and benign familial pemphigus (Hailey-Hailey disease).[12] Noneczematous contact reactions have also been reported: purpuric reactions caused by black rubber; lichen planus–like eruptions caused by color-film developers, gold, and other dental metals (oral mucosa); and granulomas caused by beryllium and zirconium.[12]

Pathogenesis

Some inflammatory immune reactions in ACD are the same as those in ICD—specifically, the two disorders have similar cytokine activity (tumor necrosis factor–α and interferon gamma) and accessory molecule activity (HLA-DR and intercellular adhesion molecule–1) producing the cascade of inflammation. However, there is no memory T cell function in ICD,[13] and the extent of reaction is directly related to the amount of irritant and duration of exposure.[14] In contrast, even small amounts of an allergen can trigger the T cell reaction in ACD. Minor variations in an allergen's physical and chemical properties may affect its ability to induce sensitization.[8] Most environmental allergens are haptens–that is, they are small (< 500 daltons) molecules that penetrate the skin and undergo in vivo conjugation with tissue, or carrier, protein. Once the complex forms, the carrier protein is no longer recognized by the immune system as self. ACD represents a delayed-type hypersensitivity reaction to this complex.

During the sensitization phase, which usually takes a minimum of 5 to 21 days, an individual acquires a specific hypersensitivity to a particular contact allergen. Sensitization not only can evoke a type 4 delayed hypersensitivity response (mediated by lymphocytes) but also can produce a type 1 immediate hypersensitivity reaction (mediated by circulating antibodies).

On reexposure to an allergen, a hapten-carrier complex capable of eliciting a specific reaction re-forms. The reaction time—the time required for a previously sensitized individual to mani-

fest a clinical dermatitis after reexposure to the antigen—is usually 12 to 48 hours but may range from 8 to 120 hours.

A spontaneous flare may occur within 10 to 21 days without reexposure, possibly because enough allergen remains at the site to cause a reaction once the sensitization phase has occurred.

Cross-sensitization occurs when a patient who is allergic to one chemical also reacts to structurally related chemicals. Examples include *Toxicodendron* antigens (poison ivy, oak and sumac Japanese lacquer, mango, and cashew nutshell oil), aromatic amines (*p*-phenylenediamine, procaine, benzocaine, and *p*-aminobenzoic acid), and perfumes or flavors (balsam of Peru, benzoin, cinnamates, and vanilla). This phenomenon may explain persistence or reactivation of dermatitis when such exposures are unknown.[8,12,15]

Diagnosis

Diagnosis of ACD is based on the patient history; on the appearance, periodicity, and localization of the eruption; and on the clinical course. The history is especially important in cases of chronic dermatitis and putative occupational contact dermatitis. The history alone may be accurate only 50% of the time, on average, ranging from 80% accuracy for nickel to 50% accuracy for moderately common allergens to about 10% accuracy for less common allergens. Even with causes that are considered obvious, the specific allergen may not be known, and ACD that is caused by other chemicals may also be present. Skillful history taking is required to differentiate ACD from contact urticaria and ICD, with differentiation being especially difficult in chronic cases [*see Table 2*].[16] Also important is detailed questioning of the patient about all topical medications (over-the-counter and prescription), systemic medication, cosmetics, other lotions and creams, occupation, hobbies, travel, and clothing. A history of hypersensitivity to one or more of the major contact allergens (e.g., nickel, rubber, topical medications, and cosmetics [fragrances, preservatives, and dyes]) or obvious occupational or avocational exposures to certain substances or chemicals (e.g., chrome, epoxy, acrylics, latex gloves, clothing, first-aid creams, preservatives, and plants) may point to the diagnosis of ACD in an otherwise unexplained eruption.[16]

Clinical features In the acute stage, papules, oozing vesicles, and crusting lesions that are surrounded by inflammation predominate. These clinical features may occur anywhere, but they are best visualized on the palms, sides of the fingers, periungual areas, and soles of the feet. Frequently occurring or persistent episodes of ACD often become chronic; lesions associated with chronic ACD may appear thickened and exhibit lichenification, scaling, and fissuring [*see Figures 2 and 3*]. Post-inflammatory hyperpigmentation or hypopigmentation may occur. In the subacute stage of ACD, features characteristic of both acute and chronic ACD may be present. All forms of contact dermatitis frequently cause pruritus. The onset of ACD is often subtle. A low-grade, subacute to chronic eczema may appear as primarily a scaly or chapped eruption, especially on the face or on the dorsa of the hands.[12,16]

The distribution of dermatitis is often the single most important clue to the diagnosis of ACD. The area of most intense dermatitis usually corresponds to the site of most intense contact with the allergen. Exceptions occur, such as nail-polish allergy, which typically appears on ectopic sites, especially the eyelids, face, and neck. In addition to the transfer of allergens to distant sites, volatile airborne chemicals may cause dermatitis on exposed body areas. Regional differences in susceptibility to contact allergens exist. Thinner eyelid and genital skin is more susceptible to both allergic and irritant contact dermatitis. Because head hair is often protective of the scalp, allergic reactions to hair cosmetics frequently involve the upper face, eyelids, postauricular area, and neck. Other areas of the body have higher or lower

Table 2 Common Misconceptions about ACD[49]

Fallacy	*Truth*
Rash quickly follows contact	Rash is often delayed 1 to 2 days and may not appear for 1 wk after contact
Allergy develops only to new substances	Allergy can develop years after contact; an induction period may last virtually a lifetime
Allergy is dose-dependent	Allergy is not, within a wide range, dose-dependent
If changes in medications or cosmetics do not lead to clearing of the rash, those products are not the cause	Many products contain the same or cross-reacting allergens; also, the composition of the product may be altered without a change in the trade name of the product
Contact allergy occurs only at the site of exposure to the offending agent	Contact allergy can spread by direct or indirect contact, airborne exposure, connubial contact, or autoeczematization
Expensive products are not allergenic	Allergy is not related to cost
Negative prick or scratch test or RAST excludes ACD	Only patch testing is diagnostic of ACD
ACD is always bilateral if allergen exposure is bilateral	Shoe and glove allergy are often bilateral but may be unilateral
ACD is of the same intensity at all areas of exposure	Body sites may differ in responsiveness to allergens; ACD may be patchy (e.g., hand dermatitis from gloves)
ACD does not affect the palms and soles	ACD may occur on the palms and soles (e.g., from gloves, topical medicaments, shoes)

ACD—allergic contact dermatitis RAST—radioallergosorbent test

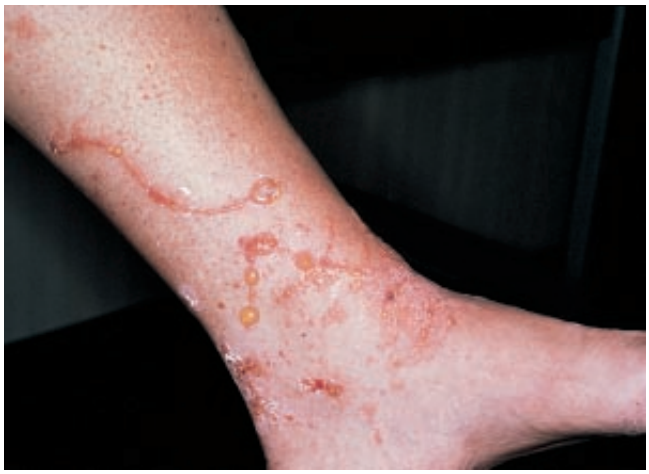


Figure 2 **Exposure to poison oak produced this acute *Toxicodendron* dermatitis with erythema, edema, and linear vesicles and bullae.**

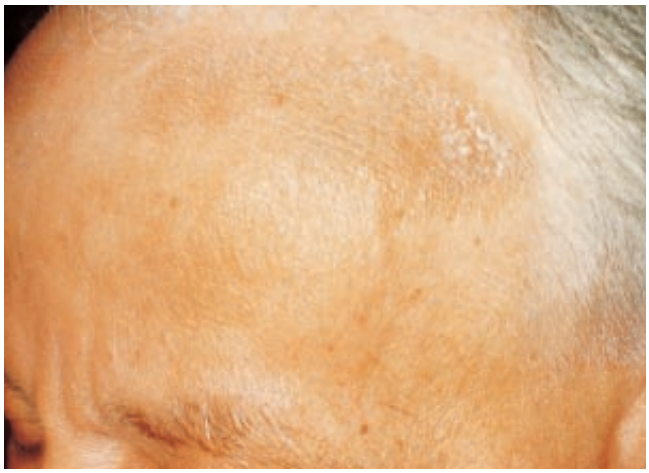


Figure 3 **Chronic eczematous dermatitis, with scaling, lichenification, and hyperpigmentation, was caused by an allergy to leather components in a hatband.**

exposures to various allergens; these exposures are not always clear and are reflected in unusual distributions of dermatitis. Allergens in lotions and creams applied all over the body sometimes produce reactions in skin folds and intertriginous areas, where the chemicals tend to concentrate. Recognition of ACD on the basis of the physical examination alone may be only partially accurate. Linear vesicular streaks are commonly seen in poison ivy, poison oak, and poison sumac dermatitis, but contact with other plants can give a similar picture. Contact with liquids may also produce linear vesicles. Failure to examine the entire skin surface may result in misdiagnosis. Eczema on the trunk and arms may in fact represent autoeczematization from contact or stasis dermatitis of the legs. Significant regional variations are associated with contact dermatitis, and knowledge of substances that cause dermatitis of specific body sites facilitates the diagnosis. Three such areas are the hands, face and neck, and feet [*see Figures 4 through 7*].

If the history and clinical presentation reveal one or more risk factors for ACD, a patch test is indicated [*see Table 3* and Patch Test, *below*]. The differential diagnosis of ACD is extensive, and a list of key points can be useful in establishing an accurate diagnosis [*see Table 4*].

Histopathology Biopsies are of limited help in diagnosing contact dermatitis. Microscopic findings vary according to the stage of the process: acute, subacute, or chronic. The hallmark of eczema is spongiosis, or intercellular edema, associated with spongiotic vesicles. Intracellular edema may cause reticular degeneration of the epidermis with multilocular bullae formation. Most types of eczema show similar pathologic changes and cannot be distinguished with certainty.[17]

Patch test The patch test is the only useful and reliable method—the gold standard—for the diagnosis of ACD. The proper performance and interpretation of this bioassay require

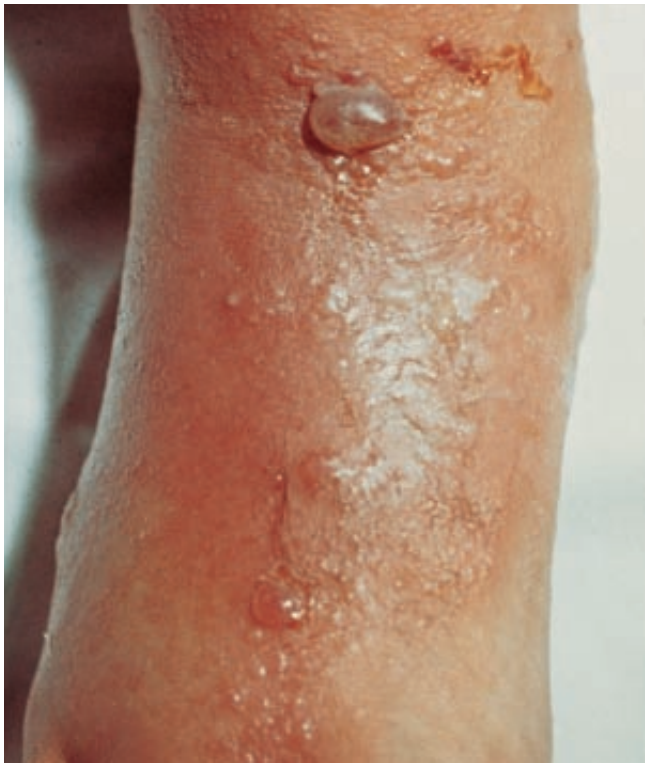


Figure 4 **Acute contact dermatitis caused by wearing sandals typically involves the dorsal surface of the feet.**

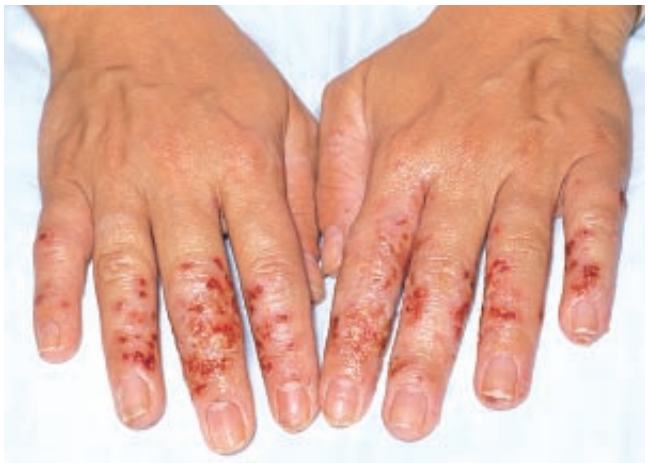


Figure 5 **Hairdresser with acute allergic contact dermatitis of the hands, caused by glyceryl thioglycolate.**

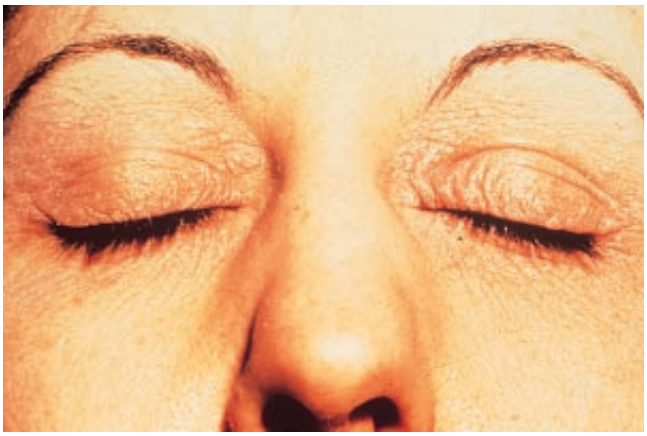


Figure 6 **Ectopic allergic contact dermatitis of the eyelids from tosylamide formaldehyde resin in nail polish.**

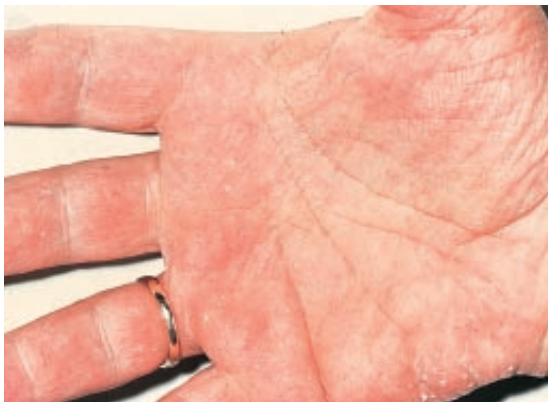
considerable experience. Because the procedure is subject to patient variability and observer error, the technique has been standardized by the North American Contact Dermatitis Group. First, the allergen is diluted in petrolatum or water to a concentration that does not produce active sensitization or irritation. A widely used patch-test system consists of strips of paper tape, onto which are fixed aluminum disks 8 mm in diameter (Finn Chambers on Scanpor tape). A small amount of allergen is placed within these disks, covering slightly more than one half of its diameter [*see Figure 8*]. Currently, the only commercially available patch-test system in the United States is the thin-layer rapid-use epicutaneous (T.R.U.E.) test. The T.R.U.E. test contains 23 preloaded allergens that are crystallized, micronized, or emulsified into gels that are affixed to paper tape.

With both systems, the tests are applied to the upper back or midback, which must be free of dermatitis. The patches are left in place and kept dry. When removed at 48 hours, the first reading is performed after 20 to 30 minutes, which allows time for pressure erythema to resolve. It is important to perform a second reading between 4 and 7 days after the patches are initially applied; otherwise, almost 20% of positive reactions will be missed. Neomycin, formaldehyde and formaldehyde-releasing preservatives, and tixocortol pivolate are often late reactors. Results at both readings are graded according to intensity of the reaction covering at least 50% of the patch-test site on a scale of 0 to 3+, as follows:

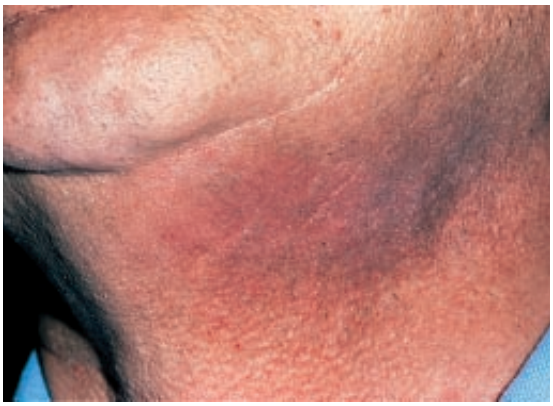
0 = no reaction
? (doubtful) = weak erythema only
1+ = erythema and edema
2+ = erythema, edema, and papules
3+ = vesicles or bullae

Both false positive and false negative reactions can result. Thus, patch testing is best done by physicians who are familiar with the intricacies of the procedure and who have been trained to advise patients about allergen substitution, relevance of the test, and prognosis. Reading test results and interpreting relevance are as important as performing the test. Any reaction must be evaluated with regard to the individual patient. Thus,

a



b



c

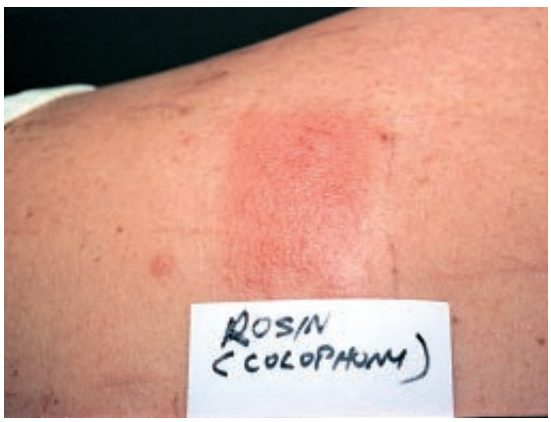


Figure 7 **Allergic contact dermatitis of the hands (*a*) and neck (*b*), with a positive patch test to rosin (colophony) (c).**

Table 3 Criteria* for Determining Which Patients with Putative ACD Should Be Patch Tested[49]

Presence of a specific type of eczema that places patient at higher risk for ACD (stasis, hand, foot, or chronic)
Patient is in a high-risk occupation
Health care worker
Cosmetologist (hairdresser)
Rubber compounder
Plastics processor
Chemical worker
Printer
Machinist
Woodworker
Specific allergen or substance is suspected
Patient has a highly suggestive history or distribution of dermatitis
Dermatitis flares or does not respond to treatment
Patient has previously undiagnosed dermatoses and erythroderma
Patient has putative occupational dermatitis
Special situation applies, such as photosensitivity or systemic contact dermatitis

*Test is ordered if any one of the risk factors is present.

when an allergen is found to be positive, it cannot always be assumed to be the cause of ACD.[8,12,15] The relevance of positive reactions to present or past episodes of ACD ranges from a low of 7.2% for thimerosal to 93.4% for dimethylol-dimethylhydantoin (DMDM hydantoin) and diazolidinyl urea [*see Table 5*]. Thus, relevance is determined by correlating the patch-test results with chemicals, products, and processes encountered in the environment. Occasionally, when patients are allergic to chemicals in products they use, the allergen may be present in only minimal amounts and may not be responsible for the dermatitis.[8] In these cases, repeat open application testing (ROAT), in which the patient applies the commercial product to normal skin twice daily for several days, can be helpful. ROAT is typically used with products that are left on rather than washed off after application.

In the United States, patch testing for ACD is often initially performed using the T.R.U.E. test; however, because there are over 3,700 environmental contact allergens and this test screens for only 23 allergens, testing with additional chemicals is imperative for a thorough evaluation. In one study, the T.R.U.E. test series of 23 allergens would have completely identified all allergens in only 25.5% of patients and clinically relevant allergens in 28% of patients.[18] Additional substances can be obtained from chemical suppliers and prepared by a compounding pharmacist in appropriate concentrations, as detailed in a standard text, for testing with the Finn Chamber system. As an alternative, many centers in the United States use individual patch-test chemicals or series (e.g., corticosteroid, plastics and glues, acrylic, dental, machinist, hairdresser) that are available in Europe but have not been approved in the United States.[11,19]

Reproducibility and validity of patch testing In a study in which 383 patients received simultaneous duplicate patch tests on opposite sides of the upper back, 8% of patients had completely discordant results: positive on one side of the back and negative on the other.[20] The intensity of the reactions was not disclosed, and clinical relevance of this problem was considered small. The most reproducible positive patch tests were for fragrance mix, nickel, and balsam of Peru. Formaldehyde and lanolin were the least reproducible positive reactors, both of which may be weak irritants.[20] The sensitivity, specificity, and validity of a standard screening series has been estimated at about 70%,[21] indicating that about 30% of these patch-test results were not valid. The patients whose screening results were negative later had positive results to other allergens. It was assumed that the earlier screening results had been false negative. A study of 500 consecutive patients who received identical patch testing reported discordant results in 5% of patients; the investigators concluded that patch testing is a reasonably reproducible procedure as long as methodological error is minimized.[22]

The positive predictive value of a diagnosis of ACD is a function of the prevalence of ACD in the population and a function of the sensitivity and specificity of the patch test.[23] A large dose-response study that tested the impact of seasonal variation on the irritant susceptibility of skin identified a stronger reaction to irritants in winter.[24]

Table 4 Key Points in the Diagnosis of ACD

Key Points	*Examples*
ACD may be identical to another disease	Tinea pedis misdiagnosed as ACD; a positive potassium hydroxide preparation made the diagnosis
	Psoriasis of the soles misdiagnosed as ACD caused by shoes; patch tests were negative
	Factitial eczema of the dorsal hand misdiagnosed as ACD; cured with an Unna Boot occlusive dressing
	ACD caused by fragrances and preservatives; misdiagnosed for 5 yr as lupus erythematosus
	ACD caused by hair tonic; misdiagnosed as seborrheic dermatitis
	ACD caused by sunscreen; misdiagnosed as sunburn
ACD may be concurrent with another disorder	ACD caused by neomycin; misdiagnosed as worsening atopic eczema
	Chronic actinic dermatitis of the face can be present with ACD caused by a fragrance
	Morphea of the leg can be present with ACD caused by a topical corticosteroid cream
ACD may be caused by an occult exposure to an allergen	Keys in pants pocket caused ACD of the lateral thigh in a man allergic to nickel
	ACD caused by the preservative imidazolidinyl urea, present in a sunscreen with a label that listed only the active ingredients
	Chronic hand eczema from ACD caused by red dye in window curtains
Diagnosis of ACD may be elusive because of inadequate or deceptive history	Patient allergic to neomycin had periorbital contact eczema caused by an ophthalmic ointment that contained tobramycin, which was not recognized as a cross-reacting allergen
	Chronic eczema worsened by use of a topical cream (doxepin) identified only from a pharmacy prescription list
Initial patch testing may not provide accurate diagnosis	Occupational contact dermatitis of the hands attributed to a false positive irritant-patch-test reaction to a cleanser
	Occupational contact dermatitis of the hands with a false negative patch-test reaction to latex surgical gloves; further patch testing indicated an allergy to thiurams, which were present as accelerators in the gloves

a

b

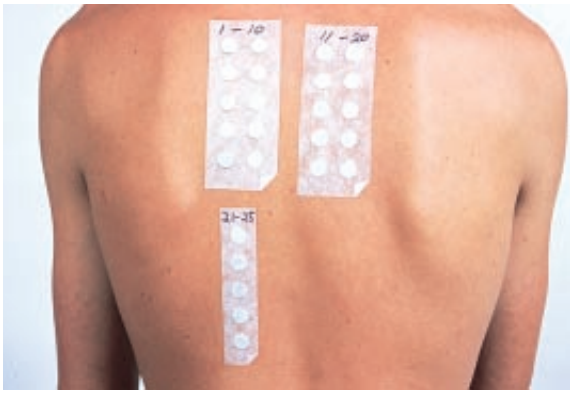


Figure 8 **Patch-test allergens to be tested, usually in petrolatum and occasionally aqueous, are placed on Finn Chambers on Scanpor tape (*a*) for application to the patient's back (*b*) for 48 hours. See patch testing in text and Figure 7c for a positive patch test.**

TREATMENT OF IRRITANT AND ALLERGIC CONTACT DERMATITIS

Most cases of contact dermatitis can be effectively treated and controlled once the offending irritant or allergen is identified and eliminated. Identifying hidden sources of allergens is important, and patients who have positive patch-test results are given exposure lists identifying various names of allergens, cross-reacting substances, lists of potential products and processes containing the allergen, and nonsensitizing substitutes. Standard texts should be consulted for detailed information[12,17]; the Internet is also a source of information on treatment of contact dermatitis [*see Sidebar* Internet Resources on Contact Dermatitis]. Examples of allergen alternatives include topical erythromycin or mupirocin ointments as substitutes for neomycin.[25] Neomycin may cross-react with gentamicin and tobramycin. Bacitracin should generally be avoided for neomycin-sensitive patients because of coreactivity.

Reasons for persistence of ACD include unidentified sources of allergens or irritants at home or at work, exposure to cross-reacting allergens, presence of underlying endogenous (e.g., atopic) eczema, and adverse reactions to therapy [*see* Topical-Medication Allergy, *below*].

Reduction of Trigger Factors

In the case of hand dermatitis, practical management must include protective measures, topical corticosteroids, and lubrication. The use of vinyl gloves with cotton liners to avoid the accumulation of moisture that often occurs during activities involving exposure to household or other irritants and foods (e.g., peeling or chopping fruits or vegetables) may be helpful. However, it is important to verify that gloves are safe to use in the workplace around machinery before recommending them. Protective devices themselves may introduce new allergic or irritant hazards in the form of rubber in gloves and solvents in waterless cleansers. Barrier creams are generally the last resort and are probably best for workers who do not have dermatitis.[26] Hand alcohol may be a superior disinfectant to soap and water in occupations that require extensive wet-work exposure.[27] A barrier agent containing quaternium-18 bentonite has been shown to be effective with exposure to a specific allergen, such as poison ivy.[28] Principles of treatment of atopic dermatitis may also be applied to treatment of contact dermatitis [*see 2:IV Eczematous Disorders, Atopic Dermatitis, and Ichthyoses*].

Topical Therapy

Treatment of contact dermatitis depends on the severity of the dermatitis. When acute serous oozing is present, cool, wet compresses should be applied for 15 minutes two or three times a day. Isotonic saline or Domeboro powder dissolved in tap water to make a 1:40 dilution (aluminum acetate) may be used. A soft cloth, such as Kerlex gauze or a towel, is immersed in the solution. The cloth is wrung slightly and applied to the affected area of the skin. The solution should not be poured directly on the dressing. Lukewarm to cool water baths or sitz baths are antipruritic and anti-inflammatory; they also aid in cleansing and removing crusts and medications. Oatmeal in the form of Aveeno Oilated Bath Treatment (colloidal powdered oatmeal with oils) may be added to the bath for its antipruritic and drying effects.

In acute vesicular dermatitis such as that caused by poison ivy, treatment with compresses and baths should be followed by the application of a topical corticosteroid spray (either triamcinolone acetonide [Kenalog aerosol] or betamethasone dipropionate [Diprosone aerosol]). A spray of 2 or 3 seconds' duration on each affected area supplies sufficient coverage, providing the container is held 6 in. from the skin. In cases in which the dermatitis is extensive or less vesicular, one of the many topical corticosteroid creams may be used. Corticosteroid creams range in potency from extremely potent (e.g., clobetasol propionate [Temovate]), to potent (e.g., betamethasone dipropionate [Diprosone topical cream]), to midstrength formulations. In addition, a lotion of camphor, menthol, and hydrocortisone (Sarnol-HC) is soothing, drying, and antipruritic. Pramoxine, a topical anesthetic in a lotion base (Prax), may also relieve pruritus.

In the subacute and chronic stages of contact dermatitis, an emollient lotion (Eucerin) or ointment (Aquaphor) may be applied to moist skin after bathing for lubrication. Oil-in-water emulsions that contain perfluoropolyethers have been shown to significantly inhibit ICD caused by a wide variety of hydrophilic and lipophilic irritants.[29] A potent or midstrength topical glucocorticosteroid cream or ointment is often used in the treatment of subacute and chronic contact dermatitis. Hydrocortisone 1% is only occasionally effective. Fluorinated corticosteroids should be used with discretion; frequent and prolonged use of these agents in skin-fold areas may cause atrophy, telangiectasia, or striae, and their use on the face may cause steroid rosacea. For patients with chronic dermatitis, crude coal tar preparations may be used to control eczema. Topical PUVA treatment may be effective for contact dermatitis of the palms and soles.[30]

Table 5 Patch-Test Results in North America from 2001 through 2002[11]

*Test Substance**	*T.R.U.E. Test Allergen*	*Use*	*Frequency of Positive Reactions (%)*	*Relevance of Patient (%)†*
Nickel sulfate 2.5%	TT	Metal	16.7	49.4
Neomycin sulfate 20%	TT	Antibiotic	11.6	32.3
Balsam of Peru (*Myroxylon pereirae*) 25%	TT	Fragrance	11.6	80.7
Fragrance mix 8%	TT	Fragrance	10.4	83.5
Thimerosal 0.1%	TT	Preservative	10.2	7.2
Gold sodium thiosulfate 0.5%		Metal	10.2	37.3
Quaternium-15 2%	TT	Preservative	9.3	84.3
Formaldehyde 1% aq	TT	Preservative	8.4	69.6
Bacitracin 20%		Antibiotic	7.9	42.6
Cobalt chloride 1%	TT	Metal	7.4	43.8
Methyldibromaglutaronitrile/phenoxyethanol 2.5%		Preservative	5.8	61.1
Carba mix 30%	TT	Rubber accelerator	4.9	76.6
p-Phenylenediamine 1%	TT	Hair dye	4.8	49.6
Thiuram 1%	TT	Rubber accelerator	4.5	78.9
Potassium dichromate 0.25%	TT	Metal	4.3	55.4
Benzalkonium chloride 0.1% aq		Preservative	4.3	26.9
Propylene glycol 30% aq		Medicine/cosmetic solvent	4.2	89.2
2-Bromo-2-nitropropane-1,3-diol 0.5%		Preservative	3.3	70.1
Diazolidinyl urea 1% aq		Preservative	3.2	91.1
Diazolidinyl urea 1%		Preservative	3.1	93.4
Imidazolidinyl urea 2%		Preservative	3.2	91.9
Tixocortol-21-pivalate 1%		Corticosteroid	3.0	86.9
Disperse blue 106 1%		Fabric dye	3.0	55.8
Ethylenediamine dihydrochloride 1%	TT	Medicine/cosmetic stabilizer	2.8	28.2
DMDM hydantoin 1%		Preservative	2.8	93.4
Cocamidopropyl betaine 1% aq		Cleanser/cosmetic solvent	2.8	89.2
Methyldibromoglutaronitrile/phenoxyethanol 4%		Preservative	2.7	70.9
Colophony (rosin) 20%	TT	Adhesive, etc.	2.6	46.1
Epoxy resin 1%	TT	Industrial coating/adhesive	2.3	60.5
Methylchloroisothiazolinone/methylisothiazolinone 100 ppm aq	TT	Preservative	2.9	83.3
Amidoamine 0.1% aq		By-product in manufacturing of cocamidopropyl betaine	2.3	83.2
Ethyleneurea melamine-formaldehyde resin 5%		Fabric-finish resin	2.3	67.6
Lanolin 30%	TT	Cosmetic emollient	2.2	82.1
DMDM hydantoin 1% aq		Preservative	2.2	88.2

Table 5 (*continued*)

*Test Substance**	*T.R.U.E. Test Allergen*	*Use*	*Frequency of Positive Reactions (%)*	*Relevance of Patient (%)†*
p-tert-Butylphenol formaldehyde resin 1%	TT	Adhesives	1.9	47.4
Glyceryl thioglycolate 1%		Permanent-wave chemical	1.9	39
Imidazolidinyl urea 2% aq		Preservative	1.8	90.8
Benzocaine 5%	TT	Anesthetic	1.7	39
Tosylamide formaldehyde resin 10%		Nail-polish resin	1.6	70.2
Methyl methacrylate 2%		Resin/adhesive	1.4	57.8
Glutaraldehyde 1%		Antibacterial	1.4	49.3
Ethyl acrylate 0.1%		Acrylic nails/resin	1.3	59.4
Cocamidopropyl betaine 0.5%		Cleanser/cosmetic solvent	1.3	74.6
DL α-Tocopherol		Vitamin E	1.1	75
Budesonide 0.1%		Corticosteroid	1.1	86.5
Dimethylol dihydroxyethylene urea 4.5%		Textile resin	1.1	61.1
Ylang ylang oil 2%		Fragrance	1.1	85.4
Black rubber mix 0.6%		Rubber accelerator	1.0	43.1
Compositae mix 6%		Plant group used in food and cosmetics	1.0	66.7
Mercaptobenzothiazole 1%	TT	Rubber accelerator	0.9	77.8
Dibucaine 2.5%	TT	Anesthetic	0.9	15.2
Thioureas 1%		Rubber accelerator	0.8	78.9
Jasmine Abs 2%		Fragrance	0.7	87.5
Mercapto mix 1%	TT	Rubber accelerator	0.7	81.8
Lidocaine 15%		Anesthetic	0.7	26.5
Paraben mix 1%	TT	Preservative	0.6	79.2
Sesquiterpene lactone mix 0.1%		Plant oleoresins	0.7	44.8
Benzophenone 3%		Sunscreen	0.6	79.3
p-Chloro-*m*-xylenol 1%		Antibacterial	0.6	71.4
Tetracaine 1%	TT	Anesthetic	0.6	21.5
Hydrocortisone-17-butyrate 1%		Corticosteroid	0.5	81.8
DL α-Tocopherol acetate		Vitamin E	0.5	72
Iodopropynyl butylcarbamate 0.1%		Preservative	0.3	61.5
Phenoxyethanol 1%		Preservative	0.2	63.6
Prilocaine 2.5%		Anesthetic	0.1	50

*Allergens in petrolatum unless noted aqueous (aq).
†Definite, probable, or possible reactions detected in percentage testing population.
TT—T.R.U.E. (thin-layer rapid-use epicutaneous) test

Systemic Therapy

Intense itching may be relieved with sedating antihistamines such as diphenhydramine hydrochloride (Benadryl), hydroxyzine hydrochloride (Atarax), and doxepin hydrochloride (Sinequan), administered at night. Most cases of ICD and ACD are effectively managed without the use of systemic corticosteroids. However, short courses of systemic corticosteroids are indicated for patients with severe vesiculobullous eruptions of the hands and feet or the face [*see Figure 9*] or with severe disseminated ACD, such as poison ivy. Strategies to reduce the side effects of corticosteroid use are especially important in patients who have diabetes, hypertension, glaucoma, latent or active tuberculosis (as indicated by a positive skin-test reaction to purified protein derivative), and diseases that could be affected by steroid therapy. Attempts at desensitization have generally been unsuccessful.[8] Secondary infection sometimes arises as a complication of ICD and ACD; in such cases, systemic antibiotics may be indicated.[29]

Specific Etiologic Forms of Contact Dermatitis

TOPICAL-MEDICATION ALLERGY

Reactions to topically applied medications include allergic and irritant contact dermatitis, photosensitivity, airborne contact dermatitis, and contact urticaria and anaphylaxis. ACD is the most common skin reaction to topically applied drugs. The three most important contact allergens are topical antibiotics, anesthetics, and antihistamines. Neomycin and bacitracin are among the most frequently prescribed medications and are common causes of ACD.[15] Mupirocin ointment infrequently causes ACD.[25] Benzocaine, the most common topical anesthetic allergen, is still widely used in topical agents, and there have been a number of reports of contact allergy to topical doxepin.[31,32]

ACD from topical corticosteroids is most often caused by the steroid itself rather than the vehicle. Studies indicate that in patients screened for contact dermatitis, the prevalence of allergy to one or more corticosteroids ranges from 0.55% to 5.98%.[33,34] Patch testing for allergy to the corticosteroid markers tixocortol pivolate and budesonide detects a great majority of cases of ACD caused by topical corticosteroids.[34] Further patch or ROAT testing using commercial preparations from the major cross-reacting classes may identify additional allergenic steroids or, alternatively, nonreacting steroids. Delayed readings are important at 5 to 7 days, because without a late reading, up to 30% of cases of contact allergy to corticosteroid markers can be missed.[34] Allergy to inhaled corticosteroids may present as perinasal or perioral itching or dermatitis, mimicking impetigo and herpes simplex or worsening asthma or allergic rhinitis. In such cases, prior sensitization by the cutaneous route is the usual occurrence, although allergy occasionally develops in response to corticosteroid inhalation.[35]

Topical-drug allergy is particularly common in patients with other forms of dermatitis, especially stasis dermatitis [*see Figure 10*]. In patients with stasis dermatitis, allergy to topical drugs often presents as a nonhealing dermatitis, which can mask the underlying cause of the eruption. A detailed history is important and should include the patient's use of nonprescription preparations, topical agents meant for animal use, medicated bandages, borrowed medications, transdermal devices, and herbal medicines. Patch testing with the standard screening tray and the patient's topical medications is invaluable in diagnosing ACD caused by topical medications.

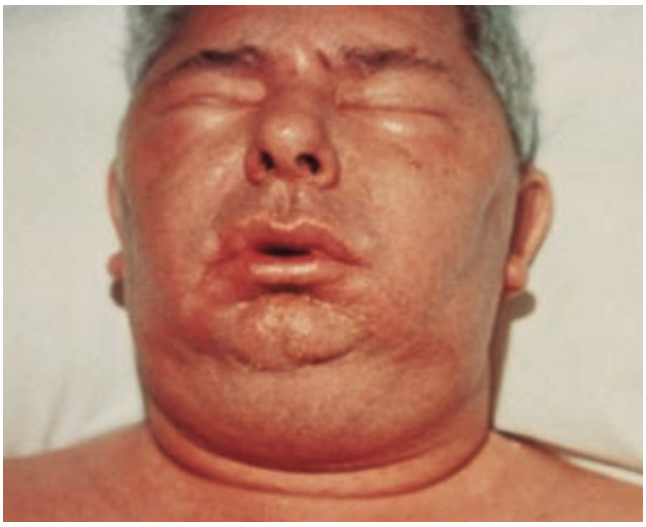


Figure 9 **For this patient with allergic contact dermatitis with marked facial edema, a short course of therapy with systemic corticosteroids is indicated.**

SYSTEMIC CONTACT DERMATITIS

Systemic contact dermatitis occurs in individuals with contact allergy to a hapten when they are exposed systemically to the hapten via the oral, subcutaneous, transcutaneous, intravenous, inhalational, intra-articular, or intravesicular route. The disorder has been caused by a number of medications, metals, and other allergens, including food components, but occurs infrequently compared with allergic and irritant contact dermatitis. Systemic contact dermatitis presents with the following clinically characteristic features[36,37]:

1. Flare-up of previous dermatitis or of prior positive-patch-test sites.
2. Skin disorders in previously unaffected skin, such as vesicular hand eczema, dermatitis in the elbow and knee flexures, nonspecific maculopapular eruption, vasculitis with palpable purpura, and the so-called baboon syndrome. This syndrome includes a pink–to–dark-violet eruption that is well demarcated on the buttocks and the genital area and is V-shaped on the inner thighs. It may occupy the whole area or only part of it.
3. General symptoms of headache, malaise, arthralgia, diarrhea, and vomiting.

Systemic contact dermatitis may start a few hours or 1 to 2 days after experimental provocation, suggesting that more than one type of immunologic reaction is involved. Documentation rests on patch testing and investigational oral-challenge studies. Well-controlled oral-challenge studies in sensitized individuals have been performed with medications but are more difficult to perform with ubiquitous contact allergens, such as metals and natural flavors. A relatively high dose of hapten is usually needed. Other variables include route of administration, bioavailability, individual sensitivity to the allergen, and interaction with amino acids and other allergens.[36] These variables can have dramatic effects on test results. For example, when 12 leg-ulcer patients with neomycin allergy were challenged with an oral dose of the hapten, 10 reacted.[37] However, of 29 patients with confirmed localized ACD caused by transdermal clonidine, only one had a skin reaction to oral clonidine.[38]

CLOTHING AND TEXTILE DERMATITIS

ACD from clothing is usually not caused by the fibers but rather by the dyes used to color the garments or by formalin finish resins added to make them wrinkle-resistant, shrinkproof, or wash-and-wear. Disperse blue dyes (especially blue 106 and blue 124) are highly valuable screening agents for diagnosing an important cause of textile dermatitis.[39] In a study in which 4,913 patients were patch tested using 65 allergens, disperse blue dye resulted in positive reactions in 3% of the study population [*see Table 5*].[11]

The distribution of dermatitis corresponds to areas where garments fit snugly, such as the upper and inner anterior thighs, popliteal fossae, buttocks, and waistband areas. Other areas include, in men, the parts of the neck that come in contact with stiff collars and, in women, the anterior or posterior axillary folds, vulva, and suprapubic area. Diagnosis is confirmed by patch testing with disperse dyes (especially blue 106 and blue 124) and formaldehyde-releasing fabric-finish resins (e.g., dimethyloldihydroxyethyleneurea and ethyleneurea melamine formaldehyde). Patch testing with the clothing (particularly acetate and polyester liners) of patients with dye allergy may yield positive results.[40]

Textile-dye dermatitis can be managed in the following ways[41]:

1. Avoiding clothes with the offending dye (especially 100% acetate or 100% polyester liners).
2. Avoiding nylon hose (especially beige tones) and tight synthetic spandex/Lycra exercise clothing.
3. Wearing 100% natural fabrics (i.e., cotton, linen, silk, wool) or 100% silk long-sleeved undershirts and slip pants.
4. Wearing loose-fitting clothing that has been washed (three times) before wearing.[42]

Many of these principles also apply to managing fabric-finish allergy, especially avoiding wrinkle-resistant, shrinkproof, and wash-and-wear clothing.

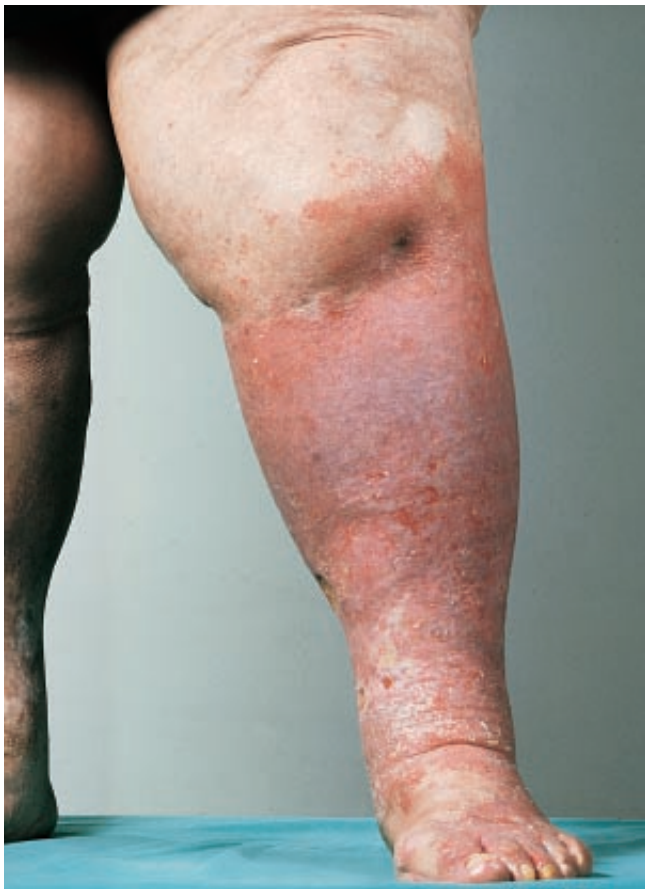


Figure 10 **Patients with stasis dermatitis are at high risk for allergic contact dermatitis, especially from topical medications. Bacitracin was the cause in this case.**

OCCUPATIONAL CONTACT DERMATITIS

Contact dermatitis, particularly of the hands, is one of the most common types of occupational skin disorders. Special issues associated with these disorders include the following:

1. Objective information on exposure history (a factory visit is ideal) is important. Direct exposure to chemicals can occur because of spills or routine work levels; indirect exposure can come from contaminated tools, rags, and gloves; and airborne exposures can result from mists, droplets, and sprays.
2. The skin is an important portal of entry for a number of toxic chemicals that may or may not have a direct effect on skin. These chemicals include aniline, carbon disulfide, ethylene glycol ethers, certain pesticides, tetrachloroethylene, and toluene.
3. Patch testing with industrial chemicals should be performed very carefully. Irritants should not be tested, and many require dilution to nonirritating concentrations. Testing with individual chemical components of mixtures is preferable in many cases.
4. Establishing occupational causation for ACD is often a challenge, and recommendations have been published.

Prevention and treatment of occupational contact dermatitis is the same as for ACD.[16]

Subtypes of Contact Dermatitis

PHOTOSENSITIVITY

Photosensitivity refers to a condition in which ultraviolet light in combination with endogenous or exogenous substances, usually drugs or chemicals, evokes an eruption on sun-exposed skin. Most cases are evoked by ultraviolet A, but on occasion, eruptions are caused by ultraviolet B (sunburn irradiation) or by visible light. The most common causes are systemic exposure to photosensitizing drugs [*see Table 6*] or cutaneous exposure, usually accidental, to psoralen in plants.

Photosensitivity reactions are of two types: phototoxicity and photoallergy. Many substances that are photoallergic at low concentrations may be phototoxic at high concentrations.[43]

Phototoxicity

Phototoxicity is analogous to irritation and occurs in any individual after one exposure to sufficient amounts of chemical and light. Phototoxicity has been likened to an exaggerated sunburn response, consisting of delayed erythema and edema followed by pigmentation and desquamation. Asphalt workers and roofers working with pitch develop the so-called smarts when exposed to sufficient sunlight. Phytophotodermatitis, or meadow dermatitis, is a particularly striking phototoxicity characterized by streaky bullae after contact, sometimes while sunbathing, with psoralen containing umbelliferones. Berloque der-

Table 6 Topical and Systemic Photosensitizers[17,44]

Agent	*PT/PA*	*Common Sources/Forms*
Topical photosensitizers		
Psoralens	PT	Plants and drugs
Pitch, creosote, and coal tar derivatives	PT	Medications/industrial products
Halogenated salicylanilides (e.g., bithionol, dibromosalicylanilide)	PA	Antibacterials in soaps and detergents
Musk ambrette	PA	Fragrance
Oxybenzone/padimate O	PA	Sunscreens
Phenothiazines	PA	Topical drugs
Ketoprofen	PA	Nonsteroidal anti-inflammatory drugs
Systemic photosensitizers		
Thiazides	PT	Diuretics
Phenothiazines	PT	Tranquilizers
Dimethylchlorotetracycline	PT	Antibiotic
Griseofulvin	PT	Antifungal
Nalidixic acid	PT	Antibiotic
Sulfonamides	PT	Antibiotic
Psoralens	PT	Photosensitizing drug
Piroxicam	PT?	Nonsteroidal anti-inflammatory drugs, especially in thimerosal-sensitive patients

PA—photoallergenic reaction PT—phototoxic reaction

matitis is a phototoxic dermatitis characterized by the appearance of hyperpigmented, droplike patches on the neck, face, and breast. This reaction is caused by exposure to 5-methoxypsoralen present in perfumes or colognes containing oil of bergamot, and the hyperpigmentation may persist for many months. Photo-onycholysis has been reported with tetracyclines, psoralen, and other phototoxic drugs. Not all cases exhibit obvious skin phototoxicity.[43] Most cases of phototoxicity are caused by administration of phototoxic systemic drugs [*see Table 6*].

Photoallergy

Photoallergy is analogous to ACD and is an immunologic reaction in which exposure of the photosensitizing compound to UV light plays a role in formation of a complete antigen. A delayed eruption, usually eczematous, appears in sun-exposed body areas, usually the face and dorsal hands, typically sparing the submental and retroauricular areas [*see Figure 11*]; shaded areas and covered areas remain relatively clear but occasionally are involved. Most cases are caused by topical photoallergens [*see Table 6*], and the most common photocontact allergens are sunscreen chemicals, which act by absorbing ultraviolet light. Oxybenzone is a common allergen; however, other sunscreen chemicals, such as padimate O and the dibenzoylmethanes, have also been reported to cause photoallergic contact dermatitis.[43]

Photoallergic contact dermatitis is reproduced and diagnosed by photopatch testing, a procedure in which ultraviolet light (usually ultraviolet A) is combined with patch testing. This form of testing is particularly helpful in differentiating eruptions caused by polymorphous light from photoallergic contact dermatitis. Photopatch testing is not indicated in phototoxic drug eruptions. In some persons, photoallergic reactions can persist as chronic actinic dermatitis (CAD), which can be difficult to treat (see below). Patients with photoallergic contact dermatitis often have contact allergy and should also be patch tested.

Treatment of Phototoxicity and Photoallergy

Elimination of exposure to the photoallergen or phototoxic agent is effective for most patients, except for a few with CAD. Broad-spectrum sunscreens or sunblocks, especially those containing micronized titanium, along with sun-protective clothing, may be helpful. Topical corticosteroids are helpful for mildly affected patients, but severely affected patients with CAD may require azathioprine, with or without systemic corticosteroids; psoralen ultraviolet A therapy and cyclosporine have also been used in some severe cases.[17]

LATEX ALLERGY

Latex allergy is an IgE-mediated hypersensitivity to one or more of a number of proteins present in raw or uncured natural rubber latex (NRL). The paradigm for immunologic contact urticaria is latex allergy, which, over the past decade, has become a significant medical and occupational health problem.

Populations at Risk

Individuals at highest risk are patients with spina bifida (30% to 65% prevalence), health care workers, and other workers with significant NRL exposure.[44] Most reported series of occupational cases of latex allergy involve health care workers; 5% to 11% of those studied are affected.[45] Studies of populations of non–health care workers are infrequent; however, evidence indicates that sensitization to NRL is more common in food handlers, construction workers, painters, hairdressers, cleaners, and miscellaneous other occupations in which NRL is utilized.[46] Children with chronic renal failure appear to be at increased risk for latex sensitization.[47]

Risk Factors and Etiology

Predisposing risk factors are hand eczema, allergic rhinitis, allergic conjunctivitis, or asthma in individuals who frequently wear NRL gloves; mucosal exposure to NRL; and multiple surgical procedures.[44,45] The majority of cases of latex allergy involve reactions from wearing NRL gloves or being examined by individuals wearing NRL gloves. Reactions from other medical and

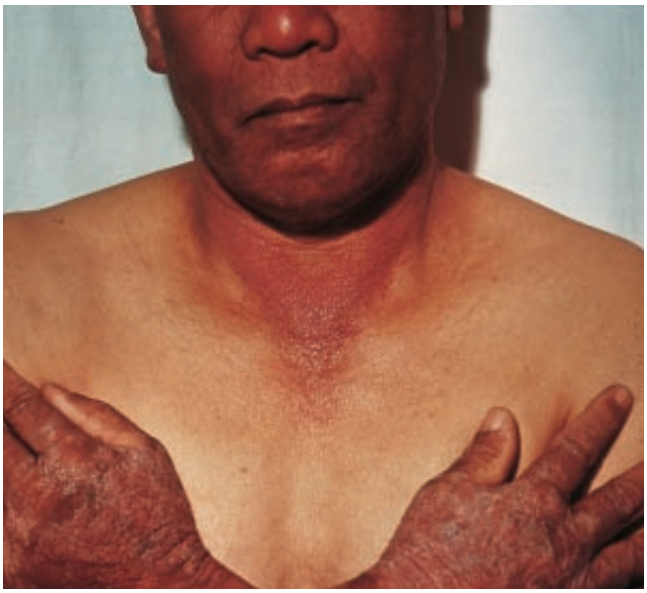


Figure 11 **Photocontact dermatitis characteristically involves areas exposed to the sun.**

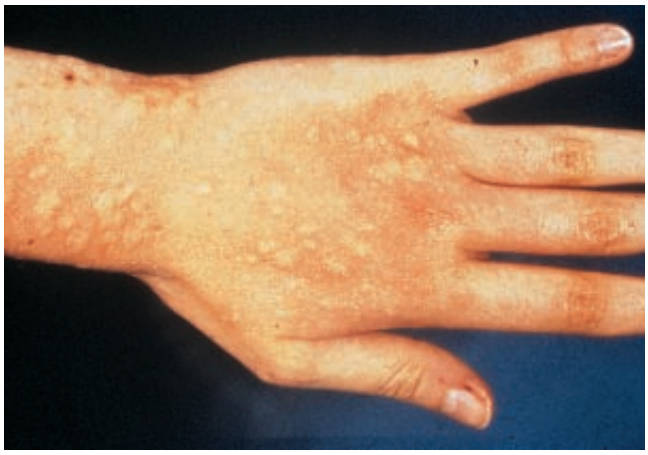


Figure 12 **Contact urticaria of the hands in a nurse allergic to her powdered natural rubber latex gloves (latex allergy). She also experienced allergic rhinitis and asthma while at work. Urticaria is often short-lived after gloves are used and may be absent at the time of examination.**

nonmedical NRL devices have occurred; these include balloons, rubber bands, condoms, vibrators, dental dams, anesthesia equipment, and toys for animals or children.

The route of exposure to NRL proteins includes direct contact with intact or inflamed skin and mucosal exposure, such as inhalation of powder from NRL gloves, especially in medical facilities and in operating rooms.[47] Most immediate-type NRL reactions result from exposure to dipped NRL products (e.g., gloves, condoms, balloons, and tourniquets). Dry-molded rubber products (e.g., syringes, plungers, vial stoppers, and baby-bottle nipples) contain lower residual protein levels or have less easily extracted proteins than do dipped NRL products.

NRL allergy is sometimes associated with allergic reactions to fruit (especially bananas, kiwi, and avocados) and to chestnuts. This allergic reaction results from cross-reactivity between proteins in NRL and those found in some fruits and nuts. Symptoms range from oral itching and angioedema to asthma, gastrointestinal upset, and anaphylaxis. Cross-reactivity to NRL may be a factor in other skin eruptions. One report suggests that the use of commercially available black henna tattoos may cause hypersensitivity to NRL.[48]

Internet Resources on Contact Dermatitis

Contact Dermatitis

American Contact Dermatitis Society
http://www.contactderm.org

American Academy of Dermatology
http://www.aad.org

Occupational Contact Dermatitis

National Institute for Occupational Safety and Health
http://www.cdc.gov/NIOSH

Canadian Centre for Occupational Health and Safety
http://www.ccohs.ca/oshanswers/diseases/allergic_derm.html

Latex Allergy

A.L.E.R.T. Inc.
http://www.latexallergyresources.org

Spina Bifida Association of America
http://www.sbaa.org

Diagnosis

Clinical signs of NRL allergy include contact urticaria [*see Figure 12*], generalized urticaria, allergic rhinitis, allergic conjunctivitis, angioedema, asthma, and anaphylaxis.[45] More than 600 serious reactions to NRL, including 16 fatal anaphylactic reactions, were reported to the Food and Drug Administration by the early 1990s.

Diagnosis of NRL allergy is strongly suggested by a history of angioedema of the lips when inflating balloons and by a history of itching, burning, urticaria, or anaphylaxis when donning gloves; when undergoing surgical, medical, and dental procedures; or after exposure to condoms or other NRL devices. Diagnosis is confirmed by a positive wear or use test with NRL gloves, a valid positive intracutaneous prick test with NRL, or a positive serum radioallergosorbent test with NRL.[44] Severe allergic reactions have occurred from prick and wear tests; epinephrine latex-safe resuscitation equipment free of NRL should be available during these procedures.[45]

Treatment and Risk Reduction

Hyposensitization to NRL is not yet feasible, and NRL avoidance and substitution are imperative. Because many patients with NRL allergy have hand eczema, have immediate allergic symptoms, or both, the most important issues for physicians are accurate diagnosis, appropriate treatment, and counseling.

Preventive measures have significantly reduced the prevalence of reported reactions to NRL.[45] Risk reduction and control of NRL allergy include latex avoidance in health care settings for affected workers and patients. Synthetic non-NRL gloves should be available to replace latex gloves. Also, in many cases, low-allergen NRL gloves should be worn by coworkers so as to minimize symptoms and decrease induction of NRL allergy in those allergic to NRL. Allergen content of gloves should be requested from manufacturers and suppliers; lists of glove allergen levels have also been published. Patients with NRL allergy should wear Medic-Alert bracelets identifying them as NRL sensitive, and they should inform health care providers of their sensitivity. These patients should be given lists of substitute gloves, other non-NRL devices, potentially allergenic fruits, latex-safe anesthesia protocols, occult sources of NRL exposure such as toys (for animals and children), and dental prophylaxis cups. Some of this information is available in published sources, government agencies, and latex-allergy support groups that publish newsletters and other relevant information. Some sources have Web sites [*see Sidebar* Internet Resources on Contact Dermatitis].

The author receives grants for clinical research from Guidant Corp. and Mekos Laboratories AS and serves as a consultant for ConvaTec and Procter & Gamble.

References

1. Rietschel RR, Adams RM, Daily AD, et al: Guidelines of care for contact dermatitis. J Am Acad Dermatol 32:109, 1995

2. Schappert SM: National Ambulatory Medical Care Survey: 1994 Summary. Adv Data 273:1, 1996

3. De Groot AC: Patch Testing: Test Concentrations and Vehicles for 3700 Chemicals, 2nd ed. Elsevier, Amsterdam, 1994

4. Wigger-Alberti W, Elsner P: Contact dermatitis due to irritation. Handbook of Occupational Dermatology. Kanerva L, Elsner P, Wahlberg JE, et al, Eds. Springer Verlag, Berlin, 2000, p 99

5. Iliev D, Elsner P: Clinical irritant contact dermatitis syndromes. Immunol Allergy Clin North Am 17:365, 1997

6. Dickel H, Kuss O, Schmidt A, et al: Importance of irritant dermatitis in occupational skin disease. Am J Clin Dermatol 3:283, 2002

7. Schwanitz HJ, Uter W: Interdigital dermatitis: sentinel skin damage in hairdressers.

Br J Dermatol 142:1011, 2000

8. Belsito DV: Allergic contact dermatitis. Fitzpatrick's Dermatology in General Medicine, 6th ed. Freedberg IM, Eisen AZ, Wolff, K, et al, Eds. McGraw-Hill, New York, 2003, p 1447

9. Condie MW, Adams RM: Influence of oral prednisone on patch test reaction to Rhus antigen. Arch Dermatol 107:540, 1973

10. Wulfhorst B: Skin hardening in occupational dermatology. Handbook of Occupational Dermatology. Kanerva L, Elsner P, Wahlberg J, et al, Eds. Springer Verlag, New York, 2000, p 115

11. Pratt MD, Belsito DV, DeLeo VA, et al: North American Contact Dermatitis Group patch-test results, 2001–2002 study period. Dermatitis 15:176, 2004

12. Rietschel RL, Fowler JF: Fisher's Contact Dermatitis, 5th ed. Williams & Wilkins, Baltimore, 2000

13. Gaspari AA: The role of keratinocytes in the pathophysiology of contact dermatitis. Immunol Allergy Clin North Am 17:377, 1997

14. Corsini E, Galli CL: Cytokines and irritant contact dermatitis. Toxicol Lett 102:277, 1998

15. Krob HA, Fleischer AB Jr, D'Agostino R Jr, et al: Prevalence and relevance of contact dermatitis allergens: a meta-analysis of 15 years of published T.R.U.E. test data. J Am Acad Dermatol 51:349, 2004

16. Taylor JS: Recognizing allergic contact dermatitis. Practical Contact Dermatitis. Guin JD, Ed. McGraw Hill, New York, 1995, p 31

17. Beck MH, Wilkinson SM: Contact dermatitis: allergic. Rook's Textbook of Dermatology, 7th ed. Burns T, Griffiths CE, Kinghorn BP, et al, Eds. Blackwell Science, London, 2004, p 20-1

18. Saripalli YV, Achen F, Belsito DV: The detection of clinically relevant contact allergens using a standard screening tray of twenty-three allergens. J Am Acad Dermatol 49:65, 2003

19. Rietschel RL: Experience with supplemental allergens in the diagnosis of contact dermatitis. Cutis 65:27, 2000

20. Bourke JF, Batta K, Prais L, et al: The reproducibility of patch tests. Br J Dermatol 140:102, 1999

21. Nethercott JR, Holness DL: Validity of patch test screening trays in the evaluation of patients with allergic contact dermatitis. J Am Acad Dermatol 21:568, 1989

22. Ale SI, Maibach HI: Reproducibility of patch test results: a concurrent right-versus-left study using TRUE Test. Contact Dermatitis 50:304, 2004

23. Diepgen TL, Coenraads PJ: Sensitivity, specificity and positive predictive value of patch testing: the more you test, the more you get? Contact Dermatitis 42:315, 2000

24. Basketter DA, Griffiths HA, Wang XM, et al: Individual, ethnic, and seasonal variation in irritant susceptibility of skin: the implications for a predictive human patch test. Contact Dermatitis 35:208, 1996

25. Zappi E, Brancaccio RR: Allergic contact dermatitis from mupirocin ointment. J Acad Derm 36:266, 1977

26. Wigger-Alberti W, Elsner P: Preventive measures in contact dermatitis. Clin Dermatol 15:661, 1997

27. Jungbauer FH, van der Harst JJ, et al: Skin protection in nursing work: promoting the use of gloves and hand alcohol. Contact Dermatitis 51:135, 2004

28. Marks JG Jr, Fowler JF Jr, Sherertz EF, et al: Prevention of poison ivy and poison oak allergic contact dermatitis by quaternium-18 bentonite. J Am Acad Dermatol 33:212, 1995

29. Elsner P, Wiggerti-Alberti W, Pantini G: Perfluoropolyethers in the prevention of irritant contact dermatitis. Dermatology 197:141, 1998

30. Taylor JS: Occupational dermatoses. Conn's Current Therapy. Rakel RE, Ed. WB Saunders, Philadelphia, 1996, p 823

31. Taylor JS, Praditsuwan P, Handel D, et al: Allergic contact dermatitis from doxepin cream: a one-year patch test clinic experience. Arch Dermatol 132:515, 1996

32. Bonnel RA, La Grenade L, Karwoski CB, et al: Allergic contact dermatitis from topical doxepin: Food and Drug Administration's postmarketing surveillance experience. J Am Acad Dermatol 48:294, 2003

33. Keegel T, Saunders H, Milne R, et al: Topical corticosteroid allergy in an urban Australian center. Contact Dermatitis 50:6, 2004

34. Isaksson M, Andersen KE, Brandao FM, et al: Patch testing with corticosteroid mixes in Europe: a multicentre study of the EECDRG. Contact Dermatitis 42:27, 2000

35. Isaksson M, Bruze M: Allergic contact dermatitis in response to budesonide reactivated by inhalation of the allergen. J Am Acad Dermatol 46:880, 2002

36. Veien NK: Ingested food in systemic contact dermatitis. Clin Dermatol 15:547, 1997

37. Menne T, Veien N, Sjolin K-E, et al: Systemic contact dermatitis. Am J Contact Dermatitis 5:1, 1994

38. Maibach HI: Oral substitution in patients sensitized by transdermal clonidine treatment. Contact Dermatitis 16:1, 1987

39. Seidenari S, Giusti F, Massone F, et al: Sensitization to disperse dyes in a patch test population over a five-year period. Am J Contact Dermatitis 13:101, 2002

40. Hatch KL, Motschi H, Maibach HI: Disperse dyes in fabrics of patients patch-test-positive to disperse dyes. Am J Contact Dermatitis 14:205, 2003

41. Mathias CGT: Contact dermatitis and worker's compensation: criteria for establishing occupational causation and aggravation. J Am Acad Dermatol 20:842, 1989

42. Pratt M, Taraska V: Disperse blue dyes 106 and 124 are common causes of textile dermatitis and should serve as screening allergens for this condition. Am J Contact Dermatitis 11:30, 2000

43. Isaksson M, Bruze M: Photocontact dermatitis: photopatch testing. Clin Dermatol 15:615, 1997

44. Taylor JS, Wattanakrai P, Charous L, et al: Latex allergy. 1999 Yearbook of Dermatology and Dermatologic Surgery. Thiers BH, Lang PG, Eds. Mosby, St. Louis, 1999, p 1

45. Taylor JS, Erkek E: Latex allergy: diagnosis and management. Dermatol Ther 17:289, 2004

46. Valks R, Conde-Salazar L, Cuevas M: Allergic contact urticaria from natural rubber latex in healthcare and non-healthcare workers. Contact Dermatitis 50:222, 2004

47. Sparta G, Kemper MJ, Gerber AC, et al: Latex allergy in children with urological malformation and renal failure. J Urol 171:1647, 2004

48. Martin JA, Hughes TM, Stone NM: 'Black henna' tattoos: an occult source of natural rubber latex allergy? Contact Dermatitis 52:145, 2005

49. Marks JG, DeLeo V: Contact and Occupational Dermatology, 2nd ed. Mosby, St. Louis, 1997

Acknowledgments

Figure 7 Courtesy of James R. Nethercott, M.D. (deceased), Department of Dermatology, University of Maryland, Baltimore.

Figure 11 Courtesy of Kristina Turjanmaa, M.D., and Arto Lahti, M.D., Tampere and Oulu, Finland.

VI CUTANEOUS ADVERSE DRUG REACTIONS

Neil H. Shear, m.d.
Sandra Knowles, b.sc. pharm.
Lori Shapiro, m.d.

An adverse drug reaction (ADR) is defined as any noxious, unintended, and undesired effect of a drug that occurs at doses used in humans for prophylaxis, diagnosis, or therapy.[1] An ADR may range from a cutaneous eruption to severe syndromes (e.g., drug hypersensitivity syndrome, Stevens-Johnson syndrome [SJS], toxic epidermal necrolysis [TEN], and serum sickness–like reaction). Over the past 20 years, a dramatic shift has occurred in our understanding of drug-induced cutaneous eruptions. It is now believed that many severe cutaneous adverse drug reactions are caused by the formation of reactive oxidative metabolites and perhaps the formation of antibodies to drug-protein complexes and skin proteins, cytochrome P-450 enzymes, or both. The predisposition to drug-induced eruptions may be genetic, and family counseling and in vitro testing are being used in certain centers to manage patients and their families. This chapter reviews the pathophysiology and clinical manifestations that are important for correct diagnosis and treatment of cutaneous ADRs.

Epidemiology

Epidemiologic studies have shown that ADRs occur in 6.7% of all hospitalized patients,[2] and 3% to 6% of all hospital admissions are the result of ADRs.[3] In the Boston Collaborative Drug Surveillance Program,[4] the prevalence of cutaneous ADRs in hospitalized patients was 2.2%. Antibiotics were responsible for 75% of detected reactions. In the Harvard Medical Practice Study, approximately 14% of ADRs in hospital patients were cutaneous or allergic in nature.[5] The cost of drug-related morbidity and mortality has been estimated at $30 billion a year,[6] and ADRs are thought to be between the fourth and sixth leading cause of death in the United States.[2,6]

Etiology

Cutaneous reactions to drugs often occur in complicated clinical scenarios that may include exposure to multiple agents. New drugs started within the preceding 6 weeks are potential causative agents, as are drugs that have been used intermittently, including over-the-counter preparations and herbal and naturopathic remedies.

Diagnosis

CLINICAL MANIFESTATIONS

The morphology of cutaneous eruptions may be exanthematous, urticarial, blistering, or pustular. The extent of the reaction is variable. For example, once the morphology of the reaction has been documented, a specific diagnosis (e.g., fixed drug eruption or acute generalized exanthematous pustulosis) can be made. The reaction may also present as a syndrome (e.g., serum sickness–like reaction or hypersensitivity syndrome reaction). Fever is associated with the more serious cutaneous ADRs.

DIFFERENTIAL DIAGNOSIS

Differential diagnoses can include viral exanthems (e.g., infectious mononucleosis and parvovirus B19 infection), bacterial infections, Kawasaki syndrome, collagen vascular disease, and neoplasia.[7]

LABORATORY TESTS

Penicillin skin testing with major and minor determinants is useful for confirmation of an IgE-mediated immediate hypersensitivity reaction to penicillin.[8] Skin tests are performed 6 weeks to 6 months after complete healing of the cutaneous drug reaction.[9] Oral rechallenges may be useful in the diagnosis of ADRs; however, they should not be used if a serious reaction, such as SJS or TEN, previously occurred. Patch testing may be helpful in the diagnosis of fixed drug eruptions or contact dermatitis.[10]

Exanthematous Eruptions

SIMPLE ERUPTIONS

Exanthematous eruptions, also known as morbilliform, maculopapular, or scarlatiniform eruptions, are the most common cutaneous ADRs.[4] Simple exanthems are erythematous changes in the skin without blistering or pustulation.

Many drugs can cause exanthematous eruptions, including the penicillins, sulfonamides, barbiturates, antiepileptic medications, nonnucleoside reverse transcriptase inhibitors (e.g., nevirapine), and antimalarials.[4,11] Exanthematous eruptions occur in 3% to 7% of patients receiving such aminopenicillins as ampicillin and amoxicillin. However, these eruptions may occur in 60% to 100% of patients taking ampicillin or amoxicillin who are receiving concurrent allopurinol therapy or who have concomitant lymphocytic leukemia, infectious mononucleosis, cytomegalovirus infection, or hyperuricemia.

Studies suggest that some exanthematous eruptions represent cell-mediated hypersensitivity.[12,13] The etiology of the ampicillin rash concurrent with a viral infection is unknown, but the rash does not appear to be IgE mediated, and patients can tolerate all β-lactam antibiotics, including ampicillin, once the infectious process has resolved. A similar reaction was seen in 50% of HIV-infected patients exposed to sulfonamide antibiotics.[14] Recent studies have shown that drug-specific T cells play a major role in exanthematous, bullous, and pustular drug reactions.[15]

Simple exanthems are symmetrical and often become generalized. Pruritus is the most frequently associated symptom. Fever is not associated with simple exanthematous eruptions. These eruptions usually occur within 1 week after the beginning of therapy and generally resolve within 7 to 14 days.[16] The exanthem's turning from bright red to brownish red marks resolution. Resolution may be followed by scaling or desquamation.[17] Some patients with ampicillin- or amoxicillin-induced exanthematous eruptions may have a positive result on a patch test or on a delayed intradermal test.[13,14] In general, however,

skin testing is not considered helpful in the diagnosis of an exanthematous eruption.

The differential diagnosis of drug-induced exanthematous eruption includes viral exanthem (patients should be tested for mononucleosis), collagen vascular disease, bacterial infection, and rickettsial infection. Hypersensitivity syndrome should be considered in the differential diagnosis.

The treatment of simple exanthematous eruptions is generally supportive. For example, oral antihistamines used in conjunction with soothing baths may help relieve pruritus. Topical corticosteroids are indicated when antihistamines do not provide relief. Systemic corticosteroids are used only in severe cases. Discontinuance of the offending agent is recommended in most cases.

COMPLEX ERUPTIONS

Hypersensitivity Syndrome Reaction

Hypersensitivity syndrome reaction is a complex drug reaction that affects various organ systems. A triad of fever, skin eruption, and internal organ involvement signals this potentially life-threatening syndrome. It occurs in approximately one in 3,000 exposures to such agents as aromatic anticonvulsants (e.g., phenytoin, phenobarbital, and carbamazepine), lamotrigine, sulfonamide antibiotics, dapsone, minocycline, and allopurinol.[18]

It has been suggested that the metabolism of aromatic anticonvulsants by cytochrome P-450 plays a pivotal role in the development of the hypersensitivity syndrome reaction with these drugs.[19] In most people, the chemically reactive metabolites that are produced are detoxified by epoxide hydroxylases. If detoxification is defective, however, one of the metabolites may act as a hapten and initiate an immune response, stimulate apoptosis, or cause cell necrosis directly.

In one study, 75% of patients with hypersensitivity syndrome reactions to one aromatic anticonvulsant showed in vitro cross-reactivity to the other two aromatic anticonvulsants.[19] In addition, in vitro testing has shown that there is a familial occurrence of hypersensitivity to anticonvulsants.[19] Although lamotrigine is not an aromatic anticonvulsant, it too can cause a hypersensitivity syndrome reaction.[20,21] There is no evidence that lamotrigine cross-reacts with the aromatic anticonvulsants. Lamotrigine and other anticonvulsants are also associated with more severe reactions (e.g., SJS and TEN) [*see* Complex Eruptions, *below*].

Sulfonamide antibiotics can cause hypersensitivity syndrome reactions in susceptible persons. The primary metabolic pathway for sulfonamides involves acetylation of the drug to a nontoxic metabolite and renal excretion. An alternative metabolic pathway, quantitatively more important in patients who are slow acetylators, engages the cytochrome P-450 mixed-function oxidase system. These enzymes transform the parent compound to reactive metabolites—namely, hydroxylamines and nitroso compounds, which produce cytotoxicity independently of preformed drug-specific antibody. In most people, detoxification of the metabolite occurs. However, hypersensitivity syndrome reactions may occur in patients who are unable to detoxify this metabolite (e.g., those who are glutathione deficient).[22] Although the detoxification defect is present in 2% of the population, only one in 10,000 people will manifest a hypersensitivity syndrome reaction in response to sulfonamide antibiotics. Siblings and other first-degree relatives of patients with the detoxification defect are at increased risk (perhaps one in four) for having a similar defect.

Other aromatic amines, such as procainamide, dapsone, and acebutolol, are also metabolized to chemically reactive compounds. We recommend that patients who develop symptoms compatible with a sulfonamide hypersensitivity syndrome reaction avoid these aromatic amines, because the potential exists for cross-reactivity. However, cross-reactivity should not occur between sulfonamides and drugs that are not aromatic amines (e.g., sulfonylureas, thiazide diuretics, furosemide, and acetazolamide).

Hypersensitivity syndrome reaction occurs most frequently on first exposure to the drug, with initial symptoms starting 1 to 6 weeks after exposure [*see Table 1*]. Fever and malaise, which can be accompanied by pharyngitis and cervical lymphadenopathy, are the presenting symptoms in most patients. This is often followed by edema and swelling of the face, especially upon rising in the morning. Atypical lymphocytosis, with subsequent eosinophilia, may occur during the initial phases of the reaction in some patients. A cutaneous eruption, which occurs in approximately 85% of patients, can range from an exanthematous eruption [*see Figure 1*] to the more serious SJS or TEN. The liver is often involved, resulting in hepatitis, although other internal organs may be affected, such as the kidney (e.g., interstitial nephritis and vasculitis), the central nervous system (e.g., encephalitis and aseptic meningitis), and the lungs (e.g., interstitial pneumonitis, respiratory distress syndrome, and vasculitis). A subgroup of patients may become hypothyroid as part of an autoimmune thyroiditis within 2 months after the initiation of symptoms.[23]

After hypersensitivity syndrome reaction has been recognized from the symptom complex of fever, rash, and lymphadenopathy, some laboratory tests can be used to evaluate internal organ involvement, which may be asymptomatic. A complete blood count, urinalysis, and measurements of liver transaminase and serum creatinine levels should be per-

Table 1 Clinical Features of Hypersensitivity Syndrome Reaction and Serum Sickness–like Reaction

	Rash	*Fever*	*Internal Organ Involvement*	*Arthralgia*	*Lymphadenopathy*
Hypersensitivity syndrome reaction	Exanthem Exfoliative dermatitis Pustular eruptions Erythema multiforme Stevens-Johnson syndrome Toxic epidermal necrolysis	Present	Present	Absent	Present
Serum sickness–like reaction	Urticaria Exanthem	Present	Absent	Present	Present

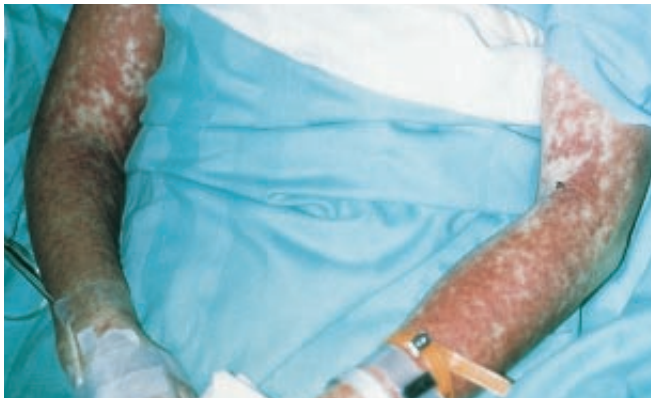


Figure 1 **This 35-year-old woman developed hypersensitivity syndrome reaction, characterized by fever, rash, and hepatitis, 14 days after starting trimethoprim-sulfamethoxazole therapy. The rash is an extensive, symmetrical, red edematous eruption.**

formed. In addition, the clinician should be guided by symptoms that may suggest specific internal organ involvement (e.g., respiratory symptoms). Thyroid function should be evaluated on presentation of hypersensitivity syndrome reaction and then 2 to 3 months after presentation. A skin biopsy may help confirm the diagnosis when the patient has a blistering or a pustular eruption. Unfortunately, diagnostic or confirmatory tests are not readily available. An in vitro test employing a mouse hepatic microsomal system is used for research purposes to characterize patients who develop hypersensitivity syndrome reaction.[19,24] Because of the severity of the reaction, oral rechallenges are not recommended.

Although the role of corticosteroids is controversial, most clinicians choose to start prednisone at a dosage of 1 to 2 mg/kg/day when symptoms are severe. Antihistamines, topical corticosteroids, or both can be used to alleviate symptoms. Because the risk of hypersensitivity syndrome reaction in first-degree relatives of patients who have had reactions is substantially higher than in the general population, counseling of family members regarding their risk of hypersensitivity syndrome reaction is advised.

Urticarial Eruptions

SIMPLE ERUPTIONS

Urticaria and Angioedema

Urticaria is characterized by pruritic red wheals of varying sizes that can occur with any medication. When deep dermal and subcutaneous tissues are also swollen, the reaction is known as angioedema.

Urticaria and angioedema usually result from a type I immediate hypersensitivity reaction. This mechanism is typified by immediate reactions to penicillin and other antibiotics. Binding of the drug or its metabolite to IgE bound to the surfaces of cutaneous mast cells leads to activation, degranulation, and release of vasoactive mediators such as histamine, leukotrienes, and prostaglandins.[25]

Urticarial reactions may also result from nonimmunologic activation of inflammatory mediators. Drugs such as acetylsalicylic acid and nonsteroidal anti-inflammatory drugs (NSAIDs),[26] radiocontrast media, and narcotic analgesics[27] may directly cause release of histamine from mast cells, independently of IgE. Angiotensin-converting enzyme (ACE) inhibitors are frequent causes of angioedema.[28] The mechanism of this reaction is unclear but may relate to accumulation of bradykinin or activation of the complement system.

Although medications tend to cause urticaria, angioedema, or both, other causal agents are food [*see 6:XVI Food Allergies*], physical factors (e.g., dermatographism and cholinergic urticaria) [*see 6:XIII Urticaria, Angioedema, and Anaphylaxis*], and idiopathic factors. Certain foods containing proteins that can cross-react with latex proteins, such as bananas, kiwifruit, avocados, and chestnuts, can cause oral itching and swelling, hives, or wheezing after ingestion. People at greatest risk for latex allergy include children with spina bifida and health care workers.[29,30] Latex allergy can present as contact urticaria at sites of latex exposure, such as lip swelling in a person who has blown up a balloon or sucked on a pacifier. Contact with aerosolized powder from latex gloves to which the latex protein has adhered may cause mucosal symptoms, such as itchy, swollen eyes; runny nose; sneezing; or wheezing. Anaphylaxis may also occur.[31]

Signs and symptoms of IgE-mediated allergic reactions are typically pruritus, urticaria, cutaneous flushing, angioedema, nausea, vomiting, diarrhea, abdominal pain, nasal congestion, rhinorrhea, laryngeal edema, and bronchospasm or hypotension or both. Fever is not associated with urticaria or angioedema reactions. In general, individual lesions of urticaria last for less than 24 hours, although new lesions can continually develop. With ACE-inhibitor therapy, the onset of the adverse reaction is usually within hours but can occur as late as 1 week to several months into therapy.[32] With treatment, the resulting angioedema usually resolves within 48 hours.

Skin testing may be helpful in cases of IgE-mediated urticaria. For example, penicillin skin testing with the major and minor determinants identifies approximately 99% of patients who have had an IgE-mediated reaction to penicillin.[10] A latex skin test is a sensitive indicator of IgE sensitization.[31] For large-molecular-weight agents, such as insulin,[33] protamine,[34] and egg-containing vaccines, positive immediate skin-test reactions identify patients at risk for IgE-mediated reactions.

Withdrawal of the causative agent is recommended. When angioedema or anaphylaxis occurs, immediate therapy with epinephrine and systemic steroids may be needed. Symptomatic relief can generally be achieved with antihistamines (H_1 receptor blockers).

Differential Diagnosis

Allergic urticaria must be differentiated from urticaria caused by physical factors. Cold urticaria, for example, is precipitated by exposure to cold, occurring within minutes after immersion of hands or body in cold water or after exposure to cold air. In severe cases, systemic symptoms, including wheezing and syncope, can occur. A rare familial form of cold urticaria that is autosomal dominant has been linked to chromosome 1q44.[35]

Cold urticaria can be differentiated from other forms of urticaria by eliciting an urticarial reaction with an ice cube applied to the skin for 5 to 10 minutes. Other physical urticarias also have distinguishing causes or features. Solar urticaria occurs within minutes of exposure to sunlight and can be produced by exposing limited areas of skin to sunlight or to appro-

priate wavelengths of ultraviolet light in a phototherapy response to physical pressure. Cholinergic urticaria, which is characterized by small urticarial papules, can be induced by exposure to heat or by exercise.

Histologically, all the urticarias are characterized by an increase in mast cells in the dermis. Edema, vascular changes, and mononuclear infiltrates are more striking in the dermis of patients with cold urticaria. Mononuclear infiltrates are also more prominent in the deep dermis of patients with delayed pressure urticaria.[36]

As with drug-induced urticaria, first-line therapy of most urticarias consists of oral antihistamines and avoidance of precipitating factors. Psoralen plus ultraviolet A (PUVA) has been used successfully to treat patients with solar urticaria. Montelukast has been used successfully to treat delayed pressure urticaria,[37] and cyclosporine is promising for cases of severe refractory chronic urticaria.[38]

COMPLEX ERUPTIONS

Serum Sickness–like Reactions

Serum sickness–like reactions are defined by fever, rash (usually urticarial), and arthralgias occurring 1 to 3 weeks after drug initiation. Other symptoms, such as lymphadenopathy and eosinophilia, may also be present. In contrast to true serum sickness, serum sickness–like reactions are without immune complexes, hypocomplementemia, vasculitis, and renal lesions [*see Table 1*].

Epidemiologic studies in children suggest that the risk of serum sickness–like reactions is greater with cefaclor than with other antibiotics, including other cephalosporins.[39,40] The overall incidence of cefaclor serum sickness–like reactions has been estimated to be 0.024% to 0.2% per course of cefaclor prescribed.

Although the pathogenesis is unknown, it has been postulated that in genetically susceptible hosts, metabolism of cefaclor produces a reactive metabolite that may bind to tissue proteins and elicit an inflammatory response that manifests as a serum sickness–like reaction.[41]

Other drugs that have been implicated in serum sickness–like reactions are cefprozil,[42] bupropion,[43] and minocycline.[18,44] The incidence of serum sickness–like reactions caused by these drugs is unknown.

Discontinuance of the culprit drug and symptomatic treatment with antihistamines and topical corticosteroids are recommended for patients with serum sickness–like reactions. A short course of oral corticosteroids may be required for patients with more severe symptoms. The drug that caused the serum sickness–like reaction should be avoided. For cefaclor and cefprozil, the risk of cross-reaction with β-lactam antibiotics is small, and the administration of another cephalosporin is usually well tolerated.[45] However, some clinicians recommend that patients who experience serum sickness–like reactions from cefaclor avoid all β-lactam drugs.[46]

Blistering Eruptions

SIMPLE ERUPTIONS

Fixed Drug Eruptions

Fixed drug eruptions usually appear as solitary pruritic, erythematous, bright-red or dusky-red macules that may evolve into an edematous plaque [*see Figure 2*]. In some patients, multiple lesions may be present. Blistering and erosion may occur on mucosal surfaces.

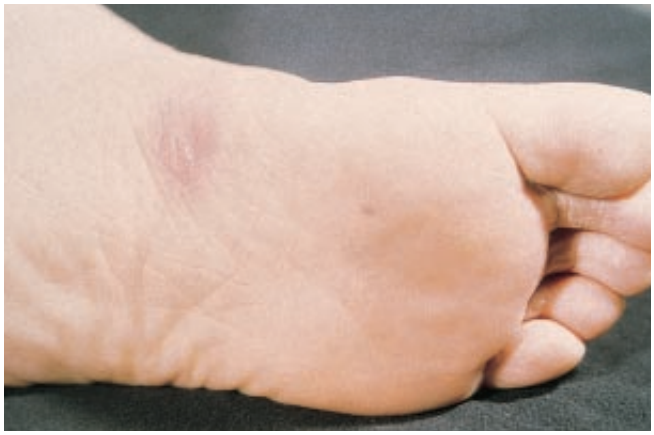


Figure 2 **This 28-year-old man taking tetracycline for acne vulgaris developed a fixed drug eruption.**

Fixed drug eruptions recur in the same skin area after readministration of the causative medication. Many drugs have been implicated in fixed drug eruptions, including phenolphthalein, ibuprofen, sulfonamides, tetracyclines, and barbiturates.[47] The pathogenesis of fixed drug eruptions has not been fully elucidated. A haplotype linkage in the setting of trimethoprim-sulfamethoxazole–induced fixed drug eruptions was recently documented.[48]

Fixed drug eruptions are most common on the genitalia and in the perianal area, although they can occur anywhere on the skin surface. The onset of a fixed drug eruption can be sudden, developing within 30 minutes to 8 to 16 hours after ingestion of the medication. In patients who continue to take the offending drug, the number of eruption sites may gradually increase.[48]

After the initial acute phase, which lasts days to weeks, residual hyperpigmentation develops. Some patients may complain of burning or stinging on the affected skin sites. Systemic manifestations, which are present in approximately 25% of cases, can include fever, malaise, and abdominal symptoms.[48]

No conclusive diagnostic tests are available, but a challenge or provocation test with the suspected drug may be useful in confirming the diagnosis. Patch testing at the site of a previous lesion yields a positive response in up to 43% of patients. Prick and intradermal skin tests are reported to yield positive reactions in 24% and 67% of patients, respectively, but results vary with different drugs and reaction patterns. Patients with maculopapular rashes are more likely to have positive patch tests than patients with urticarial rashes.[49]

Treatment includes discontinuance of the causative agent and symptomatic therapy (e.g., topical corticosteroids).

Pseudoporphyria

Pseudoporphyria is a cutaneous phototoxic disorder that can resemble either porphyria cutanea tarda (PCT) or erythropoietic protoporphyria (EPP). Tetracycline, furosemide, and naproxen have been implicated in PCT- and EPP-pseudoporphyria.[50] The eruption may begin within 1 day after initiation of therapy or may be delayed for as long as 1 year. PCT-pseudoporphyria is characterized by skin fragility, blister formation, and scarring in areas exposed to sunlight; it occurs with normal

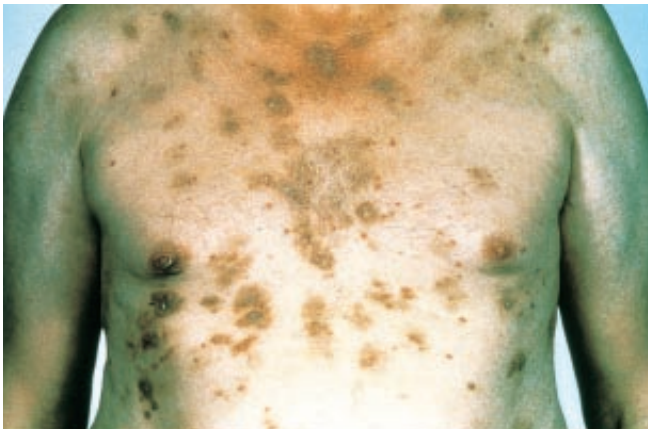


Figure 3 **Pemphigus foliaceus developed in this 64-year-old man taking enalapril.**

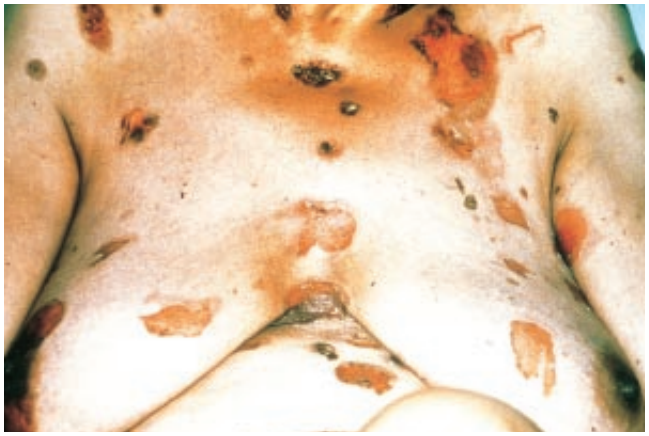


Figure 4 **Pemphigus vulgaris developed in this 59-year-old woman who took penicillamine as treatment for rheumatoid arthritis.**

porphyrin metabolism. The second clinical pattern mimics EPP and presents as cutaneous burning, erythema, vesiculation, angular scars, and waxy thickening of the skin.

Because of the risk of permanent facial scarring, the implicated drug should be discontinued when skin fragility, blistering, or scarring occurs. In addition, broad-spectrum sunscreen and protective clothing should be recommended to the patient.

COMPLEX ERUPTIONS

Drug-Induced Linear IgA Disease

Linear IgA disease is an autoimmune bullous dermatosis that is identified on the basis of the linear deposition of IgA at the basement membrane zone.[51] This disease can be induced by such drugs as vancomycin, lithium, diclofenac, and amiodarone. The drug-induced disease probably represents an immunologic response to the offending drug.

Drug-induced linear IgA disease is heterogeneous in clinical presentation. Cases have shown morphologies resembling erythema multiforme, bullous pemphigoid, and dermatitis herpetiformis. Drug-induced disease cannot be distinguished from the idiopathic variety either clinically, histologically, or immunologically; however, the clinical courses of these presentations differ. In drug-induced disease, spontaneous remission occurs once the offending agent is withdrawn; in idiopathic linear IgA disease, immune deposits disappear from the skin once the lesions resolve. Steroids and dapsone do not influence the healing process in drug-induced disease, whereas these agents have proved effective in treatment of idiopathic linear IgA disease.[52]

Drug-Induced Pemphigus

Pemphigus may be drug induced or drug triggered (i.e., the latent disease is unmasked by the drug exposure).

Drugs that cause pemphigus are penicillin, rifampin, phenylbutazone, propranolol, progesterone, piroxicam, interferon beta, interleukin-2, and levodopa.[53] An active amide group found in masked thiol drugs such as penicillin and cephalosporins and in nonthiol drugs such as enalapril may contribute to the pathogenesis of pemphigus.[53,54] Pemphigus foliaceus [*see Figure 3*] caused by penicillamine and other thiol drugs tends to resolve spontaneously in 35% to 50% of cases.[53] The average interval to onset is 1 year. Antinuclear antibodies are detected in 25% of affected patients.

Nonthiol drug–induced pemphigus manifests clinical, histologic, immunologic, and evolutionary aspects similar to those of idiopathic pemphigus vulgaris [*see Figure 4*]. Drug-induced pemphigus is associated with mucosal involvement. Spontaneous recovery after drug withdrawal occurs in 15% of affected patients.

Treatment of drug-induced pemphigus begins with drug withdrawal. Systemic corticosteroids are often required until all symptoms of active disease disappear. Vigilant follow-up is required after remission for an early relapse to be detected. The patient's serum should be monitored regularly for autoantibodies.[53]

Erythema Multiforme, Stevens-Johnson Syndrome, and Toxic Epidermal Necrolysis

The eruptions of erythema multiforme (EM), SJS, and TEN may represent variants of the same disease process. Reactions encompass a spectrum ranging from the less serious eruptions seen in EM to more serious reactions seen in SJS and TEN [*see Figure 5*].

A large percentage of EM and SJS cases are not drug related and may develop after a variety of predisposing factors, including infections, neoplasia, and autoimmune diseases. The drugs most frequently cited as causes of EM, SJS, and TEN are anticonvulsants, antibiotics (e.g., sulfonamides), allopurinol, and NSAIDs (e.g., piroxicam).[55] With anticonvulsants, risk appears to be greatest during the first 8 weeks of therapy.[56]

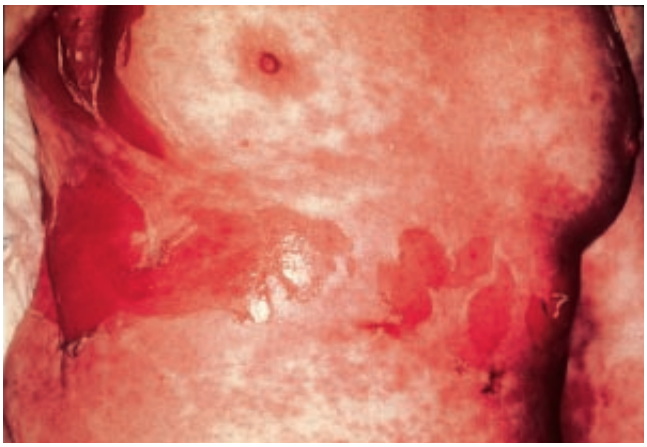


Figure 5 **This 50-year-old woman developed toxic epidermal necrolysis 17 days after starting phenytoin therapy.**

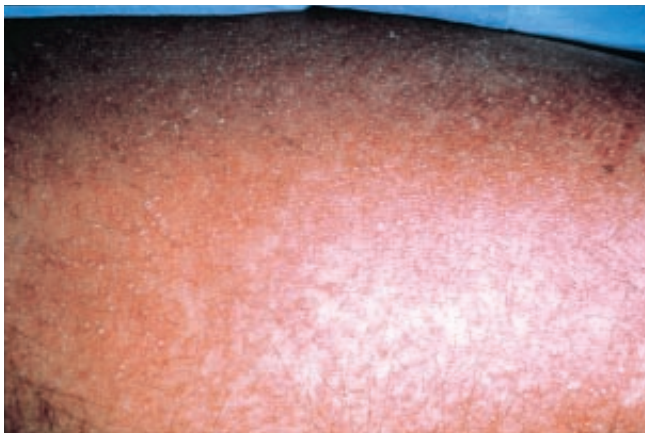


Figure 6 **Acute generalized exanthematous pustulosis (small nonfollicular pustules on a red base) in a 70-year-old man who took cloxacillin as treatment for cellulitis.**

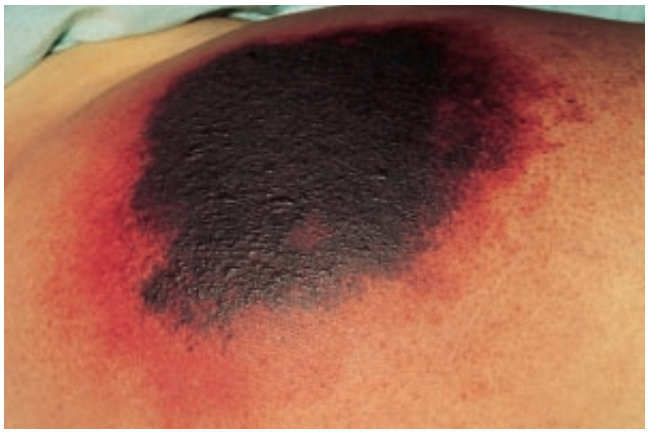


Figure 7 **Coumarin-induced skin necrosis in a 57-year-old woman who was given coumarin as treatment for atrial fibrillation.**

The pathogenesis of severe cutaneous ADRs is unknown, although a metabolic basis has been hypothesized. Sulfonamides and anticonvulsants, the two groups of drugs most frequently associated with SJS and TEN, are metabolized to toxic metabolites that are subsequently detoxified in most persons. However, in predisposed patients with a genetic defect, the metabolite may bind covalently to proteins. In some of these patients, the metabolite-protein adducts may trigger an immune response that leads to a cutaneous ADR.[57]

Clinically, the reaction patterns of EM, SJS, and TEN are characterized by the triad of mucous membrane erosions, target lesions, and epidermal necrosis with skin detachment. SJS is characterized by mucous membrane erosions and blisters on less than 10% of the total body surface area, whereas TEN involves more than 30% of the total body surface area.[58] The more severe the reaction, the more likely it is that it was drug-induced. Cases of severe cutaneous ADRs to lamotrigine (e.g., SJS and TEN) have been reported.[59] The prevalence of severe cutaneous ADRs associated with lamotrigine has been reported to be as high as one in 1,000 in adults and is higher in children. The risk is increased in the presence of valproic acid.

Complete blood counts, liver enzyme measurements, and chest x-rays should be performed to rule out concurrent internal organ involvement.

Treatment of EM, SJS, and TEN includes discontinuance of a suspected drug and such supportive measures as careful wound care, hydration, and nutritional support.[60] The use of corticosteroids in SJS and TEN is controversial.[61] Intravenous immunoglobulin (IVIg, 0.4 to 1.0 g/kg/day for 2 to 4 days), which contains naturally occurring Fas ligand (FasL)–blocking antibodies, has been shown in most reports to halt progression of TEN, especially when IVIg is started early.[62-64] Patients who have developed a severe cutaneous ADR (EM, SJS, or TEN) should not be rechallenged with the drug or undergo desensitization with the medication.

Pustular Eruptions

SIMPLE ERUPTIONS

Acneiform Eruptions

Eruptions morphologically mimicking acne vulgaris may be associated with drug ingestion. Iodides, bromides, adrenocorticotropic hormone, corticosteroids, isoniazid, androgens, lithium, dactinomycin, and phenytoin are reported to induce acne-like lesions.[65] Acne fulminans was induced by testosterone in 1% to 2% of adolescent boys who were treated for excessively tall stature.[66]

Drug-induced acne often appears on the face and back, but it may appear in atypical areas, such as arms and legs, and is usually monomorphous. Comedones are usually absent. Fever is absent. Acneiform eruptions do not affect prepubertal children, indicating that previous hormonal priming is a prerequisite. Topical tretinoin may be useful when the drug cannot be stopped.

COMPLEX ERUPTIONS

Acute Generalized Exanthematous Pustulosis

Acute generalized exanthematous pustulosis is characterized by acute onset, fever, and a cutaneous eruption with nonfollicular sterile pustules on an edematous erythema, generally starting within days of drug ingestion[67] [*see Figure 6*]; leukocy-

Table 2 Clinical Pearls to Identify Anticoagulant-Induced Skin Necrosis

	Interval to Onset	*Location*	*Other*
Coumarin-induced skin necrosis	3–5 days	Adipose-rich sites	—
Heparin-induced thrombocytopenia and thrombosis	4–14 days	Extremities	Thrombocytopenia occurs concurrently
Purple-toe syndrome	3–8 wk	Acral location	Often occurs after angiography

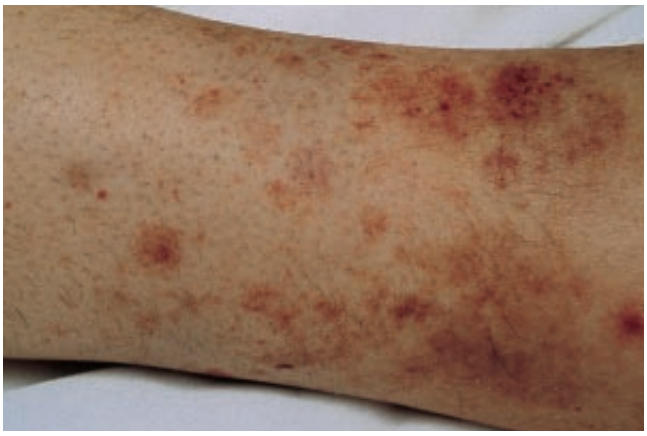


Figure 8 **Leukocytoclastic vasculitis developed in this 47-year-old woman taking hydrochlorothiazide.**

tosis is another common finding. Generalized desquamation occurs 2 weeks later. Differential diagnosis includes pustular psoriasis, subcorneal pustular dermatosis (Sneddon-Wilkinson disease), hypersensitivity syndrome reaction with pustulation, and pustular eruptions of infancy.

Acute generalized exanthematous pustulosis is most commonly associated with β-lactam and macrolide antibiotic usage. Many other drugs have been implicated, however, including calcium channel blockers and analgesics. The estimated incidence rate is approximately one to five cases per million per year.[68] Discontinuance of therapy is usually the extent of treatment necessary in most patients, although some patients may require the use of corticosteroids. Patch testing to the putative drug is often positive, resulting in a localized pustular reaction.

Other Eruptions

ANTICOAGULANT-INDUCED SKIN NECROSIS

Anticoagulant drugs may induce hypercoagulable states with subsequent vascular infarction and cutaneous necrosis [*see Figure 7*]. Both coumarin and heparin can induce skin necrosis. Clinical pearls that can help differentiate these reactions are the location, timing, platelet count, and primary diagnosis [*see Table 2*].

The pathogenesis of coumarin-induced skin necrosis is the paradoxical development of occlusive thrombi in cutaneous and subcutaneous venules caused by a transient hypercoagulable state. This condition results from the suppression of the natural anticoagulant protein C at a greater rate than natural procoagulant factors. Coumarin-induced skin necrosis is associated with protein C and protein S deficiency, but pretreatment screening is not warranted. An association with a heterozygote for the factor V Leiden mutation has been recently reported.[69]

It is estimated that one in 10,000 persons who take coumarin are at risk for this adverse event.[70] The prevalence is four times higher in women than in men. In both sexes, the peak incidence occurs in the sixth and seventh decades of life. Afflicted patients tend to be obese.

Coumarin-induced skin necrosis begins 3 to 5 days after initiation of treatment. Painful red plaques develop in adipose-rich sites such as breasts, buttocks, and hips. These plaques may blister, ulcerate, or develop into necrotic areas. An accompanying infection, such as pneumonia, viral infection, or erysipelas, may occur in as many as 25% of patients. Purple-toe syndrome occurs 3 to 8 weeks after initiation of coumarin therapy.

Treatment entails the discontinuance of coumarin, administration of vitamin K, and infusion of heparin at therapeutic doses. Fresh frozen plasma and purified protein C concentrates have been used.[71] Supportive measures for the skin are recommended. Plastic surgery for remediation is necessary in 60% of affected patients.

DRUG-INDUCED LICHENOID ERUPTIONS

Drug-induced lichen planus produces lesions that are clinically and histologically indistinguishable from those of idiopathic lichen planus. Many drugs, including beta blockers, penicillamine, NSAIDs, gold, and ACE inhibitors, especially captopril, have been reported to produce this reaction.

The latent period between the start of administration of the drug and appearance of the eruption is variable. The mean latent period is between 2 months and 3 years for penicillamine, approximately 1 year for beta-adrenergic blocking agents, and 3 to 6 months for ACE inhibitors. The latent period may be shorter if the patient was previously exposed to the drug.[72] In general, resolution usually occurs within 2 to 4 months.

Rechallenge with the culprit drug has been attempted in a few patients, with reactivation of symptoms within 4 to 15 days.[73] Patch testing has not proved helpful in most cases of drug-induced lichen planus. However, results of patch tests performed with contact inducers of lichen drug eruptions (e.g., color-film developers and dental restorative materials) are usually positive.[72]

DRUG-INDUCED VASCULITIS

Drug-induced vasculitis represents approximately 10% of the acute cutaneous vasculitides and usually affects small vessels [*see Figure 8*].[74] Drug-induced vasculitis should be considered in any patient with small vessel vasculitis that is usually confined to the skin.[75] Drugs that are most frequently associated with vasculitis include propylthiouracil, hydralazine, granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage CSF (GM-CSF), allopurinol, cefaclor, minocycline, penicillamine, phenytoin, and isotretinoin.[73] The average interval to onset of drug-induced vasculitis is 7 to 21 days.[76]

The clinical hallmark of cutaneous vasculitis is palpable purpura, classically found on the lower extremities, although any cutaneous site may be affected. Urticaria can be a manifestation of small vessel vasculitis. Unlike nonvasculitic allergic urticaria, vasculitic urticaria lasts longer than 1 day, may evolve into purpuric lesions, and may be accompanied by hypocomplementemia.[77] Other features are hemorrhagic bullae, urticaria, ulcers, nodules, Raynaud disease, and digital necrosis. The same vasculitic process may also affect internal organs, such as the liver, kidney, gut, and CNS, and is potentially life threatening.

Histologically, the small blood vessels of the dermis display fibrinoid necrosis, polymorphonuclear infiltration into the blood vessel wall, extravasation of red blood cells, and nuclear dust. Direct immunofluorescence may show deposits of IgM and C3 in the blood vessel walls. Therefore, these reactions are immune complex–dependent drug reactions. The immune complexes may be composed of antibodies directed against drug-related haptens, but this has not been proved.

Drug-induced vasculitis can be difficult to diagnose, and diagnosis is often one of exclusion.[78] Alternative causes of cuta-

neous vasculitis, such as infection or autoimmune disease, must be eliminated.

Treatment consists of drug withdrawal.Therapy for patients with severe manifestations includes hemodialysis, pulse corticosteroids, cyclophosphamide, and plasmapheresis.[73]

Neil H. Shear, M.D., has received grants for clinical research and educational purposes and served as advisor for Roche, Galderma SA, GlaxoSmithKline, Novartis AG, and Fujisawa Healthcare Inc. He has also received grants for clinical research and served as advisor for Genesoft Co. Ltd.

Sandra Knowles, B.Sc. Pharm., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Lori Shapiro, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Karch FE, Lasagna L: Adverse drug reactions: a critical review. JAMA 234:1236, 1975

2. Lazarou J, Pomeranz BH, Corey PN: Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 279:1200, 1998

3. Lakshmanan M, Hershey C, Breslau D: Hospital admissions caused by iatrogenic disease. Arch Intern Med 146:1391, 1986

4. Bigby M, Jick S, Jick H, et al: Drug-induced cutaneous reactions: a report from the Boston Collaborative Drug Surveillance Program on 15,438 consecutive inpatients, 1975 to 1982. JAMA 256:3358, 1986

5. Leape LL, Brennan TA, Laird N, et al: The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. N Engl J Med 324:377, 1991

6. Classen DC, Pestotnik SL, Evans RS, et al: Adverse drug events in hospitalized patients: excess length of stay, extra costs and attributable mortality. JAMA 277:301, 1997

7. Shear NH: Diagnosing cutaneous adverse reactions to drugs. Arch Dermatol 126:94, 1990

8. Sogn DD, Evans R, Shepherd GM, et al: Results of the National Institute of Allergy and Infectious Diseases Collaborative Clinical Trial to test the predictive value of skin testing with major and minor penicillin derivatives in hospitalized adults. Arch Intern Med 152:1025, 1992

9. Barbaud A, Goncalo M, Bruynzeel D, et al: Guidelines for performing skin tests with drugs in the investigation of cutaneous adverse drug reactions. Contact Derm 45:321, 2001

10. Alanko K, Stubb S, Reitamo S: Topical provocation of fixed drug eruption. Br J Dermatol 116:561, 1987

11. Smith HR, Croft AM, Black MM: Dermatological adverse effects with the antimalarial drug mefloquine: a review of 74 published case reports. Clin Exp Dermatol 24:249, 1999

12. Vega JM, Blanca M, Carmona MJ, et al: Delayed allergic reactions to beta-lactams. Allergy 46:154, 1991

13. Romano A, Quaratino D, Papa G, et al: Aminopenicillin allergy. Arch Dis Child 76:513, 1997

14. Coopman S, Johnson R, Platt R, et al: Cutaneous disease and drug reactions in HIV infection. N Engl J Med 328:1670, 1993

15. Pichler W, Yawalkar N, Schmid S, et al: Pathogenesis of drug-induced exanthems. Allergy 57:884, 2002

16. Nigen S, Knowles SR, Shear NH: Drug eruptions: approaching the diagnosis of drug-induced skin diseases. J Drugs Dermatol 2:278, 2003

17. Prussick R, Knowles S, Shear N: Cutaneous drug reactions. Curr Probl Dermatol 6:81, 1994

18. Knowles SR, Shapiro L, Shear NH: Serious adverse reactions induced by minocycline: a report of 13 patients and review of the literature. Arch Dermatol 132:934, 1996

19. Shear NH, Spielberg SP: Anticonvulsant hypersensitivity syndrome: in vitro assessment of risk. J Clin Invest 82:1826, 1988

20. Calabrese JR, Sullivan JR, Bowden CL: Rash in multicenter trials of lamotrigine in mood disorders: clinical relevance and management. J Clin Psychiatry 63:1010, 2002

21. Wadelius M, Karlsson T, Wadelius C, et al: Lamotrigine and toxic epidermal necrolysis. Lancet 348:1041, 1996

22. Shear NH, Spielberg SP, Grant DM, et al: Differences in metabolism of sulfonamides predisposing to idiosyncratic toxicity. Ann Intern Med 105:179, 1986

23. Gupta A, Eggo M, Uetrecht J, et al: Drug-induced hypothyroidism: the thyroid as a target organ in hypersensitivity reactions to anticonvulsants and sulfonamides. Clin Pharmacol Ther 51:56, 1992

24. Rieder MJ: In vivo and in vitro testing for adverse drug reactions. Pediatr Clin North Am 44:93, 1997

25. Anderson J: Allergic reactions to drugs and biologic agents. JAMA 268:2845, 1992

26. Simon RA, Namazy J: Adverse reactions to aspirin and nonsteroidal anti-inflammatory drugs (NSAIDs). Clin Rev Allergy Immunol 24:239, 2003

27. Fisher MM, Harle DG, Baldo BA: Anaphylactoid reactions to narcotic analgesics. Clin Rev Allergy 9:309, 1991

28. Oudit G, Girgrah N, Allard J: ACE inhibitor-induced angioedema of the intestine: case report, incidence, pathophysiology, diagnosis and management. Can J Gastroenterol 15:827, 2001

29. Archambault S, Malo JL, Infante-Rivard C, et al: Incidence of sensitization, symptoms, and probably occupational rhinoconjunctivitis and asthma in apprentices starting exposure to latex. J Allergy Clin Immunol 107:921, 2001

30. Mazon A, Nieto A, Linana JJ, et al: Latex sensitization in children with spina bifida: follow-up comparative study after two years. Ann Allergy Asthma Immunol 84:207, 2000

31. Turjanmaa K, Alenius H, Reunala T, et al: Recent developments in latex allergy. Curr Opin Allergy Clin Immunol 2:407, 2002

32. Pavletic AJ: Late angioedema in patients taking angiotensin-converting-enzyme inhibitors. Lancet 360:493, 2002

33. deShazo RD, Mather P, Grant W, et al: Evaluation of patients with local reactions to insulin with skin tests and in vitro techniques. Diabetes Care 10:330, 1987

34. Dykewicz M, Kim HW, Orfan N, et al: Immunologic analysis of anaphylaxis to protamine component in neutral protamine Hagedorn human insulin. J Allergy Clin Immunol 93:117, 1994

35. Hoffman HM, Wright FA, Broide DH, et al: Identification of a locus on chromosome 1q44 for familial cold urticaria. Am J Hum Genet 66:1693, 2000

36. Haas N, Toppe E, Henz BE: Microscopic morphology of different types of urticaria. Arch Dermatol 134:41, 1998

37. Erbagci Z: The leukotriene receptor antagonist montelukast in the treatment of chronic idiopathic urticaria: a single-blind, placebo-controlled, cross-over clinical study. J Allergy Clin Immunol 110:484, 2002

38. Ilter N, Gurer MA, Akkoca MA: Short-term oral cyclosporine for chronic idiopathic urticaria. J Eur Acad Dermatol Venereol 12:67, 1999

39. Heckbert SR, Stryker WS, Coltin KL, et al: Serum sickness in children after antibiotic exposure: estimates of occurrence and morbidity in a health maintenance organization population. Am J Epidemiol 132:336, 1990

40. Joubert GI, Hada K, Matsui D, et al: Selection of treatment of cefaclor-associated urticarial, serum sickness–like reactions and erythema multiforme by emergency pediatricians: lack of a uniform standard of care. Can J Clin Pharmacol 6:197, 1999

41. Kearns GL, Wheeler JG, Childress SH, et al: Serum sickness–like reactions to cefaclor: role of hepatic metabolism and individual susceptibility. J Pediatr 125:805, 1994

42. Lowery N, Kearns GL, Young RA, et al: Serum sickness–like reactions associated with cefprozil therapy. J Pediatr 125:325, 1994

43. McCollom RA, Elbe DH, Ritchie AH: Bupropion-induced serum sickness–like reaction. Ann Pharmacother 34:471, 2000

44. Shapiro LE, Knowles SR, Shear NH: Comparative safety and risk management of tetracycline, doxycycline and minocycline. Arch Dermatol 133:1224, 1997

45. Vial T, Pont J, Pham E, et al: Cefaclor-associated serum sickness–like disease: eight cases and review of the literature. Ann Pharmacother 26:910, 1992

46. Grammer LC: Cefaclor serum sickness. JAMA 275:1152, 1996

47. Lee AY: Fixed drug eruptions: incidence, recognition and avoidance. Am J Clin Dermatol 1:277, 2000

48. Ozkaya-Bayazit E, Akar U: Fixed drug eruption induced by trimethoprim-sulfamethoxazole: evidence for a link to HLA-A30 B13 Cw6 haplotype. J Am Acad Dermatol 45:712, 2001

49. Barbaud A, Reichert-Penetrat S, Trechot P, et al: The use of skin testing in the investigation of cutaneous adverse drug reactions. Br J Dermatol 139:49, 1998

50. Al-Khenaizan S, Schecter JF, Sasseville D: Pseudoporphyria induced by propionic acid derivatives. J Cutan Med Surg 3:162, 1999

51. Kuechle MK, Stegemeir E, Maynard B, et al: Drug-induced linear IgA bullous dermatosis: report of six cases and review of the literature. J Am Acad Dermatol 30:187, 1994

52. Neughebauer BI, Negron G, Pelton S, et al: Bullous skin disease: an unusual allergic reaction to vancomycin. Am J Med Sci 323:273, 2002

53. Brenner S, Bialy-Gohan A, Ruocco V: Drug-induced pemphigus. Clin Dermatol 16:393, 1998

54. Wolf R, Brenner S: An active amide group in the molecule of drugs that induce pemphigus: a casual or causal relationship? Dermatology 189:1, 1994

55. Roujeau JC, Kelly JP, Naldi L, et al: Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. N Engl J Med 333:1600, 1995

56. Rzany B, Correia O, Kelly JP, et al: Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis during first weeks of antiepileptic therapy: a case-control study. Study Group of the International Case Control Study on Severe Cutaneous Adverse Reactions. Lancet 353:2190, 1999

57. Wolkenstein P, Charue D, Bagot M, et al: Metabolic predisposition to cutaneous adverse drug reactions. Arch Dermatol 131:544, 1995

58. Bastuji-Garin S, Rzany B, Stern RS, et al: Clinical classification of cases of toxic epidermal necrolysis, Stevens-Johnson syndrome, and erythema multiforme. Arch Dermatol 129:92, 1993

59. Schlienger RG, Shapiro LE, Shear NH: Lamotrigine-induced severe cutaneous adverse reactions. Epilepsia 39:S22, 1998

60. Garcia-Doval I, LeCleach L, Bocquet H, et al: Toxic epidermal necrolysis and Stevens-Johnson syndrome: does early withdrawal of causative drugs decrease the risk of death? Arch Dermatol 136:323, 2000

61. Patterson R, Miller M, Kaplan M, et al: Effectiveness of early therapy with corticosteroids in Stevens-Johnson syndrome: experience with 41 cases and a hypothesis regarding pathogenesis. Ann Allergy 73:27, 1994

62. Prins C, Kerdel FA, Padilla RS, et al: Treatment of toxic epidermal necrolysis with high-dose intravenous immunoglobulins. Arch Dermatol 139:26, 2003

63. Trent JT, Kirsner RS, Romanelli P, et al: Analysis of intravenous immunoglobulin for the treatment of toxic epidermal necrolysis using SCORTEN. Arch Dermatol 139:39, 2003

64. Bachot N, Revuz J, Roujeau JC: Intravenous immunoglobulin treatment for Stevens-Johnson syndrome and toxic epidermal necrolysis. Arch Dermatol 139:33, 2003

65. Remmer H, Falk W: Successful treatment of lithium-induced acne. J Clin Psychiatry 47:48, 1986

66. Traupe H, von Muhlendahl K, Bramswig J, et al: Acne of the fulminans type following testosterone therapy in three excessively tall boys. Arch Dermatol 124:414, 1988

67. Beylot C, Doutre MS, Beylot-Barry M: Acute generalized exanthematous pustulosis. Semin Cutaneous Med Surg 15:244, 1996

68. Sidoroff A, Halevy S, Bavinck JN, et al: Acute generalized exanthematous pustulosis (AGEP): a clinical reaction pattern. J Cutan Pathol 28:113, 2001

69. Freeman BD, Schmieg RE, McGrath S, et al: Factor V Leiden mutation in a patient with warfarin-associated skin necrosis. Surgery 127:595, 2000

70. Bauer KA: Coumarin-induced skin necrosis. Arch Dermatol 129:766, 1993

71. Schramm W, Spannagel M, Bauer KA, et al: Treatment of coumarin-induced skin necrosis with a monoclonal antibody purified protein C concentrate. Arch Dermatol 19:753, 1993

72. Thompson DF, Skaehill A: Drug-induced lichen planus. Pharmacotherapy 14:561, 1994

73. ten Holder SM, Joy MS, Falk RJ: Cutaneous and systemic manifestations of drug-induced vasculitis. Ann Pharmcother 36:130, 2002

74. Sanchez NP, Van Hale HM, Su WP: Clinical and histopathologic spectrum of necrotizing vasculitis: reports of findings in 101 cases. Arch Dermatol 121:220, 1985

75. Jennette J, Falk K: Small-vessel vasculitis. N Engl J Med 337:1512, 1997

76. Merkel PA: Drugs associated with vasculitis. Curr Opin Rheumatol 10:45, 1998

77. Mehregan D, Hall M, Gibson E: Urticarial vasculitis: a histopathologic and clinical review of 72 cases. J Am Acad Dermatol 26:441, 1992

78. Wolkenstein P, Revuz J: Drug-induced severe skin reactions: incidence, management and prevention. Drug Safety 13:56, 1995

VII FUNGAL, BACTERIAL, AND VIRAL INFECTIONS OF THE SKIN

Jan V. Hirschmann, m.d.

Despite its large surface area and constant exposure to the environment, the skin resists infection well. The most important protective factor is an intact stratum corneum, the tough barrier of protein and lipid formed on the cutaneous surface by the underlying epidermis.[1] This barricade impedes invasion by environmental pathogens, and its dryness discourages colonization and growth of the many organisms that require moisture to survive, such as gram-negative bacilli. Furthermore, the constant shedding of cells of the epidermis impedes most microbes from establishing permanent residence.

Some organisms, however, can attach to skin cells and reproduce there; the normal cutaneous flora comprises primarily aerobic, gram-positive cocci and bacilli in densities ranging from about 10^2 organisms/cm^2 on dry skin to 10^7 organisms/cm^2 in moist areas, such as the axilla.[2] This resident population inhibits harmful organisms from colonizing the skin by occupying binding sites on the epidermal cells, competing for nutrients, producing antimicrobial substances, and maintaining the skin surface at a low pH (about 5.5). Anaerobes are sparse except in areas with abundant sebaceous glands, such as the face and chest; in the deeper portions of these sites, as well as in hair follicles, anaerobes reach concentrations of 10^4 to 10^6 organisms/cm^2.

Cutaneous infections occur when the skin's protective mechanisms fail, especially when trauma, inflammation, maceration from excessive moisture, or other factors disrupt the stratum corneum. The organisms causing infection may originate from the victim's own resident flora, either on the skin or on adjacent mucous membranes, but many come from other people, animals, or the environment.

Dermatophyte Infections

Dermatophytes are fungi (molds) that can infect the skin, hair, and nails. These organisms, which include *Trichophyton*, *Microsporum*, and *Epidermophyton* species, are classified as anthropophilic, zoophilic, or geophilic, depending on whether their primary source is humans, animals, or the soil, respectively.[3] Geophilic dermatophyte infections occur sporadically, primarily among gardeners and farm workers. Zoophilic dermatophytes (*Trichophyton* and *Microsporum* species) may have a restricted range of hosts (e.g., *M. persicolor* infects only voles) or may afflict many different animals (e.g., *T. mentagrophytes* can infect mice and other rodents, dogs, cats, and horses). Human infections with zoophilic species have occurred after exposure to dogs, cats, horses, cattle, pigs, rodents, poultry, hedgehogs, and voles.

Anthropophilic dermatophytes are the most common cause of fungal skin infections in humans. Transmission of these infections occurs from direct contact between people or from exposure to desquamated skin cells present in the environment—arthrospores can survive for months. Direct inoculation of the spores through breaks in the skin can lead to germination and subsequent invasion of the superficial cutaneous layers.

Dermatophyte infections occur more frequently in certain ethnic groups and in people with impaired cell-mediated immunity. Many of the anthropophilic dermatophyte infections occur more often in one gender or age group.[4]

Infection of the scalp, for example, is primarily a disease of children. Involvement of the feet and groin is most common in adolescents and young adults, especially males, but is unusual in children. Nail infection is more frequent in both men and women of advancing age. The reasons for these differences are unknown.

The anthropophilic dermatophytes also have unique geographic distribution patterns. The most common cause of scalp infection in the United States, for example, is *T. tonsurans*, but in Southeast Asia and the Middle East, it is *T. violaceum*. These differences may relate to climatic or racial factors.

The various forms of dermatophytosis, also called ringworm, are named according to the site involved. These infections include tinea capitis (scalp), tinea corporis (body), tinea barbae (beard area of men), tinea faciei (face), tinea cruris (groin), tinea pedis (feet), tinea unguium (nails), and tinea manuum (hands). The characteristic skin lesion is an annular scaly patch [*see Figure 1*], though the clinical appearance varies not only with the site involved but also with the host's immune status and the type of infecting organism. In general, anthropophilic species elicit little inflammation and cause chronic infections. Zoophilic and geophilic species, however, often provoke intense inflammation, which sometimes leads to eradication of the organisms and healing without treatment.

Clinical Presentations

Tinea Capitis

Tinea capitis occurs primarily in children but may develop in adults—especially the elderly, those who are unkempt, and the impoverished. Transmission can occur between humans by the sharing of combs, brushes, or headgear. Only *Microsporum* and *Trichophyton* species cause tinea capitis. Infection begins with invasion of the stratum corneum of the scalp skin. The hairs then become infected, in one of three microscopic patterns: ectothrix, endothrix, or favus. In ectothrix, the spores are

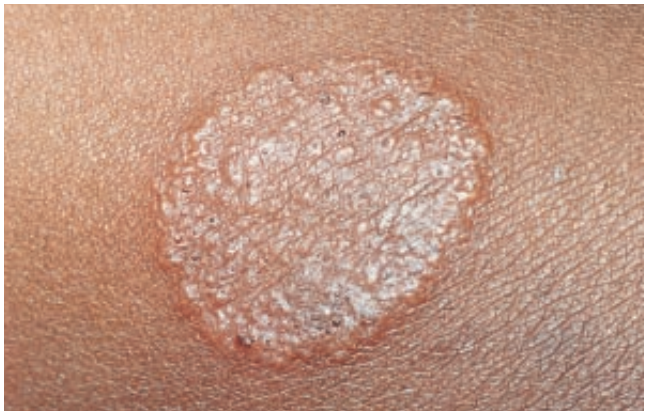


Figure 1 **Classic annular lesion of tinea corporis shows a raised or vesicular margin with central clearing.**

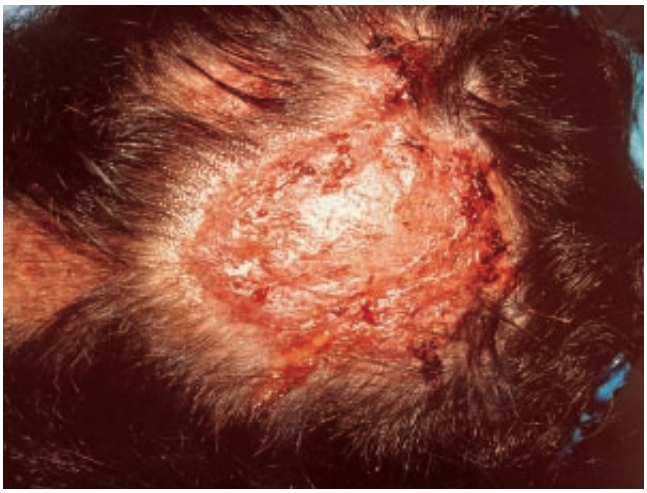


Figure 2 **A typical kerion presenting as a zoophilic *Microsporum canis* infection of the scalp (tinea capitis).**

outside the hair shaft and destroy the cuticle; in endothrix, they lie within the hair and do not affect the cuticle; and in favus, broad hyphae and air spaces form within the hair, but spores are absent. In all three types, scaling, hair loss, and inflammation of varying degrees are present.[5]

T. tonsurans, the major cause of tinea capitis in adults, characteristically produces a noninflammatory infection with either well-demarcated or irregular and diffuse areas of scaling and alopecia. Because the swollen hairs may fracture a few millimeters from the epidermis in this endothrix infection, the scalp sometimes appears to be marked by small black dots. Like all infections with *Trichophyton* species, these scalp lesions do not fluoresce under a Wood light.

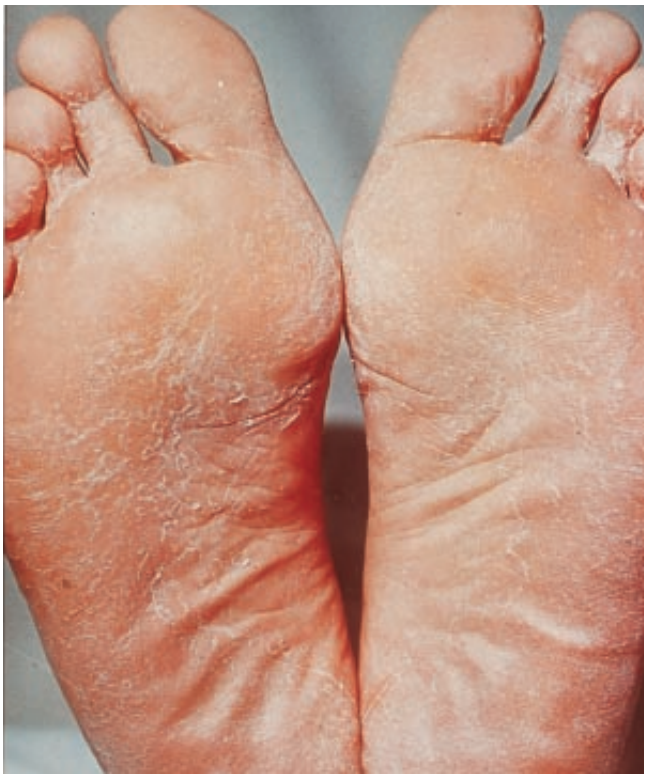
T. schoenleinii causes favus, characterized by an inflammatory crust (scutulum) in which hair appears to be matted in the dried, yellow exudate. Hair shedding late in the infection is common because the hair shaft is not damaged until the infection is well advanced.

M. audouinii, which causes an ectothrix infection, produces well-delineated, noninflammatory patches of alopecia in which the hair breaks at the epidermal surface and is often dull gray because of the presence of numerous spores on the surface of the hair shaft. As in all *Microsporum* infections, these lesions fluoresce under a Wood light. The most severe inflammation, usually from a zoophilic species, results in a kerion, a painful, boggy mass in which follicles may discharge pus and in which sinus tracts form [*see Figure 2*]. Crusting and matting of adjacent hairs are common, and cervical lymph nodes may enlarge.

Tinea Corporis

Tinea corporis typically appears as a single lesion or multiple circular lesions with scaling, well-delineated margins, and a raised, erythematous edge. Often, they have an area of central clearing. The amount of inflammation varies; when the inflammation is intense, pustules, vesicles, and even bullae may occur. Sometimes, involvement of the hair follicles in the middle of a patch of scaling erythema leads to perifollicular nodules, a condition called Majocchi granuloma. This condition usually occurs on the legs of patients infected with *T. rubrum*.

a



b

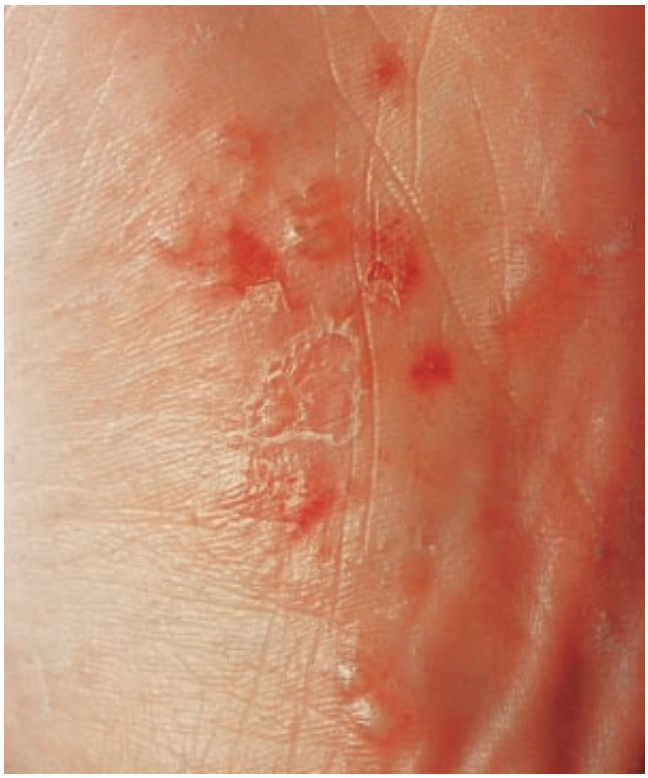


Figure 3 **(*a*) The scaling of tinea pedis appears between and under the toes and on the plantar surface. (*b*) Tinea pedis may also present as vesicles.**

In immunocompromised hosts, subcutaneous abscesses may develop.

Tinea Barbae

Tinea barbae occurs in adult men and involves the skin and coarse hairs of the beard and mustache area. The usual cause is a zoophilic species, primarily *T. verrucosum* and *T. mentagrophytes*, which are organisms that commonly infect cattle and horses. The victims are generally farm workers, and the infection usually causes erythema, scaling, and follicular pustules. Many hairs become loose and are easily removed with a forceps.

Tinea Faciei

Tinea faciei occurs as an infection of the face in women and children and infection of the area outside the mustache and beard in men. The usual causes are *T. rubrum* and *T. mentagrophytes*; these organisms reach the face through direct inoculation or by spreading from another site of infection on the body. Patients often complain of itching and burning, and symptoms may worsen after exposure to sunlight. The lesions may be scaly, annular erythematous patches, but often they are indistinct red areas with little or no scaling.

Tinea Cruris

Tinea cruris, infection of the groin, is much more common in men than women and is often associated with infection of the feet. *T. rubrum* and *E. floccosum* are the most common causes. The lesions are usually red, scaling, sharply demarcated areas with raised, erythematous borders. The infection, which affects the medial portion of the upper thighs but consistently spares the scrotum, may extend to the buttocks, abdomen, and lower back. Vesicles, nodules, pustules, and maceration may be present.

Tinea Pedis

Tinea pedis is most frequently caused by *T. rubrum*, *E. floccosum*, and *T. mentagrophytes*. The most common form consists of fissuring, scaling, and maceration in the interdigital spaces, especially between the fourth and fifth toes. A second type involves scaling, hyperkeratosis, and erythema of the soles, heels, and sides of the feet. In this kind of tinea pedis, the lesions occur in a so-called moccasin distribution pattern [*see Figure 3a*]. The plantar skin may become very thick and scaly. A third form demonstrates an inflammatory pattern characterized by vesicles, pustules, or even bullae, usually on the soles [*see Figure 3b*].

An important complication of tinea pedis is streptococcal cellulitis of the lower leg. Streptococci do not ordinarily survive on normal skin, but the presence of fungal disease apparently permits streptococci of various groups, including A, B, C, and G, to colonize the toe webs.[6] From this location, these bacteria may invade the skin damaged by the tinea pedis or migrate to locations higher up the leg and enter the skin through any defects.

Tinea Unguium

Nail involvement usually occurs from adjacent fungal infection of the hands or feet. The organisms typically invade the nail from the distal or lateral borders, and infection spreads proximally. The nails are thickened, opaque, and yellowish to brownish. They may crack or crumble, and often, subungual hyperkeratosis lifts the nail plate from the underlying bed (a condition known as onycholysis) [*see Figure 4*]. Splinter hemorrhages are common.

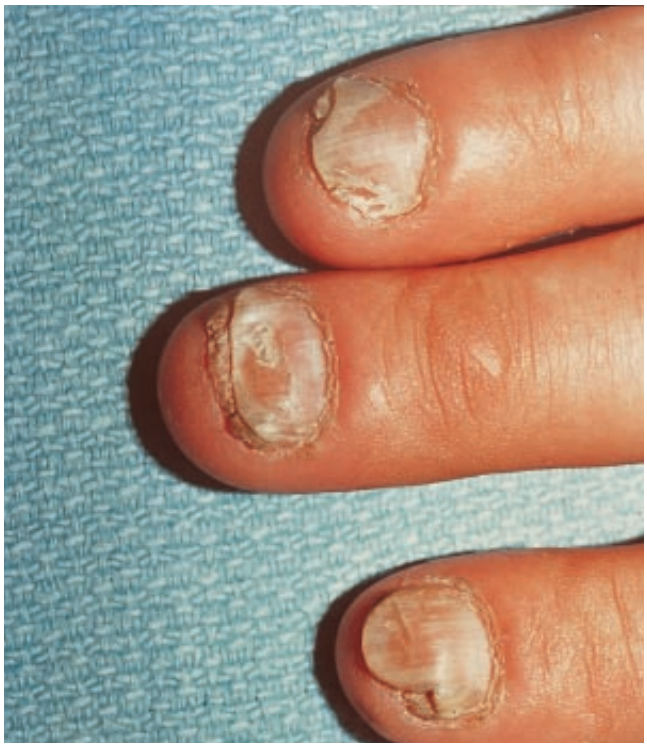


Figure 4 **Nails are usually thickened, cracked, and crumbly in tinea unguium; subungual debris may be present, as shown.**

Tinea Manuum

Tinea manuum is an infection of the hands. Most cases have accompanying involvement of the feet; inexplicably, usually only one hand is affected (so-called two-feet, one-hand disease). The most common finding is scaling or hyperkeratosis of the palms and fingers. Occasionally, vesicles, papules, or follicular nodules form on the dorsal surface of the hands.

DIAGNOSIS

Clinicians should suspect dermatophyte infection in patients with any scaling, erythematous eruption and in patients whose nails exhibit the characteristics of tinea unguium (see above). The diagnosis can be confirmed by microscopy or culture of properly obtained specimens. The optimal method of obtaining specimens from the skin is by scraping the scaly lesions; specimens from the nails are best obtained by taking fragments of subungual debris.

The specimen is prepared for microscopic examination by first placing it on a glass slide and treating it with potassium hydroxide (KOH), which digests the keratin of the skin, nails, and hair, and then heating it to hasten the process. The basic culture medium for isolating dermatophytes is an agar containing Sabouraud medium, often combined with antibiotics to eliminate bacteria and with cycloheximide to inhibit saprophytic fungi. Growth is usually apparent in 3 to 14 days. Dermatophyte test medium culture can be used in the office and is both accurate and inexpensive.[7] When both KOH preparations and cultures are negative, a biopsy may be useful in identifying the infecting organism, usually by special tissue stains such as periodic acid–Schiff or Gomori methenamine-silver stains.

TREATMENT

Tinea corporis, tinea cruris, tinea pedis, and tinea faciei respond to topical agents applied once or twice daily to the affected area, usually for 2 to 4 weeks. Good choices include azoles (e.g., miconazole, econazole, or clotrimazole) or terbinafine. The cost of the preparation can dictate which agent to prescribe. Tinea pedis often recurs after effective therapy, especially in cases of the moccasin form of the disease. When infection reappears, the previous therapy can be resumed without loss of effectiveness.

Oral therapy is necessary for extensive lesions, for infection involving the hair or hair follicles (e.g., tinea capitis and tinea barbae), for tinea unguium, and, often, for tinea manuum and various forms of dermatophytoses in immunocompromised hosts. Five oral agents are currently available: griseofulvin, ketoconazole, itraconazole, fluconazole, and terbinafine. Griseofulvin, a fungistatic agent, is the oldest oral treatment available and is still useful, primarily in infections not involving the nails. Griseofulvin reduces the serum levels of barbiturates and warfarin. Some patients receiving griseofulvin note a diminished tolerance to alcohol.

The azoles include ketoconazole, itraconazole, and fluconazole; like griseofulvin, they are fungistatic. Ketoconazole is usually well tolerated, but hepatotoxicity occurs in about 1 in 10,000 patients, typically after several weeks of use. Fluconazole and itraconazole are very expensive, but they provide protracted levels of antibiotic in the nails, allowing short or intermittent courses of therapy for tinea unguium. Both fluconazole and itraconazole can cause gastrointestinal disorders, rashes, and, occasionally, hepatotoxicity and can have serious interactions with several medications, including cyclosporine, digoxin, and quinidine. Ketoconazole, itraconazole, and fluconazole can interact with other medications; pharmacologic sources should be consulted for potential interactions.

Terbinafine, also an expensive medication, is an allylamine. Unlike both griseofulvin and the azoles, which are fungistatic, terbinafine is fungicidal. It achieves high levels of drug in the nails, and the drug persists for many weeks after discontinuance. Its few side effects include gastrointestinal reactions and, occasionally, skin rashes. Hepatotoxicity and hematologic abnormalities are rare, and drug interactions are uncommon.

These oral antifungals are quite effective for tinea capitis. The adult dosage for griseofulvin is 500 mg twice daily for 8 weeks. The other agents are effective when given for 1 to 3 weeks. Daily doses are as follows: itraconazole, 200 mg; fluconazole, 200 mg; and terbinafine, 250 mg. Of these, griseofulvin is the least expensive, but some *T. tonsurans* isolates are resistant to it. All these medications are effective in cases of tinea barbae, Majocchi granuloma, extensive tinea corporis, and tinea manuum that are unresponsive to topical agents. Griseofulvin and terbinafine appear to be superior to fluconazole and itraconazole for the treatment of tinea capitis.[8]

Tinea unguium is difficult to eradicate, particularly in the toenails. The most effective agent is terbinafine, administered at a dosage of 250 mg daily for 6 weeks for fingernail infections and for 12 weeks for toenail involvement.[9] Because terbinafine persists in the nails for many weeks, it continues to exert antifungal effects long after it is discontinued. The terbinafine regimens produce short-term eradication of infection in about 70% to 90% of patients with fingernail infection and in about 50% to 80% of patients with toenail infection. Relapse is common, and patients often require a second course of treatment. About 75% of patients who receive one or more courses of terbinafine will have a clinical cure 5 years later. This therapy is very expensive, and clinicians must decide in each case whether treatment is warranted.

Yeast Infections

Yeasts are unicellular fungi that reproduce by budding. They may form filamentous projections, which, unlike the hyphae of molds, do not contain separate cells. Accordingly, they are called pseudohyphae. *Candida* species are not part of the normal skin flora, but they commonly reside in the oropharynx, vagina, and colon. From these locations, they may cause infections in adjacent traumatized skin. Alternatively, with reduction in the other flora or with impaired host defense mechanisms, these yeasts may proliferate in large numbers to produce lesions on the mucosal surfaces of the mouth and vagina.

Malassezia furfur (also called *Pityrosporum orbiculare* or *P. ovale*) is a yeast that requires lipids for growth. It normally colonizes the skin of adults, especially of the scalp and upper trunk, where the presence of sebum is highest. For unknown reasons, these organisms, which are ordinarily commensals, can become pathogenic and cause tinea versicolor (also known as pityriasis versicolor) or folliculitis. Cogent evidence suggests that these organisms cause seborrheic dermatitis and dandruff.

CANDIDIASIS

Clinical Presentations

Oral candidiasis One form of oral candidiasis, thrush, appears as white to gray patches (pseudomembranes) on the tongue, soft palate, gingiva, oropharynx, and buccal mucosa. Removing the material from the mucosal surface reveals an underlying erythematous base. Predisposing factors in adults include diabetes mellitus, use of systemic or local corticosteroids, use of broad-spectrum antibiotics, use of radiotherapy or chemotherapy, and impaired cell-mediated immunity, especially from HIV infection. Acute atrophic candidiasis especially follows antibiotic therapy and causes painful, red, denuded lesions of the mucous membranes; the tongue may have erythematous areas with atrophic filiform papillae. In chronic atrophic candidiasis, contamination of dentures with *Candida* causes painful, red, and sometimes edematous lesions with a shiny, atrophic epithelium and well-demarcated borders where the dentures contact the mucous membranes. Poor dental hygiene and prolonged use of dentures are common predisposing factors. Some patients with these predisposing factors have angular cheilitis (perleche), characterized by erythema and fissuring of the corners of the mouth. Other contributing conditions are maceration from excessive salivation or licking, poorly fitting dentures, and a larger fold from diminished alveolar ridge height. *Candida* is present in most, but not all, patients with this disorder.

Chronic hyperplastic candidiasis (candidal leukoplakia) consists of irregular, white, persistent plaques on the tongue or mucous membranes that are difficult to remove; this form of candidiasis occurs especially in male smokers. Soreness, burning, and roughness of the affected areas are the usual symptoms. Candidiasis of the tongue can also take the form of median rhomboid glossitis, a diamond-shaped area of atrophic papillae in the central portion of the lingual surface.

Candidal intertrigo *Candida* infection may occur in any skin fold, causing soreness and itching. Obese patients are especially vulnerable. Commonly affected areas include the groin, inframammary regions, and folds of the abdominal pannus. The lesions are patches of bright erythema accompanied by maceration and an irregular, scalloped border, beyond which papules and pustules (satellite lesions) commonly form [*see Figure 5*].

Candidal vulvovaginitis and balanitis Most women with candidal vulvovaginitis have no underlying disease, but candidal vulvovaginitis may accompany diabetes mellitus and HIV infection. Candidal vulvovaginitis causes white plaques on a swollen, red vaginal mucosa; a creamy vaginal discharge; and erythema, sometimes with pustules, on the vulvar skin. Soreness and burning are common symptoms. Male sexual partners of women with candidal vulvovaginitis—especially male sexual partners who are uncircumcised—may develop balanitis, characterized by erythema, pustules, and erosions on the glans of the penis. Balanitis may occur spontaneously as well.

Candidal paronychia and nail infection Maceration of the tissue surrounding the nail, typically caused by excessive moisture, may cause paronychia, which is characterized by erythema, swelling, and pain of the nail fold with loss of the cuticle [*see Figure 6*]. *Candida* organisms often colonize the area but are probably pathogenic only when pus forms. With chronic colonization, nail involvement may occur, producing yellowish discoloration and separation of the nail plate from the nail bed (onycholysis). For chronic paronychia without purulence, topical corticosteroids, such as triamcinolone cream applied twice daily for 3 weeks, are the best therapy.[10]

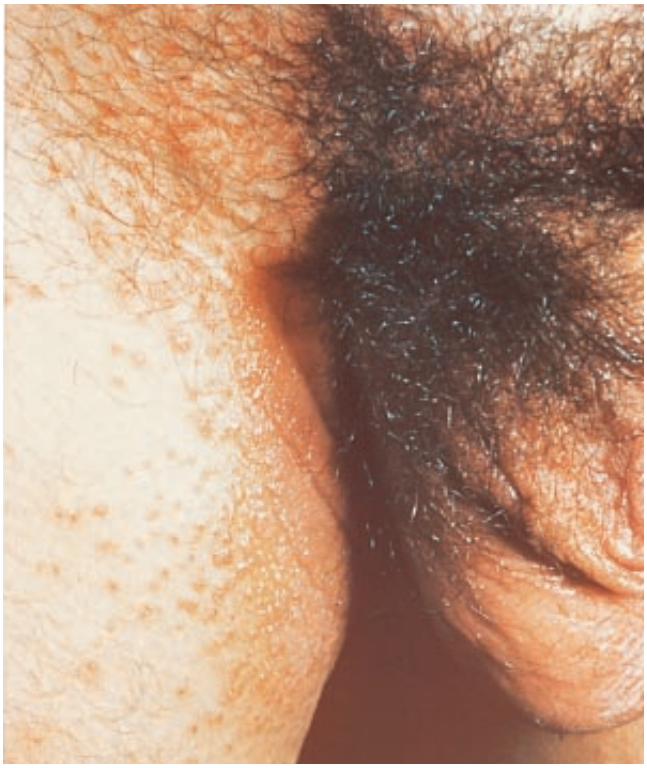


Figure 5 **Prominent satellite lesions of discrete vesicles are seen in a patient with candidiasis.**

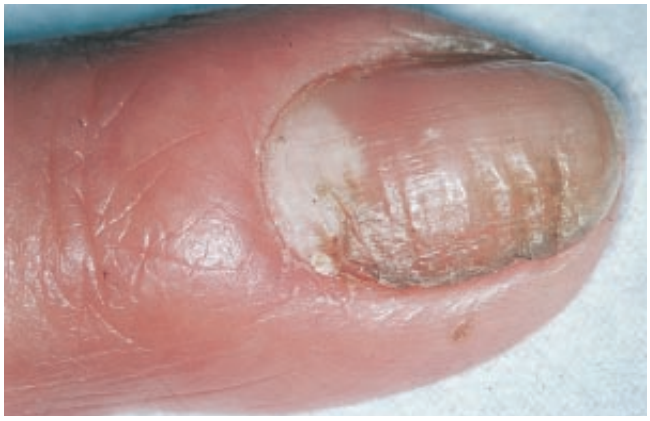


Figure 6 **In a *Candida* paronychia, seen on this patient's thumb, the nail fold becomes red, swollen, and painful. Nail dystrophy is also seen.**

Diagnosis

Scrapings from cutaneous or mucous membrane lesions may be mixed with KOH solution and examined under the microscope for budding yeasts with pseudohyphae. Gram stains of the same specimen are easier to evaluate because they disclose very large, oval, gram-positive cocci that may demonstrate budding or pseudohyphal formation. These organisms are much larger than bacteria and are much easier to see on Gram stain than on KOH preparation. Culture of specimens may be useful if the microscopy is normal or ambiguous. These organisms grow rapidly on both fungal and conventional bacterial media.

Treatment

Oral candidiasis For oral candidiasis, topical nystatin suspension, 200,000 to 400,000 units three to five times a day, is usually effective; an alternative treatment is clotrimazole troches. For patients in whom topical treatment is ineffective or poorly tolerated, systemic therapies include ketoconazole, 200 mg/day; fluconazole, 100 mg/day; and itraconazole, 100 mg twice a day. Angular cheilitis usually responds to an azole cream, such as miconazole or clotrimazole. Dentures should be cleaned carefully with an effective disinfectant, such as chlorhexidine.

Candidal intertrigo and balanitis Candidal intertrigo and balanitis respond to a topical azole cream, such as miconazole or clotrimazole.

Candidal vulvovaginitis Treatment of vulvovaginitis includes a topical azole in the form of a cream, suppository, or ointment, administered intravaginally, typically once daily for 7 days. A cream may be used for vulvar involvement. An alternative to suppositories is treatment with a single oral dose (150 mg) of fluconazole, which is at least as effective as topical therapy and is often preferred by patients.

Candidal paronychia Patients with candidal paronychia should keep their fingers dry; when wet work is unavoidable, patients should use cotton liners under rubber gloves. Prolonged topical therapy with creams or solutions of various azole preparations, such as clotrimazole, is often necessary to eradicate the infection.

Clinical Presentations

Tinea versicolor (pityriasis versicolor) Because the term tinea traditionally refers to dermatophyte infection, some clinicians prefer the term pityriasis, which means scaling, for this yeast infection. Usually asymptomatic, tinea versicolor may cause itching or skin irritation. The lesions are small, discrete macules that tend to be darker than the surrounding skin in light-skinned patients and hypopigmented in patients with dark skin. They often coalesce to form large patches of various colors (versicolor) ranging from white to tan [*see Figure 7*]. Scratching the lesions produces a fine scale. This infection most commonly involves the upper trunk, but the arms, axillae, abdomen, and groin may also be affected. Most lesions fluoresce a yellowish color under a Wood light.

***Malassezia* folliculitis (*Pityrosporum* folliculitis)** In folliculitis, inflammation of the hair follicle causes red papules and pustules that surround individual hairs. One cause of folliculitis is *M. furfur*. Lesions appear predominantly on the trunk but occasionally occur on the arms as well. The lack of comedones distinguishes the lesion from acne. Pruritus and stinging may be present.

Diagnosis

In patients with tinea versicolor, KOH preparations of scrapings from the lesions demonstrate pseudohyphae and yeasts, which resemble spaghetti and meatballs. This technique is sufficient to establish the diagnosis. The yeast form prevails in folliculitis and is easily seen on Gram stain of purulent material from a pustule, appearing as a large, oval, gram-positive coccus that is much larger than bacteria. Biopsies of these lesions show organisms around and within the hair follicle, with accompanying neutrophilic inflammation. The yeasts are best seen with periodic acid–Schiff or Gomori methenamine-silver stain. Because these yeasts form part of the normal cutaneous flora, growth of the organism on cultures from scrapings of the skin surface is not very helpful diagnostically. Culture of the yeast from the pus of folliculitis, however, is definitive, but it requires special media, such as Sabouraud agar with olive oil, to provide the necessary lipids for growth. Growth typically occurs in 3 to 5 days.

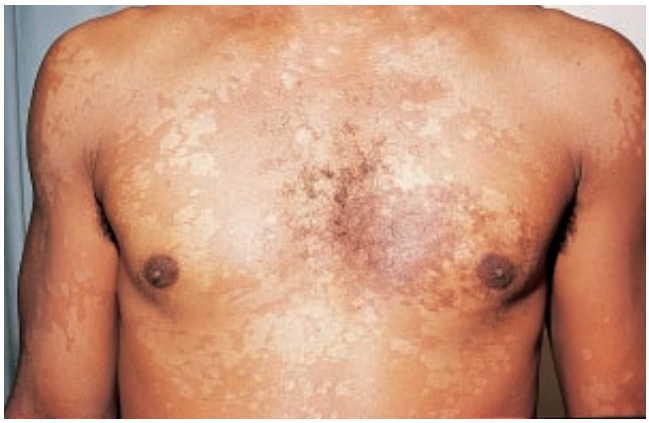


Figure 7 **Tinea versicolor appears on the chest of this patient as oval, hypopigmented, finely scaling macules.**

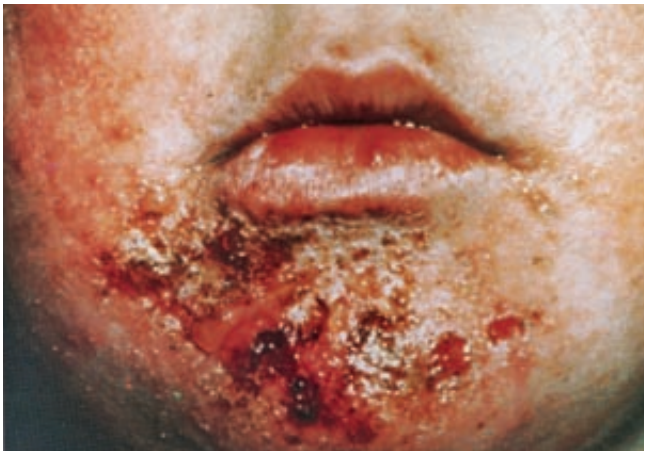


Figure 8 **Vesicopustules or bullae of impetigo rupture quickly and leave an erythematous base covered with a thin, seropurulent exudate. The exudate dries, forming layers of honey-colored crusts.**

Treatment

Simple treatment of tinea versicolor and *Malassezia* folliculitis involves applying selenium sulfide shampoo from the chin to the waist and from the shoulders to the wrist, allowing the shampoo to dry, and then washing it off after 10 to 15 minutes. Repeating this regimen after 1 week is usually effective; reapplication once every few weeks as necessary should prevent relapses, which are otherwise common. With tinea versicolor, scaling resolves promptly, but the pigmentary changes may take weeks to months to disappear. Topical azoles, such as ketoconazole, miconazole, and clotrimazole, are also effective, but the expense of these drugs makes their use impractical except for small or isolated lesions. For patients who have difficulty applying a topical agent because of physical disabilities or other factors, oral ketoconazole or fluconazole in a single 400 mg dose is an effective alternative. This oral program can be repeated for recurrences.

Bacterial Infections

SKIN INFECTIONS CAUSED BY STREPTOCOCCI, STAPHYLOCOCCI, OR BOTH

Impetigo

Initially a vesicular infection of the skin, impetigo rapidly evolves into pustules that rupture, with the dried discharge forming honey-colored crusts on an erythematous base [*see Figure 8*]. The lesions are often itchy. Impetigo characteristically occurs on skin damaged by previous trauma, such as abrasions or cuts. Exposed areas are most commonly involved, typically the extremities or the areas around the mouth and nose. Impetigo is usually a disease of young children and is more frequent in hot, humid climates than in temperate ones.

The usual cause is *Staphylococcus aureus*, but sometimes, *Streptococcus pyogenes* (group A streptococci) is also present; occasionally, *S. pyogenes* is the sole organism cultured.[11] Some strains of *S. aureus* elaborate a toxin that causes a split in the epidermis and the development of thin-roofed bullae. In this disorder, known as bullous impetigo, superficial, fragile, and flaccid vesiculopustules form and then rupture, with the exu-

date drying into a thin, brown, varnishlike crust. Sometimes, the vesiculopustules are not apparent, and erythematous erosions are the only evident disturbance.

Growth of *S. aureus*, *S. pyogenes*, or both from the skin lesions confirms the diagnosis, but cultures are unnecessary in characteristic cases. For treatment of sparse, nonbullous lesions, topical mupirocin ointment applied three times daily for 7 days is as effective as oral antimicrobials. Systemic antibiotics active against both *S. aureus* and *S. pyogenes*, such as cephalexin or dicloxacillin, represent an alternative to topical treatment. For extensive lesions, these antibiotics are preferred to topical therapy, and they are the treatment of choice for bullous impetigo. Because of the superficial nature of these infections, the lesions heal without scarring.

Ecthyma

Ecthyma (from the Greek word *ekthyma*, meaning pustule) is a deeper infection than impetigo. As with impetigo, *S. aureus*, *S. pyogenes*, or both may be the cause. Ecthyma commonly occurs in patients with poor hygiene or malnutrition or patients who have had skin trauma. The lesions, which are often multiple and are most common on the lower extremities, begin as vesicles that rupture, creating circular, erythematous lesions with adherent crusts. Beneath the scabs, which may spontaneously slough, are ulcers that leave a scar when healing occurs. Culture of the ulcer base yields the causative organisms. Treatment should be with an oral antistaphylococcal agent, such as dicloxacillin or cephalexin.

SKIN INFECTIONS CAUSED BY STREPTOCOCCI

Cellulitis and Erysipelas

Cellulitis and erysipelas are acute, spreading infections of the skin caused by streptococci of groups A, B, C, and G. Erysipelas involves the superficial dermis, especially the dermal lymphatics, and cellulitis affects the deeper dermis and subcutaneous fat. Erysipelas has an elevated, sharply demarcated border, but differences in the clinical appearances of erysipelas and cellulitis are unimportant and often unclear. The most common sites of infection are the face and lower extremities. The causative organisms may enter the skin at obvious areas, such as traumatic wounds and leg ulcers, or through cutaneous inflammation (e.g., eczema); often, however, no point of entry is apparent. Edema from any cause, including venous insufficiency, hypoalbuminemia, and lymphatic damage, is a predisposing factor. Infection commonly occurs on skin that has been permanently damaged by burns, trauma, radiotherapy, or surgery. For example, cellulitis may occur at the site of a saphenous vein removal for cardiac or vascular surgery months to years after the procedure.[12] An important predisposing factor in patients with cellulitis or erysipelas is tinea pedis, especially interdigital involvement; streptococci can invade the skin at sites of tinea pedis through adjacent skin surface disrupted by the fungal infection or can migrate to more proximal locations on the leg and enter through abnormal skin there. Obesity is also a predisposing condition.[13]

Diagnosis Cutaneous findings include rapidly expanding erythema and swelling of the skin [*see Figure 9*], sometimes accompanied by proximal streaks of redness, representing lymphangitis, and tender, enlarged regional lymph nodes. Vesicles, bullae, petechiae, and ecchymoses may occur. The cutaneous surface may resemble the skin of an orange (peau d'orange) because the hair follicles remain tethered to the deeper structures, keeping their openings below the surrounding superficial edema and creating the characteristic dimpling of the skin. On the face, the typical location is on one or both cheeks, with a butterfly pattern of erythema and swelling. Extension to the eyelids, ears, or neck is common. Systemic symptoms, such as fever, headache, and confusion, can accompany these infections; sometimes, such symptoms precede by hours any cutaneous findings on examination. Other patients have no systemic features despite severe skin abnormalities.

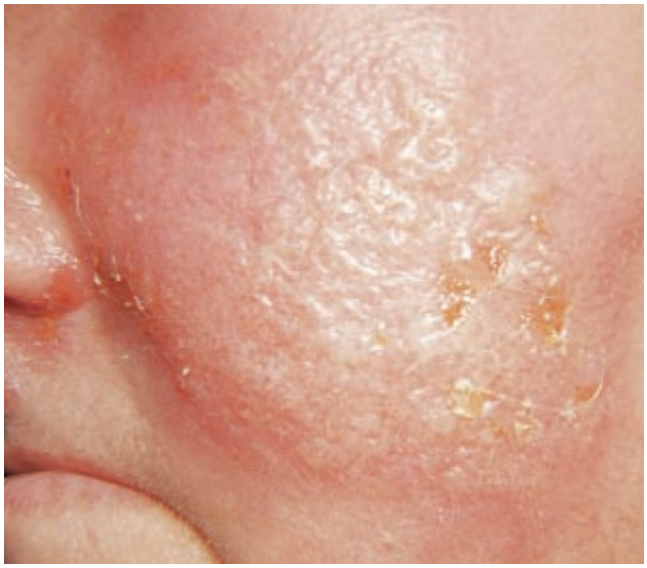


Figure 9 **Erythema, edema, and sharp demarcation of the lesion from the normal surrounding skin characterize facial erysipelas.**

The diagnosis is largely clinical; in a typical case, cultures are unnecessary and usually unrewarding. Needle aspiration of the lesion yields an isolate in about 5% of specimens, as do blood cultures in febrile patients. Because of their low yield, blood cultures are unrewarding in typical cases of cellulitis.[14] Punch biopsies of the skin are culture-positive in about 20% of cases.[15] These results, together with serum antibody tests for streptococci[16] and immunofluorescent studies of skin biopsies,[17] indicate that streptococci cause the vast majority of cases of cellulitis and erysipelas. *S. aureus* is often suspected but rarely implicated in cellulitis in the absence of an abscess or penetrating injury. Additional circumstances in which organisms other than streptococci are likely to be responsible for cases of cellulitis include immunodeficiency, penetrating trauma, immersion injuries in freshwater or saltwater, granulocytopenia, and animal bites or scratches. Cultures are appropriate in these situations.

Treatment Treatment consists of elevation of the affected area to help reduce edema and administration of systemic antibiotic therapy. For patients who do not have serious systemic illness, oral treatment is satisfactory. Penicillin is the drug of choice for streptococcal infections; for outpatients who may not take an oral medication as prescribed, I.M. benzathine penicillin G in an adult dose of 1.2 million units provides a complete course. Instead of penicillin, many clinicians prescribe an antistaphylococcal agent—either a first-generation cephalosporin or a penicillinase-resistant penicillin—because of concerns about *S.*

aureus. Patients often get worse shortly after therapy, with further extension of the cellulitis, higher fever, greater toxicity, and increased white blood cell counts, presumably because rapid killing of the organisms releases potent enzymes, such as streptokinase and hyaluronidase, that cause many of the clinical features. Oral prednisolone, taken for 8 days in doses of 30 mg, 15 mg, 10 mg, and 5 mg, with each dose taken for 2 days, decreases the duration of cellulitis and shortens hospital stay; it is a reasonable regimen in those with no contraindications to systemic corticosteroids.[18]

In patients with leg cellulitis, treatment of tinea pedis is useful in preventing further episodes, which are likely to cause permanent lymphatic damage and can lead to lymphedema and further risk of infection. Other measures to diminish the frequency of future attacks include control of edema by diuretics or mechanical means, such as elastic stockings, and, for those with frequent episodes, prophylactic antibiotics. The easiest approach is the administration of oral penicillin or erythromycin, 250 mg twice daily.[19,20]

INFECTIONS DUE TO *STAPHYLOCOCCUS AUREUS*

Furunculosis

A furuncle is a deep-seated inflammatory nodule with a pustular center that develops around a hair follicle [*see Figure 10*]. With involvement of several adjacent follicles, a mass called a carbuncle may form, with pus discharging from multiple follicular orifices. This infection typically develops on the back of the neck and appears more commonly in patients with diabetes than in the general population. Moist heat is usually adequate for small furuncles, which ordinarily drain spontaneously. Incision and drainage are appropriate for large or multiple furuncles and for all carbuncles. Systemic antibiotics are unnecessary unless there is fever or substantial surrounding cellulitis.

Some patients have recurrent episodes of furunculosis. Although a few patients have definable abnormalities in host defenses, such as neutrophil disorders, most are otherwise healthy people who, like 20% to 40% of the population, carry *S. aureus* in the anterior nares. From this site or occasionally from the perineum or axilla, organisms can spread and enter the skin, presumably through minor, usually inapparent, trauma. Successful prevention of recurrent infection requires eradication of these bacteria from their site of residence, but most systemic antibiotics do not achieve adequate levels of drug in the anterior nares. An exception is clindamycin, which, when given as a single daily dose of 150 mg for 3 months, is very effective in preventing subsequent episodes.[21] A less effective alternative is mupirocin ointment, applied in the anterior nares twice daily for 5 days each month.[22]

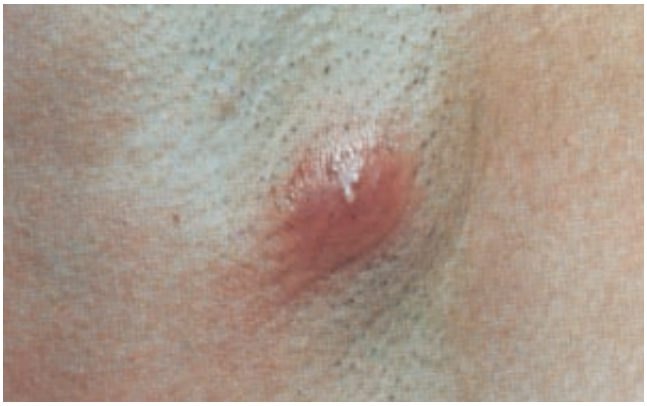


Figure 10 **A furuncle, or boil, occurs as an acute, painful, localized staphylococcal abscess surrounding a hair follicle.**

SKIN INFECTIONS CAUSED BY THE RESIDENT CUTANEOUS FLORA

The normal cutaneous flora helps prevent infection by other organisms through the mechanisms mentioned above: occupying available sites of residence, competition for nutrients, establishment of a low pH, and the elaboration of antibacterial substances. Occasionally, however, the resident skin flora causes cutaneous infections, especially with trauma or alterations in the stratum corneum. Examples are erythrasma, pitted keratolysis, trichomycosis axillaris, and most cases of cutaneous abscesses.

Cutaneous Abscesses

Cutaneous abscesses are collections of pus within the dermis and deeper skin tissues. They probably occur as a result of trauma. Sites of trauma associated with cutaneous abscesses may be apparent, as with sites of injections in illicit-drug users,[23] or they may be minor and unnoticed. *S. aureus*, usually in pure culture, causes about 25% of cutaneous abscesses, especially in the axillae, on the hand, and on the breasts of women after childbirth.[24] In other sites, however, the predominant organisms are anaerobes. Anaerobes occur either alone or in the mixture of anaerobes and aerobes that constitutes the normal regional flora; they are sometimes accompanied by microbes from adjacent mucous membranes. In anogenital infections, such as scrotal, inguinal, vaginal, buttock, and perirectal abscesses, the organisms are commonly fecal bacteria, including streptococci, anaerobic gram-positive cocci, and anaerobic gram-negative bacilli, such as *Bacteroides fragilis*. On the extremities, trunk, neck, and head, the usual microbes include coagulase-negative staphylococci, anaerobic gram-positive cocci, and *Propionibacterium acnes*, an anaerobic gram-positive bacillus. These organisms ordinarily possess little virulence, but when introduced into the dermis or subcutaneous tissue by trauma or through a disrupted cutaneous surface, they may become pathogenic.

Cutaneous abscesses usually cause a painful, fluctuant, red, tender swelling, on which may rest a pustule. Treatment is incision and drainage of the area. Gram stain and culture of the pus are ordinarily unnecessary, as are topical antimicrobials. Systemic antibiotics are reserved for patients with extensive surrounding cellulitis, neutropenia, cutaneous gangrene, or systemic manifestations of infection, such as high fever.

Erythrasma

Porphyrin-producing coryneform bacteria, which are gram-positive bacilli that constitute part of the normal cutaneous flora, cause a superficial, usually asymptomatic, skin disorder called erythrasma. One particular species, *Corynebacterium minutissimum*, has often been cited as the sole cause of this infection, but its precise role, if any, remains unclear. The most common site of erythrasma is between the toes, especially in the fourth interdigital space, where it causes fissuring, maceration, and scaling, resembling tinea pedis. Other locations are intertriginous areas, such as the axillae, groin, submammary area, and intergluteal cleft. In these regions, the lesions are usually scaly, brownish-red, sharply circumscribed patches. In hot,

humid climates, more extensive disease may occur. The definitive diagnostic technique is examination of the skin with a Wood light, which, because the organisms produce porphyrins, reveals a coral-red fluorescence. Culture of the lesions, which requires special media, is unnecessary. Because they possess some activity against gram-positive bacteria, topical azoles, such as miconazole and clotrimazole, are effective in the treatment of this infection. Topical erythromycin or clindamycin is also effective. Oral erythromycin (250 mg q.i.d. for 2 weeks) is an alternative.[25]

Pitted Keratolysis

C. minutissimum and a gram-positive coccus, *Micrococcus sedentarius*, either alone or together, cause a disorder that may affect the soles—typically in pressure-bearing areas—or, occasionally, the palms.[26] Pitted keratolysis consists of small pitted erosions about 1 to 7 mm in diameter that may be present on reddened plaques and are often more apparent after soaking in water for a few minutes. This infection occurs with increased moisture, such as caused by excessive sweating, occlusive footwear, or frequent contact with water. It appears more commonly in hot, humid climates than in more temperate ones. An impressive malodor of the feet is often apparent, and although the disorder may cause no symptoms, some patients complain of itching, tenderness, or sliminess of the feet. As in erythrasma, topical azoles, such as clotrimazole and miconazole, are effective, as are topical erythromycin and clindamycin.

Trichomycosis axillaris

Trichomycosis axillaris is characterized by colored concretions of axillary hair that result from infection of the hair shafts by large colonies of various species of *Corynebacterium*. The nodules may be yellow, black, or red; and because the organisms may invade the cuticle, the hair can become brittle. The same process occasionally affects the facial or pubic hair.[27] Excessive sweating, poor hygiene, and failure to use an axillary deodorant are predisposing factors. Shaving the hair is effective treatment; other options include topical erythromycin or clindamycin.

INFECTIONS DUE TO OTHER BACTERIA

Necrotizing Fasciitis

Necrotizing fasciitis, a necrotizing infection of the subcutaneous tissue, can be caused by streptococci; more often, however, the responsible organisms are a combination of aerobic bacteria—such as gram-negative enteric organisms (e.g., *Escherichia coli)* and gram-positive cocci—and anaerobes, including *B. fragilis*.[28] Necrotizing fasciitis usually occurs after a penetrating wound to the extremities. The injury is typically deep, but sometimes, infection occurs after apparently trivial trauma, such as abrasions or lacerations. The necrotizing process may develop from extension of an adjacent infection, especially in the second most common location, the anogenital area. There, infection typically arises from a perianal abscess; as an extension of a periurethral gland infection, especially in men with urethral strictures; through retroperitoneal suppuration from perforated abdominal viscera; or as a complication of a preceding surgery. Necrotizing infection involving the genitalia is called Fournier gangrene.

These infections typically begin with fever, systemic toxicity, severe pain in the affected site, and the development of a painful, red swelling that rapidly progresses to necrosis of the subcutaneous tissue and overlying skin. Early on, the pain may appear disproportionate to the clinical findings. In some cases involving *S. pyogenes* infection, the characteristics of the streptococcal toxic-shock syndrome may appear[29] [*see 7:I Infections Due to Gram-Positive Cocci*]. When anaerobes or certain aerobic gram-negative bacilli cause the infection, gas may form in tissues, evident as crepitus on physical examination or visible on radiographic studies. Although the disease may resemble uncomplicated cellulitis, the following signs and symptoms should suggest the presence of a necrotizing subcutaneous infection: edema beyond the apparent limits of the infection; rapid development of bullae and ecchymoses; cutaneous gangrene; fluctuance; crepitus; and radiographically visible gas. Computed tomography or magnetic resonance imaging may be helpful in some cases in detecting the infection and defining its extent. Aspiration of the affected tissue may yield purulent fluid, which on Gram stain demonstrates only gram-positive cocci in chains when *S. pyogenes* is responsible or reveals a variety of many different organisms when a mixed infection is present. The findings on Gram stain and culture of pus should dictate antibiotic choice, but a good initial program is gentamicin in combination with clindamycin. Most important is incision and drainage of the affected area, which should include removal of any necrotic tissue. Often, the amount of disease revealed at surgery is much greater than was apparent on the preoperative clinical examination, because the infection typically extends far beyond the borders of cutaneous inflammation. Repeat operation after 24 hours is typically prudent to detect new areas of infection and necrotic tissue.

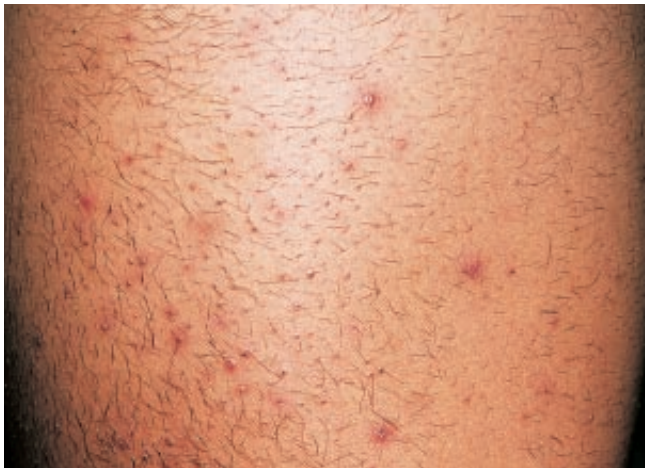
Folliculitis

Folliculitis is an inflammation at the opening of the hair follicle that causes erythematous papules and pustules surrounding individual hairs [*see Figure 11*]. The most common location is the trunk. The initiating factor seems to be occlusion of the opening of the follicle, which may occur from contact with chemicals, such as oils or cosmetics; overhydration of the skin from excessive moisture; or repetitive trauma, such as friction from tight-fitting clothing, which elicits hyperkeratosis and follicular plugging. Subsequently, inflammation develops, which may be provoked by bacteria, yeast, or other nonmicrobial substances trapped beneath the occluded ostium.

Among bacteria, *S. aureus* is often suspected but rarely found. When bacteria are present in the pustules, the culture usually yields normal skin flora. In these patients, oral erythromycin or doxycycline may be effective in eradicating the lesions. Another cause is *M. furfur*, a yeast that is a normal resident on the skin. In other patients, the avoidance of oily substances on the skin or tight clothing leads to resolution of the problem.

Occasionally, *Pseudomonas aeruginosa* is responsible, as a consequence of inadequate disinfection of hot tubs, swimming pools, or whirlpools.[30] This gram-negative bacillus grows well in hot water. Outbreaks occur an average of 48 hours after exposure, with a range of several hours to several days. Erythematous, pruritic papules, often with a pinpoint central pustule, appear in areas exposed to the contaminated water; lesions are particularly numerous in regions occluded by tight-fitting swimming suits. The lesions disappear spontaneously over several days, leaving no scars; ordinarily, no topical or systemic therapy is necessary. Some patients have sore throat,

a



b

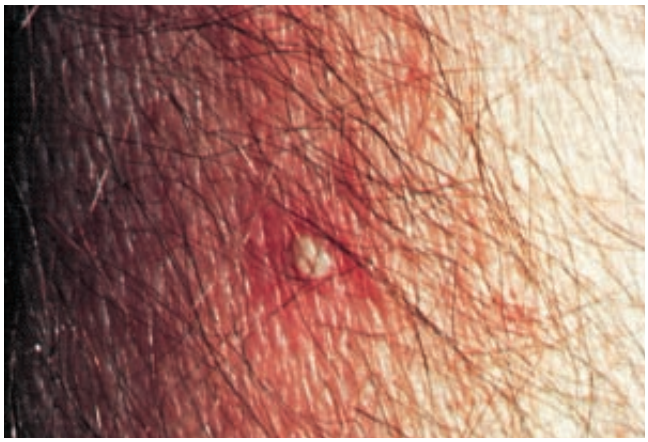


Figure 11 **Folliculitis is a superficial or deep inflammation of the hair follicles, appearing at follicular openings as small pustules surrounded by erythema (*a*). Folliculitis may also occur as an isolated lesion (*b*).**

rhinitis, earache, and headache, but fever or bacteremia is very rare. Cultures of the skin lesions and the contaminated water usually yield the organism.

Cutaneous Anthrax

Spores of *Bacillus anthracis* sent through the mail in the fall of 2001 as an act of bioterrorism caused cases of inhalational and cutaneous anthrax in several states. Otherwise, anthrax has been very rare in the United States over the past few decades. Ordinarily, this bacterium resides in the soil, where it forms spores that can persist for years. When ingested—primarily by herbivores (cattle, horses, sheep, and goats) grazing on contaminated land—these spores may cause infection. This veterinary disease is most frequent in tropical and subtropical areas, but extensive vaccination can markedly diminish its frequency.

Except for cases associated with bioterrorism [*see 8:V Bioterrorism*], humans usually develop anthrax from exposure to affected animals or their products, such as hides. Occasional laboratory-acquired cases also occur. The cutaneous form develops when spores enter the skin through abrasions and then transform into bacilli, which produce toxins that cause local tissue edema and necrosis. Macrophages can transport spores to regional lymph nodes, but bacteremia is uncommon. After an incubation period of about 1 to 7 days, a painless, pruritic papule forms at the entry site, most commonly the head, neck, and extremities. Over the next few hours the lesion enlarges, and a ring of erythema may form around it. In 1 to 2 days, vesicles appear, surrounding the papule and containing numerous bacteria but few neutrophils. Painless, gelatinous, nonpitting edema then encircles the lesion, often spreading extensively to adjacent skin and soft tissue [*see Figure 12*]. This pronounced edema is especially characteristic of anthrax. After enlarging, the vesicles become hemorrhagic and rupture. In the depressed center of the lesion, a black eschar forms and sloughs within 1 to 2 weeks, leaving a shallow ulcer that heals with minimal, if any, scarring. In the early days of illness, patients commonly have headache, malaise, and fever. Regional lymph nodes often enlarge, causing pain and tenderness.

Diagnosis *B. anthracis*, a broad, encapsulated gram-positive rod, is visible on Gram stains of material from a skin lesion as single organisms or chains of two or three bacilli. It grows readily at 37° C on blood agar media. Skin biopsies reveal necrosis, hemorrhage, and massive edema. Organisms are demonstrable with tissue Gram stain or immunohistochemical staining for the bacteria's cell wall antigen. Because it requires acute and convalescent blood specimens, serologic testing for antibodies to *B. anthracis* is unhelpful for immediate diagnosis but may establish a retrospective diagnosis of suspected but unconfirmed cases.

Treatment Treatment for cutaneous anthrax unassociated with bioterrorism is penicillin V (500 mg q.i.d. orally) or amoxicillin (500 mg t.i.d. orally) for mild cases and, for more severe disease, penicillin G (6 to 8 million units I.V. daily). For penicillin-allergic patients or cases arising from bioterrorism, the recommended therapy is oral ciprofloxacin (500 mg b.i.d.) or doxycycline (100 mg b.i.d). Antibiotic therapy does not alter the course of eschar formation and healing, but it does decrease the risk of systemic disease. Ordinarily, the duration of therapy is 7 to 10 days, but the recommended regimen for cases associated with bioterrorism is 60 days because of the possibility of simultaneous aerosol exposure.[31]

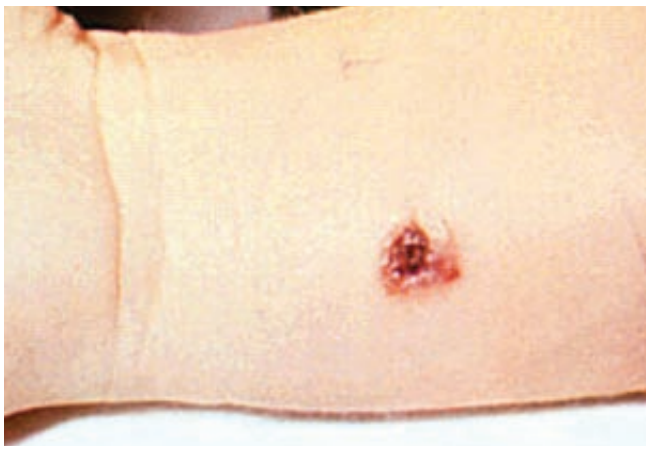


Figure 12 **Cutaneous anthrax lesion, seen on the seventh day after infection.**

WARTS

Warts, or verrucae, are caused by human papillomaviruses (HPVs), a subgroup of DNA-containing papovaviruses, of which there are more than 70 types. Humans are the only known reservoir; transmission probably occurs from close contact with infected people or possibly from exposure to sloughed, infected epidermal cells. The virus presumably enters through small breaks in the skin. The incubation period is difficult to discern but is probably several months. Autoinoculation from one portion of the body to another also occurs. Cell-mediated immunity appears important in controlling these infections, which can be very extensive and refractory to treatment in immunocompromised patients.

Verrucae vary according to location. They include the common, elevated wart (verruca vulgaris), typically appearing on the hands; the flat wart (verruca plana), on the face and legs; the moist wart (condyloma acuminatum), in the anogenital area; and the callus-covered plantar wart (verruca plantaris), on the sole of the foot. A histologic feature that distinguishes a wart from other papillomas is the presence in the upper epidermis of large, vacuolated cells that contain numerous viral particles.

Verruca Vulgaris

The common wart consists of single or multiple skin-colored papules, which often have a hyperkeratotic, papillary surface. They are commonly present on the fingers. The estimated nationwide prevalence of hand warts is 3.5% for people 18 to 64 years of age; the greatest frequency (5.5%) occurs in men 18 to 24 years of age. The warts may be filiform, with a small base and a thin projection of several millimeters, especially on the face.

Liquid nitrogen is a common initial treatment of choice for many warts. Administered with a cotton-tipped applicator or cryospray device, liquid nitrogen freezes the lesion, causing it to blister and subsequently dissolve. More than one application at 2- to 3-week intervals may be necessary for large or periungual warts. Electrodesiccation and curettage or laser surgery are effective for persistent or recurrent lesions.

Verruca Plana

The flat wart is a skin-colored or light-brown, slightly elevated, smooth papule commonly seen on the face and the dorsum of the hand. These may be difficult to treat, but freezing with liquid nitrogen, application of trichloroacetic acid, or painting the lesions with 10% salicylic acid and 10% lactic acid in flexible collodion may be effective.

Verruca Plantaris

The plantar wart is often painful and disabling. A mosaic wart, a variant of verruca plantaris, consists of multiple discrete or confluent superficial lesions and is often difficult to treat. A plantar wart that is covered by a callus can be distinguished from an ordinary callus by paring off the surface keratin; multiple, pinpoint dots, representing thrombosed vessels, or bleeding points from surface capillaries will become apparent if it is a wart. Paring of the wart can be followed by immediate treatment with liquid nitrogen, the application of strong acid (50% trichloroacetic acid), or the nightly administration of salicylic acid in plasters, an acrylic vehicle, or collodion.

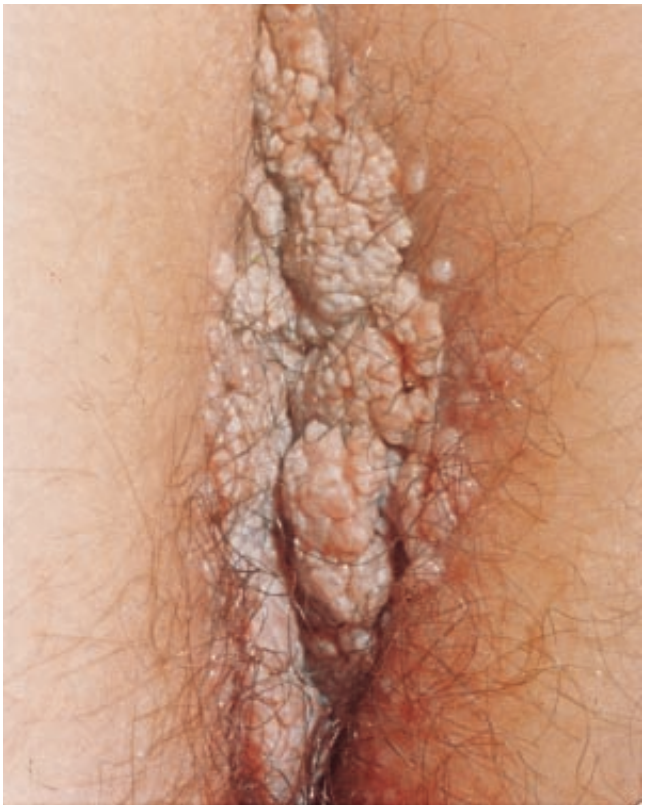


Figure 13 **Condyloma acuminatum may appear as a large cauliflower-like mass that resembles a malignant tumor.**

Condyloma Acuminatum

Anogenital warts consist of skin-colored or gray, discrete or confluent cauliflower-like excrescences that may cause no symptoms or produce itching, burning, pain, or tenderness [*see Figure 13*]. The incidence is highest in young adults; most often, it is a sexually transmitted disease, though some anogenital warts may develop from autoinoculation or may be acquired in other ways.[32]

Infection with some types of HPV predisposes to malignancy. Most cases of squamous carcinoma of the cervix are caused by HPV, especially HPV-16 and HPV-18, but fortunately, these types represent only a small percentage of the isolates from anogenital warts. Genital verrucous carcinoma, also called giant condyloma acuminatum of Buschke-Löwenstein, is a low-grade genital malignancy caused by HPV-6 and HPV-11. Squamous carcinoma of the anus is associated primarily with HPV-16.

Anogenital warts may be difficult to eradicate, and several treatments are often necessary.[33] Therapies administered by clinicians include liquid nitrogen, podophyllin resin, trichloroacetic or bichloroacetic acid, surgical removal, laser therapy, or intralesional interferon. Patient-applied treatments are podophyllotoxin, which the patient applies twice daily for 3 days, or imiquimod cream, used at bedtime three times a week for up to 16 weeks. Another approach involves fluorouracil (5-FU) cream administered twice daily for 1 to 3 weeks. This medication is particularly suitable for large wart plaques and warts of the urethral meatus, but side effects, including discomfort and painful erosions, are common.

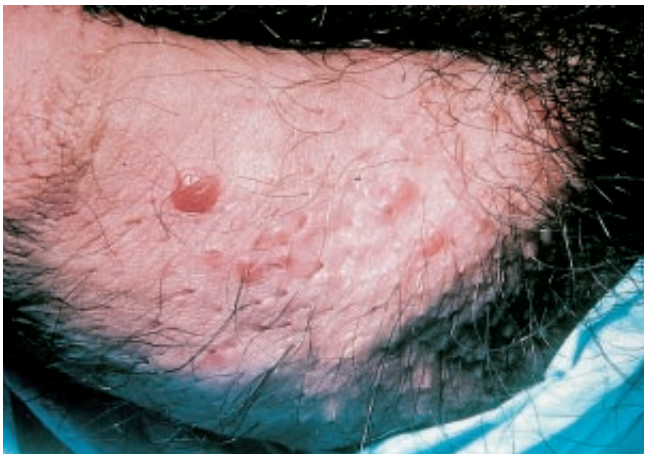


Figure 14 **Benign lesions of bowenoid papulosis, as seen on the shaft of the penis, may histologically resemble carcinoma in situ.**

Bowenoid Papulosis

Bowenoid papulosis consists of benign-appearing erythematous or pigmented papules in the anogenital area that histologically resemble Bowen disease (squamous cell carcinoma in situ) [*see Figure 14*]. Its course, however, is not aggressive, and the papules should be treated as anogenital warts (see above). HPV-16 is a common cause, however, and malignancy does occasionally develop, especially in women.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Roth RR, James WD: Microbiology of the skin: resident flora, ecology, infection. J Am Acad Dermatol 20:367, 1989

2. Leyden JJ, McGinley KJ, Nordstrom KM, et al: Skin microflora. J Invest Dermatol 88(suppl):65s, 1987

3. Macura AB: Dermatophyte infections. Int J Dermatol 32:313, 1993

4. DeVroey C: Epidemiology of ringworm. Semin Dermatol 4:185, 1985

5. Elewski BE: Tinea capitis: a current perspective. J Am Acad Dermatol 45:320, 2001

6. Semel JD, Goldin H: Association of athlete's foot with cellulitis of the lower extremities: diagnostic value of bacterial cultures of ipsilateral interdigital space samples. Clin Infect Dis 23:1162, 1996

7. Elewski BE, Leyden J, Rinaldi MG, et al: Office practice–based confirmation of onychomycosis: a US nationwide prospective survey. Arch Intern Med 162:2133, 2002

8. Gupta AK, Adam P, Dlova N, et al: Therapeutic options for the treatment of tinea capitis caused by *Trichophyton* species: griseofulvin versus the new oral antifungal agents terbinafine, itraconazole, and fluconazole. Pediatr Dermatol 18:433, 2001

9. Evans EV, Sigurgeirsson B: Double blind, randomized study of continuous terbinafine compared with intermittent itraconazole in treatment of toenail onychomycosis. The LION Study Group. BMJ 318:1031, 1999

10. Tosti A, Piraccini BM, Ghetti E, et al: Topical steroids versus systemic antifungals in the treatment of chronic paronychia: an open, randomized double-blind and double dummy study. J Am Acad Dermatol 47:73, 2002

11. Demidovich CW, Wittler RR, Ruff ME, et al: Impetigo: current etiology and comparison of penicillin, erythromycin, and cephalexin therapies. Am J Dis Child 144:1313, 1990

12. Dan M, Heller K, Shapira I, et al: Incidence of erysipelas following venectomy for coronary artery bypass surgery. Infection 15:107, 1987

13. Dupuy A, Benchikhi H, Roujeau JC, et al: Risk factors for erysipelas of the leg (cellulitis): case-control study. BMJ 318:1591, 1999

14. Perl B, Gottehrer NP, Raveh D, et al: Cost-effectiveness of blood cultures for adult patients with cellulitis. Clin Infect Dis 29:1483, 1999

15. Hook EW, Hooton TM, Horton CA, et al: Microbiologic evaluation of cutaneous cellulitis in adults. Arch Intern Med 146:295, 1986

16. Eriksson B, Jorup-Rönstrom C, Karkkonen K, et al: Erysipelas: clinical and bacteriologic spectrum and serological aspects. Clin Infect Dis 23:1091, 1996

17. Bernard P, Bedane C, Mounier M, et al: Streptococcal cause of erysipelas and cellulitis in adults: a microscopic study using a direct immunofluorescence technique. Arch Dermatol 125:779, 1989

18. Bergkvist PI, Sjöbeck K: Antibiotic and prednisolone therapy of erysipelas: a randomized, double blind placebo-controlled study. Scand J Infect Dis 29:377, 1997

19. Kremer M, Zuckerman R, Avraham Z, et al: Long-term antimicrobial therapy in the prevention of recurrent soft-tissue infections. J Infect 22:37, 1991

20. Hirschmann JV: Antimicrobial prophylaxis in dermatology. Semin Cutan Med Surg 19:2, 2000

21. Klempner MS, Styrt B: Prevention of recurrent staphylococcal skin infections with low-dose oral clindamycin therapy. JAMA 260:2682, 1988

22. Raz R, Miron D, Colodner R, et al: A 1-year trial of nasal mupirocin in the prevention of recurrent staphylococcal nasal colonization and skin infection. Arch Intern Med 156:1109, 1996

23. Ebright JR, Pieper B: Skin and soft tissue infections in injection drug users. Infect Dis Clin North Am 16:697, 2002

24. Meislin HW, Lerner SA, Graves MH, et al: Cutaneous abscesses: anaerobic and aerobic bacteriology and outpatient management. Ann Intern Med 87:145, 1977

25. Holdiness MR: Management of cutaneous erythrasma. Drugs 62:1131, 2002

26. Takama H, Tamada Y, Yano K, et al: Pitted keratolysis: clinical manifestations in 53 cases. Br J Dermatol 137:282, 1997

27. White SW, Smith J: Trichomycosis pubis. Arch Dermatol 115:444, 1979

28. Stone DR, Gorbach SL: Necrotizing fasciitis: the changing spectrum. Dermatol Clin 15:213, 1997

29. Dahl PR, Perniciaro C, Holmkvist KA, et al: Fulminant group A streptococcal necrotizing fasciitis: clinical and pathologic findings in 7 patients. J Am Acad Dermatol 47:489, 2002

30. Agger WA, Mardan A: *Pseudomonas aeruginosa* infections of intact skin. Clin Infect Dis 20:302, 1995

31. Inglesby TV, O'Toole T, Henderson DA, et al: Anthrax as a biological weapon, 2002. Updated recommendations for management. JAMA 287:2236, 2002

32. Von Krogh G, Gross G: Anogenital warts. Clin Dermatol 15:355, 1997

33. Gross G, Von Krogh G: Therapy of anogenital HPV-induced lesions. Clin Dermatol 15:457, 1997

Acknowledgment

Figure 12 Centers for Disease Control and Prevention Public Health Image Library.

VIII PARASITIC INFESTATIONS

Elizabeth A. Abel, m.d.

Ectoparasites may cause severely pruritic infectious diseases of the skin. With early detection and treatment, parasitic infestations can be cured and their spread to other persons prevented. The most common parasitic diseases of the skin that occur in nontropical environments are scabies, which is caused by itch mites, and pediculosis capitis, pediculosis corporis, and pediculosis pubis, which are caused by bloodsucking lice.

An increase in international travel, including vacation travel to tropical destinations and immigration from such areas, has led to the occurrence of parasitic disorders endemic to tropical regions in persons living in temperate climates. The differential diagnosis of skin disorders in patients treated at a tropical disease clinic in Paris over a 2-year period included cutaneous larva migrans, pyodermas, arthropod-reactive dermatitis, myiasis, tungiasis, urticaria, and cutaneous leishmaniasis.[1] The prevalence of ectoparasitoses in the general population is usually low, but it can be high in vulnerable groups. Management of some infestations (e.g., scabies and head lice) can be complicated because resistance to insecticides is spreading and unpredictable.[2]

Scabies

Scabies is caused by infestation with *Sarcoptes scabiei*, an ectoparasite that bores into the corneal layer of human skin, forming burrows in which it deposits its eggs. The incubation period is 2 to 6 weeks in a person who has not been previously exposed. During this time, the host develops delayed hypersensitivity to mite antigens. Upon reinfestation, symptoms occur in sensitized persons within 24 to 48 hours after exposure.[3]

The scabies mite does not survive for more than 48 hours away from the host. Therefore, most infestations are transmitted through direct personal skin-to-skin and sexual contact.[3] However, transfer of organisms can occur by exposure to fomites such as contaminated bedding, clothing, or furniture and is a common cause of epidemics of scabies in nursing homes and other institutions.[3,4]

Diagnosis

Clinical Features

Scabies causes severe itching, which is usually worse at night. Characteristic sites of infestation are the webs of the fingers, the flexor aspects of the wrists, the axillae, the buttocks, the umbilicus, the penis and scrotum of males, and the breasts and nipples of females. The disease is more generalized in infants and children than in adults.

The burrow of the female *Sarcoptes* may be seen as an irregular zigzag line in the stratum corneum, with a black dot at one end that indicates the presence of the mite [*see Figure 1*]. Secondary lesions represent immunologic reactions to the mites and usually appear as small erythematous papules and vesicles with surrounding edema and scratch marks [*see Figure 2*]. The type and number of lesions depend predominantly on the immune status of the host. Occasionally, nodular lesions, which may resemble lesions of histiocytosis X (Langerhans cell granulomatosis) or lymphoma, occur as a hypersensitivity reaction to retained mite parts. Fewer lesions occur in people who practice good hygiene, and the condition may be masked in those who are using topical steroids. Secondary bacterial infection with impetiginization is common, especially in children and in elderly patients who actively excoriate their lesions.

Atypical presentations of scabies have been described in immunosuppressed persons, including organ transplant recipients, patients with lymphoma or leukemia, and patients with AIDS. Itching and scratching, with elimination of mites and burrows, may be minimal in patients who lack an immunologic host response, allowing for thousands of mites to reproduce and thrive.[3] Crusted scabies, which was originally described in Norway, is associated with widespread hyperkeratotic lesions and deep fissures in the skin. Crusted scabies can develop in patients with malnutrition or severe mental deficiency and in institutionalized patients. The condition is highly contagious because of the large number of mites present in the exfoliating skin.

A severe form of scabies with unusual clinical features consisting of crusted lesions and a widespread pruritic papular dermatitis has been described in HIV-infected patients.[3,5] In these patients, multiple treatment applications may be needed because of the large mite population and the patients' impaired immunologic response.

Skin Scrapings

A skin scraping that demonstrates the presence of mite eggs or mite products can confirm a diagnosis of scabies. A No. 15 surgical blade is used to scrape across one or more burrows. Saline solution or mineral oil is used to remove scrapings from the blade. The scrapings are then placed on a glass slide with a coverslip and examined under a microscope at low-power magnification. The scraping is positive if the gravid female, eggs, or scybala (fecal pellets) are seen [*see Figure 3*]. The yield is greatest in burrows that are not yet excoriated, which may be difficult to find. For this reason, if the scraping is negative but the clinical suspicion of scabies is high, the patient should be treated empirically. Histopathologic examination of a skin biopsy sample is also diagnostic if it reveals the mite or the superficial skin bur-

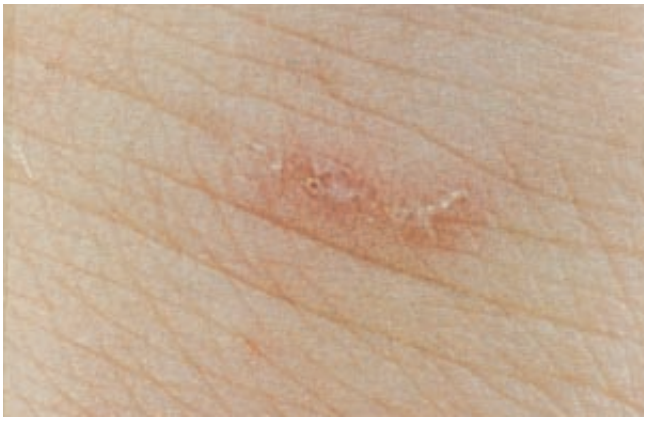


Figure 1 **Typical scabies lesions are small erythematous papules and vesicles with surrounding edema.**

row and its contents.[6] In atypical or subtle cases of scabies, epiluminescence microscopy (ELM) or polymerase chain reaction may be useful in confirming the diagnosis.[7,8] ELM, which allows visualization of the skin down to the superficial papillary dermis, is able to detect the presence of scabies within minutes, with no discomfort to the patient. PCR can amplify *S. scabiei* DNA when the number of mites is so few that diagnosis by standard means is inconclusive.

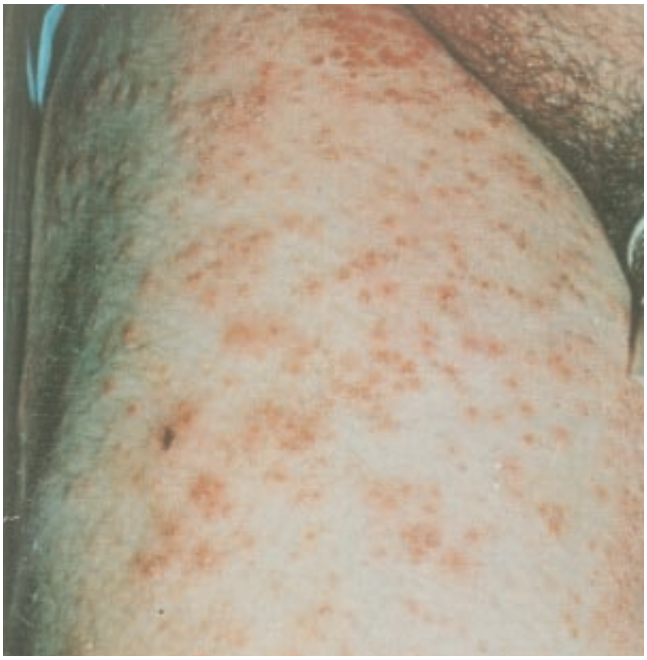


Figure 2 **The burrow of the female *Sarcoptes* frequently appears as an irregular line several millimeters to a few centimeters long in the stratum corneum.**

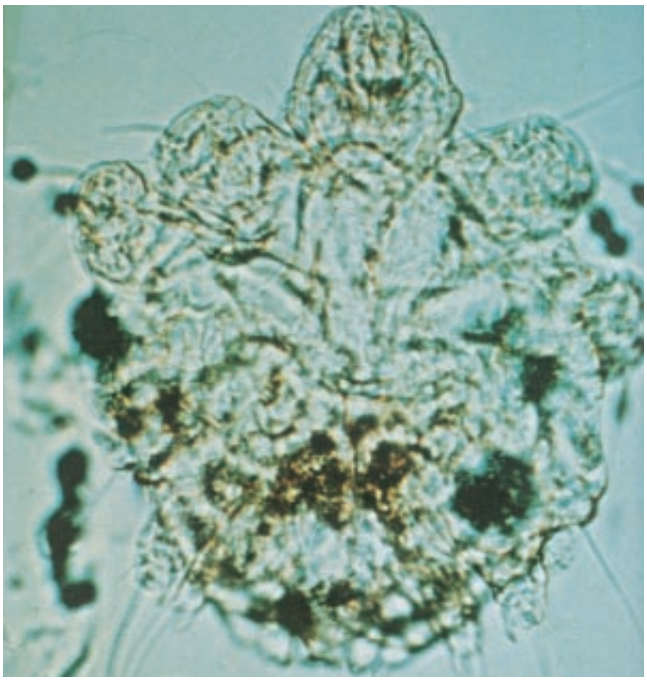


Figure 3 **Observation of the *Sarcoptes scabiei* or its eggs and feces confirms the diagnosis of scabies. Magnification is 400 times.**

DIFFERENTIAL DIAGNOSIS

Clinical differential diagnosis includes drug eruption, papular urticaria, folliculitis, atopic dermatitis, dermatitis herpetiformis, and contact dermatitis, particularly from fiberglass. Crusted scabies may be mistaken for eczema. Papular urticaria is an intensely itchy eruption caused by a hypersensitivity reaction to bites from such insects as fleas, bedbugs, and animal scabies. Lesions occur as small papules that may have a central punctum, often occurring in groups on exposed skin.

TREATMENT

Initial Treatment

After a cleansing bath or shower, the patient should allow the skin to dry and cool and then apply a scabicide over the entire body, excluding the face and scalp. Care must be taken to include skin folds such as toe webs and the skin under the nails. The medication is left in place for 8 to 12 hours, usually overnight. In the morning, the patient showers and changes clothes. All clothing worn within 2 days before treatment, in addition to towels and bed linens, is laundered in hot water or drycleaned. Chairs and mattresses should be vacuumed.

First-line treatment for uncomplicated scabies is permethrin 5% cream (Elimite, Acticin); other available scabicides include 1% lindane, or gamma benzene hexachloride, lotion or cream (Kwell, Gamene); 10% to 20% benzyl benzoate lotion; crotamiton cream (Eurax); and 6% precipitated sulfur ointment [*see Table 1*]. Ivermectin, although not approved by the Food and Drug Administration as a scabicide, has shown promising results in several trials.

Permethrin, a synthetic pyrethroid with low toxicity, has proved to be safe and effective for use in infants, children, and pregnant women.[9] Natural pyrethrins, which are derived from chrysanthemum flowers, have greater toxicity and less insecticidal activity than the synthetic pyrethroids. The low toxicity of the drug is a result of its rapid breakdown into inactive metabolites. Permethrin cream can be safely used in children and infants older than 2 months and in the elderly. Acticin is a form of permethrin in a base that has a lower viscosity to promote ease of application. Alternative scabicides that can be used for young children and pregnant or lactating women include crotamiton cream and sulfur ointment. Six-percent precipitated sulfur ointment is applied three times: at diagnosis, after 24 hours, and at 1 week. Crotamiton cream is applied for 2 or more consecutive days but is less effective than permethrin. Permethrin appears to be more effective than crotamiton in clinical and parasitic cure rates.[9]

Lindane is lipophilic and can accumulate in fat and bind to brain tissue. Toxic reactions may occur in patients who have increased absorption; infants and young children, who have a higher ratio of skin surface to body volume than do adults, are especially susceptible. Excessive treatment with lindane has been reported to cause central nervous system toxicity resulting in convulsions and seizures.[10] In 2003, the FDA issued a label change for lindane emphasizing its use as second-line therapy; treatment with lindane should be considered only if other medications have failed or cannot be tolerated (information on lindane can be found on the Internet, at http://www.fda.gov/cder/drug/infopage/lindane/default.htm). Lindane lotion should be used with caution in persons weighing less than 110 lb; its use in infants is not recommended.[11] Nevertheless, low cost, ease of application, and experience with the drug have made lindane one of

Table 1 Drug Therapy for Scabies Infestations

Route	Drug	Dosage	Relative Efficacy	Comments
Topical	Permethrin 5% cream	Single total-body application left on for 8–12 hr	First-line therapy	Effective and safe for use in infants older than 2 mo, pregnant women, and elderly patients; for a treatment failure, a second application is given 1 wk later
	Crotamiton cream	Total-body applications on 2 or more consecutive days	Alternative first-line therapy	Safe for infants and pregnant women but less effective than permethrin
	Precipitated 6% sulfur ointment	3 total-body applications: at diagnosis, 24 hr, and 1 wk	Alternative first-line therapy	Safe for infants and young children; must be compounded
	Lindane 1% lotion†	Apply for 8 hr once weekly for 2 wk	Second-line therapy; consider after failure of first-line therapy	Associated with neurotoxicity; indicated if other medications have failed or cannot be tolerated; should be used with caution in persons weighing < 110 lb; use in infants is not recommended
Oral	Ivermectin*	Single dose, 200 mg/kg	Alternative first-line therapy for uncomplicated scabies	Safe and effective in children 6 mo of age or older; advantage of oral medication is ease of use and increased compliance
		200 μg/kg initially and 2 wk later	First-line therapy for resistant scabies	Effective for treatment of crusted and resistant scabies in HIV patients; successfully used to treat outbreaks in institutional settings

*Not approved by the Food and Drug Administration for use as a scabicide.
†Gamma benzene hexachloride.

the most commonly prescribed scabicides; the FDA considers the benefits of lindane to outweigh the risks when the medication is used as directed.

Oral ivermectin is another treatment option that has been found to be safe and effective in children as young as 6 months[12] and in HIV-positive patients. A single oral dose of 200 μg/kg of ivermectin was used to treat uncomplicated scabies in 11 otherwise healthy patients and in 11 patients with HIV infection.[13] Clearing was documented by negative skin scrapings at 2 weeks and 4 weeks after treatment, and cure was achieved in all of the otherwise healthy patients and in eight of the HIV-infected patients. Advantages of an oral medication are its ease of use, lack of treatment-associated dermatitis, and increased compliance. In a comparative study of oral ivermectin and topical permethrin, a single application of permethrin was found to be superior to a single dose of ivermectin. Two doses of ivermectin were required for eradication of scabies. The lack of ovicidal activity of ivermectin may explain the difference in effectiveness between the two drugs.[14] Ivermectin is toxic to invertebrate nerve and muscle cells but may not be effective against younger stages of the parasite that do not have a developed nervous system. Permethrin acts at early stages of the life cycle of the parasite, and topical application ensures adequate drug concentration in the skin.[14]

Oral ivermectin has been used successfully to control outbreaks of scabies infestations in institutional settings[4,15]; it may prove to be the treatment of choice in nursing homes and other institutions in which topical therapy is impractical.

Topical ivermectin has also been investigated for treatment of scabies. A total of 75 patients were found to be cured, on the basis of clinical and parasitologic examinations, within 48 hours after a single application of ivermectin. Postscabies itching, which persisted in 50% of the patients, was effectively treated by a second application of ivermectin within 5 days.[16]

Postscabies Itch

Postscabies itch is thought to represent a hypersensitivity reaction to the mite or mite products and is not caused by active infestation. The pruritus may persist for weeks to months and can be treated with an antipruritic or anti-inflammatory agent, such as a low-potency to midpotency corticosteroid cream, in addition to oral antihistamines. Overtreatment with the scabicide may result in a primary irritant dermatitis that may be confused with persistent infestation. The use of bland emollients and a corticosteroid cream and avoidance of skin irritants may reduce the dermatitis. Patients should be evaluated at 4 weeks, which is the time required for viable eggs to mature to the adult stage, to determine the efficacy of treatment. If lesions are healed and no new outbreaks have occurred, the patient is considered cured.[3]

Resistant Scabies

Overuse and misuse of certain scabicides, notably lindane, have decreased their efficacy. Differing resistance patterns have been identified within a single city; more commonly, resistance corresponds to local or regional patterns.[2] If treatment failure occurs with one scabicide, the use of a different scabicide may be indicated. For example, treatment failure with topical permethrin may prompt the consideration of lindane as a second-line therapy. Pyrethroids are effective in cases of lindane-resistant scabies.[17] Treatment failures can also occur in cases involving impetiginized or crusted scabies. In these cases, treatment with the appropriate oral antibiotic is initiated along with application of the scabicide and is followed within a week by a second application of the scabicide. Keratolytics are useful as an aid in removal of the crusts.

Oral ivermectin has been used to treat resistant scabies.[18] Although it has not been approved by the FDA for this purpose, oral ivermectin is rapidly gaining acceptance as an effective therapy for resistant scabies. Combination treatment with one or two doses of ivermectin 8 days apart, in addition to permethrin and mechanical removal of subungual debris, has been advocated for outbreaks of crusted scabies in the geriatric population.[19]

In Europe, combination therapy with oral ivermectin, 200 mg/kg, and benzylbenzoate, 15% solution applied twice daily for 3 days, was found to be more effective than either agent alone for the treatment of crusted scabies in patients with HIV.[20]

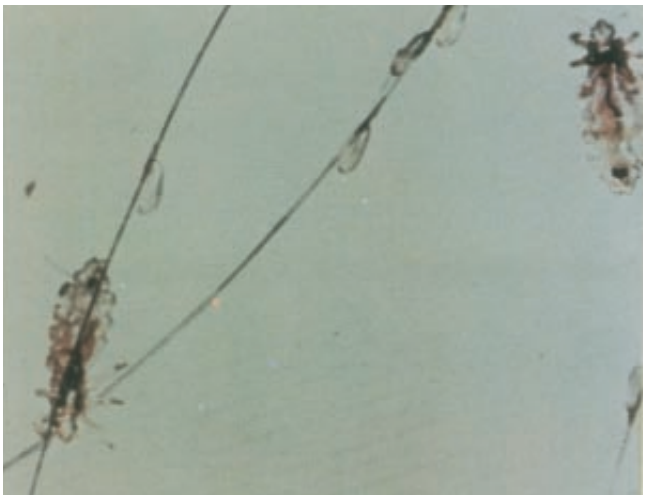


Figure 4 **Pediculosis capitis is caused by infestation of the scalp with *Pediculus humanus* var. *capitis*. Magnification is 10 times.**

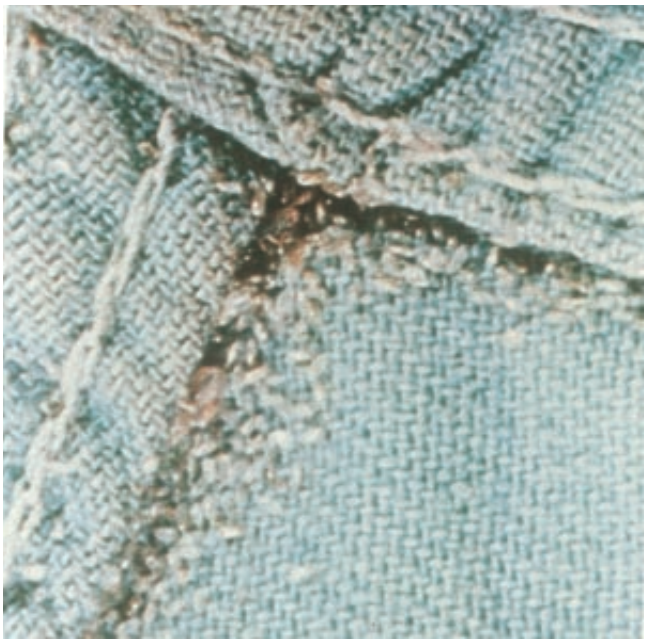


Figure 5 **Pediculosis corporis is caused by infestation with *Pediculus humanus* var. *corporis* organisms, which live in the seams of clothing.**

In the United States, a combination of oral ivermectin, total body therapy with permethrin cream, and keratolytic agents to hasten removal of crusts was used successfully to treat crusted scabies in patients with HIV.[21]

Cases of apparent resistant scabies may be the result of reinfestation. Therefore, family members and sexual partners of persons with scabies should be treated because they may be asymptomatic carriers. Scabies occurring in patients and personnel in long-term health care facilities may be difficult to diagnose and manage. In this setting, it is extremely important to treat all nursing contacts, as well as family members and other visitors of affected patients. In addition to the patients with scabies, other patients in the facility need to be assessed, and care must be coordinated to treat all affected persons simultaneously. In cases of crusted scabies, the head and neck must be treated, as well as subungual areas, which may also harbor the mites.[4]

ANIMAL SCABIES

Animal scabies is a common disorder in farm animals and domestic animals—especially dogs, in which the external ear is frequently infested with a species-specific mite. In persons who handle affected animals, an extremely pruritic papular eruption can develop that differs from ordinary scabies in several ways: distribution of lesions is proximal, with involvement of the thighs, abdomen, and forearms. Burrows are usually absent. The course is self-limited, provided there is no reexposure. Other persons in the household do not have to be treated, because human-to-human transmission of animal scabies does not occur.

The *Cheyletiella* mite is an ectoparasite that resides in the fur of dogs, cats, and rabbits. Persons who hold infested house pets, especially cats, are susceptible to a dermatitis from the mite bites. However, the mites do not live on humans, so diagnosis requires a high index of suspicion. Lesions may appear as urticarial papules, vesicles, or bullae on the arms, trunk, and legs. Cases most commonly occur in the fall or winter. An important part of the overall treatment of *Cheyletiella* infestation is treatment of the household pets by a veterinarian.[22]

Pediculosis

The three types of bloodsucking lice that cause pediculosis are *Pediculus humanus* var. *capitis* (head louse), *Pediculus humanus* var. *corporis* (body louse), and *Phthirus pubis* (pubic, or crab, louse). The first two types are closely related. The third is a separate genus and is distinctive not only in appearance and location on the body but also in its characteristic attachment to the skin for long periods. Any form of pediculosis causes intense pruritus, which is aggravated by scratching and is often complicated by secondary bacterial infection.[23]

The most common infestation is pediculosis capitis. Infestations of *P. captitis* have been reported worldwide, and an estimated 12 million cases occur annually in the United States alone.[23] Pediculosis corporis, which is usually less prevalent than pediculosis captitis, becomes widespread under conditions of overcrowding and poor sanitation or in wartime. In pediculosis capitis and pediculosis corporis, the lice may be transmitted directly from person to person or indirectly through contact with contaminated personal objects such as combs and brushes, clothing, and bedding. Pediculosis pubis (also called crabs) is usually transmitted sexually; only occasionally are the lice transmitted through contact with fomites such as contaminated bedding or toilet seats. Epidemiologic data indicate that *P. capitis* infestations are more frequent in the warmer months, whereas *P. pubis* infestations occur more frequently in the cooler months.[24]

The natural history of lice is important because it suggests specific preventive measures. The life expectancy of the organism is about 1 month. Eggs live up to 10 days but need the body heat of the host to hatch. Eggs ordinarily hatch in 7 to 8 days, and organisms reach adulthood and attain sexual reproductive capacity in 3 to 4 weeks. Lice can survive 48 hours without a blood meal.

DIAGNOSIS

Pediculosis capitis is confined to the scalp and is most prevalent in women and children. Louse infestation may present as scalp pruritus, excoriations, cervical lymphadenopathy, or conjunctivitis.[23] Examination of the itchy scalp may reveal the lice, which look like tiny black dots that are barely visible to the

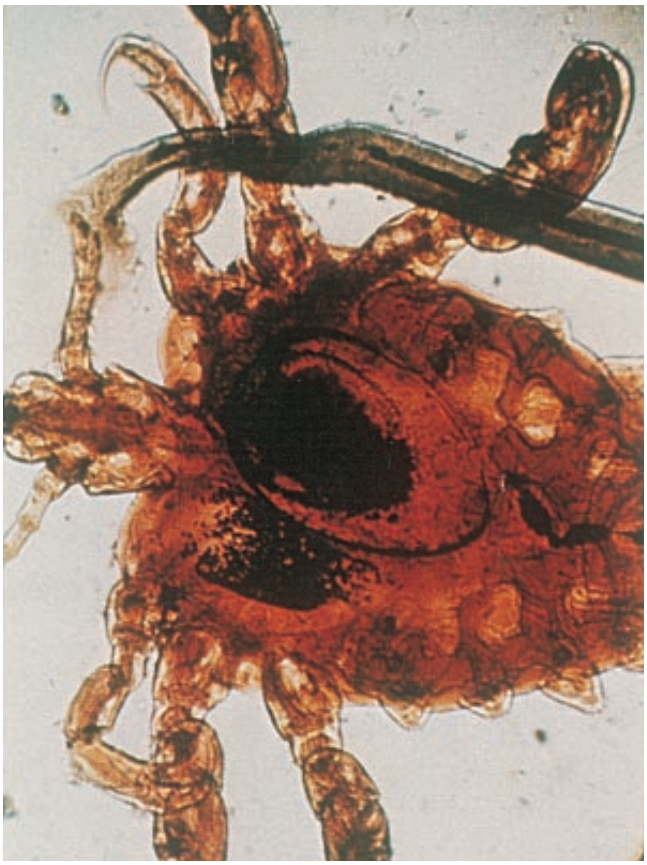


Figure 6 **Pediculosis pubis, also called crabs, is caused by infestation with *Phthirus pubis*. Magnification is 100 times.**

naked eye, and lice eggs (nits), which are white and are attached to the hair shafts [*see Figure 4*]. Except in conditions of increased warmth and high humidity, viable nits are attached close to the scalp. Those that occur several millimeters away from the surface on hairs that have grown out are empty egg cases. The hair may become matted because of exudation and secondary infection of lesions. Visual examination may not detect infestation; the use of a louse comb is recommended because it is four times more efficient than visual examination alone in the diagnosis of active pediculosis capitis.[25]

Pediculosis corporis, also called vagabond disease, affects areas of the body covered by clothing. Body lice live in the seams of clothing, and they attach to the body only to feed [*see Figure 5*]. They may serve as vectors of infectious disease under conditions of overcrowding or poor hygiene, as in wartime or during natural disasters. Characteristic lesions include erythematous macules and wheals. Lesions are most common on the shoulders, buttocks, and abdomen; furunculosis is an occasional complication. Excoriations and secondary infection may result from intense scratching. After the eggs hatch, the organisms reach adulthood in 10 days and complete their life cycle in approximately 1 month. Adult lice lay about 10 eggs a day.

Pediculosis pubis, which is caused by infestation with *Phthirus pubis* [*see Figure 6*], tends to be limited to the pubic area but occasionally affects the axillae, eyelashes, or other hairy parts of the body. Examination will reveal lice attached to the skin and lice eggs attached to the hair shafts [*see Figure 7*]. Blue macules, which are caused by the lice's sucking blood from the dermis, may be seen on the thighs or pubic area.

TREATMENT

Pediculosis Capitis

Over-the-counter preparations available for the treatment of pediculosis capitis include synergized pyrethrin products, such as RID, R&C Spray, A-200, and a 1% permethrin cream rinse (Nix) [*see Table 2*]. These products are cosmetically acceptable and require only 10 minutes to apply but may not always be effective. Repeat treatment in 7 to 10 days is advisable because the initial treatment does not kill all the eggs. If pyrethrin or permethrin fails to eradicate the infestation, the treatment of choice is malathion.[26] Malathion, which was recently reintroduced in the United States as a prescription medication for head lice, is an effective, fast-acting pediculicide and ovicide; it has not been associated with treatment resistance or notable adverse effects.[27,28] In children, proper use of malathion is safe; however, serious side effects can occur with ingestion.[29]

One of the most widely used remedies for pediculosis capitis in the United States is 1% lindane (Kwell, Gamene). However, potentially serious adverse effects associated with lindane shampoo prompted the FDA to issue a label change for this medication; treatment with lindane is now indicated only if other medications have failed or cannot be tolerated.[11,27] For the treatment of pediculosis capitis, 2 tbsp (30 ml) of the shampoo is applied to affected and adjacent areas of the scalp for at least 4 minutes, followed by thorough rinsing and drying. Adherent nits may be removed with a fine-tooth comb. Distilled white vinegar can be used to soften the nit cementing material to aid in removal of the nits.

Resistance to lindane has emerged over the past 2 decades,[30] and treatment of lice infestation has been complicated by the development of resistance to permethrin.[28] Mechanical methods of removing head lice and nits[31] and application of occlusive oils or ointments[32] have been advocated for treatment of resistant head lice. Oral ivermectin has been administered as a single dose of 12 mg (2 to 6 mg tablets), followed by a second dose 7 to 10 days lat-

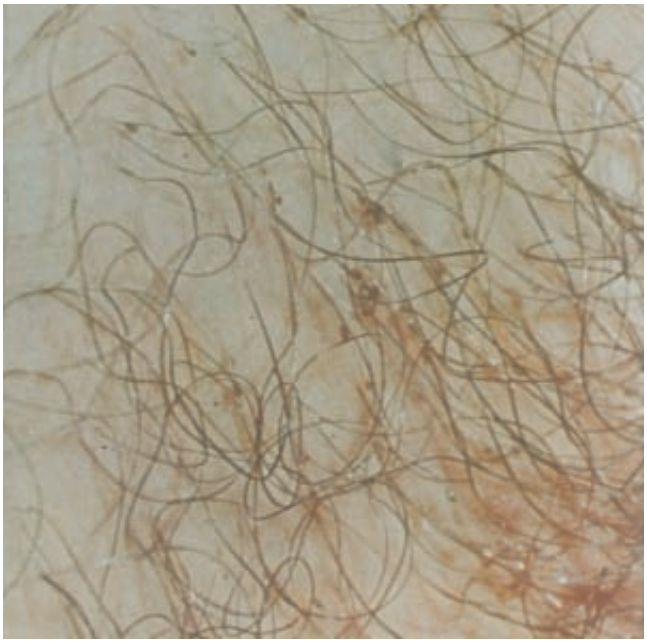


Figure 7 **Lice attached to the skin and lice eggs attached to the hair shafts can be seen on a patient with pediculosis pubis.**

Table 2 Drug Therapy for Pediculosis Infestations

Disease	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Pediculosis capitis	Pyrethrin 0.3% shampoo and permethrin 1% cream rinse	Apply shampoo for 10 min and rinse; repeat in 7–10 days	First-line therapy	Effective and safe for use in children
	Malathion 0.5% lotion	Apply for 8–12 hr daily for 1 wk	First-line therapy; most effective therapy for resistant head lice	Safe and effective in children; potential serious adverse effects if ingested
	Lindane shampoo†	2 tbsp of shampoo applied to affected and adjacent areas of scalp and allowed to stand for at least 4 min	Second-line therapy	Associated with neurotoxicity; indicated if other medications have failed or cannot be tolerated; should be used with caution in persons weighing < 110 lb; use in infants is not recommended
	Ivermectin (oral)*	Single dose of 12 mg (2–6 tablets), followed by second dose 7–10 days later	Alternative first-line therapy	Safe and effective in children 6 mo of age or older; advantage of oral medication is ease of use and increased compliance
Pediculosis pubis, pediculosis corporis	Pyrethrin 0.3% shampoo or lotion	Apply shampoo or lotion for 10 min and rinse	First-line therapy	Resistance may occur
	Lindane lotion†	1 oz applied to affected areas	Second-line therapy	Associated with neurotoxicity; caution is advised in prescribing lindane for patients with conditions that increase risk of seizure

*Not approved by the Food and Drug Administration for use as a pediculicide.
†Gamma benzene hexachloride.

er.[23] Current evidence suggests that permethrin, pyrethrin, and malathion are equally effective in the treatment of head lice; the best choice of therapy depends on local resistance.[33]

Pediculosis Pubis and Pediculosis Corpis

To treat hairy areas of the body infested with *P. pubis*, a cleansing bath or shower should first be taken and the skin dried with a towel. One ounce of lindane cream or lotion is applied to the affected and surrounding areas and left on for 12 to 24 hours. To discourage percutaneous absorption, the lotion should be applied only after the skin has become cool and dry. Lotions containing pyrethrins and piperonyl butoxide are acceptable alternatives, and their use is preferred in select patients and children (see below).[11] After another bath or shower, freshly laundered clothing should be donned; bedsheets and towels should also be changed. Lindane may be applied a second time after 1 week if infestation continues. Lindane should not be applied to the face and eyelids, because it causes irritation; eyelash infestation may be treated by local application of 0.25% physostigmine ophthalmic ointment. An alternative treatment for eyelash infestation that is effective and nonirritating is the application of a thick layer of petrolatum twice a day, followed by mechanical removal of the nits.

Neurologic complications can ensue from absorption of lindane after extensive or prolonged topical application [*see* Scabies, *above*]; severe adverse reactions have also occurred after a single use (information on lindane can be found on the Internet, at http://www.fda.gov/cder/drug/infopage/lindane/default.htm). Careful consideration should be given before prescribing lindane to patients with conditions that increase the risk of seizure (e.g., HIV infection, history of past seizure, or severe hepatic cirrhosis) or whose concomitant medications include drugs that lower risk of seizure. Alternative treatments are indicated in infants (who are especially susceptible), young children, pregnant women, and the elderly.

A combination of pyrethrins with piperonyl butoxide (RID or A-200) has been shown to be considerably less toxic than lindane in animal experiments and in clinical experience. However, this combination irritates the eyes and mucous membranes and may also cause allergic contact dermatitis in susceptible people.

General Treatment Measures

All family members should be carefully examined for pediculosis and treated, if necessary, to avoid spread or reinfection of previously treated persons. In the case of pediculosis pubis, sexual contacts should be examined and treated. Because sexually transmitted diseases are frequently present in persons infested with *P. pubis*, a serologic test for syphilis and screening for HIV are usually done. To prevent spread of pediculosis, contaminated clothing and other articles, such as towels and bedding, should be boiled, machine washed in hot water, and placed in a dryer using a 20-minute hot cycle or should be dry-cleaned. Items such as combs and brushes may be cleaned with medicated shampoo or soaked in 5% Lysol. To eradicate *P. corporis*, the patient's clothing must be put through the same decontamination process as that used for *P. pubis*. A hot iron with pressure applied, especially to the seams of clothing, may also be used to kill *P. corporis*. Systemic antibiotics should be prescribed for concomitant secondary bacterial infections such as furunculosis and impetigo, both of which are commonly associated with pediculosis capitis.

Miscellaneous Infestations

FLEA INFESTATIONS

Fleas are small (approximately 3 mm), bloodsucking, wingless ectoparasites of the insect order Siphonaptera. Fleas are medically significant because they are vectors of infectious disease [*see 7:XI Infections Due to Brucella, Francisella, Yersinia Pestis, and Bartonella and 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*]. They can also cause considerable cutaneous symptoms, particularly if the symptoms are associated with an allergic hypersensitivity reaction, as seen in papular urticaria. There are ap-

proximately 250 species of flea, 20 of which can infest humans. Two common species that infest cats and dogs are *Ctenocephalides felis* and *C. canis*. They are not host specific and can therefore infest humans as well. *Pulex irritans*, the house flea, infests humans and in most places is not a problem for pets. Flea bites appear as erythematous edematous papules with hemorrhagic puncta in clusters or groups on the lower extremities, especially on the ankles. Occasionally, vesicles and bullae appear, as well as larger urticarial lesions. Secondary impetiginization may occur because of scratching.[34]

Fleas are difficult to eradicate because of their unpredictable life cycle, which consists of egg, larva, pupa, and adult stages. The eggs are laid on the host but can drop to the ground; onto carpets, pet bedding , and furniture; and into floor cracks. Eggs hatch in 2 to 21 days into larvae. A larva molts twice and, in the third larval stage, spins a cocoon, in which it becomes a pupa. Within 7 days to 1 year or more, the adult emerges, depending on various trigger factors (e.g., a vibration caused by a nearby pet or human). The life cycle from egg to adult generally ranges from 14 to 21 days but, under ideal conditions, can be as long as 20 months.[34]

Eradication of the fleas may require consultation with a veterinarian. Pets must be treated more than once with topical agents to kill the eggs, larvae, and pupae, as well as the residual fleas. Systemic agents are available to protect pets from reinfestation. A household flea spray should be combined with a fogger to fumigate the house. A proper extermination procedure includes vacuuming the furniture and vacuuming or steam cleaning carpets or rugs. The yard should be sprayed and cleared of organic debris. Treatment of flea bites consists of cool-water compresses, application of a corticosteroid cream and an antipruritic lotion, and oral antihistamines in the case of allergic hypersensitivity reaction. Systemic antibiotics are prescribed for secondary bacterial infection.

Tungiasis

Cutaneous infestation by the sandflea *Tunga penetrans* is endemic in Central and South America, parts of Mexico, tropical Africa, Pakistan, and the west coast of India. Isolated cases have been reported in the United States, Australia, and New Zealand. Tungiasis is more prevalent in poverty-stricken areas and is associated with domestic animals such as pigs, dogs, and cattle, which serve as intermediaries in the biologic life cycle.[35] The female adult sandflea exists in sandy soils and requires a warm-blooded host to complete its life cycle. The organism penetrates the stratum corneum, resulting in erythematous nodules with a central dark spot. Common sites of skin involvement are the soles of the feet, the web spaces between fingers and toes, the ankles, the perineal area, and the buttocks.

Infestation can be prevented by wearing shoes and proper clothing and by the use of insecticides.

MYIASIS

Myiasis is caused by the larvae (maggots) of feeding flies of the order Diptera. The larvae may invade the skin primarily[36] or become secondarily implanted in a preexisting skin wound.[37] Many species of the genus *Cuterebra* can cause myiasis, but in North America, *C. cuterebra* and *C. dermatobia* cause furuncular cutaneous infestations. Mosquitoes act as vectors by transporting fly eggs from infected animals to human hosts.[38] The skin lesions appear as nonhealing single or multiple nodules on the upper trunk, usually at the site of a painful bite wound. Skin lesions may be misdiagnosed as cellulitis, boils, or sebaceous cysts. Myiasis is commonly reported in travelers to endemic areas such as Central and South America and tropical and subtropical Africa. Preventive measures include the use of insect repellents, the wearing of protective clothing to prevent mosquito bites, and the avoidance of direct skin contact with sand that may be infested with eggs. Furuncular myiasis is effectively treated by removal of the larvae by incision and drainage with debridement. Antibiotics are prescribed for secondary bacterial infection. Occlusion with such agents as liquid paraffin, lubricating jelly, and even the fatty portion of raw bacon has been suggested to cause suffocation of the larvae or migration of the larvae from the wound.[39]

CUTANEOUS LARVA MIGRANS

Cutaneous larva migrans is caused by penetration and migration of larval hookworms (usually *Ancylostoma braziliense*) within the skin. Patients are usually travelers returning from seawater beaches in tropical areas and commonly present to the dermatologist with pruritic skin lesions. The abdomen or feet are most often involved, with a characteristic eruption consisting of one or several erythematous linear to serpiginous thin lines in the skin.

Optimal management is controversial, and most treatment trials have been of low quality; however, it is generally agreed that the most effective agents in treating cutaneous larva migrans are topical or oral anthelmintics, including albendazole, thiabendazole, and ivermectin. Regimens include topical thiabendazole (10% to 15% cream) three times daily for 5 to 10 days; oral albendazole, 400 mg daily for 3 to 5 days; oral ivermectin, single dose of 12 mg; and oral thiabendazole, 50 mg/kg weekly for 2 to 3 weeks. Topical thiabendazole (10% cream) is effective and safe in children,[40,41] but it is difficult to obtain. Systemic therapy, either with albendazole or ivermectin, may be preferable to topical therapies. Oral albendazole has a high cure rate and appears to be effective in cases of multiple lesions.[42,43] Oral ivermectin is effective and reportedly safe.[43] Oral thiabendazole has been reported to cure long-standing cutaneous larva migrans, but it may cause side effects such as headaches, nausea, and vomiting.[44]

SEABATHER'S ERUPTION

Seabather's eruption, also known as sea lice by laypersons, is an acute pruritic dermatitis that occurs within 24 hours of seawater exposure and resolves spontaneously after 3 to 5 days.[45] Lesions affect areas of the skin covered by swimwear, particularly those that are subjected to pressure or friction, such as the waistline, axillae, neck, and inner thighs [*see Figure 8*]. The larvae of the thimble jellyfish *Linuche unguiculata*, which are washed ashore by ocean currents, have been identified as the cause of seabather eruption in southern Florida and the Caribbean.[46] Similar outbreaks on Long Island, New York, are thought to be caused by larvae of the sea anemone *Edwardsiella lineata*.[47] Treatment is symptomatic and includes antihistamines, topical antipruritic agents, and steroids.

SWIMMER'S ITCH

Cercarial dermatitis, known as swimmer's itch, is caused by an avian schistosome, *Microbilharzia variglandis*. The skin eruption appears approximately 12 hours after contact with seawater as a pruritic papulovesicular dermatitis on exposed skin sites.[48] The inflammatory response is attributed to dermatologic penetration by cercariae, which are the free-swimming larvae of *M. variglandis* and other bird schistosomes.

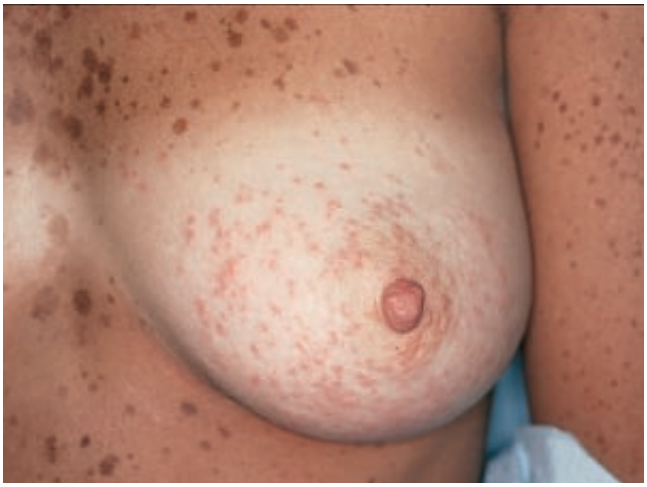


Figure 8 **Seabather's eruption is characterized by the development of pruritic papules on areas covered by the patient's bathing suit.**

Treatment is symptomatic and includes antihistamines, topical antipruritic agents, topical corticosteroids, and antibiotic treatment of superimposed bacterial infection.

CUTANEOUS AND MUCOCUTANEOUS LEISHMANIASIS

There are distinctive skin lesions associated with the cutaneous and mucocutaneous forms of leishmaniasis [*see Figure 9*]. Leishmaniasis is caused by an obligate intracellular parasite introduced by the *Phlebotomus* sandfly, which feeds on infected animals. *Leishmania braziliensis* and *L. mexicana* are the most common causes of American, or New World, leishmaniasis. *L. donovani* causes Old World leishmaniasis, which is endemic in Asia and West Africa [*see Section 7:XXXIV Protozoan Infections*]. Infection by *L. major* is the cause of cutaneous leishmaniasis in United States military personnel returning from Afghanistan and Iraq.[49]

Cutaneous leishmaniasis—the initial, or primary, form of the disease—appears as a localized, usually single, lesion involving the mouth and nose. A red-brown papule develops at the site of inoculation into a nodule that becomes verrucous or ulcerates, and satellite nodules may form. Spontaneous healing with an atrophic scar occurs in most cases. Old World leishmaniasis is usually limited to the skin, whereas New World leishmaniasis can cause mutilating mucocutaneous involvement.[50] After a period of months to years, the mucocutaneous, or secondary, form of the disease may develop, depending on host immunologic factors. Lesions in this stage range from edema of the lips and nose to perforation of the nasal cartilage. A rare form, disseminated cutaneous leishmaniasis, which has widespread nodules resembling lepromatous leprosy, may occur in immunosuppressed patients.

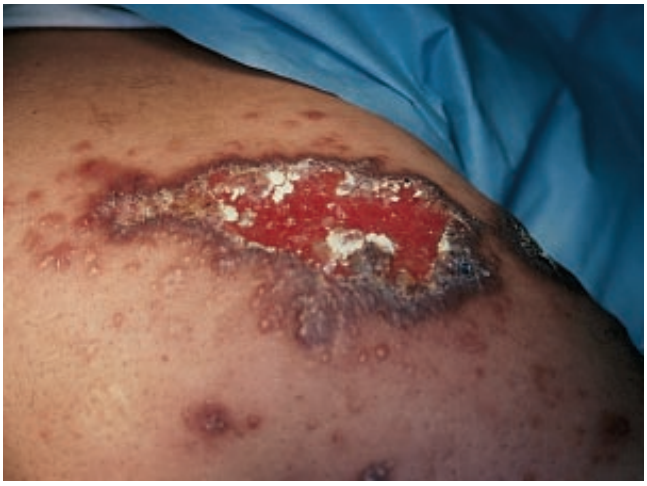


Figure 9 **Leishmaniasis can present as chronic cutaneous ulcerations.**

The differential diagnosis includes various inflammatory and neoplastic disorders, including squamous cell carcinoma. Diagnosis is made by skin biopsy with histopathologic examination. Cultures from skin biopsy may be inconclusive; PCR shows promise as a sensitive single diagnostic test.[51] Appropriate therapy depends on species identification. A pentavalent antimony compound, such as sodium stibogluconate, is the drug of choice for New World leishmaniasis, which tends to be more aggressive. Lesions acquired in the Middle East and North Africa may spontaneously involute or may respond to local therapy, including cryosurgery, heat therapy, or intralesional injection of antimonials.

Delusions of Parasitosis

Patients with delusions of parasitosis express the conviction that there are scabies, insects, lice, fleas, worms, or other vermin infesting their skin and producing a crawling, itching, or prickling sensation.[52] They may have excoriations or skin inflammation and erosions consistent with factitial dermatitis. Frequently, patients will bring small containers filled with lint, hairs, pieces of skin, fibers, or other debris for examination. Despite the lack of objective evidence for infestation—including negative results from clinical examination, microscopic examination of skin scrapings, and skin biopsy—the delusions persist. Associated underlying psychiatric disturbances may range from a phobic-obsessive state or anxiety reaction to a frank psychosis with either depression or paranoia. Not infrequently, the delusion is shared by the spouse or other family members, as in the classic folie à deux or folie à famille. The patient usually functions in a highly organized manner in other aspects of his or her life. Such patients typically resist seeking psychiatric evaluation.

Treatment with pimozide, a high-potency antipsychotic neuroleptic of the diphenylbutylpiperidine group, has been used successfully.[52] The effectiveness of the drug may be mediated by its ability to specifically block central dopamine receptors. As is characteristic of high-potency antipsychotic drugs, pimozide has fewer cardiovascular and anticholinergic effects but greater neurologic toxicity, especially with long-term use, than does low-potency antipsychotic drugs. Tardive dyskinesia, an extrapyramidal syndrome characterized by involuntary movements of facial muscles and extremities, may occur in 10% to 20% of patients on antipsychotic drugs. Other side effects may include skin discoloration, dermatitis, and blurred vision. Thorough medical and psychiatric evaluation should be obtained before antipsychotic medication is instituted.

The author has served as advisor or consultant to Amgen, Inc., Biogen Idec, Inc., Genentech, Inc., Abbott Laboratories, Ligand Pharmaceuticals, Inc., and 3M Co.

The drug ivermectin has not been approved by the FDA for uses described in this chapter.

References

1. Lucchina LC, Wilson ME, Drake LA: Dermatology and the recently returned traveler: infectious diseases with dermatologic manifestations. Int J Dermatol 36:167, 1997

2. Heukelbach J, Feldmeier H: Ectoparasites: the underestimated realm. Lancet 363:889, 2004

3. Chouela E, Abeldano A, Pellerano G, et al: Diagnosis and treatment of scabies: a practical guide. Am J Clin Dermatol 3:9, 2002

4. Scheinfeld N: Controlling scabies in institutional settings: a review of medications, treatment models, and implementation. Am J Clin Dermatol 5:31, 2004

5. Schlesinger I, Oelrich DM, Tyring SK: Crusted (Norwegian) scabies in patients with AIDS: the range of clinical presentations. South Med J 87:352, 1994

6. Head ES, Macdonald EM, Ewert A, et al: *Sarcoptes scabiei* in histopathologic sections of skin in human scabies. Arch Dermatol 126:1475, 1990

7. Argenziano G, Fabbrocini G, Delfino M: Epiluminescence microscopy: a new approach to in vivo detection of *Sarcoptes scabiei*. Arch Dermatol 133:751, 1997

8. Bezold G, Lange M, Schiener R, et al: Hidden scabies: diagnosis by polymerase chain reaction. Br J Dermatol 144:614, 2001

9. Walker GJ, Johnstone PW: Interventions for treating scabies. Cochrane Database Syst Rev (3):CD000320, 2000

10. Fischer TF: Lindane toxicity in a 24-year-old woman. Ann Emerg Med 24:972, 1994

11. Labeling changes for lindane. FDA Consum 37:6, 2003

12. Brooks PA, Grace RF: Ivermectin is better than benzyl benzoate for childhood scabies in developing countries. J Paediatr Child Health 38:401, 2002

13. Meinking TL, Taplin D, Hermida JL, et al: The treatment of scabies with ivermectin. N Engl J Med 333:26, 1995

14. Usha V, Gopalakrishnan Nair TV: A comparative study of oral ivermectin and topical permethrin cream in the treatment of scabies. J Am Acad Dermatol 42:236, 2000

15. Leppard B, Naburi AE: The use of ivermectin in controlling an outbreak of scabies in a prison. Br J Dermatol 143:520, 2000

16. Youssef MYM, Sadaka HAH, Eissa MM, et al: Topical application of ivermectin for human ectoparasites. Am J Trop Hyg 53:652, 1995

17. Purvis RS, Tyring SK: An outbreak of lindane-resistant scabies treated successfully with permethrin 5% cream. J Am Acad Dermatol 26(pt 1):1015, 1991

18. Cook AM, Romanelli F: Ivermectin for the treatment of resistant scabies. Ann Pharmacother 37:279, 2003

19. Paasch U, Haustein U-F: Management of endemic outbreaks of scabies with allethrin, permethrin, and ivermectin. Int J Dermatol 39:463, 2000

20. Alberici F, Pagani L, Ratti G, et al: Ivermectin alone or in combination with benzyl benzoate in the treatment of human immunodeficiency virus-associated scabies. Br J Dermatol 142:969, 2000

21. Taplin D, Meinking TL: Treatment of HIV-related scabies with emphasis on the efficacy of ivermectin. Semin Cutan Med Surg 16:235, 1997

22. Lee BW: *Cheyletiella* dermatitis: a report of fourteen cases. Cutis 47:111, 1991

23. Ko CJ, Elston DM: Pediculosis. J Am Acad Dermatol 50:1, 2004

24. Minouni D, Ankol OE, Gdalevich M, et al: Seasonality trends of *Pediculosis capitis* and *Phthirus pubis* in a young adult population: follow-up of 20 years. J Eur Acad Dermatol Venereol 16:257, 2002

25. Mumcuoglu KY, Friger M, Ioffe-Uspensky I, et al: Louse comb versus direct visual examination for the diagnosis of head louse infestations. Pediatr Dermatol 18:9, 2001

26. Jones KN, English JC 3rd: Review of common therapeutic options in the United States for the treatment of pediculosis capitis. Clin Infect Dis 36:1355, 2003

27. Burkhart CG: Relationship of treatment-resistant head lice to the safety and efficacy of pediculicides. Mayo Clin Proc 79:661, 2004

28. Yoon KS, Gao JR, Lee SH, et al: Permethrin-resistant human head lice, *Pediculus capitis*, and their treatment. Arch Dermatol 139:994, 2003

29. Frankowski BL: American Academy of Pediatrics guidelines for the prevention and treatment of head lice. Am J Manag Care 10:S269, 2004

30. Forrester MB, Sievert JS, Stanley SK: Epidemiology of lindane exposures for pediculosis reported to poison centers in Texas, 1998-2002. J Toxicol Clin Toxicol 42:55, 2004

31. Roberts RJ, Casey D, Morgan DA, et al: Comparison of wet combing with malathion for treatment of head lice in the UK: a pragmatic randomised controlled trial. Lancet 356:540, 2000

32. Mumcuoglu KY: Prevention and treatment of head lice in children. Paediatr Drugs 1:211, 1999

33. Dodd CS: Interventions for treating headlice. Cochrane Database Syst Rev (3)CD001165

34. Hutchins ME, Burnett JW: Fleas. Cutis 51:241, 1993

35. Campos Macias P, Mendez Sashida P: Cutaneous infestation by *Tunga penetrans*. Int J Dermatol 39:296, 2000

36. Jelinek T, Nothdurft HD, Rieder N, et al: Cutaneous myiasis: review of 13 cases in travelers from tropical countries. Int J Dermatol 34:624, 1995

37. Sherman RA: Wound myiasis in urban and suburban United States. Arch Intern Med 160:2004, 2000

38. Maier H, Honigsmann H: Furuncular myiasis caused by *Dermatobia hominis*, the human botfly. J Am Acad Dermatol 50:S26, 2004

39. Brewer TF, Wilson ME, Gonzalez MD, et al: Bacon therapy and furuncular myiasis. JAMA 270:2087, 1993

40. Jelinek T, Maiwald H, Nothdurft HD, et al: Cutaneous larva migrans in travelers: synopsis of histories, symptoms, and treatment of 98 patients. Clin Infect Dis 19:1062, 1994

41. Simon MW, Simon NP: Cutaneous larva migrans. Pediatr Emerg Care 19:350, 2003

42. Rizzitelli G, Scarabelli G, Veraldi S: Albendazole: a new therapeutic regimen in cutaneous larva migrans. Int J Dermatol 36:700, 1977

43. Caumes E, Carriere J, Datry A, et al: A randomized trial of ivermectin versus albendazole for the treatment of cutaneous larva migrans. Am J Trop Med Hyg 49:641, 1993

44. Richey TK, Gentry RH, Fitzpatrick JE, et al: Persistent cutaneous larva migrans due to *Ancylostoma* species. South Med J 89:609, 1996

45. Tomchik RS, Russell MT, Szmant AM, et al: Clinical perspectives on seabather's eruption, also known as "sea lice." JAMA 269:1669, 1993

46. Segura-Puertas L, Ramos ME, Aramburo C, et al: One *Linuche* mystery solved: all three stages of the coronate scyphomedusa *Linuche unguiculata* cause seabather's eruption. J Am Acad Dermatol 44:624, 2001

47. Freudenthal AR, Joseph PR: Seabather's eruption. N Engl J Med 329:542, 1993

48. Verbrugge LM, Rainey JJ, Reimink RL: Swimmer's itch: incidence and risk factors. Am J Public Health 94:738, 2004

49. Update: cutaneous leishmaniasis in U.S. military personnel: Southwest/Central Asia, 2002-2004. MMWR Morb Mortal Wkly Rep 53:264, 2004

50. Koff AB, Rosen T: Treatment of cutaneous leishmaniasis. J Am Acad Dermatol 31:693, 1994

51. Faber WR, Oskam L, van Gool T, et al: Value of diagnostic techniques for cutaneous leishmaniasis. J Am Acad Dermatol 49:70, 2003

52. Driscoll MS, Rothe MJ, Grant-Kels JM, et al: Delusional parasitosis: a dermatologic, psychiatric, and pharmacologic approach. J Am Acad Dermatol 29:1023, 1993

IX VESICULOBULLOUS DISEASES

Elizabeth A. Abel, m.d.
Jean-Claude Bystryn, m.d.

Vesiculobullous diseases, which number more than 50, are characterized by fluid-filled blisters in the skin. Blisters smaller than 0.5 cm are called vesicles, and larger ones are called bullae. Vesicles and bullae are reaction patterns of skin to injury and thus can be caused by a wide variety of conditions.

Most primary vesiculobullous diseases are either immunologic or genetic. They are caused by autoimmune reactions to components of skin, by allergic reactions to external agents in which the skin is the major organ system affected, and by genetic conditions in which some components of the skin are missing or abnormal. The final common pathway is disadhesion: one or more of the structures that hold the skin together separate, and a fluid-filled cavity appears. The different diseases are classified by the structure or structures affected and the mechanism or mechanisms by which disadhesion occurs [*see Table 1*]. In this subsection, several paradigmatic vesiculobullous diseases are discussed in the context of a general diagnostic approach to the patient with blistering lesions.

General Clinical Assessment

Diagnosis is based on clinical features, histologic findings, and immunologic findings. Clinical features of diagnostic importance include the following:

1. The history. Is the condition acute or chronic? Is it aggravated by sun or physical trauma?
2. The appearance of individual lesions [*see Table 2*]. Is the lesion a vesicle or bulla? Is it tense, flaccid, or umbilicated? Does the skin at the base of the blister appear normal, urticarial, or scarred? Is the border of each urticarial lesion annular or oval or is it irregular? Is the blister in the middle of urticarial plaques or on the periphery? Do more than one bulla arise from the same plaque?
3. The grouping of individual lesions. Are the lesions in closely spaced groups (as occurs in herpes simplex), or are they randomly distributed?
4. Sites of involvement. Are lesions on mucosal surfaces as well as on the skin? Are they predominantly on flexural or extensor surfaces; on the palms and soles or on the dorsa of the hands and feet; on the scalp, face, and upper torso; or on areas exposed to trauma?

The most important histologic finding is the layer of skin where the blister forms. If the blister forms in the epidermis, does it form immediately above the basal cell layer or higher up (beneath the stratum corneum)? If it forms in the basement membrane zone, is it within the lamina lucida or below the lamina densa? The precise location may be determined by immunofluorescence or by electron microscopic procedures.

The most important immunologic finding is the presence or absence of abnormal circulating or tissue-fixed antibodies to skin. These are usually detected by immunofluorescence techniques: (1) indirect immunofluorescence to detect circulating antibodies and (2) direct immunofluorescence on skin biopsy specimens to detect tissue-fixed antibodies. Recently, enzyme-linked immunosorbent assays (ELISAs) using purified antigens have become available to detect the antibodies that occur in some of the bullous diseases, such as pemphigus.

Pemphigus

definition and pathogenesis

Pemphigus is characterized by blisters that arise within the epidermis and by a loss of cohesion of the epidermal cells (acantholysis) that results in the formation of clefts above the basal cell layer. Autoantibodies directed against adhesion molecules cause epidermal keratinocytes to separate, resulting in intraepidermal bullae. There are two types of pemphigus: deep (e.g., pemphigus vulgaris) and superficial (e.g., pemphigus foliaceus). They differ in the epidermal layers that are injured, in the clinical manifestations of the diseases, and in the associated immunologic abnormalities.[1] In the deep forms, the blisters form immediately above the basal cell layer and are associated with autoantibodies to desmoglein 3; about half the cases are associated with antibodies to desmoglein 1 glycoprotein keratinocyte adhesion molecules.[2] In the superficial forms, the bullae form immediately below the stratum corneum. The superficial forms of pemphigus are associated with antibodies to desmoglein 1.

clinical features

Pemphigus Vulgaris

Pemphigus vulgaris is the most common form of pemphigus. It can develop at any age but usually occurs in persons between 30 and 60 years old. The disorder tends to affect persons of Mediterranean ancestry but can occur in persons of any ethnicity. Pemphigus is more common in persons with certain HLA allotypes. The occurrence of the disease in first-degree relatives, although rare, suggests an inherited susceptibility transferred as a dominant trait. However, other unknown factors are required for expression of the disorder in predisposed persons.[3] Studies of HLA class II alleles in Japanese patients as well as in other ethnic groups show an association with HLA-DRB1*04 and HLA-DRB1*14 in patients with pemphigus vulgaris across racial lines.[4]

Pemphigus vulgaris usually, but not invariably, begins with chronic, painful, nonhealing ulcerations in the oral cavity [*see Figure 1*]. Bullae are rarely seen because they rupture easily, leaving ulcerated bases. The ulcerations are usually multiple, superficial, and irregular in shape. Any oral mucosal surface can be involved, but the most common sites are the buccal and labial mucosae, the palate, and the gingiva. The occurrence of multiple ulcerations differentiates these lesions from ulcerated malignant tumors of the oral cavity, which are usually single. A diagnosis of pemphigus is usually considered only after lesions have been present for weeks to months.

Skin lesions can also be the initial manifestation, beginning as small fluid-filled bullae on otherwise normal-looking skin. The blisters are usually flaccid because the thin overlying epidermis cannot sustain much pressure. Bullae therefore rupture rapidly, usually in several days, and may be absent when a patient is examined. Sharply outlined, coin-sized, superficial erosions with a collarette of loose epidermis around the periphery of the erosions may appear instead. The upper chest, back,

scalp, and face are common sites of involvement, but lesions can occur on any part of the body. The condition progresses over weeks to months [*see Figure 2*]. Sites often overlooked include the periungual areas (manifested as painful, erythematous, paronychial swelling), the pharynx and larynx (pain on swallowing and hoarseness), and the nasal cavity (nasal con-

Table 1 Differentiating Features and Standard Therapy for Selected Blistering Diseases

	Disease	*Features*	*Therapy*
Epidermal	Pemphigus vulgaris Pemphigus vegetans	Chronic, painful ulcerations in the oral cavity; small, flaccid bullae *or* coin-sized superficial erosions arising from normal skin; positive Nikolsky sign; IgG and C3 at intercellular spaces; serum antidesmoglein 1 or 3 antibodies Hypertrophic proliferation of epidermis in intertriginous areas; IgG and C3 at intercellular spaces; serum antidesmoglein 3 antibodies	*Localized:* Intralesional or topical corticosteroids, low-dose systemic corticosteroids *Extensive or rapidly progressive disease:* corticosteroids, adjuvant therapy with cytotoxic and immunosuppressive agents *Refractory disease:* plasmapheresis, IVIg, pulse therapy with megadoses of I.V. methylprednisolone
	Pemphigus foliaceus Pemphigus erythematosus Fogo selvagem	Small, pruritic, crusted lesions on upper torso, face, or scalp; chronic superficial erosions; rare oral involvement; immunopathology higher in epidermis Erythematous scaly to crusted eruption on face and upper chest; lupuslike immunologic abnormalities (granular deposits of IgG and C3 at epidermal-dermal junction) Features similar to pemphigus foliaceus (primarily affects persons < 30 yr in rural areas of Brazil, Colombia, Tunisia)	*Localized:* Intralesional or topical corticosteroids, low-dose systemic corticosteroids *Extensive or rapidly progressive disease:* corticosteroids, adjuvant therapy with cytotoxic and immunosuppressive agents
	Paraneoplastic mixed bullous disease (paraneoplastic pemphigus)	Large, tense bullae; target lesions on skin; oral erosions; keratinocyte necrosis; clinical features overlap between pemphigus and erythema multiforme; subepidermal separation; IC and BMZ antibodies on direct IF	Difficult (standard treatments for autoimmune blistering diseases fail in most patients)
	Hailey-Hailey disease	Multiple vesicles on inflammatory bases in intertriginous areas and other areas subject to friction or pressure; loss of bridges between epidermal cells; no circulating or tissue-fixed autoantibodies	Involved areas kept dry and free of friction; administration of topical and systemic antibiotics; topical, intralesional corticosteroids; ablation of involved areas
Subepidermal	Bullous pemphigoid	Crops of large, tense blisters recurring from urticarial plaques on torso and flexures; negative Nikolsky sign; oral lesions (10%–25% of patients); circulating BMZ antibodies; IgG and C3 at BMZ in a linear pattern on direct IF	Administration of systemic corticosteroids at doses lower than those used for pemphigus (≤ 80 mg/day prednisone) [*see also* Bullous Pemphigoid, Treatment, *in text*]
	Cicatricial pemphigoid	Blisters on mucosal surfaces (oral cavity, esophagus, eyes) that heal with scarring, often occurring repeatedly at same site; diffuse, painful erythema and atrophy of the gingival mucosa; IgG and C3 at BMZ in a linear pattern on direct IF	Combination therapy with systemic corticosteroids and dapsone or azathioprine; long-term therapy with systemic corticosteroids, sometimes combined with immunosuppressive agents; intralesional corticosteroids
	Herpes gestationis	Pruritic urticarial plaques occurring in pregnancy (beginning around the umbilicus, spreading to abdomen and thighs); laminal blisters with linear deposits of C3 or IgG at the epidermal-dermal junction; circulating complement-fixing BMZ antibodies on indirect IF	Normally clears after delivery
	Dermatitis herpetiformis	Clusters of intensely pruritic, small, polymorphic vesicles on elbows, knees, buttocks, scapular area, and scalp; accumulations of neutrophils and eosinophils in dermal papillae; granular deposits of IgA in BMZ; no circulating antibodies to normal skin components	Administration of sulfones (dapsone, 100–200 mg/day; sulfapyridine, 1–3 g/day in divided doses; or sulfamethoxypyridazine); reduction of gluten intake
	Linear IgA dermatosis	Blisters resembling those of dermatitis herpetiformis or erythema multiforme; linear deposition of IgA in BMZ on direct IF	Administration of sulfones
	Erythema multiforme	Sudden eruption of crops of lesions on elbows, knees, hands, and feet; target papule or vesicle with halo of erythema; subepidermal edema, deep perivascular inflammatory infiltrate	Elimination of underlying causes (e.g., infectious agents, drugs); *in mild cases,* topical glucocorticoids, anti-inflammatories, antipruritics, antibiotics; *in severe cases,* prednisone 40–120 mg/day in divided doses
	Toxic epidermal necrolysis	Rapidly progressive painful denudation of epithelium (usually a drug reaction); full-thickness epidermal necrosis; absence of immune reactants within skin blood vessels; little dermal inflammation	Meticulous wound care with debridement of necrotic tissue, fluid and electrolyte replacement, and prevention of sepsis; IVIg [56]
	Staphylococcal scalded skin syndrome	Scarlatiniform eruption accompanied by skin tenderness, fever, and irritability; lack of mucous membrane involvement or target lesions	Intravenous penicillinase-resistant penicillins
	Epidermolysis bullosa	[*See* Epidermolysis Bullosa, *in text*]	Supportive therapy; counseling; promotion of wound healing; prevention of complications

BMZ—basal membrane zone IC—intercellular IF—immunofluorescence IVIg—intravenous immunoglobulin

Table 2 Pathologic Typology of Blisters[57]

Blister Type	*Mode of Formation*	*Site of Formation*	*Disease*
Subcorneal blister	Detachment of horny layer	Epidermis (subcorneal layer)	Miliaria crystallina Impetigo
Blister due to intracellular degeneration	Separation of cells from one another	Upper epidermis	Friction blisters
Spongiotic blister	Intercellular edema	Epidermis	Dermatitis (eczema) Miliaria rubra
Acantholytic blister	Dissolution of intercellular bridges	Epidermis (suprabasal layer)	Keratosis follicularis (Darier disease) Pemphigus vulgaris
		Epidermis (subcorneal layer)	Pemphigus foliaceus
Viral blister	Ballooning degeneration leading to acantholysis	Epidermis	Herpes simplex Herpes zoster Varicella
Blister due to degeneration of basal cells	Cytolysis of basal cells	Basal cell layer	Epidermolysis bullosa simplex Erythema multiforme (epidermal type)
	Loss of dermal contact by damaged basal cells	Basal cell layer	Lichen planus Lupus erythematosus
Blister due to degeneration of basement membrane zone	Damage in the structures that cause coherence of basal cells	Basement membrane zone	Bullous pemphigoid Dermatitis herpetiformis Erythema multiforme (dermal type)
Dermolytic blister	Anchoring fibrils are decreased and rudimentary	Dermis	Dystrophic epidermolysis bullosa Acquired epidermolysis bullosa

gestion and a bloody mucous discharge, particularly noticeable upon blowing the nose in the morning).

A characteristic feature of all severe active forms of pemphigus is the Nikolsky sign, in which sliding firm pressure on normal-appearing skin causes the epidermis to separate from the dermis. The Nikolsky sign is elicited most easily on clinically uninvolved skin adjacent to an active lesion.

If left untreated, the erosions and bullae of pemphigus vulgaris gradually spread, involving an increasing surface area, and can become complicated by severe infections and metabolic disturbances. Before the advent of corticosteroids, pemphigus was almost invariably fatal—approximately 75% of patients died within a year.[5] However, as better techniques have permitted the diagnosis of earlier, milder forms of the disease, the prognosis has improved significantly.[6] Mild forms may regress spontaneously, and the progression of even the most severe forms can be reversed in most cases. With treatment (see below), lesions normally heal without scarring. Most patients treated for pemphigus will enter a partial remission within 2 to 5 years. They can then be maintained lesion-free with minimal doses of corticosteroids (approximately 15 mg of prednisone daily). In a longitudinal study of outcome in 40 patients with pemphigus vulgaris, 45% entered a complete and long-term remission after 5 years and 71% after 10 years. Patients in remission remained lesion-free without any therapy.[7] The hyperpigmentation that is commonly associated with pemphigus usually resolves after several months.

In pregnancy, pemphigus appears to be associated with an increased incidence of premature delivery and fetal death.[8] The lesions of pemphigus can appear on the skin of the neonate; however, they normally resolve spontaneously in several weeks.

Pemphigus Foliaceus

Pemphigus foliaceus is the second most common form of pemphigus. It usually begins with small (approximately 1 cm), pruritic, crusted lesions resembling corn flakes on the upper torso and face. The crusts are easily removed, leaving chronic, superficial erosions.

Over weeks to months, the condition progresses, with an increasing number of lesions appearing on the upper torso, face, and scalp. In extensive cases, lesions develop over the entire body, become confluent, and can progress to an exfoliative erythroderma. In contrast to the deep forms of pemphigus, oral involvement in pemphigus foliaceus is very rare.

The prognosis of untreated pemphigus foliaceus is more favorable than that of pemphigus vulgaris. The lesions of pemphi-

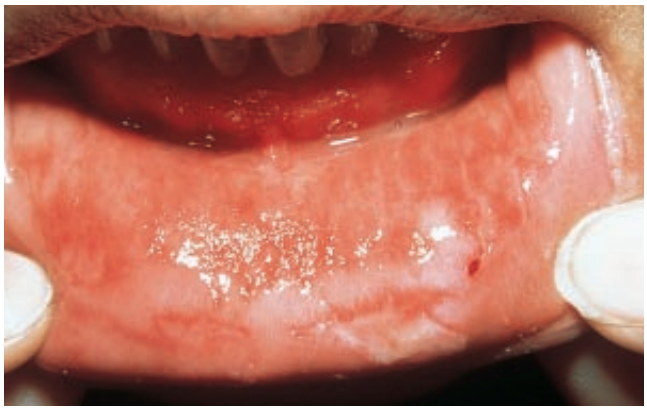


Figure 1 **Painful ulcerations or erosions in the mouth may be present many months before the onset of generalized pemphigus vulgaris.**

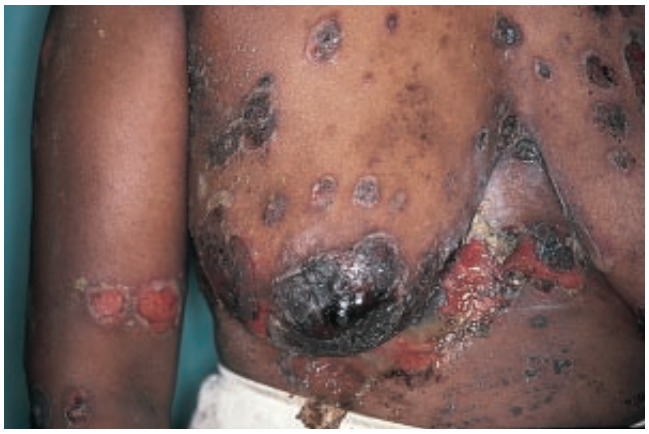


Figure 2 **Flaccid bullae of pemphigus vulgaris have broken down to form erosions and crusts, particularly under the breasts.**

gus foliaceus are not as deep, and there is less chance for infection, fluid loss, and metabolic disturbance. Although pemphigus foliaceus is less severe, the doses of medications required for control are similar to those used for pemphigus vulgaris. There are two clinical variants: pemphigus erythematosus and fogo selvagem. Pemphigus erythematosus (also known as Senear-Usher syndrome) has features of lupus erythematosus. Fogo selvagem (Portuguese for "wild fire"; also known as endemic pemphigus and Brazilian pemphigus) [*see Table 1*] may be triggered by exposure to one or more environmental antigens.[9]

Drug-Related Pemphigus

Both pemphigus vulgaris and pemphigus foliaceus can be either induced or triggered (i.e., latent disease unmasked) by certain drugs. Pemphigus that continues after a patient stops using a drug is referred to as triggered, whereas lesions that clear soon after withdrawal are referred to as induced. Although drug-related pemphigus is uncommon, its possibility must be excluded in all patients with newly diagnosed disease. The clinical, histologic,[10] and immunofluorescence abnormalities[11] of drug-induced pemphigus are similar to those of the idiopathic variety. However, pemphigus caused by drugs containing a sulfhydryl radical (thiol drugs) is clinically distinct from pemphigus caused by nonthiol drugs. The presence or absence of a sulfhydryl radical appears to influence both the type of pemphigus that is expressed and the prognosis of the drug-induced condition. Thiol drugs are more likely to induce pemphigus foliaceus, which is more likely to regress spontaneously when the drug is discontinued [*see 2:VI Cutaneous Adverse Drug Reactions*]. Nonthiol drugs are more likely to trigger pemphigus vulgaris, which can persist even after the drug is stopped. The most commonly implicated agents are thiol drugs such as penicillamine and captopril. Other responsible drugs include sulfur-containing drugs such as penicillins and cephalosporins. These undergo metabolic changes to form thiol groups and are termed masked thiol drugs. Nonthiol drugs that contain an amide group (e.g., dipyrone and enalapril) can provoke a disease that is indistinguishable from spontaneously occurring pemphigus vulgaris.[11]

Endemic Pemphigus

Epidemiologic features of fogo selvagem in rural areas of Brazil suggest that the production of pathologic antibodies to desmoglein 1 is linked to exposure to one or more environmental antigens.[9]

Paraneoplastic Mixed Bullous Disease

Paraneoplastic pemphigus is an autoimmune disease of the skin and oral mucosa that develops in patients with an underlying neoplasm. It is characterized by large, tense bullae. Unfortunately, standard treatments for autoimmune blistering diseases fail in most cases. Paraneoplastic pemphigus shares clinical features of both pemphigus and severe erythema multiforme.[12]

Hailey-Hailey Disease

Familial benign chronic pemphigus, or Hailey-Hailey disease, is an autosomal dominant disorder marked by multiple vesicles on inflammatory bases in skin subject to friction or pressure, such as intertriginous areas. In addition to pharmacologic treatment (see below), therapy includes keeping involved areas dry and free of friction.

HISTOLOGIC AND IMMUNOLOGIC FINDINGS

The diagnosis should always be confirmed by histopathologic examination and immunofluorescence studies.[13] Biopsies for pemphigus and all other bullous diseases should be performed at the edge of a lesion, so as to include clinically uninvolved adjacent skin. Acantholysis (the separation of keratinocytes from each other) is the fundamental abnormality in all forms of pemphigus.

All forms of pemphigus are associated with circulating and tissue-fixed intercellular (IC) autoantibodies that react against cell-surface keratinocyte antigens. The detection of these antibodies is very helpful in establishing the diagnosis, because they rarely appear in other conditions. Circulating IC autoantibodies are detected by indirect immunofluorescence assays on serum, and tissue-fixed IC autoantibodies are detected by direct immunofluorescence on skin biopsies. In both cases, they cause a lacelike pattern of fluorescence within the epidermis. Low titers of IC autoantibodies may also be present in burns, fungal infections, and allergic drug reactions. Antibodies against ABO blood group antigens, which are present in approximately 5% of the normal population, are the most common cause of false positive tests for IC autoantibodies. Tissue-fixed IC autoantibodies are present in lesions and adjacent normal skin in approximately 90% of patients with pemphigus and are more sensitive and specific for the diagnosis of pemphigus than are circulating IC autoantibodies. The most common autoantibodies are IgG, but IgM and IgA (with or without C3) may also be deposited.

TREATMENT

Initial Therapy

Initial therapy is determined by the extent and rate of progression of lesions. Localized, slowly progressive disease can be treated with intralesional injections of corticosteroids (triamcinolone acetonide, 10 to 20 mg/ml) or topical application of high-potency corticosteroids. New lesions that continue to appear in increasing numbers can be controlled in some cases with low-dose systemic corticosteroids (prednisone, 20 mg/day). Patients with extensive or rapidly progressive disease are treated with moderately high doses of corticosteroids (prednisone, 70 to 90 mg/day). This dose is rapidly escalated every 4 to 14 days in 50% increments until disease activity is controlled, as evidenced by an absence of new lesions and the disappearance of skin pain or itching. If the disease remains active despite high doses of corticosteroids (e.g., 120 to 160 mg/day of prednisone), one of the following approaches should be considered for rapid control:

1. Plasmapheresis, normally performed three times a week for removal of 1 to 2 L of plasma per procedure.[14]
2. Intravenous immunoglobulin (IVIg), usually given at a dosage of 400 mg/kg/day for 5 days or in higher doses for 3 days.[15] The procedure may need to be repeated every 2 to 3 weeks for several cycles. It is very expensive. The use of IVIg for the treatment of skin diseases has recently been reviewed.[16] With both IVIg and plasmapheresis it is important to concurrently administer an immunosuppressive agent such as cyclophosphamide or azathioprine to minimize rebound in the level of pemphigus antibodies,[14] and it is also important to monitor the level of these antibodies to ensure that the patient is responding to treatment.
3. Pulse therapy with megadoses of intravenous methylprednisolone, given at a dosage of 1 g/day for 5 days.[17]

No comparative studies have yet evaluated the relative effectiveness of these procedures. On the basis of such limited experience, IVIg may be preferred because it has fewer side effects than the other procedures and is associated with a significantly higher response rate. Once disease activity is controlled, the patient is maintained on the type and dose of medications required to establish control until approximately 80% of lesions are healed. Therapy should not be tapered while new lesions are appearing.

Rituximab, an anti-CD20 chimeric monoclonal antibody, is approved for use in non-Hodgkin lymphoma. However, there is a case report of partial remission from recalcitrant, life-threatening pemphigus vulgaris after treatment with rituximab.[18]

Adjuvant Therapy

The role of adjuvants in the treatment of pemphigus remains controversial. Because of a lack of controlled studies, it is not known whether the potential benefits of adjuvants outweigh the additional toxicities.[5] Indications for adjuvant therapy include the presence of relative contraindications to systemic corticosteroids, development of serious corticosteroid side effects, and repeated flares of disease activity that make it undesirable to reduce corticosteroid doses.[7] Because they require 4 to 6 weeks to become effective, adjuvants are not used to control active, rapidly progressive disease.

Adjuvant treatments for pemphigus include a variety of cytotoxic and immunosuppressive agents (e.g., cyclophosphamide, azathioprine, cyclosporine, methotrexate, and mycophenolate mofetil[19]); dapsone; anti-inflammatory agents (e.g., gold); antimalarials; and certain antibiotics (e.g., tetracycline and minocycline).

Bullous Pemphigoid

PATHOGENESIS

The immediate cause of bullous pemphigoid (BP) appears to be an autoantibody response to the 180 kd (BP180) and 230 kd (BP230) basement membrane zone antigens.[20] Passive transfer of these antibodies into animals can cause lesions of the disease[21]; anti-BP180 autoantibodies have been found to be a poor prognostic factor in a study of 94 elderly patients.[22]

CLINICAL FEATURES

BP is a nonscarring, subepidermal blistering disease that is characterized by recurrent crops of large, tense blisters arising from urticarial bases. Lesions normally appear on the torso and flexures, particularly on the inner thighs. Blisters can range in size from a few millimeters to several centimeters [*see Figure 3*]. They

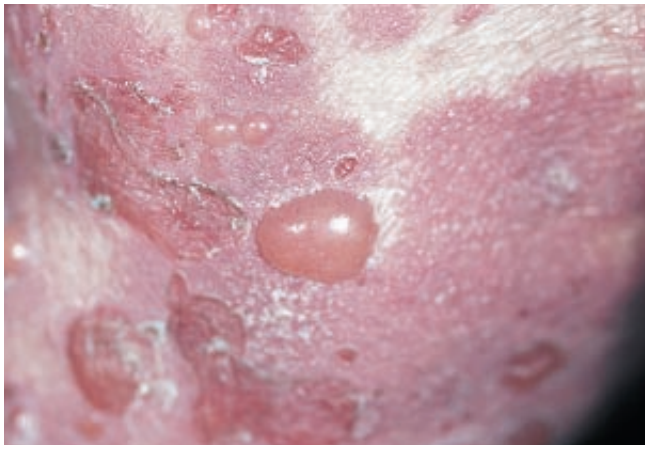


Figure 3 **Tense bullae characteristically occur in bullous pemphigoid.**

are usually filled with a clear fluid, but they can be hemorrhagic. Erosions are much less common than in pemphigus, and the Nikolsky sign is negative. A characteristic feature is that multiple bullae usually arise from large (palm-sized or larger), irregular, urticarial plaques. This is in contrast to the bullae of erythema multiforme (see below); in erythema multiforme, a single bulla arises from the center of a smaller (coin-sized) urticarial base.

In acute flares of BP, bullae may arise from normal-appearing skin. Oral lesions occur in 10% to 25% of patients; ocular involvement, however, is rare. Without treatment, the disease may become very extensive.

BP is a sporadic disease that occurs mainly in the elderly but can occur at any age and in any race. It has been reported in a 2-month-old infant.[23] Precipitating factors include trauma, burns, ionizing radiation, ultraviolet light, and certain drugs. In a case-control study of 116 incident cases, neuroleptics and diuretics—particularly aldosterone antagonists—were more commonly used by patients who developed BP than by control subjects.[24] There is still controversy as to whether BP is associated with an increased incidence of cancer[25]; however, correlations between flare in disease activity and recurrence of underlying cancer suggest such an association in individual patients.

BP is characterized by spontaneous remissions followed by flares in disease activity that can persist for years. Even without therapy, BP is often self-limited, resolving after a period of many months to years. The disease is nonetheless serious, particularly in older patients who have been treated with high doses of oral corticosteroids.[26] Mortality is low in younger persons but is significant in the elderly. In one study of patients older than 68 years, nearly a third died of the disease or complications (mainly sepsis and cardiovascular disease) within 1 year.[22]

HISTOLOGIC AND IMMUNOLOGIC FINDINGS

The earliest lesion of BP is a blister arising in the lamina lucida, between the basal membrane of keratinocytes and the lamina densa. This is associated with loss of anchoring filaments and hemidesmosomes. Histologically, there is a superficial inflammatory cell infiltrate and a subepidermal blister without necrotic keratinocytes. The infiltrate consists of lymphocytes and histiocytes and is particularly rich in eosinophils. There is no scarring.

Approximately 70% to 80% of patients with active BP have circulating antibodies to one or more basement membrane zone antigens. On direct immunofluorescence, the antibodies are de-

posited in a thin linear pattern; and on immune electron microscopy, they are present in the lamina lucida. By contrast, the antibodies to basement membrane zone antigens that are present in the skin of patients with systemic lupus erythematosus are deposited in a granular pattern.

Two less common subepidermal blistering diseases that are closely related to BP are cicatricial mucous membrane pemphigoid and herpes gestationis [*see Table 1*]. The differential diagnosis also includes dermatitis herpetiformis and acquired epidermolysis bullosa (see below). Scar formation in mucous membrane pemphigoid and acquired epidemolysis bullosa can lead to major disability.[27]

TREATMENT

Treatment of BP is generally similar to that of pemphigus.[28] The differences are as follows: (1) BP normally, but not invariably, responds to lower doses of systemic corticosteroids (alone or combined with other oral or topical agents), with most patients improving on prednisone at a dosage of 80 mg/day or less; (2) in an open prospective study of 18 cases, low-dose methotrexate was shown to be effective for maintenance of clinical remission induced by initial short-term use of potent topical steroids[29]; and (3) BP is more likely to respond to dapsone[30] or to the combination of tetracycline and niacinamide.[31,32] Considering that the prognosis of untreated BP is better than that of pemphigus, side effects of treatment are of greater concern. Two small studies of severe ocular mucous membrane pemphigoid suggest that this condition responds more favorably to treatment with cyclophosphamide combined with prednisone, whereas dapsone suppresses some cases of mild to moderate disease.[27]

Dermatitis Herpetiformis

Dermatitis herpetiformis (DH) is a rare vesiculobullous disease characterized by intensely pruritic, small vesicles that are grouped in small clusters and typically appear on the extensor aspects of the extremities and on the buttocks, scalp, and back. The condition is believed to be an immune-mediated disorder and is associated with abnormal granular deposits of IgA at the basement membrane zone and with asymptomatic, gluten-sensitive, spruelike enteropathy. The disease is chronic, with periods of exacerbation and remission. Lesions may clear if patients follow a strict gluten-free diet. Linear IgA dermatosis [*see Table 1*] is an uncommon subepidermal blistering disease that may clinically resemble DH or erythema multiforme (see below).

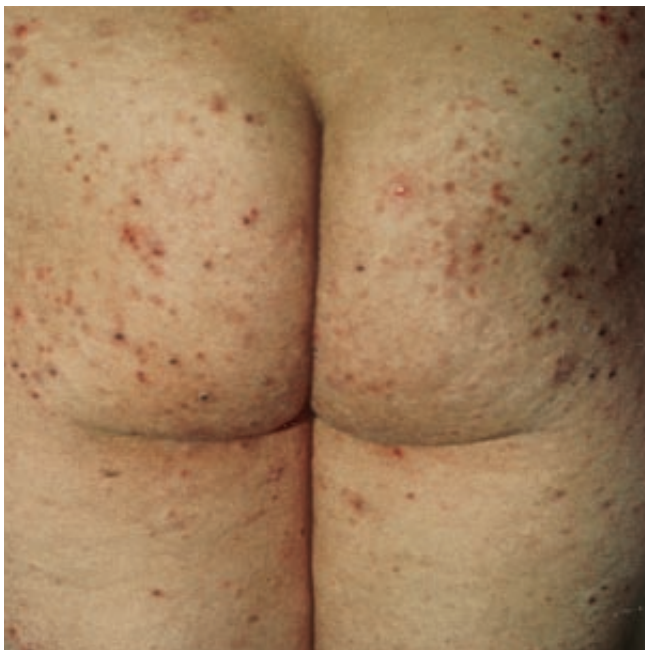


Figure 4 **Dermatitis herpetiformis, an extremely pruritic eruption, commonly presents as excoriated, grouped papulovesicles, often in a symmetrical distribution.**

PATHOGENESIS

The cause of DH is unknown. It may be related to gluten-sensitive celiac disease; there is a strong association between the two conditions, and they share a similar genetic basis (both are associated with HLA-B8 and HLA-DR3). DH is thought to result from an abnormal IgA immune response to an unidentified antigen (possibly found in gluten) that contacts the gut. Skin lesions may result from deposition of immune complexes against this antigen in skin.

CLINICAL FEATURES

Skin lesions of DH are polymorphic. They usually begin as small, very pruritic urticarial papules or vesicles that are grouped in a herpetiform pattern [*see Figure 4*]. Actual vesicles or other primary lesions are rarely seen because they are excoriated by patients' scratching. The distribution of lesions is characteristic: they occur most commonly on the elbows, knees, buttocks, scapular area, and scalp. Sometimes, lesions are scattered over the entire body. The lesions tend to appear suddenly and symmetrically, sometimes after ingestion of large amounts of gluten. Lesions heal, leaving hyperpigmentation; scarring may result from scratching or secondary infection. Involvement of mucous membranes is rare.

The disease is twice as common in men as in women. It predominantly affects persons between the ages of 20 and 50 years. There may be an associated patchy duodenal and jejunal atrophy that resembles the gluten-sensitive enteropathy of adult celiac disease.[33,34] The enteropathy is usually asymptomatic and, like celiac disease, responds to gluten restriction. Because celiac disease is associated with gastrointestinal lymphoma, there is concern that the same may be true for DH. However, although lymphomas of the small intestine have been reported in DH,[35] the association appears to be rare.

HISTOLOGIC AND IMMUNOLOGIC FINDINGS

Two characteristic laboratory features of DH are used for diagnosis. First, the disease is characterized histologically by accumulations of neutrophils and eosinophils in microabscesses at the tips of dermal papillae. In more severe cases, edema appears and can progress to subepidermal blisters appearing just below the lamina densa. Secondly, granular deposits of IgA are found at the basement membrane zone in almost all patients. These are often associated with granular deposits of C3 and, occasionally, of IgG and IgM. When found alone, IgA is one of the most sensitive and specific diagnostic markers for DH. When IgA is found with deposits of IgG, IgM, or C3, immune complex vasculitis and systemic lupus erythematosus are added to the differential diagnosis. Although basement membrane zone deposits of IgA alone also occur in linear IgA disease,[36] the deposits in that condition are linear rather than granular. There are no circulating antibodies to normal skin components in DH.

TREATMENT

DH responds rapidly and dramatically to sulfones. Dapsone at a dosage of 100 to 200 mg/day is most commonly used for treatment. Glucose-6-phosphate dehydrogenase (G6PD) deficiency must be excluded before starting therapy, because lack of this enzyme can result in severe drug-induced anemia. Sulfapyridine at a dosage of 1 to 3 g/day in divided doses (or sulfamethoxypyridazine) can be used in patients who cannot tolerate dapsone. Doses of these drugs are gradually reduced to the lowest amount that will suppress pruritus and development of new lesions. As indicated, patients also respond to a gluten-free diet; however, such diets are difficult to follow. Nevertheless, even a partial decrease in gluten intake will result in a decreased requirement for sulfones and should therefore be encouraged.

Erythema Multiforme

Erythema multiforme is an acute, recurrent, self-limiting disease that affects all age groups and races. It is characterized by the sudden eruption of crops of lesions, which represent a cell-mediated hypersensitivity reaction of the skin and mucous membranes to a variety of precipitating factors, including infectious agents and drugs [*see Table 3*].[37] Recent or recurrent infection with herpes simplex virus is a principal risk factor for erythema multiforme.[38]

CLINICAL FEATURES

Lesions may be localized or widespread and may affect both the skin and the mucous membranes. The eruption often occurs bilaterally and symmetrically on the extensor surfaces of the extremities and on both the dorsal and the volar areas of the hands and feet. Lesions vary from well-defined, red or purple, edematous macules and papules to vesicular or bullous lesions that may ulcerate, encrust, erode, and become infected. A target lesion consisting of a papule or vesicle surrounded by a region of normal skin and a halo of erythema at the periphery [*see Figure 5*] is characteristic.

Stevens-Johnson syndrome is a severe form of erythema multiforme that is usually disseminated, fulminant, and multisystemic [*see Figure 6*]. The syndrome may be accompanied by high fever, malaise, chills, headache, tachycardia, tachypnea, and prostration. Drugs are more commonly the underlying etiologic agent than infection. Some of these include long-acting sulfonamides (particularly trimethoprim-sulfamethoxazole), anticonvulsants, barbiturates, and nonsteroidal anti-inflammatory drugs. The mucous membranes in the mouth, the anus, and the vagina contain round or oval erythematous macules that form vesicles, bullae, and ulcers. Ocular lesions are bilateral yellowish-gray papules that often ulcerate and become secondarily infected, resulting in conjunctivitis. Ocular involvement has produced blindness.

Toxic epidermal necrolysis (TEN) has a potentially fatal outcome because of detachment of large areas of epidermis. TEN is considered by some to be a form of erythema multiforme, usually a reaction to medication. However, the absence of immune reactants within the blood vessels in the skin and the paucity of dermal inflammation have led other researchers to consider TEN a separate disease.

Staphylococcal scalded skin syndrome also causes large areas of epidermal necrosis. This syndrome, which results from toxins produced by *Staphylococcus aureus*,[39] is sometimes confused with TEN [*see Table 1*].

Table 3 Precipitating Factors in Erythema Multiforme

Viral diseases	Herpes simplex Hepatitis Influenza A Vaccinia Mumps
Fungal diseases	Dermatophytoses Histoplasmosis Coccidioidomycosis
Bacterial diseases	Hemolytic streptococcal infections Tuberculosis Leprosy Typhoid
Collagen vascular disorders	Rheumatoid arthritis Systemic lupus erythematosus Dermatomyositis Allergic vasculitis Polyarteritis nodosa
Malignant tumors	Carcinoma Lymphoma after radiation therapy
Hormonal changes	Pregnancy Menstruation
Drugs	Penicillins Sulfonamides Barbiturates Salicylates Halogens Phenolphthalein
Miscellaneous	*Rhus* dermatitis Dental extractions *Mycoplasma pneumoniae* infection

HISTOLOGIC AND IMMUNOLOGIC FINDINGS

Characteristic cutaneous histologic findings of erythema multiforme include subepidermal edema, bulla formation, epidermal cell necrosis, and a deep perivascular inflammatory infiltrate composed of mononuclear cells involving vessels in the upper dermis. The chemokine profile, with dominance of lymphocytic attractant chemokines at the dermoepidermal junction, is a feature of the interface dermatitis.[40] There are no specific immunofluorescence findings and no circulating antibodies, although direct immunofluorescence may show granular deposits of C3 and fibrin at the dermoepidermal junction and deposits of IgM, C3, and fibrin in the dermal blood vessels.

In vitro studies suggest that different immunopathogenic processes may be involved in herpes-mediated erythema multiforme and the drug-mediated forms of the disease.[41]

TREATMENT

Erythema multiforme eruptions may recur without warning, despite preventive measures. It is therefore important to identify and eliminate underlying causes. Mild cases are treated symptomatically with topical glucocorticoids and topical anti-inflammatory, antipruritic, or antibiotic preparations. Oral acyclovir may be effective in the prophylaxis of recurrent postherpetic erythema multiforme. In more severe cases, treatment with prednisone, 40 to 120 mg/day in divided doses, is indicated. If the eyes

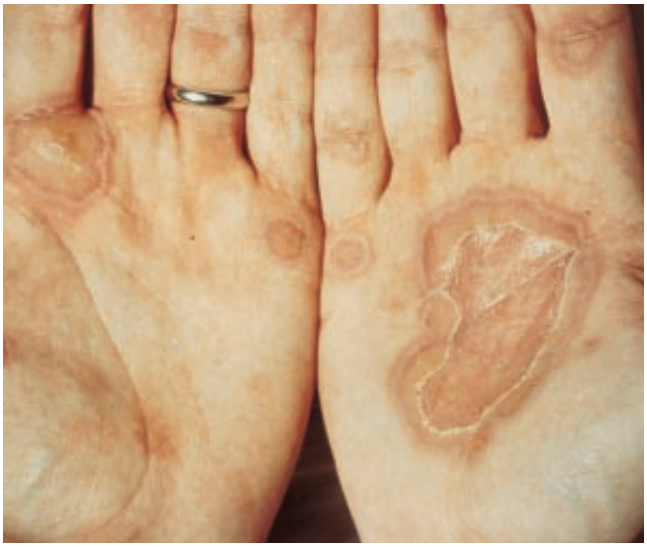


Figure 5 **Target lesions are characteristic of erythema multiforme.**

are involved, prompt ophthalmologic consultation should be obtained. Patients with large areas of epidermal necrosis (e.g., those with Stevens-Johnson syndrome) may require specialized intensive care, such as in a burn unit.

Early treatment with high-dose IVIg has been reported to be safe and effective in improving survival of patients with TEN.[42] However, there is no standard treatment of TEN that can be used as a basis for comparative studies.[43]

Epidermolysis Bullosa

Epidermolysis bullosa (EB) comprises a group of genetically based disorders with a prevalence of approximately one in 500,000 persons. There are more than 20 different phenotypes of EB, which may be inherited as an autosomal recessive trait. These disorders are characterized by blistering and erosions that arise after minor skin trauma or friction and heal with or without scarring. Extent of involvement ranges from localized blisters (e.g., on the palms and soles) to severe widespread sloughing of the skin, with a risk of severe morbidity and mortality from secondary infection, fluid and electrolyte imbalance, anemia, or other complications.

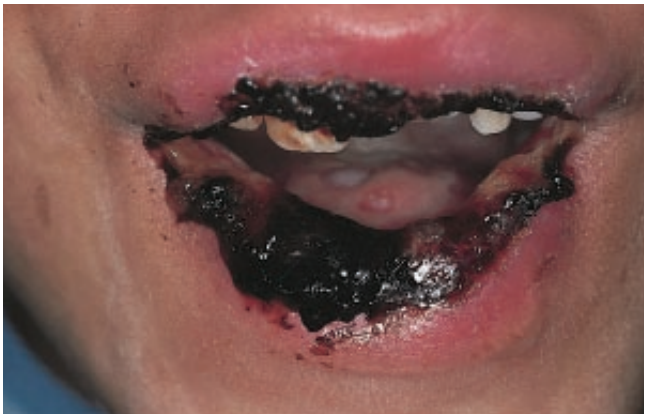


Figure 6 **Stevens-Johnson syndrome is a fulminating form of erythema multiforme associated with marked mucocutaneous involvement, eye involvement, and severe constitutional symptoms.**

EB is classified primarily on the basis of an ultrastructural level of skin cleavage in the basement membrane zone [*see Figure 7*]. Three major subtypes include EB simplex or epidermolytic (intraepidermal), junctional EB (intra–lamina lucida), and dystrophic or dermolytic EB (sub–lamina densa). Electron microscopy examination localizes the lesions to a specific layer.[44] Because this technology may not be widely available, immunofluorescence mapping with monoclonal antibodies can be used to target components of the basement membrane layers such as BP antigen (basal cell layer), laminin (lamina lucida), and type IV collagen (lamina densa).[45] The prenatal diagnosis may be made by immunocytochemical probes for antigenic components of the basement membrane in fetal skin biopsy, such as in the junctional EB pyloric atresia syndrome.[46]

EPIDERMOLYSIS BULLOSA SIMPLEX

There are three major forms of EB simplex.[47] The most common type is a mild autosomal dominant form that appears at birth or shortly thereafter as either localized or generalized blisters that do not usually result in scarring. A second type is Weber-Cockayne disease, which can be either localized or generalized. In the localized form, blisters appear acrally on the palms and soles during childhood or adolescence. In the generalized form, disease activity is usually greater in a warm climate.

The Dowling-Meara variant (EB herpetiformis) is a less common form of EB simplex that presents as severe generalized blistering in infancy; it resembles recessive junctional and dystrophic EB. EB herpetiformis becomes less severe with age.

JUNCTIONAL EPIDERMOLYSIS BULLOSA

Junctional EB is a recessively inherited group of disorders that exhibit a decreased number of hemidesmosomes and hypoplasia of hemidesmosomes, as revealed by electron microscopy, and separation at the level of the lamina lucida. Mucosal involvement and dystrophic nails are common. The most severe form, EB letalis, occurs within the first few days or months of life and has a high mortality. Patients with EB letalis have a high incidence of respiratory arrest at an early age because of laryngeal and tracheal involvement. Less severe forms of junctional EB exhibit severe generalized blistering at birth that gradually improves. Esophageal strictures may develop.

DYSTROPHIC EPIDERMOLYSIS BULLOSA

There are two forms of dystrophic EB that are inherited in an autosomal dominant fashion. Hyperplastic EB dystrophica (Cockayne-Touraine syndrome) appears in early infancy or childhood as serosanguineous blisters, predominantly on extensor aspects of the lower extremities, in association with nail dystrophy. The albopapuloid type of EB dystrophica is characterized by white papules that develop during adolescence on the trunk or extremities; however, blistering is present in the perinatal period. In both forms, ultrastructural examination reveals sublaminal dermal separation, with abnormalities in anchoring fibrils or a decrease in their number.

Recessive forms of EB dystrophica appear during the neonatal period as severe serosanguineous blistering that is either localized to sites of skin trauma or generalized. Milium formation is uncommon, but lesional scarring may result. Other complications include dental abnormalities, nail dystrophy or loss, digital fusion, flexion contractures, and esophageal strictures [*see Figure 8*]. Growth retardation, malnutrition, and chronic anemia also occur. Patients with recessive EB dystrophica are at increased

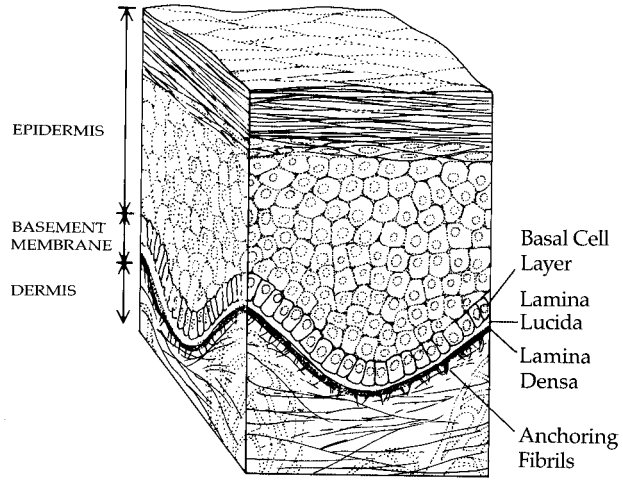


Figure 7 **Three major forms of epidermolysis bullosa (EB) have been recognized: EB simplex, in which a split occurs within the basal cell layer; junctional EB, which is characterized by separation within the lamina lucida; and dystrophic EB, in which separation occurs below the basement membrane zone.**

risk for squamous cell carcinoma, with a high incidence of fatal metastases.

Prenatal diagnosis of recessive dystrophic EB may be made by fetoscopy and skin biopsy; ultrastructural analysis of the tissue reveals dermolytic blister formation. An alternative method for prenatal diagnosis of recessive dystrophic EB involves testing of chorionic villus samples for mutation and haplotype analysis in the type VII collagen gene.[48]

Supportive treatment of EB is directed toward promotion of wound healing and prevention of complications. Daily skin care may include wet dressings or whirlpool baths, antibiotic ointment, and nonadhesive dressings, such as fine-mesh gauze (N-terface). A multidisciplinary approach that includes genetic counseling, psychological or psychiatric counseling, and support systems for the patient and family is essential, particularly for managing the severe forms of the disease.

A national registry has been established by the Dystrophic Epidermolysis Bullosa Research Association of America (http://www.debra.org) to collect epidemiologic data, to assess economic and social aspects of EB, and to register patients willing to participate in various research protocols.

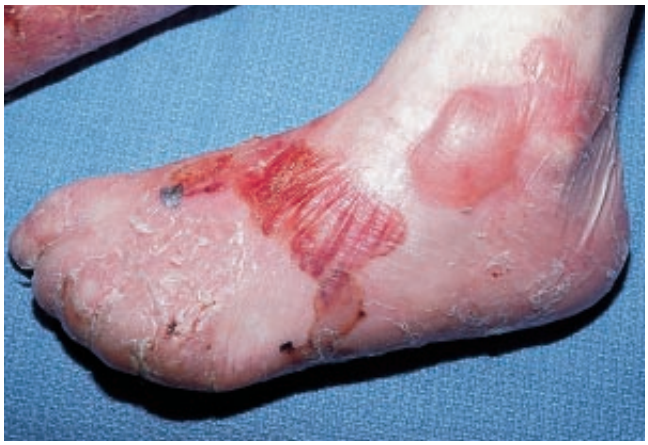


Figure 8 **Recessive dystrophic epidermolysis bullosa may cause severe scarring and syndactyly.**

ACQUIRED EPIDERMOLYSIS BULLOSA

Acquired epidermolysis bullosa, or epidermolysis bullosa acquisita (EBA), is a trauma-induced blistering disorder in adults who have no genetic basis for disease. Both circulating and tissue-bound IgG anti–basement membrane zone antibodies may be demonstrated by immunohistology. The blisters develop below the epidermis and heal with atrophic scars and malformation. They are usually confined to the extremities at sites of mechanical trauma. Oral lesions and nail dystrophy may be associated with EBA. Underlying malignant, autoimmune, and inflammatory diseases may be associated with this condition. The presence of ulcerative colitis or Crohn disease in approximately 30% of cases suggests that EBA should be included among the extraintestinal manifestations of inflammatory bowel disease.[49]

The diagnosis is made by excluding other bullous disorders, particularly BP (see above) and porphyria cutanea tarda. Immunoelectron microscopy may be used as an additional diagnostic aid, although this technique may not be widely available. Direct immunofluorescence with the use of salt-split skin to separate the lamina lucida aids in the differential diagnosis. With this method, the IgG antibodies appear on the dermal side of the split specimens in EBA and on the epidermal side in pemphigoid.[50] The antigen of EBA has been identified as the globular carboxyl terminus of type VII procollagen,[51] a major constituent of anchoring fibrils.[12] EBA may also be triggered by certain drugs, such as penicillin, cephalosporins, diclofenac, and captopril.[52]

Differential Diagnosis of Vesiculobullous Disorders

The major forms of bullous diseases occurring on an autoimmune or inherited basis have been discussed. The differential diagnosis includes a number of additional conditions in which vesicles or bullae are less common or appear secondary to other disease processes.

Acantholytic blisters occur in keratosis follicularis (Darier disease) as well as in pemphigus. Such blisters are a histologic rather than a clinical finding. Darier disease is an autosomal dominant disorder that manifests as greasy papules and plaques on seborrheic areas and in the flexures; almost all patients have nail abnormalities. Unlike pemphigus vulgaris, Darier disease is most effectively treated with oral retinoids.[53]

A fixed drug eruption may produce localized bullae that appear after ingestion of a particular drug [*see 2:VI Cutaneous Adverse Drug Reactions*]. Eczematous dermatitis results in spongiotic vesicles caused by intercellular edema [*see 2:IV Eczematous Disorders, Atopic Dermatitis, and Ichthyoses*]. This is manifested clinically by large bullae in acute allergic contact dermatitis triggered by poison ivy or poison oak. Systemic lupus erythematosus [*see 15:IV Systemic Lupus Erythematosus*] occasionally produces bullae by causing degeneration of basal cells.

A bullous eruption on the dorsa of the hands and other sun-exposed sites in patients receiving long-term hemodialysis may resemble porphyria cutanea tarda [*see 9:V The Porphyrias*].[54] Porphyrin levels are usually within normal limits. Intraepidermal or subepidermal bullae, primarily on the extremities, may be a cutaneous sign of diabetes mellitus.[55] Bacterial infections of the skin, such as impetigo, may be associated with subcorneal bulla formation. Bullae may occur on the feet in patients with severe dermatophytosis.

Various viral infections, including varicella (chickenpox), herpes simplex, and herpes zoster, also must be considered in the differential diagnosis [*see 2:VII Fungal, Bacterial, and Viral Infections of the Skin, 7:XXVI Herpesvirus Infections, and 7:XXXIII HIV and AIDS*]. Lastly, blisters from physical trauma, burns, or cold must also be considered.

Elizabeth A. Abel, M.D., has received research support from Allergan Inc. and is a consultant for Centocor, Inc. Jean-Claude Bystryn, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Korman NJ, Eyre RW, Klaus-Kovtun V, et al: Demonstration of an adhering-junction molecule (plakoglobin) in the autoantigens of pemphigus foliaceus and pemphigus vulgaris. N Engl J Med 321:631, 1989

2. Amagai M: Desmoglein as a target in autoimmunity and infection. J Am Acad Dermatol 48:244, 2003

3. Starzycki Z, Chorzelski TP, Jablonska S: Familial pemphigus vulgaris in mother and daughter. Int J Dermatol 37:211, 1998

4. Miyagawa S, Higashimine I, Iida T, et al: HLA-DRB1*04 and DRB1*14 alleles are associated with susceptibility to pemphigus among Japanese. J Invest Dermatol 109:615, 1997

5. Bystryn JC, Steinmen NM: The adjuvant therapy of pemphigus: an update. Arch Dermatol 132:203, 1996

6. Ljubojevic S, Lipozencic J, Brenner S, et al: Pemphigus vulgaris: a review of treatment over a 19-year period. J Eur Acad Dermatol Venereol 16:599, 2002

7. Herbst A, Bystryn JC: Remissions in pemphigus. J Invest Dermatol 106:850, 1996

8. Ruach M, Ohel G, Rahav D, et al: Pemphigus vulgaris and pregnancy. Obstet Gynecol Surv 50:755, 1995

9. Warren SJ, Lin MS, Giudice GJ, et al: The prevalence of antibodies against desmoglein 1 in endemic pemphigus foliaceus in Brazil. N Engl J Med 343:23, 2000

10. Landau M, Brenner S: Histopathologic findings in drug-induced pemphigus. Am J Dermatopathol 19:411, 1997

11. Brenner S, Bialy-Golan A, Anhalt GJ: Recognition of pemphigus antigens in drug-induced pemphigus vulgaris and pemphigus foliaceus. J Am Acad Dermatol 36:919, 1997

12. Anhalt GJ, Kim SC, Stanley JR, et al: Paraneoplastic pemphigus: an autoimmune mucocutaneous disease associated with neoplasia. N Engl J Med 323:1729, 1990

13. Mutasim DF, Adams BB: Immunofluorescence in dermatology. J Am Acad Dermatol 45:803, 2001

14. Turner MS, Sutton D, Sauder DN: The use of plasmapheresis and immunosuppression in the treatment of pemphigus vulgaris. J Am Acad Dermatol 43:1058, 2000

15. Bystryn JC, Jiao D, Natow S: Treatment of pemphigus with intravenous immunoglobulin. J Am Acad Dermatol 47:358, 2002

16. Dahl MV, Bridges AG: Intravenous immune globulin: fighting antibodies with antibodies. J Am Acad Dermatol 45:690, 2001

17. Werth VP: Treatment of pemphigus vulgaris with brief, high-dose intravenous glucocorticoids. Arch Dermatol 132:1435, 1996

18. Salopek TG, Logsetty S, Tredget EE: Anti-CD20 chimeric monoclonal antibody (rituximab) for the treatment of recalcitrant, life-threatening pemphigus vulgaris with implications in the pathogenesis of the disorder. J Am Acad Dermatol 47:785, 2002

19. Chams-Davatchi C, Nonahal Azar R, Daneshpazooh M, et al: Open trial of mycophenolate mofetil in the treatment of resistant pemphigus vulgaris. Ann Dermatol Venereol 129:23, 2002

20. Moll R, Moll I: Epidermal adhesion molecules and basement membrane components as target structures of autoimmunity. Virchows Arch 432:487, 1998

21. Lin MS, Mascaro JM Jr, Liu Z, et al: The desmosome and hemidesmosome in cutaneous autoimmunity. Clin Exp Immunol 107(suppl 1):9, 1997

22. Bernard P, Bedane C, Bonnetblanc JM: Anti-BP180 autoantibodies as a marker of poor prognosis in bullous pemphigoid: a cohort analysis of 94 elderly patients. Br J Dermatol 136:694, 1997

23. Cunha PR, Thomazeski PV, Hipolito E, et al: Bullous pemphigoid in a 2-month-old infant. Int J Dermatol 37:935, 1998

24. Bastuji-Garin S, Joly P, Picard-Dahan C, et al: Drugs associated with bullous pemphigoid: a case-control study. Arch Dermatol 132:272, 1996

25. Ogawa H, Sakuma M, Morioka S, et al: The incidence of internal malignancies in pemphigus and bullous pemphigoid in Japan. J Dermatol Sci 9:136, 1995

26. Joly P, Roujeau JC, Benichou J, et al: A comparison of oral and topical corticosteroids in patients with bullous pemphigoid. N Engl J Med 346:321, 2002

27. Kirtschig G, Murrell D, Wojnarowska F, et al: Interventions for mucous membrane pemphigoid/cicatricid pemphigoid and epidermolysis bullosa acquisita: a systemic literature review. Arch Dermatol 138:380, 2002

28. Yancey KB, Egan CA: Pemphigoid: Clinical, histologic, immunopathologic, and therapeutic considerations. JAMA 284:350, 2000

29. Dereure O, Bessis D, Guillot B, et al: Treatment of bullous pemphigoid by low-dose methotrexate associated with short-term potent topical steroids: an open prospective study of 18 cases. Arch Dermatol 138:1255, 2002

30. Bouscarat F, Chosidow O, Picard-Dahan C, et al: Treatment of bullous pemphigoid with dapsone: retrospective study of thirty-six cases. J Am Acad Dermatol 34:683, 1996

31. Hornschuh B, Hamm H, Wever S, et al: Treatment of 16 patients with bullous pemphigoid with oral tetracycline and niacinamide and topical clobetasol. J Am Acad Dermatol 36:101, 1997

32. Thornfeldt CR, Menkes AW: Bullous pemphigoid controlled by tetracycline. J Am Acad Dermatol 16:305, 1987

33. Brow JR, Parker F, Weinstein WM, et al: The small intestinal mucosa in dermatitis herpetiformis: I. Severity and distribution of the small intestinal lesion and associated malabsorption. Gastroenterology 60:355, 1971

34. Katz SI, Hall RP III, Lawley TJ, et al: Dermatitis herpetiformis: the skin and the gut. Ann Intern Med 93:857, 1980

35. Jenkins D, Lynde CW, Stewart WD: Histiocytic lymphoma occurring in a patient with dermatitis herpetiformis. J Am Acad Dermatol 9:252, 1983

36. Guide SV, Marinkovich MP: Linear IgA bullous dermatosis. Clin Dermatol 19:719, 2001

37. Fine JD: Management of acquired bullous skin diseases. N Engl J Med 333:1475, 1995

38. Auquier-Dunant A, Mockenhaupt M, Naldi L, et al: Correlations between clinical patterns and causes or erythema multiforme majus, Stevens-Johnson syndrome, and toxic epidermal necrolysis: results of an international prospective study. Arch Dermatol 138:1019, 2002

39. Manders SM: Toxin-mediated streptococcal and staphylococcal disease. J Am Acad Dermatol 39:383, 1998

40. Spandau U, Brocher EB, Kampgen E, et al: CC and CXC chemokines are differentially expressed in erythema multiforme in vivo. Arch Dermatol 138:1027, 2002

41. Kokuba H, Aurelian L, Burnett J: Herpes simplex virus associated erythema multiforme (HAEM) is mechanistically distinct from drug-induced erythema multiforme: interferon V is expressed in HAEM lesions and tumor necrosis factor–alpha in drug-induced erythema multiforme lesions. J Invest Dermatol 113:808, 1999

42. Prins C, Kerdel FA, Padilla RS, et al: Treatment of toxic epidermal necrolysis with high-dose intravenous immunoglobulins: multicenter retrospective analysis of 48 consecutive cases. Arch Dermatol 139:26, 2003

43. Wolff K, Tappeiner G: Treatment of toxic epidermal necrolysis. Arch Dermatol 139:85, 2003

44. Fine JD, Eady RA, Bauer EA, et al: Revised classification system for inherited epidermolysis bullosa: report of the Second International Consensus Meeting on diagnosis and classification of epidermolysis bullosa. J Am Acad Dermatol 42:1051, 2000

45. Hintner H, Stingl G, Schuler G, et al: Immunofluorescence mapping of antigenic determinants within the dermal-epidermal junction in the mechanobullous diseases. J Invest Dermatol 76:113, 1981

46. Shimizu H, Fine JD, Suzumori K, et al: Prenatal exclusion of pyloric atresia-junctional epidermolysis bullosa syndrome. J Am Acad Dermatol 31:429, 1994

47. Okulicz JF, Kihiczak NI, Janniger CK: Epidermolysis bullosa simplex. Cutis 70:19, 2002

48. Christiano AM, LaForgia S, Paller AS, et al: Prenatal diagnosis for recessive dystrophic epidermolysis bullosa in 10 families by mutation and haplotype analysis in the type VII collagen gene (COL7A1). Mol Med 2:59, 1996

49. Raab B, Fretzin DF, Bronson DM, et al: Epidermolysis bullosa acquisita and inflammatory bowel disease. JAMA 250:1746, 1983

50. Vaillant L, Bernard P, Joly P, et al: Evaluation of clinical criteria for diagnosis of bullous pemphigoid. Arch Dermatol 134:1075, 1998

51. Woodley DT, Burgeson RE, Lunstrum G, et al: The epidermolysis bullosa acquisita antigen is the globular carboxyl terminus of type VII procollagen. J Clin Invest 81:683, 1988

52. Delbaldo C, Chen M, Friedli A, et al: Drug-induced epidermolysis bullous acquisita with antibodies to type VII collagen. J Am Acad Dermatol 46(5 suppl):S161, 2002

53. Cooper SM, Burge SM: Darier's disease: epidemiology, pathophysiology, and management. Am J Clin Dermatol 4:97, 2003

54. Glynne P, Deacon A, Goldsmith D, et al: Bullous dermatoses in end-stage renal failure: porphyria or pseudoporphyria? Am J Kidney Dis 34:155, 1999

55. Perez MI, Kohn SR: Cutaneous manifestations of diabetes mellitus. J Am Acad Dermatol 30:519, 1994

56. Viard I, Wehrli P, Bullani R, et al: Inhibition of toxic epidermal necrolysis by blockade of CD95 with human intravenous immunoglobulin. Science 282:490, 1998

57. Elder D, Elenitsas R, Jaworsky C, et al: Lever's Histopathology of the Skin, 8th ed. Lippincott-Raven, Philadelphia, 1997

X MALIGNANT CUTANEOUS TUMORS

Allan C. Halpern, m.d.
Patricia L. Myskowski, m.d.

Malignant tumors can arise from cells of any layer of the skin—keratinocytes, melanocytes, fibroblasts, endothelial cells, or adipocytes—as well as from cells such as lymphocytes, which normally transit through the skin. Cutaneous metastases may also arise from other primary sites. In this chapter, we review the most common malignant cutaneous tumors in their order of frequency.

Malignant Tumors of the Epidermis

Epidermal skin cancers are the most common cancers in humans. They arise in the keratinocytes and the melanocytes of the epidermis. Epidermal skin cancers present a unique opportunity for effective intervention with both early detection and primary prevention. They are amenable to clinical diagnosis by simple visual inspection and to pathologic diagnosis by minimally invasive biopsy.

Basal cell carcinoma (BCC) and squamous cell carcinoma (SCC) originate from the keratinocytes of the epidermis. Because these two cancers share many features, they are often lumped together under the term nonmelanoma skin cancer (NMSC).

Malignant melanoma is a malignancy arising from a melanocyte. Although malignant melanomas can arise in any melanocyte of the body, including the eye, the vast majority occur in the skin. Cutaneous malignant melanoma has been categorized into four major histogenetic types: lentigo maligna melanoma, superficial spreading melanoma, nodular melanoma, and acral lentiginous melanoma.

SUN EXPOSURE AND SKIN CANCER

Several lines of evidence implicate ultraviolet (UV) radiation in the pathogenesis of all three of the major epidermal skin cancers.[1] Epidemiologic data implicate long-term cumulative sun exposure in the development of SCC and intense intermittent sun exposure in the development of BCC and melanoma. Laboratory studies indicate that both UVA (320 nm to 400 nm) and UVB (290 nm to 320 nm) radiation from sunlight can damage DNA both directly and through oxidative damage. In addition, UV radiation can suppress the cutaneous immune system.[2] The association of some SCCs with chemical carcinogens and the occurrence of acral lentiginous and mucosal melanomas in unexposed areas of the body underscore the need for studies to identify additional etiologic agents.

Recognition of the important role of sunlight in the etiology of skin cancer affords an opportunity for primary prevention through the use of sun protection. Unfortunately, the exact timing and doses of UV exposure involved in the development of skin cancer in humans are not known and likely vary among the types of skin cancer. Accordingly, patients should be educated about the deleterious effects of sun exposure and tanning. Sun-protection efforts should be geared to an overall reduction of sun exposure through the avoidance of sun-seeking behavior and the use of sun-protective clothing. Broad-spectrum sunscreens with a sun protection factor (SPF) of 15 or greater are a useful adjunct to sun protection, but they should not be used to increase the amount of time spent in direct sunlight.[3] The use of tanning beds should be avoided. The use of sunless tanning agents is safe, but the darkening of the skin that results from the use of these agents does not offer significant UV protection. For individuals who are assiduous in their sun protection efforts, attention should be given to adequate vitamin D intake through diet or supplements.[4]

NONMELANOMA SKIN CANCER

NMSC typically occurs as pink lesions on the sun-exposed skin surface. Any pink skin lesion that persists or recurs in the same location, especially if easily irritated by minor trauma, should raise the suspicion of NMSC. Some forms of NMSC will fade with changes in season (i.e., with reduced sun exposure) or with the application of topical steroids, and the clinician should advise patients that any lesion that recurs warrants further attention.

Basal Cell Carcinoma

BCC is a malignant cutaneous tumor arising from the basal keratinocytes of the epidermis.

Epidemiology BCC is the most common skin cancer. The reported incidence ranges from 3.4 per 100,000 per year in African Americans to over 1,100 per 100,000 per year in Townsville, Queensland, Australia.[5,6] Although rare, metastases and death from BCC do occur.

Etiology and risk factors UV radiation—specifically, intense intermittent sun exposure—appears to play an important role in the development of BCC. Studies of basal cell nevus syndrome (Gorlin syndrome) have yielded dramatic insights into the genetics of BCC. The patched gene, which was first recognized as a developmental gene in the fruit fly *Drosophila,* has been identified as playing a critical role in the development of BCC. Almost all patients with basal cell nevus syndrome appear to inherit a mutated copy of the patched gene, and studies of sporadic BCC suggest that mutations in the patched gene pathway (i.e., the sonic hedgehog pathway) are a necessary and often sufficient step in the development of most BCCs.[7]

Diagnosis The majority of BCCs occur on the head and neck. They occur in nodular and superficial forms, as well as in a variety of less common forms.

Nodular BCC appears as a raised, pearly, translucent, pink bump on the skin surface. It is often easily irritated, fragile, and associated with episodes of superficial ulceration or hemorrhage. When ulceration is prominent, it can lead to the appearance of a so-called rodent ulcer, in which the pearly translucent border is barely appreciable. Some nodular BCC lesions appear more white than pink and, on close observation, often demonstrate small telangiectasias. They tend to have a smoother, shinier surface and a firmer texture than common dermal nevi [*see Figure 1*].

Superficial BCC appears as a pink patch of skin. On close inspection, most superficial BCCs demonstrate a thready, translucent border, with areas of seemingly normal or slightly fibrotic skin within the lesion. Superficial BCC is usually found on the upper trunk, arms, and legs.

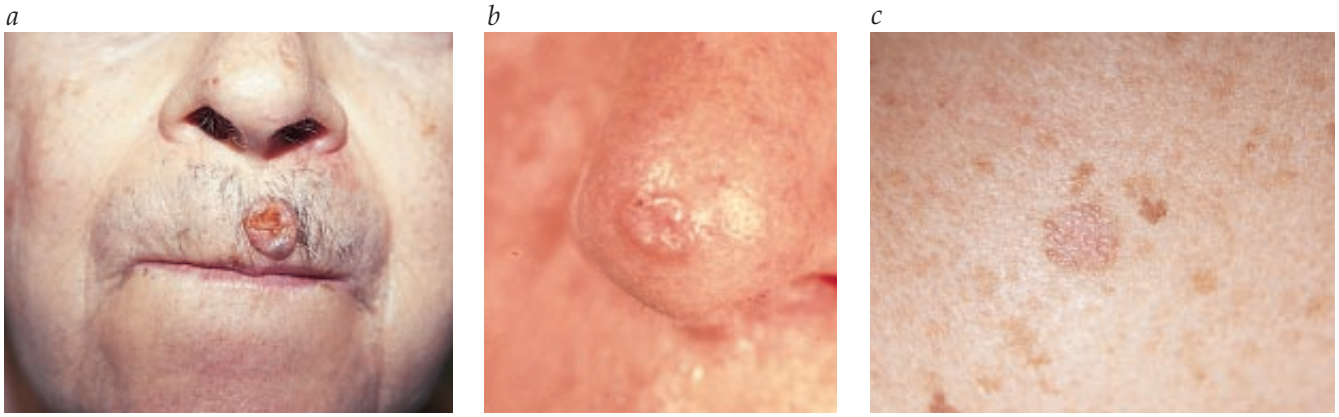


Figure 1 **Nodular basal cell carcinoma—shown here above a patient's lip, with a so-called rodent's ulcer (*a*)—commonly presents as a raised, pearly, translucent pink bump on the skin surface (*b*). A superficial form appears as a pink patch of skin (*c*).**

Less common clinical variants of BCC include morpheaform, pigmented, and cystic lesions. Morpheaform BCCs have an infiltrative pattern that histologically and clinically resembles a scar. Pigmented BCCs typically contain specks of blue-black pigment, but they may be deeply pigmented throughout. Pigmented lesions are most commonly a variant of nodular BCC. Cystic BCCs tend to be softer than typical nodular BCCs and may have a clear to blue-gray appearance.

Patient history plays a critical role in the diagnosis of BCC. When questioned about lesions that become easily irritated or bleed from minor trauma, patients can often alert the clinician to early lesions that would otherwise elude detection. With the patient under local anesthesia, a biopsy should be obtained of any suspicious lesion.

Differential diagnosis Nodular BCC can be confused with angiofibromas, dermal nevi, amelanotic melanoma, cutaneous metastases, dermatofibroma, and a host of benign adnexal tumors (e.g., trichoepithelioma). Superficial BCCs mimic several inflammatory dermatoses (e.g., eczema and tinea) and share several clinical features with actinic keratoses. Pigmented BCC can easily be confused with a primary melanocytic neoplasm. Cystic BCCs can be confused with cystic adnexal tumors and inflammatory lesions.

Treatment The goal of therapy is to adequately eradicate the lesion and ensure the best cosmetic and functional outcome. Multiple factors—such as the size, location, and histologic subtype of the lesions and attributes of the patient, including age, general health, skin color, and skin laxity—should be taken into consideration in choosing an optimal therapy.

The vast majority of BCCs are amenable to surgical treatment. The primary options include curettage and electrodesiccation, excision, and Mohs micrographic surgery. A small but significant subset of BCCs can be treated effectively with Mohs micrographic surgery, which entails microscopic examination of frozen sections of the entire undersurface of the excised specimen at the time of surgery. The technique may be indicated for recurrent lesions and lesions that have a high likelihood of recurrence. Such lesions include ill-defined lesions, large lesions (> 2 cm), lesions with a high-risk histology (i.e., aggressive growth pattern, sclerosing pattern, or perineural involvement), and lesions overlying embryonal fusion planes (e.g., ocular canthi or nasofacial sulcus). The cure rate of Mohs micrographic technique is significantly higher than the cure rates of other treatments of these high-risk lesions.[8]

Radiation therapy can be an effective, painless, and well-tolerated alternative that is typically reserved for older patients who are poor surgical candidates. Radiation therapy should be avoided, however, in patients with basal cell nevus syndrome. Cryotherapy is another therapeutic option for BCC in patients who are poor surgical candidates.

Topical therapy combined with pharmacotherapy using the immune response modifier imiquimod five times weekly for 6 weeks has been approved by the Food and Drug Administration for the treatment of superficial BCC of the trunk and extremities. One packet (250 mg) of imiquimod 5% cream is applied to 25 cm^2 of affected skin.

Experimental therapies under investigation include intralesional chemotherapy, next-generation topical immune modulators, and photodynamic therapy.

All patients treated for BCC are at risk for local recurrence, and they are at significant risk for the development of additional skin cancers. Patients should be instructed in the self-examination of their skin, as well as in methods of sun protection. In addition, they should receive routine professional follow-up.

Prognosis The risk of local recurrence relates to the lesion's size, location, and histology. Metastases are very rare: a prevalence of 0.0028% was reported in a series of 50,000 Australians.[9] Metastases occur through both the lymphatic and the hematogenous routes; risk factors include basal cell nevus syndrome, immunosuppression, and previous exposure to ionizing radiation. Metastases that are not amenable to surgical management are associated with a poor outcome.

Squamous Cell Carcinoma

Like BCC, cutaneous SCC arises from the keratinocytes of the epidermis. Histologically, the cells of well-differentiated SCC resemble the cells of the superior portion of the epidermis.

Epidemiology An estimated 150,000 to 250,000 new cases of cutaneous SCC were diagnosed in the United States in 1994.[10] The estimated mortality from SCC in the United States in 1988 was approximately 0.5 per 100,000. Several lines of data suggest significant increases in SCC incidence. In Australia, for example,

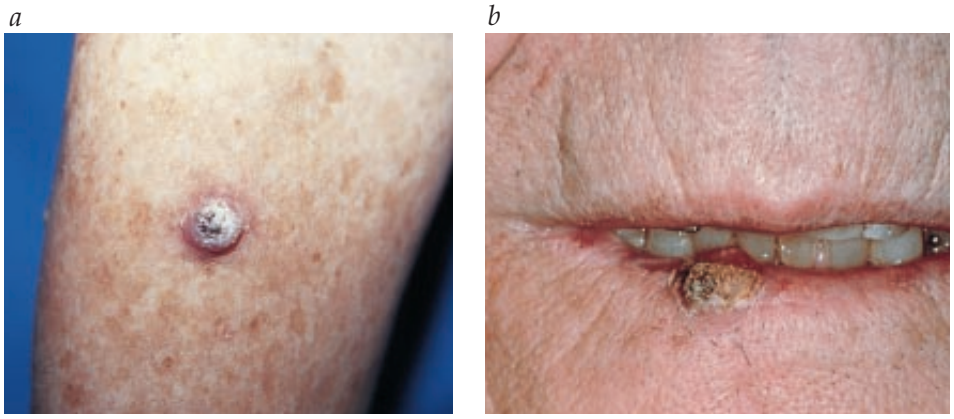


Figure 2 **A squamous cell carcinoma is shown on an arm (*a*) and lower lip (*b*).**

the incidence of SCC increased by 51% between the years 1985 and 1990.[11] In the United States, some of the highest rates of NMSC have been detected in the southwest. A population-based survey in New Mexico found the incidence of SCC doubled in both males and females between 1978 and 1999.[12]

Etiology and risk factors In addition to sunlight, other known etiologic agents that contribute to the development of cutaneous SCC are ionizing radiation, chemical carcinogens, thermal burns, and chronic nonhealing wounds. Sun-related SCCs demonstrate a lower risk of metastases and death than SCCs related to other exposures. Factors involved in predisposition to SCC from sun exposure include light skin color, a tendency to burn, and an inability to tan.

Pathophysiology and pathogenesis Sun-related SCC is often associated with a precursor lesion called an actinic keratosis. Such lesions occur on the scalp, the face, the extensor surfaces of the forearms, and the backs of the hands. They tend to be rough-surfaced, irregularly shaped, and pink. They are often more readily felt than seen. The majority of patients with actinic keratoses have multiple lesions. The risk of SCC in these individuals has been estimated to be as high as 20%.[13] SCC may also appear on normal-looking skin.

SCC of the oral or genital mucosa may arise in precursor lesions termed leukoplakia or erythroplakia. Mucosal SCCs are associated with a significant risk of metastases. Immune surveillance affects the progression of SCC. Immunosuppression, as occurs in transplant recipients and patients with lymphoma, is associated with a high incidence of SCC.[14] In these patients, infection with human papillomavirus appears to play an etiologic role in conjunction with sun exposure. SCCs tend to be more aggressive in immunosuppressed persons.

Diagnosis Most lesions occur in areas of the body that are usually exposed to the sun. The lesions are pinkish, firm plaques that often have a rough, scaly surface [*see Figure 2*]. Biopsy is required for definitive diagnosis.

Differential diagnosis The differential diagnosis of SCC includes keratoacanthoma, Bowen disease, verrucous carcinoma, BCC, hypertrophic actinic keratosis, and common warts.

Keratoacanthomas share many features with SCC, both clinically and histologically. They arise de novo on normal-looking skin and grow very rapidly. They are typically pink, dome-shaped, shiny bumps with a central crateriform keratotic plug that occur on the surface of the skin. They may become very large. Although keratoacanthomas are not associated with a risk of metastasis, they can be locally destructive. Spontaneous regression of keratoacanthoma over the course of months has been well documented.

Bowen disease is SCC that is confined to the epidermis. It appears as red, scaly, minimally elevated plaques with well-defined, irregular borders. The reported association of Bowen disease with internal malignancy has not held up to closer scrutiny.[15]

SCCs that lack a scaly keratotic surface can be confused with a host of other adnexal and dermal skin tumors.

Treatment Small SCCs evolving from an actinic keratosis can be adequately treated with simple curettage and electrodesiccation. Larger actinic lesions, as well as lesions arising in non–sun-exposed areas of skin, are best treated with definitive surgical excision with confirmation of negative margins. High-risk, ill-defined lesions, especially those occurring in the surgically sensitive areas of the face, genitalia, hands, and feet, are often best treated by Mohs micrographic surgery.

Fractionated radiation therapy is an alternative treatment of primary SCC in older patients who are poor surgical candidates. The benefits of adjuvant radiation therapy are less clear, as are the benefits of sentinel lymph node biopsy and elective lymph node dissection (ELND) for patients with high-risk SCC of the head and neck.

Cytotoxic chemotherapy and biologic response modifiers have been used in patients who have advanced SCC; this therapeutic approach has been reported to have complete response rates of up to 68%, but there are few long-term survivors.[16] Actinic keratoses are treated with cryotherapy, curettage, topical therapies (e.g., fluorouracil, imiquimod, or diclofenac), photodynamic therapy, and laser resurfacing to prevent progression to SCC.[17] Regularly updated guidelines for the treatment of SCC and BCC are available through the National Comprehensive Cancer Network (NCCN).[18]

Prognosis Regardless of the therapy employed, high-risk lesions have a significant rate of local recurrence at 5 years. High-risk SCCs include those in specific anatomic sites (e.g., ears, lips, genitalia, and other non–sun-exposed areas), those greater than 2 cm in diameter, those with aggressive histologic features (depth > 4 mm, Clark level IV and above, and poorly differentiated histology), and those in immunosuppressed patients.[19] The primary route of SCC metastasis is via lymphatic spread to regional lymph nodes. Reported rates of metastasis vary from as low as 0.3% in small, sun-derived lesions to 33% in larger, poorly differentiated lesions.[19] Reported overall 5-

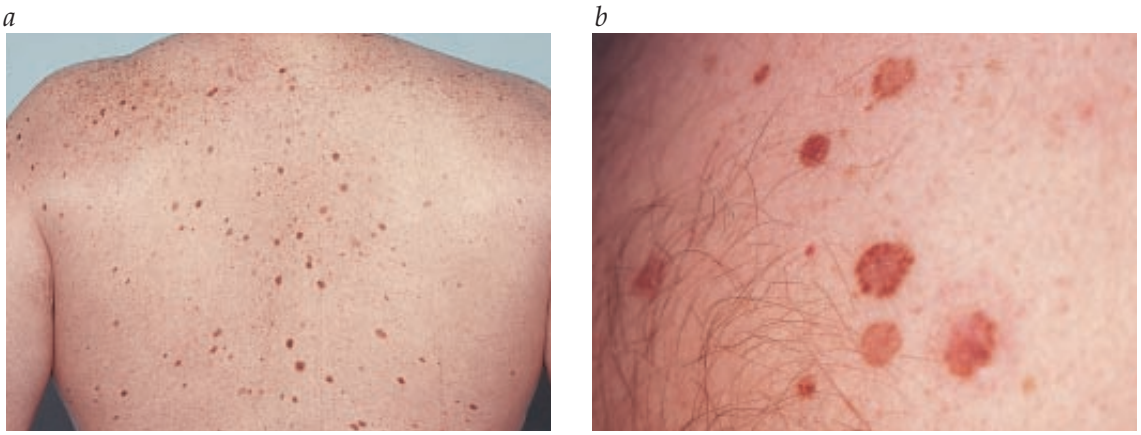


Figure 3 **Dysplastic nevi typically are larger than common moles (*a*) and have variegate pigmentation and ill-defined borders (*b*).**

year survival rates for patients with regionally metastatic SCC have ranged from 25% to 47%.[19]

MALIGNANT MELANOMA

Epidemiology

In the United States, a person's lifetime risk for developing melanoma is about 1 in 75 (1.3%).[20] Between 1973 and 1994, the incidence of melanoma rose by 121%, and the mortality rose by 39%.[21] Encouraging trends include a shift toward the detection of earlier disease, as well as a stabilization of incidence rates in some segments of the population. In terms of both morbidity and mortality, however, the burden of melanoma-related disease continues to increase. Although melanoma can occur in anyone, it is primarily a disease of whites. Melanomas occurring in blacks are more commonly of the acral lentiginous variety.

Table 1 Adjusted Estimated Relative Risks of Melanoma by Nevus Type and Number[25]

Type	*Number*	*Adjusted Relative Risk**
Nevi > 2 mm and < 5 mm	0–24	1.0
	25–49	1.8 (1.3–2.5)
	50–99	3.0 (2.1–4.4)
	≥ 100	3.4 (2.0–5.7)
Nondysplastic nevi > 5 mm	0	1.0
	1	0.9 (0.7–1.3)
	2–4	1.3 (1.0–1.8)
	5–9	1.7 (1.0–2.7)
	≥ 10	2.3 (1.2–4.3)
Dysplastic nevi	None	1.0
	Indeterminate	1.0 (0.7–1.6)
	1	2.3 (1.4–3.6)
	2–4	7.3 (4.6–12.0)
	5–9	4.9 (2.5–9.8)
	≥ 10	12.0 (4.4–31.0)

*Mutually adjusted and adjusted for age, sex, center, referral pattern, morphologic dysplastic nevi < 5 mm, sunburns, freckles, solar damage, scars, nevus excisions, and family history of melanoma (confidence interval = 95%).

Etiology and Risk Factors

Sun exposure Although strong epidemiologic and basic-science evidence supports an association between melanoma and sun exposure, the relationship appears to be complex.[22] Lentigo maligna melanoma is associated with long-term cumulative sun exposure. Superficial spreading melanoma and nodular melanoma appear to be associated with intense intermittent sun exposure, especially in youth. Acral lentiginous melanoma has no apparent association with sun exposure. Basic-science studies and animal models have implicated different wavelengths of UV in melanoma carcinogenesis; UV wavelength may vary among types of melanoma.

Skin color Melanoma can occur in all racial/ethnic groups but is much more common in lighter-skinned individuals. Among whites, several additional risk factors have been identified, such as fair complexion, a tendency to burn, an inability to tan, freckling, and a family history of melanoma.[22] Screening of the family members of patients with melanoma (particularly multiple melanomas) may be a useful preventive and diagnostic measure.[23]

Moles and dysplastic nevi The strongest phenotypic markers of melanoma risk are moles (nevi)—more specifically, increased numbers of moles and the presence of atypical moles (dysplastic nevi). Melanoma can arise in a preexisting mole or may arise de novo on normal-appearing skin.

Several epidemiologic studies have correlated dysplastic nevi with melanoma risk. Clinically, dysplastic nevi are large (> 5 mm) moles with variegate pigmentation and ill-defined borders [*see Figure 3*]. Histologically, dysplastic nevi are characterized by the presence of architectural atypia and random cytologic atypia. The degree of melanoma risk associated with dysplastic nevi depends on the genetic context. In families with familial melanoma–dysplastic nevus syndrome, the abnormal mole phenotype appears to be inherited in an autosomal dominant fashion. Members of these families with dysplastic nevi have a lifetime melanoma risk that approaches 100%.[24] Outside the context of familial melanoma, dysplastic nevi occur in approximately 5% to 15% of whites. In this general population, dysplastic nevi are markers of increased melanoma risk [*see Table 1*].[25]

Genetic Factors

Approximately 5% of patients with melanomas have a family history of melanoma. Mutations in the cell-cycle regulatory gene *p16* (cyclin-dependent kinase inhibitor–2a) are associated with melanoma in approximately 40% of familial-melanoma families, with linkage of the gene to chromosome 9p.[26] A highly specific activating somatic mutation in the *BRAF* proto-oncogene (a member of the RAF family of kinases) is found in the majority of melanomas and benign nevi, suggesting a pivotal role for this genetic pathway in melanocytic tumor progression.[27] Genomic analyses are beginning to distinguish biologically distinct subsets of melanoma.[28]

Diagnosis

As a pigmented lesion occurring on the surface of the skin, melanoma is amenable to early detection by simple visual inspection at an easily curable stage. Left untreated, melanoma is among the deadliest and most therapeutically unresponsive forms of cancer.

Physical examination Early recognition of melanoma requires attention to pigmented lesions on all body surfaces. Despite the strong association of melanoma with sun exposure, melanomas can occur anywhere on the skin or mucosa. Patients' self-examination, as well as physician examination, must therefore include all skin surfaces, including the scalp, genitalia, and soles of the feet. Any pigmented skin lesion with recent change or with features described by the ABCD mnemonic (asymmetry, border irregularity, color variation, diameter > 6 mm) warrants consideration of the possibility of melanoma. Although any mole may change gradually over time, any that change color, shape, or size relative to a patient's other moles deserve special attention [*see Figure 4*].[29] Dysplastic nevi present both opportunity and challenge in melanoma detection. On one hand, their recognition allows efficient targeting of a high-risk group. On the other, they can complicate attempts at melanoma detection by clinically mimicking early melanomas. Although some dysplastic nevi may progress to melanoma, the overwhelming majority remain benign. Furthermore, not all melanomas arising in patients with dysplastic nevi develop in a preexisting mole. Wholesale removal of dysplastic nevi is an impractical approach to melanoma prevention. In patients with dysplastic nevi, melanoma detection is predicated on specialized visual examination aided by self-examination and professional follow-up to identify changing lesions.[30]

Diagnostic aids Several specialized aids to the diagnosis of melanoma in patients with dysplastic nevi are under development. Dermoscopy entails the use of a handheld otoscope-like device to magnify a pigmented lesion while applying pressure and oil to the surface. The technique allows the visualization of pigment patterns and features not apparent with simple visual inspection. With experience and training, dermoscopy can be a useful aid in distinguishing melanoma from benign pigmented lesions; however, when used inexpertly, dermoscopy may actually decrease diagnostic accuracy.[30,31] Another aid to melanoma detection in high-risk individuals is photographically assisted follow-up.[32] A baseline set of whole-

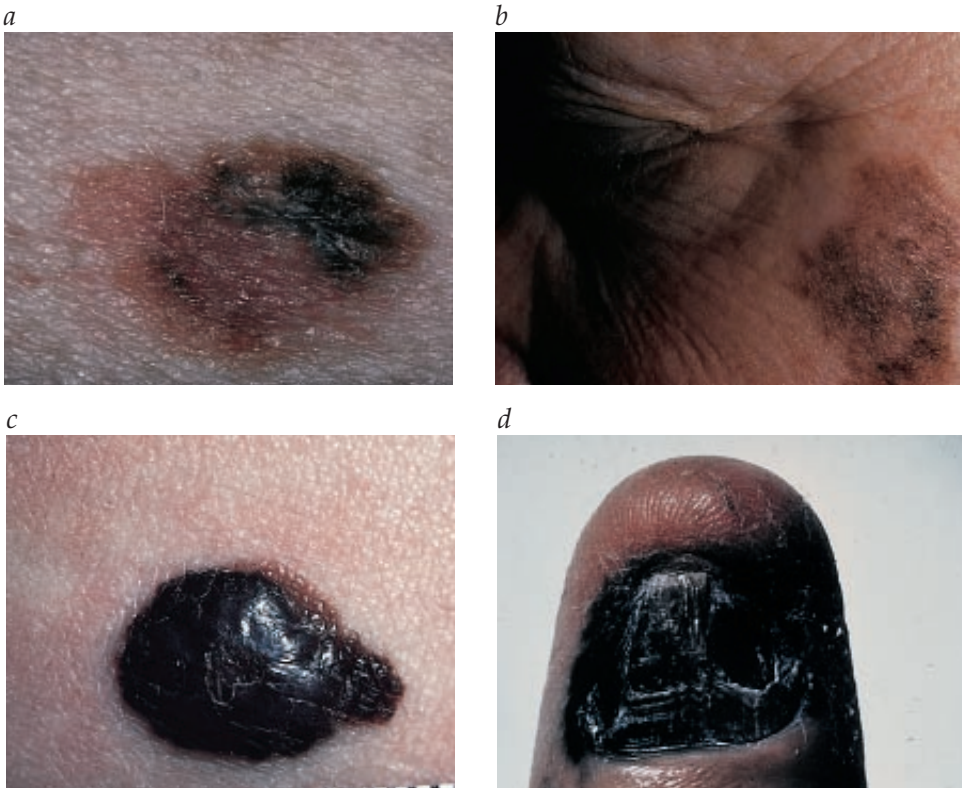


Figure 4 **Superficial spreading malignant melanoma begins as a small, irregular brown lesion (a). Variation in color and contour is characteristic of lentigo maligna melanoma (*b*). Nodular melanoma often grows more in thickness than in diameter (*c*). Acral lentiginous melanoma can resemble a hematoma under the nail (*d*).**

body photographs of the skin are used during self-examination and professional follow-up examination to assess change in the lesions. This procedure helps to prevent unnecessary excision of stable lesions and improves the sensitivity of examinations in detecting change. New imaging technologies such as in vivo confocal scanning laser microscopy hold promise for future improvements in the noninvasive diagnosis of melanoma.[33]

Full-thickness excision and biopsy Any lesion that raises a clinical suspicion of melanoma requires definitive diagnosis. Full-thickness excision is the preferred technique for biopsy of a suspicious pigmented lesion. Partial biopsy can lead to misdiagnosis through sampling error or by depriving the pathologist of a view of the overall architecture and cytology of the lesion. Incisional biopsies with good clinicopathologic correlation may be appropriate, however, in the assessment of large lesions and of lesions occurring in surgically sensitive areas. There is no evidence to suggest that incisional biopsy increases the risk of metastasis.

Differential Diagnosis

Dysplastic nevi share many features with early superficial spreading melanoma. Other common lesions that may mimic melanoma include lentigines, sunburn freckles, traumatized nevi, thrombosed angiomas, pigmented BCCs, pigmented Bowen disease, dermatofibromas, and atypical seborrheic keratoses. Two other challenges in the differential diagnosis of melanoma deserve special mention. Amelanotic melanomas (melanomas without pigment) present as pink lesions that may be misdiagnosed as BCCs or Spitz nevi. Spitz nevi can be difficult to differentiate from melanoma both clinically and histologically. Spitz nevi occur most commonly in children, but they also occur in adults. Like nodular melanomas, Spitz nevi tend to appear suddenly and range in color from red to reddish brown.

Treatment

Primary site Primary cutaneous melanoma is managed surgically with definitive reexcision. The wide excisions of the past have given way to resections with more modest margins. Multiple prospective, randomized trials have investigated the surgical resection of primary cutaneous melanoma utilizing different margins of resection; these studies have focused on varied and overlapping patient populations. On the basis of these data, the NCCN recommends resection margins of 1 cm for melanomas less than 1 mm in thickness, margins of 1 to 2 cm for melanomas between 1 and 2 mm in thickness, and margins of 2 cm for melanomas greater than 2 mm in thickness.[34,35] Primary closure and reconstructive flaps are preferable, cosmetically and functionally, to skin grafts and should be used instead of grafts whenever possible.

Lymph nodes Patients with clinically evident regional lymph node disease are treated with therapeutic lymph node dissection.[34] Elective lymph node dissection in patients with primary melanoma and who have no clinical evidence of lymph node involvement has been abandoned on the basis of the failure of multiple randomized trials to demonstrate an overall survival benefit with this procedure.

Sentinel lymph node biopsy is being increasingly used in patients with primary cutaneous melanoma. This technique utilizes lymphoscintigraphy to identify the draining regional lymph node basins for the skin at the site of the primary melanoma. At the time of definitive reexcision of the melanoma, a blue dye and radioisotope are injected into the dermis around the melanoma site. A small incision is made over the spot that has been identified on lymphoscintigraphy as the proximal area of drainage of the regional lymph node basin. The first lymph node identified as taking up the blue dye and radioisotope (i.e., the sentinel node) is then excised.

The sentinel node is then histologically evaluated, often with the use of immunohistochemical techniques and occasionally with the use of polymerase chain reaction, which is more sensitive. The absence of melanoma in the sentinel node is highly sensitive for ruling out the presence of metastases in the remainder of the lymph node basin when the procedure is performed by an experienced team. When the sentinel node is found to be positive for melanoma, a "completion" lymph node dissection of the affected basin is typically performed. Prospective studies have demonstrated sentinel node status to be strongly correlated with 5-year survival.[36] Patients with positive sentinel nodes are appropriate candidates for consideration of adjuvant therapy [*see* Adjuvant Therapy, *below*]. Several multicenter trials are currently under way to assess the clinical utility of this procedure.[37] Initial reports from the first of these trials have failed to indicate an overall survival advantage associated with the procedure.[38]

In-transit metastases In-transit metastases are metastases that establish tumors within regional dermal and subcutaneous lymphatics before reaching the regional lymph nodes. In-transit metastases can remain confined to a single limb for prolonged periods. Amputation does not appear to provide a long-term survival benefit in this setting.[34] Slow-growing individual in-transit metastases can be managed surgically. More extensive disease can be treated with sensitization therapy with dinitrochlorobenzene (DNCB), intralesional interferon, or topical agents for modifying the immune response. For extensive in-transit metastases confined to an extremity, limb perfusion therapy can result in dramatic palliation and limb salvage. The procedure entails isolation of the vasculature of the involved extremity from the systemic vasculature and perfusion of the isolated limb with chemotherapeutic agents, biologic agents, or both at doses that could not be tolerated if given systemically.[39]

Distant metastases Despite the development of several novel approaches to the treatment of patients with metastatic melanoma, including multiagent chemotherapy, biologic therapy, immunotherapy, and combinations of these treatments, no regimens have demonstrated a clear survival advantage over single-agent chemotherapy. Monotherapy with dacarbazine (2 to 4.5 mg/kg daily for 10 days, repeated every 4 weeks) or recombinant interleukin-2 (IL-2) (600,000 IU/kg every 8 hours for up to 14 doses) are the only treatment regimens approved by the FDA for the treatment of metastatic melanoma. Objective responses to dacarbazine are seen in approximately 5% to 20% of patients; durable complete responses are rare.[40] Objective responses to IL-2, a significantly more toxic agent, are seen in approximately 15% of patients; durable responses are seen in about 5%. Radiation therapy can play an important palliative role. In the absence of more effective clinically proven therapy, patients with distant metastases should be offered the opportunity to participate in clinical trials of experimental therapy. Many current experimental therapies are

Table 2 AJCC TNM Classification[45]

TNM Classification	*Tumor Thickness, Node Number, Metastases Site*	*Subclassification*
T classification T1	≤ 1.0 mm	a: Without ulceration and Clark level II or III b: With ulceration or Clark level IV or V
T2	1.01–2.0 mm	a: Without ulceration b: With ulceration
T3	2.01–4.0 mm	a: Without ulceration b: With ulceration
T4	> 4.0 mm	a: Without ulceration b: With ulceration
N classification N1	One lymph node	a: Micrometastasis* b. Macrometastasis†
N2	2–3 lymph nodes	a: Micrometastasis* b. Macrometastasis† c: In-transit met(s)/satellites(s) without metastatic lymph nodes
N3	4 or more metastatic lymph nodes, matted lymph nodes, or combination of in-transit met(s)/satellite(s) with metastatic lymph nodes	—
M classification M1a	Distant skin, subcutaneous, or lymph node mets	Normal LDH
M1b	Lung mets	Normal LDH
M1c	All other visceral mets Any distant mets	Normal LDH Elevated LDH with any M

*Micrometastases are diagnosed after sentinel or elective lymphadenectomy.
†Macrometastases are defined as clinically detectable nodal metastases confirmed by therapeutic lymphadenectomy; the term also applies to nodal metastases that exhibit gross extracapsular extension.
AJCC—American Joint Committee on Cancer LD—lactic dehydrogenase mets—metastases

predicated on decades of experience with immunotherapy of melanoma, as well as the recent availability of pharmacologic inhibitors of elements of the Ras signaling pathway that are implicated in melanoma pathogenesis.[40]

Adjuvant therapy Patients with cutaneous or regional disease who have been surgically rendered disease free but who are at high risk for recurrence or metastasis are potential candidates for adjuvant therapy.[41] Various adjuvant therapies have been used in melanoma, including immunostimulants such as bacillus Calmette-Guérin, *Corynebacterium parvum,* and levamisole. Several chemotherapeutic agents have been tried as well. More recently, immunotherapies with cytokines, such as interferons, and active immunization with vaccines have been studied**.** A high-dose regimen of interferon alfa (20 million units/m^2 I.V. daily for 1 month followed by 10 million units/m^2 S.C. three times a week for 48 weeks) has been approved by the FDA for use as adjuvant therapy for melanoma. Two studies have demonstrated a small but statistically significant improvement in overall survival with this regimen. Multiple studies have failed to demonstrate improved long-term overall survival with the use of adjuvant interferon in intermediate-dose or low-dose regimens.[42,43]

A host of novel strategies, including active immunization, passive immunization, and myriad biologic therapies, are currently being studied and may provide opportunities for patients who are appropriate candidates for trials.[44]

Prognosis

Stage The single strongest prognostic factor for melanoma is stage of disease. Various staging classifications have been used over the years. All staging systems for melanoma take into account the classic TNM classification of tumor size (T), lymph node involvement (N), and distant metastases (M). The differences across staging systems relate largely to the staging of the primary site. New staging systems attempt to use the attributes of the primary tumor that strongly correlate with outcome. These attributes include thickness, ulceration, and, in the case of thin melanomas measuring less than 1 mm thick, the Clark level of invasion. The advent of sentinel node biopsy has led to the inclusion of microstaging of lymph nodes in the staging system [*see Tables 2 and 3*].[45,46]

Attributes of the primary tumor Several attributes of the primary tumor have been identified as predictors of outcome from primary cutaneous melanoma. A strong predictor of outcome is the Breslow tumor thickness, which is measured in millimeters from the granular layer of the epidermis to the deepest tumor cell. Other important histologic parameters are the Clark level of tumor invasion, the presence or absence of ulceration, the rate of mitosis, the presence of tumor-infiltrating lymphocytes, and vascular invasion. For thin primary melanomas, one of the strongest predictors of outcome is growth phase.[47] Radial-growth-phase melanoma does not appear to metastasize, whereas vertical-growth-phase melanoma (characterized by the for-

Table 3 AJCC Staging System and Survival Rate[45]

Pathologic Stage	*TNM*	*5-Year Survival*	*10-Year Survival*
IA	T1a	95.3 ± 0.4	87.9 ± 1.0
IB	T1b	90.9 ± 1.0	83.1 ± 1.5
	T2a	89.0 ± 0.7	79.2 ± 1.1
IIA	T2b	77.4 ± 1.7	64.4 ± 2.2
	T3a	78.7 ± 1.2	63.8 ± 1.7
IIB	T3b	63.0 ± 1.5	50.8 ± 1.7
	T4a	67.4 ± 2.4	53.9 ± 3.3
IIC	T4b	45.1 ± 1.9	32.3 ± 2.1
IIIA	N1a	69.5 ± 3.7	63.0 ± 4.4
	N2a	63.3 ± 5.6	56.9 ± 6.8
IIIB	N1a	52.8 ± 4.1	37.8 ± 4.8
	N2a	49.6 ± 5.7	35.9 ± 7.2
	N1b	59.0 ± 4.8	47.7 ± 5.8
	N2b	46.3 ± 5.5	39.2 ± 5.8
IIIC	N1b	29.0 ± 5.1	24.4 ± 5.3
	N2b	24.0 ± 4.4	15.0 ± 3.9
	N3	26.7 ± 2.5	18.4 ± 2.5
IV	M1a	18.8 ± 3.0	15.7 ± 2.9
	M1b	6.7 ± 2.0	2.5 ± 1.5
	M1c	9.5 ± 1.1	6.0 ± 0.9

AJCC—American Joint Committee on Cancer

mation of a tumor nodule in the dermis) is associated with significant risk of metastasis even in lesions less than 1 mm thick.[48] Patient characteristics associated with improved survival from melanoma include young age (< 60 years), female sex, and location of the melanoma on an extremity other than the palms or soles. Multivariable models for predicting outcome from melanoma have been developed [*see Table 4*].[49]

Malignant Tumors of the Dermis

METASTATIC TUMORS

Cutaneous metastases occur in approximately 5% of patients with solid tumors and are usually associated with widespread disease. The relative frequency of skin metastases is gender specific, reflecting the rates of the primary cancers.[50] In women, two thirds of metastases are from breast cancer, but lung cancer, colorectal cancer, melanoma, and ovarian cancer are also frequent. In men, lung cancer is most common, followed by cancer of the large intestine, melanoma, SCC of the head and neck, and cancer of the kidneys.[50] The anatomic distribution of skin metastases is not random. Cutaneous metastases from breast cancer often involve the chest wall and may appear as nodules, lymphedema, or cellulitis. The scalp is a common site for metastasis, especially of cancer from the lung and kidney (in men) and breast (in women). Head and neck cancers may invade the skin by local extension, giving rise to a firm, dusky-red edema of the skin that resembles cellulitis. Abdominal wall metastases, often called Sister Joseph's nodules, may occur with gastrointestinal or ovarian malignancies.[50] Clinically, cutaneous metastases are often minimally symptomatic dermal papules or nodules and are flesh-colored or pink; dissemination occurs via lymphatic or vascular pathways. Cutaneous metastases may clinically reflect the histology of the primary tumor (e.g., black, brown, or gray nodules with metastatic melanoma, and vascular nodules with renal cell or thyroid carcinoma).

PRIMARY TUMORS

Primary malignancies of the dermis may develop from any of the myriad structures of the skin, including sebaceous glands (sebaceous carcinoma), connective tissue (dermatofibrosarcoma protuberans), smooth muscle (leiomyosarcoma), and other adnexal tissue (eccrine carcinoma). Most of these primary dermal neoplasms are rare; they may exhibit aggressive biologic behavior. Although these neoplasms are quite varied histologically, many share a common clinical presentation of a rapidly growing flesh-colored to pink or red subcutaneous nodule that occasionally resembles a sebaceous cyst.

Merkel cell carcinoma This neoplasm is a dermal malignancy of neuroendocrine origin. It usually appears as a red to violaceous dermal papule or nodule on the head and neck of elderly patients, although all age groups are affected. The treatment of choice is wide local excision with or without lymphadenectomy. Sentinel node biopsy has been proposed by some for evaluation of the regional lymph nodes. Adjuvant radiation therapy can be considered. Local recurrences are frequent, and distant metastases occur in more than one third of patients. Chemotherapy of metastases is generally disappointing.[51,52]

Paget disease A rare malignancy of the skin associated with an underlying adenocarcinoma,[50,51] Paget disease usually presents as an erythematous, often weeping unilateral dermatitis of the breast that involves the nipple and areola. The differential diagnosis includes eczema, psoriasis, contact dermatitis, and impetigo. For this reason, biopsy of an inflammatory, nonresolving dermatitis of the nipple or areola is imperative. In Paget disease, the biopsy will reveal typical pale-staining Paget cells in the epidermis. Appropriate surgical resection of the cutaneous and underlying neoplasm is the treatment of choice; lymph node metastases often occur.[51]

Extramammary Paget disease Extramammary Paget disease is even more uncommon than Paget disease. It typically presents as red, often ulcerated, plaques in the perineal areas of elderly persons.[50,51] Lesions may be pruritic or asymptomatic, are often long-standing, and may have been misdiagnosed as psoriasis, contact dermatitis, or chronic fungal infection. Underlying associated tumors include rectal and genitourinary carcinomas. Even without an associated internal malignancy, extramammary Paget disease is difficult to treat, and it is associated with a high local recurrence rate.[51]

Angiosarcoma A rare, often highly aggressive vascular malignancy,[51] angiosarcoma may appear as multicentric reddish-purple patches or nodules in a lymphedematous limb, such as on a lymphedematous arm after a mastectomy (Stewart-Treves syndrome). Another presentation is violaceous patches or plaques on the head or neck (especially scalp) of elderly persons. Patients with angiosarcoma have a poor prognosis, with pulmonary metastases frequently developing despite surgery or radiation.[51]

Dermatofibrosarcoma protuberans Dermatofibrosarcoma protuberans is a slow-growing, locally aggressive malignancy that rarely metastasizes but often recurs. Lesions typically present as firm reddish-brown or purple nodules, usually on the trunk or non–sun-exposed extremities. The differential diagnosis includes keloids and benign dermatofibroma. Young adults are most often affected, although the tumor may occur at any age. Wide local excision with or without Mohs micrographic surgery offers the best chance of cure.[51]

KAPOSI SARCOMA

Kaposi sarcoma (KS) is a multicentric cutaneous neoplasm that has four distinct clinical variants.[53-56] In spite of its name, KS is not a true sarcoma. Although the cell of origin has not been clearly established,[53,54] KS cells share phenotypic markers with lymphatic endothelium, as well as vascular smooth muscle cells, suggesting a vascular or pluripotent mesenchymal cell origin.[56] In its classic form, KS is an indolent disease of elderly men of Mediterranean or eastern European origin, in which violaceous nodules and plaques develop on the lower extremities.[53-55] A second variant, lymphadenopathic KS, is endemic to some areas of Africa. African KS, which typically affects young adults and children, pursues a more aggressive course than classic KS, with frequent bone, lymph node, and visceral involvement.[53-55] A third variant of KS occurs in iatrogenically immunosuppressed patients, especially organ transplant recipients.[57] In this variant, men are affected slightly more often than women.[53-55] The fourth variant is an aggressive epidemic KS that occurs in AIDS patients.

Epidemiology

Before the advent of AIDS, KS was rare in the United States, with an age-adjusted annual incidence of 0.29 per 100,000 population in men and 0.07 per 100,000 population in women.[54] KS was an AIDS-defining illness for 30% to 40% of patients in the earliest years of the HIV epidemic.[53] During that period, the incidence of KS in HIV-infected homosexual men was 73,000-fold higher than in the general United States population; in HIV-infected women and HIV-infected nonhomosexual men, the incidence was 10,000-fold higher.[52,53] The incidence has significantly declined since the introduction of highly active antiretroviral therapy (HAART). For example, in a large European-based study of HIV-infected patients, there was an estimated 39% annual reduction in the incidence of KS between 1994 and 2003, such that the incidence of KS in 2003 was 10% less than that reported in 1994.[58]

Etiology and Risk Factors

Human herpesvirus type 8 The epidemiology of KS has long suggested a transmissible infectious agent or cofactor.[53-55] Kaposi sarcoma–associated herpesvirus (KSHV), also known as human herpesvirus type 8 (HHV-8), has been detected in all variants of KS.[59] HHV-8 has also been found in patients with body cavity–based lymphoma, Castleman disease, and angioblastic lymphadenopathy, as well as in certain skin lesions of organ transplant recipients.[53] The mechanism by which HHV-8 infection leads to KS tumorigenesis is unclear but probably involves a complex combination of inflammation, angiogenesis, and neoplastic proliferation.[53,54] The prevalence of KS largely parallels the rate of HHV-8 infection in various populations.[54] Although the incidence of HHV-8 infection may be as high as 2% to 10% in the general population, the incidence of KS is very low, suggesting that the majority of infections are subclinical.[53,54]

Host factors Host factors, particularly immunosuppression, are crucial in some populations with KS.[47,48,50] HIV may play an indirect role in the development of KS through $CD4^+$ T cell depletion and stimulated production of growth factors and cytokines such as IL-1 and IL-6.[53,54,56] Immunosuppressive drugs, especially cyclosporine, azathioprine, and prednisone, increase the risk of developing KS, primarily in kidney and liver transplant recipients.[49]

Despite the prevalence of KS in some ethnic groups, the role of any possible genetic factors is unclear. An increased incidence of HLA-DR5 in patients with classic KS has been debated.[54] Fa-

Table 4 Estimated Probability of 10-Year Survival in Patients with Primary Cutaneous Melanoma[47]

Tumor Thickness/Age of Patient	*Probability of 10-Year Survival**			
	Tumor with Extremity Location		*Tumor with Axis Location†*	
	Female Patients	*Male Patients*	*Female Patients*	*Male Patients*
< 0.76 mm				
≤ 60 yr	0.99 (0.98–1.0)	0.98 (0.95–0.99)	0.97 (0.93–0.99)	0.94 (0.88–0.97)
> 60 yr	0.98 (0.95–0.99)	0.96 (0.89–0.98)	0.92 (0.82–0.96)	0.84 (0.70–0.93)
0.76–1.69 mm				
≤ 60 yr	0.96 (0.92–0.98)	0.93 (0.85–0.97)	0.86 (0.76–0.92)	0.75 (0.62–0.84)
> 60 yr	0.90 (0.80–0.95)	0.81 (0.64–0.91)	0.67 (0.50–0.81)	0.50 (0.33–0.67)
1.70–3.60 mm				
≤ 60 yr	0.89 (0.80–0.94)	0.80 (0.65–0.89)	0.65 (0.50–0.77)	0.48 (0.35–0.61)
> 60 yr	0.73 (0.57–0.85)	0.57 (0.38–0.75)	0.38 (0.24–0.55)	0.24 (0.14–0.37)
> 3.60 mm				
≤ 60 yr	0.74 (0.53–0.87)	0.58 (0.36–0.77)	0.39 (0.21–0.60)	0.24 (0.13–0.40)
> 60 yr	0.48 (0.28–0.69)	0.32 (0.16–0.53)	0.18 (0.08–0.35)	0.10 (0.04–0.20)

*Confidence interval = 95%.
†Axis location includes the trunk, head, neck, and volar and subungual sites.

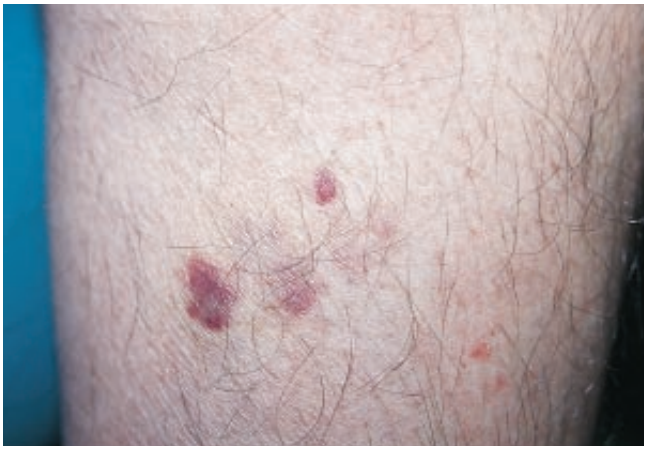


Figure 5 **HIV-associated Kaposi sarcoma lesions vary from pink patches (shown) to deep-purple plaques.**

milial KS is extremely rare, suggesting that genetic factors alone are not responsible.

Finally, gender appears to be a significant risk factor, especially in classic KS, in which the male-to-female ratio may range from 3:1 to 10:1.[53,54] The reasons for this male predominance remain unclear.[48,49]

Diagnosis

Clinical manifestations The clinical manifestations of KS differ among the variants of the disorder.[53-55] In classic KS, faint reddish-purple macules or patches or purple nodules first appear on the feet, especially the soles. Lymphadenopathy (especially inguinal) is present on rare occasions. Lesions may also occasionally develop on the arms and genital areas. As the disease progresses, the lesions coalesce into violaceous plaques.

HIV-associated KS usually presents as cutaneous lesions, but the first lesions may appear in the oral mucosa or lymph nodes. In contrast to classic KS lesions, HIV-associated KS lesions often begin on the upper body (face, trunk, or arms). Most typically, HIV-associated KS lesions are purple-red, often oval, papules that follow a pityriasis rosea–like distribution [*see 2:II Papulosquamous Disorders*].[53-55] Lesions vary from pink macules to deep-purple plaques [*see Figure 5*] or may resemble ecchymoses, especially in patients with low $CD4^+$ T cell counts. Oral lesions are typically red-purple plaques or nodules on the palate, gingiva, or buccal mucosa. Patients with darker skin may have dark-purple to black lesions or hyperpigmented plaques.[54]

As HIV-associated KS progresses, lymphedema may develop in the feet, scrotum, genitalia, and periorbital regions, and lymphadenopathy (especially inguinal) may occur. Gastrointestinal lesions are usually submucosal and asymptomatic but may result in gastrointestinal hemorrhage. Pulmonary KS carries a poor prognosis.[54]

Laboratory studies Laboratory workup of patients with KS should include HIV antibody testing, complete blood count, fecal occult blood testing, and chest radiograph. $CD4^+$ T cell counts are indicated in HIV-positive patients. A complete medical history and physical examination should be performed, with special attention paid to the presence of opportunistic infections in HIV-infected or otherwise immunosuppressed patients. Skin biopsy should be obtained in patients with suspected KS. The histopathology of KS is characterized by the presence of spindle-shaped cells in the dermis, with extravasated red blood cells present in slits between irregular vascular spaces.[55]

Differential Diagnosis

The clinical differential diagnosis of KS includes dermatofibroma, purpura, pyogenic granuloma, bacillary angiomatosis, metastatic melanoma, and BCC. Other histopathologic entities that may resemble KS include angiosarcoma and stasis dermatitis.[55]

Treatment

Classic Kaposi sarcoma The therapy for KS is palliative. In classic KS, where the disease is indolent and the patients are elderly, aggressive systemic therapy is rarely warranted.[53,54] Instead, radiation therapy is the treatment of choice.[54,60] KS is very radiosensitive: single doses of 800 cGy have been used for rapid palliation in patients with poor prognoses. Total doses of 800 to 3,500 cGy have yielded 50% complete responses and 46% partial responses, with more than half of patients needing no follow-up treatment for as long as 13 years.[60] A treatment regimen equivalent to 3,000 cGy in 10 fractions over 2 weeks has been advocated.[60]

For patients with classic KS who have only one or two papules, excisional biopsy may be sufficient for both diagnosis and treatment. Cryotherapy with liquid nitrogen may be useful for isolated papules. Systemic therapy for classic KS may be indicated in cases of extensive cutaneous disease or visceral involvement. Single-agent chemotherapy with vinca alkaloids (i.e., vincristine or vinblastine) is commonly used. Low-dose recombinant interferon alfa may also be effective in classic KS; however, side effects (e.g., fever, chills, myalgias, and fatigue) may not be well tolerated by elderly patients.[53,54]

Transplant-associated Kaposi sarcoma Spontaneous KS regression has been observed in transplant recipients after withdrawal of cyclosporine and corticosteroids.[54] Sirolimus (rapamycin), an immunosuppressive drug with antineoplastic and antiangiogenic properties, was successfully used in 15 renal transplant recipients who developed KS. After KS was diagnosed, cyclosporine and mycophenolate mofetil were discontinued and sirolimus was started. Cutaneous KS resolved in all patients, without episodes of acute rejection or changes in renal graft function.[58]

HIV-associated Kaposi sarcoma Although KS is more aggressive in HIV-infected patients, the extent of immune suppression and the presence of opportunistic infections or other systemic illnesses may be of equal importance in staging, determining prognosis, and choosing appropriate therapy.[53,54] Clinical features that were traditionally associated with a more favorable outcome included a $CD4^+$ T cell count higher than 200 cells/mm^3, a lack of systemic illness, KS limited to the skin or lymph nodes, and minimal (i.e., not nodular) oral KS; poor risk factors included a $CD4^+$ T cell count below 200 cells/mm^3, KS-associated lymphedema, visceral KS, ulcerated KS, nodular oral KS, and opportunistic infection.[61] With the advent of HAART, however, physicians treating patients with HIV-associated KS now have the opportunity to influence and even reverse immune suppression by affecting both HIV viral load and the $CD4^+$ T cell count. Regression of KS has been observed after initiation of HAART, often during the first few months of therapy[56,62]; consequently, this

is often first-line therapy for patients with limited cutaneous HIV-associated KS.[54]

Local therapy is a reasonable approach in KS patients with limited disease, those with infectious complications, and those who cannot tolerate systemic therapy.[62] Radiation therapy is effective in HIV-associated KS in doses similar to those used for classic KS (see above). Responses in HIV-associated KS are generally short-lived, however.[54] Topical alitretinoin (9-*cis*-retinoic acid) gel may be effective in HIV-associated KS and has been approved by the FDA for this use.[62] Intralesional injections of vinblastine or interferon have also been useful in selected lesions.[53,54] Cryotherapy with liquid nitrogen is effective for small lesions[62]; however, cryotherapy is contraindicated in dark-skinned patients in whom posttreatment hypopigmentation may appear much worse cosmetically than the original KS lesion.

Systemic therapy has included conventional chemotherapy and biologic response modifiers. For patients with slowly progressive, limited cutaneous KS (< 25 lesions), systemic antitumor therapy may not be necessary; HAART with or without local therapy may be sufficient.[51] However, HAART alone has not been demonstrated to be the treatment of choice for advanced HIV-associated KS.[63] Liposomal anthracyclines (e.g., doxorubicin, daunomycin) are approved by the FDA as first-line therapy of HIV-associated KS.[64-66] A reasonable approach to the treatment of advanced HIV-associated KS is the use of a combination of HAART and liposomal anthracyclines, followed by a combination of HAART plus paclitaxel if response to the first regimen is inadequate.[56,63,65] Promising investigational approaches for HIV-associated KS include antiangiogenic compounds, thalidomide, matrix metalloproteinase inhibitors, and retinoids. Prevention of HIV-associated KS may also be achieved through antiviral therapy of HHV-8.[62]

Complications

Bacterial infections and sepsis are common in patients with KS and may be associated with ulcerated tumors of the legs and feet. Opportunistic infections may intervene, especially in patients with very low $CD4^+$ T cell counts.

Prognosis

The total $CD4^+$ T cell count is the most important predictor of survival in HIV-associated KS.[61] Large tumor burdens, lymphedema, and pulmonary KS are also predictive of poorer outcomes.[61,65]

Cutaneous Lymphoma

Lymphomas may be of B cell or T cell lineage and may involve the skin primarily or secondarily [*see 12:IV Principles of Cancer Treatment*]. B cell lymphomas, particularly non-Hodgkin lymphomas, may involve the skin secondarily in advanced disease. They typically appear as reddish-purple subcutaneous plaques or nodules. Primary B cell lymphomas of the skin are even rarer. They appear as reddish nodules that often remain localized to the skin but may progress to systemic disease. The vast majority of primary cutaneous lymphomas fall into the spectrum of cutaneous T cell lymphoma (CTCL).

CTCL includes mycosis fungoides (MF) and Sézary syndrome, which is a leukemic variant of MF.[66,67] MF is the largest subset of CTCL; the two terms, however, sometimes are used interchangeably. Another variant of CTCL is associated with human T cell lymphotropic virus type I (HTLV-I) and is part of the spectra of adult T cell lymphoma/leukemia and peripheral T cell lymphoma.[66]

EPIDEMIOLOGY

CTCL is a rare disorder. In the United States, approximately 1,000 new cases of CTCL are diagnosed annually.[66] From 1973 to 1984, the incidence of CTCL rose from 0.19 per 100,000 population to 0.42 per 100,000 population. CTCL primarily affects middle-aged adults; the median age at presentation is 50 years.[68] The male-to-female ratio is approximately 2:1; blacks are twice as likely as whites to develop CTCL.[68]

ETIOLOGY

Host susceptibility and an environmental antigen, perhaps viral, are hypothesized as playing important roles in the pathogenesis of CTCL.[66] Genetic factors may be related to major histocompatibility antigens, such as an increase in HLA-DRB1*11 (formerly HLA-DR5) and HLA-DQB1*03.[69] Chronic antigenic stimulation (e.g., infection) may play an etiologic role.[66] For example, HTLV-I infection may be an etiologic factor in the development of the peripheral T cell lymphoma variant.[66]

DIAGNOSIS

Clinical Manifestations

The clinical manifestations of MF typically evolve over many months to years. In one classic study, the mean duration of symptoms before diagnosis was 7.5 years.[70] Flat, erythematous patches, often scaling and occasionally atrophic, begin most commonly on the trunk and thighs, especially in a so-called bathing-trunk distribution [*see Figure 6*]. Lesions are asymptomatic or mildly pruritic and may spontaneously remit or respond to topical corticosteroid therapy. Patients may also report improvement after sun exposure. As MF progresses, patches tend to enlarge and thicken into plaques. The color may become dark red; in dark-skinned persons, the lesions may initially be hyperpigmented or hypopigmented and may acquire an erythematous or violaceous hue. In advanced MF, tumors may develop or transform to a large-cell lymphoma.[66,67,71]

In approximately 10% of cases, tumors are the initial presentation of CTCL (tumor d'emblé). Generalized erythroderma with circulating atypical T cells (in Sézary syndrome) is the presentation in 5% of CTCL patients.[66,67]

Physical examination of patients with suspected CTCL includes complete skin examination, including classification of lesions (patch, plaque, or tumor) and extent of body surface area involved. Lymph nodes, the liver, and the spleen should be palpated.

Skin Biopsy

Skin biopsy is necessary for the definitive diagnosis of CTCL. The presence of atypical lymphoid cells with hyperconvoluted cerebriform nuclei in clusters in the epidermis (Pautrier microabscesses) and a bandlike lymphocytic infiltrate in the upper dermis are diagnostic of CTCL.[66,67] The malignant cell is a T cell, with most of the cells expressing the pan–T cell markers CD2, CD3, and CD5, as well as frequent deletion of CD7, CD26, or both.[66,67,72] The use of T cell receptor gene rearrangement studies to confirm clonality in early disease may be an aid to diagnosis.[66] Neither immunophenotypic studies nor electron microscopy may be considered to be definitively diagnostic of CTCL; clinicopathologic correlation is necessary.

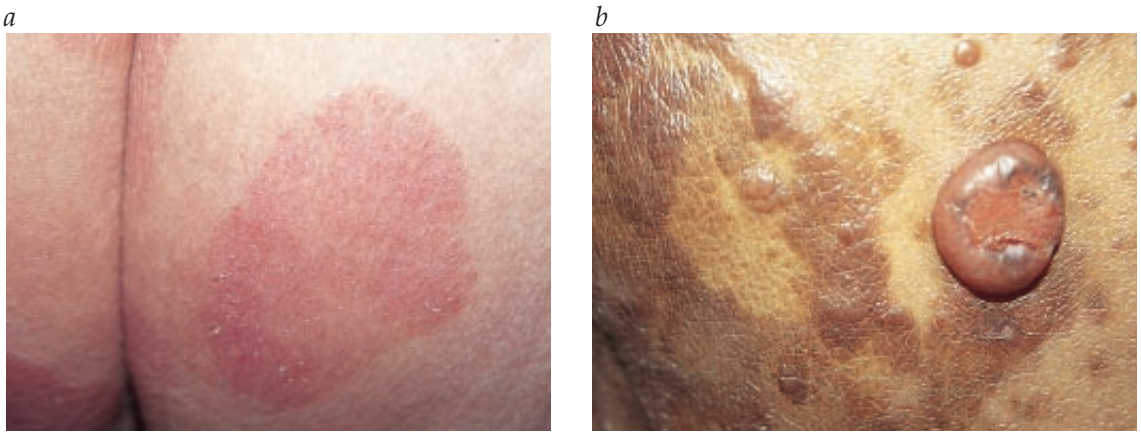


Figure 6 **Cutaneous T cell lymphoma is shown in the large-patch stage (*a*) and as tumor-stage mycosis fungoides (*b*).**

Laboratory Studies

The laboratory evaluation for CTCL includes complete blood count, eosinophil count, Sézary cell count, lactic dehydrogenase level, and liver function tests. Bone marrow biopsy is unnecessary in the absence of circulating leukemic cells. HTLV-I testing should be considered for patients with risk factors or atypical presentations. Lymph node biopsy should be considered for palpable nodes, especially those larger than 2 cm. Abdominal computed tomography or chest radiography may be important in patients with tumors or suspected visceral involvement.

DIFFERENTIAL DIAGNOSIS

In its early stages, CTCL may resemble any of a number of benign inflammatory disorders (e.g., drug reaction, eczema, psoriasis, or contact dermatitis). These disorders should be ruled out before contemplating therapy.

STAGING

The staging of CTCL is based on an evaluation of the type and extent of skin lesions and the extent of lymph node, peripheral blood, and visceral involvement.[70,71] Early disease is characterized by limited patch or plaque disease (stage IA) or generalized patch or plaque disease without evidence of extracutaneous involvement (stages IB and IIA); more advanced disease is characterized by cutaneous tumors (stage IIB), extracutaneous disease (stage III), and extracutaneous disease involving either lymph nodes (stage IVA) or viscera (stage IVB).

TREATMENT

Topical Therapy

Topical therapy is the mainstay of the treatment of early disease (stage IA, IB, and IIA). Early aggressive therapy with radiation and chemotherapy has not proved to be superior to local approaches in controlling disease or improving survival in patients with limited disease.[66,67] A rational approach for treating early limited (or histologically equivocal) disease is topical corticosteroids.[73] Topical nitrogen mustard (mechlorethamine), in either aqueous or ointment form, is the most frequently used topical chemotherapy. In one series, the overall response rate to nitrogen mustard was 83%, with a complete response rate of 50%, after a median treatment time of 12 months.[74] Median time to relapse was also 12 months.[74]

Carmustine (BCNU) solution, applied daily to lesions, is another useful regimen. Treatment generally lasts 8 to 16 weeks but has been continued for up to 6 months. Because systemic absorption can result in bone marrow suppression, complete blood counts must be monitored.[62] Bexarotene, a topical retinoid, has been shown to be effective in CTCL; it is approved by the FDA for use in CTCL.[75]

Ultraviolet Radiation

Radiation therapy for CTCL takes several forms, from ultraviolet light to ionizing radiation. UVB is useful in stage I disease. In a retrospective study of 21 patients with stage I disease, narrowband UVB led to complete remission in 81% of patients and to partial remission in 19%; the mean relapse-free interval was 24.5 months.[76]

Another effective approach to treatment of CTCL is the combination of psoralen and UVA (PUVA). In one study, 65% of patients with stage I CTCL had complete clinical clearing, with a mean relapse-free interval of 43 months; the disease-free survival rates at 5 and 10 years for stage IA were 56% and 30%, respectively.[77] In another study, complete remission was observed in 71% of early-stage patients; in this study, the mean relapse-free interval was 22.8 months.[76]

Radiation Therapy

Total skin electron beam (TSEB) radiation delivers radiotherapy to the skin surface without a significant internal dose. It is especially useful with plaque disease. Typical doses are 2,400 to 3,600 cGy, fractionated over several weeks with 4 to 9 MeV electron beam radiation.[78] Treatment responses are related to CTCL stage[79]; early-stage (stage IA) patients have a 95% response rate, but 50% will experience relapse within 10 years. TSEB may also be useful in stage IB disease (90% remission rate), but two thirds of patients treated with this modality will experience relapse within 5 years.[73] Patients with tumor-stage (stage IIB) CTCL may receive effective palliation from TSEB, especially in combination with other therapies.[79]

Systemic Therapy

Systemic therapy has been undertaken as primary therapy in advanced CTCL (stages III through IVB); in early-stage disease, systemic therapy is used as part of sequential therapy to promote more durable responses.[66,67]

Oral bexarotene has yielded response rates of up to 45% in advanced CTCL, and it is approved by the FDA for use in this disease.[80] Another systemic therapy used in the treatment of advanced CTCL is denileukin diftitox [DAB(389) IL-2].[81] This receptor-targeted cytotoxic fusion protein binds to the IL-2 receptor on T cells; it achieved a 30% response rate in heavily-pretreated patients.[81]

Extracorporeal photopheresis, which is an accepted therapy for advanced CTCL, appears most useful in erythrodermic CTCL and Sézary syndrome.[66,67] In this treatment, the patient is given a photoactivating drug (8-methoxypsoralen), the patient's white blood cells are collected via leukapheresis and irradiated with UVA, and the irradiated cells are returned to the patient intravenously. Advanced CTCL characterized by cutaneous tumors (stage III) or visceral involvement (stage IV) has also been treated with single-agent and combination chemotherapy using methotrexate, adenosine analogues, interferon alfa, and retinoids.[66,67,82]

Combination Therapy

Early aggressive treatment using TSEB followed by combination chemotherapy provides no survival advantage over sequential topical therapy.[83,84] In a randomized controlled trial, 103 patients with MF received TSEB followed by either parenteral chemotherapy with cylophosphamide, doxorubicin, etoposide, and vincristine or sequential topical treatment. Patients receiving combined therapy had a significantly higher rate of complete response than those receiving sequential topical therapy; however, there was no difference in the rates of disease-free and overall survival between the two groups after a mean follow-up of 75 months.[83] In an uncontrolled study, multimodality therapy was examined in patients with early and advanced disease. In this study, 95 CTCL patients received in consecutive phases of therapy interferon alfa and oral isotretinoin, TSEB, and maintenance therapy consisting of topical nitrogen mustard and interferon alfa. Patients with advanced disease also received six cycles of combination chemotherapy before TSEB. Although multimodality therapy resulted in high response rates (85% response, 60% complete response), the study provided no evidence that this form of combination therapy could improve the overall survival rates currently achieved with sequential topical therapy.[80] In general, the heterogeneity of reported combination therapy regimens in CTCL makes it virtually impossible to compare results.

Future Directions

A number of experimental approaches are being investigated in CTCL, including allogeneic bone marrow transplantation, histone deacetylase inhibitors, monoclonal antibodies, and fusion toxins.[67] Other investigative modalities include cytokines such as recombinant IL-12 and IL-2.[66]

COMPLICATIONS

The most serious complications of CTCL are infections. Sepsis from ulcerated cutaneous tumors is a common cause of death. Visceral CTCL may occur, as may transformation to large cell lymphoma in some CTCL patients (39% probability after 12 years).[71] In long-term survivors with early disease, local therapies (e.g., TSEB or PUVA) may contribute to the development of other skin cancers (e.g., BCC or SCC) and cataracts.[85]

PROGNOSIS

Many different attempts have been made to classify CTCL into useful prognostic groups. An early and still valid study that used the TNM system identified three major groups: good-risk patients (stages IA, IB, and IIA, with plaque-only skin disease and no lymph node, blood, or visceral involvement [median survival, > 12 years]); intermediate-risk patients (stages IIB, III, and IVA, with cutaneous tumors, erythroderma, or plaque disease and node or blood involvement but no visceral disease or node effacement [median survival, 5 years]); and poor-risk patients (stage IVB, with visceral involvement or node effacement [median survival, 2.5 years]).[70]

Eosinophilia is also associated with shortened survival.[70] Other long-term studies have revealed that stage IA patients do not have a reduced life expectancy and that fewer than 10% of these patients experience disease progression to more advanced stages.[86] Survival of patients with generalized patch/plaque MF (stage IB or IIA), at a median of 11.7 years, is significantly worse than that of a race-, age-, and sex-matched control population.[87] Gender and race appear to have no effect on survival, but older patients (> 58 years) have shorter disease-specific survivals.[68]

Allan C. Halpern, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Patricia L. Myskowski, M.D., has received grant or research support from Merck & Co., Inc., and is a consultant for Ligand Pharmaceuticals Inc.

Topical corticosteroids, topical nitrogen mustard, and carmustine have not been approved by the FDA for treatment of cutaneous T cell lymphoma; interferon alfa, sirolimus, and HAART have not been approved by the FDA for treatment of Kaposi sarcoma.

References

1. Green A, Whiteman D, Frost C, et al: Sun exposure, skin cancers and related skin conditions. J Epidemiol 9(6 suppl):S7, 1999

2. Ullrich S: Mechanisms underlying UV-induced immune suppression. Mutat Res 571:185, 2005

3. Dummer R, Maier T: UV protection and skin cancer. Recent Results Cancer Res 160:7, 2002

4. Scarlett WL: Ultraviolet radiation: sun exposure, tanning beds, and vitamin D levels. What you need to know and how to decrease the risk of skin cancer. J Am Osteopath Assoc 103: 371, 2003

5. Buettner PG, Raasch BA: Incidence rates of skin cancer in Townsville, Australia. Int J Cancer 78:587, 1998

6. Serrano H, Scotto J, Shornick G, et al: Incidence of nonmelanoma skin cancer in New Hampshire and Vermont. J Am Acad Dermatol 24:574, 1991

7. High A, Zedan W: Basal cell nevus syndrome. Curr Opin Oncol 17:160, 2005

8. Shriner DL, McCoy DK, Goldberg DJ, et al: Mohs micrographic surgery. J Am Acad Dermatol 39:79, 1998

9. Paver K, Poyzer K, Burry N, et al: Letter: the incidence of basal cell carcinoma and their metastases in Australia and New Zealand. Australas J Dermatol 14:53, 1973

10. Miller DL, Weinstock MA: Nonmelanoma skin cancer in the United States: incidence. J Am Acad Dermatol 30:774, 1994

11. Staples M, Marks R, Giles G: Trends in the incidence of non-melanocytic skin cancer (NMSC) treated in Australia 1985–1995: are primary prevention programs starting to have an effect? Int J Cancer 78:144, 1998

12. Athas WF, Hunt WC, Key CR: Changes in nonmelanoma skin cancer incidence between 1977–1978 and 1998–1999 in Northcentral New Mexico. Cancer Epidemiol Biomarkers Prev 12:1105, 2003

13. Kirkham N: Tumors and cysts of the epidermis. Lever's Histophathology of the Skin, 9th ed. Elder DE, Murphy GF, Johnson BL, et al, Eds., Lippincott Williams & Wilkins, Philadelphia, 2004, p 805

14. Lindelof B, Sigurgeirsson B, Gabel H, et al: Incidence of skin cancer in 5356 patients following organ transplantation. Br J Dermatol 143:513, 2000

15. Sarmiento JM, Wolff BG, Burgart LJ, et al: Perianal Bowen's disease: associated tumors, human papillomavirus, surgery, and other controversies. Dis Colon Rectum 40:912, 1997

16. Guthrie TH Jr, Porubsky ES, Luxenberg MN, et al: Cisplatin-based chemotherapy in advanced basal and squamous cell carcinomas of the skin: results in 28 patients including 13 patients receiving multimodality therapy. J Clin Oncol 8:342, 1990

17. Jorizzo JL: Current and novel treatment options for actinic keratosis. J Cutan Med Surg 8(suppl 3):13, 2004

18. Basal Cell and Squamous Cell Skin Cancers. Clinical Practice Guidelines in Oncology, Version 2.2005. National Comprehensive Cancer Network, Jenkintown, Pennsylvania, 2005
http://www.nccn.org/professionals/physician_gls/PDF/nmsc.pdf

19. Rowe DE, Carroll RJ, Day CL Jr: Prognostic factors for local recurrence, metastasis,

and survival rates in squamous cell carcinoma of the skin, ear, and lip. Implications for treatment modality selection. J Am Acad Dermatol 26:976, 1992

20. Jemal A, Murray T, Ward E, et al: Cancer statistics, 2005. CA Cancer J Clin 55:10, 2005

21. Hall HI, Miller DR, Rogers JD: Update on the incidence and mortality from melanoma in the United States. J Am Acad Dermatol 40:35,1999

22. Jemal A, Devesa SS, Hartge P, et al: Recent trends in cutaneous melanoma incidence among whites in the United States. J Natl Inst 93:678, 2001

23. Blackwood MA, Holmes R, Synnestvedt M, et al: Multiple primary melanoma revisited. Cancer 94:2248, 2002

24. Carey WP Jr, Thompson CJ, Synnestvedt M, et al: Dysplastic nevi as a melanoma risk factor in patients with familial melanoma. Cancer 74:3118, 1994

25. Elder DE, Clark WH Jr, Elenitsas R, et al: The early and intermediate precursor lesions of tumor progression in the melanocytic system: common acquired nevi and atypical (dysplastic) nevi. Semin Diagn Pathol 10:18, 1993

26. Haluska FG, Hodi FS: Molecular genetics of familial cutaneous melanoma. J Clin Oncol 16:670,1998

27. Tuveson DA, Weber BL, Herlyn M: BRAF as a potential therapeutic target in melanoma and other malignancies. Cancer Cell 4:95, 2003

28. Bastian BC: Understanding the progression of melanocytic neoplasia using genomic analysis: from fields to cancer. Oncogene 22:3081, 2003

29. Abbasi NR, Shaw HM, Rigel DS, et al: Early diagnosis of cutaneous melanoma: revisiting the ABCD criteria. JAMA 292:2771, 2004

30. Naeyaert JM, Brochez L: Clinical practice. Dysplastic nevi. N Engl J Med 349:2233, 2003

31. Braun RP, Rabinovitz HS, Oliviero M, et al: Dermoscopy of pigmented skin lesions. J Am Acad Dermatol 52:109, 2005

32. Halpern AC: The use of whole body photography in a pigmented lesion clinic. Dermatol Surg 26:1175, 2000

33. Marghoob AA, Swindle LD, Moricz CZ, et al: Instruments and new technologies for the in vivo diagnosis of melanoma. J Am Acad Dermatol 49:777, 2003

34. Melanoma. Clinical Practice Guidelines in Oncology, Version 2.2005. National Comprehensive Cancer Network, Jenkintown, Pennsylvania, 2005

http://www.nccn.org/professionals/physician_gls/PDF/melanoma.pdf

35. Cook J: Surgical margins for resection of primary cutaneous melanoma. Clin Dermatol 22:228, 2004

36. Roberts AA, Cochran AJ: Pathologic analysis of sentinel lymph nodes in melanoma patients: current and future trends. J Surg Oncol 85:152, 2004

37. Reintgen D, Pendas S, Jakub J, et al: National trials involving lymphatic mapping for melanoma: the Multicenter Selective Lymphadenectomy Trial, the Sunbelt Melanoma Trial, and the Florida Melanoma Trial. Semin Oncol 31:363, 2004

38. Morton DL, Thompson JF, Cochran AJ, et al: Interim results of the Multicenter Selective Lymphadenectomy Trial (MSLT-I) in clinical stage I melanoma. Proceedings of the American Society of Clinical Oncology Annual Meeting, abstract 7500. New Orleans, Jan. 20 to 22, 2005

39. Fraker DL: Management of in-transit melanoma of the extremity with isolated limb perfusion. Curr Treat Options Oncol 5:173, 2004

40. Buzaid AC: Management of metastatic cutaneous melanoma. Oncology (Williston Park) 18:1443, 2004

41. Agarwala SS, Kirkwood JM: Adjuvant interferon treatment for melanoma. Hematol Oncol Clin North Am 12:823, 1998

42. Kirkwood JM, Strawderman MH, Ernstoff MS, et al: Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. J Clin Oncol 14:7, 1996

43. Kirkwood JM, Ibrahim JG, Sosman JA, et al: High-dose interferon alfa-2b significantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: results of intergroup trial E1694/S9512/C509801. J Clin Oncol 19:2370, 2001

44. Hersey P: Adjuvant therapy for high-risk primary and resected metastatic melanoma. Int Med J 33:33, 2003

45. Balch CM, Soong SJ, Atkins, MB, et al: An evidence-based staging system for cutaneous melanoma. CA Cancer J Clin 54:131, 2004

http://caonline.amcancersoc.org/cgi/content/full/54/3/131

46. Balch CM, Soong SJ, Gershenwald JE, et al: Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. J Clin Oncol 19:3622, 2001

47. Guerry D 4th, Synnestvedt M, Elder DE, et al: Lessons from tumor progression: the invasive radial growth phase of melanoma is common, incapable of metastasis, and indolent. J Invest Dermatol 100:342S, 1993

48. Elder DE, Van Belle P, Elenitsas R, et al: Neoplastic progression and prognosis in melanoma. Semin Cutan Med Surg 15:336, 1996

49. Schuchter L, Schultz DJ, Synnestvedt M, et al: A prognostic model for predicting 10-year survival in patients with primary melanoma. The Pigmented Lesion Group. Ann Intern Med 125:369, 1996

50. Schwartz RA: Cutaneous metastatic disease. J Am Acad Dermatol 33:161, 1995

51. Demetrius RW, Randle HW: High-risk nonmelanoma skin cancers. Dermatol Surg 24:1272, 1998

52. Allen PJ, Bowne WB, Jaques DP, et al: Merkel cell carcinoma: prognosis and treatment of patients from a single institution. J Clin Oncol 23:2300, 2005

53. Antman K, Chang Y: Kaposi's sarcoma. N Engl J Med 342:1027, 2000

54. Myskowski P, Krown S: Kaposi's sarcoma. Skin Cancer. Sober AJ, Haluska FG, Eds. American Cancer Society Atlas of Clinical Oncology series. BC Decker, Philadelphia, 2001

55. Schwartz RA: Kaposi's sarcoma: an update. J Surg Oncol 87:146, 2004

56. Cheung TW: AIDS-related cancer in the era of highly active antiretroviral therapy (HAART): a model of the interplay of the immune system, virus, and cancer. "On the offensive—the Trojan Horse is being destroyed." Part A: Kaposi's sarcoma. Cancer Invest 22:774, 2004

57. Stallone G, Schena A, Infante B, et al: Sirolimus for Kaposi's sarcoma in renal-transplant recipients. N Engl J Med 352:1317, 2005

58. Mocroft A, Kirk O, Clumeck N, et al: The changing pattern of Kaposi sarcoma in patients with HIV, 1994–2003: the EuroSIDA Study. Cancer 100:2644, 2004

59. Moore PS, Chang Y: Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and without HIV infection. N Engl J Med 332:1181, 1995

60. Cooper JS, Sacco J, Newall J: The duration of local control of classic (non–AIDS-associated) Kaposi's sarcoma by radiotherapy. J Am Acad Dermatol 19:59, 1988

61. Krown SE, Testa MA, Huang J: AIDS-related Kaposi's sarcoma: prospective validation of the AIDS Clinical Trials Group staging classification. AIDS Clinical Trials Group Oncology Committee. J Clin Oncol 15:3085, 1997

62. Cattelan AM, Trevenzoli M, Aversa SM: Novel pharmacological therapies for the treatment of AIDS-related Kaposi's sarcoma. Expert Opin Investig Drugs 13:501, 2004

63. Krown SE: Highly active antiretroviral therapy in AIDS-associated Kaposi's sarcoma: implications for the design of therapeutic trials in patients with advanced, symptomatic Kaposi's sarcoma. J Clin Oncol 22:399, 2004

64. Northfelt DW, Dezube BJ, Thommes JA, et al: Pegylated-liposomal doxorubicin versus doxorubicin, bleomycin, and vincristine in the treatment of AIDS-related Kaposi's sarcoma: results of a randomized phase III clinical trial. J Clin Oncol 16:2445, 1998

65. Welles L, Saville MW, Lietzau J, et al: Phase II trial with dose titration of paclitaxel for the therapy of human immunodeficiency virus–associated Kaposi's sarcoma. J Clin Oncol 16:1112, 1998

66. Pichardo DA, Querfeld C, Guitart J, et al: Cutaneous T-cell lymphoma: a paradigm for biological therapies. Leuk Lymphoma 45:1755, 2004

67. Foss F: Mycosis fungoides and the Sézary syndrome. Curr Opin Oncol 16:421, 2004

68. Weinstock MA, Reynes JF: The changing survival of patients with mycosis fungoides: a population-based assessment of trends in the United States. Cancer 85:208, 1999

69. Hodak E, Klein T, Gabay B, et al: Familial mycosis fungoides: report of 6 kindreds and a study of the HLA system. J Am Acad Dermatol 52:393, 2005

70. Sausville EA, Eddy JL, Makuch RW, et al: Histopathologic staging at initial diagnosis of mycosis fungoides and the Sézary syndrome: definition of three distinctive prognostic groups. Ann Intern Med 109:372, 1988

71. Diamandidou E, Colome-Grimmer M, Fayad L, et al: Transformation of mycosis fungoides/Sézary syndrome: clinical characteristics and prognosis. Blood 92:1150, 1998

72. Kim EJ, Hess S, Richardson SK, et al: Immunopathogenesis and therapy of cutaneous T cell lymphoma. J Clin Invest 115:798, 2005

73. Zackheim HS, Kashani-Sabet M, Amin S: Topical corticosteroids for mycosis fungoides. Experience in 79 patients. Arch Dermatol 134: 949, 1998

74. Kim YH, Martinez G, Varghese A, et al: Topical nitrogen mustard in the management of mycosis fungoides: update of the Stanford experience. Arch Dermatol 139:165, 2003

75. Breneman D, Duvic M, Kuzel T, et al: Phase 1 and 2 trial of bexarotene gel for skin-directed treatment of patients with cutaneous T-cell lymphoma. Arch Dermatol 138:325, 2002

76. Diederen PV, van Weelden H, Sanders CJ, et al: Narrowband UVB and psoralen-UVA in the treatment of early-stage mycosis fungoides: a retrospective study. J Am Acad Dermatol 48:215, 2003

77. Querfeld C, Rosen ST, Kuzel TM, et al: Long-term follow-up of patients with early-stage cutaneous T-cell lymphoma who achieved complete remission with psoralen plus UV-A monotherapy. Arch Dermatol 141:305, 2005

78. Jones GW, Kacinski BM, Wilson LD, et al: Total skin electron radiation in the management of mycosis fungoides: consensus of the European Organization for Research and Treatment of Cancer (EORTC) Cutaneous Lymphoma Project Group. J Am Acad Dermatol 47:364, 2002

79. Duvic M, Apisarnthanarax N, Cohen DS, et al: Analysis of long-term outcomes of combined modality therapy for cutaneous T-cell lymphoma. J Am Acad Dermatol 49:35, 2003

80. Duvic M, Hymes K, Heald P, et al: Bexarotene is effective and safe for treatment of refractory advanced-stage cutaneous T-cell lymphoma: multinational phase II-III trial results. J Clin Oncol 19:2456, 2001

81. Olsen E, Duvic M, Frankel A, et al: Pivotal phase III trial of two dose levels of denileukin diftitox for the treatment of cutaneous T-cell lymphoma. J Clin Oncol 19:376, 2001

82. Zackheim HS, Kashani-Sabet M, McMillan A: Low-dose methotrexate to treat mycosis fungoides: a retrospective study of 69 patients. J Am Acad Dermatol 49:873, 2003

83. Kaye FJ, Bunn PA Jr, Steinberg SM, et al: A randomized trial comparing combination electron-beam radiation and chemotherapy with topical therapy in the initial treatment of mycosis fungoides. N Engl J Med 321:1784, 1989

84. Apisarnthanarax N, Talpur R, Duvic M: Treatment of cutaneous T cell lymphoma: current status and future directions. Am J Clin Dermatol 3:193, 2002

85. van Praag MC, Tseng LN, Mommaas AM, et al: Minimising the risks of PUVA treatment. Drug Saf 8:340, 1993

86. Kim YH, Jensen RA, Watanabe GL, et al: Clinical stage IA (limited patch and plaque) mycosis fungoides: a long-term outcome analysis. Arch Dermatol 132:1309, 1996

87. Kim YH, Chow S, Varghese A, et al: Clinical characteristics and long-term outcome of patients with generalized patch and/or plaque (T2) mycosis fungoides. Arch Dermatol 135:26, 1999

XI BENIGN CUTANEOUS TUMORS

Elizabeth A. Abel, m.d.

General Considerations

classification

Tumors of the cutaneous surface may arise from the epidermis, dermis, or subcutaneous tissue or from any of the specialized cell types in the skin or its appendages. Broad categories include tumors derived from epithelial, melanocytic, or connective tissue structures. Within each location or cell type, lesions are classified as benign, malignant, or, in certain cases, premalignant.[1,2]

Benign epithelial tumors include tumors of the surface epidermis that form keratin; tumors of the epidermal appendages; and cysts of the skin.

Melanocytic, or pigment-forming, lesions are very common. One of the most frequently encountered forms is the nevus cell nevus. The term nevus has two meanings: a malformation commonly involving the entire skin layer (tissue nevus) and a benign growth of melanocytic cells (nevus cells).

Nevus cells are closely related to melanocytes and may be defined as modified neuroectodermal melanin-producing elements. The word mole, often used as a synonym for nevus, is an imprecise term because it refers to birthmarks that may or may not contain nevus cells. Neural tumors, such as neurofibromas, are related to melanocytic tumors because both are of neuroectodermal origin.

Tumors that are derived from connective tissue include fibromas, histiocytomas, lipomas, leiomyomas, and hemangiomas.

histologic evaluation

For cases in which it is not possible to distinguish clinically between benign and malignant cutaneous tumors, histopathologic examination is extremely important. The type of biopsy performed depends on the location, size, and nature of the lesion and on cosmetic considerations. In all cases, the clinical features must be correlated with the distinctive microscopic appearance of the tumor to confirm or exclude the diagnosis on the basis of physical examination.

Epithelial Tumors

seborrheic keratosis

Diagnosis and Classification

Seborrheic keratosis (seborrheic wart) consists of a sharply circumscribed, rough or smooth papule or plaque that is 1 mm to several centimeters in size and dirty yellow or light to dark brown in color. The lesions often have the appearance of being stuck on and are characterized by prominent follicular plugging. They are most common in light-skinned races, first appearing in adults on the face and upper trunk and occurring more frequently with increasing age [*see Figure 1*].

Transient eruptive seborrheic keratoses have been associated with inflammatory skin conditions, including erythroderma associated with psoriasis and drug eruptions. These keratoses tend to resolve when the skin inflammation clears.[3] These transient keratoses should be distinguished from eruptive seborrheic keratoses—the sign of Leser-Trelat—which are associated with internal malignancy, particularly adenocarcinoma. The true value of the sign of Leser-Trelat as a marker of underlying malignancy is a subject of debate.

Dermatosis papulosa nigra is similar to seborrheic keratosis, but it is seen in dark-skinned races; it usually appears on the face and presents at an earlier age than seborrheic keratosis [*see Figure 2*].

Differential Diagnosis

The differential diagnosis of seborrheic keratosis and dermatosis papulosa nigra includes lentigo, wart, and nevus cell nevus. A biopsy may be required to rule out a pigmented basal cell carcinoma or, in the case of an inflamed seborrheic keratosis, malignant melanoma or squamous cell carcinoma. A shave biopsy that includes the base of the lesion may be performed before treatment with curettage.

Treatment

Curettage is a satisfactory treatment. When multiple lesions are present, anesthesia may be achieved by freezing the affected area with an ethyl chloride spray before performing curettage. For larger lesions, electrodesiccation is unnecessary and may cause scarring. Smaller lesions may be successfully treated with electrodesiccation, cryotherapy, or topical application of 50% trichloroacetic acid.

epidermal nevus

Diagnosis

Epidermal nevus consists of closely set, skin-colored or hyperpigmented papules that either may be localized to one side of the body and arranged in linear fashion or may be widespread. When localized, the condition is termed nevus unius lateris [*see Figure 3*]. When widespread, it is called systematized nevus. Lesions affect about one in 1,000 people; they are present at birth or appear in early childhood. The lesions have no malignant potential but may constitute a serious cosmetic problem.

Histologically, epidermal nevi exhibit hyperplasia of the epidermis; the structure or maturation of these lesions is not significantly different from that of normal epidermis. One variant, the inflammatory linear verrucous epidermal nevus, shows psori-

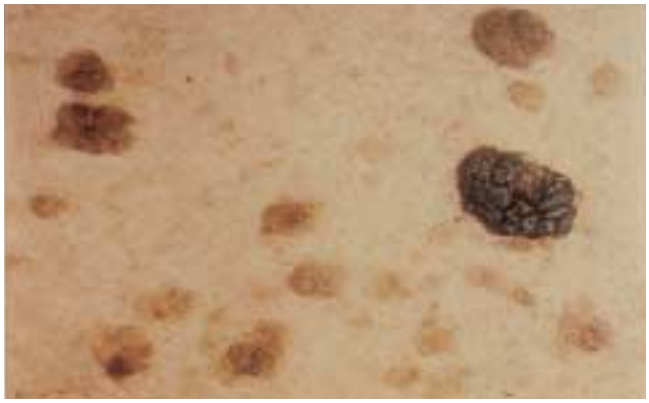


Figure 1 **Verrucous, hyperpigmented lesions of seborrheic keratosis with a stuck-on appearance are present on the trunk of this patient.**

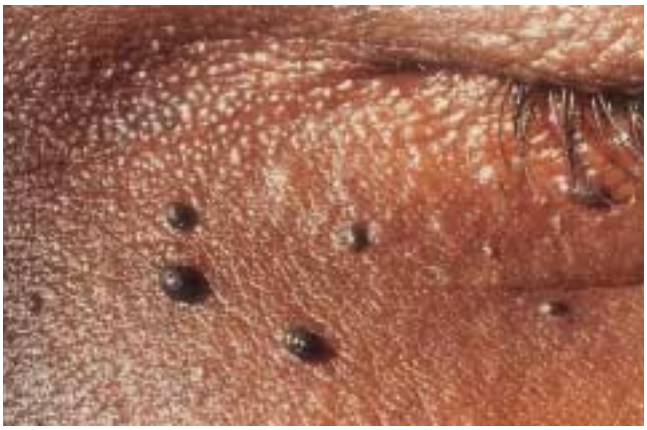


Figure 2 **Dermatosis papulosa nigra, as seen on the face, appears in dark-skinned races at a younger age than seborrheic keratosis.**

Figure 3 **Epidermal nevus with discrete and confluent brown papillomas is present in a somewhat linear arrangement.**

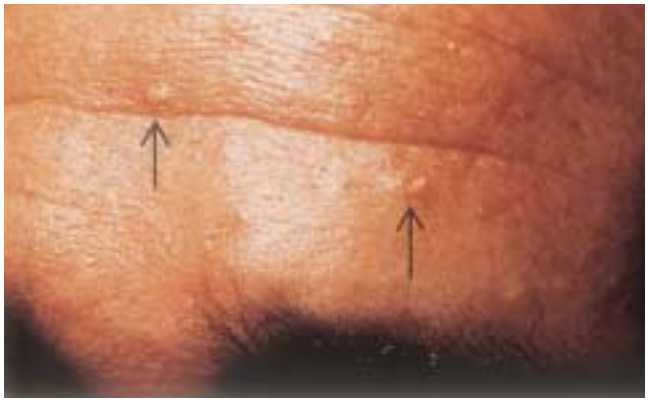


Figure 4 **Skin-colored or yellowish, often umbilicated papules of sebaceous hyperplasia, as seen on the forehead, may clinically resemble basal cell carcinomas.**

asiform hyperplasia. Another variant, which is common in systematized nevi, shows granular degeneration of epidermolytic hyperkeratosis histologically. This type of epidermal nevus is a mosaic genetic disorder of suprabasal keratin. Mutations in the *K10* gene are associated with lesions of the skin, whereas the normal gene is found in unaffected skin.[4]

Variants

The epidermal nevus syndrome involves a spectrum of different types of epidermal nevi associated with disturbances in the skeletal, urogenital, cardiovascular, and nervous systems.[5] This rare syndrome is apparent at birth; the presence of widespread epidermal nevi should trigger a search for associated anomalies.

Nevus comedonicus is a variant of an epidermal nevus affecting the pilosebaceous structures; it occurs as clusters of comedonelike papules, usually in a linear pattern on the face, neck, upper arms, and trunk.[5]

Nevus sebaceous is a benign tumor that shows sebaceous differentiation. The lesion has a yellow hue and a granular surface and occurs in a linear pattern on the face or scalp. At puberty, nevus sebaceous may become more elevated; in adulthood, there is an associated risk of basal cell carcinoma.

Treatment

Treatment of epidermal nevi with electrodesiccation and curettage is often unsuccessful and may cause scarring. Surgical or laser removal may be indicated for localized lesions. Disturbances involving other organ systems must be evaluated and managed appropriately through a multidisciplinary approach.

Tumors of the Epidermal Appendages

There are a large number of benign tumors of the hair follicles, the sebaceous glands, and the apocrine and eccrine glands. Solitary skin tumors of these epidermal appendages are typically nonhereditary, whereas multiple neoplasms may show an autosomal dominant inheritance pattern.[6]

sebaceous hyperplasia

Sebaceous hyperplasia is a common clinical condition that appears as multiple skin-colored or yellowish, often umbilicated papules or plaques, usually on the forehead, nose, or cheeks of persons after the fifth decade of life. These lesions consist of enlarged sebaceous gland lobules with a central dilated duct. Sebaceous hyperplasia may respond to cryotherapy with liquid nitrogen or the application of a dilute solution of trichloroacetic or bichloroacetic acid. Lesions may sometimes be confused clinically with basal cell carcinoma [*see Figure 4*]. In the familial form of this disorder, onset occurs in puberty; with the passage of time, the lesions increase in extent over the face, neck, and upper thorax. This condition must be distinguished from acne vulgaris, rosacea, and the angiofibromas of tuberous sclerosis. Three patients with this condition responded favorably to oral isotretinoin at a dosage of 1 mg/kg/day. To maintain the response, this dosage was tapered after 6 weeks.[7] Isotretinoin is a known teratogen that cannot be given to women of childbearing age unless strict precautions are observed [*see 2:XII Acne Vulgaris and Related Disorders*].

trichoepitheliomas

Trichoepitheliomas usually present as multiple yellowish-pink, translucent papules distributed symmetrically on the cheeks, eyelids, and nasolabial areas [*see Figure 5*]. Often inherited as an autosomal dominant trait, the papules first appear at puberty and grow slowly for years. The gene for multiple familial trichoepitheliomas has been mapped to chromosome 9p21.[8] Lesions may be confused both clinically and histologically with basal cell carcinoma, though trichoepithelioma usually shows differentiation toward the formation of hair. A single or localized trichoepithelioma may be removed by electrodesiccation and curettage. Multiple lesions are difficult to treat and may be a cosmetic problem.

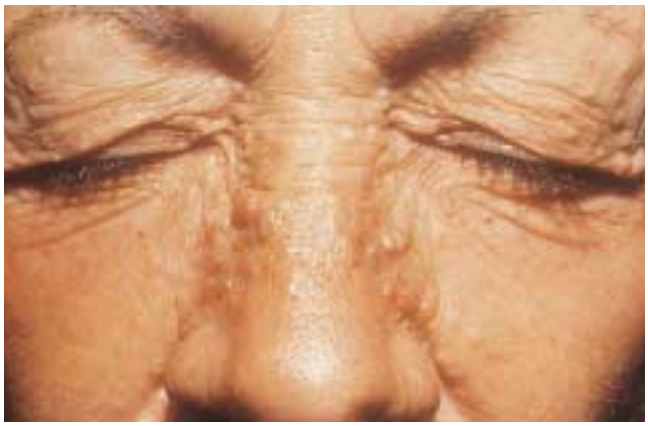


Figure 5 **Symmetrical papules of trichoepithelioma appear on the eyelids and nasolabial areas and may be inherited as an autosomal dominant trait.**

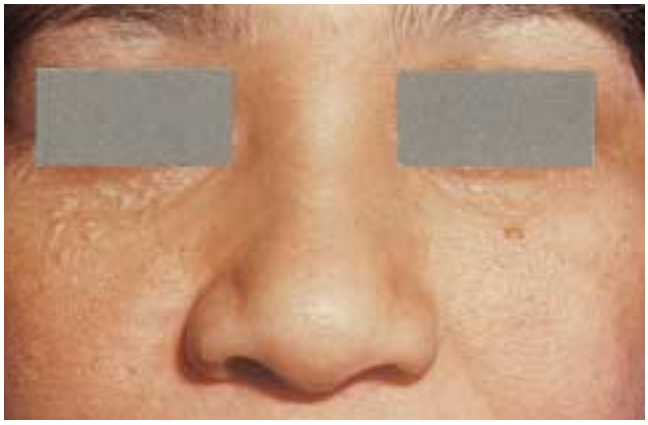


Figure 6 **Syringomas—benign tumors of eccrine ducts—are commonly seen on the face, especially on the lower eyelids.**

syringomas

Syringomas usually present in groups of multiple small papules that are distributed symmetrically over the face, especially on the lower eyelids [*see Figure 6*]. Eruptive syringoma, a rare condition, is characterized by widespread lesions.

Histologically, there is a benign proliferation of the eccrine ducts.

epidermoid cyst

Diagnosis

Commonly called wens, epidermoid cysts have a lining that resembles the epidermis. Several types of cyst exist, but they are usually clinically indistinguishable from one another. On histologic examination, most of these cysts appear to be derived from hair follicles.

The epidermoid cyst is commonly located on the back and consists of one or more slow-growing, elevated, firm nodules, often with a central pore [*see Figure 7*]. The diameters of the lesions vary from 0.2 to 5.0 cm.

Treatment

The epidermoid cyst may be incised with a pointed scalpel to express its wall and contents, which consist of a thick keratinous material. If the cyst wall is not completely removed, there may be a recurrence of the lesion. Occasionally, the entire cyst has to be excised. Preliminary treatment with a systemic antibiotic, such as erythromycin, and warm-water compresses applied three or four times daily may be instituted if the cyst is inflamed and infected. When the inflammation and infection resolve, the lesion can be removed. Repeated episodes of infection may cause fibrosis, after which the cyst may have to be surgically excised.

Other Cysts

The pilar cyst, which is less common, has a wall that contains keratin similar to that found in hair. The contents of these cysts are semifluid and often have a rancid odor.

A milium is similar to an epidermoid cyst but differs mainly in size. Milia are white, hard subepidermal keratin cysts, 1 to 2 mm in diameter, that commonly arise spontaneously on the face [*see Figure 8*]. They may also arise secondarily in scars or in association with certain bullous diseases. Incision and expression of contents with a comedo extractor may be performed.

Familial Tumor Syndromes

Multiple cutaneous neoplasms may be a feature of familial tumor syndromes that are thought to be mediated by inactiva-

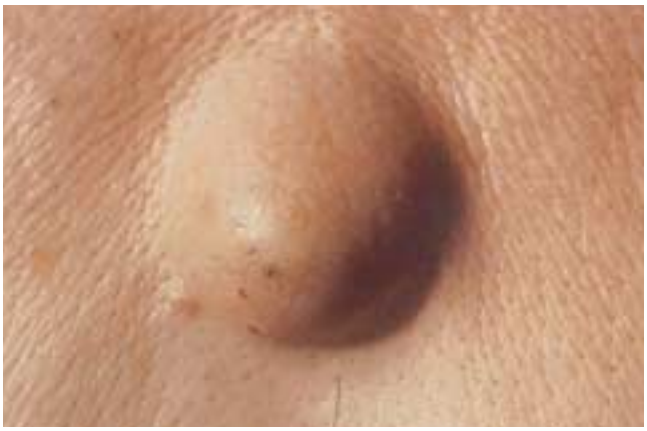


Figure 7 **This large epidermoid cyst has a central pore, contains thick keratinous material, and has a lining that resembles the epidermis.**

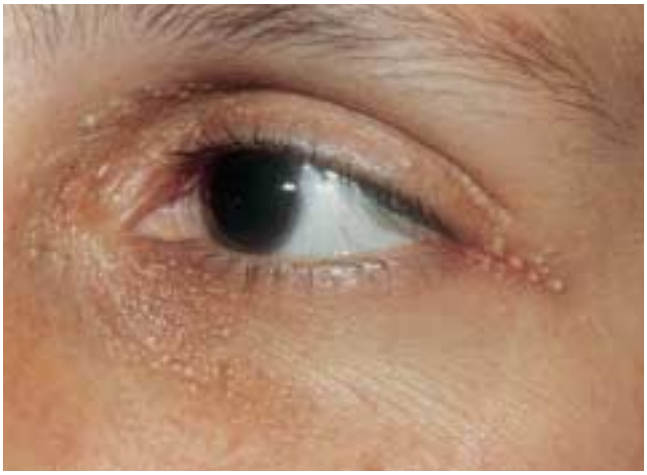


Figure 8 **Milia, which are multiple small subepidermal inclusion cysts, can be observed in the periorbital area of this patient.**

tion of tumor suppressor genes. It is important to recognize these syndromes because they may be associated with underlying malignancies.

muir-torre syndrome

Muir-Torre syndrome (MTS), previously known as Torre syndrome, consists of sebaceous gland neoplasms that are associated with visceral carcinoma and that arise from colonic epithelium. Sebaceous gland tumors may include, in decreasing order of frequency, adenomas, epitheliomas, and carcinomas.[9] Keratoacanthomas and sebaceous hyperplasia are also seen in patients with MTS. Colorectal cancer develops in 51% of patients with MTS a decade earlier than it develops in the general population. Genitourinary cancer develops in 24% of MTS patients. A germline mutation in the DNA mismatch repair gene *hMSH2* has been identified in patients with MTS. Predictive diagnosis in family members should be preceded by careful genetic counseling.[9]

gardner syndrome

Gardner syndrome consists of the triad of intestinal polyposis, bony tumors, and soft tissue lesions; it has an autosomal dominant inheritance. The colonic polyps eventually become malignant if left untreated. Soft tissue lesions include epidermoid cysts, sebaceous cysts, desmoid tumors, and scattered lentigines on the head and extremities.[10]

cowden syndrome

Cowden syndrome is characterized by facial trichilemmomas and acral fibromas, and it is associated with an increased risk of cancer of the breast, thyroid, and gastrointestinal tract. This rare genodermatosis, which is also known as multiple hamartoma syndrome, is inherited as an autosomal dominant trait. It is important to make a prompt diagnosis of this syndrome because of the high risk of malignancy, particularly cancer of the breast in women.

birt-hogg-dubé syndrome

Birt-Hogg-Dubé syndrome (BHDS) is an autosomal dominant multisystem disorder characterized by the cutaneous triad of fibrofolliculomas, trichodiscomas, and acrochordons. Fibrofolliculomas are benign tumors of the hair follicle. Fibrofolliculomas are firm, pink or skin-colored papules measuring 1 to 3 mm that appear on the face, particularly the nose, earlobes, and forehead. In the original kindred described by Birt (a dermatologist), Hogg (a pathologist), and Dubé (a pathologist), family members were afflicted with medullary carcinoma of the thyroid. Subsequently, there appeared reports of patients with BHDS who had intestinal polyps, adenocarcinoma of the colon, parathyroid adenomas, and renal cell carcinoma. The skin tumors begin in early adulthood; systemic tumors appear years later. In families with recognized renal cell carcinoma, BHDS may account for 6% of the cases.[11,12]

Melanocytic (Pigment-Forming) Tumors

Benign tumors of pigment-forming cells, including those containing nevus cells (melanocytic nevi) and those of epidermal or dermal melanocytes, are of neuroectodermal origin.

melanocytic nevus

Melanocytic nevus, also called nevus cell nevus, has a characteristic life history of evolution and involution. Melanocytic nevi are the most common of all skin tumors; each young adult has an average of 20 to 40 of them. Their incidence increases with age up to the second or third decade of life, after which they occur less commonly.

Risk Factors for Melanoma

An increase in the total number of melanocytic nevi is a risk factor for melanoma.[13] In a study of 716 patients with newly diagnosed melanoma, an increased number of small nevi (25 to 49) was associated with a twofold increase in risk of melanoma; greater numbers of nevi were associated with further increased risk.[14] The presence of one clinically dysplastic nevus was associated with a twofold increase in risk of melanoma; and 10 or more, with a 12-fold increase in risk. Criteria for dysplastic nevi included large size (over 5 mm), flatness (entirely macular or having a macular component), and at least two of the following: irregular pigmentation, asymmetry, and indistinct borders [*see 2:X Malignant Cutaneous Tumors*]. The presence of freckling conferred additional risk of melanoma for all types of nevi.

The relation between sun exposure and melanocytic nevi has been investigated to determine what environmental factors influence melanoma and to facilitate preventive measures. Studies suggest that melanocytic nevi are more common on sun-exposed skin sites and reach a peak incidence earlier in age on these sites than on covered areas of the body.[15] A study of Australian schoolchildren showed an increasing prevalence of nevi with decreasing latitude, particularly in children 6 and 9 years of age.[16] Sun exposure during childhood was considered to be a factor in the development of melanocytic nevi and an associated risk factor for melanoma.[16] In Australia, however, sun exposure may be sufficient to maximally induce nevi regardless of latitude. Further studies need to be performed on persons living at higher latitudes to see whether the relation between sun exposure and nevi continues into adulthood.

Diagnosis

A melanocytic nevus that is present at birth or appears during the first year of life is considered to be congenital. Certain syndromes are associated with congenital nevi, including epidermal (linear sebaceous) nevus syndrome, neurocutaneous melanosis, premature-aging syndrome, and occult spinal dysraphism or tethered cord syndrome.[17] Various neuroectodermal defects and multisystem abnormalities may also be present. Giant congenital melanocytic nevi are associated with an increased risk of melanoma (see below).

Acquired melanocytic nevi vary considerably in form, ranging from flat to pedunculate. They may be hairy or hairless and may be skin colored, dark brown, or even black. Nevi that are flat and darkly pigmented are called junctional nevi. Slightly raised nevi are often compound; that is, they contain both epidermal and dermal components. Nevi that are predominantly intradermal are usually more elevated and contain less pigment than compound or junctional nevi. Nevi that are papillomatous, dome shaped, or pedunculate are usually intradermal [*see Figures 9 through 11*].

Differential Diagnosis

The differential diagnosis of melanocytic nevi includes ephelis (freckle), lentigo, café au lait spot (see below), wart, seborrheic keratosis, and skin tag (a small pedunculate protrusion of skin that does not contain nevus cells). Ephelis is a tan macule, commonly seen in children after sun exposure; it often dis-

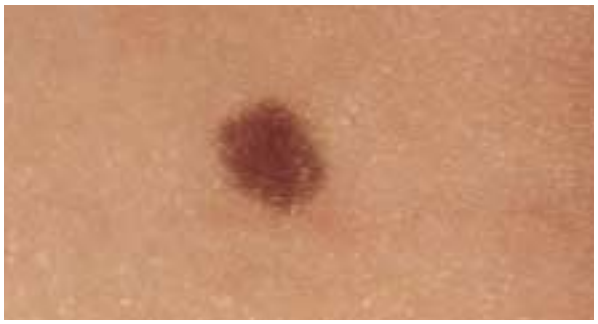


Figure 9 **A flat junctional nevus with dark pigmentation is seen in this patient.**

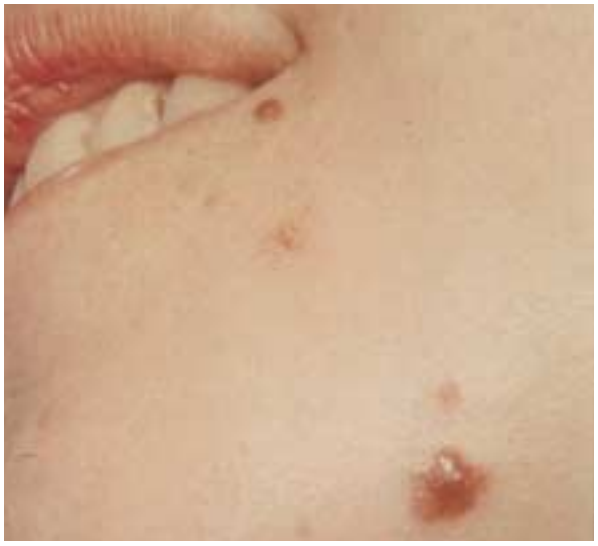


Figure 10 **This slightly raised compound nevus typically has less pigmentation than a junctional nevus.**

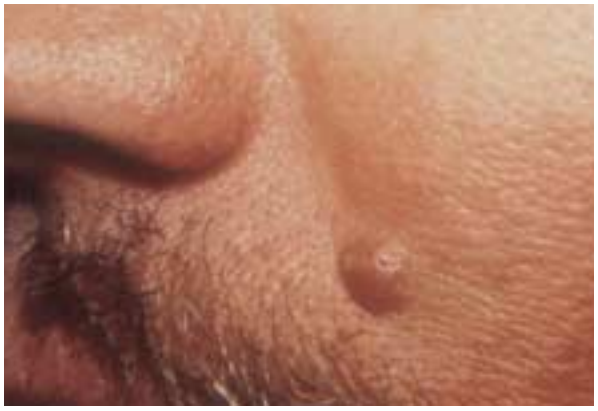


Figure 11 **A skin-colored intradermal nevus with a dome-shaped configuration is seen on the face.**

appears in the winter. Lentigo, also called senile lentigo or liver spot, is a tan or brown macule commonly seen on exposed skin areas, such as the face, the backs of the hands, and the neck. The labial melanotic macule is a distinct entity that appears in adults as a well-defined brown or black pigmented macule on the lip. In a study of 79 patients, the majority of melanocytic lesions (94%) were on the central third of the lower lip, suggesting that exposure to ultraviolet light has a causative role.[18] Patients followed for up to 13 years had no adverse developments, a finding indicative of the benign nature of this lesion.

Treatment

No treatment is required for melanocytic nevi. However, shave biopsy or excisional biopsy may be performed for cosmetic reasons or when a nevus is subject to irritation because of pressure from clothing or because it is located in an intertriginous area. Patients should be followed with serial photographs. Biopsy should be performed for nevi that appear prone to malignant transformation; nevi that show severe dysplasia should be removed. Removal of mildly or moderately dysplastic nevi is advocated by some but not all experts [*see 2:XI Malignant Cutaneous Tumors*].

café au lait spots

Café au lait spots are common benign congenital or acquired birthmarks. They are tan, round to oval macules ranging in size from several millimeters to 10 to 20 cm. They can occur on any area of the body but are more common on the trunk, buttocks, and lower extremities. The presence in a prepubertal child of five or more café au lait spots larger than 0.5 cm may be a marker for neurofibromatosis-1 (NF-1) (see below).[19] Histologically, café au lait spots show an increased number of dihydroxyphenylalanine (DOPA)-positive melanocytes that produce an increased concentration of melanosomes. The café au lait spots seen in Albright hereditary osteodystrophy are usually unilateral and show jagged rather than smooth margins. An association of juvenile xanthogranulomas with café au lait macules carries an increased risk of underlying systemic disorders, including leukemia.[20]

halo nevus

A halo nevus consists of an acquired zone of hypopigmentation surrounding a pigmented tumor, most commonly a compound nevus [*see Figure 12*]; other tumors, even malignant melanoma, may also be surrounded by a depigmented halo. The halo lesion typically involutes during a period of months in the absence of clinical signs of inflammation. Histologically, a chronic lymphocytic infiltrate surrounds the nevus cells, which may represent an autoimmune phenomenon.

spindle cell nevus

Formerly called benign juvenile melanoma, spindle cell nevus usually arises in childhood as a pink or reddish-brown, smooth or slightly scaly, firm papule with a predilection for the face, especially the cheeks [*see Figure 13*].[21] Although benign, spindle cell nevus may closely resemble a malignant melanoma. Excisional biopsy is therefore advisable in many cases.

mongolian spot

The mongolian spot is a bluish macule that is seen in newborns of dark-skinned races. The discoloration is caused by persistence of dermal melanocytes, often in the lumbosacral region [*see Figure 14*]. The lesion usually disappears by 3 or 4 years of age.

blue nevus

The common blue nevus occurs as a solitary, sharply circumscribed, blue-black papule [*see Figure 15*]. This malformation consists of a group of melanocytes with long, thin surface projections in the middle and lower thirds of the dermis and

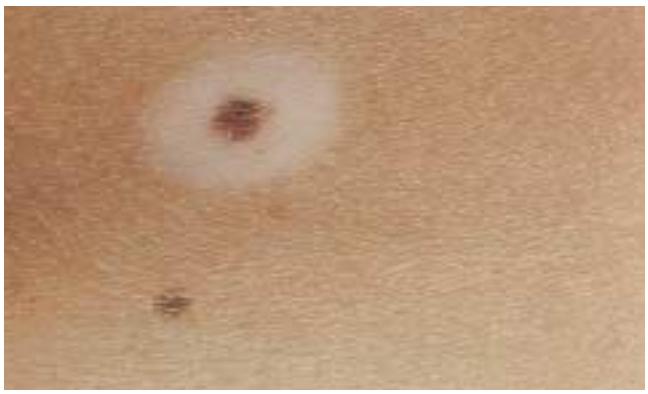


Figure 12 **The halo nevus may represent an autoimmune phenomenon; a zone of hypopigmentation may appear around a nevus, with subsequent involution of the pigmented tumor.**

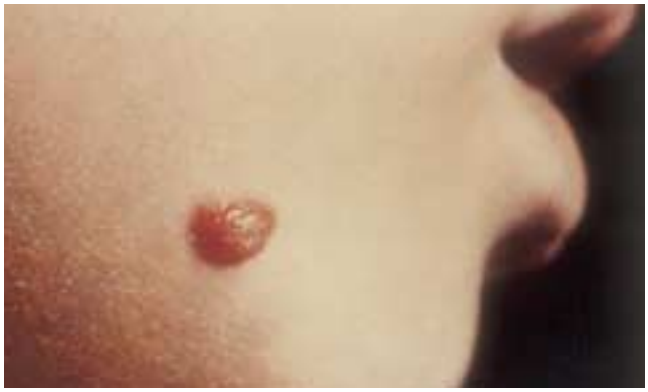


Figure 13 **The spindle cell nevus is an active compound nevus that may be difficult to distinguish histologically from a melanoma.**

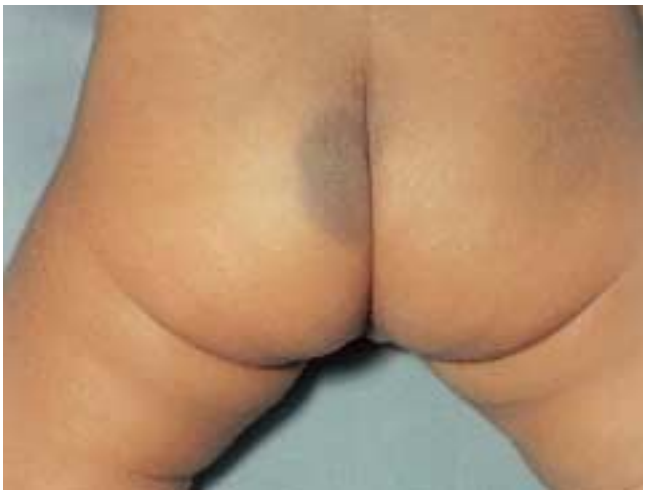


Figure 14 **The bluish pigmentation of a mongolian spot is seen in the lumbosacral area and is caused by the persistence of dermal melanocytes.**

in subcutaneous fat. The common blue nevus does not show a tendency toward malignant transformation. The cellular blue nevus, which appears as a blue-black nodule or an indurated plaque, contains two types of cells: spindle shaped and rounded. The cellular blue nevus may in rare instances become malignant.

nevus of ota

The nevus of Ota occurs in infancy or appears in adolescence as a blue-gray macule in the distribution of the trigeminal nerve. The lesion is unilateral in 90% of cases. Asian females are most commonly affected. Histologically, a benign dendritic melanocytosis is present in the papillary and upper reticular dermis. High-energy fluences of the Q-switched ruby laser results in lightening of the lesion, without scarring, after a few treatments.[22]

becker nevus

A malformation of epidermal melanocytes, Becker nevus occurs as a large area of hyperpigmentation and increased hair growth and is usually located on one shoulder. It appears most commonly in males during adolescence [*see Figure 16*]. Underlying bony and soft tissue abnormalities may be associated with this disorder.[23]

Light microscopy reveals hyperpigmentation of the basal layer of the epidermis, with melanin-containing phagocytes in the dermis but no nevus cells.

congenital giant pigmented nevus

Giant pigmented nevus is an uncommon birthmark appearing sporadically in one in 20,000 live births. Its features are different from those of an ordinary acquired nevus. Lesions are often darkly pigmented, hairy, and slightly infiltrated, eventually becoming verrucous or nodular. They tend to occur in the distribution of a dermatome and may be quite extensive, as in bathing trunk nevus [*see Figure 17*]. Satellite lesions may be present. The condition not only is of cosmetic concern but also has a high association with malignant melanoma, with a reported 10% to 15% of nevus patients developing melanoma. Histologic features of an ordinary compound nevus, an intradermal nevus, a neural nevus, or a blue nevus may be present.[1] Treatment consists of multiple operations to excise as much of the lesion as possible.

neurocutaneous melanosis

Lesions on the scalp and neck may be associated with neurocutaneous melanosis of the leptomeninges that can be complicated by epilepsy, mental retardation, or central nervous system melanoma. Large congenital melanocytic nevi (LCMN) carry a poor prognosis in the presence of CNS signs or symptoms such as abnormal reflexes, hydrocephalus, and papilledema. Posterior axial LCMN, especially in association with satellite nevi, is a risk factor for CNS melanosis. Magnetic resonance imaging should be considered in the evaluation of newborns with these findings. In one study, CNS involvement occurred in 33 of 289 patients with LCMN. All the patients with CNS involvement had nevi in the posterior axial location. Satellite nevi were present in 31 of the 33 patients.[24] These findings suggest that melanocytic malformation occurs during the migration of neural crest cells that give rise to cutaneous leptomeningeal melanocytes. Malformation resulting in LCMN on the extremities occurs after migration from the neural crest and is not associated with CNS melanosis.

Neural Tumors

Neural tumors, such as neurofibromas, are of neuroectodermal origin, as are melanocytic tumors. Neurilemmomas (also called schwannomas) are benign nerve sheath tumors that extend subcutaneously adjacent to a peripheral nerve. They usually occur in solitary form but may occur as multiple lesions in the

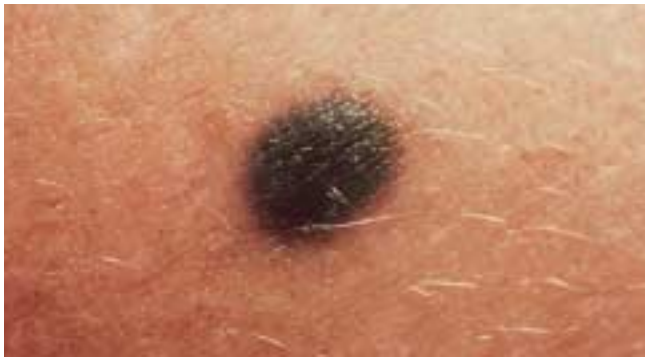


Figure 15 **The presence of melanocytes in the middle and lower dermis is responsible for the color of the blue nevus.**

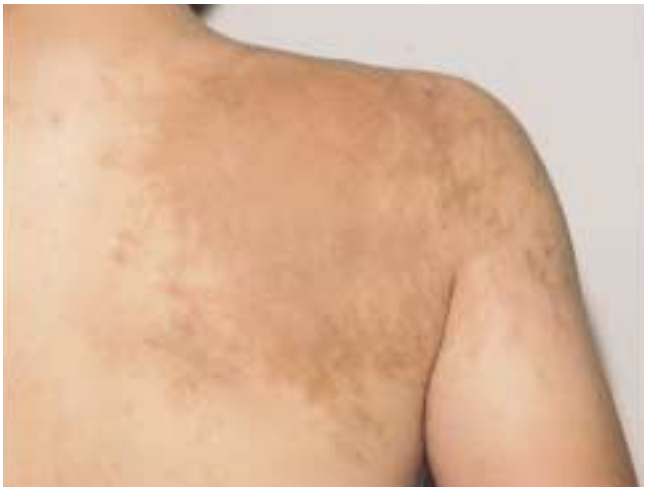


Figure 16 **Becker nevus, an acquired localized malformation of epidermal melanocytes that may be associated with hypertrichosis, is seen on the shoulder.**

syndrome of neurilemmomatosis.[25] These tumors are usually painful and may be associated with nerve compression. Other benign tumors that must be considered in the differential diagnosis of painful skin nodules are neuromas, angiolipomas and angiomyolipomas, leiomyomas, eccrine spiradenomas, glomus tumors, and the blue rubber bleb nevus.

neurofibromatosis

Neurofibromatosis represents a spectrum of disorders involving the skin, central and peripheral nervous systems, bones, and blood vessels. This neurocutaneous syndrome is transmitted via an autosomal dominant gene at an estimated frequency of one in 3,000 persons with almost complete penetrance.[26]

Diagnosis and Classification

Two distinct forms of neurofibromatosis are recognized, but variant forms also exist.

Neurofibromatosis-1 The most common form (occurring in 85% to 90% of all cases) is NF-1, or von Recklinghausen disease [*see Figure 18*]. This is a common autosomal disorder, with an incidence of one in 3,500 persons. It is characterized by the presence of café au lait spots, intertriginous freckling, multiple spinal and peripheral neurofibromas, plexiform neuromas, bilateral iris hamartomas (also known as Lisch nodules), neurologic impair-

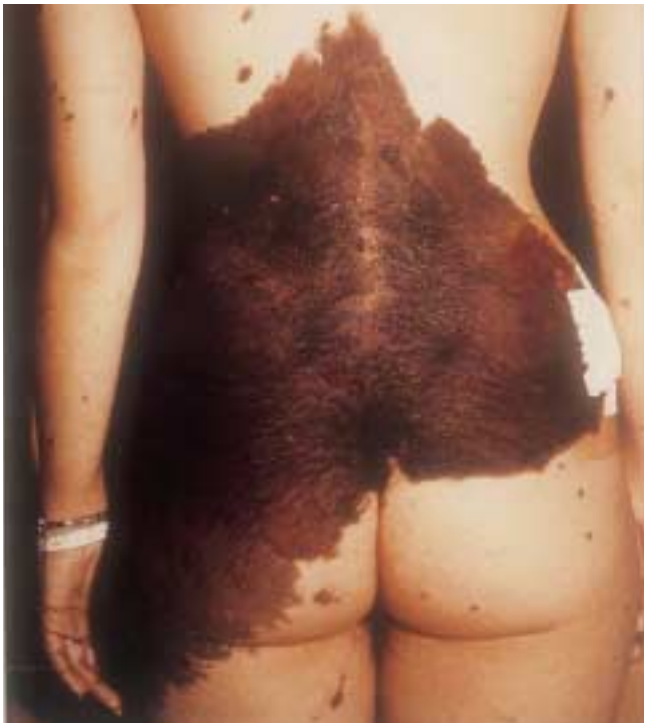


Figure 17 **This form of congenital giant pigmented hairy nevus is associated with an increased risk of malignant melanoma, which develops within the lesion.**

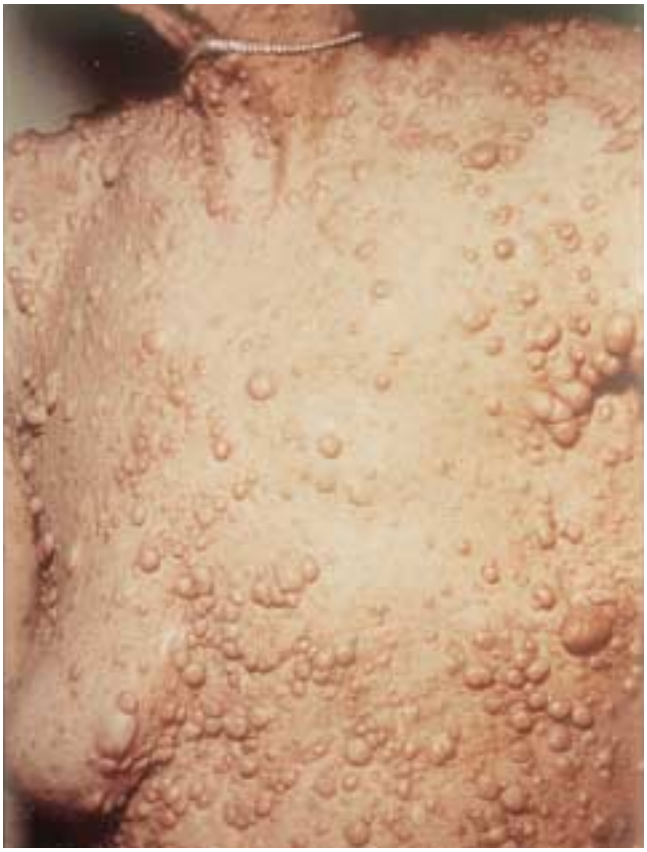


Figure 18 **Multiple neurofibromas, as seen in von Recklinghausen disease, usually appear in late childhood and increase in size and number with age.**

ment, and bone abnormalities. The disease is progressive and is associated with a predisposition to a malignant state.

Sarcomatous degeneration of skin lesions is rare but may occur in extracutaneous tumors. Café au lait spots of NF-1 may be present at birth and may be best visualized under a Wood light. Neurofibromas begin to appear at puberty as soft, globoid, and pedunculated tumors that are skin colored or violaceous. Lesions may be large and numerous, causing complications resulting from impingement on surrounding structures.

Neurofibromatosis-2 A second form of the disease, neurofibromatosis-2 (NF-2), is characterized by bilateral acoustic neuromas, which are Schwann cell tumors that arise from vestibular nerves.[27] Associated features may include meningiomas, gliomas, paraspinal neurofibromas, and subcapsular cataracts. Skin tumors and café au lait spots are less commonly seen in NF-2 than in NF-1.

Variants Other forms of neurofibromatosis include segmental cases in which café au lait spots or neurofibromas are localized to a single dermatome. The gene for NF-2 is located on chromosome 22. In this autosomal dominant disorder, the *merlin* tumor suppressor gene encoded in chromosome band 22q12 is inactivated. This results in an alteration in DNA with substitution of tyrosine for asparagine at position 220 of the merlin cytoskeletal associated protein.[28]

Genetic Counseling

Patients with either NF-1 or NF-2 should seek genetic counseling because there is a 50% risk that their offspring will also be affected with neurofibromatosis. In NF-1, optic glioma can appear in early childhood; patients with NF-1 may also have scoliosis. In NF-2, bilateral acoustic neuromas can cause deafness. The genes for the two distinct forms of neurofibromatosis have been located on two separate chromosomes. This finding may lead to improved diagnosis, which would facilitate genetic counseling and enable prenatal testing.[27]

Treatment

For treatment of selected neurofibromas, surgical excision is more successful than scalpel removal or electrodesiccation and curettage. In a preliminary study, the use of ketotifen, a benzocycloheptathiophene compound that acts as a mast cell stabilizer, was evaluated in the treatment of patients with neurofibromatosis.[29] All treated patients showed a decrease in symptoms of pruritus, pain, or skin tenderness and experienced a decreased rate of neurofibroma growth. Long-term double-blind studies are required, however, to confirm and extend these preliminary findings.

The bilateral acoustic neuromas of NF-2 may be visualized by computed tomography or MRI. Hearing loss is an early symptom that may begin in the second or third decade of life; it can be detected by an audiologic study with brain stem auditory-evoked response. Unilateral acoustic neuromas that are not associated with neurofibromatosis and that are not inherited are more common in older persons and pose fewer management problems.[27] Surgical removal of small acoustic neuromas may improve neurologic or audiologic status.

Connective Tissue Tumors

Fibroma of the skin comprises multiple conditions that may represent reactions to hemorrhage, infection, or chronic irritation.

skin tag

Skin tag, also called acrochordon, commonly occurs as multiple skin-colored or tan, filiform or smooth-surfaced papules that are 2 to 3 mm in diameter. Lesions are often located on the neck or axillae but may also appear in the groin or on the extremities, often as isolated larger polypoid growths [*see Figure 19*]. The fibrous stalk consists of loose connective tissue with dilated capillaries. Lesions may become inflamed if they are irritated or are traumatized from twisting of the stalk. Biopsy is performed if the clinical diagnosis is uncertain. Skin tags may be removed for cosmetic reasons by using scissors to clip the pedunculate lesions at the base.

dermatofibroma

Dermatofibroma, also called histiocytoma, is a firm, skin-colored or reddish-brown sessile papule or nodule that arises spontaneously or after minor trauma, usually in adults [*see Figure 20*]. A dermatofibromatous lesion may occur, for example, after an insect bite on an extremity. A solitary lesion is most common, though multiple or eruptive histiocytomas have been reported. It may be necessary to perform a biopsy when the diagnosis is uncertain. Treatment is necessary only for cosmetic reasons.

keloid and hypertrophic scar

Normal wound healing in response to tissue injury involves several integrated processes: inflammation, production of granulation tissue, formation of the extracellular matrix, wound con-

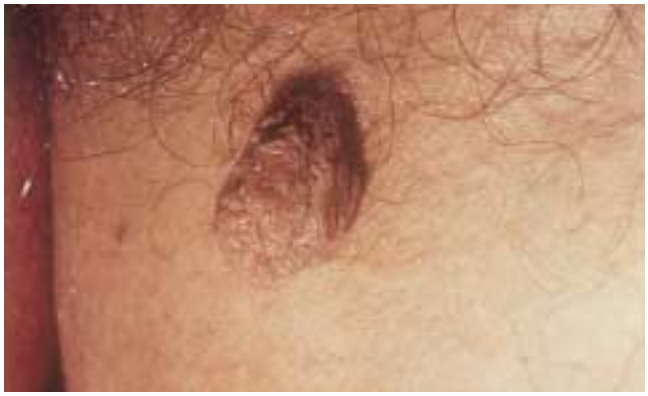


Figure 19 **Skin tags, also called acrochordons or soft fibromas, are skin-colored or tan papules. They are commonly seen in such intertriginous areas as the groin or axillae.**

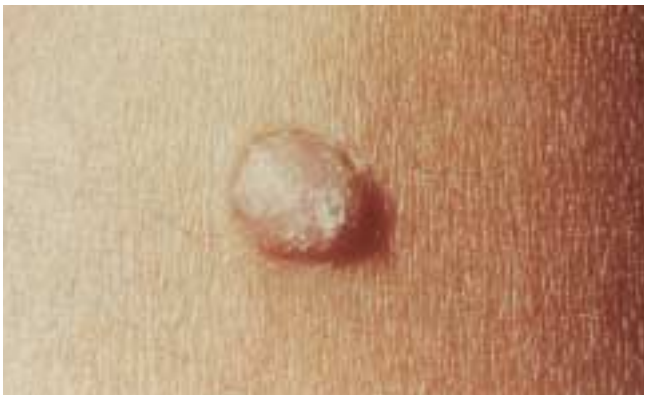


Figure 20 **Dermatofibroma appears as a firm skin-colored or reddish-brown papule and may arise spontaneously or follow minor trauma to the skin.**

traction, and, finally, scar formation. In the final phases of wound healing, fibroblasts degrade and produce bundles of collagen fibers. These bundles become thicker and are aligned along the lines of tension to which the tissues are exposed. As a result of these changes, wound tensile strength gradually increases. The resulting scar is relatively acellular and has fewer macrophages, blood vessels, and fibroblasts than the unwounded tissue.

Diagnosis and Classification

Scars may be normotrophic, atrophic, hypertrophic, or keloidal. Both hypertrophic and keloidal scars are abnormal responses to tissue injury. Hypertrophic scars mature and flatten over time, usually after 6 months. The keloid appears as a shiny, smooth, raised proliferation of scar tissue with typical crablike extensions beyond the site of the original injury [*see Figure 21*]. Keloids differ from hypertrophic scars in that their development is delayed, sometimes occurring months after tissue injury. Keloids do not regress, and they frequently cause pain, itching, and burning. Keloids are more common in African Americans, Hispanics, and persons with a personal or family history of keloids. Other factors associated with the development of keloids include wound tension, especially in skin sites such as the chest, shoulders, and back; ear piercing; healing by second intention; pregnancy; young age; and deep laceration.[30]

In atrophic scars, there is thinning of the skin and loss of normal architecture. Striae distensae, a so-called stretch mark, is a common dermal atrophic scar that tends to appear during periods of rapid weight gain and in the presence of excess glucocorticoid, as well as late in gestation.

Treatment

Treatment with intralesional steroids, 10 to 40 mg / ml once a month for up to 6 months, can effectively flatten keloid and hypertrophic scars. Cryotherapy (a 30-second application once a month for 3 months) has been found to be safe and effective.[31] Topical silicone gel sheeting, which was first used for burn scars, has been used in the treatment of keloids and hypertrophic scars.[32] There is no release of silicone into the skin, and there are no adverse side effects from this treatment. The mechanism of action is unknown. Potential side effects of intralesional corticosteroid treatment include atrophy, depigmentation, telangiectasia, and ulceration and dose-related systemic effects.

Vascular Birthmarks

Vascular proliferations are broadly classified as hyperplasias that show a tendency to regress or as benign vascular tumors that persist.[33,34] Vascular hyperplasias include pyogenic granuloma and pseudo-Kaposi sarcoma. Vascular hemangiomas can be further subdivided according to their histologic cell of origin (endothelial cell, pericyte, glomus cell), depth of tissue involvement (superficial or deep), and size of involved vessels (capillaries, venules, arterioles, veins, or arteries). Vascular birthmarks such as nevus flammeus and salmon patch may resemble angiomas but are nonproliferative malformations that usually do not involute.

epidemiology

Hemangiomas (see below) occur in a female-to-male ratio of 5:1, whereas vascular malformations occur with equal frequency in males and females. A rare familial occurrence of hemangiomas, vascular malformations, or both has been reported in six kindreds, suggesting autosomal dominant inheritance in these cases.[35]

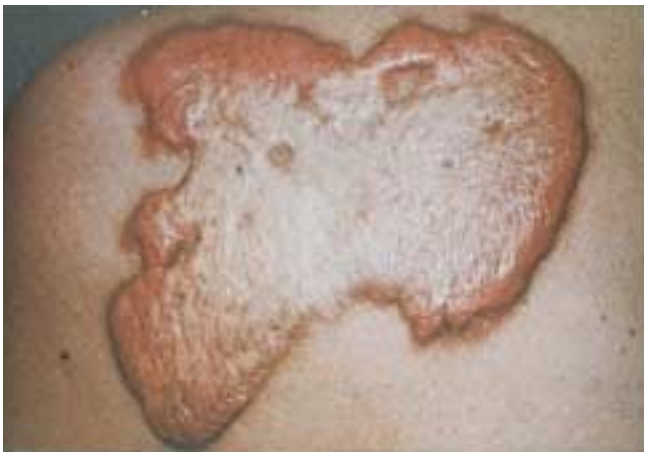


Figure 21 **The proliferation of scar tissue in a keloid may extend beyond the original site of injury.**

Vascular malformations are congenital developmental defects that are generally of unknown etiology. Port-wine stains may result from progressive ectasia of the superficial vascular plexus in the skin as a result of abnormal neural regulation of blood flow.[36] In the Klippel-Trénaunay-Weber syndrome, a mesodermal abnormality affecting differentiation of the limb bud may occur during the third to sixth week of gestation.[37]

pathogenesis

The etiopathogenesis of hemangiomas and vascular malformations is not well understood. Hemangiomas arise in response to an angiogenic stimulus that may begin in utero. Through use of immunohistochemical techniques, infantile hemangiomas and placental microvessels were found to coexpress the vascular antigens GLUT-1 and Lewis Y antigen (LeY).[38] These antigens are not present in other vascular tumors, such as pyogenic granulomas, or in vascular malformations. A pathogenic link involving aberrant differentiation of vascular precursor cells or embolization of placental cells to fetal tissue has been hypothesized.[38] These antigens are also absent in congenital nonprogressive hemangioma, a distinctive hemangioma consisting of lesions that are fully formed at birth and that either remain static or rapidly involute.[39]

overview of management

Evaluation and management of hemangiomas and malformations require a multidisciplinary approach. Specific diagnosis may be aided by imaging techniques such as CT and MRI to assess depth of involvement and extension to adjacent structures and to evaluate associated abnormalities. Laboratory evaluation for associated systemic disease may be required in addition to ophthalmologic, neurologic, and cardiologic assessment for complications of vascular tumors and dysmorphic syndromes.

hemangiomas

Hemangiomas are proliferating vascular tumors that are not necessarily present at birth. The vascular lesion may appear in neonates as a faint pink patch that subsequently undergoes rapid proliferation over a period of months to years before the lesion stabilizes and regresses.

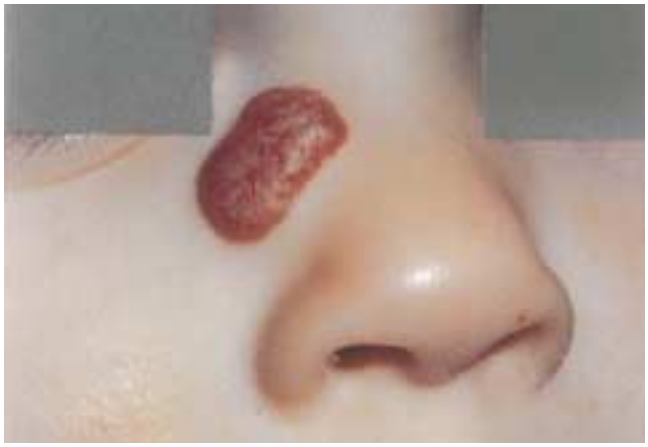


Figure 22 **The strawberry, or capillary, hemangioma appears between the second and fifth weeks of life and undergoes spontaneous involution over a period of several years.**

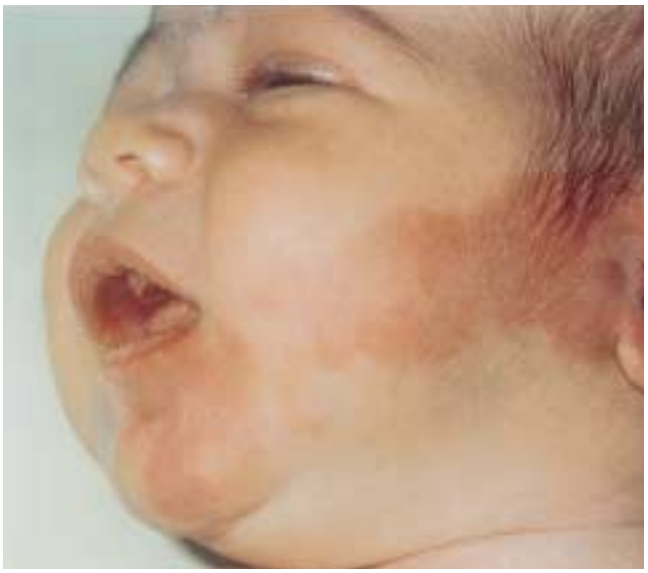


Figure 23 **A nevus flammeus is present at birth as a reddish or violaceous macular discoloration, often in a unilateral and segmental distribution; it shows little tendency to involute later in life.**

The biologic classification of hemangiomas is very different from that of vascular malformations. Vascular tumors can be classified according to their cell or origin, the size of the involved vessels, and the depth of involvement. Such classifications have led to refinement in terminology.[40] The terms strawberry hemangioma and cavernous hemangioma are descriptive clinical terms that do not specify the type of vessels that are involved.

Diagnosis and Classification

Capillary hemangioma, also known as strawberry hemangioma, appears as a single vascular lesion or multiple lesions during the second to the fifth week of life. Infantile hemangiomas are bright-red, soft, lobulated tumors that increase in size for a period of months [*see Figure 22*]. Lesions spontaneously involute, sometimes with fibrosis, over a period of several years.[33] Histologically, the capillary hemangioma shows a proliferation of endothelial cells that form many new small vessels.

Treatment

It is important to realize that most hemangiomas are uncomplicated and regress without treatment early in life with minimal residual scarring. Follow-up studies have shown that in 90% of patients, hemangiomas regress by 9 years of age.[41] Parents may require considerable reassurance that the best course is to refrain from treatment. Care must be taken to prevent trauma and infection, which may lead to scarring.

There is considerable controversy as to when to intervene in the treatment of complicated hemangiomas because of potential side effects, such as scarring. The ideal time to treat would be at the beginning of the period of rapid growth, but this is difficult to predict. Indications for treatment include involvement of a vital orifice, infection, ulceration, ocular involvement, and severe cosmetic deformity. Medical options include intralesional or systemic steroids, the latter at a dose of 1 to 3 mg/kg/day. Antimetabolites have been used for their antiproliferative effect. Interferon alfa has been used for severe hemangiomatosis, but its use is associated with systemic side effects and the potential risk of spastic diplegia. Laser surgery with 585 nm pulsed dye laser may be used to treat the superficial proliferative component.[41] Radiation therapy may lead to scarring and is discouraged in children because of long-term radiation effects, including risk of malignancy. Interventional techniques involving embolization of vessels may be required in cases involving airway obstruction or other life-threatening complications. A multidisciplinary team approach involving the dermatologist, pediatrician, radiologist, surgeon, and other specialists is needed for optimal management of complicated cases.[42]

vascular malformations

Vascular malformations are usually present at birth. They are permanent or progress in the form of ectasias but do not proliferate. Vascular malformations may be subdivided into the following groups: venous, lymphatic, combined arteriovenous, and capillary (such as port-wine stain).[43] Dysmorphic syndromes such as Sturge-Weber and Klippel-Trénaunay-Weber syndromes are more commonly associated with vascular malformations than with hemangiomas.

Diagnosis and Classification

Salmon patch The salmon patch, one of the most common vascular birthmarks, is a dull-pink macule that appears on the nape of the neck, central forehead, or eyelids. Although the salmon patch is sometimes classified as a nevus flammeus, it is distinguished from the latter by its tendency to fade in early life. The salmon patch is caused by the persistence of fetal capillary ectasia in the dermis.[33]

Port-wine stain Port-wine stain, also called nevus flammeus, appears at birth as a reddish or violaceous macular discoloration, usually in a unilateral, segmental distribution [*see Figure 23*]. Mature dilated capillaries are present in the dermis. After puberty, nevus flammeus lesions may become thickened and nodular or papular. There is little tendency toward involution. Nevus flammeus lesions may be associated with abnormalities of the larger vessels and with neurologic manifestations.

Sturge-Weber syndrome A facial port-wine stain that involves the skin innervated by the first branch of the trigeminal nerve is a feature of the Sturge-Weber syndrome (also known as encephalotrigeminal angiomatosis). Other features of the Sturge-

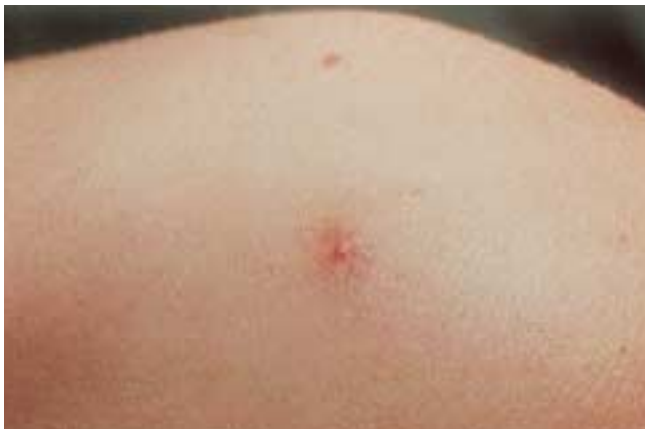


Figure 24 **A spider angioma, which has a central arteriole from which fine vessels radiate, blanches with pressure.**

Weber syndrome include ipsilateral congenital glaucoma and contralateral seizures caused by leptomeningeal angiomatosis. Ophthalmologic and neurologic evaluation may be warranted in patients with the Sturge-Weber syndrome.

Klippel-Trénaunay-Weber syndrome The triad of findings seen in Klippel-Trénaunay-Weber syndrome includes a port-wine stain, usually in a patchy distribution on the involved extremity; varicose veins; and soft tissue or bony hypertrophy. The most common site of involvement is the lower leg; the next most common sites of involvement are the arms and trunk.[37]

Venous malformation Formerly referred to as cavernous hemangiomas, vascular malformation consists of a collection of abnormal veins and venous pouches that commonly occur around the head and neck but can occur anywhere on the body. They are frequently multiple or have satellite lesions. Superficially, they appear as a subcutaneous swelling with a bluish hue on the skin surface or mucous membrane. Deeper components may be invisible on clinical examination. Lesions enlarge for several months, become stationary for an indefinite period, and spontaneously resolve.

Treatment

Because vascular malformations do not proliferate, treatment may be cosmetic and can be postponed to later in life. However, a multidisciplinary approach is needed to treat potential complications of vascular malformations associated with dysmorphic syndromes. Salmon patch tends to fade in early life and usually requires no treatment.

Treatment of port-wine stains by excision, tattooing, ionizing radiation, cryosurgery, or dermabrasion is largely unsatisfactory. Use of the argon laser has resulted in lightening of vascular lesions; however, there is wide variability in response. The effectiveness of this treatment results from the selective absorption of the monochromatic 585 nm laser light by red hemoglobin pigment, which produces thermal energy with resultant photocoagulation of tissue.[44] Thinner lesions are more responsive than thicker lesions that have undergone progressive vascular ectasia. In a study of 100 patients of different age groups who had port-wine stains of the head and neck and who were treated with a flashlamp pulsed dye laser, treatment was no more effective when given in early childhood than when given at a later date.[45]

Acquired Vascular Disorders

diagnosis and classification

Spider Angioma

Spider angioma, also called spider nevus or arterial spider, appears as a central red punctum from which fine vessels radiate; the appearance of the lesion is suggestive of a red spider [*see Figure 24*]. The central arteriole may be pulsatile. These telangiectasias (dilated capillaries) are commonly seen on the face, neck, trunk, and upper extremities and occur most commonly in middle-aged or elderly persons. They may arise spontaneously or in association with pregnancy or hepatic dysfunction. Spider angiomas may be treated with laser therapy for cosmetic reasons.

Unilateral Telangiectasia

Acquired unilateral telangiectatic nevi are uncommon, but those that have been reported resulted from mechanical or physical trauma, including sun damage.[46]

Cherry Angioma

Cherry angioma, also called senile angioma, appears as multiple bright-red, soft, dome-shaped papules on the trunk of middle-aged or older persons. Trauma produces slight bleeding. Electrodesiccation may be performed for cosmetic purposes.

Pyogenic Granuloma and Other Vascular Tumors

The pyogenic granuloma is a soft red lesion that is solitary, raised, and nonpulsatile; it often appears after minor skin trauma, such as a puncture wound. Other predisposing factors include hormonal effects, infection, viral oncogenes, microscopic arteriovenous anastomoses, and growth factors.[47] Epulis gravidarum is a variant of a pyogenic granuloma. The lesion was formerly believed to be caused by a pyogenic infection of a small wound; histologically, however, an early lesion resembles a capillary hemangioma. The thin, sometimes verrucous epidermis is friable and apt to become eroded or ulcerated. Lesions rapidly reach a size of 1 to 2 cm and then remain static. Common sites of involvement are the fingers, feet, and face [*see Figure 25*]. Biopsy is performed to rule out malignant tumors, such as Kaposi sarcoma and amelanotic melanoma.

Other benign tumors with a vascular component include angiofibroma, angioleiomyoma, and angiolipoma. Some of these can be painful. Differential diagnosis of painful skin tumors includes glomus tumor, angiolipoma, angioleiomyoma, neuromas, and eccrine spiradenoma.[34] Lesions are usually easily removed by electrodesiccation and curettage. If they recur or if satellite lesions appear after such treatment, excisional biopsy is recommended.

Kimura Disease

Kimura disease and angiolymphoid hyperplasia with eosinophilia are rare tumors of unknown cause that occur mainly on the head and neck in young adults and may resemble pyogenic granuloma.[48] Kimura disease, which was first reported in Korea, is most common in Asians. It appears as a granulomatous proliferation of lymphoid tissue that may be accompanied by peripheral eosinophilia and contiguous lymphadenopathy. Lesions may occasionally be seen on the trunk, extremities, and genitalia in addition to the head and neck. Angiolymphoid hy-

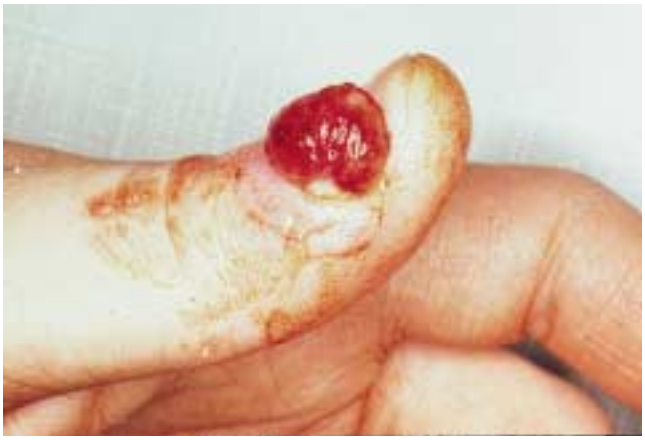


Figure 25 **The pyogenic granuloma, which may show a smooth, verrucous, eroded, or friable surface, may be confused with a malignant tumor.**

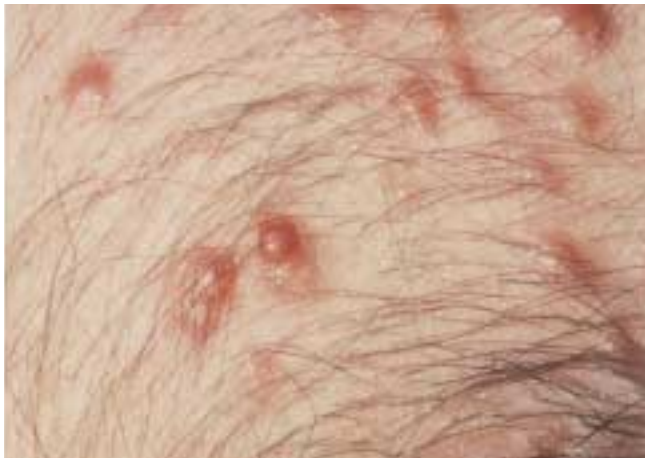


Figure 26 **Leiomyomas are sometimes painful papules that arise from smooth muscle of blood vessels or the arrector pili.**

perplasia with eosinophilia, which may or may not represent a different disease, appears as localized single or multiple nodules. Infectious, allergic, hormonal, and traumatic mechanisms have been postulated. Immunodermatopathologic studies suggest an unusual distribution of adhesion molecules, IgE, and CD23 in these angioproliferating tumors.[49]

Lipoma

The lipoma, which is a soft, rounded to lobulated subcutaneous tumor of mature fat cells, is commonly seen on the trunk, neck, or forearms. Lesions are rubbery in consistency and freely movable under the overlying skin, which appears normal. There may be a single lesion or multiple lesions, and they are usually asymptomatic unless they impinge on a nerve. Lipomas are of variable size and grow slowly. Histologically, the tumors are usually encapsulated and show fat cells that are indistinguishable from normal adipose tissue. Admixture of other tissue components may result in fibrolipomas (fibrous tissue), angiolipomas (blood vessels), and myolipomas (smooth muscle). Excision may be performed for cosmetic reasons. If a lesion grows rapidly, biopsy should be performed, though lipomas rarely become malignant.

Leiomyoma

The leiomyoma is an uncommon tumor of smooth muscle that appears as a single brownish-red papule or as multiple papules or small nodules, which are sometimes painful [*see Figure 26*]. Leiomyomas may arise from the arrector pili (the smooth muscle attached to the hair follicle sheath) or from the smooth muscle surrounding cutaneous blood vessels (angioleiomyoma). Painful lesions can be excised.

Lymphangioma Circumscriptum

Lymphangioma circumscriptum is characterized by groups of persistent localized or diffuse translucent vesicles. Indications for treatment include severe cosmetic problems, persistent leakage of lymphatic fluid or blood, and recurrent infection. The vesicles frequently recur after surgery, radiotherapy, electrocautery, or cryosurgery because of the persistence of deep lymphatic cisterns. Carbon dioxide laser in a vaporization mode has been used to ablate superficial cutaneous lesions in patients with lymphangioma circumscriptum.[50] The major advantage of this technique is that it may reduce the frequency of recurrences because it seals the communicating channels to the deeper cisterns by vaporizing the superficial lymphatics.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Lever's Histopathology of the Skin. Elder D, Elenitsas R, Jaworsky C, et al, Eds. Lippincott-Raven Publishers, Philadelphia, 1997

2. Dermatology in General Medicine. Fitzpatrick TB, Eisen AZ, Wolff K, et al, Eds. McGraw-Hill Book Co, New York, 1993

3. Flugman SL, McClain SA, Clark RF: Transient eruptive seborrheic keratoses associated with erythrodermic psoriasis and erythrodermic drug eruption: report of two cases. J Am Acad Dermatol 45:S212, 2001

4. Paller AS, Syder AJ, Chan YM, et al: Genetic and clinical mosaicism in a type of epidermal nevus. N Engl J Med 331:1408, 1994

5. Seo YJ, Piao, YJ, Suhr KB, et al: A case of nevus comedonicus syndrome associated with neurologic and skeletal abnormalities. Int J Dermatol 40:648, 2001

6. Brownstein MH: Basaloid follicular hamartoma: solitary and multiple types. J Am Acad Dermatol 27:237, 1992

7. Grimalt R, Ferrando J, Mascaro JM: Premature familial sebaceous hyperplasia: successful response to oral isotretinoin in three patients. J Am Acad Dermatol 37:996, 1997

8. Sidhu SK, Wakelin SH, Wilkinson JD: Multiple familial trichoepitheliomas. Cutis 63:239, 1999

9. Esche C, Kruse R, Lamberti C, et al: Muir-Torre syndrome: clinical features and molecular genetic analysis. Br J Dermatol 136:913, 1997

10. Parks ET, Caldemeyer KS, Mirowski GW: Radiologic images in dermatology: Gardner syndrome. J Am Acad Dermatol 45:940, 2001

11. Liu V, Kwan T, Page EH: Parotid oncocytoma in the Birt-Hogg-Dubé syndrome. J Am Acad Dermatol 43:1120, 2000

12. Lindor NM, Hand J, Burch PA, et al: Birt-Hogg-Dube syndrome: an autosomal dominant disorder with predisposition to cancers of the kidney, fibrofolliculomas, and focal cutaneous mucinosis. Int J Dermatol 40:653, 2001

13. Grob JJ, Gouvernet J, Aymar D, et al: Count of benign melanocytic nevi as a major indicator of risk for nonfamilial nodular and superficial spreading melanoma. Cancer 66:387, 1990

14. Tucker MA, Halpern A, Holly EA, et al: Clinically recognized dysplastic nevi: a central risk factor for cutaneous melanoma. JAMA 277:1439, 1997

15. Augustsson A, Stierner U, Rosdahl I, et al: Melanocytic naevi in sun-exposed and protected skin in melanoma patients and controls. Acta Derm Venereol (Stockh) 71:512, 1991

16. Kelly JW, Rivers JK, MacLennan R, et al: Sunlight: a major factor associated with the development of melanocytic nevi in Australian schoolchildren. J Am Acad Dermatol 30:40, 1994

17. Marghoob AA, Orlow SJ, Kopf AW: Syndromes associated with melanocyti nevi. J Am Acad Dermatol 29:373, 1993

18. Gupta G, Williams REA, Mackie RM: The labial melanotic macule: a review of 79 cases. Br J Dermatol 136:772, 1997

19. Cohen JB, Janniger CK, Schwartz RA: Café-au-lait spots. Cutis 66:22, 2000

20. Thami GP, Kaur S, Kanwar A: Association of juvenile xanthogranuloma with *café-au-lait* macules. Int J Dermatol 40:281, 2001

21. Mooney MA, Barr RJ, Buxton MG: Halo nevus or halo phenomenon: a study of 142 cases. J Cutan Pathol 22:342, 1995

22. Lowe NJ, Wieder JM, Sawcer D, et al: Nevus of Ota: treatment with high energy fluences of the Q-switched ruby laser. J Am Acad Dermatol 29:997, 1993

23. Glinick SE, Alper JC, Bogaars H, et al: Becker's melanosis: associated abnormalities. J Am Acad Dermatol 9:509, 1983

24. DeDavid M, Orlow SJ, Provost N, et al: Neurocutaneous melanosis: clinical features of large congenital melanocytic nevi in patients with manifest central nervous system melanosis. J Am Acad Dermatol 35:529, 1996

25. Buenger KM, Porter NC, Dozier SE, et al: Localized multiple neurilemmomas of the lower extremity. Cutis 51:36, 1993

26. Riccardi VM: Von Recklinghausen neurofibromatosis. N Engl J Med 305:1617, 1981

27. Martuza RL, Eldridge R: Neurofibromatosis 2 (bilateral acoustic neurofibromatosis). N Engl J Med 318:684, 1988

28. MacCollin M, Mohney T, Trofatter J, et al: DNA diagnosis of neurofibromatosis 2: altered coding sequence of the merlin tumor suppressor in an extended pedigree. JAMA 270:2316, 1993

29. Riccardi VM: Mast-cell stabilization to decrease neurofibroma growth: preliminary experiences with ketotifen. Arch Dermatol 123:1011, 1987

30. Sahl WJ, Clever H: Cutaneous scars: part I. Int J Dermatol 33:681, 1994

31. Zouboulis CC, Blume U, Buttner P, et al: Outcomes of cryosurgery in keloids and hypertrophic scars: a prospective consecutive trial of case series. Arch Dermatol 129:1146, 1993

32. Gold MH: Topical silicone gel sheeting in the treatment of hypertrophic scars and keloids. J Dermatol Surg Oncol 19:912, 1993

33. Requena L, Sangueza OP: Cutaneous vascular proliferations. Part II: hyperplasias and benign neoplasms. J Am Acad Dermatol 37:887, 1997

34. Requena L, Sanqueza MD: Cutaneous vascular proliferations. Part III: malignant neoplasms, other cutaneous neoplasms with significant vascular component, and disorders erroneously considered as vascular neoplasms. J Am Acad Dermatol 38:143, 1998

35. Blei F, Walter J, Orlow SJ, et al: Familial segregation of hemangiomas and vascular malformations as an autosomal dominant trait. Arch Dermatol 134:718, 1998

36. Smoller BR, Rosen R: Port-wine stains: a disease of altered neural modulation of blood vessels. Arch Dermatol 122:177, 1986

37. Meine JG, Schwartz RA, Janniger CK: Klippel-Trénaunay-Weber syndrome. Cutis 60:127, 1997

38. North PE, Waner M, Mizeracki A, et al: A unique microvascular phenotype shared by juvenile hemangiomas and human placenta. Arch Dermatol 137:559, 2001

39. North PE, Waner M, James CA, et al: Congenital nonprogressive hemangioma: a distinct clinicopathologic entity unlike infantile hemangioma. Arch Dermatol 137:1607, 2001

40. Mulliken JB, Young AE: Vascular Birthmarks: Hemangiomas and Malformations. WB Saunders Co, Philadelphia, 1988, p 77

41. Barlow RJ, Walker NJ, Markey AC: Treatment of proliferative haemangiomas with the 585 nm pulsed dye laser. Br J Dermatol 134:700, 1996

42. Donnelly LF, Adams DM, Bisset GS 3rd: Vascular malformations and hemangiomas: a practical approach in a multidisciplinary clinic. AJR Am J Roentgenol 174:597, 2000

43. Requena L, Sangueza OP: Cutaneous vascular anomalies. Part I. Hamartomas, malformations, and dilatation of preexisting vessels. J Am Acad Dermatol 37:523, 1997

44. Fitzpatrick RE, Lowe NJ, Goldman MP, et al: Flashlamp-pumped pulsed dye laser treatment of port-wine stains. J Dermatol Surg Oncol 20:743, 1994

45. van der Horst CM, Koster PL, de Borgie CM, et al: Effect of the timing of treatment of port-wine stains with the flash-lamp-pumped pulsed-dye laser. N Engl J Med 338:1028, 1998

46. Pasyk KA: Acquired lateral telangiectatic nevus: port-wine stain or nevus flammeus. Cutis 51:281, 1993

47. Mooney MA, Janniger CK: Pyogenic granuloma. Pediatr Dermatol 55:133, 1995

48. Chun SI, Ji HG: Kimura's disease and angiolymphoid hyperplasia with eosinophilia: clinical and histopathologic differences. J Am Acad Dermatol 27:954, 1992

49. von den Driesch P, Gruschwitz M, Schell H, et al: Distribution of adhesion molecules, IgE, and CD23 in a case of angiolymphoid hyperplasia with eosinophilia. J Am Acad Dermatol 26:799, 1992

50. Bailin PL, Kantor GR, Wheeland RG: Carbon dioxide laser vaporization of lymphangioma circumscriptum. J Am Acad Dermatol 14:257, 1986

XII ACNE VULGARIS AND RELATED DISORDERS

Mark Lebwohl, m.d.

Acne and its clinical variants are among the most common causes of patient visits to the physician for cutaneous disorders. Severe forms of these disorders can be disfiguring and debilitating; and because the face is the primary site of involvement, patients will often seek therapy for even mild forms. Therapeutic approaches will therefore be stressed in this chapter.

Epidemiology and Etiology

Acne vulgaris is the most common dermatologic problem of adolescent years; it usually begins in puberty. Age of onset and severity of disease are affected by sex, genetics, and external factors such as cosmetics and medications. Acne is usually more severe in males than in females and often begins earlier (i.e., in early adolescence) in males. Acne often subsides after the teenage years, but the disease can remain a problem for adults in the third and fourth decades and beyond. A significant portion of women experience premenstrual flares of acne; this phenomenon may be more common in older women.[1]

Genetic factors clearly play a role in severe acne. A family history of severe acne can often be elicited during the workup of affected patients. Various external factors, such as occlusive cosmetics, can contribute to acne, and certain medications (e.g., corticosteroids, adrenocorticotropic hormone [ACTH], phenytoin sodium, isoniazid, lithium, progestins, potassium iodide, bromides, and actinomycin D) can cause acnelike lesions [*see 2:VI Cutaneous Adverse Drug Reactions*].

Pathogenesis

Multiple factors contribute to the development of acne in susceptible persons. Among the most significant are alterations in keratinization, accumulation of sebum, and inflammation. Androgenic influences may contribute to some of these factors.

Modified keratinization of the follicular infundibulum leads to proliferation and increased cohesiveness of keratinocytes, which causes plugs to form. These plugs block follicular outlets, allowing cellular debris in sebum to form comedones (the noninflammatory lesions of acne that are the precursor lesions of inflammatory acne).

The composition of sebum does not appear to be altered in patients with acne; however, sebaceous glands are often larger and sebum production is often greater in persons affected with acne than in unaffected persons.[2] Sebum is comedogenic and inflammatory, which may account for its role in acne.[3] Inflammation in acne has also been attributed to the anaerobic diphtheroid *Propionibacterium acnes.* The presence of *P. acnes* correlates with the occurrence of acne in adolescents.[4] The microbe's role in inflammation has been attributed to lipases, proteases, and hyaluronidases, as well as to chemotactic factors. For example, *P. acnes* may activate Toll-like receptor 2 and thereby trigger inflammatory cytokine responses.[5]

Androgens play a role in the development of acne, as evidenced by increased levels of dehydroepiandrosterone sulfate (DHEAS) in girls with acne[6] and an association of acne with endocrinopathies characterized by increased levels of circulating androgens. For example, the occurrence of acne is increased in patients with congenital adrenal hyperplasia, polycystic ovaries, and some ovarian and adrenal tumors. Androgens act to increase sebum production and enlarge sebaceous glands; they may also contribute to the follicular hyperkeratinization that leads to acne. However, serum androgen levels are usually within the normal range in patients with acne. Some researchers have postulated that local production of androgens in the skin can lead to acne. Skin biopsies from patients with acne show increases in 5α-reductase activity.[7] This increased androgenic activity may result in the conversion of testosterone to dihydrotestosterone in the skin, leading to the development of acne.

Diagnosis

Clinical Features

The characteristic skin lesions of acne include open and closed comedones, erythematous papules, pustules, nodules, cysts, and scars. The most commonly affected site is the face, but in more severely affected individuals, the back and chest can be involved as well.

Comedonal Acne

Comedones consist of keratinized cells and sebum. Comedonal acne consists of a predominance of open and closed comedones. Open comedones (blackheads) are black papules measuring 0.1 to 2 mm that are easily extruded with gentle pressure. The material that is removed is greasy and has a gray-white color. Contrary to popular belief, the dark color of open comedones is caused by melanin, not by dirt or oxidized fatty acids. Closed comedones (whiteheads) consist of white papules measuring 0.1 to 2 mm; unless extracted, they persist somewhat longer than open comedones, often for weeks to months.

Inflammatory Acne

Erythematous papules, pustules, nodules, and cysts are the predominant lesions in inflammatory acne [*see Figure 1*]. Erythematous papules range in size from 3 to 10 mm and can develop into pustules or resolve into an erythematous macule that fades. Postinflammatory hyperpigmentation can occur. Pustules are superficial and usually dry in a few days. Nodules, which are 1 cm or larger, are erythematous and tender. They can be firm at onset but often become fluctuant. In severely affected individuals, these lesions form fluctuant sinuses that open to the surface through multiple tracts. Postinflammatory pigmentary changes and scarring commonly occur.

Clinical Variants of Acne

Acne conglobata Acne conglobata is a severe, scarring form of acne in which large cysts and abscesses become confluent to form draining sinus tracts. Scarring is often severe. Topical acne therapy and oral antibiotics are frequently ineffective; patients may require treatment with oral isotretinoin [*see* Treatment, *below*]. Intralesional injection of corticosteroids and drainage of abscesses are temporarily helpful.

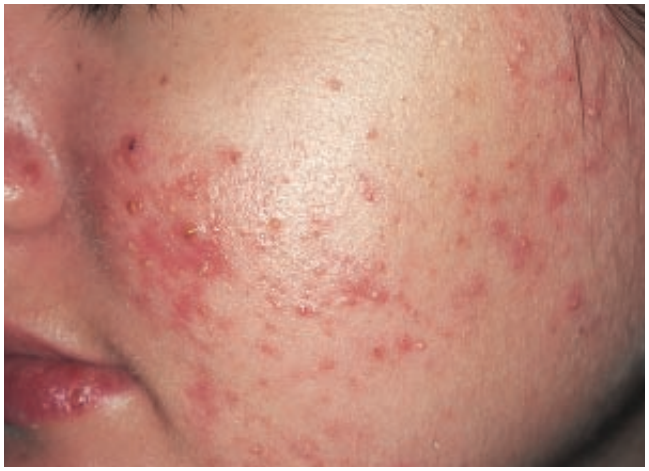


Figure 1 **Inflammatory acne is characterized by erythematous papules and pustules.**

Acne cosmetica A persistent, low-grade form of acne can result from the use of greasy, occlusive cosmetics, moisturizers, and sunscreens. Women are most commonly affected.

Acne excoriée Picking of minor acne lesions can cause large ulcers and erosions that heal with scarring. Young women are most typically affected.

Acne mechanica An acneiform eruption can result from repeated trauma associated with the wearing of sports helmets, shoulder pads, and bras and from the chin rests of violins and violas (so-called fiddler's neck).

Pomade acne A form of acne results from the use of thick oils in the hair. Comedones, papules, and pustules are usually found close to the hairline. Black men and women are most commonly affected.

Acne in neonates and children Neonatal acne has been attributed to maternal androgens, as well as androgens secreted by the neonatal adrenal gland. Erythematous papules and pustules may last for 2 to 3 months after birth but usually resolve spontaneously.

Infantile acne develops between 3 and 6 months after birth. This condition is characterized by inflamed papules and pustules; it signals early secretion of androgens by the gonads, particularly in boys. This condition may last until age 5. It has been suggested that affected infants may be predisposed to severe acne later in life.

LABORATORY TESTS

The clinical features of acne are so commonly recognized that laboratory investigation is usually not necessary. Laboratory tests should be considered, however, for female patients who have other signs of hyperandrogenism, such as hirsutism or irregular menses. Serum for determining DHEAS and free testosterone levels and for determining the ratio of luteinizing hormone to follicle-stimulating hormone (LH:FSH) should be obtained 2 weeks before the onset of menses [*see Table 1*]. Tests should also be undertaken in patients whose conditions do not respond to adequate doses of isotretinoin, the most potent treatment available for acne [*see* Treatment, *below*].

Differential Diagnosis

Clinical features of acne are sufficiently distinctive that diagnosis is usually obvious. Nevertheless, a number of disorders can be mistaken for acne.

Folliculitis The perifollicular pustules of folliculitis can be distinguished from the lesions of acne by their distribution. Folliculitis can affect the trunk and extremities and is not limited to the usual sites of acne (i.e., the face, back, and chest). Malassezia folliculitis is characterized by erythematous acneiform papules that do not respond to typical acne therapies. Gram stain of pus from the lesions reveals gram-positive budding yeast [*see 2:VII Fungal, Bacterial, and Viral Infections of the Skin*].

Gram-negative folliculitis In patients on long-term antibiotics, superficial pustules or nodules can develop at the anterior nares and spread outward on the face. This condition responds promptly to oral ampicillin; however, isotretinoin has become the treatment of choice.[8]

Milia Milia are white pinpoint cysts that resemble closed comedones. They frequently occur around the eyes but can develop anywhere on the face. If untreated, they last for months or years. Milia can be opened with a small surgical blade and their contents easily drained.

Perioral dermatitis Long-term use of topical corticosteroids on the face can result in acneiform, erythematous, inflamed papules on the chin and cheeks. Despite the name, the area immediately around the mouth is typically spared in perioral dermatitis. A similar eruption can occur in patients who have not used corticosteroids.

Chloracne Cysts and closed comedones that resemble acne lesions can be caused by exposure to halogenated hydrocarbons.

Hidradenitis suppurativa Hidradenitis suppurativa is a chronic condition in which inflamed cysts in the axillae and groin form fluctuant sinuses with draining tracts.

Favre-Racouchot disease Numerous open and closed comedones can appear around the eyes of elderly patients, especially men who have worked outdoors for much of their lives. This condition has been attributed to a lifetime of sun exposure.

Table 1 Laboratory Evaluation for Women with Acne and Signs of Hyperandrogenism

Finding	*Suspected Condition*
DHEAS	
4,000–8,000 ng/ml	Congenital adrenal hyperplasia
> 8,000 ng/ml	Adrenal tumor
LH:FSH ratio > 2.0	Polycystic ovary disease
Testosterone (unbound)	
20–40 yr, > 107.5 pmol/L	Polycystic ovary disease; ovarian tumor
41–60 yr, > 86.7 pmol/L	Polycystic ovary disease; ovarian tumor
61–80 yr, > 69.3 pmol/L	Polycystic ovary disease; ovarian tumor

DHEAS—dehydroepiandrosterone sulfate FSH—follicle-stimulating hormone
LH—luteinizing hormone

Rosacea Rosacea is a common condition that usually begins after 30 years of age. It is so similar to acne in some individuals that it has been called acne rosacea. Skin lesions consist of erythematous papules, pustules, and telangiectasia [*see Figure 2*]. Facial flushing is a common feature. In patients with a predominance of inflamed papules and pustules, differentiation from acne can be difficult. Presence of telangiectasia and the occurrence of flushing help distinguish rosacea from acne, as does the absence of comedones.

Common triggers of rosacea include alcohol, exercise, extremes of temperature, and hot or spicy foods. With long-standing disease, hypertrophy of sebaceous glands, swelling, erythema, and scarring of the nose lead to rhinophyma. Ocular involvement is common in rosacea and can include blepharitis and conjunctival hyperemia or, less commonly, iritis, episcleritis, superficial punctuate keratopathy, and corneal neovascularization.[9]

Helicobacter pylori may play a role in the pathogenesis of rosacea.[10] Further work must be done, however, to confirm the contribution of *H. pylori* to this antibiotic-responsive condition. An immunologic reaction to the mite *Demodex folliculorum* has been suggested, but not proved, as a contributing factor.

Treatment

Treatment of acne depends on the type and severity of lesions and on the patient's response to treatment. Comedonal acne is usually best managed with topical retinoids and acne surgery; inflammatory acne is treated with a range of topical therapies and may require oral therapy in moderate to severe cases. Because nodules and cysts are more likely than comedones to cause scarring, they are treated more quickly with oral antibiotics and, if necessary, isotretinoin (see below). Intralesional corticosteroids administered by dermatologists can prevent scarring from cysts. Incision and drainage of infected cysts may be necessary but can contribute to scarring. Unroofing of sinus tracts and other surgical procedures are best performed by physicians with expertise in dermatologic surgery [*see Table 2*]. Scars can be treated with dermabrasion or laser abrasion. The appearance of depressed scars can be improved by chemical peels and other resurfacing procedures, as well as by the injection of filler substances such as injectable collagen.[11]

Numerous over-the-counter cleansing agents are available to help patients remove seborrhea and oily debris from the skin, resulting in subjective improvements. Overmanipulation of lesions by picking, squeezing, or excessive washing can lead to exacerbation of lesions and even scarring.

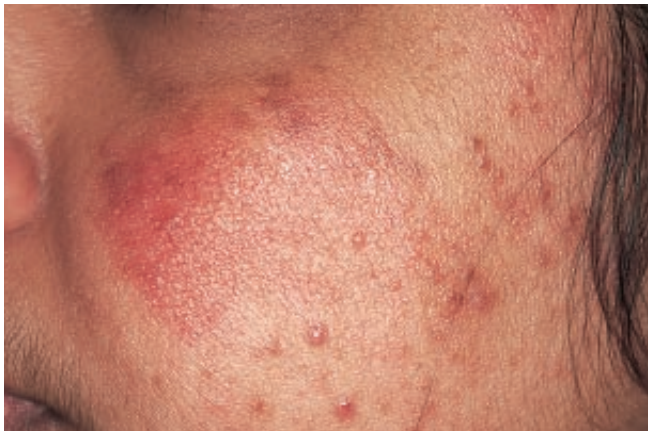


Figure 2 **Erythematous papules, pustules, telangiectasia, and flushing are features of rosacea.**

Table 2 Surgical Treatments for Acne Lesions and Acne Scars

Lesions	Extraction of comedones Drainage of pustules and cysts Intralesional injection of corticosteroids in cysts Excision and unroofing of sinus tracts and cysts
Scars	Dermabrasion Laser abrasion Acid peels Injection of filling materials (e.g., collagen) Excision Punch autografts

Topical preparations, including sunscreens, soaps, and cosmetics, should be oil-free and noncomedogenic. Many over-the-counter oil-free, noncomedogenic moisturizers are available for persons who have dry skin and acne.

There is no role for dietary change in the management of acne. Previous beliefs that chocolate or oily foods cause acne have been disproved.

TOPICAL THERAPY

Comedonal Acne

Topical retinoids are among the most effective therapies for comedonal acne; these preparations unplug follicles and allow penetration of topical antibiotics and benzoyl peroxide. Retinoids can be used in combination with antibacterial agents and are also effective in the management of inflammatory acne.[12] They are often irritating when first applied; patients can reduce the irritation by reducing the frequency of application. Significant improvement is evident within 6 weeks and can continue for 3 to 4 months, at which time the frequency of application can be reduced, depending on the patient's response.

Newer formulations of retinoids that are purportedly less irritating include a tretinoin microsponge vehicle and adapalene, but few comparative studies examining irritation have been performed.[13,14] Tazarotene, a topical retinoid used for acne and psoriasis, can be used effectively in a short-contact method, in which it is applied for seconds to minutes.[15]

Inflammatory Acne

Topical antibiotics are not as effective as retinoids or benzoyl peroxide for inflammatory acne, but they are less irritating and better tolerated. The resistance of *P. acnes* to antibiotics has been well documented; such resistance threatens the efficacy of this form of acne therapy in the future.[16,17] It is therefore useful to prescribe antibiotics in combination with benzoyl peroxide, which does not induce resistance. A combined formulation of clindamycin 1% and benzoyl peroxide 5% has been found to produce faster and greater reductions in *P. acnes* than formulations containing clindamycin alone.[18] Moreover, the combination of benzoyl peroxide and clindamycin resulted in greater improvement in acne than either of its individual components alone.[19]

A commonly used regimen includes the combined antibiotic–benzoyl peroxide gel in the morning and topical retinoid in the evening. Azelaic acid, an anticomedonal and antibacterial

Table 3 Topical Therapies for Acne

Medication	*Formulation*	*Frequency of Application*	*Primary Mechanism of Action*	*Adverse Effects*
Azelaic acid	20% cream	b.i.d.	Anticomedonal, antibacterial	Stinging, irritation
Benzoyl peroxide	2.5%, 5%, 10% creams, gels, lotions, washes	b.i.d.	Antibacterial	Dryness, irritation, allergic contact dermatitis
Antibiotics				
Clindamycin	1% solutions, lotions, gels	b.i.d.	Antibacterial	Antibiotic resistance
Erythromycin	2% solutions, creams, gels, pledgets, wipes	b.i.d.	Antibacterial	Antibiotic resistance
Erythromycin–benzoyl peroxide	3% erythromycin–5% benzoyl peroxide gel	b.i.d.	Antibacterial	Dryness, irritation, allergic contact dermatitis; deteriorates if not refrigerated
Sodium sulfacetamide–sulfur	10% sodium sulfacetamide, 5% sulfur lotions	b.i.d.	Antibacterial	Dryness, irritation, allergic contact dermatitis
Retinoids				
Adapalene	0.1% gels	q.d.	Comedolytic	Dryness, irritation, photosensitivity
Tazarotene	0.05%, 0.1% gels	q.d.	Comedolytic	Dryness, irritation, photosensitivity
Tretinoin	0.025%, 0.05%, 0.1% creams; 0.01%, 0.025% gels; 0.05% solutions	q.d.	Comedolytic	Dryness, irritation, photosensitivity
Sulfur and resorcinol	2% resorcinol, 8% sulfur lotions, creams	q.d., b.i.d.	Comedolytic	Dryness, peeling, allergic contact dermatitis
Salicylic acid	0.5%–2% gels, pads, soaps	q.d., b.i.d.	Comedolytic	Dryness, irritation

agent, offers yet another choice for the topical treatment of acne. It, too, can be used in combination with topical retinoids, benzoyl peroxide, or topical antibiotics.[20] Salicylic acid, an over-the-counter comedolytic agent, plays a minor role in the treatment of acne. Skin-colored sulfur-resorcinol lotions are available; these very effective drying and peeling agents can be useful for treating individual lesions [*see Table 3*].

SYSTEMIC THERAPY

Systemic agents are warranted for patients with nodulocystic acne or inflammatory acne that is not responsive to topical therapy. Oral antibiotics are usually the first line of systemic treatment. Isotretinoin has generally been reserved for patients whose acne is refractory to antibiotics. Isotretinoin may be used as initial therapy in patients with particularly severe acne to prevent scarring and in patients with a history of antibiotic intolerance.

Antibiotics

Antibiotics have both antibacterial and anti-inflammatory effects that are beneficial in treating acne. The antibiotics most commonly used for acne are doxycycline, erythromycin, minocycline, tetracycline, and trimethoprim-sulfamethoxazole [*see Table 4*]. Because antibiotic resistance is a major problem with many of the older antibiotics, minocycline has been prescribed for many acne patients even though it is considerably more expensive. Strains of *P. acnes* that are resistant to minocycline have begun to emerge, however, and this may limit the usefulness of this drug in the future.[21] The duration of treatment with oral antibiotics depends on patient response. Azithromycin given at a dosage of 500 mg/day for 4 days, repeated at 10-day intervals for four cycles, is as effective as minocycline given at a dosage of 100 mg/day for 6 weeks.[22] Further refinements of regimens with these newer antibiotics will undoubtedly be performed before they achieve more widespread usage.

A lupuslike syndrome has been reported in patients taking oral minocycline. Synovitis, the presence of antinuclear antibodies, and elevations in hepatic transaminase levels were reported, but renal disease and central nervous system disease do not occur.[23] Upon discontinuance of minocycline, symptoms resolve, but upon retreatment, the syndrome recurs.

Controversy about the long-term use of antibiotics for the treatment of acne was raised by a 2004 study that suggested a correlation between antibiotic use and breast cancer risk. The study found that an increase in the cumulative number of days of antibiotic use—including use of tetracyclines and macrolides, which are prescribed for acne—was associated with greater breast cancer risk.[24] Although the results of this study have been questioned because of the way the study was performed and other shortcomings of the study, the possibility of increased risk remains a concern.

Isotretinoin

Oral isotretinoin is the most effective agent available for the treatment of acne. It results in long-lasting remissions or cures in the majority of patients treated. Because of its serious potential adverse effects, however, isotretinoin is not generally used as first-line therapy except for unusual cases.

Most of the side effects of isotretinoin are dose related and affect a majority of patients treated. For example, cheilitis uniformly occurs in patients treated with significant doses. Myalgias, dryness of mucous membranes, dry eczematous skin changes, and hyperlipidemia frequently occur. Total serum cholesterol levels can rise in patients taking isotretinoin, and triglyceride levels can rise sufficiently to cause pancreatitis.

Teratogenicity occurs with the administration of even a single dose of isotretinoin to pregnant women. Birth control counseling is an essential part of the management of women for whom isotretinoin is prescribed. The use of two forms of contraception is advised. Despite major educational efforts, pregnancies in women receiving isotretinoin continue to occur, resulting in severe birth defects.[25] With the introduction of generic isotretinoin, concern over teratogenicity increased. In response, the manufacturers of isotre-

Table 4 Commonly Prescribed Systemic Therapies for Acne

Medication	*Dosage*	*Advantages*	*Adverse Effects*
Antibiotics			
Doxycycline	50–100 mg p.o., b.i.d.	Inexpensive	Photosensitivity, GI symptoms, candidiasis
Erythromycin	250–500 mg p.o., b.i.d.	Alternative to tetracyclines	GI symptoms, candidiasis
Minocycline	50 mg p.o., q.d.–100 mg p.o., b.i.d.	Highly effective; antibiotic resistance rare at 200 mg/day	GI symptoms, candidiasis, vertigo, lupuslike syndrome (rare), autoimmune hepatitis (rare)
Tetracycline	250 mg p.o., q.d.–500 mg p.o., q.i.d. (b.i.d. dosing preferred)	Inexpensive	Photosensitivity, GI symptoms, candidiasis
Trimethoprim-sulfamethoxazole	160 mg trimethoprim–800 mg sulfamethoxazole b.i.d.	Alternative to tetracyclines and erythromycin	Bone marrow suppression, drug eruption
Other Agents			
Isotretinoin	0.5–2.0 mg/kg/day, in two divided doses	Most effective treatment; long-lasting remissions	Teratogenicity, hyperlipidemia, cheilitis, alopecia, pyogenic granulomas, dry eyes, epistaxis, rare pseudotumor cerebri (especially with concomitant antibiotics)
Norgestimate–ethinyl estradiol	0.18 mg norgestimate, 0.035 mg ethinyl estradiol p.o., q.d., for 21 days; repeat every 4 wk	Alternative to antibiotics and isotretinoin; less androgenic activity than progestins in other contraceptives	Thromboembolic disorders; ?antibiotic interaction; ?increased breast carcinoma; gallbladder disease; reduced glucose tolerance; headache; fluid retention; hypertension; breakthrough bleeding; breast swelling and tenderness
Drospirenone–ethinyl estradiol	3 mg drospirenone and 0.3 mg ethinyl estradiol p.o., q.d., for 21 days, followed by 7 days of inert pills; repeat monthly	Alternative to antibiotics and isotretinoin; less androgenic activity than progestins in other contraceptives	Thromboembolic disorders; ?antibiotic interaction; ?increased breast carcinoma; gallbladder disease; reduced glucose tolerance; headaches; fluid retention; hypertension; breakthrough bleeding; breast swelling and tenderness
Estrophasic contraceptive	1 mg norethindrone acetate and increasing doses of ethinyl estradiol: 20 μg, days 1–5; 30 μg, days 6–12; 35 μg, days 13–21; then 1 wk of inert tablet; repeat cycle every 4 wk	Alternative to antibiotics and isotretinoin; less androgenic activity than progestins in other contraceptives	Thromboembolic disorders; ?antibiotic interaction; ?increased breast carcinoma; gallbladder disease; reduced glucose tolerance; headaches; fluid retention; hypertension; breakthrough bleeding; breast swelling and tenderness

tinoin started a program in which physicians and pharmacists who prescribe and administer isotretinoin must register and agree to require that patients receiving isotretinoin undergo pregnancy testing on a regular basis.[26] Unfortunately, this program failed to eliminate pregnancies in women treated with isotretinoin. Attempts to enforce guidelines on the safe use of isotretinoin[27] have been deemed inadequate, and as a result, more stringent barriers to the prescription of isotretinoin are being instituted.[28]

There have been several instances of suicide and depression occurring in patients receiving oral isotretinoin.[29,30] Teenagers with severe acne may be at increased risk for suicide, regardless of the treatment they are using. A study compared the risk of depression, psychotic symptoms, suicide, and attempted suicide in acne patients receiving isotretinoin with the risk in acne patients being treated with oral antibiotics. The relative risk of depression or psychosis for isotretinoin-treated patients was 1.0, and the relative risk of suicide and attempted suicide was 0.9, suggesting that isotretinoin does not cause depression.[31] A study of pharmacy prescriptions yielded similar results. Prescriptions for antidepressants were quantified in 2,821 patients who filled isotretinoin prescriptions for the first time, and they were again quantified for patients filling isotretinoin prescriptions for a second time. The ratio of antidepressant use with the first prescription of isotretinoin to antidepressant use with the second prescription was not significantly different from 1.0—a finding that does not support an association between the use of isotretinoin and the onset of depression.[32]

Pseudotumor cerebri is a rare side effect of isotretinoin. It occurs more commonly in patients who are concomitantly given oral antibiotics.

Extensive counseling and monitoring—including complete blood counts, chemistry screens, and pregnancy tests when appropriate—should be done before treatment with isotretinoin; such counseling and monitoring should continue at 2-week intervals during the first month of treatment and monthly thereafter. Depending on patient response, treatment with 0.5 to 1.0 mg/kg/day in two divided doses should be continued to a cumulative dose of 120 to 150 mg/kg. Some clinicians have continued low-dose isotretinoin therapy for more than 6 months. Rarely, a second course of therapy is indicated when acne recurs.

Hormone Therapy

Estrogens in the form of oral contraceptives can be beneficial for patients with acne; progestins, however, can exacerbate the condition. The newer progestins—desogestrel, norgestimate, and gestodene—have less androgenic activity and therefore are less likely to exacerbate acne. A combination of ethinyl estradiol and norgestimate has been shown to be beneficial in the treatment of acne.[33] An oral contraceptive containing ethinyl estradiol in graduated doses, along with stable doses of norethindrone acetate, has been shown to have minimal androgenic activity and is also used for the treatment of acne.[34] A combined oral contraceptive containing ethinyl estradiol and drospirenone has also been found to effectively treat acne.[35] These agents are ideal for women who are seeking birth control methods and for women who are not candidates for, or who have not responded to, oral antibiotics or isotretinoin. Oral contraceptives can be particularly helpful to women with polycystic ovary syndrome. It is noteworthy that the beneficial effects of combined oral contraceptives are diminished in patients who are obese.[36]

Some concerns have been raised about the concomitant use of antibiotics and oral contraceptives because some antibiotics may interfere with contraceptive activity. Reviews of large numbers

of patients treated concomitantly with oral contraceptives and antibiotics have not revealed significant increases in pregnancies.[37] Nevertheless, caution is advisable when a patient uses an antibiotic and an oral contraceptive together, especially one of the newer contraceptives that contain low doses of estrogen.

PHOTOTHERAPY

A number of light sources have been tested for the treatment of acne. Photodynamic therapy using topical δ-aminolevulinic acid has demonstrated efficacy for acne. Photodynamic therapy did not reduce *P. acnes* numbers or sebum excretion, so the mechanism by which it works is not entirely known.[38] A blue light administered twice weekly for 4 consecutive weeks has demonstrated efficacy for acne but not for nodulocystic lesions.[39] The 1,064 nm Q-switched neodymium:yttrium-aluminum-garnet (Nd:YAG) laser has proved useful for the treatment of acne scarring.[40]

TREATMENT OF ROSACEA

Avoidance of triggers such as alcohol, hot or spicy foods, and heat are an important part of the therapeutic regimen offered to patients with rosacea. Sunscreens are likewise important. Telangiectasia can be treated with laser therapy. Papules and pustules respond to the same topical and oral antibiotics used for acne, although benzoyl peroxide is less commonly used for rosacea. Flushing is difficult to treat. Azelaic acid may offer some benefit for the erythema associated with rosacea.[41]

Additional Information

Additional information about acne and its related disorders is available from the American Academy of Dermatology (http://www.aad.org) and the National Rosacea Society (http://www.rosacea.org).

The author has received grants from Allergan, Inc., Connetics Corp., Medicis, and Ortho-McNeil Pharmaceutical, Inc.

References

1. Stoll S, Shalita AR, Webster GF, et al: The effect of the menstrual cycle on acne. J Am Acad Dermatol 45:957, 2001

2. Harris HH, Downing DT, Stewart ME, et al: Sustainable rates of sebum secretion in acne patients and matched normal control subjects. J Am Acad Dermatol 8:200, 1983

3. Tucker SB, Rogers RS III, Winkelmann RK, et al: Inflammation in acne vulgaris: leukocyte attraction and cytotoxicity by comedonal material. J Invest Dermatol 74:21, 1980

4. Leyden JJ, McGinley KJ, Mills OH, et al: *Propionibacterium* levels in patients with and without acne vulgaris. J Invest Dermatol 65:382, 1975

5. Kim J, Ochoa MT, Krutzik SR, et al: Activation of toll-like receptor 2 in acne triggers inflammatory cytokine responses. J Immunol 169:1535, 2002

6. Lucky AW, Biro FM, Huster GA, et al: Acne vulgaris in premenarchal girls: an early sign of puberty associated with rising levels of dehydroepiandrosterone. Arch Dermatol 130:308, 1994

7. Sansone G, Reisner RM: Differential rates of conversion of testosterone to dihydrotestosterone in acne and in normal human skin: a possible pathogenic factor in acne. J Invest Dermatol 56:366, 1971

8. Boni R, Nehrhoff B: Treatment of gram-negative folliculitis in patients with acne. Am J Clin Dermatol 4:273, 2003

9. Stone DU, Chodosh J: Ocular rosacea: an update on pathogenesis and therapy. Curr Opin Ophthalmol 15:499, 2004

10. Utas S, Ozbakir O, Turasan A, et al: *Helicobacter pylori* eradication treatment reduces the severity of rosacea. J Am Acad Dermatol 40:433, 1999

11. Hirsch RJ, Lewis AB: Treatment of acne scarring. Semin Cutan Med Surg 20:190, 2001

12. Leyden JJ, Shalita A, Thiboutot D, et al: Topical retinoids in inflammatory acne: a retrospective, investigator-blinded, vehicle-controlled, photographic assessment. Clin Ther 27:216, 2005

13. Leyden J, Grove GL: Randomized facial tolerability studies comparing gel formulations of retinoids used to treat acne vulgaris. Cutis 67(6 suppl):17, 2001

14. Dunlap FE, Baker MD, Plott RT, et al: Adapalene 0.1% gel has low skin irritation potential even when applied immediately after washing. Br J Dermatol 139(suppl):52, 1998

15. Bershad S, Kranjac Singer G, Parente JE, et al: Successful treatment of acne vulgaris using a new method: results of a randomized vehicle-controlled trial of short-contact therapy with 0.1% tazarotene gel. Arch Dermatol 138:481, 2002

16. Leyden JJ: The evolving role of *Propionibacterium acnes* in acne. Semin Cutan Med Surg 20:139, 2001

17. Dreno B, Reynaud A, Moyse D, et al: Erythromycin-resistance of cutaneous bacterial flora in acne. Eur J Dermatol 11:549, 2001

18. Leyden J, Kaidbey K, Levy SF: The combination formulation of clindamycin 1% plus benzoyl peroxide 5% versus 3 different formulations of topical clindamycin alone in the reduction of *Propionibacterium acnes:* an in vivo comparative study. Am J Clin Dermatol 2:263, 2001

19. Leyden JJ, Berger RS, Dunlap FE, et al: Comparison of the efficacy and safety of a combination topical gel formulation of benzoyl peroxide and clindamycin with benzoyl peroxide, clindamycin and vehicle gel in the treatments of acne vulgaris. Am J Clin Dermatol 2:33, 2001

20. Webster G: Combination azelaic acid therapy for acne vulgaris. J Am Acad Dermatol 43(2 pt 3):S47, 2000

21. Ross JI, Snelling AM, Eady EA, et al: Phenotypic and genotypic characterization of antibiotic-resistant *Propionibacterium acnes* isolated from acne patients attending dermatology clinics in Europe, the U.S.A., Japan and Australia. Br J Dermatol 144:339, 2001

22. Gruber F, Grubisic-Greblo H, Kastelan M, et al: Azithromycin compared with minocycline in the treatment of acne comedonica and papulo-pustulosa. J Chemother 10:469, 1998

23. Lawson TM, Amos N, Bulgen D, et al: Minocycline-induced lupus: clinical features and response to rechallenge. Rheumatology (Oxford) 40:329, 2001

24. Velicer CM, Heckbert SR, Lampe JW, et al: Antibiotic use in relation to the risk of breast cancer. JAMA 291:827, 2004

25. Honein MA, Paulozzi LJ, Erickson JD: Continued occurrence of Accutane-exposed pregnancies. Teratology 64:142, 2001

26. Honein MA, Moore CA, Erickson JD: Can we ensure the safe use of known human teratogens? Introduction of generic isotretinoin in the US as an example. Drug Saf 27:1069, 2004

27. Goldsmith LA, Bolognia JL, Callen JP, et al: American Academy of Dermatology Consensus Conference on the safe and optimal use of isotretinoin: summary and recommendations. J Am Acad Dermatol 50:900, 2004 [erratum, J Am Acad Dermatol 51:348, 2004]

28. FDA announces enhancement to isotretinoin risk management program. US Food and Drug Administration, November 23, 2004
http://www.fda.gov/bbs/topics/ANSWERS/2004/ANS01328.html

29. Wysowski DK, Pitts M, Beitz J: An analysis of reports of depression and suicide in patients treated with isotretinoin. J Am Acad Dermatol 45:515, 2001

30. Ng CH, Tam MM, Hook SJ: Acne, isotretinoin treatment and acute depression. World J Biol Psychiatry 2:159, 2001

31. Jick SS, Kremers HM, Vasilakis-Scaramozza C: Isotretinoin use and risk of depression, psychotic symptoms, suicide, and attempted suicide. Arch Dermatol 136:1231, 2000

32. Hersom K, Neary MP, Levaux HP, et al: Isotretinoin and antidepressant pharmacotherapy: a prescription sequence symmetry analysis. J Am Acad Dermatol 49:424, 2003

33. Lucky AW, Henderson TA, Olson WH, et al: Effectiveness of norgestimate and ethinyl estradiol in treating moderate acne vulgaris. J Am Acad Dermatol 37:746, 1997

34. Boyd RA, Zegarac EA, Posvar EL, et al: Minimal androgenic activity of a new oral contraceptive containing norethindrone acetate and graduated doses of ethinyl estradiol. Contraception 63:71, 2001

35. Thorneycroft H, Gollnick H, Schellschmidt I: Superiority of a combined contraceptive containing drospirenone to a triphasic preparation containing norgestimate in acne treatment. Cutis 74:123, 2004

36. Cibula D, Hill M, Fanta M, et al: Does obesity diminish the positive effect of oral contraceptive treatment on hyperandrogenism in women with polycystic ovarian syndrome? Hum Reprod 16:940, 2001

37. London BM, Lookingbill DP: Frequency of pregnancy in acne patients taking oral antibiotics and oral contraceptives. Arch Dermatol 130:392, 1994

38. Pollock B, Turner D, Stringer MR, et al: Topical aminolaevulinic acid–photodynamic therapy for the treatment of acne vulgaris: a study of clinical efficacy and mechanism of action. Br J Dermatol 151:616, 2004

39. Tzung TY, Wu KH, Huang ML: Blue light phototherapy in the treatment of acne. Photodermatol Photoimmunol Photomed 20:266, 2004

40. Friedman PM, Jih MH, Skover GR, et al: Treatment of atrophic facial acne scars with the 1064-nm Q-switched Nd:YAG laser: six-month follow-up study. Arch Dermatol 140:1337, 2004

41. Elewski BE, Fleischer AB Jr, Pariser DM: A comparison of 15% azelaic acid gel and 0.75% metronidazole gel in the topical treatment of papulopustular rosacea: results of a randomized trial. Arch Dermatol 139:1444, 2003

Reviews

Feldman S, Careccia RE, Barham KL, et al: Diagnosis and treatment of acne. Am Fam Physician 69:2123, 2004

James WD: Clinical practice. Acne. N Engl J Med 352:1463, 2005

Webster GF: Acne vulgaris. BMJ 325:475, 2002

XIII DISORDERS OF HAIR

David A. Whiting, m.d.

Physiology and Evaluation of Hair Growth

A basic knowledge of the hair growth cycle is needed to evaluate disorders of hair growth.[1,2] Scalp hair follicles cycle independently of one another. On average, of 100,000 scalp hairs, approximately 90% are in the anagen (growing) phase and 10% in the telogen (resting) phase at any given time. Anagen lasts an average of 3 years, with a range of 1 to 7 years. Telogen usually lasts 3 months, after which the resting hairs are shed and new hairs grow in. The average rate of scalp hair growth is approximately 0.35 mm/day, or 1 cm/month (1 in. every 2 to 3 months).

An average loss of 100 hairs a day is normal, with larger numbers of hairs being lost on shampoo days. When obtaining a history, it is important to determine whether shedding is abnormal and whether the shed hairs break off or come out by the roots.[3] Hair normally comes out by the roots; however, trauma or excessive fragility may cause hair to break.

Examination of the patient should include a routine check for broken-off hairs and the performance of hair-pull tests on the top, sides, and back of the scalp. The hair-pull test is performed by grasping groups of 10 to 20 hairs between the index finger and thumb and pulling steadily.[4,5] Extraction of more than 20% of the grasped hairs indicates a potential for abnormal shedding, usually involving telogen hairs. Telogen hairs (club hairs) are easily recognized by their whitish club-shaped bulbs and lack of root sheaths. Anagen hairs are normally difficult to detach and have blackish, indented roots with intact root sheaths.

Androgenetic Alopecia

Androgenetic alopecia is the common type of nonscarring hair loss affecting the crown. It results from a genetically determined end-organ sensitivity to androgens. It is often referred to as common baldness, male-pattern alopecia, and female-pattern alopecia.

epidemiology and pathogenesis

Androgenetic alopecia affects at least 50% of men by 50 years of age and 50% of women by 60 years of age.[6,7] Males have more androgen than females and therefore are usually affected earlier and more severely. Male-pattern alopecia often starts between 15 and 25 years of age. Male-pattern alopecia has two characteristic components, bitemporal recession and vertex balding [*see Figure 1*], which in pronounced cases can progress to complete balding of the crown.[6,7] Female-pattern alopecia is more likely to start between 25 and 30 years of age (or sometimes later, after menopause). It is characterized by an intact frontal hairline and an oval area of diffuse thinning over the crown [*see Figure 2*]. Bitemporal recession in women is much less obvious than it typically is in men, or it can be nonexistent. In general, androgenetic alopecia in women progresses to mild, moderate, or severe thinning but not to complete baldness. The best predictor of outcome is the degree of progression in affected relatives.

Androgenetic alopecia is an autosomal dominant disorder with variable penetrance. Susceptible hairs on the crown are predisposed to miniaturize under the influence of androgens, notably dihydrotestosterone. In both sexes, miniaturization results from a shortening of the anagen cycle, from years to months or weeks. Miniaturized hairs are characterized by reduced length and diameter; this accounts for the appearance of hair loss.[8] Androgenetic alopecia largely spares the back and sides of the scalp.

diagnosis

The diagnosis of androgenetic alopecia is usually obvious from the clinical pattern of hair loss from the top of the head.[9] In some men, a female pattern of alopecia (see above) causes diagnostic confusion but has no other significance. In women, a male pattern of alopecia (i.e., bitemporal recession and vertex balding) occurring with menstrual irregularities, acne, hirsutism, and a deep voice is significant. The virilism indicates significant hyperandrogenism, the cause of which must be identified and treated [*see 3:III Ovary and 3:IV The Adrenal*].

Scalp biopsies are rarely necessary to diagnose androgenetic alopecia. Biopsies cut horizontally are sometimes useful, however, in differentiating female-pattern alopecia from chronic telogen effluvium (see below).

treatment

Depending on the severity of the condition, management of androgenetic alopecia ranges from watchful inactivity to medical and surgical treatment, or a hairpiece or wig may be used in the most refractory cases.

Topical Therapy

The Food and Drug Administration approved topical 2% minoxidil for use in men in 1987 and in women in 1989. Minoxidil is applied twice daily with a dropper, spread over the top of the scalp, and gently rubbed in. The drug should be tried for at least a year. Minoxidil acts by initiating and prolonging anagen. It produces visible hair growth in approximately one third of male and female patients, fine-hair growth in approximately one third, and no growth in approximately one third. It is more effective as a preventive agent, retarding hair loss in approximately 80% of patients.[6]

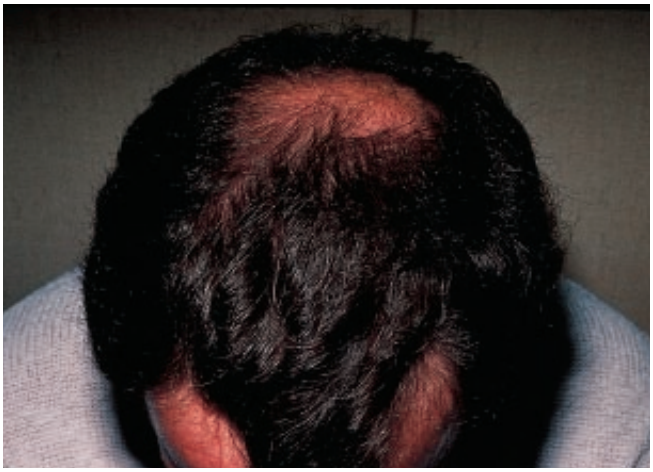


Figure 1 **Bitemporal recession and vertex balding are present in this patient with male pattern androgenetic alopecia.**

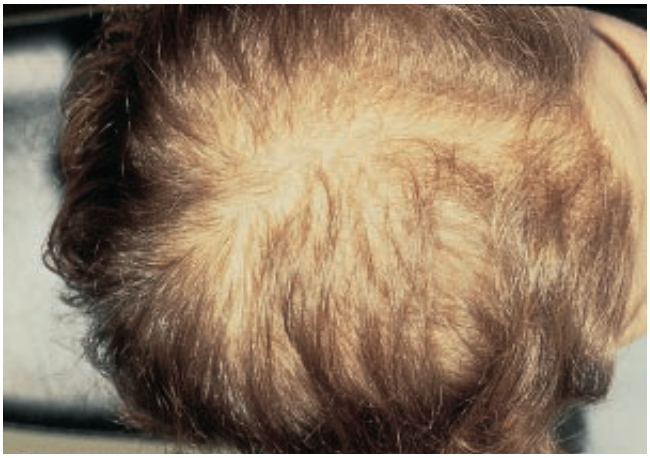


Figure 2 **Intact frontal hairline and diffuse thinning over the crown are characteristic of female pattern androgenetic alopecia.**

Topical 5% minoxidil, which was approved for use in men in 1997, produces visible hair growth in 45% of patients in less time than the 2% solution. Both concentrations are available over the counter. Side effects are not significant and include scalp irritation and increased facial hair.[10] The medication has to be continued indefinitely.[11]

Systemic Therapy

Oral finasteride, at a dosage of 1 mg/day, was approved by the FDA for the treatment of male-pattern alopecia in 1997. Finasteride is a powerful type II 5α-reductase inhibitor that prevents formation of dihydrotestosterone in the prostate gland and in the hair follicle. It reduces circulating dihydrotestosterone by 65% to 70%. When administered at a dosage of 1 mg/day for 2 years to male patients with androgenetic alopecia who are between 18 and 41 years of age, finasteride grew visible hair in 66% and prevented further hair loss in 83%.[12] The efficacy of finasteride was maintained in a 5-year study.[13] Hair-weight studies have shown that finasteride increases hair length and diameter, producing better coverage from existing hairs.[14]

Side effects in men are minimal and include lack of libido, lack of potency, and mild reduction in semen in approximately 0.5% of patients. These effects are reversed when the drug is stopped and often disappear as the drug is continued. A 1-year trial of finasteride at a dosage of 1 mg/day in postmenopausal women failed to show any positive effects.

Because of the likelihood of finasteride to cause severe side effects in the male fetus, the drug is contraindicated in premenopausal women.

Therapy for Hair Loss in Women

Topical minoxidil is currently the best available treatment for androgenetic alopecia in women.[10,15] However, various antiandrogenic drugs have been used. Oral contraceptives (e.g., ethinyl estradiol–ethynodiol diacetate [Demulen], desogestrel–ethinyl estradiol [Desogen], and ethinyl estradiol–norgestimate [Ortho Tri-Cyclen]) can reduce hair loss and occasionally lead to slight hair growth.[6] Oral spironolactone (Aldactone) in dosages of 75 mg/day to 200 mg/day can produce androgen blockade. Dexamethasone in dosages of 0.125 mg/day to 0.5 mg/day can suppress adrenal overactivity. Cyproterone acetate, which is not available in the United States, is not as effective as minoxidil in female pattern hair loss unless other signs of hyperandrogenism are present.[16]

Therapy for Refractory Cases

In patients who do not respond to the treatments listed above, the next step may be hair transplantation. Micrografts and minigrafts can produce a good cosmetic appearance in patients who have a sufficient reserve of hair on the back and sides of the scalp.[17] If all therapies fail, a hairpiece may be an option.

Diffuse Alopecia

Diffuse alopecia is generalized hair loss over the entire scalp. Because the loss is so diffuse, it is often unnoticeable until 30% to 50% of scalp hair is shed. Causes of diffuse alopecia include telogen effluvium, anagen arrest, drug reactions, and a number of systemic and nonsystemic conditions [*see Table 1*].[18,19]

TELOGEN EFFLUVIUM

Telogen effluvium is the most common form of diffuse alopecia.[20] It presents as a generalized shedding of telogen hairs from normal resting follicles. The basic cause of telogen effluvium is a premature interruption of anagen, leading to an increase in the number of hairs cycling into telogen. When the 3-month telogen period ends, new anagen hairs grow in and numerous telogen hairs fall out. Patients may need reassurance that this apparent loss of hair is actually a sign of regrowth.

Acute telogen effluvium can be caused by childbirth, febrile illnesses, surgery, chronic systemic diseases, crash diets, traction, severe emotional stress, and drug reactions [*see Table 2*]. It can also be a physiologic reaction in neonates.[21]

During acute telogen effluvium, pull tests are positive all over the scalp, yielding two to 10 club hairs. Telogen effluvium is often accompanied by bitemporal recession; this is a useful diagnostic sign in women [*see Figure 3*]. The acute form usually ends within 3 to 6 months. The diagnosis is usually made on the basis of the history of an initiating event 3 months before the onset of shedding. No treatment is needed for acute telogen effluvium, because the hair invariably regrows within a short time.

Chronic telogen effluvium has a long, fluctuating course of 6 months to 7 years or more. Very often, no identifiable cause can be found.

Diagnosis

The diagnosis of telogen effluvium is usually clinical; biopsies may be necessary to distinguish telogen effluvium from an acute onset of widespread androgenetic alopecia.[22] Other causes of

Table 1 Causes of Diffuse Alopecia[45]

Telogen effluvium (acute and chronic)
Anagen arrest
Reactions to drugs and other chemicals
Thyroid disorder
Iron deficiency and other nutritional deficiencies
Malabsorption
Renal failure
Hepatic failure
Systemic disease
Miscellaneous causes (e.g., diffuse alopecia areata, congenital hypotrichosis) and idiopathic causes

Table 2 Categories of Drugs That Can Cause Alopecia[5]

Category	*Selected Agents*
Alpha blockers	Doxazosin, prazosin, terazosin
Angiotensin converting enzyme inhibitors	Captopril, enalapril
Anticancer drugs	Bleomycin, cyclophosphamide, cytarabine, dactinomycin, daunorubicin, doxorubicin, etoposide, floxuridine, fluorouracil, methotrexate, mitomycin, mitoxantrone, procarbazine, thioguanine, vinblastine, vincristine
Anticoagulant drugs	Dicumarol, heparin, warfarin
Anticonvulsant drugs	Ethotoin, mephenytoin, paramethadione, phenytoin, trimethadione, valproate sodium
Antithyroid drugs	Carbimazole, methylthiouracil, methimazole, propylthiouracil
Beta blockers	Acebutolol, atenolol, labetalol, metoprolol, nadolol, pindolol, propranolol, timolol
Calcium channel blockers	Diltiazem, verapamil
Cholesterol reducers	Clofibrate, lovastatin
H_2 receptor blockers	Cimetidine, famotidine, ranitidine
Nonsteroidal anti-inflammatory drugs	Fenoprofen, ibuprofen, indomethacin, ketoprofen, meclomen, naproxen, piroxicam, sulindac
Retinoids and retinol	Acitretin, etretinate, isotretinoin, vitamin A overdose
Tricyclic antidepressants	Amitriptyline, amoxapine, desipramine, doxepin, imipramine, nortriptyline, protriptyline, trimipramine

hair loss should be excluded by a careful drug history and tests for iron deficiency, syphilis, and disorders of the thyroid, kidney, and liver.

Treatment

As mentioned above, no treatment is needed for acute telogen effluvium because the hair invariably regrows within a short time. In chronic telogen effluvium, topical minoxidil in a 2% or 5% solution may be indicated. The patient should be reassured that telogen effluvium rarely causes permanent baldness.

ANAGEN ARREST (ANAGEN EFFLUVIUM)

So-called anagen effluvium represents a diffuse loss of anagen hairs from growing follicles.[23] The term anagen effluvium is a misnomer. Normally, hairs pass through a brief transition phase (catagen) between the anagen and telogen phases before falling out by the roots. In anagen arrest, inhibition of cell division in the hair bulb matrix leads to a progressive narrowing of the hair shaft and sometimes failure of hair formation. As the growing hair narrows near the skin surface, it may break off. The resultant shedding can occur within a few weeks, unlike in telogen effluvium, in which shedding takes 3 months to occur.

Causes of anagen arrest include reactions to cytostatic drugs and other toxic agents, radiation therapy, endocrine diseases, alopecia areata, cicatricial alopecia, trauma and pressure, and severe protein calorie malnutrition. Because 90% of scalp hairs are in anagen at any given time, this condition causes obvious and severe baldness [*see Figure 4*].

Diagnosis

The diagnosis of anagen arrest is easily made by the history, evidence of extensive hair loss, and hair-pull tests that yield easily broken hairs with proximal tapering.

Treatment

Treatment of anagen arrest lies in elimination of the underlying cause. Once the antimitotic influence is removed, the anagen

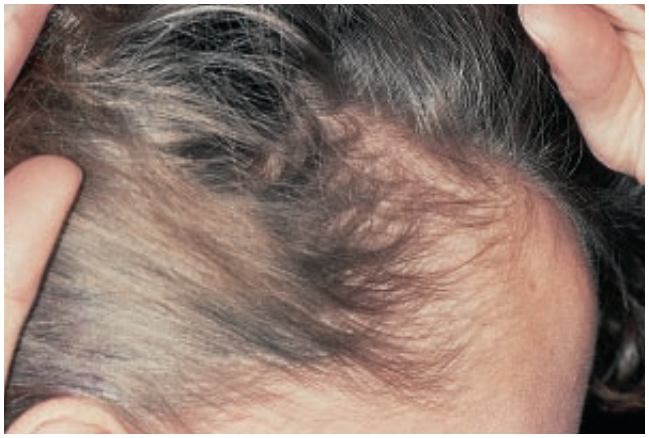


Figure 3 **In women, marked bitemporal recession is often a sign of telogen effluvium.**

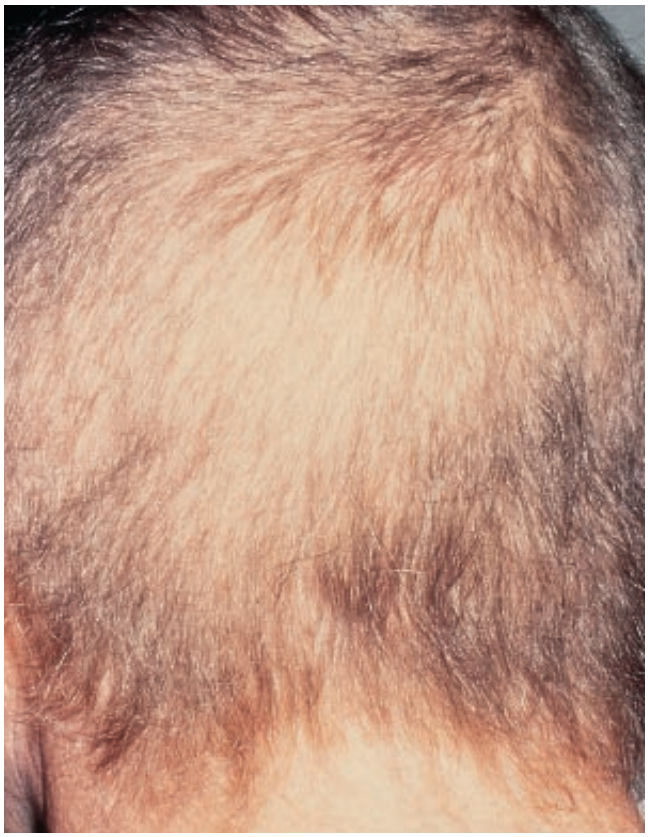


Figure 4 **Anagen arrest causes severe, diffuse hair loss.**

Table 3 Miscellaneous Chemicals That Can Cause Alopecia[45]

Chemical	*Common Source*
Abrin	Plant source (*Abrus precatorius* [rosary pea, jequirity bean, or precatory bean])
Arsenic	Pesticides
Bismuth	Old treatment for syphilis
Boric acid	Mouthwashes, occupational exposure
Chloroprene dimers	Occupational exposure (synthetic-rubber manufacturing)
Lead	Paints
Mercury	Cosmetics, teething powders, antiseptics
Mimosine	Plant source (*Leucaena glauca*)
Selenocystothione	Plant source (*Lecythis* species)
Thallium salts	Rodenticides

hair will regrow promptly with a normally tapering shaft. Unbroken hairs that regrow often show the Pohl-Pinkus deformity (i.e., a constriction that results in a dumbbell shape).

ALOPECIA CAUSED BY DRUGS AND CHEMICALS

Substance-induced alopecia is relatively common but is often hard to diagnose because of the large number of drugs and chemicals that can cause hair loss [*see Tables 2 and 3*]. It often takes time to identify the underlying cause by trial and error: many patients are exposed to several alopecia-inducing substances, and removal of the causative agent may not result in immediate regrowth of hair.

OTHER CAUSES OF DIFFUSE ALOPECIA

Hypothyroidism and iron deficiency should be excluded in patients with diffuse hair loss [*see Table 1*]. Appropriate treatment may lead to hair regrowth.

Alopecia Areata

Alopecia areata is typically characterized by patchy hair loss; however, involvement can vary from a single patch on the scalp or elsewhere to total body baldness (alopecia universalis).[24]

EPIDEMIOLOGY AND PATHOGENESIS

In the United States, alopecia areata affects 1.7% of the population younger than 50 years.[25] Some 70% to 75% of cases are not associated with any other disease. In these patients, alopecia areata often starts in the 20s and 30s, although it can occur at any age. In only about 6% of these patients with alopecia areata does the disease progress to total loss of scalp hair. Even total alopecia can reverse itself.

Alopecia areata is currently regarded as an autoimmune disease. A positive family history in 20% of alopecia areata patients indicates a genetic predisposition to this disease. Certain HLA groups have been associated with mild or severe cases of alopecia areata.[26] Although the exact cause is unknown, many researchers presume that an infectious agent such as a virus is the offending agent. Stress, seasonal factors, and infection are among the trigger factors for active episodes of hair loss.

Some 5% of alopecia areata cases—usually those occurring in middle-aged patients—are associated with autoimmune disease, either in the patient or in the patient's family. Some 10% of these patients will experience loss of all scalp hair in the course of the disorder. Approximately 20% to 25% of cases—often those occurring in childhood—may be associated with atopic disease (e.g., hay fever, asthma, or eczema). The incidence of complete scalp hair loss is much higher in these patients.

Despite its long course, often recurring over many years, the prognosis of alopecia areata is often favorable. Most patients will regrow hair at one time or another. In cases of extensive alopecia areata, alopecia totalis, and alopecia universalis, however, hair loss may be permanent.

DIAGNOSIS

Active alopecia areata is characterized by a spreading, annular area of hair loss; a smooth, depressed area of scalp that is slick to the touch is surrounded by hairs that often include so-called exclamation-point hairs (i.e., broken hairs 3 to 4 mm long, usually with an expanded tip and a telogen bulb). These hairs are not always seen but are diagnostic when present. They delineate the active spreading margin of alopecia areata. The bald patches generally affect the scalp but can also involve eyebrows, eyelashes, beard hair, and body hair. Spontaneous regrowth is common.

This condition is extremely unpredictable, often fluctuating without any obvious reason. However, seasonal outbreaks are noted in many patients. The initial patch may enlarge, or additional patches may develop and become confluent [*see Figure 5*]. The condition can progress to large irregular areas of baldness. In severe cases, patients lose all scalp hair or all body hair.

Ophiasis is a chronic and difficult to treat form of alopecia areata in which a band of baldness circles the scalp, very often around the inferior margin. This slowly extending lesion is often present for several years before any regrowth occurs. Permanent hair loss may result in some areas.

TREATMENT

The treatment of alopecia areata depends on the severity of the disease.[27] Small patches of alopecia areata often regrow hair without treatment. If not, they usually respond to medium- or high-potency topical corticosteroids or to intralesional injections of triamcinolone acetonide at a concentration of 5 mg/ml.

In more severe cases, intralesional corticosteroids may be tried; however, these may not be feasible in extensive hair loss. Daily, short-contact topical therapy with 0.25% to 1% anthralin cream for up to an hour at a time may help and is suitable for children and adults. Psoralen and ultraviolet A (PUVA) therapy has also been used with some success.

Topical 5% minoxidil can be tried to speed hair regrowth and lengthen existing hairs. Minoxidil has no effect on the course of the disease but may improve hair coverage. It has few side effects and is often used in older children; however, the FDA has approved minoxidil for use only in persons 18 years of age and older. Systemic steroids are effective; however, they have shown a potential for side effects and do not prevent future recurrences.[28] Prednisone (20 to 40 mg daily in the morning for 1 or 2 months followed by slow tapering) has controlled the disease in adults; a change to alternate-day therapy is advisable whenever possible.

Topical immunotherapy with the sensitizing chemical diphencyprone has been used in some centers; it has a response rate comparable to that of systemic corticosteroids.[29] Success with this treatment usually requires supervision in a specialized clinic.[27]

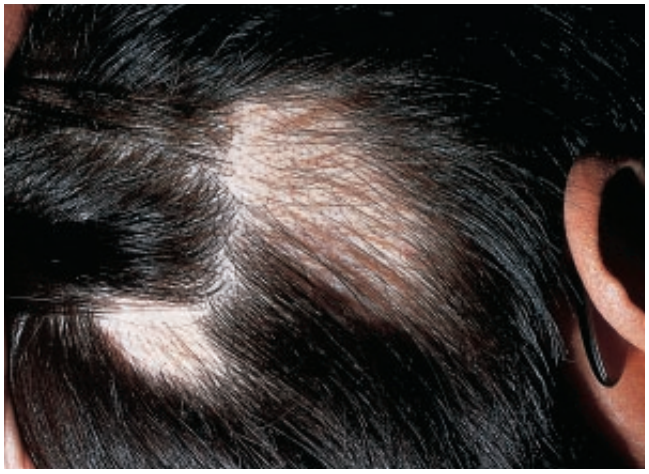


Figure 5 **Circumscribed patches of hair loss are present in alopecia areata.**

Sulfasalazine has been reported to have a 23% success rate in the treatment of severe alopecia areata.[30] Other immunosuppressive drugs, such as oral cyclosporine, have been used experimentally. Such therapies are risky and expensive, however, and have not been approved in the United States for alopecia areata.

TRAUMATIC ALOPECIA

Traumatic alopecia may be caused by a variety of physical or chemical injuries to the hair and scalp. These injuries may be deliberate or accidental, inflicted by self or others, and acute or repetitive. The cause may be obvious or unclear.[31] Potential causes include trichotillomania, habit tics, pruritic dermatoses, traction, pressure and friction, heat, radiation, and chemicals. In most cases of traumatic alopecia, management is removal of the underlying cause. In areas with permanent damage, hair transplantation may be necessary.

TRICHOTILLOMANIA

Trichotillomania is a compulsion to pull out one's hair. It is characterized by an increasing sense of tension before, and a sense of relief after, the hair is pulled. Trichotillomania is now classified as a specific disorder of impulse control.[31,32] It is more common in children, in whom it is often caused by insecurity resulting from sibling rivalry, lack of attention, divorce of parents, learning disabilities, or unhappiness or teasing at school. In adolescents and adults, trichotillomania may be accompanied by mood disorders, anxiety disorders, or mental retardation and is often harder to treat than in children.

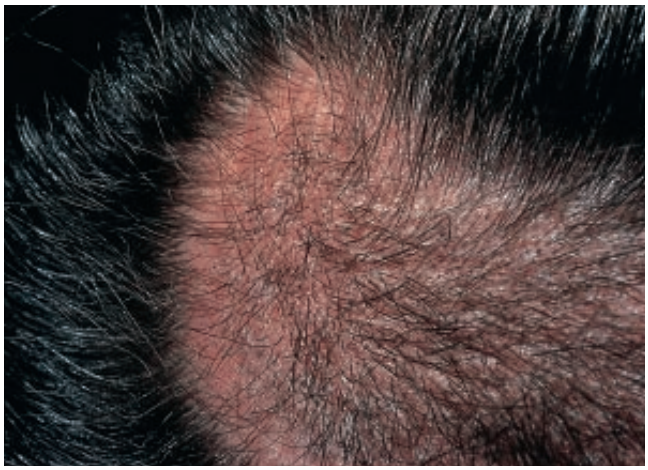


Figure 6 **Irregular, broken-off hairs are seen in trichotillomania.**

The diagnosis is based on the presence of irregular, broken-off hairs in patches on the scalp [*see Figure 6*]. The hairs are irregular in length because they are broken off at different times. The scalp itself is normal. Occasionally, biopsies are necessary to confirm the diagnosis. The best treatment is to explain cause and remedy to the patient in a nonconfrontational manner; usually, reassurance and understanding go a long way in the treatment of this condition. In more difficult cases, psychiatric consultation may be indicated.[33] If habit tics or head rolling and banging are found to be causing traumatic hair loss, those behaviors should be treated.

OTHER CAUSES OF TRAUMATIC ALOPECIA

Pruritic Dermatoses

Pruritic dermatoses such as acne necrotica, folliculitis, lichen simplex chronicus, pediculosis capitis, prurigo nodularis, psoriasis, seborrheic dermatitis, and neurotic excoriations can lead to hair loss from excoriation. They need to be identified and treated.

Traction Alopecia and Loose Anagen Syndrome

Traction alopecia may be acute (caused by accidental or deliberate avulsion of the scalp) or may arise from a familial condition, the loose anagen syndrome. Common causes of traction alopecia are excessive brushing and combing; backcombing and pulling the hair into braids, cornrows, and ponytails; weaving; and application of rollers.[31]

Loose anagen syndrome is usually seen in fair-haired children 2 to 5 years of age.[34] It often presents as patchy hair loss following an incident of hair tugging. Prompt hair regrowth is the rule. The condition becomes asymptomatic with gentle hair care. Diagnosis is made on the basis of positive hair pull tests showing many anagen hairs.

Alopecia Caused by Pressure and Friction

Prolonged pressure on a localized area of the scalp in immobilized neonates or patients under anesthesia, in coma, or with debilitating illness may result in ischemia leading to pressure alopecia. The hair usually regrows with time, but if the damage is severe, permanent hair loss and scarring may result.[31] Alopecia caused by friction from vigorous massage has been described but is easily remedied.

Alopecia Caused by Heat, Radiation, and Chemicals

Excessive heat from hot oils and pomades, hot combs, and hot rollers is a common cause of chronic hair loss. Overheated hair dryers frequently cause the fluid droplets in wet hair shafts to expand, leading to the formation of bubble hairs.[35] These brittle hairs are a frequent cause of follicle damage. The source of the overheating needs to be identified and removed.

Radiation dermatitis can cause hair loss. Permanent scarring alopecia is still seen in patients who were overtreated with x-rays for tinea capitis before oral antifungal agents became available.

Many chemicals, such as hair dyes, moisturizers, oils and pomades, permanent waves, relaxers and straighteners, setting lotions, certain cationic and detergent shampoos, and saltwater, are possible causes of hair loss.[36] A careful history of hair care and grooming is needed to uncover these causes.

Cicatricial Alopecia

Cicatricial alopecia results from permanent scarring of the hair follicles. It may be widespread or localized and is sometimes difficult to identify. The causes of cicatricial alopecia may be primary or secondary.[37] It can result from hereditary or congenital conditions, infections, injuries, neoplasms, and dermatoses [*see Table 4*].

DIAGNOSIS

On clinical examination, scarring is detected by the absence of follicular orifices and a pearly or scarred appearance of the skin. The scar may be depressed or hypertrophic. Associated lesions such as folliculitis, follicular plugs, scales, and telangiectasias may be found, along with broken, twisted, or easily extractable hairs. Other lesions may be present on skin or mucous membranes. If the disease is active, a specific diagnosis may be possible; but in an inactive case, the initial cause is often inapparent.

Clinical Variants

The common variants of primary cicatricial alopecia of the scalp include discoid lupus erythematosus, lichen planopilaris, folliculitis decalvans, and pseudopelade.[38,39] The end phases of these conditions are similar; they are characterized by a lack of pores and by inflammation in white, scarred areas. For an accurate diagnosis, an early biopsy from an area of activity might show the identifying pathology. In the final scarring stage, it is usually not possible to identify the original cause.

Discoid lupus erythematosus Lesions are often itchy at onset and lead to erythema, scaling, telangiectasia, follicular spines, and atrophy [*see Figure 7*]. They often occur centrally in bare patches of scarring with an inactive border [*see 15:IV Systemic Lupus Erythematosus*].

Lichen planopilaris Central scarring characterizes these lesions. The condition generally starts with bare, white patches that bud out from one another like pseudopods. Prominent follicular hyperkeratosis is present around the residual terminal hairs at the edges of the lesion, and varying degrees of erythema, scaling, and telangiectasia may occur. Itching may be present.

Folliculitis decalvans Crops of follicular pustules surrounding multiple, slowly expanding, and round or oval areas of alopecia characterize this condition. It may involve large areas of the skin. Secondary infection may be severe, with crusting and oozing. Eventually this condition gradually loses activity and looks like other forms of chronic cicatricial alopecia.

Pseudopelade (nonspecific cicatricial alopecia) The majority of cases classified as pseudopelade are in fact cases of nonspecific cicatricial alopecia in which the initial cause has not been established. In general, there is an insidious spread of a scarring process, which is apparently noninflammatory. It often involves the crown and occurs mainly in middle-aged women. It may represent the final common pathway of various causes, such as lichen planopilaris in particular or discoid lupus erythematosus. It is characterized by patchy areas of alopecia with irregular extensions. The affected skin is smooth, white, and devoid of erythema, scaling, or pores, causing the so-called footprints in the snow. The course is variable and may last for a few years or several decades.

The original cases were described as a specific entity in the late 19th century and were reported as pseudopelade of Brocq.

Table 4 Causes of Cicatricial Alopecia[46]

Dermatoses	Cicatricial pemphigoid, dermatomyositis, folliculitis decalvans, lichen planopilaris, lupus erythematosus, neurotic excoriations, pseudopelade, scleroderma
Hereditary and congenital disorders	Aplasia cutis, epidermal nevi, epidermolysis bullosa
Infections	
Bacterial	Acne keloidalis, dissecting cellulitis, folliculitis, syphilis
Fungal	Favus (tinea capitis), kerion, mycetoma
Protozoan	Leishmaniasis
Viral	Herpes zoster, varicella
Injuries	Burns, mechanical trauma, radiodermatitis
Neoplasms	Angiosarcoma, basal cell epithelioma, lymphoma, melanoma, metastatic tumors, squamous cell carcinoma

This eponym is rarely used nowadays except perhaps for a small cohort of cases with no inflammatory phase at all, particularly occurring in children. Most cases of so-called pseudopelade of Brocq are explainable as an end stage of lichen planopilaris or other causes.[37]

TREATMENT

Treatment depends on the level of activity of the underlying disease. Discoid lupus erythematosus and lichen planopilaris may respond to topical, intralesional, or systemic steroids or oral chloroquine therapy. Topical minoxidil is sometimes helpful in regrowing any surviving hairs, which may be normal, dystrophic, or in a resting stage. The application of a 2% or 5% solution twice daily should be tried for at least a year on scarred areas that show

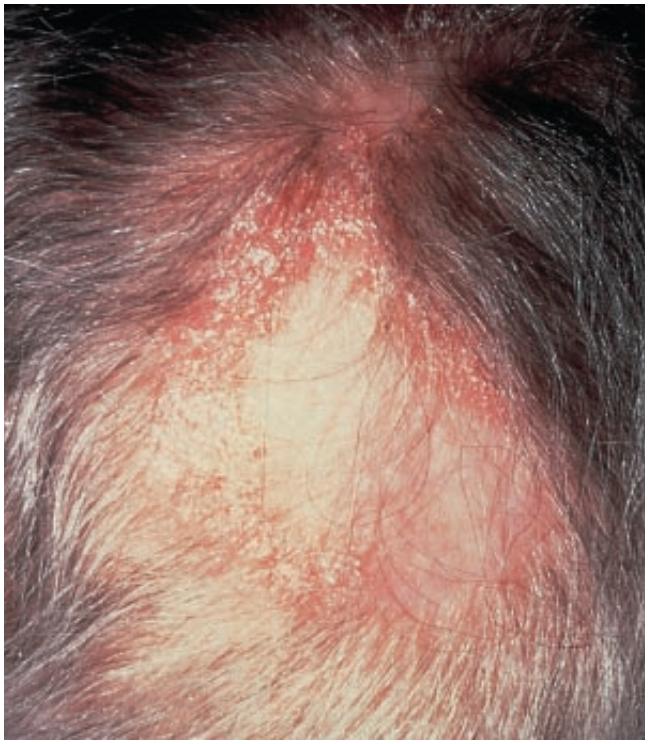


Figure 7 **Atrophic scarring with erythema, scaling, telangiectasia, and follicular spines are characteristic of discoid lupus erythematosus.**

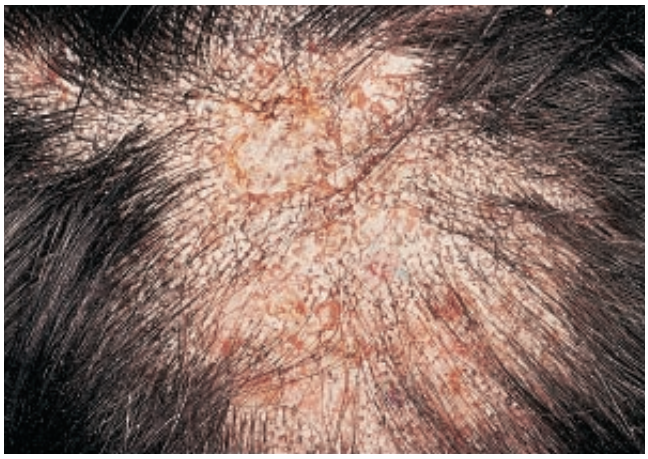


Figure 8 **In endothrix tinea capitis, black dots represent brittle, infected hairs snapped off flush with the scalp.**

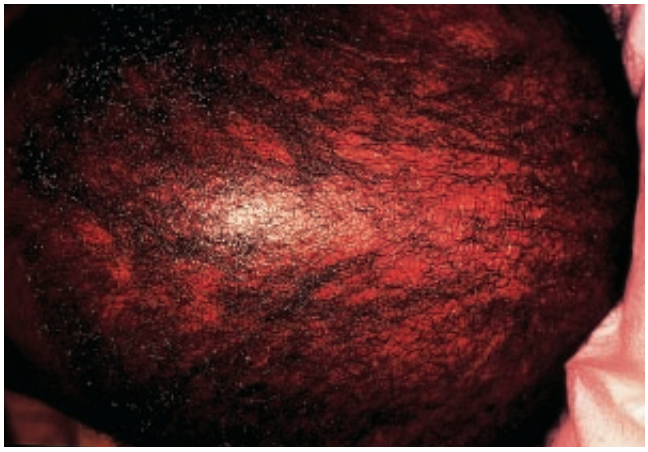


Figure 9 **The characteristic irregular, so-called moth-eaten diffuse alopecia caused by syphilis is seen here.**

some hair. Folliculitis decalvans may respond to treatment with long-term antibiotics such as tetracycline (500 mg), minocycline (100 mg), erythromycin (250 mg), trimethoprim-sulfamethoxazole (regular strength), or rifampin (300 mg, with regular-strength trimethoprim-sulfamethoxazole, given twice daily for at least 3 months). When the conditions have burned themselves out, either scalp reduction or hair transplants may be helpful.

Miscellaneous Causes of Hair Loss

As mentioned earlier, less common causes of hair loss include infections (e.g., tinea capitis),[40] infestations, hair shaft abnormalities, hereditary and congenital conditions, and various dermatoses involving the scalp.

In the United States, tinea capitis is now largely caused by *Trichophyton tonsurans*, an endothrix that infects the inside of the hair shaft. This makes the shaft brittle, which causes it to snap off flush with the skin, leaving a characteristic black dot of hair [*see Figure 8*]. The clinical diagnosis depends on this finding of black dots in patchy areas of hair loss. Removing the black dot with a small scalpel blade and dissolving it in potassium hydroxide (KOH) should reveal many spores that were packed inside the affected hair shaft. Ectothrix ringworm caused by *Microsporum canis* and *M. audouinii* is much less common; it can usually be diagnosed by Wood's light or by a finding of fungal spores around the hair shaft with KOH. Suitable oral antifungal treatments include griseofulvin, itraconazole, terbinafine, and fluconazole.

Secondary syphilis can cause a somewhat nondescript, moth-eaten type of diffuse alopecia [*see Figure 9*]. In such cases, a routine serologic test for syphilis is indicated.

Scalp lice should always be sought in cases of hair loss accompanied by pruritus. Lice are most likely to be found around and behind the ears and on the nape of the neck. Lymphadenopathy may also be present. Suitable treatment with permethrin shampoo or 0.5% malathion can be given.

Hair shaft abnormalities frequently present as broken-off hairs. Structural abnormalities of the hair shaft include fractures, irregularities, coiling and twisting, and extraneous matter.[41,42]

There are many different types of congenital and inherited hair loss.[43] These include congenital hypotrichosis with or without associated defects, congenital triangular alopecia, and many ectodermal dysplasias that affect the hair, teeth, nails, and sweat glands. One major form of congenital hypotrichosis is the Marie-Unna syndrome, which affects large families that have a dominant gene.[44] Minor forms of hypotrichosis can occur in patients with other hereditary syndromes and chromosomal abnormalities. In most of these conditions, a reduction of hair follicles accounts for the hair loss. Some patients have surviving hairs that are often in telogen and may benefit from topical minoxidil.

The author is a consultant and participates in the speakers' bureaus for Pharmacia & Upjohn, Merck & Co., Inc., and GlaxoSmithKline.

Oral contraceptives and antiandrogens discussed in this subsection have not been approved by the FDA for use in androgenetic alopecia; topical minoxidil discussed in this subsection has not been approved by the FDA for use in chronic telogen effluvium and cicatricial alopecia; diphencyprone, sulfasalazine, and cyclosporine discussed in this subsection have not been approved by the FDA for use in alopecia areata; and rifampin discussed in this subsection has not been approved by the FDA for use in folliculitis decalvans.

References

1. Abell E: Embryology and anatomy of the hair follicle. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 1

2. Stenn KS, Paus R: Controls of hair follicle cycling. Physiol Rev 81:449, 2001

3. Olsen EA: Clinical tools for assessing hair loss. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 59

4. Rietschel RL: A simplified approach to the diagnosis of alopecia. Dermatol Clin 14:691, 1996

5. Whiting DA, Howsden FL: Assessment of patient with hair loss. Color Atlas of Differential Diagnosis of Hair Loss, rev. Whiting DA, Howsden FL, Eds. Canfield Publishing, Fairfield, New Jersey, 1998, p 8

6. Olsen EA: Androgenetic alopecia. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 257

7. Sawaya ME: Clinical updates in hair. Dermatol Clin 15:37, 1997

8. Birch MP, Messenger JF, Messenger AG: Hair density, hair diameter and the prevalence of female pattern hair loss. Br J Dermatol 144:297, 2001

9. Whiting DA, Howsden FL: Androgenetic alopecia. Color Atlas of Differential Diagnosis of Hair Loss, rev. Whiting DA, Howsden FL, Eds. Canfield Publishing, Fairfield, New Jersey, 1998, p 18

10. Shapiro J, Price VH: Hair regrowth: therapeutic agents. Dermatol Clin 16:341, 1998

11. Olsen EA, DeLong ER, Weiner MS: Long-term follow-up of men with male pattern baldness treated with topical minoxidil. J Am Acad Dermatol 16:688, 1987

12. Kaufman KD, Olsen EA, Whiting DA, et al: Finasteride in the treatment of men with androgenetic alopecia. Finasteride Male Pattern Hair Loss Study Group. J Am Acad Dermatol 39:578, 1998

13. Kaufman KD: Long-term (5-year) multinational experience with finasteride 1 mg in the treatment of men with androgenetic alopecia. Eur J Dermatol 12:38, 2002

14. Price VH, Menefee E, Sanchez M, et al: Changes in hair weight and hair count in men with androgenetic alopecia after treatment with finasteride, 1 mg, daily. J Am Acad Dermatol 46:517, 2002

15. Tosti A, Piraccini BM: Androgenetic alopecia. Int J Dermatol 38(suppl 1):1, 1999

16. Vexiau P, Chaspoux C, Boudou P, et al: Effects of minoxidil 2% vs. cyproterone acetate treatment on female androgenetic alopecia: a controlled, 12-month randomized trial. Br J Dermatol 146:992, 2002

17. Unger WP: What's new in hair replacement surgery. Dermatol Clin 14:783, 1996

18. Fiedler VC, Hafeer A: Diffuse alopecia: Telogen hair loss. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 241

19. Sinclair R: Diffuse hair loss. Int J Dermatol 38(suppl 1):8, 1999

20. Headington JE: Telogen effluvium: new concepts and review. Arch Dermatol 129:356, 1993

21. Kligman AM: Pathologic dynamics of human hair loss 1: telogen effluvium. Arch Dermatol 83:175, 1961

22. Whiting DA: Chronic telogen effluvium. Dermatol Clin 14:723, 1996

23. Grossman KL, Kvedar JC: Anagen hair loss. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 223

24. Hordinsky MK: Alopecia areata. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 195

25. Safavi KH, Muller SA, Suman VJ, et al: Incidence of alopecia areata in Olmsted County, Minnesota, 1975 through 1989. Mayo Clin Proc 70:628, 1995

26. Price VH, Colombe BW: Heritable factors distinguish two types of alopecia areata. Dermatol Clin 14:679, 1996

27. Shapiro J, Madani S: Alopecia areata: diagnosis and management. Int J Dermatol 38(suppl 1):19, 1999

28. Shapiro J, Price VH: Hair regrowth: therapeutic agents. Dermatol Clin 16:341, 1998

29. Madani S, Shapiro J: Alopecia areata update. J Am Acad Dermatol 42:549, 2000

30. Ellis CN, Brown MF, Voorhees JJ: Sulfasalazine for alopecia areata. J Am Acad Dermatol 46:541, 2002

31. Whiting DA: Traumatic alopecia. Int J Dermatol 38(suppl 1):34, 1999

32. Walsh KH, McDougle CJ: Trichotillomania: presentation, etiology, diagnosis and therapy. Am J Clin Dermatol 2:327, 2001

33. Hautmann G, Hercogova J, Lotti T: Trichotillomania. J Am Acad Dermatol 46:807, 2002

34. Price VH, Gummer CL: Loose anagen syndrome. J Am Acad Dermatol 20:249, 1989

35. Detwiler SP, Carson JL, Woosley JT, et al: Bubble hair: case caused by overheating hair dryer and reproducibility in normal hair with heat. J Am Acad Dermatol 30:54, 1994

36. Wilborn WS: Disorders of hair growth in African Americans. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 389

37. Amato L, Mei S, Massi D, et al: Cicatricial alopecia: a dermatopathologic and immunopathologic study of 33 patients (pseudopelade of Brocq is not a specific clinicopathologic entity). Int J Dermatol 41:8, 2002

38. Headington JT: Cicatricial alopecia. Dermatol Clin 14:773, 1996

39. Whiting DA: Cicatricial alopecia: clinico-pathological findings and treatment. Clin Dermatol 19:211, 2001

40. DeVillez RL: Infectious, physical and inflammatory causes of hair and scalp abnormalities. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 71

41. Whiting DA: Hair shaft defects. Disorders of Hair Growth: Diagnosis and Treatment. Olsen EA, Ed. McGraw-Hill, New York, 1994, p 134

42. DeBerker D, Sinclair R: Defects of the hair shaft. Diseases of the Hair and Scalp, 3rd ed. Dawber R, Ed. Blackwell Science, Oxford, 1997, p 239

43. Sinclair R, DeBerker D: Hereditary and congenital alopecia and hypotrichosis. Diseases of the Hair and Scalp, 3rd ed. Dawber R, Ed. Blackwell Science, Oxford, 1997, p 151

44. Roberts JL, Whiting DA, Henry D, et al: Marie Unna congenital hypotrichosis: clinical description, histopathology, scanning electron microscopy of a previously unreported large pedigree. J Invest Dermatol Symp Proc 4:261, 1999

45. Dawber RR, Simpson NB, Barth JH: Diffuse alopecia: endocrine, metabolic and chemical influences on the follicular cycle. Diseases of the Hair and Scalp, 3rd ed. Dawber R, Ed. Blackwell Science, Oxford, 1997, p 123

46. Dawber RR, Fenton DA: Cicatricial alopecia. Diseases of the Hair and Scalp, 3rd ed. Dawber R, Ed. Blackwell Science, Oxford, 1997, p 370

Acknowledgments

Figures 2, 4, 7, and 8: D. A. Whiting and F. L. Howsden: *Color Atlas of Differential Diagnosis of Hair Loss, rev*. Canfield Publishing, Fairfield, New Jersey, 1998. Used with permission.

XIV DISEASES OF THE NAIL

James Q. Del Rosso, d.o.
C. Ralph Daniel iii, m.d.

The human nail is a complex unit composed of five major modified cutaneous structures: the nail matrix, nail bed, nail plate, nail folds, and cuticle (eponychium).[1] These components are structurally supported by specialized mesenchyme, which provides a ligamentlike function, anchoring the soft tissue structures of the nail to the underlying phalangeal bone. The primary function of the human nail is to provide protection for the distal digits. Nails also assist in performance of fine touch and digital dexterity. For many individuals, nails serve as an important aesthetic symbol of optimal appearance, enhanced self-image, or individuality; several cosmetic techniques are available to modify the appearance of the nail plate. The basic anatomic components of the nail unit are diagrammed [*see Figure 1*].

Nail Structure, Function, and Pathophysiology

NAIL MATRIX

The nail matrix is the dynamic, germinative portion of the nail unit that produces the nail plate.[2-4] The lunula is the visible portion of the nail matrix, appearing under the proximal nail plate as a gray-white half moon projecting just distal to the proximal nail-fold cuticle. That lunula decreases with age in approximately 20 % of persons.[5]

Nails are usually devoid of pigmentation because of the relatively sparse number of melanocytes present in matrix epithelium.[1,2] Because nail-matrix or nail-bed melanocytes tend to be more numerous in blacks, Asians, and Hispanics, persons of these racial backgrounds may present more commonly with diffuse or banded nail-plate or nail-bed hyperpigmentation.

Pathophysiology Affecting the Nail Matrix

Because of the diagonal orientation of the ventral nail matrix, the proximal portion of the nail matrix produces the superior portion of the nail plate.[6] As a result, disorders of the proximal matrix produce surface abnormalities of the nail plate. A characteristic example is nail-plate pitting secondary to psoriasis. Diseases of the distal nail matrix result in abnormalities of the undersurface of the nail plate, changes that are visible at the free edge of the nail, or both. Permanent damage to the matrix as the result of trauma, surgical intervention, or disease may result in permanent nail-plate dystrophy.

NAIL BED

The nail bed is a layer of epithelium lying between the lunula and the hyponychium (the distal epithelium at the free edge of the nail). The surface epithelium of the nail bed is longitudinally ridged, with small superficially oriented vessels coursing along the same axis, interdigitating with a complementary array of ridges on the undersurface of the nail plate.[3] This anatomic feature explains the longitudinal linearity of splinter hemorrhages, which are foci of extravasation wedged between the bed and the plate. As outgrowth of the nail plate occurs, splinter hemorrhages progress distally.

Pathophysiology Affecting the Nail Bed

The epidermis of the nail bed is thin and minimally keratinized, without a granular layer. If there is prolonged loss of nail plate as a result of disease or surgical intervention, in-

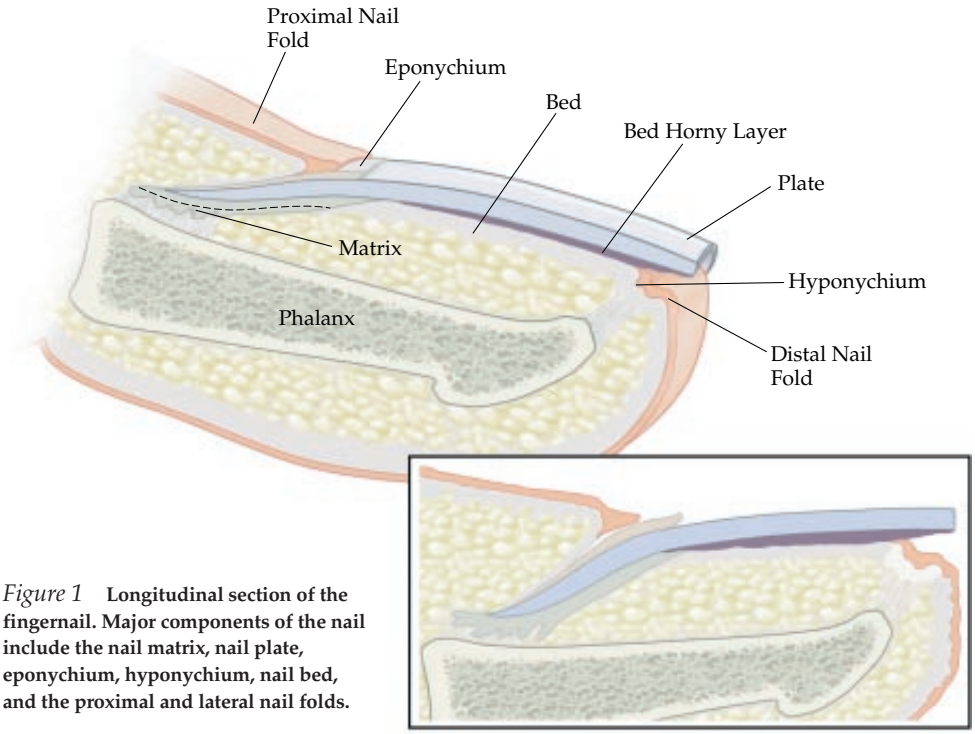


Figure 1 **Longitudinal section of the fingernail. Major components of the nail include the nail matrix, nail plate, eponychium, hyponychium, nail bed, and the proximal and lateral nail folds.**

creased nail-bed keratinization with development of a granular layer prevents the firm attachment of the ingrowing nail plate to the underlying nail bed. Melanocytes are more sparsely distributed in nail-bed epithelium than in the nail matrix [*see* Nail Matrix, *above*]. The dermal layer of the nail bed is very thin and is supported by very sparse subcutaneous tissue; it is firmly attached to the underlying bony phalanx.

NAIL PLATE

The nail plate, which is composed of densely compacted keratinized epithelial cells, is produced by the matrix and progesses distally toward the free edge of the nail as newly formed plate slowly pushes forward in a distal direction. Formation and outgrowth of the nail plate is a continual process. A fully formed nail plate extends from below the proximal nail-fold cuticle to beyond the hyponychium and extends laterally below the cuticle of the lateral nail folds. Nail-plate abnormalities frequently occur secondary to changes or disorders affecting function of the nail matrix; infections such as onychomycosis; or trauma.

Age-Related Nail-Plate Findings

The growth rate of an adult fingernail plate is approximately 3 mm/mo, with marked variability among individuals.[2] Toenail plate growth occurs at one third to one half the rate of fingernail growth. A general rule is that adult fingernails take approximately 6 months to grow out fully; adult toenails, 12 to 18 months. Nail-plate growth is faster in children, peaking between 10 and 14 years of age; there is a slowly progressive decline after the second decade of life.[2] Linear nail-plate growth decreases by 50 % over a lifetime, with periods of slow decline alternating with periods of rapid decline in approximate 7-year increments.[7] Nail-plate growth increases during pregnancy and decreases during lactation, after use of chemotherapeutic agents, and in conditions characterized by limb paralysis, persistently diminished circulation, or malnutrition.[2,3,8] Yellow nail syndrome is characterized by very slow or absent growth of nail plate; it usually affects both fingernails and toenails and is seen in association with several underlying conditions, such as lymphedema, respiratory disorders (e.g., bronchiectasis and pleural effusions), and nephrotic syndrome.[9]

Constitutional age-related findings in the nail plate include changes in nail color and luster, longitudinal ridging, changes in convexity, and brittle nails.[8,10] Nail plates, especially of toenails, often develop a yellow or gray color with a dull, opaque appearance. Longitudinal ridging may affect some or all nails and may present as slightly indented grooves or projection ridges or as beading. Over time, the surface of the nail plate may become flattened (platyonychia) or spooned (koilonychia). Temporary koilonychia, especially of the toenails, is also seen in infants.[6]

Pathophysiology Affecting the Nail Plate

Brittle nails is a common complaint; its incidence is 20% in the overall population (27% in female patients) and increases with advancing age.[10] When nail water content falls below 16%, nail plates become brittle; when the water content rises above 25%, nail plates become soft. The most common cause of brittle nails is dehydration, which can be caused or exacerbated by external factors such as use of nail-polish remover or exposure to dry climate. Onychoschizia, which presents as a layered, superficial splitting of the nail plate, may increase in incidence with age. This condition is seen much more frequently in female patients. It is likely related to recurrent exposures to water or irritants, such as during nail-care procedures.

Fingernails demonstrate a tendency to become thinner and more fragile over time. Toenails usually become thicker and harder. Onychogryphosis is a marked thickening, usually of the large toenail, resulting in a compacted mass of heaped-up dystrophic nail plate.[8] Contributing factors appear to be advanced age, poor nail care, chronic trauma, decreased peripheral circulation, and neuropathy. Poor-fitting shoewear causes long-term exposure to lateral pressure and friction, resulting in gryphotic changes (marked thickening or heaping of nail plate), usually of the first and fifth toenails.

NAIL FOLDS

The nail folds are the cutaneous soft tissue that houses the nail unit, invaginating proximally and laterally to encompass the emerging nail plate. The proximal nail fold, with the exception of the lunula, covers the underlying matrix and is devoid of sebaceous glands and dermatoglyphic skin markings.[11] The term paronychia describes inflammation of the nail folds. Paronychia may be acute or chronic and may occur secondary to a variety of conditions, including contact dermatitis, psoriasis, bacterial infection, and fungal infection.[12,13] The cuticle (eponychium) is a thin, keratinized membrane of modified stratum corneum that extends from the distal portion of the nail fold, reflecting onto the nail-plate surface. Intact cuticle serves as a seal that protects the space between the nail folds and the nail plate from exposure to external irritants, allergens, and pathogens. Loss of cuticle allows for exposure and trapping of these deleterious external agents, providing an environment in which either inflammatory or infectious paronychia can develop.

Nail Findings Associated with Disease States

Several nail findings have been associated with both underlying systemic and dermatologic conditions. The following is a review of selected, recognized associations. Diagnosis is based on proper evaluation of clinical findings; treatment is based on a confirmed etiology or the recognition of an underlying systemic association.

Special care must be taken when performing biopsy of the nail bed or matrix to avoid trauma to the tissue specimen and surrounding structures upon specimen removal. The most appropriate plane of dissection during nail-bed or nail-matrix biopsy is subdermal. The sampled tissue should be manipulated very gently throughout the biopsy procedure to avoid crush artifact, which may interfere significantly with histopathologic evaluation. It is also important to carefully dissect along the undersurface of the specimen, ensuring nontraumatic separation of the biopsy tissue from its underlying firm attachment to bone.

SPLINTER HEMORRHAGES

Splinter hemorrhages may be secondary to trauma, high altitude, primary dermatoses (i.e., psoriasis), or several underlying conditions (e.g., arterial emboli, collagen vascular disease, or thromboangiitis obliterans). The simultaneous appearance of splinter hemorrhages in several nails should raise suspicion of a possible underlying systemic disorder, especially in female patients.[14]

KOILONYCHIA

Koilonychia may be found in association with other conditions, including congenital conditions, iron deficiency anemia, cardiac disease, endocrinopathy, occupational exposures, and trauma.[3,15]

TRANSVERSE NAIL-PLATE DEPRESSIONS (BEAU LINES)

Beau lines present as well-delineated, transverse depressions in the nail plate. They are believed to occur secondary to temporary growth arrest of the nail matrix. The grooves become evident weeks after the occurrence of an abrupt, stressful event, such as an acute febrile illness. The width of the groove reflects the duration of interrupted nail-matrix function. When limited to one or a few digits, Beau lines may be associated with trauma, carpal tunnel syndrome, or Raynaud disease, or they may occur subsequent to tourniquet application during hand surgery.[15] Approximately 1 to 2 months after birth, infants may demonstrate physiologic Beau lines, which mark the transition from intrauterine to extrauterine life.[16] Multiple transverse grooves (stepladder appearance) may be seen in association with repeated cycles of chemotherapy, or they may be related to zinc deficiency. Multiple Beau lines should not be confused with the multiple transverse depressions that are stacked longitudinally along the central nail plate (washboard nails), resulting from the obsessive habit of repeatedly pushing back the cuticle or picking at the proximal nailfold margin (habit-tic deformity) [*see Figure 2*].[15,17] There is no specific treatment for Beau lines. They grow out over time after resolution of the growth-arrest period.

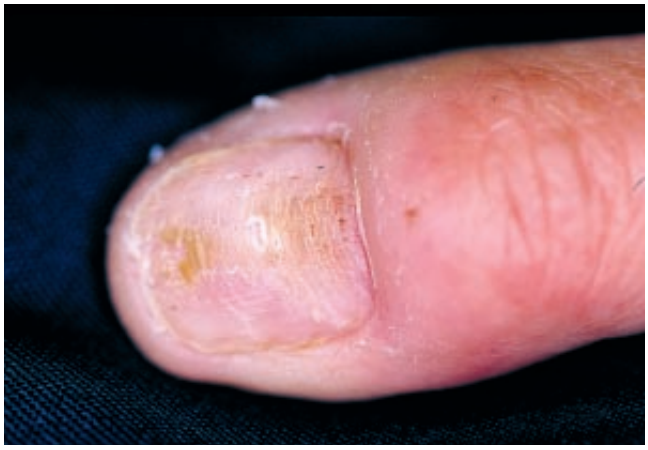


Figure 2 **Stacking of transverse linear grooves traversing the entire length of the central nail plate, resulting from the repeated picking of the proximal nail fold margin (habit-tic deformity). Note the marked hypertrophy of the lunula, which is typical of this disorder.**

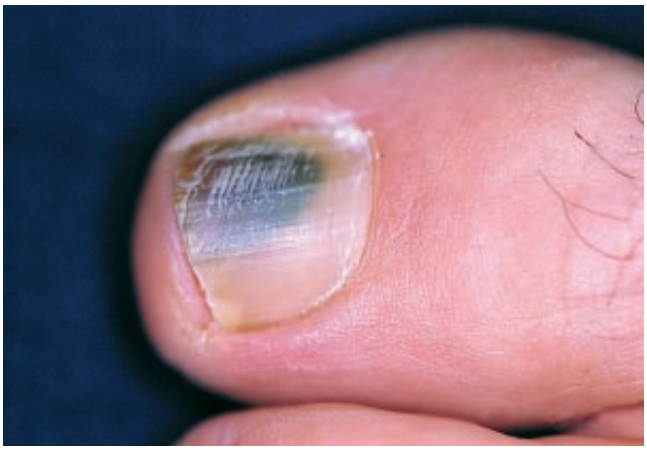


Figure 3 **Colonization of the closed space between the nail bed and nail plate with *Pseudomonas aeruginosa,* causing a green nail. Moisture trapped in the onycholytic space provides an optimal environment for proliferation of this bacterium.**

ONYCHOLYSIS

Onycholyis is defined as the separation of the nail plate from the nail bed. In most cases, onycholysis begins distally; it is often related to acute or chronic trauma that produces a lever effect, lifting the nail plate upward and away from its bed. Other causes of onycholysis are chemical exposure (allergic or irritant dermatitis), onychomycosis, and primary dermatoses (e.g., psoriasis or lichen planus).[13] Associations with underlying systemic disease (e.g., thyroid disease) have been sporadically reported but are less commonly encountered in clinical practice. When moisture accumulates under onycholytic nail plate, bacterial proliferation may occur. This can cause a green discoloration of the nail plate as a result of a pigment produced by certain organisms (e.g., *Pseudomonas aeruginosa*) [*see Figure 3*].

Treatment requires avoidance of precipating factors for onycholysis, debridement of separated nail plate, and the twice-daily topical application of diluted acetic acid solution (consisting of equal parts white vinegar and water), gentamicin, or the combination of polymyxin B and bacitracin.[18]

LEUKONYCHIA

Leukonychia, a white discoloration of the nail plate or subungual tissue, has multiple presentations. Small 1 to 3 mm white spots (punctate leukonychia) or irregular transverse streaks (leukonychia variegata) of the nail plate are the most common varieties.[15] These two presentations are generally secondary to repeated microtrauma to the matrix, growing out distally with outgrowth of the nail plate. Mee lines specifically refer to transverse 1 to 2 mm white bands, which usually are demonstrated at the same site in multiple nails and reported in association with arsenic intoxication, Hodgkin disease, sickle cell anemia, renal failure, and cardiac insufficiency. Leukonychia is also associated with systemic infection and chemotherapy.[19-21]

Half-and-half nails (Lindsay nails) present as a diffuse, dull whitening of the proximal nail bed that obscures the lunula and as a distal region of pink or reddish-brown discoloration that occupies from 20% to 60% of the nail length.[15,22] The most commonly reported association with half-and-half nails is chronic renal failure. When the distal brown band of discoloration constitutes less than 20% of the total nail length, the anomaly is known as Terry nails, which occurs in association with chronic congestive heart failure, hepatic cirrhosis, type 2 (non–insulin-dependent) diabetes mellitus, and advanced age. In both half-and-half nails and Terry nails, the proximal portion of the nail bed may be light pink, exhibiting a more normal appearance, rather than white.

Muehrcke nails present as paired, white, narrow transverse bands of the nail bed, separated by normal-appearing thin pink bands.[15,22] Muehrcke nails have been associated with chronic hypoalbuminemia. Resolution of this nail finding correlates with normalization of serum albumin levels.[15]

CLUBBING

When the normal angle between the proximal nail fold and the nail plate exceeds 180°, digital clubbing is present. The morphologic changes of clubbing typically include hypertrophy of the surrounding soft tissue of the nail folds as a result of hyperplasia of dermal fibrovasculature and edematous infiltration of the pulp tip.[21] Radiologic changes are identified in fewer than 20% of cases.[15]

Clubbing may be hereditary, or it may be seen in association with several underlying disease states, such as hypertrophic pulmonary osteoarthropathy, chronic congestive heart failure, congenital heart disease associated with cyanosis, polycythemias associated with hypoxia, Graves disease, chronic hepatic cirrhosis, lung cancer, Crohn disease, and irritable bowel disease.[15,22,23] When clubbing is unilateral, consideration should be given to underlying causes of obstructed circulation, such as aneurysm, arteriovenous fistula, and a pulmonary sulcus tumor (Pancoast tumor); disorders producing soft tissue edema; and diseases causing localized changes in underlying digital bone (e.g., sarcoidosis). Unilateral clubbing can also be found in cases of hemiplegia,[24] and a case of subungual perineurioma caused by unilateral clubbing has been reported.[25] Paronychia and distal phalangeal resorption may cause changes that simulate true clubbing (pseudoclubbing).

NAIL-PLATE PITTING

Nail-plate pitting (onychia punctata) develops as a result of focal defects in nail-plate formation from the proximal nail matrix. The number, size, and shape of the superficial depressions may vary.[15] The extent and duration of involvement with nail pitting correlates with the duration of nail-matrix abnormality. Psoriasis, the most common association with nail pitting, may produce a random array of shallow or deep pitted indentations, usually affecting one or more fingernails.[26,27]

Psoriasis of the nails often responds poorly to treatment, and it tends to recur. Topical corticosteroids, topical tazarotene, and intralesional corticosteroid injection may help in some cases.[28,29] It is a common misconception that nail pitting is pathognomonic for psoriasis.[27] Nail pitting may also be seen in association with alopecia areata, punctate keratoderma, idiopathic trachyonychia, occasionally in normal nails, and rarely in association with collagen vascular disease or syphilis. Fingernail pitting occurs in one third of children with alopecia areata; mild disease involving only a few nails is observed in approximately 20% of cases.[27] Compared to psoriasis, nail pitting seen in alopecia areata is typically more uniform and patterned, often presenting as orderly rows of shallow pitted depressions. Currently, there is no available treatment for this type of nail pitting.

LONGITUDINAL PIGMENTED BANDS

Longitudinal pigmented bands (melanonychia striata), also referred to as longitudinal melanonychia, is the presence of single or multiple longitudinally oriented brown or black bands [*see Figure 4*]. Homogeneous longitudinal bands occur in approximately 75% of African Americans older than 20 years. It usually affects the thumb and index finger.[30,31] Melanonychia striata is also commonly seen in Hispanics, may be found in up to 20% of Japanese, and is rare in whites.[30]

The deposition of melanin in the nail plate may result from increased melanin synthesis by matrix melanocytes that are usually nonfunctional; it may also occur as a result of a proliferation of matrix melanocytes.[30] Melanonychia striata affecting a single nail may result from a benign melanocytic nevus or a subungual melanoma. Thus, it is important to distinguish between a benign cause and a malignant cause of a longitudinal pigmented band. Factors suggesting the presence of melanoma or an atypical melanocytic proliferation are (1) single digit involvement; (2) periungual spread of pigment onto the nail-fold region (Hutchinson sign); (3) border irregularity or variegated color within the linear streak; and (4) changes in appearance (e.g., color or borders) involving an established longitudinal band.[32] Because of the severity of subungual melanoma and the importance of making a prompt diagnosis, the index of suspicion must be high. A simple biopsy of the nail plate is not satisfactory in establishing the diagnosis, because it will only demonstrate the presence of melanin. An appropriate biopsy inclusive of the nail matrix, as well as the nail bed if clinically indicated, should be performed by a surgeon who is familiar with the intricacies of performing a nail biopsy.[33] Because of limited experience and the difficulties that are commonly con-

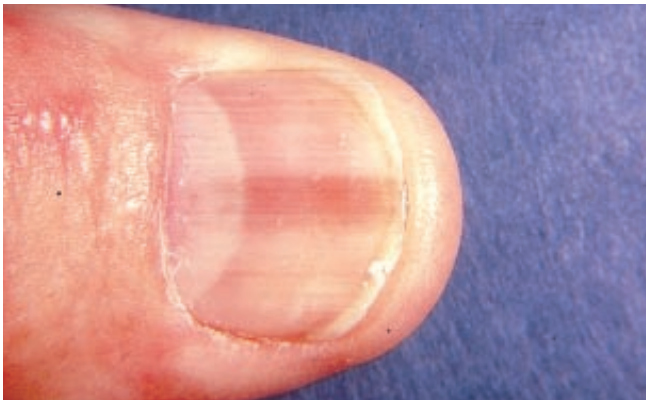


Figure 4 **Melanonychia striata (longitudinal pigmented band) produced by a melanocytic nevus of the nail matrix. A high index of suspicion for subungual melanoma is very important when a longitudinal pigmented band of the nail is identified.**

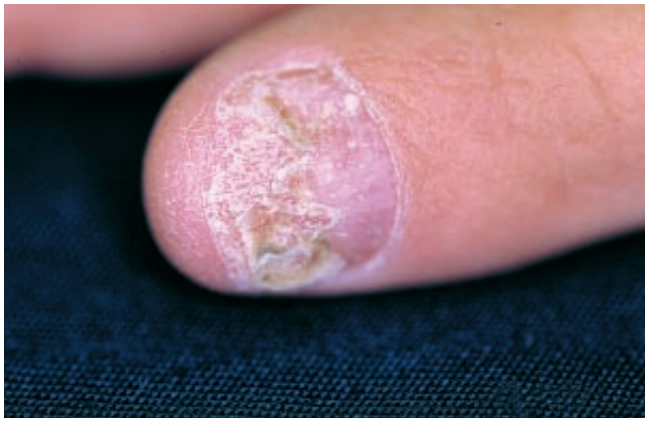
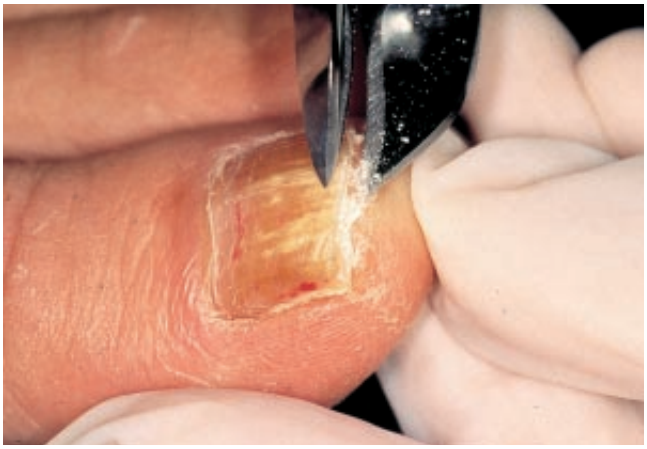


Figure 5 **Psoriasis of the nail, characterized by subungual hyperkeratosis and loss of distal onycholytic nail plate. This patient was unsuccessfully treated with oral antifungal therapy after an erroneous diagnosis of onychomycosis was made on the basis of clinical diagnosis alone. Careful examination of the proximal intact nail plate reveals pitting, a feature characteristic for psoriasis and not onychomycosis.**

a



b

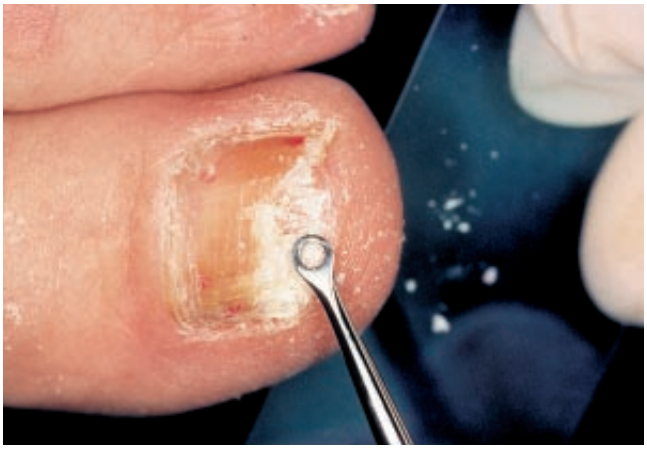


Figure 6 **(*a*) When obtaining a nail specimen for potassium hydroxide (KOH) preparation, it is important to expose the affected nail bed by first trimming away and discarding the distal, separated (onycholytic) nail plate. (*b*) Small specimen fragments of subungual hyperkeratosis of the nail bed and exposed undersurface of the nail plate are effectively obtained using a small curette. The smaller fragments are more easily dissolved by KOH, allowing for more accurate microscopic visualization, and can be easily plated on fungal culture medium.**

fronted in the histologic interpretation of nail specimens, biopsies of melanonychia striata are best interpreted by a dermatopathologist.[34]

When nail-bed pigmentation is noted, other causes such as systemic drugs (e.g., antimalarials, zidovudine, bleomycin, doxorubicin, minocycline, and hydroxyuria) or systemic disease (e.g., Addison disease and HIV infection) must be considered; however, these causes usually result in a broader, more diffuse pigmentation, often involving multiple nails.[35] Another reported association with melanonychia striata is systemic lupus erythematosus.[36] Frictional melanonychia resulting from trauma from athletic activities or poorly fitting shoewear may cause nail pigmentation, including pigmentation of the nail fold, especially in dark-skinned persons (pseudo-Hutchinson sign).[32]

Bacterial and Fungal Nail Infections

BACTERIAL PARONYCHIA

Bacterial infection of the nail folds (bacterial paronychia) is usually acute in nature. It is is characterized by swelling, erythema, discomfort, and sometimes purulence. The most common etiologic pathogen is *Staphylococcus aureus*. Treatment requires drainage of a focal abscess, if present, and oral antibiotic therapy.[37]

CHRONIC PARONYCHIA

Chronic paronychia results from persistent or frequently recurrent nail-fold inflammation, which is usually the result of chronic irritant dermatitis and loss of cuticle from trauma or nail-care practices. Secondary candidal infection may occur.[29,38]

Table 1 Oral Antifungal for Toenail Onychomycosis*[46]

Drug	*Dosage*	*Comments*
Griseofulvin tablets or liquid	500 mg – 1 g daily × 12 – 18 mo	Generally not recommended because of limited efficacy and because more effective agents are available; only active against dermatophyte organisms
Itraconazole capsules	Pulse therapy†: 200 mg twice daily × 1 wk/mo for 3 consecutive mo Continuous therapy: 200 mg daily × 3 mo	Contraindications include specific drug interactions and congestive heart failure; potential hepatotoxicity (rare); effective for dermatophytes, *Candida* species, and some nondermatophytic molds; should be administered with food; absorption may be decreased by increased gastric pH (as might result from use of H_2 blockers, antacids, proton pump inhibitors); blood clearance in 1 – 2 wk; therapeutic nail levels 9 mo posttherapy
Terbinafine tablets	250 mg daily × 3 mo	Most active for dermatophytes; some efficacy for certain nondermatophytic molds; limited activity against most *Candida* species; potential hepatotoxicity (rare); sporadic reports of blood dyscrasias (rare); reversible change or loss of taste (< 2%); blood clearance in 1 – 2 mo; therapeutic nail levels 9 mo posttherapy
Fluconazole tablets‡	150 – 300 mg × 9 – 12 mo	Effective against dermatophytes and *Candida* species; potential hepatotoxicity (rare); some significant drug interactions; limited therapeutic drug reservoir in nail posttherapy

*Topical ciclopirox 8% nail lacquer is FDA approved for onychomycosis caused by *Trichophyton rubrum*. Treatment involves application once daily for 12 mo (or until outgrowth of clear nail occurs), combined with debridement/trimming of onycholytic nail plate. Efficacy is lower than that seen with newer oral agents (e.g., itraconazole, terbinafine). No oral or topical agent is currently FDA approved for nondermatophytic onychomycosis (e.g., *Candida* species, molds).

†Pulse itraconazole is FDA approved for fingernail tinea unguium; established efficacy has been demonstrated for toenail disease.

‡Fluconazole is not FDA approved for onychomycosis; established efficacy has been demonstrated for tinea unguium.

Table 2 Selected Dermatologic Disorders Affecting the Nail Unit

Disease State	*Disease Features*	*Nail Findings*
Inflammatory diseases		
Psoriasis	Nail findings in 10% – 50% of patients; 39% of children with psoriasis with nail changes (usually pitting); nail disease present in 50% – 85% of patients with psoriatic arthritis	Proximal matrix involvement: pitting, transverse grooving, deeply ridged plate surface (onychorrhexis) Distal matrix involvement: plate thinning, lunula erythema Nail bed: subungual hyperkeratosis, oil drop sign, splinter hemorrhages Nail folds: cutaneous lesions of psoriasis Phalangeal/joint involvement: psoriatic arthritis
Lichen planus	Nail changes occur in up to 10% of patients with lichen planus; may occur in childhood or adulthood; nail involvement may be present with or without skin or mucosal disease; potentially reversible in early inflammatory stage; irreversible in cicatricial (later stage) of disease; may present as ridged, rough-surfaced, lusterless plates (trachyonychia) or 20-nail dystrophy in children	Matrix involvement: combination of nail-plate ridging, splitting, and progressive uniform thinning; distal-edge splitting, fragility, crumbling, brittleness, nail-plate shedding (onychomadesis) Focal matrix scarring: pterygium formation (scarring bridge between proximal nail fold and subungual epidermis with focal loss of nail plate) Nail-bed involvement: subungual hyperkeratosis, onycholysis Diffuse matrix/nail-bed disease: total nail-plate loss, atrophy, scarring
Alopecia areata	Nail changes in 10% of patints with alopecia areata; nail changes in over 40% of children with alopecia areata; fingernail involvement most common; may present in children as 20-nail dystrophy	Matrix involvement: orderly nail pitting arranged in a crosshatched pattern (glen-plaid sign); roughened nail-plate surface (trachyonychia); fragility; splitting; longitudinal ridging; spotted or red lunula (erythema); nail-plate shedding (onychomadesis)
Nail tumors		
Glomus tumor	75% occur on the hand, usually subungual (nail bed); a benign vascular hamartoma	Visible through plate as a light-red, reddish-blue spot; rarely exceeds 1 cm in size; characteristic symptom of intense or pulsatile pain; pain is spontaneous or provoked by slight trauma or pressure
Digital myxoid (mucus) cyst	A form of focal mucinosis; not a true cyst (no epithelial lining); contains clear, viscous, jellylike fluid; usually seen in adults	Soft, domed, translucent, pink or skin-colored, shiny, soft neoplasm of proximal nail fold or overlying distal interphalangeal joint; those over fold may compress matrix, producing flattening of plate; those over joint may connect to underlying joint space
Subungual exostoses	Outgrowths of calcified cartilage or normal bone; most seen on great toe; most frequent in adolescents and young adults; benign lesions	Emerge from the dorsal digit at distal phalanx; may erode through plate or project from under distal or lateral edge of plate; often painful; may become eroded
Periungual angiofibromas	Arise out of nail fold; often multiple; seen in 50% of cases of tuberous sclerosis (Borneville-Pringle disease); usually arise in early teenage years; benign neoplasm	Small, round, flesh-colored or pink, firm papules with shiny, smooth surface arising from nail-fold region; may partially cover nail plate; usually asymptomatic

ONYCHOMYCOSIS

Onychomycosis, the most common infection of the nail, is a fungal infection characterized by nail-bed and plate involvement. Dermatophyte onychomycosis (tinea unguium) is the most common type of fungal nail infection.[39] It is seen far more commonly in adults than in children and most frequently affects one or more toenails. The mode of fungal invasion usually presents as distal-lateral subungual onychomycosis, occurring as dermatophyte organisms migrate from pedal skin to below the nail plate and invade nail-bed tissue.[40] Tinea pedis and onychomycosis frequently coexist in a patient.[41,42]

The dermatophytes that most commonly cause onychomycosis are *Trichophyton rubrum* and *T. mentagrophytes*.[43] The tendency to harbor dermatophytes (especially *T. rubrum*), predominantly on pedal skin, has been noted in some kindreds. As a result, patients with such a tendency are prone to tinea pedis, tinea unguium, tinea cruris, and diffuse tinea corporis. They may present with dermatophyte infections earlier in life than usually seen and often experience recurrence of dermatophyte infection after completion of initially effective therapy.

The most characteristic clinical features of dermatophyte onychomycosis are distal onycholysis, subungual hyperkeratosis, and a dystrophic, discolored nail plate.[42] Because this combination of features is also seen in persons with nail psoriasis, accurate diagnosis may require performance of a potassium hydroxide (KOH) preparation and fungal culture [*see Figure 5*]. It is important that specimens be obtained from the nail bed [*see Figure 6*] and that culture specimens be transported and plated appropriately, because different culture media are required for identification of dermatophyte and nondermatophyte fungal nail pathogens.[42] Dermatophyte test medium (DTM) may be used as an in-office culture technique that has no special incubation requirements. DTM is inexpensive and accurate in the diagnosis of dermatophyte onychomycosis.[44] The clinical presentation of proximal white subungual onychomycosis, another presentation of dermatophyte onychomycosis, has been reported in association with systemic immunosuppression, including HIV disease.[45]

Candida onychomycosis is far less common than dermatophyte onychomycosis. *Candida* onychomycosis is often associated with immunosuppression (e.g., HIV disease and chronic mucocutaneous candidiasis). The *Candida* organisms may invade

the nail as a secondary pathogen, and they more frequently affect the fingernails.[42] Nondermatophyte molds, including *Aspergillus* species, *Scopulariopsis brevicaulis*, *Fusarium* species, *Scytalidium hyalinum,* and *Scytalidium dimidiatum,* have been reported to cause fingernail or toenail infection; however, such infections are relatively uncommon.[42,46] Associated paronychia may be seen when nondermatophytic fungi cause onychomycosis. Effective therapy for onychomycosis includes the use of an oral antifungal agent [*see Table 1*].[47] Because nails grow slowly, clinical response is delayed.[46] Infections with *Scytalidium* species are rare in the United States, and such infections respond poorly to currently available antifungal agents.

DERMATOLOGIC DISORDERS AFFECTING THE NAIL

Complete reviews of dermatologic, systemic, neoplastic, and exogenous disorders affecting the nail are beyond the scope of this subsection. An overview of selected dermatologic disorders affecting the nail unit and their associated clinical findings is provided [*see Table 2*].

James Q. Del Rosso, D.O., participates in the speakers' bureaus and is a consultant for Allergan, Inc., Janssen Pharmaceutica Products, L.P., Dermik Laboratories, Inc., Novartis Pharmaceuticals Corp., and Medicis Co.

C. Ralph Daniel III, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Gonzalez-Serva A: Structure and function. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 12

2. Fleckman P: Basic science of the nail unit. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 44

3. Dawber RPR, De Berker DAR, Baran R: Science of the nail apparatus. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, England, 1994, p 5

4. Fleckman P, Allan C: Surgical anatomy of the nail unit. Dermatol Surg 27:257, 2001

5. Cohen PR: The lunula. J Am Acad Dermatol 34:943, 1996

6. Tosti A, Peluso AP, Piraccini BM: Nail diseases in children. Adv Dermatol 13:353, 1998

7. Orentreich N, Markofsky J, Vogelman JH: The effect of aging on the rate of linear nail growth. J Invest Dermatol 73:126, 1979

8. Cohen PR, Scher RK: Geriatric nail disorders: diagnosis and treatment. J Am Acad Dermatol 26:521, 1992

9. Tosti A, Baran R, Dawber RPR: The nails in systemic disease and drug-induced changes: yellow nail syndrome. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, England, 1994, p 185

10. Lubach D, Cohrs W, Wurzinger R: Incidence of brittle nails. Dermatologica 172:144, 1986

11. Baran R, Dawber RPR, Tosti A: Science of the nail apparatus and relationship to foot function. A Text Atlas of Nail Disorders. Baran R, Dawber RPR, Tosti A, Eds. Martin Dunitz, London, 1996, p 3

12. Daniel CR III, Daniel MO, Gupta AK: Nonfungal infections and paronychia. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 165

13. Kern D: Occupational disease. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 285

14. Tosti A, Baran R, Dawber RPR: The nails in systemic disease and drug-induced changes: splinter hemorrhages. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, England, 1994, p 183

15. Baran R, Dawber RPR: Physical signs. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, England, 1994, p 35

16. Baran R, Dawber RPR: Physical signs: Beau's lines and transverse grooves. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, 1994, p 50

17. Habif TP: Nail diseases: habit-tic deformity. Clinical Dermatology. Habif TP, Ed. Mosby, St Louis, 1996, p 774

18. Daniel CR III, Daniel MO, Gupta AK: Appendix 2. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 368

19. Naumann R, Wozel G: Transverse leukonychia following chemotherapy in a patient with Hodgkin's disease. Eur J Dermatol 19:392, 2000

20. Cribier B, Mena ML, Rey D, et al: Nail changes in patients infected with human immunodeficiency virus: a prospective controlled study. Arch Dermatol 134:1216, 1998

21. Mautner GH, Lu I, Ort RJ, et al: Transverse leukonychia with systemic infection. Cutis 65:318, 2000

22. Daniel CR III, Sams WM, Scher RK: Nails in systemic disease. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 219

23. Myers KA, Farquhar DR: The rational clinical examination: does this patient have clubbing? JAMA 286:341, 2001

24. Siragusa M, Schepis C, Cosentino FI, et al: Nail pathology in patients with hemiplegia. Br J Dermatol 144:557, 2001

25. Baran R, Perrin C: Subungual perineurioma: a peculiar location. Br J Dermatol 146:125, 2002

26. Farber EM, Nall ML: Nail psoriasis. Cutis 50:174, 1992

27. Del Rosso JQ, Basuk P, Scher RK, et al: Dermatologic diseases of the nail unit. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 172

28. Scher RK, Stiller M, Zhu YI: Tazarotene 0.1% gel in the treatment of fingernail psoriasis: a double-blind, randomized, vehicle-controlled study. Cutis 68:355, 2001

29. Tosti A, Piraccini BM: Treatment of common nail disorders. Dermatol Clin 18:339, 2000

30. Baran R, Haneke E: Tumors of the nail apparatus and adjacent tissues: longitudinal melanonychia. Diseases of the Nails and Their Management. Baran R, Dawber RPR, Eds. Blackwell Science, Oxford, England, 1994, p 485

31. Haneke E, Baran R: Longitudinal melanonychia. Dermatol Surg 27:580, 2001

32. Baran R, Dawber RPR, Tosti A: Nail colour changes (chromonychia). A Text Atlas of Nail Disorders. Baran R, Dawber RPR, Tosti A, Eds. Martin Dunitz, London, 1996, p 147

33. Salasche SJ: Surgery. Dermatologic diseases of the nail unit. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 335

34. Fleckman P, Omura EF: Histopathology of the nail. Adv Dermatol 17:385, 2001

35. Aste N, Fumo G, Contu F, et al: Nail pigmentation caused by hydroxyurea: report of 9 cases. J Am Acad Dermatol 47:146, 2002

36. Skowron F, Combemale P, Faisant M, et al: Functional melanonychia due to involvement of the nail matrix in systemic lupus erythematosus. J Am Acad Dermatol 47(suppl):S187, 2002

37. Habif T: Nail diseases: acute paronychia. Clinical Dermatology. Habif TP, Ed. Mosby, St. Louis, 1996, p 763

38. Van Laborde S, Scher RK: Developments in the treatment of nail psoriasis, melanonychia striata, and onychomycosis: a review of the literature. Dermatol Clin 18:37, 2000

39. Gupta AK, Taborda P, Taborda V, et al: Epidemiology and prevalence of onychomycosis in HIV-positive individuals. Int J Dermatol 39:746, 2000

40. Elewski BE: Onychomycosis: treatment, quality of life, and economic issues. Am J Clin Dermatol 1:19, 2000

41. Lauritz B: Dermatoses of the feet. Am J Clin Dermatol 1:181, 2000

42. Elewski BE, Charif MA, Daniel CR III: Onychomycosis. Nails: Therapy, Diagnosis, Surgery. Scher RK, Daniel CR, Eds. WB Saunders Co, Philadelphia, 1997, p 152

43. Jennings MB, Weinberg JM, Koestenblatt EK, et al: Study of clinically suspected onychomycosis in a podiatric population. J Am Podiatr Med Assoc 92:327, 2002

44. Pariser D, Opper C: An in-office diagnostic procedure to detect dermatophytes in a nationwide study of onychomycosis patients. Manag Care 11:43, 2002

45. Baran R, Dawber RPR, Tosti A: Onychomycosis and its treatment. A Text Atlas of Nail Disorders. Baran R, Dawber RPR, Tosti A, Eds. Martin Dunitz, London, 1996, p 157

46. Del Rosso JQ: Current management of onychomycosis and dermatomycoses. Curr Infect Dis Rep 2:438, 2000

47. Crawford F, Young P, Godfrey C, et al: Oral treatments for toenail onychomycosis: a systematic review. Arch Dermatol 138:811, 2002

Acknowledgment

Figure 1 Tom Moore.

XV DISORDERS OF PIGMENTATION

Pearl E. Grimes, m.d.

Disorders of Hyperpigmentation

melasma

Definition

Melasma is a common acquired symmetrical hypermelanosis characterized by irregular light-brown to gray-brown macules involving the face. There is a predilection for the cheeks, forehead, upper lips, nose, and chin [*see Figure 1*]. Lesions may occasionally occur in other sun-exposed areas, including the forearms and back.[1-3]

Epidemiology

Melasma is most commonly observed in females. Men constitute only 10% of the cases but usually demonstrate the same clinicopathologic features as women do. The condition affects all racial and ethnic groups but is most prevalent in persons with darker complexions (skin types IV through VI). It is also more common in geographic areas with intense ultraviolet radiation (sunlight), such as tropical and subtropical regions.

Etiology and Pathogenesis

Although the precise cause of melasma is unknown, multiple factors have been implicated in the etiology and pathogenesis of this condition. These factors include genetic influences, intense ultraviolet radiation exposure, pregnancy, oral contraceptive use, hormone replacement therapy, cosmetics, and phototoxic and antiseizure medications.[1]

Endocrinologic studies of patients with melasma have shown varying results. Although a detailed study of nine women with melasma showed significantly increased levels of luteinizing hormone (LH) and low levels of estradiol, suggesting a role for subclinical mild ovarian dysfunction, a study of 26 women found no difference in LH, follicle-stimulating hormone (FSH), and α-melanocyte–stimulating hormone (α-MSH) levels between patients with melasma and control subjects.[3] Another study reported increased expression of α-MSH in the affected skin areas of 10 women with melasma. α-MSH stimulates tyrosinase and melanin synthesis.[4]

Clinically, the light-brown patches are commonly evident on the malar prominences, forehead, chin, nose, and upper lip. The patches may have a malar, centrofacial, or mandibular distribution. Histologically, an epidermal, epidermal-dermal, or dermal pattern of increased pigmentation occurs. Histologic studies document an increase in epidermal pigmentation, increased numbers of melanocytes, and increased activity of melanogenic enzymes.[5] A Wood-light examination enhances the epidermal pattern of pigment deposition. Such epidermal lesions are most amenable to treatment.

The differential diagnosis of melasma includes other conditions that cause facial hyperpigmentation, such as postinflammatory hyperpigmentation, drug-induced hyperpigmentation, lichen planus actinicus, and photosensitivity disorders.

Treatment

Current treatments for melasma include broad-spectrum sunscreens, 2% (over the counter) and 4% (prescription) hydroquinone formulations, azelaic acid, kojic acid formulations, α-hydroxyacid products, retinoic acid, retinol, superficial chemical peels, and microdermabrasion.[1,6-9] A triple-combination product containing 4% hydroquinone, 0.01% fluocinolone, and 0.05% retinoic acid has enhanced efficacy. This product was previously compounded by pharmacists, but it is now available as a commercial preparation (Tri-Luma cream, Galderma Laboratories) that has been approved by the Food and Drug Administration for treatment of melasma. Because the combination contains a fluorinated steroid, treatment should be limited to 8 weeks. Laser therapy offers minimal long-term success and, instead, may worsen the condition. Intense pulsed light therapy may offer some improvement in patients with melasma.[10,11]

Although all of these therapies reduce the severity of melasma, none are curative. Hence, it is essential for patients to rigidly adhere to a regimen of daily sun protection (e.g., using sunscreen or wearing protective clothing) to control the progression of melasma.

postinflammatory hyperpigmentation

Definition

Postinflammatory hyperpigmentation is characterized by an acquired increase in cutaneous pigmentation secondary to an inflammatory process [*see Figure 2*]. Excess pigment deposition may occur in the epidermis or in both the epidermis and the dermis.

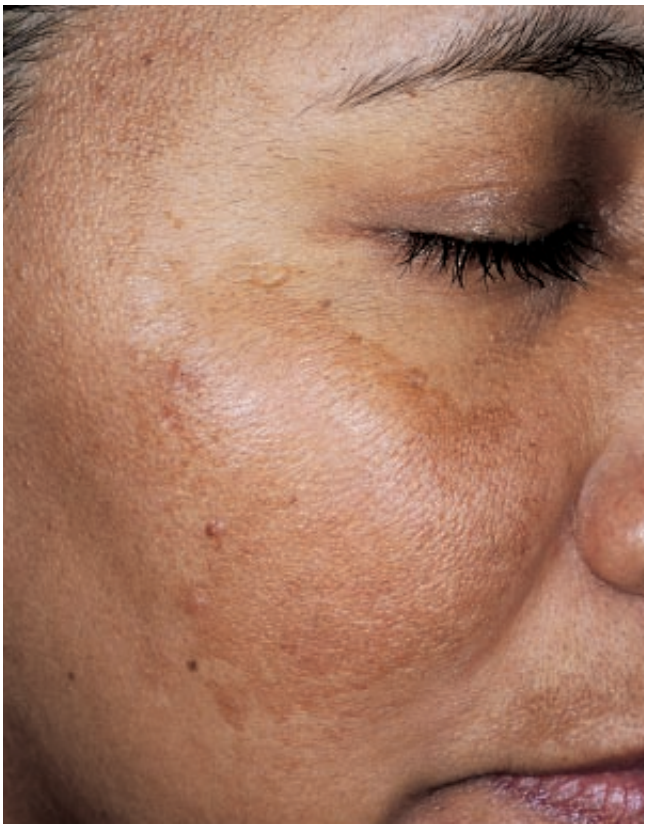


Figure 1 **Melasma is characterized by hyperpigmentation of the cheek, forehead, and upper lip.**

Epidemiology

All racial and ethnic groups are susceptible to postinflammatory hyperpigmentation, but the incidence of the condition is higher in persons with darker complexions. In a diagnostic survey of 2,000 African-American patients seeking dermatologic care, the third most common diagnosis was pigmentary disorders, of which postinflammatory hyperpigmentation was the most prevalent.[1]

Etiology and Pathogenesis

Pigmentary changes may be a result of production of inflammatory mediators and altered cytokine production.[12,13] Such changes may lead to an increase in the number and size of epidermal melanocytes. In addition, hyperpigmentation may be a consequence of pigmentary incontinence, with deposition of pigment in the upper dermis. Postinflammatory hyperpigmentation may be a sequela of conditions such as acne, allergic reactions, drug eruptions, papulosquamous disorders, eczematoid disorders, and vesiculobullous disorders.[14]

Diagnosis

Clinically, postinflammatory pigmentary changes may be localized, circumscribed, or generalized. Lesions range in color from brown to black to ashen gray and usually follow the distribution of the primary dermatosis.

Treatment

Therapies for postinflammatory hyperpigmentation include over-the-counter and prescription hydroquinone preparations. Higher concentrations are indicated for moderate to severe involvement. Other treatments include azelaic acid, kojic acid, retinoic acid [*see* Melasma, *above*], and adapalene.[15]

DRUG-INDUCED HYPERPIGMENTATION

Medications are a common cause of cutaneous hyperpigmentation. Lesions may be localized or generalized. Medications can also cause hyperpigmentation of the oral mucosa and nails. There may be some improvement upon withdrawal of the offending agent; however, drug-induced hyperpigmentation can persist for many years.

Medications causing drug-induced hyperpigmentation include oral contraceptives, hormone replacement therapies, antibiotics, antidepressants, antiviral agents, antimalarials, antihypertensives, and chemotherapeutic agents. Such medications include progesterone, estrogen, zidovudine (AZT), minocycline, tetracycline, bleomycin, hydrochlorothiazide, hydantoin, amiodarone, chlorpromazine, quinacrine, hydroxychloroquine, chloroquine, imipramine, amitriptyline, diltiazem, citalopram, hydroxyurea, doxorubicin, busulfan, daunorubicin, cisplatin, cyclophosphamide, thiotepa, vinblastine, and vincristine.[16-23]

Heavy-metal preparations can also cause hyperpigmentation. These preparations include arsenic, gold, silver, mercury, and bismuth.

Treatment with the Q-switched alexandrite laser has proven to be an effective treatment for drug-induced pigmentation.[19,24]

ERYTHEMA DYSCHROMICUM PERSTANS

Definition

Erythema dyschromicum perstans (EDP, or ashy dermatosis) is an acquired benign condition characterized by slate-gray to violaceous macules. It was first described in 1957.

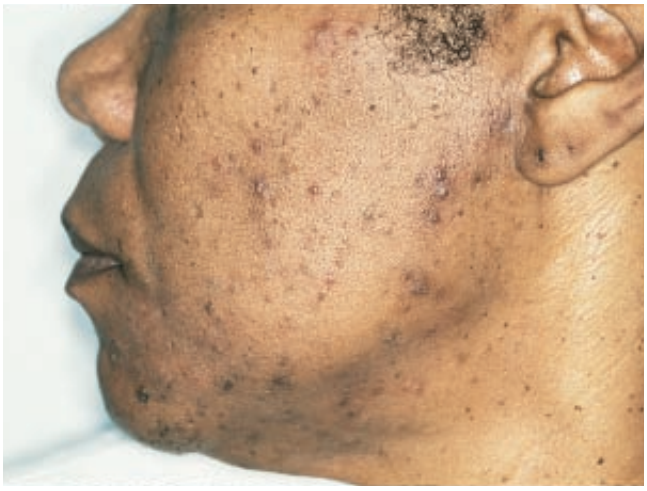


Figure 2 **Postinflammatory hyperpigmentation of the face may be secondary to acne vulgaris.**

Epidemiology

EDP is reported most commonly in dark-skinned persons. However, cases have been reported globally and in all skin types. The disease appears to have a relatively equal frequency in men and women. It has also been reported in children.

Etiology and Pathogenesis

The precise cause of EDP is unknown. Studies suggest that pollutants, pesticides, hair dyes, chemicals, and drug exposure may play a role in the pathogenesis.[23-28] Findings in light microscopic, ultrastructural, and immunofluorescent studies of EDP have been similar to those in studies of lichen planus, leading some investigators to postulate that EDP may be a variant of lichen planus. Other studies suggest that EDP is a distinct entity. Expression of intercellular adhesion molecule–I (ICAM-I) and major histocompatibility complex (MHC) class II molecules (HLA-DR) has been reported.[29] These findings suggest that aberrant cell-mediated immunity may be involved in the pathogenesis of EDP.

Diagnosis

Clinically, the macules of EDP are ashen and may have an erythematous, slightly raised border during the early stages of the disease. Erythematous macules have also been described during the early stages. Areas of erythema eventually resolve, leaving slate-gray areas of pigmentation. The lesions are usually symmetrically distributed and vary in size from small macules to very large patches. Common sites of involvement include the face, neck, trunk, and upper extremities. Mucous membranes, palms, soles, and nails are usually spared. Light microscopic findings are slight epidermal atrophic changes, spongiosis, lymphocytic exocytosis, and basal vacuolopathy in the epidermis, as well as lymphohistiocytic, lichenoid dermal infiltrates. In later stages, the lesions lack the epidermal changes and show increased deposition of dermal pigment.

Postinflammatory hyperpigmentation, idiopathic eruptive macular pigmentation, pityriasis rosea, lichen planus, fixed drug eruption, Addison disease, pinta, syphilis, macular amyloidosis, hemochromatosis, and argyria must be distinguished from EDP.

Treatment

Therapies for EDP have been minimally effective. They include sunscreens, hydroquinone, topical corticosteroids, systemic steroids, griseofulvin, clofazamine, antibiotics, and antimalarials.[30]

LENTIGINES

Definition

A lentigo is a well-circumscribed, brown to brown-black macule, usually less than 1 cm in size, that appears at birth or in early childhood. Lentigines occur in all skin types and may be found on any cutaneous surface, including the palms, soles, and mucous membranes. They do not darken with sun exposure. Lentigines can be localized and must be distinguished from freckles (ephelides). Clinical differentiating features include the later appearance of freckles (at 4 to 6 years of age) and their predominance on sun-exposed skin and increased frequency in redheads and fair-skinned persons. Freckles also tend to fade in winter and with advancing age.

Epidemiology

Multiple lentigines have been reported in 18.5% of black newborns and 0.04% of white newborns. Solar lentigines have been reported in 90% of whites older than 60 years.

Diagnosis

Several types of lentigines are recognized, including lentigo simplex, solar lentigines, nevus spilus, lentigines induced by psoralens plus ultraviolet A (PUVA), generalized lentiginosis, and syndrome-related lentiginosis.

Lentigo simplex lesions may occur as solitary localized macules or may be numerous and widespread. They often occur during the first decade of life and can be found on any cutaneous surface.

Solar (senile) lentigines, or so-called liver spots, are brown macules that appear late in adult life on chronically sun-exposed skin. These lesions are present in 90% of whites older than 70 years and occur in response to solar exposure. Solar lentigines correlate with the tendency to freckle and with two or more sunburns after 20 years of age.

Nevus spilus, or speckled lentiginous nevus, is a congenital brown patch on which dotted brown macules develop during childhood. Histologically, the brown patch has features of a lentigo, whereas the dotted brown macules most often reveal features of junctional nevi. Zosteriform patterns have also been described. Generalized lentiginosis is characterized by innumerable lentigines unassociated with other abnormalities.

Syndromes characterized by multiple lentiginosis include multiple lentigines (LEOPARD [multiple lentigines, electrocardiographic conduction abnormalities, ocular hypertelorism, pulmonary stenosis, abnormal genitalia, retardation of growth, sensorineural deafness]) syndrome, Moynahan syndrome, centrofacial lentiginosis, Carney complex, Laugier-Hunziker disease, Peutz-Jeghers syndrome, and Bannayan-Ruvalcaba-Riley syndrome.[31,32]

The histopathology of lentiginosis shows elongated rete ridges, increased numbers of basal melanocytes, and increased basal melanization. In contrast, freckles result from hypermelanization of basal melanocytes without a concomitant increase in number.

Lentigo must be distinguished from other flat, pigmented lesions, including freckles, junctional nevi, postinflammatory hyperpigmentation, and pigmented actinic keratoses.

Treatment

The treatment of lentigines includes hydroquinone-containing bleaching agents, cryotherapy, Q-switched neodymium:yttrium-aluminum-garnet (Nd:YAG) laser, and intense pulsed light.[33,34]

CONFLUENT AND RETICULATED PAPILLOMATOSIS OF GOUGEROT AND CARTEAUD

Definition

The eruption of confluent and reticulated papillomatosis was initially described by Gougerot and Carteaud in 1927 and 1932. The condition consists of 2 to 5 mm hyperpigmented papules that have a predilection for the sternal area and midline of the back and neck.

Epidemiology

Confluent and reticulated papillomatosis occurs in equal frequency in men and women and shows no racial or ethnic predilections. The disease usually begins during the third decade of life.

Etiology and Pathogenesis

The precise cause of confluent and reticulated papillomatosis is unknown. Abnormal host response to *Pityrosporum orbiculare* (*Malassezia furfur;* the fungus that causes tinea versicolor), abnormal response to colonization by follicular bacteria, and genetically determined defects of keratinization have been suggested.[35,36] The disease has been associated with Cushing syndrome, diabetes, hypopigmentation, and thyroid disorders.

Diagnosis

Patients present with 2 to 5 mm hyperpigmented, slightly verrucoid papules that have a predilection for the back, scapula, and inframammary areas. The papules become confluent near the midline and possess a reticulated pattern near the periphery. The lesions do not form a true scale but, rather, a mealy deposit that can be easily removed with the fingertips.

Histologically, studies show hyperkeratosis, decreased granular cell layers, papillomatosis, absence of sweat glands, and fragmentation of elastic fibers. Electron microscopic studies have shown increased numbers of transitional cells between the stratum granulosum and stratum corneum. This finding suggests premature keratinization. In addition, increased expression of keratin 16 has been reported, suggesting abnormal proliferation, differentiation, or both.[36]

Other conditions that simulate confluent and reticulated papillomatosis are tinea versicolor and acanthosis nigricans.

Treatment

Minocycline is reportedly beneficial.[37] Other treatments that have shown some efficacy include selenium sulfide shampoo, salicylic acid, urea, vitamin A, corticosteroids, calcipotriol, tetracycline, erythromycin, doxycycline, retinoids, and PUVA.[38-41]

DOWLING-DEGOS DISEASE

Definition

Dowling-Degos disease, or reticulated pigmented anomaly of the flexures, is an autosomal dominant disorder with variable penetrance characterized by brownish-black macules of the flexures that develop in a reticulated pattern. It may be caused by an underlying defect in follicular epithelial proliferation.

Diagnosis

Dowling-Degos disease presents as symmetrical, reticulated hyperpigmentation of the groin, axilla, antecubital area, inframammary areas, and neck.[42] The lesions begin as 1 to 3 mm macules that gradually become confluent, assuming a reticulated lacelike pattern. In addition, perinasal and facial involvement is common. Pigmented pinhead-sized comedones are frequently observed in the affected areas, and perinasal, pitted acneiform scars can occur around the mouth.

Lesions of Dowling-Degos disease begin in early adult life and are slowly progressive. The condition has been reported in association with reticulated acropigmentation of Kitamura and hidradenitis suppurativa,[43] suggesting an underlying defect in follicular epithelial proliferation. In addition, the disease has been reported in a large kindred with reticulate acropigmentation of Kitamura and acropigmentation of Dohi, suggesting an association between and overlap of these conditions.[44] Histologically, thin, pigmented epithelial strands of downgrowth extend from the epidermis and follicular wall in a filiform pattern resembling adenoid seborrheic keratoses.[42,45]

Treatment

In general, there is no effective treatment for Dowling-Degos disease. Adapalene and the erbium:YAG laser have been reported to offer some benefit.[46]

Disorders of Hypopigmentation

VITILIGO

Definition

Vitiligo is a common acquired, idiopathic skin disorder characterized by one or more patches of depigmented skin. The depigmentation results from loss of cutaneous melanocytes. These lesions are cosmetically disfiguring and usually cause emotional trauma in both children and adults [*see Figure 3*].

Epidemiology

Vitiligo affects 1% to 2% of the population. Onset may begin at any age, but peak incidence is in the second or third decade of life. The disease shows no racial or ethnic predilection, but because of the stark contrast between depigmented and darker skin tones, it is more cosmetically disfiguring in darker racial and ethnic groups. Females are affected more often than males. The disease has a familial incidence of 25% to 30%. Genetic studies suggest a polygenic inheritance pattern.

HLA studies have reported increases in a variety of haplotypes of class I and class II antigens in patients with vitiligo. However, results vary significantly by race and ethnicity of the population studied. The reported HLA associations include increased frequencies of HLA A30, CW6, CW7, DR1, DR3, DR4, and DQW3.[47]

Etiology and Pathogenesis

The precise cause of vitiligo is unknown. Multiple theories have been proposed, including genetic, autoimmune, neural, biochemical, and viral mechanisms. Reviews addressing the etiology of vitiligo suggest that vitiligo is probably a heterogeneous disease encompassing multiple etiologies.[47,48]

An immune-mediated pathogenesis is the most popular theory. This theory is predicated on the increased frequency of a plethora of immunologic diseases in patients with vitiligo, including hypothyroidism (Hashimoto thyroiditis), Graves disease, pernicious anemia, diabetes mellitus, and alopecia areata. Thyroid diseases are the most common associated diseases. Other disorders reported in association with vitiligo include Addison disease, atopic dermatitis, asthma, lichen planus, morphea, lichen sclerosus et atrophicus, mucocutaneous candidiasis, biliary cirrhosis, myasthenia gravis, Down syndrome, AIDS, and cutaneous T cell lymphoma.

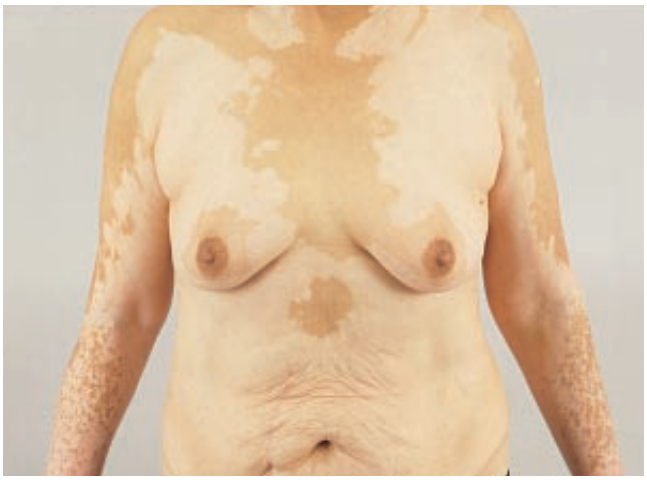


Figure 3 **Vitiligo is indicated by generalized patches of depigmentation of the trunk.**

Humoral and cell-mediated immunologic defects are a common phenomenon in vitiligo.[47,48] Numerous studies have documented an increased frequency of organ-specific autoantibodies. Antithyroid, gastric antiparietal cell, and antinuclear antibodies are most commonly demonstrated. Patients with positive organ-specific autoantibodies unassociated with autoimmune disease have an increased risk of subsequent subclinical or overt autoimmune disease.

Antimelanocyte antibodies, often demonstrated in the sera of patients with vitiligo, induce the destruction of cultured melanocytes by complement-mediated lysis and antibody-dependent cellular cytotoxicity. The presence and titer of antimelanocyte antibodies correlate with the severity and activity of vitiligo. These antibodies are directed against melanocyte cell surface antigens with molecular weights of 25, 35 to 40, 75, 90, and 150 kd. Studies suggest that the antimelanocyte antibody may mediate the destruction of melanocytes in vitiligo. Tyrosinase antibodies have also been reported in patients with localized and generalized disease.[49]

Cellular immune-mediated defects include diminished contact sensitization and quantitative and qualitative alterations in T cells and natural-killer cells. Skin-homing cytotoxic T cells have also been implicated in the destruction of melanocytes. Immunohistochemical studies have demonstrated abnormal expression of MHC class II and ICAM-I by melanocytes in vitiligo, which may contribute to the aberrant cellular immune response. In addition, there is increased expression of the antiadhesive matrix component tenasin in perilesional and lesional vitiliginous skin. Increased tenasin expression may be a consequence of elevated cytokine production and cellular infiltrates in vitiligo.[47] Studies have documented alterations in cytokine production in patients with vitiligo. Studies of affected skin showed a significantly lower expression of granulocyte-macrophage colony-stimulating factor

(GM-CSF), basic fibroblast growth factor (bFGF), and stem cell factor.[50] In contrast, expression of interleukin 6 (IL-6) and tumor necrosis factor–α (TNF-α) was greater in lesional skin than in perilesional or normal skin. Another study reported increased expression of TNF-α, interferon gamma, and IL-10 in the lesional and adjacent skin of vitiligo patients.[51]

Cytomegalovirus DNA has been demonstrated in the involved and uninvolved skin of patients with vitiligo. No viral DNA was detected in matched control subjects.[52] These findings suggest that in some cases, vitiligo may be triggered by a viral infection.

The neural theory is supported by several clinical, biochemical, and ultrastructural observations. These observations include the occurrence of segmental vitiligo; the demonstration of lesional autonomic dysfunction, such as increased sweating; and the demonstration of nerve ending–melanocyte contact. The last observation is rare in normal skin.

Several studies suggest that oxidative stress may be the initial event in the destruction of melanocytes.[53,54] Defective recycling of tetrahydrobiopterin, increased production of hydrogen peroxide, and decreased catalase have been demonstrated in the skin of patients with vitiligo.[55,56] In addition, lesional catecholamine biosynthesis and release are increased. Thus, abnormal release of catecholamines from autonomic nerve endings and oxidative stress may damage melanocytes by altering the free radical defense of the epidermis.

The self-destruction hypothesis proposes that melanocytes may be destroyed by phenolic compounds formed during the synthesis of melanin. In vivo and in vitro studies have demonstrated the destruction of melanocytes by phenols and catechols. In addition, industrial workers who are exposed to catechols and phenols may experience depigmentation of areas of skin.

A variety of environmentally ubiquitous compounds containing catechols, phenols, and sulfhydryls can induce hypopigmentation, depigmentation, or both. These compounds are most often encountered in industrial chemicals and cleaning agents. Possible mechanisms for altered pigment production by these compounds include melanocyte destruction via free radical formation, inhibition of tyrosinase activity, and interference with the production or transfer of melanosomes.

Diagnosis

Clinical manifestations Vitiliginous lesions are typically asymptomatic depigmented macules without clinical signs of inflammation. However, inflammatory vitiligo with erythematous borders has been reported. Hypopigmented lesions may coexist with depigmented lesions. The patches are occasionally pruritic. Macules frequently begin on sun-exposed or perioral facial skin and either remain localized or disseminate to other cutaneous sites. Areas of depigmentation vary in size from a few millimeters to many centimeters, and their borders are usually distinct. Trichrome lesions are most often observed in darker-complexioned persons. These lesions are characterized by zones of white, light-brown, and normal skin color. Depigmented hairs are often present in lesional skin and do not preclude repigmentation of a lesion. In addition, there is a high incidence of premature graying of scalp hair in patients with vitiligo and in their families. Vitiliginous lesions can remain stable or can slowly progress for years. In some instances, patients undergo almost complete spontaneous depigmentation over a few years.

Vitiligo is subclassified into different types on the basis of the distribution of skin lesions. These subclassifications include the generalized or vulgaris, acral or acrofacial, localized, and segmented types. The generalized pattern is characterized by symmetrical macules or patches occurring in a random distribution. Acral or acrofacial vitiligo consists of depigmented macules confined to the extremities or to the face and extremities, respectively. A subcategory of the acrofacial type is the lip-tip variety, in which lesions are confined to the lips and the tips of the digits. The generalized and acrofacial varieties are the most common. Segmental vitiligo occurs in a dermatomal or quasidermatomal distribution; lesions rarely spread beyond the affected dermatome. This type is the less common variety of vitiligo and most often occurs along the distribution of the trigeminal nerve.

Melanocytes of the eye, ear, and leptomeninges may also be involved in vitiligo. Depigmented areas of the retinal pigment epithelium and choroid have been reported in 39% of patients studied. These lesions usually do not interfere with vision. Vitiligo is also a manifestation of the Vogt-Koyanagi-Harada syndrome, which is characterized by poliosis, chronic uveitis, alopecia, dysacusis, vitiligo, and signs of meningeal irritation. It usually begins in the third decade of life, and although no race is spared, the disease tends to be more severe in darker-complexioned races, especially Asians.

The syndrome has been divided into stages. The first, or meningeal, stage, is associated with headache, nausea, vomiting, fever, confusion, cranial nerve palsies, hemiparesis, and cerebrospinal fluid pleocytosis. Usually, there are a few neurologic sequelae. In the second stage, ophthalmic and auditory changes predominate, including photophobia, ocular pain, visual loss, anterior or posterior uveitis, and sometimes retinal detachment, tinnitus, and dysacusis. Cutaneous lesions are dominant in the third, or convalescent, stage, occurring as the uveitis begins to subside. Common features are vitiligo, which frequently involves the eyelids and periorbital region [*see Figure 4*]; poliosis of the scalp, hair, eyelashes, and eyebrows; and diffuse or patchy alopecia.

Patients with malignant melanoma frequently experience a vitiligolike depigmentation surrounding melanoma lesions and at distant sites. The presence of depigmentation in melanoma patients portends a longer survival.

Laboratory findings Histologically, the predominant finding in vitiligo is an absence of melanocytes in lesional skin. Light microscopy and ultrastructural studies have also revealed vacuolar degeneration of basal and parabasal keratinocytes and revealed epidermal and dermal lymphohistiocytic cell infiltrates.

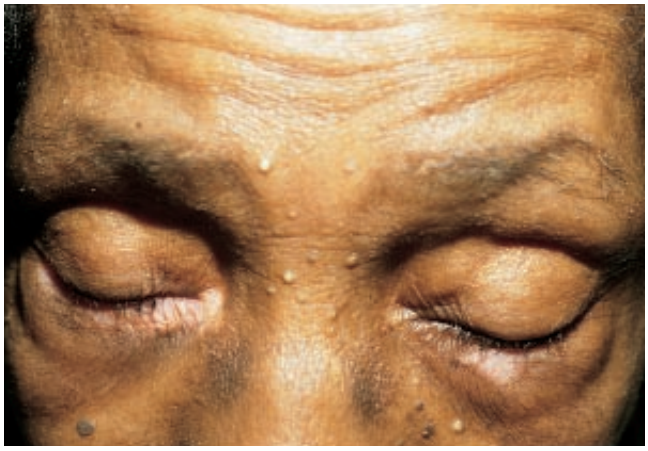


Figure 4 **A patient with Vogt-Koyanagi-Harada syndrome shows periorbital depigmentation.**

Immunohistochemical staining has confirmed the presence of a predominantly T cell infiltrate in vitiliginous and adjacent skin.

In view of the association of vitiligo with myriad other autoimmune diseases, the routine baseline evaluation of a patient should include a thorough history and physical examination. Recommended laboratory tests include a complete blood count; sedimentation rate; comprehensive metabolic panel, including liver function tests; and autoantibody tests (antinuclear antibody, thyroid peroxidase, and parietal cell antibodies).

Differential Diagnosis

Other disorders characterized by depigmentation may occasionally mimic vitiligo clinically. These include piebaldism, nevus depigmentosus, nevus anemicus, postinflammatory depigmentation or hypopigmentation, pityriasis alba, tinea versicolor, discoid lupus erythematosus, scleroderma, hypopigmented mycosis fungoides, and sarcoidosis. Therefore, in some instances, a skin biopsy may be necessary to substantiate a diagnosis of vitiligo.

Treatment Selection

Therapeutic objectives in vitiligo should include both stabilization of the disease and repigmentation of vitiliginous skin lesions. Repigmentation can be accomplished medically[57-59] or, in patients with localized stable lesions, surgically.[60] The choice of repigmentation therapies should be predicated on the age of the patient, extent of cutaneous surface involvement (severity), and activity or progression of the disease. The disease can be divided into four stages: limited (less than 10% involvement), moderate (10% to 25% involvement), moderately severe (26% to 50% involvement), and severe (greater than 50% involvement) [*see Table 1*].

Table 1 Therapeutic Approaches for Vitiligo

Stages I and II disease*	Topical steroids Topical photochemotherapy PUVA-sol In-office PUVA Bath photochemotherapy Pseudocatalase/UVB UVB phototherapy Narrow band Broad band Excimer laser Topical immunomodulators Tacrolimus Pimecrolimus L-phenylalanine/UV Topical khellin/UVA Melagenina Calcipotriol/PUVA Tar emulsions Vitamin supplementation Autologous melanocyte grafting (stable lesions)
Stages III and IV disease*	Oral photochemotherapy Systemic steroids (oral, I.M.) (for stabilization) Bath photochemotherapy UVB phototherapy Narrow band Broad band Oral khellin/UVA L-phenylalanine/UV Immunomodulators Isoprinosine Levamasole Immunosuppressives Cyclosporine Cyclophosphamide Nitrogen mustard Depigmentation (severe, recalcitrant lesions)

*Stage I, < 10% involvement; stage II, 10%–25% involvement; stage III, 26%–50% involvement; stage IV, > 50% involvement.
PUVA—psoralens plus ultraviolet A UV—ultraviolet UVA—ultraviolet A UVB—ultraviolet B

Medical Treatment

Medical therapies for vitiligo include topical and systemic steroids, topical and systemic PUVA, narrow-band ultraviolet light therapy (UVB), excimer laser therapy, nutritional vitamin supplementation, immunomodulators, calcipotriol, phenylalanine, and khellin.[57-59]

Steroids Mid- to high-potency steroids are indicated in patients with limited involvement. Low-potency topical steroids are usually ineffective. Topical mid- to high-potency steroids can be used safely for 2 to 3 months, then interrupted for 1 month or tapered to low-potency preparations. Patients must be closely monitored for topical steroid side effects, which include skin atrophy, telangiectasias, hypertrichosis, and acneiform eruptions. Since the introduction of topical immunomodulators (tacrolimus and pimecrolimus), topical steroids are used less often in vitiligo patients.

Short courses of oral prednisone for 1 to 2 weeks or intramuscular triamcinolone acetonide injections, 40 mg/month for 2 to 3 months, are often extremely helpful for stabilizing rapidly progressive vitiligo. However, prolonged use of systemic steroids is not indicated.[57-59]

Photochemotherapy Until recently, topical and systemic PUVA therapies were the mainstay for repigmenting vitiliginous lesions.[57,58] However, in the past several years, these therapies have been overshadowed by new ones, including narrow-band UVB phototherapy, lasers, and topical immunomodulators.

Topical photochemotherapy can be administered in the office or outside the office in combination with sunlight. The choice of topical PUVA is predicated on the severity of vitiligo, patient lifestyle, and convenience for the patient. Topical in-office PUVA is appropriate for patients with less than 20% cutaneous surface involvement. A thin coat of 0.01% to 0.1% methoxsalen ointment is applied to affected areas 30 minutes before UVA exposure. Treatments are weekly or twice weekly. For patients with less than 10% involvement, an alternative approach involves the use of 0.001% methoxsalen ointment applied 30 minutes before sunlight exposure. Patients are allowed to expose the affected areas for 10 minutes, gradually increasing exposure time to 30 minutes. Treatments are daily or every other day.

Oral photochemotherapy is indicated in patients with greater than 20% to 25% cutaneous surface involvement. The standard dose of 8-methoxypsoralen (8-MOP) is usually 0.3 to 0.4 mg/kg ingested 1.5 hours before UVA exposure. The treatments are administered twice weekly. Broad-spectrum sunscreen protection is essential after PUVA treatments. In addition, because of the ocular pharmacokinetics of 8-MOP, protective UVA sunglasses should be worn indoors and outdoors for 18 to 24 hours after ingestion of 8-MOP.

Contraindications to oral PUVA treatment include liver disease and photosensitivity disorders. Side effects include headaches, nausea, vomiting, xerosis, pruritus, photoaging, diffuse

hyperpigmentation, and hypertrichosis. Compared with topical PUVA, the major advantages of oral PUVA include its effectiveness in controlling the progression of active disease and its lower frequency of blistering reactions. Oral PUVA therapy has been associated with an increase in nonmelanoma and melanoma skin cancer in patients with psoriasis. However, similar documentation has not been reported in patients with vitiligo.

Factors that portend enhanced PUVA-induced repigmentation include young age (children), patient motivation, maintenance of adequate lesional phototoxicity, and location of lesions. Maximal repigmentation occurs on the face and neck, and minimal responses occur in the hands and feet. Overall, mean repigmentation of 60% to 65% of the affected areas can be achieved.[58]

Narrow-band UVB Recent studies have reported the benefits of narrow-band UVB phototherapy (NB-UVB).[61] NB-UVB treatment was shown to be as effective as topical PUVA, with fewer side effects. In a study of NB-UVB phototherapy versus oral PUVA, 56% of the UVB group had greater than 25% repigmentation, compared with 63% of the oral PUVA group. The difference was not statistically significant. Because of its efficacy and safety profile, NB-UVB has emerged as the therapy of choice for patients with moderate to severe disease.

NB-UVB phototherapy offers several advantages over oral psoralen photochemotherapy, including ease of treatment, lack of need for posttreatment ocular protection, lack of the side effects (e.g., nausea, headaches, and gastritis) associated with oral methoxsalen, and minimal phototoxic reactions. Furthermore, NB-UVB phototherapy can be used to treat young children who have extensive, progressive vitiligo. Disadvantages include the need for more treatments for maximal efficacy (three times weekly for NB-UVB, compared with twice weekly for PUVA) and the lack of data concerning the possible long-term carcinogenic effects of NB-UVB phototherapy.

Dermatologists continue to treat patients with PUVA, and it remains the gold standard despite its inherent difficulties. Patients whose vitiligo does not respond to NB-UVB phototherapy are often switched to oral PUVA.

Repigmentation occurs gradually and requires many treatments: 16 to 24 treatments are usually needed for new pigment to become evident. In general, maximal repigmentation involves 6 to 12 months of NB-UVB or PUVA therapy.

Laser therapy The excimer laser (308 nm UVB), recently approved by the FDA for treatment of psoriasis, also shows promise as a therapy for vitiligo.[62,63] This laser can be used as monotherapy or in combination with other modalities. Laser therapy targets the lesional area and theoretically reduces UV exposure. In addition, because the laser provides a focused, high-intensity dose of NB-UVB, treatment duration, in theory, may be reduced. Long-term, controlled studies are needed to further define the efficacy, risks, and benefits of the excimer laser for treatment of vitiligo.

Pseudocatalase The beneficial effects of pseudocatalase and calcium applied twice daily and UVB exposure twice weekly have also been reported. The rationale for this therapy is derived from previous studies that demonstrated aberrant catalase and calcium homeostasis in patients.[64]

Vitamins Preliminary open-label studies have documented stabilization and repigmentation in vitiligo patients treated with high-dose vitamin supplementation, including daily doses of ascorbic acid (1,000 mg), vitamin B_{12} (1,000 μg), and folic acid (1 to 5 mg).[57]

Topical immunomodulators Abnormalities of both humoral and cell-mediated immunity have been well documented in patients with vitiligo,[47-52] which explains the apparent efficacy of several immunomodulators for this disease. Preliminary investigations have reported repigmentation of vitiliginous lesions with isoprinosine, levamisole, suplatast tosilate, and cyclosporine.[57]

Tacrolimus ointment is a novel topical immunomodulatory drug for treatment of adult and pediatric atopic dermatitis. Tacrolimus exerts its therapeutic effect via inhibition of the production of proinflammatory cytokines. Moderate to excellent repigmentation was reported in five of six patients treated with tacrolimus. Patients ranged in age from 6 to 32 years. Repigmentation responses did not correlate with disease duration.[58,65]

Calcipotriol Several studies have documented the efficacy of calcipotriol for repigmentation of vitiligo. Used in combination with UV exposure, calcipotriol was well tolerated and effective in both children and adults.[58] Melanocytes are thought to express 1α,25-dihydroxyvitamin D_3 receptors, which may play a role in stimulating melanogenesis.

Depigmentation Since the 1950s, monobenzylether of hydroquinone (MBEH, or monobenzene) has been used as a depigmenting agent for patients with extensive vitiligo. In general, MBEH causes permanent destruction of melanocytes and induces depigmentation locally and remotely from the sites of application. Hence, the use of MBEH for other disorders of pigmentation is contraindicated.

Depigmentation is a viable therapeutic alternative in patients with greater than 50% cutaneous depigmentation who have demonstrated recalcitrance to repigmentation or in patients with extensive vitiligo who have no desire to undergo repigmentation therapies.[55,58] The major side effects of MBEH therapy are dermatitis and pruritus, which usually respond to topical and systemic steroids. Other side effects include severe xerosis, alopecia, premature graying, and suppression of lymphoproliferative responses.

Surgical Treatment

Surgical treatment is appropriate for patients with localized, stable areas of vitiligo that have been recalcitrant to medical treatment.[60] Such approaches are contraindicated in patients with keloids or hypertrophic scars. Techniques for surgical grafting include suction blister grafts, punch grafts, sheet grafts, pure melanocyte cultures, and cocultures of melanocytes and keratinocytes. These techniques are indeed beneficial for localized lesions.

Micropigmentation is often associated with the induction of koebnerization; therefore, its use should be limited to treatment of mucous membrane lesions.

ALBINISM

Definition

Albinism is an uncommon, complex congenital disorder characterized by hypopigmentation of the hair, eyes, and skin. Albinism is generally subclassified as oculocutaneous albinism (OCA) and ocular albinism (OA); in the latter, reduction of melanin is limited to the eye.[66-71] Sometimes, different mutations in the same gene can cause OCA or OA.

Epidemiology

OCA has been reported by investigators in all mammalian orders and in all human ethnic groups. It is one of the most widely distributed genetic abnormalities in the animal kingdom. Human albinism has been noted throughout history. OCA is the most common inherited disorder of generalized hypopigmentation.

Etiology and Pathogenesis

Albinism may result from primary defects that are specific for the melanin synthetic pathway or from defects that are not specific for melanin synthesis. Mutations in seven genes have been reported to cause OCA or OA.[67,68] They include the tyrosinase gene (*OCA1* on chromosome 12q1), the oculocutaneous albinism gene (*OCA2*, a missed mutation of the *P* gene on 15q11), the tyrosinase-related protein 1 gene (*OCA3*), the *HPS* gene (Hermansky-Pudlak syndrome at 10q23 and mutations of the β3A-adaptin gene), the *CHS* gene (Chédiak-Higashi syndrome), and the *OA1* gene (X-linked ocular albinism).

Diagnosis

Clinically, the most severe disease is observed in OCA1A, which is OCA resulting from mutations in the tyrosinase gene. It is characterized by absent tyrosinase activity, which results in complete absence of melanin in the eyes, skin, and hair. There is no improvement with age. Affected individuals have marked photophobia, nystagmus, and profound sun sensitivity because of the inability to tan.

OCA1B, or yellow albinism, is less severe. Tyrosinase activity is low or absent, and pigmentation of the hair and skin improves with age. In contrast to OCA1A, pigmented freckles and lentigines develop with age.

OCA1-MP, or minimal-pigment OCA, is characterized by white skin and hair at birth. Iris pigment is present at birth, or it appears during the first decade of life. All reported cases have been in white persons. The tyrosinase gene mutation produces a less active enzyme.

Temperature-sensitive OCA (OCA1-TS) is characterized by white skin and hair and blue eyes at birth and by development of patterned pigmentation by puberty. Darker hair develops in cooler areas (extremities), and white hair is retained in warmer areas (axilla and scalp). The pattern results from a tyrosinase mutation that causes a temperature-sensitive enzyme.

OCA2, tyrosinase-positive OCA with normal tyrosinase activity, is the most common variety. The hair darkens with age, but the skin remains white. Pigmented nevi, lentigines, and freckles develop and are especially pronounced in sun-exposed areas. This type has recently been ascribed to mutation of the *P* gene, which encodes the tyrosinase-transporting membrane protein. The *P* gene is on chromosome 15q.

OCA3 encompasses the Rufous variety and some cases of brown albinism. Clinically, there is minimal pigment reduction in the hair, eyes, and skin.

The secondary varieties of albinism in which the primary defect is not specific for the melanin synthetic pathway include Hermansky-Pudlak syndrome,[71] Chédiak-Higashi syndrome, Cross-McKusick-Breen syndrome, Prader-Willi syndrome, and Angelman syndrome.

The autosomal recessive Hermansky-Pudlak syndrome is characterized by low to absent tyrosinase activity. The *HPS* gene has been mapped to chromosome 10q23.[71-73] Skin and hair color varies from white to light brown. Freckles and lentigines develop with age. Iris pigment correlates with hair and skin pigmentation. Affected individuals experience a hemorrhagic diathesis secondary to a platelet-storage-pool deficiency. Their platelets lack storage granules (i.e., sites of storage for serotonin, calcium, and adenine nucleotides). Ceroidlike deposits are present in macrophages of the bone marrow, lungs, liver, spleen, and gastrointestinal tract. These patients bruise easily and are subject to epistaxis and gingival bleeding. Pulmonary fibrosis and granulomatous colitis develop as a consequence of the ceroid deposits.

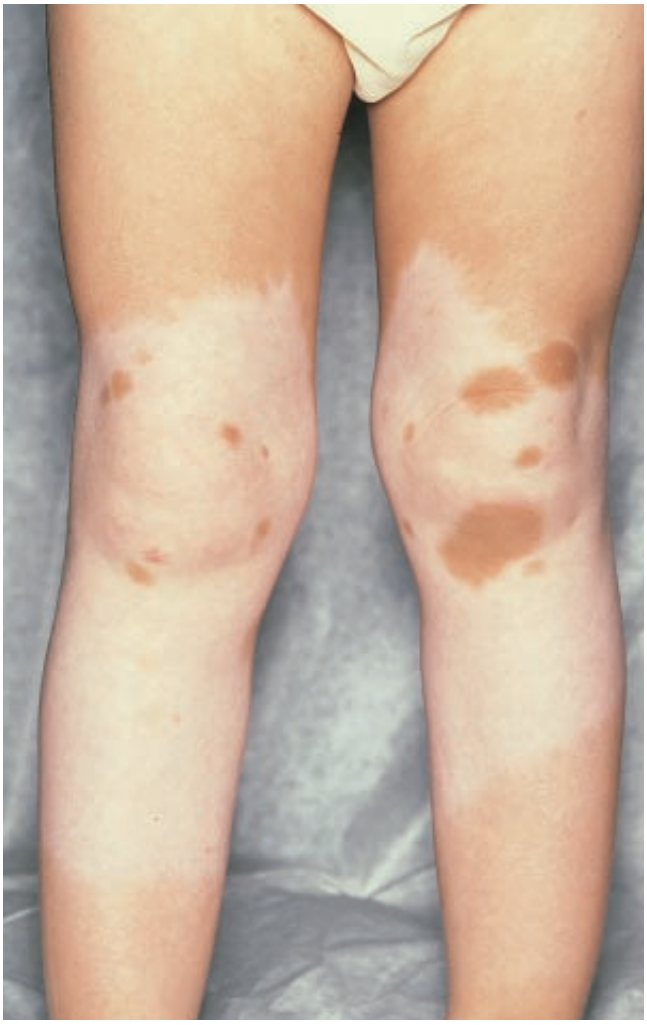


Figure 5 **A patient with piebaldism has the classic midextremity areas of depigmentation with islands of hyperpigmentation.**

Chédiak-Higashi syndrome consists of hypopigmentation, recurrent sinopulmonary bacterial infections, peripheral neuropathy, and giant lysosomal granules, with death occurring at an early age as a result of lymphoreticular malignancies. The *CHS* gene locus is on chromosome 1q29. Chédiak-Higashi syndrome must be distinguished from Griscelli syndrome, which is characterized by partial albinism, lymphohistiocytosis, immunodeficiency, neutropenia, and thrombocytopenia. Griscelli syndrome has been mapped to chromosome15q21, around the *myosin-Va* gene. However, the presence of giant lysosomal granules is pathognomonic for Chédiak-Higashi syndrome.[74,75]

Cross-McKusick-Breen syndrome includes hypopigmentation, microphthalmia, nystagmus, and severe mental and physical retardation.

Prader-Willi syndrome is a developmental syndrome characterized by mental retardation, neonatal hypotonia, and poor feeding, followed by hyperphagia and obesity later in life. Short stature, hypogonadism, and inappropriate emotional behavior constitute the syndrome. Fifty percent of patients have a deletion on the long arm of chromosome 15. Patients have ocular abnormalities and skin and hair hypopigmentation consistent with OCA.

Mutation of the *P* gene has been reported in Angelman syndrome and is also characterized by mental retardation, abnormal behavior, and hypopigmentation. The pattern of hypopigmentation is similar to that in Prader-Willi syndrome. In addition, Angelman syndrome is associated with a deletion on chromosome 15. However, in contrast to Prader-Willi syndrome, the deletion occurs on the maternal chromosome.

Treatment

The management of patients with albinism should include genetic counseling and patient education regarding the use of sunscreens and clothing for protection against ultraviolet radiation–induced damage. Magnifiers are beneficial for ocular symptoms.

Complications

The long-term consequences of albinism are solar keratoses and basal and squamous cell carcinomas. Malignant melanoma is uncommon.

PIEBALDISM

Definition

Piebaldism is a rare autosomal dominant congenital disorder of pigmentation. It is a stable leukoderma and is characterized by patches of white skin and white hair. The affected areas are principally the frontal scalp, forehead, ventral chest, abdomen, and extremities. A white forelock occurs in 80% to 90% of patients.

Epidemiology

Although rare, piebaldism occurs in all ethnic groups worldwide. Its estimated occurrence is one in 100,000 persons. It is found with equal frequency in males and females.

Etiology and Pathogenesis

Molecular genetics studies have shown that piebaldism results from mutations of the *KIT* proto-oncogene, which encodes the cell surface receptor tyrosine kinase for mast cell or stem cell factor located on chromosome segment 4q12. Mutations occur in the highly conserved tyrosinase domain of *KIT*. A number of different mutations in the *KIT* gene can cause piebaldism.[76-80] The locations of the *KIT* gene mutation correlates with severity of disease. Mutations of the intracellular tyrosine kinase domain are associated with the most severe phenotypes.[77] Reduced *KIT* function arrests the migration of melanocytes into affected hair follicles and epidermis during embryonal development.[76-78]

In general, patients with piebaldism are healthy and do not have associated systemic abnormalities. However, the disorder occasionally has been associated with heterochromia irides, mental retardation, osteopathia striata, Woolf syndrome, and Hirschsprung disease.

Diagnosis

Cutaneous depigmentation is the only manifestation of piebaldism in 10% to 20% of cases. Amelanotic macules are usually present on the ventral surface of the thorax and abdomen and extend to the back but spare the midline. Characteristic extremity lesions extend from midarm to wrist and occur on the midleg [*see Figure 5*]. White patches of the mucous membranes have also been reported. Hyperpigmented macules may appear within the areas of depigmentation.

Light and electron microscopic studies of the white macules have typically revealed an absence of melanocytes. However, melanocytes have been demonstrated in the white forelock and amelanotic skin of three patients studied.

Differential Diagnosis

Piebaldism is sometimes confused with vitiligo, but in piebaldism, the leukodermic patches are both congenital and relatively static in shape and size.

Treatment

The lesions of piebaldism are usually stable throughout life, although some patients have reported spontaneous repigmentation. In general, therapeutic approaches, including psoralen photochemotherapy and grafting, are unsatisfactory. Autologous melanocyte grafting procedures may offer some benefit for localized or limited areas of involvement.

IDIOPATHIC GUTTATE HYPOMELANOSIS

Definition

Idiopathic guttate hypomelanosis (IGH) is a common asymptomatic disorder characterized by hypopigmentation and depigmented polygonal macules ranging from approximately 2 to 8 mm in diameter.

Epidemiology

IGH appears to be a very common, benign dermatosis. It occurs in all races, with a frequency ranging from 46% to 70%, but is more prevalent in darker-skinned racial and ethnic groups. Macules may begin to appear during the third or fourth decade of life and gradually increase in number thereafter.

Etiology and Pathogenesis

The precise pathogenesis has not been established for IGH. Long-term sun exposure, trauma, genetic influences, and aging, with a gradual loss of melanocytes, have been implicated in the pathogenesis of this disorder.[81]

Diagnosis

The lesions of IGH are macules that are punctate to polygonal in shape, 2 to 8 mm in diameter, and hypopigmented to depigmented. They are most commonly observed on the lower extremities. There is no atrophy or change in the overlying skin. Histologic evaluation of lesions reveals hyperkeratosis, epidermal atrophy, and decreased epidermal melanin. Melanocytes may be normal or decreased. Immunoperoxidase studies show a markedly reduced number of melanocytes. Melanocyte differentiation appears to be unaffected.[82]

Differential Diagnosis

IGH must be differentiated from other hypopigmentary disorders, such as vitiligo, scleroderma, leukodermic guttate parapsoriasis, tinea versicolor, hypopigmented sarcoidosis, pityriasis alba, chemical depigmentation, and postinflammatory hypopigmentation.

Treatment

No definitive treatment is currently available. Patients often need reassurance regarding the banality of lesions. For patients concerned about the cosmetic appearance of lesions, clinicians have used camouflage, intralesional steroids, and topical photochemotherapy. Localized superficial dermabrasion may offer some improvement.[83]

The author has received grants for clinical research from Allergan Inc., Fujisawa Healthcare, Inc., and Galderma Laborotories, L.P., during the past 12 months.

References

1. Grimes PE: Melasma: etiologic and therapeutic considerations. Arch Dermatol 131:1453, 1995

2. O'Brien TJ, Dyall-Smith D, Hall AP: Melasma of the forearms. Australas J Dermatol 38:35, 1997

3. Boissy RE, Nordlund JJ: Molecular basis of congenital hyperpigmentary disorders in humans: a review. Pigment Cell Res 10:12, 1997

4. Im S, Kim J, On WY, et al: Increased expression of alpha-melanocyte-stimulating hormone in the lesional skin of melasma. Br J Dermatol 146:165, 2002

5. Kang WH, Yoon KH, Lee ES, et al: Melasma: histopathological characteristics in 56 Korean patients. Br J Dermatol 146:228, 2002

6. Piamphongsant T: Treatment of melasma: a review with personal experience. Int J Dermatol 37:897, 1998

7. Grimes PE: The safety and efficacy of salicylic acid peels in darker-racial ethnic groups. Dermatol Surg 25:18, 1999

8. Grimes PE: Agents for ethnic skin peeling. Dermatol Ther 13:159, 2000

9. Rajan P, Grimes PE: Skin barrier changes induced by aluminum oxide and sodium chloride microdermabrasion. Dermatol Surg 28:390, 2002

10. Negishi K, Tezuka Y, Kushikata N: Photorejuvenation for Asian skin by intense pulsed light. Dermatol Surg 27:627, 2001

11. Gerardo A, Arias M, Ferrando J: Intense pulsed light for melanocytic lesions. Dermatol Surg 27:397, 2001

12. McKenzie RC, Sauder DN: The role of keratinocyte cytokines in inflammation and immunity. J Invest Dermatol 95(6 suppl):1055, 1990

13. Kinbauer R, Kock A, Neuner P, et al: Regulation of epidermal cell interleukin-6 production by UV light and corticosteroids. J Invest Dermatol 96:484, 1991

14. Ruiz-Maldonado R, Orozco-Covarrubias ML: Postinflammatory hypopigmentation and hyperpigmentation. Semin Cutan Med Surg 16:36, 1997

15. Jacyk WK: Adapalene in the treatment of African patients. J Eur Acad Dermatol Venereol 15(suppl 3):37, 2001

16. Granstein R, Sober AJ: Drug- and heavy metal–induced hyperpigmentation. J Am Acad Dermatol 5:1, 1981

17. Lerner EA, Sober AJ: Chemical and pharmacologic agents that cause hyperpigmentation or hypopigmentation of the skin. Dermatol Clin 6:327, 1988

18. Ozog DM, Gogstetter DS, Scott G, et al: Minocycline-induced hyperpigmentation in patients with pemphigus and pemphigoid. Arch Dermatol 136:1133, 2000

19. Green D, Friedman KJ: Treatment of minocycline-induced cutaneous pigmentation with the Q-switched Alexandrite laser and a review of the literature. J Am Acad Dermatol 44(2 suppl):342, 2001

20. Mineg ME, Bhawan J, Stefanato CM, et al: Imipramine-induced hyperpigmentation: four cases and a review of the literature. J Am Acad Dermatol 40:159, 1999

21. Koppel RA, Boh EE: Cutaneous reactions to chemotherapeutic agents. Am J Med Sci 321:327, 2001

22. Alley E, Green R, Schuchter L: Cutaneous toxicities of cancer therapy. Curr Opin Oncol 14:212, 2002

23. Scherschun L, Lee MW, Lim HW: Diltiazem-associated photodistributed hyperpigmentation: a review of 4 cases. Arch Dermatol 137:179, 2001

24. Atkin DH, Fitzpatrick RE: Laser treatment of imipramine-induced hyperpigmentation. J Am Acad Dermatol 43:77, 2000

25. Combemale P, Faisant M, Guennoc B, et al: Erythema dyschromicum perstans: report of a new case and critical review of the literature. J Dermatol 25:747, 1998

26. Dominguez-Soto L, Hojyo-Tomoka T, Vega-Memije E, et al: Pigmentary problems in the tropics. Dermatol Clin 12:777, 1994

27. Penagos H, Jimenez V, Fallas V, et al: Chlorethalanil: a possible cause of erythema dyschromicum perstans (ashy dermatitis). Contact Derm 35:214, 1996

28. Spiewak R: Pesticides as a cause of occupational skin diseases in farmers. Ann Agric Environ Med 8:1, 2001

29. Baranda L, Torres-Alvarez B, Cortes-Franco R, et al: Involvement of cell adhesion and activation molecules in the pathogenesis of erythema dyschromicum perstans (ashy dermatitis): the effect of clofazimine therapy. Arch Dermatol 133:32, 1997

30. Osswald SS, Proffer LH, Sartori CR: Erythema dyschromicum perstans: a case report and review. Cutis 68:25, 2001

31. Stratakis CA: Genetics of Peutz-Jeghers syndrome, Carney complex, and other familial lentiginoses. Horm Res 54:334, 2000

32. Abdelmalek NF, Gerber TL, Menter A: Cardiocutaneous syndromes and associations. J Am Acad Dermatol 46:161, 2002

33. Todd M, Rallis TM, Gerwels JW, et al: A comparison of 3 lasers and liquid nitrogen in the treatment of solar lentigines: a randomized, controlled, comparative trial. Arch Dermatol 136:841, 2000

34. Kawada A, Shiraishi H, Asai M, et al: Clinical improvement of solar lentigines and ephelides with an intense pulsed light source. Dermatol Surg 28:504, 2002

35. Lee MP, Stiller MJ, McClain SA, et al: Confluent and reticulated papillomatosis: response to high-dose oral isotretinoin therapy and reassessment of epidemiologic data. J Am Acad Dermatol 31:327, 1994

36. Inaloz H, Patel GK, Knight AG: Familial confluent and reticulated papillomatosis. Arch Dermatol 138:276, 2002

37. Purg L, de Moragas JM: Confluent and reticulated papillomatosis of Gougerot and Carteaud: minocycline deserves trial before etretinate. Arch Dermatol 131:109, 1995

38. Angeli-Besson C, Koeppel MC, Jacquart SP, et al: Confluent and reticulated papillomatosis (Gougerot-Carteaud) treated with tetracyclines. Int J Dermatol 34:567, 1995

39. Carrozzo AM, Gatti S, Ferranti G, et al: Calcipotriol treatment of confluent and reticulated papillomatosis (Gougerot-Carteaud syndrome). J Eur Acad Dermatol Venereol 14:131, 2000

40. Gulec AT, Seckin D: Confluent and reticulated papillomatosis: treatment with topical calcipotriol. Br J Dermatol 141:1150, 1999

41. Schwartzburg J, Schwartzburg H: Response of confluent and reticulated papillomatosis of Gougerot and Carteaud to topical tretinoin. Cutis 66:291, 2000

42. Kim YC, Davis MD, Schanbacher CF, et al: Dowling-Degos disease (reticulate pigmented anomaly of the flexures): a clinical and histopathologic study of 6 cases. J Am Acad Dermatol 40:462, 1999

43. Lestringant GG, Masouye I, Frossard PM, et al: Co-existence of leukoderma with features of Dowling-Degos disease and reticulate acropigmentation Kitamura spectrum in five unrelated patients. Dermatology 195:337, 1997

44. Thami GP, Jaswal R, Kanwar AJ, et al: Overlap of reticulate acropigmentation of Kitamura, acropigmentation of Dohi and Dowling-Degos disease in four generations. Dermatology 196:350, 1998

45. Kim YC, Davis MD, Schanbacher CF, et al: Dowling-Degos disease (reticulate pigmented anomaly of the flexures): a clinical and histopathological study of 6 cases. J Am Acad Dermatol 40:462, 1999

46. Wenzel J, Tappe K, Gerdsen R, et al: Successful treatment of Dowling-Degos disease with Er:YAG laser. Dermatol Surg 28:748, 2002

47. Kovacs SO: Vitiligo. J Am Acad Dermatol 38:647, 1998

48. Le Poole C, Boissy RE: Vitiligo. Semin Cutan Med Surg 16:3, 1997

49. Baharav E, Merimski O, Shoenfeld Y, et al: Tyrosinase as an autoantigen in patients with vitiligo. Clin Exp Immunol 105:84, 1996

50. Moretti S, Spallanzani A, Amato L, et al: New insights into the pathogenesis of vitiligo: imbalance of epidermal cytokines at sites of lesions. Pigment Cell Res 15:87, 2002

51. Grimes PE, Wojdani A, Loeb LJ, et al: The effects of isoprinosine treatment on repigmentation and immunologic aberrations in vitiligo (abstr). J Invest Dermatol 98:534, 1992

52. Grimes PE, Sevall JS, Wojdani A: Cytomegalovirus DNA identified in skin biopsy specimens of patients with vitiligo. J Am Acad Dermatol 35:21, 1996

53. Dell'Anna M, Maresca V, Briganti S, et al: Mitochondrial impairment in peripheral blood mononuclear cells during the active phase of vitiligo. J Invest Dermatol 117:908, 2001

54. Picardo M, Grammatico P, Roccella F, et al: Imbalance in the antioxidant pool in melanoma cells and normal melanocytes from patients with melanoma. J Invest Dermatol 107:322, 1996

55. Schallreuter KU, Wood JM, Berger J: Low catalase levels in the epidermis of patients with vitiligo. J Invest Dermatol 97:1081, 1991

56. Schallreuter KU, Wood JM, Pittelkow MR, et al: Regulation of melanin biosynthesis in the human epidermis by tetrahydrobiopterin. Science 263:1444, 1994

57. Grimes PE: Therapies for vitiligo. Drug Therapy in Dermatology. Millikan L, Ed. Marcel Dekker, New York (in press)

58. Grimes PE: Psoralen photochemotherapy for vitiligo. Clin Dermatol 15:921, 1997

59. Jimbow K: Vitiligo: therapeutic advances. Dermatol Clin 16:399, 1998

60. Falabella R: Surgical therapies for vitiligo. Clin Dermatol 15:927, 1997

61. Westerhof W, Nieuweboer-Krobotova L: Treatment of vitiligo with UV-B radiation vs topical psoralen plus UV-A. Arch Dermatol 133:1525, 1997

62. Grimes PE: Therapeutic trends for the treatment of vitiligo. Cosmet Derm 15:21, 2002

63. Spencer JM, Nossa R, Ajmeri J: Treatment of vitiligo with the 308-nm excimer laser: a pilot study. J Am Acad Dermatol 46:727, 2002

64. Schallreuter KU, Wood JM, Lemke KR, et al: Treatment of vitiligo with a topical application of pseudocatalase and calcium in combination with short-term UVB exposure: a case study on 33 patients. Dermatology 190:223, 1995

65. Grimes PE, Soriano T, Dytoc M: The efficacy and safety of tacrolimus ointment. J Am Acad Dermatol (in press)

66. Lyle WM, Sangster JO, Williams TD: Albinism: an update and review of the literature. J Am Optom Assoc 68:623, 1997

67. Oetting WS, King RA: Molecular basis of albinism: mutations and polymorphism or pigmentation genes associated with albinism. Hum Mutat 13:99, 1999

68. Orlow SJ: Albinism: an update. Semin Cutan Med Surg 16:24, 1997

69. Sarangarajan R, Boissy RE: Tyrp 1 and oculocutaneous albinism type 3. Pigment Cell Res 14:437, 2001

70. Oetting WS: The tyrosinase gene and oculocutaneous albinism type 1 (OCA1): a model for understanding the molecular biology of melanin formation. Pigment Cell Res 13:320, 2000

71. Shotelersuk V, Gahl WA: Hermansky-Pudlak syndrome: models for intracellular vesicle formulation. Molec Genet Metab 65:85, 1998

72. Spritz RA: Hermansky-Pudlak syndrome and pale ear: melanosome-making for the new millennium. Pigment Cell Res 13:15, 2000

73. Turner M, Gahl WA, Toro J: Dermatologic manifestations of Hermansky-Pudlak syndrome in patients with and without a 16-base pair duplication in the HPS 1 gene. Arch Dermatol 135:774, 1999

74. Klein C, Philippe N, Le Deist F, et al: Partial albinism with immunodeficiency (Griscelli syndrome). J Pediatr 125:886, 1994

75. Pastural E, Barrat FJ, Dufourcq-Lagelouse R, et al: Griscelli disease maps to chromosome 15q21 and is associated with mutations in the myosin-Va gene. Nat Genet 16:289, 1997

76. Spritz RA, Holmes SA, Itin P, et al: Novel mutations of the KIT (mast/stem cell growth factor receptor) proto-oncogene in human piebaldism. J Invest Dermatol 101:22, 1993

77. Richards KA, Fukai K, Osio N, et al: A novel KIT mutation results in piebaldism with progressive depigmentation. J Am Acad Dermatol 44:288, 2001

78. Syrris P, Malik N, Murday VA, et al: Three novel mutations of the proto-oncogene KIT cause human piebaldism. Am J Med Genet 95:79, 2000

79. Spritz RA: Piebaldism, Waardenburg syndrome and related disorders of melanocyte development. Semin Cutan Med Surg 16:15, 1997

80. Spritz RA, Hearing VJ Jr: Genetic disorders of pigmentation. Adv Hum Genet 22:1, 1994

81. Falabella R: Idiopathic guttate hypomelanosis. Dermatol Clin 6:241, 1988

82. Wallace ML, Grichnik JM, Prieto VG, et al: Numbers and differentiation status of melanocytes in idiopathic guttate hypomelanosis. J Cutan Pathol 25:375, 1998

83. Hexsel DM: Treatment of idiopathic guttate hypomelanosis by localized superficial dermabrasion. Dermatol Surg 25:917, 1999

XVI APPROACH TO THE DIAGNOSIS OF SKIN DISEASE

Robert T. Brodell, m.d.
Stephen E. Helms, m.d.

Patients frequently see their primary care physician for skin disease[1]; however, compared with dermatologists, primary care physicians treat substantially fewer patients with common skin conditions, and the types of cutaneous disease they treat tend to be few in number.[2,3] One study reported that dermatologists had 728 and 352 office visits a year for acne and contact dermatitis, respectively; by contrast, internists averaged three and nine visits, a year, respectively, and family physicians averaged eight and 27 visits a year.[2] The relative inexperience with cutaneous presentations gives rise to possible error in the dermatologic care offered by nondermatologists. Some studies have reported that nondermatologists perform poorly in the diagnosis and treatment of skin disease.[4] One area of concern is the apparent low proficiency among nondermatologists in the diagnosis of skin cancer.[5,6] The root of most problems encountered by primary care physicians in the treatment of skin disease rests in establishing the accurate diagnosis of cutaneous presentations.

Diagnosis of a cutaneous disease is most reliably achieved by a stepwise approach to patient evaluation, beginning with an examination of the morphologic features of the skin lesions and frequently culminating in diagnostic testing. This chapter reviews the primary skin lesions that allow categorization of dermatologic disease (e.g., papulosquamous diseases, blistering diseases, nonscaling erythematous and infiltrative diseases, and tumors) and presents a method by which the physician can narrow the possible causes of a specific presentation and arrive at a diagnosis in a cost-effective manner.

Approach to the Patient with a Dermatologic Lesion

Dermatology is a visual specialty, and physical examination is primarily oriented toward observing the skin. Dermatologists approach skin disease in a manner that has been tested over time and perpetuated in the training of medical students and residents. A simple diagnostic evaluation based on the approach preferred by dermatologists allows primary care physicians to narrow the possible causes of a cutaneous presentation and arrive at an accurate diagnosis.

diagnostic evaluation

Diagnostic evaluation of a cutaneous presentation begins with a brief patient history that is directed at the nature of the chief complaint and its onset; factors that aggravate and alleviate symptoms; and responses to over-the-counter or prescription medications. This is followed by careful inspection of the skin. In examining a patient with a rash, the first step is to try to identify primary lesions (i.e., lesions that appear early in the disease process) [*see* Morphologic Classification of Skin Disorders, Primary Lesions, *below*]); these lesions help to categorize the disease and provide the basis for diagnosis. Information derived from the identification of primary lesions is augmented by an examination of primary lesions that have undergone change. Secondary changes to primary lesions may occur naturally or after trauma, such as scratching [*see* Morphologic Classification of Skin Disorders, Secondary Changes, *below*]. The location, distribution, and configuration of primary lesions and their secondary changes are analyzed, and the findings are categorized to promote the development of a differential diagnosis.

After an examination of the lesions, a more complete patient history is obtained, including the patient's family history, social history, and medical history. The expanded patient history is followed by a focused general medical examination. In the context of a detailed patient history, the general medical examination often provides diagnostic clues that further narrow the differential diagnosis.

The final step in a dermatologic examination comprises various forms of testing (e.g., dermatologic testing, skin biopsy, and laboratory tests) [*see* Arriving at a Diagnosis, *below*] to confirm the diagnosis or sufficiently narrow the differential diagnosis to permit selection of the most appropriate treatment. If the diagnosis remains uncertain after testing, consultation with a dermatologist may be useful in establishing the diagnosis. The patient is typically scheduled for a follow-up visit after initiation of treatment. The purpose of the follow-up visit is to assess the response to therapy and to confirm that the proper diagnosis was rendered.

suboptimal methods of diagnosis and management

Errors in dermatologic diagnosis can be classified into several categories. It is worth exploring these problems to avoid falling into predictable traps.

Treating Symptoms Rather than Diseases

Establishing the underlying cause of symptoms is a guiding principle in medicine; however, the treatment of dermatologic presentations frequently focuses on symptom management without addressing the underlying cause. This approach is seldom an efficient or effective form of management. For example, treatment of pruritus with antihistamines is a poor substitute for establishing a definitive diagnosis of the underlying condition that is the cause of the symptom. In the case of a patient with severe itching associated with dermatitis herpetiformis, treatment with a topical corticosteroid may give temporary relief from pruritus; however, a careful examination would most likely reveal grouped papulovesicles on the extensor surfaces of the extremities, and a biopsy would confirm the diagnosis of dermatitis herpetiformis. Treatment of dermatitis herpetiformis with a gluten-free diet and oral dapsone would lead to dramatic, long-lasting remission.

Snapshot Approach to Diagnosis

A snapshot diagnosis is rendered on the basis of physical appearance of a rash or other form of lesion in the absence of any other data. This method of examination is quick; however, it can lead to inaccurate diagnosis and imprecise and inadequate treatment. For example, a patient with widespread scaling, erythema, lichenification, and excoriations may appear to have eczema. On careful examination, the finger webs disclose burrows. The patient reports that itching is more severe at night, and that family members are also experiencing itching. A scabies preparation test discloses the presence of mites, confirming the diagnosis as

Table 1 Primary Lesions: Consensus Definitions of Dermatologic Morphologic Terms[7]

Morphologic Term	*DLP Proposed Definition*
Bulla	A fluid-filled blister greater than 0.5 cm in diameter; fluid can be clear, serous, hemorrhagic, or pus-filled
Comedo	An enlarged hair follicular infundibulum primarily containing keratin and lipids and having a plugged, dilated follicular opening (blackhead) or a clinically unapparent follicular opening (whitehead)
Macule	A flat area of skin or mucous membranes having a color different from the surrounding tissue and a diameter generally less than 0.5 cm; macules may have nonpalpable, fine scales
Nodule	A dermal or subcutaneous firm, well-defined lesion usually greater than 0.5 cm in diameter
Papule	A discrete, solid, elevated body usually less than 0.5 cm in diameter; papules are further classified by shape, size, color, and surface change
Patch	A flat area of skin or mucous membranes having a color different from the surrounding tissue and a diameter generally greater than 0.5 cm; patches may have nonpalpable, fine scales
Plaque	A discrete, solid, elevated body usually broader than it is thick and measuring more than 0.5 cm in diameter; plaques may be further classified by shape, size, color, and surface change
Pustule	A circumscribed elevation that contains pus; pustules are usually less than 0.5 cm in diameter
Vesicle	Fluid-filled cavity or elevation less than 0.5 cm in diameter; fluid may be clear, serous, hemorrhagic, or pus-filled
Wheal	An edematous, transitory papule or plaque

DLP—Dermatology Lexicon Project

scabies infestation. A snap diagnosis of eczema and treatment with topical steroids would have been inappropriate.

Scattershot Management

A suboptimal approach to the management of a dermatologic presentation is touted by physicians who delude themselves into believing that all rashes look alike. This leads to the scattershot approach to management, which advocates increasing the potential for successful treatment of an unknown skin disease by treating all possible causes. For example, the use of a topical steroid/antifungal preparation for a papulosquamous process may seem a prudent treatment of two possible disorders—namely, eczema and superficial fungal infection; however, if the patient has a dermatophyte-induced fungal infection, the topical steroid may decrease local immunity and slow the healing process. This management strategy is expensive, increases the risk of iatrogenic disease, and delays appropriate diagnosis and treatment.

A scattershot approach may also be used in diagnostic evaluation; such an approach entails the ordering of a broad battery of tests in the hope of stumbling upon the correct diagnosis. This inefficient method can lead to false positive results that confuse rather than confirm the diagnosis.

All rashes present physical diagnostic clues that are useful in establishing a diagnosis. A careful evaluation of the lesions' morphologic characteristics, coupled with a thorough patient history and examination, will most likely provide an accurate diagnosis. If the diagnosis remains uncertain, the morphologic condition of the lesion will suggest which specific tests are appropriate for arriving at the diagnosis.

Morphologic Classification of Skin Disorders

PRIMARY LESIONS

The first step in the diagnosis of a rash is to identify primary lesions [*see Table 1*]. Primary lesions are those physical characteristics of skin disease that appear initially and are most useful in developing a differential diagnosis. The characterizing features of primary lesions include whether they are flat or raised, solid or fluid filled, dark or light in color, large or small, smooth or rough. Lesions may be few or numerous, localized or widespread. The newly erupted and undisturbed lesions are most often helpful in categorizing skin conditions in a manner that leads to a correct diagnosis.

Primary skin lesions can be defined simply. Flat lesions are referred to as macules when they are smaller than 0.5 cm and as patches when they are greater than 0.5 cm in diameter [*see Figure 1*].

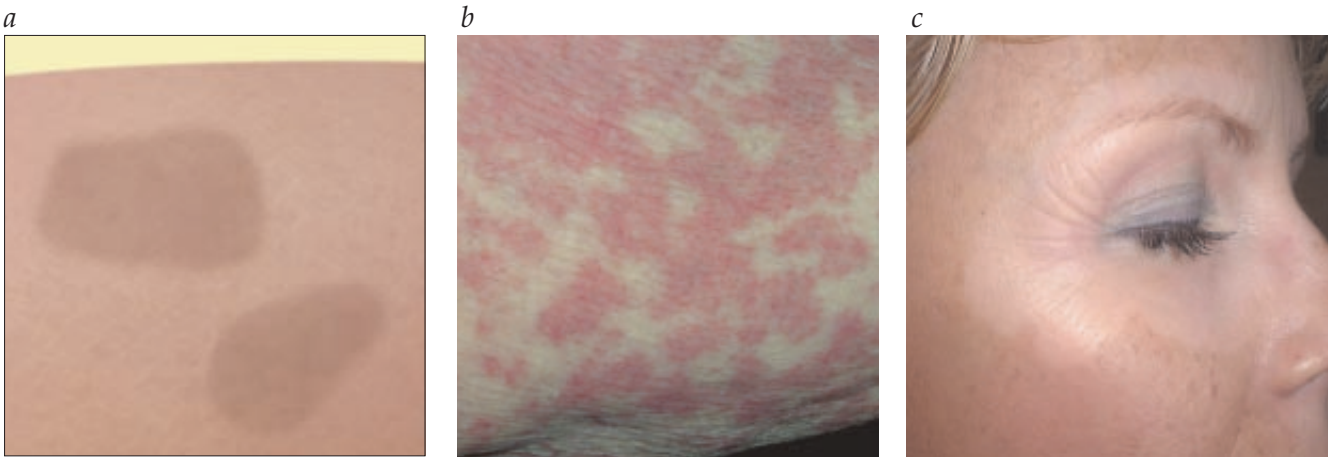


Figure 1 **(*a*) Schematic drawing of macules (lesions < 0.5 cm) or patches (lesions > 0.5 cm). Macules and patches are flat areas of skin for which the color and texture differ from that of the surrounding tissue. (*b*) Nonblanching erythematous macules and patches are present in a patient with a drug eruption. (*c*) Depigmented macules are noted on the face of a patient with vitiligo.**

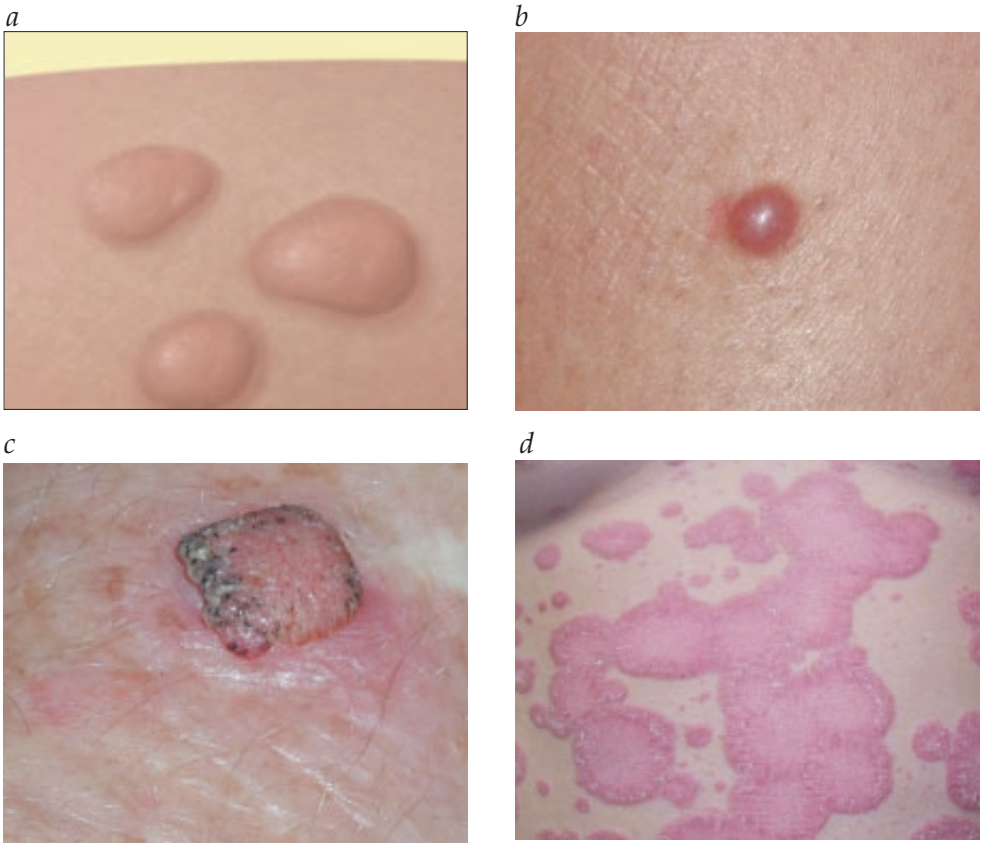


Figure 2 **(*a*) Schematic drawing of papules (lesions < 0.5 cm) or nodules (lesions > 0.5 cm). Papules and nodules are discrete, solid, elevated lesions. (*b*) A raised, dome-shaped, erythematous papule is seen in a case of dermatofibroma. (*c*) A raised, flat-topped, erythematous, and hyperkeratotic nodule with scalloped edges is present in a patient with squamous cell carcinoma. (*d*) Large, erythematous plaques of psoriasis have an annular appearance, owing to their elevated margins.**

Small raised bumps are referred to as papules [*see Figures 2a and b*], and large lesions of this type are referred to as nodules; nodules typically have a deeper dermal component [*see Figure 2c*]. Discrete, broad, raised eruptions are referred to as plaques [*see Figure 2d*]. Raised lesions containing fluid (commonly known as blisters) are referred to as vesicles when smaller than 0.5 cm [*see Figures 3a and b*] and as bullae when larger than 0.5 cm in diameter [*see Figure 3c*]. Vesicles and bullae are usually clear but may be turbid. White or yellow fluid-filled lesions are called pustules. Atrophic lesions exhibit a thinned epidermis that is often depressed; they have a scaly, shiny surface that has the texture of cigarette paper. An edematous transitory papule or plaque is called a wheal [*see Figure 4*].

The shape of primary lesions often provides diagnostic clues. Primary lesions may be round, oval, angular, or irregular; flat-topped or domed; or umbilicated or verrucous. The borders of primary lesions may be well circumscribed or poorly defined; the presence or absence of an elevated border may be a useful finding.

SECONDARY CHANGES

Secondary changes to lesions provide valuable diagnostic information; however, they are not as useful as primary lesions in arriving at a specific diagnosis [*see Table 2*]. Secondary changes may represent a late stage in the natural history of primary lesions, or they may be the result of trauma such as from scratching or rubbing of the skin. Secondary changes to lesions include the following:

- Scales: small flakes of superficial skin.
- Scale crusts: scales combined with serous exudate.
- Excoriations: abrasions resulting from the scratching of elevated lesions.
- Erosions: localized loss of epithelium.
- Ulcers: denuded areas of epidermis and some portion of dermis. Ulcers may be open or covered with a black eschar [*see Figure 5*].
- Scars: raised or depressed fibrous lesions caused by trauma or disease.
- Cutaneous horns: keratotic projections extending from a skin lesion.
- Fissures: cracks that extend through the epidermis into the dermis.

Numerous additional terms are helpful in characterizing the morphologic presentation of skin lesions [*see Table 3*]. It is critical that exacting definitions of descriptive terms be used by all clinicians if a reproducible method of diagnosis is to be promulgated. The Dermatology Lexicon Project (DLP) has provided an expert consensus of definitions for dermatologic terms [*see Tables 1 through 3*].[7]

MORPHOLOGIC PATTERNS OF PRESENTATION

Once the patient has been examined for primary and secondary lesions, the diagnostician must consider the overall presentation of the rash and determine its location, distribution, and configuration—three factors essential in determining a diagnosis.

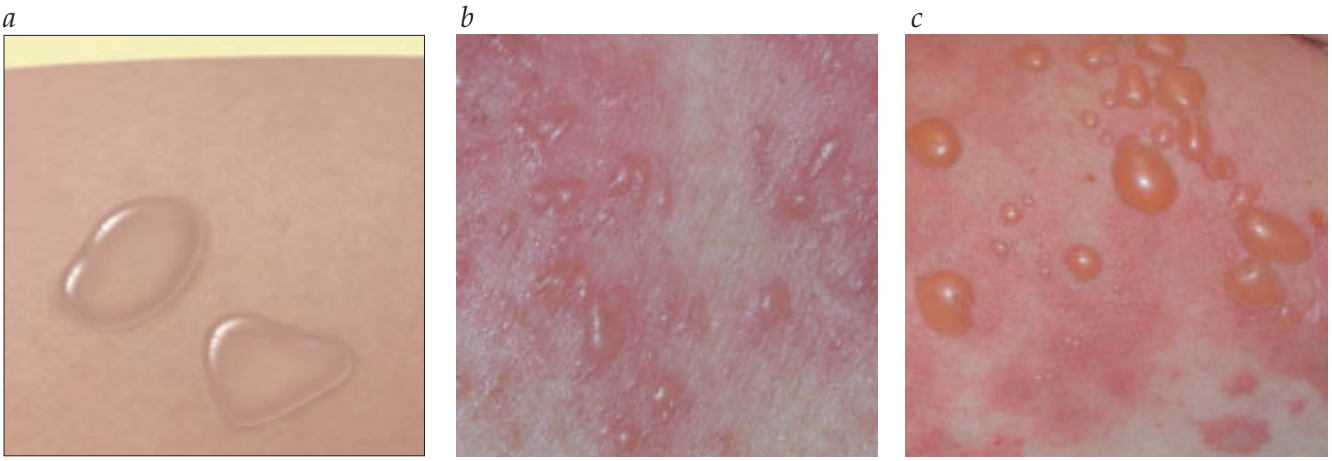


Figure 3 **(*a*) Schematic drawing of a vesicle (lesions < 0.5 cm) or bulla (lesions > 0.5 cm). Vesicles, bullae, and pustules are fluid-filled elevations or cavities in the skin. Vesicles and bullae are clear; pustules are turbid and pus filled. (*b*) Small, clear, fluid-filled vesicles on an erythematous base are present in a patient who has herpes zoster. (*c*) Large, clear, fluid-filled bullae are present in a patient with bullous pemphigoid. The bullae are associated with erythematous patches.**

Location

The location refers to the particular site where the lesion or lesions are found. The location should be carefully defined because some skin diseases target specific anatomic areas. Notation of involved anatomic sites should be as specific as possible, listing not only the involved sites but the affected aspects of those sites. For example, facial lesions may occur in the periorbital or perioral areas; hand lesions may occur on the fingers or palm; and foot lesions may occur on the toes or sole. Lesions might also be found on the upper arm or the forearm, the lower leg or the thigh, and the trunk. It is important to further describe the affected areas as being on the right or left side and on the proximal or distal, medial or lateral, dorsal or ventral, and flexural or extensor surfaces of the involved anatomic sites.

Distribution

The distribution of lesions describes the overall pattern of an eruption in relation to the entire cutaneous surface. The rash may be localized to one area of the body, may involve several areas, or may extend over much of the body surface. Rashes may be symmetrical or asymmetrical, and they may be present primarily on the trunk or on the extremities. Rashes may be present on exposed skin (i.e., skin that is not covered by clothing) or unexposed skin. These characteristics should be carefully noted because a particular distribution will narrow the differential diagnosis.

Configuration

The configuration of lesions refers to the pattern exhibited by multiple lesions within a defined area. Because the configuration of lesions may vary according to the disorder, any detectable pattern may be helpful in arriving at a definitive diagnosis. Some of the more common configurations include the following:

- Grouped or herpetiform configuration: multiple small lesions appearing within a small, defined area [*see Figure 6*].
- Zosteriform configuration: lesions occurring within a dermatone.
- Linear configuration: lesions oriented along a line [*see Figure 7*].
- Annular configuration: lesions appearing in a ringlike pattern.
- Target (iris) configuration: lesions appearing in concentric rings.

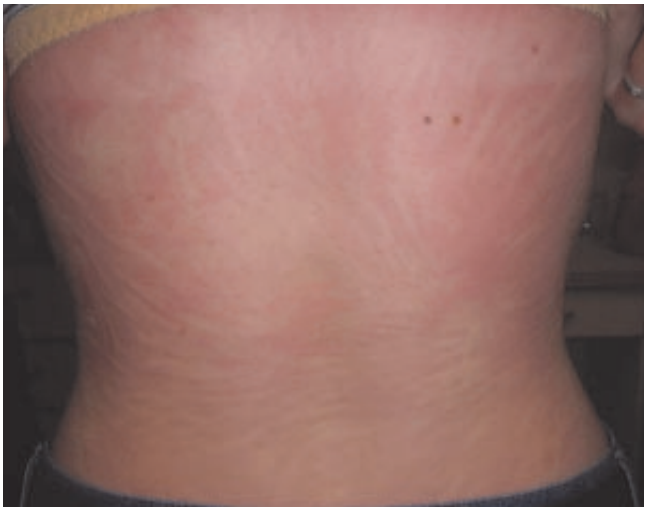
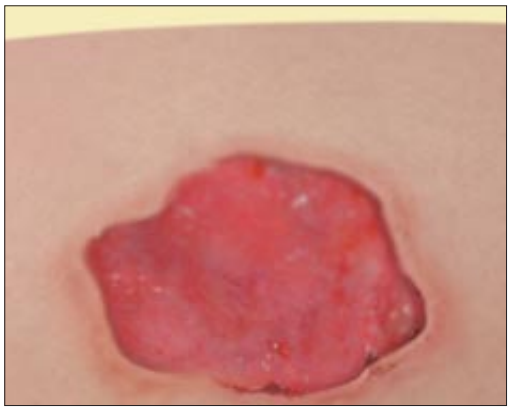


Figure 4 **Multiple linear, erythematous wheals secondary to scratching are noted on the back of a patient with chronic urticaria and dermatographism.**

Table 2 Selected Secondary Lesions: Consensus Definitions of Dermatologic Morphologic Terms[7]

Morphologic Term	*DLP Proposed Definition*
Horn	Abnormally keratinized cutaneous projection taller than it is broad
Erosion	A localized loss of the epidermal or mucosal epithelium
Fissure	A linear crack or cleavage within the skin usually found with thickened skin
Ulcer	A circumscribed loss of the epidermis and at least the upper dermis; ulcers are further classified by their depth, border/shape, edge, and tissue at their base

a



b

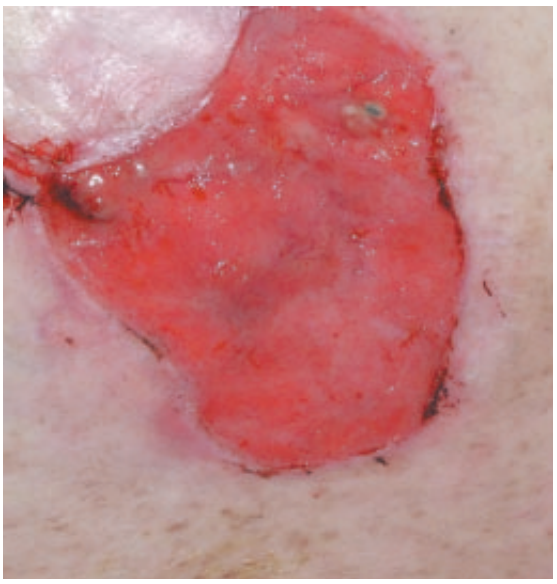


Figure 5 **(*a*) Schematic drawing of a skin ulceration. A skin ulceration is a circumscribed lesion denuded of epidermis and at least some dermis. (*b*) A well-defined 3.5 by 4.0 cm ulceration with an erythematous granulating base is present in a patient who has early evolving pyoderma gangrenosum.**

- Arcuate configuration: lesions appearing in a semicircular pattern.
- Polycyclic configuration: lesions appearing as interlocking rings.
- Serpiginous configuration: lesions appearing in snakelike whorls.
- Digitate configuration: lesions resembling the size and shape of a fingertip.

When lesions coalesce over large areas, they are termed confluent. Erythroderma describes a widespread confluence of rash covering nearly all of the cutaneous surface.

Color

The color of cutaneous lesions also provides important diagnostic clues. Lesions may be flesh-colored, hyperpigmented or hypopigmented, erythematous, or virtually any color of the rainbow. Purpuric rashes caused by the extravasation of red blood cells show no blanching on diascopy (i.e., a test in which a glass slide or lens is pressed against the skin).

Categories of Skin Diseases

The appearance of individual lesions on the skin (e.g., primary lesions and their secondary changes) classifies a rash or growth within a major category of skin disease. The most common skin diseases and many important rare conditions can be classified into one of five disease categories on the basis of their characteristic lesions. Once the category is determined, the diseases within that category are considered in the differential diagnosis of the presenting disorder [*see Table 4*].

As a clinician's dermatologic knowledge becomes more sophisticated, additional categories can be mastered, including (but not limited to) diseases of the hair, nails, or mucous membranes; photosensitivity diseases; diseases of vascular reactivity; ulcerative skin conditions; and conditions typical of specific distributions, such as diseases of the genitalia, feet and hands, and eyelids. Manuals of differential diagnosis based on the morphology of lesions and other physical features are plentiful and can be quite helpful in determining a diagnosis.

Table 3 Other Important Morphologic Terms: Consensus Definitions of Dermatologic Morphologic Terms[7]

Morphologic Term	*DLP Proposed Definition*
Abscess	A localized accumulation of pus in the dermis or subcutaneous tissue; frequently red, warm, and tender
Atrophy	A thinning of tissue defined by the location (e.g., epidermal atrophy, dermal atrophy, or subcutaneous atrophy)
Burrow	A threadlike linear or serpiginous tunnel in the skin typically made by a parasite
Carbuncle	An inflammatory nodule composed of coalescing furuncles
Ecchymosis	A discoloration of the skin or mucous membranes resulting from extravasation of blood that exhibits color change over time; the characteristic transition is from blue-black to brown-yellow to green
Erythema	Localized, blanchable redness of the skin or mucous membranes
Exfoliation	Desquamation of the superficial epidermis appearing as a fine scaling or as peeling sheets
Furuncle	A follicle-centered nodule caused by a suppurative infection characterized by pain, redness, and perhaps visible pus; usually greater than 1 cm in diameter
Induration	Hardening of the skin beneath the epidermis, usually caused by edema, dermal sclerosis, inflammation, or infiltration
Petechiae	Purpuric nonblanchable macules resulting from tiny hemorrhages, initially measuring 1 to 2 mm
Poikiloderma	An area of variegated pigmentation, atrophy, and telangiectasia
Purpura	Hemorrhaging into skin or mucous membranes that varies in size, color, and duration; types of purpura include palpable purpura, ecchymosis, and petechiae
Telangiectasia	Visible, persistent dilation of small, superficial cutaneous blood vessels

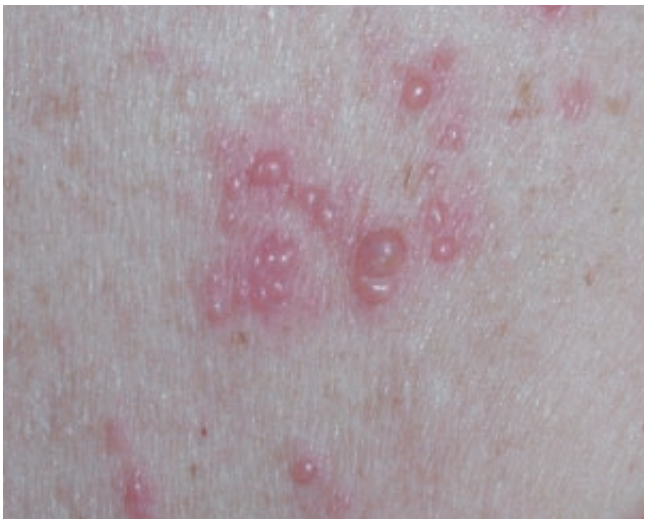


Figure 6 **Close-up view of a herpetiform pattern of vesicles on an erythematous base within a dermatome in a patient with herpes zoster.**

Arriving at a Diagnosis

It is important to begin the assessment of a skin condition with a broad differential diagnosis, noting the presentation as characteristic of one of the categories of skin disease [*see Table 4*]. Using physical findings, patient history, and diagnostic testing, the differential diagnosis is gradually narrowed until a diagnosis is determined. For example, scaling conditions, including rashes composed of both papules and plaques, are characterized as papulosquamous skin diseases; each papulosquamous condition [*see Table 4*] should be considered in the differential diagnosis of a scaling rash. The specific features of each of the papulosquamous conditions are compared with the patient's lesions to systematically identify the disorder. It is critical that the initial differential diagnosis be broadly determined. Jumping to an early conclusion and not systematically considering all of the possible papulosquamous disorders can result in an incorrect diagnosis. In fact, if the actual diagnosis is not among the diseases initially considered, it is much more difficult to determine the correct diagnosis.

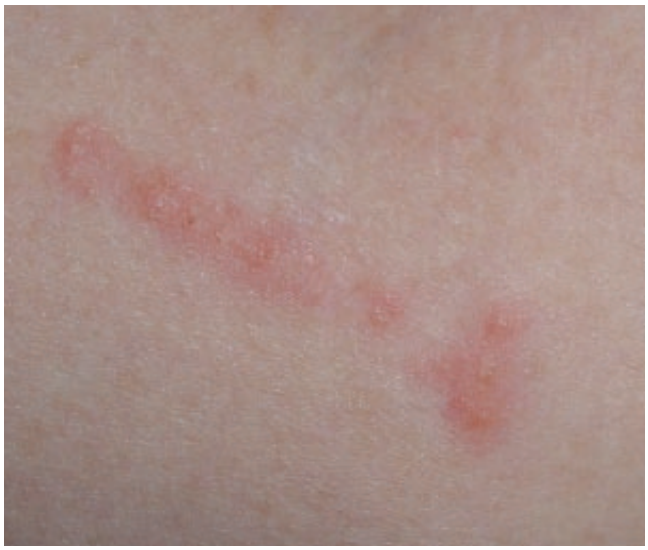


Figure 7 **Linear arrangement of papules in a patient with contact dermatitis caused by poison ivy.**

Table 4 Categories of Skin Disease

Papulosquamous diseases (discrete papules or plaques with scaling)	Psoriasis vulgaris Chronic atopic dermatitis Lichen planus Pityriasis rosea Fungal infections Secondary syphilis Mycosis fungoides
Blistering diseases (vesicles, bullae, pustules)	Acute allergic contact dermatitis Bullous pemphigoid Pemphigus vulgaris Dermatitis herpetiformis Herpesvirus infections Bacterial folliculitis
Nonscaling erythematous (macules, patches, wheals) and infiltrative diseases (plaques)	Urticaria Morbilliform drug eruptions Viral exanthems Sarcoidosis Leukemia and lymphoma cutis Amyloidosis
Diseases of pigmentation (macules or patches of various colors)	Vitiligo Tinea versicolor Pityriasis alba Café au lait macules Lentigines
Benign and premalignant tumors (macules, papules, nodules, tumors)	Actinic keratosis Basal cell carcinoma Squamous cell carcinoma Melanoma Acrochordons Dermatofibroma Neurofibroma Melanocytic nevi Adnexal tumors

Diagnosis is almost always more difficult than treatment. Consultation with a dermatologist is recommended when the diagnosis is uncertain, particularly in cases in which treatment may fail or may lead to iatrogenic disease. The dermatologist may help define the primary lesions that permit the accurate categorization of the disease process and, in turn, suggest the diagnosis.

If a diagnosis remains in doubt, a follow-up visit with the patient should be scheduled because, as the disease progresses, the development of primary lesions and lesion distribution may make the diagnosis more apparent. In addition, single, confirmatory tests often prove helpful in the diagnosis of cutaneous presentations; tests used to confirm a diagnosis include a Wood light examination, potassium hydroxide (KOH) preparation, sampling for fungal culture, scabies preparation, Tzanck preparation, patch testing, skin biopsy, dark-field microscopic examination, microscopic hair-shaft analysis, Gram stain, and viral or bacterial cultures.

A Wood light examination is performed by shining a black light on the skin in a dimly lit room. Epidermal pigmentation (e.g., lentigines) is highlighted by this examination, whereas dermal pigmentation (e.g., Mongolian spot) disappears. A search for depigmented spots such as ash-leaf macules in babies with tuberous sclerosis is also aided by the Wood lamp.

A KOH test for fungal infections involves applying KOH to scales of skin or hair shafts to clear the keratin so that fungal hyphae and spores can be identified. For example, scales can be lightly scraped onto a glass slide after placing the slide on the advancing margin of an annular plaque with central clearing. After a coverslip is applied, 2.5% KOH preparation is applied to the slide next to the coverslip. The KOH preparation spreads under the coverslip by capillary action. After gentle heating, excess KOH is blotted away, and the specimen is examined under a microscope. Fungal infections can also be confirmed by obtaining scales by lightly scraping papulosquamous lesions and sprinkling the scales onto Sabouraud dextrose agar. This agar preparation is then incubated at room temperature for several weeks, after which it is analyzed for colony growth, color, and morphology.

A scabies preparation test involves applying oil to excoriated lesions—frequently found on wrists and finger webs—and lightly scraping the burrows to obtain their contents. The scrapings are placed on a slide with coverslip; the test is positive if scabies mites, eggs, or feces are visible on microscopic examination.

Tzanck preparations are smears obtained from the base of intact vesicles and stained with one of a variety of nuclear stains. A finding of multinucleated giant cells suggests a diagnosis of herpes simplex or herpes zoster infection.

Patch testing is performed to objectively elucidate the specific cause of an allergic contact dermatitis. Standard allergens are applied under patches for 24 hours, and the reactions are read 48 hours and 1 week later.

Skin biopsies are helpful in confirming the diagnosis of a variety of inflammatory and neoplastic diseases. Dark-field microscopic examination of serous fluid from genital ulcers identifies spirochetes in lesions of primary syphilis. Hair-shaft analysis is helpful when alopecia is caused by hair-shaft abnormalities. Finally, viral and bacterial cultures using specific swabs can provide laboratory confirmation of a variety of viral and bacterial diseases.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Lowell BA, Froelich CW, Federman DG, et al: Dermatology in primary care: prevalence and patient disposition. J Am Acad Dermatol 45:250, 2001

2. Fleischer AB Jr, Herbert CR, Feldman SR, et al: Diagnosis of skin disease by nondermatologists. Am J Manag Care 6:1149, 2000

3. Feldman SR, Fleischer AB Jr, McConnell RC: Most common dermatologic problems identified by internists, 1990–1994. Arch Intern Med 158:1952, 1998

4. Federman D, Hogan D, Taylor JR, et al: A comparison of diagnosis, evaluation, and treatment of patients with dermatologic disorders. J Am Acad Dermatol 32:726, 1995

5. Whited JD, Hall RP, Simel DL, et al: Primary care clinicians' performance for detecting actinic keratoses and skin cancer. Arch Intern Med 157:985, 1997

6. Gerbert B, Maurer T, Berger T, et al: Primary care physicians as gatekeepers in managed care: primary care physicians' and dermatologists' skills at secondary prevention of skin cancer. Arch Dermatol 132:1030, 1996

7. Morphologic terminology. Dermatology Lexicon Project, 2002
http://www.dermatologylexicon.org

Reviews

Ashton R, Leppard B: Differential Diagnosis in Dermatology. Radcliffe Medical Press, Oxford, 1993

Ghatan HEY: Dermatological Differential Diagnosis and Pearls, 2nd ed. CRC Press, New York, 2002

Kusch SL: Clinical Dermatology: A Manual of Differential Diagnosis, 3rd ed. TaroPharma, 2003
http://www.taropharmadermatology.com

Lawrence CM, Cox NH: Physical Signs in Dermatology: Color Atlas and Text, 2nd ed. Mosby-Wolfe, London, 2001

Lazarus GS, Goldsmith LA: Diagnosis of Skin Disease. FA Davis Co, Princeton, 1980

Pocket Guide to Cutaneous Medicine and Surgery. Dover JS, Ed. Harcourt Brace, Philadelphia, 1996

White GM: Color Atlas of Regional Dermatology. Mosby, London, 1997

Acknowledgment

Figures 1a, 2a, 3a, and 5a Dragonfly Media Group.

I THYROID

Paul W. Ladenson, m.d.

Thyroid disorders are the most common endocrine conditions encountered in clinical practice. Persons of either sex and any age can be affected, although almost all forms of thyroid disease are more frequent in women than in men, and many thyroid ailments increase in frequency with age. The presentation of thyroid conditions can range from clinically obvious to clinically silent. Their consequences can be widespread and serious, even life-threatening. With proper testing, the diagnosis and differential diagnosis can be established with certainty, and effective treatments can be instituted for almost all patients.

Definitions

States of thyroid dysfunction include hypothyroidism and thyrotoxicosis, both of which have ubiquitous metabolic and organ-specific consequences that result in a wide variety of clinical presentations and complications. Thyrotoxicosis is sometimes referred to as hyperthyroidism, but the latter term is more properly limited to forms of thyrotoxicosis in which there is an overproduction of thyroid hormones by the gland.

Both categories of thyroid dysfunction are further classified as overt or mild. In overt thyroid dysfunction, the concentrations of thyrotropin (thyroid-stimulating hormone [TSH]) and one or both thyroid hormones are outside of their normal ranges. In mild thyroid dysfunction, the serum TSH level is abnormal, but the serum thyroid hormone concentrations remain within their reference ranges. Although the terms clinical and subclinical are often used in reference to overt and mild thyroid dysfunction, respectively, these states are actually defined on the basis of biochemical criteria, not of clinical manifestations.

Epidemiology

In the Third National Health and Nutrition Survey (NHANES III), thyroid function tests were assessed in a group of 17,353 persons 12 years of age or older whose makeup reflected the geographic and ethnic diversity of the United States population.[1] Hypothyroidism was identified in 4.6% (0.3% overt and 4.3% mild), and thyrotoxicosis was found in 1.3% (0.5% overt and 0.7% mild) [*see Figure 1*].

thyroid nodules

Thyroid nodules (masses within the gland) are relatively common in adults. In the Framingham Study, 6% of women and 2% of men had palpable thyroid nodules.[2] The prevalence of nonpalpable thyroid nodules incidentally detected by imaging studies such as sonography and CT has been reported to be as high as 27% in adults.[3] Diffuse thyroid gland enlargement (goiter) is declining in prevalence—a tendency that reflects the increase in levels of dietary iodine in the United States. Whereas goiter was identified in 3% of persons in a 10-state United States survey in the 1970s, it was self-reported by less than 0.5% of persons in the more recent NHANES III.[4,5]

thyroid cancer

Thyroid cancer is the 14th most common malignancy in the United States, with an estimated annual incidence of 23,600 new cases and a female-to-male ratio of 3 to 1.[6] However, the epidemiology of thyroid cancer is more important than this incidence ranking would imply, for two reasons. First, thyroid cancer is currently the malignancy with the fastest rising incidence in the United States, with increases of 3.8% annually from 1992 to 2001. Second, because treatment is highly effective, with 95% or more of patients surviving, there may be about 300,000 thyroid cancer survivors in the United States, all of whom require monitoring for recurrent disease.

Hypothyroidism

epidemiology

Hypothyroidism is a common disorder that occurs more commonly in women than in men; in both sexes, the incidence increases during and after middle life.[7] In the NHANES III, 2% of persons 65 years and older had overt hypothyroidism, and 14% had mild hypothyroidism.[1] Prevalences of thyroid dysfunction were also higher in whites and Mexican Americans than in blacks (5%, 4%, and 2%, respectively).

Certain individuals are at higher risk for developing hypothyroidism, including those with a family history of autoimmune thyroid disorders[8]; postpartum women[9]; those with a history of head and neck or thyroid irradiation or surgery; those with certain other autoimmune endocrine conditions[10] (e.g., type 1 diabetes mellitus, adrenal insufficiency, and ovarian failure); and those with certain nonendocrine autoimmune disorders (e.g., celiac disease, vitiligo, pernicious anemia, and Sjögren syndrome). Hypothyroidism also develops more frequently in persons with Down syndrome or Turner syndrome.

etiology and genetics

The causes of hypothyroidism vary, depending on whether the disease is congenital or acquired. In addition, the causes of primary hypothyroidism (i.e., disease of the thyroid gland itself) differ from those of secondary (central) hypothyroidism, which involves deranged hypothalamic-pituitary control of the gland.

Congenital Hypothyroidism

Endemic iodine deficiency remains an important cause of congenital hypothyroidism in certain regions of the world. Even with sufficient dietary iodine, congenital hypothyroidism affects one in 4,000 infants because of thyroid gland dysgenesis (as related, for example, to mutant *PAX8* and *TTF1* genes) or inherited defects in thyroid hormone synthesis (e.g., mutations in the genes that code for thyroid peroxidase, sodium-iodide symporter, and thyroglobulin). Absent or ineffective TSH responsiveness can be the result of mutations in the genes affecting pituitary thyrotrope differentiation (e.g., *POU1F1* and *PROP1*) or the structures of the thyrotropin-releasing hormone (TRH) receptor, the TSH β chain, and the TSH receptor. A mutation in the gene for $G_s\alpha$, which mediates adenylate cyclase activation in thyroid cells, causes hypothyroidism in pseudohypoparathyroidism. Inherited resistance to thyroid hormone can be caused by mutations in the β isoform of the nuclear triiodothyronine (T_3) receptor.

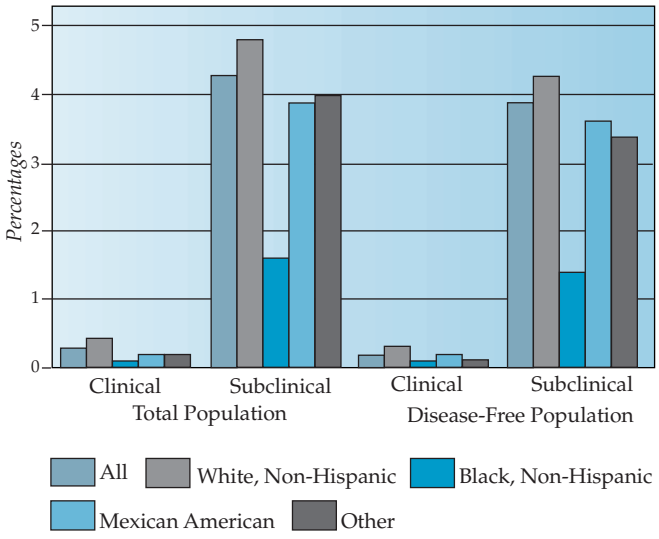


Figure 1 **Prevalences of abnormalities on thyroid function tests in different populations in the third National Health and Nutrition Examination Survey.**

Acquired Hypothyroidism

Autoimmune thyroiditis Autoimmune thyroiditis, also called Hashimoto disease, is far and away the leading cause of hypothyroidism.[11] Its autoimmune pathogenesis is evidenced by the lymphocytic infiltration of the thyroid, the presence of circulating thyroid autoantibodies and activated $CD4^+$ T cells specific for thyroid antigens, and the expression of antigen-presenting major histocompatibility complex (MHC) class II proteins by thyrocytes. There is a genetic predisposition to autoimmune thyroiditis, and a polygenic basis for this predisposition is suggested by linkage to several genetic loci in affected kindreds.[12] Because autoimmune thyroiditis is more common in populations with higher dietary-iodine content, it has been postulated that high levels of dietary iodine cause an increase in thyroglobulin antigenicity. Thyroid autoimmunity can be initiated by interferon-alfa therapy and can cause either hypothyroidism or hyperthyroidism.[13] Discontinuance of immunomodulatory therapy often reverses this effect.

Other causes of thyroid injury Thyroid surgery or thyroid irradiation—whether in the form of radioactive iodine therapy for thyrotoxicosis or external-beam radiotherapy for head and neck malignancies[14]—commonly results in hypothyroidism. In hemochromatosis, iron infiltration of the gland can cause thyroid failure. Transient primary hypothyroidism also occurs with lymphocytic thyroiditis (also known as postpartum or painless thyroiditis) and subacute thyroiditis.

Drug and toxins causing hypothyroidism Long-term administration of iodine in pharmacologic quantities, such as with amiodarone[15] or iodine-containing expectorants, can inhibit thyroid hormone production, particularly in patients with underlying autoimmune thyroiditis. Lithium carbonate interferes with hormone release from the thyroid gland, resulting in transient TSH elevation in one third of patients and sustained hypothyroidism in 10%; those with autoimmune thyroiditis are especially vulnerable.[16] The antiretroviral agent stavudine and the drugs aminoglutethimide and thalidomide have also been reported to cause hypothyroidism. Industrial exposure to polybrominated and polychlorinated biphenyls and resorcinol have been reported to produce hypothyroidism in workers. Although perchlorate is capable of inhibiting thyroid hormone synthesis, this chemical has not been shown to cause hypothyroidism at concentrations reported in contaminated drinking water.[17]

Central (secondary) hypothyroidism Diseases that interfere with TRH production by the hypothalamus or that impair pituitary TSH production can produce central hypothyroidism. The most common causes are pituitary adenomas and the surgical procedures or radiotherapy used to treat them.[18] In addition, tumors impinging on the hypothalamus or pituitary stalk, traumatic transection of the pituitary stalk,[19] and certain infiltrative diseases (e.g., sarcoidosis, hemochromatosis, and Langerhans cell histiocystosis) can interfere with hypothalamic TRH production or delivery. Pituitary thyrotrope dysfunction can be caused by lymphocytic hypophysitis; infection; metastatic disease; apoplexy (e.g., Sheehan syndrome or tumor infarction); and bexarotene, a retinoid X receptor–selective ligand used to treat cutaneous T cell lymphoma.[20]

PATHOGENESIS

Clinical hypothyroidism reflects a widespread lack of thyroid hormone actions at the genomic level in target tissues, where T_3 binds to receptors that are members of the nuclear receptor superfamily.[21] These T_3 receptors are in turn bound to thyroid-response elements located in the regulatory regions of certain genes that increase or decrease their transcription in response to thyroid hormone. Some biochemical and clinical manifestations of hypothyroidism can be explained on the basis of specific deficiencies in molecular actions. For example, reduced expression of the hepatic low-density lipoprotein (LDL) receptor gene decreases LDL cholesterol clearance, causing hypercholesterolemia; decreased expressions of the myocardial α-myosin heavy-chain genes and the sarcoplasmic reticulum adenosine triphosphatase genes impair myocardial systolic and diastolic performance, respectively. Many other clinical aspects of hypothyroidism are not yet understood in terms of specific genomic actions. Some of these may result from putative nongenomic thyroid hormone actions on G protein–coupled membrane receptors and mitochondria.[22]

DIAGNOSIS

Clinical Manifestations

Classic symptoms of hypothyroidism include fatigue, lethargy, cold intolerance, weight gain, constipation, dry skin, hoarseness, slowed mentation, and depressed mood. In a study of patients with short-term hypothyroidism, 38% to 58% of patients had one or more of these clinical findings.[23] However, the diagnostic accuracy of such symptoms is low. Of newly diagnosed hypothyroid patients in a case-control study by Canaris and colleagues, only 30% had any symptoms, and 17% of euthyroid control subjects had one or more of the same nonspecific complaints.[24] As a result, individual symptoms had a positive predictive value of only 8% to 12%.

Inaccuracy in clinical diagnosis of hypothyroidism is attributable to various factors, including the fact that many other disorders produce similar symptoms; the typically gradual onset of thyroid hormone deficiency; and, sometimes, the impaired insight that hypothyroidism produces in some patients. Symptoms that are new or that occur in combination are more likely to rep-

resent hypothyroidism. In the Canaris study, patients with seven or more new symptoms were almost ninefold more likely to be hypothyroid than those with fewer new symptoms. In addition, more hypothyroid patients than euthyroid patients reported that their symptoms had changed from the previous year.[24]

Hypothyroidism can be associated with cognitive deficits, particularly memory problems.[25] Although hypothyroidism is in the differential diagnosis of dementia and is not uncommonly detected in demented elderly patients, thyroid hormone treatment rarely reverses dementia in these patients.[26] Other neurologic findings in hypothyroid patients can include depression, psychosis, ataxia, seizures, and coma. Hypothyroidism is a potentially reversible cause of sleep apnea. It can also cause decreases in the senses of hearing, taste, and smell.

Other special manifestations of hypothyroidism have been reported in children and adolescents. Thyroid hormone deficiency can cause growth failure, delayed or precocious puberty, muscle pseudohypertrophy, and galactorrhea.

Physical Examination

Classic physical signs of hypothyroidism include bradycardia, diastolic hypertension, and hypothermia; coarse, cool, and pale skin; loss of scalp and eyebrow hair; hoarse, slow, and dysarthric speech; distant heart tones; diffuse nonpitting edema; and slowed deep tendon reflexes, particularly during the relaxation phase. However, none of these findings is sufficiently sensitive or specific for diagnosis. Additional signs may be identified when hypothyroid patients present with other unusual features, such as chronic heart failure, pericardial and pleural effusions, ileus and intestinal pseudo-obstruction, or coagulopathy.

In patients with autoimmune thyroiditis, which is the most common type of hypothyroidism, the thyroid gland can be nonpalpable, normal in size, or diffusely enlarged with an irregular contour, firm consistency, and palpable pyramidal lobe. The gland is only rarely painful and tender. There may be signs related to the other endocrine deficiency states associated with the polyendocrine failure syndromes: type 1, which includes hypoparathyroidism (Chvostek and Trousseau signs), adrenal insufficiency (hyperpigmentation), and chronic mucocutaneous candidiasis; and type 2, which includes adrenal insufficiency, type 1 diabetes mellitus, and primary ovarian failure. There can also be evidence of other associated nonendocrine autoimmune disorders, including vitiligo, atrophic gastritis, pernicious anemia, systemic sclerosis, and Sjögren syndrome.

Laboratory Tests

Routine laboratory tests Abnormalities in routine laboratory tests can be the first diagnostic clue suggesting hypothyroidism. Hypercholesterolemia and hyperhomocysteinemia are especially common in hypothyroid patients.[27] In addition, hyponatremia, hyperprolactinemia, hypoglycemia, and elevations in levels of creatine phosphokinase (predominantly MM band) can all be caused by thyroid hormone deficiency.

Serum thyroid function tests Whether it is prompted by clinical or routine laboratory test findings or performed for patient or population screening, measurement of serum TSH should usually be the first test in the diagnosis of hypothyroidism. An elevated serum TSH level identifies patients with primary hypothyroidism regardless of its cause or severity, even those with mild thyroid hormone deficiency and a serum free thyroxine (T_4) concentration within the reference range. Normal serum TSH levels in disease-free populations are typically 0.4 to 4.0 μU/L. However, values are not normally distributed; the mean TSH concentration, 1.5 μU/L, is in the lower half of the reference range.[1] Even a high-normal serum TSH level (e.g., 3.0 μU/L) may reflect very mild thyroid dysfunction, particularly in a patient who has other clinical or laboratory features of autoimmune thyroiditis. As a result, some authorities have recommended lowering the TSH assay's upper limit of normal to 2.5 μU/L.[28]

Table 1 Causes of Elevated Serum TSH Levels

Primary hypothyroidism
Central hypothyroidism*
Recovery after nonthyroidal illnesses
Renal insufficiency
Adrenal insufficiency
Drugs
Metoclopramide
Phenothiazines?
Analytic problems
Anti-TSH antibodies
Anti–mouse immunoglobulin antibodies

*Attributable to TSH with reduced biologic-to-immunologic activity ratio.

When an elevation in serum TSH is detected in a potentially hypothyroid patient, the test should be repeated, and the serum free T_4 concentration should be measured. This further testing confirms the diagnosis of hypothyroidism—an important step, because such patients will typically be committed to lifelong thyroid hormone therapy—and more fully defines the severity of hypothyroidism. The serum T_3 concentration has limited sensitivity and specificity and therefore is a poor test for hypothyroidism.

The TSH assay may fail to detect hypothyroidism in a few settings. In patients with central hypothyroidism, the serum TSH level can be low, normal, or even modestly elevated.[29] The absence of an elevation in the TSH level in a patient with a low free T_4 level is attributable to the synthesis of a TSH molecule that has a decreased ratio of biologic to immunologic activity.[30] Central hypothyroidism should be suspected in the absence of TSH elevation if the patient has clinical features of hypothyroidism; has clinical findings suggesting a sellar mass lesion or other anterior pituitary hormone deficiencies; or has a history of head trauma or conditions known to cause hypopituitarism, such as sarcoidosis. In these settings, both the serum free T_4 and TSH concentrations should be measured. Detection of a low serum free T_4 concentration, regardless of the TSH level, indicates the need for further testing, which may include cranial imaging, performance of a TRH stimulation test to assess TSH responsiveness, and other pituitary function testing.

There are also circumstances in which an elevated serum TSH level may not reflect hypothyroidism [*see Table 1*]. Euthyroid patients with renal or adrenal insufficiency may have modest TSH elevations (e.g., levels of 5 to 10 μU/L). Two rare forms of TSH-mediated hyperthyroidism that may present as clinical and biochemical hyperthyroidism with an inappropriately normal or elevated serum TSH are TSH-secreting pituitary tumors[31] and isolated pituitary resistance to thyroid hormone.[32] However, the elevation in levels of serum free T_4, T_3, or both in these patients provides a clue to the diagnosis. Circulating anti-TSH antibodies can yield falsely elevated TSH immunoassay readings.

Effects of nonthyroid illnesses and drugs Distinguishing central hypothyroidism from the thyroid function abnormalities that often accompany severe nonthyroid illnesses can be challenging. Cytokine-mediated TSH suppression can mask mild primary hypothyroidism. Furthermore, certain drugs used to treat severe illness (e.g., glucocorticoids, dopamine, and dobutamine) can normalize elevated serum TSH concentrations in patients with overt primary hypothyroidism. Conversely, false positive transient TSH elevation can be seen in patients recovering from critical illness.[33] Consequently, with severely ill patients, it is best to limit thyroid function testing to those in whom there is a significant clinical suspicion of hypothyroidism; otherwise, abnormal results are much more likely to represent false positive than true positive findings. Similarly, the antiseizure medications phenytoin and carbamazepine can cause decreases in the levels of serum total T_4, free T_4 (as measured by immunoassay), and TSH; these findings can be confused with those of central hypothyroidism.[34] In some patients who are severely ill or who are taking these antiseizure medications, free T_4 measurement by equilibrium dialysis and pituitary imaging may be required to diagnose or exclude central hypothyroidism.

DIFFERENTIAL DIAGNOSIS

Given that the clinical manifestations of hypothyroidism are quite nonspecific and can be caused by myriad other medical conditions and life circumstances, the key to diagnosis is simply for the physician to keep this condition in mind. Once the possibility of hypothyroidism is entertained, serum TSH measurement can confirm or exclude the diagnosis in almost all cases. In a survey of 1,721 primary care physicians, 80% to 90% appreciated the fact that a middle-aged woman presenting with fatigue, impaired memory, or depression might have hypothyroidism and therefore would order a serum TSH concentration for such a patient; however, only half of these physicians would screen for hypothyroidism in a hypercholesterolemic patient.[35]

The cause of primary hypothyroidism may be evident from the history alone; for example, the patient may have previously undergone thyroid surgery or radiation therapy or may currently be taking medications known to cause hypothyroidism. When the history provides no clue, sustained primary hypothyroidism can usually be assumed to be caused by autoimmune thyroiditis. Confirmatory laboratory tests are seldom required. Nonetheless, it is sometimes helpful to confirm this diagnosis by detection of thyroid autoantibodies. Anti–thyroid peroxidase antibody assay is the most sensitive test to confirm the diagnosis of autoimmune thyroiditis. Thyroid autoantibody testing can also be useful in predicting the development of hypothyroidism in patients with mild hypothyroidism and in pregnant and postpartum women.[36-38]

MANAGEMENT

Thyroid Hormone Therapy

Levothyroxine sodium (thyroxine) is the treatment of choice for patients with hypothyroidism. Thyroxine is well absorbed by the proximal small bowel. Thyroxine circulates with a 7-day half-life because of plasma protein binding, and it is metabolized in target tissues, in part by deiodination to T_3. Its long half-life permits a single daily dose; its conversion to T_3 in target tissues mimics normal physiology. The multiple dose strengths available in North America facilitate precise dose titration. Nonetheless, thyroxine and other thyroid hormone preparations have narrow therapeutic indexes and hence have the potential for adverse reactions with even modest overtreatment. Several studies examining the adequacy of thyroid hormone therapy in large populations and in patients in generalist and specialty practices have found that one fifth of patients with treated hypothyroidism are receiving an inadequate dose and one fifth an excessive dose.[39]

Dosing considerations and drug interactions The optimal thyroxine dosage for hypothyroid patients is related to body weight. In adults, this is approximately 1.8 μg/kg/day.[40] Elderly patients, whose metabolic clearance of thyroxine is reduced, have a lower dosage requirement of 0.5 μg/kg/day. The thyroxine dose is usually higher in patients who have undergone thyroidectomy than in patients with autoimmune thyroiditis, who often have residual functioning thyroid tissue. Thyroxine absorption can be decreased in patients with malabsorption from gastrointestinal disorders or previous small bowel bypass surgery. Several mineral supplements, medications, and dietary constituents can interfere with thyroxine absorption; these include iron, calcium carbonate, cholestyramine, aluminum hydroxide gel, sucralfate, soy, and perhaps dietary fiber. Metabolism of thyroxine is accelerated in the nephrotic syndrome, in other severe systemic illnesses, and with the use of phenobarbital, phenytoin, carbamazepine, and rifampin. In 75% of pregnant women, the thyroxine dose requirement is increased by 50% to 100%.[41] Postmenopausal hormone replacement therapy increases the required thyroxine dose in 35% of women.[42]

Patient noncompliance is the most common cause of inadequate thyroxine therapy. Several observations should raise suspicion that a patient is not taking thyroxine faithfully: the apparent thyroxine dose requirement is higher than expected; thyroid function test results vary without correlation with prescribed thyroxine doses; and the serum TSH concentration is elevated, yet the serum free T_4 level is in the mid- to high-normal range, reflecting improved compliance immediately before testing.

Thyroxine treatment should typically start with a dosage at the lower end of the anticipated requirement (e.g., 125 μg/day in a 70 kg adult). In otherwise healthy younger patients, there is no need to titrate the dose upward from a very low starting dose. Laboratory monitoring of treated hypothyroid patients should be performed 4 to 6 weeks after starting a new thyroxine dose or tablet formulation; thereafter, it should be performed annually. It should also be performed whenever a patient's symptoms suggest thyroid hormone deficiency or excess. The goal for most patients is to restore the TSH level to the lower half of the normal range (i.e., 1.0 to 2.0 μU/L). In patients with central hypothyroidism, the serum free T_4 concentration must be monitored; treatment should usually be targeted for a concentration in the upper half of the normal range.

Metabolism of certain other drugs can be affected by the hypothyroid state and by the initiation of thyroxine treatment. Hypothyroid patients may have increased sensitivity to anesthetic and sedative agents. Reduced digoxin clearance and drug distribution volume may predispose patients to toxicity. Sensitivity to warfarin may be decreased because of slowed metabolism of vitamin K–dependent clotting factors, and restoring euthyroidism can increase the required warfarin dose.

Adverse reactions to thyroid hormone therapy Adverse reactions to thyroxine overtreatment include symptomatic thyrotoxicosis and subclinical thyrotoxicosis with increased risks of bone loss and atrial tachyarrhythmias.[43,44] The predisposition to

osteoporosis is principally in postmenopausal women. Atrial fibrillation is more common in patients 60 years of age or older. Both of these complications have been shown to occur when the serum TSH concentration is suppressed to less than 0.1 μU/L.

Complications can also arise from restoring euthyroidism, particularly in patients with underlying ischemic heart disease[45] (see below) and borderline adrenal cortical insufficiency. Concomitant thyroid and adrenal gland failure can occur in hypopituitarism and in the type 2 polyendocrine failure syndrome (Schmidt syndrome), which is marked by autoimmune thyroiditis and idiopathic adrenal insufficiency.

A few patients experience acute sympathomimetic symptoms soon after institution of thyroxine treatment. This syndrome is poorly understood; it can be circumvented by reducing the thyroxine dose to a very low level and advancing it slowly.

Transient scalp hair loss may occur during first few weeks of thyroxine replacement therapy. Patients can be assured that this phenomenon is temporary. Treatment of hypothyroidism sometimes reveals an underlying urticarial disorder, but true allergy to thyroxine formulations has not been well documented.

Special Therapeutic Issues

Hypothyroid patients with ischemic heart disease Because thyroid hormone has positive inotropic and chronotropic effects, thyroid hormone therapy can exacerbate myocardial ischemia in hypothyroid patients with underlying coronary artery disease. In such patients, thyroxine therapy should be initiated at a low dosage (e.g., 25 μg/day) and titrated upward in increments of 12.5 to 25 μg every 4 to 6 weeks. Patients should be monitored vigilantly with clinical assessments and electrocardiography. Deliberate suboptimal dosing, which was previously advocated to limit myocardial oxygen demand, has been shown to actually increase the risk of progressive coronary atherosclerosis. Beta-blocker therapy should sometimes be initiated or intensified when thyroxine therapy is initiated. Hypothyroid patients who experience worsening myocardial ischemia despite these precautions can undergo coronary angioplasty and even surgical bypass grafting with minimal or no increased perioperative risk.[46,47]

Mild hypothyroidism Whether to identify and treat patients with mild hypothyroidism, defined by an elevated serum TSH level with a normal free T_4 level, is controversial. There is agreement that mild hypothyroidism is highly prevalent, particularly in older women, and that clinical diagnosis is inaccurate. Diagnostic serum TSH testing and thyroxine treatment of mild hypothyroidism are clearly effective and are relatively safe and inexpensive. The outstanding issue is whether mild hypothyroidism causes clinical consequences that are important enough to justify widespread screening and therapy.[48] Proponents of detection and treatment argue that it prevents progression to overt hypothyroidism in affected patients, particularly those whose serum TSH concentration is greater than 10 μU/L, who are 65 years of age or older, or who have thyroid autoantibodies, indicating underlying autoimmune thyroiditis. Advocates believe that treatment of mild hypothyroidism may reduce the risk of future atherosclerotic cardiovascular disease. They hold this view on the basis of the following observations: affected patients have higher mean cholesterol levels; most studies have shown that TSH-normalizing thyroxine therapy lowers serum total cholesterol and LDL cholesterol concentrations[49]; and some epidemiologic studies have found that persons with mild hypothyroidism have a higher risk of atherosclerotic cardiovascular disease.[50] Some proponents are persuaded by four small, controlled, double-blind trials that showed that thyroxine therapy was more effective than placebo in improving symptoms and neuropsychologic performance in patients with mild hypothyroidism.[51] On the basis of these studies, two decision and cost-effectiveness models suggested that the cost-effectiveness of screening for and treating mild hypothyroidism is comparable to that of other widely accepted preventive medicine strategies.[52,53] On the other hand, opponents of screening and treatment of mild hypothyroidism point out that these putative benefits have not been rigorously confirmed by large, randomized, controlled trials.[54] When physicians do recommend treatment for patients with mild hypothyroidism, the thyroxine dosage is typically lower than that for overt hypothyroidism—0.5 μg/kg/day.

Residual hypothyroid symptoms and T_3 therapy Compared with euthyroid patients, hypothyroid patients more often have constitutional and neuropsychological complaints, even when serum TSH measurements suggest adequate treatment.[55] This observation may represent only ascertainment bias (i.e., symptomatic patients seeking medical care are more likely to be diagnosed and treated for hypothyroidism). However, it has been postulated that the presence of residual symptoms in thyroxine-treated patients reflects a failure to replace the small amount of T_3 normally secreted by the thyroid gland. Four clinical trials in which a fraction of the thyroxine dose was replaced with a small dose of T_3 failed to confirm an earlier report of significant improvement with combination thyroxine/T_3 therapy.[56] Combination therapy has the disadvantages of a fluctuating and supraphysiologic T_3 level, a greater risk of iatrogenic thyrotoxicosis, and increased complexity and expense. Treatment with desiccated thyroid, a biologic preparation that also contains both T_4 and T_3, has the same disadvantages.

COMPLICATIONS

Severe hypothyroidism (myxedema) can become complicated by multiple organ system failure when it is profound and prolonged, especially in elderly patients who have other cardiac, pulmonary, neurologic, renal, and infectious diseases. Myxedema coma, the most severe expression of hypothyroidism, is associated with substantial mortality. Such complications of thyroid hormone deficiency can be prevented with sustained thyroxine therapy. In newly diagnosed patients, preventive measures also include giving special attention to other potentially provocative medical conditions (e.g., heart failure, renal failure, pneumonia) and medications—particularly sedative, anesthetic, and analgesic medications that suppress ventilatory drive and other central nervous system functions.

Treatment of complicated hypothyroidism includes thyroid hormone replacement and aggressive management of organ system complications that can be present. Two thyroid hormone regimens have proven efficacy for myxedema coma: (1) thyroxine in a full replacement dose (1.8 μg/kg/day), with or without a 500 μg loading dose to replete the normal body thyroxine pool[57]; and (2) T_3 in divided doses, advocated because of the impaired T_4-to-T_3 conversion that occurs in critically ill patients. No trial has rigorously compared these regimens, but one small retrospective study found a higher mortality in T_3-treated patients.[58]

PROGNOSIS

The prognosis for hypothyroid patients who are properly

treated with thyroxine should be excellent. However, discontinuance of thyroid hormone therapy predictably leads to recurrent hypothyroidism, with its potential for serious complications in the elderly. This occurs most often in settings of social neglect, poor access to health care, and associated neuropsychological impairment. Lesser degrees of suboptimal therapy are also associated with long-term risks. Inadequately treated patients may have increased risk of atherosclerotic cardiovascular disease, and iatrogenic thyrotoxicosis can predispose patients to osteoporosis and atrial tachyarrhythmias.

Patients with autoimmune thyroiditis, the most common cause of hypothyroidism, are at risk for certain associated conditions, for which they should be monitored. Pernicious anemia and gastric achlorhydria with consequent iron and calcium malabsorption affect 3% and 25% of autoimmune thyroiditis patients, respectively. Much less commonly, other autoimmune diseases (e.g., Sjögren syndrome and systemic sclerosis), endocrine deficiency states (adrenal insufficiency, type 1 diabetes, hypoparathyroidism, and hypogonadism), and primary thyroid lymphoma can occur.

Thyrotoxicosis

EPIDEMIOLOGY

The alert clinician will diagnose thyrotoxicosis several times each year. NHANES III found thyrotoxicosis in 0.5% of a surveyed cohort that reflected the demographics of the United States adult population.[1] Three disorders account for the majority of cases: diffuse toxic goiter (Graves disease), toxic nodular goiter, and iatrogenic thyrotoxicosis in thyroid hormone–treated patients. The incidence of Graves disease in one United Kingdom community survey was one to two cases per 1,000 population annually; 2.7% of women and 0.2% of men had Graves disease or a history of Graves disease.[59] The highest incidence of Graves disease is in women 30 to 60 years of age, but the disease can affect persons of virtually any age, from neonates to the very elderly. Toxic adenoma and toxic multinodular goiter are more common causes of thyrotoxicosis than Graves disease in regions where dietary iodine deficiency is prevalent; in women; and in older patients.[60] Iatrogenic thyrotoxicosis has been reported in approximately 20% of thyroid hormone–treated patients.[1,39,61]

ETIOLOGY, GENETICS, AND PATHOGENESIS

Thyrotoxicosis can be divided into three etiologic categories: abnormal stimulation of the thyroid gland, thyroid gland autonomy, and gland inflammation with unregulated thyroid hormone release. Each of these categories includes several diseases [*see Table 2*].

Table 2 Etiologic Classification of Thyrotoxicosis

Cause	*Individual Diseases*
Abnormal stimulation of the thyroid gland	Graves disease hCG-mediated thyrotoxicosis TSH-mediated thyrotoxicosis
Thyroid gland autonomy	Toxic adenoma Toxic multinodular goiter Congenital thyrotoxicosis Iodine-induced hyperthyroidism Thyroid cancer–related thyrotoxicosis
Gland inflammation with unregulated thyroid hormone release	Subacute (de Quervain) thyroiditis Lymphocytic thyroiditis Amiodarone-induced thyrotoxicosis, type 2 Acute thyroiditis

hCG—human chorionic gonadotropin

Graves Disease (Diffuse Toxic Goiter)

There is compelling evidence that there is a genetic predisposition to Graves disease, that the incidence is higher in women, that unknown environmental factors are involved in its initiation, and that gland stimulation by antibodies against the TSH receptor is the immediate precipitant of the condition. Identical twins and some families show increased incidences of Graves disease.[62] The condition has been genetically linked to certain MHC components (e.g., HLA-B8 and HLA-DR3), which are on the surface of cells that present antigenic peptide epitopes to T cell receptors. One theory is that certain HLA-DR molecules may be better able to present TSH receptor epitopes, inciting autoimmunity. Another hypothesis is that these HLA-DR recognition sequences are involved in aberrant thymic T cell selection for tolerance. Graves disease has also been linked to polymorphisms in the gene encoding CTLA-4, a T cell receptor important for interaction with antigen-presenting cells.[63] In whites, a susceptibility locus for Graves disease has been identified on chromosome 20q11.

Several environmental factors have been implicated in the initiation of Graves disease. These include stressful life events, smoking, large amounts of dietary iodine, and preceding infection with certain bacterial agents that have been postulated to induce molecular mimicry. Radiation injury to the thyroid gland may increase the risk of the condition, possibly because of increased TSH receptor exposure and immunoreactivity.

Whatever the underlying genetic and environmental factors, the vast majority of Graves disease patients have detectable antibodies that are directed against the TSH receptor and are capable of stimulating it[64] [*see Figure 2*]. Assays using thyroid cells or their membranes can detect circulating TSH receptor autoantibody species in 70% to 90% of patients with Graves disease. These autoantibodies are capable of stimulating intracellular cyclic adenosine monophosphate production (thyroid-stimulating immunoglobulins [TSI]), inhibiting TSH receptor activation (TSH receptor inhibitory immunoglobulins), and inhibiting the binding of TSH to its receptor (TSH receptor–binding inhibitory immunoglobulins [TBII]).

Toxic Adenoma and Toxic Multinodular Goiter

Solitary and multiple thyroid adenomas and diffusely hyperplastic thyroid tissue possess a growth advantage, and their constituent thyrocytes sometimes produce thyroid hormones autonomously (i.e., without regard to TSH regulation). These hyperplastic and neoplastic conditions cause hyperthyroidism when the mass and efficiency of functioning thyroid tissue are great enough to generate hormone excess in target tissues, including suppression of endogenous pituitary TSH production and function of extranodular thyroid tissue.[65] Both genetic and environmental factors are involved in the development of this autonomous function. A twin study showed that genetic factors could account for 82% of the predisposition to nodular goiter, and familial multinodular goiter has been linked to a gene locus on chromosome 14q.[66] At the same time, environmental factors (e.g., dietary iodine deficiency, goitrogens, and radiation exposure) also clearly predispose to the development of autonomous-

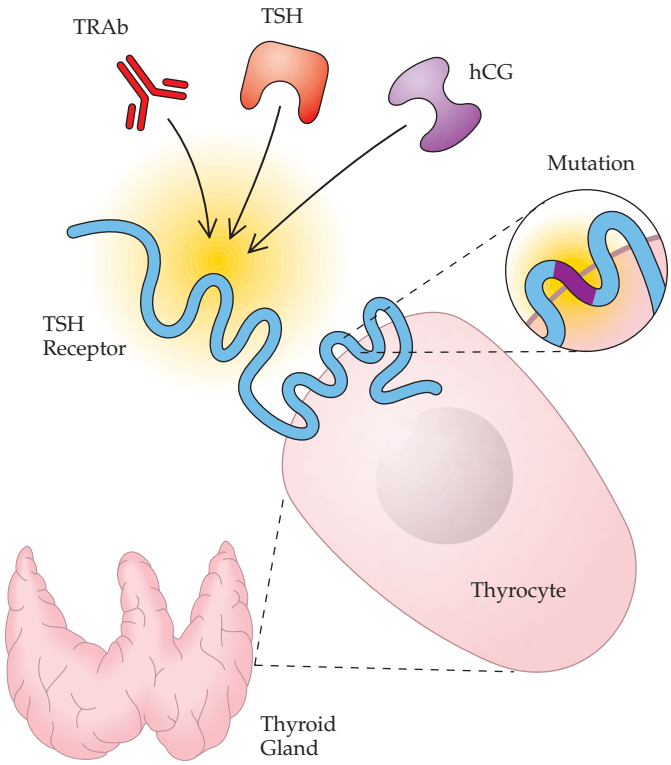


Figure 2 **Certain forms of hyperthyroidism result from deranged physiologic activation of the thyroid-stimulating hormone (TSH) receptor. In Graves disease, TSH receptor autoantibodies (TRAb) bind the TSH receptor. In patients with TSH-secreting pituitary tumors, the autonomously secreted TSH overstimulates the receptor. In molar pregnancy and choriocarcinoma, high concentrations and aberrant forms of human chorionic gonadotropin (hCG) can activate the TSH receptor. In some patients with toxic adenomas, a constitutively activating somatic mutation of the TSH receptor results in autonomous secretion of thyroid hormone.**

ly functioning thyroid tissue. Genetic and environment factors promote thyroid tissue growth by activating intraglandular growth factors (e.g., insulinlike growth factor and epidermal growth factor receptor)[67] and signaling pathway proteins (e.g., $G_s\alpha$ and ras). Constitutively activating somatic mutations of the TSH receptor and $G_s\alpha$ itself have been described in 25% to 80% of toxic adenomas, more commonly in patients from regions where dietary iodine deficiency is prevalent.[68]

Iodine-Induced Hyperthyroidism

Iodine is both a substrate and a physiologic regulator of thyroid hormone synthesis. Excessive iodine intake normally inhibits thyroid hormone production by reducing the trapping of inorganic iodide and its oxidation into an organic form (organification) and by thyroid hormone release. At the same time, exposure to pharmacologic amounts of iodine (typically 1,000-fold more than the physiologic requirement of 150 μg/day) can cause hyperthyroidism, a condition termed the Jod-Basedow effect. Patients with hyperplastic and benign neoplastic thyroid conditions and those with latent Graves disease are particularly vulnerable to iodine-induced hyperthyroidism. Epidemics of thyrotoxicosis have repeatedly been observed when iodine supplementation is instituted in regions of previous dietary iodine deficiency.[69] However, iodine-induced hyperthyroidism can also occur in patients from iodine-sufficient environments whose thyroid glands are apparently normal, especially when excess iodine exposure is substantial and sustained, as it is with long-term amiodarone therapy.[70] The precise molecular and biochemical basis for iodine-induced hyperthyroidism is poorly understood. Iodine-induced hyperthyroidism is typically transient, lasting only a few weeks, but more prolonged thyroid dysfunction can occur when iodine exposure is prolonged, as occurs with the lipid-soluble drug amiodarone and with myelographic radiocontrast agents.

Thyroiditis

Inflammation of thyroid tissue caused by infectious diseases, autoimmune processes, or pharmacologic toxicity can cause thyrocyte death, disruption of follicular architecture, and unregulated leakage of thyroid hormones from the gland into the circulation, resulting in thyrotoxicosis[71] [*see Table 3*]. Thyroiditis-related thyrotoxicosis is typically self-limited, lasting 2 to 8 weeks, with spontaneous resolution once glandular stores of thyroid hormone are exhausted. A comparable period of transient hypothyroidism often follows because of lingering impairment of thyroid hormone synthesis, but most patients ultimately become euthyroid [*see Figure 3*].

Amiodarone-Induced Thyrotoxicosis

The iodine-containing antiarrhythmic agent amiodarone can cause thyrotoxicosis by two mechanisms.[72] Type 1 amiodarone-induced thyrotoxicosis is caused by iodine, whereas type 2 amiodarone-induced thyrotoxicosis is the result of gland inflammation. Both forms can be severe, prolonged, and life-threatening, particularly because affected patients have underlying cardiac disease.

Chorionic Gonadotropin–Mediated Hyperthyroidism

Human chorionic gonadotropin (hCG), which is structurally similar to TSH, can stimulate the TSH receptor and increase thyroid function when circulating in high concentration or when variant forms of either hCG or the TSH receptor increase the affinity of their hormone receptor interaction[73] [*see Figure 3*]. In fact, during the first trimester of normal pregnancy, when a marked physiologic elevation of hCG occurs, a modest rise in the serum free T_4 level and a decline in the serum TSH level are typically seen.[74] An exaggeration of this phenomenon can cause thyrotoxicosis, as can trophoblastic tumors.

Trophoblastic tumors Women with hydatidiform mole and choriocarcinoma, as well as men with metastatic testicular choriocarcinoma, can develop hyperthyroidism as a result of very high concentrations of circulating hCG.[75] Furthermore, these tumors have been shown to produce a variant form of hCG with heightened TSH receptor stimulatory properties.

Gestational transient thyrotoxicosis Mild transient thyrotoxicosis occurs late in the first trimester of pregnancy in 1% to 3% of white women and in as many as 11% of Asian women.[76] The serum hCG level is higher in affected pregnant women than in those who remain euthyroid. Furthermore, gestational thyrotoxicosis appears to be more common in women who have hyperemesis gravidarum or twin pregnancies, both of which are characterized by higher serum hCG concentrations. A rare form of familial gestational thyrotoxicosis has been reported in which a mother and daughter both had recurrent hy-

Table 3 Characteristic Features of Thyroiditis

Form of Thyroiditis	*Presumed Etiology*	*Classic Pattern of Thyroid Dysfunction*	*Other Clinical Manifestations*	*Treatment*
Autoimmune	T cell–mediated autoimmunity	Hypothyroidism	Firm small-to-medium goiter	Thyroxine for hypothyroidism
Lymphocytic (painless, silent, postpartum)	T cell–mediated autoimmunity?	Transient thyrotoxicosis followed by hypothyroidism	Painless small goiter	Observation, beta blockade for thyrotoxicosis, thyroxine for hypothyroidism
Subacute (de Quervain)	Viral infection?	Transient thyrotoxicosis followed by hypothyroidism	Painful and tender hard goiter	NSAID or glucocorticoid, beta blockade for thyrotoxicosis, thyroxine for hypothyroidism
Acute (suppurative)	Bacterial, fungal, and protozoal infections	Thyroid dysfunction (rare)	Painful, tender, and inflamed goiter	Antibiotic therapy, surgical drainage
Amiodarone-induced type 1	Iodine-induced hyperthyroidism	Thyrotoxicosis	Normal-size nontender gland	Thionamide antithyroid medication
Amiodarone-induced type 2	Inflammatory thyroiditis, precise cause unknown	Transient thyrotoxicosis	Normal-size nontender gland	Glucocorticoids
Riedel (invasive fibrous)	Idiopathic fibrosis?, autoimmune?	Hypothyroidism in one third of patients	Enlarging, hard, fixed mass	Surgery, glucocorticoids, tamoxifen

NSAID—nonsteroidal anti-inflammatory drug

perthyroidism during their pregnancies and were found to have a mutant TSH receptor with increased affinity and signaling responsiveness to hCG.[77]

TSH-Mediated (Central) Hyperthyroidism

Hyperthyroidism can be caused by excessive TSH secretion in two rare conditions: TSH-secreting pituitary adenoma and the syndrome of isolated central resistance to thyroid hormone.[78] Excessive and relatively autonomous TSH production by pituitary tumors predictably results in goitrous hyperthyroidism.[79] The TSH produced by TSH-secreting pituitary tumors has increased bioactivity, and normal inhibition of TSH release by dopamine has been shown to be defective in these patients. However, the fundamental cause of these tumors, which often cosecrete other pituitary hormones, is unknown.

Isolated central resistance to thyroid hormone is a rare inherited condition in which impaired negative feedback of thyroid hormone on pituitary thyrotropes leads to TSH hypersecretion and hyperthyroidism.[80] In one patient with isolated central resistance to thyroid hormone, a novel mutation in the thyroid hormone receptor–β gene was identified. In patients with this syndrome, unlike those with generalized resistance to thyroid hormone, other target tissues for thyroid hormone, such as the brain, heart, and liver, respond normally to the resulting thyrotoxicosis.

Exogenous Thyrotoxicosis

Iatrogenic thyrotoxicosis is relatively common, occurring in 20% of thyroid hormone–treated patients.[1,39,61] Possible explanations for this condition include improved patient compliance with therapy, decreased metabolic clearance of thyroid hor-

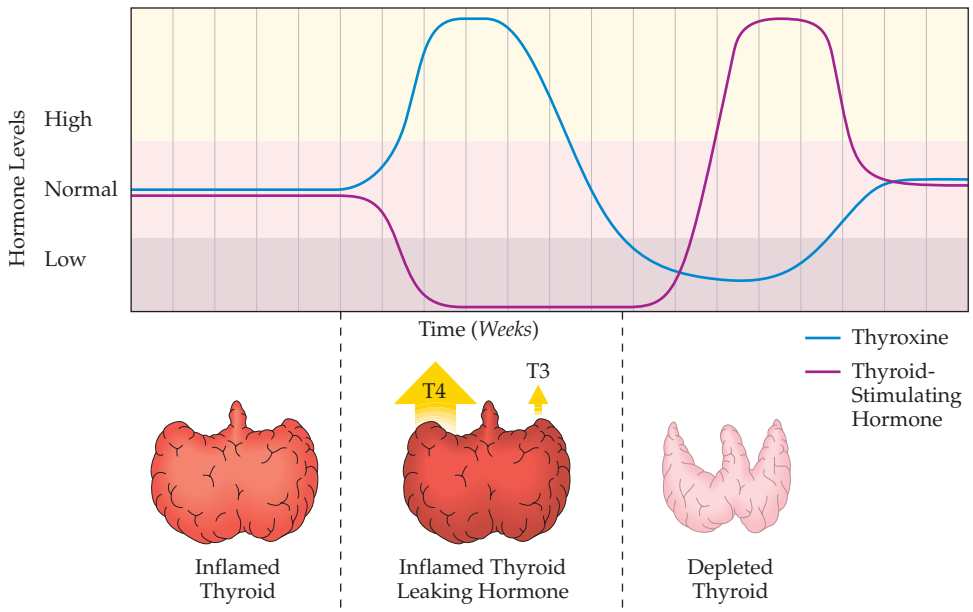


Figure 3 **In acute thyroiditis, inflammation of thyroid tissue leads to unregulated leakage of thyroid hormones from the gland into the circulation. The resulting thyrotoxicosis typically lasts 2 to 8 weeks and ends spontaneously as the glandular hormone stores are exhausted. A comparable period of hypothyroidism often follows, because of impaired thyroid hormone synthesis, but in most patients, the gland gradually returns to normal function.**

mones with aging, a substantial decrease in body weight, an increase in underlying gland function in patients with treated Graves disease or nodular goiter, and discontinuance of medications that interfere with thyroid hormone absorption or that accelerate its metabolism. Factitious thyrotoxicosis is sometimes prompted by a desire to enhance energy and weight loss; it can also occur through the complex psychopathology of Munchausen syndrome. Accidental or suicidal thyroid hormone intoxication can be life-threatening; its clinical manifestations may take 12 to 48 hours to become fully expressed, necessitating close observation even of asymptomatic patients, especially children.[81]

DIAGNOSIS

Clinical Manifestations

The classic symptoms of thyrotoxicosis are familiar to every third-year medical student—weight loss despite good appetite, heat intolerance, tremor, palpitations, and anxiety—yet even experienced clinicians are often slow to make the diagnosis, for several reasons. First, many common symptoms of thyroid hormone excess are nonspecific, such as fatigue, insomnia, dyspnea, and atypical chest pain. Second, patients can present with atypical chief complaints: weight gain; anorexia, nausea, and vomiting; muscle weakness; headache; urticaria; and, in elderly patients, apathy without sympathomimetic symptoms. Severe thyrotoxicosis may also present as heart failure, delirium, or an apparent febrile illness. Third, thyrotoxicosis can occur without the full complement of findings associated with Graves disease (e.g., prominent goiter and ocular findings). For example, thyrotoxicosis can be overlooked in a postpartum woman with weight loss and anxiety from acute lymphocytic thyroiditis, in a middle-aged man with bilateral earache reflecting radiation pain from subacute thyroiditis, and in the older patient with "failure to thrive" related to toxic nodular goiter. Fourth, new thyrotoxic complaints often arise in patients who have been otherwise entirely well and in whom the symptoms can potentially be discounted as a minor intercurrent illness or life stress. Finally, the spectrum of thyrotoxicosis includes entirely asymptomatic disease that nonetheless can have potential health consequences related to mild thyroid hormone excess.

A history of exposure to certain drugs, radiocontrast dye, homeopathic or traditional medicines, and dietary supplements can sometimes be the key to diagnosis. For example, a history of therapy with thyroid hormone, amiodarone, or interferon alfa suggests both the possibility and the likely cause of thyrotoxicosis. A history of recent radiocontrast studies or the recent ingestion of kelp may suggest iodine-induced hyperthyroidism. The family history is also often important in revealing a predisposition to autoimmune thyroid disease, nodular goiter, or, in rare cases, inherited forms of thyrotoxicosis.

Physical Examination

Physical signs related to thyrotoxicosis are often the key to diagnosis and differential diagnosis. Classic signs accompanying thyrotoxicosis can include an anxious, hyperactive demeanor and pressured speech; tachycardia, systolic hypertension, and widened pulse pressure; velvety, warm, and moist skin; onycholysis; flaxen, oily hair; staring gaze and lid lag; prominent apical impulse and systolic flow murmur; and proximal leg muscle weakness and tremor.

Certain findings on physical examination are characteristic of specific etiologies of thyrotoxicosis [*see Table 3*]. In Graves disease, patients typically have a symmetrical, rubbery goiter that is nontender and smooth or subtly lobulated; an audible bruit is sometimes noted. They may also have subtle or prominent eye findings, including episcleral injection, conjunctival swelling, periorbital edema, proptosis, limitation of extraocular motility, and impaired visual acuity or color vision. Less commonly, these patients may have pretibial myxedema, an orange peel–like thickening of the soft tissues of the anterior aspect of the lower leg from subcutaneous mucopolysacharide deposition; rarely, they may have clubbing of the fingers. Graves disease patients may also have physical signs of associated disorders, such as vitiligo and prematurely gray hair, which often escape detection without specific inquiry.

Other findings may suggest other etiologies for thyrotoxicosis. A solitary palpable thyroid nodule or multinodular goiter suggests the possibility of toxic adenoma or toxic multinodular goiter, respectively. Modest thyroid enlargement with an exquisitely tender, wood-hard gland may represent subacute thyroiditis. Thyrotoxic symptoms and signs in a pregnant woman may reflect hCG-related hyperthyroidism or, in a postpartum woman, acute lymphocytic thyroiditis. Signs of an expanding sellar mass lesion or other syndromes of pituitary hormone excess (e.g., acromegaly, Cushing syndrome, or galactorrhea) may suggest the presence of a TSH-secreting pituitary adenoma.

Laboratory Tests

Routine laboratory tests Abnormalities on routine laboratory studies can be the first clue to thyrotoxicosis. Such abnormalities include hypercalcemia, an elevated serum alkaline phosphatase concentration, and a serum total or LDL cholesterol concentration that is either low or that is lower than previously documented for that patient. Serum ferritin, angiotensin-converting enzyme, and testosterone-binding globulin concentrations are all increased in thyrotoxicosis, and such increases may suggest the diagnosis. New significant atrial arrhythmias detected by electrocardiography, particularly atrial fibrillation, mandate testing for thyrotoxicosis.

Serum thyroid function tests Serum TSH measurement is a highly sensitive way to diagnose or exclude all common forms and degrees of thyrotoxicosis.[82] Physiologic inhibition of pituitary thyrotrope function by thyroid hormones results in a serum TSH concentration that is low—almost invariably, less than 0.1 μU/L in patients with thyrotoxicosis. When the TSH assay is employed to diagnose thyrotoxicosis, it must have a detection limit low enough to distinguish normal from low values; a functional sensitivity to less than 0.02 μU/L has been recommended.[83] The serum TSH concentration is so sensitive in detecting thyroid hormone excess that it can be suppressed even when a patient's serum thyroid hormone concentration rises but remains within the reference range for that population—so-called subclinical thyrotoxicosis (see below).

In a few circumstances, however, TSH measurement can be inaccurate in the diagnosis of thyrotoxicosis. First, in patients with rare forms of TSH-mediated thyrotoxicosis (see above), the serum TSH concentration can be elevated, inappropriately normal, or only modestly decreased (i.e., 0.1 to 0.5 μU/L). Second, spurious elevations of the measured TSH level, masking thyrotoxicosis, can occur with rare analytic problems, such as the presence of interfering anti-TSH autoantibodies. Third, there are other causes of a low serum TSH level, including central hypothyroidism and severe nonthyroidal illnesses. Whenever one of

Table 4 Causes of Elevated Serum Total Thyroxine Level

- Thyrotoxicosis
- Increased serum protein binding
  - Increased serum thyroxine-binding globulin concentrations
    - Inherited
    - Estrogen (pregnancy, exogenous, tumor produced)
    - Hepatitis
    - Hepatoma
    - HIV infection
    - Drugs (methadone, heroin, clofibrate, 5-fluorouracil)
  - Familial dysalbuminemic hyperthyroxinemia
  - Increased serum transthyretin binding or concentrations
    - Inherited
    - Carcinoma of the pancreas
    - Hepatoma
- Inhibition of T_4-to-T_3 conversion
  - Medical illnesses
  - Drugs (high-dose propranolol, amiodarone)
- Test artifacts (assay interference from anti-T_4 immunoglobulins)

these circumstances is suspected, serum free T_4 and T_3 levels should be obtained to rule out thyrotoxicosis definitively.

Serum T_4 and T_3 measurements are useful to confirm the diagnosis of thyrotoxicosis, define its severity, and monitor the response to treatment. However, elevated serum total thyroid hormone concentrations are not specific for thyrotoxicosis[84] [*see Table 4*]. Because most of the circulating thyroid hormones are bound to plasma proteins (e.g., thyroxine-binding globulin, transthyretin [thyroxine-binding prealbumin], and albumin), conditions that increase the concentration or binding affinity of these proteins can cause euthyroid hyperthyroxinemia—an increase in the total serum T_4 level without elevation of the small fraction (0.03% for T_4) of biologically active free hormone. The most common such condition is the estrogen-induced increase in thyroxine-binding globulin level that occurs in women who are pregnant or who are taking estrogen preparations. Conversely, a decrease in binding of thyroid hormone by plasma proteins, such as occurs with nephrotic syndrome or androgen use, can mask the diagnosis of thyrotoxicosis on the basis of total T_4 measurement.

The serum free (or unbound) T_4 concentration can help distinguish thyrotoxicosis from euthyroid hyperthyroxinemia. Although equilibrium dialysis is the most accurate approach to free T_4 measurement, it is technically demanding and few laboratories perform it. Free T_4 immunoassays are now widely available and relatively inexpensive. They provide much the same information and have largely supplanted the free T_4 index, which provides an estimate of the unbound T_4 concentration on the basis of partition of radiolabeled thyroid hormone between plasma proteins and a binding resin. Both the free T_4 immunoassay and free T_4 index can reliably differentiate between the hyperthyroxinemia of thyrotoxicosis and that associated with thyroxine-binding globulin elevation.

Certain other conditions causing euthyroid hyperthyroxinemia still cannot be reliably differentiated from thyrotoxicosis with conventional methods of measuring free T_4. For example, free T_4 immunoassays often report falsely elevated values in patients with familial dysalbuminemic hyperthyroxinemia, in which a mutant albumin binds T_4 with increased affinity.[85] Similarly, increased transthyretin binding of thyroxine caused by a mutant transthyretin gene or acquired transthyretin overproduction by hepatic or pancreatic neoplasms can yield deceptively elevated free T_4 immunoassay values.[86] T_4-binding autoantibodies, which occasionally develop in patients with autoimmune thyroiditis, can cause spurious serum T_4 elevation.[87] Hyperthyroxinemia can also occur with disorders and medications that reduce T_4 clearance, including acute systemic illnesses, psychosis, and treatment with amiodarone or high-dose propranolol. Finally, patients with the syndrome of generalized resistance to thyroid hormone typically have elevated serum total and free T_4 and T_3 concentrations.

In summary, hyperthyroxinemia is not pathognomonic of thyrotoxicosis. Clinical information—such as the presence of symptoms and signs of thyrotoxicosis or other conditions or the use of medications associated with hyperthyroxinemia—often permit a straightforward differentiation of thyrotoxicosis from euthyroid hyperthyroxinemia. Serum TSH measurement is invaluable in distinguishing all common forms of thyrotoxicosis, in which serum TSH is low, from euthyroid hyperthyroxinemia, in which serum TSH is usually normal.

Serum total and free T_3 concentrations are elevated in most patients with thyrotoxicosis caused by increased thyroid T_3 production and increased extrathyroid conversion of T_4 to T_3. Less than 5% of hyperthyroid patients have T_3 thyrotoxicosis (i.e., a high serum T_3 concentration and a normal serum T_4 concentration). An elevated serum total T_3 concentration is not entirely specific for thyrotoxicosis, because it can also occur with thyroxine-binding globulin excess, a rare form of familial dysalbuminemic hypertriiodothyroninemia, and anti-T_3 autoantibodies. Serum T_3 assays are useful clinically for fully defining the severity of certain forms of hyperthyroidism, particularly Graves disease; in addition, they are useful, along with the free T_4 concentration, for monitoring the response to treatment of thyrotoxicosis.

DIFFERENTIAL DIAGNOSIS

It is vital for the physician to establish the underlying cause of thyrotoxicosis, because the etiology determines the therapy. In many patients, the history and physical examination alone are sufficient for specific diagnosis. For example, a thyrotoxic woman with a diffuse goiter and exophthalmos almost certainly has Graves disease, whereas a febrile patient with an extremely tender, wood-hard thyroid gland probably has subacute thyroiditis. In other patients, however, the underlying cause may be less certain. For example, a woman with postpartum thyrotoxicosis could have painless (postpartum) thyroiditis, Graves disease, or even factitious thyrotoxicosis—each of which would be treated quite differently. In such patients, further laboratory or radionuclide studies are needed to define the cause and optimal treatment.

The relative degrees of serum T_3 and T_4 elevations can provide a clue to the form of thyrotoxicosis. Predominant T_3 overproduction is common in Graves hyperthyroidism and, to a lesser extent, in toxic nodular goiter (i.e., a serum T_3-to-T_4 [ng/dl:µg/dl] ratio greater than 20). In contrast, T_4-predominant thyrotoxicosis (i.e., a serum T_3-to-T_4 ratio less than 15) suggests thyroiditis (subacute or lymphocytic), iodine-induced thyrotoxicosis, or exogenous T_4 ingestion.

Determining the fractional thyroid uptake of radioactive iodine or pertechnetate and thyroid imaging are required for etiologic diagnosis in some patients. Hyperthyroidism from excessive thyroid hormone synthesis, as in Graves disease, is typically accompanied by increased fractional uptake of the tracer in func-

tioning tissue. In contrast, thyrotoxicosis caused by thyroid inflammation, exogenous thyroid hormone ingestion, and iodine exposure are all associated with a low thyroid uptake. Radionuclide imaging of the thyroid gland often permits differentiation of Graves disease from toxic nodular goiter, because tracer distribution is homogeneous in the former and focal in the latter. Radionuclide imaging can also localize ectopic thyroid tissue that may be hyperfunctioning, such as substernal toxic multinodular goiter and struma ovarii, an ovarian teratoma in which a toxic adenoma can arise.

Anti–TSH receptor immunoglobulin assays have limited clinical uses in differential diagnosis and management.[83] They may be helpful in confirming the diagnosis of Graves disease in clinically and biochemically euthyroid patients who have ophthalmopathy or when differentiation of Graves disease from toxic multinodular goiter is otherwise difficult and important for treatment. In pregnant women with Graves disease, the level of thyroid-stimulating immunoglobulins can predict the likelihood of fetal and neonatal thyrotoxicosis.

Other tests may be helpful in diagnosing certain other causes of thyrotoxicosis. Patients with subacute thyroiditis usually have an elevated erythrocyte sedimentation rate (ESR) and C-reactive protein level, whereas patients with lymphocytic (silent) thyroiditis do not.[88] In patients with thyrotoxicosis caused by thyroid hypersecretion or inflammation, the serum thyroglobulin concentration is high, whereas it is low in patients with factitious thyrotoxicosis. Measurements of the serum glycoprotein hormone α subunit may be useful to confirm the diagnosis of TSH-secreting pituitary adenoma, in which the molar ratio of the α subunit to intact TSH is higher than normal.

Differentiating the two causes of amiodarone-induced thyrotoxicosis is often very difficult, if not impossible. Both the iodine-induced type (type 1) and the inflammatory type (type 2) can be severe; both can be T_4 predominant, and both can be associated with a low thyroid radioiodine uptake. Early reports of higher interleukin-2 levels in the inflammatory form have not been confirmed. Glandular blood flow, as defined by Doppler sonography, is decreased in some patients with the inflammatory form, but this has also proved to be an imperfect distinguishing feature in many affected patients.

MANAGEMENT

Optimal treatment of patients with thyrotoxicosis depends on the underlying cause and severity of their condition and sometimes on the presence of complications that result from hyperthyroidism itself or the patient's other medical disorders. Transient thyrotoxicosis (e.g., exogenous and thyroiditis-related thyrotoxicosis) may require only symptomatic therapy with a beta-adrenergic blocking agent while awaiting spontaneous restoration of euthyroidism. Hyperthyroid Graves disease can be treated with antithyroid medication, radioiodine, or surgery; most of these patients ultimately require an ablative treatment. The hyperthyroidism caused by a toxic adenoma or toxic multinodular goiter will also respond to thionamides (e.g., methimazole and propylthiouracil [PTU]), but it almost never remits spontaneously, so radioiodine or surgery is always required. Fortunately, these three most commonly required therapies are quite comparable with regard to cost-effectiveness, and the vast majority of patients are satisfied with the treatment that they have chosen. Certain special forms of thyrotoxicosis, such as TSH-secreting pituitary adenoma and thyroid hormone intoxication, require other modes of treatment tailored to the responsible cause.

Beta-Adrenergic Blocking Agents

Beta blockers provide prompt relief from some symptoms of thyrotoxicosis, including tremor, palpitations, and anxiety. However, constitutional complaints, such as fatigue and weakness, and hypermetabolic manifestations, such as heat intolerance and weight loss, are unrelieved by beta-adrenergic blockade. These drugs are often valuable for temporary control of symptoms while awaiting a response to more definitive therapies or spontaneous remission of thyrotoxicosis. Beta blockers can be used—optimally, in combination with other drugs—to prepare patients with thyrotoxicosis for surgery. They are also useful for ventricular rate control in patients with thyrotoxic atrial fibrillation (see below). When used judiciously, beta blockers can be a component of treatment for some patients with thyrotoxic heart failure. Some beta blockers, such as propranolol, also partially inhibit extrathyroid T_4-to-T_3 conversion, but these agents do not otherwise address the underlying pathogenesis of thyrotoxicosis.

Antithyroid Drugs

The thionamide antithyroid drugs inhibit iodination and coupling, which are key steps in thyroid hormonogenesis.[89] Consequently, methimazole and PTU are effective treatments for forms of hyperthyroidism caused by excess thyroid hormone production by the gland, such as Graves disease. However, they are ineffective when thyrotoxicosis is caused by unregulated release of hormone from an inflamed gland, such as that which occurs in subacute thyroiditis. Overtreatment with thionamides causes hypothyroidism, so the dose must be titrated or thionamide use must be accompanied by thyroxine replacement. Compared with ablative therapies, antithyroid drugs have the advantage of lowering the long-term incidence of hypothyroidism. However, Graves disease is associated with a 25% long-term incidence of hypothyroidism, which results from the destructive effects of gland inflammation and occurs even with prolonged antithyroid drug treatment.

Thionamide treatment is an appropriate choice for patients with mild Graves disease, in whom the absence of severe manifestations (e.g., a large goiter, a very elevated serum T_4 level, or ophthalmopathy) predicts a higher spontaneous remission rate. Typically, antithyroid drugs are prescribed for 6 to 24 months, after which the dose is tapered to determine whether a remission has occurred. The antithyroid drugs are also useful in four other circumstances: (1) for temporary treatment of patients with Graves disease who are unwilling to accept definitive radioiodine therapy immediately, (2) for preliminary control of hyperthyroidism before definitive radioiodine or surgical treatment, (3) for the management of pregnant women with hyperthyroidism and neonatal Graves disease, and (4) to determine whether nonspecific symptoms are in fact related to mild thyrotoxicosis.

The antithyroid drugs have several limitations. First, they typically take 3 to 8 weeks to restore euthyroidism. Although they inhibit new thyroid hormone synthesis, they do not block the gland's release of existing hormone stores, which can be plentiful in patients with goitrous Graves disease, toxic nodular goiter, and amiodarone-induced thyrotoxicosis. Second, the antithyroid drugs' actions end when the drug is discontinued. As a result, virtually all patients with toxic nodular goiter and the majority of those with Graves disease will experience relapse when the medication is stopped. Third, antithyroid drugs have potential side effects. In addition to hypothyroidism resulting from overtreatment, they cause rash, pruritus, and fever in approxi-

mately 5% of treated patients. Much less commonly, severe adverse reactions can occur, including potentially fatal granulocytosis, vasculitis, or, with PTU, hepatic failure. All thionamide-treated patients must be warned of the symptoms and signs of these problems and be advised that if such manifestations occur, they should report the manifestations and immediately discontinue the medication. The likelihood of another adverse reaction occurring if a patient is switched from one thionamide to another is poorly documented, but such cross-reactivity does occur. Consequently, it is generally advisable in this circumstance to recommend radioiodine or surgery.

For most hyperthyroid patients, methimazole, 10 to 30 mg a day, is more effective and safer than PTU. Methimazole's longer half-life permits a single daily dose, which improves compliance; in contrast, PTU must be given three times daily for sustained effect. Methimazole at dosages of less than 40 mg a day is less likely than PTU to cause agranulocytosis and is not associated with hepatic failure. In certain circumstances, however, PTU, 100 to 200 mg a day in divided doses, may be preferred. In patients with severe and complicated thyrotoxicosis (see below), PTU has the advantage of also blocking extrathyroid T_4-to-T_3 conversion; a benefit in pregnant women is that PTU crosses the placenta less readily than methimazole and therefore has less of an effect on fetal thyroid function.

Radioiodine

Iodine-131 (^{131}I) is a highly effective, safe, and convenient treatment for hyperthyroid patients with hyperthyroid Graves disease, toxic multinodular goiter, and toxic adenoma. This radioisotope of iodine is preferentially concentrated in thyrocytes, where it emits beta particles (electrons) with a short path-length that limits the field of its destructive effects to the thyroid gland. With dosing regimens that are based on estimated gland size and preliminary thyroid radioiodine fractional uptake determinations, or even with empirical doses, a single dose will provide effective treatment for approximately 75% of Graves disease patients and 50% of patients with toxic nodular goiter. Almost all of the remaining patients are cured with a second radioiodine treatment, which is usually best held until the initial dose has proved ineffective after 6 months. For patients with toxic multinodular goiter, there is limited experience with the use of recombinant thyrotropin to increase thyroid uptake of radioiodine—a strategy that might be expected to improve the cure rate and permit a reduction in the administered dose of radioiodine.

Radioiodine has limitations. It takes 1 to 2 months before irradiated thyrocytes die and hyperthyroidism resolves; during this time, patients must often be treated with adjunctive beta-adrenergic blockade, antithyroid drugs, or stable iodide (see below). Approximately 25% of patients develop a transient worsening of thyrotoxicosis 2 to 4 weeks after treatment because of radiation thyroiditis, which can also cause mild and short-lived gland discomfort. The principal side effect of radioiodine is postablative hypothyroidism, which occurs within 3 months in more than half of radioiodine-treated Graves disease patients. Furthermore, in Graves disease patients who remain euthyroid, gland failure continues to occur at a rate of 3% annually, so lifelong follow-up of their thyroid function is essential. In patients with hyperthyroidism resulting from toxic multinodular goiter and toxic adenoma, the incidence of postablative hypothyroidism is lower (approximately 25%) because the suppressed normal extranodular tissue receives much less irradiation. During the more than 65 years that radioiodine has been used to treat hyperthyroidism, the preponderance of evidence has shown no higher long-term incidence of thyroid or other malignancies. In radioiodine-treated women, no higher incidences of subsequent infertility or spontaneous abortion has been found, nor has there been a higher incidence of teratogenesis in their children. In one large follow-up study, radioiodine treatment of children and adolescents was associated with a higher subsequent incidence of benign thyroid nodules. As a result, many experts prefer to treat pediatric patients with antithyroid drugs for several years before resorting to radioiodine. However, radioiodine should not be withheld from children when hyperthyroidism is poorly controlled or side effects occur with thionamide therapy.

Radioiodine is inappropriate in several circumstances. It is absolutely contraindicated in pregnant women. All women of childbearing age should be advised to avoid pregnancy until euthyroidism is restored; this typically requires 3 to 6 months. Radioiodine is not indicated in transient forms of thyrotoxicosis, such as subacute and lymphocytic thyroiditis. Furthermore, it is ineffective in these and other forms of thyrotoxicosis in which the thyroid uptake of radioiodine is decreased, including amiodarone-induced thyrotoxicosis.

Surgery

Thyroidectomy by an experienced surgeon who has a demonstrated low incidence of complications is a highly effective, prompt, and relatively safe alternative. However, transient pain and scarring are universal after thyroidectomy, and postanesthetic symptoms are common. Partial thyroidectomy has an unacceptably high rate of residual hyperthyroidism; thus, gland resection that is extensive enough to ensure success can be expected to result in postsurgical hypothyroidism. Even in the most skilled surgical hands, there is a small risk (approximately 2% to 5%) of hypoparathyroidism or injury to the recurrent laryngeal nerve. Finally, for most patients, surgery is less convenient and entails greater interruption of life commitments than does radioiodine therapy.

Despite its drawbacks, surgery is the best choice in several settings. Pregnant women with severe thionamide side effects have no alternative. In some hyperthyroid patients, other aspects of their condition may make neck surgery necessary (e.g., the patient may have a cytologically suspicious thyroid nodule or hyperparathyroidism). Amiodarone-induced thyrotoxicosis can require surgery when medical treatment is ineffective and the patient's cardiovascular and metabolic status is deteriorating. For some patients, such as those planning prolonged travel in the near future and those who cannot or will not take medications reliably, surgery is attractive because it provides prompt and certain cure. Thyroidectomy is also preferred in countries that require prolonged hospitalization for even low-dose radioiodine therapy.

Other surgical procedures play a role in the treatment of certain rare forms of hyperthyroidism. Transsphenoidal pituitary adenomectomy is typically the first step in treatment of patients with TSH-secreting adenomas, but it is curative in only one third of patients.[90] Oophorectomy is appropriate when hyperthyroidism results from a toxic adenoma arising in teratomatous ovarian tissue in a patient with struma ovarii.

Other Agents

Stable iodide, given as either potassium iodide or Lugol solution in pharmacologic amounts (i.e., 30 mg or more a day), blocks thyroid hormone release from the gland and inhibits organifica-

tion of iodide in patients with Graves disease. When combined with antithyroid drug therapy, iodide can accelerate the decline in circulating thyroid hormone concentrations. However, its effects are only temporary, dissipating after 10 to 14 days, after which hyperthyroidism recurs. Consequently, it is useful in only two settings. First, it can be employed as a short-term measure to prepare patients for thyroidectomy. Second, it can be started several days after radioiodine treatment to accelerate restoration of euthyroidism. In such patients, the irradiated gland is incapable of escaping from iodide's inhibitory effects. Iodide has infrequent side effects of rash, gastritis, and sialadenitis.

Iodinated radiocontrast agents, such as sodium iopanoate, can have two salutary effects in patients with severe thyrotoxicosis. First, they are an abundant source of iodide. Second, they inhibit T_4-to-T_3 conversion.

Lithium carbonate also inhibits hormone release from the thyroid gland. It is used in rare cases to accelerate recovery from severe thyrotoxicosis, as an adjunct to antithyroid medication in type 1 amiodarone-induced thyrotoxicosis, or as a short-term treatment for patients who have experienced severe thionamide side effects. Potassium perchlorate, which blocks iodide uptake by the thyroid, is a rarely used treatment of type 1 amiodarone-induced thyrotoxicosis.

Nonsteroidal anti-inflammatory drugs (NSAIDs) such as aspirin and ibuprofen are first-line agents for controlling thyroid pain and constitutional symptoms in subacute thyroiditis. They do not, however, accelerate recovery of normal thyroid function.

Glucocorticoids (e.g., prednisone, 40 mg/day) can be used to treat thyrotoxicosis under a few specific circumstances. When subacute thyroiditis is resistant to NSAID therapy, steroids are almost invariably effective. However, their side effects and the fact that they may prolong the overall course of subacute thyroiditis make them second-line agents. Glucocorticoid therapy is also useful in the treatment of type 2 amiodarone-induced thyrotoxicosis. Finally, when combined with high-dose antithyroid medication and iodinated radiocontrast agents, glucocorticoids can often be effective in controlling even severe thyrotoxic Graves disease within 1 week after initiation of therapy.

Other agents are employed to treat rare causes of thyrotoxicosis. Cholestyramine can be an adjunct in the treatment of patients with exogenous thyroid hormone intoxication; it interrupts the enterohepatic circulation of thyroid hormones and increases their fecal disposal. Somatostatin analogues are useful in the medical management of patients with inoperable TSH-secreting pituitary tumors or of patients whose tumors were incompletely resected.[91]

COMPLICATIONS

Atrial Fibrillation

Atrial tachyarrhythmias can occur with even mild thyrotoxicosis. Atrial fibrillation is the most common and potentially the most serious of these dysrhythmias; atrial flutter and paroxysmal atrial tachycardia also occur. Thyrotoxic atrial fibrillation occurs more often in older persons and men, as well as in persons with intrinsic cardiac diseases, particularly when these diseases cause left atrial enlargement. Thyrotoxic atrial fibrillation can be complicated by heart failure and thromboembolism. Beta blockers are useful for ventricular rate control; they are less likely to cause hypotension than calcium channel blockers, which remain an alternative. Anticoagulation is generally advisable. Exceptions are in patients in whom thyrotoxic atrial fibrillation has been of short duration and rapidly reverts to sinus rhythm or in those with contraindications to anticoagulation. Cardioversion should generally be deferred until the patient is euthyroid. It should be attempted if spontaneous reversion to sinus rhythm does not occur by 3 months after euthyroidism has been restored[92] [*see 1:IV Atrial Fibrillation*].

Thyrotoxic Heart Failure

When thyrotoxicosis is severe and prolonged or when the patient has intrinsic cardiac disease, heart failure can occur.[93] Contributing factors can include atrial fibrillation; ventricular hypertrophy and dilatation with impaired diastolic function; failure of mitral valve leaflet apposition and resulting regurgitation; and tachycardia-induced cardiomyopathy. Typically, one or more of these factors occurs in the context of increased peripheral tissue demands, vasodilatation, and an expanded blood volume. The left ventricular ejection fraction may be normal at rest but deteriorates with exertion. Therapy includes aggressive treatment of thyrotoxicosis, ventricular rate control, restoration of sinus rhythm when possible, and optimization of blood volume and ventricular filling pressures.

Thyroid Crisis

Thyroid crisis—so-called thyroid storm—refers to the life-threatening constellation of fever; heart failure, often with atrial fibrillation; delirium or psychosis; and fluid and electrolyte depletion resulting from poor oral intake and gastrointestinal losses caused by vomiting or diarrhea.[94] Patients with severe hyperthyroidism, underlying cardiac disease or superimposed infection, and poor access to health care are at increased risk for these complications.[95] Treatment entails medical management of each individual complication and aggressive therapy for thyrotoxicosis, typically with multiple agents appropriate for the underlying cause. Because Graves disease is the most common cause of thyroid storm, treatment typically includes PTU, sodium iopanoate, and glucocorticoids.

Graves Orbitopathy

For some patients with Graves disease, ocular and orbital involvement represents the most disabling aspect of their condition. Exposure keratitis resulting from proptosis, eyelid retraction, and lagophthalmos (inability to close the eye) causes symptoms and can be complicated by infection and ulceration. Treatment includes moistening eyedrops and lubricant ointments, sunglasses, taping the eyelids shut for sleeping, and sometimes blepharoplasty, orbital decompression surgery, and irradiation. Extraocular muscle swelling and fibrosis can cause diplopia, which can require prisms and sometimes corrective surgery. Optic nerve compression can threaten vision; it is treated acutely with high-dose glucocorticoids and definitively by orbital decompression surgery.

PROGNOSIS

Although the short-term morbidity of thyrotoxicosis can be disabling, the long-term outlook is generally bright, given accurate etiologic diagnosis and appropriate therapy targeted to the specific cause. However, many patients with Graves disease ultimately require lifelong treatment of postablative hypothyroidism, and their ophthalmopathy may be an ongoing source of discomfort, cosmetic concern, and, rarely, visual impairment. Mortality and severe long-term disability are rare events that typically result from either cardiovascular complications of thy-

rotoxicosis or side effects of antithyroid medication or thyroid surgery.

Thyroiditis

There are several types of thyroiditis, each of which has distinct causes, clinical manifestations, and treatments [*see Table 3*]. Some types cause thyroid dysfunction. When thyrotoxicosis occurs in these conditions, it is a transient result of the unregulated release of hormone from the gland, whereas hypothyroidism can be transient or permanent. Goiter occurs in some of these disorders (e.g., autoimmune thyroiditis and Reidel thyroiditis), whereas it is not a prominent feature of others. Pain is characteristic of subacute thyroiditis but is an uncommon feature of the other types.

AUTOIMMUNE THYROIDITIS

Autoimmune thyroiditis (Hashimoto disease) is a common condition, particularly in women, who are affected over 10 times more often than men. Its incidence increases with age; during and after middle life, approximately 20% of women have serologic evidence of the condition (i.e., thyroid autoantibodies), and 10% to 15% have an elevated serum TSH level secondary to thyroid hormone insufficiency. Its etiology, genetics, and pathogenesis are discussed elsewhere [*see* Hypothyroidism, *above*].

Clinical manifestations of autoimmune thyroiditis, when present, include hypothyroidism and a goiter. However, most affected patients have either no symptoms or only nonspecific ones, and they do not have significant gland enlargement. When present, the goiter is typically diffuse, modest in size, nontender, and firm with a roughened contour. Patients with the fibrous variant can have more substantial gland enlargement. Pain and tenderness are rarely present.

The presence of thyroid autoantibodies confirms the diagnosis, which can often be established on clinical grounds alone. Immunoassay for antithyroid peroxidase antibodies, which are present in 90% of patients, is the most sensitive single test[96]; antithyroglobulin antibodies are present in only 60% of patients. An elevated serum TSH concentration indicates associated primary hypothyroidism. The differential diagnosis of diffuse goiter includes simple euthyroid goiter, Graves disease, iodine-deficiency goiter, and, rarely, diffusely infiltrating malignancies (i.e., lymphoma, papillary cancer, and anaplastic cancer).

The management of patients with autoimmune thyroiditis includes thyroid hormone replacement for those who are hypothyroid and periodic thyroid function testing for those who are euthyroid but remain at risk for the development of hypothyroidism later in life. Women who have experienced transient hypothyroidism are particularly at risk. Associated goiter rarely requires any treatment, but thyroid hormone replacement sometimes results in partial gland shrinkage. The rare patient with thyroid pain may benefit from NSAIDs. The prognosis for properly treated and monitored patients is excellent. Certain other autoimmune disorders occur more often in affected patients; these can include vitiligo, pernicious anemia, adrenal insufficiency, Sjögren syndrome, and systemic sclerosis. Thyroid lymphoma is more common as well, but it is still a very rare event.

LYMPHOCYTIC THYROIDITIS

Lymphocytic thyroiditis (also known as painless thyroiditis, silent thyroiditis, or postpartum thyroiditis) is believed to be caused by cell-mediated autoimmunity. This belief is based on the fact that the gland is infiltrated with lymphocytes and the condition's incidence is highest postpartum—a time when autoimmune disorders are more common.[97] The condition is relatively common in the postpartum period, affecting approximately 6% of women between 2 and 12 months after delivery or abortion. Women with thyroid autoantibodies, previous episodes of postpartum thyroiditis, or type 1 diabetes mellitus are at markedly increased risk. The condition also occurs during treatment with immunomodulatory agents (see below). Rarely, lymphocytic thyroiditis may present in women or men at other times.

Lymphocytic thyroiditis can cause several patterns of transient thyroid dysfunction: thyrotoxicosis alone, thyrotoxicosis followed by hypothyroidism, hypothyroidism alone, or, rarely, hypothyroidism followed by thyrotoxicosis. The pathogenesis of these derangements is described elsewhere [*see* Thyrotoxicosis, *above*]. In postpartum women, symptoms of thyroid dysfunction can often be overlooked or mistaken for depression. Affected patients have a modest goiter or no goiter.

Serum TSH measurement is the best first-line test to detect thyroid dysfunction in patients with lymphocytic thyroiditis. In thyrotoxic patients, the condition can be differentiated from Graves disease, which can also present postpartum, by the absence of significant goiter or eye involvement; by relatively greater concentration of serum T_4 than of serum T_3 (ratio > 20:1 in μg/dl:ng/dl); and by a low thyroid radioisotope uptake. In hypothyroid patients, postpartum thyroiditis can best be distinguished from autoimmune thyroiditis by whether it remits spontaneously or not. Thyroid autoantibodies are detected in many postpartum thyroiditis patients.

Management can often be expectant, without drug treatment. Symptomatic thyrotoxicosis and hypothyroidism can be treated with temporary beta-adrenergic blockade and thyroxine, respectively. One quarter of affected patients go on to develop typical autoimmune thyroiditis and permanent hypothyroidism.

SUBACUTE THYROIDITIS (DE QUERVAIN OR GRANULOMATOUS THYROIDITIS)

Subacute thyroiditis is believed to be the result of a viral infection, because of its association with prodromal symptoms, the presence of circulating viral antibody titers, and electron microscopic evidence of viral particles. Classically, episodes of subacute thyroiditis have three clinical components. First are the systemic manifestations: symptoms suggesting a viral upper respiratory tract infection, followed by malaise, fever, and chills. The second is a painful goiter, which is characteristically moderate in size, wood-hard, and extremely tender. The third is transient thyrotoxicosis, which can be quite severe. The thyrotoxicosis lasts for 2 to 8 weeks and is typically followed by transient hypothyroidism. Although the clinical features of subacute thyroiditis are usually sufficient to suggest the diagnosis, constitutional symptoms (especially high fever) can mimic other infections, and thyroid pain radiating to the ears can be confused with otitis. Thyrotoxicosis is confirmed by a low serum TSH and a high serum T_4 concentration. Marked elevation of the ESR is detectable during the acute phase of the illness. Other causes of painful thyroid enlargement include hemorrhage into a thyroid nodule, which is usually asymmetrical; acute thyroiditis (see below); thyroid lymphoma or anaplastic thyroid cancer; and, very rarely, autoimmune thyroiditis or Graves disease.

Management of subacute thyroiditis entails prescription of an NSAID in high dosage. In approximately 20% of patients, this provides inadequate relief of pain and constitutional symptoms;

these patients can be treated with glucocorticoid therapy (e.g., prednisone, 40 mg/day with a slow taper over 3 to 8 weeks). Transient thyrotoxic symptoms are treated with beta blockers; transient hypothyroid symptoms are treated with thyroxine. In more than 80% of cases, normal thyroid function returns and the condition does not recur.

ACUTE THYROIDITIS (SUPPURATIVE THYROIDITIS)

Acute or suppurative thyroiditis is a rare condition caused by either untreated bacterial infections of the upper respiratory tract or cervical soft tissues or by opportunistic agents in immunocompromised hosts. Hematogenous spread to the thyroid of fungal, mycobacterial, and parasitic infections have all been reported. Piriform sinus fistula, multinodular goiter, or autoimmune thyroiditis may predispose patients to acute thyroiditis.

Patients with suppurative infections typically are extremely ill, with high fever; a painful, tender, swollen thyroid gland; and erythema and warmth of overlying soft tissues. Glands infected with opportunistic pathogens (e.g., *Pneumocystis jiroveci*) may have more subtle signs of gland infection. Treatment requires aggressive, often parenteral, antibiotic therapy and sometimes surgical drainage.

DRUG-INDUCED THYROIDITIS

Several drugs have been associated with painless thyroiditis and thyroid dysfunction, including amiodarone (see above). Treatment of hepatitis C with interferon alfa and interleukin-2 causes thyroid dysfunction in as many as 15% of patients; such dysfunction includes transient thyrotoxicosis with or without subsequent hypothyroidism, as well as persistent hypothyroidism and persistent hyperthyroidism (i.e., Graves disease). Lithium carbonate can exacerbate thyroid dysfunction and cause hypothyroidism in patients with underlying autoimmune thyroiditis. Rarely, pharmacologic doses of iodide can cause transient thyroiditis.

REIDEL THYROIDITIS

Reidel thyroiditis is an extremely rare form of invasive fibrous thyroiditis that can cause substantial goiter with compression and infiltration of adjacent structures. Some patients with this idiopathic disorder also develop hypothyroidism. It may be encountered in patients with retroperitoneal and mediastinal fibrosis. Treatment options are limited. Surgery is challenging and is limited to palliation of obstructive complications. There are reports of effective drug treatment with glucocorticoids and tamoxifen.[98]

Goiter, Thyroid Nodules, and Thyroid Cancer

GOITER

Epidemiology

The prevalence of goiter in populations varies inversely with dietary iodine intake. There are estimated to be 100 million persons with dietary iodine deficiency and one billion with borderline iodine sufficiency. Consequently, in some regions, goiter is almost universal, particularly in women, whereas in regions with an adequate iodine intake, goiter affects less than 5% of the population.

Etiology and Pathogenesis

Worldwide, dietary iodine deficiency is the most common cause of thyroid gland enlargement. In addition, populations have been described in which exposure to goitrogens in drinking water or dietary substances has caused endemic goiter.[99] Although modest thyroid gland enlargement occurs during pregnancy, the so-called goiter of pregnancy is largely restricted to women with associated dietary iodine deficiency. Tobacco smoking has been associated with goiter development, especially in populations without ample dietary iodine. Mutations in genes encoding key proteins involved in thyroid hormone synthesis (e.g., thyroglobulin and thyroid peroxidase), can result in a compensatory goiter. Activation of the TSH receptor by TSH predictably causes diffuse goiter, as occurs in a variety of conditions, including primary hypothyroidism from autoimmune thyroiditis or drugs, TSH-secreting pituitary adenoma, and resistance to thyroid hormone. The TSH receptor can be aberrantly stimulated when thyroid-stimulating immunoglobulins or high levels of hCG bind to it or when a mutation in the TSH receptor gene itself leads to constitutive activation. For the majority of patients with hyperplastic thyroid glands arising despite sufficient dietary iodine—a condition that is sometimes apparently inherited and sometimes sporadic—the precise molecular cause remains unknown. Activation of the biochemical pathways signaling thyrocyte growth or abnormal local levels or activity of intrathyroid growth factors seems likely to be involved. Goiter can also be the result of gland infiltration with inflammatory cells (e.g., leukocytes and multinucleated giant cells in subacute thyroiditis) or tumor cells (e.g., anaplastic or diffusely infiltrating papillary thyroid cancers).

In summary, goiter can be associated with hypothyroidism, euthyroidism (nontoxic goiter), or hyperthyroidism (toxic goiter). It can be a manifestation of benign hyperplastic or neoplastic conditions, malignancy, and inflammation. In many of these disorders, thyroid gland enlargement is initially diffuse, but with time, the thyroid becomes multinodular, and it may become asymmetrical.

Diagnosis and Differential Diagnosis

For the clinician evaluating a patient with a goiter, three pragmatic questions are typically more important than the specific etiologic diagnosis. First, is the enlarged thyroid gland producing pain or other local symptoms as a result of obstruction or invasion of adjacent structures, or is it so large as to be unsightly? Second, is the goiter a manifestation of a disorder causing hypothyroidism or hyperthyroidism that requires treatment? Third, is the gland enlarged because of malignancy?

Clinical manifestations Symptoms related to a goiter reflect gland impingement on adjacent structures. A sensation of cervical fullness, tightness, or pain can occur when the enlarging gland stretches its capsule, which has sensory innervation, and compresses adjacent tissues and structures. Thyroid pain can radiate to the jaw or ears. Tracheal compression can cause cough and difficulty clearing mucus; tracheal invasion from thyroid malignancy can cause hemoptysis. Esophageal compression can produce dysphagia and, rarely, odynophagia. Compression of the recurrent laryngeal nerve results in hoarseness, a weak voice, and dysphagia for fluids; bilateral nerve dysfunction can also cause dyspnea from airway obstruction.

Physical examination Inspection during deglutition often provides the first clue to the presence of a goiter. In addition, it helps distinguish true enlargement of the thyroid, which moves cephalad with swallowing, whereas subcutaneous fat does not.

Any tracheal deviation, cervical vein engorgement, or visible adenopathy should also be noted. On palpation, the dimensions of the gland and its symmetry, contour, consistency, mobility, and tenderness should all be noted.

Other physical findings can provide important information. A bruit suggests either a hypervascular Graves disease gland or compression of cervical blood vessels. Venous engorgement and facial plethora that develops when the patient touches the hands together above the head (Pemberton sign) implies near obstruction of the thoracic outlet by a goiter. Signs of hypothyroidism or hyperthyroidism should, of course, be sought.

Laboratory tests and imaging studies In all patients with goiter, the serum TSH and free T_4 concentrations should be measured to assess gland function. Serologic testing for thyroid autoantibodies, especially anti–thyroid peroxidase antibody, can help establish the diagnosis of autoimmune thyroiditis, the most common cause of diffuse goiter in populations with sufficient dietary iodine.

Ultrasonography can be useful in confirming that a neck mass is, in fact, a goiter. It can define the gland size; determine whether there is diffuse heterogeneity typical of autoimmune thyroiditis or discrete nodules; and identify potentially related cervical adenopathy. A chest and lower cervical x-ray can show tracheal deviation and suggest mediastinal extension of a goiter, but computed tomography or magnetic resonance imaging provides a fuller depiction of the gland's substernal extent and its relationship to the trachea and intrathoracic structures. The decision whether or not to use iodinated radiocontrast agents with CT imaging should be thoughtfully addressed, because these agents can precipitate hyperthyroidism in patients with multinodular goiter and can interfere with subsequent postoperative radioiodine therapy in patients who prove to have thyroid cancer. Radionuclide imaging is seldom required in the initial assessment of patients with goiter, but it can confirm that a superior mediastinal mass concentrating radioiodine is thyroidal in origin. Radioiodine fractional uptakes and imaging can be useful in the differential diagnosis of goitrous thyrotoxicosis (see above). Assessment of ventilatory flow-volume loops can be useful in determining whether dyspnea in a patient with goitrous tracheal compression is caused by the thyroid condition. Although cytologic evaluation plays a central role in the differential diagnosis of thyroid nodules, it is required in only a small minority of patients with diffuse goiters that cannot be readily characterized with clinical, laboratory, and imaging findings.

Management

Management of goiter addresses three key clinical issues: size, function, and potential malignancy. Treatment for thyroid dysfunction or thyroid cancer is the same as in patients without goiter. Large nontoxic, multinodular goiters causing obstructive symptoms or cosmetic concerns can be surgically excised or, if obstruction is not severe, treated with radioiodine.[100] Surgery provides more prompt relief and excludes cancer with certainty, but it is associated with greater short-term morbidity than radioiodine therapy. Even goiters with substantial substernal extension can often be removed through a cervical incision. Recombinant thyrotropin can successfully augment ^{131}I concentration by nodular goiters, which typically have only a normal fractional radioiodine uptake.[101] TSH-suppressive thyroxine therapy has limited value in the treatment of most patients with goiters of significant dimensions. Published experience shows that no more than half of patients have a response, which is often only partial.[102] Furthermore, long-term TSH suppression is associated with risks of bone loss and atrial fibrillation.

Complications and Prognosis

The type and probability of complications in patients with goiter depend on the underlying cause. Most patients with gland enlargement of benign cause never suffer local compressive symptoms and require treatment only if they have associated thyroid dysfunction (i.e., hypothyroidism in patients with autoimmune thyroiditis or hyperthyroidism in patients with multinodular goiter). However, large multinodular goiters can cause dyspnea from tracheal narrowing or dysphagia from esophageal compression. Recurrent laryngeal nerve impingement and dysfunction is very unusual and should raise concern of malignancy. Rarely, thyroid substernal goiter extension can cause superior vena cava obstruction. Goiter from papillary or anaplastic thyroid cancer or lymphoma can cause all of the complications associated with goiters from benign causes, as well as pain from invasion of adjacent structures, hemoptysis from tracheal invasion, and vocal cord paresis from recurrent laryngeal nerve involvement.

THYROID NODULES

Epidemiology and Etiology

Thyroid nodules are palpable in 6% of women and 2% of men.[2] Nonpalpable thyroid nodules can be detected by sonography in one third of women. The prevalence of nodules increases with age. The genetic and environmental factors associated with thyroid nodule development are essentially the same as those for goiter. Nodules may represent solid tissue composed of thyroid cells or colloid or represent cysts from accumulated serous fluid or blood, often from hemorrhage within a solid nodule in the gland. The majority of thyroid nodules are benign; in adults, 5% to 10% of thyroid nodules are cancerous.

Diagnosis

Three clinical issues must be addressed in patients with thyroid nodules: the nodule's size and the resulting potential for local complications, the possibility of associated thyroid dysfunction, and malignancy.[103] The same principles and approach apply to palpable thyroid nodules as to incidentally detected nodules (so-called thyroid incidentalomas) that are greater than 1.0 to 1.5 cm in diameter.[104] After proper assessment, the majority of patients will be found to have none of these three problems, and they can be monitored conservatively.

Clinical Manifestations

Thyroid nodules are usually detected incidentally by asymptomatic patients themselves or by their physicians. Rapidly enlarging or invasive nodules may cause pain in the anterior neck, jaw, or ear. Tracheal or esophageal compression can cause cough or dysphagia, respectively. Invasion of the trachea or recurrent laryngeal nerve by tumor can produce the worrisome symptoms of hemoptysis or hoarseness, respectively. Symptoms of thyrotoxicosis suggest the possibility of toxic adenoma, whereas complaints consistent with hypothyroidism may reflect autoimmune thyroiditis and an asymmetrical goiter that is mimicking a true nodule.

The presence of pulmonary, skeletal, or neurologic symptoms suggesting metastatic disease increases concern about a primary

thyroid cancer. An increased risk of thyroid cancer is also suggested by a history of childhood or adolescent irradiation; irradiation was employed until the early 1950s for thymic enlargement, tonsillitis, adenoiditis, cutaneous hemangiomas, and acne. A family history of thyroid cancer is also an indication for thorough assessment, especially if the familial disease is medullary or papillary thyroid cancer. A history of hyperparathyroidism or pheochromocytoma raises the possibility of the multiple endocrine neoplasia type II (MEN II) syndrome, which includes these disorders and medullary thyroid cancer.[105] Hypercalcitoninemia in patients with metastatic medullary thyroid cancer can cause flushing, pruritus, and diarrhea.

Physical Examination

Nodules that are fixed or associated with ipsilateral cervical adenopathy are worrisome for thyroid cancer. Nodule size and consistency are not reliable features for distinguishing benign from malignant lesions. Although multiple thyroid nodules are typical of benign multinodular goiter, a nodule that is larger, that is growing more rapidly, or that is more symptomatic than others in the gland requires the same assessment as a solitary thyroid nodule. Signs of hyperthyroidism or hypothyroidism suggest toxic adenoma or autoimmune thyroiditis, respectively. Patients with the MEN IIB (or MEN III) syndrome, which includes medullary thyroid cancer and pheochromocytoma, can have a Marfanoid body habitus and submucosal neuromas that are visible as lumps beneath the buccal mucosa and conjunctivae.

Laboratory Tests

The serum TSH concentration should be measured: a low serum TSH suggests a possible toxic adenoma; a high TSH, autoimmune thyroiditis. Antithyroid antibody screening can corroborate the diagnosis of autoimmune thyroiditis in patients who actually have diffuse but asymmetrical thyroid gland enlargement. Serum calcitonin should be measured in patients with clinical features that suggest hypercalcitoninemia or the MEN II syndrome. The serum thyroglobulin assay is not helpful in distinguishing benign from malignant thyroid nodules.

For most thyroid nodules, the definitive diagnostic procedure is fine-needle aspiration to provide cytologic material for examination by an experienced pathologist. Sonographic guidance of aspiration can be useful when nodules are poorly localized by palpation or for the assessment of lesions that have a cystic component; in addition, it can sometimes be useful in identifying additional nonpalpable nodules requiring assessment.[106] Radionuclide imaging currently has only a secondary role in thyroid nodule assessment. In thyroid nodule patients whose serum TSH concentration is low, a iodine-123 or technetium-99m pertechnetate scan that indicates that the nodule is "hot" (i.e., that shows a concentration of tracer in the nodule, with suppression of uptake in the remainder of the gland) provides assurance that the nodule is benign. CT and MRI have no role in the evaluation of the typical patient with a thyroid nodule unless there is substernal extension of a nodular goiter.

Differential Diagnosis

Clinical features of thyroid nodules are sometimes helpful but seldom provide definitive diagnosis. Sonography can confirm that an ambiguous neck mass is thyroidal. Radionuclide imaging should be limited to patients with a suppressed serum TSH concentration. The sensitivity and specificity of cytologic assessment of thyroid nodules are 97% and 95%, respectively. However, approximately 20% of nodules are cytologically indeterminate, among which approximately 15% are malignant.[107] Some thyroid nodules can ultimately be definitively diagnosed only by surgical excision and histopathologic examination.

Management, Complications, and Prognosis

Most cytologically benign thyroid nodules can be managed with observation only, unless the lesion is so large as to cause discomfort, other local compressive complications, or cosmetic concern. Because cytology is not 100% sensitive for cancer exclusion, patients should be followed up for 12 to 24 months. For palpable lesions, follow-up should consist of physical examination; for nonpalpable lesions, sonography is usually advisable. Toxic adenomas diagnosed on the basis of a low serum TSH level and radionuclide confirmation of hot-nodule status can usually be assumed to be benign; treatment of associated thyrotoxicosis is, of course, indicated. Patients with cytologically malignant nodules should undergo thyroidectomy unless their general medical condition contraindicates it. Most patients with cytologically indeterminate nodules require surgery for definitive diagnosis, although if the suspicion of cancer is low, thyroid lobectomy may be worth considering. A subset of patients with cytologically indeterminate thyroid nodules (e.g., older women with multinodular goiter and no clinical features suggesting malignancy) can be followed with sonographic monitoring of the nodules' dimensions.

THYROID CANCER

Epidemiology

Thyroid cancers represent approximately 2% of clinically detected malignancies.[108] The most common tumor types, arising from follicular epithelium (i.e., papillary, follicular, and Hürthle cell cancers), occur three times more often in women and increase in incidence with age. Medullary thyroid cancer arising from parafollicular C cells represents less than 10% of all thyroid cancers but has special importance because of its common familial occurrence. In the United States, the reported incidence of thyroid cancer is rising, perhaps because these tumors are being detected more easily with contemporary diagnostic tools.[109]

Etiology and Pathogenesis

The cause of most thyroid cancers is unknown. Genetic predisposition to papillary thyroid cancer is seen in the familial syndromes of familial adenomatous polyposis (*APC* gene mutation) and Cowden syndrome (*PTEN* gene mutations). It can also occur as familial isolated papillary thyroid cancer. Overall, however, familial cases represent less than 10% of all thyroid cancers.[110] Thyroid irradiation from external sources and accidental radioiodine exposure predisposes to the development of malignant and benign thyroid tumors, as has been observed after radiotherapy (e.g., for recurrent tonsillitis and lymphoma) and after exposure to radioactive iodine fallout from a nuclear-weapon detonation or nuclear-reactor accident. Radioiodine exposure has been shown to produce a characteristic chromosomal rearrangement that creates the *RET/PTC* oncogene.

Progression and clinical aggressiveness of thyroid malignancies have been associated with a sequence of molecular events, including *BRAF* gene mutation and loss of the *p53* tumor suppressor gene.[111] Controversy surrounds evidence relating the development of thyroid cancer to preexisting benign thyroid conditions; to parity and estrogen therapy in women; to previous therapeutic radioiodine exposure; and to dietary factors, includ-

ing iodine intake. Dietary iodine does clearly influence the distribution of thyroid cancer types, with more papillary cancers in populations with generous dietary iodine content.

Diagnosis

Clinical manifestations and physical examination Most thyroid cancers present as a thyroid nodule in an otherwise asymptomatic and euthyroid patient. Enlargement of the mass over weeks or months is more suspicious for cancer than long-standing stable size or very rapid appearance, which can represent hemorrhage into a preexisting benign nodule. Less commonly, patients develop complaints related to local invasion (e.g., pain, hoarseness, or hemoptysis) or distant metastatic disease (e.g., dyspnea, bone pain, or neurologic symptoms). On physical examination, nodule fixation or ipsilateral cervical adenopathy suggests thyroid cancer.

Laboratory tests Cytologic diagnosis of thyroid cancer can often be established from material obtained by fine-needle aspiration of suspicious thyroid nodules [*see* Thyroid Nodules, *above*]. Patients with cytologically indeterminate lesions usually require surgery for definitive diagnosis. Novel molecular markers, such as *BRAF* in papillary thyroid cancers, may in the future permit more accurate preoperative diagnosis and exclusion of cancer in these lesions. In thyroid cancer patients who present with metastatic disease in cervical nodes, lungs, bone, and other sites, biopsy of the identified lesion often establishes the diagnosis, which can be confirmed by thyroglobulin immunostaining. Subsequent careful examination and sonographic imaging of the thyroid gland typically reveal a nodule that can then itself be subject to biopsy.

Management

The therapeutic modalities commonly employed to treat patients with epithelial thyroid cancers are surgical thyroidectomy, radioiodine, and TSH-suppressive thyroid hormone therapy. Physicians and surgeons face the challenge of determining how aggressively these treatments should be applied to these patients, who exhibit a wide spectrum of disease behavior.[112]

When the diagnosis has been made preoperatively, total or near-total thyroidectomy is the procedure of choice. The rationale for bilateral thyroid excision is that papillary cancer, the most common thyroid malignancy, is often multifocal, involving the contralateral lobe in at least 20% of cases. Bilateral surgery has been shown to be associated with a lower papillary cancer recurrence rate.[113] In addition, removal of all, or nearly all, normal thyroid tissue positions patients for more accurate long-term monitoring with serum thyroglobulin measurement and radioiodine imaging. When thyroid cancer is unexpectedly diagnosed after lobectomy, these potential benefits must be balanced against the risks and inconvenience of completion thyroidectomy. For patients with microscopic papillary and minimally invasive follicular cancers, unilateral surgery may be deemed to have been adequate.[114] Unless regional node metastases are recognized before thyroidectomy, selective central neck compartment node excision is generally advisable, although modified radical neck dissection is justifiable when extensive nodal involvement is identified before or at the time of initial surgery.

Postoperatively, ^{131}I is often recommended for patients with epithelial thyroid cancers, with the rationale of eradicating residual disease and ablating remnant thyroid tissue that will otherwise limit the accuracy of long-term monitoring with serum thyroglobulin and radioiodine scans. The value of adjunctive radioiodine treatment to reduce risk of tumor recurrence has been shown in retrospective and observational trials for patients with more advanced stages of disease.[115] The principal factors related to an increased risk of recurrence are older patient age, larger tumor size, extrathyroidal invasion, incomplete tumor resection, and extensive and nodal metastases. In addition, certain histologic subtypes of thyroid cancer are more likely to recur, including the tall cell, columnar cell, and insular variants of papillary thyroid cancer, as well as follicular cancers with vascular invasion. Traditionally, patients have been withdrawn from thyroid hormone therapy postoperatively to effect a rise in endogenous TSH and to facilitate radioiodine uptake by residual thyroid tissue. This is effective but predictably causes clinical hypothyroidism. It has now been shown that radioiodine therapy can also be effective in euthyroid patients who are given recombinant thyrotropin.[116]

Patients with epithelial thyroid carcinoma require long-term thyroxine therapy both to replace thyroid hormone and to suppress TSH to reduce the risk of tumor recurrence. Typically, the thyroxine dose is increased until the lowest dosage capable of suppressing the serum TSH concentration to less than 0.1 μU/L is identified. In patients with symptoms of thyrotoxicosis or patients at low risk for tumor recurrence, the target TSH concentration may adjusted to the 0.1 to 0.5 μU/L range.

Patients treated for epithelial thyroid cancers require long-term follow-up to detect recurrent disease, which can present years after initial therapy. Serum thyroglobulin measurement has become the first-line choice for tumor detection.[117] This thyroid-specific protein should be undetectable in the blood of patients who have no residual thyroid cancer and who have undergone complete ablation of normal thyroid tissue. When circulating thyroglobulin is detected—whether during thyroid hormone therapy or after TSH stimulation by either thyroid hormone withdrawal or recombinant TSH administration—imaging techniques should be employed to localize the residual disease. Cervical sonography and fine-needle aspiration of suspicious adenopathy is the most productive initial step, followed by CT of the chest and, in patients with a serum thyroglobulin concentration greater than 10 ng/ml, 18-fluorodeoxyglucose positron emission tomography. Radioiodine scanning with ^{123}I or ^{131}I is another monitoring technique that is particularly helpful in identifying residual normal thyroid tissue as the source of circulating thyroglobulin and in localizing iodine-avid residual cancer tissue that may be amenable to radioiodine therapy. Radioiodine imaging also requires TSH stimulation by thyroid hormone withdrawal or recombinant TSH administration.[118]

Additional treatment modalities may be required for patients with advanced epithelial thyroid cancers.[119] Radioiodine can be employed for iodine-avid metastatic disease that is nonresectable, such as pulmonary metastases in younger patients with papillary thyroid cancer and metastatic follicular thyroid cancer in older patients. Repeat surgery may be indicated to excise recurrent cervical disease and, occasionally, other distant metastatic lesions that are solitary or that threaten to cause complications. External-beam radiotherapy can be used to treat nonresectable cervical disease, painful bone lesions, or pulmonary metastases causing airway obstruction or hemoptysis. Chemotherapy for these tumors has only a partial response rate; moreover, the response rate is relatively low, and the risk of side effects is significant. Nonetheless, chemotherapy may be offered when other alternatives have been exhausted.

Less common and more aggressive thyroid malignancies are managed with some of the same therapeutic modalities. Medullary thyroid cancer is treated with initial thyroidectomy.[120] Thyroid hormone therapy for replacement, but not TSH suppression, is then prescribed. Patients are followed with serial calcitonin and carcinoembryonic antigen measurements. Repeat surgery, external-beam radiotherapy, and chemotherapy (which is relatively ineffective) are sometimes employed. Thyroid lymphoma, which is primary to the thyroid gland in half of cases, is diagnosable with tissue biopsy and is treated, sometimes rather effectively, with combined chemotherapy and radiation therapy. Anaplastic thyroid cancer is typically nonresectable and is also treated with combined external-beam radiotherapy and chemotherapy.[121] Only in exceptional cases, however, do these interventions significantly alter the grim prognosis.

Complications

Complications can occur in thyroid cancer patients as a result of the malignancy or its treatment. Local cervical invasion of the recurrent laryngeal nerve can cause temporary or permanent hoarseness, dysphagia, and dyspnea. Progression of tumor in the neck can lead to strangulation or esophageal obstruction and malnutrition. Pulmonary failure can occur with pulmonary metastases, fractures with bony involvement, paraparesis with paraspinal lesions, and other neurologic consequences with brain dissemination. Functioning metastases can cause thyrotoxicosis in patients with follicular cancer and, rarely, in patients with papillary carcinoma.

Thyroidectomy may be complicated by recurrent laryngeal nerve injury or hypoparathyroidism. Radioiodine treatment can cause gastritis with short-term symptoms, and it can cause sialadenitis, whose symptoms are dry mouth, loss of taste, and dental caries. High cumulative doses of therapeutic ^{131}I have been associated with an increase in the risk of leukemia. External-beam radiotherapy and chemotherapy have their usual potential for adverse reactions.

Prognosis

The indolent growth of most thyroid tumors and the efficacy of available therapies result in a low mortality, with survival rates of 98% for papillary, 92% for follicular, and 80% for medullary cancers.[122] Clinical recurrence is common, however; for example, almost one third of papillary thyroid cancers recur. Extracervical medullary cancer is incurable but generally follows a slowly progressive course. Poorly differentiated epithelial cell thyroid cancers and anaplastic thyroid cancers unfortunately can be among the most aggressive and treatment-resistant malignancies known.

The author has received grants for clinical research or educational activities from or served as advisor or consultant to Abbott Laboratories and Genzyme Corporation during the past year.

Recombinant thyroid-stimulating hormone has not been approved by the FDA for treatment of nodular goiter and postoperative ablation of thyroid gland remnant tissue.

References

1. Hollowell JG, Staehling NW, Flanders WD, et al: Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab 87:489, 2002

2. Vander JB, Gaston EA, Dawber TR: The significance of nontoxic thyroid nodules: final report of a 15 year study of the incidence of thyroid malignancy. Ann Intern Med 69:537, 1968

3. Brander AE, Viikinkoski VP, Nickels JI, et al: Importance of thyroid abnormalities detected at US screening: a 5-year follow-up. Radiology 215:801, 2000

4. Trowbridge FL, Hand KE, Nichaman MZ: Findings relating to goiter and iodine in the Ten-State Nutrition Survey. Am J Clin Nutr 28:712, 1975

5. Soldin OP, Soldin SJ, Pezzullo JC: Urinary iodine percentile ranges in the United States. Clin Chim Acta 328:185, 2003

6. NCI Fact Book. National Cancer Institute, Washington DC, 2004, p C-2 http://www3.cancer.gov/admin/fmb/03Factbk.pdf

7. Roberts C, Ladenson PW: Hypothyroidism. Lancet 363:793, 2004

8. Chopra IJ, Solomon DH, Chopra U, et al: Abnormalities in thyroid function in relatives of patients with Graves disease and Hashimoto's thyroiditis: lack of correlation with inheritance of HLA-B8. J Clin Endocrinol Metab 45:45, 1977

9. Muller AF, Drexhage HA, Berghout A: Postpartum thyroiditis and autoimmune thyroiditis in women of childbearing age: recent insights and consequences for antenatal and postnatal care. Endocr Rev 22:605, 2001

10. Falorni A, Laureti S, Santeusanio F: Autoantibodies in autoimmune polyendocrine syndrome type II. Endocrinol Metab Clin North Am 31:369, 2002

11. Dayan CM, Daniels GH: Chronic autoimmune thyroiditis. N Engl J Med 335:99, 1996

12. Allen EM, Hsueh W-C, Sabra M, et al: A genome-wide scan for autoimmune thyroid disease in the Old Order Amish: replication of loci on chromosomes 5q11.2-q14.3 and 18p11.31 and modest evidence for linkage in four other regions. J Clin Endocrinol Metab 88:1292, 2003

13. Koh LK, Greenspan FS, Yeo PP: Interferon-alpha induced thyroid dysfunction: three clinical presentations and a review of the literature. Thyroid 7:891, 1997

14. Sklar C, Whitton J, Mertens A, et al: Abnormalities of the thyroid in survivors of Hodgkin's disease: data from the Childhood Cancer Survivor Study. J Clin Endocrinol Metab 85:3227, 2000

15. Martino E, Safran M, Aghini-Lombardi F, et al: Environmental iodine intake and thyroid dysfunction during chronic amiodarone therapy. Ann Intern Med 101:28, 1984

16. Lazarus JH: The effects of lithium therapy on thyroid and thyrotropin-releasing hormone. Thyroid 8:909, 1998

17. Li FX, Squartsoff L, Lamm SH: Prevalence of thyroid diseases in Nevada counties with respect to perchlorate in drinking water. J Occup Environ Med 43:630, 2001

18. Rose SR: Cranial irradiation and central hypothyroidism. Trends Endocrinol Metab 12:97, 2001

19. Segal-Lieberman G, Karasik A, Shimon I: Hypopituitarism following closed head injury. Pituitary 3:181, 2000

20. Sherman SI: Etiology, diagnosis, and treatment recommendations for central hypothyroidism associated with bexarotene therapy for cutaneous T-cell lymphoma. Clin Lymphoma 3:249, 2003

21. Yen PM: Physiological and molecular basis of thyroid hormone action. Physiol Rev 81:1097, 2001

22. Davis PJ, Davis FB: Nongenomic actions of thyroid hormone. Thyroid 6:497, 1996

23. Ladenson PW, Braverman LE, Mazzaferri EL, et al: Comparison of administration of recombinant human thyrotropin with withdrawal of thyroid hormone for radioactive iodine scanning in patients with thyroid carcinoma. N Engl J Med 337:888, 1997

24. Canaris GJ, Steiner JF, Ridgway EC: Do traditional symptoms of hypothyroidism correlate with biochemical disease? J Gen Intern Med 12:544, 1997

25. Nickel SN, Frame B: Neurological manifestations of myxedema. Neurology 8:511, 1958

26. Clarnette RM, Patterson CJ: Hypothyroidism: does treatment cure dementia? J Geriatr Psychiatry Neurol 7:23, 1994

27. Morris MS, Bostom AG, Jacques PF, et al: Hyperhomocysteinemia and hypercholesterolemia associated with hypothyroidism in the Third US National Health and Nutrition Examination Survey. Atherosclerosis 155:195, 2001

28. Demers LM, Spencer CA: Laboratory support for the diagnosis and monitoring of thyroid disease. The National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines. National Academy of Clinical Biochemistry, Washington DC, 2002 http://www.nacb.org/lmpg/thyroid_lmpg_pub.stm

29. Faglia G, Bitensky L, Pinchera A, et al: Thyrotropin secretion in patients with central hypothyroidism: evidence for reduced biological activity of immunoreactive thyrotropin. J Clin Endocrinol Metab 48:989, 1979

30. Beck-Peccoz P, Amr S, Menezes-Ferreira MM, et al: Decreased receptor binding of biologically inactive thyrotropin in central hypothyroidism: effect of treatment with thyrotropin-releasing hormone. N Engl J Med 312:1085, 1985

31. Beck-Peccoz P, Brucker-Davis F, Persani L, et al: Thyrotropin-secreting pituitary tumors. Endocr Rev 17:610, 1996

32. Beck-Peccoz P, Chatterjee VK: The variable clinical phenotype in thyroid hormone resistance syndrome. Thyroid 4:225, 1994

33. Hamblin PS, Dyer SA, Mohr VS, et al: Relationship between thyrotropin and thyroxine changes during recovery from severe hypothyroxinemia of critical illness. J Clin Endocrinol Metab 62:717, 1986

34. Smith PJ, Surks MI: Multiple effects of 5,5′-diphenylhydantoin on the thyroid hormone system. Endocr Rev 5:514, 1984

35. Meyer CM, Ladenson PW, Scharfstein JA, et al: Evaluation of common problems in primary care: effects of physician, practice, and financial characteristics. Am J Manag Care 6:457, 2000

36. Vanderpump MP, Tunbridge WM, French JM, et al: The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. Clin Endocrinol (Oxf) 43:55, 1995

37. Glinoer D, Riahi M, Grun JP, et al: Risk of subclinical hypothyroidism in pregnant

women with asymptomatic autoimmune thyroid disorders. J Clin Endocrinol Metab 79:197, 1994

38. Premawardhana LD, Parkes AB, John R, et al: Thyroid peroxidase antibodies in early pregnancy: utility for prediction of postpartum thyroid dysfunction and implications for screening. Thyroid 14:610, 2004

39. Parle JV, Franklyn JA, Cross KW, et al: Thyroxine prescription in the community: serum thyroid stimulating hormone level assays as an indicator of undertreatment or overtreatment. Br J Gen Pract 43:107, 1993

40. Fish LH, Schwartz HL, Cavanaugh J, et al: Replacement dose, metabolism, and bioavailability of levothyroxine in the treatment of hypothyroidism: role of triiodothyronine in pituitary feedback in humans. N Engl J Med 316:764, 1987

41. Alexander EK, Marqusee E, Lawrence J, et al: Timing and magnitude of increases in levothyroxine requirements during pregnancy in women with hypothyroidism. N Engl J Med 351:241, 2004

42. Arafah BM: Increased need for thyroxine in women with hypothyroidism during estrogen therapy. N Engl J Med 344:1743, 2001

43. Quan ML, Pasieka JL, Rorstad O: Bone mineral density in well-differentiated thyroid cancer patients treated with suppressive thyroxine: a systematic overview of the literature. J Surg Oncol 79:62, 2002

44. Sawin CT, Geller A, Wolf PA, et al: Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. N Engl J Med 331:1249, 1994

45. Keating FR Jr, Parkin TW, Selby JB, et al: Treatment of heart disease associated with myxedema. Prog Cardiovasc Dis 3:364, 1961

46. Sherman SI, Ladenson PW: Percutaneous transluminal angioplasty in hypothyroidism. Am J Med 90:367, 1991

47. Ladenson PW, Levin AA, Ridgway EC, et al: Complications of surgery in hypothyroid patients. Am J Med 77:261, 1984

48. Surks MI, Ortiz E, Daniels GH, et al: Subclinical thyroid disease: scientific review and guidelines for diagnosis and management. JAMA 291:228, 2004

49. Danese MD, Ladenson PW, Meinert CL, et al: Effect of thyroxine therapy on serum lipoproteins in patients with mild thyroid failure: a quantitative review of the literature. J Clin Endocrinol Metab 85:2993, 2000

50. Hak AE, Pols HAP, Visser TJ, et al: Subclinical hypothyroidism is an independent risk factor for atherosclerosis and myocardial infarction in elderly women: the Rotterdam Study. Ann Intern Med 132:270, 2000

51. Nystrom E, Caidahl K, Fager G, et al: A double-blind crossover 12-month study of L-thyroxine treatment of women with "subclinical" hypothyroidism. Clin Endocrinol (Oxf) 29:62, 1988

52. Danese MD, Powe NR, Sawin CT, et al: Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA 276:285, 1996

53. Bona M, Santini F, Rivolta G, et al: Cost effectiveness of screening for subclinical hypothyroidism in the elderly: a decision-analytical model. Pharmacoeconomics 14:209, 1998

54. Screening for thyroid disease: recommendation statement. U.S. Preventive Services Task Force. Ann Intern Med 140:125, 2004

55. Saravanan P, Chau WF, Roberts N, et al: Psychological well-being in patients on "adequate" doses of L-thyroxine: results of a large, controlled community-based questionnaire study. Clin Endocrinol (Oxf) 57:577, 2002

56. Walsh JP, Shiels L, Lim EM, et al: Combined thyroxine/liothyronine treatment does not improve well-being, quality of life, or cognitive function compared to thyroxine alone: a randomized controlled trial in patients with primary hypothyroidism. J Clin Endocrinol Metab 88:4543, 2003

57. Holvey DN, Goodner CJ, Nicoloff JT, et al: Treatment of myxedema coma with intravenous thyroxine. Arch Intern Med 113:89, 1964

58. Yamamoto T, Fukuyama J, Fujiyoshi A: Factors associated with mortality of myxedema coma: report of eight cases and literature survey. Thyroid 9:1167, 1999

59. Tunbridge WM, Evered DE, Hall R, et al: The spectrum of thyroid disease in a community: the Wickham Survey. Clin Endocrinol 7:481, 1977

60. Laurberg P, Pedersen KM, Vestergaard H, et al: High incidence of multinodular toxic goitre in the elderly population in a low iodine intake area vs. high incidence of Graves' disease in the young in a high iodine intake area: comparative surveys of thyrotoxicosis epidemiology in East-Jutland Denmark and Iceland. J Intern Med 229:415, 1991

61. Canaris GJ, Manowitz NR, Mayor G, et al: The Colorado thyroid disease prevalence study. Arch Intern Med 160:526, 2000

62. Brix TH, Christensen K, Holm NV, et al: A population-based study of Graves' disease in Danish twins. Clin Endocrinol (Oxf) 48:397, 1998

63. Yanagawa T, Hidaka Y, Guimaraes V, et al: CTLA-4 gene polymorphism associated with Graves' disease in a Caucasian population. J Clin Endocrinol Metab 80:41, 1995

64. Saravanan P, Dayan CM: Thyroid autoantibodies. Endocrinol Metab Clin North Am 30:315, 2001

65. Derwahl M, Studer H: Hyperplasia versus adenoma in endocrine tissues: are they different? Trends Endocrinol Metab 13:23, 2002

66. Brix TH, Kyvik KO, Hegedus L: Major role of genes in the etiology of simple goiter in females: a population-based twin study. J Clin Endocrinol Metab 84:3071, 1999

67. Bidey SP, Hill DJ, Eggo MC: Growth factors and goitrogenesis. J Endocrinol 160:321, 1999

68. Parma J, Duprez L, Van Sande J, et al: Diversity and prevalence of somatic mutations in the thyrotropin receptor and Gs alpha genes as a cause of toxic thyroid adenomas. J Clin Endocrinol Metab 82:2695, 1997

69. Stanbury JB, Ermans AE, Bourdoux P, et al: Iodine-induced hyperthyroidism: occurrence and epidemiology. Thyroid 8:83, 1998

70. Roti E, Uberti ED: Iodine excess and hyperthyroidism. Thyroid 11:493, 2001

71. Pearce EN, Farwell AP, Braverman LE: Thyroiditis. N Engl J Med 348:2646, 2003; erratum in N Engl J Med 349:620, 2003

72. Bartalena L, Wiersinga WM, Tanda ML, et al: Diagnosis and management of amiodarone-induced thyrotoxicosis in Europe: results of an international survey among members of the European Thyroid Association. Clin Endocrinol (Oxf) 61:494, 2004

73. Hershman JM: Physiological and pathological aspects of the effect of human chorionic gonadotropin on the thyroid. Best Pract Res Clin Endocrinol Metab 18:249, 2004

74. Glinoer D: The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. Endocr Rev 18:404, 1997

75. Hershman JM: Human chorionic gonadotropin and the thyroid: hyperemesis gravidarum and trophoblastic tumours. Thyroid 9:653, 1999

76. Yeo CP, Khoo DH, Eng PH, et al: Prevalence of gestational thyrotoxicosis in Asian women evaluated in the 8th to 14th weeks of pregnancy: correlations with total and free beta human chorionic gonadotrophin. Clin Endocrinol (Oxf) 55:391, 2001

77. Rodien P, Bremont C, Sanson ML, et al: Familial gestational hyperthyroidism caused by a mutant thyrotropin receptor hypersensitive to human chorionic gonadotropin. N Engl J Med 339:1823, 1998

78. McDermott MT, Ridgway EC: Central hyperthyroidism. Endocrinol Metab Clin North Am 27:187, 1998

79. Brucker-Davis F, Oldfield EH, Skarulis MC, et al: Thyrotropin-secreting pituitary tumors: diagnostic criteria, thyroid hormone sensitivity, and treatment outcome in 25 patients followed at the National Institutes of Health. J Clin Endocrinol Metab 84:476, 1999

80. Gershengorn MC, Weintraub BD: Thyrotropin-induced hyperthyroidism caused by selective pituitary resistence to thyroid hormone. J Clin Invest 56:633,1975

81. White JD, Ladenson PW: Thyroid hormone overdose. Clinical Management of Poisoning and Drug Overdose. Haddad LM, Winchester JE, Eds. WB Saunders Co, Philadelphia, 1988

82. Ladenson PW: Diagnosis of hyperthyroidism. The Thyroid, 9th ed. Braverman LE, Utiger RD, Eds. JB Lippincott, Philadelphia (in press)

83. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. National Academy of Clinical Biochemistry Guidelines Committee. Thyroid 13:3, 2003

84. Borst GC, Eil C, Burman KD: Euthyroid hyperthyroxinemia. Ann Intern Med 98:366, 1983

85. Rushbrook JI, Becker E, Schussler GC, et al: Identification of a human serum albumin species associated with familial dysalbuminemic hyperthyroxinemia. J Clin Endocrinol Metab 80:461, 1995

86. Moses AC, Rosen HN, Moller DE, et al: A point mutation in transthyretin increases affinity for thyroxine and produces euthyroid hyperthyroxinemia. J Clin Invest 86:2025, 1990

87. Sakata S: Autoimmunity against thyroid hormones. Crit Rev Immunol 14:157, 1994

88. Pearce EN, Bogazzi F, Martino E, et al: The prevalence of elevated serum C-reactive protein levels in inflammatory and noninflammatory thyroid disease. Thyroid 13:643, 2003

89. Cooper DS: Antithyroid drugs for the treatment of hyperthyroidism caused by Graves' disease. Endocrinol Metab Clin North Am 27:225, 2003

90. Brucker-Davis F, Oldfield EH, Skarulis MC, et al: Thyrotropin-secreting pituitary tumors: diagnostic criteria, thyroid hormone sensitivity, and treatment outcome in 25 patients followed at the National Institutes of Health. J Clin Endocrinol Metab 84:476, 1999

91. Beck-Peccoz P, Persani L: Medical management of thyrotropin-secreting pituitary adenomas. Pituitary 5:83, 2002

92. Nakazawa HK, Sakurai K, Hamada N, et al: Management of atrial fibrillation in the post-thyrotoxic state. Am J Med 72:903, 1982

93. Kahaly GJ, Dillmann WH: Thyroid hormone action in the heart. Endocr Rev Jan 4, 2005 [Epub ahead of print]

94. Dillmann WH: Thyroid storm. Curr Ther Endocrinol Metab 6:81, 1997

95. Sherman SI, Simonson L, Ladenson PW: Clinical and socioeconomic predispositions to complicated thyrotoxicosis: a predictable and preventable syndrome. Am J Med 101:192, 1996

96. Knobel M, Barca MF, Pedrinola F, et al: Prevalence of anti-thyroid peroxidase antibodies in autoimmune and nonautoimmune thyroid disorders in a relatively low-iodine environment. J Endocrinol Invest 17:837, 1994

97. Muller AF, Drexhage HA, Berghout A: Postpartum thyroiditis and autoimmune thyroiditis in women of childbearing age: recent insights and consequences for antenatal and postnatal care. Endocr Rev 22:605, 2001

98. Yasmeen T, Khan S, Patel SG, et al: Clinical case seminar: Riedel's thyroiditis: report of a case complicated by spontaneous hypoparathyroidism, recurrent laryngeal nerve injury, and Horner's syndrome. J Clin Endocrinol Metab 87:3543, 2002

99. Gaitan E: Goitrogens in food and water. Annu Rev Nutr 10:21, 1990

100. Hegedus L, Bonnema SJ, Bennedbaek FN: Management of simple nodular goiter: current status and future perspectives. Endocr Rev 24:102, 2003

101. Huysmans DA, Nieuwlaat WA, Hermus AR: Towards larger volume reduction of nodular goitres by radioiodine therapy: a role for pretreatment with recombinant human thyrotropin? Clin Endocrinol (Oxf) 60:297, 2004

102. Castro MR, Caraballo PJ, Morris JC: Effectiveness of thyroid hormone suppressive therapy in benign solitary thyroid nodules: a meta-analysis. J Clin Endocrinol Metab 87:4154, 2002

103. Hegedus L: Clinical practice: the thyroid nodule. N Engl J Med 351:1764, 2004

104. Tan GH, Gharib H: Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. Ann Intern Med 126:226, 1997

105. Gertner ME, Kebebew E: Multiple endocrine neoplasia type 2. Curr Treat Options

Oncol 5:315, 2004

106. Mandel SJ: Diagnostic use of ultrasonography in patients with nodular thyroid disease. Endocr Pract 10:246, 2004

107. Chow LS, Gharib H, Goellner JR, et al: Nondiagnostic thyroid fine-needle aspiration cytology: management dilemmas. Thyroid 11:1147, 2001

108. Sherman SI: Thyroid carcinoma. Lancet 361:501, 2003

109. Schneider AB, Ron E: Carcinoma of the follicular epithelium: epidemiology and pathogenesis. The Thyroid, 9th ed. Braverman LE, Utiger RD, Eds. JB Lippincott, Philadelphia, 2005, p 889

110. Jemal A, Tiwari RC, Murray T, et al: Cancer statistics, 2004. CA Cancer J Clin 54:8, 2004

111. Fagin JA: Challenging dogma in thyroid cancer molecular genetics: role of RET/PTC and BRAF in tumor initiation. J Clin Endocrinol Metab 89:4264, 2004

112. Ringel MD, Ladenson, PW: Controversies in the follow-up and management of well-differentiated thyroid cancer. Endocr Relat Cancer 11:97, 2004

113. Hay ID, Grant CS, Bergstralh EJ, et al: Unilateral total lobectomy: is it sufficient surgical treatment for patients with AMES low-risk papillary thyroid carcinoma? Surgery 124:958, 1998

114. Pearce EN, Braverman LE: Papillary thyroid microcarcinoma outcomes and implications for treatment. J Clin Endocrinol Metab 89:3710, 2004

115. Mazzaferri EL, Jhiang SM: Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med 97:418, 1994; erratum in Am J Med 98:215, 1995

116. Lippi F, Capezzone M, Angelini F, et al: Radioiodine treatment of metastatic differentiated thyroid cancer in patients on L-thyroxine, using recombinant human TSH. Eur J Endocrinol 144:5, 2001

117. Mazzaferri EL, Robbins RJ, Spencer CA, et al: A consensus report of the role of serum thyroglobulin as a monitoring method for low-risk patients with papillary thyroid carcinoma. J Clin Endocrinol Metab 88:1433, 2003

118. Haugen BR, Pacini F, Reiners C, et al: A comparison of recombinant human thyrotropin and thyroid hormone withdrawal for the detection of thyroid remnant or cancer. J Clin Endocrinol Metab 84:3877, 1999

119. Wilson PC, Millar BM, Brierley JD: The management of advanced thyroid cancer. Clin Oncol (R Coll Radiol) 16:561, 2004

120. Massoll N, Mazzaferri EL: Diagnosis and management of medullary thyroid carcinoma. Clin Lab Med 24:49, 2004

121. De Crevoisier R, Baudin E, Bachelot A, et al: Combined treatment of anaplastic thyroid carcinoma with surgery, chemotherapy, and hyperfractionated accelerated external radiotherapy. Int J Radiat Oncol Biol Phys 60:1137, 2004

122. Gilliland D, Hunt WC, Morris DM, et al: Prognostic factors for thyroid carcinoma: a population-based study of 15,698 cases from the Surveillance, Epidemiology and End Results (SEER) program 1973–1991. Cancer 79:564, 1997

Acknowledgment

Figures 2 and 3 Seward Hung.

II TESTES AND TESTICULAR DISORDERS

Peter J. Snyder, m.d.

The testes begin to function early in utero and continue to function into senescence, but the consequences of their function differ at different stages of life. Diseases that affect testicular function, therefore, also have different consequences at different stages of life. Testicular function can be affected by diseases of the hypothalamus and pituitary, as well as by diseases of the testes themselves. The diseases may be either congenital or acquired [*see Tables 1 and 2*].

Normal Testicular Function

The testes have two functions, the secretion of testosterone by the Leydig cells and the production of sperm by the seminiferous tubules. The cumulative effect of testosterone is to produce and maintain a phenotypic male. Testicular function is stimulated by the gonadotropins, follicle-stimulating hormone (FSH) and luteinizing hormone (LH), which are secreted by the gonadotroph cells of the pituitary gland, which in turn are stimulated by gonadotropin-releasing hormone (GnRH) from the hypothalamus.

gonadotropin synthesis and secretion

GnRH is a decapeptide cleaved from a larger precursor peptide that is synthesized in the arcuate nucleus of the hypothalamus. FSH and LH are synthesized in the gonadotroph cells of the pituitary. Each is a heterodimeric glycopeptide consisting of a common α subunit and a unique β subunit.

GnRH travels via the hypothalamic-pituitary portal circulation to the pituitary, where it binds to G protein–coupled receptors on the surface of the gonadotroph cell. This triggers a cascade of intracellular signaling pathways and stimulates LH and FSH release. Although GnRH cannot be measured readily in the portal or peripheral circulation of humans, its secretion is thought to be pulsatile, because LH secretion is pulsatile. In addition, administration of GnRH to men who have GnRH deficiency increases LH secretion to normal only if GnRH is administered in pulses.[1]

Table 1 Causes of Primary Hypogonadism

Congenital	Chromosomal abnormalities
	Klinefelter syndrome
	46 XX male
	Microdeletions of the long arm of the Y chromosome
	Cryptorchidism
	Disorders of androgen biosynthesis
	Myotonic dystrophy
Acquired	Orchitis (e.g., mumps)
	Ionizing radiation
	Drugs
	Alkylating agents
	Ketoconazole
	Alcohol
	Trauma
	Testicular torsion
	Autoimmune damage
	AIDS

testosterone

Synthesis and Secretion

LH stimulates testosterone synthesis by binding to a surface receptor on the Leydig cells and activating a cyclic adenosine monophosphate–mediated mechanism that increases cholesterol side-chain cleavage and conversion to pregnenolone and eventually to testosterone. Testosterone is synthesized at a rate of 5 to 7 mg a day. Testosterone secretion is episodic, like that of LH, and follows a diurnal pattern. In normal young men, the highest concentrations of testosterone occur at about 8 A.M.; the lowest, at about 8 P.M.[2]

Plasma Binding

Circulating testosterone is 98% to 99% bound. About 40% is bound to sex hormone–binding globulin (SHBG) with high affinity; 60% is bound to albumin with low affinity. Testosterone bound to SHBG is not available to tissues, but that bound to albumin probably is available. SHBG synthesis is stimulated by estrogens and decreased by androgens and obesity.

Actions

Testosterone has many different effects in many different tissues, at least partly because it can act on a cellular level as three

Table 2 Causes of Secondary Hypogonadism

Congenital	Isolated gonadotropin deficiency
	Isolated gonadotropin-releasing hormone (GnRH) deficiency
	With anosmia (Kallmann syndrome)
	With other abnormalities (Prader-Willi, Lawrence-Moon-Biedl syndromes)
	Without other abnormalities
	Mutations of the GnRH receptor, *LH*β, or *DAX-1* genes
	Multiple hypothalamic and pituitary hormone deficiencies
Acquired	Benign tumors and cysts
	Pituitary adenomas
	Craniopharyngiomas, dysgerminomas, Rathke pouch cysts
	Malignant tumors
	Metastases from lung, breast, and other malignancies
	Meningiomas, gliomas
	Lymphomas
	Infiltrative diseases (e.g., sarcoidosis, Langerhans cell histiocytosis, hemochromatosis
	Infectious diseases (e.g., tuberculosis, histoplasmosis)
	Infarction of the pituitary (e.g., Sheehan syndrome)
	Lymphocytic hypophysitis
	Trauma
	Surgery
	Radiation
	Systemic illness (starvation, anorexia, acute and chronic illness)
	Medications (glucocorticoids, megestrol acetate, suramin)
	Drugs of abuse (alcohol, opiates)
	Hyperprolactinemia
	Isolated acquired GnRH deficiency

distinct hormones: testosterone itself and its two metabolites, dihydrotestosterone (DHT) and estradiol. Both testosterone and DHT act by binding to the androgen receptor, which is encoded by a gene that is located on the X chromosome and belongs to the steroid-retinoid-thyroid hormone superfamily of receptors.[3] Although there is only one androgen receptor, the presence of coactivators or corepressors of transcription in certain types of cells could explain why the effects of testosterone vary in different tissues.[4]

The direct effect of testosterone on cells is mediated by its passive diffusion into cells, its binding to the androgen receptor, the binding of the testosterone–androgen receptor complex to DNA, and subsequent stimulation of messenger RNA (mRNA) and protein synthesis. This mechanism appears to be responsible for testosterone's stimulation of the wolffian ducts to become the male internal genitalia during embryonic development[5] and for testosterone's inhibition of gonadotropin secretion. Testosterone has a probable role in the stimulation of erythropoiesis; the growth of muscle; an increase in linear bone growth; and, to some degree, an increase in bone mineral density.

In tissues that express the enzyme 5α-reductase, testosterone is irreversibly converted to DHT in the target cell cytoplasm. Two forms of 5α-reductase have been identified: type 1, which is found predominantly in nongenital skin and the liver; and type 2, which is found predominantly in urogenital tissue in both men and women. DHT binds to the androgen receptor with greater affinity than does testosterone and so has a greater effect than testosterone. After DHT binds to the androgen receptor, the DHT-receptor complex binds to DNA, stimulating mRNA and protein synthesis. This mechanism appears to be responsible for male differentiation of the external genitalia in utero,[5] enlargement of the male external genitalia during puberty, and the development of sexual hair during puberty.

In tissues that express the enzyme complex aromatase—especially some hypothalamic nuclei, adipose tissue, liver, and perhaps bone—testosterone is converted to estradiol, which binds to an estrogen receptor. This mechanism appears to mediate several effects of testosterone. One effect is on bone: in males who have mutations of the gene coding for the estrogen receptor[6] or the aromatase enzyme,[7] the epiphyses do not close and the bones are osteoporotic. Another effect is on libido, as suggested by the case report of a man who lacked aromatase and had poor libido until he was treated with estradiol.[8] Aromatase appears to partially mediate the inhibition of LH secretion by testosterone and to entirely mediate the inhibition of FSH secretion by testosterone.[9]

SPERMATOGENESIS

Spermatogenesis occurs in the seminiferous tubules, stimulated principally by testosterone. The concentration of testosterone within the testes is 100 times that in the peripheral circulation. This high concentration, which is essential for spermatogenesis, probably results from both the LH-stimulated production of testosterone in the nearby Leydig cells and the FSH-stimulated binding of testosterone by androgen-binding protein, produced by the Sertoli cells of the seminiferous tubules. FSH also stimulates the Sertoli cells to secrete activin, which stimulates spermatogenesis. In addition, the Sertoli cells secrete inhibin, which inhibits FSH secretion by the pituitary.

Spermatogenesis takes approximately 3 months. Maturation of a spermatogonium to a mature spermatozoon takes approximately 75 days; passage through the epididymis, where motility is acquired, takes another 14 days.

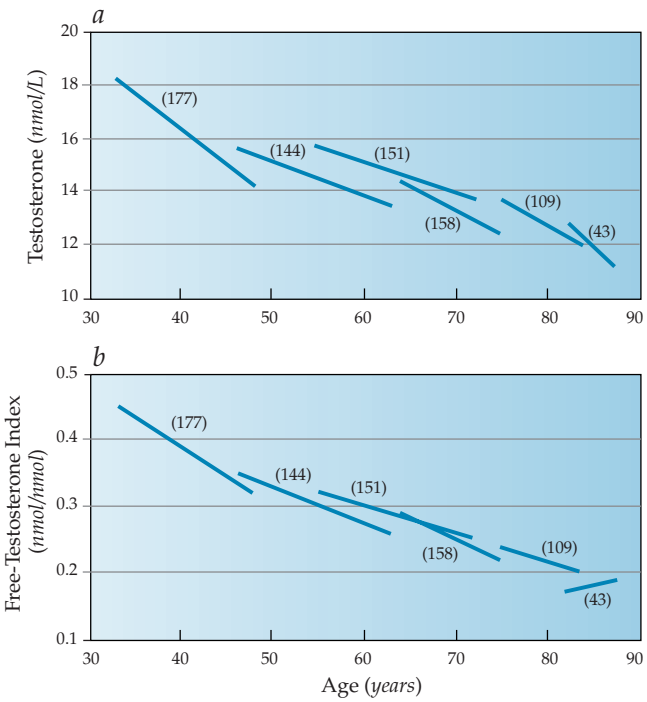


Figure 1 **Serum concentration of testosterone (*a*) and the free-testosterone index (*b*) versus age in healthy men who were followed longitudinally.[12] Each line represents the mean for a cohort of men, and the numbers in parentheses represent the number of men in each cohort.**

TESTICULAR FUNCTION THROUGHOUT LIFE

Testicular function begins in utero, increases briefly in the first few months of infancy, develops fully during puberty, and declines gradually during adulthood.

In Utero

Sexual differentiation occurs during the first trimester in utero.[10] In the presence of the *SRY* gene, which is the sex-determining region of the Y chromosome, the undifferentiated gonads become testes. The Sertoli cells of the testes secrete antimüllerian hormone, which suppresses the müllerian ducts, thereby preventing the development of female internal genitalia.[11] Human chorionic gonadotropin (hCG) from the placenta stimulates the Leydig cells of the testes to secrete testosterone, which in turn stimulates the nearby wolffian ducts to become the male internal genitalia—the vas deferens and seminal vesicles. In addition, testosterone is converted to DHT by the anlage of the external genitalia, and DHT influences the anlage to become the penis, scrotum, and prostate. During the third trimester, LH from the fetal pituitary stimulates the fetal testes to secrete testosterone, which results in penile growth. During the first few months post partum, there is a third testosterone elevation, to levels approximating that of a midpubertal boy. The consequences of this elevation are not known. Thereafter, the serum testosterone concentration falls to relatively low values until puberty.

Puberty

Puberty in boys begins with an increase in the secretion of LH and FSH by the pituitary, which is presumably stimulated by an increase in GnRH secretion by the hypothalamus. This takes place at a mean age of 11.4 years (with a standard deviation of ± 1.1 years). The initial consequence is an increase in testicular volume, from 2 ml toward its adult volume of 20 to 25 ml. Rising levels of circulating testosterone cause an increase in the size of the phallus and promote the growth of pubic, axillary, and eventually body and facial hair, along with regression of temporal scalp hair. Testosterone also causes an increase in long bone growth and body height, thickening of the vocal cords and lowering of the voice, and an increase in hemoglobin concentration. Most of these changes are completed within 4 to 5 years, but full development of body hair and beard may take several more years, and temporal scalp hair regression continues for decades.

Senescence

As men age, their serum total testosterone concentration decreases [*see Figure 1*].[12-14] The decrease in the serum concentration of total testosterone is very gradual and of relatively small magnitude. SHBG, however, increases with increasing age, so the free-testosterone concentration decreases to a greater degree than the total. By 80 years of age, according to cross-sectional studies, the free-testosterone concentration is one half to one third that at 20 years of age.[15,16] The decrease in testosterone appears to result from both decreased LH secretion and decreased responsiveness of the Leydig cells.[17] The parallels between male senescence and male hypogonadism—both of which are marked by decreases in libido, energy, muscle mass and strength, and bone mineral density—suggest that testosterone deficiency could be a cause of the changes in male senescence.[17] Serum estradiol concentration also decreases with increasing age, however. Several studies show a better correlation of some consequences of aging, such as the decrease in bone mineral density, with the estradiol level than with the testosterone level. Nevertheless, there is preliminary evidence that increasing the serum testosterone concentration of elderly men to that of young men increases bone mineral density[18] and muscle mass and decreases fat mass,[19] which supports the notion that the decrease in testosterone does have adverse consequences.

Male Hypogonadism

ETIOLOGY

Male hypogonadism can occur as a consequence of a disease of the testes (primary hypogonadism) or as a consequence of a disease of the pituitary or hypothalamus (secondary hypogonadism). Certain clinical findings suggest hypogonadism, but these are usually nonspecific, so the diagnosis must be confirmed by laboratory tests.

DIAGNOSIS

Clinical Findings

The clinical findings of hypogonadism result from either decreased spermatogenesis or decreased testosterone secretion. The sole clinical finding of decreased spermatogenesis is infertility. In contrast, decreased testosterone secretion causes a wide variety of clinical findings; specific findings depend on the stage of life in which the deficiency occurs. When testosterone deficiency occurs in the first trimester in utero, male sexual differentiation is incomplete. Complete lack of testosterone during this period results in female external genitalia (i.e., clitoris and labia). Incomplete testosterone deficiency causes partial virilization, ranging from posterior labial fusion when testosterone deficiency is severe to hypospadias when testosterone deficiency is mild. Testosterone deficiency that begins in the third trimester in utero results in normal male sexual differentiation but microphallus at birth. When testosterone deficiency occurs in childhood but before puberty, the result is incomplete puberty. When testosterone deficiency develops after puberty, some pubertal changes regress; such changes usually occur slowly, and the effects can occur at different rates. Energy and libido diminish within days to weeks of the fall in testosterone, and the hemoglobin concentration and hematocrit decline within a few months. Decreases in sexual hair, muscle mass, and bone mineral density are usually not recognized for several years.

Physical Examination

The physical examination focuses primarily on whether sexual development is consistent with the patient's age. If the patient is an adult, he should have facial, chest, and other body hair; temporal scalp hair should be receding appropriately for the patient's age and family pattern; and pubic hair should be dense and in a diamond pattern. The voice should be appropriately deep. Musculature should be normal for a man. Subcutaneous fat should be less than that of a boy or a woman. The testes should be 4 to 7 cm in length (20 to 25 ml in volume). If the patient is an adolescent, development should be appropriate for his age. If the patient is a child, the testes should be descended, and no hypospadias should be present.

The physical examination should also include evaluation for possible eunuchoid proportions and gynecomastia. An adult male usually has an upper body segment approximately equal to his lower segment and an arm span equal to his height. The absence of testosterone and the continued presence of growth hormone during puberty, as occurs in primary hypogonadism and isolated secondary hypogonadism, causes a delay in epiphyseal closure and an increase in the length of the long bones. In such patients, the lower body segment becomes longer than the upper and the arms become longer than the legs—a relationship known as eunuchoid proportions. This relationship persists even after testosterone treatment. Consequently, a man of any age who has a heel-to-pubis measurement more than 2 cm longer than his pubis-to-crown measurement and an arm span more than 2 cm longer than his height was probably hypogonadal during adolescence. Gynecomastia often occurs in hypogonadism; it is especially common in patients with primary hypogonadism.

Laboratory Findings

Once the diagnosis of hypogonadism has been suspected on the basis of symptoms and signs, the diagnosis must be confirmed by documenting decreased production of sperm or testosterone. If hypogonadism is confirmed, the next step is to measure LH and FSH. Elevated serum concentrations of LH and FSH indicate primary hypogonadism, whereas subnormal or normal values indicate secondary hypogonadism.

Spermatogenesis Sperm production can be assessed most readily by counting the sperm in an ejaculated semen specimen. Generally accepted normal values for ejaculated sperm are a density of greater than 20×10^6 sperm/ml of ejaculate and a total

count of more than 40×10^6 sperm/ejaculate. More than 60% of the sperm should be motile, and more than 30% should be normal in morphology. A recent study of the male partners in 765 infertile couples and 696 fertile couples showed that fertility was associated with a sperm density of greater than 48×10^6 sperm/ml, sperm motility of greater than 63%, and normal morphology in more than 12% of sperm.[20] Low fertility was associated with a sperm density of less than 13.5×10^6 sperm/ml, sperm motility of less than 32%, and normal morphology in less than 9%. Indeterminate fertility was associated with intermediate values. A severely subnormal sperm count (e.g., $< 5 \times 10^6$ sperm/specimen) can result from either primary or secondary hypogonadism. A normal or mildly subnormal sperm count (e.g., 35×10^6 sperm/specimen) associated with markedly abnormal sperm motility more likely indicates a primary spermatogenic abnormality and less likely indicates secondary hypogonadism.

Testicular biopsy usually provides no more information about spermatogenesis than a semen analysis, because the variety of histologic responses to testicular injury is very limited. Testicular biopsy is likely to be helpful only when the ejaculated semen contains no sperm but the testicular size is normal and the serum concentrations of testosterone, LH, and FSH are normal. Such a patient may have obstruction of the ejaculatory outflow, or he may have suffered damage to the seminiferous tubules sufficient to impair spermatogenesis but not sufficient to cause an elevation in the serum FSH concentration. A testicular biopsy showing normal seminiferous tubules would favor the former diagnosis.

Testosterone concentration Testosterone secretion is best evaluated by measuring the serum concentration of total testosterone, because the total testosterone level is usually an accurate reflection of the free-testosterone level. Also, most of the current assay techniques for free testosterone are not as accurate as those for total testosterone. Testosterone is secreted into the circulation episodically, in a diurnal pattern; the serum testosterone concentration is highest at about 8 A.M. and lowest at about 8 P.M.[2] Therefore, the serum testosterone concentration should be measured at 8 A.M. If the result is low or borderline, the test should be repeated. Measurement of free testosterone and SHBG may be helpful in situations in which the total testosterone level does not accurately reflect the free-testosterone level, such as would be the case with obese patients. If free testosterone is measured, the assay method should be equilibrium dialysis.

Gonadotropins If the testosterone concentration is low, serum LH and FSH concentrations should be measured. If those values are high, the patient has primary hypogonadism; otherwise, he has secondary hypogonadism. In a patient with a distinctly subnormal sperm count but a normal serum testosterone concentration, the combination of an elevated FSH concentration and a normal LH concentration indicates that there has been damage to the seminiferous tubules but that the Leydig cells have not been affected.

In patients with secondary hypogonadism, magnetic resonance imaging of the sellar region is indicated. The MRI scan will show whether the patient has a mass lesion and, if so, whether it is in the pituitary, the hypothalamus, or the parasellar region. Pituitary and hypothalamic lesions cannot be distinguished on the basis of the LH response to a single dose of exogenous GnRH. Administration of repeated doses of exogenous GnRH, however, will result in a normal LH response to an individual dose of GnRH in patients who have hypothalamic disease, but not in patients who have pituitary disease. In patients with hypothalamic disease, the length of time required for LH response to become normal varies widely.

DISEASES THAT CAUSE HYPOGONADISM

Overall, primary hypogonadism [*see Table 1*] is more common than secondary hypogonadism [*see Table 2*]. Once a patient's hypogonadism has been identified as primary or secondary, the specific etiology can be sought.

Primary Hypogonadism

Primary hypogonadism may be congenital or acquired. Many cases of primary hypogonadism have no identifiable cause, however. Presumably, many causes are yet unknown.

Congenital Of the congenital abnormalities that cause primary hypogonadism, the most common is Klinefelter syndrome,[21] which occurs in approximately 0.2% of newborns. It is the phenotypic presentation of a male with more than one X chromosome. The most common genotype is 47 XXY, but additional X chromosomes (e.g., 48 XXXY) and mosaics (e.g., 46 XY/47 XXY) have also been reported. The 47 XXY genotype results from nondisjunction of the sex chromosomes of either parent during meiotic division. Mosaicism probably results from nondisjunctive mitotic division after conception. The severity of the phenotypic consequences usually increases with the number of extra X chromosomes. The gonadal consequences are usually severe damage to the seminiferous tubules and variable damage (minimal to severe) to the Leydig cells. Consequently, men with Klinefelter syndrome usually have very small testes, no sperm in their ejaculate, infertility, and markedly high serum FSH concentrations. Their serum testosterone concentrations vary from normal to subnormal; correspondingly, their virilization varies from normal to low and their serum LH concentrations vary from normal to elevated. Klinefelter syndrome is also usually marked by abnormalities of behavior and of the long bones. These abnormalities are not directly related to the gonadal abnormalities. The behavioral abnormality is manifested as difficulty in social interactions that is recognized in childhood, and it leads to problems in school and eventually in work. The long-bone abnormality is increased length of the legs but not the arms; this abnormality occurs independently of increased length of both the arms and legs as a result of testosterone deficiency.

The diagnosis of Klinefelter syndrome can usually be made by determining the karyotype of the peripheral leukocytes. Testosterone deficiency, if present, can be treated with testosterone replacement (see below). The behavioral abnormality cannot be treated satisfactorily, but a support group can be helpful for the patient's family, and school counselors should be advised of the diagnosis.

Cryptorchidism, or undescended testes, is also associated with damage to the testes and with greater damage to the seminiferous tubules than to the Leydig cells. More than one mechanism may be involved: testosterone deficiency in utero may inhibit descent, and the heat of the abdomen may cause further damage to the undescended testis. The clinical consequences depend partly on whether one or both testes are undescended. If only one testis is undescended, there is a 25% to 33% likelihood that the sperm count will be subnormal and the serum FSH level slightly high.[22] If both testes are undescended, the sperm count will likely be severely subnormal and the patient infertile; the

serum testosterone concentration may be subnormal, and the patient may be undervirilized as well. Neoplasms are two to five times more likely to develop in cryptorchid testes.[23] The diagnosis is made in patients younger than 1 year by failure to palpate a testis that either is within the scrotum or can be manipulated manually from the inguinal canal into the scrotum.

Varicocele—a varicosity of the venous plexus within the scrotum—has for decades been considered a possible cause of infertility. The proposed mechanism is that varicocele causes an increase in blood flow, which impairs spermatogenesis by raising scrotal temperature above normal. However, scrotal temperatures are similar in infertile men with and without varicoceles, and varicoceles are not much more common in infertile than fertile men, so it is not certain that varicocele can cause infertility. More important, in a randomized trial of the surgical treatment of varicocele in men who were infertile, fertility was not found to be improved as a result of treatment.[24] Therefore, surgical treatment of a varicocele cannot be recommended as a means of improving fertility.

Congenital deficiency of testosterone production can also result from mutations of genes that encode enzymes necessary for androgen biosynthesis. These disorders are rare. The cholesterol side-chain cleavage enzymes 3β-hydroxysteroid dehydrogenase and 17α-hydroxylase occur in the adrenal as well as in the testes, so deficiencies of either of these enzymes lead to deficient cortisol secretion as well. Deficiency of 17β-hydroxysteroid oxidoreductase affects only the testes. All of these disorders result in deficient testosterone secretion, beginning in the first trimester in utero, and subsequent incomplete virilization. The degree of incompleteness, especially of phallic development, influences whether these babies are raised as boys or girls. The testosterone deficiency itself can be treated in the same way as testosterone deficiency from any other cause.

Deletions on the long arm of the Y chromosome appear to be associated with infertility. Azoospermia is more common than oligospermia in such cases.[25]

Acquired Many acquired illnesses can cause primary hypogonadism. These include infections—notably, mumps orchitis. Orchitis is an uncommon complication of mumps and may be unilateral. In bilateral cases, both testes initially become markedly swollen and severely painful, then gradually atrophy. Diminished sperm production is common; decreased testosterone secretion is less common. The diagnosis is made by eliciting a history of painful swelling of the testes during systemic mumps infection.

Treatment of neoplasms with chemotherapeutic drugs (especially alkylating agents) or with radiation therapy to the inguinal lymph nodes often damages the seminiferous tubules; less often, it damages the Leydig cells. Radiation causes damage despite shielding of the testes, because of radiation scatter. The degree of damage is usually proportionate to the radiation dose. In cases of less extensive treatment, the damage may be reversible. No specific remedy for such damage is available, however.

Medications and drugs of abuse can produce hypogonadism. The antifungal agent ketoconazole impairs testosterone production. Heavy alcohol ingestion damages the testes.

HIV infection and AIDS wasting are commonly associated with hypogonadism.[26] Several mechanisms appear to be involved in these cases. Some men with HIV infection and subnormal serum testosterone concentrations have inappropriately low serum concentrations of LH. This may be the result of conditions such as malnutrition, opiate abuse, and megestrol acetate administration, all of which are known to cause secondary hypogonadism. Other men with HIV infection lack known risk factors for secondary hypogonadism but have elevated serum concentrations of LH, indicating primary hypogonadism. Hypogonadism in HIV-infected men has been observed less commonly since the introduction of retroviral therapy.

Testicular torsion can cause permanent damage if not treated promptly. Trauma to the testes can sometimes be sufficiently severe to damage them.

Hypogonadism may be induced (surgically or chemically) as a therapeutic strategy in cases of advanced prostate cancer [*see 12:IX Prostate Cancer*]. Bilateral orchiectomy is used as a treatment for bilateral testicular cancer. In testicular cancer patients, however—unlike those treated with castration for prostate cancer—there is no reason to withhold testosterone replacement.

Secondary Hypogonadism

Like primary hypogonadism, secondary hypogonadism (also called hypogonadotropic hypogonadism) has both congenital and acquired causes in men [*see Table 2*]. Unlike primary hypogonadism, secondary hypogonadism often has a cause that is amenable to specific treatment. For that reason, finding the cause carries particular importance. Pituitary adenomas, other benign tumors and cysts of the sellar area, and malignancies that arise in the sellar region or metastasize there can usually be detected by MRI. Infiltrative diseases (e.g., sarcoidosis, hemochromatosis) usually produce manifestations in other organ systems that suggest the diagnosis. Tumors, cysts, and infiltrative lesions are often accompanied by deficiencies of other hypothalamic or pituitary hormones.

Some cases of secondary hypogonadism are not associated with any other hormonal abnormalities and are called isolated. Some cases appear to be caused by a deficiency of GnRH secretion by the hypothalamus; such cases can be congenital or acquired. When congenital, they may or may not be a part of Kallmann syndrome.[27] Patients with Kallmann syndrome have deficient GnRH secretion, variably associated with anosmia, cryptorchidism, red-green color blindness, and long-bone and urogenital tract abnormalities. Kallmann syndrome may occur sporadically or in families; familial cases can be inherited in an autosomal dominant pattern, with expression mostly limited to males, or in an X-linked recessive pattern. The genetic defect responsible both for the deficiency in GnRH secretion and for anosmia in some patients who have the X-linked recessive form of Kallmann syndrome is a mutation in the *KAL-1* gene, which encodes a neural cell adhesion protein, anosmin. When this protein is not present during embryogenesis, GnRH-secreting neurons do not migrate from the olfactory placode to the olfactory bulb and then to the hypothalamus, resulting in both anosmia and hypogonadotropic hypogonadism.

Another cause of isolated secondary hypogonadism is a mutation of the GnRH receptor. In these cases, GnRH is secreted by the hypothalamus but does not stimulate LH secretion by the pituitary. A third cause is a mutation of the *DAX-1* gene, which leads to hypogonadotropic hypogonadism and to adrenal hypoplasia congenita.

Gonadotropin secretion can be reversibly inhibited by any systemic illness or by hyperprolactinemia. Inhibition from medications, such as glucocorticoids, suramin, and opiates, is also reversible. Since the introduction and widespread use of controlled-release forms of opioids for chronic-pain management,

hypogonadism from these medications has become more common.[28] Heroin addicts may experience hypogonadism by the same mechanism. Damage to the pituitary from surgery or radiation, in contrast, usually results in permanent inhibition of gonadotropin secretion.

Delayed puberty is diagnosed in any boy whose pubertal development does not begin by more than two standard deviations past the mean age. In some cases, this delay represents a normal variant; these patients eventually enter puberty spontaneously. In other cases, the delay is caused by secondary hypogonadism. Distinguishing a normal variant from pathologic delay can be difficult. The degree of hypogonadism is usually not helpful in making this distinction, nor is any biochemical test. A family history of delayed puberty or constitutional short stature increases the likelihood of physiologic delayed puberty. Anosmia, symptoms of a chiasmal lesion, or other signs of a specific hypothalamic or pituitary disease increase the likelihood of an organic lesion as the cause. In many cases, the diagnosis can be made only by continued observation.

In an otherwise healthy elderly man, an unequivocally subnormal serum testosterone concentration, along with an LH concentration that is not elevated, can be considered a form of secondary hypogonadism.

TREATMENT

Testosterone Replacement

Testosterone can be replaced whether the hypogonadism is primary or secondary. Unlike estrogen, testosterone itself is not suitable for oral replacement, because it is catabolized rapidly during its first pass through the liver. Derivatives of testosterone that are alkylated in the 17α position do not undergo this rapid hepatic catabolism; however, these agents appear to lack the full virilizing effect of testosterone, and they may cause hepatic toxicity, including cholestatic jaundice, a cystic condition of the liver called peliosis, and, possibly, hepatocellular carcinoma. Consequently, the 17α-alkylated androgens should not be used to treat testosterone deficiency.

Currently, replacement therapy is instead delivered by the intramuscular or transdermal routes. The intramuscular formulations, testosterone enanthate and testosterone cypionate, are long-acting esters of testosterone produced by esterifying the hydroxyl group in the 17β position with a fatty acid. These do produce full virilization. They are usually administered in doses of 150 to 200 mg by deep intramuscular injection every 2 weeks. With this regimen, serum testosterone values peak within 1 to 2 days after the injection and fall to a nadir just before the next injection [*see Figure 2*].[29] These fluctuations are noticed by some patients as fluctuations in energy, mood, and libido.

Transdermal testosterone is now available in both patch[30] and gel[31] form [*see Figure 2*]. In most hypogonadal men, these preparations usually produce serum testosterone concentrations that are within the normal range and that fluctuate no more than physiologically, resulting in reasonable stability of energy, mood, and libido. The relatively physiologic pattern of serum testosterone concentrations and the infrequency of side effects make transdermal preparations the best means of testosterone replacement for most hypogonadal men.

During replacement therapy, clinicians should monitor patients for the efficacy and side effects of testosterone. Efficacy is determined by measurement of the serum testosterone concentration, which should be in the middle of the normal range midway between injections of testosterone esters and at any time after application of a transdermal preparation. Serum testosterone concentrations can vary with any of these preparations, however, so testosterone should be measured more than once to determine whether the initial dose is optimal. Serum testosterone should be measured again after a dose is changed and then once or twice a year. If the serum testosterone concentration is maintained within the normal range, the patient should experience reversal of the consequences of testosterone deficiency. Specifically, energy, libido, hemoglobin concentration, muscle mass, and bone density will increase.[32]

Men older than 40 years who are receiving testosterone replacement should be monitored for testosterone-dependent diseases, such as prostate cancer, benign prostatic hyperplasia, and erythrocytosis. However, there is as yet no evidence that exogenous testosterone is more likely to exacerbate any of these conditions than is endogenous testosterone.

Stimulation of Spermatogenesis

When sperm production is impaired by damage to the seminiferous tubules, no treatment can improve fertility. However, if some mature sperm are produced, they may be used for in vitro fertilization. When the sperm count is low because of pituitary or hypothalamic disease, sperm production can often be stimulated to within the normal range by administration of exogenous gonadotropins. If the hypogonadism occurred postpubertally, usually only LH need be replaced. If the hypogonadism occurred prepubertally, usually both LH and FSH need to be replaced.[33] In hypogonadism secondary to hypothalamic disease, spermatogenesis can also be stimulated by pulsatile administration of GnRH.

Androgen Insensitivity

Generalized tissue insensitivity to the action of androgens results in abnormalities similar to those of testosterone deficiency. Such insensitivity can be caused by abnormalities of either the androgen receptor [34] or 5α-reductase type 2 enzyme.[35] Both conditions result from genetic mutations, both are rare, and both result in incomplete virilization of the external genitalia.

Many different mutations of the androgen receptor gene have been described. Some of these mutations interfere with binding of androgen to the receptor, and others interfere with binding of the androgen-receptor complex to the DNA of the androgen-responsive cell.

The clinical presentations of the different receptor abnormalities also vary. In the most severe clinical presentation, called complete androgen insensitivity, the affected person is born with testes in the inguinal canals, no internal genitalia, and female external genitalia. At puberty, the serum testosterone concentration increases to a high-normal or slightly above-normal value, but sexual hair does not develop. Breasts do develop, because testosterone can still be converted to estradiol. In incomplete androgen insensitivity, there is variable partial fusion of the labial folds: a lesser degree of fusion results in genitalia that are still more female than male, and a greater degree results in genitalia that are more, but still incompletely, male. The least severe form of androgen resistance is manifested only by infertility and sometimes gynecomastia. Serum testosterone and LH concentrations are high normal to slightly high in all of these forms of androgen resistance.

Kennedy disease consists of a relatively mild form of androgen insensitivity but progressively severe spinal bulbar neuropa-

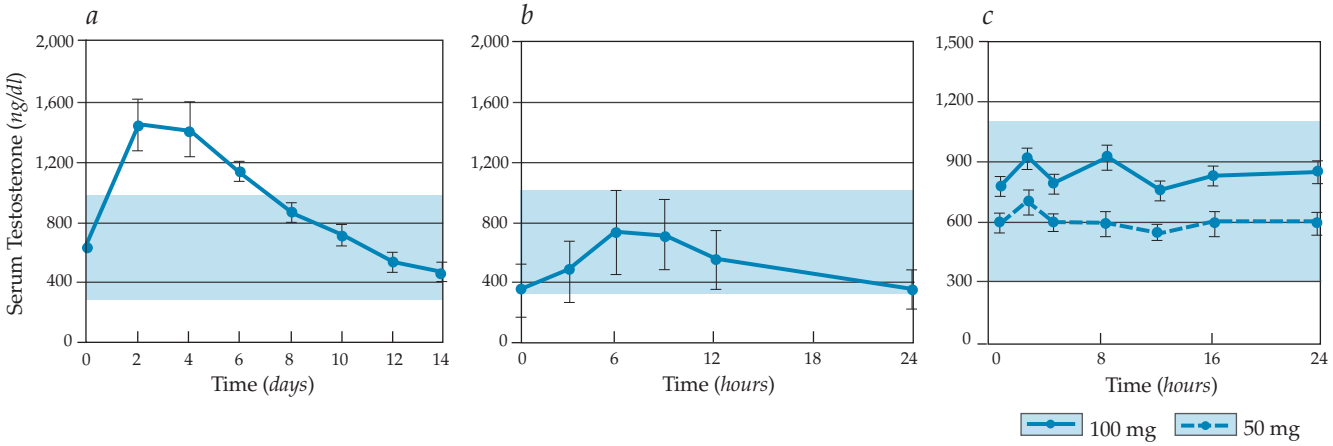


Figure 2 **Serum testosterone concentrations during the course of chronic administration of three different testosterone preparations to hypogonadal men. (*a*) Concentrations during 14 days after the injection of 200 mg of testosterone enanthate.[29] (*b*) Concentrations during the 24 hours after application of a testosterone patch that delivers approximately 5 mg of testosterone.[41] (*c*) Concentrations during the 24 hours after application of a testosterone gel containing 50 or 100 mg of testosterone.[31]**

thy. It is caused by a mutation leading to doubling of the number of N-terminal glutamines in the androgen receptor.

Other mutations of the androgen receptor have been described in metastatic prostate cancer that has become resistant to androgen deprivation. These mutations may allow ligands other than testosterone to bind to the receptor and stimulate receptor-dependent functions even in the absence of testosterone.

Mutations of the 5α-reductase type 2 gene also lead to incomplete virilization of the external genitalia because virilization requires conversion of testosterone to DHT by this enzyme.[35] At birth, most males with 5α-reductase type 2 deficiency have external genitalia that are predominantly female, so the sex of rearing is usually female. At puberty, the serum testosterone concentration increases and overcomes the lack of DHT to some degree; binding of testosterone to the androgen receptor causes phallic enlargement and hair growth in an adult male pattern. If the patient is being raised as a female, the testes can be removed before puberty to prevent these events.

Gynecomastia

Gynecomastia is the development of glandular breast tissue in a man. In most cases of gynecomastia, the stimulation of glandular tissue appears to result from an increased ratio of estrogen to androgen.

EPIDEMIOLOGY

Gynecomastia is common at all ages of life but is more common in infancy, puberty, and from middle age on. In most large series, gynecomastia can be detected in 10% to 20% of midpubertal boys and in more than 50% of men older than 50 years. These high rates probably reflect the hormonal changes that typically occur at those ages and do not represent disease.

ETIOLOGY AND PATHOGENESIS

The common mechanism for gynecomastia appears to be an increased estrogen-to-androgen effect on the breast. This may involve an increase in estrogen effect or a decrease in androgen effect, or both, and may result from changes in hormonal production or in hormonal action at the cellular level. Normally, most of the estrogen in the peripheral circulation in men is produced by the conversion of testosterone to estradiol or of androstenedione to estrone by the enzyme complex aromatase, which is concentrated in adipose tissue and the liver.

Exposure to Exogenous Estrogen

Gynecomastia from exogenous estrogens is uncommon. Reported cases have involved exposure to a partner's vaginal cream, application of antibalding creams, dietary intake, and occupational exposure.

Increased Estrogen Secretion

The most common cause of increased endogenous estrogen secretion is increased gonadotropin stimulation of the testes, which increases intratesticular aromatase levels and thereby increases the amount of estradiol secreted relative to testosterone. This is the likely mechanism by which gynecomastia occurs in normal males during puberty; with refeeding after starvation; and after successful treatment of severe illness, such as chronic cardiac, hepatic, or renal disease. In these situations, a period of secondary hypogonadism is followed by normal gonadotropin secretion. Increased gonadotropin (specifically, LH) secretion is also the cause of gynecomastia in primary hypogonadism. Increased secretion of hCG is the cause in patients who have tumors of the testes or liver. Administration of hCG therapeutically (e.g., to stimulate spermatogenesis) acts similarly, especially when the dose is excessive.

Increased Peripheral Conversion of Androgens to Estrogens

Increased peripheral conversion of testosterone to estradiol or of androstenedione to estrone can occur via several mechanisms: (1) an increased rate of conversion, as in hyperthyroidism or cirrhosis of the liver; (2) an increase in the amount of aromatase, as occurs in obesity; and (3) an increased substrate for aromatization, as occurs when an adrenal carcinoma secretes large amounts of androstenedione or when an aromatizable androgen, such as a long-acting testosterone ester, is administered in excessive doses.

Inhibition of Androgen Binding

Many drugs that cause gynecomastia appear to do so by binding to the androgen receptor and thereby blocking endogenous testosterone. As a consequence, endogenous androgens have

less androgenic action but are still converted to estrogens. Drugs that can block the androgen receptor include spironolactone, cimetidine, flutamide, bicalutamide, and cyproterone acetate. Inherited disorders of the androgen receptor produce similar results, although in these cases, the inhibition of binding is, of course, irreversible.

DIAGNOSIS

The diagnosis of gynecomastia is confirmed by physical examination. The examiner places a spread thumb and forefinger above and below the patient's nipple and draws them together, like calipers, toward the nipple, hugging the chest wall. Subcutaneous tissue feels soft as the fingers are drawn together, whereas gynecomastia feels firm. The diameter of the gynecomastia can be measured with a ruler. Mammography is usually unnecessary. Gynecomastia is generally bilateral, although it is occasionally unilateral. It is often asymmetrical. If the tissue is tender, the gynecomastia is more likely to be of recent origin.

Having confirmed that gynecomastia exists, the clinician then needs to find the cause. This involves inquiring about medications and searching for diseases known to cause gynecomastia [*see* Etiology and Pathogenesis, *above*].[36]

DIFFERENTIAL DIAGNOSIS

Gynecomastia must be distinguished from carcinoma of the male breast, which is rare, and from adiposity, which is common. Breast cancer should be suspected when the breast enlargement is unilateral, nontender, not centered directly under the nipple, and hard. The diagnosis can be confirmed by mammography. Breast adiposity is bilateral and can usually be distinguished from gynecomastia by the absence of palpable glandular tissue on physical examination.

TREATMENT

Gynecomastia is not physically harmful, and it usually regresses once the cause has been removed, although regression may take many years. Therefore, treatment is indicated only if the gynecomastia is causing psychological distress. The only accepted treatment is surgical removal, which is best performed by a plastic surgeon. Small series suggest that the antiestrogen drug tamoxifen and the aromatase inhibitor testolactone can reduce gynecomastia. Anastrozole is a more potent aromatase inhibitor than testolactone, but no results of its efficacy in treating gynecomastia have been published. None of these drugs have been approved by the Food and Drug Administration for this purpose.

Erectile Dysfunction

Erectile dysfunction is the inability to achieve or maintain an erection sufficient for intercourse. Although occasional erectile dysfunction does not indicate disease, its occurrence on most attempts to engage in sexual activity may indicate disease and is usually very troubling to the patient and his partner. Erectile dysfunction is not the same as a decrease in libido, which is decreased sexual interest. The two usually have different causes and, therefore, different treatments.

EPIDEMIOLOGY

In a cross-sectional survey of noninstitutionalized men 40 to 70 years of age, all degrees of erectile dysfunction, from minimal to complete, occurred in 50% of men. Older men were more commonly affected, however. For example, impotence occurred in 5% of 40-year-old men and in 15% of 70-year-old men.[37] When men from the same population were followed longitudinally, the incidence of new cases was 2.5% a year.[38]

PATHOPHYSIOLOGY

Development of an erection requires intact psychological, neurologic, and vascular mechanisms.[39,40] Erotic stimuli result in neural impulses that are carried from the cerebral cortex to the penis via the spinal cord, and stimulation of the penis results in neural impulses that loop to the spinal cord and back to the penis via parasympathetic nerves. These stimuli trigger blood flow into the corpora cavernosa. The inflow of blood is mediated by relaxation of arteriolar smooth muscle under the influence of nitric oxide, the production of which is catalyzed by the enzyme nitric oxide synthetase. Nitric oxide, in turn, promotes the production of cyclic guanosine monophosphate (cGMP), which also relaxes arteriolar smooth muscle and increases blood flow. Outflow of blood from the engorged corpora cavernosa is impeded by an increase in venous resistance.

Disruption of any of these steps can lead to erectile dysfunction. The neural mechanisms can be disrupted mechanically (such as by radical prostatectomy, surgery for an abdominal aortic aneurysm, or spinal cord trauma) or pathologically (such as by diabetic autonomic neuropathy or autonomic insufficiency syndromes). They also can be disrupted functionally, such as by drugs. The vascular mechanism can be disrupted by large vessel (atherosclerotic) or small vessel (diabetic) disease. The mechanism by which hyperprolactinemia causes impotence is not known, although it does not appear to be via hypogonadism.

DIAGNOSIS

History, physical examination, and laboratory testing all contribute to finding the cause of erectile dysfunction. From the history, the physician can determine whether the patient is taking a drug or has a disease associated with this disorder. Several drugs that are used to treat hypertension may occasionally cause erectile dysfunction; these include thiazide diuretics and alpha- and beta-blocking agents. Other drugs include tranquilizers and antidepressants. Excessive alcohol ingestion can also cause erectile dysfunction. The physician should ask about a history of diabetes of long duration, including other manifestations of diabetic neuropathy, and explore psychogenic factors, including depression, anxiety, fatigue, interpersonal stresses, and chronic illness. It is useful to ask whether the patient can obtain an erection under any circumstances. If the patient has an erection on awakening in the morning or can get an erection on some occasions but not others, the cause is more likely to be psychogenic than organic.

On physical examination, the absence of peripheral pulses and the presence of femoral bruits indicate vascular disease. Neurologic disease is indicated by diminished touch sensation and proprioception and a diminished cremasteric reflex (retraction of the testis on stroking of the ipsilateral inner thigh).

Laboratory evaluation should include measurement of the serum prolactin concentration, because hyperprolactinemia can be detected in no other way. The serum testosterone concentration should be measured in a man who has decreased libido as well as erectile dysfunction, but it will rarely be helpful in a man who has erectile dysfunction but normal libido.

DIFFERENTIAL DIAGNOSIS

Erectile dysfunction should be differentiated from decreased libido, if possible, because they often have different causes and

different treatments. The two conditions may occur together, however. Decreased libido is decreased sexual interest, of which the principal hormonal cause is hypogonadism. Hypogonadism does not by itself impair erectile ability. A man who has a normal libido but has difficulty getting an erection probably does not have hypogonadism, whereas a man who has decreased libido and potency should be evaluated for hypogonadism as well as erectile dysfunction.

TREATMENT

Treatment of erectile dysfunction depends on the cause. If a medication is the source of the problem, it may be possible to substitute a drug that does not affect erectile function. For example, if the patient is being treated for hypertension with a thiazide diuretic, it might be replaced with an angiotensin-converting enzyme inhibitor or a calcium channel blocker—agents that do not seem to interfere with erectile function.

Underlying disorders should be corrected whenever possible. Hyperprolactinemia can be treated with a dopamine agonist (e.g., cabergoline).

The most effective treatment for erectile dysfunction of psychogenic, vascular, or neurologic origin is sildenafil. This agent inhibits phosphodiesterase, the enzyme that degrades cGMP, and therefore enhances smooth muscle relaxation in the corpora cavernosa, increasing arteriolar blood flow and promoting erection. In one randomized study of 329 men with erectile dysfunction who were randomized to receive either sildenafil or placebo, the men who took sildenafil were eventually able to get an erection on 69% of attempts, compared with 22% of attempts in men who received placebo.[39,40]

Other treatments include intraurethral instillation of alprostadil, intracavernous injection of alprostadil or prostaglandin E_1, and application of a vacuum pump to the penis. These are more cumbersome and therefore less popular than sildenafil, even when they are effective.

The author has received research support from, and has been a consultant to, Solvay Pharmaceuticals, Inc., during the past year.

References

1. Crowley WF Jr, Filicori M, Spratt DI, et al: The physiology of gonadotropin-releasing hormone (GnRH) secretion in men and women. Recent Prog Horm Res 41:473, 1985

2. Bremner WJ, Vitiello V, Prinz PN: Loss of circadian rhythmicity in blood testosterone levels with aging in normal men. J Clin Endocrinol Metab 56:1278, 1983

3. Zhou ZX, Wong CI, Sar M, et al: The androgen receptor: an overview. Recent Prog Horm Res 49:249, 1994

4. Heinlein CA, Chang C: Androgen receptor (AR) coregulators: an overview. Endocr Rev 23:175, 2002

5. Wilson JD, Griffin JE, Leshin M, et al: Role of gonadal hormones in development of the sexual phenotypes. Hum Genet 58:78, 1981

6. Smith EP, Boyd J, Frank GR: Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. N Engl J Med 331:1056, 1994

7. Bilezikian JP, Morishima A, Bell J, et al: Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. N Engl J Med 339:599, 1998

8. Carani C, Rochira V, Faustini-Fustini M, et al: Role of oestrogen in male sexual behaviour: insights from the natural model of aromatase deficiency. Clin Endocrinol (Oxford) 51:517, 1999

9. Hayes FJ, DeCruz S, Seminara SB, et al: Differential regulation of gonadotropin secretion by testosterone in the human male: absence of a negative feedback effect of testosterone on follicle-stimulating hormone secretion. J Clin Endocrinol Metab 86:53, 2001

10. Parker KL, Schedl A, Schimmer BP: Gene interactions in gonadal development. Annu Rev Physiol 61:417, 1999

11. Lee MM, Donahoe PK: Mullerian inhibiting substance: a gonadal hormone with multiple functions. Endocr Rev 14:152, 1993

12. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab 86:724, 2001

13. Morley JE, Kaiser FE, Perry HM 3rd, et al: Longitudinal changes in testosterone, luteinizing hormone, and follicle-stimulating hormone in healthy older men. Metabolism 46:410, 1997

14. Feldman HA, Longcope C, Derby CA, et al: Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. J Clin Endocrinol Metab 87:589, 2002

15. Deslypere JP, Vermeulen A: Leydig cell function in normal men: effect of age, lifestyle, residence, diet, and activity. J Clin Endocrinol Metab 59:955, 1984

16. Purifoy FE, Koopmans LH, Mayes DM: Age differences in serum androgen levels in normal adult males. Hum Biol 53:499, 1981

17. Snyder PJ: Effects of age on testicular function and consequences of testosterone treatment. J Clin Endocrinol Metab 86:2369, 2001

18. Snyder PJ, Peachey H, Hannoush P, et al: Effect of testosterone treatment on bone mineral density in men over 65 years of age. J Clin Endocrinol Metab 84:1966, 1999

19. Snyder PJ, Peachey H, Hannoush P, et al: Effect of testosterone treatment on body composition and muscle strength in men over 65 years of age. J Clin Endocrinol Metab 84:2647, 1999

20. Guzick DS, Overstreet JW, Factor-Litvak P, et al: Sperm morphology, motility, and concentration in fertile and infertile men. N Engl J Med 345:1388, 2001

21. Smyth CM, Bremner WJ: Klinefelter syndrome. Arch Intern Med 158:1309, 1998

22. Lipshultz LI, Caminos-Torres R, Greenspan CS, et al: Testicular function after orchiopexy for unilaterally undescended testis. N Engl J Med 295:15, 1976

23. Stang A, Ahrens W, Bromen K, et al: Undescended testis and the risk of testicular cancer: importance of source and classification of exposure information. Int J Epidemiol 30:1050, 2001

24. De Kretser DM, Baker HW: Infertility in men: recent advances and continuing controversies. J Clin Endocrinol Metab 84:3443, 1999

25. Foresta C, Moro E, Ferlin A: Y chromosome microdeletions and alterations of spermatogenesis. Endocr Rev 22:226, 2001

26. Sellmeye DE, Grunfield C: Endocrine and metabolic disturbances in human immunodeficiency virus infection and acquired immunodeficiency syndrome. Endo Rev 17:518, 1996

27. Seminara SB, Hayes FJ, Crowley WF Jr: Gonadotropin-releasing hormone deficiency in the human (idiopathic hypogonadotropic hypogonadism and Kallmann's syndrome): pathophysiological and genetic considerations. Endocr Rev 19:521, 1998

28. Abs R, Verhelst J, Maeyaert J, et al: Endocrine consequences of long-term intrathecal administration of opioids. J Clin Endocrinol Metab 85:2215, 2000

29. Snyder PJ, Lawrence DA: Treatment of male hypogonadism with testosterone enanthate. J Clin Endocrinol Metab 51:1535, 1980

30. Findlay JC, Place VA, Snyder PJ: Treatment of primary hypogonadism in men by the transdermal administration of testosterone. J Clin Endocrinol Metab 68:369, 1989

31. Swerdloff RS, Wang C, Cunningham G, et al: Long-term pharmacokinetics of transdermal testosterone gel in hypogonadal men. J Clin Endocrinol Metab 85:4500, 2000

32. Snyder PJ, Peachey H, Berlin JA, et al: Effects of testosterone replacement in hypogonadal men. J Clin Endocrinol Metab 85:2670, 2000

33. Finkel DM, Phillips JL, Snyder PJ: Stimulation of spermatogenesis by gonadotropins in men with hypogonadotropic hypogonadism. N Engl J Med 313:651, 1985

34. McPhaul MJ: Molecular defects of the androgen receptor. Recent Prog Horm Res 57:181, 2002

35. Wilson JD, Griffin JE, Russell DW: Steroid 5 alpha-reductase 2 deficiency. Endocr Rev 14:577, 1993

36. Braunstein GD, Glassman HA: Gynecomastia. Curr Ther Endocrinol Metab 6:401, 1997

37. Feldman HA, Goldstein I, Hatzichristou DG, et al: Impotence and its medical and psychosocial correlates: results of the Massachusetts Male Aging Study. J Urol 151:54, 1994

38. Johannes CB, Araujo AB, Feldman HA, et al: Incidence of erectile dysfunction in men 40 to 69 years old: longitudinal results from the Massachusetts male aging study. J Urol 163:460, 2000

39. Lue TF: Erectile dysfunction. N Engl J Med 342:1802, 2000

40. Oral sildenafil in the treatment of erectile dysfunction. Sildenafil Study Group. N Engl J Med 338:1397, 1998

41. Dobs AS, Meikle AW, Arver S, et al: Pharmacokinetics, efficacy, and safety of a permeation-enhanced testosterone transdermal system in comparison with bi-weekly injections of testosterone enanthate for the treatment of hypogonadal men. J Clin Endocrinol Metab 84:3469, 1999

IV THE ADRENAL

D. Lynn Loriaux, m.d., ph.d.

Anatomy

There are two adrenal glands, which are adjacent to the rostral pole of each kidney. The glands weigh 3 to 5 g each and are about 5 cm in largest dimension. They are the shape of a flattened sphere, with one side invaginated. Hence, on tomographic views, the gland has an upside-down Y shape, with a trunk and two limbs [*see Figure 1*]. The thickness of the gland in a given person should be no greater than that of the ipsilateral crus of the diaphragm.

Histologically, the adrenal gland is composed of a cortex and a medulla. The cortex has three zones [*see Figure 2*]. From the capsule inward, these zones are the glomerulosa, the fasciculata, and the reticularis. The glomerulosa is a thin, discontinuous zone in which the cells are arranged in a fashion similar to that of glomeruli. In the process of fixation, lipid is lost from these cells disproportionately; as a result, they are histologically clear. The glomerulosa accounts for about 5% of the cortical volume. The cells of the fasciculata are arranged in linear fashion, vertically, similar to that of the fascicles of pages in the spine of a book. The fasciculata accounts for about 70% of the volume of the adrenal cortex. The reticularis is intensely eosinophilic, with cells arranged in a poorly organized netlike fashion. This zone of the cortex is referred to as the x-zone in fetal life, when it accounts for most of the cortical volume. The reticularis disappears in childhood and then reappears at the time of adrenarche, which usually takes place 2 to 3 years before the onset of puberty. In the mature adrenal cortex, the reticularis accounts for 10% to 20% of the volume of the gland. The innermost zone of the adrenal gland is the medulla. Its structure is analogous to that of a sympathetic ganglion, being composed of chromaffin cells innervated by presynaptic sympathetic axons.

Physiology

The adrenal gland makes three principal hormones: hydrocortisone, or cortisol, which is necessary for life; aldosterone, which promotes salt retention and thereby permits maintenance of salt balance in a salt-poor environment; and adrenaline, or epinephrine, which features prominently in the fight-or-flight response (etymologically, the two terms are identical, with the former deriving from the Latin *ad renal* and the latter from the Greek *epi nephros*). Additionally, the adrenal gland secretes small amounts of estrogen and androgen, as well as two androgen precursor steroids, androstenedione and dehydroepiandrosterone (DHEA). These hormones and prohormones are products of distinct zones of the gland: aldosterone from the glomerulosa, cortisol from the fasciculata, DHEA from the reticularis, and epinephrine from the medulla.

Cortisol levels are regulated by a feedback loop [*see Figure 3*]. The synthesis and secretion of cortisol are stimulated by adrenocorticotropic hormone (ACTH) from the pituitary gland. Once in the bloodstream, cortisol levels rapidly regulate the synthesis and secretion of ACTH by the pituitary. ACTH release is also dependent on corticotropin-releasing hormone (CRH) from the hypothalamus. CRH is also regulated by cortisol, but at a much slower tempo.

Aldosterone levels are also regulated by feedback loops [*see Figure 3*]. Renin, a polypeptide hormone secreted by the juxtaglomerular cells of the kidney, is converted to angiotensin II, notably in the lungs, and stimulates adrenal aldosterone synthesis and secretion. Aldosterone acts on the thick ascending limb of the loop of Henle to enhance salt retention and thus expand vascular volume. The juxtaglomerular cells monitor vascular volume in the afferent artery of the glomerulosa and decrease renin secretion in response to expanded volume. DHEA and epinephrine have no known feedback regulation. All the adrenal hormones can now be measured specifically and accurately by radioimmunoassay, greatly facilitating the study of adrenal physiology and pathophysiology.

Diseases of the adrenal gland can be conveniently categorized as conditions associated with increased or decreased activity of the key hormones: hypercortisolism and hypocortisolism, hyperaldosteronism and hypoaldosteronism, virilization and feminization from sex hormone secretion, and catecholamine excess or deficiency.

Hypercortisolism (Cushing syndrome)

Cortisol is normally secreted at a rate of 6.5 mg/m^2/day.[1] Secretion of cortisol in excess of this rate can, with time, lead to Cushing syndrome. Harvey Cushing described this syndrome in his book *The Pituitary Body and Its Disorders*, published in 1919.[2] The classic clinical presentation of Cushing syndrome includes central obesity, striae, moon facies, supraclavicular fat pads, diabetes mellitus, hypertension, hirsutism and oligomenorrhea in women, and erectile dysfunction in men [*see Table 1*].[3]

The causes of cortisol excess can best be categorized as ACTH dependent or ACTH independent. The two categories can be distinguished by the measurement of ACTH in the blood. ACTH-dependent causes account for about 90% of cases of noniatrogenic Cushing syndrome, and they most often result from

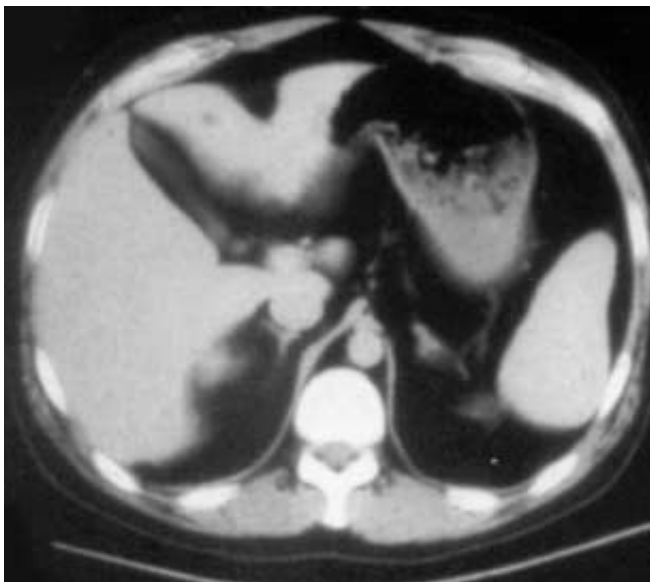


Figure 1 **A computed tomographic image shows a normal adrenal gland.**

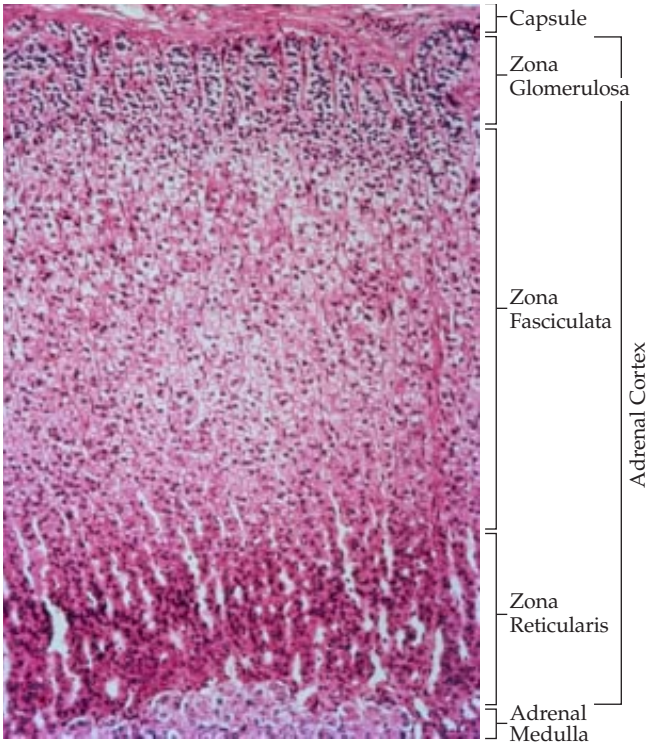


Figure 2 **Hematoxylin and eosin staining distinguishes the zones of the adrenal cortex.**

ACTH secretion by a pituitary microadenoma that is relatively insensitive to the feedback effects of cortisol (i.e., Cushing disease). The most common cause of ACTH-independent Cushing syndrome is iatrogenic long-term glucocorticoid administration. The most common naturally occurring cause of ACTH-independent Cushing syndrome is cortisol secretion from a benign adrenal adenoma [*see Table 2*].

DIAGNOSIS

The diagnosis of Cushing syndrome is principally clinical. The more signs and symptoms of the syndrome that are present, the more confident of the diagnosis the physician can be. The firmer the clinical diagnosis, the less important is biochemical confirmation. Typically, patients will have some, but not all, of the clinical manifestations of Cushing syndrome, and the diagnosis will require confirmation by demonstrating an elevated urinary free cortisol excretion on 24-hour urine testing, which is the single best biochemical marker of Cushing syndrome. The upper limit of the normal range of urinary free cortisol in unstressed persons is generally agreed to be about 100 μg/day as measured in commercial laboratories. If the patient has florid signs of Cushing syndrome, the diagnosis can be made even in the absence of convincing biochemical confirmation. On the other hand, if the patient has few clinical signs of Cushing syndrome, the urinary free cortisol must be greater than 300 μg/day to permit the diagnosis on that basis alone.

Taken together, the number of clinical manifestations and the level of urinary free cortisol define three general categories of Cushing disease: atypical, anorexia-associated, and classic [*see Figure 4*]. Atypical Cushing syndrome is characterized by low levels of urinary free cortisol but many clinical manifestations. Anorexia-associated Cushing syndrome is characterized by high levels of urinary free cortisol and few clinical manifestations. Classic Cushing syndrome is characterized by high levels of urinary free cortisol and many clinical manifestations.

Each of these general categories of Cushing syndrome has an associated differential diagnosis. The most common cause of atypical Cushing syndrome is glucocorticoid use, whether iatrogenic or factitious, and that of anorexia-associated Cushing syndrome is small cell carcinoma of the lung. The most powerful tool in the identification of these less common causes of Cushing syndrome is the clinical history.

In patients with classic Cushing syndrome, a search for the cause should not be undertaken unless the diagnosis is secure. Once that is accomplished, the first step in the differential diagnosis is to determine whether the condition is ACTH dependent or ACTH independent. This is most easily done by measuring the level of circulating plasma ACTH. There is no need to perform dexamethasone suppression testing.[4]

Although an ACTH level greater than 10 pg/ml indicates ACTH dependence, this threshold will fail to identify 5% of ACTH-dependent cases. Consequently, patients with a random plasma ACTH level of less than 10 pg/ml should undergo a CRH challenge. In this test, ACTH levels are measured 10 to 30 minutes after injection of an intravenous bolus of CRH.[5] If the plasma ACTH is less than 10 pg/ml after a CRH challenge and if the urinary free cortisol level is normal or high, the disorder is adrenal in origin; in other words, it is ACTH independent. Either CT or MRI scans will reveal these lesions with a high degree of accuracy. If no tumor is found on adrenal imaging studies, micronodular adrenal dysplasia should be suspected.[6] The familial form of micronodular adrenal dysplasia, Carney syndrome, is an autosomal dominant disorder characterized by pigmented lentigines, blue nevi, and multiple tumors.

ACTH-secreting pituitary microadenomas are often too small to be visible on MRI scans [*see 3:V Pituitary*]. Consequently, patients with ACTH-dependent Cushing syndrome and a normal pituitary MRI should undergo an inferior petrosal sampling procedure to search for a gradient in ACTH levels between blood draining the pituitary gland (inferior petrosal sinus blood) and peripheral antecubital blood. An ACTH gradient greater than 3 between simultaneously sampled central and peripheral blood confirms a pituitary etiology for Cushing syndrome.[7] If the gradient is less than 3, the search for an ectopic source of ACTH should be undertaken. Chest CT and MRI scans (with CT scans typically performed first) are central to this investigation, because 95% of these tumors are intrathoracic. The most common offending tumor is a bronchial carcinoid.

TREATMENT

Except for atypical Cushing syndrome, in which the treatment is to discontinue exogenous glucocorticoid, all treatments for Cushing syndrome are surgical. ACTH-independent cases should be treated with adrenal tumor resection or, in the case of micronodular adrenal dysplasia, bilateral adrenalectomy. ACTH-dependent cases should be treated with transsphenoidal microadenomectomy or ablation of the ectopic ACTH-secreting tumor. When an ectopic source for ACTH cannot be identified, the patient can be treated temporarily with a cortisol synthesis inhibitor, most commonly ketoconazole, until the lesion is found. If it is not found within 18 to 24 months, bilateral adrenalectomy should be performed.

Effective surgery for an ACTH-secreting tumor renders the patient adrenally insufficient. This is documented by measurement of the plasma cortisol level on the morning after surgery. If the

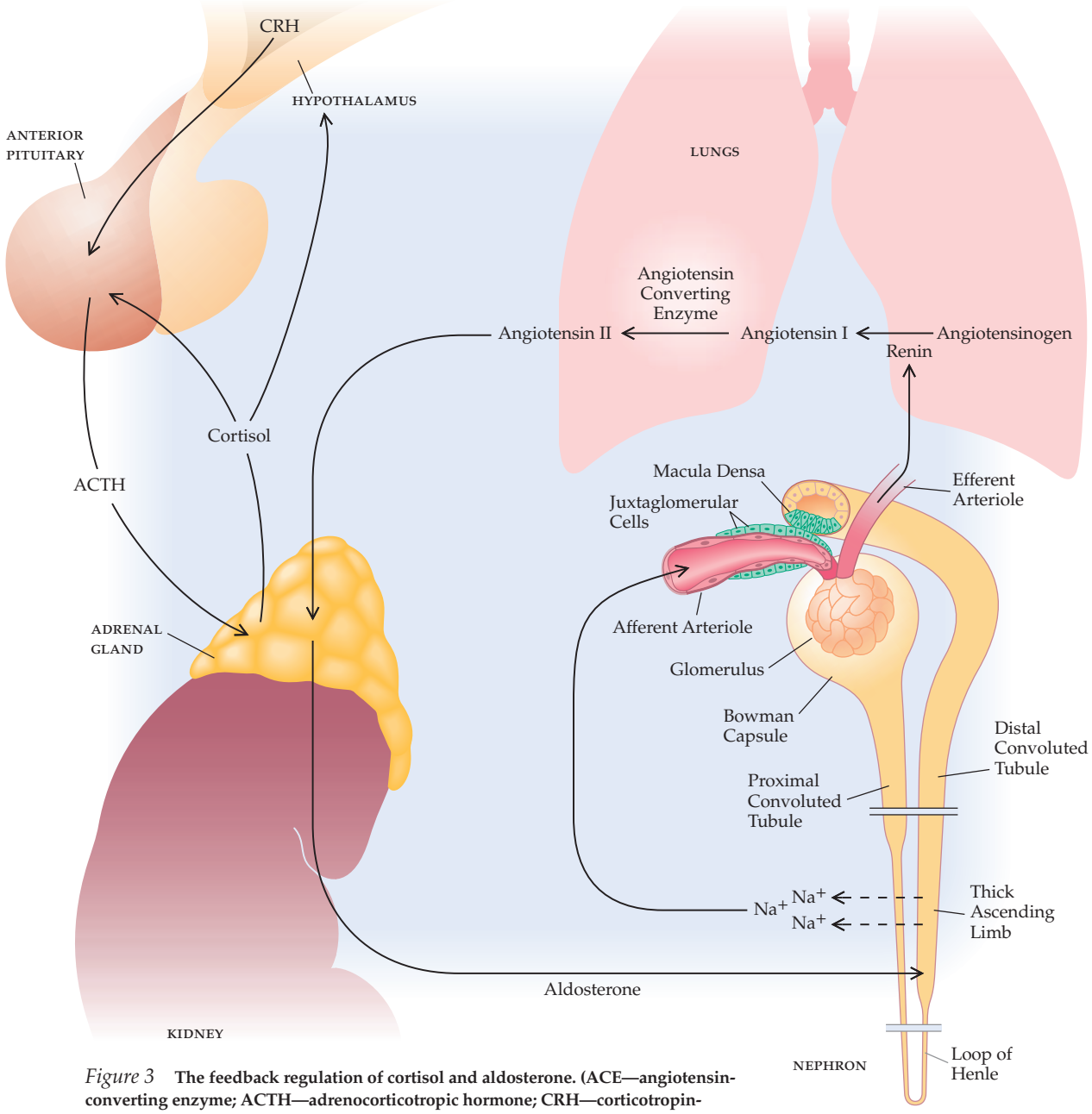


Figure 3 **The feedback regulation of cortisol and aldosterone. (ACE—angiotensin-converting enzyme; ACTH—adrenocorticotropic hormone; CRH—corticotropin-releasing hormone)**

level is greater than 20 μg/dl, the operation has not been successful and no cortisol therapy is indicated. If the plasma cortisol level is below the normal range (i.e., less than 5 μg/dl), the operation has been successful, and steroid replacement therapy will be required until the hypothalamic-pituitary-adrenal (HPA) axis can recover—possibly 1 year or longer. During this period, the patient should be treated with hydrocortisone at a rate of 12 mg/m^2 in a single daily dose with breakfast. Adrenal function is tested with a synthetic ACTH (cosyntropin) stimulation test at 3-month intervals. In patients with cortisol values between 5 and 20 μg/dl after cosyntropin stimulation, cortisol replacement therapy can be withdrawn; but 3-month testing should continue, and cortisol replacement therapy should be reinstituted if signs and symptoms of adrenal insufficiency appear. As soon as the cortisol level reaches 20 μg/dl or greater after cosyntropin stimulation, discontinuance of cortisol replacement without continued testing is safe.

If the operation has not been curative, Cushing syndrome will reappear with time. In addition to cortisol replacement, patients who have undergone bilateral adrenal resection will also require aldosterone replacement in the form of fludrocortisone, 0.1 mg each morning.

Without treatment, Cushing syndrome is fatal. Except for adrenal cancer, however, all causes of Cushing syndrome can be cured. Successful treatment restores normal life expectancy.

Adrenal Cancer

Adrenocortical cancer is a rare disease, with an annual incidence of 1 in 600,000.[8] Thus, about 500 new cases are diagnosed in the United States each year. This disease usually presents clinically as a combination of steroid hormone excess syndromes, the most common being Cushing syndrome with virilization.

Table 1 Sensitivity and Specificity of Selected Findings in Cushing Syndrome

Finding	*Sensitivity (%)*	*Specificity (%)*
Central obesity	90	71
Glucose intolerance	88	23
Plethora	82	69
Proximal muscle weakness	65	93
Striae	40	78
Hyperkalemia	25	96
Osteoporosis	64	97

Differential diagnostic tests will generally show ACTH-independent Cushing syndrome associated with increased serum testosterone levels in hirsute or virilized women. Occasionally, increased serum estradiol levels will be found in feminized men with Cushing syndrome. In rare instances, patients will present with virilization or feminization only, and some will have hypertension caused by tumor secretion of aldosterone precursors. CT or MRI scans usually show a large unilateral adrenal mass. Masses greater than 6 cm have a greater chance of being malignant. These tumors tend to be large when first discovered because the adrenal steroidogenic cells, in the course of dedifferentiation, become less efficient in cortisol synthesis. As a result, the tumor burden has to be large to yield the same clinical effect as a small, highly differentiated, benign adenoma.

The treatment of adrenocortical cancers is surgical. The first operation should focus on complete resection with clear margins. However, the incidence of surgical cure is unknown and is believed to be low, perhaps zero. With recurrence, each subsequent operation should focus on removing all visible disease, including accessible metastases. If left untreated, about half the patients with adrenocortical cancer will die in a few months. With aggressive surgery, their survival can be extended to about 48 months. When surgery is no longer feasible, treatment with steroid hormone synthesis blockers is indicated. Ketoconazole is typically used for this purpose. In addition, ortho,para'-DDD (mitotane) can be considered. At full dosages (> 2 g/day in divided doses), mitotane can produce remission in about 25% of patients. The average remission is 7 months long. Lengthened life span has not been demonstrated, and mitotane has severe side effects—nausea, vomiting, lethargy, and vertigo—so some physicians argue that the cost-to-benefit ratio (quality of life versus days of life gained) does not justify the use of this drug. Other chemotherapeutic regimens are under development, and patients with this rare disease are best served by referral to a center engaged in this research.[9]

Table 2 Differential Diagnosis of Cushing Syndrome

ACTH-dependent causes	Pituitary ACTH-producing tumor Ectopic ACTH-producing tumor
ACTH-independent causes	Glucocorticoid use (factitious or iatrogenic) Adrenal adenoma Adrenal carcinoma Micronodular adrenal disease

Incidental Adrenal Masses

About 300,000 abdominal CT and MRI procedures are done annually in the United States for indications unrelated to the adrenals. An incidental adrenal mass (incidentaloma) is found in about 4% of these procedures, or 12,000 newly discovered masses each year.[10] Assuming that this is the incidence in the general population, there would be roughly 12 million such masses in the United States population. Assuming that adrenal cancers are detectable by MRI 6 years before they become clinically evident, the prevalence of adrenal cancer would be about 1 in 3,000 population. Therefore, 1 in 4,000 incidentally discovered adrenal masses will be an adrenal cancer.

Measurement of serum potassium and bicarbonate levels is appropriate in patients with an incidentaloma. Plasma free metanephrine measurement may disclose a pheochromocytoma. Dexamethasone suppression testing has been advocated,[10] but its predictive value is only 0.5—no better than tossing a coin.

The central question in the management of the incidental adrenal mass is whether surgical removal is indicated. Certainly, the tumor should be removed if it is metabolically functional, as manifested by Cushing syndrome, the syndrome of mineralocorticoid excess, or virilization or feminization for which there is no other explanation. If there is no evidence of functionality, the tumor should be studied with a CT contrast washout study. Benign adrenal adenomas are lipid rich, whereas malignant ones tend to contain much more cellular and intercellular water. Thus, water-soluble contrast agents tend to wash out of benign lesions much faster than they do from malignant ones. The accuracy of this procedure in experienced hands is very high, with greater than 90% specificity and sensitivity.[11,12]

A mass that displays slow washout can be assumed to be a metastasis. If the patient has a known malignancy, the adrenal mass can be treated as part of the primary process. If there is no known primary malignancy, the mass could be the first manifestation of metastasis or a rare nascent adrenocortical carcinoma. Percutaneous needle biopsy can readily differentiate between these two possibilities, but it does involve risks, such as pneumothorax and tumor seeding. Alternatively, the mass can be removed laparoscopically and a pathologic analysis done.

In most series, 5% to 6% of incidental adrenal masses are pheochromocytomas.[13] This possibility should be excluded before biopsy or operation so as to avoid the hypertensive crises that can be associated with surgery. The safest course is to assume that the lesion is a pheochromocytoma and to prepare all such patients for surgery with adequate alpha blockade [*see* Pheochromocytoma, *below*].

Adrenal Insufficiency

Adrenal insufficiency (Addison disease) is categorized as primary or secondary. Primary adrenal insufficiency results from destruction of the adrenal cortex. There is a long list of causes of primary adrenal insufficiency [*see Table 3*]; worldwide, tuberculosis is the most common cause, and in the industrialized nations, idiopathic or autoimmune adrenal destruction is the most common cause. Secondary adrenal insufficiency results from disruption of

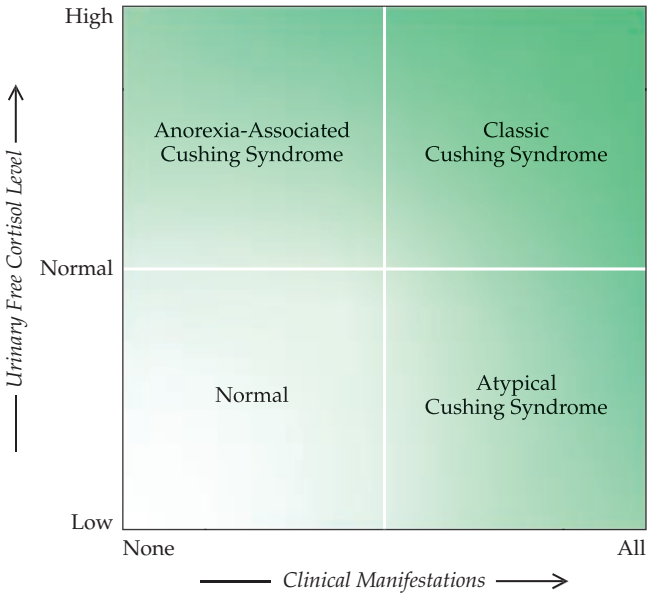


Figure 4 **The severity of the clinical manifestations and the level of urinary free cortisol can be used to define three categories of Cushing syndrome: classic, atypical, and anorexia associated.**

pituitary secretion of ACTH, which by far is most commonly caused by prolonged treatment with exogenous glucocorticoids. With time, doses of exogenous glucocorticoids sufficient to suppress ACTH secretion will lead to dysfunction of CRH-secreting neurons and attendant ACTH deficiency; subsequent withdrawal of glucocorticoids, for whatever reason, will then unmask the deficiency. Recovery of function may require a year or more. A far less common cause of disrupted ACTH secretion is destructive lesions in and around the pituitary gland and hypothalamus [*see Table 4*].

The symptoms and signs of adrenal insufficiency can be grouped into chronic and acute syndromes. The chronic syndrome is characterized by anorexia, weight loss, fatigue, and orthostatic hypotension. In patients with primary disease, the predominant signs are weight loss and hyperpigmentation of the skin, especially of the sun-exposed areas and extensor surfaces. The acute syndrome is closely analogous to cardiogenic or septic shock, with reduced cardiac output into a dilated and unresponsive vascular system. Symptoms include prostration and all of the signs and symptoms of the shock syndrome. Shock in this setting tends to be unresponsive to volume replacement and vasoconstrictor therapy.

Table 3 Causes of Primary Adrenal Insufficiency

Autoimmunity (70% of cases)
Polyendocrine deficiency syndrome
Tuberculosis (20% of cases)
Other (10% of cases)
Fungal infection
Adrenal hemorrhage
Adrenomyeloneuropathy
Adrenoleukodystrophy
Sarcoidosis
Amyloidosis
Congenital adrenal hyperplasia
Congenital unresponsiveness to ACTH
Metastatic cancer
AIDS

Table 4 Causes of Secondary Adrenal Insufficiency

Iatrogenic suppression of the hypothalamic-pituitary-adrenal axis (90% of cases)
Other (10% of cases)
Hypophysectomy
Pituitary irradiation
Head trauma
Hypophysitis
Hemochromatosis
Infection
Actinomycosis
Nocardiosis
Intracranial tumor
Pituitary tumor

With both chronic and acute syndromes, the diagnosis should be suspected on clinical grounds, but it requires laboratory confirmation. The critical test for the diagnosis of chronic adrenal insufficiency is the cosyntropin stimulation test. Synthetic ACTH (cosyntropin) is administered in a 250 μg intravenous bolus, and plasma cortisol levels are then measured after 45 and 60 minutes. Values greater than 20 μg/dl exclude adrenal insufficiency as a cause of the clinical findings. Values less than 20 μg/dl suggest that adrenal compromise could be a contributing factor. In this situation, treatment with glucocorticoids is mandatory until the clinical situation is clarified with more precision.

In acute adrenal insufficiency, the most useful test is measurement of the plasma cortisol level. Cosyntropin stimulation testing is not necessary; the illness, which is sufficiently severe to merit admission to an intensive care unit, represents an endogenous source of maximal physiologic stress. Plasma cortisol levels in acute adrenal insufficiency are greater than 20 μg/dl, with the only exception being in patients who have a low plasma albumin concentration, which lowers the total cortisol concentration.[14] Unfortunately, there are no published data on the interpretation of plasma cortisol values in patients with low albumin concentrations, so most clinicians adhere to the 20 μg/dl standard. Currently, if the cortisol value is less than 20 μg/dl, it should be confirmed with a standard cosyntropin stimulation test.

The differential diagnosis of adrenal insufficiency requires the discrimination of primary and secondary causes; the most useful test is measurement of the circulating plasma ACTH level. ACTH levels greater than normal define primary disease; values in the normal range or below define secondary disease.

Patients with primary adrenal disease should have the adrenal glands imaged with CT or MRI. Infectious, malignant, and vascular causes of adrenal insufficiency all result in enlargement of the adrenal glands. In idiopathic or autoimmune adrenal insufficiency, the glands are normal or small in size. Patients with secondary adrenal insufficiency should first be assessed for exogenous glucocorticoid use. If that can be eliminated as a cause, they should undergo CT or MRI scanning of the hypothalamus and pituitary gland to exclude destructive lesions in this area.

TREATMENT

The goal in treating adrenal insufficiency is to replace the

Table 5 Differential Diagnosis of Pheochromocytoma

Panic attacks
Thyrotoxicosis
Amphetamine use
Cocaine use
Over-the-counter cold medicines containing phenylephrine or pseudoephedrine
Monoamine oxidase inhibitors
Hypoglycemia
Insulin reaction
Brain tumor
Subarachnoid hemorrhage
Menopausal hot flashes
Toxemia of pregnancy
Selective serotonin reuptake inhibitors

missing hydrocortisone and aldosterone in quantities calibrated to the clinical situation. Hydrocortisone can be replaced with oral or intravenous hydrocortisone. Aldosterone is replaced with oral fludrocortisone. Exogenous hydrocortisone and fludrocortisone are both equipotent with the endogenously secreted hormone. Unstressed persons secrete hydrocortisone at a rate of 6.5 mg/m^2 daily. In the face of stress, such as a surgical procedure or serious trauma, hydrocortisone secretion can rise more than 10-fold. The secretion rate of aldosterone is 100 μg/day in persons consuming large amounts of sodium (i.e., a typical United States diet).

Primary chronic adrenal insufficiency is treated with oral hydrocortisone, 12 to 15 mg/m^2/day. This is roughly double the amount of hydrocortisone that is normally secreted; the added amount is needed to compensate for first-pass hepatic metabolism. Hydrocortisone is best given as a single daily dose with breakfast. Fludrocortisone is given at a dose of 0.1 mg/day. When moderate stress is anticipated (e.g., a root canal procedure), the dose of hydrocortisone is temporarily doubled, beginning the day before the stress and continuing until 2 days afterward. It is not necessary to alter the fludrocortisone dose. With anticipated major stress (e.g., appendectomy with general anesthesia), the hydrocortisone dosage is increased to 100 mg every 6 hours from the day before the procedure until 2 days afterward. Hydrocortisone dosage increases are not required for periods of psychological stress, such as major depression, psychosis, or grief.

These replacement regimens roughly reproduce the patterns of cortisol and aldosterone secretion in persons with normal adrenal function. The need for these temporary dosage increases has not been clearly established, on either clinical or biologic grounds, but this has become the standard of practice and is not likely to change. Chronic secondary adrenal insufficiency is treated in the same way as chronic primary disease but with replacement of hydrocortisone only, not aldosterone.

Patients with acute adrenal insufficiency are treated in the same fashion as those with chronic adrenal insufficiency who are experiencing major stress. Treatment is monitored clinically. Signs of Cushing syndrome indicate overtreatment; hyponatremia, orthostasis, and anorexia indicate undertreatment. There is no good clinical evidence to suggest that the dosage regimens ever need to be exceeded. If a patient on recommended replacement doses of hydrocortisone and fludrocortisone fails to do as well as expected, the reason is something other than the adrenal replacement regimen.

All patients with adrenal insufficiency should wear a medical-alert bracelet imprinted with the words "adrenal insufficiency" and carry a similar wallet card at all times.

Pheochromocytoma

The adrenal medulla accounts for about 10% of the weight of the adrenal gland. It is composed primarily of chromaffin cells, which are named for the yellow-brown color they take on when stained with chromatic salts. The cells of the medulla are directly innervated by preganglionic sympathetic nerve cells. Hence, these epinephrine-secreting cells are analogous to the postganglionic neurons in the other areas of the sympathetic nervous system. These cells are not neurons, however, and have no dendrites or axons. In addition, the primary secretory product of the adrenal medulla is epinephrine, whereas the remainder of the sympathetic nervous system employs norepinephrine as the neurotransmitter. The reason for this difference is that the blood supply to the adrenal medulla is derived from the capillary plexus draining the adrenal cortex. This capillary blood is extremely rich in cortisol—perhaps the highest concentration of cortisol in the human body is in the adrenal medulla—and cortisol induces catechol-O-methyl transferase, the enzyme that converts norepinephrine to epinephrine. The primary disease of the adrenal medulla is pheochromocytoma; 90% of pheochromocytomas occur in the adrenal medulla. Extra-adrenal tumors of the chromaffin cell are known as paraganglions or chemodectomas, depending on the location. All have similar clinical presentations, and all are treated in the same way (see below).

The main clinical manifestation of pheochromocytomas is hypertension. The hypertension can be sustained or episodic; the two forms occur with equal frequency. Paroxysmal hypertension is associated with tachycardia, diaphoresis, anxiety, and a sense of foreboding. Patients also complain of nausea and abdominal pain. The association of headache, palpitations, and sweating with hypertension has a high (> 90%) sensitivity and specificity for pheochromocytoma. The differential diagnosis for pheochromocytoma is extensive and includes anxiety and panic attacks, thyrotoxicosis, amphetamine and cocaine use, and use of over-the-counter cold medicines that depend upon catecholamines for effect, such as atomizers for nasal congestion [*see Table 5*]. Pheochromocytomas are usually benign (90%) and usually unilateral (90%). The incidence of pheochromocytoma is markedly increased in several genetic syndromes: multiple endocrine neoplasia types 2a and 2b and the phakomatoses, including neurofibromatosis, cerebelloretinal hemangioblastosis, tuberous sclerosis, and Sturge-Weber syndrome.

DIAGNOSIS

The traditional tests for diagnosing pheochromocytoma are measurements of the urinary fractionated catecholamines and urinary metanephrine excretion in 24-hour urine samples. Total catecholamine excretion is normally less than 100 μg/day, with no more than 25% being epinephrine. Urinary metanephrine excretion is normally less than 1.3 mg/day. The urine for these tests must be collected in an acid medium (laboratories typically provide appropriate containers) and need not be refrigerated. Creatinine should also be measured, as an indicator of completeness of collection. The patient should be taken off all medications when possible. If the hypertension must be treated, diuretics, vasodila-

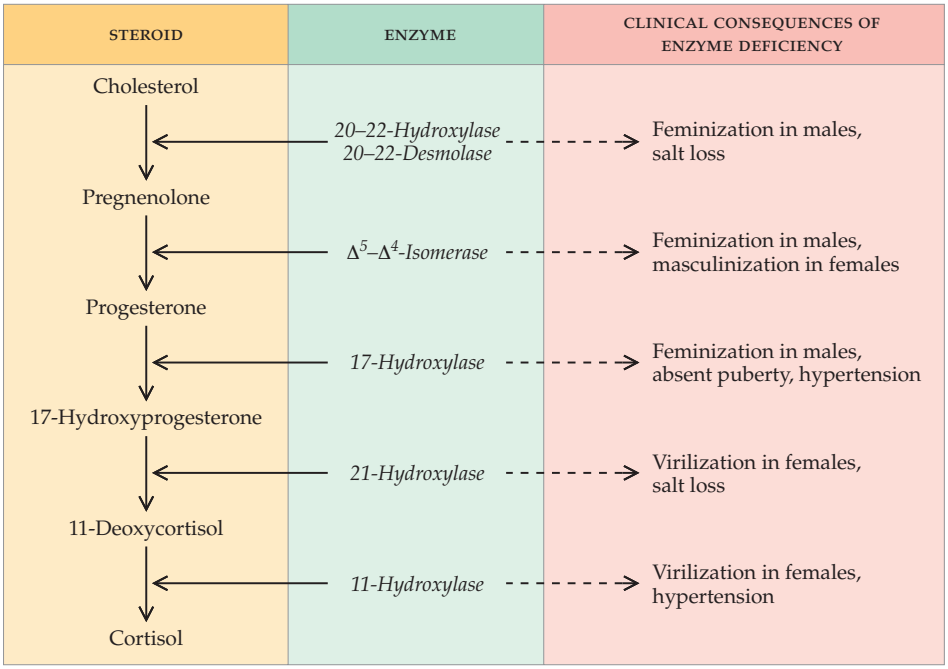


Figure 5 **Congenital adrenal hyperplasia may result from mutations that inactivate any of the six enzymatic steps in the biosynthesis of cortisol from cholesterol. The clinical manifestations of the disorder vary with the enzyme deficiency.**

tors, calcium channel blockers, and angiotensin-converting enzyme (ACE) inhibitors interfere minimally with the assays. When there is concordance between the clinical picture and the biochemical tests, CT or MRI scans should be employed to localize the tumor. MRI is particularly useful because these tumors almost always "brighten" with T_2-weighted images. If CT and MRI fail to reveal an adrenal tumor, radiolabeled meta-iodobenzylguanidine (MIBG) can be a useful scanning technique for locating tumors outside of the adrenal gland, such as those in the carotid body, heart, urinary bladder, and the organ of Zuckerkandl.

TREATMENT

The treatment of pheochromocytoma is surgical. The surgery should be undertaken only by a team experienced and skilled in the management of pheochromocytoma. Before the surgical procedure, complete alpha blockade should be induced to avoid intraoperative hypertensive crisis. Preparation should begin 7 days before the planned procedure, using phenoxybenzamine at an initial dosage of 10 mg by mouth twice daily. The dose should be increased daily, and by the seventh day, the patient should be taking at least 1 mg/kg/day in three divided doses. Adequate blockade is associated with reduced blood pressure and reduced orthostatic hypotension as the vascular volume is restored.

Malignant pheochromocytoma should be treated with surgical debulking, ongoing alpha blockade with phenoxybenzamine, and comanagement with an oncologist. Radiation therapy is useful for bone pain, and some success has been achieved with combination chemotherapy, including cyclophosphamide, vincristine, and dacarbazine.

Congenital Adrenal Hyperplasia

There are six enzymatic steps in the biosynthesis of cortisol from cholesterol, and all can be affected by inactivating mutations [*see Figure 5*]. Because cortisol is essential for life, cortisol concentrations are maintained in the normal range at the expense of adrenal hypertrophy and increased adrenal secretion of the steroid biosynthetic intermediate in the step immediately before the affected enzyme. Depending on which enzyme is blocked, the increased concentrations of the steroid biosynthetic intermediate can lead to virilization in females and to hypertension. In some cases, primarily because of reduced androgen secretion in utero—a time when there is no feedback regulation of testosterone—male fetuses can be feminized.

The most common underlying disorder in congenital adrenal hyperplasia is 21-hydroxylase deficiency. The virilizing form of this disease is thought to be the most common autosomal recessive disorder.

21-Hydroxylase deficiency is categorized according to two clinical distinctions: (1) the classic form, which can be salt losing or non–salt losing, and (2) the nonclassic form.

The degree to which a person with 21-hydroxylase deficiency loses salt in a salt-poor environment correlates with the degree of expression of the enzyme defect in the zona glomerulosa. In persons with mild expression, salt loss is sufficiently minimal that a standard United States diet will maintain a normal salt balance.

The classic form of the disease is usually diagnosed in the neonatal period and is characterized by failure to thrive as a result of the salt loss and by male pseudohermaphroditism in female infants. The nonclassic form of the disease, which is sometimes referred to as adult onset or attenuated, usually becomes clinically apparent in adolescence. It is manifested by a slightly earlier age at puberty (approximately 1 year) and, in females, oligomenorrhea and androgen-mediated hirsutism. Adults who present with the classic form of 21-hydroxylase deficiency usually have a well-documented diagnosis since infancy, have a gender assignment, and have completed a series of genital reconstructive plastic surgical procedures. The usual clinical questions are whether ongoing treatment is necessary and, if so, whether the current regimen is appropriate.

The typical adult patient with the nonclassic or attenuated form is a young woman with oligomenorrhea, infertility, and hirsutism. The most common confounding diagnosis is the polycystic ovary syndrome [*see 16:V Polycystic Ovary Syndrome*].

The diagnostic test for 21-hydroxylase deficiency is a cosyntropin stimulation test: synthetic ACTH is administered in a 250 μg intravenous bolus, and plasma levels of 17-hydroxyprogesterone are measured after 45 and 60 minutes. 17-Hydroxyprogesterone is the steroid biosynthetic intermediate immediately proximal to the enzyme defect. In normal patients, 17-hydroxyprogesterone levels will rise to no higher than 340 ng/dl after cosyntropin stimulation; in patients with 21-hydroxylase deficiency, 17-hydroxyprogesterone levels will be no lower than 1,000 ng/dl. CT or MRI scanning in these patients will show that the adrenal glands are larger than normal and, in some cases, nodular.

TREATMENT

All patients with 21-hydroxylase deficiency should be considered to have some degree of salt loss. Fludrocortisone, 0.2 mg every morning, should be the first therapy. Hydrocortisone, 12 to 15 mg/m^2 as a single morning dose, should be initiated several days later. After 2 weeks of combined therapy, a morning 17-hydroxyprogesterone level should be measured. If the target level of 400 to 600 ng/dl is achieved, the fludrocortisone dose can be reduced by half. Two weeks later, the 17-hydroxyprogesterone should be measured again; if it is still below 600 ng/dl, that establishes the fludrocortisone dose as the patient's maintenance dose. If the 17-hydroxyprogesterone level has risen above 600 ng/dl, the fludrocortisone dose should be restored to the initial 0.2 mg/day, which likely will be the maintenance dose. Reduction of 17-hydroxyprogesterone levels to within the normal range is not recommended. Achieving this level often requires doses of fludrocortisone that produce adrenal suppression and lead to Cushing syndrome.

Lifelong treatment is required in patients with 21-hydroxylase deficiency to prevent the appearance of adrenal rest tumors, which are nodules of ectopic adrenal tissue that become hypertrophic because of ongoing ACTH stimulation. These tumors are usually found in the broad ligament in women and in the testes in men. In women, hemorrhage or necrosis of adrenal rest tumors occasionally necessitates emergency pelvic surgery; in men, these tumors can result in testicular pain, testicular masses, and infertility. Testicular pain may be so severe and intractable that castration is required.

Hyperaldosteronism

Hyperaldosteronism can be primary or secondary. In primary hyperaldosteronism, there is disordered function of the renin-aldosterone feedback axis; in secondary hyperaldosteronism, the renin-aldosterone axis is responding normally to chronic intravascular volume deficiency, which may result from such conditions as heart failure or ascites associated with cirrhosis of the liver.

Aldosterone acts on the epithelial cells of the renal collecting tubule to promote reabsorption of sodium and excretion of potassium and hydrogen. Other tissues similarly affected include sweat glands, salivary glands, and intestinal epithelium. Clinically, the result of excess aldosterone is the so-called mineralocorticoid excess syndrome, characterized by hypokalemia, metabolic alkalosis, and, sometimes, hypertension.

PRIMARY HYPERALDOSTERONISM

Primary hyperaldosteronism is caused by benign adrenal adenomas, which are typically unilateral, are usually less than 2.5 cm in diameter, and secrete aldosterone independently of renin-angiotensin stimulation. Patients with primary hyperaldosteronism present with hypertension; in fact, primary adrenal hypersecretion of aldosterone is thought to account for about 2% of cases of hypertension. Laboratory testing shows hypokalemia and metabolic alkalosis, with a serum sodium level that is usually in the high-normal range [*see Figure 6*]. Diagnosis of this disorder is confirmed by demonstrating normal or elevated plasma aldosterone levels (> 14 ng/dl) along with suppression of stimulated plasma renin activity (PRA) to less than 2 ng/ml/hr. Stimulated PRA is determined by measuring the plasma renin activity level after 2 hours of upright posture (standing or walking).

The differential diagnosis of primary hyperaldosteronism also includes dexamethasone-suppressible hyperaldosteronism, in which aldosterone is secreted in response to ACTH rather than angiotensin [*see* Dexamethasone-Suppressible Hyperaldosteronism, *below*], and idiopathic bilateral adrenal hyperplasia, in which the hypertrophic zona glomerulosa secretes aldosterone independent of renin-angiotensin stimulation [*see* Idiopathic Bilateral Adrenal Hyperplasia, *below*]. Dexamethasone-suppressible hyperaldosteronism is confirmed by the suppression of aldosterone levels with dexamethasone administration, 2 mg/day in divided doses for 7 days. In most cases, aldosterone levels decrease by the third day of treatment. If dexamethasone fails to suppress plasma aldosterone levels and to ameliorate the associated hypertension, CT or MRI should be employed to search for an adrenal adenoma. If an adenoma is not found by CT or MRI, simultaneous adrenal venous sampling for the measurement of aldosterone and cortisol will be needed to define the source of aldosterone secretion.[15] If the venous sampling identifies unilateral aldosterone secretion, the patient should be treated as if primary adrenal hypoaldosteronism is present, despite the absence of a visible adenoma. The surgeon, at the time of operation, can define unilateral versus bilateral disease.

The treatment of primary adrenal hyperaldosteronism is unilateral adrenalectomy, preferably by a laparoscopic procedure. The cure rate, defined as correction of hyperaldosteronism and hypertension, is about 75%.[16] Patients whose blood pressure remains elevated postoperatively will require ongoing antihypertensive therapy, which is managed as if essential hypertension were present.

Idiopathic Bilateral Adrenal Hyperplasia

The clinical presentation of idiopathic bilateral adrenal hyperplasia is indistinguishable from that of primary hyperaldosteronism caused by an adrenal adenoma. However, patients with idiopathic bilateral adrenal hyperplasia have no dominant adrenal adenoma, and aldosterone secretion from both adrenal glands can be documented by bilateral adrenal venous sampling. Adrenalectomy in these patients does not correct the hypertension. Thus, treatment is directed at the hypertension. Interestingly, antagonizing aldosterone activity with spironolactone is usually ineffective. Calcium channel blockers, however, are effective antihypertensive agents in these patients, as are ACE inhibitors. If hypokalemia persists during the treatment of hypertension, it can usually be managed by the addition of a potassium-sparing diuretic.

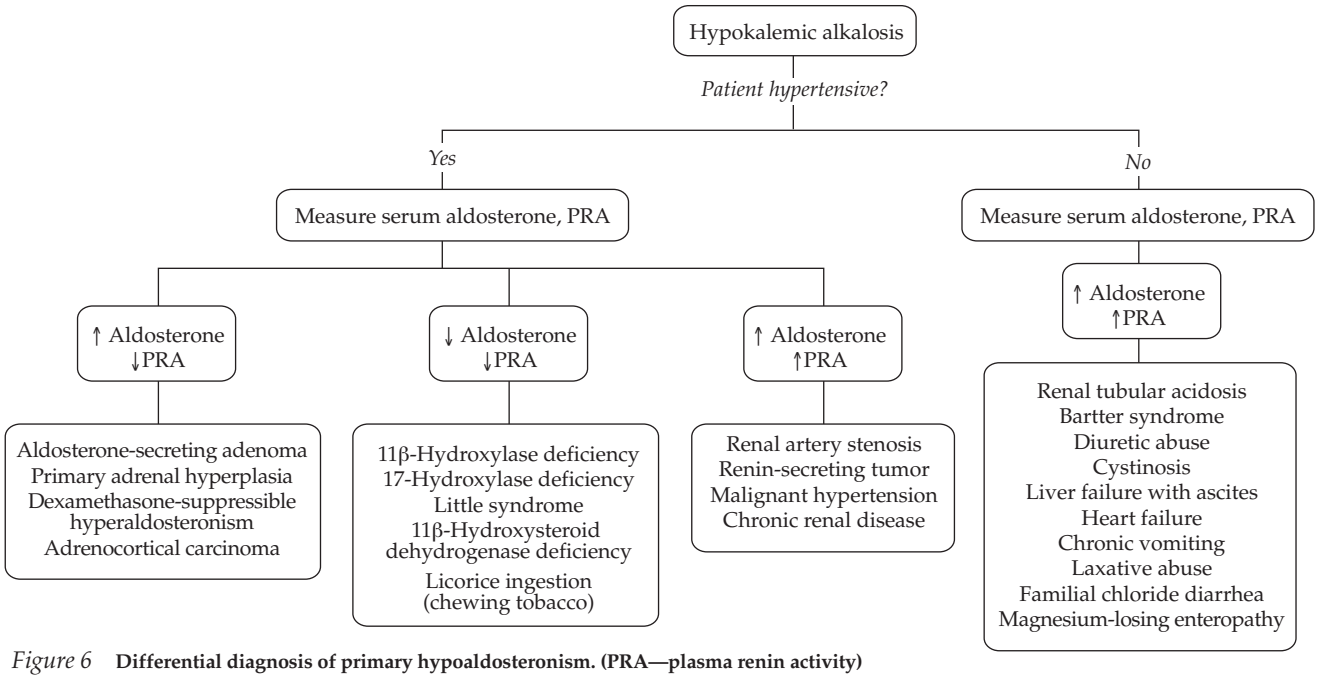


Figure 6 **Differential diagnosis of primary hypoaldosteronism. (PRA—plasma renin activity)**

Dexamethasone-Suppressible Hyperaldosteronism

Dexamethasone-suppressible hyperaldosteronism is a rare familial cause of hyperaldosteronism and is transmitted as an autosomal dominant trait. The cause of the disorder is a fusion gene in which the coding region for ACTH-responsive regulation of 11-β hydrolase is coupled with the coding region for aldosterone synthase. Thus, aldosterone secretion becomes entrained to ACTH secretion and is "blind" to renin-angiotensin levels. Because ACTH secretion is not modulated by aldosterone, aldosterone secretion becomes independent of salt balance, blood potassium levels, and vascular volume.

Treatment for this disorder starts with the use of a potassium-sparing diuretic such as amiloride or triamterene. This regimen has the advantage of not suppressing the HPA axis. If it is unsuccessful, ACTH secretion can be suppressed with dexamethasone, usually 0.5 mg in a single daily dose.

SECONDARY HYPERALDOSTERONISM

Secondary hyperaldosteronism may or may not be associated with hypertension. Patients with hypertension usually have underlying renal pathology, including renal artery stenosis, renin-secreting tumors, and chronic renal failure. Both plasma renin activity and aldosterone are elevated in such cases. Treatment should be directed at the underlying cause.

Secondary hyperaldosteronism that is not associated with hypertension occurs in disorders characterized by decreased vascular volume. Renal causes include chronic nephritis, renal tubular acidosis, and calcium- and magnesium-losing nephropathies. Chronic diuretic abuse also is a cause. Gastrointestinal causes include chronic vomiting, laxative abuse, and chronic diarrhea of any kind. Probably the most common causes are chronic heart failure and cirrhosis of the liver with ascites. Again, treatment is best directed at the underlying disorder.

Finally, there are two forms of congenital adrenal hyperplasia in which overproduction of mineralocorticoids other than aldosterone leads to the syndrome of mineralocorticoid excess. These two disorders are 11-hydroxylase deficiency and 17-hydroxylase deficiency. Both renin and aldosterone levels are low in these disorders. Treatment is the same as that for 21-hydroxylase deficiency (see above), but without fludrocortisone.

Bartter syndrome is associated with hypokalemic alkalosis, hyperreninemia, and hyperaldosteronism, with normal blood pressure. This pattern can be seen in a number of disorders causing secondary hyperaldosteronism. Bartter syndrome is caused by a deficit in chloride transport in the thick ascending limb of the loop of Henle. Diagnosis is difficult because the pattern of electrolyte abnormalities mimics that seen in diuretic abuse. A more detailed discussion of Bartter syndrome is provided elsewhere [*see 10:II Disorders of Acid-Base and Potassium Balance*].

Hypoaldosteronism

PRIMARY HYPOALDOSTERONISM

Primary hypoaldosteronism is defined as aldosterone deficiency of adrenal cause. Hypoaldosteronism manifests as an inability to conserve sodium, leading to a negative salt balance in a salt-poor environment. This leads to hypotension, hyperkalemia, dehydration, and volume depletion associated with a mild metabolic acidosis. The disorder can be corrected by a high-salt diet or by replacement of aldosterone with fludrocortisone.

Primary adrenal insufficiency is the most common cause of primary hypoaldosteronism. Diagnosis and treatment are the same as those for adrenal insufficiency (see above). Two rare autosomal recessive disorders, corticosterone methyl oxidase (CMO) deficiency types I and II, can result in markedly reduced adrenal secretion of aldosterone. CMO deficiency type I is recognized by the syndrome of mineralocorticoid deficiency and low aldosterone levels associated with high plasma corticosterone concentration. CMO deficiency type II is similar, except that high levels of 18-hydroxycorticosterone will be associated with low levels of aldosterone. These are primarily diseases of childhood, becoming less severe with age and free access to salt.

SECONDARY HYPOALDOSTERONISM

The syndrome of hyporeninemic hypoaldosteronism is the most common form of secondary hypoaldosteronism. The disorder is often referred to as renal tubular acidosis type 4. It has been described in almost every disorder of renal function. Chronic renal disease is present in 80% of patients with the disorder. The clinical picture is that of hyperkalemia, hyponatremia, and metabolic acidosis in association with a low plasma renin activity and a low plasma aldosterone level. The most direct and rational therapy for this syndrome is replacement of aldosterone with fludrocortisone at a dosage of 0.1 to 0.2 mg/day.

PSEUDOHYPOALDOSTERONISM (MINERALOCORTICOID RESISTANCE)

Pseudohypoaldosteronism type I and type II are syndromes of end-organ resistance to the effects of aldosterone. Type I is caused by an inactivating mutation in the mineralocorticoid receptor, and type 2 is ascribed to an ill-defined defect in aldosterone action distal to its binding to the mineralocorticoid receptor. Pseudohypoaldosteronism type 1 is characterized by salt wasting that is resistant to mineralocorticoid replacement. It is best treated with a high-salt diet, 10 to 40 mEq/kg/day. Pseudohypoaldosteronism type II (Gordon syndrome) is a non–salt-wasting disorder that can be associated with hypertension, metabolic acidosis, and hyperkalemia. Plasma renin activity and aldosterone are both low, and administration of mineralocorticoid fails to correct the hyperkalemia and acidosis. The basic defect is thought to be a chloride shunt disorder in the nephron. Treatment is with a potassium-wasting diuretic; hydrochlorothiazide and furosemide are most often used.

Glucocorticoid Therapy

Glucocorticoids can be valuable, even lifesaving, in the treatment of many inflammatory and neoplastic diseases. Although cortisol accounts for about half of the mineralocorticoid effect produced by the adrenal gland, the synthetic steroids that are customarily used for glucocorticoid therapy (e.g., prednisone and dexamethasone) have virtually no salt-retaining activity and, therefore, do not cause unacceptable salt retention. On the other hand, their glucocorticoid effect is far more powerful than that of cortisol. Gram for gram, prednisone has four times the glucocorticoid potency of cortisol; dexamethasone has about 25 times the potency.

The target tissues in glucocorticoid-responsive diseases are glucocorticoid resistant. The basis for this resistance remains unknown, but the prevailing hypothesis is that the chaperone proteins produced in stressed cells, particularly the heat shock proteins, in some way attenuate glucocorticoid action. Overcoming glucocorticoid resistance may require dosages of prednisone as high as 100 mg/day and dosages of dexamethasone as high as 20 mg/day. These high doses expose the rest of the tissues in the patient's body, which have normal responsiveness to glucocorticoid, to an extremely enhanced glucocorticoid effect. Over time, this leads to Cushing syndrome, whose potentially lethal effects may force the tapering or even discontinuance of glucocorticoid therapy.

An invariable aspect of Cushing syndrome induced by exogenous glucocorticoid is suppression of ACTH secretion. In contrast to the recovery of pituitary secretion of other hormones, such as thyroid-stimulating hormone or luteinizing hormone and follicle-stimulating hormone, recovery of ACTH secretion is very slow; the return to normal may require a year or more. Thus, the physician must ensure that the HPA axis is intact before completely withdrawing long-term glucocorticoids.

Pharmacologic glucocorticoid therapy is typically initiated at a high dose (e.g., prednisone, 60 mg daily in divided doses). As soon as the disease process is controlled, the dose is reduced in 5% increments weekly in an attempt to find the lowest effective dose as quickly as possible. The ultimate goal is to taper to normal replacement doses of the glucocorticoid. When the glucocorticoid dose approximates the replacement level, the preparation is changed to an equivalent dose of hydrocortisone given at a dosage of 12 mg/m^2 once a day in the morning. This dose remains unchanged until it is safe to withdraw glucocorticoid therapy completely or until the disease reactivates, in which case the process is begun anew. Patients receiving hydrocortisone at the replacement dose should undergo cosyntropin stimulation testing every 3 months. When the plasma cortisol response to cosyntropin exceeds 20 μg/dl, hydrocortisone can be discontinued safely. In the event that the dose cannot be lowered to replacement levels because of recurrent disease activity, alternative and adjunctive non–glucocorticoid-based therapies must be aggressively pursued in the hope that they might permit tapering of the glucocorticoid to replacement dose before the ravages of Cushing syndrome demand cessation of glucocorticoid treatment in the setting of an uncontrolled inflammatory or neoplastic illness.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Esteban NV, Loughlin T, Yergey AL, et al: Daily cortisol production rate in men determined by stable isotope dilution/mass spectrometry. J Clin Endocr Metab 72:39, 1991

2. Cushing H: The Pituitary Body and Its Disorders. JB Lippincott, Philadelphia, 1912

3. Danese RD, Aron DC: Principles of clinical epidemiology and their application to the diagnosis of Cushing's syndrome: Rev. Bayes meets Dr. Cushing. Endocrinologist 4:339, 1994

4. Reimondo G, Paccotti P, Minetto M, et al: The corticotrophin-releasing hormone test is the most reliable noninvasive method to differentiate pituitary from ectopic ACTH secretion in Cushing's syndrome. Clin Endocrinol (Oxf) 58:718, 2003

5. Nieman LK, Loriaux DL: Corticotropin releasing hormone: clinical applications. Annu Rev Med 40:331, 1989

6. Hodge BO, Froesch TA: Familial Cushing's syndrome: micronodular adrenocortical dysplasia. Arch Intern Med 148:1133, 1988

7. Oldfield EH, Chrousos GP, Schulte HM, et al: Preoperative lateralization of ACTH secreting pituitary microadenomas by bilateral and simultaneous inferior petrosal sinus sampling. N Engl J Med 312:100, 1985

8. Ng L, Libertino JM: Adrenocortical carcinoma: diagnosis, evaluation and treatment. J Urol 169:5, 2003

9. Schteingart DE: Current perspective in the diagnosis and treatment of adrenocortical carcinoma. Rev Endocr Metab Disord 2:323, 2001

10. NIH state-of-the-science statement on management of the clinically inapparent adrenal mass ("incidentaloma"). NIH Consens State Sci Statements 19:1, 2002

11. Pena CS, Boland GW, Hahn PF, et al: Characterization of intermediate (lipid poor) adrenal masses: use of washout characteristics of contrast-enhanced CT. Radiology 217:798, 2000

12. Dunnick NR, Korobkin M: Imaging of adrenal incidentalomas: current status. AJR Am J Roentgenol 179:559, 2002

13. A survey on adrenal incidentaloma in Italy. Study Group on Adrenal Tumors of the Italian Society of Endocrinology. J Clin Endocrinol Metab 85:637, 2000

14. Hamrahian AH, Oseni TS, Arafah BM: Measurement of serum free cortisol in critically ill patients. N Engl J Med 350:1629, 2004

15. Dunnick NR, Doppman JL, Gill JR Jr, et al: Localization of functional adrenal tumors by computed tomography and venous sampling. Radiology 142:429, 1982

16. Young WF Jr, Horgan MJ, Klee GG, et al: Primary aldosteronism: diagnosis and treatment. Mayo Clin Proc 65:96, 1990

Acknowledgment

Figures 3, 4, and 5 Seward Hung.

V PITUITARY

Shlomo Melmed, m.d.

Functional Anatomy of the Pituitary

The pituitary gland regulates the critical hormonal functions of growth, development, reproduction, stress homeostasis, and metabolic control. Because of its prominent role in these processes, the pituitary has been termed the master gland.

The pituitary is situated within the sella turcica at the base of the brain and weighs about 600 mg. It comprises functionally distinct anterior and posterior lobes. The blood supply to the anterior pituitary is predominantly derived from the hypothalamic-pituitary portal vessels. The posterior lobe is supplied directly by the systemic inferior hypophyseal arteries.

Anatomically and functionally, the pituitary is closely linked with the hypothalamus [*see Figure 1 and Table 1*]. Neural cell bodies in the hypothalamus synthesize releasing and inhibiting hormones that control pituitary hormone secretion. These hypothalamic hormones are secreted into the portal vessels of the pituitary stalk and are transported to the anterior pituitary cell surface receptors.

The anterior pituitary synthesizes and secretes adrenocorticotropic hormone (ACTH), growth hormone (GH), prolactin (PRL), thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), and luteinizing hormone (LH).[1,2] The posterior pituitary secretes vasopressin (also known as antidiuretic hormone [ADH]) and oxytocin, both of which are synthesized in the hypothalamus.

Pituitary tropic hormones elicit responses from their respec-

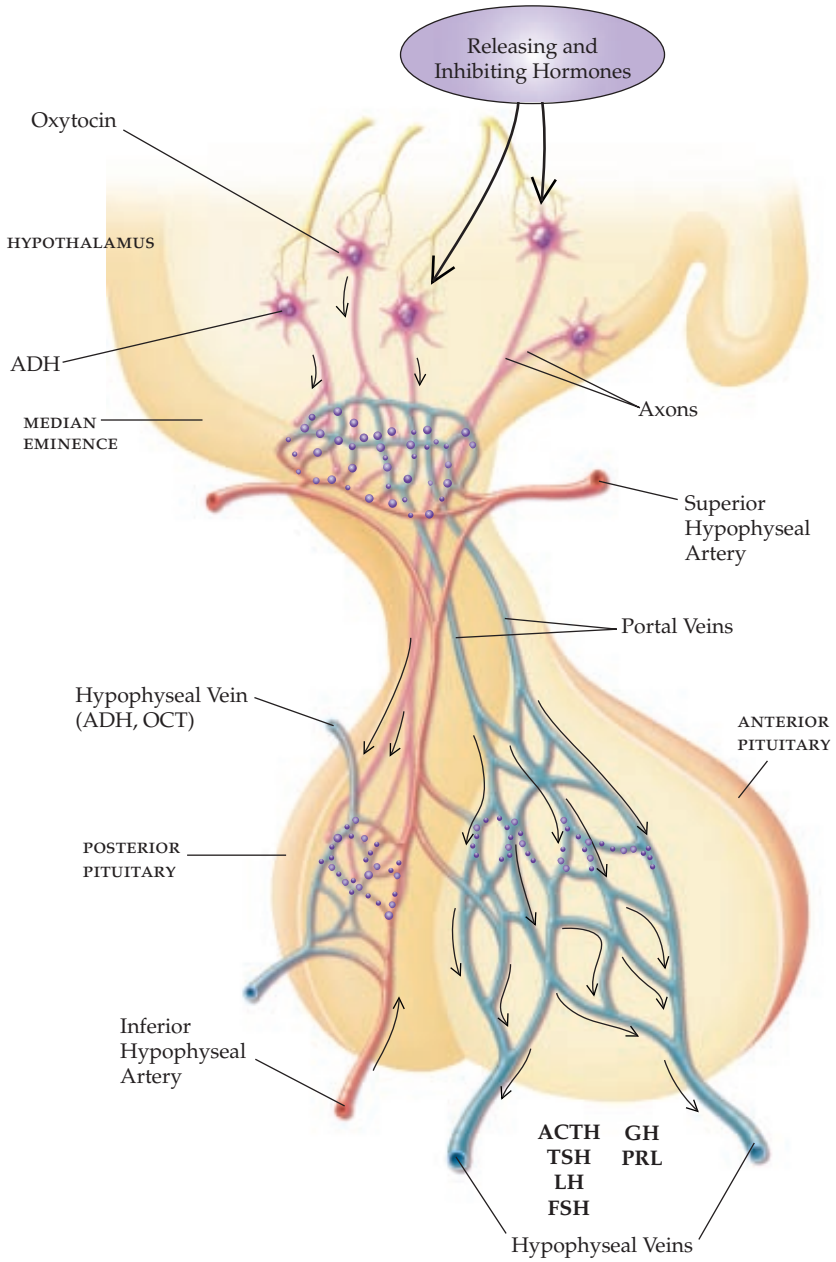


Figure 1 **The anterior pituitary and the hypothalamus are connected by the hypophyseal portal vasculature. Releasing or inhibiting hormones secreted by hypothalamic neurons enter the primary plexus of the hypophyseal portal vasculature. They flow down the long portal veins in the pituitary stalk to the secondary plexus, a capillary network that enmeshes the cells of the anterior pituitary. The anterior pituitary cells secrete their hormones in response to the releasing hormones. Because neither the hypothalamus nor the anterior pituitary is isolated by the blood-brain barrier, feedback signals have direct access to both sites of regulation. The posterior pituitary is made up of the terminal portions of neurons whose origin is the hypothalamus. (ACTH—adrenocorticotropic hormone; ADH—antidiuretic hormone; FSH—follicle-stimulating hormone; GH—growth hormone; LH—luteinizing hormone; PRL—prolactin; TSH—thyroid-stimulating hormone)**

Table 1 Hypothalamic and Related Pituitary Hormones

Hypothalamic Hormones	*Pituitary Hormones*
Growth hormone–releasing hormone	Growth hormone (GH)
Growth hormone release–inhibiting hormone (somatostatin)	GH
Prolactin release inhibitory factor (dopamine)	Prolactin
Gonadotropin-releasing hormone	Follicle-stimulating hormone Luteinizing hormone
Corticotropin-releasing hormone Vasopressin (arginine vasopressin; antidiuretic hormone)	Adrenocorticotropic hormone (corticotropin)
Thyrotropin-releasing hormone	Thyrotropin (thyroid-stimulating hormone)

tive target glands; the latter secrete endocrine hormones that activate specific tissue receptors. Circulating levels of these peripheral hormones influence secretion of their respective pituitary tropic hormone by negative feedback [*see Table 2*].

Pituitary Masses

LOCAL MASS EFFECTS

Pituitary masses can cause symptoms by secreting hormones, by impinging on adjacent structures, or both. These masses may also compress adjacent normal pituitary tissue, leading to pituitary failure. Expanding intrasellar lesions can exert significant compressive effects on surrounding vascular and neurologic structures, including the cavernous sinuses, cranial nerves, and optic chiasm. Intrasellar lesions may invade contiguous local structures and may compress central structures, depending on their anatomic location [*see Figure 2*]. The sellar roof presents the least resistance to soft tissue expansion from within the confines of the bony sella; this accounts for the vulnerability of the optic chiasm to sellar mass expansion. Small changes in intrasellar pressure may stretch the dural plate and cause headache, the severity of which does not necessarily correlate with mass size or extension. Chiasmic pressure can result in bilateral or unilateral visual defects. Pituitary stalk compression encroaches on the portal vessels, with resultant hyperprolactinemia and concurrent failure of other pituitary tropic hormones. Cavernous sinus invasion may lead to palsies of the third, fourth, and sixth cranial nerves, as well as lesions of the ophthalmic and maxillary branches of the fifth cranial nerve. Inferior extension through the bony sellar floor involves the sphenoid sinus; further extension into the palate roof may result in nasopharyngeal invasion and, rarely, cerebrospinal fluid leakage. Tumor invasion of the temporal or frontal lobe can cause seizures and personality disorders. Hypothalamic encroachment by an invasive pituitary mass may have metabolic sequelae, including precocious puberty or hypogonadism, diabetes insipidus, dysthermia, appetite disorders, and sleep disturbances.

PITUITARY TUMORS

Pituitary Adenomas

Pituitary adenomas account for about 15% of all intracranial neoplasms. They arise from one of the specific anterior pituitary cell types as benign monoclonal expansions. Loss of heterozygosity of regions of chromosome 11q13, 13, and 9 occurs in up to 20% of larger sporadic pituitary tumors, suggesting the presence of tumor suppressor genes at these loci. Other factors involved in initiation and promotion of pituitary adenoma growth include loss of negative feedback inhibition, as seen with thyroidal or gonadal failure; intrapituitary paracrine growth factors (angiogenesis factors), mainly mediated by estrogen; and activation of any of several oncogenes.

Pituitary adenomas are usually diagnosed when they hypersecrete pituitary hormones or compress adjacent structures. Tumors arising from lactotroph, somatotroph, corticotroph, and thyrotroph cells hypersecrete PRL, GH, ACTH, or TSH, respectively [*see Table 3*]. Functional tumors exhibit autonomous tropic hormone secretion, leading to hyperprolactinemia, acromegaly, Cushing disease, or, rarely, TSH hypersecretion. Plurihormonal tumors may produce mixed clinical features. About one third of adenomas do not actively secrete hormones and are clinically nonfunctional. On autopsy, up to one quarter of patients are found to harbor an unsuspected microadenoma (diameter < 10 mm) with no apparent clinical sequelae. Rarely, ectopic secre-

Table 2 Pituitary Hormones, Their Mediators, and Their Effects

Pituitary Hormones	*Stimulators*	*Inhibitors*	*Target Glands*	*Tropic Effects*
Gonadotropins: follicle-stimulating hormone, luteinizing hormone	Gonadotropin-releasing hormone	Sex steroids, inhibin	Ovary, testis	Sex steroid production, reproductive activity
Thyroid-stimulating hormone	Thyrotropin-releasing hormone (TRH)	Triiodothyronine (T_3), thyroxine (T_4), dopamine, somatostatin, glucocorticoids	Thyroid	T_3, T_4 synthesis and secretion
Prolactin	Estrogen, TRH	Dopamine	Breast, other tissues	Milk production
Growth hormone (GH)	GH-releasing hormone, GH secretagogue	Somatostatin, insulinlike growth factor (IGF)	Liver, bones, other tissues	IGF-1 production, growth induction, insulin antagonism
Adrenocorticotropic hormone	Corticotropin-releasing hormone, vasopressin, cytokines	Glucocorticoids	Adrenal	Steroid production

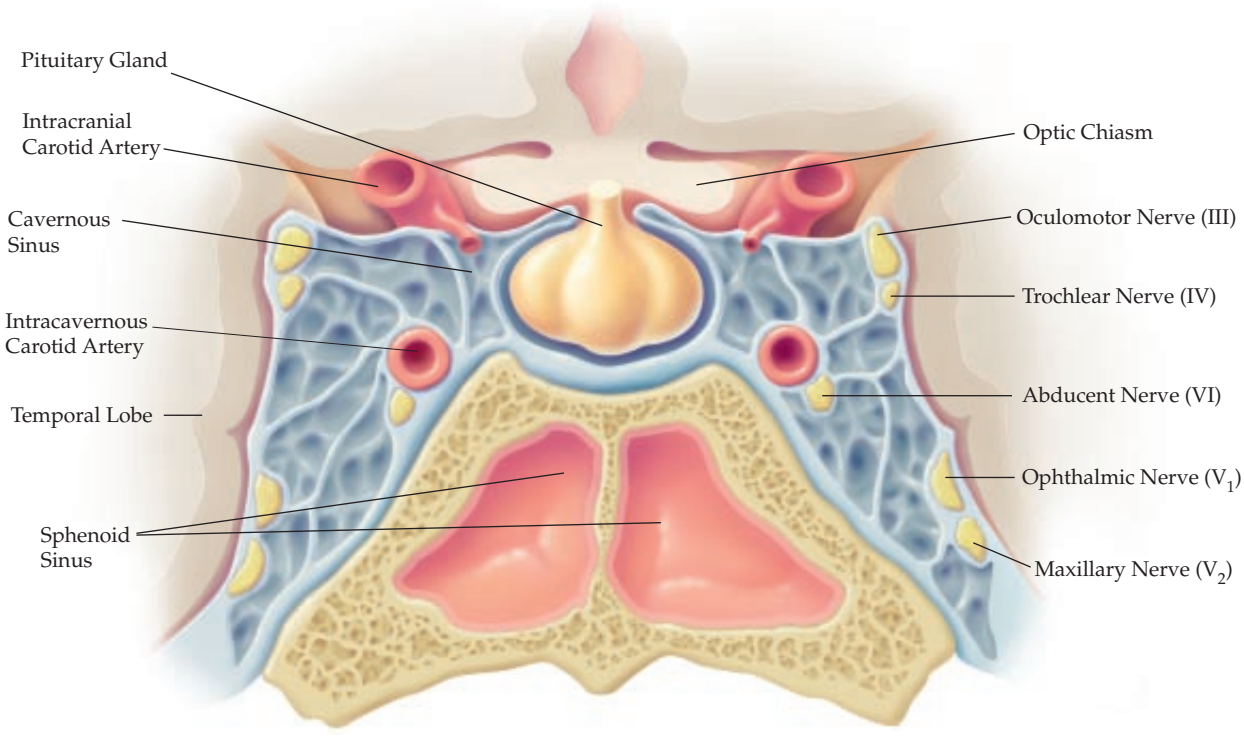


Figure 2 **Cross section of the pituitary gland and adjacent structures.**

tion of GH-releasing hormone (GHRH) or corticotropin-releasing hormone (CRH) elaborated by abdominal or chest tumors results in hyperplasia of the cells that secrete GH or ACTH; these patients may present with pituitary hyperplasia and acromegaly or Cushing syndrome.

Genetic Syndromes Associated with Pituitary Adenomas

Multiple endocrine neoplasia type I Multiple endocrine neoplasia type I (MEN I) is an autosomal dominant syndrome caused by an inactivating mutation in the coding region of *menin*, a tumor suppressor gene located at the q13 locus of chromosome 11. The syndrome comprises parathyroid, pancreatic, and pituitary adenomas, including prolactinomas, and may present as acromegaly or Cushing syndrome.

Carney syndrome Carney syndrome is an autosomal dominant syndrome associated with activated protein kinase activity. It comprises spotty skin pigmentation; myxomas; and testicular, adrenal, and pituitary adenomas.

McCune-Albright syndrome The McCune-Albright syndrome is associated with chromosome 20q13.2 mosaicism and constitutive activation of cyclic adenosine monophosphate (cAMP). This syndrome manifests as polyostotic fibrous dysplasia (cancellous bone is replaced with immature woven bone and fibrous tissue), pigmented skin patches, precocious puberty, and acromegaly.

Familial acromegaly Affected persons with this rare syn-

Table 3 Effects of Pituitary Adenomas

General Effect	*Adenoma Cell Origin*	*Hormone Product*	*Clinical Syndrome*
Hormone hypersecretion	Lactotroph	PRL	Hypogonadism, galactorrhea
	Somatotroph	GH	Acromegaly/gigantism
	Corticotroph	ACTH	Cushing disease
	Mixed growth hormone and prolactin cell	GH, PRL	Acromegaly, hypogonadism, galactorrhea
	Acidophil stem cell, mammosomatotroph	PRL, GH	Hypogonadism, acromegaly
	Thyrotroph	TSH	Hyperthyrotoxinemia
	Other plurihormonal cell	Any	Mixed
Hypopituitarism	Gonadotroph	FSH, LH, subunits	Silent or hypogonadism
	Null cell	None	Pituitary failure from mass effect
	Oncocytoma	None	Pituitary failure from mass effect

Note: all tumors may cause local pressure effects, including visual disturbances, cranial nerve palsy, and headache.
ACTH—adrenocorticotropic hormone FSH—follicle-stimulating hormone GH—growth hormone LH—luteinizing hormone PRL—prolactin TSH—thyroid-stimulating hormone

drome have acromegaly or gigantism and exhibit loss of heterozygosity at an 11q13 chromosomal locus distinct from that of *menin*.

Anterior Pituitary Hormones and Associated Disorders

PROLACTIN

Synthesis

Lactotrophs comprise about 20% of the anterior pituitary cells. Estrogen causes lactotroph cell hyperplasia, which occurs transiently during pregnancy and lactation. Central inhibitory control of PRL secretion is mediated predominantly by dopamine from the hypothalamus. Physiologic, pharmacologic, or pathologic alterations in dopamine availability or action disrupt PRL regulation. For example, if the hypophyseal-portal system is disrupted by pituitary compression or pituitary stalk damage and the flow of hypothalamic dopamine to the anterior pituitary is compromised, the resulting loss of lactotroph inhibition leads to PRL hypersecretion.[3]

Secretion

Normal serum PRL levels are 10 to 25 µg/L. The PRL level rises approximately 10-fold during pregnancy, as does the estrogen level. The PRL level declines rapidly within 2 weeks after delivery and returns to normal during the subsequent 3 months. Basal levels remain elevated during breastfeeding, and suckling induces a transient (approximately 30 minutes) reflex rise in PRL level.

Actions

PRL induces and maintains puerperal lactation. It also attenuates reproductive function, thus helping to ensure that lactation is not interrupted by pregnancy. In the primed puerperal breast, integration of multihormonal signals—from PRL, placental lactogens, progesterone, and local paracrine growth factors—leads to lactation. PRL also enhances milk production by improving calcium absorption and mobilization.

Hyperprolactinemia

Etiology Hyperprolactinemia has many possible causes; it may be physiologic, pathologic, or iatrogenic in origin [*see Table 4*]. Pregnancy, lactation, nipple stimulation, and chest wall lesions (including surgical incisions and herpes zoster) are associated with hyperprolactinemia. PRL-secreting pituitary adenomas (prolactinomas) produce the highest elevations of serum PRL levels (*see* Prolactinomas, *below*). Medications, compromised pituitary stalk function, hypothyroidism, and renal failure typically produce lesser elevations in PRL level [*see Table 1*]. Hypothalamic dopamine delivery may be disrupted by hypothalamic tumors, cysts, infiltrations, and radiation-induced damage. Plurihormonal tumors commonly hypersecrete PRL, and clinically nonfunctioning pituitary tumors may also compromise stalk integrity and cause hyperprolactinemia.

Table 4 Causes of Hyperprolactinemia[39]

Physiologic hypersecretion	Pregnancy Lactation Chest wall lesions Sleep Stress
Hypothalamic-pituitary damage	Masses Craniopharyngioma Suprasellar pituitary mass extension Rathke cyst Meningioma Dysgerminoma Metastases Granulomas Infiltration Lymphocytic hypophysitis Trauma Pituitary stalk section Suprasellar surgery Cranial irradiation
Pituitary hypersecretion	Prolactinoma Acromegaly Empty sella syndrome
Systemic disorders	Chronic renal failure Hypothyroidism Cirrhosis Epileptic seizures
Drug-induced hypersecretion	Dopamine receptor blockers Phenothiazines (e.g., chlorpromazine, perphenazine) Butyrophenones (e.g., haloperidol) Thioxanthenes Metoclopramide Dopamine synthesis inhibitor α-Methyldopa Catecholamine depletor Reserpine Opiates H_2 antagonists (e.g., cimetidine, ranitidine) Imipramines Amitriptyline, amoxapine Selective serotonin reuptake inhibitors (e.g., fluoxetine) Calcium channel blockers (e.g., verapamil) Hormones Estrogens Antiandrogens

Diagnosis The clinical features of hyperprolactinemia vary by the sex of the patient. In males, PRL attenuates LH secretion, leading to low testosterone levels. Men with hyperprolactinemia present with diminished libido and diminished sexual potency, oligospermia, and lowered ejaculate volume; up to about 30% may have galactorrhea. In women, hyperprolactinemia leads to loss of pulsatile LH secretion, blunting of the LH peak, hypoestrogenism, and anovulation [*see Figure 3*]. Women with hyperprolactinemia develop oligomenorrhea and amenorrhea. Anovulation and estrogen deprivation result in vaginal dryness, dyspareunia, loss of libido, and infertility. Hyperprolactinemia is also associated with enhanced risk of bone loss, which is further exacerbated by associated hypoestrogenemia.

In patients with clinical complaints consistent with hyperprolactinemia, a careful history and physical examination may reveal the source of the problem. Laboratory studies are indicated to exclude hypothyroidism, which can cause hyperprolactin-

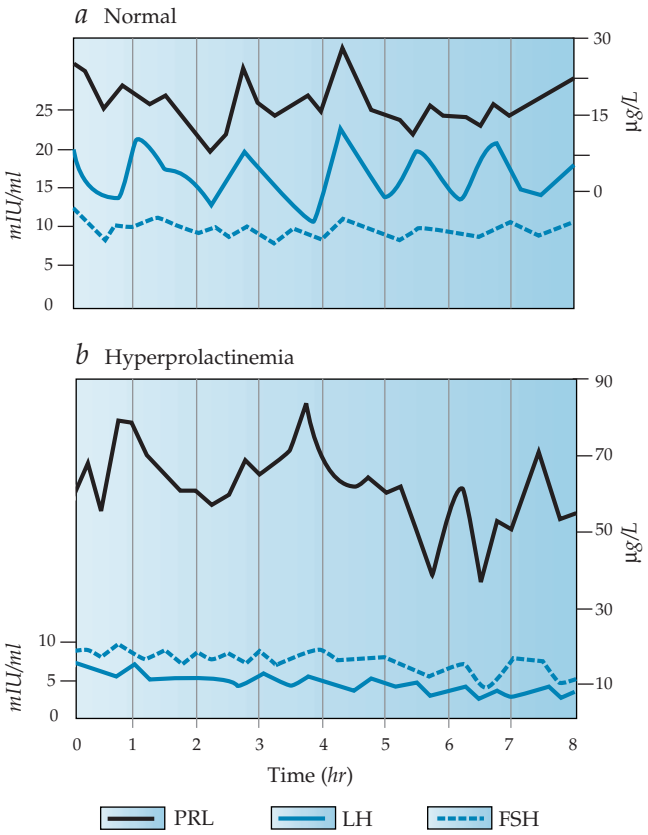


Figure 3 **In women, hyperprolactinemia results in loss of pulsatile luteinizing hormone (LH) secretion and lowering of follicle-stimulating hormone (FSH) levels.[39] (PRL—prolactin)**

emia. Alternatively, hypothyroidism and hyperprolactinemia can result from pituitary disease.

The degree of PRL elevation may offer a clue to the source of the prolactinemia. Prolactinomas account for most elevations of prolactin level higher than 100 µg/L; serum prolactin levels greater than 200 µg/L almost invariably indicate a prolactinoma.

All patients with symptoms of hyperprolactinemia and PRL levels above 30 µg/L should undergo MRI imaging of the pituitary. Small microadenomas (< 2 mm), which are undetectable on MRI scanning, may account for most cases of idiopathic hyperprolactinemia.

Treatment Treatment of hyperprolactinemia is aimed at normalizing PRL levels, alleviating gonadal dysfunction and galactorrhea, and preserving bone mineral density. Medications known to alter PRL levels should be discontinued, if possible. Dose titration of critical neuroleptic drugs with a dopamine agonist can normalize serum PRL levels and alleviate reproductive dysfunction. Hyperprolactinemia usually resolves after thyroid hormone replacement in hypothyroid patients and after renal transplantation in patients with chronic renal failure who are on dialysis. Hypothalamic or nonadenomatous sellar mass lesions should be removed surgically. Spontaneous resolution of hyperprolactinemia occurs in up to 30% of patients, whether or not they have a visible pituitary microadenoma.

Prolactinoma

Prolactinomas arising from lactotrophs are the most common functional pituitary tumors, with an annual incidence of about three per 100,000 population. Microadenomas are less than 1 cm in diameter and do not invade the parasellar region. Macroadenomas are more than 1 cm in diameter, are locally invasive, and may compress vital structures, leading to symptoms such as headaches and visual defects. Microprolactinomas have a female preponderance (20:1). Macroadenomas occur equally in both sexes, although men usually present with larger tumors. Tumor size correlates with PRL concentrations—serum PRL values above 200 µg/L are invariably associated with larger adenomas.

Diagnosis Prolactinoma should be suspected in patients with clinical signs of hyperprolactinemia (see above) and high random PRL levels. Men with prolactinomas tend to have relatively higher PRL levels than do women with prolactinomas. Diagnosis is confirmed by visualizing a pituitary adenoma on MRI.

Treatment Prolactinomas can be treated medically, with cabergoline or bromocriptine [*see Table 5*], or, rarely, surgically.

Cabergoline is a long-acting dopamine agonist that suppresses prolactin for more than 14 days after a single oral dose and shrinks prolactinomas in most patients.[4] The dosage is 0.5 to 1.0 mg twice weekly. Normal serum prolactin levels are achieved in about 80% of patients with microadenomas; normal gonadal function is restored and galactorrhea improves or resolves in 90% of patients. In patients with macroadenomas, cabergoline normalizes prolactin levels and shrinks the tumor in about 70% of cases. Cabergoline may be more effective in patients resistant to bromocriptine. Adverse effects and drug intolerance are less commonly encountered with cabergoline than with bromocriptine.

Bromocriptine mesylate is a D_2 dopamine receptor agonist that normalizes prolactin secretion in up to 70% of patients with microadenomas; it decreases tumor size and restores gonadal function. Prolactin levels normalize in 70% of patients with macroadenomas, and tumor mass shrinkage of 50% or more is achieved in about 50% of patients. Headaches and visual disorders usually improve or resolve within days, and sexual function improves. Therapy is initiated with 0.625 to 1.25 mg given at bedtime with a

Table 5 Dopamine Agonists in the Treatment of Prolactinomas

Treatment Response	*Bromocriptine Patient Response (%)**	*Cabergoline Patient Response (%)†*
Microadenomas		
Prolactin level normalized	70	80
Menses resumed	70	80
Macroadenomas		
Prolactin level normalized	65	70
Menses resumed	85	80
Tumor shrinkage		
≥ 50%	40	25
< 50%	40	55
None	20	20
Visual-field improvement	90	70
Drug intolerance	15	5

*2.5–7.5 mg/day; bromocriptine is preferred for infertility because it is short acting and can be discontinued immediately on pregnancy confirmation.
†0.5–1 mg twice weekly; cabergoline offers better compliance because it is long acting and has fewer gastrointestinal side effects.

snack, and the dosage is gradually increased. Successful control is usually achieved with a daily dose of less than 7.5 mg (i.e., 2.5 mg t.i.d.). About 20% of patients are resistant to the drug.

Side effects of dopamine agonists include transient nausea, vomiting, and postural hypotension with faintness; these symptoms occur in about 25% of patients. Other side effects include reversible constipation, nasal stuffiness, nightmares, and insomnia. For women who cannot tolerate orally administered agonists, intravaginal administration of bromocriptine tablets is often effective.

Indications for surgical resection of prolactinomas include resistance to or intolerance of pharmacologic treatment and the presence of an invasive macroadenoma that causes compromised vision and that fails to rapidly improve with dopamine agonists.[5] Initial attempts at resection lead to normalization of prolactin levels in about 70% of patients with microprolactinomas but only 30% of patients with macroadenomas. Prolactinomas recur in up to 20% of patients within the first year after surgery; long-term recurrence rates for macroadenomas exceed 50%.

Therapeutic goals in patients with prolactinomas include control of hyperprolactinemia; reduction of tumor size; resolution of galactorrhea; and restoration of menses, fertility, or both [*see Figure 4*].[6] Dopamine agonists suppress PRL secretion and synthesis and lactotroph cell proliferation. Patients are monitored with measurement of serum PRL levels, pituitary MRI scans, and visual field examinations. Once controlled, PRL levels can be measured every 6 months and MRI can be performed every 2 years. Medication doses are titrated to the lowest levels required to normalize PRL levels, restore reproductive function, and shrink the tumor mass.

If fertility is not desired, no treatment of microprolactinoma may be needed. Such patients should be monitored through regular serial PRL measurements, pituitary MRI scans, and assessment of bone mineral density. For patients with macroadenomas, visual field testing is performed before initiating dopamine agonists. MRI results and visual fields should be assessed serially until the mass shrinks, and annually thereafter. Reduction in PRL levels invariably precedes radiographically evident tumor shrinkage, and failure to lower PRL levels usually portends lack of tumor shrinkage. Radiotherapy is reserved for the rare patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists or surgery.

Prolactinomas and pregnancy Women with prolactinomas who wish to become pregnant should receive bromocriptine and use barrier methods of contraception until they have had regular menses for 3 months; this will permit accurate conception dating. Contraception may then be discontinued. When pregnancy is confirmed, bromocriptine should be discontinued and PRL levels followed serially. The patient should be carefully monitored for headaches or visual field disturbance. Cabergoline is not approved for restoration of fertility.

During pregnancy, the pituitary swells and there is an increased risk of prolactinoma growth; in particular, up to 30% of macroadenomas may grow during pregnancy. In women harboring macroadenomas, bromocriptine is restarted if visual field defects develop. Although pituitary MRI is considered safe during pregnancy, it is reserved for patients who develop severe headache or documented visual field defects. In the rare cases in which vision is threatened during the third trimester, surgical decompression may be indicated. Bromocriptine can be safely restarted during pregnancy. Comprehensive surveillance data do not indicate an adverse impact on the fetus; nevertheless, this approach should be undertaken cautiously, and only with the patient's informed consent.

GROWTH HORMONE

Synthesis and Secretion

GH is the most abundant anterior pituitary hormone; GH-secreting somatotroph cells constitute about 50% of the pituitary cell population. GH is encoded by five distinct genes situated on chromosome 17q22. The pituitary GH gene gives rise to a circulating form of GH that is 22 kilodaltons in size, and to a less abundant, cleaved 20-kd GH molecule. Placental syncytiotrophoblast cells express a GH variant, as well as chorionic somatotrophin. Somatotroph development and pituitary GH expression are largely determined by the Pit-1 nuclear transcription factor, whose mutations may also account for rare cases of hereditary pituitary failure.

GH secretion is controlled by complex hypothalamic and peripheral factors. Hypothalamic GHRH and somatostatin release-inhibitor factor (SRIF) stimulate and inhibit GH secretion, respectively. Ghrelin is synthesized predominantly in the gastrointestinal tract and stimulates GH secretion by binding to a specific pituitary GH secretagogue receptor.[7] SRIF is also expressed in extrahypothalamic tissues, including the GI tract and the pancreas. SRIF binds to five distinct receptor subtypes (SSTR1 to SSTR5), of which SSTR2 and SSTR5 are expressed on the surface membranes of pituitary cells. Signaling through the SSTR2 and SSTR5 subtypes preferentially suppresses secretion of GH (and also TSH). Insulinlike growth factor–1 (IGF-1), the peripheral target hormone for GH, inhibits GH via negative feedback.

GH secretion occurs in pulsatile peaks, interspersed by periods during which GH may be undetectable.[8] GH secretion peaks during puberty and declines by middle age, in parallel with age-related decline in muscle mass. Mean integrated nocturnal GH levels are at least twice that of daytime levels. GH levels rise within 1 hour after onset of deep sleep, as well as after exercise and trauma. GH secretion is low in the elderly and the obese and is higher in women; GH secretion is enhanced by estrogen replacement therapy. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition and prolonged fasting. Glucose loading suppresses GH to below 0.7 µg/L in women and below 0.07 µg/L in men. A complex interaction of nutritional factors and hypothalamic appetite-regulating peptides, including leptin, mediate GH secretion. Therefore, random measurements of GH levels do not readily identify adult patients with GH deficiency.[9] Differences in linear growth patterns in males and females may reflect differences in GH pulsatility. Higher GH pulses observed in males, as compared with the relatively continuous GH secretion patterns in females, may determine liver enzyme induction and postreceptor activity levels of GH-signaling molecules.

Actions

Peripheral GH receptors are most abundant in the liver. The extracellular domain of GH receptors is a soluble form—GH binding protein (GHBP)—which circulates in the blood. Binding of GH to its receptor induces intracellular signaling that is mediated by a phosphorylation cascade involving the Janus kinase-signal transducer and activator of transcription (JAK/STAT) pathway.[10]

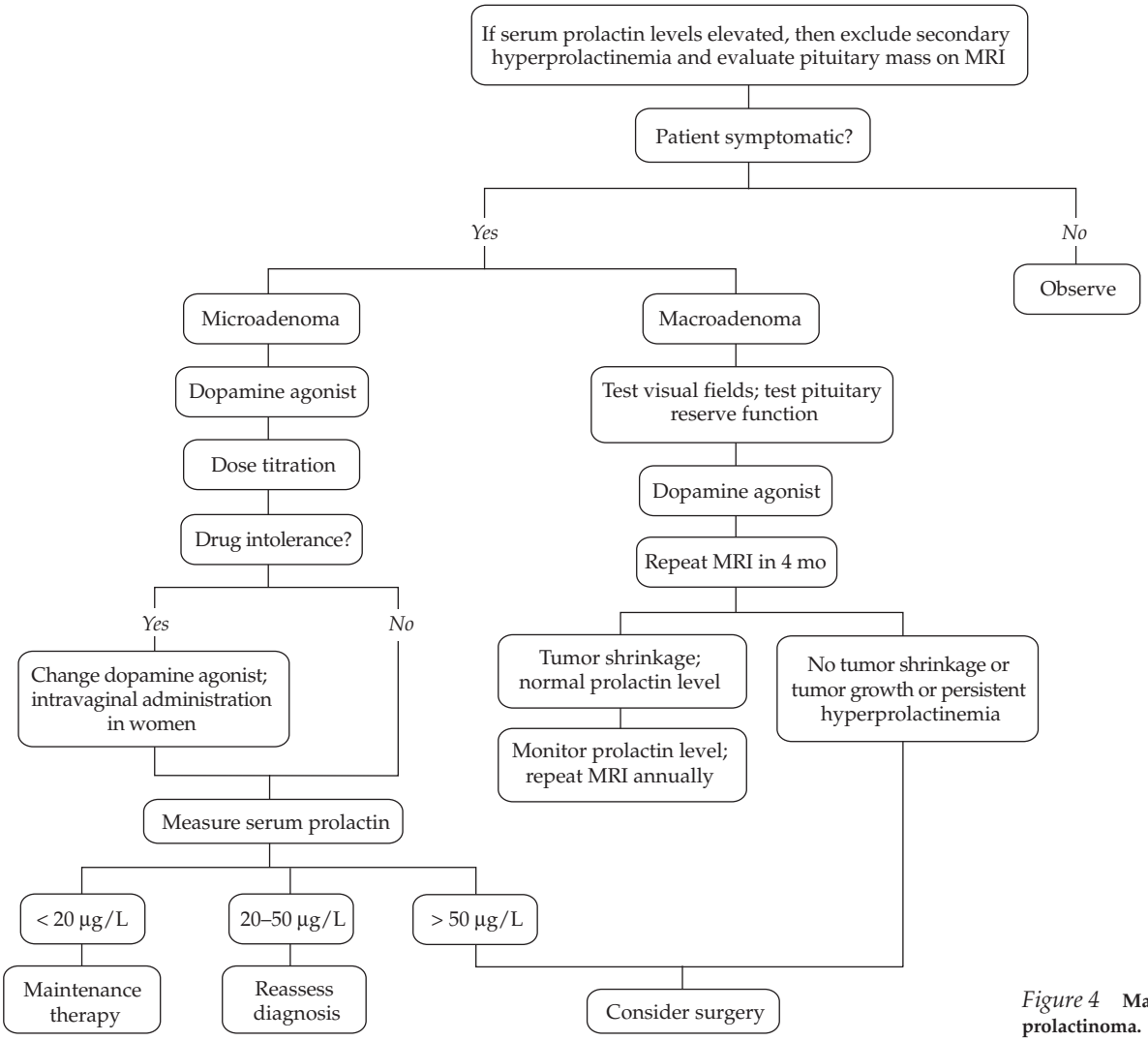


Figure 4 **Management of prolactinoma.**

In children and adolescents, GH stimulates the differentiation of epiphyseal prechondrocytes into IGF-1–responsive cells. GH also induces local IGF-1 and chondrocyte expansion. Linear growth is maintained by complex endocrine and paracrine mechanisms. In persons of all ages, GH antagonizes insulin action, impairs glucose tolerance, induces protein synthesis, enhances lipolysis, and increases lipid oxidation.

Insulinlike growth factors The IGF family of polypeptide growth factors comprises IGF-1, IGF-2, and proinsulin. IGF-1 in peripheral tissue exerts local paracrine actions, which are both GH dependent and independent. GH induces increases in circulating levels and tissue levels of IGF-1. Both IGF-1 and IGF-2 are bound to one of six high-affinity circulating IGF binding proteins (IGFBPs) that also regulate IGF bioactivity. IGFBP3 is GH dependent and is the major carrier protein for circulating IGF-1. GH deficiency, GH insensitivity, and malnutrition are associated with low IGFBP3 levels. Serum IGF-1 levels increase throughout puberty, peak at 16 years of age, and decline thereafter. Concentrations are higher in female subjects, especially during puberty. IGF-1 levels are lower in patients with GH deficiency, cachexia, malnutrition, or sepsis; they are invariably high in patients with acromegaly.

Adult GH Deficiency

Somatotroph damage and the subsequent development of pituitary tropic hormone deficiency follows a sequential pattern in which loss of adequate GH reserve usually foreshadows subsequent deficits of other pituitary hormones. The presence of central hypogonadism, hypothyroidism, or hypoadrenalism invariably implies concomitant GH deficiency. About half of all patients with pituitary insufficiency will already manifest GH deficiency if specifically tested.

Diagnosis Clinically, GH deficiency in adults is marked by impaired quality of life, body composition changes, and decreased exercise capacity [*see Table 6*]. Cardiovascular risk factors increase in patients with GH deficiency; indeed, the increase in mortality associated with adult hypopituitarism, and, possibly, GH deficiency in particular, is primarily from cardiovascular and cerebrovascular disease.[11] Because adult GH deficiency is rare and its symptoms are largely nonspecific, patients should be carefully selected for evaluation on the basis of well-defined risk criteria. These criteria include a history of pituitary surgery; pituitary or hypothalamic mass lesions; cranial irradiation; the need for GH replacement therapy in childhood, or the finding of a low IGF-1 level, as compared to the age- and sex-

Table 6 Findings in Adult Growth Hormone Deficiency[39]

Clinical Manifestations
- Impaired quality of life
  - Decreased energy and drive
  - Poor concentration
  - Low self-esteem
  - Social isolation
- Body-composition changes
  - Increased body fat mass
  - Abdominal fat
  - Increased waist-hip ratio
  - Decreased lean body mass

Imaging Studies
- Pituitary: mass or structural damage
- Bone: reduced density
- Abdomen: excess omental adiposity

Laboratory Tests
- Evoked GH level < 3 ng/ml
- IGF-1 and IGFBP3 levels low or normal
- Lipid disorders
- Concomitant deficits in gonadotropin, TSH, or ACTH reserve

Reduced Exercise Capacity
- Reduced maximum O_2 uptake
- Impaired cardiac function
- Reduced muscle mass

Cardiovascular Risk Factors
- Impaired cardiac structure and function
- Abnormal lipid profile
- Decreased fibrinolytic activity
- Atherosclerosis
- Omental obesity

ACTH—adrenocorticotropic hormone GH—growth hormone
IGF—insulinlike growth factor TSH—thyroid-stimulating hormone

matched population.[12] A subnormal evoked GH response (i.e., < 3 µg/ml) to a standard GH stimulation test establishes the diagnosis of adult GH deficiency. If other pituitary tropic hormone deficits are present, GH deficiency will be an inevitable concomitant finding; for that reason, some experts have recommended that GH testing not be required in this setting.[13] About 25% of GH-deficient adults have normal IGF-1 levels.

Treatment GH replacement is indicated for adult patients with unequivocal GH deficiency.[14] The decision to treat is also determined by informed patient perception of therapeutic benefits, including prevention of future ischemic heart disease and skeletal fractures, improved exercise capacity and energy levels, and enriched quality of life. For replacement therapy, GH is started at a dosage of 0.15 to 0.2 mg/day and titrated to a maximum of 1.25 mg/day, to maintain midrange age- and sex-matched IGF-1 levels. Women require higher GH doses than men, and elderly patients require lower doses.[15]

Contraindications to therapy include the presence of an active neoplasm or both uncontrolled diabetes and retinopathy. The risks of pituitary tumor regrowth are currently being assessed in long-term surveillance studies.

The side effects of GH replacement include reversible dose-related fluid retention, joint pain, and myalgia and parethesia associated with carpal tunnel syndrome. These side effects occur in up to 30% of patients.[16] Patients with type 2 (non–insulin-dependent) diabetes mellitus will initially experience increased insulin resistance. However, glycemic control may improve in association with sustained loss of abdominal fat during long-term GH replacement.

If after 6 months there is no clinical response to GH replacement, treatment should be discontinued. In patients who show a response, GH replacement is continued in conjunction with regular monitoring of IGF-1 levels, lipids, and bone density.

GH is not indicated for adults with intact pituitary function, except for those with AIDS-related cachexia. The hormone should not be used for nonapproved indications, because the risk of side effects—especially glucose intolerance and fluid retention—outweigh potential benefits ascribed to improved muscle energy and anti-aging properties. Results of prospective controlled trials for these potential indications are not yet at hand.

Acromegaly

Etiology GH hypersecretion usually results from a GH-secreting pituitary adenoma [*see Table 7*]. Occasionally, patients with partially empty sella may harbor a small GH-secreting adenoma within the compressed rim of pituitary tissue. Rarely, GH is secreted ectopically by abdominal or chest tumors. GHRH may be elaborated by hypothalamic tumors or carcinoid tumors in the chest or abdomen, causing acromegaly through chronic somatotroph overstimulation.

Diagnosis The manifestations of GH and IGF-1 hypersecretion are protean and develop slowly; they are often not diagnosed for 10 years or more [*see Table 8*]. Acral bony overgrowth results in frontal bossing, increased hand and foot size, and mandibular enlargement with prognathism and a widening of incisor spaces. GH hypersecretion that occurs before epiphyseal long-bone closure causes pituitary gigantism. Soft tissue swelling results in coarse facial features; increased heel pad thickness; and enlargement of the feet and hands, evidenced by increased shoe or glove size and ring tightening. Hyperhidrosis; oily skin; a deepening of the voice; arthropathy; kyphosis; carpal tunnel syndrome; proximal muscle weakness and fatigue; skin tags; and visceromegaly, including macroglossia, cardiomegaly, thyroid, and salivary gland enlargement, may be encountered. About 30% of patients develop coronary artery disease, cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, or hypertension. Sleep apnea, caused by soft tissue laryngeal airway obstruction or central sleep dysfunction, is an important comorbidity. Diabetes develops in 25% of patients, because GH is a potent insulin an-

Table 7 Causes of Acromegaly

- Excess growth hormone (GH) secretion
  - Pituitary (~98% of cases)
    - GH cell adenoma
    - Mixed GH cell and prolactin cell adenoma
    - Mammosomatotroph cell adenoma
    - Plurihormonal adenoma
    - GH cell carcinoma or metastases
    - Multiple endocrine neoplasia type 1 (GH cell adenoma)
    - McCune-Albright syndrome (rarely, adenoma)
    - Ectopic sphenoid or parapharyngeal sinus pituitary adenoma
  - Extrapituitary tumor (< 1% of cases)
    - Pancreatic islet cell tumor
- Excess GH-releasing hormone secretion
  - Central (< 1% of cases)
    - Hypothalamic hamartoma, choristoma, ganglioneuroma
  - Peripheral (~1% of cases)
    - Bronchial carcinoid
    - Pancreatic islet cell tumor
    - Small cell lung cancer
    - Adrenal adenoma
    - Medullary thyroid carcinoma
    - Pheochromocytoma

Table 8 Features of Acromegaly

Enlarged hands and feet	Profuse sweating/hot
Coarsening of facial features/prognathism	Headache
Bite problems	Carpal tunnel syndrome
Skin tags	Arthritis/arthralgias
Frontal bossing	Hypertension and heart disease
Cystic acne	Sleep apnea and snoring
Colonic polyps	Glucose intolerance
Deepening of voice	Hyperprolactinemia
Oily skin	Visual problems
	Sexual dysfunction

tagonist; most patients with elevated GH levels are intolerant of glucose. Colon polyps are present in up to one third of patients. Overall mortality is enhanced about threefold, primarily as a result of cardiovascular and cerebrovascular disorders and respiratory disease. Unless GH levels are tightly controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

Measurement of serum IGF-1 can be used for case finding in patients with possible acromegaly; in patients with GH hypersecretion, IGF-1 levels are invariably elevated, as compared with the levels in the age- and sex-matched population. Single random GH measurements are not useful for diagnosis. Instead, diagnosis is confirmed by demonstrating a failure to suppress GH levels to below 1 μg/L within 1 to 2 hours after an oral glucose load (75 g); about 20% of patients exhibit a paradoxical glucose-induced rise in GH. PRL levels are elevated in about 25% of patients. Thyroid function studies and assays of gonadotropin and sex steroid levels may show attenuation, which is the result of the compressive effects of an expanding pituitary mass.

Treatment Control of acromegaly can be achieved by a judicious application of multimodal therapeutic approaches.[17,18] Therapeutic interventions include surgery, somatostatin analogues, and dopamine agonists. Transsphenoidal surgical resection by an experienced surgeon is indicated for both microadenomas and macroadenomas. Resection results in control of disease in about 70% of patients with microadenomas but in less than 50% of patients with macroadenomas. GH levels fall rapidly after tumor resection, and IGF-1 levels return to normal within 3 to 4 days. The disorder recurs in about 10% of patients, and pituitary failure develops in up to 15% of patients after surgery. Persistent postoperative GH hypersecretion necessitates adjuvant therapy, typically with somatostatin analogues.

Octreotide acetate is an 8–amino acid synthetic somatostatin analogue that binds mainly to SSTR2 receptors and effectively controls GH hypersecretion.[19] Octreotide is given in a dosage of 50 to 400 μg subcutaneously every 8 hours. Within an hour of receiving an injection, most patients experience an 80% reduction in GH level. About 10% of patients show no response. Rapid relief of headache and soft tissue swelling occurs, with amelioration of excessive perspiration, obstructive apnea, and cardiac failure. Significant pituitary tumor shrinkage occurs in about 40% of patients.[20] A long-acting octreotide formulation, Sandostatin LAR Depot, provides sustained GH suppression, with effects lasting for up to 6 weeks after a 30 mg intramuscular injection. Long-term treatment with monthly injections of 20 to 40 mg maintains GH and IGF-1 suppression in about 70% of patients, and pituitary tumor size is controlled. Because it is effective, and well tolerated and is less inconvenient for the patient than subcutaneous preparations, the long-acting formulation is the medical treatment choice for these patients.[21,22]

Side effects of somatostatin analogues are typically minor and transient; they are mostly related to suppression of GI motility and secretion. Nausea, abdominal discomfort, diarrhea, and flatulence occur in one third of patients but usually remit within 2 weeks. In the United States, up to 30% of patients receiving long-term treatment develop echogenic gallbladder sludge or asymptomatic cholesterol gallstones. Mild glucose intolerance, hypothyroxinemia, asymptomatic bradycardia, and local pain at the injection site have been reported.

Bromocriptine may suppress GH secretion in some patients. High doses (i.e., 20 mg/day or more) are usually required. About 10% of patients receiving bromocriptine have normalized IGF-1 levels. Cabergoline suppresses GH when given at a relatively high dosage (i.e., 0.5 mg/day). Combination treatment with octreotide plus cabergoline offers additive biochemical control compared with either drug alone.

GH antagonists have been developed. These new GH analogues antagonize GH action by blocking peripheral GH receptor binding. Pegvisomant, administered in daily subcutaneous injections, lowers serum IGF-1 levels and so may block the deleterious peripheral effects of GH. GH levels may remain elevated, but the excess hormone is effectively inactive. Long-term monitoring of pituitary adenoma size and liver function testing are suggested. The drug is particularly useful in those patients with persistently elevated IGF-1 levels and controlled GH levels.

External radiation therapy or high-energy radiosurgery suppresses GH levels to below 5 μg/L, although 50% of patients require at least 8 years of therapy for this outcome.[23] Interim medical therapy is required in the years before patients attain maximal radiation benefits. Most patients also develop gonadotropin, ACTH, or TSH deficiency within 10 years of therapy. Rarely, visual deficits, brain necrosis, or new tumor formation are encountered. Stereotactic ablation of GH-secreting adenomas by gamma-knife radiosurgery is promising, but compelling long-term results are not yet available, and long-term side-effect profiles have not been established.

The initial treatment option for well-circumscribed GH-secreting tumors is surgical resection. Somatostatin analogues reduce GH hypersecretion and are used for preoperative shrinkage of large, invasive macroadenomas; for immediate relief of debilitating symptoms in frail patients experiencing morbidity; for patients who decline surgery; and for patients in whom surgery fails to result in biochemical control, as is inevitable in cases of invasive adenoma.[24] Irradiation or repeat surgery is indicated for patients for whom medical therapy fails. The main disadvantages of radiotherapy are the slow rate of biochemical response (i.e., 5 to 15 years) and the high rate of hypopituitarism. Comorbid features of acromegaly, including cardiovascular disease, diabetes, and arthritis, should be aggressively treated. Maxillofacial surgery may be indicated for mandibular repair.

Adrenocorticotropic Hormone Synthesis

Up to 20% of the pituitary consists of ACTH-secreting corticotroph cells. These cells express products of the *POMC* (pro-opiomelanocortin) gene, which include 1-39 ACTH, β-lipotropin, and endorphins. β-Lipotropin gives rise to α-lipotropin and β-endorphin; the latter contains the sequence for met-enkephalin. The *POMC* gene, located on chromosome 2, possesses different promot-

er regions that determine pituitary-specific and peripheral tissue–specific POMC expression, respectively. Ectopic ACTH/POMC transcripts are expressed in gonads, placenta, GI tissues, kidney, adrenal medulla, lung, and lymphocytes; POMC products also arise from peripheral neuroendocrine tumors.

Secretion ACTH synthesis and release are stimulated by CRH. In addition, ACTH release is induced by vasopressin, cytokines, physical stress, exercise, acute illness, and hypoglycemia.

ACTH secretion is pulsatile and follows a circadian rhythm that is highest in early morning and declines at night. This rhythm is paralleled by a diurnal pattern of adrenal glucocorticoid secretion. ACTH levels peak at 6 A.M., with values ranging from 8 to 25 pg/ml; peak values are approximately fourfold higher than the nadir levels measured between 11 P.M. and 3 A.M. Glucocorticoids suppress CRH and ACTH release. The loss of cortisol inhibition that occurs with primary adrenal failure results in extremely high compensatory ACTH levels.

Actions

The hypothalamic-pituitary-adrenal (HPA) axis maintains metabolic homeostasis and mediates the neuroendocrine stress response. The pituitary affects the pattern and quantity of adrenal cortisol secretion by integrating peripheral and central signals. The neuroendocrine stress response reflects the net result of sensitively integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals, resulting in cortisol production. The HPA axis is triggered by acute inflammatory or septic insults that mediate release of inflammatory cytokines, bacterial toxins, and neural signals. ACTH stimulates steroidogenesis by maintaining adrenal cell proliferation and function. Cortisol elevation curtails the inflammatory response and provides host protection.

Pro-opiomelanocortin peptides and appetite control Several lines of experimental and clinical evidence implicate the POMC system in appetite control. The melanocortin receptor family comprises important regulators of central appetite control. Inactivation of MC-2 receptors leads to obesity, hypoadrenalism, and red hair pigmentation. Disruption of MC4 receptors is associated with childhood obesity and elevations in the level of circulating leptins; disruption is also genetically linked to the POMC gene locus [*see 3:III Obesity*].

HPA Axis Testing

Insulin-induced hypoglycemia and cortisol levels Intravenous administration of insulin (0.05 to 0.3 U/kg) lowers blood glucose levels to 50% of baseline within 30 minutes. This evokes plasma cortisol increases of 7 mg/dl or greater; peak cortisol levels of at least 20 μg/dl are evoked within 30 to 45 minutes of nadir blood glucose levels and indicate intact pituitary ACTH reserve production.

Metyrapone The pituitary response to a decrease in the serum cortisol level can be assessed with metyrapone testing. A 3 g oral dose of metyrapone administered at 11 P.M. with a snack blocks conversion of the cortisol precursor 11-deoxycortisol (compound S) to cortisol. The resulting fall in serum cortisol level normally stimulates ACTH secretion, raising the compound S level to above 8 μg/dl at 8 A.M. the next morning. Cortisol inhibition by metyrapone can be confirmed by finding that the plasma cortisol level is less than 5 μg/dl.

Synthetic ACTH Injection of synthetic ACTH (Cortrosyn) at a dose of 250 μg intravenously or intramuscularly evokes adrenal cortisol reserve after 30 and 60 minutes. Cortisol levels should rise to at least twice the baseline value, rise at least 7 μg/dl, or peak at above 20 μg/dl; any one of those three reactions indicates normal reserve. Blunted cortisol responses to ACTH reflect compromised pituitary ACTH reserve, primary adrenal failure, or steroid ingestion.

CRH Intravenous CRH, 1 μg/kg, directly stimulates ACTH secretion during the 60 minutes after injection. Pituitary damage prevents an evoked response. Patients with Cushing disease associated with an ACTH-secreting corticotroph cell adenoma often have exaggerated ACTH responses to CRH. CRH injection does not stimulate a further rise in ACTH secretion by ectopic ACTH-secreting tumors.

ACTH Deficiency

Diagnosis Clinically, pituitary ACTH deficiency results in secondary hypocortisolism with tiredness, weakness, anorexia, nausea, and vomiting; occasionally, hypoglycemia results from diminished counterregulation of insulin. Stressful acute illness may unmask the presence of partial ACTH deficiency and cause life-threatening hypocortisolism.

On laboratory testing, ACTH deficiency is characterized by inappropriately low ACTH levels in conjunction with low cortisol levels. Low basal serum cortisol levels or blunted cortisol responses to provocative ACTH stimulation reflect diminished adrenal reserve caused by prolonged insufficient ACTH tropic action on the adrenal cortex.

Treatment Hydrocortisone replacement reverses most clinical and biochemical features of cortisol deficiency. Hydrocortisone is given two or three times daily. The total daily dose should usually not exceed 20 mg. Doses should be increased severalfold during periods of acute illness or stress.

ACTH-Secreting Adenoma (Cushing Disease)

ACTH-producing adenomas account for about 10% to 15% of all pituitary adenomas and are usually well-differentiated microadenomas. Cushing syndrome is also caused by ectopic ACTH production by tumors, including small-cell lung carcinomas and bronchial and thymic carcinoids. In contrast to ACTH secretion by pituitary tumors, which can be suppressed by high-dose glucocorticoids, ectopic ACTH secretion by neoplasms is usually not suppressible, a fact that highlights the unrestrained malignant gene expression.

Diagnosis Unrestrained ACTH secretion causes hypercortisolemia, which results in thin, brittle skin; central obesity; hypertension; plethoric moon facies; purple striae and susceptibility to bruising; glucose intolerance or diabetes; gonadal dysfunction; osteoporosis; proximal muscle weakness; acne; hirsutism; and labile depression, mania, or psychosis [*see Table 9*]. Leukocytosis, lymphopenia, and eosinopenia also may develop. In young women, osteoporosis may be particularly prominent. Cardiovascular disease is the primary cause of death.

The differential diagnosis of ACTH-secreting pituitary tumor includes other causes of hypercortisolism: iatrogenic glucocorticoid administration, ectopic ACTH-secreting tumor, and cortisol-secreting adrenal tumor. In ectopic Cushing syndrome, manifestations usually develop acutely: patients present with

Table 9 Clinical Features of Cushing Syndrome[40]

Symptoms and Signs	*Frequency (%)*
Obesity or weight gain (>115% ideal body weight)	80
Thin skin	80
Moon facies	75
Hypertension	75
Purple skin striae	65
Hirsutism	65
Abnormal glucose tolerance	55
Impotence	55
Menstrual disorders (usually amenorrhea)	60
Proximal muscle weakness	50
Truncal obesity	50
Acne	45
Bruising	45
Mental changes	45
Osteoporosis	40
Edema of lower extremities	30
Hyperpigmentation	20
Hypokalemic alkalosis	15
Diabetes	15

Note: manifestations seen in patients of all ages.

florid skin hyperpigmentation, severe myopathy, hypertension, hypokalemic alkalosis, glucose intolerance, and edema. Serum potassium levels are below 3.3 mmol/L in most patients with ectopic ACTH secretion.

The diagnosis of pituitary Cushing disease requires documentation of hypercortisolism in the presence of pituitary-derived ACTH elevation. Reproducible markers for hypercortisolism include the failure to experience suppression of the cortisol level after a dose of dexamethasone and an elevation in 24-hour urinary free cortisol level. Urinary cortisol levels greater than 300 μg/day indicate the presence of Cushing syndrome. Urinary 17-hydroxysteroid levels reflect secretion of cortisol metabolites.

In general, ACTH-secreting pituitary tumors retain feedback responsiveness to circulating glucocorticoids. Basal ACTH levels are usually about eightfold higher in patients with ectopic ACTH secretion, but considerable overlap with pituitary adenoma–derived ACTH may preclude an accurate biochemical distinction of the two disorders. In patients with endogenous (adrenal) or exogenous (iatrogenic) Cushing syndrome, ACTH levels are suppressed. Elevated concentrations of circulating ACTH and cortisol measured at midnight usually indicate the presence of Cushing syndrome.

Dynamic testing should be undertaken when hypercortisolemia has been rigorously documented.[25] Dexamethasone suppression of ACTH and ultimately of cortisol levels is the standard test for diagnosis of ACTH-dependent Cushing disease. Ingestion of 1 mg oral dexamethasone at 11 P.M. should result in suppression of serum cortisol levels to below 7 μg/dl at 8 A.M. the next morning, unless obesity, chronic depression, or alcoholism is present. In patients with ACTH-secreting pituitary adenomas or ectopic tumors, overnight dexamethasone does not suppress plasma ACTH or serum cortisol levels, and longer-term dexamethasone suppression testing is required. Baseline pretesting of 24-hour urinary free cortisol and 17-ketosteroids or 17-hydroxysteroid values is followed by administration of low-dose dexamethasone (0.5 mg every 6 hours) for 2 days. Plasma ACTH, serum cortisol, and 24-hour urinary free cortisol levels remain elevated in patients with Cushing disease. High-dose dexamethasone (2 mg every 6 hours) for the subsequent 2 days will usually suppress 17-hydroxysteroid levels by 50% or less and suppress urinary free cortisol to less than 90% of baseline in patients with pituitary ACTH-secreting tumors, but this result will be seen in only 10% of those patients with ectopic ACTH secretion.

MRI scanning is indicated for patients with documented hypercortisolemia and nonsuppressed ACTH levels. If a pituitary mass is clearly visible on MRI, transsphenoidal surgical resection should be undertaken after rigorous biochemical confirmation of pituitary-derived ACTH hypersecretion. However, most ACTH-secreting tumors are less than 5 mm in diameter; about half are less than 2 mm in diameter, and so are undetectable even by sensitive MRI. Therefore, MRI has only limited ability to visualize ACTH-secreting pituitary tumors. Bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may distinguish pituitary from ectopic ACTH hypersecretion.[26] Because most ectopic ACTH-secreting tumors are located in the chest or abdomen, imaging studies of those areas are indicated for diagnosis. The diagnosis of ectopic ACTH secretion is ultimately confirmed by four measures: (1) rigorous exclusion of a pituitary lesion; (2) demonstration of an arteriovenous ACTH gradient over the tumor bed; (3) resolution of hypercortisolism with excision of the tumor; and (4) confirmation of *POMC* gene expression in excised tumor tissue.

Adrenal imaging is indicated when suppressed ACTH levels point to an adrenal origin of hypercorticolism. Bilateral adrenal hyperplasia with cortical thickening usually indicates tropic effects of ACTH hypersecretion. Adrenal adenomas causing Cushing syndrome are usually clearly visible, and adrenal carcinomas are larger than adenomas (> 2 cm). The contralateral gland may be normal or atrophic. Adrenal nodularity may occur unilaterally or bilaterally, with approximately 50% of glands appearing normal [*see 3:IV The Adrenal*].

Treatment Selective transsphenoidal resection after careful preoperative localization is the preferred treatment for ACTH-secreting pituitary adenomas.[27] Remission rates are about 80% for microadenomas but less than 50% for the less common ACTH-secreting macroadenomas. After successful surgery, patients may experience a period of compensatory adrenal insufficiency for up to 6 months and may require low-dose cortisol replacement during that time. Within 5 years of the operation, approximately 5% of patients in whom surgery was initially successful will experience biochemical recurrence.

Patients with ACTH hypersecretion that is not controlled by surgery require pituitary irradiation. Cortisol-lowering agents (i.e., mitotane, ketoconozole, or aminoglutethimide) are administered after irradiation to achieve earlier biochemical remission. Rarely, all these measures fail, and bilateral adrenalectomy is required.

FOLLICLE-STIMULATING HORMONE AND LUTEINIZING HORMONE

Synthesis

Gonadotroph cells comprise up to 10% of anterior pituitary cells. The gonadotropins FSH and LH (along with TSH and human chorionic gonadotropin) are glycoprotein hormones comprising a common α and a specific β subunit. Gonadotroph cells exhibit cytoplasmic immunostaining for both FSH and LH β

subunits, as well as for the common α subunit. Primary gonadal failure, resulting from gonadal damage, is associated with hyperplastic gonadotroph cells with accumulation of hormone secretory granules, reflecting loss of negative feedback by peripheral sex steroids.

Hypothalamic gonadotropin-releasing hormone (GnRH) regulates both LH and FSH secretion. GnRH, under positive feedback control by peripheral estrogens, is secreted in a pulsatile fashion every 60 to 120 minutes; it regulates the complex reproductive cycles. Activins also induce gonadotropins, whereas inhibins suppress their secretion.

ACTIONS

Gonadotropins interact with their respective cell surface receptors on the ovary and testis, thereby controlling the development and maturation of germ cells and the synthesis of steroid hormones. In women, LH mediates ovulation and the maintenance of the corpus luteum, and FSH mediates ovarian follicle development and induces ovarian estrogen production. In men, LH induces testosterone secretion by the Leydig cells, and FSH regulates seminiferous tubule development and stimulates spermatogenesis.

Gonadotropin Deficiency

Gonadotropin secretion is sensitive to pituitary damage, and hypogonadism is the most common presenting feature of adult hypopituitarism. Congenital or acquired central hypogonadotropic hypogonadism results from a pituitary or hypothalamic disorder that disrupts GnRH availability. Hypothalamic defects causing hypogonadism include Kallmann syndrome and a mutation in the *DAX-1* gene that is associated with deficient GnRH and pituitary gonadotropin synthesis. Inactivating mutations in the LH and FSH β-subunit gene cause hypogonadism by disrupting gonadotropin formation and function.

Diagnosis Clinical features of hypogonadism depend on the age at onset of the disorder. Primary amenorrhea, immature internal and external genitalia, absent secondary sex characteristics, and eunuchoidal body proportions occur in adolescent girls. In premenopausal women, decreased ovarian function presents as oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, breast atrophy, and hot flushes [*see 3:III Ovary*]. The onset of hypogonadotropism during male adolescence results in sexual infantilism, with a smooth scrotum and small penis, diminished or absent postpubertal sex drive, absent secondary sexual characteracteristics, central obesity, eunuchoid proportions, delayed epiphyseal closure, and a characteristic high-pitched prepubertal voice. In men, testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, attenuated beard and body hair growth, soft testes, and fine facial wrinkles [*see 3:II Testes and Testicular Disorders*]. Prolonged hypogonadism results in osteoporosis in both females and males.

Central hypogonadism is diagnosed by a finding of low-normal or low serum gonadotropin levels and low sex hormone concentrations (testosterone in males, estradiol in females). Male patients have abnormal results on semen analysis. Normal values for circulating FSH and LH in menstruating women are 4 to 20 mIU/ml, depending on the menstrual phase; levels rise considerably with menopause. FSH and LH levels in men are 1 to 12 mIU/ml. Normal total testosterone levels in men are above 280 ng/100 ml; the serum testosterone concentration should be measured at about 8 A.M., when it is at its peak [*see 3:II Testes and Testicular Disorders*]. In women, circulating estradiol levels vary with the menstrual cycle.

Pituitary gonadotropin deficiency can sometimes be confirmed with GnRH stimulation testing. Gonadotrophs are stimulated by intravenous injection of 100 μg GnRH; evoked LH levels peak within 30 minutes, and FSH plateaus during the subsequent 60 minutes. Normal responses vary with the age and sex of the subject and, in women, the menstrual cycle stage. However, a robust gonadotropin response does not necessarily exclude pituitary gonadotroph damage, and the absence of a response does not reliably distinguish pituitary from hypothalamic causes of hypogonadism. In patients with documented central hypogonadism, pituitary MRI and pituitary function testing are required.

Treatment In premenopausal women, estrogen and progesterone replacement therapy results in the maintaining of secondary sexual characteristics and genitourinary tract integrity and prevents osteoporosis. Gonadotropin therapy is used for ovulation induction. Pulsatile GnRH is effective for treating hypothalamic hypogonadism [*see 3:III Ovary*]. In women who exercise vigorously and maintain a low body mass index (e.g., athletes and ballet dancers), caloric replacement may restore menses. In males, testosterone replacement therapy will result in the attaining and maintaining of growth and development of the external genitalia and secondary sexual characteristics, and patients will maintain libido, muscle mass, and bone density [*see 3:II Testes and Testicular Disorders*].

Nonfunctioning Pituitary Adenomas

So-called nonsecreting adenomas arising from gonadotroph cells are the most common pituitary adenomas. Because these adenomas are clinically nonfunctional, they usually produce no distinct hypersecretory syndrome.[28] Some adenomas express gonadotropin α-subunits but not intact FSH or LH molecules, and administration of thyrotropin-releasing hormone (TRH) may inappropriately evoke gonadotropins or subunit secretion. Clinically inactive, asymptomatic pituitary microadenomas are commonly encountered as incidental findings on MRI; these are termed pituitary incidentalomas.

Diagnosis Nonsecreting adenomas may be incidentally discovered on an MRI performed for another indication. Mass effects, including optic chiasm pressure and other neurologic symptoms, are the usual initial presenting symptoms of large tumors. Gradual onset of visual defects with progressive bitemporal field defects, scotoma, or impaired acuity may occur. Compression of surrounding pituitary tissue by an adenoma may disrupt gonadotropin secretion, resulting in hypogonadism. Amenorrhea and infertility occur in women, whereas men present with progressively decreased potency and low testosterone levels. Rarely, excess FSH or LH secretion results in ovarian hyperstimulation or the downregulation of the reproductive axis.

On laboratory testing, circulating gonadotropin α-subunit levels are elevated in about 15% of male patients. TRH administration evokes LH β-subunit levels in most patients of both sexes. The serum prolactin level should be measured: an elevation suggests a prolactinoma; a PRL level below 100 μg/L in a hypogonadal patient harboring a pituitary mass suggests pituitary stalk compression by a nonfunctioning adenoma. In postmenopausal women, physiologic elevations of FSH concentra-

tions may be difficult to distinguish from tumor-derived FSH elevations. Primary ovarian or testicular failure may lead to compensatory gonadotroph cell hyperplasia and uniformly elevated LH and FSH levels.

Treatment Nonfunctioning microadenomas have a benign natural history. They are slow-growing and can safely be followed with annual imaging and visual testing, as long as the patient remains asymptomatic. Nonfunctioning pituitary masses greater than 1 cm in diameter should be resected. These larger masses should be distinguished from nonadenomatous lesions by MRI characteristics and histologic evaluation of resected tissue.[29] After resection, the tissue diagnosis of a clinically nonsecreting gonadotroph adenoma should be confirmed.[30] Visual improvement occurs in 70% of patients with preoperative visual field defects. Hypopituitarism resulting from compression of normal pituitary tissue improves and may resolve completely. Early complications of surgery include diabetes insipidus, inappropriate antidiuretic hormone secretion, or both. Approximately 15% of tumors recur within 5 to 6 years after initially successful surgical resection.[31] Adjuvant pituitary radiotherapy after transsphenoidal surgery has been advocated to prevent future tumor regrowth in patients with residual adenoma tissue.[32]

THYROID-STIMULATING HORMONE

Synthesis

TSH-secreting thyrotroph cells constitute 5% of the anterior pituitary cell population. Hypothalamic TRH stimulates TSH synthesis and secretion. TRH also stimulates lactotroph cells to secrete PRL. Thyroid hormones, dopamine, SRIF, and glucocorticoids suppress TSH and override TRH induction. Thyroid damage, including surgical thyroidectomy, radiation-induced hypothyroidism, chronic thyroiditis, or prolonged goitrogen exposure are associated with reversible thyrotroph hypertrophy and hyperplasia with prominent TSH secretory granules and sellar enlargement. Thyrotroph cells regress with thyroid hormone treatment and hormone-mediated TSH suppression.

TSH Deficiency

Hypothyroidism from TSH deficiency has the same clinical features as primary hypothyroidism [*see 3:I Thyroid*]. On thyroid function testing, however, patients with pituitary hypothyroidism have low levels of both TSH and thyroxine (T_4). Patients with hypothyroidism of hypothalamic origin have normal, low, or slightly elevated TSH levels and low T_4 values.

Testing of TSH reserve is used to confirm the diagnosis of central hypothyroidism. Twofold to threefold increases in TSH levels occur within 30 minutes of an intravenous injection of TRH (200 μg) in normal patients. Primary hypothyroidism is associated with an exaggerated TSH response to TRH because of release of the thyrotroph from negative feedback inhibition. Hyperthyroidism, exogenously administered thyroid hormone, and pituitary damage result in blunted TSH responses to TRH.

Thyrotropin-Secreting Adenomas

TSH-producing pituitary adenomas are very rare. Patients with these tumors usually present with a goiter and mild or frank hyperthyroidism.[33] The diagnosis is made by demonstrating elevated serum T_4 levels, inappropriately high TSH or α-subunit secretion, and evidence of a pituitary adenoma on MRI. These tumors are usually large and locally invasive. Administration of thyroid hormone fails to suppress TSH secretion, wheras administration of TRH evokes a blunted TSH response. In such cases, it is important to exclude thyroid hormone resistance, which can produce abnormalities in TSH and T_4 levels identical to those seen with TSH-producing adenomas.

Treatment TSH-producing pituitary adenomas are debulked surgically. Total resection is often not achieved because most of these tumors are large and locally invasive. Postoperative treatment with a somatostatin analogue controls residual TSH and α-subunit hypersecretion, shrinks the tumor mass in approximately 50% of patients, and improves visual fields in about 75% of patients.

Posterior Pituitary Hormones and Associated Disorders

Vasopressin and oxytocin are stored in the posterior pituitary and released in response to appropriate stimuli. Serum vasopressin levels (and, thus, urinary concentrations) vary in response to changes in serum osmolality. The sensitivity and, to a lesser extent, the threshold of vasopressin response to a change in tonicity show considerable variability from one person to another; at least part of this variation is hereditary. Chronic heart failure lowers the osmotic threshold for vasopressin release, whereas aging and other factors reduce sensitivity of vasopressin release (i.e., the rate of vasopressin release per unit change in osmolality). Shifts in blood volume and pressure of greater than 10% affect vasopressin release significantly. Hypotension and hypovolemia stimulate vasopressin release by lowering the osmotic threshold; hypertension and hypervolemia inhibit release by raising the threshold. These influences are mediated by baroreceptor pathways that have left atrial afferents.

Nausea, but not vomiting, is a powerful stimulus to vasopressin release; it raises the serum vasopressin up to 1,000 times the level required for maximal antidiuresis. Pain, however, is not an important stimulant of vasopressin release. Many neural pathways influence vasopressin release in response to nonosmotic stimuli. In general, alpha-adrenergic pathways stimulate and beta-adrenergic pathways inhibit vasopressin release.

The principal disorders of vasopressin secretion consist of partial or complete deficiency (diabetes insipidus) and the syndrome of inappropriate antidiuretic hormone (SIADH) excess [*see 10:I Renal Function and Disorders of Water and Sodium Balance*].

DIABETES INSIPIDUS

Polyuria is a common clinical problem. A patient passing large quantities of urine generally has one of three abnormalities: an osmotic diuresis (e.g., from glycosuria), resistance to vasopressin, or deficient vasopressin secretion. Resistance to vasopressin (i.e., nephrogenic diabetes insipidus) is discussed elsewhere [*see 10:I Renal Function and Disorders of Water and Sodium Balance*]. Deficiency of vasopressin (i.e., neurogenic diabetes insipidus) reflects either functional or structural disease of the supraoptic hypothalamic neurons that secrete the hormone. Brain tumors, craniopharyngiomas, metastatic cancer, hypothalamic-pituitary surgery or trauma, pituitary stalk damage, histiocytosis, and lymphocytic hypophysitis account for most cases. Rare familial polyuric syndromes may also present as hypothalamic diabetes insipidus.

Diagnosis

Two clinical clues suggest vasopressin deficiency: sudden

onset of polyuria and a preference for iced beverages. However, neurogenic diabetes insipidus must be distinguished from primary polydipsia, because overdrinking also results in polyuria and suppressed vasopressin secretion.

Neurogenic and nephrogenic diabetes insipidus can usually be differentiated by means of clinical testing. After confirmation that the blood glucose level is normal, the patient is deprived of water until 3% to 5% of body weight is lost and the serum tonicity is higher than 295 mOsm/kg. If polyuria disappears and the urine concentration rises above 500 mOsm/kg, vasopressin secretion is adequate. If polyuria and dilute urine (< 300 mOsm/kg) persist, then 20 mg desmopressin acetate (DDAVP), a synthetic vasopressin analogue, is given intranasally; alternatively, 300 μU of DDAVP can be administered intravenously. If urine flow decreases and urine concentration rises, vasopressin deficiency can be inferred. If, however, the serum becomes concentrated and the urine remains dilute despite administration of DDAVP, the patient has nephrogenic diabetes insipidus.

Some cautions should be kept in mind when conducting dehydration tests. First, the term partial diabetes insipidus describes a patient who, when deprived of water, achieves a urine concentration greater than the serum osmolality but less than that obtained after administration of vasopressin. Functional testing can be misleading in patients with neurogenic or nephrogenic partial diabetes insipidus. In such patients, who have a urine concentration between 300 and 500 mOsm/kg, measurement of the serum vasopressin level can be extremely helpful. A high vasopressin level in the presence of concentrated serum and relatively dilute urine points to nephrogenic diabetes insipidus; a low value points to hormone deficiency. Conversely, partial resistance to vasopressin can result from chronic overdrinking, with secondary dilution of the medullary concentration in the kidney. If such patients control their excess water intake, they recover a normal renal medullary concentration and, at the same rate, a normal response to vasopressin. Finally, water deprivation appears to produce less thirst in older men than in younger men. Men older than 80 years must be watched carefully after testing to ensure that they resume appropriate water intake.

Granulomas, trauma, infection, and other infiltrations can all produce diabetes insipidus. Metastatic tumor seldom produces insufficiency in other endocrine glands, but secondary tumors arising from lung, breast, and other organs can all produce insufficiency in the posterior pituitary. The sensitivity of MRI has considerably refined the approach to the diagnosis of diabetes insipidus.

Diabetes insipidus can develop suddenly after neurosurgery or external trauma. Cases that develop after neurosurgery may be marked by a triphasic sequence of vasopressin deficiency, vasopressin excess, and vasopressin deficiency. In postoperative or posttraumatic diabetes insipidus, a dilute polyuria with a serum sodium level greater than 145 mEq/L allows a presumptive diagnosis, and parenteral DDAVP should be given immediately. Conversely, hyponatremia from increased vasopressin secretion after transsphenoidal surgery should also be anticipated by following serum sodium levels. Explosive and fatal central diabetes mellitus and diabetes insipidus have been reported in young women with postoperative hyponatremia that was not aggressively treated. The pathogenesis of the disorder is not understood, but the pathologic sequence included cerebral edema and herniation, compression of the third cranial nerve, hypoxic infarction of the pituitary and hypothalamus, respiratory arrest, and coma. The rapidity of deterioration in these patients indicates that the hyponatremia in such cases should be promptly corrected, even though fixed pupillary dilatation, secondary to compression of the oculomotor nerve, may suggest brain death.

Treatment

There are several approaches to the treatment of diabetes insipidus. If the polyuria is mild and does not interfere with sleep, no treatment may be needed. Chlorpropamide potentiates the effect of vasopressin on renal concentrating ability and can be used to treat partial diabetes insipidus. It is given in a dosage of 250 to 375 mg once a day and usually does not produce hypoglycemia in normal persons. However, if patients do not eat regularly or if they have unsuspected anterior pituitary insufficiency, chlorpropamide can be hazardous.

For patients with severe diabetes insipidus, intranasal or oral DDAVP provides excellent control of polyuria and polydipsia. Intranasal DDAVP is effective, nontoxic, and nonirritating. Tablets of DDAVP are given in a dose of 0.1 or 0.2 mg, taken one to three times daily. All patients with diabetes insipidus should be warned that in circumstances of extreme water loss or unconsciousness, they are exposed to added risk unless they are under the care of a physician who is aware of the diagnosis.

Pituitary Failure

Attenuated pituitary secretory reserve can develop as a result of impingement and compression of an expanding mass on adjacent functioning pituitary cells or because of acquired or inherited pituitary cell damage.[34] Tropic hormone failure associated with pituitary compression or destruction usually occurs sequentially, with GH; then FSH, LH, and TSH; and finally ACTH. In childhood, growth retardation is often the presenting feature; in adults, hypogonadism is the earliest symptom. Pressure effects may impair synthesis or secretion of hypothalamic hormones, with pituitary failure [*see Table 10*].

DEVELOPMENTAL PITUITARY DYSFUNCTION

Developmental pituitary dysfunction occurs with aplastic, hypoplastic, or ectopic pituitary gland development. Midline craniofacial disorders may be associated with structural pituitary dysplasia. Birth trauma—including cranial hemorrhage, asphyxia, and breech delivery—can cause acquired pituitary failure in the newborn.

Transcription Factor Mutations

Tissue-specific transcription factors, including Pit-1 and PROP-1, determine tissue-specific development and expression of pituitary hormones and are critical for maintaining anterior pituitary cell function. Hereditary transcription factor mutations may result in disruption of pituitary function, which may manifest during infancy, childhood, puberty, or early adulthood. Autosomal dominant or recessive Pit-1 mutations result in combined deficiency of GH, PRL, and TSH. Pituitary imaging may reveal a normal or hypoplastic gland. PROP-1 is an early transcription factor that appears to be necessary for Pit-1 function. PROP-1 mutations result in combined deficiency of GH, TSH, gonadotropins, and sometimes ACTH. Most afflicted patients have growth retardation and do not enter puberty spontaneously. By adulthood, most patients are deficient in TSH and gonadotropins, and some have an enlarged pituitary gland. *T-Pit* mutations result in isolated ACTH deficiency.

Dysgenesis of the septum pellucidum or corpus callosum

Table 10 Causes of Pituitary Failure[39]

Development/ structural	Transcription factor defect Pituitary dysplasia/aplasia Congenital central nervous system masses, encephalocele Primary empty sella Congenital hypothalamic disorders (e.g., septo-optic dysplasia, Prader-Willi syndrome, Laurence-Moon-Biedl syndrome, Kallmann syndrome)
Traumatic	Surgical resection Radiation damage Accidental
Neoplastic	Pituitary adenoma Parasellar mass (meningioma, germinoma, ependymoma, glioma) Rathke cyst Craniopharyngioma Hypothalamic hamartoma, gangliocytoma Pituitary metastases Lymphoma and leukemia Meningioma
Infiltrative/ inflammatory	Lymphocytic hypophysitis Sarcoidosis Histiocytosis X Hemochromatosis Granulomatous hypophysitis
Vascular	Pituitary apoplexy Pregnancy-related infarction Sickle cell disease Arteritis
Infections	Fungal (histoplasmosis) Parasitic (toxoplasmosis) Tuberculosis *Pneumocystis carinii*

may lead to hypothalamic dysfunction and hypopituitarism, with manifestations that include diabetes insipidus, GH deficiency, and, occasionally, TSH deficiency. Affected children harbor a mutation in the *HESX-1* gene. Clinical features include cleft palate, syndactyly, ear deformities, and hypertelorism.

ACQUIRED PITUITARY FAILURE

Rarely, infiltration of the hypothalamus by diseases such as sarcoidosis, histiocytosis X, amyloidosis, or hemochromatosis may disrupt hypothalamic and pituitary function.[35] This hypothalamic infiltration may result in diabetes insipidus and, if GH attenuation occurs before pubertal bone closure, in growth retardation. Hypogonadotropic hypogonadism and, rarely, hyperprolactinemia occur with disrupted gonadotropin secretion. Pituitary damage may be directly caused by accidental or neurosurgical trauma; pituitary or hypothalamic neoplasms, including pituitary adenomas, craniopharyngioma, Rathke cysts, chordomas, or metastatic deposits; inflammatory disease, such as lymphocytic hypophysitis; or pituitary irradiation. Tuberculosis, opportunistic fungal infections associated with HIV infection, and tertiary syphilis may destroy pituitary tissue.

Cranial Irradiation

Cranial irradiation results in long-term compromise of hypothalamic and pituitary function. Children and adolescents who have undergone therapeutic irradiation of the brain or head and neck are at especially high risk. The resulting hormonal abnormalities correlate strongly with radiation dosage, as well as the time elapsed since completion of radiotherapy. By 10 to 15 years after therapy, GH deficiency invariably occurs, whereas central hypogonadism and ACTH deficiency less commonly occur. Anterior pituitary function should be tested in previously irradiated patients, and replacement therapy instituted when required.

Lymphocytic Hypophysitis

Lymphocytic hypophysitis usually occurs in pregnant or postpartum women. It presents as hyperprolactinemia and a pituitary mass resembling an adenoma on MRI; PRL levels are often mildly elevated.[36] Transient pituitary failure and symptoms of progressive sellar compression, such as headache and visual disturbance, may occur, and the erythrocyte sedimentation rate may be elevated. Because its appearance on MRI may be indistinguishable from that of a pituitary adenoma, lymphocytic hypophysitis should be excluded in a postpartum woman with a newly diagnosed pituitary mass. Pituitary surgery is unnecessary in such cases: glucocorticoid treatment usually restores pituitary function within 6 months, and the mass invariably resolves.

Pituitary Apoplexy

Acute intrapituitary hemorrhage may result in catastrophic vascular compression of parasellar structures.[37] Hemorrhage may occur in a preexisting adenoma, often in association with diabetes or hypertension, or in the postpartum period (Sheehan syndrome). During pregnancy, swelling of the pituitary increases the risk of intrapituitary hemorrhage and infarction. Hypoglycemia, hypotension, shock, apoplexy, and death may follow. Severe headache with signs of meningeal irritation, visual loss, dynamically changing ophthalmoplegia, cardiovascular collapse, and loss of consciousness portend acutely progressive intrasellar bleeding. Pituitary imaging may reveal signs of intratumoral or sellar hemorrhage, with deviation of vital structures, including compression of noninvolved pituitary tissue. If vision is intact and consciousness is unimpaired, patients can be treated conservatively with observation and high-dose steroid infusions. Visual loss or decreased consciousness is an indication for urgent surgical decompression. Subsequent pituitary hormone replacement will be required for the inevitable pituitary damage.

Empty Sella Syndrome

Clinically silent pituitary mass infarction may result in development of a partially or totally empty sella. CSF fills the dural herniation. Pituitary function often remains intact, because the surrounding tissue is fully functional. Hypopituitarism may develop insidiously, however. A partially or apparently totally empty sella is usually an incidental MRI finding. Rarely, small functional pituitary adenomas may arise within the rim.

DIAGNOSIS

Pituitary failure is characterized by the clinical impact of single or multiple tropic hormone loss. Growth disorders and abnormal body composition result from GH loss in children and adults, respectively; menstrual disorders and infertility in women and decreased sexual function, infertility, and loss of secondary sexual characteristics in men are caused by gonadotropin deficits; hypothyroidism is caused by TSH loss; hypocortisolism with hypoglycemia is caused by ACTH loss; and failed lactation is caused by PRL loss. Polyuria and poly-

Table 11 Replacement Therapy for Hypopituitarism in Adults

Tropic Hormone Deficit	*Hormone Replacement*
ACTH	Hydrocortisone, 10–15 mg q. A.M., 5 mg q. P.M. Cortisone acetate, 25 mg q. A.M., 12.5 mg q. P.M.
TSH	Levothyroxine, 0.075–0.15 mg daily
FSH/LH	Males Testosterone enanthate, 200 mg I.M. q. 2 wk Testosterone skin patch, 5–7.5 mg/day Females Conjugated estrogen, 0.65–1.25 mg daily for 25 days Ethinyl estradiol, 0.02–0.05 mg Progesterone on days 16–25 to facilitate uterine shedding Estradiol skin patch, 4–8 mg, twice weekly
GH	Somatotropin, 0.15–1.0 mg S.C. daily
Vasopressin	Desmopressin Intranasal: 5–20 μg, b.i.d. Oral: 300–600 μg, q.d.

Note: Doses should be individualized and should be reassessed during stress, surgery, or pregnancy. Treatment for infertility (gonadotropins or gonadotropin-releasing hormone) should be individualized.
ACTH—adrenocorticotropic hormone FSH—follicle-stimulating hormone GH—growth hormone LH—luteinizing hormone TSH—thyroid-stimulating hormone

dipsia reflect loss of ADH secretion. These features may occur selectively or may be sequential and ultimately result in panhypopituitarism. Enhanced mortality in patients with long-standing pituitary damage is caused mainly by increased cardiovascular and cerebrovascular disease.[38]

On laboratory tests, patients with pituitary insufficiency demonstrate lack of normal hormonal feedback responses, with low tropic hormone levels in conjunction with low target hormone concentrations. Provocative tests confirm lack of pituitary hormone reserve.

TREATMENT

Replacement of pituitary hormones or their respective target hormones usually results in clinical homeostasis with few side effects [*see Table 11*]. Hormone replacement therapy for pituitary failure includes glucocorticoids, thyroid hormone, sex steroids, GH, and vasopressin. Rational replacement regimens ensure a normal and safe quality of life. Patients receiving glucocorticoid replacement require dose increases during stressful events, including dental procedures, trauma, and hospitalizations for acute illness.

The author has received research support and been a scientific consultant for Eli Lilly and Co. and Novartis Pharmaceuticals Corp. for the past 12 months.

References

1. Prezant TR, Melmed S: Molecular pathogenesis of pituitary disorders. Curr Opin Endocrinol Diabetes 9:61, 2002

2. Faglia G, Spada A: Genesis of pituitary adenomas: state of the art. J Neurooncol 54:95, 2001

3. Goffin V, Binart N, Touraine P, et al: Prolactin: the new biology of an old hormone. Annu Rev Physiol 64:47, 2002

4. Verhelst J, Abs R, Maiter D, et al: Cabergoline in the treatment of hyperprolactinemia: a study in 455 patients. J Clin Endocrinol Metab 84:2518, 1999

5. Losa M, Mortini P, Barzaghi R, et al: Surgical treatment of prolactin-secreting pituitary adenomas: early results and long-term outcome. J Clin Endocrinol Metab 87:3180, 2002

6. Colao A, di Sarno A, Pivonello R, et al: Dopamine receptor agonists for treating prolactinomas. Expert Opin Investig Drugs 11:787, 2002

7. Cunha SR, Mayo KE: Ghrelin and growth hormone (GH) secretagogues potentiate GH-releasing hormone (GHRH)-induced cyclic adenosine 3′,5′-monophosphate production in cells expressing transfected GHRH and GH secretagogue receptors. Endocrinology 143:4570, 2002

8. Giustina A, Veldhuis JD: Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. Endocr Rev 19:717, 1998

9. Biller BM, Samuels MH, Zagar A, et al: Sensitivity and specificity of six tests for the diagnosis of adult GH deficiency. J Clin Endocrinol Metab 87:2067, 2002

10. Herrington J, Carter-Su C: Signaling pathways activated by the growth hormone receptor. Trends Endocrinol Metab 12:252, 2001

11. Colao A, Cuocolo A, Di Somma C, et al: Impaired cardiac performance in elderly patients with growth hormone deficiency. J Clin Endocrinol Metab 84:3950, 1999

12. Shalet SM, Toogood A, Rahim A, et al: The diagnosis of growth hormone deficiency in children and adults. Endocr Rev 19:203, 1998

13. Hartman ML, Crowe BJ, Biller BM, et al: Which patients do not require a GH stimulation test for the diagnosis of adult GH deficiency? J Clin Endocrinol Metab 87:477, 2002

14. Gibney J, Wallace JD, Spinks T, et al: The effects of 10 years of recombinant human growth hormone (GH) in adult GH-deficient patients. J Clin Endocrinol Metab 84:2596, 1999

15. Carroll PV, Christ ER, Sonksen PH: Growth hormone replacement in adults with growth hormone deficiency: assessment of current knowledge. Trends Endocrinol Metab 11:231, 2000

16. Attanasio AF, Bates PC, Ho KK, et al: Human growth hormone replacement in adult hypopituitary patients: long-term effects on body composition and lipid status—3-year results from the HypoCCS Database. J Clin Endocrinol Metab 87:1600, 2002

17. Friend KE: Acromegaly: a new therapy. Cancer Control 9:232, 2002

18. Melmed S, Casanueva FF, Cavagnini F, et al: Guidelines for acromegaly management. J Clin Endocrinol Metab 87:4054, 2002

19. Lamberts SW, Reubi JC, Krenning EP: Somatostatin analogs in the treatment of acromegaly. Endocrinol Metab Clin North Am 21:737, 1992

20. Colao A, Ferone D, Marzullo P, et al: Long-term effects of depot long-acting somatostatin analog octreotide on hormone levels and tumor mass in acromegaly. J Clin Endocrinol Metab 86:2779, 2001

21. Caron P, Beckers A, Cullen DR, et al: Efficacy of the new long-acting formulation of lanreotide (lanreotide Autogel) in the management of acromegaly. J Clin Endocrinol Metab 87:99, 2002

22. Davies PH, Stewart SE, Lancranjan L, et al: Long-term therapy with long-acting octreotide (Sandostatin-LAR) for the management of acromegaly. Clin Endocrinol (Oxf) 48:311, 1998

23. Chon BH, Loeffler JS: Efficacy and risk for radiotherapy for pituitary tumors. Endocrinologist 12:525, 2002

24. Melmed S, Jackson I, Kleinberg D, et al: Current treatment guidelines for acromegaly. J Clin Endocrinol Metab 83:2646, 2002

25. Ruff H, Kindling F: A physiologic approach to diagnosis of the Cushing syndrome. Ann Intern Med 138:980, 2003

26. Nieman LK: Diagnostic tests for Cushing's syndrome. Ann N Y Acad Sci 970:112, 2002

27. Swearingen B, Biller BM, Barker FG II, et al: Long-term mortality after transsphenoidal surgery for Cushing disease. Ann Intern Med 130:821, 1999

28. Greenman Y, Melmed S: Diagnosis and management of nonfunctioning pituitary tumors. Annu Rev Med 47:95, 1996

29. King JT Jr, Justice AC, Aron DC: Management of incidental pituitary microadenomas: a cost-effectiveness analysis. J Clin Endocrinol Metab 82:3625, 1997

30. Wilson CB: Surgical management of pituitary tumors. J Clin Endocrinol Metab 82:2381, 1997

31. Cappabianca P, Cavallo LM, Colao A: Surgical complications associated with the endoscopic endonasal transsphenoidal approach for pituitary adenomas. J Neurosurg 97:293, 2002

32. Gittoes NJ, Bates AS, Tse W, et al: Radiotherapy for non-function pituitary tumours. Clin Endocrinol (Oxf) 48:331, 1998

33. Beck-Peccoz P, Brucker-Davis F, Persani L, et al: Thyrotropin-secreting pituitary tumors. Endocr Rev 17:610, 1996

34. Vance ML: Hypopituitarism. N Engl J Med 330:1651, 1994

35. Saeger W: Tumor-like lesions of the pituitary and sellar region. Endocrinologist 12:300, 2002

36. Cheung CC, Ezzat S, Smyth HS, et al: The spectrum and significance of primary hypophysitis. J Clin Endocrinol Metab 86:1048, 2001

37. Randeva HS, Schoebel J, Byrne J, et al: Classical pituitary apoplexy: clinical features, management and outcome. Clin Endocrinol (Oxf) 51:181, 1999

38. Association between premature mortality and hypopituitarism. West Midlands Prospective Hypopituitary Study Group. Lancet 357:425, 2001

39. Tolis G: Prolactin: physiology and pathology. Hosp Pract 15:85, 1980

40. Melmed S: Disorders of the anterior pituitary and hypothalamus. Harrison's Textbook of Medicine, 15th ed, Braunwald E, Fauci AS, Kasper DL, et al, Eds. McGraw-Hill, New York, 2001, p 2029

Acknowledgment

Figures 1 and 2 Alice Y. Chen.

VI DISEASES OF CALCIUM METABOLISM AND METABOLIC BONE DISEASE

ELIZABETH H. HOLT, M.D., PH.D.
SILVIO E. INZUCCHI, M.D.

Calcium Metabolism

The precise regulation of body calcium stores and of the calcium concentration in both extracellular and intracellular compartments is critically important, for the following reasons: calcium is the chief mineral component of the skeleton; calcium serves major roles in neurologic transmission, muscle contraction, and blood coagulation; and it is a ubiquitous intracellular signal. A typical laboratory range for serum calcium concentration is between 8.8 and 10.5 mg/dl; 50% to 60% of the calcium in the blood is bound to plasma proteins or is complexed with citrate and phosphate. The remaining ionized (free) calcium controls physiologic actions. The body regulates not only ionized calcium concentrations but also the entry and exit of calcium into its main storage site, bone, through the activity of parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D_3 (1,25-$(OH)_2D_3$) [*see Figure 1*]. PTH, secreted by the parathyroid glands, is an 84–amino acid peptide with a very short plasma half-life (2 to 4 minutes). Cholecalciferol (vitamin D_3) is generated by the skin, upon exposure to ultraviolet light; it is also supplied by dietary sources (chiefly fortified liquid milk products). In the liver, vitamin D_3 is hydroxylated to 25-$(OH)D_3$, which is in turn hydroxylated in the kidney to 1,25-$(OH)_2D_3$ (calcitriol), markedly increasing its potency. In concert, this hormonal system expresses its action at the level of the gastrointestinal tract, bone, and the kidney and maintains circulating ionized calcium concentrations under extremely tight control (variation < 0.1 mg/dl), despite significant variations in calcium supply.

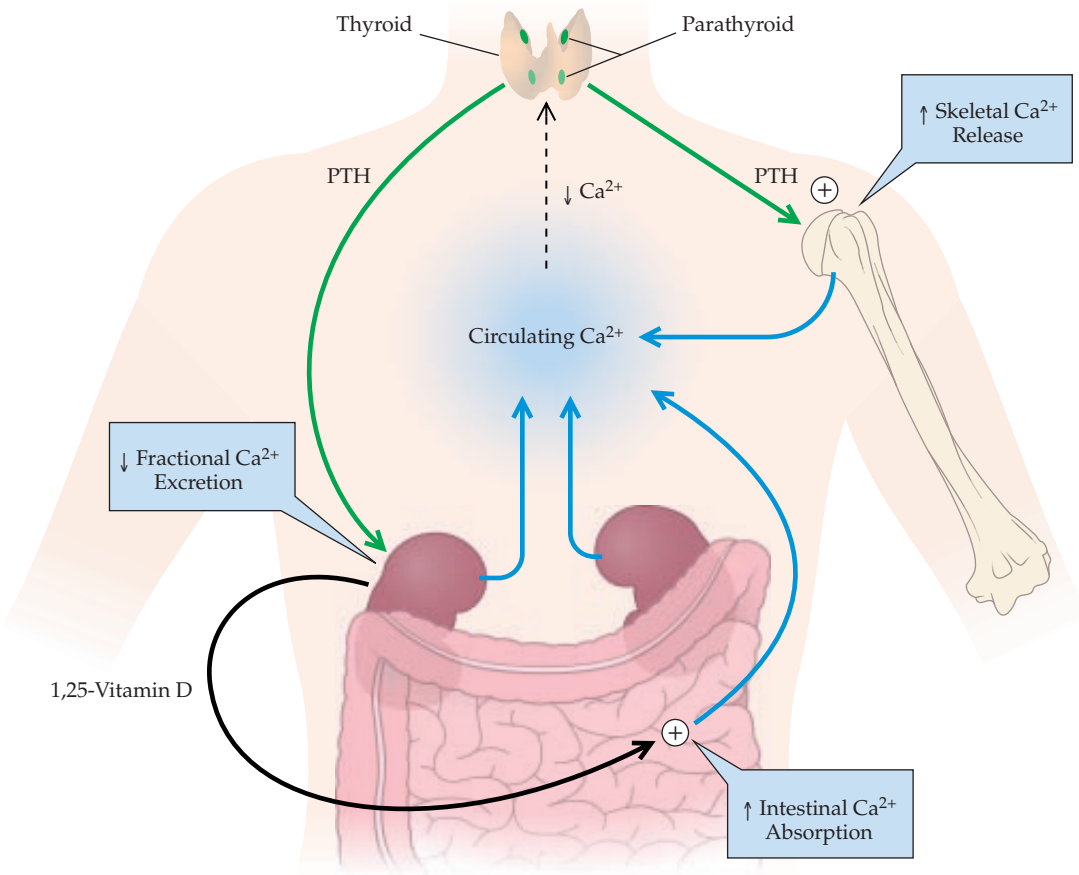


Figure 1 **Circulating concentrations of ionized calcium are maintained under extremely tight control by parathyroid hormone (PTH) and the vitamin D axis. Absorption of dietary calcium by the gastrointestinal tract, reduction of calcium excretion by the kidneys, and release of stored calcium from bones serve as sources for circulating calcium. Decreases in circulating calcium trigger the release of PTH, which promotes release of calcium into the extracellular space by increasing bone resorption; the release of PTH also causes an increase in calcium reabsorption in the distal nephron, resulting in a decrease in urinary calcium loss. PTH also augments renal production of 1,25-dihydroxyvitamin D, which secondarily increases calcium absorption in the gut.**

Under normal conditions, despite ranges in dietary calcium consumption that can vary from 400 to 2,000 mg daily, net calcium absorption from the GI tract averages about 150 to 200 mg/day. In steady state, this equals the amount of calcium excreted by the kidneys. Ongoing remodeling of bone results in the consumption and release of approximately 500 mg of calcium a day. Through humoral regulation, this calcium reservoir can be exploited to maintain extracellular calcium levels in a narrow range despite increased physiologic need or decreased intake, such as results from severe curtailment of the dietary calcium supply or from impairment of intestinal calcium absorption.

Changes in the extracellular ionized calcium concentration are registered by parathyroid cells via the cell surface calcium-sensing receptor (CaSR).[1] Interaction of calcium ions with the extracellular domain of the CaSR triggers a series of intracellular signaling events, which ultimately govern PTH secretion. As circulating concentrations of calcium fall, PTH secretion rises, and vice versa.

PTH increases bone resorption and distal nephron calcium reabsorption, the former promoting calcium release into the extracellular space and the latter decreasing urinary calcium losses. PTH also augments renal production of calcitriol, which then increases fractional calcium absorption in the gut. If calcium intake increases beyond the body's needs, PTH secretion decreases, leading to decreased calcitriol production and decreased calcium absorption by the gut. If calcium is absorbed in excess of requirements, it will be promptly excreted. In this elegant manner, circulating ionized calcium concentration is guarded closely, albeit sometimes at the expense of skeletal calcium stores. Disturbances of PTH, vitamin D action, or both are most often manifested by altered serum calcium or phosphate concentration and by abnormal bone turnover. In some cases, bone mineral density (BMD) is decreased.

Measurement of Calcium

Diagnosis of a calcium disorder depends first on accurate measurement of serum or ionized calcium or both. Serum measurements are usually performed by spectrophotometry or by atomic absorption spectrophotometry, which yields more accurate measurements. Spurious readings may occur with tourniquet stasis (i.e., if the tourniquet is in place too long before the

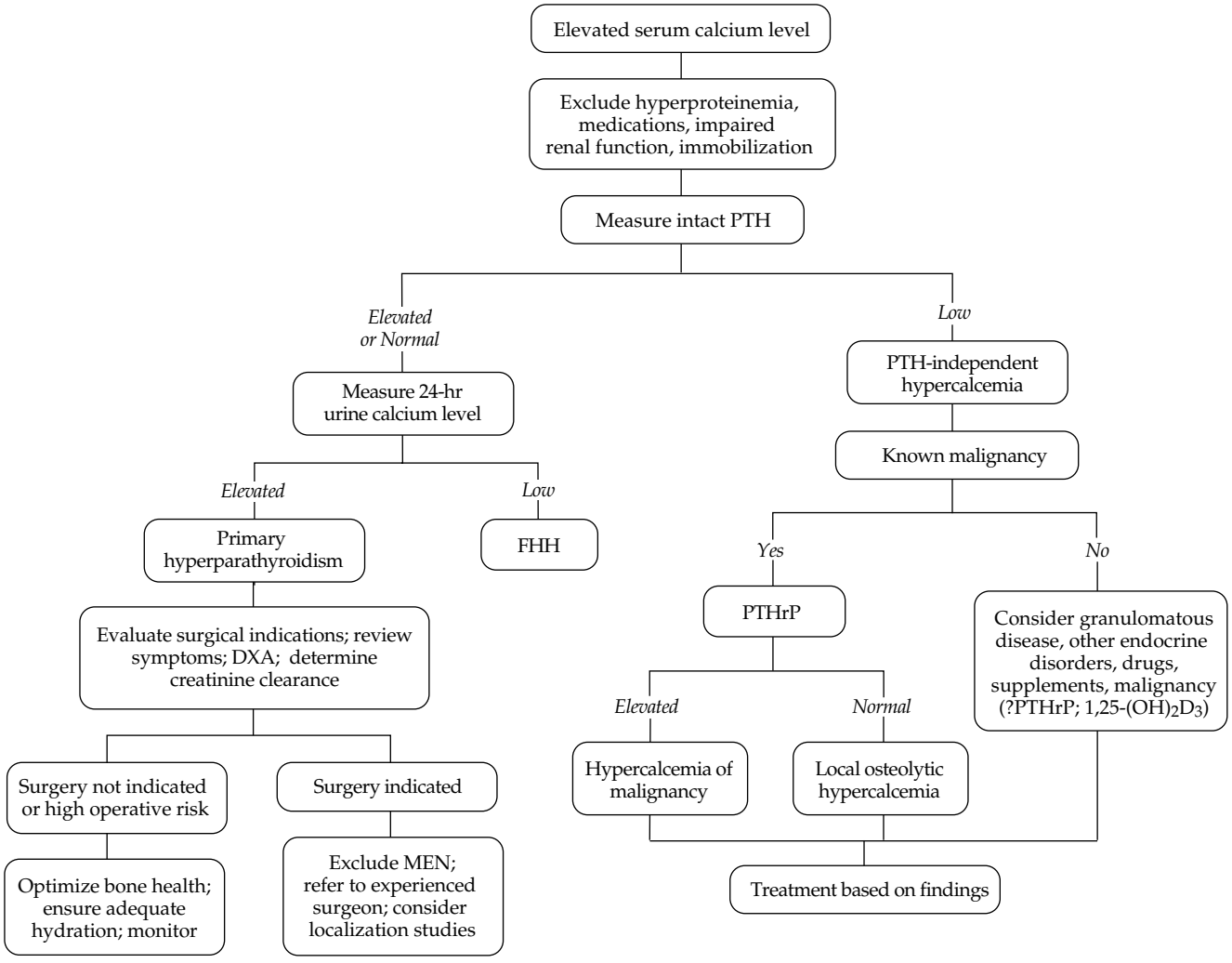


Figure 2 **Evaluation and management of hypercalcemia. (DXA—dual-energy x-ray absorptiometry; FHH—familial hypocalciuric hypercalcemia; MEN—multiple endocrine neoplasia; PTH—parathyroid hormone; PTHrP— parathyroid hormone–related protein)**

Table 1 Differential Diagnosis of Hypercalcemia

Parathyroid hormone–mediated hypercalcemia	Primary hyperparathyroidism Parathyroid adenoma Parathyroid hyperplasia Parathyroid carcinoma Tertiary hyperparathyroidism
Parathyroid hormone–independent hypercalcemia	Humoral hypercalcemia of malignancy Parathyroid hormone–related protein mediated Squamous cell carcinoma of the lung Carcinoma of the oropharynx, nasopharynx, larynx, and esophagus Cervical carcinoma Ovarian carcinoma Renal cell carcinoma Transitional cell carcinoma of the bladder Pheochromocytoma Islet cell neoplasms of the pancreas T cell lymphoma Others 1,25-$(OH)_2$ D_3 mediated B cell lymphoma Local osteolytic hypercalcemia Multiple myeloma Breast carcinoma Lymphoma Others Medications/supplements Vitamin D Vitamin A Lithium Thiazides Calcium antacids (milk-alkali syndrome) Granulomatous diseases Sarcoidosis Tuberculosis Histoplasmosis Leprosy Other conditions Increased plasma protein levels (factitious hypercalcemia) Acute renal failure Thyrotoxicosis Adrenal insufficiency Immobilization Familial hypocalciuric hypocalcemia (benign familial hypercalcemia)

blood is drawn), which can elevate serum calcium values by up to 1 mg/dl. Dilution of blood by drawing samples from indwelling intravenous catheters is a common error that leads to spuriously low calcium readings. Ionized calcium measurements should be considered accurate only when performed on samples collected anaerobically (i.e., in a blood gas syringe) and placed on ice, with immediate analysis.

Hypercalcemia

Hypercalcemia is a common metabolic abnormality. Signs and symptoms of hypercalcemia vary significantly from patient to patient and correlate somewhat with the degree of calcium elevation and its rate of change. The diagnostic workup of hypercalcemia is straightforward [*see Figure 2*].[2] The etiology of hypercalcemia is usually discovered after a comprehensive history, physical examination, focused laboratory assessment, and, occasionally, diagnostic imaging studies.[3]

CLINICAL MANIFESTATIONS

Most patients with mild hypercalcemia (serum calcium level < 11 mg/dl) are asymptomatic, although some may experience mild fatigue, vague changes in cognitive function, depression, or polyuria and polydipsia (from decreased urine concentrating ability caused by a high calcium level). Those with moderate hypercalcemia (serum calcium levels of 11 to 14 mg/dl) are more likely to be symptomatic. The likelihood of classic manifestations of hypercalcemia increases sharply when calcium levels rise to 12 to 14 mg/dl. These symptoms include anorexia, nausea, vomiting, abdominal pain, constipation, muscle weakness, and altered mental status. Severe hypercalcemia (i.e., serum calcium levels greater than 14 mg/dl) may cause progressive lethargy, disorientation, and even coma.

In addition to the degree of elevation, the rate of increase in serum calcium may influence the clinical picture. Chronically hypercalcemic patients can function and feel reasonably well with serum calcium values even as high as 15 to 16 mg/dl. In contrast, patients whose calcium level has risen abruptly will often experience symptoms at lesser calcium elevations. Elderly or debilitated patients are also more likely to be symptomatic.

HISTORY AND PHYSICAL EXAMINATION

The history and physical examination are directed at uncovering signs or symptoms of hypercalcemia, as well as signs of the most common causes of hypercalcemia: hyperparathyroidism, malignancy, granulomatous diseases, and certain endocrinopathies. Evidence of any related condition, such as osteoporosis or urinary tract stones, should also be sought. The medical record should be reviewed to determine the duration of the hypercalcemia. The most common cause of hypercalcemia, primary hyperparathyroidism, presents as stable or gradually progressive elevation of the serum calcium level over a period of years. In contrast, malignancy typically causes a more acute rise in serum calcium. All recent medications, foods, and nutritional supplements should be thoroughly reviewed for possible culprits. A careful family history should be performed to identify disorders of calcium metabolism; renal stones; fragility fractures; and any related endocrinopathies, such as diseases of the pituitary, adrenal, thyroid, or endocrine pancreas.

Aside from mental status deficits and signs of dehydration, physical examination findings are generally normal in patients with hypercalcemia, especially if calcium levels are only mildly to moderately elevated. Rarely, severe and prolonged hypercalcemia results in a visible horizontal calcium deposit on the cornea, a condition known as band keratopathy. Other signs and symptoms depend on the etiology of the elevation [*see Table 1*]. Patients with hyperparathyroidism classically have osteopenia, bone pain, or nephrolithiasis. Currently, however, most cases of primary hyperparathyroidism are identified before the patient becomes symptomatic. Patients whose hyperparathyroidism is associated with multiple endocrine neoplasia (MEN) syndromes may have specific manifestations of the other conditions that are part of these syndromes. Patients with sarcoidosis may present with fever, lymphadenopathy, skin rashes, or pulmonary symptoms. Hypercalcemia of malignancy develops only when a substantial tumor burden is present; consequently, most of these patients have an established cancer diagnosis and

clinical features associated with the specific tumor type and extent of disease.

LABORATORY STUDIES

The first step in the laboratory assessment is to exclude factitious hypercalcemia, which may result from an increase in circulating concentrations of plasma proteins. About 50% to 60% of circulating calcium is bound to these proteins, so elevation in their concentrations (as occurs in HIV infection, chronic viral hepatitis, and multiple myeloma) will produce a proportionate rise in the total calcium concentration. The ionized calcium concentration, however, remains normal. To adjust for elevations in plasma protein, the serum calcium level should be lowered by 0.8 mg/dl for every 1 g/dl of albumin (or protein) above the normal range. When performed correctly, ionized calcium measurement is more accurate than adjusted total calcium. Because acute renal failure may occasionally lead to hypercalcemia, renal function should also be assessed.

Once hypercalcemia is confirmed, the next step is measurement of the serum PTH concentration. This is the most important test for determining the cause of hypercalcemia.[3] Several PTH assays are commercially available. The most commonly utilized is the two-site immunochemiluminometric assay (ICMA, or so-called bio-intact PTH). Earlier assays could not distinguish between full-length PTH and inactive molecular fragments that circulate in significant concentrations. The ICMA measures only the intact PTH molecule and is therefore the preferred test in most instances, especially in patients whose serum creatinine level is elevated.

Other helpful tests include measurement of serum creatinine and alkaline phosphatase, as well as inorganic phosphorus assays and an electrolyte panel. Assessment of 24-hour urinary calcium excretion is usually performed. Serum creatinine may be elevated in patients with nephrocalcinosis secondary to prolonged hypercalcemia. The alkaline phosphatase level may be elevated in patients with hypercalcemic states involving increased bone turnover. Patients with hypercalcemia caused by malignancy may demonstrate biochemical or hematologic findings consistent with the site of neoplasia and the degree of its dissemination. Most causes of hypercalcemia are also accompanied by hypercalciuria (24-hour urinary calcium excretion > 4 mg/kg/day), which may lead to nephrocalcinosis or renal stone formation. A serum calcium × phosphate product greater than 70 suggests the patient is at risk for calciphylaxis, and efforts to lower the serum phosphate level (e.g., with phosphate binders) should accompany the interventions to lower serum calcium.

Other diagnostic studies may be dictated by clinical circumstances. Electrocardiographic abnormalities of severe hypercalcemia include shortening of the QT_c interval and, rarely, atrioventricular blocks. In addition, many hypercalcemic conditions cause a decrease in BMD, which may be noted on plain x-rays but is best quantified by measurement of bone density (see below). Abdominal x-rays may identify renal stones or nephrocalcinosis. Specific bone radiographic findings are few, and in primary hyperaparathyroidism, specific bony abnormalities are now rare, thanks to early detection of hypercalcemia.

DIFFERENTIAL DIAGNOSIS

The results of PTH measurement indicate whether hypercalcemia is or is not mediated by PTH and thus provide a broad indication of the cause of hypercalcemia [*see Table 1*]. When PTH levels are high or, in some cases, inappropriately normal, the hypercalcemia is PTH mediated; this is commonly referred to as hyperparathyroidism. When PTH levels are suppressed, the hypercalcemia is said to be non–PTH mediated, or PTH independent. In turn, this distinction guides subsequent patient assessment.

HYPERPARATHYROIDISM (PTH-MEDIATED HYPERCALCEMIA)

Classification

Primary hyperparathyroidism Primary hyperparathyroidism is the most common cause of hypercalcemia in outpatients. Current estimates place the annual incidence at approximately four per 100,000 population; the incidence peaks in the fifth to sixth decade of life, and there is a female-to-male ratio of 3:2.[4] The most common clinical presentation is that of asymptomatic mild hypercalcemia. Pathologically, a solitary parathyroid adenoma is present in 80% to 85% of cases; hyperplasia involving multiple glands is found in 15% to 20% of cases, and parathyroid carcinoma is found in less than 1%. Occasionally, double adenomas are found. Patients with type I MEN (MEN I) or MEN II usually have parathyroid hyperplasia.[5]

Lithium therapy can change the set point for the calcium-sensing receptor such that a higher serum calcium concentration is needed to inhibit PTH secretion. This can lead to biochemical abnormalities (e.g., high levels of calcium and high-normal to elevated PTH levels) that mimic primary hyperparathyroidism. Patients on lithium will often have very low urinary calcium excretion.

Secondary hyperparathyroidism Conditions that tend to decrease serum calcium increase PTH secretion as a corrective measure. This increase of PTH secretion is termed secondary hyperparathyroidism. Once circulating PTH is elevated, the serum calcium may return to normal or remain low. Common causes of secondary hyperparathyroidism include chronic renal insufficiency, vitamin D deficiency, intestinal malabsorption, renal calcium losses, and severe dietary inadequacy. Correction of the underlying calcium abnormality will return serum PTH concentrations to normal.

Familial hypocalciuric hypercalcemia Familial hypocalciuric hypercalcemia (FHH), also referred to as benign familial hypercalcemia, is a rare inherited condition caused by various inactivating mutations in the CaSR. This results in inappropriately increased PTH secretion and a higher set point for the extracellular ionized calcium concentration. Patients with FHH have chronic asymptomatic hypercalcemia associated with relatively depressed urinary calcium excretion.

Tertiary hyperparathyroidism In some patients with prolonged secondary hyperparathyroidism, hyperplasia or neoplasia of the parathyroid glands develops. These parathyroids no longer respond appropriately to serum calcium; instead, they produce excess PTH at all times, leading to chronic hypercalcemia. This is most often seen in patients with chronic kidney disease. More than one parathyroid gland is usually affected.

Diagnosis

Clinical manifestations The clinical manifestations of hyperparathyroidism depend, in part, on the severity of the hypercalcemia. When hyperparathyroidism was first described more than 50 years ago, most patients presented with late-stage com-

plications of prolonged and severe hypercalcemia, such as abnormalities of bone (osteitis fibrosa cystica)[6] or kidneys (nephrocalcinosis, renal failure). Since the development more than 30 years ago of laboratory equipment for measuring serum chemistry, hyperparathyroidism is often diagnosed by routine blood testing, before the development of symptoms. It also may be uncovered during the evaluation of osteoporosis or during the workup of renal stone disease.

When symptomatic, patients with hyperparathyroidism demonstrate clinical manifestations of hypercalcemia (see above).

Physical examination In general, parathyroid tumors are too small to be palpable. Indeed, a palpable parathyroid tumor should be suspected as a malignancy until proved otherwise. Evidence of the consequences of hyperparathyroidism should be sought, such as osteoporosis (kyphosis) or nephrolithiasis (costovertebral angle tenderness).

Laboratory tests Currently, most patients with hyperparathyroidism have a serum calcium concentration of less than 12 mg/dl (unless coexisting volume contraction is present), and they may have mild to moderate hypophosphatemia and a non–anion gap metabolic acidosis (from renal tubular acidosis). Urinary calcium excretion is often increased; in these patients, the reduction of fractional calcium excretion by PTH is overcome by the high filtered calcium load. This may result in nephrocalcinosis or nephrolithiasis.

Renal stones in patients with hyperparathyroidism are usually composed of calcium oxalate and tend to occur bilaterally, especially when urinary calcium excretion is high. Rarely, nephrocalcinosis and azotemia develop, usually in those with the most severe and protracted hypercalcemia, especially if dehydration or other renal insult is superimposed. Because PTH increases both osteoclast and osteoblast activity, there are increases in serum and urinary concentrations of biochemical markers of bone turnover, including bone alkaline phosphatase.

Elevation of both the serum calcium and the PTH concentrations (in the absence of low urinary calcium excretion) supports a diagnosis of primary hyperparathyroidism. PTH levels are usually increased to less than five times the upper limit of normal. In certain mild cases, the calcium level is only slightly high, and the PTH is minimally elevated or inappropriately normal. Rarely, patients with primary hyperparathyroidism have serum calcium levels in only the high-normal range. In fact, most such patients have elevated serum ionized calcium values and therefore are not actually normocalcemic. The diagnosis in such patients can be extremely challenging.

When the PTH level is normal or mildly elevated and the 24-hour urinary calcium level is low, consideration should be given to the possibility of FHH.[7] The relatively low urinary calcium output seen in FHH may help distinguish this condition from primary hyperparathyroidism, although low urinary calcium excretion may also occur in hyperparathyroidism.

The possibility of FHH is raised when there is a strong family history of symptomatic, stable hypercalcemia, especially in patients younger than 40 years; when family members have undergone unsuccessful parathyroid surgery; or when the patient's urinary calcium output is unexpectedly low. When this diagnosis is suspected, further evaluation is necessary, such as the screening of other family members. Unfortunately, specific genetic testing is not currently widely available from commercial laboratories. In some cases, FHH cannot be distinguished confidently from primary hyperparathyroidism. However, in most such patients, expectant management is safe and avoids unnecessary parathyroid exploration. When there is trouble distinguishing between primary hyperparathyroidism and FHH, parathyroid imaging is sometimes useful. In primary hyperparathyroidism, enlarged parathyroid glands are easily found, whereas parathyroid size is usually normal in FHH.

Once the diagnosis of primary hyperparathyroidism is secured, it will usually already be apparent whether the patient is a candidate for parathyroidectomy. If the patient does not meet criteria for surgery on the basis of age, renal function, urinary calcium excretion, history of fractures, or renal stones/nephrocalcinosis, then measurement of bone density with a dual-energy x-ray absorptiometry (DXA) scan may be useful. In addition to the standard left hip and lumbar spine measurements, assessment at the distal radius may be particularly helpful, because hyperparathyroidism may affect this predominantly cortical site more than the other locations, which have a greater percentage of trabecular bone.[6]

Other diagnostic studies are usually not necessary. Consideration should also be given to the possibility of one of the MEN syndromes, particularly if the patient is young or has a personal or family history of a related endocrinopathy.[5] This information will be helpful to the surgeon, because the patient with primary hyperparathyroidism in the setting of a MEN syndrome usually has multigland parathyroid hyperplasia, and in such patients a surgical procedure beyond a single parathyroidectomy is necessary. If MEN II is suspected, medullary thyroid cancer should be considered, and pheochromocytoma must be excluded before the patient goes to surgery.

Treatment

Previous controversy over which patients with hyperparathyroidism require surgical intervention has been largely resolved. Those without symptoms or complications clearly related to hyperparathyroidism can be followed safely for long periods. Treatment of the patient with primary hyperparathyroidism must take into account the degree of the hypercalcemia, the presence of symptoms, and the severity of any end-organ damage.[8] Understandably, it is widely agreed that patients with symptoms clearly attributable to hypercalcemia should undergo surgery.

Guidelines for surgical intervention in patients with primary hyperparathyroidism were developed at a National Institutes of Health workshop in 2002.[9] The indications for surgical intervention are as follows:

1. Significant bone, renal, gastrointestinal, or neuromuscular symptoms typical of primary hyperparathyroidism.
2. Elevation of serum calcium by 1 mg/dl or more above the normal range (i.e., ≥ 11.5 mg/dl in most laboratories).
3. Marked elevation of 24-hour urine calcium excretion (e.g., > 400 mg).
4. Decreased creatinine clearance (i.e., reduced by ≥ 30% compared with age-matched normal persons.
5. Significant reduction in bone density (i.e., > 2.5 standard deviations below peak bone mass [T score < –2.5 at the lumbar spine, proximal femur, or distal radius]).
6. Consistent follow-up is not possible or is undesirable because of coexisting medical conditions.
7. Age younger than 50 years.

Those patients with mild hypercalcemia who are truly asymptomatic can be followed clinically for the subsequent de-

velopment of surgical indications. Most will likely remain asymptomatic and will not require intervention.[9]

Preoperative localization Imaging studies to locate parathyroid adenomas have become more widely used, particularly as more centers have started offering minimally invasive surgery with intraoperative PTH assays (see below).[10,11] In most cases of adenoma in a single gland, precise knowledge of the location of the adenoma may decrease operative time by allowing the surgeon to direct attention to the area of suspicion. It is important to remember, however, that in good hands, parathyroidectomy for primary hyperparathyroidism has a cure rate in the range of 90% to 95%, even without such localization studies. Thus, it is unlikely that preoperative localization will ever be demonstrated to improve overall surgical outcomes. Localization studies are mandatory before minimally invasive parathyroidectomy, in the setting of a second neck exploration for persistent or recurrent hyperparathyroidism, or if previous thyroid surgery has been performed. The localization test of choice is technetium-99m sestamibi scintigraphy.[12,13] This is often followed by a neck ultrasound of the region demonstrating scintigraphic activity to confirm the location of an enlarged parathyroid gland. An additional benefit of ultrasound at this stage in the evaluation is to provide the opportunity for any coexisting thyroid abnormalities to be addressed.

Surgical management The surgical procedure required in patients with hyperparathyroidism resulting from a solitary parathyroid adenoma is resection of that gland. If intraoperative PTH assays show a drop in the PTH level by more than 50% a few minutes after resection, no further neck exploration is required. Intraoperative measurement of PTH is considered by some experts to be critical in the case of ectopic parathyroid adenoma (which would not be easily found during routine neck exploration) and in reoperations. If an intraoperative PTH assay is not used, the other three glands must be directly inspected to ensure that a second adenoma or generalized hyperplasia is not present.[14] If a second adenoma is found, it too should be excised. If hyperplasia is encountered, the surgeon performs a subtotal parathyroidectomy: removal of approximately three to three and one half glands. In some centers, this is followed by autotransplantation of remaining parathyroid tissue to the forearm, which may simplify follow-up surgical exploration in the event of recurrent hypercalcemia. Additional parathyroid tissue may be frozen in case future need develops. Parathyroid autotransplantation is a controversial treatment, because some patients experience aggressive regrowth of parathyroid tissue within the forearm muscles. This can require challenging and disfiguring surgery to correct.

At certain centers, so-called minimally invasive parathyroidectomy is being offered in conjunction with intraoperative PTH measurements.[11] This approach is best suited for a good surgical candidate in whom both history and preoperative imaging studies suggest a single adenoma (which is, in fact, the most common situation in primary hyperparathyroidism). With information from scintigraphy and ultrasound already in hand, the diseased gland can be excised through a smaller, unilateral incision, under local nerve block, in an ambulatory setting. Success is gauged by the drop in PTH levels intraoperatively. This approach usually provides a better cosmetic result, quicker recovery time, and a lower incidence of postoperative hypocalcemia. Minimally invasive surgery may not be appropriate in suboptimal surgical candidates, in patients who may have multigland disease, and in reoperative cases. However, it is quite likely that the majority of parathyroidectomies will be performed in this fashion in the future.

Table 2 2002 NIH Working Group Recommendations Regarding Follow-up Testing for Patients with Primary Hyperparathyroidism Who Do Not Undergo Surgery[9]

Measurement	*Frequency*
Serum calcium	Biannually
24-hour urine calcium	At initial evaluation only
Creatinine clearance	At initial evaluation only
Serum creatinine	Annually*
Bone mineral density	Annually (lumbar spine, femur, and forearm)
Abdominal radiograph (or ultrasound)	At initial evaluation only

* If the serum creatinine suggests a change in renal function, measurement of creatinine clearance is recommended.

Hyperparathyroidism occasionally persists after operative intervention, usually because of failure to identify the culprit gland, occasionally because of undiagnosed multigland hyperplasia, and rarely because of undiagnosed parathyroid carcinoma.[15-17] Scar tissue and the sometimes unexpected location of remaining pathologic parathyroid tissue make second surgeries notoriously more challenging and prone to complications. Consequently, preoperative imaging studies are invaluable in patients undergoing repeat surgery for persistent hyperparathyroidism. Catheterization studies with venous sampling may also be helpful in certain difficult cases. The identity of putative parathyroid glands can be confirmed by fine-needle aspiration with real-time PTH assay.

Nonsurgical management Although there is as yet no recognized medical therapy for primary hyperparathyroidism, patients who do not meet the criteria for surgical intervention or who refuse surgery can be followed expectantly. This involves periodic monitoring of serum and urine calcium levels, renal function, and BMD, as well as evaluation for nephrocalcinosis or nephrolithiasis. The extent and frequency of this monitoring should be tailored to the individual patient's disease and comorbidities. [*see Table 2*].[9] Drugs that have a tendency to raise serum calcium levels, such as thiazides and lithium, should be avoided. Calcium and excessive vitamin D supplementation should generally be avoided. Dietary calcium should not be restricted, because such restriction may lead to further elevation of PTH and may possibly have detrimental effects on bone mass. Vitamin D deficiency should be identified and treated with gradual supplementation, because vitamin D deficiency will enhance the adverse effects of hyperparathyroidism on bone. Good hydration should be maintained at all times to avoid the development of renal insufficiency and renal stones, especially in patients with hypercalciuria. In patients with low BMD, a bisphosphonate will help to slow bone loss. In patients who are very hypercalcemic but cannot or will not have surgery, calcimimetic agents have been used to control hypercalcemia, although they

are not approved by the Food and Drug Administration for use in this particular setting. For example, the calcimimetic agent cinacalcet is approved for the treatment of secondary hyperparathyroidism in patients with chronic kidney disease who are on dialysis, as well as for the treatment of hypercalcemia in patients with parathyroid cancer. Calcimimetic agents activate the CaSR and thus diminish PTH production. The high cost of these agents and the relative ease of parathyroid surgery make their widespread use in the future unlikely.[18]

PTH-INDEPENDENT HYPERCALCEMIA

Cancer remains the most common cause of PTH-independent, persistent, substantial hypercalcemia and is most frequently to blame when an acutely elevated calcium level is discovered in a hospitalized patient. Other causes include sarcoidosis, certain endocrine disorders, and various drugs and supplements.

Etiology

Malignancy Malignancy-associated hypercalcemia has two forms: humoral hypercalcemia of malignancy (HHM) and local osteolytic hypercalcemia (LOH).

HHM results from the elaboration by the tumor of a circulating factor that has systemic effects on skeletal calcium release, renal calcium handling, or GI calcium absorption. Rarely, it can be caused by the unregulated production of calcitriol (usually by B cell lymphomas). However, the best-recognized mediator responsible for HHM is parathyroid hormone–related protein (PTHrP).[19] Normally, PTHrP appears to serve as a paracrine factor in a variety of tissues (e.g., bone, skin, breast, uterus, and blood vessels); it is involved in cellular calcium handling, smooth muscle contraction, and growth and development. The amino terminus of PTHrP is homologous with that of PTH, and they share a common receptor. When PTHrP circulates in supraphysiologic concentrations, it induces most of the metabolic effects of PTH, such as osteoclast activation, decreased renal calcium output, and increased renal phosphate clearance.

Tumors that produce HHM by secreting PTHrP are usually squamous cell carcinomas (e.g., lung, esophageal, laryngeal, oropharyngeal, nasopharyngeal, or cervical carcinomas).[20] Other tumor types that occasionally produce PTHrP include adenocarcinoma of the breast and ovary, renal cell carcinoma, transitional cell carcinoma of the bladder, islet cell tumors of the pancreas, T cell lymphomas, and pheochromocytoma. All tumors that elaborate PTHrP do so in relatively small amounts; thus, the syndrome typically develops in patients with a large tumor burden. It is also unusual for HHM to be the presenting feature of the cancer.

LOH occurs when a tumor growing within bone itself causes the local release of calcium through the production of cytokines that activate osteoclasts; there is no production of a systemic factor in these cases. The classic tumor associated with this syndrome is multiple myeloma, although other neoplasms, such as adenocarcinoma of the breast and various lymphomas, may also cause LOH. Local factors produced by bone cells may further enhance the growth of such tumors; this results in the skeleton inadvertently working in concert with the tumor to promote progressive bone resorption and calcium release and further advancement of the cancer. (This is the basis of the success of bisphosphonates in the treatment of multiple myeloma.)

Other causes PTH-independent hypercalcemia may be caused by sarcoidosis and other granulomatous diseases, such as tuberculosis, in which granulomas produce calcitriol. Endocrine conditions that may occasionally lead to hypercalcemia include hyperthyroidism (which stimulates bone turnover) and Addison disease (in which volume contraction reduces calcium clearance). Immobilization may increase calcium levels, usually in persons with active bone turnover, such as adolescents or those with previously unrecognized hyperparathyroidism or Paget disease of bone (see below). Use of drugs and dietary supplements (e.g., vitamin D and vitamin A) may be associated with hypercalcemia. The association of thiazides with hypercalcemia is now thought to occur when a thiazide-induced reduction in calcium excretion unmasks previously unrecognized primary hyperparathyroidism. Although only rarely encountered today, the so-called milk-alkali syndrome results from the long-term consumption of large quantities of milk and antacids; milk and antacids were the standard treatment of peptic ulcers in the days before the development of H_2 receptor blockers and proton pump inhibitors.

Diagnosis

If the serum calcium concentration is elevated but the PTH level is very low, the patient has PTH-independent hypercalcemia. Possible causes include malignancy, granulomatous disease, thyrotoxicosis, and vitamin D intoxication. These cases require further laboratory assessment, with the choice of tests depending on the clinical situation.

In malignancy-associated hypercalcemia, the degree of calcium elevation is usually moderate or severe. Evidence of significant volume depletion and generalized debility may dominate the clinical picture, along with other cancer-related symptoms. Typically, the diagnosis of malignancy has already been established. The diagnosis of malignancy-associated hypercalcemia should be suspected in cancer patients with hypercalcemia who have abnormally low PTH concentrations. In patients with tumors associated with HHM, measurement of PTHrP is indicated. Radioimmunoassays for PTHrP are commercially available; an elevation of PTHrP concentration will essentially confirm the diagnosis of most cases of HHM. Special care should be taken to ensure that blood for PTHrP levels is drawn and handled correctly to avoid spuriously low results. In HHM from B cell lymphomas, circulating plasma concentrations of calcitriol are increased. In local osteolytic disease, PTHrP and calcitriol are within normal ranges, and there is definitive evidence of bony metastases.

When the PTH is low and the patient is not known to have a malignancy, diagnostic possibilities include granulomatous diseases, other endocrine disorders, drugs or dietary supplements, and immobilization. Possible laboratory studies in such patients might include measurement of vitamin D metabolites, thyroid hormone levels, or 24-hour urine calcium excretion. If investigation of these diagnoses proves unrewarding, the very rare possibility of unrecognized malignancy may be considered, especially if measurement of PTHrP is performed and shows elevated values. Further imaging studies are indicated in such cases, including a plain chest radiograph or a computed tomographic scan of the thorax as the initial study. If the results are negative, consideration should be given to a comprehensive otolaryngoscopic examination, esophagoscopy, or CT of the abdomen. Should such further assessment be unrevealing, radiographic or endoscopic assessment of the genitourinary tract should be considered.

Treatment

Acute hypercalcemia A nonparathyroid disorder, often a malignancy, is responsible for most cases of acute hypercalcemia [*see Table 1*]. When the serum calcium level is substantially elevated, treatment includes attempts to increase renal calcium excretion while simultaneously attenuating either bone resorption or intestinal calcium absorption, depending on which is the primary source of calcium. Because most patients have at least moderate volume contraction, which further exacerbates their ability to excrete calcium, the initial intervention should be expansion of the intravascular volume with an intravenous infusion of normal saline [*see Table 3*]. This will augment the delivery of sodium and water to the distal nephron, both of which will, in turn, increase urinary calcium excretion. Once the intravascular volume is repleted, adding a loop diuretic such as furosemide will allow continued aggressive saline hydration and may further increase calcium excretion. If the serum calcium concentration does not normalize quickly with intravenous fluid and diuresis, pharmacologic therapy is indicated.[21] Because almost all causes of severe hypercalcemia involve some degree of increased osteoclast activation, drugs that decrease bone turnover are favored. The treatment of choice is a bisphosphonate, such as pamidronate or zoledronic acid, both of which are available for intravenous infusion. Pamidronate is given in a dosage of 60 to 90 mg intravenously over several hours; it is generally well tolerated. Typically, serum calcium levels begin to decrease within 24 to 48 hours of the infusion, although the peak effect may not occur for several days. The action of pamidronate may persist for up to several weeks; treatment can be repeated as needed if renal function will allow. Zoledronic acid is given at a dosage of 4 mg intravenously over no less than 15 minutes. It appears to have a higher potency and an even longer duration of action than pamidronate. A repeat dose may be provided after 7 days. Use of intravenous bisphophonates, especially zoledronic acid, is often associated with an acute-phase response after the first dose, with flulike symptoms. Caution should be employed with these agents in the setting of renal dysfunction. In addition, if parathyroidectomy is planned, use of bisphosphonates should be considered carefully, because they may make postoperative hypocalcemia management more difficult. When more rapid action is desired, subcutaneous injection of calcitonin can be tried, either alone or in conjunction with a bisphosphonate. Calcitonin is given at a dosage of 4 IU/kg twice daily. Calcitonin is a relatively weak hypocalcemic agent; tachyphylaxis to the effects of calcitonin is common and limits its use to a few days. Other possible therapies are plicamycin and gallium nitrate, although certain toxicities limit their use as first-line agents. In severe or refractory cases, hemodialysis against a low-calcium bath may also be undertaken.

In the more unusual situation of hypercalcemia resulting from an increase in gut calcium absorption, such as in vitamin D intoxication or granulomatous diseases, glucocorticoid therapy may have an integral role. Glucocorticoids directly impede intestinal calcium transport and also decrease renal or granulomatous 1α-hydroxylase activity, which results in a decrease in concentrations of calcitriol. In patients with lymphoma, steroids may also have an antineoplastic effect.

Contributing factors to hypercalcemia, such as the use of oral calcium or vitamin D supplements, diuretic therapy, or immobilization, should be corrected, if possible.

In malignancy-associated hypercalcemia, effective surgery, chemotherapy, or radiotherapy targeted at the tumor itself will reduce the hypercalcemia. However, because hypercalcemia is often an end-stage complication, further chemotherapy or radiotherapy may be neither possible nor desired.

Table 3 Therapy for Acute Hypercalcemia

- Fluids
  - 0.9% NaCl I.V.
  - Loop diuretic (forced diuresis)
- Medications
  - Bisphosphonates
    - Pamidronate (60–90 mg I.V.)
    - Zolendronic acid (4 mg I.V.)
  - Calcitonin (4 IU/kg S.C. q. 12 hr)
  - Plicamycin (15–25 μg/kg I.V.)
  - Gallium nitrate (200 mg/m²/day continuous infusion for 5 days)
  - Glucocorticoids (20–100 mg of prednisone a day)
- Other
  - Primary therapy directed at tumor
    - Surgery
    - Chemotherapy
    - Radiation
  - Decrease calcium and vitamin D intake
  - Maintain adequate hydration
  - Mobilize patient

Hypocalcemia

Hypocalcemia is defined as a serum calcium level below the reference range for the laboratory. As with hypercalcemia, an ionized calcium determination on a correctly collected sample is the best way to confirm hypocalcemia.

ETIOLOGY

An abnormally low level of serum calcium on laboratory testing is most often factitious, resulting from a decrease in plasma protein concentration secondary to decreased protein synthesis or hemodilution. Because circulating calcium is so highly protein bound, decreases in serum albumin concentrations—such as occurs with malnourishment, liver disease, or nephrotic syndrome—produce proportionate reductions in total serum calcium. In such situations, the serum calcium level may be corrected by adding 0.8 mg/dl for each 1 g/dl reduction in the serum albumin level below 4 g/dl. An accurate ionized calcium measurement will circumvent many of these pitfalls.

True hypocalcemia is most often related to vitamin D deficiency or impaired parathyroid gland function. Removal of or vascular injury to the parathyroids during neck surgery can result in hypoparathyroidism, which is manifested by hypocalcemia, hyperphosphatemia, and inappropriately low concentrations of PTH. However, unless all four parathyroids are removed or their blood supply is severely impaired, hypocalcemia after parathyroidectomy is usually a transient phenomenon. Normal parathyroid function typically returns after a period of several days to weeks. Patients who experience prolonged, severe primary hyperparathyroidism and significant bone resorption before undergoing parathyroidectomy may experience protracted hypocalcemia and hypophosphatemia after surgery, as a result of the deposition of large quantities of mineral into the skeleton. This is referred to as the "hungry bone syndrome."

Automimmune destruction of the parathyroid glands may be

seen in certain conditions, including autoimmune polyglandular syndrome type 1, a condition marked by hypoparathyroidism, premature ovarian failure, Addison disease, and mucocutaneous candidiasis.[22] Certain infiltrative diseases, such as hemochromatosis, may also adversely affect parathyroid function, as may external-beam irradiation of the neck. Functional hypoparathyroidism may also result from hypomagnesemia, because magnesium is necessary for both PTH release and PTH action. This is often seen in hospitalized alcoholic patients. Pseudohypoparathyroidism is caused by inherited PTH resistance, which results in hypocalcemia and secondary marked elevations of PTH levels.

Because vitamin D ultimately regulates intestinal calcium absorption, disorders of its supply, production, or activity may lead to hypocalcemia**.** In such conditions, serum calcium concentrations are usually not severely affected, thanks to compensatory increases in PTH levels. Indeed, the primary clinical manifestations are in the skeleton (e.g., rickets in children and osteomalacia in adults). Dietary vitamin D deficiency in the elderly is common, but it is often overlooked.[23] At-risk adults include the elderly and darker-skinned persons with poor dietary habits who avoid liquid milk products and have little sun exposure, particularly in northern climates. However, recent reports suggest that vitamin D deficiency may be more frequent than traditionally considered, even in persons not previously thought to be at risk.[24]

Hypocalcemia may occur in patients with acute pancreatitis, when fatty acids released through the action of pancreatic lipase complex with calcium. The complexing of phosphate with calcium also occurs in severely hyperphosphatemic states, such as acute renal failure, rhabdomyolysis, and the tumor lysis syndrome, and it may result in a decrease in serum calcium concentrations. Hypocalcemia may also be caused by large-volume blood transfusions using red blood cells to which calcium chelators have been added to prevent clotting.

Table 4 Differential Diagnosis of Hypocalcemia

Abnormal supply or action of parathyroid hormone	Hypoparathyroidism Surgical External-beam irradiation (to neck) Autoimmune Polyendocrine syndromes Congenital Infiltrative Hemochromatosis Thalassemia Wilson disease Magnesium deficiency DiGeorge syndrome PTH resistance Pseudohypoparathyroidism
Abnormal supply or action of vitamin D	Vitamin D–dependent rickets (VDDR)/ osteomalacia Nutritional deficiency Malabsorption Altered vitamin D metabolism Cirrhosis Renal failure Anticonvulsant medications Vitamin D pseudodeficiency (VDDR I) Abnormal vitamin D receptor (VDDR II) Vitamin D–resistant hypophosphatemic rickets/osteomalacia Oncogenic osteomalacia
Medications/ supplements	Phosphate Calcitonin Bisphosphonates Plicamycin
Other conditions	Hypoalbuminemia (factitious) Acute pancreatitis Rhabdomyolysis Calcium malabsorption Hyperphosphatemia Large transfusions of citrate-containing blood products Osteoblastic metastases (prostatic or breast carcinoma)

DIAGNOSIS

Chronic mild to moderate hypocalcemia is usually well tolerated. However, when the serum calcium level falls below 7.5 to 8 mg/dl (assuming that plasma protein levels are normal), the patient may develop symptoms of neuromuscular irritability, such as tremor, muscle spasms, or paresthesias. On examination, Chvostek and Trousseau signs may be present. If the serum calcium level drops further, tetany or seizures may result. Prolongation of the QT_C interval may also occur, predisposing the patient to cardiac arrhythmias.

As with hypercalcemia, the cause of hypocalcemia can usually be discerned after a careful history (including a review of medications, previous surgeries, and dietary and social habits) and by the measurement of the circulating concentrations of calcium, phosphorus, PTH, and 25-$(OH)D_3$. The differential diagnosis consists principally of conditions that result in an abnormal supply or the abnormal action of PTH or vitamin D, but the use of medications or supplements and the presence of other conditions must be considered [*see Table 4*].

TREATMENT

In patients with symptoms of marked hypocalcemia (e.g., neuromuscular irritability), calcium should be infused slowly (e.g., as calcium gluconate) to raise the serum calcium level until symptoms are relieved. Individual boluses of intravenous calcium will not achieve this effect. Concurrently, any deficiency in magnesium stores should be corrected. In severe cases, hypocalcemia may recur quickly after discontinuance of calcium infusion, so oral calcium should be administered concurrently. In less severe cases, calcium can be administered orally as calcium carbonate or calcium citrate in doses of 1,000 to 2,000 mg of elemental calcium daily in divided doses. If appropriate, vitamin D should also be provided. If dietary deficiency of vitamin D is suspected, cholecalciferol (vitamin D_3) or, if cholecalciferol is unavailable, ergocalciferol (vitamin D_2), may be adequate. Because hydroxylation of cholecalciferol may take several days, however, a brief course of calcitriol may also be necessary. In cases of hypoparathyroidism, long-term administration of calcitriol is needed, because renal 1α-hydroxylase may not be active in the absence of PTH. In hypoparathyroid patients, it is important to not fully normalize the serum calcium level, because this often results in hypercalciuria and hyperphosphaturia, increasing the risk of nephrocalcinosis or renal stones. Instead, serum calcium should be kept around the lower limit of the normal range, at a level sufficient to relieve symptoms and reverse tetanic signs (e.g., Trousseau sign). Periodic monitoring for nephrocalcinosis in these patients may be appropriate.

OSTEOPOROSIS

Osteoporosis is defined as a loss of bone mass, including loss of trabecular bone microarchitecture and connectivity and a thinning of cortical bone, leading to an increased risk of fracture. Clinically, osteoporosis is usually diagnosed by measuring BMD, which reflects the bone calcium content and is a surrogate for bone mass. The diagnostic criteria of the World Health Organization are based on the results of standardized bone density measurements: osteoporosis is present when the BMD is more than 2.5 standard deviations (SDs) below that of a normal, young adult control population (in whom bone density is at its peak). Osteopenia is present when the BMD falls between –1.0 and –2.5 SDs from peak bone density.

Epidemiology

Peak bone mass occurs in persons who are in their late 20s or early 30s; after this age, bone density decreases slowly. Consequently, the incidence of osteoporosis increases with age, becoming most common in persons older than 60 years. Because peak bone mass is lower in women than in men, women generally have lower bone density at each succeeding stage of life.[25] Therefore, women experience higher rates of fracture. The most common sites of so-called fragility fractures are the hip, the distal forearm, and vertebrae.

The lifetime risk of experiencing any fragility fracture for white women is 40%, whereas for white men it is 13%. By site, the respective risks for women and men are as follows: 18% and 6% for hip fracture, 16% and 3% for distal radius fracture, and 18% and 6% for vertebral fracture.[26] The incidence of hip fracture in women and men aged 65 years is approximately 300 and 150 per 100,000 person-years, respectively. These rates increase to approximately 3,000 and 2,000 per 100,000, respectively, by 85 years of age.[25] African Americans generally have higher BMD and are at lower risk for fracture than their white, Hispanic, or Asian counterparts.

Osteoporosis produces enormous burdens, both for patients and for society at large. In the United States, the direct costs of treatment of osteoporotic fractures alone are estimated to be $10 billion to $15 billion annually. Hip fracture carries the highest morbidity and mortality of all fractures. Deaths may occur from associated complications, such as pulmonary embolism or pneumonia. Only one third of patients with hip fracture return to their previous level of functioning. Of the remainder, one third will be placed in long-term nursing care facilities. Rates of depression and anxiety also increase after osteoporotic fracture.[10] Because of the high cost to patients and the insurance system, prevention of hip fracture has been a major focus of osteoporosis prevention and treatment.

Pathogenesis

Bone remodeling occurs continuously in adults; at any given time, as much as 5% to 10% of the skeleton is in a state of turnover. The cells involved in this remodeling process are the osteoclasts, which resorb bone, and the osteoblasts, which form new bone. A cycle of bone remodeling begins with the recruitment of osteoclasts. Osteoclast-mediated resorption of bone from the site releases mineral and collagen breakdown products into the circulation. Through local osteoclast-derived cytokine signals, osteoblasts are then recruited to the site and create new bone matrix to fill the resorption pit left behind by the osteoclasts. The matrix is then mineralized through the physiochemical crystallization of hydroxyapatite. Each bone turnover cycle lasts approximately 3 months.

Through the process of bone remodeling, the skeleton is constantly rejuvenated. In an accelerated form, the bone remodeling process allows for the healing of fractures. In addition, the massive mineral stores of the skeleton are continuously made available to the body for systemic needs, especially during times of decreased calcium supply.

With advancing age, slightly less bone is formed than was resorbed during each remodeling cycle, presumably because of a gradual decline in osteoblast activity. As a result, net bone loss occurs with each cycle, resulting in the gradual decline in bone mass with aging. Therefore, bone loss is to some degree linked to the rate of bone turnover. Any process that increases bone resorption without increasing bone formation will result in a decrease in bone mass and in a concomitant increase in the risk of fracture.

Bone mass accumulates during the first 2 decades of life, achieves its peak in the late third or early fourth decade, stabilizes during the next 1 or 2 decades, and then declines slowly. The age-related decline in bone mass occurs at a rate of approximately 0.1% to 0.5% a year in both sexes. In women, however, the rate of bone loss accelerates during the relatively abrupt loss of gonadal steroids during menopause, especially just before and during the first 6 or 7 years after the cessation of menses. During this period, bone mass may actually fall by up to 4% a year. Thus, by the end of this period, a woman may have lost one quarter to one third of her total skeletal mass.[26,27] Subsequently, bone loss tends to slow to a rate similar to that seen in aging men.

Risk Factors and Pathologic Causes

The most important risk factors for decreased bone density and osteoporotic fracture are advanced age, female gender, postmenopausal status, white or Asian race, personal or family history of fragility fracture, and low body weight.[28,29] These risk factors assist in identifying patients who are at increased risk for bone loss and consequent fracture. Those patients warrant prophylactic measures to help maintain bone mass, and they may benefit from formal bone density measurement to more precisely quantify risk. Other factors contributing to bone loss include cigarette smoking, ethanol abuse, insufficient dietary calcium, and lack of physical exercise.

Diseases or conditions associated with low bone density include Cushing syndrome, glucocorticoid therapy, thyrotoxicosis, excessive thyroid hormone replacement, primary hyperparathyroidism, hypogonadism, intestinal malabsorption, chronic obstructive pulmonary disease, chronic renal or hepatic failure, multiple myeloma and other malignancies, hypopituitarism (growth hormone deficiency), rheumatoid arthritis and other connective tissue diseases, and organ transplantation.

Diagnosis

Although risk-factor analysis assists in determining which patients are at greatest risk for osteoporosis and fracture, the measurement of bone density remains an essential tool to assess risk. Several modalities for measuring bone density are currently in use, including DXA, quantitative CT (QCT), and ultrasound. QCT is not yet widely used clinically. Ultrasound is used for screening purposes, but selected patients must be followed with central DXA measurements. DXA has the highest accuracy

and precision of any densitometric method and is currently the diagnostic tool preferred by most authorities.[30,31] It is also the method most widely employed in large clinical trials of osteoporosis treatment regimens and is both widely available and safe. DXA should therefore be used for the initial screening and follow-up.

In a typical DXA report, the bone density measurements (expressed in g/cm^2) are converted to T scores and Z scores. The T score is the number of SDs the patient's BMD falls above or below the mean value for young, healthy persons at peak bone density. The Z score represents the number of SDs the patient's BMD falls above or below the average value for persons of the patient's age and sex. The T score is the best indicator of fracture risk. Bone density reference databases are currently available only for whites and African Americans; this limits the value of the results obtained for other ethnic groups. The most common sites measured by DXA are the proximal femur and lumbar spine, although the distal nondominant radius can also be assessed. Pitfalls in the interpretation of DXA scans include incorrect patient positioning and improper selection of the region of interest for analysis. Degenerative disease or scoliosis in the lumbar spine can make the bones appear denser, leading to falsely reassuring results. Each SD below peak bone mass represents a loss of 10% to 12% of bone mineral content and corresponds to an approximate twofold to 2.5-fold increase in fracture risk at that site. It should be noted, however, that factors other than bone density play important roles in the risk of fracture. Recent attention has been directed to so-called bone quality, which refers to the microarchitecture and fracture resistance of the bone, and which may not correspond to BMD as measured by DXA.[32] In elderly persons, in particular, additional factors that increase the risk of fractures include low visual acuity, impaired neuromuscular function, decreased mobility, cognitive decline, sedative drug use, and residence in a nursing home.[33] In a review of the risk of fracture in almost 8,000 women enrolled in a longitudinal study of osteoporosis, the clinical factors found to be most important for risk of fracture included a history of fracture after 50 years of age, maternal history of hip fracture, weight less than 125 lb, current cigarette smoking, and the inability to raise oneself from the seated position without use of the arms. By combining these factors with the T score at the hip, these researchers were able to create a fracture risk index that predicted the patient's likelihood of hip fracture over the subsequent 5 years with greater accuracy than seen with the bone density result alone.[34]

In 1998, the National Osteoporosis Foundation recommended that DXA be used as a screening modality in women with established osteoporotic fractures (to establish a baseline for follow-up measurements) and in women without established osteoporotic fractures who are 65 years of age or older or who are younger than 65 years but have one or more accepted risk factors for fragility fracture. These risk factors include low body weight (< 128 lb); current smoking; and personal history of, or a first-degree family relative with, a low-trauma fracture.[29] Indications for bone density measurement in any patient include fracture from mild or moderate trauma, evidence of osteopenia on plain radiography, pending organ transplant, and ongoing or anticipated long-term corticosteroid therapy. Bone density measurements are also useful in the evaluation of patients with conditions that might adversely affect bone mass (e.g., hyperparathyroidism) and for monitoring patients who are receiving therapy for osteoporosis**.** When follow-up bone density studies are performed, it is important that they be done in a reproducible fashion, so that an accurate comparison can be made. It is best to use the same densitometer at the same facility from one year to the next. Identical patient positioning for each scan will also help eliminate error.

Once the diagnosis of osteoporosis or osteopenia is made, the clinician may wish to undertake a selective evaluation to exclude causes of secondary osteoporosis (other than estrogen deficiency). In a premenopausal woman or a man with decreased bone density, such investigations are imperative. A comprehensive history and physical examination will reveal many of the causes of secondary bone loss. The evaluation should explore symptoms of chronic illness, hyperthyroidism, hyperparathyroidism, intestinal disease, and glucocorticoid use.[35] Lifelong calcium and vitamin D intake should be reviewed. In women, the menstrual history should also be discussed, because even relatively short periods of amenorrhea (reflecting estrogen deficiency) in the past may have a detrimental effect on bone mass.[36] Lifestyle factors such as physical activity level, eating disorders, cigarette smoking, and alcohol abuse should also be addressed. In men, osteoporosis is more often associated with a secondary cause, the more common ones being alcoholism, steroid use, and hypogonadism.[37] However, in about half of men with osteoporosis, the disorder is idiopathic.

An extensive biochemical assessment of the patient, other than that indicated by the clinical evaluation, is not necessary. It is reasonable, however, to perform routine blood chemistry studies, including measurement of levels of serum calcium and phosphorus, serum creatinine, and alkaline phosphatase along with a complete blood count. Immunofixation electrophoresis can also be performed to rule out early myeloma if there is any suspicion of malignancy. Subclinical hyperthyroidism can be ruled out with a thyroid-stimulating hormone determination. Measurement of PTH and vitamin D levels is often helpful, because asymptomatic disease is common. Measurement of 24-hour urinary calcium excretion will evaluate for calcium malabsorption or excessive renal losses of calcium.

Treatment

Modifiable risk factors Lifestyle modification is the first step in the prevention or treatment of osteoporosis. Smoking cessation and moderation of alcohol consumption are important first steps. Exercise is an important aspect of osteoporosis management. Weight-bearing physical activity attenuates bone loss; exercise also helps maintain the proximal muscle strength and balance necessary to avoid falls.[38] A physical therapy evaluation is appropriate for patients considered at high risk for falls. Physical therapists can conduct home safety evaluations to identify and address conditions that might promote falls (e.g., trailing electric cords, throw rugs). Physical therapy can also improve strength and gait stability, thus decreasing fall risk. For frail elderly patients, it is prudent to review medication lists and try to eliminate medications that may cause dizziness or sedation and thus predispose patients to falls. Hip protectors have been shown in some, but not all, studies to prevent hip fracture if a patient falls.[39]

Nutritional therapy Any patient being treated for bone loss must consume adequate amounts of both calcium and vitamin D. The recommended daily dietary intake of elemental calcium is 1,000 to 1,500 mg, depending on age and menopausal status [*see Table 5*],[40] and recommended vitamin D intake ranges from

Table 5 Dietary Reference Intakes for Calcium*[76]

Population	*Age (yr)*	*DRI (mg)*
Children	1–3	500
	4–8	800
Males and females	9–18	1,300
	19–50	1,000
	> 50	1,200
Pregnant/lactating women	≤ 18	1,300
	≥ 19	1,000

*Recommended Daily Allowances (RDA) are being replaced with dietary reference intakes (DRI).

400 to 800 IU a day. Patients taking medications that increase vitamin D metabolism (e.g., phenytoin) may need higher vitamin D doses. Substantial and prolonged deficiencies in calcium intake may lead to secondary hyperparathyroidism with reduction of bone mass, as bone is resorbed to release calcium for systemic requirements. Several investigators have demonstrated that calcium and vitamin D supplementation has a beneficial effect on postmenopausal bone loss, although the effects are not as dramatic as those seen with antiresorptive or anabolic therapies.[41] In addition, vitamin D supplementation in the elderly appears to be associated with as much as a 22% decreased risk of falls.[42] Skeletal muscle has vitamin D receptors, and it is thought that vitamin D sufficiency is necessary for optimal muscle strength.

Preferably, calcium and vitamin D should be from dietary sources.[40] Unless milk products are a major component of the diet, however, achieving adequate intake may be difficult; therefore, commercially available supplements should be used. In patients who are elderly, in patients who are taking proton pump inhibitors or H_2-blockers, or in patients who have pernicious anemia, calcium citrate may be better absorbed than calcium carbonate. In all major clinical trials of osteoporosis therapies, participants were also provided basal calcium and vitamin D supplements. Thus, the efficacy of currently available pharmacologic agents for osteoporosis generally has been demonstrated only in persons with adequate calcium and vitamin D intake.

Antiresorptive therapy Osteoporosis is most often treated with antiresorptive agents. FDA-approved agents for the treatment of established osteoporosis include the bisphosphonates (e.g., alendronate, risedronate), selective estrogen receptor modulators (SERMs) (e.g., raloxifene), calcitonin,[28] and estrogen.[43-46] All of these agents reduce vertebral and nonvertebral fracture rates (on the order of 30% to 60%) over 2 to 3 years,[28,37,40,43,44,47-53] but only estrogen[54] and the bisphosphonates have been shown to reduce hip fracture risk.[47-49] All of these agents tend to be more effective for increasing bone density and lowering fracture risk at the spine than at the hip. Estrogen is less widely used for prevention of postmenopausal bone loss since the publication of the Women's Health Initiative (WHI) results. The WHI studies showed an increased risk of cardiovascular disease, breast cancer, stroke, and pulmonary embolism in patients treated with estrogen in combination with progestin,[55] as well as an increase in stroke risk for women treated with estrogen alone.[56]

Antiresorptive therapy should be considered for postmenopausal women for the prevention of osteoporosis, particularly in those with established fracture and those who are at high risk for fracture. For the prevention of osteoporosis, the antiresorptives currently approved by the FDA are alendronate, risedronate, raloxifene, and estrogen. These agents increase BMD at both the hip and the spine during the first 2 to 3 years of use, with subsequent stabilization.

Bisphosphonates Given their demonstrated safety record and their ability to prevent hip fractures, the bisphosphonates should be considered for initial prevention and treatment of osteoporosis. Bisphosphonates bind to skeletal hydroxyapatite and decrease osteoclast activity, thus slowing bone resorption while new bone formation continues. Over a period of 2 to 3 years, they produce a 6% increase in bone density and a 30% to 50% reduction in fracture risk at both vertebral and nonvertebral sites, with greater effectiveness at the former.[48-51] Alendronate is dosed at 70 mg once weekly for osteoporosis treatment and at 35 mg once weekly for osteoporosis prevention.[57] Risedronate is given at 35 mg once weekly for prevention or treatment of osteoporosis.[58] Risks of bisphosphonate therapy include esophagitis; these agents are contraindicated in patients with active esophagitis, achalasia, or esophageal stricture, and they should be used with caution in anyone with a history of esophagitis or gastroesophageal reflux disease. Caution should be taken in prescribing these agents for women of childbearing age: they persist in bone matrix and can be measured in the blood for years after discontinuance, so there is a risk of passage to the fetus even in patients no longer taking them.

Because they are poorly absorbed, bisphosphonates should be taken on an empty stomach immediately upon awakening in the morning. After taking the agent, the patient should remain upright and should not consume food for at least 1 hour. How long bisphosphonate therapy should be provided is an area of active interest in osteoporosis research. Alendronate use for up to 10 years provides a sustained benefit and is well tolerated.[59] The bone-preserving effects of alendronate may persist for up to 2 years after the drug is discontinued, presumably because it remains in bone matrix. There has been concern that prolonged therapy with bisphosphonates might lead to adynamic bone, a condition associated with low bone turnover, microfractures, and chronic pain. It may be appropriate to consider a 1- to 2-year drug holiday after 5 to 10 years of bisphosphonate therapy to avoid these theoretical risks, although no guidelines currently exist for duration of bisphosphonate therapy for osteoporosis.

Raloxifene Raloxifene, a SERM, can be used for osteoporosis prevention or treatment.[60] Raloxifene acts as an estrogen agonist in bone, but it acts as an estrogen antagonist in breast and uterus. Thus, its use is not associated with endometrial hyperplasia, and concurrent treatment with progestins is not required. Raloxifene also does not increase the risk of breast cancer. In fact, it is under active investigation for its potential role in preventing breast cancer in high-risk patients.[61] Raloxifene increases bone density by only about 1% over 12 to 24 months, but data on vertebral fracture are comparable to those of estrogen replacement therapy (ERT) or bisphosphonate therapy. Raloxifene has not been shown to reduce the incidence of hip fracture. It is generally well tolerated but may exacerbate menopausal hot flashes. Patients should be advised that raloxifene carries a threefold increased risk of thromboembolic disease, which is similar to that seen with estrogen. Preliminary reports suggest no detrimental effect on cardiovascular risk.[62] Raloxifene is a good choice for patients who have bone loss primarily in the spine, for

women with relatively low risk of hip fracture, and for patients who are unable to tolerate oral bisphosphonates. Raloxifene is not approved for use in premenopausal women or men.

Calcitonin Calcitonin has much less potent effects on BMD and fracture than do the bisphosphonates, raloxifene, or estrogen.[63] This hormone, which is normally produced by the parafollicular cells (C cells) of the thyroid, typically circulates in low concentration in humans. The precise role of calcitonin in the body is not fully understood, but it appears to be a weak regulator of serum calcium concentrations and bone turnover. Commercially available products include calcitonin injections and nasal spray. Both are approved for the treatment of established osteoporosis but not for its prevention. In most of the calcitonin trials, the average increase in bone density was only 1% to 2% over 2 years. Calcitonin has been primarily shown to prevent vertebral fractures and has not been shown to prevent nonvertebral or hip fractures.[52] Calcitonin is generally safe; occasional flushing, headaches, anosmia, or nasal irritation is observed with the nasal spray. There is a concern about tachyphylaxis, or decreased effectiveness over time, with calcitonin use. Because of the availability of other safe and more potent drugs for osteoporosis, calcitonin is rarely used except when other options are lacking.

Estrogen Until recently, ERT was widely recommended as first-line therapy for both the prevention and treatment of osteoporosis, although it is approved by the FDA for prevention only.[43] Advocates argued that estrogen directly corrected the chief pathophysiologic defect of the menopause: estrogen deficiency. However, use of ERT to maintain bone health has fallen out of favor because of data indicating that it may actually increase the risk of cardiovascular disease, as well as the risk of breast cancer and ovarian cancer. The multicenter WHI was created to study the effects of hormone replacement therapy in healthy postmenopausal women. Women receiving estrogen in combination with progestin had a lower incidence of fracture, but this arm of the study was stopped prematurely because of a 26% increase in the risk of breast cancer and a lack of overall benefit in this treatment group. The WHI also found that, compared with women taking a placebo, women taking the combination of estrogen and progestin had a 29% increase in myocardial infarction, a 41% increase in stroke, and a doubling of thromboembolic events. For hysterectomized women taking estrogen without progestin, the risk of stroke was increased compared with the group receiving placebo. As a result of the WHI findings, estrogen should probably no longer be considered the optimal first-line preventive or therapeutic agent for bone loss in postmenopausal women.[55] On the basis of these data, the use of estrogen for osteoporosis prevention or treatment should be limited to women who require its beneficial effects for menopausal symptoms. For other women, there are equally effective and probably safer alternatives. When used, ERT should be accompanied by a comprehensive screening program consisting of regular lipid profiles, breast examinations, mammography, and gynecologic assessments.[64]

Anabolic therapy with recombinant human PTH (1-34) Currently, PTH (1-34)—teriparatide—is the only available anabolic agent for osteoporosis in the United States, although other anabolic agents have been developed [*see* Future Therapies, *below*]. Teriparatide was approved by the FDA in late 2002 for use in osteoporosis. Although chronic elevation in PTH results in bone loss, brief increases in PTH have anabolic effects on bone. A daily injection of recombinant human PTH (hPTH) provides a brief rise in PTH, resulting in increased BMD. To date, teriparatide is the anabolic agent with the most potent effects on BMD. Its effects are particularly dramatic in the spine: in one study comparing teriparatide and alendronate treatment, at the end of 14 months, alendronate-treated patients had a 5.6% increase in lumbar spine BMD, whereas teriparatide-treated patients had a 12.2% increase.[65] It should be noted, however, that teriparatide has not been demonstrated to prevent hip fracture, and it does not have FDA approval for that purpose.

Teriparatide is administered nightly in a standard subcutaneous dose by a pen delivery system. Side effects include flushing, hypercalcemia, and hypercalciuria. Patients must be monitored with measurement of serum calcium and 24-hour urine calcium levels, and calcium intake must be adjusted as needed to keep urine and serum calcium in a normal range. Teriparatide comes with a black-box warning from the FDA concerning an association with risk of sarcomas, based on its effects in rats. For this reason, it is administered for no longer than 2 years, and its use is contraindicated in patients with active malignancy. Unfortunately, the high cost of teriparatide has limited its use, because much of its target population is on Medicare and does not have prescription drug coverage.

Selection of patients for this costly and potent drug should be done with care. Osteoporosis experts have developed a consensus opinion, published in the spring of 2004, to help clinicians identify appropriate patients for teriparatide therapy. Indications for its use were as follows: (1) history of vertebral fracture, T score –3.0 or below, or age greater than 69 years; (2) fracture or unexplained bone loss in patients on antiresorptive therapy; and (3) intolerance of oral bisphosphonate therapy. Contraindications listed include hypercalcemia, Paget disease, history of irradiation to the skeleton, sarcoma, or malignancy involving bone.[66]

Combining teriparatide and bisphosphonates Simultaneous administration of teriparatide and oral bisphosphonates impairs the anabolic effects of teriparatide.[67] When these agents are given sequentially, however, the increase in BMD seen with initial teriparatide therapy is followed by additional gain during subsequent bisphosphonate treatment.[68] This increase in BMD in the setting of bisphosphonate use is thought to reflect mineralization of bone matrix laid down during the teriparatide treatment. Interestingly, when teriparatide is administered to patients who have previously been treated with bisphosphonate, their response to teriparatide is blunted.[69] This finding is of concern, because many patients currently being considered for teriparatide therapy have already been treated with bisphosphonates.

National Osteoporosis Foundation guidelines The most comprehensive set of guidelines for the management of osteoporosis comes from the National Osteoporosis Foundation (NOF).[70] These guidelines, which were updated in 2004, recommend that any postmenopausal women with a prior vertebral or hip fracture should receive pharmacologic therapy to reduce fracture risk. Therapy should also be initiated in women with a hip T score lower than –1.5 and one or more of the following risk factors for osteoporotic fracture: (1) a family history of osteoporosis in a first-degree relative, (2) a personal history of any fracture as an adult, (3) low body weight (< 127 lb), (4) current smoking,

or (5) use of oral corticosteroid therapy for more than 3 months. Women with no risk factors for osteoporotic fracture (other than age, gender, and menopausal status) should be treated if the hip T score is below –2.0. In the NOF guidelines, the choice of antiosteoporotic therapy follows current FDA indications, without specific recommendations of one agent over another [*see Table 6 for other sources of information*].

Follow-up For purposes of monitoring, a follow-up bone density study is indicated no sooner than 1 to 2 years after the initial determination, depending on the results and whether any therapy is initiated. Subsequent measurements may be made at similar or longer intervals, depending on the patient's progress and any further therapeutic alterations. For more precise comparison, follow-up studies should be performed using the same DXA unit, if possible. It should be borne in mind that with many of these therapies, the expected increase in BMD seen on follow-up DXA studies will be minimal in comparison with the improvement in fracture risk. However, patients who continue to lose bone despite ongoing antiresorptive therapy should be evaluated for previously unrecognized causes of secondary osteoporosis. Biochemical bone turnover markers (e.g., N-telopeptide, pyridinoline cross-links) may be helpful to document patient response to antiresorptive therapy, but they show significant variability within individuals, and therefore large changes are necessary for proper interpretation. In addition, their concentrations may be influenced by timing of collection, diet, and other factors. Accordingly, widespread consensus is lacking on their precise role in osteoporosis management.

Future therapies Recombinant hPTH (1-84) is currently in phase III trials and appears to have effects similar to those of hPTH (1-34) (see above). Additional agents in the SERM class include lasofoxifene (currently being evaluated for FDA approval) and bazedoxifene (in phase III trials). Strontium ranelate is an anabolic agent that appears to act on the CaSR to induce osteoblast differentiation. It enhances both bone resorption and formation, with an emphasis on formation, resulting in significantly decreased risk of vertebral fracture.[71] It was approved for use in Europe in late 2004. Preclinical studies of the anabolic effects of vitamin D on bone are under way[72] and may lead to the development of vitamin D analogues for osteoporosis therapy.

OSTEOMALACIA

Osteomalacia is a condition in which the bone matrix is normal in quantity but is weakened by insufficient mineral content. Osteomalacia in the growing skeleton is termed rickets. Causes of osteomalacia include nutritional deficiencies of calcium, phosphate, or vitamin D; intestinal disease affecting the absorption of these substances; abnormalities in vitamin D metabolism, such as occurs in liver disease, renal failure, or through the use of antiepileptic drugs; vitamin D resistance; renal phosphate leak; and oncogenic osteomalacia (a humoral syndrome of increased urinary phosphate loss associated with rare tumors of mesenchymal origin).[73] In adults, severe osteomalacia presents as fatigue, proximal muscle weakness, and diffuse or focal skeletal pain. Mild osteomalacia is common and often asymptomatic.[74] Decreased or low-normal concentrations of both calcium and phosphorus are noted on biochemical testing, and the alkaline phosphatase concentration is elevated. Depending on the cause, decreased levels of either 25-$(OH)D_3$ or calcitriol may be seen. Plain films may demonstrate osteopenia and pseudofractures. When necessary, the diagnosis can be confirmed with bone biopsy processed and analyzed by an experienced bone pathologist. The treatment of osteomalacia depends on the pathogenesis of the condition. For the majority of cases in which the problem is dietary insufficiency or malabsorption of vitamin D, administration of high doses of cholecalciferol and calcium will rapidly correct deficits and heal the bone. Underlying conditions such as celiac disease (which may be asymptomatic) should be identified and treated. Other conditions, such as tumor-induced osteomalacia, represent special cases and require a different approach (e.g., phosphate supplementation).

Table 6 Internet Resources for Osteoporosis and Bone Metabolism

American Dietetic Association Nutrition Resources
http://www.eatright.org
American Society for Bone and Mineral Research
http://www.asbmr.org
BoneKEY-Osteovision Site of the International Bone and Mineral Society (IBMS)
http://www.bonekey-ibms.org
International Osteoporosis Foundation
http://www.osteofound.org

PAGET DISEASE OF BONE

Paget disease is a relatively common condition in which abnormal osteoclast function leads to accelerated and disordered bone remodeling, producing highly disorganized bone microarchitecture in affected areas. This sometimes leads to deformity of affected bones, increased vascularity, nerve impingement syndromes, and a propensity to fracture. Paget disease is commonly seen in the elderly and may be familial. The precise etiology is not yet known, although a viral origin is suspected. Many persons with Paget disease are asymptomatic. The sole manifestation may be increased serum alkaline phosphatase activity, detected incidentally on blood testing. If the disease is severe or extensive, pain syndromes may result. Very often, the discomfort originates not from bone itself but from arthritic changes in adjacent joints caused by altered biomechanics. The skull may be enlarged, or there may be significant bowing of the long bones of the legs. Bony overgrowth may lead to local impingement on spinal nerve roots, with pain or neurologic deficits; overgrowth in the inner ear can lead to sensorineural hearing loss. Rare complications include high-output chronic heart failure (from multiple vascular shunts in bone) and transformation to osteosarcoma.

Diagnosis

The diagnosis of Paget disease is typically made after finding isolated elevation of the serum alkaline phosphatase level without evidence of liver disease. (Fractionation of alkaline phosphatase isoenzymes can confirm bone as the source). A nuclear bone scan is performed next to identify involved areas. The results of the bone scan will identify which bones should be evaluated by plain x-ray to exclude signs of metastatic disease and confirm pagetic findings.

Treatment

Treatment of Paget disease is indicated for patients with bone pain; it is also indicated, regardless of symptoms, if there is involvement of a weight-bearing bone or a joint. Antiresorptive

agents, such as high-dose oral or injectable bisphosphonates (see above) or injectable calcitonin, can be used to treat this disorder.[75] Treatment of any vitamin D deficiency is essential before starting intravenous bisphosphonate therapy to avoid hypocalcemia. Disease activity and response to therapy are assessed with serial measurement of alkaline phosphatase or other bone turnover markers. The goal of treatment is normalization of alkaline phosphatase. Re-treatment is indicated if the alkaline phosphatase level begins to rise above normal. Patients with skull involvement should have periodic audiometry to exclude hearing loss.

Elizabeth H. Holt, M.D., Ph.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Silvio E. Inzucchi, M.D., has received research funding from Eli Lilly Co. and speakers' honoraria from Merck & Co., Inc.

References

1. Brown EM: Physiology and pathophysiology of the extracellular calcium–sensing receptor. Am J Med 106:238, 1999

2. Marx SJ: Hyperparathyroid and hypoparathyroid disorders. N Engl J Med 343:1863, 2000

3. Inzucchi SE: Management of hypercalcemia: diagnostic workup, therapeutic options for hyperparathyroidism and other common causes. Postgrad Med 115:27, 2004

4. Wermers RA, Khosla S, Atkinson EJ, et al: The rise and fall of primary hyperparathyroidism: a population-based study in Rochester, Minnesota, 1965–1992. Ann Intern Med 126:433, 1997

5. Brandi ML, Gagel RF, Angeli A, et al: Guidelines for diagnosis and therapy of MEN type 1 and type 2. J Clin Endocrinol Metab 86:5658, 2001

6. Khan A, Bilezikian J: Primary hyperparathyroidism: pathophysiology and impact on bone. CMAJ 163:184, 2000

7. Brown EM: Familial hypocalciuric hypercalcemia and other disorders with resistance to extracellular calcium. Endocrinol Clin North Am 29:503, 2000

8. Silverberg SJ, Shane E, Jacobs TP, et al: A 10-year prospective study of primary hyperparathyroidism with or without parathyroid surgery. N Engl J Med 341:1249, 1999

9. Bilezikian JP, Potts JT Jr, Fuleihan Gel-H, et al: Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century. J Bone Miner Res 17(suppl 2):N2, 2002

10. Westerdahl J, Lindblom P, Bergenfelz A: Measurement of intraoperative parathyroid hormone predicts long-term operative success. Arch Surg 137:186, 2002

11. Monchik JM, Barellini L, Langer P, et al: Minimally invasive parathyroid surgery in 103 patients with local/regional anesthesia, without exclusion criteria. Surgery 131:502, 2002

12. Civelek AC, Ozalp E, Donovan P, et al: Prospective evaluation of delayed technetium-99m sestamibi SPECT scintigraphy for preoperative localization of primary hyperparathyroidism. Surgery 131:149, 2002

13. Dackiw AP, Sussman JJ, Fritsche HA Jr, et al: Relative contributions of technetium Tc 99m sestamibi scintigraphy, intraoperative gamma probe detection, and the rapid parathyroid hormone assay to the surgical management of hyperparathyroidism. Arch Surg 135:550, 2000

14. Irvin GL 3rd, Carneiro DM: Management changes in primary hyperparathyroidism. JAMA 284:934, 2000

15. Udelsman R: Six hundred fifty-six consecutive explorations for primary hyperparathyroidism. Ann Surg 235:665, 2002

16. Kearns AE, Thompson GB: Medical and surgical management of hyperparathyroidism. Mayo Clin Proc 77:87, 2002

17. Shepherd JJ, Burgess JR, Greenaway TM, et al: Preoperative sestamibi scanning and surgical findings at bilateral, unilateral, or minimal reoperation for recurrent hyperparathyroidism after subtotal parathyroidectomy in patients with multiple endocrine neoplasia type 1. Arch Surg 135:844, 2000

18. Peacock M, Bilezikian JP, Klassen, PS, et al: Cinacalcet hydrochloride maintains long-term normocalcemia in patients with primary hyperparathyroidism. J Clin Endocrinol Metab 90:135, 2005

19. Dunbar ME, Wysolmerski JJ, Broadus AE: Parathyroid hormone–related protein: from hypercalcemia of malignancy to developmental regulatory molecule. Am J Med Sci 312:287, 1996

20. Rankin W, Grill V, Martin TJ: Parathyroid hormone–related protein and hypercalcemia. Cancer 80(8 suppl):1564, 1997

21. Ziegler R: Hypercalcemic crisis. J Am Soc Nephrol 12(suppl 17):S3, 2001

22. Betterle C, Greggio NA, Volpato M: Clinical review 93: autoimmune polyglandular syndrome type 1. J Clin Endocrinol Metab 83:1049, 1998

23. Thomas MK, Lloyd-Jones DM, Thadhani RI, et al: Hypovitaminosis D in medical inpatients. N Engl J Med 338:777, 1998

24. Meyer C: Scientists probe role of vitamin D: deficiency a significant problem, experts say. JAMA. 292:1416, 2004

25. Cummings SR, Melton LJ: Epidemiology and outcomes of osteoporotic fractures. Lancet 359:1761, 2002

26. Seeman E: Pathogenesis of bone fragility in women and men. Lancet 359:1841, 2002

27. Melton LJ III, Thamer M, Ray NF, et al: Fractures attributable to osteoporosis: report from the National Osteoporosis Foundation. J Bone Miner Res 12:16, 1997

28. Heinemann DF: Osteoporosis: an overview of the National Osteoporosis Foundation clinical practice guide. Geriatrics 55:31, 2000

29. Fitzpatrick LA: Secondary causes of osteoporosis. Mayo Clin Proc 77:453, 2002

30. Kanis JA: Diagnosis of osteoporosis and assessment of fracture risk. Lancet 359:1929, 2002

31. Miller PD, Zapalowski C, Kulak CA, et al: Bone densitometry: the best way to detect osteoporosis and to monitor therapy. J Clin Endocrinol Metab 84:1867, 1999

32. Ammann P, Rizzoli R: Bone strength and its determinants. Osteoporo Int 14(S3):S13, 2003

33. Slemenda C: Prevention of hip fractures: risk factor modification. Am J Med 103(2A):65S, 1997

34. Black DM, Steinbach M, Palermo L, et al: An assessment tool for predicting fracture risk in postmenopausal women. Osteoporos Int 12:519, 2001

35. Canalis E, Giustina A: Glucocorticoid-induced osteoporosis: summary of a workshop. J Endocrinol Metab 86:5681, 2001

36. Miller KK, Klibanski A: Clinical review 106: amenorrheic bone loss. J Endocrinol Metab 84:1775, 1999

37. Bilezikian JP: Osteoporosis in men. J Endocrinol Metab 84:3431, 1999

38. Messinger-Rapport BJ, Thacker HL: Prevention for the older woman: a practical guide to prevention and treatment of osteoporosis. Geriatrics 57:16, 2002

39. Kannus P, Parkkari J, Niemi S, et al: Prevention of hip fracture in elderly people with use of a hip protector. N Engl J Med 343:1506, 2000

40. Weaver CM: Calcium requirements of physically active people. Am J Clin Nutr 72(2 suppl):579S, 2000

41. Atkinson SA, Ward WE: Clinical nutrition: 2. The role of nutrition in the prevention and treatment of adult osteoporosis. CMAJ 165:1511, 2001

42. Bischoff-Ferrari HA, Dawson-Hughes B, Willett WC, et al: The effect of vitamin D on falls: a meta-analysis. JAMA 291:1999, 2004

43. Manson JE, Martin KA: Clinical practice: postmenopausal hormone-replacement therapy. N Engl J Med 345:34, 2001

44. Altkorn D, Vokes T: Treatment of postmenopausal osteoporosis. JAMA 285:1415, 2001

45. Delmas PD: Treatment of postmenopausal osteoporosis. Lancet 359:2018, 2002

46. NIH Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy. JAMA 285:785, 2001

47. Liberman UA, Weiss SR, Broll J, et al: Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. N Engl J Med 333:1437, 1995

48. Hosking D, Chilvers CED, Christiansen C, et al: Prevention of bone loss with alendronate in postmenopausal women under 60 years of age. N Engl J Med 338:485, 1998

49. Fracture risk reduction with alendronate in women with osteoporosis: the Fracture Intervention Trial. FIT Research Group. J Clin Endocrinol Metab 85:4118, 2000

50. Randomized trial of the effects of risedronate on vertebral fractures in women with established postmenopausal osteoporosis. Vertebral Efficacy with Risedronate Therapy (VERT) Study Group. Osteoporos Int 11:83, 2000

51. Effect of risedronate on the risk of hip fracture in elderly women. Hip Intervention Program Study Group. N Engl J Med 344:333, 2001

52. Rico H, Revilla M, Hernandez ER, et al: Total and regional bone mineral content and fracture rate in postmenopausal osteoporosis treated with salmon calcitonin: a prospective study. Calcif Tissue Int 56:181, 1995

53. Maricic M, Adachi JD, Sarkar S, et al: Early effects of raloxifene on clinical vertebral fractures at 12 months in postmenopausal women with osteoporosis. Arch Intern Med 162:1140, 2002

54. Effects of hormone therapy on bone mineral density: results from the postmenopausal estrogen/progestin interventions (PEPI) trial. PEPI Writing Group. JAMA 276:1389, 1996

55. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy post-menopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

56. Anderson GL, Limacher M, Assaf AR, et al: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 291:1701, 2004

57. Vasikaran SD: Bisphosphonates: an overview with special reference to alendronate. Ann Clin Biochem 38(pt 6):608, 2001

58. Crandall C: Risedronate: a clinical review. Arch Intern Med 161:353, 2001

59. Bone HG, Hosking D, Devogelaer JP, et al: Ten years' experience with alendronate for osteoporosis in postmenopausal women. N Engl J Med 350:1189, 2004

60. Khovidhunkit W, Shoback DM: Clinical effects of raloxifene hydrochloride in women. Ann Intern Med 130:431, 1999

61. Continued breast cancer risk reduction in postmenopausal women treated with raloxifene: 4-year results from the MORE trial. Multiple Outcomes of Raloxifene Evaluation. Breast Cancer Res Treat 65:125, 2001

62. Barrett-Connor E, Grady D, Sashegyi A, et al: Raloxifene and cardiovascular events in osteoporotic postmenopausal women: four-year results from the MORE (Multiple

Outcomes of Raloxifene Evaluation) randomized trial. JAMA 287:847, 2002

63. Downs RW Jr, Bell NH, Ettinger MP, et al: Comparison of alendronate and intranasal calcitonin for treatment of osteoporosis in postmenopausal women. J Clin Endocrinol Metab 85:1783, 2000

64. Burkman RT, Collins JA, Greene RA: Current perspectives on benefits and risks of hormone replacement therapy. Am J Obstet Gynecol 185(2 suppl):S13, 2001

65. Body JJ, Gaitch GA, Scheele WH, et al: A randomized double-blind trial to compare the efficacy of teriparatide [recombinant human parathyroid hormone (1-34)] with alendronate in postmenopausal women with osteoporosis. J Clin Endocrinol Metab 87:4528, 2002

66. Miller PD, Bilezikian JP, Deal C, et al: Clinical use of teriparatide in the real world: initial insights. Endocr Practice 10:139, 2004

67. Finkelstein JS, Hayes A, Hunzelman JL, et al: The effects of parathyroid hormone, alendronate, or both in men with osteoporosis. N Engl J Med 349:1216, 2003

68. Rittmaster RS, Bolognese M, Ettinger MP, et al: Enhancement of bone mass in osteoporotic women with parathyroid hormone followed by alendronate. J Clin Endocrinol Metab 85:2129, 2000

69. Ettinger B, San Martin J, Crans G, et al: Differential effects of teriparatide on BMD after treatment with raloxifene or alendronate. J Bone Miner Res 19:745, 2004

70. Physician's Guide to Prevention and Treatment of Osteoporosis. National Osteoporosis Foundation, Washington, DC, 2003
www.nof.org/physguide

71. Meunier PJ, Roux C, Seeman E, et al: The effects of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. N Engl J Med 350:459, 2004

72. Gardiner EM, Baldock PA, Thomas GP, et al: Increased formation and decreased resorption of bone in mice with elevated vitamin D receptor in mature cells of the osteoblast lineage. FASEB J 14:1908, 2000

73. Drezner MK: Tumor-induced osteomalacia. Rev Endocrinol Metab Disord 2:175, 2001

74. Reginato AJ, Falasca GF, Pappu R, et al: Musculoskeletal manifestations of osteomalacia: report of 26 cases and literature review. Semin Arthr Rheum 28:287, 1999

75. Lyles KW, Siris ES, Singer FR, et al: A clinical approach to diagnosis and management of Paget's disease of bone. J Bone Miner Res 16:1379, 2001

76. Dietary Reference Intakes (DRI) and Recommended Dietary Allowances (RDA). Food and Nutrition Information Center, USDA/ARS/National Agriculture Library, Beltsville, Maryland, December 2002
http://www.nal.usda.gov/fnic/etext/000105.html

Acknowledgment

Figure 1 Seward Hung.

VII GENETICS FOR THE CLINICIAN

Robb Moses, m.d.
Jone E. Sampson, m.d.

The human genome consists of approximately three billion pairs of nucleotides (bases) that encode about 40,000 genes in a string array consisting of a DNA polymer duplex. The information in the protein-coding genes is converted to functional elements through the copying of base sequences to RNA. RNA may itself be active—heterogeneous ribonucleotide proteins can interact with newly synthesized RNA to regulate structural changes or conformation—or it may be copied to protein. Whereas DNA is a relatively stable molecule, RNA is much less so, and RNA molecules are replaced by other RNA molecules rapidly in the cell. Models of how the genome is packaged within the nucleus, copied, and read by the cell have evolved at the molecular level. Science has also come to recognize how the genome is partitioned during cell division and during the formation of germ cells. At all levels, the information content is protected by mechanisms safeguarding the stability of the genome. Information is transferred out of the genome along two paths: information is transferred within the cell for defined functions specific to cell type (horizontal information transfer) and is transferred from one generation of cells to another through cell division, either for cell multiplication or for reproduction (vertical transfer).

The advances in the past century, from the verification of Mendel's observations[1] to having in hand the essentially complete sequence of the human genome, occurred in bursts in understanding or in technology. During the first 50 years of the 20th century, the principle of inheritance by means of packets of genetic information that were stable and that persisted independently of other units of inheritance from generation to generation was verified in animals and plants, with the fruit fly *Drosophila melanogaster* being a notable organism of study. In the 1940s, DNA was unequivocally shown to be the chemical basis of the gene.[2] Within another dozen years, the structure of DNA at the chemical level was proposed by Watson and Crick.[3] The model had immediate implications for the copying of genes and the mode of transfer of information. During the next 2 decades, the fundamental rules and mechanisms of these processes were determined; these advances relied heavily on the study of bacteria and their viruses, the phages. In the 1970s, three different technological advances catapulted genetics to the point at which the human DNA sequence was determined by century's end. These disparate techniques were as follows: the development of the ability to determine DNA sequence information in a relatively simple and reproducible manner; the ability to move and duplicate isolated segments of DNA; and the evolution of computers with adequate power for storing and comparing large amounts of sequence information. Refinements in these basic advances, coupled with automation, led to the sequencing of the human genome during the last 15 years before the turn of the millennium.

With progress in technology, genetics has become applicable to clinical practice. The physician needs to be responsive to patterns of inheritance suggestive of genetic disease in a family member; the physician should also be aware of diagnostic capabilities for inherited diseases and be able to interpret the results of genetic testing for the patient or refer the patient for counseling. Given the pace of advances and the complexity of techniques, it is important that clinicians have knowledge of resources for patient referral and other relevant information. These resources include the Internet and other resources for technical and medical information and patient referral.

Diagnostic capabilities continue to improve, but therapies based on genetic technologies still lag. Prenatal diagnosis and neonatal screening are already powerful tools for the prevention of disease. However, researchers in genetic medicine are striving to address and develop treatments. New diagnostic powers based on microarray analysis, the hope of gene therapy, and the tailoring of drug therapies to maximize responsiveness and sensitivity for individuals seem feasible.

This subsection addresses basic knowledge of the human genome for the physician, gives reference points for the nongenetic physician, and focuses on the pathophysiology of genetic disease.

Genome Structure and Function

DNA STRUCTURE

Understanding the structure of the DNA duplex led at once to recognition that the polymer strand, composed only of four bases, could encode information in a linear format. DNA is composed of two polymer strands with a sugar-phosphate backbone, with the bases attached to each deoxyribose moiety [*see Figure 1*]. The two strands of DNA are stabilized by the bonding of hydrogen between the bases—the purine adenine (A) pairs with the pyrimidine thymine (T), and the purine guanine (G) pairs with the pyrimidine cytosine (C). The discovery of this was pivotal in modeling the structure of DNA. Alternative forms of pairing can be projected, but the pairings based on the most common structures of the bases and the strongest hydrogen bonds are predominant. The sugar-phosphate backbones of the strands have a chemical polarity, and the strands are antiparallel with respect to the polarity of the bonds between deoxyribose components. The information can be duplicated using each separate strand as a template. The linear code of the four-base alphabet allows enormous potential for information content—each cell contains about 1 m of DNA packaged within it.

DNA REPLICATION

The replication of DNA is a critical step in information storage. Although the structure is suggestive of a model for the replication of DNA, the definition of the apparatus took decades [*see Figure 2*]. Because the information is contained in the sequence of the bases, any change of a base could result in a subsequent change (mutation); mutations may be silent, in which case they do not lead to a change in protein function, or mutations may result in the inactivation of the product of the gene. Cells possess mechanisms to protect against changes in the DNA sequence, although natural variety—an advantage in selection—requires a low rate of change.

The replication machinery is centered on a DNA polymerase. DNA polymerase is an unusual enzyme in that it catalyzes a reaction directed by the base sequence of the strand be-

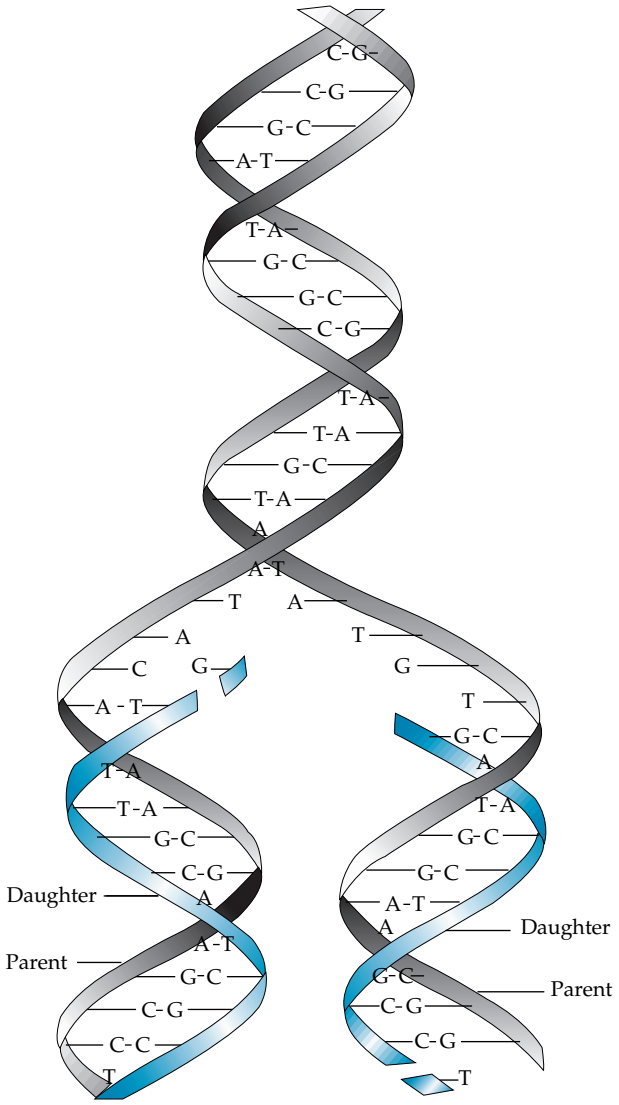


Figure 1 **In replication, the two strands of the parent DNA molecule (gray) separate as the base pairs detach. The daughter strands (blue) form when guanine (G) pairs with cytosine (C) and when adenine (A) pairs with thymine (T). The orientation of the two strands is antiparallel, so the strands grow in opposite directions.**

ing copied; that is, the base inserted into the new strand is directed by the base opposite. The enzyme uses all four bases with equal affinity, and the reaction is determined by the template. DNA polymerases are part of an assembly of proteins termed the replisome. The synthesis of new DNA is faithful; only about one base in 100,000 is a misincorporation. The replisome is actively proofreading the product, an exonuclease, which can remove a mispaired base from the growing end of the DNA strand (all DNA is made only in the 5′ to 3′ direction, based on the convention for the sugar-phosphate backbone). This proofreading by the replisome removes 99% of misincorporations. A last line of defense for integrity of information during replication is the system of mismatch repair. A complex of proteins tracks DNA synthesis, recognizes mismatches in the DNA that occur as a result of misincorporation, and corrects those mismatches. About 99% of mismatches are removed; the overall mistake rate in replication is about one in one billion to 10 billion.

Information is copied from the long-lived DNA molecule to the less stable messenger RNA (mRNA) molecule, which serves to translate the information into proteins. mRNA represents only 1% or so of the RNA in the cell; other categories of RNA synthesize proteins or act as catalytic units in the processing of genetic information. For protein synthesis, the mRNA must be read and the information transferred in a three-base code (codon). This allows for more triplet codons than the 20 amino acids used to build proteins would require. Several of the codons are delegated to terminating protein synthesis, and there is redundancy in the codons for the amino acids.[4]

mRNA is initially transcribed as a long, exact copy of the DNA strand, but posttranslational processing results in retention of the essential information needed for translation to protein. The genome contains introns, which interrupt the coding sequences, the exons. The intron material does not contain information for translation, so it must be removed for translation. This occurs during the RNA processing step, which is regulated by conserved sequences that identify the end and start of the exons [*see Figure 3*], allowing the coding portions to be spliced together. Because the signals that lead to removal are not stringent, it is possible for exons to be occasionally skipped and for splice variants and splice mutations to occur. This gives a great degree of flexibility for the final gene product and allows tissue-specific products.

The primary RNA transcript has other notable features. The 5′ end is capped with a protective methylated base; the 3′ end contains a region rich in pyrimidines and ends in a stretch of A residues, the poly-A region. After processing, the mRNA is transported to the cytoplasm, where translation on the ribosome protein synthesizing machine occurs.

After a protein is made, it may be modified through the attachment of sugars, acetyl groups, phosphates, or other modifying groups. Specific sequences direct the export or subcellular localization of the protein.

CHROMOSOME STRUCTURE AND IDENTIFICATION

The DNA double helix is packaged into the chromosome, a structure recognizable by light microscopy. The packaging is very efficient; the DNA is in the form of supercoils—like a rubber band that is tightly wound until it compacts upon itself. It is then folded into the chromatin assembly by the binding of basic histone proteins. The resulting structure resembles beads on a string, with the DNA wound tightly around a core of histone proteins—two H2A, two H2B, two H3, and two H4 residues—to form the nucleosome. Nucleosomes are spaced approximately 80 bases apart. The DNA structure is further condensed by the addition of other proteins. The DNA in chromatin also may be modified by the addition of methyl groups to certain positions, and the histones may be modified by phosphorylation, acetylation, or the addition of ubiquitin. These modifications to the chromatin are related to the regulation of gene expression.

There are 23 chromosome packages of genes in a cell; two of these chromosomes, X and Y, are the sex chromosomes. In females, there are two X chromosomes; in males, an X and a Y chromosome. The remaining 22 pairs of chromosomes are the autosomes. In the process of cell division, or mitosis, the chromosomes condense and are duplicated, with a complete, new set going to each daughter cell [*see Figure 4*]. In producing germ cells for reproduction, the number of chromosomes is halved to a haploid number of autosomes through the process of meiosis,

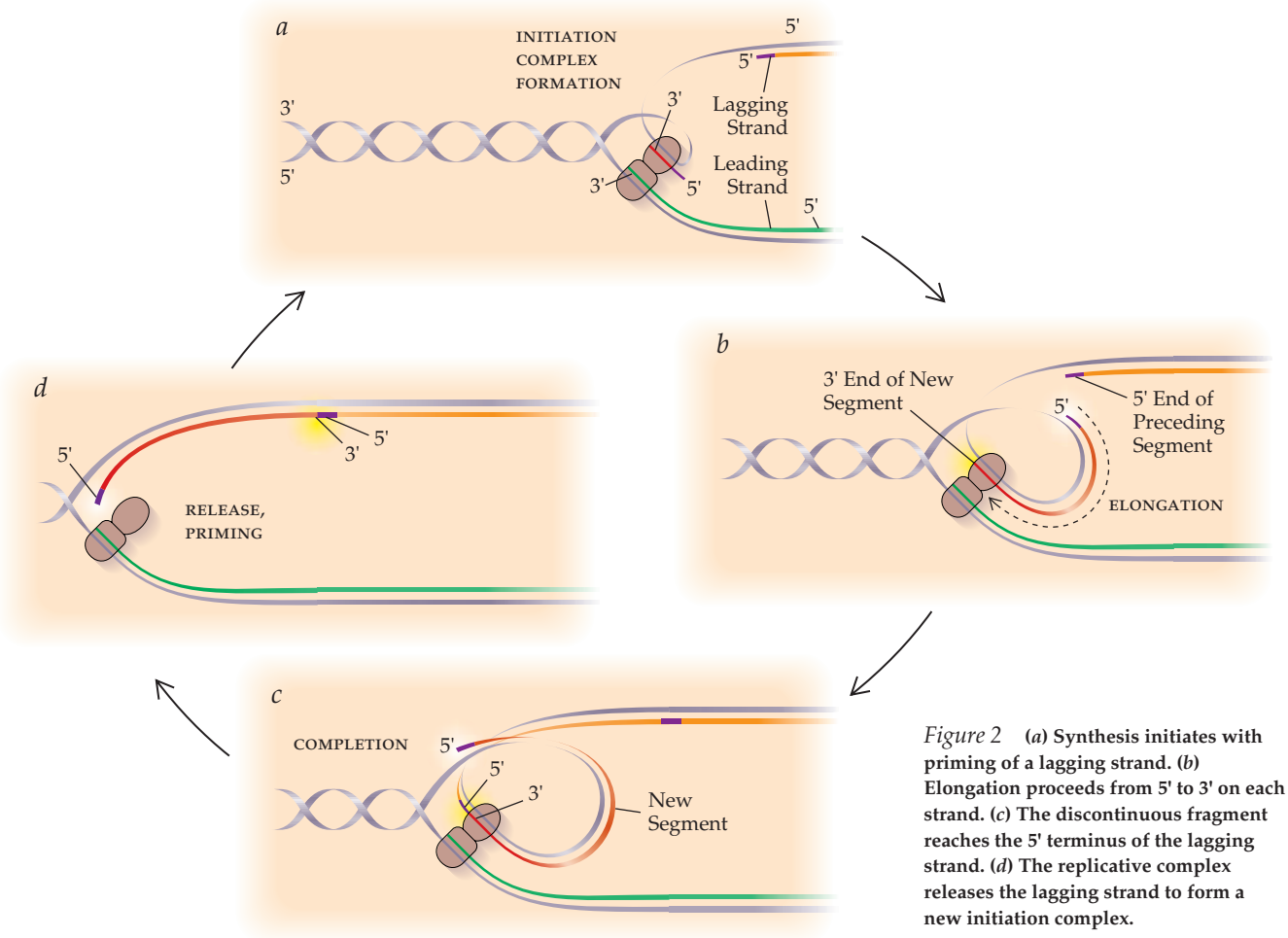


Figure 2 **(*a*) Synthesis initiates with priming of a lagging strand. (*b*) Elongation proceeds from 5' to 3' on each strand. (*c*) The discontinuous fragment reaches the 5' terminus of the lagging strand. (*d*) The replicative complex releases the lagging strand to form a new initiation complex.**

with either an X or a Y chromosome in spermatozoa and an X chromosome in oocytes. Telomeres are special DNA sequences at the ends of chromosomes that maintain integrity at chromosome termini and require telomerase enzyme activity for maintenance of normal length. The chromosomes contain a region of repeated sequence DNA, which is the centromere. This is the portion that anchors the replicated duplexes (chromatids) together at the time of cell division. The centromere is not centrally located; this results in a long arm, termed the q arm, and a short arm, termed the p arm, for petite. There is a standard system of nomenclature for describing the number of chromosomes and recognizable alterations or rearrangements of the chromosomes. For example, the normal male karyotype is listed as 46,XY and the normal female as 46,XX.

Chromosome identification and characterization was much improved by the development of staining or banding techniques. This process involves partial denaturation of the DNA and proteins, followed by staining. The resulting preparations allow identification of up to 800 bands, which may then be evaluated for structural changes. These techniques have led to the recognition of many rearrangements, leading to localization of genes.

Additional staining techniques based on binding to complementary short stretches of DNA tagged with fluorescent probes have been developed. Further development has led to chromogenic stains, which allow the identification of individual chromosomes and of certain regions of the chromosome—for example, the centromere or telomere.

Mutations in Clinical Conditions

Mutations are changes in the sequence of nucleotides within DNA, which may result in an abnormal or deleterious function of the gene product. There are different types of mutations leading to genetic disease. These mutations range in size from single base changes that alter the gene product to the addition or deletion of whole chromosomes. Intermediate structural rearrangements may involve segments that are large enough to be able to be detected microscopically, or they may involve segments that are so small as to require detection by molecular labeling methods. Genetic diseases resulting from single gene mutations are inherited in classic Mendelian fashion, although there is always the possibility of new mutations occurring in individuals with unaffected parents. Several disorders were originally thought to be genetic in nature, but the pedigrees of affected individuals were not consistent with known patterns of Mendelian inheritance. Understanding the mechanisms by which mutations occur in these disorders led to an understanding of other factors that influence disease; such factors include the effect of imprinting on phenotype expression, the role of trinucleotide expansion in genetic diseases, and the role of mitochondrial DNA (mtDNA) mutations in disorders of energy metabolism.

The following discussions describe the roles of the different mechanisms of mutation in genetic conditions seen more commonly in the general population.

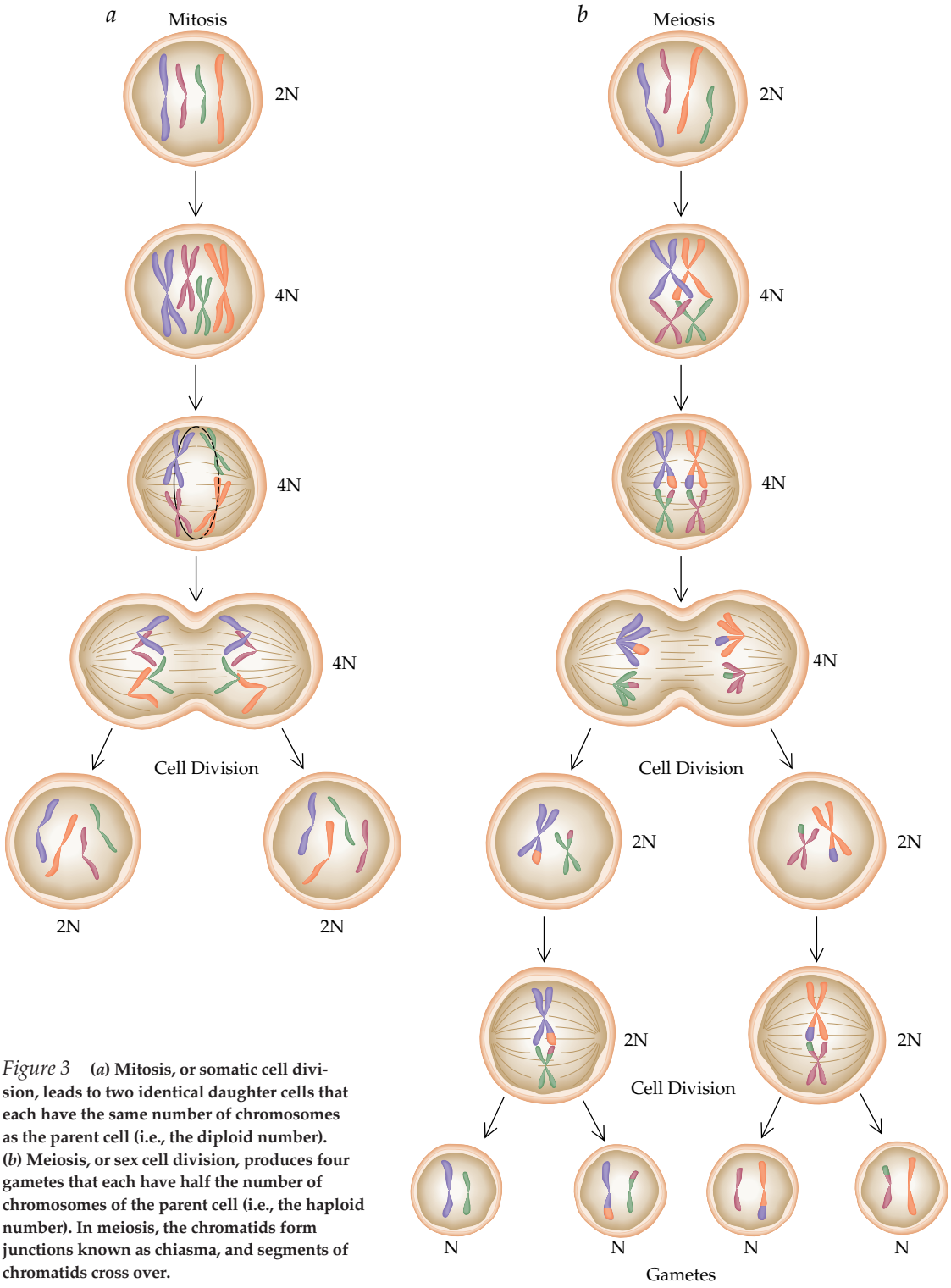


Figure 3 **(*a*) Mitosis, or somatic cell division, leads to two identical daughter cells that each have the same number of chromosomes as the parent cell (i.e., the diploid number). (*b*) Meiosis, or sex cell division, produces four gametes that each have half the number of chromosomes of the parent cell (i.e., the haploid number). In meiosis, the chromatids form junctions known as chiasma, and segments of chromatids cross over.**

DISORDERS CAUSED BY AN ABNORMAL NUMBER OF CHROMOSOMES

Our ability to associate clinical disease with detectable changes in genetic material was established in patients who were identified as having an abnormal number of chromosomes, including trisomy 21 and the sex chromosomes. Before the development of techniques for identifying and separating individual chromosomes from cell preparations, these patients were described clinically on the basis of a shared constellation of congenital anomalies and dysmorphic features (i.e., a syndrome). Individuals with Down syndrome have characteristic facial features; they experience hypotonia in infancy, delayed development, and cognitive impairment, as well as a pattern of congenital malformations. With the ability to karyotype individuals, rarer abnormalities of whole chromosomes were detected in dysmorphic stillborn infants or in live-born infants who subsequently died early in infancy (for example, from trisomy 18 or 13 syndrome) and were detected in analyses of

first-trimester abortuses, which in general have a 50% rate of chromosome abnormalities.

The gain or loss of an entire chromosome is generally the result of nondisjunction or the missegregation of chromosomes at the time of cell division (i.e., in meiosis or mitosis). This results in one daughter cell having two copies of a particular chromosome and in the other daughter cell having no copy. Fertilization of a germ cell with two copies of a single chromosome results in a zygote trisomic for that chromosome. Three autosomal trisomy syndromes have been described in live-born infants: the syndromes associated with trisomies 13, 18, and 21. Occasionally, trisomies of other autosomes occur, but there is usually a normal cell line present as well (mosaicism). In most cases, this is the result of a postzygotic segregation error in mitosis that occurs early in embryogenesis. Extra chromosomal material is better tolerated than missing material; there are no viable autosomal monosomy syndromes. The lack of one of the sex chromosomes is deleterious; the lack of a single X chomosome is lethal. Having a single X chromosome without a Y chromosome (Turner syndrome, which is associated with karyotype 45,X) results in a high proportion of fetal wastage. Triploidy and tetraploidy result in abnormal embryogenesis, and a haploid conceptus has never been reported.[5]

The development of techniques for chromosome banding allowed the identification of individual chromosomes by banding pattern rather than merely by size. This enabled the detection of the addition, loss, or rearrangement of large groups of genes by means of changes in chromosome appearance; thus, translocations, inversions, duplications, isochromosomes, and ring or marker chromosomes were described. These changes may or may not have an effect on phenotype, depending on whether there is a net conservation of genetic material, but they can have profound effects on reproductive fitness, affecting the process of chromosome segregation in meiosis. In approxi-

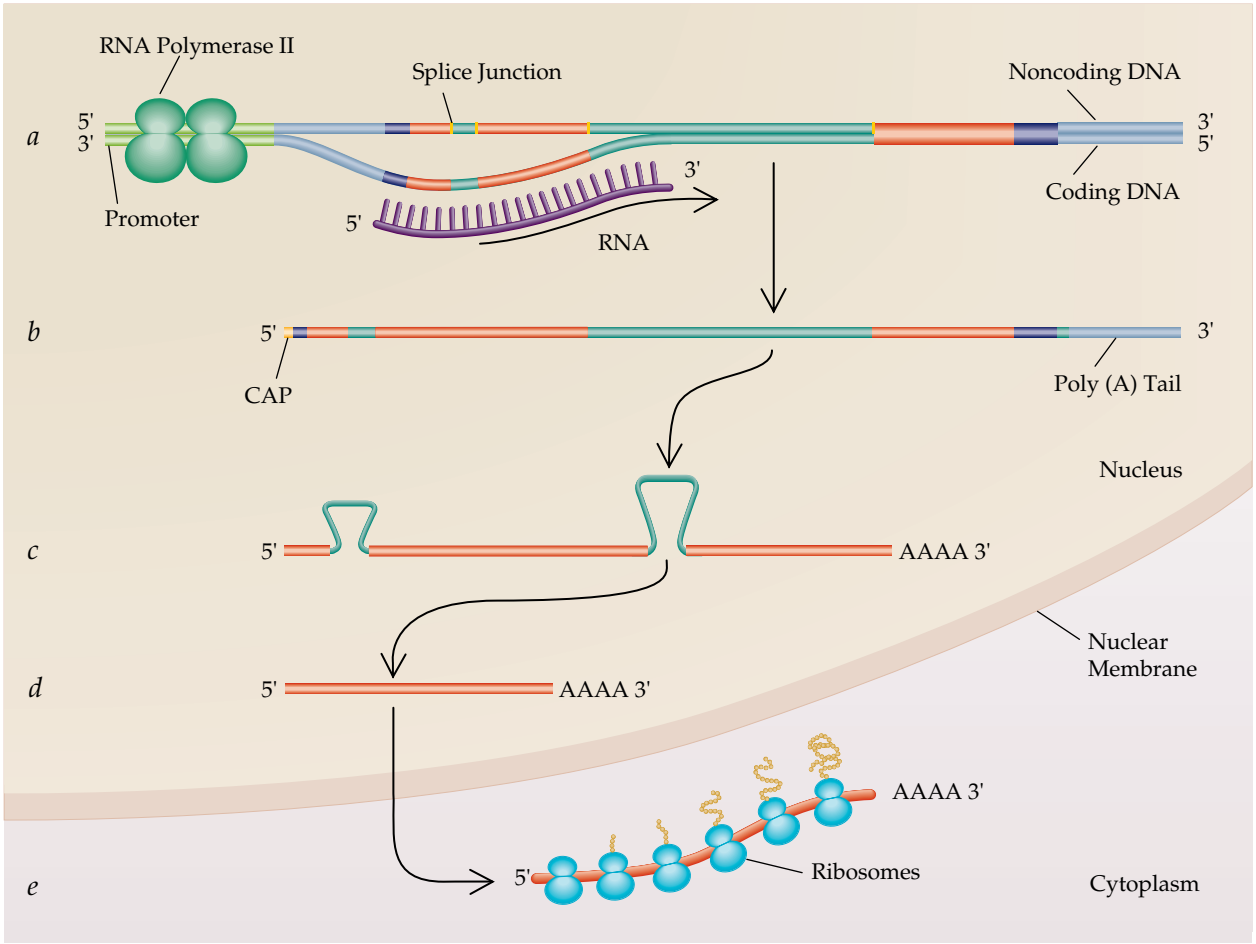


Figure 4 **(*a*) The β-globin gene contains three exons (orange) separated by two introns (green). The boundaries between exons and introns are known as splice junctions and contain specific nucleotide sequences that are required for proper joining of the exons. The synthesis of messenger RNA (mRNA) from the β-globin gene proceeds in a 5′ to 3′ direction. The enzyme RNA polymerase II (dark green) binds to a promoter region (light green) located 200 to 300 base pairs in the 5′ direction or located upstream of the point at which mRNA synthesis begins. (*b*) mRNA begins with a 7-methylguanosine residue, referred to as the CAP site, and includes a 5′ untranslated region (light purple), a coding region of exons and introns, and a 3′ untranslated region (light purple). Nearly all mRNAs that encode proteins terminate at their 3′ ends with a string of approximately 200 adenine residues [known as the poly (A) tail], which are added 18 to 20 base pairs downstream from an AAUAAA signal in the 3′ untranslated region. (*c*) After mRNA is synthesized but before it leaves the nucleus, the introns are excised and the exons are spliced together to form mature mRNA (*d*). (*e*) Once the mature mRNA reaches the cytoplasm, it attaches to ribosomes and is translated into protein.**

mately 5% of couples with a history of three or more first-trimester losses, one of the partners will be found to have a chromosome abnormality; thus, karyotype analysis is indicated in such persons.[6]

DISORDERS OF PARTIAL CHROMOSOME DELETION

22q11 deletion syndrome is a microdeletion syndrome that is common (occurring in one in 4,000 persons) and unique, in that most cases are de novo, not inherited from an affected individual. Persons with 22q11 deletion syndrome have variable clinical features, including (1) congenital heart disease (occurring in 74% of patients), particularly conotruncal malformations, such as tetralogy of Fallot, interrupted aortic arch, and truncus arteriosus; (2) palatal abnormalities (69%), notably velopharyngeal incompetence, submucosal cleft, and cleft palate; (3) characteristic facial features, including auricular abnormalities, hypoplastic alae nasi with a bulbous nasal tip, prominent nasal root, malar flatness, and hooded eyelids (> 50%); and (4) learning disabilities (70% to 90%).

Before the identification of the 22q11 microdeletion, patients were diagnosed on the basis of clinical features. The condition went under several names, including velocardiofacial syndrome, DiGeorge syndrome, Shprintzen syndrome, CATCH-22 (*c*ardiac defects, *a*bnormal facies, *t*hymic hypoplasia, *c*left palate, *h*ypocalcemia), Cayler syndrome, and conotruncal anomaly face syndrome. Velocardiofacial syndrome was originally described as the combination of velopharyngeal incompetence, congenital heart disease, characteristic facial features, and developmental delay. DiGeorge syndrome, which includes the previously mentioned features as well as parathyroid deficiency and immune dysfunction from thymic aplasia or hypoplasia, was thought to be a developmental field defect of the third and fourth pharyngeal pouches.

In 1992, the first report of a microdeletion of chromosome 22 at the 11.2q band was reported; this was subsequently confirmed in other cases. In approximately 15% of cases, a visible deletion can be seen. Deletion 22q11 is diagnosed in individuals with submicroscopic deletions by use of fluorescence in situ hybridization (FISH) using DNA probes from the DiGeorge chromosome region [*see Figure 3*]. Fewer than 5% of patients with clinical features of deletion 22q11 have normal results on cytogenetic studies and negative results on FISH testing.[7]

The typical deletion encompasses three million base pairs, with smaller deletions of several hundred thousand base pairs reported. However, there is no correlation between the size of the deletion and the expression of the syndrome. It is still unknown whether the syndrome is a contiguous gene syndrome or whether the majority of the phenotype is the result of a single gene deletion that is variably expressed in affected individuals. The broadness of the phenotype and the unification of the various above-mentioned syndromes under the umbrella of deletion 22q11 have created some confusion. All cases of velocardiofacial syndrome, DiGeorge syndrome, Cayler syndrome, and conotruncal anomaly face syndrome that are associated with a deletion at 22q11 represent the same disorder.

The high occurrence of de novo deletions of 22q11 suggests some instability in this region. Because the overwhelming majority of patients have the same deletion in the 3 Mb (megabase) region, this area has been sequenced and carefully examined. This region contains four copies of duplicated sequence or low copy repeats, located nearer to the end points of the region. Each low copy repeat contains one or more duplicated modules, which contain duplicated markers. The presence of these low copy repeats at this typically deleted region suggests that sometimes these areas misalign; during cell division and homologous recombination, this leads to duplication of the region on one chromatid and deletion on the other. The presence of these low copy repeats, therefore, gives us some insight into the mechanisms responsible for the recurrence of this common de novo deletion involving chromosome 22.[8] Repetitive sequences contribute to the inherent instability of some chromosome regions.

DISORDERS OF SINGLE GENE MUTATIONS

Anemia, which is a common clinical problem, is an excellent example of a condition that has many causes, both genetic and environmental. As monogenic disorders, the hemoglobinopathies are varied and complex. Approximately 7% of the world's population are carriers of different inherited disorders of hemoglobin, including structural hemoglobin variants and the thalassemias, which are disorders that result from defective synthesis of the globin chains. Hemoglobin is a tetramer of two pairs of dissimilar globin chains, commonly α-globin and β-globin chains in hemoglobin A (HbA) or α-globin and δ-globin chains in HbA_2. Healthy adults can have a residual amount of HbF (fetal hemoglobin, composed of two α-globin and two γ-globin chains), which is produced during fetal life and then replaced by adult hemoglobin in the first year of life. Since the discovery of a single point mutation that leads to the amino acid substitution of valine for glutamine, resulting in sickle cell anemia, over 700 structural hemoglobin variants have been identified, the most common of which are sickle hemoglobin HbS, HbC, and HbE. The thalassemias are generally classified on the basis of the particular globin chain or chains that are inefficiently synthesized.[9]

β-Thalassemia results from defective β-globin synthesis, which leads to an excess of α-globin chains. Over 200 mutations in β-globin genes have been identified in patients with β-thalassemia. The majority of these are point mutations—the loss of one or two bases that results in the disruption of gene function at the transcriptional, translational, or posttranslational level and in the decreased synthesis of the β-globin chain. Clinically, one would expect the severity of the condition to correlate with the amount of β-globin chain produced, with homozygotes or compound heterozygotes being profoundly anemic and requiring lifelong blood transfusions and heterozygotes having a milder or silent condition. Sibship studies have demonstrated phenotypic diversity in family members with the same genotype. This diversity may be a reflection of the inheritance of mutations in other loci involved in globin synthesis, because mutations for thalassemias and structural hemoglobinopathies occur together at a higher frequency in many populations. Combinations of structural hemoglobinopathies may positively alter the phenotype of thalassemia and reduce the concurrence of α- and β-thalassemia mutations in an individual; the occurrence of this process can vary from individual to individual in families. In addition, there are several mutations in the β-globin gene cluster and in the promoter region of the γ-globin genes that result in the persistence of fetal hemoglobin; such persistence produces a milder phenotype overlaying either a structural hemoglobinopathy or a β-thalassemia. Finally, mutations or polymorphisms in genes involved in bilirubin, iron, and bone metabolism may play a role in the clinical course of the disease in affected individuals.[10]

IMPRINTING DEFECTS

Gregor Mendel reported that the outcomes of reciprocal crosses were independent of the parental origin of a trait. In the late 1980s, however, researchers discovered that the two parental genomes are not equivalent in mammals. In the mouse zygote, the two pronuclei are distinct from one another and can be individually removed from the cell, and a zygote containing two female pronuclei or two male pronuclei can be created (a phenomenon known as uniparental disomy). Early embryonic development in such zygotes is abnormal. Purely female-derived embryos have poorly developed extraembryonic tissues, and purely male-derived embryos demonstrate abnormal embryo development. This phenomenon occurs sporadically in human conception, when a sperm fertilizes an egg without a pronucleus. This causes a doubling of the sperm chromosomes. The resulting diploid conceptus is a hydatidiform mole, which is a mass of extraembryonic membranes without an embryo. In contrast, ovarian dermoid cysts are derived from the spontaneous division of an oocyte; this results in the duplication of the maternal genome.

This phenomenon whereby progeny phenotypes differ according to whether the genetic material is maternal or paternal in origin is called genomic imprinting. This represents an extreme situation wherein the genetic material is derived entirely from one parent. In studying this phenomenon, investigators focused on the chromosomal regions responsible for the genomic imprinting effects observed in mouse embryos. Certain regions of distinct chromosomes were found to produce markedly different phenotypes, depending on whether the two copies were inherited from one parent, resulting in duplication or deficiency of one parental complement. An imprinted allele is one whose expression is changed or silenced as it passes through a particular sex. An allele is paternally imprinted if it is not expressed when it is inherited from the father. It is maternally imprinted if it is not expressed when it is inherited through the mother. Imprinted regions have been identified in both mouse and human chromosomes; alterations in normal imprinting patterns are associated with disorders of growth and development, cell proliferation, and behavior.

It is thought that during gamete formation in mammals, some genes are altered by the methylation of certain cytosine groups in DNA. This process tends to prevent access by transcription machinery to that region of the chromosome for transcription, thus resulting in the "silencing" of that gene or genes. Whatever the process, the imprinting procedure would have to be erased during embryogenesis so that an individual could reimprint its genes according to its own sex during gametogenesis. Demethylation in embryonic cells occurs in the early cleavage divisions. Shortly after implantation, the embryonic somatic cells are methylated again, whereas the germ cells in the developing embryo are methylated later, as they develop in the gonads. An imprinting center on chromosome 15 may play a role in this process.[11]

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are two clinically distinct genetic diseases associated with deletions of the same region of chromosome 15. These syndromes are characterized by deficiencies in growth and sexual development, behavioral abnormalities, and mental retardation. Major diagnostic criteria for PWS include hypotonia; hyperphagia with resulting obesity; hypogonadism; and developmental delay. Patients with AS may have ataxia; sleep disorders; seizures; and hyperactivity with severe mental retardation. They may exhibit characteristic outbursts of inappropriate laughter.

Approximately 70% of patients with PWS and AS have a de novo 3 to 4 Mb deletion in the q11–q13 region of chromosome 15. Because this region is imprinted, the phenotypes that result from this deletion differ, depending on the allele upon which the deletion occurred. When the deletion occurs on the paternal chromosome, it results in PWS; when it occurs in the maternal copy, it results in AS. This suggests that the normal *PWS* gene is expressed from the paternal chromosome and that the normal *AS* gene is expressed from the maternal chromosome. Most of the remaining cases of PWS are the result of maternal uniparental disomy; paternal uniparental disomy accounts for only 4% of AS cases.[12] In uniparental disomy, an individual inherits both copies of a chromosome from either the mother or the father through a nondisjunction error in meiosis. Again, lack of paternal 15q11–15q13 results in PWS; lack of maternal 15q11–15q13 results in AS. Imprinting defects have been implicated in some individuals with these syndromes.

Defects in a region termed the imprinting center, located within 15q11–15q13, can change the DNA methylation and transcription activity of certain genes that reside in the region. Thus, if there is a mutation in the imprinting center, the process of activation or inactivation of the imprinted region may not occur. Because the different mechanisms are associated with different risks of recurrence, it is recommended that the diagnostic workup begin with a search for a deletion by use of chromosome analysis and FISH; if the results are normal, DNA methylation should be performed. Although the risk of recurrence of a deletion is low, a mutation of the imprinting center is associated with a recurrence risk of 50%.

TRINUCLEOTIDE REPEAT DISORDERS

Several inherited disorders are known to have a worsening phenotype in each subsequent generation of family members affected by the disease.

Fragile X Syndrome, Huntington Disease, and Friedreich Ataxia

In the early 1990s, molecular geneticists discovered a new type of mutation, first in fragile X syndrome and then in a series of inherited neurologic disorders, including myotonic dystrophy, Huntington disease, and Friedreich ataxia. The mutation involves a repeat expansion of a DNA triplet, a trinucleotide repeat in an exon, or an intron of the gene. In patients with these conditions, the normal number of repeats (in an unaffected individual) is expanded. The number of nucleotide repeats can increase in successive generations, causing disease symptoms to appear at an earlier age. The molecular basis of repeat instability is not well understood, but increased severity of the phenotype and earlier age of onset in successive generations (a phenomenon termed anticipation) are generally associated with larger repeat length. The parental origin of the disease allele can also influence expression; for most of these disorders, there is greater risk of repeat expansion with paternal transmission, although in fragile X syndrome and congenital myotonic dystrophy (see below), the maternally transmitted alleles are more prone to expansion, thereby causing more severe phenotypes. Most of the trinucleotide repeat disorders are inherited in an autosomal dominant or X-linked fashion, with the exception of Friedreich ataxia, which is an autosomal recessive disorder [*see Table 1*].[13]

Table 1 Trinucleotide Repeat Disorders

Disease	*Gene Locus/Protein*	*Repeat*	*Location*
Fragile X syndrome	Xq27.3/FMR-1 protein	CGG	Noncoding
Fragile XE syndrome	Xq28/FMR-2 protein	GCC	Noncoding
Friedreich ataxia	9q13–9q21.1/frataxin	GAA	Noncoding
Myotonic dystrophy 1	19q13/myotonic dystrophy protein kinase	CTG	Noncoding
Myotonic dystrophy 2	3q21	CCTG	Noncoding
Spinobulbar muscular atrophy	Xq13–Xq21/androgen receptor	CAG	Coding
Huntington disease	4p16.3/huntington	CAG	Coding
Dentatorubral-pallidoluysian atrophy	12p13.31/atrophin-1	CAG	Coding
SCA type 1	6p23/ataxin-1	CAG	Coding
SCA type 2	12q24/ataxin-2	CAG	Coding
SCA type 3 (Machado-Joseph disease)	14q32.1/ataxin-3	CAG	Coding
SCA type 6	19p13/α-1A (voltage-dependent calcium channel subunit)	CAG	Coding
SCA type 7	3p12–3p13/ataxin-7	CAG	Coding
SCA type 8	13q12/none identified	CTG	?
SCA type 12	5q31–5q33	CAG	Noncoding

SCA—spinocerebellar ataxia

Myotonic Dystrophy

Myotonic dystrophy is a trinucleotide repeat disorder resulting in multisystem involvement of skeletal and smooth muscle, as well as involvement of the eye, heart, endocrine system, and central nervous system. The disorder represents a continuum of clinical findings; it has been classified for diagnostic purposes into three somewhat overlapping phenotypes: mild, classic, and congenital. Mild myotonic dystrophy is characterized by the development of cataracts in early adulthood and mild myotonia (difficulty relaxing the muscles after contraction). The symptoms may be so subtle that diagnosis is made retrospectively, after the birth of an affected offspring. Classic myotonic dystrophy is characterized by muscle weakness and wasting, myotonia, cataract formation, and cardiac conduction abnormalities that occur in adulthood. The life span of these patients may be somewhat shorter than normal.

Congenital myotonic dystrophy is a disease of the neonate characterized by generalized hypotonia, respiratory insufficiency requiring ventilatory support, and mental retardation if the infant survives to childhood. Individuals have characteristic facial features, which include drooping eyelids, facial weakness resulting in an open-mouthed appearance, and wasting of the muscles in the jaw and neck. The overall incidence of myotonic dystrophy is estimated to be one in 20,000 persons.

The diagnosis of myotonic dystrophy is confirmed by detection of an expansion of the CTG trinucleotide repeat that affects the noncoding regions of two adjacent genes (*DMPK* and *SIX5*) on chromosome 19q13. Normal individuals have a repeat of 37 trinucleotides or fewer. The trinucleotide repeat is located at the 3′ end of the gene (the transcription occurs in the 5′ to 3′ direction), but it is in a part of the gene that is transcribed but not translated into the final protein product. Unaffected individuals have a polymorphic repeat length of 5 to 37 CTG repeats; this repeat length is stable when passed from generation to generation. Stability is disrupted, however, when the number of repeats exceeds 37. When the number exceeds 37, this repeat expansion not only disrupts the function of the gene but also engenders further instability and larger expansions. This tendency accounts for the phenomenon of anticipation seen in families with this disorder, in which a mildly affected adult can give birth to a child with the congenital form of the disease. In rare cases, the region will contract, with the CTG repeat being smaller in an offspring. In affected individuals, further expansion can occur during somatic cell division, resulting in mosaicism from tissue to tissue.[14]

The gene product of *DMPK* is a protein kinase. It is expressed in the different organs involved in the disease: skeletal muscle, the heart, the brain, and the testes. The function of the protein is unknown, and it is unclear how the expansion in the untranslated region leads to the phenotype. There is evidence, however, that the mutant *DMPK* transcripts accumulate abnormally in the nuclei and bind to RNA-binding proteins, thus disrupting RNA splicing and metabolism. A second form of myotonic dystrophy has recently been described. This form involves a chromosome 3 trinucleotide repeat, which also results in the accumulation of RNA in cells.[15]

This disorder presents a unique genetic counseling problem. Individuals who are mildly affected have a 50% chance of passing on the expanded allele to their offspring. However, because of the instability of the expanded region, it is impossible to predict the severity of the condition in an affected child. There is a risk of having a child with the severe form of the disease—congenital myotonic dystrophy—only if the mutation is transmitted through the mother. Approximately 20% of the offspring of an affected mother who inherit the mutation manifest the severe form, depending on the size of the expansion in the mother. Although prenatal testing for the expansion is available, often the diagnosis of mild myotonic dystrophy in the mother is established only after the birth of an infant with the congenital form of the disease.

Mitochondrial Inheritance

In general, the inheritance (autosomal dominant, recessive, or X-linked) of a condition can be determined by pedigree analysis of large families. Mechanisms such as imprinting and

anticipation affect the expression of mutations, but transmission still abides by classic Mendelian patterns of inheritance. Recently, mutations in mtDNA have been associated with a number of disorders with a unique inheritance pattern, termed maternal transmission. In maternal transmission, a condition affects individuals in each generation, suggesting dominant inheritance. Males and females may be affected, but men never transmit the disorder to their offspring. Women pass the trait on to all of their children, although there is great variability in expression.

Mitochondria are the cellular organelles responsible for the generation of energy in the form of adenosine triphosphate (ATP) through aerobic metabolism. Some mtDNA molecules are encoded in the nuclear genome and are transported out to the mitochondria, but a minority are encoded and synthesized in the mitochondria. mtDNA is a circular, double-stranded structure without introns; it resembles a prokaryotic genome. It contains 16,569 base pairs that encode at least 13 proteins required for oxidative phosphorylation. In addition, it contains the transfer RNA (tRNA) and ribosomal RNA (rRNA) involved in the translation of these proteins in the organelle.

The manner in which mitochondria are passed from one generation to the next accounts for the phenomena of maternal transmission. At the time of fertilization, the sperm sheds its cytoplasm, and only the nuclear DNA enters the egg. Therefore, all mitochondria in the zygote are contributed by the egg cell. However, there are hundreds of copies of mtDNA in each cell. During cell division, each mtDNA replicates, but unlike nuclear DNA, the newly synthesized mitochondria segregate passively to the daughter cells. A mitochondrial mutation arises randomly, and chance segregation leads to an accumulation of mutant mitochondria in a cell. The phenotypic expression of a mutation in mtDNA depends on the relative proportion of normal functioning product above a certain threshold value for manifesting the phenotype. This unpredictability in phenotype from individual to individual is the result of this random segregation of mitochondria; such random segregation of mitochondria is termed heteroplasmy. Additionally, different types of mtDNA mutations are inherited differently; deletions occur only sporadically and are not transmitted from affected females to their offspring. Examples of disorders associated with such deletions are chronic progressive external ophthalmoplegia, Kearns-Sayre syndrome, and Pearson syndrome. Point mutations result in mitochondrial encephalomyopathy with lactic acidosis and strokelike episodes (MELAS), myoclonic epilepsy with ragged-red fibers (MERRF), and neuropathy, ataxia, and retinitis pigmentosa (NARP).[16]

The number of mitochondria per cell is dependent on the energy production in the cell. The brain, retina, and muscle cells have a relatively higher demand for energy than other cell types. Thus, mutations in mtDNA tend to result in disorders with muscle and neurologic dysfunction, such as myopathies, cardiomyopathy, ophthalmoplegia, encephalopathies, and encephalomyopathies. Defects in oxidative phosphorylation should be considered in the differential diagnosis of any patient with unexplained multisystem involvement, including a progressive myopathy or neurologic problem. Although routine laboratory testing in such patients may reveal hypoglycemia, abnormal liver function, or elevated blood lactate levels, muscle biopsy is often necessary to make a diagnosis. The abnormal pathologic appearance of the cellular mitochondria is apparent on electron microscopy and through the use of special staining. Some examples of conditions resulting from mutations in mtDNA are listed [*see Table 2*]. Defects in oxidative phosphorylation may also result from mutations in nuclear DNA that code for mitochondrial proteins. Diagnosis of these disorders is difficult, and subsequent counseling issues can be complex.

Cancer Genetics

Thus far, we have focused on genetic diseases with specific phenotypes associated with the presence of germline mutations that lead to expression of one or more abnormally functioning proteins. This type of mutation is presumably present in all the cells of an individual from birth, although there can be a degree of mosaicism, depending on the stage of development at which the mutation occurs. Investigation into the control of cell growth has given new insight into genetic changes that occur in both germ cells and somatic cells and that can lead to malignancy. Mutations in three types of genes that regulate cell growth are involved in the development of cancer: tumor suppressor genes, proto-oncogenes, and DNA repair genes. Somatic mutations in these genes may result in unchecked proliferation or clonal expansion of a single cell with subsequent loss of cellular organization; somatic mutations may also confer the ability to metastasize. This process generally requires a number of mutations, because there is an elaborate backup system in place to prevent faulty cell proliferation. Although sporadic mutations arise in individual somatic cells and ultimately play a role in cancer development, the study of familial or inherited cancer syndromes has contributed to our understanding of the genetic changes responsible for the development of some of the more common cancers.

It is perhaps easiest to understand how a germline mutation in a tumor suppressor gene could lead to a predisposition to

Table 2 Disorders Resulting from Mitochondrial Mutations

Disease	*Phenotype*
Chronic progressive external ophthalmoplegia	Progressive weakness of extraocular muscles; ptosis
Kearns-Sayre syndrome	Progressive external ophthalmoplegia before age 20; pigmentary retinopathy and CSF protein > 1g/L; cerebellar ataxia or heart block
Leber hereditary optic neuropathy	Rapid bilateral optic nerve death resulting in loss of central vision in early adulthood
Mitochondrial encephalopathy with lactic acidosis and strokelike episodes	Strokelike episodes before age 40, seizures, dementia, ragged-red fibers, lactic acidosis
Myoclonic epilepsy with ragged-red fibers	Myoclonus, epilepsy, ataxia, myopathy, sensorineural deafness
Pearson syndrome	Sideroblastic anemia, pancytopenia, pancreatic insufficiency, lactic acidosis
Leigh disease	Neuropathy, ataxia, retinitis pigmentosa, developmental delay, lactic acidemia
Deafness	Progressive sensorineural deafness often induced by aminoglycoside antibiotics

cancer. Such is the case in families with an inherited mutation in *BRCA1* and *BRCA2*. Such individuals have only one functional copy of the gene; a subsequent somatic mutation in the normal copy in a single cell gives rise to a population of cells that have no *BRCA1* or *BRCA2* gene and have therefore lost the tumor suppressor activity that limits cell proliferation. In individuals with a germline mutation in *BRCA1,* the chance of developing breast cancer over one's lifetime is estimated to be 80%; the chance of developing ovarian cancer is 15%. The cancer usually develops at an earlier age than is seen in the general population, and there can be multiple primary sites. Nevertheless, most breast cancer disease is sporadic, and the disease is common enough that family history may be misleading, particularly in a large family. There are algorithms for assessing a patient's risk of developing breast cancer, as well as the risk of carrying a germline mutation. Such risk-assessment algorithms are based on personal health history and family history of breast cancer, ovarian cancer, or both. With regard to family history, important factors include the age of onset in affected individuals and whether there was more than one primary site. Verification of the family member's medical records is imperative.[17]

Proto-oncogenes are recessively acting genes that regulate the cell cycle. Mutant dominant genes, called oncogenes, are usually gain-of-function mutations; the altered products of such mutations cause uncontrolled cell proliferation. Oncogenes were discovered by transformation experiments in tissue culture. In these experiments, normal cells were made into malignant cells by the insertion of a mutant piece of DNA (the oncogene). A number of proto-oncogenes have been located in the human genome, and mutations in them have been implicated in the development of leukemias, lymphomas, breast and ovarian carcinomas, and cancer of the colon, thyroid, lung, and pancreas. Many oncogenes are caused by chromosomal changes that result from breakage and translocations occurring as cell proliferation becomes more disorganized. The so-called Philadelphia chromosome seen in chronic myelogenous leukemia is a translocation between chromosomes 9 and 22. The breakpoint in chromosome 9 occurs in the cellular proto-oncogene *ABL,* which normally codes for a tyrosine kinase that binds to DNA. The breakpoint in chromosome 22 is in a gene called *BCR,* or breakage cluster region, which codes for a serine kinase. The fused *BCR-ABL* gene in the Philadelphia chromosome makes a novel protein, which leads to unregulated proliferation of hematopoietic stem cells and chronic myeloid leukemia.[18]

Although most cases of colorectal cancer are sporadic, there are two more common forms of autosomal, dominantly inherited colorectal cancer, familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer, which together account for about 10% of cases of colorectal cancer. FAP is caused by a germline mutation in a tumor suppressor gene, the *APC* gene. Loss of function of the second *APC* allele leads to adenoma formation and progression to cancer through the accumulation of other somatic mutations.

Supporting the concept that accumulated genetic changes underlie the development of neoplasia, the defect in hereditary nonpolyposis colorectal cancer (HNPCC) was found to be in a group of genes that function in DNA mismatch repair. In the course of normal cell division, DNA replication is subject to error, although, as discussed earlier, the fidelity of DNA polymerase is quite good. The mismatch repair genes, of which at least six are known (*MSH2, MSH6, MLH1, MLH3, PMS1,* and *PMS2*) function in DNA replication errors resulting from misincorporation. The mismatch repair proteins function as a complex to recognize the deformation in the double helix and to then recruit enzymes to correct the error. Without these proteins, errors are propagated in successive generations of cells. Individuals with a germline mutation in the mismatch repair system, again, may undergo loss of function in the second mismatch repair allele, resulting in the characteristic microsatellite instability (MSI). Microsatellites are repeating DNA sequences of unknown function that are found throughout the genome. These repetitive sequences are more prone to errors in replication. Loss of mismatch repair mechanisms permits expansion of these repeats, as may be demonstrated in tumor specimens. The presence of MSI indicates an increased likelihood of HNPCC, although MSI is seen in 15% of sporadic colorectal cancers. Families with HNPCC have an increased incidence of other types of cancers, including endometrial, ovarian, upper GI, renal, pelvis, and brain cancers.[19]

Human Genome Project

The Human Genome Project was undertaken to determine the sequence of the entire human genome. It is a massive effort that will have major effects on medical research and practice. With the advent of recombinant DNA technology and the development of DNA sequencing, the question arose in the 1980s as to whether the genome might be sequenced. By 1990, the National Institutes of Health had founded a National Institute of Human Genome Research. The projected date of completion of the sequencing of the human genome was 2005; advances in technology, chiefly in terms of sequencing, were incorporated into the time estimate. As it turned out, the genome was for the most part completed by 2000; some 3% to 5% of the genome remained uncertain because of high redundancy.[20,21] However, the remaining sequences are not thought to contain many meaningful coding sequences.

The earlier-than-expected completion was in part the result of a strong organization and cooperation among the centers. Investigators around the world who had linked constructs of DNA sent them to genome centers for sequence determination. The government-sponsored project required that such sequences be made public within a day, thus aiding dissemination of information. A second factor in timely completion was that technology improved even faster than anticipated. A third factor was the cooperative interaction of private industry, which participated under the original NIH rules. A fourth factor contributing to early completion was the competition of private industry using a modified approach to sequence determination. By 1998, 3% of the sequence was actually known. Three years later, more than 95% of the sequence had been determined.

Hybridization, or complementary pairing, of single strands of DNA to each other, allows identification of overlapping DNA fragments by labeling the short fragment and hybridizing to a library. The genome of a person contains regions that vary in a manner specific to that individual. By definition, a variation present in 1% of the population is termed a polymorphism. For the Human Genome Project, the most useful of these turned out to be blocks of CA sequence, which vary in the number of times the dinucleotide is repeated.[22] The repeats serve as signposts in the genome of an individual. As the genome project progressed, these, along with

other markers, served to build a map of signposts along the genome.

Amplification of DNA sequences for cloning and for utilization of the markers depended on the polymerase chain reaction technique.[23] If the sequence of a region of DNA is known, then a large amount of the sequence can be cloned and mapped by the use of excess primers, which hybridize to the specific sequence. Usually, the primers are about 20 nucleotides long; this length is sufficient to give adequate specificity in binding. The amplification depends on the repeated denaturing and renaturing of the DNA, so that the excess primers create a site for DNA synthesis in each cycle. With a thermoresistant DNA polymerase, the progress can be automated [*see Figure 5*].

Another major technical advance facilitating positional cloning and the genome project was the cloning of large blocks of human genome sequence. Stable vectors containing up to 1 Mb of genome sequence were developed using yeast artificial chromosomes.[24] However, these tend to recombine the DNA frequently; bacterial artificial chromosomes (BACs) proved to be more useful for the genome project because they are more stable. As libraries of BACs were made, the assignment of markers to each BAC allowed the development of a map for signposts and the establishment of contiguous sequences (contigs) as subunits of the overall genome. These were frequently verified by cytogenetics with the use of FISH.

As the Human Genome Project developed many markers along the genome, a second technique became feasible—

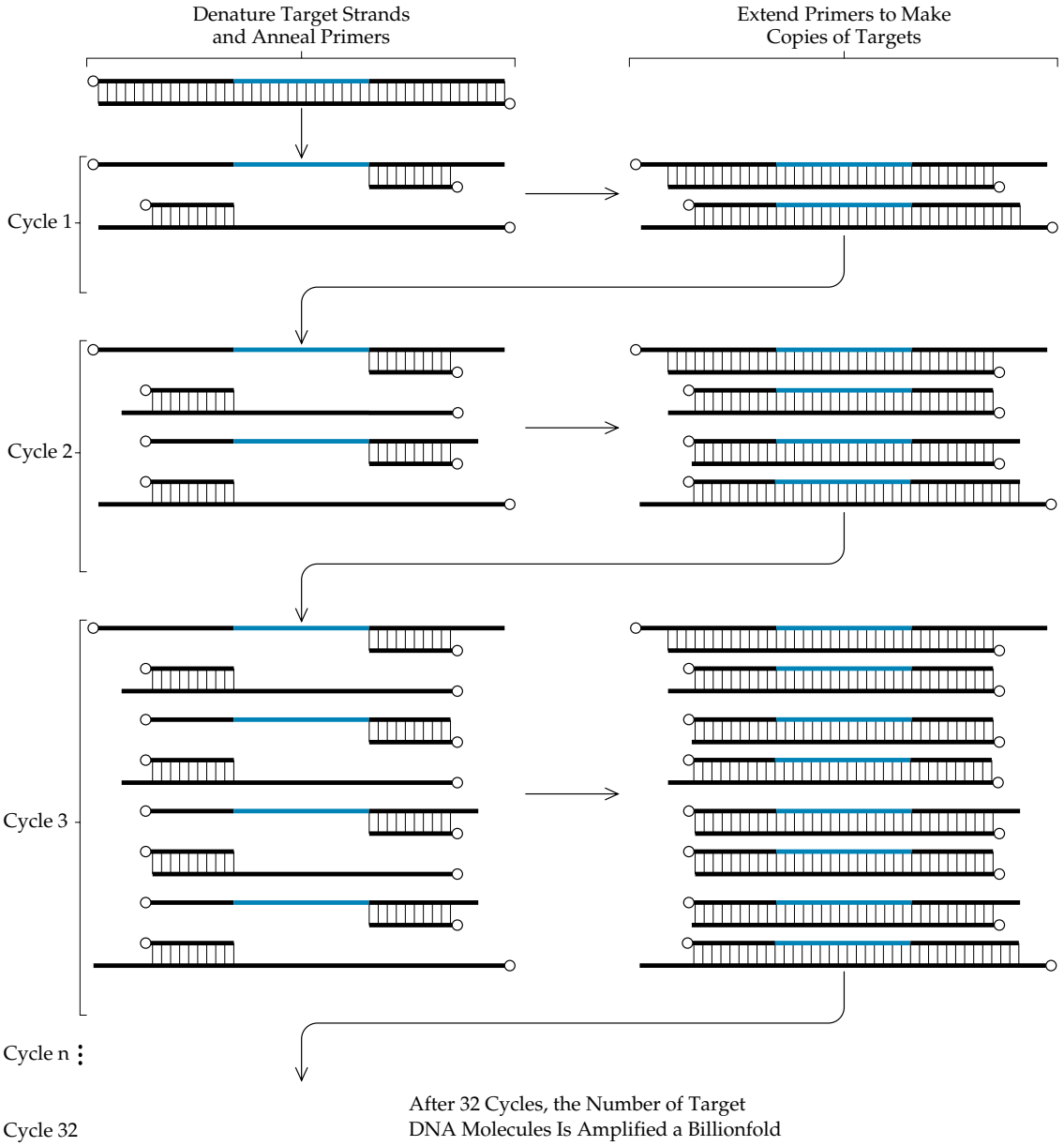


Figure 5 **The polymerase chain reaction. The strands in each targeted DNA duplex are separated by heating and then cooled to allow single-stranded oligonucleotide primers that are complementary to the end sequences of the opposite strands to bind to those sequences. DNA polymerase extends the primers (i.e., adds nucleotides), using the target DNA strands as templates. In this way, duplicates of the original DNA strands are produced in each cycle, and the quantity of the target DNA duplex increases exponentially.**

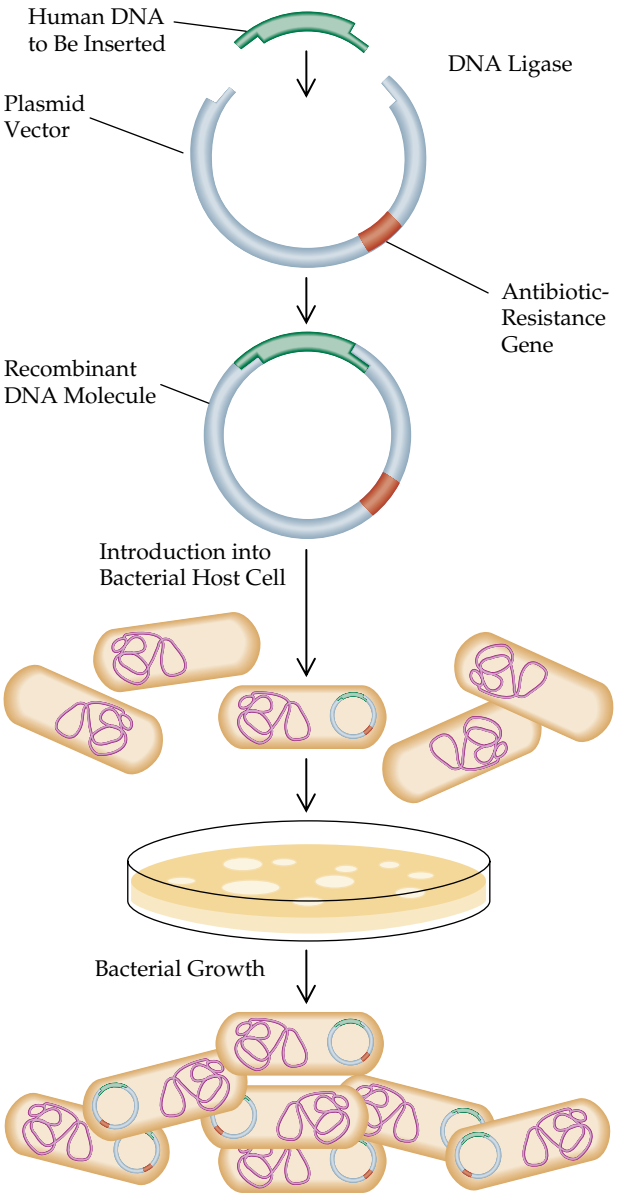


Figure 6 **Restriction fragments of human DNA generated by digestion of chromosomal DNA with a restriction endonuclease (e. g., *Eco*RI) can be inserted into a cloning vector, such as an *Escherichia coli* plasmid. The plasmid contains a single recognition site for *Eco*RI and a bacterial gene that confers antibiotic resistance. After EcoRI cuts the plasmid, the fragments of human DNA and the linear plasmid can be covalently joined by the action of the enzyme DNA ligase. The plasmid can replicate within a bacterial host to produce many copies of the DNA fragment, and the bacteria themselves reproduce. The bacteria are grown on plates containing antibiotics to select for those cells that contain copies of the plasmid with the human DNA insert. The human DNA (e.g., a gene) can then be recovered when the harvested plasmid DNA is digested with *Eco*RI.**

shotgun sequencing. In that approach, the genome was sequenced repetitively in small fragments; with appropriate computer algorithms, it was then possible to reconstruct the sequence. This method was aided by the existence of the knowledge of the markers that had already been obtained. Regions of redundancy present problems for the combined methodologies.

UNRESOLVED QUESTIONS

Currently, the human genome appears to contain approximately 38,000 genes. In terms of complexity of function, it is thought that protein modification may be the basis for varied activities of gene products. Only about 1% of the genome codes directly for information. About 24% is intronic sequence; the remaining 75% is intergenic. The genes are not distributed at an average frequency but are clustered in gene-rich regions, and there is significant variability among the chromosomes as to gene content per unit length. The human genome contains many retroviruses, or transposable elements, most of which do not appear to be active; however, many can function as mobile genetic elements in the genome. The results emphasize the presence of introns and lead to the question of whether there is additional information encoded in them, other than by the triplet code. For example, do these introns contain topologic or conformational information that is important in gene regulation?

Another remaining difficulty is that we do not understand the rules of genomic "punctuation." With genes embedded in small exons occupying less than 1% of the genome, recognition of the coding regions is difficult. Identification of genes is based on the assumption that genes are expressed and usually converted to mRNA. Because mRNA has a poly-A 3′ tail, it is possible to capture portions of messages. These snippets are termed expressed sequence tags; they were used to identify the signposts for genes in the genome project. The portion of DNA coding for the desired sequence was expanded by cloning the DNA into plasmid vectors and then growing these vectors in bacteria [*see Figure 6*]. Cloning depends on restriction enzymes that cut double-stranded DNA at specific sites; recognition is usually based on a sequence of 4 to 8 bases, and the cleavage may produce overhang or blunt termini in the DNA. If the same enzyme is used to cut the plasmid vector, the DNA piece may be joined to the plasmid by ligase enzyme, and the plasmid will thus be able to replicate with the complementary

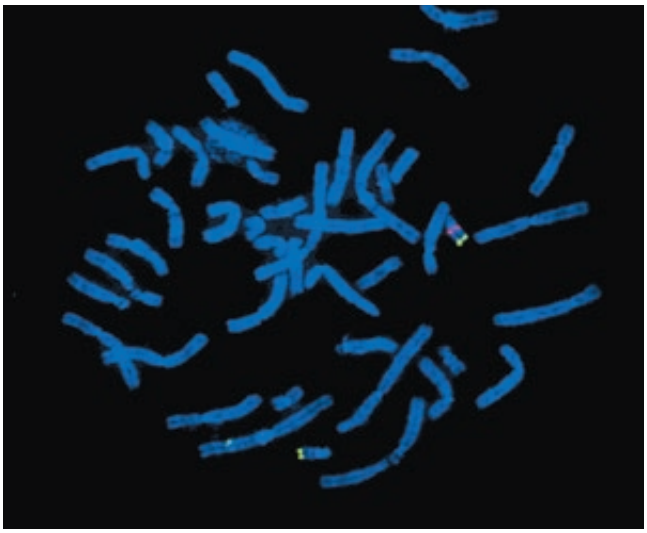


Figure 7 **Fluorescence in situ hybridization (FISH) was performed using the TUPLE1 probe (red) for the VCFS/DGS region at 22q11.2 along with an identifier probe (green) (Vysis, Inc., Downers Grove, Illinois). One homologue has both probe signals (red and green); however, the other homologue has only the identifier signal (green), which indicates that this homologue is deleted.**

Selected Internet Resources for Genetic Information

PubMed

http://www.ncbi.nlm.nih.gov/PubMed

A convenient entry point to search for publications or to search the genome database for sequence information. Published by the National Library of Medicine.

The Online Mendelian Inheritance of Man (OMIM)

http://www.ncbi.nlm.nih.gov/Omim

Provides a compilation of heritable diseases and a summary of the clinical and molecular information relating to them. It is organized under the direction of the National Center for Biotechnology Information.

GeneClinics

http://www.geneclinics.org

Provides information for clinicians regarding molecular testing laboratories. This resource is helpful in patient evaluation and in locating laboratory testing sites for families with known diagnoses or risk of genetic diseases.

American Society of Human Genetics

http://gopher.faseb.org/genetics/ashg/ashgmenu.htm

Provides access to electronic publications and links to other sources of information.

DNA (cDNA) insert [*see Figure 7*]. Plasmids were engineered that would exist in the bacterial host cell in groups of several hundred copies each; in this way, the DNA insert was amplified, and such amplifed inserts were relatively easy to rescue for study.

Our knowledge of human genetics and its application to clinical medicine is constantly evolving. We have progressed from inferring inheritance modes by pedigree analysis and from inferring risk to future offspring by probability calculations to molecular testing based on the identification of mutations in a gene or genes involved in a specific disorder. The human genome has been sequenced, but it still must be deciphered as to the genetic bases of the remaining single-gene disorders, the genetic component of multifactorial inheritance conditions, and the function of noncoding DNA. Aided by improvements and advances in molecular technology, scientists will have this task in the 21st century.

FUTURE APPLICATIONS OF GENETICS TO MEDICINE

Scanning the Genome for Risk of Disease

Technology holds the promise of detailed analysis of the genomes of individuals with attention to particular areas. By combining computer-chip design with DNA hybridization techniques, arrays of DNA sequences containing many thousands of specified sequence variations can be made.[13] This will allow searching for disease-specific mutations or associated polymorphisms in a person. However, because it appears that most common diseases have a genetic component but are made manifest on the basis of other factors (multifactorial disease), so-called array analysis offers a new tool. The human genome contains single base variations—single nucleotide polymorphisms (SNPs) that occur at a rate of about one per 1,000 bases; there are close to three million SNPs in the human genome. Of these, perhaps 1% are in exons and can be used to identify disease risk by linkage. As associations with multifactorial disease are made, scanning for markers linked to risk—even though, at the molecular level, the basis of the risk remains unknown—will allow determination of the apparent risk of multifactorial disease in an individual patient. It is expected that array analysis will prove useful in assessing the risk of diabetes, heart disease, cancers, and other common diseases.

Identifying Drug Responsiveness

A second avenue of use for the complex analysis of individual genomes may come with regard to drug prescription. Associations of drug responsiveness and genome markers will develop. It seems likely that genome variations affect a patient's response to drug therapy and that such variations may thus have a role in drug selection and dosage schedules. Medical practice may thus come to utilize an array analysis for a given drug.

Privacy Issues in Genetics

The testing capabilities and the ability to store and compare sequence data raise ethical concerns. To a large degree, these questions are not new in medicine, but the extent of the knowledge and the possible predictive nature of the information make the issue one of new focus and attention. Collection, storage, and dissemination of an individual's genetic information have become topics for discussion at the state and national level. Already a number of states have revised statutes regarding privacy. The question of privacy in a time of electronic records is in itself a difficult one for health care providers. Access to records is a thorny issue. Added to this are concerns over the availability of health care insurance and life insurance for individuals with a family history of genetic disease. With patient profiles that include a large number of disease-causing sequence alterations now a reality, the problem has only become larger.

Online Resources for Genetic Information

Several Web-based sites for information regarding the genome or genetic diseases are available [*see Sidebar* Selected Internet Resources for Genetic Information].

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Henig RM: The Monk in the Garden. Houghton Mifflin Company, Boston, 2000

2. Avery OT, MacLeod CM, McCarty M: Induction of transformation by a deoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med 79:137, 1944

3. Watson JD, Crick FC: Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature 171:737, 1953

4. Crick F: The genetic code. Sci Am 207:66, 1962

5. Korf BR: Chromosomes and chromosomal abnormalities. Human Genetics: A Problem Based Approach, 2nd ed. Malden MA, Ed. Blackwell Science, Oxford, 2000, p 181

6. Olson SB, Magenis RE: Cytogenetic aspects of recurrent pregnancy loss. Semin Reprod Endocrinol 6:191, 1988

7. McDonald-McGinn DM, Emanuel BS, Zackai EH: 22q11 deletion syndrome, September 1999 http://www.geneclinics.org

8. Shaikh TH, Kurahashi H, Emanuel BS: Evolutionarily conserved low copy repeats (LCRs) in 22q11 mediate deletions, duplications, translocations, and genomic instability: an update and literature review. Genet Med 3:6, 2001

9. Tuzmen S, Schechter AN: Genetic diseases of hemoglobin: diagnostic methods for elucidating β-thalassemia mutations. Blood Rev 15:19, 2001

10. Weatherall DJ: Phenotype-genotype relationships in monogenic disease: lessons from the thalassemias. Nat Rev Genet 2:245, 2001

11. Mange EJ, Mange AP: Complicating factors. Basic Human Genetics, 2nd ed. Sinauer Associates, Inc., Sunderland, Massachusetts, 1999, p 209

12. Falls JG, Pulford DJ, Wylie AA, et al: Genomic imprinting: implications for human disease. Am J Pathol 154:635, 1999

13. Cummings CJ, Zoghbi HY: Trinucleotide repeats: mechanisms and pathophysiology. Annu Rev Genomics Hum Genet 1: 281, 2000

14. Adams C: Myotonic dystrophy, August 2001 http://www.geneclinics.org

15. Alfred J: Myotonic dystrophy comes into focus. Nat Rev Genet 2:736, 2001

16. Chinnery PF, Turnbull DM: Mitochondrial DNA and disease. Lancet 354(suppl 1):SI17, 1999

17. Frank TS: Hereditary cancer syndromes. Arch Pathol Lab Med 125:85, 2000

18. Faderl S, Talpaz M, Estrov Z, et al: The biology of chronic myeloid leukemia. N Engl J Med 341:164, 1999

19. Kinzler KW, Vogelstein B: Lessons from hereditary colorectal cancer. Cell 87:159, 1996

20. Lander ES, Linton LM, Birren B, et al: Initial sequencing and analysis of the human genome. Nature 409:860, 2001

21. Venter JC, Adams MD, Myers EW, et al: The sequence of the human genome. Science 291:1304, 2001

22. Mullis KB: The unusual origin of the polymerase chain reaction. Sci Am 262:56, 1990

23. Botstein D, White RL, Skoltnick M, et al: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314, 1980

24. Kim UJ, Birren BW, Slepak T, et al: Construction and characterization of a human bacterial artificial chromosome library. Genomics 34:213, 1996

Acknowledgments

Figure 1 George Kelvin.

Figure 2 Seward Hung.

Figures 4, 6, and 7 Dimitry Schidlovsky.

Figure 5 Tom Moore.

VIII GENETIC DIAGNOSIS AND COUNSELING

Roberta A. Pagon, m.d.

Genetic testing is an increasingly useful, cost-effective, sensitive, noninvasive tool that allows clinicians to identify disease in symptomatic persons, predict the probability of disease in asymptomatic at-risk persons, detect carriers of heritable disorders, and diagnose genetic disease in fetuses.

Although genetic tests involve the analysis of DNA, RNA, chromosomes, proteins, and certain metabolites, the most widely used of these tests over the past decade have been DNA-based tests for the diagnosis of heritable disorders and for genetic counseling. The effectiveness of genetic tests depends on the technical skill with which they are performed, the clinical skill with which they are interpreted, and the patient's interest in the results. As with all other medical testing, genetic testing is context specific. Before recommending a test, the clinician must be able to answer the question, "Why am I testing this patient at this time?"

Genetic testing differs from traditional medical testing in two ways: genetic test results always have implications for disease risk for the individual patient and, frequently, for his or her family; and genetic testing may be used for the sole purpose of personal decision making rather than medical care.

The focus of this chapter is on germline (inherited) mutations that are present at conception and have phenotypic (clinical) effects, implications for reproduction, or both. Germline mutations contrast with somatic (acquired) mutations, which are the basis of certain acquired disorders, such as many cancers. Somatic mutations, with the exception of so-called germline mosaicism in which the mutations affect the gonads, are not heritable and are not discussed here.

Molecular Genetic Testing

Direct and indirect testing are the two broad categories of DNA-based testing. Direct testing is synonymous with mutation detection—the positive identification of disease-causing genetic alterations that establish a person's genetic status independent of knowledge of family history or prior risk status. The most commonly used methods for direct testing include the following:

1. Targeted mutation analysis, which screens for the presence of a specific mutation, a specific type of mutation, or a specific set of mutations.
2. Mutation scanning (mutation screening), which screens a segment of DNA by one of a variety of methods to identify a variant gene region or regions, which are further analyzed by sequence analysis or mutation analysis to identify the sequence alteration.
3. Sequence analysis (sequencing), which determines the nucleotide sequence for a segment of DNA. The nucleotide sequence may involve either an entire gene or a portion of a gene. Sequence analysis is considered the gold-standard genetic test by many; however, interpretation of the significance of a sequence alteration may not be straightforward, and failure to detect a sequence alteration may not indicate the absence of a disease-causing mutation [*see Table 1*].
4. Other direct testing methods, including Southern blotting, protein truncation testing, and methylation testing, have more specific applications.

The methodology selected depends on such parameters as the size of the gene, possible mutation types, and the presence of recurrent (common) mutations. The mutation-detection rate in a given disorder is likely to vary by methodology. Thus, the clinician needs to be aware of the test method used by a given laboratory and its relevance to the questions being addressed in the testing of a given patient. For example, in testing for familial adenomatous polyposis (FAP), three test types are used in varying combinations by different laboratories, each with a different mutation-detection rate [*see Table 2*].

Once a disease-causing mutation is identified in one affected family member, others at risk in the family need only be tested by targeted mutation analysis for that exact mutation. Conversely, testing at-risk family members when the disease-causing mutation has not been identified in an affected family member is problematic: detection of a disease-causing mutation is informative, but failure to detect a mutation is not informative, because one cannot distinguish between true absence of a disease-causing mutation and failure to detect the disease-causing mutation by the laboratory method employed.

Indirect testing relies on linkage analysis, in which DNA sequences serve as markers to track a gene mutation within a family. However, linkage analysis is now rarely used in patient care, having been replaced by gene sequencing and mutation scanning, which provide improved mutation detection.

In the evaluation of patients and their families, genetic-testing information can be used in a medical-testing paradigm, a genetic-counseling paradigm, or both [*see Table 3*].

Table 1 Interpretation of Sequence-Alteration Results*

Types of Sequence Alterations That May Be Detected
Pathogenic sequence alteration reported in the literature
Sequence alteration predicted to be pathogenic but not reported in the literature
Unknown sequence alteration of unpredictable clinical significance
Sequence alteration predicted to be benign but not reported in the literature
Benign sequence alteration reported in the literature
Possible Significance if a Sequence Alteration Is Not Detected
Patient does not have a mutation in the tested gene (e.g., a sequence alteration exists in another gene at another locus)
Patient has a sequence alteration that cannot be detected by sequence analysis (e.g., a large deletion)
Patient has a sequence alteration in a region of the gene (e.g., an intron or regulatory region) not covered by the laboratory's test

*Adapted from Reference 42. For more details, see http://www.acmg.net.

Table 2 Molecular Genetic Testing: Familial Adenomatous Polyposis

Test Type	*Mutation Detection Rate*
Full gene sequencing	Up to 90%
Mutation screening and protein truncation testing	~80%–90%
Protein truncation testing	~80%

Medical Paradigm of Genetic Testing

In the medical paradigm of genetic testing, genetic tests provide patients and their physicians with information that directly influences medical care. Issues of sensitivity, specificity, cost, and the risk of recurrence (i.e., the probability that a disease will recur in a family) are relevant. Sensitivity refers to the frequency with which a test yields a positive result when the person being tested is affected and has a disease-causing mutation in the gene in question. Specificity refers to the frequency with which a test yields a negative result when the person being tested is unaffected and does not have the gene mutation. Positive predictive value refers to the likelihood that a person with a disease-causing mutation will develop the condition. Diagnostic testing establishes or confirms a diagnosis in a symptomatic person.

DIAGNOSTIC TESTING IN SYMPTOMATIC PERSONS

A DNA-based diagnostic test relies on knowledge of the disease-causing genetic alterations and on the ability to detect them in readily obtainable tissue samples, usually blood.

High Sensitivity, Specificity, and Positive Predictive Value

Trinucleotide repeat diseases, caused by the presence of an abnormally large number of tandem trinucleotide repeats within a gene, are examples of diseases for which DNA-based testing is highly sensitive [*see Table 4*].[1,2] Targeted mutation analysis and, when necessary, Southern blotting measure the repeat size (i.e., number of trinucleotide repeats present). The cost of these methods is low because of the straightforward laboratory methodologies used in testing.

Because the molecular genetic basis of trinucleotide repeat diseases is known, the disease spectrum for a number of these disorders has been redefined. Thus, establishing the diagnosis requires molecular testing; the sensitivity and specificity of molecular testing for these diseases approaches 100%.

In spinocerebellar ataxia type 3 (Machado-Joseph disease), four overlapping but age-related phenotypes are recognized. The spectrum of clinical involvement ranges from spasticity or a predominance of extrapyramidal findings (i.e., rigidity, dystonia, or involuntary movements) with cerebellar findings (i.e., ataxia or ophthalmoplegia) in young patients to a predominance of parkinsonism and neuropathy in patients older than 40 years. The different clinical phenotypes all derive from mutations of the same gene.[3]

Another example of a spectrum of clinical phenotypes resulting from trinucleotide repeat mutations in the same gene is Friedreich ataxia (FDRA). In a study that examined the predictive value of the molecular test for FDRA in 187 patients with autosomal recessive childhood-onset ataxia, only 60% had findings that were considered typical of FDRA by strict diagnostic criteria.[4] All of the patients with typical findings and 46% of the patients with atypical presentations had GAA expansions in the *FDRA* gene, which encodes the protein frataxin; such findings were consistent with the diagnosis of FDRA. To accommodate the molecular diagnostic criteria, the phenotypic spectrum of FDRA was broadened to include older age at onset and preservation of deep tendon reflexes.

A rough correlation exists between the number of trinucleotide repeats and the severity and age at onset of disease in all of these disorders; however, the positive predictive value of the number of repeats for these findings is less than 100%. This value is not relevant when the test that measures repeat number is used for the diagnosis of symptomatic persons, but it becomes relevant when the test is used for predictive testing in asymptomatic at-risk relatives and for recurrence-risk counseling of the offspring of an affected person. Recurrence-risk counseling for trinucleotide repeat disorders depends not only on the usual mendelian genetics but also on the empirical risk of further gene expansion (i.e., the increase in length of the trinucleotide repeat) during meiosis. For unknown reasons, expansion can be influ-

Table 3 Use of DNA-Based Testing for Certain Inherited Disorders

Disease Name	*Medical Paradigm*		*Counseling-Only Paradigm*		
	Diagnosis	*Predictive*	*Predictive*	*Carrier Detection*	*Prenatal Diagnosis*
Huntington disease	X		X		X*
Friedreich ataxia	X		X	X	X
Factor V Leiden	X	X			X*
Duchenne muscular dystrophy Becker muscular dystrophy	X		X	X	X
Familial adenomatous polyposis	X†	X			X
Retinoblastoma		X			X
Breast cancer		X	X		X*
Cystic fibrosis		X		X	X

*It is not common for parents to request prenatal testing for conditions that typically manifest in adulthood. Differences in perspective may exist among medical professionals and within families regarding the use of prenatal testing when the testing is being considered for the purpose of pregnancy termination or for early diagnosis. Most centers would consider decisions about prenatal testing to be the choice of the parents; however, careful discussion of these issues is appropriate.

†Molecular testing may be necessary to establish the diagnosis in individuals with the mild variant form of familial adenomatous polyposis.

Table 4 Trinucleotide Repeat Diseases[1,2]

Disease Name	*Type of Repeat*	*Mode of Inheritance*	*Normal Allele Size**	*Intermediate Allele Size*†*	*Abnormal Allele Size*‡*	*Stabilization by Repeat Interruption*	*Parental Effect*	*Somatic Mosaicism*
X-linked spinal and bulbar muscular atrophy	CAG	XLR	≤ 34		≥ 38	No	pat > mat	No
Huntington disease	CAG	AD	9–26	27–35	36–121	No	pat > mat	Yes
Dentatorubral-pallidoluysian atrophy	CAG	AD	≤ 35		49–93	No	pat > mat	Yes
Spinocerebellar ataxia type 1 (SCA1)	CAG	AD	6–44		39–91	Yes (CAT)	pat > mat	Yes
SCA2	CAG	AD	≤ 30		33–500 (or more)	Yes (CAA)	pat = mat	ND
SCA3	CAG	AD	< 47	48–51	53–86	No	pat = mat	Yes
SCA6	CAG	AD	< 18	19	20–23	No	?	ND
SCA7	CAG	AD	4–35		36–450 (or more)		pat > mat	
SCA8	CTG	AD	15–50		80–800 (or more)			
SCA12	GAG	AD	7–31		55–78			
Fragile X syndrome locus A	CGG	XLD	5–44	59–200	> 200	Yes (AGG)	mat > pat	Yes
Fragile X syndrome locus E	GCC	XLD	6–25		> 200			
Myotonic dystrophy	CTG	AD	5–35		> 50		mat > pat	
Friedreich ataxia	GAA	AR	≤ 33	34–65	66–1,700			Occasional
Oculopharyngeal muscular dystrophy	GCG	AD	6		8–13		No	No
Oculopharyngeal muscular dystrophy	GCG	AR	6		7		No	No

*Allele size is described by the number of CAG trinucleotide repeats.
†A person with an allele in this range of CAG repeats is not at risk of developing symptoms of the disorder but may be at risk of having a child with an allele in the abnormal range.
‡Alleles of abnormal size fall into two categories: (1) reduced-penetrance alleles comprising 36–40 CAG repeats and (2) full-penetrance alleles, comprising 41 CAG repeats or more. Reduced-penetrance alleles may or may not lead to symptoms of HD within normal life expectancy.[20] Full-penetrance alleles are associated with the development of HD.
AD—autosomal dominant AR—autosomal recessive mat—maternal ND—no data pat—paternal XLR—X-linked recessive

enced by the sex of the transmitting parent; for example, further expansion is probable when a mother transmits the expanded allele in fragile X syndrome and myotonic dystrophy type 1 and when the father transmits the abnormal allele in Huntington disease, Kennedy disease, dentatorubral-pallidoluysian atrophy (DRPLA), or spinocerebellar ataxia types 1 and 3.[1] The risk of further expansion may depend on the total length of the trinucleotide repeat region and the presence of different stabilizing sequences within or adjacent to the gene.

High Sensitivity and Low Positive Predictive Value

Factor V Leiden targeted mutation analysis is the most commonly ordered genetic test. The specific mutation in coagulation factor V, named factor V Leiden, is a glutamine substituted for an arginine at codon 506.[5] By definition, the test that detects this mutation is 100% sensitive. Factor V Leiden causes resistance to activated protein C, a natural anticoagulant that allows extravascular blood to clot while maintaining intravascular fluidity.[5] Epidemiologic data support a predisposition to primary and recurrent venous thrombosis in factor V Leiden heterozygotes. It is estimated that 5% of whites are heterozygous for factor V Leiden and that approximately 20% of all persons with venous thromboembolism are heterozygous for factor V Leiden.[6] Heterozygotes with the mutation have a 2.4-fold greater risk of recurrent thromboembolism than patients without the mutation. Factor V Leiden is present in more than half of families with a so-called thrombophilic tendency; that is, several family members have deep vein thromboses that often are multiple, are of early onset, or occur in the absence of clear risk factors. It is presumed that in these families, other risk factors—some genetic and some environmental—are present.

Factors that increase the risk for venous thrombosis include the presence of two factor V Leiden alleles; the presence of other inherited and acquired thrombophilic disorders, including protein C deficiency, protein S deficiency, antithrombin deficiency, the prothrombin gene mutation, and hyperhomocysteinemia[7]; age; surgery; the use of oral contraceptives[8]; hormone replacement therapy[9]; and pregnancy.[10]

Increased Sensitivity by Use of Tests in Combination

Molecular genetic testing of leukocyte DNA in the diagnosis of Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) is an example of diagnostic molecular genetic testing in which sensitivity has increased because of new test methodologies. Allelic heterogeneity reduces the ability of one test method to detect all disease-causing mutations. Allelic heterogeneity (sometimes called mutational heterogeneity) refers to the situation in which more than one disease-causing mutation (allele) at one locus causes a given phenotype. About 65% of males with DMD have a deletion in the *DMD* gene, which encodes the protein dystrophin. Deletions are detected by a combination of multiplex polymerase chain reaction, Southern blotting, and fluorescence in situ hybridization testing. In DMD, deletions result in a frameshift and an absence of production of the protein dystrophin. The detection of an out-of-frame mutation is sufficient to establish the diagnosis of DMD. When no

deletion is detected, mutation scanning and sequence analysis can be used to detect an additional 30% of DMD-causing mutations.[11,12] Thus, muscle biopsy using immunohistochemical and immunoelectron microscopic techniques to visualize the protein dystrophin in the subsarcolemmal area, which was the diagnostic test previously used when no deletion was detected, is now necessary to establish the diagnosis of DMD in a small percentage of patients. In muscle-biopsy testing, the absence of dystrophin is diagnostic of DMD.

BMD is allelic to DMD; that is, the two disorders are caused by different mutations at the same gene (*DMD*). BMD is rarer than DMD, and BMD has a later onset and a milder course than DMD. Skeletal muscle function and lifespan are better for persons with BMD than for those with DMD; however, life-threatening dilated cardiomyopathy is common in patients with BMD. Deletions of the *DMD* gene occur in about 85% of males with BMD, but these deletions are in-frame mutations that lead to the production of a truncated dystrophin protein. Sequence analysis and, if necessary, muscle biopsy can be diagnostic in patients suspected of having BMD who have no discernible *DMD* deletion. In such cases, dystrophin is detectable in the subsarcolemma but is reduced in quantity.

The use of *DMD* molecular genetic testing exemplifies the positive predictive value of molecular genetic testing through genotype-phenotype correlation of frameshift (DMD phenotype) and in-frame (BMD phenotype) mutations.

The cost of molecular genetic testing on leukocytes varies by method; however, in general, genetic testing is less expensive than an open muscle biopsy, which requires a surgeon, an anesthesiologist, a pathologist, and related staff. Furthermore, use of molecular genetic testing to clarify the genetic status of female relatives of patients with DMD and BMD necessitates that the specific disease-causing mutation be identified through use of molecular genetic testing in at least one affected male relative.

PREDICTIVE TESTING IN ASYMPTOMATIC AT-RISK PERSONS

Predictive testing refers to testing of at-risk asymptomatic relatives for a disease-causing gene alteration known to be present in the family to clarify the relatives' genetic status. Predictive testing is considered presymptomatic when it is certain that all persons who have the altered gene will become symptomatic; it is considered predispositional when penetrance of the gene is reduced (i.e., fewer than 100% of persons with the altered gene will be affected).

For predictive testing to be useful, specificity and positive predictive value must be high. Cost-effectiveness of predictive testing is realized by reducing morbidity and mortality in patients at high risk through (1) early detection and treatment of those with a disease-causing mutation and (2) removal of persons who do not have the mutation from screening protocols, which can be expensive and invasive.[13] The disorders in this category are primarily autosomal dominant cancers [*see Table 5*].[14] The distinctions between presymptomatic testing and predispositional testing may not be meaningful when the test is for a mutation associated with a high risk of a disorder for which the general population is at relatively low risk (e.g., hereditary nonpolyposis colon cancer [HNPCC]). Molecular genetic testing may not be required to establish the diagnosis in the proband when the disorder is diagnosed by clinical findings. However, testing of an affected family member is required to identify the family-specific mutation for the purpose of testing asymptomatic at-risk relatives. The interest of at-risk relatives in pursuing testing for certain disorders is largely unknown. Hadley and colleagues determined that 51% of first-degree relatives at risk for HNPCC chose to undergo predispositional testing after education and genetic counseling to clarify the risk to their children. The potential effect on health insurance was the single most common reason to decline testing.[15]

Table 5 Autosomal Dominant Cancer Syndromes for Which Molecular Genetic Testing Is Available

Von Hippel–Lindau disease	Multiple endocrine neoplasia type 2
Familial adenomatosis polyposis	Multiple endocrine neoplasia type 1
Hereditary nonpolyposis colorectal cancer	Breast cancer
Retinoblastoma	Melanoma

Presymptomatic Testing

FAP is an autosomal dominant disorder in which penetrance of the disease-causing gene mutations is 100%. Persons with an *APC* gene mutation develop adenomas in the colorectum starting at around 16 years of age; in these individuals, the number of adenomas increases to hundreds or thousands, and colorectal cancer develops at a mean age of 39 years. The mean age at death is 42 years in those who go untreated. Early diagnosis via presymptomatic testing reduces morbidity and increases life expectancy through improved surveillance and timely prophylactic colectomy.[16] Testing of the *APC* gene has been shown to be cost-effective when used to identify individuals with the disease-causing *APC* mutation among at-risk relatives of persons with FAP.[17] For years, the mainstay of FAP testing was protein truncation testing (PTT). However, the mutation-detection rate with PTT was only about 80%; the introduction of other test methods—namely, gene sequencing—has increased the mutation-detection rate to 98%.

Predispositional Testing

Retinoblastoma is an example of a disorder in which penetrance of disease-causing gene mutations is less than 100%. It is caused by mutations in the *RB1* gene and can be inherited in an autosomal dominant manner. On average, penetrance of *RB1* gene mutations is 90% (i.e., 90% of persons with a germline disease-causing gene mutation will develop retinoblastoma). Cost-effectiveness and improved outcome through the use of predispositional gene testing have been demonstrated.[18]

Improved outcome is defined as the preservation of vision in at-risk persons through early detection and treatment of ocular tumors, as well as the reduction of morbidity through early detection of nonocular secondary tumors. Early detection of retinoblastoma, while the tumor is small, allows less aggressive treatments that ablate tumors but preserve vision. Before the availability of molecular genetic testing, recurrence-risk counseling for the parents of a child with retinoblastoma or for an adult with retinoblastoma was empirical; counseling was offered on the basis of a positive or negative family history and the presence of a single tumor or multiple tumors. The surveillance protocol is required whether a child has a 6% risk of retinoblastoma (parent or sibling with unilateral, sporadic retinoblastoma), a 40% risk (parent with bilateral retinoblastoma), or a 90% risk (person known to have a germline *RB1* mutation).

Sequencing of the *RB1* gene detects mutations in over 80% of patients with bilateral or hereditary retinoblastoma.[19] Although it is both labor intensive and expensive, gene sequencing is required to establish the molecular diagnosis in a proband. Because of extensive allelic heterogeneity, gene sequencing is the gold standard for detection of *RB1* gene mutations. Thus, an adult proband who has had retinoblastoma can undergo *RB1* sequence analysis in hopes of identifying the disease-causing mutation so that molecular genetic testing can be used in the management of his or her at-risk offspring. When a germline *RB1* mutation is identified in the proband, the offspring can be tested prenatally or at birth to determine the genetic status and whether there is a need for frequent ophthalmologic examinations. When no mutation is identified in the adult, the risk of recurrence is determined empirically, and all offspring must be evaluated regularly by an ophthalmologist. Cost-effectiveness results from not subjecting at-risk children to unnecessary and expensive screening protocols after they test negative for an *RB1* germline mutation known to be in their family.[13]

Counseling-Only Paradigm of Genetic Testing

In the counseling-only paradigm of genetic testing, genetic tests provide persons with information pertaining to disease risk for the purpose of personal decision making, which may include reproductive planning. Issues of test sensitivity, specificity, positive predictive value, and recurrence risk are as relevant in the genetic-counseling paradigm as they are in the medical model, but cost-effectiveness cannot be assessed when testing is used only for personal decision making.

PREDICTIVE TESTING

Predictive testing used for presymptomatic and predispositional diagnosis of persons at risk for disorders for which no medical interventions exist falls into the genetic-counseling paradigm.

Presymptomatic Diagnosis

Huntington disease is an example of a disorder for which no medical intervention exists. Huntington disease is caused by a CAG trinucleotide repeat expansion in the *HD* gene. When the CAG expansion is greater than 41 repeats, the penetrance is 100%—that is, all persons with an allele that size will eventually develop Huntington disease. Clarification of genetic status in persons at risk for Huntington disease allows those who have inherited the altered gene and those who have not inherited the altered gene to make informed personal and social decisions. Such decisions may include matters of lifestyle, employment, personal finance, and family planning.

Offering genetic testing to persons at risk for an untreatable, debilitating, fatal disorder requires careful forethought. The molecular diagnosis must always be confirmed in a symptomatic relative before testing can be offered to asymptomatic family members who are at risk. Because no medical intervention can be offered, anticipating and addressing the patient's psychoemotional needs are paramount. In a position paper, the National Society of Genetic Counselors emphasized that pretest education and genetic counseling are necessary and that posttest follow-up care must be in place at the time of genetic testing.[20] Informing the patient of normal results requires as much preparation and counseling as the relaying of abnormal results. The pretest counseling with the patient must address the positive predictive value of the test, particularly as relating to age at onset and severity of the disease. Greater repeat length is usually associated with earlier onset and more severe disease, but repeat length is not always a predictor of disease onset or severity. Furthermore, patients with an intermediate (or moderately abnormal) number of trinucleotide repeats (i.e., 36 to 40 CAG repeats) may have an indeterminate genetic risk.[21] Trinucleotide repeat sizes in this range are considered to have reduced penetrance because they can cause disease symptoms but do not always do so within a normal life expectancy.[2] Thus, the patient who is prepared to hear a negative or a positive result may be in the same uncertain position after testing as before.

Confidentiality of test results and possible discrimination in employment and health insurance coverage[22] may be issues for an asymptomatic person who has undergone genetic testing. Predictive genetic testing of asymptomatic at-risk individuals younger than 18 years is strongly discouraged in the genetic-counseling paradigm because of concerns that children will be inappropriately labeled at a time when they cannot be expected to use this information for personal planning or reproductive decision making.[23] Diagnostic testing of symptomatic at-risk children is always appropriate.

Predispositional Testing

Predispositional testing for a disorder may not be appropriate when the disorder is highly prevalent in the general population and when the efficacy of measures to reduce risk in persons with disease-predisposing mutations is unknown. Predispositional testing for breast cancer through molecular genetic testing of the genes *BRCA1* and *BRCA2* can be considered in this category, because the efficacy of measures to reduce cancer risk for individuals with *BRCA1* or *BRCA2* cancer-predisposing mutations is unknown.[24] Furthermore, the high prevalence of breast cancer in the general population means that the rigorous screening for early breast cancer identification recommended for all women cannot be relaxed even when an at-risk woman does not have the *BRCA1* or *BRCA2* cancer-predisposing mutation identified in a relative.

The dilemma posed by the indeterminate role of *BRCA1* and *BRCA2* molecular genetic testing in reducing morbidity from breast cancer may turn out to be a recurring issue in genetic testing for common diseases. Breast cancer, like such other common disorders as coronary artery disease, diabetes mellitus, and Alzheimer disease, is regarded as a complex disorder. Complex disorders have multiple etiologies, including heritable single genes, multiple genes with an additive effect that interact with often undefined environmental influences, and acquired environmental or genetic changes. With regard to the overall incidence and morbidity of common diseases, the contribution of single heritable genes is relatively small. For example, breast cancer affects one in nine women, yet only 5% to 10% of cases of breast cancer are attributed to mutations in single genes, including *BRCA1* and *BRCA2*. For a woman whose relatives have a known *BRCA1* mutation but who has tested negative for the mutation known to be in the family, the chance of breast cancer developing is still one in nine. She therefore has the same need for close surveillance as women in the general population. Furthermore, detection of a *BRCA1* or *BRCA2* mutation may not alter the surveillance protocol for breast cancer that is recommended for all women, but it may increase the utilization of mammography, breast self-examination, and oophorectomy.[25] The options for breast cancer prevention (e.g., bilateral mastectomy), however, might be considered in a different light for

women with a *BRCA1* or *BRCA2* mutation. Such a prevention strategy should be undertaken with caution because the positive predictive value of a *BRCA1* or *BRCA2* mutation for the development of breast cancer may not be fully understood and may be biased upward as a result of higher risks and different disease spectra in the high-risk families studied initially.

Serious issues surround testing for *BRCA1* and *BRCA2* mutations, including appropriate pretest counseling for at-risk women[26]; appropriate interpretation of positive and negative test results [*see Table 1*]; the high probability of missense mutations, which are considered indeterminate test results; recommendations for surveillance; and consideration of prophylactic mastectomy [*see 12:VII Breast Cancer*].

CARRIER TESTING

Carrier testing is used primarily to identify carriers of autosomal recessive gene mutations and X-linked gene mutations. There are no health-related issues for carriers of an autosomal recessive gene mutation, because all are expected to be asymptomatic. Health-related issues are a concern for a subset of female carriers of an X-linked gene mutation.

Autosomal Recessive Disorder

In cases of autosomal recessive disorders, testing may be used for diagnosis in a symptomatic person, for evaluation of at-risk asymptomatic persons, and for detection of carriers and affected fetuses. Although the gene causing an autosomal recessive disorder may be well characterized, allelic heterogeneity may reduce the sensitivity of molecular genetic testing below levels acceptable for diagnostic use, because it may not be technically feasible to identify all possible disease-causing mutations. In other cases, carrier detection and prenatal testing may be possible only through molecular genetic testing, which may provide information for reproductive decision making that would not otherwise be accessible [*see Table 6*].

Cystic fibrosis (CF) is an example of such an autosomal recessive disorder. Although discovery of disease-causing mutations in the *CFTR* gene has led to new tests for CF and redefinition of the disease spectrum, the traditional diagnostic criteria for classic CF are still valid.[27] The diagnosis of CF is established when the amount of sweat chloride is greater than 60 mEq/L in the presence of one or more characteristic clinical findings (e.g., typical gastrointestinal or sinopulmonary disease or obstructive azoospermia) or when the family history is positive for the disease. In questionable cases, *CFTR* molecular genetic testing can be helpful in establishing the diagnosis; however, genotyping alone rarely establishes the diagnosis of CF. Some individuals may have classic CF without a detectable *CFTR* disease-causing mutation because of allelic heterogeneity. Allelic heterogeneity in CF is extensive, with over 1,000 known disease-causing mutations. The American College of Medical Genetics (ACMG) has recommended a panel of 25 mutations for routine testing in clinical laboratories.[28] With the use of this panel, mutation detection rates vary by ethnicity; in white Europeans, 2% of patients with CF have no detectable abnormal alleles and 26% have only one detectable abnormal allele.

When two disease-causing alleles are identified in the proband, both parents can be tested to determine which parent carries which allele. Then, relatives of the mother can be tested for the presence of her disease-causing allele, and relatives of the father can be tested for the presence of his disease-causing allele. Any relative who is found to be a carrier of a disease-causing allele has the option of having his or her spouse tested with the clinically available panel of 25 common disease-causing alleles. Couples in which both partners are carriers of disease-causing alleles have a 25% chance of having a child who inherits two *CFTR* disease-causing mutations; however, the clinical manifestations and severity of the disease cannot necessarily be predicted by the specific mutations present. When the spouse has no identifiable disease-causing mutations, carrier risk can be calculated using Bayesian analysis.[29]

Table 6 Autosomal Recessive Disorders for Which Genetic Molecular Testing Permits Carrier Detection*

Cystic fibrosis	Congenital disorders of glycosylation
Phenylketonuria	β-Thalassemia
21-Hydroxylase deficiency	Canavan disease
Tay-Sachs disease	α-Thalassemia

* Partial list of disorders.

The use of DNA-based testing is sensitive in high-risk families. Its use in preconceptual counseling for carrier detection is more complicated because of its low sensitivity; however, such testing was endorsed by the American College of Obstetrics and Gynecology (ACOG), the ACMG, and the National Human Genome Research Institute in 2001.

X-Linked Disorders

In X-linked disorders, molecular genetic testing may be used to diagnose symptomatic males and the occasional symptomatic female and to detect carrier females and affected male fetuses. As in autosomal recessive disorders, molecular genetic testing is often the only option for carrier detection and prenatal testing. Certain factors make the testing for carriers of X-linked disorders more complicated; these include the high frequency of new gene mutations in males who are the only affected family member, as well as the possibility of germline mosaicism in the mother of a male who is the only affected family member. New gene mutations are borne by only a single egg or a single sperm. Germline mosaicism is the presence in some germline cells (eggs or sperm) of a mutation that is not found in other germline cells or somatic cells. Germline mosaicism for an X-linked disorder is surmised to be present in the mother of two or more affected males when there is no evidence of their disease-causing mutation present in her leukocytes.

DMD is an example of an X-linked disorder in which these issues must be considered in genetic counseling. Carrier detection in DMD can be problematic because a significant number of cases of DMD in males are simplex cases (i.e., single occurrences in a family). The following three equally probable possibilities exist for males with DMD who have a negative family history:

1. The affected boy has a new (de novo) gene mutation. In this case, his mother does not carry a disease-causing allele, and her female relatives are not at risk to be carriers of the altered allele.
2. The mother carries a de novo mutation, which places her daughters but not her sisters at risk for being carriers of the altered allele.
3. The maternal grandmother carries a de novo mutation, which places all her daughters at risk for being carriers of the altered allele.

Thus, in families in which DMD occurs in one male only, recurrence-risk counseling depends on establishing which, if any, of the women are carriers of a disease-causing mutation. The following testing and recurrence-risk counseling paradigm is used:

- DNA testing is performed on the male proband with DMD to identify the causative *DMD* gene mutation. When a *DMD* gene mutation is identified, a blood sample from the proband's mother is tested for the same mutation.
- If she has the same mutation as her son, she is counseled regarding the 50% risk of other sons being affected and the 50% risk of daughters being carriers; it is appropriate to test the proband's maternal grandmother for the same disease-causing mutation.
- If the proband's mother tests negative for his mutation, two possibilities exist: the son has a new gene mutation or the mother has germline mosaicism, which occurs in about 20% of women in this situation.
- If the son has a new gene mutation, the mother is not at increased risk for having other affected sons, and other women in the family are not at increased risk for being carriers.
- If the mother has germline mosaicism, she is at risk for having carrier daughters and additional affected sons. Her sisters, however, are not at increased risk for being carriers.

PRENATAL TESTING

Prenatal testing is used to evaluate a fetus at high risk for a genetic disorder on the basis of family history or to evaluate a fetus at no known increased risk but who is suspected of having a genetic disorder because of suggestive findings during the pregnancy.

Positive Family History

Testing of fetuses using molecular genetic testing can be offered to couples at risk for having a child with an autosomal dominant, autosomal recessive, or X-linked disorder for which the specific gene mutation (or mutations) has been identified in the family. Genetic counseling must be offered to provide the family an opportunity to review their reproductive options. Molecular genetic testing can be performed on tissues obtained by chorionic villus sampling at 9 to 11 weeks' gestation or from amniocentesis at 16 to 18 weeks' gestation to provide timely information should pregnancy termination be considered.

Findings Suggestive of a Genetic Disorder

Prenatal molecular genetic testing can be a part of the diagnostic evaluation of a fetus not known to be at increased risk for a genetic disorder that is being evaluated further because of abnormalities detected during routine monitoring of the pregnancy. When such findings are detected early in the pregnancy, DNA-based diagnosis may be undertaken if pregnancy termination is being considered. When findings are not apparent until the third trimester, diagnosis may be initiated for the purpose of perinatal management. For example, ultrasound findings of intestinal obstruction with hyperechoic meconium would warrant *CFTR* molecular genetic testing because of the association of *CFTR* with cystic fibrosis. Such testing can be performed on DNA extracted from amniocytes obtained from amniocentesis after 16 weeks' gestation, when timing is not an issue, or from white cells obtained by percutaneous umbilical blood sampling (PUBS) when results are needed urgently.

Genetic Consultation

Genetic consultation is as essential to the care of the patient with a genetic disorder as the testing itself; it is required for persons considering either the medical paradigm or the counseling-only paradigm of genetic testing. A positive genetic test result always raises the consideration of referral for genetic counseling.[30-33] Genetic counseling is the process of helping patients understand the nature and cause of the inherited disorder, of outlining the advantages and disadvantages of genetic testing to allow them to make informed medical and personal decisions, and of offering necessary psychosocial support and referral.[30,31,34] Genetic evaluation is the process of information gathering regarding a patient or family with a known or suspected genetic disorder. Genetic evaluation and genetic counseling are integral to genetic testing.

Genetic evaluation involves the gathering of information before a clinic visit and during the initial portion of the visit, which usually lasts 1 hour. The following information is obtained from the patient or family: the reason for referral; a family history, including the history of first- and second-degree relatives of the consultand; additional directed family history based on the known or suspected diagnosis and information provided by the patient or other family members; medical records of affected relatives; prenatal and perinatal history; past medical history; and information on growth, development, education, and employment. In addition, family functioning is assessed, potential ethical issues are identified, and a physical examination is performed on the patient and other family members as needed.

Once the gathering of information is complete, genetic counseling is provided. Discussion with the patient or family includes a summary of information obtained; the possible diagnosis and the degree of certainty of that diagnosis, determined on the basis of available information; recommended tests and evaluations necessary to establish the diagnosis or for management of the patient; the sensitivity and positive predictive value of such tests; the natural history of the disorder, including prognosis; inheritance pattern, including penetrance and variable expressivity (i.e., the variation in the type and severity of a genetic disorder between affected individuals, even within the same family); and recurrence risk for affected persons and for at-risk persons, including reproductive options and options for prenatal diagnosis. Medical management and referrals to appropriate medical specialists are discussed. Psychosocial issues discussed include anticipatory guidance of the patient and family, the availability of community support services, and the availability of regional or national disease-specific or umbrella organizations, many of which can be identified through the Genetic Alliance (www.geneticalliance.org). Genetic-counseling issues for the extended family are addressed. Geographically dispersed family members are referred to local genetic services that can be identified through the GeneTests Clinic Directory, which is available on the GeneTests Web site; the Clinic Directory can be searched by location within the United States, disease specialty, and services offered. Clinic visits and genetic-counseling sessions are documented with detailed summaries suitable for distribution to the family and health care providers. Summary letters are often sent to the family. Short-term follow-up is planned for conveying outstanding test results or other information; long-term follow-up at 2- to 5-year intervals is planned for routine management and updating of genetic-counseling issues.

Difficulties Encountered in DNA-Based Testing for Inherited Disorders

LACK OF AWARENESS OF TEST AVAILABILITY

For many inherited disorders, molecular genetic testing is not available, because the causative gene (or genes) is not known. In other instances, the gene is known but test sensitivity is less than that of clinical evaluation, and testing is done only in a research context. For rarer disorders, the gene may be known and clinical testing may be theoretically possible, but clinical laboratories do not offer the test because the cost of low-volume, highly complex testing is prohibitive. For other inherited diseases, the causative gene or genes are known, research testing is currently available, and clinical testing is expected to be available in the near future.[35,36] The rapid transition of testing from research laboratory to clinical practice makes it difficult even for those who are familiar with genetic testing to keep abreast of new developments. For those not familiar with genetic testing and its applications to patient care, the task is even more daunting.[37]

GeneTests is a genetic-testing information resource, funded by the National Institutes of Health (NIH) and maintained at the University of Washington in Seattle, that is designed to facilitate awareness of test availability and use.[38] The Web site (http://www.genetests.org) includes a laboratory directory that serves to help health care providers identify clinical and research laboratories offering testing of heritable disorders. As of July 2005, the Laboratory Directory contained listings of about 1,100 diseases for which clinical (~ 800) and research-only (~ 300) testing was available from approximately 575 laboratories. Clinical laboratories are defined as those that examine human specimens and report results for the purpose of diagnosis, prevention, or treatment in the care of individual patients; such laboratories must be licensed according to the Clinical Laboratory Improvement Amendments (http:\\www.cms.hhs.gov/clia). The Laboratory Directory can be searched by disease name, gene symbol, protein name, clinical features, the laboratory director's name, and the laboratory's geographic location.

COMPLEXITY OF TESTING METHODOLOGIES AND INTERPRETATION OF TEST RESULTS

Physicians may not be familiar with the use and limitations of molecular genetic tests in patient care. For example, Giardiello and colleagues determined that almost 20% of clinicians ordering *APC* molecular genetic testing for FAP used the wrong testing strategy.[31] These investigators also determined that 34% of clinicians ordering *APC* molecular genetic testing were unable to identify and interpret false negative results. Several genetic concepts that are intrinsic to the correct use of testing may be confusing. These concepts include (1) locus heterogeneity, in which the identical phenotype can be caused by a single mutation in one of two or more genes (e.g., mutation in either the *TSC1* or *TSC2* gene can cause tuberous sclerosis complex), which means that negative testing of the gene at only one locus does not rule out the disease; (2) allelic heterogeneity, in which multiple disease-causing mutations at a locus reduce the sensitivity of molecular testing below an acceptable level for use in diagnosis but not for recurrence-risk counseling; and (3) redefinition of phenotypes on the basis of molecular genetic findings (e.g., trinucleotide repeat diseases, CF, and the dystrophinopathies, including DMD, BMD, and X-linked dilated cardiomyopathy).

In a survey of genetic counselors in the United States, McGovern and colleagues determined that genetic counselors contacted laboratory-testing personnel for 58% of tests ordered regarding details of test ordering or interpretation of test results.[39] Only 72% felt that laboratory reports contain enough information to explain results to patients. There are concerns that nongeneticist health care providers face even greater challenges in using genetic-testing laboratories.

The *GeneReviews* portion of the GeneTests Web site contains current information on the use of genetic testing in diagnosis, management, and genetic counseling for specific inherited disorders. Entries on over 300 diseases (as of July 2005) provide expert-authored, peer-reviewed information for health care professionals. A context-sensitive illustrated glossary familiarizes nongeneticists with genetic counseling and genetic testing terms.

CONFUSION BETWEEN TESTING PARADIGMS

The intertwining of genetic testing for medical management and testing for personal decision making may lead to confusion about medical necessity, privacy, and discrimination. Just as the application of testing to patient care is context specific, consideration of these social issues is also context specific.

UNDERUTILIZATION OF GENETIC SERVICES

Giardiello and colleagues determined that only 18% of patients undergoing predictive testing for FAP, an autosomal dominant disorder that has 100% penetrance and that is associated with a 100% risk of cancer by 40 years of age, received genetic counseling.[31] A survey of 600 primary care physicians in Oregon revealed that 20% of internists did not know of any genetic services available to them for consultation.[40] Furthermore, most felt they did not need to refer patients for genetic consultation, preferring to offer risk-assessment and recurrence-risk counseling themselves, even though they were unfamiliar with the specific disorders and genetic-counseling concepts.[40] The need for primary care physicians to understand and use genetic services has been emphasized.[33,37,41,42] Possible explanations of underutilization of genetic services are the so-called therapeutic gap between diagnosis and prediction of diseases and the ability to treat or prevent them[43]; real or perceived restrictive reimbursement policies of health care payers; and concern about ethical and social issues that would seem to create "genetic exceptionalism" (the justification, provided by genetic testing, for special consideration regarding issues of informed consent and privacy),[44] which would move genetic testing beyond the purview of traditional medical care.[45] The change in the use of some genetic tests from a diagnostic role to a screening role (e.g., testing for factor V Leiden) may shift the emphasis of testing away from the evaluation of at-risk family members and genetic counseling to a broader role related to population-based health care.

The author receives contract support from the National Institutes of Health.

References

1. La Spada AR: Trinucleotide repeat instability: genetic features and molecular mechanisms. Brain Pathol 7:943, 1997

2. Potter NT, Nance MA: Genetic testing for ataxia in North America. Mol Diagn 5:91, 2000

3. Nance MA: Clinical aspects of CAG repeat diseases. Brain Pathol 7:881, 1997

4. Dürr A, Cossee M, Agid Y, et al: Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med 335:1169, 1996

5. Ridker PM, Hennekens CH, Lindpaintner K, et al: Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. N Engl J Med 332:912, 1995

6. Simioni P, Prandoni P, Lensing AW, et al: Risk for subsequent venous thromboembolic complications in carriers of the prothrombin or the factor V gene mutation with a first episode of deep-vein thrombosis. Blood 96:3329, 2000

7. Ridker PM, Hennekens CH, Selhub J, et al: Interrelation of hyperhomocyst(e)inemia, factor V Leiden, and risk of future venous thromboembolism. Circulation 95:1777, 1997

8. Legnani C, Palareti G, Guazzaloca G, et al: Venous thromboembolism in young women: role of thrombophilic mutations and oral contraceptive use. Eur Heart J 23:984, 2002

9. Grady D, Wenger NK, Herrington D, et al: Postmenopausal hormone therapy increases risk for venous thromboembolic disease. The Heart and Estrogen/progestin Replacement Study. Ann Intern Med 132:689, 2000

10. Ridker PM, Glynn RJ, Miletich JP, et al: Age-specific incidence rates of venous thromboembolism among heterozygous carriers of factor V Leiden mutation. Ann Intern Med 126:528, 1997

11. Mendell JR, Buzin CH, Feng J, et al: Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology 57:645, 2001

12. Dent KM, Dunn DM, von Niederhausern AC, et al: Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort. Am J Med Genet A 134:295, 2005

13. Noorani HZ, Khan HN, Gallie BL, et al: Cost comparison of molecular versus conventional screening of relatives at risk for retinoblastoma. Am J Hum Genet 59:301, 1996

14. Offit K: Clinical Cancer Genetics: Risk Counseling and Management. Wiley-Liss, New York, 1998

15. Hadley DW, Jenkins J, Dimond E, et al: Genetic counseling and testing in families with hereditary nonpolyposis colorectal cancer. Arch Intern Med 163:573, 2003

16. Winawer S, Fletcher R, Rex D, et al: Colorectal cancer screening and surveillance: clinical guidelines and rationale: update based on new evidence. Gastroenterology 124:544, 2003

17. Cromwell DM, Moore RD, Brensinger JD, et al: Cost analysis of alternative approaches to colorectal screening in familial adenomatous polyposis. Gastroenterology 114:893, 1998

18. Gallie BL: Predictive testing for retinoblastoma comes of age. Am J Hum Genet 61:279, 1997

19. Lohmann DR: RB1 gene mutations in retinoblastoma. Hum Mutat 14:283, 1999

20. McKinnon WC, Baty BJ, Bennett RL, et al: Predisposition genetic testing for late-onset disorders in adults: a position paper of the National Society of Genetic Counselors. JAMA 178:1217, 1997

21. Langbehn DR, Brinkman RR, Falush D, et al: A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. Clin Genet 65:267, 2004

22. Pokorski RJ: Insurance underwriting in the genetic era. Am J Hum Genet 60:205, 1997

23. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. ASHG/ACMG Report. Am J Hum Genet 57:1233, 1995

24. Burke W, Daly M, Garber J, et al: Recommendations for follow-up care of individuals with an inherited predisposition to cancer: II. BRCA1 and BRCA2. JAMA 277:997, 1997

25. Botkin JR, Smith KR, Croyle RT, et al: Genetic testing for a BRCA1 mutation: prophylactic surgery and screening behavior in women 2 years post testing. Am J Med Genet 118A:201, 2003

26. Burke W, Culver JO, Bowen D, et al: Genetic counseling for women with an intermediate family history of breast cancer. Am J Med Genet 90:361, 2000

27. Stern RC: The diagnosis of cystic fibrosis. N Engl J Med 336:487, 1997

28. Richards CS, Bradley LA, Amos J, et al: Standards and guidelines for CFTR mutation testing. Genet Med 4:379, 2002

29. Curnow RN: Carrier risk calculations for recessive diseases when not all the mutant alleles are detectable. Am J Med Genet 52:108, 1994

30. Assessing Genetic Risks: Implications for Health and Social Policy. Andrews LB, Fullarton JE, Holtzman NA, et al, Eds. National Academy Press, Washington, DC, 1994

31. Giardiello FM, Brensinger JD, Petersen GM, et al: The use and interpretation of commercial *APC* gene testing for familial adenomatous polyposis. N Engl J Med 336:823, 1997

32. Marymee K, Dolan CR, Pagon RA, et al: Development of critical elements of genetic evaluation and genetic counseling for genetics professionals and perinatologists in Washington State. J Genet Couns 7:133, 1998

33. Ensenauer RE, Michels VV, Reinke SS: Genetic testing: practical, ethical, and counseling considerations. Mayo Clin Proc 80:63, 2005

34. Schneider KA: Counseling about Cancer: Strategies for Genetic Counselors, 2nd ed. John Wiley & Sons, Boston, 2001

35. Maron BJ: Hypertrophic cardiomyopathy. Lancet 350:127, 1997

36. Yu B, French JA, Jeremy RW, et al: Counseling issues in familial hypertrophic cardiomyopathy. J Med Genet 35:183, 1998

37. Cotton P: Prognosis, diagnosis, or who knows? Time to learn what gene tests mean. JAMA 273:93, 1995

38. Burke W: Genetic testing. N Engl J Med 347:1867, 2002

39. McGovern MM, Benach M, Zinberg R: Interaction of genetic counselors with molecular genetic testing laboratories: implications for non-geneticist health care providers. Am J Med Genet 119A:297, 2003

40. Stephenson J: As discoveries unfold, a new urgency to bring genetic literacy to physicians. JAMA 278:1225, 1997

41. Touchette N, Holtzman NA, Davis JG, et al: Toward the 21st Century: Incorporating Genetics into Primary Health Care. Cold Spring Harbor Laboratory Press, Woodbury, NY, 1997

42. Seashore MR, Wappner RS: Genetics in Primary Care and Clinical Medicine. Appleton & Lange, Stamford, Connecticut, 1996

43. Holtzman NA, Watson MS: Promoting safe and effective genetic testing in the United States: final report of the task force on genetic testing. National Institutes of Health/Department of Energy, September 1997

44. Green MJ, Botkin JR: "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. Ann Intern Med 138:571, 2003

45. McCrary SV, Allen B, Moseley R, et al: Ethical and practical implications of the human genome initiative for family medicine. Arch Fam Med 2:1158, 1993

IX HYPOGLYCEMIA

F. John Service, m.d., ph.d.

Definition

Hypoglycemia is a clinical syndrome of diverse etiologies characterized by episodes of low blood glucose. These episodes are typically marked by autonomic manifestations such as trembling, sweating, nausea, and, in more severe episodes, central nervous system manifestations (neuroglycopenia) such as dizziness, confusion, and headache.

Classification

Hypoglycemic disorders have long been categorized as fasting or postprandial (reactive). This classification lacks practical value. Insulinoma, which is the archetypal cause of fasting hypoglycemia, may produce symptoms postprandially and, indeed, in some cases solely postprandially. Patients with factitious hypoglycemia evince symptoms irrespective of meals.

A more useful approach for the practitioner is a classification based on the patient's clinical characteristics. Persons who appear otherwise healthy have hypoglycemic disorders different from those of persons who are ill.

hypoglycemia in apparently healthy patients

In apparently healthy persons, single episodes of hypoglycemia may result from accidental drug ingestion (e.g., ethanol in children). In addition to ethanol, salicylates and quinine can lower blood glucose levels; the combined effects of ethanol and quinine are responsible for so-called gin-and-tonic hypoglycemia. The healthy-appearing adult patient with a history of repeated episodes of neuroglycopenia usually has a disorder involving excessive insulin production, such as insulinoma; rarely, the hypoglycemia is factitious, caused by surreptitious or inadvertent use of a hypoglycemic agent (e.g., insulin or a sulfonylurea) [*see* Conditions That Cause Hypoglycemia, *below*].

Hypoglycemia may occur in patients who have coexistent disease but whose disease is being controlled with medical treatment. Typically, the hypoglycemia in these cases is a side effect of the medication being used to treat the coexistent disease, or it results from the mistaken dispensing of a sulfonylurea instead of the prescribed drug.

hypoglycemia in ill patients

Illness can lead to hypoglycemia through a variety of mechanisms, only some of which involve insulin and not all of which are known. Many illnesses (e.g., renal failure and sepsis) are known to pose the risk of low blood glucose levels [*see Table 1*]; hypoglycemia in a patient with one of these illnesses requires little if any investigation of its cause. However, not all patients with a disease that has a proclivity to generate hypoglycemia actually experience low blood glucose levels. Why only some ill patients experience hypoglycemia is unknown.

Hospitalized patients are at increased risk for hypoglycemia, often from iatrogenic factors. In any inpatient with hypoglycemia, medication should be considered a potential cause.

Low blood glucose levels may be found on laboratory testing of ill patients who have no symptoms of hypoglycemia. In patients with leukemia or severe hemolysis, the hypoglycemia may be an artifact resulting from consumption of glucose in the blood collection tube by large numbers of leukocytes or by nucleated red blood cells, respectively.[1,2] Patients with glycogen storage disease may be asymptomatic because they have adapted to lifelong hypoglycemia from their disease.[3]

Diagnosis

Although the diagnosis of hypoglycemia requires the measurement of blood glucose, such measurement often is not feasible when symptoms arise during activities of ordinary life. Under these circumstances, the physician must take a detailed history to determine whether to proceed with further evaluation. The history should include a full description of the patient's symptoms and the circumstances under which they occur.

A medication history is also an important aspect of the evaluation in a patient with clinical manifestations of hypoglycemia, especially if the onset coincides with the filling of a new prescription. Because of the potential for drug error, all medications taken by the patient should be identified by a medical professional, such as a physician or pharmacist.

clinical manifestations

The symptoms of hypoglycemia have been classified into two major groups: autonomic and neuroglycopenic. In a study of experimentally induced hypoglycemia in diabetic and nondiabetic persons, a principal-components analysis assigned sweating, trembling, feelings of warmth, anxiety, and nausea to the autonomic group and dizziness, confusion, tiredness, difficulty in speaking, headache, and inability to concentrate to the neuroglycopenic group. Hunger, blurred vision, drowsiness, and weakness could not be confidently assigned to either group.[4] In a retrospective analysis of 60 patients with insulinomas, 85% had various combinations of diplopia, blurred vision, sweating, palpitations, and weakness; 80% had confusion or abnormal behavior; 53% had amnesia or went into coma during the episode; and 12% had generalized seizures.[5]

The symptoms of hypoglycemia differ between persons but are nevertheless consistent from episode to episode in any one person.[6,7] There is no consistent chronologic order to the evolution of symptoms; autonomic symptoms do not always precede neuroglycopenic ones. In many patients, neuroglycopenic symptoms are the only ones observed.[7] Patients who have autonomic symptoms only are unlikely to have a hypoglycemic disorder. An additional factor that influences the generation of symptoms in hypoglycemia is their blunting by earlier hypoglycemic episodes.

physical examination

In patients who appear healthy, with or without coexistent compensated disease, the physical examination is normal or reveals only minor abnormalities that are unlikely to be germane to the underlying hypoglycemic disorder. In patients suspected of having factitious hypoglycemia from injection of insulin, a search for needle-puncture sites is fruitless. In ill patients with a primary disorder that can cause hypoglycemia, the results of physical examination will reflect that disease. For the patient observed while hypoglycemic, findings may include diaphoresis,

Table 1 Causes of Hypoglycemia

Drugs	Disopyramide Ethanol Haloperidol Quinine Salicylates
Drugs in specific illnesses	Pentamidine in *Pneumocystis* pneumonia Propoxyphene in renal failure Quinine in malaria Trimethoprim-sulfamethoxazole in renal failure Topical salicylates in renal failure
Endogenous hyperinsulinism	Insulinoma Islet hyperplasia/nesidioblastosis Persistent hyperinsulinemic hypoglycemia of infancy Noninsulinoma pancreatogenous hypoglycemia syndrome Insulin autoimmune hypoglycemia
Conditions that predispose to hypoglycemia	Neonatal Infant small for gestational age Erythroblastosis fetalis Infant of diabetic mother Cyanotic congenital heart disease Beckwith-Wiedemann syndrome Inherited Defects in amino acid and fatty acid metabolism Glycogen storage disease Hereditary fructose intolerance Isolated adrenocorticotropic hormone (ACTH) deficiency Isolated growth hormone deficiency Acquired Addison disease Carnitine deficiency Intense exercise Hypopituitarism Heart failure Lactic acidosis Severe liver disease Postoperative status Renal failure Reye syndrome Sepsis Shock Spinal muscular atrophy Starvation Anorexia nervosa Large mesenchymal tumors (fibroma, sarcoma, small cell carcinoma, mesothelioma)

widened pulse pressure, and neurologic abnormalities ranging from slowed mentation or withdrawal from spontaneous communication to more overt confusion, erratic behavior, coma, seizure, and hypothermia.

LABORATORY TESTS

Serum Glucose

Studies of acute insulin-induced hypoglycemia in healthy persons have shown that the threshold for the development of symptoms is a serum glucose concentration of approximately 60 mg/dl; the threshold for impairment of brain function is approximately 50 mg/dl.[8,9] These measurements were taken from arterialized venous blood (i.e., blood drawn from a vein in a heated hand [the application of heat shunts arterial blood into the venous system]); comparable levels in venous blood would probably be about 3 mg/dl lower. The rate of decrease in the serum glucose level does not influence the occurrence of the symptoms and signs of hypoglycemia.

Because symptoms of hypoglycemia are nonspecific, it is necessary to verify their origin. This is accomplished by applying a set of criteria first proposed by Whipple in 1938. The Whipple triad comprises spontaneous symptoms consistent with hypoglycemia, a low serum glucose concentration at the time the symptoms occur, and relief of the symptoms through normalization of the glucose level.[10]

A normal serum glucose concentration, reliably obtained during the occurrence of spontaneous symptoms, eliminates the possibility of a hypoglycemic disorder. Capillary glucose measurements that patients take themselves with a blood glucose meter during the occurrence of spontaneous symptoms are often unreliable, because nondiabetic patients usually are not experienced in this technique and because the measurements are obtained under adverse circumstances. Patients with a confirmed low serum glucose level (< 50 mg/dl) or a history of neuroglycopenic symptoms should undergo further testing. This is best accomplished with a prolonged fast.

The Prolonged (72-Hour) Fast

The prolonged (72-hour) fast is the classic diagnostic test for hypoglycemia. It should be conducted in a standardized manner [*see Table 2*]. The fast may be undertaken to demonstrate the Whipple triad and thereby establish that hypoglycemia is the basis for the patient's symptoms. If the Whipple triad has already been documented, the fast may be conducted for the purpose of determining the mechanism of the hypoglycemia, through measurement of beta cell polypeptide and plasma sulfonylurea levels. In the latter case, the fast can be terminated when the serum glucose level drops to 55 mg/dl or less (or, better yet, ≤ 50 mg/dl), which is the concentration at which beta cell polypeptides should be suppressed. Not all patients will need the full 72 hours to accomplish the purpose for the fast. In a study of 170 patients with surgically proven insulinomas, termination of the fast occurred within 12 hours in 33% of patients, within 24 hours in 65%, within 36 hours in 84%, within 48 hours in 93%, and within 72 hours in 99%.[11] Truncation of the fast at 48 hours, if hypoglycemia has not occurred by then, risks misdiagnosis.

Starting the fast overnight has allowed 40% of patients (including those with insulinoma and other causes of hypoglycemia) to conclude their fast in the outpatient endocrine-testing unit. Patients whose fast is not completed by the end of the business day are admitted to the hospital to complete the fast.

The decision whether to end the fast may not be easy to make when the Whipple triad is the goal. Because of delays in the availability of glucose measurements, the bedside glucose meter may have to serve as a guide. Some patients have slightly depressed glycemic levels without symptoms or signs of hypoglycemia. In other patients, fasting evokes the symptoms they experience in ordinary life but their serum glucose levels are not in the hypoglycemic range. In such instances, symptoms cannot be attributed to hypoglycemia. To complicate matters, young, lean, healthy women—and, to a lesser degree, some men—may have serum glucose concentrations in the range of 40 mg/dl or even lower during prolonged fasting.[12] Careful examination and

testing for subtle signs or symptoms of hypoglycemia should therefore be conducted repeatedly when the patient's serum glucose level is near or in the hypoglycemic range. To end the fast solely on the basis of a low serum glucose level, in the absence of symptoms or signs of hypoglycemia, may jeopardize accurate diagnosis. On the other hand, failing to appreciate the manifestations of neuroglycopenia and, hence, concluding that the results of the fast are negative is an equally egregious error. It is essential to monitor patients closely during the fast and to be vigilant for subtle signs of neuroglycopenia.

Beta Cell Polypeptides and Their Surrogates

Concentrations of beta cell polypeptides (insulin, C-peptide, and proinsulin) are interpreted in the context of the concomitant serum glucose concentration. The normal overnight fasting ranges for these polypeptides do not apply when the serum glucose level is low. When immunochemiluminometric assays (ICMA) are used, the criteria for endogenous hyperinsulinemia are as follows: serum insulin, 3 µU/ml or greater; C-peptide, 200 pmol/L or greater; and proinsulin, 5 pmol/L or greater [*see Figure 1*].[13]

Insulin concentrations rarely exceed 100 µU/ml in patients with insulinomas. Values above this level suggest recent insulin administration or the presence of insulin antibodies.

Ratios of glucose to insulin, and vice versa, have no diagnostic utility [*see Figure 1*]. The molar ratio of insulin to C-peptide is the same for patients with insulinomas and healthy persons (approximately 0.2). The molar ratio of proinsulin to insulin appears to be higher in persons with insulinoma, but it provides poor diagnostic utility.

Because insulin has an antiketogenic effect, serum levels of the ketone body β-hydroxybutyrate can be used as a surrogate for measurement of insulin. The serum β-hydroxybutyrate level is low—2.7 mmol/L or less—in patients with insulin-mediated hypoglycemia; normal persons and those with non–insulin-mediated hypoglycemia have higher levels [*see Figure 1*].[13]

At the end of the fast, the patient is given an intravenous dose of 1 mg of glucagon, and the subsequent glucose response is measured. Because insulin is glycogenic and antiglycogenolytic, the glucagon injection results in an increase in the serum glucose level of 25 mg/dl or greater in patients with insulin-mediated hypoglycemia, whereas normal persons or those with non–insulin-mediated hypoglycemia have lesser increases [*see Figure 1*].[13] An exuberant serum insulin response to intravenous glucagon has been considered an indication of insulinomas, but unfortunately, no normative data have been generated for this test. Measurement of beta cell polypeptides and insulin surrogates (β-hydroxybutyrate and glucose response to intravenous glucagon) has diagnostic utility only when the serum glucose level is 60 mg/dl or lower at the end of the fast.

Table 2 Protocol for Prolonged Supervised Fast

1. Date the onset of the fast as of the last ingestion of calories. Discontinue all nonessential medications.
2. Allow the patient to drink calorie-free and caffeine-free beverages.
3. Ensure that the patient is active during waking hours.
4. Measure plasma glucose, insulin, C-peptide, and, if an assay is available, proinsulin in the same specimen. Repeat measurements every 6 hr until the plasma glucose drops below 60 mg/dl; then repeat the measurements every 1–2 hr.
5. When the plasma glucose is less than 45 mg/dl and the patient has symptoms or signs of hypoglycemia, measure plasma glucose, insulin, C-peptide, proinsulin, β-hydroxybutyrate, and sulfonylurea in the same specimen; then inject 1 mg of glucagon I.V. and measure plasma glucose after 10, 20, and 30 min.
6. Feed the patient.

Sulfonylureas and Meglitinides

Persons with hypoglycemia from inappropriate use of sulfonylureas or meglitinides (e.g., repaglinide) have concentrations of beta cell polypeptides that are identical to those observed in persons with insulinoma. Consequently, plasma assays for these drugs is an essential aspect of the evaluation. I use a highly sensitive and accurate liquid chromatographic tandem mass spectroscopy method to identify these drugs. A positive assay suggests either covert or inadvertent usage.

Insulin Antibodies

An assay for insulin antibodies should be done in every patient with clear evidence of hypoglycemia. The detection of insulin antibodies in a nondiabetic patient was once considered to be firm evidence of insulin factitious hypoglycemia, especially when animal insulin was the only commercially available type. Currently, most patients with factitious hypoglycemia have no detectable insulin antibodies, possibly because of the use of human insulin, which is less antigenic than beef or pork insulin. Rather, the presence of insulin antibodies, especially in high titers, is diagnostic of insulin autoimmune hypoglycemia (IAH) (see below).[13] Very low titers of insulin antibodies may sometimes be detected in persons without hypoglycemia[14] and, in rare instances, in patients with insulinomas.

Glycated Hemoglobin

Measurement of glycated hemoglobin is not a standard aspect of the clinical evaluation of hypoglycemia. Concentrations of glycated hemoglobin are statistically significantly lower in patients with insulinomas than in normal persons, but there is too much overlap between the two groups for this test to provide a diagnostic criterion.[6]

Oral Glucose Tolerance Test

The oral glucose tolerance test should not be used for the evaluation of hypoglycemia, because it is fraught with risk of misdiagnosis. At least 10% of healthy persons have serum glucose nadirs below 50 mg/dl, and the results of the test do not correlate with serum glucose responses to a mixed meal (i.e., a meal containing a balance of proteins, carbohydrates, and fat).

Mixed-Meal Test

For persons with a history of neuroglycopenic symptoms within 5 hours after food ingestion, a mixed-meal test may be conducted. The test is considered to be positive if the patient experiences neuroglycopenic symptoms when a concomitant serum glucose level measures 50 mg/dl or less. A positive mixed-meal test does not provide a diagnosis, only biochemical confirmation of the history.

C-Peptide Suppression Test

C-peptide is formed during the conversion of proinsulin to insulin by the pancreatic beta cells. In the C-peptide suppression

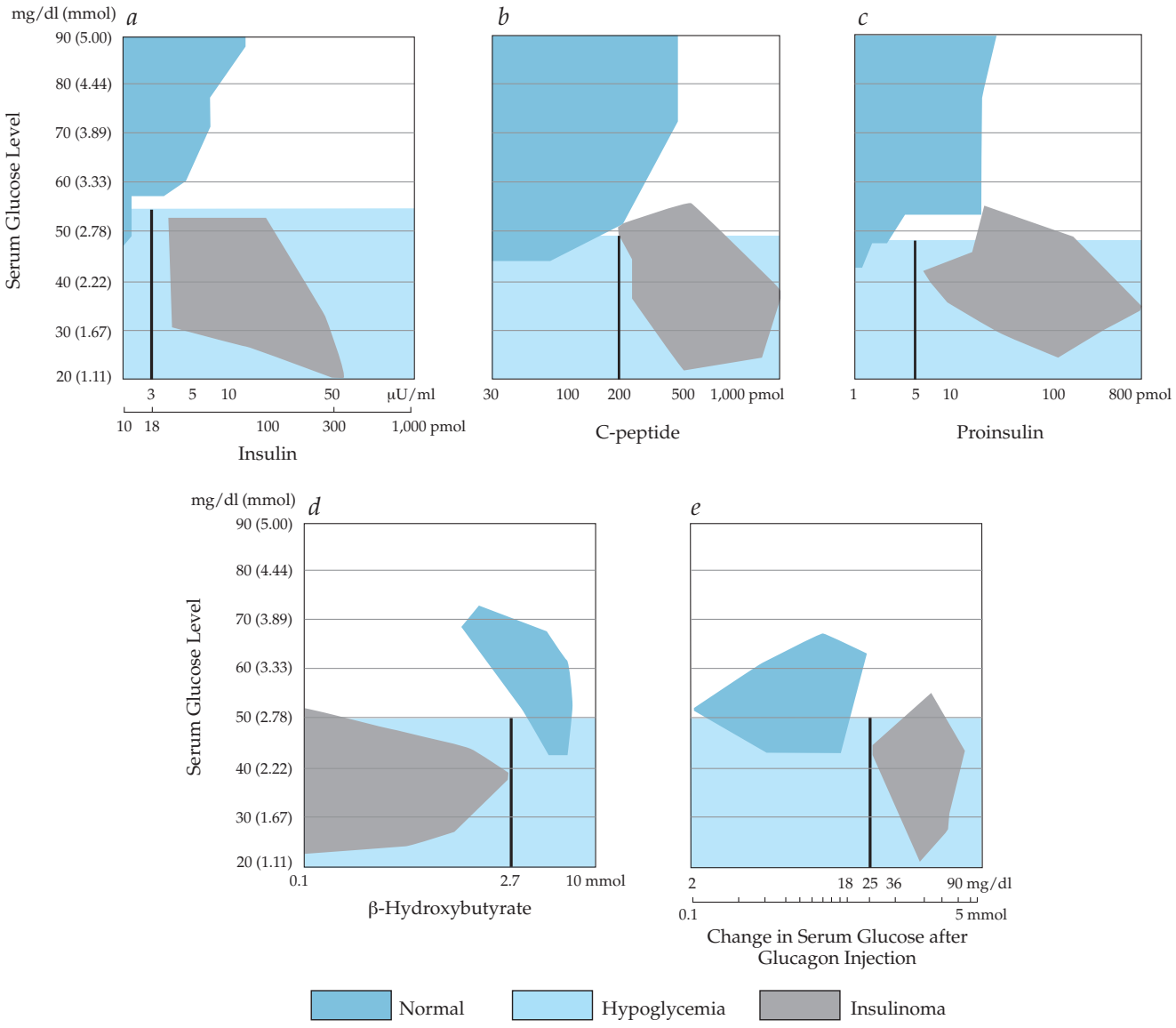


Figure 1 **During the 72-hour fast, levels of serum glucose are compared with serum levels of insulin (*a*), C-peptide (*b*), proinsulin (*c*), and β-hydroxybutyrate (*d*). At the end of the fast, 1 mg of glucagon is injected intravenously, and its effect on glucose levels is measured (*e*). Normal patients may have glucose levels that drop into the hypoglycemic range, so careful documentation of hypoglycemic symptoms is necessary.**

test, the patient fasts overnight and then receives an hour-long intravenous infusion of insulin, during which levels of serum glucose and C-peptide are measured. In normal patients, hypoglycemia from the exogenous insulin results in suppression of C-peptide production; patients with insulinomas have higher levels of C-peptide. When the likelihood of a hypoglycemic disorder is not high, a normal result on the C-peptide suppression test may preclude the need for a 72-hour fast. Interpretation of the C-peptide suppression test requires normative data appropriately adjusted for the patient's body mass index and age.[7]

Intravenous Tolbutamide Test

In the past, serum glucose response to an intravenous injection of tolbutamide was used in the diagnosis of insulinoma. This test is potentially dangerous and is less accurate than other tests for insulinoma and, therefore, has been rarely used in recent years.

Conditions That Cause Hypoglycemia

The causes of hypoglycemia in healthy-appearing adults encompass the following conditions: insulinoma, factitious hypoglycemia from insulin or sulfonylurea, noninsulinoma pancreatogenous hypoglycemia syndrome (NIPHS), and insulin autoimmune hypoglycemia.

INSULINOMA

Epidemiology

Between 1927 and 1986, 224 hypoglycemic patients underwent their first pancreatic exploration at the Mayo Clinic and were found to have insulinoma. Because of the relatively large number of cases of insulinoma treated at the Mayo Clinic, in comparison with other medical centers, and the comprehensive epidemiologic database that the Mayo Clinic maintains for Olm-

sted County, Minnesota, it was possible to determine the population-based incidence of insulinoma, the risk of recurrence, and the survival in patients with insulinoma.[15]

The median age of the Mayo Clinic insulinoma patients was 47 years, with a range of 8 to 82 years; 59% were female. The incidence in Olmsted County was 4 cases per 1 million person-years. Of the 224 patients, 7.6% had multiple endocrine neoplasia type I (MEN I) and 5.8% had malignant insulinoma. The risk of recurrence was greater in patients with MEN I (21%) than in those without this condition (7%). Over a 45-year period, overall survival of the total cohort was similar to the expected survival (78% versus 81%).

Insulinomas have been found in pregnant patients and in patients with type 2 (non–insulin-dependent) diabetes mellitus. One case of insulinoma in type 1 (insulin-dependent) diabetes mellitus has been reported.[16]

Diagnosis

Laboratory tests Insulinoma is characterized by hypoglycemia caused by elevated levels of endogenous insulin. Confirmation of the diagnosis requires exclusion of hypoglycemia from exogenous sources.

Localization Once a biochemical diagnosis of insulinoma has been made, the next step is localization. Success with the various modalities reflects local skill and experience. Great success has been seen with transabdominal ultrasonography and triple-phase spiral computed tomography. Magnetic resonance imaging and scintigraphy with indium-111 (In-111)–pentetreotide (OctreoScan) can also be used. I reserve endoscopic ultrasonography for complex cases. Percutaneous transhepatic portal venous sampling has been abandoned even by its former proponents.

In patients whose tumor is not found by ultrasonography or CT, the selective arterial calcium stimulation test provides a means to both regionalize and confirm endogenous hyperinsulinemia. This test involves serial injections of calcium into the splenic, gastroduodenal, and superior mesenteric arteries. Subsequent doubling of serum insulin concentrations in the right hepatic vein indicates hyperfunctioning beta cells in the part of the pancreas served by that artery.

There is general agreement that the best localization of insulinomas is achieved with intraoperative ultrasonography and careful mobilization and palpation of the pancreas by a surgeon experienced with insulinoma surgery. This approach has seen a 98% success rate in the identification of insulinoma. After this test, these patients go straight to surgery.

Management

The treatment of choice for insulinomas is surgical removal. Depending on the lesion, the surgery required may range from enucleation of the insulinoma to subtotal pancreatectomy. It is advisable for the surgery to be performed at an institution with expertise in the management of insulinoma.

Medical therapy is less effective than tumor resection, but the former can be used in patients who are not candidates for surgery, who refuse surgery, or whose surgery is unsuccessful. The most effective medication for controlling symptomatic hypoglycemia in these patients is diazoxide, which lowers insulin secretion. Diazoxide is given in divided doses of up to 1,200 mg daily. Side effects include edema, which may require high doses of loop diuretics, and hirsutism. Other medications for insulinomas include verapamil, phenytoin, and octreotide.

FACTITIOUS HYPOGLYCEMIA

The term factitious (or factitial) has been used in medical parlance to imply covert patient activity. The consideration of such a possibility often changes the patient-physician relationship, leading the physician to feel deceived and the patient to feel mistrusted. However, the pejorative connotation with which factitious illness has been encumbered requires softening because some patients with factitious disease suffer through no fault of their own.

Epidemiology

Factitious hypoglycemia is more common in women and occurs most often in the third or fourth decade of life. Many of these patients work in health-related occupations.

Factitious hypoglycemia in patients with diabetes is probably more common than the incidence noted in published series.[17] Confirmation of the diagnosis in these cases can be very difficult. When deprived of access to hypoglycemic agents, diabetic patients with factitious hypoglycemia become hyperglycemic.

Etiology

Factitious hypoglycemia results from the use of insulin or the use of sulfonylureas or meglitinides that stimulate insulin secretion. The most common form of factitious hypoglycemia is the covert self-administration of a hypoglycemic drug or insulin by a patient without diabetes or the inappropriate manipulation of hypoglycemic drugs or insulin by a patient with diabetes. Less often, a parent may administer a hypoglycemic agent to a child; this is a form of child abuse.[18] In all reported cases, the alleged perpetrator was the patient's mother, who had ready access to insulin. Insulin has also been used to attempt suicide or homicide.[19]

There are increasing numbers of patients who, by taking a prescribed medication in good faith, incur hypoglycemia because a sulfonylurea was mistakenly dispensed.[20] In most instances, confusion in dispensing the drug arose because of similarity in spelling between the intended medication and the sulfonylurea. In some cases, however, the dispensing error was a result of negligence. On occasion, cases have arisen in which a nondiabetic person mistakenly takes hypoglycemic medication belonging to another member of the household.

Diagnosis

The possibility of factitious hypoglycemia should be considered in every patient undergoing evaluation for a hypoglycemic disorder, especially when the hypoglycemia has a chaotic occurrence—that is, when it has no relation at all to meals or fasting. All medications should be identified; the assistance of a pharmacist is desirable. The practice of searching personal effects and labeling insulin with a traceable substance that can be detected in blood or urine is probably unacceptable in the current climate of patients' rights.

The diagnosis of factitious hypoglycemia can usually be established by measuring serum insulin, sulfonylurea, and C-peptide when the patient is hypoglycemic. If a spontaneous episode of hypoglycemia is not observed, the patient should undergo a 72-hour fast. The results of the fast may be negative, however, should the patient not take the offending agent.

In a patient whose hypoglycemia results from covert use of a hypoglycemic agent, the agent will be present in the blood. A sensitive method such as liquid chromatography linked to mass

spectroscopy should be used for the detection of sulfonylureas and meglitinides.

In insulin-related factitious hypoglycemia, the serum insulin level is high and the C-peptide level is suppressed, usually being close to the lower limit of detection. This observation applies both to nondiabetic patients and to those with type 2 diabetes. Patients with type 1 diabetes are characteristically severely insulin deficient and have low or undetectable serum concentrations of C-peptide. Although the C-peptide values in these patients cannot be further suppressed, confirmation that the values are low during a hypoglycemic episode eliminates any consideration of endogenous hyperinsulinism.

Management

Treatment of factitious hypoglycemia is simple: the patient stops taking the offending medication. The difficulty involved when medication is taken in error is identification of the drug. In the case of deliberate covert use, psychiatric referral is indicated.

NONINSULINOMA PANCREATOGENOUS HYPOGLYCEMIA SYNDROME

There have been cases of adults who do not have insulinomas but have hypoglycemia resulting from postprandial hypersecretion of insulin by pancreatic beta cells. Because of the unique clinical, diagnostic, radiologic, surgical, and histologic features of this disorder, it warrants designation as a new syndrome. We have termed it noninsulinoma pancreatogenous hypoglycemia syndrome, or NIPHS.[21]

Epidemiology

Like insulinoma, NIPHS affects patients across a broad age range—16 to 78 years, in one series—and causes severe neuroglycopenia, with loss of consciousness and, in some cases, generalized seizures. Unlike insulinoma, NIPHS occurs predominantly in males (70%).

Pathophysiology and Pathogenesis

Histologic analysis of pancreatic tissue from patients with NIPHS shows cells budding off ducts, which is best seen by chromogranin A and insulin immunohistochemical staining. Islet cell hypertrophy is also evident. No gross or microscopic tumor has been identified on hematoxylin-eosin–stained sections in any NIPHS patients.

Whether islet hypertrophy, nesidioblastosis, or both are pathogenic in these patients is open to question, as is the case with persistent hyperinsulinemic hypoglycemia of infancy (PHHI). However, a role for some form of diffuse islet cell dysfunction appears well established in these cases. Whatever the pathologic process may be, it is nonfocal, yet it does not necessarily involve the entire pancreas uniformly.

The histologic findings in NIPHS are similar to those in PHHI. Although familial forms of PHHI may be associated with mutations in the *Kir6.2* and *SUR1* genes, analysis of these genes in NIPHS patients has not shown such mutations.[22] However, these patients may have common mutations at another, as yet unspecified, locus.

Diagnosis

Clinical manifestations Symptoms of NIPHS occur primarily in the postprandial state 2 to 4 hours after eating. Although insulinoma patients may experience symptoms postprandially, they also have symptoms during food deprivation. It is extremely rare for insulinoma patients to have symptoms solely in the postprandial state.

Laboratory tests Patients with NIPHS have low serum glucose levels and elevated serum insulin levels in the postprandial period. Because of the short half-life of insulin, the criteria for hyperinsulinemia used in the fasting state appear to apply in the postprandial state, as long as the low glucose level occurs more than 30 minutes from the peak postprandial insulin level. Supervised 72-hour fasts have shown normal results in patients with NIPHS, whereas a negative 72-hour fast in a patient with insulinoma is a rare occurrence.

The selective arterial calcium stimulation test has shown positive results for patients with NIPHS.[23,24] All radiologic localizing studies in patients with NIPHS (transabdominal ultrasonography, triple-phase CT, celiac axis angiography, and intraoperative ultrasonography) have been negative for insulinoma.

Management

Gradient-guided partial pancreatectomy has been effective in relieving symptoms in patients with NIPHS. The pancreas is resected to the left of the superior mesenteric vein when results of the selective arterial calcium stimulation test are positive only for the splenic artery, and the pancreas is resected to the right of the superior mesenteric vein when the test is positive for an additional artery. Fortunately, gradient-guided debulking of the pancreas can ameliorate the symptoms of NIPHS even in patients whose disease would appear to have involved the whole pancreas. In rats, the mechanism for this effect may be related to decreased insulin secretion, attributed to reduced glucose transporter GLUT2, in remnant pancreas after partial pancreatectomy.[25] Unfortunately, recurrence of hypoglycemia after a few symptom-free years has developed in a few of the NIPHS patients.

INSULIN AUTOIMMUNE HYPOGLYCEMIA

Epidemiology

IAH is an extraordinarily rare disorder that is observed primarily, although not exclusively, in persons of Japanese and Korean ethnicity. The disorder may occur at any age. IAH tends to be be self-limited in Asians, but it may be persistent in whites. There is no gender predilection. Many patients have an ongoing autoimmune disorder or a history of treatment with a sulfhydryl-containing drug such as antithyroid medication. No patients have had a history of exposure to insulin.[14]

Pathogenesis

IAH is characterized by the presence of autoantibodies to insulin or the insulin receptor. There is speculation that meal ingestion in these patients may result in the unbinding of insulin from these antibodies. However, measurements of total insulin and free insulin have shown no postprandial alteration in their relative concentrations. The mechanism for the generation of insulin antibodies is unknown but may involve enhanced immunogenicity resulting from an effect of the disulfide bond in drugs with a sulfhydryl component.

Diagnosis

Clinical manifestations Patients with IAH typically experience postprandial hypoglycemia resulting in neuroglycopenia. The symptomatic severity of IAS appears to vary greatly. Whites may become more seriously debilitated than Asians.

Laboratory tests Serum insulin levels are markedly elevated in IAH, because the insulin antibodies interfere with this assay. Values can be as high as 1,000 μU/ml. Oddly, C-peptide levels are usually not suppressed. Insulin antibody titers are very high, higher than those seen in insulin-treated diabetic patients. The antibodies may bind only to human insulin or to both human insulin and beef and pork insulin. The antibodies may be polyclonal or monoclonal, and they usually have characteristics similar to those that occur in patients with type 1 diabetes mellitus. It should be noted that very low titers of insulin antibodies may also be observed in healthy persons without hypoglycemia and occasionally in persons with insulinoma.

Management

Supportive treatment, such as frequent small meals, may be effective in IAH, especially for mild cases. For more severely affected patients, a variety of approaches have been tried, including glucocorticoids, immunosuppressants, plasmapheresis, octreotide, and diazoxide. Unfortunately, all these treatments usually fail. Use of partial pancreatectomy and splenectomy has led to amelioration but not complete resolution of symptoms.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Goodenow TJ, Malarkey WB: Leukocytosis and artifactual hypoglycemia. JAMA 237:1961, 1977

2. Macaron CI, Kadri A, Macaron Z: Nucleated red blood cells and artifactual hypoglycemia. Diabetes Care 4:113, 1981

3. Service FJ, Veneziale CM, Nelson RA, et al: Combined deficiency of glucose-6-phosphate and fructose-1,6-diphosphate: studies of glucagon secretion and fuel utilization. Am J Med 64:698, 1978

4. Hepburn DA, Deary IJ, Frier BM, et al: Symptoms of acute insulin-induced hypoglycemia in humans with and without IDDM: factor-analysis approach. Diabetes Care 14:949, 1991

5. Service FJ, Dale AJD, Elveback LR, et al: Insulinoma: clinical and diagnostic features of 60 consecutive cases. Mayo Clin Proc 51:417, 1976

6. Hassoun AAK, Service FJ, O'Brien PC: Glycated hemoglobin in insulinoma. Endocr Pract 4:181, 1998

7. Service FJ, O'Brien PC, Kao PC, et al: C-peptide suppression test: effects of gender, age and body mass index: implications for the diagnosis of insulinoma. J Clin Endocrinol Metab 74:204, 1992

8. Schwartz NS, Clutter WE, Shah SD, et al: Glycemic thresholds for activation of glucose counterregulatory systems are higher than the threshold for symptoms. J Clin Invest 79:777, 1987

9. Mitrakou A, Ryan C, Veneman T, et al: Hierarchy of glycemic thresholds for counterregulatory hormone secretion, symptoms and cerebral dysfunction. Am J Physiol 260:E57, 1991

10. Whipple AE: The surgical therapy of hyperinsulinism. J Int Chir 3:237, 1938

11. Service FJ, Natt N: Clinical perspective: the prolonged fast. J Clin Endocrinol Metab 85:3973, 2000

12. Merimee TJ, Fineberg SE: Homeostasis during fasting: II. Hormone substrate differences between men and women. J Clin Endocrinol Metab 37:698, 1973

13. Service FJ: Diagnostic approach to adults with hypoglycemic disorders. Endocrinol Metab Clin North Am 28:519, 1999

14. Redmon JB, Nuttall FQ: Autoimmune hypoglycemia. Hypoglycemic Disorders 28:603, 1999

15. Service FJ, McMahon MM, O'Brien PC, et al: Functioning insulinoma—incidence, recurrence, and long-term survival of patients: a 60-year study. Mayo Clin Proc 66:711, 1991

16. Svartberg J, Stridsberg M, Wilander E, et al: Tumour-induced hypoglycaemia in a patient with insulin-dependent diabetes mellitus. J Intern Med 239:181, 1996

17. Tattersall RB, Gregory R, Selby C, et al: Course of brittle diabetes: 12 year follow up. BMJ 302:1240, 1991

18. Mayefsky JH, Sarnaik AP, Postellon DC: Factitious hypoglycemia. Pediatrics 69:804, 1982

19. Marks V, Teale JD: Hypoglycemia: factitious and felonious. Endocrinol Metab Clin North Am 28:579, 1999.

20. Hooper PL, Tello RJ, Burstein PH, et al: Pseudoinsulinoma: the Diamox-Diabinese switch. N Engl J Med 323:448, 1990

21. Service FJ, Natt N, Thompson GB, et al: Noninsulinoma pancreatogenous hypoglycemia: a novel syndrome of hyperinsulinemic hypoglycemia in adults independent of mutations in *Kir6.2* and *SUR1* genes. J Clin Endocrinol Metab 84:1582, 1999

22. Thomas PM, Cote GJ, Wohlik N, et al: Mutations in the sulfonylurea receptor gene in familial hyperinsulinemic hypoglycemia of infancy. Science 268:426, 1995

23. Doppman JL, Chang R, Fraker DL, et al: Localization of insulinomas to regions of the pancreas by intraarterial stimulation by calcium. Ann Intern Med 123:269, 1995

24. O'Shea D, Rohrer-Theus A, Lynn JA, et al: Localization of insulinomas by selective intraarterial calcium injection. J Clin Endocrinol Metab 81:1623, 1996

25. Zangen DH, Bonner-Weir S, Lee CH, et al: Reduced insulin, GLUT2, and IDX-1 in beta-cells after partial pancreatectomy. Diabetes 46:258, 1997

X OBESITY

Jonathan Q. Purnell, m.d.

Obesity and its associated disorders are leading causes of morbidity and premature mortality around the world. Obese persons are also vulnerable to low self-esteem and depression because of the psychological and social stigmata that can be associated with obesity. Despite societal prejudicial perceptions that obesity develops because of deficient self-control, research has provided insight into the physiology behind unwanted weight gain. Indeed, during the past decade, the field of body-weight regulation (the study of the homeostatic mechanisms controlling body weight and fat content and the pathophysiology leading to unwanted weight gain or weight loss) has undergone an explosion in research, particularly in the area of neuroendocrine control of appetite and energy expenditure. As with other leading diseases in developed countries, such as hypertension and diabetes, obesity is recognized as a chronic condition resulting from an interaction between environmental influences and an individual's genetic predisposition to weight gain.

The initial evaluation of overweight and obese patients begins with the exclusion of secondary causes of weight gain and the identification of comorbid disorders such as hypertension, diabetes, heart disease, and sleep apnea. Once screening is completed, the approach to the treatment of overweight and obesity is similar to that of other chronic diseases: begin with lifestyle improvements, and then consider medical and surgical options. Although the weight loss that accompanies current therapeutic options is modest on average, the future promises better diagnostic and treatment options for obesity that are based on research into the mechanisms of weight regulation and their role in unwanted weight gain and maintenance of the obese state.

Definition of Obesity

Obesity is an abnormal accumulation of body fat in proportion to body size. Overweight persons have a body-fat proportion that is intermediate between normal and obese. Ideally, an obesity classification system would be based on a practical measurement of body fat that could be performed in the office, would accurately predict disease risk, and would apply to patients from diverse ethnic backgrounds. The most direct measures of body fat, such as underwater weighing or dual-energy x-ray absorptiometry (DXA) scanning, are impractical for use in a clinical setting. Indirect estimates of body fat are clinically more practical.

Classification of Obesity

BODY MASS INDEX

Body mass index (BMI), which is calculated by dividing the body weight in kilograms by height in meters squared, is a classification system that attempts to allow comparison of weights independent of stature across populations. Except in persons who have increased lean weight as a result of intense exercise (e.g., bodybuilders), BMI does correlate with percentage of body fat, but this relationship is independently influenced by sex, age, and race.[1] In the United States, data from the second National Health and Nutrition Examination Survey (NHANES II) were used to define obesity in adults as a BMI of 27.3 kg/m^2 or more for women and a BMI of 27.8 kg/m^2 or more for men.[2] These definitions were based on the gender-specific 85th-percentile values of BMI for persons 20 to 29 years of age. In 1998, however, the National Institutes of Health (NIH) Expert Panel on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults adopted the World Health Organization (WHO) classification for overweight and obesity.[3] The WHO classification, which predominantly applies to people of European ancestry, assigns an increasing risk for comorbid conditions—including hypertension, type 2 diabetes mellitus, and cardiovascular disease—to persons with higher BMIs [*see Table 1*] relative to persons of normal weight (i.e., those with a BMI between 18.5 kg/m^2 and 25 kg/m^2). Asian populations, however, are known to be at increased risk for diabetes and hypertension at lower BMI ranges than those for non-Asian groups.[4] Consequently, the WHO has suggested lower cutoff points for consideration of therapeutic intervention in Asians: a BMI of 18.5 to 23 kg/m^2 represents acceptable risk, 23 to 27.5 kg/m^2 represents increased risk, and 27.5 kg/m^2 or higher represents high risk.[5]

FAT DISTRIBUTION

In addition to an increase in total body fat, a proportionally greater amount of fat in the abdomen or trunk compared with fat in the lower extremities or hips has been associated with increased risk for diabetes, hypertension, and heart disease in both men and women.[6] Abdominal obesity is commonly reported as a waist-to-hip ratio, but it is most easily quantified by a single circumferential measurement obtained at the level of the superior iliac crest.[3] Current guidelines categorize men at increased relative risk for coronary artery disease, diabetes, and hypertension if they have a waist circumference greater than 40 inches (102 cm); women are at increased risk if their waist circumference exceeds 35 inches (88 cm) [*see Table 1*]. Thus, an overweight person with abnormal fat patterning may be at high risk for these diseases even if that person is not obese by BMI criteria. In those of Asian descent, abdominal (central) obesity is recognized to be a better predictor of comorbidity than BMI.[7] Therefore, the WHO has recommended lower waist circumference cutoffs to assign increased risk for comorbidities in this population: 36 inches (90 cm) or more in men and 32 inches (80 cm) or more in women.[4]

Table 1 Classification of Weight and Risk for Comorbid Conditions[2]

Classification	*Body Mass Index (kg/m^2)*	*Risk for Diabetes, Hypertension, and Cardiovascular Disease*	
		*Normal Waist Circumference**	*Increased Waist Circumference**
Underweight	< 18.5	Average	Average
Normal	18.5–24.9	Average	Average
Overweight	25–29.9	↑	↑↑
Obese	30–34.9	↑↑	↑↑↑
	35–39.9	↑↑↑	↑↑↑
	≥ 40	↑↑↑↑	↑↑↑↑

*Normal waist circumference is ≤ 102 cm (40 in) in men, ≤ 88 cm (35 in) in women.

Epidemiology

In the United States, the prevalence of overweight has been increasing over the past several decades [*see Figure 1*]. In the most recently published United States data (1999 to 2002), 65% of adults are overweight (BMI 25 to 30 kg/m^2), 30% of the total population are obese (BMI 30 to 40 kg/m^2), and 5% have a BMI of 40 kg/m^2 or higher.[8,9] The prevalence of obesity has also risen in some minority populations, with the highest rates found in some Native American groups, Hispanics, and African Americans; the lowest rates have been found in populations of Asian ancestry [*see Figure 2*].[9-12] Prevalence rates for obesity in the United States are also highest in populations with less education and lower income levels.[12] Internationally, obesity rates are generally lower than those in the United States.[13] However, even in societies that traditionally had the lowest prevalence of overweight and obesity, the rates of weight gain are beginning to meet or exceed those of Western societies.[14]

The age at which obesity is most prevalent has also increased. Until NHANES III (1988 through 1994), obesity in the United States peaked between the ages of 40 and 59 years, then declined in the older-age groups.[9] According to the most recent NHANES data (1999 to 2002), the prevalence of obesity now remains high past the age of 60 years, reaching 30.5% in men and 34.7% in women.[9] In studies that have measured body composition in unselected populations, fat mass also peaks just past middle age in men and women, but percent body fat continues to increase past this age, particularly in men, because of a proportionally greater loss in lean mass.[15-17] The menopausal period has also been associated with an increase in percent body fat and propensity for central fat distribution, even though total body weight changes very little during this time.[18] A propensity for greater abdominal adiposity has also been demonstrated in men,[19] in older individuals,[20] and in persons with impaired glucose tolerance or type 2 diabetes.[21]

Etiology and Genetics

Studies of populations, families, adoptions, and twins have established a strong genetic role in determining body weight. Estimates of the genetic contribution to the variance of relative body weight and adiposity range from a low of approximately 20% to a high of 90%.[22] The largest study to date to address the contribution of nature versus nurture to body weight, which used a dataset that included over 25,000 twin pairs and 50,000 biologic and adoptive family members, found that genetic factors accounted for 67% of the variance in adiposity in men and women.[22] Rarely, childhood-onset obesity will manifest itself as a result of a single-gene obesity syndrome, such as Prader-Willi syndrome or Bartlett-Biedel syndrome, or from a mutation in one of the genes encoding proteins involved with body-weight regulation, such as the pro-opiomelanocortin (*POMC*) gene, which makes α-melanocortin–stimulating hormone (α-MSH); the melanocortin receptor (MC4); leptin; the leptin receptor; and prohormone convertase enzymes.[23]

Although the genetics explaining the tendency toward overweight and obesity in the majority of the population remains to be elucidated, over 600 genetic markers have been described in association with obesity-related variables in humans (e.g., BMI, skin-fold thickness, waist-to-hip ratio, fat mass, and percent fat mass).[23] With time, discoveries of specific gene products, the role that these proteins play in the pathophysiology of weight regulation, and their interaction with the environment in the expression of unwanted weight gain should lead to more specific pharmacologic treatments for overweight and obese patients who fail to respond adequately to lifestyle measures alone.

Epidemiologic studies have identified several environmental factors that contribute to the continued weight gain documented over the past several decades in westernized countries. The foremost among these factors are an increasingly sedentary lifestyle (e.g., increased car use, community and work environments that discourage activity, and more time spent watching television) and the availability of energy-dense (high-fat, concentrated-sugar), low-fiber foods.[24-28] In children, the increased consumption of sugar-added beverages and reduction of dairy intake have also been associated with greater weight gain in prospective studies.[29,30] Similar environmental predictors of weight gain have been described in societies adopting Western lifestyles in the transition to First World economies.[14,31] Additional societal trends that are thought to have contributed to the increasing weight gain in the United States include smoking cessation (cigarette smoking is known to reduce body weight)[25,32] and eating a greater proportion of food away from home, particularly at fast-food restaurants, where food is typically very calorically dense.[28,33]

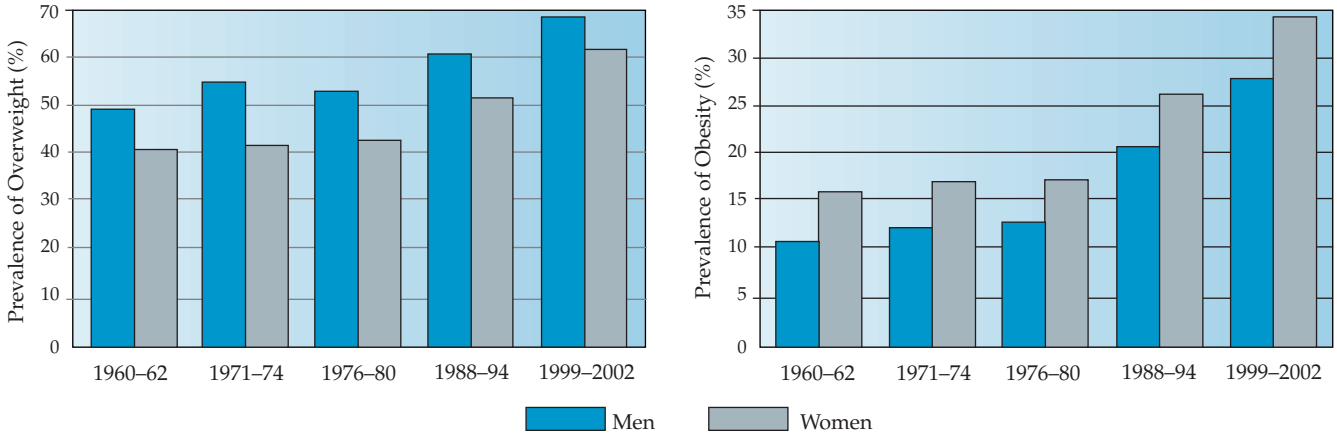


Figure 1 **Time trends of age-adjusted prevalence of overweight (BMI ≥ 25 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) in United States men and women who are 20 years of age and older.[8]**

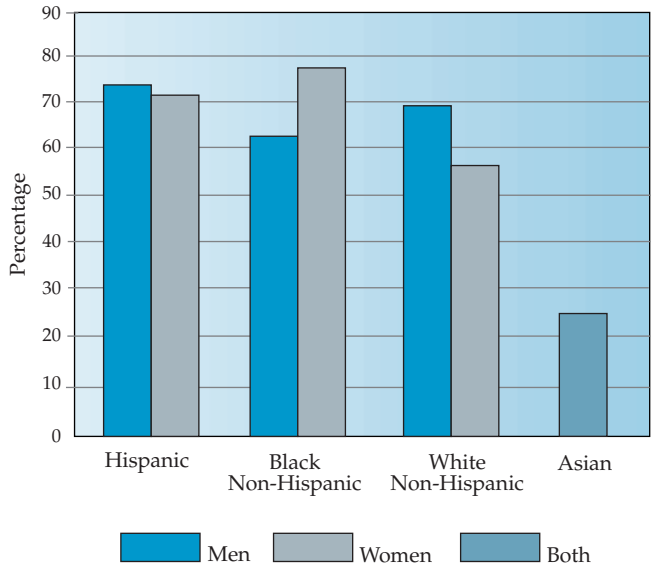


Figure 2 **Age-adjusted percentage of United States adults who were overweight (BMI ≥ 25 kg/m^2), by sex and race/ethnicity, 1999–2002.[8] Data for Asians include both men and women in 2003.[11]**

Pathophysiology and Pathogenesis

Arguably the most significant recent advances in the science of obesity have been in the area of neuroendocrine control of energy homeostasis, including the understanding of the mechanisms that lead to unwanted weight gain and the counterregulatory systems that restore weight lost after caloric restriction. At its most basic level, body weight is the end result of a balance between energy taken in and energy expended. Weight gain ensues when more energy is consumed than expended. Weight loss occurs through restriction of energy intake, increased energy output, or both. However, this simple model fails to incorporate what are now known to be complex homeostatic systems that counteract voluntary energy perturbations, whether they be forced overfeeding or caloric restriction.

A homeostatic model of weight regulation is conceptually identical to other tightly regulated systems in the body. For example, blood glucose levels reflect input from meals and hepatic stores balanced against clearance through uptake by peripheral tissues and excretion by the kidneys. Glucose levels are kept within a normal range by complex, integrated responses from insulin, glucagon, and other so-called counterregulatory hormones such as catecholamines, cortisol, and growth hormone, which regulate production and clearance. Elevated blood glucose levels and diabetes result when the secretion of primary regulators (i.e., insulin and glucagon) is impaired, when resistance to insulin signaling develops, or both.

Like glucose, body weight is regulated at multiple levels to maintain a normal range or set point through an interaction between systems that control meal-to-meal intake (satiety) and those that control relative fat mass (adiposity) [*see Figure 3*].[34] Although short-term (meal-to-meal) signals such as cholecystokinin have been studied for decades, a long-term afferent signal from the fat tissue, leptin, was not discovered until 1994.[35] Leptin is a hormone that is secreted by fat cells in direct proportion to total fat mass, is transported across the blood-brain barrier, and has receptors in hypothalamic nuclei that control appetite and energy expenditure.[34] When leptin levels decline with weight loss from caloric restriction or when they increase with overfeeding, altered signaling in central hypothalamic centers become integrated with other input signals (e.g., insulin and ghrelin) to set in motion systems that restore body weight to baseline. Therefore, most obese patients fail to sustain long-term weight loss with calorie restriction alone because of activation of these counterregulatory systems and their promotion of a positive energy balance. In this feedback-loop model of weight regulation, primary obesity results when leptin signaling to central centers is reduced (leptin resistance),[34] resulting in uncompensated weight gain that eventually reestablishes energy homeostasis at a higher body-weight set point and blood level of leptin, analogous to insulin levels rising in compensation for acquired insulin resistance.

Although this model oversimplifies the complex nature of body-weight regulation, it nonetheless provides a starting point for clinicians in their education of patients about the pathophysiology of obesity and in the rationale for medical management. To achieve sustained weight reduction in overweight and obese patients, interventions must prevent activation of counterregulatory systems that act to restore lost weight by increasing appetite or reducing energy expenditure. Future medical therapies will be based on an understanding of the body's weight regulatory system and will have greater promise for success in maintaining weight loss.

With aging, dysregulation of a number of hypothalamic-pituitary systems may contribute to increased fat mass and sarcopenia. For example, growth hormone secretion diminishes with age.[36] Prospective trials in older adults that involved replacing growth hormone and targeting levels of insulinlike growth factor–1 (IGF-1) to the midnormal to upper-normal range have demonstrated improved body composition (less fat, more lean tissue) and, in some studies, reduced central fat.[36-38] In addition, the decline in testosterone levels in men, the drop in estrogen levels in women at menopause, and increased levels of cortisol in both sexes may also contribute to reduced muscle mass, central fat distribution, or both.[18,39-41]

Diagnosis

The history, physical examination, and laboratory evaluation of overweight and obese patients are directed toward three goals: first, to identify secondary causes of obesity [*see* Differential Diagnosis, *below*]; second, to identify comorbid conditions [*see Figure 4*]; and third, to establish the patient's dietary and activity habits.

HISTORY

A number of the symptoms associated with diseases that can cause or contribute to unwanted weight gain, such as hypothyroidism or Cushing disease, occur frequently in overweight patients. These include fatigue, aches, cold intolerance, constipation, poor exercise tolerance, central obesity, loss of libido, and depression. Deciding when to screen a patient for secondary causes of obesity, therefore, can be a challenge for the practitioner. Establishing a pattern of weight gain may be helpful. A patient with a lifelong history of being heavy and a stable adult weight is unlikely to have a secondary cause of obesity. A sudden or rapid weight gain over a few months or years, however, especially when accompanied by onset of comorbid conditions, may correspond to the prescription of medications that contribute to excess weight gain (especially steroids and newer an-

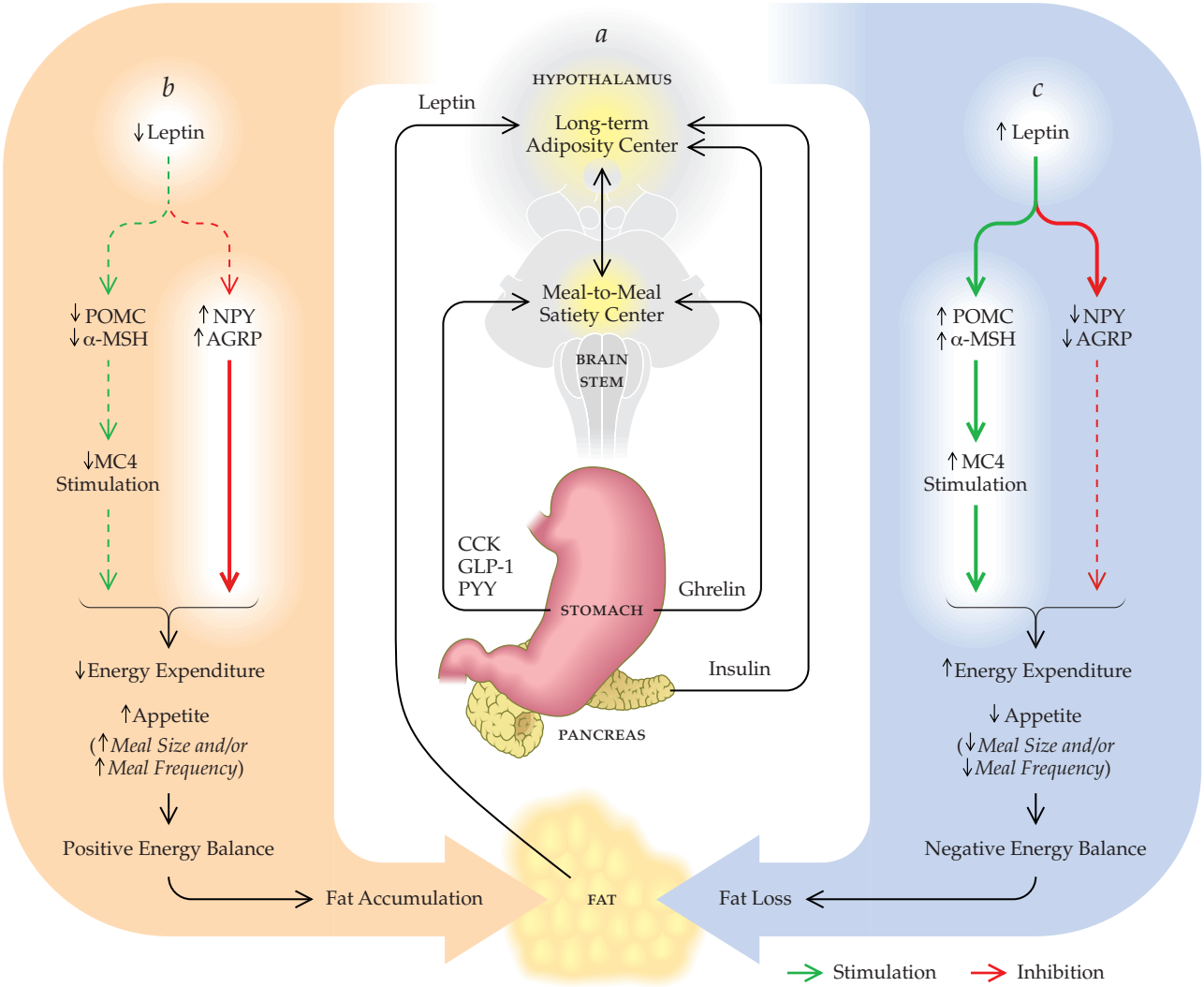


Figure 3 **(*a*) A feedback model for body-weight regulation in humans based on data from animal models.[34] Hypothalamic centers that control long-term energy homeostasis sense fat stores through circulating levels of leptin and insulin. Satiety signals (short-term or meal-to-meal regulators) from the gut are relayed through the brain stem to the hypothalamus, where they are integrated with signals reflecting fat stores. These integrated signals then affect appetite and energy expenditure so as to maintain body weight within a set point range. (*b*) Leptin controls appetite and energy expenditure in the hypothalamus by alternatively stimulating production of pro-opiomelanocortin (POMC) and α-melanocortin (α-MSH) and inhibiting production of neuropeptide-Y (NPY) and agouti-related protein (AGRP). α-MSH binds to the melanocortin-4 receptor (MC4), which inhibits appetite and increases energy expenditure. NPY and AGRP stimulate appetite while decreasing energy expenditure. Reduced leptin secretion, such as that which occurs after voluntary caloric restriction,[59,146] leads to enhanced NPY/AGRP signaling, diminished MC4 signaling, and positive energy balance once caloric restriction ceases.[147,148] On the other hand, overfeeding leads to increased fat mass and leptin secretion,[149] reduced NPY/AGRP signaling, increased MC4 signaling, and negative energy balance[148,150] until body weight is restored to baseline. (CCK—cholecystokinin; GLP-1—glucagonlike peptide–1; PYY—peptide YY)**

tipsychotics) or indicate onset of an illness that requires further evaluation.

The history should include questions about diseases for which overweight and obese patients are at higher risk, including hypertension, impaired glucose tolerance or diabetes, hyperlipidemia, heart disease, pulmonary disease, and sleep apnea. These conditions may cause minimal or no symptoms and therefore may be present for months or years before a diagnosis is made. Sleep apnea in particular is a common cause of fatigue and poor concentration or work performance in obese patients; these symptoms are often mistakenly ascribed to an abnormally functioning thyroid gland (despite normal results on thyroid function tests) or a so-called altered metabolism. This diagnosis may be missed unless the clinician specifically asks about characteristic symptoms: restless sleep at night, snoring or observed apnea, fatigue or headache upon awakening and during the daytime, and spontaneous daytime sleep when inactive or while driving. In severely obese patients, increasing peripheral edema, orthopnea, and worsening exercise tolerance may be symptoms of congestive heart failure or pulmonary hypertension and right-sided heart failure from severe sleep apnea. New-onset headaches may indicate normal-pressure hydrocephalus. Gastroesophageal reflux disease usually results in heartburn or an acid taste in the throat. During a period of weight gain, women may develop ir-

regular periods or symptoms of androgen excess. Although commonly diagnosed as polycystic ovary syndrome (PCOS), these findings differ from classic PCOS in that they occur after menarche and are not usually associated with polycystic ovaries.

Finally, inquiring about past and present dietary and activity habits is important for subsequent discussions of medical and surgical management. Most overweight and obese patients will have made numerous attempts to lose weight, through diets, exercise regimens, or commercial weight-loss programs. Because of unrealistic expectations and the inevitable weight regain that occurs, patients are often discouraged or leery of new advice. These failures can also compound feelings of guilt or inadequacy, fueling cycles of worsening self-image and depression. A broad survey of the types of foods people eat can often be accomplished in an office visit; in particular, asking about intake of calorically dense foods, including sodas, and frequency of meals outside the home may identify habits that can be improved. More detailed dietary analysis requires a visit with a nutritionist. Physical impediments such as arthritis, back pain, and asthma should be identified and treated so as to optimize daily activity and adherence to exercise recommendations.

PHYSICAL EXAMINATION AND LABORATORY TESTS

Height and weight measurements in the office are used to classify patients as overweight or obese according to BMI criteria [*see Table 1*]; however, these criteria may not apply to patients who have gained weight as the result of increased muscle mass from intensive exercise. Evaluation of abdominal obesity requires the use of a tape measure. A waist circumference (obtained at the level of the superior iliac crest) greater than 40 inches (102 cm) in a man or greater than 35 inches (88 cm) in a woman is considered abnormal.

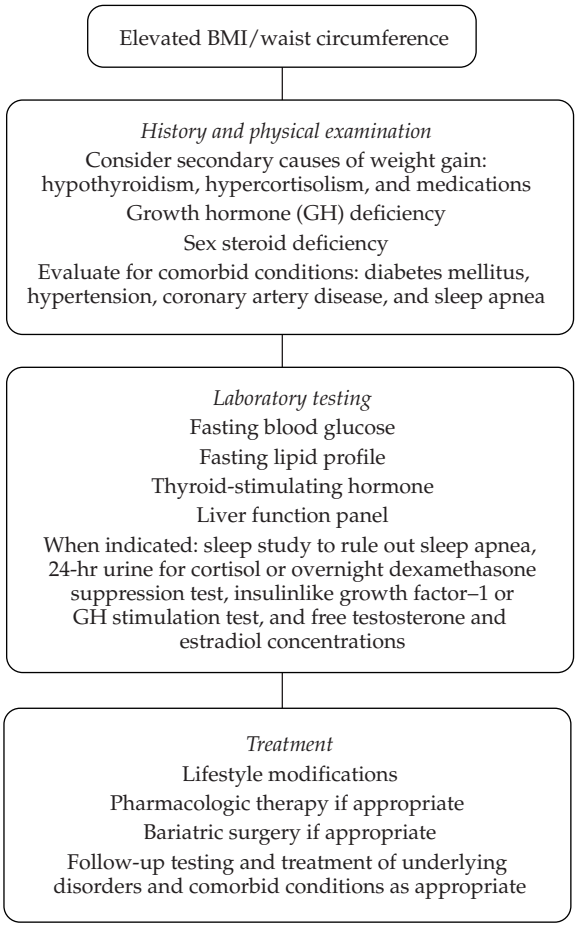


Figure 4 **Evaluation, laboratory testing, and treatment of overweight and obese patients.**

Specific physical findings that might indicate secondary causes of obesity include pretibial edema and delayed tendon reflexes (hypothyroidism), purple striae, supraclavicular fat pad enlargement, and muscle weakness (Cushing syndrome). Other aspects of the clinical evaluation focus on comorbid conditions. Documentation of hypertension requires properly obtained blood pressure measurements (i.e., using the correctly sized cuff for larger persons). Insulin resistance and type 2 diabetes may manifest themselves as acanthosis nigricans—patches of feathery-pigmented skin (hyperkeratotic and hyperpigmented) on the extensor surfaces of the hands and elbows, in the axilla, or on the neck [*see 2:I Cutaneous Manifestations of Systemic Diseases*]. Hepatomegaly can indicate hepatosteatosis, especially in centrally obese subjects.

With or without acanthosis nigricans, impaired glucose tolerance may be diagnosed by a fasting plasma glucose level between 100 and 125 mg/dl or a 2-hour glucose level between 140 and 200 mg/dl during an oral glucose tolerance test. Type 2 diabetes is diagnosed by two fasting blood glucose measurements of 126 mg/dl or greater, a 2-hour glucose level of 200 mg/dl or more during an oral glucose tolerance test, or a random glucose level of 200 mg/dl or greater and symptoms of diabetes.

Screening for macrovascular risk involves obtaining an electrocardiogram when appropriate and carefully examining the patient for xanthomata, which can indicate the presence of elevated blood levels of chylomicrons (eruptive xanthoma), type III hyperlipidemia (palmar xanthoma or tuberoeruptive xanthoma), or familial hypercholesterolemia (tendon xanthoma). Each of these physical manifestations of hyperlipidemia, although rare in a primary care practice, indicates a severe or potentially life-threatening condition that requires urgent diagnosis and treatment. A fasting lipid profile should be obtained to complete the cardiovascular risk assessment, and if necessary, treatment should be instituted according to guidelines from the National Cholesterol Education Program Expert Panel.[42] This panel also incorporated several nonlipid risk factors for cardiovascular disease into its recommendations for clinical care by defining criteria for a condition that has become known as the metabolic syndrome (also called syndrome X, the deadly quartet, and the insulin-resistance syndrome). The metabolic syndrome includes the most common abnormalities of lipid and glucose metabolism that accompany abdominal obesity [*see Table 2*]. Identifying these abnormalities in a patient allows the practitioner to better assign that patient's risk for diabetes and coronary artery disease.[43] Similar criteria are now recognized by the Centers for Disease Control and Prevention (CDC) as the dysmetabolic syndrome X and have been assigned a diagnosis code (277.7) in the International Classification of Diseases, 9th Revision (ICD-9). Screening laboratory tests for hepatosteotosis include a liver panel. In addition, all overweight patients should have documentation of normal thyroid function with a thyroid-stimulating hormone level.

Although obesity is associated with abnormal levels of a number of hormones and cytokines, including leptin, ghrelin, interleukins, and tumor necrosis factor, measurement of these variables should be limited to research protocols and are not currently recommended for general clinical practice.

Table 2 Criteria for Metabolic Syndrome[42,151]

Any Three of the Following:

Increased waist circumference
- Men: > 102 cm (40 in)
- Women: > 88 cm (35 in)

Fasting plasma glucose ≥ 100 mg/dl

Elevated blood pressure
- Systolic ≥ 130 mm Hg
- Diastolic ≥ 85 mm Hg

Serum triglyceride level ≥ 150 mg/dl

Decreased high-density lipoprotein (HDL) cholesterol level
- Men < 40 mg/dl
- Women < 50 mg/dl

Differential Diagnosis

It is important for clinicians to be alert for secondary medical causes of obesity but also to be aware that, in most cases, treatment of these coexisting diseases rarely leads to complete reversal of the obese state. As an example, hypothyroidism is relatively common in the general population and may be present in an obese patient, but the weight loss that might be expected with thyroid hormone replacement is limited and variable.

Hypercortisolemia of Cushing syndrome is a rare cause of unwanted weight gain, but clinicians should have a low threshold for screening for this disease when patients experience large amounts of weight gain in a short period, especially when the weight gain is accompanied by hypertension, diabetes, or muscle weakness. Deficiencies of growth hormone or gonadal steroids are also associated with modest increases in body adiposity. Growth hormone deficiency can lead to reduced muscle mass and increased fat mass, which is improved with hormone replacement therapy.[44] Similar changes in body composition have been described in hypogonadal men[45] and in postmenopausal women.[18] Unfortunately, obesity is often accompanied by low levels of IGF-1 and, in men, low testosterone levels because of low sex-hormone–binding globulin levels. To distinguish obesity-associated low testosterone from a true deficiency state, free testosterone levels can be measured. In addition, weight loss will increase both IGF-1 and total testosterone levels in obese patients but not in patients with true deficiencies.

A number of medications can lead to unwanted weight gain and obesity; if possible, such patients should be switched to alternative agents [*see Table 3*]. Drug-related weight gain occurs most commonly during long-term glucocorticoid treatment of inflammatory conditions (e.g., asthma and inflammatory arthritis), with immunosuppression after transplantation, and with cancer chemotherapy. When possible, reducing or discontinuing a glucocorticoid in favor of an alternative medication can reverse this weight gain. Patients with type 1 or type 2 diabetes often gain weight after starting therapy; this weight gain is proportional to the degree of improved glycemic control and results from a reduction in glucosuria and improvement in metabolic efficiency.[46] Long-term studies have shown that intensive insulin treatment of type 1 diabetes can result in excessive weight gain and obesity in up to 25% of patients; in type 2 diabetes, intensive glycemic control with insulin, a sulfonylurea, or one of the thiazolidinediones may also result in greater weight gain than that predicted by improved glycemic control alone.[47] Therapy with metformin plus nighttime long-acting insulin may reduce or prevent this extra weight gain,[48] and newer diabetes medications, such as pramlintide and exenatide, can improve glycemic control in both type 1 and type 2 diabetes with a modest weight loss.[47] Neuropsychotropic drugs, particularly newer antipsychotic and antiseizure medications, have been associated with weight gain (sometimes massive), obesity, and diabetes.[49,50]

Table 3 Medications Commonly Associated with Weight Gain and Obesity, with Possible Alternative Agents[47,49,152]

Medication Class	*Agents*	*Alternatives*
Steroids	Glucocorticoids	Asthma: inhalers Cancer chemotherapy: non–glucocorticoid-based regimens Rheumatoid arthritis: methotrexate and remitting agents
Antidiabetic drugs	Insulin Sulfonylureas Thiazolidinediones	Metformin, acarbose, pramlintide, exenatide
Antiepileptic drugs	Gabapentin Valproic acid	Lamotrigine Topiramate Zonisamide
Antipsychotic agents	Clozapine Olanzapine Quetiapine Risperidone Sertindole	Aripiprazole Haloperidol Ziprasidone
Antidepressants	Tricyclic antidepressants Monoamine oxidase inhibitors Mirtazapine	Bupropion Nefazodone Selective serotonin reuptake inhibitors Venlafaxine

Treatment

As a first step in the management of obesity, appropriate follow-up testing and treatment should be provided for any secondary causes of obesity and comorbid conditions identified during screening. Then, the approach to the treatment of obesity is similar to that of other chronic conditions, such as hypertension, hypercholesterolemia, and diabetes. Intervention starts with lifestyle measures for 3 to 6 months. For obesity, these lifestyle interventions include improved diet and increased activity. For patients whose weight does not change with lifestyle intervention alone or whose weight loss is insufficient to lower their long-term health risk, consideration is then given to pharmacologic or surgical management. An NIH expert panel has suggested that patients whose BMI is 30 or more or who have a BMI of 27 or more plus obesity-related risk factors (i.e., diabetes, hypertension, or hyperlipidemia) could be considered for pharmacologic therapy.[3] Patients with a BMI of 40 or more or a BMI of 35 or more plus obesity-related risk factors could be considered for surgical therapy.

The weight-loss goal for the treatment of obesity is sustained weight loss of 5% or more of initial body weight. Although this goal does not result in attainment of a normal body weight (BMI of 19 to 25) in the majority of patients, it still represents a weight loss that can be achieved with available intervention modalities and that has been associated with lower morbidity, including reductions in risk for diabetes and heart disease.[51,52]

For some patients who are experiencing a period of weight gain, weight stability may be their primary goal. This is especially common in patients who have just completed a low-calorie weight-loss program and are struggling to remain below their initial weight.

NONMEDICAL (LIFESTYLE) THERAPY

Diet Modification

Caloric restriction Hypocaloric diets have been a mainstay recommendation by the medical community for obese patients. These diets range from a moderate reduction in daily intake (200 to 500 fewer calories a day) to more stringent, very low calorie diets (600 to 800 total calories a day), which require careful follow-up by a nutritionist and a physician to avoid life-threatening electrolyte disorders and symptomatic cholelithiasis. Although it is possible to achieve short-term weight loss with these strategies, long-term weight loss is poor even when behavior-modification weight-maintenance programs are continued. Analyses of published data on long-term weight-loss maintenance showed that approximately 50% of the initial lost body weight is regained within the first 1 to 2 years, and 95% or more is regained by 5 years after the completion of the calorie-restriction phase.[53,54] This restoration of lost body weight after a period of calorie-restriction–induced weight loss can be explained by the reduction of fat-dependent feedback signals to the brain, such as leptin, which then activate counterregulatory systems to restore body weight to baseline [*see* Pathophysiology and Pathogenesis, *above*].

The long-term failure to maintain weight loss after caloric restriction also indicates that the central set point for body weight is not reset at a lower body weight with the passage of time. Another important implication of these data is that obesity treatments lacking mechanisms that interfere with this counterregulatory system will likely fail to allow long-term weight-loss maintenance. Some existing therapies do result in limited weight loss without activation of appetite (see below) and can be combined with caloric restriction for improved long-term weight-loss maintenance; these include a low-fat diet,[55,56] exercise,[56] and pharmacologic treatments.[57,58]

Dietary-fat restriction An increase in dietary-fat intake leads to obesity in animal studies and has been associated with a higher prevalence of overweight and obesity in many human-population studies.[25] Prospectively randomizing overweight and obese persons to ad libitum feeding (eating until one feels full, then stopping) of a fat-restricted diet results in a spontaneous reduction in caloric intake and subsequent modest weight loss, compared with results in persons on a diet that contains a higher amount of fat typical of Western societies.[59] This calorie reduction occurs despite a concomitant fall in leptin levels, in contrast to the increase in appetite that follows a fall in leptin levels with weight loss from caloric restriction.[59-61] By implication, an increase in dietary fat results in a state of central leptin resistance, requiring a higher level of body fat and leptin levels to attain a new body-weight equilibrium,[62] whereas dietary-fat restriction leads to partial improvement in leptin signaling, resulting in spontaneous reduction in appetite and body weight.

The average amount of weight loss attributable to a low-fat diet in these studies, however, is only on the order of 3 to 4 kg (6.6 to 8.8 lb).[59] In addition, the weight-loss responses of persons to a low-fat diet can vary tremendously, with some individuals losing 13 kg (28.7 lb) or more and others losing no weight or even gaining weight.[63] This variable response to a lifestyle intervention, such as a change in a specific diet component, is common in chronic diseases whose expression results from an interaction between a genetic predisposition and environmental influence. In patients with hypertension, for example, blood pressure reductions in response to restriction of dietary salt are also heterogeneous.

An apparent paradox has been reported in that the average percent fat content of the American diet is dropping, yet the weight of the American population keeps increasing. Although it is true that the average percentage of total calories from fat in the American diet has declined over the past several decades (from 36% to 34%, according to the most recent NHANES data),[64] this did not occur because Americans have been eating less fat (daily dietary fat intake was 81.9 g in 1972 and increased to 85.5 g in 1990) but, rather, because total calories increased, leading to a lower fat percentage.[65] In contrast, studies that documented weight loss with a lower fat intake did so by lowering the absolute amount of fat in the diet. It is worth noting that the levels of fat restriction leading to weight loss in these studies were not severe. Severe fat restriction (< 20% of total calories) may not be sustainable for many patients because of limited food options and palatability.

Dietary-carbohydrate changes Increasing dietary-carbohydrate intake while lowering total fat intake results in modest spontaneous reduction in caloric intake and weight loss in overweight and obese persons. In the studies that documented this effect, the additional carbohydrates were derived from fruits, vegetables, and grain products, and the resulting increase in dietary fiber also may have played a role in greater satiety and weight loss. In society (especially in young people), however, dietary carbohydrates have increasingly been consumed in the form of processed foods sweetened with sucrose or fructose. These simple carbohydrates (especially fructose) may potential-

ly have deleterious effects on insulin resistance, lipid levels, and body weight when consumed in large amounts.[29,66]

Paradoxically, severe carbohydrate restriction (< 30 g/day) may also lead to modest spontaneous weight loss without initial activation of appetite. Such severe carbohydrate restriction initially mobilizes glycogen stores in the liver and induces ketogenesis, and the resulting diuresis accounts for some of this weight loss.[67] At one time, the ability to draw meaningful conclusions about the longer-term safety and efficacy of low-carbohydrate diets was impeded by the paucity of controlled studies and the variability of carbohydrate restriction from study to study (from < 20 to ≥ 200 g/day).[68] Although methodological issues remain, randomized, controlled studies have now shown that during the first 6 months of diet treatment, persons placed on a low-carbohydrate diet lose weight more rapidly than those placed on a low-calorie, low-fat diet.[69-71] Subsequently, however, individuals on the low-carbohydrate diet either stop losing weight or regain weight, and by 1 year, weight loss is the same with the two diets.[70-72] The average 1-year weight loss ranged from 2.5 to 5.1 kg in these studies, and both diets had a high dropout rate, of approximately 30% to 40%.[70-72] Contrary to popular beliefs about the potential adverse effects of consuming diets high in fat or carbohydrates, lipid levels and glucose metabolism improved with both diets in proportion to weight loss.[69-73]

Dietary-protein changes Increasing dietary-protein intake has also been associated with weight loss. In one of the few prospective, randomized studies of an ad libitum high-protein diet (fat restricted), obese patients experienced significantly greater weight loss than obese control subjects who followed a regular diet or a low-fat, high-carbohydrate diet over a period of 6 months.[74] After 12 months, however, total weight loss on the high-protein diet was attenuated and no longer differed from that seen with the high-carbohydrate diet.[74]

Most nutrition societies recommend limiting protein intake to approximately 10% to 15 % of daily calories because of concerns regarding long-term health consequences of high intake of protein (especially animal protein). These concerns include the possible association of increased protein intake with intestinal cancers, bone disease, and renal disease. To date, prospective studies have shown that increasing dietary protein increases the glomerular filtration rate,[75] which may be harmful to patients with existing renal disease or diabetes, but the long-term effect of increased glomerular filtration rate in otherwise healthy persons is not known.

Dietary fiber Increased dietary fiber has been shown to improve body weight and cardiovascular risk factors .[76,77] Typical high-fiber foods include fruits, vegetables, oat and wheat bran, and legumes, which are also low in fat. Even after controlling for low-fat content, however, diets higher in fiber result in reduced intake and a weight loss of approximately 2 kg.[76] Although it is possible to increase fiber through the use of supplements such as psyllium or methylcellulose, the current intake of fiber in the United States of about 15 g/day could be increased to 25 to 30 g/day by avoiding calorically dense, refined-sugar foods and increasing consumption of fruits, vegetables, and whole-grain products.

Summary of dietary recommendations Overall, an initial recommendation to lower dietary-fat intake and increase dietary-fiber intake for weight loss is reasonable and supported by the scientific literature. Long-term studies of greater than 1 year have not been conducted to show that this weight loss is sustained. Nevertheless, animal models of obesity, population studies, and prospective studies of 1 year or less of low-fat, high-fiber diets versus high-fat diets have documented that a high-fat diet is detrimental to body weight and that restriction of dietary fat to 25% to 30% of calories and an increase of 10 to 15 g of fiber a day result in a significant, albeit limited, weight loss for the average patient.

It is important that clinicians discourage unrealistic expectations about weight loss from a low-fat, high-fiber diet so that patients do not become disillusioned with this therapy. Also, patients should be informed that low-fat, high-fiber diets have been shown to reduce numerous health risks, especially when instituted as part of an overall lifestyle change that includes exercise.[50,78] For this dietary advice to be effective, however, it is often necessary to refer patients to a nutritionist for evaluation and follow-up.

A low-fat diet can be achieved by substituting either carbohydrate or protein for fat; this allows tailoring of the diet to the individual patient. Some patients may respond better to a high-carbohydrate diet in terms of food preferences and weight loss, whereas others might have better responses to a high-protein diet, although all these diet variations have shown only moderate weight loss.[72] A low-carbohydrate diet cannot currently be recommended for clinical practice, because it has not been shown to be superior to other diets[70,72]; the long-term health outcomes of sustained ketosis are uncertain; and increased intake of saturated fat and *trans*-fatty acids may negate the benefit of weight loss by increasing serum cholesterol and triglyceride levels in some patients.[72,73] Many questions concerning the effects of a higher protein intake (up to 30% of total calories) on patients' renal function remain unanswered. For this reason, high-protein diets should be avoided in patients with existing renal disease and diabetes.

Exercise

Increasing energy expenditure through exercise has been another mainstay of obesity therapy. Without counterregulation (alteration in appetite or non–exercise-based energy expenditure), an increase in activity should lead to continued and sustained weight loss. Prospective intervention studies have shown, however, that the average amount of weight loss attributable to exercise alone (no caloric restriction in addition to the exercise) is small, ranging from 1 to 4 kg (2.2 to 8.8 lb).[79,80] Further weight loss presumably is limited by alterations in nonexercise energy expenditure that compensate for the increase in exercise-induced energy expenditure; increased appetite can be discounted as the source of counterregulation, because most studies show little change in energy intake.[81]

Current recommendations are to participate in 3 to 5 hours of moderate to vigorous activity per week. For many patients, especially those with limited mobility, simply increasing activity may be an initial goal. As with alterations in the macronutrient content of the diet, exercise leads to variable degrees of weight loss, ranging from little or none to a substantial amount. Even with moderate weight loss, which is typical, patients should be encouraged to continue with increased activity because of the numerous health benefits attributable to being fit.[82,83]

Combined Diet and Exercise

A number of studies have prospectively examined the effect of combined diet and exercise interventions on health outcomes.

Sustained weight loss with caloric restriction can be improved with a low-fat diet[84,85] or regular exercise.[79,86] In a national survey seeking to determine the characteristics of persons who have been able to sustain a weight loss of at least 13.6 kg (30 lb) for 1 year, responders reported that on average, they followed a diet consisting of approximately 24% fat[87] and expended an average of 2,827 kcal/wk in exercise (roughly the equivalent of walking 28 miles/wk).[88] This level of activity is nearly three times more than the 1,000 kcal/wk recommended by the American College of Sports Medicine for the minimum weekly exercise for the purposes of reducing body weight.[88] Prospective studies of dietary-fat restriction and exercise have also demonstrated reduced progression, or even reversal, of heart disease in patients with known cardiovascular disease[78] and up to a 58% reduction in the incidence of type 2 diabetes in overweight patients with impaired glucose tolerance.[50]

Because of the overall health benefits, the initial treatment of overweight and obese patients should include increased activity in addition to dietary-fat restriction and increased dietary fiber. If caloric restriction is recommended, the likelihood of long-term weight-loss maintenance may be improved by the addition of these interventions. Dietary and activity advice should be implemented using individualized, sustainable behavioral and lifestyle changes. Again, it is important for both clinicians and patients not to have unreasonable expectations for weight loss from diet and exercise. For some patients, the best that may be achieved through lifestyle improvements is prevention of further weight gain. Patients who fail to meet unreasonable goals may become frustrated and return to a less healthy lifestyle. Even when little or no weight loss ensues, these lifestyle changes offer many health benefits, including improved lipid levels, increased insulin sensitivity, and reduced risk for progression of cardiovascular disease and for onset of type 2 diabetes.[89] After institution of lifestyle measures, if a patient's body weight remains above the guideline cutoff points, medical and surgical treatment options remain.[90]

Prevention

Once patients have become overweight or obese, lifestyle interventions play an important role in reducing comorbidities, but they typically have modest effects on body weight (see above). These patients will then require lifelong medical or surgical management to achieve more meaningful weight loss (see below). This failure of the body-weight set point to remain at a lower level in the average patient greatly increases the importance of prevention of unwanted weight gain. Prevention requires targeted improvements in food choices for pregnant mothers and children and, more importantly, a reversal of the societal trend toward reduced activity levels.

The hurdles for implementing these simple recommendations are, however, considerable. Increasing intake of healthy foods includes not only overcoming personal and cultural preferences for higher-fat foods but also consideration of the economic costs of food at home and at schools (calorically dense foods are often cheaper than fruits and vegetables) and the impact of family and work demands on ability to eat at home or purchase prepared foods.[91] Increasing activity does not necessarily require going to a gym on a regular basis. Rather, it includes maintaining physical activity in routine daily events, such as physical education in schools, designing work spaces and buildings to promote walking and stair use, planning urban environments to promote more pedestrian activity (e.g., safer streets, more sidewalks, increased density of commercial and residential properties), and overcoming a cultural reliance on automobiles for routine travel.[92] Even with immediate institution of lifestyle improvements, though, it is likely that reversing the current decades-long trend in weight gain will take several generations.

PHARMACOLOGIC THERAPY

The Food and Drug Administration has approved several prescription medications for the treatment of obesity [*see Table 4*].[90] These medications fall into two categories: centrally acting drugs, which suppress appetite, and peripherally acting drugs, which reduce fat absorption. For example, phentermine and sibutramine act centrally, reducing appetite by promoting the release of norepinephrine from presynaptic terminals (phentermine) and inhibiting the uptake of both norepinephrine and serotonin (sibutramine) in central nuclei. Orlistat acts peripherally, inhibiting the action of lipases in the brush border of the intestine and thereby reducing lipid absorption.

Clinicians who treat obese patients with medical therapy should keep in mind four important principles. First, most studies have included an initial treatment phase in which patients are placed on a hypocaloric diet (usually, daily caloric intake is reduced by 500 to 1,000 kcal) at the time of randomization or just before drug treatment. Second, the weight loss with obesity agents varies considerably: some patients lose a dramatic amount of weight; others lose only a little weight; and still others lose some weight only to regain it despite continuation of the medication. In any case, average weight loss with currently approved medications does not usually exceed 10% of the baseline weight. Third, weight loss is greatest during the first 3 to 6 months, followed by a plateau at a new lower weight even with continuation of the therapy. Intermittent therapy (i.e., 3 months on therapy followed by 3 months off, 3 months on, etc.) does not increase weight loss[93,94] and has been associated with increased side effects (e.g., dry mouth).[93] Finally, every drug-treatment study that has included posttreatment follow-up has shown rapid weight regain toward baseline after discontinuance of the medication. The plateauing of weight after initial weight loss and the regaining of weight after medication discontinuance have been interpreted to indicate that the medication became ineffective with time or failed because weight loss was neither sufficient nor sustained after cessation of the treatment. However, pharmacologic therapy for obesity is no different from therapy for other common chronic diseases. For instance, patients with hypertension often have a variable response to a first-line agent. In some cases, blood pressure lowering may be insufficient and may require the addition of a second medication, and hypertension returns when the medication is discontinued. Once a person begins receiving medical therapy for obesity, continued efficacy is evidenced by sustained weight-loss maintenance. Treatment should therefore be continued indefinitely unless the weight is regained or significant side effects develop.

Agents Approved for Short-Term Use

Benzphetamine, phendimetrazine, diethylpropion, mazindol, and phentermine are approved by the FDA for the short-term treatment of obesity (weeks). All but phentermine are rarely, if ever, used in clinical practice today. Benzphetamine and phendimetrazine are both Drug Enforcement Administration (DEA) schedule III drugs, with higher abuse potential than the others. Diethylpropion and mazindol have indications and side effects identical to those of phentermine (all are DEA schedule IV), but

Table 4 Pharmacologic Agents Approved by the Food and Drug Administration for the Treatment of Obesity

Duration of Treatment	*Drug (Trade Name)*	*Dose*	*Average Weight Loss*	*Comments*
Short term	Phentermine	30 mg resin; 15 or 37.5 mg tablets	8.7 kg	—
Long term*	Orlistat (Xenical)	250 mg with each meal (two or three times a day)	7–13 kg	Patients must already be on a low-fat diet (< 30% of total calories from fat); supplemental multivitamins are needed to prevent reduction in fat-soluble vitamins
	Sibutramine (Meridia)	5 mg 10 mg 15 mg	3.7 kg 5.7 kg 7.0 kg	On average, blood pressure and pulse rise slightly with treatment; contraindicated in patients with untreated hypertension and coronary artery disease

*Data are for treatment duration of 6 mo to 1 yr when combined with a low-calorie diet.

they are less well studied and have not had the same acceptance in clinical practice as phentermine; consequently, diethylpropion and mazindol are not extensively discussed in this chapter.

Phentermine Phentermine inhibits appetite and causes an average weight loss of 8.7 kg (19.2 lb) (net weight loss of 5.1 kg [11.2 lb] when compared with placebo).[95] The agent is available as a 30 mg resin and in 15 mg and 37.5 mg tablets. Doses in excess of 37.5 mg are not recommended because of unacceptable side effects.

Little information is available about subsequent improvement in health outcomes with phentermine treatment alone, because studies of this drug have historically been short term, and the drug was frequently used in combination with fenfluramine. A group of postmenopausal women treated with phentermine and a low-calorie diet experienced a 14% weight loss, along with reduction in low-density lipoprotein (LDL) cholesterol and triglyceride levels and an increase in high-density lipoprotein (HDL) cholesterol levels over 9 months.[96] In patients with diabetes, despite a net loss of 3.8 kg (8.4 lb) compared with placebo over 6 months, use of phentermine produced no improvement in glycemic control or glycosuria and no significant reduction in hypoglycemic drug use.[97]

As a result of central nervous system activation by phentermine, patients may experience anxiousness, insomnia, palpitations, and dry mouth. In case reports, phentermine treatment has been associated with vasospasm, psychosis, and ischemic events,[98-100] although in a larger cohort study, phentermine was not associated with stroke.[101]

Agents Approved for Long-Term Use

Sibutramine The FDA approved sibutramine for the medical treatment of obesity in 1997. The recommended duration for therapy was 1 year, but treatment may be continued beyond that time if no significant side effects occur and sustained weight loss is documented.

The average weight loss on the highest currently approved dose, 15 mg, is 7.0 kg (15.4 lb) (net weight loss of 5.7 kg [12.6 lb] when compared with placebo).[102] In a 2-year study of sibutramine therapy, a weight loss of 10.2 kg (22.5 lb) (net weight loss of 5.5 kg [12.1 lb] when compared with placebo) persisted for up to 18 months, at which time a slight upward trend became evident.[57] With placebo, in contrast, patients began to regain weight immediately after completion of the hypocaloric phase, which was at 6 months. After 2 years, significantly greater proportions of patients taking sibutramine, compared with control subjects, had maintained 5% and 10% weight loss. These results may have limited applicability, however, because only persons who completed the initial 6-month weight-loss phase and lost at least 5% of their initial body weight by caloric restriction were subsequently randomized to placebo or continued sibutramine therapy. This represents a 23% dropout rate before randomization to drug therapy. In addition, approximately 30% of the patients treated with sibutramine and 50% of the patients given placebo withdrew from the study. A higher pretreatment body weight[103] and weight loss of at least 1.8 kg (4 lb) during the first month of therapy[102] have been shown to predict continued weight-loss response with longer treatment.

Sibutramine treatment is also associated with improvements in lipid levels, including lower triglyceride levels and higher HDL levels.[57,94,104] Sibutramine has been safely used in patients with type 2 diabetes, with improvements in lipid levels being similar to those reported in nondiabetic patients.[105] Glycemic control also improves in diabetic patients on sibutramine therapy, with the greatest reductions in hemoglobin A_{1c} occurring in those patients who lose more than 10% of their initial body weight.[105] Common side effects of sibutramine treatment include dry mouth, constipation, insomnia, palpitations, and headache. Sibutramine consistently raises average blood pressures slightly and pulse rate more so, even with weight loss.[57,102] In populations not selected for hypertension, studies have documented increases in average diastolic pressure of 0 to 3.4 mm Hg, increases in systolic blood pressure of 0 to 2.7 mm Hg, and increases of average pulse rate of 4.1 to 6 beats/min.[57,94,102] Similar results with sibutramine have been shown for hypertensive patients receiving a variety of antihypertensive medications.[106,107] Therefore, sibutramine treatment can be used if blood pressure is controlled, but blood pressure and pulse rate should be monitored routinely.[108] However, because of the potential for added cardiovascular demand, sibutramine should not be used in patients with a diagnosis of cardiovascular disease, heart failure, arrhythmia, or stroke. To date, sibutramine has not been associated with either valvular disease or pulmonary hypertension.

Orlistat Orlistat inhibits lipases in the gastrointestinal lumen, thereby antagonizing triglyceride hydrolysis and reducing fat absorption by roughly 30%. Because orlistat is not absorbed to any significant extent, its primary mechanism of action is thought to

be through providing what is in effect a low-fat diet, thereby promoting lower caloric intake and weight loss [*see* Dietary-Fat Restriction, *above*], as well as improved weight-loss maintenance, when combined with a low-calorie diet.[109] As with sibutramine, orlistat treatment should be continued as long as the patient maintains weight loss and avoids significant side effects.

To minimize side effects related to fat malabsorption, candidates for orlistat treatment are first placed on a diet containing only 30% of calories from fat. When combined with a calorie-restricted diet, orlistat treatment (120 mg with each meal) results in an average weight loss of 7.2 to 13 kg (16 to 28.7 lb) (net weight loss of 1.3 to 5.6 kg [2.9 to 12.3 lb] when compared with placebo). After 1 year, patients taking orlistat maintain greater weight loss than those taking placebo [*see Table 4*].[58,110,111] During continued follow-up for another year, regaining of weight is seen but remains less than that in patients given placebo.[57,110,111]

Patients who lose weight with orlistat also experience a significant reduction in levels of total and LDL cholesterol (approximately 4% to 8%, which is significantly lower than reductions in patients given placebo).[110,111] Levels of triglycerides and HDL cholesterol are either reduced or left unchanged; these results are not different from those in the placebo group. Blood pressure and insulin levels also decrease with weight loss in patients on orlistat therapy.[58,110,111] Similar improvements in lipid levels, along with improvements in glycemic control, have been documented in obese persons with type 2 diabetes who lose weight with orlistat.[112]

Gastrointestinal side effects may occur in up to 80% of patients when they begin therapy with orlistat (such side effects are also seen in 50% to 60% of patients given placebo), but this incidence diminishes with time. Symptoms include abdominal discomfort, flatus, fecal urgency, oily spotting, and fecal incontinence. When administered to patients who adhere to a low-fat diet, orlistat is generally well tolerated. The fat malabsorption that accompanies orlistat treatment can also lead to reductions in fat-soluble vitamins, but in prospective studies, the average levels for vitamins A, D, E, and β-carotene remained in the normal range.[58,110,113] In one study, only 2.4% of orlistat-treated patients had documented below-normal levels of β-carotene; only 3.1%, below-normal levels of vitamin D; and only 1.6%, below-normal levels of vitamin E.[110] Nevertheless, orlistat should not be given to patients with existing malabsorptive states, and it is recommended that patients take a daily multivitamin supplement during therapy.

Non–FDA-Approved Medical Therapy for Obesity

Observational data from studies of patients with depression suggested that treatment with selective serotonin reuptake inhibitors (SSRIs) may result in weight loss, but this effect is slight and short-lived. Moreover, in prospective studies, SSRI treatment is sometimes associated with mild weight gain.[49] Bupropion is currently approved as an antidepressant and for smoking cessation. The mechanism of action of bupropion is not precisely known, but it includes weak inhibition of norepinephrine and dopamine reuptake. Combining bupropion (300 to 400 mg/day) with moderate daily caloric restriction has been shown to result in greater weight loss than placebo (net weight loss of 3 to 4 kg [6.6 to 8.8 lb], on average) after 6 months in obese patients with no symptoms or mild symptoms of depression.[114-116] Bupropion is contraindicated in patients with seizures, anorexia nervosa, and bulimia.

Deficiencies of growth hormone and sex steroids have been associated with higher body-fat content and greater central obesity. Replacing growth hormone reduces fat mass, increases lean mass, and reduces central obesity. Similar improvements in body composition have been reported with testosterone replacement in hypogonadal men. Although obese persons have documented abnormalities in the hypothalamic-pituitary-gonadal and somatotropic axes, treatment of overweight or centrally obese persons with growth hormone, sex steroids, or both has not been shown to produce clear improvements in body composition, fat distribution, or comorbid conditions.

Surgical Therapy

With improved safety from technical advances and demonstrated efficacy, bariatric (obesity) operations clearly have a role in the current management of severely obese patients.[90] The mechanisms whereby these operations result in sustained weight loss are poorly understood, but they likely include alterations in the gut-derived hormonal and neural inputs to the central nervous system.[117] In general, these procedures can be classified into one of three types: restriction of food passage, malabsorption of nutrients, or a combination of the two [*see Table 5*]. As an example of a purely restrictive procedure, vertical-banded gastroplasty involves the formation of a small stoma for the passage of food. The rationale behind this procedure is to increase a sense of fullness and reduce food intake. Procedures resulting in malabsorption typically involve bypassing sections of small or large intestine. Early procedures in which large sections, or even the entire length, of the small intestine were bypassed often resulted in severe malabsorption and sometimes hepatic failure and death. Subsequent modifications have led to bypass of shorter sections and lower morbidity. Patients must understand that bypass surgery is anatomically irreversible in most cases and has a potentially high postoperative complication rate.

Several large-scale studies have demonstrated the efficacy of various bariatric procedures in severely obese patients. One study of gastric bypass reported sustained loss of approximately 50% of baseline body weight (at 5- to 10-year follow-up).[118] In the ongoing Swedish Obesity Subjects (SOS) Intervention Study of 1,157 severely obese persons (average BMI of 42 kg/m^2), weight loss in the first postoperative year ranged from 21% after gastric banding to 38% after gastric bypass.[119] Some weight regain was found on 10-year follow-up, but the gastric banding patients remained 13% below and the gastric bypass patients 25% below their initial body weight.[119]

A more recent technique known as laparoscopic gastric banding, in which a restrictive band is placed around the upper stomach, has also shown efficacy in sustained weight loss in studies up to 4 years. The weight loss with laparoscopic gastric banding is generally felt to be similar to that with vertical-banded gastroplasty but not as great as that with gastric bypass.[120]

An important benefit of bariatric surgery is that the large weight losses bring improvements in the comorbid conditions that accompany obesity,[121,122] including lowering of lipid levels and blood pressure and reduced rates of diabetes, progression to diabetes, and sleep apnea.[119,122] Moreover, cohort and population-based studies have demonstrated that weight loss after bariatric surgery is associated with both reduced mortality and reduced use of health care resources.[123]

Complication rates for bariatric surgery will vary by surgeon, site, and technique. The most common operative-related complications, occurring in up to 30% of patients, are wound infections, atelectasis or pneumonia, and hernia.[118,124-127] Serious but rare complications include anastomotic stenosis or leakage, throm-

Table 5 Most Commonly Used Bariatric Procedures[118,153,154]

Type of Surgery	*Procedure*	*Average Weight Loss*	*Medical Complications*	*Management*
Restrictive	Vertical-banded gastroplasty	~17% (5 yr)	Nausea, vomiting Gastric distention	Reversal of procedure if complications unacceptable
	Laparoscopic adjustable gastric banding	17%–21% (3 yr)	Nausea, vomiting Slippage Erosion and leakage	Adjustment of band to minimize nausea; reversal of procedure if necessary
Malabsorptive	Biliopancreatic diversion	~27% (5 yr)	Diarrhea, fatty stools Protein-calorie malnutrition Anemia (low iron) Deficiency of vitamin D (and other fat-soluble vitamins) Hypocalcemia Hyperparathyroidism Metabolic bone disease Vitamin B_{12} deficiency, Wernicke encephalopathy (rare)	Monitoring of levels and replacement of nutritional deficiencies; patients often require high oral or parenteral doses
Combination	Gastric bypass	~27% (5 yr)	Dumping syndrome Reductions in levels of iron, vitamin B_{12}, folate, calcium, vitamin D Nesidioblastosis (rare)	Small, frequent meals; monitoring of nutrient levels and oral supplementation with iron tablets (325 mg), vitamin B_{12} (500 μg), folate (1 mg), calcium (500–1,000 mg), and vitamin D

boembolism, and bowel obstruction. Perioperative death rates have ranged from 0.2% to 1.3%.[118,125-127] The laparoscopic techniques offer greater safety than open procedures without sacrificing efficacy.[128,129] Nesidioblastosis has been described in a small series of obese patients after gastric bypass operations.[130] These patients experienced serious or life-threatening postprandial hypoglycemic episodes. A number of questions remain, however, regarding the frequency of this finding, the validity of the link between the gastric bypass and beta cell hypertrophy, and how to best do a workup of patients who may have this condition. Nevertheless, clinicians should include nesidioblastosis in the differential diagnosis of gastric bypass patients who complain of symptoms that have heretofore been ascribed to the so-called dumping syndrome (i.e., rapid transit of food into the small intestine associated with rapid increases in blood volume to the gut and insulin secretion): weakness, dizziness, palpitations, and near-syncope.

Because of the potential for rapid weight loss after bariatric surgery, follow-up by a nutritionist and physician is important to ensure that patients preserve lean mass and maintain hydration and to monitor for symptomatic cholelithiasis. Lifelong alteration of dietary habits may be required to ensure the continued success of the surgery, and dietary supplements may be needed to prevent nutritional deficiencies [*see Table 5*].

Surgical therapy for obesity results in significant, sustained weight loss in a majority of patients with severe obesity.[121,122] With regard to total weight loss and improved disease outcomes, the gastric bypass procedure has the greatest support from published studies,[121,122] whereas the gastric banding procedure has the lowest published rates of morbidity (11%) and mortality (0.05%).[120] With the considerable improvements in comorbidity and quality of life that result from this weight loss, bariatric surgery is not only efficacious but also cost-effective in the management of severe obesity.[122,123,131]

Complications of Obesity

Persons who are overweight or obese and have central adiposity are at increased risk for hyperlipidemia, hypertension, and cardiovascular disease mortality.[6,132-135] In addition, obesity and central adiposity are both strong risk factors for the development of type 2 diabetes.[11,136] Other diseases with a higher incidence in obese persons include gallstones, high uric acid levels and gout, hepatic steatosis, osteoarthritis, obesity hypoventilation, atrial fibrillation, nephrolithiasis, and certain cancers.[137,138] Sleep apnea is likely underdiagnosed in overweight and obese patients[139] and should be strongly considered in patients with complaints of fatigue, daytime somnolence, snoring, restless sleep, and morning headaches. People who are overweight and obese also carry a significant psychosocial burden. Obesity is accompanied by lower self-esteem in children[140] and prospectively predicts depression in adults.[141] Young adults who were overweight as adolescents are less likely to marry, will complete less schooling, and have lower incomes than nonoverweight peers.[142] The economic costs of the treatment of obesity and its comorbidities are high: estimates from 1995 United States data put the costs at $99.2 billion, and of this, $52 billion, or 5.7% of the United States national health expenditure, went to pay direct medical costs.[143]

Prognosis

Mortality increases when men and women become overweight or obese, in large part because of comorbid conditions (see above). This impact of obesity on mortality is greatest in younger age groups and diminishes with age[144,145] and may be significantly influenced by the patient's fitness status. Prospective observational studies suggest that mortality and risk associated with coronary artery disease are highest in the least fit and lowest in the most fit, independent of body weight.[82,83] Evidence of lower mortality after specific lifestyle, medical, or surgical in-

terventions for obesity is currently lacking. Instead, decreased severity of a number of diseases and cardiovascular risk factors associated with increased mortality have been demonstrated after even modest weight loss.[89] Insulin resistance, hyperlipidemia, hypertension, sleep apnea, diabetes, and cardiovascular disease have all been shown to improve with weight loss that follows intensive lifestyle interventions (e.g., low-fat, high-fiber diet and exercise) and medical therapy.[51,52,58,78,89,104,105] With the greater weight loss that accompanies bariatric surgery, resolution of diseases such as diabetes and sleep apnea[119,123] and increased survival[124] have been reported. Optimal therapy for obesity, however, remains elusive. Simple caloric restriction alone results in short-term weight loss (months to years), but without additional interventions such as a low-fat diet, exercise, pharmacologic therapy, or a combination of these, 95% or more of the weight initially lost will be regained within 5 years.[54] Lifestyle interventions alone, including a low-fat, high fiber diet and increased activity, have only small effects on body weight but result in significant improvements in obesity-related morbidity.[51] Current pharmacologic therapy is effective in achieving weight loss of up to 10% of initial body weight for at least up to 2 years.[57,110] Bariatric surgery has been shown to increase weight loss to up to 50% of initial weight for at least 10 years,[118,119] but it carries the risks of surgery and, with most operations, irreversible anatomic modification. Nonetheless, for patients with severe obesity who are at high risk for morbidity and mortality, bariatric surgery can offer hope for improved survival while more effective medical therapies are being developed.

The author has served as a consultant for Amylin Pharmaceuticals, Inc.

The drug bupropion, which is discussed in this chapter, has not been approved by the FDA for use in obesity.

References

1. Jackson AS, Stanforth PR, Gagnon J, et al: The effect of sex, age and race on estimating percentage body fat from body mass index: the Heritage Family Study. Int J Obes Relat Metab Disord 26:789, 2002

2. Najjar MF, Rowland M: Anthropometric reference data and prevalence of overweight, United States, 1976–80. Vital Health Stat 11(238):1, 1987

3. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults—The Evidence Report. National Institutes of Health. Obes Res 6(suppl 2):51S, 1998

4. The Asia-Pacific Perspective: Redefining Obesity and Its Treatment. World Health Organization, Western Pacific Region; International Association for the Study of Obesity; International Obesity Task Force, 2000

5. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. WHO Expert Consultation. Lancet 363:157, 2004

6. Janssen I, Katzmarzyk PT, Ross R: Waist circumference and not body mass index explains obesity-related health risk. Am J Clin Nutr 79:379, 2004

7. Fujimoto WY, Bergstrom RW, Boyko EJ, et al: Susceptibility to development of central adiposity among populations. Obes Res 3(suppl 2):179S, 1995

8. Health, United States, 2004. National Center for Health Statistics, Hyattsville, Maryland, 2004, p 241
http://www.cdc.gov/nchs/data/hus/hus04.pdf

9. Hedley AA, Ogden CL, Johnson CL, et al: Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. JAMA 291:2847, 2004

10. Knowler WC, Pettitt DJ, Saad MF, et al: Obesity in the Pima Indians: its magnitude and relationship with diabetes. Am J Clin Nutr 53(6 suppl):1543S, 1991

11. Lethbridge-Çejku M, Vickerie J: Summary health statistics for U.S. adults: National Health Interview Survey, 2003. National Center for Health Statistics. Vital Health Stat 10(225):1, 2005

12. Paeratakul S, Lovejoy JC, Ryan DH, et al: The relation of gender, race and socioeconomic status to obesity and obesity comorbidities in a sample of US adults. Int J Obes Relat Metab Disord 26:1205, 2002

13. York DA, Rossner S, Caterson I, et al: American Heart Association. Prevention Conference VII: Obesity, a worldwide epidemic related to heart disease and stroke: Group I: worldwide demographics of obesity. Circulation 110:e463, 2004

14. Popkin BM, Horton S, Kim S, et al: Trends in diet, nutritional status, and diet-related noncommunicable diseases in China and India: the economic costs of the nutrition transition. Nutr Rev 59:379, 2001

15. Mott JW, Wang J, Thornton JC, et al: Relation between body fat and age in 4 ethnic groups. Am J Clin Nutr 69:1007, 1999

16. Gallagher D, Ruts E, Visser M, et al: Weight stability masks sarcopenia in elderly men and women. Am J Physiol Endocrinol Metab 279:E366, 2000

17. Hughes VA, Frontera WR, Roubenoff R, et al: Longitudinal changes in body composition in older men and women: role of body weight change and physical activity. Am J Clin Nutr 76:473, 2002

18. Trémollieres FA, Pouilles JM, Ribot CA: Relative influence of age and menopause on total and regional body composition changes in postmenopausal women. Am J Obstet Gynecol 175:1594, 1996

19. Lemieux S, Prud'homme D, Bouchard C, et al: Sex differences in the relation of visceral adipose tissue accumulation to total body fatness. Am J Clin Nutr 58:463, 1993

20. Matsuzawa Y, Shimomura I, Nakamura T, et al: Pathophysiology and pathogenesis of visceral fat obesity. Obes Res 3(suppl 2):187S, 1995

21. Shuman WP, Morris LL, Leonetti DL, et al: Abnormal body fat distribution detected by computed tomography in diabetic men. Invest Radiol 21:483, 1986

22. Maes HH, Neale MC, Eaves LJ: Genetic and environmental factors in relative body weight and human adiposity. Behav Genet 27:325, 1997

23. Perusse L, Rankinen T, Zuberi A, et al: The human obesity gene map: the 2004 update. Obes Res 13:381, 2005

24. Bray GA, Popkin BM: Dietary fat intake does affect obesity! Am J Clin Nutr 68:1157, 1998

25. Coakley EH, Rimm EB, Colditz G, et al: Predictors of weight change in men: results from the Health Professionals Follow-up Study. Int J Obes Relat Metab Disord 22:89, 1998

26. Ludwig DS, Pereira MA, Kroenke CH, et al: Dietary fiber, weight gain, and cardiovascular disease risk factors in young adults. JAMA 282:1539, 1999

27. Gordon-Larsen P, Adair LS, Popkin BM: Ethnic differences in physical activity and inactivity patterns and overweight status. Obes Res 10:141, 2002

28. Drewnowski A, Specter SE: Poverty and obesity: the role of energy density and energy costs. Am J Clin Nutr 79:6, 2004

29. Ludwig DS, Peterson KE, Gortmaker SL: Relation between consumption of sugar-sweetened drinks and childhood obesity: a prospective, observational analysis. Lancet 357:505, 2001

30. Pereira MA, Jacobs DR Jr, Van Horn L, et al: Dairy consumption, obesity, and the insulin resistance syndrome in young adults: the CARDIA Study. JAMA 287:2081, 2002

31. Stookey JD, Adair L, Stevens J, et al: Patterns of long-term change in body composition are associated with diet, activity, income and urban residence among older adults in China. J Nutr 131:2433S, 2001

32. Flegal KM, Troiano RP, Pamuk ER, et al: The influence of smoking cessation on the prevalence of overweight in the United States. N Engl J Med 333:1165, 1995

33. Pereira MA, Kartashov AI, Ebbeling CB, et al: Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. Lancet 365:36, 2005

34. Schwartz MW, Woods SC, Porte D Jr, et al: Central nervous system control of food intake. Nature 404:661, 2000

35. Zhang Y, Proenca R, Maffei M, et al: Positional cloning of the mouse obese gene and its human homologue. Nature 372:425, 1994

36. Lamberts SW, van den Beld AW, van der Lely AJ: The endocrinology of aging. Science 278:419, 1997

37. Munzer T, Harman SM, Hees P, et al: Effects of GH and/or sex steroid administration on abdominal subcutaneous and visceral fat in healthy aged women and men. J Clin Endocrinol Metab 86:3604, 2001

38. Blackman MR, Sorkin JD, Munzer T, et al: Growth hormone and sex steroid administration in healthy aged women and men: a randomized controlled trial. JAMA 288:2282, 2002

39. Van Cauter E, Leproult R, Kupfer DJ: Effects of gender and age on the levels and circadian rhythmicity of plasma cortisol. J Clin Endocrinol Metab 81:2468, 1996

40. Anawalt BD, Merriam GR: Neuroendocrine aging in men: andropause and somatopause. Endocrinol Metab Clin North Am 30:647, 2001

41. Purnell JQ, Brandon DD, Isabelle LM, et al: Association of 24-hour cortisol production rates, cortisol-binding globulin, and plasma-free cortisol levels with body composition, leptin levels, and aging in adult men and women. J Clin Endocrinol Metab 89:281, 2004

42. Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 285:2486, 2001

43. Eckel RH, Grundy SM, Zimmet PZ: The metabolic syndrome. Lancet 365:1415, 2005

44. Carroll PV, Christ ER, Sonksen PH: Growth hormone replacement in adults with growth hormone deficiency: assessment of current knowledge. Trends Endocrinol Metab 11:231, 2000

45. Wang C, Swedloff RS, Iranmanesh A, et al: Transdermal testosterone gel improves sexual function, mood, muscle strength, and body composition parameters in hypogonadal men. Testosterone Gel Study Group. J Clin Endocrinol Metab 85:2839, 2000

46. Makimattila S, Nikkila K, Yki-Jarvinen H: Causes of weight gain during insulin therapy with and without metformin in patients with type II diabetes mellitus. Diabetologia 42:406, 1999

47. Purnell JQ, Weyer C: Bodyweight effects of current and experimental drugs for diabetes mellitus. Treat Endocrinol 2:33, 2003

48. Yki-Jarvinen H: Combination therapies with insulin in type 2 diabetes. Diabetes Care 24:758, 2001

49. Vanina Y, Podolskaya A, Sedky K, et al: Body weight changes associated with psychopharmacology. Psychiatr Serv 53:842, 2002

50. Casey DE: Metabolic issues and cardiovascular disease in patients with psychiatric disorders. Am J Med 118(suppl 2):15S, 2005

51. Knowler WC, Barrett-Connor E, Fowler SE, et al: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346:393, 2002

52. Blackburn G: Effect of degree of weight loss on health benefits. Obes Res 3(suppl 2):211s, 1995

53. Tsai AG, Wadden TA: Systematic review: an evaluation of major commercial weight loss programs in the United States. Ann Intern Med 142:56, 2005

54. Safer DJ: Diet, behavior modification, and exercise: a review of obesity treatments from a long-term perspective. South Med J 84:1470, 1991

55. Toubro S, Astrup A: Randomised comparison of diets for maintaining obese subjects' weight after major weight loss: ad lib, low fat, high carbohydrate diet v fixed energy intake. BMJ 314:29, 1997

56. Wing RR, Hill JO: Successful weight loss maintenance. Annu Rev Nutr 21:323, 2001

57. James WP, Astrup A, Finer N, et al: Effect of sibutramine on weight maintenance after weight loss: a randomised trial. STORM Study Group. Sibutramine Trial of Obesity Reduction and Maintenance. Lancet 356:2119, 2000

58. Davidson MH, Hauptman J, DiGirolamo M, et al: Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat: a randomized controlled trial. JAMA 281:235, 1999

59. Astrup A, Grunwald GK, Melanson EL, et al: The role of low-fat diets in body weight control: a meta-analysis of ad libitum dietary intervention studies. Int J Obes Relat Metab Disord 24:1545, 2000

60. Weigle DS, Cummings DE, Newby PD, et al: Roles of leptin and ghrelin in the loss of body weight caused by a low fat, high carbohydrate diet. J Clin Endocrinol Metab 88:1577, 2003

61. Weigle DS, Breen PA, Matthys CC, et al: A high-protein diet induces sustained reductions in appetite, ad libitum caloric intake, and body weight despite compensatory changes in diurnal plasma leptin and ghrelin concentrations. Am J Clin Nutr 82:41, 2005

62. El-Haschimi K, Pierroz DD, Hileman SM, et al: Two defects contribute to hypothalamic leptin resistance in mice with diet-induced obesity. J Clin Invest 105:1827, 2000

63. Schaefer EJ, Lichtenstein AH, Lamon-Fava S, et al: Body weight and low-density lipoprotein cholesterol changes after consumption of a low-fat ad libitum diet. JAMA 274:1450, 1995

64. Daily dietary fat and total food-energy intakes—Third National Health and Nutrition Examination Survey, Phase 1, 1988–91. MMWR Morb Mortal Wkly Rep 43:116, 1994

65. Ernst ND, Sempos CT, Briefel RR, et al: Consistency between US dietary fat intake and serum total cholesterol concentrations: the National Health and Nutrition Examination Surveys. Am J Clin Nutr 66(4 suppl):965S, 1997

66. Elliott SS, Keim NL, Stern JS, et al: Fructose, weight gain, and the insulin resistance syndrome. Am J Clin Nutr 76:911, 2002

67. Freedman MR, King J, Kennedy E: Popular diets: a scientific review. Obes Res 9(suppl 1):1S, 2001

68. Bravata DM, Sanders L, Huang J, et al: Efficacy and safety of low-carbohydrate diets: a systematic review. JAMA 289:1837, 2003

69. Brehm BJ, Seeley RJ, Daniels SR, et al: A randomized trial comparing a very low carbohydrate diet and a calorie-restricted low fat diet on body weight and cardiovascular risk factors in healthy women. J Clin Endocrinol Metab 88:1617, 2003

70. Foster GD, Wyatt HR, Hill JO, et al: A randomized trial of a low-carbohydrate diet for obesity. N Engl J Med 348:2082, 2003

71. Stern L, Iqbal N, Seshadri P, et al: The effects of low-carbohydrate versus conventional weight loss diets in severely obese adults: one-year follow-up of a randomized trial. Ann Intern Med 140:778, 2004

72. Dansinger ML, Gleason JA, Griffith JL, et al: Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. JAMA 293:43, 2005

73. Yancy WS Jr, Olsen MK, Guyton JR, et al: A low-carbohydrate, ketogenic diet versus a low-fat diet to treat obesity and hyperlipidemia: a randomized, controlled trial. Ann Intern Med 140:769, 2004

74. Due A, Toubro S, Skov AR, et al: Effect of normal-fat diets, either medium or high in protein, on body weight in overweight subjects: a randomised 1-year trial. Int J Obes Relat Metab Disord 28:1283, 2004

75. Skov AR, Toubro S, Bulow J, et al: Changes in renal function during weight loss induced by high vs low-protein low-fat diets in overweight subjects. Int J Obes Relat Metab Disord 23:1170, 1999

76. Howarth NC, Saltzman E, Roberts SB: Dietary fiber and weight regulation. Nutr Rev 59:129, 2001

77. Davy BM, Melby CL: The effect of fiber-rich carbohydrates on features of syndrome X. J Am Diet Assoc 103:86, 2003

78. Ornish D, Scherwitz LW, Billings JH, et al: Intensive lifestyle changes for reversal of coronary heart disease. JAMA 280:2001, 1998

79. Wing RR: Physical activity in the treatment of the adulthood overweight and obesity: current evidence and research issues. Med Sci Sports Exerc 31(11 suppl):S547, 1999

80. Votruba SB, Horvitz MA, Schoeller DA: The role of exercise in the treatment of obesity. Nutrition 16:179, 2000

81. Blundell JE, King NA: Exercise, appetite control, and energy balance. Nutrition 16:519, 2000

82. Wei M, Kampert JB, Barlow CE, et al: Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. JAMA 282:1547, 1999

83. Farrell SW, Braun L, Barlow CE, et al: The relation of body mass index, cardiorespiratory fitness, and all-cause mortality in women. Obes Res 10:417, 2002

84. Pascale RW, Wing RR, Butler BA, et al: Effects of a behavioral weight loss program stressing calorie restriction versus calorie plus fat restriction in obese individuals with NIDDM or a family history of diabetes. Diabetes Care 18:1241, 1995

85. Toubro S, Astrup A: Randomised comparison of diets for maintaining obese subjects' weight after major weight loss: ad lib, low fat, high carbohydrate diet v fixed energy intake. BMJ 314:29, 1997

86. Jakicic JM, Winters C, Lang W, et al: Effects of intermittent exercise and use of home exercise equipment on adherence, weight loss, and fitness in overweight women: a randomized trial. JAMA 282:1554, 1999

87. Shick SM, Wing RR, Klem ML, et al: Persons successful at long-term weight loss and maintenance continue to consume a low-energy, low-fat diet. J Am Diet Assoc 98:408, 1998

88. Klem ML, Wing RR, McGuire MT, et al: A descriptive study of individuals successful at long-term maintenance of substantial weight loss. Am J Clin Nutr 66:239, 1997

89. Pi-Sunyer FX: A review of long-term studies evaluating the efficacy of weight loss in ameliorating disorders associated with obesity. Clin Ther 18:1006, 1996

90. Snow V, Barry P, Fitterman N, et al: Pharmacologic and surgical management of obesity in primary care: a clinical practice guideline from the American College of Physicians. Ann Intern Med 142:525, 2005

91. Drewnowski A, Darmon NR: The economics of obesity: dietary energy density and energy cost. Am J Clin Nutr 82(1 suppl):265S, 2005

92. Booth KM, Pinkston MM, Poston WS: Obesity and the built environment. J Am Diet Assoc 105(5 suppl 1):S110, 2005

93. Weintraub M, Sundaresan PR, Schuster B, et al: Long-term weight control study. II (weeks 34 to 104): an open-label study of continuous fenfluramine plus phentermine versus targeted intermittent medication as adjuncts to behavior modification, caloric restriction, and exercise. Clin Pharmacol Ther 51:595, 1992

94. Wirth A, Krause J: Long-term weight loss with sibutramine: a randomized controlled trial. JAMA 286:1331, 2001

95. Bray GA: Use and abuse of appetite-suppressant drugs in the treatment of obesity. Ann Intern Med 119:707, 1993

96. Cordero-MacIntyre ZR, Lohman TG, Rosen J, et al: Weight loss is correlated with an improved lipoprotein profile in obese postmenopausal women. J Am Coll Nutr 19:275, 2000

97. Campbell CJ, Bhalla IP, Steel JM, et al: A controlled trial of phentermine in obese diabetic patients. Practitioner 218:851, 1977

98. Kim I, Whitsett TL: Acute vasospasm associated with anorexiant use. J Okla State Med Assoc 81:395, 1988

99 Devan GS: Phentermine and psychosis. Br J Psychiatry 156:442, 1990

100. Kokkinos J, Levine SR: Possible association of ischemic stroke with phentermine. Stroke 24:310, 1993

101. Derby LE, Myers MW, Jick H: Use of dexfenfluramine, fenfluramine and phentermine and the risk of stroke. Br J Clin Pharmacol 47:565, 1999

102. Bray GA, Blackburn GL, Ferguson JM, et al: Sibutramine produces dose-related weight loss. Obes Res 7:189, 1999

103. Hansen D, Astrup A, Toubro S, et al: Predictors of weight loss and maintenance during 2 years of treatment by sibutramine in obesity: results from the European multicentre STORM trial. Sibutramine Trial of Obesity Reduction and Maintenance. Int J Obes Relat Metab Disord 25:496, 2001

104. Dujovne CA, Zavoral JH, Rowe E, et al: Effects of sibutramine on body weight and serum lipids: a double-blind, randomized, placebo-controlled study in 322 overweight and obese patients with dyslipidemia. Am Heart J 142:489, 2001

105. Finer N, Bloom SR, Frost GS, et al: Sibutramine is effective for weight loss and diabetic control in obesity with type 2 diabetes: a randomised, double-blind, placebo-controlled study. Diabetes Obes Metab 2:105, 2000

106. Bray GA: Sibutramine and blood pressure: a therapeutic dilemma. J Hum Hypertens 16:1, 2002

107. Sramek JJ, Leibowitz MT, Weinstein SP, et al: Efficacy and safety of sibutramine for weight loss in obese patients with hypertension well controlled by beta-adrenergic blocking agents: a placebo-controlled, double-blind, randomised trial. J Hum Hypertens 16:13, 2002

108. Jordan J, Scholze K, Matiba B, et al: Influence of sibutramine on blood pressure: evidence from placebo-controlled trials. Int J Obesity Relat Metab Disord 29:509, 2005

1090. Hutton B, Fergusson D: Changes in body weight and serum lipid profile in obese patients treated with orlistat in addition to a hypocaloric diet: a systematic review of randomized clinical trials. Am J Clin Nutr 80:1461, 2004

110. Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. Lancet 352:167, 1998

111. Weight loss, weight maintenance, and improved cardiovascular risk factors after 2 years treatment with orlistat for obesity. European Orlistat Obesity Study Group. Obes Res 8:49, 2000

112. Kelley DE, Jneidi M: Orlistat in the treatment of type 2 diabetes mellitus. Expert Opin Pharmacother 3:599, 2002

113. Efficacy and tolerability of orlistat in the treatment of obesity: a 6-month dose-ranging study. Orlistat Dose-Ranging Study Group. Eur J Clin Pharmacol 54:125, 1998

114. Gadde KM, Parker CB, Maner LG, et al: Bupropion for weight loss: an investigation of efficacy and tolerability in overweight and obese women. Obes Res 9:544, 2001

115. Jain AK, Kaplan RA, Gadde KM, et al: Bupropion SR vs. placebo for weight loss in obese patients with depressive symptoms. Obes Res 10:1049, 2002

116. Anderson JW, Greenway FL, Fujioka K, et al: Bupropion SR enhances weight loss: a 48-week double-blind, placebo-controlled trial. Obes Res 10:633, 2002

117. Cummings DE, Overduin J, Foster-Schubert KE: Gastric bypass for obesity: mechanisms of weight loss and diabetes resolution. J Clin Endocrinol Metab 89:2608, 2004

118. Pories WJ, MacDonald KG Jr, Morgan EJ, et al: Surgical treatment of obesity and its effect on diabetes: 10-y follow-up. Am J Clin Nutr 55(2 suppl):582S, 1992

119. Sjöström L, Lindroos AK, Peltonen M, et al: Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. N Engl J Med 351:2683, 2004

120. Chapman AE, Kiroff G, Game P, et al: Laparoscopic adjustable gastric banding in the treatment of obesity: a systematic literature review. Surgery 135:326, 2004

121. Brolin RE: Bariatric surgery and long-term control of morbid obesity. JAMA 288:2793, 2002

122. Buchwald H, Avidor Y, Braunwald E, et al: Bariatric surgery: a systematic review and meta-analysis. JAMA 292:1724, 2004 (erratum in JAMA 293:1728, 2005)

123. Christou NV, Sampalis JS, Liberman M, et al: Surgery decreases long-term mortality, morbidity, and health care use in morbidly obese patients. Ann Surg 240:416, 2004

124. Flum DR, Dellinger EP: Impact of gastric bypass operation on survival: a population-based analysis. J Am Coll Surg 199:543, 2004

125. Wolf AM, Kortner B, Kuhlmann HW: Results of bariatric surgery. Int J Obes Relat Metab Disord 25(suppl 1):S113, 2001

126. Schauer PR, Ikramuddin S, Gourash W, et al: Outcomes after laparoscopic Roux-en-Y gastric bypass for morbid obesity. Ann Surg 232:515, 2000

127. Podnos YD, Jimenez JC, Wilson SE, et al: Complications after laparoscopic gastric bypass: a review of 3464 cases. Arch Surg 138:957, 2003

128. Lujan JA, Frutos MD, Hernandez Q, et al: Laparoscopic versus open gastric bypass in the treatment of morbid obesity: a randomized prospective study. Ann Surg 239:433, 2004

129. Nguyen NT, Goldman C, Rosenquist CJ, et al: Laparoscopic versus open gastric bypass: a randomized study of outcomes, quality of life, and costs. Ann Surg 234:279, 2001

130. Service GJ, Thompson GB, Service FJ, et al: Hyperinsulinemic hypoglycemia with nesidioblastosis after gastric-bypass surgery. N Engl J Med 353:249, 2005

131. Clegg A, Colquitt J, Sidhu M, et al: Clinical and cost effectiveness of surgery for morbid obesity: a systematic review and economic evaluation. Int J Obes Relat Metab Disord 27:1167, 2003

132. Rimm EB, Stampfer MJ, Giovannucci E, et al: Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men. Am J Epidemiol 141:1117, 1995

133. Manson JE, Willett WC, Stampfer MJ, et al: Body weight and mortality among women. N Engl J Med 333:677, 1995

134. Lakka HM, Laaksonen DE, Lakka TA, et al: The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. JAMA 288:2709, 2002

135. Jonsson S, Hedblad B, Engstrom G, et al: Influence of obesity on cardiovascular risk: twenty-three-year follow-up of 22,025 men from an urban Swedish population. Int J Obes Relat Metab Disord 26:1046, 2002

136. Mokdad AH, Ford ES, Bowman BA, et al: Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. JAMA 289:76, 2003

137. Bray GA: Medical consequences of obesity. J Clin Endocrinol Metab 89:2583, 2004

138. Must A, Spadano J, Coakley EH, et al: The disease burden associated with overweight and obesity. JAMA 282:1523, 1999

139. Peppard PE, Young T, Palta M, et al: Longitudinal study of moderate weight change and sleep-disordered breathing. JAMA 284:3015, 2000

140. Strauss RS: Childhood obesity and self-esteem. Pediatrics 105:e15, 2000

141. Roberts RE, Kaplan GA, Shema SJ, et al: Are the obese at greater risk for depression? Am J Epidemiol 152:163, 2000

142. Gortmaker SL, Must A, Perrin JM, et al: Social and economic consequences of overweight in adolescence and young adulthood. N Engl J Med 329:1008, 1993

143. Wolf AM, Colditz GA: Current estimates of the economic cost of obesity in the United States. Obes Res 6:97, 1998

144. Bender R, Jockel KH, Trautner C, et al: Effect of age on excess mortality in obesity. JAMA 281:1498, 1999

145. Fontaine KR, Redden DT, Wang C, et al: Years of life lost due to obesity. JAMA 289:187, 2003

146. Kolaczynski JW, Considine RV, Ohannesian J, et al: Responses of leptin to short-term fasting and refeeding in humans: a link with ketogenesis but not ketones themselves. Diabetes 45:1511, 1996

147. Keim NL, Stern JS, Havel PJ: Relation between circulating leptin concentrations and appetite during a prolonged, moderate energy deficit in women. Am J Clin Nutr 68:794, 1998

148. Leibel RL, Rosenbaum M, Hirsch J: Changes in energy expenditure resulting from altered body weight. N Engl J Med 332:621, 1995

149. Kolaczynski JW, Ohannesian JP, Considine RV, et al: Response of leptin to short-term and prolonged overfeeding in humans. J Clin Endocrinol Metab 81:4162, 1996

150. Sims EA, Danforth E Jr, Horton ES, et al: Endocrine and metabolic effects of experimental obesity in man. Recent Prog Horm Res 29:457, 1973

151. Diagnosis and classification of diabetes mellitus. American Diabetes Association. Diabetes Care 27(suppl 1):S5, 2004

152. Gadde KM, Franciscy DM, Wagner HR 2nd, et al: Zonisamide for weight loss in obese adults: a randomized controlled trial. JAMA 289:1820, 2003

153. Ren CJ, Horgan S, Ponce J: US experience with the LAP-BAND system. Am J Surg 184:46S, 2002

154. Anthone GJ, Lord RV, DeMeester TR, et al: The duodenal switch operation for the treatment of morbid obesity. Ann Surg 238:618, 2003

Acknowledgment

Figure 3 Seward Hung.

I ESOPHAGEAL DISORDERS

Stuart Jon Spechler, m.d.

Approach to the Patient with Dysphagia

The esophagus is susceptible to three types of diseases that cause dysphagia [*see Table 1*]: (1) mucosal (intrinsic) diseases that narrow the lumen of the esophagus through inflammation, fibrosis, or neoplasia; (2) mediastinal (extrinsic) diseases that encase and obstruct the esophagus by direct invasion or through lymph node enlargement; and (3) diseases affecting the esophageal muscle or its innervation (esophageal motility disorders) that disrupt peristalsis, interfere with sphincter relaxation, or both. The American Gastroenterological Association has endorsed algorithms summarizing the approach to the patient with dysphagia [*see Figures 1 through 4*].[1]

clinical evaluation

History

On the basis of a careful history alone, the astute clinician can determine the cause of dysphagia in approximately 80% of patients.[2] Eight questions form the key elements of the history.

Is the dysphagia for solid foods, liquids, or both? Mucosal and mediastinal diseases that involve the esophagus cause dysphagia by narrowing the lumen. Such mechanical narrowing usually does not impede the passage of liquids, and consequently, such a disease causes dysphagia for solid foods only. Diseases that disrupt peristalsis, however, may cause dysphagia for both solids and liquids. Of the esophageal motility disorders, achalasia is the one most likely to cause dysphagia for liquids. In achalasia, chronic contraction of the lower esophageal sphincter (LES) causes complete mechanical obstruction of the esophagus that persists until either the sphincter relaxes or the hydrostatic pressure of the retained material exceeds the pressure generated by the sphincter muscle. Even in the absence of peristalsis, gravity often can empty the esophagus of liquid effectively, provided that the LES is relaxed. Therefore, patients who have disordered peristalsis with an LES that is profoundly hypotensive often experience no dysphagia or experience dysphagia only for solid foods.

Where does the patient perceive that ingested material sticks? Patients with esophageal strictures often perceive that swallowed material sticks at a point that is either above or at the level of the stricture.[3] It is uncommon for patients to perceive that swallowed material is stuck below the obstructing lesion. Thus, the history that a swallowed bolus sticks above the suprasternal notch is of little value in localizing the obstruction, because this sensation could be caused by a lesion located anywhere from the pharynx to the gastroesophageal junction. If the patient localizes the obstruction to a point below the suprasternal notch, however, then it is highly likely that the dysphagia is caused by an esophageal disorder.

Are there symptoms of oropharyngeal dysfunction? Oropharyngeal dysfunction often results from diseases that affect the striated muscles of the oropharynx or their innervation. Examples include muscular dystrophies, dermatomyositis, myasthenia gravis, and cerebrovascular accidents.[4] Patients with these neuromuscular diseases may experience difficulty in initiating a swallow, and swallowing may be accompanied by nasopharyngeal regurgitation, pulmonary aspiration, and a sensation that residual material remains in the pharynx. If any of these symptoms are prominent, evaluation for oropharyngeal dysfunction (e.g., with videofluoroscopy) is appropriate.

Is the dysphagia intermittent or progressive? Patients who have a lower esophageal mucosal (Schatzki) ring (see below) typically complain of discrete episodes of dysphagia for solid foods that are intermittent and nonprogressive. The episodes often occur during meals in restaurants (hence the term steakhouse syndrome) or at social functions. Dysphagia episodes may be separated by a period of weeks, months, or years, and the patient typically experiences no swallowing difficulty between episodes. In contrast, esophageal strictures usually cause dysphagia that is progressive in frequency and severity. With benign strictures, the progression is typically slow and insidious (over a period of months to years), and weight loss is minimal. Malignant esophageal strictures usually cause dysphagia that progresses rapidly (over a period of weeks to months), and weight loss may be profound.

Is there a history of chronic heartburn? Heartburn is the cardinal symptom of gastroesophageal reflux disease (GERD). Thus, a history of chronic heartburn supports the possibility that dysphagia may result from a peptic esophageal stricture. However, up to 25% of patients with peptic strictures have no antecedent history of heartburn. Furthermore, the majority of patients with dysphagia from adenocarcinoma in Barrett esophagus have a history of long-standing heartburn. Also, around 30% of patients with achalasia complain of heartburn. Therefore, conclusions regarding the etiology of dysphagia should not be based primarily on the presence or absence of heartburn.

Has the patient taken medications likely to cause pill esophagitis? A number of medications taken in pill form are

Table 1 Diseases of the Esophagus That Cause Dysphagia

Mucosal diseases	Gastroesophageal reflux disease (peptic stricture) Esophageal rings Esophageal tumors Caustic injury (e.g., lye ingestion, pill esophagitis, sclerotherapy) Radiation injury Infectious esophagitis
Mediastinal diseases	Tumors (e.g., lung cancer, lymphoma) Infections (e.g., tuberculosis, histoplasmosis)
Diseases affecting smooth muscle and its innervation	Achalasia Scleroderma Other motility disorders

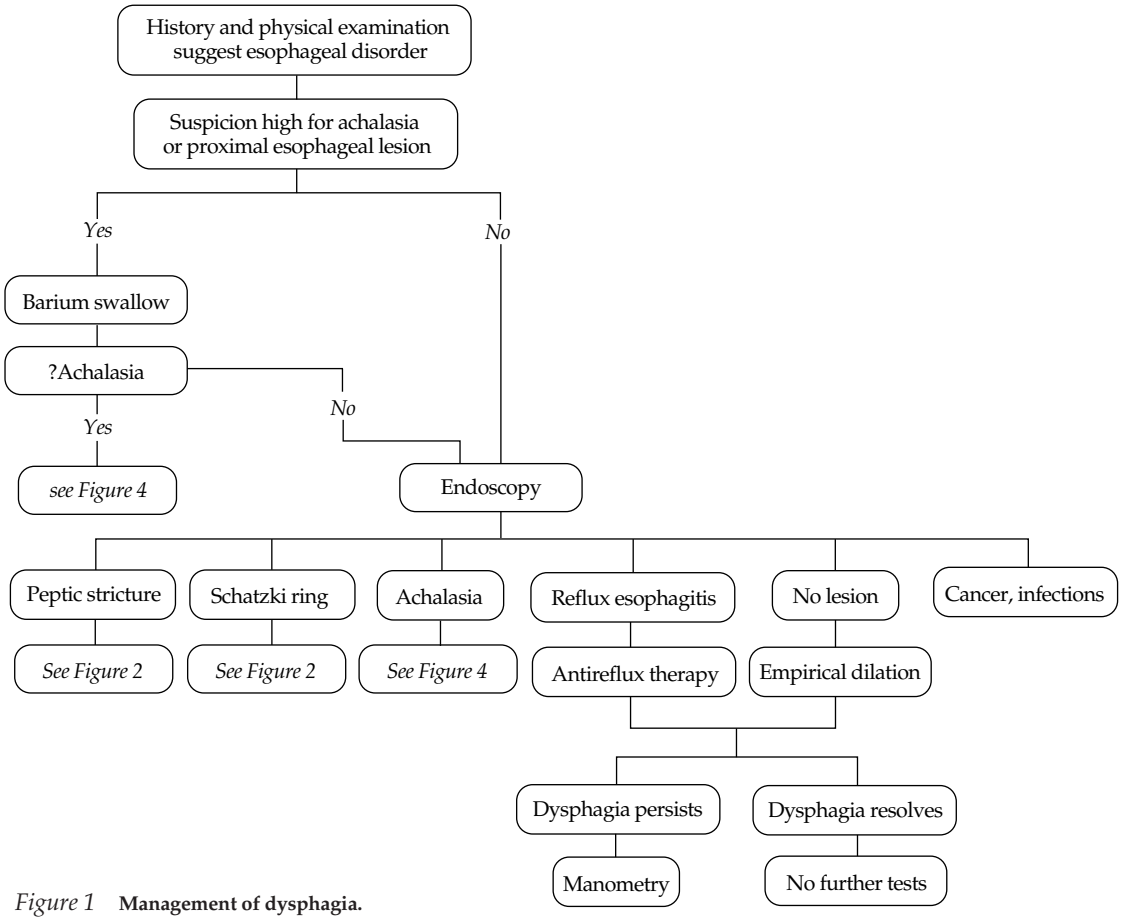


Figure 1 **Management of dysphagia.**

potentially caustic to the esophagus and can cause ulceration with stricture formation if they have prolonged contact with the esophageal mucosa. Although a large number of medications have been reported to cause pill esophagitis, most cases have been caused by antibiotics (e.g., doxycycline), potassium chloride preparations, nonsteroidal anti-inflammatory drugs (NSAIDs), and quinidine.

Is there a history of collagen vascular disease? Collagen vascular diseases such as scleroderma, rheumatoid arthritis, and systemic lupus erythematosus can cause disordered esophageal motility that is often associated with Raynaud phenomenon.[5] In scleroderma and related collagen vascular disorders, fibrosis and vascular obliteration in gut smooth muscle cause poor esophageal contractility and weakness of the lower esophageal sphincter that predisposes to severe GERD. Also, patients with collagen vascular disease often are treated with medications, such as NSAIDs, that can cause pill esophagitis. Consequently, dysphagia associated with collagen vascular disease may be the result of disordered esophageal motility, severe GERD, pill esophagitis, or some combination thereof.

Is the patient immunosuppressed? Infectious esophagitis occurs most often in patients whose immune system has been compromised by infection with HIV, by advanced malignancy, or by organ transplantation with the administration of potent immunosuppressive drugs. Odynophagia is usually the predominant symptom in infectious esophagitis, but most patients with this disease also experience dysphagia.[6] In rare cases, esophageal stricture can be a late complication of infectious esophagitis.

Physical Examination

The physical examination of the patient with dysphagia is important primarily for assessing the patient's nutritional status and ability to tolerate the invasive procedures that may be considered to treat the esophageal disorder. Only infrequently does the physical examination provide specific clues to the etiology of dysphagia. For patients with dysphagia caused by collagen vascular disease, physical examination may reveal characteristic features such as joint abnormalities, calcinosis, telangiectasias, sclerodactyly, proximal muscle weakness, and rashes. A palpable left supraclavicular (Virchow) lymph node suggests dysphagia from a malignancy in the abdomen (e.g., adenocarcinoma of the esophagogastric junction). Also, the physical examination may reveal evidence of a neuromuscular disorder that can interfere with swallowing (e.g., Parkinson disease).

DIAGNOSTIC TESTS

Testing in patients with dysphagia generally starts with barium swallow or endoscopy. If initial testing discloses an esophageal motility disorder, manometry may then be done. Videofluoroscopy, in which a motion recording is made while the patient swallows barium suspensions and barium-coated materials, is an excellent technique for assessing oropharyngeal function but generally is not needed for evaluating esophageal disorders.

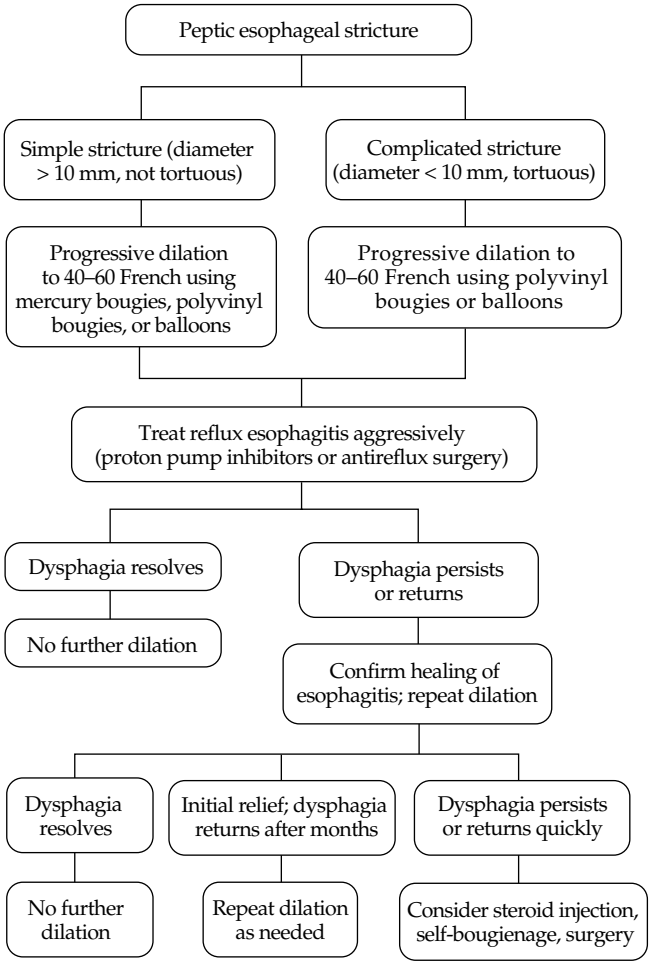


Figure 2 **Management of peptic esophageal stricture.**

There is an unresolved debate about whether to start the evaluation of dysphagia with a barium swallow or with esophageal endoscopy. Proponents of the latter approach argue that endoscopy is almost always required to evaluate dysphagia, for both diagnostic and therapeutic purposes, and that a barium swallow usually does not provide sufficient additional information to justify its expense and inconvenience. On the other side of the argument are those who contend that a barium swallow can provide valuable anatomic information about the esophagus that may help direct therapy and prevent procedural complications. In the absence of studies validating the cost-effectiveness of either approach, this debate continues.

Barium Swallow

A barium contrast examination can be more sensitive than endoscopy for detecting subtle narrowings of the esophagus (e.g., rings, peptic strictures greater than 10 mm in diameter) and for identifying esophageal dysmotility.[7] A barium swallow may be especially helpful in suggesting the diagnoses of achalasia and diffuse esophageal spasm, conditions that may be difficult to identify endoscopically. The early radiographic demonstration of achalasia may spare the patient repeat endoscopy, a situation that can occur because the endoscopist either did not recognize the disorder on the initial evaluation or was not prepared to perform endoscopic therapy at that time. A barium swallow can identify lesions that may pose potential hazards or confuse the endoscopist, such as a large Zenker diverticulum or an epiphrenic diverticulum. For patients with an esophageal stricture, a barium swallow can provide information on the extent and severity of the lesion that may help in choosing the type of dilator to be used for treatment. Finally, an initial barium swallow provides an objective baseline record of the esophagus that can be useful in assessing the response to therapy or progression of disease.

Endoscopy

For virtually all patients with dysphagia of esophageal origin, endoscopy is recommended to establish or confirm a diagnosis, seek evidence of esophagitis and malignancy, and implement therapy when appropriate. The endoscopist can obtain biopsy and brush cytology specimens of esophageal lesions that may establish the diagnosis of neoplasms or specific infections. Endoscopy also is more sensitive than radiology for identifying subtle mucosal lesions of the esophagus (e.g., mild esophagitis).

Esophageal Manometry

Esophageal manometry is the gold standard test for esophageal motility disorders. Esophageal manometry has been shown to be especially useful for establishing the diagnoses of achalasia and diffuse esophageal spasm and for detecting esophageal motor abnormalities associated with collagen vascular diseases.[8]

For patients with dysphagia, the history and the results of the barium swallow or endoscopy can be used to decide whether esophageal manometry is indicated. An esophageal motility study usually is not needed for patients with mechanical causes of dysphagia, such as peptic strictures or rings, unless their dysphagia persists despite appropriate treatment. For patients thought to have dysphagia caused by motility abnormalities associated with collagen vascular diseases, manometry need not be performed routinely if dysphagia responds to treatment of any associated reflux esophagitis and esophageal stenoses. If the dysphagia persists despite such treatment, manometry can establish the nature of the motility problem.

Dysphagia from Benign Esophageal Strictures and Rings

BENIGN ESOPHAGEAL STRICTURES

Strictures develop from severe esophageal inflammation, usually associated with ulceration, that stimulates fibrous tissue production and collagen deposition. Approximately two thirds of all cases of benign esophageal stricture in the United States are caused by reflux esophagitis (so-called peptic strictures).[9] The remainder are the result of caustic ingestions (e.g., lye), pill esophagitis, infectious esophagitis, and radiation esophagitis.

Treatment

Benign esophageal strictures usually are treated with dilation [*see Figure 2*]. Three major types of esophageal dilating devices are used commonly: (1) mercury-filled bougies that are passed blindly through the mouth (e.g., tapered-tipped Maloney dilators, blunt-tipped Hurst dilators); (2) polyvinyl bougies that can be passed over a fine guide wire that is positioned in the stricture, under either fluoroscopic or endoscopic guidance (e.g., Savary dilators), and (3) balloon dilators that are passed either over a guide wire or through the endoscope (so-called through-the-scope [TTS] balloons). Usually, the physician passes a series of dilators of increasing diameter to stretch

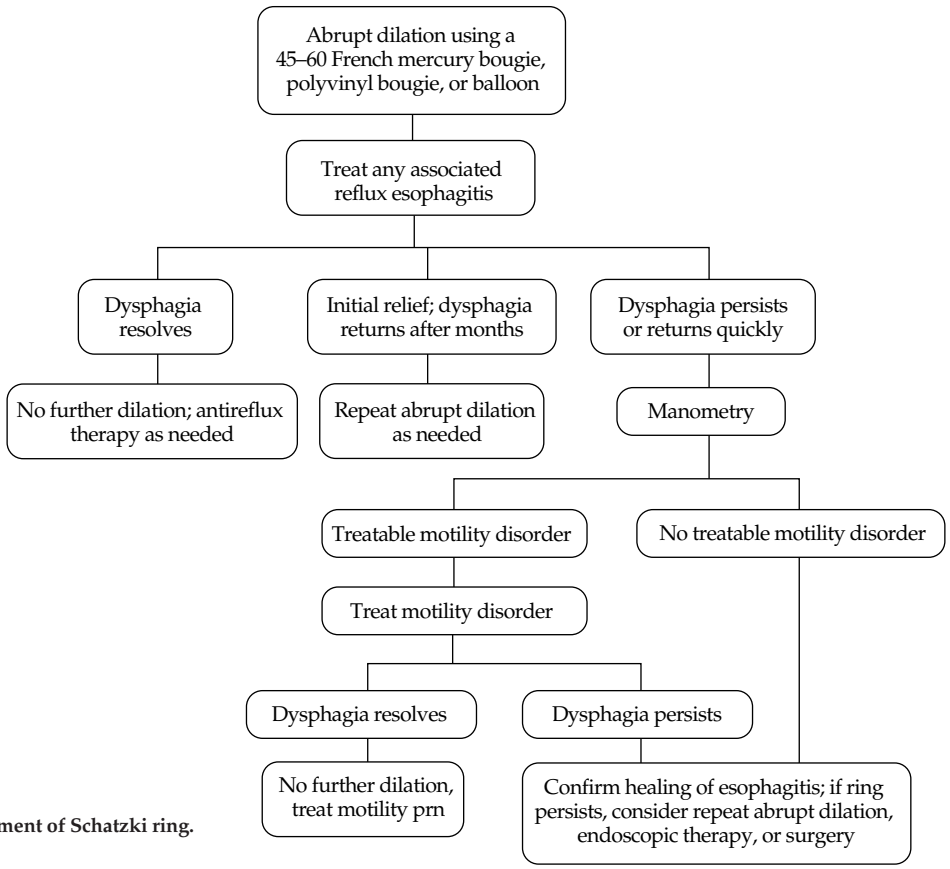


Figure 3 **Management of Schatzki ring.**

the stricture gradually. No study has established the superiority of one type of dilator over another. Serious complications such as perforation and bleeding occur in approximately 0.5% of all esophageal dilation procedures.[2]

For patients with peptic strictures caused by GERD, aggressive treatment with proton pump inhibitors both improves dysphagia and decreases the need for subsequent esophageal dilations.[10] Surgical therapy can be used for esophageal strictures that do not respond to dilation and antisecretory therapy. There are two major surgical approaches: (1) antireflux surgery with intraoperative stricture dilation for patients with peptic strictures or (2) resection of the stenotic esophagus with esophageal reconstruction (e.g., by interposing a loop of bowel between the remaining esophagus and the stomach).

lower esophageal (schatzki) rings

The lower esophageal mucosal (Schatzki) ring is a thin, diaphragmlike, circumferential fold of mucosa that protrudes into the lumen of the distal esophagus, thereby posing a physical barrier to the passage of solid material.[11] Mucosal rings usually are located at the squamocolumnar junction and have squamous epithelium lining their upper surface and columnar epithelium lining the lower aspect. With careful radiologic technique aimed at distending the distal esophagus, a lower esophageal ring can be found in approximately 15% of all patients who have barium swallows. Only a minority of these rings cause dysphagia, however.

The pathogenesis of lower esophageal mucosal rings is disputed. It is not clear whether they are congenital or acquired structures. Lower esophageal mucosal rings often are associated with hiatal hernias and GERD, and some authorities have suggested that the rings are in fact thin peptic strictures. Data on the role of GERD in the pathogenesis of Schatzki rings are inconclusive and contradictory, however.

Treatment

Dilation therapy is recommended for patients who have dysphagia from Schatzki rings [*see Figure 3*]. Traditionally, this involves the passage of a single large bougie or balloon (45 to 60 French) aimed at fracturing (rather than merely stretching) the mucosal fold. This approach differs from that for peptic strictures, which are treated by gradual stretching for fear of rupturing the fibrotic esophagus with a single, abrupt dilation. Most patients experience immediate relief of dysphagia after dilation, but recurrence is common and many patients require repeated dilations.

Esophageal Motility Abnormalities

Spechler and Castell have recently proposed a classification system for esophageal motility disorders. This system categorizes such disorders according to four major patterns of esophageal manometric abnormalities: inadequate LES relaxation, uncoordinated contraction, hypercontraction, and hypocontraction [*see Table 2*].[12] Most esophageal motility abnormalities fall predominantly into one of these four major categories, although there can be considerable overlap.

Processes that affect the inhibitory innervation of the LES (e.g., achalasia) can interfere with LES relaxation and thereby delay esophageal clearance. In the body of the esophagus, abnormal motility is characterized by uncoordinated contraction, hypercontraction, and hypocontraction. Uncoordinated esopha-

Table 2 Classification of Esophageal Motility Abnormalities[12]

Inadequate relaxation of lower esophageal sphincter (LES)
Classic achalasia
Atypical disorders of LES relaxation
Uncoordinated contraction
Diffuse esophageal spasm
Hypercontraction
Nutcracker esophagus
Isolated hypertensive LES
Hypocontraction
Ineffective esophageal motility

geal contractions (i.e., contractions that are not peristaltic and directed toward the stomach) can delay esophageal clearance. Such uncoordinated contractions are the hallmark of diffuse esophageal spasm. Hypercontraction abnormalities are those that are characterized by contractions that are of high amplitude, long duration, or both. The putative disorders of hypercontraction (e.g., nutcracker esophagus, isolated hypertensive LES) are perhaps the most controversial of the abnormal esophageal motility patterns because it is not clear whether esophageal hypercontraction has any pathophysiologic significance. In contrast, hypocontraction abnormalities that result from weak (low-amplitude) muscle contractions can cause ineffective esophageal motility that delays esophageal clearance, and LES hypotension can result in GERD.

ACHALASIA

Primary achalasia is the best characterized of all the esophageal motility disorders.[13] In achalasia, there is degeneration of neurons in the wall of the esophagus, especially the nitric oxide–producing inhibitory neurons that effect the relaxation of esophageal smooth muscle necessary for opening of the LES and for coordinated esophageal contraction. Degenerative changes also may be found in brain stem ganglion cells and their efferent fibers, but the disordered motility appears to result primarily from the degeneration of intramural neurons. The loss of inhibitory innervation in the LES causes basal sphincter pressures to rise and interferes with sphincter relaxation. In the body of the esophagus, the loss of intramural neurons results in aperistalsis.

Primary achalasia has an annual incidence of approximately one case per 100,000 population. The disorder affects men and women equally and usually is diagnosed in patients who are between 25 and 60 years of age.

Secondary achalasia, or pseudoachalasia, which can be

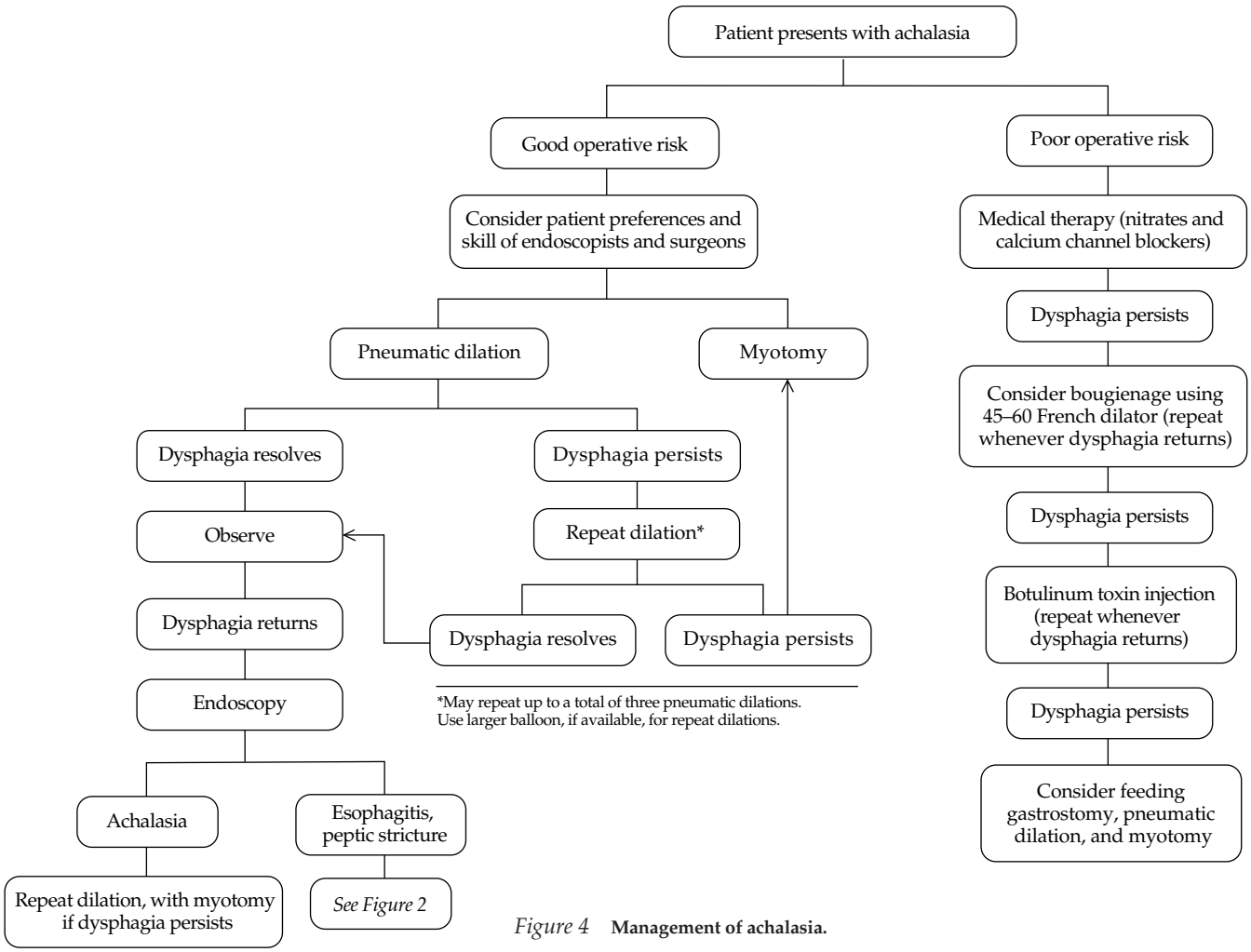


Figure 4 **Management of achalasia.**

caused by certain diseases, exhibits esophageal motor abnormalities identical to those of primary achalasia. In Chagas disease, for example, esophageal infection with *Trypanosoma cruzi* can destroy intramural ganglion cells and cause aperistalsis with incomplete LES relaxation. Malignancies can cause pseudoachalasia by invading the esophageal neural plexuses directly or, very rarely, by releasing uncharacterized humoral factors that disrupt esophageal function as part of a paraneoplastic syndrome.

Diagnosis

Clinical features Dysphagia for both solid foods and liquids is the primary symptom of achalasia. Moderate weight loss, regurgitation, and chest pain also are common clinical features. For reasons that are not clear, approximately one third of patients complain of heartburn, and achalasia occasionally can be confused with GERD. The symptoms of achalasia can be insidious in onset and gradual in progression, and patients frequently experience symptoms for years before seeking medical attention.

Diagnostic studies Achalasia can be confirmed with radiographic, manometric, and endoscopic evaluation. Occasionally, the diagnosis is suggested on a plain radiograph of the chest that shows widening of the mediastinum from the dilated esophagus and reveals absence of the normal gastric air bubble, because LES contraction prevents swallowed air from entering the stomach. Barium swallow, which has a diagnostic accuracy for achalasia of 95%, typically shows a dilated esophagus that terminates in a beaklike narrowing caused by persistent contraction of the LES [*see Figure 5*].

Esophageal manometry is the gold standard for the diagnosis of achalasia. The requisite manometric features are (1) incomplete relaxation of the LES (defined as a mean swallow-induced fall in resting LES pressure to a nadir value more than 8 mm above gastric pressure) and (2) aperistalsis characterized either by simultaneous esophageal contractions with amplitudes less than 40 mm Hg or by no apparent esophageal contractions. Other common manometric features of classic achalasia include LES hypertension (resting pressure greater than 45 mm Hg) and resting pressure in the esophageal body that exceeds resting pressure in the stomach.

Diagnostic endoscopy is recommended for patients with achalasia, primarily to exclude malignancy at the esophagogastric junction. In primary achalasia, the endoscopist sees a dilated esophagus that often contains residual food. The LES does not open spontaneously, but in contrast to cases of obstruction caused by neoplasms or fibrotic strictures, the contracted LES of achalasia usually can be traversed easily with gentle pressure on the endoscope.

Treatment

Treatments for achalasia are aimed at decreasing the resting pressure in the LES to the point that the LES no longer impedes the passage of swallowed material [*see Figure 4*]. There is no therapy that can halt or reverse the degeneration of enteric neurons, and no treatment reliably restores peristaltic function in the body of the esophagus.

Pharmacotherapy Nitrates and calcium channel blockers relax the smooth muscle of the LES, and these agents have been used to treat achalasia, with limited success. The drugs usually are taken sublingually 10 to 30 minutes before meals. Unfortunately, pharmacotherapy for achalasia is inconvenient, often ineffective, and frequently associated with side effects (e.g., headache and hypotension) and tachyphylaxis. Consequently, pharmacotherapy is used primarily for patients who are unwilling or unable to tolerate invasive therapies.

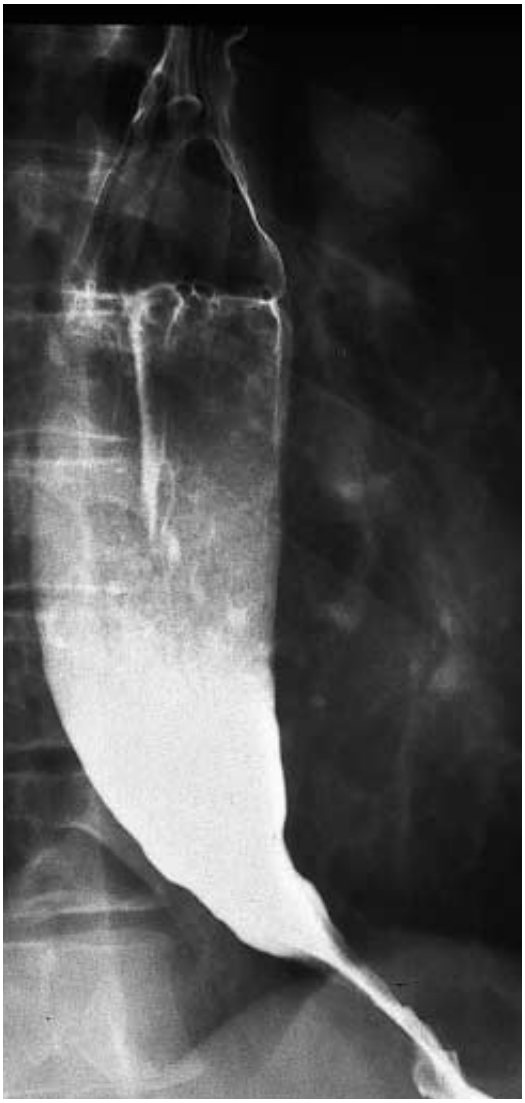


Figure 5 **Barium swallow in a patient with achalasia shows dilatation of the esophagus with a beaklike narrowing at the esophagogastric junction.**

Pneumatic dilation therapy In pneumatic dilation therapy for achalasia, a large deflated balloon is passed through the mouth to the LES, where the balloon is inflated rapidly to tear the sphincter muscle and thereby weaken the LES.[14] Most studies describe good short-term relief of dysphagia in 60% to 85% of patients who are treated with a single session of pneumatic dilation. Esophageal perforation is the most common serious complication of the procedure, occurring in 2% to 6% of cases in most large series. Approximately 50% of patients who are treated with pneumatic dilation will require further therapy within 5 years, and subsequent pneumatic dilations are progressively less likely to result in a sustained remission.

Surgical therapy Achalasia also can be treated by surgical myotomy, in which the surgeon weakens the LES by cutting its muscle fibers. The procedure now can be performed laparoscopically, and good to excellent relief of symptoms has been described in 70% to 90% of patients. There are few serious complications, although reflux esophagitis develops postoperatively in approximately 10% of patients. Some surgeons recommend that an antireflux (fundoplication) procedure be performed at the time of myotomy to prevent this complication. The few long-term studies available suggest that surgical myotomy results in sustained remission rates of approximately 85% at 10 years and 65% at 20 years. In a prospective, randomized comparison of myotomy with pneumatic dilation, excellent results were found after median follow-up of approximately 5 years in 40 of 42 (95%) patients in the surgical group, compared with 24 of 37 (65%) patients who had pneumatic dilation.[15] Currently, the decision between pneumatic dilation and myotomy as initial therapy for achalasia should take into consideration the patient's preferences and the experience of available personnel.

Botulinum toxin therapy Endoscopic injection of botulinum toxin into the LES poisons the excitatory (acetylcholine-releasing) neurons that contribute to LES smooth muscle tone, thereby decreasing LES pressure and relieving achalasia.[16] The procedure is safe, and most patients treated with botulinum toxin injection experience immediate symptomatic improvement. Unfortunately, this effect usually is short lived. With repeated injections, approximately two thirds of patients with achalasia can be maintained in remission for 6 months. Of those patients in remission at 6 months, however, only about two thirds are still in remission at 1 year, despite repeated injections. Botulinum toxin injection is used primarily to treat patients with serious comorbidities, for whom pneumatic dilation and surgical myotomy pose inordinate risks.

DIFFUSE ESOPHAGEAL SPASM

Diffuse esophageal spasm is a rare condition of unknown etiology characterized by uncoordinated (spastic) motor activity in the smooth muscle portion of the esophagus. The esophageal spasms are manifested clinically by episodes of dysphagia and chest pain and on radiography and manometry by tertiary (nonperistaltic) contractions of the esophagus. The requisite manometric features of diffuse esophageal spasm are (1) simultaneous contractions associated with more than 10% of wet swallows and (2) a mean simultaneous contraction amplitude greater than 30 mm Hg.[12] A common manometric pattern is intermittent normal peristalsis alternating with periods of spontaneous, repetitive, and multiple-peaked contractions.

Diffuse esophageal spasm is treated with agents that relax esophageal smooth muscle, such as nitrates and calcium channel blockers, although few reports document the efficacy of this therapy. Psychotropic agents such as tricyclic antidepressants may help relieve the pain of diffuse esophageal spasm. Surgical treatment by long myotomy of the esophagus has been reported, but the efficacy of this operation is poorly documented.

NUTCRACKER ESOPHAGUS

In nutcracker esophagus, manometry reveals peristaltic waves in the distal esophagus that have mean amplitudes more than 2 standard deviations above normal values.[17] Although high-amplitude peristalsis is the most common motility abnormality observed in patients with noncardiac chest pain,[18] the clinical and physiologic importance of this manometric finding is disputed. It is not clear whether the hypercontraction of esophageal muscle is a cause of chest pain or is merely an epiphenomenon that is associated with the chest pain syndrome. The clinical response to smooth muscle–relaxing agents (nitrates and calcium channel blockers) is often disappointing. Treatment with psychotropic agents such as tricyclic antidepressants can be effective in controlling the chest pain.

ISOLATED HYPERTENSIVE LES

In patients with isolated hypertensive LES, esophageal motility is normal and the LES relaxes appropriately, but the mean resting LES pressure is abnormally high (i.e., > 45 mm Hg). Although this condition has been reported in patients with noncardiac chest pain, it is unlikely that the isolated hypertensive LES has any clinical consequences.

SCLERODERMA

In 80% of patients with scleroderma (progressive systemic sclerosis), fibrosis and ischemia damage the esophageal muscles and nerves. This damage causes manometric abnormalities, including weak contractions in the body of the esophagus and hypotension of the LES.[19] The LES hypotension can result in GERD. When the amplitude of peristaltic contractions falls below 30 mm Hg, esophageal clearance is compromised.

In addition to weak peristalsis, patients with scleroderma often exhibit abnormalities in the progression of peristalsis, including (1) failed peristalsis, in which the peristaltic wave progresses through the pharynx and proximal esophagus but fails to traverse the entire length of the distal esophagus; (2) simultaneous esophageal contractions; and (3) absent esophageal contractions.

The manometric features of scleroderma are not specific for the disorder. Identical manometric abnormalities can be found in patients with the CREST variant (limited scleroderma [calcinosis, Raynaud phenomenon, esophageal involvement, sclerodactyly, and telangiectasias]) and other collagen vascular disorders (e.g., mixed connective tissue disease, rheumatoid arthritis, and systemic lupus erythematosus) and certain nonrheumatic diseases (e.g., diabetes mellitus and amyloidosis). Furthermore, otherwise healthy patients who have GERD often exhibit scleroderma-like motility disturbances, yet such patients infrequently develop rheumatic diseases. For these reasons, some authorities discourage use of the term scleroderma esophagus and recommend use of the term ineffective esophageal motility to describe the constellation of manometric abnormalities typical of scleroderma.[20]

There is currently no effective treatment for ineffective esophageal motility. Proton pump inhibitors can treat and prevent complications from associated GERD. Prokinetic agents such as metoclopramide are often used but are rarely effective, especially when the disease is advanced.

Gastroesophageal Reflux Disease

In GERD, the reflux of gastric juice into the esophagus or oropharynx causes symptoms, tissue injury, or both. Approximately 20% of adults in the United States experience GERD symptoms such as heartburn and acid regurgitation at least once a week.[21] Severe GERD can result in ulceration of the esophagus,

which can lead to fibrosis and esophageal stricture formation. In some cases, the ulcerated squamous epithelium of the distal esophagus is replaced by a metaplastic, intestinal-type mucosa (a condition called Barrett esophagus) that predisposes to cancer. Consequently, GERD is a strong risk factor for esophageal adenocarcinoma,[22] a tumor whose frequency has increased profoundly in western countries over the past 2 decades.

PATHOPHYSIOLOGY

The development of GERD is a multifactorial process involving dysfunction of the antireflux mechanisms that normally prevent gastric juice from entering the esophagus and of the clearance mechanisms that normally rid the esophagus of refluxed material.[23]

Antireflux Mechanisms

The normal barrier to gastroesophageal reflux comprises three major components: (1) the LES, (2) the crural diaphragm, and (3) anatomic features of the gastroesophageal junction.

Lower esophageal sphincter The LES normally prevents reflux by maintaining a resting pressure that is 10 to 30 mm Hg higher than ambient pressure in the stomach.[24] In the 1980s, Dodds showed that episodic collapse of LES pressure, a phenomenon called transient LES relaxation (TLESR), is the major mechanism for gastroesophageal reflux both in normal individuals and in patients with GERD.[25] Unlike the brief (2- to 8-second) LES relaxations that normally accompany primary (swallow-induced) peristalsis, TLESRs are not preceded by swallowing and last from 10 to 45 seconds.[26] When LES pressure becomes identical to gastric pressure during a TLESR, the sphincter can no longer function as a barrier to acid reflux.

TLESRs occur two to six times per hour in normal persons, especially after meals, and 40% to 50% of these TLESRs are accompanied by brief episodes of acid reflux. In patients with GERD, TLESRs occur three to eight times per hour, and 60% to 70% of these are associated with acid reflux. In severe GERD, approximately 70% of reflux episodes are the result of TLESRs; the remaining reflux episodes are from feeble basal LES pressure, swallow-induced LES relaxation, and sudden elevations in abdominal pressure.[24]

Gaseous distention of the stomach normally triggers a TLESR that allows the gas to escape into the esophagus (the belch reflex).[26] The nucleus tractus solitarius in the medulla mediates the belch reflex by integrating sensory information from the stomach and by controlling the neural circuits that induce the TLESR. Activation of medullary neurons with γ-aminobutyric acid B ($GABA_B$) receptors inhibits TLESRs, as does cholinergic blockade with atropine. The sphincter relaxation of the TLESR is effected by the activation of cholecystokinin-A receptors in LES muscle.

Crural diaphragm The right crus of the diaphragm normally encircles the distal esophagus. During inspiration, contraction of the diaphragmatic crura pinches the distal esophagus to prevent reflux. Crural contraction also helps minimize reflux during the sudden increases in abdominal pressure that accompany events such as coughing, sneezing, and straining. Thus, the crural muscle functions as an external esophageal sphincter that buttresses the LES.[27] TLESRs often are accompanied by relaxation of the crura, and studies in dogs have shown that gastroesophageal reflux does not occur during a TLESR unless the episode is attended by neural inhibition of the crural diaphragm.

Anatomic features of the gastroesophageal junction The acute angle formed by the junction of the esophagus and the stomach (the angle of His) can function as a one-way flap valve that stops reflux. Also, a portion of the distal esophagus normally is located within the abdomen, where the high ambient pressure pushes the esophageal walls together to prevent reflux.[27]

Disruption of the Antireflux Barrier by Hiatal Hernia

Most patients with severe GERD have a hiatal hernia in which the proximal portion of the stomach herniates into the chest through the diaphragmatic hiatus formed by the right crus of the diaphragm. With a large hiatal hernia, the LES muscle is completely dissociated from the crural diaphragm. In this situation, contraction of the crural diaphragm does not pinch the esophagus; rather, it creates an intrathoracic pouch of stomach whose contents may reflux readily into the esophagus.[28,29] With no buttressing of the internal sphincter by the crural diaphragm, sudden elevations in abdominal pressure caused by inspiration, coughing, and straining can far exceed LES pressure, resulting in reflux. Reduction of the angle of His and loss of the intra-abdominal portion of the esophagus also may compromise the antireflux barrier. Furthermore, patients with large hiatal hernias have an abnormally high frequency of TLESRs induced by gastric distention.[30]

Disruption of Esophageal Clearance

To injure the esophagus, caustic refluxed material must remain in contact with the mucosa for a sufficient period. The duration of this contact is determined by the efficacy of the esophageal clearance mechanisms, which include gravity, peristalsis, salivation, and bicarbonate secretion by the submucosal glands of the esophagus. When gastric juice refluxes into the esophagus, most of the material is cleared by the combined effects of gravity and peristalsis.[31] The small quantity of residual acidic material is neutralized by alkaline saliva and, to a lesser extent, by bicarbonate secreted by the submucosal glands.

Peristaltic abnormalities that can interfere with esophageal emptying (e.g., failed peristalsis, hypotensive peristalsis) have been found in 25% to 48% of patients with reflux esophagitis.[32] Patients with large hiatal hernias frequently have impaired esophageal clearance because of the retrograde flow of material from the hernia back into the esophagus. Cigarette smoking also has been shown to increase esophageal acid exposure both by increasing the frequency of acid reflux events and by decreasing salivary flow.[33]

Normal persons regularly experience brief episodes of acid reflux that do not cause esophageal injury. Most patients with reflux esophagitis have prolonged esophageal acid exposure that overwhelms the normal epithelial defenses. In some patients with reflux esophagitis, however, 24-hour pH monitoring studies demonstrate a normal daily duration of acid reflux. It is conceivable that these patients have uncharacterized defects that render the esophageal epithelium vulnerable to normal acid reflux.

ETIOLOGY

NSAIDs

There is evidence to suggest that aspirin and other NSAIDs may contribute to GERD.[34,35] Some NSAID preparations are caustic to the esophageal mucosa, and esophagitis can result if the tablet lingers in the esophagus (pill esophagitis). Patients

with esophageal strictures may be especially susceptible to NSAID-induced esophageal injury.

Helicobacter pylori

Some studies suggest that gastric infection with *H. pylori* protects the esophagus from GERD and its complications.[36] In general endoscopy units, *H. pylori* infection has been found significantly less often in patients with reflux esophagitis than in control patients without GERD.[37] Reflux esophagitis has developed in some patients with duodenal ulcers after their *H. pylori* infections were eradicated with antibiotics.[38] There is a negative association between esophageal adenocarcinoma and *H. pylori* infections, particularly for infections with *cagA*-positive strains that may be especially likely to cause severe pangastritis.[39,40] Graham and others have proposed that *H. pylori* infections that cause pangastritis also cause a decrease in gastric acid production that may protect against GERD.[41] This issue remains highly controversial, however, and the role of *H. pylori* infection in GERD is not yet clear.

DIAGNOSIS

Clinical Features

Heartburn, the cardinal symptom of GERD, is an uncomfortable, burning sensation that is located behind the sternum. The sensation often originates in the epigastrium and radiates up the chest. Patients who describe heartburn often wave an open hand vertically over the sternum, whereas patients with angina pectoris typically hold a clenched fist stationary over the chest while describing their pain. Heartburn may be associated with the ingestion of foods that predispose to reflux by decreasing pressure in the LES, such as chocolate, onions, and fat. Spicy foods, citrus products, and tomato products may cause the sensation of heartburn by irritating the inflamed esophageal mucosa directly. Some patients experience heartburn when they bend over or exercise, presumably because these activities induce acid reflux by increasing intra-abdominal pressure. Characteristically, heartburn caused by gastroesophageal reflux is relieved immediately by antacids and can be eliminated by the administration of potent acid-suppressing agents.

When refluxed gastric juice reaches the oropharynx, patients may complain of regurgitation of sour (acid) or bitter-tasting (bilious) material. Dysphagia may result from a peptic stricture or from reflux esophagitis alone. With ulcerative esophagitis, patients may complain of odynophagia. Some patients experience water brash, in which the mouth suddenly fills with saliva as a result of reflex salivation stimulated by acid in the esophagus.

In addition to these typical symptoms, GERD can have a variety of so-called atypical manifestations.[42-45] Esophageal irritation by acid reflux may result in chest pain that can mimic ischemic heart disease. If gastric juice reaches the oropharynx, it can cause globus, sore throat, and burning tongue, and the acid can erode dental enamel. Laryngitis and pulmonary problems such as chronic cough and asthma can result from aspiration of material into the airway. Asthma may also be a consequence of acid in the esophagus that triggers reflex bronchoconstriction.

Diagnostic Studies

Patients who have a typical history of heartburn that disappears with antisecretory therapy can be assumed to have GERD, and diagnostic tests are not necessary merely to confirm the diagnosis. Moreover, by definition, patients with GERD can have symptoms without objective evidence of esophageal damage, so normal test results often cannot exclude GERD as a cause of symptoms. However, diagnostic tests may be needed to evaluate atypical symptoms or to look for complications of GERD. For such patients, several options are available, such as radiography, endoscopy, biopsy, acid perfusion test, and pH monitoring.

Radiography Radiography has a limited role in the evaluation of GERD. A barium swallow can show signs of reflux esophagitis (e.g., thickening of the esophageal folds, erosions, and ulcerations) that support the diagnosis of GERD. A barium swallow can also identify peptic strictures. Radiography is considerably less sensitive than endoscopy for demonstrating esophagitis, however; and endoscopic examination has the added advantage of permitting biopsy of specimens from any abnormal areas.

Endoscopy An endoscopic examination is the most sensitive test for establishing the diagnosis of reflux esophagitis and Barrett esophagus. However, the clinician should appreciate that a normal endoscopic examination does not eliminate GERD as a cause of symptoms. Gastroesophageal reflux can cause disabling symptoms without causing visible esophageal damage.[46] Endoscopy shows reflux esophagitis in only 50% to 70% of patients who complain of frequent heartburn, and heartburn severity is not a reliable index for the severity of esophagitis.[47] Furthermore, the esophagus typically appears normal on endoscopy in patients who have only extraesophageal symptoms of GERD, such as chronic cough and laryngitis.

The Practice Parameters Committee of the American College of Gastroenterology recommends that endoscopy be reserved for patients with uncomplicated GERD in whom empirical therapy is unsuccessful. Endoscopic evaluation without an empirical trial of therapy is appropriate in patients who have symptoms suggesting complicated disease, including fever, anorexia, weight loss, dysphagia, odynophagia, and bleeding.[48] In patients with long-standing GERD symptoms, particularly those 50 years of age and older, endoscopy is indicated to look for Barrett esophagus.[49] It is important to appreciate that these guidelines are merely committee recommendations whose efficacy has not been established by clinical studies.

Biopsy Biopsy specimens of the squamous epithelium in the distal esophagus of patients with GERD frequently show the histologic abnormalities of reflux esophagitis, including lengthening of the papillae, hyperplasia of cells in the basal zone, and infiltration of the epithelium with eosinophils and polymorphonuclear cells. The importance of these histologic changes of GERD is disputed, however; and routine biopsy of the squamous epithelium to seek evidence of reflux esophagitis generally is not recommended for clinical purposes.

Acid perfusion test The acid perfusion (Bernstein) test has been used in patients who have atypical chest pain. The esophagus is perfused with 0.1N hydrochloric acid; reproduction of the chest pain implicates GERD as the etiology. This test has limited sensitivity and specificity, however, and has largely been replaced by ambulatory esophageal pH monitoring.

pH monitoring Ambulatory monitoring of esophageal

pH is used to document the pattern, frequency, and duration of acid reflux and to seek a correlation between reflux episodes and symptoms. In most ambulatory systems, an episode of acid reflux is defined as a drop in esophageal pH below 4. Esophageal pH monitoring records a number of different variables, such as the total number of reflux episodes, the number of episodes that last longer than 5 minutes, and the duration of the longest episode. The most useful variable appears to be the percentage of the monitoring period in which esophageal pH remains below 4 (in normal individuals, this is less than 4.5% of the 24-hour test). Ambulatory esophageal pH monitoring usually is not needed for patients with typical signs and symptoms of GERD. For patients with atypical or unresponsive symptoms, however, the test can be very useful in documenting an association between symptoms and acid reflux, as well as the efficacy of antisecretory therapy in reducing acid reflux.

TREATMENT

The efficacy of antireflux therapy is inversely related to the severity of the underlying reflux esophagitis.[50] A treatment that is highly effective for mild esophagitis may be virtually useless for patients with severe disease. Ulcerative esophagitis does not respond reliably to medical therapy with agents other than proton pump inhibitors (PPIs), the most effective antisecretory medications. For patients with mild or moderate esophagitis, some authorities advocate a step-up approach to therapy that begins with antireflux lifestyle modifications and progresses eventually to the most potent medications (i.e., PPIs) only when the disease does not respond to lesser treatments.[51] Others advocate a step-down approach that begins with the most effective therapy (PPIs).[52] The optimal approach remains disputed. However, for patients who are known to have severe, ulcerative reflux esophagitis, it is appropriate to begin therapy immediately with PPIs rather than proceeding stepwise through trials of agents unlikely to effect healing.

Lifestyle Modifications

A number of lifestyle modifications have been proposed to decrease esophageal acid exposure [*see Table 3*]. Few published data support the efficacy of these lifestyle modifications in controlling GERD, however, and it is unclear how many patients for whom such modifications are prescribed actually comply with them.

Antacids and Alginic Acid

Antacids and alginic acid can provide temporary relief of episodic heartburn. Despite the wide use of these over-the-counter products, surprisingly few data are available on their utility for healing reflux esophagitis or for the long-term management of GERD symptoms.

H_2 Receptor Blocking Agents

H_2 receptor antagonists are safe medications that relieve GERD symptoms and heal esophagitis within 12 weeks in approximately one half to two thirds of all patients.[50,53] The H_2 blockers are most useful for patients with GERD of mild to moderate severity, in whom the highest rates of healing can be anticipated. However, healing rates with these agents are poor in patients who have severe reflux esophagitis. High doses of H_2 receptor blockers (up to eight times the conventional dose) have been used effectively to treat esophagitis in severe cases of GERD, but this approach generally is not recommended. Few data document the long-term efficacy of H_2 receptor blockers used in any dosage, and many patients develop tolerance to the antisecretory effects of these agents. For patients with severe GERD, most authorities prescribe PPIs rather than high-dose H_2 receptor blocker therapy.

Table 3 Lifestyle Modifications for Gastroesophageal Reflux Disease

- Elevate the head of the bed on 4- to 6-in blocks
- Weight loss for obese patients
- Avoid recumbency for several hours after meals
- Avoid bedtime snacks
- Avoid fatty foods, chocolate, peppermint, onions, and garlic
- Avoid cigarettes and alcohol
- Avoid drugs that decrease lower esophageal sphincter pressure and delay gastric emptying, such as calcium channel blocking agents
- Avoid nonsteroidal anti-inflammatory drugs

Prokinetic Agents

In theory, prokinetic agents may decrease gastroesophageal reflux by increasing LES pressure and by enhancing esophageal and gastric clearance. Currently, metoclopramide is the only prokinetic agent available in the United States for the treatment of GERD. Metoclopramide is a dopamine antagonist, and its use is limited by side effects such as agitation, restlessness, somnolence, and extrapyramidal symptoms, which occur in up to 30% of patients. Cisapride, a serotonin-4 (5-HT_4) receptor agonist, demonstrated efficacy in mild GERD, but this agent was withdrawn when it was found to cause lethal cardiac arrhythmias in patients with a number of predisposing conditions.

Sucralfate

Sucralfate, a complex metal salt of sulfated sucrose, is an exceptionally safe medication that has some demonstrated efficacy in the treatment of mild reflux esophagitis. Few published data are available on the use of sucralfate in GERD, however, and the drug has never achieved popularity as an antireflux therapy.

Proton Pump Inhibitors

The PPIs are substituted benzimidazoles that decrease gastric acid secretion through inhibition of H^+,K^+-ATPase, the proton pump of the parietal cell. These agents are clearly the most effective inhibitors of gastric acid secretion available. Five PPIs are used widely for the treatment of GERD: omeprazole, esomeprazole (the S-optical isomer of omeprazole), lansoprazole, pantoprazole, and rabeprazole. All of these preparations are similar in efficacy and side-effect profiles, although when used in conventional dosages, esomeprazole may effect marginally higher rates of healing of reflux esophagitis than lansoprazole.[54] Patients with mild to moderately severe reflux esophagitis who are treated with PPIs in conventional dosages achieve healing rates of 80% to 100% within 8 to 12 weeks.[50,55] Very severe reflux esophagitis may persist despite conventional-dose PPI therapy in up to 40% of cases, however.[56] In most of these resistant cases, the esophagitis usually can be healed by increasing the dose of the PPI.[57] Recent studies also have shown

that aggressive acid suppression with PPIs improves dysphagia and decreases the need for esophageal dilation in patients who have peptic esophageal strictures.[10]

Most patients with severe GERD who respond to PPIs require maintenance therapy, because GERD recurs shortly after stopping the drug.[58] For most patients, the dose of PPI necessary to maintain remission is at least the dose required to heal the acute esophagitis. In some patients with severe GERD, furthermore, PPI maintenance-dose requirements may increase with time. One long-term study of patients with severe GERD who were given a maintenance dose of 20 mg of omeprazole daily found that relapses occurred at the rate of 1 per 9.4 treatment-years and that patients often required increasing doses of omeprazole (up to 120 mg/day) to maintain GERD in remission.[57]

The profound acid suppression that can be achieved with PPIs has raised theoretical concerns regarding their long-term safety.[53] Nevertheless, there are no reports of tumors or nutritional deficiencies clearly attributable to the use of PPIs after more than a decade of extensive clinical experience with these agents.

Antireflux Surgery

Antireflux operations share four fundamental features: (1) reduction of the hiatal hernia, (2) restoration of an intra-abdominal segment of esophagus, (3) approximation of the diaphragmatic crura, and (4) fundoplication, in which the surgeon wraps a portion of the gastric fundus around the distal esophagus. The operations differ primarily in the approach (e.g., transthoracic versus transabdominal) and in the extent of fundoplication.

The precise mechanisms whereby these operations prevent reflux are not clear, but a number of potential ones have been proposed.[59] The fundoplication may prevent the distention of the gastric fundus that ordinarily triggers TLESRs. Restoration of the distal esophagus to the positive pressure environment of the abdomen may prevent reflux, and the anatomic rearrangement of the gastroesophageal junction may create an antireflux flap-valve effect. Also, reduction of the hiatal hernia and approximation of the diaphragmatic crura may restore the normal antireflux function of the crural diaphragm.

Antireflux surgery can be performed laparoscopically.[60] The laparoscopic approach has become popular not because it is safer or produces a better functional result than the open procedure, but because it supposedly offers the advantages of less postoperative discomfort, shorter hospital stay, and better cosmetic outcome. Two recent randomized trials of laparoscopic versus open Nissen fundoplication found no significant differences in the functional results of the two procedures (i.e., relief of GERD symptoms and reduction in esophageal acid exposure).[61,62] However, one of the studies was terminated prematurely because an interim analysis showed an excess of adverse outcomes (primarily dysphagia) in the group treated laparoscopically.[62] At least one study has shown that the primary factor involved in overall patient satisfaction with antireflux surgery is the relief of GERD symptoms. These observations suggest that the primary decision for the clinician to make is whether the patient should have an antireflux operation, not how the operation should be performed.

Uncontrolled retrospective studies of antireflux surgery generally have described excellent results, with success rates exceeding 80%.[60] Few randomized trials have done direct comparisons of medical and surgical antireflux therapies, however. In the late 1980s (before the release of PPIs for clinical use in the United States), the Department of Veterans Affairs conducted a large cooperative study that prospectively compared the efficacy of medical therapy with that of surgical therapy in 247 patients with complicated GERD.[63] Patients were prescribed antireflux lifestyle modifications and randomly assigned to receive one of three types of treatment: (1) continuous medical therapy (antacid tablets and ranitidine taken on a daily basis regardless of symptoms, with metoclopramide and sucralfate added if necessary to control symptoms); (2) symptomatic medical therapy (the same drugs as in the continuous-medical-therapy group but used only when necessary for control of symptoms); or (3) surgical therapy (open Nissen fundoplication). All three groups showed significant improvements in the symptoms and endoscopic signs of GERD for up to 2 years. However, surgical therapy was significantly better than both medical therapies during that period, and surgical patients had higher overall satisfaction.

In the 1990s, a Scandinavian group conducted a randomized trial of omeprazole versus open antireflux surgery in 310 patients with erosive esophagitis.[64] In patients who received a fixed dose of 20 mg of omeprazole a day, antireflux surgery was superior in maintaining GERD in remission for the 3-year duration of the study. In clinical practice, however, the dose of a PPI is typically titrated to control symptoms. In this study, when the physician was permitted to titrate the PPI dose, there was no statistically significant difference between the medical group and the surgical group in maintaining remission for 3 years.

Relatively few reports have described the long-term outcome of antireflux surgery. Some uncontrolled studies have found success rates that exceed 90% at 10 to 20 years after open fundoplication,[65] whereas others have described breakdown of the operation and the return of reflux esophagitis in more than 50% of cases.[66] A follow-up study was conducted on the patients who participated in the VA cooperative study (see above).[67] During the follow-up period of 10 to 13 years, surgical patients were significantly less likely to have taken antireflux medications regularly, and when antireflux medications were discontinued, the GERD symptoms in these patients were significantly less severe than those in the medical group. However, 62% of the surgical patients took antireflux medications on a regular basis, and there were no significant differences between the groups in the rates of neoplastic and peptic complications of GERD, overall physical and mental well-being scores, and overall satisfaction with antireflux therapy. There were 79 deaths, involving 33 (40%) of the 82 surgical patients and 46 (28%) of the 165 medical patients ($P = 0.047$). For reasons that are not clear, the excess deaths in the surgical group were from heart disease. This and other studies suggest that antireflux surgery does not effect a permanent cure for GERD in the majority of patients and that surgery is no better than medication for preventing the peptic and neoplastic complications of GERD.

Endoscopic Antireflux Procedures

Two endoscopic therapies for GERD have been approved by the Food and Drug Administration. One system uses an endoscopic sewing-machine device to plicate the gastroesophageal junction; the other system delivers radiofrequency (microwave) energy to create thermal lesions in the LES muscle, which may narrow the lumen and destroy nerves that mediate TLESRs. A number of other endoscopic antireflux procedures are under in-

vestigation. Small studies describe promising results,[68,69] but the safety and efficacy of these procedures are not yet known, and their role in the treatment of GERD is not clear.

Barrett Esophagus

Barrett esophagus is a sequela of chronic GERD in which the stratified squamous epithelium that normally lines the distal esophagus is replaced by an abnormal columnar epithelium.[70] The abnormal columnar epithelium typical of Barrett esophagus is an incomplete form of intestinal metaplasia called specialized intestinal metaplasia. This metaplastic epithelium predisposes to esophageal adenocarcinoma, which develops in patients with Barrett esophagus at the rate of approximately 0.5% a year.

DIAGNOSIS

Endoscopy

The diagnosis of Barrett esophagus is established when the endoscopist sees columnar epithelium lining the distal esophagus. Columnar epithelium has a characteristic dull, reddish appearance that is readily distinguished from squamous epithelium, which is normally glossy and pale [*see Figure 6*]. The diagnosis is confirmed by esophageal biopsy specimens showing specialized intestinal metaplasia.

Barrett esophagus can be further categorized according to the extent of esophageal involvement. In traditional, or long-segment, Barrett esophagus (LSBE), specialized intestinal metaplasia extends for 3 cm or more into the distal esophagus. LSBE is usually found in patients who have severe GERD. Less than 3 cm of metaplasia constitutes short-segment Barrett esophagus (SSBE), a condition often associated with only mild GERD. It is not clear whether these two types of Barrett esophagus have the same pathogenesis and natural history, nor is it clear whether SSBE progresses to LSBE. Currently, however, SSBE and LSBE are managed similarly.

Regular endoscopic surveillance for esophageal cancer has been recommended in patients with Barrett esophagus. Retrospective studies have documented that cancers discovered during surveillance endoscopies tend to be less advanced than those detected during endoscopies performed because of cancer symptoms (e.g., dysphagia and weight loss). There is no direct proof that surveillance reduces cancer mortality in Barrett esophagus, however.

Biopsy

Esophageal biopsy specimens are taken during surveillance endoscopy primarily to identify dysplasia, a histologic diagnosis suggesting that one or more clones of epithelial cells have acquired genetic alterations rendering them neoplastic and predisposed to malignancy. Unfortunately, dysplasia is an imperfect predictor of malignancy in Barrett esophagus. The histologic abnormalities of low-grade dysplasia are not specific for neoplasia, and interobserver agreement for the diagnosis of low-grade dysplasia in Barrett esophagus may be less than 50%. For high-grade dysplasia, in contrast, interobserver agreement is approximately 85%. Dysplasia has no distinctive gross features, so endoscopists must rely on random biopsy sampling techniques to find it. Consequently, biopsy sampling error is a major problem. Of patients with Barrett esophagus whose biopsy specimens show high-grade dysplasia, approximately one third already have an invasive cancer that was missed because of sampling error. Extensive biopsy sampling can reduce, but not eliminate, this problem. Finally, the natural history of dysplasia is not well defined.

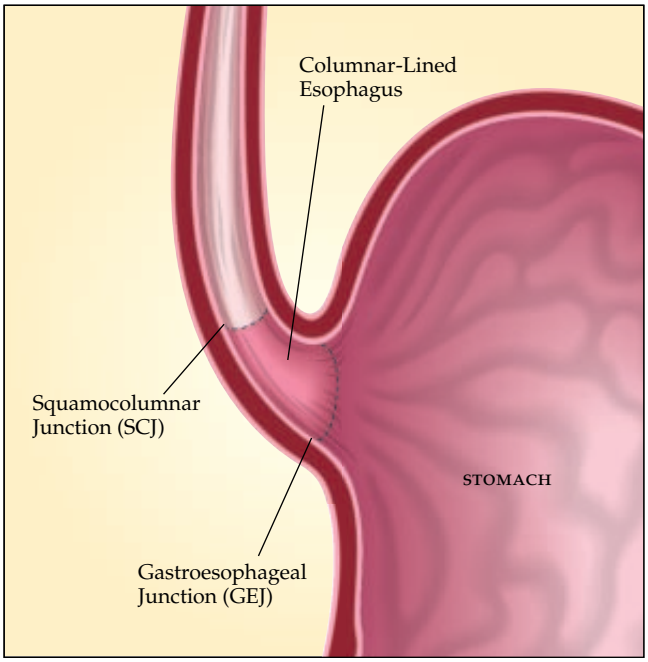


Figure 6 **In Barrett esophagus, columnar epithelium extends proximal to the gastroesophageal junction (the imaginary line at which the esophagus ends and the stomach begins, which corresponds to the most proximal extent of the gastric folds).**

Researchers have been searching for better alternatives to random biopsy sampling for dysplasia in Barrett esophagus. Other markers for malignant potential (e.g., flow cytometry and abnormalities in p53 expression) have been studied, as have endoscopic techniques to target dysplastic areas for biopsy. Despite some promising preliminary data, none of these tests and techniques has yet been shown to provide sufficient clinical information to justify its routine application in practice.

TREATMENT

Fit patients with verified high-grade dysplasia in Barrett esophagus have three management options: esophagectomy, endoscopic ablative therapy, and intensive surveillance (withholding invasive therapy until biopsy specimens show adenocarcinoma). Esophagectomy, the only therapy that clearly can prevent the progression from dysplasia to invasive cancer, is associated with operative mortality of 3% to 12% and with a 30% to 50% rate of serious operative complications. Endoscopic ablative therapies (e.g., laser photoablation and photodynamic therapy) use thermal or photochemical energy to destroy the metaplastic esophageal epithelium.[71] No study has shown that endoscopic ablation decreases the long-term risk for cancer development, so at present these therapies should be considered experimental. Although intensive surveillance for high-grade dysplasia (e.g., endoscopy every 3 months) has been endorsed as a management option by the American College of Gastroenterology, few published data directly support the safety and efficacy of the practice. The esophagus has an extensive lymphatic system and no confining serosa, features that can facilitate the spread of malignant cells and contribute to the

dismal prognosis for patients with esophageal cancer. The concern about surveillance for high-grade dysplasia is that by the time biopsy specimens reveal adenocarcinoma, the tumor already may be incurable because of systemic metastases.

Management Recommendations

A management algorithm for patients with Barrett esophagus has been developed [*see Figure 7*].[70] This strategy assumes that patients have had an initial endoscopic examination in which four-quadrant biopsy specimens are taken at intervals of 2 cm or less throughout the columnar-lined esophagus. If biopsy sampling during the initial endoscopic procedure is not adequate or if there is any question regarding the degree of dysplasia, endoscopy should be repeated to resolve these issues. If inflammation interferes with the histologic assessment of dysplasia, the patient should be treated with intensive antireflux therapy (e.g., a PPI administered at least twice daily) for 8 to 12 weeks before repeating the endoscopy. Data regarding the safety of intensive endoscopic surveillance for patients with high-grade dysplasia are limited, and available studies have involved primarily older patients. The clinician should be especially cautious in applying the results of these studies to the management of high-grade dysplasia in younger patients. Intensive endoscopic surveillance may be a valid alternative to immediate esophagectomy for older patients with high-grade dysplasia who can comply with the program. For patients who are too old, infirm, or unwilling to assume the risks of esophagectomy, endoscopic ablative therapy may be a reasonable alternative if the procedure is performed as part of a study protocol. Finally, the clinician should bear in mind that these recommendations have not been validated by studies demonstrating that this strategy prolongs survival or enhances quality of life.

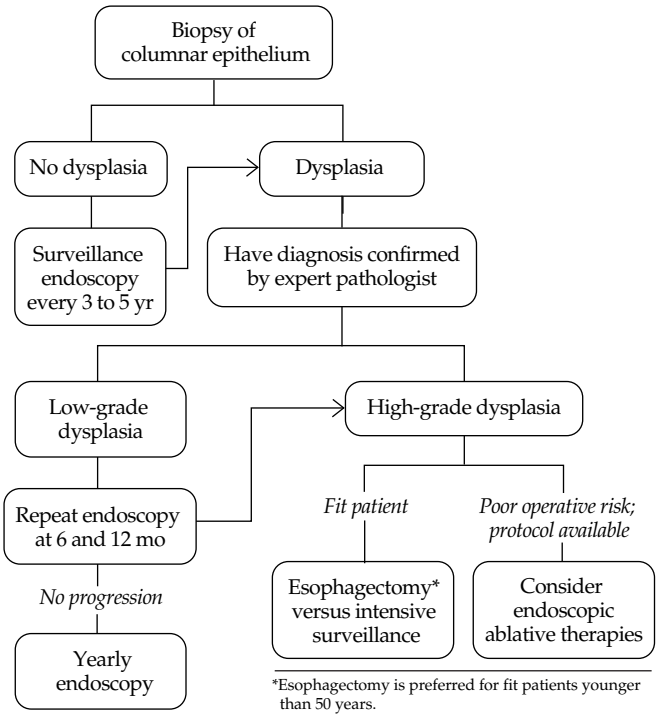


Figure 7 **Management of patients with Barrett esophagus.[73]**

Infectious Esophagitis

Most esophageal infections are caused by *Candida*, herpes simplex virus (HSV), or cytomegalovirus (CMV), alone or in combination.[6] These organisms rarely infect the esophagus of normal persons, but they often cause esophagitis in patients whose immune system has been compromised by AIDS, advanced malignancy, or the immunosuppressive drugs used to prevent rejection of organ transplants. Immune dysfunction that can accompany diabetes mellitus, alcoholism, and advanced age also may predispose to esophageal infection, especially by *Candida*. Antibiotic therapy that alters the normal microbial flora of the oropharynx and esophagus and corticosteroid therapy that suppresses immune function also can result in candidal esophagitis, as can abnormalities that delay the clearance of *Candida* from the esophagus, such as progressive systemic sclerosis (scleroderma), achalasia, and esophageal strictures.

The clinical manifestations of esophageal infections are similar, regardless of the pathogen. Dysphagia and odynophagia are the presenting symptoms in 60% to 95% of patients with infectious esophagitis, and weight loss is reported by 35%.[6] CMV esophagitis often is only one component of a generalized CMV infection, and 20% to 40% of patients with CMV esophagitis have systemic symptoms [*see* Cytomegalovirus Esophagitis, *below*]. In contrast, *Candida* and HSV esophagitis usually are not associated with infection in other organs, and systemic symptoms are uncommon. However, oral lesions (e.g., thrush and focal ulcerations) are found frequently in patients who have *Candida* and HSV esophagitis, but they are not found in patients with CMV esophagitis.

A surprising number of esophageal infections are asymptomatic. In published series, approximately one quarter of all cases of candidal esophagitis were discovered incidentally during radiographic or endoscopic examinations performed for the evaluation of extraesophageal symptoms.

CANDIDAL ESOPHAGITIS

Candida, a yeast that is part of normal oropharyngeal flora, is the most frequent cause of esophageal infections in immunocompromised patients. Most cases of candidal esophagitis are caused by *C. albicans*, although other candidal species, such as *C. tropicalis* and *C. glabrata*, occasionally infect the esophagus. Approximately 85% of patients with candidal esophagitis have oral thrush, and the combination of oral thrush and esophageal symptoms has a high positive predictive value for candidal esophagitis.

Typically, endoscopic evaluation of patients with candidal esophagitis reveals raised, white plaques that resemble cottage cheese clinging to the esophageal mucosa. On barium swallow, coating of the raised plaques and their interstices with barium gives the esophageal mucosa a characteristic irregular, shaggy appearance. Confirmation of the diagnosis requires the demonstration of budding yeast cells, hyphae, or pseudohyphae in brush cytology or biopsy specimens of the esophagus.

Several antifungal agents are available for the treatment of *Candida* infections. The decision regarding which agent to choose is influenced principally by the severity of the infection and the severity of the patient's immunocompromise. A patient who has a mild esophageal infection and minimal immunocompromise (e.g., a young patient who develops mild candidal esophagitis during a limited course of steroid therapy for asthma) often can be treated effectively with a topical anti-

fungal agent such as clotrimazole. In contrast, a patient with moderately severe candidal esophagitis and substantial immunocompromise (e.g., a patient with troublesome odynophagia and AIDS) usually requires the oral administration of a systemic antifungal agent such as fluconazole. Patients who have severe infection and profound immunocompromise (e.g., disseminated candidiasis or candidal esophagitis in the setting of severe granulocytopenia) generally require treatment with intravenous amphotericin B [*see 7:XXXVIII Mycotic Infections in the Compromised Host*].

herpes simplex virus esophagitis

Primary HSV infections of the oropharynx are common in the general population, and they result when oral mucous membranes or breaks in the facial skin are exposed to secretions from a person with an active HSV infection. The virus enters the nerves that supply the infected epithelium, where it remains in latent form after healing of the primary infection. The latent virus in the neurons can be reactivated and spread to epithelial cells through the nerve fibers. In immunocompetent persons, HSV reactivation commonly causes cold sores of the lips (herpes labialis). When reactivation of latent virus occurs in the setting of immunodeficiency, however, HSV can spread to involve the squamous epithelium of the oropharynx and esophagus.

The endoscopic findings in persons with HSV esophagitis vary with the duration of infection.[72] The earliest lesions are small (1 to 3 mm), rounded vesicles that usually involve the middle to distal esophagus. Sloughing of the vesicles results in small, sharply demarcated ulcers that have raised margins and a yellowish base. In severe cases, the small ulcers coalesce to form large ulcers that can be covered with dense exudates resembling candidal plaques.

Histologic diagnosis is best accomplished by examining biopsy and brush cytology specimens from the squamous epithelium at the edges of ulcerated areas. Specimens obtained from the ulcer base often contain only nonspecific granulation tissue and exudates. Typical histologic changes in HSV infection include multinucleated giant cells and intranuclear Cowdry type A inclusion bodies.

In immunocompetent patients, HSV esophagitis usually is a short-lived illness that may require no therapy other than supportive care and expectant management. HSV esophagitis often does not resolve spontaneously in immunocompromised patients, however, so such patients should receive systemic antiviral therapy. Acyclovir currently is the drug of choice for HSV infections of the esophagus.

cytomegalovirus esophagitis

CMV is a ubiquitous herpesvirus that usually is transmitted from person to person by exposure to infected secretions. The virus can also be transmitted through transfused blood that carries infected leukocytes or through transplanted infected organs. CMV can infect virtually any tissue in the body, and after recovery from the primary infection, evidence of latent CMV infection can be found in most organs. With the development of immunodeficiency, the latent virus can reactivate and cause esophagitis.[73] Immunocompromised patients can also develop CMV esophagitis during primary CMV infections. CMV esophagitis is extremely uncommon in immunocompetent persons.

Patients who have CMV esophagitis often have widespread CMV infection, with systemic symptoms such as fever, weight loss, nausea, vomiting, and diarrhea. CMV tends to cause discrete, shallow esophageal ulcerations that are very elongated (up to 15 cm in length) and surrounded by normal-appearing esophageal mucosa. Tissue sampling for histologic examination and culture is necessary to distinguish these giant CMV ulcerations from the giant idiopathic esophageal ulcerations that can be associated with HIV infection.

Histologic examination of cells infected with CMV reveals distinctive abnormalities that include cellular enlargement and inclusion bodies in both the nucleus and the cytoplasm. Although the virus is found most often in fibroblasts and endothelial cells, biopsy specimens from granulation tissue in the base of the esophageal ulcer have a higher yield on histology and culture than specimens from squamous epithelial cells at the edges of the ulcer. Unfortunately, no single test for CMV infection is highly sensitive. In a study of 14 bone marrow transplant recipients who developed CMV disease, for example, conventional and centrifugation cultures of endoscopic biopsy specimens identified the organism in only 57% of patients, and conventional histologic examination of the specimens revealed characteristic findings in only 30%.[74] For patients with negative test results, therefore, repeated diagnostic testing may be necessary if the suspicion of CMV infection is high. However, evidence of CMV infection is not proof of the presence of CMV disease. The mere identification of CMV in an inflamed esophagus does not establish that CMV is the cause of the inflammation.

CMV disease can respond to treatment with ganciclovir. Maintenance therapy with ganciclovir may be indicated for patients who have recurrences of CMV disease or who have a high risk of recurrence (e.g., patients with advanced AIDS). Prophylactic antiviral therapy is commonly recommended for recipients of solid-organ and bone marrow transplants.

esophageal disease in hiv infection

Within 2 to 3 weeks after primary exposure to HIV, some patients develop a self-limited, infectious mononucleosis–like illness with malaise, fever, myalgias, pharyngitis, and rash. This acute HIV seroconversion syndrome can be complicated by the development of esophageal ulcerations that cause odynophagia.[75] Endoscopically, the ulcers are typically multiple, round, 3 to 15 mm in diameter, well demarcated, and surrounded by normal-appearing esophageal mucosa. Usually, the ulcers heal and the symptoms of the acute HIV seroconversion syndrome resolve spontaneously within 2 weeks. Patients then may remain asymptomatic for years until the development of AIDS.

Symptoms of esophageal disease occur in 30% to 40% of AIDS patients.[76] Although the symptoms are usually from infections with *Candida*, HSV, or CMV, these patients can also have large esophageal ulcerations in which no pathogenic microorganism can be identified by culture or by histologic and immunohistochemical tests. Radiographically and endoscopically, HIV-associated idiopathic ulcerations of the esophagus closely resemble the large esophageal ulcerations caused by CMV. HIV-associated idiopathic esophageal ulcerations can be found in approximately 10% of patients with AIDS who complain of esophageal symptoms and in up to 40% of such patients who have discrete esophageal ulcerations on endoscopic examination.

HIV-associated idiopathic ulcerations generally do not respond to therapy with antimicrobial agents. Rather, patients with these lesions usually experience symptomatic relief and ulcer healing during treatment with systemic corticosteroids. Al-

though corticosteroid therapy entails substantial risk for patients who already are profoundly immunosuppressed, the treatment is surprisingly well tolerated in most cases. The injection of methylprednisolone through the endoscope directly into idiopathic ulcerations also has resulted in relief of esophageal symptoms in some cases, but experience with this treatment is limited. Finally, thalidomide, which has immunomodulatory effects, has been used successfully to treat idiopathic ulcerations.

Candida is by far the most common cause of esophageal infection in AIDS, and candidal esophagitis is found in more than 50% of AIDS patients who have esophageal symptoms. Although CMV and HSV esophagitis also occur commonly in AIDS patients, these viruses often are not the sole pathogens that can be identified in the inflamed esophagus. CMV usually is discovered in biopsy specimens from an esophagus that is also infected by *Candida*, and most patients with coexistent CMV and candidal esophagitis respond well to antifungal therapy alone. Consequently, authorities have recommended that patients with AIDS who have esophageal symptoms should be treated empirically with antifungal therapy, usually fluconazole; endoscopy is reserved for patients who fail to respond to empirical treatment.

Esophageal Cancer

EPIDEMIOLOGY

The two major histologic types of esophageal cancer, squamous cell carcinoma and adenocarcinoma, differ profoundly in their epidemiologic features.[77] Squamous cell carcinoma of the esophagus has a strong predilection for blacks and Asians, whereas esophageal adenocarcinoma is predominantly a disease of whites. In the United States, the incidence of squamous cell carcinoma of the esophagus is six times greater in African Americans than in whites, whereas esophageal adenocarcinoma is at least four times more frequent in whites than in African Americans. Worldwide, more than 90% of all esophageal cancers are squamous cell carcinomas; this tumor ranks among the world's 10 most frequent malignancies. Exceptionally high incidence rates of squamous cell carcinoma are found in the Transkei region of South Africa, in southern Brazil, in parts of northern France and Italy, and throughout an esophageal cancer belt that extends from the shores of the Caspian Sea of Iran across northern China. In the Henan province of China, the incidence of esophageal squamous cell carcinoma exceeds 100 per 100,000. In the United States, in contrast, the incidence of this tumor in the general population is less than 4 per 100,000. In most countries, cancer of the esophagus affects men two to four times more often than women.

RISK FACTORS

GERD and Barrett esophagus are the major risk factors for adenocarcinoma of the esophagus.[70] For squamous cell carcinoma, cigarette smoking and alcoholism are the major risk factors.[78] The combination of cigarette smoking and alcoholism appears to have a synergistic (rather than merely additive) effect in esophageal carcinogenesis, but the mechanism of this synergy is not known. Generalized malnutrition and a variety of specific nutritional deficiencies, including deficiencies in vitamin A, vitamin C, magnesium, selenium, and zinc, have been associated with squamous cell carcinoma. In contrast, obesity is a risk factor for adenocarcinoma. Carcinogens such as *N*-nitroso compounds can be formed from the nitrates and amines in pickled vegetables and cured meats, and ingestion of these foods has been linked to esophageal cancer. Regional practices such as opium smoking and the long-term ingestion of very hot foods and beverages may contribute to the pathogenesis of squamous cell cancers. Also, some high-incidence areas for squamous cell carcinoma have soils that are deficient in certain elements, such as molybdenum and zinc.

Local differences in endemic microflora have been proposed as underlying reasons for some of the regional variations in the incidence of esophageal squamous cell carcinoma. For example, the food and water in some high-incidence areas are contaminated with fungi and bacteria that promote the formation of *N*-nitroso compounds from dietary nitrates. The human papillomavirus (HPV) can infect squamous epithelial cells, and HPV infection has been implicated in the development of squamous cell carcinoma of the esophagus. In high-incidence regions for esophageal cancer, such as China and South Africa, researchers have found HPV DNA in more than 20% of squamous cell carcinomas. In low-incidence areas, such as the United States, however, esophageal tumors generally do not show evidence of HPV infection.

A number of medical conditions predispose to the development of esophageal squamous cell carcinoma. Patients with tylosis, a rare heritable disorder characterized by hyperkeratosis of the palms and soles, are at very high risk for development of the esophageal tumor. These patients have mutations in the tylosis esophageal cancer gene, a putative tumor suppressor gene located on the long arm of chromosome 17. Achalasia, lye stricture of the esophagus, and Plummer-Vinson, or Paterson-Kelly, syndrome also are risk factors for squamous cell cancers, perhaps because these conditions are associated with stasis of esophageal contents that leads to chronic inflammation of the mucosa. Squamous cell cancer of the esophagus is strongly associated with malignancies of the head, neck, and lungs, probably because these tumors share the strong risk factor of cigarette smoking. Finally, celiac sprue has been associated with esophageal cancer, for reasons that are not clear.

DIAGNOSIS

Clinical Features

Most patients with cancer of the esophagus present with dysphagia and weight loss. The dysphagia usually involves solid foods only and progresses rapidly in severity (over a period of weeks to months). Approximately 60% of patients who have esophageal adenocarcinoma have a long-standing history of GERD symptoms. Proximal esophageal tumors can invade the recurrent laryngeal nerve, causing vocal cord paralysis with hoarseness. The development of coughing associated with swallowing may indicate that the tumor has invaded the airway and caused an esophagobronchial fistula. Ulcerated tumors can cause odynophagia, and tumor necrosis occasionally causes esophageal hemorrhage. Local tumor invasion can cause chest pain, and metastatic disease can cause bone pain. Symptoms of esophageal cancer generally develop only when the tumor has grown to the extent that it has narrowed the lumen of the esophagus substantially, has invaded local structures, or has metastasized. Therefore, the presence of symptoms usually indicates advanced disease and a poor prognosis.

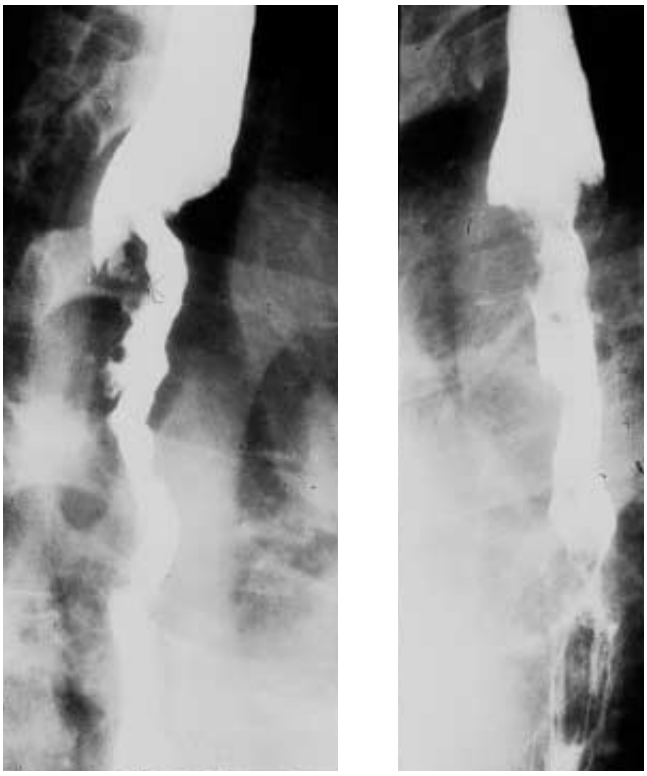


Figure 8 **Barium swallow showing an extensive cancer of the esophagus.**

Diagnostic Studies

Barium swallow and endoscopy Both barium swallow and endoscopy are useful for the evaluation of patients with esophageal cancer. Radiographic features that suggest malignancy include irregular borders and sharp angles [*see Figure 8*]. Endoscopically, esophageal cancers typically appear as nodular lesions that protrude into the lumen of the esophagus [*see Figure 9*]. In Asian countries where there is a high incidence of esophageal cancer, endoscopists often recognize early esophageal cancers that cause either slight elevations or shallow depressions in the mucosal surface. Staining of the esophagus with vital dyes such as toluidine blue or Lugol iodine (chromoendoscopy) can be useful for finding such early lesions during endoscopic evaluation. These superficial esophageal cancers are diagnosed infrequently in western countries.

Imaging studies Computed tomography of the chest and abdomen generally is recommended to assess the extent of disease within the chest and to look for metastases. However, the sensitivity and specificity of CT for determining the depth of esophageal tumor penetration (the T level) and the presence of regional lymph node metastases (the N status) are poor. Endoscopic ultrasonography (EUS) is superior to CT in this regard, accurately predicting the T level and N status in 70% to 80% of patients.

Invasive modalities In addition to EUS, invasive diagnostic modalities sometimes used for the staging of esophageal cancer include bronchoscopy, laparoscopy, thoracotomy, and thoracoscopy. There is little consensus regarding the need for these procedures in the routine evaluation of patients with esophageal cancer, and usage of the procedures varies widely among different institutions.

MANAGEMENT

Cancer of the esophagus usually is disseminated at the time of diagnosis, and because there is no treatment that reliably eradicates metastatic disease, cure is not possible in most cases. Furthermore, patients are often elderly, and many have severe comorbidities (e.g., malnutrition or pulmonary, cardiac, or liver disease) that further limit their treatment options. Initial treatment usually involves a choice between surgery, radiation therapy, chemotherapy, and some combination of these three modalities.[79,80] Squamous cell cancer and adenocarcinoma are treated similarly, with similarly poor survival rates. Despite recent advances in therapeutic options, overall cure rates for cancer of the esophagus remain below 10%.

Surgical therapy Esophagectomy, with or without lymphadenectomy, can provide immediate palliation of symptoms and, arguably, the best potential for cure of esophageal cancer. Mortality for esophagectomy ranges from 3% to more than 12%, and serious complications of the operation (e.g., pneumonia, atelectasis, arrhythmias, myocardial infarction, heart failure, wound infections, and anastomotic leaks) can be expected in 30% to 50% of patients. Cure rates vary widely among institutions. Prognostic factors include tumor stage and the number of positive lymph nodes. Surgery generally is not recommended for patients who have metastatic disease.

Radiation therapy The acute mortality of radiation therapy is low, and radiation can cover a wider treatment area than is practical with surgery (to eradicate local and regional disease). However, radiation therapy usually takes 2 to 8 weeks to complete; palliation can be delayed for weeks; there can be substantial radiation damage to surrounding normal tissues; and the overall cure rate is low. Trials of radiation therapy as the

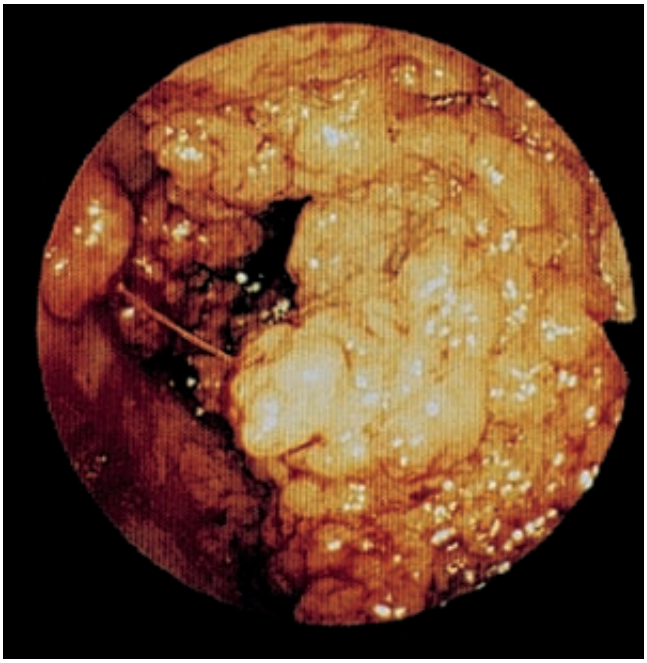


Figure 9 **Endoscopic photograph of an esophageal adenocarcinoma.**

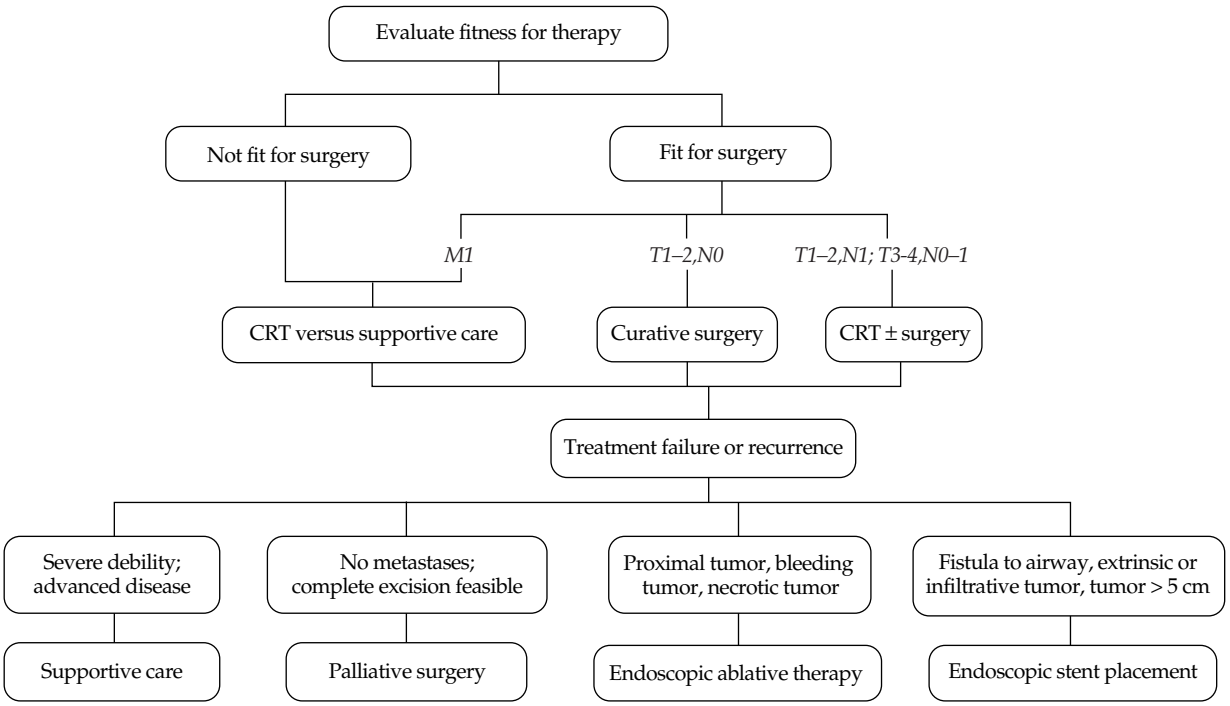


Figure 10 **Management of patients with cancer of the esophagus. (CRT—chemoradiotherapy)**

sole treatment modality for esophageal cancer have involved primarily patients with advanced squamous cell carcinomas that were deemed unresectable. Results appear to be comparable to surgery, but there are no randomized trials directly comparing radiation therapy alone with surgery alone.

Chemotherapy Chemotherapy has the potential to reach the disseminated disease that usually is present in symptomatic patients. Unfortunately, chemotherapy is associated with substantial morbidity and considerable mortality; it is often ineffective; and the tumor response, if any, is often brief. Studies of chemotherapy as the sole treatment modality have included primarily patients with unresectable tumors. Modern studies have used cisplatin-based regimens, and response rates appear to be better with combination regimens than with single agents. Chemotherapy alone does not appear to improve survival, however.

A number of studies have explored the role of radiation therapy or chemotherapy used either before (neoadjuvant) or after (adjuvant) definitive surgery for squamous cell carcinoma of the esophagus. Unfortunately, most randomized, controlled trials have shown no convincing benefit for neoadjuvant or adjuvant treatment with either radiation therapy or chemotherapy for patients with potentially resectable tumors. One recent trial showed a modest survival benefit for patients who received preoperative chemotherapy in a relatively low dose, however.[80]

Combination therapy Much recent interest has focused on the role of combining chemotherapy with radiation therapy (chemoradiotherapy) for esophageal cancer.[81] In some studies of patients treated with chemoradiotherapy followed by esophagectomy, complete histologic response (defined as no histologic evidence of tumor in the resected specimen) has been observed in almost 30% of cases. Complete histologic response is not tantamount to cure, however, and even complete responders frequently succumb to recurrent disease. Furthermore, chemoradiotherapy is associated with serious toxicity. Some randomized trials of chemoradiotherapy for patients with potentially resectable tumors have shown significant improvements in survival, whereas others have not. Consequently, the role of chemoradiotherapy remains unclear. Preliminary studies suggest that patients who have locally advanced tumors might benefit from preoperative chemoradiotherapy.

Palliative therapy Purely palliative therapies include esophageal dilatation and the placement of intraluminal stents. There also are palliative techniques designed to ablate the portion of the neoplasm that obstructs the esophageal lumen. These ablative therapies include endoscopic laser irradiation, the application of tumor probes that burn the neoplasm directly, the injection of caustic chemicals directly into the tumor body, and photodynamic therapy that uses photochemical energy to destroy the tumor.

Given that the optimal treatment for cancer of the esophagus is not clear, patients should be treated according to well-designed, established research protocols whenever possible. If the initial use of research protocols is not feasible, management should be individualized [*see Figure 10*]. After staging of the tumor that includes at least EUS and a CT scan of the chest and abdomen, the next step is to decide whether the patient is fit enough to undergo surgery. If surgery is not a viable option because of advanced age or comorbidity, primary therapy might include chemoradiation or supportive care alone. In general, surgery is not indicated for patients with metastatic disease. For tumors that do not invade beyond the muscularis propria and do not involve local lymph nodes, surgery appears to offer the best hope for cure. For lesions that are more advanced (i.e., with lymph node involvement or invasion to the esophageal adventitia and beyond), the choices for primary therapy include chemoradiation with or without surgery.

If these primary treatments fail or if the tumor recurs, there are a number of palliative treatment options. For patients who are severely debilitated and have advanced disease, the most humane option may be supportive care only, with careful attention to pain control. If there are no apparent metastases and complete excision of the tumor is possible, surgery can be considered for palliation. The other options are ablative therapies or stents. Stents may not provide good palliation for patients with proximal tumors, bleeding tumors, and necrotic tumors; ablative therapy may be preferable in these circumstances. Alternatively, ablative therapy has little to offer for a patient with an esophagobronchial fistula or for a patient with a tumor that is extrinsic or infiltrative. Also, ablative therapy may be difficult and time-consuming for patients with very long tumors. Stenting may be preferable in these circumstances.

The author has received grants for clinical research or educational activities from AstraZeneca, Janssen Pharmaceutica Products LP, TAP Pharmaceuticals, Inc., and Wyeth-Ayerst.

References

1. Spechler SJ: American Gastroenterological Association medical position statement on treatment of patients with dysphagia caused by benign disorders of the distal esophagus. Gastroenterology 117:229, 1999

2. Castell DO, Knuff TE, Brown FC, et al: Dysphagia. Gastroenterology 76:1015, 1979

3. Wilcox CM, Alexander LN, Clark WS: Localization of an obstructing esophageal lesion: is the patient accurate? Dig Dis Sci 40:2192, 1995

4. Cook IJ, Kahrilas PJ: AGA technical review on management of oropharyngeal dysphagia. Gastroenterology 116:455, 1999

5. Lapadula G, Muolo P, Semeraro F, et al: Esophageal motility disorders in the rheumatic diseases: a review of 150 patients. Clin Exp Rheumatol 12:515, 1994

6. Baehr PH, McDonald GB: Esophageal infections: risk factors, presentation, diagnosis, and treatment. Gastroenterology 106:509, 1994

7. Ott DJ, Richter JE, Chen YM, et al: Esophageal radiography and manometry: correlation in 172 patients with dysphagia. AJR Am J Roentgenol 149:307, 1987

8. Kahrilas PJ, Clouse RE, Hogan WJ: American Gastroenterological Association technical review on the clinical use of esophageal manometry. Gastroenterology 107:1865, 1994

9. Marks RD, Shukla M: Diagnosis and management of peptic esophageal strictures. Gastroenterologist 4:223, 1996

10. A comparison of omeprazole and ranitidine in the prevention of recurrence of benign esophageal stricture. The Restore Investigator Group. Gastroenterology 107:1312, 1994

11. DeVault KR: Lower esophageal (Schatzki's) ring: pathogenesis, diagnosis, and therapy. Dig Dis 14:323, 1996

12. Spechler SJ, Castell DO: Classification of oesophageal motility abnormalities. Gut 49:145, 2001

13. Wong RKH, Maydonovitch CL: Achalasia. The Esophagus, 2nd ed. Castell DO, Ed. Little, Brown, Boston, 1995, p 219

14. Spiess AE, Kahrilas PJ: Treating achalasia: from whalebone to laparoscope. JAMA 280:638, 1998

15. Csendes A, Braghetto I, Henriquez A, et al: Late results of a prospective randomised study comparing forceful dilatation and oesophagomyotomy in patients with achalasia. Gut 30:299, 1989

16. Pasricha PJ, Rai R, Ravich WJ, et al: Botulinum toxin for achalasia: long-term outcome and predictors of response. Gastroenterology 110:1410, 1996

17. Dalton CB, Castell DO, Richter JE: The changing faces of the nutcracker esophagus. Am J Gastroenterol 83:623, 1988

18. Richter JE, Bradley LA, Castell DO: Esophageal chest pain: current controversies in pathogenesis, diagnosis, and therapy. Ann Intern Med 110:66, 1989

19. Lock G, Holstege A, Lang B, et al: Gastrointestinal manifestations of progressive systemic sclerosis. Am J Gastroenterol 92:763, 1997

20. Leite LP, Johnston BT, Barrett J, et al: Ineffective esophageal motility (IEM): the primary finding in patients with nonspecific esophageal motility disorder. Dig Dis Sci 42:1859, 1997

21. Locke GR III, Talley NJ, Fett SL, et al: Prevalence and clinical spectrum of gastroesophageal reflux: a population-based study in Olmsted County, Minnesota. Gastroenterology 112:1448, 1997

22. Lagergren J, Bergström R, Lindgren A, et al: Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 340:825, 1999

23. Kahrilas PJ: Gastroesophageal reflux disease. JAMA 276:983, 1996

24. Holloway RH, Dent J: Pathophysiology of gastroesophageal reflux: lower esophageal sphincter dysfunction in gastroesophageal reflux disease. Gastroenterol Clin North Am 19:517, 1990

25. Dodds WJ, Dent J, Hogan WJ, et al: Mechanisms of gastroesophageal reflux in patients with reflux esophagitis. N Engl J Med 307:1547, 1982

26. Mittal RK, Holloway RH, Penagini R, et al: Transient lower esophageal sphincter relaxation. Gastroenterology 109:601, 1995

27. Mittal RK, Balaban DH: The esophagogastric junction. N Engl J Med 336:924, 1997

28. Sloan S, Rademaker AW, Kahrilas PJ: Determinants of gastroesophageal junction incompetence: hiatal hernia, lower esophageal sphincter, or both? Ann Intern Med 117:977, 1992

29. Kahrilas PJ, Lin S, Chen J, et al: The effect of hiatus hernia on gastro-oesophageal junction pressure. Gut 44:476, 1999

30. Kahrilas PJ, Shi G, Manka M, et al: Increased frequency of transient lower esophageal sphincter relaxation induced by gastric distention in reflux patients with hiatal hernia. Gastroenterology 118:688, 2000

31. Helm JF, Dodds WJ, Pele LR, et al: Effect of esophageal emptying and saliva on clearance of acid from the esophagus. N Engl J Med 310:284, 1984

32. Kahrilas PJ, Dodds WJ, Hogan WJ, et al: Esophageal peristaltic dysfunction in peptic esophagitis. Gastroenterology 91:897, 1986

33. Kahrilas PJ, Gupta RR: The effect of cigarette smoking on salivation and esophageal acid clearance. J Lab Clin Med 114:431, 1989

34. Lanas A, Hirschowitz BI: Significant role of aspirin use in patients with esophagitis. J Clin Gastroenterol 13:622, 1991

35. Cryer B, Spechler SJ: Effects of non-steroidal anti-inflammatory drugs (NSAIDs) on acid reflux in patients with gastroesophageal reflux disease (GERD). Gastroenterology 118:A862, 2000

36. O'Connor HJ: *Helicobacter pylori* and gastro-oesophageal reflux disease—clinical implications and management. Aliment Pharmacol Ther 13:117, 1999

37. Werdmuller BFM, Loffeld RJLF: *Helicobacter pylori* infection has no role in the pathogenesis of reflux esophagitis. Dig Dis Sci 42:103, 1997

38. Labenz J, Blum AL, Bayerdörffer E, et al: Curing *Helicobacter pylori* infection in patients with duodenal ulcer may provoke reflux esophagitis. Gastroenterology 112:1442, 1997

39. Chow WH, Blaser MJ, Blot WJ, et al: An inverse relation between *cagA*+ strains of *Helicobacter pylori* infection and risk of esophageal and gastric cardia adenocarcinoma. Cancer Res 58:588, 1998

40. Vicari JJ, Peek RM, Falk GW, et al: The seroprevalence of *cagA*-positive *Helicobacter pylori* strains in the spectrum of gastroesophageal reflux disease. Gastroenterology 115:50, 1998

41. Graham DY, Yamaoka Y: *H. pylori* and *cagA*: relationships with gastric cancer, duodenal ulcer, and reflux esophagitis and its complications. Helicobacter 3:145, 1998

42. Richter JE: Chest pain and gastroesophageal reflux disease. J Clin Gastroenterol 30 (3 suppl):S39, 2000

43. Jailwala JA, Shaker R: Oral and pharyngeal complications of gastroesophageal reflux disease: globus, dental erosions, and chronic sinusitis. J Clin Gastroenterol 30(3 suppl):S35, 2000

44. Harding SM, Richter JE: The role of gastroesophageal reflux in chronic cough and asthma. Chest 111:1389, 1997

45. Ormseth EJ, Wong RKH: Reflux laryngitis: pathophysiology, diagnosis and management. Am J Gastroenterol 94:2812, 1999

46. Richter JE, Peura D, Benjamin SB, et al: Efficacy of omeprazole for the treatment of symptomatic acid reflux disease without esophagitis. Arch Intern Med 160:1810, 2000

47. Spechler SJ: Epidemiology and natural history of gastro-oesophageal reflux disease. Digestion 51(suppl 1):24, 1992

48. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. The Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol 94:1434, 1999

49. DeVault KR, Castell DO: Guidelines for the diagnosis and treatment of gastroesophageal reflux disease. Arch Intern Med 155:2165, 1995

50. Eggleston A, Wigerinck A, Huijghebaert S, et al: Cost effectiveness of treatment for gastro-oesophageal reflux disease in clinical practice: a clinical database analysis. Gut 42:13, 1998

51. Inadomi JM, Jamal R, Murata GH, et al: Step-down management of gastroesophageal reflux disease. Gastroenterology 121:1095, 2001

52. Spechler SJ: Peptic ulcer disease and its complications. Sleisenger & Fordtran's Gastrointestinal and Liver Disease. Feldman M, Friedman LS, Sleisenger MH, Eds. WB Saunders Co, Philadelphia, 2002, p 747

53. Chiba N, De Gara CJ, Wilkinson JM, et al: Speed of healing and symptom relief in grade II to IV gastroesophageal reflux disease: a meta-analysis. Gastroenterology 112:1798, 1997

54. Castell DO, Kahrilas PJ, Richter JE, et al: Esomeprazole (40 mg) compared with lansoprazole (30 mg) in the treatment of erosive esophagitis. Am J Gastroenterol 97:575, 2002

55. Hetzel DJ, Dent J, Reed WD, et al: Healing and relapse of severe peptic esophagitis after treatment with omeprazole. Gastroenterology 95:903, 1988

56. Long-term omeprazole treatment in resistant gastroesophageal reflux disease: efficacy, safety, and influence on gastric mucosa. Long-Term Study Group. Gastroenterology 118:661, 2000

57. Dent J, Yeomans ND, Mackinnon M, et al: Omeprazole v ranitidine for prevention of relapse in reflux oesophagitis: a controlled double blind trial of their efficacy and safety. Gut 35:590, 1994

58. Rydberg L, Ruth M, Lundell L: Mechanism of action of antireflux procedures. Br J

Surg 86:405, 1999

59. Hinder RA, Libbey JS, Gorecki P, et al: Antireflux surgery: indications, preoperative evaluation, and outcome. Gastroenterol Clin North Am 28:987, 1999

60. Laine S, Rantala A, Gullichsen R, et al: Laparoscopic vs conventional Nissen fundoplication: a prospective randomized study. Surg Endosc 11:441, 1997

61. Laparoscopic or conventional Nissen fundoplication for gastro-oesophageal reflux disease: randomised clinical trial. The Netherlands Antireflux Surgery Study Group. Lancet 355:170, 2000

62. Spechler SJ: Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. N Engl J Med 326:786, 1992

63. Long-term management of gastro-oesophageal reflux disease with omeprazole or open antireflux surgery: results of a prospective, randomized clinical trial. The Nordic GORD Study Group. Eur J Gastroenterol Hepatol 12:879, 2000

64. DeMeester TR, Bonavina L, Albertucci M: Nissen fundoplication for gastroesophageal reflux disease: evaluation of primary repair in 100 consecutive patients. Ann Surg 204:9, 1986

65. Brand DL, Eastwood IR, Martin D, et al: Esophageal symptoms, manometry, and histology before and after antireflux surgery: a long-term follow-up study. Gastroenterology 76:1393, 1979

66. Spechler SJ, Lee E, Ahnen D, et al: Long-term outcome of medical and surgical treatments for gastroesophageal reflux disease: follow-up of a randomized controlled trial. JAMA 285:2331, 2001

67. Filipi CJ, Lehman GA, Rothstein RI, et al: Transoral, flexible endoscopic suturing for treatment of GERD: a multicenter trial. Gastrointest Endosc 53:416, 2001

68. Triadafilopoulos G, DiBaise JK, Nostrant TT, et al: The Stretta procedure for the treatment of GERD: 6 and 12 month follow-up of the U.S. open label trial. Gastrointest Endosc 55:149, 2002

69. Spechler SJ: Barrett's esophagus. N Engl J Med 346:836, 2002

70. Van den Boogert J, van Hillegersberg R, Siersema PD, et al: Endoscopic ablation therapy for Barrett's esophagus with high-grade dysplasia: a review. Am J Gastroenterol 94:1153, 1999

71. McBane RD, Gross JB Jr: Herpes esophagitis: clinical syndrome, endoscopic appearance, and diagnosis in 23 patients. Gastrointest Endosc 37:600, 1991

72. Meyers JD: Prevention and treatment of cytomegalovirus infection. Ann Rev Med 42:179, 1991

73. Hackman RC, Wolford JL, Gleaves CA, et al: Recognition and rapid diagnosis of upper gastrointestinal cytomegalovirus infection in marrow transplant recipients: a comparison of seven virologic methods. Transplantation 57:231, 1994

74. Rabeneck L, Popovic M, Gartner S, et al: Acute HIV infection with painful swallowing and esophageal ulcers. JAMA 263:2318, 1990

75. Laine L, Bonacini M: Esophageal disease in human immunodeficiency virus infection. Arch Intern Med 154:1577, 1994

76. Blot WJ, McLaughlin JK: The changing epidemiology of esophageal cancer. Semin Oncol 26(suppl 15):2, 1999

77. Stoner G, Gupta A: Etiology and chemoprevention of esophageal squamous cell carcinoma. Carcinogenesis 22:1737, 2001

78. Siewert JR, Stein HJ, Sendler A, et al: Esophageal cancer: clinical management. Gastrointestinal Oncology: Principles and Practice. Kelsen D, Daly JM, Kern SE, et al, Eds. Lippincott Williams & Wilkins, Philadelphia, 2002, p 261

79. Surgical resection with or without preoperative chemotherapy in oesophageal cancer: a randomised controlled trial. Medical Research Council Oesophageal Cancer Working Group. Lancet 359:1727, 2002

80. Bosset JF, Gignoux M, Triboulet JP, et al: Chemoradiotherapy followed by surgery compared with surgery alone in squamous-cell cancer of the esophagus. N Engl J Med 337:161, 1997

81. Spechler SJ: The role of gastric carditis in metaplasia and neoplasia at the gastroesophageal junction. Gastroenterology 117:218, 1999

Acknowledgment

Figures 4, 8, and 9 American Gastroenterological Association.

Figure 6 Seward Hung.

II PEPTIC ULCER DISEASES

MARK FELDMAN, M.D.

Definition

Peptic ulcers are holes in the inner lining of the gastrointestinal (GI) tract that are attributed to exposure of the mucosa to gastric acid and pepsin. Peptic ulcers extend through the mucosa and the muscularis mucosae, a thin layer of smooth muscle separating the mucosa from the deeper submucosa, muscularis propria, and serosa. Most peptic ulcers are round or oval, but some are linear, triangular, or irregular in shape. Ulcers have depth when viewed through an endoscope. Typically, only a single ulcer is present. An erosion is a focal loss of superficial epithelial cells and glands, without extension through the muscularis mucosae. On endoscopy, erosions appear as breaks in the mucosal lining without depth. At the other extreme, a peptic ulcer may burr itself entirely through the wall of the GI tract, thus connecting the GI lumen with the peritoneal cavity (perforated ulcer), a solid organ such as the pancreas (penetrating ulcer), or another hollow organ such as the intestine or bile duct (fistulizing ulcer).

Epidemiology

In the United States, peptic ulcer disease affects 10% of men and 4% of women at some time in their lives. The incidence is influenced by age (older persons are more susceptible than younger persons) and gender (males are more susceptible than females). Because ulcer disease is often recurrent, its prevalence exceeds its incidence. Eradication of *Helicobacter pylori* from the stomach markedly reduces recurrence of ulcer disease. With the widespread use of treatment regimens for *H. pylori*, the prevalence of peptic ulcer is decreasing in the United States. Reinfection with *H. pylori* remains an uncommon event in the United States (approximately one reinfection per 100 patients a year).

Pathogenesis and Etiologic Factors

The normal stomach and duodenum are able to resist autodigestion by acid-pepsin. However, high rates of acid-pepsin secretion or impaired mucosal resistance factors, such as prostaglandin deficiency, can predispose to duodenal ulcer formation (typically in the most proximal part of the duodenum, the bulb) or to gastric ulcer formation (typically in the most distal part of the stomach, the antrum).

On rare occasions, peptic ulcers occur in the second, third, or fourth portion of the duodenum (postbulbar ulcer) or even in the proximal jejunum. Ordinarily, alkaline secretions from the duodenum, biliary tract, and pancreas neutralize gastric acid in the duodenum, but high rates of gastric acid secretion (e.g., in Zollinger-Ellison syndrome) can overwhelm these endogenous alkaline secretions and lead to postbulbar or jejunal ulcerations.

In patients with pathologic amounts of gastroesophageal reflux of acid-pepsin and in many patients with gastric acid hypersecretion, erosions and ulcers may develop in the lower esophagus [*see 4:I Esophageal Disorders*]. Peptic ulcers may also occur where acid and pepsin are secreted heterotopically, such as in a congenital ileal (Meckel) diverticulum.

Regardless of location and etiology, chronic peptic ulcers are similar pathologically. In addition to the focal loss of mucosal epithelial cells, these ulcers have four characteristic layers at their base: fibrinoid necrosis, exudate, granulation tissue, and a fibrous scar, the deepest layer. A layer of granulation tissue and fibrosis may be absent in acute ulcers that occur in settings of serious trauma or severe surgical or medical illnesses [*see* Acute Stress Ulcers, *below*].

Why a peptic ulcer is such a focal lesion is unclear. Although peptic ulcers require the presence of acid-pepsin, acid-pepsin alone is only rarely sufficient to produce an ulcer—such as in Zollinger-Ellison syndrome, in which marked gastric hypersecretion is present. In the majority of patients, there must be another predisposing factor, such as *H. pylori* infection of the stomach,[1,2] use of nonsteroidal anti-inflammatory drugs (NSAIDs), or stress [*see Figure 1*].

H. PYLORI ULCERS

The prevalence of *H. pylori* infection of the stomach is much higher in duodenal ulcer patients and, to a somewhat lesser extent, in gastric ulcer patients than in age-matched control subjects [*see Figure 2*].[3] In addition, cure of *H. pylori* infection with antimicrobial therapy markedly reduces recurrences of duodenal and gastric ulcers.[4,5] The correlation of *H. pylori* infection with peptic ulcers is not consistent, however. Duodenal ulcers develop in some infected persons and gastric ulcers in others, but most infected persons experience no ulcers at all. Patients with duodenal ulcers tend more often to be infected with *cagA*-positive strains than do *H. pylori*–infected patients without ulcer,[6] but how this is mediated is not clear.

The etiologic mechanism linking *H. pylori* infection with ulcer development is not yet absolutely established, for the following reasons: (1) voluntary ingestion of *H. pylori* led to gastric *H. pylori* infection and to gastritis but not to ulcers; (2) duodenal or gastric ulcers develop in only 10% to 20% of individuals with *H. pylori* gastritis, implying that only certain persons with additional genetic, anatomic, physiologic, or environmental risk factors are predisposed to ulcers or that only certain *H. pylori* strains are ulcerogenic; (3) *H. pylori* induces diffuse inflammation in the stomach, yet the strongest link between *H. pylori* and peptic ulcer is with focal duodenal bulbar ulcer; and (4) gastric *H. pylori* infection is as common in women as in men, yet duodenal ulcer is two to three times less common in women. Currently, *H. pylori* can be considered the most important risk factor for duodenal and gastric ulcers, but it is clear that the mere presence of *H. pylori* in the stomach is not sufficient to cause peptic ulcers [*see Figure 1a*].

NSAID ULCERS

The ulcerogenicity of NSAIDs has been established experimentally by exposing animals, human volunteers, and patients to these drugs. Experimental studies have been corroborated by numerous case-control studies and autopsy studies. Unlike *H. pylori*–related peptic ulcers, which more often occur in the duodenal bulb, NSAID ulcers typically occur in the stomach. A gastric or duodenal ulcer associated with NSAID use is classified as a peptic ulcer, and it usually heals with potent acid antisecretory therapy, even if NSAID use is continued. NSAIDs can also cause ulcers in the jejunum, ileum, or colon, areas where there is little

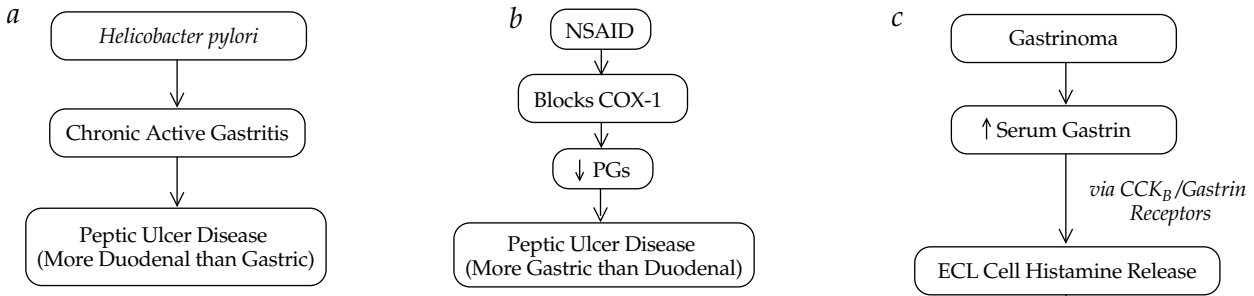


Figure 1 **Etiopathogenesis of peptic ulcers. (*a*) *Helicobacter pylori* induces a diffuse, chronic, active superficial gastritis, usually throughout the stomach. Exactly how this infectious gastritis results in peptic ulcer disease is unknown. (*b*) Nonselective nonsteroidal anti-inflammatory drugs (NSAIDs) block cyclooxygenase-1 (COX-1) to reduce the amount of gastroduodenal prostaglandins (PGs) synthesized from their precursor, arachidonic acid. COX-2 selective NSAIDs produce a lesser reduction in prostaglandins and are associated with fewer peptic ulcers than nonselective COX-1– or COX-2–inhibiting NSAIDs. (*c*) Gastrinoma cells in the pancreas or duodenum secrete large amounts of gastrin into the circulation. Elevated serum gastrin levels promote the release of histamine by acting on receptors for cholecystokinin$_B$ (CCK_B) and for gastrin, which are located on gastric enterochromaffin-like (ECL) cells. Histamine acts on H_2 receptors on parietal and chief cells to augment hydrochloric acid (HCl) and pepsin secretion.**

or no acid-pepsin. These ulcers are not actually peptic ulcers.

Although the pathogenesis of NSAID ulcers is multifactorial, by far the most important mechanism appears to be inhibition of cyclooxygenase-1 (COX-1), the rate-limiting enzyme in GI prostaglandin synthesis [*see Figure 1b*]. Prostaglandins normally protect the GI mucosa from damage by maintaining mucosal blood flow and increasing mucosal secretion of mucus and bicarbonate. Blockade of COX-1 activity by NSAIDs reduces prostaglandin synthesis and thus lowers GI mucosal blood flow and secretion of mucus and bicarbonate. Evidence continues to accumulate that in NSAID users, *H. pylori* gastritis produces about a twofold increase in the risk of gastroduodenal ulcer formation.[7] Moreover, clinically symptomatic peptic ulcers occur much less commonly if *H. pylori* gastritis is treated before the start of long-term NSAID therapy.[8] Clinically diagnosed peptic ulcers will develop in approximately 2% to 4% of persons taking NSAIDs per year of exposure. The extent to which the damaging effects of NSAIDs on the stomach are topical rather than systemic is unclear. Many NSAIDs, such as aspirin, are acidic and thus nonionized in the acidic stomach, where they can be absorbed and initiate gastric mucosal damage. However, NSAIDs (e.g., ketorolac) given by parenteral injection and aspirin given transdermally are ulcerogenic, as are so-called NSAID prodrugs, such as sulindac and nabumetone (neither drug inhibits gastric prostaglandins until it is metabolized to its active form after GI absorption). Evidence suggests that acute mucosal damage by NSAIDs (i.e., hemorrhages and erosions, but seldom ulcers) is mainly caused by the topical damaging effects of NSAIDs. Chronic ulcer formation, often with complications such as bleeding and perforation, is mainly the result of the systemic effect of NSAIDs on prostaglandin synthesis by the GI mucosa.

Epidemiologic studies suggest that NSAIDs vary in their ability to cause ulcers,[9] but this issue is complicated by the difficulty of comparing equipotent doses of NSAIDs. All prescription or over-the-counter NSAIDs should be considered ulcerogenic, with the risk of ulcer dependent on dosages and other patient-related factors, particularly advanced age and previous ulcer history. Even low doses of aspirin used for prophylaxis of cardiovascular disease (75 to 325 mg/day) are ulcerogenic.[10] Neither buffering of aspirin nor enteric coating appears to reduce the incidence of clinically detected ulcer formation.[11] Nonacetylated salicylates such as salicylsalicylic acid (salsalate) do not block COX-1 and are not ulcerogenic. Epidemiologic studies indicate that the greatest risk of NSAID ulcers is early in the course of treatment (between day 7 and day 30 after initiation), with the risk decreasing thereafter.

Most NSAIDs, including aspirin, block both COX-1 and COX-2. Unlike COX-1, COX-2 is induced and expressed at inflammatory sites but not in the normal GI tract.[12] Selective COX-2 inhibitors (coxibs) such as celecoxib and valdecoxib

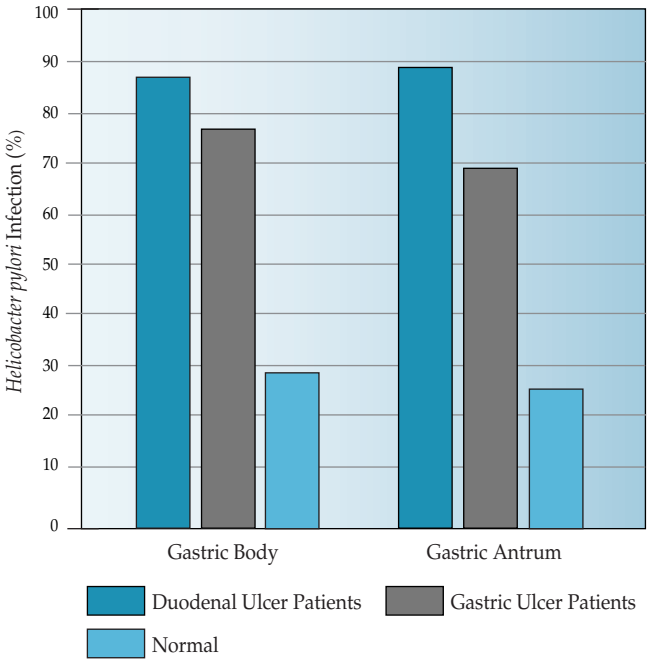


Figure 2 **Prevalence of *H. pylori* infection in either the gastric body or the gastric antrum, as assessed by endoscopic biopsy and mucosal histology. Normal control subjects were matched by age and sex to patients with duodenal ulcers and to patients with gastric ulcers. None of the ulcer patients were receiving aspirin or NSAIDs.[3]**

are analgesic and anti-inflammatory but seem to cause less GI ulcer formation than currently available cyclooxygenase inhibitors when used in recommended doses.[13,14] Whether routine use of coxibs in place of nonselective NSAIDs will be cost-effective remains to be demonstrated.

Corticosteroids, which block COX-2 but not COX-1,[12] are not ulcerogenic when used alone, although they impair healing of preexisting ulcers. When corticosteroids are used in combination with NSAIDs, the risk of ulcer formation is much greater than when NSAIDs are used alone.

ulcers in gastrinoma or other hypersecretory states

A gastrinoma is an endocrine tumor of the pancreas or duodenum (usually malignant) consisting of gastrin (G) cells. Gastrinoma causes less than 1% of all peptic ulcers. Peptic ulcers develop in 95% of patients with gastrinoma (Zollinger-Ellison syndrome); ulcers occur most commonly in the duodenal bulb but are also seen in the postbulbar duodenum, jejunum, lower esophagus, and stomach. Multiple ulcers are present in up to 25% of cases of Zollinger-Ellison syndrome.

Patients with a gastrinoma have high circulating levels of gastrin [*see Figure 1c*], which acts on receptors for cholecystokinin$_B$ (CCK_B) and gastrin located on enterochromaffin-like (ECL) cells within the mucosa of the gastric body. ECL cells then release histamine, which acts on H_2 receptors present on the membrane of neighboring parietal cells to stimulate (via an adenylate cyclase–cyclic adenosine monophosphate [cAMP]–mediated pathway) the secretion of hydrochloric acid by a unique proton pump, the H^+,K^+–ATPase pump. Of less physiologic importance, gastrin also acts directly on CCK_B/gastrin receptors on parietal cells, increasing cytosol calcium levels in the parietal cells.

Hypergastrinemia in Zollinger-Ellison syndrome results in a continuous high rate of secretion of hydrochloric acid and pepsin, even under basal (fasting) conditions. These secretions overwhelm the buffering and neutralizing capacity of food and upper digestive secretions, as well as mucosal defense factors. Peptic ulceration results, and in many cases, diarrhea (with or without malabsorption) occurs.

Approximately 20% to 30% of patients with gastrinomas have features suggesting a multiple endocrine neoplasia type I (MEN I) syndrome, such as hypercalcemia secondary to hyperparathyroidism, a pituitary adenoma, or both. MEN I is inherited as an autosomal dominant disorder.

Some patients with duodenal ulcer have marked acid hypersecretion but normal serum gastrin levels. A few of these patients have hyperhistaminemia caused by systemic mastocytosis or chronic basophilic leukemia. However, the majority of patients have no known reason for the acid hypersecretion (idiopathic basal acid hypersecretion), although some are infected with *H. pylori*. Eradication of *H. pylori* in these individuals may reduce basal acid hypersecretion.

idiopathic ulcers

In the United States, up to 20% of cases of chronic gastric and duodenal ulcers occur in patients who have no evidence of *H. pylori* infection, deny taking NSAIDs, and have normal serum gastrin concentrations. These ulcers are referred to as idiopathic peptic ulcers. Some patients with this disorder may be taking NSAIDs surreptitiously or are unaware that they are taking these drugs. In others, emotional stress, perhaps associated with gastric acid hypersecretion, may be a contributing factor.[15] Cigarette smoking is also a risk factor for peptic ulcers.

acute stress ulcers

Acute gastroduodenal erosions and ulcers are very common in patients with serious medical and surgical conditions.[16] Such conditions include severe head injury (Cushing ulcers); burn injury (Curling ulcers); major surgical procedures; and life-threatening illnesses such as septic shock, respiratory failure requiring mechanical ventilation, hepatic failure, renal failure, and multiorgan failure. Unlike peptic ulcers, stress ulcers are typically asymptomatic, rarely causing dyspepsia or epigastric pain. Approximately 10% to 25% of patients with acute stress ulcers experience painless upper GI bleeding of variable severity. Bleeding may manifest itself in the intensive care unit as a dark (so-called coffee-ground) or bloody nasogastric aspirate, as a declining hematocrit, as an increasing transfusion requirement, or as unexplained hypotension.

The pathogenesis of stress ulcers is not well understood. The common denominator seems to be tissue hypoxia and acidosis, precipitated by mucosal vasoconstriction and ischemia. Systemic hypoxia, metabolic acidosis, anemia, and reduced cardiac output often are contributing factors. Once mucosal hypoxia develops, mucosal defense factors are impaired and the cells lining the stomach and duodenum become vulnerable to damage by acid-pepsin. Acute stress ulcers have become less common because of the routine use of effective prophylactic medications in patients at high risk for this condition (see below).

cameron ulcers

Linear gastric erosions that occur in a hiatus hernia are known as Cameron ulcers.[16] The erosions are thought to be related either to traumatic injury of the stomach by the surrounding diaphragm or to mucosal ischemia at the point where the stomach herniates through the diaphragm. Like acute stress ulcers, Cameron ulcers tend to present as bleeding without dyspepsia. Both acute and chronic GI blood loss are possible outcomes of Cameron ulcers.

Diagnosis

clinical manifestations

Peptic ulcers produce a variety of symptoms but none specific for the disease. Also, symptoms of duodenal ulcer are indistinguishable from those of gastric ulcer. Patients with uncomplicated ulcers typically experience mild to moderate abdominal pain, usually in the epigastrium. However, the pain may be localized to the left or right upper quadrant of the abdomen, to the lower chest (subxiphoid or substernal), the midabdomen, or the back. The pain is often gnawing or burning. It may occur in the middle of the night; rarely, it occurs upon first awakening in the morning. Discomfort is typically relieved by food or an antacid.

Severe pain or a rapid increase in pain suggests an ulcer complication (e.g., perforation or penetration) or another diagnosis (e.g., acute pancreatitis). Associated dyspeptic symptoms include nausea, bloating, heartburn, and belching. Although vomiting may occur with uncomplicated peptic ulcers and may temporarily relieve pain, repeated vomiting suggests an ulcer complication (e.g., gastric outlet obstruction) or another diagnosis (e.g., intestinal obstruction).

Peptic ulcers are the most common cause of acute upper GI bleeding. Therefore, hematemesis, melena, or both, even in a patient with no history of ulcer and no dyspeptic symptoms, should suggest the possibility of a bleeding peptic ulcer. Patients

who develop ulcers while taking prescription or over-the-counter NSAIDs or low (cardiovascular) doses of aspirin often have no history of ulcerlike pain. Other patients with bleeding ulcers will have experienced dyspeptic symptoms for the preceding days or weeks, only to have these symptoms wane when bleeding ensues.

In addition to a review of the patient's symptoms and ulcer risk factors (particularly NSAID use and smoking), a family history should be obtained. A family history of ulcer can usually be attributed to within-family infection by *H. pylori*, to NSAID use, or to smoking. However, a family history of ulcer, hyperparathyroidism, kidney stones, or endocrine tumor should alert the physician to the possibility of gastrinoma (Zollinger-Ellison syndrome), with or without autosomal dominant MEN I syndrome.

PHYSICAL EXAMINATION FINDINGS

In uncomplicated peptic ulcer disease, the examination is generally normal. The presence of epigastric tenderness does not distinguish dyspepsia caused by peptic ulcer from other types of dyspepsia.

Patients who have complicated ulcers often have tachycardia and hypotension, which are exaggerated when the patient assumes an upright position. These findings may indicate a bleeding ulcer, a perforated ulcer with peritonitis, or an obstructing ulcer with protracted vomiting and volume depletion. Pulse and blood pressure measurements may give misleading information about the extent of volume contraction if the patient has preexisting hypertension, has cardiovascular disease, or is receiving medication that can affect these parameters (e.g., a beta-adrenergic blocker or a calcium channel blocker). Fever and tachypnea suggest ulcer perforation with peritonitis.

Special attention should be given to the patient's mental status, skin and mucous membranes, heart and lungs, and, of course, abdomen and rectum. Involuntary guarding, rigidity, rebound tenderness, and a paucity or absence of bowel sounds suggests ulcer perforation with peritonitis. These findings may be less prominent or even absent in the very young, the elderly, and patients on corticosteroids or analgesics. Abdominal distention suggests gastric outlet obstruction or ileus. In a patient who has not eaten in 6 hours, a splashing sound over the stomach when the body is shaken (succussion) suggests gastric outlet obstruction or delayed gastric emptying caused by ileus. Melena or a positive fecal occult blood test suggests ulcer bleeding. Hematochezia or maroon-colored stool may be present if bleeding is voluminous and intestinal transit is rapid. Detection of melena, hematochezia, or maroon-colored stool should prompt placement of a nasogastric tube to obtain an aspirate of gastric contents. If this aspirate is grossly bloody, the diagnosis of upper GI bleeding is confirmed and the likelihood of a bleeding ulcer is increased.

LABORATORY STUDIES

Laboratory results are normal in most patients with uncomplicated ulcer. A complete blood count should be done if blood loss or ulcer perforation is suspected; and serum electrolytes, blood urea nitrogen (BUN), and serum creatinine should be measured if the patient has poor oral intake, nausea, or vomiting. An elevated serum calcium level suggests the possibility of hyperparathyroidism and MEN I with Zollinger-Ellison syndrome, but the pretest probability of this condition is too low in patients presenting with ulcerlike symptoms to recommend routine measurement of serum calcium. If the patient has a

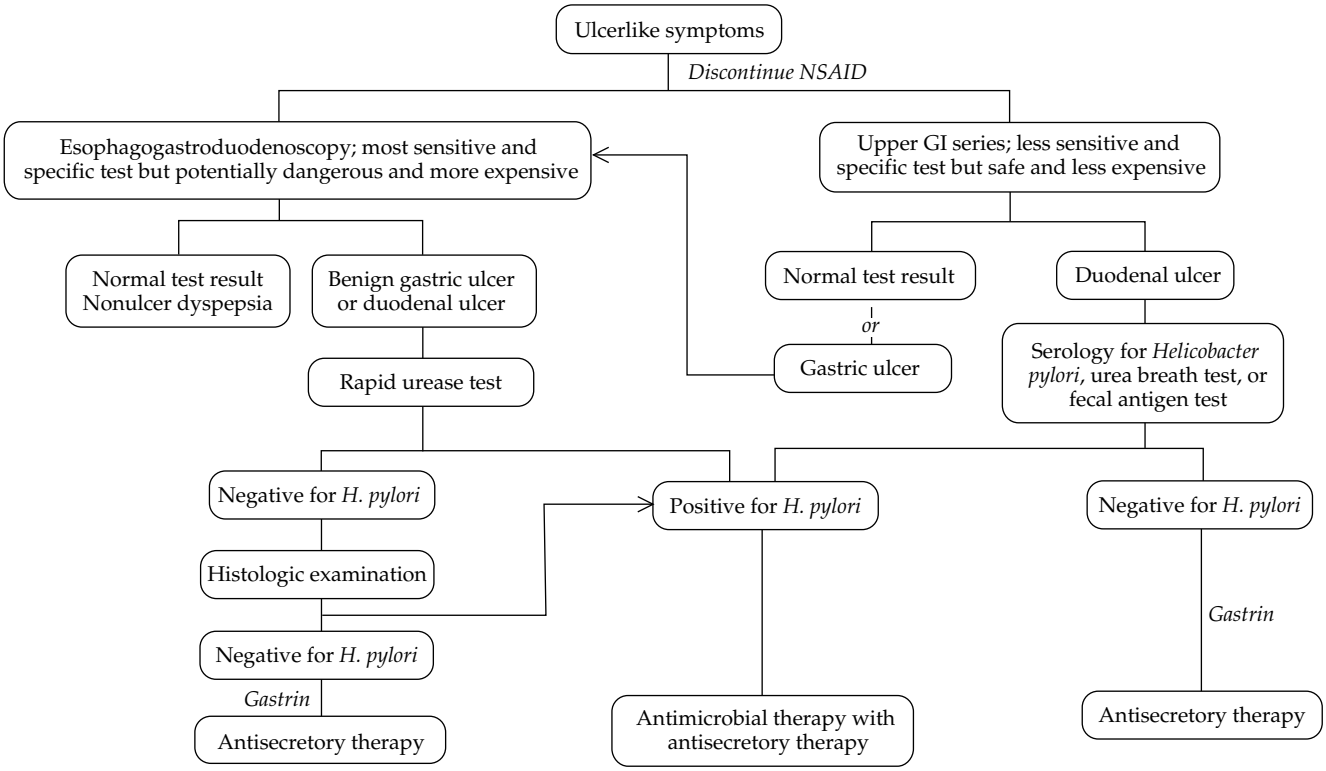


Figure 3 **Approach to a patient with new and undiagnosed ulcerlike symptoms refractory to a trial of antisecretory therapy with an H_2 receptor blocker or a proton pump inhibitor at customary doses or a patient with recurrent ulcerlike symptoms when the antisecretory therapy is stopped.**

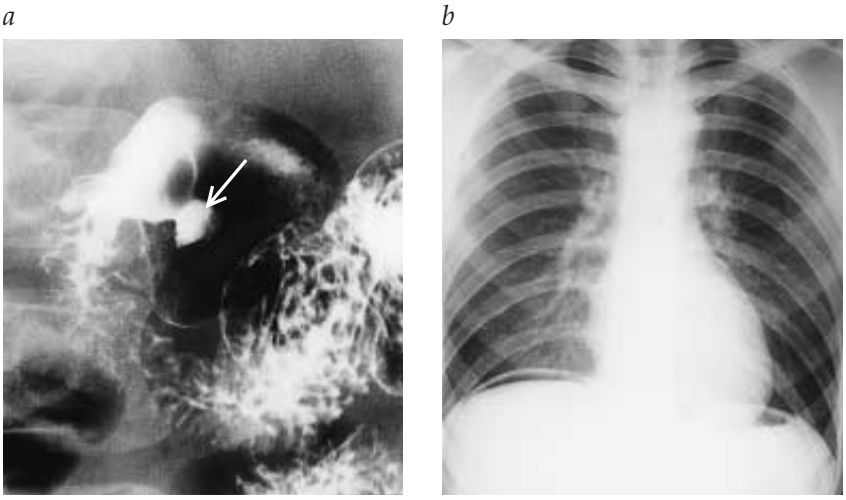


Figure 4 **(*a*) Upper GI series in which double contrast (barium and air) is used, showing rounded collection of barium in an ulcer (arrow) in the duodenal bulb of a patient presenting with dyspepsia (uncomplicated duodenal ulcer). (*b*) Upright chest x-ray showing air beneath the right hemidiaphragm (pneumoperitoneum) of a patient presenting with an acute abdomen caused by a perforated duodenal ulcer.**

strong family history of ulcer disease or of renal stones or has a personal history of renal stones, measurement of serum calcium is warranted, as is measurement of fasting serum gastrin once ulcer disease is confirmed and other causes of ulcer are excluded. If the calcium level is elevated, a serum parathyroid hormone measurement should be ordered.

Patients with complicated ulcers often have significant laboratory abnormalities, but these abnormalities are not specific for ulcer disease. Patients with bleeding ulcers have anemia and may have leukocytosis. The red cell indices (e.g., mean corpuscular volume) are typically normal. In the first several hours after an acute ulcer bleed, the hemoglobin concentration will not completely reflect the severity of the blood loss until compensatory hemodilution occurs or until intravenous fluids such as isotonic saline are administered. Thus, the pulse rate and blood pressure in the supine and upright positions are better initial indicators of extent of blood loss than are red cell counts. Patients with bleeding ulcers typically have azotemia, with ratios of BUN to serum creatinine concentrations exceeding 20:1, resulting from digestion and intestinal absorption of nitrogenous blood components in concert with reduced renal perfusion.

In patients with perforated ulcers and peritonitis, exudation of plasma into the peritoneal cavity (so-called third space) may result in an increased hemoglobin concentration from hemoconcentration. The presence of leukocytosis, elevated band forms, or leukopenia should raise suspicion of intra-abdominal sepsis. Lactic acidosis with an increased anion gap may ensue as a consequence of a sepsis syndrome or hypovolemia.

Patients with gastric outlet obstruction typically exhibit a hypokalemic, hypochloremic metabolic alkalosis. If volume loss is extreme, a coexistent metabolic lactic acidosis with an increased anion gap may be present, which may cause an elevated serum bicarbonate level to drop toward normal or even to low levels. Likewise, mild to moderate hyponatremia often develops in patients with vomiting from gastric outlet obstruction. Prerenal azotemia and a BUN–serum creatinine ratio greater than 20:1 are typical.

IMAGING STUDIES

Although ulcer disease can be suggested by history, physical examination, and laboratory studies, none of these has sufficient specificity to confirm the diagnosis. Ulcers are diagnosed endoscopically, radiologically, or surgically. Once an ulcer is diagnosed, additional studies can help in determining the cause of the ulcer (e.g., *H. pylori* infection, NSAID use, gastrinoma, or cancer masquerading as benign ulcer).

Endoscopy

Endoscopy is the most accurate way to diagnose a peptic ulcer [*see Figure 3*]. Most patients require local anesthesia of the pharynx and conscious sedation with an intravenous agent such as midazolam. The advantages of endoscopy are its nearly 100% specificity (rare false positives), greater than 90% sensitivity, portability (i.e., it can be performed in the intensive care unit, emergency department, or operating room), and ability to obtain tissue samples to help determine the etiology of the ulcer. The disadvantages of endoscopy are its cost and its potential for serious side effects. The most serious complications of endoscopy are respiratory depression and perforation of the GI tract. When a bleeding or obstructing ulcer is suspected, the stomach should be intubated and emptied with a large-bore tube before endoscopy to decrease the possibility of bronchopulmonary aspiration of gastric contents and to facilitate endoscopic visualization of mucosal lesions. Endoscopy is contraindicated in cases of suspected ulcer perforation.

Radiology

Despite having a lower sensitivity and specificity than endoscopy, an upper GI series using barium and air (double contrast) may be favored by primary care physicians and patients over referral for endoscopy for suspected uncomplicated ulcer. An upper GI series offers lower cost, wider availability, and fewer complications [*see Figure 3*]. However, for troublesome and undiagnosed dyspepsia, an upper GI series may be superfluous, because a normal result will often necessitate endoscopy (endoscopy is more sensitive than radiography) and because an upper GI series showing a gastric ulcer will also necessitate endoscopy for biopsy of the ulcer to exclude gastric malignancy. In many patients, only a finding of a duodenal bulbar ulcer on an upper GI series will preclude endoscopy [*see Figure 4*].

Plain films of the abdomen, abdominal sonography, and computed tomographic scans may be helpful in patients presenting with suspected complicated ulcers, particularly perforated or obstructing ulcers. Upright chest x-rays of a patient with a perforated ulcer may show free intraperitoneal air [*see Figure 4b*], typically beneath the right hemidiaphragm. When

plain films are negative or equivocal, pneumoperitoneum may be diagnosed by abdominal sonography or CT scan. Such studies should be performed only if the diagnosis of perforation is unclear; if physical signs of peritonitis are obvious, the patient should be referred to a surgeon. Patients with gastric outlet obstruction may have an enlarged stomach with old food debris visible on plain film of the abdomen, upper GI series, abdominal sonography, or CT scan.

SURGICAL DIAGNOSIS

Certain patients will not have ulcers diagnosed until surgery is performed. Such patients include those presenting with an acute abdomen, in whom the diagnosis of perforated ulcer is made at exploratory laparotomy; those presenting with copious upper GI bleeding, in whom it is difficult for the endoscopist to visualize and treat the ulcer; and those with an obstructing ulcer who have a pinpoint pylorus or a duodenal stricture that prevents passage of the endoscope beyond the stenosis.

Tests to Establish the Etiology of the Ulcer

ENDOSCOPIC TESTS

The endoscopist can take a biopsy sample of the stomach of an ulcer patient to determine whether *H. pylori* organisms are present [*see Figure 5*]. *H. pylori* organisms contain abundant amounts of urease, which splits urea into carbon dioxide and ammonia. If the biopsy sample is placed on a urea-containing medium that also contains a pH-sensitive dye, a change in color indicates that ammonia is being produced. This so-called rapid urease test has a high sensitivity and specificity (> 90%) for *H. pylori*. If the rapid urease test is negative, a separate biopsy specimen should be sent to a pathology laboratory in formalin for histology. *H. pylori* can be detected with routine hematoxylin and eosin stains [*see Figure 5*] or, if necessary, by special stains. Moreover, the presence of diffuse, active, chronic gastritis is highly suggestive of *H. pylori* infection, and its absence excludes *H. pylori* infection.

Another useful endoscopic procedure is to obtain multiple biopsies from the edges and the base of the ulcer to exclude malignancy. This is routinely done in cases of gastric ulcer because 2% to 4% of benign-appearing gastric ulcers are in actuality an ulcer within a malignancy, usually an adenocarcinoma. Duodenal ulcers need not be biopsied unless the ulcer is located in a mass distal to the duodenal bulb.

Endoscopy may also demonstrate a neuroendocrine tumor, compatible with a gastrinoma on special stains. Such a tumor is usually located in the proximal duodenum.

In an ulcer patient with a negative rapid urease test and no *H. pylori*–related gastritis or gastric malignancy on histology, further history regarding NSAID use should be obtained from the patient or the patient's family. Many patients with NSAID-related ulcers have erosions, subepithelial hemorrhages, or both, which clue the endoscopist to the possibility of occult or surreptitious NSAID use; these lesions may occur with or without gastric or duodenal ulcers.[7]

SEROLOGIC TESTS

A number of serum antibody tests for *H. pylori* are available that have a greater than 90% sensitivity and specificity if the patient has not yet received therapy for *H. pylori*.[17] In patients with active ulcers diagnosed by radiology or surgery in whom gastric tissue is not available, *H. pylori* serology can confirm infection with high accuracy [*see Figure 3*].

In ulcer patients with no evidence of *H. pylori* infection or NSAID use, the fasting serum gastrin concentration should be measured [*see Figure 3*] to screen for gastrinoma (Zollinger-Ellison syndrome). If the serum gastrin concentration is greater than 1,000 pg/ml in a patient with duodenal ulcer, the diagnosis of gastrinoma is confirmed. A modest elevation in fasting serum gastrin concentration (> 150 pg/ml but < 1,000 pg/ml) is suggestive of gastrinoma, but a provocative test should be performed using intravenous secretin (2 IU/kg as a bolus).[18] A rise in serum gastrin concentration of more than 200 pg/ml after secretin administration has a greater than 90% sensitivity and specificity for gastrinoma. Because achlorhydria can produce marked hypergastrinemia as a result of the loss of negative feedback of gastric acid on gastrin release, basal acid output or pH should be measured to confirm that the stomach secretes acid in ulcer patients with fasting hypergastrinemia. The combination of achlorhydria, hypergastrinemia, and duodenal ulcer is exceedingly rare, whereas the combination of achlorhydria, hypergastrinemia, and gastric ulcer is sometimes encountered and should suggest gastric adenocarcinoma or NSAID use. A fasting gastric pH measurement will almost invariably distinguish gastrinoma (pH 1 to 2) from achlorhydria (pH 6 to 8), unless the pa-

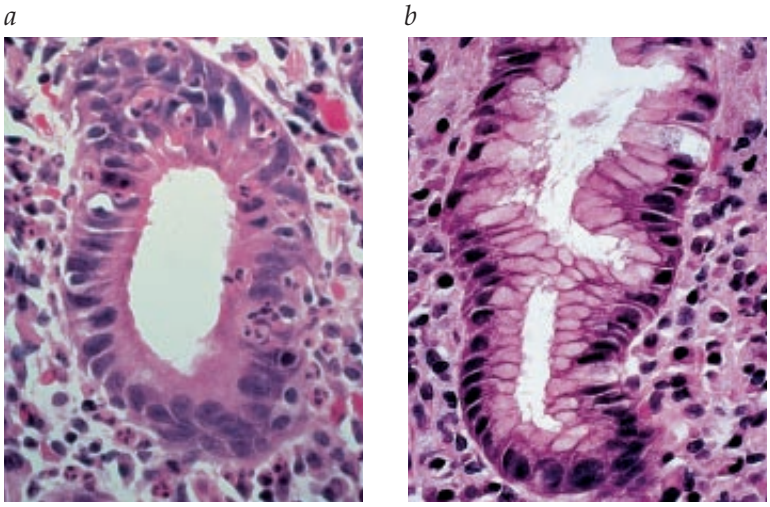


Figure 5 **Gastric biopsy samples stained with hematoxylin and eosin demonstrating (*a*) chronic active gastritis with a few *H. pylori* organisms faintly seen in the lumen of a gland and (*b*) chronic active gastritis with *H. pylori* organisms more abundant.**

Table 1 Differential Diagnosis of Peptic Ulcer Disease

Presentation	*Diagnosis*
Suspected uncomplicated ulcer	Nonulcer dyspepsia, gastroesophageal reflux, biliary colic, pancreatic disease, angina pectoris, gastric cancer
Bleeding ulcer	Varices, Mallory-Weiss tear, esophagitis, vascular lesion (arteriovenous malformation, Dieulafoy lesion, angiodysplasia)
Perforated ulcer	Appendicitis, pancreatitis, cholecystitis, spontaneous bacterial peritonitis, bowel ischemia or infarction, diverticulitis
Penetrating ulcer	Pancreatitis, muscle strain, herniated vertebral disk, ureteral stone
Fistulizing ulcer	Gallstones, GI malignancy, Crohn disease, intra-abdominal abscess

tient has received a potent acid antisecretory agent before pH measurement.

Measurement of serum thromboxane B_2 (platelet COX-1 activity) has been used in research laboratories to demonstrate occult or surreptitious NSAID use.[19] However, this assay is not widely available.

BREATH TESTS

A noninvasive method for detecting *H. pylori* in the stomach, the urea breath test, begins with oral ingestion of urea that has been labeled with carbon-13 (^{13}C) or carbon-14 (^{14}C). If *H. pylori*, with its abundant urease, is present in the stomach, the labeled urea will be rapidly converted to $^{13}CO_2$ or $^{14}CO_2$, which can be detected in breath samples collected during the first 30 to 60 minutes after urea ingestion. Sensitivity and specificity of the breath test are comparable to those of serology.[16] In a patient for whom there is no clinical indication for endoscopy, a urea breath test is an alternative to serology for documenting *H. pylori* infection [*see Figure 3*]. However, because proton pump inhibitors can suppress *H. pylori* without eradicating it, use of these drugs should be avoided for 2 weeks before the urea breath test is administered, to minimize false negative results.

Because serology is quicker, it is preferred to breath testing for initial diagnosis. Breath testing is more useful than serology in diagnosing failure of eradication of *H. pylori* or reinfection in patients who were previously treated for *H. pylori* infection, because the serology will usually remain positive for several months even after successful treatment.[20]

FECAL ANTIGEN TEST

Stool testing for *H. pylori* antigen compares favorably with urea breath tests.[21] Like breath testing, stool testing can distinguish current infection (antigen present in stool) from past infection (antigen not present in stool).

Differential Diagnosis

The most common disorder confused with uncomplicated peptic ulcer is nonulcer dyspepsia; the most serious GI disorder confused with uncomplicated peptic ulcer is gastric cancer [*see 12:VI Pancreatic, Gastric, and Other Gastrointestinal Cancers*].

Nonulcer, or functional, dyspepsia is a symptom complex similar to that experienced by patients with peptic ulcers. However, no ulcers or other lesions are visible on endoscopy. Nonulcer dyspepsia is a heterogeneous, poorly understood group of disorders. *H. pylori* gastritis probably causes dyspepsia in a few of these patients, and many physicians treat all dyspeptic patients who are infected with *H. pylori*. The cost-effectiveness of this approach has not been established, however; studies show that eradication of *H. pylori* is unlikely to relieve symptoms of nonulcer dyspepsia.[22-24] Many patients with nonulcer dyspepsia appear to suffer from a dysmotility of the upper GI tract that is akin to irritable bowel syndrome of the lower GI tract. Such individuals may complain of abdominal fullness, postprandial bloating, early satiety, and nausea, all suggestive of delayed gastric emptying. In some of these patients, gastric prokinetic agents such as domperidone or metoclopramide may help relieve symptoms.

Complicated ulcers may be confused with a variety of disorders. These include both intra-abdominal and musculoskeletal processes [*see Table 1*].

Treatment

The goals of ulcer therapy are rapid relief of symptoms; healing the ulcer; preventing ulcer recurrences; and reducing ulcer-related complications, morbidity (including the need for endoscopic therapy or surgery), and mortality. The general strategy in a patient with an ulcer should be to treat complications aggressively if present; to determine the etiology of the ulcer; to discontinue NSAID use if possible; to eradicate *H. pylori* infection if present or strongly suspected, even if other risk factors (e.g., NSAID use) are also present; and to use acid antisecretory therapy to heal the ulcer if *H. pylori* infection is not present. Smoking cessation should be encouraged. If duodenal ulcer is diagnosed by endoscopy, rapid urease testing of endoscopically obtained gastric biopsy samples, with or without histologic examination, should reliably establish the presence or absence of *H. pylori*. If duodenal ulcer is diagnosed by x-ray, then a serologic, urea breath, or fecal antigen test to diagnose *H. pylori* infection is recommended before treating the patient for *H. pylori*.

TREATMENT OF UNCOMPLICATED DUODENAL ULCERS

H. pylori–*Related Duodenal Ulcer*

Duodenal ulcer associated with *H. pylori* infection should be treated with antimicrobial therapy because successful therapy is associated with markedly reduced ulcer recurrences [*see Figure 6*][4,5] Antimicrobial therapy is usually empirical rather than based on results of culture and in vitro antimicrobial sensitivity testing. No single antimicrobial agent has an acceptably high success rate against *H. pylori*. Combinations of antimicrobial agents are required, and some regimens that have been approved by the Food and Drug Administration can be recommended [*see Table 2*].

H. pylori has adapted to the acidic stomach, and potent acid antisecretory agents facilitate eradication of *H. pylori* by antimicrobial agents. Bismuth compounds, like proton pump inhibitors, suppress the growth of *H. pylori* but usually do not by themselves eradicate it from the stomach. For this reason, they are frequently employed together with antibiotics.

Antimicrobial agents with activity against *H. pylori* include metronidazole, tetracycline, amoxicillin, and clarithromycin. Most popular are 10- to 14-day regimens, although 7-day courses

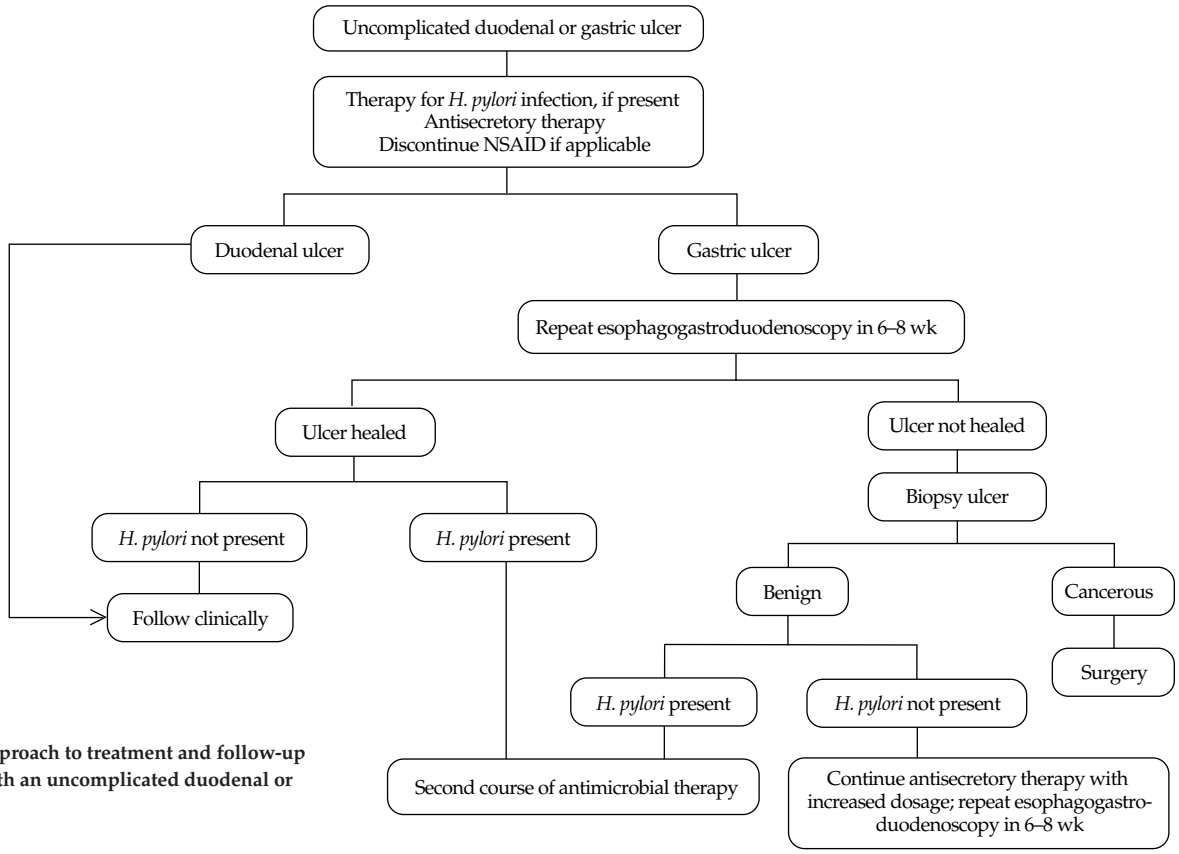


Figure 6 **Approach to treatment and follow-up of a patient with an uncomplicated duodenal or gastric ulcer.**

may be effective and are especially favored in Europe. A 2-week course of a three-drug regimen that includes a proton pump inhibitor, clarithromycin, and amoxicillin has a success rate approaching 90%. The major causes of treatment failure are poor compliance with the regimen and clarithromycin resistance; the latter occurs in around 10% of current strains and is increasing with more macrolide use in the population. Metronidazole resistance occurs in 30% to 40% of strains. However, unlike resistance to clarithromycin, which is usually absolute, resistance to metronidazole is relative and can be overcome in some patients.

If a patient carefully complies with one of the clarithromycin-based regimens yet the treatment fails, clarithromycin resistance is likely. In such cases, the retreatment regimen should not include clarithromycin. Most physicians choose a regimen consisting of metronidazole, tetracycline, bismuth (e.g., Pepto Bismol), and a proton pump inhibitor or H_2 receptor blocker. Because of the frequency of metronidazole resistance, other antimicrobials with activity against *H. pylori* are being used more often.[25-28] These agents include azithromycin, the quinolones norfloxacin and levofloxacin, and rifabutin [*see Table 3*]. In one study, a 7-day rescue treatment for persistent *H. pylori* infection using the proton pump inhibitor rabeprazole plus rifabutin and levofloxacin had a 95% success rate, as did a four-drug regimen consisting of rabeprazole, bismuth subcitrate, metronidazole, and tetracycline.[29]

It may also be possible to predict resistance to clarithromycin or to metronidazole by taking a careful history to look for prior exposure to these drugs. Such a history might help the physician choose a first-line regimen that will be more likely to be successful.

Side effects of *H. pylori*–directed therapy are not uncommon but are generally mild. Physicians should be aware of potential drug-drug interactions if the patient is receiving other medications. If the patient has an active, symptomatic ulcer, an antisecretory drug should be continued at a reduced (standard) dosage for 2 to 5 weeks after completion of antimicrobial agents.

After a patient has completed a course of ulcer therapy for an *H. pylori*–related uncomplicated duodenal ulcer, it is acceptable

Table 2 Selected Clarithromycin-Based Regimens to Eradicate *Helicobacter pylori*

Esomeprazole, amoxicillin, clarithromycin (EAC)
Esomeprazole, 40 mg b.i.d. for 10 days; then 40 mg q.d. for 18 days if an active ulcer is present
Amoxicillin, 1 g b.i.d. for 10 days
Clarithromycin, 500 mg b.i.d. or t.i.d. for 10 days
Lansoprazole, amoxicillin, clarithromycin (LAC)
Lansoprazole, 30 mg b.i.d. for 10–14 days; then 15 mg q.d. for 14–18 days if an active ulcer is present
Amoxicillin, 1 g b.i.d. for 10–14 days
Clarithromycin, 500 mg b.i.d. for 10–14 days
Omeprazole, amoxicillin, clarithromycin (OAC)
Omeprazole, 20 mg b.i.d. for 10 days; then 20 mg q.d. for 18 days if an active ulcer is present
Amoxicillin, 1 g b.i.d. for 10–14 days
Clarithromycin, 500 mg b.i.d. for 10 days
Rabeprazole, amoxicillin, clarithromycin (RAC)
Rabeprazole, 20 mg b.i.d. for 7 days; then 20 mg q.d. for 21 days if an active ulcer is present
Amoxicillin, 1 g b.i.d. for 7 days
Clarithromycin, 500 mg b.i.d. for 7 days

Table 3 Additional Antimicrobial Agents with Activity against *Helicobacter pylori*

*Agent**	*Dosage*	*Comments*
Azithromycin	500 mg q.d.	Combined with either amoxicillin or metronidazole, plus a proton pump inhibitor, for 3–7 days
Norfloxacin	400 mg b.i.d.	Combined with a proton pump inhibitor for 14 days
Levofloxacin	500 mg q.d.	Combined with either amoxicillin or metronidazole, plus a proton pump inhibitor, for 7 days
Rifabutin	300 mg q.d.	Combined with amoxicillin and a proton pump inhibitor for 7–10 days

*These agents are marketed in the United States, but they are not yet approved for *H. pylori* therapy.

to follow the patient clinically without confirming eradication, because most compliant patients will be successfully cured of their *H. pylori* infection [*see Figure 6*]. A patient with an *H. pylori*–related duodenal ulcer that does not recur symptomatically within 2 years after antimicrobial therapy is probably cured. Serology has often reverted to negative by this time.[20]

Those in whom recurrent ulcer symptoms develop during the first 2 years after therapy should be assessed either by endoscopy (for ulcer recurrence and for *H. pylori* persistence or reinfection) or by a urea breath test or fecal antigen test. The most common cause of recurrent ulceration in patients treated for *H. pylori*–related duodenal ulcer is failure to eradicate the organism. Retreatment of these patients is indicated [*see* Treatment of Intractable Duodenal Ulcers or Gastric Ulcers, *below*]. Rarer causes of duodenal ulcer recurrence include an acid hypersecretory state (e.g., Zollinger-Ellison syndrome), NSAID use, and reinfection with *H. pylori*.

H. pylori–*Negative Duodenal Ulcer*

In a duodenal ulcer patient who is *H. pylori* negative, the physician should consider NSAID use and gastrinoma (Zollinger-Ellison syndrome).[30] Patients with duodenal ulcer who are taking NSAIDs should discontinue the NSAID, if possible. At the same time, an acid antisecretory drug should be administered for 4 to 8 weeks [*see Table 4*]. The anticipated healing rate with this regimen is 85% to 95%.

Table 4 FDA-Approved Antisecretory Drugs for Active Peptic Ulcer Disease*

Class	*Drugs*	*Dosage*	*Drug Interactions*‡
H_2 receptor blockers†	Cimetidine	800 mg h.s. *or* 400 mg b.i.d.	Warfarin, theophylline, phenytoin, benzodiazepines, itraconazole, ketoconazole, atazanavir, cefpodoxime, cefditoren, gefitinib, memantine, metformin, and many others
	Ranitidine	300 mg h.s. *or* 150 mg b.i.d.	Warfarin, atazanavir, itraconazole, ketoconazole, cefpodoxime, cefditoren, enoxacin, gefitinib, memantine, metformin, tolazoline
	Nizatidine	300 mg h.s. *or* 150 mg b.i.d.	Atazanavir, itraconazole, ketoconazole, cefpodoxime, cefditoren
	Famotidine	40 mg h.s. *or* 20 mg b.i.d.	Atazanavir, itraconazole, ketoconazole, cefpodoxime, cefditoren
Proton pump inhibitors§	Omeprazole	20 mg q.d., a.c.	Benzodiazepines, ampicillin, atazanavir, digoxin, iron, itraconazole, ketoconazole, voriconazole, methotrexate, tacrolimus
	Lansoprazole	15 mg q.d., a.c.	Ampicillin, atazanavir, digoxin, iron, itraconazole, ketoconazole
	Esomeprazole	40 mg q.d., a.c.	Atazanavir, digoxin, iron, ketoconazole
	Rabeprazole	20 mg q.d., a.c.	Warfarin, ampicillin, atazanavir, digoxin, iron, itraconazole, ketoconazole
	Pantoprazole	40 mg q.d., a.c.	Ampicillin, atazanavir, iron, itraconazole, ketoconazole

*Patients with gastrinoma (Zollinger-Ellison syndrome) will usually require much higher dosages of antisecretory drugs than listed here.
†Use for 4–8 wk in the treatment of duodenal ulcer and 6–12 wk in the treatment of gastric ulcer. Duodenal ulcers that do not heal by 8 wk and gastric ulcers that do not heal by 12 wk are considered intractable. Dosage of H_2 receptor blockers should be reduced in patients with renal failure.
‡Micromedex Health Care Services. Most of these drug interactions are minor and not clinically relevant; nevertheless, caution is advised.
§Omeprazole and lansoprazole are approved for gastric ulcers; omeprazole, lansoprazole, and rabeprazole are approved for duodenal ulcers.

Table 5 Treatment and Prevention of Peptic Ulcers

Type of Ulcer	*Treatment*	*Prevention*	*Comments*
Helicobacter pylori–related ulcers	Antibiotics [*see Tables 2 and 3*] ± antisecretory agents [*see Table 4*]	None needed if *H. pylori* eradicated	Highly cost-effective; document healing in gastric ulcer; document *H. pylori* eradication in complicated duodenal or gastric ulcer and in intractable duodenal or gastric ulcer
NSAID-related ulcers	Antisecretory agents (proton pump inhibitors have greater efficacy than H_2 receptor blockers) [*see Table 4*] Discontinue NSAID use, if possible	Misoprostol (600–800 µg/day) or proton pump inhibitor (e.g., omeprazole, 20–40 mg q.d., or lansoprazole, 15–30 mg q.d.) along with an NSAID	Diarrhea may limit compliance in patients treated with misoprostol; avoid misoprostol during pregnancy (abortifacient); proton pump inhibitors are not yet approved by the FDA for prevention of NSAID-related ulcers
Ulcers associated with Zollinger-Ellison syndrome and other hypersecretory states	High-dose proton pump inhibitor	Proton pump inhibitor, adjusted to keep basal acid output < 5–10 mEq/hr	Consider exploratory laparotomy (guided by abdominal imaging studies) to remove easily resectable gastrinomas, if feasible; consider MEN I syndrome (present in 20%–30% of cases)[21]
Idiopathic ulcers	H_2 receptor blocker or proton pump inhibitor	Nocturnal H_2 receptor blocker or A.M. proton pump inhibitor	Parietal cell vagotomy for intractable duodenal ulcer and antrectomy for intractable gastric ulcer
Stress ulcers (ICU)	I.V. H_2 receptor blocker or proton pump inhibitor (e.g., pantoprazole) ?Angiography ?Surgery	I.V. H_2 receptor blocker or proton pump inhibitor; intragastric sucralfate; or intragastric antacid	Maintain pH above 4 with H_2 receptor blocker, proton pump inhibitor or antacid; continuous I.V. infusion is superior to I.V. boluses of antisecretory drug
Cameron ulcers (linear gastric erosions in a hiatal hernia)	Iron salts; packed red cell transfusions; endoscopic hemostasis ?Angiography ?Surgery	Hiatal hernia repair, laparoscopic or open	Roles of H_2 receptor blockers and proton pump inhibitors are unproved

MEN I—multiple endocrine neoplasia type I NSAID—nonsteroidal anti-inflammatory drug

Patients with duodenal ulcer as part of the Zollinger-Ellison syndrome should be managed initially with a high dose of a proton pump inhibitor, followed by a maintenance dose guided by gastric acid measurements. If there is no evidence of hepatic metastasis on abdominal CT scan, then exploratory laparotomy for gastrinoma resection, with or without parietal cell vagotomy, is warranted.[31] Radionuclide scintigraphy with octreotide, an analogue of somatostatin, is a highly sensitive and specific preoperative test for detecting and staging gastrinoma, as is endoscopic ultrasonography.

The vast majority of duodenal ulcers, regardless of cause, heal after 8 weeks of antisecretory therapy with a proton pump inhibitor or an H_2 receptor blocker. Antacids are often prescribed as needed to relieve ulcer symptoms.

In rare cases of idiopathic duodenal ulcer, it is prudent that, after the ulcer has been healed by an acid antisecretory agent, the patient be placed on a maintenance dose of an H_2 receptor blocker given at bedtime to reduce ulcer recurrences. Proton pump inhibitors are also effective in preventing duodenal ulcer recurrences. It is not necessary to confirm duodenal ulcer healing by endoscopy or x-ray before reducing the antisecretory drug dose to a maintenance level.

TREATMENT OF UNCOMPLICATED GASTRIC ULCERS

H. pylori–*Related Gastric Ulcer*

Gastric ulcer associated with *H. pylori* should be treated with antibiotics [*see Table 2*]. Because they are larger than duodenal ulcers, gastric ulcers take longer to heal. Thus, after antibiotic administration, the patient should be treated with an acid antisecretory agent [*see Tables 3 and 4*] for an additional 4 to 8 weeks. Patients with gastric ulcers should be followed endoscopically until complete healing has been achieved so that an ulcerated gastric cancer is not missed. Gastric biopsies should be obtained during follow-up endoscopy to determine whether eradication of *H. pylori* has occurred [*see Figure 6*]. Patients with a history of an uncomplicated gastric ulcer that is currently quiescent should be screened for *H. pylori* infection, and if the result is positive, they should be treated for *H. pylori* infection to prevent ulcer recurrences.

NSAID-Related Gastric Ulcer

The therapy for an active NSAID-related gastric ulcer is administration of a proton pump inhibitor [*see Tables 4 and 5*], as well as discontinuance of the NSAID. Healing rates with H_2 receptor blockers are nearly as high as with proton pump inhibitors if the NSAID can be stopped.

TREATMENT OF INTRACTABLE DUODENAL ULCERS OR GASTRIC ULCERS

Ulcers That Fail to Heal

Ulcers refractory to pharmacotherapy are rare, and in most cases, prolonging the course of the gastric antisecretory drug, increasing the dose, or taking both measures will lead to healing. The causes of nonhealing include poor compliance with medications, an acid hypersecretory state requiring higher-than-customary doses of antisecretory drugs, continued NSAID use, and persistent *H. pylori* infection. Often, combinations of these factors and others (e.g., smoking and stress) are present.

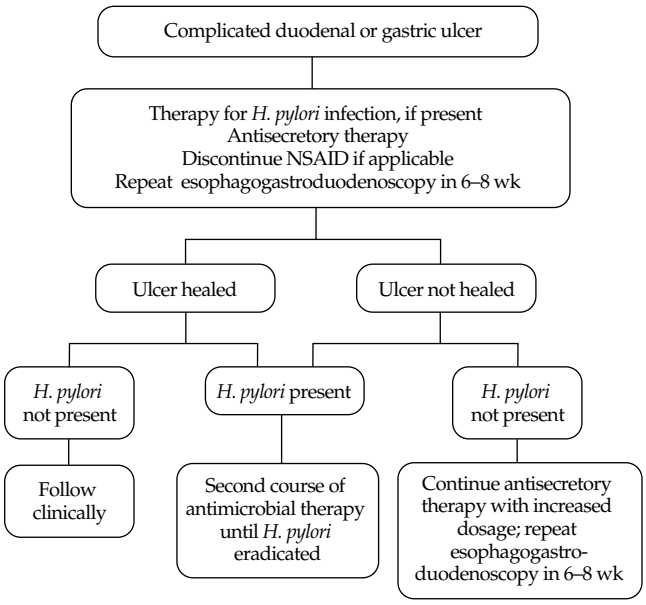


Figure 7 **Approach to treatment and follow-up of a patient with a complicated duodenal or gastric ulcer.**

Poor compliance with medications necessitates patient education and consideration of elective ulcer surgery. A fasting serum gastrin concentration can be used to screen for an acid hypersecretory state resulting from Zollinger-Ellison syndrome. Physicians should be aware that antisecretory drugs (especially proton pump inhibitors) can also raise serum gastrin levels modestly (to 150 to 600 pg/ml). Definitive documentation of an acid hypersecretory state requires quantitative gastric acid measurement (gastric analysis). NSAID use should be discontinued if at all possible. Persistent *H. pylori* infection is the result of poor compliance with medications or is caused by drug-resistant strains.[25-29] *H. pylori* has proved to be resistant to metronidazole in 30% to 40% of cases and to clarithromycin in 10% of cases; resistance to tetracycline or amoxicillin occurs in 1% of strains or less.[25] Combined resistance to macrolides (e.g., clarithromycin) and imidazoles (e.g., metronidazole) occurs in approximately 5% of patients, in whom infection may prove difficult to eradicate. Culture of gastric biopsy material for *H. pylori,* followed by antimicrobial drug-susceptibility testing when available, can guide retreatment. In the absence of this information, the patient should be retreated for 2 weeks with a proton pump inhibitor, with amoxicillin or tetracycline, and with either clarithromycin or metronidazole (whichever antimicrobial agent the patient did not receive initially). Some physicians use a bismuth preparation (e.g., colloidal bismuth subcitrate, bismuth subsalicylate, or ranitidine bismuth citrate) in place of a proton pump inhibitor [*see Table 2*]. Several other antibiotics have activity against *H. pylori* and may prove to be useful in rescue therapy for patients in whom treatments have failed; many such patients harbor antibiotic-resistant strains.[25-29] Agents that are available in the United States include the macrolide azithromycin, the quinolones norfloxacin and levofloxacin, and rifabutin [*see Table 3*].

Frequently Recurring Ulcers

Another type of intractability is frequent ulcer recurrences (at least three a year). This type of intractability occurs most often when *H. pylori* has not been eradicated (necessitating retreatment) or, less often, when NSAID use is resumed or when an acid hypersecretory state is present. Some idiopathic ulcers recur frequently and require lifelong maintenance therapy with an H_2 receptor blocker or a proton pump inhibitor. Alternatively, patients may choose ulcer surgery (parietal cell vagotomy for duodenal ulcer or antrectomy for gastric ulcer) over lifelong medication.

TREATMENT OF COMPLICATED PEPTIC ULCERS

Bleeding Ulcers

The first priority in a patient with a suspected bleeding peptic ulcer is to stabilize the vital signs with volume resuscitation, ideally in an intensive care unit. Such an approach is associated with improved patient outcomes.[32] Hemodynamic monitoring may assist in fluid and blood replacement, particularly if the patient has significant (New York Heart Association class III or IV) cardiac disease.

After the patient becomes clinically stable, diagnostic upper GI endoscopy is performed [*see Figure 7*]. If an actively bleeding ulcer or an ulcer with a visible vessel is found, the lesion is treated endoscopically, by injection of epinephrine, by thermal application with a heater probe or a bipolar electrode, or by a combination of these methods. Endoscopic therapy is successful in controlling bleeding in approximately 90% of patients; the other 10% are referred for surgery if major bleeding continues. Random gastric biopsies are obtained at the time of endoscopy to detect *H. pylori* by rapid urease testing and, if necessary, gastric histology.

Once an ulcer is demonstrated, intravenous gastric antisecretory therapy, usually with a proton pump inhibitor (e.g., pantoprazole) can be started. Oral therapy with a proton pump inhibitor is superior to no therapy in reducing early rebleeding if endoscopic therapy is not attempted[33]; proton pump inhibitors also reduce ulcer rebleeding after endoscopic therapy.[34] The combination of endoscopic therapy and an intravenous proton pump inhibitor is superior to the proton pump inhibitor alone in patients with visible vessels or with clots adherent to the ulcer.[35] However, endoscopic therapy may fail, necessitating surgery, or result in a perforation.[35] Patients who have been on intravenous therapy should be switched to oral therapy with a proton pump inhibitor as soon as they resume oral intake. If the rapid urease test or gastric histology is positive for *H. pylori*, the patient should receive at least two effective antibiotics (e.g., clarithromycin and amoxicillin) for 14 days, along with a proton pump inhibitor [*see Figure 7*]. The proton pump inhibitor is continued at the same dosage until week 6 for duodenal ulcer or week 8 for gastric ulcer. If a subsequent endoscopy shows complete ulcer healing and disappearance of *H. pylori* by both rapid urease testing and gastric histology, the risk of rebleeding is low[34,36,37] and therapy can be stopped. Future use of NSAIDs or aspirin is almost always prohibited. Although the use of coxibs instead of an NSAID is attractive, freedom from rebleeding is not guaranteed. If an NSAID or low-dose aspirin is absolutely necessary, it should be coprescribed with the prostaglandin E_1 analogue misoprostol[38] or a proton pump inhibitor.[39] If the ulcer heals but *H. pylori* organisms are still present, the patient should be treated again for *H. pylori* or left on maintenance therapy with an H_2 receptor blocker or proton pump inhibitor. If the ulcer is not healed, persistent *H. pylori* infection is likely. Under such circumstances, gastric tissue can be cultured for *H. pylori*, if available facilities exist, so that antibiotic sensitivities can be deter-

mined before retreatment. Finally, if the ulcer has not healed even though *H. pylori* has been eradicated, then NSAID use, an acid hypersecretory state, or cancer should be considered. Biopsy samples of the ulcer should be obtained, especially if the ulcer is in the stomach.

A patient with a bleeding peptic ulcer that is negative for *H. pylori* by rapid urease testing and gastric histology usually has an NSAID-related ulcer. The NSAID is stopped if possible, and a high-dose proton pump inhibitor (e.g., 40 mg of omeprazole or 30 mg of lansoprazole) is prescribed for 8 weeks for gastric ulcer or 4 to 6 weeks for duodenal ulcer. Repeat endoscopy is usually indicated to assess healing of a gastric ulcer; if the ulcer has not healed, it is biopsied. Whether to perform a repeat endoscopy to assess healing 4 to 8 weeks after an NSAID-related bleeding duodenal ulcer is controversial. If the ulcer is shown to have healed after the patient is off the NSAID, no further therapy is required unless the patient is placed back on an NSAID or aspirin, even a low dose of aspirin. In such cases, misoprostol (200 µg q.i.d.) is modestly protective against subsequent bleeding.[38] There is also evidence that maintenance therapy with a proton pump inhibitor such as lansoprazole or omeprazole is associated with a low rate of NSAID-related ulcer rebleeding. In general, however, NSAIDs should be avoided in patients with ulcers that have bled. In a study of high-risk arthritis patients with prior ulcer bleeding, use of a coxib (celecoxib) proved to be as effective as the combination of a nonselective NSAID (diclofenac) and a proton pump inhibitor (omeprazole).[40] With either therapy, however, 5% to 6% of such patients rebled over the ensuing 6 months.[40]

Patients with bleeding ulcers that are idiopathic and that heal on an antisecretory drug should receive long-term maintenance therapy with an H_2 receptor blocker or proton pump inhibitor. There is evidence that maintenance therapy with the H_2 receptor blocker ranitidine is effective in preventing rebleeding from duodenal ulcers.[41]

Perforated Ulcers

When a perforated viscus is documented or strongly suspected, the patient is started on broad-spectrum intravenous antibiotics covering gram-negative aerobic bacilli, enterococci, and anaerobes such as *Bacteroides* species and is then taken to surgery for closure of the perforation with a patch of omentum. If the surgeon does not obtain an intraoperative gastric biopsy sample, the patient should undergo postoperative testing for *H. pylori* by serology, urea breath test, or fecal antigen test, and the infection should be treated if present.[42] Many perforated ulcers are associated with NSAID use rather than with *H. pylori* infection.[43] Regardless of the patient's *H. pylori* status, antisecretory drugs should be administered for 6 to 8 weeks postoperatively. Endoscopy is then performed to assess healing, and success of eradication of *H. pylori* is ascertained by gastric biopsy with rapid urease testing and histology.

Up to 10% of perforated gastric ulcers are in fact perforated gastric cancers. If no biopsy or resection of the ulcer is done at the time of repair of the perforation, postoperative endoscopy with biopsy is imperative before the patient is discharged from the hospital or soon thereafter.

The mortality in patients with perforated peptic ulcers is 5% to 10%. Factors associated with higher mortality include delayed diagnosis and treatment of perforation; advanced age; comorbid conditions, such as cardiac, pulmonary, or liver disease; immunodeficiency; and advanced malignancy.

A small number of patients with suspected or probable perforated peptic ulcer improve rapidly before surgery is performed. Others refuse surgery or are poor surgical candidates. An alternative to surgery in these situations is nonoperative therapy consisting of nothing by mouth; intravenous fluids and electrolytes; a broad-spectrum antibiotic such as ticarcillin–clavulanic acid or piperacillin-tazobactam; and nasogastric suction. Compared with surgical therapy, nonoperative therapy is associated with more abdominal complications (e.g., abscesses), fewer pulmonary complications (e.g., atelectasis), and similar mortality.[44]

Obstructing Ulcers

The patient with an obstructing ulcer is initially placed on nasogastric suction, intravenous fluids and electrolytes, and an intravenously administered proton pump inhibitor. If the obstruction is the result of edema associated with an active ulcer, the gastric outlet may open as edema subsides and the ulcer heals, over several days to weeks. If, on the other hand, obstruction is the result of scarring from previous ulcers, it will not resolve with these measures. In some patients, it is difficult to determine whether edema or fibrosis is the primary cause of gastric outlet obstruction. Because obstruction may resolve with time, early consideration should be given to parenteral hyperalimentation. This intervention prevents or minimizes tissue catabolism during the waiting period and also induces a positive nitrogen balance, which will be beneficial if the gastric outlet fails to open up and the patient requires surgery.

A saline load test can be used to guide management. Thirty minutes after 750 ml of isotonic saline is infused into the stomach through the nasogastric tube, gastric contents (saline plus secretions) are aspirated. A return of less than 200 ml indicates normal gastric emptying of liquids and a good prognosis; 200 to 400 ml is indeterminate; and more than 400 ml is suggestive of a high-grade obstruction that will likely require intervention. Repeating the saline load test every day or two may also provide information about whether the obstruction is resolving.

If the obstruction resolves within 3 to 7 days, the nasogastric tube is removed and the patient is fed and observed clinically. An oral proton pump inhibitor or H_2 receptor blocker is started as soon as it is feasible. Gastric prokinetic agents (e.g., metoclopramide) should not be used. At least 50% of patients whose obstruction resolves with conservative medical therapy will experience another obstruction in about a year. Whether routine treatment of *H. pylori* infection will reduce this high recurrence rate is unknown. Unlike bleeding or perforation, obstruction is usually a late complication of ulcer disease. Thus, *H. pylori* eradication in ulcer patients is more likely to be effective in primary prevention of obstruction than in secondary prevention. NSAIDs may cause gastric outlet obstruction as well and should therefore be avoided. Endoscopic therapy is an option for obstruction that does not resolve with conservative therapy.[45] Using inflatable balloons placed over guide wires, the endoscopist can dilate a stenotic pylorus or duodenum under fluoroscopic guidance, although complications such as perforation may occur. Endoscopic balloon dilatation, when feasible, is a temporizing measure and rarely obviates surgery. Thus, obstruction that recurs after medical therapy or after endoscopic therapy is an indication for surgery. Pyloroplasty, gastroenterostomy, and resection plus gastroenterostomy are the most popular operations for an obstructing ulcer. Pyloroplasty and gastroenterostomy are typically combined with a vagotomy to reduce the likelihood of recurrent ulceration.

Fistulizing Ulcers

Gastric or duodenal ulcers associated with fistulas must be biopsied to exclude malignancy. Initially, benign ulcers are treated as described for an uncomplicated ulcer [*see Figure 6*]. An antisecretory agent—ideally, a proton pump inhibitor—should be prescribed, along with antibiotics, if *H. pylori* organisms are present. Ulcer healing may be associated with closure of the fistula. If the fistula persists, surgical resection of the fistula is warranted only if significant symptoms are present (e.g., troublesome diarrhea in a patient with a gastrocolonic fistula or cholangitis in a patient with a duodenocholedochal fistula).

TREATMENT OF ACUTE STRESS ULCERS

Therapy for bleeding acute stress ulcers and erosions involves blood transfusion if necessary and attempts to treat the underlying disease state. The role of intravenous H_2 receptor blockers or proton pump inhibitors is unproved. Endoscopic therapy is not usually curative, because multiple bleeding lesions are often present. In rare cases, visceral angiography with embolization of the major bleeding site is attempted. Gastrectomy for continuous bleeding or significant rebleeding is used as a last resort and is associated with a very high mortality.

Because of the dire consequences of stress ulcers and the lack of an effective therapy, high-risk patients in intensive care units should be placed on stress ulcer prophylaxis.[46] The patients at highest risk for bleeding are those with multiorgan failure and those who receive ventilatory assistance for more than 24 hours. The incidence of significant bleeding in high-risk patients is reduced from about 10% to 25% to about 1% to 5% with the use of prophylactic intragastric or oral antacids or sucralfate or with the use of intravenous H_2 receptor blockers or proton pump inhibitors given by continuous infusion. I prefer intravenous antisecretory therapy because of ease of administration, the ability to monitor gastric pH to assess effectiveness (the goal is a pH > 4), and proven efficacy in clinical trials.

TREATMENT OF CAMERON ULCERS

Although acid secretory inhibitors are often used in the treatment of linear erosions in hiatal hernias (Cameron ulcers), their value is uncertain.[16] Standard therapy consists of packed red cell transfusions for acute bleeding, oral iron replacement for chronic bleeding, and laparoscopic or open repair of the hiatal hernia when medical therapy fails. Because many of the patients with Cameron ulcers are elderly or at high risk, surgery should be undertaken only when medical therapy fails or becomes cumbersome.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Blaser MJ: The bacteria behind ulcers. Sci Am 274:104, 1996

2. *Helicobacter pylori* in peptic ulcer disease. NIH Consensus Development Panel on *Helicobacter pylori* in Peptic Ulcer Disease. JAMA 272:65, 1994

3. Cryer B, Faust TW, Goldschmiedt M, et al: Gastric and duodenal mucosal prostaglandin concentrations in gastric or duodenal ulcer disease: relationships with demographics, environmental, and histological factors, including *Helicobacter pylori*. Am J Gastroenterol 87:1747, 1992

4. Forbes GM, Glaser ME, Cullen DJE, et al: Duodenal ulcer treated with *Helicobacter pylori* eradication: seven-year follow-up. Lancet 343:258, 1994

5. Sung JJY, Chung SCS, Ling TKW, et al: Antibacterial treatment of gastric ulcers associated with *Helicobacter pylori*. N Engl J Med 332:139, 1995

6. Spechler SJ, Fischbach L, Feldman M: Clinical aspects of genetic variability in *Helicobacter pylori*. JAMA 283:1264, 2000

7. Papatheodoridis GV, Papadelli D, Cholongitas E, et al: Effect of *Helicobacter pylori* infection on the risk of upper gastrointestinal bleeding in users of nonsteroidal anti-inflammatory drugs. Am J Med 116:601, 2004

8. Chan FLK, To KF, Wu JCY, et al: Eradication of *Helicobacter pylori* and risk of peptic ulcers in patients starting long-term treatment with non-steroidal anti-inflammatory drugs: a randomised trial. Lancet 359:9, 2002

9. Henry D, Lim LL, Rodriguez LAG, et al: Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. BMJ 312:1563, 1996

10. Cryer B, Feldman M: Effects of very low dose daily, long-term aspirin therapy on gastric, duodenal, and rectal prostaglandin levels and on mucosal injury in healthy humans. Gastroenterology 117:17, 1999

11. Kelly JP, Kaufman DW, Jurgelson JM, et al: Risk of aspirin-associated major upper-gastrointestinal bleeding with enteric-coated or buffered product. Lancet 348:1413, 1996

12. Cryer B, Feldman M: Cyclooxygenase-1 and cyclooxygenase-2 selectivity of widely used NSAIDs and other anti-inflammatory or analgesic drugs: studies in whole blood and gastric mucosa of healthy humans. Am J Med 104:413, 1998

13. Silverstein F, Faich G, Goldstein JL, et al: Gastrointestinal toxicity with celecoxib versus nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: The CLASS Study: A Randomized Controlled Trial. JAMA 284:1247, 2000

14. Bombardier C, Laine L, Reicin A, et al: Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 343:1520, 2000

15. Aoyama N, Kinoshita Y, Fujimoto S, et al: Peptic ulcers after the Hanshin-Awaji earthquake: increased incidence of bleeding gastric ulcer. Am J Gastroenterol 93:311, 1998

16. Spechler SJ: Peptic ulcer disease and its complications. Gastrointestinal and Liver Disease, 7th ed. Feldman M, Friedman LF, Sleisenger MS, Eds. WB Saunders Co, Philadelphia, 2002, p 747

17. Cutler AF, Havstad S, Ma CK, et al: Accuracy of invasive and noninvasive tests to diagnose *Helicobacter pylori* infection. Gastroenterology 109:136, 1995

18. Metz DC, Buchanan M, Purich E, et al: A randomized controlled crossover study comparing synthetic porcine and human secretins with biologically derived porcine secretin to diagnose Zollinger-Ellison syndrome. Aliment Pharm Ther 15:669, 2001

19. Lanas A, Sekar MC, Hirschowitz BI: Objective evidence of aspirin use in both ulcer and nonulcer upper and lower gastrointestinal bleeding. Gastroenterology 103:862, 1992

20. Feldman M, Cryer B, Lee E, et al: Role of seroconversion in confirming cure of *Helicobacter pylori* infection. JAMA 280:363, 1998

21. Vaira D, Nimish V, Menegatti M, et al: The stool test for detection of *Helicobacter pylori* after eradication therapy. Ann Intern Med 136:280, 2002

22. Blum AL, Talley NJ, O'Morain C, et al: Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. N Engl J Med 339:1875, 1998

23. McColl K, Murray L, El-Omar E, et al: Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. N Engl J Med 339:1869, 1998

24. Friedman LS: *Helicobacter pylori* and nonulcer dyspepsia. N Engl J Med 339:1928, 1998

25. Meyer JM, Silliman NP, Wang W, et al: Risk factors for *Helicobacter pylori* resistance in the United States: the surveillance of *H. pylori* antimicrobial resistance partnership study. Ann Intern Med 136:13, 2002

26. Guslandi M: Alternative antibacterial agents for *Helicobacter pylori* eradication. Aliment Pharmacol Ther 15:1543, 2001

27. Gisbert JP, Pajares JM: *Helicobacter pylori* 'rescue' regimen when proton pump inhibitor-based therapies fail. Aliment Pharmacol Ther 16:1047, 2002

28. Perri F, Festa V, Clemente R, et al: Randomized study of two 'rescue' therapies for *Helicobacter pylori*-infected patients after failure of standard triple therapies. Am J Gastroenterol 96:58, 2001

29. Wong WM, Gu Q, Lam SK, et al: Randomized controlled study of rabeprazole, levofloxacin and rifabutin triple therapy vs. quadruple therapy as a second-line treatment for *Helicobacter pylori* infection. Aliment Pharm Ther 17:553, 2003

30. McColl KEL, El-Nujumi AM, Chittajallu RS, et al: A study of pathogenesis of *Helicobacter pylori* negative chronic duodenal ulceration. Gut 34:762, 1993

31. McArthur KE, Richardson CT, Barnett CC, et al: Long-term outcome after exploratory laparotomy and proximal gastric vagotomy in Zollinger-Ellison syndrome. Am J Gastroenterol 91:1104, 1996

32. Baradarian R, Ramdhaney S, Chapalamadugu R, et al: Early intensive resuscitation of patients with upper gastrointestinal bleeding decreases mortality. Am J Gastroenterol 99:619, 2004

33. Khuroo M, Yattoo GN, Javid G, et al: A comparison of omeprazole and placebo for bleeding peptic ulcer. N Engl J Med 336:1054, 1997

34. Lau JW, Sung JJY, Lee KKC, et al: Effect of intravenous omeprazole on recurrent bleeding after endoscopic treatment of bleeding peptic ulcers. N Engl J Med 343:310, 2000

35. Sung JJY, Chan FKL, Lau JYW, et al: The effect of endoscopic therapy in patients receiving omeprazole for bleeding ulcers with nonbleeding visible vessels or adherent clots. Ann Intern Med 139:237, 2003

36. Rokkas T, Karameris A, Mavrogeorgis A, et al: Eradication of *Helicobacter pylori* reduces the possibility of rebleeding in peptic ulcer disease. Gastrointest Endosc 41:1, 1995

37. Macri G, Milani S, Surrenti E, et al: Eradication of *Helicobacter pylori* reduces the rate

of duodenal ulcer rebleeding: a long-term follow-up study. Am J Gastroenterol 93:925, 1998

38. Silverstein FE, Graham DY, Senior JR, et al: Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal antiinflammatory drugs: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 123:241, 1995

39. Lai K, Lam SK, Chu KM, et al: Lansoprazole for the prevention of recurrences of ulcer complications from long-term low-dose aspirin use. N Engl J Med 346:2033, 2002

40. Chan FKL, Hung LCT, Suen BY, et al: Celecoxib versus diclofenac and omeprazole in reducing the risk of recurrent ulcer bleeding in patients with arthritis. N Engl J Med 347:2104, 2002

41. Jensen DM, Cheng S, Kovacs TOG, et al: A controlled study of ranitidine for the prevention of recurrent hemorrhage from duodenal ulcer. N Engl J Med 330:382, 1994

42. Ng EK, Lam YH, Sung JJ, et al: Eradication of *Helicobacter pylori* prevents recurrence of ulcer after simple closure of duodenal ulcer perforation: randomized controlled trial. Ann Surg 231:153, 2000

43. Reinbach DH, Cruickshank G, McColl KEL: Acute perforated duodenal ulcer is not associated with *Helicobacter pylori* infection. Gut 34:1344, 1993

44. Crofts TJ, Park KGM, Steele RJC, et al: A randomized trial of nonoperative treatment for perforated peptic ulcer. N Engl J Med 320:970, 1989

45. DiSario JA, Fennerty MB, Tietze CC, et al: Endoscopic balloon dilation for ulcer-induced gastric outlet obstruction. Am J Gastroenterol 89:868, 1994

46. Cook D, Guyatt G, Marshall J, et al: A comparison of sucralfate and ranitidine for prevention of upper gastrointestinal bleeding in patients requiring mechanical ventilation. N Engl J Med 338:791, 1998

Acknowledgments

Figures 1 and 3 Marcia Kammerer.

Figure 5 Courtesy of Edward Lee, M.D.

The author thanks Tracy Thornburg for help in preparing this chapter.

III DIARRHEAL DISEASES

LAWRENCE R. SCHILLER, M.D., F.A.C.P.

Definition and Epidemiology

The word diarrhea is derived from the Greek words for "flowing through." For most persons, diarrhea means the frequent passage of loose stools.[1] This definition includes two major components: loose-stool consistency (pourable stools) and increased stool frequency (more than two bowel movements daily). Physicians often include a third component: increased stool weight (> 200 g/24 hr), but patients are poor estimators of stool output. In addition, some patients report diarrhea when they have fecal incontinence, even if stools are solid; therefore, every patient complaining of diarrhea should be asked about incontinence.

Diarrhea is a universal human experience. Most persons have had acute infectious diarrhea at some time during their lives. The incidence of acute diarrhea is roughly 5% to 7% annually in the United States.[2] Infectious diarrhea is associated with contaminated food and water and typically is spread via fecal-oral transmission. Chronic diarrhea (i.e., lasting more than 4 weeks) is also common, with a prevalence of approximately 5% in the United States.[3] It is less likely to be caused by infection and more likely to be a symptom of other disorders, such as inflammatory bowel disease, celiac disease, or lactose intolerance.

Pathophysiology and Classification

Diarrhea results from excess water in the stool.[4] To understand the pathophysiology of diarrhea, it is necessary to briefly review how water is transported across the mucosa of the gastrointestinal tract. Water moves in response to osmotic gradients that are established by the absorption of salts (mainly sodium chloride but also potassium and bicarbonate salts) and nutrients (monosaccharides, amino acids, and fatty acids). Salts and nutrients move both passively in response to electrochemical gradients across the mucosa and actively in response to molecular pumps located in the enterocyte membranes.[5]

Each day, a typical person ingests about 2 L of fluid and produces 7 to 8 L of secretions (i.e., saliva, gastric juice, bile, pancreatic juice, and succus entericus). Thus, a total volume of 9 to 10 L enters the upper intestine daily. Most of the water is absorbed in the jejunum, along with nutrients. Absorption of residual nutrients and salts in the ileum results in a reduction of the volume of luminal contents entering the colon to only 1 to 1.5 L daily, a 90% reduction in the volume of fluid entering the intestine each day. The colonic mucosa can absorb salt against large electrochemical gradients and can reclaim 90% of the fluid passing the ileocecal valve each day, making the overall efficiency of small bowel and colonic water absorption about 99%.

Diarrhea develops if the overall efficiency of absorption declines by as little as 1%. This can occur under the following circumstances: the rate of intestinal nutrient and salt absorption decreases; net electrolyte secretion develops (an unusual circumstance except in cases of severe secretory diarrhea such as cholera, in which stool output can exceed 10 L/day); transit through the intestine speeds up, thereby limiting the time available for absorption; or poorly absorbable substances are ingested and increase intraluminal osmotic activity, causing the retention of water within the intestine.[6]

Common problems that primarily cause a reduction in the rate of intestinal nutrient and salt absorption include mucosal diseases, such as celiac disease; inflammatory diseases that disrupt the integrity of the intestinal mucosa (e.g., Crohn disease); and infections that cause diarrhea as the result of toxins that affect enterocyte function.

Isolated acceleration of intestinal transit is a poorly recognized mechanism of diarrhea, although historically, diarrhea was always attributed to it. Some patients with so-called functional diarrhea have rapid intestinal transit, which is likely to be important in the pathogenesis of their condition. Many patients with chronic idiopathic diarrhea have normal rates of fluid and electrolyte absorption when measured under perfusion conditions during which motility effects are neutralized, suggesting that motility must be playing a role in the pathogenesis of their diarrhea under ordinary circumstances.[7] Accelerated transit is also a major factor in diarrhea that is associated with some endocrine diarrheas (e.g., hyperthyroidism, carcinoid syndrome, and other peptide-secreting tumors) or with irritable bowel syndrome.

Poorly absorbed substances that can induce osmotic diarrhea include lactose in lactose-intolerant individuals. Osmotic diarrhea can also occur with ingestion of excess quantities of other poorly absorbed carbohydrates (e.g., fructose and the sugar alcohols mannitol and sorbitol) and ions such as magnesium, phosphate, and sulfate.

Mechanisms that reduce the overall efficiency of absorption may coexist in various disease states. For instance, in celiac disease, loss of intestinal villi results in reduced salt and water absorption, as well as reduced nutrient absorption. Thus, increased stool water in this condition results from both a reduced rate of electrolyte absorption and the increased intraluminal osmotic activity of poorly absorbed substances. Intestinal transit may accelerate in many diarrheal states because of stimulation of peristalsis by increased intraluminal volumes.

FECAL OSMOTIC GAP

As the rate of intestinal salt absorption decreases, the concentration of salts in stool rises to the point at which the concentration approaches plasma osmolality (290 mOsm/kg), which is defined as the osmolality that intestinal contents must maintain beyond the proximal jejunum. If the rate of salt absorption is unimpaired but either nutrients are malabsorbed or poorly absorbable substances are ingested, fecal salt concentrations decrease because most of the available osmotic space is occupied by the poorly absorbed substance. This is the basis for calculation of the fecal osmotic gap [*see Figure 1*].[8] In this calculation, the contribution of electrolytes to stool osmolality is estimated by doubling the concentration of sodium and potassium (the predominant cations in stool water) to account for unmeasured anions (mostly fatty anions, bicarbonate, or chloride). This value is then subtracted from 290 mOsm/kg (the putative osmolality of gut contents) to determine the contribution of nonelectrolytes to fecal osmolality. When electrolytes constitute most of luminal osmolality, the calculated fecal osmotic gap will be low (< 50 mOsm/kg). When poorly absorbable substances are present, the fecal osmotic gap will be large (> 100 mOsm/kg). Watery diarrhea with a low osmotic gap is classified as secretory diarrhea; diarrhea with a large fecal osmotic gap is classified as osmotic di-

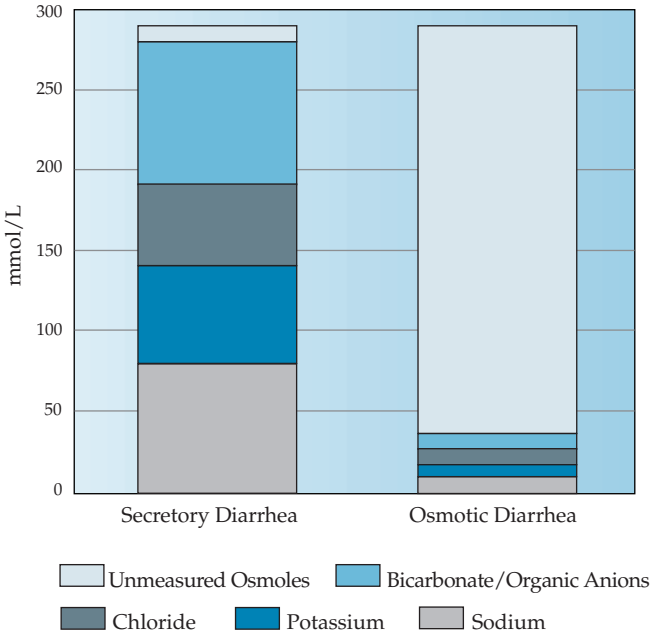


Figure 1 **Fecal electrolyte concentrations in secretory diarrhea (left column) and in osmotic diarrhea (right column). Note that most of the fecal osmolality can be attributed to fecal electrolytes in secretory diarrhea, whereas most of the osmolality in osmotic diarrhea results from the unmeasured (nonelectrolyte) osmoles. Calculation of the fecal osmotic gap allows an estimate of the contribution of unmeasured osmoles to fecal osmolality.[82]**

arrhea. These categories are most helpful in the evaluation of patients with chronic diarrhea.

CLASSIFICATION OF DIARRHEA

For clinical purposes, diarrhea can be classified as either acute (< 4 weeks' duration) or chronic (> 4 weeks' duration). Chronic diarrhea is further divided into watery, inflammatory, and fatty on the basis of stool characteristics.[3] The value of this classification is that it allows the physician to direct evaluation and management more effectively, because diarrheal diseases can be distinguished by the duration of illness and the type of stools produced.

Acute Diarrhea

ETIOLOGY

Infectious Causes of Acute Diarrhea

Most forms of acute diarrhea (i.e., those lasting less than 4 weeks) are caused by infections and are self-limiting; the majority are caused by viruses (e.g., adenovirus, Norwalk agent, rotovirus), but some are caused by bacteria (e.g., *Salmonella, Shigella,* and *Escherichia coli*) and others by protozoa (e.g., *Giardia,* amebas) [*see Table 1*].[2] The disease course of most viral and bacterial diarrheas lasts less than 1 week; therefore, infectious diarrhea lasting more than 7 days is more likely to be caused by protozoa.[9]

The epidemiology of acute infectious diarrhea depends on the circumstances of the infection and where one contracts the infection. For example, a history of recent travel, particularly to developing countries, makes a diagnosis of traveler's diarrhea likely. Previous antibiotic use and residence in an institution where antibiotic use is common (e.g., hospitals and nursing homes) are risk factors for *Clostridium difficile* infection. Children in day care facilities and their contacts, people engaging in promiscuous sexual activity, and users of illicit intravenous drugs are all at increased risk of contracting infectious diarrhea. Consumption of potentially contaminated food and drink is another risk factor for infectious diarrhea. With the globalization of commerce and mass processing of food, esoteric infections from overseas and large outbreaks of diarrhea have become more common.[10]

Pathogenic infections cause diarrhea by one of four mechanisms: (1) enterotoxins that subvert the regulatory mechanisms of enterocytes, (2) cytotoxins that destroy enterocytes, (3) adherence to the mucosa by organisms (so-called enteroadherent organisms) that alter enterocyte function as a result of physical proximity to the mucosa, and (4) invasion of the mucosa by organisms that provoke an inflammatory response by the immune system.[11] In general, patients with cytotoxin-mediated diarrhea and those with invasive organisms experience more toxicity and have more abdominal pain than patients with enterotoxin-mediated diarrhea or enteroadherent infections.

Toxic Causes of Acute Diarrhea

Another mechanism for acute diarrhea is ingestion of a preformed toxin.[12] Several species of bacteria, such as *Staphylococcus aureus*, *Clostridium perfringens,* and *Bacillus cereus*, can produce toxins that in turn cause so-called food poisoning (i.e., vomiting and diarrhea within 4 hours after ingestion). In such cases, the bacteria do not need to establish an intraluminal infection; ingestion of the toxin alone can produce the disease. Symptoms subside after the toxin is cleared, usually by the next day, and evidence of toxicity (e.g., fever) is minimal.

Other Causes of Acute Diarrhea

Other potential causes of acute diarrhea include food allergies and medication reactions. Food allergies are rarely recognized as causes of diarrhea in adults in the United States unless the diarrhea is associated with urticaria or other allergic symptoms. Medications often produce diarrhea as a side effect; this association is typically recognized by the patient because of the temporal relation between drug ingestion and diarrhea.

Finally, acute diarrhea may represent the initial stages of chronic diarrhea. However, patients with chronic diarrhea often do not seek help during the initial weeks of their illness unless the diarrhea is severe or is complicated by dehydration, symptomatic electrolyte disorders, or fever.

DIAGNOSIS

Medical History

A careful medical history is the key to the diagnosis of diarrhea. The acuity and severity of the process should be determined. Frequency of defecation is the easiest parameter for patients to relate, but frequency does not necessarily correlate with stool weight, which is a more meaningful measure of the physiologic impact of diarrhea. Manifestations of dehydration or volume depletion, such as orthostasis, thirst, decreased urine output, and weakness, suggest voluminous diarrhea. Acute weight loss can also be a guide to the severity of diarrhea; voluminous diarrhea produces substantial weight loss if rehydration efforts are suboptimal.

Stool characteristics are also quite important. The presence of blood or pus in the stool raises the issue of inflammatory diarrhea, such as that from colitis or enteroinvasive bacteria. Watery

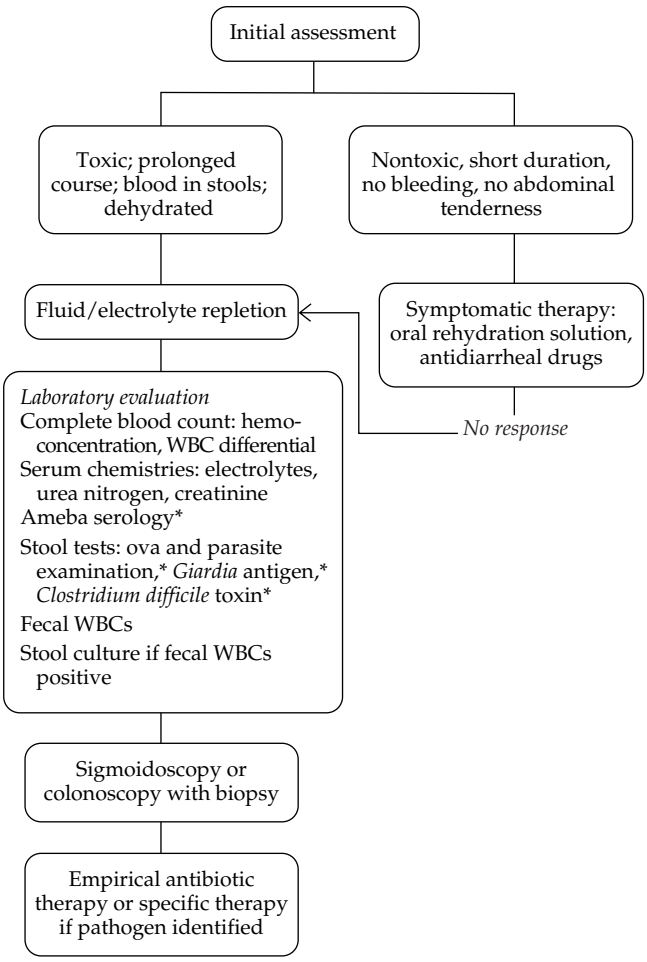
Table 1 Selected Infectious Diarrheas

Organism	*Vehicle*	*Mechanism*	*Classic Characteristics*	*Complications*
Campylobacter	Food (poultry); animal-to-person	Invasion; inflammation	Watery or bloody diarrhea; ileitis and/or colitis, ulceration	Guillain-Barré syndrome; reactive arthritis
Salmonella	Food (poultry, eggs, seafood); animal-to-person	Invasion; inflammation	Gastroenteritis, ileitis, colitis; enteric fever (*S. typhi*)	Endovascular infection; osteomyelitis; sepsis
Shigella	Food (poultry); day care centers	Cytotoxin; inflammation	Two phases: enteritis (fever, cramps, diarrhea), followed by colitis (ulcers, inflammation)	Seizures, encephalopathy; reactive arthritis
E. coli O157:H7	Food (beef); fruit juices	Cytotoxin	Hemorrhagic colitis	Hemolytic-uremic syndrome
Enteroinvasive *E. coli*	Food (various); water	Invasion; inflammation	Colitis	Fever, sepsis
Enterotoxigenic *E. coli*	Food (various); water	Enterotoxin	Watery diarrhea	Dehydration, shock
Enteropathogenic and enteroadherent *E. coli*	Food (various); water	Contact with brush border	Watery diarrhea; may be prolonged	Dehydration
Vibrio cholerae	Water, seafood	Enterotoxin	Voluminous watery diarrhea	Dehydration, shock
Clostridium difficile	Person-to-person	Cytotoxin	Nosocomial infection; antibiotic-associated diarrhea; toxicity	Toxic megacolon; protein-losing enteropathy
Aeromonas, Plesiomonas	Water	Enterotoxin	Watery diarrhea; may be prolonged	
Yersinia	Raw milk	Invasion; inflammation	Acute diarrhea or chronic ileocolitis-like Crohn disease	Reactive arthritis, extraintestinal infection, Guillain-Barré syndrome
Bacillus cereus	Fried rice	Exotoxin	Acute gastroenteritis	Fulminant liver failure
Staphylococcus	Fatty foods	Exotoxin	Acute gastroenteritis	
Clostridium perfringens	Fatty foods	Exotoxin	Acute gastroenteritis	
Viruses	Person-to-person; water	Inflammation; ?toxins	Acute gastroenteritis; watery diarrhea	
Giardia	Person-to-person; animal-to-person; water; day care	Contact	Watery diarrhea, dyspepsia	
Cryptosporidium	Water; day care; animal-to-person	Contact	Watery diarrhea, may be prolonged; epidemics	
Cyclospora	Imported fruit	Inflammation	Watery diarrhea, flatulence, pain, fatigue; may be prolonged	
Entamoeba histolytica	Person-to-person	Invasion; inflammation	Variable: asymptomatic to dysentery; may mimic irritable bowel syndrome, inflammatory bowel disease	Liver abscess
Strongyloides	Larvae invade skin	Invasion; inflammation	Abdominal pain and diarrhea	Hyperinfection in immunosuppressed hosts

stools are more in keeping with a secretory process. The relationship of defecation to meals or fasting and the occurrence of nocturnal diarrhea, fecal urgency, or incontinence are other points of potential significance. Urgency and incontinence do not necessarily indicate voluminous diarrhea; more often, they reflect independent defects in the continence mechanisms. Additional symptoms of diarrhea that should be noted are abdominal pain or cramps; flatulence; bloating or distention; fever; and weight loss. A list of all prescription, over-the-counter, and herbal medications being taken by the patient should be compiled, and previous surgeries or radiation therapy should be discussed. The patient's diet should be scrutinized, and epidemiologic features (e.g., family members or other contacts with diarrhea, recent travel, water source, occupation, sexual activity, and illicit drug use) should be investigated.

Physical Examination

The physical examination is more useful for judging the severity of diarrhea than for determining its cause. Volume status should be assessed by looking for orthostatic change in blood



*In appropriate epidemiologic circumstances.

Figure 2 **Initial evaluation of acute diarrhea.**[83]

pressure and pulse. Fever and other signs of toxicity should be recorded. A careful abdominal examination, with emphasis on bowel sounds and the presence of distention or tenderness, should be conducted.

Laboratory Testing

Blood tests Extensive laboratory testing is not necessary for most patients with acute diarrhea; it should be reserved for those with toxicity, dehydrating diarrhea, or persistence of diarrhea for longer than would be expected, given its probable cause [*see Figure 2*]. In patients requiring extensive laboratory tests, a complete blood count should be obtained to assess for hemoconcentration, anemia, or leukocytosis. Patients with viral diarrhea typically have normal white blood cell (WBC) counts and differentials, although lymphocytosis may be seen. Invasive bacterial infections typically produce leukocytosis with many immature WBCs, but salmonellosis can induce leukopenia. Serum electrolytes and renal tests can define the metabolic impact of diarrhea.

Stool tests Stool testing is of value for patients with blood in their stools, dehydrating diarrhea, prolonged diarrhea, or dysentery and for patients who present as part of an outbreak of diarrhea. Stool cultures are sensitive and specific, but they are expensive. Some experts recommend obtaining stool cultures only for patients who have leukocytes (or the leukocyte marker lactoferrin) in the stool, because the yield of pathogenic bacteria will be higher in this group.[13] Other researchers dispute this recommendation.[14] Laboratories routinely test for *Salmonella*, *Shigella*, *Campylobacter*, and *E. coli* serotype O157:H7. Special cultures for tuberculosis, *Yersinia*, *Aeromonas*, or *Plesiomonas* may need to be requested in appropriate patients. Polymerase chain reaction (PCR) testing with primers that are based on bacterial DNA is proving to be an accurate and sensitive technique for specific diagnosis.

Examination of stool for ova and parasites has variable utility, depending on the pretest probability of certain infections. For example, such testing might be very useful in a day care worker with diarrhea, but it would be of little help in a patient with hospital-acquired diarrhea. Enzyme-linked immunosorbent assay (ELISA) testing for giardiasis and serologic testing for amebiasis are more accurate tests for such specific infections in most settings. Patients who were treated with antibiotics during the 3 months before the onset of diarrhea or patients who develop diarrhea in institutional settings should have a stool sample analyzed for *C. difficile* toxin.

Imaging and Endoscopic Tests

In patients who have toxicity, blood in their stools, or persistent acute diarrhea, sigmoidoscopy or colonoscopy should be considered. In most patients without rectal bleeding, sigmoidoscopy is probably adequate as an initial evaluation, because most patients with colitis will have involvement of the left side of their colon. In patients with bleeding or those with AIDS and diarrhea, colonoscopy is preferable because some opportunistic infections and lymphomas are seen only in the right colon.[15] Mucosal biopsies should be obtained in either case, particularly if the colon is grossly inflamed, because the pathologist can readily distinguish self-limited colitis from chronic ulcerative colitis even early on in the course of the disease.[16] Patients with toxicity should undergo abdominal x-rays or computed tomography to confirm a diagnosis of colitis, to determine its extent, and to look for evidence of ileus or megacolon.

TREATMENT

Nonspecific Therapy

Because most cases of acute diarrhea are self-limited, most patients do not require specific therapy. Instead, judicious replacement of fluid and electrolyte losses is sufficient. This can be accomplished by intravenous fluids or oral rehydration solutions. Oral rehydration solutions are based on the concept that nutrient absorption accelerates sodium and fluid absorption by the jejunum.[17] Initially, rehydration formulas used glucose as the absorbable nutrient; more recently, cereal-based oral rehydration solutions have been found to be more efficient. Oral rehydration solution does not reduce fecal losses (it may actually increase stool output); instead, it increases net fluid and electrolyte absorption. These solutions cannot be used if vomiting precludes ingestion; in such situations, intravenous rehydration must be used. Sports drinks (e.g., Gatorade) are designed to offset fluid and electrolyte losses from sweating and do not contain sufficient amounts of sodium to replace fecal losses. Solutions that more closely approximate World Health Organization rehydration solution are now commercially available (e.g., Rehydralyte, Resol, Ricalyte).

Diet Most patients seek advice about altering their diets when suffering from diarrhea. Other than the provision of adequate water and salt, no specific instructions are needed. Some physicians routinely restrict dairy products in patients with diarrhea on the theory that these patients may have temporary lactase deficiency. This precaution is not necessary unless there is clinical evidence of lactose intolerance (e.g., exacerbation of diarrhea or flatus with ingestion of dairy products).

Antibiotics Empirical antibiotic therapy for acute diarrhea may be appropriate under certain circumstances (e.g., diarrhea in travelers, outbreaks of bacterial or protozoan diarrhea, patient frailty, and patient toxicity). However, experts discourage routine use of empirical antibiotic therapy because of its lack of demonstrable efficacy in many infections and because of concerns about precipitating complications, such as hemolytic-uremic syndrome in patients with *E. coli* serotype O157:H7.[18] A meta-analysis suggests that this latter point is not supported by the literature.[19] When indicated, fluoroquinolones or trimethoprim-sulfamethoxazole is commonly used as empirical therapy. Rifaximin, a nonabsorbable antibiotic with activity against most bacterial enteric pathogens, is useful in patients with infections that are limited to the lumen (e.g., traveler's diarrhea), but it may not be effective against invasive organisms.[20] In cases of persistent diarrhea in which protozoan infection is more likely, nitazoxanide, metronidazole, or tinadazole may be more appropriate.[21]

Nonspecific antidiarrheal agents, such as opiates, can reduce stool frequency and stool weight, and they may reduce associated symptoms, such as abdominal cramps.[22] Concerns about slowing the clearance of pathogens from the intestine by reducing peristalsis have largely not been borne out. Intraluminal agents, such as bismuth subsalicylate (Pepto-Bismol) and adsorbents (e.g., kaolin), are also sometimes used [*see Table 2*].

Therapy for Specific Infections and Syndromes

Campylobacter A frequent cause of acute ileocolitis in the United States,[23] *Campylobacter* is usually acquired by eating undercooked chicken; it has an incubation period of up to 1 week. Ulceration of the colonic mucosa and bloody diarrhea may occur with this infection. Antibiotics, such as erythromycin, shorten the course of the illness if given within the first few days of symptoms.

Table 2 Nonspecific Treatment of Diarrhea

Category	*Treatment*	*Typical Adult Dose*
Rehydration	Intravenous fluid	1–5 L/24 hr
	Oral rehydration solution	1–5 L/24 hr
Intraluminal agents	Adsorbents (kaolin-pectin)	15–60 ml q.i.d.
	Bismuth subsalicylate	30 ml q.i.d.
	Texture modifiers (psyllium)	18–30 g/24 hr
Drugs that inhibit transit	*Opiates*	
	Deodorized tincture of opium (10 mg morphine/ml)	5–20 drops q.i.d.
	Paregoric (0.4 mg morphine/ml)	5–10 ml q.i.d.
	Morphine sulfate (20 mg/ml)	2–10 drops q.i.d.
	Codeine phosphate or sulfate	15–60 mg q.i.d.
	Diphenoxylate with atropine	1–2 tablets q.i.d.
	Difenoxin with atropine	1–2 tablets q.i.d.
	Loperamide (2 mg)	1–2 tablets q.i.d.
	Others	
	Clonidine	0.1–0.3 mg t.i.d.
	Octreotide injection	50–200 mg t.i.d.

Salmonella enteritidis* and *S. choleraesuis *S. enteritidis* and *S. choleraesuis* are spread via contaminated food or water and cause acute gastroenteritis, ileocolitis, or colitis characterized by watery diarrhea.[24] Antibiotic therapy with a fluoroquinolone, ampicillin, or trimethoprim-sulfamethoxazole should be reserved for severely ill patients or patients with compromised immunity (e.g., infants, elderly patients, pregnant women, and AIDS patients).[25]

Salmonella typhi *S. typhi* causes typhoid fever, a form of enteric fever.[26] The propensity of *S. typhi* to produce bacteremia distinguishes it from other enteric pathogens. When the infection is limited to the intestine of an otherwise healthy individual, no specific therapy is indicated, because antibiotics may paradoxically prolong excretion of the organism and increase relapses. When the infection becomes systemic and the patient is very ill, therapy is necessary, especially if the organism produces a metastatic endovascular infection. Fluoroquinolones are most often used. The diagnosis of a carrier state is made when stool cultures are positive for over 1 year.

Shigella *Shigella* species are invasive organisms, but they also produce an enterotoxin that reduces water and electrolyte absorption.[27] Shigellosis commonly causes a watery diarrhea initially (this watery diarrhea is most likely related to the enterotoxin). Watery diarrhea is followed by bloody diarrhea, which results from colitis produced by invasion of the colonic mucosa. Because of growing resistance to the fluoroquinolones in the United States, trimethoprim-sulfamethoxazole is the recommended initial treatment for most patients with shigellosis. Shigellosis contracted overseas is initially treated with fluoroquinolones, because those strains are more likely to be resistant to trimethoprim-sulfamethoxazole.

***E. coli* serotype O157:H7** The O157:H7 organism has become a common cause of food-borne infection in the United States.[28] It produces toxins similar to those produced by *Shigella*.[29] Infection with this organism causes a hemorrhagic segmental colitis. The disease often occurs in large outbreaks from contamination of widely distributed foods, such as hamburger meat. Patients can become quite ill; hemolytic-uremic syndrome is a well-recognized complication. Antibiotics do not seem to improve the course of the illness and may cause hemolytic-uremic syndrome in children, although this theory is controversial.[18,19]

Clostridium difficile *C. difficile* has become the most common cause of nosocomial diarrhea in many institutions.[30] In nonhospitalized adults, carriage rates for this organism are low, but it is spread easily from person to person by spores. Suppression of the normal bacterial flora of the colon can result in an overgrowth of *C. difficile,* if it is present. The organisms produce toxin A and toxin B; these cytotoxins inactivate small guanosine triphosphate (GTP)–binding proteins in the enterocytes, resulting in apoptosis.[31] In institutional settings, the organism can be distributed efficiently to a large pool of susceptible persons by health care workers who do not wash their hands. The disease produced can range from a simple, self-limited diarrhea to a fulminant colitis.

Treatment for 2 weeks with metronidazole, 250 mg four times daily, or vancomycin, 125 to 500 mg four times daily, is effective against *C. difficile*. Relapses occur in up to 25% of patients, probably because of residual spores.[32] Ingestion of probiotic bacteria or the nonpathogenic yeast *Saccharomyces boulardii* may reduce relapse rates.[33] In most instances of relapse, longer periods of antibiotic therapy are indicated.

Other nosocomial diarrheas Noninfectious causes of nosocomial diarrhea include medications (particularly elixirs that contain sorbitol or mannitol as noncaloric sweeteners and cancer chemotherapeutic drugs) and enteral feeding; in addition, nosocomial paradoxical diarrhea can occur in patients with fecal impaction. Infections with organisms other than *C. difficile* also occur in institutions, particularly extended-stay facilities. An important cohort of hospital patients that may develop infectious diarrhea are those who are immunocompromised by diseases such as AIDS or by drugs that are used to treat transplant rejection or inflammatory diseases. These patients are often infected with opportunistic pathogens, including viruses (e.g., cytomegalovirus and herpesvirus), bacteria (e.g., *Mycobacterium avium* complex), and parasites (e.g., *Cryptosporidium* and *Strongyloides*).[34,35] In addition, bone marrow transplant recepients may develop acute diarrhea from graft versus host disease.

Parasites Acute diarrhea in noninstitutionalized patients can be caused by parasites.[36] The likelihood of parasitic disease as a cause of acute diarrhea is profoundly influenced by geography and epidemiologic features. Giardiasis, for example, is a common infection in some areas but not others, probably because of variability in the effectiveness of water treatment. Ingestion of as few as a dozen cysts of *G. lamblia* may establish an infection, which accounts for the frequency of person-to-person transmission of this disease. ELISA for *Giardia* antigen is superior to microscopic inspection of stool (so-called ova and parasites testing) for the detection of giardiasis. Therapy with tinidazole, metronidazole, or nitazoxanide is effective in most patients, but reinfection can occur.[37]

Amebiasis is also common in some areas. Persons with amebiasis may be asymptomatic or may be extremely ill from invasion and spread of the organism to other organs, such as the liver.[38] Diagnosis is typically made by microscopic examination of fresh stools, but ELISA shows promise in distinguishing the pathogenic species, *Entamoeba histolytica,* from nonpathogenic amebas. The colonoscopic appearance of amebiasis is often distinctive, and the organism can be identified in colonic biopsy specimens.

Cryptosporidiosis is a common but unappreciated cause of diarrhea.[39] *Cryptosporidium* is resistant to chlorination, and it can cause large outbreaks when water supplies are contaminated. Microscopic inspection of stools has poor sensitivity for this organism, and many cases go undiagnosed. Treatment with nitazoxanide reduces the duration of diarrhea in children and adults with this infection.[39]

Other parasites that may cause acute diarrhea include *Isospora*, *Cyclospora*, *Trichuris trichiura* (whipworm), and *Strongyloides*. Special tests that may be necessary to identify these parasites include concentration of stool samples and mucosal biopsy. If these organisms are suspected, consultation with the laboratory staff allows use of the proper diagnostic tests.

Chronic Diarrhea

In contrast to acute diarrhea, in which infection is the overwhelmingly likely cause of illness, chronic diarrhea has an extensive and daunting list of possible causes [*see Table 3*].[3] The simplest approach to making a diagnosis is to classify chronic diar-

Table 3 Major Causes of Chronic Diarrhea

- Osmotic diarrhea
  - Osmotic laxative abuse
    - Mg^{2+}, SO_4^{2-}, PO_4^{3-}, lactulose, mannitol, sorbitol, polyethylene glycol
  - Carbohydrate malabsorption
    - Lactose, fructose, others
- Fatty diarrhea
  - Malabsorption syndromes
    - Mucosal diseases
    - Short bowel syndrome
    - Postresection diarrhea
    - Small bowel bacterial overgrowth
    - Mesenteric ischemia
  - Maldigestion
    - Pancreatic insufficiency
    - Reduced luminal bile acid
- Inflammatory diarrhea
  - Inflammatory bowel disease
    - Ulcerative colitis
    - Crohn disease
    - Diverticulitis
    - Ulcerative jejunoileitis
- Infections
  - Invasive bacterial infection
    - *Clostridium, E. coli,* tuberculosis, others
  - Ulcerating viral infection
    - Cytomegalovirus
    - Herpes simplex
  - Invasive parasites
    - Amebiasis
    - Strongyloides
- Ischemic colitis
- Radiation enterocolitis
- Neoplasia
  - Carcinoma of the colon
  - Lymphoma
- Secretory diarrhea
  - Congenital chloridorrhea
  - Chronic infections
  - Inflammatory bowel disease
    - Ulcerative colitis
    - Crohn disease (ileum)
    - Microscopic colitis
      - Lymphocytic colitis
      - Collagenous colitis
    - Diverticulitis
  - Drugs and poisons
    - Stimulant laxative abuse
- Disordered regulation
  - Postvagotomy
  - Postsympathectomy
  - Diabetic neuropathy
  - Irritable bowel syndrome
- Ileal bile acid malabsorption
- Endocrine diarrhea
  - Hyperthyroidism
  - Addison disease
- Neuroendocrine tumors
  - Gastrinoma
  - VIPoma
  - Somatostatinoma
  - Mastocytosis
  - Carcinoid syndrome
  - Medullary carcinoma of the thyroid
- Other neoplasia
  - Colon carcinoma
  - Lymphoma
  - Villous adenoma
- Idiopathic secretory diarrhea
  - Epidemic (Brainerd)
  - Sporadic

rhea by the characteristics of the stools. Three categories of chronic diarrhea are recognized: watery, inflammatory, and fatty. Watery diarrhea can be subdivided further into osmotic and secretory diarrhea on the basis of stool analysis.

WATERY DIARRHEA

Osmotic Diarrhea

Osmotic diarrhea results from ingestion of an osmotically active, poorly absorbable substance that necessitates the retention of water intraluminally to maintain isosmotic conditions.[40] In practical terms, osmotic diarrhea is caused by ingestion of osmotic laxatives (magnesium, phosphate, and sulfate salts; sugar analogues, such as lactulose; sugar alcohols, such as mannitol or sorbitol; and polyethylene glycol) and carbohydrate malabsorption. The ingestion of osmotic laxatives may be purposeful [*see* Laxative Abuse, *below*] or accidental, as when excess magnesium is ingested as part of an antacid, mineral supplement, or multivitamin tablet. Carbohydrate malabsorption is most often the result of acquired lactase deficiency (a normal development in adult mammals) or mucosal disease, such as celiac sprue, that interferes with nutrient absorption.

Secretory Diarrhea

Secretory diarrhea has a much larger list of possible causes than does osmotic diarrhea [*see Table 3*].

Congenital chloridorrhea Rarely, congenital absence of a transporter mechanism results in diarrhea. This is the case in congenital chloridorrhea, in which the chloride-bicarbonate exchanger in the ileum is not active.[41] Under such conditions, chloride becomes poorly absorbable in the distal bowel and obligates water retention intraluminally.

Chronic infections Some bacterial infections can last long enough to produce chronic secretory diarrhea.[42] These include *Aeromonas*, *Plesiomonas*, enteropathogenic *E. coli*, *C. difficile*, *M. tuberculosis*, and *Yersinia enterocolitica*. A special situation is small bowel bacterial overgrowth syndrome, in which structural problems, such as jejunal diverticulosis, or motility problems, such as those seen in scleroderma, result in proliferation of bacteria in the jejunum.[43] Although this bacterial overgrowth disrupts digestive processes and may produce fatty diarrhea, it also may reduce water and salt absorption, producing secretory diarrhea. Infection with parasites, such as *G. lamblia*, *E. histolytica*, and *Cryptosporidium*, also can produce chronic diarrhea.[44]

Inflammatory bowel disease Typically, inflammatory bowel diseases (e.g., ulcerative colitis and Crohn disease) produce inflammatory diarrhea, with blood and pus in the stool. Watery diarrhea can occur, especially when the distal colon is not involved. One form of inflammatory bowel disease that typically produces a watery diarrhea is microscopic colitis syndrome (lymphocytic colitis and collagenous colitis), in which the mucosa is inflamed but not ulcerated.[45] Colonic diverticulitis is sometimes associated with a secretory diarrhea, which is probably mediated by inflammation-linked cytokines. Vasculitis and systemic inflammatory diseases may also be associated with secretory diarrhea.

Drugs Drug therapy is a key cause of secretory diarrhea.[46] Many drugs have diarrhea as a side effect. These include antibiotics; cardiovascular agents, such as beta-adrenergic antagonists, digitalis, and quinidine; cancer chemotherapy; nonsteroidal anti-inflammatory drugs (NSAIDs); and colchicine. Thus, in taking the history of a patient with chronic diarrhea, it is critical to formulate a detailed drug list, including over-the-counter and alternative medications. A special category of drug-induced secretory diarrhea is surreptitious ingestion of stimulant laxatives.

Other causes Disordered motility or regulation can produce secretory diarrhea. Secretory diarrhea associated with disordered motility can occur in patients who have undergone vagotomy or sympathectomy, patients with autonomic neuropathy from diabetes or amyloidosis, and many patients with irritable bowel syndrome.[47,48] In the United States, irritable bowel syndrome is the most common diagnosis made in patients with chronic diarrhea. This diagnosis is often incorrect, however, and may delay accurate diagnosis and treatment.

Malabsorption of bile acid in the ileum occurs in many diarrheal diseases as a result of ileal disease or resection and may be secondary to other processes, such as vagotomy, cholecystectomy, and rapid transit past the ileum. In a relatively small group of patients, idiopathic bile acid malabsorption is the cause of diarrhea.[49]

Endocrine causes of secretory diarrhea include hyperthyroidism, Addison disease, and a group of rare tumors of the endocrine cells of the gut, including gastrinomas, carcinoid tumors, vasoactive intestinal peptide tumors (VIPomas), somatostatinomas, and medullary carcinoma of the thyroid.[50,51] These tumors produce peptides and other mediators that affect intestinal mucosal and muscle function and thereby produce diarrhea. In most cases, rapid intestinal transit seems to be the major mechanism producing diarrhea in these disorders, although this remains controversial.

Other tumors that produce secretory diarrhea include colon cancer (mechanism uncertain), villous adenoma of the rectum, lymphoma, and mastocytosis. Mastocytosis (and probably some lymphomas) produce diarrhea by release of histamine or other mediators that affect gut function. Infiltration of the mucosa also may play a role in some cases.

Secretory diarrhea can also be idiopathic.[52] Idiopathic secretory diarrhea occurs in both sporadic and epidemic forms and may be caused by an as-yet unidentified infection.

INFLAMMATORY DIARRHEA

Inflammatory diarrhea is characterized by the presence of blood and pus in the stools, which usually occurs as a result of ulceration of the mucosa. Inflammatory bowel diseases, such as Crohn disease and ulcerative colitis, are in this category [*see 4:IV Inflammatory Bowel Diseases*]. Some patients with diverticulitis and diarrhea may have blood and pus in the stool, as do patients with the rare condition ulcerative jejunoileitis. Ulcerating infectious diseases may also produce inflammatory diarrhea. Such infections include pseudomembranous colitis from *C. difficile* infection; invasive bacterial infections, such as tuberculosis and yersiniosis; ulcerating viral infections, such as those caused by cytomegalovirus or herpesvirus; and invasive parasitic infections, such as amebiasis and *Strongyloides*. Inflammatory diarrhea also may be seen with ischemic colitis and radiation colitis, as well as colon cancer and lymphoma.

FATTY DIARRHEA

Fatty diarrhea may be caused by fat malabsorption resulting from mucosal diseases, such as celiac disease or Whipple dis-

ease; short bowel syndrome secondary to extensive surgical resection of the small intestine; small bowel bacterial overgrowth syndrome; and mesenteric ischemia. Fatty diarrhea also may be the consequence of maldigestion of fat caused by pancreatic exocrine deficiency or inadequate luminal bile acid concentration [*see 4:XI Diseases Producing Malabsorption and Maldigestion*].

Diagnosis

Medical history An accurate medical history is even more important in cases of chronic diarrhea than in acute diarrhea. In addition to all the issues that should be discussed with patients who have acute diarrhea [*see* Acute Diarrhea, *above*], the history of patients with chronic diarrhea should include long-term trends in body weight, current appetite and food intake, review of previous medical problems and surgeries, potential secondary gains from illness, previous evaluations and treatments for diarrhea, and a detailed review of systems to look for clues to systemic illnesses [*see Table 4*].

A principal diagnostic distinction in chronic diarrhea is between diarrhea associated with irritable bowel syndrome and diarrhea associated with other functional or organic problems. Irritable bowel syndrome is characterized by abdominal pain associated with defecation and an altered bowel habit.[53] Variable stool consistency and intermittent constipation are common. Painless diarrhea should no longer be considered to be a type of irritable bowel syndrome; other causes of diarrhea should be sought in such cases.

Physical examination The physical examination may provide clues to the diagnosis of chronic diarrhea. Characteristic skin changes may be seen in mastocytosis, glucagonoma, Addison disease, amyloidosis, carcinoid syndrome, Degos disease, and celiac disease. Amyloidosis may produce orthostatic hypotension and hepatosplenomegaly. Thyroid nodules or findings of hyperthyroidism may suggest medullary carcinoma of the thyroid or thyroid adenoma causing hyperthyroidism. Carcinoid syndrome may produce hepatosplenomegaly, edema, and a right-sided heart murmur in addition to flushing. Arthritis may be a clue to inflammatory bowel disease, Whipple disease, and some enteric infections. Lymphadenopathy could be present in patients with AIDS or lymphoma. The absence of peripheral arterial pulses or bruits suggests the possibility of mesenteric vascular disease. Rectal examination may disclose defective functioning of the anal sphincter or pelvic floor muscle, which could produce fecal incontinence. The physical findings that reflect the severity of diarrhea should also be recorded [*see* Acute Diarrhea, *above*].

Laboratory tests As in acute diarrhea, routine laboratory testing is indicated to help determine the severity of chronic diarrhea [*see* Acute Diarrhea, *above*]. Unlike acute diarrhea, in which stool analysis is typically not used, stool analysis plays a key role in the assessment of chronic diarrhea by allowing adequate categorization of the type of diarrhea, thereby limiting the number of conditions to be considered.[3] The stool analysis can be obtained on either a random sample or a timed collection. The value of a timed collection is that it allows the physician to quantitate stool output accurately. However, stool analysis obtained on a random sample can still provide many diagnostic clues.

Table 4 Steps in the Evaluation and Classification of Chronic Diarrhea[3]

Step	*Elements*	*Findings/Considerations*
History	Onset	Congenital, abrupt, gradual
	Pattern	Continuous, intermittent
	Duration	—
	Epidemiologic features	Travel, food, water
	Stool characteristics	Watery, bloody, fatty
	Fecal incontinence	—
	Abdominal pain	Occurs in inflammatory bowel disease, irritable bowel syndrome, ischemia
	Weight loss	May be severe in malabsorption or neoplasm
	Aggravating factors	Diet, stress
	Mitigating factors	Diet, over-the-counter drugs, prescription drugs
	Previous medical evaluation	—
	Iatrogenic diarrhea	From drugs, radiation, surgery
	Factitious diarrhea	Laxatives; may be surreptitious
	Systemic disease	Diarrhea may complicate hyperthyroidism, diabetes mellitus, collagen vascular disease, tumor syndromes, AIDS, immunoglobulin deficiencies
Routine laboratory tests	CBC	Anemia, leukocytosis
	Serum chemistry	Fluid/electrolyte status, nutritional status, serum protein/globulin
Stool analysis	Weight	—
	Electrolytes	For calculating fecal osmotic gap
	pH	Acid stools suggest carbohydrate malabsorption
	Stool leukocytes	Found in inflammatory diarrhea
	Fat output	Can be assessed by Sudan stain or quantitatively
	Laxative screen	
Categorization	Watery diarrhea (secretory or osmotic)	—
	Inflammatory diarrhea	
	Fatty diarrhea	

Stool tests Stool characteristics to measure include stool sodium and potassium concentrations, osmolality, and pH. Fecal occult blood testing and examination of stool for WBCs (or a surrogate chemical test, such as fecal lactoferrin concentration) should be conducted. Stool fat output should be measured quantitatively or assessed qualitatively with a Sudan stain of a fecal smear.

Measurement of stool electrolyte concentrations allows calculation of the fecal osmotic gap [*see* Fecal Osmotic Gap, *above*]. This can be used to identify watery diarrhea as being osmotic or secretory. Measurement of actual stool osmolality is only of value in detecting samples that have been contaminated with water or dilute urine and therefore have an osmolality less than 290 mOsm/kg. Stool osmolality rises rapidly in vitro because of bacterial fermentation, so the actual measurement should not be used to calculate the fecal osmotic gap. The pH of stool water can indicate whether or not carbohydrate malabsorption is present. Carbohydrates (or sugar alcohols) that are not absorbed in the small bowel and that reach the bacterial flora of the colon are fermented into short-chain fatty acids that reduce fecal pH, usually to less than 6. Thus, acid stools suggest carbohydrate malabsorption.[8]

Fatty diarrhea can be identified by measurement of stool fat, although careful interpretation of the results is sometimes necessary [*see* Steatorrhea, *below*]. When appropriate, a laxative screen should be obtained. Measurement of laxatives by chemical or chromatographic methods can detect surreptitious laxative ingestion.

Completion of the stool analysis allows the clinician to characterize chronic diarrhea as being watery (whether secretory or osmotic diarrhea), inflammatory, or fatty. The subsequent evaluation depends on this categorization.

Evaluation of Watery Secretory Diarrhea

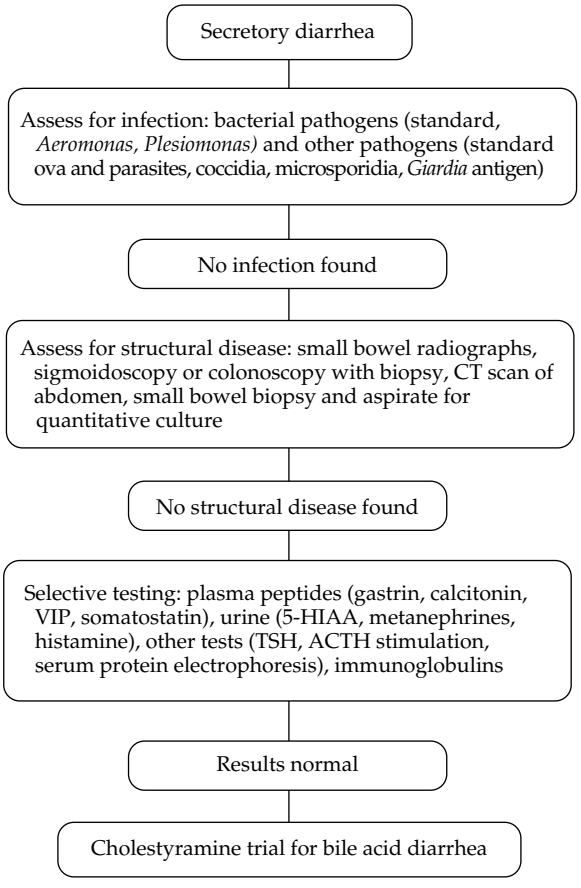
Secretory diarrhea is associated with many disorders; a thorough evaluation is therefore needed to identify the underlying cause [*see Figure 3*].

Stool tests Infection should be excluded by stool culture for bacteria, stool assay for *C. difficile* toxin, and other tests for parasites, including ELISA for giardiasis. Biopsies of the small bowel or colon may be necessary to find the pathogens, especially in patients with AIDS or other immunodeficiencies. Small bowel aspirate for quantitative culture is the best test for detecting small bowel bacterial overgrowth.

Imaging and endoscopic tests Structural diseases (e.g., short bowel syndrome or fistula, mucosal diseases, inflammatory bowel disease, and tumors) should be sought by radiographic and endoscopic testing. Small bowel radiography remains an important test in such cases. CT scans can detect small bowel and colonic disease, as well as problems extrinsic to the gut that may cause diarrhea, such as endocrine tumors. Inspection of the colonic mucosa by colonoscopy or sigmoidoscopy is essential in patients with secretory diarrhea, both to evaluate for gross changes and to obtain biopsy samples to look for evidence of microscopic colitis syndrome. Biopsies should be obtained even if the gross appearance of the colon is normal, because of the prevalence of microscopic colitis syndrome in patients with chronic watery diarrhea. A long endoscope that can reach the jejunum to obtain biopsy samples and aspirates is a valuable adjunct when other studies are unrevealing. The role of capsule endoscopy in the evaluation of patients with chronic diarrhea is under investigation; studies suggest that it may be helpful in detecting Crohn disease and, perhaps, celiac disease.[54-56] However, capsule endoscopy does not allow for biopsy of abnormalities that are visualized during the procedure, which limits its utility. Double-balloon enteroscopy offers the possibility of visualizing and obtaining biopsies from the entire small intestine; this technique may find a place in the evaluation of patients with watery secretory diarrhea.[57]

Serum peptide measurement Because diarrheagenic endocrine tumors are very rare, the measurement of serum peptides (e.g., gastrin, vasoactive intestinal polypeptide, calcitonin, and glucagons) or urinary excretion of secretagogue metabolites (e.g., 5-hydroxyindoleacetic acid or metanephrine) should be restricted to patients with symptoms consistent with tumor syndromes or those in whom a diagnosis remains elusive after initial testing.[58] More common endocrine problems, such as diabetes, hyperthyroidism, or Addison disease, should be excluded with appropriate blood tests.

Bile acid absorption measurement Ileal resection or ileal disease can result in the escape of sufficient bile acid into the colon to increase luminal bile acid concentrations above 3 to 5 mmol. At these concentrations, bile acids reduce colonic mucosal water and electrolyte absorption; alternatively, they stimulate secretion, resulting in increased stool water. In most circum-



ACTH—adrenocorticotropic hormone
TSH—thyroid-stimulating hormone
5-HIAA—5-hydroxyindole acetic acid
VIP—vasoactive intestinal peptide

Figure 3 **Evaluation of chronic secretory diarrhea.[3] Every test does not need to be done for every patient.**

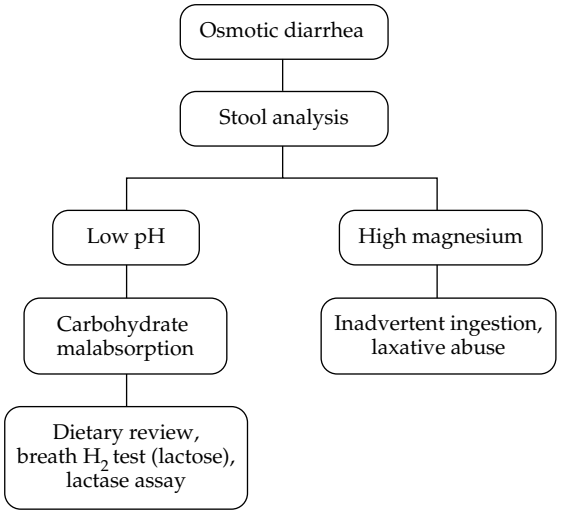


Figure 4 **Evaluation of chronic osmotic diarrhea.[3] Every test does not need to be done for every patient.**

stances, bile acid malabsorption can be inferred from a history of ileal resection or disease. More controversial is the concept that bile acid malabsorption occurring in the absence of ileal resection or obvious ileal disease is responsible for idiopathic secretory diarrhea.[59] Although bile acid malabsorption can be documented in many of these patients, administration of bile acid–binding resins does not always mitigate the diarrhea, casting doubts on bile acid malabsorption as the cause of the diarrhea.[49] Therefore, in patients with secretory diarrhea that appears to be idiopathic, it is more practical to give a therapeutic trial of bile acid–binding resins than to measure bile acid malabsorption directly.

Evaluation of Watery Osmotic Diarrhea

Because osmotic diarrhea has fewer potential causes than secretory diarrhea, the evaluation is simpler [*see Figure 4*]. If stool water has low electrolyte concentrations (and therefore a high fecal osmotic gap), some other substance is taking up the osmotic space and is holding water in the lumen. In practice, this substance is usually magnesium ingestion or carbohydrate malabsorption.

Magnesium ingestion Magnesium can be measured accurately in stool water. Excretion of more than 15 mmol (30 mEq) daily or concentrations greater than 45 mmol/L (90 mEq/L) strongly suggest magnesium-induced diarrhea.[60] This diarrhea may be intentional (surreptitious laxative ingestion) or accidental (magnesium-containing antacids or mineral supplements).

Carbohydrate malabsorption Carbohydrate malabsorption can occur from ingestion of poorly absorbable carbohydrates, such as lactose in someone with lactase deficiency, or from reduced carbohydrate absorption as a result of small bowel mucosal disease. In addition to ingestion of lactose, common causes of osmotic diarrhea include excessive ingestion of fructose (often used as a sweetener in commercial products),[61] ingestion of poorly absorbed sugar alcohols (such as mannitol and sorbitol, which are used as low-calorie sweeteners), and use of inhibitors of carbohydrate absorption, such as acarbose. Because malabsorbed carbohydrate is rapidly fermented by colonic bacteria, gas and bloating are frequent symptoms. Diagnosis is made on the basis of a finding of low stool pH (typically less than 6) and a thorough dietary history.

Evaluation of Chronic Inflammatory Diarrhea

Patients with WBCs or blood in the stool are classified as having inflammatory diarrhea. Causes may include inflammatory bowel disease, infections, ischemia, radiation enteritis, and neoplasia [*see Table 3*]. Sometimes, these conditions produce a watery, secretory diarrhea without blood or pus in the stool; therefore, they must also be considered in the evaluation of that type of diarrhea [*see* Evaluation of Watery Secretory Diarrhea, *above*].

Imaging and endoscopic tests Evaluation of patients with chronic inflammatory diarrhea should start with radiographic and endoscopic tests to look for structural problems [*see Figure 5*]. Sigmoidoscopy or colonoscopy should be considered first, because colitis is a common cause of inflammatory diarrhea. Biopsies should be performed to properly categorize colitis. CT has proved useful in many patients with inflammatory diarrhea because of the ability of CT to visualize inflammatory changes in the small bowel and colon and to identify complications of inflammation, such as abscess.

Infections that may produce chronic diarrhea, such as *C. difficile*, cytomegalovirus, amebiasis, and tuberculosis, need to be excluded by culture, biopsy, or serologic testing. It is important to realize that infection may complicate the courses of established problems, such as ulcerative colitis or Crohn disease. Patients with AIDS need an especially careful search for opportunistic infections.

Evaluation of Chronic Fatty Diarrhea

Steatorrhea Excessive fat in the stool, or steatorrhea, implies a problem with fat solubilization, digestion, or absorption in the small intestine. Steatorrhea is usually defined as stool fat output of more than 7 g over 24 hours or daily output of more than 9% of the intake of fat. These criteria may not be valid in patients with diarrhea, however, because voluminous stools per se may increase fat excretion. In one study, artificially induced diarrhea produced mild steatorrhea of up to 14 g/24 hr in 35% of normal persons.[62] Thus, in patients with diarrhea, fecal fat excretion of up to 14 g/24 hr has a low specificity for the diagnosis of defective fat absorption. The threshold for the diagnosis of steatorrhea also should be corrected for fat intake, because some patients

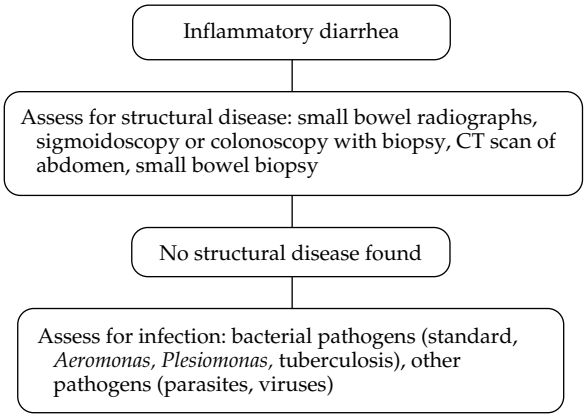


Figure 5 **Evaluation of chronic inflammatory diarrhea.[3] Every test does not need to be done for every patient.**

with diarrhea have anorexia and some patients with steatorrhea have hyperphagia. When possible, fat intake should be estimated from diet diaries that are maintained during the collection period. Finally, measurement of fat excretion can be compromised by ingestion of poorly absorbed fat substitutes, such as olestra.

Qualitative estimation of fat excretion by Sudan stain of a fecal smear can be used when a timed collection or quantitative analysis is not possible. Semiquantitative methods employing assessment of the number and size of fat globules correlate well with quantitative analysis of fat excretion.

The fecal fat concentration may provide a clue to the etiology of steatorrhea. The major causes of steatorrhea are mucosal diseases (e.g., celiac disease), pancreatic exocrine insufficiency (e.g., chronic pancreatitis), and lack of bile acids (e.g., advanced biliary cirrhosis). Mucosal diseases are often associated with reduced fluid and electrolyte absorption; as a result, fat is diluted by unabsorbed water. Furthermore, in mucosal disease, fat still can be digested to fatty acids, which can inhibit water absorption in the colon. In contrast, diseases that alter fat solubilization or digestion typically do not alter mucosal water and electrolyte absorption; as a result, unabsorbed fat is disbursed in a smaller stool volume. Fecal fat concentrations of more than 9.5 g/100 g strongly suggest pancreatic or biliary steatorrhea. Assessment of patients with chronic fatty diarrhea should therefore begin with measurement of fecal fat excretion and concentration [*see Figure 6*].

Imaging and endoscopic tests If the cause of steatorrhea is not obvious from the patient's history and the results of fecal fat assessment, the next step is evaluation of the absorptive surface of the small intestine by endoscopic, histologic, and radiographic tests. During endoscopy, small bowel biopsies should be obtained for histologic analysis, and small bowel contents should be aspirated for quantitative culture to assess for small bowel bacterial overgrowth. Indirect tests, such as measurement of antigluten (antiendomysial) antibodies or tissue transglutaminase antibodies for the diagnosis of celiac disease or breath tests for bacterial overgrowth, have not displaced endoscopic testing as the gold standard for diagnosis of these conditions. Such tests, however, may be useful in some cases. Small bowel radiography and CT are valuable adjuncts for structural assessment in patients with steatorrhea.

If the absorptive surface is normal, attention should shift to luminal problems with fat solubilization or digestion. Testing for pancreatic exocrine insufficiency is rarely done, because of unwillingness to use duodenal intubation tests. An indirect test, such as measurement of stool chymotrypsin activity, has limited sensitivity and specificity. The best test for pancreatic exocrine insufficiency may be a therapeutic trial of pancreatic enzyme supplementation. If this is done, a large dose of enzymes should be administered and objective measurement of fat excretion should be monitored to assess the response to therapy. Likewise, testing for the adequacy of bile salt solubilization of fat is rarely done. If necessary, duodenal bile salt concentration can be measured.

TREATMENT

Nonspecific Therapy

Nonspecific therapy is used in patients with chronic diarrhea in three situations: (1) as a temporizing or initial therapy before diagnostic testing, (2) after diagnostic testing has failed to result in a diagnosis, and (3) when a diagnosis has been made, but no specific treatment is available or specific treatment has failed.[22]

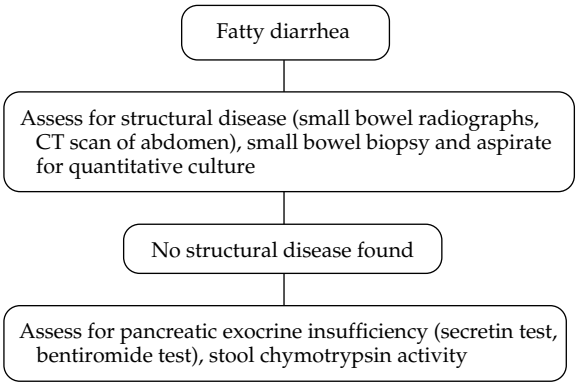


Figure 6 **Evaluation of chronic fatty diarrhea.[3] Every test does not need to be done for every patient.**

Antibiotic therapy Antibiotics are less useful in chronic diarrhea than in acute diarrhea because bacterial infection is less likely to be the cause of chronic diarrhea. Nevertheless, many clinicians try an empirical course of metronidazole or a fluoroquinolone before starting an extensive evaluation.

Symptomatic therapy with antidiarrheal drugs is often required in patients with chronic diarrhea [*see Table 2*]. Loperamide or diphenoxylate with atropine can be tried initially. In patients with chronic diarrhea, routine dosing (e.g., two tablets before each meal or at bedtime) is more effective than as-needed dosing after passing loose stools. More potent opiates, such as codeine, opium, and morphine, are underutilized in patients who do not respond to loperamide or diphenoxylate with atropine. Although these are controlled substances because of the possibility of abuse, abuse is unlikely in closely monitored patients with chronic diarrhea. Dosing should be started at a low level (e.g., codeine, 30 mg q.i.d.; deodorized tincture of opium, 3 drops q.i.d.; or morphine, 2 mg q.i.d.) and titrated up gradually to an effective dose. Stool-modifying agents, such as psyllium, can alter stool consistency but do not reduce stool weight. They may be of special help in patients with coexisting fecal incontinence.

Treatment of Specific Diseases and Syndromes

Osmotic diarrhea Osmotic diarrhea should abate with fasting or elimination of the offending agent from the diet. This response may be incomplete if other diarrhea-producing mechanisms are still active, such as short bowel syndrome or diseases of small bowel mucosa.

Irritable bowel syndrome and functional diarrhea Patients with chronic diarrhea in whom no other etiology is established are commonly diagnosed with irritable bowel syndrome or functional diarrhea. Irritable bowel syndrome is characterized chiefly by abdominal pain that is associated with altered bowel function, including constipation, diarrhea, or alternating diarrhea and constipation.[53] A diagnosis of functional diarrhea is made when patients do not have prominent abdominal pain and have no evidence of other specific causes of diarrhea. Obviously, these diagnoses are only as firm as the evaluation that is used to exclude other causes of diarrhea. For example, most cases of diarrhea from malabsorption of bile acid or carbohydrates are characterized as functional diarrhea or irritable bowel syndrome because specific testing for those malabsorption disorders is not done. Thus, careful consideration of alternative diagnoses should

precede a diagnosis of irritable bowel syndrome or functional diarrhea in patients with chronic diarrhea.

Nevertheless, there are certain clues to the diagnosis of irritable bowel syndrome or functional diarrhea that should be sought by the physician. Features that suggest a diagnosis of irritable bowel syndrome include a long history of diarrhea dating back to adolescence or young adulthood; passage of mucus; and exacerbation of symptoms with stress. Historical points that argue against irritable bowel syndrome include recent onset of diarrhea, especially in older individuals; nocturnal diarrhea; weight loss; blood in stools; voluminous stools (> 400 g/24 hr); and blood tests indicating anemia, leukocytosis, a low serum albumin concentration, or a high erythrocyte sedimentation rate.

New treatments for irritable bowel syndrome are being developed[63] in response to current theories about the pathogenesis of this disorder; pathogenetic processes that may yield new treatment strategies include dysregulation by the enteric nervous system,[64] food allergies,[65] small bowel bacterial overgrowth,[66] and changes in the colonic bacterial flora.[67]

Microscopic colitis syndrome Microscopic colitis syndrome, which subsumes the diagnoses of lymphocytic colitis and collagenous colitis, is a frequent cause of chronic diarrhea.[45,68,69] This disorder is characterized by chronic watery diarrhea and microscopic evidence of mucosal inflammation in the presence of normal gross colonoscopic findings. Histologic findings in both lymphocytic colitis and collagenous colitis include intraepithelial lymphocytic infiltration and chronic inflammation in the lamina propria without crypt destruction. Collagenous colitis and lymphocytic colitis are distinguished by the presence or absence of a thickened subepithelial collagen layer.

The cause of microscopic colitis syndrome is uncertain. It is associated with many autoimmune disorders and immunologically mediated diseases, such as celiac disease, which suggests that immune dysregulation is important. Bacterial antigens within the colonic lumen may also play a role. NSAIDs have been implicated in some reports.

Women are more likely than men to have collagenous colitis; lymphocytic colitis is equally likely in men and women. Diarrhea is of moderate severity (typically, 500 to 1,000 g/24 hr) and is characteristically secretory in nature because it results from failure of the colonic mucosa to absorb water and salt. Diagnosis is made by obtaining biopsy material from normal-appearing mucosa at the time of sigmoidoscopy or colonoscopy.

Treatment options include budesonide, bismuth subsalicylate, 5-aminosalicylate drugs, prednisone, and azathioprine.[45,68,69] Bile acid–binding drugs also have been reported to be successful in reducing diarrhea.[70] Microscopic colitis can have a remitting and relapsing course, and symptomatic therapy with opiate antidiarrheal drugs may be all that is needed. There is no evidence that microscopic colitis is a risk factor for colon carcinoma, and no surveillance program is currently recommended.

Laxative abuse Although rarely suspected, laxative abuse occurs regularly in four groups of patients: (1) those with anorexia or bulimia, (2) those who obtain a secondary gain from illness (e.g., disability payments, attention from relatives), (3) those with Munchausen syndrome, and (4) those who are dependent on others for their health care and who are poisoned by their caregivers (caregivers who do this are usually motivated by the desire to demonstrate their devotion to the patients).[71] Physicians need to consider surreptitious laxative abuse in patients who confound diagnosis and who are in one of the categories.

Detection of laxative abuse depends on having a high index of suspicion. Clues include the presence of hypokalemia in a patient who is able to eat (suggesting stimulant laxative abuse or concurrent ingestion of diuretics), melanosis coli (brownish pigmentation in the colonic mucosa caused by ingestion of anthraquinone laxatives) in a patient being evaluated for chronic diarrhea, or a large fecal osmotic gap (seen with magnesium ingestion). Most laxatives can be detected in stool water by chemical techniques. Adulteration of stool by added water or hypotonic urine can be detected by finding a low measured stool osmolality (< 280 mOsm/kg). The addition of hypertonic urine can be detected by impossibly high stool osmolality (> 600 mOsm/kg) and the presence of a negative fecal osmotic gap resulting from high urinary sodium or potassium concentrations. Negative fecal osmotic gaps may also be calculated in patients ingesting laxatives containing phosphate or sulfate.

Before patients are confronted with the diagnosis of laxative abuse, testing should be confirmed on another stool specimen, and appropriate psychiatric consultation should be available, because some of these patients become suicidal when confronted, and all of them need counseling. In cases of laxative poisoning by a caregiver, legal proceedings need to be instituted to separate the patient from the caregiver. Outcome studies in laxative-abuse patients are few. One study suggested that nearly half of the patients sought further medical attention elsewhere for chronic diarrhea.[72]

Postsurgical diarrhea Diarrhea can occur after several different kinds of operations. Peptic ulcer surgery is less common than it used to be, but new kinds of gastric operations, such as gastric bypass for obesity, produce similar complications. Dumping syndrome is the term used to describe a condition characterized by postprandial flushing, hypotension, diarrhea, and hypoglycemia.[73] This syndrome results from unregulated gastric emptying, osmotic shifts of fluid into the gut, and the rapid release of peptide hormones from the small intestine. Dumping syndrome can occur after vagotomy (intentional or accidental), pyloroplasty, gastrojejunostomy, and gastric resection. It can be treated with dietary modifications, antidiarrheal drugs [*see Table 2*], and the somatostatin analogue octreotide. Gastric surgery may also predispose patients to bacterial overgrowth in the small intestine, abnormally rapid intestinal transit, bile acid malabsorption, and pancreatic exocrine insufficiency from inadequate stimulation of the pancreas.

Bowel resection can result in loss of surface area sufficient to impair absorption of nutrients or water and salt. Lesser degrees of resection can result in diarrhea if an area of specialized function is removed.[74] For example, resection of the terminal ileum and right colon reduces bile acid absorption and the ability to absorb sodium against a large electrochemical gradient; these defects cannot be overcome by other areas of the intestine. With time, intestinal adaptation can overcome impaired electrolyte absorption, but intestinal adaptation cannot reverse loss of these specialized functions.

Ileostomy diarrhea is said to occur when stoma output exceeds 1,000 ml/24 hr. It may be caused by loss of absorptive surface area, if a substantial length of bowel has been resected; it may result from stomal stenosis, partial bowel obstruction, bacterial overgrowth, recurrent disease, medications, or intraperitoneal infection.[75] A special situation occurs in patients with ul-

cerative colitis who have had an ileoanal anastomosis with creation of an ileal reservoir pouch. These patients may develop inflammation of the pouch (so-called pouchitis) caused by bacterial overgrowth or recurrent inflammatory bowel disease.[76] Pouchitis can be treated with antibiotics such as metronidazole, anti-inflammatory drugs such as mesalamine, or ingestion of probiotic bacteria. Ordinary ileostomy diarrhea can be treated successfully with antidiarrheal opiate drugs.

Postcholecystectomy diarrhea occurs in as many as 20% of patients. It may be delayed in onset, and it is rarely severe. Diarrhea may occur as a result of ileal bile acid malabsorption at night, when the migrating motor complex may sweep bile acid past the absorptive sites in the terminal ileum, but some cases may have other causes.[77] Postcholecystectomy diarrhea is best treated with bile acid–binding agents given at bedtime. Opiate antidiarrheal drugs may be needed in refractory cases.

Diabetic diarrhea Up to 30% of patients with long-standing diabetes mellitus may experience chronic diarrhea.[78] This diarrhea has been attributed to autonomic neuropathy and dysregulation of motility, but definitive evidence of neuropathy is not always evident. If steatorrhea is present, three conditions that occur with increased prevalence in diabetics should be considered: (1) small bowel bacterial overgrowth, (2) pancreatic exocrine insufficiency, and (3) celiac disease. Other causes that need to be considered are medications, such as acarbose, and ingestion of dietetic foods containing sugar alcohols (e.g., sorbitol or mannitol).

When watery diarrhea is present, treatment with clonidine, an alpha$_2$-adrenergic agonist drug, may have special value. When clonidine cannot be tolerated because of its hypotensive effect or when it is not effective, opiate antidiarrheal drugs may be used. Fecal incontinence related to diabetic sensorimotor neuropathy may complicate diarrhea; this form of diarrhea needs to be evaluated, because therapies to mitigate incontinence, such as biofeedback training, may have a dramatic effect on quality of life.[79]

Diarrhea in patients with AIDS Diarrhea in AIDS patients is likely to result from opportunistic infections or lymphoma. A careful search for the cause of diarrhea can result in targeted therapy that may cure the diarrhea.[80] Colonoscopy is preferable to sigmoidoscopy because it allows visualization and biopsy of the right colon and ileum, which are often the sites of infection. It is possible that HIV-1 may directly produce diarrhea (so-called AIDS enteropathy), but in most cases, a specific infection can be identified.

Idiopathic secretory diarrhea The diagnosis of idiopathic secretory diarrhea can be made when an exhaustive evaluation fails to reveal a cause of chronic secretory diarrhea. This condition often begins suddenly in previously normal individuals, and it is distinguished from the acute secretory diarrhea by its persistence for more than 4 weeks. It occurs in two forms, epidemic and sporadic.

Epidemic idiopathic secretory diarrhea occurs in outbreaks that are seemingly related to contaminated food or water.[81] The initial description of this condition involved an epidemic of chronic diarrhea in Brainerd, Minnesota, and the condition has consequently become known as Brainerd diarrhea. Several outbreaks have been described in detail since the initial epidemic, and although the epidemiology suggests an infectious cause, no organism has been isolated.

Sporadic idiopathic secretory diarrhea affects individuals in an identical fashion as the epidemic form, but it does not seem to be acquired easily by family members or others.[52] Many patients describe a history of travel to local lakes or recreational sites, but they are the only members of their parties that become ill.

Both forms of idiopathic secretory diarrhea begin abruptly and reach maximum intensity shortly thereafter. Fever is unusual. Weight loss of up to 20 lb characteristically occurs in the first few months of the illness, but it does not become progressive thereafter. Empirical trials of antibiotics and bile acid–binding drugs are ineffective, but nonspecific opiate antidiarrheal drugs provide some relief. Idiopathic secretory diarrhea is self-limited and usually disappears within 2 years of onset. The offset of diarrhea is gradual, occurring over 2 to 3 months.

Diarrhea of obscure origin Diarrhea of obscure origin is said to be present when chronic diarrhea has evaded diagnosis in spite of an evaluation for structural problems. Patients are often referred commonly to centers interested in diarrheal diseases, where a specific cause for their diarrhea is often identified. Common diagnoses in these patients include fecal incontinence, drug-induced diarrhea, surreptitious laxative ingestion, microscopic colitis syndrome, bile acid–induced diarrhea, pancreatic exocrine insufficiency, carbohydrate malabsorption, sporadic chronic idiopathic secretory diarrhea, and, rarely, endocrine tumors. Most of these conditions can be recognized with a careful history, an appropriate index of suspicion, proper testing, or a well-conducted therapeutic trial. Failure to make a diagnosis is usually the result of not thinking through the differential diagnosis of chronic diarrhea and not appreciating the evidence at hand.

The author has received grants for educational activities from and served as an advisor for Novartis Pharmaceuticals Corp., GlaxoSmithKline, Romark Laboratories, Salix Pharmaceuticals, Inc., Santarus, Inc., and Takeda Pharmaceuticals North America, Inc.; has received grants for clinical research from Novartis Pharmaceuticals Corp. and GlaxoSmithKline; and has served as an advisor to TAP Pharmaceutical Products, Inc.

References

1. Talley NJ, Weaver AL, Zinsmeister AR, et al: Self-reported diarrhea: what does it mean? Am J Gastroenterol 89:1160, 1994

2. Thielman NM, Guerrant RL: Clinical practice. Acute infectious diarrhea. N Engl J Med 350:38, 2004

3. Fine KD, Schiller LR: AGA technical review on the evaluation and management of chronic diarrhea. Gastroenterology 116:1464, 1999

4. Wenzl HH, Fine KD, Schiller LR, et al: Determinants of decreased fecal consistency in patients with diarrhea. Gastroenterology 108:1729, 1995

5. Sellin JH: Intestinal electrolyte absorption and secretion. Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management, 7th ed. Feldman M, Friedman LS, Sleisenger MH, Eds. WB Saunders Co, Philadelphia, 2002, p 1693

6. Schiller LR, Sellin JH: Diarrhea. Gastrointestinal and Liver Disease: Pathophysiology/Diagnosis/Management, 7th ed. Feldman M, Friedman LS, Sleisenger MH, Eds. WB Saunders Co, Philadelphia, 2002, p 131

7. Fordtran JS: Pathophysiology of chronic diarrhoea: insights derived from intestinal perfusion studies in 31 patients. Clin Gastroenterol 15:477, 1986

8. Eherer AJ, Fordtran JS: Fecal osmotic gap and pH in experimental diarrhea of various causes. Gastroenterology 103:545, 1992

9. Guerrant RL, Van Gilder T, Steiner TS, et al: Practice guidelines for the management of infectious diarrhea. Clin Infect Dis 32:331, 2001

10. Brito GA, Alcantara C, Carneiro-Filho BA, et al: Pathophysiology and impact of enteric bacterial and protozoal infections: new approaches to therapy. Chemotherapy 51:23, 2005

11. Vazquez-Torres A, Fang FC: Cellular routes of invasion by enteropathogens. Curr Opin Microbiol 3:54, 2000

12. Crane JK: Preformed bacterial toxins. Clin Lab Med 19:583, 1999

13. Silletti RP, Lee G, Ailey E: Role of stool screening tests in the diagnosis of inflammatory bacterial enteritis and in selection of specimens likely to yield invasive enteric pathogens. J Clin Microbiol 34:1161, 1996

14. Savola KL, Baron EJ, Tompkins LS, et al: Fecal leukocyte stain has diagnostic value for outpatients but not inpatients. J Clin Microbiol 39:266, 2001

15. Bini EJ, Cohen J: Diagnostic yield and cost-effectiveness of endoscopy in chronic human immunodeficiency virus–related diarrhea. Gastrointest Endosc 48:354, 1998

16. Surawicz CM, Haggitt RC, Husseman M, et al: Mucosal biopsy diagnosis of colitis: acute, self-limited colitis and idiopathic inflammatory bowel disease. Gastroenterology 107:755, 1994

17. Desjeux JF, Briend A, Butzner JD: Oral rehydration solution in the year 2000: pathophysiology, efficacy and effectiveness. Baillieres Clin Gastroenterol 11:509, 1997

18. Wong CS, Jelacic S, Habeeb RL, et al: The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N Engl J Med 342:1930, 2000

19. Safdar N, Said A, Gangnon RE, et al: Risk of hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 enteritis: a meta-analysis. JAMA 288:1014, 2002

20. Huang DB, DuPont HL: Rifaximin: a novel antimicrobial for enteric infections. J Infect 50:97, 2005

21. Cohen SA: Use of nitazoxanide as a new therapeutic option for persistent diarrhea: a pediatric perspective. Curr Med Res Opin 21:999, 2005

22. Schiller LR: Antidiarrhoeal pharmacology and therapeutics. Aliment Pharmacol Ther 9:87, 1995

23. Allos BM: *Campylobacter jejuni* infections: update on emerging issues and trends. Clin Infect Dis 32:1201, 2001

24. Edwards BH: *Salmonella* and *Shigella* species. Clin Lab Med 19:469, 1999

25. Oldfield EC 3rd, Wallace MR: The role of antibiotics in the treatment of infectious diarrhea. Gastroenterol Clin North Am 30:817, 2001

26. House D, Bishop A, Parry C, et al: Typhoid fever: pathogenesis and disease. Curr Opin Infect Dis 14:573, 2001

27. Sandvig K: Shiga toxins. Toxicon 39:1629, 2001

28. Tarr PI, Neill MA: *Escherichia coli* O157:H7. Gastroenterol Clin North Am 30:735, 2001

29. Nakao H, Takeda T: *Escherichia coli* Shiga toxin. J Nat Toxins 9:299, 2000

30. Moyenuddin M, Williamson JC, Ohl CA: *Clostridium difficile*–associated diarrhea: current strategies for diagnosis and therapy. Curr Gastroenterol Rep 4:279, 2002

31. Voth DE, Ballard JD: *Clostridium difficile* toxins: mechanisms of action and role in disease. Clin Microbiol Rev 18:247, 2005

32. Bricker E, Garg R, Nelson R, et al: Antibiotic treatment for *Clostridium difficile*–associated diarrhea in adults. Cochrane Database Syst Rev (1):CD004610, 2005

33. Cremonini F, Di Caro S, Nista EC, et al: Meta-analysis: the effect of probiotic administration on antibiotic-associated diarrhoea. Aliment Pharmacol Ther 16:1461, 2002

34. Monkemuller KE, Wilcox CM: Investigation of diarrhea in AIDS. Can J Gastroenterol 14:933, 2000

35. Ziring D, Tran R, Edelstein S, et al: Infectious enteritis after intestinal transplantation: incidence, timing, and outcome. Transplantation 79:702, 2005

36. Schuster H, Chiodini PL: Parasitic infections of the intestine. Curr Opin Infect Dis 14:587, 2001

37. Ali SA, Hill DR: *Giardia intestinalis*. Curr Opin Infect Dis 16:453, 2003

38. Stauffer W, Ravdin JI: *Entamoeba histolytica*: an update. Curr Opin Infect Dis 16:479, 2003

39. Chappell CL, Okhuysen PC: Cryptosporidiosis. Curr Opin Infect Dis 15:523, 2002

40. Hammer HF, Santa Ana CA, Schiller LR, et al: Studies of osmotic diarrhea induced in normal subjects by ingestion of polyethylene glycol and lactulose. J Clin Invest 84:1056, 1989

41. Canani RB, Terrin G, Cirillo P, et al: Butyrate as an effective treatment of congenital chloride diarrhea. Gastroenterology 127:630, 2004

42. Lee SD, Surawicz CM: Infectious causes of chronic diarrhea. Gastroenterol Clin North Am 30:679, 2001

43. Attar A, Flourie B, Rambaud JC, et al: Antibiotic efficacy in small intestinal bacterial overgrowth–related chronic diarrhea: a crossover, randomized trial. Gastroenterology 117:794, 1999

44. Thielman NM, Guerrant RL: Persistent diarrhea in the returned traveler. Infect Dis Clin North Am 12:489, 1998

45. Schiller LR: Microscopic colitis syndrome: lymphocytic colitis and collagenous colitis. Semin Gastrointest Dis 10:145, 1999

46. Cappell M: Colonic toxicity of administered drugs and chemicals. Am J Gastroenterol 99:1175, 2004

47. Sellin JH, Hart R: Glucose malabsorption associated with rapid intestinal transit. Am J Gastroenterol 87:584, 1992

48. Camilleri M: Motor function in irritable bowel syndrome. Can J Gastroenterol 13(suppl A):8A, 1999

49. Schiller LR, Bilhartz LE, Santa Ana CA, et al: Comparison of endogenous and radiolabeled bile acid excretion in patients with idiopathic chronic diarrhea. Gastroenterology 98:1036, 1990

50. Warner RR: Enteroendocrine tumors other than carcinoid: a review of clinically significant advances. Gastroenterology 128:1668, 2005

51. Modlin IM, Kidd M, Latich I, et al: Current status of gastrointestinal carcinoids. Gastroenterology 128:1717, 2005

52. Afzalpurkar RG, Schiller LR, Little KH, et al: The self-limited nature of chronic idiopathic diarrhea. N Engl J Med 327:1849, 1992

53. Thompson WG, Longstreth GF, Drossman DA, et al: Functional bowel disorders and functional abdominal pain. Gut 45(suppl 2):II43, 1999

54. Petroniene R, Dubcenco E, Baker JP, et al: Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. Am J Gastroenterol 100:685, 2005

55. Kalantzis N, Papanikolaou IS, Giannakoulopoulou E, et al: Capsule endoscopy: the cumulative experience from its use in 193 patients with suspected small bowel disease. Hepatogastroenterology 52:414, 2005

56. Sturniolo GC, Di Leo V, Vettorato MG: Clinical relevance of small-bowel findings detected by wireless capsule endoscopy. Scand J Gastroenterol 40:725, 2005

57. Matsumoto T, Moriyama T, Esaki M, et al: Performance of antegrade double-balloon enteroscopy: comparison with push enteroscopy. Gastrointest Endosc 62:392, 2005

58. Schiller LR, Rivera LM, Santangelo WC, et al: Diagnostic value of fasting plasma peptide concentrations in patients with chronic diarrhea. Dig Dis Sci 39:2216, 1994

59. Brydon WG, Nyhlin H, Eastwood MA, et al: 7-Alpha-hydroxy-4-cholesten-3-one and selenohomocholyltaurine (SeHCAT) whole body retention in the assessment of bile acid induced diarrhoea. Eur J Gastroenterol Hepatol 8:117, 1996

60. Fine KD, Santa Ana CA, Fordtran JS: Diagnosis of magnesium-induced diarrhea. N Engl J Med 324:1012, 1991

61. Johlin FC Jr, Panther M, Kraft N: Dietary fructose intolerance: diet modification can impact self-rated health and symptom control. Nutr Clin Care 7:92, 2004

62. Fine KD, Fordtran JS: The effect of diarrhea on fecal fat excretion. Gastroenterology 102:1936, 1992

63. Schoenfeld P: Efficacy of current drug therapies in irritable bowel syndrome: what works and does not work. Gastroenterol Clin North Am 34:319, 2005

64. Gershon MD: Nerves, reflexes, and the enteric nervous system: pathogenesis of the irritable bowel syndrome. J Clin Gastroenterol 39(suppl 3):S184, 2005

65. Zar S, Benson MJ, Kumar D: Food-specific serum IgG4 and IgE titers to common food antigens in irritable bowel syndrome. Am J Gastroenterol 100:1550, 2005

66. Walters B, Vanner SJ: Detection of bacterial overgrowth in IBS using the lactulose H2 breath test: comparison with 14C-D-xylose and healthy controls. Am J Gastroenterol 100:1566, 2005

67. Malinen E, Rinttila T, Kajander K, et al: Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. Am J Gastroenterol 100:373, 2005

68. Pardi DS: Microscopic colitis: an update. Inflamm Bowel Dis 10:860, 2004

69. Chande N, Driman DK, Reynolds RP: Collagenous colitis and lymphocytic colitis: patient characteristics and clinical presentation. Scand J Gastroenterol 40:343, 2005

70. Ung KA, Gillberg R, Kilander A, et al: Role of bile acids and bile acid binding agents in patients with collagenous colitis. Gut 46:170, 2000

71. Ewe K, Karbach U: Factitious diarrhoea. Clin Gastroenterol 15:723, 1986

72. Slugg PH, Carey WD: Clinical features and follow-up of surreptitious laxative users. Cleveland Clin Q 51:167, 1984

73. Hasler WL: Dumping syndrome. Curr Treat Options Gastroenterol 5:139, 2002

74. Arrambide KA, Santa Ana CA, Schiller LR, et al: Loss of absorptive capacity for sodium chloride as a cause of diarrhea following partial ileal and right colon resection. Dig Dis Sci 34:193, 1989

75. Metcalf AM, Phillips SF: Ileostomy diarrhoea. Clin Gastroenterol 15:705, 1986

76. Heuschen UA, Allemeyer EH, Hinz U, et al: Diagnosing pouchitis: comparative validation of two scoring systems in routine follow-up. Dis Colon Rectum 45:776, 2002

77. Sauter GH, Moussavian AC, Meyer G, et al: Bowel habits and bile acid malabsorption in the months after cholecystectomy. Am J Gastroenterol 97:1732, 2002

78. Saslow SB, Camilleri M: Diabetic diarrhea. Semin Gastrointest Dis 6:187, 1995

79. Wald A: Incontinence and anorectal dysfunction in patients with diabetes mellitus. Eur J Gastroenterol Hepatol 7:737, 1995

80. Cohen J, West AB, Bini EJ: Infectious diarrhea in human immunodeficiency virus. Gastroenterol Clin North Am 30:637, 2001

81. Mintz E: A riddle wrapped in a mystery inside an enigma: Brainerd diarrhoea turns 20. Lancet 362:2037, 2003

82. Schiller LR: Chronic diarrhea. GI/Liver Secrets, 2nd ed. McNally PR, Ed. Hanley and Belfus, Philadelphia, 2002, p 411

83. Schiller LR: Diarrhea. Med Clin North Am 84:1259, 2000

IV INFLAMMATORY BOWEL DISEASES

Stephen B. Hanauer, m.d., f.a.c.p.

Ulcerative colitis (UC) and Crohn disease (CD) constitute the two major idiopathic inflammatory bowel diseases (IBDs). Gastroenterologists recognize that there is a spectrum of IBDs, encompassing varying types and degrees of intestinal inflammation, and that these idiopathic disorders must be distinguished from inflammation caused by infections, drugs (particularly nonsteroidal anti-inflammatory drugs [NSAIDs]), ischemia, and radiation. UC and CD are the most common and best understood IBDs, with a well-defined epidemiology and pathogenesis; their etiologies, however, remain elusive.[1]

Epidemiology

UC and CD share most epidemiologic characteristics.[2] These diseases are relatively common in developed nations and infrequent in countries with poor sanitation. In North America and Europe, the incidence is approximately five cases per 100,000 population for each disease, with a combined prevalence of approximately 100 per 100,000 population. The diseases can affect persons of any age, but onset most commonly occurs in the second and third decades of life. Much smaller, secondary peaks in incidence occur in the sixth and seventh decades. Males and females are affected equally. Risk of disease is higher in some ethnic groups than in others. Ashkenazi Jews have a higher risk of IBD than Africans, African Americans, and Asians; the incidence of IBD increases in these lower-risk groups, however, when they emigrate to developed nations or adopt Western culture and diet.

Etiology

The cause of IBD has yet to be established. IBD probably has multiple causes, involving an interplay between genetically mediated susceptibility, environmental factors, and aberrant immune function. These disorders may well prove to be a series of syndromes with overlapping features.[1] As yet, no dietary factor has been identified, despite case-control studies suggesting a possible association with the ingestion of large amounts of refined sugar and, perhaps, fat (essentially, the Western diet).[3] Extrapolation from animal models of IBD suggests that commensal (rather than pathogenic) flora can initiate inflammation in genetically susceptible individuals,[1,4] but no specific infectious agent has been identified as causative of either UC or CD.

There are two major clues to the etiology of IBDs. The first is the familial association of IBD, which suggests a genetic predisposition.[5] Both UC and CD are more common in families with an affected relative. Once a proband has been identified, the risk of the disease occurring in a second family member is approximately 20% (40% if the proband is a child). The risk is distributed throughout families, with an estimated risk for individuals of 3% to 5% spread among first-, second-, and third-degree relatives. If both parents have IBD, the risk of disease in an offspring is nearly 50%. There is a concordance for disease type (and subtype, in the case of CD) within families, although either UC or CD may be seen.[5] Risk of disease is highest in identical twins, with a concordance of 20% for UC and 60% for CD. This degree of risk suggests a polygenic causation with higher penetrance in CD.[6] Genetic loci for IBD have been found: for example, the *NOD2/CARD15* gene on chromosome 16 (*IBD1*) increases susceptibility to CD; other potential loci are being investigated.[5]

The second etiologic clue is the relationship between cigarette smoking and UC and CD.[7] Case-control studies from around the world have demonstrated that cigarette smokers are less likely to develop UC and more likely to develop CD. In contrast, ex-smokers are more likely to develop UC. Cigarette smoking also influences the course of IBD; for example, ex-smokers with UC are more likely to have refractory disease and to require surgery than are patients who have never smoked.[8] Cigarette smoking also protects against the development of primary sclerosing cholangitis associated with UC.[9] Conversely, current cigarette smokers with CD are more likely to have disease that is refractory to medical therapies and that recurs more rapidly after surgical resection; these effects can be reversed by smoking cessation.[10] It has not been established whether nicotine is the primary factor in the associations of cigarette smoking with UC and CD. Nicotine delivered by transdermal patch appears to have a modest therapeutic potential for UC,[11] although not as much as resumption of smoking.

Pathogenesis

There are numerous animal models for IBD, including the cotton-top tamarin, a New World monkey that, in captivity, develops a so-called spontaneous colitis with many features that are similar to those of human UC. These shared features include the development of antiepithelial antibodies, response to anti-inflammatory medications, and the development of dysplasia and adenocarcinoma.

Other animal models more closely mimic CD. These include genetically engineered transgenic rats that overexpress human HLA-B27 and β_2-microglobulin molecules and knock-out mice with targeted deletions of interleukin 2 (IL-2), IL-10, T cell receptor chains, and transforming growth factor–β. The transfer of enriched populations of functional T helper type 1 (Th1) lymphocytes into severe combined immunodeficient mice induces colitis and wasting that can be prevented by transfer of unfractionated $CD4^+$ T cells.[1]

Regardless of species or immune status, animals raised in germ-free environments do not develop intestinal inflammation.[12] The significantly higher rate of autoimmune diseases in developed versus developing countries suggests that infectious agents (bacterial, viral, and parasitic) act as antigenic triggers and that exposure to numerous microorganisms early in life may protect against autoimmunity later in life; this is the so-called hygiene hypothesis.[13] The loss of immune tolerance to specific (bacterial) antigenic triggers has been demonstrated in patients with CD.[14] Hypothetically, environmental factors interact with a genetically determined, defective innate immune response. The innate immune defects may involve dendritic cells[15] and Toll-like receptors[16] [*see 6:II Innate Immunity*]; in addi-

tion, bacterial invasion of the mucosa may be facilitated by impairment of the mucosal protection mediated by mucins, trefoil peptides, or defensins.[17] Subsequently, the onset of IBD symptoms may be triggered by acute infections (e.g., traveler's diarrhea or acute gastroenteritis), antibiotic exposure, or other environmental factors in genetically susceptible hosts. Whether a trigger (specific or nonspecific) modifies the disease subtype cannot as yet be discerned.

Once intestinal inflammation begins, the primary difference between patients with IBD and unaffected persons is an impaired ability to downregulate mucosal inflammation.[1] Chronic inflammatory cells are normal in the intestinal mucosa; they comprise the gut component of the mucosa-associated lymphoid tract. The number of lymphoid elements in the mucosa is proportional to enteric exposure to bacteria. Persons raised in a sanitary environment (as is typical of developed countries) have less chronic inflammation than those raised in countries with poor sanitation. An extreme example is that of tropical sprue, in which mucosal inflammation is extensive and is associated with atrophy and ulceration of the small bowel villi.

In IBD, most immune elements (including tissue macrophages and mucosal T cells) respond to an exaggerated degree when triggered by an antigen. Activated macrophages and T cells are prominent in the recruitment of nonspecific inflammatory cells—primarily neutrophils, which are the final mediators of tissue damage.[1] The cytokine responses in UC and CD seem to differ, which may account for differences in disease phenotypes.[18] Many studies have demonstrated increases in levels of IL-1, IL-6, and tumor necrosis factor–α (TNF-α) in the mucosa of patients with UC and CD, although it is becoming apparent that the balance of cytokines may be different in the two diseases. CD is marked by a higher Th1 cytokine profile (interferon gamma, IL-2, IL-12, and TNF-α), whereas in UC, the balance is more consistent with a Th2 profile, with increased proportions of mucosal B cells, plasma cells, and antibodies.[1] In UC, increased production of both antineutrophil cytoplasmic antibodies and IgG antibodies reacting with a 40 kd tropomysin protein have been identified, although the pathogenic consequences of this interaction have not been defined.[19] Conversely, patients with CD have a greater likelihood of developing antibodies to a common brewer's yeast (*Saccharomyces cerevisiae*).[19]

A final pathway of tissue destruction is through the recruitment and activation of macrophages and neutrophils.[20] Activation of the arachidonic acid cascade leads to increased tissue levels of cyclooxygenase products (prostaglandins and thromboxanes), lipoxygenase products (primarily leukotriene B_4), and platelet activating factor. These compounds and other nonspecific mediators (e.g., nitric oxide, neutrophil tissue proteases, and reactive oxygen species) contribute to tissue destruction and can be targeted for specific and nonspecific anti-inflammatory therapy.[1,4]

Ulcerative Colitis

CLASSIFICATION

UC is marked by diffuse, superficial inflammation of the colonic mucosa, beginning in the rectum and extending proximally to involve any contiguous length of colon. The small intestine is not involved, except in the setting of extensive colitis, in which the most distal terminal ileum may exhibit similar superficial inflammation, termed backwash ileitis. Because the extent of colitis usually remains constant from the onset, the length of involved colon defines the classification of UC: proctitis (limited to the rectum), proctosigmoiditis or left-sided colitis (extending up to the splenic flexure), or pancolitis (extending into the transverse colon). Proximal extension occurs in approximately one third of patients with distal disease, and regression from pancolitis is also possible.[21,22] The extent of involvement does not necessarily imply severity but does pertain to prognosis (e.g., the risk of cancer) and to treatment selection. The symptoms and course of UC relate to both the extent and the severity of inflammation within the involved segment of colon.

DIAGNOSIS

The diagnosis of UC is made on the basis of clinical, endoscopic, and histologic findings. The presence of rectal bleeding or diarrhea should raise the suspicion of UC. Symptoms are often chronic, but they may also be intermittent or progressive. The easiest way to exclude UC is by direct examination of the rectosigmoid colon with a proctoscope or flexible sigmoidoscope. Radiography (barium enemas) has been almost completely replaced by more sensitive endoscopic examinations. Because UC always involves the rectum, inflammatory changes should be visible with a limited examination. In newly diagnosed patients, stool cultures are performed to rule out infectious diseases that may mimic or complicate UC, such as infection with *Salmonella, Shigella, Campylobacter,* hemorrhagic *Escherichia coli,* and *Clostridium difficile*.

Clinical Manifestations

The onset of UC typically is insidious rather than abrupt, although the disease occasionally presents acutely after infectious colitis or traveler's diarrhea.[23] Rectal bleeding is the most consistent feature. Bleeding may be gross or may be noted with evacuation of mucopus. Associated rectal urgency and tenesmus are related to diminished compliance of the rectum. Diarrhea, distinguished from the passage of mucopus without stool, relates to the extent of colonic involvement. Patients with proctitis often present with constipated bowel movements with interim passage of blood or mucus. Abdominal cramps preceding bowel movements are common, although abdominal pain or tenderness (related to transmural inflammation) signifies progressive, severe disease. In severe or fulminant colitis, systemic symptoms of night sweats, fever, nausea and vomiting, and weight loss accompany diarrhea. Extraintestinal manifestations can include inflammation of the eyes, skin, joints, and liver.

Ulcerative proctitis In the most common variant of UC, accounting for approximately 25% to 30% of cases and usually being the mildest, inflammation is limited to the distal 15 to 20 cm of rectum. Patients with ulcerative proctitis typically present with hematochezia, a sense of rectal urgency, and constipated bowel movements because of delayed transit of fecal material in the right colon. Systemic manifestations are uncommon, but skin or joint symptoms can occur.[24] The disease usually remains confined to the rectum but may advance proximally in as many as 30% to 40% of patients.[21]

Proctosigmoiditis Left-sided colitis is an intermediate syndrome of UC, accounting for about one third of cases. Patients present with either constipation or diarrhea accompanied by tenesmus, urgency, and rectal bleeding. Left lower quadrant

cramping abdominal pain is more common than with proctitis, as are extraintestinal symptoms. The proximal disease margin usually remains fixed throughout the course but can spread more proximally or even retract distally.

Extensive colitis (pancolitis) In pancolitis, inflammation extends into the transverse or right colon. Patients are more likely to present with diarrhea because of diminished absorptive capacity of the colon, accompanied by rectal bleeding and urgency. Abdominal cramps may be diffuse or localized, and patients are more likely to have weight loss, systemic or extraintestinal symptoms, and anemia.

Toxic megacolon Toxic megacolon refers to the most severe manifestation of UC, which occurs when the inflammation extends from the superficial mucosa into the submucosa and muscular layers of the colon.[25] Toxic megacolon occurs more commonly with extensive colitis but can also occur with severe distal colitis. The colonic wall becomes tissue-paper thin as the colon dilates and becomes hypomotile; perforation may occur. The patient often has fever, prostration, severe cramps, abdominal distention, and abdominal tenderness; the tenderness may be localized, diffuse, or rebound.

Clinical Severity

The severity of UC depends on both the length of colon involved and the severity of colonic inflammation. The set of criteria most commonly used to define the severity of disease was created by Truelove and Witts.[26-28] Although the criteria were developed to assess improvement in clinical trials, they remain useful in classifying severity in clinical practice, and they have been modified to include fulminant colitis [*see Table 1*]. Severity criteria are as follows:

Mild In mild UC, patients have less than four bowel movements daily, with minimal cramps and urgency. Usually, most of the bowel movements occur early in the day; and after the morning evacuations, the patient is able to proceed with activities of daily life.

Moderate Patients with moderate UC have four to eight bowel movements daily, more frequent rectal urgency, and postprandial cramping and bowel movements. Blood is present in most stools, and nocturnal wakening for bowel movements is common. The disease can interfere with daily work or school activities and social life.

Severe Patients with severe UC have more than eight bowel movements daily, nocturnal bowel movements, severe urgency with or without incontinence, and systemic signs that include low-grade fever, night sweats, weakness, and weight loss. Abdominal tenderness, tachycardia, anemia, leukocytosis, and hypoalbuminemia are common.

Fulminant Patients with fulminant colitis have more than 10 bowel movements a day, nocturnal bowel movements, severe abdominal pain or relentless tenesmus, and rebound tenderness or distention with tympanic bowel sounds. They also have prostration, high fever, and hypotension. Radiographic studies show evidence of mucosal edema, intramural air (pneumatosis coli), colonic dilatation (toxic megacolon), or free abdominal air (perforation).

Physical Examination

In most patients with UC, the physical examination results are normal. There may be mild abdominal tenderness to deep palpation, particularly in the left colon, but significant abdominal findings are limited to patients with moderate to severe disease, in whom tenderness is more prominent. Surprisingly, despite the frequent diarrhea, perianal manifestations are absent. Any significant perianal findings (e.g., large hemorrhoids, skin tags, fissures, abscesses, or fistulas) suggest CD rather than UC.

Table 1 Classification of Ulcerative Colitis

Feature	*Mild Disease*	*Moderate Disease*	*Severe Disease*	*Fulminant Disease*
History				
Stools	< 4/day (usually early in the day); minimal abdominal cramps and urgency	4–8/day; abdominal cramps and urgency	> 8/day; nocturnal bowel movements; severe urgency with or without incontinence	> 10/day; nocturnal bowel movements; severe urgency; tenesmus; severe abdominal pain
Blood in stool	Intermittent	Frequent	Frequent	Continuous
Physical examination				
Abdominal findings	Nontender abdomen; normal bowel sounds	Nontender or minimally tender over sigmoid	Tender abdomen; no rebound tenderness	Distended abdomen; decreased bowel sounds; rebound tenderness
Temperature	Normal	Normal	> 37.5° C (99.5° F)	> 37.5° C
Pulse	Normal	Normal	> 90	> 90
Other physical findings	Normal	Pallor	Weakness, pallor, weight loss	Prostration, hypotension
Laboratory tests				
Hemoglobin	Normal	Mild anemia	< 75% of normal	Transfusion required
Erythrocyte sedimentation rate	< 30 mm/hr	< 30 mm/hr	> 30 mm/hr	> 30 mm/hr
Radiography	Normal gas pattern	Absent stool in involved segment	Edematous colon wall, thumbprinting	Dilated colon, mucosal edema, intramural air, free abdominal air

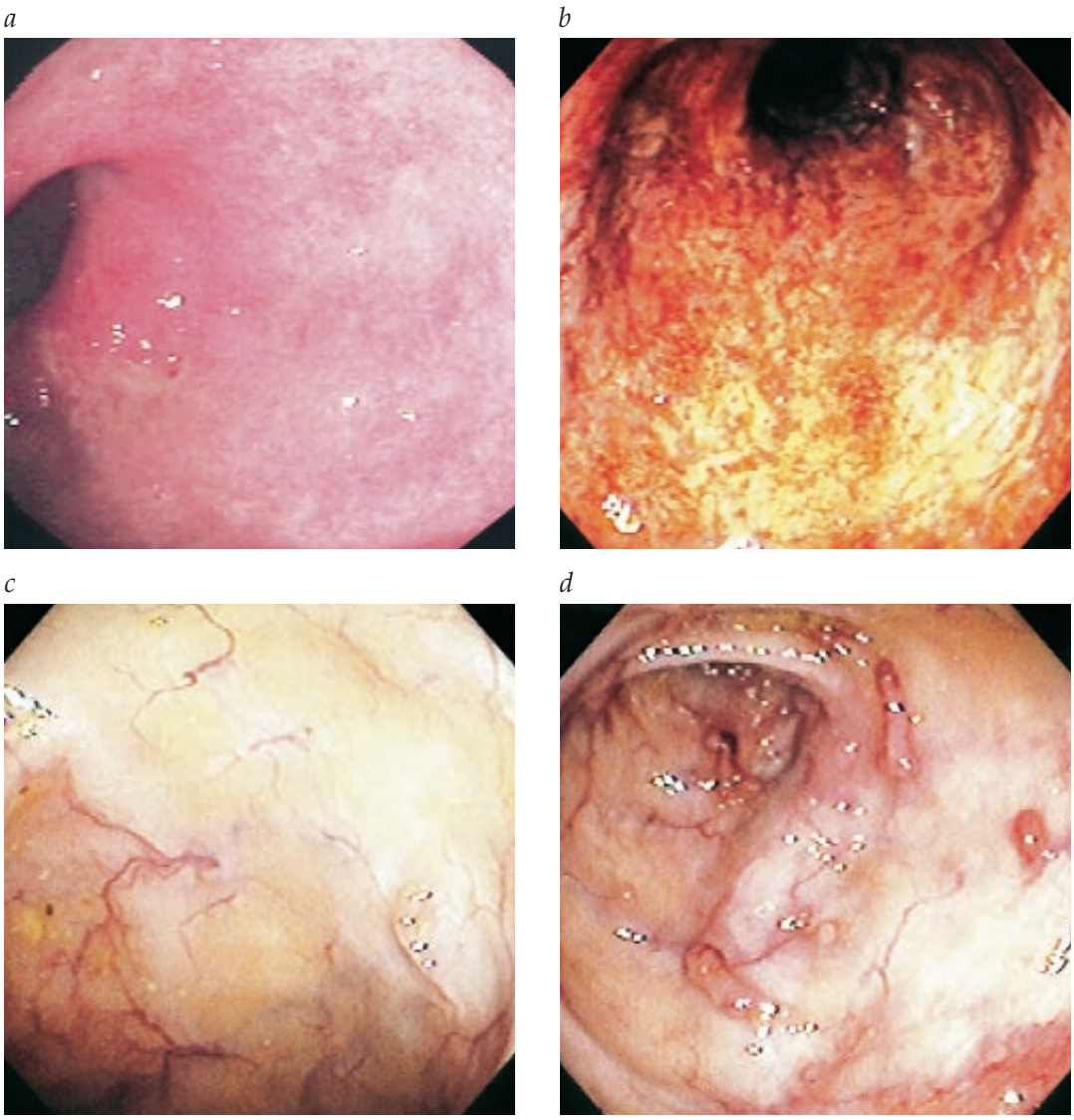


Figure 1 **The endoscopic spectrum of ulcerative colitis includes (*a*) mucosal edema, erythema, loss of vasculature; (*b*) granular mucosa with pinpoint ulceration and friability; (*c*) regenerated (i.e., healed) mucosa with distorted mucosal vasculature; and (*d*) regenerated mucosa with typical postinflammatory pseudopolyps.**

Conjunctival pallor is common because of anemia. Patients may present with ocular inflammation, erythema nodosum, pyoderma gangrenosum, or arthritis of larger joints.[29] Low back pain with diminished range of motion, or sacroiliac tenderness, is uncommon. Hepatomegaly, splenomegaly, or evidence of chronic liver disease is rare and limited to patients with end-stage primary sclerosing cholangitis.

Endoscopy

UC can almost always be diagnosed by endoscopic examination of the rectum and sigmoid colon.[30] The disease presents as diffuse and continuous inflammation beginning in the rectum, with proximal extension that varies among individual patients. In most cases, it is advisable to determine the proximal margin of disease, from the standpoints of prognosis[24] and therapy.[26,27] The initial examination should be performed without enema preparation, to avoid confusion with trauma or inflammation from administration of the enema. Patients with active colitis rarely have any fecal material in the involved lumen.

Healthy colonic mucosa is smooth and glistening, reflects light back from the scope, and demonstrates a branching mucosal vascular pattern. With inflammation, the mucosa becomes erythematous and more granular, which breaks apart the light reflection, and the vascular pattern becomes obscured by edema [*see Figure 1*]. The granularity of the mucosa may be fine or coarse. Coarse granularity represents microscopic or pinpoint ulcerations and is associated with friability (hemorrhage from the mucosa that may be spontaneous or induced by scope trauma). Exudates of mucopus are a common associated finding. Gross ulcerations represent more severe disease and, although usually shallow, can progress to a total denudation of the mucosa with exposure of the underlying circular musculature. These changes can be continuous up to a distinct margin where the mucosa appears normal, or they may extend diffusely to the cecum and, occasionally, into the distal ileum (backwash ileitis). In the setting of pancolitis, the ileocecal valve is usually wide open (patulous), allowing easy entry into the terminal ileum. Some patients with distal UC (involving the rec-

tum or sigmoid) also may have limited inflammatory changes around the appendix in the cecum (cecal red patch).[31]

As ulcerative colitis heals, the mucosal changes may become more focal.[32] The colonic mucosa regenerates from ulceration to granularity, with gradual restitution of a distorted mucosal vascular pattern with less distinct branching or irregular, pruned-appearing vessels. In areas that had been more severely inflamed, granulation tissue may protrude and become reepithelialized as so-called pseudopolyps. These postinflammatory changes can arise in a variety of sizes and shapes and are more likely to become fingerlike projections, or even mucosal bridges, in areas that had severe, undermining ulcerations [*see Figure 1*]. Pseudopolyps have no neoplastic potential but can be difficult to differentiate from adenomatous polyps. When pseudopolyps are extensive, they can totally carpet the mucosa, making it impossible to discern distinct, potentially neoplastic polyps.

Histology

Samples for histologic analysis are typically taken during endoscopy. The histologic features of UC parallel the endoscopic appearance of diffuse, continuous mucosal inflammatory changes. The principal components are disruption of glandular architecture and an inflammatory infiltrate.[33] A hallmark distinction between chronic IBDs such as UC and acute self-limited (infectious) colitis is architectural distortion [*see Figure 2*]. In UC, the normal vertical (so-called test-tube) alignment of glands is distorted; the glands are separated by expanded lamina propria lymphocytes, plasma cells, and eosinophils, as well as by neutrophils, which normally are sparse. The glands themselves become irregular in shape and, often, branched. The neutrophil infiltrate is localized to the base of the glandular crypts and invades the crypts, producing crypt abscesses. In more severe disease, the epithelial lining is destroyed, leaving ulcerations over the lamina propria. The inflammatory changes are usually superficial, limited by the muscularis mucosae. Despite severe superficial changes, deeper inflammation is uncommon, except in the setting of fulminant colitis. In fulminant colitis, the muscular layers are breached by expanding inflammatory ulceration that can leave the bowel wall tissue-paper thin and protected only by the serosa.

As the mucosa heals, the glands may become atrophied or shortened and irregularly shaped, with a thinned-out lamina propria. Inflammatory polyps are composed of vascular granulation tissue with a thin colonic epithelium. In quiescent colitis, the architectural distortion is present but acute inflammation (neutrophils and crypt abscesses) is absent [*see Figure 2*].

Both acute inflammation and regeneration of the colonic epithelium produce cellular atypia that must be distinguished from epithelial dysplasia.[34] In regenerating mucosa, the glandular epithelium can become irregularly shaped and hyperchromatic, with depletion of normal apical mucus. Stratification of nuclei and loss of polarity are manifestations of neoplastic transformation (i.e., dysplasia).

Imaging Studies

Radiography Radiographs have largely been supplanted by endoscopic examinations for the diagnosis of UC, but radiography remains a valuable adjunct to endoscopy in specified clinical situations. Plain abdominal radiographs are useful in the setting of severe colitis. These examinations outline the air-filled colon and can demonstrate the presence or absence of haustrations or dilatation of the colon (to rule out toxic megacolon) [*see Figure 3*]. Extraluminal gas under the diaphragm (free air) and evidence of an ileus pattern are additional features of severe colitis. Plain abdominal radiographs also provide a view of the sacroiliac joints and lumbosacral spine as a gross assessment of sacroiliitis or ankylosing spondylitis for patients presenting with low back pain.

Contrast barium studies, once the primary diagnostic modality, are less commonly employed because endoscopic examinations provide higher diagnostic sensitivity and specificity and

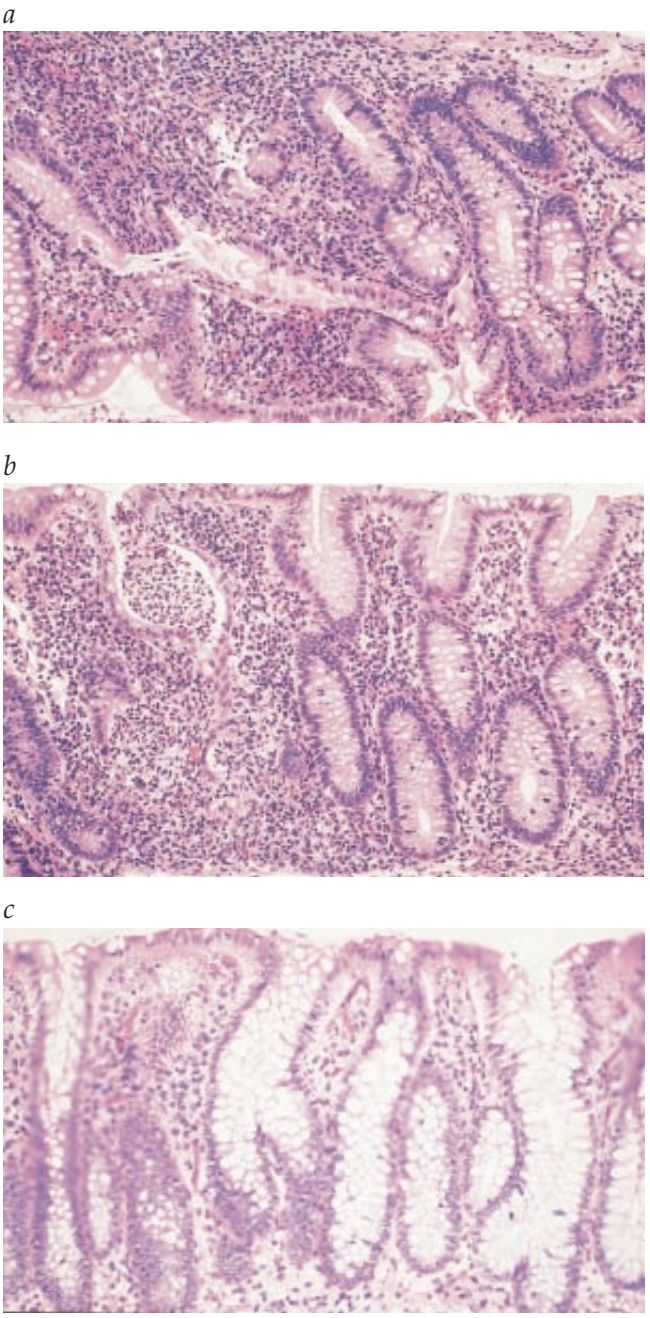


Figure 2 **Pathologic changes in ulcerative colitis include (*a*) acute superficial inflammation with distortion of the normal crypt architecture, (*b*) crypt abscesses, and (*c*) quiescent colitis without acute inflammation but with distortions of crypt architecture (abnormal branched crypts).**

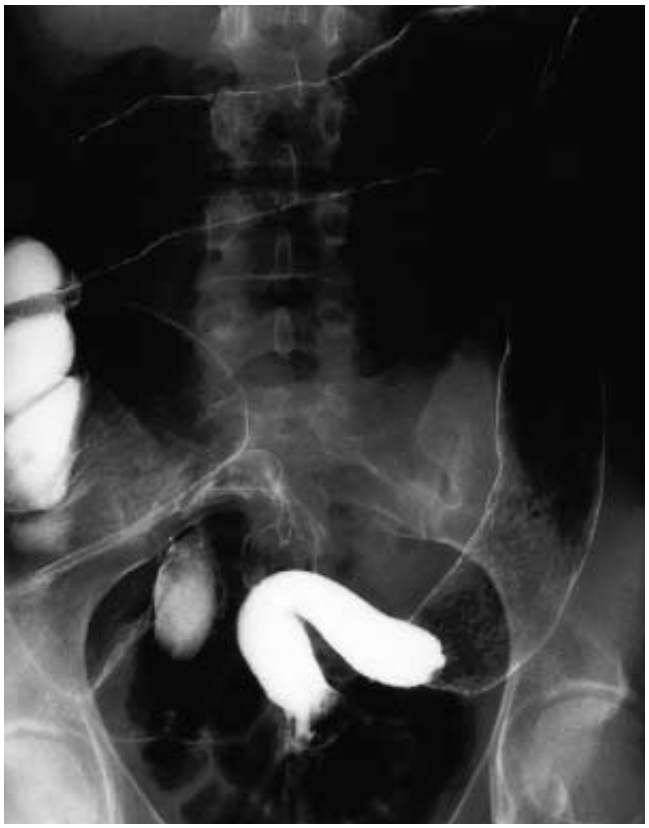


Figure 3 **In this air-contrast radiograph of ulcerative colitis, the mucosal pattern is granular with loss of normal haustrations in a diffuse, continuous pattern.**

permit histologic sampling.[26] Air-contrast barium enemas demonstrate the fine or coarse mucosal granularity of microscopic ulcerations or the diffuse, continuous, and symmetrical pattern of ulceration involving the rectum to the proximal extent of disease [*see Figure 3*]. Other features of UC that may be visible on barium enemas are the loss of haustration in inflamed segments, foreshortening of the colon, and an increase in the space between the sacrum and the rectum. Barium enema examinations are contraindicated in severely ill patients because of the potential for perforation or the induction of a toxic megacolon.

Scintigraphy Scanning with indium-labeled or technetium-labeled leukocytes is occasionally indicated for severely ill adults or children when the extent of colitis is uncertain or when small bowel disease has not been excluded.[35] These studies provide relatively rapid determination of the extent, severity, and continuity of intestinal inflammation. Scintigraphy is noninvasive, is sensitive and specific for intestinal inflammation, and is occasionally helpful in discriminating UC from CD.

Crohn Disease

CD is manifested by focal, asymmetrical, and transmural inflammation of the digestive tract, at times accompanied by granuloma formation. In contrast to the inflammation of UC, which is diffuse, continuous, superficial (mucosal), and typically limited to the colon, the inflammation of CD is more patchy, may be transmural, and can involve any segment of the gastrointestinal tract from mouth to anus. Because the inflammation may be transmural, CD can lead to intestinal complications of stenoses (strictures) and fistulas. Although a hallmark of CD is the histologic finding of noncaseating granulomas, these granulomas are identified in only about 30% of patients and are not necessary to make the diagnosis.

Because CD may involve any segment of the gastrointestinal tract, the presentation is more heterogeneous than that of UC and is determined by the location, extent, severity of inflammation, and inflammatory pattern. The location and pattern tend to remain constant for each patient.[36] CD produces a spectrum of inflammatory patterns: from superficial inflammation similar to that of UC, to formation of fibrostenosing strictures, to penetration of the bowel wall and fistula formation accompanied by a mesenteric inflammatory mass or perienteric abscess. An attempt to classify CD on the basis of inflammatory patterns[37] has been compromised by the tendency of inflammatory patterns to progress to stenoses over time.[38] In contrast to UC, CD is usually not curable by surgery; intestinal resection and anastomosis are almost inevitably followed by recurrence of the disease involving the anastomotic site and proximal intestine.[39]

Table 2 Key Distinguishing Features of Ulcerative Colitis and Crohn Disease

Feature	*Ulcerative Colitis*	*Crohn Disease*
History		
Smoking status	Nonsmoker or ex-smoker	Smoker
Physical examination		
Symptoms	Rectal bleeding, cramps	Diarrhea, abdominal pain, weight loss, nausea, vomiting
Signs	Normal perianal findings, no abdominal mass	Perianal skin tags, fistulas, abscesses; abdominal mass; clubbing of digits
Laboratory tests		
Endoscopy	Rectal involvement; continuous superficial inflammation with granular, friable mucosa; terminal ileum normal or showing backwash ileitis	Rectal sparing; local ulceration with normal intervening mucosa; aphthous, linear, or stellate ulcers; terminal ileum inflamed with aphthous or linear ulcers
Radiology	Diffuse, continuous superficial ulceration; ahaustral (lead-pipe) colon; backwash ileitis	Focal, asymmetrical, transmural ulceration; strictures, inflammatory masses, fistulas; small bowel disease
Histology	Diffuse, continuous, superficial inflammation; crypt architectural deformity	Focal inflammation, aphthous ulcers, lymphoid aggregates, transmural inflammation, granulomas (15%–30% of patients)
Serology	Elevated p-ANCA (60%–80% of patients)	Elevated ASCA (~ 30% of patients)

p-ANCA—perinuclear antineutrophil cytoplasmic antibody ASCA—anti–*Saccharomyces cerevisiae* antibody

In clinical trials, the instrument most commonly used to quantify disease activity has been the CD Activity Index (CDAI).[40] However, because of its complex derivation and lack of discrimination between symptoms and inflammation, the CDAI is not used in clinical practice. Instead, patients require individualized assessments of the severity of disease according to inflammatory symptoms, obstruction, fistulization, abscess formation, systemic complications, and effect on the patient's quality of life.[41]

DIAGNOSIS

CD is diagnosed on the basis of clinical, radiographic, endoscopic, and histologic criteria. As with UC, there is no pathognomonic marker. The clinical presentation and key features of the history, physical examination, and laboratory studies determine the diagnostic workup and serve to differentiate CD from UC [*see Table 2*].

Clinical Manifestations

CD most commonly involves the terminal ileum and cecum. However, the pattern of CD can be quite varied [*see Figure 4*]. The presentation depends on the site, extent, severity, and complications of intestinal and extraintestinal disease.[42,43] Patients usually present with chronic disease, but CD can be acute, with severe abdominal pain, intestinal blockage, or hemorrhage. Abdominal pain is a more common feature of CD than of UC because the transmural extension of CD results in stimulation of pain receptors in the serosa and peritoneum. Abdominal cramping and postprandial pain are common symptoms that often are accompanied by diarrhea, rectal bleeding, nocturnal bowel movements, fevers, night sweats, and weight loss. Nausea and vomiting occur in the presence of intestinal strictures that produce partial or complete bowel obstructions. Transmural disease commonly manifests in the perianal region as skin tags or perirectal abscesses or fistulas,[44] but it also can present as an inflammatory mass in the right lower quadrant. In children and adolescents, the presentation often is more insidious, with weight loss, failure to grow or to develop secondary sex characteristics, arthritis, or fevers of undetermined origin. Skin lesions, primarily erythema nodosum, may precede intestinal symptoms.[45,46]

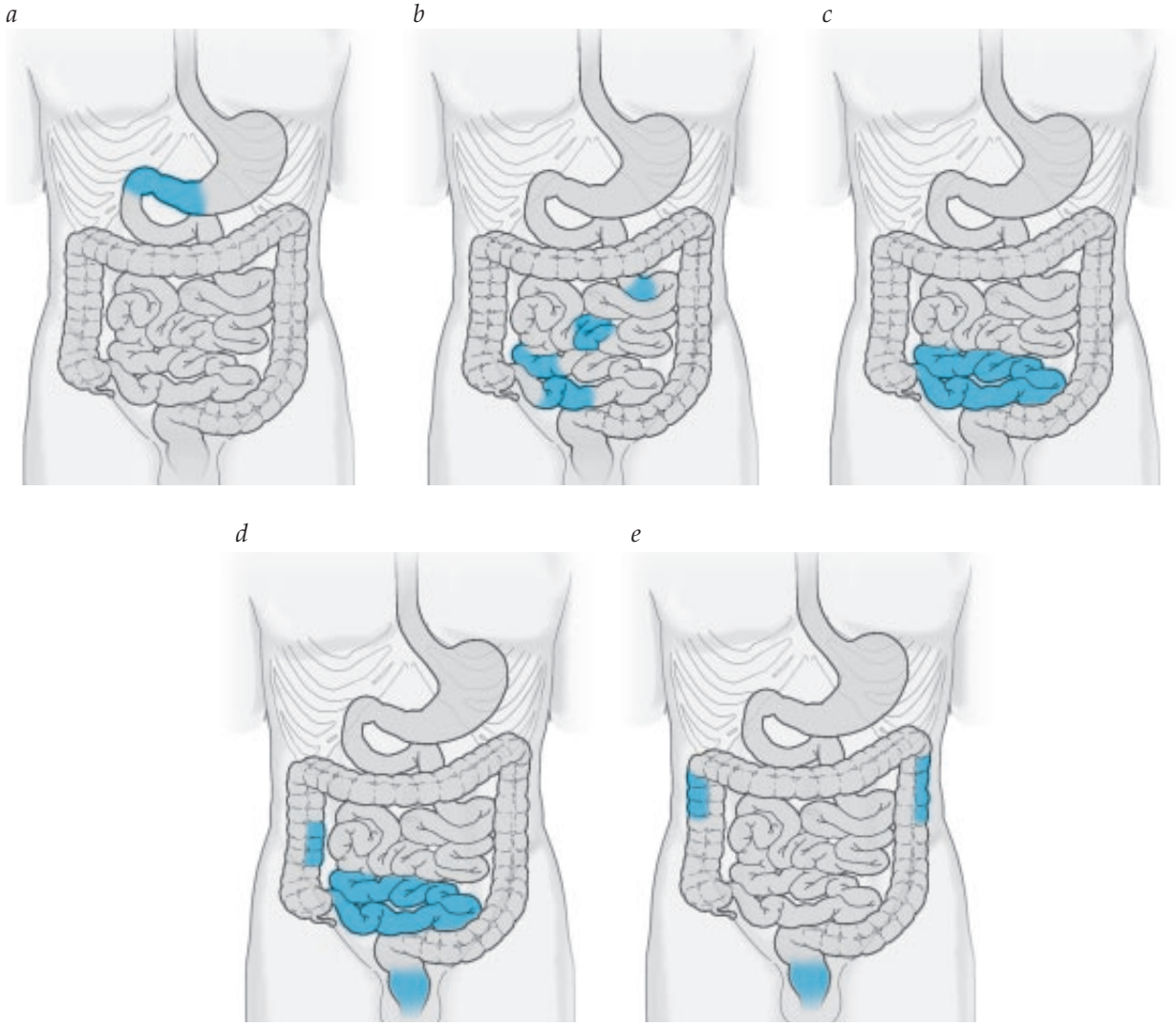


Figure 4 **The spectrum of Crohn disease presentations includes (*a*) gastroduodenitis (7% of patients), (*b*, *c*) jejunoileitis and ileitis (33% of patients), (*d*) ileocolitis (45% of patients), and (*e*) colitis (15% of patients).**

Crohn disease of the esophagus, stomach, and duodenum Infrequently, primary manifestations of CD mimic gastroesophageal reflux or peptic ulcer disease.[47,48] Heartburn, dysphagia, nausea, dyspepsia, epigastric pain, and early satiety or postprandial vomiting typically accompany other systemic inflammatory symptoms such as fever, night sweats, and rectal bleeding.

Jejunoileitis Another relatively uncommon presentation of CD, jejunoileitis most often presents with vomiting and diarrhea, cramping abdominal pain, and weight loss.[49] Patients describe borborygmi related to focal, segmental strictures compromising the passage of enteric contents. Diarrhea is multifactorial and can be secondary to malabsorption as a consequence of inflammation, protein-losing enteropathy, or stasis and small bowel bacterial overgrowth proximal to strictures.

Ileitis and ileocecal Crohn disease CD most commonly presents as right lower quadrant abdominal pain and tenderness (often accompanied by an inflammatory mass), diarrhea with or without rectal bleeding, weight loss, fevers, chills, and night sweats. An acute presentation may mimic appendicitis; occasionally, Crohn ileitis will be diagnosed at exploratory laparotomy for presumed appendicitis.

Crohn colitis Approximately 15% of CD cases are limited to the colon. Distinguishing these cases from UC can be difficult, because the clinical manifestations—diarrhea, rectal bleeding, and urgency—overlap with those of UC [*see Table 2*]. However, CD of the colon is more likely than UC to be accompanied by perianal manifestations (skin tags and perirectal abscess or fistulas), and the rectum often is spared, whereas UC always involves the rectum. In approximately 10% to 20% of patients presenting with colitis, the classification may be indeterminate in the setting of diffuse or severe inflammation or of questionable focal inflammation.[50]

Perianal Crohn disease Perianal involvement in CD most often accompanies colonic disease and begins within the anal crypts.[51] Small fistulas from the anorectal junction progress through or around the anal sphincter and present as perirectal abscesses or fistulas. Often, perianal tissue becomes hypertrophied, producing skin tags [*see Figure 5*]; these may be misdiagnosed as hemorrhoids. At times, perianal manifestations are the primary presentation, and in extreme situations, the anal sphincter and perineum can become grossly deformed.

Physical Examination

Key findings on physical examination of patients with CD include both abdominal and general systematic abnormalities. The abdominal examination may be significant for distension and abnormal bowel sounds in the presence of intestinal strictures producing partial intestinal obstruction. Tenderness in the area of involvement and the presence of an inflammatory mass are common. It is important to examine the perianal region and rectum for evidence of abscess, fistula, skin tags, or anal stricture.

Patients with CD often are chronically ill and can present with weight loss and pallor. The eye exam may demonstrate episcleritis or uveitis. Aphthous ulcerations in the mouth are common, and in extreme cases, patients may exhibit evidence of nutritional deficiencies (e.g., cheilosis or tongue atrophy). Examination of the musculoskeletal system may demonstrate swelling or redness of large joints (e.g., knees, ankles, or wrists) or clubbing of the fingers. Skin examination can reveal erythema nodosum or, rarely, pyoderma gangrenosum.

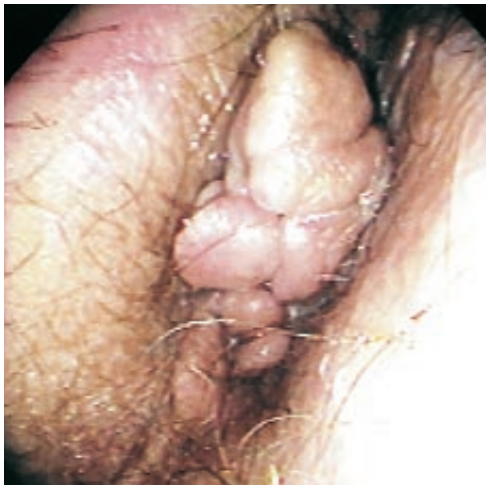


Figure 5 **The typical perianal skin tag of CD differs from the typical hemorrhoid tag.**

Laboratory Studies

Anemia is common in CD. Anemia can result from deficiencies of iron, vitamin B_{12}, or folic acid or may be the anemia of chronic disease. Serum ferritin levels correlate better than iron and iron-binding protein levels with bone marrow iron stores in IBD.[52] Leukocytosis is common, depending on the severity of inflammation and the presence of suppurative complications. Thrombocytosis also is common and is related to inflammation or iron deficiency. Elevated erythrocyte sedimentation rates and C-reactive protein levels reflect nonspecific acute-phase reactions.[53] Electrolyte disturbances depend on the severity of diarrhea and dehydration. Serum albumin levels often are reduced as a result of malnutrition and enteric protein losses. Patients with severe weight loss may have prolonged clotting times because of vitamin K deficiency. Urinalysis commonly demonstrates calcium oxalate crystals.

Quantitative stool examinations are useful in the setting of diarrhea to assess fecal leukocytes (confirming inflammatory diarrhea), stool volume, and fecal fat. Quantification of either fecal calprotectin[54] or lactoferrin[55] is a surrogate for the presence of fecal leukocytes. The presence of the serologic markers anti–*Saccharomyces cerevisiae* antibody and an antibody to the outer core membrane of *E. coli* (OmpC) have high specificity for CD.[56]

Imaging Studies

Radiography Barium contrast studies are the most commonly used diagnostic tools to assess and confirm CD of the small intestine and are useful for assessing the upper digestive tract and colon. In colonic disease, barium studies can define intestinal complications (e.g., stricture formation or fistulas) that cannot be adequately assessed by endoscopy. Features of CD that are shown with barium examinations include mucosal edema, aphthous and linear ulcerations, asymmetrical narrowing or strictures, and separation of adjacent loops of bowel caused by mesenteric thickening. Abnormalities are focal and asymmetrical, with ulcerations most often involving the antimesenteric border. Cobblestoning of the mucosa represents

networks of linear ulcerations outlining islands of residual normal mucosa. Pseudodiverticula formation or dilated loops of bowel are common proximal to strictures. There may be evidence of fistulas extending from any involved segment to an adjacent loop of bowel, the mesentery, or the urinary bladder or from the rectum to the vagina or perineum. CT scanning after direct injection of barium into the small bowel through a nasogastric tube (enteroclysis) provides excellent discrimination between intestinal and extra-intestinal disease.[57]

Other imaging studies Ultrasound examinations or CT scans are useful to assess for abscess in patients who have an inflammatory abdominal mass or who have fever, leukocytosis, or abdominal tenderness. Ultrasound or CT scan is also warranted to assess for hydronephrosis in the setting of an inflammatory mass in the right lower quadrant, because these have the potential to obstruct the right ureter. Transrectal ultrasound, CT scan, and MRI also are useful to assess the extent of perianal and sphincter involvement in patients presenting with perianal or perirectal pain.[51] Scintigraphy using leukocytes labeled with indium or technetium can be helpful to define locations of intestinal inflammation when barium studies are not possible or the results are indeterminate.[35,58]

Endoscopy Colonoscopic examinations have become a primary means of diagnosing CD that involves the colon. Endoscopy in these patients typically reveals sparing of the rectum, with focal inflammatory changes in the more proximal colon and terminal ileum. Other typical features include the presence of aphthous, linear, or irregularly shaped ulcerations with normal intervening mucosa [*see Figure 6*]. Inflammatory strictures may preclude examination of proximal segments of bowel. Inflammatory pseudopolyps may be seen, as in UC. In some patients, polypoid or masslike inflammatory changes may be difficult to differentiate from neoplastic masses; biopsy and histologic analysis may be required. Similar endoscopic

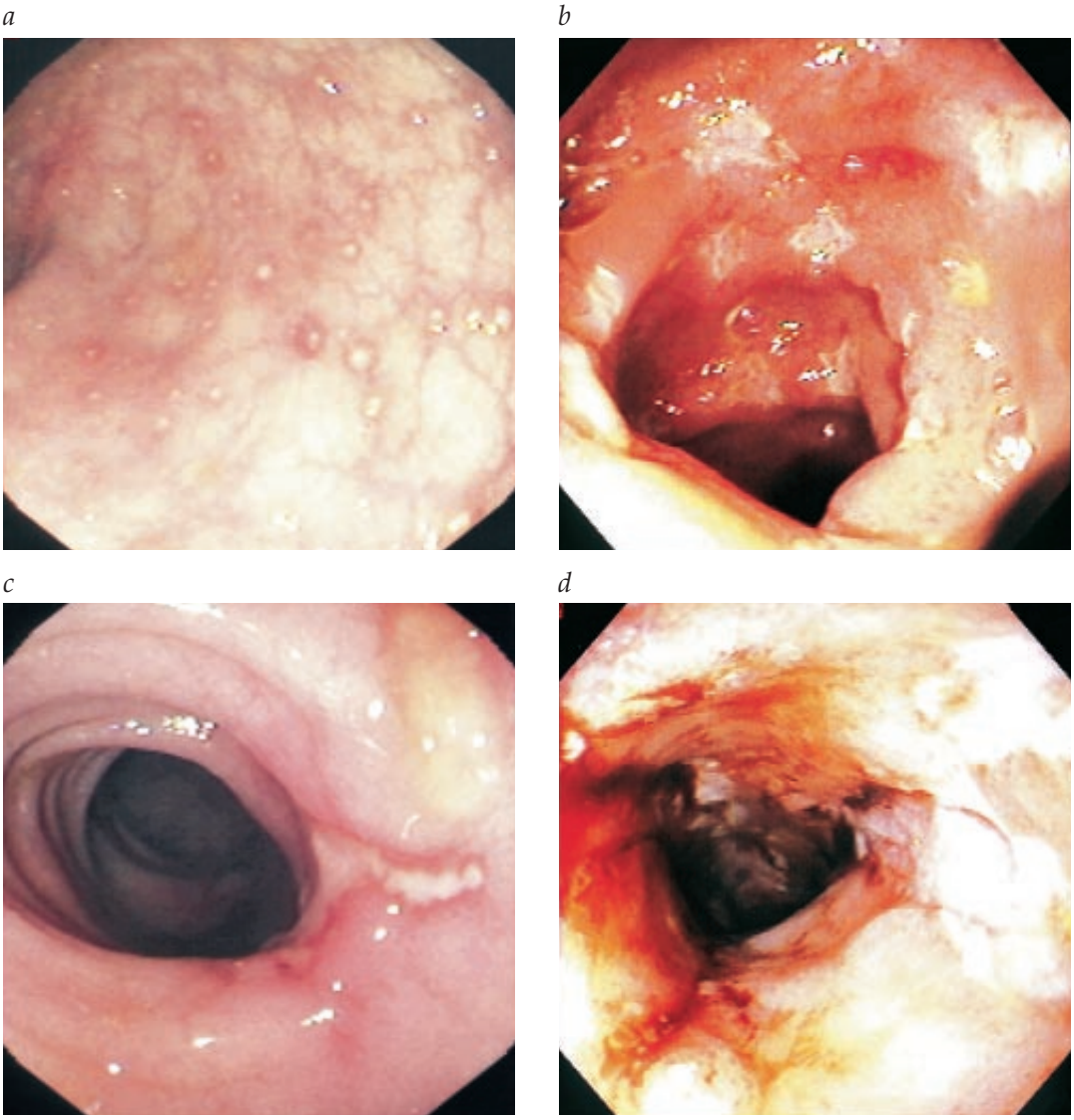


Figure 6 **Endoscopic spectrum of Crohn disease includes (*a*) aphthous ulcerations amid normal colonic mucosal vasculature; (*b*) deeper, punched-out ulcers in ileal mucosa; (*c*) a single colonic linear ulcer; and (*d*) deep colonic ulcerations forming a stricture.**

features may be present in the esophagus, stomach, or duodenum. An important aspect of endoscopic examinations is the ability to obtain samples for pathologic interpretation.

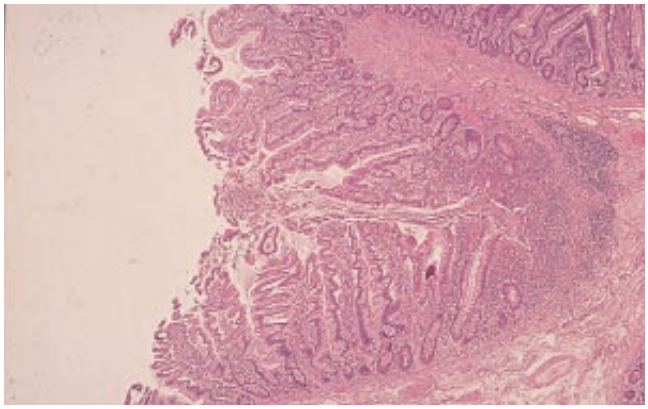
Wireless capsule endoscopy is now being used to diagnose CD. This technique offers access to parts of the small bowel that cannot be reached by standard endoscopy and may be more sensitive than conventional radiographic studies for identifying subtle lesions.[59]

Histology

Pathologic findings in CD reflect the gross pattern of focal and asymmetrical intestinal involvement.[32,33,50] The primary histologic lesion is an aphthous ulcer [*see Figure 7*]. These begin as erosions overlying lymphoid aggregates. As the minute ulceration extends, in either a linear or a transmural pattern, the microscopic and macroscopic changes that develop include a mixed acute and chronic inflammatory cell infiltrate composed of lymphocytes, plasma cells, and neutrophils. Crypt abscesses are common, and the inflammatory infiltrates often are located adjacent to normal epithelium. Noncaseating granulomas, which may be identified in mucosal biopsies or in resected specimens, are characteristic of CD; however, they are not necessary for confirming the diagnosis.

Granulomas may be found in mucosal specimens that appear grossly normal. Specimens from resected intestine demonstrate transmural inflammatory changes extending from the mucosa into the serosa (which is hyperemic, with creeping mesenteric fat). At times, there may be paradoxical involvement of the deeper layers of the bowel wall, with lymphoid aggregates overlying normal-appearing epithelium. Submucosal fibrosis, deep fissuring ulcerations, and fistulizing ulcerations communicate between loops of bowel or into the adjacent mesentery.

a



b

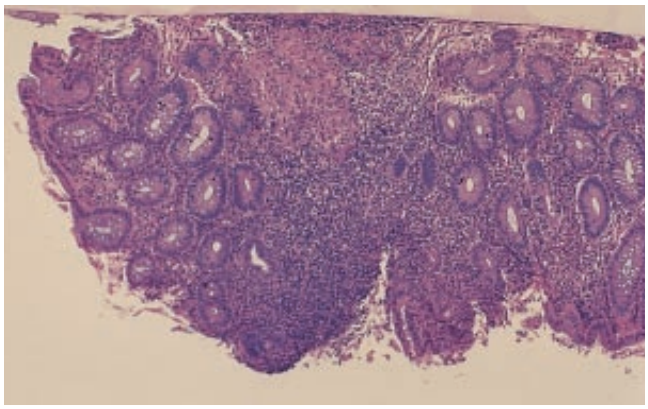


Figure 7 **Pathologic changes in Crohn disease include (*a*) ileal aphthous ulceration overlying a lymphoid aggregate and (*b*) focal colonic ulcer with noncaseating granuloma in lymphoid tissue.**

Differential Diagnosis

IBD should be considered in any patient who presents with rectal bleeding or diarrhea. Identification of fecal leukocytes is the simplest means of discerning an inflammatory process of the intestine. Other causes of rectal bleeding are either traumatic or neoplastic. Diarrhea is nonspecific and has a large differential schema [*see 4:III Diarrheal Diseases*]. The primary chronic diarrheal illness that requires differentiation from UC or CD is irritable bowel syndrome (IBS). IBS is never associated with rectal bleeding, and nocturnal symptoms are uncommon. The presence of occult blood or fecal leukocytes excludes IBS.

Patients with gross or occult blood in the stool require endoscopic evaluation. Colonic neoplasia is a prominent consideration for patients older than 50 years, whereas hemorrhoids or anal fissures are common in younger patients. NSAID-induced colitis is common and may contribute to ischemic colitis in persons in older age groups.[60] Ischemic colitis presents acutely in elderly patients after precipitating events such as dehydration or heart failure; in younger patients, the condition is associated with oral contraceptive use, vasculitis, and hypercoagulable states.[61] Endoscopic examination demonstrates focal hemorrhagic or ulcerated mucosa in the so-called watershed segments of the sigmoid colon or splenic flexure. Diverticular hemorrhage is typically profuse and painless. Some patients, however—particularly elderly persons taking NSAIDs—may present with less vigorous rectal bleeding from diverticulosis involving a segment of the sigmoid colon.[62]

Inflammatory diarrhea can be infectious or noninfectious.[63] The infectious colitides are caused by bacteria such as *Salmonella, Shigella, Campylobacter*, and hemorrhagic *E. coli*. Most of these diseases are acute and self-limited and need be considered only in patients who present with sudden onset of bloody diarrhea and fever. UC and CD develop insidiously, over weeks. *C. difficile* colitis can also mimic ulcerative colitis and may be more chronic, lasting weeks [*see 7:V Anaerobic Infections*]. In immunocompetent hosts, viral or parasitic infections rarely mimic UC. The exception is amebiasis, which may cause acute or subchronic symptoms. Amebiasis can often be distinguished from IBD by wet-mount examination of the stool for motile ameba and the more typical focal (so-called collar button) ulcerations in the colon [*see 7:XXXIV Protozoan Infections*]. Although most cases of infectious diarrhea are acute and self-limited, intercurrent infections and traveler's diarrhea can initiate flare-ups of IBD.[64] Consequently, patients presenting with new-onset IBD or acute exacerbations of IBD should be evaluated for a complicating enteric infection.

Management

The treatment of UC or CD is based on the location, extent, and severity of disease, as well as the patient's response to past therapy.[26,27] Factors that contribute to exacerbations of activity or refractoriness to therapy should be addressed. Such factors may include concomitant medications (e.g., NSAIDs or antibi-

otics), intercurrent infections (e.g., with *C. difficile*), menstruation, and dietary or lifestyle changes.

Treatment follows a sequential approach: induction of remission and, then, maintenance of remission. Clinicians can now choose from a variety of medication classes for treatment of IBD, and both medication selection and dosage may vary according to whether the therapeutic intent is induction or maintenance. Surgical treatment is indicated in selected patients to treat severe disease activity or specific complications.

ANTI-INFLAMMATORY AGENTS

Aminosalicylates

Aminosalicylates are the primary therapies for mild or moderate UC and CD.[26,27] These agents have a long history of clinical use and have been extensively studied in clinical trials for both UC and CD.[65,66] Sulfasalazine, the prototype aminosalicylate [*see Figure 8*], was developed with the intention of providing both an antibacterial agent (sulfapyridine) and an anti-inflammatory agent (5-aminosalicylic acid [5-ASA], mesalamine, or mesalazine) into the connective tissues. It was subsequently recognized that sulfasalazine remains intact through the stomach and small intestine, with minimal enteric absorption. On reaching the colon, the azo bond between sulfapyridine and 5-ASA is cleaved by colonic bacteria. Released sulfapyridine is almost completely absorbed from the colon and undergoes hepatic acetylation and subsequent renal excretion. In contrast, the 5-ASA released into the colon is poorly absorbed and is primarily eliminated in the feces.[67] Therefore, sulfasalazine primarily serves as a carrier for 5-ASA to the colon. The 5-ASA moiety accounts for the primary therapeutic benefits, whereas sulfapyridine causes the majority of side effects attributed to sulfasalazine. These attributes have led to the development of a series of sulfa-free aminosalicylates (e.g., olsalazine, balsalazide, and formulations of mesalamine) that can be targeted to specific sites along the gastrointestinal tract. A basic premise regarding the aminosalicylates is that the effects of 5-ASA are topical (mucosal), rather than systemic, and that the active moiety needs to be delivered to the site of intestinal inflammation.

Sulfasalazine and mesalamine have multiple anti-inflammatory effects, including inhibition of the arachidonic acid cascade along the cyclooxygenase, lipoxygenase, and platelet-activating factor pathways.[65,66] In addition, the aminosalicylates inhibit oxygen radical production and scavenge free radicals.[68] They inhibit lymphocyte and monocyte function and production of immunoglobulin by plasma cells. Sulfasalazine also has been shown to inhibit production of IL-1 and nuclear factor κB.[69]

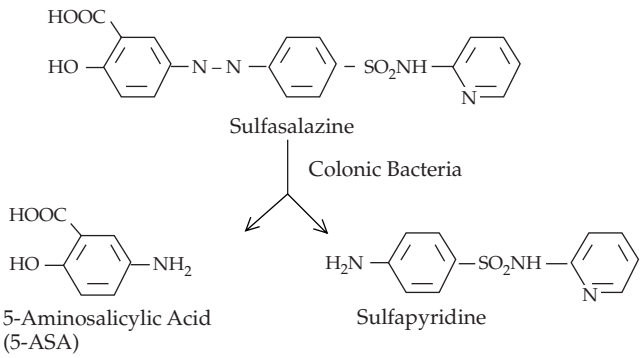


Figure 8 **Sulfasalazine is composed of sulfapyridine and 5-aminosalicylic acid (mesalamine), linked by an azo bond.**

Adverse effects of sulfasalazine are common and are primarily related to plasma sulfapyridine concentrations; these concentrations depend on the rate of hepatic acetylation of sulfapyridine, which is genetically determined.[70] Intolerance side effects (i.e., nausea, vomiting, malaise, anorexia, dyspepsia, and headaches) are dose related. In contrast, hypersensitivity reactions (i.e., rash, fever, hemolytic anemia, agranulocytosis, hepatitis, hypersensitivity pneumonitis, pancreatitis, and worsening of colitis) are independent of dose. Reversible sperm abnormalities and folate malabsorption are unique complications related to sulfasalazine.

Mesalamine has relatively few side effects, and dose-related toxicities are unusual in patients taking up to 4.8 g/day of delayed-release formulations.[70,71] Rare idiosyncratic reactions, including pancreatitis, interstitial nephritis, and worsening of colitis, have been reported. Eighty percent of patients who are unable to tolerate sulfasalazine can tolerate a nonsulfa aminosalicylate containing mesalamine. One unique complication of olsalazine is dose-related diarrhea.

Sulfasalazine and the oral aminosalicylates are equally effective for treatment of mild to moderate UC.[26,27,72] Oral aminosalicylates are effective for both proximal and distal colitis. Mesalamine suppositories and enemas effectively treat distal colitis, provided the formulation reaches the proximal extent of disease.[73] A dose-response relationship for the oral aminosalicylates is well defined for up to 4.8 g/day of mesalamine, with higher doses being more effective for moderately active disease.

Prevention of relapse and the prolongation of remission have been primary indications for all of the aminosalicylates in UC.[26,27,74] All the nonsulfa formulations provide comparable efficacy to sulfasalazine for maintenance therapy.[74] In ulcerative proctitis and distal colitis, topical mesalamine is also effective at preventing relapse and is more effective than oral treatment when continued on a long-term basis.[73]

The efficacy of aminosalicylates in CD is less definitive than that in UC and is more dependent on location of disease activity.[27,66] In CD involving the colon, the efficacy of sulfasalazine is determined by the presence of the colonic bacteria needed to cleave the azo bond and liberate 5-ASA. Aminosalicylates are commonly used as maintenance therapy for patients with quiescent CD, despite differing interpretations of the clinical trials of this indication.[27,65] The specific release characteristics [*see Table 3*] should correspond to the disease location.

Corticosteroids

Corticosteroids are the primary therapy for moderate to severe and fulminant UC and moderate to severe active CD.[27,75] They are ineffective at maintaining remissions of UC and CD, however.[76] The mechanisms of action of corticosteroids in IBD are multifactorial and are similar to their mechanisms of action in other inflammatory diseases.[77] Like aminosalicylates, corticosteroids can be targeted to specific sites within the digestive tract. Newer glucocorticoids (e.g., budesonide) both enhance potency and minimize systemic exposure.[78]

Ulcerative colitis Oral corticosteroids are the primary treatment for outpatients with moderately severe UC.[26,27,79] Prednisone, 20 to 60 mg daily, is administered once or in divided doses. In general, 40 mg is the optimal dose; the modest benefits of higher doses are offset by increasing side effects.[80]

Table 3 Aminosalicylate Preparations for Management of Ulcerative Colitis and Crohn Disease

Preparation	*Formulation*	*Delivery*	*Dosing*
Oral Agents			
Azo bond			
Sulfasalazine (Azulfidine, 500 mg)	Sulfapyridine carrier	Colon	3–6 g/day (acute) 2–4 g/day (maintenance)
Olsalazine (Dipentum, 250 mg)	5-Aminosalicylic acid dimer	Colon	1–3 g/day
Balsalazide (Colazal, 750 mg)	Aminobenzoyl-alanine carrier	Colon	6.75 g/day
Delayed Release			
Mesalamine (Asacol, 400 mg, 800 mg)	Eudragit S (pH, 7)	Distal ileum–colon	2.4–4.8 g/day (acute) 0.8–4.8 g/day (maintenance)
(Claversal,* Mesasal,* Salofalk,* 250 mg, 500 mg)	Eudragit L (pH, 6)	Ileum-colon	1.5–3.0 g/day (acute) 0.75–3.0 g/day (maintenance)
Sustained Release			
Mesalamine (Pentasa, 250 mg, 500 mg, 1,000 mg*)	Ethylcellulose granules	Stomach-colon	2–4 g/day (acute) 1.5–4.0 g/day (maintenance)
Rectal Agents			
Mesalamine suppository (Canasa, 400 mg,* 500 mg, 1,000 mg*)	—	Rectum	1.0–1.5 g/day (acute) 500–1,000 mg/day (maintenance)
Mesalamine enema (Rowasa, 1 g,* 4 g)	60 ml, 100 ml suspension	Rectum–splenic flexure	1–4 g/day (acute) 1 g q.d.–1 g t.i.w. (maintenance)

*Not available in the United States.

Rectal (i.e., topical) administration of systemically absorbed glucocorticoids (e.g., hydrocortisone) and of rapidly metabolized glucocorticoids (e.g., budesonide) is effective therapy for active distal colitis and has been incorporated into the treatment of severe colitis as an adjunct to parenteral steroids.[81]

Parenteral corticosteroids are the mainstay of therapy for hospitalized patients with severe or fulminant UC. Although controlled trial data are limited, there is consensus supporting the use of intravenous hydrocortisone, methylprednisolone, or prednisolone in dosages equivalent to 40 to 60 mg/day of prednisone; these agents may be given either in a continuous infusion or in divided doses.[26,27,79,80] There is no evidence supporting dosages higher than 60 mg of prednisone. Corticosteroids are not effective in preventing relapse of quiescent UC.[26,27,79,80]

Crohn disease Corticosteroids are the primary therapy for moderate to severe CD.[27,82] Both uncontrolled and controlled trials demonstrate a response rate of approximately 75%. Parenteral steroids have not been formally assessed in the setting of severe CD, but there is a consensus that they are as effective in severe CD as in severe UC. As a first-line therapy for mild to moderate CD, enteric-coated, delayed-release formulations of budesonide can deliver topically active steroids to targeted sites (ileum and proximal colon).[78,83] However, as with corticosteroids in UC, neither conventional nor topically active steroids have proved effective for preventing relapse in CD.[76,82]

IMMUNOMODULATORY AGENTS

Immunomodulating therapies have had an expanding role in the treatment of IBD. These agents can be used either to induce or to maintain remission in UC and CD.

Azathioprine and 6-Mercaptopurine

Although azathioprine (AZA) and 6-mercaptopurine (6-MP) have been used to treat IBD for over 30 years, their mechanisms of action and optimal use remain incompletely known. AZA is rapidly absorbed and converted to 6-MP, which is metabolized to thioinosinic acid, an inhibitor of purine ribonucleotide synthesis and cell proliferation.[84] Although AZA is considered a purine antimetabolite, its exact mechanism of action has not been defined. Presumably, AZA inhibits some function of long-lived lymphocytes, which accounts for the 3- to 6-month delay in onset of action. There is increasing evidence that these agents promote apoptosis of T cells.[85] In addition, a genetic polymorphism has been recognized in the enzyme thiopurine methyltransferase, which metabolizes the purine analogues into 6-thioguanine.[84,86] One in 300 persons lacks this enzyme, and an additional 11% of the population has depressed levels of it. Homozygotes are susceptible to increased accumulation of 6-thioguanine nucleotides and bone marrow suppression.

AZA and 6-MP usually are well tolerated.[80] Pancreatitis occurs in 3% to 15% of patients, usually within the first few weeks of therapy, and resolves completely upon withdrawal of the drug. Other potential side effects include nausea, fever, rash, and hepatitis. Bone marrow suppression, particularly leukopenia, is dose related and may be delayed, necessitating long-term monitoring of blood counts. IBD patients treated with purine analogues may experience a slight increase in relative risk of neoplasia, but the absolute risk remains extremely small.[80,87] There is a growing consensus that these agents are effective and safe for use through pregnancy and lactation.[88]

Ulcerative colitis Both controlled trials with AZA and uncontrolled series with 6-MP have supported the role of purine analogues for the long-term (maintenance) treatment of UC.[26,27,79] There have not been adequate comparative studies between AZA and 6-MP or dose-ranging trials for these agents, and to date, there are no standard guidelines for their use in UC.[26] Most authorities agree that if patients are to be started on full-dose therapy (i.e., 2.5 mg/kg/day of AZA or 1.5 mg/kg/day of 6-MP), activity of the enzyme thiopurine methyltransferase

should be measured before therapy is initiated.[80,89] However, many clinicians start therapy at lower doses and monitor the white blood cell count. Alternatively, in the event of therapeutic unresponsiveness, elevated liver enzymes, or potential noncompliance, these clinicians measure levels of the thiopurine metabolites 6-methylmercaptopurine and 6-thioguanine.

Crohn disease AZA and 6-MP can induce remission in active CD, but prolonged therapy is necessary: 56% of patients with active CD respond after 4 months of treatment with either AZA (2.0 to 2.5 mg/kg/day) or 6-MP (1.0 to 1.5 mg/kg/day).[82,90] Because of their delayed onset of action, AZA and 6-MP are most often used to maintain remission or as steroid-sparing agents.[91] They can also be used to treat CD fistulas and perianal disease.[51]

Cyclosporine

Cyclosporine is a potent inhibitor of T cells, primarily via inhibition of IL-2 production by helper T cells, and inhibits recruitment of cytotoxic T cells and production of IL-3, IL-4, interferon gamma, and TNF-α. Treatment with cyclosporine can provide dramatic results in severe IBD, particularly in UC.[79,92] Cyclosporine has a much more rapid onset of action than AZA or 6-MP; its effects are usually evident within the first week. There is controversy regarding the long-term benefits of cyclosporine, however. Cyclosporine is metabolized by a cytochrome P-450 enzyme, and interactions with a number of drugs can increase or decrease cyclosporine levels.[80]

UC trials have used intravenous cyclosporine as a continuous infusion of 2 to 4 mg/kg/day.[93] However, a dosage of 2 mg/kg/day appears to be as effective as higher doses and may reduce toxicity.[94] Correlations between response and blood levels have not been defined; similarly, the correlation between blood levels and toxicity is poor.[80]

The narrow therapeutic margin and significant potential toxicity of cyclosporine remain obstacles for its use outside of centers with transplantation expertise. Major toxicities include nephrotoxicity and opportunistic infections. Nephrotoxicity can manifest as hypertension or elevations in blood urea nitrogen and creatinine levels. Because of the increased risk of opportunistic infections, including *Pneumocystis* pneumonia, prophylaxis with trimethoprim-sulfamethoxazole has been recommended for patients receiving cyclosporine in conjunction with high-dose steroids.

The primary use of cyclosporine for IBD is in hospitalized patients with severe UC in whom therapy with oral or intravenous steroids has failed[26,27]; 50% to 80% of such patients respond to short-term treatment with intravenous cyclosporine.[79] Duration of use is limited to 3 to 6 months. The long-term prognosis after cyclosporine therapy is controversial, but approximately 40% to 50% of responders to cyclosporine may avoid eventual surgical colectomy.[93] This response improves to greater than 60% when patients are transitioned to long-term therapy with AZA or 6-MP.[95]

Cyclosporine also has been successfully used for steroid-refractory and fistulizing CD.[96] Beneficial results are primarily achieved with intravenous cyclosporine. Oral cyclosporine has not been found to be effective for maintaining remission in CD, possibly because of poor and variable absorption.

Methotrexate

Only limited evidence supports the use of methotrexate in UC. Despite early optimism from uncontrolled clinical experience, methotrexate has not been effective therapy for UC in a small number of clinical trials.[97]

In studies of steroid-dependent CD patients, approximately 40% of patients were able to achieve clinical remission while tapering steroids during a 16-week trial of methotrexate, given parenterally (intramuscularly or subcutaneously) in a weekly dose of 25 mg.[82,98] In a subsequent study, approximately two thirds of patients who achieved remission and remained on parenteral methotrexate, at a dosage of 15 mg/wk, continued in remission for 40 weeks.[99]

Methotrexate is well tolerated in IBD patients. Toxicity, which includes bone marrow suppression and hepatic fibrosis, is uncommon provided that blood counts and liver enzyme levels are monitored.[97] Hypersensitivity pneumonitis is a rare but potentially irreversible complication. Methotrexate is a known teratogen and abortifacient, precluding its use in women anticipating pregnancy.

Other Immunomodulatory Agents

Tacrolimus has been used as therapy for refractory fistulas in CD[100] and as an oral therapy for refractory UC.[101] The indications for tacrolimus are very limited, however, now that infliximab has become available for CD [*see* Biologic Therapies, *below*].

ANTIBIOTICS

Antibiotic therapy has been used selectively in both UC and CD.[102] Although a specific therapeutic role for antibiotics in UC remains unproved, most centers continue to advocate broad-spectrum antibiotics as a component of the intensive intravenous therapy used in patients with fulminant colitis and toxic megacolon.[26,27] Antibiotics are also effective in the treatment of pouchitis after ileoanal anastomoses in patients with UC.[103]

In CD, the role for antibiotic therapy as a first-line agent for mild to moderate disease continues to be debated.[104-106] In mild to moderate CD, metronidazole has proved to be comparable to sulfasalazine and superior to placebo at doses of 20 mg/kg/day.[107] Metronidazole is also effective for the treatment of perianal CD[51] and can reduce the likelihood of relapse after intestinal resection.[108] Ciprofloxacin is comparable to mesalamine for mild to moderate CD and has been used successfully in combination with metronidazole for ileal disease and perianal CD.[109-111] Combinations of antimycobacterial therapies for CD have not had consistent results in the treatment of active CD or as maintenance therapies.[112]

BIOLOGIC THERAPIES

The introduction of biologic agents has opened a new era in the treatment of IBD. Current biologic agents target cellular messengers, including cytokines, chemokines, and adhesion molecules.

The first biologic agent approved by the Food and Drug Administration for CD was infliximab, a chimeric monoclonal antibody of the IgG1 subclass that targets TNF-α. In clinical trials, infliximab has been shown to induce and maintain clinical remissions in patients with moderate to severe active luminal or fistulizing CD refractory to therapy with aminosalicylates, corticosteroids, and immunomodulators.[113-115] Infliximab is administered intravenously in a dose of 5 mg/kg. After the initial dose, repeat doses are given 2 weeks and 6 weeks later and, then, every 8 weeks on average. Infliximab is similarly effective in UC. Two large trials that enrolled patients with moderately active UC that was refractory to aminosalicylates, cortico-

steroids, and immunomodulators also demonstrated positive results.[116]

Infliximab is generally well tolerated. The primary risk with infliximab therapy is infection with intracellular organisms (e.g., tuberculosis, histoplasmosis, or cryptococcosis) in exposed or endemic populations.[117] Pretreatment skin testing with purified protein derivative and chest x-rays is recommended, but false negative results are possible because many patients with CD are anergic.[118] For that reason, clinicians must exercise clinical judgment regarding potential tuberculosis exposure. Infliximab is contraindicated in patients with active infections.

Unique adverse events with infliximab include the development of antibodies to the drug that reduce the effectiveness of the agent and are associated with infusion reactions (e.g., acute infusion reactions and serum sickness–like reactions).[119] Additionally, antinuclear antibodies and anti-DNA antibodies develop in approximately 10% of patients with CD who are receiving infliximab therapy. Drug-induced lupus reactions have been reported but are uncommon. Delayed hypersensitivity (i.e., serum sickness–like reactions) has been observed in patients retreated after a long hiatus between doses (3 months to 4 years) but not in patients who have received continuous retreatment at 8-week intervals. A small increase in the risk of lymphomas has been observed with anti-TNF therapy in IBD,[120] but it is not clear whether the increase reflects a small underlying risk in patients with CD or is the result of immune suppression.[121] Other biologic therapies under development to treat IBD include inhibitors of so-called selective adhesion molecules[122]; anti–IL-12, anti–IL-2, and anti-CD3 monoclonal antibodies[123,124]; epidermal growth factor[125]; and granulocyte-macrophage colony-stimulating factor.[126]

MISCELLANEOUS THERAPIES

The recognition that cigarette smoking can protect against the development of UC has led to trials utilizing nicotine as adjunctive therapy. Although trials have demonstrated a role for nicotine in the symptomatic management of UC, nicotine therapy has not been shown to be effective at inducing remissions.[127] Currently, nicotine is not a proven therapy for UC, but it may be a useful adjunctive measure in patients with UC that develops after smoking cessation.

Omega-3 fatty acids inhibit synthesis of leukotriene B_4 and, at high doses, have shown a modest benefit in the treatment of active UC or as maintenance therapy.[128,129] An enteric-coated fish oil preparation has proved to be effective in reducing relapse rates in CD[130] and may eventually offer an alternative therapeutic option in IBD. However, current treatment guidelines do not include these agents.

Other novel therapeutic approaches to IBD that are currently under investigation include phosphodiesterase inhibitors,[131] small molecules targeting mitogen-activated protein kinases,[132] probiotics,[133] apheresis,[134] and targeting of costimulatory molecules.[135]

NUTRITIONAL THERAPIES

Hypotheses regarding dietary intraluminal antigens as important stimuli of the mucosal immune response have led to the investigation of nutritional therapies for IBD.[3,136] Dietary manipulations have not been effective in treating UC, but patients with active CD have responded to several nutritional approaches.[136,137] Bowel rest and total parenteral nutrition (TPN) are as effective as corticosteroids at inducing short-term remissions in active CD. In contrast, enteral nutrition in the form of elemental or liquid polymeric preparations has been shown to be less effective than corticosteroids.[138] It has been suggested that elemental diets may provide the small intestine with nutrients vital to cell growth (e.g., glutamine) while avoiding complications related to TPN. However, despite their efficacy in active CD, neither enteral nor parenteral nutrition is effective at maintaining remissions.[136]

Patients with IBD are susceptible to nutritional deficits as a consequence of blood loss, protein-losing enteropathy, small bowel bacterial overgrowth, surgical resections (in CD), or inanition. IBD does not increase the risk of lactose intolerance, but ingestion of lactose can contribute to diarrhea in IBD patients with an inflamed small bowel or impaired colonic absorptive capacity. Similarly, consumption of nonabsorbable carbohydrates (e.g., sorbitol) or fats (e.g., olestra) can lead to excess flatus, bloating, or diarrhea. Occasionally, patients with proctitis present with constipation that improves with additional dietary fiber.

SUPPORTIVE THERAPIES

Many symptoms of IBD are not related to active inflammation and can therefore be treated separately from the inflammation. The management of these symptoms, which include pain and diarrhea, is as important to the patient's well-being as the treatment of mucosal inflammation. Treatment should be individualized according to symptoms and clinical disease state.

IBS is as common in IBD patients as it is in the general population.[139] A dietary history is important to identify potentially aggravating components contributing to digestive symptoms. Although the stress of day-to-day living does not impact on the inflammatory activity of IBD, many patients identify stressful aspects in life as being associated with worsening of symptoms. IBS often responds to antispasmodics, antidiarrheals, fiber supplementation, or low doses of tricyclic antidepressants [*see 4:XIV Gastrointestinal Motility Disorders*]. Antispasmodics, primarily anticholinergic agents (e.g., dicyclomine, clidinium bromide, hyoscyamine, propantheline, or belladonna alkaloids), can treat cramping abdominal discomfort or symptoms of IBS that accompany UC and CD. Similarly, antidiarrheal preparations (e.g., diphenoxylate, loperamide, or codeine) can be utilized in patients with mild or moderate IBD to reduce the frequency of bowel movements and rectal urgency. Antimotility agents should be avoided in patients with severe or fulminant IBD because of the risk of inducing toxic megacolon.

In women with IBD, flares of disease activity are often related to the menstrual cycle, occurring more often during the premenstrual and menstrual phases.[140] If menstrual cycles impact greatly on symptoms, ablation of the menstrual cycle with progesterone or leuprolide may be warranted. Pregnancy is associated with both exacerbations and remissions of IBD.[141]

There is no predisposing psychiatric personality profile in IBD, and there is no routine role for sedative, anxiolytic, antidepressant, or antipsychotic therapy. Psychopharmacologic therapies are reserved for individual patients as needed, usually after consultation with a psychiatrist.

Treatment of IBD patients with narcotic analgesia is rarely indicated. Pain in UC is related either to visceral hyperalgesia and muscle spasm or to transmural inflammation. The former condition is treated with antispasmodics and the latter with specific anti-inflammatory therapy. In CD, abdominal pain may be related to transmural inflammation or stenosis, but given the chronic nature of the disease, addictive analgesics should be avoided because of the risk of tolerance. Attempts should be made to re-

duce the inflammatory component of symptoms and to treat irritability with antispasmodics or nonaddictive analgesics.

NSAIDs (both nonspecific agents and cyclooxygenase-2 inhibitors) may exacerbate disease activity in IBD and can contribute to refractory disease.[139,142-144] Minor pain, fever, menstrual symptoms, or arthralgias should be treated with alternatives to NSAIDs. If these agents are used, it should be with great caution and continued observation for their potential to exacerbate IBD.

Approximately half of patients with IBD use complementary therapies.[145] Consequently, the history should include a careful review of nonprescription vitamins, health foods, homeopathic agents, or herbs, which may identify factors contributing to changes in bowel habits.

Patients with CD who have undergone bowel resection often have increased diarrhea, related to the length of bowel removed. Bile salt malabsorption may complicate resections of less than 100 cm; the diarrhea in these patients often responds to cholestyramine or alternative bile-salt sequestrants. Longer resections result in steatorrhea, which is managed with a low-fat diet.

MEDICAL TREATMENT OF ULCERATIVE COLITIS

Ulcerative Proctitis

Induction therapies Topical aminosalicylates are the most effective treatments for distal UC.[73] A daily dose of 1,000 to 4,000 mg is administered nightly as an enema or in divided doses as a suppository or foam. Topical corticosteroids (given via suppository, enema, or foam) are acceptable alternatives to mesalamine.[26,27] Foam preparations are easier to retain and are better tolerated, allowing maintenance of daily activities despite twice-daily administration.

Oral aminosalicylates can be used to treat mild to moderate symptoms of proctitis but are less effective than topical therapies.[146] Sulfasalazine, 2 to 6 g/day in divided doses, is the most cost-effective aminosalicylate, but sulfa intolerance, toxic reactions, or allergy can compromise therapy. Mesalamine, olsalazine, or balsalazide formulations are preferable for patients with a history of sulfa allergy or for patients who develop sulfa-related side effects.

Inductive therapy is continued until the patient is asymptomatic. Although improvement should begin within a week, a complete response may require 4 to 12 weeks. Clinicians should recognize that patients with treated UC are capable of, as well as expected to, achieve a clinical remission, which is defined by the resolution of all inflammatory symptoms and a regeneration of the colonic mucosa.

Maintenance therapies Maintenance therapy is indicated for the majority of patients with UC. In patients with proctitis, however, the limited nature of the disease permits treatment of any recurrent attacks on an as-needed basis.

Once remission has been achieved, the daily dose of mesalamine can be tapered according to the initial response; nevertheless, continuation of the inductive therapy, excluding steroids, is most effective for maintenance treatment. Mesalamine suppositories (or enemas) administered nightly, with gradual tapering to every other night and then every third night, will maintain remission in most patients. An oral aminosalicylate is added if patients continue to experience flares despite attempts to wean them from topical therapy; it may also be added after induction and then tapering of topical steroids.[133]

Left-Sided Colitis

Induction therapies Mesalamine enemas are the most effective therapy for left-sided colitis, with steroid enemas being an alternative.[73] Oral aminosalicylates also are effective, with improvement generally noted by 2 to 4 weeks.[147] The oral amnosalicylates are generally equivalent in their efficacy for distal colitis.[67,148]

Patients with moderate to severe disease and those in whom therapy with topical and oral aminosalicylates has failed are treated as outpatients with oral steroids to induce remission, in a manner similar to that for patients with extensive colitis (see below). As with extensive colitis (see below), severe left-sided colitis requires hospitalization and treatment with systemic steroids.

Maintenance therapies Inductive therapy is continued until the patient achieves clinical remission (normal bowel movements without bleeding, urgency, tenesmus, or inability to evacuate flatus). The transition to a maintenance regimen is then begun. Neither oral steroids nor topical steroids are effective at maintaining remissions. Patients who have responded to rectal mesalamine can continue with this therapy or switch to oral treatment.[26,148] The combination of oral and topical mesalamine has advantages over either therapy alone for maintenance therapy for left-sided UC.[149]

If patients who are taking an oral aminosalicylate experience relapse, the dose should be increased to up to 4.8 g of mesalamine. Maintenance of remission in such patients may require topical mesalamine. After inductive treatment with steroid enemas, patients should be transitioned to an oral aminosalicylate, with gradual tapering of the topical therapy. Those requiring systemic steroids should be maintained on an oral aminosalicylate, with or without topical mesalamine.

Extensive Colitis

Induction therapies Oral aminosalicylates are the primary therapy for outpatients with mild to moderate extensive colitis, but these agents may be supplemented with topical mesalamine or steroids.[26,27,108] The dose of the oral aminosalicylate is more important than the specific formulation.[67,72] Response rates of up to 80% can be anticipated with 4 to 6 g of sulfasalazine or 2 to 4.8 g of a mesalamine formulation given over 6 to 8 weeks. Therapy is continued as long as the patient is improving, to the point of clinical remission (i.e., normal bowel movements without blood or urgency). In the absence of a complete response, the dose of the aminosalicylate should be increased to a maximum of 4.8 g of mesalamine. An antispasmodic or antidiarrheal preparation may be added to treat abdominal cramping or mild diarrhea.

In patients who fail to improve or whose condition worsens, steroid therapy should be added in the form of prednisone, 40 to 60 mg/day.[75] Once the patient has achieved a clinical remission, which generally takes 2 to 4 weeks, steroids are tapered according to the time course to improvement. Prednisone can be tapered by approximately 5 mg every week down to 20 mg daily. Below 20 mg of prednisone, the daily dosage is reduced 2.5 to 5 mg every 1 to 2 weeks. Aminosalicylate therapy is continued as steroids are reduced.

Patients who respond to steroids but who are unable to completely taper without relapse despite optimal doses of an aminosalicylate should be started on AZA or 6-MP, which typically permits steroid withdrawal.[26,27] Achieving the therapeutic benefits of these agents requires 3 to 6 months, during which

time steroids are maintained at the lowest dose needed to prevent recurrence of symptoms. Calcium and vitamin D supplementation is indicated during steroid therapy to prevent metabolic bone disease. Reduced bone density is an indication for additional therapy with a bisphosphonate, estrogen replacement in postmenopausal women, or calcitonin.[150]

Hospitalization is indicated for patients who have significant weight loss, fever, disabling extraintestinal manifestations, frequent nocturnal bowel movements, severe anemia, or progressive symptoms despite outpatient therapy with corticosteroids.[26,27] A low-residue diet is prescribed, to minimize abdominal cramps and bowel movements; the diet should contain sufficient protein and calories to counter the catabolic influence of active inflammation and steroids. Antispasmodics or antidiarrheals should be used with caution, and patients should be monitored for worsening symptoms.

Intravenous steroids are indicated for severely ill patients with fever, orthostasis, evidence of dehydration, more than 10 to 12 stools daily, rectal bleeding necessitating transfusion, protein depletion, or abdominal tenderness or distention. Prompt correction of fluid and electrolyte imbalances is critical. In patients with active bleeding, transfusions of packed red blood cells should be given to maintain the hematocrit above 30%. Anticholinergics, antidiarrheals, and narcotic analgesics are contraindicated because they can worsen colonic dilatation and mask peritoneal signs in debilitated, steroid-treated patients.

The intensive intravenous steroid regimen consists of prednisolone (40 to 60 mg/day), methylprednisolone (32 to 48 mg/day), or hydrocortisone (300 to 400 mg/day) administered in divided doses or as a continuous infusion. Steroid enemas (e.g., with 100 mg of hydrocortisone) can be used as adjunctive treatment to reduce rectal urgency or tenesmus. Oral aminosalicylates are discontinued because their anti-inflammatory effects are minor compared with those of high-dose steroids, as well as because of the potential for intolerance and the rare instances in which 5-ASA can worsen colitis.

When vital signs normalize, the hematocrit stabilizes, and the patient is able to tolerate a full (low-residue) diet with formed bowel movements without blood or urgency, treatment can be transitioned to an oral regimen. Full-dose therapy with an aminosalicylate is resumed and intravenous steroids are replaced with oral steroids.

If the patient is not improving after 5 to 7 days of intensive intravenous steroid therapy, the likelihood of improvement is small and the patient should be considered a candidate for cyclosporine therapy or surgery. Intravenous cyclosporine has been an important advance in the therapy of severe UC, but its use should be limited to clinicians experienced in the monitoring of immune suppression. A response is anticipated within 4 to 5 days, but if there is no significant improvement within 1 week, the patient should be referred for surgery. When clinical remission is achieved with intravenous cyclosporine, the regimen is replaced with both oral cyclosporine and prednisone. The daily dose of cyclosporine is doubled and administered in two divided doses (e.g., if the patient was receiving 200 mg daily, the oral dosing is 200 mg twice daily). Because of the high relapse rate after intravenous cyclosporine therapy, AZA or 6-MP is usually added to the oral regimen.[26,151] In addition, trimethoprim-sulfamethoxazole is given three times weekly as prophylaxis against *Pneumocystis* pneumonia.[26]

Outpatient monitoring of cyclosporine levels and other laboratory measures are repeated weekly for the first month and then less often. Steroids are tapered (see above), generally over 8 to 12 weeks; cyclosporine is gradually discontinued; and maintenance therapy is begun.

Maintenance therapy Maintenance therapy for extensive UC is determined by the intensity of therapy needed to induce remission. If aminosalicylate therapy has been sufficient to induce remission, continuation of the same dosage is optimal for maintenance. Patients treated with steroids require a more individualized approach, with the rate of steroid tapering determined by the rapidity of response as maximum doses of aminosalicylate are continued. Patients receiving AZA or 6-MP also are continued on maximum aminosalicylate therapy. The optimum doses of immunomodulators, as well as the doses that will cause leukopenia, have not been clarified.[26,89] Complete blood counts should be obtained on at least a quarterly basis to detect delayed bone marrow suppression.[152]

Fulminant Colitis and Toxic Megacolon

Fulminant colitis, with or without colonic dilatation (toxic megacolon), is a medical emergency that is best managed by an experienced team of gastroenterology specialists and surgeons. Management is similar to that for severe colitis but with several modifications. Patients take nothing by mouth until they show clinical improvement. In the presence of small bowel ileus, a nasogastric tube should be inserted and maneuvers undertaken to reduce colonic distention and to allow passage of colonic gas by rectum (i.e., by rolling the patient from side to side, inserting a rectal tube, or placing the patient in the knee-elbow position).[26] Intravenous steroids are continued and broad-spectrum antibiotic coverage is added for presumed transmural extension of disease, risk of microperforation, and systemic bacteremia. Cyclosporine in this setting is controversial but has been used in selected cases.

Aggressive medical management is successful in 40% to 50% of patients with fulminant colitis or toxic megacolon. Unfortunately, many patients are destined to develop complications or resistant disease, including recurrent toxic megacolon.[25] Persisting peritoneal signs, any deterioration, or failure to improve within 24 to 72 hours is an indication for immediate colectomy.

SURGICAL TREATMENT OF ULCERATIVE COLITIS

UC is cured by proctocolectomy, and the quality of life after such surgery is generally excellent. The advantages of surgery include the elimination of the drawbacks of medical treatment, which include continued morbidity, adverse reactions to therapy, and the risk of neoplasia.[26]

Indications for Surgery

Indications for colectomy in UC are emergent, urgent, or elective. Emergent indications include exsanguinating hemorrhage, perforation, and unresponsive fulminant colitis or toxic megacolon. Urgent indications are chronic refractory colitis and significant complications of the disease or medical therapy (e.g., hemolytic anemia, pyoderma gangrenosum, and steroid-induced psychosis).

Surgical indications are often less acute and allow preparation and education of the patient and family to optimize timing and minimize physical and emotional consequences. Patients with quiescent colitis but with dysplasia diagnosed during colonoscopic surveillance are often feeling well and must adjust to the need for a major operation.

The most common indications for surgery are medically in-

tractable disease, poor quality of life, or chronic complications from colitis or medical therapy. Given the availability of a surgical cure, it is not acceptable for patients to suffer physical debility, psychosocial dysfunction, or intolerable side effects.

Surgical Procedures

Proctocolectomy Removal of the colon with an end ileostomy cures UC. This is the standard procedure with which all other treatments must be compared. Proctocolectomy and ileostomy are usually performed in a single procedure, even in the most urgent of settings. This approach has the least likelihood of complications. The primary drawback is the need for a permanent stoma. Quality of life after proctocolectomy is usually good, although many patients have difficulty adjusting to the cosmetic and functional aspects of the ileostomy. An unanticipated complication of proctocolectomy has been reduced fecundity in women.[153]

Sphincter-saving procedures Because UC is essentially limited to the colonic mucosa, alternative procedures have been developed that involve removing the proximal colon, stripping the rectal mucosa off the distal rectal musculature, and sparing the anal sphincter.[154] These so-called sphincter-saving procedures afford the opportunity of curing colitis and reestablishing continuity between the ileum and anus via an ileoanal anastomosis. Although a direct communication is technically feasible, it would result in intolerable postoperative diarrhea. Thus, these procedures include the provision of an ileal pouch to provide reservoir function. These J-, S-, or W-shaped pouches are created by folding the distal ileum and anastomosing the outlet to the anal canal.

Additional surgical modifications include the actual surgical stripping of the distal rectal mucosa or the stapling of the distal ileum to a short strip of residual rectal mucosa. In experienced surgical centers, the outcomes of stripping and stapling procedures are comparable.

Most often, sphincter-saving procedures are performed in stages: first, the surgeon removes the colon and performs an ileostomy, leaving the distal rectum as a Hartmann pouch; then, the surgeon creates the ileoanal anastomosis with a diverting ileostomy; and finally, the surgeon closes the ileostomy, which allows continuity of enteric flow through the pouch. Depending on surgeon preference and patient status, these procedures can be performed in one, two, or three stages.

Quality of life after an ileoanal anastomosis is excellent. Most patients describe full continence, with an average of six unformed (but not urgent) bowel movements daily.[155] Approximately 10% of patients develop small bowel obstructions, either between stages or after completion of the procedure. The most common complication after colectomy and ileoanal anastomosis is the development of so-called pouchitis, a superficial inflammation within the pouch that is similar to the inflammation of UC.[156] Pouchitis presents as increased urgency and evacuations that may be associated with bleeding and extraintestinal manifestations such as arthralgias, fever, and malaise. Most episodes of pouchitis respond to a course of an antibiotic, such as metronidazole or ciprofloxacin.[103] Approximately 15% of patients who develop pouchitis will have a more chronic course requiring long-term antibiotics or oral budesonide.[156] There is also evidence that high doses of the probiotic agent VSL#3 can prevent the onset of pouchitis or maintain remission of pouchitis after antibiotic therapy for recurrent or refractory pouchitis.[157-159] As with proctocolectomy, ileoanal anastomoses have been associated with reduced fecundity in women.[153]

TREATMENT OF CROHN DISEASE

Gastroduodenal Crohn Disease

Dyspepsia, epigastric burning, or nausea in CD patients with gastroduodenal involvement usually responds to acid reduction therapy with an H_2 receptor antagonist or a proton pump inhibitor.[160,161] More profound nausea or vomiting responds to corticosteroids, followed by an immunomodulator for steroid-sparing effects. Gastric outlet obstruction that does not respond to steroids or immunomodulators, or both, is an indication for surgical decompression with a gastrojejunostomy.

Jejunoileitis

In patients with isolated proximal small bowel CD, diarrhea should be evaluated from a mechanistic standpoint. Malabsorption from short bowel syndrome or resection is treated with a low-fat diet, whereas small bowel bacterial overgrowth is managed with antibiotics. Patients presenting with prominent pain or small bowel obstruction are treated with short-term corticosteroids and then usually with AZA or 6-MP. Bowel obstructions that do not respond to short-term steroid therapy require surgical resection or, more commonly, stricturoplasty.

Ileitis, Ileocolitis, and Colitis

In patients with limited ileal or ileocolonic CD, therapy should be staged to alleviate presenting symptoms, then to maintain long-term well-being, while minimizing chronic complications related to the disease or therapy.

Mild to moderate disease Outpatient therapy with anti-inflammatory agents, symptom-specific medications, and diet is utilized in patients who have abdominal pain and tenderness, diarrhea, low-grade fevers, weight loss without obstruction, painful mass, or severe malnutrition. Aminosalicylates, including sulfasalazine and mesalamine, have been effective at relatively higher doses than those that are used in UC. Sulfasalazine is effective in dosages of 3 to 6 g/day for ileocolonic and colonic CD but has not had significant benefits in limited small bowel disease.[27] In contrast, mesalamine, 4 g/day, provides modest benefits for small bowel and colonic involvement if the formulation releases the mesalamine at involved segments.[162] Benefits of sulfasalazine and mesalamine are modestly better than those achieved with placebo and less than those achieved with corticosteroids.[163] Enteric-coated, delayed-release budesonide, 9 mg daily, is an alternative first-line therapy for mild to moderate CD affecting the ileum, right colon, or both; it is more effective than mesalamine therapy.[163] Despite the limited potency of aminosalicylates, however, their potential long-term efficacy and absence of side effects make them first-line agents. As long as the patient continues to respond, the medication should be continued at the same dose used for induction. For patients with ileal or right colon disease who fail to respond within a short time (i.e., 2 to 4 weeks), budesonide therapy is a logical alternative.

Antibiotic therapy with metronidazole or ciprofloxacin, alone or in combination, is an alternative to aminosalicylates for ileocolonic and colonic CD.[27,102] Although there are no long-term data regarding antibiotic therapy for ileal or ileocolonic CD, clinical observations have suggested that maintenance therapy is

likely to be necessary. Patients receiving long-term metronidazole therapy should be monitored for peripheral neuropathy.

Dietary and nutritional therapy should focus on reduction of symptoms, prevention or correction of nutritional deficits, and avoidance of long-term complications. Elemental diets have short-term efficacy but are not practical for the majority of adult patients.[27,138] The disease location, complications, and surgical history will direct attention to potential nutritional deficiencies. Calorie and protein requisites are the primary concern. Secondary considerations include maintenance of iron stores and levels of water-soluble vitamins in the setting of proximal small bowel disease, as well as levels of vitamins B_{12}, A, D, and E with ileal disease or resection. Adequate calcium and vitamin D intake are of particular importance to avoid metabolic bone disease.

Moderate to severe disease A different therapeutic approach is required in patients who fail to respond to aminosalicylates or steroids or who present with fever, greater than 10% weight loss, and abdominal pain accompanied by tenderness (without obstruction) but who are able to maintain oral intake. Corticosteroids are required to induce a clinical remission; however, clinicians must exclude perforating complications (i.e., abscesses) before starting corticosteroids. Oral treatment with prednisone, 0.5 to 1 mg/kg, reduces symptoms in most patients.[27,75] However, the clinical response usually does not persist after steroid tapering; approximately 70% of patients will have a relapse or become steroid dependent within 1 year.[164]

Prednisone is continued at the initial dose until the patient responds completely (i.e., resolution of inflammatory symptoms). The dose is then decreased according to the time course to response. Tapering can usually proceed by 5 to 10 mg/wk, until a daily dose of 20 mg is reached, and then by 2.5 to 5 mg/wk. Calcium and vitamin D supplements reduce the risk of accelerated osteoporosis; in patients with reduced bone density, bisphosphonate or calcitonin should also be considered.[150] Clinical monitoring is continued, with attention paid to relapse of inflammatory symptoms. Persisting noninflammatory symptoms (e.g., nonbloody diarrhea or abdominal cramps) can be treated with dietary modifications, antispasmodics, and antidiarrheals without intensifying anti-inflammatory therapy.

Severe disease Infliximab may be beneficial in CD patients whose symptoms persist despite use of oral corticosteroids.[119] A single infusion of 5 mg/kg provides significant improvement and clinical remission for patients who have not responded to aminosalicylates, antibiotics, steroids, or immune suppressants. However, a three-dose induction regimen followed by maintenance infusions given every 8 weeks affords a more optimal long-term approach. Concomitant treatment with an immune suppressant minimizes the risk of developing antibodies to infliximab that are associated with loss of response and infusion reactions.[119] Before receiving infliximab, patients should be interviewed regarding exposure to tuberculosis and should undergo skin testing and chest x-rays.

Patients presenting with dehydration, high fever, cachexia, GI bleeding, obstructive symptoms, rebound tenderness, or an abscess require hospitalization and resuscitation with intravenous fluids, electrolytes, or transfusion. Acute obstructive symptoms, in the absence of chronic symptoms, mandate assessment for a mechanical cause (i.e., adhesions) rather than an inflammatory narrowing. Parenteral nutritional support is indicated as a supplement for patients unable to tolerate sufficient caloric intake and is mandatory for patients with profound malnutrition and an inability to eat.

Intravenous corticosteroids are indicated for severe manifestations of CD, once an abscess has been ruled out.[27] Prednisolone (40 to 60 mg), hydrocortisone (200 to 300 mg), or methylprednisolone (32 to 48 mg) is given in an intermittent or continuous infusion and continued until the patient is free of pain and is passing flatus and stool or until the patient no longer has diarrhea. Oral steroids are then substituted at an equivalent dose. Failure to improve with intravenous steroids should cause consideration of surgical intervention, prolonged total parenteral nutrition and bowel rest, use of intravenous cyclosporine, or use of infliximab. Broad-spectrum antibiotics are added for febrile patients and those with abdominal tenderness or an inflammatory mass. These agents are continued until defervescence unless a specific pathogen has been identified and a narrow-spectrum agent can be substituted.

Maintenance Therapy

Maintenance therapy for CD was once discounted because of poor results from early studies that evaluated low-dose sulfasalazine and corticosteroids. However, it is now apparent that other forms of maintenance therapy can reduce the possibility of clinical relapse in certain patients.[165] As with UC, steroids are not effective and should not be routinely used as maintenance agents for CD.

The aminosalicylates are useful maintenance agents when continued after inductive therapy, but they have limited value after steroid-induced remissions.[27,66] In contrast, immunomodulators have been shown to have steroid-sparing and maintenance benefits. AZA and 6-MP are effective steroid-sparing agents for patients who cannot be weaned from steroids. Initial therapy is AZA, 2 to 2.5 mg/kg, or 6-MP, 1 to 1.5 mg/kg. The dosage is adjusted at 2-week intervals according to the leukocyte count, which must be maintained above leukopenic levels. The efficacy of these agents may not be evident until after 3 to 6 months of treatment, but benefits have been demonstrated to last for at least 4 years. To avoid unanticipated bone marrow suppression, monitoring of blood counts on a quarterly basis must continue once the patient is off steroids.

Maintenance therapy to delay postsurgical relapse has been evaluated according to different end-point criteria (e.g., endoscopic evidence of relapse, clinical symptoms, or repeat surgery).[39] There is evidence that mesalamine, 3 to 4 g/day, can prevent postoperative recurrence, particularly when therapy is initiated shortly after surgery[27]; conversely, postponing therapy for more than 3 months circumvents any benefits. In addition to mesalamine, metronidazole is effective for reducing postoperative recurrence when administered at high doses (20 mg/kg) for 3 months after resection. There are no data regarding lower doses of metronidazole, prolongation of therapy beyond 3 months, or the use of other antibiotics.[27,102] Finally, 6-MP at doses of 50 mg/day may reduce postoperative relapse for at least 2 years after resection.[166] In view of the negative impact of cigarette smoking on the postoperative course of CD, however, all patients who smoke should be advised to stop.[27]

Surgical Therapy

Unlike UC, CD is not cured by surgery, except in the case of CD confined to the colon, for which proctocolectomy and ileostomy provide similar likelihoods of cure; sphincter-saving

procedures are not advocated because of the high likelihood of recurrence after anastomoses. With CD in other locations, disease recurrence at the anastomotic site is virtually inevitable. Therefore, surgery is undertaken to treat refractory disease or complications, rather than to cure the disease. Nevertheless, in view of the excellent quality of life after limited surgery and the evolving capability to reduce or delay recurrence, it is imperative that surgery not be deferred because of fear of recurrence.

Purulent complications (e.g., abscesses) require percutaneous or surgical drainage. Surgery in CD patients is also indicated for intractable hemorrhage, perforation, persisting or recurrent obstruction, or toxic megacolon. The most common indications for surgery are intractable disease, failure of medical therapy, or complications related to treatment (e.g., steroid dependence). Many of these indications are subjective, requiring experienced clinical judgment and cooperative consultation between medical and surgical specialists. Surgical resections should be limited to macroscopic disease, and in general, primary anastomoses should not be performed in the setting of uncontained purulent complications.[27]

IBD AND PREGNANCY

Fertility is usually normal in both men and women with IBD, although an increase in disease activity correlates with a decrease in libido and with menstrual irregularities.[167] Risk of early miscarriage is increased in women with active disease. When disease activity is controlled, the course and outcome of pregnancy do not differ substantially from those in the general population. Therefore, the best means of ensuring normal fetal outcome is to time conception when disease is under control, aggressively treat disease activity during pregnancy, and maintain the health of the mother.[27] Aminosalicylates, steroids, and immunomodulators are safe during pregnancy and lactation, but they should be added only when necessary to maintain maternal well-being.[88,168,169] Conversely, because of the risk of maternal worsening, neither acute nor maintenance therapy should be withdrawn during or after pregnancy. Attention to the mother's nutritional status is essential throughout pregnancy and during lactation. Neonates born to women on high doses of steroids should be monitored for adrenal suppression.

Complications

INTESTINAL COMPLICATIONS

The intestinal complications of IBD include hemorrhage, stricture, fistulas, toxic megacolon, and neoplasia. Chronic blood loss, with subsequent iron deficiency anemia, is common in both UC and CD. Profuse bleeding, however, is uncommon, particularly in UC, because the inflammation is superficial. Occasionally, patients with CD experience severe lower gastrointestinal bleeding when deep ulcerations erode into large vessels.

Strictures are more common in CD than in UC and result from transmural inflammation and fibrosis. These strictures remain fixed and lead to progressive bowel obstruction. In UC, narrowing of the lumen can occur from smooth muscle hypertrophy; the narrowing is related to disease activity and is reversible with treatment of acute inflammation. Fixed strictures in UC are almost always dysplastic or malignant. Toxic megacolon, although more common in UC, is not unique to UC and can occur in infectious colitis or CD.[170]

EXTRAINTESTINAL COMPLICATIONS

The extraintestinal complications of IBD can result from inflammation or from an HLA-related autoimmune process that underlies the intestinal disease.[171] Complications may also occur as a metabolic consequence of intestinal disease or its treatment.[29]

Mucocutaneous complications include eye changes of episcleritis or scleritis. These complications most commonly parallel colonic disease activity. Involvement of the anterior or posterior chambers with iritis or uveitis is related to HLA-B27 and follows a course that is independent of disease activity in the bowel.[172] Skin lesions of erythema nodosum and pyoderma gangrenosum usually accompany or herald the onset of colitis and respond to treatment of bowel inflammation.

Musculoskeletal lesions can either be independent of or correlate with intestinal disease activity. Peripheral arthralgias and arthritis commonly involve larger joints (e.g., the hips, knees, ankles, elbows, and wrists) in an asymmetrical pattern. The inflammation usually accompanies intestinal disease activity and is almost never deforming, progressive, or associated with rheumatoid nodules. In contrast, arthritis of the central spine, ankylosing spondylitis, and sacroiliitis are associated with HLA-B27 and progress independently of intestinal disease. Metabolic bone disease is most often a consequence of long-term steroid use and, in CD, can be accelerated by malabsorption or inadequate dietary supplementation with vitamin D or calcium.[150]

A spectrum of hepatobiliary involvement occurs in both UC and colonic CD.[173] Inflammation of the intrahepatic and extrahepatic bile ducts may take the form of a mild, periportal inflammatory infiltrate (pericholangitis or small-duct sclerosing cholangitis) that is asymptomatic, nonprogressive, and manifested only as mild elevations of γ-glutamyltransferase (GGT), alkaline phosphatase, and transaminases. At the other extreme, such inflammation may lead to full-blown sclerosing cholangitis with progressive secondary biliary cirrhosis. Of interest, cigarette smoking, which protects against UC, also protects against primary sclerosing cholangitis.[9] Hepatic steatosis commonly manifests as a mild elevation of biliary enzymes in the presence of malnutrition or steroid therapy. Patients with CD who have ileal involvement or have had ileal resections are at increased risk for gallstones because of reduced enterohepatic circulation of bile salts, which increases biliary cholesterol saturation.

Urinary tract complications are more common in CD than UC.[174] Kidney stones may reflect dehydration from diarrhea or an ileostomy.[175] In the setting of ileitis or after ileal resections, the mechanism of nephrolithiasis is hyperoxaluria caused by steatorrhea. Normally, oxalate in the diet binds to free calcium in the colonic lumen and is excreted in the feces as calcium oxalate crystals. In patients with steatorrhea, free luminal calcium preferentially binds to fatty acids, creating soaps that are similarly excreted in the feces. Lacking calcium to bind it, the free oxalate is instead absorbed by the colon and excreted by the kidneys in abnormally high amounts. In the urine, oxalate complexes with urinary calcium to form calcium oxalate crystals. Patients with CD are also more susceptible to nephrolithiasis because their kidneys excrete low amounts of citrate, which acts as a nonspecific solubilizer in the urine. Calcium oxalate stones may result from either hyperoxaluria or idiopathic hypercalciuria; the two conditions can be differentiated by measuring calcium and oxalate levels in a 24-hour urine sample. If hyperoxaluria is identified, the treatment is to reduce fat intake (to reduce steatorrhea) and increase calcium supplementation.

Hematologic complications of IBD include anemia and clotting abnormalities. Anemia is most often caused by iron loss from bleeding or, in CD, from impaired iron absorption because of proximal small bowel disease.[176] Folic acid deficiency is most often related to concurrent use of sulfasalazine, but occasionally, it is related to inadequate dietary consumption of folic acid or extensive jejunal disease. Vitamin B_{12} deficiency from extensive ileal disease or resection can lead to macrocytic anemia. Hypercoagulability is a nonspecific complication of active IBD that results from increased production of acute-phase reactants and that increases the risk of venous thrombosis.[177] Rarely, enteric losses of anticoagulant factors are a consequence of protein-losing enteropathy. In contrast, hypocoagulability may be a complication of vitamin K malabsorption or prolonged antibiotic administration.

Management of Extraintestinal Complications

Treatment of extraintestinal complications of IBD varies according to whether the manifestations are dependent or independent of intestinal (usually colonic) inflammation.[29,171] Peripheral arthritis, erythema nodosum, pyoderma gangrenosum, and episcleritis occur in the presence of active disease and require intensification of anti-inflammatory therapy.

In some patients, complications can be treated independently, along with therapy for intestinal inflammation. For example, inflamed joints can be drained or injected with steroids, and pyoderma gangrenosum can be approached with a combination of topical and systemic approaches. Infliximab therapy has been very effective for the treatment of pyoderma gangrenosum associated with CD.[178] Ocular complications should be evaluated by an ophthalmologist to prevent irreversible damage. Erythema nodosum usually responds to more aggressive therapy for colitis and should not be treated with NSAIDs. Peripheral articular manifestations of colitis can be treated with acetaminophen and increased doses of sulfasalazine; again, NSAIDs should be avoided.

Ankylosing spondylitis, sacroiliitis, and iritis (uveitis) are HLA-B27–associated manifestations that follow a course independent from colitis. The same is true of primary sclerosing cholangitis. Physical therapy is critical for patients with ankylosing spondylitis and sacroiliitis. Concurrent immunomodulatory therapy with methotrexate, hydroxychloroquine, or infliximab is indicated for inflammatory arthropathies.

NEOPLASIA

Adenocarcinoma of the intestines is a potential long-term complication of IBD, with features that are distinct from those of spontaneous adenocarcinomas in the general population.[179] The risks of colonic neoplasia are similar in UC and CD and historically have been related to the extent and duration of disease, age at onset, and stricture formation, with primary sclerosing cholangitis indicating an increased risk for cholangiocarcinoma. In addition, evidence now suggests that in UC, severity of inflammation is also a risk factor for development of neoplasia.[180] In contrast to colorectal cancer in the general population, which develops from adenomatous polyps, dysplasia is the precursor to cancer in IBD patients.[34] Dysplasia is a neoplasia and has been defined on a pathologic basis and categorized as indefinite, low grade, or high grade (carcinoma in situ).[34] Patients with dysplasia have an increased risk of cancer elsewhere in the colon if they have high-grade lesions (up to 50% of these patients have other colonic cancers) or low-grade lesions that are multifocal or in a raised plaque (dysplasia-associated lesion or mass).[181]

The ability to identify histologic dysplasia in UC makes it possible to perform surveillance colonoscopic examinations. Despite the absence of prospective data regarding colonoscopic surveillance for dysplasia in UC, most North American centers recommend that, after 8 to 10 years of UC, patients who are at increased risk for adenocarcinoma be entered into a colonoscopic surveillance program.[26,181,182] Surveillance colonoscopies are recommended every 2 to 3 years for patients with 10 to 20 years of disease and every 1 to 2 years for patients with over 20 years of disease. The confirmation of low-grade dysplasia on surveillance colonoscopy strongly predicts progression to advanced neoplasia and warrants a recommendation for colectomy.[183] Patients with indefinite dysplasia are treated aggressively to control inflammation and should undergo repeat colonoscopy after 3 to 6 months.

The risk of cancer in CD is also related to the location and chronicity of inflammation and, thus, includes a risk for small bowel adenocarcinoma in long-standing small bowel disease.[184] However, because of the varied locations and segmental involvement in CD, there are no standardized guidelines for surveillance. Although there is no means of screening for small bowel dysplasia or cancer in CD, patients with colonic CD can be followed with colonoscopic surveillance in a manner similar to that used in patients with UC. Unfortunately, stricture formation in these patients often prevents visualization of the entire colon.

In addition to being associated with adenocarcinomas, IBD is associated with a small increase in the relative risk of lymphomas.[185] This increase may be associated with immunosuppressive therapy.[186-188] However, it is difficult to tease out any small increase in the relative risk from immunomodulatory therapy from the underlying risk associated with the chronic inflammatory disease and a potential association with Epstein-Barr virus infection.[189,190]

Prognosis and Conclusion

The diagnosis and treatment of IBD challenge the physician to guide patients through chronic illness. It is important to recognize the perceptions and concerns of the patient and family members confronted with a chronic, medically incurable, socially embarrassing, and potentially disfiguring condition.[27] Despite the absence of known causes, medical therapy is usually effective, and surgical techniques have improved to the point that both longevity and quality of life can be preserved. The life expectancy of patients without fulminant disease is the same as that of the general population. Patience, optimism, and empathy are required to balance the concerns and misinformation that surround the disease, as well as the guilt associated with misconceptions that IBD is a so-called neurotic, psychosomatic, or self-induced disorder.

Patient information and support groups constitute a valuable part of management in IBD. Resources for high-quality patient information and support are available through national organizations such as the Crohn's and Colitis Foundation of America (http://www.ccfa.org).

The author has received grant support from Centocor, Inc, The Procter & Gamble Company, and Prometheus Laboratories, Inc.; has been a consultant for Centocor, Inc., The Procter & Gamble Company, Prometheus Laboratories,

Inc., Salix Pharmaceuticals, Inc., and Shire Pharmaceuticals Group; and has participated in speakers' bureaus for Centocor, Inc., The Procter & Gamble Company, and Salix Pharmaceuticals, Inc.

References

1. Podolsky DK: Inflammatory bowel disease. N Engl J Med 347:417, 2002

2. Loftus EV Jr, Sandborn WJ: Epidemiology of inflammatory bowel disease. Gastroenterol Clin North Am 31:1, 2002

3. Cashman KD, Shanahan F: Is nutrition an aetiological factor for inflammatory bowel disease? Eur J Gastroenterol Hepatol 15:607, 2003

4. Hendrickson BA, Gokhale R, Cho JH: Clinical aspects and pathophysiology of inflammatory bowel disease. Clin Microbiol Rev 15:79, 2002

5. Mathew CG, Lewis CM: Genetics of inflammatory bowel disease: progress and prospects. Hum Mol Genet 13:R161, 2004

6. Halfvarson J, Bodin L, Tysk C, et al: Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. Gastroenterology 124:1767, 2003

7. Cosnes J: Tobacco and IBD: relevance in the understanding of disease mechanisms and clinical practice. Best Pract Res Clin Gastroenterol 18:481, 2004

8. Beaugerie L, Massot N, Carbonnel F, et al: Impact of cessation of smoking on the course of ulcerative colitis. Am J Gastroenterol 96:2113, 2001

9. Loftus EV Jr, Sandborn WJ, Tremaine WJ, et al: Primary sclerosing cholangitis is associated with nonsmoking: a case-control study. Gastroenterology 110:1496, 1996

10. Cosnes J, Beaugerie L, Carbonnel F, et al: Smoking cessation and the course of Crohn's disease: an intervention study. Gastroenterology 120:1093, 2001

11. Sandborn WJ: Nicotine therapy for ulcerative colitis: a review of rationale, mechanisms, pharmacology, and clinical results. Am J Gastroenterol 94:1161, 1999

12. Shanahan F: The host-microbe interface within the gut. Best Pract Res Clin Gastroenterol 16:915, 2002

13. Danese S, Sans M, Fiocchi C: Inflammatory bowel disease: the role of environmental factors. Autoimmun Rev 3:394, 2004

14. Landers CJ, Cohavy O, Misra R, et al: Selected loss of tolerance evidenced by Crohn's disease–associated immune responses to auto- and microbial antigens. Gastroenterology 123:689, 2002

15. Stagg AJ, Hart AL, Knight SC, et al: The dendritic cell: its role in intestinal inflammation and relationship with gut bacteria. Gut 52:1522, 2003

16. Sartor RB: Innate immunity in the pathogenesis and therapy of IBD. J Gastroenterol 38(suppl 15):43, 2003

17. Fellermann K, Wehkamp J, Herrlinger KR, et al: Crohn's disease: a defensin deficiency syndrome? Eur J Gastroenterol Hepatol 15:627, 2003

18. Bouma G, Strober W: The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 3:521, 2003

19. Shanahan F: Inflammatory bowel disease: immunodiagnostics, immunotherapeutics, and ecotherapeutics. Gastroenterology 120:622, 2001

20. Papadakis KA, Targan SR: The role of chemokines and chemokine receptors in mucosal inflammation. Inflamm Bowel Dis 6:303, 2000

21. Langholz E, Munkholm P, Davidsen M, et al: Changes in extent of ulcerative colitis: a study on the course and prognostic factors. Scand J Gastroenterol 31:260, 1996

22. Moum B, Ekbom A, Vatn MH, et al: Change in the extent of colonoscopic and histological involvement in ulcerative colitis over time. Am J Gastroenterol 94:1564, 1999

23. Schumacher G, Sandstedt B, Kollberg B: A prospective study of first attacks of inflammatory bowel disease and infectious colitis: clinical findings and early diagnosis. Scand J Gastroenterol 29:265, 1994

24. Farmer RG, Easley KA, Rankin GB: Clinical patterns, natural history, and progression of ulcerative colitis: a long-term follow-up of 1116 patients. Dig Dis Sci 38:1137, 1993

25. Gan SI, Beck PL: A new look at toxic megacolon: an update and review of incidence, etiology, pathogenesis, and management. Am J Gastroenterol 98:2363, 2003

26. Kornbluth A, Sachar DB: Ulcerative colitis practice guidelines in adults (update): American College of Gastroenterology, Practice Parameters Committee. Am J Gastroenterol 99:1371, 2004

27. Carter MJ, Lobo AJ, Travis SP: Guidelines for the management of inflammatory bowel disease in adults. Gut 53(suppl 5):V1, 2004

28. Rizzello F, Gionchetti P, Venturi A, et al: Review article: monitoring activity in ulcerative colitis. Aliment Pharmacol Ther 16(suppl 4):3, 2002

29. Su CG, Judge TA, Lichtenstein GR: Extraintestinal manifestations of inflammatory bowel disease. Gastroenterol Clin North Am 31:307, 2002

30. Chutkan RK, Scherl E, Waye JD: Colonoscopy in inflammatory bowel disease. Gastrointest Endosc Clin N Am 12:463, 2002

31. D'Haens G, Geboes K, Peeters M, et al: Patchy cecal inflammation associated with distal ulcerative colitis: a prospective endoscopic study. Am J Gastroenterol 92:1275, 1997

32. Geboes K, Dalle I: Influence of treatment on morphological features of mucosal inflammation. Gut 50(suppl 3):III37, 2002

33. Tanaka M, Riddell RH, Saito H, et al: Morphologic criteria applicable to biopsy specimens for effective distinction of inflammatory bowel disease from other forms of colitis and of Crohn's disease from ulcerative colitis. Scand J Gastroenterol 34:55, 1999

34. Riddell RH, Goldman H, Ransohoff DF, et al: Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. Hum Pathol 14:931, 1983

35. Martin-Comin J, Prats E: Clinical applications of radiolabeled blood elements in inflammatory bowel disease. Q J Nucl Med 43:74, 1999

36. Farmer RG, Whelan G, Fazio VW: Long-term follow-up of patients with Crohn's disease: relationship between the clinical pattern and prognosis. Gastroenterology 88:1818, 1985

37. Gasche C, Scholmerich J, Brynskov J, et al: A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 6:8, 2000

38. Cosnes J, Cattan S, Blain A, et al: Long-term evolution of disease behavior of Crohn's disease. Inflamm Bowel Dis 8:244, 2002

39. Lewis JD, Schoenfeld P, Lichtenstein GR: An evidence-based approach to studies of the natural history of gastrointestinal diseases: recurrence of symptomatic Crohn's disease after surgery. Clin Gastroenterol Hepatol 1:229, 2003

40. Hanauer SB: Measurement of disease activity. Inflammatory Bowel Disease: From Bench to Bedside, 1st ed. Targan SR, Shanahan F, Eds. Williams & Wilkins, Baltimore, 1994, p 429

41. Cohen RD: The quality of life in patients with Crohn's disease. Aliment Pharmacol Ther 16:1603, 2002

42. Bernstein CN, Blanchard JF, Rawsthorne P, et al: The prevalence of extraintestinal diseases in inflammatory bowel disease: a population-based study. Am J Gastroenterol 96:1116, 2001

43. Loftus EV Jr, Schoenfeld P, Sandborn WJ: The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. Aliment Pharmacol Ther 16:51, 2002

44. Schwartz DA, Loftus EV Jr, Tremaine WJ, et al: The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. Gastroenterology 122:875, 2002

45. Langholz E, Munkholm P, Krasilnikoff PA, et al: Inflammatory bowel diseases with onset in childhood: clinical features, morbidity, and mortality in a regional cohort. Scand J Gastroenterol 32:139, 1997

46. Mamula P, Telega GW, Markowitz JE, et al: Inflammatory bowel disease in children 5 years of age and younger. Am J Gastroenterol 97:2005, 2002

47. Decker GA, Loftus EV Jr, Pasha TM, et al: Crohn's disease of the esophagus: clinical features and outcomes. Inflamm Bowel Dis 7:113, 2001

48. Wagtmans MJ, Verspaget HW, Lamers CB, et al: Clinical aspects of Crohn's disease of the upper gastrointestinal tract: a comparison with distal Crohn's disease. Am J Gastroenterol 92:1467, 1997

49. Higuero T, Merle C, Thiefin G, et al: Jejunoileal Crohn's disease: a case-control study. Gastroenterol Clin Biol 28:160, 2004

50. Guindi M, Riddell RH: Indeterminate colitis. J Clin Pathol 57:1233, 2004

51. Sandborn WJ, Fazio VW, Feagan BG, et al: AGA technical review on perianal Crohn's disease. Gastroenterology 125:1508, 2003

52. Oldenburg B, Koningsberger JC, Van Berge Henegouwen GP, et al: Iron and inflammatory bowel disease. Aliment Pharmacol Ther 15:429, 2001

53. Vermeire S, Van Assche G, Rutgeerts P: C-reactive protein as a marker for inflammatory bowel disease. Inflamm Bowel Dis 10:661, 2004

54. Wassell J, Dolwani S, Metzner M, et al: Faecal calprotectin: a new marker for Crohn's disease? Ann Clin Biochem 41:230, 2004

55. Kane SV, Sandborn WJ, Rufo PA, et al: Fecal lactoferrin is a sensitive and specific marker in identifying intestinal inflammation. Am J Gastroenterol 98:1309, 2003

56. Nakamura RM, Matsutani M, Barry M: Advances in clinical laboratory tests for inflammatory bowel disease. Clin Chim Acta 335:9, 2003

57. Zalis M, Singh AK: Imaging of inflammatory bowel disease: CT and MR. Dig Dis 22:56, 2004

58. Alberini JL, Badran A, Freneaux E, et al: Technetium-99m HMPAO-labeled leukocyte imaging compared with endoscopy, ultrasonography, and contrast radiology in children with inflammatory bowel disease. J Pediatr Gastroenterol Nutr 32:278, 2001

59. Kornbluth A, Legnani P, Lewis BS: Video capsule endoscopy in inflammatory bowel disease: past, present, and future. Inflamm Bowel Dis 10:278, 2004

60. Cipolla G, Crema F, Sacco S, et al: Nonsteroidal anti-inflammatory drugs and inflammatory bowel disease: current perspectives. Pharmacol Res 46:1, 2002

61. Alapati SV, Mihas AA: When to suspect ischemic colitis: why is this condition so often missed or misdiagnosed? Postgrad Med 105:177, 1999

62. Peppercorn MA: The overlap of inflammatory bowel disease and diverticular disease. J Clin Gastroenterol 38(5 suppl):S8, 2004

63. Sands BE: From symptom to diagnosis: clinical distinctions among various forms of intestinal inflammation. Gastroenterology 126:1518, 2004

64. Stallmach A, Carstens O: Role of infections in the manifestation or reactivation of inflammatory bowel diseases. Inflamm Bowel Dis 8:213, 2002

65. Gisbert JP, Gomollon F, Mate J, et al: Role of 5-aminosalicylic acid (5-ASA) in treatment of inflammatory bowel disease: a systematic review. Dig Dis Sci 47:471, 2002

66. Lim WC, Hanauer SB: Controversies with aminosalicylates in inflammatory bowel disease. Rev Gastroenterol Disord 4:104, 2004

67. Sandborn WJ, Hanauer SB: The pharmacokinetic profiles of oral mesalazine formulations and mesalazine pro-drugs used in the management of ulcerative colitis. Aliment Pharmacol Ther 17:29, 2003

68. Allgayer H: Review article: mechanisms of action of mesalazine in preventing colorectal carcinoma in inflammatory bowel disease. Aliment Pharmacol Ther 18(suppl 2):10, 2003

69. Wahl C, Liptay S, Adler G, et al: Sulfasalazine: a potent and specific inhibitor of nu-

clear factor kappa B. J Clin Invest 101:1163, 1998

70. Walker AM, Szneke P, Bianchi LA, et al: 5-Aminosalicylates, sulfasalazine, steroid use, and complications in patients with ulcerative colitis. Am J Gastroenterol 92:816, 1997

71. Cunliffe RN, Scott BB: Monitoring for drug side-effects in inflammatory bowel disease. Aliment Pharmacol Ther 16:647, 2002

72. Sutherland L, MacDonald JK: Oral 5-aminosalicylic acid for induction of remission in ulcerative colitis. Cochrane Database Syst Rev (3):CD000543, 2003

73. Cohen RD, Woseth DM, Thisted RA, et al: A meta-analysis and overview of the literature on treatment options for left-sided ulcerative colitis and ulcerative proctitis. Am J Gastroenterol 95:1263, 2000

74. Sutherland L, Roth D, Beck P, et al: Oral 5-aminosalicylic acid for maintenance of remission in ulcerative colitis. Cochrane Database Syst Rev (4):CD000544, 2002

75. Katz JA: Treatment of inflammatory bowel disease with corticosteroids. Gastroenterol Clin North Am 33:171, 2004

76. Steinhart AH, Ewe K, Griffiths AM, et al: Corticosteroids for maintenance of remission in Crohn's disease. Cochrane Database Syst Rev (4):CD000301, 2003

77. Farrell RJ, Kelleher D: Glucocorticoid resistance in inflammatory bowel disease. J Endocrinol 178:339, 2003

78. McKeage K, Goa KL: Budesonide (Entocort EC Capsules): a review of its therapeutic use in the management of active Crohn's disease in adults. Drugs 62:2263, 2002

79. Bebb JR, Scott BB: How effective are the usual treatments for ulcerative colitis? Aliment Pharmacol Ther 20:143, 2004

80. Ardizzone S, Porro GB: Comparative tolerability of therapies for ulcerative colitis. Drug Saf 25:561, 2002

81. Marshall JK, Irvine EJ: Rectal corticosteroids versus alternative treatments in ulcerative colitis: a meta-analysis. Gut 40:775, 1997

82. Bebb JR, Scott BB: How effective are the usual treatments for Crohn's disease? Aliment Pharmacol Ther 20:151, 2004

83. Kane SV, Schoenfeld P, Sandborn WJ, et al: The effectiveness of budesonide therapy for Crohn's disease. Aliment Pharmacol Ther 16:1509, 2002

84. Sandborn WJ: Azathioprine: state of the art in inflammatory bowel disease. Scand J Gastroenterol Suppl 225:92, 1998

85. Tiede I, Fritz G, Strand S, et al: CD28-dependent Rac1 activation is the molecular target of azathioprine in primary human CD4+ T lymphocytes. J Clin Invest 111:1133, 2003

86. Louis E, Belaiche J: Optimizing treatment with thioguanine derivatives in inflammatory bowel disease. Best Pract Res Clin Gastroenterol 17:37, 2003

87. Vial T, Descotes J: Immunosuppressive drugs and cancer. Toxicology 185:229, 2003

88. Polifka JE, Friedman JM: Teratogen update: azathioprine and 6-mercaptopurine. Teratology 65:240, 2002

89. Lichtenstein GR: Use of laboratory testing to guide 6-mercaptopurine/azathioprine therapy. Gastroenterology 127:1558, 2004

90. Sandborn W, Sutherland L, Pearson D, et al: Azathioprine or 6-mercaptopurine for inducing remission of Crohn's disease. Cochrane Database Syst Rev (2):CD000545, 2000

91. Pearson DC, May GR, Fick G, et al: Azathioprine for maintaining remission of Crohn's disease. Cochrane Database Syst Rev (2):CD000067, 2000

92. Loftus CG, Egan LJ, Sandborn WJ: Cyclosporine, tacrolimus, and mycophenolate mofetil in the treatment of inflammatory bowel disease. Gastroenterol Clin North Am 33:141, 2004

93. Hanauer SB: Medical therapy for ulcerative colitis 2004. Gastroenterology 126:1582, 2004

94. Van Assche G, D'Haens G, Noman M, et al: Randomized, double-blind comparison of 4 mg/kg versus 2 mg/kg intravenous cyclosporine in severe ulcerative colitis. Gastroenterology 125:1025, 2003

95. Sandborn WJ: Cyclosporine in ulcerative colitis: state of the art. Acta Gastroenterol Belg 64:201, 2001

96. Egan LJ, Sandborn WJ, Tremaine WJ: Clinical outcome following treatment of refractory inflammatory and fistulizing Crohn's disease with intravenous cyclosporine. Am J Gastroenterol 93:442, 1998

97. Feagan BG, Alfadhli A: Methotrexate in inflammatory bowel disease. Gastroenterol Clin North Am 33:407, 2004

98. Feagan BG, Rochon J, Fedorak RN, et al: Methotrexate for the treatment of Crohn's disease. The North American Crohn's Study Group Investigators. N Engl J Med 332:292 1995

99. Feagan BG, Fedorak RN, Irvine EJ, et al: A comparison of methotrexate with placebo for the maintenance of remission in Crohn's disease. North American Crohn's Study Group Investigators. N Engl J Med 342:1627, 2000

100. Sandborn WJ, Present DH, Isaacs KL, et al: Tacrolimus for the treatment of fistulas in patients with Crohn's disease: a randomized, placebo-controlled trial. Gastroenterology 125:380, 2003

101. Hogenauer C, Wenzl HH, Hinterleitner TA, et al: Effect of oral tacrolimus (FK 506) on steroid-refractory moderate/severe ulcerative colitis. Aliment Pharmacol Ther 18:415, 2003

102. Isaacs KL, Sartor RB: Treatment of inflammatory bowel disease with antibiotics. Gastroenterol Clin North Am 33:335, 2004

103. Sandborn W, McLeod R, Jewell D: Pharmacotherapy for inducing and maintaining remission in pouchitis. Cochrane Database Syst Rev (2):CD001176, 2000

104. Greenberg GR: Antibiotics should be used as first-line therapy for Crohn's disease. Inflamm Bowel Dis 10:318, 2004

105. Shanahan F, Bernstein CN: Antibiotics as a first-line therapy for Crohn's disease: is there any consensus? Inflamm Bowel Dis 10:324, 2004

106. Wild GE: The role of antibiotics in the management of Crohn's disease. Inflamm Bowel Dis 10:321, 2004

107. Hanauer SB, Sandborn W: Management of Crohn's disease in adults. Am J Gastroenterol 96:635, 2001

108. Achkar JP, Hanauer SB: Medical therapy to reduce postoperative Crohn's disease recurrence. Am J Gastroenterol 95:1139, 2000

109. Colombel JF, Lemann M, Cassagnou M, et al: A controlled trial comparing ciprofloxacin with mesalazine for the treatment of active Crohn's disease. Groupe d'Etudes Therapeutiques des Affections Inflammatoires Digestives (GETAID). Am J Gastroenterol 94:674, 1999

110. Prantera C, Zannoni F, Scribano ML, et al: An antibiotic regimen for the treatment of active Crohn's disease: a randomized, controlled clinical trial of metronidazole plus ciprofloxacin. Am J Gastroenterol 91:328, 1996

111. Greenbloom SL, Steinhart AH, Greenberg GR: Combination ciprofloxacin and metronidazole for active Crohn's disease. Can J Gastroenterol 12:53, 1998

112. Borgaonkar M, MacIntosh D, Fardy J, et al: Anti-tuberculous therapy for maintaining remission of Crohn's disease. Cochrane Database Syst Rev (2):CD000299, 2000

113. Akobeng AK, Zachos M: Tumor necrosis factor–alpha antibody for induction of remission in Crohn's disease. Cochrane Database Syst Rev (1):CD003574, 2004

114. Hanauer S, Lukas M, Macintosh D, et al: A randomized, double-blind, placebo-controlled trial of the human anti–TNF-alpha monoclonal antibody adalimumab for the induction of remission in patients with moderate to severely active Crohn's disease (abstr). Gastroenterology 127:332, 2004

115. Sands BE, Anderson FH, Bernstein CN, et al: Infliximab maintenance therapy for fistulizing Crohn's disease. N Engl J Med 350:876, 2004

116. Olson A: Personal communication, 2005

117. Colombel JF, Loftus EV Jr, Tremaine WJ, et al: The safety profile of infliximab in patients with Crohn's disease: the Mayo Clinic experience in 500 patients. Gastroenterology 126:19, 2004

118. Mow WS, Abreu-Martin MT, Papadakis KA, et al: High incidence of anergy in inflammatory bowel disease patients limits the usefulness of PPD screening before infliximab therapy. Clin Gastroenterol Hepatol 2:309, 2004

119. Rutgeerts P, Van Assche G, Vermeire S: Optimizing anti-TNF treatment in inflammatory bowel disease. Gastroenterology 126:1593, 2004

120. Khanna D, McMahon M, Furst DE: Safety of tumour necrosis factor–alpha antagonists. Drug Saf 27:307, 2004

121. Aithal GP, Mansfield JC: Review article: the risk of lymphoma associated with inflammatory bowel disease and immunosuppressive treatment. Aliment Pharmacol Ther 15:1101, 2001

122. Sandborn WJ, Yednock TA: Novel approaches to treating inflammatory bowel disease: targeting alpha-4 integrin. Am J Gastroenterol 98:2372, 2003

123. Mannon PJ, Fuss IJ, Mayer L, et al: Anti–interleukin-12 antibody for active Crohn's disease. N Engl J Med 351:2069, 2004

124. Lim WC, Hanauer SB: Emerging biologic therapies in inflammatory bowel disease. Rev Gastroenterol Disord 4:66, 2004

125. Sinha A, Nightingale J, West KP, et al: Epidermal growth factor enemas with oral mesalamine for mild-to-moderate left-sided ulcerative colitis or proctitis. N Engl J Med 349:350, 2003

126. Dieckgraefe BK, Korzenik JR: Treatment of active Crohn's disease with recombinant human granulocyte-macrophage colony-stimulating factor. Lancet 360:1478, 2002

127. McGrath J, McDonald J, Macdonald J: Transdermal nicotine for induction of remission in ulcerative colitis. Cochrane Database Syst Rev (4):CD004722, 2004

128. Aslan A, Triadafilopoulos G: Fish oil fatty acid supplementation in active ulcerative colitis: a double-blind, placebo-controlled, crossover study. Am J Gastroenterol 87:432, 1992

129. Loeschke K, Ueberschaer B, Pietsch A, et al: n-3 fatty acids only delay early relapse of ulcerative colitis in remission. Dig Dis Sci 41:2087, 1996

130. Belluzzi A, Brignola C, Campieri M, et al: Effect of an enteric-coated fish-oil preparation on relapses in Crohn's disease. N Engl J Med 334:1557, 1996

131. Banner KH, Trevethick MA: PDE4 inhibition: a novel approach for the treatment of inflammatory bowel disease. Trends Pharmacol Sci 25:430, 2004

132. van Deventer SJ: Small therapeutic molecules for the treatment of inflammatory bowel disease. Gut 50(suppl 3):III47, 2002

133. Shanahan F: Probiotics in inflammatory bowel disease—therapeutic rationale and role. Adv Drug Deliv Rev 56:809, 2004

134. Takazoe M, Tanaka T, Kondo K, et al: The present status and the recent development of the treatment for inflammatory bowel diseases: desirable effect of extracorporeal immunomodulation. Ther Apher 6:305, 2002

135. Danese S, Sans M, Fiocchi C: The CD40/CD40L costimulatory pathway in inflammatory bowel disease. Gut 53:1035, 2004

136. Graham TO, Kandil HM: Nutritional factors in inflammatory bowel disease. Gastroenterol Clin North Am 31:203, 2002

137. Gassull MA: Nutrition and inflammatory bowel disease: its relation to pathophysiology, outcome and therapy. Dig Dis 21:220, 2003

138. Zachos M, Tondeur M, Griffiths AM: Enteral nutritional therapy for inducing remission of Crohn's disease. Cochrane Database Syst Rev (3):CD000542, 2001

139. Miner PB Jr: Factors influencing the relapse of patients with inflammatory bowel disease. Am J Gastroenterol 92(12 suppl):1S, 1997

140. Kane SV, Sable K, Hanauer SB: The menstrual cycle and its effect on inflammatory bowel disease and irritable bowel syndrome: a prevalence study. Am J Gastroenterol 93:1867, 1998

141. Kane S, Kisiel J, Shih L, et al: HLA disparity determines disease activity through pregnancy in women with inflammatory bowel disease. Am J Gastroenterol 99:1523, 2004

142. Felder JB, Korelitz BI, Rajapakse R, et al: Effects of nonsteroidal antiinflammatory drugs on inflammatory bowel disease: a case-control study. Am J Gastroenterol 95:1949, 2000

143. Biancone L, Tosti C, Geremia A, et al: Rofecoxib and early relapse of inflammatory bowel disease: an open-label trial. Aliment Pharmacol Ther 19:755, 2004

144. Bonner GF: Exacerbation of inflammatory bowel disease associated with use of celecoxib. Am J Gastroenterol 96:1306, 2001

145. Hilsden RJ, Scott CM, Verhoef MJ: Complementary medicine use by patients with inflammatory bowel disease. Am J Gastroenterol 93:697, 1998

146. Gionchetti P, Rizzello F, Venturi A, et al: Comparison of oral with rectal mesalazine in the treatment of ulcerative proctitis. Dis Colon Rectum 41:93, 1998

147. Safdi M, DeMicco M, Sninsky C, et al: A double-blind comparison of oral versus rectal mesalamine versus combination therapy in the treatment of distal ulcerative colitis. Am J Gastroenterol 92:1867, 1997

148. Haghighi DB, Lashner BA: Left-sided ulcerative colitis. Gastroenterol Clin North Am 33:271, 2004

149. d'Albasio G, Pacini F, Camarri E, et al: Combined therapy with 5-aminosalicylic acid tablets and enemas for maintaining remission in ulcerative colitis: a randomized double-blind study. Am J Gastroenterol 92:1143, 1997

150. Bernstein CN, Leslie WD, Leboff MS: AGA technical review on osteoporosis in gastrointestinal diseases. Gastroenterology 124:795, 2003

151. Cohen RD, Stein R, Hanauer SB: Intravenous cyclosporin in ulcerative colitis: a five-year experience. Am J Gastroenterol 94:1587, 1999

152. Colombel JF, Ferrari N, Debuysere H, et al: Genotypic analysis of thiopurine S-methyltransferase in patients with Crohn's disease and severe myelosuppression during azathioprine therapy. Gastroenterology 118:1025, 2000

153. Ording Olsen K, Juul S, Berndtsson I, et al: Ulcerative colitis: female fecundity before diagnosis, during disease, and after surgery compared with a population sample. Gastroenterology 122:15, 2002

154. Larson DW, Pemberton JH: Current concepts and controversies in surgery for IBD. Gastroenterology 126:1611, 2004

155. Michelassi F, Lee J, Rubin M, et al: Long-term functional results after ileal pouch anal restorative proctocolectomy for ulcerative colitis: a prospective observational study. Ann Surg 238:433, 2003

156. Mahadevan U, Sandborn WJ: Diagnosis and management of pouchitis. Gastroenterology 124:1636, 2003

157. Gionchetti P, Rizzello F, Helwig U, et al: Prophylaxis of pouchitis onset with probiotic therapy: a double-blind, placebo-controlled trial. Gastroenterology 124:1202, 2003

158. Gionchetti P, Rizzello F, Venturi A, et al: Oral bacteriotherapy as maintenance treatment in patients with chronic pouchitis: a double-blind, placebo-controlled trial. Gastroenterology 119:305, 2000

159. Mimura T, Rizzello F, Helwig U, et al: Once daily high dose probiotic therapy (VSL#3) for maintaining remission in recurrent or refractory pouchitis. Gut 53:108, 2004

160. Tremaine WJ: Gastroduodenal Crohn's disease: medical management. Inflamm Bowel Dis 9:127, 2003

161. van Hogezand RA, Witte AM, Veenendaal RA, et al: Proximal Crohn's disease: review of the clinicopathologic features and therapy. Inflamm Bowel Dis 7:328, 2001

162. Hanauer SB, Stromberg U: Oral Pentasa in the treatment of active Crohn's disease: a meta-analysis of double-blind, placebo-controlled trials. Clin Gastroenterol Hepatol 2:379, 2004

163. Sandborn WJ, Feagan BG: Review article: mild to moderate Crohn's disease: defining the basis for a new treatment algorithm. Aliment Pharmacol Ther 18:263, 2003

164. Faubion WA Jr, Loftus EV Jr, Harmsen WS, et al: The natural history of corticosteroid therapy for inflammatory bowel disease: a population-based study. Gastroenterology 121:255, 2001

165. Brookes MJ, Green JR: Maintenance of remission in Crohn's disease: current and emerging therapeutic options. Drugs 64:1069, 2004

166. Hanauer SB, Korelitz BI, Rutgeerts P, et al: Postoperative maintenance of Crohn's disease remission with 6-mercaptopurine, mesalamine, or placebo: a 2-year trial. Gastroenterology 127:723, 2004

167. Friedman S, Regueiro MD: Pregnancy and nursing in inflammatory bowel disease. Gastroenterol Clin North Am 31:265, 2002

168. Moskovitz DN, Bodian C, Chapman ML, et al: The effect on the fetus of medications used to treat pregnant inflammatory bowel-disease patients. Am J Gastroenterol 99:656, 2004

169. Norgard B, Pedersen L, Fonager K, et al: Azathioprine, mercaptopurine and birth outcome: a population-based cohort study. Aliment Pharmacol Ther 17:827, 2003

170. Stein R, Hanauer SB: Life-threatening complications of IBD: how to handle fulminant colitis and toxic megacolon. J Crit Illness 13:518, 1998

171. Orchard T: Extraintestinal complications of inflammatory bowel disease. Curr Gastroenterol Rep 5:512, 2003

172. Orchard TR, Chua CN, Ahmad T, et al: Uveitis and erythema nodosum in inflammatory bowel disease: clinical features and the role of HLA genes. Gastroenterology 123:714, 2002

173. Ahmad J, Slivka A: Hepatobiliary disease in inflammatory bowel disease. Gastroenterol Clin North Am 31:329, 2002

174. Pardi DS, Tremaine WJ, Sandborn WJ, et al: Renal and urologic complications of inflammatory bowel disease. Am J Gastroenterol 93:504, 1998

175. Worcester EM: Stones from bowel disease. Endocrinol Metab Clin North Am 31:979, 2002

176. Wilson A, Reyes E, Ofman J: Prevalence and outcomes of anemia in inflammatory bowel disease: a systematic review of the literature. Am J Med 116(suppl 7A):44S, 2004

177. van Bodegraven AA: Haemostasis in inflammatory bowel diseases: clinical relevance. Scand J Gastroenterol Suppl (239):51, 2003

178. Regueiro M, Valentine J, Plevy S, et al: Infliximab for treatment of pyoderma gangrenosum associated with inflammatory bowel disease. Am J Gastroenterol 98:1821, 2003

179. Itzkowitz SH, Yio X: Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. Am J Physiol Gastrointest Liver Physiol 287:G7, 2004

180. Rutter M, Saunders B, Wilkinson K, et al: Severity of inflammation is a risk factor for colorectal neoplasia in ulcerative colitis. Gastroenterology 126:451, 2004

181. Itzkowitz SH: Cancer prevention in patients with inflammatory bowel disease. Gastroenterol Clin North Am 31:1133, 2002

182. Mpofu C, Watson AJ, Rhodes JM: Strategies for detecting colon cancer and/or dysplasia in patients with inflammatory bowel disease. Cochrane Database Syst Rev (2):CD000279, 2004

183. Ullman T, Croog V, Harpaz N, et al: Progression of flat low-grade dysplasia to advanced neoplasia in patients with ulcerative colitis. Gastroenterology 125:1311, 2003

184. Lichtenstein GR: Reduction of colorectal cancer risk in patients with Crohn's disease. Rev Gastroenterol Disord 2(suppl 2):S16, 2002

185. Loftus EV Jr, Tremaine WJ, Habermann TM, et al: Risk of lymphoma in inflammatory bowel disease. Am J Gastroenterol 95:2308, 2000

186. Bebb JR, Logan RP: Review article: does the use of immunosuppressive therapy in inflammatory bowel disease increase the risk of developing lymphoma? Aliment Pharmacol Ther 15:1843, 2001

187. Farrell RJ, Ang Y, Kileen P, et al: Increased incidence of non-Hodgkin's lymphoma in inflammatory bowel disease patients on immunosuppressive therapy but overall risk is low. Gut 47:514, 2000

188. Fraser AG, Orchard TR, Robinson EM, et al: Long-term risk of malignancy after treatment of inflammatory bowel disease with azathioprine. Aliment Pharmacol Ther 16:1225, 2002

189. Dayharsh GA, Loftus EV Jr, Sandborn WJ, et al: Epstein-Barr virus–positive lymphoma in patients with inflammatory bowel disease treated with azathioprine or 6-mercaptopurine. Gastroenterology 122:72, 2002

190. Loftus EV Jr, Sandborn WJ: Lymphoma risk in inflammatory bowel disease: influences of referral bias and therapy. Gastroenterology 121:1239, 2001

Acknowledgment

Figure 4 Tom Moore.

V DISEASES OF THE PANCREAS

Peter Draganov, m.d.
Chris E. Forsmark, m.d., f.a.c.p.

Definitions of Disease Presentations

acute and chronic pancreatitis

Acute pancreatitis has traditionally been defined as an acute inflammatory process of the pancreas that (1) is associated with abdominal pain and elevations in serum levels of pancreatic enzymes and (2) disrupts normal pancreatic architecture and function only until the illness resolves. Chronic pancreatitis, on the other hand, is traditionally described as associated with permanent and irreversible damage to the gland. These definitions, which were developed at a series of international meetings, have limited applicability for clinicians. For example, some patients with severe acute pancreatitis develop substantial necrosis of the gland during the acute attack and sustain permanent abnormalities of both pancreatic architecture and pancreatic function. Likewise, many patients with an acute attack of alcoholic pancreatitis have already developed histologic changes of chronic pancreatitis at the onset of symptoms. Both types of patient are at risk for the complications of acute pancreatitis and are best managed as having acute pancreatitis, although they do not fit the traditional classification schemes.

In addressing these problems, the most recent consensus conference, the International Symposium on Acute Pancreatitis,[1] defined acute pancreatitis in a more clinically useful manner—that is, as an acute inflammatory process of the pancreas with variable involvement of regional tissues and remote organ systems. Because it may not be possible at the time of the attack to determine whether permanent architectural or functional changes are present or will develop, the disease may subsequently be reclassified, on the basis of additional clinical information, as chronic pancreatitis or as an acute exacerbation of chronic pancreatitis.

The International Symposium on Acute Pancreatitis also defined severe acute pancreatitis as the presence of organ failure (e.g., shock, pulmonary insufficiency, renal failure, or gastrointestinal bleeding) or pancreatic or peripancreatic complications (e.g., necrosis, abscess, or pseudocyst), or both, along with unfavorable early prognostic signs (e.g., using the Ranson criteria or the APACHE II score) [*see* Determining Disease Severity, *below*]. Although not perfect, these clinical definitions more closely fit the approach to management.

complications of pancreatitis

As part of the consensus conference on acute pancreatitis,[1] more precise definitions were developed to describe the local and systemic complications of acute pancreatitis. An acute fluid collection was defined as a collection of fluid occurring in or around the pancreas early in the course of acute pancreatitis. This collection of fluid is composed of both pancreatic juice and inflammatory fluid; it is poorly circumscribed and lacks a visible wall of fibrosis or granulation tissue. On computed tomography, these collections are seen as low-attenuation areas without a visible capsule. They are quite common in acute pancreatitis, occurring in 30% to 50% of cases.[2] Many of these acute fluid collections resolve, but some may persist and develop a visible capsule, at which time they are termed pseudocysts.

Pseudocysts are defined as collections of fluid (pancreatic juice) surrounded by a fibrous capsule. It takes at least 4 to 6 weeks for an acute fluid collection to develop a capsule and become a pseudocyst. Pseudocysts may remain sterile or may become secondarily infected.

Pancreatic necrosis is a pathologic finding but is clinically defined on contrast-enhanced CT (CECT) as areas of pancreatic parenchyma that show no enhancement with the infusion of intravenous contrast. Acute necrotizing pancreatitis (i.e., CECT findings of pancreatic necrosis) may be subclassified as sterile necrosis or infected necrosis on the basis of the presence or absence of bacteria in an aspiration or surgical sample. Acute interstitial pancreatitis is defined by the absence of CECT findings of necrosis. Finally, pancreatic abscess is defined as a circumscribed collection of pus containing little necrotic tissue. What formerly was called infected pseudocyst is now referred to as a pancreatic abscess. The term phlegmon was omitted from the report because there was no consensus on its definition.

Acute Pancreatitis

epidemiology

Estimates of the incidence of acute pancreatitis range from about 5 to 25 cases per 100,000 population. In the United States, between 166,000 and 224,000 patients are admitted each year with a primary diagnosis of acute pancreatitis.[3] The number of patients discharged from hospitals with a diagnosis of acute pancreatitis has steadily increased over the past 20 years.[4,5] Similar trends have been seen in other developed countries.[6] The reason for the increased incidence of acute pancreatitis in the United States is unclear, but the increase may be related to the increased incidence of gallstones (one of the major causes of acute pancreatitis)[5] in association with the epidemic of obesity.[7] In large series from referral hospitals, the mortality associated with acute pancreatitis has ranged from 5% to 10%; however, this range is probably high because of referral patterns, as recent estimates using more comprehensive hospital databases have documented an overall mortality of about 2%.[8] Mortality varies with etiology, the development of complications or necrosis, and the number and severity of comorbid medical conditions.[9,10] The cost of care is substantial, with estimates of total direct and indirect costs ranging from $3.6 billion to $6 billion annually.[1,5,8]

etiology

Many factors have been implicated as causes of acute pancreatitis [*see Table 1*]. Together, gallstone disease and alcohol abuse account for 70% to 80% of all cases of acute pancreatitis.[1] The prevalence of acute pancreatitis varies from population to population, depending on the relative prevalence of alcohol abuse and gallstone disease.

Gallstone Disease

The exact mechanism by which gallstone disease causes acute pancreatitis is not completely understood. It is clear that the pas-

sage of a gallstone through the ampulla of Vater is an important initiating event for gallstone pancreatitis, most likely by the gallstone's causing transient obstruction of the pancreatic duct or by edema resulting after stone passage. The association between obesity and gallstone disease is well established,[4] but abdominal obesity may be a more specific risk factor. A large prospective study indicated that abdominal adiposity in men carries a relative risk of gallstone disease of 2.29.[11]

Alcohol Abuse

The mechanism by which alcohol consumption produces acute (and chronic) pancreatitis remains obscure. In most patients, long-standing abuse of alcohol is required, and in such patients, histologic chronic pancreatitis is usually present at the onset of a clinically apparent acute attack.[12] In a minority of patients, a large alcoholic binge is the initiating event for acute pancreatitis, and no evidence is found of preexisting chronic damage to the gland.

Obstruction of the Pancreatic Duct

A number of disorders appear to cause acute pancreatitis by a process that obstructs the pancreatic duct. The most common of these is the presence of gallstones (see above). The other conditions are relatively uncommon. One such condition is sphincter of Oddi dysfunction, in which elevations of basal pancreatic sphincter pressure (more than 40 mm Hg above duodenal baseline pressures) produce pancreatic duct obstruction and acute pancreatitis. In addition, both benign and malignant strictures of the pancreatic duct can produce acute pancreatitis, as can malignancy of the ampulla of Vater. Given this, a search for underlying pancreatic or ampullary malignancy is warranted in patients at higher risk for malignancy (e.g., those older than 40 to 45 years) with unexplained pancreatitis. Less common causes of pancreatic duct obstruction and acute pancreatitis include choledochal cysts, periampullary duodenal diverticula, and worms migrating through the ampulla (*Ascaris lumbricoides*, *Clonorchis sinensis*).

Pancreas divisum, which occurs in 5% to 7% of the population, is a rare cause of acute pancreatitis.[13,14] In this congenital condition, the fetal dorsal and ventral pancreatic buds fail to fuse, and the majority of pancreatic secretions enter the duodenum through the smaller minor papilla. In a small subset of patients with pancreas divisum, the minor papilla may be inadequate to allow free drainage of pancreatic juice, creating a blockage that may lead to acute or chronic pancreatitis.[13-15]

Drugs and Toxins

Drug-induced acute pancreatitis is a relatively rare event and is usually idiosyncratic.[16,17] The antimetabolites 6-mercaptopurine and azathioprine have the highest attack rate, causing acute pancreatitis in up to 4% of patients who take these drugs. Many additional drugs have been reported to cause acute pancreatitis, the most common being pentamidine, didanosine, sulfonamides, valproic acid, furosemide, and aminosalicylates. In addition to ethyl alcohol [*see* Alcohol Abuse, a*bove*], a number of toxins may injure the pancreas and cause acute pancreatitis; these include methyl alcohol, organophosphate insecticides, and the venom from certain Central and South American scorpions. Scorpion venom and insecticides appear to cause acute pancreatitis by hyperstimulating pancreatic secretion via a cholinergic mechanism.

Table 1 Causes of Acute Pancreatitis

Gallstones and microlithiasis
Alcohol abuse
Obstruction of pancreatic duct
Sphincter of Oddi dysfunction
Pancreas divisum with stenotic minor papilla
Ampullary or pancreatic tumors
Trauma
Post-ERCP trauma
Blunt or penetrating trauma
Toxins
Methyl alcohol
Scorpion venom
Organophosphate insecticides
Drugs
Azathioprine
6-Mercaptopurine (6-MP)
Pentamidine
Didanosine
Sulfonamides
Valproic acid
Furosemide
Aminosalicylates
Infections
Viral (mumps, rubella, coxsackie B, cytomegalovirus, HIV)
Bacterial (*Klebsiella, Escherichia coli*)
Fungal (*Candida*)
Hypertriglyceridemia
Genetic mutation
Hereditary pancreatitis
Cystic fibrosis
Surgery

ERCP—endoscopic retrograde cholangiopancreatography

Infection

A number of infections have been reported to cause acute pancreatitis; among them are a variety of viral infections such as cytomegalovirus, mumps, rubella, and coxsackie B. Patients with AIDS commonly have increased serum amylase levels in the absence of acute pancreatitis and less commonly develop acute pancreatitis secondary to opportunistic infections (e.g., cytomegalovirus, *Cryptosporidium*, or *Mycobacterium* infections) or as a side effect of a medication.[18]

Metabolic Factors

Metabolic causes of acute pancreatitis include hypertriglyceridemia and hypercalcemia. Serum triglycerides generally need to be in excess of 1,000 mg/dl to produce acute pancreatitis.[19] Serum triglyceride levels in excess of 1,000 mg/dl are most commonly seen in type V hyperlipoproteinemia and are usually associated with diabetes mellitus. Acute pancreatitis can itself raise triglyceride levels but not to this degree. The use of estrogens in postmenopausal women with underlying hypertriglyceridemia is associated with increased levels of triglyceride and the induction of pancreatitis, particularly if the fasting triglyceride level before initiating estrogen treatment is more than 750 mg/dl.[20] Hypercalcemia, usually associated with hyperparathyroidism, is a very rare metabolic cause of acute pancreatitis.

Trauma

Trauma to the pancreas or pancreatic duct may cause acute

pancreatitis. Blunt trauma to the abdomen may cause contusion, laceration, or complete transection of the gland. In most cases of major trauma affecting the pancreas, damage occurs at the mid-body of the pancreas, where the pancreas is crushed against the vertebral bodies; acute pancreatitis develops rapidly in most of these patients. Patients with less extensive injuries may experience a delayed onset of symptoms up to several months or even longer after the trauma. Iatrogenic trauma during endoscopic retrograde cholangiopancreatography (ERCP) causes acute pancreatitis in about 3% to 5% of cases, although in certain subgroups (e.g., those suspected of having sphincter of Oddi dysfunction), the risk may be as high as 20% to 25%.[21] ERCP appears to cause pancreatitis as a consequence of obstruction, inflammation, and edema of the pancreatic duct orifice and by barotrauma to the acinar cells. Ischemic injury to the pancreas may occur in the setting of many surgical procedures, because the pancreatic vasculature has very limited ability for vasodilatation. In such cases, postoperative pancreatitis is often quite severe and most commonly occurs after cardiac surgery or cardiopulmonary bypass.

Genetic Factors

A number of mutations have been described in association with acute and chronic pancreatitis. These include mutations in the cationic trypsinogen (*PRSS1*), cystic fibrosis transmembrane conductance regulator (*CFTR*), and secretory trypsin inhibitor (or serine protease inhibitor Kazal type 1 [*SPINK1*]) genes.[22] These conditions are most commonly associated with chronic pancreatitis [see Chronic Pancreatitis, Etiology, below] but in some cases may also produce acute flares. A number of studies have examined polymorphisms in cytokines involved in the inflammatory response to determine whether such polymorphisms might be predictors of the severity of acute pancreatitis. To date, these studies have been unrevealing.

Autoimmune Pancreatitis

Autoimmune pancreatitis is a benign disease characterized by irregular narrowing of the pancreatic duct, swelling of the parenchyma, and lymphoplasmacytic infiltration and fibrosis. Autoimmune pancreatitis can present clinically as an attack of acute pancreatitis. Patients with autoimmune pancreatitis may have high antinuclear antibody and serum IgG4 concentrations, providing a useful means of distinguishing this disorder from other diseases of the pancreas or biliary tract.[23,24] Patients with autoimmune pancreatitis generally have a favorable response to corticosteroid treatment.

Undetermined Causes

After evaluation, about 25% of all patients with acute pancreatitis do not have a specific definable etiology. In fact, after gallstones and alcohol, idiopathic acute pancreatitis is the most common diagnosis. Some of these patients may be surreptitious alcoholics, but many more appear to have a forme fruste of gallstone disease. Two series have documented the presence of microscopic gallstones (so-called microlithiasis) in two thirds to three fourths of patients with apparent idiopathic acute pancreatitis.[25,26] The importance of microlithiasis is underscored by the fact that cholecystectomy, ERCP with sphincterotomy, and agents used to dissolve gallstones (e.g., ursodeoxycholic acid) all reduced the frequency of recurrent attacks of acute pancreatitis in patients participating in these studies. Unfortunately, there is as yet no standardized method to determine the presence of microlithiasis.

PATHOGENESIS

The pathophysiology of acute pancreatitis, irrespective of cause, remains poorly understood. All etiologies appear to converge on a final common pathway that allows the premature activation of digestive enzymes within the pancreas.[27,28] The conversion of the inactive proenzyme trypsinogen to its active form trypsin appears to be a critical early step because trypsin can then activate most of the other digestive proenzymes. The release of activated digestive enzymes into the pancreas and surrounding tissues can produce tissue damage and necrosis of the pancreas, its surrounding fat, and adjacent structures. This chemical "burn" of the retroperitoneum leads to substantial fluid loss into this area—so-called third-space fluid losses. Not all patients with acute pancreatitis develop necrosis of the pancreas itself; necrosis is most commonly seen in severe attacks of acute pancreatitis. Substantial pancreatic necrosis (acute necrotizing pancreatitis) is usually distinguished from the milder form in which necrosis is absent (interstitial pancreatitis).

The release of activated digestive enzymes into the systemic circulation can overwhelm normal protective mechanisms (e.g., antiproteases) and cause direct damage to distant organs and other systemic enzyme systems (e.g., complement and kinin systems). Finally, a number of inflammatory mediators and cytokines can be released from inflammatory cells to produce a systemic immune response syndrome (SIRS) or sepsislike syndrome.[29,30] The combination of activated digestive enzymes in the systemic circulation, the activation of other enzyme systems, and the release of inflammatory cytokines can produce the severe systemic complications associated with severe acute pancreatitis [*see Table 2*]. Recognition of the role of these inflammatory mediators has not only improved our understanding of the pathophysiology of acute pancreatitis but also provided potential new targets for therapy.

DIAGNOSIS

Clinical Findings

The diagnosis of acute pancreatitis is usually suspected on the basis of compatible signs and symptoms and confirmed by laboratory tests and radiographic imaging. Pain is the most common symptom of acute pancreatitis, occurring in up to 95% of patients. The pain of acute pancreatitis is most commonly felt in the epigastrium and radiates to the back in up to two thirds of patients. Pain may be felt more diffusely across the abdomen. The pain is usually quite severe, reaches its maximum intensity within 30 minutes, and lasts hours to days. In some cases, pain may not be the dominant symptom, particularly if it is masked by multiorgan failure, delirium, or coma; in rare cases, pain may be absent altogether. Nausea and vomiting are commonly associated with the pain of pancreatitis. No relief of the abdominal pain is achieved by vomiting.

The physical examination usually reveals epigastric or diffuse tenderness on palpation, with rebound tenderness and guarding present in the most severe cases. The abdomen is often distended and tympanic, and bowel sounds may be decreased or absent. Vital signs may be normal; more commonly, tachycardia, hypotension, tachypnea, and low-grade fever are noted. Orthostatic hypotension, tachycardia, and shock early in the course of acute pancreatitis are markers for substantial third-space fluid losses and indicate both a poor prognosis and probable need for admission to an intensive care unit. Dyspnea or tachypnea may occur because of muscular splinting secondary to abdominal

pain, pleural effusions, or a pulmonary capillary leak syndrome (i.e., acute respiratory distress syndrome [ARDS]). Generally, the presence of tachypnea, dyspnea, or oxygen desaturation merits ICU admission. Rare physical findings include ecchymoses of the flank (Grey Turner sign) and umbilicus (Cullen sign) and eruptive xanthomas in patients with hyperlipidemic pancreatitis; signs of alcoholic liver disease may be present in patients with alcoholic pancreatitis. Altered mental status may be present and usually has multiple causes (alcohol withdrawal, hypotension, electrolyte imbalance, and hypoxemia). Jaundice may be present, either from obstruction of the bile duct by a gallstone or from extrinsic compression of the bile duct by a large peripancreatic fluid collection. Purtscher retinopathy presenting as retinal hemorrhage is a very rare complication of acute pancreatitis.

Laboratory Tests

A history and physical examination suggestive of acute pancreatitis may be seen in a wide variety of intra-abdominal diseases. Therefore, the diagnosis is usually confirmed with a combination of laboratory tests and imaging studies.

Serum amylase The serum amylase level has long been the most widely used confirmatory laboratory measurement for acute pancreatitis. At least 75% of patients with acute pancreatitis will have increased levels of serum amylase at the time of initial evaluation.[31] Levels greater than three times the upper limit of normal are highly suggestive of acute pancreatitis. Amylase is cleared by the kidney; in patients with renal failure, a higher threshold of five times the upper limit of normal should be used. Normal levels of serum amylase, however, do not rule out the presence of pancreatitis. Serum amylase levels may be normal in some patients with acute alcoholic pancreatitis and in patients with hyperlipidemic pancreatitis (marked elevations in triglyceride levels can interfere with the laboratory assay for amylase). More generally, elevated levels of serum amylase may have already returned to normal if testing is delayed until several days after the onset of symptoms. In several large series of fatal pancreatitis, 10% to 30% of patients who died of acute pancreatitis were undiagnosed before autopsy.[32] The diagnosis of acute pancreatitis is generally missed in such patients for one of two reasons: serum amylase levels are normal (or are not measured) or the presenting symptoms are atypical (e.g., coma or multiorgan failure rather than abdominal pain). The true sensitivity of the serum amylase measurement as a diagnostic test for acute pancreatitis is therefore difficult to determine.

Elevations of the serum amylase level are not specific for acute pancreatitis and may be associated with a very wide variety of nonpancreatic conditions [*see Table 3*]. Although many of these other conditions would not be mistaken for acute pancreatitis, a number of intra-abdominal conditions can produce increased serum amylase levels and mimic both the signs and the symptoms of acute pancreatitis. Such disorders include intestinal ischemia and perforation, bowel obstruction, choledocholithiasis, cholelithiasis with cholecystitis, tubo-ovarian disease (ectopic pregnancy, acute salpingitis), and acute appendicitis.

Serum lipase The serum lipase level is often used as an adjunct to or in place of serum amylase testing as a confirmatory test for acute pancreatitis. Accurate measurement of serum lipase was difficult in the past, but new methods provide high levels of precision. The lipase level is in fact slightly more sensitive and somewhat more specific for acute pancreatitis than the amylase level.[31] In addition, serum lipase stays elevated longer and can confirm a diagnosis of acute pancreatitis up to 5 to 10 days after the onset of symptoms, by which time amylase levels have generally returned to normal. Like amylase, lipase may be elevated in other intra-abdominal conditions (with the exception of tubo-ovarian disease) and may be elevated in renal failure. Elevations that are more than three times the upper limit of normal have the greatest diagnostic sensitivity and specificity, but again, this threshold may need to be increased to five times the upper limit of normal in patients with renal failure. Lipase is probably preferable to amylase as a confirmatory test because in addition to its greater specificity, it is no more costly and, in most hospitals, has equally rapid availability.

Table 2 Complications of Acute Pancreatitis

- Local complications
  - Peripancreatic fluid collection
  - Pseudocyst
  - Pancreatic necrosis (sterile or infected)
  - Abscess
  - Duodenal obstruction
  - Biliary obstruction
- Systemic complications
  - Cardiovascular
    - Hypotension and shock
    - Pericardial effusion and tamponade
    - ECG changes
  - Pulmonary
    - Hypoxia
    - Atelectasis, pneumonia
    - Pleural effusion
    - Acute respiratory distress syndrome
  - Metabolic
    - Hypocalcemia
    - Hypertriglyceridemia
    - Hyperglycemia
  - Renal
    - Oliguria and azotemia
    - Acute tubular necrosis
  - Hematologic
    - Disseminated intravascular coagulation (DIC)
    - Vascular thrombosis (particularly splenic vein)
  - Gastrointestinal bleeding
  - Other
    - Encephalopathy
    - Distant fat necrosis
    - Retinopathy

Other tests Leukocytosis is frequently present in acute pancreatitis. The hematocrit may be normal, but in patients with severe pancreatitis and substantial third-space fluid loss, hemoconcentration is present. There are a number of methods to gauge the severity of pancreatitis, but the presence of hemoconcentration is a reasonably accurate marker of severe pancreatitis.[33] Hyperglycemia and hypocalcemia may also be present. Tetany is rare because ionized calcium levels are usually normal in pancreatitis despite the presence of hypocalcemia. Liver chemistries may be elevated in persons with gallstone pancreatitis or with intrinsic liver disease (e.g., alcoholic hepatitis). Elevations of alanine aminotransferase levels to three times the normal level strongly suggest gallstone disease as the etiology; however, any significant abnormality of liver chemistries should raise the suspicion of gall-

stone pancreatitis, particularly if the abnormalities rapidly return to the normal range over the course of a few days.[34] The differentiation of gallstone pancreatitis from other forms of pancreatitis is important because specific therapy may be required [*see* Removal of Common Bile Duct Stones, *below*].

Imaging Studies

Imaging studies, particularly ultrasound and CT, can be useful in confirming a diagnosis of acute pancreatitis, determining etiology, and assessing the severity of the attack.

Radiology Plain abdominal radiographs may help in the evaluation of acute abdominal pain by documenting the presence of conditions (e.g., an ileus or free intraperitoneal air) that cause acute pain, but the findings are never specific enough to confirm a diagnosis of acute pancreatitis. Similarly, barium or water-soluble contrast studies of the upper gastrointestinal tract are not helpful in confirming a diagnosis of acute pancreatitis.

Ultrasonography Abdominal ultrasonography (US) is a highly useful test in the evaluation of suspected acute pancreatitis. Diagnostic abnormalities of the pancreas, including pancreatic enlargement, changes in echotexture, and peripancreatic fluid collections, can be seen in up to two thirds of patients; in the remaining third, overlying bowel gas limits the ability of sound transmission, thus preventing adequate visualization of the pancreas.

US is the most sensitive test for detecting stones in the gallbladder in patients with gallstone pancreatitis. The presence of gallstones or a dilated bile duct visualized on US is highly predictive of gallstone disease as the etiology of acute pancreatitis. If the gallbladder and biliary tree cannot be imaged on initial ultrasonography, a repeat ultrasound several days later may prove diagnostic of gallstone pancreatitis.

Table 3 Nonpancreatic Causes of Elevated Amylase and Lipase Levels

Amylase	*Lipase*
Biliary disease Common bile duct obstruction Acute cholecystitis	Biliary disease Common bile duct obstruction Acute cholecystitis
Intestinal ischemia, obstruction, or perforation	Intestinal ischemia, obstruction, or perforation
Acute appendicitis	Acute appendicitis
Gynecologic conditions Ectopic pregnancy Acute salpingitis Ovarian cysts and malignancies	Renal insufficiency
Renal insufficiency	
Macroamylasemia	
Salivary gland disease, including mumps	
Miscellaneous causes Anorexia nervosa Diabetic ketoacidosis Lung cancer Head trauma	

Computed tomography Computed tomography is much more accurate than US in confirming the presence of acute pancreatitis, although CT is less accurate in evaluating the biliary tree and gallbladder for stones.[35-37] The two tests are therefore often used together in patients with acute pancreatitis. CT results may be normal in a small subset of patients with very mild acute pancreatitis (10% of patients), but the test is reliably diagnostic in moderate or severe disease. CT is also quite useful in assessing conditions that mimic severe acute pancreatitis. In addition, CT plays a very important role in determining severity of the attack [*see* CT Findings, *below*].

The use of a rapid bolus of intravenous contrast coupled with rapid scanning of the pancreas by use of CECT can provide a diagnosis of acute pancreatitis and, very importantly, assess the severity of disease and the extent of pancreatic necrosis. As visualized on CECT, viable pancreatic parenchyma is enhanced by uptake of the contrast medium, and necrotic areas of the gland are unenhanced. The extent of necrosis is a very important indicator of prognosis.[35-37]

CT scans are not required in every patient with acute pancreatitis, but they should be performed in patients with a first attack of pancreatitis, with moderate or severe symptoms, with systemic complications, in whom there is a suspicion of a complication (e.g., pancreatic pseudocyst), with smoldering pancreatitis that is slow to improve, or when the diagnosis is unclear.

Magnetic resonance imaging Gadolinium-enhanced dynamic magnetic resonance imaging can be used to grade the severity of acute pancreatitis if there are contraindications to intravenous contrast–enhanced CT, such as renal failure or iodine sensitivity.[38] Furthermore, magnetic resonance cholangiopancreatography (MRCP) is an accurate way to test for the presence of common bile duct stones.[39] It may be difficult, however, to perform MRI or MRCP in very sick patients.

Endoscopy ERCP and endoscopic ultrasonography (EUS) are not used as diagnostic tests for acute pancreatitis, although they may be useful in determining etiology. ERCP is accurate in evaluating many of the less common causes of acute pancreatitis, including microlithiasis, sphincter of Oddi dysfunction, pancreas divisum, and pancreatic duct strictures (benign and malignant). As a diagnostic test, ERCP is generally reserved for patients who have experienced a second attack of unexplained pancreatitis, although use of ERCP may be considered as a diagnostic option after a single attack of unexplained pancreatitis in patients at risk for malignant pancreatic duct strictures (e.g., those older than 40 to 45 years). ERCP certainly has value as a therapeutic tool (e.g., for finding and removing common bile duct stones in patients with gallstone pancreatitis).

EUS is also useful in the documentation of gallstones, microlithiasis, pancreatic tumors, and pancreas divisum. Although it is used less frequently than ERCP in patients with acute pancreatitis (primarily because it is not widely available), EUS has a significantly lower risk of complications. EUS is highly accurate in both documenting stones and visualizing tumors (more sensitive than ERCP) and will be used more commonly in the future [*see Figure 1*].

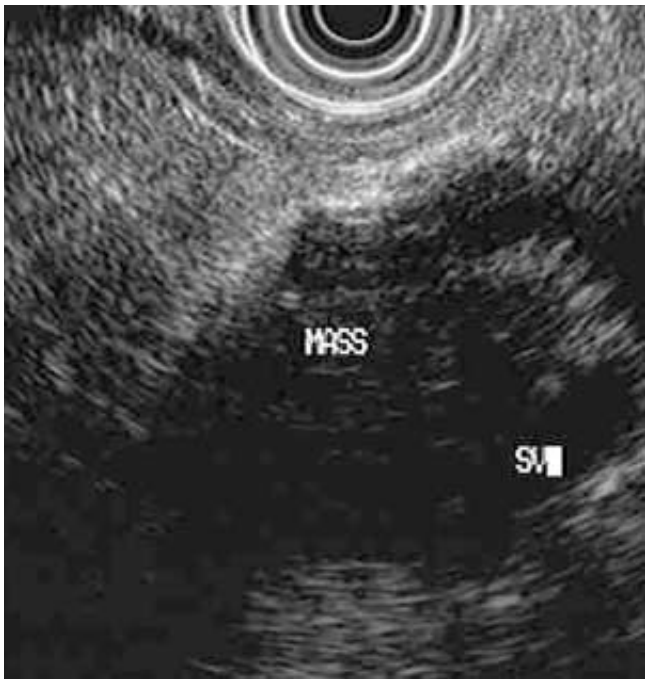


Figure 1 **An endoscopic ultrasound image of a pancreatic mass in a patient with unexplained acute pancreatitis. The mass is labeled and sits on the splenic vein (SV). The mass was not visible on computed tomography.**

Determining Disease Severity

Three quarters of patients with acute pancreatitis have a benign course and recover rapidly. The highest morbidity and mortality are in patients with necrotizing pancreatitis. The mortality associated with acute pancreatitis is 2% to 10%, with the lower estimates of mortality coming from database analyses and the higher estimates coming from more selected or referral populations. Mortality within the first week of the illness is most commonly from ARDS, multiorgan failure, or a sepsislike syndrome. Patients with persisting organ failure are at greatest risk of mortality; transient organ failure is generally not associated with mortality. Mortality after the first or second week is most commonly caused by infection of pancreatic tissue (i.e., pancreatic abscess or infected necrosis).

After making a diagnosis of acute pancreatitis, the clinician's next goal is to estimate prognosis and severity of disease. The accurate assessment of prognosis and severity allows more accurate decision making regarding ICU admission, measurement of central venous pressure or pulmonary capillary wedge pressure, and administration of prophylactic antibiotics. The assessment of severity of acute pancreatitis may be based on clinical features, laboratory tests, or imaging studies or a combination of the three. Frequently, careful evaluation by an experienced clinician is a very helpful method of gauging severity and detecting complications. The presence of delirium or coma, hypoxia, or features suggestive of massive third-space fluid loss (i.e., hypotension, tachycardia, oliguria, azotemia, or hemoconcentration) within the first 24 hours suggests a severe attack and the need for ICU admission.[40] Many patients will not develop such a dramatic illness, and a number of multiple-factor scoring systems are available to assist the clinician in determining severity and estimating prognosis.

Multiple-factor scoring systems Several multiple-factor scoring systems have been developed that use a combination of clinical and laboratory features to determine disease severity; these include the Ranson criteria [*see Table 4*], modified Glasgow criteria, and APACHE II criteria [*see Table 5*]. The most widely quoted system developed by Ranson utilizes two systems: one for gallstone pancreatitis and one for nongallstone (alcoholic) pancreatitis. In the Ranson system, the presence of one or two criteria is very specific, though not sensitive, for clinically mild pancreatitis.[40,41] The presence of three or four risk factors is associated with a mortality of 15%, although many patients with three or four criteria also recover rapidly without sequelae. The presence of a high number of criteria (e.g., six or seven Ranson criteria) is associated with a high likelihood of substantial morbidity and mortality (> 50% mortality).[40,41] All of the multiple-factor scoring systems suffer from the limitation of a high false positive rate; in other words, the presence of a moderate number of criteria is generally associated with moderate risk of morbidity and mortality, but the scoring systems do not identify whether that individual patient is likely to suffer morbidity or mortality. The APACHE II system has an advantage over the Ranson and Glasgow systems in that it can be applied at any point in the clinical illness; the Ranson and Glasgow scores require up to 48 hours before they can be calculated. Three or more Ranson criteria or eight or more APACHE II points are commonly considered an indication of an unfavorable prognosis.

CT findings CT is a useful adjunct for estimating disease severity. The initial scoring systems correlated the severity of the pancreatic and peripancreatic inflammatory processes with the prognosis. Subsequently, CT was combined with a rapidly administered bolus of intravenous contrast (CECT) to define the presence and extent of pancreatic necrosis; this technique has become widely used.[36,37,41] The lack of vascular contrast enhancement of the pancreas on CECT corresponds in a general way to the presence of necrosis. Pancreatic necrosis complicates about 25% of all cases of acute pancreatitis and is generally associated with more clinically severe disease; it is particularly associated with the development of the late complications of pancreatic abscess and infected necrosis.[37,41,42]

As with all systems used in assessing severity, CECT produces a significant number of false positive results, in that many patients who have evidence of necrosis on dynamic CT scanning have a mild clinical course.[43] Despite that, the presence of substantial pancreatic necrosis (more than one third of the gland) is a useful marker of severity because nearly all patients with a clinically severe course have necrosis and almost all cases of serious pancreatic infection occurs in this group.[42,43]

At the International Symposium on Acute Pancreatitis, an attempt was made to consolidate the various methods for determining disease severity into a unified approach.[1] The resulting system defines severe pancreatitis by a combination of clinical features (organ failure), multiple-factor scoring systems, and the presence of local pancreatic complications [*see Table 6*]. This system is useful in that it consolidates the various methods, but it does not replace frequent and experienced clinical observation.

MANAGEMENT

The treatment of acute pancreatitis has four goals: (1) provide supportive care; (2) minimize or reduce the local necrosis and the systemic inflammatory process; (3) recognize and treat complications; and (4) prevent subsequent attacks.

Supportive Care

Mild acute pancreatitis The foundations of supportive care include making the patient nil per os (NPO); providing relief from pain and nausea; replacing fluid losses; providing nutrition, if needed; and monitoring for the development of complications. This is relatively straightforward in patients with mild pancreatitis, because fluid losses are modest and complications are rare. Pain and nausea can usually be controlled by the use of moderate dosages of intravenous analgesics and antiemetics. Even in mild pancreatitis, fluid losses may be significant because of third-space fluid losses, vomiting, and insensible losses; appropriate fluid resuscitation is critical in minimizing complications. Patients can generally be fed when bowel sounds have returned and pain has resolved.

Severe acute pancreatitis When pancreatitis is severe or is predicted to be severe (based on CT or CECT findings, multiple-factor scoring systems, early evidence of significant third-space fluid losses, or early respiratory insufficiency), supportive care is more challenging and usually requires the resources of an ICU. Prompt and vigorous fluid replacement is critical in the early phases of severe acute pancreatitis and can minimize or prevent early complications, including renal failure and cardiovascular collapse.[44] The pancreas itself is prone to ischemic injury in the setting of intravascular fluid volume depletion. The pancreatic microcirculation has little capacity to respond to diminished blood supply, and intravascular volume depletion may worsen the degree of pancreatic necrosis. For all these reasons, early and vigorous fluid resuscitation is important in the management of severe pancreatitis.[40,44]

Hemoconcentration is a common and readily available marker of substantial third-space fluid losses. Measurement of central venous pressure or, if required, pulmonary capillary wedge pressure allows accurate assessment of fluid needs. Fluid needs of 5 to 10 L/day are not uncommon. Treatment with crystalloid solutions is usually appropriate, although colloid solutions (albumin or blood) may be appropriate when albumin levels are extremely low (< 2.0 mg/dl) or when the hematocrit is below 25%.

Admission to an ICU, in addition to facilitating the monitoring of fluid resuscitation, allows for intensive monitoring of respiratory and metabolic complications. Pulmonary capillary leak syndrome (i.e., ARDS) is one of the most serious complications of severe pancreatitis. Hypoxia and dyspnea are usually noted, but ARDS must be distinguished from fluid overload or congestive heart failure. This is best done with the use of a Swan-Ganz catheter. A variety of early metabolic complications (e.g., hyperglycemia, hypocalcemia, hypertriglyceridemia, and hypomagnesemia) are also most easily managed in an ICU setting.

Table 4 Ranson Prognostic Scoring System for Pancreatitis

Type	*On Admission*	*Within 48 Hours*
Nongallstone pancreatitis	Age > 55 yr WBC count > 16,000/mm^3 Glucose > 200 mg/dl LDH > 350 IU/L AST > 250 U/L	Decrease in Hct > 10 points Increase in BUN > 5 mg/dl Serum calcium < 8 mg/dl P_aO_2 < 60 mm Hg Base deficit > 4 mmol/L Fluid deficit > 6 L
Gallstone pancreatitis	Age > 70 yr WBC count > 18,000/mm^3 Glucose > 220 mg/dl LDH > 400 IU/L AST > 500 U/L	Decrease in Hct > 10 points Increase in BUN > 2 mg/dl Serum calcium < 8 mg/dl Base deficit > 5 mmol/L Fluid deficit > 4 L

AST—aspartate aminotransferase BUN—blood urea nitrogen Hct—hematocrit LDH—lactate dehydrogenase P_aO_2—arterial oxygen tension WBC—white blood cell

Nutritional support is useful for patients with severe pancreatitis and for those with milder pancreatitis who nonetheless are unable to eat for more than 5 to 7 days. The preferred route of providing exogenous nutrients has changed. For years, total parenteral nutrition (TPN) has been the standard practice. Accumulating evidence suggests that enteral feeding is comparable or superior to TPN.[45-47] Prospective, randomized trials have demonstrated that enteral feeding infused distal to the ligament of Treitz is associated with fewer complications (infection and hyperglycemia) and is cheaper than TPN.[45] Although the evidence is not definitive,[47] the accumulating data supporting this method of enteral feeding have led to a shift in the preferred method of providing nutrition to patients with acute pancreatitis. The main practical challenge in using enteral jejunal feeding is placing and maintaining position of the nasojejunal tube.

Treatment of Necrosis and Inflammation

Pancreatic rest No treatment has been proved to interrupt the inflammatory process effectively. Many early studies focused on strategies that were thought to "rest" the pancreas beyond the rest associated with maintaining the patient NPO. These have included nasogastric suction, H_2 receptor antagonists, atropine, somatostatin and its analogue octreotide, glucagon, and even fluorouracil. None of these approaches appear to have any benefit on the outcome of acute pancreatitis, although meta-analyses of somatostatin and octreotide suggest a slight trend toward benefit.[48] That is not to say that nasogastric suction is not useful if the patient has substantial nausea and vomiting or that administration of H_2 receptor antagonists does not prevent stress erosions and ulcers; however, neither of these therapies improves the overall outcome of the acute pancreatitis itself.

Protease removal or inhibition A second strategy to interrupt the inflammatory process is to remove proteases by peritoneal lavage or inhibit circulating proteases by administration of antiproteases (e.g., aprotinin or gabexate). However, neither method of protease control has been shown to be of benefit in acute pancreatitis. One potential reason for lack of efficacy is that these therapies can generally be administered only after the initiation of acute pancreatitis. In animal models, these therapies have been administered before the initiation of pancreatitis and have been shown to be nearly uniformly beneficial.[49] Pancreatitis induced by ERCP offers a unique opportunity to administer therapy in humans before the onset of acute pancreatitis. Although the data are inconclusive, meta-analyses have identified a reduction in post-ERCP pancreatitis in patients receiving the protease inhibitor gabexate or the antisecretory hormone somatostatin (but not, interestingly, its analogue octreotide).[50,51] The effect of these agents is only modest, and they are not available for clinical use in the United States. The effect of other methods, particularly the use of temporary pancreatic duct stents, appears to be far superior.

Anticytokine therapy Some studies have focused on control of the systemic immune response through the modulation of inflammatory cytokines. Because this cytokine cascade is felt to

Table 5 APACHE II Severity of Disease Classification System*

Physiologic Variable	*Physiologic Points*								
	Range								
Rectal temperature (°C)	≥ 41°	39.0°–40.9°	—	38.5°–38.9°	36.0°–38.4°	34.0°–35.9°	32.0°–31.9°	30.0°–31.9°	≤ 29.9°
Mean arterial pressure (mm Hg)	≥ 160	130–159	110–129	—	70–109	—	50–69	—	≤ 49
Heart rate (ventricular response)	≥ 180	140–179	110–139	—	70–109	—	55–69	40–54	≥ 39
Respiratory rate (nonventilated or ventilated)	≥ 50	35–49	—	25–34	12–24	10–11	6–9	—	≤ 5
A-aPo_2 (mm Hg) F_IO_2 ≥ 0.5 (record A-aPo_2) F_IO_2 < 0.5 (record only P_aO_2)	≥ 500 —	350–499 —	200–349 —	— —	< 200 Po_2 > 70	— Po_2 61–70	— —	— Po_2 55–60	— Po_2 < 55
Arterial pH	≥ 7.7	7.6–7.69	—	7.5–7.59	7.33–7.49	—	7.25–7.32	7.15–7.24	< 7.15
Serum sodium (mmol/L)	≥ 180	160–179	155–159	150–154	130–149	—	120–129	111–119	< 110
Serum potassium (mmol/L)	≥ 7.0	6.0–6.9	—	5.5–5.9	3.5–5.4	3.0–3.4	2.5–2.9	—	< 2.5
Serum creatinine (mg/dl)†	≥ 3.5	2.0–3.4	1.5–1.9	—	0.6–1.4	—	< 0.6	—	—
Hematocrit (%)	≥ 60	—	50.0–59.9	46.0–49.9	30.0–45.9	—	20.0–29.9	—	< 20
White blood cell count 1,000/mm^3	≥ 40	—	20.0–39.9	15–19.9	3.0–14.9	—	1.0–2.9	—	< 1
Serum HCO_3 (mmol/L)‡	≥ 52	41.0–51.9	—	32.0–40.9	22.0–31.9	—	18.0–21.9	14.0–17.9	< 15
Individual variable points	+4	+3	+2	+1	0	+1	+2	+3	+4

Total acute physiology score = sum of the individual variable points for all 12 variables.
*APACHE II Score = Physiologic points + Glasgow Coma points + Age points + Chronic Health points.

(*continued*)

underlie the development of multiorgan failure, cytokines are attractive targets for therapy. Platelet-activating factor (PAF) has been considered to be a major proinflammatory cytokine, and several small randomized trials using an antagonist of PAF have suggested that this agent may reduce the severity of pancreatitis if administered early in the disease course. The results of these small trials, however, have not been confirmed in a large randomized trial.[52] PAF antagonists have also been tested as therapies to prevent post-ERCP pancreatitis but have not shown significant benefit. It is likely that interfering with the cytokine cascade will require multiple agents, and further testing of these and similar therapies will clarify the role they play in the treatment of severe acute pancreatitis.

Removal of common bile duct stones A therapy that has been tested as a strategy to reduce local or systemic inflammation is the removal of common bile duct stones in patients with gallstone pancreatitis. In the vast majority of patients with gallstone pancreatitis, the offending bile duct stone has already passed into the duodenum at the onset of disease. Evaluation of the common bile duct for stones early in the clinical course of gallstone pancreatitis detects the presence of stones in up to 78% of patients. This level drops to between 3% and 33% if the evaluation is undertaken later in the clinical course.[53] The vast majority of patients thus pass the stone spontaneously; however, in a small subset of patients, a persistent common bile duct stone remains, and anecdotal observations suggest that these patients seem to be at risk for more severe pancreatitis (e.g., more organ failure and a greater degree of necrosis) and concomitant cholangitis. It is well established that cholangitis complicates up to 10% of cases of gallstone pancreatitis. Furthermore, it may be difficult in some patients to distinguish the presence of concomitant cholangitis from severe pancreatitis, because the two diseases may present similar features (e.g., fever, leukocytosis, abdominal pain, and abnormal liver chemistries). Therefore, the strategy was proposed to remove common bile duct stones in patients with gallstone pancreatitis as a means to reduce severity of disease and to prevent or treat concomitant cholangitis.

Early attempts to remove persistent common bile duct stones by surgery were associated with a mortality higher than that associated with conservative management. Subsequently, endoscopic techniques (e.g., ERCP) were used to remove stones. Three randomized trials assessed the utility of early ERCP and stone removal in patients with suspected gallstone pancreatitis.[54-56]

The initial study reported that the morbidity in patients with gallstone pancreatitis who underwent ERCP and stone removal within 72 hours was lower than the morbidity in a group of patients managed conservatively.[54] This benefit included a reduction in complications (organ failure and others) and a trend (not statistically significant) toward lower mortality. These benefits were restricted to a subgroup of patients who were predicted to have a severe attack. The second randomized trial noted a reduction in biliary sepsis but no reduction in organ failure or other complications associated with severe gallstone pancreatitis.[55] One of the two studies therefore suggested that early ERCP reduced the severity of pancreatitis, whereas the other study found that ERCP had no effect on the severity of pancreatitis but merely prevented or treated concomitant cholangitis caused by common bile duct stones. A third randomized trial attempted to reconcile these results by excluding patients with cholangitis or those at high risk for cholangitis (i.e., patients with jaundice). This study demonstrated no re-

Table 5 (*continued*)

Glasgow Coma Points

	Response	*Points*
Eyes open	Spontaneous	+4
	To voice	+3
	To pain	+2
	None	+1
Verbal response	Oriented	+5
	Confused conversation	+4
	Inappropriate words	+3
	Incomprehensible sounds	+2
	None	+1
Best motor response	Obeys commands	+6
	Localizes pain	+5
	Flexion-withdrawal to pain	+4
	Abnormal flexion (decorticate)	+3
	Abnormal extension (decerebrate)	+2
	None/flaccid	+1

Total Glasgow Coma points = 15 – Glasgow Coma score.

Age Points

Age (yr)	*Points*
< 44	0
45–54	2
55–64	3
65–74	5
≥ 75	6

Chronic Health Points

Hepatic	Biopsy-proven cirrhosis and documented portal hypertension; past episodes of upper GI bleeding attributed to portal hypertension; or prior episodes of hepatic failure, encephalopathy, or coma
Cardiovascular	New York Heart Association class IV status
Respiratory	Chronic restrictive, obstructive, or vascular disease resulting in severe exercise restriction (e.g., unable to climb stairs or perform household duties) or documented chronic hypoxia, hypercapnia, secondary polycythemia, severe pulmonary hypertension (> 40 mm Hg), or respirator dependency
Renal	Recurring long-term dialysis
Immunocompromised	The patient has received therapy that suppresses resistance to infection (e.g., immunosuppression, chemotherapy, radiation, long-term or recent high-dose steroids) or has a disease that is sufficiently advanced to suppress resistance to infection (e.g., leukemia, lymphoma, or AIDS) If the patient has a history of severe organ system insufficiency[§] or is immunocompromised,[§] assign points as follows: Nonoperative or emergency postoperative patients, 5 points Elective postoperative patients, 2 points

[†]Double point score for acute renal failure.

[‡]Venous; not preferred use if there are no arterial blood gases.

[§]Organ insufficiency or immunocompromised state must have been evident before hospital admission.

A-aPo_2—alveolar-arterial oxygen tension difference F_IO_2—fraction of inspired oxygen P_aO_2—arterial oxygen tension

duction in morbidity or mortality in patients with gallstone pancreatitis but without jaundice.[56]

Taken together, these three studies suggest that early ERCP is indicated in patients with evidence of biliary sepsis (i.e., fever, jaundice, and right upper quadrant pain) and in those with a high likelihood of developing biliary sepsis. High risk of biliary sepsis might be clinically defined as findings highly suggestive of a persistent obstructing common bile duct stone and could be defined by the presence of a stone in the common bile duct as visualized on radiographic imaging; by persistently abnormal liver chemistries; or by radiographic evidence of a persistently dilated bile duct. Early ERCP may also be considered in patients with early and progressive organ system failure, in whom it may be difficult to determine whether the downhill course is caused by severe pancreatitis or by associated cholangitis. Undertaking ERCP in this situation can be challenging, and sedating these critically ill individuals is not without risk.

Treatment of Complications

Systemic complications Systemic complications of acute pancreatitis can occur in a wide variety of organ systems [*see Table 2*].[44,57] Systemic complications, particularly shock, ARDS, and multiorgan failure, are the most common causes of death from acute pancreatitis within the first week of the illness. In patients with severe pancreatitis, fluid losses into the retroperitoneum can be massive and can produce intravascular volume depletion, hypotension and shock, and renal failure. The development of renal failure, shock, or massive volume depletion is an indication of severe disease and is associated with increased mortality. Most substantial fluid losses occur early in the course of acute pancreatitis; hence, attention must be paid to adequate and aggressive fluid resuscitation early in the disease course.[44]

Hypoxia Hypoxia is not uncommon during the initial stages of acute pancreatitis. A subset of patients will develop more substantial or prolonged hypoxia and go on to develop ARDS. Thus, it is reasonable to monitor patients with acute pancreatitis, especially those with severe acute pancreatitis, by means of pulse oximetry to detect the development of hypox-

Table 6 Atlanta Criteria for Severity

Organ failure
- Shock (supine systolic blood pressure < 90 mm Hg)
- Pulmonary insufficiency (P_aO_2 < 60 mm Hg)
- Renal failure (serum creatinine > 2 mg/dl)
- Gastrointestinal tract bleeding (> 500 ml in 24 hr)

and/or

Local complications
- Necrosis
- Abscess
- Pseudocyts

and/or

Unfavorable prognostic signs
- Ranson score > 3
- APACHE II score > 8

emia. Because patients who develop hypoxia are also at risk for MOF, persistent hypoxemia merits ICU admission. Fluid management in these patients can be difficult and is best done with monitoring of pulmonary capillary wedge pressure. Mechanical ventilation, usually with positive end-expiratory pressure (PEEP), is often necessary.

Cardiac complications A variety of cardiac complications may occur in severe acute pancreatitis, including congestive heart failure, myocardial infarction, cardiac arrhythmias, and cardiogenic shock. Hypotension is most commonly caused by third-space fluid losses and intravascular volume depletion. Cardiac dysfunction may, however, occur as part of SIRS seen in severe acute pancreatitis, which is characterized by high cardiac output and low systemic vascular resistance. Hypotension that is not responsive to fluid resuscitation may require the use of pressor agents.

Metabolic complications A number of metabolic complications may also occur in acute pancreatitis, including hypocalcemia, hyperglycemia, and hyperlipidemia. Hypocalcemia is most commonly the result of hypoalbuminemia and is uncommonly associated with a reduction in ionized calcium or symptoms of hypocalcemia. Calcium replacement is usually not needed in the absence of decreased ionized calcium or signs of neuromuscular instability (e.g., tetany, the Chvostek sign, and the Trousseau sign). Calcium should, nonetheless, be monitored carefully, as it can be a marker of severe pancreatitis.

Hyperglycemia, like hypocalcemia, is one of the Ranson criteria indicating a poor prognosis. Treatment of mild hyperglycemia is not necessary, but significant increases in blood glucose levels (i.e., levels > 200 mg/dl) should be treated with sliding-scale insulin to minimize associated fluid losses caused by glycosuria and to prevent any detrimental effect on white cell function.

Hyperlipidemia is associated with acute pancreatitis, both as an etiologic factor and as a consequence. Many patients with acute pancreatitis may develop a modest elevation in serum triglyceride levels (i.e., levels > 300 to 400 mg/dl) as a consequence of acute pancreatitis. These elevations in triglyceride levels are usually short-lived and do not require therapy. Levels above 1,000 mg/dl indicate hypertriglyceridemia as the cause, rather than a consequence, of acute pancreatitis. These levels will usually drop rapidly while the patient is NPO. Marked elevations in triglyceride levels (> 10,000 mg/dl) or failure of triglyceride levels to drop as expected may occasionally necessitate the use of plasmapheresis to rapidly clear triglycerides from the serum. After recovery from pancreatitis, patients with hyperlipidemic pancreatitis should be started on appropriate medications and dietary therapy to control lipids.

Gastrointestinal bleeding Gastrointestinal bleeding may complicate acute pancreatitis and is a marker of a severe attack.[57,58] Bleeding may occur from stress erosions, peptic ulceration, pseudoaneurysm, or varices developing as a consequence of splenic vein thrombosis. Splenic vein thrombosis, which may occur as a consequence of inflammatory and neoplastic pancreatic diseases, causes a left-sided portal hypertension characterized by gastric varices out of proportion to esophageal varices. These varices may bleed and, if they do, may be managed by splenectomy, which is curative. Bleeding from a pseudoaneurysm is usually associated with a pseudocyst [*see* Chronic Pancreatitis, Treatment of Other Complications, *below*].

Other systemic complications Rare complications of acute pancreatitis include other vascular thromboses, disseminated intravascular coagulation, distant fat necrosis in the skin (resembling erythema nodosum), encephalopathy, and sudden blindness.

Pancreatic infection and abscess Infected pancreatic necrosis, the most serious form of pancreatic infection, occurs in about 1% to 4% of patients overall with pancreatic infection and in 15% to 30% of those with pancreatic necrosis.[42,59,60] The mortality of necrotizing pancreatitis is about 10%, but this rate triples when infection supervenes. Most commonly, infection occurs during the second and third weeks of an attack of severe pancreatitis. The patient develops fever (often, temperatures > 102º F [38.9º C]), leukocytosis, and recurrent or worsening abdominal pain. The infecting organisms usually seed the necrotic area from the gut and are most commonly gram-negative rods (e.g., *Klebsiella* or *Escherichia coli*) and *Staphylococcus aureus*. *Candida*, *Enterococcus*, and anaerobic organisms are seen less commonly as causal agents of pancreatic infection. When infection is suspected, a CT scan should be obtained to evaluate the extent of necrosis and identify optimal locations for percutaneous sampling.[40,42,61] The finding of gas within the pancreatic parenchyma is highly specific but quite insensitive for serious pancreatic infection [*see Figure 2*]. However, a clinical suspicion of infection together with a finding of necrosis is an indication for percutaneous aspiration of suspicious areas, and a Gram stain and a culture of the collected tissue should be done. Experience over the past 15 years has demonstrated that percutaneous aspiration is both highly accurate and safe.[40,42,59,60] If pancreatic infection is demonstrated on percutaneous aspiration, the therapy of choice is prompt surgical debridement. Successful management of necrotic collections by percutaneous or endoscopic catheterization has also been reported,[62] but these techniques require further study because it is difficult to remove the necrotic tissue through catheters and, hence, difficult to cure the infection by this means.

Pancreatic abscess may also complicate severe acute pancreatitis, usually as a consequence of superinfection of a preexisting fluid collection or pseudocyst. Less commonly, an abscess develops secondary to superinfection of necrotic sites in the pancreas. The clinical presentation of pancreatic abscess is indistinguishable from that of infected necrosis, although therapy may differ.

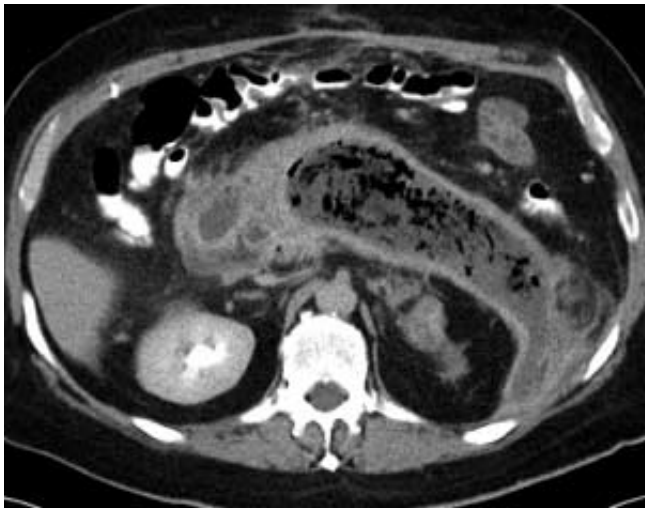


Figure 2 **A CT scan demonstrating a large amount of air within the necrotic pancreas in a patient with infected pancreatic necrosis.**

Therapy for pancreatic abscess usually consists of antibiotics and drainage. Unlike infected necrosis, which is typically treated by surgical debridement, pancreatic abscess may allow treatment by percutaneous or endoscopic tube drainage, because typically there is little solid necrotic tissue in the abscess cavity. The decision whether to treat with tube drainage or surgical exploration depends on whether the collections contain solid necrotic tissue (indicating infected necrosis, which typically necessitates open surgical drainage) or primarily pus (indicating abscess, which may allow drainage by endoscopic or percutaneous tube). However, distinguishing infected necrosis from pus may be difficult, and the most accurate picture of the collection contents is probably provided by MRI and EUS. EUS in particular is valuable because the collection not only can be assessed for the character of its contents but also can be drained safely in the same setting if appropriate. If doubt exists, it is better to err on the side of caution and opt for open surgical drainage.

Because the consequences of abscess and infected pancreatic necrosis are often severe, numerous studies have been directed toward identifying effective measures to prevent these infections. Currently, however, the role of antibiotics in the prevention of pancreatic infection is controversial. Early studies using prophylactic ampicillin demonstrated no reduction in pancreatic infections; it was later demonstrated that ampicillin does not penetrate the necrotic pancreas at adequate concentrations. More recent studies have identified agents that have adequate penetration of the necrotic pancreas; effective therapies include imipenem,[63,64] cefuroxime,[65] ofloxacin and metronidazole,[66] and ciprofloxacin and metronidazole.[67] A Cochrane Database review concluded that intravenous prophylactic antibiotic therapy for 10 to 14 days reduced the risk of superinfection of pancreatic necrosis and mortality.[68] This analysis, however, did not include the only relevant double-blind, randomized trial,[67] which demonstrated that there was no benefit with the use of prophylactic antibiotics for necrotic pancreatitis. These divergent findings have led to a difference of opinion on the overall utility of prophylactic antibiotics.

Generally, antibiotic use should be limited to patients at risk for serious pancreatic infection (i.e., if more than 30% of the gland appears necrotic on CT), and the selected antibiotic should be one that penetrates the necrotic tissue (i.e., imipenem or a fluoroquinolone plus metronidazole). Treatment should be continued for a maximum of 10 to 14 days. Patients receiving these broad-spectrum antibiotics are at risk for infection with resistant bacteria and fungal superinfection. Some researchers have advocated the concomitant use of antifungal agents such as fluconazole to reduce the risk of fungal infection, but the effectiveness of this strategy is unproved.

Fluid collections and pseudocysts Fluid collections in and around the pancreas occur commonly in patients with acute pancreatitis. Peripancreatic fluid collections are generally amorphous and not encapsulated. Most fluid collections resolve spontaneously, but some develop into pseudocysts.[57,69,70] A pseudocyst is a rounded collection of pancreatic fluid enclosed by a wall of fibrous or granulation tissue that can usually be seen on a CT scan. Most collections occur in the lesser sac or pararenal spaces, but they may develop anywhere and may even penetrate adjacent solid organs (e.g., the liver or spleen). Pseudocysts complicate 1% to 8% of all cases of acute pancreatitis.[69,70] Pseudocysts may persist, resolve, be asymptomatic, or be associated with symptoms or complications. The most common symptom associated with a pseudocyst is abdominal pain; however, other symptoms may develop if the pseudocyst obstructs an adjacent hollow viscus. For example, obstruction of the duodenum causes nausea and vomiting, whereas obstruction of the bile duct causes jaundice. Up to two thirds of all pseudocysts ultimately resolve, but spontaneous resolution is unlikely to occur with pseudocysts that are larger than 5 to 6 cm or are present for more than 6 weeks. In addition, pseudocysts larger than 6 cm are somewhat more likely to produce complications, including obstruction of an adjacent hollow viscus (e.g., duodenum or bile duct), infection, bleeding, or rupture. Infection of a pseudocyst (i.e., a pancreatic abscess) is usually relatively easy to manage with antibiotics and endoscopic or percutaneous catheter drainage; however, pseudocysts characterized by bleeding and rupture are associated with much greater morbidity and mortality. Bleeding may occur from a large artery that has formed a pseudoaneurysm from the pressure exerted by a contiguous pseudocyst, and the resulting blood flow can reach the gut through the pancreatic duct or can enter the peritoneum through a rupture of the pseudocyst. An initial bleed may be self-limited, but any unexplained drop in hemoglobin or change in pain pattern in a patient with a pseudocyst is an indication for an emergency CT scan. If any evidence of bleeding is found, emergency angiography with embolization can be lifesaving.[58]

Asymptomatic pseudocysts generally pose little risk of complications, even if they are large.[70] Symptomatic or complicated pseudocysts require therapy, and emergency surgery is required when bleeding or rupture is detected. Otherwise, elective surgical, percutaneous, or endoscopic techniques can be successful, depending on the location of the pseudocyst and the availability of expertise in these modalities. In the past, endoscopic drainage could be applied only to pseudocysts that produced a visible bulging impression in the lumen of the stomach or duodenum. Today, endoscopic pseudocyst drainage can be accomplished by using real-time endoscopic ultrasound guidance without the need of visually observing a bulge.[71] Percutaneous tube drainage can also treat pseudocysts that are farther from the gut lumen; however, tube drainage can produce a chronically draining external pancreatic fistula. Surgical treatment of pseudocysts probably has the best long-term results, but it also carries the most significant morbidity. Endoscopic ultrasound-guided transmural pseudocyst drainage appears to offer the best risk-to-benefit ratio for pseudocysts in anatomically amenable locations, but studies directly comparing surgical, endoscopic, and percutaneous drainage are lacking.

Prevention of Subsequent Attacks

Preventing subsequent attacks of acute pancreatitis requires elimination of the cause of the disease. In patients with acute alcoholic pancreatitis, cessation of alcohol consumption appears to have some benefit in reducing relapse, although unfortunately, the disease may continue to progress to symptomatic chronic pancreatitis despite abstinence. In patients with gallstone pancreatitis, cholecystectomy virtually eliminates recurrence. Similarly, the detection of microlithiasis followed by appropriate therapy (i.e., cholecystectomy, endoscopic biliary sphincterotomy, and possibly the use of ursodeoxycholic acid) can prevent recurrent pancreatitis. Aggressive control of serum lipid levels can prevent recurrent attacks of hyperlipidemic pancreatitis. In patients with a disorder that obstructs the pancreatic duct (e.g., benign or malignant pancreatic duct stricture, pancreas divisum, sphincter of Oddi dysfunction, and ampullary tumor), removal

of the obstruction by surgical or endoscopic means is generally effective in preventing relapse.

Chronic Pancreatitis

Chronic pancreatitis is characterized by irreversible damage to the pancreas and the development of histologic evidence of fibrosis and destruction of exocrine (acinar cell) and endocrine (islets of Langerhans) tissue. As with acute pancreatitis, the definition of chronic pancreatitis was developed at international symposia.[72] Several variants of chronic pancreatitis were defined, including chronic calcified pancreatitis (the most common form, which is commonly caused by excessive alcohol intake), chronic obstructive pancreatitis (caused by long-standing pancreatic duct obstruction), and chronic inflammatory pancreatitis (associated with inflammatory and, particularly, autoimmune diseases). Unfortunately, it is often not possible to make the distinction between these variants on the basis of clinical findings or imaging studies, and the distinctions are not very useful to clinicians. Recent classification schemes have focused more on etiology and are somewhat more useful to clinicians.[22]

EPIDEMIOLOGY

The prevalence of chronic pancreatitis varies with the population. Estimates of annual incidence in several studies range from three to nine cases per 100,000 population.[5,73] One study estimated an overall prevalence of 27.4 per 100,00 population. In nonfederal hospitals in the United States, this accounts for 122,000 outpatient visits and more than 20,000 hospitalizations annually.[3,4] The natural history can be quite variable and is clearly affected by the presence of ongoing alcoholism in persons with chronic alcoholic pancreatitis. In one large multicenter study,[74] the standardized mortality ratio was 3.6 (those with a diagnosis of chronic pancreatitis died at 3.6 times the rate of age-matched control subjects). Older persons and those with alcoholic chronic pancreatitis have the lowest survival. Overall, 10-year survival for patients with chronic pancreatitis has been shown to be about 70%, and 20-year survival about 45%.[5,74] Chronic pancreatitis is a strong risk factor for pancreatic adenocarcinoma, which partly explains the increased mortality associated with chronic pancreatitis.

ETIOLOGY

Alcohol Abuse

Alcohol is the cause of chronic pancreatitis in 70% to 90% of all cases. In general, at least 5 years of alcohol intake exceeding 150 g/day is required to develop symptomatic chronic pancreatitis, although some patients develop chronic pancreatitis with less alcohol intake. Only 5% to 15% of heavy drinkers ultimately develop chronic pancreatitis, suggesting that cofactors play an important role in pathogenesis.[12] Predisposing factors may include genetic abnormalities and a diet high in fat and protein. The mechanism by which alcohol causes chronic pancreatitis remains undefined, but it may be related to a change in pancreatic secretion leading to (1) protein plug formation in the pancreatic duct, (2) a direct toxic effect of alcohol or its metabolites, or (3) repeated attacks of acute alcoholic pancreatitis that eventually produce chronic irreversible damage. It has been observed that the vast majority of patients who develop an acute attack of alcoholic pancreatitis already have preexisting chronic damage to the gland.

Tropical Pancreatitis

Tropical pancreatitis is seen in certain areas of Indonesia, India, and Africa. The disease typically presents in childhood, with diabetes, abdominal pain, steatorrhea, malnutrition, and diffuse pancreatic calcifications. Malnutrition appears to be an important cofactor in this disease, as may be the presence of toxic metabolites of the dietary staple cassava. Studies also suggest a strong genetic component, with mutations in the *SPINK1* gene occurring in more than one third of patients.[75]

Genetic Factors

Hereditary pancreatitis is an autosomal dominant disease that typically presents in childhood or early adulthood and frequently is accompanied by steatorrhea, diabetes mellitus, and diffuse pancreatic calcifications. Pain and acute episodes of pancreatitis flares may also occur but are somewhat less common in hereditary pancreatitis than in alcoholic chronic pancreatitis. The initially identified genetic abnormality is a defect in the *PRSS1* gene on chromosome 7.[22,76] Multiple mutations have been described, but two are more common.[76] The two more common mutations appear to produce a trypsinogen that, once activated, is difficult or impossible to inactivate. The activated enzyme, trypsin, can in turn activate all the other pancreatic enzymes. Chronic pancreatitis appears to be caused in this situation by prolonged low-grade pancreatic injury from the activated proteases. Pancreatic adenocarcinoma frequently complicates the condition; patients with chronic pancreatitis have a 30% risk of developing pancreatic adenocarcinoma by age 70.[77] The risk may be substantially higher in patients with paternal inheritance.

Mutation of the *SPINK1* gene increases the propensity to develop chronic pancreatitis,[22,75] although mutation of this gene appears to act as a cofactor. Increased frequency of mutations in *SPINK1* is seen in patients with pancreatitis of different etiologies—namely, tropical pancreatitis (> 33% of patients), alcoholic chronic pancreatitis (about 6% of patients), hereditary pancreatitis (in a few kindreds in addition to their *PRSS1* mutation), and idiopathic chronic pancreatitis (sometimes in association with additional mutations in the *CFTR* gene). Mutations in *SPINK1* are common in the general population, but in most cases, pancreatic disease does not occur in these individuals; therefore, these mutations provide only a predisposition to chronic pancreatitis and are only some of the many factors contributing to formation of the disease.

CFTR mutations are also associated with chronic pancreatitis. Patients with classic cystic fibrosis commonly develop pancreatic insufficiency that requires supplementation of pancreatic enzyme. Several studies have also suggested that less common cystic fibrosis gene mutations, particularly when they occur as a mixed heterozygote (different mutations on the two alleles), are associated with relapsing pancreatitis and chronic pancreatitis in the absence of obvious sinopulmonary disease.[22,78] Patients with both *CFTR* and *SPINK1* mutations are at exceedingly high risk for chronic pancreatitis.

Undetermined Causes

Some patients may be misdiagnosed as having idiopathic pancreatitis if appropriate genetic studies are not performed or if a careful history of alcohol use is not obtained. Even if genetic studies are done, not all mutations may be identified (e.g., many commercially available screens look for only a few hundred of the more than 1,200 known *CFTR* mutations), and many identified *CFTR* and *SPINK1* mutations produce only a predisposition

to disease (unlike *PRSS1* mutations, which cause disease). Previous studies of patients with so-called idiopathic chronic pancreatitis that did not assess for the presence of genetic mutations appeared to identify two forms: early-onset and late-onset idiopathic chronic pancreatitis. The early-onset form presents just before or during the second decade, and it is typically associated with severe pain in the absence of diabetes, steatorrhea, or pancreatic calcification.[79] The late-onset form presents at a mean age of 56 years and is more commonly associated with exocrine or endocrine insufficiency and less commonly associated with severe pain.[79] The relative role of genetic influences on these two phenotypic variations remains to be determined.

PATHOGENESIS

The pathophysiology of chronic pancreatitis remains poorly understood. The events that occur in hereditary chronic pancreatitis suggest that one common underlying theme may be multiple subclinical episodes of acute injury that ultimately produce chronic pancreatitis.[80] A recent hypothesis (the so-called sentinel acute pancreatitis event [SAPE] hypothesis)[81] suggests that in a patient with an underlying susceptibility (e.g., genetic background), a sentinel event (e.g., alcohol exposure) can trigger the disease process, producing acute inflammation and infiltration of inflammatory cells. The acute pancreatitis may heal or, with repeated episodes, may lead to activation of pancreatic stellate cells and the development of fibrosis (i.e., chronic pancreatitis). This hypothesis, while attractive, is as yet unproved.

DIAGNOSIS

Clinical Findings

The diagnosis of chronic pancreatitis is suspected on the basis of suggestive signs and symptoms and confirmed by further tests of pancreatic structure or function. The disease is usually suspected on the basis of the presence of abdominal pain.

The vast majority of patients with chronic pancreatitis will experience pain at some point during their illness.[73,79] The pain tends to be episodic initially, but it may become more constant or continuous as the disease progresses. During acute attacks, the patient may be thought to have acute pancreatitis until the diagnosis of chronic pancreatitis can ultimately be established. Although there is no pathognomonic character of the pain, it is most commonly felt in the epigastrium, with radiation to the back. In severe episodes, nausea and vomiting are common. The natural history of the pain is quite variable, and it may worsen, stabilize, or even resolve over time. In some patients, the onset is gradual and evolves into constant abdominal pain. However, a minority of patients with chronic pancreatitis have no pain. In these patients, the disease may be suspected on the basis of the development of exocrine insufficiency (steatorrhea, weight loss, and malnutrition) or endocrine insufficiency (diabetes mellitus).

Laboratory Tests

The clinical features suggestive of chronic pancreatitis (e.g., abdominal pain, steatorrhea, weight loss, and malnutrition) are not specific for chronic pancreatitis; the diagnosis requires confirmatory tests. Diagnostic tests are usually separated into tests that detect abnormalities of pancreatic function and tests that detect abnormalities of pancreatic structure [*see Table 7*]. Chronic pancreatitis is a slowly progressive disease, and the abnormalities of pancreatic structure or function may take years to develop or may not develop at all. Hence, all of the diagnostic tests are most accurate in far-advanced disease, when obvious structural or functional abnormalities have developed.

Table 7 Diagnostic Tests for Chronic Pancreatitis*

Structural Tests	*Functional Tests*
Endoscopic ultrasonography	Direct hormonal stimulation test (secretin or secretin-CCK test)
Endoscopic retrograde pancreatography	Fecal elastase
Computed tomography	Serum trypsin
Magnetic resonance imaging/magnetic resonance cholangiopancreatography	Fecal fat
Abdominal ultrasound	Serum glucose
Plain abdominal radiograph	—

*Ranked in approximate order of decreasing sensitivity.
CCK—cholecystokinin

Structural abnormalities that can be diagnostic of chronic pancreatitis include changes in the main pancreatic duct (dilatation, strictures, irregularity, and pancreatic duct stones), side branches of the pancreatic duct (dilatation and irregularity), or pancreatic parenchyma (diffuse pancreatic calcifications). These findings can be visualized utilizing the diagnostic tests that evaluate pancreatic structure [*see Table 7*].

Functional abnormalities in chronic pancreatitis include a decrease in stimulated secretory capacity, exocrine insufficiency (malabsorption and steatorrhea), and endocrine insufficiency (diabetes mellitus).[82] Patients with alcoholic chronic pancreatitis, hereditary chronic pancreatitis, tropical pancreatitis, and late-onset idiopathic chronic pancreatitis are most likely to develop these abnormalities, although the course of development may take many years. Patients with early-onset idiopathic chronic pancreatitis may not develop these abnormalities at all. This observation has led to a general classification of chronic pancreatitis as either big-duct or small-duct disease [*see* Imaging Studies, Disease Classification, *below*].

Serum tests Serum amylase or lipase levels may be elevated during acute exacerbations of chronic pancreatitis, but these elevations are usually only modest and are neither routinely present nor diagnostic for chronic pancreatitis. Serum trypsinogen (often called serum trypsin) can also be measured. Low levels of serum trypsinogen (< 20 ng/ml) are highly specific for chronic pancreatitis,[73,83] but such low levels occur only in advanced disease (in the presence of steatorrhea). Very low levels of serum trypsinogen may also be seen occasionally in patients with pancreatic adenocarcinoma. Serum trypsinogen levels are in the normal range in most patients with less advanced chronic pancreatitis.

Stool tests A 72-hour stool collection for fat is the gold standard to detect steatorrhea but is cumbersome and unpleasant to perform. Steatorrhea is seen only in far-advanced chronic pancreatitis. More than 7 g of fat in the stool per 24 hours is considered abnormal. Of note, the patient has to be placed on a diet containing 100 g of fat a day for the results of the stool collection to be valid. At least 90% of the pancreatic enzyme secretory capacity needs to be lost before steatorrhea will develop.[84] Qualitative stool stains for fat (e.g., Sudan III) are far less accurate than a

72-hour collection but are easily performed. They should also be performed only while the patient is on a high-fat diet. Fecal levels of elastase and chymotrypsin may be reduced in patients with chronic pancreatitis, but only in cases of more advanced chronic pancreatitis.[82] Measurement of fecal elastase in a random stool sample is of reasonable accuracy[83,85] in these patients and is now available from reference laboratories in the United States. Values of fecal elastase of less than 200 μg/g of stool are seen in patients with more advanced chronic pancreatitis.

Direct pancreatic function tests The direct pancreatic function tests involve placing a tube into the duodenum to collect pancreatic juice and are complex and cumbersome. These tests directly measure pancreatic output of enzymes or bicarbonate after stimulation with a secretagogue (e.g., secretin or cholecystokinin or its analogue). Although these tests are able to detect severe decreases in pancreatic secretory output, their strength is in detecting moderate decreases in maximal stimulated secretory capacity. This decrease in maximal stimulated secretory capacity occurs before secretory failure (exocrine insufficiency) in chronic pancreatitis, and the direct function tests are felt to be the most sensitive tests available to detect chronic pancreatitis at an early stage (earlier than any other test).[73,82,86] They are particularly useful in making the diagnosis in patients with small-duct chronic pancreatitis, in whom alternative diagnostic tests (e.g., CT, ERCP) are likely to miss the diagnosis. Unfortunately, direct pancreatic function tests are available only at a few referral centers in the United States. Alternatives to cumbersome traditional pancreatic function testing have been studied. The collection of pure pancreatic juice for 15 minutes at the time of ERCP—the so-called intraductal secretin test—proved to be an inaccurate way to evaluate pancreatic function.[87] Another alternative involves administering a secretagogue (secretin) and collecting pancreatic secretions through an endoscope during upper endoscopy with sedation. Although accurate, the 1-hour collection of pancreatic secretions after secretin stimulation at the time of upper endoscopy appears to be perhaps too impractical for widespread application.[88] It is hoped that refinements will eventually allow these more sensitive tests to be used in a wider population of patients.

Imaging Studies

Radiology Simple plain abdominal radiographs may detect diffuse pancreatic calcification in very far advanced chronic pancreatitis [*see Figure 3*]. This finding is highly specific but quite insensitive.

Ultrasonography Abdominal US is most likely to detect advanced abnormalities of pancreatic structure; however, US is diagnostic in only 60% of patients.[73] The pancreas is often not well visualized on transabdominal US. New techniques of contrast-enhanced US and tissue harmonic imaging may provide better diagnostic accuracy.[89]

Computed tomography CT is much more sensitive than US (CT, 75% to 90%) because of its capacity to detect more focal abnormalities, such as calcification, a dilated pancreatic duct, fluid collections, and focal enlargements. CT may also demonstrate gland atrophy, which is seen in patients of advanced age in the absence of chronic pancreatitis. The use of multislice CT produces images of exceptional quality; this imaging technique should improve diagnostic accuracy, although it has not been adequately studied.[90] Like US, CT can be falsely negative in early or less advanced chronic pancreatitis.

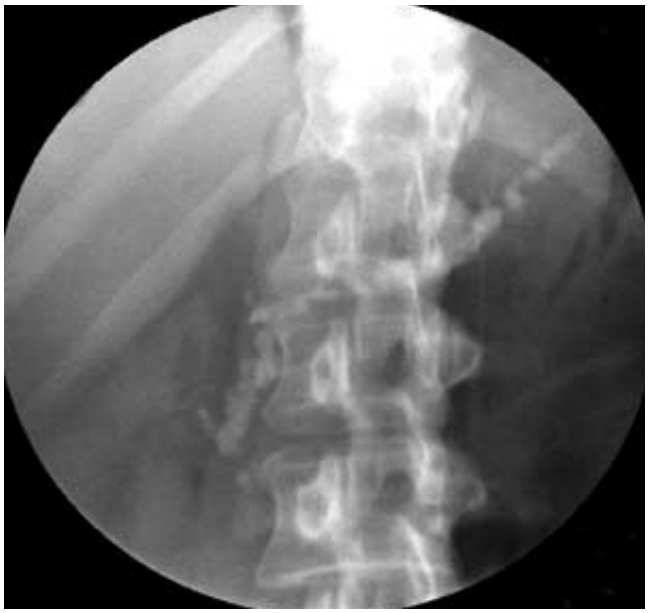


Figure 3 **A plain film of the abdomen demonstrating multiple calcified stones in the pancreatic duct in a patient with advanced chronic pancreatitis.**

Magnetic resonance imaging An improvement in magnetic resonance technology has allowed more accurate imaging of both the pancreatic parenchyma and the pancreatic duct with MRCP. It is not clear whether MRI and MRCP are superior to CT, but they do appear to be at least equivalent to CT in overall accuracy.[90] The use of secretin before MRCP allows improved imaging of the pancreatic duct and, theoretically, calculation of pancreatic secretory volume.[91] Secretin-stimulated MRCP is being used at a number of centers and may improve overall accuracy of MRI for chronic pancreatitis.

Endoscopy Two endoscopic tests are used to diagnose chronic pancreatitis: ERCP and EUS. ERCP has a reported sensitivity of 75% to 90%.[73,86] With ERCP, radiographic contrast is injected into the pancreatic duct. Changes in the pancreatic duct consistent with chronic pancreatitis include ductal dilatation, strictures, irregularity, and filling defects (stones) in the pancreatic duct [*see Figures 4 and 5*]. The changes associated with chronic pancreatitis that are seen on ERCP are not specific, as they can also be seen in other clinical presentations—namely, (1) in elderly patients with pancreatic duct dilatation caused by aging, (2) in patients with resolving acute pancreatitis, (3) in some patients with pancreatic carcinoma, and (4) in patients who have previously undergone pancreatic duct stenting.[86] In addition to its diagnostic ability, ERCP may have therapeutic application in a subset of patients [*see* Management, Endoscopic and Surgical Therapy, *below*].

EUS, which allows a highly detailed examination of the pancreatic parenchyma and pancreatic duct, routinely detects abnormalities in patients with chronic pancreatitis (high sensitivity). The test is interpreted on the basis of documented changes in both the pancreatic duct and pancreatic parenchyma; a system of

grading EUS findings usually assesses nine specific features.[92,93] In addition, a normal EUS examination essentially rules out chronic pancreatitis. The specificity of the test requires some further study in that many patients without clinical chronic pancreatitis may have modest numbers of abnormalities on EUS [*see Figure 6*].[92,93]

Disease classification Depending on the findings on imaging studies, patients may be classified as having so-called big-duct or small-duct chronic pancreatitis. This distinction has both diagnostic and therapeutic implications. Big-duct disease implies substantial abnormalities of the pancreatic duct (generally, dilatation visible on US, CT, ERCP, or EUS, often with pancreatic calcifications), and small-duct disease implies the absence of these findings (e.g., a normal or near-normal US, CT, or ERCP) [*see Figures 4 and 5*]. The diagnosis of big-duct disease is much simpler; the disease usually results from alcohol abuse, and the therapeutic options include treatments aimed at decompressing the dilated pancreatic duct. The diagnosis and management of small-duct disease may prove to be more difficult than those of big-duct disease because with small-duct disease, imaging studies may be normal, disease is more commonly idiopathic, and treatment options focus on medical therapy rather than surgical or endoscopic attempts to decompress the pancreatic duct.

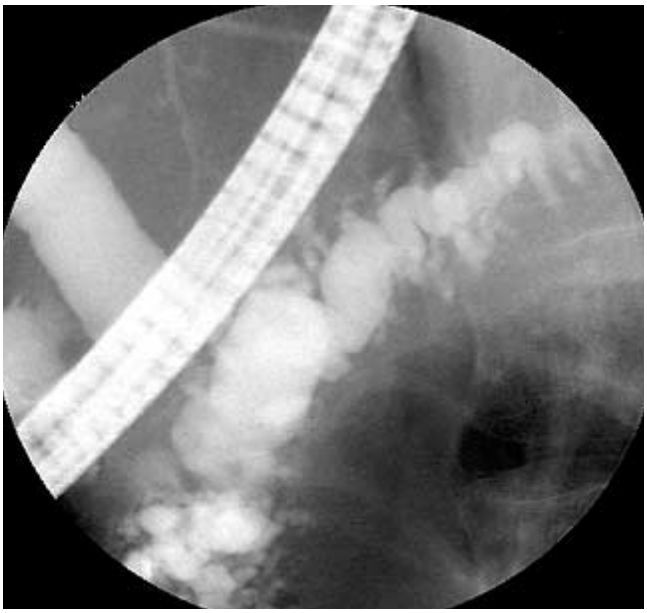


Figure 4 **An endoscopic retrograde cholangiopancreatography image demonstrating massive pancreatic duct dilatation in a patient with big-duct chronic pancreatitis.**

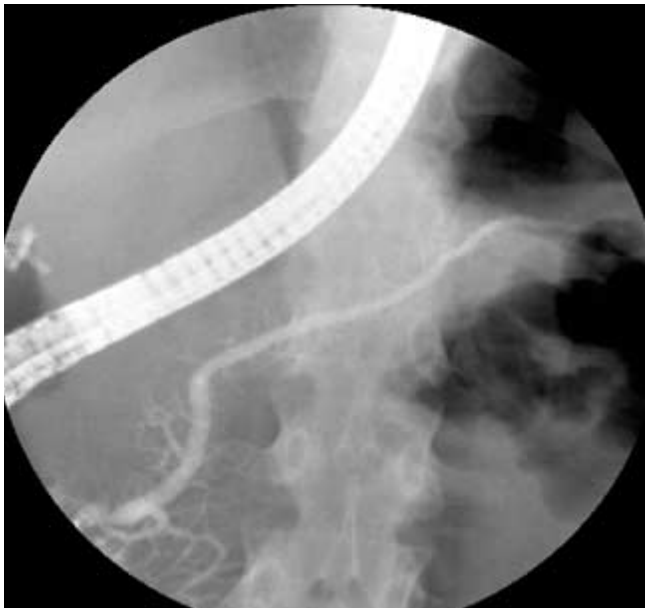


Figure 5 **An endoscopic retrograde cholangiopancreatography image demonstrating minimal pancreatic duct abnormalities in a patient with painful small-duct chronic pancreatitis.**

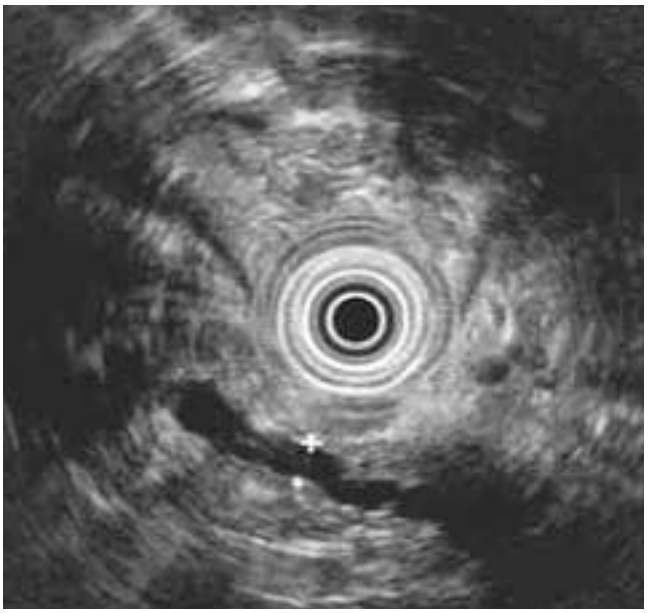


Figure 6 **An endoscopic ultrasound image demonstrating a dilated pancreatic duct (markers) in a patient with advanced chronic pancreatitis.**

Diagnostic Approach

The diagnostic approach to chronic pancreatitis should begin with tests that are safe, inexpensive, and able to detect relatively far advanced disease. Diagnostic tests that fit in this category include serum trypsinogen, fecal elastase, and abdominal US. If these tests do not lead to a diagnosis, riskier or more expensive tests will generally need to be employed (e.g., MRI/MRCP, CT, ERCP, or EUS). Direct pancreatic function testing, if available, logically should be used after the initial tests and before the more expensive or invasive tests, because direct pancreatic function tests are the most sensitive tests available and are lower in cost and less risky than the other options. Because most clinicians do not have access to pancreatic function testing, the riskier, more expensive tests are usually performed, starting with a good-quality CT using pancreatic protocol.

MANAGEMENT

Abdominal Pain

Pain is the most common symptom of chronic pancreatitis requiring medical care. There are a number of potential causes of pain in chronic pancreatitis, including inflammation of pan-

creatic nerves, pancreatic tissue ischemia, increased pressure in the gland or an associated pseudocyst, obstruction of a surrounding hollow viscus (e.g., duodenum or bile duct), and coexistent pancreatic carcinoma. The initial evaluation should focus on identifying conditions for which specific therapy exists. These conditions include pancreatic pseudocyst, duodenal or bile duct compression, and superimposed pancreatic carcinoma. This evaluation is most commonly done by performing a high-quality CT scan of the abdomen. If such a condition is identified, specific therapy is required [*see* Treatment of Other Complications, *below*].

Analgesia and cessation of alcohol use Nonspecific measures to reduce pain in chronic pancreatitis include cessation of alcohol consumption and use of analgesics. Cessation of alcohol consumption can reduce pain in some patients with alcoholic chronic pancreatitis and may prolong life by preventing other alcohol-induced diseases. Unfortunately, abstinence will not halt the progression of chronic pancreatitis, although it may slow it. Analgesics are generally needed, but it is important to start with the least potent agents first (e.g., propoxyphene napsulate or tramadol) because narcotic addiction can occur in up to 30% of patients. Many patients will require more potent narcotics. Adding an antidepressant (a selective serotonin reuptake inhibitor or a tricyclic) may allow potentiation of the narcotic effect. If these simple measures fail, further therapy utilizing medical, endoscopic, or surgical techniques can be considered.

Administration of pancreatic enzymes Several controlled trials have attempted to delineate the effectiveness of orally administered pancreatic enzymes to decrease pain. The concept behind this therapy is that proteases that are present in the duodenum may reduce the stimulus for pancreatic secretion by a negative-feedback mechanism. Only conventional (non–enteric-coated) preparations of pancreatic enzymes can deliver proteases to the duodenum; enteric-coated enzyme preparations deliver proteases too far distally to achieve a negative-feedback effect. Two studies utilizing non–enteric-coated enzymes at high dosages (8 tablets total, in four divided doses with meals and at night, coupled with an agent to suppress gastric acid to prevent premature inactivation of the enzymes) demonstrated a reduction in pain. Four other studies utilizing enteric-coated enzymes showed no effect. A meta-analysis of these studies suggested enzymes are of no benefit in treating pain.[94] In the two studies that demonstrated effectiveness, it appears that persons with less advanced disease (small-duct chronic pancreatitis) respond best, and females with idiopathic chronic pancreatitis seem to have the highest response rate. The role of enzymes in treating pain is controversial, although a consensus review advocated their use.[95] Non–enteric-coated enzyme preparations are of modest effectiveness, but they may be worth a trial in patients with less advanced disease in whom other simple medical measures have failed. A trial of enzymes for pain is generally pointless in patients with advanced or big-duct chronic pancreatitis (mainly alcoholic chronic pancreatitis). These enzyme products are inactivated by gastric acid; thus, concomitant therapy with an agent to reduce gastric acid (H_2 receptor antagonist or proton pump inhibitor) is needed.

Neurolysis Neurolysis via a celiac plexus block or thoracoscopic splanchnicectomy has been evaluated in a number of small studies. The use of percutaneous celiac plexus block has largely been abandoned in patients with chronic pancreatitis because of its transitory effectiveness. EUS-guided celiac plexus block is simpler and safer than percutaneous techniques and appears to last longer.[96] Thoracoscopic splanchnicectomy involves sectioning the splanchnic nerves at thoracoscopy. The short-term response is good (60% to 80% response), but the long-term response has been disappointing.[97]

Endoscopic and surgical therapy Endoscopic therapy and surgical therapy are most useful in patients with advanced or big-duct chronic pancreatitis. Endoscopic therapy has a general goal of relieving obstruction in the pancreatic duct by dilating or stenting a stricture or by removing an obstructing stone. Only a subset of patients with chronic pancreatitis are candidates for such therapy; generally, endoscopic therapy may be considered an option in patients with a dilated pancreatic duct (big-duct chronic pancreatitis) and an obstructing stricture or stone in the head of the gland. The results of large case series indicate that endoscopic therapy may improve pain in 70% to 80% of carefully selected patients.[73,95,98,99] The only randomized, controlled trial that compared endoscopic therapy with surgery showed that the two therapies provide equivalent short-term pain relief, but surgery provided better long-term pain relief.[100] Even if endoscopic therapy is not possible, however, an ERCP may provide useful information for planning surgical therapy.

Surgical procedures that are commonly used to treat pain include decompression of the main pancreatic duct, with or without resection of a portion of the pancreas. Surgery may also be indicated for treatment of a complication such as a pseudocyst, duodenal or common bile duct obstruction, or pancreatic fistula. For patients with intractable pain and a dilated duct, the most commonly performed procedure is the lateral pancreaticojejunostomy (i.e., modified Puestow procedure), in which the pancreatic duct is widely incised along its length and overlaid with a defunctionalized loop of small intestine for drainage of pancreatic juice directly into the small bowel. In some patients, resection of a portion of the pancreatic head may also be needed to decompress smaller-duct branches in the pancreatic head and uncinate process. Substantial pain relief is obtained in 65% to 85% of patients and appears to be relatively long lasting (in some studies, pain relief was sustained for more than 7 years); over time, the response declines to about 50%.[95,101] Mortality for these procedures is about 3% in experienced hands.

Table 8 Enzyme Supplements for Chronic Pancreatitis

Preparation	*Units of Lipase per Tablet or Capsule*
Nonenteric-coated (conventional) preparations	—
Viokase 8, 16	8,000 and 16,000, respectively
Kuzyme-HP	8,000
Generic pancrealipase	8,000
Enteric-coated preparations	—
Creon 10, 20	10,000 and 20,000, respectively
Pancrease MT 4, 10, 16	4,000, 10,000, and 16,000, respectively
Ultrase MT 6, 12, 18, 20	6,000, 12,000, 18,000, and 20,000, respectively

Steatorrhea

The pancreas possesses a 90% functional reserve for the secretion of digestive enzymes.[84] Hence, only 10% of normal maximal output of enzymes is needed to prevent malabsorption of nutrients. Steatorrhea is therefore a late complication of chronic pancreatitis requiring the presence of substantial pancreatic damage. Steatorrhea takes, on average, 13 and 26 years to develop after a diagnosis of alcoholic and idiopathic chronic pancreatitis, respectively.[79] Fat malabsorption tends to occur earlier than protein or carbohydrate malabsorption. Steatorrhea is most precisely established by measuring fecal fat excretion over 72 hours while the patient is on a diet that contains 100 g of fat a day. This test, however, is cumbersome to perform and unpleasant for both patient and staff. In practice, steatorrhea is more commonly diagnosed by identification of clinical features of oily or floating stool, diarrhea, and weight loss and is confirmed by the response to pancreatic enzyme supplementation. Unlike for the treatment of pain, enteric-coated enzyme preparations are commonly selected over non–enteric-coated preparations for the treatment of steatorrhea because of the higher potency and the need for fewer pills with the enteric-coated agents. However, non–enteric-coated enzyme preparations can be used effectively to treat steatorrhea if they are administered in sufficient doses and coadministered with an agent that reduces gastric acid. At least 30,000 units of lipase must be delivered to the small intestine during the prandial and postprandial period to reduce steatorrhea to a manageable level [*see Table 8*]. The clinical goal of this enzyme replacement therapy is to reduce the diarrhea and the losses in stool of fat, protein, and carbohydrate, thereby allowing maintenance or improvement of weight and nutritional status. In patients who do not achieve these end points, explanations for therapeutic failure need to be considered. Such explanations include patient noncompliance, inadequate dosage of enzyme therapy, destruction of exogenous enzymes by gastric acid (if non–enteric-coated preparations were prescribed), poor diet (particularly in chronic alcoholics), and the presence of a second disease causing malabsorption (e.g., small bowel bacterial overgrowth).[102]

Endocrine Insufficiency

Diabetes mellitus, like steatorrhea, is a late complication of chronic pancreatitis. Progressive destruction of the islets of Langerhans can destroy both insulin- and glucagon-secreting cells. The inadequate glucagon reserves predispose patients with this disorder to treatment-induced hypoglycemia. Complications of diabetes such as neuropathy, retinopathy, and nephropathy occur at the same rate as in other patients with diabetes mellitus.[103] Therapy is usually directed at controlling urinary losses of glucose rather than at maintaining tight control of blood glucose levels. Overvigorous attempts at tight control of blood glucose levels are often associated with disastrous complications of treatment-induced hypoglycemia.[104] However, attempts at tight control of blood glucose levels are indicated in patients with hyperlipidemic pancreatitis because in this group, the diabetes is usually a primary illness and tight control of blood glucose levels makes control of serum lipid levels possible.

Other Complications

Pancreatic pseudocyst In chronic pancreatitis, as in acute pancreatitis [*see* Acute Pancreatitis, Treatment of Complications, *above*], asymptomatic pseudocysts less than 6 cm can be safely observed. However, unlike in acute pancreatitis, most pseudocysts that occur in the setting of chronic pancreatitis are generally mature at the time of their diagnosis, and therapy therefore need not be delayed if therapy is indicated. Symptomatic, complicated, or enlarging pseudocysts require therapy by percutaneous, endoscopic, or surgical techniques. Surgical therapy usually involves cyst decompression into a loop of small bowel, and this is often coupled with a pancreatic duct drainage procedure (e.g., modified Puestow procedure). Surgical therapy has a long-term success rate of more than 90%, with an operative mortality of less than 3%.[69,70,73] Percutaneous tube drainage of pseudocysts is also a management option and is immediately successful in 95% of patients. The long-term success rate of percutaneous drainage is still unknown, but it is certainly less than that of surgical techniques. Endoscopic therapy has short-term success rates greater than 90%.[69,70] The limited number of studies that have evaluated the long-term success rate of endoscopic drainage suggest excellent results, with complete resolution of the pseudocyst observed in 90% of patients.[69,70,105]

Complicated pseudocysts may require specific types of therapy. An infected pseudocyst (pancreatic abscess) generally responds to antibiotics and drainage (e.g., endoscopic, percutaneous, or surgical). Bleeding from a pseudocyst may occur in small vessels in its wall or from an associated large arterial pseudoaneurysm. Bleeding from a pseudoaneurysm requires urgent angiography with embolization, sometimes followed by surgical therapy.[58] If no pseudoaneurysm is present on angiography, surgical therapy remains the best choice of therapy for bleeding pseudocysts. Some pseudocysts may rupture and produce a pancreatic fistula that drains into the peritoneal cavity (producing pancreatic ascites) or into the pleural space (producing a pancreatic pleural effusion). In such cases, patients may not complain of symptoms of chronic pancreatitis but may instead note abdominal distention or shortness of breath. The diagnosis can be established by documenting high levels of amylase in the leaked fluid, typically more than 4,000 U/L.[73,106] Treatment may require surgery, and ERCP is used preoperatively to delineate the location of the leak. In many patients, endoscopic therapy and placement of a stent across the fistula site will prove curative.

Other cystic lesions A number of other cystic lesions may occur in the pancreas, including true cysts and cystic neoplasms. Serous cystic neoplasms are benign, but mucin-producing cystic neoplasms may follow a more malignant course. Mucinous cystic neoplasms present as large cystic collections (cystadenomas and cystadenocarcinomas) and may be relatively asymptomatic. Most cystic neoplasms occur in middle-aged patients, particularly in women.[107] They are often mistaken for pseudocysts and inappropriately treated as such. These cystic neoplasms may follow an initially benign course; but when they undergo malignant degeneration, they have poor outcomes equivalent to those of standard adenocarcinoma.

The presence of a pancreatic cystic collection in a middle-aged person (particularly female) without a previous history of pancreatitis should immediately suggest a cystic neoplasm, not a pseudocyst. The diagnosis of a cystic neoplasm requires histologic evidence of epithelial or neoplastic tissue in the cyst wall. Analysis of the cyst fluid obtained by EUS is becoming a highly useful method of differentiating pseudocysts from cystic neoplasms.[108] When these cystic collections are mistaken for pseudocysts, they are treated with drainage, and no tissue is obtained to allow differentiation of a cystic neoplasm from a pseudocyst. Therefore, the therapy of choice for cystic neoplasms is surgical resection, not drainage.

Intraductal papillary mucinous tumors (IPMT, formerly called mucinous ductal ectasia) are characterized by superficially spreading neoplastic tissue along the wall of the pancreatic duct. This neoplastic tissue produces mucin, and patients usually present with a markedly dilated pancreatic duct filled with gelatinous mucin. The appearance is often pathognomonic on ERCP but is occasionally mistaken for chronic pancreatitis. The natural history of the lesions is variable. In general, resection is attempted if the patient is a fit surgical candidate. Depending on the extent of the neoplastic tissue along the pancreatic duct, extensive resection may be needed to eliminate the lesion.[109]

Pancreatic cancer Chronic pancreatitis is a risk factor for pancreatic carcinoma, and the two diseases can be difficult to distinguish in some patients. The risk is about 4% after 20 years of disease.[110] The cancer risk may be as high as 40% in patients with hereditary chronic pancreatitis.[77] There is no effective method of surveillance in these patients and no absolutely reliable method to distinguish cancer from chronic pancreatitis. EUS with directed biopsy, ERCP with cytologic brushings, CT scan, and tumor markers such as CA19-9 are most commonly used.

Common bile duct or duodenal obstruction Fibrosis and inflammation in the head of the pancreas may compress surrounding hollow structures, particularly the common bile duct and duodenum. Compression of the common bile duct produces jaundice, and duodenal compression produces symptoms similar to those of gastric outlet obstruction. Both duodenal compression and gastric outlet obstruction generally require surgical repair with a biliary bypass or gastrojejunostomy, respectively.

Peter Draganov, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Chris E. Forsmark, M.D., F.A.C.P., has received grants for clinical research from AstraZeneca Pharmaceuticals LP and TAP Pharmaceutical Products, Inc., and has received grants for educational activities from Cook Endoscopy and the Olympus Corporation.

Octreotide acetate and pancreatic enzymes discussed in this chapter have not been approved by the FDA for use in the treatment of chronic pancreatic pain.

References

1. Bradley EL 3rd: A clinically based classification system for acute pancreatitis. Arch Surg 128:586, 1993

2. DiMagno EP, Chari S: Acute pancreatitis. Sleisinger and Fordtran's Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management, 7th ed. Feldman M, Friedman LS, Sleisenger MH, Eds. WB Saunders Co, Philadelphia, 2002, p 913

3. Go VLW, Everhart JE: Pancreatitis. Digestive Diseases in the United States: Epidemiology and Impact. Everhart JE, Ed. (NIH Publication No. 94-1447.) U.S. Dept of Health and Human Services, Public Health Service, National Institutes of Health, Washington, DC, 1994, p 693

4. DeFrances CJ, Hall MJ, Podgornik MN: 2003 National Hospital Discharge Survey. Advance data from vital and health statistics (No. 359). National Center for Health Statistics, Hyattsville, Maryland, 2005
http://www.cdc.gov/nchs/data/ad/ad359.pdf

5. Lowenfels AB, Sullivan T, Fioranti J, et al: The epidemiology and impact of pancreatic diseases in the United States. Curr Gastroenterol Rep 7:90, 2005

6. Lindkvist B, Appelros S, Manjer J, et al: Trends in incidence of acute pancreatitis in a Swedish population: is there really an increase? Clin Gastroenterol Hepatol 2:83, 2004

7. Overweight and Obesity: Obesity Trends: U.S. Obesity Trends 1985–2004 (Online). Centers for Disease Control and Prevention, Atlanta, 2005
http://www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/index.htm

8. Grendell JH: Clinical and economic impact of acute pancreatitis in the United States (abstr). Pancreas 19:422, 1999

9. Mann DV, Hershman MJ, Hittinger R, et al: Multicentre audit of death from acute pancreatitis. Br J Surg 81:890, 1994

10. Talamini G, Bassi C, Falconi M, et al: Risk of death from acute pancreatitis: role of early simple "routine" data. Int J Pancreatol 19:15, 1996

11. Tsai CJ, Leitzmann MF, Willett WC, et al: Prospective study of abdominal adiposity and gallstone disease in U.S. men. Am J Clin Nutr 80:38, 2004

12. Hanck C, Whitcomb DC: Alcoholic pancreatitis. Gastroenterol Clin North Am 33:751, 2004

13. Klein SD, Affronti JP: Pancreas divisum, an evidence-based review: part I, pathophysiology. Gastrointest Endosc 60:419, 2004

14. Klein SD, Affronti JP: Pancreas divisum, an evidence-based review: part II, patient selection and treatment. Gastrointest Endosc 60:585, 2004

15. Lans JI, Geenen JE, Johanson JF, et al: Endoscopic therapy in patients with pancreas divisum and acute pancreatitis: a prospective, randomized, controlled clinical trial. Gastrointest Endosc 38:430, 1992

16. Lankisch PG, Droege M, Gottesleben F: Drug-induced acute pancreatitis: incidence and severity. Gut 37:565, 1995

17. Trivedi CD, Pitchumoni CS: Drug-induced pancreatitis: an update. J Clin Gastroenterol 239:709, 2005

18. Reisler RB, Murphy RL, Redfield RR, et al: Incidence of pancreatitis in HIV-1-infected individuals enrolled in 20 adult AIDS clinical trials group studies: lessons learned. J Acquir Immune Defic Syndr 39:159, 2005

19. Yadav D, Pitchumoni CS: Issues in hyperlipidemic pancreatitis. J Clin Gastroenterol 36:54, 2005

20. Glueck CJ, Lang J, Hamer T, et al: Severe hypertriglyceridemia and pancreatitis when estrogen replacement therapy is given to hypertriglyceridemic women. J Lab Clin Med 123:59, 1994

21. Freeman ML, DiSario JA, Nelson DB, et al: Risk factors for post-ERCP pancreatitis: a prospective, multicenter study. Gastrointest Endosc 54:425, 2001

22. Etemad B, Whitcomb DC: Chronic pancreatitis: diagnosis, classification, and new genetic developments. Gastroenterology 120:682, 2001

23. Hamano H, Kawa S, Horiuchi A, et al: High serum IgG4 concentrations in patients with sclerosing pancreatitis. N Engl J Med 344:732, 2001

24. Lara LP, Chari ST: Autoimmune pancreatitis. Curr Gastroenterol Rep 7:101, 2005

25. Ros E, Navarro S, Bru C, et al: Occult microlithiasis in "idiopathic" acute pancreatitis: prevention of relapses by cholecystectomy or ursodeoxycholic acid therapy. Gastroenterology 101:1701, 1991

26. Lee SP, Nichols JF, Park HZ: Biliary sludge as a cause of acute pancreatitis. N Engl J Med 326:589, 1992

27. Bhatia M, Wong FL, Cao Y, et al: Pathophysiology of acute pancreatitis. Pancreatology 5:132, 2005

28. Halangk W, Lerch MM: Early events in acute pancreatitis. Gastroenterol Clin North Am 33:717, 2004

29. Norman J: The role of cytokines in the pathogenesis of acute pancreatitis. Am J Surg 175:76, 1998

30. Raraty MG, Connor S, Criddle DN, et al: Acute pancreatitis and organ failure: pathophysiology, natural history, and management strategies. Curr Gastroenterol Rep 6:99, 2004

31. Al-Bahrani AZ, Ammori BJ: Clinical laboratory assessment of acute pancreatitis. Clin Chim Acta 362:26, 2005 [Epub July 18, 2005]

32. Lankisch PG, Schirren CA, Kunze E: Undetected fatal pancreatitis. Why is the disease so frequently overlooked? Am J Gastroenterol 86:322, 1991

33. Baillargeon JD, Orav J, Ramagopal V, et al: Hemoconcentration as an early risk factor for necrotizing pancreatitis. Am J Gastroenterol 93:2130, 1998

34. Tenner S, Dubner H, Steinberg W: Predicting gallstone pancreatitis with laboratory parameters: a meta-analysis. Am J Gastroenterol 89:1863, 1994

35. Balthazar EJ, Freeny PC, vanSonnenberg E: Imaging and intervention in acute pancreatitis. Radiology 193:297, 1994

36. Balthazar EJ: Staging of acute pancreatitis. Radiol Clin North Am 40:1129, 2002

37. Balthazar EJ: Complications of acute pancreatitis: clinical and CT evaluation. Radiol Clin North Am 40:1211, 2002

38. Arvanitakis M, Delhaye M, DeMaertelaere V, et al: Computed tomography and magnetic resonance imaging in the assessment of acute pancreatitis. Gastroenterology 126:715, 2004

39. Halla AH, Amortegui JD, Jeroukhimov IM, et al: Magnetic resonance cholangiopancreatography accurately detects common bile duct stones in resolving gallstone pancreatitis. J Am Coll Surg 200:869, 2005

40. Banks PA: Practice guidelines in acute pancreatitis. Am J Gastroenterol 92:377, 1997

41. Papachristou GI, Whitcomb DC: Predictors of severity and necrosis in acute pancreatitis. Gastroenterol Clin North Am 33:871, 2004

42. Baron TH, Morgan DE: Acute necrotizing pancreatitis. N Engl J Med 340:1412, 1999

43. Tenner S, Sica G, Hughes M, et al: Relationship of necrosis to organ failure in severe acute pancreatitis. Gastroenterology 113:899, 1997

44. Tenner S: Initial management of acute pancreatitis: critical issues during the first 72 hours. Am J Gastroenterol 99:2489, 2004

45. McClave SA, Dryden GW: Issues of nutritional support for the patient with acute pancreatitis. Semin Gastrointest Dis 13:154, 2002

46. Kaushik N, O'Keefe SJ: Nutritional support in acute pancreatitis. Curr Gastroenterol Rep 6:320, 2004

47. Al-Omran M, Groof A, Wilke D: Enteral versus parenteral nutrition for acute pancreatitis. Cochrane Database Syst Rev (1):CD002837, 2003

48. Andriulli A, Leandro G, Clemente R, et al: Meta-analysis of somatostatin, octreotide, and gabexate mesilate in the therapy of acute pancreatitis. Aliment Pharmacol Ther 12:237, 1998

49. Steinberg WM, Schlesselman SE: Treatment of acute pancreatitis: comparison of animal and human studies. Gastroenterology 93:1420, 1987

50. Andriulli A, Caruso N, Quitadamo M, et al: Antisecretory vs. antiproteasic drugs in the prevention of post-ERCP pancreatitis: the evidence-based medicine derived from a meta-analysis study. JOP 4:41, 2003

51. Antruilli A, Leandro G, Niro G, et al: Pharmacologic treatment can prevent pancreatic injury after ERCP: a meta-analysis. Gastrointest Endosc 55:100, 2000

52. Johnson CD, Kingsnorth AN, Imrie CE, et al: A double blind, randomized, placebo controlled study of a platelet activating factor antagonist lexipafant in the treatment and prevention of organ failure in predicted severe acute pancreatitis. Gut 48:62, 2001

53. Neoptolemos JP, Ogunbiyi O, Wilson PG, et al: Etiology, pathogenesis, natural history and treatment of biliary acute pancreatitis. The Pancreas. Warshaw AL, Buchler MW, Carr-Locke DL, et al, Eds. Blackwell Science, Malden, Massachusetts, 1998, p 521

54. Neoptolemos JP, London NJ, James D, et al: Controlled trial of urgent endoscopic retrograde cholangiopancreatography and endoscopic sphincterotomy versus conservative management for acute pancreatitis due to gallstones. Lancet 2:979, 1988

55. Fan ST, Lai ECS, Mok FPT, et al: Early treatment of acute biliary pancreatitis by endoscopic papillotomy. N Engl J Med 328:228, 1993

56. Folsch UR, Nitsche R, Ludtke R, et al: Early ERCP and papillotomy compared with conservative management for acute biliary pancreatitis. German Study Group on Acute Biliary Pancreatitis. N Engl J Med 336:237, 1997

57. Forsmark CE, Grendell JH: Complications of pancreatitis. Semin Gastrointest Dis l2:165, 1991

58. Forsmark CE, Wilcox CM, Grendell JH: Endoscopy-negative upper gastrointestinal hemorrhage in a patient with chronic pancreatitis. Gastroenterology 102:320, 1992

59. Gerzof SG, Banks PA, Robbins AH, et al: Early diagnosis of pancreatic infection by computed tomography–guided aspiration. Gastroenterology 93:1315, 1987

60. Beger HG, Bittner R, Block S, et al: Bacterial contamination of pancreatic necrosis: a prospective clinical study. Gastroenterology; 91:433, 1986

61. Kendrick ML, Sarr MG: Pancreatic necrosis and infections in patients with acute pancreatitis. Pancreatitis and Its Complications. Forsmark CE, Ed. Humana Press, Totowa, New Jersey, 2004, p 99

62. Baron TH, Harewood GC, Morgan DG: Outcome differences after endoscopic drainage of pancreatic necrosis, acute pancreatic pseudocysts, and chronic pancreatic pseudocysts. Gastrointest Endosc 56:150, 2002

63. Pederzoli P, Bassi S, Vesenti S, et al: A randomized multicenter clinical trial of antibiotic prophylaxis of septic complications in acute necrotizing pancreatitis with imipenem. Surg Gynecol Obstet 176:480, 1993

64. Nordback I, Sand J, Saaristo R, et al: Early treatment with antibiotics reduces the need for surgery in acute necrotizing pancreatitis—a single-center randomized study. J Gastrointest Surg 5:113, 2001

65. Sainio V, Kemppainen E, Puolakkainen P, et al: Early antibiotic treatment in acute necrotizing pancreatitis. Lancet 346:663, 1995

66. Schwarz M, Isenmann R, Meyer H, et al: Antibiotic use in necrotizing pancreatitis: results of a controlled study. Dtsch Med Wochenschr 122:356, 1997

67. Isenmann R, Runzi M, Kron M, et al: Prophylactic antibiotic treatment in patients with predicted severe acute pancreatitis: a placebo-controlled, double-blind trial. Gastroenterology 126:997, 2004

68. Bassi C, Larvin M, Villatoro E: Antibiotic therapy for prophylaxis against infection of pancreatic necrosis in acute pancreatitis. Cochrane Database Syst Rev (4):CD002941, 2003
http://www.cochrane.org/reviews/en/ab002941.html

69. Baillie J: Pancreatic pseudocysts (part I). Gastrointest Endosc 59:873, 2004

70. Baillie J: Pancreatic pseudocysts (part II). Gastrointest Endosc 60:105, 2004

71. Giovannini M: Endoscopic ultrasound-guided pancreatic pseudocyst drainage. Gastrointest Endosc Clin N Am 15:179, 2005

72. Sarles H, Adler G, Dani R, et al: Classifications of pancreatitis and definition of pancreatic diseases. Digestion 43:234, 1989

73. Forsmark CE: Chronic pancreatitis. Gastrointestinal and Liver Disease. Pathophysiology, Diagnosis, Management, 7th ed. Feldman M, Friedman LS, Sleisenger MH, Eds. WB Saunders Co, Philadelphia, 2002, p 940

74. Lowenfels AB, Maisonneuve P, Cavallini G, et al: Prognosis of chronic pancreatitis: an international multicenter study. Am J Gastroenterol 89:1467, 1994

75. Schneider A: Serine protease inhibitor Kazal type 1 mutations and pancreatitis. Gasteroenterol Clin North Am 33:789, 2004

76. Howes N, Greenhalf W, Stocken DD, et al: Cationic trypsinogen mutations and pancreatitis. Gastroenterol Clin North Am 33:767, 2004

77. Lowenfels AB, Maisonneuve P, Whitcomb DC: Risk factors for cancer in hereditary pancreatitis. International Hereditary Pancreatitis Study Group. Med Clin North Am 84:565, 2000

78. Cohn JA, Mitchell RM, Jowell PS: The role of cystic fibrosis gene mutations in determining susceptibility to chronic pancreatitis. Gastroenterol Clin North Am 33:817, 2004

79. Layer P, Yamamoto H, Kalthoff L, et al: The different courses of early- and late-onset idiopathic and alcoholic chronic pancreatitis. Gastroenterology 107:1481, 1994

80. Stevens T, Conwell DL, Zuccaro G: Pathogenesis of chronic pancreatitis: an evidence-based review of past theories and recent developments. Am J Gastroenterol 99:2256, 2004

81. Whitcomb DC: Genetic predisposition to alcoholic chronic pancreatitis. Pancreas 27:321, 2003

82. Chowdhury RS, Forsmark CE: Pancreatic function testing. Aliment Pharmacol Ther 17:733, 2003

83. Jacobsen DG, Currington C, Connery K, et al: Trypsin-like immunoreactivity as a test for pancreatic insufficiency. N Engl J Med 310:1307, 1984

84. DiMagno EP, Go VLW, Summerskill WHJ: Relations between pancreatic enzyme outputs and malabsorption in severe pancreatic insufficiency. N Engl J Med 288:813, 1973

85. Lankisch PG: Now that fecal elastase is available in the United States, should clinicians start using it? Curr Gastroenterol Rep 6:126, 2004

86. Forsmark CE, Toskes PP: What does an abnormal pancreatogram mean? Gastrointest Endosc Clin N Am 5:105, 1995

87. Draganov P, Patel A, Fazel A, et al: Prospective evaluation of the accuracy of the intraductal secretin test in the diagnosis of chronic pancreatitis. Clin Gastroenterol Hepatol 3:695, 2005

88. Conwell DL, Zuccaro G Jr, Vargo JJ, et al: An endoscopic pancreatic function test with synthetic porcine secretin for the evaluation of chronic abdominal pain and suspected chronic pancreatitis. Gastrointest Endosc 57:37, 2003

89. Kwon RS, Brugge WR: New advances in pancreatic imaging. Curr Opin Gastroenterol 21:561, 2005

90. Del Frate C, Zanardi R, Mortele K, et al: Advances in imaging for pancreatic disease. Curr Gastroenterol Rep 4:140, 2002

91. Ball MA, Sztantics A, Metens T, et al: Quantification of pancreatic exocrine function with secretin-enhanced magnetic resonance cholangiopancreatography: normal values and short-term effects of pancreatic duct drainage procedures in chronic pancreatitis: initial results. Eur Radiol l15:2110, 2005

92. Wallace MB, Hawes RH: Endoscopic ultrasound in the diagnosis and treatment of chronic pancreatitis. Pancreas 23:26, 2001

93. Forsmark CE: The diagnosis of chronic pancreatitis. Gastrointest Endosc 52:293, 2000

94. Brown A, Hughes M, Tenner S, et al: Does pancreatic enzyme supplementation reduce pain in patients with chronic pancreatitis: a meta-analysis. Am J Gastroenterol 92:2032, 1997

95. Warshaw AL, Banks PA, Fernandez-Del Castillo C: AGA technical review: treatment of pain in chronic pancreatitis. Gastroenterology 115:763, 1998

96. Gress F, Schmidt C, Sherman S, et al: A prospective randomized comparison of endoscopic ultrasound– and computed tomography–guided celiac plexus block for managing chronic pancreatitis pain. Am J Gastroenterol 94:900, 1999

97. Bradley EL 3rd, Bem J: Nerve blocks and neuroablative surgery for chronic pancreatitis. World J Surg 27:1241, 2003

98. Delhaye M, Arvanitakis M, Verset G, et al: Long-term clinical outcome after endoscopic pancreatic ductal drainage for patients with painful chronic pancreatitis. Clin Gastroenterol Hepatol 2:1096, 2004

99. Rösch T, Daniel S, Scholz M, et al: Endoscopic treatment of chronic pancreatitis: a multicenter study of 1,000 patients with long-term follow-up. Endoscopy 34:765, 2002

100. Dite P, Ruzicka M, Zboril V, et al: A prospective, randomized trial comparing endoscopic with surgical therapy for chronic pancreatitis. Endoscopy 35:553, 2003

101. Sakorafas GH, Farnell MB, Farley DR, et al: Long-term results after surgery for chronic pancreatitis. Int J Pancreatol 27:131, 2000

102. Pongprasobchai S, DiMango EP: Treatment of exocrine pancreatic insufficiency. Pancreatitis and Its Complications. Forsmark CD, Ed. Humana Press, Totowa, New Jersey, 2004, p 295

103. Levitt NS, Adams G, Salmon J, et al: The prevalence and severity of microvascular complications in pancreatic diabetes and IDDM. Diabetes Care 18:971, 1995

104. Linde J, Nilsson LH, Barany FR: Diabetes and hypoglycemia in chronic pancreatitis. Scand J Gastroenterol 12:369, 1977

105. Giovannini M, Pesenti C, Rolland AL, et al: Endoscopic ultrasound-guided drainage of pancreatic pseudocysts or pancreatic abscesses using a therapeutic echo endoscope. Endoscopy 33:473, 2001

106. Rockey DC, Cello JP: Pancreaticopleural fistula: report of 7 cases and a review of the literature. Medicine (Baltimore) 69:332, 1990

107. Mishra G, Forsmark CE: Cystic neoplasms of the pancreas. Curr Treat Options Gastroenterol 3:355, 2000

108. Levy MJ, Clain JE: Evaluation and management of cystic pancreatic tumors: emphasis on the role of EUS FNA. Clin Gastroenterol Hepatol 2:639, 2004

109. Fernandez-del Castillo C, Warshaw AL: Cystic neoplasms of the pancreas. Pancreatology 1:641, 2001

110. Lowenfels AB, Maisonneuve P, Cavallini G, et al: Pancreatitis and the risk of pancreatic cancer. N Engl J Med 328:1433, 1993

VI GALLSTONES AND BILIARY TRACT DISEASE

Kimberly M. Persley, m.d.
Rajeev Jain, m.d., f.a.c.p.

Gallstone and biliary tract diseases constitute a common and costly health problem in the United States.[1] The prevalence of gallstones increases with age in all racial groups. Increased body weight, rapid weight loss, pregnancy, alcoholic cirrhosis, and a family history of gallstone disease also appear to be risk factors.[2-4]

Incidence and Prevalence of Gallstones

In one epidemiologic study of persons 30 years of age and older, new gallstones were found to develop in 2.2% of men and 2.9% of women over a 5-year period.[2] In the United States, gallstones occur in approximately 10% of persons older than 40 years, but the prevalence is significantly higher in women, increasing to 20% to 25% in women older than 50 years. Fortunately, only 20% to 30% of gallstones are symptomatic, with biliary colic being the most common symptom. A report from the Third National Health and Nutrition Examination Survey (NHANES III) stated that an estimated 6.3 million men and 14.2 million women 20 to 74 years of age had gallbladder diseases.[5]

Gallstone Formation

Two principal types of stone, the cholesterol stone and the pigment stone, form in the gallbladder and biliary tract. The cholesterol stone is composed mainly of cholesterol (> 50% of stone composition) and comprises multiple layers of cholesterol crystals and mucin glycoproteins. Mixed gallstones contain 20% to 50% cholesterol. The pigment stone contains a wide variety of organic and inorganic components, including calcium bilirubinate (40% to 50% of dry weight). In Europe and the United States, 90% of gallstones are of the cholesterol or mixed type; the remainder are pigment gallstones. Multiple risk factors for cholesterol and pigment gallstone formation have been identified [*see Table 1*]. In cholesterol gallstone formation, a genetic predisposition has been proposed on the basis of murine models and epidemiologic studies that show ethnic and geographic differences, as well as familial clustering of cholesterol gallstone disease.[6]

pigment gallstone formation

The pathogenesis of pigment gallstones is not completely understood, but bacteria may play a central role.[7] Black pigment stones are most often seen in patients with cirrhosis or hemolytic anemia and are found predominantly in the gallbladder. Brown pigment stones, which are common in Asians, are the most common stone to appear de novo in the bile duct and are associated with biliary tract infection. The prevalence of gallstones and gallbladder disease in Asians ranges from 5% to 20%.[8] Pigment stones, in contrast to cholesterol stones, are often radiopaque and can be seen on plain abdominal x-rays [*see Figure 1*].

cholesterol gallstone formation

Cholesterol is a minor but clinically significant component of bile. The other components of bile are bile salts, phospholipids, conjugated bilirubin, fatty acids, water, electrolytes, and other organic and inorganic substances. Cholesterol is a hydrophobic molecule that is relatively insoluble in water and precipitates unless it is maintained in solution by bile salts. Bile salt molecules possess hydrophilic (water-soluble) and hydrophobic (fat-soluble) regions that maintain cholesterol in a soluble state.

When bile salt molecules in water reach concentrations of 2 to 4 mM, they form spherical complexes called micelles; the concentration at which micelles form is known as the critical micellar concentration. In micelles, the negatively charged hydrophilic ends of the molecules face outward, toward the water, and the uncharged hydrophobic regions face the center of the sphere, toward one another. Cholesterol molecules are enclosed in the hydrophobic interiors.

A pure bile salt micelle must comprise at least 50 molecules to enclose a single molecule of cholesterol. The intercalation of phospholipids (principally lecithin) between bile salt molecules of a micelle improve the efficiency with which the micelle solubilizes cholesterol. Such a mixed micelle [*see Figure 2*], which is the type that exists in bile, needs only seven bile salt molecules to solubilize one cholesterol molecule. Free bile salt molecules exist in equilibrium with mixed micelles in a water solution. The combined molar concentration of bile salt and phospholipid is about 11 times that of cholesterol.

Cholesterol Saturated Bile

Cholesterol gallstone formation is potentiated by hepatic secretion of bile containing excess cholesterol relative to the concentration of bile salt.[3,9] This occurs most often because of an increase in the biliary concentration of cholesterol but may also result from decreased bile acid secretion in certain disease states. Excess cholesterol is solubilized in micelles and in vesicles composed of phospholipid bilayers. Cholesterol crystal formation seems to occur at the surface of these vesicles.

Nucleation of Crystals and Gallbladder Stasis

In addition to supersaturated bile, nucleation of crystals and gallbladder stasis are also important factors in gallstone formation. Microscopic crystals initially precipitate from a supersatu-

Table 1 Risk Factors for Cholesterol and Pigment Gallstone Formation

Cholesterol Gallstones	*Pigment Gallstones*
Increasing age	Increasing age
Female gender	Chronic hemolysis
Obesity	Alcoholic liver disease
Rapid weight loss	Biliary infection
Native-American heritage	Asian heritage
Hyperalimentation (gallbladder stasis)	Hyperalimentation (gallbladder stasis)
Elevated triglyceride levels	Duodenal diverticulum
Medications (e.g., fibric acid derivatives, estrogens, octreotide)	Truncal vagotomy
Ileal disease, resection, or bypass	Primary biliary cirrhosis

rated bile in a process called nucleation, which is influenced by several pronucleating and antinucleating proteins.[3,9] Protein mucins, which are secreted by the gallbladder, and calcium are crucial promoters of the nucleation process. Prostaglandins stimulate the synthesis and secretion of mucins. Antinucleating factors, such as certain apolipoproteins, have been less well studied. Gallbladder stasis, with concentration and acidification of bile, is also an important factor in gallstone formation, promoting the growth of cholesterol crystals into stones. Cholesterol stones rarely recur in patients after cholecystectomy.

Development of Biliary Sludge

Biliary sludge (or microlithiasis) is a term that is often applied to cholesterol crystals of sufficient number to be visualized on ultrasonography.[10] Biliary sludge, the precursor of most gallstones, is a mix of mucus, cholesterol monohydrate microcrystals, and calcium bilirubinate granules.[4] Gallbladder sludge has been shown to precipitate biliary colic, acute cholecystitis, or pancreatitis and should be regarded as part of the spectrum of gallstone disease.

Gallbladder sludge and gallstones occur in 10% of women during pregnancy and in the early postpartum period, with significant spontaneous regression.[11] Gallbladder sludge is also associated with fasting, rapid weight loss, parenteral nutrition, cirrhosis, and certain medications, such as ceftriaxone, cyclosporine, and octreotide.[10]

Choledocholithiasis

Common bile duct stones may form de novo in bile ducts (so-called primary choledocholithiasis, constituting 5% of bile duct stones) or migrate from the gallbladder to the biliary tract (secondary choledocholithiasis, constituting 95% of bile duct stones). Stones in the biliary tract usually have the same composition as those in the gallbladder, although some are softer and brownish. The brown color is a result of deposition of calcium bilirubinate and other calcium salts as a result of bacterial deconjugation of bilirubin and hydrolysis of phospholipids.

Cholecystitis and Cholelithiasis

Patients who have stones in the gallbladder or the biliary tree display syndromes that range from acute disease to chronic symptomatic or silent disease. Gallstone disease (cholelithiasis) may remain silent throughout a person's lifetime. At any stage of disease, obstruction of the cystic duct or common bile duct by a gallstone that has passed from the gallbladder may cause pain, with or without acute inflammation (cholecystitis).

SYMPTOMATIC CHOLELITHIASIS

With the exception of biliary colic, most symptoms of gallstones are not specific for gallstone disease; a meta-analysis indicated that 80% of patients with gallstones presented with other abdominal symptoms.[12] Biliary colic is a misnomer, because the pain is steady and not colicky. The pain of biliary colic is caused by functional spasm of the cystic duct obstructed by a stone. Biliary colic often develops without any precipitating events. Typically, the pain is localized to the epigastrium, has a sudden onset, and increases rapidly in intensity to a plateau that can last as long as 3 hours before subsiding. Some patients describe the pain as excruciating or lancinating, whereas others describe it as a deep ache or cramp. The pain may radiate to the interscapular region or to the right shoulder, and it may be associated with nausea or vomiting. The pain is less frequently located in the left upper quadrant, precordium, or lower abdomen. Pain lasting longer than 6 hours or pain that is associated with fever suggests acute cholecystitis. Gastrointestinal symptoms, such as dyspepsia, heartburn, bloating, and fatty-food intolerance, are common whether or not gallstones are present. Thus, the diagnosis of biliary colic is based on clinical judgment. Once an episode of biliary colic has occurred, repeated attacks of pain are common.

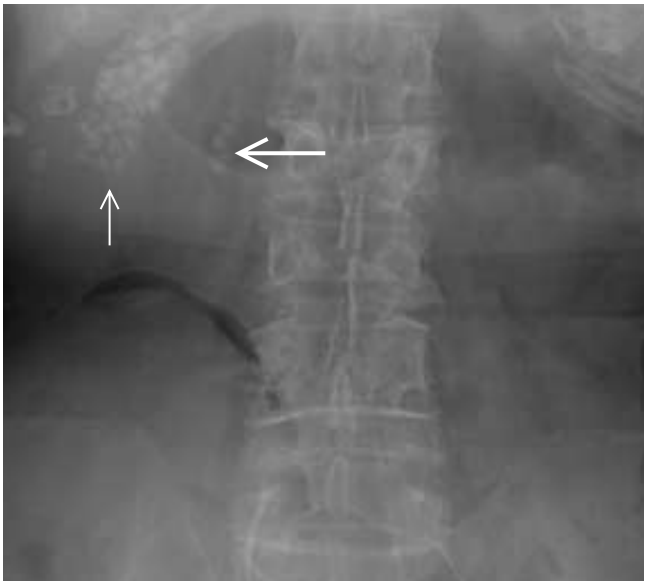


Figure 1 **Radiograph of the right upper abdominal quadrant showing radiopaque stones in the gallbladder (small arrow) and common bile duct (large arrow).**

ACUTE CHOLECYSTITIS

Acute cholecystitis refers to a syndrome of abdominal pain, fever, and leukocytosis associated with gallbladder inflammation, which is usually related to gallstone disease. Cholelithiasis is present in 90% to 95% of patients with acute cholecystitis, and most patients have had previous attacks of biliary colic. Acute cholecystitis may present as an acalculous cholecystitis in 5% to 10% of patients. It is predominantly noted in older men who are critically ill after major surgery, severe trauma, or extensive burn injury.[13] In rare cases, acute cholecystitis can result from a specific infection, such as that caused by *Salmonella* species. *Salmonella* organisms can also colonize the gallbladder epithelium without inflammation (carrier state). Cytomegalovirus and cytosporidia can infect the biliary system, resulting in cholecystitis or cholangitis in immunocompromised patients, such as those with AIDS or those who have undergone bone marrow transplantation.

Diagnosis

Clinical manifestations An episode of prolonged right upper quadrant pain (> 6 hours), especially if associated with fever, should arouse suspicion of acute cholecystitis, as opposed to simple biliary colic. Clinical features of acute cholecystitis include anorexia, nausea, vomiting, fever, and abdominal pain that initially may localize to the epigastrium before shifting to the right upper quadrant. The severe abdominal pain associated with acute cholecystitis is caused by inflammation of the gallbladder wall.[9] Most patients who present with jaundice have stones in the common bile duct. Patients are ill for several days

to a week before the acute attack completely subsides. Acalculous cholecystitis, an acute necroinflammatory disease, is clinically identical to acute cholecystitis but is not associated with gallstones.

Physical examination Physical examination may reveal upper quadrant subcostal tenderness and pain on inspiration, often with inspiratory arrest (the Murphy sign). Of all physical examination findings, the Murphy sign has the highest positive likelihood ratio (LR) for acute cholecystitis (LR, 2.8; 95% confidence interval [CI], 0.8 to 8.6).[14] The gallbladder may be palpable, especially at the time of the first attack, before fibrosis has reduced its distensibility. Tenderness, guarding, and rebound pain in the area of an inflamed gallbladder are important findings. Generalized rebound tenderness in a patient who has been ill for several days may reflect a perforation; however, localized tenderness may indicate secondary pancreatitis or an abscess in the area of the gallbladder.

Laboratory evaluation Laboratory tests frequently reveal leukocytosis and mild hyperbilirubinemia, which may occur in the absence of biliary obstruction secondary to reduced hepatic excretion of bile.[15] In one prospective study, 25% of patients with acute cholecystitis had a serum bilirubin level of 2 to 5 mg/dl and had no common bile duct abnormalities; 4% had an elevated amylase level without pancreatitis.[16] No single laboratory parameter has a sufficient positive or negative LR to establish the diagnosis of acute cholecystitis or rule it out.[14]

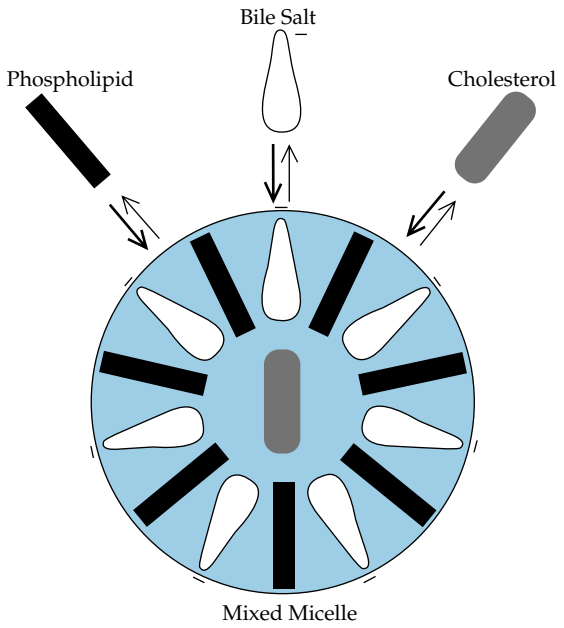


Figure 2 **Cholesterol is solubilized in bile by the formation of mixed micelles that consist predominantly of bile salt and phospholipid. Micelles form when the concentration of bile salts in water is between 2 and 4 mM, the so-called critical micellar concentration (CMC). The negatively charged hydrophilic region of the bile salt molecule faces outward into the water phase, whereas the uncharged hydrophobic region is directed inward. These three components—bile salts, phospholipids, and cholesterol—exist in equilibrium between the free state and micelle constituents. At the CMC for bile salts, the equilibrium shifts strongly in the direction of the micelle. If bile salt concentrations are insufficient, the hydrophobic cholesterol molecules will precipitate to form a nidus for a gallstone.**

Imaging studies Acute cholecystitis should be suspected when a patient presenting with certain clinical manifestations (see above) is found to have gallstones on an imaging study. However, the mere presence of gallstones is not confirmation of acute cholecystitis, because asymptomatic cholelithiasis [*see* Asymptomatic Cholelithiasis, *below*] is a common condition in the general population. Ultimately, the combination of physical findings and laboratory results suggests the diagnosis of acute cholecystitis, which can then be confirmed by diagnostic imaging.[14]

Transabdominal ultrasonography (TUS) is the diagnostic procedure of choice for a patient with suspected gallstones and acute cholecystitis. Meta-analysis indicates that TUS has a sensitivity of 88% to 90% and a specificity of 97% to 98% for the diagnosis of gallstones larger than 2 mm.[17] Gallbladder ultrasonography should ideally be preceded by an 8-hour fast, because gallstones are best visualized in a distended, bile-filled gallbladder. In addition to detecting gallstones, TUS can be used to identify other causes of right upper quadrant pain, such as hepatic abscess or malignancy, and it may reveal biliary duct obstruction. However, specific evidence of acute cholecystitis (i.e., the presence of pericholecystic fluid, edema of the gallbladder wall, or both) is found infrequently. Occasionally, a sonographic Murphy sign will be elicited when the ultrasound probe is positioned below the right costal margin.

Cholescintigraphy is the best method of confirming the clinical diagnosis of acute cholecystitis.[17] This procedure takes only 60 to 90 minutes and involves the intravenous injection of technetium-99m (^{99m}TC)–labeled hepatoiminodiacetic acid (HIDA, or lidofenin), which is selectively excreted into the biliary tree and enters the gallbladder. In the presence of acute cholecystitis, radiolabeled material enters the common bile duct and duodenum but not the gallbladder. Meta-analysis suggests that radionuclide scanning is the most accurate method of diagnosing acute cholecystitis.[17] Occasionally, the scan gives false positive results in patients who have alcoholic liver disease or who are fasting or receiving total parenteral nutrition; however, false negative results are rare. Radionuclide scanning may not be useful for patients with deep jaundice, because the labeled agent fails to enter the biliary tree.

Direct examination of bile is more sensitive than ultrasonography in the diagnosis of biliary sludge. Ideally, gallbladder bile, rather than hepatic and ductal bile, should be obtained to maximize sensitivity for detecting microlithiasis; gallbladder bile is most reliably obtained by cholecystokinin-induced stimulation of the gallbladder [*see* Chronic Cholecystitis, Bile Collection and Examination, *below*]. Bile collected at endoscopic retrograde cholangiopancreatography (ERCP) after the injection of contrast may lead to false positive findings of crystals.[18] Bile must be centrifuged and examined under polarizing or light microscopy for detection of crystals. Plain abdominal x-rays are much less useful than cholescintigraphy or ultrasonography, because only 15% to 20% of stones are radiopaque; oral cholecystography is also less useful and is now rarely performed because it requires 24 to 48 hours to perform and is less accurate than ultrasonography.

Differential Diagnosis

Because no single clinical or laboratory measurement carries sufficient weight to establish or exclude the diagnosis of acute cholecystitis,[14] the differential diagnosis is broad; it includes diseases that are characterized by severe epigastric symptoms and transient abnormal results on liver function testing.

Severe acute viral hepatitis or alcoholic hepatitis may be associated with moderately severe right upper quadrant pain, fever, and leukocytosis. A history of acute alcoholism, the finding of an enlarged liver, or markedly elevated aminotransferase levels should help distinguish one of these diagnoses from acute cholecystitis.

A patient with a penetrating or perforating ulcer may have severe epigastric pain and usually has a history of ulcer; free air may be evident on a plain abdominal x-ray if the ulcer has perforated. Early in its course, acute appendicitis may produce symptoms similar to those of acute cholecystitis, particularly if the appendix is retrocecal or the cecum is malpositioned in the subhepatic area. Acute pyelonephritis of the right kidney may produce anterior pain similar to the pain that occurs with acute cholecystitis. Pneumonia or infarction of the right lung may also cause abdominal symptoms.

Acute pancreatitis may be nearly impossible to distinguish from acute cholecystitis. Patients with either disorder may exhibit moderate signs on physical examination, with tenderness or localized rebound pain in the epigastrium. Serum amylase and lipase levels can be high in either condition, but the higher these enzyme levels are, the more likely it is that pancreatitis is present. Cholelithiasis occasionally causes pancreatitis, which further complicates the diagnosis. At times, only the clinical course distinguishes pancreatitis from cholecystitis.

Treatment

Medical therapy Patients with a clinical diagnosis of acute cholecystitis should not be fed and should be given intravenous fluids and electrolytes. It is usually necessary to give a narcotic analgesic such as morphine or meperidine to alleviate severe pain. Febrile patients who have leukocytosis or bandemia (elevated circulating band forms) should be given a broad-spectrum antibiotic, such as a third-generation cephalosporin or, for broader coverage against *Enterococcus,* ampicillin-sulbactam or piperacillin-tazobactam. Nasogastric tube decompression may be required in patients who present with vomiting or with evidence of an ileus. The usual course is one of gradual improvement for several days. Persistence of severe symptoms may indicate pericholecystic abscess or perforation.

Surgery In acute cholecystitis, laparoscopic cholecystectomy should be performed within 96 hours of onset of symptoms because the increasing inflammatory changes that occur over time have been implicated in bile duct injury; these changes may necessitate converting the procedure to an open cholecystectomy.[19,20] Early laparoscopic cholecystectomy is recommended for acute cholecystitis, because a delay in surgery does not reduce morbidity, mortality, rate of conversion to open surgery, or mean hospital stay.[21,22] In skilled hands, the laparoscopic procedure carries approximately the same risk as that of open cholecystectomy, but it is associated with much less postoperative pain and a shorter convalescence.[19] In patients with cirrhosis, laparoscopic cholecystectomy is performed for more emergent reasons and is associated with higher morbidity; however, the laparoscopic approach offers advantages of less blood loss, shorter operative time, and shorter length of hospitalization in patients with compensated cirrhosis.[23] In addition, laparoscopic cholecystectomy can be safely performed during pregnancy.[24] Laparoscopic cholecystectomy is less expensive than minilaparotomy or open cholecystectomy in high-volume surgery.[25]

In the United States, approximately 75% of all cholecystectomies are performed laparoscopically; in 5% to 10% of these patients, the procedure has to be converted to open cholecystectomy.[26] The complication rate of laparoscopic cholecystectomy is less than 5%, which is comparable to the rate reported for conventional cholecystectomy. Complications of laparoscopic cholecystectomy (bleeding and injury to the common bile duct, vasculature, and bowel) are more common when the surgeon is inexperienced.[27] Although mortality appears to be lower for laparoscopic cholecystectomy than for open cholecystectomy, the total number of cholecystectomy-related deaths has not decreased over the years, because more procedures are being performed.[28-30] This suggests that the benefits of laparoscopic cholecystectomy have expanded the indications for cholecystectomy.

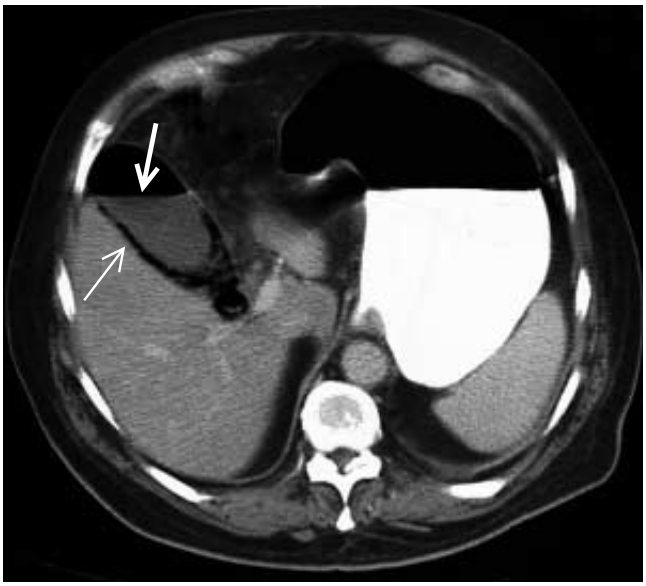


Figure 3 **CT scan from an elderly man with nausea and abdominal pain following an acute myocardial infarction, angioplasty, and stent placement. The scan shows air in the wall of the gallbladder and an air-fluid level within the gallbladder, diagnostic of emphysematous cholecystitis. The patient was treated with antibiotics and a percutaneous cholecystostomy tube; laparoscopic cholecystectomy was planned in a few months.**

Cholangiography can be performed during laparoscopic biliary surgery. However, because 10% to 15% of patients with acute cholecystitis have common duct stones, the physician should consider preoperative ERCP in patients with suspected choledocholithiasis (e.g., patients with jaundice, cholangitis, or a dilated common bile duct, as seen on ultrasonography).[31] Common duct stones can be removed endoscopically. If endoscopic common duct stone removal is not possible, the operative procedure of choice is cholecystectomy, either open or laparoscopic, for common bile duct exploration and stone removal.

Some patients (e.g., patients with septic shock, peritonitis, severe pancreatitis, portal hypertension, or marked clotting disorders) are not candidates for laparoscopic cholecystectomy. These patients should generally undergo either open cholecystectomy, if their condition permits, or simple cholecystostomy. Cholecystostomy, either operative or percutaneous under ultrasound guidance, involves extracting the stones and draining the biliary tree through a catheter left in the gallbladder. Percutaneous cholecystostomy is superior to gallbladder aspiration in severe acute cholecystitis.[32] Cholangiography can be carried out later through this drainage catheter. For patients who respond to cho-

lecystostomy and improve enough to become candidates for elective surgery, interval cholecystectomy is recommended, because the risk of recurrent symptoms is significant.[20]

Surgery is contraindicated for some patients because of the presence of other serious medical problems. In these cases, conservative medical therapy, including the use of antibiotics, may be the only possible approach for the acute attack.

Complications

The major complications of acute cholecystitis are related to severe inflammation and necrosis of gallbladder tissue.[33] Jaundice in the absence of choledocholithiasis can be noted in 15% of patients with acute cholecystitis; the stone impacted in the cystic duct results in edema and swelling, leading to extrinsic compression of the common hepatic duct, the common bile duct, or both [*see* Mirizzi Syndrome, *below*].[34]

Localized perforation and abscess Localized perforation and abscess formation are commonly found in patients who have severe symptoms that persist for many days. Such patients usually show localized right upper quadrant tenderness and rebound pain. In patients who have a delayed or subacute presentation, perforation typically occurs in the gallbladder fundus, as a consequence of ischemia that leads to gangrene and necrosis.[35] In patients who present with acute symptoms, the clinical differentiation between uncomplicated cholecystitis and gallbladder perforation is difficult; in this setting, ultrasonography should be the initial diagnostic modality.[36] These patients often have diabetes or other immuncompromising conditions; therefore, acute symptoms in diabetic or immunocompromised patients may heighten the suspicion for perforation. Free perforation extending into the peritoneum occurs in 2% to 10% of patients with acute cholecystitis; it is associated with peritonitis and a mortality of 10% to 30%.[37]

Empyema Empyema of the gallbladder occurs in 2% to 3% of patients with acute cholecystitis.[38] Typically, abdominal pain is severe and lasts for more than 7 days. The physical examination is not distinctive. Mortality approaches 25%; death often occurs as a result of septicemia.

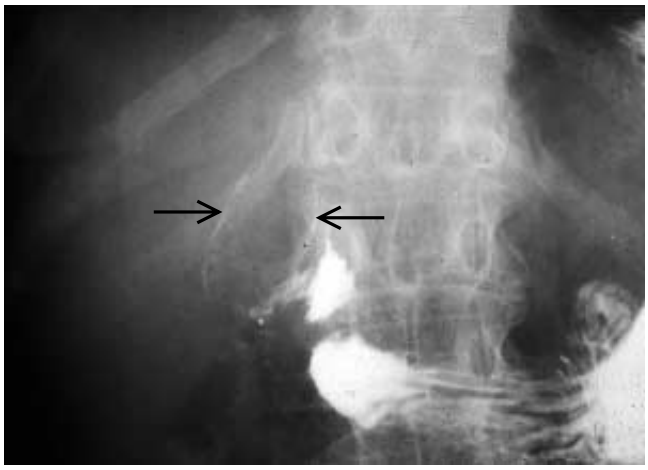


Figure 4 **Radiograph of the right upper abdominal quadrant during upper GI barium study showing a calcified wall of the gallbladder ("porcelain" gallbladder), indicating chronic cholecystitis and a high risk of gallbladder cancer.**

Emphysematous cholecystitis Emphysematous cholecystitis, which has a higher morbidity than uncomplicated acute cholecystitis, is usually caused by gas-forming bacteria, such as *Clostridium perfringens* and other clostridia, *Escherichia coli,* or anaerobic streptococci. Patients who have such infections are often very ill, and up to 50% also have diabetes.[35] Emphysematous cholecystitis occurs three times more often in men than in women.[39] Many cases of this type of cholecystitis are not associated with cholelithiasis. An ultrasound or computed tomography scan frequently reveals gas within the gallbladder; however, one study indicated that a plain abdominal x-ray is not sensitive as a diagnostic study in emphysematous cholecystitis [*see Figure 3*].[39]

Cholecystenteric fistula Another possible complication of acute cholecystitis is a cholecystenteric fistula, in which the gallbladder is connected either to the duodenum or to the hepatic flexure of the colon. In rare cases, the gallbladder communicates directly with the stomach or jejunum. A large gallstone (> 2.5 cm in diameter) will erode through the gallbladder wall into the duodenum. Subsequently, the stone may become impacted at the terminal ileum, causing small bowel obstruction, or in the duodenal bulb, resulting in gastric outlet obstruction (Bouveret syndrome). A cholecystenteric fistula is suspected when a plain abdominal x-ray shows pneumobilia, an ectopic stone, and mechanical obstruction. A barium upper gastrointestinal series can delineate a fistulous tract between the gallbladder and the intestines. CT scanning can define the site of obstruction and visualize the cholecystenteric fistula in 11% of patients.[40] Treatment of cholecystenteric fistula usually consists of one-stage cholecystectomy, exploration of the common bile duct, closure of the fistula, and extraction of the impacted stone.

Gallstones and malignancy Gallstones are present in 65% to 90% of patients with gallbladder cancer, although it is not clear whether gallstones themselves are the causal factor in oncogenesis.[41] A palpable gallbladder is usually found in malignant obstruction of the common bile duct (Courvoisier's law); however, this sign is uncommon in cases in which obstruction is caused by gallstones.

CHRONIC CHOLECYSTITIS

Chronic cholelithiasis is usually accompanied by evidence of chronic cholecystitis. The wall of the gallbladder is often thickened, fibrotic, and rigid, and the gallbladder is thus prevented from contracting and expanding normally. This condition may arise from a series of attacks of acute cholecystitis, from chronic mechanical irritation by calculi, or from both. The gallbladder wall may calcify and appear as the so-called porcelain gallbladder on plain abdominal x-ray [*see Figure 4*].

Diagnosis

Clinical manifestations It is difficult to attribute any symptom to chronic cholecystitis per se. Complaints of flatulence, heartburn, and nonspecific postprandial distress are common in patients with chronic cholecystitis, but such symptoms are also common in patients with no evidence of gallbladder disease. It is possible, however, to elicit a history of discrete attacks of abdominal pain resembling those of acute cholecystitis.

Physical examination Findings on physical examination are usually normal unless the patient is experiencing an acute attack of cholecystitis. The gallbladder is rarely palpable be-

cause scarring associated with chronic cholecystitis prevents expansion.

Laboratory evaluation Results of routine laboratory tests are usually normal; occasionally, the serum alkaline phosphatase level is modestly elevated.

Imaging studies TUS is the procedure of choice for the diagnosis of chronic gallbladder disease. In 90% to 95% of cases of cholelithiasis, ultrasonography demonstrates the echo of the calculus and the acoustic shadow behind the calculus [*see Figure 5*]. When the ultrasound is nondiagnostic, oral cholecystography may still be used to evaluate a patient with suspected gallbladder disease. If a double dose of the oral contrast agent fails to cause gallbladder opacification, cholelithiasis and chronic cholecystitis are almost certainly present. Cholescintigraphy is not helpful in diagnosing chronic cholelithiasis or chronic cholecystitis [*see* Acute Cholecystitis, *above*].

ERCP may reveal gallstones in the gallbladder of patients who have biliary tract pain and whose oral cholecystograms and gallbladder sonograms are normal. In one study, small gallstones were found with ERCP in 29 of 206 such patients (14%); the presence of these stones was confirmed during surgery.[42] CT or magnetic resonance imaging may also detect gallstones, but these techniques are unlikely to demonstrate stones not detected by ultrasonography. Endoscopic ultrasonography (EUS), with or without duodenal bile aspiration, may be a promising diagnostic approach in patients who have typical biliary symptoms but normal findings on transabdominal ultrasound.[43-45]

Bile collection and examination Gallbladder bile can be obtained by nasogastric tube or by endoscopic aspiration of the duodenum after infusing cholecystokinin intravenously to promote gallbladder emptying. In patients with biliary sludge, examination of bile under light and polarizing microscopy can show cholesterol crystals, which appear as rhomboid plates with a notch in one corner. Bile collected from the common bile duct during ERCP rarely contains precipitate, because hepatic bile transits rapidly through the biliary system,[10] and injection of contrast can lead to false positive results.[18] In one small study, EUS-guided aspiration of gallbladder bile was complicated by bile peritonitis in two of three patients.[46] Although some experts consider bile microscopy to be a gold standard in the diagnosis of biliary sludge, crystal analysis is limited by the need for invasive evaluation, meticulous sample processing, and institutional expertise.

Treatment

Surgery Elective cholecystectomy is indicated for patients who have symptomatic gallstones and chronic cholecystitis. Recurrent pain is to be expected in these patients if cholecystectomy is not performed. As many as 50% of patients with symptomatic gallstones who do not undergo cholecystectomy experience serious complications within 20 years after initial onset of symptoms.[47]

It is occasionally difficult to determine whether abdominal symptoms are secondary to documented gallbladder disease. A history of typical recurrent pain makes this determination easier. In certain cases, elective cholecystectomy is performed as a last diagnostic procedure when a thorough search for other causes of abdominal symptoms has proved negative. All too often, the symptoms recur postoperatively.

Dissolution therapy Oral bile acids, such as ursodeoxycholic acid (8 to 12 mg/kg daily) and chenodeoxycholic acid (13 to 15 mg/kg daily), can decrease biliary cholesterol levels; and when administered for months to years, ursodeoxycholic acid and chenodoxycholic acid can result in complete gallstone dissolution in 30% and 14% of patients, respectively.[48] A randomized, controlled trial found that combination therapy using these two agents was not superior to monotherapy with ursodeoxycholic acid.[49] Chenodeoxycholic acid has largely been replaced by the safer ursodeoxycholic acid; however, these drugs are effective only in patients with small cholesterol stones and a functioning gallbladder. A high rate of gallstone recurrence is noted after cessation of therapy. Infusing methyl *tert*-butyl ether through a transhepatic catheter directly into the gallbladder can rapidly dissolve cholesterol stones.[50] The rapid infusion and removal of this ether, which remains liquid at body temperature, results in the dissolution of most cholesterol gallstones within 4 to 31 hours. Dissolution therapy has limited value, except in patients who are poor candidates for surgery.

Extracorporeal biliary lithotripsy Stones in the gallbladder or common bile duct have been successfully fragmented using extracorporeal shock wave lithotripsy (ESWL), a technique widely employed for the nonsurgical fragmentation of kidney stones. Patients undergoing biliary ESWL are carefully positioned and monitored so that the shock waves are targeted at the gallstones. The highest success rates of biliary ESWL are seen in patients with a radiolucent solitary gallstone less than 2 cm in diameter; 60% to 84% of these patients are free of stones after 6 to 12 months of therapy.[51] ESWL has been associated with low rates of adverse events such as pancreatitis, biliary pain, hepatic hematoma, and hematuria. The administration of oral ursodeoxycholic acid after fragmentation of stones has been associated with an increase in the percentage of patients who are free of gallbladder stones 6 months after ESWL.[52,53] In one study, 21% of patients who received 10 to 12 mg/kg of ursodeoxycholic acid daily for 6 months after ESWL were free of gallbladder stones at the end of the treatment period; in contrast, only 9% of patients who received placebo for 6 months were free of stones.[53] Stone

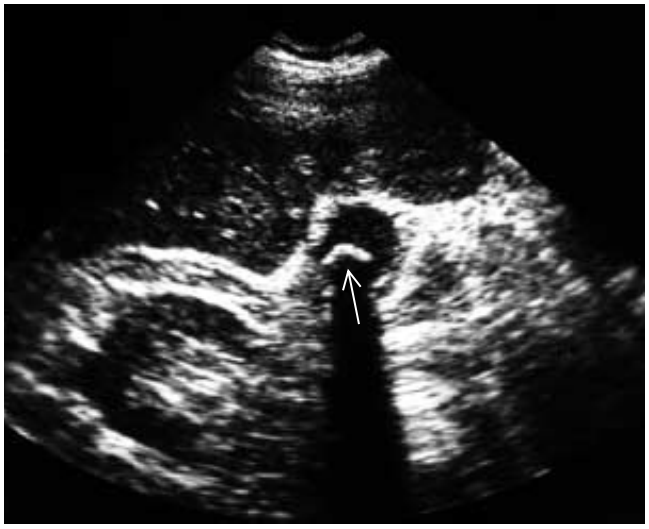


Figure 5 **Ultrasound of the gallbladder showing, in the center of the image, a stone within the gallbladder with a triangular area of acoustic attenuation ("shadowing") behind the gallstone.**

fragments in the common bile duct may pass spontaneously after endoscopic sphincterotomy or can be extracted with a basket. The usefulness of biliary ESWL is limited by its high rate of gallstone recurrence and the widespread availability of laparoscopic cholecystectomy.

ASYMPTOMATIC CHOLELITHIASIS

Most gallstones are asymptomatic (silent gallstones). In one prospective study, gallstones were present or there was evidence of cholecystectomy in 291 of 1,701 persons (17%) at the time of postmortem examination.[54] Of these 291 persons, only 31 had undergone cholecystectomy, presumably because of symptomatic disease. Ten deaths were directly attributable to the gallstones; four of these deaths occurred after cholecystectomy.

Natural History

Silent gallstones seldom lead to problems. In a long-term follow-up study of patients with asymptomatic gallstones, the cumulative risk of the development of symptoms was 10% at 5 years, 15% at 10 years, and 18% at 15 years or later.[55] Nineteen percent of patients who experienced symptoms (2.5% of the patients enrolled in the study) subsequently developed acute cholecystitis or pancreatitis. No patients died of gallbladder disease during a mean follow-up period of more than 10 years.

Diagnosis

Asymptomatic gallstones are usually identified incidentally on transabdominal or pelvic ultrasonography performed for other diagnostic purposes, such as the evaluation of gynecologic symptoms or findings on physical examination.

Treatment

Patients who have asymptomatic gallstones should generally be managed conservatively without surgery. Exceptions may be made for patients at increased risk for gallbladder cancer, such as Pima Indians, patients with calcified gallbladders (porcelain gallbladder), patients with very large gallstones (> 3 cm), and patients with an associated gallbladder polyp greater than 10 mm in diameter.[56]

In the past, prophylactic cholecystectomy was recommended for diabetic patients who had asymptomatic gallstones; anecdotal reports suggested that such patients did poorly when cholecystectomy was performed as an emergency procedure. However, two well-controlled, retrospective studies of patients undergoing surgery for acute cholecystitis and a decision analysis showed that diabetes was not an independent risk factor of operative mortality or serious postoperative complications, and prophylactic cholecystectomy resulted in a shortened life span.[57,58] Thus, prophylactic cholecystectomy cannot be recommended for patients with diabetes.

CHOLEDOCHOLITHIASIS

Choledocholithiasis, a condition in which a stone lodges in the common bile duct after passage from the gallbladder through the cystic duct, develops secondary to chronic cholelithiasis in 15% to 20% of patients.[9,59] Primary common bile duct stones are more commonly seen in Asian populations than in populations of the Western world. This increased incidence of primary common bile duct stones is attributed to the increased prevalence of flukes and parasitic infections (e.g., clonorchiasis, fascioliasis, and ascariasis) in Asia, because of the prevalent use of uncooked seafood in the diet. Other risk factors for choledocholithiasis include the presence of periampullary diverticula and advancing age.[60]

Diagnosis

Clinical manifestations The signs and symptoms associated with choledocholithiasis vary. Some patients have no symptoms, whereas others may present with an acute illness. Pain is a common feature and is often located in the right upper quadrant or midepigastrium, with radiation of the pain to the interscapular region. Pain may be associated with nausea, vomiting, or both; it can be indistinguishable from biliary colic. Cholangitis may present as the Charcot triad (fever, pain, and jaundice) or the Reynold pentad (Charcot triad of symptoms, hypotension, and a change in mental status). Patients may also present with pancreatitis.

Physical examination Vital signs may reveal an elevated temperature. In more acutely ill patients, hypotension and tachycardia may occur. Physical exam may reveal tenderness and guarding in the right upper quadrant and midepigastrium. Hepatomegaly may be found when common bile duct obstruction has been present for some time. Scleral icterus may also be seen.

Laboratory evaluation Both serum bilirubin and alkaline phosphatase levels can be markedly elevated. However, when the stones do not obstruct the duct, the serum bilirubin level may be only slightly elevated or may be normal, and the alkaline phosphatase level may be substantially elevated. Typically, serum aminotransferase levels are only modestly elevated. It would be unusual to see aminotransferase levels higher than 1,000 IU/L. In some instances, aminotransferase levels rise and fall rapidly early in the course of bile duct obstruction.

Imaging studies TUS may detect only 50% of common bile duct stones[61]; however, it can often detect dilatation of common bile duct and intrahepatic ducts. The sensitivity of TUS for detecting common duct stones increases to 76% when ductal dilatation of more than 6 mm is used as the primary end point for choledocholithiasis. CT is no more sensitive or specific than TUS. Cholescintigraphy may show common bile duct obstruction, particularly when symptoms are of recent onset, but not all common bile duct stones will cause complete bile duct obstruction. Magnetic resonance cholangiopancreatography (MRCP) and EUS have similar accuracies in detecting common bile duct stones. MRCP [*see Figure 6*] is noninvasive and may be preferred in cases where the suspicion of choledocholithiasis is mild to moderate.[62,63]

ERCP allows radiographic visualization of the biliary tree [*see Figure 7*] and the option of therapeutic intervention.[64,65] EUS has greater sensitivity and specificity than ERCP in the detection of common bile duct stones but lacks the therapeutic option available with ERCP.[66] Therefore, ERCP is the technique of choice if common bile duct stones are highly suspected on the basis of the history, physical examination findings, and laboratory and imaging studies. When ERCP is unavailable or is unsuccessful in detecting bile duct stones, percutaneous transhepatic cholangiography (PTC) allows for direct imaging of bile ducts and offers the potential for therapeutic intervention. PTC involves accessing the bile ducts via a small needle.[67] The success rate of PTC in patients with dilated ducts is close to 100%; nondilated ducts are entered successfully about 70% of the time. Complica-

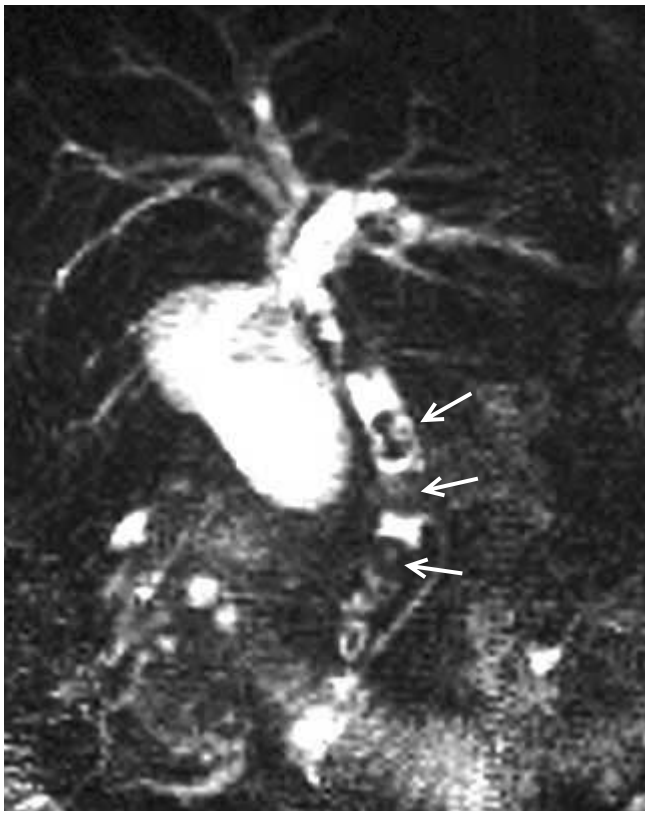


Figure 6 **This magnetic resonance cholangiopancreatogram shows multiple gallstones (arrows) in the common bile duct (choledocholithiasis).**

tion rates for both ERCP and PTC approach 5%. ERCP has replaced PTC as the technique of choice.

Treatment

Endoscopic sphincterotomy is the initial treatment for the patient with choledocholithiasis. In one large study, sphincterotomy was successful in 97.5% of patients with common bile duct stones, although more than one attempt was necessary in some patients. The overall rate of clearance of bile duct stones was 84.5%. The remaining patients required either surgery or permanent placement of a biliary endoprosthesis. The overall complication rate was 6.9%, and the complications included bleeding, cholangitis, pancreatitis, and perforation. The 30-day procedure-related mortality was 0.6%.[68] Follow-up studies have shown a low rate of recurrence of biliary duct problems and a low incidence of papillary stenosis.[69] Operative exploration of the common duct should be reserved for the few patients in whom endoscopic sphincterotomy is unsuccessful. Laparoscopic removal of biliary stones may be an alternative to preoperative ERCP.[70,71]

Endoscopic sphincterotomy is also the treatment of choice for patients with retained bile duct stones after gallbladder or biliary tract surgery. If sphincterotomy fails and if the patient has a T tube in place, instrumental extraction through the mature T-tube tract may be successful. Surgical exploration of the biliary tree is indicated if nonsurgical treatments fail.

MIRIZZI SYNDROME

Mirizzi syndrome refers to an obstruction of the common hepatic duct caused by a stone impacted at the neck of the gallbladder or the cystic duct. Mirizzi syndrome is classified into type I and type II.[72] In Mirizzi syndrome type I, there is only an extrinsic compression of the common hepatic duct by the gallstone and accompanying inflammation. In Mirizzi syndrome type II, a cholecystocholedochal fistula is established by the mechanism of pressure-induced necrosis from the gallstone.

Diagnosis

The clinical presentation of individuals with Mirizzi syndrome varies greatly.[73] Obstructive jaundice is commonly seen. However, 20% to 40% of patients may present without jaundice or have normal serum aminotransferase levels.[74] Biliary imaging tests often fail to demonstrate the features of Mirizzi syndrome; therefore, successful management of patients with Mirizzi syndrome is a challenge and relies heavily upon clinical suspicion and early recognition by the treating physician.

Treatment

Nonsurgical treatment of Mirizzi syndrome is limited and suboptimal. Long-term biliary stenting has a relatively high incidence of complications, including cholangitis and secondary biliary cirrhosis.[73] Nonsurgical lithotripsy and stone removal is restricted to patients with Mirizzi syndrome type II.[75] In Mirizzi syndrome type I, the offending stones are not accessible for clearance via bile ducts. Cholecystectomy is the treatment of choice. If the gallbladder is not removed, patients with Mirizzi syndrome are left at significant risk for complications from continued gallstone disease, including acute cholangitis, cholecystitis, suppurative cholangitis, liver abscess, secondary biliary cholangitis, and, perhaps, gallbladder carcinoma.[73,75,76] Nonsurgical treatment of patients with Mirizzi syndrome should be limited to those patients who are unfit for surgery or who have a shortened life expectancy.[34]

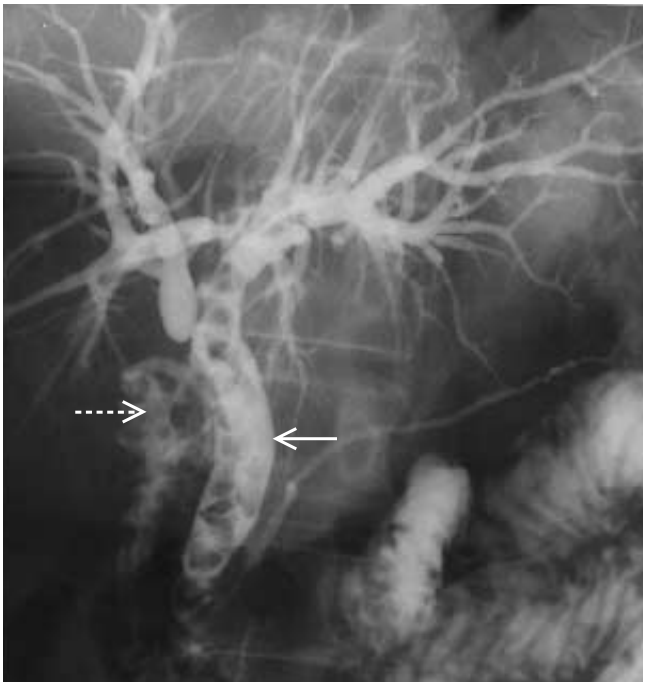


Figure 7 **Endoscopic retrograde cholangiopancreatography reveals abnormalities in a patient with gallstones. Multiple radiolucent areas establish the diagnosis of stones in the gallbladder (broken arrow) and common bile duct (solid arrow).**

Chronic Biliary Tract Disease

Chronic inflammation of biliary ducts is usually caused by partial or complete obstruction of the biliary tree. Some patients with chronic cholelithiasis or other chronic diseases of the biliary ducts will experience associated chronic inflammation or stricturing of the biliary tree.

DIAGNOSTIC OVERVIEW

Clinical Manifestations

Patients with chronic inflammation of the biliary tree may complain of fatigue, intermittent fever and chills, anorexia, pruritus, and weight loss. The physical examination may be fairly unremarkable; jaundice, excoriations of the skin related to marked pruritus, and stigmata of chronic liver disease may raise the level of suspicion.

Laboratory Evaluation

Laboratory tests will often reveal chronically elevated serum alkaline phosphatase levels and increased levels of serum 5´-nucleotidase, leucine aminopeptidase, and γ-glutamyl transpeptidase (GGT). Transient elevations of the total bilirubin level may also been seen.

Imaging Studies

Direct visualization of the biliary tree is important in determining whether the symptoms and signs result from an anatomic defect that can be corrected by endoscopic therapy or surgery. Use of MRCP or ERCP usually leads to identification of the obstructive site.

SPECIFIC PRESENTATIONS

Common Bile Duct Stricture

Benign and malignant strictures are similar in appearance, as imaged by ERCP or MRCP. Epithelial samples of biliary strictures for evaluation can be obtained by brush cytology; fine-needle aspiration; endoscopic pinch biopsy; or a combination of the three. The sensitivity of brush cytology is as high as 70% for the diagnosis of a malignant stricture of the bile duct; specificity is as high as 100%.[77] Simple bile duct aspiration alone is not as reliable. Common bile duct stricture, which may result from biliary tract surgery, can be treated endoscopically with balloon dilatation or with the placement of an endoprosthesis. If these treatments are unsuccessful, surgical intervention may prove beneficial for selected patients.

Primary Sclerosing Cholangitis

Primary sclerosing cholangitis (PSC) is a disease of unknown etiology that is characterized by an irregular inflammatory fibrosis of both the intrahepatic and extrahepatic bile ducts [*see Figure 8*].[78] It usually occurs in men between 20 and 50 years of age.

Patients may present with jaundice, pruritus, nonspecific pain, fever, and weight loss. Approximately 75% of patients will have chronic ulcerative colitis. Liver function tests show cholestatic abnormalities.[78]

Ursodeoxycholic acid therapy will usually result in an improvement in the biochemical markers of cholestasis, but it has not been shown to increase survival.[79] Endoscopic treatment of significant ductal strictures may also improve biochemical markers of cholestasis and reduce the number of episodes of cholangitis.[80] A combined approach using therapeutic stricture dilatation and ursodeoxycholic acid therapy may benefit a select group of patients.

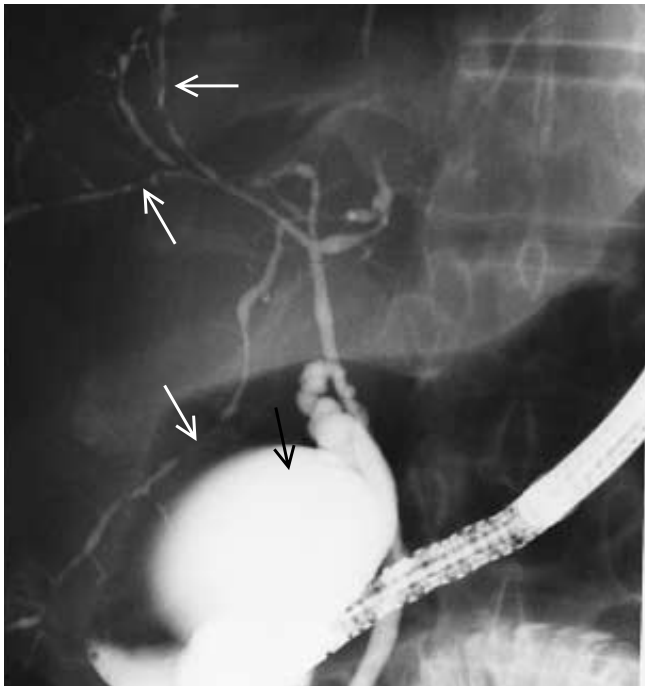


Figure 8 **This cholangiogram, obtained during endoscopic retrograde cholangiopancreatography, shows a normal gallbladder (black arrow) and a narrowed biliary tree with many areas of segmental stenosis (white arrows), diagnostic of primary sclerosing cholangitis.**

Patients with PSC are at increased risk for biliary tract cancer. The incidence of cholangiocarcinoma in patients with PSC is as high as 30%, and there is an increased risk of gallbladder and pancreatic cancer.[81] A substantial number of patients with PSC may have undetected cholangiocarcinoma at the time of liver transplantation.

Recurrent Pyogenic Cholangitis

Recurrent pyogenic cholangitis (RPC), as its name suggests, is a condition characterized by recurrent bouts of inflammation of the bile ducts. It most commonly affects patients of Asian descent. The exact etiology of RPC is unclear. Some experts propose a dietary or infectious cause. The parasites *Opisthorchis sinensis* and *Ascaris lumbricoides* are commonly found in the stools of affected patients.[82]

Patients typically present with repeated attacks of fever, chills, abdominal pain, and jaundice. Laboratory tests usually demonstrate an elevation in serum bilirubin and alkaline phosphatase levels. Elevations in serum aminotransferase levels and the prothrombin time signify hepatocyte injury, although the prothrombin time may also be prolonged because of vitamin K malabsorption.[83]

Imaging studies, such as ultrasonography, may be somewhat confusing in patients with RPC, because there may be areas of intrahepatic biliary dilatation without common bile duct dilatation. Evaluation using CT or MRCP usually defines the areas of intrahepatic and extrahepatic biliary dilatation more clearly than does ultrasonography, and these techniques also provide three-dimensional information.[84] ERCP is often required to confirm areas of stricture and dilatation. ERCP also allows for possible

therapeutic intervention. PTC provides access to peripheral ducts that may be inaccessible by ERCP.[85]

Treatment usually consists of antibiotic therapy and endoscopic or surgical stone clearance to improve biliary drainage.

Choledochal Cyst

Biliary cystic disease includes choledochal cyst disease and the less common gallbladder cysts and cystic duct cysts.[86] Choledochal cyst is an ectasia of the common bile duct that may present in late childhood or in adult life as obstructive jaundice. The cause of the disorder is not fully defined, and both congenital and acquired etiologies are postulated.[87]

Diagnosis Clinical manifestations of choledochal cyst in children include abdominal pain, cholangitis, and an abdominal mass. A palpable mass is unusual in adults, because adults tend to present with recurrent cholangitis, pancreatitis, or, rarely, portal hypertension. Choledochal cysts may involve any segment of the bile duct and are categorized according to the classification proposed by Todani and colleagues [*see Table 2*].[88] An abnormal pancreatobiliary duct junction is more common in patients with choledochal cysts and could expose the bile ducts to pancreatic juices, which could result in progressive injury to the ductal system. Type I cysts are the most common, accounting for 40% to 60% of all cases, followed by type IV. Types II, III, and V are rare.

A combination of imaging studies may establish the diagnosis. Ultrasonography may delineate the cyst and intrahepatic portions of the disease. CT and MRCP may provide useful information in regard to the extent of disease and the potential for malignancy. ERCP, PTC, and intraoperative cholangiography are important for diagnostic evaluation and surgical planning.

Table 2 Modified Classification System for Choledochal Cysts and Surgical Procedure of Choice

Classification	*Type*	*Procedure of Choice*
Type IA	Choledochal cyst	Roux-en-Y hepaticojejunostomy
Type IB	Segmented choledochal dilatation	
Type IC	Diffuse or cylindrical duct dilatation	
Type II	Extrahepatic duct diverticulum	Excision of diverticulum
Type III	Choledochocele	Endoscopic sphincterotomy
Type IVA	Multiple intrahepatic and extrahepatic duct cysts	Roux-en-Y hepaticojejunostomy
Type IVB	Multiple extrahepatic duct cysts	
Type V	Intrahepatic duct cysts (Caroli disease and Caroli syndrome)	Hepatic resection, liver transplantation

*All patients with choledochal cysts must undergo cholecystectomy to decrease the risk of malignancy, with the possible exception of patients with type III cysts. For the much rarer gallbladder and cystic duct cysts, treatment is cholecystectomy.

Table 3 Clinical Classification System for Biliary-Specific Abdominal Pain Associated with SOD*

Criteria

A. Typical biliary-type pain
B. Elevated liver enzyme levels (AST, alkaline phosphatase, or both more than two times normal on at least two occasions)
C. Delayed drainage of contrast injection during ERCP (> 45 min)
D. Dilated common bile duct (> 12 mm)

Classification Based on above Criteria

Biliary type I: criteria A through D are present; SOD is present in 80%–90% of patients
Biliary type II : criterion A plus one or two other criteria are present; SOD is present in 50% of patients
Biliary type III: only criterion A is present; SOD is uncommon

*A similar classification for SOD and pancreatic-type abdominal pain exists but is not included in this table.
AST — aspartate aminotransferase ERCP — endoscopic retrograde cholangiopancreatography SOD — sphincter of Oddi dysfunction

Treatment The initial treatment of choledochal cysts depends on the age of the patient, the presentation, and the type of the cyst. In terms of definitive treatment, pharmacologic or endoscopic management offers little benefit in that these forms of therapy do not address the well-described malignant potential of bile duct cysts.[89] Therefore, the primary role of endoscopic procedures is in the initial evaluation and diagnosis of bile duct cysts. However, endoscopic interventions such as lithotripsy, stone extraction, and laser ablation have proved successful in the treatment of intrahepatic and extrahepatic biliary stones in patients with Caroli disease, a congenital disorder associated with renal cystic disease of varying severity.[90] Endoscopic therapy, such as stone extraction, can be a definitive treatment for patients with recurrent pyogenic cholangitis. It would be the chosen therapy in elderly patients or in patients considered to be poor candidates for surgery. The current standard for surgical treatment in the patient who is a reasonable surgical risk is excision of the cyst with free biliary drainage into the gastrointestinal tract. The classical surgical reconstruction is a hepaticojejunostomy with a Roux-en-Y [*see Table 2*] reconstruction.[91]

Sphincter of Oddi Dysfunction

Sphincter of Oddi dysfunction (SOD) is a benign condition of intermittent or permanent obstruction of biliary drainage, pancreatic drainage, or both that is caused either by a stenosis or by smooth muscle dysfunction of the sphincter muscle.[92] Biliary SOD is classified into three types on the basis of clinical parameters using the modified Milwaukee criteria [*see Table 3*].

Diagnosis Biliary SOD is usually seen in women in the fourth to sixth decades of life. The symptoms arise typically after cholecystectomy, although SOD may occur in patients with an intact gallbladder.[93,94] The clinical presentation of biliary SOD is episodic abdominal pain in the epigastric region or the right upper quadrant that may radiate to the back or shoulders. It may be associated with nausea or vomiting that worsens with eating. Laboratory tests may reveal elevated liver function. Right upper quadrant ultrasonography and CT may reveal a dilated common bile duct.

ERCP with sphincter of Oddi manometry is the gold stan-

dard for diagnosis of SOD. A basal sphincter pressure of more than 40 mm Hg is abnormal and indicative of SOD.[95] Other tests that are noninvaive and less reliable may also indicate the presence of SOD; such tests include a provocation test with morphine (or neostigmine), which produces biliary pain and elevation of the serum aminotransferase level; ultrasound evaluation of dilatation and emptying of the common bile duct after secretin stimulation; or the kinetics of ductal emptying studied by scintigraphy.[96]

Sphincter of Oddi manometry is not required to confirm the diagnosis of type I SOD disease. However, patients classified with type II disease should undergo sphincter of Oddi manometry because only 50% of patients in this group have SOD. In patients classified as having type II disease, only patients whose SOD is confirmed by sphincter of Oddi manometry should undergo endoscopic sphincterotomy. Sphincter of Oddi manometry, endoscopic sphincterotomy, or both have low efficacy in patients with type III disease.

Treatment A low-fat diet may decrease biliary or pancreatic stimulation, although the efficacy of this approach is unknown. Endoscopic sphincterotomy is the primary treatment for patients with SOD type I disease and for patients with types II and III disease in which the presence of SOD has been confirmed by manometry. Over 90% of patients with type I disease will have a favorable response to endoscopic sphincterotomy; therefore, manometry should not be performed in these patients.

Pharmacologic therapies (i.e., calcium channel blockers and nitrates) are primarily used in patients with type III disease, because in these patients, sphincter of Oddi manometry has the greatest risk of complication and the smallest diagnostic yield. Treatment with calcium channel blockers and nitrates decreases pain by relaxing the sphincter smooth muscle.[92]

Other endoscopic therapies such as balloon dilatation, injection of botulinum toxin, temporary stent placement, and surgical sphincteroplasty are not widely used in the treatment of SOD.[97-99]

Kimberly M. Persley, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Rajeev Jain, M.D., F.A.C.P., is an advisor or consultant for AstraZeneca Pharmaceuticals LP.

References

1. Sandler RS, Everhart JE, Donowitz M, et al: The burden of selected digestive diseases in the United States. Gastroenterology 122:1500, 2002

2. Jensen KH, Jorgensen T: Incidence of gallstones in a Danish population. Gastroenterology 100:790, 1991

3. Donovan JM: Physical and metabolic factors in gallstone pathogenesis. Gastroenterol Clin North Am 28:75, 1999

4. Ko CW, Sekijima JH, Lee SP: Biliary sludge. Ann Intern Med 130(4 pt 1):301, 1999

5. Everhart JE, Khare M, Hill M, et al: Prevalence and ethnic differences in gallbladder disease in the United States. Gastroenterology 117:632, 1999

6. Wang DQ, Afdhal NH: Genetic analysis of cholesterol gallstone formation: searching for Lith (gallstone) genes. Curr Gastroenterol Rep 6:140, 2004

7. Stewart L, Oesterle AL, Erdan I, et al: Pathogenesis of pigment gallstones in Western societies: the central role of bacteria. J Gastrointest Surg 6:891, 2002

8. Shaffer EA: Epidemiology and risk factors for gallstone disease: has the paradigm changed in the 21st century? Curr Gastroenterol Rep 7:132, 2005

9. Johnston DE, Kaplan MM: Pathogenesis and treatment of gallstones. N Engl J Med 328:412, 1993

10. Jain R: Biliary sludge: When should it ot be ignored? Curr Treat Options Gastroenterol 7:105, 2004

11. Ko CW, Beresford SA, Schulte SJ, et al: Incidence, natural history, and risk factors for biliary sludge and stones during pregnancy. Hepatology 41:359, 2005

12. Berger MY, van der Velden JJ, Lijmer JG, et al: Abdominal symptoms: do they predict gallstones? A systematic review. Scand J Gastroenterol 35:70, 2000

13. Owen CC, Jain R: Acute acalculous cholecystitis. Curr Treat Options Gastroenterol 8:99, 2005

14. Trowbridge RL, Rutkowski NK, Shojania KG: Does this patient have acute cholecystitis? JAMA 289:80, 2003

15. Edlund G, Kempi V, van der Linden W: Jaundice in acute cholecystitis without common duct stones. Acta Chir Scand 149:597, 1983

16. Kurzweil SM, Shapiro MJ, Andrus CH, et al: Hyperbilirubinemia without common bile duct abnormalities and hyperamylasemia without pancreatitis in patients with gallbladder disease. Arch Surg 129:829, 1994

17. Shea JA, Berlin JA, Escarce JJ, et al: Revised estimates of diagnostic test sensitivity and specificity in suspected biliary tract disease. Arch Intern Med 154:2573, 1994

18. Parasher VK, Romain K, Sukumar R, et al: Can ERCP contrast agents cause pseudomicrolithiasis? Their effect on the final outcome of bile analysis in patients with suspected microlithiasis. Gastrointest Endosc 51(4 pt 1):401, 2000

19. Zacks SL, Sandler RS, Rutledge R, et al: A population-based cohort study comparing laparoscopic cholecystectomy and open cholecystectomy. Am J Gastroenterol 97:334, 2002

20. Bhattacharya D, Ammori BJ: Contemporary minimally invasive approaches to the management of acute cholecystitis: a review and appraisal. Surg Laparosc Endosc Percutan Tech 15:1, 2005

21. Shikata S, Noguchi Y, Fukui T: Early versus delayed cholecystectomy for acute cholecystitis: a meta-analysis of randomized controlled trials. Surg Today 35:553, 2005

22. Papi C, Catarci M, D'Ambrosio L, et al: Timing of cholecystectomy for acute calculous cholecystitis: a meta-analysis. Am J Gastroenterol 99:147, 2004

23. Puggioni A, Wong LL: A metaanalysis of laparoscopic cholecystectomy in patients with cirrhosis. J Am Coll Surg 197:921, 2003

24. Barone JE, Bears S, Chen S, et al: Outcome study of cholecystectomy during pregnancy. Am J Surg 177:232, 1999

25. Nilsson E, Ros A, Rahmqvist M, et al: Cholecystectomy: costs and health-related quality of life: a comparison of two techniques. Int J Qual Health Care 16:473, 2004

26. Livingston EH, Rege RV: A nationwide study of conversion from laparoscopic to open cholecystectomy. Am J Surg 188:205, 2004

27. See WA, Fisher RJ, Winfield HN, et al: Laparoscopic surgical training: effectiveness and impact on urological surgical practice patterns. J Urol 149:1054, 1993

28. Legorreta AP, Silber JH, Costantino GN, et al: Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy. JAMA 270:1429, 1993

29. Steiner CA, Bass EB, Talamini MA, et al: Surgical rates and operative mortality for open and laparoscopic cholecystectomy in Maryland. N Engl J Med 330:403, 1994

30. Escarce JJ, Chen W, Schwartz JS: Falling cholecystectomy thresholds since the introduction of laparoscopic cholecystectomy. JAMA 273:1581, 1995

31. Tse F, Barkun JS, Barkun AN: The elective evaluation of patients with suspected choledocholithiasis undergoing laparoscopic cholecystectomy. Gastrointest Endosc 60: 437, 2004

32. Ito K, Fujita N, Noda Y, et al: Percutaneous cholecystostomy versus gallbladder aspiration for acute cholecystitis: a prospective randomized controlled trial. AJR Am J Roentgenol 183:193, 2004

33. Abou-Saif A, Al-Kawas FH: Complications of gallstone disease: Mirizzi syndrome, cholecystocholedochal fistula, and gallstone ileus. Am J Gastroenterol 97:249, 2002

34. Gomez G: Mirizzi syndrome. Curr Treat Options Gastroenterol 5:95, 2002

35. Bennett GL, Balthazar EJ: Ultrasound and CT evaluation of emergent gallbladder pathology. Radiol Clin North Am 41:1203, 2003

36. Sood BP, Kalra N, Gupta S, et al: Role of sonography in the diagnosis of gallbladder perforation. J Clin Ultrasound 30:270, 2002

37. Menakuru SR, Kaman L, Behera A, et al: Current management of gall bladder perforations. ANZ J Surg 74:843, 2004

38. Thornton JR, Heaton KW, Espiner HJ, et al: Empyema of the gall bladder: reappraisal of a neglected disease. Gut 24:1183, 1983

39. Gill K, Chapman A, Weston M: The changing face of emphysematous cholecystitis. Br J Radiol 70:986, 1997

40. Lassandro F, Gagliardi N, Scuderi M, et al: Gallstone ileus analysis of radiologic findings in 27 patients. Eur J Radiol 50:23, 2004

41. Misra S, Chaturvedi A, Misra NC, et al: Carcinoma of the gallbladder. Lancet Oncol 4:167, 2003

42. Venu RP, Geenen JE, Toouli J, et al: Endoscopic retrograde cholangiopancreatography. Diagnosis of cholelithiasis in patients with normal gallbladder x-ray and ultrasound studies. JAMA 249:758, 1983

43. Thorboll J, Vilmann P, Jacobsen B, et al: Endoscopic ultrasonography in detection of cholelithiasis in patients with biliary pain and negative transabdominal ultrasonography. Scand J Gastroenterol 39:267, 2004

44. Dill JE, Hill S, Callis J, et al: Combined endoscopic ultrasound and stimulated biliary drainage in cholecystitis and microlithiasis—diagnoses and outcomes. Endoscopy 27: 424, 1995

45. Dahan P, Andant C, Levy P, et al: Prospective evaluation of endoscopic ultrasonography and microscopic examination of duodenal bile in the diagnosis of cholecystolithiasis in 45 patients with normal conventional ultrasonography. Gut 38:277, 1996

46. Jacobson BC, Waxman I, Parmar K, et al: Endoscopic ultrasound-guided gallbladder bile aspiration in idiopathic pancreatitis carries a significant risk of bile peritonitis. Pancreatology 2:26, 2002

47. Wenckert A, Robertson B: The natural course of gallstone disease: eleven-year review of 781 nonoperated cases. Gastroenterology 50:376, 1966

48. Broughan TA: Gallstones. Curr Treat Options Gastroenterol 2:154, 1999

49. Petroni ML, Jazrawi RP, Pazzi P, et al: Ursodeoxycholic acid alone or with chenodeoxycholic acid for dissolution of cholesterol gallstones: a randomized multicentre trial. The British-Italian Gallstone Study group. Aliment Pharmacol Ther 15:123, 2001

50. Thistle JL, May GR, Bender CE, et al: Dissolution of cholesterol gallbladder stones by methyl *tert*-butyl ether administered by percutaneous transhepatic catheter. N Engl J Med 320:633, 1989

51. Paumgartner G, Sauter GH: Extracorporeal shock wave lithotripsy of gallstones: 20th anniversary of the first treatment. Eur J Gastroenterol Hepatol 17:525, 2005

52. Tsumita R, Sugiura N, Abe A, et al: Long-term evaluation of extracorporeal shock-wave lithotripsy for cholesterol gallstones. J Gastroenterol Hepatol 16:93, 2001

53. Schoenfield LJ, Berci G, Carnovale RL, et al: The effect of ursodiol on the efficacy and safety of extracorporeal shock-wave lithotripsy of gallstones. The Dornier National Biliary Lithotripsy Study. N Engl J Med 323:1239, 1990

54. Godrey PJ, Bates T, Harrison M, et al: Gall stones and mortality: a study of all gall stone related deaths in a single health district. Gut 25:1029, 1984

55. Gracie WA, Ransohoff DF: The natural history of silent gallstones: the innocent gallstone is not a myth. N Engl J Med 307:798, 1982

56. Guidelines for the Treatment of Gallstones. American College of Physicians. Ann Intern Med 119:620, 1993

57. Ransohoff DF, Miller GL, Forsythe SB, et al: Outcome of acute cholecystitis in patients with diabetes mellitus. Ann Intern Med 106:829, 1987

58. Friedman LS, Roberts MS, Brett AS, et al: Management of asymptomatic gallstones in the diabetic patient: a decision analysis. Ann Intern Med 109:913, 1988

59. Ahmed A, Cheung RC, Keeffe EB: Management of gallstones and their complications. Am Fam Physician 61:1673, 2000

60. Houdart R, Perniceni T, Darne B, et al: Predicting common bile duct lithiasis: determination and prospective validation of a model predicting low risk. Am J Surg 170:38, 1995

61. Lichtenbaum RA, McMullen HF, Newman RM: Preoperative abdominal ultrasound may be misleading in risk stratification for presence of common bile duct abnormalities. Surg Endosc 14:254, 2000

62. Kondo S, Isayama H, Akahane M, et al: Detection of common bile duct stones: comparison between endoscopic ultrasonography, magnetic resonance cholangiography, and helical-computed-tomographic cholangiography. Eur J Radiol 54:271, 2005

63. Aube C, Delorme B, Yzet T, et al: MR cholangiopancreatography versus endoscopic sonography in suspected common bile duct lithiasis: a prospective, comparative study. AJR Am J Roentgenol 184:55, 2005

64. Ramesh H: A balanced approach to choledocholithiasis. Surg Endosc 15:1494, 2001

65. NIH state-of-the-science statement on endoscopic retrograde cholangiopancreatography (ERCP) for diagnosis and therapy. NIH Consens State Sci Statements 19:1, 2002

66. Palazzo L, O'Toole D: EUS in common bile duct stones. Gastrointest Endosc 56(4 suppl):S49, 2002

67. Mueller PR, vanSonnenberg E, Simeone JF: Fine-needle transhepatic cholangiography: indications and usefulness. Ann Intern Med 97:567, 1982

68. Vaira D, D'Anna L, Ainley C, et al: Endoscopic sphincterotomy in 1000 consecutive patients. Lancet 2:431, 1989

69. Hawes RH, Cotton PB, Vallon AG: Follow-up 6 to 11 years after duodenoscopic sphincterotomy for stones in patients with prior cholecystectomy. Gastroenterology 98:1008, 1990

70. Hawasli A, Lloyd L, Cacucci B: Management of choledocholithiasis in the era of laparoscopic surgery. Am Surg 66:425, 2000

71. Ponsky JL, Heniford BT, Gersin K: Choledocholithiasis: evolving intraoperative strategies. Am Surg 66:262, 2000

72. Csendes A, Diaz JC, Burdiles P, et al: Mirizzi syndrome and cholecystobiliary fistula: a unifying classification. Br J Surg 76:1139, 1989

73. England RE, Martin DF: Endoscopic management of Mirizzi's syndrome. Gut 40:272, 1997

74. Curet MJ, Rosendale DE, Congilosi S: Mirizzi syndrome in a Native American population. Am J Surg 168:616, 1994

75. Tsuyuguchi T, Saisho H, Ishihara T, et al: Long-term follow-up after treatment of Mirizzi syndrome by peroral cholangioscopy. Gastrointest Endosc 52:639, 2000

76. Redaelli CA, Buchler MW, Schilling MK, et al: High coincidence of Mirizzi syndrome and gallbladder carcinoma. Surgery 121:58, 1997

77. Foutch PG, Kerr D, Harlan JR, et al: Endoscopic retrograde wire-guided brush cytology for diagnosis of patients with malignant obstruction of the bile duct. Am J Gastroenterol 85:791, 1990

78. Angulo P, Lindor KD: Primary sclerosing cholangitis. Hepatology 30:325, 1999

79. Lindor KD: Ursodiol for primary sclerosing cholangitis. Mayo Primary Sclerosing Cholangitis–Ursodeoxycholic Acid Study Group. N Engl J Med 336:691, 1997

80. Johnson GK, Geenen JE, Venu RP, et al: Endoscopic treatment of biliary duct strictures in sclerosing cholangitis: follow-up assessment of a new therapeutic approach. Gastrointest Endosc 33:9, 1987

81. Buckles DC, Lindor KD, Larusso NF, et al: In primary sclerosing cholangitis, gallbladder polyps are frequently malignant. Am J Gastroenterol 97:1138, 2002

82. Fan ST, Choi TK, Wong J: Recurrent pyogenic cholangitis: current management. World J Surg 15:248, 1991

83. Wilson MK, Stephen MS, Mathur M, et al: Recurrent pyogenic cholangitis or "oriental cholangiohepatitis" in occidentals: case reports of four patients. Aust N Z J Surg 66:649, 1996

84. Kim MJ, Cha SW, Mitchell DG, et al: MR imaging findings in recurrent pyogenic cholangitis. AJR Am J Roentgenol 173:1545, 1999

85. Jeyarajah DR: Recurrent pyogenic cholangitis. Curr Treat Options Gastroenterol 7:91, 2004

86. Liu CL, Fan ST, Lo CM, et al: Choledochal cysts in adults. Arch Surg 137:465, 2002

87. Cheney M, Rustad DG, Lilly JR: Choledochal cyst. World J Surg 9:244, 1985

88. Todani T, Watanabe Y, Narusue M, et al: Classification, operative procedures, and review of thirty-seven cases including cancer arising from choledochal cyst. Am J Surg 134:263, 1977

89. Postema RR, Hazebroek FW: Choledochal cysts in children: a review of 28 years of treatment in a Dutch children's hospital. Eur J Surg 165:1159, 1999

90. Shemesh E, Czerniak A, Klein E, et al: The role of endoscopic retrograde cholangiopancreatography in the diagnosis and treatment of adult choledochal cyst. Surg Gynecol Obstet 167:423, 1988

91. Terblanche J, Worthley CS, Spence RA, et al: High or low hepaticojejunostomy for bile duct strictures? Surgery 108:828, 1990

92. Menees S, Elta GH: Sphincter of Oddi dysfunction. Curr Treat Options Gastroenterol 8:109, 2005

93. Black NA, Thompson E, Sanderson CF: Symptoms and health status before and six weeks after open cholecystectomy: a European cohort study. ECHSS Group. European Collaborative Health Services Study Group. Gut 35:1301, 1994

94. Luman W, Adams WH, Nixon SN, et al: Incidence of persistent symptoms after laparoscopic cholecystectomy: a prospective study. Gut 39:863, 1996

95. Toouli J, Roberts-Thomson IC, Dent J, et al: Manometric disorders in patients with suspected sphincter of Oddi dysfunction. Gastroenterology 88(5 pt 1):1243, 1985

96. Pineau BC, Knapple WL, Spicer KM, et al: Cholecystokinin-stimulated mebrofenin (99mTc-Choletec) hepatobiliary scintigraphy in asymptomatic postcholecystectomy individuals: assessment of specificity, interobserver reliability, and reproducibility. Am J Gastroenterol 96:3106, 2001

97. Geenen JE, Hogan WJ, Dodds WJ, et al: The efficacy of endoscopic sphincterotomy after cholecystectomy in patients with sphincter-of-Oddi dysfunction. N Engl J Med 320:82, 1989

98. Wehrmann T, Schmitt TH, Arndt A, et al: Endoscopic injection of botulinum toxin in patients with recurrent acute pancreatitis due to pancreatic sphincter of Oddi dysfunction. Aliment Pharmacol Ther 14:1469, 2000

99. Moody FG, Becker JM, Potts JR: Transduodenal sphincteroplasty and transampullary septectomy for postcholecystectomy pain. Ann Surg 197:627, 1983

Acknowledgments

Figure 1 Courtesy of William E. Stevens, M.D.

Figure 2 Alan Iselin.

Figure 3 Courtesy of Laura Thomas, M.D., and Mark Feldman, M.D.

Figures 4 and 5 Courtesy of Mark Feldman, M.D.

Figure 6 Courtesy of David Riepe, M.D.

Figures 7 and 8 Courtesy of Malcolm F. Anderson, M.D.

VII ACUTE VIRAL HEPATITIS

Emmet B. Keeffe, m.d.

Most cases of acute hepatitis are caused by one of the hepatotrophic viruses, but drug-induced hepatitis and hepatitis that is secondary to other viruses may at times mimic typical acute viral hepatitis. Classic acute viral hepatitis is caused by one of five etiologic agents: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis D virus (HDV), or hepatitis E virus (HEV).[1,2] In the United States in 1999, 59% of reported cases of acute viral hepatitis were caused by HAV infection and 26.5% by HBV infection; 14.5% were classified as non-A, non-B hepatitis or were unspecified.[3] HDV infection (delta hepatitis) occurs either as a superinfection in chronic HBV carriers or as a coinfection during acute HBV infection. HEV infection occurs predominantly outside the United States, but a few cases have been reported in travelers returning to the United States. All five viruses can cause acute hepatitis, but only three—HBV, HCV, and HDV—can lead to chronic infection. A new hepatitis virus, hepatitis G virus (HGV), has been identified but does not appear to be pathogenic, nor does it account for those cases that have been termed non-A–E viral hepatitis, as some investigators originally suggested. Finally, a number of viruses that cause systemic illnesses may also affect the liver—for example, cytomegalovirus (CMV) and Epstein-Barr virus (EBV).

All five types of viral hepatitis are similar and cannot be distinguished reliably by clinical features or routine laboratory tests. Infection either may occur asymptomatically or may be associated with nonspecific flulike symptoms; some patients experience jaundice. The characteristic laboratory abnormality in acute hepatitis is an elevated aminotransferase level, typically greater than 300 IU/L and, occasionally, 1,000 to 3,000 IU/L. The specific etiology of viral hepatitis is determined by serologic testing [*see Table 1*].

Classification and Pathology

hepatitis a virus

HAV is a picornavirus similar to poliovirus and rhinovirus.[4] HAV was initially discovered in stool but has also been found in the serum of patients with acute HAV infection and in the cytoplasm of liver cells and bile of animals infected with HAV.[5] It is a nonenveloped, positive-stranded RNA virus that has at least seven genotypes but only one serotype.[4] The antigenic compositions of HAV throughout the world are remarkably similar, which explains the global efficacy of immune globulin and of hepatitis A vaccine. IgM antibody to HAV is detectable at the onset of clinical illness and usually disappears within 60 to 120 days. IgG antibody reaches a high titer during convalescence, persists indefinitely, and confers immunity.

hepatitis b virus

HBV, the only member of the family Hepadnaviridae that infects humans, has a diameter of 42 nm and consists of a 28 nm core surrounded by a protein coat; the core contains protein, circular double-stranded DNA, and DNA polymerase [*see Figure 1*].[6] Immunofluorescent antibody studies have detected HBV in the nuclei of infected liver cells. The core moves through the nuclear membrane into the cytoplasm, where it acquires its surface coat. HBV is found in the serum of almost all patients early in the course of acute HBV infection.

Two additional particles appear in the liver cell cytoplasm and serum of patients with HBV: a 22 nm–diameter sphere and a rod-shaped filament of the same diameter. These particles are found at the onset of jaundice in nearly all patients with acute HBV infection. The surface coat of the hepatitis B virion and the spheres and filaments are composed of pre-S1, pre-S2, and S polypeptides, in both glycosylated and unglycosylated forms. The S polypeptide is the major hepatitis B surface antigen (HBsAg).

Although there is only one major serotype of HBV, HBsAg has five major subtype determinants, termed *a, d, y, w,* and *r,* which are primarily of epidemiologic interest. All HBsAg-positive sera contain determinant *a*; determinants *d* and *y* are mutually exclusive, as are *w* and *r*. Hence, four subtype patterns are possible: *adw, ayw, adr,* and *ayr*. The first three subtype patterns occur frequently; *ayr* is rare. Many studies have attempted to correlate subtype with clinical course. It appears, however, that the subtypes are associated with different geographic distributions of HBV rather than with different degrees of virulence. Subtype *adw* is most common in the Americas and Europe; *adr* prevails in most of the Far East.

HBV can be classified into seven genotypes (A to G) on the basis of an intergroup divergence of 8% or more in the complete nucleotide sequence.[7,8] Genotypes A (serotype *adw*) and D (*ayw*) are most common in the United States and Europe; genotypes B (*adw*) and C (*adr*) are most frequent in China and Southeast Asia.[9] There are also several variations or mutations in the nucleotide sequence of HBV. Core promoter and precore variants produce HBV virions that do not produce hepatitis B e antigen (HBeAg). These variants are most commonly seen in patients with genotypes B, C, and D. These differences in genotypes and the presence or absence of variants may account for variations in clinical manifestations of chronic HBV infection in the United States and other parts of the world.

hepatitis c virus

HCV is a single-stranded, positive-sense RNA virus that is 9.4 kb in length and accounts for most cases of non-A, non-B hepatitis.[10,11] It is most closely related to the pestiviruses and flaviviruses and is believed to be a distinct genus in the Flaviviridae family.

Table 1 Serologic Diagnosis of Acute Viral Hepatitis

Disease	*Serology*	*Comments*
Hepatitis A	IgM anti-HAV	Reasonably specific
Hepatitis B	HBsAg IgM anti-HBc	May be negative late Indicates acute hepatitis
Hepatitis C	Anti-HCV HCV RNA	Appears late Appears early
Hepatitis D	HBsAg and anti-HDV + IgM anti-HBc – IgM anti-HBc	Anti-HDV may appear late Coinfection Superinfection
Hepatitis E	Anti-HEV	Not licensed in the United States

HAV—hepatitis A virus HBc—hepatitis B virus core HBsAg—hepatitis B surface antigen HCV—hepatitis C virus HDV—hepatitis D virus HEV—hepatitis E virus

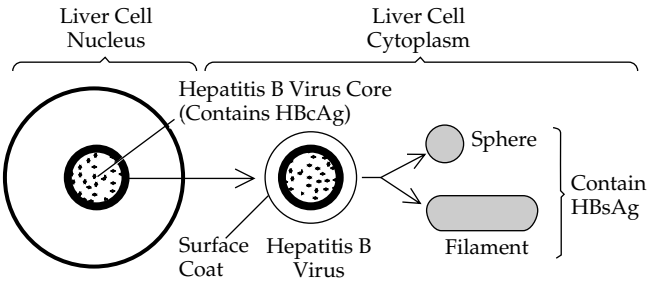


Figure 1 **The hepatitis B virus exists in the cytoplasm of parenchymal liver cells of persons with hepatitis B and constitutes the infective virus. The core of this particle is found in the nucleus of parenchymal cells (left), but as it passes through the cytoplasm, it acquires a surface coat (middle). The core contains hepatitis B core antigen (HBcAg). Spheres and filaments, also in the cytoplasm (right), appear to be excess surface coat material. They are the main source of hepatitis B surface antigen (HBsAg) in serum.**

Structural proteins are encoded at the 5′ end, which is highly conserved [*see Figure 2*]. Further downstream are the HCV core protein, the envelope proteins, and the four nonstructural proteins located at the 3′ end. On the basis of nucleic acid sequence analysis, at least six major genotypes and various subtypes have been identified worldwide. The HCV genotypes are divided into types (1, 2, 3, etc.) and more closely related subtypes (1a, 1b, 2a, etc.). Specific HCV genotypes exhibit different degrees of responsiveness to interferon therapy (e.g., genotypes 2 and 3 respond better to interferon therapy than does genotype 1). In the United States, genotype 1 is the most common, accounting for 70% to 80% of cases; genotype 2 accounts for about 15% of cases, and genotype 3 accounts for 5%. Other genotypes appear to be uncommon in the United States. Coinfection with more than one genotype may occur, particularly in patients with hemophilia or in other patient groups who have been repeatedly exposed to HCV. Quasispecies of HCV also exist. This genetic heterogeneity develops with a longer duration of infection; it may be associated with a poorer response to interferon therapy.

HEPATITIS D VIRUS

HDV, or delta agent, is a single-stranded, circular, negative-polarity, defective RNA virus that requires HBV for its expression.[12] HDV is smaller than any known animal virus and resembles certain plant viruses known as viroids. It circulates in the blood in association with hepatitis D antigen, and the RNA genome has an external coat composed of HBsAg. Although hepatitis D antigen, HDV RNA, and IgM anti–hepatitis D virus (anti-HDV) can be found in the plasma of infected persons, the only commercially available serologic marker for this infection is the total IgG antibody to hepatitis D virus antigen (anti-HDV). When anti-HDV is present in serum, markers of the HBV replication, such as HBeAg and HBV DNA, are usually absent. Although HDV infection is present worldwide, it is most prevalent in Mediterranean countries, the Middle East, and northern Africa. The virus is responsible for epidemics of fulminant hepatitis in South America. HDV infections are uncommon in the United States and northern Europe, except in I.V. drug abusers and persons frequently exposed to blood products. HDV is also uncommon in Southeast Asia and China, areas where HBV is common. Successful vaccination against HBV will prevent HDV infection.

HEPATITIS E VIRUS

HEV is a single-stranded, positive-sense RNA virus of approximately 7.5 kb that causes enterically transmitted non-A, non-B hepatitis.[13] The diagnosis of HEV infection can be made by serologic identification of anti-HEV antibodies (not yet commercially available in the United States). HEV has also been identified in the stool of patients with acute HEV infection through use of immune electron microscopy. Strain variation in HEV has been noted in different parts of the world (e.g., the HEV [B or Burma] and HEV [M or Mexico] strains have only 76% sequence similarity), but there is only one serotype. Acute HEV infection is observed in developing countries; sporadic cases in the United States have been diagnosed in travelers returning from endemic areas.

HEPATITIS F AND G VIRUSES

Between 5% and 20% of cases of acute and chronic hepatitis are not caused by the five known hepatitis viruses and have been presumed to be caused by non-A–E agents. A virus identified in stool extracts from French patients with non-A–E hepatitis was tentatively called the hepatitis F virus (HFV),[14] but the existence of this virus has not been confirmed and is doubtful. HGV was identified as a coinfection in patients with HCV infection and also in persons with non-A–E hepatitis.[15] HGV is an RNA virus closely related to but distinct from HCV. HGV appears to be similar to a GB virus (GBV).[16] Two viruses, GBV-A and GBV-B, are probably nonhuman viruses that are contaminants of serially passed human serum in tamarins. A third virus, GBV-C, is a human virus that is a closely related genotype of HGV. HGV (GBV-C) does not appear to be pathogenic or an independent cause of acute or chronic hepatitis.[17]

Epidemiology

It is particularly important to consider the epidemiology of types A, B, C, D, and E hepatitis and the special tests that may

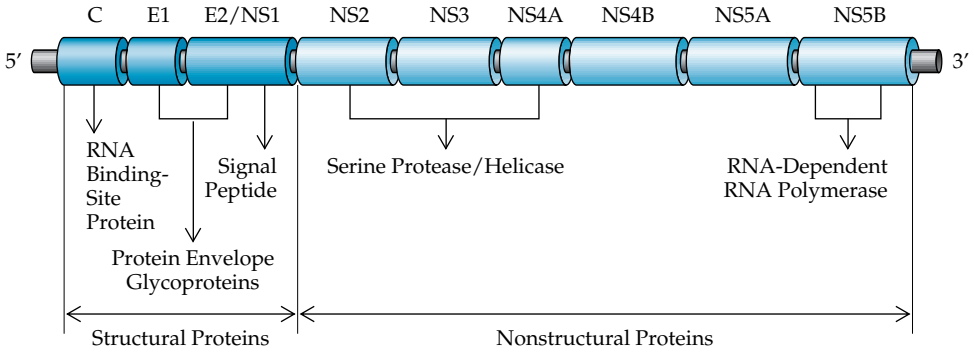


Figure 2 **The hepatitis C virus genome consists of a single, long, open reading frame. The three structural proteins (the RNA binding-site protein [C, core protein] and two envelope glycoproteins [E1 and E2]) are encoded on the 5′ end, and the nonstructural proteins (protease, helicase, and RNA-dependent RNA polymerase [NS2, NS3/NS4A, and NS5B]) are encoded on the 3′ end.**

Table 2 Features of Type A, Type B, Type C, Type D, and Type E Acute Viral Hepatitis

	Type A	*Type B*	*Type C*	*Type D*	*Type E*
Mode of transmission	Fecal-oral, sewage-contaminated shellfish	Percutaneous, sexual	Percutaneous and community	Percutaneous	Fecal-oral
Incubation period (days)	20–37 (15–49)*	60–110 (25–160)*	35–70 (21–84)*	Appears to be comparable to type B	10–56
Results of serum antigen and antibody tests	Development of IgM antibody early and IgG antibody in convalescence	Antigen (HBsAg) and antibody (anti-HBc) appear early and persist in carriers	Anti-HCV appears in 6 wk to 9 mo	Anti-HDV appears late and may be short-lived	IgM antibody usually detected within 26 days of jaundice; IgG antibody persists indefinitely
Immunity	45% of the U.S. population has hepatitis A antibodies in serum and is probably immune	5%–15% of the population has anti-HBs	Unknown	Patients immune to hepatitis B are also protected against hepatitis D	Unknown
Prevalence	Seen with increasing frequency in adults	Increasing in the United States	80%–90% of post-transfusion hepatitis; 12%–25% of sporadic acute hepatitis	Unusual in the United States but common in I.V. drug abusers	Rare in the United States
Course	Does not progress to chronic liver disease	Chronic liver disease develops in 1%–5% of adults and 80%–90% of children	Chronic liver disease develops in 85%	> 95% resolution of delta coinfection with acute hepatitis B; chronic infection common if delta superinfection is present in chronic hepatitis B carrier	Does not progress to chronic liver disease
Prevention of the disease after exposure	Pooled γ-globulin (0.02 ml/kg) decreases the occurrence of clinical disease 7- to 8-fold	Hepatitis B immune globulin and hepatitis B vaccine prevent clinical disease in adults and the carrier state in infants	The efficacy of pooled γ-globulin is uncertain	Unknown	Uncertain
Mortality	0%–0.2% with fulminant hepatitis	0.3%–1.5%	Uncertain; may approximate rate for type B	2%–20% for acute icteric hepatitis	1%–2%; may be as high as 10%–15% in pregnant women

*Usual range, with outside limits given in parentheses.

permit their differentiation, because their prognoses are considerably different [*see Tables 1 and 2*].[1,2]

HEPATITIS A VIRUS

In HAV infection, virus is shed in the stool 14 to 21 days before the onset of jaundice. Although patients may continue shedding virus for the first 1 to 2 weeks of clinical illness, they are usually no longer infectious 21 days after the illness has begun. However, virus may be detected in the stool again if the patient experiences a relapse of the acute illness. HAV is transmitted via food that has been contaminated by feces-soiled hands of infected persons. The disease is quite contagious; transmission in families is common, and several large point-source epidemics have been reported.[18] Outbreaks of HAV infection have been reported in day care centers, with young children being the most commonly infected.[19] Employees of the day care centers and household contacts and close relatives of the infected children contracted the disease with alarming frequency. Ingestion of sewage-contaminated shellfish has resulted in several epidemics of HAV,[20] as has contamination of raw produce, as in the recent outbreak of HAV infection associated with green onions.[21] The disease is sexually transmitted in men who have sex with men.[22,23] HAV is also common in I.V. drug abusers, but the method of transmission is uncertain. Viremia is present from 1 to 25 days before the onset of symptoms, but transmission by serum or blood products seldom occurs. Patients with HAV develop immunity to the disease—approximately one third of the population of the United States have serum antibodies to HAV. There is no known human or nonhuman reservoir of HAV.

HEPATITIS B VIRUS

HBV is transmitted primarily through percutaneous inoculation of infected serum or blood products. HBsAg and HBV DNA are found in a wide variety of bodily secretions, but the importance of these factors in the spread of HBV is unknown. The most common mode of transmission in men who have sex with men may be by oral or genital contact with asymptomatic bleeding lesions in the rectal mucosa. HBV may also be transmitted to the fetus during pregnancy. An appreciable segment of the population has serum antibodies to HBsAg (anti-HBs). Prevalence of anti-HBs varies among subpopulations: middle-class whites have a 5% prevalence; middle-class African Americans, 12%; Chinese Americans, 37%; and white homosexual men, 48%. This antibody confers immunity to HBV.

HEPATITIS C VIRUS

Before the advent of tests that could screen for HCV and thus help eliminate this virus from the blood supply, HCV caused most cases of posttransfusion hepatitis.[24] Since that time, the rate of HCV transmission by transfusion has declined, and the risk of posttransfusion HCV infection is estimated to be between 0.01% and 0.001% per unit transfused.[25] HCV is responsible for more than 80% of non-A, non-B hepatitis.

HCV is transmitted by parenteral means (e.g., transfusions, I.V. drug use, or occupational exposure to blood or blood products).[26] The risk of transmission from a single needle-stick accident averages 1.8% in prospective studies (range, 0% to 7%).[25] It is estimated that the risk of sexual transmission of HCV in monogamous couples is about 5%, well below the risk of sexual transmission of HBV (about 30%) or HIV (about 10% to 15%). However, the rate of HCV infection is higher in persons who have frequent sexual contact with numerous partners, and in this setting, the risk is higher for female partners of men with anti-HCV.[27] Most studies, particularly those from the United States, have failed to demonstrate any serologic or virologic evidence of HCV transmission to nonsexual partners within households. Perinatal transmission of HCV infection is unusual, except in babies born to mothers with very high levels of HCV RNA, such as mothers with concomitant HIV infection. The risk of perinatal transmission is estimated to be between 5% and 6%. There appears to be no increase in HCV infection in breast-fed babies. In summary, barrier precautions are not recommended for monogamous partners, but persons with multiple sexual partners should practice safe sex and use latex condoms. An additional commonsense precaution is to avoid shared percutaneous exposures, such as razors and toothbrushes. Finally, there is no reason to advise against pregnancy for a woman with HCV infection, because the rate of perinatal infection is low.

Recipients of organs from donors who have antibodies to HCV have a high probability of becoming infected.[28] Transplantation of an organ from an infected donor is controversial.

HEPATITIS D AND E VIRUSES

HDV occurs only in patients with HBV and is transmitted percutaneously. Simultaneous infection with HBV and HDV may produce a more severe acute hepatitis than that caused by HBV alone.[12]

HEV is a common cause of large epidemics of acute viral hepatitis in developing countries. It characteristically affects adults and may be associated with an unusually high mortality in pregnant women.[13] Epidemics have occurred in rural Mexico, with a high attack rate and with jaundice occurring in more than 5% of the local popoulation. HEV infection has also been found in immigrants to the United States and in travelers to Mexico and the Indian subcontinent.[29]

Table 3 Incidence of Symptoms in Acute Viral Hepatitis

Symptom	*Percentage of Patients*
Dark urine	94
Fatigue	91
Anorexia	90
Nausea	87
Fever	76
Emesis	71
Headache	70
Abdominal discomfort	65
Light stools	52
Myalgia	52
Drowsiness	49
Irritability	43
Itching	42
Diarrhea	25
Arthralgia	21

Diagnosis

CLINICAL MANIFESTATIONS

The onset of viral hepatitis may be gradual or sudden. The symptoms are protean [*see Table 3*].[1,2] The most common early symptoms are fatigue, lassitude, drowsiness, anorexia, nausea, and dark urine. Dehydration may result from repeated vomiting. Low-grade fever is common; shaking chills are rare. Frank pain may occur in the right upper quadrant, but vague, generalized abdominal discomfort is more common. Itching may occur but is seldom severe. Diarrhea occurs in some cases. About half of patients have myalgias or arthralgias, and some have acute arthritis with local pain, redness, swelling, and effusions. Joint symptoms are usually associated with HBV and HEV infections. Many of these early symptoms abate when jaundice develops or shortly thereafter. In the case of severe hepatitis, which is unusual, confusion, stupor, or even coma may develop. Fetor hepaticus and asterixis are usually present in these patients.

PHYSICAL EXAMINATION

The sclerae and skin may be icteric. The liver is often enlarged and tender. The spleen is palpable in about 10% of patients. Asterixis, marked peripheral edema, or ascites implies that the disease is unusually severe and suggests a poor prognosis.

LABORATORY TESTS

General Laboratory Findings

Most patients have mild anemia and relative lymphocytosis. The leukocyte count is usually normal but may be greater than 12,000/mm^3. The serum bilirubin level generally does not exceed 15 to 20 mg/dl; levels greater than 30 mg/dl imply severe disease or associated hemolysis. Serum aspartate aminotransferase (AST) and serum alanine aminotransferase (ALT) levels rise 7 to 14 days before the onset of jaundice and begin to fall shortly after jaundice occurs. The degree of aminotransferase elevation does not necessarily parallel severity, but levels less than 500 IU/L usually reflect mild illness.

The alkaline phosphatase level is slightly increased but may be markedly elevated in the few patients in whom prominent cholestasis develops later in the course of acute illness. Serum γ-globulin levels are normal or slightly elevated; concentrations greater than 3 g/dl suggest chronic active hepatitis rather than acute viral hepatitis. The serum albumin level and prothrombin time reflect the liver cells' synthetic capacity and are depressed in patients with severe, acute viral hepatitis.

Serologic and Virologic Assays

Serologic assays are used to identify each type of viral hepatitis [*see Table 1*].

Hepatitis A virus The IgM antibody to HAV (IgM anti-HAV) appears early and is quite specific for acute HAV infection. It typically persists for an average of 3 months and is then replaced by the IgG anti-HAV, which lasts throughout life and confers immunity to future infection.

Hepatitis B virus A number of serologic tests are useful to physicians who are caring for patients with acute HBV infection [*see Table 4*]. HBsAg is present on the surface of the hepatitis B virion and in the circulation as spheres and filaments [*see Figure 1*]. It appears in the serum of infected persons as early as 1 to 2

Table 4 Tests for Hepatitis B Virus Infection

Symbol	*Characteristics*
HBsAg	Hepatitis B surface antigen; present in surface coat of the hepatitis B virus and in the 22 nm diameter filaments and spheres; purified surface antigen expressed in yeast cells is used in the recombinant hepatitis B vaccines
Anti-HBs	Antibody to HBsAg; present during convalescent phase of acute hepatitis B infection and after successful hepatitis B vaccination; hepatitis B immune globulin derived from serum with high anti-HBs titers is effective in preventing clinical hepatitis B infection
Anti-HBc	Antibody to the hepatitis B virus core; present in all patients with any form of hepatitis B infection; presence of this antibody in the sera is evidence that the patient has been infected in the past or is currently infected with the hepatitis B virus
IgM anti-HBc	Antibody to the hepatitis B core antigen; present in high titers in patients with acute hepatitis B; may be the only marker of acute infection when HBsAg is no longer detectable
Anti-HDV	Antibody to hepatitis D; serologic marker of coinfection or superinfection by hepatitis D of patients with hepatitis B
HBeAg	A soluble protein derived from the hepatitis B virus core; reflects presence in the blood of circulating hepatitis B virus; sera positive for this antigen are highly infectious
Anti-HBe	Anti-HBe appears weeks to months after HBeAg (and the circulating hepatitis B virus) is no longer detectable in the blood; sera positive for this antibody are substantially less infectious

weeks after parenteral injection of infectious virus and may persist for months. If HBsAg remains in the serum of an infected person for 6 months after an episode of acute HBV infection, it will probably persist indefinitely. The antibody to HBsAg usually appears in the blood 2 to 4 months after an attack of acute HBV infection that resolves—in most cases, after HBsAg is no longer detectable. Antibody to the core antigen (anti-HBc) appears promptly in the blood of infected persons and persists indefinitely. High titers of IgM anti-HBc are found in patients with acute disease and may be the only marker of acute HBV infection if HBsAg is no longer detectable. Detection of HBeAg, a soluble protein derived from the core particle, correlates with the presence of HBV DNA and indicates that the HBV is actively replicating. Serum that is positive for HBeAg is highly infectious. A pregnant woman who is positive for HBsAg is much more likely to transmit HBV to her offspring if her blood also contains HBeAg or HBV DNA. The detection of antibody to HBeAg (anti-HBe) in association with the absence of HBV DNA is evidence that viral replication is minimal and the blood is substantially less infectious.

Hepatitis C virus Detection of antibodies to HCV (anti-HCV) remains the most practical way to diagnose acute and chronic HCV infection.[30,31] In 1992, a second-generation anti-HCV enzyme-linked immunosorbent assay (ELISA) replaced the original assay. The second-generation assay employs several viral antigens, making it more sensitive and specific than the first-generation test. The most commonly used supplemental assay for specificity is the second-generation recombinant immunoblot assay (RIBA-2), which incorporates four antigens as separate bands. Reaction of two or more of the four bands is considered a positive RIBA-2 test result; reaction of only one band is indeterminate. Third-generation ELISAs and RIBAs have been developed and should soon be available in the United States.

The detection of serum HCV RNA can be used to establish viremia and has been widely employed in the management of chronic HCV infection.[30,31] The diagnosis of acute HCV infection can also be established early, before the appearance of anti-HCV, by the use of HCV RNA assays. Assays for serum HCV RNA have recently become standardized, and serum HCV RNA can be detected by both qualitative and quantitative assays. The results of a quantitative HCV RNA assay may be useful in predicting the subsequent response to interferon therapy; patients with lower levels of viremia respond better than those with higher levels. Finally, in patients with normal serum aminotransferase levels in whom the results of RIBA-2 are positive for anti-HCV, the presence or absence of serum HCV RNA can distinguish between active infection with viremia and recovery from previous HCV infection.

Hepatitis D and E viruses HDV infection is diagnosed by the detection of anti-HDV with HBsAg. Patients with acute HDV infection will have HBsAg with IgM anti-HBc along with anti-HDV (i.e., coinfection), whereas patients with chronic HBV infection who are superinfected with HDV will have a negative IgM anti-HBc. HEV is diagnosed by the detection of antibodies to HEV (anti-HEV). Anti-HEV is found in acute and convalescent serum from patients with acute HEV infection. This assay has yet to be licensed in the United States.

Liver Biopsy

Liver biopsy is not usually performed in patients with acute viral hepatitis, because serologic tests are generally diagnostic. Spotty necrosis of liver cells and an inflammatory cell reaction that consists primarily of lymphocytes and histiocytes are the typical histologic findings. Acidophils (dying liver cells) and bile plugs are common. Biopsy performed late in the course of the disease reveals prominent evidence of hepatic cell regeneration (rosette formation and multinucleated cells) and pigment-filled histiocytes. Although usually more marked in the pericentral areas, the inflammatory reaction and cell necrosis appear throughout the parenchyma. In severe hepatitis, necrotic zones link portal areas to one another or to central areas, or they may involve whole lobules (bridging necrosis). The portal tracts contain a mild to moderate mononuclear cell inflammatory reaction, and the limiting plate, which demarcates portal areas from parenchymal cells, may be disrupted.

CLINICAL COURSE

Hepatitis typically produces symptoms for 1 to 2 weeks before the onset of dark urine and jaundice. As icterus deepens, appetite begins to return and malaise lessens. The serum bilirubin level rises for 10 to 14 days and then declines over 2 to 4 weeks. Aminotransferase levels usually begin to decline just before peak jaundice occurs and fall quite rapidly thereafter. The patient often feels much better by the time bilirubin levels have begun to decline. Usually, the clinical course is uneventful and recovery is complete, with liver function returning to normal.

In a small percentage of patients, the clinical course is atypical. Acute viral hepatitis may be protracted in elderly patients or in those infected with either HBV or HCV; the disease may last several months, and full recovery may not occur for a year. Between

6% and 15% of patients with acute viral hepatitis will have recurrent symptoms and worsening of liver function before recovery from the initial attack is complete. This relapse is usually milder than the original attack and is short-lived. In a few patients, the disease has an acute fulminant course leading to hepatic coma and even death. These events appear to be more common in pregnant women infected with HEV. HCV appears to be an unusual cause of fulminant disease.[32] In some countries, a mutant form of HBV that is incapable of encoding for e antigen is associated with fulminant disease.[9] Often, no virus can be identified, leading to speculation that other viruses may be involved—either mutant viruses or as yet unidentified viruses. Some patients do not recover completely from the initial attack, and chronic hepatitis develops. Chronic hepatitis does not occur after HAV or HEV infection. It ensues in 1% to 5% of cases of acute HBV infection and 85% of cases of acute HCV infection.[1,2]

A study in Italy has found that carriers of HBsAg who are symptom-free and whose liver function tests are normal have an excellent prognosis. In this study, the risk of hepatocellular carcinoma was low over the mean follow-up period of about 11 years.[33] The natural history of HCV infection is still being studied. The general view is that chronic HCV infection usually progresses but does so slowly over many decades.[26,34] In two histologic studies of the natural history of chronic HCV infection that resulted from blood transfusion, the times to the presence of chronic HCV infection averaged 12 years; to the presence of cirrhosis, 21 years, and to the presence of hepatocellular carcinoma, 29 years.[35,36] Both of these studies reported experiences from tertiary liver centers, however, which may have introduced referral bias. More recent studies have suggested that chronic HCV infection is more benign than originally reported and is associated with a low rate of cirrhosis.[37-40] In a large, prospective study of 568 patients with posttransfusion non-A, non-B hepatitis (mostly HCV) who were followed for an average of 18 years, there was no increase in mortality from all causes, but there was a small increase in the number of deaths related to liver disease.[39] Another large, cross-sectional study from Europe of 2,235 patients with chronic HCV infection found that the median time from infection to cirrhosis was 30 years,[40] which is about 10 years longer than the 21-year interval from infection to cirrhosis reported earlier from tertiary liver centers.[35,36] Analysis of liver biopsy specimens showed that the rate of fibrosis progression was not normally distributed, with approximately one third of persons progressing to cirrhosis in less than 20 years and another third not appearing to progress to cirrhosis for at least 50 years. Factors associated with an increase in the rate of fibrosis progression were age at which infection occurred (> 40 years of age), daily alcohol use (> 50 g), and male gender. The role of heavy alcohol abuse in exacerbating the risk of cirrhosis has been confirmed.[41] Once cirrhosis develops in patients with chronic HCV infection, the 10-year rates of decompensation of cirrhosis and development of hepatocellular carcinoma are 29% (3.9% yearly) and 14% (1.4% yearly), respectively.[42] Chronic HBV infection also predisposes the infected person to the development of primary hepatocellular carcinoma, perhaps through the integration of viral DNA into the genome of the host's hepatocytes.[43]

Unusual and sometimes fatal complications of acute viral hepatitis include aplastic anemia, hemolytic anemia, hypoglycemia, and polyarteritis. The risk appears to be higher if infection occurs at a very early age or if chronic liver disease is also present.

Differential Diagnosis

At the time of initial presentation with symptoms and elevated aminotransferase levels, before the results of serologic tests are known, it is worthwhile to consider the differential diagnosis of acute viral hepatitis.

EPSTEIN-BARR VIRUS

EBV, a herpesvirus, usually produces mild hepatitis associated with nausea and vomiting; jaundice occurs in only 10% to 20% of patients.[44] Serum aminotransferase levels are moderately elevated (300 to 500 IU/L). In most instances, the hepatitis is part of the typical clinical syndrome of infectious mononucleosis. In rare instances, hepatic dysfunction is severe and proves to be fatal, particularly in immunodeficient patients.[45] The virus appears to be transmitted during oral-oral contact through infected saliva and may be transmitted parenterally; the incubation period is about 28 days. A rise in titer of specific fluorescent antibodies to EBV or detection and quantitation of viral levels confirm the diagnosis.

CYTOMEGALOVIRUS

CMV, which is also a member of the herpesvirus group, is ubiquitous. About 80% of adults have serum complement-fixation reactivity for CMV. This virus can also produce a disease similar to infectious mononucleosis but without adenopathy or tonsillopharyngeal involvement.[46] Liver involvement may mimic that of the more common forms of viral hepatitis, but it is usually mild and does not progress to chronic liver disease. Diagnosis requires inoculation of an appropriate tissue culture with blood to demonstrate viremia. Polymerase chain reaction to assess quantitative viral loads is the best test and is becoming more widely used.

OTHER VIRUSES

Acute hepatitis caused by herpes simplex virus or varicella-zoster virus, usually accompanied by typical skin lesions, has occurred in immunocompromised patients.

DRUG-INDUCED HEPATITIS

Hundreds of drugs can cause hepatitis that may be indistinguishable from acute viral hepatitis.[47] These idiosyncratic drug reactions are infrequent, unpredictable, and not dose dependent. Clinical onset usually occurs within 2 to 6 weeks after therapy is started but may occur on the first day that the drug is administered or not until 6 months later. The disease may progress despite withdrawal of the drug; failure to withdraw the drug promptly may result in death.

One well-documented drug reaction is the hepatic necrosis that occurs in one in 9,000 to 10,000 patients given halothane. The hepatitis is often fatal and is more common in overweight women or in persons exposed a second time to the anesthetic. Fever, malaise, and elevated aminotransferase levels develop 1 to 12 days after initial exposure to the drug. Onset may be sooner after multiple exposures; the average delay is 3 days. Signs of hepatic necrosis include marked eosinophilia, marked elevation of serum aminotransferase and bilirubin levels, reduced serum albumin levels, and a prolonged prothrombin time. Other common drugs that cause hepatitis are isoniazid, methyldopa, phenytoin, and the sulfonamides. Because most drugs will injure the liver on rare occasions, hepatitis that develops shortly after initiation of a new medication should suggest a drug reaction. The treatment of choice is discontinuance of the medication.

Idiosyncratic drug reactions differ from hepatitis that results from drug overdose. Drug reactions of the latter type are rare be-

cause a clear potential for hepatotoxicity usually precludes release of the drug. Acetaminophen, however, is an exception: more than 25 g orally, usually in a suicide attempt, will cause profound hepatocellular necrosis in most persons.[48] Cell injury occurs because the liver produces a toxic metabolite that is usually rendered harmless by conjugation with glutathione. When the drug dose is high, hepatic glutathione stores are depleted and the toxic metabolite accumulates and destroys liver cells. Oral *N*-acetylcysteine, given in a loading dose of 140 mg/kg followed by 70 mg/kg every 4 hours for a total of 18 doses, reduces hepatotoxicity and mortality in cases of acetaminophen overdose.[49] The drug is most effective when given within 8 hours after the overdose but appears to have some effect as long as 24 hours after the ingestion of acetaminophen. Maximal medical support during the 1- to 2-week illness is mandatory. Acetaminophen can also cause severe hepatotoxicity when taken in ordinary doses if hepatic glutathione stores are low as a result of alcoholism with malnutrition (the so-called Tylenol-alcohol syndrome). The beneficial role of *N*-acetylcysteine in this clinical setting is less certain.

Treatment

Many treatments have been recommended for acute viral hepatitis, but it is unlikely that any of them alters the course of the disease. When the patient feels ill, it seems reasonable to reduce physical activity to a tolerable level. For some patients, bed rest may be indicated during the initial phase of illness. Once the patient feels better, there is no reason to restrict activity. Two large, controlled studies of young servicemen with viral hepatitis have shown convincingly that even heavy physical exercise in the recovery period does not result in more frequent relapse or chronic disease.[50,51]

Patients should be encouraged to eat whatever they can; there is no evidence that a low-fat diet is beneficial. At times, nausea and vomiting are so severe that hospitalization and intravenous fluid and electrolyte replacement become necessary. Abstention from alcohol is advised during the acute phase, although alcohol has not been shown to adversely affect the course of viral hepatitis.

Acute HCV infection is usually silent and thus not commonly seen in clinical practice. However, meta-analyses of published studies of interferon therapy for acute HCV infection support its efficacy in this setting.[52,53] It is generally recommended that standard doses of interferon (e.g., 3 million units of interferon alfa-2b three times a week) be administered for 3 to 6 months, which, in comparison with no therapy, increases the likelihood of sustained biochemical (normal ALT levels) and virologic (undetectable HCV RNA) responses.

Many forms of treatment have been recommended for the patient with severe acute viral hepatitis who becomes encephalopathic, but no regimen is clearly effective. In controlled clinical trials, corticosteroids,[54-57] cimetidine,[58] hyperimmune γ-globulin,[59] and exchange transfusions[60] had no effect on the course of acute hepatitis. Although no controlled trials have been completed, liver transplantation improves survival in patients with acute severe viral hepatitis and stage IV hepatic encephalopathy.[61]

Therapy for encephalopathic patients should be supportive, with evaluation for liver transplantation. Bacterial infections should be treated with suitable antibiotics. Bleeding warrants the administration of appropriate clotting factors (fresh frozen plasma, platelets, or both) and transfusions. Clotting abnormalities without bleeding do not justify massive transfusions of fresh frozen plasma, because congestive heart failure can result. Encephalopathy should be treated with oral lactulose (30 ml every 4 hours), although there is little evidence that acute encephalopathy responds to treatment.

Antiviral therapy is available for chronic HBV and chronic HCV infections [*see 4:VIII Chronic Hepatitis*].

Prevention

Prevention of viral hepatitis entails avoidance of exposure to the virus, passive immunization with globulin products, and active immunization with specific vaccines.

PASSIVE IMMUNIZATION

Immune globulin is prepared from human plasma; when given intramuscularly, it decreases the clinical attack rate for HAV by sevenfold to eightfold. The official recommendation of the U.S. Public Health Service is to administer 0.02 ml/kg to contacts as soon as possible after exposure to a confirmed case of HAV infection.[62] Administration more than 2 weeks after exposure is not protective. The usual dose for adults is 2.0 ml; for children weighing up to 25 kg (55 lb), 0.5 ml; and for children between 25 and 50 kg (110 lb), 1.0 ml. Immune globulin should be given to all persons who share a household, hospital room, or dormitory room with an HAV patient. It should also be given to staff and children in day care centers where cases of HAV infection have been identified. When a food handler with acute HAV infection has been identified, immune globulin should be given to coworkers and considered for patrons of the eating establishment if they can be identified and treated within 2 weeks after exposure. Immune globulin need not be given to all contacts at work or school unless there is clear evidence of spread. Classmates and neighborhood children who play together frequently, however, probably should be immunized. Travelers to developing countries are at risk of acquiring HAV, particularly those who plan to visit extensively or to reside in areas with poor sanitation [*see Clinical Essentials:VII Health Advice for International Travelers*]. A single dose of 0.02 ml/kg of immune globulin will be protective for as long as 2 months; a dose of 0.06 ml/kg will be protective for 5 months.[62]

Immune globulin contains anti-HBs at low titer (approximately 1:100 by radioimmunoassay), whereas HBV immune globulin has an anti-HBs titer of greater than 1:100,000. When the source is known to be HBsAg positive, persons exposed (by percutaneous or mucous membrane routes) should be given HBV immune globulin (0.06 ml/kg) and HBV vaccine within 24 hours.[63,64] When the HBsAg status of the source is unknown, the first dose of HBV vaccine should be given promptly and the series completed as recommended. If the source is subsequently found to be HBsAg positive, HBV immune globulin (0.06 ml/kg) should be administered, provided that it can be given within 7 days after exposure. Infants who are born to HBsAg-positive mothers should receive 0.5 ml of HBV immune globulin intramuscularly and 0.5 ml of HBV vaccine intramuscularly at another site within 12 hours after birth.

Prophylaxis against HCV infection with γ-globulin is more problematic. Its effect in household or casual contacts and after a needle-stick injury is unknown. The value of immune globulin in the prevention of HEV infection is also uncertain.

ACTIVE IMMUNIZATION

Two HBV vaccines that are produced by recombinant DNA techniques are available (Recombivax HB, Merck & Co., and Engerix-B, SmithKline Beecham Biologicals) [*see Table 5*].[63,64] Both

Table 5 Administration Schedules and Dosing of Hepatitis B Vaccines[63,64]

Patients	*Schedule*	*Engerix-B*	*Recombivax HB*
Infants			
HBsAg-negative mother	0–2, 1–4, and 6–18 mo	10 μg/0.5 ml	2.5 μg/0.5 ml
HBsAg-positive mother	At birth,* 1–2 mo, and 6 mo	10 μg/0.5 ml	5.0 μg/0.5 ml
Children and adolescents (0–19 yr)	0, 1–2, and 4–6 mo	10 μg/0.5 ml	5.0 μg/0.5 ml
Alternative two-dose regimen for adolescents (11–15 yr)	0 and 4–6 mo	—	10 μg/1.0 ml
Adults (≥ 20 yr)	0, 1–2, and 4–6 mo	20 μg/1.0 ml	10 μg/1.0 ml
Immunocompromised adults (hemodialysis)	0, 1, and 6 mo	40 μg/2.0 ml	40 μg/1.0 ml

*Immunization should occur within 12 hr with hepatitis B immune globulin.
HBsAg—hepatitis B surface antigen

vaccines are highly effective in inducing antibody to HBV and preventing HBV infection in infants, children, and adults. The recommendations of the Centers for Disease Control and Prevention for the use of HBV vaccine are outlined [*see Table 6*].

The vaccines are given in three I.M. doses into the deltoid muscle in young children and adults and into the anterolateral thigh muscle in infants and neonates. A suboptimal response has been observed when the vaccine was injected into the buttocks.[65] The second dose is given 1 month after the first; the third dose is usually given 6 months after the first. For healthy adults, depending on the vaccine preparation, each dose should contain 10 or 20 μg of HBsAg; for patients undergoing hemodialysis and for other immunosuppressed patients, each dose should contain 40 μg; and for infants and children younger than 10 years, each dose should contain 2.5 to 10 μg. The vaccine should be given to groups at substantial risk for HBV infection: hospital staff and other health care workers with frequent exposure to blood products, clients and staff of institutions for the mentally retarded, hemodialysis patients, homosexually active males, users of I.V. drugs, recipients of certain blood products, contacts of HBV carriers, infants born to HBsAg-positive mothers, special high-risk populations (e.g., emigrants from areas with highly endemic disease), and prisoners. It should be strongly considered for travelers who plan to reside in areas with high levels of endemic HBV infection.

Approximately 3% to 4% of healthy people have little or no antibody response to the vaccine. They appear to lack an immune response gene in the major histocompatibility complex that accounts for the ability to mount a normal antibody response to HBsAg.[66] Repeat vaccination induces a protective level of antibody in less than 50% of such people.[67] Those who do not respond to vaccine and who later become infected with HBV do not have an unusual clinical course.

After successful vaccination, titers of antibody to HBsAg begin to decline, and in 5 years, 20% to 30% of patients lack protective levels.[68] These persons will respond immediately to a booster dose of vaccine, but the need for routine booster vaccination has not been determined.[69] HBV infection may develop in a few persons when the antibody titer falls to low levels, but the infection is invariably asymptomatic and is usually identified only by the development of antibody to the core antigen. Thus, several countries and certain individuals have a policy of administering a booster injection to certain risk groups if the anti-HBs level falls below 10 mIU/ml.[70]

Vaccine is useless in HBV carriers, and it is unnecessary for those already immune to HBV; however, it has no ill effects on these groups. Therefore, the decision to screen people for susceptibility to HBV infection before vaccination is primarily based on the relative costs of the two procedures. In general, if the expected prevalence of immune persons in a particular population is high (> 20%) and the cost of screening is low (< $30 a person), screening should be done. However, it is difficult to decide which antibody (anti-HBs or anti-HBc) should be used. Anti-HBc provides definitive proof of HBV infection, but it does not discriminate between HBV carriers and noncarriers. Anti-HBs is usually not present in HBsAg carriers, but even when anti-HBs is present in the general population, it may not indicate immunity to HBV infection. Several studies have identified persons who have low serum titers of anti-HBs but who do not have anti-HBc.[71,72] This low-titer anti-HBs is predominantly of the IgM type and may not be immunoprotective.[72,73] Thus, if anti-HBs is used as the screening test, only a titer of 10 radioimmunoassay sample ratio units or higher or a positive enzyme immunoassay should be considered evidence of immunity.

Several different approaches to the development of active immunization against HAV have been attempted. A successful approach has been the preparation of inactivated HAV vaccines [*see Table 7*]. Other approaches to vaccine development, including a live attenuated and recombinant vaccine, have been slower in development because of a number of technical problems.

A large, randomized, double-blind efficacy trial demonstrating protection against HAV infection was carried out in Thailand.[74] A total of 40,119 children 1 to 16 years of age received Hav-

Table 6 Recommendations for Use of Hepatitis B Vaccine[63,64]

Routine immunization
- All infants and previously unvaccinated children and adolescents 1–18 years of age

Persons at increased risk for HBV infection
- Persons with multiple sexual partners
- Sexual partners or household contacts of HBsAg-positive persons
- Men who have sex with men
- Injecting drug users
- Travelers to regions of high HBV endemicity
- Persons with occupational exposure to blood or body fluids
- Clients and staff of institutions for developmentally disabled persons
- Patients with chronic renal failure
- Patients receiving clotting factor concentrates

HBsAg—hepatitis B surface antigen HBV—hepatitis B virus

Table 7 Administration Schedules and Dosing of Hepatitis A Vaccines

Vaccine	*Patients*	*Dosage*	*Schedule (months)*
Havrix	Children (2–18 yr) Adolescents and adults (> 18 yr)	720 ELU/0.5 ml 1,440 ELU/1.0 ml	0, 6–12 0, 6–12
VAQTA	Children (2–17 yr) Adolescents and adults (> 17 yr)	25 U/0.5 ml 50 U/1.0 ml	0, 6–18 0, 6

ELU—enzyme-linked immunosorbent assay unit

rix or a control HBV vaccine (Engerix-B) at 0, 1, and 12 months. Patients were crossed over to the alternative vaccine at 18 months. Side effects were minor. The efficacy of the HAV vaccine was 94% before the month-12 booster injection and 99% after it. In an earlier trial, researchers evaluated a different inactivated HAV vaccine.[75] They studied 1,037 children in upstate New York and also found that this HAV vaccine was safe and 100% effective in preventing HAV infection 50 to 137 days after administration of a single dose. These studies led to the licensing of a second HAV vaccine, VAQTA (Merck & Co.). Both Havrix and VAQTA are administered to adults in an initial 1 ml dose I.M. followed by a booster in 6 to 18 months.

Compared with the short-term protection afforded by immune globulin, inactivated HAV vaccine will probably induce protection lasting from 5 to 10 years and perhaps much longer.[76] In the United States, the overall incidence of HAV infection has decreased, leading to a higher proportion of adults who are susceptible. Older individuals are known to experience a more severe clinical course, and thus, the costs of HAV infection in the United States remain substantial.[77] These facts underline the importance of ensuring compliance with the current recommendations for HAV vaccination.[62] The vaccine will be particularly useful in preventing HAV infection in persons at high risk for the disease, such as travelers and immigrants to highly endemic regions. The risk of symptomatic HAV infection in travelers staying in Western-style accommodations in high-risk countries is three per 1,000 persons per month.[78] Backpackers or travelers in areas with poor hygienic conditions have a higher risk (20 per 1,000 persons per month). In unprotected travelers from the United States, the incidence of HAV infection is 10 to 100 times that of typhoid fever and 1,000 times that of cholera. Other persons at risk for HAV infection are listed [*see Table 8*].[62] Finally, patients with chronic liver disease may experience a more severe illness with acute HAV infection.[79,80] HAV vaccination has been shown to be safe and effective in patients with chronic viral liver disease.[81]

Table 8 Recommendations for Use of Hepatitis A Vaccine[62]

Routine immunization
- Children living in areas where rates of HAV infection are at least twice the national average (i.e., ≥ 20 cases per 100,000 population); vaccination should also be considered for children in areas where rates are greater than the national average, which is 10–20 cases per 100,000 population

Persons at increased risk for HAV infection
- Persons traveling to or working in countries with high or intermediate HAV endemicity, such as Mexico, the Caribbean, Southeast Asia, South and Central America, and Africa
- Men who have sex with men
- Illegal drug users
- Individuals who work with HAV-infected primates or with HAV in research laboratories
- Persons with clotting factor disorders
- Outbreaks in communities with high or intermediate rates of HAV infection

Persons at increased risk for more severe disease
- Persons with chronic liver disease

HAV—hepatitis A virus

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Ryder SD, Beckingham IJ: ABC of diseases of liver, pancreas, and biliary system: acute hepatitis. BMJ 322:151, 2001

2. Younossi ZM: Viral hepatitis guide for practicing physicians. Cleveland Clinic of Medicine. Cleve Clin J Med 67(suppl 1):SI6, 2000

3. Summary of notifiable diseases, United States, 1999. MMWR Morb Mortal Wkly Rep 48:1, 2001

4. Yokosuka O: Molecular biology of hepatitis A virus: significance of various substitutions in the hepatitis A virus genome. J Gastroenterol Hepatol 15(suppl):D91, 2000

5. Feinstone SM, Kapikian AZ, Purcell RH: Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness. Science 182:1026, 1973

6. Lee WM: Hepatitis B virus infection. N Engl J Med 337:1733, 1997

7. Norder H, Courouce AM, Magnius LO: Complete genomes, phylogenetic relatedness, and structural proteins of six strains of the hepatitis B virus, four of which represent two new genotypes. Virology 198:489, 1994

8. Stuyver L, De Gendt S, Van Geyt C, et al: A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. J Gen Virol 81(pt 1):67, 2000

9. Lok AS, Heathcote EJ, Hoofnagle JH: Management of hepatitis B: 2000—summary of a workshop. Gastroenterology 120:1828, 2001

10. Choo QL, Kuo G, Weiner AJ, et al: Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244:359, 1989

11. Bukh J, Miller RH, Purcell RH: Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. Semin Liver Dis 15:41, 1995

12. Hadziyannis SJ: Review: hepatitis delta. J Gastroenterol Hepatol 12:289, 1997

13. Krawczynski K, Aggarwal R, Kamili S: Hepatitis E. Infect Dis Clin North Am 14:669, 2000

14. Deka N, Sharma MD, Mukerjee R: Isolation of the novel agent from human stool samples that is associated with sporadic non-A, non-B hepatitis. J Virol 68:7810, 1994

15. Linnen J, Wages J Jr, Zhang-Keck ZY, et al: Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science 271:505, 1996

16. Simons JN, Leary TP, Dawson GJ, et al: Isolation of novel virus-like sequences associated with human hepatitis. Nat Med 1:564, 1995

17. Cheung RC, Keeffe EB, Greenberg HB: Hepatitis G virus: is it a hepatitis virus? West J Med 167:23, 1997

18. Hutin YJ, Pool V, Cramer EH, et al: A multistate, foodborne outbreak of hepatitis A. National Hepatitis A Investigation Team. N Engl J Med 340:595, 1999

19. Staes CJ, Schlenker TL, Risk I, et al: Sources of infection among persons with acute hepatitis A and no identified risk factors during a sustained community-wide outbreak. Pediatrics 106:E54, 2000

20. Ruddy SJ, Johnson RF, Mosley JW, et al: An epidemic of clam-associated hepatitis. JAMA 208:649, 1969

21. Dentinger CM, Bower WA, Nainan OV, et al: An outbreak of hepatitis A associated with green onions. J Infect Dis 183:1273, 2001

22. Corey L, Holmes KK: Sexual transmission of hepatitis A in homosexual men: incidence and mechanism. N Engl J Med 302:435, 1980

23. Katz MH, Hsu L, Wong E, et al: Seroprevalence of and risk factors for hepatitis A infection among young homosexual and bisexual men. J Infect Dis 175:1225, 1997

24. Kuo G, Choo Q-L, Alter HJ, et al: An assay for circulating antibodies to a major etio-

logic virus of human non-A, non-B hepatitis. Science 244:362, 1989

25. Alter MJ: Epidemiology of hepatitis C. Hepatology 26(suppl 1):62S, 1997

26. Liang TJ, Rehermann B, Seeff LB, et al: Pathogenesis, natural history, treatment, and prevention of hepatitis C. Ann Intern Med 132:296, 2000

27. Dienstag JL: Sexual and perinatal transmission of hepatitis C. Hepatology 26(suppl 1): 66S, 1997

28. Pereira BJ, Wright TL, Schmid CH, et al: A controlled study of hepatitis C transmission by organ transplantation. The New England Organ Bank Hepatitis C Study Group. Lancet 345:484, 1995

29. Hepatitis E among U.S. travelers, 1989–1992. MMWR Morb Mortal Wkly Rep 42:1, 1993

30. Gretch DR: Diagnostic tests for hepatitis C. Hepatology 26(suppl 1):43S, 1997

31. Lok ASF, Gunaratnam NT: Diagnosis of hepatitis C. Hepatology 26(suppl 1):48S, 1997

32. Hoofnagle JH: Hepatitis C: the clinical spectrum of disease. Hepatology 26(suppl 1): 15S, 1997

33. de Franchis R, Meucci G, Vecchi M, et al: The natural history of asymptomatic hepatitis B surface antigen carriers. Ann Intern Med 118:191, 1993

34. Seeff LB: Natural history of hepatitis C. Hepatology 26(suppl 1):21S, 1997

35. Kiyosawa K, Sodeyama T, Tanaka E, et al: Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 12:671, 1990

36. Tong MJ, el-Farra NS, Reikes AR, et al: Clinical outcomes after transfusion-associated hepatitis C. N Engl J Med 332:1463, 1995

37. Wiese M, Berr F, Lafrenz M, et al: Low frequency of cirrhosis in a hepatitis C (genotype 1b) single-source outbreak in Germany: a 20-year multicenter study. Hepatology 32: 91, 2000

38. Rodger AJ, Roberts S, Lanigan A, et al: Assessment of long-term outcomes of community-acquired hepatitis C infection in a cohort with sera stored from 1971 to 1975. Hepatology 32:582, 2000

39. Seeff LB, Buskell-Bales Z, Wright EC, et al: Long-term mortality after transfusion-associated non-A, non-B hepatitis. The National Heart, Lung, and Blood Institute Study Group. N Engl J Med 327:1906, 1992

40. Poynard T, Bedossa P, Opolon P: Natural history of liver fibrosis progression in patients with chronic hepatitis C. The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. Lancet 349:825, 1997

41. Harris DR, Gonin R, Alter HJ, et al: The relationship of acute transfusion-associated hepatitis to the development of cirrhosis in the presence of alcohol abuse. Ann Intern Med 134:120, 2001

42. Fattovich G, Giustina G, Degos F, et al: Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. Gastroenterology 112:463, 1997

43. Martin P: Hepatocellular carcinoma: risk factors and natural history. Liver Transpl Surg 4(5 suppl 1):S87, 1998

44. Schiff GM: Hepatitis caused by viruses other than hepatitis A, hepatitis B, and non-A, non-B hepatitis viruses. Diseases of the Liver, 7th ed. Schiff L, Schiff ER, eds. JB Lippincott Co, Philadelphia, 1993, p 578

45. Markin RS, Linder J, Zuerlein K, et al: Hepatitis in fatal infectious mononucleosis. Gastroenterology 93:1210, 1987

46. Weller TH: The cytomegaloviruses: ubiquitous agents with protean clinical manifestations. I. N Engl J Med 285:203, 1971

47. Farrell GC: Drug-induced hepatic injury. J Gastroenterol Hepatol 12:S242, 1997

48. Makin AJ, Wendon J, Williams R: A 7-year experience of severe acetaminophen-induced hepatotoxicity (1987–1993). Gastroenterology 109:1907, 1995

49. Smilkstein MJ, Knapp GL, Kulig KW, et al: Efficacy of oral *N*-acetylcysteine in the treatment of acetaminophen overdose: analysis of the National Multicenter Study (1976 to 1985). N Engl J Med 319:1557, 1988

50. Chalmers TC, Eckhardt RD, Reynolds WE, et al: The treatment of acute infectious hepatitis: controlled studies of the effects of diet, rest, and physical reconditioning on the acute course of the disease and on the incidence of relapses and residual abnormalities. J Clin Invest 34:1163, 1955

51. Repsher LH, Freebern RK: Effects of early and vigorous exercise on recovery from infectious hepatitis. N Engl J Med 281:1393, 1969

52. Quin JW: Interferon therapy for acute hepatitis C viral infection: a review by meta-analysis. Aust N Z J Med 27:611, 1997

53. Orland JR, Wright TL, Cooper S: Acute hepatitis C. Hepatology 33:321, 2001

54. Gregory PB, Knauer CM, Kempson RL, et al: Steroid therapy in severe viral hepatitis: a double-blind, randomized trial of methyl-prednisolone versus placebo. N Engl J Med 294:681, 1976

55. Redeker AG, Schweitzer IL, Yamahiro HS: Randomization of corticosteroid therapy in fulminant hepatitis (letter). N Engl J Med 294:728, 1976

56. A double-blinded, randomized trial of hydrocortisone in acute hepatic failure (abstr). Acute Hepatic Failure Study Group. Gastroenterology 76:1297, 1979

57. Ware AJ, Cuthbert JA, Shorey J, et al: A prospective trial of steroid therapy in severe viral hepatitis. Gastroenterology 80:219, 1981

58. MacDougall BRD, Bailey RJ, Williams R: H_2-receptor antagonists and antacids in the prevention of acute gastrointestinal hemorrhage in fulminant hepatic failure. Lancet 1:617, 1977

59. Failure of specific immunotherapy in fulminant type B hepatitis. Acute Hepatic Failure Study Group. Ann Intern Med 86:272, 1977

60. Redeker AG, Yamahiro HS: Controlled trial of exchange-transfusion therapy in fulminant hepatitis. Lancet 1:3, 1973

61. Wall W, Adams PC: Liver transplantation for fulminant hepatic failure: North American experience. Liver Transplant Surg 1:178, 1995

62. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 48(RR-12):1, 1999

63. Hepatitis B virus: a comprehensive strategy for eliminating transmission in the United States through universal childhood vaccination: Recommendations of the Immunization Practices Committee (ACIP). MMWR Morb Mortal Wkly Rep 40(RR-13):1, 1991

64. Update: recommendations to prevent hepatitis B virus transmission—United States. MMWR Morb Mortal Wkly Rep 48:33, 1999

65. Suboptimal response to hepatitis B vaccine given by injection into the buttock. MMWR Morb Mortal Wkly Rep 34:105, 1985

66. Alper CA, Kruskall MS, Marcus-Bagley D, et al: Genetic prediction of nonresponse to hepatitis B vaccine. N Engl J Med 321:708, 1989

67. Weissman JY, Tsuchiyose MM, Tong MJ, et al: Lack of response to recombinant hepatitis B vaccine in nonresponders to the plasma vaccine. JAMA 260:1734, 1988

68. Wainwright RB, McMahon BJ, Bulkow LR, et al: Duration of immunogenicity and efficacy of hepatitis B vaccine in a Yupik Eskimo population. JAMA 261:2362, 1989

69. Horowitz MM, Ershler WB, McKinney WP, et al: Duration of immunity after hepatitis B vaccination: efficacy of low-dose booster vaccine. Ann Intern Med 108:185, 1988

70. Are booster immunisations needed for lifelong hepatitis B immunity? European Consensus Group on Hepatitis B Immunity. Lancet 355:561, 2000

71. Perrillo RP, Parker ML, Campbell C, et al: Prevaccination screening of medical and dental students: should low levels of antibody to hepatitis B surface antigen preclude vaccination? JAMA 250:2481, 1983

72. Kessler HA, Harris AA, Payne JA, et al: Antibodies to hepatitis B surface antigen as the sole hepatitis B marker in hospital personnel. Ann Intern Med 103:21, 1985

73. Werner BG, Dienstag JL, Kuter BJ, et al: Isolated antibody to hepatitis B surface antigen and response to hepatitis B vaccination. Ann Intern Med 103:201, 1985

74. Innis BL, Snitbhan R, Kunasol P, et al: Protection against hepatitis A by an inactivated vaccine. JAMA 271:1328, 1994

75. Werzberger A, Mensch B, Kuter B, et al: A controlled trial of formalin-inactivated hepatitis A vaccine in healthy children. N Engl J Med 327:453, 1992

76. Lemon SM: Inactivated hepatitis A vaccines (editorial). JAMA 271:1363, 1994

77. Berge JJ, Drennan DP, Jacobs RJ, et al: The cost of hepatitis A infections in American adolescents and adults in 1997. Hepatology 31:469, 2000

78. Steffen R, Kane MA, Shapiro CN, et al: Epidemiology and prevention of hepatitis A in travelers. JAMA 272:885, 1994

79. Keeffe EB: Is hepatitis A more severe in patients with chronic hepatitis B and other chronic liver diseases? Am J Gastroenterol 90:201, 1995

80. Vento S, Garofano T, Renzini C, et al: Fulminant hepatitis associated with hepatitis A virus superinfection in patients with chronic hepatitis C. N Engl J Med 338:286, 1998

81. Keeffe EB, Iwarson S, McMahon BJ, et al: Safety and immunogenicity of hepatitis A vaccine in patients with chronic liver disease. Hepatology 27:881, 1998

Acknowledgments

Figure 1 Alan D. Iselin.

Figure 2 Seward Hung.

VIII CHRONIC HEPATITIS

Peter F. Malet, m.d., f.a.c.p.

Definition

Chronic hepatitis is a term that encompasses an etiologically diverse group of clinical and pathologic diseases. Chronic hepatitis is characterized by the presence of hepatic inflammation on liver biopsy and elevation of serum liver enzymes, especially transaminases.[1] Chronic hepatitis is generally defined as disease that has lasted for 6 months or longer; in many cases, however, the diagnosis can be established earlier.

Etiology

The most important diseases that cause chronic hepatitis are (1) autoimmune hepatitis (AIH) (previously called autoimmune chronic active hepatitis), (2) chronic hepatitis B, which is caused by infection with hepatitis B virus (HBV), and (3) chronic hepatitis C, caused by hepatitis C virus (HCV) [*see Table 1*].

The hepatitis D virus (HDV) may also be present in some patients with HBV infection. Chronic overuse of alcohol may result in chronic hepatic inflammation (alcohol-induced liver disease or alcoholic hepatitis). Nonalcoholic steatohepatitis (NASH) is similar histologically to alcohol-induced liver disease and is discussed more fully elsewhere [*see 4:IX Cirrhosis of the Liver*]. Less commonly, chronic hepatitis is cryptogenic or caused by drugs. Wilson disease, α_1-antitrypsin deficiency, primary biliary cirrhosis (PBC), and primary sclerosing cholangitis (PSC) are also chronic liver diseases characterized by hepatocellular or bile duct inflammation, but the term chronic hepatitis is not generally used to describe these conditions. Over the past decade, international working groups have substantially modified the terminology of chronic hepatitis to reflect an etiologic basis rather than a pathologic basis.[1] As a result, previously used terms, such as chronic active hepatitis and chronic persistent hepatitis, are no longer used.

Table 1 Major Types of Autoimmune and Chronic Viral Hepatitis

Type/Subtype	*Diagnosis*
Autoimmune	
Type 1 (classic)	(+) ANA
	(+) ASMA
Type 2	(+) anti-LKM1
Type 3	(+) anti-SLA
	(+) anti-LP
Chronic hepatitis B	(+) HBsAg
	(+) anti-HBc
	(-) anti-HBs
	(+) HBV DNA > 100,000 copies/ml
(+) e antigen	(+) HBeAg
	(-) anti-HBe
(-) e antigen	(-) HBeAg
	(+) anti-HBe
Chronic hepatitis C	(+) anti-HCV
	(+) HCV RNA

ANA—antinuclear antibody ASMA—anti–smooth muscle antibody HBc—hepatitis B c antigen HBeAG—hepatitis B e antigen HBsAG—hepatitis B s antigen HBV—hepatitis B virus HCV—hepatitis C virus LKM—liver/kidney microsome LP—liver/pancreas antigen SLA —soluble liver antigen

Approach to Chronic Hepatitis

DIAGNOSIS

Clinical Manifestations

Clinical manifestations of chronic hepatitis are diverse, ranging from asymptomatic disease characterized by mildly elevated aminotransferase levels to severe, rapidly progressive illness and fulminant hepatic failure. The most common symptoms of chronic hepatitis are fatigue, malaise, and mild abdominal discomfort. Patients with mild chronic hepatitis are usually asymptomatic or have minimal symptoms with no stigma of chronic liver disease on physical examination. In more advanced cases, when hepatic synthetic function begins to diminish (a condition referred to as hepatic decompensation), the symptoms and signs may include anorexia, jaundice, spider angiomas, palmar erythema, ascites, edema, and encephalopathy. Pruritus may occur but is unusual; pruritus is more characteristic of PBC or PSC. A small proportion of patients with AIH have an acute fulminant course and are critically ill at initial presentation. Some extrahepatic manifestations of chronic hepatitis include arthralgias, arthritis, glomerulonephritis, and skin rashes.

Standard Laboratory Tests

The serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels are usually elevated in patients with chronic hepatitis; however, a minority of patients, especially those who have chronic hepatitis B or C, may have either persistently or transiently normal aminotransferase levels. Therefore, even a mild elevation of aminotransferase levels (5 to 10 IU/L higher than the upper limit of normal) should lead physicians to consider the presence of chronic hepatitis. Elevations of more than 400 IU/L are not unusual in cases of AIH and may occasionally be seen with chronic hepatitis B. Aminotransferase levels that are twofold to threefold higher than the upper limit of normal are common with chronic hepatitis C. The serum bilirubin level is usually normal in chronic viral hepatitis, unless hepatic decompensation has occurred. The serum bilirubin level is higher than 3 mg/dl in patients with moderately severe AIH. A characteristic feature of AIH is an increased γ-globulin level (> 1.6 g/dl), which may sometimes be markedly increased (3 to 7 g/dl). In the most severe forms of chronic hepatitis, hepatic synthetic function is impaired; this is manifested by a decreased serum albumin level and a prolonged prothrombin time.

Imaging studies (i.e., ultrasonography and computed tomography) of the abdomen may be normal in early chronic hepatitis or may show variable degrees of hepatomegaly with or without splenomegaly. When more advanced hepatic fibrosis is present, irregularity of liver texture or contour may be seen. Evidence of portal hypertension with portal collateral vessels or ascites may be seen with advanced disease.

Liver Biopsy

The specific etiology of chronic hepatitis can usually be determined by clinical evaluation combined with serologic testing.

Table 2 METAVIR Scoring System for Grading Severity of Chronic Hepatitis

Grade/Stage	*Grade of Inflammation*	*Stage of Fibrosis*
1	Minimal	Portal
2	Mild	Periportal, rare bridging
3	Moderate	Bridging
4	Severe	Cirrhosis

Liver biopsy can confirm or exclude certain diagnoses and can establish the grade of inflammatory activity and the stage of fibrosis and cirrhosis.

The grade and stage of chronic hepatitis can be assessed with various semiquantitative scoring systems.[1,2] In the histology activity index (HAI) (also known as the Knodell score), the grades of inflammation range from 0 to 18, and the stages of fibrosis range from 0 to 4. The HAI is sometimes used in clinical research studies but is no longer commonly used in clinical practice.[2] The most popular scoring system for chronic hepatitis is the METAVIR system.[3] This system generates two scores from 0 to 4: one for the degree of inflammation and the other for the degree of fibrosis [*see Table 2*]. Other scoring systems are available but are not commonly used in clinical practice. In addition to the characterization of the degree of inflammation and fibrosis, the evaluation of a liver biopsy specimen from a patient with chronic hepatitis includes a description of all the findings present (e.g., steatosis, bile duct changes, granulomas, type of inflammatory cells, presence of iron, and Mallory hyaline) or a notation of their absence.

DIFFERENTIAL DIAGNOSIS

Primary Biliary Cirrhosis

On liver biopsy, PBC may have features similar to those of AIH but is differentiated by the presence of bile duct inflammation or ductopenia. Marked elevation of serum alkaline phosphatase and high titers of antimitochondrial antibody (AMA) (> 1:160) are useful in making the diagnosis of PBC.

Primary Sclerosing Cholangitis

PSC can sometimes mimic chronic hepatitis. Prominent elevations of the serum alkaline phosphatase level and accompanying inflammatory bowel disease will, in most cases, distinguish this disorder from other types of chronic hepatitis. The definitive diagnosis is made by endoscopic retrograde or magnetic resonance cholangiography.

Alcoholic Liver Disease

Liver enzyme elevations caused by alcohol overuse are very commonly encountered in clinical practice. When the patient admits to the overuse of alcohol, the diagnosis is straightforward; however, careful and repeated questioning of the patient and family members may sometimes be required to obtain an accurate assessment of the patient's alcohol use.

Fatty Liver and Nonalcoholic Steatohepatitis

Fatty liver and nonalcoholic steatohepatitis (NASH) are increasingly common liver diseases that are usually associated with obesity or being overweight. Fatty liver is characterized by macrovesicular and, sometimes, microvesicular steatosis without inflammation; serum transaminase levels may be normal or elevated, sometimes severalfold or more. NASH is characterized by hepatocellular inflammation, steatosis, and, usually, elevated serum transaminase levels. NASH may lead to fibrosis or to cirrhosis. Although radiologic studies are useful in the evaluation of suspected fatty liver and NASH, a definitive diagnosis may only be made with liver biopsy.

Drug-Induced Chronic Hepatitis

Drug-induced chronic hepatitis constitutes a small but important category of chronic hepatitis.[4] α-Methyldopa, nitrofurantoin, and isoniazid are well-recognized causes. In addition, cases have occasionally been reported after therapy with sulfonamides, propylthiouracil, diclofenac, terbinafine, and dantrolene; a number of other drugs have also been implicated. Thus, it is reasonable to discontinue as many medications as possible when chronic hepatitis is first diagnosed. If a patient's hepatitis is drug related, liver function abnormalities and the clinical course of disease frequently improve after the causative agent has been withdrawn, although improvement may take weeks or even months.

Wilson Disease

When neurologic abnormalities are absent, Wilson disease can present as chronic hepatitis. It is critical to establish the diagnosis of Wilson disease, because specific treatment with penicillamine, trientine, or zinc is available. In Wilson disease, the serum ceruloplasmin level is low and the 24-hour urinary copper level is elevated. Slit-lamp examination for Kayser-Fleischer rings should also be performed. Measurement of the hepatic copper content in a needle-biopsy specimen is diagnostic.

α_1-Antitrypsin Deficiency

α_1-Antitrypsin deficiency, which is usually associated with the presence of homozygous *ZZ* alleles, is associated with progressive liver disease, which evolves into cirrhosis. Liver disease associated with α_1-antitrypsin deficiency can be distinguished from chronic hepatitis by a reduced serum α_1-antitrypsin level and by inclusions in the liver parenchyma that are positive on periodic acid–Schiff (PAS) staining.

Autoimmune Hepatitis

AIH is characterized by portal and periportal inflammation and fibrosis, autoantibodies, and hypergammaglobulinemia.[5] The early recognition of AIH is important because the condition generally responds well to treatment; if left untreated, it can progress to cirrhosis and, occasionally, liver failure and death.

DIAGNOSIS

The diagnosis of AIH rests on the presence of characteristic findings combined with the exclusion of other causes of chronic liver disease. The presentation may be acute or subacute but is more commonly chronic. Other autoimmune diseases may be present concurrently.

Elevated levels of serum transaminase and γ-globulin are typical in AIH. Aminotransferase levels may be slightly elevated or more than 10 times higher than normal. The presence of antinuclear antibody (ANA) or one of the other autoantibodies is common.

A liver biopsy is always useful to establish histologic grading

and staging; it is essential in cases in which the diagnosis is not clear on the basis of clinical and laboratory data. Interface hepatitis (formerly called piecemeal necrosis) is the histologic hallmark of AIH; however, this finding is not specific for AIH. The inflammatory component of AIH consists mainly of mononuclear cells; typically, plasma cells are present, but occasionally, they may not be a prominent feature.

In addition to its diagnostic use, liver biopsy is also an important prognostic tool. Patients with portal or mild periportal hepatitis (i.e., hepatitis that extends outside the limiting plate of the portal tract) generally respond well to therapy, whereas those with bridging necrosis (hepatitis that extends or bridges from one portal tract to another), multilobular necrosis, or cirrhosis respond less well to therapy and are at greater risk for progressive liver disease.

The International Autoimmune Hepatitis Group (IAIHG) has published a diagnostic scoring system for atypical AIH for use in adults when the diagnosis is uncertain.[6] The pretreatment score is based on 12 features, and the posttreatment score includes the response to therapy. A pretreatment score of 10 to 15 signifies probable AIH, whereas a score greater than 15 indicates definite AIH. A posttreatment score of 12 to 17 signifies probable AIH, whereas a score greater than 17 indicates definite AIH.

Disease Types

The wide spectrum of clinical and serologic manifestations of AIH has led investigators to propose several types of AIH[5,7] The distinctions in type are based on seropositive findings; however, it should be noted that in some cases of AIH, no serologic markers are present. All the proposed serologic types of AIH respond similarly to immunosuppressive therapy. Type 1, or classic, AIH is the most common form of the disease in the United States. It is characterized by hypergammaglobulinemia and the presence of ANA, anti–smooth muscle antibody (ASMA), or both. Type 2 AIH is characterized by the absence of ANA and ASMA and by the presence of antibody to liver/kidney microsome (anti-LKM1); it is much less common than type 1 and has been observed primarily in Europe. Type 3 is the least well characterized form of AIH, and it is distinguished by the presence of antibody to a soluble liver antigen (anti-SLA), to a recently characterized liver/pancreas antigen (anti-LP), or to both.

Table 3 Typical Regimens for Treatment of Autoimmune Hepatitis

	Schedule	*Dose*
Monotherapy	Initial dose	Prednisone, 60 mg p.o., q.d.
	Second week	Prednisone, 40 mg p.o., q.d.
	Third and fourth weeks	Prednisone, 30 mg p.o., q.d.
	Thereafter until end point is reached	Prednisone, 20 mg p.o., q.d.
Combination Therapy	Initial dose	Prednisone, 30 mg p.o., q.d. Azathioprine, 50 mg p.o., q.d.
	Second week	Prednisone, 20 mg p.o., q.d. Azathioprine, 50 mg p.o., q.d.
	Third and fourth weeks	Prednisone, 15 mg p.o., q.d. Azathioprine, 50 mg p.o., q.d.
	Thereafter until end point is reached	Prednisone, 10 mg p.o., q.d. Azathioprine, 50–100 mg p.o., q.d.

Overlap Syndromes

Overlap syndromes of AIH and either PBC or PSC have been recognized but are uncommon.[8,9] The term autoimmune cholangitis has been proposed to characterize patients who have biochemical or histologic cholestasis that resembles PBC. These patients test negative for AMA and have a normal IgM level; however, they have high titers of ANA and elevated levels of IgG. Results of cholangiography are normal.

TREATMENT

The standard treatment of AIH is immunosuppressive therapy. Three large randomized, controlled trials evaluated immunosuppressive treatment in patients with severe AIH[10-12]; however, data regarding treatment of patients with mild to moderate AIH are less extensive. The potential benefit of immunosuppressive therapy has to be balanced against the risks, particularly in cases of mild AIH. Most patients who have the clinical, biochemical, and histologic features of AIH should be treated,[5] but treatment may not be appropriate for some patients; for example, patients with advanced cirrhosis and relatively mild abnormalities of serum aminotransferase levels (less than twice the upper limit of normal) are probably not good candidates for such treatment.

The purpose of treatment is to relieve symptoms, decrease hepatic inflammation, and prevent the progression of hepatic fibrosis. The decision whether to initiate treatment can sometimes be problematic, particularly in cases that are atypical (e.g., when test results are negative for all autoantibodies). In such cases, the IAIHG scoring system can be useful in decision making.

Prednisone and Azathioprine Regimens

The usual regimen for most AIH patients consists of a combination of prednisone and azathioprine [*see Table 3*].[5] The starting doses are usually 30 mg of prednisone daily and 50 mg of azathioprine daily. Maintenance doses (until remission is achieved) are usually 20 mg of prednisone daily and 50 to 150 mg of azathioprine daily. Prednisone alone may be used with equal effectiveness; it is administered at a starting dose of 60 mg daily and is tapered over 4 weeks to a maintenance dose of 20 mg daily. Use of azathioprine may not be appropriate in patients who have severe cytopenia (because of the drug's bone marrow suppressive effect), who are pregnant (because of its potential to cause birth defects), or who have active malignancy (because of its potential to interfere with standard cancer treatment regimens). Azathioprine alone will not induce remission, but long-term azathioprine therapy is effective in maintaining remission.[13]

Response to Therapy

Clinical, biochemical, and histologic remission occurs in 65% of patients within 18 months and in 80% of patients within 3 years after starting treatment. In general, symptoms resolve (i.e., clinical remission) within 3 months, serum transaminase levels improve to normal or less than twice normal (i.e., biochemical remission) within 3 to 6 months, and histologic improvement (i.e., histologic remission) occurs within 18 months to 3 years. The presence of serum autoimmune markers, such as ANA, does not influence the initial response to therapy. Failure to achieve remission may represent either incomplete response to therapy or treatment failure. Patients who have an incomplete response to therapy may experience some improvement in clinical, laboratory, and histologic features when adequate therapy is administered. In cases of treatment failure, patients will expe-

rience worsening of disease despite adequate doses and compliance with therapy. Drug toxicities that necessitate the reduction of drug doses may contribute to the failure to achieve remission.

Discontinuance of Therapy

It is not clear-cut when to taper medication doses for the purpose of discontinuance of therapy. One commonly used approach is to begin tapering drugs when clinical and biochemical remissions have been achieved. Some authorities recommend a repeat liver biopsy to determine whether histologic remission has been achieved; however, this approach is not routine practice. When prednisone and azathioprine are used in combination, the dose of prednisone is usually tapered first while that of azathioprine is kept constant at 50 to 150 mg daily. Gradual tapering should occur over several months; however, there is no agreement on the rapidity of the tapering process. Once prednisone is completely discontinued, gradual tapering of azathioprine may begin.

Management of Relapse

Despite every precaution, relapse occurs within 3 to 6 months in approximately 20% to 90% of patients. Relapse is less likely if the histologic findings before tapering show that the hepatic inflammatory activity has completely resolved. If relapse occurs, medication doses should be increased in an attempt to induce remission. Once a patient has a relapse, however, the risk of future relapses is significant.

When a patient has had two or more relapses, a change in treatment approach is warranted[5]; with a change of medications, lower doses can be used to induce remission. After relapse, the aim of therapy is to keep disease activity as quiescent as possible. An acceptable therapeutic end point is a reduction of the serum transaminase level to two times normal or less. Such a serum transaminase level may be achieved with prednisone alone or in combination with azathioprine. Between 80% and 90% of patients can be maintained on a daily dosage of 10 mg of prednisone or less. When both drugs are used, the dose of prednisone can be gradually tapered while that of azathioprine is held constant at approximately 2 mg/kg/day; this combination regimen has a success rate in maintaining disease quiescence that is similar to that of prednisone alone. Limited data are available on the use of drugs such as 6-mercaptopurine and mycophenolate mofetil for the treatment of AIH.

Chronic Hepatitis B

EPIDEMIOLOGY

Chronic hepatitis B is a major global health care problem: 5% of the world's population, or approximately 350 million persons, are chronically infected.[14,15] In the United States, it is estimated that 1.25 million persons are chronically infected.[14] Approximately 0.2% to 0.5% of the United States population is positive for hepatitis B surface antigen (HBsAg); however, chronic HBV infection rates five to 10 times higher have been identified for certain groups, including Asian Americans, immigrants from endemic areas, persons who have received multiple blood transfusions or hemodialysis, intravenous drug users, men who have sex with men, and persons with HIV infection. Age at the time of initial HBV infection is the major determinant of chronicity. As many as 90% of infected neonates develop chronic infection; however, only 3% to 5% of newly infected adults develop chronic infection. Another important risk factor for chronicity is the presence of intrinsic or iatrogenic immunosuppression. Gender is also a well-established but poorly understood determinant of chronicity; women are more likely than men to clear HBsAg. As a result, men predominate in all populations with chronic HBV infection.

Table 4 Typical Serologic Findings in Hepatitis B Infection

Inactive Carrier State (low viral replication)	*Chronic Hepatitis B (high viral replication)*
(+) HBsAg	(+) HBsAg
(+) anti-HBc	(+) anti-HBc
(–) anti-HBs	(–) anti-HBs
(–) HBeAg	(+) HBeAg or (–) HBeAg
(+) anti-HBe	(+) HBeAg and (–) anti-HBe
HBV DNA < 100,000 copies/ml	HBV DNA > 100,000 copies/ml, typically in the range of 1 million to 10 million copies/ml
ALT and AST normal	ALT and AST elevated

ALT—alanine aminotransferase AST—aspartate aminotransferase

Ongoing HBV infection is an important risk factor for the development of hepatocellular carcinoma (HCC). The relative risk of HCC may be 200 times higher in patients with chronic HBV infection than in the general population. Patients with advanced fibrosis and cirrhosis are at highest risk. Evidence suggests that the use of screening tests, including ultrasonography and serum α-fetoprotein, may be useful in the early diagnosis of HCC.[14,16]

DIAGNOSIS

Chronic HBV infection can be readily diagnosed with serologic testing [*see Table 4*]. The diagnosis of chronic HBV infection is made when HBsAg remains detectable for more than 6 months. In a small percentage of patients with chronic HBV infection, mainly those who are inactive HBsAg carriers (i.e., 1% to 2% of patients a year), HBsAg clears spontaneously.

Categories of Hepatitis Infection

There are two broad categories of chronic HBV infection. The terminology to describe these states varies; the American Association for the Study of Liver Disease (AASLD) practice guidelines refer to these categories as the inactive carrier state and chronic hepatitis B.[14,15] A person who is positive for HBsAg, has normal aminotransferase levels, and has little or no necroinflammatory hepatic activity is an inactive HBsAg carrier. In inactive HBsAg carriers, the serum HBV DNA usually is relatively low ($< 10^5$ copies/ml). However, there are some inactive carriers in whom the serum HBV DNA may be much higher (often in the range of 10^7 to 10^{10} copies/ml), and this has been termed the immune-tolerant state. The immune-tolerant state is most characteristic of patients who were infected perinatally, most commonly in southern Asia.

The other broad disease category defined by AASLD practice guidelines is chronic hepatitis B. A person who is positive for HBsAg, has elevated aminotransferase levels, and has significant necroinflammatory hepatic activity is classified as having chronic hepatitis B. In patients with chronic hepatitis B, the serum HBV DNA usually is relatively high ($> 10^5$ copies/ml). In chronic hepatitis B, the HBeAg is usually positive (and anti-HBe negative); however, there is an increasing prevalence of HBeAg-negative disease. HBeAg is negative in patients with chronic he-

patitis B because of a genetic mutation in the HBV DNA genome affecting the production of HBeAg. Chronic hepatitis B is subdivided into HBeAg-positive and HBeAg-negative disease. Typically, HBV DNA levels are somewhat lower in HBeAg-negative disease, and seroconversion from HBeAg to anti-HBe cannot be used as an end point for therapy in these patients.

TREATMENT

The ultimate goal of treatment of chronic hepatitis B is to eradicate HBV infection and prevent the development of cirrhosis or HCC. Interferon, lamivudine, and adefovir dipivoxil [*see Table 5*] can suppress HBV replication and lead to improvement in the clinical, biochemical, and histologic features of chronic hepatitis B.[14,15] The two oral agents, lamivudine and adefovir, are well tolerated. Interferon has a number of potential side effects, and careful consideration must be given to its use.

All patients with chronic hepatitis B should be considered for treatment.[17] In the majority of patients who are not considered suitable candidates for therapy, the severity of disease will be deemed too mild to warrant treatment. The most severely decompensated cirrhotic patients should be treated with lamivudine because of its extremely good tolerability. For those without severely decompensated cirrhosis, the decision whether to use interferon or oral therapy with lamivudine or adefovir is based on a variety of factors, not the least of which is patient preference. Oral therapy is very well tolerated; however, it typically requires several years of treatment to achieve a seroconversion rate comparable to that which can be achieved by interferon therapy in 6 to 9 months. On the other hand, interferon therapy has many more potential side effects and is considerably more expensive. Generally, interferon treatment is appropriate for those patients able to tolerate its potentially adverse effects.

Terminology Used in Describing Response to Treatment

In HBeAg-positive chronic hepatitis B, the term seroconversion indicates the loss of HBeAg and the appearance of anti-HBe. Seroconversion from HBsAg to anti-HBs can also occur, either spontaneously or after interferon treatment. Response to treatment can be (1) biochemical, as evidenced by the normalization of ALT levels; (2) virologic, as evidenced by a marked decrease in HBV DNA levels ($< 10^5$ copies/ml); or (3) histologic, as evidenced by improvement in or resolution of hepatic necroinflammatory activity. Such responses may be seen either during therapy or after therapy has been stopped. The response to treatment is considered a sustained response if remission is maintained for at least 6 months post treatment.

Interferon Therapy

Therapy with standard interferon has been most extensively studied; however, pegylated interferon (interferon alfa-2a and alfa-2b) is being used increasingly in clinical practice because of its once-weekly administration and seroconversion rates that are as good as or better than those of standard interferon regimens.

A large number of trials have demonstrated the efficacy of interferon alfa in the treatment of HBeAg-positive chronic hepatitis B.[14] Doses of interferon alfa have been in the range of 5 million units daily to 10 million units three times a week. Treatment for 16 to 24 weeks results in seroconversion from HBeAg to anti-HBe, a low HBV DNA replicative state in about 35% of patients, and loss of HBsAg and the appearance of anti-HBs in approximately 8% of patients.[17] Predictors of a response to interferon therapy include lower pretreatment HBV DNA levels, higher ALT levels (optimally greater than five times the upper limit of normal), and short duration of infection; patients with such characteristics are optimal candidates for treatment with interferon. Relapse and reappearance of HBeAg occur in about 20% of patients. Patients who fail to seroconvert on interferon therapy should be treated with oral agents. It should be noted that interferon therapy is not equally effective in Asian patients, particularly in those whose ALT levels are only minimally elevated.

Patients with mildly decompensated cirrhosis can be treated with low, titrated doses of interferon, although the safety and tolerability of lamivudine or adefovir would make either of these drugs the first choice for treatment. One third of such patients will respond to therapy and have a sustained loss of HBV DNA and HBeAg, which may be associated with resolution of cirrhotic symptoms. These patients must be monitored closely, however, because bacterial infections and exacerbation of hepatitis are potentially serious complications.

Interferon treatment of HBeAg-negative chronic hepatitis B is less well studied. After 12 months of treatment, a sustained response (normalization of ALT levels and low levels of HBV DNA) can be achieved in about 15% to 30% of patients.[17,18]

The use of pegylated interferon has recently been examined in patients with HBeAg-positive and HBeAg-negative chronic hepatitis B.[19,20] In one study, 194 HBeAg-positive patients were randomized to receive 24 weeks of therapy with 90, 180, or 270 μg of pegylated interferon alfa-2a once a week or with standard interferon alfa-2a, 4.5 mIU three times a week.[19] At 24 weeks' follow-up, the HBeAg seroconversion rates were 37%, 35%, and 29%, respectively, for the three pegylated interferon alfa-2a treatment groups and 25% for the standard-interferon group. In another study of HBeAg-negative chronic hepatitis B, 537 patients were randomized to receive 48 weeks of therapy with 180 μg of pegylated interferon alfa-2a; lamivudine, 100 mg daily; or the two regimens combined.[20] At 24 weeks after cessation of therapy, the rates of biochemical and virologic remission were significantly higher in the groups receiving pegylated interferon; the addition of lamivudine did not improve the response rate. The combination of pegylated interferon and lamivudine continues to be studied, but it is too early to draw conclusions about the efficacy of this treatment.[21]

Table 5 Treatment of Chronic Hepatitis B

Drug	*Dose*
Lamivudine	100 mg p.o., q.d.
Adefovir dipivoxil	10 mg p.o., q.d.
Interferon alfa-2a or alfa-2b	5 million units S.C. daily or 10 million units S.C. three times weekly
Pegylated interferon alfa-2a	180 mg/wk S.C.*
Pegylated interferon alfa-2b	1.5 mg/kg/wk S.C.*

*Optimal dose and duration of treatment are under study.

Side effects of interferon therapy Interferon commonly causes side effects, but these are usually manageable. Among these are influenzalike symptoms (i.e., fever, myalgia, arthralgia, and headache), hematologic toxicity (i.e., granulocytopenia, leukopenia, and thrombocytopenia), systemic symptoms (i.e., fatigue and hair thinning), neurologic signs (i.e., decreased concentration, depression, and irritability), and thyroid dysfunction.

Contraindications to interferon therapy Patients with a history of hypersensitivity to interferon, decompensated cirrhosis,

immunosuppression associated with organ transplantation, active autoimmune disease, or severe psychiatric disease or patients who are elderly or frail are not good candidates for treatment.

Lamivudine Therapy

Lamivudine is a nucleoside analogue that inhibits HBV DNA synthesis. A dose of 100 mg/day achieves maximal suppression of HBV DNA. Lamivudine is cleared mainly in urine, and thus, dose adjustments are required for patients with significant renal failure. It has very few side effects and could be considered for use in virtually any patient with chronic hepatitis B.[14]

Lamivudine and HBeAg-positive chronic hepatitis B In three placebo-controlled studies, improved liver histology occurred in a significantly higher percentage of patients given lamivudine than in patients who received placebo.[22-24] Improvements in liver histology were similar in treatment-naive patients and patients who experienced relapse after interferon therapy or who did not respond to it; the improvements occurred independently of HBeAg seroconversion.[25,26] Serum HBV DNA levels fell rapidly and remained at least 94% below baseline values; serum ALT levels also decreased during therapy, with 50% of patients achieving and maintaining normal ALT levels 2 years post treatment. Patients receiving lamivudine for 1 year experienced a 17% rate of seroconversion from HBeAg to anti-HBe. The seroconversion rate increased progressively with additional years of lamivudine therapy, resulting in 27% seroconversion after 2 years, 33% after 3 years, 47% after 4 years, and 50% after 5 years.[27-30] The cumulative HBeAg seroconversion rate was higher in patients who had elevated baseline ALT levels before treatment. The responses to lamivudine are similar in Asian and non-Asian patients; however, the durability of HBeAg seroconversion appears to be lower in Asian patients (i.e., 60% to 80%). Patients who achieve HBeAg seroconversion can discontinue lamivudine therapy. However, on the basis of limited data, relapse rates may be lower (i.e., seroconversion is more durable) if lamivudine is continued for 3 to 6 months after initial HBeAg seroconversion is documented. The serum levels of ALT and HBV DNA return to pretreatment levels if lamivudine is discontinued before HBeAg seroconversion is achieved. After treatment, some patients may experience serum ALT levels that are transiently higher than pretreatment levels. Generally, no adverse effects have been associated with these elevations, although there are rare reports of severe flares of hepatitis B.

Lamivudine and HBeAg-negative chronic hepatitis B Lamivudine is effective in lowering ALT and HBV DNA levels and improving hepatic histology in patients with HBeAg-negative chronic hepatitis B. Several studies have demonstrated a biochemical and virologic response rate of about 70% at 1 year of therapy[14,17]; however, this response rate decreases to 50% after 2 years and 40% after 3 years. The end point for treatment is unknown, but it is the consensus that treatment beyond 1 year is warranted, provided the patient continues to exhibit a response.[15] The relapse rate after discontinuance of therapy is much higher in these patients than in patients with HBeAg-positive chronic hepatitis B.

Viral resistance to lamivudine therapy Resistant strains of HBV may appear within the first year of lamivudine therapy. The most common mutation imparting resistance occurs near the YMDD (the amino acid sequence tyrosine-methionine-aspartate-aspartate) motif of the HBV DNA polymerase. Resistance to lamivudine therapy occurs in 14% to 32% of patients after 1 year of therapy, 38% after 2 years, 49% after 3 years, 66% after 4 years, and 69% after 5 years.[30] The appearance of lamivudine resistance is manifested by rising HBV DNA and serum transaminase levels. A serologic test detecting HBV mutations imparting lamivudine resistance is available. Generally, patients experiencing lamivudine resistance can be switched to adefovir, which is effective against HBV DNA mutant strains. If patients who develop lamivudine resistance are continued on lamivudine, the initial beneficial effect of treatment on disease activity is usually lost.

Lamivudine and end-stage liver disease Lamivudine plays a role in patients with chronic hepatitis B who have end-stage liver disease. Stabilization and even improvement of biochemical and clinical features of cirrhosis may be seen in a majority of patients. Occasionally, improvement in liver function may result in deferral of liver transplantation.[17,31,32] Lamivudine therapy may serve as a bridge to transplantation for patients with decompensated cirrhosis who are awaiting a donor liver.

Lamivudine and liver transplantation Liver transplantation can be performed for liver failure associated with chronic hepatitis B, but HBV infects the allograft in 80% to 100% of cases if antiviral prophylaxis is not given, and long-term graft survival is only 45% to 50% (compared with 80% to 85% in liver transplant patients with other types of cirrhosis).[32] The HBV reinfection often is accelerated and progresses to cirrhosis. As a result, most transplant centers now implement prophylactic antiviral strategies to reduce reinfection. Before undergoing transplantation, patients should be treated with either lamivudine or adefovir to reduce the HBV viral load. Subsequently, two prophylaxis strategies are (1) the intraoperative, immediately postoperative, and long-term administration of high-dose hepatitis B immune globulin (HBIG) to maintain an anti-HBs level at 100 to 200 mIU/ml or greater, and (2) the administration of HBIG in combination with lamivudine or adefovir.[32] A number of trials suggest that these prophylaxis strategies can reduce the HBV reinfection rate to 10% to 20% and improve 1-year and 3-year graft survival rates.

Adefovir Dipivoxil Therapy

Adefovir dipivoxil is a nucleotide analogue of adenosine monophosphate. Adefovir inhibits HBV DNA polymerase and reverse transcriptase. It is effective against both wild-type and lamivudine-resistant HBV and in both HBeAg-positive and HBeAg-negative chronic hepatitis B.[14,15] Typically, adefovir therapy results in a 3 to 4 $\log_{10}$ drop in serum HBV DNA levels.

In a randomized study involving 515 patients who had HBeAg-positive chronic hepatitis B, treatment with adefovir 10 mg daily for 48 weeks resulted in a 12% HBeAg seroconversion rate.[33] As with lamivudine, the rate of treatment response is higher in patients who have higher pretreatment ALT values. Longer-term studies of adefovir are ongoing, and preliminary results indicate that the HBeAg seroconversion rate increases after 2 years. Adefovir, particularly at higher doses, may result in some renal impairment; therefore, it is prudent to periodically monitor renal function in patients receiving adefovir therapy. Adefovir has not been well studied in patients with chronic hepatitis B and decompensated cirrhosis.

Adefovir therapy in patients with HBeAg-negative chronic hepatitis B results in about a 46% biochemical and virologic response rate after 48 weeks and a 51% response rate after 96

weeks. If therapy is stopped after 48 weeks, relapse occurs in more than 90% of patients.[34,35] Development of drug resistance is much less of a problem with adefovir than with lamivudine. Resistance to adefovir is seen in about 2% of patients after 2 years and 4% of patients after 3 years.

Entecavir Therapy

Entecavir was approved by the Food and Drug Administration in early 2005 for use in the treatment of chronic hepatitis. The efficacy of entecavir is at least comparable to that of lamivudine (at a dose of 0.5 mg orally daily) in previously untreated patients. It is also effective, at a dose of 1 mg orally daily, in patients who develop lamivudine resistance.

Tenofovir Therapy

Tenofovir is related to adefovir. Tenofovir has been approved by the FDA for treatment of HIV infection. Its use in the treatment of chronic hepatitis B is being evaluated, but very limited data have been published.

MANAGEMENT OF HBV REACTIVATION

In some inactive HBsAg carriers, biochemical, clinical, and histologic exacerbations of disease activity have been noted; such exacerbations are characterized by elevated serum transaminase levels, the presence of HBV DNA, and the reversion from anti-HBe to HBeAg. These exacerbations (so-called reactivations) appear spontaneously in about 20% of inactive HBsAg carriers. Repeated reactivations, which are usually asymptomatic, may occur and may lead to progressive fibrosis. Therefore, even inactive HBsAg carriers should be periodically monitored (i.e., every 6 to 12 months) for resurgence of disease activity.

Reactivation may also occur in patients with malignancy during or after cessation of chemotherapy, especially when the chemotherapy regimens include corticosteroids. It is important to be aware of such a possibility because reactivations in these patients, although generally mild, may at times be severe and even fatal. Preemptive therapy with lamivudine in patients with malignancy has been found to decrease the incidence and severity of reactivations.[36,37] Lamivudine therapy should be strongly considered in inactive-HBsAg-carrier patients who are scheduled to have chemotherapy. Patients with chronic hepatitis B who are scheduled to have chemotherapy would presumably already be receiving lamivudine or adefovir; if not, they should be started on lamivudine therapy before initiation of chemotherapy.

Chronic Hepatitis C

EPIDEMIOLOGY

Approximately 30,000 new cases of acute HCV infection are reported annually to the Centers for Disease Control and Prevention. Of these patients, about 80% develop chronic HCV infection.[38] Approximately four million persons in the United States (1.8%) have been infected with HCV, and 74% of these individuals (1.4% of the population) are viremic. The high chronicity rate of HCV infection makes chronic hepatitis C a much more prevalent disease than chronic hepatitis B (0.2% to 0.5% of the general population).

DIAGNOSIS

The diagnosis of chronic hepatitis C is typically made by a positive anti-HCV on enzyme-linked immunosorbent assay and detectable HCV RNA.[38] Serum transaminase levels are usually elevated; about 20% of patients have persistently normal values. Most infected patients are asymptomatic. Liver biopsy may demonstrate the full spectrum of disease severity, ranging from mild portal tract inflammation and no fibrosis to cirrhosis.

Table 6 Treatment of Chronic Hepatitis C

HCV Genotypes	*Drug Regimen*
For HCV genotype 1 (48-wk regimen)	Pegylated interferon alfa-2a, 180 µg/wk S.C., plus ribavirin, 1,000-1,200 mg/day *or* Pegylated interferon alfa-2b, 1.5 µg/kg/wk S.C., plus ribavirin, 800-1,200 mg/day
For HCV genotype 2 or 3 (24-wk regimen)	Pegylated interferon alfa-2a, 180 µg/wk S.C., plus ribavirin, 800 mg/day *or* Pegylated interferon alfa-2b, 1.5 µg/kg/wk S.C., plus ribavirin, 800 mg/day

NATURAL HISTORY

After the onset of acute HCV infection, the infection resolves in 15% to 30% of patients and there is a loss of HCV RNA, although anti-HCV remains detectable. The natural history of chronic hepatitis C typically spans several decades. In general, liver disease progresses insidiously, and cirrhosis may not develop for 2 or more decades, if ever. The natural history may be more prolonged when HCV infection occurs earlier in life. In a Japanese study, the mean interval from blood transfusion to development of chronic hepatitis was 10 years; to development of cirrhosis, 21 years; and to development of HCC (a late risk), 29 years.[39] Similar results were found in a population of patients seen in a referral liver center in the United States: the mean interval from transfusion to cirrhosis was 21 years; and for progression to hepatocellular carcinoma, the mean interval was 28 years.[40]

In a European study designed to examine the natural progression of hepatic fibrosis in patients with chronic hepatitis C, the median interval from the presumed time of infection to cirrhosis, identified by liver biopsy, was 30 years.[41] The rate of progression to fibrosis and cirrhosis was not normally distributed; findings suggested at least three populations of patients with chronic hepatitis C: those with rapid progression of fibrosis (median time to cirrhosis < 30 years), those with intermediate progression, and those with no or slow progression of fibrosis. Independent factors associated with an increased rate of progression to cirrhosis are age greater than 40 years at the time of infection, daily alcohol consumption of 50 g or more, and male gender. Fibrosis progression is not related to HCV RNA level or HCV genotype.

TREATMENT

The current standard of care for chronic hepatitis C involves the use of pegylated interferon (interferon alfa-2a or alfa-2b) in combination with ribavirin [*see Table 6*].[38] Polyethylene glycol (PEG) is a water-soluble polymer that is covalently linked to interferon, which markedly increases the half-life, resulting in sustained serum levels and allowing once-weekly administration. Combination therapy with standard (nonpegylated) interferon and ribavirin is still used occasionally. In patients with certain

comorbid conditions, monotherapy with interferon (either a pegylated or standard form) is advised because of the risk of severe anemia associated with ribavirin use [*see* Side Effects of Combination Therapy, *below*].

Pegylated Interferon and Ribavirin Therapy

The persistence of undetectable HCV RNA for 6 months or more after cessation of therapy is the definition of a sustained virologic response (SVR). The major pretreatment determinant of response is HCV genotype.[38] Patients infected with HCV genotype 1 virus (the most common genotype in the United States) have lower SVR rates (in the range of 42% to 52%) than patients with HCV genotypes 2 and 3 (in the range of 76% to 84%). There are varying responses to treatment of HCV genotypes 4, 5, and 6.

Varying study designs and ribavirin doses have been used to evaluate the efficacy of pegylated interferon (interferon alfa-2a and alfa-2b) and ribavirin combination therapy. Therefore, results of treatment have been reported that often do not reflect comparable treatment regimens, particularly with regard to ribavirin doses. Some studies have used a fixed dose of ribavirin, whereas others have used doses based on weight.[38,42,43] Despite these differences, SVR rates for patients with HCV genotypes 1, 2, and 3 have been similar using either interferon alfa-2a or interferon alfa-2b plus ribavirin at varying doses.

One randomized, double-blind trial clearly addressed the issues of length of therapy and ribavirin doses.[44] In that study, 1,284 patients with HCV genotypes 1, 2, and 3 were randomized to receive pegylated interferon alfa-2a (180 μg/wk) in combination with ribavirin in either a low-dose (800 mg/day) or standard weight-based dose (1,000 or 1,200 mg/day) for 24 or 48 weeks. The SVR for HCV genotype 1 patients was 52% after 48 weeks of treatment using the regimen with 1,000 to 1,200 mg of ribavirin daily versus 41% using the 800 mg ribavirin daily dose; the SVR rates for 24 weeks of treatment were much lower. The SVRs for patients with HCV genotypes 2 and 3 were similar, ranging from 79% to 84%, regardless of which treatment regimen was used. On the basis of the study's findings, 24 weeks of treatment with pegylated interferon and ribavirin at a dosage of 800 mg daily (which resulted in an SVR of 84%) is sufficient to achieve an optimal SVR in patients with HCV genotype 2 or 3. However, patients with HCV genotype 1 must be treated for 48 weeks using pegylated interferon and ribavirin at a dosage of 1,000 to 1,200 mg daily to achieve an optimal SVR. This and other studies have led to the practice of administering treatment for 48 weeks to patients with genotype 1 and for 24 weeks to those with genotype 2 or 3 [*see Table 6*]. Among factors shown to lessen the likelihood of successful therapy are high viral load ($> 2 \times 10^6$ copies/ml), male gender, African-American ethnicity, long duration of infection, and advanced fibrosis.

Early virologic response Because pegylated interferon and ribavirin combination therapy has many side effects, strategies have been studied to determine whether continuation of treatment after an early virologic response (EVR) will likely result in an SVR.[45,46] Such studies have mostly centered on patients with HCV genotype 1, because the standard treatment length is longer and the SVR rates are lower for genotype 1 than for genotypes 2 and 3. The values of HCV RNA at two time points (12 and 24 weeks) have been closely examined. If a patient with HCV genotype 1 does not have an undetectable level of HCV RNA after 24 weeks of treatment, the chance of achieving an SVR with another 24 weeks of treatment is extremely low (i.e., in the range of 1% or less). It is strongly recommended that treatment be halted in such patients, unless individual circumstances dictate otherwise.

Recent studies have focused on the 12-week HCV RNA value and have examined the negative predictive value for achieving an SVR based on various declines in HCV RNA, as compared with baseline values. For HCV genotype 1 patients, failure to achieve a drop of 2 $\log_{10}$ or greater in HCV RNA values after 12 weeks of treatment has a negative predictive value of approximately 97% to 100% for achieving SVR after a full course of 48 weeks of treatment. Clinical judgment must be used in deciding whether to discontinue therapy after failure to achieve a 12-week SVR. The value of determining a 12-week EVR in patients who have HCV genotype 2 or 3 is unclear because virtually all these patients achieve a response at 12 weeks. The utility of determining a 4-week EVR for HCV genotype 2 or 3 patients is under study.

Side effects of combination therapy Side effects are varied and common with interferon and ribavirin combination therapy[38,47] and may necessitate discontinuance of therapy in 10% to 15% of patients. The major side effect of ribavirin is a dose-dependent hemolytic anemia, which is reversible and usually stabilizes after 6 weeks of treatment. For milder degrees of anemia, the ribavirin dose can be temporarily decreased. If the anemia is not corrected by a reduction in the ribavirin dose, epoetin can be used. If severe anemia develops, ribavirin must be discontinued, either temporarily or permanently.

Patients with preexisting moderate to severe anemia usually cannot tolerate the degree of hemolysis that occurs with ribavirin therapy, which can be dangerous. Moreover, patients with cardiovascular disease are particularly at risk should severe anemia develop during therapy, and very careful consideration would have to be given to the decision to initiate interferon monotherapy because of the drug's side effects. Patients with any degree of symptomatic cardiovascular disease are not candidates for interferon monotherapy. Patients with chronic hepatitis C and comorbid conditions preventing the use of ribavirin can be treated with interferon monotherapy, although this treatment has an SVR rate considerably lower than that of combination therapy. Ribavirin is teratogenic; therefore, patients must be advised against its use during pregnancy. Other side effects attributed to ribavirin are rash and nasal or sinus problems.

Interferon can cause or exacerbate depression and other psychiatric symptoms. Caution is warranted in using interferon therapy for patients with depression; however, these patients can often be managed with antidepressants, with or without reduction in interferon dosage. Interferon has many other potential side effects, including leukopenia and thrombocytopenia, thyroid dysfunction, insomnia, hair thinning, headaches, weight loss, various neurologic dysfunctions, and irritability. Deaths have occurred, usually because of sepsis, suicide, or cardiovascular disease.

Sustained virologic response In 5- to 10-year follow-up studies of patients who had an initial posttreatment 6-month SVR, relapse occurred in 1% to 2 % of patients.[48] Following an SVR, there is usually improvement of liver histology, including significant regression of hepatic fibrosis in about 25% of patients. Interferon therapy may delay or prevent liver decompensation resulting from cirrhosis, as well as the development of HCC, particularly in patients who show a sustained response.[49]

Relapse after Treatment

Relapse after a course of therapy is defined by detectable levels of HCV RNA, which were undetectable during treatment. Relapse usually occurs within the first 6 months after treatment. No effective treatment is available to patients in whom relapse occurs after the standard regimen of pegylated interferon and ribavirin.

Nonresponse to Therapy

Nonresponse to antiviral therapy is defined as detectable serum HCV RNA levels during therapy. A variant of nonresponse called breakthrough is characterized by an initial disappearance of HCV RNA, with subsequent reappearance of HCV RNA while the patient is still being treated. There are no effective treatments available for patients who experience a nonresponse to current therapies.

Liver Transplantation

Cirrhosis caused by chronic hepatitis C is the most common indication for liver transplantation. Although hepatitis C virus reinfects the allograft in nearly all cases, the subsequent illness is usually mild, but a small percentage of cases progress to cirrhosis and liver failure. There is no consensus of opinion regarding guidelines for treatment of chronic hepatitis C after liver transplantation.[38] Usually, posttransplant treatment is more difficult because of more severe cytopenia and complications related to immunosuppression.

The author has received grants for clinical research from Roche Pharmaceuticals and Schering-Plough Corporation.

The drugs pegylated interferon alfa-2a, pegylated interferon alfa-2b, and azathioprine have not been approved by the FDA for uses described in this chapter.

References

1. Desmet VJ, Gerber M, Hoofnagle JH, et al: Classification of chronic hepatitis: diagnosis, grading and staging. Hepatology 19:1513, 1994

2. Brunt EM: Grading and staging the histopathological lesions of chronic hepatitis: the Knodell histology activity index and beyond. Hepatology 31:241, 2000

3. Bedossa P, Poynard T: An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. Hepatology 24:289, 1996

4. Lee WM: Drug-induced hepatotoxicity. N Engl J Med 349:474, 2003

5. Czaja AJ, Freese DK: Diagnosis and treatment of autoimmune hepatitis. AASLD Practice Guidelines. Hepatology 36:479, 2002

6. Alvarez F, Berg PA, Bianchi FB, et al: Review of criteria for diagnosis of autoimmune hepatitis. International Autoimmune Hepatitis Group report. J Hepatol 31:929, 1999

7. Czaja AJ, Manns MP: The validity and importance of subtypes of autoimmune hepatitis: a point of view. Am J Gastroenterol 90:1206, 1995

8. Czaja AJ: Frequency and nature of the variant syndromes of autoimmune liver disease. Hepatology 28:360, 1998

9. Ben-Ari Z, Czaja AJ: Autoimmune hepatitis and its variant syndromes. Gut 49:589, 2001

10. Soloway RD, Summerskill WHJ, Baggenstoss AH, et al: Clinical, biochemical, and histological remission of severe chronic active liver disease: a controlled trial of treatments and early prognosis. Gastroenterology 63:820, 1972

11. Murray-Lyon IM, Stern RB, Williams R: Controlled trial of prednisone and azathioprine in active chronic hepatitis. Lancet 1:735, 1973

12. Cook CG, Mulligan R, Sherlock S: Controlled prospective trial of corticosteroid therapy in active chonic hepatitis. QJ Med 40:159, 1971

13. Johnson PJ, McFarlane IG, Williams R: Azathioprine for long-term maintenance of remission in autoimmune hepatitis. N Engl J Med 333:958, 1995

14. Lok ASF, McMahon BJ: Chronic hepatitis B. AASLD Practice Guidelines. Hepatology 34:1225, 2001

15. Lok ASF, McMahon BJ: Chronic hepatitis B: update of recommendations. AASLD Practice Guidelines. Hepatology 39:1, 2004

16. McMahon BJ, Bulkow L, Harpster A, et al: Screening for hepatocellular carcinoma in Alaska natives infected with chronic hepatitis B: a 16-year population-based study. Hepatology 32:842, 2000

17. Keefe EB, Dietrich DT, Han S-HB, et al: A treatment algorithm for the management of chronic hepatitis B virus infection in the United States. Clin Gastroenterol Hepatol 2:87, 2004

18. Manesis EK, Hadziyannis SJ: Interferon alpha treatment and retreatment of hepatitis B e antigen-negative chronic hepatitis B. Gastroenterology 121:101, 2001

19. Cooksley W, Piratvisuth T, Lee SD, et al: Peginterferon alfa-2a (40 Kda): an advance in the treatment of hepatitis B e antigen-positive chronic hepatitis B. J Viral Hepatitis 10:298, 2003

20. Marcellin P, Lau GK, Bonino F, et al: Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. N Engl J Med 351:1206, 2004

21. Chan HL, Leung NW, Hui AY, et al: A randomized, controlled trial of combination therapy for chronic hepatitis B: comparing pegylated interferon-alpha2b and lamivudine with lamivudine alone. Ann Intern Med 142:240, 2005

22. Dienstag JL, Schiff ER, Wright TL, et al: Lamivudine as initial treatment for chronic hepatitis B in the United States. N Engl J Med 341:1256, 1999

23. Lai CL, Chien RN, Leung NW, et al: A one-year trial of lamivudine for chronic hepatitis B. Asia Hepatitis Lamivudine Study Group. N Engl J Med 339:61, 1998

24. Schalm SW, Heathcote J, Cianciara J, et al: Lamivudine and alpha interferon combination treatment of patients with chronic hepatitis B infection: a randomized trial. Gut 46:562, 2000

25. Honkoop P, de Man RA, Zondervan PE, et al: Histological improvement in patients with chronic hepatitis B virus infection treated with lamivudine. Liver 17:103, 1997

26. Suzuki Y, Kumada H, Ikeda K, et al: Histological changes in liver biopsies after one year of lamivudine treatment in patients with chronic hepatitis B infection. J Hepatol 30:743, 1999

27. Liaw YF, Leung NWY, Chang TT, et al: Effects of extended lamivudine therapy in Asian patients with chronic hepatitis B. Gastroenterology 119:172, 2000

28. Leung NWY, Lai CL, Chang TT, et al: Extended lamivudine treatment in patients with chronic hepatitis B enhances hepatitis B e antigen seroconversion rates: results after 3 years of therapy. Hepatology 33:1527, 2001

29. Chang TT, Lai CL, Liaw YF, et al: Incremental increases in HBeAg seroconversion and continued ALT normalization in Asian chronic HBV (CHB) patients treated with lamivudine for four years. Antiviral Therapy 16(suppl 1):44, 2000

30. Guan R, Lai CL, Liaw YF, et al: Efficacy and safety of 5-years lamivudune treatment of Chinese patients with chronic hepatitis B. J Gastroenterol Hepatol 16(suppl 1):A60, 2001

31. Villeneuve JP, Condreay LD, Willems B, et al: Lamivudine treatment for decompensated cirrhosis resulting from chronic hepatitis B. Hepatology 31:207, 2000

32. Keeffe EB: End-stage liver disease and liver transplantation: role of lamivudine therapy in patients with chronic hepatitis B. J Med Virol 61:403, 2000

33. Marcellin P, Chang TT, Lim SG, et al: Adefovir dipivoxil for the treatment of hepatitis B e antigen-positive chronic hepatitis B. N Engl J Med 348:808, 2003

34. Hadziyannis SJ, Tassopoulos N, Heathcote EJ, et al: Adefovir dipivoxil for the treatment of hepatitis B e antigen-negative chronic hepatitis B. N Engl J Med 348:800, 2003

35. Hadziyannis SJ, Tassopoulos N, Heathcote EJ, et al: Two year results from a double-blind, randomized, placebo-controlled study of adefovir dipivoxil (ADV) for presumed precore mutant chronic hepatitis B. J Hepatol 38(suppl 2):143, 2003

36. Lau GKK, He ML, Fong DYT, et al: Preemptive use of lamivudine reduces hepatitis B exacerbation after allogenic hematopoietic cell transplantation. Hepatology 36:702, 2002

37. Rossi G, Pelizzari A, Motta M, et al: Primary prophylaxis with lamivudine of hepatitis B reactivation in chronic HBsAg carriers with lymphoid malignancies treated with chemotherapy. Br J Haematol 115:58, 2001

38. Strader DB, Wright T, Thomas DL, et al: Diagnosis, management, and treatment of hepatitis C. AASLD Practice Guideline. Hepatology 39:1147, 2004

39. Kiyosawa K, Sodeyama T, Tanaka E, et al: Interrelationship of blood transfusion, non-A, non-B hepatitis and hepatocellular carcinoma: analysis by detection of antibody to hepatitis C virus. Hepatology 12:671, 1990

40. Tong MJ, El-Farra NS, Reikes AR, et al: Clinical outcomes after transfusion-associated hepatitis C. N Engl J Med 332:1463, 1995

41. Poynard T, Bedossa P, Opolon P, et al: Natural history of liver fibrosis progression in patients with chronic hepatitis C. Lancet 349:825, 1997

42. Fried MW, Shiffman ML, Reddy KR, et al: Combination of peginterferon alfa-2a plus ribavirin in patients with chronic hepatitis C virus infection. N Engl J Med 347:975, 2002

43. Manns MP, McHutchison JG, Gordon SC, et al: Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomized trial. Lancet 358:958, 2001

44. Hadziyannis SJ, Sette H, Morgan TR, et al: Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose. Ann Intern Med 140:346, 2004

45. Russo MW, Fried MW: Hepatitis C therapy: guidelines for stopping therapy. Curr Gastroenterol Rep 6:17, 2004

46. Davis GL, Wong JB, McHutchison JG, et al: Early virologic response to treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C. Hepatology 38:645, 2003

47. Russo MW, Fried MW: Side effects of therapy for chronic hepatitis C. Gastroeneterology 124:1711, 2003

48. Lau DTY, Kleiner DE, Ghany MG, et al: 10-Year follow-up after interferon-alpha therapy for chronic hepatitis C. Hepatology 28:1121, 1998

49. Yoshida H, Shiratori Y, Moriyama M, et al: Interferon therapy reduces the risk for hepatocellular carcinoma: national surveillance program of cirrhotic and noncirrhotic patients with chronic hepatitis C in Japan: IHIT Study Group: inhibition of hepatocarcinogenesis by interferon therapy. Ann Intern Med 131:174, 1999

IX CIRRHOSIS OF THE LIVER

Ramón Bataller, m.d.
Pere Ginès, m.d.

Cirrhosis is the most advanced stage of most types of chronic liver disease. It is defined as a diffuse disorganization of normal hepatic structure by extensive fibrosis associated with regenerative nodules. Fibrosis is potentially reversible if the causative agent is removed. However, advanced cirrhosis comprises major alterations in the hepatic vascular bed and is usually irreversible.[1] Clinically, cirrhosis is associated with high morbidity and mortality. It leads to a wide spectrum of characteristic clinical manifestations, mainly from hepatic insufficiency and portal hypertension.[2] Major complications include ascites, renal failure, gastrointestinal bleeding, encephalopathy, bacterial infections, and coagulopathy. Cirrhosis is also a risk factor for developing hepatocellular carcinoma (HCC). Decompensated cirrhosis carries a poor prognosis, in both the short and the long term, and orthotopic liver transplantation (OLT) is often indicated.

Epidemiology

Cirrhosis is the ninth leading cause of death in the United States.[3] Chronic liver disease and cirrhosis cause 4% to 5% of deaths in persons 45 to 54 years of age and result in about 30,000 deaths each year. The incidence of newly diagnosed cases of chronic liver disease in the United States is 72.3 per 100,000 population. The prevalence of chronic liver disease and cirrhosis is 5.5 million cases. Over 60% of patients are male. Cirrhosis is more common in Hispanic whites and Native Americans; it is the sixth leading cause of death in those two populations. The economic impact of cirrhosis is considerable, with $1.5 billion in direct costs and $234 million in indirect costs in 2000.[4] In 2002, there were 421,000 hospitalizations for chronic liver disease and cirrhosis.[5]

Etiology and Genetic Factors

In the United States, the main causes of cirrhosis are hepatitis C virus (HCV) infection and alcoholic liver disease, which account for two thirds of all cirrhosis cases. Other major causes are hepatitis B virus (HBV) infection, autoimmune hepatitis, chronic cholestasis (primary biliary cirrhosis [PBC] and primary sclerosing cholangitis [PSC]), and genetic metabolic diseases (hemochromatosis and Wilson disease) [*see Table 1*]. With the current epidemic of obesity, nonalcoholic steatohepatitis (NASH) is increasingly being recognized as a major cause of cirrhosis. Many patients diagnosed with cryptogenic cirrhosis have a history of metabolic syndrome, suggesting a role for NASH in the pathogenesis of their cirrhosis.

Many genes interact with environmental factors to cause cirrhosis.[6] Nongenetic factors that influence progression to cirrhosis include age, alcohol intake, immunosuppressive therapy, and HIV infection. Genetic factors involved in the pathogenesis of cirrhosis are not well known, but they may explain the broad spectrum of responses to the same etiologic agent found in patients with chronic liver disease. Polymorphisms in genes encoding immunoregulatory proteins, inflammatory cytokines, and fibrogenic mediators influence the occurrence of conditions that may cause chronic liver injury (e.g., alcohol abuse, chronic HCV infection, and autoimmune disorders), as well as modulate the progression of chronic hepatitis to cirrhosis.

Pathogenesis

EARLY PHASE: LIVER FIBROGENESIS

Cirrhosis is the end stage of many forms of chronic liver disease that are characterized by progressive fibrosis. Hepatic fibrosis is the result of the wound-healing response of the liver to repeated injury.[7] It consists of the accumulation of extracellular matrix (ECM) proteins, mainly fibrillar collagen, from both increased ECM synthesis and decreased degradation. Myofibroblasts, mostly derived from hepatic stellate cells, are the main ECM-producing cells in the injured liver. Chronic injury promotes the activation of stellate cells into fibrogenic myofibroblasts [*see Figure 1*]. Key mediators of this process include inflammatory cytokines, transforming growth factor–1 (TGF-1), and angiotensin II. The pathogenesis of liver fibrosis varies with the underlying cause. In alcohol-induced liver disease, lipopolysaccharide levels are elevated in portal blood; the lipopolysaccharide activates Kupffer cells to release reactive oxygen species and cytokines, activating stellate cells and sensitizing hepatocytes to undergo apoptosis. The pathogenesis of HCV-induced liver fibrosis is poorly understood. HCV infects hepatocytes, causing oxidative stress and inducing the recruitment of inflammatory cells. Both factors lead to stellate cell activation. In chron-

Table 1 Main Causes of Cirrhosis

Viral diseases	Hepatitis B Hepatitis C Hepatitis D
Autoimmune diseases	Autoimmune hepatitis Primary biliary cirrhosis Primary sclerosing cholangitis Graft versus host disease
Hepatotoxic agents	Alcohol abuse Drugs (e.g., methotrexate, α-methyldopa, amiodarone) Vitamin A intoxication
Acquired metabolic diseases	Nonalcoholic steatohepatitis
Vascular diseases	Chronic right-sided heart failure Budd-Chiari syndrome Veno-occlusive disease Inferior vena cava thrombosis
Genetic diseases	Wilson disease Hemochromatosis Type IV glycogen storage disease Tyrosinemia α_1-Antitrypsin deficiency
Miscellaneous	Secondary biliary cirrhosis Cryptogenic

a *b*

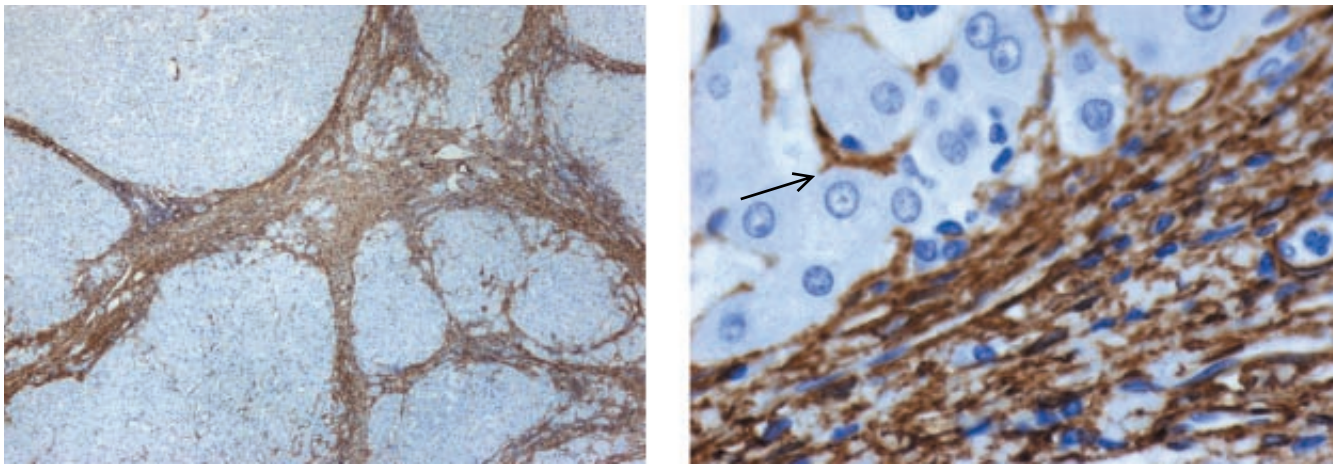


Figure 1 **Immunohistochemical analysis of accumulation of fibrogenic myofibroblasts (smooth muscle [α-actin–positive] cells) in a liver biopsy specimen from a 56-year-old man with liver cirrhosis from chronic hepatitis C infection. The patient was admitted for the study of new-onset ascites. Myofibroblasts mainly accumulate in fibrous septa. Some activated hepatic stellate cells can be observed around hepatic sinusoids (arrow). (*a*) Magnification: ×40; (*b*) magnification: ×600.**

ic cholestatic disorders, such as PBC, T cells and cytokines mediate persistent bile duct damage. Biliary cells secrete fibrogenic mediators that activate neighboring portal myofibroblasts to secrete ECM. Eventually, perisinusoidal stellate cells become activated and fibrotic bands develop. The pathogenesis of liver fibrosis in NASH is poorly understood, but a so-called two-hit model has been proposed: first, hyperglycemia and insulin resistance lead to elevated serum levels of free fatty acids, resulting in hepatic steatosis; second, oxidative stress and inflammatory cytokines promote hepatocyte apoptosis and the recruitment of inflammatory cells, leading to progressive fibrosis.

CIRRHOSIS

Bridging fibrosis is associated with profound abnormalities in hepatic microcirculation.[8] Capillarization of the hepatic sinusoids occurs, and new vessels form within the fibrous sheath. There is a local predominance of vasoconstrictors over vasodilators, resulting in a tonic contraction of perisinusoidal stellate cells that increases vascular resistance. Moreover, thrombosis in small vessels occurs and intrahepatic arterial shunts develop. Hepatocytes proliferate in ischemic areas in a disorganized manner, forming regenerative nodules. Pressure in the portal venous system progressively increases, leading to the development of portocollateral veins and esophageal varices.[9] The resulting portal hypertension leads to splanchnic vasodilatation, which increases hepatic venous blood flow. Systemic vascular resistance is decreased, and eventually, there is a marked activation of systemic vasoconstrictor systems that worsen portal hypertension and favor ascites formation. Hepatocellular function is progressively impaired, and there is decreased function of the reticuloendothelial system, leading to endotoxemia and the increased risk of bacterial infections. Eventually, hepatocellular function fails, resulting in severe coagulopathy and hepatic encephalopathy. A profound circulatory dysfunction from impaired myocardial function and decreased systemic vascular resistance is frequently seen. In very late stages of cirrhosis, renal vasoconstriction develops, leading to the hepatorenal syndrome (HRS). In this phase of the disease, most patients die unless an OLT is rapidly performed.

Diagnosis

The diagnostic process in a patient with suspected cirrhosis is intended to determine the presence, severity, and cause of the condition. Data obtained from the history, physical examination, laboratory tests, and liver biopsy are used to identify the etiology of cirrhosis [*see Table 2*].

CLINICAL MANIFESTATIONS

Cirrhosis can be clinically silent, and some cases are discovered incidentally at laparotomy or autopsy. In many patients, symptoms are insidious in onset and include generalized weakness, anorexia, malaise, and weight loss. Skeletal muscle mass is frequently reduced. So-called compensated cirrhosis is defined by the absence of symptoms or the presence of only minor symptoms. Eventually, most patients exhibit the clinical manifestations of hepatocellular dysfunction and portal hypertension, including progressive jaundice, bleeding from gastroesophageal varices, ascites, and neuropsychiatric symptoms. The abrupt onset of one of these complications may be the first manifestation of cirrhosis. Coagulopathy and subsequent mucosal bleeding typically occur in patients with advanced cirrhosis. Progressive obstruction to bile flow, which is especially common in patients with PBC and PSC, leads to skin hyperpigmentation, jaundice, pruritus, and xanthelasmas. Patients who have progressed to such conditions often experience malnutrition secondary to anorexia, fat malabsorption, and increased catabolism. Deficiency of fat-soluble vitamins is also frequently found in patients with cirrhosis. In patients with alcohol-induced liver disease, extrahepatic symptoms related to the nervous system, the heart, and the pancreas can also be present.

PHYSICAL EXAMINATION FINDINGS

Physical examination can be normal in patients with early cirrhosis. More commonly, the liver is enlarged initially and is palpable. In advanced cirrhosis, liver size usually decreases. Splenomegaly is a common finding. Ascites, peripheral edema, or both may be present, and collateral venous circulation can be observed in the abdomen. Patients with hepatic encephalopathy have altered mental status, decreased consciousness, and asterix-

is. Other signs typical of cirrhosis include muscle wasting, palmar erythema, vascular spiders, gynecomastia, axillary hair loss, testicular atrophy, and fetor hepaticus. In alcoholic patients, Dupuytren contractures, parotid gland enlargement, and peripheral neuropathy can be noted. Skin hyperpigmentation is typical of patients with cholestatic disorders (e.g., PBC) or hemochromatosis. Advanced cirrhosis is commonly marked by severe malnourishment, prominent ascites, and neuropsychiatric symptoms [*see Figure 2*].

LABORATORY STUDIES

Blood Tests

Liver function test results are commonly abnormal in patients with cirrhosis. Serum aspartate aminotransferase (AST) levels are frequently elevated, but levels above 300 U/L are uncommon. Serum levels of alanine aminotransferase (ALT) may be relatively low (AST/ALT ratio greater than 2). Serum prothrombin time is frequently prolonged, reflecting reduced synthesis of clotting proteins, most notably the vitamin K–dependent factors. Serum albumin levels are decreased, mainly because of poor hepatocellular synthesis. Total serum globulin concentration increases in advanced cirrhosis, as a result of poor reticuloendothelial function and increased blood levels of bacterial products. The alkaline phosphatase concentration is usually only moderately increased, except in patients with biliary diseases (i.e., PBC or PSC), who show markedly increased levels of alkaline phosphatase and γ-glutamyl transpeptidase, which in some cases are associated with increased bilirubin levels. Anemia is fairly common; it is usually normocytic, but it may be microcytic, hypochromic from chronic GI bleeding, macrocytic from folate deficiency (in alcoholism), or hemolytic. Hypersplenism can lead to leukopenia and thrombocytopenia. Cholesterol and triglyceride levels may be increased in patients with biliary obstruction, whereas they are low in patients with advanced cirrhosis of nonbiliary origin. Cirrhotic patients may develop glucose intolerance and diabetes mellitus, mainly because of insulin resistance. Central hyperventilation may lead to respiratory alkalosis. Dietary deficiency and increased urinary losses cause hypomagnesemia and hypophosphatemia. Renal failure, as indicated by elevated creatinine and blood urea nitrogen levels, and hyponatremia can be observed in cirrhotic patients with ascites.

Table 2 Identification of the Main Causes of Cirrhosis

Cause	*Diagnostic Method*
Alcohol-related	Medical history (also obtained from relatives), urinary alcohol levels, histologic findings
Hepatitis C virus (HCV)	Anti-HCV antibodies, HCV RNA assay
Hepatitis B virus (HBV)	HBsAg, HBV DNA assay
Hepatitis D virus (HDV)	Anti-delta IgM or IgG
Autoimmune disease	Antitissue antibodies (ANA, LKM, ASMA), hypergammaglobulinemia, histologic findings
Primary biliary cirrhosis	AMA, histologic findings
Primary sclerosing cholangitis	Severe cholestasis; detection of biliary tract abnormalities by magnetic resonance imaging/magnetic resonance cholangiopancreatography (MRI/MRCP) or ERCP; ANCA; presence of inflammatory bowel disease; histologic findings
Wilson disease	Serum ceruloplasmin levels, Kayser-Fleischer rings, copper content in the liver, genetic studies
Hemochromatosis	Serum ferritin levels, total iron binding capacity, iron content in the liver, genetic studies
Nonalcoholic steatohepatitis	Metabolic syndrome, histologic findings (may be absent in cirrhosis), absence of alcohol abuse

AMA—antimitochondrial antibodies ANA—antinuclear antibody ANCA—antineutrophil cytoplasmic autoantibodies ASMA—anti–smooth muscle antibodies ERCP—endoscopic retrograde cholangiopancreatography HBsAg—hepatitis B surface antigen LKM—liver/kidney microsomal antibody

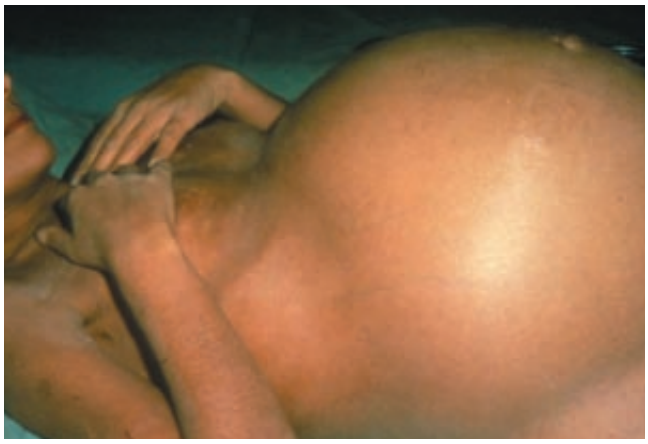


Figure 2 **Photograph of a 45-year-old patient with advanced cirrhosis from alcohol-induced liver disease. The patient was admitted because of tense ascites, and a superimposed acute alcoholic hepatitis was diagnosed. A large-volume paracentesis followed by albumin administration was performed.**

Imaging Studies

Real-time ultrasonography, in combination with color flow Doppler, is the most useful tool in the evaluation of patients with cirrhosis.[10] Ultrasonography is useful for demonstrating the morphologic characteristics of cirrhosis, including irregular or nodular liver edges, altered structure, and signs of portal hypertension such as portocollateral veins. It is also useful to detect hepatic steatosis, ascites, splenomegaly, and portal vein thrombosis. In patients with cholestasis, ultrasonography helps rule out extrahepatic causes of jaundice. Doppler ultrasonography provides useful information on portal hemodynamics and can detect reversal of portal blood flow [*see Figure 3*]. Ultrasound examination is particularly helpful for detecting hepatic tumors such as HCC. Demonstration of tumor vascularization by Doppler ultrasonography, with or without injection of ultrasound contrast, is valuable in the differentiation of regenerating nodules from HCC. Dynamic studies using computed tomography and magnetic resonance imaging are also useful in the assessment of cirrhosis and the diagnosis of hepatic tumors previously detected by ultrasonography. The use of CT or MRI to screen for HCC in patients with cirrhosis is limited by the high cost of these techniques.

Liver Biopsy

Liver biopsy can unequivocally establish the presence of cir-

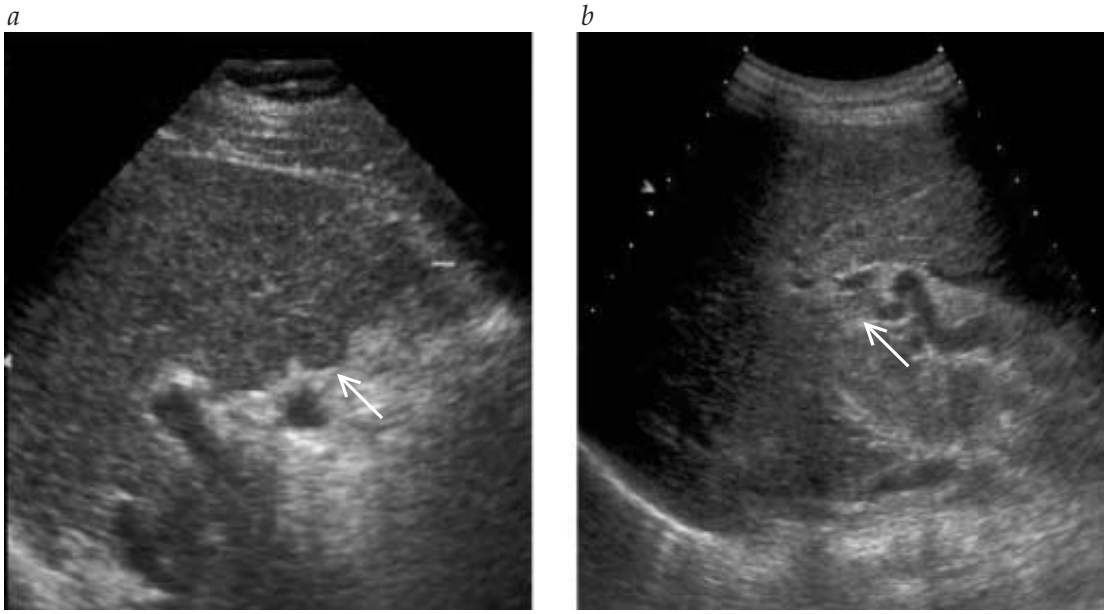


Figure 3 **Real-time ultrasound images of a 56-year-old man with liver cirrhosis from chronic hepatitis C. The patient had a compensated cirrhosis and was undergoing liver ultrasonography plus determination of α-fetoprotein serum levels every 6 months to screen for hepatocellular carcinoma. (*a*) The liver showed irregular edges (arrow) and an altered structure. (*b*) A patent portal vein thrombosis was detected (arrow).**

rhosis.[11] Liver biopsy helps determine the cause of cirrhosis, as well as provides information on the extent of liver damage. The biopsy is usually performed using a percutaneous approach, but percutaneous biopsy should not be used in patients with severe coagulopathy (i.e., those with an international normalized ratio [INR] greater than 1.5 or a platelet count less than 50,000/μl), and it must be used with caution in patients with ascites or severe obesity. Limitations of liver biopsy are that it is an invasive procedure and that sampling error can occur (i.e., false negative results), especially in patients with macronodular cirrhosis.

Transjugular liver biopsy offers an alternative to percutaneous biopsy. Transjugular liver biopsy can be used in patients with ascites; is indicated in patients with severe coagulopathy; and allows the measurement of portal pressure.[12] However, the amount of tissue obtained is limited, and often, the diagnosis of cirrhosis cannot be made. In selected cases, liver biopsy can be performed during laparoscopy. This approach is generally reserved for the staging of cancer or for ascites of unknown origin.

Histologic findings that define cirrhosis include extensive fibrosis and regenerative nodules. The degree of infiltration of inflammatory cells depends on the activity of the underlying disease. Micronodular cirrhosis is characterized by the presence of uniformly small nodules (diameter < 3 mm), whereas in macronodular cirrhosis, nodules vary in size (diameter 3 mm to 5 cm) and contain some normal lobular structure (e.g., portal tracts or terminal hepatic venules).

In some cases, histologic findings help identify the causative agent of cirrhosis, such as periportal lymphocyte infiltration in HCV-induced cirrhosis; Mallory bodies, polymorphonuclear leukocyte (PMN) infiltration, and steatosis in alcohol-induced cirrhosis and NASH; biliary involvement in PBC; and massive iron deposition in hemochromatosis. In advanced cirrhosis, however, different underlying diseases may have similar histologic findings.

Management

GENERAL MEASURES

Cirrhotic patients should undergo regular follow-up. Patients with compensated cirrhosis should be seen two or three times a year. At diagnosis, an extensive medical history should be taken and laboratory tests, including viral serologies, performed to identify the causative agent. Endoscopic examination should be done to assess the presence and size of esophageal varices. Abdominal ultrasonography and α-fetoprotein serum measurements should be performed at diagnosis and every 6 months thereafter to detect early HCC. Criteria for OLT should be reviewed periodically, and major clinical complications (i.e., bacterial infections, renal impairment, and GI bleeding) should be actively prevented.[13] Many patients complain of anorexia, and care should be taken to ensure that patients take in adequate calories and protein. Nutritional supplements are often beneficial. Zinc deficiency is common and should be treated. Zinc sulfate (50 to 200 mg/day) may be effective in the treatment of muscle cramps and is adjunctive therapy for hepatic encephalopathy. Pruritus is a common complaint in cirrhotic patients, especially in those with chronic cholestasis (i.e., PBC and PSC). Drugs that may provide relief for pruritus include cholestyramine, ursodeoxycholic acid, naltrexone, rifampicin, and ondansetron. Some men suffer from hypogonadism; those with severe symptoms can be treated with topical testosterone preparations, although their safety and efficacy are not well studied. Patients with cirrhosis may develop osteoporosis. Supplementation with calcium and vitamin D is important in patients at high risk for osteoporosis, especially patients with chronic cholestasis and those receiving corticosteroids for autoimmune hepatitis. Evidence of decreased bone mineralization from bone densitometry studies also may prompt institution of therapy with a bisphosphonate (e.g., alendronate). Mild exercise, including walking or swimming, should be encouraged in patients with compensated cirrhosis. Debilitated patients frequently benefit from formal exercise programs su-

pervised by a physical therapist. Patients with cirrhosis should be vaccinated against hepatitis A. Other protective measures include vaccination against hepatitis B, pneumococcal infection, and influenza. Potential hepatotoxic medications should be avoided. Patients with ascites should not receive nonsteroidal anti-inflammatory drugs (NSAIDs) or nephrotoxic antibiotics (e.g., aminoglycosides). NSAID use may predispose patients with cirrhosis to development of renal failure or GI bleeding. Surgery and general anesthesia carry increased risks in patients with cirrhosis, particularly those with portal hypertension, and may lead to hepatic decompensation.

COMPENSATED CIRRHOSIS

Specific medical therapies may be applied to different liver diseases to diminish disease progression. However, these therapies may become progressively less effective if chronic liver disease evolves into cirrhosis. In patients with compensated cirrhosis, specific therapies prevent the development of clinical complications and therefore delay the need for liver transplantation. Treatment with pegylated interferon plus ribavirin should be considered in patients with compensated cirrhosis from HCV infection, although the rate of sustained response is lower than in noncirrhotic patients.[14] Moreover, antiviral treatment may worsen existing anemia or thrombocytopenia, and drug discontinuance is frequent. In patients with HBV-related cirrhosis, lamivudine appears to be a safe and effective antiviral agent, which may improve or stabilize liver disease in selected patients with advanced cirrhosis and active HBV replication.[15] However, viral resistance can develop with prolonged treatment. Adefovir and entecavir are newer antiviral agents that have activity against both wild-type and lamivudine-resistant HBV. The most effective measure for patients with alcohol-induced cirrhosis is to stop drinking.[16] Abstinence can stabilize and may dramatically improve liver function. Psychological support is highly recommended to help patients achieve prolonged alcohol abstinence. Nutritional support is advisable in all alcoholic patients. Although small clinical trials have shown improvement in survival and reversal of cirrhosis with colchicine treatment, a randomized, controlled trial found that in patients with advanced alcoholic cirrhosis, there was no reduction in overall or liver-specific mortality with colchicine; although liver histology improved to septal fibrosis in a minority of patients after 24 months of treatment, rates of improvement were similar with placebo and colchicine.[17]

In cases of superimposed alcoholic hepatitis, treatment with glucocorticoids (40 mg/day for 4 weeks followed by tapering of therapy for 1 or 2 weeks) or pentoxifylline (400 mg three times daily) increases short-term survival.[18] In patients with PBC, ursodiol (13 to 15 mg/kg/day) relieves pruritus and improves blood chemistry test results.[19] Although ursodiol may decrease the need for OLT, its usefulness in cirrhotic patients is limited. Other treatments (e.g., glucocorticoids, colchicine, azathioprine) are not indicated, because they are associated with severe side effects. No specific therapy improves the outcome of patients with PSC, but ursodiol does have beneficial effects on biochemical parameters. In patients with cirrhosis from autoimmune hepatitis, immunosuppressant therapy (e.g., glucocorticoids) should be used with caution because it may favor infections, and necroinflammatory injury at this stage of the disease is usually mild. Patients with cirrhosis resulting from hemochromatosis benefit from phlebotomies to reduce iron stores [*see 5:II Red Blood Cell Function and Disorders of Iron Metabolism*]; those with Wilson disease benefit from treatment with copper chelators (i.e., D-penicillamine or trientine) or zinc [*see 11:XV Parkinson Disease and Other Movement Disorders*].

DECOMPENSATED CIRRHOSIS

Ascites

Ascites is the most frequent complication of cirrhosis.[20] It impairs quality of life and increases the risk of bacterial infections. It is caused primarily by splanchnic vasodilatation from increased synthesis of vasodilators (e.g., nitric oxide). Severe splanchnic vasodilatation decreases effective arterial blood volume, which activates systemic vasoconstrictor and sodium-retaining factors. In advanced cirrhosis, solute-free water excretion is also impaired and renal vasoconstriction develops, leading to dilutional hyponatremia and HRS, respectively. Ascites can be graded into three groups: grade 1 ascites is clinically silent and detectable only by ultrasonography; grade 2 ascites is moderate, with patent distention of the abdomen; and grade 3 ascites is tense, with marked abdominal distention.

The first step in the evaluation of patients with new-onset ascites is to rule out extrahepatic causes (e.g., tuberculosis and malignancies). Besides serum tests, ultrasonography is useful to confirm signs of cirrhosis, rule out HCC, and detect portal vein thrombosis. Ascitic fluid should be examined in patients with new-onset ascites, suspected spontaneous bacterial peritonitis (SBP), encephalopathy, or GI bleeding. Measurements should be done of cell counts, albumin and total protein concentrations, and culture in blood culture bottles. Renal function and circulatory status should also be assessed in all patients.

The initial management of ascites includes reduction of sodium intake to 60 to 90 mEq/day. In patients with dilutional hyponatremia (i.e., a serum sodium concentration of 130 mmol/L in the presence of ascites or edema), fluid intake should be restricted to less than 1,000 ml/day, although compliance is problematic. Patients with moderate-volume ascites can achieve a negative sodium balance and loss of ascitic fluid with spironolactone (50 to 200 mg/day) or amiloride (5 to 10 mg/day). Low doses of furosemide (20 to 40 mg/day) may be also added; however, patients should be followed closely to avoid excessive diuresis. The recommended weight loss to prevent renal failure is 300 to 500 g/day in patients without peripheral edema and 800 to 1,000 g/day in those with peripheral edema. Patients with large-volume ascites should be treated initially with large-volume paracentesis. Plasma expanders should be given to prevent paracentesis-induced circulatory dysfunction and renal failure.[21] Albumin is the plasma expander of choice if more than 5 L of ascitic fluid is removed (8 g of I.V. albumin for each 1 L of ascitic fluid removed). Spironolactone (100 to 400 mg/day), with or without furosemide (40 to 160 mg/day), can be given to prevent recurrence of ascites. Doses of diuretics should be adjusted according to diuretic response.

Refractory ascites, which is defined as a lack of response to high doses of diuretics or the occurrence of side effects (e.g., renal failure, encephalopathy, hyponatremia, or hyperkalemia) that preclude the use of diuretics, occurs in 5% to 10% of patients with ascites. Current therapeutic strategies for patients with refractory ascites include repeated large-volume paracentesis with plasma expanders and transjugular intrahepatic portosystemic shunting (TIPS) [*see Figure 4*]. TIPS is effective in preventing ascites recurrence, but it does not improve survival.[22] The principal drawbacks of TIPS are the high rates of shunt stenosis and hepatic en-

cephalopathy. The use of polytetrafluoroethylene-covered prostheses for TIPS can improve patency rates and decrease clinical relapses and the need for reintervention, without increasing the risk of encephalopathy.[23] TIPS is indicated for patients without severe liver failure or encephalopathy who have loculated fluid that cannot be treated with paracentesis and for those who do not tolerate repeated paracentesis. Patients with ascites should be evaluated for OLT, because their 5-year survival rate is only 30% to 40%. Patients with refractory ascites, SBP, or HRS have a worse prognosis, and prioritization in the waiting list should be considered.

Dilutional hyponatremia is present in 30% to 35% of hospitalized patients with cirrhosis and ascites.[24] It reflects impaired excretion of renal solute free water caused by nonosmotic hypersecretion of antidiuretic hormone. Although dilutional hyponatremia is commonly asymptomatic, it may favor the development of hepatic encephalopathy. Management consists of fluid restriction (1,000 ml/day) and discontinuance of diuretics; however, these measures do not correct hyponatremia in many cases. Vasopressin type 2 receptor antagonists are being evaluated for the management of hyponatremia, but they are not available in the United States.

Hepatorenal Syndrome

HRS is the most severe complication of patients with cirrhosis.[25] It is a functional renal failure resulting from extreme renal vasoconstriction [*see 10:VI Acute Renal Failure*]. HRS may occur spontaneously or after precipitating conditions such as SBP, acute alcoholic hepatitis, or large-volume paracentesis without plasma expansion. Diagnostic criteria for HRS have been established [*see Table 3*].[26]

There are two clinical types of HRS. Type 1 is characterized by progressive oliguria and a rapid rise of the serum creatinine concentration to more than 2.5 mg/dl. Survival of patients with type 1 HRS is extremely poor. Type 2 is defined by a moderate and stable increase in the serum creatinine concentration and is frequently associated with refractory ascites.

In type 1 HRS, the use of vasoconstrictors (e.g., terlipressin, midrodine, and norepinephrine) plus intravenous albumin improves renal function in more than half of patients.[27] TIPS is effective for patients with HRS, but its use is not recommended for patients with severe liver dysfunction. These treatments may serve as a bridge to OLT. Liver transplantation is the treatment of choice, but its applicability is limited by the poor survival of these patients.

Spontaneous Bacterial Peritonitis

SBP is a severe infection found in 15% to 25% of cirrhotic patients hospitalized with ascites.[28] Predisposing factors include severe liver insufficiency and low protein content in ascitic fluid (< 1 g/dl). SBP appears to be related to the translocation of GI tract bacteria from the mesenteric lymph nodes. Clinical manifestations are variable, ranging from no symptoms to a severe picture of peritonitis [*see 7:XXI Peritonitis and Intra-abdominal Abscesses*]. SBP should also be suspected in cirrhotic patients with impairment of renal or liver function that has no apparent cause.

The most common causative organisms are *Escherichia coli, Streptococcus pneumoniae, Klebsiella* species, and other gram-negative enteric organisms. SBP is diagnosed when the ascitic fluid has more than 250 PMNs/μl or is positive for leukocyte esterase (3+ or 4+) on urine dipstick testing. Culture results of ascitic fluid are positive in fewer than 50% of patients.

SBP should be treated empirically. The most commonly used regimen is a 5- to 7-day course of a third-generation cephalosporin (e.g., cefotaxime, 2 g every 8 to 12 hours, or ceftriaxone, 1 g/day).[29] Alternatives include oral ofloxacin and other intravenous antibiotics with activity against gram-negative enteric organisms. Development of renal failure during SBP is common and is the most important predictor of mortality. Administration of albumin at a dose of 1.5 g/kg at diagnosis and 1 g/kg 48 hours later prevents renal failure and reduces mortality from 30% to 10%.[30] Response to therapy is indicated by decreases in the signs of infection and in the PMN count in ascitic fluid. After SBP resolution, patients have a 70% chance of recurrence within 1 year. Prophylactic antibiotic therapy can reduce the recurrence rate of SBP to 20%.[31] Current prophylactic regimens include norfloxacin, 400 mg/day, and trimethoprim-sulfamethoxazole, one double-strength tablet 5 days a week. The 1-year survival probability after an episode of SBP is only 40%. Accordingly, eligible patients should be evaluated for OLT after resolution of SBP.

a

b

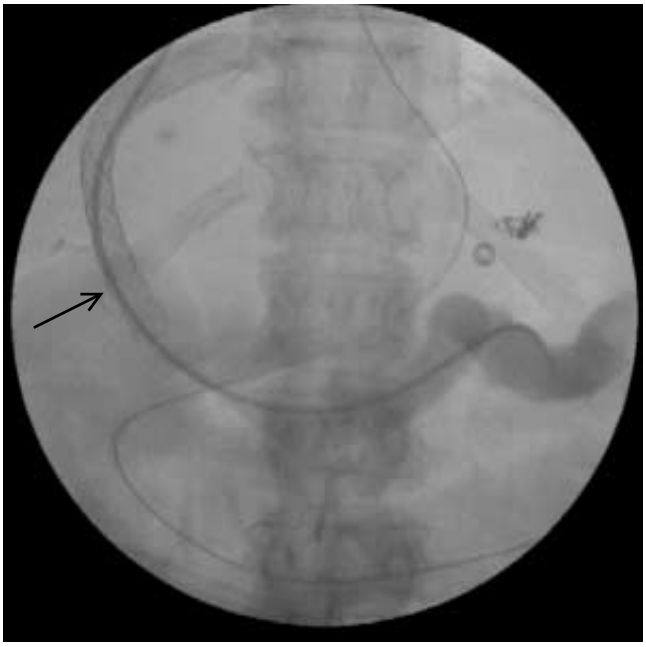


Figure 4 **Transjugular intrahepatic portosystemic shunting is basically indicated for patients with variceal bleeding and refractory ascites. It consists of (*a*) an autoexpandable stent, which is inserted using a transjugular approach. (*b*) The stent (arrow) creates a shunt between a portal vein branch and the inferior vena cava.**

Table 3 Diagnostic Criteria for Hepatorenal Syndrome*

Major Criteria
Low glomerular filtration rate, as indicated by serum creatinine level > 1.5 mg/dl or 24-hr creatinine clearance < 40 ml/min
Absence of shock, ongoing bacterial infection, fluid losses, and current treatment with nephrotoxic drugs
No sustained improvement in renal function (decrease in serum creatinine to ≤ 1.5 mg/dl or increase in creatinine clearance to ≥ 40 ml/min) after diuretic withdrawal and expansion of plasma volume with 1.5 L of a plasma expander
Proteinuria < 500 mg/day and no ultrasound evidence of obstructive uropathy or parenchymal renal disease
Additional Criteria
Urinary volume < 500 ml/day
Urinary sodium level < 10 mEq/L
Urinary osmolality > plasma osmolality
Urinary red blood cells < 50 cells/high-power field
Serum sodium concentration < 130 mEq/L

*According to the International Ascites Club.[26]

Variceal Bleeding

Rupture of gastroesophageal varices because of portal hypertension is a severe and frequent complication of cirrhosis.[32] Portal hypertension is caused by increased intrahepatic vascular resistance, as well as increased portal blood flow secondary to splanchnic vasodilatation. It is recommended that all patients with cirrhosis be screened for gastroesophageal varices and that those with large varices be offered primary prophylaxis. If esophageal varices are very small or absent, it is recommended that an endoscopic examination be performed every 2 years.

Primary prophylaxis of variceal bleeding should be initiated in patients with medium-size to large varices. Nonselective beta blockers (e.g., nadolol or propranolol) are the treatment of choice. They should be given in a stepwise fashion, with the dose being increased until the resting heart rate decreases by 25% of the baseline value. However, there are a number of limitations to the use of beta blockers in such patients (e.g., hypotension).[33] Alternatively, varices can be eradicated by repeated sessions of endoscopic variceal band ligation, although this is more commonly done in Europe than in the United States. Pharmacologic therapy and endoscopic therapy are similarly effective; both reduce the risk of bleeding by 40% to 50%.[34] Endoscopic treatment should be offered to cirrhotic patients in whom the use of beta blockers is contraindicated. The combination of variceal band ligation and beta blockers seems to be more effective than beta blockers alone and is being evaluated in large clinical trials.[35]

Acute variceal bleeding Initial therapy for acute variceal bleeding should be directed at correcting hypovolemia, achieving hemostasis, and preventing severe complications (e.g., renal failure, bacterial infections, and hepatic encephalopathy).[32] Volume replacement, as well as the need for blood transfusion, should be considered. Excessive transfusion should be avoided, because it may increase portal pressure and favor variceal rebleeding. In patients with hepatic encephalopathy and those requiring aggressive sedation for endoscopic examination, endotracheal intubation should be considered. Antibiotics (norfloxacin, 400 mg/day, or cefotaxime, 2 g every 12 hours; both for 7 days) decrease the rate of bacterial infections and improve outcome.[28] Hemostatic treatments include vasoactive drugs, endoscopic band ligation, and surgical portosystemic shunts or TIPS. Vasoactive drugs that are effective in controlling variceal bleeding include octreotide (100 μg bolus, followed by 50 μg/hr for 5 days); alternatives currently unavailable in the United States are terlipressin (2 mg every 4 hours for the first 48 hours, then 1 mg every 4 hours for up to 5 days) and somatostatin (bolus of 250 μg, followed by an infusion of 250 μg/hr for 5 days).[36]

Pharmacologic therapy controls variceal bleeding in 75% to 80% of cases. Cirrhotic patients with upper gastrointestinal bleeding should be initially treated with a vasoactive drug. If the endoscopic examination confirms that esophageal varices are the source of the hemorrhage, variceal band ligation should be performed and drug therapy maintained for 5 days to prevent early variceal rebleeding. This approach controls bleeding in most patients. In patients with massive bleeding, balloon tamponade may temporarily help in controlling the hemorrhage. A repeat session of therapeutic endoscopy can be performed in patients who rebleed. In patients who are hemodynamically unstable or who experience several rebleeding episodes, TIPS or surgical portosystemic shunts or both should be considered.[37,38] TIPS controls bleeding in more than 90% of cases and is preferred over shunt surgery because it is associated with lower morbidity and mortality. However, TIPS can impair liver function in patients with advanced cirrhosis. Patients with preserved liver function (Child-Pugh class A) may also benefit from shunt surgery (i.e., H-graft portacaval shunt or mesocaval shunt).

Because of the high rate of rebleeding (60%), secondary prophylaxis is recommended. Drug therapy with nonselective beta blockers and repeated sessions of variceal band ligation are similarly effective.[39] The beneficial effect of beta blockers should be confirmed by the hepatic venous pressure portal gradient (HVPG), if this is available. A reduction of HVPG to less than 12 mm Hg or by 20% protects patients from variceal rebleeding.[40] The combination of beta blockers and endoscopic band ligation seems to be more effective than either treatment used alone; this combined approach is being evaluated. TIPS, surgical portosystemic shunting, or both should be considered for patients who rebleed despite drug therapy and endoscopic treatment.

Hepatopulmonary Syndrome

The hepatopulmonary syndrome (HPS), which is characterized by hypoxemia from intrapulmonary shunting, a ventilation-perfusion mismatch, or both, develops in some patients with cirrhosis.[41] Patients with HPS have no apparent parenchymal lung disease but have orthodeoxia, the unusual finding of increased hypoxemia with the change from a supine to a standing position. Other typical manifestations include exertional dyspnea, platypnea, and digital clubbing.[42] The diagnostic workup includes arterial blood gas measurements, contrast-enhanced echocardiography, and scanning with technetium-99m–labeled macroaggregated albumin. Pulmonary angiography may be necessary to detect discrete arteriovenous communications. Pharmacologic agents, such as almitrine bismesylate, prostaglandin $F_{2\alpha}$, indomethacin, somatostatin, and methylene blue, have been used to treat HPS, but results have been disappointing. Although TIPS may improve oxygenation, OLT is the only curative treatment; by 6 months after OLT, about 80% of patients with HPS have improved oxygenation.[43]

Hepatic Encephalopathy

Hepatic encephalopathy is a syndrome observed in patients

with advanced cirrhosis that is marked by personality changes, intellectual impairment, neuromuscular dysfunction, and a depressed level of consciousness.[44] The pathogenesis involves altered brain-energy metabolism and increased permeability of the blood-brain barrier, facilitating the passage of neurotoxins.[45] Putative neurotoxins include short-chain fatty acids, mercaptans, false neurotransmitters (e.g., tyramine, octopamine, and β-phenylethanolamines), ammonia, and γ-aminobutyric acid.[46]

The diagnosis of hepatic encephalopathy is made on the basis of altered mental status and neuromuscular signs in the absence of any specific mental or neurologic disease. Hepatic encephalopathy is classified into five grades, according to clinical severity [*see Table 4*]. In addition, hepatic encephalopathy can be classified as episodic, persistent, or minimal. Minimal encephalopathy refers to patients with subtle manifestations of hepatic encephalopathy that cannot be detected by standard clinical examination.[47]

Typical findings on physical examination include asterixis and fetor hepaticus. The serum ammonia level (arterial or free venous) is commonly elevated. Electroencephalography usually shows high-amplitude low-frequency waves and triphasic waves. CT scan and MRI studies of the brain may be important in ruling out neurologic diseases.

Common precipitating factors of hepatic encephalopathy include diuretic therapy, renal failure, GI bleeding, bacterial infections, and constipation. Dietary protein overload is an infrequent cause of worsening encephalopathy. Medications—notably opiates, benzodiazepines, antidepressants, and antipsychotic agents—also may worsen encephalopathy symptoms. Surgical portosystemic shunts and TIPS favor the development of encephalopathy. The differential diagnosis for hepatic encephalopathy includes intracranial lesions, central nervous system infections, metabolic encephalopathy, toxic encephalopathy from alcohol or drugs, organic brain syndrome, and postseizure encephalopathy. In the initial management of hepatic encephalopathy, precipitants should be identified and corrected.[48] Lactulose, lactitol (not available in the United States), or both are helpful in patients with the acute onset of severe encephalopathy symptoms and in patients with milder, chronic symptoms.[49] Lactulose stimulates the passage of ammonia from tissues into the gut lumen and inhibits intestinal ammonia production. The initial lactulose dosage is 30 ml orally once or twice a day. The dose is increased until the patient has two to four loose stools a day. The dose should be reduced if the patient complains of diarrhea, abdominal cramping, or bloating. In hospitalized patients with severe encephalopathy, higher doses of lactulose may be administered via either a nasogastric tube or a rectal tube.[50] Neomycin (2 to 6 g/day), metronidazole (250 mg/day), rifaximin (1,200 mg/day), and other antibiotics (e.g., oral vancomycin, paromomycin, and oral quinolones) serve as second-line agents.[51] Antibiotics work by decreasing the colonic concentration of ammoniagenic bacteria. Other chemicals capable of decreasing blood ammonia levels are L-ornithine–L-aspartate (available in Europe) and sodium benzoate. Low-protein diets are not recommended, because they worsen the catabolic status of these patients and may cause malnutrition. In patients with portosystemic shunts, including TIPS, shunt-diameter reduction can be considered when hepatic encephalopathy is severe and does not respond to medical therapy. Because hepatic encephalopathy carries a poor prognosis, patients with episodic or permanent encephalopathy should be evaluated for OLT. The specific prognosis of patients with minimal encephalopathy is still unknown.

Table 4 Grading of Hepatic Encephalopathy

Grade	*Clinical Manifestations*
0 (subclinical)	Normal mental status, but minimal changes in memory, concentration, intellectual function, and coordination
1	Mild confusion, euphoria or depression, decreased attention, slowing of ability to perform mental tasks, irritability, disorder of sleep pattern (i.e., inverted sleep cycle)
2	Drowsiness, lethargy, gross deficits in ability to perform mental tasks, obvious personality changes, inappropriate behavior, intermittent disorientation (usually for time)
3	Somnolent but arousable, unable to perform mental tasks, disorientation to time and place, marked confusion, amnesia, occasional fits of rage, speech is present but incomprehensible
4	Coma, with or without response to painful stimuli

Hepatocellular Carcinoma

HCC is currently the main cause of mortality in cirrhotic patients.[52] The annual incidence of HCC in cirrhosis from HCV is 3% to 5%. Surveillance to detect early HCC involves the use of ultrasound examination and serum α-fetoprotein measurement every 6 months. In patients with nodules smaller than 1 cm, which are malignant in less than 50% of cases, close follow-up is recommended. HCC diagnosis is based on elevated serum α-fetoprotein levels, ultrasonography, helical CT and MRI findings, and positive cytohistology. The prognosis in patients with early-stage HCC depends on tumor status, liver function, and the treatment applied. Different staging systems (e.g., Barcelona Clinic Liver Cancer [BCLC] or Okuda) use tumor characteristics and liver function to classify patients with HCC.[53] Unfortunately, many HCC patients are diagnosed at advanced stages of disease that preclude the use of curative treatments. The 3-year survival rates of patients at intermediate and advanced stages of HCC are 65% and 16%, respectively. Curative treatments for HCC include surgical resection, OLT, and percutaneous ablation. In well-se-

Table 5 Indications for Liver Transplantation

Disease	*Criteria*
Hepatocellular liver disease	Serum bilirubin > 3 mg/dl Serum albumin < 2.5 g/dl Prothrombin time > 5 sec above control
Cholestatic liver disease	Serum bilirubin > 5 mg/dl Intractable pruritus Progressive bone disease Recurrent bacterial cholangitis
Both hepatocellular and cholestatic liver disease	Recurrent or severe hepatic encephalopathy Refractory ascites Spontaneous bacterial peritonitis Recurrent portal hypertensive bleeding Progressive malnutrition
Hepatocellular carcinoma	< 3 nodules No nodule > 5 cm No portal invasion

lected patients, resection and OLT achieve the best outcomes, with 5-year survival rates of 60% to 70%, whereas 5-year survival rates with percutaneous treatments are only 40% to 50%. Transplantation is the ideal treatment for patients with one tumor and decompensated cirrhosis or multicentric small tumors.[54] Arterial embolization may improve quality of life and, in some cases, even increase survival. Tamoxifen does not seem to have a significant beneficial effect.

Indications for Liver Transplantation

OLT is a central tool for the management of advanced cirrhosis.[55] In the United States, more than 3,000 liver transplants are performed each year. However, because there are many more candidates for transplantation than there are available donor livers, the selection and timing of patient referral are critical. The general indications for OLT are broadly categorized as clinical and biochemical [*see Table 5*]. Biochemical indexes vary, depending on whether liver disease is caused by hepatocellular conditions or chronic cholestatic disorders. Patients should be referred for transplant workup if the serum bilirubin level is greater than 3 mg/dl in noncholestatic disease or greater than 5 mg/dl in cholestatic disorders; if the prothrombin time is prolonged by more than 5 seconds; or if the serum albumin level is below 2.5 g/dl. Clinical criteria include HCC, hepatic encephalopathy, refractory ascites, recurrent variceal bleeding, SBP, and intractable pruritus. The clinical complications of cholestatic liver disease, such as intractable pruritus, recurrent bacterial cholangitis, and progressive bone disease, often warrant liver transplantation before hepatic encephalopathy or variceal hemorrhage develops. HCV-infected patients with decompensated cirrhosis awaiting OLT can be treated with pegylated interferon plus ribavirin; in these patients, treatment can be initiated several months before OLT to prevent graft reinfection.[56]

Contraindications for OLT include severe cardiovascular or pulmonary disease, active drug or alcohol abuse, malignancy outside the liver, sepsis, or psychosocial problems that may jeopardize a patient's ability to follow medical regimens after transplantation. The presence of HIV infection was considered a contraindication to transplantation, but successful liver transplantations are now being performed in patients in whom antiretroviral therapy has eliminated any detectable HIV viral load. Additional clinical study is required before OLT can be offered routinely to such patients.

In the United States, the Model for End-Stage Liver Disease (MELD) is the scoring system used by most liver transplant centers for determining priority for OLT.[57] MELD relies primarily on the bilirubin level, INR, and creatinine level to determine a patient's risk of dying within 3 months if OLT is not performed.[58] Patients' scores are calculated continuously while they are on the waiting list for OLT. Scores typically range from 6 (less ill) to 40 (gravely ill). A MELD calculator is available on the Internet (http://www.unos.org/resources/MeldPeldCalculator.asp?index=98).

Advances in surgical technique, organ preservation, and immunosuppression have resulted in dramatic improvements in postoperative survival over the past 2 decades.[59] In the early 1980s, 1-year and 5-year survival after liver transplantation were only 70% and 15%, respectively; the current rates are 85% and more than 70%. In most cases, patients can anticipate a good quality of life after liver transplantation.

Approximately 15% of patients listed as candidates for liver transplantation die before a donor organ becomes available. Strategies to improve the current donor organ shortage include programs to increase public awareness of the importance of organ donation, increased utilization of living-donor liver transplantation for pediatric recipients, and exploration of the efficacy and safety of living-donor liver transplantation in adults.[60]

Prognosis

The prognosis of patients with cirrhosis depends on the underlying disease, the occurrence of major complications (i.e., ascites, GI bleeding, encephalopathy, HRS, or bacterial infections), the degree of liver insufficiency, and the existence of HCC. In patients with compensated cirrhosis, the 10-year probability of major clinical complications is 58% and that of survival is 47%. For patients with decompensated cirrhosis, prognosis can be estimated by the older Child-Pugh classification and by the MELD score.[61] The variables included in the Child-Pugh score reflect the synthetic (albumin and prothrombin time) and elimination (bilirubin) functions of the liver, as well as major complications (ascites and encephalopathy). In contrast, the MELD score includes only numeric variables that reflect liver function (INR and bilirubin level) and renal function (creatinine level). The principal advantages of the MELD score are that it is based on objective variables selected for their influence on prognosis and that continuous recalculation helps in scoring individuals more precisely among large populations.[58] However, the MELD score has not been validated in some clinical situations. For example, in patients with type 1 HRS, the MELD score may underestimate survival.[62]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Desmet VJ, Roskams T: Cirrhosis reversal: a duel between dogma and myth. J Hepatol 40:860, 2004

2. Afdhal NH: The natural history of hepatitis C. Semin Liver Dis 24(suppl 2):3, 2004

3. Vong S, Bell BP: Chronic liver disease mortality in the United States, 1990–1998. Hepatology 39:476, 2004.

4. Sandler RS, Everhart JE, Donowitz M, et al: The burden of selected digestive diseases in the United States. Gastroenterology 122:1500, 2002

5. Kozak LJ, Owings MF, Hall MJ: National Hospital Discharge Survey: 2002 annual summary with detailed diagnosis and procedure data. National Center for Health Statistics. Vital Health Stat 13:1, 2005

6. Bataller R, North KE, Brenner DA: Genetic polymorphisms and the progression of liver fibrosis: a critical appraisal. Hepatology 37:493, 2003

7. Bataller R, Brenner DA: Liver fibrosis. J Clin Invest 115:209, 2005

8. Roskams T, Baptista A, Bianchi L, et al: Histopathology of portal hypertension: a practical guideline. Histopathology 42:2, 2003

9. Groszmann RJ, Abraldes JG. Portal hypertension: from bedside to bench. J Clin Gastroenterol 39:S125, 2005

10. Murakami T, Mochizuki K, Nakamura H: Imaging evaluation of the cirrhotic liver. Semin Liver Dis 21:213, 2001

11. Dienstag JL: The role of liver biopsy in chronic hepatitis C. Hepatology 36:S152, 2002

12. McCormack G, Nolan N, McCormick PA: Transjugular liver biopsy: a review. Ir Med J 94:11, 2001

13. Cardenas A, Gines P: Management of complications of cirrhosis in patients awaiting liver transplantation. J Hepatol 42(suppl):S124, 2005

14. Wright TL: Treatment of patients with hepatitis C and cirrhosis. Hepatology 36:S185, 2002

15. Lai CJ, Terrault NA: Antiviral therapy in patients with chronic hepatitis B and cirrhosis. Gastroenterol Clin North Am 33:629, 2004

16. Levitsky J, Mailliard ME: Diagnosis and therapy of alcoholic liver disease. Semin Liver Dis 24:233, 2004

17. Morgan TR, Weiss DG, Nemchausky B, et al: Colchicine treatment of alcoholic cirrhosis: a randomized, placebo-controlled clinical trial of patient survival. Gastroenterology 128:882, 2005

18. O'Shea RS, McCullough AJ: Treatment of alcoholic hepatitis. Clin Liver Dis 9:103, 2005

19. Oo YH, Neuberger J: Options for treatment of primary biliary cirrhosis. Drugs 64:2261, 2004

20. Gines P, Cardenas A, Arroyo V, et al: Management of cirrhosis and ascites. N Engl J Med 350:1646, 2004

21. Gines P, Tito L, Arroyo V, et al: Randomized comparative study of therapeutic paracentesis with and without intravenous albumin in cirrhosis. Gastroenterology 94:1493, 1988

22. Gines P, Uriz J, Calahorra B, et al: Transjugular intrahepatic portosystemic shunting versus paracentesis plus albumin for refractory ascites in cirrhosis. Gastroenterology 123:1839, 2004

23. Bureau C, Garcia-Pagan JC, Otal P, et al: Improved clinical outcome using polytetrafluoroethylene-coated stents for TIPS: results of a randomized study. Gastroenterology 126:469, 2004

24. Cardenas A, Gines P: Pathogenesis and treatment of fluid and electrolyte imbalance in cirrhosis. Semin Nephrol 21:308, 2001

25. Gines P, Guevara M, Arroyo V, et al: Hepatorenal syndrome. Lancet 362:1819, 2003

26. Arroyo V, Gines P, Gerbes AL, et al: Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. International Ascites Club. Hepatology 23:164, 1996

27. Ortega R, Gines P, Uriz J, et al: Terlipressin therapy with and without albumin for patients with hepatorenal syndrome: results of a prospective, nonrandomized study. Hepatology 36:941, 2002

28. Garcia-Tsao G: Bacterial infections in cirrhosis. Can J Gastroenterol 18:405, 2004

29. Fernandez J, Bauer TM, Navasa M, et al: Diagnosis, treatment and prevention of spontaneous bacterial peritonitis. Baillieres Best Pract Res Clin Gastroenterol 14:975, 2000

30. Sort P, Navasa M, Arroyo V, et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 341:403, 1999

31. Gines P, Rimola A, Planas R, et al: Norfloxacin prevents spontaneous bacterial peritonitis recurrence in cirrhosis: results of a double-blind, placebo-controlled trial. Hepatology 12:716, 1990

32. Bosch J, Abraldes JG: Management of gastrointestinal bleeding in patients with cirrhosis of the liver. Semin Hematol 41:8, 2004

33. Bernard B, Lebrec D, Mathurin P, et al: Propranolol and sclerotherapy in the prevention of gastrointestinal rebleeding in patients with cirrhosis: a meta-analysis. J Hepatol 26:312, 1997

34. Jutabha R, Jensen DM, Martin P, et al: Randomized study comparing banding and propranolol to prevent initial variceal hemorrhage in cirrhotics with high-risk esophageal varices. Gastroenterology 128:870, 2005

35. Chalasani N, Boyer TD: Primary prophylaxis against variceal bleeding: beta-blockers, endoscopic ligation, or both? Am J Gastroenterol 100:805, 2005

36. de Franchis R: Somatostatin, somatostatin analogues and other vasoactive drugs in the treatment of bleeding oesophageal varices. Dig Liver Dis 36(suppl 1):S93, 2004

37. Vangeli M, Patch D, Burroughs AK: Salvage tips for uncontrolled variceal bleeding. J Hepatol 37:703, 2002

38. Neuhaus P, Blumhardt G: Surgery for portal hypertension. Hepatogastroenterology 38:355, 1991

39. Patch D, Sabin CA, Goulis J, et al: A randomized, controlled trial of medical therapy versus endoscopic ligation for the prevention of variceal rebleeding in patients with cirrhosis. Gastroenterology 123:1013, 2002

40. Merkel C, Bolognesi M, Sacerdoti D, et al: The hemodynamic response to medical treatment of portal hypertension as a predictor of clinical effectiveness in the primary prophylaxis of variceal bleeding in cirrhosis. Hepatology 32:930, 2000

41. Hoeper MM, Krowka MJ, Strassburg CP: Portopulmonary hypertension and hepatopulmonary syndrome. Lancet 363:1461, 2004

42. Lima BL, Franca AV, Pazin-Filho A, et al: Frequency, clinical characteristics, and respiratory parameters of hepatopulmonary syndrome. Mayo Clin Proc 79:42, 2004

43. Swanson KL, Wiesner RH, Krowka MJ: Natural history of hepatopulmonary syndrome: impact of liver transplantation. Hepatology 41:1122, 2005

44. Ferenci P, Lockwood A, Mullen K, et al: Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology 35:716, 2002

45. Butterworth RF: Pathogenesis of hepatic encephalopathy: new insights from neuroimaging and molecular studies. J Hepatol 39:278, 2003

46. Jones EA: Ammonia, the GABA neurotransmitter system, and hepatic encephalopathy. Metab Brain Dis 17:275, 2002

47. Ortiz M, Jacas C, Cordoba J: Minimal hepatic encephalopathy: diagnosis, clinical significance and recommendations. J Hepatol 42(suppl):S45, 2005

48. Blei AT: Treatment of hepatic encephalopathy. Lancet 365:1383, 2005

49. Morgan MY, Hawley KE: Lactitol vs. lactulose in the treatment of acute hepatic encephalopathy in cirrhotic patients: a double-blind, randomized trial. Hepatology 7:1278, 1987

50. Bongaerts G, Severijnen R, Timmerman H: Effect of antibiotics, prebiotics and probiotics in treatment for hepatic encephalopathy. Med Hypotheses 64:64, 2005

51. Mas A, Rodes J, Sunyer L, et al: Comparison of rifaximin and lactitol in the treatment of acute hepatic encephalopathy: results of a randomized, double-blind, double-dummy, controlled clinical trial. J Hepatol 38:51, 2003

52. Llovet JM, Burroughs A, Bruix J: Hepatocellular carcinoma. Lancet 362:1907, 2003

53. Llovet JM, Fuster J, Bruix J: The Barcelona approach: diagnosis, staging, and treatment of hepatocellular carcinoma. Liver Transpl 10:S115, 2004

54. Schwartz M: Liver transplantation for hepatocellular carcinoma. Gastroenterology 127:S268, 2004

55. Brown KA: Liver transplantation. Curr Opin Gastroenterol 21:331, 2005

56. Forns X, Garcia-Retortillo M, Serrano T, et al: Antiviral therapy of patients with decompensated cirrhosis to prevent recurrence of hepatitis C after liver transplantation. J Hepatol 39:389, 2003

57. Freeman RB Jr: MELD and liver allocation: continuous quality improvement. Hepatology 40:787, 2004

58. Kamath PS, Wiesner RH, Malinchoc M, et al: A model to predict survival in patients with end-stage liver disease. Hepatology 33:464, 2001

59. Tra TT, Nissen N, Poordad FF, et al: Advances in liver transplantation: new strategies and current care expand access, enhance survival. Postgrad Med 115:73, 2004

60. White SA, Al-Mukhtar A, Lodge JP, et al: Progress in living donor liver transplantation. Transplant Proc 36:2720, 2004

61. Durand F, Valla D: Assessment of the prognosis of cirrhosis: Child-Pugh versus MELD. J Hepatol 42(suppl):S100, 2005

62. Alessandria C, Ozdogan O, Guevara M, et al: MELD score and clinical type predict prognosis in hepatorenal syndrome: relevance to liver transplantation. Hepatology 41:1282, 2005

X GASTROINTESTINAL BLEEDING

Elizabeth Rajan, m.d.
David A. Ahlquist, m.d.

Gastrointestinal (GI) bleeding occurs commonly, has many causes, and ranges from trivial to torrential and life-threatening in severity. Practical classification of GI bleeding—based on the presentation, site, and mechanism of the bleed—aids the clinician in selecting an appropriate management algorithm.

GI bleeding is defined as overt when visible red or altered blood is noted in emesis or feces. Overt bleeding is considered major when accompanied by hemodynamic instability and considered minor when not. Occult bleeding is visibly inapparent but is detected directly by stool testing or suggested indirectly by iron deficiency anemia.

GI bleeding occurs when a pathologic process such as ulceration, inflammation, or neoplasia causes erosion of a blood vessel. The size of the eroded artery is an important determinant of the rate of bleeding, the risk of rebleeding, and the clinical outcome. Blood flow and, thus, rate of blood loss vary directly with the diameter of the vessel; small changes in vessel diameter have dramatic effects on bleeding rates. Most GI bleeds result from erosion into small vessels and are trivial and self-limited. Erosion of larger vessels can produce lesions that exceed the capacity of normal hemostasis and result in overt major bleeding. A study of the external diameter of arteries in gastric ulcers that bled recurrently showed a range of 0.1 to 1.8 mm, with a mean of 0.7 mm.[1] Deep, large ulcers are more likely to erode into large blood vessels. Recurrent or persistent bleeding may result from inadequate vasoconstriction because of large vessel size or inflammatory necrosis of the vessel wall, from pseudoaneurysm formation at the bleeding site, or from systemic coagulopathies.

Overt Bleeding

EPIDEMIOLOGY

The reported incidence of GI bleeding varies widely, in part because of varying definitions. Overt minor bleeding, such as from anorectal hemorrhoids, is exceedingly common. Most major bleeding arises from upper GI lesions, and the estimated annual incidence ranges from 40 to 150 episodes per 100,000 population.[2,3] Mortality from upper GI bleeding has remained at 8% to 10% over the past 50 years.[4,5] The fact that, over this period, mortality has failed to decrease substantially despite advances in patient care and technology may reflect the increasing number of elderly patients with complicated comorbidities. Cases in individuals older than 60 years account for 35% to 45% of all cases of acute upper GI bleeding but nearly all of the associated mortality.[6] Lower GI sources account for an estimated 15% to 20% of all major GI bleeds. The incidence of lower GI bleeds increases with age.[7,8]

ETIOLOGY

The causes of GI bleeding are protean [*see Table 1*]. The most common etiologies are briefly elaborated in this subsection.

Upper GI Bleeding

Upper GI bleeding is arbitrarily defined as hemorrhage from a source proximal to the ligament of Treitz (i.e., the esophagus, stomach, or duodenum). Hematemesis essentially always reflects upper GI bleeding, and stools may range from black (melena) to bright red (hematochezia), depending on rates of bleeding and intestinal transit.

Peptic ulcer disease (PUD) The most common cause of upper GI bleeding is PUD, accounting for 60% of cases found at emergency endoscopy.[9] About 50% of cases have a clean-based ulcer with a low probability of rebleeding, so that only pharmacologic intervention is required.[10] Adherent clots, visible vessels, or active bleeding [*see Figure 1*] portend progressively less favorable outcomes unless endoscopic or surgical treatment is applied. Although nonsteroidal anti-inflammatory drug (NSAID) use and *Helicobacter pylori* infection are the two most important risk factors for bleeding in PUD, heavy alcohol ingestion and smoking are also associated with increased risk.[11-13]

Drugs Aspirin and other NSAIDs are responsible for most drug-induced GI bleeding. In the United States, more than 30 billion NSAID tablets are consumed annually. Except for sodium salicylate, all NSAIDs can cause bleeding. Acetaminophen is not associated with GI bleeding.

The elderly are especially susceptible to NSAID-induced GI bleeding.[14] NSAIDs may cause bleeding at any level of the GI tract, but they most commonly do so in the stomach or duodenum. Although the bleeding risk increases in proportion to NSAID dose, any amount (including low-dose aspirin taken for

Table 1 Major Causes of Gastrointestinal Bleeding

Inflammatory
Peptic ulcer disease
Esophagitis or esophageal ulceration
Diaphragmatic hernia (Cameron erosions)
Diverticular disease
Inflammatory bowel disease
Meckel diverticulum
Cancers and neoplasms
Primary lesion at any site
Metastatic deposits at any site
Large polyps
Gastrointestinal stromal tumors
Vascular anomalies
Gastroesophageal varices
Angiodysplasia
Dieulafoy lesion
Watermelon stomach
Radiation proctopathy
Drugs
Aspirin
Nonsteroidal anti-inflammatory drugs
Miscellaneous
Postpolypectomy
Mallory-Weiss tear

cardiovascular prophylaxis) may cause bleeding. Use of selective serotonin reuptake inhibitors (SSRIs) has recently been found to be associated with a higher risk of upper GI bleeding, especially in patients who are also taking NSAIDs or low-dose aspirin.[15] Anticoagulants do not cause GI bleeding per se, but they can unmask or aggravate hemorrhage from preexisting lesions.

Variceal bleeding Gastroesophageal variceal bleeding accounts for 10% to 30% of all upper GI hemorrhage. Patients present with overt major bleeding that is sudden in onset. Variceal bleeding is distinctive, with large-volume hematemesis of bright-red blood or clots and associated severe hemodynamic instability [*see Figure 2*]. Because of the cathartic nature of blood, patients may also present with hematochezia. A prospective review found that the distribution of bleeding varices was 75% esophageal and 25% gastric.[16] The most common site of bleeding is the distal 5 cm of the esophagus, because of relatively greater variceal distention and thinner supporting tissue surrounding the veins in this region, compared with the upper and the middle esophagus. Varices are present in 40% to 60% of patients with cirrhosis, and hemorrhage occurs in 25% to 35% of them.[17-19] Approximately one third of first variceal bleeds are fatal.[19] Physicians should bear in mind that up to half of patients with portal hypertension bleed from a nonvariceal cause.[16]

Mallory-Weiss tear Mallory-Weiss tear is a longitudinal mucosal laceration at the gastroesophageal junction or gastric cardia caused by forceful retching or vomiting. Most tears occur within 2 cm of the gastroesophageal junction on the lesser curvature aspect of the cardia. Mallory-Weiss tears account for 5% to 11% of all major upper GI hemorrhages.[20] Most patients present with hematemesis, often associated with alcohol use. Typically, overt bleeding is minor and bleeding ceases spontaneously. Mallory-Weiss tears can also occur with upper GI endoscopy when a patient struggles or retches during the procedure.

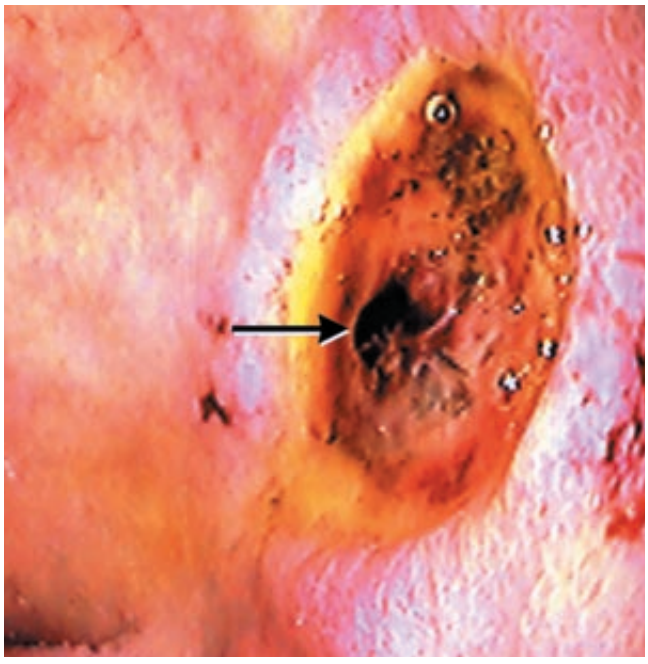


Figure 1 **High-risk posterior duodenal bulb ulcer with nonbleeding visible vessel (arrow).**

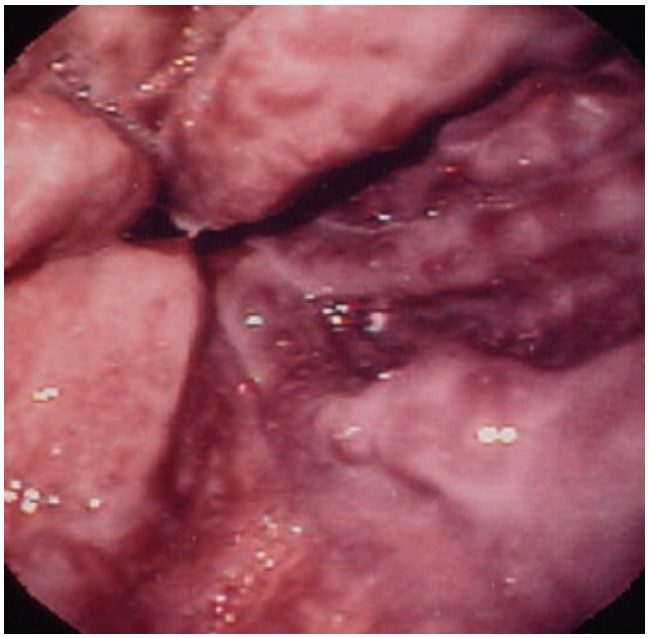


Figure 2 **High-risk esophageal varices with red wale marking.**

Dieulafoy lesions Dieulafoy lesions account for approximately 5% of cases of major upper GI bleeding.[21] Their characteristic feature is the presence of a large-caliber, tortuous artery in the submucosa close to the mucosal surface, which bleeds upon erosion of the overlying mucosa and artery wall [*see Figure 3*]. They can be extremely difficult to detect endoscopically unless actively bleeding. Dieulafoy lesions are usually single lesions located in the proximal stomach. However, these lesions can occur anywhere throughout the GI tract. A review of 90 Dieulafoy lesions identified 34% of lesions as extragastric.[22]

Lower GI Bleeding

Diverticulosis Diverticular disease is one of the most common causes of lower GI bleeding, particularly in the elderly. Diverticulosis is uncommon in persons younger than 40 years, but it affects roughly two thirds of persons older than 80 years.[23,24] The mean age period for diverticular hemorrhage is the sixth decade of life. The true incidence of diverticular bleeding is difficult to ascertain, given the different definitions and evaluations used in various studies. Bleeding occurs from an arteriole at either the dome or neck of a diverticulum. Typically, there is no associated diverticulitis. Diverticula are most commonly found in the left colon, but many bleeds arise from diverticula in the right colon. Patients typically present with painless, large-volume hematochezia. Because diverticular bleeding tends to stop spontaneously, the diagnosis is often presumptive and based on exclusion of other sources of bleeding in a patient with diverticulosis.[25]

Angiodysplasia Angiodysplasia is an acquired vascular ectasia that is considered to be degenerative in origin, given its propensity to occur in the elderly. Typically, patients present between 60 and 80 years of age. The pathogenesis of angiodysplasia remains unclear, but a proposed cause is chronic, intermittent, low-grade obstruction of submucosal veins, leading to dilatation of mucosal capillaries [*see Figure 4*]. The lesions of angiodysplasia are usually small (2 to 5 mm in diameter) and can be single or multiple. These lesions can occur anywhere

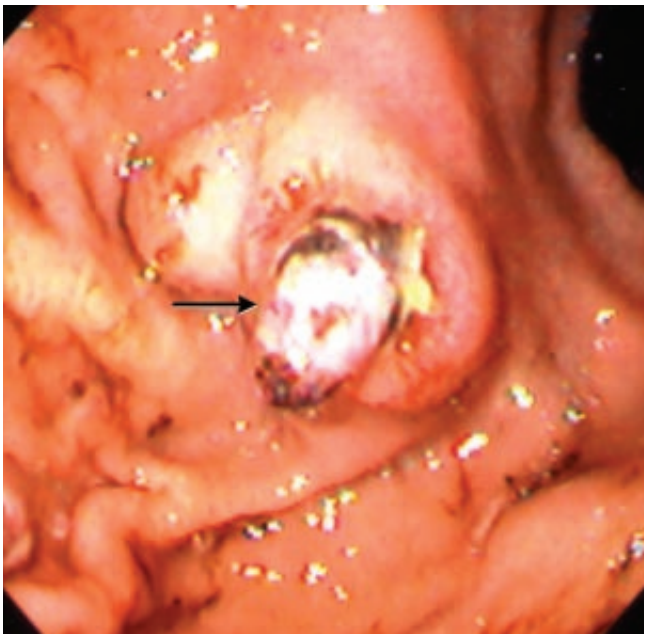


Figure 3 **Dieulafoy lesion in gastric fundus (arrow).**

along the GI tract but are most commonly found in the proximal colon (approximately 80%), particularly the cecum.[26] Angiodysplasia is an incidental finding at colonoscopy in 2% of nonbleeding patients older than 65 years.[27,28] Fewer than 10% of patients with angiodysplasia will bleed.[27] Bleeding stops spontaneously in the majority of patients, but rebleeding is common.

Polypectomy Colonoscopic polypectomy is generally considered a safe procedure, but hemorrhage is reported to occur in 0.3% to 6.0% of cases.[29] A retrospective study of 83 patients and 274 polypectomy sites found that bleeding occurred at a median of 5 days (range, 0 to 17 days) after the procedure.[30] Bleeding was associated with advanced age, polyps greater than 1 cm in diameter, sessile polyps, and polyps in the cecum. The prognosis for these patients is favorable. Most cases are managed with observation or endoscopic hemostasis.

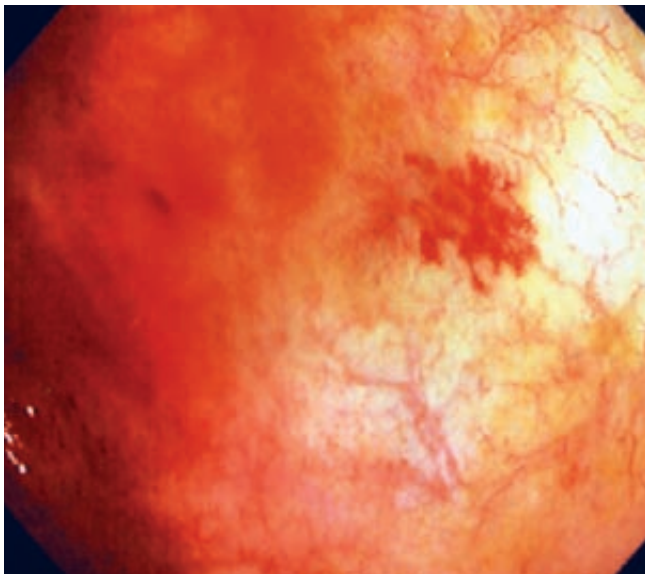


Figure 4 **The lesions of angiodysplasia are usually small (2 to 5 mm in diameter) and can be single or multiple. These lesions can occur anywhere along the GI tract but are most commonly found in the proximal colon, particularly the cecum.**

Diagnosis

EMERGENT EVALUATION AND MANAGEMENT

Management of GI bleeding is determined by the severity of the bleed; algorithms differ with major bleeding [*see Figure 5*] and minor bleeding [*see Figure 6*]. Patients with overt major bleeding require immediate hospitalization with intensive monitoring. Patients are initially stabilized with fluid and blood component replacement and with correction of any coagulopathy or electrolyte imbalances. Endotracheal intubation may be necessary. Stabilization is followed by immediate endoscopic evaluation and therapy as indicated. If hemorrhage control is ineffective and the patient continues to be hemodynamically unstable, radiologic or surgical interventions are considered.

CLINICAL AND LABORATORY ASSESSMENT

The history and physical examination provide vital information on the location, severity, and duration of bleeding and can help identify patients at increased risk of exsanguination and rebleeding [*see Table 2*]. It is important to remember that patients with overt major bleeding from an upper GI source can present

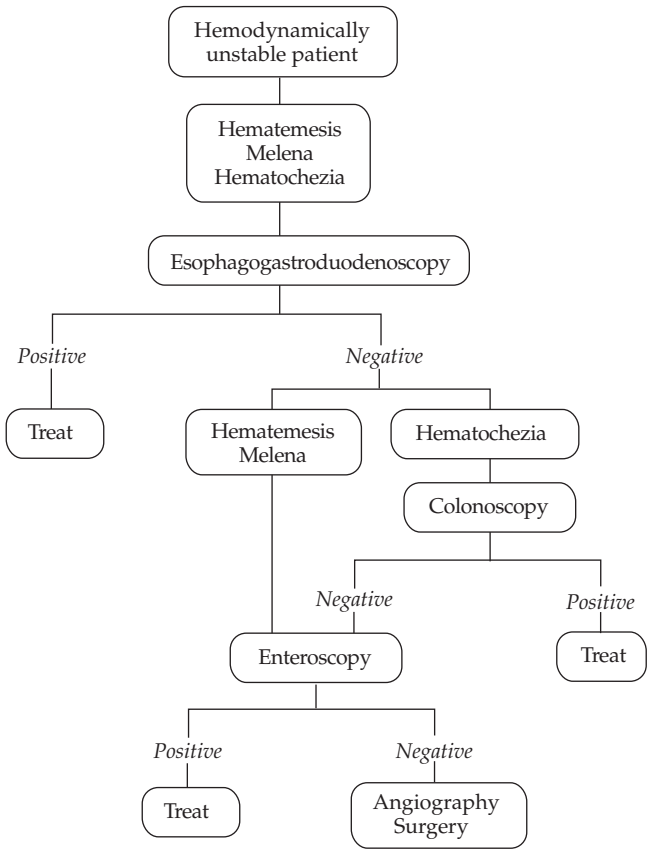


Figure 5 **Evaluation and management of overt major gastrointestinal bleeding.**

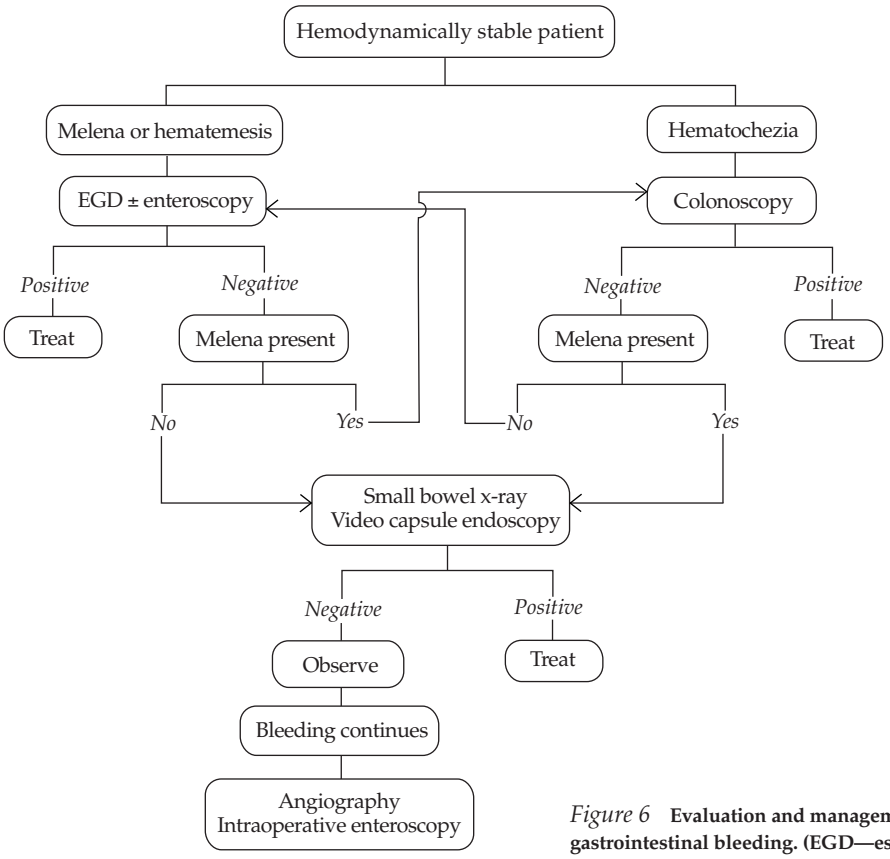


Figure 6 **Evaluation and management of overt minor gastrointestinal bleeding. (EGD—esophagogastroduodenoscopy)**

with hematochezia. These patients can experience visceral discomfort and orthostatic symptoms shortly after the onset of bleeding. Abdominal pain—especially periumbilical cramping and gaseous distention—usually indicates rapid intestinal transit of blood and suggests a major bleed.

The physician should look for evidence of liver disease, PUD, coagulopathy, previous abdominal aortic aneurysm repair, and significant comorbidities such as heart disease and diabetes mellitus. A history of drug or alcohol ingestion may suggest a diagnosis.

After cessation of active upper GI bleeding, patients may experience melena for 2 to 3 days. This does not indicate rebleeding, especially if the hemoglobin level does not decrease.

Serial recording of vital signs is crucial in determining whether an overt major bleed has occurred. Significant volume loss is indicated by hypotension (systolic blood pressure less than 100 mm Hg), orthostasis (a decrease in systolic pressure of more than 20 mm Hg or an increase in heart rate of more than 20 beats/min), tachycardia (heart rate greater than 100 beats/min), or a drop in hemoglobin of more than 2 g/dl. Further assessment of skin pallor, features of liver disease or portal hypertension, and stool color from rectal examination can also help with diagnosis or management.

A nasogastric tube can be placed if there is uncertainty about the location of the bleed in a patient with hematochezia or if bleeding persists in a patient with hematemesis. Aspiration of blood indicates a recent upper GI bleed, but absence of blood in the aspirate does not exclude a recent bleed.

The most important laboratory measurement to assess severity of the initial bleed and to monitor rebleeding is the hemoglobin level. An abrupt drop of more than 2 g/dl indicates a significant bleeding episode. An increase in the ratio of blood urea nitrogen to creatinine to more than 25:1 strongly suggests an upper GI source. Measurement of serum electrolyte concentrations, coagulation indices, platelet count, and liver enzyme levels aids in the diagnosis and guides management.

ENDOSCOPY

In most patients, the location of the bleed is identified by upper GI endoscopy or colonoscopy. Endoscopy also provides therapeutic options and essential information on the risk of rebleeding [*see Table 3*]. There are established visual criteria, based on stigmata of recent hemorrhage, that the endoscopist can use to identify patients at high or low risk for rebleeding.

During upper GI endoscopy, if massive active bleeding is encountered, it is prudent to discontinue the procedure and protect the airway by endotracheal intubation before proceeding. If vi-

Table 2 Clinical High-Risk Criteria for Rebleeding and Mortality

Advanced age (≥ 70 yr)
Major organ comorbidities
In-hospital bleed
Bright-red hematemesis in patient with liver cirrhosis
Hypotension (systolic blood pressure < 100 mm Hg)
Tachycardia (heart rate >100 bpm)
Orthostasis (BP drop > 20 mm Hg; HR rise > 20 bpm)
Hemoglobin < 10 g/dl or drop of ≥ 2 g/dl
≥ 4 units of blood transfused in 24 hr

Table 3 Endoscopic High-Risk Stigmata for Rebleeding and Indications for Endoscopic Therapy

Nonvariceal bleeding	Arterial spurting Oozing bleed Nonbleeding visible vessel Adherent clot
Variceal bleeding	Large varices (> 5 mm esophageal, > 1 cm gastric) Red wale marks (longitudinal dilated venules resembling whip marks) Cherry-red spots (< 2 mm diameter) Hematocystic spots (> 4 mm diameter)

sualization is impaired, use of large-bore orogastric lavage or a jumbo-channel (6 mm) therapeutic endoscope to evacuate blood and clots may be effective. Erythromycin lactobionate (125 mg intravenously) can also be used to promote quick intestinal transit of blood when active bleeding has stopped.

Before colonoscopy, whenever possible, patients should receive a rapid colonic lavage with 2 to 3 L of a nonabsorbable polyethylene glycol solution administered through a nasogastric tube over 2 hours to cleanse the colon and facilitate adequate visualization.

RADIOLOGY

Selective visceral angiography is considered when endoscopic therapy for an established lesion has failed and surgery is not an option or when the site of an active bleed remains obscure after endoscopy. An optimal examination with a high positive yield is best obtained when there is active bleeding at rates exceeding 0.5 to 1 ml/min. Significant complications—including contrast reaction, acute renal failure, and femoral artery thrombosis—have been reported in approximately 9% of cases.[31,32] The reported sensitivity of angiography varies from 22% to 87%. The specificity approaches 100%.[33]

Radionuclide Technetium Scan

A technetium-99m–labeled red cell scan should be considered when active bleeding is suspected but endoscopy has been negative. Nuclear scans can detect bleeding at rates that exceed 0.1 ml/min. On scans, however, pooled blood may sometimes be mistaken for active bleeding, which contributes to a reported false positive rate of about 22%.[34] Upper GI bleeding may be misdiagnosed as lower because of pooling in the distal ileum or right colon. A positive result is more reliable when the scan is done early rather than delayed (several hours later).

Other Measures

Endoscopic techniques that are currently available for examination of the small bowel include push enteroscopy, wireless capsule endoscopy, and intraoperative enteroscopy. Push enteroscopy typically reaches into the proximal jejunum only, whereas wireless capsule enteroscopy and intraoperative enteroscopy reach the entire small bowel.

Enteroscopy is currently performed using a pediatric colonoscope with or without an overtube. In one study, the diagnostic yield of enteroscopy in overt GI bleeding was 46%; the most common lesions seen were angiodysplasia and ulcers.[35]

Wireless capsule endoscopy represents a new technology involving an easily swallowed 11 × 30 mm capsule. No sedation is required. The capsule contains a color video chip, light source, and transmitter. The patient wears an antenna array on a belt. While transiting through the intestine by peristalsis, the capsule takes color photos and sends them to the antenna array. These images are then downloaded onto a computer after the examination. There is a total of 6 to 8 hours of recording time. This technique may be beneficial in patients with recurrent or occult GI bleeding of obscure origin, but it is not appropriate in unstable patients with major active bleeding.

Intraoperative enteroscopy, performed during exploratory laparotomy, through single or multiple enterotomy sites, is indicated for the occasional patient with active or recurrent major bleeding of obscure origin. Complications include mucosal laceration, intramural hematoma, mesenteric hemorrhage, and intestinal ischemia.[36]

Treatment

NONVARICEAL BLEEDING

Endoscopic Therapy

A variety of endoscopic modalities are currently available for the management of GI bleeding. These can be categorized into thermal, mechanical, and injection devices. Thermal devices are either contact (e.g., heater probe, multipolar electrocautery) or noncontact (e.g., argon plasma coagulator, laser). These devices generate sufficient heat to create a hemostatic bond through tissue desiccation. The heater probe consists of a Teflon-coated hollow aluminum cylinder with a heating coil. Only heat (no electrical current) is delivered to the tissue. Multipolar or bipolar cautery works by completion of an electrical circuit between two electrodes on the probe tip. The argon plasma coagulator utilizes high-frequency monopolar alternating current delivered to target tissue through ionized argon gas. The conduit of argon gas is called the argon plasma. Electrons flow through a channel of electrically activated, ionized argon gas from the probe electrode to the tissue, causing a thermal effect at the interface. In laser photocoagulation, which is less frequently used, the conversion of light to heat results in coagulation or vaporization of tissue. The neodymium:yttrium-aluminum-garnet (Nd:YAG) laser is the one most commonly used. Mechanical devices for hemostasis include metallic clips and rubber-band ligators. An injection solution that is generally used to achieve hemostasis is saline mixed with epinephrine at a 1:10,000 concentration.

These therapeutic modalities are used alone or in combination. A common practice is to start by injecting epinephrine and saline submucosally in the region of active bleeding so as to stop or slow hemorrhaging and therefore allow for adequate inspection. Thermal or mechanical modalities are then used to achieve definitive hemostasis. Prospective, controlled studies have confirmed the benefit of endoscopic intervention in achieving initial hemostasis and in prevention of rebleeding.[37] Combination therapy (i.e., injection plus thermal therapy) has been demonstrated to reduce rebleeding rates more successfully than single therapy.[38,39] Currently, combination therapy using injection followed by either a thermal or a mechanical intervention is the most effective approach. Rebleeding after endoscopic therapy occurs in approximately 20% of cases, typically within 48 to 72 hours after treatment. However, rebleeding can occur as late as 7 days after therapy.

Pharmacotherapy

Initial drug therapy for major nonvariceal upper GI bleeding is directed at gastric acid suppression. In a randomized, double-blind study of high-dose omeprazole versus placebo, rebleeding after endoscopic therapy occurred less frequently in the omeprazole group (7% versus 23%).[40] In general, proton pump inhibitors are administered in doses that reduce gastric acidity. Blood clot stability depends on intragastric pH, with optimum stability at a pH of 6 or higher.[41]

In patients with PUD, long-term acid suppression and eradication of *H. pylori* infection after endoscopic intervention promote ulcer healing, including ulceration at the treatment site, and reduce rebleeding substantially. GI bleeding from NSAIDs is best prevented by avoiding these drugs or by using a cyclooxygenase-2 (COX-2) inhibitor plus a proton pump inhibitor.

Radiologic Intervention

Selective arterial embolization and selective vasoconstriction with intra-arterial infusion of vasopressin are the methods currently available for the control of major nonvariceal GI bleeding. The proponents of embolization favor this form of therapy because it reduces the need for intensive care observation and it eliminates indwelling arterial lines, the risk of catheter dislodgement, and problematic systemic side effects of intravenous vasopressin [*see* Pharmacotherapy, *below*]. Advances in catheter design have allowed for superselective embolization of vasa recta; in experienced units, this modality is probably the treatment of choice. A study of superselective embolization in 48 patients with lower GI bleeding showed that embolization was the definitive treatment in 44% of patients, with a 27% technical failure rate.[42] The risks associated with embolization include misplacement of embolic material, inadvertent distal reflux of embolic agent, and excessive devascularization of an organ leading to ischemia and eventual luminal stenosis. Endoscopy can be helpful in determining ischemic injury if suspected. Microcoils (e.g., stainless steel, platinum), gelfoam pledgets, polyvinyl alcohol particles, and collagen suspensions have been used for embolization.

Intra-arterial vasopressin is the drug of choice for selective vasoconstrictive therapy and is generally infused for a minimum of 24 hours. It is associated with a 70% rate of bleeding control and an 18% rate of rebleeding.[43-45] Vasopressin may be ineffective when bleeding arises from large arteries that do not constrict in response to therapy. A study comparing embolization with vasopressin showed similar initial hemostasis rates but a higher rebleeding rate with vasopressin.[2] The use of intra-arterial provocative mesenteric angiography with heparin and tissue plasminogen activator (t-PA) to aid in diagnosis has been described but is still in the experimental stage.

Surgery

Despite the high overall success rate of endoscopic therapy in the treatment of major GI bleeding, surgery is still indicated when (1) initial hemostatic control cannot be achieved, (2) rebleeding occurs despite repeated endoscopic sessions, (3) a large (> 2 cm) penetrating ulcer is present, (4) a vessel larger than 2 mm in diameter is visible within the culprit lesion, (5) the ulcer is located in the posterior duodenal bulb (this location is associated with the large gastroduodenal artery), and (6) the patient requires substantial transfusion (i.e., four or more units of blood over 24 hours). The choice of surgery depends on the location of the bleed and the presence of comorbidities. Localization of the site of bleeding is critical for surgical planning.

VARICEAL BLEEDING

Endoscopic Therapy

With variceal bleeding, endoscopic treatment is used primarily for esophageal varices, and the techniques include sclerotherapy and band ligation. Sclerotherapy utilizes a variety of sclerosants to induce variceal thrombosis, with sodium tetradecyl sulfate and ethanolamine oleate used most frequently. Intravariceal injections are more effective than paraesophageal injections in controlling bleeding. Compared with a sham injection, sclerotherapy is significantly more likely to stop bleeding (91% versus 60%), reduce mortality during hospitalization (mortality, 25% versus 49%), reduce rebleeding rates (rebleeding, 20% versus 51%), and reduce transfusion need (four versus eight units).[46] Complications of sclerotherapy include retrosternal chest pain, fever, ulceration (usually deep ulcers that heal within 3 weeks), dysphagia, delayed perforation (1 to 4 weeks later), and stricture formation. Complication rates vary from 19% to 35%.[47-49] The popularity of sclerotherapy has diminished as a result of these complications.

Band ligation is now considered the first-line endoscopic therapy for esophageal varices. The band ligator is readily attached to the distal end of the endoscope, which is advanced to the varix; the endoscopist then suctions the varix into the ligator cap and deploys a rubber band around the varix. This results in the plication of the varices and surrounding submucosal tissue, with fibrosis and eventual obliteration of varices. Comparative studies report a better initial control of bleeding (control rates, 91% versus 77%) and lower rebleeding rates (rebleeding, 24% versus 47%) with band ligation than with sclerotherapy.[47] Complications of banding include retrosternal chest pain, dysphagia from compromise of the esophageal lumen, band ulceration (usually superficial ulcers that heal within 2 weeks), esophageal injury from the overtube, or esophageal perforation. Complication rates vary from 2% to 19%.[50,51]

If bleeding continues despite endoscopic therapy or if endoscopic therapy cannot be initiated, then a modified Sengstaken-Blakemore (Minnesota) tube should be inserted. However, this is only a temporary measure until more definitive treatment—endoscopic, radiologic, or surgical—can be undertaken.

Preventive measures may be indicated in patients with esophageal varices. Preventive measures are generally offered to patients who have a history of a bleed and to those who have large esophageal varices without a prior bleeding event. Currently, the accepted preventive measures for variceal bleeding include endoscopic band ligation, beta-blocker therapy, or a combination of both. Ligation is performed every 14 to 21 days until varices are completely eradicated, which typically requires three or four sessions.

Pharmacotherapy

In acute variceal bleeding, splanchnic blood flow and portal pressure can be reduced by intravenous infusion of vasoconstrictors such as vasopressin, terlipressin, somatostatin, and octreotide. Vasopressin is a potent vasoconstrictor that has a reported overall success rate of 50% but a high rebleeding rate when treatment is discontinued.[52] It has a short half-life and therefore is given as a continuous infusion. Vasopressin-induced hypertension and bradycardia have the potential to confound hemodynamic monitoring and may give false reassurance in the face of active bleeding. Because the systemic vasoconstrictive side effects associated with vasopressin may lead to myocardial

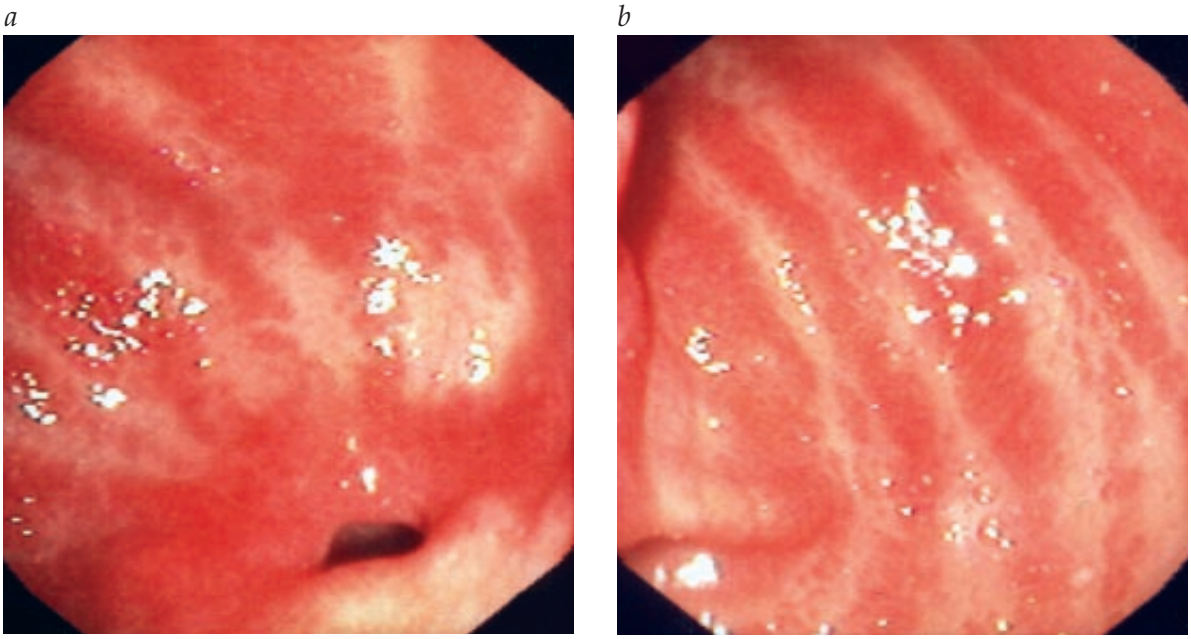


Figure 7 **Watermelon stomach with (*a*) typical spokes of vascular ectasia radiating from the pylorus into the antrum and (*b*) close-up view.**

or mesenteric ischemia, it is rarely used alone. To minimize these side effects, it is given in conjunction with nitroglycerin. The nitroglycerin can be administered as a continuous infusion, sublingually, or by transdermal patch. Terlipressin is a synthetic analogue of vasopressin that has fewer side effects and a longer half-life and is given in bolus injections; however, terlipressin has not yet been approved for use in the United States.

Somatostatin, a naturally occurring peptide, is reported to stop variceal bleeding in 80% of patients.[17,53] Side effects are few and include hyperglycemia and abdominal pain. Octreotide is a synthetic analogue of somatostatin that is preferred because of its longer half-life. The combination of pharmacologic treatment (e.g., octreotide for 5 days) and endoscopic therapy appear to offer better control of acute bleeding than either alone.

The role of beta blockers is primarily prophylactic. These agents are not used in the acute management of variceal bleeding. The use of isosorbide mononitrate with beta-blocker therapy does not offer a survival advantage and in fact reduces the tolerability of therapy.

Radiologic Intervention

The radiologic intervention available for variceal bleeding is transjugular intrahepatic portosystemic shunt (TIPS). The accepted indications for TIPS are bleeding or rebleeding that cannot be controlled by either pharmacologic or endoscopic therapy. TIPS is contraindicated in patients with severe hepatic failure, chronic heart failure, hepatic encephalopathy, bile duct obstruction, or cholangitis. TIPS is reported to control bleeding in at least 90% of patients, with rebleeding rates of 12% to 26% at 1 year and 16% to 44% at 2 years.[54,55] Patients require close surveillance for shunt dysfunction (evidenced by reduced flow by Doppler ultrasound or reappearance of varices) because primary shunt patency rates are poor (reported cumulative patency rates of 50% at 1 year and 21% at 3 years) but cumulative secondary shunt patency rates can be as satisfactory as 85% and 55% at 1 and 3 years, respectively.[55] TIPS should not be undertaken lightly, because the overall procedure-related mortality can be as high as 1% to 2%,[54] largely from intraperitoneal hemorrhage. Other complications include hepatic encephalopathy, portal vein thrombosis, renal failure, sepsis, and stent migration or stenosis.

Surgical Intervention

Surgical intervention is rarely used for variceal bleeding; it is considered when other measures have proved ineffective. Surgical treatments include portosystemic venous shunt operations and esophageal devascularization. A variety of surgical shunts are available. These are generally classified as total, partial, or selective, depending on the intended impact of portal flow diversion. The end-to-side portacaval shunt is a total shunt that diverts all portal blood flow into the inferior vena cava. The side-to-side portacaval shunt diverts only a part of the portal blood flow. Selective shunts decompress variceal flow while preserving portal blood flow. The distal splenorenal shunt is a selective shunt designed to prevent encephalopathy, which is often seen with total shunts. Surgical shunts are used for both esophageal and gastric varices. Encephalopathy, accelerated progression of liver failure, and perioperative morbidity can occur with surgical intervention. Esophageal devascularization may be an effective means of controlling acute variceal bleeding, but bleeding can recur as additional varices develop.

Occult Bleeding

The critical metabolic sequela of occult GI bleeding is iron deficiency.[56] Occult GI bleeding causes most cases of iron deficiency in adults, especially in men and postmenopausal women.

ETIOLOGY

Most of the many lesions that cause overt bleeding can also produce occult blood loss. However, variceal and diverticular hemorrhage invariably bleed overtly, whereas lesions such as watermelon stomach (gastric antral vascular ectasia) and diaphragmat-

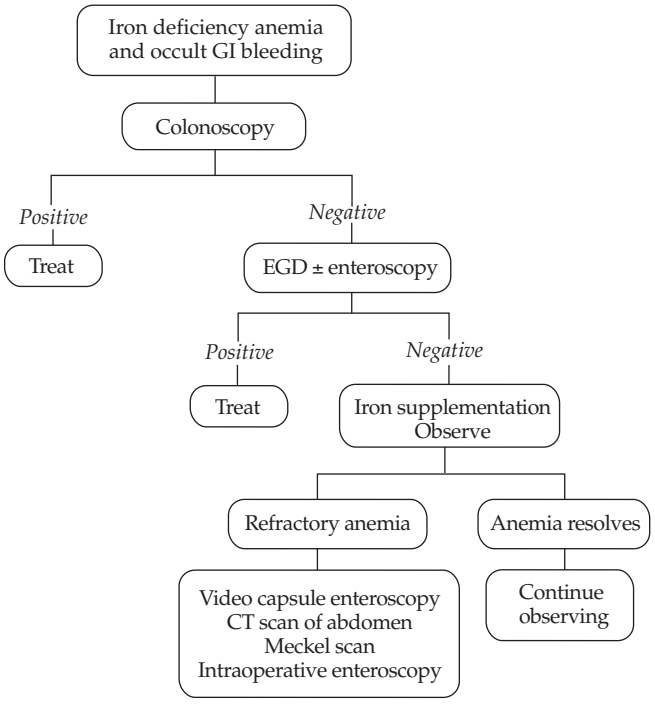


Figure 8 **Evaluation and management of occult gastrointestinal bleeding.**

ic hernia with Cameron erosions tend to bleed occultly. Occult GI bleeding in most patients is suspected only when manifested by fatigue, pallor, or the finding of iron deficiency.

Inflammation

In Western countries, erosive or ulcerative diseases of the esophagus, stomach, and duodenum are the most common GI lesions associated with occult bleeding and iron deficiency anemia. Most peptic disease is caused by either *H. pylori* infection or use of drugs such as aspirin or other NSAIDs. The association between large diaphragmatic hernias and iron deficiency anemia has long been known. A large diaphragmatic hernia is found in about 10% of iron-deficient patients.[57] Blood loss in these patients is generally caused by longitudinal mucosal erosions (Cameron erosions) located proximally in the hernia and believed to be secondary to repeated mechanical trauma from respiration.

Cancers and Neoplasms

In adults from Western countries, GI tumors are second only to PUD as a cause of occult bleeding leading to iron deficiency anemia.[58] Colorectal cancer is currently the most common source of occult bleeding from GI tract malignancies.

Vascular Causes

Vascular malformations are found in approximately 6% of adults with iron deficiency anemia.[59,60] This may be acquired or hereditary (hereditary hemorrhagic telangiectasia). An increasingly recognized and endoscopically treatable vascular lesion is watermelon stomach [*see Figure 7*], which typically presents as iron deficiency anemia in older women.

TREATMENT

When a patient is found to have iron deficiency and occult GI bleeding, it is critical to conduct a thorough GI investigation. Such an evaluation may disclose a health-threatening lesion, in which case specific therapy can be given to prevent associated morbidity and further iron loss. Only after a specific lesion has been treated or has been ruled out, is it appropriate to place patients on iron therapy and monitor them [*see Figure 8*].

Whatever the culpable lesion, treatment with iron supplementation is important to correct iron deficiency. With conditions such as Cameron erosions, iron supplementation is the mainstay of treatment. Most patients can be managed as outpatients. Oral iron therapy with ferrous sulfate is preferred because it is inexpensive, effective, and, in most cases, well tolerated [*see 5:II Red Blood Cell Function and Disorders of Iron Metabolism*]. A maximal adult dose of ferrous sulfate is 325 mg three times a day. Absorption is not appreciably increased with higher doses. Oral iron is as effective as parenteral iron in repleting iron stores, except in patients with a malabsorption syndrome, and is safer.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Swain CP, Storey DW, Bown SG, et al: Nature of the bleeding vessel in recurrently bleeding gastric ulcers. Gastroenterology 90:595, 1986

2. Peter DJ, Dougherty JM: Evaluation of the patient with gastrointestinal bleeding: an evidence based approach. Emerg Med Clin North Am 17:239, 1999

3. Vreeburg EM, Snel P, de Bruijne JW, et al: Acute upper gastrointestinal bleeding in the Amsterdam area: incidence, diagnosis, and clinical outcome. Am J Gastroenterol 92:236, 1997

4. Friedman LS, Martin P: The problem of gastrointestinal bleeding. Gastroenterol Clin North Am 22:717, 1993

5. Pianka JD, Affronti J: Management principles of gastrointestinal bleeding. Prim Care 28:557, 2001

6. Farrell JJ, Friedman LS: Gastrointestinal bleeding in the elderly. Gastroenterol Clin North Am 30:377, 2001

7. Gostout CJ, Wang KK, Ahlquist DA, et al: Acute gastrointestinal bleeding: experience of a specialized management team. J Clin Gastroenterol 14:260, 1992

8. Gostout CJ: Acute lower GI bleeding. Current Medicine: Clinical Practice of Gastroenterology. Brandt LJ, Daum F, Eds. Churchill Livingstone, Philadelphia, 1998, p 651

9. Hay JA, Lyubashevsky E, Elashoff J, et al: Upper gastrointestinal hemorrhage—clinical guideline determining the optimal hospital length of stay. Am J Med 100:313, 1996

10. Rockall TA, Logan RF, Devlin HB, et al: Selection of patients for early discharge or outpatient care after acute upper gastrointestinal haemorrhage. National Audit of Acute Upper Gastrointestinal Haemorrhage. Lancet 347:1138, 1996

11. Andersen IB, Jorgensen T, Bonnevie O, et al: Smoking and alcohol intake as risk factors for bleeding and perforated peptic ulcers: a population-based cohort study. Epidemiology 11:434, 2000

12. Bardhan KD, Graham DY, Hunt RH, et al: Effects of smoking on cure of *Helicobacter pylori* infection and duodenal ulcer recurrence in patients treated with clarithromycin and omeprazole. Helicobacter 2:27, 1997

13. Cohen H: Peptic ulcer and *Helicobacter pylori*. Gastroenterol Clin North Am 29:775, 2000

14. Solomon DH, Gurwitz JH: Toxicity of nonsteroidal anti-inflammatory drugs in the elderly: is advanced age a risk factor? Am J Med 102:208, 1997

15. Dalton SO, Johansen C, Mellemkjaer L, et al: Use of selective serotonin reuptake inhibitors and risk of upper gastrointestinal tract bleeding: a population-based cohort study. Arch Intern Med 163:59, 2003

16. Gostout CJ: Patient assessment and resuscitation. Gastrointest Endosc Clin N Am 9:175, 1999

17. Sharara AI, Rockey DC: Gastroesophageal variceal hemorrhage. N Engl J Med 345:669, 2001

18. Cales P, Zabotto B, Meskens C, et al: Gastroesophageal endoscopic features in cirrhosis: observer variability, interassociations, and relationship to hepatic dysfunction. Gastroenterology 98:156, 1990

19. Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices: a prospective multicenter study. The North Italian Endoscopic Club for the Study and Treatment of Esophageal Varices. N Engl J Med 319:983, 1988

20. Sugawa C, Steffes CP, Nakamura R, et al: Upper GI bleeding in an urban hospital: etiology, recurrence, and prognosis. Ann Surg 212:521, 1990

21. Kasapidis P, Georgopoulos P, Delis V: Endoscopic management and long-term follow-up of Dieulafoy's lesions in the upper GI tract. Gastrointest Endosc 55:527, 2002

22. Norton ID, Petersen BT, Sorbi D: Management and long-term prognosis of Dieulafoy lesion. Gastrointest Endosc 50:762, 1999

23. Buttenschoen K, Buttenschoen DC, Odermath R, et al: Diverticular disease-associated hemorrhage in the elderly. Langenbecks Arch Surg 386:8, 2001

24. Freeman SR, McNally PR: Diverticulitis. Med Clin North Am 77:1149, 1993

25. McGuire HH Jr: Urgent colonoscopy for the diagnosis and treatment of severe diverticular hemorrhage. N Engl J Med 342:1609, 2000

26. Hodgson H: Hormonal therapy for gastrointestinal angiodysplasia. Lancet 359:1630, 2002

27. Imdahl A: Genesis and pathophysiology of lower gastrointestinal bleeding. Langenbecks Arch Surg 386:1, 2001

28. Hochter W, Weingart J, Kuhner W, et al: Angiodysplasia in the colon and rectum: endoscopic morphology, localisation and frequency. Endoscopy 17:182,1985

29. Rosen L, Bub DS, Reed JF 3rd, et al: Hemorrhage following colonoscopic polypectomy. Dis Colon Rectum 36:1126, 1993

30. Sorbi D, Norton I, Conio M, et al: Postpolypectomy lower GI bleeding: descriptive analysis. Gastrointest Endosc 51:690, 2000

31. Lingenfelser T, Ell C: Gastrointestinal bleeding in the elderly. Best Pract Res Clin Gastroenterol 15:963, 2001

32. Egglin TK, O'Moore PV, Feinstein AR, et al: Complications of peripheral arteriography: a new system to identify patients at increased risk. J Vasc Surg 22:787, 1995

33. Koval G, Benner KG, Rosch J, et al: Aggressive angiographic diagnosis in acute lower gastrointestinal hemorrhage. Dig Dis Sci 32:248, 1987

34. Ng DA, Opelka FG, Beck DE, et al: Predictive value of technetium Tc 99m-labeled red blood cell scintigraphy for positive angiogram in massive lower gastrointestinal hemorrhage. Dis Colon Rectum 40:471, 1997

35. Feitoza A, Rajan E, Gostout CJ: Overt gastrointestinal bleeding: diagnostic and therapeutic yield of push-enteroscopy. Am J Gastroenterol 96:S104, 2001

36. Lopez MJ, Cooley JS, Petros JG, et al: Complete intraoperative small-bowel endoscopy in the evaluation of occult gastrointestinal bleeding using the Sonde enteroscope. Arch Surg 131:272, 1996

37. Fullarton GM, Birnie GG, Macdonald A, et al: Controlled trial of heater probe treatment in bleeding peptic ulcers. Br J Surg 76:541, 1989

38. Lin HJ, Tseng GY, Perng CL, et al: Comparison of adrenaline injection and bipolar electrocoagulation for the arrest of peptic ulcer bleeding. Gut 44:715, 1999

39. Bleau BL, Gostout CJ, Sherman KE, et al: Recurrent bleeding from peptic ulcer associated with adherent clot: a randomized study comparing endoscopic treatment with medical therapy. Gastrointest Endosc 56:1, 2002

40. Palmer KR: Intravenous omeprazole after endoscopic treatment of bleeding peptic ulcers. Gut 49:610, 2001

41. Green FW Jr, Kaplan MM, Curtis LE, et al: Effect of acid and pepsin on blood coagulation and platelet aggregation: a possible contributor to prolonged gastroduodenal mucosal hemorrhage. Gastroenterology 74:38, 1978

42. Bandi R, Shetty PC, Sharma RP, et al: Superselective arterial embolization for the treatment of lower gastrointestinal hemorrhage. J Vasc Interv Radiol 12:1399, 2001

43. Rosen RJ, Sanchez G: Angiographic diagnosis and management of gastrointestinal hemorrhage: current concepts. Radiol Clin North Am 32:951, 1994

44. Gostout CJ: The evaluation and management of acute upper gastrointestinal bleeding. Clinical Practice of Gastroenterology. Brandt LJ, Daum F, Eds. Churchill Livingstone, Philadelphia, 1998, p1514

45. Zuccaro G Jr: Management of the adult patient with acute lower gastrointestinal bleeding. American College of Gastroenterology. Practice Parameters Committee. Am J Gastroenterol 93:1202, 1998

46. Sclerotherapy for actively bleeding esophageal varices in male alcoholics with cirrhosis. Veterans Affairs Cooperative Variceal Sclerotherapy Group. Gastrointest Endosc 46:1, 1997

47. Bleau BL: Endoscopic management of the acute variceal bleeding event. Gastrointest Endosc Clin N Am 9:189, 1999

48. Stiegmann GV, Goff JS, Michaletz-Onody PA, et al: Endoscopic sclerotherapy as compared with endoscopic ligation for bleeding esophageal varices. N Engl J Med 326:1527, 1992

49. Lo GH, Lai KH, Cheng JS, et al: Emergency banding ligation versus sclerotherapy for the control of active bleeding from esophageal varices. Hepatology 25:1101, 1997

50. Sarin SK, Govil A, Jain AK, et al: Prospective randomized trial of endoscopic sclerotherapy versus variceal band ligation for esophageal varices: influence on gastropathy, gastric varices and variceal recurrence. J Hepatol 26:826, 1997

51. Laine L, el-Newihi HM, Migikovsky B, et al: Endoscopic ligation compared with sclerotherapy for the treatment of bleeding esophageal varices. Ann Intern Med 119:1, 1993

52. Nader A, Grace ND: Pharmacological intervention during the acute bleeding episode. Gastrointest Endosc Clin N Am 9:301, 1999

53. Double-blind randomized controlled trial comparing terlipressin and somatostatin for acute variceal hemorrhage. Variceal Bleeding Study Group. Gastroenterology 111:1291, 1996

54. McKusick MA: Interventional radiology for the control and prevention of bleeding. Gastrointest Endosc Clin N Am 9:311, 1999

55. Zhuang ZW, Teng GJ, Jeffery RF, et al: Long-term results and quality of life in patients treated with transjugular intrahepatic portosystemic shunts. AJR Am J Roentgenol 179:1597, 2002

56. Ahlquist DA: Approach to the patient with occult gastrointestinal bleeding. Textbook of Gastroenterology, 4th ed. Yamada T, Alpers DH, Kaplowitz N, et al, Eds. Lippincott Williams & Wilkins, Philadelphia (in press)

57. Descamps C, Schmit A, Van Gossum A: "Missed" upper gastrointestinal tract lesions may explain "occult" bleeding. Endoscopy 31:452, 1999

58. Rockey DC, Celo JP: Evaluation of the gastrointestinal tract in patients with iron deficiency anemia. N Engl J Med 329:1691, 1993

59. Kendrick ML, Buttar NS, Anderson MA: Contribution of intraoperative enteroscopy in the management of obscure gastrointestinal bleeding. J Gastrointest Surg 5:162, 2001

60. Gostout CJ, Ahlquist DA, Radford CM, et al: Endoscopic laser therapy for watermelon stomach. Gastroenterology 96:1462, 1989

Acknowledgment

Figures 1 through 4 and 7 © 2003 Mayo Foundation for Medical Education and Research. Used by permission.

XI DISEASES PRODUCING MALABSORPTION AND MALDIGESTION

Charles M. Mansbach II , m.d.

Definition

Malabsorption refers to the impaired intestinal absorption of nutrients. It can result from congenital defects in the transport of nutrients or from acquired defects in the absorptive surface of the intestinal epithelium. Maldigestion, another factor in nutrient absorption, refers to the impaired digestion of nutrients within the intestinal lumen. Although these two processes are pathophysiologically distinct, they are interdependent; and in clinical practice, malabsorption has come to signify derangements in both processes.

Overview of Diseases Producing Malabsorption

Malabsorption is clinically defined as impaired absorption of fat (steatorrhea), because measuring fat absorption is the best indicator of the normality of the overall process of nutrient absorption. Under certain conditions, however, fat absorption may be normal but other substances may be poorly absorbed, such as iron, calcium, bile salts, or, in certain hereditary conditions, specific amino acids, disaccharides, or monosaccharides.

etiology

Generally, there are three possible causes of fat malabsorption: small bowel disease, liver or biliary tract disease, and pancreatic exocrine insufficiency [*see Table 1*].

Small Bowel Disease

Small bowel disease can result in moderate amounts of fat in the stool (7 to 30 g/day on a diet containing 100 g of fat). Patients with small bowel disease may leak protein (protein-losing enteropathy) through a diseased intestinal mucosa, which results in a reduced serum albumin concentration. Deficiencies of fat-soluble vitamins (i.e., vitamins A, D, E, and K) may be present in small bowel disease. Patients may malabsorb vitamin B_{12} because of a very diseased or previously resected (usually over 60 cm) terminal ileum. Folic acid may also be malabsorbed, and hypocalcemia and hypomagnesemia may also be present.

Liver or Biliary Tract Disease

Patients with liver or biliary tract disease usually have only small increases in fat in the stool (7 to 15 g/day) and may also malabsorb fat-soluble vitamins. The association of cholestatic liver disease, especially primary biliary cirrhosis, with osteoporosis is well known. Osteoporosis may be the presenting symptom of the liver disease. Vitamin K deficiency, as shown by a prolonged prothrombin time, may also occur. Administration of vitamin K corrects the clotting defect in cases where the extent of the liver disease is not severe enough to impede clotting factor synthesis.

Pancreatic Exocrine Insufficiency

Patients with pancreatic exocrine insufficiency may have up to 80 g of fat/day in the stool. That they absorb fat at all is the result of the action of gastric lipases. Gastric lipase is present in the chief cells of the human stomach[1] and is thought to account for any lipid absorbed in the setting of chronic pancreatitis as exemplified by cystic fibrosis. Indeed, in cystic fibrosis, an increase in gastric lipase has been reported.[2]

clinical manifestations

The symptoms of malabsorption are protean. In the most obvious case, the patient complains of weight loss despite a good appetite. In these cases, there is a clear change in the quality of the stool and usually an increase in stool number. The consistency of the stool softens, and in the presence of excess fat, the stool becomes more malodorous and is difficult to flush down the toilet. Oil drops or a lipid sheen may appear on the water. Excess gas in the stool causes the stool to float.[3]

Depending on other dietary constituents that are malabsorbed, patients may experience a distended abdomen, borborygmi, abdominal cramps (lactose intolerance), easy bruising (vitamin K deficiency), osteopenia or tetany (vitamin D deficiency and calcium malabsorption), iron deficiency, or night blindness (vitamin A deficiency). The most challenging cases are those in which the question of malabsorption is not raised because of lack of change in the quality of the stools.

The diarrhea of malabsorption is classified as an osmotic diarrhea and usually stops during fasting. In fat malabsorption, the diarrhea is caused not only by the excessive osmotically active

Table 1 Causes of Malabsorptive Syndromes

Diseases of the small intestine	Gluten-sensitive enteropathy Tropical sprue Collagenous sprue Eosinophilic enteritis Radiation enteritis Amyloidosis Mastocytosis Abetalipoproteinemia Whipple disease Intestinal lymphangiectasia Immunoproliferative small intestinal disease Ischemic bowel disease *Giardia lamblia* infection AIDS Short bowel syndrome Ileal resection Ileitis (e.g., Crohn disease)
Diseases of the liver and biliary tract	Cirrhosis/parenchymal liver disease Intrahepatic cholestasis syndrome Cholestasis due to extrahepatic obstruction
Diseases of the pancreas	Chronic pancreatitis Cystic fibrosis Cancer
Combined or multiple defects in digestion and absorption	Hyperthyroidism Diabetes mellitus Carcinoid syndrome Zollinger-Ellison syndrome Postgastrectomy (Billroth II type)

Table 2 Tests of Digestive-Absorptive Function

	Characteristics	*Clinical Use*
Fecal fat analysis		
Qualitative	Simple microscopic study for increase in fat globules	A good screening test for moderate increase in stool fat, but quantitative fecal fat analysis is preferable
Quantitative	Chemical analysis for fat excretion during a 72 hr period by titration with NaOH; most sensitive test for malassimilation of fat; normal is < 6 g/day; does not distinguish between small intestine, pancreatic, or luminal abnormalities	The most important test to identify maldigestion or malabsorption; indicated in all patients suspected of having malassimilation
Xylose absorption	As a pentose not requiring luminal or intestinal surface digestion, xylose allows assessment of small intestine function; normally, > 4 g/5 hr is excreted in urine after ingestion of 25 g; plasma xylose should be 10–20 mg/dl/1.73 m^2 of body surface area at 60–75 min	Indicated whenever quantitative fecal fat is abnormal; not as sensitive as fat analysis but localizes the abnormality to the small intestine
Small intestine x-ray	Allows analysis of continuity of small intestine and identification of diverticula or alteration of mucosa; diseased pancreas may impinge on duodenum	Indicated when quantitative fecal fat excretion is increased
Small intestine peroral biopsy	Permits direct histologic examination of mucosa; characteristic alterations occur in several diseases producing malabsorption	Indicated when fecal fat excretion is increased, particularly if the xylose test or small intestine x-ray is abnormal; a portion of biopsy may be assayed for disaccharidases
Bile acid breath test	In small intestine bacterial overgrowth or ileal disease that produces malabsorption, ^{14}C–glycocholic acid (5 μCi) will be deconjugated, metabolized, and excreted via the lungs as $^{14}CO_2$	Indicated in patients with documented steatorrhea caused by suspected bacterial overgrowth or ileal dysfunction
Bentiromide test	The peptide bond in this nonabsorbable arylamine is cleaved specifically by intraluminal chymotrypsin to yield PABA, which is then readily absorbed and excreted by the intestine	Indicated when fecal fat excretion is increased; less sensitive than quantitative fat analysis but, when positive, establishes insufficiency of intraduodenal pancreatic digestive enzyme levels Not available in U. S.

particles but also by fatty acids, which stimulate cyclic adenosine monophosphate (cAMP)–dependent Cl^- secretion.

Specific physical findings of various diseases may accompany the malabsorptive state and assist in making the diagnosis. For example, the skin changes of scleroderma or dermatitis herpetiformis may be present. Signs of diabetic neuropathy may be disclosed. Although thyrotoxicosis may be associated with excessive fat in the stool, patients with thyrotoxicosis usually eat gluttonously but absorb a normal percentage of dietary fat eaten (95%) and therefore do not malabsorb in the true sense.

TESTS FOR SUSPECTED MALABSORPTION

The tests for malabsorption involve determining whether there is excessive fecal fat excretion [*see Table 2*]. Protein is produced in large quantities by the digestive tract, especially the pancreas, making creatorrhea difficult to interpret. Malabsorbed carbohydrate delivered to the colon may be metabolized by colonic bacteria to short-chain fatty acids, which are in part absorbed by the colon. Thus, the quantitative measurement of carbohydrate absorption is inaccurate, although a fall in stool pH occurs, which is indicative of excessive amounts of the short-chain fatty acids that are excreted under these conditions.

Measurement of Fecal Fat

Fecal fat can be measured qualitatively and quantitatively. The qualitative measurement of fecal fat using Sudan III staining has been shown to be surprisingly accurate,[4] especially if clinically significant amounts of fat are being excreted. One group reported that counting and measuring the size of fat globules present in the stool significantly improved the sensitivity and specificity of the Sudan assay.[4] As with many qualitative tests, however, accuracy varies with test performance and interpretation, making the skill of the observer crucial to success.

The quantitative measurement of fecal fat is the benchmark by which all other tests are ranked. It is important to remember that the test cannot be performed unless the patient is able to eat at least 80 g, preferably 100 g, of fat a day.

Xylose Absorption Test

The absorptive surface area of the intestine is measured by the ability of the patient to absorb xylose, a sugar. Unlike glucose, xylose is not actively absorbed by the intestine but is absorbed by the slower process of passive diffusion. In the xylose absorption test, 25 g of xylose is given by mouth and the urine is collected for 5 hours. The normal urinary excretion of xylose is greater than 4 g over 5 hours. For an adequate urinary flow to be ensured, the patient should drink 500 ml of water after drinking the xylose. This intake should result in a urine volume of at least 300 ml during the collection period. Xylose excretion can be falsely low in patients with reduced renal function or in patients with ascites in which the xylose is diluted in the ascitic fluid. To avoid falsely low results, it is advisable to measure the concentration of xylose in blood [*see Table 2*]. Malabsorbed xylose reaches the colon and can be metabolized by the resident bacteria to hydrogen. Hydrogen may be quantitated in the breath; this test is reported to be as accurate as the measurement of xylose in the serum or urine.

Imaging Studies

A plain film or ultrasonogram of the abdomen is usually not helpful in most cases of malabsorption. However, 30% of cases of chronic pancreatitis have visible calcifications on an abdominal plain film. Detection of pancreatic calcification can be increased if computed tomography or ultrasonography is used. CT or ultrasonography also can identify dilatated pancreatic ducts, another characteristic sign of chronic pancreatitis. Endoscopic retrograde pancreatography (ERP) can also be helpful when ductular changes indicative of chronic pancreatitis are seen [*see 4:V Diseases of the Pancreas*].

Radiographic studies of the small intestine after oral ingestion of barium can aid in the diagnosis of several abnormalities. The

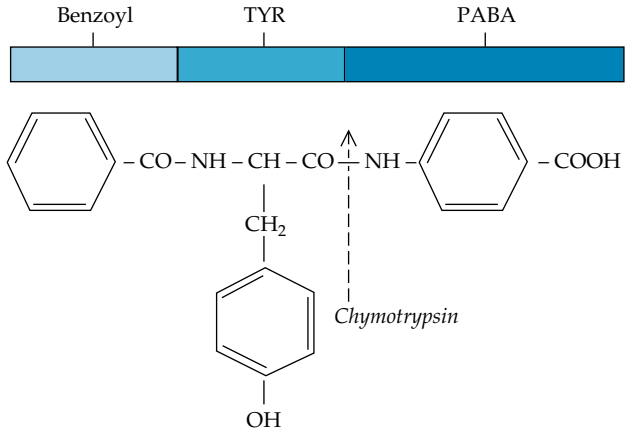


Figure 1 **Cleavage of the bentiromide molecule by the intraduodenal enzyme chymotrypsin yields two fragments, benzoyl-tyrosine and aminobenzoic acid (PABA). Bentiromide is composed of benzoyl (light blue), tyrosine (TYR) (blue), and PABA (dark blue). The released PABA is absorbed across the intestine and excreted in significant quantities in the urine. Absence of chymotrypsin, as a result of pancreatic disease or duct obstruction, will result in failure of release, absorption, and urinary excretion of PABA.**

presence of diverticula of the small intestine or of impaired peristalsis, as seen in scleroderma or idiopathic intestinal dysmotility, can be an indicator of bacterial overgrowth. A careful examination of the terminal ileum can identify Crohn disease. Stricturing may be identified in some patients with radiation injury or injury caused by nonsteroidal anti-inflammatory drugs. Hypoalbuminemia affecting the small intestine may lead to the so-called stack-of-coins sign.

Small Bowel Biopsy

An experienced pathologist can be helpful in supporting the diagnoses of gluten-sensitive enteropathy (with or without dermatitis herpetiformis), hypogammaglobulinemic sprue, tropical sprue, Whipple disease, *Mycobacterium avium* complex disease, stasis syndrome, amyloidosis, and intestinal lymphangiectasia.

Assessment of Pancreatic Exocrine Function

More than 90% of pancreatic exocrine function needs to be destroyed before symptomatic malabsorption results [*see 4:V Diseases of the Pancreas*].[5] The most sensitive test of pancreatic exocrine function requires the passage of a double-lumen tube.[6] Cholecystokinin (CCK) or secretin is administered intravenously, and gastric and duodenal secretions are collected separately. However, secretin became unavailable in the United States when the manufacturer discontinued production in 1999. If CCK is given, lipase or trypsin activity is determined using appropriate substrates. When secretin is administered, duodenal fluid volume and bicarbonate concentration are measured.

The noninvasive bentiromide test is based on the action of trypsin on bentiromide to yield *p*-aminobenzoic acid (PABA) and benzoyl-tyrosine [*see Figure 1*]. PABA is readily absorbed by the intestine and excreted into the urine. In healthy persons, when 500 mg of bentiromide is ingested, 57% or more of the PABA appears in the urine within 6 hours. In patients with chronic pancreatitis, the amount of PABA excreted is significantly less, averaging 42%. Using the 57% excretion as a cutoff, the sensitivity of the bentiromide test is 67% to 80% and the specificity is 95%.[7] PABA may also be measured in the plasma 120 minutes after ingestion of bentiromide, which may enhance the sensitivity of the test.[8] Plasma measurements are helpful in cases of impaired renal excretion, which may be seen in the elderly. PABA is identified colorimetrically, and thus, other arylamines can interfere with its determination (e.g., acetaminophen, lidocaine, procainamide, sulfonamides, and thiazide diuretics).[9] In cases in which intestinal absorption is impaired, such as with sprue, the absorption of released PABA may be reduced, leading to a falsely low urinary recovery. Unfortunately, the bentiromide test becomes positive only when the pancreatic gland is more than 90% destroyed. Nevertheless, in considering the workup of a patient with steatorrhea, the test may be useful because it takes an equal amount of glandular destruction to generate steatorrhea.

Although the vast majority of pancreatic proteases and lipases are stored in zymogen granules and are released from the apical portion of the pancreatic exocrine cell into the pancreatic duct, a small percentage leaks into the interstitium of the gland, is carried into the circulation, and can be measured by the serum trypsinogen assay. Because the activation peptide of trypsin is not yet released and any active trypsin is quickly bound by α_1-antitrypsin, the free-circulating form of trypsin is trypsinogen. In patients who have chronic pancreatitis with exocrine insufficiency, the serum concentration of trypsinogen is lower than that in healthy persons (2 to 18 ng/ml, compared with 29 to 79 ng/ml in healthy persons).[10] A low serum trypsinogen level appears to have a high degree of specificity for chronic pancreatitis but is only modestly sensitive.

Bile Acid Absorption Tests

Bile acids are synthesized from cholesterol in the liver and require conjugation by either glycine or taurine before they are excreted into the intestine via the common bile duct. The bile acid conjugates solubilize the products of triacylglycerol hydrolysis into complex micelles, which facilitate the rapid absorption of dietary lipid. Bile acids are not absorbed in the proximal intestine with dietary lipid but in the distal ileum. The bile acid pool recirculates six times a day. About 95% of bile acids are reabsorbed and recirculate in the enterohepatic circulation each day; approximately 0.5 g of bile acids appears in the stool daily, which equals the hepatic synthetic rate under steady-state conditions. If bile acids are not adequately absorbed, diarrhea results (choleretic enteropathy). In the complete absence of bile salts, fatty acids are less efficiently absorbed, with up to 25% to 50% of ingested lipid appearing in the stool. In patients with idiopathic diarrhea or with diarrhea after ileal resection (≥ 30 cm), the malabsorption of bile acids is an etiologic possibility. Also, children who have unexplained diarrhea may have a congenital defect of the sodium-dependent bile salt transporter in the terminal ileum.[11]

To test for the presence of bile acid malabsorption, two methods are available, although not widely used. The first is the ^{14}C-glycocholic acid breath test, and the second is the selenium-75–labeled homocholic acid–taurine (^{75}SeHCAT) absorption test. In the former test [*see Figure 2*], a trace amount of ^{14}C-glycocholic acid is given by mouth. Many bacteria are capable of hydrolyzing the amide bond and releasing the ^{14}C-glycine; either it is absorbed and $^{14}CO_2$ is produced in the liver or it is further metabolized in the intestinal lumen to $^{14}CO_2$. In either event, the $^{14}CO_2$ appears in the breath in measurable amounts. The percentage of the ingested dose excreted in the breath increases if the intestinal lumen contains more bacteria than normal or if an excess of bile acids is delivered to the colon (ileal dysfunction). A gastric anti-

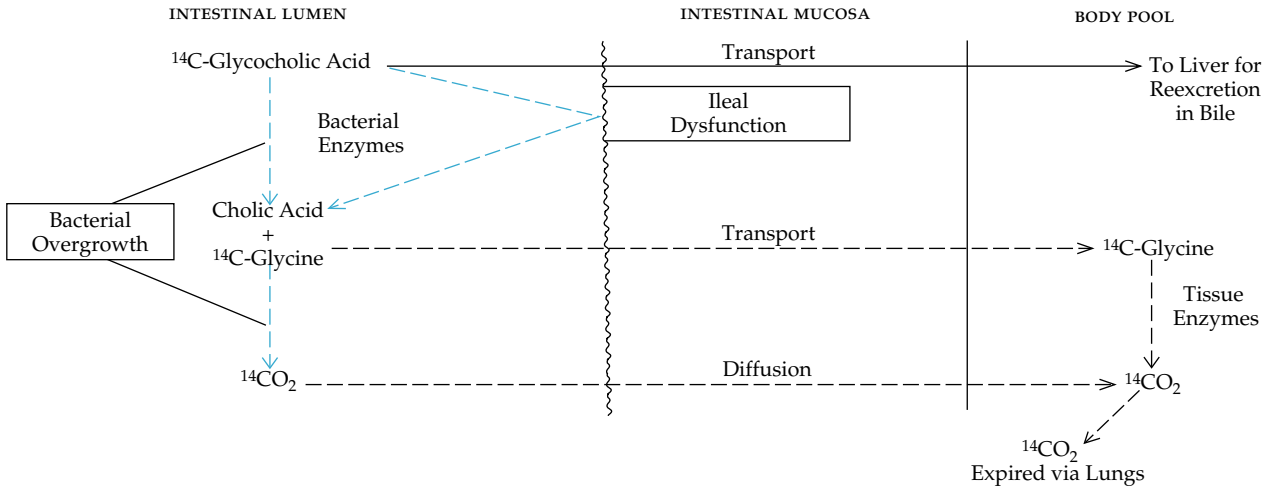


Figure 2 **In the bile acid breath test, a small dose of ^{14}C–glycocholic acid is ingested and its fate determined by measurement of $^{14}CO_2$ excretion in breath. In a normal person, little of the ^{14}C–glycocholic acid is metabolized for excretion in breath because it passes intact to the ileum for absorption and return to the enterohepatic circulation. If there is either intestinal bacterial overgrowth or ileal dysfunction, however, bacterial enzymes will deconjugate the bile acid (broken blue lines), releasing cholic acid and ^{14}C-glycine. The radioactive glycine may be transported across the intestinal mucosa (upper broken gray line) and subsequently degraded to $^{14}CO_2$ by tissue enzymes; alternatively, the ^{14}C-glycine may be metabolized within the intestinal lumen to $^{14}CO_2$, which then diffuses (lower broken gray line) into the circulation and is carried to the lungs. Consequently, $^{14}CO_2$ excretion is 10 times greater in either intestinal bacterial overgrowth or ileal dysfunction than it is in the normal state.**

secretory drug may also increase the resident population of bacteria in the intestine to a level that results in an abnormal breath test.[12] The usefulness of this test as an indicator of bile acid malabsorption is therefore limited. The ^{75}SeHCAT test has more potential clinical usefulness because of its strong correlation with cholate excretion and the ease of measurement of ^{75}Se retention by the whole-body gamma camera. Normal persons retain greater than 19% of an orally administered dose of ^{75}Se after 7 days, whereas patients with significant ileal dysfunction or resection retain less than 12%.[13]

Now that the human sodium-dependent bile acid transporter has been cloned, congenital defects are being discovered that lead to bile acid malabsorption resulting in diarrhea.[11] Such defects may be the cause of primary bile acid malabsorption.

Diseases Producing Malabsorption

GLUTEN-SENSITIVE ENTEROPATHY

Gluten-sensitive enteropathy (GSE) was once called celiac disease in children and idiopathic or nontropical sprue in adults. In 1960, it was recognized that the diseases were the same, caused by the major wheat protein gluten and, more specifically, its alcohol-soluble component, gliadin.[14]

GSE is generally considered less common in the United States than in Western Europe. However, a recent large multicenter study indicated that the prevalence of GSE in the United States in symptomatic patients (1 in 56) and in not-at-risk persons (1 in 133) is similar to that reported in Europe.[15]

Genetic and Etiologic Factors

GSE is associated with haplotypes HLA-DQ2 (DQA1*501, DQB1* 201) and HLA-DQ8 (DQA1*031, DQB1*302). In sets of monozygotic twins, 30% of the opposite twins from the incidence case do not have GSE. This condition leads to the belief that there is another, unknown (nongenetic) factor that is important in disease causation. A 33-amino-acid peptide part of gliadin has been shown to be poorly digested by proteases and to cause T cell activation in GSE patients.[16]

Pathogenesis

GSE is associated with impaired cholecystokinin release. CCK cells are either reduced in number or so defective that the amount of CCK present in the duodenal mucosa is greatly reduced.[14] This CCK deficiency leads to a reduced amount of pancreatic lipase and bile acids delivered to the intestinal lumen in response to dietary lipid. The intestinal crypt cells are the major fluid-secreting cells of the intestine, via their cAMP-dependent Cl^- secretion with attendant water secretion. In GSE, the cryptal portion of the villous complex is greatly expanded, leading to increased water secretion. Because the villous tip cells, which normally absorb the water, are diseased and reduced in number, water and electrolyte absorption is not as effective as normal, and the intestine becomes secretory.[14] Thus, the concentration of bile acids in the intestinal lumen is reduced below that expected simply from the impaired CCK release. The ability of bile acids to solubilize the products of lipolysis depends on the presence of bile salts at a concentration greater than their critical micellar concentration (CMC) of 1.4 mM.[17] Normally, the intestine has a postprandial bile salt concentration of 10 mM.[18] The brush borders at the surface of mature enterocytes are severely affected in GSE. Further, the villous structures are flattened. These two conditions lead to a severely reduced surface area that limits lipid absorption. The amount of reduction in surface area can be estimated by the D-xylose absorption test. The enterocytes that are at the surface of the intestine are not as mature as normal enterocytes, because their turnover rate is greatly increased, which probably results in a reduced capacity to process absorbed lipid.

Diagnosis

Clinical manifestations Although GSE may start in childhood and respond to gluten withdrawal, children with the dis-

ease undergo a remission in their teenage years even if they ingest a diet containing gluten. As adults, these patients, 25% or more of whom were symptomatic in childhood, may present with a variety of complaints; usually, weight loss, fatigue, abdominal cramps, distention, bloating, and diarrhea (steatorrhea) are prominent, although there may be no loss of appetite. In some patients, the disease is insidious in onset and the symptoms are mild. It is only after these patients have been treated that they realize, in retrospect, how ill they were. In population studies in which the presence of disease was determined by intestinal biopsy, people with a biopsy consistent with GSE were often asymptomatic but sometimes of shorter stature than unaffected siblings. Because nothing specifically leads to the diagnosis, especially in the absence of clinically evident steatorrhea, the realization that the patient has GSE may be delayed. This problem is most likely to occur with patients who do not have steatorrhea but do have osteoporosis, easy bruising as a result of vitamin K deficiency, or unexplained iron deficiency anemia. Iron deficiency anemia is common in adult patients with GSE,[19] and one study suggested that fractures may be the only sign of malabsorption in patients with undetected GSE.[20]

Laboratory tests In a patient in whom the suspicion of GSE is high (e.g., a first-degree relative of a known GSE patient; a patient who has a history of a childhood disease that caused diarrhea, was evaluated by a specialist, and was treated with a special diet; or a patient with malabsorption who is not an alcoholic and does not have another obvious reason for malabsorbing fat), a positive tissue transglutaminase antibody test makes the diagnosis almost certain [*see Figure 3*].[21] Alternatively, the diagnosis might rest on small bowel biopsy findings [*see Figure 4*]. Classic features include partial or complete villous atrophy, abnormal-appearing enterocytes at the villous tips, an increase in intraepithelial lymphocytes, a lamina propria infiltrate consisting predominantly of lymphocytes and macrophages, an increase in the size of the crypts both vertically and horizontally, and an increase in the number of mitotic figures.[14] These features, although typical, are not pathognomonic. For the diagnosis to be definitive, the patient must respond to dietary therapy. Symptomatic improvement can be expected in 80% of patients within 1 month, but histologic improvement lags behind considerably. Another 10% of patients do not respond until after 2 months, and the remainder may take up to 2 years. Even under the strictest dietary control, the biopsy findings might not return to normal. Most often, the patient remains under good control because of the symptomatic improvement while on the diet, but many patients will eventually either test whether they are cured or be in a situation that forces them to commit a dietary indiscretion. This lapse inevitably results in recurrence of symptoms, further securing the diagnosis.

Another helpful test is the identification of an antiendomysial antibody. This antibody is present in up to 95% of cases and is rarely present in control subjects.[22] Other tests of malabsorption, such as the D-xylose absorption test or stool fat studies, may be abnormal. Low clotting factors, the presence of anemia caused by folate or iron deficiency, or osteoporosis may also be present. None of these conditions are specific for GSE, however.

Treatment

The treatment of GSE is a strict gluten-free diet. Because a gluten-free diet is a lifelong commitment, is more expensive than a normal diet, and may carry social limitations, it should not be recommended until the diagnosis of GSE is firmly established. A gluten-free diet prohibits the intake of wheat, rye, and barley. Oats are thought to be safe but are usually avoided during the early stage, when the clinical response to the diet is being judged. Keeping the patient on the diet is sometimes difficult because many foods have hidden gluten content. Maintaining a gluten-free diet is important because intestinal lymphomas are more likely to develop in patients who do not follow such a diet.[23]

Support groups, such as those organized by the Celiac Society of America, can be helpful, especially when the disease is newly diagnosed. Information such as what to look for on package labels and interesting recipes can be very instrumental in helping the patient maintain the gluten-free diet. During the trial period, beer, ale, and whiskey, which may contain enough gluten to sensitize the patient, should not be consumed. After the dietary response is clear, these drinks may be tried, if desired, to determine whether the patient is sensitive. Other products that are not usually thought of as containing gluten, but often do, are ice

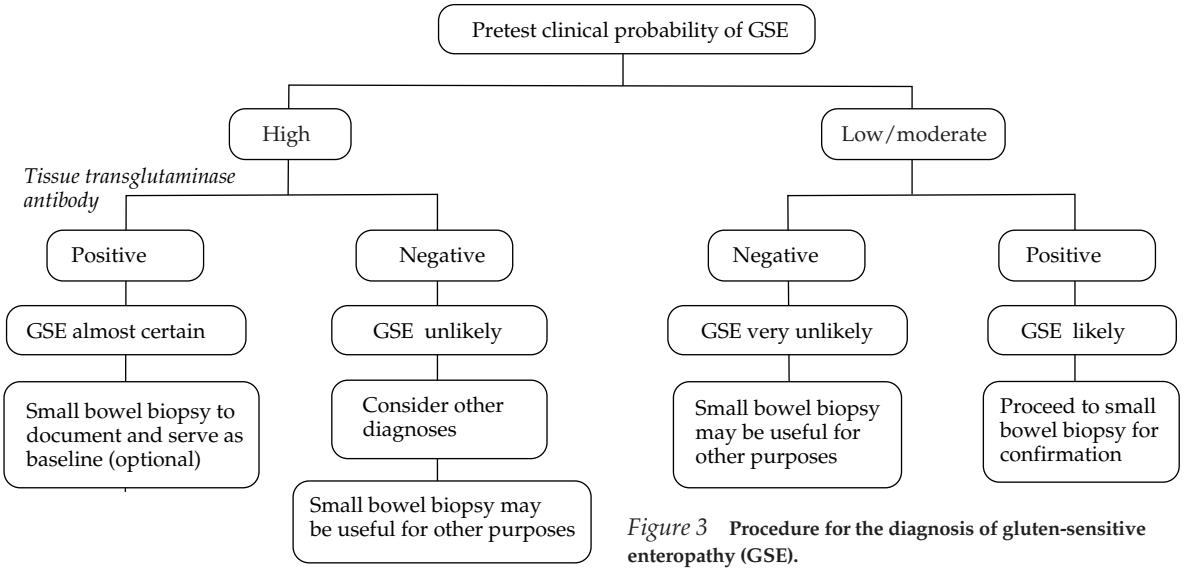


Figure 3 **Procedure for the diagnosis of gluten-sensitive enteropathy (GSE).**

a *b*

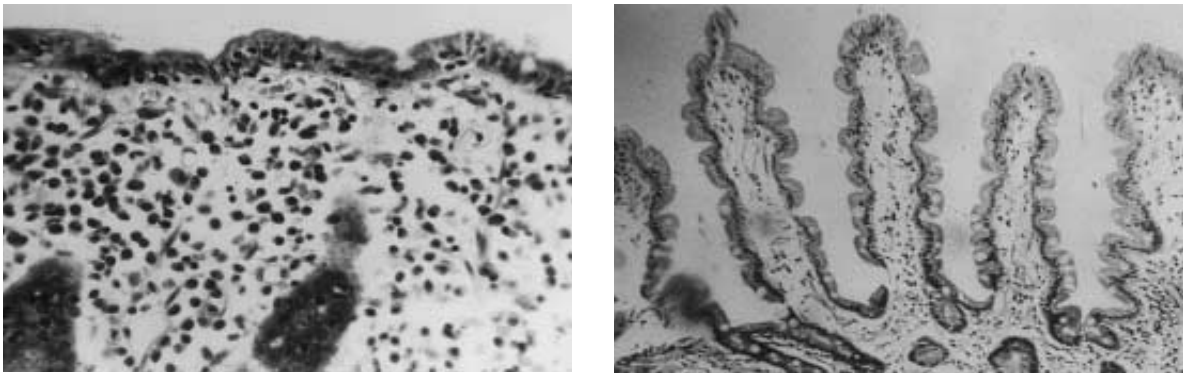


Figure 4 **Biopsy specimen from the small intestine of a patient with untreated celiac sprue (*a*) demonstrates a flat surface with plasmocytic infiltration of the subepithelial region (magnification: 400×). In contrast, a biopsy sample taken from a patient with pancreatic exocrine insufficiency (*b*) is indistinguishable from a normal specimen and shows tall, scalloped villi and minimal subepithelial mononuclear infiltration (magnification: 100×).**

cream, communion wafers, and even some drugs (as a filler). Despite the restrictions, many dietary options are open to the patient, including certain breakfast cereals, milk, cheese, eggs, meat, chicken, fish, chocolate, and products made from corn, rice, or potato flour.

If the patient does not respond, the most likely reason is that the patient is not accurately following the diet. In such cases, it is helpful to have a dietitian carefully go over the patient's dietary history.[14] Less often, the patient will have an intestinal stasis syndrome or pancreatic insufficiency. When these subsidiary problems are diagnosed and successfully treated, the patient usually shows a response to the diet. However, a minority of patients who do not respond to a gluten-free diet have what is called refractory sprue.[24] Patients with refractory sprue may require treatment with corticosteroids and other immunosuppressants, including azathioprine.[25]

GSE-Associated Disorders

Dermatitis herpetiformis Many patients with dermatitis herpetiformis have GSE.[21] The intensely pruritic, blistering lesions appear on the knees, elbows, shoulders, and buttocks [*see 2:I Cutaneous Manifestations of Systemic Diseases*]. Skin biopsies of dermatitis herpetiformis lesions have characteristic immunoglobulin A (IgA) deposits. On a gluten-exclusion diet, both the dermatologic and the intestinal lesions improve, indicating a linkage between the two. However, the skin lesions respond to dapsone treatment and the intestinal lesions do not, which indicates that there are differences between the two diseases as well.

Type 1 diabetes mellitus Patients with type 1 diabetes mellitus are at increased risk of developing GSE. With the use of antiendomysial antibodies as a screen, gluten-sensitive enteropathy has been found in three of 47 diabetic patients (6%), a much higher incidence than would be expected by chance.[26] Even when patients with type 1 diabetes mellitus show no apparent signs of malabsorption, they are at risk for developing celiac disease; screening for celiac disease in these patients therefore may be advisable.[27] One study reported a 5.7% prevalence of celiac sprue in patients with type 1 diabetes and noted a finding of increased autoimmune diseases in these patients.[28] The prevalence of autoimmune disorders in patients with celiac disease seems related to duration of gluten exposure, which provides further rationale for the early diagnosis and treatment of GSE.[29]

OTHER SPRUELIKE DISORDERS

Tropical Sprue

Tropical sprue is a malabsorptive illness that appears in certain areas of the world, especially the tropics, in both the indigenous populations and tourists. In two carefully studied populations, 5% to 13% of North Americans living in Puerto Rico for 6 months or longer experienced symptoms of tropical sprue. Expatriates from the United States who return from the tropics or other areas endemic for tropical sprue may experience symptoms of tropical sprue more than 10 years after their return.[30] Peace Corps volunteers from the United States who spent time in Pakistan had demonstrable small bowel lesions and functional abnormalities that reverted to normal over several months after returning home.[31] Indians and Pakistanis living in the United States may take a longer time (up to 4 years) to excrete normal amounts of D-xylose.[32] Exactly what causes these changes in the small bowel is not clear, but the tropical sprue syndrome is thought to be caused by one or more species of coliform bacteria, such as *Klebsiella* species,[33] which colonize the upper intestinal tract.

Diagnosis The symptoms of tropical sprue differ from those of GSE. Weight loss caused mostly by anorexia is very prominent, as is diarrhea. A sore tongue (70% of patients), pedal edema (25%), folate and vitamin B_{12} deficiency (75% to 100%), or an abnormal result on the Schilling test (96% to 100%) is much more common in tropical sprue than in nontropical sprue.[33] The symptoms can be quite severe, sometimes leading to death in endemic areas. However, the prognosis, in general, is excellent for patients either remaining in the tropics or returning to the United States.

The diagnosis of tropical sprue is made by performing a small bowel biopsy in patients with a compatible clinical presentation and travel history. Villi are leaflike or blunt, and the lamina propria are packed with inflammatory cells [*see Figure 5*]. Thin villous structures are seen in North Americans and Europeans [*see Figure 4*]. In considering this disease, it should be noted that intestinal biopsy results in residents of endemic areas or in tourists who do not stay in mainstream hotels in endemic areas would be classified as abnormal in persons living in the United States or Europe.

Treatment Treatment of tropical sprue should begin with folic acid (5 mg/day).[33] This therapy is associated with rapid improvement in appetite, and it eliminates most of the clinical

symptoms. In patients with a short duration of symptoms (< 4 months), folate given for 6 months to 1 year may suffice. For patients with a longer duration of symptoms (> 4 months), antibiotics, such as tetracycline (2 g/day for 1 year), should be added. Most patients returning to the United States gain weight quickly even if the results of absorption tests or intestinal biopsies are not normalized.

Collagenous Sprue

Collagenous sprue is a rare, devastating disease in which there is a layer of collagen underneath the enterocytes of the small bowel. The relation of collagenous sprue to collagenous colitis is unclear, but the basic histologic feature of subepithelial collagen deposition is the same. The origin of collagenous colitis is unknown, but it develops in approximately half the patients who have refractory celiac disease (those unresponsive to the gluten-exclusion diet).[34] Although it is known that type 6 collagen is deposited in the more commonly diagnosed collagenous colitis, the type of collagen laid down in the small bowel in collagenous sprue is unknown. In collagenous colitis, the symptoms (primarily diarrhea) are usually modest, but in collagenous sprue, symptoms are more severe and include obvious malabsorption. This severity of symptoms is probably caused by the diffusion barrier presented by the collagen, which prevents nutrients from diffusing either into the portal capillaries or into the lymphatics.

Diagnosis The diagnosis of collagenous sprue is made by the classic histologic picture of villous atrophy and subepithelial collagen deposition. If the diagnosis is missed, however, and the patient is thought to have GSE on the basis of the flat villous structure, the patient will usually not be responsive to the gluten-free diet.

Treatment Therapy for collagenous sprue is uncertain. The most common problem is the osmotic diarrhea caused by the gross malabsorptive state induced by the disease process. In this event, the patient is treated as if he or she had the short bowel syndrome. Some patients respond to steroid therapy. A few respond to steroids and a gluten-exclusion diet, with the patient's improved condition eventually making it possible to taper the steroid dosage.[35]

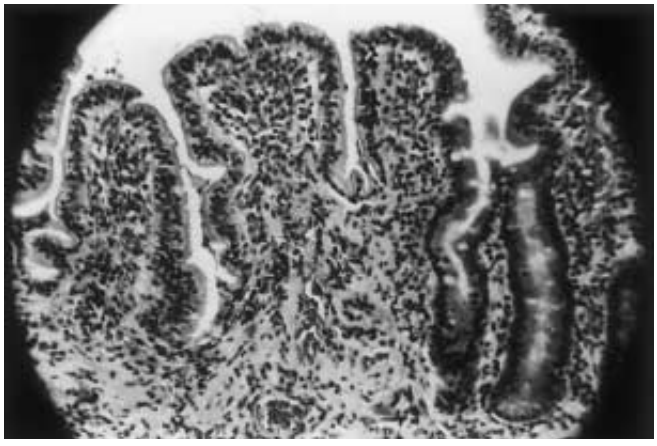


Figure 5 **In a biopsy specimen from the small intestine of a patient with tropical sprue, villi are broadened and shortened and the crypts are deepened; these changes yield a villus-to-crypt ratio of 1:1 (magnification: 100×).**

Hypogammaglobulinemic Sprue

The gastrointestinal tract is the largest lymphoid organ in the body. The environment to which this immune system is exposed is filled with foreign antigens that must be sorted, identified, and, if necessary, reacted to. Thus, it is not surprising that intestinal dysfunction may develop in patients who are immune deficient, particularly those with IgA deficiency, because IgA is the most important immunoglobulin of the intestine. Some patients who have one of the hypogammaglobulinemic syndromes may experience malabsorption.[36] Patients with IgA deficiency also usually have a history of recurrent respiratory infections,[36] which further distinguishes them from patients who have gluten-sensitive enteropathy. The most common cause of malabsorption seen in this condition is giardiasis.

Diagnosis The diagnosis of hypogammaglobulinemic sprue is suspected if the patient has signs and symptoms of malabsorption and low levels of serum immunoglobulins, especially IgA. Intestinal biopsy specimens lack plasma cells and thus are easily distinguishable from those of patients with gluten-sensitive enteropathy, in which plasma cell types are abundant. Plasma cells are readily seen in normal biopsy specimens as well. *Giardia lamblia* organisms may also be present in hypogammaglobulinemic sprue.

Treatment Frequently, the intestinal symptoms associated with hypogammaglobulinemic sprue improve if metronidazole is given at 750 mg/day for 10 days to treat giardiasis.

SMALL BOWEL DISEASE SECONDARY TO SURGERY AND RADIATION

Short Bowel Syndrome

Massive small bowel resection is used to treat various diseases, including mesenteric ischemia, volvulus, and Crohn disease. Because the intestine requires a certain surface area over which absorption can occur, reducing the area below a critical value results in malabsorption. Depending on the amount of bowel resected, the results can range from mildly inconvenient to catastrophic. Retention of the ileocecal valve lessens symptoms. The ileum responds to jejunal resection by hyperplasia much more effectively than the jejunum responds to an ileal resection. There are also specialized mechanisms present in the ileum that are not available to the jejunum, such as bile salt and vitamin B transporters. The maintenance of an adequate bile acid pool is important for fat absorption because the reduced absorptive surface area in patients who have undergone bowel resection makes it necessary for fat absorption to be as efficient as possible. Alternatively, the ileum can perform most of the functions of the jejunum except for folic acid, Ca^{2+}, and Fe^{2+} absorption. However, these can be replenished by appropriate medication.

Diagnosis The diagnosis of short bowel syndrome is made by a history of bowel resection in combination with clinical manifestations of the syndrome, such as diarrhea, steatorrhea, weight loss, trace-element deficiencies, hyponatremia, and hypokalemia.

Treatment Treatment of patients with short bowel syndrome depends on the part and the amount of the bowel that has been resected. Protein requires the greatest surface area for absorption.[37] Thus, achieving adequate assimilation may become

problematic, despite the water solubility of proteins and their hydrolytic products. Vitamins and minerals also need to be added to any therapeutic regimen, depending on what part of the bowel is missing. Treatment can include eating multiple small meals each day, eating quickly absorbed foods such as canned caloric supplements, having food finely chopped or ground, and eating foods containing medium-chain triglycerides, which can be absorbed in the absence of bile salts.[37] Foods rich in polyunsaturated fatty acids, such as vegetable oils, are more easily absorbed than meats, which have more saturated fat. Finally, completely hydrolyzed dietary supplements are rapidly absorbed. To slow bowel transit, diphenoxylate-atropine, loperamide, or deodorized tincture of opium can be used effectively. An alternative method is to have the patient drink a small amount of safflower oil just before a meal. The lipid quickly goes to the ileum (if present), the colon, or both[38] and elaborates peptide YY (PYY),[39] which is the putative ileal break, slowing gastric emptying. Having patients try different diets will often enable them to ingest food orally rather than receive total parenteral nutrition (TPN), which is less desired.

Radiation Enteritis

Injury of the intestine is an all too common result of delivery of ionizing radiation as oncologic treatment. Injury to the small bowel is more common if the patient has had previous abdominal surgery, which may restrict the movement of the small bowel. The terminal ileum may become involved during pelvic irradiation. To prevent radiation injury, it may be advisable to irradiate patients on a turntable so that more than one part of the intestine receives the maximal amount of radiation.

WHIPPLE DISEASE

Whipple disease is a rare multisystem disease caused by the bacterium *Tropheryma whippelii*.[40] The pathogenesis is obscure. The bacillus may be widespread throughout the body, but the sites of invasion show little sign of inflammation, suggesting that an autoimmune deficiency may create a predisposition to the disease.[41] Accurate diagnosis is imperative because mortality approaches 100% without antibiotic treatment.

Diagnosis

Clinical manifestations Classically, Whipple disease begins in a middle-aged male with a nondeforming arthritis that usually starts years before the onset of the intestinal symptoms. Arthralgias, diarrhea, abdominal pain, and weight loss are the cardinal manifestations of Whipple disease. Other complaints include fever, abdominal distention, lymphadenopathy, hyperpigmentation of the skin, and steatorrhea.[42] Many patients express the HLA-B27 isotype. Occasionally, intestinal symptoms are absent, even in some patients with central nervous system involvement.[43] In a well-documented but unusual case, intestinal involvement was not identified, even after extensive biopsies in two laboratories, despite the fact that the patient otherwise had typical symptoms of the disease.[44] Cardiac and pulmonary involvement may also be found.[45]

Laboratory tests The recognition of Whipple disease in patients without intestinal symptoms or involvement by the disease has been increasing since the advent of polymerase chain reaction techniques that identify the unique 16S ribosomal RNA of *T. whippelii*.[46]

The diagnosis rests on identifying the classic periodic acid–Schiff (PAS)–positive macrophages, which contain sickle-form particles.[42] By far the most common site of biopsy that yields positive results is the intestine. The histologic lesion shows distended villi (clubbed villi) filled with the foamy, PAS-positive macrophages and lymphatic dilatation. A flat villous surface can be seen in extreme cases. These findings need to be differentiated in the appropriate clinical setting from those of *M. avium* complex disease, in which PAS-positive macrophages are also found. A stain for acid-fast bacilli should differentiate between them. Central nervous system involvement, occasionally associated with typical macrophages in the cerebrospinal fluid as substantiated by the more sensitive PCR technique, may be present in the absence of neurologic symptoms,[47] and a negative result on PCR may predict a low likelihood of clinical relapse.[48] Occasionally, a brain biopsy is required, which can be guided by magnetic resonance imaging.

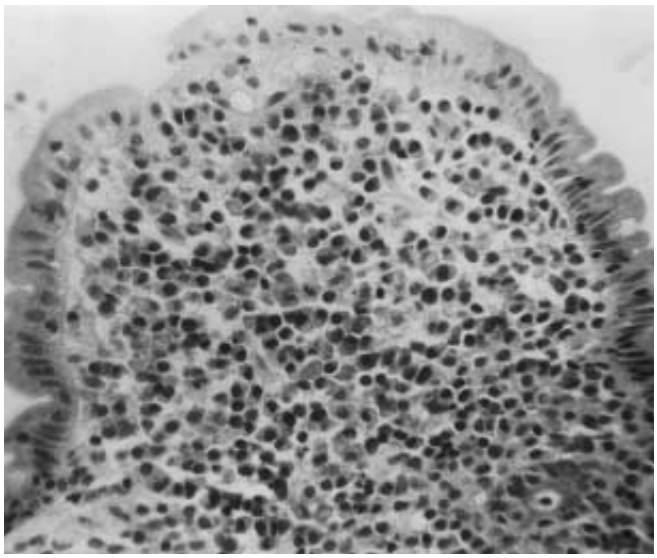


Figure 6 **Small intestinal biopsy specimen from a patient with primary intestinal lymphoma shows a single broadened villus (magnification: 400×). The epithelium is composed of normal columnar cells, but the lamina propria is packed with plasma cells and other mononuclear cells. Surgical biopsies in this patient revealed evidence of generalized subepithelial histiocytic lymphoma.**

Treatment Because Whipple disease is so uncommon, a well-defined treatment plan is difficult to establish. The originally proposed treatment was penicillin (250 mg q.i.d.) and streptomycin (1 g I.M.) for 2 weeks, followed by tetracycline (1 g) for 1 year. Typically today, trimethoprim-sulfamethoxazole (one double-strength tablet b.i.d.) is given for 1 year. All antimicrobial agents are used in customary doses.

Although the intestinal and systemic symptoms respond readily to either treatment, the major concern is treatment of CNS manifestations. Usually, in those patients who do not have CNS involvement initially, CNS symptoms appear a year or more after treatment of the systemic and intestinal symptoms, especially if an antibiotic is administered that does not cross the blood-brain barrier. A progressive dementia may be seen, but the pathognomonic signs of CNS disease, when present, are oculomasticatory myorhythmia and oculofacial-skeletal myorhythmia.[49] Antibiotics that cross the blood-brain barrier are therefore required. Interestingly, the short period of penicillin-streptomycin administration is enough to block CNS symptoms, whereas even long-

term trimethoprim-sulfamethoxazole therapy occasionally may not prevent CNS manifestations of Whipple disease.[50] Tetracycline alone does not eradicate CNS disease and should not be given by itself, even though it is effective in treating the intestinal and systemic symptoms. An important aspect to keep in mind is that in 50% of patients, the CSF may contain Whipple disease macrophages or PCR-positive material even in the absence of CNS symptoms.[46] Once CNS involvement occurs, treatment is usually not helpful, although with treatment, some improvement may be noted and the disease may not progress.

IMMUNOPROLIFERATIVE SMALL INTESTINAL DISEASE

Immunoproliferative small intestinal disease (IPSID), previously known as primary intestinal lymphoma, is a condition in which the lamina propria of the small bowel is intensely infiltrated with lymphocytes and the overlying enterocytes are normal morphologically [*see Figure 6*]. It is a B cell disorder involving the mucosa-associated lymphoid tissue (MALT). The disease is rare in developed nations and more common in underprivileged populations, primarily in persons in the second and third decades of life, with a male predominance. In a series of Chinese patients, six of 45 patients with intestinal lymphoma had IPSID.[51] These patients presented with severe malabsorption. Among the lymphoma patients without IPSID, 65% had abdominal pain, weight loss, abdominal masses, obstruction, and perforation. IPSID is associated with heavy chains (from IgA), with paraprotein present in the serum, urine, or jejunal fluid. Duodenography shows thickened folds and many nodular elevations without ulceration. The diagnosis may be made by small bowel biopsy in 85% of cases.[52] Early in its course, the disease appears to be treatable with antibiotics. If allowed to progress, however, it may develop into more aggressive forms of lymphoma.[53]

INTESTINAL LYMPHANGIECTASIA

Intestinal lymphangiectasia is often a congenital condition in which deformed lymphatics impair the transport of chylomicrons from the enterocytes to the mesenteric lymph duct. A similar pathophysiologic picture occurs in certain cases of IPSID, granulomatous enteritis, tuberculous enteritis, or Whipple disease in which normal lymphatic drainage is blocked.

Diagnosis

The blockage of lymphatic drainage may result in chylous ascites, chyluria, or chylometrorrhea.[54] Protein-losing enteropathy and lymphopenia are prominent features. Modest steatorrhea is also present, with fat excretion commonly reaching 20 g/day. In the congenital form of the disease, lymphedema of the legs or of one leg and one arm is seen. With endoscopic examination, white villi, white nodules, and submucosal elevations may be noted.[55] The white appearance of the mucosa is undoubtedly caused by retained chylomicron triacylglycerol. Double-contrast barium x-ray examination shows smooth nodular protrusions and thick mucosal folds without ulceration.[56] On histologic examination, dilated lymphatics with club-shaped villi are seen, sometimes in asymptomatic patients, in whom the outcome is benign.

Treatment

Treatment is directed toward any identified causative process. In patients with the congenital condition, in whom improvement of the lymphatics is not expected, a low-fat diet supplemented with medium-chain triglycerides is usually helpful. Surgery can be used to remove isolated areas of lymphatic dysfunction if these areas can be identified or to anastomose a lymph duct to the venous system. Sometimes a peritoneovenous (LeVeen) shunt is helpful.

ABETALIPOPROTEINEMIA

In the rare congenital condition of abetalipoproteinemia, postprandial chylomicronemia does not develop in patients, because they are unable to adequately couple apolipoprotein B to the developing chylomicron. Because lipid and lipid-soluble vitamins are transported from the intestine in chylomicrons, the consequent reduction in lipid and lipid-soluble vitamin absorption results in symptomatic steatorrhea, neurologic abnormalities, a variant of retinitis pigmentosa, and spiculated red cells. In contrast to earlier theories about the etiology of abetalipoproteinemia, patients with this disease have normally transcribed apolipoprotein B messenger RNA (mRNA) from which the protein is adequately translated. Nevertheless, apolipoprotein B is not secreted from the intestinal cell. The defect in this condition is in various mutations in the gene that encodes the M component of the microsomal triglyceride transport protein.[57] This chaperonelike protein complex consisting of a 97 kd M component and a 55 kd P component (protein disulfide isomerase) helps translocate the apolipoprotein across the membrane of the endoplasmic reticulum.[58] Without this step, the apolipoprotein is degraded by cytosolic and microsomal peptidases. The result of this defect is that both the intestine and the liver are unable to produce and secrete their triacylglycerol-rich lipoproteins, chylomicrons, and very low density lipoproteins. Because chylomicrons cannot transport the fat out of the enterocyte, it is presumed, but not proved, that the 80% of the lipid that is absorbed is absorbed via the portal vein.[59]

Diagnosis

In addition to having intestinal symptoms, patients with abetalipoproteinemia have severe neurologic problems. These neurologic problems may be caused in part by essential fatty acid deficiency and in part by either the impaired delivery of lipid to nerves or an interference with the local synthesis of lipids. The result is a demyelinating condition that causes a sensory ataxia because of the loss of position and vibratory sensations. The symptoms are similar to but less severe than those of Friedreich ataxia.[60] Patients may have muscle weakness and athetoid movements. Patients also experience retinitis pigmentosa, usually with mild loss of visual acuity but preservation of central vision. In addition to having neurologic abnormalities, patients have acanthocytes in their blood. Acanthocytes are spiculated red cells that have a near-normal life span but that demonstrate an increased susceptibility to mechanical trauma on in vitro testing.

These patients have low plasma triacylglycerol and cholesterol levels. On histologic examination, the enterocytes are seen to be laden with fat. Despite this phenotype, the amount of steatorrhea is modest (about 20 g/day).

Abetalipoproteinemia is usually discovered in childhood because patients with the disease fail to thrive and have steatorrhea. In adults, the disease can be recognized by the combination of neurologic and ophthalmologic findings, the red cell morphology, the very low levels of plasma lipids, and the modest steatorrhea. On small bowel biopsy, the enterocytes are seen to be stuffed with lipid even after an overnight fast, indicating that the absorbed lipid cannot exit the enterocytes.[61]

Treatment

Treatment should include vitamin E as well as the other fat-soluble vitamins and medium-chain triglycerides to reduce the steatorrhea, if required.

EOSINOPHILIC GASTROENTERITIS

Eosinophilic gastroenteritis is a rare disease that is characterized by the presence of eosinophilic infiltration of one or more portions of the gastrointestinal tract, anywhere from the esophagus to the colon, in conjunction with gastrointestinal symptoms. No identifiable cause of the eosinophilic infiltrate, such as parasitic infestation, is present. Many patients have an underlying allergic diathesis (e.g., hay fever, asthma, atopic dermatitis, or drug allergies).

It is not known why eosinophils congregate in the GI tract in eosinophilic gastroenteritis, but evidence suggests that eosinophils, once activated, can produce cytokines that self-perpetuate the accumulation of additional eosinophils. These cytokines are interleukin-3 (IL-3), IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF), which have been identified in eosinophils of patients but not in control subjects with irritable bowel syndrome. Local production of these cytokines is suggested by the finding that serum levels of IL-5 are normal in patients with eosinophilic gastroenteritis, in contrast to patients with the hypereosinophilic syndrome, who have increased levels of IL-5 in their blood.[62]

Diagnosis

Although eosinophils are a normal constituent of the GI tract, in eosinophilic gastroenteritis the eosinophils appear more numerous than normal and are more invasive. For example, eosinophilic invasion of the crypts in the small intestine is a hallmark of eosinophilic gastroenteritis. A peripheral eosinophilia is often seen but is not always present.

Eosinophilic gastroenteritis can be divided into two basic forms: a tumorous mass of eosinophils producing a granulomatous-type lesion and a more diffusely infiltrative form. In the former case, the lesions are most often in the distal stomach, which may produce obstructive symptoms, or the masses may be found in the more proximal stomach, small bowel, or colon. When lesions are in the small bowel or colon, the condition needs to be differentiated from a lymphoma or Crohn disease.[63] In the case of diffuse disease involving the small bowel, the infiltration can be mucosal, with symptoms of protein-losing enteropathy or malabsorption. If the infiltration is primarily in the muscle layers of the intestine, obstructive symptoms are common. Finally, the disease may be found in the subserosal area of the intestine, with resultant eosinophilic ascites.[64]

Treatment

Most patients respond to conservative measures and steroids. Surgery should be avoided unless it is needed to relieve persistent or small bowel obstruction.

Prednisone, 40 mg orally every morning and tapered slowly over 2 weeks, is the most effective therapy for patients with obstructive symptoms and ascites. If high-dose steroids are needed to maintain remission, azathioprine can be added for its steroid-sparing effect.

Diet elimination therapy may be beneficial in patients with mucosal layer involvement.

CROHN DISEASE

The prevalence of Crohn disease in North America ranges from 26.0 to 198.5 cases per 100,000 persons, with a maximal incidence in people of Jewish decent.[65,66]

Pathogenesis

Crohn disease, a stenosing, fistulizing disease of the intestine, may impair intestinal absorption by at least two mechanisms, ileal dysfunction and the stasis syndrome [*see* Stasis (Bacterial Overgrowth) Syndrome, *below*]. In the case of either ileal resection or severe ileal involvement with Crohn disease, the ileum cannot absorb bile salts normally. In that event, postprandial bile salt deficiency occurs in the upper intestine; this condition may become more severe the later in the day a meal is eaten.[67] Postprandial bile salt deficiency occurs despite the liver's response to bile acid loss from the enterohepatic circulation, which is to increase bile acid synthesis. The increase in bile salt synthesis is not adequate, because each time the gallbladder contracts in response to a meal, most of the bile salt pool is lost to the colon[68] if significant amounts of the ileum have been resected. Thus, the liver does not have time to generate enough replacement bile salts for the complete absorption of the meal just eaten or the next one. The colonic perfusion of bile acids may result in diarrhea. This condition has been termed choleretic enteropathy and may occur when more than 30 cm of the terminal ileum is resected. The excess fluid in the colon is caused by cAMP-driven Cl^- secretion, specifically by the dihydroxylated bile acids, chenodeoxycholate and deoxycholate, not trihydroxylated cholic acid.[69]

Diagnosis

Crohn disease with malabsorption is suggested by the history (e.g., a prior ileal resection); by physical examination in which thickened, tender bowel may be felt; or by imaging studies (e.g., a small bowel barium study showing the absence of the terminal ileum and the presence of strictures, stellate ulcers, or cobblestoning of the mucosa). Ileal dysfunction secondary to ileal resection is suggested by the presence of diarrhea without steatorrhrea, but it may difficult to differentiate this presentation from the stasis syndrome. Lipid malabsorption would be suggested by an increase in fat in the stool.

The loss of bile acids to the colon and thus to the enterohepatic circulation can be associated with no or minimal steatorrhea.[70]

With more extensive (100 cm or greater) ileal resection, however, the diarrhea is caused not only by bile acids but also by malabsorbed fatty acids (steatorrhea).[71] Thus, the diarrhea associated with Crohn disease may be caused not by active disease but by the results of ileal resection. This scenario is suggested by the finding that the diarrhea occurs when the patient first eats after surgery, a time when disease activity may be low secondary to active disease resection, or by the fact that the patient had no or minimal diarrhea before surgery, with diarrhea becoming more prominent afterward.

Because of the stenosis present in some patients with Crohn disease, the stasis syndrome can develop [*see* Stasis (Bacterial Overgrowth) Syndrome, *below*].

Treatment

When the diarrhea is caused by bile acid loss, the treatment is cholestyramine (4 g a.c. and h.s.).[71] This resin preferentially binds the dihydroxylated bile acids, reducing their aqueous concentration and reducing their proportion in the total bile acid pool.

Both effects are beneficial. In the case of larger ileal resections in which steatorrhea is prominent, cholestyramine may actually provoke more diarrhea and malabsorption because it reduces the aqueous bile acid concentration in the upper intestine when taken before meals. In this case, medium-chain fatty acids are used as a replacement for the long-chain fatty acids. The results of this strategy are often not as good as desired. Vitamin B_{12} absorption should also be evaluated in all patients with ileal resection, and if absorption is found to be abnormal, vitamin B_{12} should be given parenterally.

Some patients with severe Crohn disease undergo extensive intestinal resection, resulting effectively in short bowel syndrome. Similarly, patients who have numerous enteroenteric fistulas also have symptoms of short bowel syndrome because the fistulas cause the chyme to bypass large sections of the small intestine. Both types of patients should be treated as if they had short bowel syndrome [*see* Short Bowel Syndrome, *above*].

STASIS (BACTERIAL OVERGROWTH) SYNDROME

The stasis (bacterial overgrowth) syndrome occurs when intestinal stasis leads to the opportunity for bacteria to proliferate locally. This condition has a multiplicity of causes. The most prominent causes are diabetes, scleroderma, intestinal diverticulosis, afferent loop of a gastrojejunostomy, and intestinal obstruction caused by strictures, adhesions, or cancer. These disorders may be present years before the development of symptoms. Symptoms may appear in an otherwise stable patient because of the administration of a proton pump inhibitor that reduces gastric acid production, allowing gastric and small bowel overgrowth, or the administration of an opiate that further reduces intestinal motility.

Intestinal dysfunction in the stasis syndrome is probably caused by bacterial glycosidases that hydrolyze the carbohydrate moieties that form the extensive glycosylation of the apical brush-border proteins.[72] Although bile acid deconjugation occurs in the stasis syndrome, which may theoretically lead to impaired solubilization of the products of triglyceride hydrolysis, studies have shown that in fact the fatty acid concentration in the aqueous phase of postprandial intestinal content is normal.[73] Electron micrographs, however, show that there is damage to the enterocytes, in that absorbed lipid collects in the endoplasmic reticulum and does not progress normally to the Golgi apparatus.[73]

Diagnosis

Clinical manifestations Symptoms of the stasis syndrome are similar to those of other malabsorptive states and include steatorrhea and anemia. The patient may have vitamin B_{12} deficiency, which has several causes, including binding of the vitamin to bacteria[74,75] and bacterial metabolism of the vitamin to metabolically ineffective metabolites. Folic acid levels are usually high secondary to bacterial production of folate.[76] Serum albumin levels may be low secondary to protein-losing enteropathy and remain low for months after adequate treatment. The diagnosis is usually made in a patient with malabsorption in the appropriate clinical setting. Intestinal (usually jejunal) diverticulosis is most often unsuspected until a small bowel x-ray is performed.

Laboratory tests Establishing the diagnosis of the stasis syndrome is not simple. The most accurate way is to pass an aspiration tube into the intestine. The fluid must be quantitatively cultured both aerobically and anaerobically. In most cases, more than 10^5 anaerobes will be found. Alternatively, the noninvasive hydrogen breath test may be used. A high resting hydrogen level or a quick increase in the breath hydrogen in response to a fermentable substrate, such as glucose or lactulose, can be used. Another breath test is the 1 g (^{14}C)-D-xylose test, in which the breath $^{14}CO_2$ is measured.

Treatment

The first choice of treatment for the stasis syndrome is surgical correction of defects, such as an afferent loop that is harboring bacteria, or a jejunocolic fistula. If the surgical option is not available, then recurrent dosing of an antibiotic is required. Tetracycline, at a dosage of 1 to 2 g/day for a 7- to 10-day course, gives good results, or another antibiotic that is active against anaerobic bacteria may be used (e.g., trimethoprim-sulfamethoxazole, one double-strength tablet b.i.d.). The patient will need to be re-treated if clinical symptoms reappear, or the patient can receive treatment for 1 week every month.

AMYLOIDOSIS

The intestine is often involved in patients with systemic amyloidosis, especially if they have polyneuropathy. In patients older than 85 years, 36% have intestinal involvement with amyloidosis,[77] although most are asymptomatic. Endoscopically, mucosal erosion, friability, or polypoid protrusions can be seen.[78] The diagnosis is made by either full-thickness or peroral intestinal biopsies. If a peroral biopsy is performed, it must be deep enough to have arteries visible, so that amyloid, if present, can be demonstrated. Congo red–stained arterioles that become apple green under polarized light confirm the diagnosis. Small bowel follow-through x-rays may show swollen intestinal plicae, possibly with separated loops of bowel. If steatorrhea is present, it may be the result either of bacterial overgrowth caused by intestinal dysmotility or of impaired bile acid absorption.[79] No specific effective therapy is available. If bacterial overgrowth is present, then appropriate antibiotics should be given.

SYSTEMIC MASTOCYTOSIS

In this rare condition, the skin (99% of cases), bones (9%), liver (12%), spleen (11%), lymph nodes, and GI tract are involved with proliferating mast cells. Diarrhea or abdominal pain or both (23% of cases), peptic ulceration (4%), and itching and flushing (36%) may be seen. Headache, fatigue, and malaise are seen in 12% of cases. There may also be cognitive dysfunction. Eosinophilia is seen in 12% to 50% of cases.[69] Many of these manifestations of the disease are secondary to histamine, which is released from the mast cells. Histamine release may be precipitated by alcohol, aspirin, narcotics, and nonsteroidal anti-inflammatory drugs, causing episodic disturbances of flushing, diarrhea, abdominal pain, and hypotension that may progress to syncope.[80]

Excess histamine is excreted into the urine in excess in approximately 75% of patients, making the urinary excretion test useful for diagnostic purposes.[80] The urinary excretion of a metabolite of prostaglandin D_2 from mast cells may be an even better test.[81] X-ray studies of the small intestine may show thickened folds or nodulation. These findings are not diagnostic but may point to a diseased small bowel.

Histamine-mediated overproduction of gastric acid may lead to peptic ulceration. In that event, H_2 receptor blockers or proton pump inhibitors are effective in controlling symptoms. In the skin, urticaria pigmentosa may be effectively treated with H_1 re-

ceptor antagonists such as diphenhydramine (25 μg every 6 to 8 hours). If diarrhea persists, cromolyn sodium may be given at a dosage of 100 mg orally four times a day.

CHRONIC PANCREATITIS WITH EXOCRINE INSUFFICIENCY

Most chronic pancreatitis is caused by alcoholism. In rare cases, the disease is inherited. Patients experience weight loss resulting from malabsorption of food. Malabsorption caused by pancreatitis is discussed elsewhere [*see 4:5 Diseases of the Pancreas*].

POSTGASTRECTOMY STEATORRHEA

One of the consequences of gastric surgery is steatorrhea, primarily in patients who have the Billroth II gastric resection with a gastrojejunostomy. In this operation, the antrum and a variable portion of the body of the stomach are resected, the stomach is sutured closed, and a gastrojejunostomy is created. Thus, food bypasses the duodenum and most proximal jejunum, which are the sites of maximal cholecystokinin and secretin concentrations and the active sites for folate, calcium, and iron absorption. Approximately one half of patients who have undergone the Billroth II procedure have steatorrhea of 10 to 15 g of fat/day. This condition is thought to result from food entering the jejunum without the hormone-sensitive sites in the duodenum receiving the appropriate signals for hormone release. Thus, there is poor gallbladder contraction and reduced release of pancreatic digestive enzymes to the intestine resulting in poor admixing of the chyme with pancreatic enzymes and bile acids. The afferent loop, which drains the duodenum and proximal jejunum, may become blocked or atonic and harbor bacteria. The stasis syndrome may occur if enough bacteria are present. Because of their small stomachs, the affected patients cannot eat as much as they previously could. This decrease in food consumption, in combination with steatorrhea, causes many patients who undergo the Billroth II procedure to stabilize at a lower weight than they were before surgery. Osteopenia and iron deficiency anemia are also found. Constant small amounts of blood loss from the gastric ostomy site combined with impaired iron absorption contribute to the iron-deficient state, which is the most common form of anemia. Folate deficiency secondary to the inability to generate absorbable monoglutamyl folate from nonabsorbable heptaglutamyl folate (the common form of folate found in the diet) is also found.[82] Least commonly seen is vitamin B_{12} deficiency caused by hypochlorhydria and resection of intrinsic factor–containing gastric parietal cells. Treatment of the steatorrhea is usually not necessary, because it is not clinically significant. Iron, calcium, or vitamin B_{12} and folic acid must be replaced as indicated. If the patient has early satiety, multiple small meals may be efficacious.

Symptoms of GSE may develop in patients after gastric surgery.[83] It is likely that these patients had clinically silent GSE before the operation. The operation itself causes modest steatorrhea (10 to 15 g fat/day) in 50% of cases, even in patients whose intestine is otherwise normal. In the previously compensated GSE patient, however, surgery is enough to cause clinical symptomatology. Therefore, if postgastrectomy patients exhibit excessive steatorrhea, an evaluation for GSE is warranted. Inflammatory bowel disease that develops in patients after gastrectomy may likewise be an indication of the presence of previously silent GSE.[84]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Moreau H, Gargouri Y, Bernadal A, et al: Etude biochemique et physiologique des lipases préduodéales d'origines animale et humaine. Revue Française des Corps Gras 35:169, 1988

2. Roulet M, Weber A, Roy C: Perspectives in Cystic Fibrosis. Canada Cystic Fibrosis Foundation, Toronto, 1980, p 172

3. Levitt MD, Duane WC: Floating stools: flatus versus fat. N Engl J Med 286:973, 1972

4. Fine KD, Ogunji F: A new method of quantitative fecal fat microscopy and its correlation with chemically measured fecal fat output. Am J Clin Pathol 113:528, 2000

5. DiMagno EP, Go VLW, Summerskill WHJ: Relation between pancreatic enzyme outputs and malabsorption in severe pancreatic insufficiency. N Engl J Med 288:813, 1973

6. Chowdhury RS, Forsmark CE: Pancreatic function testing. Aliment Pharmacol Ther 17:733, 2003

7. Kato H, Nakao A, Kishimoto W, et al: ^{13}C-labeled trioctanoin breath test for exocrine pancreatic function test in patients after pancreatoduodenectomy. Am J Gastroenterol 88:64, 1993

8. Lang C: Value of serum PABA as a pancreatic function test. Gut 25:508, 1984

9. Bando N, Ogawa T, Tsuji H: Enzymatic method for selective determination of 4 aminobenzoic acid in urine. Clin Chem 36:1937, 1990

10. Jacobson DG, Curlington C, Connery K, et al: Trypsin-like immunoreactivity as a test for pancreatic insufficiency. N Engl J Med 310:1307, 1984

11. Oelkers P, Kirby LC, Heubi JE, et al: Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene. J Clin Invest 99:1880, 1997

12. Shindo K, Yamazaki R, Koide K, et al: Alteration of bile acid metabolism by cimetidine in healthy humans. J Investig Med 44:462, 1996

13. Nyhlin H, Merrick MV, Eastwood MA, et al: Evaluation of ileal function using 23-selena-25-homotaurocholate, a γ-labeled conjugated bile acid. Gastroenterology 84:63, 1983

14. Maki M, Collin P: Coeliac disease. Lancet 349:1755, 1997

15. Fasano A, Berti I, Gerarduzzi T, et al: Prevalenc of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. Arch Intern Med 163:286, 2003

16. Shan L, Molberg O, Parrot I, et al: Structural basis for gluten intolerance in celiac sprue. Science 297:2218, 2002

17. Hofmann AF: The function of bile salts in fat absorption. Biochem J 89:57, 1963

18. Mansbach CM II, Cohen RS, Leff PB: Isolation and properties of the mixed micelles present in intestinal content during fat digestion in man. J Clin Invest 56:781, 1975

19. Farrell RJ, Kelly CP: Celia sprue. N Engl J Med 346:180, 2002

20. Vasquez H, Mazure R, Gonzalez D, et al: Risk of fractures in celiac disease patients: a cross-sectional, case-control study. Am J Gastroeneterol 95:183, 2000

21. Dieterich W, Laag E, Schópper H, et al: Autoantibodies to tissue transglutaminase as predictors of coeliac disease. Gastroenterology 115:1317, 1998

22. Volta V, Molinaro N, deFranceschi L, et al: IgA antiendomysial antibodies on human umbilical cord tissue for celiac disease screening. Dig Dis Sci 40:1902, 1995

23. Holmes GKT, Prior P, Lane MR, et al: Malignancy in coeliac disease: effect of a gluten-free diet. Gut 30:333, 1989

24. Ryan BM, Kelleher D: Refractory celiac disease. Gastroenterology 119:243, 2000

25. Gawkrodger DJ, Vestey JP, O'Mahouny S: Dermatitis herpetiformis and established coeliac disease. Br J Dermatol 129:694, 1993

26. Rensch MJ, Merenich JA, Lieberman M, et al: Gluten-sensitive enteropathy in patients with insulin-dependent diabetes mellitus. Ann Intern Med 124:564, 1996

27. Holmes GK: Coeliac disease and type 1 diabetes mellitus, the case for screening. Diabet Med 18:169, 2001

28. Not T, Tommasini A, Tonini G, et al: Undiagnosed celiac disease and risk of autoimmune disorders in type 1 diabetes mellitus. Diabetologia 44:151, 2001

29. Ventura A, Magazzu G, Greco L: Duration of exposure to gluten and risk for autoimmune disorders in patients with celiac disease. SIGEP Study Group for Autoimmune Disorders in Celiac Disease. Gastroenterology 117:297, 1999

30. Klipstein FA, Falaiye JM: Tropical sprue in expatriates from the tropics living in the continental United States. Medicine (Baltimore) 48:475, 1969

31. Lindenbaum J, Gerson CD, Kent TH: Recovery of small intestinal structure and function after residence in the tropics: I. Studies in Peace Corps Volunteers. Ann Intern Med 74:218, 1971

32. Gerson CD, Kent TH, Saha JR, et al: Recovery of small intestinal structure and function after residence in the tropics: II. Studies in Indians and Pakistanis living in New York City. Ann Intern Med 75:41, 1971

33. Haghighi P, Wolf PL: Tropical sprue and subclinical enteropathy: a vision for the nineties. Crit Rev Clin Lab Sci 34:313, 1997

34. Robert ME, Ament ME, Weinstein WM: The histologic spectrum and clinical outcome of refractory and unclassified sprue. Am J Surg Pathol 24:676, 2000

35. McCashland TM, Donovan JP, Strobach RS, et al: Collagenous enterocolitis: a manifestation of gluten-sensitive enteropathy. J Clin Gastroenterol 15:45, 1992

36. Hermaszewski RA, Webster AD: Primary hypogammaglobulinaemia: a survey of clinical manifestations and complications. Q J Med 86:31, 1993

37. Ladefoged K, Hessov I, Jarnum S: Nutrition in short-bowel syndrome. Scand J Gastroenterol 216(suppl):122, 1996

38. Lin HC, Zhao X-T, Wang L: Fat absorption is not complete by midgut but is dependent on load of fat. Am J Physiol 271(1 pt 1):G62, 1996

39. Lin HC, Zhao X-T, Wong H: Fat-induced ileal brake in the dog depends on peptide

YY. Gastroenterology 110:1491, 1996

40. Redman DA, Schmidt TM, McDermott RP, et al: Identification of the uncultured bacillus of Whipple's disease. N Engl J Med 327:393, 1992

41. Bentley SD, Maiwald M, Murphy LD, et al: Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. Lancet 361:637, 2003

42. Durand DV, Lecomte C, Cathebras P, et al: Whipple disease: clinical review of 52 cases. The SNFMI Research Group on Whipple Disease. Medicine (Baltimore) 76:170, 1997

43. Dobbins WO III: HLA antigens in Whipple's disease. Arthritis Rheum 30:102, 1987

44. Mansbach CM II, Shelburne J, Stevens RD, et al: Lymph node bacilliform bodies morphologically resembling those of Whipple's disease in a patient without intestinal involvement. Ann Intern Med 89:64, 1978

45. Kelly CA, Egan M, Rawlinson J: Whipple's disease presenting with lung involvement. Thorax 51:343, 1996

46. Swartz MN: Whipple's disease—past, present and future. N Engl J Med 342:648, 2000

47. von Herbay A, Ditton H-J, Schumacher F, et al: Whipple's disease: staging and monitoring by cytology and polymerase chain reaction analysis of cerebrospinal fluid. Gastroenterology 113:434, 1997

48. Ramzan NN, Loftus E Jr, Burgart LJ, et al: Diagnosis and monitoring of Whipple disease by polymerase chain reaction. Ann Intern Med 126:520, 1997

49. Louis ED, Lynch T, Kaufmann P, et al: Diagnostic guidelines in central nervous system Whipple's disease. J Ann Neurol 40:561, 1996

50. Feurle GE, Marth T: An evaluation of antimicrobial treatment for Whipple's disease: tetracycline versus trimethoprim-sulfamethoxazole. Dig Dis Sci 39:1642, 1994

51. Shih LY, Liaw SJ, Dunn P, et al: Primary small-intestinal lymphomas in Taiwan: immunoproliferative small-intestinal disease and nonimmunoproliferative small-intestinal disease. J Clin Oncol 12:1375, 1994

52. Halphen M, Najjar T, Jaafoura H, et al: Diagnostic value of upper intestinal fiber endoscopy in primary small intestinal lymphoma: a prospective study by the Tunisian-French Intestinal Lymphoma Group. Cancer 58:2140, 1986

53. Khojasteh A, Haghighi P: Immunoproliferative small intestinal disease: portrait of a potentially preventable cancer from the Third World. Am J Med 89:483, 1990

54. Fox C, Lucani G: Disorders of the intestinal mesenteric lymphatic system. Lymphology 26:61, 1993

55. Aoyagi K, Iida M, Yao T, et al: Characteristic endoscopic features of intestinal lymphangiectasia: correlation with histological findings. Hepatogastroenterology 44:133, 1997

56. Aoyagi K, Iida M, Yao T, et al: Intestinal lymphangiectasia: value of double-contrast radiographic study. Clin Radiol 49:814, 1994

57. Sharp D, Blinderman L, Combs KA, et al: Cloning and gene defects in microsomal triglyceride transfer protein associated with abetalipoproteinaemia. Nature 365:65, 1993

58. Gordon DA, Jamil H, Gregg RE, et al: Inhibition of the microsomal triglyceride transfer protein blocks the step of apolipoprotein B lipoprotein assembly but not the addition of bulk core lipids in the second step. J Biol Chem 271:33047, 1996

59. Mansbach CM II, Dowell RF, Pritchett D: Portal transport of absorbed lipids in the rat. Am J Physiol 261:G530, 1991

60. Isselbacher KJ, Scheig R, Plotkin GR, et al: Congenital β-lipoprotein deficiency: an hereditary disorder involving a defect in the absorption and transport of lipids. Medicine (Baltimore) 43:347, 1964

61. Ways PO, Parmentier CM, Kayden HJ, et al: Studies on the absorptive defect for triglyceride in abetalipoproteinemia. J Clin Invest 46:35, 1967

62. Desreumaux P, Blogot F, Seguy D, et al: Interleukin 3, granulocyte-macrophage colony-stimulating factor, and interleukin 5 in eosinophilic gastroenteritis. Gastroenterology 110:768, 1996

63. Madhotra R: Eosinophilic gastroenteritis masquerading as ampullary adenoma. J Clin Gastroenterol 34:240, 2002

64. Klein NC, Hargrove RL, Sleisenger MH, et al: Eosinophilic gastroenteritis. Medicine (Baltimore) 49:299, 1970

65. Loftus EV Jr, Schoenfeld P, Sandborn WJ: The epidemiology and natural history of Crohn's disease in population-based patient cohorts from North America: a systematic review. Aliment Pharmacol Ther 16:51, 2002

66. Sugimura K, Taylor KD, Lin YC, et al: A novel NOD2/CARD15 haplotype conferring risk for Crohn disease in Ashkenazi Jews. Am J Hum Genet 72:509, 2003

67. Van Deest BW, Fordtran JS, Morawski SG, et al: Bile salt and micellar fat concentration in proximal small bowel contents of ileectomy patients. J Clin Invest 47:1314, 1968

68. Low-Beer TS, Wilkins RM, Lack L, et al: Effect of one meal on enterohepatic circulation of bile salts. Gastroenterology 67:490, 1974

69. Merhjian HS, Phillips SF, Hofmann AF: Colonic secretion of water and electrolytes induced by bile acids: perfusion studies in man. J Clin Invest 50:1569, 1971

70. Mansbach CM II, Newton DF, Stevens RD: Fat digestion in patients with bile acid malabsorption but minimal steatorrhea. Dig Dis Sci 25:353, 1980

71. Hofmann AF, Poley JR: Role of bile acid malabsorption in pathogenesis of diarrhea and steatorrhea in patients with ileal resection: I. Response to cholestyramine or replacement of dietary long chain triglyceride by medium chain triglyceride. Gastroenterology 62:918, 1972

72. Riepe S, Goldstein J, Alpers DH: Effect of secreted *Bacteroides* proteases on human intestinal brush border hydrolases. J Clin Invest 66:314, 1980

73. Ament ME, Shimoda SS, Saunders DR, et al: Pathogenesis of steatorrhea in three cases of small intestinal stasis syndrome. Gastroenterology 63:728, 1972

74. Gianella RA, Broitman SA, Zamcheck N: Vitamin B_{12} uptake by intestinal microorganisms: mechanisms and relevance to syndromes of bacterial overgrowth. J Clin Invest 50:1100, 1971

75. Gregg CR: Enteric bacterial flora and bacterial overgrowth syndrome. Semin Gastrointest Dis 13:2000, 2002

76. Hoffbrand AV, Tabaqchali S, Booth CC, et al: Small intestinal bacterial flora and folate status in gastrointestinal disease. Gut 12:27, 1971

77. Rocken C, Saeger W, Linke RP: Gastrointestinal amyloid deposits in old age: report of 110 consecutive autopsical patients and 98 retrospective bioptic specimens. Pathol Res Pract 190:641, 1994

78. Tada S, Iida M, Yao KK, et al: Endoscopic features in amyloidosis of the small intestine: clinical and morphologic differences between chemical types of amyloid protein. Gastrointest Endosc 40:45, 1994

79. Suhr O, Danielsson A, Steen L: Bile acid malabsorption caused by gastrointestinal motility dysfunction? An investigation of gastrointestinal disturbances in familial amyloidosis with polyneuropathy. Scand J Gastroenterol 27:201, 1992

80. Golkar L, Bernhard JD: Mastocytosis. Lancet 349:1379, 1997

81. Morrow JD, Guzzo C, Lazarus G, et al: Improved diagnosis of mastocytosis by measurements of the major urinary metabolite of prostaglandin D2. J Invest Dermatol 104:937, 1995

82. Rosenberg IH: Folate absorption and malabsorption. N Engl J Med 293:1303, 1975

83. Bai J, Moran C, Martinez C: Celiac sprue after surgery of the upper gastrointestinal tract: report of 10 patients with special attention to diagnosis, clinical behavior, and follow-up. J Clin Gastroenterol 13:521, 1991

84. Kitis G, Holmes GTK, Cooper BT: Association of coeliac disease and inflammatory bowel disease. Gut 21:636, 1980

Acknowledgments

Figure 1 Janet Betries.

Figure 2 Dana Burns-Pizer.

XII DIVERTICULOSIS, DIVERTICULITIS, AND APPENDICITIS

WILLIAM V. HARFORD, M.D., F.A.C.P.

Colonic Diverticular Disease

Colonic diverticula are herniations of colonic mucosa and submucosa that extend through the muscularis propria. They occur where perforating arteries traverse the circular muscle layer and form parallel rows between the mesenteric and antimesenteric taenia. Diverticulosis describes the presence of diverticula, whereas diverticulitis refers to the inflammation of diverticula. Diverticulosis is a common condition; of persons with known diverticulosis, about 10% to 20% will develop diverticulitis or diverticular bleeding.[1]

DIVERTICULOSIS

Epidemiology

There are no population-based studies of the prevalence of diverticulosis. About 1% of the United States population reported having diverticulosis in the 1983–1987 National Health Interview Survey (NHIS). Women were two to three times more likely than men to report having diverticulosis, and whites were more likely than African Americans. The prevalence of self-reported diverticulosis increased with age. It was 0.1% at 45 years of age or younger and 4.4% at 75 years of age or older. Unrecognized diverticulosis is more common than known diverticulosis. It is estimated that 10% to 20% of persons older than 50 years have diverticulosis.[2] In Western countries, diverticula occur predominantly in the left colon, particularly the sigmoid colon, which is involved in 95% of cases. In the Orient, including Japan, diverticula occur predominantly in the right colon.[3,4]

About 85% of persons with self-reported diverticulosis in the NHIS were asymptomatic or reported no limitations resulting from diverticulosis. Patients who are asymptomatic at the time of diagnosis are unlikely to develop diverticulitis. In the First National Health and Nutrition Examination Survey (NHANES I) Epidemiologic Follow-up Study, a cohort of physicians with asymptomatic diverticulosis were followed for a 10-year period. The probability of hospitalization for diverticular disease was less than 1% for physicians who were 25 to 44 years of age at the beginning of the follow-up period and was about 5% for those who were 65 to 74 years of age.[2] In English and Finnish populations, the risk of acute diverticulitis was about four per 100,000 population per year.[5,6] The risk of diverticulitis increases with age and with the use of nonsteroidal anti-inflammatory drugs (NSAIDs), steroids, and opioids.[1,5,7]

Pathogenesis

Reduced colonic diameter and reduced colonic wall compliance are felt to predispose persons to diverticulosis. A reduced colonic diameter causes the formation of closed segments during colonic contractions, thereby increasing intraluminal pressure. Diverticulosis is common in countries where a low-fiber diet is consumed, because a low-fiber diet leads to reduced stool volume and colonic diameter, particularly in the sigmoid colon. A high-fiber diet reduces the risk of diverticular disease.[8,9] Patients with diverticulosis have an age-related increase in elastin deposition and collagen cross-linking.[10] Increased proline absorption from Western diets may be a factor contributing to increased elastin deposition.[11] Changes in elastin and collagen lead to thickening and shortening of the taenia and circular muscle layers (myochosis) in many patients with diverticulosis, increasing the possibility of segmentation. These elastin and collagen changes also result in reduced compliance of the colonic wall, so that for any colonic diameter, intraluminal pressure is higher than it is in patients with normal compliance.[12] Increased intraluminal pressure leads to herniation of mucosa through the defects in the muscularis of the colon associated with perforating arteries.

Only mucosa and submucosa separate the lumen of diverticula from the colonic serosa. Diverticulitis may result from abrasion of the mucosa by inspissated stool. Changes in bacterial colonic microflora have been reported in patients with diverticulosis. It has been proposed that these changes may lead to low-grade chronic inflammation, predisposing to the development of diverticulitis.[13,14] Chronic intermittent use of oral rifaximin, a poorly absorbed antibiotic, and mesalazine, an anti-inflammatory agent, appears to reduce the risk of diverticulitis.[15,16]

Diverticula form where medium-sized perforating arteries penetrate the muscularis propria to enter the submucosa. Pathologic examination has been reported to reveal evidence of chronic injury to the internal elastic lamina and media of these arteries. This injury can cause arterial rupture into the lumen of the colon. Diverticular bleeding is rarely associated with acute diverticulitis.

DIVERTICULITIS

Diagnosis

Diverticula are most often discovered incidentally during investigation of another condition. Diverticulitis varies in presentation and severity. The diagnosis of acute diverticulitis is often made on the basis of the history and physical examination, which includes abdominal, rectal, and pelvic examinations; imaging studies are used to confirm the diagnosis. Computed tomography has become the optimal method of investigation for patients suspected of having diverticulitis. The modified Hinchey classification, which takes into account both clinical and CT findings, is useful for prognosis and management [*see Table 1*].[17]

Clinical presentation Patients with mild diverticulitis (Hinchey stage 0 or 1a) have limited inflammation or phlegmon in the area of the involved diverticulum. They typically present with left-sided lower abdominal pain and localized tenderness, low-grade fever, anorexia, and nausea without vomiting. They may have mild leukocytosis. Patients with mild diverticulitis can often be managed without hospitalization.[4] Patients with more severe diverticulitis usually must be hospitalized. They often have a diverticular abscess (stage 1b or 2), which is usually contained in the pericolic fat, mesentery, or pelvis but may extend beyond the pelvis. Patients with an abscess (or large phlegmon) commonly have systemic toxicity, high fever, severe localized abdominal tenderness, and leukocytosis. The phlegmon or abscess may be palpable. Rupture of a diverticular abscess results

in purulent peritonitis (stage 3), which usually leads to diffuse abdominal tenderness. Free perforation of a diverticulum with fecal soiling of the abdominal cavity leads to feculent peritonitis (stage 4). Feculent peritonitis causes severe acute generalized peritonitis and sepsis.

Colonic inflammation associated with diverticulitis may cause either diarrhea or constipation. Acute diverticulitis may lead to colonic or small bowel obstruction. Repeated episodes of diverticulitis with fibrosis may cause colonic stricture.

Diverticulitis may cause fistula formation, most commonly from the sigmoid colon to the bladder. Inflammation adjacent to the bladder may lead to dysuria even if no fistula is present. Overt lower gastrointestinal bleeding is rarely associated with acute diverticulitis. Other causes of bleeding (e.g., angiodysplasia, a neoplasm, or inflammatory bowel disease) must be excluded in patients with diverticulitis who present with overt bleeding or who have positive fecal occult blood tests.

In a large retrospective study of patients who required hospitalization for acute diverticulitis, 72% of patients had no abscess (Hinchey stage 0 or 1a); 19%, an abscess (Hinchey stage 1b or 2); 5%, prurulent peritonitis (Hinchey stage 3); 1%, feculent peritonitis (Hinchey stage 4); 1%, obstruction; and 2%, fistula. Overall, surgery was required in 26% of patients.[17] Comparable distribution of stages has been reported in other studies.[18-20]

Diverticulitis in areas other than the sigmoid colon is uncommon in Western countries. In such cases, clinical presentation may be atypical and confusing. Cases of right-sided colonic and cecal diverticulitis are often clinically indistinguishable from appendicitis.[21-23]

Diverticulitis may lead to episodic abdominal pain. In a study of patients previously hospitalized for acute diverticulitis, 70% subsequently experienced new, recurrent episodes of abdominal pain, usually lasting less than 4 hours.[24] After the first episode of diverticulitis, 30% to 50% of patients have subsequent episodes. About 20% to 25% of patients have subsequent episodes of complicated diverticulitis within the first several years of follow-up.[18,19] Patients who initially have a large diverticular abscess have an increased risk of recurrence of diverticulitis, even if the abscess is treated by antibiotics and drainage by interventional radiology.[17]

Table 1 Modified Hinchey Classification of Acute Diverticulitis[17]

Stage	*Characteristic Symptoms*
0	Mild clinical diverticulitis (left lower quadrant abdominal pain, low-grade fever, leukocytosis, no imaging information)
1a	Confined pericolic inflammation, no abscess
1b	Confined pericolic abscess (abscess or phlegmon may be palpable; fever; severe, localized abdominal pain)
2	Pelvic, retroperitoneal, or distant intraperitoneal abscess (abscess or phlegmon may be palpable, fever, systemic toxicity)
3	Generalized purulent peritonitis, no communication with bowel lumen
4	Feculent peritonitis, open communication with bowel lumen
Complications	Fistula, obstruction (large bowel or small bowel)

Imaging studies Patients with symptoms and signs of mild, uncomplicated diverticulitis who respond promptly to outpatient medical treatment do not necessarily require an imaging study immediately. Confirmation of the diagnosis can be delayed for 4 to 6 weeks, when active inflammation has resolved. If there is uncertainty about the diagnosis, outpatient CT is performed to exclude conditions that mimic diverticulitis [*see* Differential Diagnosis, *below*].

On CT, diverticula are seen as collections of gas or contrast measuring 5 to 10 mm and protruding from the wall of the colon. Symmetrical thickening of the colonic wall may be noted. In diverticulitis, a phlegmon is marked by streaky enhancement of pericolic or perirectal soft tissue and the mesentery [*see Figure 1*]. Perforation and fistula may be visualized by air or contrast. Abscess is seen as one or more discrete fluid collections.[25] If the abscess communicates with the colonic lumen, contrast may enter the abscess cavity. CT readily detects remote abscess. When abscess is detected, the feasibility of CT-guided drainage can be determined.[8,26,27]

Diagnostic tests The objectives of diagnostic testing in suspected acute diverticulitis are to exclude other important conditions, to confirm the diagnosis of diverticulitis, to determine if complications have occurred, and to plan treatment.

Leukocytosis is usually present in acute diverticulitis. The urine may contain a modest number of white cells or red blood cells. Recurrent or polymicrobial urinary tract infections should suggest the possibility of a colovesical fistula. Plain abdominal x-rays are most useful to exclude other abdominal conditions, such as intestinal obstruction. Occasionally, an inflammatory mass with gas may be noted, confirming the presence of an abscess. Free air in the abdominal cavity is unusual in diverticulitis.

Colonoscopy is generally not required for diagnosis in suspected acute diverticulitis, and air insufflation caused by introduction of air may worsen a contained perforation. Colonoscopy can be performed with relative safety if no fluid or free air is noted on abdominal CT.[28] After treatment and resolution of an acute epsisode of diverticulitis, patients should have an elective examination either by colonoscopy or by fiberoptic sigmoidoscopy after barium enema ; the purpose of elective colonoscopy is to exclude the presence of colon cancer and inflammatory bowel disease.[29] Barium enema should be avoided in suspected acute diverticulitis because of the risk of barium contamination of the peritoneum if a perforation is present.

Differential Diagnosis

A number of conditions may mimic acute diverticulitis [*see Table 2*]; among the differential diagnoses less frequently considered are epiploic appendagitis and omental torsion/infarction, which may be distinguished from diverticulitis by CT. Epiploic appendices are small peritoneal pouches filled with fat that are situated along the margin of the colon. These structures can become inflamed, resulting in acute epiploic appendagitis, which may mimic diverticulitis or appendicitis. The clinical presentation consists of acute abdominal pain and tenderness. Peritoneal signs sometimes occur, as do low-grade fever and mild leukocytosis. On CT scan, epiploic appendagitis has a characteristic appearance: an oval, fatty mass surrounded by mesenteric stranding, and mural thickening of adjacent colon is typically present.[30-32]

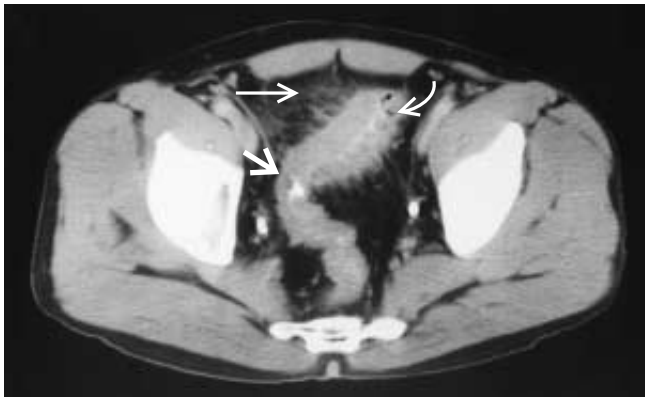


Figure 1 **CT diverticulitis. The wall of the sigmoid colon is thickened (broad arrow). Air is seen within a diverticulum (curved arrow). Streaky enhancement of pericolis fat (horizontal arrow) is caused by inflammation.**

In rare cases, the omentum may undergo spontaneous torsion, causing ischemia or infarction. The clinical presentation may mimic acute diverticulitis, appendicitis, or cholecystitis. The diagnosis of omental torsion may be determined preoperatively by the use of abdominal CT.[33,34]

Diverticulitis in Specific Patient Groups

Immunocompromised patients Immunocompromised patients, such as those on glucocorticoids or those who have had an organ transplant, may not manifest the usual signs of diverticulitis, and diagnosis in these patients may therefore be delayed. The severity of diverticulitis may also be underestimated. The threshold for diagnostic evaluation should be low in such patients.

Abdominal CT with rapid helical technique is the most useful diagnostic study when complicated acute diverticulitis is suspected or when the diagnosis is not clear. Specificity and sensitivity are reported to be over 95%.[35] The colon should be filled with water-soluble contrast given either orally (most commonly) or by gentle enema. CT can confirm the diagnosis, identify complications, and aid in planning of treatment. If the patient does not have acute diverticulitis, abdominal CT will usually lead to the correct diagnosis. Conditions other than diverticulitis are found in up to 25% of CT studies.[35]

Women suspected of gynecologic conditions Abdominal ultrasonography may be most useful for women when gynecologic conditions are part of the differential diagnosis. Abdominal ultrasonography may be an alternative test when CT is not readily available or is contraindicated. On ultrasound, diverticula are echogenic and produce acoustic shadowing. On a graded compression ultrasound, the colonic wall in diverticulitis is thickened, noncompressible, and hypoechoic. The involved segment is hypoperistaltic. A phlegmon causes irregular enhancement of pericolic soft tissue, whereas an abscess appears as a fluid collection, within which gas is readily appreciated, if present. When an abscess is identified on ultrasound, a CT scan should be performed to evaluate the potential for radiographic drainage.[26,36]

Magnetic resonance imaging has been reported to be useful in the diagnosis of acute diverticulitis in pregnant women. MRI has the advantage over CT of avoiding fetal exposure to radiation.[37]

Table 2 Differential Diagnosis of Acute Diverticulitis

Inflammatory bowel disease: Crohn disease, ulcerative colitis
Perforated colon cancer
Ischemic colitis
Infectious colitis
Mesenteric appendagitis, omental torsion
Gynecologic conditions: pelvic inflammatory disease; ovarian torsion, ruptured follicle or cyst; endometriosis
Appendicitis (situs inversus)

TREATMENT

Outpatient Management

Fewer than 20% of patients with acute diverticulitis require hospitalization.[38] Patients who present with mild diverticulitis should be placed on a regimen of oral fluid/electrolyte solution (e.g., a sports drink) and oral antibiotics. Patients should eat no solid foods during this period. The antibiotic regimen should provide coverage against gram-negative and anaerobic bacteria. For example, amoxicillin–clavulanic acid at a dosage of 875/125 mg twice daily is acceptable monotherapy; a suitable combination therapy is a quinolone (e.g., levofloxacin, 750 mg once daily) combined with metronidazole (500 mg twice daily).[39] The patient should be instructed to report back at once if symptoms worsen. Reevaluation is scheduled for 48 to 72 hours after the office visit. If improvement is satisfactory, the diet is advanced to full liquids, antibiotics are continued, and another office visit is scheduled at 7 days. If improvement is evident at 7 days, the patient can resume a regular diet and discontinue antibiotics.[4] Patients whose symptoms worsen or who do not have a favorable response within 48 to 72 hours should be hospitalized. If there is uncertainty about the diagnosis, outpatient CT is performed. An elective colonoscopy or sigmoidoscopy and barium enema exam is scheduled for 6 weeks after the acute illness, unless such a study was performed within the past 5 years.

Inpatient Management

Patients should be hospitalized if there are signs of severe or complicated diverticulitis, such as systemic toxicity, temperature exceeding 101° F (38.3° C), vomiting, an abdominal mass, or signs of peritonitis; patients should also be hospitalized if they fail to respond within 2 to 3 days to outpatient management. Hospitalized patients should be placed on bowel rest and given intravenous fluids and antibiotics. Antibiotic coverage must include both aerobic and anaerobic gram-negative bacteria. An example of a suitable monotherapy regimen is ampicillin-sulbactam (1.5 to 3.0 g every 6 hours); an acceptable combination therapy regimen is levofloxacin (750 mg I.V. once daily) combined with metronidazole (1 g I.V. every 12 hours).[39] A surgical consultation should be obtained upon admission. CT scanning should be done promptly. If a phlegmon or small abscess (< 3 cm) is found, antibiotic treatment alone may suffice. Abscesses larger than 5 cm should be drained by interventional radiology, unless radiologic drainage is contraindicated by the location of the abscess or the presence of multiple abscesses.[27] Antibiotic treatment and radiologic abscess drainage often allow control of infection in cases of complicated diverticulitis. Control of infection im-

proves the possibility of elective single-stage resection and reanastomosis. Urgent surgery should be considered for patients with large abscesses that are not amenable to radiologic drainage or with multiple abscesses; for patients failing to respond within 48 to 72 hours; and for patients with evidence of free rupture of an abscess or a large perforation with fecal spillage. About 20% to 30% of patients hospitalized for the first time with acute diverticulitis require either urgent or elective surgery.[17]

Surgical Treatment

Open resection The optimal surgical treatment of acute diverticulitis involves resection of the involved segment at the initial operation whenever this is technically possible. Leaving the diseased colon in place and performing only a diverting colostomy is associated with a higher rate of complications than primary resection. Primary reanastomosis is generally possible.[40] If there is concern that the anastomosis is at undue risk of disruption, a temporary diverting ileostomy may be performed. Alternatively, the distal rectal segment can be closed (Hartmann procedure) and a descending colostomy created. Patients undergoing surgery for diverticulitis should be informed about the possibility that an ostomy, if created, may be permanent. Because of comorbidities and other issues, about 35% of ostomies performed for diverticulitis will still be in place 4 years after surgery.[41] In the case of sigmoid diverticulitis, it is important to extend the resection to the rectum—to include the entire segment involved with diverticula—because failure to do so markedly increases the probability of recurrent diverticulitis.[42,43]

Laparoscopic surgery Laparoscopic surgical techniques are increasingly being used for diverticular disease. Results of laparoscopic resection for diverticulitis are the same as those of open resection if the resection extends to the rectum and no sigmoid colon is left in place.[44,45] Laparoscopic resection is safe and effective; its advantages over open surgery include decreased blood loss, faster recovery of bowel function, and shorter hospital stay. There are no differences in operative time or mortality with the two procedures.[45] Less than 10% of cases require conversion from laparoscopic to open resection.[44]

Elective surgery Factors considered in the recommendation of elective surgery for diverticulitis include the general health of the patient, the number and severity of episodes, and the degree to which symptoms resolve between episodes. As mentioned, about 20% to 25% of patients will have a subsequent episode of complicated diverticulitis within several years after their first episode.[18,20] Surgery is often recommended after one episode of complicated diverticulitis or two episodes of uncomplicated diverticulitis; however, a recent decision analysis study suggested that the best overall outcome—taking into account mortality, morbidity, the number of surgical procedures, and the number of ostomies—may be achieved if elective colectomy is recommended after the fourth episode of uncomplicated diverticulitis, which is a more conservative recommendation than is generally practiced.[46]

Young, overweight men have been reported in some series to have a higher risk of complicated diverticulitis and recurrent diverticulitis than other patients. Some surgeons therefore recommend surgery after the first episode of diverticulitis in such patients[29,47,48]; however, this increased risk has not been confirmed in other series, and some surgeons have suggested that the recommendation for surgery after the first episode be tempered.[20,49,50]

Preventive Treatment

A diet high in insoluble fiber and low in fat and red meat appears to reduce the risk of diverticular disease. Higher levels of physical activity are also associated with reduced risk of diverticulitis.[8,9]

The intermittent use of rifaximin, a poorly absorbed oral antibiotic, has been reported to provide a greater reduction of symptoms of diverticular disease and risk of recurrence than the use of fiber alone.[15] Rifaximin is given in a dosage of 400 mg by mouth twice daily for 1 week of each month. Mesalamine has been used for the same purpose in a dosage of 800 mg by mouth twice daily for 1 week of each month. The combination of rifaximin and mesalamine has been reported to be more effective than mesalamine alone.[16]

It is prudent to advise patients with a history of diverticular disease to avoid NSAIDs if possible, as these medications have been associated with an increased risk of complications.[5,7]

COMPLICATIONS OF DIVERTICULAR DISEASE

Fistula

Fistula is the presenting complication in 10% to 15% of patients who require surgery for diverticular disease. There is often no history of acute diverticulitis. Patients may present with symptoms related primarily to the organ involved with the diverticular fistula, which is the bladder in most cases. Diverticular disease is the most common cause of colovesical fistula, followed by colon cancer and Crohn disease. Colovesical fistulas are much less common in women than in men, presumably because the uterus is interposed between the sigmoid colon and the bladder. Patients with colovesical fistulas usually present with recurrent polymicrobial bladder infections, pneumaturia, or both. Fistulas may also connect with other parts of the colon, the small bowel, the uterus, or the vagina. Colocutaneous fistulas are unusual and generally occur after surgery for diverticulitis.[51]

Colovesical fistulas are difficult to visualize. CT scanning may be the most useful single study. Even though fistulas are rarely directly visualized, CT is very sensitive for detection of air in the bladder, which is virtually diagnostic of an enterovesical fistula in the absence of prior bladder catheterization. CT demonstrates the presence of diverticula, and thickening of the bladder wall adjacent to an area of diverticulitis supports the diagnosis.[52] Colovesical fistulas are often not identified by a barium enema study, although secondary changes of diverticulitis are usually apparent. Cystoscopy often reveals focal mucosal inflammation in the area of a fistula, even though the opening is not apparent. Fistulas may be visualized by contrast cystography.[53] Colonoscopy should be done at some point to exclude cancer or inflammatory bowel disease, but it rarely reveals the fistula. CT, barium enema, cystoscopy, and colonoscopy are complementary studies for the evaluation of a suspected colovesical fistula.[52] Patients with a diverticular fistula should have elective surgical resection of the involved segment of colon once the acute inflammatory process has been controlled. In the case of colovesical fistulas, the adherent colon can usually be dissected off the bladder and the involved bladder oversewn, rather than resected.[54]

Obstruction and Stricture

Acute obstruction during an episode of diverticulitis is often self-limited. It may be caused by a large phlegmon or abscess. Secondary ileus may mimic obstruction. Chronic stricture is an uncommon presentation of diverticular disease. It usually occurs

after repeated episodes of diverticulitis. A diverticular stricture may be difficult to distinguish from a malignant stricture, particularly by CT or a barium enema study. Colonoscopy with biopsies is the best means of making this distinction, although visualization may be limited even with colonoscopy.

Segmental Colitis Associated with Diverticular Disease

In about 1% of cases, chronic diverticular disease is associated with patchy mucosal hemorrhage, congestion, and granularity in the sigmoid colon. On microscopic examination, a lymphocytic infiltrate, lymphoglandular complexes, mucin depletion, mild cryptitis, crypt distortion, Paneth cell metaplasia, and ulceration can be seen. These changes may mimic Crohn colitis, ulcerative colitis, or ischemic colitis. However, in segmental colitis associated with diverticular disease (SCAD), the changes are limited to areas of diverticulosis.[55,56]

Risk of Colon Cancer in Diverticular Disease

Some studies have shown an increased risk of sigmoid colon cancer in patients with a history of diverticulitis, but other studies have not confirmed this finding.[57-59] In addition to the difficulty in distinguishing a diverticular stricture from a malignant stricture, perforated colon cancer can mimic acute diverticulitis accompanied by a phlegmon or abscess. Thus, it seems prudent to recommend elective colonoscopy to patients with diverticular disease.

DIVERTICULAR BLEEDING

Diverticular bleeding is less frequent than diverticulitis as a complication of diverticulosis. It is difficult to arrive at a firm estimate of what proportion of major lower gastrointestinal bleeding is caused by colonic diverticula. Estimates have ranged from 15% to 56%.[60-62] In many cases, the source is presumed to be diverticular when no other cause is found.

Diverticular bleeding is usually sudden and painless. Typically, moderate to large amounts of bright-red blood, clots, or maroon stool are passed. Bleeding stops spontaneously in 70% to 80% of cases; bleeding eventually recurs in 25% to 35% of cases.[63,64] Lower gastrointestinal bleeding (occult bleeding in particular) should not be attributed to diverticulosis unless other causes have been excluded. The most common causes of lower gastrointestinal bleeding in adults are diverticulosis, inflammatory bowel disease, neoplasm, angioectasias, benign anorectal disease, and upper gastrointestinal sources.[65]

Practice guidelines for the evaluation and management of lower gastrointestinal bleeding have been published.[64]

Emergent Management and Evaluation

The first priority in lower gastrointestinal bleeding is volume restoration with intravenous fluids and blood. During resuscitation, a directed history is taken and a physical examination is performed; the physical examination includes anoscopy and proctoscopy to exclude an anorectal source. A nasogastric aspirate should be obtained, because about 10% to 15% of apparent major lower gastrointestinal bleeding is, in actuality, from an upper gastrointestinal source.[61] Lack of blood in the nasogastric aspirate, however, is not conclusive proof of a lower gastrointestinal source. Bleeding from an upper gastrointestinal source may have stopped, and the stomach may have evacuated residual blood. Blood from a duodenal source may not reflux into the stomach. Some clinicians perform an upper gastrointestinal endoscopy before colonoscopy even if the nasogastric aspirate is negative for blood, particularly for apparent major lower gastrointestinal bleeding.

Colonoscopy Many clinicians consider colonoscopy to be the most useful initial study for the evaluation of patients presenting with major lower gastrointestinal bleeding of unknown cause.[66] Lower gastrointestinal bleeding is usually intermittent and often stops before hospital admission or shortly thereafter. Adequate colonic purging followed by colonoscopy is almost always possible even in spite of moderate active lower gastrointestinal bleeding.[61,66] The most important contribution of colonoscopy is to establish a presumptive diagnosis of diverticular bleeding by excluding other causes and to determine the distribution of diverticula, which is important if colonic resection becomes necessary. Some authors recommend urgent colonoscopy and report a high success rate in localizing and treating diverticula responsible for bleeding.[66] In another report, however, there was no correlation between the timing of colonoscopy and success in localizing the bleeding diverticulum over an interval of 7 to 29 hours.[67] If active bleeding or an adherent clot is found at the time of colonoscopy, endoscopic therapy can be performed with epinephrine injection, cautery, hemostatic clips, or banding.[68,69]

Labeled blood cell scintigraphy Technetium-99m (^{99m}Tc)–labeled red blood cell scintigraphy is another tool for the evaluation of acute lower gastrointestinal bleeding. After injection with ^{99m}Tc, imaging is done continuously for 60 to 120 minutes. Dynamic scans are viewed in a computer-generated cinematic format. Continuous scanning improves the probability of detecting intermittent bleeding.[70] The radioactive label remains active for over 24 hours; thus, if the scan is not positive initially, scanning can be repeated at any time during the first day if there are indications of recurrent bleeding. A scan is considered positive if there is a focus of increased activity that changes in location and intensity over time [*see Figure 2*]. ^{99m}Tc-labeled red blood cell scanning is the most sensitive study for detection of acute gastrointestinal bleeding. It can theoretically detect bleeding rates as low as 0.5 ml/min. Localization of bleeding to the small bowel, right colon, or left colon is generally possible, but reports of the accuracy of ^{99m}Tc-labeled red blood cell scanning have been inconsistent.[70-72] False positive readings may occur with vascular neoplasms, inflammatory conditions, vascular grafts, varices, splenosis, and bladder or penile activity.

^{99m}Tc scanning does not elucidate the cause of bleeding; rather, it demonstrates active bleeding and the approximate location of its source. Colonoscopy is still necessary to determine the presence of diverticula and exclude other potential colonic sources. Once diverticula have been established as the probable cause of bleeding, scanning has a role in detecting recurrent bleeding. A positive ^{99m}Tc scan also provides prognostic information; patients with positive scans are much more likely to have positive arteriograms and are more likely to require surgery than those with negative scans.[72]

Treatment

Visceral angiography and vasopressin infusion Visceral angiography may be useful in persistent moderate to severe lower gastrointestinal bleeding, particularly when the site has not been discovered by colonoscopy and upper gastrointestinal endoscopy or when colonoscopy is not feasible. Angiography can provide precise localization of bleeding in patients who have a positive bleeding scan. If active bleeding is detected and local-

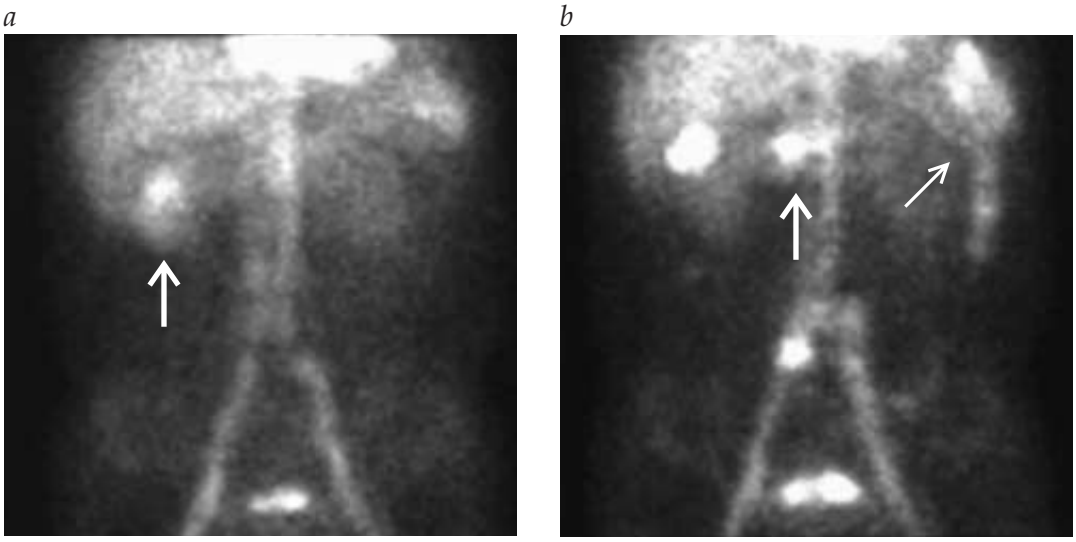


Figure 2 ***(a)*** **Technetium-99m red cell scanning shows increased tracer activity in the area of the hepatic flexure (arrow).** ***(b)*** **Later, tracer activity has progressed to the transverse colon (broad arrow), as well as the splenic flexure and descending colon (small arrow), confirming that the right colon is the source of the bleeding.**

ized, it may then be treated with intra-arterial vasopressin infusion. Rebleeding occurs in 25% of patients when the vasopressin infusion is stopped. Even temporary control may provide time to stabilize the patient and prepare for elective surgery.[64,71]

Arterial embolization Arterial embolization of the bleeding artery has been reported, but it is associated with a substantial risk of intestinal infarction, particularly if followed by vasopressin infusion.[71] Bowel or myocardial ischemia, as well as other complications inherent to contrast arteriography, may complicate both vasopressin infusion and embolization.

Surgery About 20% of patients hospitalized for diverticular bleeding require surgery during that hospitalization.[63] Patients who initially require four units or more of blood have a 50% risk of continued bleeding, as compared with a 2% risk of continued bleeding for patients who require two or more units of blood. Persistent instability despite aggressive resuscitation demands surgical intervention.[63]

Elective surgery for recurrent bleeding The risk of recurrent diverticular bleeding after the first episode is about 10% after 2 years and 25% after 4 years.[60] The risk is higher after a second episode. Elective surgery is usually recommended after two or more episodes of bleeding. If diverticula are limited to the left side of the colon, left hemicolectomy is appropriate. The decision regarding the extent of resection is more difficult when diverticula are distributed throughout the colon. If all previous episodes of diverticular bleeding have been localized to either the right or the left side by scintigraphy or angiography, some clinicians would advocate a corresponding hemicolectomy. The risk of recurrent bleeding from diverticula in the remaining colon is not known. Other clinicians would recommend a subtotal colectomy. Occasionally, persistent or recurrent severe lower gastrointestinal bleeding cannot be localized. Blind segmental resection is associated with an unacceptable recurrence rate, and subtotal colectomy is the favored procedure.[61]

Appendicitis

EPIDEMIOLOGY

The lifetime risk of appendicitis is about 9% for males and 7% for females.[73] Appendicitis rarely occurs in infants; it increases in frequency between the ages of 2 and 4; and it reaches a peak between the ages of 10 and 20. About 80% of cases occur before the age of 45. Nevertheless, there is a steady low incidence in older individuals. The mortality associated with acute appendicitis declined between 1945 and 1960, coincident with advances in antibiotic treatment. In 1990, the mortality associated with acute uncomplicated appendicitis was approximately equal to that associated with general anesthesia. However, the mortality associated with gangrenous appendicitis is about 0.5 %, and that of perforated appendicitis is 5%. Most deaths from acute appendicitis occur in persons older than 65 years.[73]

ANATOMY AND PATHOGENESIS

The adult appendix is a tubular structure that is 4 to 25 cm long and arises from the medial posterior wall of the cecum several centimeters below the ileocecal valve. Its location in the peritoneal cavity varies. Atypical locations, such as the pelvis, retrocecal area, and right upper quadrant, lead to atypical clinical presentations [*see* Atypical Presentations, *below*].

Appendicitis is generally caused by obstruction of the lumen of the appendix, followed by infection. The appendix has abundant lymphoid tissue. Appendicitis increases in frequency during the period of lymphoid hyperplasia in childhood. During periods of childhood enteric infection, lymphoid tissue may obstruct the appendiceal lumen. About one third of cases of appendicitis are associated with obstruction by fecaliths. Foreign bodies, tumor (e.g., carcinoid or cecal adenocarcinoma), barium, and adhesions may also cause obstruction. Obstruction leads to bacterial overgrowth. Mucus accumulates in the lumen proximal to the obstruction, and intraluminal pressure increases. Impairment of lymphatic and venous drainage leads to mucosal ulceration, bacterial invasion, transmural inflammation, and ischemia. During the first 24 hours after obstruction, most patients

have only inflammation. The incidence of necrosis and perforation increases markedly after that. Patients who present with a history of symptoms for 48 hours should be strongly suspected of having perforation and abscess. Free perforation causes generalized peritonitis.

DIAGNOSIS

Clinical Presentation

Appendicitis usually causes a distinctive sequence of symptoms and signs. More than 90% of patients with appendicitis complain of pain. The pain of appendicitis is initially caused by obstruction of the appendiceal lumen. It has the qualities of midgut visceral pain and is referred to the periumbilical or epigastric areas. It may be cramping or aching in nature, but it is often difficult for patients to describe. Within 12 to 24 hours, inflammation becomes transmural, involving the adjacent parietal peritoneum. Pain then becomes somatic in quality: sharper and more localized. At this time, patients may note exacerbation of pain by coughing, sneezing, or movement.

Anorexia is present in 80% to 90% of patients. Vomiting, when it occurs, does not occur initially but follows the onset of pain. Prominent vomiting is unusual and suggests the possibility of another diagnosis, such as gastroenteritis or small bowel obstruction. Fever is usually low grade. High fever or rigors suggest perforation.

Tenderness in the right lower quadrant can be elicited in more than 90% of patients. Proximity of the inflammatory process to the retroperitoneal muscles produces the psoas and obturator signs. The psoas sign is present when pain occurs as the patient raises the right leg against resistance or, alternatively, when the physician passively extends the right hip with the patient lying on the left side. The obturator sign is present when pain occurs upon internal rotation of the hip. Local hyperesthesia of the skin in the right lower quadrant may be noted. Voluntary guarding progresses to involuntary muscle rigidity as the inflammatory process worsens. Diffuse abdominal tenderness and rigidity suggest perforation. An abdominal mass suggests phlegmon or abscess formation.

The presentation of appendicitis may mimic a broad range of diseases [*see Table 3*].

There is evidence that appendicitis may resolve spontaneously and recur. Occasionally, the patient history includes previous episodes of acute appendicitis that resolved without treatment. Examination of appendixes removed incidentally at surgery or at autopsy sometimes show fibrosis and obliteration of a portion of the lumen, suggesting previous episodes of appendicitis.[74,75]

Atypical presentations Atypical presentations of appendicitis are as common as the classic presentation. Atypical location of the appendix leads to atypical symptoms and signs. An inflamed retrocecal appendix is relatively shielded from the parietal peritoneum. Pain may be less severe and abdominal tenderness less pronounced. The characteristic shift in pain location to the right lower quadrant may be delayed. A pelvic appendix may cause symptoms resulting from inflammation of the bladder or rectum, such as dysuria or tenesmus. In such cases, tenderness may be best elicited on pelvic or rectal examination. Incomplete intestinal rotation and third-trimester pregnancy displace the appendix toward the right upper quadrant, causing confusion with cholecystitis or perforated peptic ulcer.

Tip appendicitis and stump appendicitis In appendicitis that involves only the tip of the appendix, inflammation may be less severe and may resolve spontaneously; cases of recurrent appendicitis may involve such a pathophysiology. Partial visualization of the appendix on CT may be mistaken for complete filling and lead to a false negative reading.[76] Appendicitis has been reported to occur in the stump of the appendix that is left behind after a previous appendectomy; this can lead to confusion and a delay in diagnosis.[77]

Appendicitis and appendiceal tumors Tumor of the appendix, such as carcinoids, may obstruct the lumen and lead to appendicitis. In elderly men with appendicitis, there is a relatively high incidence of appendiceal tumors.[13,69]

Appendicitis in special groups of patients The diagnosis of appendicitis is difficult in certain groups of patients. Young children, for example, often do not express their symptoms clearly; they may present with only lethargy, irritability, and anorexia.[78] Elderly patients may have a reduced inflammatory reaction; in such cases, pain may be vague, and there may be less fever or abdominal tenderness.[79] Consideration of other diagnoses, such as diverticulitis, may delay surgery. Delayed diagnosis of appendicitis contributes to an incidence of perforation that is close to 20% in the elderly and small children.[80]

In women of childbearing age, gynecologic conditions and pregnancy cause difficulty in diagnosing appendicitis.[81] In one study, the preoperative diagnosis was incorrect in 25% of cases.[82] Complications of pregnancy attributable to appendicitis are considerable. Appendicitis in pregnancy has been reported to result in fetal loss in 33% of first-trimester cases and 14% of second-trimester cases.[82]

Diagnosis of appendicitis may also be difficult in immunosuppressed patients; for example, patients on glucocorticoids often have attenuated symptoms, leading to delay in presentation and a high incidence of complications. In patients with AIDS, symptoms may be typical, but concerns about a multiplicity of other diagnoses may delay surgery.

Diagnostic Evaluation

In the management of suspected appendicitis, it is important to minimize the delay to surgery—and thus reduce the risk of perforation—while at the same time minimizing the number of

Table 3 Differential Diagnosis of Appendicitis

Crohn disease
Gynecologic conditions: ovarian torsion (especially during pregnancy), ovarian vein thrombosis, endometriosis
Perforated right-sided colon cancer
Cecal diverticulitis
Foreign-body perforation of right colon
Meckel diverticulitis
Omental torsion
Epiploic appendagitis
Infections (particularly those involving the ileum): yersinosis, brucellosis, salmonellosis, tuberculosis, amebiasis
Vasculitis
Appendiceal tumors
Psoas abscess

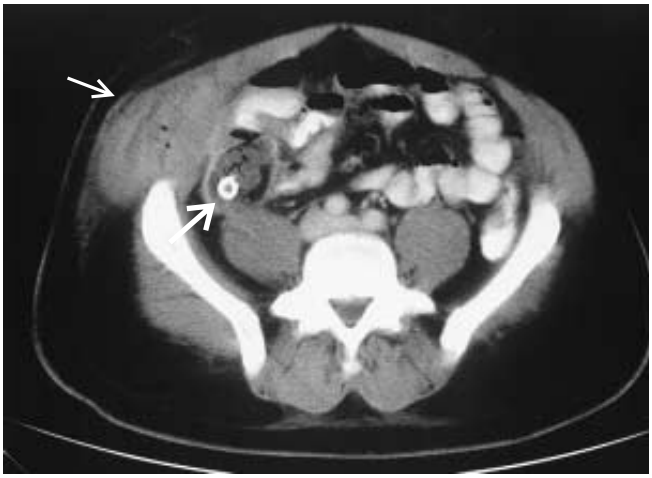


Figure 3 **CT of appendicitis. A calcified appendicolith is noted within a complex abscess containing fluid and air (broad arrow). The abscess involves the soft tissues of the abdominal wall (small arrow).**

unnecessary appendectomies. Taking into account the serious consequences of progression to perforation and the relatively low morbidity of appendectomy, most surgeons adopt a relatively aggressive approach to early surgery, accepting a 10% to 15% rate of negative appendectomies (i.e., negative exploration in patients with clinically suspected acute appendicitis).

Some factors that lead to a delay in diagnosis are not under physician control, such as a patient's delay in seeking medical care. Factors that are under physician control include decisions to perform additional diagnostic studies or to observe the patient for the purpose of improving diagnostic accuracy and reducing the number of unnecessary appendectomies. Clinical scoring systems have generally not been found to improve preoperative diagnostic accuracy.[83]

Abdominal examination The abdominal examination is very important for the diagnosis of appendicitis. This has led some physicians to delay the administration of narcotic analgesics until a surgeon has had the opportunity to examine the patient. It has been proved that this practice is unnecessary: morphine does not change the physical examination in acute appendicitis, and early pain relief does not affect the decision for surgery in adults.[84,85]

Diagnostic Tests

The role of diagnostic studies for suspected appendicitis depends on the particular clinical presentation. In cases characterized by a classic presentation, it is standard practice to base the decision to operate primarily on the history and physical examination. A complete blood count, a urinalysis, and plain x-rays of the chest and abdomen may be obtained, but these serve the purpose of excluding other conditions rather than confirming the diagnosis of appendicitis. The total white blood cell count, the differential count, or both are abnormal in more than 90% of cases, but the decision to perform surgery should not be delayed if the white blood cell count is not elevated.[86]

In women of childbearing age, a pregnancy test is mandatory. In many cases, no further diagnostic studies are performed. Among women of childbearing age, the rate of negative appendectomy can be as high as 40%.[81] In other patients, particularly young children or the elderly, the presentation may be atypical and the diagnosis uncertain. In such patients, additional diagnostic studies may be appropriate and helpful.

Abdominal ultrasound Examination by transabdominal or transvaginal ultrasound or both is useful in pregnant and nonpregnant women of childbearing age to exclude a gynecologic cause of symptoms.[81,87] Ultrasound examination is also useful for evaluation of children in cases in which the diagnosis is doubtful. Sonography is widely available, fast, safe, and inexpensive.

On graded compression ultrasonography, the inflamed appendix is a noncompressible, aperistaltic tubular structure that is greater than 6 mm in diameter and located in the right lower quadrant. It has a target appearance, and the lumen is filled with anechoic or hyperechoic material. An appendicolith may be visualized in up to 30% of cases. Pericecal inflammation or phlegmon is seen as prominent fat; abscess appears as loculated fluid. In experienced hands, the sensitivity and specificity for the diagnosis of appendicitis are 85% and 95%, respectively.[88] Despite abdominal tenderness, most patients find the examination tolerable if it is performed gently. Marked peritonitis or abdominal gas may compromise the examination.

Abdominal ultrasound is not as accurate as CT for the diagnosis of appendicitis in adults and adolescents; in addition, it is not as useful as CT for the evaluation of phlegmon or abscess.[89]

Abdominal CT Abdominal CT scanning can increase diagnostic accuracy in many cases of suspected appendicitis. The development of rapid scanning techniques has made CT readily available and practical even in emergency room settings. CT is useful when the diagnosis is unclear, as in elderly patients in whom diverticulitis and perforated colon cancer are important considerations. Abdominal CT may be helpful in women if pregnancy has been ruled out.[90] The routine use of appendiceal CT in the emergency room setting has been reported to reduce both the number of unnecessary appendectomies and delays before necessary appendectomies.[91] CT is helpful when an appendiceal abscess is suspected. It provides information about the size, location, and number of abscesses, as well as the feasibility of percutaneous drainage under radiologic guidance.[92]

CT for suspected appendicitis has been reported to have an accuracy of about 95%.[93,94] CT scan for appendicitis is best done using thin collimation helical scanning. The terminal ileum and cecum must be opacified with contrast that is administered by mouth or by rectum. Rectal administration of contrast has been reported to have advantages over oral contrast for assessing appendiceal filling.[84] The diagnosis of appendicitis is established when pericecal inflammation , phlegmon, or abscess is seen with either an appendicolith or an abnormal appendix. In appendicitis, the appendix is enlarged to more than 6 mm in diameter and fails to fill with contrast. If intravenous contrast has been given, the inflamed appendix will also show enhancement.[71] CT is useful for the evaluation of the degree and extent of periappendiceal inflammation [see *Figure 3*]. Inflammation may cause thickening of the adjacent cecum or ileum. Streakiness of periappendiceal fat is seen with phlegmon. Loculated fluid and, sometimes, gas bubbles are seen with abscess; if gas bubbles are not present, an air-fluid level may be apparent.

Barium enema Since the advent of abdominal ultrasound and CT examination, barium enema has had little place in the evaluation of suspected appendicitis. Appendicitis is most likely to be discovered on barium enema exam when the study is done

for determination of another diagnosis. In acute appendicitis, the appendix fails to fill with contrast on barium enema exam; this finding is more valuable in children than in adults, because the appendix fails to fill with contrast in 15% to 20% of normal adults. Partial filling of the appendix is often difficult to distinguish from complete filling, given the marked anatomic variations in the length of the appendix. Appendicitis may produce a mass effect or inflammatory changes in the adjacent cecum or ileum that can be appreciated by barium enema, but CT provides much more information than barium enema in this regard. Barium enema exam may be helpful in the evaluation of chronic or recurrent right lower quadrant abdominal conditions that mimic "chronic" appendicitis, such as Crohn disease.

TREATMENT

Preoperative Management

The treatment of appendicitis is prompt appendectomy. Preoperative preparation consists of intravenous volume repletion and antibiotics. For simple appendicitis, one dose of a broad-spectrum antibiotic given before surgery and one dose given postoperatively is sufficient.[95] One example of an appropriate regimen for adults is cefazolin (1.5 g I.V.) preoperatively and metronidazole (500 mg I.V.) postoperatively.[39]

If perforated appendicitis is suspected, antibiotic coverage should be broadened. An example of an acceptable regimen for adults is levofloxacin (750 mg I.V. once daily) combined with metronidazole (1 g I.V. every 12 hours).[39] If the presence of a phlegmon or abscess is confirmed, antibiotics are customarily continued for 7 to 14 days postoperatively. However, a prospective, randomized study found that adding a 7-day regimen of oral antibiotics after a course of intravenous antibiotics made no difference in outcome, either in complicated or uncomplicated appendicitis.[96]

Patients with free, unconfined perforation should have abdominal saline lavage during surgery. A prolonged ileus should be anticipated.

Appendectomy

Appendectomy can be performed by a traditional open incision or by laparoscopy. Compared to open appendectomy, laparoscopic appendectomy results in decreased wound infections but slightly increased intra-abdominal infections. Postoperative pain is reduced in laparoscopic appendectomy, and patients return to normal activity sooner. Overall, the benefits of laparoscopic appendectomy over open appendectomy are modest; the greatest benefits occur in women and obese patients.[97-99] The laparoscopic approach is useful in women because it allows for accurate diagnosis of gynecologic conditions if appendicitis is not responsible for the symptoms.[100,101]

Complications occur in fewer than 5% of cases of simple appendicitis but can be anticipated in 30% to 50% of cases of appendicitis after perforation. The most common complications are wound infections, intra-abdominal abscess, intestinal obstruction, and prolonged ileus. Postoperative abscesses are heralded by recurrent malaise, anorexia, and fever, and they are best evaluated by CT.

Interval Appendectomy

Some patients with a contained perforation can be managed by interval appendectomy after treatment with antibiotics or CT-guided percutaneous drainage. The information from preoperative CT or sonography is useful in planning therapy. If imaging shows a phlegmon or small abscess and the patient responds to antibiotic treatment within 48 hours, appendectomy may be postponed for 6 weeks, until the inflammatory process has subsided. If the abscess is large but well circumscribed and accessible, CT-guided percutaneous catheter drainage may be used to reduce the abscess before surgery. The catheter is placed in the abscess and left until drainage from the abscess becomes minimal.[92] If drainage is successful, appendectomy can be postponed for 6 weeks. However, if a patient with a contained perforation does not respond promptly to antibiotic treatment or drainage, surgery should not be delayed. Catheter drainage is not possible if imaging shows a poorly defined or multilocular abscess or if the abscess is not accessible to percutaneous drainage.

If a patient is operated on for suspected appendicitis and the appendix is found to be normal at surgery, it should be removed to prevent future confusion. The cecum and terminal ileum should be examined for evidence of Crohn disease or other acute inflammatory bowel disease, for tumor, or for Meckel diverticulitis. Lymph nodes in the area should be inspected for evidence of mesenteric adenitis, and biopsies should be performed if the lymph nodes appear abnormal. The gallbladder and duodenum should be palpated. If necessary, the incision should be extended to permit wider exposure. Before closing, the surgeon must feel confident that the cause of the clinical presentation has been explained and that there is no other acute abdominal condition.

Incidental Appendectomy

Incidental appendectomy, performed during surgery for another cause, may be justified in individuals younger than 30 years if the primary surgery would not be compromised. Appendectomy does not increase morbidity when performed under these circumstances.[102]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Floch MH, Bina I: The natural history of diverticulitis: fact and theory. J Clin Gastroenterol 38(5 Suppl):S2, 2004

2. Mendeloff AI, Everhart JE: Diverticular disease of the colon. Digestive Diseases in the United States: Epidemiology and Impact. Everhart JE, Ed. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. Washington, DC, 1994, p 551

3. Nakada I, Ubukata H, Goto Y, et al: Diverticular disease of the colon at a regional general hospital in Japan. Dis Colon Rectum 38:755, 1995

4. Mizuki A, Nagata H, Tatemichi M, et al: The out-patient management of patients with acute mild-to-moderate colonic diverticulitis. Aliment Pharmacol Ther 21:889, 2005

5. Morris CR, Harvey IM, Stebbings WS, et al: Anti-inflammatory drugs, analgesics and the risk of perforated colonic diverticular disease. Br J Surg 90:1267, 2003

6. Makela J, Kiviniemi H, Laitinen S: Prevalence of perforated sigmoid diverticulitis is increasing. Dis Colon Rectum 45:955, 2002

7. Hart AR, Kennedy HJ, Stebbings WS, et al: How frequently do large bowel diverticula peforate? An incidence and cross-sectional study. Eur J Gastroenterol Hepatol 12:661, 2000

8. Ambrosetti P, Rober J, Witzig JA: Prognostic factors from computed tomography in acute left colonic diverticulitis. Br J Surg 79:117, 1992

9. Aldoori W, Ryan-Harshman M: Preventing diverticular disease. Review of recent evidence on high-fibre diets. Can Fam Physician 48:1632, 2002

10. Wess L, Eastwood MA, Wess TJ, et al: Cross linking of collagen is increased in colonic diverticulosis. Gut 37:91, 1995

11. Ludeman L, Warren BF, Shepherd NA: The pathology of diverticular disease. Best Pract Res Clin Gastroenterol 16:543, 2002

12. Whiteway J, Morson BC: Elastosis in diverticular disease of the sigmoid colon. Gut 26:258, 1985

13. Bucher P, Mathe Z, Demirag A, et al: Appendix tumors in the era of laparoscopic appendectomy. Surg Endosc 18:1063, 2004

14. Colecchia A, Sandri L, Capodicasa S, et al: Diverticular disease of the colon: new

perspectives in symptom development and treatment. World J Gastroenterol 9:1385, 2003

15. Tursi A: Preventive therapy for complicated diverticular disease of the colon: looking for a correct therapeutic approach. Gastroenterology 127:1865, 2004

16. Tursi A, Brandimarte G, Daffina R: Long-term treatment with mesalazine and rifaximin versus rifaximin alone for patients with recurrent attacks of acute diverticulitis of colon. Dig Liver Dis 34:510, 2002

17. Kaiser AM, Jeng-Kae J, Lake JP, et al: The management of complicated diverticulitis and the role of computed tomography. Am J Gastroenterol 100:910, 2005

18. Macias LH, Haukoos JS, Dixon MR, et al: Diverticulitis: truly minimally invasive management. Am Surg 70:932, 2004

19. Bahadursingh AM, Virgo KS, Kaminski DL, et al: Spectrum of disease and outcome of complicated diverticular disease. Am J Surg 186:696, 2003

20. Biondo S, Pares D, Marti RJ, et al: Acute colonic diverticulitis in patients under 50 years of age. Br J Surg 89:1137, 2002

21. Lo CY, Chu KW: Acute diverticulitis of the right colon. Am J Surg 171:244,1996

22. Junge K, Marx A, Peiper C, et al: Caecal-diverticulitis: a rare differential diagnosis for right-sided lower abdominal pain. Colorectal Dis 5:241, 2003

23. Fang JF, Chen RJ, Lin BC, et al: Aggressive resection is indicated for cecal diverticulitis. Am J Surg 185:135, 2003

24. Simpson J, Neal KR, Scholefield JH, et al: Patterns of pain in diverticular disease and the influence of acute diverticulitis. Eur J Gastroenterol Hepatol 15:1005, 2003

25. Kircher MF, Rhea JT, Kihiczak D, et al: Frequency, sensitivity, and specificity of individual signs of diverticulitis on thin-section helical CT with colonic contrast material: experience with 312 cases. AJR Am J Roentgenol 178:1313, 2002

26. McKee RF, Deignan RW, Krukowski ZH: Radiological investigation in acute diverticulitis. Br J Surg 80:560, 1993

27. Kaiser AM, Jiang JK, Lake JP, et al: The management of complicated diverticulitis and the role of computed tomography. Am J Gastroenterol 100:910, 2005

28. Sakhnini E, Lahat A, Melzer E, et al: Early colonoscopy in patients with acute diverticulitis: results of a prospective pilot study. Endoscopy 36:504, 2004

29. Wong WD, Wexner SD, Lowry A, et al: Practice parameters for sigmoid diverticulitis-supporting documentation. The Standards Task Force. The American Society of Colon and Rectal Surgeons. Dis Colon Rectum 43:290, 2000

30. Zissin R, Hertz M, Osadchy A, et al: Acute epiploic appendagitis: CT findings in 33 cases. Emerg Radiol 9:262, 2002

31. Legome EL, Belton AL, Murray RE, et al: Epiploic appendagitis: the emergency department presentation. J Emerg Med 22:9, 2002

32. Blinder E, Ledbetter S, Rybicki F: Primary epiploic appendagitis. Emerg Radiol 9:231, 2002

33. Young TH, Lee HS, Tang HS: Primary torsion of the greater omentum. Int Surg 89:72, 2004

34. Abadir JS, Cohen AJ, Wilson SE: Accurate diagnosis of infarction of omentum and appendices epiploicae by computed tomography. Am Surg 70:854, 2004

35. Werner A, Diehl SJ, Farag-Soliman M, et al: Multi-slice spiral CT in routine diagnosis of suspected acute left-sided colonic diverticulitis: a prospective study of 120 patients. Eur Radiol 13:2596, 2003

36. Zielke A, Hasse C, Nies C, et al: Prospective evaluation of ultrasonography in acute colonic diverticulitis. Br J Surg 84:385,1997

37. Schreyer AG, Furst A, Agha A, et al: Magnetic resonance imaging based colonography for diagnosis and assessment of diverticulosis and diverticulitis. Int J Colorectal Dis 19:474, 2004

38. Schecter S, Mulvey J, Eisenstat TE: Management of uncomplicated acute diverticulitis. Dis Colon Rectum 42: 470, 1999

39. Sanford JP, Gilbert DN, Sande MA: The Sanford Guide to Antimicrobial Therapy. Antimicrobial Therapy, Inc., Hyde Park, Vermont, 2004

40. Wedell J, Banzhaf G, Chaoui R, et al: Surgical management of complicated colonic diverticulitis. Br J Surg 84:380, 1997

41. Maggard MA, Zingmond D, O'Connell JB, et al: What proportion of patients with an ostomy (for diverticulitis) get reversed? Am Surg 70:928, 2004

42. Thaler K, Weiss EG, Nogueras JJ, et al: Recurrence rates at minimum five-year follow-up: laparoscopic versus open sigmoid resection for uncomplicated diverticulitis. Acta Chir Iugosl 51:45, 2004

43. Thaler K, Baig MK, Berho M, et al: Determinants of recurrence after sigmoid resection for uncomplicated diverticulitis. Dis Colon Rectum 46:385, 2003

44. Schwandner O, Farke S, Bruch HP: Laparoscopic colectomy for diverticulitis is not associated with increased morbidity when compared to non-diverticular disease. Int J Colorectal Dis 20:165, 2005

45. Gonzalez R, Smith CD, Mattar SG, et al: Laparoscopic vs open resection for the treatment of diverticular disease. Surg Endosc 18:276, 2004

46. Salem L, Veenstra DL, Sullivan SD, et al: The timing of elective colectomy in diverticulitis: a decision analysis. J Am Coll Surg 199:904, 2004

47. Spivak H, Weinrauch S, Harvey JC: Acute colonic diverticulitis in the young. Dis Colon Rectum 40:570, 1997

48. Schauer PR, Ramos R, Ghiatas AA, et al: Virulent diverticular disease in young obese men. Am J Surg 164:443, 1992

49. Chautems RC, Ambrosetti P, Ludwig A, et al: Long-term follow-up after first acute episode of sigmoid diverticulitis: is surgery mandatory?: a prospective study of 118 patients. Dis Colon Rectum 45:962, 2002

50. Guzzo J, Hyman N: Diverticulitis in young patients: is resection after a single attack always warranted? Dis Colon Rectum 47:1187, 2004

51. Lavery IC: Colonic fistulas. Surg Clin North Am 76:1183, 1996

52. Jarrett TW, Vaughan ED Jr: Accuracy of computerized tomography in the diagnosis of colovesical fistula secondary to diverticular disease. J Urol 153:44, 1995

53. Kirsh GM, Hampel N, Shuck JM, et al: Diagnosis and management of vesicoenteric fistulas. Surg Gynecol Obstet 173:91, 1991

54. Wilson RG, Smith AN, Macintyre IM: Complications of diverticular disease and non-steroidal anti-inflammatory drugs: a prospective study. Br J Surg 77:1103, 1990

55. West AB, Losada M: The pathology of diverticulosis coli. J Clin Gastroenterol 38:S11, 2004

56. Ghorai S, Ulbright TM, Rex DK: Endoscopic findings of diverticular inflammation in colonoscopy patients without clinical acute diverticulitis: prevalence and endoscopic spectrum. Am J Gastroenterol 98:802, 2003

57. Stefansson T, Ekbom A, Sparen P, et al: Association between sigmoid diverticulitis and left-sided colon cancer: a nested, population-based, case control study. Scand J Gastroenterol 39:743, 2004

58. Krones CJ, Klinge U, Butz N, et al: The rare epidemiologic coincidence of diverticular disease and advanced colonic neoplasia. Int J Colorectal Dis 2005

59. Kieff BJ, Eckert GJ, Imperiale TF: Is diverticulosis associated with colorectal neoplasia? A cross-sectional colonoscopic study. Am J Gastroenterol 99:2007, 2004

60. Longstreth GF: Epidemiology and outcome of patients hospitalized with acute lower gastrointestinal hemorrhage: a population-based study. Am J Gastroenterol 92:419, 1997

61. Jensen DM, Machicado GA: Colonoscopy for diagnosis and treatment of severe lower gastrointestinal bleeding: routine outcomes and cost analysis. Gastrointest Endosc Clin North Am 7:477, 1997

62. Peura DA, Lanza FL, Gostout CJ: The American College of Gastroenterology bleeding registry: preliminary findings. Am J Gastroenterol 92:924, 1997

63. Bokhari M, Vernava AM, Ure T, et al: Diverticular hemorrhage in the elderly: is it well tolerated? Dis Colon Rectum 39:191, 1996

64. Zuccaro G: Management of the adult patient with acute lower gastrointestinal bleeding. Am J Gastroenterol 93:1202, 1998

65. Zuckerman GR, Prakash C: Acute lower intestinal bleeding: part II: etiology, therapy, and outcomes. Gastrointest Endosc 49:228, 1999

66. Jensen DM, Machicado GA, Jutahbha R, et al: Urgent colonoscopy for the diagnosis and treatment of severe diverticular hemorrhage. N Engl J Med 342:78, 2000

67. Smoot RL, Gostout CJ, Rajan E, et al: Is early colonoscopy after admission for acute diverticular bleeding needed? Am J Gastroenterol 98:1996, 2003

68. Farrell JJ, Graeme-Cook F, Kelsey PB: Treatment of bleeding colonic diverticula by endoscopic band ligation: an in-vivo and ex-vivo pilot study. Endoscopy 35:823, 2003.

69. Simpson PW, Nguyen MH, Lim JK, et al: Use of endoclips in the treatment of massive colonic diverticular bleeding. Gastrointest Endosc 59:433, 2004

70. Maurer AH: Gastrointestinal bleeding and cine-scintigraphy. Semin Nucl Med 26:43, 1996

71. O'Neill BB, Gosnell JE, Lull RJ, et al: Cinematic nuclear scintigraphy reliably directs surgical intervention for patients with gastrointestinal bleeding. Arch Surg 135:1076, 2000

72. Suzman MS, Talmor M, Jennis R, et al: Accurate localization and surgical management of active lower gastrointestinal hemorrhage with technetium-labeled erythrocyte scintigraphy. Ann Surg 224:29, 1996

73. Everhart JE, Mendeloff AI: Acute appendicitis. Digestive Diseases in the United States: Epidemiology and Impact. Everhart JE, Ed. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Washington, DC, 1994, p 457

74. Barber MD, McLaren J, Rainey JB: Recurrent appendicitis. Br J Surg 84:110, 1997

75. Babb RR, Trollope ML: Recurrent appendicitis. uncommon, but it does occur. Postgrad Med 106:135, 1999

76. Levine CD, Aizenstein O, Wachsberg RH: Pitfalls in the CT diagnosis of appendicitis. Br J Radiol 77:792, 2004

77. Watkins BP, Kothari SN, Landercasper J: Stump appendicitis: case report and review. Surg Laparosc Endosc Percutan Tech 14:167, 2004

78. Mason JD: The evaluation of acute abdominal pain in children. Emerg Med Clin North Am 14:629, 1996

79. Franz MG, Norman J, Fabri P J: Increased morbidity of appendicitis with advancing age. Am Surg 61:40, 1995

80. Korner H, Sondenaa K, Soreida JA, et al: Incidence of acute nonperforated and perforated appendicitis: age-specific and sex-specific analysis. World J Surg 21:313, 1997

81. Rothrock SG, Green SM, Dobson M, et al: Misdiagnosis of appendicitis in nonpregnant women of childbearing age. J Emerg Med 13:1, 1995

82. Andersen B, Nielsen TF: Appendicitis in pregnancy: diagnosis, management and complications. Acta Obstet Gynecol Scand 78:758, 1999

83. Ohmann C, Yang Q, Franke C: Diagnostic scores for acute appendicitis. Abdominal Pain Study Group. Eur J Surg 161:273, 1995

84. Wolfe JM, Smithline HA, Phipen S, et al: Does morphine change the physical examination in patients with acute appendicitis? Am J Emerg Med 22:280, 2004

85. Vermeulen B, Morabia A, Unger PF, et al: Acute appendicitis: influence of early pain relief on the accuracy of clinical and US findings in the decision to operate—a randomized trial. Radiology 210:639, 1999

86. Cardall T, Glasser J, Guss DA: Clinical value of the total white blood cell count and temperature in the evaluation of patients with suspected appendicitis. Acad Emerg Med 11:1021, 2004

87. Barloon T J, Brown BP, Abu-Yousef MM, et al: Sonography of acute appendicitis in

pregnancy. Abdom Imaging 20:149, 1995

88. Kaiser S, Frenckner B, Jorulf HK: Suspected appendicitis in children: US and CT—a prospective randomized study. Radiology 223:633, 2002

89. Terasawa T, Blackmore CC, Bent S, et al: Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med 141:537, 2004

90. Antevil J, Rivera L, Langenberg B, et al: The influence of age and gender on the utility of computed tomography to diagnose acute appendicitis. Am Surg 70:850, 2004

91. Rao PM, Rhea JT, Novelline RA, et al: Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. N Engl J Med 338:141, 1998

92. Jeffrey RB, Federle MP, Tolentino CS: Periappendiceal inflammatory masses: CT-directed management and clinical outcome in 70 patients. Radiology 167:13, 1988

93. Rao PM, Rhea JT, Novelline RA, et al: Helical CT combined with contrast material administered only through the colon for imaging of suspected appendicitis. AJR Am J Roentgenol 169:1275, 1997

94. Stroman DL, Bayouth CV, Kuhn JA, et al: The role of computed tomography in the diagnosis of acute appendicitis. Am J Surg 178:485,1999

95. Gorecki WJ, Grochowski JA. Are antibiotics necessary in nonperforated appendicitis in children? A double blind randomized controlled trial. Med Sci Monit 7:289, 2001

96. Taylor E, Berjis A, Bosch T, et al: The efficacy of postoperative oral antibiotics in appendicitis: a randomized prospective double-blinded study. Am Surg 70:858, 2004

97. Hellberg A, Rudberg C, Kullman E, et al: Prospective randomized multicentre study of laparoscopic versus open appendicectomy. Br J Surg 86:48, 1999

98. Pedersen AG, Petersen OB, Wara P, et al: Randomized clinical trial of laparoscopic versus open appendicectomy. Br J Surg 88:200, 2001

99. Sauerland S, Lefering R, Neugebauer E: Laparoscopic versus open surgery for suspected appendicitis. Cochrane Database Syst Rev 18:CD001546, 2004

100. Borgstein PJ, Gordjin RV, Eijsbouts QA, et al: Acute appendicitis: a clear-cut case in men, a guessing game in young women: a prospective study of the role of laparoscopy. Surg Endosc 11:923, 1997

101. Larsson PG, Henriksson G, Olsson M, et al: Laparoscopy reduces unnecessary appendicectomies and improves diagnosis in fertile women: a randomized study. Surg Endosc 15:200, 2001

102. Fisher KS, Ross DS: Guidelines for therapeutic decision in incidental appendectomy. Surg Gynecol Obstet 171:95, 1990

XIII ENTERAL AND PARENTERAL NUTRITION

Khursheed N. Jeejeebhoy m.b.b.s., ph.d.

Definitions

Enteral nutrition is the process of nourishing a patient with a liquid diet of defined composition, usually given through a nasogastric, nasointestinal, gastrostomy, or jejunostomy tube. Parenteral nutrition is the administration of nutrients directly into the bloodstream through a central venous catheter or by peripheral infusion. When the only source of nutrient intake is via the parenteral route, it is called total parenteral nutrition (TPN). The term nutritional support refers to the use of enteral or parenteral nutrition rather than to an oral diet, with or without supplements.

Etiology of Malnutrition

In circumstances in which food is available, malnutrition has three main causes: (1) insufficient intake of food, as a result of conditions such as anorexia, coma, dysphagia, gastric lesions, and psychological factors; (2) heightened metabolic requirements, as may occur in burns, trauma, sepsis, and neoplasia; and (3) intestinal failure, which comprises all conditions that prevent the proper intake, digestion, or absorption of a normal oral diet. Malnutrition from reduced food intake or gastrointestinal failure is most amenable to treatment or prevention with nutritional support. Although nutritional support may overcome some of the effects of trauma, burns, sepsis, or cancer, nutritional support alone may be unable to prevent the development of critical malnutrition in such cases.

Effects of Malnutrition

Even in the absence of disease, malnutrition adversely influences function and survival. A study of Irish hunger strikers found a 30% mortality in strikers who lost 35% to 40% of their body weight.[1] Similarly, in patients with cancer, weight loss of about 30% preceded death.[2] In 12 human volunteers, semistarvation (with a 15% to 20% weight loss over 24 weeks) led to a 60% decrease in function on the basis of a fitness score.[1] Even after 20 weeks of refeeding, the fitness score and handgrip strength in these individuals did not return to normal. Other studies have shown that lack of food intake results in substantial loss of muscle function in addition to loss of body mass.[3] Surgical patients who had weight loss greater than 10% and clinical evidence of dysfunction of two or more organ systems (including skeletal and respiratory muscles) preoperatively had significantly more postoperative complications than did normal patients or those with weight loss but no physiologic dysfunction.[4]

The presence of various diseases compounds the effects of malnutrition. In ill patients, malnutrition results in nutritionally associated complications such as poor wound healing, increased infections, delayed rehabilitation, and increased mortality.

Evidence Regarding Nutritional Support

Well-nourished patients are unlikely to benefit from nutritional support. However, in patients with initial malnutrition and poor function who have continued inability to eat or to absorb ingested food, randomized controlled trials have demonstrated that nutritional support favorably influences outcome by reducing nutritionally associated complications.

parenteral nutrition

Three large meta-analyses of parenteral nutrition have given inconsistent results. In a comparison of parenteral nutrition with standard care in 26 trials, Heyland and colleagues[5] found that parenteral nutrition did not influence overall mortality but did reduce complications in malnourished patients. Benefit from TPN was observed in studies performed before 1988, in studies deemed to be of less statistical quality, and in patients who did not receive lipid. These researchers found only six trials of parenteral nutrition in critical illness; in these trials, complications and mortality were significantly higher than in trials done in surgical patients. Another meta-analysis showed that in malnourished patients, standard care, compared with parenteral nutrition, was associated with increased mortality and a trend toward increased infectious complications; in well-nourished patients, infections were more frequent with parenteral nutrition than with standard care or enteral nutrition.[6] These authors speculated that the increased infectious complications in patients on parenteral nutrition were attributable to hyperglycemia. Not all the studies included in this meta-analysis mentioned blood glucose, but of the seven that did, six found both hyperglycemia and increased infectious complications.

Koretz and colleagues[7] have done a technical review and made recommendations to the American Gastroenterological Association about parenteral nutrition. They found that overall, mortality with parenteral nutrition was no lower than mortality with standard care. In contrast to the meta-analysis by Heyland and colleagues,[5] this analysis showed that total complications and length of stay were lower only in studies in which lipid was a component of TPN. Infectious complications were increased with TPN, especially in cancer patients. Benefit from parenteral nutrition was seen only in patients with upper GI cancer, who had significantly fewer complications when given perioperative parenteral nutrition.

enteral nutrition

Enteral nutrition has not been compared with standard care in the same systematic way as has parenteral nutrition. However, comparisons of enteral nutrition with parenteral nutrition have consistently shown fewer infectious complications with enteral nutrition than with parenteral nutrition.[6] Data from a large controlled trial in intensive care unit patients showed that keeping blood glucose levels below 127 mg/dl (7 mmol/L) significantly reduced mortality from sepsis-related multisystem organ failure.[8] Hyperglycemia probably was more frequent with parenteral nutrition because patients randomized to parenteral nutrition received more calories than those on enteral nutrition,[9]

despite the intent to make both groups isocaloric. None of these studies prove that enteral nutrition is better than standard therapy; rather, they show that enteral nutrition is less likely than parenteral nutrition to cause infection. In a 562-patient trial of enteral nutrition versus TPN that mirrored the conventional practice of nutritional support, Woodcock and colleagues[10] concluded that TPN did not increase sepsis, enteral nutrition delivered less than the target nutritional intake, and procedure-related complications were greater with enteral nutrition.

Determining the Need for Nutritional Support

Unfortunately, for many clinical situations there are no data from randomized, controlled trials to help clinicians determine how to identify patients who are likely to progress to critical weight loss and to determine when to start nutritional support in patients who are at risk. In the absence of reliable data, clinicians have to make decisions about nutritional support at the bedside. Obviously, a previously healthy person who does not eat for 1 or 2 days does not need nutritional support. On the other hand, if inadequate nutritional intake persists for weeks, weight loss will continue; the loss will accelerate if there is added trauma or sepsis; and when loss of body weight exceeds 30%, there is an increased likelihood of death.

The risk of malnutrition can be assessed with a clinical tool called the Subjective Global Assessment (SGA).[11] The SGA, which can be used by physicians, dietitians, or nurses after brief training, is based on a focused history and physical examination that includes the degree and progression of any weight loss, dietary intake, ability to take and absorb food (state of the GI tract), the degree of stress from comorbidity, and functional status.[12] This information is used to classify the patient into one of three groups: A (normally nourished and unlikely to progress to a malnourished state), B (normally nourished but likely to progress to a malnourished state), or C (malnourished and progressing to increasing malnutrition).

The SGA not only provides an assessment of the patient's current nutritional status but also predicts the possible nutritional outcome if nutritional support is not instituted. More important, it allows the clinician to weigh the role of disease severity versus limited nutrient intake as the cause of malnutrition.

Two controlled studies of the SGA have shown that the likelihood of nutritionally associated complications progressively increased from grades A to C. Patients who are classified as SGA C are very likely to develop nutritionally associated complications and therefore should benefit from nutritional support. These studies also found that the SGA grade correlated with other objective measures of nutritional status but was more likely to predict nutritionally associated complications than several of the objective measures taken individually.[13] SGA has been shown to be a valid predictor of nutritionally associated complications in general surgical patients, patients on dialysis, and liver transplant patients. In two large studies, SGA independently identified increased mortality and morbidity from malnutrition, even when the data were adjusted for other factors influencing survival and complications.[14,15]

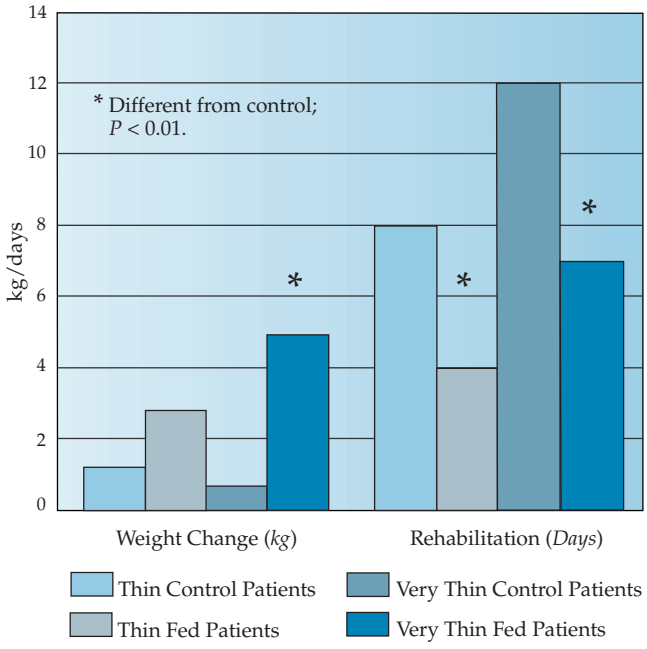


Figure 1 **In elderly women with femoral neck fractures, weight increase was greater and rehabilitation time was shorter in those who received overnight supplementary enteral feeding than in control subjects, who were given a normal hospital diet. The effect was evident in thin patients and was particularly marked in very thin patients.[22]**

Nutritional Support in Specific Clinical Conditions

INSUFFICIENT ORAL INTAKE DESPITE A NORMAL GUT

Well-nourished Patients

In general, most patients with serious illness have reduced food intake, partly from the illness itself and partly as a result of iatrogenic factors. Most hospital inpatients eat insufficient food or are prevented from eating. Several studies have indicated that a significant number of hospital patients have signs of malnutrition. Patients likely to have an inadequate intake of food are those with critical illness (e.g., trauma, burns, severe sepsis, respiratory failure); coma and neurologic diseases; or major psychiatric illnesses. Although many hospital patients fit these categories, there are no controlled trials to provide guidelines that can be confidently used to guide nutritional support in such patients and to confirm that nutritional support can reduce the occurrence of nutritionally associated complications. Clinically, it is a common practice to start nutritional support if the period of reduced intake exceeds 7 to 10 days or weight loss exceeds 10%.[16] Unfortunately, this practice has no supporting data except consensus and expert opinion.

Early enteral feeding has been recommended on the basis of a randomized trial in trauma patients who were to undergo a laparotomy and had an abdominal trauma index greater than 15.[17] This subset constituted 20% of all trauma patients admitted during the period of study. These patients, who were well nourished on admission, were randomized to a group who received early (12 to 18 hours after surgery) institution of enteral feeding through a jejunostomy tube inserted at surgery or to a control group in whom TPN was started 5 days after surgery if the patient was not yet on a regular oral diet. There was no difference in overall complications between the groups, but septic complications were significantly lower in the early-fed group (4%, versus 26% in the TPN group). Such data are subject to the criticism that there was no control group receiving standard care. However, there are other reasons to support early feeding. In a randomized trial of postoperative supplemental sip

feeding of a liquid formulation, grip strength significantly improved and the occurrence of serious infections was reduced.[18] In a trial of 501 hospitalized elderly patients randomized to oral supplements or a ward diet, Larsson and colleagues[19] showed that irrespective of their initial nutritional status, the supplemented patients had lower mortality, better mobility, and a shorter hospital stay. The difference between ward diet and supplementation was even more pronounced in a secondary analysis of patients with weight loss.

Recommendations The available data suggest that well-nourished patients who are admitted with major trauma should receive enteral nutrition. Elderly patients should receive supplemental feeding or enteral nutrition if they are incapable of eating adequately. However, all hospitalized patients should have their SGA assessed so that possible future outcome without nutritional support can be documented and considered. For example, if a previously healthy patient has a severe head injury and is likely to remain comatose (and therefore unable to eat) for an indefinite period, it is easy to predict that malnutrition will occur in the absence of nutritional support. Such a patient should be started on enteral nutrition. Similarly, major burns, the hypermetabolic state, anorexia, and ileus all result in rapid weight loss unless nutritional support is given. Each case needs to be assessed individually, however. Repeated evaluation of the SGA allows the clinician to determine any impediment to the intake of food, the presence of GI dysfunction, and progressive functional loss and weight loss, which signal the need to start nutritional support.

The purely scientific approach would be to avoid nutritional support in all situations for which proof of benefit from randomized, controlled studies is lacking; however, in patients without adequate oral intake, this approach could in some cases result in starvation and death. The pragmatic approach is to evaluate the patient, using the SGA, and start nutritional support if the clinical evidence shows that otherwise the patient is likely to progress to critical malnutrition.[20]

Malnourished Patients

There are no controlled trials to show that nutritional support will reduce complications in all patients classified as SGA C. However, there are several indirect lines of evidence suggesting that nutritional support in such patients will reduce complications and improve outcome.

A multicenter, randomized, controlled trial undertaken by the Veterans Affairs Total Parenteral Nutrition Cooperative Study Group[21] stratified patients into three nutritional groups. In the group of severely malnourished patients, the rate of major noninfectious complications was significantly lower in patients randomized to TPN than in control subjects (5.3% versus 42.9%). Overall rates of complications and infectious complications in TPN-treated patients in this trial were not different from those in control patients, however.

Other studies have shown that nutritional supplementation can significantly reduce rehabilitation time in patients with hip fractures who had severe weight loss [*see Figure 1*][22] and that elderly patients with hip fractures, especially those with weight loss, benefit the most from supplemental feeding.[19]

Recommendations Elderly patients, especially those with weight loss, should receive nutritional supplements in the hospital. Despite the lack of data from well-designed controlled trials, patients who are classified as SGA C should be given nutritional support.

SURGERY

In a meta-analysis of perioperative parenteral nutrition, Detsky and colleagues[23] combined the results of 14 randomized or quasi-randomized trials and showed that absolute morbidity was reduced by 5.2% and the relative risk reduction was 20.7%. These differences were not statistically significant, however ($P = 0.21$). Of the 14 studies, only one showed a significant reduction in complication and fatality rates with TPN. These authors concluded that perioperative TPN did not influence outcome. On the other hand, only three of the 14 trials were limited to malnourished patients (who were the most likely to benefit from TPN), so the negative result may simply reflect the fact that the trials were weighted by patients who were unlikely to benefit from nutritional support. In contrast, Twomey[24] concluded that the pooled estimate in malnourished surgical patients shows a 7.1% reduction in morbidity with TPN. In the VA trial,[21] secondary analysis showed that the severely malnourished patients had a reduction in overall morbidity from 47% to 26% with perioperative TPN.

Fan and colleagues[25] conducted a controlled trial of perioperative nutritional support in 124 patients undergoing major hepatic resection for hepatocellular carcinoma. The patients were randomized to parenteral nutrition plus oral diet or to diet only. Patients in the treatment arm received 1.5 g/kg of amino acids, of which 35% were branched-chain amino acids (BCAA), with 30 kcal/kg of a glucose-lipid mixture for energy. Medium-chain triglycerides (MCT) constituted 50% of the lipid infused. The parenteral formulation was given for 14 days. At least 20% of the patients had a preoperative weight loss of greater than 10% and therefore were likely to be malnourished, but 80% did not have weight loss. Overall morbidity, morbidity from sepsis, and diuretic use for ascites all were lower in patients who received nutritional support.

Although the benefits of parenteral nutrition in the perioperative state are controversial, randomized trials of postoperative enteral feeding have shown improved outcome. In hip fracture patients,[22] supplemental feeding of a liquid formula diet reduced recovery time. In general surgical patients,[26] the rate of infectious complications with early enteral feeding was lower than that with nil per os (NPO).

Recommendations

Postoperatively, patients who have undergone major surgery should receive supplemental liquid formula feeding. The data do not support the routine use of parenteral nutrition for perioperative nutritional support, but parenteral nutrition clearly reduces complications in patients undergoing hepatic resection. It is not clear whether standard parenteral formulations will reduce complications in patients undergoing hepatic resection or whether it is necessary to give BCAA or MCT. Patients with hip fracture and weight loss will benefit from enteral feeding. Despite the lack of proven benefit, other severely malnourished patients (i.e., those classified as SGA C) should receive perioperative nutritional support.

SERIOUS COMPROMISE OF BOWEL FUNCTION

In patients with massive small bowel resection (i.e., less than 60 cm remaining), chronic bowel obstruction, extensive bowel disease, severe radiation enteritis, or end jejunostomy in which

oral feeding results in uncontrolled fluid and electrolyte losses, parenteral nutrition is needed because oral feeding is very unlikely to provide sufficient nourishment. An economic analysis of such patients showed that provision of parenteral nutrition at home was associated with improved quality of life and was cost-effective.[27] The outlook was especially good for those with chronic intestinal failure from benign disease.[28]

Recommendations

Initially, all patients with a short bowel (see above) need parenteral nutrition. Later, about 30% (especially those with an intact or partially intact colon) can be treated with oral diet and supplements. Enteral nutrition is not necessary in these patients; controlled studies have shown that enteral nutrition was no better than an oral diet in patients with a short bowel and end jejunostomy.[29] Patients with a massive resection can absorb 50% to 60% of an oral diet.[30] By using oral rehydration solution, supplements, and a high-calorie oral diet, about 30% of such patients can reduce or stop home parenteral nutrition. The remaining patients will require supplemental fluid and electrolytes or parenteral nutrition to maintain a normal weight and electrolyte-fluid status.

BOWEL REST

Parenteral Nutrition

Bowel rest is widely used in pancreatitis, intestinal fistulas, and inflammatory bowel disease. The bowel is rested by keeping the patient NPO. Malnutrition is avoided by instituting parenteral nutrition.

Parenteral nutrition is used in pancreatitis because eating often induces pain in such cases. The only controlled trial of parenteral nutrition versus oral diet in patients with mild pancreatitis showed that TPN did not influence recovery.[31] In two trials comparing parenteral nutrition with enteral nutrition in patients with mild or acute pancreatitis, the trial of patients with mild pancreatitis[32] found no difference in septic complications, whereas the trial of patients with severe pancreatitis[33] found less sepsis with enteral nutrition. However, in the latter trial, twice the number of patients on parenteral nutrition were hyperglycemic, a factor known to increase septic complications.[8] Again, these trials do not prove that enteral nutrition is better than standard care.

Parenteral nutrition is useful in patients with intestinal fistulas, in whom eating increases output and fasting reduces output by 30% to 50%. However, there are no controlled trials comparing the effect of bowel rest plus parenteral nutrition with that of oral intake in the healing of fistulas.

In inflammatory bowel disease, bowel rest reduces abdominal discomfort and diarrhea. Controlled trials have not shown that bowel rest aids recovery in these patients, however.[34]

Recommendations Because pancreatitis, intestinal fistulas, and inflammatory bowel disease may prevent the ingestion or absorption of oral nutrients and result in malnutrition, the use of bowel rest and parenteral nutrition is a reasonable strategy in some of these cases, despite the lack of evidence that bowel rest alters the course of the disease. Specifically, enteral or parenteral nutrition should be given to prevent or treat malnutrition when a patient cannot take in or absorb nutrients for 7 to 10 days, when a patient loses nutrients because of a fistula for 7 to 10 days, or when a patient is clearly malnourished (SGA C). The route of administration selected should be capable of delivering the ideal nutrient intake successfully. For example, enteral nutrition is unlikely to be successful in a patient with a high jejunal fistula who is putting out large volumes of intestinal contents.

Enteral Nutrition in Crohn Disease

Controlled trials in Crohn disease have shown that enteral nutrition reduces the activity of the disease and, in children, promotes growth.[35] However, a recent meta-analysis of eight randomized, controlled trials of 413 patients with Crohn disease showed that enteral nutrition was not as effective as corticosteroids in inducing a remission (odds ratio of enteral nutrition/corticosteroids, 0.35; confidence interval, 0.23–0.53). In addition, there was no difference between elemental and polymeric diets in inducing clinical remission.[36] Regrettably, there are no placebo-controlled trials to show whether enteral nutrition is an effective modality for treatment of active Crohn disease.

Recommendations Enteral nutrition is not a replacement for routine drug treatment of active Crohn disease, but under certain circumstances it has definite benefits. Enteral nutrition is especially useful in promoting growth and reducing disease activity in children with growth failure. In such children, enteral nutrition can be given on a long-term basis at home, along with other treatment to promote growth.

In line with other recommendations for nutritional support, patients with active Crohn disease who are SGA C should be treated with enteral nutrition and other modalities as required. However, if they are SGA C and are unable to tolerate enteral nutrition, parenteral nutrition should be used until they can tolerate adequate nutrition by the oral route. Nutritional support is also necessary when serial SGA determinations show evidence of poor intake and the patient has severe GI symptoms and continued functional impairment that could lead to critical malnutrition. The route used depends on the capacity of the GI tract to absorb nutrients.

CANCER MALNUTRITION

Malnutrition in metastatic cancer has been used as an indication for parenteral nutrition. Controlled trials have failed to substantiate that nutritional support is beneficial in patients with metastatic cancer,[37] however, and in fact have suggested that parenteral nutrition may have adverse effects. On the other hand, parenteral nutrition has been shown to favorably influence graft survival in patients receiving a bone marrow transplant.[38]

Recommendations

In cancer patients, nutritional support with enteral or parenteral nutrition is appropriate for preventing or treating malnutrition that is not caused by the tumor per se. For example, patients whose colon cancer has been eradicated but who suffer from short bowel because of extensive radiation enteritis should respond to parenteral nutrition. Criteria for nutritional support in cancer patients are as follows: (1) there is no evidence of tumor or its progression; (2) the patient has a GI complication, such as radiation enteritis or resection; and (3) as a result of this GI complication, critical malnutrition has occurred or will predictably occur (i.e., the patient is SGA C, or serial evaluation of SGA indicates progression toward SGA C).

The most difficult ethical question concerns the use of parenteral nutrition for patients in whom tumor progression causes intestinal obstruction or cachexia. Parenteral nutrition is being increasingly used for this indication [*see* Home Parenteral Nutrition, *below*].

RENAL FAILURE

Because patients with renal disease cannot excrete nitrogen normally, parenteral nutrition in which the source of nitrogen is limited to essential amino acids (EAA) has been used to reduce urea production. A meta-analysis has concluded that parenteral nutrition with EAA does not improve survival to discharge; when the trials were adjusted for quality, there was no effect of EAA.[39]

Recommendations

Patients with renal failure who cannot meet their nutritional requirements by the oral route should be given nutritional support and have fluid, electrolytes, and nitrogenous metabolites removed by dialysis or continuous arteriovenous hemofiltration. Fluid intake is minimized by using enteral nutrition with a calorie density of 2 kcal/ml or parenteral nutrition containing 35% dextrose or 20% lipid as the source of energy. Sodium intake should be restricted to 40 to 70 mmol/day, and other electrolytes should be added if their plasma levels fall. Acidosis should be controlled by appropriate dialysis. Trace elements and vitamin supplements need not be curtailed.

HEPATIC FAILURE AND ALCOHOLIC LIVER DISEASE

The discovery that hepatic encephalopathy is associated with reduced BCAAs and increased aromatic amino acids in plasma has led to the use of parenteral nutrition formulas enriched in BCAAs and reduced in aromatic amino acids. Meta-analysis of trials comparing BCAA-enriched mixtures with standard therapy has shown significant improvement in encephalopathy and, possibly, in short-term mortality.[40] On the other hand, there is no evidence that standard amino acid mixtures or enteral nutrition providing 0.8 to 1 g/kg/day of protein or amino acids has precipitated encephalopathy. In fact, 75 g/day of supplementary amino acids with 400 kcal/day of dextrose improved liver function and was tolerated by patients with severe alcoholic hepatitis.[41]

Recommendations

Patients with hepatic failure who are unable to be on a normal diet need enteral or parenteral nutrition. The protein intake should be about 0.8 to 1 g/kg/day of a high-quality protein or balanced amino acids. Carbohydrates and fat should be given in equal proportions because these patients are carbohydrate intolerant but utilize fat well, and fat infusions increase the levels of BCAA in plasma.[42] Because these patients are sodium and water overloaded, they should receive a total of about 1,500 ml of water daily, and their sodium intake should be restricted to 20 mmol/day. Supplemental potassium, vitamins (A, D, and B complex), and zinc should be given.

Practice of Nutritional Support

GENERAL PRINCIPLES OF NUTRITIONAL CARE

At hospital admission, all patients should be interviewed by a dietitian and have their SGA calculated to determine whether they can be maintained on a normal or modified oral diet (with supplements) in sufficient quantities or whether nutritional support is indicated and, if so, how urgently. In patients requiring nutritional support, the physician and the dietitian should define nutrient intake, route of administration, and goals. The most important objective is maintenance of uninterrupted nutrient intake, to avoid weeks of starvation followed by the urgent institution of parenteral nutrition to an iatrogenically malnourished patient.

Table 1 Procedure for Nasogastric or Nasoenteral Tube Placement

1. Explain the procedure to the patient, to obtain cooperation.
2. Seat the patient comfortably at the edge of the bed, sitting upright.
3. Check nostrils for painful lesions and obstruction.
4. Insert stylet into tube and lubricate.
5. Measure approximate length of tube to be passed by the distance between the tip of the nose to the ear and down to the midepigastrium. Add about 25 cm to this distance.
6. Flex neck slightly.
7. Pass tube through an unobstructed nostril. If the patient finds this very uncomfortable, spray nostril with lidocaine 4% topical solution.
8. Ask the patient to swallow water as the tube is passed.
9. If the patient coughs or chokes, withdraw tube into the pharynx and reinsert.
10. Aspirate gastric contents to confirm position of tube.
11. Air may be injected into the tube while auscultating to determine the intragastric location of the tube.
12. For nasogastric feeding, confirm the tube position by x-ray before infusing.
13. For nasoenteral feeding, place the patient in right lateral position and gradually advance tube. Metoclopramide, 10 mg I.V., may be used to propel the tube.
14. If tube has not passed into the bowel by 24 hours, endoscopic or fluoroscopic guidance may be used.

Oral Nutrition

In patients who can eat, close attention to maintenance of oral dietary intake—and use of supplements, where required—should be the standard of care. Enteral nutrition should be considered if it becomes clear that this approach does not permit sufficient intake to meet requirements.

Enteral Nutrition

Enteral nutrition is applicable to all patients, but it should be used with caution in patients with (1) clinically significant gastroesophageal reflux; (2) intestinal obstruction; (3) GI fistula or recent surgical anastomosis, unless the tube can be inserted distal to the area in question or threaded at operation past the area; and (4) cardiovascular instability with shock. Gastric retention is a relative contraindication. In patients who accumulate secretions in the stomach and then aspirate, it may be possible to pass a feeding tube into the small intestine and aspirate the stomach with a second tube. However, in such cases the relative discomfort of two tubes versus parenteral nutrition should be considered. A recent survey showed that patients preferred parenteral nutrition over enteral nutrition.[43]

Short-term enteral access Nasogastric or nasoenteric placement of a feeding tube provides short-term enteral access. The tube should be small bore (9 to 12 French) and 105 to 110 cm long [*see Table 1*]. These tubes are usually made of Silastic or

polyurethane. The latter become very slippery when wet, thus aiding insertion. I prefer intestinal placement of the tube, because controlled trials have shown better achievement of nutrient intake[44] and, possibly, reduced risk of aspiration when the tube is placed beyond the ligament of Treitz.

Long-term feeding The definition of long-term feeding is arbitrary. Children with Crohn disease have been fed for months by teaching them to pass a nasogastric tube each night, receive a nocturnal feeding, and then remove the tube in the morning before going to school. However, in many instances nasal tubes become uncomfortable, and a gastrostomy tube can be placed endoscopically by a gastroenterologist or an interventional radiologist. This method has been shown to be safer and more cost-effective than a surgically placed gastrostomy. There are two methods of percutaneous endoscopic gastrostomy (PEG): the pull (Ponsky-Gauderer) method and the push (Russell) method.

Feeding into the small bowel can be performed after the insertion of a percutaneous endoscopic jejunostomy (PEJ). After the tract of the PEG tube is established, a PEJ tube with two arms can replace the tube. One arm remains in the stomach and can be used to drain this organ; the other arm is advanced under endoscopic guidance through the pylorus into the small intestine. In this way, the stomach can be decompressed, and simultaneously, the patient can be fed into the small bowel.

To eliminate the inconvenience of the bulky feeding tube, patients with long-term gastrostomies can be fitted with a so-called button device, which lies flush with the abdominal wall. Between feedings, a valve in the device closes off access to the stomach; during feedings, the feeding tube is inserted past the valve, permitting access to the stomach.

Parenteral Nutrition

The intravenous route is used as a supplement to oral or enteral nutrition or is used as the sole source of nutrition (TPN) when it becomes clear that the patient is not receiving sufficient nutrients by the other routes. Regular evaluation of SGA should be performed during TPN to ensure that the patient's nutrient requirements are being met.

Short-term parenteral feeding Short-term infusions are best given through a peripherally inserted central catheter (PICC). These catheters are inserted into an arm or forearm vein and advanced into the superior vena cava. PICCs are comfortable and avoid the risks of subclavian puncture or the difficulties of maintaining sterility of the exit sites of jugular catheters. In addition, full TPN with hypertonic mixtures can be given through these catheters without risk of thrombosis. Despite the designation "short term," these catheters can be used for months.

Long-term parenteral feeding Patients with intestinal failure often require parenteral feeding for years. To permit long-term parenteral feeding, an interventional radiologist advances a specially designed catheter through a subcutaneous tunnel via the jugular vein to the superior vena cava. The tip of this catheter should lie just above the right atrium, to avoid thrombotic complications. Near the exit site, within the subcutaneous tunnel, the catheter is surrounded by a Dacron cuff. Fibroblasts will grow into the cuff, sealing and anchoring the skin exit site.

NUTRIENT REQUIREMENTS

Protein

Protein requirements are met by giving whole proteins, peptides, or amino acids in enteral nutrition and by infusing an amino acid mixture in parenteral nutrition. The goal is to promote nitrogen retention and protein synthesis. Although limiting glucose and lipid (energy) intake will maximize nitrogen retention, dietary protein has an anabolic effect independent of energy intake, and will reduce nitrogen losses when infused alone.[45] Thus, the amount of amino acids given appears to be a very important determinant of nitrogen balance.

About 1 to 1.5 g/kg of ideal body weight of protein or amino acids will be sufficient for most patients with normal renal function. Additional amounts should be added for losses from prior depletion or current hypercatabolism. In patients with hepatic failure, protein intake should be restricted to 0.8 to 1.0 g/kg a day.

Glutamine

Glutamine is an amino acid released by muscle and used by immune cells and enteral cells for energy. In malnutrition and after trauma, muscle glutamine and muscle protein synthesis are reduced. The infusion of glutamine normalizes muscle glutamine and restores protein synthesis.[46] Clinically, bone marrow transplant patients were noted to have fewer episodes of sepsis and a shorter hospital stay if they received a glutamine-supplemented amino acid solution.[47] Because glutamine does not have a long shelf-life in solution, dipeptides containing glutamine have been used as a substitute. Infusion of solutions containing such dipeptides has been found to increase muscle glutamine and improve protein synthesis.[46]

Immunonutrition

Enteral formulations enriched in arginine, omega-3 fatty acids, and glutamine nucleotides are considered to enhance the immune response; treatment with these formulations is referred to as immunonutrition. These formulations vary in composition, but they are distinguished by high (12 to 15 g/L) or low (4 to 6 g/L) arginine content, presence or absence of glutamine and nucleotides, and different concentrations of omega-3 fatty acids. A recent summit on immune-enhancing enteral therapy[48] concluded, on the basis of published literature, that immunonutrition should be given to malnourished patients undergoing elective GI surgery and to trauma patients with an injury severity score of 18 or greater or an abdominal trauma index of 20 or greater. Immunonutrition was also recommended, despite lack of evidence, in patients undergoing head and neck surgery or aortic reconstruction, as well as in patients with severe head injury or burns, and in ventilator-dependent nonseptic patients. It was not recommended for patients with splanchnic hypoperfusion or bowel obstruction distal to the access site or after major upper GI hemorrhage.

A systematic review of immunonutrition by Heyland and colleagues[49] showed that it reduced septic complications but did not reduce mortality. Their analysis of 22 randomized, controlled trials covering 2,419 critically ill or surgical patients indicated that only high-arginine formulations reduced infectious complications and length of stay. These authors concluded that in patients undergoing elective surgery, immunonutrition may reduce complications and reduce length of stay. Pending further studies, however, immunonutrition was not recommend-

ed in patients with critical illness. Because many trauma and septic patients may be critically ill, these authors' recommendations are at variance with those of the immunonutrition summit (see above). The finding that benefit is seen only with the formulation containing higher amounts of arginine raises the question whether arginine per se or the higher nitrogen intake is responsible for the benefit.

Energy (Glucose and Lipids)

In healthy persons, basal energy expenditure (BEE), or basal metabolic rate (BMR), in kilocalories a day can be predicted with the Harris-Benedict equation:

$$\text{BEE in males} = 66.5 + (13.8 \times \text{weight in kg}) + (5.0 \times \text{height in cm}) - (6.8 \times \text{age in yr})$$

$$\text{BEE in females} = 655.1 + (9.6 \times \text{weight in kg}) + (1.8 \times \text{height in cm}) - (4.7 \times \text{age in yr})$$

A calculator for determining BEE according to the Harris-Benedict equation can be found on the Internet, at www-users.med.cornell.edu/~spon/picu/calc/beecalc.htm.

For patients substantially on bed rest, about 30% should be added to the BEE to meet their metabolic requirements. In practice, this calculates as a daily expenditure of about 31 kcal/kg. An expert group has suggested a daily intake of 25 kcal/kg in ICU patients.[50] Therefore, 25 to 30 kcal/kg/day will meet the needs of most patients, except those with burns. Malnutrition reduces the expected BEE by as much as 35%; injury, sepsis, and, especially, burns increase requirements.[51] Baker and colleagues[52] found that in critically sick patients in respiratory failure, the maximal degree of hypermetabolism was about 30%.

Energy requirements during TPN can be met by infusing glucose or lipid emulsions. These nonprotein energy sources enhance nitrogen retention. The most striking increase in nitrogen balance has been found to occur when energy was increased from 0 kcal/kg to 30 kcal/kg of ideal body weight. Increases above that provided only slight improvement. In obese persons, a high-protein formulation with only about 14 kcal/kg/day meets nitrogen requirements[53] and is associated with satisfactory wound healing.[54]

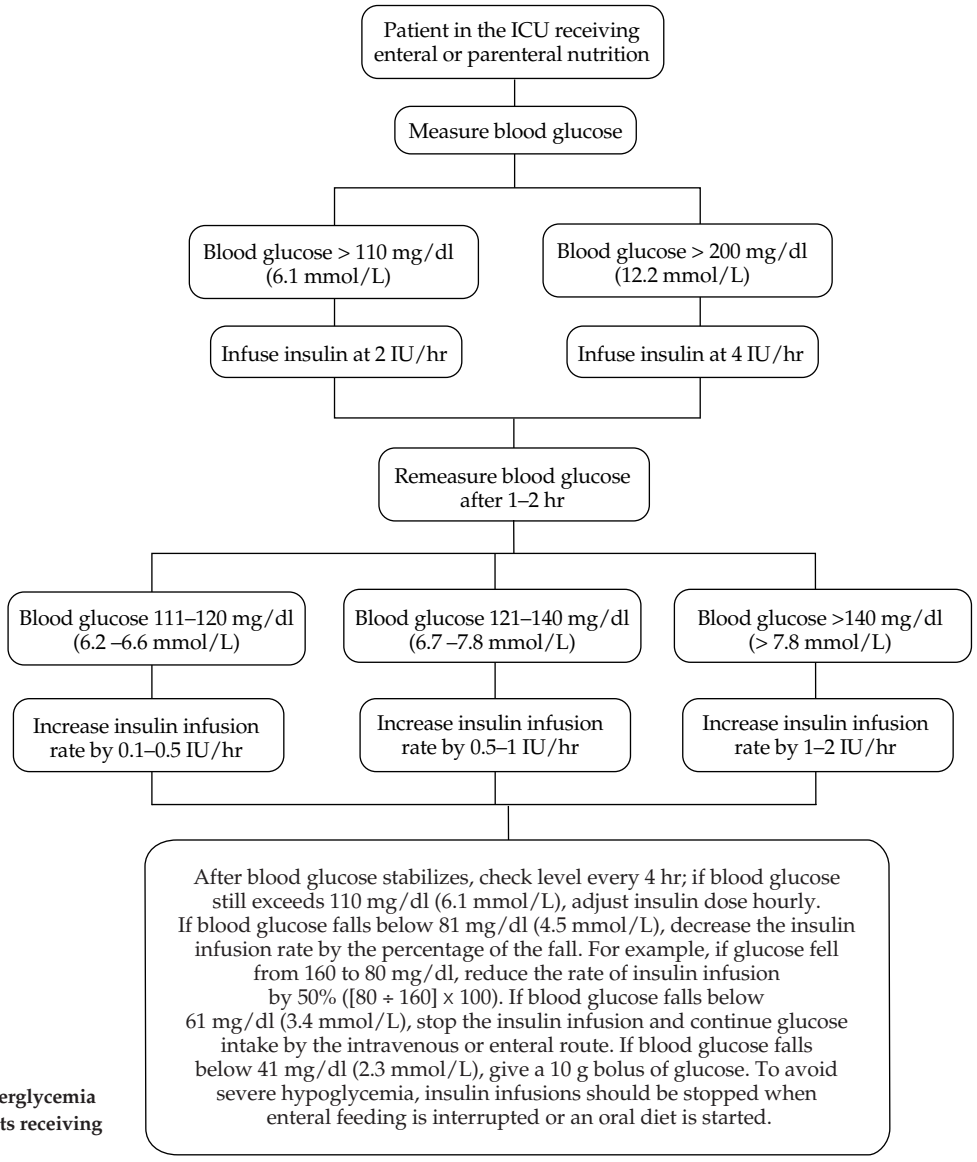


Figure 2 **Controlling hyperglycemia in intensive care unit patients receiving nutritional support.**

Table 2 Daily Electrolyte and Trace Element Requirements for Adults on Total Parenteral Nutrition

Element	*Normal*	*Increased GI Losses*	*Renal Failure*	*Comments*
Sodium (mmol)	80–120	Meet losses	20–40	Reduce in heart failure
Potassium (mmol)	40–-80	80–120	0–20	Correct hypokalemia before starting nutrition
Magnesium (mmol)	5–10	10–20	0–5	Correct hypomagnesemia before starting nutrition
Phosphorus (mmol)	10–15	10–15	0–5	Risk of dangerously low serum levels when feeding patients with severe malnutrition
Zinc (mg)	TPN: 3–4 Enteral: 15–20	TPN: 12–25 Enteral: 50–100	No change	—
Copper (mg)	TPN: 0.25–0.3 Enteral: 2–4	TPN: 0.5–0.7 Enteral: 4–8	No change	Reduce to 0.1 in hepatic failure

Because glucose spares nitrogen in fasting persons, it has been advocated as the main source of energy for parenteral nutrition. However, recent studies have shown that in malnourished patients and septic patients, lipids can promote nitrogen retention and increase total body nitrogen to the same extent as glucose, provided amino acids are given.[55] Fats constitute about 30% of total energy in most enteral formulas. Furthermore, glucose-lipid mixtures facilitate the control of severe hyperglycemia in septic patients with insulin resistance.[50]

Infusion of glucose at rates that exceed energy requirements elevates O_2 consumption, CO_2 production, resting energy expenditure, and urinary norepinephrine excretion. However, the magnitude of increased CO_2 production is small if total calories infused conform to levels recommended for the patient's clinical situation.[51]

The exact amount of lipid to include in the parenteral nutrition regimen is controversial. In a randomized, controlled trial of 512 bone marrow transplant patients receiving TPN, sepsis was no more frequent in patients who received 30% of energy as lipid than in those who received only sufficient lipids to meet essential fatty acid (EFA) needs (6% to 8% of energy intake).[56] In addition, EFA deficiency developed in some of the latter patients, and in some, this small amount of lipid was insufficient to meet energy requirements without induction of hyperglycemia from the glucose component. These authors recommend giving 25% to 30% of energy as long-chain triglycerides (LCTs). In contrast, a study in 57 trauma patients found that TPN with added lipid increased sepsis and hospital stay.[57] It was not clear whether the adverse effect was from the lipid per se or the increased energy intake while on lipid.

Because of their glucose content, both enteral nutrition and TPN enhance the risk of sepsis if the blood glucose level is allowed to rise above 127 mg/dl (7 mmol/L).[8] Therefore, insulin should be infused in patients receiving nutritional support to keep them as close to normoglycemia as possible [*see Figure 2*].

Whereas the major concern with glucose-based formulations is hyperglycemia, the key concern with lipid emulsions is hypertriglyceridemia, which may induce pancreatitis. Lipid particles also reduce gas diffusion in the lungs and inhibit the reticuloendothelial system. Provided that lipid emulsions are infused continuously at a rate that does not exceed 110 mg/kg/hr, hypertriglyceridemia does not occur. When these principles are followed, 30% to 50% of nonprotein calories can be given as fat, especially in glucose-intolerant patients.

Electrolytes, Trace Elements, and Vitamins

In patients receiving nutritional support, levels of electrolytes and trace elements should be adjusted to fit the clinical circumstances [*see Table 2*]. Carbohydrate feeding induces sodium retention, resulting in refeeding edema. In malnourished patients, great care should be taken to prevent salt and water overload.

Body potassium is disproportionately reduced relative to nitrogen in malnourished patients. Positive nitrogen balance does not occur unless potassium, phosphorus, and magnesium are given.[58,59] During enteral and parenteral nutrition, serum phosphorus may drop precipitously and cause dangerous neurologic symptoms.[60]

Micronutrients comprise vitamins and trace elements. The former are complex organic compounds; the latter are inorganic elements. Trace elements important to nutritional support include zinc, copper, chromium, and selenium. Diarrhea increases zinc requirements markedly and copper requirements modestly [*see Table 2*]. Oral chromium requirements have not been precisely determined, but deficiency occurs in patients receiving TPN; in one of my patients, the daily chromium needs were increased to 10 to 20 μg. Patients receiving parenteral nutrition may develop selenium deficiency, with muscle pains and cardiomyopathy. Increased losses of selenium can occur from the GI tract and from wounds. The recommended dose of selenium for stable patients is 40 μg/day. Patients depleted of selenium may require as much a 120 μg/day to regain normal levels.

The current recommendations for vitamins [*see Table 3*] specify the amounts required to maintain normal plasma or blood levels in patients on long-term home parenteral nutrition. There are no clearly defined recommendations for critically sick or septic patients.

HOME PARENTERAL NUTRITION

Patients with intestinal failure from a short bowel, chronic bowel obstruction, radiation enteritis, or untreatable malabsorption can be nourished by parenteral nutrition given at home. Arteriovenous shunts were initially used for long-term venous access in these patients, but success was limited because of clotting or disruption of the shunt. Long-term success has been achieved with a tunneled silicone rubber catheter or an implanted reservoir. Premixed nutrients are infused overnight. The catheter is then disconnected and a heparin lock applied, leaving the patient free to attend to daily activities. We have used home parenteral nutrition for more than 20 years in

Table 3 Recommendations for Vitamins in Adults on Total Parenteral Nutrition

Vitamin	*Recommended Daily Dose*
A	3,300 IU
D_2	200 IU
E	10 IU
K_1	150 mg
Ascorbate	200 mg
Thiamin	6 mg
Riboflavin	3.6 mg
Pyridoxine	6 mg
Niacin	40 mg
Pantothenate	15 mg
Biotin	60 μg
Folate	600 μg
Cobalamin	5 μg

two patients with total jejunoileal resection; one continues to receive it after 30 years. Survival of patients with short bowel from treatment for Crohn disease or pseudo-obstruction is excellent. Home parenteral nutrition increases quality-adjusted years of life in these patients and is cost-effective. On the other hand, mean survival in AIDS patients or those with metastatic cancer who receive home parenteral nutrition is about 3 months. There is no evidence that home parenteral nutrition prolongs their survival or enhances their quality of life. Trials are urgently required to justify the use of home parenteral nutrition in terminal cancer and AIDS.

Complications of Long-term Home Parenteral Nutrition

At the start of nutritional support, patients are vulnerable to complications related to venous and enteral access and to metabolic complications. Careful and frequent monitoring and adjusting of nutrient intake will prevent these complications. Over the longer term, patients receiving TPN are vulnerable to three organ-specific complications: hepatic disease, bone disease, and gallstones.

Hepatic disease The most serious form of hepatic disease related to TPN is chronic cholestasis with fibrosis. This condition is most common in patients with a very short bowel. The exact cause is unknown, but absorption of endotoxin or alteration in bile salts by bacterial dehydroxylation are possible factors. Successful treatment with metronidazole and with ursodeoxycholic acid has been reported. In some patients, carnitine infusions have corrected cholestasis.

Bone disease Bone loss during long-term TPN is a complex issue. In a prospective longitudinal study, patients were noted to have a high bone turnover before the institution of home parenteral nutrition, but during TPN this changed to osteomalacia and slow bone turnover. This process has been attributed to aluminum toxicity but occurs in its absence[61] and seems to respond to withdrawal of vitamin D from the TPN formula. In a prospective 4-year study of patients on home parenteral nutrition, withdrawal of vitamin D increased spinal bone mass.[62] On the other hand, patients on home parenteral nutrition can lose bone mass as a result of factors such as active inflammatory bowel disease, corticosteroid therapy, and inactivity. Some clinicians are treating reduced bone mineral density in these patients with intravenous bisphosphonates such as pamidronate and clodronate (the latter is not available in the United States). Although there are no controlled trials of bisphosphonates in patients receiving home parenteral nutrition, there are anecdotal reports of improvement of bone mass with this therapy.

Gallstones The short bowel state results in bile salt deficiency and increased biliary cholesterol secretion. In addition, sludge composed of bilirubin and calcium forms in the gallbladder. Consequently, the incidence of gallstones is high in these patients. These stones are mixed cholesterol and pigment.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Allison SP: The uses and limitations of nutritional support. Clin Nutr 11:319, 1992

2. DeWys WD, Begg D, Lavin PT: Prognostic effect of weight loss prior to chemotherapy in cancer patients. Am J Med 69:491, 1980

3. Jeejeebhoy KN: Rhoads lecture—1988. Bulk or bounce—the object of nutritional support. JPEN J Parenter Enteral Nutr 12:539, 1988

4. Windsor JA, Hill GL: Weight loss with physiologic impairment: a basic indicator of surgical risk. Ann Surg 207:290, 1988

5. Heyland DK, MacDonald S, Keefe L, et al: Total parenteral nutrition in the critically ill patient: a meta-analysis. JAMA 280:2013, 1998

6. Braunschweig CL, Levy P, Sheean PM, et al: Enteral compared with parenteral nutrition: a meta-analysis. Am J Clin Nutr 74:534, 2001

7. Koretz RL, Lipman TO, Klein S: AGA Technical Review on Parenteral Nutrition. American Gastroenterological Association. Gastroenterology 121:970, 2001

8. van den Berghe G, Wouters P, Weekers F, et al: Intensive insulin therapy in the critically ill patient. N Engl J Med 345:1359, 2001

9. Jeejeebhoy KN: TPN: potion or poison. Am J Clin Nutr 74:160, 2001

10. Woodcock NP, Zeigler D, Palmer MD, et al: Enteral versus parenteral nutrition: a pragmatic study. Nutrition 17:1, 2000

11. Baker J, Detsky AS, Wesson DE, et al: Nutritional assessment: a comparison of clinical judgment and objective measurements. N Engl J Med 306: 969, 1982

12. Detsky AS, McLaughlin JR, Baker JP, et al: What is subjective global assessment of nutritional status? JPEN J Parenter Enteral Nutr 11:8, 1987

13. Detsky AS, Baker JP, O'Rourke K, et al: Predicting nutrition-associated complications for patients undergoing gastrointestinal surgery. JPEN J Parenter Enteral Nutr 11:440, 1987

14. Perman M, Crivelli A, Khoury M: Nutritional prognosis in hospitalized patients. Am J Clin Nutr 75:426S, 2002

15. Pirlich M, Schütz T, Gastell S, et al: Malnutrition affects long-term prognosis in hospitalized patients. Gastroenterology 122(suppl):A636, 2002

16. Guidelines for the use of parenteral and enteral nutrition in adult and pediatric patients. JPEN J Parenter Enteral Nutr 26(1 suppl):1SA, 2002

17. Moore EE, Jones TN: Benefits of immediate jejunostomy feeding after major abdominal trauma: a prospective, randomized study. J Trauma 26:874, 1986

18. Rana SK, Bray J, Menzis-Gow N, et al: Short term benefits of post-operative oral dietary supplements in surgical patients. Clin Nutr 11:337, 1992

19. Larsson J, Unosson M, Ek AC, et al: Effect of dietary supplement on nutritional status and clinical outcome in 501 geriatric patients: a randomized study. Clin Nutr 9:179, 1990

20. Detsky AS: Parenteral nutrition: is it helpful? N Engl J Med 325:573, 1991

21. Perioperative total parenteral nutrition in surgical patients. The Veterans Affairs Total Parenteral Nutrition Cooperative Study Group. N Engl J Med 325:525, 1991

22. Bastow MD, Rawlings J, Allison SP: Benefits of supplementary tube feeding after fractured neck of femur: a randomised controlled trial. BMJ 287:1589, 1983

23. Detsky AS, Baker J, O'Rourke K, et al: Perioperative parenteral nutrition: a meta-analysis. Ann Intern Med 107:195, 1987

24. Twomey PL: Cost-effectiveness of total parenteral nutrition. Clinical Nutrition, Parenteral Nutrition. Rombeau JL, Caldwell MD, Eds. WB Saunders Co, Philadelphia, 1993, p 401

25. Fan S-T, Lo C-M, Lai ECS, et al: Perioperative nutritional support in patients undergoing hepatectomy for hepatocellular carcinoma. N Engl J Med 331:1547, 1994

26. Lewis SJ, Egger M, Sylvester PA, et al: Early enteral feeding versus "nil by mouth" after gastrointestinal surgery: systematic review and meta-analysis of controlled trials. BMJ 323:773, 2001

27. Detsky AS, McLaughlin JR, Abrams HB, et al: A cost-utility analysis of the home parenteral nutrition program at Toronto General Hospital: 1970–1982. JPEN J Parenter Enteral Nutr 10:49, 1986

28. Messing B, Lemann M, Landis P, et al: Prognosis of patients with nonmalignant chronic intestinal failure receiving long-term home parenteral nutrition. Gastroenterology 108:1005, 1995

29. McIntyre PB, Fitchew M, Lennard-Jones JE: Patients with a high jejunostomy do not need a special diet. Gastroenterology 91:25, 1986

30. Woolf GM, Miller C, Kurian V, et al: Diet for patients with a short bowel: high fat or high carbohydrate? Gastroenterology 84:823, 1983

31. Sax HC, Warner BW, Talamini MA, et al: Early total parenteral nutrition in acute pancreatitis: lack of beneficial effects. Am J Surg 153:117, 1987

32. McClave SA, Greene LM, Snider HL, et al: Comparison of the safety of early enteral vs parenteral nutrition in mild acute pancreatitis. JPEN J Parenter Enteral Nutr 21:14, 1997

33. Kalfarentzos F, Kehagias J, Mead N, et al: Enteral nutrition is superior to parenteral nutrition in severe acute pancreatitis: results of a randomized trial. Br J Surg 84:1665, 1997

34. Greenberg GR, Fleming CR, Jeejeebhoy KN, et al: Controlled trial of bowel rest and nutritional support in the management of Crohn's disease. Gut 29:1309, 1988

35. Polk DB, Hattner JAT, Kerner JA: Improved growth and disease activity after intermittent administration of a defined formula diet in children with Crohn's disease. JPEN J Parenter Enteral Nutr 16:499, 1992

36. Griffiths AM, Ohlsson A, Sherman PM, et al: Meta-analysis of enteral nutrition as primary treatment of active Crohn's disease. Gastroenterology 108:1056, 1995

37. McGeer AJ, Detsky AS, O'Rourke K: Parenteral nutrition in cancer patients undergoing chemotherapy: a meta-analysis. Nutrition 6:233, 1990

38. Weisdorf SA, Lysne J, Wind D, et al: Positive effect of prophylactic total parenteral nutrition on long-term outcome of bone marrow transplantation. Transplantation 43:833, 1987

39. Naylor CD, Detsky AS, O'Rourke K, et al: Does treatment with essential amino acids and hypertonic glucose improve survival in acute renal failure? A meta-analysis. Ren Fail 10:141, 1987

40. Naylor CD, O'Rourke K, Detsky AS, et al: Parenteral nutrition with branched-chain amino acids in hepatic encephalopathy: a meta-analysis. Gastroenterology 97:1033, 1989

41. Bonkovsky HL, Fiellin DA, Smith GS, et al: A randomized, controlled trial of treatment of alcoholic hepatitis with parenteral nutrition and oxandrolone. I: Short-term effects on liver function. Am J Gastroenterol 86:1200, 1991

42. Glynn MJ, Powell-Tuck J, Reaveley DA: High lipid parenteral nutrition improves portosystemic encephalopathy. JPEN J Parenter Enteral Nutr 12:457, 1988

43. Scolapio JS, Picco MF, Tarrosa VB: Enteral versus parenteral nutrition: the patient's preference. JPEN J Parenter Enteral Nutr 26:248, 2002

44. Montecalvo MA, Steger KA, Farber HW, et al: Nutritional outcome and pneumonia in critical care patients randomized to gastric versus jejunal tube feedings. Crit Care Med 20:1377, 1992

45. Greenberg GR, Marliss EB, Anderson GH, et al: Protein-sparing therapy in postoperative patients: effects of added hypocaloric glucose or lipid. N Engl J Med 294:1411, 1976

46. Hammarqvist F, Wernerman J, von der Decken A, et al: Alanyl-glutamine counteracts the depletion of free glutamine and the postoperative decline in protein synthesis in skeletal muscle. Ann Surg 212:637, 1990

47. Ziegler TR: Glutamine supplementation in bone marrow transplantation. Br J Nutr 87(suppl 1):S9, 2002

48. Proceedings from Summit on Immune-Enhancing Enteral Therapy. May 25-26, 2000, San Diego, California, USA. JPEN J Parenter Enteral Nutr 25(2 suppl):S1, 2001

49. Heyland DK, Novak F, Drover JW, et al: Should immunonutrition become routine in critically ill patients? JAMA 286:944, 2001

50. Applied nutrition in ICU patients. Consensus statement of the American College of Chest Physicians. Chest 111:769, 1997

51. Allard JP, Pichard C, Hoshino E, et al: Validation of a new formula for calculating the energy requirements of burn patients. JPEN J Parenter Enteral Nutr 14:115, 1990

52. Baker JP, Detsky AS, Stewart S, et al: Randomized trial of total parenteral nutrition in critically ill patients: metabolic effects of varying glucose-lipid ratios as the energy source. Gastroenterology 87:53, 1984

53. Choban PS, Burge JC, Scales D, et al: Hypoenergetic nutrition support in hospitalized obese patients: a simplified method for clinical application. Am J Clin Nutr 66:546, 1997

54. Dickerson RN, Rosato EF, Mullen JL: Net protein anabolism with hypocaloric parenteral nutrition in obese stressed patients. Am J Clin Nutr 44:747, 1986

55. MacFie J, Smith RC, Hill GL: Glucose or fat as a non-protein energy source? A controlled clinical trial in gastroenterological patients requiring intravenous nutrition. Gastroenterology 80:103, 1981

56. Lenssen P, Bruemmer BA, Bowden RA, et al: Intravenous lipid dose and incidence of bacteremia and fungemia in patients undergoing marrow transplantation. Am J Clin Nutr 67:927, 1998

57. Battistella FD, Widergren JT, Anderson JT, et al: A prospective, randomized trial of intravenous fat emulsion administration in trauma victims requiring total parenteral nutrition. J Trauma 43:52, 1997

58. Rudman D, Millikan WJ, Richardson TJ, et al: Elemental balances during intravenous hyperalimentation of underweight adult subjects. J Clin Invest 55:94, 1975

59. Freeman JB, Wittime MF, Stegink LD, et al: Effects of magnesium infusions on magnesium and nitrogen balance during parenteral nutrition. Can J Surg 25:570, 1982

60. Silvis SE, DiBartolomeo AG, Aaker HM: Hypophosphatemia and neurological changes secondary to oral caloric intake: a variant of hyperalimentation syndrome. Am J Gastroenterol 73:215, 1980

61. Karton MA, Rettmer R, Lipkin EW, et al: D-Lactate and metabolic bone disease in patients receiving long-term parenteral nutrition. JPEN J Parenter Enteral Nutr 13:132, 1989

62. Verhage AH, Cheong WK, Allard JP, et al: Vars Research Award. Increase in lumbar spine bone mineral content in patients on long-term parenteral nutrition without vitamin D supplementation. JPEN J Parenter Enteral Nutr 19:431, 1995

XIV GASTROINTESTINAL MOTILITY DISORDERS

Michael Camilleri, m.d.

Motility disorders of the stomach, small intestine, and colon are characterized by the acute, recurrent, or chronic presentation of symptoms of stasis or rapid transit in the absence of mucosal disease or any obstruction within the lumen of the gut.[1]

The most common syndromes associated with disorders of motility are nonulcer dyspepsia,[2] irritable bowel syndrome (IBS), functional constipation,[3] and outlet obstruction to defecation (evacuation disorders)[4]; the prevalence of gastroparesis and chronic intestinal pseudo-obstruction[1] is far lower. These disorders result from impaired neurologic or muscular control of the gut or from incoordination of defecation dynamics. Motility disorders are sometimes caused by a process that influences the extrinsic autonomic nerves that supply the gut. Other disorders infiltrate the GI smooth muscle and extraintestinal organs, particularly the urinary bladder.[1]

Physiology

GASTRIC, SMALL BOWEL, AND COLONIC MOTILITY

GI motor functions are characterized by distinct patterns of contractile activity in the fasting and postprandial periods. The fasting period is characterized by a cyclic motor phenomenon called the interdigestive migrating motor complex [*see Figure 1*]. In healthy people, it occurs approximately once every 60 to 90 minutes and comprises a period of quiescence (phase I), a period of intermittent pressure activity (phase II), and an activity front, during which the stomach and small intestine contract at their highest frequency (phase III). These contraction frequencies reach three a minute in the stomach and 11 or 12 a minute in the proximal small intestine. The interdigestive activity front migrates a variable distance down the small intestine; there is a gradient in the frequency of contractions during phase III, from 11 or 12 a minute in the duodenum to as low as five a minute in the ileum. The distal small intestine also demonstrates another characteristic motor pattern—a propagated prolonged contraction, or power contraction, that serves to empty residue from the ileum to the colon in bolus transfers.

In the postprandial period, the fasting cyclic activity of the stomach and small intestine is replaced by irregular, fairly frequent contractions in those regions of the stomach and small bowel that come in contact with food [*see Figure 1*]. The caloric content of the meal is the major determinant of the duration of this so-called fed pattern. The maximum frequency of contractions is below that noted during phase III of the interdigestive migrating motor complex. After meals, segments of the small intestine that are not exposed to digesta may still show the interdigestive complex. Thus, there may be simultaneous patterns of interdigestive activity in the distal small bowel at a time when the proximal small bowel is in contact with intraluminal digesta and is responding with the irregular contractile activity that characterizes the fed pattern.

Solid and liquid food empty from the stomach at different rates. Liquids empty from the stomach in an exponential manner. For nonnutrient liquids, the healthy stomach tends to empty liquids with a half-emptying time of 20 minutes or less. On the other hand, solids are initially retained selectively within the stomach until particles have been triturated to a size smaller than 2 mm, at which point they can be emptied in a linear fashion from the stomach. Thus, the gastric emptying of solids consists of an initial lag period, followed by a more linear postlag gastric-emptying phase. The small intestine transports solids and liq-

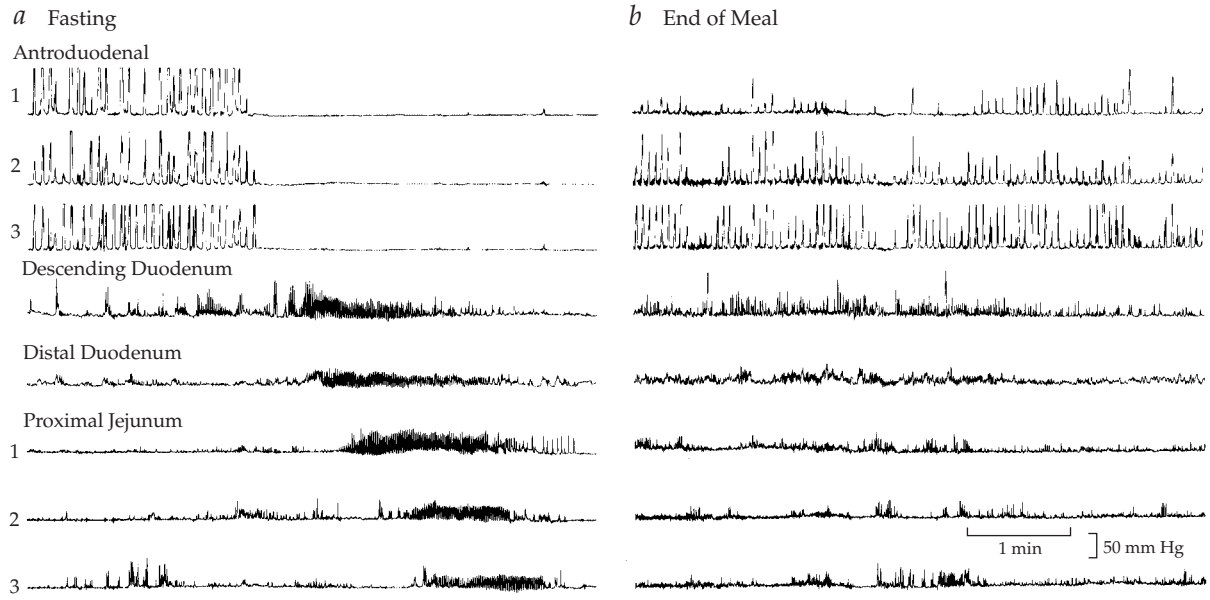


Figure 1 **Normal gastrointestinal motility. Note the normal interdigestive migrating motor complex during the fasting phase (*a*) and the irregular but persistent antral and intestinal phasic pressure activity in the fed phase (*b*).**[57]

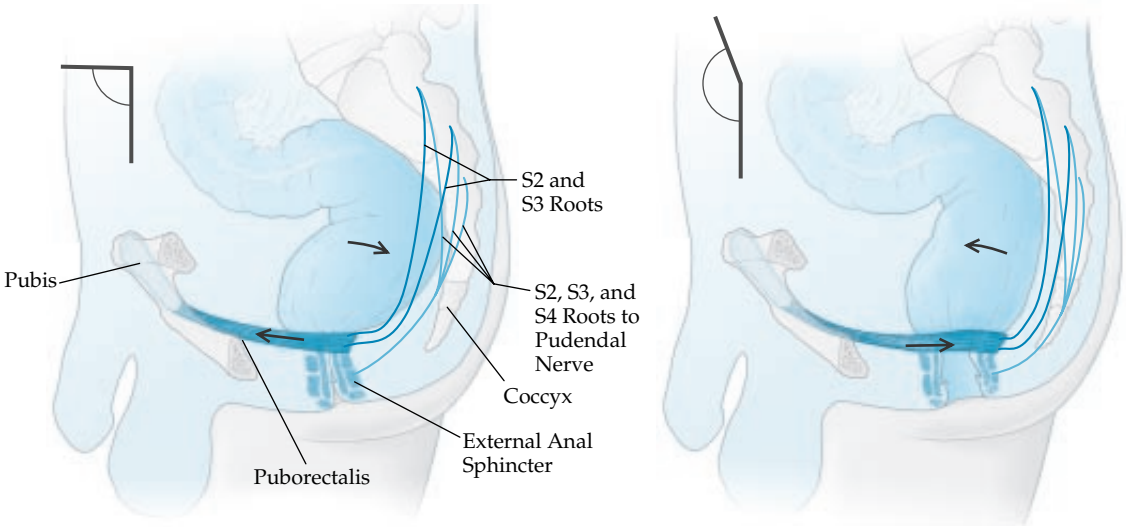


Figure 2 **Normal defecation requires relaxation of the puborectalis and external anal sphincter, straightening of the rectoanal angle, and an increase in intraluminal pressure, usually induced by the Valsalva maneuver. Defecation may be obstructed if any of these functions is impaired.[58]**

uids at approximately the same rate. Because there is a separation of the two phases in the stomach, it is clear that liquids may arrive in the colon before the head of the solid phase of the meal. Ileal emptying of chyme is characterized by bolus transfers.

Another important function of the stomach is the initial relaxation response that occurs after the ingestion of food. This response, also called accommodation, is mediated by the vagus nerves and involves the activation of intrinsic nitrergic inhibitory nerves in the wall of the stomach. Failure of gastric accommodation results in symptoms such as early fullness, satiety, and bloating and may contribute to nausea, indigestion (dyspepsia), or discomfort postprandially.[5,6]

The colon serves as a reservoir to facilitate absorption of fluids, electrolytes, and short-chain fatty acids produced by bacterial metabolism of unabsorbed carbohydrates. This reservoir function is centered predominantly in the ascending and transverse colonic regions. The descending colon functions as a conduit for the relatively rapid transit of feces to the sigmoid colon, which acts as a second reservoir. Emptying of the sigmoid colon is largely under volitional control. The defecatory process requires the Valsalva maneuver to raise intra-abdominal pressure, which is transmitted to the rectal contents, and relaxation of the puborectalis (or pelvic floor) and external anal sphincter,[4] which necessitates a coordinated series of functions [*see Figure 2*]. This facilitates the opening or straightening of the rectoanal angle and expulsion of stool. The control and function of contractions in the colon are not fully understood; some irregular contractions serve to mix its contents, whereas high-amplitude propagated contractions (HAPCs), which occur on average four to six times a day, are sometimes associated with mass movement of colonic residue and lead to defecation. After meals of at least 500 kcal, there is a greater propensity for HAPCs to develop and for the tone (background state of contractility) of the colon to increase and lead to bowel movements in the first 2 hours after meals.[7]

CONTROL OF GI MOTOR AND SENSORY FUNCTIONS

Motor function of the gut depends on the contraction of the smooth muscle cells and their integration and modulation by enteric and extrinsic nerves.[1] Smooth muscle contraction results from fluxes of ions that alter the electrical potential of the cell membrane. The enteric nervous system—approximately 100 million (10^8) neurons organized in ganglionated plexi (submucosal and myenteric being the predominant plexi)—is organized in intricate excitatory and inhibitory programmed circuits [*see Figure 3*]. These circuits play essential roles in controlling peristalsis and the migrating motor complex. Among the enteric plexi, there are also interstitial cells of Cajal, which are thought to serve as pacemakers. Enteric nerves are also important in mediating sensation from the gut. Visceral sensation, as with somatic sensation, is mediated by A-delta fibers (which respond to short, sharp stimuli) and polymodal C-unmyelinated fibers (which tend to respond to more prolonged stimuli). The latter nerves mediate pain as well as the autonomic and emotional responses that are commonly noted in patients with functional GI diseases.[8]

Extrinsic neural control is subdivided into the craniosacral parasympathetic outflow and the thoracolumbar sympathetic supply. The cranial outflow is predominantly through the vagus nerve, which supplies neural control from the stomach down to the right side of the colon, and the sacral parasympathetic supply, which provides neural control to the left colon and, through ascending intracolonic fibers, the more proximal regions of the colon. Parasympathetic supply is excitatory to nonsphincteric muscle. Sympathetic fibers to the GI tract arise from levels T5 to L1 of the intermediolateral column of the spinal cord. Sympathetic fibers are stimulatory to sphincteric muscle and relaxatory to nonsphincteric muscle. The prevertebral sympathetic ganglia integrate afferent impulses from the gut and sympathetic supply from the central nervous system. Derangement of any of these intrinsic or extrinsic control mechanisms may lead to altered gut motor function [*see* Gastroparesis and Chronic Intestinal Pseudo-obstruction, *below*].[1]

Structural Diseases and Their Effects on GI Motility

Disturbances of gastric and proximal small bowel motility are frequently observed in symptomatic patients after gastric surgery. Uncoordinated phasic pressure waves occur in the Roux limb after Roux-en-Y partial gastrectomy.[9] In these pa-

tients, the vagotomized gastric remnant may also contribute to the development of symptoms, because relaxation and contraction of the gastric remnant are deranged after vagotomy and partial gastric resection. In practice, pharmacologic agents are generally ineffective in this situation; further resection of the gastric remnant may relieve the symptoms resulting from upper gut stasis in about two thirds of these patients.[10]

Another frequently encountered postoperative disorder is the postfundoplication syndrome. An excessively tight repair of hiatal hernia may result in dysphagia; the increase in the use of laparoscopic fundoplication has led to a greater appreciation of the frequency with which this procedure results in postprandial upper abdominal pain, gas, bloating, and a dyspeptic condition that seems further aggravated by the patient's inability to belch as a result of the effective wrap.[11]

In subacute mechanical obstruction, proximal small bowel manometry shows simultaneous prolonged contractions separated by periods of quiescence. This pattern was shown to have a positive predictive value of 80%,[12] and patients showing this pattern should undergo further careful assessment with enteroclysis, laparoscopy, or laparotomy to exclude obstruction. The increased availability and experience of laparoscopic surgeons have led to less need for motility tests to diagnose mechanical obstruction.

Small bowel fistulas, diverticula, and postsurgical blind loops are all associated with bacterial overgrowth, but the pathogenic sequence is not always clear. Experimentally, bacterial toxins induce migrating action potential complexes in the rabbit ileum, as well as abnormal motility, rapid transit through the small bowel, diarrhea, and steatorrhea. It has also been suggested that multiple jejunal diverticula may result from abnormal neuromuscular function.[13]

Volvulus of the stomach, small bowel loops, cecum, and sigmoid colon may present as acute or subacute symptoms caused by mechanical obstruction. These need to be differentiated from conditions that primarily affect the motor apparatus, such as gastroparesis and pseudo-obstruction.

Functional Gastrointestinal Disorders

Functional GI disorders are characterized by disturbances of motor or sensory functions in the absence of mucosal or structural abnormality or of known biochemical or metabolic disorders. These syndromes affect one or more regions of the GI tract and include functional dysphagia, nonulcer dyspepsia,[2] IBS,[3] slow-transit constipation,[4] and outlet obstruction to defecation[4] (also termed evacuation disorders).

Functional GI disorders share common pathogenetic features, including abnormal motility, heightened visceral sensation, and psychosocial disturbance.[2,3] In some patients, these syndromes are preceded by an episode of gastroenteritis. The abnormal motility may be characterized by rapid or slow transit of food or residue through the bowel or abnormal gastric relaxation to accommodate the meal.[5,6] Abnormal contractile patterns have been described, but more important, patients perceive a sensation of excessive gut contractions.[8] In patients with these conditions, there is frequently evidence of psychological comorbidity, such as anxiety, depression, or obsessive-compulsive disorder.[3] These factors appear to influence the decision of patients to consult their physicians.

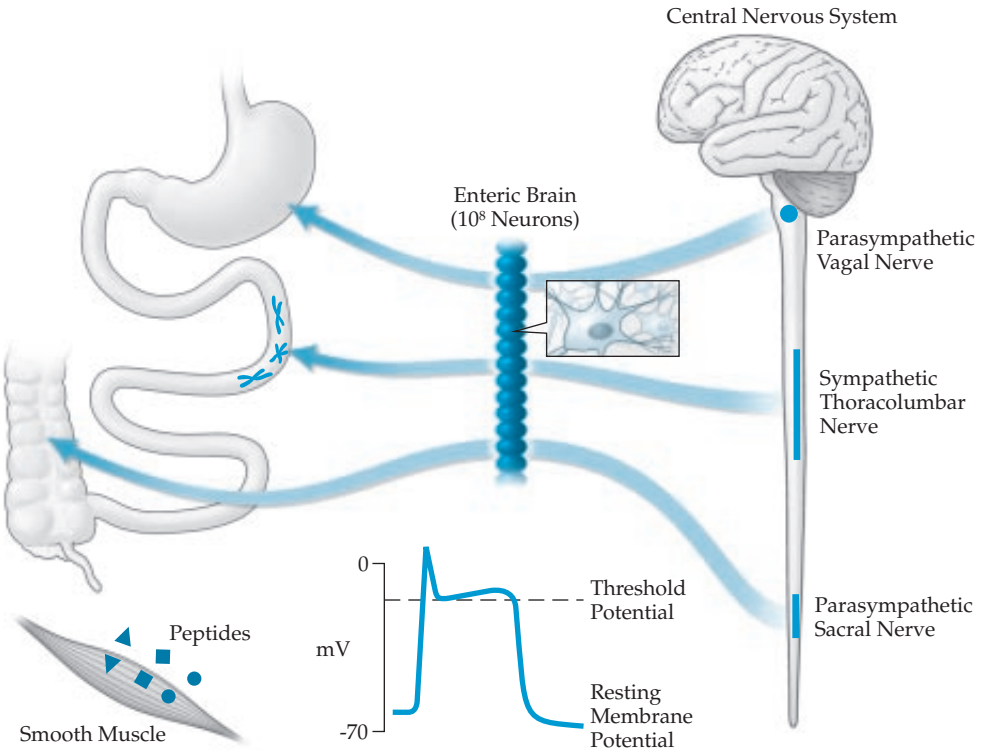


Figure 3 **Control of gut motility. Interactions between extrinsic neural pathways and the intrinsic nervous system (enteric brain) modulate contractions of gastrointestinal smooth muscle. Peptide-receptor interactions alter muscle membrane potentials by stimulating bidirectional ion fluxes. In turn, membrane characteristics dictate whether or not the muscle cell contracts.[59]**

Dyspepsia (from the Greek term for bad digestion) that is not caused by ulcers is characterized by upper abdominal symptoms in the postprandial period, such as nausea, vomiting, pain, bloating, anorexia, and early satiety. It affects about 20% of the population of the United States.

Pathogenesis

Factors other than altered gastric emptying, increased gastric sensitivity, and psychosocial distress may contribute to the development of nonulcer dyspepsia, but the pathogenesis remains unclear. A subgroup of patients may suffer nonerosive reflux esophagitis; others may have *Helicobacter pylori* gastritis. The role *H. pylori* infection plays in dyspepsia is uncertain, but current epidemiologic evidence and the results of eradication studies do not support a causal relationship.[14,15] Dyspepsia is also associated with impaired gastric relaxation or accommodation.[5,6]

Diagnosis

The history usually provides information on the specific symptoms or the spectrum of symptoms experienced by the patient. However, the symptoms have little discriminative value for predicting the physiologic alterations in an individual patient. Studies have suggested that the presence of postprandial fullness or satiation soon after starting the meal may be indicative of delayed gastric emptying and reduced gastric accommodation, respectively.[5,6,16] Moreover, weight loss of more than 5 kg is more frequent in patients with reduced gastric accommodation.[5] However, weight loss should not be dismissed on the basis of being a functional alteration; it mandates performance of endoscopy to exclude ulceration or cancer. The physical examination is usually normal. On rare occasions, there may be a succussion splash in the epigastrium from the retention of food in the stomach. An epigastric mass, hepatomegaly, or supraclavicular lymphadenopathy may suggest that the dyspepsia is the result of malignancy.

In the presence of alarm features such as dysphagia, bleeding, or weight loss in association with dyspepsia, it is essential to exclude mucosal diseases, such as ulcer or cancer.[2] Cancer may still be present, however, even when these alarm features are absent. Patients are reassured by a negative endoscopic examination.

In most cases, the underlying cause of dyspepsia will not be obvious from the history and physical examination. For new-onset dyspepsia, endoscopy and testing for *H. pylori* infection are generally recommended.

The presence of heartburn suggests a component of gastroesophageal reflux. Reflux needs to be differentiated from rumination,[17] which results in the effortless regurgitation of undigested food within 30 minutes after oral ingestion; rumination occurs with virtually every meal and is not associated with nausea. The symptoms that appear to be most closely related to impaired gastric relaxation or accommodation are early satiety and weight loss.[5] This impairment of accommodation may also contribute to hypersensitivity of the stomach of such patients to intraluminal stimuli. There is great interest in identifying ways to demonstrate this hypersensitivity before therapy is initiated, because such knowledge would have a bearing on choice of therapy. Tests in which the patient drinks water or a nutrient beverage have been devised to evaluate the maximal tolerated volume and the symptoms of fullness, satiety, bloating, nausea, and pain at a defined period after ingestion (typically 30 minutes).[5,18,19] These tests are noninvasive and inexpensive, and they have been introduced into clinical practice in some centers. However, they do not necessarily differentiate disturbances in the accommodation response from hypersensitivity per se. Until recently, measurement of accommodation required the placement of an intragastric balloon to measure fasting and postprandial volumes.[5] Recently, a novel imaging approach was developed to measure accommodation noninvasively by use of single-photon emission computed tomography.[6]

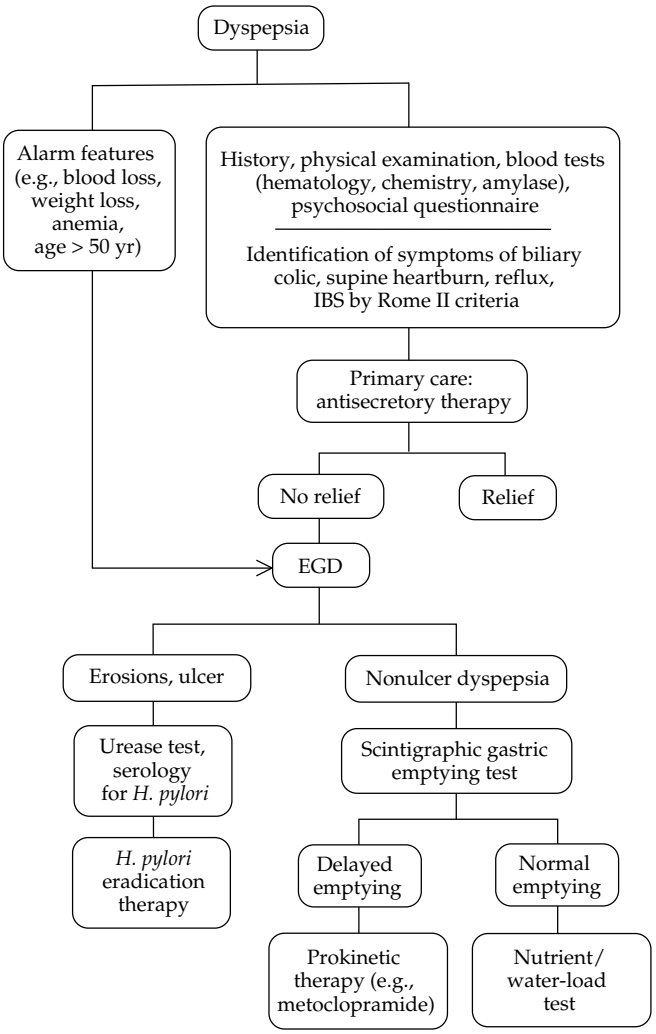


Figure 4 **Algorithm for the treatment of dyspepsia.[2] (IBS—irritable bowel syndrome; EGD—esophagogastroduodenoscopy)**

Simple, cost-effective tests for mucosal disease, gastric emptying,[20-22] and gastric accommodation[6] provide a rational alternative to the use of sequential empirical trials for identifying the mechanism causing dyspepsia[2] [*see Figure 4*].

Treatment

In clinical practice, dyspepsia is often treated with acid-suppressing regimens consisting of proton pump inhibitors or H_2 receptor antagonists, though the evidence in favor of this approach is limited, and of all patients treated, a cure is achieved in less than one in 10.[23] Temporary acid suppression with a proton pump inhibitor or an H_2 receptor agonist may delay diagnosis of cancer.[24]

In cases of dyspepsia associated with *H. pylori* infection, eradication of the *H. pylori* infection results in resolution of the syn-

drome in only a small minority of patients[25,26]; the current consensus is that in the absence of erosions or ulcers, eradication of *H. pylori* is not indicated for treatment of dyspepsia,[13] though it is usually treated anyway because of concern with development of atrophy or gastric cancer in the long term.[27]

IRRITABLE BOWEL SYNDROME

Diagnosis

Clinical features IBS is characterized by abdominal pain and alteration in bowel movements. The pain is often worse after eating (experienced by about 30% of patients), is usually located in the lower quadrants or hypogastrium, and is aggravated before and relieved after a bowel movement. Bowel function is disturbed with either diarrhea (increased frequency or stools of loose consistency) or constipation (decreased frequency, abnormally hard stools, need to strain to complete bowel evacuation). Most patients also experience abdominal bloating, and some have a sense of incomplete rectal evacuation or pass mucus with stools. These symptoms constitute the Manning criteria or the Rome criteria[3] and are helpful in diagnosing IBS.

The sense of incomplete evacuation may also suggest a component of outlet obstruction to defecation or increased rectal sensitivity.[4] In clinical practice, the diarrhea may present as either of two variants: (1) loose to watery stools often associated with borborygmi, abdominal cramps, and a borderline-high 24-hour stool weight (i.e., 200 to 300 g/day) or (2) small, pelletlike, repetitive stools that are misinterpreted as diarrhea. Constipation may be reflected in a reduction in the frequency of bowel movements (fewer than three a week), in incomplete rectal evacuation, or in the need for excessive straining.

The rectal examination is important in excluding anorectal or pelvic floor spasms that obstruct defecation.

IBS may be associated with lactose intolerance and colonic diverticulosis in some patients. It is unclear whether the associated conditions actually contribute to the clinical syndrome. Some patients who manifest clinical features of IBS have celiac disease; the proportion of IBS patients with celiac disease is unclear from the literature.[28] The effect of the exclusion of gluten on IBS symptoms in these patients has not been evaluated.

Laboratory tests The diagnosis of IBS is facilitated by recognition of the symptom complex described above. However, it is a diagnosis of exclusion, requiring selected tests to exclude organic disease, such as stool examination for ova, parasites, and occult blood; blood tests; serologic testing for celiac disease; flexible sigmoidoscopy; and, in patients older than 45 years, barium colon x-ray or colonoscopy [*see Figure 5*].

Treatment

Constipation in patients with IBS responds to treatment with fiber or simple laxatives, including osmotic agents[29,30]; psychotherapy may also be beneficial.[31] The serotonin (5-HT_3) antagonist alosetron was found to provide adequate relief of pain, improved stool frequency, decreased urgency, and consistency for many patients whose predominant bowel disturbance was diar-

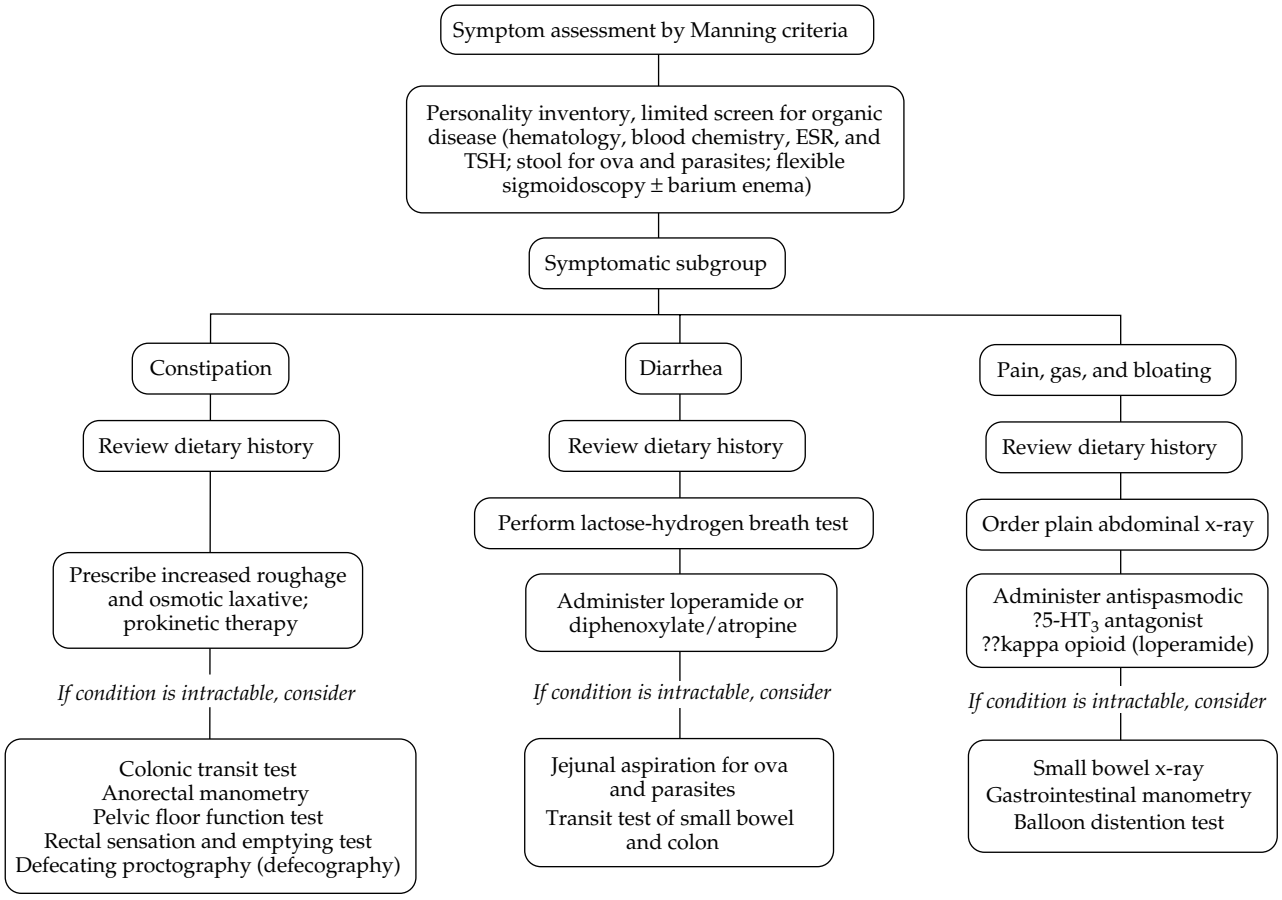


Figure 5 **Algorithm for irritable bowel syndrome.[60] (ESR—erythrocyte sedimentation rate; TSH—thyroid stimulating hormone)**

rhea.[32] The medication was withdrawn from the market, but it is possible that it will be reintroduced for specific indications. Other compounds in the same class (e.g., cilansetron) are undergoing trials. Other serotoninergic agents that activate the 5-HT_4 receptor may be approved for the treatment of constipation-predominant IBS.[33-35] IBS patients tend to use complementary and alternative medicine (CAM) more frequently than patients with organic bowel diseases. IBS patients who used CAM also were found to have significantly poorer quality-of-life scores for emotional and social factors than patients who did not use CAM.[36] Physicians need to be aware of this, both with regard to the potential for adverse interactions and as an indication of emotional unease in these patients.

SLOW-TRANSIT CONSTIPATION

Slow-transit constipation[4] is a motility disorder of the colon that results in prolonged transit.

Diagnosis

The diagnosis of slow-transit constipation should be made only after exclusion of mucosal diseases, such as tumors and strictures. It is most conveniently diagnosed by assessing mean colonic transit time through use of abdominal radiography and radiopaque markers. There are two commonly used variations of this method. The first type involves ingestion of 24 radiopaque markers in a soluble medication capsule on 4 successive days; plain abdominal radiography is performed on day 5. The number of markers in the colon approximates the mean colonic transit time in hours (normal: < 72 hours). The second variation requires that the patient ingest 20 markers on day 1; an abdominal x-ray is obtained on day 5. Normally, there should be fewer than five markers remaining in the colon. In all patients with delayed colonic transit, the possibility of outlet obstruction to defecation or a pangastrointestinal motility disorder must be ruled out.

Treatment

Treatment of slow-transit constipation consists of increasing dietary bulk or fiber and administering osmotic laxatives (e.g., magnesium salts when not contraindicated) and stimulant laxatives or colonic prokinetic agents.

A more severe variant of slow-transit constipation is colonic inertia. In this disorder, the colon fails to produce a motor response to physiologic stimuli, such as a meal, or to pharmacologic stimulation, as would occur, for example, after administration of neostigmine, 0.5 mg I.M., or intraluminal bisacodyl, 2 to 4 mg.

OUTLET OBSTRUCTION TO DEFECATION

Outlet obstruction to defecation (evacuation disorders) occurs when defecation dynamics [*see Figure 2*] function poorly and the patient is unable to expel stool.[4]

Diagnosis

The patient may present with constipation or the inability to have spontaneous and complete bowel movements; bloating; and left-sided abdominal pain. The syndrome may thus mimic IBS and is commonly associated with it. A careful clinical history is useful in identifying failure of evacuation; specifically, patients may experience the need for digital disimpaction of the rectum or digital pressure on the posterior wall of the vagina or the perineum to expel stool. Enemas may not be emptied. The rectal examination identifies an immobile perineum during the process of straining and a tight, unyielding puborectalis sling muscle abutting the rectum posteriorly. This tight pelvic floor persists during attempts to evacuate. In rare instances, the anal sphincter itself is spastic or the entire perineum balloons or herniates down as a result of years of straining or of multiple childbirths, which weaken the ligaments and muscles that normally support the pelvic floor and rectoanal angle.

Treatment

Occasionally, outlet obstruction is caused by an anatomic defect such as a rectocele or rectal internal mucosal prolapse; these are amenable to surgical correction. A spastic pelvic floor or spastic anal sphincter muscles usually respond to biofeedback and muscle relaxation exercises. Some patients with outlet obstruction to defecation have a profound psychological disorder or a history of abuse that requires identification and subsequent therapy.

Gastroparesis and Chronic Intestinal Pseudo-obstruction

PATHOGENESIS

Although several etiologic factors are involved in the development of gastric or small bowel motility disturbances [*see Table 1*], these can generally be grouped as disorders of the extrinsic nervous system, the enteric nervous system (including the interstitial cells of Cajal or intestinal pacemakers), or smooth muscle.[1]

Extrinsic Neuropathic Disorders

Extrinsic neuropathic processes include vagotomy, diabetes, amyloidosis, and a paraneoplastic syndrome usually associated with small cell carcinoma of the lung. Another common neuropathic problem in clinical practice is the effect of medications such as anticholinergics on neural control.

Enteric or Intrinsic Neuropathic Disorders

Disorders of the enteric nervous system are usually the result of a degenerative, immune, or inflammatory process.[37] The etiology can only rarely be ascertained in these disturbances; gastroparesis and pseudo-obstruction may be caused by viruses (including rotavirus, Norwalk virus, cytomegalovirus, and Epstein-Barr virus) or degenerative disorders associated with infiltration of the myenteric plexus with inflammatory cells [*see Figure 6*], including eosinophils. Idiopathic chronic intestinal pseudo-obstruction is a condition in which there is no disturbance of the extrinsic neural control and no underlying etiology for the enteric nervous system abnormality.

Full-thickness biopsies of the intestine may be required to evaluate the myenteric plexus[37] and interstitial cells of Cajal.[38] Regrettably, other than resection of the affected region (e.g., colectomy in slow-transit constipation), there are few therapeutic regimens that can be proposed on the basis of the information from the biopsy. The benefits of biopsy need to be weighed against the risk of complications associated with full-thickness intestinal biopsy, which include adhesion formation, with the potential for mechanical obstruction superimposed on episodes of pseudo-obstruction.

Smooth Muscle Disorders

Disturbances of smooth muscle may result in significant disorders of gastric emptying and small bowel transit as well as, occasionally, colonic transit. These disturbances include systemic sclerosis and amyloidosis. Dermatomyositis, dystrophia myotonica, and metabolic muscle disorders such as mitochondrial

Table 1 Classification of Gastroparesis and Chronic Intestinal Pseudo-obstruction

Type	*Myopathic*	*Neuropathic*	*Comments*
Familial	Familial visceral myopathies (autosomal dominant or recessive)	Familial visceral neuropathies, von Recklinghausen disease	Rare, often present in neonatal period or childhood; neurofibromata may also cause mechanical obstruction
Sporadic			
Infiltrative	Progressive systemic sclerosis	Early progressive systemic sclerosis	Manometry essential to differentiate pathophysiology (neuropathic vs myopathic)
	Amyloidosis	Amyloidosis	
General neurologic disease	Myotonic and other dystrophies	Diabetes, porphyria, spinal cord transection, dysautonomias, multiple sclerosis, brain stem tumor	For review, see reference 1
Infectious	—	Chagas disease, Norwalk virus, cytomegalovirus, Epstein-Barr virus	Nonspecific postviral causes appear to be common
Drug-induced	—	Tricyclic antidepressants, narcotics, anticholinergics, antihypertensives, vincristine	Adverse effects of medications to be excluded in all patients
Neoplastic	—	Paraneoplastic (small cell lung cancer, carcinoid lung tumors)	May require computed tomography to exclude tumor if chest x-ray is negative
Idiopathic	Sporadic hollow visceral myopathy	Chronic idiopathic intestinal pseudo-obstruction	Variable manifestations and severity

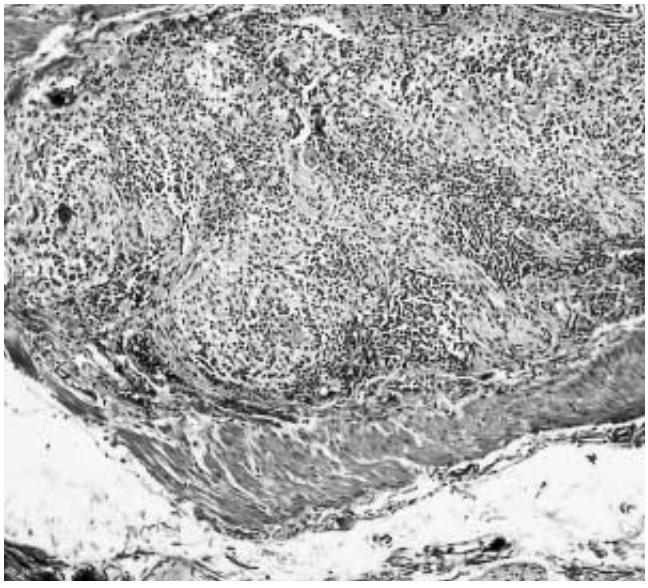
myopathy are seen infrequently and are suggested by the presence of ptosis, external ocular paralysis, acidosis, and peripheral neuromyopathy.[39] Hollow visceral myopathy may occur either sporadically or, rarely, in families. Motility disturbances may be the result of metabolic disorders such as hypothyroidism and hyperparathyroidism, but patients with these disorders more commonly present with constipation.

DIAGNOSIS

Clinical Features

The clinical features of gastroparesis and chronic intestinal pseudo-obstruction are similar and include nausea, vomiting, early satiety, abdominal discomfort, distention, bloating, and anorexia. In patients in whom stasis and vomiting are significant problems, there may be considerable weight loss, and disturbances of mineral and vitamin stores may result. The severity of the motility problem often manifests itself most clearly in the degree of nutritional and electrolyte depletion. Disturbances of bowel movements, such as diarrhea and constipation, indicate that the motility disorder is more extensive than gastroparesis. Significant vomiting may be complicated by aspiration pneumonia or Mallory-Weiss tears, which may result in acute GI hemorrhage. When patients have a more generalized motility disorder, there may also be symptoms referable to abnormal swallowing or delayed colonic transit.

a



b

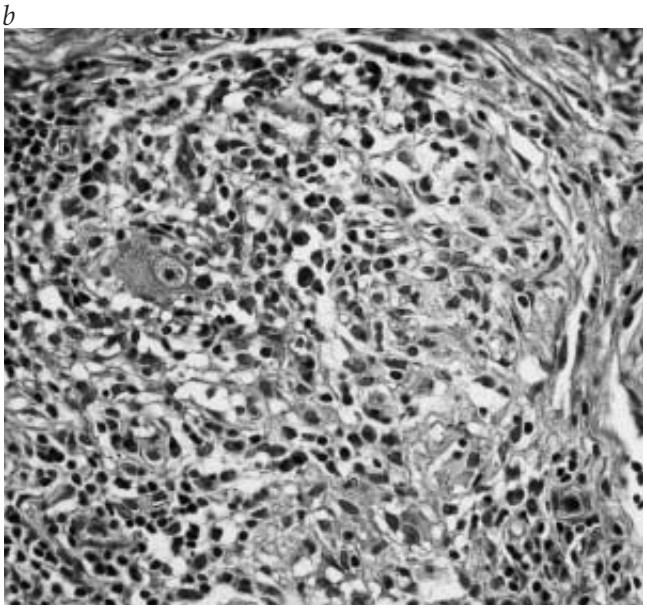


Figure 6 **Mononuclear infiltration in the gastric enteric plexus from a patient with small cell lung cancer and a paraneoplastic gastroparesis. The portion of the plexus rich in ganglion cells was expanded, but no necrosis was observed. Intact nerve fiber bundles can be seen (*a*) (original magnification: ×100). High-power view of the same field (*b*) shows mature small lymphocytes and abundant plasma cells. Although neurons are decreased in number, a normal-appearing ganglion cell can be seen just above the center of the image (original magnification: ×400).[61]**

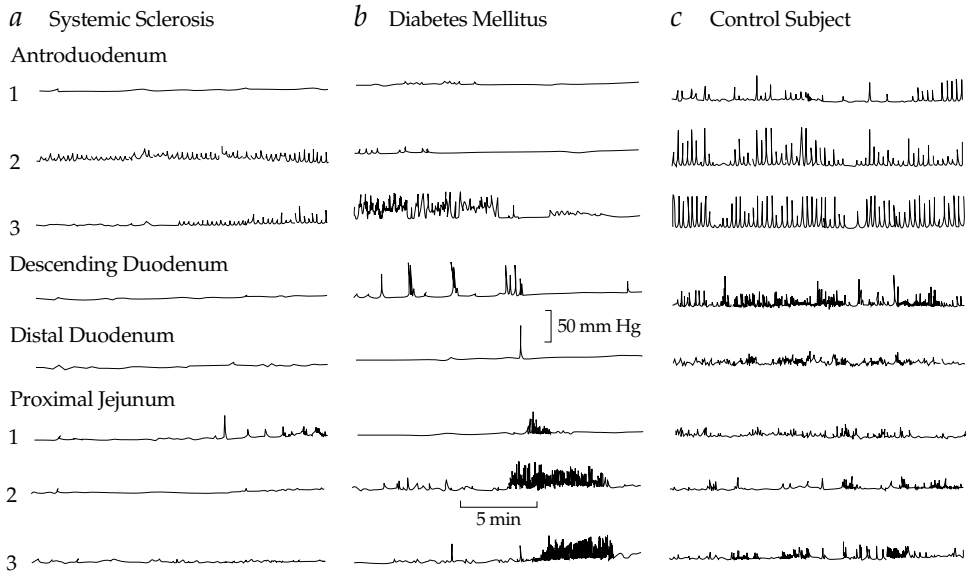


Figure 7 **Postprandial manometric profiles of patients with small bowel dysmotility caused by myopathy (*a*) and neuropathy (*b*), compared with a healthy control subject (*c*).**

Family history and medication history are essential to identify underlying etiologic factors such as diabetes mellitus that may result in gastric or small bowel motor disorders. A careful review of systems will help reveal an underlying collagen vascular disease (e.g., scleroderma) or disturbances of extrinsic neural control that also may be affecting the abdominal viscera. Such symptoms include orthostatic dizziness; difficulties with erection or ejaculation; recurrent urinary tract infections; dry mouth, eyes, or vagina; difficulties with visual accommodation in bright lights; and absence of sweating.

On physical examination, the presence of a succussion splash is usually indicative of a region of stasis within the GI tract, typically the stomach. The hands and mouth may show signs of Raynaud phenomenon or scleroderma. Testing of pupillary responses (to light and accommodation), blood pressure in the lying and standing positions, general features of a peripheral neuropathy, and external ocular movements can identify patients with an associated neurologic disturbance, such as those with a long history (usually longer than 10 years) of diabetes mellitus or oculogastrointestinal dystrophy.

The conditions to be differentiated are mechanical obstruction—which may occur because of peptic stricture or Crohn disease in the small intestine—functional GI disorders, and eating disorders such as anorexia nervosa and rumination syndrome. The degree of impairment of gastric emptying in eating disorders is relatively minor compared with diabetic and postvagotomy gastric stasis. Rumination syndrome is characterized by postprandial, effortless regurgitation of undigested food within 30 minutes after virtually every meal. This condition occurs in mentally challenged children (e.g., those with Down syndrome), and it is also being increasingly recognized in adolescents and adults of normal intelligence.[17] It is treatable by behavioral modification, including diaphragmatic breathing in the postprandial period.

Neonatal pseudo-obstruction rarely occurs alone; it is more often found in association with other anomalies requiring surgical correction, including gastroschisis, duodenal atresia, or megacystis. Prokinetic medications are usually ineffective, and many patients require parenteral nutrition and bowel decompression, including gastrostomies or enterostomies.[40]

Laboratory Tests

A motility disorder of the stomach or small bowel should be suspected whenever large volumes are aspirated from the stomach, particularly after an overnight fast, or when undigested solid food or large volumes of liquids are observed during an esophagogastroduodenoscopy. Barium studies rarely identify the etiology of the motor disorder except in small bowel systemic sclerosis, which is characterized by megaduodenum and packed valvulae conniventes in the small intestine. Barium x-ray, however, serves the important function of excluding mechanical obstruction. The diagnosis of a gastric or small bowel motility disorder, therefore, depends on a careful history and confirmation by transit tests.

The emptying of solids provides the best way to distinguish between healthy and disease states. If the patient's history includes an obvious etiologic factor, such as long-standing diabetes mellitus, it is usually unnecessary to pursue further investigations. When the cause of the gastric or small bowel transit disorder is unclear and the patient does not respond to treatment with a prokinetic agent, referral to a specialized center for autonomic tests and upper GI manometry may be needed [*see Figure 7*]. Transit tests, which can now be performed relatively simply and inexpensively,[20-22] enable good discrimination between healthy and disease states. The two most widely available approaches are the carbon-13 breath tests and scintigraphy with scans taken immediately after ingestion of the radiolabeled meal, as well as 1, 2, 4, and 6 hours later.[20] Manometry is generally available only in specialized centers; it may identify a myopathic or neuropathic disorder or an unsuspected mechanical obstruction resulting from simultaneous prolonged contractions at several levels of the intestine.[12]

In patients presenting with diarrhea, it is important to assess nutritional status (essential element and vitamin levels) and to exclude bacterial overgrowth by culture of small bowel aspirates. It is also important to exclude celiac sprue by small bowel biopsies. Bacterial overgrowth is relatively uncommon in neuropathic disorders but is more often found in myopathic conditions, such as scleroderma, that are more often associated with bowel dilatation.

TREATMENT

Four questions should be considered in the management of each patient. First, is the presentation acute or chronic? Second, is there evidence of a systemic disorder indicative of a neuropathy or myopathy? Third, what is the patient's state of hydration and nutrition? Fourth, which regions of the digestive tract are affected? The principal methods of management include correction of dehydration and nutritional deficiencies, the use of prokinetic and antiemetic medications [*see Table 2*], suppression of bacterial overgrowth, decompression of dilated segments, and surgery.[41]

Correction of Dehydration and Nutritional Deficiencies

Correction of dehydration and electrolyte and nutritional depletion is particularly important during acute exacerbations of gastroparesis or chronic intestinal pseudo-obstruction syndromes. Nutritional support should be tailored to the severity of the deficiencies of trace elements and dietary constituents in each patient. Dietary measures include the use of low-fiber and low-fat caloric supplements that contain iron, folate, calcium, and vitamins D, K, and B_{12}. Patients who have more severe symptoms, such as severe diabetic gastroparesis or severe myopathic pseudo-obstruction, may need parenteral or enteral nutrition supplementation.[31] For patients in whom supplementation of nutrition may be required for more than 3 months, it is usually best to place the enteral feeding tube via laparoscopy or minilaparotomy to secure the location of the tube in the intestine. Although severely affected patients may need parenteral nutrition, many patients continue to tolerate some oral feeding.

Prokinetic Therapy

Prokinetic medications (e.g., metoclopramide, 10 to 20 mg up to four times a day) are often used for the treatment of neuromuscular motility disorders.[41] Unfortunately, there is little evidence that they are effective in myopathic disturbances. Domperidone, a D_2 dopamine antagonist with antiemetic properties, relieves symptoms of diabetic gastroparesis,[42] but it is not approved for use in the United States.

Erythromycin, a macrolide antibiotic that stimulates motilin receptors partly through a cholinergic mechanism, results in the dumping of nondigestible and digestible solids from the stomach. Erythromycin lactobionate at a dosage of 3 to 6 mg/kg every 8 hours clears bezoars from the stomach in patients with diabetic gastroparesis.[43,44] The effect of oral erythromycin appears to be restricted by tachyphylaxis; there is little evidence that continued therapy is effective beyond 2 weeks, and GI upset may develop in some patients.[44]

Before its withdrawal in 2000, cisapride, a substituted benzamide that acts as a serotonin agonist, was used to treat altered motility, such as impaired gastric emptying, in patients with both gastroparesis and chronic intestinal pseudo-obstruction.[41]

Metoclopramide, with its antinausea and indirect cholinomimetic actions, is the current drug of choice for the treatment of motility disorders, though evidence for its efficacy is limited. Neuropsychiatric side effects such as dystonias are not infrequent, and rare cases of tardive dyskinesia have been reported. The usual dosage is 10 mg four times a day. Metoclopramide is also available for parenteral use; the usual dose is 10 mg I.M. or I.V. It should be used with caution, and a test dose (1 to 2 mg) is often used to exclude dystonic reactions resulting from an idiosyncratic reaction.

The peripheral dopaminergic antagonist domperidone has been shown to be efficacious in diabetic gastroparesis[45]; its efficacy is generally similar to that of metoclopramide. This agent suppresses emesis at the chemoreceptor trigger zone, which is outside the blood-brain barrier. Domperidone is not approved in the United States. The usual dosage is 30 to 80 mg/day in three or four divided doses. Novel prokinetics that are currently undergoing trial include the partial or full 5-HT_4 agonists such as tegaserod, levosulpiride, renzapride, and mosapride. Tegaserod accelerates gastric and small bowel transit in healthy persons[46] and small bowel transit in patients with constipation-predomi-

Table 2 Medications Used in the Treatment of Gastrointestinal Motility Disorders

Drug	*Dose*	*Efficacy Rating*	*Comments*
Gastroparesis			
Prokinetics			
Metoclopramide I.M. or p.o	10 mg t.i.d. + h.s.	Moderate	Central side effects; no evidence of efficacy below stomach level
Domperidone, p.o	10–30 mg t.i.d. + h.s.	Moderate	Not approved
Erythromycin I.V. or p.o.	40–200 mg t.i.d.	First choice if I.V. agent needed	Oral administration results in abdominal side effects after 2 wk
Cisapride p.o	10 mg t.i.d. + h.s.	Greater	Limited access because of drug interactions and potential for cardiac dysrhythmia
Octreotide s.c.	25–50 mg h.s.	Modest	For induction of MMCs to avoid bacterial overgrowth; if given with meals, retards transit
Antiemetics			
Prochlorperazine	5–12.5 mg p.r.n.	Significant for adult emesis	I.M., p.o., or rectal suppository
5-HT_3 antagonists (e.g., ondansetron)	0.15 mg/kg I.V. or p.o.	Modest	Less effective for dysmotility than for chemotherapy-induced emesis
Dumping, Diarrhea, or Short Bowel			
Octreotide	25 µg t.i.d., a.c.	Moderate	Adjunct to total parenteral nutrition and fluid replacement; prescribe multidraw vial; store in refrigerator
Irritable Bowel Syndrome Diarrhea			
Loperamide	2 mg up to 6 mg/day	First choice	No benefit for pain; very effective for diarrhea

5-HT_3 — serotonin MMCs — mucosal mast cells

nant irritable bowel syndrome.[47] It is chemically different from the benzamides and does not cause cardiac dysrhythmias.[35]

Octreotide, a cyclized octapeptide analogue of somatostatin, has been shown to induce activity fronts (phase III of the migrating motor complex) in the small intestine.[48] In an open trial, octreotide appeared to alleviate symptoms in patients with small bowel scleroderma who received the drug for up to 3 weeks.[49] However, it is unclear whether small bowel transit really improves with use of octreotide. In healthy persons, low doses of octreotide markedly retard small bowel transit.[50] Its therapeutic efficacy needs to be further assessed in clinical trials.

Symptomatic Relief

Standard antiemetics can be used for symptom relief. The more expensive 5-HT_3 antagonists (e.g., ondansetron) have not been proved to be of greater benefit than the less expensive antiemetic agents. Some patients have significant pain; if pain is associated with gut dilatation, decompression may be needed (see below). In the absence of dilatation, narcotics should be avoided, because these agents may aggravate the motility disorder. Parenteral ketorolac may be useful during acute exacerbations of pain.

Antibacterial Therapy

Bacterial overgrowth must be suppressed in infected patients with pseudo-obstruction. A common practice is to use a different antibiotic for 7 to 10 days each month to avoid development of bacterial resistance, although no trials have been performed to study this approach. Typical antibiotics used are doxycycline (100 mg b.i.d.), metronidazole (500 mg t.i.d.), ciprofloxacin (500 mg b.i.d.), or double-strength trimethoprim-sulfamethoxazole (two tablets b.i.d.). These measures usually result in significant symptomatic relief in those patients with diarrhea and steatorrhea.

Decompression

Decompression is rarely necessary in patients with chronic intestinal pseudo-obstruction and should be restricted to patients with severe motility disorders who are being cared for at tertiary care centers. Venting enterostomy creates a means to relieve abdominal distention and bloating and has been shown to significantly reduce the frequency of nasogastric intubations or hospitalizations for acute exacerbations in patients with severe intestinal pseudo-obstruction that require central parenteral nutrition.[40] Access to the small intestine may also provide a way to deliver nutrients by the enteral route. Enteral tubes are available that facilitate both aspiration and feeding with the same apparatus.

Electrical Stimulation and Gastric Pacing

Electrical stimulation and gastric pacing is an evolving treatment option for patients who do not respond to standard medical therapy[51]; however, the evidence for efficacy of electrical stimulation is controversial, and the mechanism for relieving symptoms is unknown.

Surgical Treatment

Surgical treatment should be considered whenever the motility disorder is localized to a portion of the gut that can be resected. In clinical practice, the two most common surgical therapies are (1) colectomy and ileoproctostomy for intractable symptoms associated with slow-transit constipation, colonic inertia, or pseudo-obstruction and (2) completion gastrectomy for patients with stasis syndrome after gastric surgery.[40] The role of small bowel transplantation in patients with pseudo-obstruction is unclear. Successful transplants have been performed in children with pseudo-obstruction,[52] but the experience in adults is limited.[52]

PROGNOSIS

The prognosis depends on the severity of the case. Patients with suspected postviral gastroparesis appear to have a good overall prognosis, with restoration of nutrition and reduction of symptoms within 2 years. On the other hand, patients with myopathic and dilated bowel have persistent symptoms, are more prone to develop bacterial overgrowth, and usually require long-term parenteral nutrition, which carries inherent morbidity. Between these extremes are patients with mild to moderately severe motility disorders who have recurrent or chronic symptoms. These patients can usually be managed as outpatients with dietary supplementation to maintain nutrition (including liquid formula supplements) and medications (e.g., prokinetics or antiemetics) and decompression to relieve symptoms.

Dumping Syndrome and Accelerated Gastric Emptying

Rapid gastric emptying results from impaired relaxation of the stomach upon ingestion of food. Postprandial intragastric pressure is relatively high and results in active propulsion of liquid foods from the stomach. A high caloric (usually carbohydrate) content of the liquid phase of the meal evokes a rapid insulin response with secondary hypoglycemia. These patients may also have impaired antral contractility and gastric stasis of solids, which may paradoxically result in a clinical picture of both gastroparesis (for solids) and dumping (for liquids). Typically, these conditions follow truncal vagotomy and gastric drainage procedures; with the use of more selective vagotomies in the treatment of peptic ulceration, it is likely that the prevalence of these problems may decrease. The most useful means of assessment is a dual-phase (solid and liquid) radioisotopic gastric emptying test.

Management of dumping[53] includes patient education (particularly regarding the avoidance of high-nutrient liquid drinks) and, possibly, the addition of guar gum or pectin to retard liquid emptying. If these measures are ineffective, pharmacologic approaches, such as use of subcutaneous octreotide (50 to 100 mg) 15 minutes before meals, decreases many of the vasomotor symptoms and also retards gastric emptying and small bowel transit, thereby relieving associated hypoglycemia and diarrhea.[54,55]

Rapid-Transit Dysmotilities of the Small Bowel

Rapid transit through the small bowel is a minor component of IBS in some patients.[4] However, it is a major component of other diseases and results in a significant loss of fluid and osmotically active solutes that overwhelm colonic capacitance and reabsorptive capacity and result in severe diarrhea. Examples include postvagotomy diarrhea, short bowel syndrome, diabetic diarrhea, and carcinoid diarrhea.[56] These disturbances of small bowel transit can best be identified by use of scintigraphy or, if scintigraphy is not available, by use of the lactulose-hydrogen breath test.

The objectives of treatment are restoration of hydration and nutrition and retardation of small bowel transit. Dietary interventions include avoidance of hyperosmolar drinks (e.g., virtually all soft drinks), use of iso-osmolar or hypo-osmolar rehydration solutions, and reduction of the fat content in the diet to

around 50 g a day to avoid delivery of unabsorbed fat to the colon (where their metabolites are cathartic). Correction of nutritional deficiencies (commonly, calcium, magnesium, potassium, and water- and fat-soluble vitamins) is often required.

Pharmacotherapy should be delivered in a stepwise fashion. First, an opioid agent in high dosage (e.g., loperamide, 4 mg) is given one-half hour before each meal and at bedtime to suppress the small bowel transit and colonic response to feeding. Next, verapamil (40 mg b.i.d.) or clonidine (0.1 to 0.2 mg orally or by patch) should be given, and if these are ineffective or produce unacceptable side effects (usually hypotension), subcutaneous octreotide, starting at 50 μg before meals, should be prescribed.[54] Patients with less than 1 m of residual small bowel may be unable to sustain fluid and electrolyte homeostasis without parenteral support. However, it is almost invariably possible to maintain patients with more than 1 m of residual small bowel with oral nutrition, pharmacotherapy, and supplements.

Michael Camilleri, M.D., has received grant or research support from and is a consultant for Novartis Pharmaceuticals Corp., GlaxoSmithKline PLC, Janssen Pharmaceutica Products LP, and Solvay Pharmaceuticals Inc.

References

1. Camilleri M, Bharucha AE: Gastrointestinal dysfunction in neurologic disease. Semin Neurol 16:203, 1996

2. Camilleri M: Nonulcer dyspepsia: a look into the future. Mayo Clin Proc 71:614, 1996

3. Drossman DA, Whitehead WE, Camilleri M: Irritable bowel syndrome: a technical review for practical guideline development. Gastroenterology 112:2120, 1997

4. Camilleri M, Thompson WG, Fleshman JW, et al: Clinical management of intractable constipation. Ann Intern Med 121:520, 1994

5. Tack J, Piessevaux H, Coulie B, et al: Role of impaired gastric accommodation to a meal in functional dyspepsia. Gastroenterology 115:1346, 1998

6. Kim DY, Delgado-Aros S, Camilleri M, et al: Noninvasive measurement of gastric accommodation in patients with idiopathic nonulcer dyspepsia. Am J Gastroenterol 96:3099, 2001

7. Camilleri M, Ford MJ: Review article: colonic sensorimotor physiology in health, and its alteration in constipation and diarrhoeal disorders. Aliment Pharmacol Ther 12:287, 1998

8. Camilleri M, Saslow SB, Bharucha AE: Gastrointestinal sensation: mechanisms and relation to functional gastrointestinal disorders. Gastroenterol Clin North Am 25:247, 1996

9. Miedema BW, Kelly KA, Camilleri M, et al: Human gastric and jejunal transit and motility after Roux gastrojejunostomy. Gastroenterology 103:1133, 1992

10. Karlstrom L, Kelly KA: Roux-Y gastrectomy for chronic gastric atony. Am J Surg 157:44, 1989

11. Wijnhoven BP, Salet GA, Roelofs JM, et al: Function of the proximal stomach after Nissen fundoplication. Br J Surg 85:267, 1998

12. Frank JW, Sarr MG, Camilleri M: Use of gastroduodenal manometry to differentiate mechanical and functional intestinal obstruction: an analysis of clinical outcome. Am J Gastroenterol 89:339, 1994

13. Krishnamurthy S, Kelly MM, Rohrmann CA, et al: Jejunal diverticulosis: a heterogenous disorder caused by a variety of abnormalities of smooth muscle or myenteric plexus. Gastroenterology 85:538, 1983

14. Danesh J, Lawrence M, Murphy M, et al: Systematic review of the epidemiological evidence on *Helicobacter pylori* infection and nonulcer or uninvestigated dyspepsia. Arch Intern Med 160:1192, 2000

15. Laine L, Schoenfeld P, Fennerty MB: Therapy for *Helicobacter pylori* in patients with nonulcer dyspepsia: a meta-analysis of randomized, controlled trials. Ann Intern Med 134:361, 2001

16. Stanghellini V, Tosetti C, Paternico A, et al: Predominant symptoms identify different subgroups in functional dyspepsia. Am J Gastroenterol 94:2080, 1999

17. O'Brien MD, Bruce BK, Camilleri M: Rumination syndrome: clinical features rather than manometric diagnosis. Gastroenterology 108:1024, 1995

18. Boeckxstaens GE, Hirsch DP, van den Elzen BD, et al: Impaired drinking capacity in patients with functional dyspepsia: relationship with proximal stomach function. Gastroenterology 121:1054, 2001

19. Chial HJ, Camilleri C, Delgado-Aros S, et al: A nutrient drink test to assess maximum tolerated volume and postprandial symptoms: effects of gender, body mass index and age in health. Neurogastroenterol Motil (in press)

20. Thomforde GM, Camilleri M, Phillips SF, et al: Evaluation of an inexpensive screening scintigraphic test of gastric emptying. J Nucl Med 36:93, 1995

21. Ghoos YF, Maes BD, Geypens BJ, et al: Measurement of gastric emptying rate of solids by means of a carbon-labeled octanoic acid breath test. Gastroenterology 104:1640, 1993

22. Viramontes BE, Kim DY, Camilleri M, et al: Validation of a stable isotope gastric emptying test for normal, accelerated or delayed gastric emptying. Neurogastroenterol Mot 13:567, 2001

23. Delaney BC, Innes MA, Deeks J, et al: Initial management strategies for dyspepsia. Cochrane Database Syst Rev (3):CD001961, 2001

24. Bramble MG, Suvakovic Z, Hungin AS: Detection of upper gastrointestinal cancer in patients taking antisecretory therapy prior to gastroscopy. Gut 46:464, 2000

25. McColl K, Murray L, El-Omar E, et al: Symptomatic benefit from eradicating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. N Engl J Med 339:1869, 1998

26. Blum AL, Talley NJ, O'Morain C, et al: Lack of effect of treating *Helicobacter pylori* infection in patients with nonulcer dyspepsia. Omeprazole Plus Clarithromycin and Amoxicillin Effect One Year after Treatment (OCAY) Study Group. N Engl J Med 339:1875, 1998

27. Uemura N, Okamoto S, Yamamoto S, et al: *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med 345:784, 2001

28. Sanders DS, Carter MJ, Hurlstone DP, et al: Association of adult coeliac disease with irritable bowel syndrome: a case-control study in patients fulfilling ROME II criteria referred to secondary care. Lancet 358:1504, 2001

29. Tramonte SM, Brand MB, Mulrow CD, et al: The treatment of chronic constipation in adults: a systematic review. J Gen Intern Med 12:15, 1997

30. Voderholzer WA, Schatke W, Muhldorfer BE, et al: Clinical response to dietary fiber treatment of chronic constipation. Am J Gastroenterol 92:95, 1997

31. Guthrie E, Creed F, Dawson D, et al: A controlled trial of psychological treatment for the irritable bowel syndrome. Gastroenterology 100:450, 1991

32. Camilleri M, Northcutt AR, Kong S, et al: Efficacy and safety of alosetron in women with irritable bowel syndrome: a randomised, placebo-controlled trial. Lancet 355:1035, 2000

33. Müller-Lissner SA, Fumagalli I, Bardhan KD, et al: Tegaserod, a 5-HT_4 receptor partial agonist, relieves symptoms in irritable bowel syndrome patients with abdominal pain, bloating and constipation. Aliment Pharmacol Ther 15:1655, 2001

34. Camilleri M: Management of the irritable bowel syndrome. Gastroenterology 120:652, 2001

35. Camilleri M: Review article: tegaserod. Aliment Pharmacol Ther 15:277, 2001

36. Langmead L, Chitnis M, Rampton DS: Use of complementary therapies by patients with IBD may indicate psychological distress. Inflamm Bowel Dis 8:174, 2002

37. Singaram C, SenGupta A: Histopathology of the enteric neuropathies: from silver staining to immunohistochemistry. Gastroenterol Clin North Am 25:183, 1996

38. He CL, Burgart L, Wang L, et al: Decreased interstitial cell of Cajal volume in patients with slow-transit constipation. Gastroenterology 118:14, 2000

39. Mueller LA, Camilleri M, Emslie-Smith AM: Mitochondrial neurogastrointestinal encephalomyopathy: manometric and diagnostic features. Gastroenterology 116:959, 1999

40. Murr MM, Sarr MG, Camilleri M: The surgeon's role in the treatment of chronic intestinal pseudo-obstruction. Am J Gastroenterol 90:2147, 1995

41. Camilleri M: Appraisal of medium- and long-term treatment of gastroparesis and chronic intestinal dysmotility. Am J Gastroenterol 89:1769, 1994

42. Soykan I, Sarosiek I, McCallum RW: The effect of chronic oral domperidone therapy on gastrointestinal symptoms, gastric emptying, and quality of life in patients with gastroparesis. Am J Gastroenterol 92:976, 1997

43. Janssens J, Peeters TL, Vantrappen G, et al: Improvement in gastric emptying in diabetic gastroparesis by erythromycin: preliminary studies. N Engl J Med 322:1028, 1990

44. Richards RD, Davenport K, McCallum RW: The treatment of idiopathic and diabetic gastroparesis with acute intravenous and chronic oral erythromycin. Am J Gastroenterol 88:203, 1993

45. Patterson D, Abell T, Rothstein R, et al: A double-blind multicenter comparison of domperidone and metoclopramide in the treatment of diabetic patients with symptoms of gastroparesis. Am J Gastroenterol 94:1230, 1999

46. Degen L, Matzinger D, Merz M, et al: Tegaserod, a 5-HT4 receptor partial agonist, accelerates gastric emptying and gastrointestinal transit in healthy male subjects. Aliment Pharmacol Ther 15:1745, 2001

47. Prather CM, Camilleri M, Zinsmeister AR, et al: Tegaserod accelerates orocecal transit in patients with constipation-predominant irritable bowel syndrome. Gastroenterology 118:463, 2000

48. Haruma K, Wiste JA, Camilleri M: Effect of octreotide on gastrointestinal pressure profiles in health and in functional and organic gastrointestinal disorders. Gut 35:1064, 1994

49. Soudah HC, Hasler WL, Owyang C: Effect of octreotide on intestinal motility and bacterial overgrowth in scleroderma. N Engl J Med 325:1461, 1991

50. von der Ohe MR, Camilleri M, Thomforde GM, et al: Differential regional effects of octreotide on human gastrointestinal motor function. Gut 36:743, 1995

51. WAVESS Study Group: Gastric electrical stimulation for gastroparesis: a multi-center double blind cross over study by the WAVESS study group (abstr). Gastroenterology 118:2060, 2000

52. Sigurdsson L, Reyes J, Kocoshis SA, et al: Intestinal transplantation in children with chronic intestinal pseudo-obstruction. Gut 45:570, 1999

53. Hasler WL: Dumping syndrome. Curr Ther Options Gastroenterology 5:139, 2002

54. Farthing MJ: Octreotide in dumping and short bowel syndromes. Digestion 54:47, 1993

55. Vecht J, Lamers CB, Masclee AA: Long-term results of octreotide-therapy in severe dumping syndrome. Clin Endocrinol 51:619, 1999

56. von der Ohe M, Camilleri M, Kvols LK, et al: Motor dysfunction of the small bowel and colon in patients with the carcinoid syndrome and diarrhea. N Engl J Med 329:1073, 1993

57. Malagelada JR, Camilleri M, Stanghellini V: Manometric Diagnosis of Gastrointestinal Motility Disorders. Thieme Medical Publishers, New York, 1986

58. Camilleri M: Four patients with intractable constipation. Gastrointestinal Diseases Today 2:7, 1993

59. Camilleri M, Phillips SF: Disorders of small intestinal motility. Motility Disorders. Gastroenterology Clinics of North America, Vol 18, No. 2. Ouyang A, Ed. WB Saunders Co, Philadelphia, 1989, p 405

60. Camilleri M, Prather CM: The irritable bowel syndrome: mechanisms and a practical approach to management. Ann Intern Med 116:1001, 1992

61. Lennon VA, Sas DF, Busk MF, et al: Enteric neuronal autoantibodies in pseudoobstruction with small-cell lung carcinoma. Gastroenterology 100:137, 1991

Acknowledgments

I wish to thank Mrs. Cindy Stanislav for her excellent secretarial assistance with the manuscript.

Figures 2 and 3 Tom Moore.

XV LIVER AND PANCREAS TRANSPLANTATION

Robert L. Carithers, Jr., m.d.
James D. Perkins, m.d.

Liver Transplantation

More than 4,500 liver transplants are performed annually in the United States.[1] Enhancements in surgical technique and the availability of powerful immunosuppressive agents have resulted in steady improvement in patient survival. As a result, liver transplantation has been accepted as the standard of care for patients with severe acute or chronic liver disease in whom conventional modalities of therapy have failed. The major obstacle to the procedure is the critical shortage of donor organs.

candidates for transplantation

Any patient with acute or chronic liver failure is a potential candidate for liver transplantation; there are a number of common indications [*see Table 1*].[2]

The methods of evaluating candidates for transplantation include careful history and physical examination, echocardiography, color flow Doppler imaging of the portal vein, computed tomographic angiography of the hepatic arterial supply, and a thorough evaluation of social factors and support. Echocardiography is useful in assessing left ventricular function and detecting pulmonary hypertension, which is seen in as many as 5% of cirrhotic patients.[3] Color flow Doppler studies of the portal vein are used to gauge the integrity of portal vein flow. If portal vein thrombosis is detected, the transplant surgeon can obtain extra donor vessels to bypass the blockade if necessary. CT angiography permits detection of small hepatocellular carcinomas, detection of aberrant arterial blood supply to the liver, and visualization of the splenic artery. Rigorous evaluation of the patient for any addictive behavior and assessment of the patient's social support system allow the transplant team to plan in advance for any needed services, which may include counseling, specialized addiction treatment, housing, transportation, and financial assistance for medications and other expenses.

Table 1 Common Indications for Liver Transplantation

- Chronic hepatitis
  - Hepatitis C
  - Hepatitis B
  - Autoimmune hepatitis
- Cholestatic liver disease
  - Biliary atresia (in children)
  - Primary biliary cirrhosis
  - Sclerosing cholangitis
- Metabolic diseases
  - Wilson disease
  - α_1-Antitrypsin deficiency
  - Hemochromatosis
- Malignancy
  - Primary hepatocellular carcinoma
  - Hemangioendothelioma
- Alcoholic liver disease
- Cryptogenic cirrhosis
- Miscellaneous conditions
  - Hepatic veno-occlusive disease
  - Nonalcoholic steatohepatitis
  - Tyrosinemia
  - Crigler-Najjar syndrome
- Fulminant hepatic failure
  - Hepatitis B
  - Hepatitis A
  - Acetaminophen overdose
  - Other drug-induced hepatitis
  - Toxin-induced hepatitis
  - Other viral hepatitides

contraindications to transplantation

Patients with severe neurologic or cardiopulmonary disease cannot withstand the stress of transplantation surgery. Patients with cirrhosis who have severe pulmonary hypertension rarely survive the operation and perioperative period.[4] Other contraindications to transplantation include severe or morbid obesity, extrahepatic malignancies, systemic infection, and cholangiocarcinoma.[5] The most common surgical contraindication to liver transplantation is thrombosis of the portal vein and other splanchnic veins to such an extent that no viable portal blood flow can be achieved.[6] Finally, the most frequent contraindications to liver transplantation are ongoing destructive behavior resulting from drug or alcohol addiction and the inability of the patient to comply with the complex medical regimen required after the operation.

timing of transplantation

Determining the optimal timing for liver transplantation can be as important as patient selection. A few simple clinical approaches have proved useful in determining the prognosis of patients with liver disease. These approaches include use of the Child-Turcotte-Pugh (CTP) classification [*see Table 2*], use of the Mayo model for predicting survival in patients with end-stage liver disease (MELD), the determination of degree of ascites, and the identification of other complications of cirrhosis.[7,8]

The MELD model, which employs a scoring system based on the serum bilirubin level, the serum creatinine level, and the international normalized ratio (INR) for prothrombin time, is now used for the allocation of donor organs in patients on liver transplant waiting lists in the United States.[9,10] MELD scores range from 6 to 40, with higher scores representing sicker patients; waiting time in the MELD system serves as a tiebreaker between patients with the same score.[11] The United Network for Organ Sharing provides on their Web site a resource for calculating MELD scores for individual patients.[12] The MELD score is able to accurately predict 3-month mortality of patients with chronic liver disease who are on the liver waiting list [*see Figure 1*].[11] Under the CTP system, sicker patients often did not survive the 1- to 2-year wait for a donor organ. By replacing the CTP scoring system with the MELD system and by eliminating patient waiting time as a major factor in the allocation of donor organs, the sickest patients are granted earlier access to donor organs. CTP classification and assessment of the complications of cirrhosis remain valuable tools for determining the optimal referral of patients to transplant centers.[8] For example, transplantation improves survival only in those patients with pretransplant CTP scores of 7 or greater. As a result, a CTP score of 7 or higher remains a reasonable threshold for referral of patients to centers for potential transplantation. New-onset ascites and development of other, more ominous complications of cirrhosis (e.g., spontaneous bacterial peritonitis and hepatorenal syndrome) indicate the need for immediate referral of patients to a transplant center.

OPERATIVE PROCEDURES

Most liver transplants are performed using a whole cadaveric liver placed in the orthotopic position. To increase the overall organ supply and especially to aid young children, for whom there is a chronic shortage of donor organs, a cadaveric liver can be divided into parts for more than one recipient [*see Figure 2*]. The same techniques can be used with living donors, with only part of the liver being removed for transplantation. Living related donor transplantation for children is a well-established procedure.[13] Living related donor transplantation for adults is also being performed at many transplantation centers, although donor safety remains a major concern.[14,15]

Liver transplantation is a complex, time-consuming operation that requires vascular reconstruction of the hepatic venous drainage to the inferior vena cava, to the hepatic artery, and to the portal vein. The hepatic vein of the donor organ is anastomosed to the inferior vena cava of the recipient; the donor hepatic artery is anastomosed to the recipient hepatic artery; and the portal vein is reconstructed by a vein graft or patch. Biliary reconstruction is usually accomplished by use of an end-to-end anastomosis of the proximal donor bile duct attached to the distal recipient duct; however, in recipients with diseased ducts, the donor duct is usually anastomosed to the jejunum by way of a Roux-en-Y loop.

A number of complications can be anticipated after liver transplantation, including perioperative and surgical complications, immunologic and infectious disorders, and a variety of medical complications.

COMPLICATIONS OF TRANSPLANTATION

Perioperative and Surgical Complications

The most serious immediate complication seen after liver transplantation is nonfunction of the transplanted liver, which occurs in 5% to 10% of cases. In these cases, patients fail to recover neurologic function; coagulopathy fails to improve spontaneously; and there is progressive jaundice and acidosis. Emergent retransplantation is the only recourse for these patients.

Other important surgical complications encountered after liver transplantation include hepatic artery thrombosis, portal vein thrombosis, and biliary tract complications (e.g., bile leaks and obstruction). Biliary tract complications are the most common; fortunately, most can be managed effectively with endoscopic techniques.[16] Hepatic artery thrombosis is a much more serious complication that can result in the need for retransplantation. A variety of surgical and nonsurgical factors are associated with increased risk of hepatic artery thrombosis. Included among the nonsurgical factors are immunologic status, hypercoagulable states, tobacco use, and cytomegalovius infection.[17]

Table 2 Child-Turcotte-Pugh (CTP) Classification*

	Score		
	1	*2*	*3*
Encephalopathy (grade)	None	1–2	3–4
Ascites	Absent	Slight	Moderate
Bilirubin (mg/dl)	1–2	2–3	> 3
Albumin (g/dl)	> 3.5	2.8–3.5	< 2.8
Prothrombin time (seconds prolonged)	1–4	4–6	> 6

*Minimum CTP score, 5 points; maximum CTP score, 15 points. CTP class A: 5 to 6 points. CTP class B: 7 to 9 points. CTP class C: 10 to 15 points.

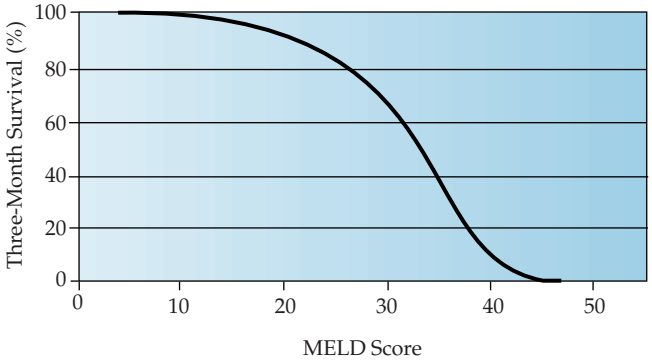


Figure 1 **Estimated 3-month survival as a function of the MELD score.[11]**

Immunologic Complications (Graft Rejection)

Two types of allograft rejection are seen after liver transplantation: cellular and ductopenic. Cellular rejection, which is usually manifested by elevated aminotransferase levels, is most commonly seen 6 to 10 weeks after transplantation. The diagnosis is confirmed by liver biopsy, which reveals cellular invasion of small bile ducts and vascular endothelium. Most patients respond rapidly to increased immunosuppression. Ductopenic rejection is a more indolent process that usually presents as progressive jaundice months to years after transplantation. Liver biopsies reveal gradual disappearance of intrahepatic bile ducts. Most patients with this condition ultimately require retransplantation.

Infectious Complications

Infections remain among the most serious complications encountered after liver transplantation. Many potential pathogens (e.g., *Pneumocystis carinii* and cytomegalovirus) can usually be prevented with aggressive prophylaxis. In the early postoperative period, the most common pathogens are fungal and nosocomial bacterial infections. Candidiasis and aspergillosis, which remain the most serious infections encountered after liver transplantation, often occur in malnourished, critically ill patients.[18] After the first few months following surgery, cytomegalovirus infection and recurrent hepatitis B and C virus infections become much more prominent. Infection with antimicrobial-resistant bacteria, such as methicillin-resistant *Staphylococcus aureus* (MRSA) or vancomycin-resistant *Enterococcus faecium* (VREF), is associated with increased postoperative mortality.[19]

Complications of Immunosuppressive Therapy

A number of immunosuppressive agents are now available for use after solid-organ transplantation. These agents include cyclosporine, tacrolimus, azathioprine, mycophenolate mofetil, sirolimus, and corticosteroids, as well as various polyclonal or monoclonal antilymphocyte preparations.[20] Most liver transplant recipients receive either cyclosporine or tacrolimus in combination with one or more other immunosuppressive agents.

Complications from cyclosporine and tacrolimus Cyclosporine and tacrolimus are both associated with a number of

complications, including renal dysfunction, neurologic toxicity, hypertension, pancreatic injury, and a variety of metabolic abnormalities. Renal failure occurs in 10% of patients who take cyclosporine or tacrolimus within 10 years after transplantation.[21] Patients with a glomerular filtration rate of less than 40 ml/min/kg body surface area 1 year after transplantation are at high risk for subsequent renal failure. Replacing calcineurin inhibitors with other immunosuppressive agents, such as mycophenolate mofetil, may improve renal function in some patients, although monotherapy with mycophenolate mofetil is associated with a slight increase in the risk of rejection.[22] Some patients receiving cyclosporine or tacrolimus experience severe neuropsychiatric complications, including psychosis, seizures, and apraxia.[23] Many patients who take these drugs complain of headaches, tremors, and severe musculoskeletal pains. Hypertension, which is quite common in patients who take either cyclosporine or tacrolimus, is thought to result from peripheral and renal vasoconstriction [24] Pancreatic damage with development of type 1 (insulin-dependent) diabetes mellitus is more common after the use of tacrolimus. Patients who take either drug can experience hyperkalemia, hyperuricemia, and elevated cholesterol and triglyceride levels.[25] Switching patients from cyclosporine to tacrolimus appears to reduce the severity of hyerlipidemias in some patients.[26,27] Treatment with low-dose cerivastatin or pravastatin also has been shown to significantly improve lipid profiles without adversely affecting liver function.[28] Cyclosporine, but not tacrolimus, is associated with gingival hyperplasia and excessive hair growth, particularly on the arms and face.

Complications from azathioprine and mycophenolate mofetil Azathioprine and mycophenolate mofetil can cause bone marrow depression with leukopenia, thrombocytopenia, and anemia. A number of patients who take mycophenolate mofetil also experience gastrointestinal side effects, including nausea, abdominal pain, and diarrhea. Long-term corticosteroid therapy is associated with obesity, hypertension, glucose intolerance, cataracts, osteoporosis, and hypercholesterolemia. Side effects of sirolimus include gastrointestinal symptoms and marked elevations of serum lipids, particularly when sirolimus is used in combination with cyclosporine.[29]

Complications from drug-drug interactions Both cyclosporine and tacrolimus are extensively metabolized in the liver, primarily via the cytochrome P-450 IIIA enzyme. As a result, both drugs are prone to numerous drug-drug interactions.[20] The most dramatic examples include interactions with ketoconazole and phenytoin. Ketoconazole inhibits the P-450 IIIA enzyme and can result in marked increases in circulating levels of cyclosporine and tacrolimus. In contrast, phenytoin induces the enzyme, resulting in enhanced metabolism of cyclosporine and tacrolimus and difficulty maintaining adequate circulating levels of both drugs. A number of other commonly used drugs have lesser but important effects on cyclosporine and tacrolimus metabolism. Awareness of these interactions is important in managing patients after transplantation.

Delayed complications from immunosuppressive drugs Most of the delayed complications seen after liver transplantation are secondary to the long-term use of immunosuppressive drugs. The most common of these complications include renal dysfunction, hypertension, diabetes, hyperkalemia and hyperuricemia, hyperlipidemia, obesity, and malignancies.[30] Hypertension can usually be effectively managed with a combination of calcium channel blockers and beta blockers.[31] Transient hyperkalemia can be managed effectively with sodium poly-

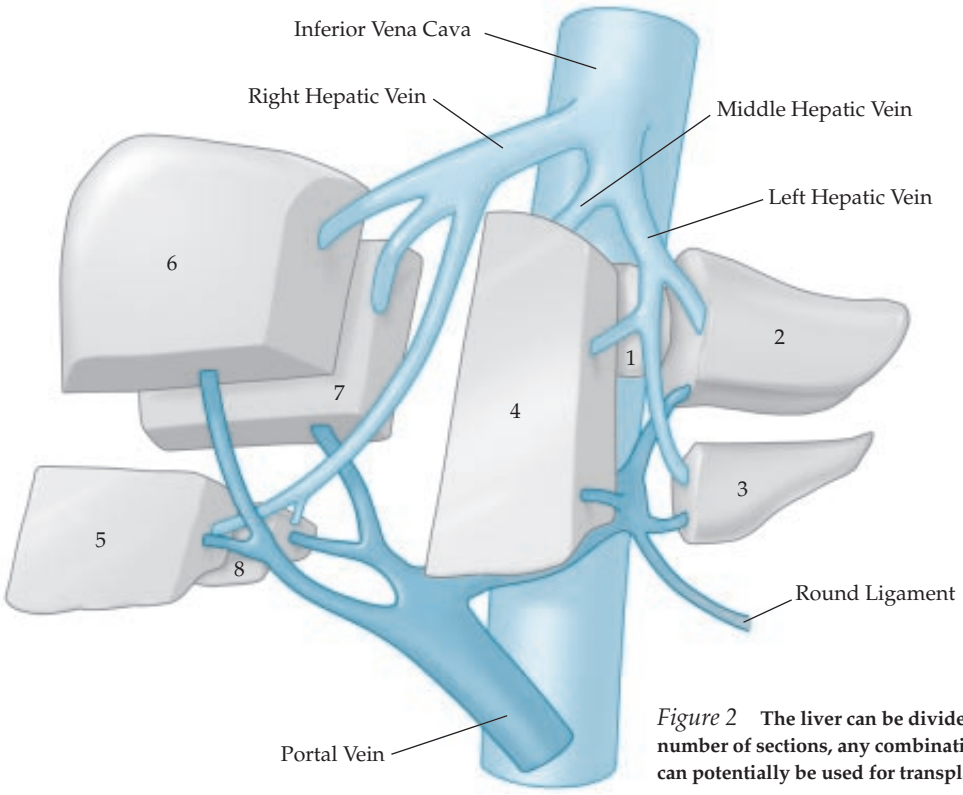


Figure 2 **The liver can be divided into a number of sections, any combination of which can potentially be used for transplantation.**

styrene sulfonate. If hyperkalemia is sustained, fludrocortisone can be used. Although many patients experience hyperuricemia after liver transplantation, very few experience gout. Treatment of gout is difficult because allopurinol can interfere with azathioprine metabolism, which can result in profound, life-threatening leukopenia, and because nonsteroidal anti-inflammatory drugs often worsen renal dysfunction. The necessity for treatment of hyperlipidemia after liver transplantation remains unclear. Obese patients who have undergone liver transplantation need a regular exercise program, limited caloric intake, and reduction or discontinuance of corticosteroids.[32] After age-related cardiovascular complications, malignancies are the leading cause of late death in liver transplant recipients. The most common tumors seen in these patients are lymphoproliferative disorders associated with chronic viral infections and skin cancers (e.g., squamous cell carcinoma and Kaposi sarcoma).[33] Many more recipients of liver transplantation are now receiving the bulk of their care from general internists, gastroenterologists, and primary care physicians. As a result, recognition of potential long-term complications and the need for appropriate immunizations and regular screening visits have become increasingly important.[34]

Disease-Specific Complications

Certain patients require specific management after liver transplantation because of potential disease-specific complications. For example, progressive liver disease can develop rapidly in patients with hepatitis B and can become fatal within a year after transplantation. However, if they are treated with aggressive antiviral therapy before and after transplantation, these patients have an excellent outcome, with minimal risk of severe recurrent disease.[35] Most potential transplant candidates now receive lamivudine therapy before the operation to reduce levels of circulating virus. Some patients with decompensated cirrhosis have such a dramatic response that transplantation can be postponed indefinitely.[36] After surgery, most patients now receive continuous treatment with hepatitis B immune globulin plus lamivudine to prevent recurrent disease.[37] There is concern about the emergence of YMDD (tyrosine-methionine-aspartate-aspartate) mutations after long-term therapy with lamivudine; however, patients resistant to lamivudine therapy have been successfully treated with adefovir dipivoxil, which has recently received approval by the Food and Drug Administration.[38]

Patients with chronic hepatitis C who undergo liver transplantation invariably have persistent infection after the operation. However, the long-term survival of these patients approximates that seen in other transplant recipients.[39] Nevertheless, some patients develop severe disease that can progress rapidly to cirrhosis and liver failure within the first few years after liver transplantation.[40] The optimal management of these patients, which may include pretransplantation and posttransplantation antiviral therapy and retransplantation, remains unclear. Because chronic liver disease secondary to hepatitis C is the leading indication for liver transplantation, management of such cases is an issue of increasing importance.

Patients with liver disease caused by sclerosing cholangitis often have associated inflammatory bowel disease. Although the transplant effectively addresses their liver disease, these patients remain at high risk for colon cancer. As a result, they require careful monitoring with colonoscopy and biopsies at least annually. If severe dysplasia is detected, these patients can be effectively treated with colectomy.

Liver transplantation has emerged as the optimal treatment for most patients with hepatocellular carcinoma (HCC). Excellent disease-free survival after transplantation is associated with patients who have (1) a single tumor no greater than 5 cm in diameter; (2) no more than three lesions, none of which is greater than 3 cm in diameter; (3) no radiographic evidence of vascular invasion; and (4) no evidence of metastases on head and chest CT scans and bone scans.[41] The major obstacles to effective utilization of transplantation have been the availability of donor organs and long delays before transplantation could be accomplished. The issue of long waiting periods before transplantation has been addressed in the new MELD system for allocation of donor organs, which gives patients with HCC who are optimal candidates for transplantation elevated scores to facilitate early transplantation.[10]

OUTCOMES AFTER TRANSPLANTATION

Survival after liver transplantation has improved steadily over the past 10 years. Most centers now report 1-year survival rates of 85% to 90% and 5-year survival rates of 75% to 80%.[1,2] During the same interval, the costs have progressively decreased as the result of reduced hospitalization for most patients.[42] The quality of life for most patients after successful transplantation is quite good. Most patients have been able to return to work, and physically active recipients have returned to vigorous endeavors, including marathon running and mountain climbing.

Pancreas Transplantation

Pancreas transplantation, which aims at providing physiologic insulin replacement, is a therapy that reliably achieves euglycemia in patients with type 1 diabetes mellitus. Islet transplantation (engrafting only the insulin-producing B cells of the pancreas) is an exciting alternative that is still in its clinical infancy.[43,44] One of the major difficulties preventing this technique from becoming widespread is that more than one pancreas is required to provide the recipient with enough islets to become euglycemic.

Since the first vascularized pancreas transplantation in 1966, more than 17,000 have been performed worldwide.[45,46] Approximately 83% of pancreas transplantations have been performed with simultaneous kidney transplantations from the same donors (i.e., simultaneous pancreas and kidney, or SPK, transplantation), with the recipients being those in whom renal failure is imminent or those who are already on dialysis.[46] Of the remaining transplantations, 12% have been performed as a pancreas after kidney (PAK) transplantation in diabetic patients who have had a previous kidney transplant, and 5% have been performed as a pancreas transplantation alone (PTA) in diabetic patients who have not yet experienced significant renal failure.[46]

The goals of pancreas transplantation are to improve the quality of life for patients with type 1 diabetes mellitus, reverse the metabolic abnormalities caused by the disease, and prevent the secondary complications of the disease. Despite these lofty goals, postoperative complications and the need for long-term immunosuppression have rendered pancreas transplantation controversial except in a select subpopulation of patients.

CANDIDATES FOR TRANSPLANTATION

During evaluation, it is essential to confirm the diagnosis of type 1 diabetes mellitus, to confirm that secondary complica-

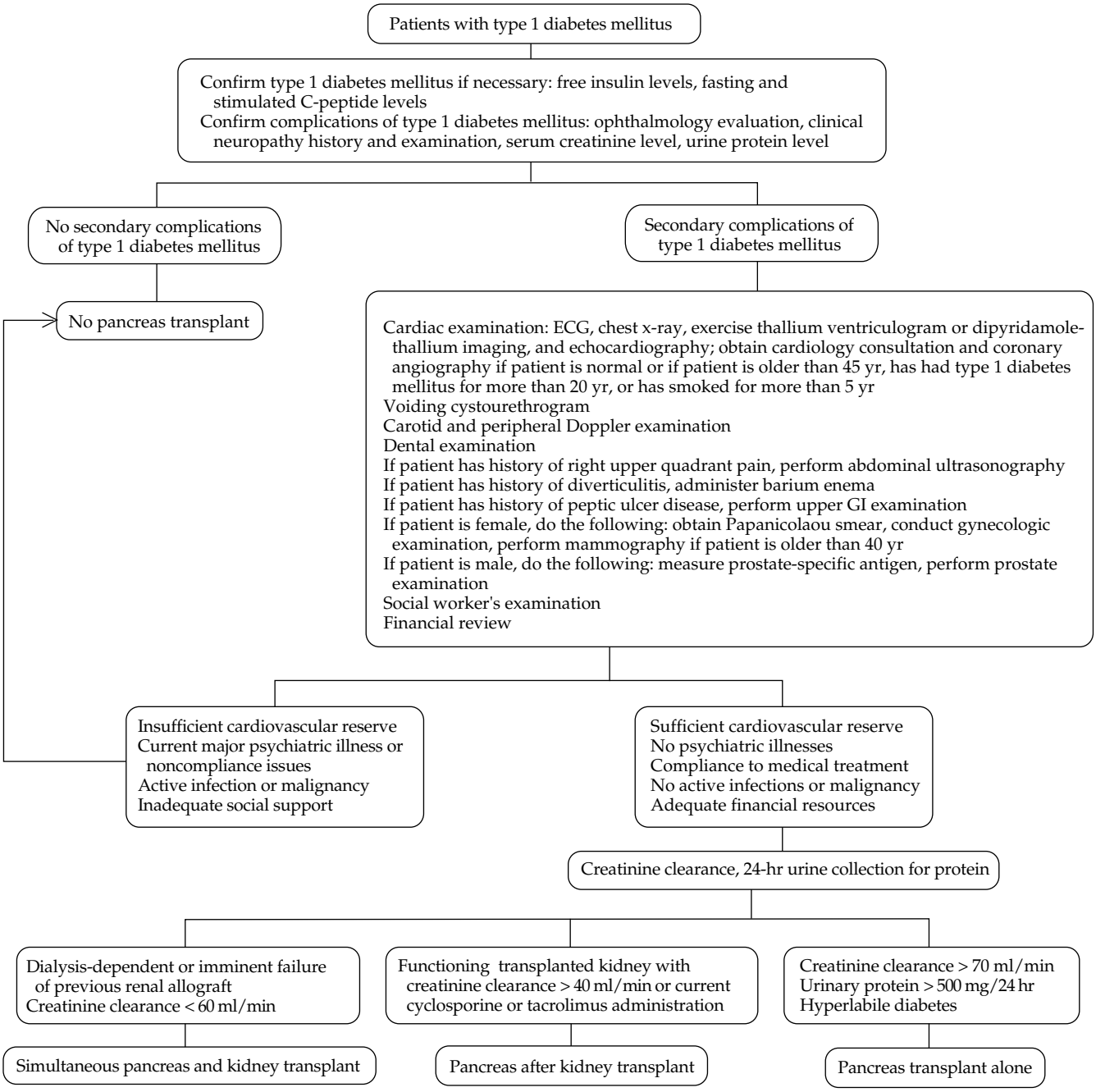


Figure 3 **Algorithm for evaluation of patients with type 1 (insulin-dependent) diabetes mellitus being considered for pancreas transplantation.**

tions of diabetes are present, to determine the candidate's ability to undergo a major operation, and to rule out any contraindications to the operation.[47] The type of procedure to be performed is determined by the renal function status of the potential recipient [*see Figure 3*].

CONTRAINDICATIONS TO TRANSPLANTATION

Patients with insufficient cardiovascular reserve (e.g., recent myocardial infarction), with a left ventricular ejection fraction below 50%, or with coronary angiographic evidence of significant uncorrectable coronary artery disease should not undergo pancreas transplantation[47] [*see Figure 3*]. Unnecessary loss of pancreas grafts is avoided by excluding patients with current major psychiatric illness or evidence of significant noncompliance. In addition, transplantation should not be considered in patients with an active infection or malignancy.

Other contraindications are controversial and depend on the individual transplantation center. Extremity amputations because of vascular disease usually indicate severe generalized vasculopathy and suggest a condition in which pancreas transplantation would not be beneficial. Patients whose weight is greater than 130% of their ideal body weight often have insulin resistance and, as a result, are not helped by transplantation.[47] Continued cigarette use often indicates poor compliance in patients who have already been strongly encouraged to stop smoking. Severe neurogenic bladder dysfunction usually pre-

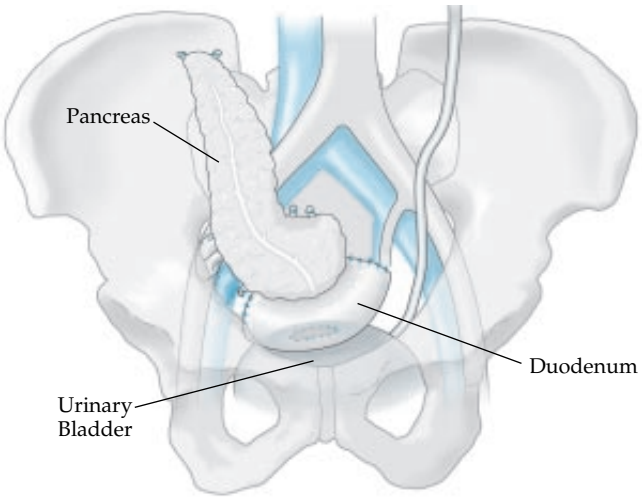


Figure 4 **During the pancreas transplantation procedure, the pancreaticoduodenal graft is anastomosed to the posterolateral aspect of the urinary tract. The arteries and veins of the graft are anastomosed to the common iliac vessels.**

dicts a complicated postoperative course and is considered a contraindication at some centers.

OPERATIVE PROCEDURES

The operation of pancreas transplantation includes placement of the pancreas graft, usually in the right lower quadrant, with the reconstructed arteries of the pancreas anastomosed to the common iliac artery [*see Figure 4*].[48] To provide drainage for pancreatic exocrine excretions, the increasingly favored procedure is to anastomose the duodenum of the graft to the recipient's small bowel as opposed to the mobilized urinary bladder.[46,48,49] The venous drainage of the graft is achieved by anastomosing the portal vein either to the mobilized common iliac vein or to the portal vein.[48,49] In SPK transplantation, the kidney is placed in the left lower quadrant.

PERIOPERATIVE CARE

In the immediate postoperative period, specific care should be directed toward monitoring cardiovascular function.[48] Insulin infusions are generally given for a few days to rest the transplanted islets. Because many patients have some form of diabetic gastropathy, a nasogastric tube is required for 4 to 7 days postoperatively. A urinary catheter is required for an extended period to reduce the risk of complications from neurogenic bladder dysfunction.

COMPLICATIONS OF TRANSPLANTATION

Surgical Complications

Vascular thrombosis, the leading cause of nonimmunologic pancreatic graft failure, occurs in 12% of transplants (5% arterial and 7% venous), usually within the first week after transplantation.[50] With rare exception, salvage of the graft is impossible, and delaying removal of the nonfunctional graft can risk the recipient's life.

Immunologic Complications (Graft Rejection)

Rejection is the leading cause of graft loss after a successful pancreas transplant procedure.[46] Urinary amylase determinations remain the most useful biochemical parameter for indicating rejection in bladder-drained allografts; however, the gold standard for diagnosis of rejection is histopathologic evaluation of the graft.[51] Rejection can be confirmed histologically because tissue samples of the graft can be obtained by percutaneous biopsies.[51]

Complications of Medical Therapy

Dehydration and metabolic acidosis are the next most common adverse effects that follow bladder-drained pancreas transplantations. There is obligatory loss of pancreatic secretions rich in sodium and bicarbonate into the urinary tract. All transplant recipients must increase their fluid and salt intake, and most require continuous oral bicarbonate supplementation. The dehydration and metabolic acidosis can be avoided by use of enteric-drained pancreas transplantation, which is a major reason for the popularity of this technique.[49]

Graft pancreatitis, which is also a common side effect, is manifested by hyperamylasemia, abdominal pain, and graft tenderness. Most bladder-drained grafts can be successfully treated with bladder drainage by catheter for a few days, although a few require conversion to enteric drainage.

Tacrolimus has become the mainstay of immunosuppressive therapy for pancreas transplantation, permitting steroid withdrawal in selected patients.[52] Complications associated with immunosuppressive therapy are similar to those associated with hepatic transplantation.

OUTCOMES AFTER TRANSPLANTATION

Metabolic Outcomes

Successful pancreas transplantation results in normalization of glucose and hemoglobin A_{1c} levels.[47] Glucose tolerance tests are normal or near normal; however, insulin levels are much higher than normal in recipients of pancreas transplantation. The systemic venous drainage of the graft causes elevated plasma levels of insulin, which is known to be a potent regulator of plasma lipoprotein metabolism. As a result, SPK transplantation recipients have a more favorable lipid profile than patients with type 1 diabetes mellitus who have kidney transplants.[53]

Effect on Disorders Associated with Type 1 Diabetes Mellitus

Diabetic nephropathy A transplanted pancreas can prevent or reduce the nephropathy that eventually develops in diabetic patients with a kidney graft. Pancreas transplantation also reduces the risk of recurrent diabetic nephropathy in the kidneys of SPK transplant recipients.[54] In contrast, diabetic glomerular lesions are improved by PTA, but reversal requires more than 5 years of normoglycemia.[55]

Retinopathy Pancreas transplantation appears to have a stabilizing effect on retinopathy. In one study, most diabetic patients who underwent SPK transplantations experienced no progression of diabetic retinopathy but did require maintenance laser therapy.[56]

Neuropathy Reestablishment of the euglycemic state by successful pancreas transplantation halts or reverses diabetic neuropathy. In patients who undergo kidney transplantations, the neuropathy initially improves with elimination of uremia but tends to worsen during longer follow-up. In patients who

also receive pancreas transplants, the severe and prevalent peripheral neuropathy shows a rapid initial improvement followed by stabilization.[57]

Vasculopathy Pancreas transplantation has at least a partial beneficial effect on the macroangiopathy of the carotid artery in patients with type 1 diabetes mellitus.[58] Also, compared with type 1 diabetes mellitus patients who receive kidney transplants, SPK recipients show improvement of diabetic microangiopathy.[56,59] The progression of coronary atherosclerosis in patients with functioning pancreas grafts is reduced.

Survival Outcomes

Patient survival exceeds 95% at 1 year and 90% at 3 years. Graft survival (complete insulin independence) exceeds 85% at 1 year and 75% at 3 years.[46] Patients who undergo SPK transplantation have a markedly improved 10-year survival, compared with diabetic patients who undergo kidney transplantation alone.[60,61]

Quality of Life

Quality of life in terms of general health perception, physical ability, and sexual activity is higher for SPK transplant recipients than for patients with type 1 diabetes mellitus who receive kidney transplants and is far higher for SPK transplant recipients than for patients who remain on hemodialysis.[62]

The authors have no commercial relationships with manufacturers of products or providers of services described in this chapter.

References

1. Keefe EB: Liver transplantation: current status and novel approaches to liver replacement. Gastroenterology 120:749, 2001

2. Carithers RL Jr: Liver transplantation: American Association for the Study of Liver Diseases practice guidelines. Liver Transpl 6:122, 2000

3. Kim WR, Krowka MJ, Plevak DJ, et al: Accuracy of Doppler echocardiography in the assessment of pulmonary hypertension in liver transplant candidates. Liver Transpl 6:453, 2000

4. Krowka MJ, Plevak DJ, Findlay JY, et al: Pulmonary hemodynamics and perioperative cardiopulmonary-related mortality in patients with portopulmonary hypertension undergoing liver transplantation. Liver Transpl 6:443, 2000

5. Nair S, Verma S, Thuluvath PJ: Obesity and its effect on survival in patients undergoing orthotopic liver transplantation in the United States. Hepatology 35:105, 2002

6. Manzanet G, Sanjuan F, Orbis P, et al: Liver transplantation in patients with portal vein thrombosis. Liver Transpl 7:125, 2001

7. Malinchoc M, Kamath PS, Gordon FD, et al: A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. Hepatology 31:864, 2000

8. Lucey MR, Brown KA, Everson GT, et al: Minimal criteria for placement of adults on the liver transplant waiting list: a report of a national conference organized by the American Society of Transplant Physicians and the American Association for the Study of Liver Diseases. Liver Transpl Surg 3:628, 1997

9. Kamath PS, Wiesner RH, Malinchoc M, et al: A model to predict survival in patients with end-stage liver disease. Hepatology 33:464, 2001

10. Wiesner RH, McDiarmid SV, Kamath PS, et al: MELD and PELD: application of survival models to liver allocation. Liver Transpl 7:567, 2001

11. Wiesner R, Edwards E, Freeman R, et al: Model for end-stage liver disease (MELD) and allocation of donor livers. Gastroenterology 124:91, 2003

12. United Network for Organ Sharing (UNOS): Resources: MELD/PELD calculator.

http://222.unos.org/resources/MeldPeldCalculator.asp

13. Otte JB, Ville-de-Goyet J, Reding R, et al: Pediatric liver transplantation: from the full-size liver graft to reduced, split, and living related liver transplantation. Pediatr Surg Int 13:308, 1998

14. Trotter JF, Wachs M, Everson GT, et al: Adult-to-adult transplantation of the right hepatic lobe from a living donor. N Engl J Med 346:1074, 2002

15. Surman OS: The ethics of partial-liver donation. N Engl J Med 346:1038, 2002

16. Moser MA, Wall WJ: Management of biliary problems after liver transplantation. Liver Transpl 7(suppl 1):S46, 2001

17. Pastacaldi S, Teixeira R, Montalto P, et al: Hepatic artery thrombosis after orthotopic liver transplantation: a review of nonsurgical causes. Liver Transpl 7:75, 2001

18. Rabkin JM, Oroloff SL, Corless CL, et al: Association of fungal infection and increased mortality in liver transplant recipients. Am J Surg 179:426, 2000

19. Singh N, Gayowski T, Rihs JD, et al: Evolving trends in multiple-antibiotic-resistant bacteria in liver transplant recipients: a longitudinal study of antimicrobial susceptibility patterns. Liver Transpl 7:22, 2001

20. Levy GA: Long-term immunosuppression and drug interactions. Liver Transpl 7(suppl 1):S53, 2001

21. Cohen AJ, Stegall MD, Rosen CB, et al: Chronic renal dysfunction late after liver transplantation. Liver Transpl 8:916, 2002

22. Schlitt HJ, Barkmann A, Boker KH, et al: Replacement of calcineurin inhibitors with mycophenolate mofetil in liver-transplant recipients with renal dysfunction: a randomized controlled study. Lancet 357:587, 2001

23. Beresford TP: Neuropsychiatric complications of liver and other solid organ transplantation. Liver Transpl 7(suppl 1):S36, 2001

24. Textor SC, Taler SJ, Canzanello VJ: Posttransplantation hypertension related to calcineurin inhibitors. Liver Transpl 6:521, 2000

25. Charco R, Cantarell C, Vargas V, et al: Serum cholesterol changes in long-term survivors of liver transplantation: a comparison between cyclosporine and tacrolimus therapy. Liver Transpl Surg 5:204, 1999

26. Manzarbeitia C, Reich DJ, Rothstein KD, et al: Tacrolimus conversion improves hyperlipidemic states in stable liver transplant recipients. Liver Transpl 7:93, 2001

27. Neal DA, Gimson AE, Gibbs P, et al: Beneficial effects of converting liver transplant recipients from cyclosporine to tacrolimus on blood pressure, serum lipids, and weight. Liver Transpl 7:533, 2001

28. Zachoval R, Gerbes AL, Schwandt P, et al: Short-term effects of statin therapy in patients with hyperlipoproteinemia after liver transplantation: results of a randomized cross-over trial. J Hepatol 35:86, 2001

29. Trotter JF, Wachs ME, Trouillot TE, et al: Dyslipidemia during sirolimus therapy in liver transplant recipients occurs with concomitant cyclosporine but not tacrolimus. Liver Transpl 7:401, 2001

30. Sheiner PA, Magliocca JF, Bodian CA, et al: Long-term medical complications in patients surviving > or = 5 years after liver transplant. Transplantation 69:781, 2000

31. Gonwa TA: Hypertension and renal dysfunction in long-term liver transplant recipients. Liver Transpl 7(suppl 1):S26, 2001

32. Reubin A: Long-term management of the liver transplant patient: diabetes, hyperlipidemia, and obesity. Liver Transpl 7(suppl 1):S13, 2001

33. Fung JJ, Jain A, Kwak EJ, et al: De novo malignancies after liver transplantation: a major late cause of late death. Liver Transpl 7(suppl 1):S109, 2001

34. McCashland TM: Posttransplantation care: role of the primary care physician versus transplant center. Liver Transpl 7(suppl 1):S2, 2001

35. Steinmuller T, Seehofer D, Rayes N, et al: Increasing applicability of liver transplantation for patients with hepatitis B-related liver disease. Hepatology 35:1528, 2002

36. Fontana RJ, Hann HW, Perrillo RP, et al: Determinants of early mortality in patients with decompensated chronic hepatitis B treated with antiviral therapy. Gastroenterology 123:719, 2002

37. Vargas HE, Dodson FS, Rakela J: A concise update on the status of liver transplantation for hepatitis B virus: the challenges in 2002. Liver Transpl 8:2, 2002

38. Terrault NA: Treatment of recurrent hepatitis B infection in liver transplant recipients. Liver Transpl 8(suppl 1):S74, 2002

39. Shuhart MC, Bronner MP, Gretch DR, et al: Histological and clinical outcome after liver transplantation for hepatitis C. Hepatology 26:1646, 1997

40. Berenguer M, Prieto M, Rayon JM, et al: Natural history of clinically compensated hepatitis C virus-related graft cirrhosis after liver transplantation. Hepatology 32:852, 2000

41. Bruix J, Llovet JM: Prognostic prediction and treatment strategy in hepatocellular carcinoma. Hepatology 35:519, 2002

42. Best JH, Veenstra DL, Geppert J: Trends in expenditures for Medicare liver transplant recipients. Liver Transpl 7:858, 2001

43. Worldwide Transplant Center Directory: Islet transplants. Clin Transpl 383: 2001

44. Ryan EA, Lakey JR, Paty BW, et al: Successful islet transplantation: continued insulin reserve provides long-term glycemic control. Diabetes 51:2148, 2002

45. Kelly WD, Lillehei RC, Merkel FK, et al: Allotransplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. Surgery 61:827, 1967

46. University of Minnesota: International Pancreas Transplant Registry. IPTR Newsletter 14:1, 2002

47. Robertson RP, Sutherland DE, Lanz KJ: Normoglycemia and preserved insulin secretory reserve in diabetic patients 10–18 years after pancreas transplantation. Diabetes 48:1737, 1999

48. Perkins JD, Fromme GA, Narr BJ, et al: Pancreas transplantation at Mayo: II. Operative and perioperative management. Mayo Clin Proc 65:483, 1990

49. Stratta RJ, Gaber AO, Shokouh-Amiri MH, et al: A prospective comparison of systemic-bladder versus portal-enteric drainage in vascularized pancreas transplantation. Surgery 127:217, 2000

50. Troppmann C, Gruessner AC, Benedetti E, et al: Vascular graft thrombosis after pancreatic transplantation: univariate and multivariate operative and nonoperative risk factor analysis. J Am Coll Surg 182:285, 1996

51. Laftavi MR, Gruessner AC, Bland BJ, et al: Significance of pancreas graft biopsy in detection of rejection. Transplant Proc 30:642, 1998

52. Jordan ML, Shapiro R, Gritsch HA, et al: Long-term results of pancreas transplantation under tacrolimus immunosuppression. Transplantation 67:266, 1999

53. Foger B, Konigsrainer A, Palos G, et al: Effect of pancreas transplantation on lipoprotein lipase, postprandial lipemia, and HDL cholesterol. Transplantation 58:899, 1994

54. Wilczek HE, Jaremko G, Tyden G, et al: Evolution of diabetic nephropathy in kidney grafts. Transplantation 59:51, 1995

55. Fioretto P, Steffes MW, Sutherland DE, et al: Reversal of lesions of diabetic nephropathy after pancreas transplantation. N Engl J Med 339:69, 1998

56. Pearce IA, Ilango B, Sells RA, et al: Stabilisation of diabetic retinopathy following simultaneous pancreas and kidney transplant. Br J Ophthalmol 84:736, 2000

57. Nankivell BJ, Al Harbi IS, Morris J, et al: Recovery of diabetic neuropathy after pancreas transplantation. Transplant Proc 29:658, 1997

58. Larsen JL, Ratanasuwan T, Burkman T, et al: Carotid intima media thickness decreases after pancreas transplantation. Transplantation 73:936, 2002

59. Jukema JW, Smets YF, van der Pijl JW, et al: Impact of simultaneous pancreas and kidney transplantation on progression of coronary atherosclerosis in patients with end-stage renal failure due to type 1 diabetes. Diabetes Care 25:906, 2002

60. Tyden G, Tollemar J, Bolinder J: Combined pancreas and kidney transplantation improves survival in patients with end-stage diabetic nephropathy. Clin Transplant 14:505, 2000

61. Reddy KS, Stablein D, Taranto S, et al: Long-term survival following simultaneous kidney-pancreas transplantation versus kidney transplantation alone in patients with type 1 diabetes mellitus and renal failure. Am J Kidney Dis 41:464, 2003

62. Gross CR, Limwattananon C, Matthees B, et al: Impact of transplantation on quality of life in patients with diabetes and renal dysfunction. Transplantation 70:1736, 2000

Acknowledgment

Figures 2 and 4 Tom Moore.

I APPROACH TO HEMATOLOGIC DISORDERS

David C. Dale, m.d.

Hematology deals with the normal functions and disorders of the formed elements in the blood (i.e., erythrocytes, leukocytes, and platelets) and the plasma factors governing hemostasis. The blood sustains life by transporting oxygen and essential nutrients, removing waste, and delivering the humoral and cellular factors necessary for host defenses. Platelets and coagulation factors, together with vascular endothelial cells, maintain the integrity of this system. Some hematologic disorders such as anemia, leukocytosis, and bleeding are quite common, occurring secondary to infectious, inflammatory, nutritional, and malignant diseases. Other disorders, including the hematologic malignancies, are far less common. This subsection presents the general principles for understanding the hematopoietic system [*see other subsections under Hematology for a more detailed description of the pathophysiology of specific hematologic diseases and their treatment*].

Hematopoiesis

Hematopoiesis begins in the fetal yolk sac and later occurs predominantly in the liver and the spleen.[1] Recent studies demonstrate that islands of hematopoiesis develop in these tissues from hemangioblasts, which are the common progenitors for both hematopoietic and endothelial cells.[2] These islands then involute as the marrow becomes the primary site for blood cell formation by the seventh month of fetal development.[3] Barring serious damage, such as that which occurs with myelofibrosis or radiation injury, the bone marrow remains the site of blood cell formation throughout the rest of life. In childhood, there is active hematopoiesis in the marrow spaces of the central axial skeleton (i.e., the ribs, vertebrae, and pelvis) and the extremities, extending to the wrists, ankles, and the calvaria. With normal growth and development, hematopoiesis gradually withdraws from the periphery. This change is reversible, however; distal marrow extension can result from intensive stimulation, as occurs with severe hemolytic anemias, long-term administration of hematopoietic growth factors, and hematologic malignancies. The term medullary hematopoiesis refers to the production of blood cells in the bone marrow; the term extramedullary hematopoiesis indicates blood cell production outside the marrow in the spleen, liver, and other locations.

Organization of Hematopoietic Tissues

In its normal state, the medullary space in which hematopoietic cells develop contains many adipocytes and has a rich vascular supply [*see Figure 1*].[4] Vascular endothelial cells, marrow fibroblasts, and stromal cells are important sources of the matrix proteins that provide structure to the marrow space; these cells also produce the hematopoietic growth factors and chemokines that regulate blood cell production.[5] The vascular endothelial cells also form an important barrier that keeps immature cells in the marrow and permits mature hematopoietic elements to enter the blood. The abundant adipocytes may influence hematopoiesis by serving as a localized energy source, by synthesizing growth factors, and by affecting the metabolism of androgens and estrogens.[6] Marrow macrophages remove effete or apoptotic cells and clear the blood of foreign materials when they enter the marrow. Osteoblasts and osteoclasts maintain and remodel the surrounding cancellous bone and the calcified lattice, which crisscrosses the marrow space.[4]

The thymus, lymph nodes, mucosa-associated lymphatic tissues, and the spleen have multiple hematopoietic functions. Early in development, they are major sites of hematopoiesis. In adulthood, they are principally sites of lymphocyte development, processing of antigens, development of effector T cells, and antibody production [*see 6 Immunology/Allergy*]. In leukemia and the myeloproliferative disorders, the size and cellular architecture of these tissues are deranged, leading to many of the clinical manifestations of these disorders [*see 12:XVI Acute Leukemia and 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*].

Hematopoietic Stem Cells

All cells of the hematopoietic system are derived from common precursor cells, the hematopoietic stem cells.[7] These cells are difficult to identify, in part because they normally represent only about 0.05% of marrow cells. Through self-renewal, this population is maintained at a constant level.[8] Through the use of monoclonal antibodies that recognize specific cell surface molecules expressed selectively on developing hematopoietic cells and other specialized techniques, the stem cells can now be separated from other marrow cells. With these methods, very primitive hematopoietic stem cells have been found to be positive for c-kit and thy-1 but negative for CD34, CD38, CD33, and HLA-DR.[8] For clinical purposes, $CD34^+$ progenitor cell populations, which contain stem cells and some more mature cells, are often used for hematopoietic stem cell transplantation[9] [*see 5:XI Hematopoietic Stem Cell Transplantation*].

Stem cells give rise to daughter cells, which undergo irreversible differentiation along various hematopoietic cell lineages [*see Figure 2*].[10] Many aspects of the earliest steps in this differentiation process are not well understood. With lineage commitment, however, differentiation, maturation, and release of cells to the blood come under the control of well-defined hematopoietic growth factors. In the early phases of differentiation, the regulatory roles played by these growth factors overlap.[11] Later in development, some growth factors are lineage specific, meaning that they govern the maturation and deployment of single lineages. Erythropoietin (EPO) (erythrocytes), thrombopoietin (TPO) (platelets), granulocyte colony-stimulating factor (G-CSF) (neutrophils), and macrophage colony-stimulating factor (M-CSF) (monocytes) are the best-characterized lineage-specific factors.

Hematopoietic Growth Factors

The hematopoietic growth factors, also referred to as hematopoietic cytokines, are a family of glycoproteins produced in the bone marrow by endothelial cells, stromal cells, fibroblasts, macrophages, and lymphocytes; they are also produced at distant sites, from which they are transported to the marrow through the blood [*see Table 1*]. The naming of these factors is somewhat confusing. Erythropoietin and throm-

bopoietin derive part of their names from the Greek word poiesis, meaning "to make." The colony-stimulating factors were first recognized because of their capacity to stimulate early hematopoietic cells to grow into clusters and large colonies in tissue culture systems. The term interleukin denotes factors that are produced by leukocytes and that affect other leukocytes. This is a large family of factors that predominantly govern lymphocytopoiesis, but many members also have broad effects on other lineages. The discovery of new growth factors and of the biologic consequences of deficiencies or excesses of these factors continues to evolve rapidly.

Hematopoietic cells have distinctive patterns of expression of growth factor receptors, and the patterns evolve as the cells differentiate [*see Figure 2*].[11] Each growth factor binds only to its specific receptor.[12] It is now known that some growth factors share components of the receptor (e.g., interleukin-3 [IL-3], IL-5, and granulocyte-macrophage colony-stimulating factor [GM-CSF] share a common β chain of their receptor); specificity comes from other unique or private components of the receptor. Binding of the ligand to the receptor leads to a conformational change, activation of intracellular kinases, and, ultimately, the triggering of cell proliferation.[13,14] For some growth factors, these pathways are well defined; for others, the pathways are still unclear [*see Figure 3*].

Hematopoietic growth factors not only stimulate cell proliferation but also prolong cell survival; that is, they have antiapoptotic effects.[15] For some lineages, such as neutrophils and monocytes, growth factor receptors occur on fully mature cells; exposure of these cells to the factors primes the cells for an enhanced responsiveness to bacteria or other stimulators of their metabolic activity. Thus, for cells of the neutrophil lineage, the growth factors G-CSF and GM-CSF can stimulate early hematopoietic cell proliferation, increase the number of cells produced by the marrow, prolong the life span of these cells, and augment cell functions.[16]

Erythropoietin

The peritubular interstitial cells located in the inner cortex and outer medulla of the kidney are the primary site for erythropoietin production.[17] In response to hypoxia, transcription of the erythropoietin gene in these cells increases, resulting in increased secretion of erythropoietin. The protein is then transported through the blood to the marrow to stimulate erythropoiesis. With renal failure, erythropoietin production is severely impaired. In infections and many chronic inflammatory conditions, the erythropoietin response is blunted, and erythropoietin levels are low.[18]

Erythropoietin is a glycosylated protein that modulates erythropoiesis by affecting several steps in red cell development. The most primitive identifiable erythroid cells, the burst-forming unit–erythroid cells (BFU-E), are relatively insensitive to erythropoietin. More mature cells, the colony-forming unit–erythroid cells (CFU-E), are very sensitive. Erythropoietin treatment prolongs survival of erythroid precursors, shortens the time between cell divisions, and increases the number of cells produced from individual precursors.[19]

Erythropoietin can be administered intravenously or subcutaneously for the treatment of anemia caused by inadequate endogenous production of erythropoietin.[20] Treatment is maximally effective when the marrow has a generous supply of iron and other nutrients, such as cobalamin and folic acid.[21] For patients with renal failure, who have very low erythropoietin levels, the starting dosage is 50 to 100 units S.C. three times a week. The most easily monitored immediate effect of increased endogenous or exogenous erythropoietin is an increase in the blood reticulocyte count. Normally, as red cell precursors mature, the cells extrude their nucleus at the normal blast stage. The resulting reticulocytes, identified by the supravital stain of their residual ribosomes, persist for about 3 days in the marrow and 1 day in the blood. Erythropoietin shortens the transit time through the marrow, leading to an increase in the number and proportion of blood reticulocytes within a few days.

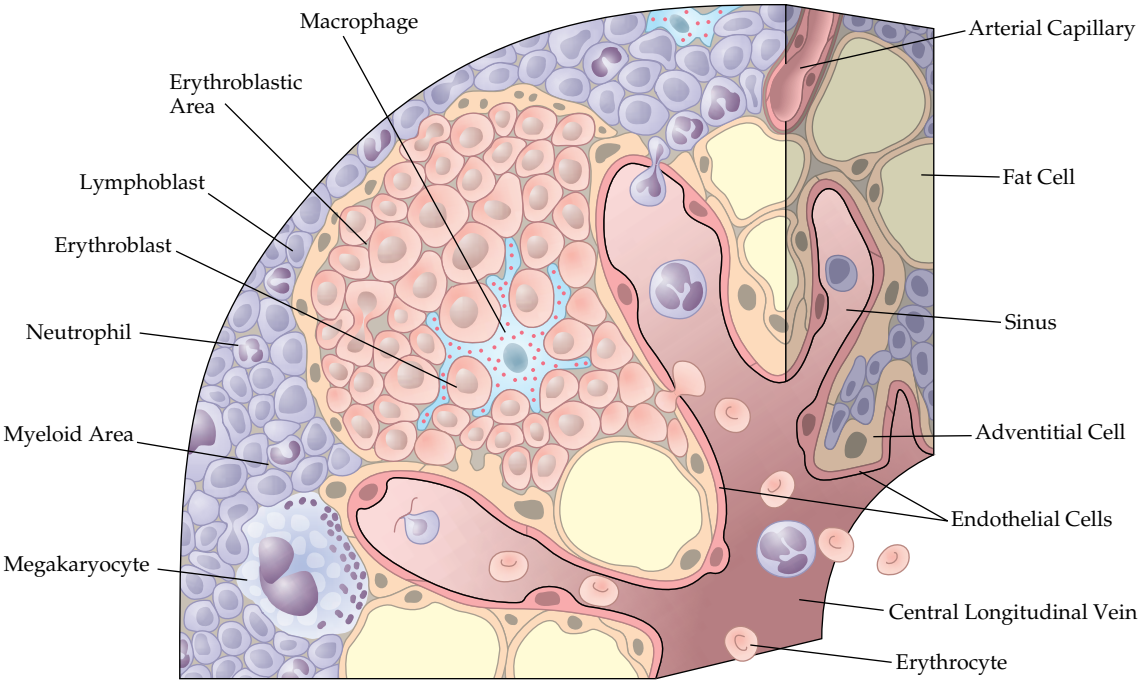


Figure 1 **The architecture of the bone marrow showing the various types of cells.**

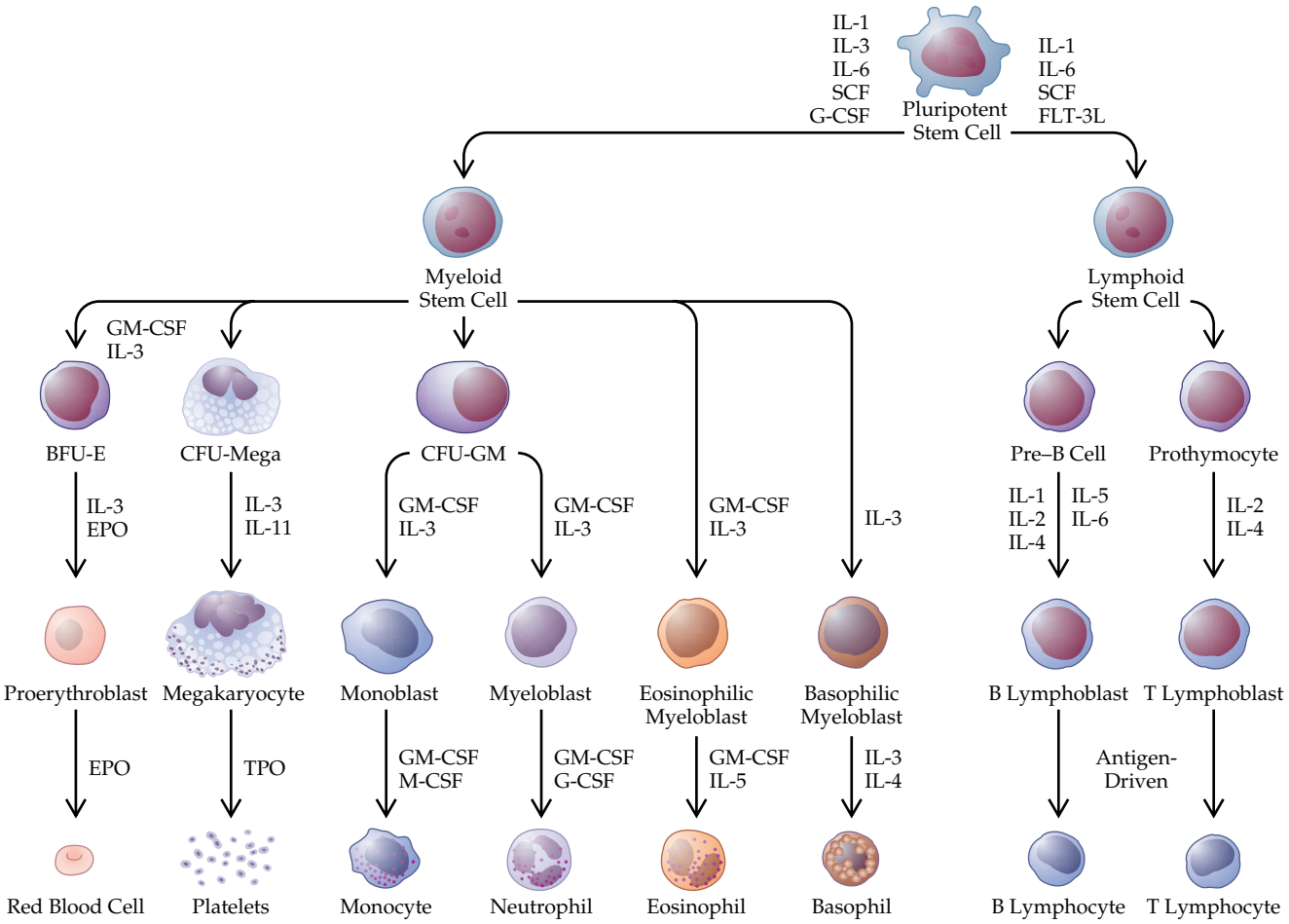


Figure 2 **The pattern for development of various types of blood cells in the bone marrow. (BFU-E—burst-forming unit–erythroid; CFU-GM—colony-forming unit–granulocyte-macrophage; CFU-mega—colony-forming unit–megakaryocyte; EPO—erythropoietin; EPOR—surface component of the erythropoietin receptor; FLT-3L—fms-like tyrosine kinase 3 ligand; G-CSF—granulocyte colony-stimulating factor; GM-CSF—granulocyte-macrophage colony-stimulating factor; IL—interleukin; M-CSF—macrophage colony-stimulating factor; TPO—thrombopoietin; SCF—stem cell factor)**

In some conditions, particularly chronic inflammatory diseases, the effectiveness of erythropoietin can be predicted from measurement of the serum erythropoietin level by immunoassay.[17,18] It may be cost-effective to measure the level before initiating treatment in patients with anemia attributable to suppressed erythropoietin production, such as patients with HIV infection, cancer, and chronic inflammatory diseases. Several studies have shown that erythropoietin treatment decreases the severity of anemia and improves the quality of life for these patients. In patients with anemia caused by cancer and cancer chemotherapy, current guidelines recommend erythropoietin treatment if the hemoglobin level is less than 10 g/dl.[22]

Thrombopoietin

The development of megakaryocytes from hematopoietic stem cells and the level of platelets in the blood are governed by thrombopoietin.[23] Thrombopoietin is produced primarily by the liver and is similar to erythropoietin in structure. However, thrombopoietin has broader biologic effects than erythropoietin, stimulating the proliferation and release of hematopoietic stem cells from the bone marrow and prolonging survival of these cells.[24] Thrombopoietin signals through its specific receptor, called cMpL, expressed on hematopoietic cells. Plasma thrombopoietin levels are inversely related to the blood platelet count. Deficiencies in thrombopoietin cause thrombocytopenia, and excesses in thrombopoietin cause thrombocytosis. Recombinant human thrombopoietin is being studied for use in the treatment of thrombocytopenia of diverse causes. Thrombopoietin is not yet approved for clinical use.[24]

Granulocyte Colony-Stimulating Factor

G-CSF is a glycosylated protein produced by monocytes, macrophages, fibroblasts, stromal cells, and endothelial cells throughout the body.[25] It stimulates the growth and differentiation of neutrophils both in vitro and in vivo. G-CSF levels are normally very low or undetectable but increase with bacterial infections or after administration of bacterial endotoxin.[16] G-CSF (the synthesized form is known as filgrastim or lenograstim) administration causes a dose-dependent increase in the blood neutrophil count in healthy persons. Studies in animals have shown that G-CSF deficiency causes neutropenia.[26] As with erythropoietin, administration of G-CSF leads to an acceleration in the development of neutrophils in the bone marrow, with the neutrophils shifting at an earlier stage than normal from the marrow to the blood.[27]

Table 1 Hematopoietic Growth Factors

Factor	*Other Names*	*Cell Source*	*Chromosome Location*	*Function*
EPO	Erythropoietin	Juxtaglomerular cells	7q	Stimulates erythrocyte formation and release from marrow
TPO	Thrombopoietin; megakaryocyte growth and development factor (MGDF)	Hepatocytes, renal and endothelial cells, fibroblasts	3q27	Stimulates megakaryocyte proliferation and platelet formation
G-CSF	Granulocyte colony-stimulating factor; filgrastim; lenograstim	Endothelial cells, monocytes, fibroblasts	17q11.2-q21	Stimulates formation and function of neutrophils
GM-CSF	Granulocyte-macrophage colony-stimulating factor	T cells, monocytes, fibroblasts	5q23-q31	Stimulates formation and function of neutrophils, monocytes, and eosinophils
M-CSF	Macrophage colony-stimulating factor; colony stimulating factor–1 (CSF-1)	Endothelial cells, macrophages, fibroblasts	5q33.1	Stimulates monocyte formation and function
IL-1α and IL-1β	Interleukin-1α and -1β, endogenous pyrogen hemopoietin-1	Monocytes, keratinocytes, endothelial cells	2q13	Proliferation of T cells, B cells, and other cells; induces fever and catabolism
IL-2	T cell growth factor	T cells ($CD4^+$, $CD8^+$), large granular lymphocytes (natural killer, or NK, cells)	4q	T cell proliferation, antitumor and antimicrobial effects
IL-3	Multi–colony stimulating factor; mast cell growth factor	Activated T cells; large granular lymphocytes (NK cells)	5q23-q31	Proliferation of early hematopoietic cells
IL-4	B cell growth factor; T cell growth factor II; mast cell growth factor II	T cells	5q23-q31	Proliferation of B cells and T cells; enhances cytotoxic activities
IL-5	Eosinophil differentiation factor; eosinophil colony-stimulating factor	T cells	5q23.3-q32	Stimulates eosinophil formation; stimulates T cell and B cell functions
IL-6	B cell stimulatory factor II; hepatocyte stimulatory factor	Monocytes, tumor cells, B cells and T cells, fibroblasts, endothelial cells	7p	Stimulates and inhibits cell growth; promotes B cell differentiation
IL-7	Lymphopoietin 1; pre–B cell growth factor	Lymphoid tissues and cell lines	8q12-q13	Growth factor for B cells and T cells
IL-11	Plasmacytoma stimulating factor	Fibroblasts, trophoblasts, cancer cell lines	19q13.3-q13.4	Stimulates proliferation of early hematopoietic cells; induces acute-phase protein synthesis
IL-12	Natural killer cell stimulating factor	Macrophages, B cells	5q31-q33; 3p12-q13.2	Stimulates T cell expansion and interferon-gamma; synergistically promotes early hematopoietic cell proliferation
LIF	Leukemia inhibitory factor	Monocytes and lymphocytes; stomal cells	22q	Stimulates hematopoietic cell differentiation
SCF	Stem cell factor; kit ligand; steel factor	Endothelial cells; hepatocytes	4q11-q20	Stimulates proliferation of early hematopoietic cells and mast cells
FLT-3 ligand	fms-like tyrosine kinase 3; STK-1	T cells, stromal cells, and fibroblasts	19q13.3	Stimulates early hematopoietic cell differentiation; increases blood dendrite cells

G-CSF is approved for the treatment of neutropenia after cancer chemotherapy, for acceleration of neutrophil recovery after bone marrow transplantation, for mobilization of hematopoietic progenitor cells from the marrow to the blood in hematopoietic transplantation, and for the treatment of severe chronic neutropenia. The usual dosage is 5 mg/kg S.C. daily; higher doses are used to mobilize progenitor cells, and lower doses are used for long-term treatment of neutropenia. A new formulation of G-CSF, in which G-CSF is conjugated to polyethylene glycol to reduce renal clearance, was recently approved for marketing. Its principal advantage is that one injection is sufficient to stimulate marrow recovery after standard doses of cancer chemotherapy. Side effects of either form of G-CSF are principally musculoskeletal pain and headaches during the period of rapid marrow expansion soon after therapy is initiated. Other side effects are uncommon.

Granulocyte-Macrophage Colony-Stimulating Factor

GM-CSF is a glycosylated protein produced by many types of cells, including T cells.[28] GM-CSF stimulates formation of neutrophils, monocytes, and eosinophils and may also enhance the growth of early cells of other lineages. In contrast to G-CSF, GM-CSF levels generally do not increase with infections or acute inflammatory conditions,[29] and neutropenia does not result from deficiencies of GM-CSF.[30] The marrow effects of G-CSF and GM-CSF are similar, but GM-CSF is less potent in elevating the blood neutrophil count.[31] GM-CSF (the synthesized form is known as sargramostim or molgramostim) is approved

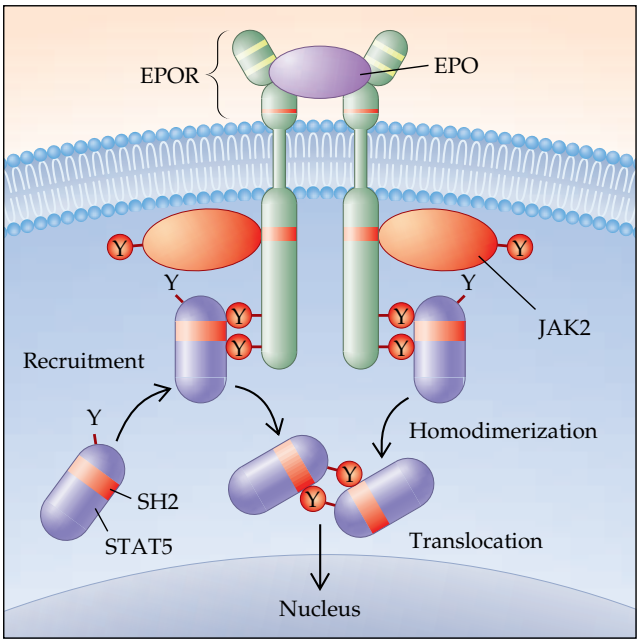


Figure 3 **A model of how hematopoietic growth factors interact with their receptors to initiate cell proliferation. (EPO—erythropoietin; JAK2—Janus kinase 2; SH2—Src homology 2; STAT5—signal transducer and activator of transcription 5)**

in the United States for acceleration of marrow recovery after bone marrow transplantation or chemotherapy and for mobilization of progenitor cells from the marrow. The usual dosage is 250 mg/m^2/day S.C. Its side effects include bone and musculoskeletal pain, myalgias, and injection-site reactions.

Interleukin-11

IL-11 (oprelvekin) is a pleiotropic cytokine that is expressed by, and active in, many tissues.[32] IL-11 acts synergistically with other growth factors, including thrombopoietin, to stimulate megakaryocyte development and platelet formation. It is approved for use in the prevention of severe thrombocytopenia and for patients who need platelet transfusions after chemotherapy. The usual dosage is 50 mg/kg/day S.C. Its side effects include edema, tachycardia, and dyspnea.

Other Growth Factors

Several other hematopoietic growth factors have potential clinical uses. IL-3 acts at an early phase in hematopoiesis to stimulate cell proliferation but has relatively little effect on peripheral counts. IL-3 has been molecularly coupled to other growth factors, including GM-CSF, G-CSF, and TPO, to produce hybrid molecules that are under investigation. Stem cell factor (SCF) and fms-like tyrosine kinase 3 (FLT-3) ligand are other early-acting factors under investigation. M-CSF is a selective factor for monocytes and macrophage formation. IL-5 is a similar selective factor for the generation of eosinophils.

It is presumed that normally, hematopoietic cell formation is governed by combinations of factors, released in a cascade, that closely coordinate the development of these cells. The details of how this process occurs, however, are not yet clear. Numerous laboratory and clinical studies have investigated combinations of factors, but the therapeutic benefit of using multiple growth factors is not yet proved.

DYNAMICS OF HEMATOPOIESIS

In the marrow, blood cells develop in two phases, the proliferative and the maturational phases. During cell proliferation, the precursors of blood cells normally undergo cell division at intervals of about 18 to 24 hours. In the maturational phase, cell division ceases, but final features are added before the cells enter the blood. During this phase, erythrocytes normally lose all their nuclear material, acquire their biconcave shape, and develop their final content of enzymes necessary for maintaining the biconcave shape and resisting destruction by oxidative stress. Normally, it takes 7 to 10 days for erythrocytes to develop from their early precursors, but this process can be accelerated by erythropoietin therapy.[33]

Neutrophils acquire most of their granules (known as the primary, secondary, and tertiary granules), which are necessary for their microbicidal activities, during the proliferative phase.[34] During maturation, their nuclear chromatin condenses, the glycogen content of the cytoplasm increases, and the surface properties governing the circulation, adherence, and migration to tissues are added.[35] Neutrophils reach a fully mature state in the marrow before they are released into the blood. These mature marrow cells are called the marrow neutrophil reserve. Quantitatively, this neutrophil pool is substantially larger—probably

Table 2 Causes of Lymphadenopathy[42]

Infections
- Bacterial: streptococci,* *Staphylococcus aureus*,* syphilis,*† cat-scratch disease,* *Mycobacterium tuberculosis* and other mycobacteria,† brucellosis,† leptospirosis,† meliodiosis,† chancroid, plague, tularemia, rat-bite fever
- Viral: adenovirus,* HIV,*† infectious mononucleosis,*† herpes simplex,* measles,† rubella,† cytomegalovirus,† hepatitis,† Kawasaki disease
- Mycotic: sporotrichosis, histoplasmosis,† coccidioidomycosis†
- Rickettsial: Rocky Mountain spotted fever,*† scrub typhus†
- Chlamydial: *Chlamydia trachomatis*, lymphogranuloma venereum
- Protozoan: toxoplasmosis,† trypanosomiasis,† kala-azar†
- Helminthic: filariasis,† onchocerciasis

Immunologic Causes
- Stings and bites*
- Drug reactions*†: phenytoin, hydralazine
- Serum sickness*†
- Collagen vascular diseases: rheumatoid arthritis,† dermatomyositis,† angioimmunoblastic lymphadenopathy†

Malignancies
- Hematologic: Hodgkin disease,* acute leukemia,† chronic lymphocytic leukemia,† chronic myelogenous leukemia,† lymphoma,† myelofibrosis†
- Other: metastatic carcinoma, sarcomas

Endocrine Diseases
- Hyperthyroidism†

Histiocytic Disorders
- Lipid storage disease,† malignant histiocytosis,† Langerhans (eosinophilic) histiocytosis

Miscellaneous
- Sarcoidosis, amyloidosis,† chronic granulomatous disease, lymphomatoid granulomatosis, necrotizing lymphadenitis

*Most common causes in general practice in the United States.

†Usually cause generalized lymphadenopathy.

five to 10 times larger—than the total circulating supply of neutrophils. Normally, it takes 10 to 14 days for blood neutrophils to develop from early precursors, but this process is accelerated in the presence of infections and by treatment with G-CSF or GM-CSF [*see 5:VII Nonmalignant Disorders of Leukocytes*].

Platelets form from the breaking apart of the cytoplasm of the fully mature megakaryocytes, which are also derived from hematopoietic stem cells.[23] Megakaryocytes undergo reduplication of their nuclear chromatin without cell division, which results in the production of extremely large cells. When marrow damage occurs from chemotherapeutic agents and after hematopoietic transplantation, the megakaryocytes are often the slowest cells to recover, and thrombocytopenia is often the last cytopenia to resolve.

There are important differences in the dynamics or kinetics of erythrocytes, platelets, and leukocytes in the blood. For instance, neutrophils have a blood half-life of only 6 to 8 hours; essentially, a new blood population of neutrophils is formed every 24 hours.[35] Erythrocytes last the longest by far: the normal life span is about 100 days.[33] These differences partially account for why neutrophils and their precursors are the predominant marrow cells, whereas in the blood, erythrocytes far outnumber neutrophils. Similarly, the short half-life and high turnover rate of neutrophils account for why neutropenia is the most frequent hematologic consequence when bone marrow is damaged by drugs or radiation. Finally, transfusion of erythrocytes and platelets is feasible because of their relatively long life span, whereas the short life span of neutrophils has greatly impeded efforts to develop neutrophil transfusion therapy.

Clinical Manifestations of Hematologic Disorders

The following signs and symptoms are frequently observed in patients with hematologic diseases.

WEAKNESS, FATIGUE, AND PALLOR

Weakness and fatigue are common complaints of patients with anemia, especially if it is of recent onset, such as anemia caused by recent blood loss or acute hemolysis.[36] Anemia that develops gradually, particularly in inactive persons, may cause only fatigue. Fatigue is a very common complaint of patients with infections, inflammatory diseases, and malignancies. Several recent studies document that raising hemoglobin levels with erythropoietin injection decreases fatigue in patients with cancer or HIV infection.[37,38] Other common causes of fatigue include chronic lung diseases, congestive heart failure, endocrine disorders, and depression.

Pallor is recognized by examining the conjunctiva, mucous membranes, nail beds, and palmar creases—tissues lacking melanin pigmentation. The World Health Organization has developed a simple clinical scale to measure pallor for diagnosing anemia when blood counts are not available. The sensitivity and specificity of this scale vary between 70% and 90%, depending on the population and severity of the anemia.[39,40] Other causes of pallor include edema (including myxedema) and vasoconstriction caused by cold temperatures, hemorrhage, hypoglycemia, or shock.

PAIN

Pain, particularly bone pain, is an important marker of hematologic disease. Pain is usually generalized in patients with acute leukemia and multiple myeloma,[41] but most frequently, it is felt in the back or pelvis. With metastatic breast, colon, or lung cancer, the pain is more often localized and asymmetrical. In sickle cell disease, severe bone pain and pain in many other tissues occur with vascular obstruction and infarction caused by obstruction of blood flow by the aggregation of abnormal cells [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*]. Bone pain mimicking these disorders occurs with marrow expansion in response to treatment with hematopoietic growth factors.

FATIGUE, PHARYNGITIS, AND FEVER

Fatigue, pharyngitis, and fever are a frequently observed sequence in patients with acutely developing neutropenia, occurring as an idiosyncratic or toxic reaction to many drugs. In cases of severe neutropenia, cough and respiratory symptoms, perianal pain and tenderness, or acute abdominal pain often occurs and necessitates immediate medical assessment [*see 5:VII Nonmalignant Disorders of Leukocytes*].

MOUTH ULCERS, GINGIVITIS, AND CERVICAL ADENOPATHY

Mouth ulcers, gingivitis, and cervical adenopathy are common problems of patients with chronic neutropenia [*see 5:VII Nonmalignant Disorders of Leukocytes*]. Gingivitis is a serious problem, often leading to periodontal disease and tooth loss.

LYMPHADENOPATHY AND SPLENOMEGALY

Lymphadenopathy is a common presentation of infectious, inflammatory, and hematologic diseases, particularly the lymphomas and leukemias [*see Table 2*]. Lymphadenopathy may occur without associated symptoms, but often, fatigue and intermittent fever (e.g., Pel-Ebstein fever) occur. In contrast to acute infectious diseases leading to tender lymphadenopathy, in most hematologic disorders the lymph nodes and spleen are nontender, with a soft to rubbery consistency. Splenomegaly is often more difficult to detect than lymphadenopathy; most of the diseases causing lymphadenopathy can also cause splenic enlargement.

BLEEDING

Bleeding occurs as a consequence of thrombocytopenia, deficiencies of coagulation factors, or both [*see 5:XIII Hemorrhagic Disorders*]. Thrombocytopenia usually presents as petechial bleeding that is first observed in the lower extremities. Coagulation factor deficiencies more often cause bleeding into the gastrointestinal tract or joints. Intracranial bleeding, however, can occur with a deficiency of platelets or coagulation factors and can be catastrophic.

THROMBOSIS

Thrombosis can be either venous or arterial [*see 5:XIV Thrombotic Disorders*]. With venous thrombosis, swelling, tenderness, and pain beyond the obstruction usually occur, and embolization to the lungs is a frequent concern. Venous thrombosis usually occurs after inactivity or obstruction of venous flow or with imbalances of coagulation factors. On the other hand, arterial thrombosis usually occurs because of abnormalities of the arterial wall from atherosclerosis or acute vascular injury, as in thrombotic thrombocytopenic purpura, or from thrombocytosis in the myeloproliferative disorders.

Laboratory Evaluation

The following basic tests are widely used to diagnose hematologic disorders.

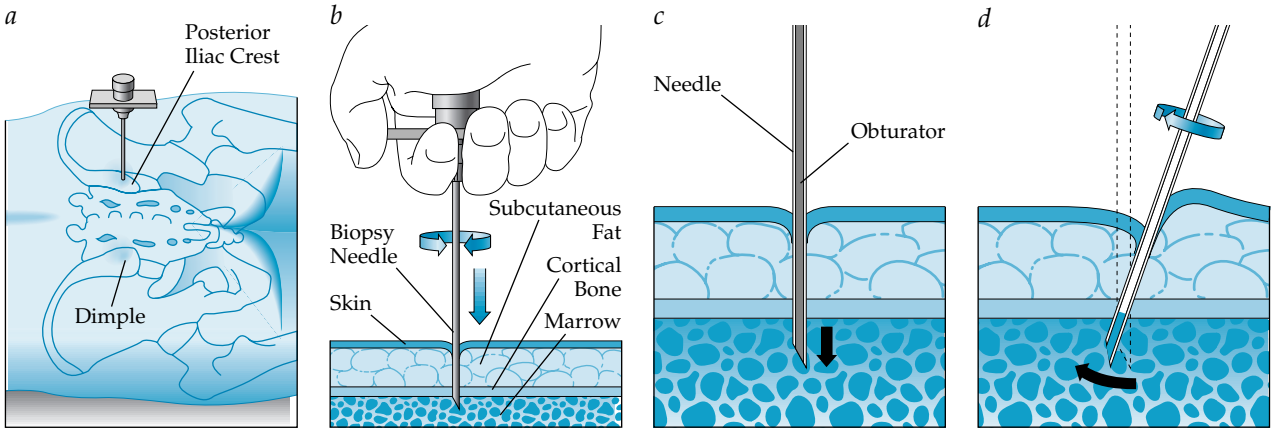


Figure 4 **Bone marrow aspirate and biopsy procedure. (*a*) The posterior iliac crest is the usual site for sampling; (*b*) the needle is placed through the skin to the marrow space; (*c*) the marrow sample is aspirated; and (*d*) the biopsy sample is carefully removed.**

COMPLETE BLOOD CELL COUNTS

CBCs are routinely performed in most laboratories through the use of an electronic particle counter, which determines the total white blood cell and platelet counts and calculates the hematocrit and hemoglobin levels from the erythrocyte count and the dimensions of the red cells. Abnormalities in the CBC are described in other Hematology subsections [*see also the* Normal Laboratory Values *section*].

PERIPHERAL BLOOD SMEARS

Peripheral blood smears usually stain with Wright stain. When examined by light microscopy, they reveal the size and shape of blood cells, which allows an estimate to be made of the amount of hemoglobin in erythrocytes. Differential leukocyte counts, enumerating the number of neutrophils, monocytes, lymphocytes, eosinophils, and basophils, are made by manually counting cells on the blood smears or by using an automated cell counter [*see the* Normal Laboratory Values *section*]. The morphology of the leukocytes often provides a clue for the diagnosis of leukemia and for recognizing some disorders of leukocytes that lead to susceptibility to infections [*see 5:VII Nonmalignant Disorders of Leukocytes*].

RETICULOCYTE COUNTS

Reticulocyte counts are useful for evaluating the marrow response to anemia [*see the* Normal Laboratory Values *section*]. Normally, during their first 24 to 36 hours in the circulation, young red cells contain residual ribosomal RNA, which precipitates with certain dyes such as methylene blue. An increase in the proportion or absolute number of reticulocytes occurs a few days after significant blood loss or in response to red blood cell destruction in hemolytic anemias. Low reticulocyte counts in chronic anemia suggest either an endogenous erythropoietin deficiency or a marrow abnormality.

BONE MARROW EXAMINATION

Hematopoietic cells of the bone marrow can be removed by aspiration or by needle biopsy. In adults, the best site is the posterior iliac crest, with the patient in a prone position [*see Figure 4*]. Under special circumstances and in children, other sites can be used, such as the anterior iliac crest, the sternum, or the long bones. With local anesthesia and sterile technique, the patient experiences only transient pain. Bleeding or infection at the injection site is quite uncommon. The aspirate yields cells for morphologic examination, and differential counts reveal the ratio of myeloid cells to erythroid cells (M:E ratio) [*see the* Normal Laboratory Values *section*]. A biopsy reveals the cellularity of the marrow at the site sampled. Biopsies are particularly useful for examination of the marrow for infiltrative cells (e.g., in lymphomas or carcinomas involving the marrow) and for diagnosing leukemia, characterized by the marrow's being so densely packed with cells that none of the bone marrow can be aspirated. Biopsies take longer for interpretation because they must be decalcified and stained before examination.

Imaging Studies

Radionuclide scanning (e.g., using technetium-99m) reveals the extent of the hematopoietic tissue in the marrow because the phagocytic cells of the marrow take up the radiolabeled particles. Marrow scanning is sometimes used to determine the extensiveness of the hematopoietic tissue; more often, it is useful in determining whether there are localized areas of increased uptake resulting from infection or a malignancy that has metastasized to the marrow. Computed tomography and ultrasonography are useful in determining the size of lymph nodes and the spleen, but they are not particularly useful for marrow examination. The marrow is seen well with magnetic resonance imaging. This technique is principally used to look for infiltrative processes in the marrow space, such as those that occur in malignancies and infections.

The author is a consultant for, receives research support from, and is a member of the speakers' bureau of Amgen, Inc.

References

1. Nishikawa SI: A complex linkage in the developmental pathway of endothelial and hematopoietic cells. Cur Opin Cell Biol 13:673, 2001

2. Choi K: The hemangioblast: a common progenitor of hematopoietic and endothelial cells. J Hematother Stem Cell Res 11:91, 2002

3. Tavassoli M: Embryonic and fetal hemopoiesis: an overview. Blood Cells 1:269, 1991

4. Verfaillie CM: Anatomy and physiology of hematopoiesis in hematology. Hematology: Basic Principles and Practice, 3rd ed. Hoffman R, Benz EJ, Shattil SJ, et al, Eds. Churchill Livingstone, New York, 2000, p 139

5. Youn BS, Mantel C, Broxmeyer HE: Chemokines, chemokine receptors and hematopoiesis. Immunol Rev 177:150, 2000

6. Gimble JM, Robinson CE, Wu X, et al: The function of adipocytes in the bone marrow

stroma: an update. Bone 19:421, 1996

7. Weissman IL: Stem cells: units of development, units of regeneration, and units in evolution. Cell 100:157, 2000

8. Williams DA: Stem cell model of hematopoiesis. Hematology: Basic Principles and Practice, 3rd ed. Hoffman R, Benz EJ Jr, Shattil SJ, et al, Eds. Churchill Livingstone, New York, 2000, p 126

9. Kim H: Whartenby KA, Georgantas RW III, et al: Human CD34+ hematopoietic stem/progenitor cells express high levels of FLIP and are resistant to Fas-mediated apoptosis. Stem Cells 20:174, 2002

10. Metcalf D: Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. Blood 92:345, 1998

11. Metcalf D: Regulatory mechanisms controlling hematopoiesis: principles and problems. Stem Cells 16(suppl 1):3, 1998

12. Drachman JG, Kaushansky K: Structure and function of the cytokine receptor superfamily. Curr Opin Hematol 2:22, 1995

13. McLemore ML, Grewal S, Liu F, et al: STAT-3 activation is required for normal G-CSF-dependent proliferation and granulocytic differentiation. Immunity 4:193, 2001

14. Haq R, Halupa A, Beattie BK, et al: Regulation of erythropoietin-induced STAT serine phosphorylation by distinct mitogen-activated protein kinases. J Biol Chem 277:17359, 2002

15. Tehranchi R, Fadeel B, Forsblom AM, et al: Granulocyte colony-stimulating factor inhibits spontaneous cytochrome c release and mitochondria-dependent apoptosis of myelodysplastic syndrome hematopoietic progenitors. Blood 101:1080, 2002

16. Hubel K, Dale DC, Liles WC: Therapeutic use of cytokines to modulate phagocyte function for the treatment of infectious diseases: current status of granulocyte colony-stimulating factor, granulocyte-macrophage factor, and interferon-gamma. J Infect Dis 185:1490, 2002

17. Spivak JL: The biology and clinical applications of recombinant erythropoietin. Semin Oncol 25(3 suppl 7):7, 1998

18. Means RT Jr: Advances in the anemia of chronic disease. Int J Hematol 70:7, 1999

19. Spivak JL: Erythropoietin use and abuse: when physiology and pharmacology collide. Adv Exp Med Biol 502:207, 2001

20. Adamson JW, Eschbach JW: Erythropoietin for end-stage renal disease. N Engl J Med 339:625, 1998

21. Goodnough LT: The role of iron in erythropoiesis in the absence and presence of erythropoietin therapy. Nephrol Dial Transplant 17(suppl 5):14, 2002

22. Rizzo JD, Lichtin AE, Woolf SH, et al: Use of epoetin in patients with cancer: evidence-based clinical practice guidelines of the American Society of Clinical Oncology and the American Society of Hematology. Blood 100:2303, 2002

23. Kaushansky K: Thrombopoietin: from theory to reality. Int J Hematol Suppl 1:343, 2002

24. Kuter DJ, Begley CG: Recombinant human thrombopoietin: basic biology and evaluation of clinical studies. Blood 100:3457, 2002

25. Welte K, Gabrilove J, Bronchud MH, et al: Filgrastim (r-metHuG-CSF): the first 10 years. Blood 88:1907, 1996

26. Molineux G: Physiologic and pathologic consequences of granulocyte colony-stimulating factor deficiency. Curr Opin Hematol 9:199, 2002

27. Price TH, Chatta GS, Dale DC: Effect of recombinant granulocyte colony-stimulating factor on neutrophil kinetics in normal and elderly humans. Blood 88:335, 1996

28. Kwon EM, Sakamoto KM: The molecular mechanism of action of granulocyte-macrophage colony-stimulating factor. J Investig Med 44:442, 1996

29. Cebon J, Layton JE, Maher D, et al: Endogenous haemopoietic growth factors in neutropenia and infection. Br J Haematol 86:265, 1994

30. Stanley E, Lieschke GJ, Grail D, et al: Granulocyte/macrophage colony-stimulating factor–deficient mice show no major perturbation of hematopoiesis but develop a characteristic pulmonary pathology. Proc Natl Acad Sci USA 91:5592, 1994

31. Dale DC, Liles WC, Llewellyn C, et al: Effects of granulocyte-macrophage colony-stimulating factor (GM-CSF) on neutrophil kinetics and function in normal human volunteers. Am J Hematol 57:7, 1998

32. Du X, Williams DA: Interleukin-11: review of molecular, cell biology, and clinical use. Blood 89:3897, 1997

33. Adamson J: Erythropoietin, iron metabolism, and red blood cell production. Semin Hematol 33:5, 1996

34. Gullberg U, Andersson E, Garwicz D, et al: Biosynthesis, processing and sorting of neutrophil proteins: insight into neutrophil granule development. Eur J Haematol 58:137, 1997

35. Dale DC, Liles WC: Neutrophils and monocytes: normal physiology and disorders of neutrophil and monocyte production. Blood: Principles and Practice of Hematology. Handin RI, Lux SE, Stossel TP, Eds. Lippincott Williams & Wilkins, Philadelphia, 2003

36. Eichner ER: Fatigue of anemia. Nutr Rev 59:S17, 2001

37. Yount S, Lai JS, Cella D: Methods and progress in assessing the quality of life effects of supportive care with erythropoietin therapy. Curr Opin Hematol 9:234, 2002

38. Moyle G: Anaemia in persons with HIV infection: prognostic marker and contributor to morbidity. AIDS Rev 4:13, 2002

39. Ingram CF, Lewis SM: Clinical use of WHO haemoglobin colour scale: validation and critique. J Clin Pathol 53:933, 2000

40. Chowdhury ME, Chongsuvivatwong V, Geater AF, et al: Taking a medical history and using a colour scale during clinical examination of pallor improves detection of anaemia. Trop Med Int Health, 7:133, 2002

41. Kyle RA: Clinical aspects of multiple myeloma and related disorders including amyloidosis. Pathol Biol (Paris) 47:148, 1999

42. Dale DC: Lymphadenopathy. Outpatient Medicine. Fihn SD, DeWitt DE, Eds. Saunders, Philadelphia, 1998

Acknowledgment

Figures 1 through 4 Seward Hung.

II RED BLOOD CELL FUNCTION AND DISORDERS OF IRON METABOLISM

Gary M. Brittenham, m.d.

Red Blood Cell Function

The red blood cell, or erythrocyte, carries oxygen from the lungs to peripheral tissues for utilization and brings carbon dioxide from tissues to the lungs for excretion.[1] The mature erythrocyte dedicates more than 95% of its intracellular protein, as hemoglobin, to these tasks. Hemoglobin, the oxygen transport molecule, binds oxygen molecules at the high oxygen tensions of the pulmonary alveoli and releases oxygen molecules at low oxygen tensions to peripheral tissues.[2,3] Hemoglobin also acts as a carrier of nitric oxide (NO), a third respiratory gas that seems to regulate oxygen delivery.[4] The exact role of NO in the cardiorespiratory cycle is debated,[5] but NO is a potent vasorelaxant that reportedly is released during arteriovenous transit, increasing blood flow and therefore oxygen transport in hypoxic tissue.[4]

The cell membrane of the erythrocyte is a flexible structure composed of a lipid bilayer with integral proteins. These proteins anchor the membrane to an underlying protein skeleton that maintains the biconcave discoid form of the cell.[6] This shape optimizes passage of the cell through the circulatory system and permits apposition of erythrocytes and parenchymal cells across the thin endothelium of capillaries, facilitating exchange of oxygen and carbon dioxide.

In the erythrocyte, a variety of metabolic pathways maintain the iron of hemoglobin in the ferrous state, protect against oxidant damage, generate 2,3-diphosphoglycerate (2,3-DPG) to help regulate oxygen affinity, and maintain osmotic stability through a series of membrane pumps.[7] Without erythrocytes, blood plasma can carry only about 5 ml O_2/L; with erythrocytes containing normal hemoglobin, whole blood can transport about 200 ml O_2/L or more.[1]

Structure of Hemoglobin

Hemoglobin is a spherical molecule composed of two pairs of dissimilar globin chains, with a heme group, ferroprotoporphyrin IX, bound covalently at a specific site in each chain.[3] The major adult hemoglobin, hemoglobin A, is formed from a pair of α chains (each containing 141 amino acids) and a pair of β chains (each containing 146 amino acids) and is written as $\alpha_2\beta_2$ [*see Figure 1*].

The configuration of hemoglobin shifts with oxygenation and deoxygenation.[2,8] The deoxy configuration of hemoglobin is stabilized through the binding of protons and 2,3-DPG, a highly charged anion. With oxygenation of one subunit, these bonds are sequentially broken, and the resulting change in tertiary structure increases oxygen affinity of the remaining unliganded subunits.[2] This phenomenon is termed cooperativity or heme-heme interaction. As oxygen is released in the tissues, a reversal of this process decreases oxygen affinity, facilitating the release of oxygen. Conformational changes also contribute to the decrease in oxygen affinity with decreasing pH.[2] This effect, called the Bohr effect, is physiologically beneficial both in the lungs, where elimination of carbon dioxide raises the pH, enhancing

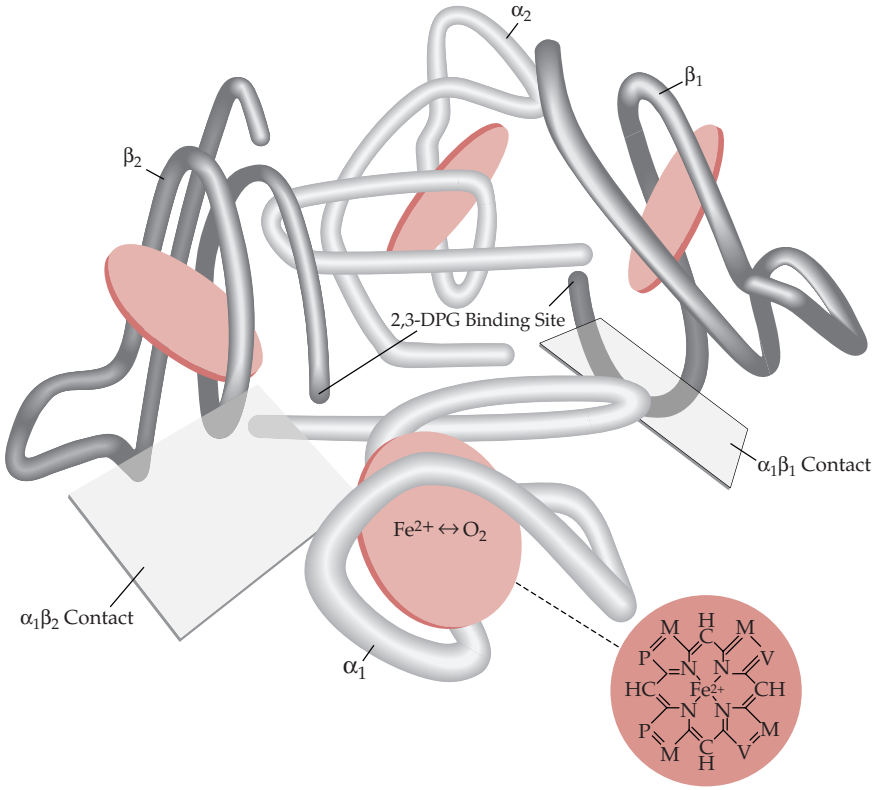


Figure 1 **A model of the hemoglobin molecule shows the relative alignment of the α chains (light gray) and β chains (dark gray). 2,3-Diphosphoglycerate (2,3-DPG), a glycolytic intermediate, binds in the central cavity of the hemoglobin and stabilizes the deoxygenated form by cross-linking the β chains, thus reducing the oxygen affinity of hemoglobin. Note that the α and β chains are in contact at two points. On oxygenation, movement of the iron atom into the plane of the heme group (colored disks) apparently triggers other structural changes in the α and β subunits as the molecule assumes the oxygenated conformation. Sliding occurs at the $\alpha_1\beta_2$ interface, and the spacing between the two β chains is reduced in oxyhemoglobin. In the detail showing the structure of heme, M is methyl, V is vinyl, and P is propionic acid.**

oxygen affinity and uptake, and in tissues, where carbon dioxide uptake decreases the pH, lowering oxygen affinity and facilitating oxygen release.

FACTORS AFFECTING THE OXYGEN-CARRYING CAPACITY OF HEMOGLOBIN

Oxygen-Hemoglobin Dissociation Curve

The oxygen-hemoglobin dissociation curve [*see Figure 2*] is a plot of the equilibrium between oxygen and hemoglobin at various oxygen tensions (P_{O_2}).[3] At sea level and at a partial pressure of oxygen of about 90 mm Hg, hemoglobin is 97% saturated in the lungs. After unloading oxygen to tissues, at a P_{O_2} of about 40 mm Hg in mixed venous blood, the hemoglobin saturation is about 75%. The P_{50} (i.e., the partial pressure of oxygen at which hemoglobin is 50% saturated) is a useful measure of the oxygen affinity of hemoglobin: the higher the affinity, the lower the P_{50}. Under normal physiologic conditions (i.e., a temperature of 37° C [98.6° F]; pH of 7.4; 2,3-DPG level of 5 mmol/L; and carbon dioxide tension [P_{CO_2}] of 40 mm Hg), the P_{50} of normal adult blood is 26 ± 1 mm Hg. The P_{50} is decreased (shifted to the left on the oxygen-hemoglobin dissociation curve) by increasing pH, decreasing 2,3-DPG, or decreasing temperature.

Effects of 2,3-Diphosphoglycerate

The glycolytic intermediate 2,3-DPG, which is present in mature erythrocytes at approximately the same intracellular concentration as hemoglobin, is the most important allosteric regulator of oxygen affinity.[3] With acute hypoxia, 2,3-DPG concentrations increase within hours, which shifts the oxygen-hemoglobin dissociation curve to the right. The increase in the 2,3-DPG concentration promotes delivery of oxygen to tissues but also impedes the acquisition of oxygen in the lungs. Short-term adaptation to hypoxic stress may be helped if the supply of oxygen is plentiful and the cardiopulmonary reserve is robust. At high altitude, with the cardiovascular system unable to effectively meet increased circulatory demands, or in other pathologic circumstances, increased amounts of 2,3-DPG may be counterproductive.[3]

OXYGEN TRANSPORT

Several other physiologic factors function in an integrated manner to provide an adequate supply of oxygen, including blood volume, blood viscosity, pulmonary and cardiac function, and regional blood flow.[9] The concentration of circulating red blood cells depends on the production of erythropoietin by the kidney and the erythropoietic response of the erythroid marrow.[10] Hemoglobin transport of NO, which binds to both heme iron and globin, may help match regional blood flow and oxygen requirements. In peripheral tissues, the erythrocyte may release NO, which would relax the microvasculature, improve blood flow, and enhance oxygen delivery.[4,5] An intriguing development has been the identification of a new class of so-called hexacoordinate hemoglobins (expressed in nerve cells[11] or a wide array of tissues[12]), whose functions are not yet established but which may facilitate oxygen transport, help protect against hypoxia, or scavenge reactive oxygen species.

CARBON DIOXIDE TRANSPORT

After delivering oxygen, hemoglobin binds with carbon dioxide.[13] Deoxyhemoglobin has a higher binding affinity for carbon dioxide than does oxyhemoglobin, facilitating unloading of carbon dioxide from tissues and pulmonary excretion.[13] Most of the carbon dioxide from tissue capillaries is transported to the lungs as bicarbonate, with about 10% carried as a carbamino complex reversibly bound to the globin chains.

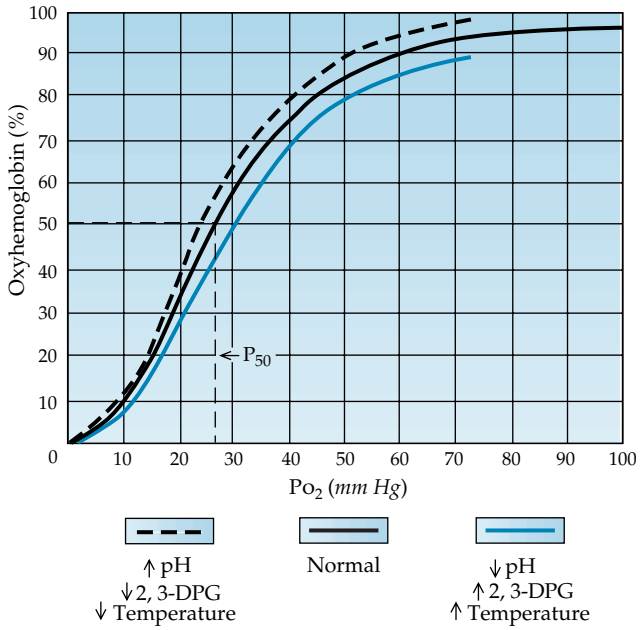


Figure 2 **The normal oxygen-hemoglobin dissociation curve (solid black line) is shifted by changes in temperature, pH, and the intracellular concentration of 2,3-DPG. P_{50} stands for 50% oxygen saturation; PO_2 stands for oxygen tension.**

Iron Metabolism

Remarkable progress continues to be made in understanding disorders of iron metabolism and in improving the diagnosis and management of both iron deficiency and iron overload.[14,15] In the body, iron transports and stores oxygen, carries electrons, catalyzes reactions in oxidative metabolism, and sustains cellular growth and proliferation. With iron deficiency, the body is unable to produce sufficient amounts of heme, other iron-porphyrin complexes, metalloenzymes, and other iron-containing compounds to sustain normal functions. With iron overload, excess iron can catalyze free radical reactions that can damage cellular membranes, proteins, and nucleic acids, resulting in progressive cellular and organ damage and eventual death.

PATTERNS OF IRON BALANCE AND METABOLISM

The concentration of iron in the human body is carefully regulated and is normally maintained at about 40 mg Fe/kg in women and about 50 mg Fe/kg in men. Iron balance is the result of the difference between the amount of iron taken up by the body and the amount lost [*see Figure 3*]. Because humans are unable to excrete excess iron, iron balance is physiologically regulated by the control of iron absorption. The two major factors that influence iron absorption are the level of body iron stores and the extent of erythropoiesis.[16] If iron stores increase, absorption decreases; if stores decrease, absorption increases. Absorption also increases with increased erythropoietic activity, especially with ineffective erythropoiesis. Most of the iron in the body is located in the erythron, which consists of the totality of circulating erythrocytes and their precursors in the bone mar-

row. The predominant pathway of internal iron flux is a one-way flow from the plasma iron transport protein, transferrin, to the erythron, and then through the monocyte-macrophage system back to transferrin [*see Figure 3*]. The erythron uses about 80% of the iron passing through the transferrin compartment each day. Normally, the majority of this iron is used for hemoglobin synthesis and returned to the circulation within red blood cells. Small quantities of iron are stored in ferritin, enter the iron-containing enzymes of immature erythroid cells, or are lost in the products of ineffective erythropoiesis. At the end of their life span, senescent red cells are phagocytized by specialized macrophages in the spleen, bone marrow, and liver, which then return most of the iron to the transferrin compartment, where the cycle begins again. The phagocytosis of flawed and aged erythrocytes accounts for almost all of the storage iron normally found in the macrophages of the liver, bone marrow, and spleen. By contrast, the parenchymal cells of the liver may either take iron from, or give iron to, plasma transferrin. Under normal physiologic conditions, iron recycling is very efficient; less than 0.05% of the total body iron is acquired or lost each day.

MOLECULAR BASIS OF IRON METABOLISM

A number of proteins are now known to be involved in the absorption, transport, utilization, and storage of iron. Some of these proteins have more than one function.

Systemic Iron Transport

Transferrin is the physiologic carrier of iron through the plasma and extracellular fluid.[15] Apotransferrin, transferrin without attached iron, is a single-chain glycoprotein with two structurally similar lobes. Binding of a ferric ion to one of these lobes yields monoferric transferrin; binding of ions to both yields diferric transferrin. The transferrin saturation is the proportion of the available iron-binding sites on transferrin that are occupied by iron atoms, expressed as a percentage. In humans, almost all the circulating plasma apotransferrin is synthesized by the hepatocyte.[17] After delivering iron to cells, apotransferrin is promptly returned to the plasma to again function as an iron transporter, completing 100 to 200 cycles of iron delivery during its lifetime in the circulation.[15]

Cellular Iron Uptake

Specific receptors for transferrin, which are found on the surface membrane of all nucleated cells, provide the route of entry for transferrin-bound iron. The affinity of transferrin receptors is greatest for diferric transferrin, intermediate for monoferric transferrin, and almost negligible for apotransferrin. These differences in affinity contribute to the efficiency of iron delivery. Two forms of the transferrin receptor have now been identified and are designated as transferrin receptor 1 and transferrin receptor 2.[15] Both consist of paired subunits, each of which can

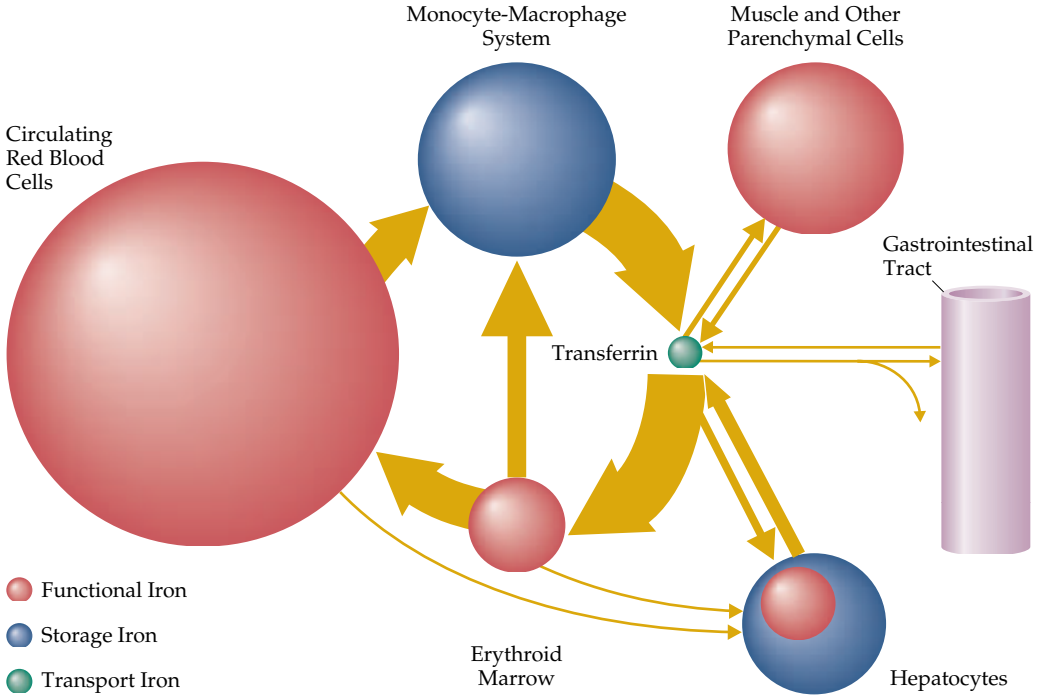


Figure 3 **Body iron supply and storage. The figure shows a schematic representation of the routes of iron movement in the adult.[95] The area of each circle is proportional to the amount of iron contained in the compartment, and the width of each arrow is proportional to the daily flow of iron from one compartment to another. The major portion of iron is found in the erythron as hemoglobin iron (28 mg/kg in women; 32 mg/kg in men) dedicated to oxygen transport and delivery. Small amounts of erythron iron (< 1 mg/kg) are also present in heme and nonheme enzymes in developing red cells. The remainder of functional iron is found as myoglobin iron (4 mg/kg in women; 5 mg/kg in men) in muscle and as iron-containing and iron-dependent enzymes (1 to 2 mg/kg) throughout the cells of the body. Small amounts of iron are deposited within ferritin in erythroid cells, but most storage iron (5 to 6 mg/kg in women; 10 to 12 mg/kg in men) is held in reserve by hepatocytes and macrophages in the liver, bone marrow, spleen, and muscle. The small fraction of transport iron (about 0.2 mg/kg) in the plasma and extracellular fluid is bound to the protein transferrin, which carries iron to meet tissue needs throughout the body.**

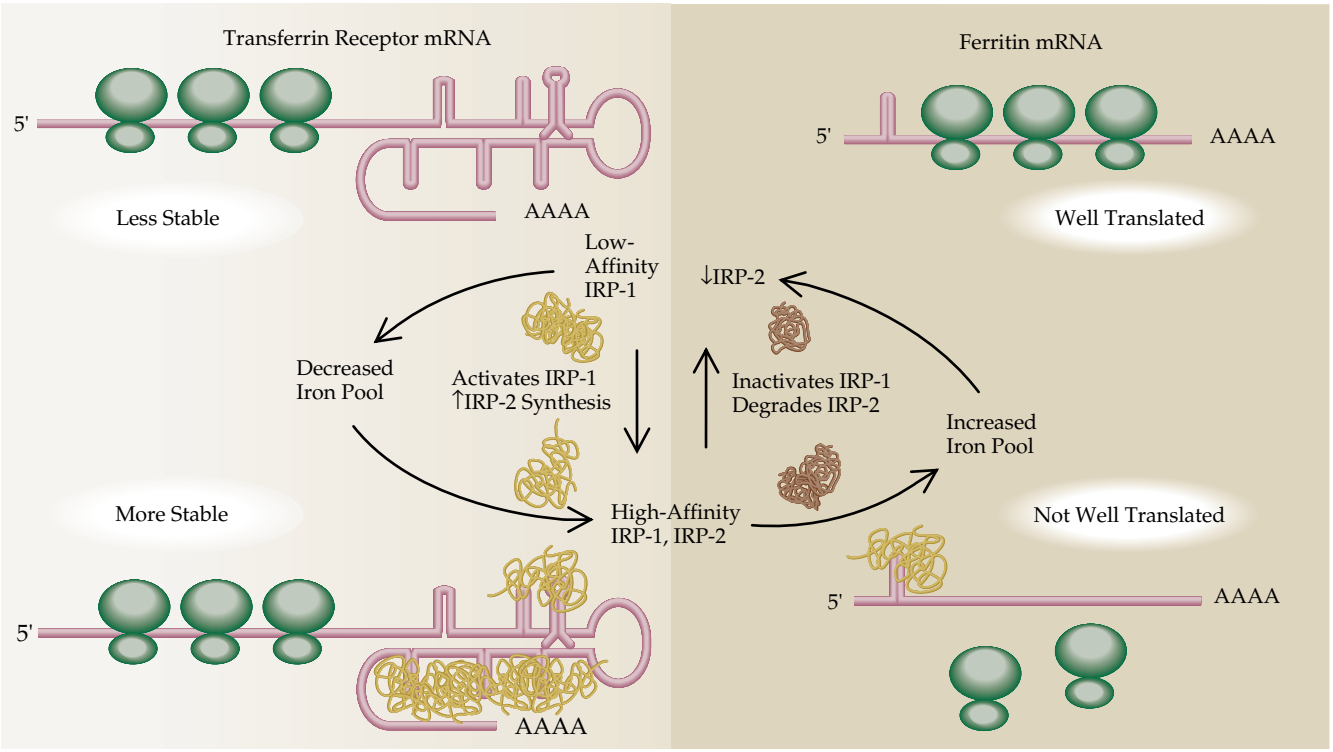


Figure 4 **Regulation of transferrin receptor and ferritin expression by the iron regulatory proteins IRP-1 and IRP-2.**

bind a molecule of transferrin. Although their extracellular structures are quite similar,[18] the two forms of the receptor have important functional differences. The binding affinity of transferrin receptor 1 for diferric transferrin is about 25-fold to 30-fold greater than that of transferrin receptor 2. Expression of transferrin receptor 1 is regulated by intracellular iron levels (see below), but that of transferrin receptor 2 is not. Transferrin receptor 1 is expressed by all iron-requiring cells, whereas transferrin receptor 2 is most highly expressed on hepatocytes and developing erythroid cells. Transferrin receptor 1 is required for life in mammals[19]; transferrin receptor 2 cannot compensate for the absence of transferrin receptor 1. Transferrin receptor 2 seems to have other functions in maintaining iron homeostasis (see below).

The delivery of transferrin-bound iron begins with the binding of two molecules of monoferric or diferric transferrin to a transferrin receptor on the cell surface.[15] The iron-transferrin–transferrin receptor complex then moves to the interior of the cell within an endosome, where it releases its iron. The dissociated iron is then taken up by the divalent metal transporter–1 (DMT1) and transported across the endosomal membrane for utilization or for storage in the cell.[19] Freed of its iron, the transferrin—now apotransferrin—binds avidly to the transferrin receptor and is carried back to the cell membrane and released into the plasma. [15]

Cellular Iron Storage

Cellular iron storage utilizes ferritin, a protein found in the cytoplasm of virtually all cells. Ferritin is a spherical shell that can store as many as 4,500 atoms of iron in its interior. Ferritin functions both as a safe storage site for iron and as a readily accessible reserve for iron that has been acquired by the cell in excess of its immediate needs.[20] Accordingly, the greatest amounts of ferritin are found in cells dedicated to iron storage (e.g., macrophages and hepatocytes) and in cells with the highest iron requirements for the synthesis of iron-containing compounds (e.g., developing erythroid cells). Apoferritin, or ferritin without attached iron, is composed of 24 oblong subunits that are designated as H (heavy) and L (light). Ferritin molecules with a greater proportion of H subunits seem to be more active in iron metabolism; ferritin molecules with a greater abundance of L subunits apparently are used for the longer-term storage of iron. A ferritin H–like protein assembled into ferritin shells has been identified within iron-loaded mitochondria of patients with impairment of heme synthesis.[21] In patients with sideroblastic anemia, most of the iron in ringed sideroblasts is sequestered in mitochondrial ferritin.[22]

Regulation of Cellular Iron Uptake and Storage

Cellular iron uptake and storage are regulated through the synthesis of transferrin receptors and ferritin. This synthesis is coordinated by the iron-regulatory proteins IRP-1 and IRP-2 [*see Figure 4*]. When intracellular iron levels are low, IRP-1 and IRP-2 bind to messenger RNA (mRNA) stem-loop elements known as iron-responsive elements (IREs) in transcripts.[23] Transferrin receptor synthesis is regulated by controlling the stability of cytoplasmic transferrin receptor mRNA, whereas ferritin synthesis is regulated by controlling translation of ferritin mRNA without changing the amount of ferritin mRNA in the cytoplasm. As a result, changes in the amounts of the IRPs have opposite effects on the production of transferrin receptor and ferritin, allowing iron to self-regulate its intracellular availability. IRPs also regulate mRNAs for other proteins involved in iron uptake,[24] availability, release,[25] and utilization.[26]

Macrophage Hemoglobin Catabolism and Iron Release

Specialized macrophages in the bone marrow, liver, and spleen selectively recognize and phagocytize erythrocytes that are senescent or damaged.[27] On average, each of these macrophages can

phagocytize one erythrocyte a day.[28] After ingesting the erythrocyte, the macrophage lyses the erythrocyte membrane. The hemoglobin within then undergoes oxidative precipitation and rapid catabolism into heme.

Any hemoglobin released into plasma by intravascular hemolysis is rapidly bound by haptoglobin. Macrophages remove the haptoglobin-hemoglobin complex from plasma by binding[29] and endocytosis; the complex is then digested in lysosomes, liberating heme.

The heme from both sources is degraded by the microsomal enzyme heme oxygenase, yielding biliverdin IXa, carbon monoxide, and iron[30,31]; the iron is either stored in ferritin or returned to plasma, apparently via ferroportin.[27,32] The outpouring of iron from macrophages in the bone marrow, liver, and spleen to plasma apotransferrin normally constitutes the largest single flux of iron from cells in the body.[30] Unsaturated transferrin is not required for the release of iron from the macrophages; apotransferrin does not enter the macrophage and accepts iron only after the release of iron from the cell. Rather, a major determinant of the rate of iron exit from the macrophage is ceruloplasmin, which establishes a rate of oxidation of ferrous iron.[33] After oxidation, the ferric iron can be bound and transported by transferrin back to the erythron and other iron-requiring tissues.

Other Pathways of Iron Exchange

Heme that is released into the plasma as a result of intramedullary or intravascular hemolysis is bound by hemopexin. The heme-hemopexin complex is delivered to the hepatocyte via specific receptors and internalized, and the heme is catabolized to liberate the iron.[34] The hepatocyte may either donate iron to plasma transferrin or receive iron from it.[35] At high transferrin saturations, iron moves from the plasma to the liver; at low saturations, iron is mobilized from hepatocyte stores and supplied to plasma transferrin. Normally, the overall magnitude of iron exchange by hepatocytes is only about one fifth that by macrophages. Other pathways of iron movement involve approximately equal exchanges: about 1 mg Fe/day is absorbed and lost, about 3 mg Fe/day is transferred between the plasma and extravascular transferrin compartments, and about 2 mg Fe/day moves between extravascular transferrin and parenchymal tissues.

Intestinal Iron Absorption

The amount of iron in the body is regulated by control of iron absorption in the proximal small intestine. Both heme iron and nonheme iron enter through the brush border of intestinal enterocytes.[36] The exact means by which heme iron is absorbed are still uncertain, but nonheme iron seems to be taken up via the apical iron transporter DMT1 (the same iron transporter that provides an exit for iron from the endosome; see above) and perhaps via other routes.[37] DMT1 is a proton-coupled symporter with a broad substrate range that includes other metallic cations, but its physiologic function appears to be the uptake of Fe^{2+}.[19] DMT1 is a ferrous iron transporter, but most dietary iron is in the ferric form. Duodenal cytochrome b, a heme protein highly expressed in duodenal brush border membrane, reduces luminal iron to the ferrous state for transport into enterocytes via DMT1.[38] Once within the enterocyte, the absorbed iron may be transported across the basolateral membrane into the plasma or stored within ferritin and then lost when the enterocyte is exfoliated. The details of the handling of iron in the enterocyte remain obscure, but as in the macrophage, ferroportin seems to serve as the transmembrane channel for the transfer of ferrous iron into the plasma,[32] and hephaestin is a ceruloplasmin homologue that is required for the efficient exit of iron from the basolateral membrane of the enterocyte into the systemic circulation.[39]

Regulation of Body Iron Absorption and Storage

The molecular mechanisms responsible for the regulation of body iron stores through control of iron absorption are still not understood, but some of the proteins involved have been identified. The gene *HFE*, which is mutated in most patients with hereditary hemochromatosis (see below), has been identified,[40] but the means by which the normal gene product, HFE, regulates iron balance remains uncertain. HFE is found in crypt cells in the duodenum, in tissue macrophages, and in Kupffer cells in the liver. Two mutually exclusive functions have been suggested for HFE: decreasing iron uptake by binding to transferrin receptor (thereby competing with transferrin) and inhibiting iron release from macrophages.[41,42] The relative effects of these offsetting activities may be determined by the balance between serum transferrin saturation and serum transferrin–receptor concentrations.[43] Other models give the liver a central role in the regulation of iron homeostasis.[44] The observations that tissue iron overload can result from mutations in two other proteins, transferrin receptor 2[45] (see above) and hepcidin,[46] indicate that these proteins are also involved in the pathway for regulation of iron balance. Hepcidin, an antimicrobial peptide produced by hepatocytes, seems to act as an iron-regulatory hormone whose expression is inversely related to both iron absorption and macrophage iron release. In HFE-associated hereditary hemochromatosis, expression of hepcidin is not increased despite hepatic iron overload, suggesting that HFE may be involved in the regulation of hepcidin expression in response to changes in body iron stores.[47] Despite these intriguing observations, elucidation of the molecular mechanisms underlying the regulation of body iron absorption and storage will likely await the identification of other key proteins involved.

Iron Deficiency

DEFINITIONS

Iron deficiency designates conditions in which the body's iron requirements exceed iron supply. Iron is needed to restore physiologic loss, which is just under 1 mg/day in men and is about 1.5 mg/day in menstruating women[35]; iron is also needed for growth and pregnancy and to replace pathologic losses. Sequential stages of decreases in body iron can be identified [*see Table 1*]. A decrease in iron stores without a change in the amounts of functional iron compounds is designated as reduced iron stores. When iron stores are exhausted, patients may be described as having iron depletion. Further decrements in the level of body iron result in limited production of hemoglobin and other iron-containing functional compounds; this stage is termed iron-deficient erythropoiesis. Still further decreases in body iron produce iron deficiency anemia.

EPIDEMIOLOGY

Iron deficiency is the most common nutritional deficiency worldwide. Its prevalences are highest in developing countries, where 30% to 70% of the population may be affected[48]; in comparison, the overall prevalence of iron deficiency is less than 20% in the industrialized countries of Europe and North America.[48] In

Table 1 Changes in Iron Stores and Distribution with Increased or Decreased Body Iron Content

Condition	*Marrow Iron (0–6+)*	*Liver Iron (μmol/g, dry weight)*	*Plasma Ferritin (μg/L)*	*Plasma Transferrin Receptor (mg/L)*	*Plasma Iron (μg/dl)*	*Transferrin Saturation (%)*	*Protoporphyrin (μg/dl Red Blood Cells)*	*Red Blood Cells*
Iron deficiency anemia	0	< 3.0	< 12	10	< 40	< 10	150	Microcytic, hypochromic
Iron stores depletion	0 to trace amounts	3.0	< 20	5.5	< 115	< 30	30	Normal
Reduced iron stores	1+	< 10.0	< 25	5.5	< 115	30	30	Normal
Normal	2–3+	15.0 ± 5.0	100 ± 60	5.5 ± 1.5	115 ± 50	35 ± 15	30	Normal
Increased iron stores, hereditary hemochromatosis	2–3+	100	1,000	5.5	> 150	> 60	30	Normal
Increased iron stores; transfusional iron overload	4+	200	1,000	5.5	> 150	> 50	30	Normal
Massive iron overload, hereditary hemochromatosis	3–4+	400	4,000	5.5	200	> 60	30	Normal
Massive iron overload; transfusional iron overload	6+	800	4,000	5.5	200	> 90	30	Normal

the United States, the Centers for Disease Control and Prevention estimate that the prevalence of iron deficiency is greatest in toddlers 1 to 2 years of age (7%) and in adolescent girls and adult women 12 to 49 years of age (9% to 16%).[49]

ETIOLOGY

Iron deficiency can result from increased iron requirements, inadequate iron supply, or both [*see Table 2*]. Blood loss is the most common cause of increased iron requirements that lead to iron deficiency. In men and postmenopausal women, iron deficiency is almost always the result of gastrointestinal blood loss.[50] In menstruating women, genitourinary blood loss often accounts for increased iron requirements. Oral contraceptives tend to decrease menstrual blood loss, whereas intrauterine devices tend to increase menstrual bleeding. Other causes of genitourinary bleeding and respiratory tract bleeding can also increase iron requirements [*see Table 2*]. For blood donors, each donation results in the loss of 200 to 250 mg of iron. During periods of growth in infancy, childhood, and adolescence, iron requirements may outstrip the supply of iron available from diet and stores.[51] Iron loss from tissue growth during pregnancy and from bleeding during delivery and post partum averages 740 mg.[35,51] Breast-feeding increases iron requirements by about 0.5 to 1 mg /day.

An insufficient supply of iron may contribute to the development of iron deficiency. In infants and in women with high iron requirements, diets containing inadequate amounts of bioavailable iron increase the risk of iron deficiency.[51] In older children, men, and postmenopausal women, a poor supply of dietary iron is almost never the only factor responsible for iron deficiency; therefore, other etiologic factors must be sought, especially blood loss.[50,52,53] Impaired absorption of iron is an uncommon cause of iron deficiency. In some patients, intestinal malabsorption of iron is only one aspect of more generalized malabsorption[54] [*see Table 2*]. Gastric surgery, especially partial or total gastric resection or gastroenterostomy for bypass of the duodenum, may result in iron deficiency. Although absorption of dietary iron may be poor in such patients, therapeutic iron salts are usually well absorbed, and the iron deficiency can be readily corrected.

The risk of iron deficiency is especially high when iron requirements are increased and the supply of iron is inadequate. For example, infants who are fed cow's milk often become iron deficient because of the combination of increased iron losses from cow's milk–induced gastrointestinal bleeding and the small amounts of bioavailable iron in cow's milk.[51,55] Women with high iron requirements because of menstruation often have diets that contain little bioavailable iron and contain inhibitors of iron absorption, such as calcium. A common mutation of the transferrin gene (designated as G277S) has been associated with a reduction in the circulating transferrin concentration and may predispose menstruating women to iron deficiency.[56] The mechanism underlying this effect is unknown.

DIAGNOSIS

Clinical Manifestations

Patients with iron deficiency may be asymptomatic and their disorder recognized only because of abnormal results of laboratory tests.[51] Other patients may come to medical attention because of the manifestations of the underlying disorder that produced iron deficiency, but they may have no findings resulting from the iron deficiency. Still other patients may present with the signs and symptoms common to all anemias, such as weakness, dizziness, easy fatigability, pallor, irritability, and other indefinite and nonspecific complaints. Iron deficiency may also be associated with signs and symptoms that are unrelated to anemia, such as angular stomatitis, glossitis, postcricoid esophageal stricture or web, and gastric atrophy. In addition, a high prevalence of iron deficiency has been found in patients with the restless legs syndrome, a neurologic disorder characterized by a distressing, irresistible urge to move the legs (akathisia).[57] Finally, some patients present with one or more of the limited number of signs and symptoms thought to be highly specific for iron deficiency, which include blue sclerae[58] and koilonychia. Pagophagia, or pica with ice, is thought to be another highly specific symptom of iron deficiency and disappears shortly after iron therapy is begun.[59] Other types of pica may accompany iron deficiency, but

Table 2 Causes of Iron Deficiency

INCREASED IRON REQUIREMENTS
- Blood loss
  - Gastrointestinal tract
    - Hemorrhagic lesions (e.g., hiatal hernia, esophageal varices, gastritis, duodenitis, peptic ulcer, cholelithiasis, intrahepatic bleeding, inflammatory bowel disease, diverticulosis, hemorrhoids, or adenomatous polyp)
    - Occult gastrointestinal malignancy
    - Chronic ingestion of drugs (e.g., alcohol, salicylates, steroids, and nonsteroidal anti-inflammatory drugs)
    - Helminthic infections (e.g., hookworm, *Schistosoma mansoni, S. japonicum*, or severe *Trichuris trichiura*)
    - Other (e.g., vascular purpura with scurvy, aberrant pancreas, Meckel diverticulum, hereditary hemorrhagic telangiectasia, other vascular ectasia of the bowel, or colonic polyposis)
  - Genitourinary tract
    - Menstrual blood loss
    - Other (e.g., uterine malignancies or fibroids, stones, infarction, infection with *S. haematobium*, inflammatory disease, malignancy of the urinary tract, or chronic hemoglobinuria or hemosiderinuria resulting from paroxysmal nocturnal hemoglobinuria or chronic intravascular hemolysis)
  - Respiratory tract
    - Chronic recurrent hemoptysis
    - Idiopathic pulmonary siderosis
    - Goodpasture syndrome
  - Blood donation
- Growth
  - Infants, premature infants
  - Children
  - Adolescents
- Pregnancy and lactation

INADEQUATE IRON SUPPLY
- Diets with insufficient amounts of bioavailable iron
- Impaired absorption of iron
  - Intestinal malabsorption (e.g., steatorrhea, sprue, celiac disease, diffuse enteritis, atrophic gastritis with achlorhydria, or pica)
- Gastric surgery

none is as specific a symptom as pagophagia. The nonhematologic consequences of iron deficiency are diminished exercise tolerance and work performance and impaired immunity and resistance to infection.[60,61] In children, iron deficiency seems to adversely affect growth, motor development, behavior, and cognitive function[51,62]; these abnormalities may not be reversible with later treatment.

Laboratory Tests

Iron deficiency anemia is the only microcytic hypochromic anemia associated with lack of iron stores. In other microcytic hypochromic disorders, marrow iron stores are normal or increased. Indirect measures of body iron can be used to identify a characteristic sequence of changes that occur as body iron decreases from the iron-replete normal levels to levels found in iron deficiency anemia [*see Table 2*]. Measurement of the serum ferritin concentration is the most useful test for the detection of iron deficiency, because serum ferritin concentrations decrease as body iron stores decline.[50] A serum ferritin concentration below 12 µg/L is virtually diagnostic of absent iron stores. In contrast, a normal serum ferritin concentration does not confirm the presence of storage iron, because serum ferritin concentration may be increased independently of body iron by infection, inflammation, liver disease, malignancy, and other conditions.[63] Because the serum transferrin receptor concentration seems to be unaffected by these conditions, determination of this value (or the ratio of the serum transferrin receptor concentration to the serum ferritin concentration) provides a means of distinguishing between the anemia of iron deficiency and the anemia associated with chronic inflammatory disorders.[64-66] The concentration ratio of the serum transferrin receptor to serum ferritin can also provide a quantitative estimate of body iron stores that may be useful in monitoring iron status in patients who are highly susceptible to iron deficiency, such as infants, preschool children, and pregnant women.[67] The serum iron level and serum transferrin saturation are decreased in both iron deficiency and infectious or inflammatory states and therefore are of little practical assistance in distinguishing between these conditions. An alternative approach to the detection of an impaired iron supply for erythropoiesis is the use of hematologic indices derived from automated analyzers, such as the proportion of hypochromic cells and reticulocyte cellular indices.[68,69] These measurements of erythrocyte and reticulocyte indices may be particularly useful in the evaluation of iron-restricted erythropoiesis in patients with chronic renal failure or chronic disease who are treated with erythropoietin.[70] An empirical trial of iron therapy may also be an effective means of establishing the diagnosis of iron deficiency.

Although bone marrow examination is now seldom performed solely for the assessment of iron status, the diagnosis of iron deficiency can almost always be verified by direct assessment of marrow iron stores. If no iron stores are present, the diagnosis of iron deficiency is established; if hemosiderin (an intracellular granule that stores iron-containing molecules) is found, iron deficiency is excluded. In addition, with iron deficiency, marrow sideroblasts will be absent or present in low numbers (less than 10% of the number of normoblasts).

TREATMENT

Therapy for iron deficiency anemia should both correct the hemoglobin deficit and replace storage iron [*see Sidebar*, Iron Replacement Therapy]. Oral and parenteral replacement therapies yield similar results,[50,51] but for almost all patients, oral iron is the treatment of choice.[50] Oral iron therapy is effective, safe, and inexpensive.[71] Because of the risk of local and systemic adverse reactions, parenteral iron should be used only in the small number of patients who cannot absorb or tolerate oral iron or whose iron requirements cannot be met by oral therapy because of chronic, uncontrollable bleeding or other blood loss. In severe iron deficiency anemia, red cell transfusions are needed in rare instances to prevent cardiac or cerebral ischemia. Red cell transfusions may also sometimes be necessary for patients whose chronic rate of iron loss exceeds the rate of replacement possible with parenteral therapy. Although the majority of patients take oral iron without difficulty, 10% to 20% experience side effects related to iron—most commonly, gastrointestinal complaints. Despite manufacturers' claims, there are no clinically significant differences between different iron salts.

Iron deficiency can almost always be treated effectively. Alleviation of symptoms often occurs within the first few days of treatment. With uncomplicated iron deficiency, the initial hematologic response—a mild reticulocytosis—usually begins within 3 to 5 days after the start of therapy, reaches a maxi-

Iron Replacement Therapy

ORAL IRON THERAPY

- Indication
  - Treatment of choice for iron deficiency anemia
- Initial therapy to correct iron deficiency anemia
  - Ferrous iron salt (e.g., ferrous sulfate) given separately from meals in two or three divided doses; for example, ferrous sulfate tablets, 325 mg three times a day, or ferrous gluconate tablets, 300 mg two or three times a day
- Continued therapy to replace iron stores
  - Ferrous iron salt given as a single daily dose of approximately 60 mg of elemental iron until the plasma ferritin concentration is > 50 µg/L (often requires 6 mo or more of treatment)
- Management of side effects
  - Gastrointestinal side effects are the most common (10%–20% of patients) and usually can be managed symptomatically by (1) giving iron with or immediately after meals, (2) reducing the amount of iron in each dose, or (3) reducing the dose frequency to once daily

PARENTERAL IRON THERAPY

- Indications
  - Chronic, uncontrollable blood loss producing iron needs that cannot be met by oral iron therapy
  - Malabsorption of iron
  - Intolerance of oral iron despite repeated modifications in dosage regimen
- Risks
  - Immediate, life-threatening anaphylactic reactions
  - Delayed but severe serum sickness–like reactions with fever, urticaria, adenopathy, myalgias, and arthralgias
  - Exacerbation of rheumatoid arthritis and related conditions
  - Local reactions with intramuscular iron (skin staining, muscle necrosis, phlebitis, and persistent pain at injection site)
- Precautions
  - Iron dextran is the only currently available parenteral preparation; a 0.5 ml test dose is to be given at least 1 hr before every intramuscular or intravenous injection of iron dextran, but the value of this precaution is limited because anaphylaxis is not dose dependent and can occur with the test dose
- Administration and dosage
  - Parenteral iron may be administered either intramuscularly (limited to 2 ml or 100 mg of iron per injection) or intravenously (as an undiluted injection, as a total-dose infusion, or as an additive to total parenteral nutrition); because of the risks of therapy, recommendations of the manufacturers and recent-study recommendations for treatment should be reviewed carefully before parenteral iron is given

mum within 8 to 10 days, and declines thereafter. After the first week, the hemoglobin concentration begins to increase and is usually normal within 6 weeks. Microcytosis may not resolve completely for as long as 4 months. If the iron deficiency is treated with oral iron at a dosage of 200 mg/day or less, the serum ferritin concentration remains below 12 µg/L until the anemia is corrected and then gradually rises as storage iron is replenished. If the response to iron therapy is not complete and characteristic, review and reevaluation of the patient is mandatory.[50] One of the most common problems is mistaking the anemia of chronic disease for the anemia of iron deficiency. Recovery may be retarded by coexistent disorders, including other nutritional deficiencies; liver or kidney disease; infectious, inflammatory, or malignant disorders; or continued occult blood loss. In the event of incomplete recovery in a patient who is being treated with oral iron, the form and dosage of iron used should be reviewed, compliance evaluated, and the possibility of malabsorption considered.

Iron Overload

DEFINITIONS

Iron overload arises from a sustained excess of iron supply over iron requirements and causes characteristic patterns of changes in functional, transport, and storage iron. The amount of body iron is normally controlled by regulation of dietary iron absorption. Iron overload develops with conditions that modify or circumvent the regulation of intestinal iron absorption. Because humans have no physiologic means of eliminating excess iron, any persistent increase in intake may eventually result in iron overload. When the extent of iron accumulation exceeds the body's ability to safely sequester the surplus iron, characteristic patterns of tissue damage develop. The precise manifestations of iron overload depend on the underlying abnormality responsible but generally are governed by the magnitude of the body iron burden; the rate at which the increase in body iron has occurred; the distribution of the excess iron between storage sites in macrophages and potentially more harmful deposits in parenchymal cells; and the coexistence of conditions that may ameliorate (e.g., ascorbate deficiency) or worsen (e.g., alcohol use or hepatitis) the outcome. The most common consequences of iron overload are liver disease, pancreatic disease (associated with diabetes mellitus), cardiac dysfunction, endocrine disorders (associated with gonadal insufficiency), arthropathy, and, with some forms of iron overload, specific neurologic abnormalities.

Increased iron absorption may develop because of primary disorders leading to abnormal control of iron absorption, such as hereditary hemochromatosis, or as a secondary consequence of acquired or inherited conditions, such as chronic liver disease or the iron-loading anemias. Iron overload may also result from chronic red blood cell transfusion, which bypasses intestinal control of iron uptake [*see Table 3*].

PRIMARY IRON OVERLOAD

By far the most common forms of primary iron overload are those related to mutations in the *HFE* gene. Some of the genes responsible for less common forms of primary iron overload have now been identified, providing valuable insights into the control of iron metabolism. Testing for mutations in the *HFE* gene has become an essential step in the evaluation of primary iron overload, and continued progress is anticipated in genetic characterization of these disorders.

HFE-Associated Hereditary Hemochromatosis

Epidemiology Hereditary HFE-associated hemochromatosis (hemochromatosis type 1), an autosomal recessive disease, is the most common genetic disorder in persons of northern European descent.[72] Data from the Third National Health and Nutrition Examination Survey (NHANES III) suggest that in the United States, 9.54% of the non-Hispanic white population is heterozygous and 0.30% is homozygous for the most common mutation in *HFE,* C282Y (see below).[73]

Etiology and genetics The discovery of *HFE*, the gene on chromosome 6p21.3 that is mutated in most cases of hereditary hemochromatosis,[40] has revolutionized both the understanding and the diagnosis of this disorder. Although the underlying mechanism is still not understood, a defect in the HFE protein results in an inappropriate increase in iron absorption that leads to a progressive buildup of body iron. Initially, the overload has a predominantly parenchymal pattern of deposition, with iron first

accumulating in hepatocytes; subsequently, the iron builds up in the pancreas, heart, and other organs.[74,75] Characteristically, macrophage iron levels in the bone marrow may be normal or even decreased despite severe parenchymal iron deposition [*see Table 1*].

Two missense mutations in the *HFE* gene, usually designated as C282Y and H63D, are responsible for up to 85% of the cases of hereditary hemochromatosis in the United States[73]; in other areas of the world, the percentage ranges from about 60% to almost 100%.[75] In the United States, 15% or more of patients with primary iron overload have neither of these mutations but are clinically indistinguishable from patients who do have one of these mutations.[40] Some of these patients are found to have other mutations in the *HFE* gene. Patients without evidence of mutations on chromosome 6p are classified as having non–HFE-associated hereditary hemochromatosis. The proportion of patients with non–HFE-associated hemochromatosis is higher in populations from southern Europe.[76]

Clinical manifestations Homozygotes for hereditary hemochromatosis may have no distinctive clinical manifestations, especially at younger ages. In homozygotes who present with hereditary hemochromatosis in middle age or later, the classic tetrad of clinical signs is liver disease, diabetes mellitus, skin pigmentation, and gonadal failure.[74,75] Arthropathy may be an early manifestation of the disease and is frequently present in patients with advanced disease. Cardiac failure may develop and may even be the presenting symptom in untreated homozygotes. Body iron stores have usually increased from the normal amount of 1 g or less to 15 to 20 g or more by the time symptoms of parenchymal damage occur, usually in middle or late life. Additional increases in body iron may be fatal, although some patients are able to tolerate a total iron accumulation of as much as 40 to 50 g or more.[77] Men are affected at younger ages than women, presumably because of iron losses during menstruation and childbearing. Environmental factors (e.g., dietary iron content, blood donation or loss, and alcohol use) and coexisting disorders (e.g., viral hepatitis) may greatly influence the rate and severity of organ damage.[74] The penetrance of HFE-associated hereditary hemochromatosis is a subject of controversy. In a study of 41,038 patients attending a health-appraisal clinic in the United States, the results were interpreted as showing that less than 1% of homozygotes develop frank clinical hemochromatosis,[78,79] although this interpretation has been questioned.[80] The Hemochromatosis and Iron Overload Screening (HEIRS) study of more than 100,000 adults should help clarify both the penetrance and the prevalence of HFE-associated hemochromatosis and other forms of iron overload in the United States.[81]

Table 3 Causes of Iron Overload

Primary	Hereditary hemochromatosis HFE associated (type 1) Non–HFE associated Transferrin receptor 2 associated (type 3) Juvenile hemochromatosis (type 2) Chromosome 1q21 associated (type 2A) Hepcidin associated (type 2B) Autosomal dominant hemochromatosis Ferroportin 1 associated (type 4) Ferritin H–subunit mRNA A49U mutation associated Atransferrinemia Aceruloplasminemia
Secondary	Iron-loading anemias (refractory anemias with hypercellular erythroid marrow) Chronic liver disease Porphyria cutanea tarda Insulin resistance–associated hepatic iron overload African dietary iron overload* Medicinal iron ingestion* Parenteral iron overload Transfusional iron overload Inadvertent iron overload from therapeutic injections
Perinatal	Neonatal hemochromatosis Hereditary tyrosinemia (hypermethionemia) Cerebrohepatorenal syndrome GRACILE (Fellman) syndrome
Focal sequestration of iron	Idiopathic pulmonary hemosiderosis Renal hemosiderosis Associated with neurologic abnormalities Pantothenate kinase–associated neurodegeneration (formerly Hallervorden-Spatz syndrome) Neuroferritinopathy Friedreich ataxia

*May have a genetic component.
GRACILE—growth retardation, aminoaciduria, cholestasis, iron overload, lactacidosis, and early death

Screening and diagnostic tests Screening for hereditary hemochromatosis can use both phenotypic and genotypic methods and is indicated for patients with chronic liver disease or symptoms and signs associated with iron overload.[75,82] Phenotypic screening can provide biochemical evidence of iron overload, but no single test or combination of tests will identify all patients who are genetically susceptible to iron loading.[78] Genotypic screening for the most common *HFE* mutations, C282Y and H63D, in populations of northern European ancestry can identify a majority of those patients at risk for the development of primary iron overload. In pedigree studies, genotyping should replace HLA typing in the assessment of siblings of a C282Y homozygote.[75] In addition, genotyping the spouse of a C282Y homozygote is a cost-efficient strategy that leads to a more selective investigation of children for the hemochromatosis gene. Nonetheless, such limited genetic screening will not detect other mutations associated with iron loading; is less useful in other population groups, such as those originating from southern Europe,[76] Africa,[83] or Asia; and provides no indication of the extent of iron excess. Population screening has been advocated[82,84] but has not been undertaken generally because of uncertainties about disease penetrance[78]; the disease burden and natural history of iron overload; and a variety of ethical, legal, and social concerns.[75,78]

Measurement of the serum transferrin saturation is usually recommended as the initial phenotypic screening determination.[75,82,84,85] Although individual laboratories may have their own reference ranges, a persistent value of 45% or higher is often recommended as a threshold value for further investigation. The serum ferritin concentration is then used as a biochemical indicator of iron overload, and in the absence of complicating factors, increased concentrations suggest increased iron stores.[74] Genetic testing should then be considered in patients with abnormal elevations in transferrin saturation, serum ferritin, or both. The exact role of genetic testing in screening and diagnosis depends in part on the population being examined because of variations in the proportion of patients with hereditary hemochromatosis who have *HFE* mutations.

Once genetic testing has identified a patient as homozygous for the C282Y mutation (i.e., C282Y/C282Y), an elevated transferrin saturation establishes the diagnosis of hereditary hemochromatosis.[85] Liver biopsy, formerly a standard part of the diagnostic process, is no longer needed in most cases. Liver biopsy is indicated to detect cirrhosis if the serum ferritin level is above 1,000 μg/L[86]; biopsy may also be considered in patients with hepatomegaly, patients with abnormal findings on liver function tests, and patients who are older than 40 years.[74,85,87]

Patients who are heterozygous for the C282Y mutation and wild type (i.e., C282Y/wild type) have serum transferrin saturations and serum ferritin concentrations that are similar to those of wild-type homozygotes (i.e., wild type/wild type),[88,89] and these patients do not develop clinically important iron overload.[88] Consequently, finding a persistently elevated level of transferrin saturation, serum ferritin concentration, or both in a C282Y/wild type heterozygote should lead the clinician to search for other causes of iron overload.[85]

In patients who are heterozygous for the C282Y mutation and H63D, the other major *HFE* mutation (i.e., C282Y/H63D), mild to moderate iron overload may develop, but the penetrance of this genotype is even less than that of C282Y/C282Y homozygotes.[74,75,85,87] Heterozygotes for H63D (i.e., H63D/wild type) may have elevated transferrin saturation levels[90] but do not develop iron overload. Homozygotes for H63D (i.e., H63D/H63D) also may have elevated transferrin saturation levels, but the risk of clinically important iron overload is slight.[90] Less common mutations of the *HFE* gene have been identified[91]; S65C heterozygotes with either C282Y (i.e., S65C/C282Y) or H63D (i.e., S65C/H63D) may develop mild iron overload.

Several diagnostic approaches can be pursued in patients with phenotypic evidence of iron overload who are neither homozygous for C282Y (C282Y/C282Y) nor heterozygous for C282Y and H63D (C282Y/H63D). These approaches include further genetic testing for less common *HFE* mutations and for non-*HFE* mutations associated with iron loading; noninvasive assessment of the liver iron concentration[92]; and liver biopsy, which permits a definitive diagnosis of hereditary hemochromatosis regardless of genotype.[74,85,87] Evaluation of the biopsy specimen should include quantitative determination of the nonheme iron concentration, histochemical evaluation of the pattern of iron deposition, and pathologic assessment of tissue injury. Calculation of the hepatic iron index—the hepatic iron concentration (expressed as μmol Fe/g of liver, dry weight) divided by the age of the patient in years—may be helpful in distinguishing homozygotes for hereditary hemochromatosis from heterozygotes or from patients with increased body iron associated with chronic (usually alcoholic) liver disease.[74,87] In the absence of other causes of iron overload, a hepatic iron index greater than 1.9 is evidence for hereditary hemochromatosis. In patients with evidence of increased body iron levels, further evaluation should be directed toward detecting complications of iron overload and may include liver function tests, assessment of glucose tolerance and hormonal function, cardiac examination, joint and bone x-rays, and, especially if cirrhosis is present, screening for hepatocellular carcinoma.[74,85,87]

Treatment The treatment of choice for hereditary hemochromatosis is phlebotomy to reduce the body iron levels to normal or near-normal and maintain them in that range.[74,85,87] In patients with hereditary hemochromatosis who develop cardiac failure, the use of both phlebotomy and chelation therapy has been suggested. Phlebotomy therapy should be started as soon as the diagnosis of the homozygous state for hereditary hemochromatosis has been established; postponement only increases the risk of organ damage from iron overload. The phlebotomy program should remove 500 ml of blood (containing 200 to 250 mg of iron) once weekly or, for heavily loaded patients, twice weekly until the patient is iron deficient.[85] Before each phlebotomy, the hematocrit or hemoglobin concentration should be measured. Initially, the hematocrit and hemoglobin levels will decline by about 10% of their initial values but may then rise as the rate of erythropoiesis increases to match the demands of phlebotomy. Measurements of serum ferritin, iron, and transferrin saturation should be done regularly to follow the progress of iron removal. As iron is removed, the serum ferritin concentration will decrease progressively but the serum transferrin saturation will remain raised until iron stores are almost exhausted. Finally, when all the storage iron has been removed, the ferritin concentration will fall to less than 12 μg/L, the serum iron concentration and transferrin saturation will drop, and the hemoglobin concentration will decrease to less than 10 g/dl for 2 weeks without further phlebotomy. In patients with hereditary hemochromatosis, prolonged treatment is often needed. For example, if the initial body iron burden is 25 g, complete removal of the iron burden with weekly phlebotomy may require 2 years or more. After the iron load has been completely removed, a lifelong program of maintenance phlebotomy is required to prevent reaccumulation of the iron burden.[85] Typically, phlebotomy of 500 ml of blood every 3 to 4 months is needed. The goal of maintenance phlebotomy should be to maintain a serum ferritin concentration of less than about 50 μg/L.

If phlebotomy therapy removes the iron load before diabetes mellitus or cirrhosis develops, the patient's life expectancy is normal.[93] If cirrhosis develops, however, the risk of hepatocellular carcinoma is increased by more than 200-fold.[74] In hereditary hemochromatosis, hepatomas develop almost exclusively in patients with hepatic cirrhosis and are the ultimate cause of death in 20% to 30% of these patients, even after successful removal of the iron burden. Phlebotomy therapy is almost always indicated for patients with hereditary hemochromatosis, even when cirrhosis or organ damage is already present, because further progression of the disease can be stopped and alleviation of some organ dysfunction is possible.

Non–HFE-Associated Hereditary Hemochromatosis

Clinically, non–HFE-associated hereditary hemochromatosis is indistinguishable from hereditary hemochromatosis associated with *HFE* mutations. Genetically, the non-HFE disorder is heterogeneous. One subset of this autosomal recessive disorder, designated as hemochromatosis type 3, is caused by mutations of the gene encoding transferrin receptor 2 on chromosome 7q22.[45]

Juvenile Hemochromatosis

Juvenile hemochromatosis, designated as hemochromatosis type 2, is a rare autosomal recessive disorder in which severe iron overload develops before age 30. The two sexes are affected equally, and patients may present with cardiomyopathy, hypogonadism, impaired glucose tolerance, or some combination of these manifestations.[94] Genetically, two subtypes have so far been distinguished. Type 2A shows linkage to chromosome 1q21, but the responsible gene has not yet been identified.[95] Type 2B is caused by mutations in the gene for hepcidin on chromosome 19q13.[46]

Autosomal Dominant Hemochromatosis

Hemochromatosis with an autosomal dominant pattern of inheritance is also genetically heterogeneous, and at least one variety has been designated as hemochromatosis type 4. Several families have been identified with iron overload associated with mutations in the ferroportin gene (*SLC11A3*) on chromosome 2q32. Characteristically, initial iron accumulation is predominantly reticuloendothelial and the serum ferritin concentration is increased, with relatively low transferrin saturation; mild anemia early in life has been reported in a number of those affected. Some, but not all, of those with ferroportin mutations have also developed parenchymal iron overload.[96] A single Japanese family has been described with autosomal dominant iron overload ascribed to a point mutation (A49U) in the iron-responsive element (IRE) of H-ferritin mRNA.[97] Autosomal dominant hemochromatosis has also been reported in a Melanesian kindred, but the genetic basis has not been determined.[98]

Atransferrinemia and Aceruloplasminemia

Iron overload may also result from two rare autosomal recessive disorders in which the synthesis of plasma proteins vital for iron transport is absent or almost absent. In atransferrinemia, dietary iron is readily absorbed and circulates as nontransferrin plasma iron but cannot be used for erythropoiesis because of the lack of a physiologic means of transport into developing erythroid cells; affected persons die unless they receive transferrin infusion or blood transfusion.[99] In aceruloplasminemia, the deficiency of ceruloplasmin ferroxidase activity results in iron accumulation in the liver, pancreas, and brain, with smaller amounts of excess iron found in the spleen, heart, kidney, thyroid, and retina. Patients present with progressive neurodegeneration of the retina and basal ganglia and with diabetes mellitus in middle age.[100]

SECONDARY IRON OVERLOAD

Secondary iron overload may result from increased gastrointestinal absorption of iron, from transfusion of red blood cells, from inadvertent iatrogenic parenteral administration of iron, or from some combination thereof. Despite the progress made in understanding genetically determined increases in intestinal iron uptake, the pathophysiologic mechanisms responsible for the increased absorption of iron in secondary iron overload are still obscure.

Iron-Loading Anemias

The iron-loading anemias include congenital dyserythropoietic anemia, pyruvate kinase deficiency, thalassemia major (Cooley anemia) and thalassemia intermedia, hemoglobin E–β-thalassemia, a variety of forms of sideroblastic anemia, some myelodysplastic anemias, and other anemic disorders in which the incorporation of iron into hemoglobin is impaired. In patients with iron-loading anemias, severe iron overload may develop as a result of increased gastrointestinal iron absorption. Any red cell transfusions these patients receive will contribute to the iron loading. Because the extent of ineffective erythropoiesis, not the severity of the anemia, seems to determine the rate of iron loading, severe iron overload may develop in patients with only slight or mild anemia.[101] The clinical manifestations and pathology that may develop in patients with iron-loading anemias are similar to those seen in hereditary hemochromatosis, including liver disease, diabetes mellitus, endocrine disorders, and cardiac dysfunction.[102] Suppression of hepcidin synthesis by anemia, hypoxia, or both has been suggested as a potential mechanism for the increased iron absorption,[103] but the distinctive influence of ineffective erythropoiesis remains to be explained.

Other Causes and Forms of Absorption-Related Iron Overload

Chronic liver disease Some patients with chronic liver disease, including those with alcoholic cirrhosis and those with portacaval shunting, may experience minor or modest degrees of iron loading as a result of increased dietary iron absorption.[74] The mechanisms responsible for the increased gastrointestinal iron uptake have not been identified, although ineffective erythropoiesis and hyperferremia associated with alcohol-induced folate and sideroblastic abnormalities have been proposed as etiologic factors.[104] Body iron stores are increased only to a minor degree, typically to 2 to 4 g, but in alcoholic cirrhosis, the higher the liver iron, the shorter the survival.

Porphyria cutanea tarda Symptomatic patients with porphyria cutanea tarda, a hepatic porphyria, usually have a modest increase in body iron levels that almost always is the result of increased gastrointestinal absorption.[102] In patients who are of European ancestry, *HFE* mutations are common and may contribute to the pathogenesis of both the familial and the sporadic forms of the disorder.[105]

Insulin resistance–associated hepatic iron overload An iron-overload syndrome characterized by an increased serum ferritin level with a normal transferrin saturation level in association with glucose or lipid metabolic abnormalities, or both, was first described in 1997[106] and has come to be known as insulin resistance–associated hepatic iron overload.[107] The iron overload is typically mild or moderate, and the histologic appearance is distinct from that of HFE-associated hemochromatosis.

African dietary iron overload In sub-Saharan Africa, iron overload in association with greatly increased dietary iron intake from a fermented maize beverage home-brewed in steel drums has been described. Iron burdens may be as great as those found in hereditary hemochromatosis, and patients may develop liver disease (with cirrhosis and hepatoma), pancreatic disease (with diabetes mellitus), endocrine disorders, and cardiac dysfunction. Although increased dietary iron intake was long considered the sole cause of the increased iron absorption in this disorder, a series of pedigree studies has suggested that a genetic component may be involved and may be common in populations of African ancestry.[83]

Medicinal iron ingestion Ingestion of iron supplements can undoubtedly contribute to the body iron burden of patients with iron-loading disorders, but the extent to which orally administered iron can increase the body iron stores of normal individuals remains uncertain. Although some case reports have described iron accumulation in patients who have taken medicinal iron for long periods, the potential involvement of an unrecognized iron-loading mutation in these individuals cannot be excluded.

Transfusional and Other Parenteral Iron Overload

Etiology and diagnosis An adequate transfusion program can sustain life in patients with severe chronic refractory anemia, but transfusion therapy alone produces a progressive accumulation of the iron contained in transfused red cells.[108,109]

Iron accumulation from transfusion initially occurs predominantly in macrophage sites, followed by redistribution to parenchymal tissues. In patients with severe congenital anemias, such as thalassemia major and the Blackfan-Diamond syndrome, regular transfusions can prevent death from anemia in infancy and permit normal growth and development during childhood. Treatment of acquired transfusion-dependent anemias, such as aplastic anemia, pure red cell aplasia, and hypoplastic or myelodysplastic disorders, among others, may result in the development of marked iron overload. If the transfusion-dependent anemia includes erythroid hyperplasia with ineffective erythropoiesis, increased gastrointestinal iron absorption may add to the iron loading. In such cases, dietary iron uptake may be minimized by suppression of erythropoiesis with an adequate transfusion program. Although sickling disorders (e.g., sickle cell anemia and sickle cell–β-thalassemia) are not transfusion-dependent, these patients may acquire a considerable iron load from repeated transfusions for the prevention of stroke, painful crises, and other recurrent complications.[110] Because humans lack a physiologic means of eliminating excess iron, iron contained in transfused red cells progressively accumulates and eventually damages the liver, heart, pancreas, and other organs; death usually occurs from cardiac failure. In younger patients, the iron burden results in growth failure and, in adolescence, delayed or absent sexual maturation. Parenteral medicinal iron may needlessly add to the iron burden in patients with refractory microcytic anemias who are misdiagnosed as iron deficient.

Treatment About 200 to 250 mg of iron is added to the body iron load with each unit of transfused red cells. Most transfusion-dependent patients require 200 to 300 ml/kg of blood a year; for example, a 70 kg adult requires about two to three units of blood every 3 to 4 weeks, adding about 6 to 10 g of iron a year. The severity of iron toxicity seems to be related to the magnitude of the body iron burden. Almost all patients who have been treated with transfusion alone and have received 100 or more units of blood (about 20 to 25 g of iron) have developed cardiac iron deposits, often in association with signs of hepatic, pancreatic, and endocrine damage.[108,109] For patients who are transfusion dependent or severely anemic, the only way to prevent iron overload is treatment with a chelating agent capable of complexing with iron and permitting its excretion. The only iron-chelating agent now available for clinical use in North America is deferoxamine B, a siderophore produced by *Streptomyces pilosus*. Clinical trials with deferoxamine have documented the effectiveness of iron chelation as therapy for iron overload, demonstrating that regular iron chelation can decrease the body iron burden, alleviate organ dysfunction, and improve survival.[108] Although toxic side effects can occur, especially with intensive therapy, deferoxamine has been a remarkably safe drug, even with near-lifelong use in some patients.[109]

Iron chelation therapy should be started early to prevent the accumulation of toxic amounts of iron in vulnerable tissues and to maintain body iron stores at concentrations associated with a low risk of early death and clinical complications. The longer chelation therapy is delayed, the greater the risk of iron toxicity. Because deferoxamine is poorly absorbed after oral administration and rapidly eliminated from the circulation, deferoxamine must be given by slow subcutaneous or intravenous infusion over 9 to 12 hours each day at least 5 days a week to be optimally useful in the treatment of patients with transfusional iron overload.[108,109] In patients with thalassemia major and other congenital refractory anemias who have been transfusion dependent from early infancy, chelation therapy is best started after about 10 to 20 transfusions, usually when the patient is 3 or 4 years of age.[108,109] Deferoxamine is administered by slow subcutaneous infusion at a dosage not exceeding 25 mg/kg/day to minimize the risk of growth retardation. In older patients and adults with acquired refractory anemias who require regular transfusion and in patients with sickle cell disease who are chronically transfused for prevention of complications, early therapy also seems prudent, beginning after transfusion of 10 to 20 units of blood. The usual dosage of deferoxamine in these older patients is not more than 50 mg/kg/day, given over 9 to 12 hours by slow subcutaneous infusion at least 5 days a week. In some patients who are unable to tolerate the local pain and discomfort of subcutaneous infusion or who need rapid reduction of high body iron burdens, deferoxamine may be administered intravenously through implantable venous access ports. Compliance with these near-daily regimens of prolonged parenteral infusions may be difficult, and lack of compliance is the chief obstacle to effective iron chelation therapy. Administration of ascorbic acid can enhance deferoxamine-induced iron excretion but carries the risk of an internal redistribution of iron from relatively benign storage sites in macrophages to a potentially toxic pool in parenchymal cells. Although the evidence is anecdotal, large doses of ascorbic acid should be regarded as hazardous in patients with iron overload. Although deferoxamine is a generally safe and nontoxic drug in the iron-loaded patient, systemic complications have been reported, including allergic anaphylactoid reactions, infections, visual abnormalities and auditory dysfunction, and growth retardation.[108,109] As a result, regular evaluation for drug toxicity should be included in the management of any patient receiving deferoxamine, including annual audiograms, retinal examination, and assessment of growth in children and adolescents. The risk of many of these complications may be minimized by adjusting the deferoxamine dose to the magnitude of the body iron load. New and promising chelating agents are in development[111] or in clinical trials,[112] but none is yet available for clinical use.

Additional Rare or Uncommon Forms of Iron Overload

Perinatal iron overload occurs in several forms. Neonatal hemochromatosis is a heterogeneous group of disorders associated with severe congenital hepatic disease and deposits of iron in the liver, pancreas, heart, and other extrahepatic sites, with evidence of autosomal recessive inheritance in some cases.[113] Several rare metabolic abnormalities of the neonate may be associated with abnormal iron deposition, including hereditary tyrosinemia (hypermethioninemia); Zellweger cerebrohepatorenal syndrome[114]; the tricho-hepato-enteric syndrome; and the GRACILE (or Fellman) syndrome.[113]

Focal sequestration of iron is found in other rare disorders, including idiopathic pulmonary hemosiderosis and renal hemosiderosis. Such abnormal iron deposition is associated with neurologic abnormalities in Friedreich ataxia, in pantothenate kinase–associated neurodegeneration (formerly Hallervorden-Spatz syndrome), and in neuroferritinopathy.[115] Finally, hyperferritinemia with autosomal dominant congenital cataract[116] is a disorder of iron metabolism in which affected family members present with early-onset, bilateral nuclear cataracts and moderately elevated serum ferritin concentrations.[116] The body iron level is

normal in these patients, but overload is often suspected because of the elevated serum ferritin concentrations.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Klinken S: Red blood cells. Internat J Biochem Cell Biol 34:1513, 2002

2. Yonetani T, Park SI, Tsuneshige A, et al: Global allostery model of hemoglobin: modulation of O(2) affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors. J Biol Chem 277:34508, 2002

3. Hsia CC: Respiratory function of hemoglobin. N Engl J Med 338:239, 1998

4. McMahon TJ, Moon RE, Luschinger BP, et al: Nitric oxide in the human respiratory cycle. Nat Med 8:711, 2002

5. Schechter AN, Gladwin MT: Hemoglobin and the paracrine and endocrine functions of nitric oxide. N Engl J Med 348:1483, 2003

6. Discher DE, Carl P: New insights into red cell network structure, elasticity, and spectrin unfolding—a current review. Cell Mol Biol Lett 6:593, 2001

7. Jacobasch G: Biochemical and genetic basis of red cell enzyme deficiencies. Baillieres Best Pract Res Clin Haematol 13:1, 2000

8. Henry ER, Bettati S, Hofrichter J, et al: A tertiary two-state allosteric model for hemoglobin. Biophys Chem 98:149, 2002

9. Leach RM, Treacher DF: The pulmonary physician in critical care: 2: oxygen delivery and consumption in the critically ill. Thorax 57:170, 2002

10. Fisher JW: Erythropoietin: physiology and pharmacology update. Exp Biol Med (Maywood) 228:1, 2003

11. Trent JT 3rd, Watts RA, Hargrove MS: Human neuroglobin, a hexacoordinate hemoglobin that reversibly binds oxygen. J Biol Chem 276:30106, 2001

12. Burmester T, Ebner B, Weich B, et al: Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol Biol Evol 19:416, 2002

13. Geers C, Gros G: Carbon dioxide transport and carbonic anhydrase in blood and muscle. Physiol Rev 80:681, 2000

14. Andrews NC: Disorders of iron metabolism. N Engl J Med 341:1986, 1999

15. Aisen P, Enns C, Wessling-Resnick M: Chemistry and biology of eukaryotic iron metabolism. Int J Biochem Cell Biol 33:940, 2001

16. Finch C: Regulators of iron balance in humans. Blood 84:1697, 1994

17. Zakin MM, Baron B, Guillou F: Regulation of the tissue-specific expression of transferrin gene. Dev Neurosci 24:222, 2002

18. Trinder D, Baker E: Transferrin receptor 2: a new molecule in iron metabolism. Int J Biochem Cell Biol 35:292, 2003

19. Andrews NC: A genetic view of iron homeostasis. Semin Hematol 39:227, 2002

20. Arosio P, Levi S: Ferritin, iron homeostasis, and oxidative damage. Free Radic Biol Med 33:457, 2002

21. Levi S, Corsi B, Bosisio M, et al: A human mitochondrial ferritin encoded by an intronless gene. J Biol Chem 276:24437, 2001

22. Cazzola M, Invernizzi R, Bergamaschi G, et al: Mitochondrial ferritin expression in erythroid cells from patients with sideroblastic anemia. Blood 101:1996, 2003

23. Rouault TA: Systemic iron metabolism: a review and implications for brain iron metabolism. Pediatr Neurol 25:130, 2001

24. Hubert N, Hentze MW: Previously uncharacterized isoforms of divalent metal transporter (DMT)-1: implications for regulation and cellular function. Proc Natl Acad Sci USA 99:12345, 2002

25. Abboud S, Haile DJ: A novel mammalian iron-regulated protein involved in intracellular iron metabolism. J Biol Chem 275:19906, 2000

26. Cairo G, Recalcati S, Pietrangelo A, et al: The iron regulatory proteins: targets and modulators of free radical reactions and oxidative damage. Free Radic Biol Med 32:1237, 2002

27. Knutson M, Wessling-Resnick M: Iron metabolism in the reticuloendothelial system. Crit Rev Biochem Mol Biol 38:61, 2003

28. Kondo H, Saito K, Grasso JP, et al: Iron metabolism in the erythrophagocytosing Kupffer cell. Hepatology 8:32, 1988

29. Kristiansen M, Graversen JH, Jacobsen C, et al: Identification of the haemoglobin scavenger receptor. Nature 409:198, 2001

30. Wilks A. Heme oxygenase: evolution, structure, and mechanism. Antioxid Redox Signal 4:603, 2002

31. Baranano DE, Wolosker H, Bae BI, et al: A mammalian iron ATPase induced by iron. J Biol Chem 275:15166, 2000

32. Pietrangelo A: Physiology of iron transport and the hemochromatosis gene. Am J Physiol Gastrointest Liver Physiol 282:G403, 2002

33. Hellman NE, Gitlin JD: Ceruloplasmin metabolism and function. Annu Rev Nutr 22:439, 2002

34. Tolosano E, Altruda F: Hemopexin: structure, function, and regulation. DNA Cell Biol 21:297, 2002

35. Bothwell TH, Charlton RW, Cook JD, et al: Iron Metabolism in Man. Blackwell, London, England, 1979

36. Miret S, Simpson RJ, McKie AT: Physiology and molecular biology of dietary iron absorption. Annu Rev Nutr 23:283, 2003

37. Conrad ME, Umbreit JN: Pathways of iron absorption. Blood Cells Mol Dis 29:336, 2002

38. McKie AT, Latunde-Dada GO, Miret S, et al: Molecular evidence for the role of a ferric reductase in iron transport. Biochem Soc Trans 30:722, 2002

39. Anderson GJ, Frazer DM, McKie AT, et al: The expression and regulation of the iron transport molecules hephaestin and IREG1: implications for the control of iron export from the small intestine. Cell Biochem Biophys 36:137, 2002

40. Feder JN, Gnirke A, Thomas W, et al: A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399, 1996

41. Fleming RE, Sly WS: Mechanisms of iron accumulation in hereditary hemochromatosis. Annu Rev Physiol 64:663, 2002

42. Drakesmith H, Sweetland E, Schimanski L, et al: The hemochromatosis protein HFE inhibits iron export from macrophages. Proc Natl Acad Sci USA 99:15602, 2002

43. Townsend A, Drakesmith H: Role of HFE in iron metabolism, hereditary haemochromatosis, anaemia of chronic disease, and secondary iron overload. Lancet 359:786, 2002

44. Frazer DM, Anderson GJ: The orchestration of body iron intake: how and where do enterocytes receive their cues? Blood Cells Mol Dis 30:288, 2003

45. Camaschella C, Roetto A, Cali A, et al: The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. Nat Genet 25:14, 2000

46. Roetto A, Papanikolaou G, Politou M, et al: Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. Nat Genet 33:21, 2003

47. Bridle KR, Frazer DM, Wilkins SJ, et al: Disrupted hepcidin regulation in HFE-associated haemochromatosis and the liver as a regulator of body iron homoeostasis. Lancet 361:669, 2003

48. Ramakrishnan U, Yip R: Experiences and challenges in industrialized countries: control of iron deficiency in industrialized countries. J Nutr 132:820S, 2002

49. Iron Deficiency—United States, 1999–2000. MMWR Morb Mortal Wkly Rep 51:897, 2002

50. Goddard AF, McIntyre AS, Scott BB: Guidelines for the management of iron deficiency anaemia. British Society of Gastroenterology. Gut 46(suppl 3-4):IV1, 2000

51. Recommendations to prevent and control iron deficiency in the United States. MMWR Recomm Rep 47(RR-3):1, 1998

52. Willoughby JM, Laitner SM: Audit of the investigation of iron deficiency anaemia in a district general hospital, with sample guidelines for future practice. Postgrad Med J 76:218, 2000

53. Mukhopadhyay D, Mohanaruban K: Iron deficiency anaemia in older people: investigation, management and treatment. Age Ageing 31:87, 2002

54. Annibale B, Capurso G, Chistolini A, et al: Gastrointestinal causes of refractory iron deficiency anemia in patients without gastrointestinal symptoms. Am J Med 111:439, 2001

55. Sandoval C, Berger E, Ozkaynak MF, et al: Severe iron deficiency anemia in 42 pediatric patients. Pediatr Hematol Oncol 19:157, 2002

56. Lee PL, Halloran C, Trevino R, et al: Human transferrin G277S mutation: a risk factor for iron deficiency anaemia. Br J Haematol 115:329, 2001

57. Earley CJ: Clinical practice. Restless legs syndrome. N Engl J Med 348:2103, 2003

58. Kalra L, Hamlyn AN, Jones BJ: Blue sclerae: a common sign of iron deficiency? Lancet 2:1267, 1986

59. Rector WG Jr: Pica: its frequency and significance in patients with iron-deficiency anemia due to chronic gastrointestinal blood loss. J Gen Intern Med 4:512, 1989

60. Haas JD, Brownlie T 4th: Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. J Nutr 131:676S, 2001

61. Oppenheimer SJ: Iron and its relation to immunity and infectious disease. J Nutr 131:616S, 2001

62. Grantham-McGregor S, Ani C: A review of studies on the effect of iron deficiency on cognitive development in children. J Nutr 131:649S, 2001

63. Cook J: The nutritional assessment of iron status. Arch Latinoam Nutr 49:11S, 1999

64. Cook JD: The measurement of serum transferrin receptor. Am J Med Sci 318:269, 1999

65. Malope BI, MacPhail AP, Alberts M, et al: The ratio of serum transferrin receptor and serum ferritin in the diagnosis of iron status. Br J Haematol 115:84, 2001

66. Beguin Y: Soluble transferrin receptor for the evaluation of erythropoiesis and iron status. Clin Chim Acta 329:9, 2003

67. Cook JD, Flowers CH, Skikne BS: The quantitative assessment of body iron. Blood 101:3359, 2003

68. Brugnara C: Reticulocyte cellular indices: a new approach in the diagnosis of anemias and monitoring of erythropoietic function. Crit Rev Clin Lab Sci 37:93, 2000

69. Thomas C, Thomas L: Biochemical markers and hematologic indices in the diagnosis of functional iron deficiency. Clin Chem 48:1066, 2002

70. Goodnough LT, Skikne B, Brugnara C: Erythropoietin, iron, and erythropoiesis. Blood 96:823, 2000

71. McDiarmid T, Johnson ED: Are any oral iron formulations better tolerated than ferrous sulfate? J Fam Pract 51:576, 2002

72. Milman N, Pedersen P: Evidence that the Cys282Tyr mutation of the HFE gene originated from a population in Southern Scandinavia and spread with the Vikings. Clin Genet 64:36, 2003

73. Steinberg KK, Cogswell ME, Chang JC, et al: Prevalence of C282Y and H63D mutations in the hemochromatosis (HFE) gene in the United States. JAMA 285:2216, 2001

74. Powell LW: Hereditary hemochromatosis and iron overload diseases. J Gastroenterol Hepatol 17(suppl):S191, 2002

75. Bomford A: Genetics of haemochromatosis. Lancet 360:1673, 2002

76. Barosi G, Salvaneschi L, Grasso M, et al: High prevalence of a screening-detected, HFE-unrelated, mild idiopathic iron overload in Northern Italy. Haematologica 87:472, 2002

77. Adams PC, Deugnier Y, Moirand R, et al: The relationship between iron overload, clinical symptoms, and age in 410 patients with genetic hemochromatosis. Hepatology 25:162, 1997

78. Beutler E, Felitti VJ, Koziol JA, et al: Penetrance of 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. Lancet 359:211, 2002

79. Beutler E: The HFE Cys282Tyr mutation as a necessary but not sufficient cause of clinical hereditary hemochromatosis. Blood 101:3347, 2003

80. Ajioka RS, Kushner JP: Clinical consequences of iron overload in hemochromatosis homozygotes. Blood 101:3351, 2003

81. McLaren CE, Barton JC, Adams PC, et al: Hemochromatosis and Iron Overload Screening (HEIRS) study design for an evaluation of 100,000 primary care-based adults. Am J Med Sci 325:53, 2003

82. Niederau C, Strohmeyer G: Strategies for early diagnosis of haemochromatosis. Eur J Gastroenterol Hepatol 14:217, 2002

83. Gordeuk VR: African iron overload. Semin Hematol 39:263, 2002

84. Adams P, Brissot P, Powell LW: EASL International Consensus Conference on Haemochromatosis. J Hepatol 33:485, 2000

85. Brissot P, Guyader D, Loreal O, et al: Clinical aspects of hemochromatosis. Transfus Sci 23:193, 2000

86. Morrison ED, Brandhagen DJ, Phatak PD, et al: Serum ferritin level predicts advanced hepatic fibrosis among U.S. patients with phenotypic hemochromatosis. Ann Intern Med 138:627, 2003

87. Pietrangelo A: Haemochromatosis. Gut 52(suppl 2):ii23, 2003

88. Moirand R, Guyader D, Mendler MH, et al: HFE based re-evaluation of heterozygous hemochromatosis. Am J Med Genet 111:356, 2002

89. Beutler E, Felitti V, Gelbart T, et al: Haematological effects of the C282Y HFE mutation in homozygous and heterozygous states among subjects of northern and southern European ancestry. Br J Haematol 120:887, 2003

90. Gochee PA, Powell LW, Cullen DJ, et al: A population-based study of the biochemical and clinical expression of the H63D hemochromatosis mutation. Gastroenterology 122:646, 2002

91. Pointon JJ, Wallace D, Merryweather-Clarke AT, et al: Uncommon mutations and polymorphisms in the hemochromatosis gene. Genet Test 4:151, 2000

92. Brittenham GM, Badman DG: Noninvasive measurement of iron: report of an NIDDK workshop. Blood 101:15, 2003

93. Niederau C, Fischer R, Purschel A, et al: Long-term survival in patients with hereditary hemochromatosis. Gastroenterology 110:1107, 1996

94. De Gobbi M, Roetto A, Piperno A, et al: Natural history of juvenile haemochromatosis. Br J Haematol 117:973, 2002

95. Roetto A, Totaro A, Cazzola M, et al: Juvenile hemochromatosis locus maps to chromosome 1q. Am J Hum Genet 64:1388, 1999

96. Cazzola M, Cremonesi L, Papaioannou M, et al: Genetic hyperferritinaemia and reticuloendothelial iron overload associated with a three base pair deletion in the coding region of the ferroportin gene (SLC11A3). Br J Haematol 119:539, 2002

97. Kato J, Fujikawa K, Kanda M, et al: A mutation, in the iron-responsive element of H ferritin mRNA, causing autosomal dominant iron overload. Am J Hum Genet 69:191, 2001

98. Eason RJ, Adams PC, Aston CE, et al: Familial iron overload with possible autosomal dominant inheritance. Aust N Z J Med 20:226, 1990

99. Sheth S, Brittenham GM: Genetic disorders affecting proteins of iron metabolism: clinical implications. Annu Rev Med 51:443, 2000

100. Nittis T, Gitlin JD: The copper-iron connection: hereditary aceruloplasminemia. Semin Hematol 39:282, 2002

101. Cazzola M, Beguin Y, Bergamaschi G, et al: Soluble transferrin receptor as a potential determinant of iron loading in congenital anaemias due to ineffective erythropoiesis. Br J Haematol 106:752, 1999

102. Bottomley SS: Secondary iron overload disorders. Semin Hematol 35:77, 1998

103. Nicolas G, Chauvet C, Viatte L, et al: The gene encoding the iron regulatory peptide hepcidin is regulated by anemia, hypoxia, and inflammation. J Clin Invest 110:1037, 2002

104. Conrad ME, Barton JC: Anemia and iron kinetics in alcoholism. Semin Hematol 17:149, 1980

105. Bygum A, Christiansen L, Petersen NE, et al: Familial and sporadic porphyria cutanea tarda: clinical, biochemical and genetic features with emphasis on iron status. Acta Derm Venereol 83:115, 2003

106. Moirand R, Mortaji AM, Loreal O, et al: A new syndrome of liver iron overload with normal transferrin saturation. Lancet 349:95, 1997

107. Chitturi S, George J: Interaction of iron, insulin resistance, and nonalcoholic steatohepatitis. Curr Gastroenterol Rep 5:18, 2003

108. Olivieri NF, Brittenham GM: Iron-chelating therapy and the treatment of thalassemia. Blood 89:739, 1997

109. Porter JB: Practical management of iron overload. Br J Haematol 115:239, 2001

110. Files B, Brambilla D, Kutlar A, et al: Longitudinal changes in ferritin during chronic transfusion: a report from the Stroke Prevention Trial in Sickle Cell Anemia (STOP). J Pediatr Hematol Oncol 24:284, 2002

111. Wong C, Richardson DR: beta-Thalassaemia: emergence of new and improved iron chelators for treatment. Int J Biochem Cell Biol 35:1144, 2003

112. Nisbet-Brown E, Olivieri NF, Giardina PJ, et al: Effectiveness and safety of ICL670 in iron-loaded patients with thalassaemia: a randomised, double-blind, placebo-controlled, dose-escalation trial. Lancet 361:1597, 2003

113. Kelly AL, Lunt PW, Rodrigues F, et al: Classification and genetic features of neonatal haemochromatosis: a study of 27 affected pedigrees and molecular analysis of genes implicated in iron metabolism. J Med Genet 38:599, 2001

114. Goldfischer S, Grotsky HW, Chang CH, et al: Idiopathic neonatal iron storage involving the liver, pancreas, heart, and endocrine and exocrine glands. Hepatology 1:58, 1981

115. Ponka P: Rare causes of hereditary iron overload. Semin Hematol 39:249, 2002

116. Cazzola M: Hereditary hyperferritinaemia/cataract syndrome. Best Pract Res Clin Haematol 15:385, 2002

Acknowledgments

Figure 1 Seward Hung.

Figure 2 Marcia Kammerer.

Figure 3 Seward Hung.

Figure 4 Dimitry Schidlovsky.

III ANEMIA: PRODUCTION DEFECTS

STANLEY L. SCHRIER, M.D., F.A.C.P.

Classification of Production Defects

Red blood cell production defects cause anemia that is marked by a low absolute reticulocyte count. Examination of the peripheral blood count and the bone marrow aids in classifying these disorders. The marrow characteristically shows one of the following:

1. A normal ratio of myeloid cells to erythroid cells (M:E ratio), normal overall cellularity, and a normal pattern of erythroid maturation.
2. Virtual absence of normal bone marrow elements caused by aplasia (absence of marrow cells) or by replacement of normal marrow elements by fibrosis, solid tumors, granulomas, or leukemia.
3. Erythroid hyperplasia with increased cellularity. Because of defects of erythroid maturation, there is ineffective erythropoiesis or intramedullary hemolysis. Erythroid precursors die in the marrow, and few cells reach the periphery.

Production Defects Associated with Apparently Normal Bone Marrow

ANEMIA OF CHRONIC DISEASE

Definition

The anemia of chronic disease occurs secondary to neoplastic, infectious, and inflammatory diseases and other chronic illnesses, including liver disorders, congestive heart failure, and diabetes mellitus.[1,2] Hematocrit values usually range from 27% to 35%, although 20% of patients have hematocrit values below 25%.[2]

Pathophysiology

The anemia of chronic disease usually results from a combination of slightly shortened red blood cell survival, the sequestration of iron in the reticuloendothelial system, and erythropoietin levels that are less than expected for the degree of anemia.[1,2] Red blood cells usually have a normal morphologic appearance, although they may occasionally be mildly hypochromic and microcytic. The serum iron and transferrin levels are low, and iron saturation is frequently as low as 15%.[1,2] The serum ferritin level is usually normal or elevated.[2,3] All these changes can be induced by the inflammatory cytokines (e.g., interleukin-1 [IL-1]; tumor necrosis factor–α; interferons alfa, beta, and gamma; and perhaps transforming growth factor–β).[4] Under experimental conditions, these cytokines reduce erythropoietin production, cause hypoferremia, increase serum ferritin levels, impair erythropoiesis, and block release of iron from reticuloendothelial cells.[5] Hepcidin, a newly described mediator of iron metabolism, may be the major mediator of the anemia of chronic disease[6]; hepcidin production is increased up to 100-fold with inflammation. Hepcidin seems to be the long-sought mediator that transmits iron stores to the gut. Hepcidin is secreted when iron stores, primarily in the liver, are increased, and it blocks iron absorption from the gut and causes iron to be trapped in macrophages.[6]

Diagnosis

Mild anemia, with normal or elevated levels of leukocytes and platelets, in a patient with a chronic illness suggests the diagnosis of anemia of chronic disease. This normocytic or hypochromic and microcytic anemia is easily misdiagnosed as iron deficiency anemia, thalassemia trait, or a sideroblastic anemia. If the diagnosis is uncertain after careful examination of the blood smear, the most useful tests for making the diagnosis are measurement of the serum ferritin level and, in rare cases, bone marrow examination that includes an iron stain [*see Table 1 and Figure 1*]. In some cases, there is more than one cause of the anemia, and thorough examination of the patient may be required to establish the primary cause. For example, a patient who has anemia of chronic disease resulting from carcinoma of the colon may also be iron deficient because of intestinal bleeding. HIV infection produces complex hematologic effects, including Coombs-positive autoimmune hemolytic anemia, but it also

Table 1 Differential Diagnosis of Hypochromic Anemias

	Anemia of Chronic Disorders	*Iron Deficiency Anemia*	*Thalassemia Trait*	*Sideroblastic Anemias*
Smear	Usually normochromic, normocytic but can be mildly hypochromic, microcytic	Varies with the degree of the anemia [*see 5:II Red Blood Cell Function and Disorders of Iron Metabolism*]	Hypochromia, target cells, microcytes, basophilic stippling	May be similar to that of the thalassemia trait
Serum iron level	Low	Low	Normal	High
Iron-binding capacity	Low	High	Normal	Normal
Percent saturation	5–16	0–16	Normal (20–40)	60–90
Serum ferritin level	Normal or high	Low	Normal	High
Marrow iron in reticuloendothelial cells	+ + to + + +	0	+ + to + + +	+ + + +
Marrow iron in sideroblasts	0	0	+ + to + + +	+ + + + with ringed sideroblasts
Marrow erythroid precursors	Normal	Generally normal; cytoplasm may be scanty	Usually mild erythroid hyperplasia	Intense erythroid hyperplasia with dyserythropoiesis

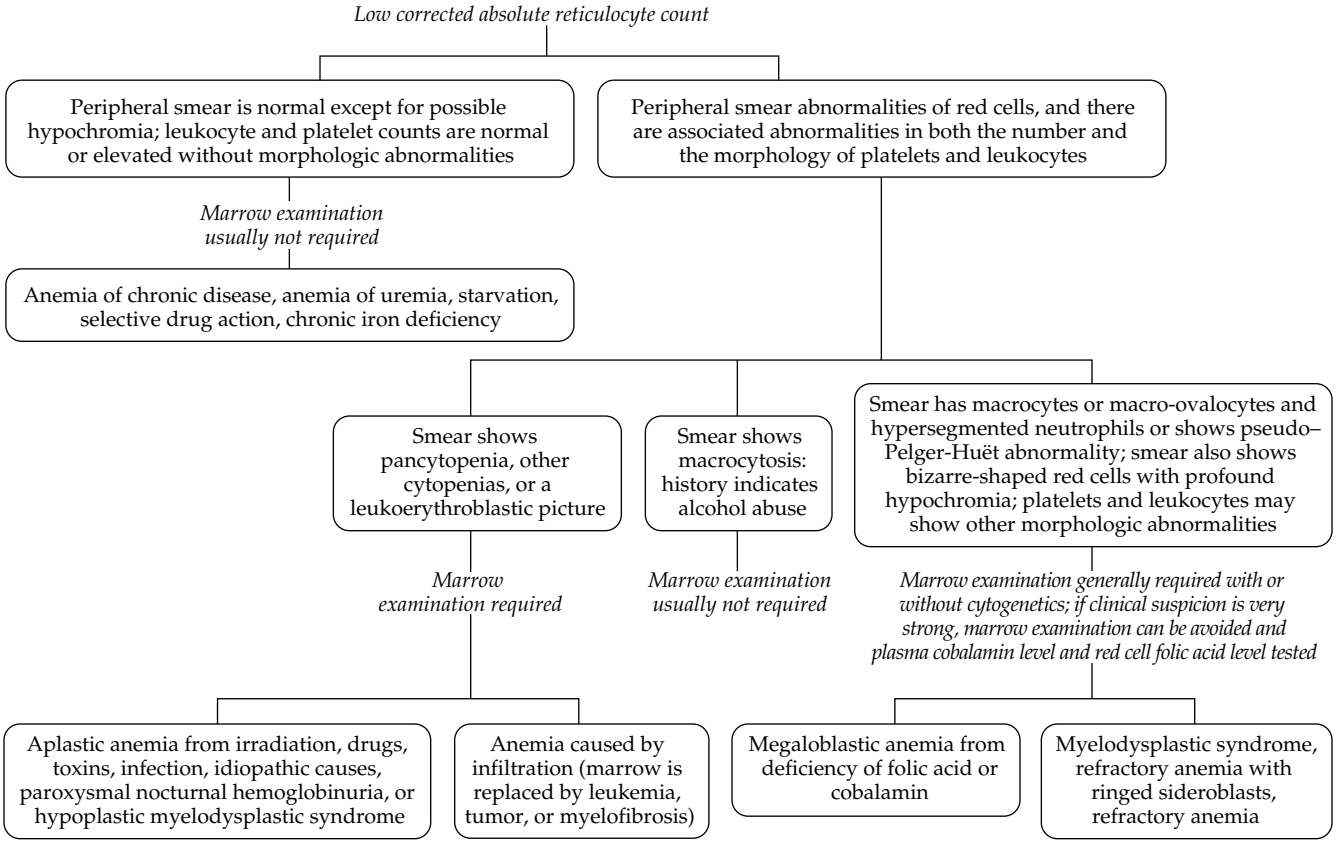


Figure 1 **Flowchart shows steps in the diagnosis of anemia caused by production defects. This type of anemia is suggested by a low corrected reticulocyte count or the finding of associated leukocyte or platelet abnormalities on the peripheral blood smear.**

causes anemia of chronic disease in the majority of patients with AIDS.[7]

Treatment

Identifying and treating the primary disease is the most important part of managing the anemia of chronic disease. Oral or parenteral iron administration is usually not helpful. Erythropoietin is the standard treatment for patients with anemia of chronic disease. For many patients, administration of pharmacologic doses of erythropoietin corrects the anemia of chronic disease by overriding the defect in erythropoietin production. It is useful to obtain a baseline measurement of the plasma erythropoietin level, because a response to erythropoietin is unlikely in patients whose endogenous levels are above 500 mU/ml. Erythropoietin responses have been reported in patients with rheumatoid arthritis,[8] AIDS,[9] inflammatory bowel disease,[10,11] and cancer.[12] To respond optimally, the patient must have adequate available iron stores (i.e., normal or elevated ferritin level or marrow iron stain) [*see 5:I Approach to Hematologic Disorders*]. Previously, the recommendation was to start the patient on 100 to 150 U/kg subcutaneously three times weekly; however, most physicians give a single subcutaneous dose of 40,000 units of erythropoietin weekly.[13]

If the hemoglobin level does not rise after 12 weeks, erythropoietin should be discontinued. A longer-acting form of erythropoietin, darbepoietin alfa, can be given subcutaneously at doses of 100 μg weekly or 200 μg every other week.

ANEMIA IN SEVERE RENAL DISEASE

Pathophysiology and Etiology

The predominant cause of anemia in renal disease is a deficiency of erythropoietin production by the diseased kidneys. If underlying inflammatory renal disease is present, there may be a component of anemia of chronic disease.[14] Anorexia and poor iron intake, frequent blood sampling, and loss of erythrocytes during hemodialysis may produce iron deficiency. Folic acid deficiency, hypersplenism, and secondary hyperparathyroidism with marrow fibrosis[4] may also promote anemia.

Anemia in hemodialysis patients can be caused by aluminum toxicity, as well. This anemia was initially identified in patients who had so-called dialysis dementia. Very high plasma aluminum levels probably result from aluminum contamination of the dialysis fluid or gastrointestinal absorption of the aluminum gels taken to bind dietary phosphates. In vitro experiments have shown that aluminum inhibits the growth of the erythroid precursors colony-forming unit–erythroid (CFU-E) and burst-forming unit–erythroid (BFU-E).[15]

Diagnosis

The blood smear should be examined for erythrocyte fragmentation or echinocytosis to exclude other causes of the anemia. The presence of Heinz bodies suggests that oxidative hemolysis has occurred, perhaps caused by oxidants in the hemodialysis fluid.

Treatment

Erythropoietin is the standard treatment for anemic patients with renal disease. Erythropoietin therapy can eliminate the transfusion requirement for patients on hemodialysis and in patients with progressive renal disease who do not yet require hemodialysis. Such treatment significantly improves their quality of life.[16] Side effects, such as hyperkalemia and hypertension, occur infrequently. It is customary to start therapy with 50 U/kg of erythropoietin three times weekly, either intravenously or subcutaneously, and to increase the dosage as necessary to bring the hemoglobin level to the desired value. Parenteral iron supplementation improves the response. ImFed (a form of iron dextran) can be given intramuscularly or intravenously at doses ranging from 100 to 500 mg, with an anticipated frequency of reaction of 4.7%. Ferrlecit (a form of sodium ferric gluconate) can be infused intravenously (125 mg over 1 hour), with the occasional occurrence of hypotension and rash.[12] In a study of patients with anemia caused by aluminum toxicity, treatment with I.V. deferoxamine (30 mg/kg I.V. at the end of each dialysis session) produced substantial improvement.[17]

ANEMIA SECONDARY TO OTHER CONDITIONS

Alcohol Abuse

Excessive alcohol ingestion—either acute or chronic—has profound hematologic effects.[18] Ingestion of about 80 g of alcohol (one bottle of wine, six pints of beer, or one-third bottle of whiskey) daily may produce macrocytosis,[19] stomatocytosis,[20] thrombocytopenia,[21] vacuolization of proerythroblasts, ringed sideroblasts,[20] a sharp drop in serum folic acid levels, and a rise in serum iron levels; it may also impair the reticulocyte response to administered folic acid in a patient known to be folic acid deficient. Acute alcohol ingestion itself does not produce a megaloblastic anemia.[18] It has been postulated that alcohol-induced hematologic toxicity is mediated through acetaldehyde, the major metabolite of ethanol, which is far more toxic and reactive than ethanol. The mechanism for these alcohol-induced abnormalities may be the formation of antibodies against acetaldehyde–hemoglobin adducts.[20] Megaloblasts, macro-ovalocytes, and hypersegmented polymorphonuclear neutrophils (PMNs) usually appear when concomitant folic acid deficiency is present. Chronic alcohol abuse often results in concomitant folic acid or iron deficiency, severe liver disease, GI bleeding, hypersplenism, and the anemia of chronic disease.

Starvation

Starvation resulting from anorexia nervosa or protein deficiency can cause anemia and even pancytopenia. Hemolysis may also be present [*see Figure 2*]. The bone marrow biopsy is hypocellular, with a characteristic gelatinous background material consisting of acid mucopolysaccharides. The anemia can occur despite normal folic acid and cobalamin (vitamin B_{12}) levels and can be corrected with proper nutrition.

Hypothyroidism

Hypothyroidism impairs erythrocyte production. The presence of macrocytosis in a hypothyroid patient suggests concomitant dietary folic acid deficiency or pernicious anemia.

Panhypopituitarism

The mild anemia that is associated with severe panhypopituitarism can be corrected by replacement of adrenal, thyroid, and gonadal hormones; the enhancing effect of androgens on the action of erythropoietin is well known.

Aging

The hemoglobin levels, red blood cell indices, and leukocyte and platelet counts of healthy older people are similar to those of younger adults; this finding was confirmed in a study of patients who were 84 years of age or older.[22] Thus, a workup is required when anemia occurs in such older patients. The evaluation and treatment of anemia in the aged has become increasingly important because the presence of anemia (hemoglobin concentration [Hgb] < 12 g/dl in women and < 13 g/dl in men) is an independent risk factor for decline in quality of life.[23]

Production Defects Associated with Marrow Aplasia or Replacement

The combination of anemia and neutropenia or thrombocy-

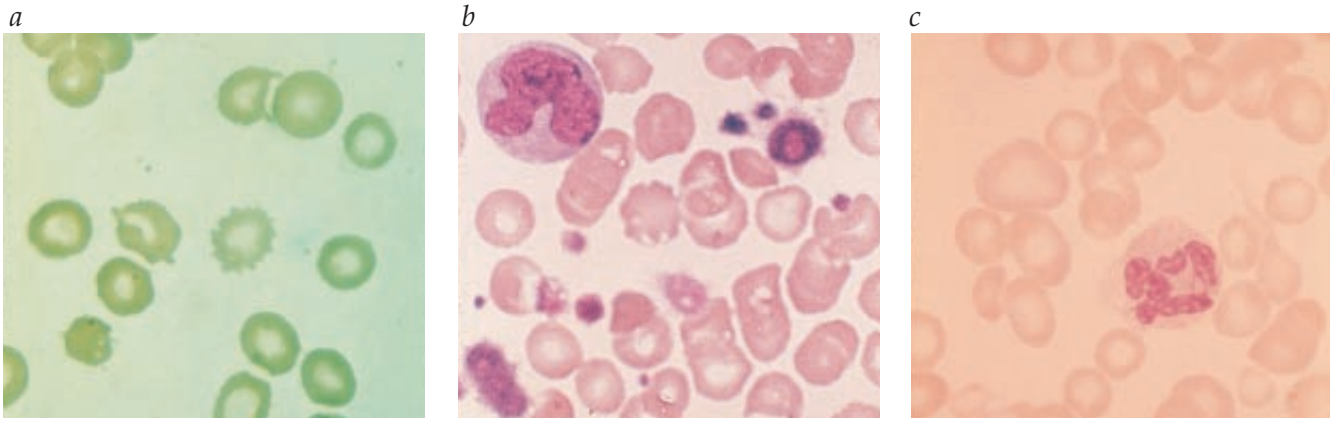


Figure 2 **The peripheral smear changes seen in severe liver disease or starvation (*a*) include distinct variation in size and shape of red blood cells; both sharply spiculed cells (spur cells) and scalloped erythrocytes are prominent. The leukoerythroblastic blood smear (*b*) indicates marrow replacement with extramedullary hematopoiesis. It is characterized by variation in the size and shape of red blood cells, by the presence of nucleated red blood cells in the peripheral blood, by giant platelets, and by immaturity in the myeloid series. In folic acid or cobalamin deficiency (*c*), the smear is characterized by variation in erythrocyte size and by distinct macrocytosis. Occasionally, fish-tailed erythrocytes are present, along with hypersegmented neutrophils.**

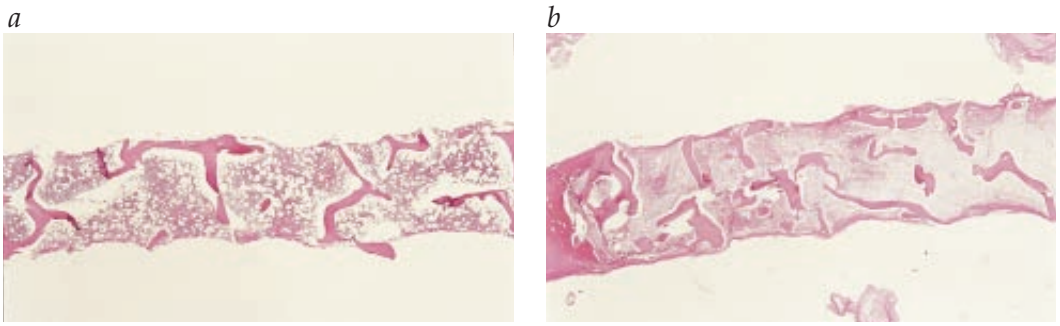


Figure 3 **Shown are (*a*) biopsy of normal bone marrow and (*b*) biopsy of bone marrow from a patient with aplastic anemia showing almost complete aplasia.**

topenia or the combination of all three of these abnormalities (i.e., pancytopenia) usually indicates that the hematopoietic marrow is damaged. If the marrow cavity is infiltrated but pluripotent stem cells are intact, extramedullary hematopoiesis will often develop in the organs of fetal hematopoiesis (i.e., spleen, liver, and distal bones).

Pancytopenia can be congenital or acquired. The finding of combined cytopenias or of immature cells in the blood (myelocytes, metamyelocytes, and erythroblasts)—that is, a leukoerythroblastic blood smear—suggests extramedullary hematopoiesis [*see Figure 2*]. These findings are an indication for bone marrow aspiration and biopsy.

APLASTIC ANEMIA

Definition

Pancytopenia (i.e., anemia, neutropenia, and thrombocytopenia) and aplastic marrow on biopsy examination [*see Figure 3*] establish a working diagnosis of aplastic anemia. The biopsy specimen must not be taken from a marrow site that has been irradiated. It is essential to determine the severity of aplastic anemia. Severe aplastic anemia (SAA) is defined by (1) marrow of less than 25% normal cellularity or marrow of less than 50% normal cellularity in which fewer than 30% of the cells are hematopoietic, and (2) two out of three abnormal peripheral blood values (absolute reticulocyte count < 40,000/μl, absolute neutrophil count [ANC] < 500 μl, or platelet level < 20,000/μl). These criteria have been criticized as being relatively insensitive. Some investigators prefer to identify a cohort of patients with very severe aplastic anemia (VSAA) as those who had an ANC less than 200/μl.[24]

Etiology

Aplastic anemia has a number of causes [*see Table 2*], although in many cases the exact cause cannot be determined.

Ionizing irradiation and chemotherapeutic drugs used in the management of malignant and immunologic disorders have the capacity to destroy hematopoietic stem cells. With careful dosing and scheduling, recovery is expected. Certain drugs, such as chloramphenicol, produce marrow aplasia that is not dose dependent. Gold therapy and the inhalation of organic solvent vapors (e.g., benzene or glue) can also cause fatal marrow failure.

In 2% to 10% of hepatitis patients, severe aplasia occurs 2 to 3 months after a seemingly typical case of acute disease, usually in young men. Often, the hepatitis has no obvious cause, and tests for hepatitis A, B, and C are negative.[25] There is a high incidence of aplastic anemia after liver transplantation in patients with severe non-A, non-B hepatitis.[26]

Several lines of evidence support the possibility that immune disorders can lead to aplasia. Marrow aplasia occurs in graft versus host disease (GVHD).[27] Immunosuppressive preconditioning improves the chances of successful transplantation of syngeneic marrow into patients with aplastic anemia,[28] and immunosuppressive therapy has been used successfully to treat idiopathic aplastic anemia.[27,28] The blood of some patients with aplastic anemia appears to contain suppressor T cells that suppress the growth of the committed progenitor cells known as colony-forming unit–granulocyte-macrophage (CFU-GM). The suppressor T cells may act by producing interferon gamma.[28] The result of these complex immune mechanisms involving suppressor T cells is a profound decrease in primitive hematopoietic cells as measured by both the long-term culture–initiating cell (LTC-IC) assay and the ability to form secondary colonies from the colonies surviving 5 weeks of marrow culture.[29]

Aplasia can also be part of a prodrome to hairy-cell leukemia [*see 12:XV Chronic Lymphoid Leukemias and Plasma Cell Disorders*],

Table 2 Causes of Aplastic Anemia

IRRADIATION	
DRUGS	
	Agents whose use regularly causes myelosuppression
	Alkylating agents: melphalan, cyclophosphamide, chlorambucil, busulfan
	Antimetabolites: azathioprine, 6-mercaptopurine, hydroxyurea, methotrexate
	Other antitumor agents: daunorubicin, doxorubicin, carmustine, lomustine, amsacrine
	Agents whose use occasionally causes myelosuppression
	Chloramphenicol, gold compounds, arsenic, sulfonamides, mephenytoin, trimethadione, phenylbutazone, quinacrine, indomethacin, diclofenac, felbamate
TOXINS	Benzene, glue vapors
INFECTIONS	Non-A, non-B, non-C hepatitis, infectious mononucleosis, parvovirus infection (attacks erythroid precursors), HIV
MALIGNANT DISEASES	Hairy-cell leukemia, acute lymphocytic leukemia, acute myeloid leukemia (rarely), myelodysplastic syndromes
CLONAL DISORDERS	Paroxysmal nocturnal hemoglobinuria
IMMUNE-MEDIATED APLASIA	Eosinophilic fasciitis
INHERITED DISORDERS	Fanconi anemia
PREGNANCY	

acute lymphoblastic leukemia [*see 12:XVI Acute Leukemia*], or, in rare cases, acute myeloid leukemia; or it can develop in the course of myelodysplasia [*see 12:XVI Acute Leukemia*].

Diagnosis

The patient with aplastic anemia may seek medical attention because of fatigue and shortness of breath. Accompanying thrombocytopenia may cause petechiae, oral blood blisters, gingival bleeding, and hematuria depending on the level of the platelet count. By far the major problem associated with aplastic anemia is the recurrent bacterial infections caused by the profound neutropenia. Sepsis, pneumonia, and urinary tract infections are common among patients with aplastic anemia. Invasive fungal infections may cause death, especially in patients with severe neutropenia.

The diagnosis of aplastic anemia requires a marrow aspirate and biopsy [*see Figure 3*], as well as a thorough history of drug exposures, infections, and especially symptoms suggesting viral illnesses and serologic test results for hepatitis, infectious mononucleosis, HIV, and parvovirus [*see Figure 4*]. Measurement of red cell CD59 is helpful in the diagnosis of paroxysmal nocturnal hemoglobinuria.

It is also important to determine the severity of aplastic anemia [*see* Aplastic Anemia, Definition, *above*]. Severe cases are associated with a very low rate of spontaneous remission and a mortality of 70% within 1 year. In contrast, 80% of patients who have milder forms of aplastic anemia survive for 1 year.[24]

Differential Diagnosis

The differential diagnosis of pancytopenia includes chronic lymphocytic leukemia, systemic lupus erythematosus, and congestive splenomegaly. In these diseases, however, the marrow is not aplastic but rather shows hyperplasia of the involved cell lines. Other conditions that cause pancytopenia include hypoplastic myelodysplastic syndrome, acute leukemia, megaloblastosis, and large granular lymphocytic leukemia.[30]

Treatment of Mild Aplastic Anemia

Treatment of milder forms of aplastic anemia involves removing the offending agent and providing supportive therapy, primarily transfusion therapy, anticipating that the remaining pluripotent stem cells will repopulate the marrow.

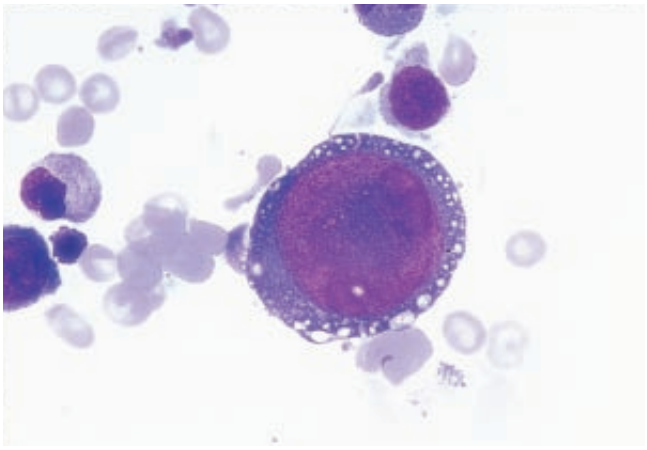


Figure 4 **Giant pronormoblast, evident on this marrow smear, strongly suggests a diagnosis of parvovirus infection.**

Supportive therapy Thrombocytopenia is often a major problem associated with aplastic anemia. It should be managed by platelet transfusion as needed to control or prevent bleeding. Usually, a threshold of 10,000 platelets/μl is used for transfusion, but conservative treatment is best, and as few transfusions as possible are given. Extensive platelet replacement may result in allosensitization to platelets and may complicate future allogeneic bone marrow transplantation. Red blood cell transfusions are given as required to control the symptoms and signs of anemia.

Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) have been given to patients to raise the absolute neutrophil count and help combat infection. They are usually ineffective when used alone, because of the severe deficiency in precursor cells, which are the target for the actions of G-CSF and GM-CSF.[31] It is generally preferable to proceed to definitive treatment: immunosuppressive therapy or preferably allogeneic bone marrow transplantation if a matched sibling donor is available [*see 5:XI Hematopoietic Cell Transplantation*].[32]

Definitive therapy Transplantation from a matched sibling after a preparative regimen of high-dose cyclophosphamide and antithymocyte globulin, together with the use of methotrexate and cyclosporine for GVHD prophylaxis, is a very effective regimen for patients with aplastic anemia. Current results suggest a cure rate greater than 90%.[33] Results with mismatched or unrelated matched donors are somewhat worse; therefore, patients with aplastic anemia who are without sibling donors are often given a trial of immunosuppressive therapy before transplantation.

Three forms of immunosuppression have been shown to produce partial remission in aplastic anemia.[31,32,34] Antithymocyte globulin (ATG) produced sustained remission in about half of the patients in a randomized trial.[32] High-dose corticosteroids improved blood counts in about 40% of treated patients, and cyclosporine was also shown to be beneficial.[32] (Androgens such as oxymetholone may have a role in the treatment of severe aplastic anemia but are not given alone.[31,34])

Although each of these agents can be used individually or consecutively in the treatment of aplastic anemia, a controlled study suggests that results are better when all three are used simultaneously.[31,32] The combination of ATG, a corticosteroid, and cyclosporine resulted in an actuarial survival of 62% at 36 months. The first signs of response occurred at about 4 weeks; the median time to remission was 60 to 82 days.[32] In this study, patient outcome was related to the quality of hematologic response. An 11-year follow-up report confirmed the effectiveness of the combination of ATG, corticosteroids, and cyclosporine. The relapse rate was 38%, and clonal or malignant diseases developed in 25% of patients.[35]

One recommendation, based on the usual availability of horse ATG in the United States,[31,32] is to administer horse ATG at a dosage of 40 mg/kg/day in 500 ml of saline for 4 days over a period of 4 to 5 hours through an I.V. line equipped with a microaggregate in-line filter. The toxic side effect of ATG is serum sickness, which can usually be controlled with corticosteroids. Prednisone (60 to 100 mg/day) is given orally in divided doses, or methylprednisolone (40 mg) is added to the infusion bottle, and the dose can be increased to 1 mg/kg/day. Corticosteroid therapy is adjusted to control serum sickness, but it can usually be tapered after 2 weeks and stopped after 30 days. Because ATG can lower platelet counts, platelet transfusions are given as

needed to maintain the platelet count at more than 20,000 μl.

Cyclosporine (10 to 12 mg/kg/day) is given orally in two divided doses, with the aim of achieving whole blood trough levels of 500 to 800 ng/ml or a serum level of 100 to 200 ng/ml. After 29 days, the cyclosporine dosage can be tapered for a trough whole blood level of 200 to 500 ng/ml.[31,32] The cyclosporine is continued for at least 6 months. Cyclosporine can cause hypertension, renal toxicity, hypomagnesemia, vitiligo, tremors, hypertrichosis, susceptibility to *Pneumocystis carinii* pneumonia (PCP), and gingival hyperplasia.[31,32] In one study, 300 mg of aerosolized pentamidine was given every 4 weeks as PCP prophylaxis.[32]

In another study, G-CSF (5 μg/kg/day) was given subcutaneously for the first 90 days, along with I.V. methylprednisolone (2 mg/kg/day on days 1 through 5, followed by 1 mg/kg/day on days 6 through 10, and tapered off in 30 days), with good results.[36]

In contrast to patients who undergo allogeneic bone marrow transplantation, patients who respond to immunosuppressive therapy are not actually cured. Many of these patients continue to have moderate cytopenia[37]; 20% to 36% experience relapses of aplastic anemia,[31,32,37] and as many as 20% to 36% eventually develop clonal disorders, such as paroxysmal nocturnal hemoglobinuria, myelodysplastic syndrome, and acute leukemia.[31,32] Patients also are at increased risk for the development of solid tumors after treatment of aplastic anemia, but the risk is the same for patients who underwent immunosuppressive therapy as it is for those who underwent allogeneic bone marrow transplantation.[38] More than 50% of patients who have relapses of aplastic anemia after initially responding to immunosuppressive therapy may respond to a second course of therapy. [31,32] For unresponsive patients, a trial of rabbit ATG may work. The rabbit ATG (3.5 mg/kg/day diluted in saline and infused over 6 to 8 hours for 5 consecutive days)[39] is given along with cyclosporine (5 mg/kg/day p.o. on days 1 through 180, then tapered) and G-CSF (5 μg/kg/day on days 1 through 90).

An intriguing report concerns 10 patients with severe aplastic anemia who were treated with high-dose I.V. cyclophosphamide (45 mg/kg/day) for 4 consecutive days.[40] Some patients also received cyclosporine. Only one course of I.V. cyclophosphamide was given. Seven of 10 patients had a complete hematologic response, and six were still alive after a median follow-up of 10.8 years (range, 7.3 to 17.8 years). However, a trial comparing high-dose cyclophosphamide with ATG was ended early because of excessive cyclophosphamide-induced morbidity and mortality.[41] Therefore, the role of high-dose cyclophosphamide in the treatment of aplastic anemia needs extensive clarification.

Treatment of Severe Aplastic Anemia

The choice of appropriate therapy for patients with SAA is influenced by age and disease severity. The European Group for Blood and Marrow Transplantation reported on the results of immunosuppressive therapy in 810 patients subdivided into three age groups: younger than 49, 50 through 59, and older than 60. The 5-year survival rates for those with SAA were 86%, 72%, and 54%, respectively; for those with VSAA, the comparable rates were 49%, 40%, and 21%.[42] Older patients had more bleeding and infections.

Patients younger than 20 years Allogeneic bone marrow transplantation should be performed in patients younger than 20 years if a matched sibling donor is available. Although there are risks, including chronic GVHD and organ dysfunction caused by the conditioning program,[31] 50% to 80% of patients may be cured; the incidence of later clonal disorders is very low.[34] Allogeneic bone marrow transplantation, along with conditioning programs consisting of cyclophosphamide and ATG, produced an actuarial survival rate of 69% after 15 years.[34] Patients younger than 20 years who do not have a matched sibling donor should consider transplantation from a matched unrelated donor. Allogeneic transplantation from a matched unrelated donor initially produced a 2-year survival rate of only 29% because of severe GVHD.[31] In a study of 15 patients who received unrelated-donor transplantations, all were reported alive at 2 to 86 months (mean follow-up, 51 months); only one patient developed extensive GVHD, and five (33%) developed moderate to acute GVHD. These results suggest that conditioning regimens that contain ATG and cyclophosphamide are improving the treatment outcomes for unrelated donor transplantation in this patient group.[43]

Patients between 20 and 45 years of age Patients between 20 and 45 years of age who are in excellent health and have a fully matched sibling donor may be able to tolerate GVHD and thus benefit from the curative potential of an allogeneic bone marrow transplant. Some experts propose that allogeneic bone marrow transplantation should be considered for patients in this age group,[34] particularly because newer conditioning programs seem to be capable of reducing the severity of GVHD.[31,44] In a study of 154 patients younger than 46 years who received allogeneic transplantation, the median survival was 29 months, and the probability of overall survival at 5 years was 56%.[45]

Patients older than 45 years Previously it was thought that the impact of GVHD was too severe for patients older than 45 years, and it was suggested that these patients receive immunosuppressive therapy.[31,34] However, conditioning programs containing ATG and cyclophosphamide seem to be more tolerable, and even heavily pretreated patients as old as 59 years have done well after allogeneic marrow transplantation.[46]

ACQUIRED PURE RED CELL APLASIA

Definition

In adults, pure red cell aplasia (PRCA) is an acquired disorder. The anemia is severe (hematocrit usually less than 20%), reticulocytopenia is profound (often 0%), the absolute reticulocyte count is usually less than 10,000/μl, and marrow erythroid precursors are virtually absent. Marrow myeloid and megakaryocytic elements are preserved, however, and the peripheral platelet and white blood cell counts are also normal.

Pathophysiology

In PRCA, erythropoiesis is thought to be inhibited primarily by immune mechanisms, including autoantibody-mediated and T cell–mediated suppression of erythroid progenitors, usually at a stage after the CFU-E stage of erythroid differentiation and before formation of proerythroblasts. T cells, particularly of the large granular lymphocyte (T-LGL) class, may be involved in the suppression of erythropoiesis, and in some cases, there is evidence that the suppression is caused by clonal T cells.[47] Autoantibody inhibition of erythropoietin has also been described, but it is quite uncommon.[48] Two other mechanisms probably cause

Table 3 Causes of Acquired Pure Red Cell Aplasia

Primary

- Associated with thymoma in 10%–15% of cases[51]
- Idiopathic causes

Secondary

- Neoplasia: chronic lymphocytic leukemia, chronic myeloid leukemia, Hodgkin and non-Hodgkin lymphomas; large granular lymphocytic proliferative disorders; prodrome to myelodysplastic syndromes[51]
- Systemic lupus erythematosus or rheumatoid arthritis
- Associated with pregnancy
- Associated with autoimmune hemolytic anemia
- Drugs: those most commonly associated are phenytoin, chlorpropamide, zidovudine,[57] trimethoprim-sulfamethoxazole, isoniazid[51]
- Multiple endocrine gland insufficiency
- Primary amyloidosis
- Infections: infectious mononucleosis, viral hepatitis, parvovirus infection, HIV[51]
- ABO-incompatible bone marrow transplantation

PRCA: (1) a specific attack on erythroid precursors by the parvovirus B19 (one report indicated that 14% of cases were caused by this virus[49]) and (2) an underlying hematopoietic clonal abnormality that may be a prodrome to myelodysplastic syndrome.[48]

Etiology

PRCA may be caused by a variety of processes, including neoplasia, autoimmune disorders, drugs, and infections [*see Table 3*].

The association of PRCA with LGL proliferation and leukemia is increasingly being recognized.[30] The routine use of T cell receptor gene rearrangement studies in one series showed that nine of 14 patients had a clonal LGL disorder.[50] Presumably, these LGL cells directly mediate inhibition of erythropoiesis.[49,50] In perhaps as many as 20% of cases, PRCA may be a prodrome to the myelodysplastic syndromes or acute myeloid leukemia.[49,51]

Erythroblastopenia also occurs in a small percentage of patients with autoimmune hemolytic anemia [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*] and may be caused by autoantibody attack on maturing normoblasts.

The treatment of HIV infection with zidovudine (AZT) produces, in virtually all patients, an anemia that is usually marked by significant macrocytosis.[52] Moderate erythroid hypoplasia is the usual cause of this anemia, which can progress to PRCA.

Parvovirus infection is the cause of the transient aplastic crises that occur in patients who have severe hemolytic disorders. The marrow in patients with such disorders must compensate for the peripheral hemolysis by increasing its production up to sevenfold and thus typically shows an intense erythroid hyperplasia. Although parvovirus can affect all precursor cells, the red cell precursors are the most profoundly affected.[49]

PRCA can complicate ABO-incompatible allogeneic bone marrow transplantation; the recipient's serum continues to express anti-A or anti-B isohemagglutinins against donor A or B antigen expressed on the surface of erythroid progenitors.[51] With PRCA of pregnancy, antibodies against BFU-E usually disappear after delivery, coinciding with clinical remission.[53]

Diagnosis

The patient with PRCA presents with symptoms characteristic of anemia—namely, weakness, fatigue, and shortness of breath. White blood cell and platelet counts are normal morphologically and functionally. A very low reticulocyte count—either a relative reticulocyte value of less than 0.2% or a very low absolute reticulocyte count of less than 10,000 μl—should prompt the physician to order a bone marrow aspirate. In a patient with PRCA, a bone marrow aspirate and biopsy typically show normal myelopoiesis, lymphopoiesis, and megakaryocytopoiesis; erythropoiesis is virtually absent. In the absence of any apparent cause of PRCA, four conditions must be considered: idiopathic PRCA, thymoma, hypoplastic myelodysplastic syndromes (MDS), and LGL leukemia. The workup to diagnose PRCA usually includes computed tomography of the chest to evaluate the possibility of thymoma, immunophenotypic analysis of circulating blood or marrow lymphocytes to identify LGL proliferation, marrow cytogenetics to evaluate the possibility of MDS, and antibody tests for parvovirus.[49] A diagnostic hallmark of parvovirus infection is the appearance of giant pronormoblasts in the marrow [*see Figure 4*]. The distinction between PRCA associated with the myelodysplastic syndromes and acute myeloid leukemia may be difficult to determine at the time of diagnosis unless a typical myelodysplastic cytogenetic abnormality is detected during a bone marrow examination.

Treatment

Two general principles of management in PRCA are transfusions for symptomatic anemia and cessation of possible offending drugs. No specific therapy is indicated in those forms of PRCA that are self-limited, such as pregnancy, ABO-incompatible bone marrow transplantation, and some cases of parvovirus infection.[51,53] Treatment of PRCA depends on the identified cause. If a thymoma is present, it should be removed surgically[52]; this procedure leads to patient improvement in about one third of such cases.[51] When surgery is impossible, one should consider a course of prednisone combined with octreotide, a somatostatin analogue that binds to thymomas and may inhibit the function of thymic immune cells.[54]

Treatment of other causes of PRCA is based on the supposition that the attack is immune mediated and therefore will respond to immunosuppressive therapy. Treatment can begin with the administration of 60 mg of oral prednisone daily in divided doses; this regimen should be continued for 1 to 3 months.[49] If a patient fails to respond, as indicated by a rise in the reticulocyte count, cyclophosphamide or azathioprine should be added at a dosage of 2 to 3 mg/kg/day orally. Patients with marrow cytogenetic abnormalities suggestive of myelodysplastic syndrome respond poorly.[49,50] Some patients who are refractory to other forms of therapy have responded well to I.V. IgG (0.4 g/kg/day for 5 days).[55]

Patients with LGL proliferation as the underlying cause respond well to cyclophosphamide.[50,56] Usually, low doses of cyclophosphamide (50 to 100 mg/day p.o.) for 3 to 6 months suffice to produce remission, which is sometimes associated with disappearance of LGL proliferation.[50,57] Patients who respond poorly usually respond to oral cyclosporine.[50,57] Cyclosporine (12 mg/kg/ day) has been shown to produce responses of approximately 65%, even in patients who did not respond to corticosteroids, plasmapheresis, cyclophosphamide, or azathioprine therapy.[51,58]

For patients in whom parvovirus infection is the cause of PRCA, I.V. IgG works well; the standard dosage is 0.4 g/kg/day for 5 days.[49] For AIDS patients with parvovirus infection and PRCA, I.V. IgG may have to be continued.[57] Recovery

from the transient crises of parvovirus infection occurs spontaneously in 1 to 2 weeks after onset of the infection.

ATG therapy for patients with refractory PRCA is similar to that for patients with aplastic anemia (40 mg/kg/day I.V. for 4 days).[51] Other drugs that have been used in refactory cases are azathioprine (2 to 3 mg/kg/day), antilymphocyte globulin, and anti-CD20 monoclonal antibody.[31] In very refractory cases, allogeneic bone marrow transplantation can be effective.[59]

Production Defects with Marrow Erythroid Hyperplasia and Ineffective Erythropoiesis

DEFINITION

Anemia with a low reticulocyte count may occur despite intense marrow erythroid hyperplasia. This paradoxic situation is the hallmark of ineffective erythropoiesis or intramedullary hemolysis. Generalized erythroid impairment may be present, or specific subpopulations of erythroid precursors may be involved. Some of these subpopulations escape death in the marrow, but their progeny are so severely damaged that they are rapidly removed from the circulation, thus giving the picture of peripheral hemolysis. Other signs of ineffective erythropoiesis include jaundice, a very high serum lactic dehydrogenase level, and 75% to 90% saturation of serum iron-binding capacity. The classic ferrokinetic picture shows rapid plasma iron clearance, which indicates intense erythroid precursor activity. The delivery of labeled red blood cells to the peripheral circulation, however, is dramatically reduced, which suggests that the precursors are being destroyed by intramedullary hemolysis.

The differential diagnosis includes megaloblastic anemias, sideroblastic anemias, thalassemia [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*], myelodysplastic syndromes [*see 12:XVI Acute Leukemia*], and agnogenic myeloid metaplasia [*see 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*].

MEGALOBLASTIC ANEMIAS

Etiology

Megaloblastic anemias are caused by cobalamin or folic acid deficiency, by drugs that interfere with the synthesis of DNA or with the absorption or metabolism of cobalamin, and by genetic disorders that interfere with DNA metabolism or with the absorption or distribution of cobalamin.

Pathophysiology

Megaloblastic erythropoiesis is characterized by defective DNA synthesis and arrest at the G_2 phase, with impaired maturation and a buildup of cells that do not synthesize DNA and that contain anomalous DNA. This condition leads to asynchronous maturation between the nucleus and cytoplasm.[60] RNA production and protein synthesis continue; thus, larger cells, or megaloblasts, are produced. Ineffective erythropoiesis results, and there is disagreement about the presence of increased apoptosis.[61,62] It is presumed that similar defects in DNA synthesis characterize the mucosal abnormalities of the stomach and tongue. In the granulocytic line, the presence of giant metamyelocytes represents ineffective granulopoiesis.[60]

The role of folic acid and cobalamin The interactions between folic acid and cobalamin are critical in the metabolism of single carbon units, mainly methylene and formyl analogues, which have a key role in the synthesis of DNA and purines [*see Figure 5a*].[63] There are two major coenzymes of cobalamin, adenosylcobalamin and methylcobalamin. Adenosylcobalamin is the coenzyme for methylmalonyl–coenzyme A mutase, which catalyzes a step in the catabolism of propionic acid [*see Figure 5b*].[63] Methylcobalamin is the coenzyme for methionine synthase, which functions as a methyltransferase in the reaction that converts 5-methyltetrahydrofolate (CH_3-THF_1) to tetrahydrofolate (THF_1) [*see Figure 5a*].[63] Cobalamin and folic acid [*see Figure 6*] combine in the methionine synthase reaction [*see Figure 5a*], in which the methyl group of CH_3-THF_1 is transferred to cobalamin to form methylcobalamin. Methylcobalamin then transfers its methyl group to homocysteine to form methionine. The monoglutamated THF_1, which is formed by this reaction, is polyglutamated by the enzyme folylpolyglutamate synthase, and a methylene group is added to it by the serine-glycine methyltransferase to form 5,10-methylene THF_n. 5,10-Methylene THF_n provides its methylene to convert deoxyuridylate to thymidylate, a key step in DNA synthesis. 5,10-Methylene THF_n can also be directly converted to CH_3-THF_1 by the enzyme 5,10-methylene tetrahydrofolate reductase, thereby making its methyl group available.

Formyl THF_n (also called leucovorin, folinic acid, or citrovorum factor) has an important role in purine synthesis and DNA metabolism. It can be generated by oxidation of 5,10-methylene THF_n or directly from THF_n by the enzyme formyl THF synthase, with methionine providing the formate group [*see Figure 5a*].[63] When cobalamin is deficient, CH_3-THF_1 cannot transfer its methyl group to cobalamin; therefore, THF_1 is not free to be polyglutamated by folylpolyglutamate synthase [*see Figure 5a*]. The polyglutamated form is required for synthesis of either 5,10-methylene THF_n or formyl THF_n; thus, DNA synthesis and purine synthesis are blocked. This hypothesis, the methylfolate trap hypothesis, is supported by the finding of increased levels of CH_3-THF_1 in the plasma of cobalamin-deficient patients. An alternative explanation is the formate starvation hypothesis, wherein cobalamin deficiency impairs methionine generation, which therefore cannot provide the methyl groups needed by the enzyme formyl THF_n synthase to produce formyl THF_n.

Other aspects of folic acid and cobalamin metabolism Neither folic acid nor cobalamin is produced by humans in adequate amounts; both must be absorbed from food. Cobalamin, in particular, is derived from microbial sources and is ingested in the form of meat or eggs.

Most of the dietary folic acid is in the polyglutamate form and is absorbed at the intestinal mucosa. Absorption of radioactively labeled folic acid approaches 80% of a 200 μg dose.[60,63] The serum folic acid level appears to be maintained by folic acid absorbed from food. Enterohepatic circulation of folic acid has been observed in which folic acid passing into the bile and small intestine is quantitatively reabsorbed. In an animal model, ethanol administration blocks the entry of folic acid into the bile. This effect could account, in part, for the sharp fall in the serum folic acid level seen 8 hours after alcohol consumption. A similar fall in serum folic acid level follows phenytoin ingestion. The daily requirement of cobalamin is about 1 μg, and the amount usually provided by the Western diet, which is rich in animal products, is about 5 to 15 μg.[64]

R proteins are a class of cobalamin-binding glycoproteins found in saliva and gastric juice; they are produced by granulocytes and other tissues. Intrinsic factor (IF) is a 45 kd glycopro-

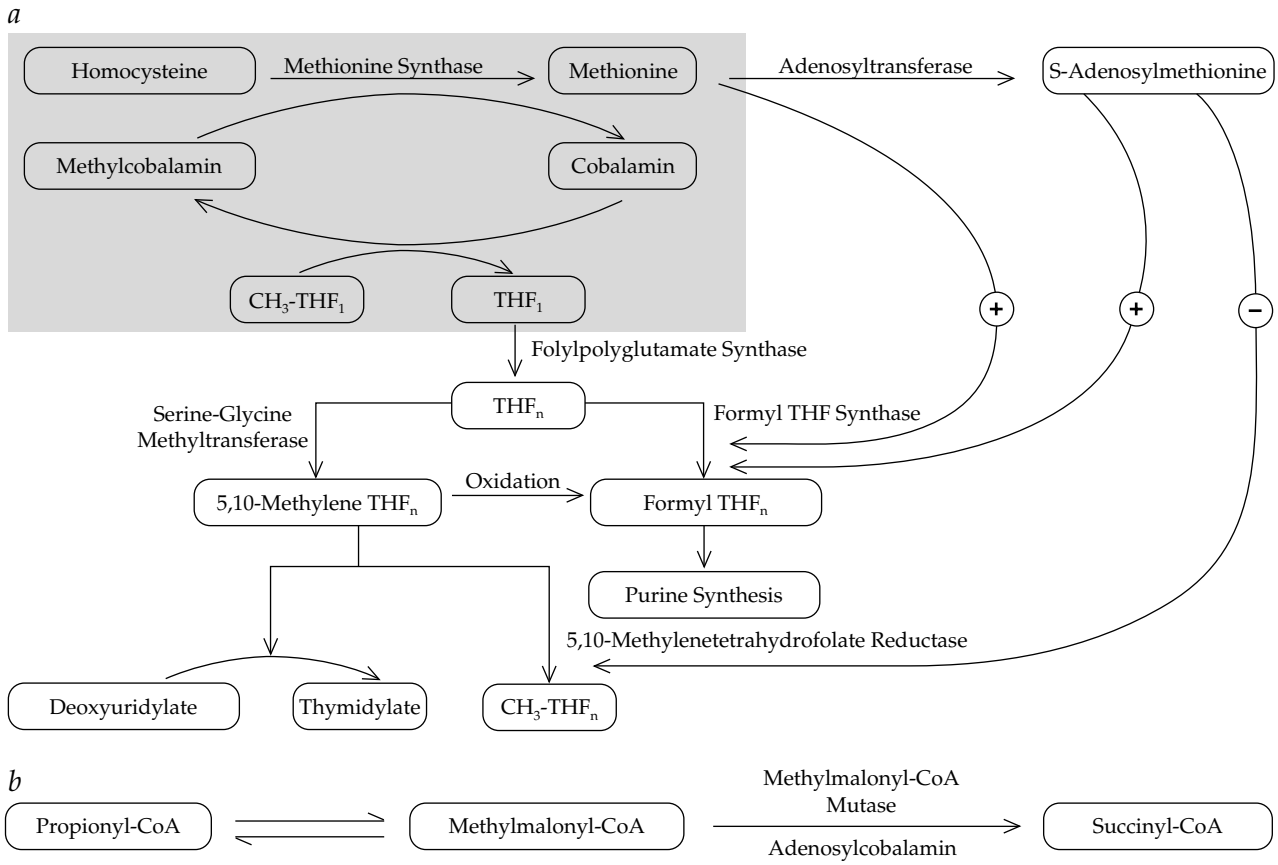


Figure 5 **(*a*) Intracellular interdependent cofactor activity of cobalamin and folic acid is essential in DNA synthesis and metabolism.[63] (*b*) Adeno-sylcobalamin is a cofactor in the synthesis of succinyl–coenzyme A from methylmalonyl–coenzyme A.[63] (CoA—coenzyme A)**

tein, secreted by gastric parietal cells, that is highly specific for unaltered cobalamin. The R protein–cobalamin complex does not bind to ileal receptors and thus is not absorbed. In the stomach, cobalamin binds preferentially to R proteins rather than to IF[60,63,64]; thus, it is the physiologically inactive R protein–cobalamin complex that is discharged into the duodenum. In the duodenum and small intestine, however, the pancreatic proteases along with pepsin degrade the R proteins, freeing cobalamin and allowing it to bind to IF. Thus, gastric atrophy and pancreatic insufficiency contribute to cobalamin malabsorption.[63,64] The IF-cobalamin complex, in the presence of Ca^{2+} and at a pH level greater than 5.4, binds specifically to a limited number of sites on the microvilli of mucosal cells in the terminal portion of the ileum, where absorption takes place [*see Figure 7*].[63,64]

In the plasma, most of the cobalamin is bound to the physiologically unimportant R proteins, transcobalamins I and III (TC-I and TC-III), which are about 70% saturated with cobalamin.[65] The physiologically important transport protein is transcobalamin II (TC-II), which has considerable specificity for cobalamin and is only 5% to 10% saturated with cobalamin. Receptors for the TC-II–cobalamin complex are present on many cell membranes. TC-II binds about 90% of a newly injected dose of cobalamin; and the complex is rapidly cleared, with a half-life of 6 to 9 minutes.[66,67] In persons with congenital TC-II deficiency, which results in severe megaloblastic anemia, both plasma cobalamin transport and cobalamin absorption are impaired. Impaired cobalamin absorption implies that TC-II has a role within the ileal enterocyte, where cobalamin is transferred from IF to TC-II.

The elevation of cobalamin levels seen in patients with chronic granulocytic leukemia or significant granulocytosis is caused by increases in TC-I and, to a lesser extent, TC-III, which are produced in granulocytes.

MEGALOBLASTIC ANEMIA CAUSED BY COBALAMIN DEFICIENCY (PERNICIOUS ANEMIA)

Pathophysiology

Cobalamin deficiency in pernicious anemia is thought to result from an autoimmune gastritis and an autoimmune attack on gastric intrinsic factor. There are two types of anti-IF antibodies: one of these antibodies blocks attachment of cobalamin to IF, and the other blocks attachment of the IF-cobalamin complex to ileal receptors.[66] Clinically, highly specific anti-IF antibodies are found in about 70% of patients with pernicious anemia. A second component of pernicious anemia is chronic atrophic gastritis that leads to a decline in IF production. The chronic atrophic gastritis in pernicious anemia is also associated with an increased risk of intestinal-type gastric cancer and of gastric carcinoid tumors.[67] Pernicious anemia occurs in association with other autoimmune disorders. In one study, autoimmune thyroid disorders were observed in 24% of 162 patients with pernicious anemia.[68]

Diagnosis

Clinical manifestations In addition to macrocytic and megaloblastic anemia, the patient with cobalamin deficiency

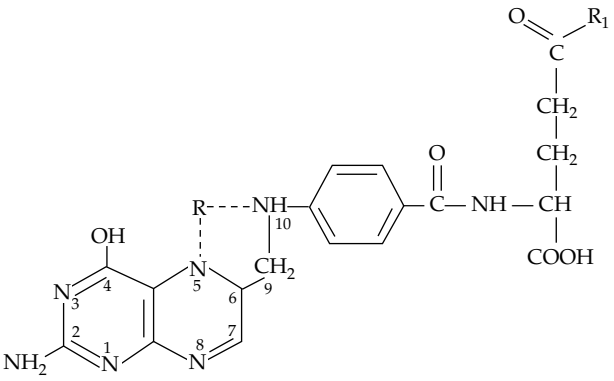


Figure 6 **Folic acid functions as a coenzyme in single-carbon transfer reactions. It is not physiologically active until it is reduced at positions 5, 6, 7, and 8 to tetrahydrofolate (THF). Single-carbon groups (R) such as methyl analogues and formate are added at either position 5 or position 10, or they may bridge from 5 to 10, as shown. There may be several glutamates attached in sequence (R_1), which convert the monoglutamate to the polyglutamate form. Enzymes of the intestinal mucosa split polyglutamates back to monoglutamate, whereas liver enzymes add glutamate to tetrahydrofolate or to other reduced folic acids.**

may present with weakness, lethargy, jaundice, and dementia, as well as atrophy of the lingual papillae and glossitis. Neuropathy is the presenting feature in about 12% of patients with cobalamin (vitamin B_{12}) deficiency without concomitant anemia.[69] Patients with severe cobalamin deficiency initially complain of paresthesia. The sense of touch and temperature sensitivity may be minimally impaired. Memory impairment and depression may be prominent.[69] The disease may progress, involving the dorsal columns, causing ataxia and weakness. The physical examination reveals a broad-base gait, Romberg sign, slowed reflexes, and loss of sense of position and feeling of vibration (especially when tested with a 256 Hz tuning fork). If the disorder is not detected and treated, the lateral columns become involved, resulting in weakness, spasticity, inability to walk, sustained clonus, hyperreflexia, and Babinski sign. Because the peripheral nerves, as well as the dorsal and lateral columns, are involved, these neurologic manifestations are sometimes termed subacute combined degeneration or subacute combined system disease.

Cobalamin deficiency appears to be the cause of various neuropsychiatric disorders, with such symptoms as paresthesia, ataxia, limb weakness, gait disturbance, memory defects, hallucinations, and personality and mood changes.[65] These symptoms, however, cannot be easily accounted for by the type of spinal cord lesions that occur in patients with cobalamin deficiency. Investigators have tried to determine whether a defect in methionine synthesis or an abnormality in propionic acid metabolism accounts for the neuropathy associated with cobalamin deficiency [*see Figure 5b*], but the exact mechanism remains obscure. Accruing evidence supports the impairment of methionine synthase as the cause of the neuropathy.[60,69] A recent study measured various metabolites of cobalamin and discovered that only high levels of plasma cysteine were predictive of neurologic dysfunction.[70]

Diagnostic workup The evaluation of suspected cobalamin deficiency generally proceeds in two stages: documenting the presence of the vitamin deficiency and determining its cause (e.g., pernicious anemia, malabsorption, dietary lack). The diagnosis can often be established by (1) measurement of the serum cobalamin concentration, (2) evaluation of specific metabolites, and (3) use of the Schilling test to establish malabsorption of cobalamin.

Macrocytosis (mean corpuscular volume [MCV] greater than 100 fl) is a hallmark of cobalamin deficiency, but it may be masked by concurrent disorders, such as iron deficiency. If macrocytocis is not apparent on examination of the peripheral smear, it is easily detected when red blood cell counts are made with an electronic particle counter. The peripheral smear shows macro-ovalocytes, fish-tailed red blood cells, hypersegmented neutrophils, and, occasionally, nucleated red blood cells [*see Figure 2*]. The finding that a single polymorphonuclear neutrophil has six lobes or that 5% of PMNs have five lobes constitutes strong evidence of megaloblastic anemia. In severe cases, granulocytopenia and thrombocytopenia are present. Examination of the bone marrow is usually not necessary, but if performed, it reveals megaloblastic erythroid hyperplasia and giant metamyelocytes.[60] If severe iron deficiency is concurrent with macrocytosis, the full morphologic expression of megaloblastosis is blocked, although the giant metamyelocytes in the marrow and hypersegmented PMNs in the peripheral blood will still be present.

Plasma cobalamin levels and red blood cell folic acid levels should be measured if the MCV is greater than 100 fl. If performed, a bone marrow aspirate and biopsy typically reveal

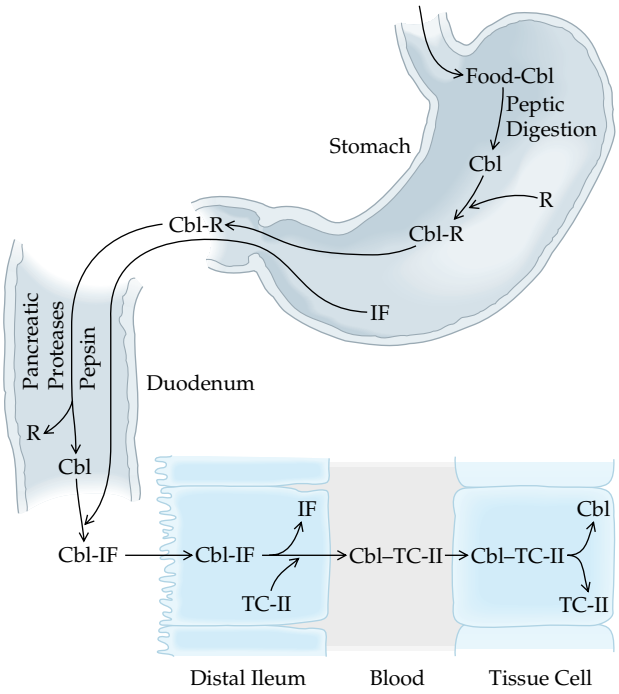


Figure 7 **Cobalamin assimilation. Dietary cobalamin (Cbl) enters the stomach and binds to R protein. This physiologically inactive complex enters the duodenum. In the small intestine, pancreatic enzymes and pepsin digest the R protein, and Cbl binds to intrinsic factor (IF). The Cbl-IF complex passes through the intestine until it reaches receptors on the microvilli of mucosal cells in the distal ileum. The Cbl is then transferred to transcobalamin II (TC-II), which circulates in the blood until it binds to receptors on cells in the body and is internalized.**

enormous megaloblastic erythroid hyperplasia with giant metamyelocytes.[60] The hypercellularity detected on bone marrow examination can be so dramatic and megaloblasts so immature that clinicians still sometimes make the erroneous diagnosis of leukemia.[60,71]

The standard approach to determining the cause of proven cobalamin deficiency has traditionally relied on the Schilling test, which is becoming difficult to order. The Schilling test measures the absorption of cobalamin labeled with cobalt-57 (^{57}Co). After 1 μg of radioactively labeled cobalamin is given orally, 1,000 μg of unlabeled cobalamin is given parenterally. The parenteral dose saturates transcobalamins I, II, and III, so that a significant portion of the absorbed material is flushed and excreted in the urine. If the amount of ^{57}Co-labeled cobalamin measured in an accurately collected 24-hour urine sample is less than 10% of the dose that was administered orally, cobalamin absorption is poor.

There are increasing numbers of reports of patients with proven pernicious anemia who have low or borderline serum cobalamin levels but normal Schilling test results. As the gastric atrophic lesion of pernicious anemia progresses, the ability to produce acid-pepsin is lost before all IF activity disappears. Thus, the ability to cleave the R protein–cobalamin complex, freeing cobalamin to bind to IF, is impaired. Coexisting infection with *Helicobacter pylori* may further impair production of acid-pepsin.[72] However, there may be sufficient IF to bind the free oral cobalamin administered in the Schilling test and therefore yield a normal value.

Malabsorption of cobalamin can be demonstrated by means of a food Schilling test, which is not available clinically. This test is performed with eggs from chickens that have been injected with radioactive cobalamin[73] and indicates whether there is insufficient acid-pepsin to split the cobalamin-enzyme complex and release free cobalamin to be bound by IF. If pernicious anemia is strongly suspected in a patient whose Schilling test result is apparently normal and whose plasma cobalamin is not diagnostically low, other steps should be taken to confirm the diagnosis, including examination of the blood cell morphology, measurement of the anti-IF antibody, or performance of a therapeutic trial with parenterally administered cobalamin. Measurement of the serum levels of homocysteine and methylmalonic acid is increasingly being used, because both are elevated as a consequence of cobalamin deficiency [*see Figures 5a and b*].[74]

If the initial Schilling test demonstrates reduced excretion of cobalamin, a second phase of the test may be conducted, aimed at correcting cobalamin absorption caused by pernicious anemia. In this phase of the test, supplementary oral IF is administered and will normalize the cobalamin absorption unless the supplementary IF is not fully active, the patient secretes antibodies to IF, or the patient is taking drugs that interfere with cobalamin absorption. In no case, however, will supplementary cobalamin be effective in patients with intestinal malabsorption. It is important to recognize that prolonged cobalamin deficiency impairs intestinal epithelial cells and thus impairs absorption. Therefore, the second stage of the test should only be performed after several weeks of cobalamin replacement therapy. If the result of the second stage of the Schilling test is abnormally low, this suggests the presence of generalized malabsorption, such as may occur in sprue, pancreatic insufficiency, or blind loop syndromes.

Factors affecting test results Concurrent α-thalassemia may minimize the macrocytosis of pernicious anemia.[75] This possibility should be considered particularly in patients of African descent, among whom there is a high incidence of α-thalassemia (about 30%). Anemia of chronic disease or anemia resulting from blood loss and iron deficiency can also reduce the degree of macrocytosis but will not affect the hypersegmentation of neutrophils. In one study, iron deficiency was discovered in 20% of 121 patients with pernicious anemia[76]; in another study, 19% of patients with pernicious anemia were not anemic, and 33% did not have macrocytosis.[77]

Falsely low serum cobalamin levels occur during pregnancy and in folic acid deficiency states.[74] In the past, a decline in the serum cobalamin level was usually not considered important unless the value was very low (i.e., < 150 pg/ml). It has become clear, however, that patients with serum cobalamin levels as high as 250 pg/ml and perhaps higher may have cobalamin deficiency.[73,77,78] Fortunately, the finding of macro-ovalocytes or hypersegmented PMNs on the peripheral smear remains a sensitive indicator for the presence of cobalamin deficiency.

Determining the underlying cause After the presence of macrocytosis and a reduced cobalamin level have been identified, the cause of these conditions must be determined. It is important to remember that macrocytosis can be caused by conditions other than pernicious anemia, including folic acid deficiency, liver disease, alcohol abuse, reticulocytosis, and ingestion of drugs such as antimetabolites, alkylating agents, and zidovudine.[52,75] Cobalamin deficiency can be caused by inadequate absorption resulting from gastric abnormalities (e.g., pernicious anemia, gastritis) and small bowel disease (e.g., tropical sprue, Crohn disease), and pancreatic insufficiency [*see Table 4*].

Gastric surgery in which the IF, pepsin, and acid-secreting components are removed often results in cobalamin deficiency (it occurred in 31% of patients in one study[79]). Patients who have undergone gastric surgery should be regularly screened by measurements of plasma cobalamin or homocysteine levels and supplemented with lifelong cobalamin therapy if the levels are low.[79]

Pancreatic insufficiency can result in malabsorption of cobalamin if the damaged pancreas does not produce enough trypsin and chymotrypsin for digesting the R protein–cobalamin complex and freeing the vitamin to form the complex with IF [*see Figure 7*] [*see 4:V Diseases of the Pancreas*].

There are other causes of cobalamin deficiency. In vegetarians, especially vegans, profound nutritional megaloblastic anemia can develop as a result of very low cobalamin intake. Deficiencies of folic acid and iron have also been observed in vegans.[80] A careful patient history should indicate the possibility of inadequate dietary intake of cobalamin. Infants of vegan mothers can become severely cobalamin deficient, particularly when they are breast-fed.[65] Cobalamin deficiency is surprisingly common in less well developed countries where people are not strict vegans.[65] The incidence is particularly high in pregnant women and in preschool-age children.[65]

Treatment

Specific replacement should be started promptly after the diagnosis has been made and serum samples have been taken to determine cobalamin levels. Patients who have a low serum cobalamin level and macrocytic anemia should undergo a trial of parenteral cobalamin therapy. The diagnosis of cobalamin deficiency is confirmed if cobalamin therapy produces a reticulocytosis in 3 to 4 days that is associated with a rise in the hemoglobin level and a fall in the MCV.

Table 4 Causes of Cobalamin Deficiency

Inadequate Diet
Strict vegetarianism
Inadequate Absorption
Gastric abnormalities that produce deficient or defective intrinsic factor
Pernicious anemia
Total gastrectomy
Gastritis
Small bowel disease
Ileal resection or bypass
Blind loop syndrome with abnormal gut flora
Malabsorption
Tropical sprue
Crohn disease
Pancreatic insufficiency
Interference with Cobalamin Absorption
Drugs
Neomycin
Biguanides
Colchicine
Ethanol
Aminosalicylic acid
Omeprazole
Fish tapeworm competing for cobalamin
Degradation of Cobalamin Coenzymes
N_2O anesthesia
Rare Congenital Disorders
Transcobalamin II deficiency
Defective intrinsic factor production

If the patient has symptoms of severe anemia, packed red blood cells can be transfused; the transfusion should be administered very slowly to avoid precipitating or aggravating congestive heart failure. This circumstance is one of the few in which a single-unit transfusion may be justified, because it may produce a 25% increase in oxygen-carrying capacity. A large dose of cobalamin should be given because the retention of parenterally administered cobalamin is poor but variable; the vitamin is inexpensive and has no harmful side effects. The reticulocyte response begins in 4 to 6 days, and the granulocyte count, if low, begins to increase at the same time. The hypersegmentation of PMNs disappears after 10 to 14 days, which suggests that in the megaloblastic anemias, granulopoiesis is affected by cobalamin deficiency at two different steps: (1) the lobe number of the PMNs is determined, and (2) granulocytes mature and leave the marrow.[55] Weekly dosages of 1,000 μg of parenteral cobalamin for 6 weeks should be followed by parenteral dosages of 1,000 μg monthly for life. The standard parenteral preparation is cyanocobalamin. For pancreatic insufficiency, cobalamin can be given parenterally or pancreatic enzymes can be administered orally. Specific therapy must be designed for patients with intestinal forms of malabsorption.

Because a small amount of cobalamin is absorbed even in the absence of IF and because only 1 μg/day is required, oral cobalamin has proved adequate for replacement in patients with pernicious anemia, freeing the patient from monthly injections (2,000 μg/day p.o. is recommended).[81]

Diagnosis

Clinical manifestations The patient with folic acid deficiency has a clinical presentation that is distinct from that of the patient with cobalamin deficiency.[82] The patient may abuse alcohol or other drugs and have poor dietary intake of folic acid. Patients with folic acid deficiency are more often malnourished than those with cobalamin deficiency. The gastrointestinal symptoms in folic acid deficiency are similar to those in cobalamin deficiency but may be more severe than those in pernicious anemia. Diarrhea is often present. The hematologic manifestations of folic acid deficiency are the same as those of cobalamin deficiency: severe macrocytic anemia, a low absolute reticulocyte count, and a characteristic blood smear showing macro-ovalocytes, occasional megaloblasts, and hypersegmented neutrophils. Patients with megaloblastic anemia who do not have glossitis, a family history of pernicious anemia, or the neurologic features described for cobalamin deficiency may have folic acid deficiency.

Diagnostic workup A meticulous dietary history is important because food faddism, poor dietary intake, and alcoholism are the usual causes of severe folic acid deficiency [*see Table 5*]. Cobalamin and folic acid deficiencies frequently coexist and are not easily distinguished. In evaluating patients for folic acid deficiency, values for the levels of serum folic acid, serum cobalamin, and red blood cell folic acid must be obtained. The red blood cell folic acid level reflects tissue stores[83] but may be reduced in patients with severe cobalamin deficiency. In isolated cases, the serum folic acid level of cobalamin-deficient patients is usually normal or elevated. Severe, long-standing cobalamin deficiency leads to anorexia and GI disturbances, which may cause dietary folic acid deficiency. As a result, both serum cobalamin and folic acid levels are low, producing a double-deficiency state.

A serum folic acid level less than 2 ng/ml is consistent with folic acid deficiency, as is a red blood cell folic acid level less than 150 ng/ml. If the test results are inconclusive or if it is necessary to distinguish the megaloblastosis of folic acid deficiency from that of cobalamin deficiency, measurements of the serum methylmalonate and homocysteine levels are helpful. If both metabolite tests are normal (i.e., methylmalonate level of 70 to 270 nmol/L and total homocysteine level of 5 to 14 μmol/L), deficiency of both vitamins is ruled out. If the methylmalonate level is normal but the total homocysteine level is increased, folic acid deficiency is likely and investigation into the underlying cause is appropriate.[83]

Determining the underlying cause Folic acid deficiency is most frequently caused by poor dietary intake, but it may also result from inadequate absorption secondary to disease or drug administration [*see Table 5*]. Ingestion of ethanol by well-nourished individuals does not produce megaloblastosis, but in patients with borderline folic acid stores, ethanol can lower serum folic acid levels and block the reticulocyte response to folic acid administration. Alcohol may block release of folic acid from tissues to the serum.

Megaloblastic anemia occurring as a consequence of drug administration or pregnancy is likely to be caused by folic acid deficiency. Many of the antineoplastic and immunosuppressive agents produce megaloblastosis; these include fluorouracil, hy-

Table 5 Causes of Folic Acid Deficiency

Mechanism	*Cause*
Absolutely inadequate intake	Alcoholism Nutritional deficiencies
Relatively inadequate intake (resulting from increased folic acid requirements)	Pregnancy Severe hemolysis Chronic hemodialysis or peritoneal dialysis
Inadequate absorption	Tropical sprue Gluten-sensitive enteropathy (nontropical sprue) Crohn disease Lymphoma or amyloidosis of small bowel Diabetic enteropathy Intestinal resections or diversions
Drug-induced interference with folic acid metabolism	Action of dihydrofolate reductase blocked by methotrexate, trimethoprim, pyrimethamine Reduced folate absorption and tissue folate depletion caused by sulfasalazine Interference of unknown mechanism caused by phenytoin, ethanol, antituberculosis drugs, ?oral contraceptives

droxyurea, mercaptopurine, thioguanine, cytarabine, and azathioprine. In pregnant women, the presence of megaloblastosis may not be initially apparent. Because the combination of folic acid and iron deficiency is common, full expression of megaloblastosis is often blocked, and the patient will have a dimorphic anemia rather than the easily identifiable macro-ovalocytosis. Hypersegmentation of PMNs persists.[60,83]

An abnormality in folate metabolism can be caused by a chromosomal mutation, and women who are homozygous for this defect are thought to be at higher risk for pregnancies affected by neural tube defects. One of the enzymes that regulates homocysteine levels, 5,10-methylenetetrahydrofolate reductase has a genetic variant, C677T. Individuals homozygous for this variant have increased plasma homocysteine levels that are lowered by folate supplementation. About 5% to 10% of the general population are homozygous for this variant. Both pregnant and nonpregnant women who are homozygous for the C677T mutation have significantly lower red blood cell folic acid levels.[84] These women may be susceptible to cardiovascular disease and stroke and may bear children with neural tube defects.[84,85] It would be advisable to know before pregnancy that a woman is homozygous for this variant, and genetic testing would be helpful if a woman has a family history of this defect.

A number of intestinal disorders cause folic acid deficiency. These include severe pancreatic disease and small bowel disease, including malabsorption, ileal disease, Crohn disease, resection, and bypass [*see Table 5*]. When there is no apparent cause of cobalamin deficiency, it may be practical to suspect an undiagnosed disease of malabsorption. In one prospective study of patients who had laboratory-defined folate deficiency, 10.9% were positive for celiac disease antibodies and 4.7% had histologically confirmed celiac disease.[86]

Treatment

Standard therapy for folic acid deficiency is 1 mg/day orally. The response, manifested by reticulocytosis in 4 to 6 days, loss of megaloblastosis, and the return of normal blood counts, confirms the diagnosis of folic acid deficiency. Neutrophil hypersegmentation disappears only after 10 to 14 days, however.[60] Patients with megaloblastosis and severe bone marrow depression secondary to administration of drugs that block dihydrofolate reductase, such as pyrimethamine and methotrexate, may be treated with folinic acid. In the case of toxicity after single large doses of methotrexate, a single equivalent dose of I.M. folinic acid (i.e., milligram for milligram) will suffice. For toxicity after chronic pyrimethamine therapy, 1 to 5 mg of folinic acid daily can be given without blocking the antimalarial effects of pyrimethamine. Megaloblastosis caused by anticonvulsant therapy can be treated with 1 mg of folic acid daily. Supplementation during pregnancy is advised and may also be useful for patients who have severe chronic hemolysis.

In most patients (i.e., those who do not require a large amount of folic acid because of conditions such as hemolysis or pregnancy), a hematologic response occurs after administration of 200 μg of folic acid daily. The increased demand of folic acid during pregnancy requires administration of about 200 to 300 μg/day.[87] Furthermore, folic acid supplementation seems to prevent fetal neural tube defects.[88] Such neural tube defects may occur in the embryo or very early in gestation—even before the pregnancy is confirmed.[89,90] Therefore, it is recommended that women of childbearing age or those who plan to become pregnant receive about 400 μg of folic acid a day. Women who are homozygous for the C677T mutation should also take folic acid supplements. Staple foods such as flour and cereal grains can be fortified with folic acid. Concern has been expressed, however, that folic acid supplementation may mask the megaloblastosis of pernicious anemia, causing the development of severe neuropathy rather than anemia.[89]

SIDEROBLASTIC ANEMIAS

Definition

The sideroblastic anemias are a heterogeneous group of disorders characterized by anemia, ringed sideroblasts in the marrow, and ineffective erythropoiesis.[91] There are hereditary and acquired forms; the latter are subdivided into benign and malignant variants. A fairly common form is the myelodysplastic syndrome called refractory anemia with ringed sideroblasts. Other than alcohol and drugs (e.g., isoniazid), the secondary causes of these diseases remain largely unknown.

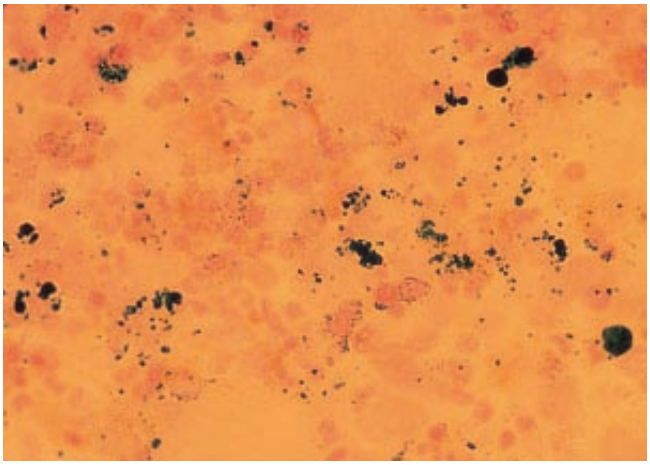


Figure 8 **Prussian blue stain shows ringed sideroblasts in the bone marrow of a patient who has idiopathic sideroblastic anemia.**

Table 6 Sideroblastic Anemias

Type	*Disorders*
Hereditary variant, probably benign	Sex-linked disorders, autosomal disorders
Acquired variant, probably benign	Mitochondrial DNA deletions (Pearson syndrome)
Probably benign variant	Induced by drugs (e.g., isoniazid or other antituberculosis drugs) or by lead intoxication; alcoholic sideroblastosis; pyridoxine-responsive anemia
Clonal disorder (myelodysplastic syndrome)	Refractory anemia with ringed sideroblasts, acquired idiopathic sideroblastic anemia

Pathophysiology

Abnormalities of heme synthesis are probably the most frequent cause of the hereditary sideroblastic anemias. Molecular defects of the enzyme 5-aminolevulinate synthase have been described as the cause of this abnormality.[90,92] This enzyme initiates the heme synthetic pathway, and its impairment profoundly affects heme synthesis. In other cases, there are major deletions in mitochondrial DNA. Iron enters erythroid precursors, but because heme synthesis is impaired, the iron cannot be incorporated into heme and accumulates on the cristae of mitochondria.[90]

Diagnosis

The principal feature common to all sideroblastic anemias is a refractory or progressive anemia. However, mild, lifelong anemia may go unnoticed. The diagnosis of sideroblastic anemia is established by reticulocytopenia; the red blood cells on smear are frequently profoundly hypochromic and microcytic, and distorted red blood cells and basophilic stippling may be noted.[93,94] Occasionally, Pappenheimer bodies (deposits of iron that stain with the Prussian blue reagent) are present in the red blood cells. There are ringed sideroblasts seen on the marrow aspirate (bone marrow normoblasts with heavy incrustations of nonferritin iron on the mitochondria) [*see Figure 8*]. Because of ineffective erythropoiesis, there is saturation of serum iron-binding capacity (usually approaching 80%) and elevation of the serum lactate dehydrogenase level. Cytogenetic study of the bone marrow may reveal one of the typical patterns seen in the myelodysplastic syndromes. The sideroblastic anemias can be classified into four groupings: hereditary (probably benign), acquired (probably benign), probably benign, and clonal disorder [*see Table 6*].

Treatment

For prognostic purposes, it is important to decide whether the patient has a benign or malignant form of sideroblastic anemia. It is also important to recognize reversible forms of sideroblastic anemia (e.g., those caused by alcoholism, folic acid deficiency, and drugs such as isoniazid and chloramphenicol) and to discontinue any potentially offending agents.

Indicators of myelodysplasia include granulocytopenia, thrombocytopenia, dysplastic marrow granulopoiesis, bilobed megakaryocytes, and typical cytogenetic abnormalities. In rare cases, patients have a reticulocyte and hemoglobin response to pyridoxine (200 to 600 mg/day), with or without folic acid.[95]

The author has served as a consultant for Tularik, Inc., and Receptron, Inc.

References

1. Means RT Jr, Krantz SB: Progress in understanding the pathogenesis of the anemia of chronic disease. Blood 80:1639, 1992

2. Cash JM, Sears DA: The anemia of chronic disease: spectrum of associated diseases in a series of unselected hospitalized patients. Am J Med 87:638, 1989

3. Means RT Jr: Pathogenesis of the anemia of chronic disease: a cytokine-mediated anemia. Stem Cells 13:32, 1995

4. Lacombe C: Resistance to erythropoietin. N Engl J Med 334:660, 1996

5. Voulgari PV, Kolios G, Papadopoulos GK, et al: Role of cytokines in the pathogenesis of anemia of chronic disease in rheumatoid arthritis. Clin Immunol 92:153, 1999

6. Ganz T: Hepcidin, a key regulator of iron metabolism and mediator of anemia of inflammation. Blood 102:783, 2003

7. Weiss G: Iron and anemia of chronic disease. Kidney Int 55(suppl 69):12, 1999

8. Pincus T, Olsen NJ, Russell IJ, et al: Multicenter study of recombinant human erythropoietin in correction of anemia in rheumatoid arthritis. Am J Med 89:161, 1990

9. Henry DH, Beall GN, Benson CA, et al: Recombinant human erythropoietin in the treatment of anemia associated with human immunodeficiency virus (HIV) infection and zidovudine therapy. Ann Intern Med 117:739, 1992

10. Dowlati A, R'Zik S, Fillet G, et al: Anaemia of lung cancer is due to impaired erythroid marrow response to erythropoietin stimulation as well as relative inadequacy of erythropoietin production. Br J Haematol 97:297, 1997

11. Schreiber S, Howaldt S, Schnoor M, et al: Recombinant erythropoietin for the treatment of anemia in inflammatory bowel disease. N Engl J Med 334:619, 1996

12. Ludwig H, Fritz E, Leitgeb C, et al: Prediction of response to erythropoietin treatment in chronic anemia of cancer. Blood 84:1056, 1994

13. Goodnough LT, Skikne B, Brugnara C: Erythropoietin, iron, and erythropoiesis. Blood 96:823, 2000

14. Allen DA, Breen C, Yaqoob MM, et al: Inhibition of CFU-E colony formation in uremic patients with inflammatory disease: role of IFN-γ and TNF-α. J Investig Med 47:204, 1999

15. Mladenovic J: Aluminum inhibits erythropoiesis in vitro. J Clin Invest 81:1661, 1988

16. Eschbach JW, Egrie JC, Downing MR, et al: Correction of the anemia of end-stage renal disease with recombinant human erythropoietin: results of a combined phase I and II clinical trial. N Engl J Med 316:73, 1987

17. Altmann P, Plowman D, Marsh F, et al: Aluminum chelation therapy in dialysis patients: evidence for inhibition of haemoglobin synthesis by low levels of aluminum. Lancet 1:1012, 1988

18. Coleman N, Herbert V: Hematologic complications of alcoholism: overview. Semin Hematol 17:164, 1980

19. Lindenbaum J: Folate and vitamin B_{12} deficiencies in alcoholism. Semin Hematol 17:119, 1980

20. Latvala J, Parkkila S, Melkko J, et al: Acetaldehyde adducts in blood and bone marrow of patients with ethanol-induced erythrocyte abnormalities. Mol Med 7:401, 2001

21. Conrad ME, Barton JC: Anemia and iron kinetics in alcoholism. Semin Hematol 17:149, 1980

22. Baldwin JG Jr: Hematopoietic function in the elderly. Arch Intern Med 148:2544, 1988

23. Penninx BW, Guralnik JM, Onder G, et al: Anemia and decline in physical performance among older persons. Am J Med 115:104, 2003

24. Bacigalupo A, Hows J, Gluckman E, et al: Bone marrow transplantation (BMT) versus immunosuppression for the treatment of severe aplastic anaemia (SAA): a report of the EBMT SAA working party. Br J Haematol 70:177, 1988

25. Brown KE, Tisdale J, Barrett AJ, et al: Hepatitis-associated aplastic anemia. N Engl J Med 336:1059, 1997

26. Tzakis AG, Arditi M, Whitington PF, et al: Aplastic anemia complicating orthotopic liver transplantation for non-A, non-B hepatitis. N Engl J Med 319:393, 1988

27. Young NS: Autoimmunity and its treatment in aplastic anemia. Ann Intern Med 126:166, 1997

28. Young NS, Maciejewski J: The pathophysiology of acquired aplastic anemia. N Engl J Med 336:1365, 1997

29. Maciejewski JP, Selleri C, Sato T, et al: A severe and consistent deficit in marrow and circulating primitive hematopoietic cells (long-term culture-initiating cells) in acquired aplastic anemia. Blood 88:1983, 1996

30. Go RS, Lust JA, Phyliky RL: Aplastic anemia and pure red cell aplasia associated with large granular lymphocyte leukemia. Semin Hematol 40:196, 2003

31. Young NS: Aplastic anemia. Lancet 346:228, 1995

32. Rosenfeld SJ, Kimball J, Vining D, et al: Intensive immunosuppression with antithymocyte globulin and cyclosporine as treatment for severe acquired aplastic anemia. Blood 85:3058, 1995

33. Deeg HJ, Leisenring W, Storb R, et al: Long-term outcome after marrow transplantation for severe anemia. Blood 91:3637, 1998

34. Doney K, Leisenring W, Storb R, et al: Primary treatment of acquired aplastic anemia: outcomes with bone marrow transplantation and immunosuppressive therapy. Ann Intern Med 126:107, 1997

35. Frickhofen N, Heimpel H, Kaltwasser JP, et al: Antithymocyte globulin with or without cyclosporin A: 11-year follow-up of a randomized trial comparing treatments of aplastic anemia. Blood 101:1236, 2003

36. Bacigalupo A, Bruno B, Saracco P, et al: Antilymphocyte globulin, cyclosporine, prednisolone, and granulocyte colony-stimulating factor for severe aplastic anemia: an update of the GITMO/EBMT study on 100 patients. Blood 95:1931, 2000

37. Frickhoffen N, Kaltwasser JP, Schrezenmeier H, et al: Treatment of aplastic anemia with antilymphocyte globulin and methylprednisolone with or without cyclosporine. N Engl J Med 324:1297, 1991

38. Socié G, Henry-Amar M, Bacigalupo A, et al: Malignant tumors occurring after treatment of aplastic anemia. N Engl J Med 329:1152, 1993

39. Di Bona E, Rodeghiero B, Bruno A, et al: Rabbit antithymocyte globulin (r-ATG) plus cyclosporine and granulocyte colony stimulating factor is an effective treatment for aplastic anaemia patients unresponsive to a first course of intensive immunosuppressive therapy. Br J Haematol 107:330, 1999

40. Brodsky RW, Sensenbrenner LL, Jones RJ: Complete remission in severe aplastic anemia after high-dose cyclophosphamide without bone marrow transplantation. Blood 87: 491, 1996

41. Tisdale JF, Maciejewski JP, Nunez O, et al: Late complications following treatment for severe aplastic anemia (SAA) with high-dose cyclophosphamide (Cy): follow-up of a randomized trial. Blood 100:4668, 2002

42. Tichelli A, Socie G, Henry-Amar M, et al; Effectiveness of immunosuppressive therapy in older patients with aplastic anemia. Ann Intern Med 130:193, 1999

43. Kojima S, Inaba J, Yoshimi A, et al: Unrelated donor marrow transplantation in children with severe aplastic anaemia using cyclophosphamide, antithymocyte globulin and total body irradiation. Br J Haematol 114:706, 2001

44. Paquette RL, Tebyani N, Frane M, et al: Long-term outcome of aplastic anemia in adults treated with antithymocyte globulin: comparison with bone marrow transplantation. Blood 85:283, 1995

45. Kojima S, Matsuyama T, Kato S, et al: Outcome of 154 patients with severe aplastic anemia who received transplants from unrelated donors: the Japan Marrow Donor Program. Blood 100:799, 2002

46. Storb R, Blume KG, O'Donnel MR, et al: Cyclophosphamide and antithymocyte globulin to condition patients with aplastic anemia for allogeneic marrow transplantations: the experience in four centers. Biol Blood Marrow Transplant 7:39, 2001

47. Maung ZT, Norden J, Middleton PG, et al: Pure red cell aplasia: further evidence of T cell clonal disorder. Br J Haematol 87:189, 1994

48. Casadevall N, Dufuy E, Molho-Sabatier P, et al: Brief report: autoantibodies against erythropoietin in a patient with pure red-cell aplasia. N Engl J Med 334:630, 1996

49. Charles RJ, Sabo KM, Kidd PG, et al: The pathophysiology of pure red cell aplasia: implications for therapy. Blood 87:4831, 1996

50. Lacy MQ, Kurtin PJ, Tefferi A: Pure red cell aplasia: association with large granular lymphocyte leukemia and the prognostic value of cytogenetic abnormalities. Blood 87: 3000, 1996

51. Fisch P, Handgretinger R, Schaefer HE: Pure red cell aplasia. Br J Haematol 111:1010, 2000

52. Walker RE, Parker RI, Kovacs JA, et al: Anemia and erythropoiesis in patients with the acquired immunodeficiency syndrome (AIDS) and Kaposi sarcoma treated with zidovudine. Ann Intern Med 108:372, 1988

53. Baker RI, Manoharan A, De Luca E, et al: Pure red cell aplasia of pregnancy: a distinct clinical entity. Br J Haematol 85:619, 1993

54. Palmieri G, Lastoria S, Colao A, et al: Successful treatment of a patient with thymoma and pure red-cell aplasia with octreotide and prednisone. N Engl J Med 336:263, 1997

55. McGuire WA, Yang HH, Bruno E, et al: Treatment of antibody-mediated pure red-cell aplasia with high-dose intravenous gamma globulin. N Engl J Med 317:1004, 1987

56. Yamada O, Mizoguchi H, Oshimi K: Cyclophosphamide therapy for pure red cell aplasia associated with granular lymphocyte-proliferative disorders. Br J Haematol 97:392, 1997

57. Ramratnam B, Gollerkeri A, Schiffman FJ, et al: Management of persistent B19 parvovirus infection in AIDS. Br J Haematol 91:90, 1995

58. Means RT Jr, Dessypris EN, Krantz SB, et al: Treatment of refractory pure red cell aplasia with cyclosporine A: disappearance of IgG inhibitor associated with clinical response. Br J Haematol 78:114, 1991

59. Miller BU, Tichelli A, Passweg JR, et al: Successful treatment of refractory acquired pure red cell aplasia (PRCA) by allogeneic bone marrow transplantation. Bone Marrow Transplant 23:1205, 1999

60. Wickramasinghe S: Morphology biology and biochemistry of cobalamin- and folate-deficient bone marrow cells. Baillieres Clin Haematol 8:441, 1995

61. Koury MJ, Horne DW: Apoptosis mediates and thymidine prevents erythroblast destruction in folate deficiency anemia. Proc Natl Acad Sci USA 91:4067, 1994

62. Igram CF, Davidoff AN, Marais E, et al: Evaluation of DNA analysis for evidence of apoptosis in megaloblastic anaemia. Br J Haematol 96:576, 1997

63. Tefferi A, Pruthi RK: The biochemical basis of cobalamin deficiency. Mayo Clin Proc 69:181, 1994

64. Pruthi RK, Tefferi A: Pernicious anemia revisited. Mayo Clin Proc 69:144, 1994

65. Allen LH: Vitamin B_{12} metabolism and status during pregnancy, lactation and infancy. Adv Exp Med Biol 352:173, 1996

66. Guéant JL, Safi A, Aimone-Gastin I, et al: Autoantibodies in pernicious anemia type I patients recognize sequence 251-256 in human intrinsic factor. Proc Assoc Am Physicians 109:462, 1997

67. Hsing AW, Hansson L-E, McLaughlin JK, et al: Pernicious anemia and subsequent cancer. Cancer 71:745, 1993

68. Carmel R, Spencer CA: Clinical and subclinical thyroid disorders associated with pernicious anemia: observations on abnormal thyroid-stimulating hormone levels and on a possible association of blood group O with hyperthyroidism. Arch Intern Med 142:1465, 1982

69. Weir DG, Scott JM: The biochemical basis of the neuropathy in cobalamin deficiency. Baillieres Clin Haematol 8:479, 1995

70. Carmel R, Melnyk S, James SJ: Cobalamin deficiency with and without neurologic abnormalities: differences in homocysteine and methionine metabolism. Blood 101:3302, 2003

71. Dokal IS, Cox TM, Galton DAG: Vitamin B_{12} and folate deficiency presenting as leukaemia. BMJ 300:1263, 1990

72. Annibale B, Lahner E, Bordi C, et al: Role of *Helicobacter pylori* infection in pernicious anaemia. Dig Liver Dis 32:756, 2000

73. Carmel R, Sinow RM, Siegel ME, et al: Food cobalamin malabsorption occurs frequently in patients with unexplained low serum cobalamin levels. Arch Intern Med 148:1715, 1988

74. Dharmarajan TS, Adiga GU, Norkus EP: Vitamin B_{12} deficiency: recognizing subtle symptoms in older adults. Geriatrics 58:30, 2003

75. Chanarin I, Metz J: Diagnosis of cobalamin deficiency: the old and the new. Br J Haematol 97:695, 1997

76. Carmel R, Weiner JM, Johnson CS: Iron deficiency occurs frequently in patients with pernicious anemia. JAMA 257:1081, 1987

77. Carmel R: Pernicious anemia: the expected findings of very low serum cobalamin levels, anemia, and macrocytosis are often lacking. Arch Intern Med 148:1712, 1988

78. Green R: Screening for vitamin B_{12} deficiency: caveat emptor. Ann Intern Med 124: 509, 1996

79. Sumner AE, Chin MM, Abrahm JL, et al: Elevated methylmalonic acid and total homocysteine levels show high prevalence of vitamin B_{12} deficiency after gastric surgery. Ann Intern Med 124:469, 1996

80. Pippard MJ: Megaloblastic anaemia: geography and diagnosis. Lancet 334:6, 1994

81. Kuzminski AM, Del Giacco EJ, Allen RH, et al: Effective treatment of cobalamin deficiency with oral cobalamin. Blood 92:1191, 1998

82. Carmel R, Green R, Rosenblatt DS: Update on cobalamin, folate, and homocysteine. Hematology (Am Soc Hematol Educ Program) 62:2003

83. Amose RJ, Dawson DW, Fish DI, et al: Guidelines on the investigation and diagnosis of cobalamin and folate deficiencies. Clin Lab Haematol 16:101, 1994

84. Molloy AM, Daly S, Mills JL, et al: Thermolabile variant of 5,10-methylenetetrahydrofolate reductase associated with low red-cell folates: implications for folate intake recommendations. Lancet 349:1591, 1997

85. Wilcken DE: MTHFR 677CT mutation, folate intake, neural-tube defect, and risk of cardiovascular disease. Lancet 350:603, 1997

86. Howard MR, Turnbull AF, Morley P, et al: A prospective study of the prevalence of undiagnosed celiac disease in laboratory defined iron and folate deficiency. J Clin Pathol 55:754, 2002

87. McPartlin J, Halligan A, Scott JM, et al: Accelerated folate breakdown in pregnancy. Lancet 341:148, 1993

88. Wald NJ, Bower C: Folic acid, pernicious anaemia, and prevention of neural tube defects. Lancet 343:307, 1994

89. Carmel R: Subtle cobalamin deficiency. Ann Intern Med 124:338, 1996

90. Cotter PD, May A, Li L, et al: Four new mutations in the erythroid-specific 5-aminolevulinate synthase (ALAS2) gene causing X-linked sideroblastic anemia: increased pyridoxine responsiveness after removal of iron overload by phlebotomy and coinheritance of hereditary hemochromatosis. Blood 93:1757, 1999

91. Alcindor T, Bridges KR: Sideroblastic anaemias. Br J Haematol 116:733, 2002

92. Nakajima O, Takahashi S, Harigae H, et al: Heme deficiency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. EMBO J 18:6282, 1999

93. Bottomley SS, Healy HM, Brandenburg MA, et al: 5-Aminolevulinate synthase in sideroblastic anemias: mRNA and enzyme activity levels in bone marrow cells. Am J Hematol 41:76, 1992

94. Bottomley SS, May BK, Cox TC, et al: Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia. J Bioenerg Biomembr 27:161, 1995

95. Cotter PD, May A, Fitzsimons EJ, et al: Late-onset X-linked sideroblastic anemia: missense mutations in the erythroid-aminolevulinate synthase (ALAS2) gene in two pyridoxine-responsive patients initially diagnosed with acquired refractory anemia and ringed sideroblasts. J Clin Invest 96:2090, 1995

Acknowledgments

Figures 1 and 6 Talar Agasyan.

Figure 5 Alan D. Iselin.

Figure 7 Tom Moore.

IV HEMOGLOBINOPATHIES AND HEMOLYTIC ANEMIAS

Stanley L. Schrier, m.d.

Alteration of the erythrocyte membrane usually signals the reticuloendothelial macrophages to remove the damaged red blood cell (RBC) from the circulation. In extraordinary circumstances, however, the damage to the membrane is so great that the erythrocyte undergoes hemolysis, and its intracellular contents, including hemoglobin, are liberated into the plasma. This chapter describes structural and functional features of normal erythrocytes and diseases involving membrane architecture, RBC proteins, and extracorpuscular factors that can lead to shortened RBC survival.

Development, Structure, and Physiology of the Erythrocyte

Erythroid precursor cells undergo four or five cell divisions in the bone marrow and then extrude their nuclei and become reticulocytes. As these enucleate cells mature, hemoglobin synthesis decreases. The cells lose most of their transferrin receptors and enter the peripheral blood; they survive in the circulation for about 4 months.

As they move through the circulation, erythrocytes must withstand severe mechanical and metabolic stresses, deform to traverse capillaries with diameters half their own, resist high shearing forces while moving across the cardiac valves, survive repeated episodes of stasis-induced acidemia and substrate depletion, and avoid removal by the macrophages of the reticuloendothelial system. They must also maintain an internal environment that protects hemoglobin from oxidative attack and sustain the optimum concentration of 2,3-bisphosphoglycerate (2,3-BPG) needed for hemoglobin function.

HEMOGLOBIN

The normal adult RBC contains three forms of hemoglobin (Hb): HbA (96%), HbA_2 (2% to 3%), and HbF (< 2%). Normal HbA ($\alpha_2\beta_2$) is composed of two α chains, coded by four genes on chromosome 16, and two β chains, coded on chromosome 11. HbA_2 is composed of two α chains and two δ chains ($\alpha_2\delta_2$), and fetal hemoglobin (HbF) is composed of two α chains and two γ chains ($\alpha_2\gamma_2$). The genes for the β, δ, and γ chains are closely linked to one another on chromosome 11. The extraordinarily high concentration of hemoglobin in the RBC—33 to 35 g/dl (the mean corpuscular hemoglobin concentration, or MCHC)—produces a viscous solution intracellularly.

NONHEMOGLOBIN CYTOSOL

Erythrocytes principally utilize glucose to maintain the reducing power that protects the cell against oxidative attack, to generate the 2,3-BPG required to modulate the function of hemoglobin, and to control the salt and thus the water content of the RBC by the actions of adenosine triphosphate (ATP) and the transport adenosine triphosphatases (ATPases) [*see Table 1*]. The water and the hemoglobin content of the RBC determine the mean corpuscular volume (MCV) and the MCHC.

PLASMA MEMBRANE

The RBC normally has a discoid shape with a diameter of 7 to 8 μm, an MCV of 85 to 90 femtoliters (fl) (1 fl = 10^{-15} L), and a surface area of 140 μm^2 [*see Figure 1*]. Its unique shape enables it to squeeze through capillaries as narrow as 3 μm in diameter.

Lipids (phospholipids and cholesterol) account for 50% of the weight of the surface membrane. The phospholipids are distributed asymmetrically in the membrane bilayer, with positively charged ones in the outer half and relatively negatively charged ones predominantly in the inner half. This asymmetry permits the selective intercalation of small charged molecules into either the outer or inner half of the bilayer, producing echinocytes or stomatocytes [*see Figure 1*].

The RBC membrane proteins include integral and peripheral proteins. Integral proteins interact with and span the hydrophobic phospholipid bilayer [*see Figure 2*]. The major integral proteins of the erythrocyte membrane are the glycophorins (which contain most of the membrane sialic acid and carry the MNSs blood group antigens) and band 3, which is the anion and bicarbonate transporter.

Table 1 Erythrocyte Metabolism

Pathway	*Product*	*Functions of Metabolic Products*
Glycolysis by Embden-Meyerhof pathway	ATP	Serves as a substrate for all kinase reactions, for the ATPase-linked sodium-potassium pump, for the ATPase-linked calcium efflux pump, and for other ATPases of the RBC membrane, including aminophospholipid translocase Maintains deformable state of RBC membrane
	2,3-DPG	Interacts with deoxyhemoglobin, shifting equilibrium to favor unloading of O_2 from oxyhemoglobin Acts as an intracellular anion that cannot cross the RBC membrane
	NADH	Acts as a substrate for a methemoglobin reductase, enabling it to reduce methemoglobin (Fe^{3+}) to hemoglobin (Fe^{2+})
Pentose phosphate pathway (hexose monophosphate shunt)	NADPH	Serves as a substrate for another methemoglobin reductase in methemoglobin reduction (a fail-safe mechanism) Serves as a coenzyme for glutathione reductase in reduction of oxidized glutathione; reduced glutathione (GSH) protects RBC against oxidative denaturation

ATP—adenosine triphosphate 2,3-DPG—2,3-diphosphoglycerate NADH—reduced nicotinamide-adenine dinucleotide NADPH—reduced nicotinamide-adenine dinucleotide phosphate

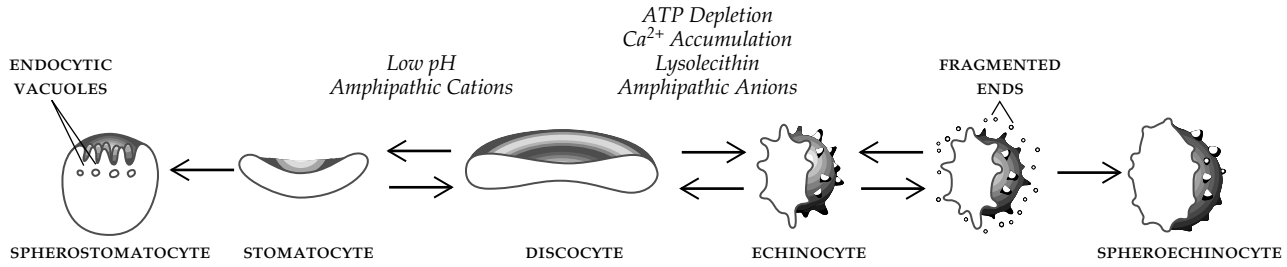


Figure 1 **The normal erythrocyte, or discocyte, undergoes shape changes in response to conditions created by treatment with certain agents. Most changes are reversible if inducing agents are removed before the permanent loss of membrane material.**

The peripheral proteins are all found at the cytosol face of the membrane. The interaction of these peripheral proteins, which include spectrin and actin, results in the tough but resilient cytoskeleton of the erythrocyte. The peripheral cytoskeleton, in turn, is connected to the integral proteins [*see Figure 2*].[1,2]

The membrane carbohydrates contribute to the external negative charge of the membrane and function partly as blood group antigens. Some of these glycolipids associate with phosphatidylinositol to form a glycolipid anchor, called the glycosylphosphatidylinositol (GPI) anchor. These GPI anchors provide the membrane-anchoring site for several classes of proteins that have important biologic functions at membrane surfaces, including several that serve to control complement action [*see* Paroxysmal Nocturnal Hemoglobinuria, *below*].[3]

CONTROL OF HYDRATION AND VOLUME

Control of RBC volume has considerable pathophysiologic importance because the water and cation contents of RBCs determine intracellular viscosity and the ratio of surface area to volume. The Na^+ and K^+ content is determined by passive diffusion and by active transport, primarily through Na^+,K^+-ATPase. The major intracellular anion is Cl^-, which enters the RBC with high permeability through band 3. The K^+-Cl^- cotransporter drives the K^+-Cl^- gradient and is activated by RBC swelling and low intracellular pH, causing a net loss of K^+ and Cl^-. The Ca^{2+}-ATPase actively pumps Ca^{2+} out of the RBC, making the free cytosolic Ca^{2+} content less than 0.1 μM—four orders of magnitude lower than the plasma concentration of 1 mM. The Gardos channel, which is a Ca^{2+}-activated K^+ efflux channel, plays an important role in volume regulation. Water enters and exits through a water channel called CHIP 28 (28 kd channel-forming integral membrane protein) or aquaporin. Other important intracellular anions are 2,3-BPG and hemoglobin, neither of which penetrates the cell membrane. When the concentration of free cytosolic Ca^{2+} rises to levels even as low as 0.3 μM, the channel is activated and results in a net loss of K^+. If such a loss is not corrected, the affected RBC becomes dehydrated.[4]

SHAPE CHANGES

ATP depletion, calcium ion accumulation, or treatment with lysolecithin or with anionic amphipathic compounds transforms the normal erythrocyte, or discocyte, into an echinocyte—a crenated spiculated cell sometimes called a burr cell [*see Figure 1*]. Calcium, acting either alone or in concert with the calcium-binding protein calmodulin, can effect the echinocytic shape change. If the echinocytic process persists, fragmentation or budding of the tips of the echinocyte leads to loss of membrane components, particularly of band 3 and phospholipids. This results in loss of surface area, a reduction in the ratio of surface area to volume, and the formation of poorly deformable spheroechinocytes.

PRINCIPLES OF BLOOD FLOW

The major determinants of blood flow are the hematocrit; the plasma concentration of proteins such as fibrinogen and immunoglobulins, which influence the degree of rouleau formation or aggregation; RBC deformability; the caliber of blood vessels; and the shear rate (the ratio of flow rate to tube radius). At the low shear rates that exist in postcapillary venules, the RBCs tend to clump in asymmetrical masses, with a consequent increase in blood viscosity and resistance to flow.

CELL AGING AND DEATH

In the bone marrow, the developing reticulocyte progressively loses its residual RNA over a 4-day period after nuclear extrusion. At the conclusion of this stage, the reticulocyte can no longer engage in protein synthesis. The active K^+-Cl^- cotransporter functions to reduce cell volume. With the membrane protein assembly complete, the resulting mature cell enters the circulation and survives for a period of 100 to 120 days.[5] Erythrocyte death is an age-dependent phenomenon and may be related to mechanical and chemical stresses the cell encounters in the circulation. As the erythrocyte ages, it loses water and its surface area diminishes. The ratio of surface area to volume decreases and the mean corpuscular hemoglobin concentration increases, impairing cell deformability. In addition, decreased enzymatic activity lowers the cell's ability to withstand metabolic stress. Aging may be manifested by changes at the erythrocyte's surface, such as a decrease in the density or type of surface charge or the appearance of a senescence neoantigen, perhaps oxidatively clustered band 3 [*see Figure 2*], that binds specific immunoglobulins and complement components.[6] By such changes, the age-worn erythrocyte signals its incapacity to the reticuloendothelial system, triggering removal by macrophages.

Under physiologic conditions, slightly less than 1% of the RBCs are destroyed each day and are replaced by a virtually identical number of new cells. For a 70 kg (154 lb) man with a blood volume of about 5 L, about 50 ml of whole blood, containing approximately 22 ml of packed erythrocytes, is destroyed and replaced each day. Inasmuch as one third of each erythrocyte is hemoglobin, the replacement of these cells requires the synthesis of about 7 g of hemoglobin each day. Normal adult bone marrow can readily increase its erythroid output fivefold. After extensive and prolonged anemic stress, erythroid production can be raised by as much as seven or eight times. The supply of iron, however, places an important limit on RBC replacement: three fourths of the iron used in the synthesis of cells in a day comes from cells that were destroyed on the previous day.

General Features of Hemolytic Anemias

The severity of anemia is determined both by the rate of RBC destruction and by the marrow's capacity to increase erythroid production. When a person has a healthy marrow, erythrocyte survival time can be reduced from 120 days to 20 days without inducing anemia or jaundice; however, a substantial reticulocytosis will be present in such cases.

Most forms of hemolysis are extravascular; the damaged cell signals its changed status to the reticuloendothelial system via its membrane and is removed. In unusual circumstances in which damage to the erythrocyte is devastating—as in some forms of complement-mediated lysis—or in circumstances in which the reticuloendothelial system cannot cope with the burden of damaged cells, intravascular lysis develops and leads to hemoglobinemia.

Hemoglobin released to the plasma is degraded to αβ dimers, which bind to haptoglobin. The hemoglobin-haptoglobin complexes are removed by the reticuloendothelial system. When the haptoglobin-binding capacity is exceeded, αβ dimers pass into the glomerular filtrate. Some of the αβ dimers are excreted into the urine directly, producing hemoglobinuria, whereas others are taken up by renal tubule cells. Iron-containing renal tubule cells may be excreted for several days after an episode of intravascular hemolysis. Hemosiderinuria can be identified with Prussian blue stain. Free plasma hemoglobin can dissociate into globin and hemin. Hemin may bind to hemopexin and may reach the renal tubule cells in that form, or it may bind to plasma albumin, producing methemalbuminemia.

Intravascular hemolysis may produce severe anemia acutely. In addition, erythrocytic membrane particles released into the plasma may act as potent stimuli for disseminated intravascular coagulation. Acute severe hemolysis is also a cause of acute renal failure [*see 10:VI Acute Renal Failure*]. When a patient compensating for a marked increase in hemolysis has an infection that sharply impairs marrow erythroid activity,[7] the hemoglobin level may fall dramatically—a condition called aplastic crisis. With chronic hemolysis, pigment stones often develop in the gallbladder.

Causes of hemolysis may be classified as either extracorpuscular or intracorpuscular. The intracorpuscular causes, which are essentially erythrocyte defects, comprise membrane abnormalities, metabolic disturbances, and disorders of hemoglobin structure or biosynthesis. Extracorpuscular causes represent abnormal elements within the vascular bed that attack and destroy normal erythrocytes. Because erythrocytes with intracorpuscular defects that cause hemolysis are intrinsically abnormal, when they are transfused into normal recipients, their survival time is characteristically short. Of the intracorpuscular defects, only one disorder, paroxysmal nocturnal hemoglobinuria, is not hereditary.

Erythrocyte Membrane Defects

DISORDERS OF SALT AND WATER METABOLISM

Hydrocytosis (Hereditary Stomatocytosis)

Hydrocytosis is a hereditary disorder that usually presents early in life as partly compensated hemolytic anemia; occasionally, the spleen is palpable. The MCV is usually elevated. The peripheral smear shows stomatocytes [*see Figure 3*]. Passive flux of both Na^+ and K^+ increases greatly. The Na^+,K^+-ATPase is overwhelmed; the cation concentration and thus the water content of the RBC increase, accounting for the increase in MCV and the decrease in the ratio of surface area to volume. Stomatocytes appear to adhere more avidly than normal RBCs, a finding that may account for the reported increase in thromboembolic events.[8] Perhaps more importantly, the number of RBCs with phosphatidylserine exposed on the outer membrane surface is increased. Phosphatidylserine—a relatively negatively charged

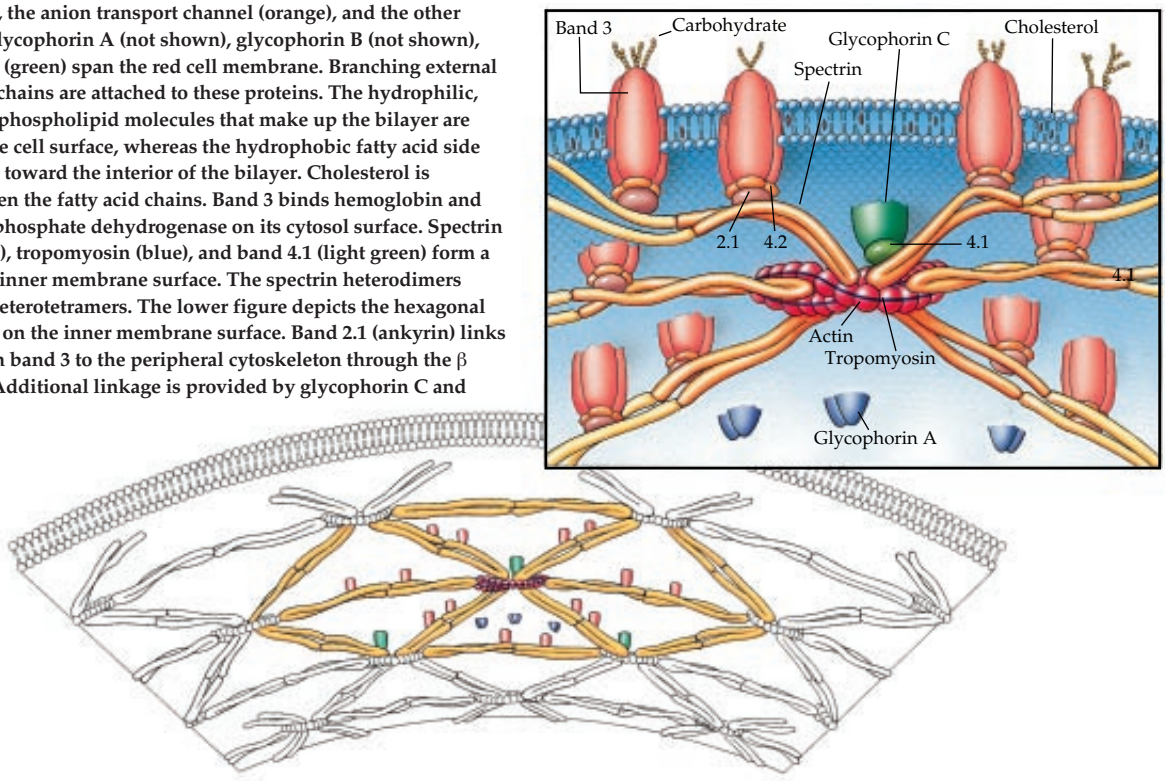


Figure 2 **Band 3, the anion transport channel (orange), and the other integral proteins glycophorin A (not shown), glycophorin B (not shown), and glycophorin C (green) span the red cell membrane. Branching external carbohydrate side chains are attached to these proteins. The hydrophilic, polar heads of the phospholipid molecules that make up the bilayer are oriented toward the cell surface, whereas the hydrophobic fatty acid side chains are directed toward the interior of the bilayer. Cholesterol is intercalated between the fatty acid chains. Band 3 binds hemoglobin and glyceraldehyde-3-phosphate dehydrogenase on its cytosol surface. Spectrin (yellow), actin (red), tropomyosin (blue), and band 4.1 (light green) form a latticework on the inner membrane surface. The spectrin heterodimers associate to form heterotetramers. The lower figure depicts the hexagonal cytoskeletal lattice on the inner membrane surface. Band 2.1 (ankyrin) links the integral protein band 3 to the peripheral cytoskeleton through the β chain of spectrin. Additional linkage is provided by glycophorin C and band 4.1.**

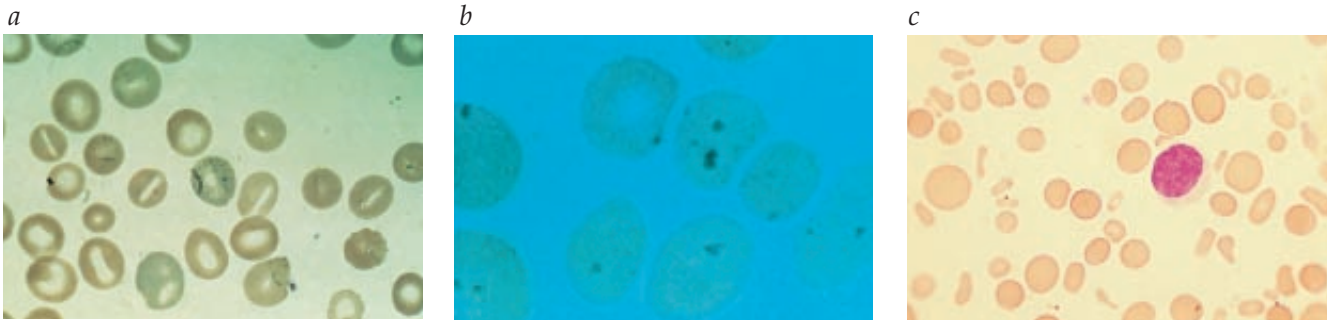


Figure 3 **Stomatocytes are identified by slitlike areas of central pallor (*a*); the smear also shows microspherocytes, which are a more advanced stage of stomatocytosis. On scanning electron microscopy or examination of wet preparations, the microspherocytes are shown to be stomatocytes. Microspherocytes are seen in hereditary spherocytosis and in autoimmune hemolytic anemia, as well as in other conditions characterized by relatively selective loss of membrane material or increase in cell volume. Supravital stain of erythrocytes (*b*) shows single and multiple blue-staining Heinz bodies within counterstained erythrocytes. Phase microscopy can also be used to demonstrate Heinz bodies. Elliptocytes are visualized in a smear from a patient with hereditary elliptocytosis (*c*).**

phospholipid that is normally found predominantly in the inner membrane layer—provides a nidus for thrombin formation and thus may also contribute to the tendency to thrombosis.[9] Splenectomy may lead to improvement in the anemia. Other therapies may eventually prove useful; vaso-occlusive events were controlled in one patient by long-term RBC transfusion and in another by therapy with pentoxifylline.[8]

Xerocytosis

Xerocytosis, another hereditary hemolytic disorder, is characterized by a membrane defect that leads to loss of cations, particularly K^+. Dehydration of erythrocytes occurs because the K^+ leak exceeds the Na^+ influx, possibly as a result of an overactive K^+-Cl^- cotransporter. Patients present with variably compensated hemolysis. Splenomegaly is not a prominent feature. The peripheral smear is variable, showing target cells, stomatocytes, echinocytes, or so-called hemoglobin puddling (i.e., hemoglobin collected around the circumference of the cell). MCHC is increased. Because these rigid cells are removed in many parts of the reticuloendothelial system, splenectomy is of little benefit.[10] In rare instances, xerocytosis can cause nonimmune hydrops fetalis.[11]

PROTEIN ABNORMALITIES

Hereditary Elliptocytosis

There are perhaps 250 to 500 cases of hereditary elliptocytosis per million population.[10] Three morphologic variants are seen: (1) common hereditary elliptocytosis, (2) spherocytic hereditary elliptocytosis, and (3) stomatocytic hereditary elliptocytosis.[12] Most patients with common hereditary elliptocytosis are heterozygous for this autosomal dominant disorder and have only elliptical RBCs or, at worst, compensated hemolysis. Homozygotes for the disorder may have severe uncompensated hemolytic anemia.

Under applied shear stress, erythrocytes assume an elliptical shape; when the stress is removed, the cell normally recoils to its discoid shape. It has been hypothesized that membrane defects in hereditary elliptocytosis interfere with normal recoil. The membrane defect appears to be a lesion in the membrane cytoskeleton; RBC membranes from patients with hereditary elliptocytosis are almost invariably mechanically fragile.

The diagnosis is made in patients with extravascular intracorpuscular hemolysis who have elliptocytes on the peripheral smear. Elliptocytosis can also be seen in severe iron deficiency, myeloproliferative and myelodysplastic disorders, and, occasionally, cobalamin and folate deficiencies.[12] Results of the osmotic fragility test are usually normal. Splenectomy has been useful in patients with severe common hereditary elliptocytosis.

Hereditary Propoikilocytosis

The syndrome of hereditary (autosomal recessive) pyropoikilocytosis, a variant of hereditary elliptocytosis, causes severe hemolysis in young children. It is caused by an abnormal α or β spectrin mutation. The blood smear shows extreme microcytosis and extraordinary variation in the size and shape of erythrocytes [*see Figure 3*]. Splenectomy may reduce the rate of hemolysis.

Hereditary Spherocytosis

Hereditary spherocytosis is usually inherited as an autosomal dominant trait and affects about 220 per million people worldwide. A rare autosomal recessive variant of hereditary spherocytosis has been described.

Because of a loss of surface membrane, RBCs assume a microspherocytic shape and thus cannot deform sufficiently to pass through the splenic vasculature; splenic trapping of RBCs, hemolysis, and a compensatory increase in RBC production result. The underlying membrane defects lead to budding of membrane vesicles under conditions of metabolic depletion. These membrane vesicles are enriched in phospholipids from the bilayer, as well as in associated transmembrane proteins [*see Figure 2*]. The underlying molecular lesions appear to consist of deficiencies of spectrin, spectrin-ankyrin, band 3, and band 4.2 (palladin).[12,13]

About 25% of patients with hereditary spherocytosis have completely compensated hemolysis without anemia; their disorder is diagnosed only when a concomitant condition, such as infection or pregnancy, increases the rate of hemolysis or reduces the marrow's compensatory capacity. In other patients, mild anemia, pigmented gallstones, leg ulcers, and splenic rupture may develop. Aplastic crises may be precipitated by ordinary respiratory tract infections, especially by parvovirus infection.[7] It is important to remember that this disease can become apparent during the first year of life, when increased splenic maturation resulting in RBC removal combined with a sluggish erythropoietic response can result in anemia severe enough to require RBC transfusion.[14]

This diagnosis is suggested by a predominance of microspherocytes on the peripheral smear [*see Figure 3b*], an MCHC of 35 g/dl or greater, reticulocytosis, mild jaundice, splenomegaly,

and a positive family history, although at least half of newly diagnosed patients have no family history. Confirmation of the diagnosis is made by a 24-hour incubated osmotic fragility test. A negative Coombs test and a family history positive for hereditary spherocytosis rule against a diagnosis of acquired autoimmune hemolytic anemia. Splenectomy eradicates clinical manifestations of the disorder, including aplastic crises.

PAROXYSMAL NOCTURNAL HEMOGLOBINURIA

Paroxysmal nocturnal hemoglobinuria (PNH) is a somatic clonal disorder of hematopoietic stem cells. PNH involves the *PIG-A* gene, which maps to the short arm of the X chromosome.[15] The mutation results in a deficiency of the membrane-anchoring protein phosphatidylinositol glycan class A; the resulting mature hematopoietic cells are usually chimeric. Normal human erythrocytes, and probably platelets and neutrophils, modulate complement attack by at least three GPI membrane-bound proteins: DAF (CD55), C8-binding protein (C8BP), and MIRL (CD59). In the absence of the GPI anchor, all of the proteins that use this membrane anchor will be variably deficient in the blood cells of persons with PNH.[16] Because the defective synthesis of GPI affects all hematopoietic cells, patients with PNH may have variable degrees of anemia, neutropenia, or thrombocytopenia, or they may have complete bone marrow failure.[17]

Diagnosis

Classically, acute episodes of intravascular hemolysis are superimposed on a background of chronic hemolysis. The patient typically notes hemoglobinuria on voiding after sleep.[18,19] Recurrent venous occlusions lead to pulmonary embolism and hepatic and mesenteric vein thrombosis, possibly resulting from release of procoagulant microparticles derived from platelets.[20] A literature review found that thrombotic events accounted for 22% of deaths in patients with PNH.[21] Occasionally, PNH patients with thrombosis are mistakenly thought to have psychosomatic disorders because they complain of recurrent severe pain in the abdomen and back that has no obvious cause. In these cases, the associated anemia and hemolysis may be very mild, and episodes of hemolysis do not necessarily correlate with bouts of pain.

A diagnosis of PNH should be considered in any patient with chronic or episodic hemolysis. The diagnosis should also be considered for patients with recurrent venous thromboembolism, particularly if the thrombus occurs in a site such as the inferior vena cava or the portal mesenteric system or if it produces Budd-Chiari syndrome. Evidence of intravascular hemolysis, such as hemoglobinemia; reduced serum haptoglobin; increased serum methemalbumin; hemoglobinuria; or hemosiderinuria, suggests the diagnosis. The combination of marrow hypoplasia and hemolysis is an important clue. PNH may occur in association with aplastic anemia. Erythrocyte morphology is usually normal. Diagnosis is made by specific tests based on fluorescence-activated cell sorter analysis using antibodies that quantitatively assess DAF (CD55) and particularly MIRL (CD59) on the erythrocyte or on the leukocyte surface.[22]

Treatment

In PNH, the anemia is occasionally so severe (hemoglobin level < 8 g/dl) that the patient needs transfusions regularly[19]; therefore, the choice of transfusion component is critical. It is believed that infusion of blood products containing complement may enhance hemolysis. Infusion of donor white blood cells (WBCs), which are ordinarily present in a unit of packed RBCs, into an HLA-immunized recipient may provide the antigen-antibody reaction that activates complement by the classical pathway. In such a case, the use of special leukocyte-poor units may be helpful [*see 5:X Transfusion Therapy*].

A trial of prednisone (e.g., 60 mg a day with rapid tapering, or 20 to 60 mg every other day) may reduce transfusion requirements and may be helpful in alleviating the anemia. Splenectomy is of very questionable benefit. Surgery is risky in patients with PNH because stasis and trauma accentuate hemolysis and venous occlusion. If surgery is to be performed, prophylactic anticoagulation with warfarin in the perioperative period should be considered.

Patients with PNH are frequently iron deficient. The simple administration of iron to correct this defect, however, often aggravates hemolysis because iron therapy produces a cohort of new cells, many of which are susceptible to complement-mediated lysis. Transfusion before iron therapy may help circumvent this problem because it will decrease the erythropoietic stimulus to the marrow.

Thrombocytopenia resulting from poor platelet production may necessitate platelet transfusions [*see 5:X Transfusion Therapy*].[18] Budd-Chiari syndrome and inferior vena cava thrombosis must be diagnosed and treated quickly with heparin, followed by long-term administration of warfarin. If heparinization is ineffective, thrombolytic therapy (e.g., streptokinase) may be used.[23] Children and adolescents with PNH that is complicated by aplastic anemia should be considered for allogeneic bone marrow transplantation.[19,24] In case reports, the anemia associated with PNH responded to erythropoietin,[25] and four patients with severe neutropenia and thrombocytopenia responded to combinations of granulocyte–colony-stimulating factor (G-CSF) and cyclosporine.[26]

Prognosis

A study of 80 patients with PNH indicated that median survival was 10 years.[18] The causes of PNH-related death were thrombocytopenia, PNH hemolysis, thromboses, or PNH-associated aplastic anemia [*see 5:III Anemia: Production Defects*]. Of interest is that 15% of patients experienced spontaneous remission.[18] In rare instances, prolonged and severe iron loss may occur as a result of chronic hemosiderinuria, producing iron deficiency; some patients develop transfusion-associated hemochromatosis.[19]

Acute myeloid leukemia may develop during the course of PNH. In one series, this occurred in three of 80 patients; in another series, of 220 patients, the incidence of myelodysplastic syndromes was 5% and the incidence of acute leukemia was 1%.[19]

Abnormalities of Erythrocyte Metabolism

DEFECTIVE REDUCING POWER

The reducing power of the erythrocyte is provided by reduced glutathione (GSH) and the reduced coenzymes nicotinamide adenine dinucleotide (NADH) and nicotinamide-adenine dinucleotide phosphate (NADPH) [*see Table 1*]. When erythrocytic stores of these materials are inadequate, hemoglobin and membrane-associated proteins can be oxidized, leading to the production of Heinz bodies, which consist predominantly of oxidative degradation products of hemoglobin [*see Figure 3b*]. Erythrocytes containing Heinz bodies are rigid and are therefore selectively removed by the reticuloendothelial system.

Defective Glutathione Synthesis

Deficiencies of certain enzymes involved in GSH synthesis lead to oxidative attacks on erythrocytes and to hemolysis. Several reports have described families whose members show almost negligible GSH synthesis and have hemolysis associated with the production of Heinz bodies. Glutathione peroxidase deficiency apparently contributes to hemolysis in newborn infants.

GLUCOSE-6-PHOSPHATE DEHYDROGENASE DEFICIENCY

Glucose-6-phosphate dehydrogenase (G6PD) is the first enzyme in the pentose phosphate pathway, or hexose monophosphate shunt. It catalyzes the conversion of $NADP^+$ to NADPH, a powerful reducing agent. NADPH is a cofactor for glutathione reductase and thus plays a role in protecting the cell against oxidative attack. RBCs deficient in G6PD are therefore susceptible to oxidation and hemolysis.[27,28]

G6PD deficiency is one of the most common disorders in the world; approximately 10% of male blacks in the United States are affected, as are large numbers of black Africans and some inhabitants of the Mediterranean littoral. This disorder confers some selective advantage against endemic malaria. For example, in a study in Ghana on pregnant women (who are highly susceptible to falciparum malaria and its consequences), the prevalence of infection was 66% in normal women, 58% in G6PD heterozygotes, and 50% in homozygotes.[29]

The gene for G6PD is on the X chromosome at band q28; males carry only one gene for this enzyme, so those males that are affected by the disorder are hemizygous. Females are affected much less frequently because they would have to carry two defective G6PD genes to show clinical disease of the same severity as that in males. However, expression of a defective G6PD gene is not completely masked in heterozygous women; in fact, such women exhibit highly variable G6PD enzyme activity. According to the X-inactivation, or Lyon-Beutler, hypothesis,[28] females heterozygous for G6PD have two cell lines: one that contains an active X chromosome with a gene for normal G6PD and another that contains an active X chromosome with a gene for deficient G6PD. Chance partly determines the relative proportions of the two cell lines, which in turn control the clinical severity of the defect.

Classification

There are three clinical classes of G6PD deficiency: class I, which is the uncommon chronic congenital nonspherocytic hemolytic anemia; class II, in which the enzyme deficiency is severe but hemolysis tends to be episodic; and class III, the most common variant, in which the enzyme deficiency is moderate and hemolysis is caused by oxidant attack. The severity of the hemolysis and the anemia is directly related to the magnitude of the enzyme deficiency, which is determined by the half-life of the enzyme. The normal G6PD half-life is 62 days; in class III G6PD deficiency, the enzyme has a half-life of 13 days; and in class II deficiency, G6PD has a half-life of several hours. The cloning and sequencing of the G6PD gene have clarified the classification of G6PD deficiency; before the sequencing of the G6PD gene, more than 300 variants of G6PD deficiency had been described.[28]

Etiology

Hemolysis occurs in persons with class III G6PD deficiency after exposure to a drug or substance that produces an oxidant stress. Ingestion of, or exposure to, fava beans may cause a devastating intravascular hemolysis (known as favism) in G6PD-deficient patients, but it usually occurs only in those with the Mediterranean variant of class II deficiency. Fava beans contain isouramil and divicine, two strong reducing agents whose actions eventuate in the oxidation of membrane proteins. This produces a rigid cell in which hemoglobin is confined to one part of the cytosol; the other part of the cytosol appears as a clear ghost (i.e., the classic bite, hemiblister, or cross-bonded cell) [*see Figure 4*]. These membrane defects cause extravascular and intravascular hemolysis.[27] Severe infections, diabetic ketoacidosis, and renal failure also reportedly trigger hemolysis.

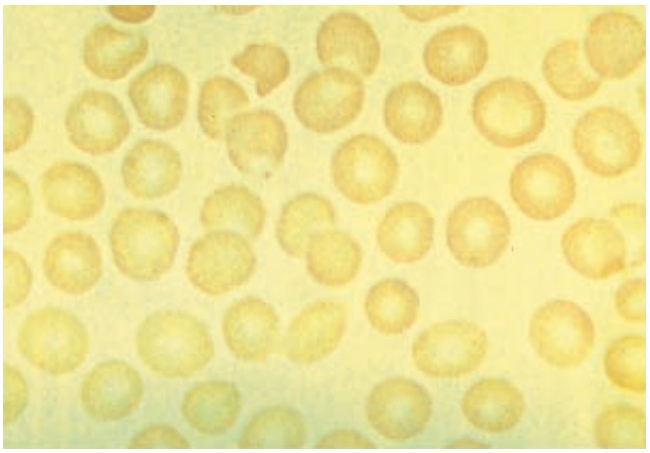


Figure 4 **Bite, hemiblister, or cross-bonded cells are indicative of oxidative attack leading to oxidative hemolysis.**

Diagnosis

Hemolytic anemia characterized by the appearance of bite cells and Heinz bodies after administration of certain drugs suggests the possibility of G6PD deficiency [*see Table 2*]. Dapsone, which is capable of inducing oxidant-type hemolysis, has increasingly come into use as prophylaxis for *Pneumocystis carinii* pneumonia in patients infected with HIV [*see 7:XXXIII HIV and AIDS*]. Therefore, it is important to screen potential users of dapsone for G6PD deficiency with the standard enzymatic tests. Other agents with oxidative potential, such as amyl nitrite ("poppers"), can cause hemolysis.[30]

Other disorders to be considered in the differential diagnosis of oxidative hemolysis include unstable hemoglobinopathy, he-

Table 2 Drugs That Produce Hemolysis in G6PD-Deficient Patients

Class	*Example*
Antimalarials	Primaquine Chloroquine
Sulfonamides	Sulfamethoxazole Sulfapyridine
Sulfones	Dapsone
Analgesics	Acetanilid Phenacetin Acetylsalicylic acid (10 g/day)
Nitrofurans	Nitrofurantoin Furazolidone
Water-soluble vitamin K derivatives	Menadiol

moglobin M disease, and deficiencies of other enzymes essential to glutathione metabolism. A G6PD screening test or direct enzyme assay usually resolves the question. Patients with A-type G6PD (class III) deficiency and brisk reticulocytosis, however, may have a near-normal G6PD level because young RBCs have relatively high G6PD levels. In such cases, it is best to repeat the tests when the reticulocyte count returns to normal. Information on genetic testing for G6PD deficiency can be found on the Internet at http://www.geneclinics.org.

Treatment

Avoidance of drugs that may produce hemolysis is critical in management. Acute favism requires circulatory support, maintenance of good renal blood flow, and transfusions with erythrocytes that are not G6PD deficient. The physician must also be alert to the possible onset of disseminated intravascular coagulation.

DEFECTS IN GLYCOLYSIS

The series of reactions constituting the glycolytic pathway generates several products, such as ATP, that have various essential functions in erythrocyte metabolism [*see Table 1*]. Defects involve the major glycolytic pathway (Embden-Meyerhof pathway) and generally interfere with ATP production.

Pyruvate kinase (PK) catalyzes the formation of pyruvate, a reaction associated with ATP synthesis. After G6PD deficiency, PK deficiency (autosomal recessive) is the second most common hereditary enzymopathy. Hemolysis, mild jaundice, and, occasionally, palpable splenomegaly are the presenting problems. The peripheral smear usually reveals normal RBCs, but in a few cases, the RBCs show extreme spiculation. Aplastic crises may occur.[7]

Congenital nonspherocytic hemolysis raises the possibility of PK deficiency. An enzyme assay establishes the diagnosis. Splenectomy should be considered for patients who require transfusions.

Glucose-6-phosphate isomerase deficiency is the third most common enzymopathy that leads to hemolysis. Other enzymopathies are quite rare. Screening tests and specific assays are available for deficiencies of such enzymes as hexokinase, phosphofructokinase, triose phosphate isomerase, phosphoglycerate kinase, and aldolase.

DEFECTS IN NUCLEOTIDE METABOLISM

In hemolytic anemia associated with pyrimidine 5′-nucleotidase deficiency, coarse basophilic stippling persists in mature erythrocytes, presumably because the enzyme deficiency prevents degradation of reticulocyte RNA. This accumulation results in expansion of the total RBC nucleotide pool to a level five times normal. Pyrimidine nucleotides accumulate, and adenine nucleotides are decreased. Glycolysis is impaired by an undetermined mechanism.

Disorders Involving Hemoglobin

CLASSIFICATION OF THE HEMOGLOBINOPATHIES

The clinically important hemoglobinopathies are classified into five categories on the basis of the underlying defect. The defects are as follows:

1. Hemoglobin tends to gel or crystallize (e.g., sickle cell anemia or hemoglobin C disease).
2. Hemoglobin is unstable (e.g., the congenital Heinz body anemias).
3. Hemoglobin has abnormal oxygen-binding properties (e.g., the disorder caused by hemoglobin Chesapeake).
4. Hemoglobin is readily oxidized to methemoglobin (e.g., methemoglobinemia).
5. Hemoglobin chains are synthesized at unequal rates (e.g., the thalassemias).

SICKLE CELL DISEASE

Sickle Cell Anemia

Definition Sickle cell anemia is an autosomal recessive disease caused by the substitution of the amino acid valine for glutamine at the sixth position of the β-hemoglobin chain, which results in the production of HbS.

Epidemiology From 8% to 10% of African Americans and a lesser percentage of persons with eastern Mediterranean, Indian, or Saudi Arabian ancestry have the sickle (HbS) gene. Disease develops in persons who are homozygous for the sickle gene (*HbSS*), in whom 70% to 98% of hemoglobin is of the S type. About 0.2% of African Americans have sickle cell anemia. The fact that the sickle gene occurs in populations living in regions endemic for falciparum malaria suggests that sickle heterozygosity confers a protective advantage against malaria.[31]

Restriction endonuclease analyses indicate that the sickle gene mutation probably arose spontaneously in at least five geographic locations. These variations are called Senegal, Benin, Central African Republic (or Bantu), Saudi-Asian, Cameroon, and Indian (which may be the same as the Saudi-Asian variant). These variants are important clinically because some variants are associated with higher output of γ-globin chains (and thus higher HbF levels); others are associated more often with the gene for α-thalassemia-2 [*see* The Thalassemias, *below*]. Either of these associations may alleviate some aspects of the sickling process.[31]

Pathophysiology Two major clinical features characterize sickle cell anemia: (1) chronic hemolysis and (2) acute, episodic vaso-occlusive crises that cause organ failure and account for most of the morbidity and mortality associated with the disease.

HbS liganded to oxygen or carbon monoxide shows near-normal solubility. When the molecule gives up its oxygen and changes to the deoxy S form, however, its solubility decreases. In an environment with reduced oxygen, HbS polymerizes into long tubelike fibers that induce erythrocytic sickling.[32]

The deoxyhemoglobin S polymer is in equilibrium with surrounding soluble molecules of deoxyhemoglobin S. An increase in the concentration of HbS, a decrease in pH, or an increase in the concentration of 2,3-BPG tends to stabilize the deoxy S form and enhances gelation.[32] In addition, sickled erythrocytes retain the K^+-Cl^- cotransport function and have sufficient intracellular calcium to activate the Gardos efflux channel[33] [*see* Control of Hydration and Volume, *above*]. These two mechanisms act together to produce a population of very dense sickled erythrocytes with MCHCs ranging up to 50 g/dl.[33] HbF inhibits polymerization,[33] so patients with high HbF values, such as those with the Saudi-Asian variant of sickle cell anemia, have milder disease.[31] When hypoxemia and the MCHC reach a critical level, polymerization occurs after a variable delay[33]; this delay represents the pe-

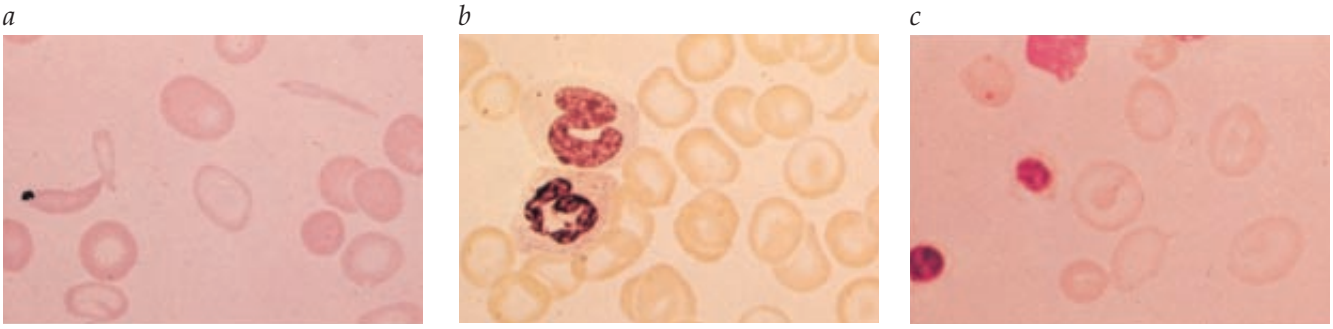


Figure 5 **Sickle cell anemia is characterized by markedly distorted sickle cells, including elongated forms (*a*). Target cells (*b*) are seen in a variety of conditions, including hypochromia caused by iron deficiency, hemoglobinopathies such as HbC variants and the thalassemias, and liver disease. Cooley anemia (*c*), or β-thalassemia major, is indicated by profound hypochromia, targeting, variation in size and shape of erythrocytes, and the presence of nucleated red cells.**

riod during which the deoxyhemoglobin S tetramers are slowly associating to form a nucleus. When the nucleus reaches a critical size, rapid, almost explosive gelation occurs. Free deoxyhemoglobin S tetramers rapidly attach to the nucleus to produce the long tubelike fibers that align to form parallel tubelike structures that distort the cell and produce the sickle shape [*see Figure 5*].

Most cells in the venous circulation are not sickled. However, sickling will occur if the time to polymerization is shortened to less than 1 second or if RBCs become trapped in the microcirculation. Some RBCs contain polymerized sickle hemoglobin even in the arterial circulation. Another manifestation of membrane damage in sickle cells is the irreversibly sickled cell, which retains its sickle shape even when reoxygenated.[34] Some of these poorly deformable RBCs are directly derived from a subpopulation of reticulocytes that are low in HbF[30] and are removed predominantly in the reticuloendothelial system. The rapid removal of these young cells, as well as older, dense, rigid cells that cannot traverse the monocyte-macrophage system, results in chronic extravascular hemolysis.

Because of the extreme sensitivity of sickling to the local environment, attention has been focused on cellular factors. The extreme hyperosmolality of the renal medulla (1,200 mOsm) dehydrates RBCs and raises the MCHC. Consequently, sickling sufficient to abolish the renal medullary concentrating ability may occur even in patients who have only the sickle trait.

Sickle Crisis and Ischemic Infarction

Sickle crisis is a potentially life-threatening vaso-occlusive complication of sickle disease. The initiating event in the sickle crisis is not known, nor is it clear why some patients have severe crises and others do not.

Clusters of increasingly rigid sickle cells will occlude the microvasculature in the followng circumstances: (1) the pH falls, deoxygenation increases, or the MCHC rises; (2) nitric oxide production decreases or nitric oxide is trapped and removed by free hemoglobin in plasma[35]; (3) microvascular disease is present; or (4) capillary transit time is prolonged. Thrombosis may also play a role in sickle occlusion. There is some disorganization of the membrane phospholipid bilayer, with phosphatidylserine moving to the outer leaflet, possibly enhancing the thromboembolic manifestations of sickle disease.[36] In sickle cell anemia, there also appears to be an increase in circulating endothelial cells, which abnormally express tissue factor and may provide an additional basis for thromboembolism.[37]

Blockage leads to ischemic infarction, the release of inflammatory cytokines, and an amplifying sequence of stasis-induced occlusion, which may progress to sickle crisis. Portal circulations in which oxygen tension is low, such as those in the liver or the kidney, are at particular risk for occlusion.

Risk factors predisposing to painful crises include a hemoglobin level greater than 8.5 g/dl, pregnancy, cold weather, and a high reticulocyte count. Nocturnal hypoxemia is an important risk factor in children.[38] Conversely, the low hematocrit in sickle cell anemia reduces blood viscosity and is protective. Sickle cell patients also characteristically have a high plasma fibrinogen level, which enhances the aggregation of already rigid erythrocytes and increases viscosity, particularly at the low shear rates encountered in the microcirculation.[39] Sickled RBCs also have a greater tendency to adhere to endothelial cells than do normal RBCs.[40] The role of leukocytes in this adhesion process is becoming clearer. Adminstration of G-CSF has led to sickle crises and even death.[41,42] Granulocyte-macrophage CSF (GM-CSF) has caused similar crises. The severity of sickle disease appears to parallel the level of the WBC count, and WBC cell-adhesion molecules seem to be critical to sickle vaso-occlusion.[43,44]

Diagnosis of Sickle Cell Disease

In the past, the diagnosis of sickle cell anemia was usually made on the basis of clinical manifestations occurring in childhood; the affected child was seen to have limitation in exercise tolerance, shortness of breath, tachycardia, frequent severe infections, and episodes of very painful dactylitis. Currently, many cases are identified on screening tests, which may be prompted by the diagnosis in a family member or performed as a routine neonatal procedure; in California and many other states, every fetal cord blood sample is examined by high-performance liquid chromatography (HPLC). Rarely, the disorder is diagnosed in adult life, occasionally during a first pregnancy, when prenatal screening reveals anemia. The general symptoms are limited exercise tolerance, exertional dyspnea, painful crises, bouts of jaundice, and even biliary colic.

The clinical appearance of the patient and a blood smear showing sickled cells, holly leaf cells, and erythrocytes with Howell-Jolly bodies are fairly suggestive of sickle cell anemia. Howell-Jolly bodies represent cytoplasmic remnants of nuclear chromatin that are normally removed by the spleen. Platelet and WBC counts are usually high. Unless an aplastic crisis is in

progress, causing a virtual absence of normoblasts, the marrow shows erythroid hyperplasia. Diagnosis is confirmed by performing a sickle preparation: a drop of blood is incubated with fresh 2% sodium metabisulfite, and the proportion of sickle cells is measured immediately and then 1 hour later. Commercial testing sets such as Sickledex rely on the relative insolubility of HbS in 1.0 M phosphate buffers to make the diagnosis. The most definitive tests for sickle cell anemia, however, are hemoglobin electrophoresis or HPLC, which indicate the relative percentages of HbS and HbF. All of these tests are also useful in screening family members for sickle cell trait. Patients who are heterozygous for both the HbS gene and the β-thalassemia gene may appear to be homozygous for HbS. Other varieties of sickling hemoglobin are observed very infrequently. DNA-based methods can also be used to pinpoint the specific genetic abnormality and to identify the subpopulations from which the patient descended[31]; further description and information on diagnostic testing is available online at http://www.geneclinics.org. Persons with sickle cell anemia and α-thalassemia have higher hemoglobin levels, lower reticulocyte counts, a lower MCHC, a lower MCV, and less-dense RBCs than persons who have sickle cell anemia alone. Such patients may have increased life expectancy and perhaps a different pattern of manifestations of veno-occlusive complications.[45] The combination of G6PD deficiency and sickle cell anemia has neither beneficial nor harmful effects.[46,47]

Management of Sickle Cell Disease

Sickle crisis Standard conservative management of sickle crisis centers on rest, hydration, and analgesia. In demonstrably acidotic patients, mild alkalinization should be induced by administration of a bicarbonate solution, which is prepared by addition of an ampule of sodium bicarbonate to 1 L of either 5% dextrose in water or half-normal saline. The bicarbonate solution should be infused at a rate of 5 to 7 ml/kg/hr for the first 4 hours and at 4 ml/kg/hr for the next 20 hours. The role of supplemental oxygen in patients with normal arterial oxygen tension (P_aO_2) and no cardiopulmonary problems is untested.

Pain management Pain [*see 11:XIV Pain*] is the major concern for 10% to 20% of patients with sickle cell anemia. Avascular necrosis of bone marrow produces excruciating pain that can last as long as 8 to 10 days. The need for pain relief sometimes results in habituation or addiction. Because there are few objective ways to monitor the sickle crisis, the physician may not know whether a demand for narcotics is a manifestation of drug-seeking behavior.

The patient who has sickle cell anemia should be provided with oral analgesics for use at home in an attempt to abort the pain crisis at its onset. Nonsteroidal anti-inflammatory drugs (NSAIDs), such as naproxen (500 mg) and ketorolac (10 mg), can be used initially. If NSAIDs alone are not sufficient, they can be followed by a narcotic-analgesic combination, such as hydrocodone and acetaminophen or oxycodone and aspirin. Adjuvants such as oral diphenhydramine (50 mg) or lorazepam (1 to 2 mg) may calm the patient and perhaps antagonize the actions of released histamine.[48] If these measures, perhaps repeated every 6 hours, do not control the pain, the patient usually requires parenteral treatment. Care from the patient's regular physicians is far preferable to reliance on unfamiliar providers in emergency departments.[48] The patient needs rapid evaluation for possible infection, acute chest syndrome, bone infarction, and other complications, and the pain should be treated either with 10 mg of intravenous morphine along with 50 mg of intramuscular diphenhydramine every 2 hours or with 4 mg of intramuscular hydromorphone along with 50 mg of intramuscular diphenhydramine every 2 hours. If there is no pain relief or inadequate pain relief 30 minutes after the first dose, 50% of the initial dose of opiates can be administered; the respiratory rate should be monitored closely, particularly if it approaches 10 respirations a minute. Some units have used patient-controlled analgesia with good results. It is important to continue to administer parenteral analgesia at regular intervals and to provide increased doses for breakthrough pain. The patient will probably need a laxative and may need an antiemetic, such as prochlorperazine (10 mg p.o. or I.M.). If the patient responds, home therapy with oral controlled-release morphine, such as MS Contin, is usually effective. If pain continues for more than 8 to 12 hours, the patient will probably need to be hospitalized to receive extended therapy with increased doses of analgesia and parenteral fluids, along with observation.[48]

Alteration of sickle cell pathophysiology A clearer understanding of the kinetics of sickling suggests some future prospects for the therapy of sickle cell anemia. Decreasing the MCHC should diminish gelation. An approach that attempts to block the Ca^{2+}-dependent K^+ efflux (Gardos channel) [*see* Control of Hydration and Volume, *above*] has been tested in a sickle mouse model and shows promise in preventing RBC dehydration.[49,50]

Therapies to interfere with sickling are being actively pursued. The presence of 20% to 30% HbF in sickle RBCs markedly delays gelation, so a mechanism that would switch on the genes that control fetal hemoglobin synthesis and thus lessen the severity of sickle disease appears feasible.[51,52] Hydroxyurea produces an increase in F reticulocyte and HbF levels. In a phase III trial, patients treated with hydroxyurea (starting dosage, 15 mg/kg/day) had fewer painful crises, admissions for crisis, and episodes of acute chest syndrome, as well as required fewer transfusions, than patients given a placebo.[53] There was no effect on stroke; however, after 8 years of follow-up, mortality was reduced by 40%.[54] The beneficial effect of hydroxyurea accrued after about 8 weeks of therapy and was accompanied by an increase in MCV and an increase in the proportion of F cells; in addition, there was a decrease in neutrophils and a decrease in sickle RBC adhesion to endothelial cells.[55] Trials are also being conducted with butyrate, which can increase γ-chain production, thereby increasing HbF levels and interfering with gelation.[56,57] Demethylating agents such as 5-azacytidine and decitabine can also increase HbF to therapeutically useful levels. Because sickle cells adhere abnormally to the endothelium, attempts have been made to block adhesion; thus far, these efforts have not proved useful.

Inflammatory cytokines appear to play an important role in the sickle crisis, as evidenced by the fact that a predictor of success in hydroxyurea therapy is a decrease in the WBC count.[54,55] Other investigators are studying the possible vasodilatory role of nitric oxide.

Sibling-donor allogeneic bone marrow transplantation can result in cure or can lead to a substitution of sickle trait for sickle cell anemia. Bone marrow transplantation resulted in apparent cure in 15 of 22 carefully selected patients; there were two deaths (9%), and the remaining five patients had complications such as graft failure. Of the 22 patients, 12 had a history of stroke, five had a history of recurrent episodes of acute chest syndrome, and five had recurrent painful crises.[58]

Long-term transfusion therapy Long-term transfusion therapy has been found to prevent stroke.[59] Some investigators have shown that preventive transfusions reduce or eliminate pain crisis, episodes of acute chest syndrome, bacterial infection, and hospitalization.[60,61] Other authors, however, warn against the dangers of iron overload,[62,63] transfusion hepatitis, problems with venous access, and RBC alloimmunization.[64] Further studies may clarify the role of long-term transfusion therapy.

Complications and Their Management

Skeletal problems Aseptic necrosis (osteonecrosis) of the femoral head occurs in about 10% of patients, particularly those who also have α-thalassemia. Arthroplasty has been relatively ineffective, partly because of the presence of adjacent hard bone, which interferes with the placement of the prosthesis, and because of the increased risk of infection.[65]

Cardiopulmonary problems Cardiac complications associated with anemia are the result of a large increase in cardiac output. Such complications include chamber enlargement, cardiomegaly, left ventricular hypertrophy, and flow murmurs.[66] Acute myocardial infarction has occurred in relatively young adults who do not have coronary disease.[67] The incidence of pulmonary hypertension is unknown, but its presence markedly shortens survival.[68]

Acute pulmonary complications are a major cause of morbidity and mortality; such complications include local infection, vascular occlusions in the pulmonary vessels (both in situ thrombosis and embolism), and pulmonary fat embolism from ischemic marrow fat necrosis.[69] A large study of acute chest syndrome found that adult patients were afebrile but had shortness of breath, chills, and pain in the chest and in at least one extremity.[70] Infarctions of the thoracic vertebrae contribute substantially to the pain.[64] Physical examination frequently shows no abnormal chest findings. In one study, the P_aO_2 was found to be low, averaging 71 mm Hg but falling below 60 mm Hg in 25% of patients. In this study, the death rate in adults was 4.3%; death was preceded by a lower hemoglobin value, a higher WBC count, and multilobe involvement. Autopsy of 16 cases showed that nine patients had pulmonary embolism and fat emboli and possibly 20% had bacterial infections. In patients with acute chest syndrome and pulmonary infection, the most common infecting organism was *Chlamydia pneumoniae* (30%), followed by *Mycoplasma pneumoniae* (21%), respiratory syncytial virus (10%), *Staphylococcus aureus* (4%), and *Streptococcus pneumoniae* (3%).[71]

Usually, therapy for acute chest syndrome should include incentive spirometry,[64] antimicrobial therapy for patients with evidence of infection, the cautious use of analgesia, aggressive fluid replacement, and consideration of bronchoalveolar lavage to identify microbial infection or the fat-laden macrophages of fat emboli. Meticulous monitoring is required; repeat measurements of oxygenation should be made, and transfusions should be performed when clinically necessary. One of the most important benefits of hydroxyurea therapy is its ability to reduce the frequency of acute chest syndrome.[53,72] Children may also need supplementary penicillin prophylaxis.[73]

Hepatobiliary disease Cholelithiasis occurs in 30% to 70% of patients, some of whom exhibit signs and symptoms of cholecystitis.[74] There are conflicting data regarding frequency of cholecystitis or obstruction of the common bile duct.[74,75] If cholecystectomy is to be done, one should wait until the painful crisis is over. Transfusions should be given to raise the hemoglobin to 10 g/dl before surgery, if necessary, and the procedure should be done laparoscopically.[74]

Hepatic complications include congestive hepatopathy secondary to heart failure and viral hepatitis from frequent transfusions. Sickling in the liver can also produce hepatopathy. Often, serum bilirubin levels exceed 30 mg/dl in patients with intrahepatic cholestasis, and coagulation abnormalities may lead to hemorrhagic complications and death.

Renal and urologic complications Water loss as a result of an inability to concentrate urine may enhance the sickling process. The extremely hypertonic milieu of the renal medulla induces severe sickling and destruction of the vasa recta. Hematuria and papillary necrosis ensue. These complications are also observed in patients with sickle trait and in those who have sickle cell–hemoglobin C disease. The defect in renal concentrating ability appears to depend on the amount of HbS polymer contained in cells and is thus less severe in patients who also have α-thalassemia variants.[76]

Complications include renal tubular acidosis, hyperkalemia, and proteinuria. Treatment with enalapril reduces proteinuria, suggesting the presence of a component of glomerular capillary hypertension.[77] Renal failure, in association with worsening anemia, contributes to the death of about one fifth of patients older than 40 years who have homozygous sickle disease.

Priapism is an extraordinarily painful complication of sickle cell anemia and may result in impotence.[78] A United Kingdom study reported a good response in 13 of 18 patients treated for priapism with the alpha-adrenergic agonist etilefrine; however, this agent is not available in the United States.[79]

Neurologic disorders Neurologic complications of sickle cell disease include stroke, subarachnoid hemorrhage, and isolated functional losses that suggest a focal occlusion. The pathogenesis of occlusion of the large cerebral arteries is probably different from that of the microvascular occlusive events that occur in hypoxic capillary beds. The most likely underlying causes are damage to the vascular endothelium, followed by extensive intimal proliferation and then thrombosis of the damaged vascular bed.[45] In a multi-institutional study of 4,082 patients, the prevalence of cerebrovascular accidents (CVAs) was 4% to 5%; the incidence was 0.61 per 100 patient-years.[80] Of the CVAs, 54% were infarcts, 34% were hemorrhagic in nature, 11% were transient ischemic attacks (TIAs), and 1% had both infarctive and hemorrhagic features. Of the patients who survived, the recurrence rate of CVA was 14%. Mortality was 11%. Virtually all patients who died had hemorrhagic CVAs.

In a prospective study in which transcranial Doppler ultrasonography was used to pinpoint children at risk for stroke, treatment with standard care or transfusion therapy (to reduce the HbS concentration to < 30%) resulted in only one CVA, compared with 10 CVAs and one intracerebral hematoma in the 65 control subjects ($P < 0.002$). The trial was terminated early.[59] The success of this trial raises many serious questions about the necessity of ultrasonographic devices for successful management; the optimum duration of transfusion therapy; the inevitable consequences of transfusional hemochromatosis [*see* β-Thalassemia major (Cooley anemia), *below*] and the necessity for ethnically matched blood to minimize allotransfusion reaction; the willingness of patients and families to accept transfusion therapy; and the role of allogeneic bone marrow transplantation as a potential

alternative.[59,81] The risk of recurrent cerebrovascular events is increased in patients receiving long-term transfusion therapy who have multiple cerebral collateral vessels as a result of moyamoya disease (hazard ratio, 2.40).[82]

Ocular complications The major ocular problems associated with sickle cell anemia are retinopathy, vitreous hemorrhage, and neovascularization. Annual ophthalmologic evaluations are recommended. The efficacy of laser photocoagulation in treating sickle-induced ocular changes is currently being investigated.

Dermatologic complications Poorly healing leg ulcers can be an important cause of morbidity in patients with sickle cell anemia. The degree of anemia does not seem to correlate with the presence or severity of these ulcers, but incompetence of venous valves and the resulting venous insufficiency have been associated with ulceration.[83] Standard management includes debridement, control of local infection, use of wet-dry dressings, and possibly RBC transfusion. Local treatment with GM-CSF enhances healing, perhaps by stimulating the local growth of macrophages.[84] GM-CSF can be either injected perilesionally or added topically to the wound, but the more successful application method involves the subcutaneous injection of 100 μg of GM-CSF in each of four sites circumferentially around the ulcer at a distance of 5 mm from its edge (resulting in a total dose of 400 μg in the wound). In some circumstances, one treatment sufficed, whereas in others, weekly treatments for 4 to 12 weeks were necessary. This therapy has not been approved by the Food and Drug Administration.

Aplastic crisis Aplastic crisis rapidly lowers hemoglobin and hematocrit levels and produces reticulocytopenia, as it does in any chronic hemolytic state. Parvovirus infection has been found to cause aplastic crisis,[7] as has bone marrow necrosis.[85]

Susceptibility to infections Patients with sickle cell anemia are hyposplenic and exhibit complement system abnormalities. Deficient serum opsonizing activity for *Salmonella* organisms may confer an increased susceptibility to those infections, including osteomyelitis.

Anesthesia complications The hypoxemia and vascular stasis that may occur during general anesthesia enhance sickling and may lead to a sickle crisis in the postoperative period. In an analysis of almost 4,000 patients, 12 deaths were associated with 1,079 procedures, and there were more complications after regional anesthesia than after general anesthesia.[86] A simple transfusion program to raise the hemoglobin level to 10 g/dl was as effective as more aggressive preoperative programs in reducing the rate of complications.[87]

Pregnancy and contraception The dangers of pregnancy for women with sickle disease include pulmonary problems and an increased incidence of urinary tract infection, hematuria, preeclampsia, and maternal death. Presumably, pelvic hypoxemia and the vascular overload associated with pregnancy lead to enhanced sickling, with its attendant complications. Vaso-occlusion in the placenta may account for fetal death and low birth weight.

Experienced clinicians differ in their approach to the pregnant patient with sickle disease. Some advocate only meticulous conservative care, whereas others recommend prophylactic transfusions. A controlled study has indicated that there is no advantage to the use of prophylactic transfusions.[88]

Chorionic villus sampling (which can provide DNA for analysis in the first trimester of pregnancy), DNA amplification techniques, and probes that identify the specific nucleotide change of sickle cell anemia can give a relatively safe and very reliable prenatal diagnosis.[89]

Oral contraceptives may pose a special hazard to women with sickle cell anemia, because they have been associated with a slight increase in the incidence of stroke, venous thromboembolism, and myocardial infarction. However, the emerging evidence that daily use of oral contraceptives containing less than 50 mg of synthetic estrogens is relatively safe suggests that patients with sickle disease can take such medication with reasonable confidence. The use of the Norplant implantable contraceptive device is another alternative for some patients. In any event, pregnancy or abortion in sickle disease carries significant risk.[88]

Genetic Counseling

A key element to be considered in the provision of genetic counseling to patients with sickle trait or sickle disease is the significant morbidity in affected children and adults. Couples with sickle disease or sickle trait may want to have children despite the associated fetal and maternal risks. There are about 4,000 to 5,000 such pregnancies in the United States each year.[89] In one study, 286 of 445 pregnancies (64%) in mothers with sickle cell anemia proceeded to delivery; 21% of the infants were small and thus would be expected to require additional care, which the mother might have difficulty providing. In this study, there was one maternal death caused by sickle cell disease[90] [*see* Genetic Counseling and Prenatal Diagnosis, *below*].

Prognosis in Sickle Cell Disease

Whereas it was once assumed that most patients with sickle cell anemia would die by 20 years of age, the median age of death is now 42 years for men and 48 years for women.[51] This life expectancy is 25 to 30 years less than that of the general African-American population. Of the identified causes of death, only 18% involved organ failure—predominantly renal disease, heart failure, or the consequences of chronic strokes. Thirty-three percent of patients died during acute pain crises; these crises were frequently associated with the acute chest syndrome and were less often associated with stroke. The presence of α-thalassemia had no measurable effect. Predictors of poor outcome were a white cell count greater than 15,000/μl; a low HbF level; and organ involvement manifested by renal disease, acute chest syndrome, and neurologic events. Taking hydroxyurea had a significant impact on prognosis, with a 40% decrease in mortality and a reduction in painful crises.[91]

SICKLE VARIANTS

Sickle Trait

Heterozygosity for the sickle cell gene results in sickle trait (HbAS). The RBCs of persons with sickle trait have an HbS concentration of less than 50%; frequently, the level is as low as 30%.

Generally, persons with sickle trait lead normal, healthy lives. A few complications occur: hyposthenuria; renal hematuria; and, during pregnancy, bacteriuria and pyelonephritis. Splenic infarction occurs under conditions of hypoxia; it also occurs at high altitudes, predominantly in nonblack persons who have sickle cell anemia.

Sickle trait has been identified as a major risk factor for sudden death during basic training in the military[92]; death has resulted from unexplained cardiac arrest, heatstroke, heat stress, or rhabdomyolysis. Increasing age has been correlated with an increased risk of sudden death. However, these events have occurred under extreme conditions: very strenuous physical activity, usually in untrained persons, occasionally at high altitudes or in extreme heat. Usually, persons with sickle trait who are accustomed to physical activity do not have an increased risk of sudden death. For example, the incidence of sudden death in African-American football players with sickle trait is not higher than in other players.[93]

Therapeutic options for renal hematuria include the administration of diuretics, parenteral bicarbonate, transfusions, or ε-aminocaproic acid.

Sickle Cell–β-Thalassemia

When combined with sickle trait, a defect in the β-thalassemia gene produces a disease very similar to sickle cell anemia. The β-thalassemia gene reduces the rate of synthesis of the β^A chain, resulting in a predominance of β^S in patients with sickle trait. Depending on whether the patient has a β^0 or a β^+ thalassemia, the RBCs contain varying amounts of HbS, HbA, HbA_2, and HbF. Patients with β^0 thalassemia have no HbA, but only HbS, HbF, and HbA_2; thus, disease is severe in these patients. Diagnosis is based on an elevation in the level of HbA_2, HbF, or both on hemoglobin electrophoresis, as well as a positive family history of thalassemia and the sickle gene. In a study of 55 Greek patients, treatment with hydroxyurea resulted in distinct clinical improvement.[94,95] Further description and information on diagnostic testing is available online at http://www.geneclinics.org.

Sickle Cell–Hemoglobin C Disease

In sickle cell–hemoglobin C (HbSC) disease, almost equal amounts of HbS and HbC are formed. Between 1% and 2% of hemoglobin is HbF, and small amounts of HbA_2 are also present; however, HbA is absent. The increased sickling seen in these patients results from the pathologic effect of HbC [*see* Hemoglobin C Disease, *below*].[96] As many as 30% to 50% of patients with this disorder are not anemic and have only modest reticulocytosis. Patients may not be identified until the disorder manifests itself in the form of a vaso-occlusive crisis during surgery, pregnancy, or a medical emergency.[97] Splenomegaly, proliferative retinopathy, aseptic necrosis of long bones, and the acute chest syndrome[96] also occur. The peripheral smear [*see Figure 5*] shows irreversibly sickled cells in addition to target cells, stomatocytes, and erythrocytes with eccentric hemoglobin depositions, probably representing HbC aggregates or crystals. Diagnosis is confirmed by hemoglobin electrophoresis or HPLC.[97] Further information on diagnostic testing is available online at http://www.geneclinics.org.

Management is the same as that for sickle cell anemia. In a study of six patients with HbSC disease, treatment with hydroxyurea at a dosage of 1,000 mg/day resulted in an increase in MCV, a decrease in so-called stress reticulocytes, an increase in hemoglobin, and probably a reduction in cell density. Although not definitive, this small study suggests that hydroxyurea benefits patients with HbSC disease.[98] Life expectancy for patients with HbSC disease is almost 20 years greater than that for patients with HbSS disease.[51]

OTHER HEMOGLOBINOPATHIES

Hemoglobin C Disease

The HbC molecule is $\alpha_2\beta_2^{6glu\rightarrow lys}$; the gene mutation probably originated at a single site in Burkina Faso, in West Africa.[97] The presence of this hemoglobin produces almost no illness in the heterozygous state but causes mild compensated hemolysis and palpable splenomegaly in the homozygous state.

The relative insolubility of HbC is responsible for the pathologic changes associated with its presence. HbC probably interacts with the K^+-Cl^- cotransporter, which keeps it active, whereas the K^+-Cl^- cotransporter normally shuts off in RBCs after the reticulocyte stage. The result is a loss of K^+, cellular dehydration with elevated MCHC, and then aggregation and crystallization of the poorly soluble HbC.[97] The relative insolubility of HbC causes erythrocytes to become rigid and thereby subject to fragmentation and to loss of membrane material, resulting in the microspherocytes seen on a peripheral blood smear [*see Figure 3*].

Target cells, an important morphologic finding, constitute about 80% of the erythrocytes. HbC crystals are in the oxyhemoglobin state and dissolve when the RBCs are deoxygenated, probably accounting for the absence of vaso-occlusive episodes.

Diagnosis of hemoglobin C disease is based on blood-smear findings and the absence of evidence of either iron deficiency or thalassemia; the diagnosis is confirmed by hemoglobin electrophoresis. No therapy is required.

Hemoglobin E Disease

In hemoglobin E disease, lysine is substituted for glutamic acid at position 26 of the β-globin chain, resulting in an oxidatively unstable molecule. Hemoglobin E trait is found predominantly in Southeast Asia. It came to clinical attention in the United States as a result of the influx of Southeast Asians, in whom the incidence of this trait is about 10%.

Patients heterozygous for HbE have normal hemoglobin values, microcytosis, and no splenomegaly. Electrophoresis reveals that 70% of the hemoglobin is HbA, 25% is HbE, and the remainder is HbA_2 or HbF. Inexperienced laboratories may mistake HbE for HbA_2; the clue to this error is that HbA_2 never accounts for more than 8% of the total hemoglobin. A laboratory report of an HbA_2 level of 25% should prompt a review of the data.

Patients homozygous for HbE have mild anemia, with a hemoglobin level of about 12 to 13 g/dl, a low mean corpuscular volume, and an elevated RBC count but no reticulocytosis; they exhibit microcytes and target cells. Electrophoresis shows only HbE. Chronic hemolysis does not occur. Oxidant drugs such as dapsone should be avoided in both heterozygotes and homozygotes.

A serious clinical problem occurs when a patient is doubly heterozygous for HbE and β-thalassemia trait. Such patients present with β-thalassemia intermedia, characterized by severe anemia and splenomegaly (see below). These patients occasionally require transfusions of blood and even allogeneic bone marrow transplantation.[99] Further information on diagnostic testing is available online at http://www.geneclinics.org.

Unstable Hemoglobinopathies

Many individual variants make up the unstable hemoglobinopathies. The hemoglobin instabilities stem from amino acid substitutions that deprive the molecule of its heme group, alter the heme pocket, loosen the link between its α and β chains, or weaken the subunit structure [*see 5:I Approach to Hematologic Dis-*

orders]. The result is disruption and precipitation of hemoglobin, particularly when it is subjected to oxidant attack. Precipitated hemoglobin forms Heinz bodies, which are observed even in persons heterozygous for the unstable hemoglobin variant. Because of the deleterious effects of Heinz bodies on the erythrocyte and its membrane, significant hemolysis can occur even in the heterozygous state.

Diagnosis of an unstable hemoglobinopathy is suggested by the presence of a partly compensated chronic nonspherocytic hemolysis. Heinz bodies are observed in the erythrocytes of patients who have undergone splenectomy. Erythrocytes from patients who have not undergone splenectomy demonstrate Heinz bodies on incubation with brilliant cresyl blue dye. The differential diagnosis of a hemoglobinopathy includes G6PD deficiency; this disorder can usually be ruled out by direct assay for the enzyme.

Management includes avoidance of oxidant drugs. Splenectomy may be considered when hemolysis is severe and inadequately compensated.

Hemoglobin with Abnormal Oxygen Affinity

The presence of hemoglobin with increased oxygen affinity should be considered in the differential diagnosis of unexplained erythrocytosis, particularly if there is a familial association [*see 5:V The Polycythemias*]. Hemoglobin electrophoresis may reveal the disorder, but in suspected cases, measurement of the oxyhemoglobin dissociation curve [*see 5:I Approach to Hematologic Disorders*] is preferable as a basis for diagnosis. Hemoglobin Chesapeake and hemoglobin Rainier are examples of forms with particularly increased oxygen affinity.

The rare instances of hemoglobin with low oxygen affinity, such as hemoglobin Kansas, represent mutations. Patients with low-oxygen-affinity hemoglobinopathy are sometimes cyanotic because of enhanced oxygen unloading.

Methemoglobinemia

Methemoglobin is an oxidation product of hemoglobin in which iron is in the ferric form; thus, the molecule cannot bind oxygen reversibly. Ordinarily, 1% of hemoglobin is in the ferric state. Between 0.5% and 3% of deoxyhemoglobin is normally spontaneously oxidized to methemoglobin every day. The normal reducing power of erythrocytes [*see Table 1*] maintains the balance between oxidation and reduction. The enzyme system that reduces 95% of methemoglobin to hemoglobin involves two proteins, NADH-cytochrome b_5 reductase and cytochrome b_5, and also requires NADH. As the name suggests, NADH-cytochrome b_5 reductase uses NADH to reduce cytochrome b_5. Reduced cytochrome b_5 then reduces methemoglobin.[100,101] Novel mutations in the affected gene have been described.[102]

Most often, methemoglobinemia is acquired by ingestion of or exposure to oxidants that oxidize Fe^{2+} so fast that the reducing systems are overwhelmed [*see* Mechanism of Oxidative Attack, *below*].

There are two congenital forms of methemoglobinemia. In the hereditary enzymopenic form of methemoglobinemia, patients are homozygous or doubly heterozygous for a deficiency of NADH-cytochrome b_5 reductase.[103] These patients appear blue even when only about 10% of their hemoglobin is in the form of methemoglobin, but they are not sick and easily tolerate methemoglobin levels of 25% or more. In contrast, the presence of about 5 g/dl of reduced, deoxygenated hemoglobin produces cyanosis. Patients with this form of methemoglobinemia do not exhibit hemolysis and generally do not require treatment. Assay of NADH-cytochrome b_5 reductase, done by a special laboratory, can establish the diagnosis. If desired, methylene blue at a dosage of 100 to 300 mg/day orally can be used, but it may produce urinary discomfort.[100] Methylene blue transfers electrons from NADPH to methemoglobin.

The other hereditary form of methemoglobinemia is caused by HbM, of which there are five rare variants. Each of these variants contains an amino acid substitution in the heme pocket, which allows stable bonds to be formed between the heme iron and the amino acid side chains. These bonds keep hemoglobin in the Fe^{3+} form—a form that is unable to bind oxygen and is inaccessible to the reducing enzymes. The disorder is seen only in heterozygotes; about 30% of hemoglobin is abnormal, as detected by electrophoresis. Cyanosis is noted at birth. Hemolysis is minimal, and therapy is not needed.

THE THALASSEMIAS

The thalassemias have a worldwide distribution; in many regions, they are responsible for major medical, social, and economic perturbations. Throughout the world, the regions in which the thalassemias occur are contiguous with regions endemic for malaria, indicating that the heterozygous forms of thalassemia provide protection against malaria.[104] The techniques of molecular biology have helped elucidate the pathophysiology of these syndromes,[104] which in turn has enabled investigators to make unambiguous antenatal diagnoses. Using these data, expectant parents can make thoughtful, informed choices regarding the outcome of pregnancies in which the fetus is severely affected.

Molecular Genetics

Thalassemias result from gene deletion, abnormalities in transcription and translation [*see 9:VII Basic Genetics for the Clinician*], and instability of the mRNA directing globin synthesis or of the globin itself. The genes controlling the synthesis of the α and non-α chains of hemoglobin are located on chromosomes 11 and 16 [*see Figure 6*].

Pathophysiology

In a healthy person, the synthesis of α and β chains is meticulously coordinated to produce adult HbA ($\alpha_2\beta_2$). In contrast, patients with thalassemia usually demonstrate imbalanced synthesis of normal globin chains. Occasionally, however, thalassemia-like syndromes can result from diminished production of a structurally abnormal chain.[105] Because one of the globin chains is present in reduced amounts, the unpaired chain accumulates in the developing erythroid precursor cell, and toxicity results [*see Figure 6*]. Consequently, erythroid cells die in the marrow, giving rise to a classic form of ineffective erythropoiesis [*see 5:III Anemia: Production Defects*]; affected erythrocytes undergo hemolysis in the peripheral blood.

The β-thalassemias are characterized by diminished production of β-globin chains, causing unmatched α-globin chains to accumulate and aggregate. These aggregates of α chains precipitate, causing decreased ATP synthesis, potassium leak, and reduced amounts of surface sialic acid; the affected erythrocytes are misshapen and relatively rigid. The membrane Ca^{2+} barrier is breached, allowing Ca^{2+} to enter. These α-globin aggregates also appear to keep the K^+-Cl^- cotransporter functioning; as a result, in severe forms of β-thalassemia, dehydration of varying degree is seen.[106] The RBC membranes are unstable and fragment easily; there is evidence of oxidation of RBC membrane proteins 4.1 and

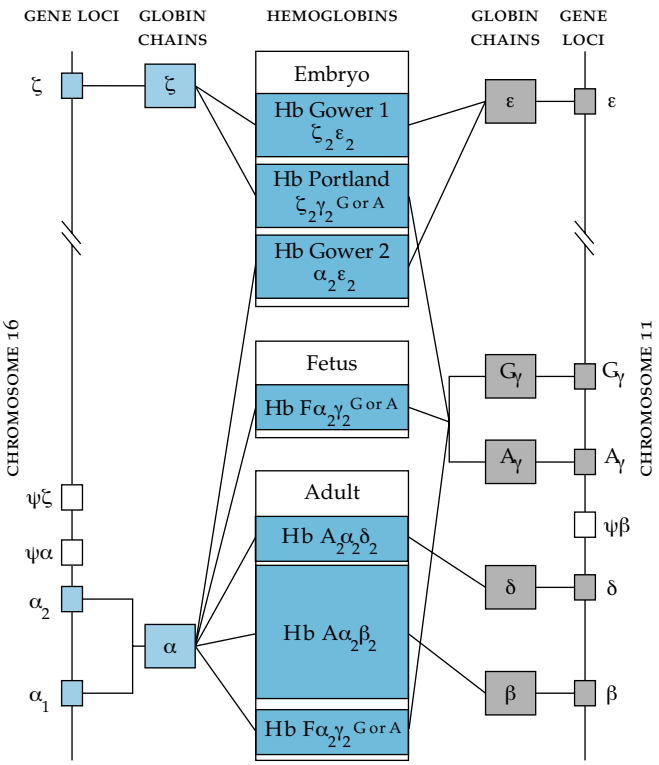


Figure 6 **The genes encoding the α and non-α chains that come together to form the hemoglobin tetramer lie on chromosomes 16 and 11, respectively. The α genes are present at duplicated loci. Six distinct species of normal hemoglobin have been described. Three of these hemoglobins are synthesized only during embryonic stages of development (Hb Gower 1, Hb Portland, and Hb Gower 2). HbF predominates during fetal development, and a small amount continues to be synthesized in adult life. HbA and HbA_2 constitute the major forms of adult hemoglobin. Different hemoglobin genes are activated at various stages of development. In the embryo, ξ chains combine with ε chains to yield Hb Gower 1 and with γ chains to form Hb Portland; α and ε chains are linked to form Hb Gower 2. There are two varieties of γ chains that are derived from separate loci and that differ in a single amino acid; $^{G}\gamma$ contains glycine at position 136, whereas $^{A}\gamma$ contains alanine at this position. The genes coding for the two other non-α chains, β and δ, which are required for the synthesis of adult hemoglobins, are switched on late in fetal development. The factors regulating this precisely coordinated sequence of changes in hemoglobin production are poorly understood; some evidence suggests that DNA segments intervening between the various hemoglobin genes may control the relative rates of synthesis of the adjacent gene products.**

spectrin [*see Figure 2*]; and phosphatidylserine migrates to the outer membrane layer, perhaps forming a nidus for thromboembolic events.[107,108] These destructive alterations of the membrane, which can be detected by macrophages, may in part be caused by local oxidation.[109,110] Abnormal accumulations of α chains probably account for the accelerated apoptosis and ineffective erythropoiesis seen in marrow erythroid precursors.[110] The overall decrease in hemoglobin synthesis per cell accounts for the observed hypochromia and target cell formation.

Patients with α-thalassemia demonstrate accumulations of excess β chains that, if present in sufficient amounts, form β_4 tetramers (HbH) [*see Figure 6*]. Such tetramers have high oxygen affinity and are unstable, aggregating in the presence of oxidative stresses such as infection. β-Globin aggregates also become attached to the erythrocyte's membrane skeleton, but they produce lesions different from those produced by α-globin aggregates. In the severe α-thalassemias, ineffective erythropoiesis is less prominent[85]; rather, destruction of peripheral RBCs is the critical characteristic. RBCs in severe α-thalassemia are rigid, but in contrast to those in severe β-thalassemia, the membranes in severe α-thalassemia are more stable than normal.[107,108] Also in contrast to β-thalassemia, the RBCs in severe α-thalassemia are uniformly overhydrated.[106]

Both α-thalassemia and β-thalassemia are characterized by variable degrees of anemia. This variation is attributable to varying degrees of ineffective erythropoiesis and hemolysis.[110] When the anemia is severe, the associated hypoxia induces a vigorous compensatory erythropoiesis, leading to expansion of the marrow cavity, osteopenia, and enlargement of reticuloendothelial organs; tumors may arise at sites of extramedullary erythropoietic activity. Destruction of erythroblasts and erythrocytes may predispose to cholelithiasis and obstructive jaundice. Patients with the more severe forms of thalassemia require regular transfusions, which may eventually generate clinically significant iron overload [*see 5:II Red Blood Cell Function and Disorders of Iron Metabolism*].

Diagnosis and Treatment of Thalassemia

The HbF and HbA_2 measurements that aid in the diagnosis of the β-thalassemias are readily available from clinical laboratories using hemoglobin electrophoresis and, more recently, HPLC. In contrast, the tests required to diagnose the α-thalassemias are quite sophisticated and in the past were performed only in institutions specifically engaged in thalassemia research. Currently, specialized laboratories can detect the number and position of deleted α-globin genes. Further description and information on diagnostic testing is available online at http://www.geneclinics.org.

The clinical diagnostic tools used to assess patients suspected of having α-thalassemia include clinical history, smear evaluation, calculation of indices, brilliant cresyl blue staining, and family studies. In practice, α-thalassemia trait is diagnosed on the basis of a finding of microcytosis in an iron-replete patient who has normal HbA_2 and HbF levels.

The diagnosis of either α- or β-thalassemia should be suspected when the MCV is less than 75 fl and the RBC count is greater than 5 million cells/μl. A patient with these two findings has an 85% chance of having a thalassemia syndrome.[111] In one study, diagnosis of thalassemia was not considered in about half of the patients with the disease.

β-*Thalassemia*

The deficient synthesis of β-globin characteristic of β-thalassemia leads to accumulation of unmatched α chains. A diagnostically significant development in β-thalassemia is the partial compensatory increase of the δ and γ chains that yields elevated levels of HbA_2 ($\alpha_2\delta_2$) and HbF ($\alpha_2\gamma_2$), respectively [*see Figure 5*]. The β-thalassemia variants produce three clinical syndromes: β-thalassemia major, β-thalassemia minor, and thalassemia intermedia.

β-Thalassemia major (Cooley anemia) β-Thalassemia major is usually a homozygous or doubly heterozygous condition; both parents of an affected individual carry a β-thalassemia trait. In β^0-thalassemia, the most severe variant, no β chains are synthesized; only HbF and HbA_2 are found. β^+-Thalassemia is somewhat less

severe. It is characterized by small amounts of β chains and small quantities of HbA in addition to HbF and HbA_2. δβ-Thalassemia is yet milder; it is caused by deletion of the δ-globin and β-globin genes. This mutation prohibits production of HbA_2 and HbA, permitting synthesis of fetal hemoglobin alone.

β-Thalassemia major is characterized by severe anemia that appears in the first year of life. Patients also have jaundice, hepatosplenomegaly, expansion of the erythroid marrow with secondary body changes (including retarded growth), and an increased susceptibility to infection.

Diagnosis is not difficult; no other condition closely resembles Cooley anemia. The peripheral smear shows nucleated RBCs, distorted hypochromic erythrocytes, and basophilic stippling, which represents aggregates of ribosomal RNA [*see Figure 5*]. Supravital staining reveals accumulations of excess unmatched α chains.

Management consists of aggressive transfusion therapy. The strategy involves transfusing to a hemoglobin level of about 12 g/dl, then allowing the hemoglobin level to fall to about 9 g/dl just before the next transfusion; this prevents such complications as heart failure, fluid overload, and skeletal deformity. Splenectomy is usually necessary to enhance survival of the patient's own RBCs as well as transfused RBCs.[112] Vaccination with pneumococcal vaccine is indicated because of the risk of pneumococcal sepsis after splenectomy.

Long-term transfusions eventually generate iron overload, which if untreated leads to death from cardiac hemochromatosis during adolescence. Iron overload should be managed prophylactically by infusion of subcutaneous deferoxamine, an iron chelator, before iron buildup occurs. Subcutaneous deferoxamine at a dosage of 50 mg/kg/day can effect iron losses of 50 to 200 mg/day but only if infused continuously over 8 to 12 hours for 5 days each week.[113] Such therapy not only prevents left ventricular dysfunction but also reverses already established abnormalities.[114] The beneficial effects of iron chelation have improved the prognosis for persons with Cooley anemia[114]: it is no longer inevitable that patients die in their 20s of arrhythmia and left ventricular failure. With current deferoxamine therapy, 61% of patients born before 1976 have had no cardiovascular disease. Compliant patients whose ferritin levels are mostly below 2,500 ng/ml have a survival rate of 91% after 15 years.[115] However, compliance with deferoxamine is a problem, and the cost of the drug, together with the cost of the pump and tubing that are required for administration, takes it out of reach of most patients in developing countries. Use of the oral iron chelator deferiprone remains a very contentious issue with regard to its safety and efficacy.[116] A new orally active tridentate iron chelator (ICL670) looked very promising in early trials, but it is currently available only on fixed protocols.[117] Bone marrow transplantation has been performed with HLA-matched sibling donors. More than 1,000 patients have now undergone allogeneic bone marrow transplantation from sibling donors who either were normal or had β-thalassemia trait.[118] Some patients with hemoglobin E β-thalassemia have a phenotype fully as severe as β-thalassemia major and require the same therapy, including allogeneic bone marrow transplantation.[99] Experience with cord blood transplantation is more limited.[119] Depending on the condition of the patient at the time of transplantation, the rate of transplantation-related mortality was 5% to 19%; the cure rate was 54% to 90%.[120,121] Other approaches to the treatment of severe β-thalassemia are still experimental.[56] However, two small clinical trials have shown hydroxyurea to be of benefit. This treatment probably works by increasing the production of γ chains, which combine with and remove the excess α chains, and by causing an increase in the production of HbF, a useful hemoglobin.[122,123]

β-Thalassemia minor (β-thalassemia trait) Patients with β-thalassemia minor are usually heterozygous for a β-globin mutation and have either mild or no anemia. The peripheral smear shows distinct hypochromia and microcytosis with basophilic stippling. Splenomegaly is occasionally found.

The HbA_2 level is elevated above 5% in 90% of patients, and the HbF level is raised above 2% in 50% of patients. This increase in fetal hemoglobin occurs in varying proportions per RBC (a phenomenon known as heterocellular distribution), as shown by the Kleihauer-Betke stain. Patients with higher HbF levels have less severe anemia. Heterozygotes for δβ-thalassemia produce increased amounts of HbF but only normal amounts of HbA_2.

Iron deficiency anemia should be excluded from the differential diagnosis of β-thalassemia trait [*see 5:II Red Blood Cell Function and Disorders of Iron Metabolism*]. Generally, it is easy to distinguish the two disorders. Both are associated with hypochromia and microcytosis, but iron deficiency produces hypoproliferation of RBCs, whereas β-thalassemia minor causes only a minimal reduction in their number. At a hemoglobin level of 9 g/dl, an iron-deficient patient has an RBC count of about 3 million cells/μl, whereas a patient with β-thalassemia trait has an RBC count of about 5 million cells/μl. If the diagnosis remains in doubt, measurement of the serum iron and iron-binding capacity or of the serum ferritin level can be used to distinguish these disorders. It is important to remember, however, that a patient with thalassemia trait may also be iron deficient as a consequence of vaginal bleeding, gastrointestinal bleeding, or both.

Thalassemia intermedia As the term implies, thalassemia intermedia is characterized by clinical manifestations of moderate severity. Patients with this syndrome have distinct anemia, with hemoglobin levels as low as 6 to 7 g/dl; they exhibit variable degrees of hepatosplenomegaly, but they usually do not require regular transfusions. During infections or other erythropoietic insults, however, transfusions may be needed transiently. In two small clinical trials, isobutyramide was found to be of benefit.[124,125]

β-Thalassemia–like variants The hemoglobinopathy associated with hemoglobin Lepore represents another β-thalassemia variant. Patients who are homozygous for this disorder present with Cooley anemia or thalassemia intermedia, and their RBCs contain only hemoglobin Lepore and HbF.[105]

Hereditary persistence of fetal hemoglobin The RBCs of patients heterozygous for hereditary persistence of fetal hemoglobin (HPFH) contain about 50% of HbF, whereas homozygotes have 100% of HbF. It was once believed that patients with HPFH were well and had minimal or no anemia, but some clinical variants of HPFH associated with distinct anemia have been described.

α-Thalassemia

The α-globin gene and the β-globin gene differ in two major respects. First, there are no fetal, neonatal, or adult substitutes for the α-globin genes; second, there are only two β-globin genes but four α-globin genes—two α-globin genes on each chromosome 16 [*see Figure 6*]. The normal α-globin genotype is designated

αα/αα. Patients who carry the α-thalassemia–1 variant exhibit a deletion of two α-chain genes from the same chromosome and thus have the --/ or α^0 haplotype; this deletion is common among Asian patients. Patients who have the α-thalassemia–2 variant have lost one α gene on one chromosome and show the -α/ or α^+ haplotype. Although this mutation is particularly frequent among blacks, it is also observed in Asian and Mediterranean populations. Five clinically distinct syndromes have been recognized among patients who carry different genotypes for the α-globin genes: hemoglobin Barts or hydrops fetalis (--/--); hemoglobin H disease (--/-α); heterozygous α-thalassemia–1 (--/αα); homozygous α-thalassemia–2 (-α/-α); and the silent carrier syndrome (-α/αα).

Hemoglobin Barts (hydrops fetalis) Children with hemoglobin Barts syndrome are homozygous for α-thalassemia–1 (--/--) and therefore produce no α chains. The unmatched γ chains form γ_4 tetramers (hemoglobin Barts). All infants with this condition are born hydropic, and most die unless rescued by intrauterine stem cell transplantation. The parents are usually heterozygous for α-thalassemia–1 (--/αα).

Hemoglobin H disease The clinical picture of hemoglobin H disease is that of variable hemolytic anemia occurring in patients of Asian, Middle Eastern, or Mediterranean origin. HbH, which precipitates on staining with brilliant cresyl blue, can usually be detected in the patient's freshly drawn RBCs. The molecular mechanisms may involve deletion of three α genes, as would be the case if the patient were doubly heterozygous for α-thalassemia–1 (--/) and α-thalassemia–2 (-α/), yielding a --/-α genotype. Splenomegaly is common. Patients usually do not require regular transfusions, but transient RBC support may be necessary when the patient has infection or experiences other oxidative stresses that lead to the precipitation of the unstable HbH and enhanced hemolysis. During pregnancy, anemia may become clinically severe and require RBC transfusion. Partners of pregnant patients with HbH disease should be screened because if the partner carries an α-thalassemia trait, the fetus may have homozygous α-thalassemia hydrops fetalis. Occasionally there is associated growth retardation and even iron accumulation in the absence of RBC transfusions.[126]

Heterozygous α-thalassemia Heterozygous α-thalassemia–1 (--/αα), a common genotype among Asians, causes mild or no anemia; rather, it engenders distinctly hypochromic, microcytic RBCs. Patients homozygous for α-thalassemia–2 (-α/-α), a common genotype among blacks, lack two α genes; the clinical manifestations of patients with this genotype resemble those of patients heterozygous for α-thalassemia–1. The heterozygous state for α-thalassemia–2 (-α/αα) is clinically undetectable and thus represents the silent carrier syndrome.

α-Thalassemia–like syndrome Hemoglobin constant spring (hemoglobin CS) is a structurally abnormal hemoglobin common in some Asian populations. The α-globin gene contains a mutation in the termination codon, resulting in the synthesis of an α-globin that contains an additional 31 amino acids. Patients heterozygous for this defect have a clinical picture similar to that of a patient homozygous for α-thalassemia–2. Patients homozygous for hemoglobin CS tend to have slightly more severe clinical manifestations than patients heterozygous for α-thalassemia–1. In patients who are doubly heterozygous for α-thalassemia–1 and alpha CS (--/αCS α) and who have HbH/HbCS, disease is slightly more severe than in those with hemoglobin H disease.[110]

Genetic Counseling and Prenatal Diagnosis

Parents who have had a stillborn hydropic infant or a child with Cooley anemia are justifiably reluctant to repeat the experience. Adults from thalassemia families who know themselves to be heterozygous for thalassemia are often eager to receive genetic counseling when starting their own families. Genetic counseling entails screening prospective parents on the basis of routine diagnostic tests and family studies. In addition, advances in molecular genetics can now provide accurate, unambiguous prenatal diagnoses of the thalassemias. In the first trimester, chorionic villus sampling combined with the use of polymorphic DNA markers and synthetic oligonucleotide probes can provide the definitive diagnosis in about 80% of cases of β-thalassemia [*see 9:VII Basic Genetics for the Clinician*].[127,128] Indeed, the incidence of births of infants with thalassemia major has fallen in several parts of the world. Different ethnic groups respond differently to genetic counseling.

Extracorpuscular Defects

Erythrocytes can be damaged through trauma or by antibodies, drugs, abnormally functioning organs, and toxins. These causes of an extracorpuscular defect should be considered whenever hemolysis develops in a patient who has no personal or family history of anemia.

MECHANICAL INJURY: MICROANGIOPATHIC HEMOLYSIS

Microangiopathic hemolysis is characterized by the appearance of bizarre, fragmented erythrocytes (e.g., schistocytes, or helmet cells) on a peripheral smear and by signs of intravascular and extravascular hemolysis.

Pathophysiology

The normal erythrocyte can withstand considerable elongation and twisting, but it disintegrates when subjected to strong stretching or shearing forces. Stresses of this magnitude have been observed to occur in jets produced by deformed aortic valves, by arteriovenous shunts, by ventricular septal defects, or by the older valvular prostheses.

Localized intravascular coagulation, in which fibrin strands bridge the arteriolar lumen, is thought to occur in arterioles supplying inflamed or neoplastic tissues. Fibrin strands lop off fragments of RBCs, whose membranes promptly reseal. Some of the erythrocyte contents leak out, however, producing varying degrees of intravascular hemolysis. The distorted RBCs are then removed by the reticuloendothelial system.

Diagnosis

Hemolysis in conjunction with typical blood smear findings is diagnostic of microangiopathic hemolysis [*see Figure 5*]. If the angiopathy is extensive, thrombocytopenia and disseminated intravascular coagulation develop. Causes include hemodynamic jets, vasculitis,[129] giant hemangiomas, thrombotic thrombocytopenic purpura, metastatic cancer,[130] certain infections (especially meningococcemia, rickettsial diseases, and hantavirus infection), hemolytic-uremic syndrome, disseminated intravascular coagulation, drugs (cocaine, cyclosporine, mitomycin, and tacrolimus), and even subclavian catheters.[130,131] Quinine has been identified

as a fairly common cause of drug-induced thrombotic thrombocytopenic purpura (TTP) and hemolytic-uremic syndrome.[132] A single case of microangiopathic hemolysis has been described in an infant with cutaneous anthrax.[133]

Treatment

In treating microangiopathic hemolysis, clinicians must focus primary attention on the underlying disease. Patients may become iron deficient and require iron therapy. Supplementation of depleted folate stores may stimulate erythropoiesis. In rare cases, anemia caused by an old prosthetic aortic valve may be severe enough to warrant valve replacement. Plasmapheresis provides effective therapy for TTP [*see 5:X Transfusion Therapy and 5:XIII Hemorrhagic Disorders*].

March hemoglobinuria March hemoglobinuria, a disorder that somewhat resembles microangiopathic hemolysis, usually occurs in young persons after prolonged marching or running or playing on bongo drums. The severe and repetitive trauma to the feet or hands is thought to destroy RBCs circulating in the vessels of the soles and palms. The patient notices red urine that clears in 1 day or less after the activity. Transient hemoglobinemia and hemoglobinuria without anemia, smear abnormalities, or reticulocytosis confirm the diagnosis. The use of padded shoes and the avoidance of paved surfaces may prevent recurrences in persons who continue running.

IMMUNE HEMOLYSIS

General Mechanisms

A classic, well-delineated example of immune (not autoimmune) hemolysis involves fetomaternal incompatibility at Rh locus D, in which the D-negative mother, after contact with D-positive erythrocytes, may produce an IgG anti-D antibody; the antibody crosses the placenta and attacks and destroys fetal erythrocytes. The fetus becomes jaundiced and has spherocytic erythrocytes.

The fetal RBCs, now coated with maternal IgG anti-D antibody, attach to fetal macrophages and monocytes that contain receptors for the Fc portion of these IgG molecules. Macrophagic digestion of portions of the erythrocytic membrane leads to the loss of considerable surface area. The resulting rigid spherocyte returns to the circulation and becomes trapped in the fetal reticuloendothelial system, particularly in the spleen. Hemolysis results. The IgG antibody is maximally active at 37° C; it generally cannot extensively activate the complement pathway, and it cannot agglutinate attacked RBCs suspended in saline.

The direct Coombs antiglobulin test [*see Figure 7*] is used clinically to detect IgG coating of RBCs. This test is negative in the mother, because her erythrocytes lack D antigen and thus are not coated with anti-D antibody. The indirect Coombs test [*see Figure 7*], which detects the presence of free serum antibody that reacts with RBCs, is positive for the mother's serum because she has circulating anti-D antibody. In the fetus, in contrast, the direct Coombs test is strongly positive because the fetus's RBCs, which express D antigen, are coated with maternal anti-D antibody. The results of the fetus's indirect Coombs test may be positive or negative, depending on the amount of anti-D antibody that has been transferred by the mother, the avidity of the anti-D antibody for fetal D-positive RBCs, and the availability of D antigen sites on fetal RBCs.

These antibodies are described as warm (maximum activity at 37° C [usually IgG1 or IgG3]) or cold (maximum activity at 5° C [usually IgM]). Antibodies have also been classified as complete (i.e., capable of agglutinating saline-suspended RBCs) and incomplete (i.e., incapable of agglutinating saline-suspended RBCs); their detection requires the use of techniques such as the direct Coombs antiglobulin test [*see Figure 7*] or enzyme treatment of RBCs.[134] Warm autoantibodies are usually incomplete, whereas cold agglutinins, which are for the most part IgM, are usually complete.[135]

Autoimmune Hemolytic Anemia

Autoimmune hemolytic anemia is generally an acute disorder characterized by extravascular hemolysis. Intravascular hemolysis in this condition is rare and indicates that an extremely rapid rate of erythrocyte destruction is occurring or that the extravascular removal mechanisms have been overwhelmed.

Pathophysiology In autoimmune hemolytic anemia, for reasons that are unclear, autoantibodies form and are directed against central components of the erythrocyte (e.g., Rh antigen, Kell antigen,[134] glycophorin A).[136] Alternatively, the patient's RBCs are sensitized with both an IgG antibody and a complement component, usually C3d. In other circumstances, however, it appears that complement is fixed to the RBC surface by an IgM antibody that is subsequently washed away. Occasionally, the RBCs exhibit only complement components, and no IgG can be detected by the Coombs test. Complement fixation in such cases may be explained by the continued presence of IgG at a level below that detectable in the usual direct antiglobulin test; alternatively, a complement-fixing IgG or IgM antibody had been attached to the cell but was eluted in the testing procedure.[135]

The severity of hemolysis correlates with the number and class of IgG and, in rare cases, IgA molecules attached to the RBC surface. Antibody-coated RBCs attach to the macrophages' receptors (FcRI, FcRII, or FcRIII) by the antibody's Fc portion. The firm binding of RBCs to these macrophage receptors is then followed either by removal of a portion of the RBC membrane, which results in the production of a spherocyte, or by phagocytosis of the entire RBC.[134] Relatively low levels of IgG1 attachment to RBCs produce a positive result on direct Coombs antiglobulin testing without evidence of hemolysis (approximately 1,000 molecules per RBC), whereas much higher levels of IgG1 autoantibody per RBC are associated with frank hemolysis.[134] The combined presence of IgG and complement components may enhance the severity of hemolysis.

Erythrocytes sensitized to IgG alone are usually removed in the spleen, whereas RBCs sensitized to IgG and complement or to complement alone are generally destroyed in the liver, because hepatic Kupffer cells carry receptors specific for complement component C3b.

Differential diagnosis Both an idiopathic variety of autoimmune hemolytic anemia and a variety that occurs secondary to other disorders have been described. Such primary disorders include systemic lupus erythematosus, non-Hodgkin lymphoma (especially chronic lymphocytic leukemia), Hodgkin disease, cancer, myeloma, dermoid cyst, HIV infection, angioimmunoblastic lymphadenopathy with dysproteinemia, hepatitis C,[137] and chronic ulcerative colitis.

Diagnosis Patient presentations vary markedly, from asymptomatic to severe. A person may be found to have a positive Coombs test result when undergoing blood bank or blood

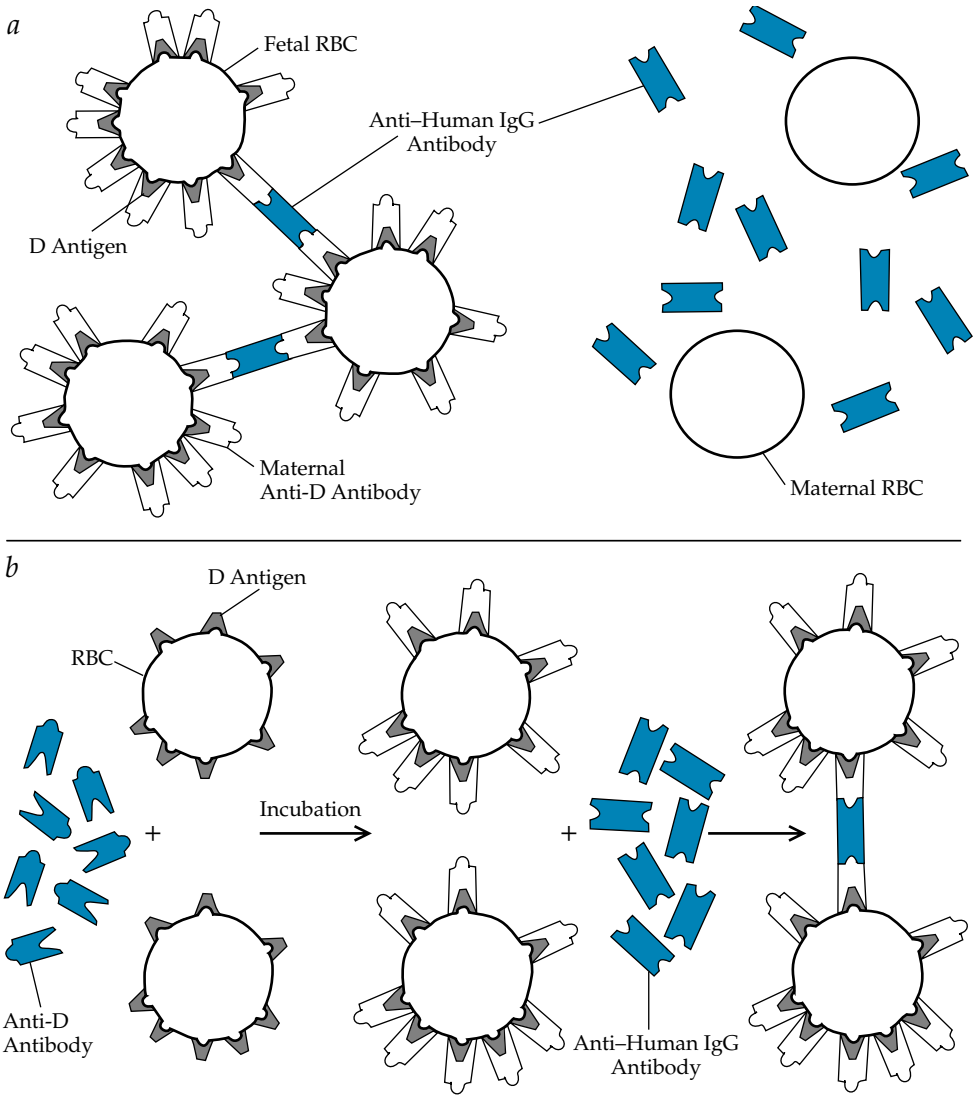


Figure 7 **The Coombs test detects the presence of human antibodies or complement components on erythrocytes or the presence of antibodies in serum. The test is useful in diagnosing Rh hemolytic disease of the newborn, autoimmune hemolysis, or potential hemolytic transfusion reactions. The figure illustrates Rh hemolytic disease of the newborn.**

In the direct Coombs test (*a*), fetal erythrocytes (RBCs) are shown with D antigen attached to their surfaces. Maternal anti-D antibody binds to the fetal erythrocytes at the D antigen sites in utero. Coombs antiserum, which contains antibody to human IgG, binds to the anti-D antibody on a sample of washed fetal erythrocytes, causing them to clump (a positive reaction). Washed maternal erythrocytes, having no D antigen, will have no attached anti-D antibody and are therefore not clumped by Coombs serum (a negative reaction).

In the indirect Coombs test (*b*), maternal or fetal serum is added to the red cells of another person or to panels of erythrocytes of known antigenic specificity; Coombs antiserum is then added. Clumping in this case occurs only if the test serum, such as the maternal serum, contains anti-D antibody and if the red cells chosen have the D antigen.

The direct Coombs test is used to detect immunoglobulin molecules already attached to erythrocytes, such as those found on the fetal erythrocytes in Rh hemolytic disease of the newborn or in autoimmune hemolytic anemia. Therefore, the test is done on the patient's thoroughly washed erythrocytes. The indirect Coombs test is used for determining whether specific antibodies are present in a serum sample, and it is performed on the patient's serum.

donation screening. Such persons can usually be shown to have complement or the combination of complement and IgG (usually IgG1 or IgG4) on their RBCs, but they are generally not undergoing hemolysis. By contrast, an acute hemolytic episode can lower the hematocrit from 45% to 15% in 2 days. With this extreme presentation, severe fatigue and cardiorespiratory symptoms will develop, together with jaundice, lymphadenopathy, and hepatosplenomegaly.

In severe cases, the blood smear shows macrocytosis, polychromatophilia, variable spherocytosis, and autoagglutination of RBCs. The platelet count is also occasionally depressed (Evans syndrome), and there may be leukopenia. One third of patients

may have reticulocytopenia at presentation.[138] The direct Coombs test will be positive. Any or all of these findings may be absent in mild disease.

Whether complement, IgG, or both are present on RBCs should be determined by the use of Coombs reagents that are specifically directed against IgG, IgA, or complement components. Occasionally, an autoimmune hemolytic anemia is suspected, but the direct Coombs test is repeatedly negative; in such cases, the level of autoantibody may be below the level of detectability for very active autoantibodies, such as subclass IgG3 autoantibodies, or the autoantibody may be IgA or IgM.[134]

Patients with evidence of hemolytic anemia should be screened for autoimmune diseases (e.g., systemic lupus erythematosus) and other forms of hemolysis, such as paroxysmal nocturnal hemoglobinuria, cold agglutinin disease, and paroxysmal cold hemoglobinuria.

Treatment Treatment of clinically affected patients is directed at decreasing autoantibody production and reducing the macrophagic attack on the RBCs. Initial therapy usually consists of 60 to 100 mg of prednisone a day, given in divided doses. This approach usually produces a slow decrease in antibody coating of RBCs and is thought to interfere with phagocytic attack on coated erythrocytes. A good response to corticosteroid use—indicated by a rise in the reticulocyte count and an improvement in hemoglobin and hematocrit—may be apparent within 1 or 2 days. Supplementation with 1 mg of folic acid a day is recommended.

After the initial response to therapy, which is usually satisfactory, the hemoglobin level and reticulocyte count may return to normal. The Coombs test is then repeated to determine whether the response has become weaker; if so, the prednisone dosage is tapered cautiously. Approximately 20% of patients remain well indefinitely, but the majority suffer from a chronic, treacherous disease that can produce sudden relapses with abrupt anemia. The prednisone should be titrated in accordance with the hemoglobin level, the reticulocyte count (elevation indicates continued hemolysis), and the direct Coombs titer; alternate-day therapy should be considered to minimize steroid side effects. If patients do not respond to standard prednisone therapy, high-dose dexamethasone (e.g., 40 mg/day orally for 4 consecutive days in 28-day cycles[139]) may be effective.

If the corticosteroid dose required for long-term therapy produces significant morbidity, one can proceed empirically either to splenectomy or to the use of immunosuppressive agents. Measurements of splenic sequestration of chromium-51 (^{51}Cr)-labeled erythrocytes do not reliably indicate the benefits of splenectomy. Splenectomy rarely results in extended remission but is valuable as a prednisone-sparing measure. After splenectomy, low-dose prednisone (5 to 10 mg/day) may stabilize the hemoglobin concentration.

The immunosuppressive agent azathioprine or cyclophosphamide can be used as an alternative to splenectomy. There is no reliable evidence to support the use of one of these agents over the other. For patients with very aggressive disease, cyclosporine has been used successfully.[140] Azathioprine should be started at a dosage of 100 to 200 mg a day; the peripheral blood count should be monitored with a view toward preventing reticulocytopenia or neutropenia. Cyclophosphamide is started at a dosage of 100 to 200 mg a day, with monitoring of blood counts and urine; however, because cyclophosphamide can cause therapy-related acute myeloid leukemia or myelodysplastic syndrome, its use should be limited [*see 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*].

Azathioprine or cyclophosphamide doses have to be adjusted to reduce the white cell count to about 3,000/mm^3. Improvement usually comes in 3 to 4 weeks. When a response occurs, the prednisone dose can be reduced and the hemoglobin level, reticulocyte count, and Coombs titer monitored to determine the minimally required therapy. For patients with very refractory disease, therapy with intravenous cyclophosphamide at doses used for allogeneic bone marrow transplantation has been tried. This approach is clearly myelotoxic, and its usefulness awaits further confirmation. High-dose intravenous IgG has been used to treat autoimmune hemolytic anemia. In one report, only one third of patients had a transient response, and doses larger than those used in idiopathic thrombocytopenic purpura (i.e., 1g/kg/day for 5 days) were required.[141,142]

There are anecdotal reports that rituximab is useful in treating refractory and relapsing cases.[143,144] However, a case of automimmune hemolytic anemia occurring after rituximab therapy for lymphoproliferative disorder has been reported.[145]

Patients with symptomatic anemia require an RBC transfusion, but often the blood bank reports an incompatibility. Many blood banks regularly perform a direct Coombs antiglobulin test on the recipient's RBCs. A patient who has free antibodies in the serum will exhibit very extensive and broad reactivity against donor panels of RBCs and will usually produce an incompatible major cross-match when tested with the antiglobulin reagent. If transfusions are needed to support cardiorespiratory and central nervous system functions, immediate consultation with the transfusion medicine service is recommended.[135] No patient should be allowed to die because the blood bank does not have a perfectly compatible unit of RBCs. If transfusion is clinically indicated, the physician should administer the best units of blood that are available, because it has been shown that these patients can tolerate even imperfectly matched RBCs.[146]

Drug-Related Immune Hemolysis

Drug-initiated immune hemolysis is often indistinguishable from autoimmune hemolytic anemia. There are two variants: the hapten type and the hemolysis that results from alteration of a membrane antigen.[147]

Hapten type Drugs such as the penicillins and the cephalosporins bind firmly to the erythrocyte membrane. In rare circumstances in which massive dosages of the drug (e.g., more than 10 million units of penicillin a day) are required, the protein-bound drug may act as a hapten and elicit an immune response. An IgG antibody that appears to be directed against the drug–RBC complex is produced[148,149]; this leads to a positive result on direct Coombs testing with the anti-IgG reagent and a negative result with the anti-C3d reagent. When the offending drug is stopped, hemolysis ends in a few days. In contrast, the drug may be bound loosely to produce a neoantigen that generates the immune response.[147] In this circumstance, the result of direct Coombs testing with the anti-C3d reagent is usually positive, and the result of testing with the anti-IgG reagent may be negative.

If the patient's serum is tested against normal cells (i.e., the indirect Coombs test is used), no reaction occurs unless the offending drug and a source of complement are first added to the normal RBCs. Stopping or switching the drug is effective in eliminating the hemolysis because the antibody is usually very specific.

Alteration of a membrane antigen Some drugs may alter a membrane antigen, thereby stimulating the production of IgG antibodies that cross-react with the native antigen. Methyldopa is the classic example of a drug that causes autoimmune hemolytic anemia. Other examples are levodopa, mefenamic acid, and procainamide. Drug administration leads to a positive direct Coombs test with anti-IgG reagents in 15% to 20% of treated patients, but hemolysis occurs in fewer than 1%. The eluted antibody is seen to be a classic IgG autoantibody directed against Rh components. The mechanism of hemolysis is identical to that of autoimmune hemolytic anemia.[150]

The NSAID diclofenac sodium has been reported to cause a devastating acute hemolytic anemia, with evidence of intravascular and extravascular hemolysis accompanied occasionally by shock, organ failure, and even disseminated intravascular coagulation.[151] Patients develop both RBC autoantibodies and drug-dependent antibodies. It is thought that diclofenac sodium binds to the surface of RBCs, forming neoantigens that lead to the generation of true autoantibodies, as well as drug-dependent antibodies. The direct Coombs test is positive with both the IgG and the C3d reagents. Additional antibody reactivity occurs with the addition of diclofenac sodium metabolites obtained from the urine of patients treated with the drug. Therapy consists of recognizing the cause, stopping the diclofenac sodium, and supporting the patient for several days until the process stops.[151]

Delayed Hemolysis of Transfused Erythrocytes

Blood is usually typed only for ABO and Rh-D antigens, but other antigens are also present on RBCs. Thus, a patient who receives extensive transfusions over 1 to 2 weeks may develop an antibody response to one or more of these other antigens. Kell, Duffy, Kidd, and Rh antigens other than D are the usual offenders. When the patient with antibodies receives RBCs expressing these antigens, an acute self-limited hemolysis, usually extravascular, may ensue. Clues are a history of transfusion, spherocytosis on peripheral smear, a positive direct Coombs test, and the recent appearance of an antibody in the patient's serum (positive indirect Coombs test). Usually, no therapy is required, but further transfusions should be cross-matched with the patient's serum [*see 5:X Transfusion Therapy*]. Similar problems arise with transplantation of bone marrow and other tissue.[152]

Cold Agglutinin Disease

Cold agglutinin disease has several variants. One rare variant affects young adults and usually occurs after infection with *Mycoplasma pneumoniae* or infectious mononucleosis, although several cases have also been reported in association with chronic falciparum malaria. A more common variant affects persons about 60 years of age and may present as idiopathic cold agglutinin disease, as a prodrome to a lymphoproliferative or an immunoproliferative disorder, or in association with an already established lymphoproliferative disorder.[153]

Pathophysiology Serologically, cold agglutinin disease is characterized by the presence of high titers of IgM agglutinins (> 1:1,000 and usually > 1:10,000) in serum. These antibodies are maximally active at 4° C, are capable of activating the complement sequence, and are directed against the polysaccharide antigens. Presumably, IgM reacts with erythrocytes circulating in the cooled blood of the nose, ears, and shins, where it fixes complement and then dissociates from the RBCs when they reach warmer areas of the body.

In the postinfectious variety of this disorder, IgM cold agglutinin is oligoclonal and short-lived. Conversely, the IgM is monoclonal in chronic idiopathic cold agglutinin disease or in cases associated with Waldenström macroglobulinemia, chronic lymphocytic leukemia, or other lymphomas. IgM predominantly contains λ light chains in patients with chronic idiopathic cold agglutinin disease or Waldenström macroglobulinemia; in patients with lymphoma, however, the IgM mainly contains κ light chains. Occasionally, the IgM cold agglutinin is detectable as an M protein spike on serum protein electrophoresis [*see 12:XV Chronic Lymphoid Leukemias and Plasma Cell Disorders*].

In the post-*Mycoplasma* variant, the mycoplasmas appear to bind to the RBC surface at the Ii antigen site. This receptor-ligand interaction results in the presentation of the I antigen in an immunogenic form.[154] *Listeria monocytogenes* contains the I antigen,[153] further supporting the idea that some infectious agents stimulate the naturally occurring cold agglutinins, as well as cause the postinfectious cold agglutinin disease.

Diagnosis The clinical syndrome of cold agglutinin disease is quite variable. Patients occasionally show only low titers of cold agglutinins and have no other symptoms or have a history of recent pneumonia. In patients with warm-and-cold autoimmune hemolytic anemia, the associated hemolysis tends to be severe and chronic. The RBCs of these patients are coated with IgG and complement components, whereas their serum contains a relatively low titer of cold agglutinin that acts at 30° C and perhaps even at temperatures up to 37° C.

The diagnosis is suggested by hemolytic anemia with acral signs and symptoms. It may be difficult to draw blood, and the RBCs may visibly agglutinate in a cold syringe and on the blood smear. Automated blood cell counters may count the agglutinated RBCs as single cells and thus report absurdly high values for the MCV and MCHC. Usually, the laboratory detects a broadly active cold agglutinin. The direct Coombs test is positive with anticomplement reagents but infrequently positive with anti-IgG.

Findings that support the diagnosis of idiopathic cold agglutinin disease include a high IgM cold agglutinin titer with broad thermal reactivity[134] and I specificity (reacting with erythrocytes from adults but not with cord erythrocytes), pure κ light-chain composition, occasionally an absolute serum IgM elevation, and an M protein pattern on serum protein electrophoresis. Investigation should be directed at discovering a possible lymphoma or other underlying disorder in these patients. Conversely, post-*Mycoplasma* and post–infectious mononucleosis cold agglutinins are polyclonal. The post–infectious mononucleosis antibody is usually directed against i antigens (cord RBCs).

Treatment The post-*Mycoplasma* or the post–infectious mononucleosis variant is usually mild and self-limited and requires no specific treatment. Patients with the idiopathic variety who have acral symptoms must change their way of life, either by moving to a warmer climate or by keeping their ears, nose, hands, and feet covered during cold weather. In severely anemic patients, transfusions with packed RBCs may be required; in such patients, careful cross-matching and warming of the blood is necessary to minimize cold agglutination.

Splenectomy and corticosteroids are generally of no benefit in controlling hemolysis associated with cold agglutinin disease. Presumably, complement-coated cells are removed to a substantial degree by hepatic rather than splenic macrophages, and the cells that produce IgM are relatively insensitive to the effects of

corticosteroids. Occasionally, however, high doses of corticosteroids (e.g., 100 mg of prednisone a day) have resulted in a reduction in the hemolytic rate in patients with relatively low titers of cold agglutinins. In the relatively rare variant caused by IgG cold agglutinins, corticosteroids and splenectomy may be of benefit. Use of penicillamine or other reducing agents containing sulfhydryl groups produces no benefit. Good responses are occasionally obtained by the use of chlorambucil at a dosage of 4 to 6 mg/day. Exchange transfusion and plasmapheresis appear to be logical therapies for acute disease, but further clinical studies are needed to evaluate these techniques. Interferon alfa, at a dosage of 3 million U/m^2 three times weekly, was reported to produce an impressive drop in cold agglutinin titer, with a decrease in serum IgM monoclonal protein and in acral symptoms over a 1-month period.[134] Treatment with rituximab in the doses used to treat non-Hodgkin lymphoma has been beneficial.[155,156]

Paroxysmal Cold Hemoglobinuria

Patients with the rare disorder of paroxysmal cold hemoglobinuria have cold-induced signs and symptoms of intravascular hemolysis. The hemolysis is associated with the presence of an IgG serum antibody that is directed against the RBC's P system. The IgG antibody is best demonstrated by the Donath-Landsteiner test; the serum is mixed either with the patient's own blood cells or with blood cells from a normal person. The mixture is chilled to 4° C. If the IgG antibody associated with this disorder is present, hemolysis occurs after warming to 37° C. In the past, paroxysmal cold hemoglobinuria was usually seen as a complication of syphilis, but it has recently been observed in association with viral infections and non-Hodgkin lymphoma.[157]

HYPERSPLENISM

Hypersplenic disorders constitute a diverse group of clinical conditions sharing the common features of splenomegaly and hemolysis. Splenic enlargement and hemolysis occur in many disorders, including hepatic cirrhosis with congestive splenomegaly, Gaucher disease, lymphoma, connective tissue disorders, Felty syndrome, sarcoidosis, tuberculosis, and other infectious diseases.

Pathophysiology

The spleen's unique structure accounts for several of the pathophysiologic features of hypersplenism. Splenic arterioles have a few direct branches leading to the sinusoids, but most of the terminal arterioles open into the splenic cords. Blood cells pass from the cords to the pulp through slits in the sinus walls; the slits have dimensions of about 1 by 3 μm.[158] Blood cells must squeeze through the longitudinal spaces, which are lined with reticular fibers, and between adventitial cells that are located outside the sinus. Macrophages and endothelial cells line the inside of the sinus. Repeated intimate contact occurs between blood cells and these macrophages.

Blood flow in the spleen is slow. The erythrocyte's pH and oxygen tension level fall, glucose is consumed, and the cell's metabolism is impaired. The hematocrit may increase, further elevating viscosity and resistance to flow. As a consequence, the blood cells are exposed to metabolic and mechanical stresses in the presence of macrophages and other leukocytes that can recognize cell membrane damage. As erythrocytes age, phagocytes remove defective surface areas, transforming the biconcave erythrocytes into rigid spherocytes or RBC fragments; these particles are later trapped and removed by the reticuloendothelial system. A big spleen has a greater than normal blood flow and exposes an unusually large proportion of blood cells to its culling activities. Thus, the problem in hypersplenism is essentially a quantitative one. A vicious circle may evolve in patients undergoing hemolysis, because hemolysis itself may cause splenomegaly.

Diagnosis

If the spleen is not palpable but the clinical situation is strongly suggestive of splenomegaly, ultrasonography or CT scanning may prove useful. Because blood cells other than erythrocytes are affected by a large spleen, the patient may be pancytopenic. Unless the underlying disease specifically involves the bone marrow, the marrow of patients with hypersplenism is generally hyperplastic because of rapid regeneration of all affected cell lines. The peripheral blood smear is not diagnostic of hypersplenism.

Treatment

If hypersplenism is producing clinically significant complications and if therapy for the patient's primary disease does not shrink the spleen, splenectomy may be necessary. Anemia, however, is not necessarily attributable to hypersplenism, irrespective of the size of the spleen. Hemodilution is another possible mechanism. Patients with massive splenomegaly who have very low hematocrit and hemoglobin values may have a normal RBC mass as assessed with the ^{51}Cr technique. Massive splenomegaly often is associated with an increase in plasma volume that results in extraordinary hemodilution. Moreover, greatly enlarged spleens may contain a pool of erythrocytes that constitutes as much as 25% of the total RBC mass—in contrast to normal spleens, which have no such RBC pool. In patients with splenomegaly who have a true decrease in RBC mass, the underlying disease may act to reduce RBC production by suppressing erythropoietin production rather than by accelerating destruction. Therefore, it is prudent to determine RBC mass before making the diagnosis of hypersplenism.

DRUGS AND TOXINS AS CAUSES OF HEMOLYSIS

Drugs Causing Oxidative Attack

Pathogenesis Dapsone, sulfasalazine, phenacetin, sodium perchlorate, nitroglycerin, phenazopyridine, primaquine,[100] paraquat, and vitamin K analogues can insert themselves into the oxygen-binding cleft of hemoglobin. By this action, such agents can generate oxidizing free radicals, such as superoxide, hydroxyl free radical, and peroxide. If the erythrocyte's protective reducing mechanisms are overwhelmed [*see Table 1*], hemoglobin is oxidized to form Heinz bodies and methemoglobin. Sulfhemoglobin is also produced by oxidative attack. The molecule contains a sulfur atom in the porphyrin ring, which gives it a blue-green color. The source of the sulfur atom is not clear, but the presence of sulfur in the heme ring makes it a poor oxygen transporter.[159] The RBC membrane may also suffer from oxidative attack. Damaged cells are removed in the reticuloendothelial system. Hemolysis is usually, but not invariably, extravascular, and Heinz bodies can be seen on a specially stained blood smear. The smear may also show the bite, hemiblister, or cross-bonded cells typical of oxidative attack on erythrocytes [*see Figure 4*]. Severe oxidative damage apparently causes hemoglobin to puddle at one side of the RBC, leaving a plasma membrane–enclosed hemighost in the remainder. Such hemighosts can be detected in the peripheral blood. Severe oxidative destruction is associated with increased methemoglobin levels and a decrease in RBC levels of GSH. The methemoglobin level is elevated. As little as 1.5 g/dl of methemoglobin or

0.5 g/dl of sulfhemoglobin can produce the physical finding of cyanosis. By contrast, 5 g/dl of reduced deoxyhemoglobin is required to produce comparable cyanosis.[100]

Nitrites can oxidize hemoglobin to methemoglobin. Consequently, the recreational use of butyl and isobutyl nitrites as stimulants, psychedelics, and aphrodisiacs has led to clinical problems. When inhaled in usual amounts, these agents may produce a mild to modest increase in methemoglobin, raising its concentration from the normal level of 1% to 2% to as much as 20%. More extensive inhalation or ingestion of these agents has induced severe methemoglobinemia, characterized by methemoglobin levels approaching 62%. Because methemoglobin does not carry oxygen, these high levels are accompanied by manifestations of tissue hypoxia such as headache, shortness of breath, lethargy, and stupor. Physical examination shows tachycardia, postural hypotension, and cyanosis; the venous blood is purple-brown.[160] If untreated, it is likely to be fatal.

Diagnosis Diagnosis is based on a history of exposure to an oxidant drug or other toxin, together with characteristic peripheral blood smear findings and elevated methemoglobin measurements.

Treatment Treatment should restore normal methemoglobin levels. Management starts with the identification and withdrawal of the offending agent. Patients who have severe methemoglobinemia should be treated immediately with 1 to 2 mg/kg of methylene blue; the agent is infused intravenously in a 1 g/dl solution over a 5-minute period. In the presence of the RBC enzyme NADPH-methemoglobin reductase and adequate amounts of the electron donor NADPH [*see Table 1*], methylene blue is rapidly reduced to leukomethylene blue. This product in turn quickly reduces methemoglobin to hemoglobin. Cyanosis is thereby reversed, and the patient should turn pink immediately after the infusion. Several hours later, however, the patient may again become cyanotic, presumably because nitrates released from tissues reenter the peripheral blood at that time. Readministration of methylene blue at a dosage of 1 mg/kg intravenously over a 5-minute period should restore normal hemoglobin levels.

Successful methylene blue therapy requires adequate supplies of NADPH. Patients who have abnormalities of the pentose phosphate pathway, such as G6PD deficiency, will not respond to this approach and should receive emergency exchange transfusions.[160] Patients with very high levels of methemoglobin (at least 60%) or those whose smears contain many hemighosts should undergo exchange transfusion, perhaps with hemodialysis.[100,160]

Lead-Induced Hemolysis

Lead exposure results in hypertensive encephalopathy, neuropathy, and hemolytic anemia characterized by coarse basophilic stippling in RBCs. The mechanism of lead-induced hemolysis is complex because the metal has several actions: it blocks heme synthesis, thus causing a buildup of RBC protoporphyrin; it produces a deficiency of pyrimidine 5′-nucleotidase[161]; and it attacks erythrocyte membrane phospholipids, producing potassium leak and interfering with Na^+,K^+-ATPase activity.

Diagnosis Screening for lead poisoning entails measuring the free erythrocyte protoporphyrin level (sometimes called the zinc protoporphyrin level), which is elevated because lead blocks the last step in heme synthesis. The diagnosis is confirmed by measuring blood and urine lead levels.

Treatment After the exposure to lead is stopped, use of a chelating agent such as edetate calcium disodium ($CaNa_2EDTA$) may be considered. Treatment is started with 0.5 to 1 g of intravenous $CaNa_2EDTA$, given over a period of 6 to 8 hours; the compound is given daily for 5 days.

After this initial course, 0.5 g of $CaNa_2EDTA$ is given as an intravenous bolus or intramuscular injection every 2 days for 2 weeks, during which time the urine lead levels are monitored. Alternatively, the initial 5-day course of $CaNa_2EDTA$ can be followed with oral penicillamine: 1 g a day is given for the first 7 days; the drug is withheld for the next 7 days; and during the final 7 days of the regimen, the dosage of 1 g a day is resumed and the urine lead level is measured at the end of the final day. Another study recommends giving 500 mg of penicillamine a day and continuing this dosage for 60 days after the patient has become asymptomatic.[162]

VENOMS AND PHYSICAL AGENTS AS CAUSES OF HEMOLYSIS

Agents Causing Enzymatic Attack

Classic examples of attacking enzymes are the snake-venom or clostridial lecithinases (e.g., phospholipase C). Such enzymes attack the phospholipids of the membrane bilayer and produce RBC fragmentation, spherocytosis, and intravascular and extravascular hemolysis. Disseminated intravascular coagulation and shock may occur. Prompt recognition and management of the primary disorder is critical, as is supportive therapy.

Venom from the brown spider, *Loxosceles intermedia,* releases sphingomyelinases and metalloproteinases that cleave the RBC membrane glycophorins. This in turn facilitates complement activation and lysis of affected RBCs.[163]

Physical Causes of Hemolysis

Freshwater drowning and accidental intravenous administration of sterile water can cause intravascular hemolysis by osmotic lysis. In such cases, RBCs swell and become spheroidal. Saltwater drowning can induce hemolysis by desiccating RBCs. Burns cause temperature-mediated denaturation of erythrocyte membrane polypeptides, resulting in hemolysis.

Infectious Diseases Causing Hemolysis

Malaria is the most important infectious cause of hemolysis. The resulting severe anemia causes the death of large numbers of pregnant women and 2- to 5-year-old children in sub-Saharan Africa. *Plasmodium* species, particularly *P. falciparum*, directly parasitize and destroy RBCs, but the anemia is a complex blend of impaired RBC production, hemolysis of parasitized and nonparasitized[130] RBCs, and ineffective erythropoiesis.[164] The diagnosis is made by pathognomonic findings on the blood smear; treatment is directed against the malarial parasite, with support of the circulation with RBC transfusions if required [*see 7:XXXIV Protozoan Infections*].

Other infectious causes of hemolysis Infection with *M. pneumoniae* and infectious mononucleosis can cause cold agglutinin hemolysis. Infection with *H. influenzae* type b can cause hemolysis. The major virulence factor of *H. influenzae*, polyribose ribosyl phosphate (PRRP), allows the organism to escape phagocytosis. When PRRP is released into the circulation, it binds to RBCs. The binding of anti-PRRP antibodies then leads to complement-dependent hemolysis.[165] Patients infected with

HIV or cytomegalovirus may have autoimmune hemolytic anemia (see above).[166]

Clostridial sepsis can be devastating; the appearance of free plasma hemoglobin or hemoglobinuria should suggest this often fatal infection. *Clostridium* species are capable of sudden, explosive growth; they can release many enzymes, including phospholipases and proteases, that digest RBCs, producing intravascular hemolysis.

Some infections can cause splenomegaly and hypersplenic hemolysis. Meningococcemia or overwhelming gram-negative septicemia often produces disseminated intravascular coagulation and microangiopathic hemolysis.

Babesiosis is caused by a parasite that invades RBCs and that is transmitted from its rodent reservoir by the same ixodid tick that carries Lyme disease and human granulocytic ehrlichiosis. This disease is being more frequently diagnosed, particularly in New England. Immunocompromised persons, such as those with HIV, are more likely to have chronic and severe infections. The diagnosis has been made on peripheral blood smears, but polymerase chain reaction methods are more sensitive.[167]

HEMOLYSIS ASSOCIATED WITH LIVER DISEASE

Anemia in patients with liver disease is often the result of a production defect rather than hemolysis, but cirrhotic patients may have congestive splenomegaly with hypersplenic hemolysis. Macrocytes (with or without B_{12} or folate deficiency) and target cells (caused by cholesterol elevation) are also common findings in such cases.

Spur cell anemia Severe liver disease, including alcoholic cirrhosis, may result in the formation of irregularly spiculated RBCs known as spur cells (acanthocytes).[168] Spur cells have alterations in their membranes (a decreased ratio of phospholipids to cholesterol[169]) that shorten their survival, resulting in hemolytic anemia.

OTHER CAUSES OF HEMOLYSIS

Copper Accumulation

In rare instances, Wilson disease, a metabolic disorder associated with excessive copper deposition, is first detected during a coincident episode of dramatic, acute hemolysis. The release of free copper into the serum and its subsequent entry into RBCs are thought to be the underlying hemolytic mechanism. In addition to affecting hexokinase levels, the intracellular copper appears to cause formation of oxygen radicals that react with and oxidize membrane components. Although no successful therapeutic intervention has been reported, penicillamine can be given at a dosage of 2 to 4 g once a day orally to reduce the free copper level. The administration of 1,000 to 2,000 IU of vitamin E (α-tocopherol) a day for several days may also be helpful if oxidative attack is an important factor.

Cardiopulmonary Bypass

Free plasma hemoglobin increases after cardiopulmonary bypass. The increase is thought to be caused by activation of the complement pathway, leading to deposition of the C5b-C9 attack complex on the RBC surface [*see 6:IV Disorders of the Complement System*].[170]

The author has served as a consultant for Tularek Corp. and Receptron, Inc., during the past 12 months.

References

1. Liu S-C, Derick LH: Molecular anatomy of the red blood cell membrane skeleton: structure-function relationships. Semin Hematol 29:231, 1992

2. Cohen CM, Gascard P: Regulation and post-translational modification of erythrocyte membrane and membrane-skeletal proteins. Semin Hematol 29:244, 1992

3. Schwartz RS: PIG-A: the target gene in paroxysmal nocturnal hemoglobinuria (editorial). N Engl J Med 330:283, 1994

4. Canessa M: Red cell volume–related ion transport systems in hemoglobinopathies. Hematol Oncol Clin North Am 5:495, 1991

5. Hanspal M, Palek J: Biogenesis of normal and abnormal red blood cell membrane skeleton. Semin Hematol 29:305, 1992

6. Turrini F, Mannu F, Arese P, et al: Characterization of the autologous antibodies that opsonize erythrocytes with clustered integral membrane proteins. Blood 81:3146, 1993

7. Potter CG, Potter AC, Hatton CSR, et al: Variation of erythroid and myeloid precursors in the marrow and peripheral blood of volunteer subjects infected with human parvovirus (B19). J Clin Invest 79:1486, 1987

8. Smith BD, Segel GB: Abnormal erythrocyte endothelial adherence in hereditary stomatocytosis. Blood 89:3451, 1997

9. Gallagher PG, Chang SH, Rettig MP, et al: Altered erythrocyte endothelial adherence and membrane phospholipid asymmetry in hereditary hydrocytosis. Blood 101:4625, 2003

10. Vives Corrons JL, Besson I, Aymerich M, et al: Hereditary xerocytosis: a report of six unrelated Spanish families with leaky red cell syndrome and increased heat stability of the erythrocyte membrane. Br J Haematol 90:817, 1995

11. Ogburn PL Jr, Ramin KD, Danilenko-Dixon D: In utero erythrocyte transfusion for fetal xerocytosis associated with non-immune hydrops fetalis. Am J Obstet Gynecol 185,238, 2001

12. Palek J, Jarolim P: Clinical expression and laboratory detection of red blood cell membrane protein mutations. Semin Hematol 30:249, 1993

13. Dhermy D, Galand C, Bournier O, et al: Heterogenous band 3 deficiency in hereditary spherocytosis related to different band 3 gene defects. Br J Haematol 98:32, 1997

14. Delhommeau F, Cynober T, Schischmanoff, et al: Natural history of hereditary spherocytosis during the first year of life. Blood 95:393, 2000

15. Rosse WF: Hematopoiesis and the defect in paroxysmal hemoglobinuria. J Clin Invest 100:953, 1997

16. Ohashi H, Hotta T, Ichikawa A, et al: Peripheral blood cells are predominantly chimeric of affected and normal cells in patients with paroxysmal nocturnal hemoglobinuria: simultaneous investigation on clonality and expression of glycophosphatidylinositol-anchored proteins. Blood 83:853, 1994

17. RostiV: The molecular basis of paroxysmal nocturnal hemoglobinuria. Haematologica 85:82, 2000

18. Hillmen P, Lewis SM, Bessler M, et al: Natural history of paroxysmal nocturnal hemoglobinuria. N Engl J Med 333:1253, 1995

19. Socie G, Marie J-Y, de Gramont A, et al: Paroxysmal nocturnal haemoglobinuria: long-term follow-up and prognostic factors. Lancet 348:573, 1996

20. Hugel B, Socié G, Vu T, et al: Elevated levels of circulating procoagulant microparticles in patients with paroxysmal nocturnal hemoglobinuria and aplastic anemia. Blood 93:3451, 1999

21. Ray JG, Burows RF, Ginsberg JS, et al: Paroxysmal nocturnal hemoglobinuria and the risk of venous thrombosis: review and recommendations for management of the pregnant and nonpregnant patient. Haemostasis 30:103, 2000

22. Hall SE, Rosse WF: The use of monoclonal antibodies and flow cytometry in the diagnosis of paroxysmal nocturnal hemoglobinuria. Blood 87:5332, 1996

23. Sholar PW, Bell WR: Thrombolytic therapy for inferior vena cava thrombosis in paroxysmal nocturnal hemoglobinuria. Ann Intern Med 103:539, 1985

24. Saso R, Marsh J, Cevreska L, et al: Bone marrow transplants for paroxysmal nocturnal haemoglobinuria. Br J Haematol 104:392, 1999

25. Balleari E, Gatti AM, Mareni C, et al: Recombinant human erythropoietin for long-term treatment of anemia in paroxysmal nocturnal hemoglobinuria. Haematologica 81:143, 1996

26. Schubert J, Scholz C, Geissler RG: G-CSF and cyclosporine induce an increase of normal cells in hypoplastic paroxysmal nocturnal hemoglobinuria. Ann Hematol 74:225, 1997

27. Arese P, De Flora A: Pathophysiology of hemolysis in glucose-6-phosphate dehydrogenase deficiency. Semin Hematol 27:1, 1990

28. Beutler E: G6PD deficiency. Blood 84:3613, 1994

29. Mockenhaupt FP, Mandelkow J, Till H: Reduced prevalence of *Plasmodium falciparum* infection and of concomitant anaemia in pregnant women with heterozygous G6PD deficiency. Trop Med Int Health 8:118, 2003

30. Neuberger A, Flishman S, Golic A: Hemolytic anemia in G6PD-deficient man after inhalation of amyl nitrite ("poppers"). Isr Med Assoc J 4:1085, 2002

31. Nagel RL, Ranney HM: Genetic epidemiology of structural mutations of the beta-globin gene. Semin Hematol 27:342, 1990

32. Bunn HF: Pathogenesis and treatment of sickle cell disease. N Engl J Med 337:762, 1997

33. Steinberg MH, Rodgers GP: Pathophysiology of sickle cell disease: role of cellular and genetic modifiers. Semin Hematol 38:299, 2001

34. Hebbel RP: Beyond hemoglobin polymerization: the red blood cell membrane and sickle disease pathophysiology. Blood 77:214, 1991

35. Gladwn MT, Lancaster JR Jr, Freeman BA, et al: Nitric oxide's reactions with hemoglobin: a view through the SNO-storm. Nat Med 9:496, 2003

36. Kuypers FA, Lewis RA, Hua M, et al: Detection of altered membrane phospholipid asymmetry in subpopulations of human red blood cells using fluorescently labeled annexin V. Blood 87:1179, 1996

37. Solovey A, Gui L, Key NS, et al: Tissue factor expression by endothelial cells in sickle cell anemia. J Clin Invest 101:1899, 1998

38. Hargrave DR, Wade A, Evans JP, et al: Nocturnal oxygen saturation and painful sickle cell crises in children. Blood 101:846, 2003

39. Singhal A, Doherty JF, Raynes JG: Is there an acute-phase response in steady-state sickle cell disease? Lancet 341:651, 1993

40. Hebbel RP: Adhesive interactions of sickle erythrocytes with endothelium. J Clin Invest 99:2561, 1997

41. Grigg AP: Granulocyte colony-stimulating factor-induced sickle cell crisis and multiorgan dysfunction in a patient with compound heterozygous sickle cell/beta+ thalassemia. Blood 97:3998, 2001

42. Adler BK, Salzman DE, Carabasi MH, et al: Fatal sickle cell crisis after granulocyte colony-stimulating factor administration. Blood 97:3313, 2001

43. Okpala I, Daniel Y, Haynes R, et al: Relationship between the clinical manifestations of sickle cell disease and the expression of adhesion molecules on white blood cells. Eur J Haematol 69:135, 2002

44. Turhan A, Weiss LA, Mohandas N: Primary role for adherent leukocytes in sickle cell vascular occlusion: a new paradigm. Proc Natl Acad Sci USA 99:3047, 2002

45. Francis RB, Johnson CS: Vascular occlusion in sickle cell disease: current concepts and unanswered questions. Blood 77:1405, 1991

46. Steinberg MH, West MS, Gallagher D, et al: Effects of glucose-6-phosphate dehydrogenase deficiency upon sickle cell anemia. Blood 71:748, 1988

47. Nagel RL, Steinberg MH: Role of epistatic (modifier) genes in the modulation of the phenotypic diversity of sickle cell anemia. Pediatr Pathol Mol Med 20:123, 2001

48. Ballas SK: Neurobiology and treatment of pain. Sickle Cell Disease: Basic Principles and Clinical Practice. Embury SH, Hebbel RP, Mohandas N, et al, Eds. Raven Press, New York, 1994, p 745

49. Brugnara C, de Franceschi L, Alper SL: Inhibition of Ca^{2+}-dependent K^+ transport and cell dehydration in sickle erythrocytes by clotrimazole and other imidazole derivatives. J Clin Invest 92:520, 1993

50. Bennekou P, de Franceschi L, Pedersen O, et al: Treatment with NS3623, a novel Cl^- conductance blocker, ameliorates erythrocyte dehydration in transgenic SAD mice: a possible new therapeutic approach for sickle cell disease. Blood 97:1451, 2001

51. Platt OS, Brambilla DJ, Rosse WF, et al: Mortality in sickle cell disease: life expectancy and risk factors for early death. N Engl J Med 330:1639, 1994

52. Bunn HF: Induction of fetal hemoglobin in sickle cell disease. Blood 93:1787, 1999

53. Charache S, Terrin ML, Moore RD, et al: Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. N Engl J Med 332:1317, 1995

54. Steinberg MH, Barton F, Castro O, et al: Hydroxyurea (HU) is associated with reduced mortality in adults with sickle cell anemia (abstr). J Am Soc Hematol 96:485a, 2000

55. Bridges KR, Barabino GD, Brugnara C, et al: A multiparameter analysis of sickle erythrocytes in patients undergoing hydroxyurea therapy. Blood 88:4701, 1996

56. Perrine SP, Ginder GD, Faller DV, et al: A short-term trial of butyrate to stimulate fetal-globin-gene expression in the beta-globin disorders. N Engl J Med 328:81, 1993

57. Atweh GF, Sutton M, Nassif I, et al: Sustained induction of fetal hemoglobin by pulse butyrate therapy in sickle cell disease. Blood 93:1790, 1999

58. Walters MC, Patience M, Leisenring W, et al: Bone marrow transplantation for sickle cell disease. N Engl J Med 335:369, 1996

59. Adams RJ, McKie VC, Hsu L, et al: Prevention of a first stroke by transfusions in children with sickle cell anemia and abnormal results on transcranial Doppler ultrasonography. N Engl J Med 339:5, 1998

60. Vichinsky EP: Understanding the morbidity of sickle cell disease (correspondence). Br J Haematol 99:974, 1997

61. Ohene-Frempong K: Indications for red cell transfusion in sickle cell disease. Semin Hematol 38(1 suppl 1):5, 2001

62. Olivieri NF: Progression of iron overload in sickle cell disease. Semin Hematol 38(1 suppl 1):57, 2001

63. Cohen AR, Martin MB: Iron chelation therapy in sickle cell disease. Semin Hematol 38(1 suppl 1):69, 2001

64. Bellet PS, Kalinyak KA, Shukla R, et al: Incentive spirometry to prevent acute pulmonary complications in sickle cell diseases. N Engl J Med 333:699, 1995

65. Milner PF, Kraus AP, Sebes JI, et al: Sickle cell disease as a cause of osteonecrosis of the femoral head. N Engl J Med 325:1476, 1991

66. Koate P: Cardiovascular pathology in sickle cell anemia. Bull Acad Natl Med 175:1055, 1991

67. Norris S, Johnson CS, Haywood LJ: Sickle cell anemia: does myocardial ischemia occur during crisis? J Natl Med Assoc 83:209, 1991

68. Castro O, Hoque M, Brown BD: Pulmonary hypertension in sickle cell disease: cardiac catheterization results and survival. Blood 101:1257, 2003

69. Vichinsky E, Williams R, Das M, et al: Pulmonary fat embolism: a distinct cause of severe acute chest syndrome in sickle cell anemia. Blood 83:3107, 1994

70. Vichinsky EP, Styles LA, Colangelo LH, et al: Acute chest syndrome in sickle cell disease: clinical presentation and course. Blood 89:1787, 1997

71. Dean D, Neumayr L, Kelly DM, et al: *Chlamydia pneumoniae* and acute chest syndrome in patients with sickle cell disease. J Pediatr Hematol Oncol 25:46, 2003

72. Styles LA, Schalkwijk CG, Aarsman AJ, et al: Phospholipase A_2 levels in acute chest syndrome of sickle cell disease. Blood 87: 2573, 1996

73. Gaston MH, Verter JI, Woods G, et al: Prophylaxis with oral penicillin in children with sickle cell anemia: a randomized trial. N Engl J Med 314:1593, 1986

74. Haberkern CM, Neumayr LD, Orringer EP, et al: Cholecystectomy in sickle cell anemia patients: perioperative outcome of 364 cases from national preoperative transfusion study. Blood 89:1533, 1997

75. Serjeant GR: Sickle-cell disease. Lancet 350:725, 1997

76. Gupta AK, Kirchner KA, Nicholson R, et al: Effects of α-thalassemia and sickle polymerization tendency on the urine-concentrating defect of individuals with sickle cell trait. J Clin Invest 88:1963, 1991

77. Falk RJ, Scheinman J, Phillips G, et al: Prevalence and pathologic features of sickle cell nephropathy and response to inhibition of angiotensin-converting enzyme. N Engl J Med 326:910, 1992

78. Bruno D, Wigfall DR, Zimmerman SA, et al: Genitourinary complications of sickle cell disease. J Urol 166:803, 2001

79. Opkala I, Westerdale N, Jegede T: Etilefrine for the prevention of priapism in adult sickle cell disease. Br J Haematol 118:918, 2002

80. Ohene-Frempong K, Weiner SJ, Sleeper LA, et al: Cerebrovascular accidents in sickle cell disease: rates and risk factors. Blood 91:288, 1998

81. Cohen AR: Sickle cell disease: new treatments, new questions. N Engl J Med 339:42, 1998

82. Dobson SR, Holden KR, Nietert PJ: Moyamoya syndrome in childhood sickle cell disease: a predictive factor for recurrent cerebrovascular events. Blood 99:3144, 2002

83. Mohan JS, Vigilance JE, Marshall JM, et al: Abnormal venous function in patients with homozygous sickle cell (SS) disease and chronic leg ulcers. Clin Sci (Colch) 98:667, 2000

84. Groves RW, Schmidt-Lucke JA: Recombinant human GM-CSF in the treatment of poorly healing wounds. Adv Skin Wound Care 13:107, 2000

85. Serjeant GR, Serjeant BE, Thomas PW, et al: Human parvovirus infection in homozygous sickle cell disease. Lancet 341:1237, 1993

86. Koshy M, Weiner SJ, Miller ST, et al: Surgery and anesthesia in sickle cell disease. Blood 86:3676, 1995

87. Vichinsky EP, Haberkern CM, Neumayr L, et al: A comparison of conservative and aggressive transfusion regimens in the perioperative management of sickle cell disease. N Engl J Med 333:206, 1995

88. Koshy M, Burd L: Management of pregnancy in sickle cell syndromes. Hematol Oncol Clin North Am 5:585, 1991

89. Embury SH: Prenatal diagnosis. Sickle Cell Disease: Basic Principles and Clinical Practice. Embury SH, Hebbel RP, Mohandas N, et al, Eds. Raven Press, New York, 1994, p 485

90. Smith JA, Espeland M, Bellevue R, et al: Pregnancy in sickle cell disease: experience of the Cooperative Study of Sickle Cell Disease. Obstet Gynecol 87:199, 1996

91. Steinberg MH, Barton F, Castro O, et al: Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. JAMA 289:1645, 2003

92. Kark JA, Posey DM, Schumacher HR, et al: Sickle-cell trait as a risk factor for sudden death in physical training. N Engl J Med 317:781, 1987

93. Sullivan LW: The risks of sickle-cell trait: caution and common sense. N Engl J Med 317:830, 1987

94. Loukopoulos D, Voskaridou E, Kalotychou V, et al: Reduction of the clinical severity of sickle cell/beta-thalassemia with hydroxyurea: the experience of a single center in Greece. Blood Cells Mol Dis 26:453, 2000

95. Rogers ZR: Hydroxyurea therapy for diverse pediatric populations with sickle cell disease. Semin Hematol 34(3 suppl 3):42, 1997

96. Nagel RL, Fabry ME, Steiberg MH: The paradox of hemoglobin SC disease. Blood Rev 17:167, 2003

97. Nagel RL, Lawrence C: The distinct pathobiology of sickle cell–hemoglobin C disease. Hematol Oncol Clin North Am 5:433, 1991

98. Steinberg MH, Nagel RL, Brugnara C: Cellular effects of hydroxyurea in Hb SC disease. Br J Haematol 98:838, 1997

99. Rees DC, Styles L, Vichinsky EP, et al: The hemoglobin E syndromes. Ann NY Acad Sci 850:334, 1998

100. Jaffe ER: Methemoglobinemia in the differential diagnosis of cyanosis. Hosp Pract (Off Ed) 20:92, 1985

101. Charache S: Methemoglobinemia-sleuthing for a new cause (editorial). N Engl J Med 314:776, 1986

102. Manabe J, Arya R, Sumimoto H, et al: Two novel mutations in the reduced nicotinamide adenine dinucleotide (NADH)-cytochrome b5 reductase gene of a patient with generalized type, hereditary methemoglobinemia. Blood 88:3208, 1996

103. Percy MJ, Gillespie MJ, Savage G, et al: Familial idiopathic methemoglobinemia revisited: original cases reveal 2 novel mutations in NADH-cytochrome b5 reductase. Blood 100:3447, 2002

104. Weatherall DJ, Clegg JB: Thalassemia: a global public health problem. Nat Med 2:847, 1996

105. Adams JG III, Coleman MB: Structural hemoglobin variants that produce the phenotype of thalassemia. Semin Hematol 27:229, 1990

106. Schrier SL: Thalassemia: pathophysiology of red cell changes. Annu Rev Med 45:211, 1994

107. Schrier SL: Pathophysiology of thalassemia erythrocytes. Curr Opin Hematol 4:75, 1997

108. Schrier SL: Pathophysiology of thalassemia. Curr Opin Hematol 9:123, 2002

109. Amer J, Goldfarb A, Fibach E: Flow cytometric measurement of reactive oxygen

species production by normal and thalassaemic red blood cells. Eur J Haematol 70:84, 2003

110. Pootrakul P, Sirankapracha P, Hemsorach S, et al: A correlation of erythrokinetics, ineffective erythropoiesis, and erythroid precursor apoptosis in Thai patients with thalassemia. Blood 96:2606, 2000

111. Hansen RM, Hanson G, Anderson T: Failure to suspect and diagnose thalassemic syndromes: interpretation of RBC indices by the nonhematologist. Arch Intern Med 145: 93, 1985

112. Piomelli S: The management of patients with Cooley's anemia: transfusions and splenectomy. Semin Hematol 32:262, 1995

113. Hoffbrand AV: Oral iron chelation. Semin Hematol 33:1, 1996

114. Brittenham GM, Griffith PM, Nienhuis AW, et al: Efficacy of deferoxamine in preventing complications of iron overload in patients with thalassemia major. N Engl J Med 331:567, 1994

115. Olivieri NF, Nathan DG, MacMillan JH, et al: Survival in medically treated patients with homozygous β-thalassemia. N Engl J Med 331:574, 1994

116. Cohen AR, Galanello R, Piga A: Safety and effectiveness of long-term therapy with the oral iron chelator deferiprone. Blood 102:583, 2003

117. Nick H, Acklin P, Lattmann R, et al: Development of tridentate iron chelators: from desferrithiocin to ICL670. Curr Med Chem 10:1065, 2003

118. Winterbourne CC: Oxidative denaturation in congenital hemolytic anemias: the unstable hemoglobins. Semin Haematol 27:41, 1990

119. Issaragrisil S, Visuthisakchai S, Suvatte V, et al: Brief report: transplantation of cord-blood stem cells into a patient with severe thalassemia. N Engl J Med 332:367, 1995

120. Lucarelli G, Galimberti M, Giardini C, et al: Bone marrow transplantation in thalassemia. Ann NY Acad Sci 850:270, 1998

121. Gaziev J, Lucarelli G: Stem cell transplantation for hemoglobinopathies. Curr Opinion Pediatr 15:24, 2003

122. Bradai M, Abad MT, Pissard S, et al: Hydroxyurea can eliminate transfusion requirements in children with severe beta-thalassemia. Blood 102:1529, 2003

123. De Paula EV, Lima CS, Arruda VR, et al: Long-term hydroxyurea therapy in beta-thalassemia patients. Eur J Haematol 70:151, 2003

124. Domenica Cappellini M, Graziadei G, Ciceri L, et al: Oral isobutyramide therapy in patients with thalassemia intermedia: results of a phase II open study. Blood Cells Mol Dis 26:105, 2000

125. Reich S, Buhrer C, Henze G, et al: Oral isobutyramide reduces transfusion requirements in some patients with homozygous beta-thalassemia. Blood 96:3357, 2000

126. Chui DHK, Fucharoen S, Chan V: Hemoglobin H disease: not necessarily a benign disorder. Blood 101:791, 2003

127. Kazazian HH Jr: The thalassemia syndromes: molecular basis and prenatal diagnosis in 1990. Semin Hematol 27:209, 1990

128. Dover GJ, Valle D: Therapy for β-thalassemia: a paradigm for the treatment of genetic disorders. N Engl J Med 331:609, 1994

129. Ross CN, Reuter H, Scott D, et al: Microangiopathic haemolytic anaemia and systemic vasculitis. Br J Rheumatol 35:377, 1996

130. Rytting M, Worth L, Jaffe N: Hemolytic disorders associated with cancer. Hematol Oncol Clin North Am 10:365, 1996

131. Mach-Pascual S, Samii K, Beris P, et al: Microangiopathic hemolytic anemia complicating FK 506 (tacrolimus) therapy. Am J Hematol 52:310, 1996

132. Kojouri K, Vesely SK, George JN: Quinine-associated thrombotic thrombocytopenic purpura–hemolytic uremic syndrome: frequency, clinical features, and long-term outcomes. Ann Intern Med 135:1047, 2001

133. Freedman A, Afonja O, Chang MW, et al: Cutaneous anthrax associated with microangiopathic hemolytic anemia and coagulopathy in a 7-month-old infant. JAMA 287:869, 2002

134. Engelfriet CP, Overbeeke MAM, von dem Borne AEGK: Autoimmune hemolytic anemia. Semin Hematol 29:3, 1992

135. Jefferies LC: Transfusion therapy in autoimmune hemolytic anemia. Hematol Oncol Clin North Am 8:1087, 1994

136. Leddy JP, Falany JL, Kissel GE, et al: Erythrocyte membrane proteins reactive with human (warm-reacting) anti-red cell autoantibodies. J Clin Invest 91:1672, 1993

137. Ramos-Casals M, Garcia-Carrasco M, Lopez-Medrano F, et al: Severe autoimmune cytopenias in treatment-naïve hepatitis C virus infection: clinical description of 35 cases. Medicine 82:87, 2003

138. Liesveld JL, Rowe JM, Lichtman MA: Variability of the erythropoietic response in autoimmune hemolytic anemia: analysis of 109 cases. Blood 69:820, 1987

139. Meyer O, Stahl D, Beckhove P, et al: Pulsed high-dose dexamethasone in chronic autoimmune haemolytic anaemia of warm type. Br J Haematol 98:860, 1997

140. Emilia G, Messora C, Longo G, et al: Long-term salvage treatment by cyclosporin in refractory autoimmune haematological disorders. Br J Haematol 93:341, 1996

141. Collins PW, Newland AC: Treatment modalities of autoimmune blood disorders. Semin Hematol 29:64, 1992

142. Flores G, Cunningham-Rundles C, Newland AC, et al: Efficacy of intravenous immunoglobulin in the treatment of autoimmune hemolytic anemia: results in 73 patients. Am J Hematol 44:237, 1993

143. Gottardo NG, Baker DL, Willis FR: Successful induction and maintenance of long-term remission in a child with chronic relapsing autoimmune hemolytic anemia using rituximab. Pediatr Hematol Oncol 20:557, 2003

144. Trape G, Fianchi L, Lai M, et al: Rituximab chimeric anti-CD20 monoclonal antibody treatment for refractory hemolytic anemia in patients with lymphoproliferative disorders. Haematologica 88:223, 2003

145. Jourdan E, Topart D, Richard B, et al: Severe autoimmune hemolytic anemia following rituximab therapy in a patient with a lymphoproliferative disorder. Leuk Lymphoma 44:889, 2003

146. Salama A, Berghofer H, Mueller-Eckhardt C: Red blood cell transfusion in warm-type autoimmune haemolytic anaemia. Lancet 340:1515, 1992

147. Salama A, Mueller-Eckhardt C: Immune-mediated blood cell dyscrasias related to drugs. Semin Hematol 29:54, 1992

148. Garratty G: Immune cytopenia associated with antibiotics. Transfus Med Rev 7:255, 1993

149. Kopicky JA, Packman CH: The mechanisms of sulfonylurea-induced immune hemolysis: case report and review of the literature. Am J Hematol 23:283, 1986

150. Petz LD: Drug-induced autoimmune hemolytic anemia. Transfus Med Rev 7:242, 1993

151. Salama A, Kroll H, Wittmann G, et al: Diclofenac-induced immune haemolytic anaemia: simultaneous occurrence of red blood cell autoantibodies and drug-dependent antibodies. Br J Haematol 95:640, 1996

152. Ramsey G: Red cell antibodies arising from solid organ transplants. Transfusion 31:76, 1991

153. Silberstein LE: B-cell origin of cold agglutinins. Adv Exp Med Biol 347:193, 1994

154. Loomes LM, Uemura K, Childs RA, et al: Erythrocyte receptors for *Mycoplasma pneumoniae* are sialylated oligosaccharides of Ii antigen type. Nature 307:560, 1984

155. Berentsen S, Tjonnfjord GE, Brudevold R: Favourable response to therapy with the anti-CD20 monoclonal antibody rituximab in primary chronic cold agglutinin disease. Br J Haematol 115:79, 2001

156. Mori A, Tamaru J, Sumi H, et al: Beneficial effects of rituximab on primary cold agglutin disease refractory to conventional therapy. Eur J Haematol 68:243, 2002

157. Sharara AI, Hillsley RE, Wax TD, et al: Paroxysmal cold hemoglobinuria associated with non-Hodgkin's lymphoma. South Med J 87:397, 1994

158. Rosse WF: The spleen as a filter. N Engl J Med 317:704, 1987

159. Lu HC, Shih RD, Marcus S, et al: Pseudomethemoglobinemia: a case and review of sulfhemoglobinemia. Arch Pediatr Adolesc Med 152:803, 1998

160. Coleman MD, Coleman NA: Drug-induced methaemoglobinaemia: treatment issues. Drug Saf 14:394, 1996

161. Valentine WN, Paglia DE, Fink K, et al: Lead poisoning: association with hemolytic anemia, basophilic stippling, erythrocyte pyrimidine 5′-nucleotidase deficiency, and intraerythrocytic accumulation of pyrimidines. J Clin Invest 58:926, 1976

162. Carton JA, Maradona JA, Arribas JM: Acute-subacute lead poisoning: clinical findings and comparative study of diagnostic tests. Arch Intern Med 147:697, 1987

163. Tambourgi DV, De Sousa Da Silva M, Billington SJ, et al: Mechanism of induction of complement susceptibility of erythrocytes by spider and bacterial sphingomyelinases. Immunology 107:93, 2002

164. Nagel RL: Malarial anemia. Hemoglobin 26:329, 2002

165. Shurin SB, Anderson P, Zollinger J, et al: Pathophysiology of hemolysis in infections with *Hemophilus influenzae* type b. J Clin Invest 77:1340, 1986

166. van Spronsen DJ, Breed WPM: Cytomegalovirus-induced thrombocytopenia and haemolysis in an immunocompetent adult. Br J Haematol 92:218, 1996

167. Krause PJ, Spielman A, Telford SR III, et al: Persistent parasitemia after acute babesiosis. N Engl J Med 339:160, 1998

168. Owen JS, Brown DJC, Harry DS, et al: Erythrocyte echinocytosis in liver disease. J Clin Invest 76:2275, 1985

169. Allen DW, Manning N: Cholesterol-loading of membranes of normal erythrocytes inhibits phospholipid repair and arachidonoyl-CoA:1-palmitoyl-sn-glycero-3-phosphocholine acyl transferase: a model of spur cell anemia. Blood 87:3489, 1996

170. Salama A, Hugo F, Heinrich D, et al: Deposition of terminal C5b-complement complexes on erythrocytes and leukocytes during cardiopulmonary bypass. N Engl J Med 318:408, 1988

V THE POLYCYTHEMIAS

Virginia C. Broudy, m.d.

Classification of the Polycythemias

Polycythemia, also called erythrocytosis, is an increase in the number of circulating red blood cells per volume of blood, as reflected by an elevated hematocrit or hemoglobin level. The three major categories of polycythemia are (1) relative polycythemia, (2) secondary polycythemia, and (3) primary polycythemia, or polycythemia vera.

In relative polycythemia, the red blood cell mass is normal but the plasma volume is decreased. Secondary polycythemia is caused by an elevated erythropoietin level. Polycythemia vera is a neoplastic stem cell disorder characterized by an autonomous overproduction of red blood cells and, often, of white blood cells and platelets [*see 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*].

Initial Evaluation

Patients are often asymptomatic, and the elevated hemoglobin or hematocrit level is usually discovered accidentally. When such an increase is found, it should be promptly evaluated to determine its cause [*see Figure 1*]. Any family history of polycythemia and the results of any previous hematocrit determinations should be obtained. History and physical examination findings suggestive of congenital heart disease, severe chronic obstructive pulmonary disease (COPD), or sleep apnea syndrome should be sought, and the presence or absence of splenomegaly should be determined. The results of the complete blood count, including the platelet count and white blood cell differential, should be critically reviewed for abnormalities. Findings of leukocytosis, thrombocytosis, an occasional circulating immature white blood cell, or increased basophils are suggestive of polycythemia vera and argue against secondary causes of erythrocytosis.

A hematocrit level of 60% or higher in a man or 57% or higher in a woman virtually always indicates a true increase in red blood cell mass (i.e., primary or secondary polycythemia). If the patient's hematocrit is below these values but above normal, a red blood cell mass study is required.[1] To perform this study, a sample of the patient's red blood cells is labeled with radioactive chromium (^{51}Cr) ex vivo and injected back into the patient. A second blood sample is then obtained to quantitate the concentration of ^{51}Cr-labeled red blood cells among the unlabeled red blood cells. In parallel, the patient is given an injection of albumin labeled with radioactive iodine (^{125}I) to measure the plasma volume. If a contraction of plasma volume is noted, the patient has relative polycythemia; if an increase in red blood cell mass is noted, the patient has true (i.e., primary or secondary) polycythemia. Once a relative or an absolute increase in red blood cell mass is documented, an exact diagnosis should be determined [*see Figure 1*].

Relative Polycythemia

Patients with relative polycythemia (Gaisböck syndrome) are often obese, hypertensive men who may also be heavy smokers[2]; they often are 45 to 55 years of age—a decade younger than typical for polycythemia vera patients [*see 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*]. It has been estimated that 0.5% to 0.7% of the healthy male population in the United States have relative polycythemia. Diuretic use for treatment of hypertension may exacerbate the deficit in plasma volume, and smoking-induced high carboxyhemoglobin levels or hypoxemia may also play a role.

Relative polycythemia is usually mild (hematocrit lower than 55%). In patients with a hematocrit lower than 60%, this diagnosis should be considered, and the red blood cell mass and plasma volume should be measured [*see Figure 1*] to avoid an extensive and ultimately frustrating workup for other causes of polycythemia. Patients with relative polycythemia fall into two major groups: (1) those with normal red blood cell mass and clearly decreased plasma volume and (2) those with red blood cell mass and plasma volume at the upper and lower range of normal, respectively. Behavior modification (e.g., an exercise regimen and smoking cessation) is recommended for these patients. Hematocrit returns to normal over time in approximately one third of patients.

Secondary Polycythemia

Secondary polycythemia occurs when erythropoietin production is increased as a result of chronic tissue hypoxia. Causes of tissue hypoxia include life at high altitude, high-affinity hemoglobin, cardiopulmonary disease, obstructive sleep apnea, obesity-hypoventilation syndrome, and high serum levels of carboxyhemoglobin. Polycythemia also occurs in some renal and hepatic disorders, in rare genetic disorders, and from treatment with androgens or erythropoietin.

Polycythemia caused by appropriate increases in erythropoietin production

Life at High Altitude

Initial human adaptation to high altitude includes increases in the respiratory rate, cardiac output, and the level of 2,3-bisphosphoglycerate to facilitate oxygen unloading from hemoglobin to the tissues [*see Figure 2*]. Within 6 to 24 hours after a person has ascended to a high altitude, erythropoietin levels increase, resulting in reticulocytosis within 24 to 48 hours. Over several days, serum erythropoietin levels return to normal, but the increase in hematocrit is sustained. In addition to the increase in red blood cell mass, a modest decrease in plasma volume occurs. A patient's travel history should be taken to determine the likelihood of high-altitude effect and thus possibly avoid having to conduct an extensive workup.

Life at high altitude, such as in the Rocky Mountains of North America or the Andes of South America, may result in chronic mountain sickness characterized by headaches, dizziness, mental slowing, dyspnea, and weakness. In such cases, individuals may have a hematocrit as high as 63% and are at risk for the development of pulmonary hypertension, early onset cardiovascular disease, and proteinuria. Treatment with low-dose angiotensin-converting enzyme (ACE) inhibitors (e.g., enalapril, 5 mg orally each day) can improve the hematocrit and renal function over 1 to 2 years.[3]

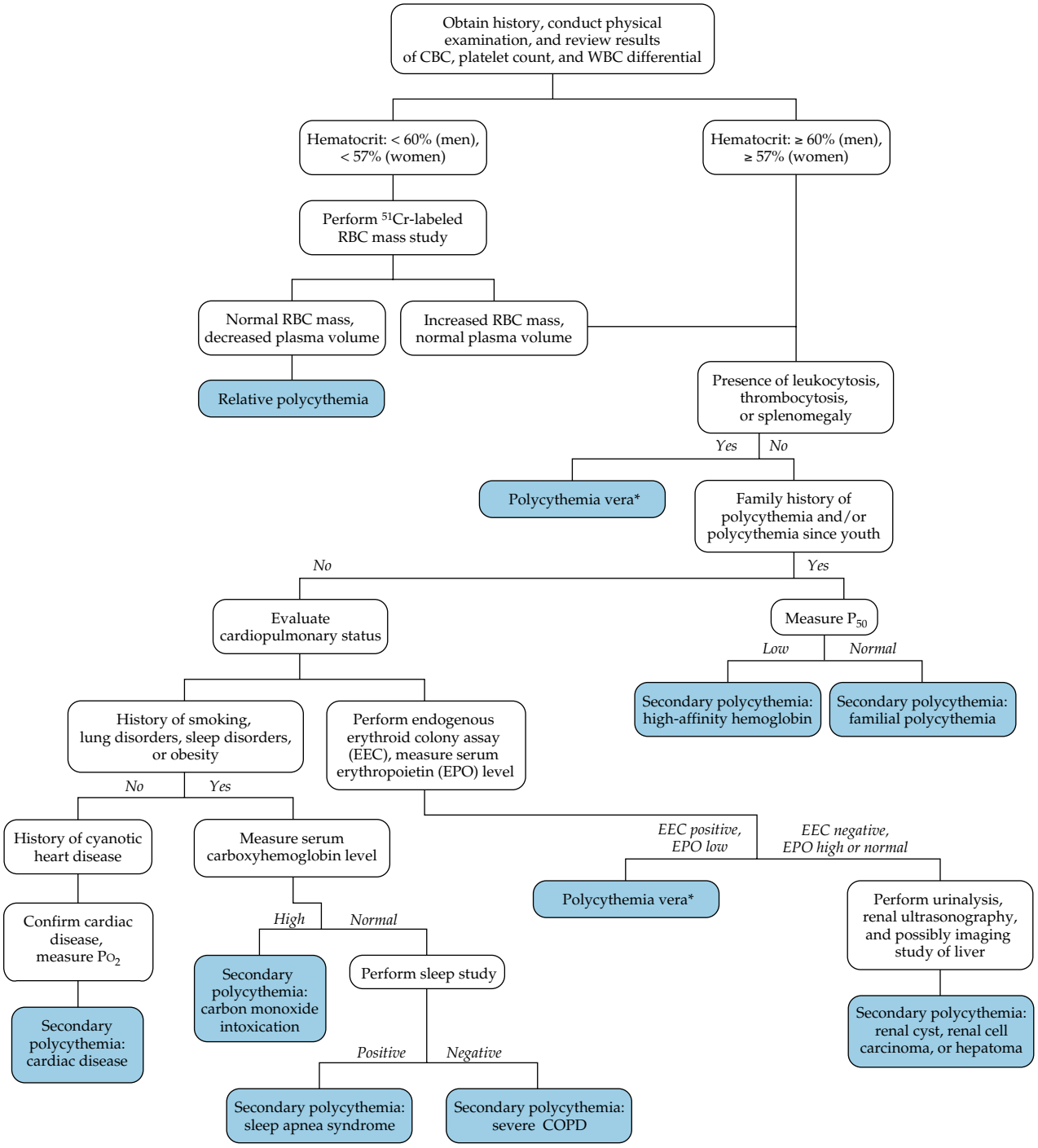


Figure 1 **This flowchart depicts an approach to the evaluation of a patient with polycythemia, as evidenced by an elevated hematocrit or hemoglobin level on routine complete blood count. (CBC—complete blood cell count; COPD—chronic obstructive pulmonary disease; EEC—endogenous erythroid colony; EPO—erythropoietin; RBC—red blood cell count; WBC—white blood cell count)**
***For coverage of polycythemia vera, see *12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*.**

High-Affinity Hemoglobin

High-affinity hemoglobin is caused by an amino acid substitution in either the α chain or, more commonly, the β chain of globin that impedes the normal conformational change during oxygen loading and unloading. This condition results in an impaired ability to release oxygen in the tissues, causing tissue hypoxia and increased erythropoietin production. More than 100 mutations causing high-affinity hemoglobin have been described. They are usually familial and are inherited in an autosomal dominant manner but are occasionally the result of spontaneous mutation. A review of the patient's medical history should show evidence of lifelong polycythemia. The hematocrit is usually less than 60%, and the white blood cell and platelet counts are normal. The partial pressure of oxygen at which hemoglobin is 50% saturated (P_{50}) should be measured; it is reduced in patients with high-affinity hemoglobin [*see Figure* 2].

Hemoglobin electrophoresis is usually not helpful, because many of the mutations that result in high-affinity hemoglobin are electrophoretically silent. However, the mutations can be identified by DNA sequencing. In rare instances, patients have congenital 2,3-bisphosphoglycerate mutase deficiency; the presentations of these patients are similar to those of patients with high-affinity hemoglobin. Patients with high-affinity hemoglobin or congenital 2,3-bisphosphoglycerate mutase deficiency usually have no symptoms of hyperviscosity and require no therapy. Phlebotomy decreases exercise tolerance in these patients and should not be used.

Cardiopulmonary Disease

Polycythemia caused by cardiopulmonary defects (e.g., Eisenmenger complex, univentricular heart, and tetralogy of Fallot) results from a failure to load oxygen onto hemoglobin adequately in the lungs.[4,5] The hematocrit may range from 60% to 75% and cause profound symptoms of hyperviscosity, including headache, dizziness, visual disturbances, fatigue, paresthesias, irritability, and decreased mental acuity. Platelet microparticles are overproduced in cyanotic congenital heart disease, especially when the hematocrit is greater than 60%; overabundance of these microparticles may contribute to the hemostatic abnormalities in these patients.[6] Some adults with cyanotic congenital heart disease have decompensated erythrocytosis, which is characterized by unstable, rising hematocrit and symptomatic hyperviscosity; these patients may benefit from phlebotomy.[5] Other adults with cyanotic congenital heart disease have compensated erythrocytosis, in which a stable (though elevated) hematocrit is maintained without overt symptoms of hyperviscosity; these patients do not require phlebotomy.[5] A practical approach is to cautiously phlebotomize patients whose hematocrits range from 60% to 65% and who have symptoms of hyperviscosity.[5] The extent of phlebotomy should be guided by the patient's symptoms. Acute dehydration, which exacerbates polycythemia, should be excluded from the diagnosis before phlebotomy is performed, and the volume of blood withdrawn should be replaced with isotonic saline. Iron deficiency should be avoided by the use of oral iron therapy if necessary because severe iron deficiency may alter red blood cell rheology and increase the risk of stroke.[7]

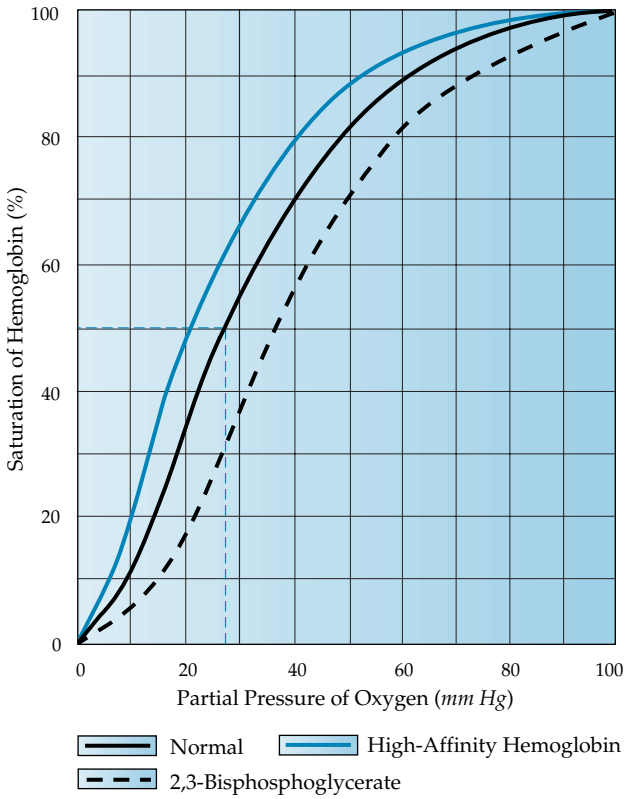


Figure 2 **Depicted is the oxygen-hemoglobin dissociation curve (solid black line). The partial pressure of oxygen at which hemoglobin is 50% saturated (P_{50}) is normally 27 mm Hg (broken blue lines). The presence of high-affinity hemoglobin shifts the curve to the left, reflecting impaired oxygen unloading in the tissues (solid blue line). An increase in the level of 2,3-bisphosphoglycerate—a feature of adaptation to high altitude—shifts the curve to the right, reflecting increased oxygen unloading in the tissues (broken black line).**

Severe COPD can be associated with polycythemia, although the clinical features of COPD usually predominate. In patients who continue to smoke, both hypoxemia and elevated carboxyhemoglobin levels may contribute to the development of polycythemia. Reduction of hematocrit in patients with significant polycythemia caused by COPD results in increased cerebral blood flow, relief from the symptoms of dizziness and headache that are associated with hyperviscosity, and dramatic improvement in mental alertness. In a study of seven patients with severe COPD and pulmonary hypertension, serial phlebotomy reduced pulmonary arterial pressure and improved exercise capacity.[8]

Obstructive Sleep Apnea

Sleep apnea syndrome is estimated to occur in 4% of middle-aged men and 2% of women and is underdiagnosed.[9] Risk factors include obesity, male sex, central body fat distribution, and a family history of obstructive sleep apnea.[10] The prevalence of sleep apnea syndrome increases as the body mass index increases. Recurrent episodes of upper airway collapse during sleep obstruct air movement, resulting in intermittent nocturnal hypoxemia. Patients may have a history of loud snoring, alternating with periods of silence lasting 10 seconds to 1 minute, followed by gasping sounds. Fragmented sleep results in excessive daytime sleepiness and impaired work performance and may increase the risk of motor vehicle accidents.[11] The hematocrit may be modestly increased in patients with severe obstructive sleep apnea, and this syndrome should be considered in patients with unexplained polycythemia. Nocturnal polysomnography with quantitation of the apnea-hypopnea index can establish the diagnosis. Management of this condition may include weight loss, nasal continuous positive airway pressure, and surgery[12] [*see 11:XIII Disorders of Sleep and 14:VI Ventilatory Control during Wakefulness and Sleep*].

Obesity-Hypoventilation Syndrome

Obesity-hypoventilation syndrome is also known as pickwickian syndrome, in reference to Charles Dickens' astute description of the obese coachboy who had excessive daytime sleepiness. Patients with this syndrome are usually morbidly obese (body mass index of 40 kg/m^2) and have chronic daytime hypoxemia and hypercapnia, in part because of a blunted ventilatory response to these stimuli.[13] Many of these patients also have nocturnal obstructive sleep apnea.[14] Hypoxemia provides the stimulus for increased erythropoietin production and polycythemia. Other clinical features associated with obesity-hypoventilation syndrome are daytime hypersomnolence and cor

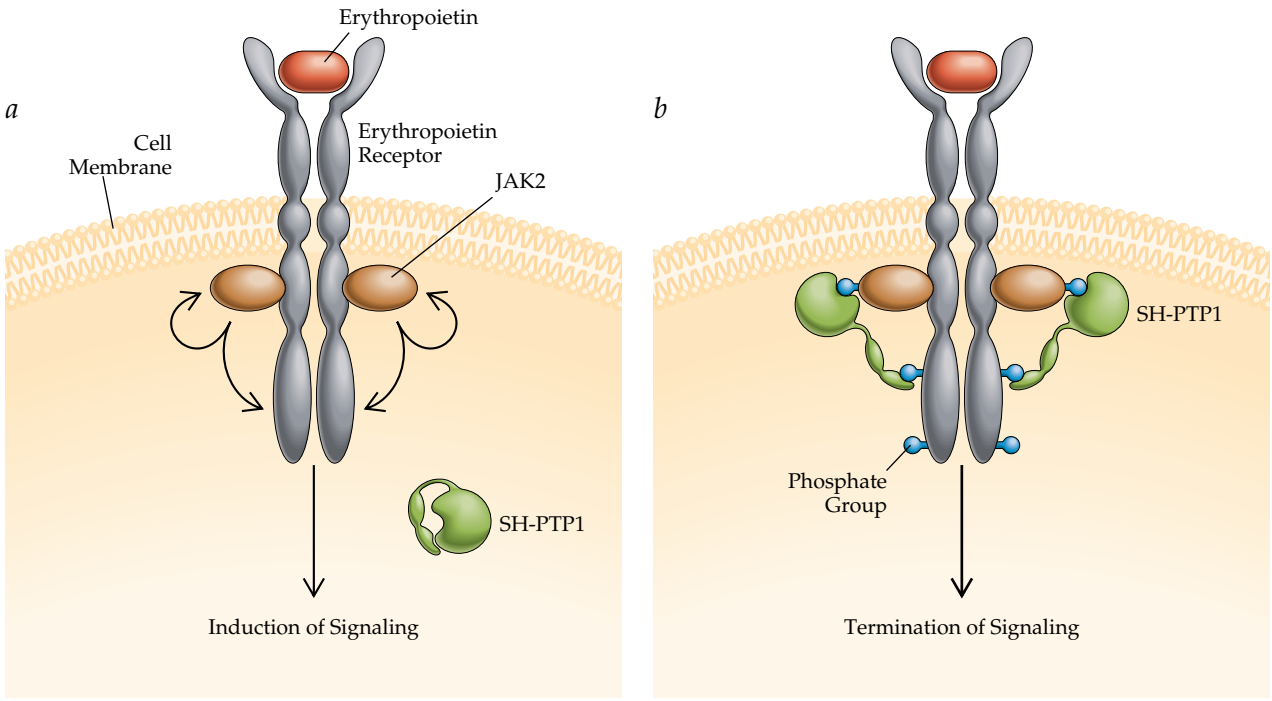


Figure 3 **(*a*) Binding of erythropoietin to the erythropoietin receptor on an erythroid progenitor cell triggers association and activation of the protein-tyrosine Janus kinase-2 (JAK2) and the initiation of signal transduction, stimulating growth of the erythroid progenitor cell. (*b*) Binding of the protein-tyrosine phosphatase SH-PTP1 results in dephosphorylation of JAK2 and termination of signal transduction.**

pulmonale. Management of this condition includes weight loss and progesterone therapy to stimulate the central respiratory drive [*see 14:VI Ventilatory Control during Wakefulness and Sleep*].

High Carboxyhemoglobin Levels

Long-term exposure to carbon monoxide results in chronic high carboxyhemoglobin levels [*see 8:I Management of Poisoning and Drug Overdose*]. Carbon monoxide binds to hemoglobin with an affinity 210 times greater than that of oxygen, decreasing the quantity of hemoglobin available for oxygen transport. Carbon monoxide binding also increases the affinity of the remaining heme groups for oxygen, shifting the oxygen-hemoglobin dissociation curve to the left [*see Figure 2*] and impairing the unloading of oxygen in the tissues. By these mechanisms, long-term carbon monoxide exposure can cause polycythemia. Cigarette and cigar smokers and persons with long-term occupational exposure to automobile exhaust in poorly ventilated areas (e.g., toll-booth operators, underground-garage attendants, and truck loaders) are at risk. The average carboxyhemoglobin level in the blood of nonsmokers is approximately 1% or less, whereas it is 4% in smokers and as high as 15% in heavy smokers.

Symptoms may include subtle neuropsychiatric abnormalities and exacerbation of angina (likely as a result of impaired myocardial oxygen delivery). The diagnosis can be established by measuring the percentage of carboxyhemoglobin in the blood. Because the half-life of carboxyhemoglobin is approximately 5 hours, the test should be done late in the day, when the patient has smoked the usual number of cigarettes or spent several hours in the work environment. Polycythemic smokers usually have both elevated red blood cell mass and decreased plasma volume. For smokers, the most effective therapy is smoking cessation; abnormal blood and plasma levels revert to normal within 3 months. No therapy is available for persons with occupational polycythemia, with the exception of avoidance of the workplace.

POLYCYTHEMIA CAUSED BY RENAL AND HEPATIC DISORDERS

Polycythemias arise when erythropoietin production is increased because of renal or, less often, hepatic disorders. In adults, approximately 90% of erythropoietin production occurs in the kidney, and 10% occurs in the liver. Because of the intricate regulation of erythropoietin production in the kidney, distortion of renal anatomy can result in polycythemia. Case reports document that renal cysts, hydronephrosis, focal glomerulonephritis, and Bartter syndrome can cause polycythemia. After renal transplantation, approximately 10% to 20% of patients have transient or persistent polycythemia. It is important to identify these patients, because they are at increased risk for arterial or venous thrombotic events and may require phlebotomy or ACE inhibitors.[15] In addition, primary malignancies of the kidney or liver can cause polycythemia. Polycythemia develops in approximately 3% of patients with renal cell carcinoma. Erythropoietin production by primary renal cell carcinoma or hepatoma tissues is the likely cause of polycythemia in these patients. In rare instances, focal nodular hyperplasia of the liver, hepatic or cerebral hemangiomas or hemangioblastomas,[16] uterine fibroids, adrenal adenomas, and pheochromocytomas have been reported to cause polycythemia. Mutations in the von Hippel–Lindau gene have been associated with cerebral hemangioblastomas or renal cell cancer, either as a part of the von Hippel–Lindau syndrome or as an acquired somatic mutation.[17]

FAMILIAL POLYCYTHEMIA

The familial polycythemias are rare diseases resulting from inborn mutations affecting hematopoietic or nonhematopoietic cells. The molecular mechanisms causing familial polycythemia may be different in different families, and mutations may be inherited in an autosomal dominant or recessive fashion.

A high frequency of autosomal recessive familial polycythemia is found in the Chuvash region of Russia.[18] Elegant

studies have demonstrated that a point mutation in the von Hippel–Lindau gene results in enhanced stability of hypoxia-inducible factor-1α, which regulates transcription of the erythropoietin gene.[19] Thus, Chuvash polycythemia is a congenital disorder of oxygen homeostasis. Chuvash polycythemia has also been identified in families of European or Asian descent.[20] Patients with Chuvash polycythemia present during the teenage years with headache, dizziness, fatigue, and dyspnea on exertion.[21] Affected members of these families have a high hematocrit (approximately 60%) and elevated levels of erythropoietin, and they have thromboembolism and cerebrovascular disorders, which shorten survival. Treatment is with phlebotomy.

In autosomal dominant familial polycythemia, abnormalities in the erythropoietin receptor have been identified in a small proportion of patients.[22-25] Erythropoietin initiates its biologic effects by binding to a specific receptor that is found on the surface of erythroid progenitor cells, precursor cells, and certain other types of cells. Binding triggers a cascade of events, including activation of the protein tyrosine Janus kinase-2 (JAK2) [*see Figure 3*]. Tyrosine phosphorylation of the erythropoietin receptor creates docking sites for other signal transduction molecules and for the protein-tyrosine phosphatase SH-PTP1, which dephosphorylates JAK2 and terminates signal transduction. In one large Finnish family with polycythemia, a point mutation in the erythropoietin receptor affecting SH-PTP1 rendered the erythroid progenitor cells hypersensitive to erythropoietin. Interestingly, one member of this family who had a hematocrit of 60% won three gold medals in cross-country skiing at the 1964 Winter Olympics. Another proportion of patients with autosomal dominant familial polycythemia have been found to have mutations in the erythropoietin receptor, resulting in deletion of the carboxyl terminus negative regulatory region of the receptor.[23-25] Individuals with autosomal dominant familial polycythemia have erythrocytosis that remains stable over time; they do not experience leukocytosis or thrombocytosis, and no long-term clinical consequences have been described.

POLYCYTHEMIA CAUSED BY DRUG USE

Androgens (e.g., testosterone) can cause polycythemia by stimulating erythropoietin production.[26] The elevation in hematocrit is usually mild, and hematocrit returns to normal 2 to 3 months after discontinuance of anabolic steroid use. Since recombinant human erythropoietin and darbepoetin have become available, concern has been raised that competitive athletes involved in endurance sports, such as bicycle racing, cross-country skiing, and long-distance running, might surreptitiously self-inject this drug to improve athletic performance.[27-29] Phlebotomy followed by blood doping is known to improve performance in runners and skiers. A similar increase in hematocrit can be achieved with erythropoietin injections, which can increase maximal exercise capacity. The unmonitored increase in red blood cell production may cause significant polycythemia, which, when coupled with exercise-induced dehydration, can have tragic consequences. Erythropoietin abuse has been linked to the deaths of competitive bicyclists.[27] The use of erythropoietin or darbepoetin to improve athletic performance is banned by the International Olympics Committee. Recombinant human erythropoietin can be detected in the urine by isoelectric focusing.[30]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Pearson TC, Guthrie DL, Simpson J, et al: Interpretation of measured red cell mass and plasma volume in adults: expert panel on radionuclides of the International Council for Standardization in Haematology. Br J Haematol 89:748, 1995

2. Messinezy M, Pearson TC: Apparent polycythaemia: diagnosis, pathogenesis and management. Eur J Haematol 51:125, 1993

3. Plata R, Cornejo A, Arratia C, et al: Angiotensin-converting-enzyme inhibition therapy in altitude polycythaemia: a prospective randomised trial. Lancet 359:663, 2002

4. Vongpatanasin W, Brickner ME, Hillis LD, et al: The Eisenmenger syndrome in adults. Ann Intern Med 128:745, 1998

5. Thorne SA: Management of polycythaemia in adults with cyanotic congenital heart disease. Heart 79:315, 1998

6. Horigome H, Hiramatsu Y, Shigeta O, et al: Overproduction of platelet microparticles in cyanotic congenital heart disease with polycythemia. J Am Coll Cardiol 39:1072, 2002

7. Ammash N, Warnes CA: Cerebrovascular events in adult patients with cyanotic congenital heart disease. J Am Coll Cardiol 28:768, 1996

8. Borst MM, Leschke M, König U, et al: Repetitive hemodilution in chronic obstructive pulmonary disease and pulmonary hypertension: effects on pulmonary hemodynamics, gas exchange, and exercise capacity. Respiration 66:225, 1999

9. Stradling JR, Davies RJO: Sleep 1: Obstructive sleep apnoea/hypopnoea syndrome: definitions, epidemiology, and natural history. Thorax 59:73, 2004

10. Young T, Skatrud J, Peppard P: Risk factors for obstructive sleep apnea in adults. JAMA 291:2013, 2004

11. Teran-Santos J, Jimenez-Gomez A, Cordero-Guevara J: The association between sleep apnea and the risk of traffic accidents. N Engl J Med 340:847, 1999

12. McNicholas WT: Obstructive sleep apnea syndrome: who should be treated? Sleep 23(suppl 4):S187, 2000

13. Martin TJ, Sanders MH: Chronic alveolar hypoventilation: a review for the clinician. Sleep 18:617, 1995

14. Kessler R, Chaouat A, Schinkewitch P, et al: The obesity-hypoventilation syndrome revisited: a prospective study of 34 consecutive cases. Chest 120:369, 2001

15. Vlahakos D, Marathias K, Agroyannis B: Perspectives in renal medicine: posttransplant erythrocytosis. Kidney Int 63:1187, 2003

16. Kuhne M, Sidler D, Hofer S, et al: Diagnosis in oncology: challenging manifestations of malignancies. J Clin Oncol 22:3639, 2004

17. Wiesener MS, Seyfarth M, Warnecke C, et al: Paraneoplastic erythrocytosis associated with an inactivating point mutation of the von Hippel-Lindau gene in a renal cell carcinoma. Blood 99:3562, 2002

18. Sergeyeva A, Gordeuk VR, Tokarev YN, et al: Congenital polycythemia in Chuvashia. Blood 89:2148, 1997

19. Ang SO, Chen H, Hirota K, et al: Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nat Genet 32:614, 2002

20. Percy MJ, McMullin MF, Jowitt SN, et al: Chuvash-type congenital polycythemia in 4 families of Asian and western European ancestry. Blood 102:1097, 2003

21. Gordeuk VR, Sergueeva AI, Miasnikova GY, et al: Congenital disorder of oxygen sensing: association of the homozygous Chuvash polycythemia VHL mutation with thrombosis and vascular abnormalities but not tumors. Blood 103:3924, 2004

22. de la Chapelle A, Traskelin AL, Juvonen E: Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. Proc Natl Acad Sci USA 90:4495, 1993

23. Arcasoy MO, Degar BA, Harris KW, et al: Familial erythrocytosis associated with a short deletion in the erythropoietin receptor. Blood 89:4628, 1997

24. Gregg XT, Prchal JT: Erythropoietin receptor mutations and human disease. Semin Hematol 34:70, 1997

25. Kralovics R, Prchal JT: Genetic heterogeneity of primary familial and congenital polycythemia. Am J Hematol 68:115, 2001

26. Besa EC: Hematologic effects of androgens revisited: an alternative therapy in various hematologic conditions. Semin Hematol 31:134, 1994

27. Tokish JM, Kocher MS, Hawkins RJ: Ergogenic aids: a review of basic science, performance, side effects, and status in sports. Am J Sports Med 32:1543, 2004

28. Vogel G: A race to the starting line. Science 305:632, 2004

29. Noakes TD: Tainted glory: doping and athletic performance. N Engl J Med 351:847, 2004

30. Breidbach A, Catlin DH, Green GA, et al: Detection of recombinant human erythropoietin in urine by isoelectric focusing. Clin Chem 49:901, 2003

Acknowledgments

Figures 1 and 2 Marcia Kammerer.

Figure 3 Jared Schneidman.

VII NONMALIGNANT DISORDERS OF LEUKOCYTES

David C. Dale, m.d.

Leukocytes, or white blood cells, protect the body against infections and participate in many types of immunologic and inflammatory responses. There are two main types of leukocytes: lymphocytes, which are responsible for antibody production and cell-mediated immunity, and phagocytes, which are responsible for the ingestion and killing of microorganisms. Neutrophils, monocytes, macrophages, and eosinophils are all phagocytes [*see Figure 1*]. Leukocytes interact with one another and modulate immune responses through the release of cytokines (interleukins and growth factors), enzymes, and vasoactive substances. This chapter covers the diagnosis of disorders of neutrophils, monocytes, and eosinophils and the treatment of neutropenia; the functions and disorders of lymphocytes are discussed elsewhere [*see Section 6 Immunology and Allergy*].

The White Blood Cell Count

The total white blood cell (WBC) count and differential count are often the first studies performed in evaluating a patient with a suspected infection or with susceptibility to infections. Most laboratories measure the WBC count using automated cell-counting techniques.[1] The normal WBC count ranges from 4,300 to 10,000/mm^3, with a median of 7,000/mm^3 [*see Table 1*]. A differential count gives the percentage for each type of leukocyte. The absolute count is determined by multiplying the total WBC count by this percentage (e.g., WBC × percent neutrophils = absolute neutrophil count). Because the blood level of each type of leukocyte is separately regulated, it is always better to use the absolute count rather than the percentage in assessing abnormalities.

Indications of the Presence of a Phagocytic Cell Disorder

Because the phagocytes, particularly neutrophils, represent the first line of defense against invading microorganisms, disorders in the number or function of these cells often result in an increased susceptibility to infection. A quantitative or qualitative disorder of phagocytic cells should be suspected when a patient has an increased number of bacterial or fungal infections, increasingly severe infections, or infections with unusual organisms.

Neutrophil Physiology

neutrophil production

Neutrophils are derived from the common stem cell, which also gives rise to erythrocytes, platelets, and other leukocytes. The proliferation and differentiation of the neutrophil precursors are governed by a family of regulatory cytokines. Granulocyte colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF) are two important cytokines affecting neutrophil production and function. G-CSF selectively stimulates progenitor cells to differentiate into neutrophils and rapidly increases blood neutrophils in hematologically normal individuals [*see Figure 2*].[2,3] GM-CSF stimulates progenitor cells to differentiate into neutrophils, eosinophils, monocytes, macrophages, and dendritic cells.[4] The life cycle of the neutrophil consists of bone marrow, blood, and tissue phases. Neutrophil production in the bone marrow takes approximately 10 to 14 days, and the bone marrow produces approximately 1×10^9 neutrophils/kg/day.[5] Most of the body's neutrophils are found in the bone marrow. The mitotic compartment, which contains about 20% of the total neutrophil pool, consists of myeloblasts (the earliest morphologically recognizable precursors), promyelocytes, and myelocytes. The postmitotic pool or maturation compartment—the metamyelocytes, bands, and mature neutrophils—contains about 70% of the body's neutrophils. The marrow neutrophils and bands are sometimes called the storage compartment or marrow reserve. As neutrophils mature, they develop the capacity to enter the blood through increasing deformability and through changes in the adhesion proteins on their surface membranes. Entry into the blood involves interactions of the mature cells and the endothelial cells of the marrow sinusoids that are not yet well understood. Agents that stimulate release of neutrophils from the marrow (e.g., G-CSF, GM-CSF, corticosteroids, or endotoxin administration) can result in a doubling of the blood neutrophil count within 3 to 5 hours. The peripheral blood contains fewer than 10% of the body's neutrophils. In the blood, the neutrophils are divided approximately evenly between the circulating pool and the marginating pool; these pools are in dynamic equilibrium. Cells in the marginating pool can be swept rapidly (within minutes) into the circulation by endogenous or exogenous epinephrine or as a result of exercise or any cause of rapid increase in cardiac output. This response, called demargination, can double the blood neutrophil count very rapidly and is also quickly reversible. The blood half-life of the neutrophils is approximately 6 to 10 hours. Neutrophils leave the blood and enter the tissues by migrating between endothelial cells and penetrating the capillary basement membrane. It is now believed that neutrophils that do not leave the circulatory system die by apoptosis and are removed by mononuclear phagocytes in the spleen, liver, and other tissues.[6]

neutrophil structure

As neutrophil precursors mature, their nuclear chromatin becomes condensed and segmented. Mature cells have no nucleoli, few mitochondria, and very little endoplasmic reticulum. The cytoplasm is filled with granules and glycogen. The primary granules, which appear at the myeloblast and promyelocyte stages, contain myeloperoxidase (MPO), proteases, defensins, and other antibacterial substances.[7,8] Secondary granules, produced primarily during the myelocyte stage, predominate in mature cells. They contain collagenase, lactoferrin, lysozyme, vitamin B_{12}–binding protein, and several other proteins. Small tertiary granules are also found in mature neutrophils. Neutrophils also may have cytoplasmic vesicles containing lactases, alkaline phosphatases, and components of nicotinamide-adenine dinucleotide phosphate (NADPH) oxidase.

The surface of the neutrophil is replete with deep folds and ruffles. On the neutrophil surface, there are numerous receptors, including receptors for immunoglobulins (e.g., FcγRII [CD32],

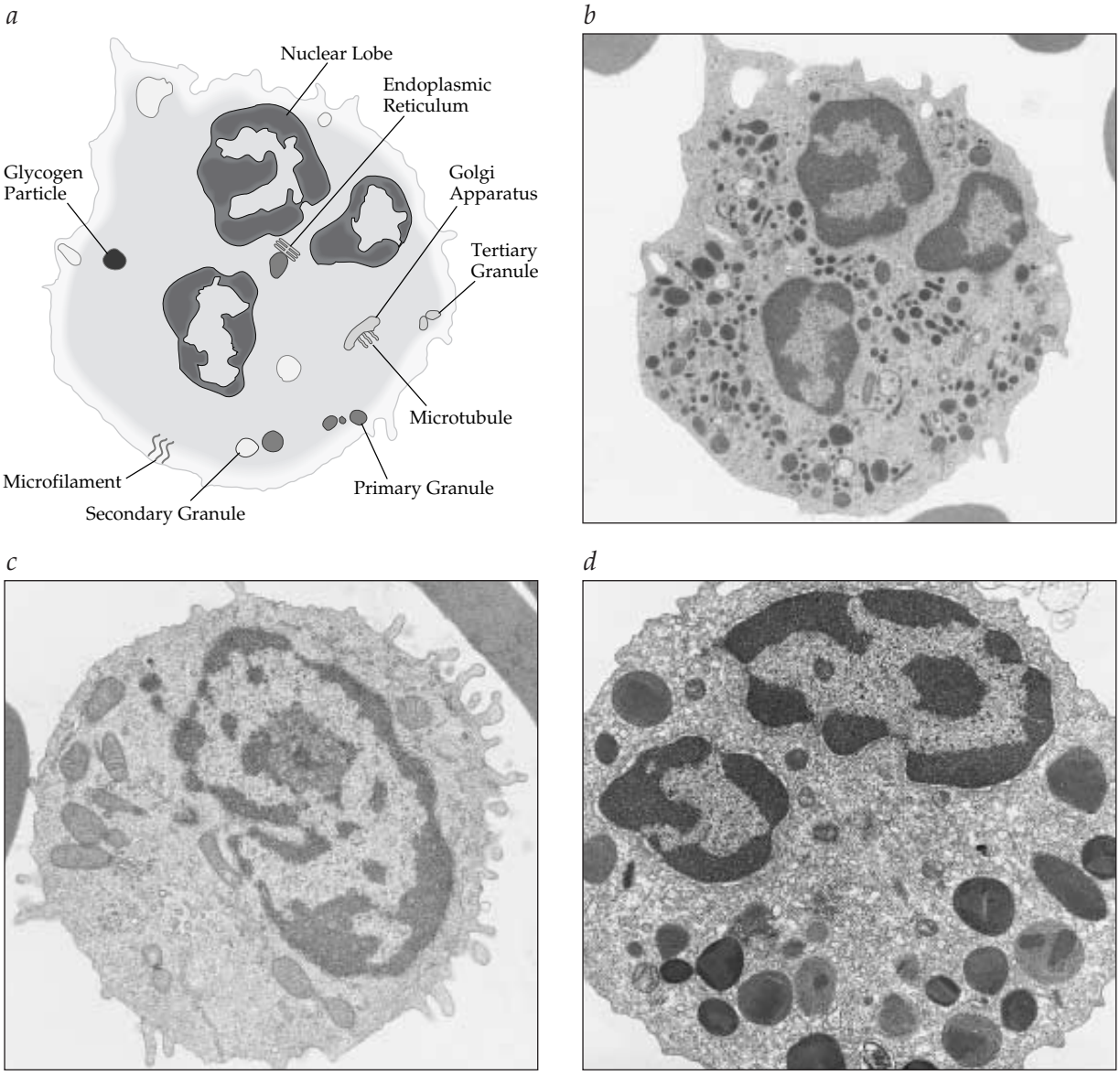


Figure 1 **Shown are a schematic diagram of a neutrophil (*a*), a corresponding electron micrograph of a neutrophil (*b*), and electron micrographs of a monocyte (*c*) and an eosinophil (*d*).**

FcγRIII [CD16]), complement (e.g., CR3 [CD11b18], CR1 [CD35]), chemokines, the colony-stimulating factors G-CSF and GM-CSF, Fas, tumor necrosis factor receptor (TNF-R), and the apoptosis-related receptors.[9]

The cytoskeleton of the neutrophil is composed of microtubules and microfilaments that are critical for phagocytic shape and movement, including migration through the vascular endothelium. The microfilaments, which consist primarily of actin polymers, are dispersed throughout the cytoplasm.[10]

NEUTROPHIL FUNCTION

The major function of neutrophils is to respond rapidly to microbial invasion to kill the invaders. This response has several distinct steps—adherence, migration, recognition, phagocytosis (or ingestion), degranulation, oxidative metabolism, and bacterial killing [*see Figure 3*]. Susceptibility to infection results from abnormalities in any one or a combination of these processes.

Adherence

For neutrophils to move to an inflammatory site, they must first adhere to a capillary wall.[11] Loose adherence is facilitated by L-selectins, such as sialyl-Lewis^x (sLe^x), on the neutrophil and E-selectin and P-selectin on capillary endothelial cells [*see Figure 4*]. Bacterial invasion increases local selectin expression and neutrophil accumulation. Other neutrophil surface proteins, called β_2 integrins, facilitate firmer adhesion to endothelial cells and interact with actin, myosin, and actin-binding proteins to initiate movement of neutrophils to the tissue.[11] The three proteins in this family have a common β subunit (CD18) and a different α subunit (CD11a, CD11b, or CD11c). There is generally increased expression of these proteins (e.g., CD 11b/C18) on neutrophils in response to inflammation. Concomitantly, there is increased expression of the intracellular adherence molecules (ICAMs) on the endothelial cells, with a net result of increased trafficking of neutrophils to the inflammatory focus.

Table 1 Normal Leukocyte Values in Peripheral Blood

Cell Type	Cells/mm³*		Percentage of Total Differential Count
	Median	Range	
All leukocytes (white blood cells)	7,000	4,300–10,000	100
Total neutrophils	4,000	1,800–7,200	55
Band neutrophils	500	100–2,000	10
Segmented neutrophils	3,500	1,000–6,000	45
Lymphocytes	2,500	1,500–4,000	36
Monocytes	450	200–900	6
Eosinophils	150	0–700	2
Basophils	30	0–150	1

*To calculate the number of cells/L, multiply by 10^6.

Chemotaxis

Chemotaxis, the directed movement of cells, occurs when neutrophils detect a chemoattractant at low concentrations and move up the concentration gradient toward its source, which is usually a site in the extravascular spaces.[12] Well-characterized stimulators of neutrophil chemotaxis are the complement proteins C5a, leukotriene B4, interleukin-8 (IL-8), and a family of small peptides, the chemokines. The trafficking of neutrophils from the blood is unidirectional; they do not return from the tissues to the circulation.

Recognition and Phagocytosis

At the site of inflammation, neutrophils utilize their immunoglobulin and complement receptors to recognize bacteria and other particles coated or opsonized by immunoglobulins or complement. Inflammation stimulates neutrophils to express increased numbers of the high-affinity IgG receptor FcγRI (CD64).[13] As the neutrophil internalizes a particle, a phagocytic vesicle, or phagosome, develops around it. This process stimulates degranulation and activates a burst of oxidative metabolism.

Degranulation

When the neutrophil is activated, the granule membranes come in contact with the plasma membranes surrounding the phagosome. The membranes fuse, which leads to the release of granule proteins into the phagosome and to the reorganization of the components of the critical NADPH oxidase system.[14]

Oxidative Metabolism and Bacterial Killing

Resting granulocytes are primarily anaerobic cells that rely on anaerobic glycolysis for adenosine triphosphate (ATP) production. Although chemotaxis, ingestion, and degranulation require

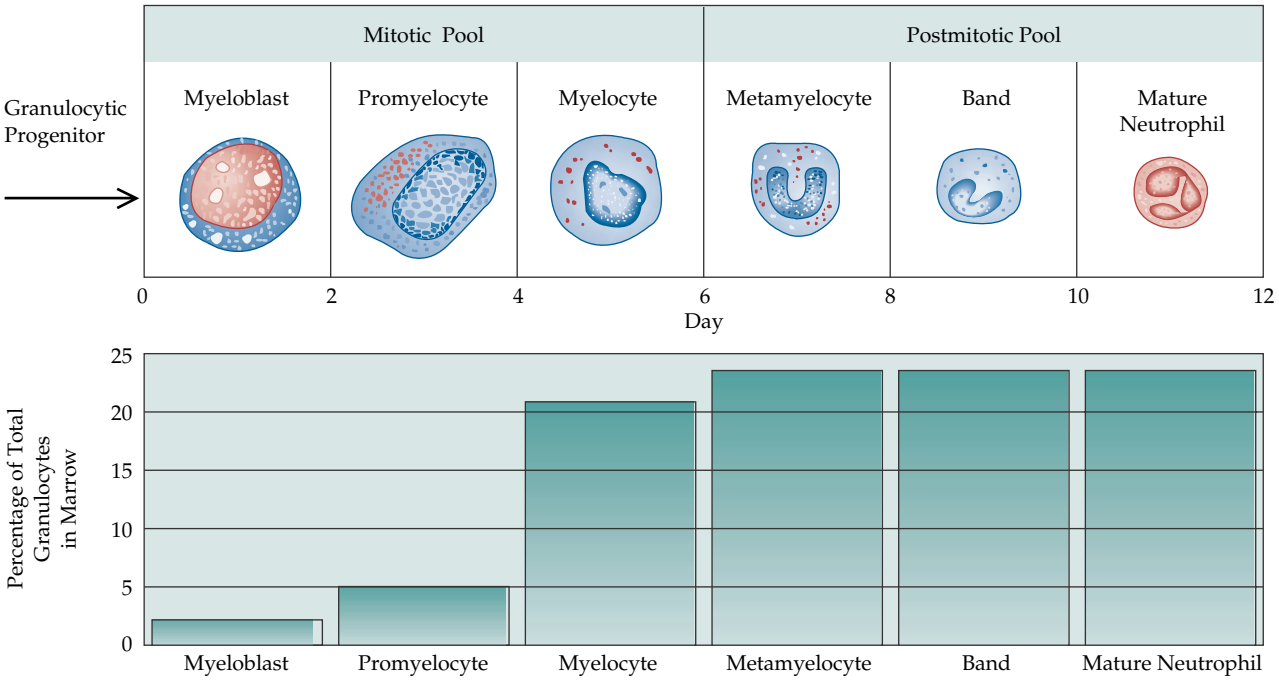


Figure 2 **The process of neutrophil maturation begins in the bone marrow (top). After about 12 days, approximately 10% of the mature neutrophils are released into the peripheral blood, where they have a half-life of approximately 6 to 10 hours. Eventually, the neutrophils migrate into the tissues by diapedesis. The percentage of neutrophils at each stage of development (bottom) ranges from about 2% at the myoblast stage to almost 25% at the mature neutrophil stage.**

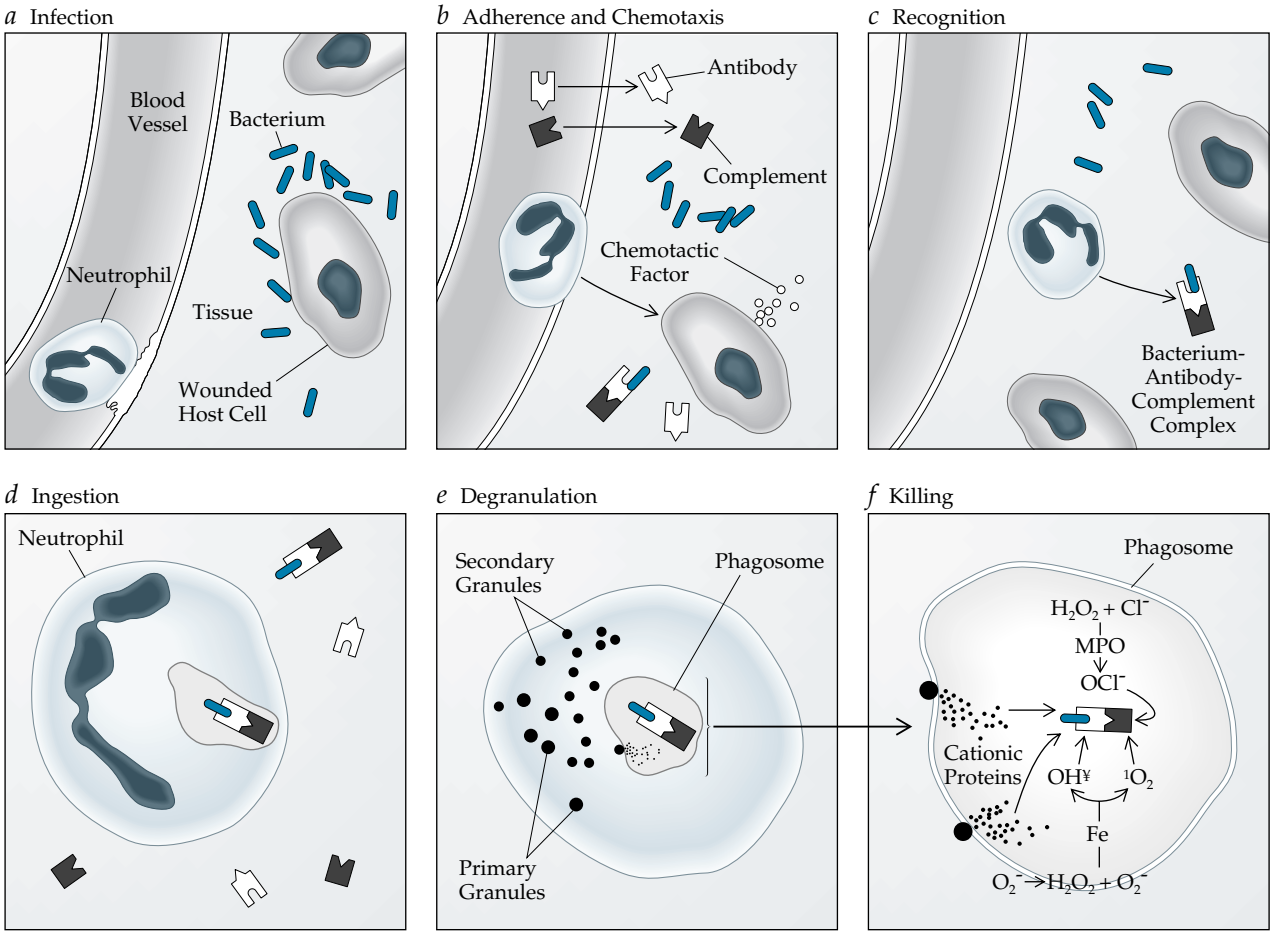


Figure 3 **The neutrophil response to bacterial invasion involves several stages. A bacterium infects a host cell and injures it (*a*). Bacterial products, antibodies, and complement cause the release of chemotactic factors, which activate a neutrophil in the adjacent blood vessel. The neutrophil adheres to the vessel wall and undergoes chemotaxis and diapedesis into tissue (*b*) to follow the chemoattractants to their sites of generation or expression. The neutrophil recognizes (*c*) and ingests (*d*) the bacterium-antibody-complement complex, forming a phagosome. The neutrophil then undergoes degranulation, a process in which granule membranes fuse with the plasma membrane (*e*). Degranulation releases various enzymes and enhances oxidative metabolism, the products of which are bactericidal (*f*). For example, hydrogen peroxide (H_2O_2), produced from superoxide (O_2^-), can interact with O_2^- in the presence of iron (Fe) to produce hydroxyl radicals ($OH\bullet$) and singlet oxygen ($^1O_2$), both of which are highly toxic to bacteria. In addition, H_2O_2 and chloride (Cl^-) combine in the presence of the myeloperoxidase (MPO) released in the phagosome to produce hypochlorite (OCl^-), which is also bactericidal.**

some energy, they also proceed quite well anaerobically. However, bacterial killing generally is associated with a rapid increase (within seconds) in oxygen use. This respiratory burst occurs as a result of the activation of an NADPH oxidase.[15] Before activation, the components of the oxidase are located separately in the plasma and granule membranes and in the cytosol. The membranes contain two components: $gp91^{phox}$ and $p22^{phox}$. The cytosol includes a p47 protein and a p67 protein. When the neutrophil is activated, the cytosolic proteins first associate and then combine with the membrane components to produce the complete NADPH oxidase. NADPH oxidase can reduce oxygen by one electron to superoxide O_2^-; in the process, NADPH is converted to $NADP^+$. The NADPH is then regenerated through the hexose monophosphate shunt. Dismutation of the superoxide in the presence of superoxide dismutase produces hydrogen peroxide (H_2O_2), which can then be converted to hydroxyl radical ($OH\bullet$). H_2O_2, O_2^-, and $OH\bullet$ are highly toxic. In addition, within the phagocyte vacuole, hydrogen peroxide and chloride (Cl^-) can combine in the presence of myeloperoxidase to produce hypochlorous acid (HOCl), which is bactericidal.[15] These products of the respiratory burst can also be released from the activated neutrophil and subsequently damage the surrounding cells and tissues.

Responses to and Production of Cytokines

The growth factors that affect neutrophil production, such as G-CSF and GM-CSF, also influence neutrophil function.[16] These cytokines upregulate stimulus-dependent NADPH oxidase activity and can enhance bactericidal and fungicidal activities. Although neutrophils contain very few ribosomes, they can respond to bacterial stimuli by synthesizing and secreting proinflammatory cytokines such as IL-1, IL-6, and tumor necrosis factor–α (TNF-α); monocytes, however, produce much larger quantities of these substances.[17]

Disorders of Neutrophil Number

NEUTROPHILIA

Neutrophilia, or granulocytosis, is usually defined as a neutrophil count greater than 10,000/mm^3.

Etiology

Neutrophilia most often occurs secondary to inflammation, stress, or corticosteroid therapy. Cigarette smoking commonly causes neutrophilia as a result of inflammation in the airways and lungs. Malignancies, hemolytic anemia, and lithium therapy are less common causes. Neutrophilia is also associated with splenectomy. Extreme neutrophilia (i.e., neutrophil counts of more than 30,000 to 50,000/mm^3), often called a leukemoid reaction, occurs with severe infections, sepsis, hemorrhagic shock, and severe tissue injury of any cause. Neutrophilia is also seen in patients with leukocyte adhesion deficiency (LAD), a rare disease in which neutrophils accumulate in the blood because they lack either the integrin CD11b18 or the selectin sLe^x (CD15s) required to leave the circulation.[18]

Serious bacterial infections and chronic inflammation are usually associated with changes in both the number of circulating neutrophils and their morphology. Characteristic changes include increased numbers of young cells (bands), of cells with residual endoplasmic reticulum (Döhle bodies), and of cells with more prominent primary granules (toxic granulation). These changes are probably caused by the endogenous production of G-CSF or GM-CSF and are also seen with administration of these growth factors.

Primary neutrophilia (i.e., neutrophilia attributed to defects in proliferation and maturation of neutrophil precursors) occurs in patients with myeloproliferative disorders, such as chronic myeloid leukemia (CML) and polycythemia vera [*see 5:VI The Polycythemias*]. Hereditary and idiopathic neutrophilias have been described; they are benign and quite rare. One such uncommon idiopathic condition is Sweet syndrome, which is an acute febrile illness with painful cutaneous plaques and associated neutrophilia of any cause.[19] Neutrophilia is also associated with congenital abnormalities. For example, infants with Down syndrome can have transient leukemoid reactions that must be distinguished from congenital leukemia.[20]

Diagnosis

When neutrophilia cannot be readily attributed to an infection or inflammatory condition or to glucocorticosteroid therapy, the possibility of a myeloproliferative disease should be considered. The presence of splenomegaly, metamyelocytes, and myelocytes in the blood, together with increased basophils or eosinophils and a low leukocyte alkaline phosphatase (LAP) score, suggests CML [*see 12:XVII Chronic Myelogenous Leukemia and Other Myeloproliferative Disorders*]. A high LAP score or the presence of toxic granulations usually suggests an underlying infection. When there is uncertainty, bone marrow aspiration and biopsy, chromosomal analysis, as well as marrow cultures for bacteria (e.g., *Salmonella, Brucella, Mycobacterium*, and fungi), are warranted. The results of these tests will enable the clinician to make a diagnosis of CML (or another myeloproliferative disorder), a granulomatous infection, inflammatory disease, or metastatic malignancy. If no such cause

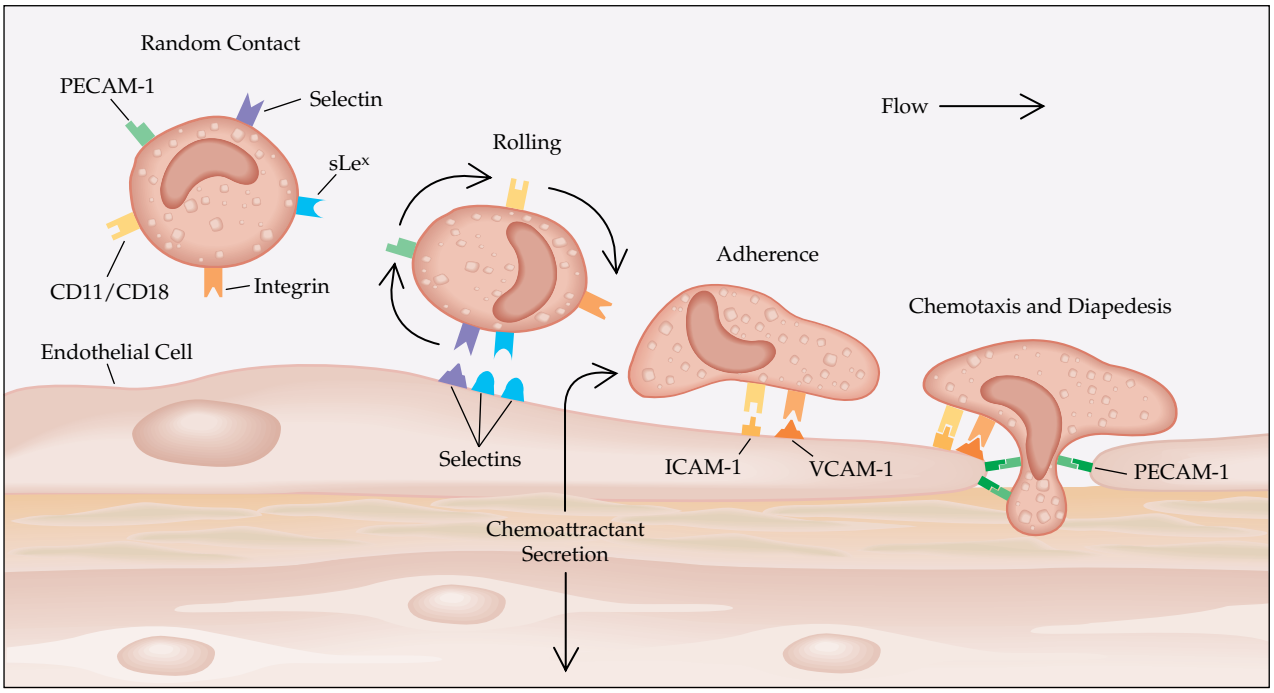


Figure 4 **Neutrophils in the peripheral blood exist in either the circulating or the marginating pool. The marginated neutrophils roll along a vessel wall, where their surface carbohydrates interact with selectins on the endothelial cells. After activation by chemotactic agents, the neutrophils change shape and change the affinity of their integrin molecules for endothelial cell intercellular adhesion molecules. The neutrophils then crawl and undergo diapedesis by interacting with platelet–endothelial cell adhesion molecules on the endothelial surface and by liberating hydrolases that permit passage of the neutrophils through the capillary basement membrane. (PECAM-1—platelet–endothelial cell adhesion molecule–1; sLe^x—sialyl-Lewis^x carbohydrate; ICAM-1—intercellular adhesion molecule–1; VCAM-1—vascular cell adhesion molecule–1)**

can be found in an otherwise healthy-appearing person, a diagnosis of idiopathic or familial neutrophilia may be considered, and repeated neutrophil counts can be performed at monthly intervals until the diagnosis is clarified.

Treatment

Except for the myeloproliferative syndromes, treatment of neutrophilia is not indicated. Neutrophil levels will return to normal when the inflammatory process is resolved.

NEUTROPENIA

Neutropenia is generally defined as a neutrophil count of less than 1,800/mm^3, which is two standard deviations below the normal mean. In some populations (e.g., Africans, African Americans, and Yemenite Jews), neutrophil counts as low as 1,000/mm^3, or 1.0×10^9/L, are probably normal.[21,22]

In otherwise healthy persons, the risk of bacterial infections is relatively low if the neutrophil count is greater than 500 mm^3, or 0.5×10^9/L—the level usually defined as severe neutropenia. When neutropenia develops after myelotoxic chemotherapy, the risk of infection is much greater, particularly in patients whose age and medical history (e.g., diabetes, heart failure, renal failure, previous chemotherapy, and HIV infection) also predispose them to infection.[22,23] Patients with neutropenia are also at greater risk for serious infections if they have disrupted mucosal or cutaneous barriers or are taking corticosteroids. Patients with neutropenia are at risk for infection by those pathogenic organisms that normally colonize body surfaces, particularly the skin, the oropharynx, and the GI tract. Thus, infections from staphylococci occur in neutropenic patients after breaks in the skin. Infection by mixtures of aerobic and anaerobic organisms of the oropharynx frequently causes gingivitis, pharyngitis, and sinusitis with neutropenia. Gram-negative bacilli often invade the blood from the GI tract in these patients. Antibiotic therapy, particularly therapy involving broad-spectrum antibiotics and protracted treatments, leads to colonization by resistant bacteria and to fungal infections.[24]

Etiology

Neutropenia may be a primary or secondary condition. Primary neutropenia is caused by abnormalities of neutrophil formation derived from hematopoietic stem cells in the bone marrow; disorders with this underlying pathogenesis include the myeloid malignancies and several congenital disorders [*see* Primary Forms of Neutropenia, *below*]. Secondary neutropenia may be caused by drug therapy, infections, and immunologic disorders, including autoimmune diseases. Secondary neutropenia is far more common than primary neutropenia. In all of these conditions, the risk of infection depends on the level of blood neutrophils and the capacity of the marrow to respond to an inflammatory stimulus and increase production of these cells. Usually if blood neutrophils are greater than 0.5×10^9/L, the risk of serious infection is relatively low.

Drug-induced neutropenia In aggregate, drug reactions are probably the most common cause of neutropenia in adults [*see Table 2*].[25,26] Many cancer chemotherapy drugs, some of which are also used as cytotoxic immunosuppressive agents (e.g., cyclophosphamide, methotrexate, and azathioprine), predictably cause dose-dependent neutropenia. The use of these agents requires careful attention to medical history, dosages, treatment schedules, and serial neutrophil counts to avoid serious and life-threatening toxicity. Other drugs cause neutropenia idiosyncratically. Many of these reactions probably occur because drugs can act as immunogens or as haptens, causing immunologic injury to neutrophils and their precursors. Other mechanisms of drug-induced neutropenia may involve direct toxicity of marrow cells in susceptible persons. Most patients recover from drug-induced neutropenia; the time for recovery can vary from 2 days to 2 weeks or more.

Infection-associated neutropenia Viral infections often cause mild neutropenia, especially in children. Such infections include measles and other viral exanthems, infectious mononucleosis, hepatitis, and HIV infection. The mechanisms are diverse. For example, in HIV infection, possible mechanisms include infections of the hematopoietic precursor cells and the marrow stromal cells, which lead to decreased production; induction of autoantibodies, which leads to accelerated turnover of mature neutrophils; and accelerated apoptosis of mature cells.[27] HIV-associated neutropenia generally develops late in AIDS and is often compounded by the use of antiviral agents (e.g., zidovudine, ganciclovir), antibiotics (e.g., sulfonamides), or the presence of hematologic malignancies (e.g., lymphoma, Kaposi sarcoma).[28] With other viral infections, the neutropenia is usually mild and without serious consequences. In rare instances, infectious mononucleosis causes severe hypoplasia, which has more severe consequences.[29] Neutropenia and anemia are common features of human parvovirus B19 infection.[30]

With severe bacterial infections, neutropenia occurs as a consequence of endotoxemia, which results in rapid neutrophil mobilization and turnover, especially in patients with a marrow reserve that is impaired because of previous chemotherapy, other drugs, or alcohol. In this setting, neutropenia generally portends a grave prognosis. Neutropenia occurs in parasitic infections associated with splenomegaly, such as kala-azar and acute malaria, presumably as a result of splenic trapping of the cells.

Autoimmune and idiopathic neutropenia Autoimmune neutropenia occurs as an isolated phenomenon or secondary to other autoimmune disorders.[31] For example, in patients with Evans syndrome, autoimmune neutropenia may be associated with immune thrombocytopenia and hemolytic anemia. The bone marrow cellularity in patients with autoimmune neutropenia is either normal or increased, with a relative decrease in the number of cells in the late stages of the neutrophil formation. The diagnosis of autoimmune neutropenia requires specific antineutrophil antibody tests.[32] The specificity of these tests probably varies considerably, and they are performed by a limited number of laboratories. It is often difficult to distinguish autoimmune neutropenia from cases otherwise categorized as idiopathic neutropenia. Neutropenia with antineutrophil antibodies also occurs in systemic lupus erythematosus,[33] Sjögren syndrome,[34,35] rheumatoid arthritis,[36] and Felty syndrome (i.e., rheumatoid arthritis, splenomegaly, and neutropenia).[37]

Patients with rheumatoid arthritis and neutropenia may have clonal expansion of large granular lymphocytes (usually $CD2^+$, $CD3^+$, $CD8^+$, and $CD57^+$ cells), which impair neutrophil production by excessive Fas ligand or interferon-gamma production.[38] Recent studies indicate that this same mechanism may be involved in cases diagnosed as idiopathic neutropenia.[39] The marrow typically shows increased lymphocytes with reduced neutrophils in the later stages of development. In most patients, the lymphocytosis is clonal and may evolve very gradually into a lymphoid malignancy.[40]

Table 2 Drugs Associated with Neutropenia

ANALGESICS
- Aminopyrine
- Dipyrone

ANTIBIOTICS
- Cephalosporins
- Chloramphenicol
- Clindamycin
- Doxycycline
- Flucytosine
- Gentamicin
- Griseofulvin
- Isoniazid
- Lincomycin
- Metronidazole
- Nitrofurantoin
- Penicillins
- Rifampin
- Streptomycin
- Sulfonamides
- Vancomycin

ANTICONVULSANTS
- Carbamazepine
- Ethosuximide
- Mephenytoin
- Phenytoin
- Primidone
- Trimethadione

ANTIHISTAMINES
- Brompheniramine
- Cimetidine
- Ranitidine
- Thenalidine
- Tripelennamine

ANTI-INFLAMMATORY AGENTS
- Fenoprofen
- Gold salts
- Ibuprofen
- Indomethacin
- Phenylbutazone

ANTIMALARIALS
- Amodiaquine
- Dapsone
- Hydroxychloroquine
- Pyrimethamine
- Quinine

ANTITHYROID DRUGS
- Carbimazole
- Methimazole
- Methylthiouracil
- Potassium perchlorate
- Propylthiouracil

CARDIOVASCULAR AGENTS
- Captopril
- Diazoxide
- Hydralazine
- Methyldopa
- Pindolol
- Procainamide
- Propranolol
- Quinidine

DIURETICS
- Acetazolamide
- Bumetanide
- Chlorothiazide
- Chlorthalidone
- Hydrochlorothiazide
- Methazolamide
- Spironolactone

HYPOGLYCEMIC AGENTS
- Chlorpropamide
- Tolbutamide

PHENOTHIAZINES
- Chlorpromazine
- Methylpromazine
- Prochlorperazine
- Promazine
- Thioridazine
- Trifluoperazine
- Trimeprazine

SEDATIVES AND NEUROPHARMACOLOGIC AGENTS
- Chlordiazepoxide
- Clozapine
- Desipramine
- Diazepam
- Imipramine
- Meprobamate
- Metoclopramide

MISCELLANEOUS AGENTS
- Allopurinol
- Colchicine
- Ethanol
- Levamisole
- Levodopa
- Penicillamine

In sarcoidosis, cirrhosis, and congestive splenomegaly of diverse causes, neutropenia and thrombocytopenia often occur concomitantly, presumably because of splenic sequestration. In most instances, the neutropenia is mild and without recognizable consequences.

Other secondary causes of neutropenia Neonates can have severe neutropenia because of transplacental transfer of maternal IgG antibodies to the FcγRIII (CD16) isotype (previously called NA-1 or NA-2) that is inherited from the infant's father.[41] This abnormality is transient, usually lasting less than 3 months. Transient severe neutropenia also can occur in an infant as a result of transplacental transfer of other antibodies (e.g., transfer of IgG) from a mother with autoimmune neutropenia. Pure white cell aplasia is a rare acquired condition characterized by a complete absence of myeloid precursors.[42] Pure white cell aplasia may be associated with a thymoma; if so, the aplasia may respond on removal of the thymoma. The short-term consequences of all these conditions depend primarily upon the level of blood neutrophils and the proliferative response of the marrow when inflammation or infections occur. The causes of other forms of neutropenia in children and adults are often difficult to establish and usually require referral to an expert hematologist.

Primary forms of neutropenia There are a number of congenital and inherited causes of neutropenia [*see Table 3*].

Diagnosis

Neutropenia is easily diagnosed by performing a white blood cell count and differential count. Patients with acute, severe neutropenia are often febrile and are frequently referred to as having acute febrile neutropenia. In this circumstance, attention is immediately focused on determining whether the patient has an infection, as well as focused on instituting empirical antibiotic therapy. Hematologic studies (e.g., bone marrow examination) are generally not necessary, because the cause of neutropenia is recognized from the patient's history, and it will resolve if the inciting cause has been eliminated.

Initial evaluation of patients with chronic neutropenia should include a careful family history and review of the incidence and severity of infections, including oral ulcers, gingivitis, cellulitis, and more serious problems. A complete blood count will reveal whether the neutropenia is isolated or associated with other hematologic abnormalities. Medications should be discontinued if they can be implicated as causes of the neutropenia. A bone marrow biopsy and aspirate are indicated if there is any question of a primary disease affecting the marrow (e.g., metastatic carcinoma, tuberculosis) or if myelodysplasia or a hematologic malignancy is suspected. Serologic testing for infectious mononucleosis, hepatitis, and HIV is often warranted, as is measurement of antinuclear antibodies and rheumatoid factor titers. Broader immunologic assessments (i.e., lymphocyte subtypes and immunoglobulin levels) are warranted if the history suggests a susceptibility to infections by viruses, parasites, or bacteria; and they are also useful to detect clonal proliferation of lymphocytes and to diagnose the large granular lymphocyte syndrome. Neutrophil mobilization with corticosteroids and demargination tests with epinephrine are rarely helpful.

Treatment of Neutropenia

Evidence-based guidelines for management and prevention of acute febrile neutropenia associated with cancer chemotherapy have been developed by the Infectious Diseases Society of America (www.idsociety.org) and the American Society of Clinical Oncology (www.asco.org). Other guidelines are also available (www.guideline.gov) [*see Table 4*]. In general, acute management of severe, idiosyncratic, drug-induced neutropenia should be similarly managed.[43,44]

Table 3 Intrinsic Disorders of Neutrophils That Cause Neutropenia

Disorder	*Inheritance*	*Clinical Features*	*Diagnosis*	*Treatment*
Congenital neutropenia (also known as infantile genetic agranulocytosis and Kostmann syndrome)	AD, AR, S Locus: 19p13.3	From birth, upper respiratory, lung, liver, and skin infections; mild anemia; thrombocytosis; a normal immune system; possible development of leukemia	Selective, severe neutropenia; marrow promyelocytes but few more mature cells; marrow eosinophils; normal chromosomes; possible G-CSF receptor defect Genetic testing: research only*	G-CSF (effective in most cases); bone marrow transplantation; prophylactic antibiotics
Myelokathexis	AD, S	Recurrent infections; severe leukopenia and neutropenia	Marrow cellularity normal with maturing, often binucleate neutrophils	G-CSF
Cyclic neutropenia (also known as cyclic hematopoiesis)	AD, S Locus: 19p13.3	Regular oscillations of blood cell counts, most prominently of neutrophil and monocyte counts	Serial CBCs show severe neutropenia that recurs regularly, usually every 21 days Genetic testing: research only*	G-CSF
Shwachman-Diamond syndrome	AR Locus: 7q11	Neutropenia with pancreatic insufficiency and sometimes with anemia or thrombocytopenia	Neutropenia with malabsorption caused by pancreatic enzyme deficiency; tests for cystic fibrosis negative Clinical testing available*	G-CSF; pancreatic enzymes
Chédiak-Higashi syndrome	AR, S Locus: 1q42	Recurrent infections; partial albinism; lymphoproliferative syndrome; neutropenia; thrombocytopenia	Giant cytoplasmic granules; defective neutrophil migration and bacterial killing Genetic testing: research only*	Antibiotics; vitamin C; bone marrow transplantation
Reticular dysgenesis and congenital immunodeficiency syndromes with neutropenia	AR, S	From birth, severe infections with severe leukopenia	Neutropenia; hypogammaglobulinemia; T cell and B cell deficiencies	Bone marrow transplantation; immunoglobulin therapy; G-CSF for neutropenia
Dyskeratosis congenita	AR Locus: Xq28; 3q21-q28	Severe infections; skin hyperpigmentation; dystrophic nails; leukoplakia	Skin changes associated with severe neutropenia Clinical testing available*	Prophylactic antibiotics

* See GeneTests (http://www.geneclinics.org) for laboratory directory.

AD—autosomal dominant AR—autosomal recessive CBC—complete blood count G-CSF—granulocyte colony-stimulating factor S—sporadic cases

Treatments for chronic neutropenia vary with the severity of neutropenia and the pattern of susceptibility to infection. Mild or moderate neutropenia (i.e., counts above 0.5×10^9/L, determined by serial counts over several weeks) rarely requires treatment. The neutropenia in Felty syndrome often responds to splenectomy and weekly doses of methotrexate.[45,46] With few other exceptions, long-term use of corticosteroids, γ-globulin injections, androgens, and splenectomy is not indicated for management of chronic neutropenia. With suspected infections, short-term, broad-spectrum antibiotic therapy is indicated, usually initiated after culture of blood and other body fluids for bacteria. Long-term antibiotic therapy is of unproven benefit in preventing infections, and it carries the risk of colonization by antibiotic-resistant organisms. G-CSF, usually in doses of 1 to 5 mg/kg/day, is of proven benefit for the treatment of congenital, idiopathic, and cyclic neutropenia and hastens the recovery of marrow from neutropenia after cancer chemotherapy.[47] G-CSF and GM-CSF have been widely used to treat other forms of chronic neutropenia, including the neutropenia associated with HIV infection.

Disorders of Neutrophil Function

In patients who have recurrent, severe, or unusual infections but who have a normal number of neutrophils, the presence of a neutrophil function disorder must be considered. Neutrophil function disorders are caused by defects in neutrophil adherence, chemotaxis, degranulation, or oxidative metabolism [*see Table 5*].

The evaluation of patients with confirmed, recurrent, or unusual infections is first to review the family history and then to examine the patient [*see Figure 5*]. A complete blood count and examination of the granulocytes in a blood smear can show neutrophilia or neutropenia, specific granule deficiency, or giant granules such as those that occur in Chédiak-Higashi syndrome. Evaluation of immunoglobulin levels (IgG, IgM, IgA, and IgE) and complement levels (C3 and CH_{50}) are also potentially helpful, especially if there is a pattern of infection by encapsulated bacteria or unusual organisms. After these considerations, neutrophil function should be evaluated with the nitroblue tetrazolium (NBT) test, superoxide production assays, and chemotactic assays. The NBT test and superoxide assays can determine whether a patient has chronic granulomatous disease (CGD), severe glucose-6-phosphate dehydrogenase (G6PD) deficiency, or a glutathione-pathway disorder[14]; chemotactic assays can be used to confirm the diagnosis of Chédiak-Higashi syndrome and acquired chemotactic defects.[12] Leukocyte adhesion deficiency types I and II are diagnosed by flow cytometry.[11] If the results of all of these tests are normal, ingestion assays using the patient's serum and cells and staining for MPO may be helpful. In this manner, all of the known neutrophil function abnormalities can be diagnosed, often with the aid of specialty consultations and a research laboratory.[48]

Monocytes and Macrophage Physiology

Monocytes and macrophages play critical roles in homeostasis and in host defense mechanisms. Monocytes and macro-

Table 4 Guidelines for Management and Prevention of Febrile Neutropenia[43]

Management

- Take careful history and conduct thorough physical examination of the patient
- Examine patient carefully for portal for bacterial or fungal infections
- Culture blood and other appropriate body fluids
- Start antibiotics immediately
  - Monotherapy (e.g., ceftazidime or imipenem) or duotherapy (e.g., an aminoglycoside, such as gentamicin, with a β-lactam drug that is effective against *Pseudomonas,* such as piperacillin
  - Add vancomycin if there is a significant risk of gram-positive sepsis
- Adjust antibiotic therapy after 3 days, depending on the results of cultures and the patient's clinical status
- Switch low-risk patients to oral therapy
- Continue broad-spectrum therapy for severely ill patients*
  - Consider antifungal treatments
  - Consider colony-stimulating factors as an adjunct to antibiotics for febrile neutropenia in severely ill, high-risk patients†

Prevention

- Primary prophylaxis with G-CSF reduces incidence of febrile neutropenia by ~50% when the risk of febrile neutropenia is ~40%‡
- Use G-CSF or GM-CSF as a preventive strategy for patients who have had their treatment reduced or experienced a delay in treatment because of an episode of febrile neutropenia or a prolonged period of neutropenia
  - Consider reducing the intensity of chemotherapy

Note: further information can be found at the following Web sites: www.idsociety.org, www.asco.org, and www.guidelines.gov.
*Resolution of illness generally follows resolution of neutropenia.
†For most patients with febrile neutropenia, CSF therapy has no proven benefit.
‡Administration of G-CSF or GM-CSF is not routinely indicated in previously untreated patients.
G-CSF—granulocyte colony-stimulating factor GM-CSF—granulocyte-macrophage colony-stimulating factor

phages perform tissue maintenance functions, such as clearance of particles—including bacteria—from the blood and removal of old red blood cells. They process antigens by interacting with T cells and B cells and are essential for containment of mycobacterial, parasitic, fungal, and viral infections.

MONOCYTE-MACROPHAGE DEVELOPMENT

Monocytes develop from hematopoietic progenitor cells in the bone marrow. Once the progenitor cells are committed to a monocyte lineage, they develop morphologically into monoblasts, then promonocytes, and then monocytes. Monocytes are present in the bone marrow and blood. They are the precursors for the tissue mononuclear phagocyte system (including alveolar, peritoneal, and splenic macrophages), Kupffer cells, osteoclasts, dendritic cells, and Langerhans cells. In addition to having phagocytic capabilities, monocytes and macrophages play a central role in the immune response through the generation of numerous cytokines, including growth factors for white blood cells.

With the exception of the alveolar macrophages, which are uniquely dependent on aerobic metabolism for energy production, monocytes and macrophages are facultative anaerobes. Phagocytosis by monocytes and macrophages is associated with an oxidative burst and stimulation of the hexose monophosphate shunt. Adhesion, chemotaxis, and activation are similar for monocytes and neutrophils, although macrophages are better than neutrophils at phagocytosis and perform chemotaxis less rapidly and efficiently. Macrophages are also capable of oxygen-independent bactericidal activity that may depend on lytic activity. Stimulated macrophages are capable of producing nitric oxide. Macrophages are capable of secreting many cytokines, growth factors, and acute-phase reactants.

Monocytes and macrophages present antigen to T cells in association with major histocompatibility complex (MHC) class II molecules. This association occurs in the lysosomes of a mononuclear cell before the MHC class II molecules are expressed on the cell surface. Monocytes and macrophages are involved in antibody-dependent and antibody-independent cell-mediated cytotoxicity. The cytotoxicity involves oxidative metabolism, the production of nitric oxide and cytokines, and the secretion of cytotoxic mediators.

Macrophages play a key role in metabolizing high-molecular-weight proteins, glycoproteins, and other material and are intimately involved in the destruction of senescent and killed cells. They also are required for angiogenesis and wound healing and are able to induce neovascularization and endothelial cell proliferation. Given these diverse products and functions, macrophages are involved in many metabolic, infectious, inflammatory, and degenerative diseases.

Increases in blood monocytes (usually less than two times the normal level or less than $1.0 \times 10^9/L$) are a common feature of chronic inflammatory diseases and malignancies. Higher counts should raise concern about a hematologic malignancy (e.g., monocytic or myelomonocytic leukemia) [*see 12:XVI Acute Leukemia*].

Disorders of Monocytes and Macrophages

HISTIOCYTIC SYNDROMES

Histiocytic syndromes are a group of malignant and nonmalignant disorders in which the macrophages and dendritic (Langerhans) cells are the principal cells of abnormality.[49] The malignant disorders include acute monocytic leukemia, monocytic sarcoma, and histiocytic sarcoma. The nonmalignant disorders include the Langerhans cell histiocytosis (LCH) syndromes and the hemophagocytic syndromes, such as sinus histiocytosis with massive lymphadenopathy, hemophagocytic lymphohistiocytosis (HL), and infection-associated hemophagocytic syndrome (IAHS).

Langerhans Cell Histiocytosis Syndromes

The LCH syndromes include solitary eosinophilic granuloma, multifocal eosinophilic granuloma, Hand-Schüller-Christian disease, and Letterer-Siwe disease.[49,50] These disorders predominantly affect children 1 to 15 years of age but also occur in young adults. The LCH syndromes represent a continuum of disease that has been divided on the basis of histologic studies, age at diagnosis, extent of disease, and organ involvement. The signs and symptoms of the LCH syndromes depend on the specific organs involved.[51] The bones, skin, teeth, gingival tissue, ears, endocrine organs, lungs, liver, spleen, lymph nodes, and bone marrow can all be involved and become dysfunctional as a result of cellular infiltration.[52] For example, diabetes insipidus is caused by histocyte infiltration of the pituitary gland,[53] and Erdheim-Chester disease is a multisystem disease characterized by histiocyte infil-

Table 5 Selected Disorders of Neutrophil Function

	Disorder	*Inheritance*	*Clinical Features*	*Diagnosis*	*Treatment*
Adherence defects	Leukocyte adhesion deficiency I	AR Locus: 21q22.3	Neutrophilia with recurrent severe infections; failure of pus formation; delayed umbilical cord separation	Decreased neutrophil adherence and migration; CD11/CD18 deficiency Clinical testing available*	Bone marrow transplantation; antibiotics
	Leukocyte adhesion deficiency II	S, possibly AR	Neutrophilia with recurrent infections	Neutrophils lack surface sLe^x and have deficient adherence	Bone marrow transplantation; antibiotics
	Actin polymerization defect	AR, S	Recurrent severe infections	Defective neutrophil migration and ingestion of bacteria	Bone marrow transplantation; antibiotics
Granule defects	Chédiak-Higashi syndrome	AR, S Locus: 1q42	Recurrent infections; partial albinism; lymphoproliferative syndrome; neutropenia; thrombocytopenia	Giant cytoplasmic granules; defective neutrophil migration and bacterial killing Genetic testing: research only*	Antibiotics; vitamin C; bone marrow transplantation
	Specific granule deficiency	S, possibly AR	Recurrent infections, especially sinopulmonary and skin infections	Absence of specific (secondary) granules in neutrophils; abnormal neutrophil migration and respiratory burst	Antibiotics
Respiratory burst defects	Chronic granulomatous disease	AR or X-linked CYBA locus: 16q24 CYBB locus: Xp21.1 NCF1 locus: 7q11.23 NCF2 locus: 1q25	Recurrent skin, pulmonary, and liver abscesses	Severely defective respiratory burst; NBT test; abnormality in one of four subunits of NADPH oxidase Clinical testing available*	Interferon gamma; antibiotics
	Glucose-6-phosphate dehydrogenase deficiency	X-linked Locus: Xq28	Recurrent bacterial infections; hemolytic anemia	Reduced levels of glucose-6-phosphate dehydrogenase Clinical testing available*	Antibiotics
	Myeloperoxidase deficiency	AR	Mild, if any, susceptibility to infection	Reduced levels of myeloperoxidase	Generally none indicated

* See GeneTests (http://www.geneclinics.org) for laboratory directory.
AR—autosomal recessive NADPH—nicotinamide-adenine dinucleotide phosphate NBT—nitroblue tetrazolium S—sporadic cases sLe^x—sialyl-Lewis^x

tration of many tissues.[54] LCH with solitary and multifocal eosinophilic granuloma is found predominantly in older children and young adults; more infiltrative disease is common in younger patients.[55] On presentation, patients with solitary lesions may have an inability to bear weight, or they may have tender swelling caused by tissue infiltrates that overlie a sharply marginated bony lesion. Diagnosis is usually made by demonstration of dendritic cells, eosinophils, and giant cells present in a biopsy specimen; electron microscopy and immunostaining may be helpful for further classification.

Treatment of local LCH is sometimes unnecessary; when it is necessary, surgery or local radiation therapy can be curative.[50-55] LCH syndromes respond to chemotherapeutic agents, including vinblastine, methotrexate, 6-mercaptopurine, etoposide, or 2-chlorodeoxyadenosine (cladribine). There is a long-term risk of secondary or treatment-related malignancies in these patients.

Sinus Histiocytosis with Massive Lymphadenopathy

Sinus histiocytosis with massive lymphadenopathy, or Rosai-Dorfman disease, is characterized by chronic, painless, massive lymphadenopathy that usually involves the cervical nodes and less frequently involves the axillary, hilar, peritracheal, or inguinal nodes.[56] It occurs in both adults and children. Extranodal disease in the respiratory tract, bones, orbits, skin, liver, and kidneys is present in almost 30% of patients. The disease is usually benign, but significant morbidity and even death may result if massive tissue invasion of the liver, kidneys, lungs, and other critical structures occurs. Patients are usually of African descent, and the incidence of this disease is highest in Africa and the West Indies.

The affected lymph nodes show marked sinusoidal dilatation and follicular hyperplasia with proliferation of foamy histiocytes and multinucleated giant cells in the sinuses. The etiology of this disorder is unknown and may be related to abnormal immune regulation. Attempts at treatment should be reserved for special circumstances that are potentially life threatening. Surgery, irradiation, corticosteroids, vinblastine, and cyclophosphamide have all been administered with varying degrees of success.

Hemophagocytic Lymphohistiocytosis

HL is a rapidly fatal disorder, occurring as a familial or acquired condition; it is characterized by fever, pancytopenia, hepatic dysfunction, and activated macrophages, which overproduce inflammatory cytokines.[57] Family studies suggest that a portion of cases are attributable to mutations in the perforin gene.[58,59] The disease is usually diagnosed in young children; however, secondary forms of HL account for numerous cases in adults and occur in association with bacterial, fungal, and parasitic infections and exposure to various drugs. Treatment is diffi-

cult. Chemotherapy may be helpful. If the disease is associated with infection, treatment with appropriate antimicrobials may resolve the disorder.

LYSOSOMAL STORAGE DISEASES

Monocytes and macrophages play a role in tissue remodeling and the removal of senescent cellular debris, and lysosomes are the organelles that perform these functions; therefore, enzymatic abnormalities that involve lysosomal constituents result in disorders of storage that are related to macrophage function. These disorders, usually diagnosed in early childhood, include the mucopolysaccharidoses, the glycoproteinoses, the sphingolipidoses, and the neutral lipid storage diseases [*see Table 6*]. Enzymatic defects have been described for most of these disorders, and diagnosis depends on demonstrating the enzymatic abnormality in macrophages or histiocytes. Most of these defects result from

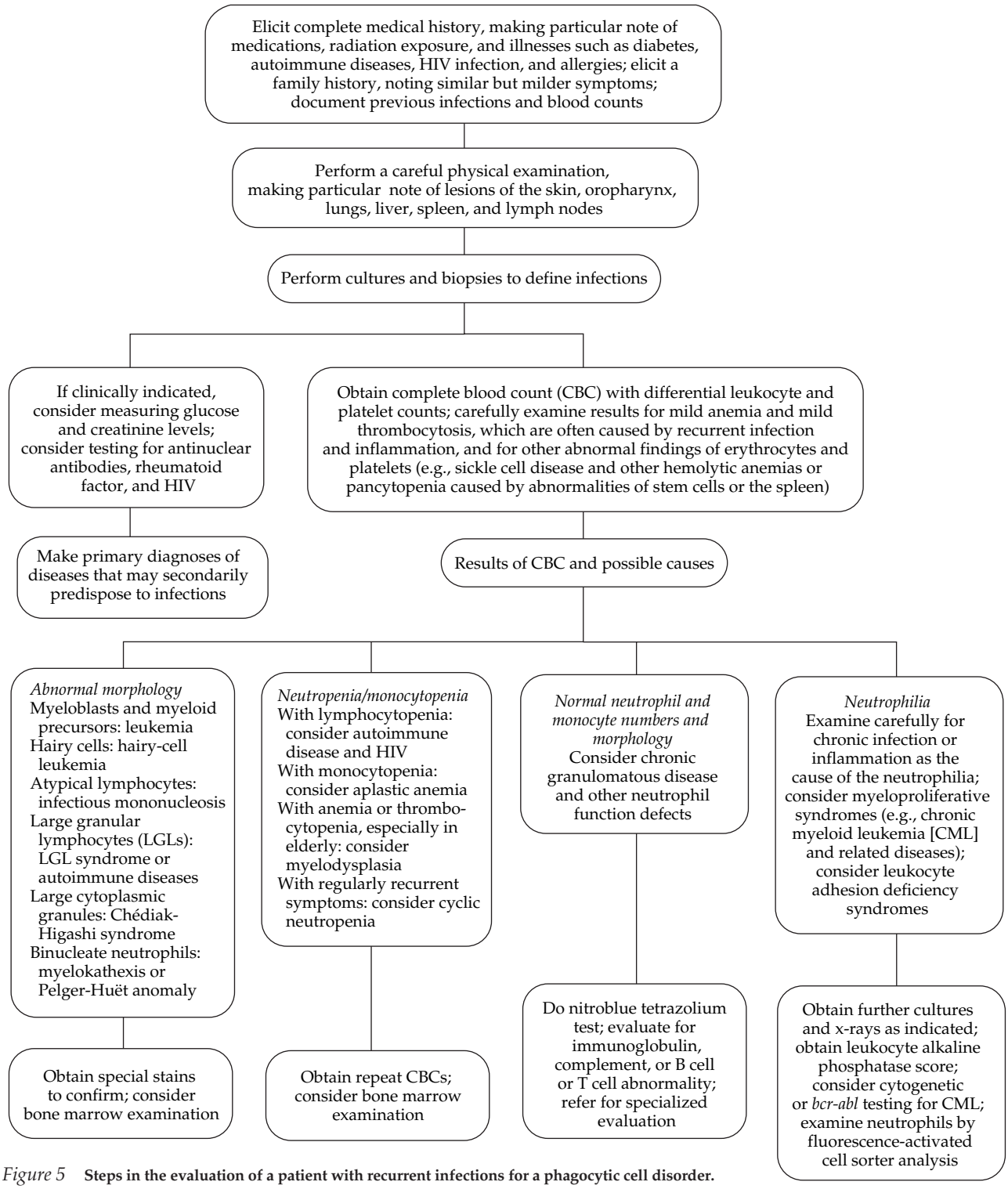


Figure 5 **Steps in the evaluation of a patient with recurrent infections for a phagocytic cell disorder.**

Table 6 Lysosomal Storage Diseases

Disease (Common Name)	*Inheritance*	*Enzymatic Defect*	*Organs and Tissues Involved*	*Stored Material*
Mucopolysaccharidoses (MPS)				
MPS IH (Hurler syndrome)	AR; locus: 4p16.3	α-L-Iduronidase	Liver, spleen, brain, heart, cornea, bone (mild and severe variants)	Dermatan sulfate, heparan sulfate
MPS II (Hunter syndrome)	Locus: Xq28	Iduronate-2-sulfatase	Liver, spleen, brain, heart, bone	Dermatan sulfate, heparan sulfate
MPS III				
(Sanfilippo A syndrome)	Locus: 17q25.3	Heparan *N*-sulfatase	Brain, liver, spleen, heart, bone	Heparan sulfate
(Sanfilippo B syndrome)	Locus: 17q21	α-*N*-Acetylglucosaminidase	Brain, liver, spleen, heart, bone	Heparan sulfate
(Sanfilippo C syndrome)	Chromosome 14	Acetyl-coenzyme A:α-glucosaminide *N*-Acetyltransferase		
MPS IV				
(Sanfilippo D syndrome)	Locus: 12q14	*N*-Acetylglucosamine-6-sulfatase		
(Morquio A syndrome)	Locus: 16q24.3	*N*-Acetylgalactosamine-6-sulfatase	Bone, cornea	Keratan sulfate, chondroitin 6-sulfate
(Morquio B syndrome)	Locus: 3p2	β-Galactosidase	Bone, cornea	Keratan sulfate
MPS VI (Maroteaux-Lamy syndrome)	Locus: 5q11-q13	*N*-Acetylgalactosamine-4-sulfatase	Bone, cornea, liver, spleen, heart (moderate and severe variants)	Dermatan sulfate
MPS VII (Sly syndrome)	Locus: 7q21	β-Glucuronidase	Brain, liver, spleen, bone, coronary arteries	Dermatan sulfate, heparan sulfate, chondroitin 4-sulfate, chondroitin 6-sulfate
Glycoproteinoses				
Mannosidosis	AR; locus: 19cen-q12	Lysosomal α-mannosidase	Brain, liver, spleen, bone (several variants)	Mannose-rich oligosaccharides
Fucosidosis	Locus: 1p34	Glycoprotein α-fucosidase	Brain, liver, spleen, heart, skin (several variants)	Fucose-containing oligosaccharides
Aspartylglucosaminuria	Locus: 4q32-q33	Aspartylglucosaminidase	Brain, liver, spleen, bone, heart	Aspartylglucosamine-containing peptides
Sialidosis	Locus: 6p21.3	Glycoprotein neuraminidase (sialidase)	Brain, liver, spleen, bone, retina (several variants)	Sialylated glycopeptides
Galactosialidosis	Locus: 20q13	Protector protein deficiency, combined neuraminidase (sialidase) and β-galactosidase deficiency	Brain, liver, spleen, bone (several variants)	GM_1 ganglioside, sialylated glycopeptides
Mucolipidosis II (I-cell disease)	Locus: 4q21-q23	*N*-Acetylglucosamine-1-phosphotransferase	Brain, bone, connective tissue	Glycoproteins, glycolipids
Mucolipidosis III (Pseudo-Hurler polydystrophy)	?	*N*-Acetylglucosamine-1-phosphotransferase	Brain, bone, connective tissue	Glycoproteins, glycolipids
Mucolipidosis IV (sialolipidosis)	Locus: 19p13.3-p13.2	Mucolipin 1	?	?
Sphingolipidoses				
(Gaucher disease type I [nonneuronopathic])	AR; locus: 1q21	Acid β-glucosidase, glucocerebrosidase	Liver, spleen, bone, bone marrow (highly variable phenotype)	Glucosylceramide
(Gaucher disease type 2 [acute neuronopathic])		Acid β-glucosidase, glucocerebrosidase	Brain, brain stem, liver, spleen, bone marrow, lungs	Glucosylceramide, glucosylsphingosine
(Gaucher disease type 3 [sub-acute neuronopathic])		Acid β-glucosidase, glucocerebrosidase	Brain, liver, spleen, bone marrow, lungs (variable phenotype)	Glucosylceramide, glucosylsphingosine

(*continued*)

point mutations or genetic rearrangements at a single locus of the gene that codes for a single lysosomal hydrolase.

The two types of therapy for lysosomal storage diseases that are currently available are cellular transplantation and enzyme therapy.[60] Gaucher disease was formerly treated with bone marrow transplantation, but it is currently treated with alglucerase, an α-mannosyl–terminated glucocerebrosidase. Bone marrow transplantation for the other lysosomal storage diseases is investigational and has yielded mixed results.

Eosinophil Physiology

Eosinophils can enhance or suppress acute inflammatory reactions and mediate responses to helminthic infection, allergy, and certain tumors.[61] Like neutrophils, eosinophils are capable of phagocytosis, but eosinophils are primarily secretory cells. Most of the functions they perform require the release of granule contents or reactive oxygen species. The eosinophils respond to unique chemotactic agents and growth factors that permit their accumulation at sites of inflammation.

EOSINOPHIL STRUCTURE

The granules of eosinophils contain strongly basic proteins and stain intensely with acid dyes. They have a striking and unique appearance on electron microscopy [*see Figure 1*]. The granules consist of an electron-dense core surrounded by a relatively radiolucent matrix; eosinophil peroxidase is active in the matrix. The dense core has a crystalloid structure and contains eosinophil cationic proteins (ECPs), major basic proteins (MBPs), and eosinophil-derived neurotoxins. MBPs and ECPs are capa-

Table 6 (continued)

Disease (Common Name)	*Inheritance*	*Enzymatic Defect*	*Organs and Tissues Involved*	*Stored Material*
Metachromatic leukodystrophy (MLD)				
Infantile MLD	Locus: 22q13.3-qter	Arylsulfatase A	Brain, peripheral nerves	Sulfatide
Juvenile MLD	Locus: 10q22	Arylsulfatase A	Brain, peripheral nerves	Sulfatide
Adult MLD		Arylsulfatase A, saposin B deficiency	Brain, peripheral nerves	Sulfatide
Pseudodeficiency		Partial arylsulfatase A	Normal	None
Multiple sulfatase deficiency		Unknown primary defect, multiple lysosomal and nonlysosomal sulfatase deficiencies	Brain, liver, spleen, bone	Sulfatide, dermatan sulfate, heparan sulfate
Gangliosidoses				
GM_2 gangliosidoses				
Infantile Tay-Sach disease (TSD)	Locus: 15q23-q24	Hexosaminidase A (α chain)	Brain	GM_2 ganglioside
Juvenile TSD		Hexosaminidase A (α chain)	Brain	GM_2 ganglioside
Adult TSD		Hexosaminidase A (α chain)	Brain	GM_2 ganglioside
Activator deficiency	Locus: 5q31-q33	GM_2 activator	Brain	GM_2 ganglioside
Sandhoff disease		Hexosaminidase B and A (β chain)	Brain, liver, spleen, bone	GM_2 ganglioside, globoside
GM_1 gangliosidoses	Locus: 3p21	β-Galactosidase	Brain, liver, spleen, bone	GM_1 ganglioside, keratan sulfate
Neutral sphingolipidoses				
Fabry disease	Locus: Xq22	α–Galactosidase A	Kidney, vascular endothelial system, heart, central nervous system vessels	Globotriaosylceramide
Schindler disease	Locus: 22q11	α-*N*-Acetylgalactosaminidase	Brain (probably several variants)	*N*-Acetylgalactosamine–linked oligosaccharides
Krabbe disease	Locus: 14q31	Galactocerebrosidase	Brain	Galactocerebroside
Niemann-Pick disease				
(Niemann-Pick A disease [infantile])	Locus: 11p15.4-p15.1	Sphingomyelinase	Brain, liver, spleen, lungs	Sphingomyelin
(Niemann-Pick B disease [late-onset])	Locus: 18q11-q12 (C1)	Sphingomyelinase	Liver, spleen, lungs	Sphingomyelin
(Niemann-Pick C disease)	Locus: 18q11-q12 (C2)			
Neutral lipid storage diseases				
Wolman disease	Locus: 10q24-q25	Lysosomal acid lipase	Liver, spleen, adrenal glands, bone marrow	Cholesteryl esters, triglycerides
Cholesterol ester storage disease	Locus: 10q23	Lysosomal acid lipase	Liver, spleen, blood vessels	Cholesteryl esters
Farber disease		Ceramidase	Brain, joints, tendons, skin, liver	Ceramide

ble of inflicting considerable damage to parasites such as schistosomula by binding to and disrupting their cell membranes. In addition, MBPs enhance the adherence of eosinophils and neutrophils to schistosomula.[62]

EOSINOPHIL FUNCTION

Eosinophils respond to a variety of chemotactic factors that enable them to enter tissues and carry out their functions. Some chemokines and chemotactic factors, such as C5a, *N*-formylmethionyl–containing peptides, and leukotriene B_4, stimulate both eosinophils and neutrophils. Several chemotactic stimuli, however, are highly specific for eosinophils. Among these eosinophil-specific stimuli are platelet-activating factor (PAF), eosinophil chemotactic factor of anaphylaxis, and a variety of parasite-derived factors. Responses to PAF, one of the most potent activators of normal eosinophils, include chemotaxis, adherence, enhanced binding of IgE, production of superoxide, release of granule proteins, and synthesis of prostanoids.

Both the production and activation of eosinophils are affected by GM-CSF, IL-5, and IL-3. IL-5 appears to be critical for eosinophil production and deployment.[63] Exposure to low doses of IL-5 also specifically primes eosinophils for later actions by other stimulants. Once activated, the eosinophils have enhanced generation of reactive oxygen species, enhanced glucose utilization and transport, increased oxygen consumption, a reduced cell surface charge, and activation of acid phosphatases in specific granules.

Eosinophils enhance the immune response to helminths. They perform this function by binding to the surface of both larval and adult forms, by damaging target cells through oxygen-dependent mechanisms that are similar to those of neutrophils, and by damaging cell surfaces by releasing granule proteins such as MBP and ECP. Although the release of these proteins similarly damages normal tissues and tumor cells, these interactions between eosinophils and host cells are less well understood. Eosinophils also produce cytokines that enhance the in-

flammatory response.[64] The presence of eosinophilia in patients with Hodgkin disease appears to be a function of the production of IL-5 by Reed-Sternberg cells. Eosinophils contribute to the fibrosis of the nodular sclerosis type of Hodgkin disease by producing transforming growth factor–β1.

Disorders of Eosinophil Number

EOSINOPHILIA

Evaluation of the patient with eosinophilia (eosinophil count > 700/mm^3) is difficult because the causes of this disorder are multiple and diverse.[65] Common causes of secondary eosinophilia include allergic disorders, infections caused by parasites and other organisms, dermatologic diseases, pulmonary diseases, collagen vascular disease, neoplasms, and immunodeficiency diseases. There are also myriad uncommon causes, such as eosinophilic gastroenteritis, inflammatory bowel disease, chronic active hepatitis, pancreatitis, and hypopituitarism.

HYPEREOSINOPHILIC SYNDROME

The term hypereosinophilic syndrome (HES) is often used for patients with chronic eosinophilia of unknown cause.[66] The criteria used to diagnose HES are an unexplained eosinophil count of greater than 1,500/mm^3 for longer than 6 months and signs or symptoms of infiltration of eosinophils into tissues. Recent evidence points to a mutation in chromosome 4 that results in linkage of the *Rhe* gene and the *PDGFRα* gene.[67]

The clinical features of HES are rash, fever, cough, dyspnea, diarrhea, and peripheral neuropathy. Patients may have chronic congestive heart failure, valvular abnormalities, and distinctive, fibrous, biventricular endocardial thickening with mural thrombi.[66] The blood smear of a patient with HES usually reveals normal mature eosinophils of typical morphology; however, the presence of hypogranulation and cytoplasmic vacuoles has been reported. The total leukocyte count is typically 10,000 to 30,000/mm^3, 30% to 70% of which are eosinophils. The bone marrow is generally hypercellular, with eosinophils constituting 25% to 75% of the marrow elements.

HES can usually be distinguished from malignant disorders associated with eosinophilia, such as acute or chronic eosinophilic leukemias.[68] Allergic reactions must also be excluded; the exclusion of such a reaction is usually based on the history, physical examination, and review of current medications. Because many drugs may generate an allergic reaction accompanied by eosinophilia, all nonessential medication should be discontinued before the patient is evaluated.

Parasitic infections, most commonly with such tissue-invasive helminths as filariae and *Strongyloides, Trichinella, Schistosoma,* and *Toxocara* species, frequently present with eosinophilia. To eliminate parasitosis as the cause of eosinophilia, multiple stool samples and a small bowel aspirate are recommended, particularly in patients who are at particular risk for infection (e.g., those who frequently travel, those who are exposed to animals, and those who have immunodeficiencies). If these test results are negative, serologic assays, radiologic tests, and peripheral blood and bone marrow smears should be performed to exclude the presence of connective tissue diseases, occult lymphoproliferative syndromes and solid tumors, and hematologic malignancies, respectively. In patients with possible cardiac involvement, an echocardiogram should be performed.

Therapy is directed toward lowering the eosinophil count and correcting specific symptoms. If symptoms involving the lungs or the heart are present, prednisone at a dosage of 1 mg/kg/day should be given for 2 weeks, followed by 1 mg/kg every other day for 3 months or longer. If this treatment fails or if an alternative is necessary to avoid steroid side effects, hydroxyurea at a dosage of 0.5 to 1.5 g/day should be given to lower the WBC count to less than 10,000/mm^3 and the eosinophil count to less than 5,000/mm^3. Study findings suggest that treatment with imatinib mesylate is effective.[69] Alternative agents include interferon alfa, cyclosporine, and etoposide.

Table 7 Causes of Lymphocytosis

Lymphoproliferative disorders (primary lymphocytosis)
Leukemia
Acute lymphocytic leukemia
Chronic lymphocytic leukemia
Hairy-cell leukemia
Large granular lymphocyte leukemia
Lymphoma
Monoclonal B cell lymphocytosis
Reactive (secondary) lymphocytosis
Viral infection (most likely with EBV, CMV, HIV, HSV, VZV, rubella, adenovirus, or hepatitis virus)
Toxoplasmosis
Pertussis
Stress
Acute
Cardiovascular collapse
Septic shock
Sickle cell crisis
Status epilepticus
Trauma
Surgery
Drugs
Hypersensitivity
Chronic
Autoimmune disorders
Cancer
Hyposplenism
Sarcoidosis
Cigarette smoking

CMV—cytomegalovirus EBV—Epstein-Barr virus HSV—herpes simplex virus VZV—varicella-zoster virus

Basophil and Mast Cell Physiology

Basophils and mast cells are important in immediate hypersensitivity reactions, asthma, urticaria, allergic rhinitis, and anaphylaxis.[70] They are derived from a common hematopoietic progenitor cell in the bone marrow and are stimulated by soluble mediators, primarily IgE, to release granule contents and arachidonic acid metabolites from their plasma membranes.

The cytoplasmic granules of both basophils and mast cells contain sulfated glycosaminoglycans; in normal basophils, the sulfated glycosaminoglycans are predominantly heparin. The sulfated glycosaminoglycans are the granule contents that are primarily responsible for the intense staining of the basophil. Most, if not all, of the circulating histamine in the body is synthesized by the basophil and stored in its granules. Degranulation causes the release of histamine, which mediates many immediate hypersensitivity effects and which, because it is a potent

eosinophil chemoattractant, draws eosinophils to the site of degranulation. Other substances that are released on basophil degranulation include additional eosinophil chemotactic factors and a variety of arachidonic acid metabolites, the most important of which is leukotriene C_4. In addition, the cell membranes of basophils contain high-affinity IgE receptors, the number of which tends to be increased in allergic persons.

Disorders of Basophil Number

BASOPHILIA

Basophilia (basophil count > 150/mm^3) is seen in myeloproliferative disorders, such as CML, polycythemia vera, and myeloid metaplasia; after splenectomy; in some hemolytic anemias; and in Hodgkin disease. The basophil count can also be increased in patients with ulcerative colitis or varicella infection. Although basophils and mast cells are involved in immediate hypersensitivity reactions and basophils are often seen in areas of contact dermatitis, basophilia is not seen in patients with these disorders.

Lymphocyte Physiology

Lymphocytes (e.g., B cells and T cells) are also derived from hematopoietic stem cells. These cells develop and mature in the bone marrow, thymus, spleen, and lymph nodes and in other specialized lymphoid tissues [*see Section 6 Immunology and Allergy*].

Disorders of Lymphocytes

LYMPHOCYTOSIS

Lymphocytosis in adults is defined as an absolute lymphocyte count greater than 4,000/mm^3. In children with the disease, lymphocyte counts are higher than in adults and may be as high as 20,000/mm^3 in the first year of life. The blood film of any patient with lymphocytosis should be carefully examined to determine the morphology and diversity of the lymphocytes (e.g., reactive lymphocytes, large granular lymphocytes, blasts, or smudge cells).

Lymphocytosis can be either primary or secondary. Primary lymphocytosis, often called lymphoproliferative disease, is caused by dysregulation in the production of lymphocytes. The primary lymphocytoses include the leukemias (e.g., chronic lymphocytic leukemia, acute lymphocytic leukemia, hairy-cell leukemia, or large granular lymphocyte leukemia), the lymphomas, and monoclonal B cell lymphocytosis [*see Table 7*].

The reactive, or secondary, lymphocytoses are conditions that involve absolute increases in lymphocytes caused by physiologic or pathophysiologic responses to infection, inflammation, toxins, cytokines, or unknown agents. The most common causes of reactive lymphocytosis are viral infections: Epstein-Barr virus, cytomegalovirus, herpes simplex virus, varicella-zoster virus, rubella, human T cell lymphotropic virus type I (HTLV-I), HIV, adenovirus, or one of the hepatitis viruses is frequently responsible for the disease. Other pathogens that produce lymphocytosis are *Toxoplasma gondii* and, in children, *Bordetella pertussis* (which causes the lymphocyte count to rise to as high as 70,000/mm^3). Lymphocytosis is also associated with stress and consequent release of epinephrine, such as that seen in patients who have had cardiovascular collapse, septic shock, sickle cell crisis, status epilepticus, trauma, major surgery, drug reactions, or hypersensitivity. Persistent lymphocytosis may be seen in patients with autoimmune disorders, sarcoidosis, hyposplenism, or cancer and in those who are long-term cigarette smokers.

Table 8 Causes of Lymphocytopenia

- Inherited
  - Congenital immunodeficiency diseases
    - Severe combined immunodeficiency
      - Adenosine deaminase deficiency
      - Purine-nucleoside phosphorylase deficiency
      - Reticular dysgenesis
    - Ataxia-telangiectasia
    - Wiskott-Aldrich syndrome
    - Cartilage-hair hypoplasia
  - Idiopathic CD4$^+$ T lymphocytopenia
- Acquired
  - Infection
    - Viral (e.g., with HIV, a hepatitis virus, influenza virus, or respiratory syncytial virus)
    - Bacterial (e.g., typhoid fever, pneumonia, sepsis, or tuberculosis)
  - Aplastic anemia
  - Autoimmune diseases
  - Hodgkin disease
  - Sarcoidosis
  - Renal failure
  - Protein-losing enteropathies
  - Chylous ascites
  - Zinc deficiency
  - Chronic alcohol ingestion
  - Immunosuppressive agents (e.g., antithymocyte globulin, corticosteroids, chemotherapeutic agents, and radiation)

LYMPHOCYTOPENIA

Lymphocytopenia is defined as a total lymphocyte count less than 1,000/mm^3. Because in adults 80% of lymphocytes are T cells, most cases of lymphocytopenia are caused by a reduction in the T cell count. The mechanisms of lymphocytopenia are often unknown, and the causes are usually differentiated as inherited or acquired.

Inherited lymphocytopenias are usually caused by congenital immunodeficiency diseases. These diseases include severe combined immunodeficiency (e.g., adenosine deaminase deficiency, purine-nucleoside phosphorylase deficiency, and reticular dysgenesis), ataxia-telangiectasia, Wiskott-Aldrich syndrome, and cartilage-hair hypoplasia [*see Table 8*]. In addition, some persons have idiopathic CD4$^+$ T cell lymphocytopenia.

Acquired lymphocytopenia can be seen in patients with viral infections, such as HIV infection, hepatitis, influenza, and respiratory syncytial virus infection; in patients with certain bacterial infections, such as typhoid fever, pneumonia, sepsis, and tuberculosis; and in patients with aplastic anemia, autoimmune diseases, Hodgkin disease, sarcoidosis, renal failure, protein-losing enteropathies, and chylous ascites. Zinc deficiency and long-term alcohol ingestion are also associated with lymphocytopenia. Finally, immunosuppressive agents, such as antithymocyte globulin, corticosteroids, chemotherapeutic agents, and radiation, also produce lymphocytopenia.

The author has received grants for clinical research from and served as consultant for Amgen Inc.

References

1. Pierre RV: Peripheral blood film review: The demise of the eyecount leukocyte differential. Clin Lab Med 22:279, 2002

2. Hubel K, Engert A: Clinical applications of granulocyte colony-stimulating factor: an update and summary. Ann Hematol 82:207, 2003

3. Root RK, Dale DC: Granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor: comparisons and potential for use in the treatment of infections in nonneutropenic patients. J Infect Dis 179:S342, 1999

4. Dranoff G: GM-CSF-based cancer vaccines. Immunol Rev 188:147, 2002

5. Dancey JT, Deubelbeiss KA, Harker LA, et al: Neutrophil kinetics in man. J Clin Invest 58:705, 1976

6. Simon HU: Neutrophil apoptosis pathways and their modifications in inflammation. Immunol Rev 93:101, 2003

7. Gullberg U, Bengtsson N, Bulow E, et al: Processing and targeting of granule proteins in human neutrophils. J Immunol Methods 17:201, 1999

8. Lehrer RI, Ganz T: Cathelicidins: a family of endogenous antimicrobial peptides. Curr Opin Hematol 9:18, 2002

9. Fais S, Malorni W: Leukocyte uropod formation and membrane/cytoskeleton linkage in immune interactions. J Leukoc Biol 73:556, 2003

10. Lindbom L, Werr J: Integrin-dependent neutrophil migration in extravascular tissue. Semin Immunol 14:115, 2002

11. McIntyre TM, Prescott SM, Weyrich AS, et al: Cell-cell interactions: leukocyte-endothelial interactions. Curr Opin Hematol 10:150, 2003

12. Cicchetti G, Allen PG, Glogauer M: Chemotactic signaling pathways in neutrophils: from receptor to actin assembly. Crit Rev Oral Biol Med 13:220, 2002

13. Hogarth PM: Fc receptors are major mediators of antibody-based inflammation in autoimmunity. Curr Opin Immunol 14:798, 2002

14. Babior BM, Lambeth JD, Nauseef W: The neutrophil NADPH oxidase. Arch Biochem Biophys 397:342, 2002

15. Babior BM: Phagocytes and oxidative stress. Am J Med 109:33, 2000

16. Hubel K, Dale DC, Liles WC: Therapeutic use of cytokines to modulate phagocyte function for the treatment of infectious diseases: current status of granulocyte colony-stimulating factor, granulocyte-macrophage colony-stimulating factor, macrophage colony-stimulating factor, and interferon-gamma. J Infect Dis 185:1490, 2002

17. Xing L, Remick DG: Relative cytokine and cytokine inhibitor production by mononuclear cells and neutrophils. Shock 20:10, 2003

18. Bunting M, Harris ES, McIntyre TM, et al: Leukocyte adhesion deficiency syndromes: adhesion and tethering defects involving beta 2 integrins and selectin ligands. Curr Opin Hematol 9:30, 2002

19. Callen JP: Neutrophilic dermatoses. Dermatol Clin 20:409, 2002

20. Creutzig U, Ritter J, Vormoor J: Myelodysplasia and acute myelogenous leukemia in Down's syndrome: a report of 40 children of the AML-BFM Study Group. Leukemia 10:1677, 1996

21. Bain BJ, Phillips D, Thomson K, et al: Investigation of the effect of marathon running on leucocyte counts of subjects of different ethnic origins: relevance to the aetiology of ethnic neutropenia. Br J Haematol 108:483, 2000

22. Scott S: Identification of cancer patients at high risk of febrile neutropenia. Am J Health Syst Pharm 59:S16, 2002

23. Chrischilles E, Delgado DJ, Stolshek BS, et al: Impact of age and colony-stimulating factor use on hospital length of stay for febrile neutropenia in CHOP-treated non-Hodgkin's lymphoma. Cancer Control 9:203, 2002

24. Zinner SH: New pathogens in neutropenic patients with cancer: an update for the new millennium. Int J Antimicrob Agents 16:97, 2000

25. van der Klauw MM, Wilson JH, Stricker BH: Drug-associated agranulocytosis: 20 years of reporting in The Netherlands (1974-1994). Am J Hematol 57:206, 1998

26. van Staa TP, Boulton F, Cooper C, et al: Neutropenia and agranulocytosis in England and Wales: incidence and risk factors. Am J Hematol 72:248, 2003

27. Pitrak DL, Tsai HC, Mullane KM, et al: Accelerated neutrophil apoptosis in the acquired immunodeficiency syndrome. J Clin Invest 15:2714, 1996

28. Kuritzkes DR: Neutropenia, neutrophil dysfunction, and bacterial infection in patients with human immunodeficiency virus disease: the role of granulocyte colony-stimulating factor. Clin Infect Dis 30:256, 2000

29. Sato T, Hirasawa A, Kawabuchi Y, et al: Cellular expressions and serum concentrations of Fas ligand and Fas receptor in patients with infectious mononucleosis. Int J Hematol 72:329, 2000

30. Honda K, Ishiko O, Tsujimura A, et al: Neutropenia accompanying parvovirus B19 infection after gynecologic surgery. Acta Haematol 103:186, 2000

31. Palmblad JE, von dem Borne AE: Idiopathic, immune, infectious, and idiosyncratic neutropenias. Semin Hematol 39:113, 2002

32. Bux J: Molecular nature of antigens implicated in immune neutropenias. Int J Hematol 76:399, 2002

33. Keeling DM, Isenberg DA: Haematological manifestations of systemic lupus erythematosus. Blood Rev 7:199, 1993

34. Ramakrishna R, Chaudhuri K, Sturgess A, et al: Haematological manifestations of primary Sjogren's syndrome: a clinicopathological study. Q J Med 83:547, 1992

35. Starkebaum G: Chronic neutropenia associated with autoimmune disease. Semin Hematol 39:121, 2002

36. Bowman SJ: Hematological manifestations of rheumatoid arthritis. Scand J Rheumatol 31:251, 2002

37. Rosenstein ED, Kramer N: Felty's and pseudo-Felty's syndromes. Semin Arthritis Rheum 21:129, 1991

38. Liu JH, Wei S, Lamy T, et al: Chronic neutropenia mediated by fas ligand. Blood 15:3219, 2000

39. Papadaki HA, Eliopoulos AG, Kosteas T, et al: Impaired granulocytopoiesis in patients with chronic idiopathic neutropenia is associated with increased apoptosis of bone marrow myeloid progenitor cells. Blood 101:2591, 2003

40. Lamy T, Loughran TP Jr: Clinical features of large granular lymphocyte leukemia. Semin Hematol 40:185, 2003

41. Maheshwari A, Christensen RD, Calhoun DA: Immune-mediated neutropenia in the neonate. Acta Paediatr Suppl 91:98, 2002

42. Fumeaux Z, Beris P, Borisch B: Complete remission of pure white cell aplasia associated with thymoma, autoimmune thyroiditis and type 1 diabetes. Eur J Haematol 70:186, 2003

43. Beauchesne MF, Shalansky SJ: Nonchemotherapy drug-induced agranulocytosis: a review of 118 patients treated with colony-stimulating factors. Pharmacotherapy 19:299, 1999

44. Andres E, Kurtz JE, Perrin AE, et al: Haematopoietic growth factor in antithyroid-drug-induced agranulocytosis. QJM 94:423, 2001

45. Rashba EJ, Rowe JM, Packman CH: Treatment of the neutropenia of Felty syndrome. Blood Rev 10:177, 1996

46. Hellmich B, Schnabel A, Gross WL: Treatment of severe neutropenia due to Felty's syndrome or systemic lupus erythematosus with granulocyte colony-stimulating factor. Semin Arthritis Rheum 29:82, 1999

47. Dale DC, Cottle TE, Fier CJ, et al: Severe chronic neutropenia: treatment and follow-up of patients in the Severe Chronic Neutropenia International Registry. Am J Hematol 72:82, 2003

48. Dinauer MC, Lekstrom-Himes JA, Dale DC: Inherited Neutrophil Disorders: Molecular Basis and New Therapies. Hematology (Am Soc Hematol Educ Program) 303: 2000

49. Harris NL, Jaffe ES, Diebold J, et al: The World Health Organization classification of neoplasms of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting—Airlie House, Virginia, November, 1997. Hematol J 1:53, 2000

50. Ladisch S: Langerhans cell histiocytosis. Curr Opin Hematol 5:54, 1998

51. Lieberman PH, Jones CR, Steinman RM, et al: Langerhans cell (eosinophilic) granulomatosis: a clinicopathologic study encompassing 50 years. Am J Surg Pathol 20:519, 1996

52. Sundar KM, Gosselin MV, Chung HL, et al: Pulmonary Langerhans cell histiocytosis: emerging concepts in pathobiology, radiology, and clinical evolution of disease. Chest 123:1673, 2003

53. Kaltsas GA, Powles TB, Evanson J: Hypothalamo-pituitary abnormalities in adult patients with Langerhans cell histiocytosis: clinical, endocrinological, and radiological features and response to treatment. J Clin Endocrinol Metab 85:1370, 2000

54. Shamburek RD, Brewer HB Jr, Gochuico BR: Erdheim-Chester disease: a rare multisystem histiocytic disorder associated with interstitial lung disease. Am J Med Sci 321:66, 2001

55. Ghanem L, Tolo VT, D'Ambra P, et al: Langerhans cell histiocytosis of bone in children and adolescents. J Pediatr Orthop 23:124, 2003

56. Pulsoni A, Anghel G, Falcucci P, et al: Treatment of sinus histiocytosis with massive lymphadenopathy (Rosai-Dorfman disease): report of a case and literature review. Am J Hematol 69:67, 2002

57. Ravelli A: Macrophage activation syndrome. Curr Opin Rheumatol 14:548, 2002

58. Goransdotter Ericson K, Fadeel B, Nilsson-Ardnor S, et al: Spectrum of perforin gene mutations in familial hemophagocytic lymphohistiocytosis. Am J Hum Genet 68:590, 2001

59. Feldmann J, Le Deist F, Ouachee-Chardin M, et al: Functional consequences of perforin gene mutations in 22 patients with familial haemophagocytic lymphohistiocytosis. Br J Haematol 117:965, 2002

60. Schiffmann R, Brady RO: New prospects for the treatment of lysosomal storage diseases. Drugs 62:733, 2002

61. Brito-Babapulle F: The eosinophilias, including the idiopathic hypereosinophilic syndrome. Br J Haematol 121:203, 2003

62. Gleich GJ: Mechanisms of eosinophil-associated inflammation. J Allergy Clin Immunol 105:651, 2000

63. Menzies-Gow A, Robinson DS: Eosinophils, eosinophilic cytokines (interleukin-5), and antieosinophilic therapy in asthma. Curr Opin Pulm Med 8:33, 2002

64. Rumbley CA, Sugaya H, Zekavat SA, et al: Activated eosinophils are the major source of Th2-associated cytokines in the schistosome granuloma. J Immunol 162:1003, 1999

65. Rothenberg ME: Eosinophilia. N Engl J Med 338:1592, 1998

66. Roufosse FE, Goldman M, Cogan E: Hypereosinophilic syndrome. N Engl J Med 348:2687, 2003

67. Griffin JH, Leung J, Bruner RJ, et al: Discovery of a fusion kinase in EOL-1 cells and idiopathic hypereosinophilic syndrome. Proc Natl Acad Sci USA 100:7830, 2003

68. Bain BJ: Hypereosinophilia. Curr Opin Hematol 7:21, 2000

69. Cools J, DeAngelo DJ, Gotlib J, et al: A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. N Engl J Med 348:1201, 2003

70. Prussin C, Metcalfe DD: IgE, mast cells, basophils, and eosinophils. J Allergy Clin Immunol 111:S486, 2003

Acknowledgments

Figure 1 Tom Moore. Electron micrographs courtesy of Dr. E. Chi, University of Washington School of Medicine, Seattle.

Figure 3 Tom Moore.

X TRANSFUSION THERAPY

W. Hallowell Churchill, m.d.

Transfusion medicine developed rapidly, owing to several key discoveries and technical advances. These include the discovery of blood group antigens and the understanding of the host immune response to these antigens, the development of methods of anticoagulation and storage of blood, and the creation of plastic bags that allow sterile fractionation of whole blood into components. The potential of blood to act as an agent of disease transmission has heavily shaped both the donation process and transfusion practice.[1] Decisions about whether to transfuse must involve weighing the benefits against the risks. This chapter provides a basis for these decisions, including indications for blood-component use, complications of transfusion therapy, and methods of reducing risks during the collection, processing, and preparation of blood components.

Blood Donation

The donation process for either whole blood or special products, such as single-donor platelets (SDPs) obtained by apheresis, is designed to protect both the donor and the recipient. For example, persons weighing less than 110 lb (49.9 kg) are excluded from donation because they have too small a blood volume to donate blood safely. Donors taking drugs that would impair recovery from a vasovagal donor reaction may also be excluded in some locales. Recipient safety is promoted by excluding donors who are at risk for viral or bacterial infections or are taking medications that could cause reactions or impair the function of donated blood products.

autologous and directed donation

Autologous donations and directed donations are two strategies adopted by patients seeking to minimize their real or perceived risk of infection from blood products.

Autologous Donation

In autologous donation, patients deposit their own blood and then receive that blood if they need transfusion therapy. This eliminates the infectious and sensitization risks associated with allogeneic blood. Absolute contraindications to autologous donation are tight aortic stenosis, unstable angina, and active bacterial infection.[2] Low hemoglobin levels and poor venous access frequently limit the number of units that can be collected. With the increasing safety of allogeneic blood, the rationale for autologous donation may ultimately depend on the importance of the possible modulation of recipient immune function associated with allogeneic transfusion rather than the avoidance of blood-borne infections.

Directed Donation

Directed donation is donation for a specific recipient. It usually involves donations made by friends or family members of the intended recipient. It is based on the assumption that transfusions involving donors selected by the recipient carry lower risk of infections than transfusions involving donors from the general population. However, available prevalence data show that the risk of infectious disease from directed donors is no different from that of first-time donors.[3] The current risk of infection via transfusion is so low [*see Table 1*] that justification for directed donor programs depends primarily on patient preferences or on the need for a selected donor serving as the only source of blood products to reduce the recipient's risk from exposure to multiple donors. The latter form of directed donation is most appropriate for neonatal transfusions, in which one of the biologic parents may provide all the needed blood products.[4]

screening procedures

The combination of improved donor selection and postdonation testing has greatly decreased the infectious risks of allogeneic blood [*see Table 1*]. Predonation donor screening to identify clinical and lifestyle characteristics associated with higher incidences of infection has produced the biggest decrease in the risk of transfusion-transmitted disease.

postdonation testing

Postdonation testing is essential in identifying donors likely to transmit blood-borne infections who are missed in the initial screening process.

Screening for Hepatitis Viruses

Hepatitis C Screening for hepatitis C began in 1990 with the availability of a single antigen-based enzyme-linked immunosorbent assay (ELISA). This assay, together with second- and third-generation assays, their associated confirmatory tests, and nucleic acid testing (NAT) for viral RNA or DNA, has reduced the per-unit risk of hepatitis C virus (HCV) transmission to less than 0.0001% (1 per 1.6 million to 1 per 1.935 million).[5,6] Before these tests were available, the risk per unit was about 4%. Improved hepatitis C testing has eliminated the need for surrogate tests, such as the measurement of alanine aminotransferase (ALT) levels and testing for antibody to hepatitis B virus (HBV) core antigen. However, the test for antibody to HBV core antigen is still used to detect recently infected donors who lack measurable circulating HBV antigen.[7]

The epidemiology of HCV is still poorly understood. Approximately 20% to 25% of persons found to be HCV positive have no known risk factors.[8] Sexual transmission occurs with enough frequency to warrant evaluation of partners and appropriate use

Table 1 Estimated Risk of Infection per Transfused Blood Product[5,6]

Virus	*Risk in Year 2000*	*Risk in Year 2003*
Human immunodeficiency virus type 1/type 2 (HIV-1/2)*	1 in 660,000	1 in 2,135,000
Human T cell lymphotropic virus type I/type II (HTLV-I/II)†	1 in 641,000	1 in 2,993,000
Hepatitis B	1 in 63,000	1 in 205,000
Hepatitis C	< 1 in 103,300	1 in 1,935,000

*Risk with p24 antigen testing and HIV antibody testing.
†Approximately 67% of infections resulting from transfused blood products are infections with HTLV-II.

of methods of barrier protection.[9] Heterosexual transmission of HCV may be asymptomatic; a donor who was infected via sexual contact but has not yet developed detectable antibodies is a potential risk to the blood supply. Therefore, persons who are sexual partners of known HCV-infected persons may be excluded from donation. Donors found to be ELISA positive for HCV should have supplemental tests, such as the second-generation and third-generation recombinant immunoblot assays (RIBA-2, RIBA-3). Donors with positive supplemental test results are likely to have a chronic HCV infection and require further clinical evaluation.[10] Donors with negative supplemental test results probably had false positive screening results and may be eligible for reentry into the allogeneic donor pool after 6 months.[11] The infection status of donors with indeterminate supplemental results is best resolved by testing for HCV RNA; those with only a single band on the most sensitive supplemental test (RIBA-3) have a less than 4% chance of having circulating HCV RNA.[12]

Gene amplification methods for detecting HCV RNA are used on all blood products before release for transfusion. These tests directly detect the presence of virus before antibody development and are responsible for the current minuscule risk of HCV transmission.[6] Correlation studies have shown that only 80% of samples with confirmed serologic positive results for HCV are also NAT positive. This is consistent with previous estimates of the prevalence of HCV-positive persons who have cleared the virus.[6]

Hepatitis B Transmission of other forms of hepatitis by blood products is extremely rare. Modern testing methods and eliminating the practice of paying whole blood donors have reduced HBV infections to about one in 205,000 units transfused.

Hepatitis A Because the viremic phase of hepatitis A lasts about 17 days in humans before signs and symptoms develop, hepatitis A transmission from single-donor products is extremely rare. Pooled products, such as factor concentrates, however, carry a substantially higher risk.[13]

Hepatitis D Hepatitis D is a defective virus that requires HBV to produce fulminating hepatitis; it is a concern only for patients already infected with HBV.

Hepatitis G The flavivirus hepatitis G, now shown to include several strains, is present in about 4.5% of normal donors. It is transmitted from mother to infant, sexually, and by blood.[14] Circulating hepatitis G RNA is removed after the recipient develops antibodies, which makes the recipient resistant to further infection.[15] HGV has not been linked to any form of clinical hepatitis in children or adults. It is an example of a blood-borne virus without known pathogenicity; as yet, there is no clear reason to remove it from the blood supply.

Screening for Retroviruses

All blood products are screened for HIV-1, HIV-2, human T cell lymphotropic virus type I (HTLV-I), and HTLV-II. Data obtained nationally from American Red Cross donors indicate that the infection risk has been reduced from two per 100 transfusions to about one per 2 million transfusions because of the exclusion of high-risk donors and the postdonation testing for HIV-1 and HIV-2 antibodies and NAT for viral RNA or DNA.[5]

To have predictive value, the ELISA screening test for HIV must be confirmed by Western blot assay. Studies based on polymerase chain reaction (PCR) data, culture data, and donor review all indicate that donors with negative or indeterminate Western blot results are seldom, if ever, HIV positive.[16] A follow-up study of donors who were ELISA positive and whose Western blot results were indeterminate demonstrated that positive ELISA results persisted in about 45% of cases. Of these, 84% still had indeterminate Western blot results, but none were shown to be HIV positive by PCR.[17,18] There have been occasional false positive results of Western blot assays in low-risk donors.[19] The possibility of a false positive result should be remembered when one is counseling low-risk donors who have had unexplained positive results on Western blot testing; these false positive results must always be confirmed by careful clinical follow-up. Data since the introduction of NAT show that less than 6% of confirmed serologic positive results will be NAT negative. On the other hand, only 1.5% of NAT screen reactives were Western blot indeterminate or negative. Virtually none of these NAT reactives were confirmed by discriminatory NAT testing or PCR.[6]

The prevalence of HTLV-I and HTLV-II in United States donors was about 0.03% in 1995. Data from 2001 suggest that the prevalence has been reduced to about 0.01%; about two thirds of these HTLV-positive patients have HTLV-II.[5] Several longitudinal studies have defined the clinical consequences of HTLV-I/II infection[20-22]; they are useful in advising donors who have had positive or indeterminate test results. In a prospective, longitudinal study comparing seropositive blood donors with seronegative blood donors, both viruses were associated with an increase in the incidence of some infectious diseases. No cases of adult T cell leukemia or lymphoma were identified; myelopathies, though rare, were associated with both HTLV types.[22] The risk of HTLV-I/II transmission by blood products is one per 2,993,000.[5] As with HIV, laboratory studies and epidemiologic investigations of HTLV-I/II indicate that patients with positive screening-test results and negative or indeterminate supplemental-test results are unlikely to have clinical sequelae and that the positive results are most likely false positives.[23]

False Positive Test Results during Donor Screening

The causes of false positive test results for HCV and retroviruses are poorly understood. Flu vaccines administered in 1992 were associated with an increase in false positive results for these viruses.[24] However, the proteins responsible for cross-reactivity have not yet been identified. Tests for low-prevalence infections, even tests with excellent specificity and sensitivity, will always be associated with a substantial proportion of false positive results. Consequently, test characteristics, as well as culture and PCR results, can provide reassurance for donors who are not at risk but who have had positive screening-test results and negative or indeterminate confirmatory-test results. As PCR technology improves, it will probably become the most reliable means of establishing whether a positive result represents infection or is a false positive result.

Emerging Infectious Diseases

Until either screening tests or sterilization procedures become available, epidemiologic considerations are the only possible protective strategy against newly recognized infections.[25] For example, transmission of West Nile virus by blood products has led to new donor questions to eliminate donors at risk for this disease, and a nucleic acid–based test for all donated units was introduced in June 2003.[26] In the case of prion diseases such as

variant Creutzfeldt-Jakob disease (vCJD), the restrictions put in place have led to a loss of 4% to 5% of active blood donors and caused transient shortages of certain products such as albumin and immune globulin. In Great Britain, the first case of possible transfusion-transmitted vCJD was reported in December 2003.[27] The report identified 48 recipients of blood from a total of 15 donors who had developed vCJD. In one recipient, symptoms of the disease developed 6.5 years after the possible exposure, a time frame consistent with human-to-human vCJD transmission. The authors estimated the chance of this patient having contracted vCJD independent of the transfusion to be between one in 15,000 and one in 30,000. Recent estimates of the incidence of new vCJD diagnoses in the United Kingdom suggest that there will be fewer than 200 cases from 2001 to 2005 and fewer than 100 cases from 2006 to 2010.[28] Thus, although the current donor restrictions seem prudent, it is important that donors rejected because of epidemiologic risk for vCJD understand the low probability that they actually have a health problem.

Pretransfusion Testing

ANTIGEN PHENOTYPING

Blood recipients are routinely tested to establish their ABO phenotype and Rh type. Establishing ABO type is essential because isoagglutinins (antibodies) against A or B antigens not present on a person's red cells are acquired during the first 2 years of life. These IgM antibodies will cause an immediate hemolytic reaction if ABO-incompatible red cells are transfused.

The terminal carbohydrate on these antigens determines specificity in the ABO system, with type A being associated with *N*-acetylgalactosamine and type B being associated with a terminal galactose. Persons with type O lack both of these terminal sugars. These residues are added by a glycosyltransferase, which was thought to be either nonfunctional or absent in type O persons. Yamamoto and colleagues[29] used molecular techniques to prove that glycosyltransferase in type O persons is very similar to the transferase in type A persons. The type O glycosyltransferase is nonfunctional because of a single base deletion that produces a frameshift and a downstream stop codon.

All methods of ABO typing depend on demonstrating that the antigens found on the red cells are consistent with the expected isoagglutinins [*see Table 2*]. Molecular methods to determine ABO genotype are available.[30] D antigen specificity typing in the Rh system is done because of this antigen's potency as an immunogen. Antibodies to the D antigen are the most important cause of isoimmune hemolytic disease of newborns. Rh antigens are membrane glycolipids or glycoproteins. Antibodies against antigens of this class, which includes the Rh, Duffy, Kell, Kidd, and Lutheran systems, will usually cause shortened red cell survival. In contrast to antigens with carbohydrate-mediated specificity, glycolipid and glycoprotein antigens do not stimulate antibody formation unless the transfusion recipient was previously exposed to allogeneic red cells either from transfusion or from fetal red cells during pregnancy or delivery.

D antigen typing is also done using agglutination techniques. In some cases, less antigenic forms of the D antigen, called weak D, require an antiglobulin reagent to enhance detection. Structural studies of the complementary DNA associated with the major Rh antigens (D, Cc, and Ee) have provided probes for direct genotyping.[31] Molecular methods of prenatal Rh type determination have revealed that most Rh-negative persons lack the *D* gene. Some persons with the weak D phenotype have mosaic *D* genes because of exchange with some of the exons of the *CcEe* gene.

Because the genotypes of many of the clinically relevant red cell antigens are now known, it should now be possible to predict red cell phenotype by DNA analysis. Reid and colleagues[32] were able to correctly predict the red cell phenotype in 60 multitransfused patients by DNA analysis of each patient's white blood cells. This approach, although not yet generally available, will be useful for recently transfused patients, for whom circulating allogeneic red cells complicate antigen phenotyping.

SCREENING FOR ANTIBODIES

In addition to identifying patient ABO and D red cell phenotypes, blood banks must screen serum for red cell–specific antibodies, which can cause serious reactions with transfused red cells. Screening involves testing serum against indicator type O red cells displaying all the clinically important red cell antigens. Positive reactions are detected by adding an antiglobulin reagent (Coombs reagent) to the incubated mixture of type O red cells after it has been washed free of serum. Any observed agglutination is from the reaction of the antiglobulin reagent with antibody adsorbed on the surface of the indicator red cells [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*]. Agglutination of the indicator red cells indicates the presence of other antibodies, which require identification. The absence of agglutination excludes all antibodies except those against antigens so rare that they are not displayed on the indicator red cells.

Use of type-specific blood removes the risk of ABO incompatibility. There is, however, a residual risk of an immunologic reaction from the antibodies to other red cell antigens; such antibodies are present in about 3% to 5% of a random population and in 10% to 15% of persons who were recently transfused or women with a history of pregnancy. Screening for antibodies reduces the frequency of reactions to about 0.06%. Performing a full crossmatch, in which the recipient's serum is tested against the red cells actually being transfused, is of negligible additional benefit, because it excludes only technical errors and the rare antibody that is not detected by the screening. Therefore, a full crossmatch is performed only for persons already known to have made antibodies, because such persons are more likely to form additional antibodies if they are further stimulated by red cell transfusion.

Patients who may receive allogeneic red cells who either have had a transfusion or have become pregnant within the past 3 months must be tested for new antibodies every 3 days. There is no consensus concerning how long the interval should be between patient specimen collection and use of the specimen in pretransfusion testing for patients not recently exposed to red cells. Commonly, specimens are accepted 14 to 28 days before the

Table 2 ABO Typing

Blood Type	*Erythrocytes plus Anti-A Serum*	*Erythrocytes plus Anti-B Serum*	*Antibodies in Patient's Serum*
A	+	0	Anti-B antibodies
B	0	+	Anti-A antibodies
AB	+	+	No antibodies
O	0	0	Anti-A and anti-B antibodies

+—agglutination 0—no agglutination

date for use. However, one study showed that no new antibodies appeared in paired specimens collected up to 1 year apart, suggesting that a longer acceptance interval may be possible.[33]

Blood Components

Most blood donations undergo a fractionation process that allows each component to be used for specific indications. Whole blood can be fractionated into red cells (which contain most of the leukocytes), platelet concentrates (which contain some leukocytes), and plasma [*see Table 3*]. Plasma can be further subdivided into coagulation components and albumin. Each whole-blood unit can potentially support many recipients and clinical needs, maximizing use of each donation.

After 24 hours' storage, whole blood contains no active platelets, and after 2 days, it is deficient in factors V and VIII. Therefore, except for some autologous blood programs that use whole blood rather than packed red cells, use of whole blood has now been almost completely supplanted by therapy employing specific blood components.

RED BLOOD CELLS

The anticoagulant used determines the shelf life of red cells [*see Table 3*]. Citrate-phosphate-dextrose (CPD) with the addition of adenine (CPDA-1) increases storage time from 28 days to 35 days. Most red cells are now stored in CPD to which extra nutrients have been added, which increases storage time to 42 days. This additive solution sometimes contains additional saline, which can be removed if units with very high hematocrits (~70%) are needed.

To prevent transfusion reactions or to delay alloimmunization, red cells are further processed by leukocyte reduction (see below) or washing to remove plasma proteins. Current filter technology reduces white cell counts to less than 5×10^6 cells per unit, a concentration that is sufficient to reduce febrile transfusion reactions and delay alloimmunization and platelet refractoriness. Washing red cells removes the plasma, leaving less than 0.5 ml per unit, a degree of plasma depletion usually effective in treating allergic transfusion reactions. Leukocyte filtration and washing red cells usually shorten the product shelf life to 24 hours, because these procedures require breaking the seal on the plastic bag that contains the red cells, thereby increasing the risk of bacterial contamination. Leukocyte reduction can be accomplished during collection, immediately after collection in the blood bank, or at the bedside during product infusion. Prestorage or laboratory filtration is preferred to bedside filtration.[34] Universal leukoreduction has been implemented in Canada and Europe, but it is not yet required in the United States because of concerns regarding cost-effectiveness.

Table 3 Characteristics of Blood Products and Indications for Use

Product	*Volume (One Unit)*	*Hematocrit (Hct) or Platelet Count*	*White Cell Count*	*Shelf Life*	*Donors per Unit*	*Storage Outside Blood Bank*	*Indication*
Whole blood	450–500 ml (± 10%)	Hct 35–45	$3–5 \times 10^9$	With additive solution, 42 days; with CPDA-1, 35 days; with CPD, 28 days	1	2°–6° C	Massive transfusion if available; exchange transfusions in newborns younger than 3 days
Red cells	With additive solution, 350 ml; with CPD, 250 ml	If additive solution used, Hct 55; if CPD used, Hct 70	$1–2 \times 10^9$	Same as whole blood	1	Same as whole blood	To increase oxygen-carrying capacity; to maintain volume and oxygen-carrying capacity when bleeding
Platelet concentrates	40 ml	$8–9 \times 10^{10}$	$2–6 \times 10^7$	5 days	1/U; given as pool of 5–6 U	Room temperature	For major bleeding or surgical procedures, when platelet count is < 50,000–100,000 μl; for prophylaxis in nonbleeding patients, when platelet count is < 10,000 μl; for bleeding that is refractory to SDPs, when HLA-matched or crossmatched platelets are not available
Single-donor platelets	200–250 ml	$3–5$ ml $\times 10^{11}$	Depends on method of collection	5 days	1	Room temperature	Same as platelet concentrates, but SDPs are preferred because of lower donor exposure
Fresh frozen plasma (FFP)	200–250 ml	—	$< 1 \times 10^5$	1 yr; 24 hr when thawed	1	2°–6° C	Multiple coagulation factor deficiency from bleeding or DIC; reversal of warfarin therapy; factor XI deficiency when factor XI concentrates are unavailable; treatment of TTP
Cryoprecipitate	10–20 ml; pool of 10 U ~ 200 ml	—	$< 1 \times 10^5$	1 yr; 24 hr when thawed	1/U; given as pool of 10 U	2°–6° C wet ice	Replacement of fibrinogen when acutely depleted or when patient cannot tolerate volume load of equivalent amount of FFP (1 pool of cryoprecipitate = 4 FFP); fibrin glue (usually only 1 unit); replacement of von Willebrand factor if concentrate is not available; replacement of factor XIII; replacement for qualitatively abnormal fibrinogen

CPD—citrate-phosphate-dextrose CPDA-1—CPD with adenine DIC—disseminated intravascular coagulation FFP—fresh frozen plasma HLA—human leukocyte antigen SDPs—single-donor platelets TTP—thrombotic thrombocytopenic purpura

Freezing is an alternative method for storing red cells. Red cells can be kept in a cryoprotectant (usually glycerol) for 10 years. Freezing is therefore ideal for storing rare units or autologous units from persons with rare blood types, for whom it is difficult to find compatible allogeneic red cells. When a unit is at the end of its liquid storage shelf life, the cells can be rejuvenated with fresh media and nutrients; they can then be frozen and stored. To be used, frozen red cells must be thawed and the glycerol removed, so preparation time for this product is longer than for products stored in the liquid state. Thawed, deglycerolized red cells generally must be transfused within 24 hours.

PLATELETS

Platelets can be provided either as platelet concentrates from a number of blood donors or from a single donor. SDPs are collected by a continuous apheresis process that removes platelets and returns all other blood components. A single transfusion of platelet concentrates usually consists of platelets derived from four to six units of donated whole blood, which is about the same number of platelets contained in one SDP product. The advantage of SDP therapy is the reduced risk of blood-borne infection and antigen exposure, because the product is from one donor rather than from four to six; disadvantages are a longer collection time, greater cost, and often limited supply. The potential advantages of each of these products have been summarized in a review.[35] ABO Rh–compatible platelets should be used when possible, because studies have shown significantly better therapeutic results from compatible transfusions.[36]

PLASMA

Fresh plasma, frozen within 8 hours of collection (FFP), contains all the procoagulants at normal plasma concentrations. After thawing, it can be kept for 24 hours at 2° to 6° C and will retain 3 to 4 mg/ml of fibrinogen and 1 IU/ml of all the other coagulation components.

Solvent/detergent-treated plasma (S/D plasma) was formerly available as an alternative to FFP with lower infectious risks. However, in 2002 this product was withdrawn from the market by the manufacturer, presumably because of lack of demand for the product and evidence of selective inactivation of certain plasma components. Alternative methods of postcollection sterilization of single units of plasma are becoming available. The main advantage of a postdonation sterilization process is that it protects patients from blood-borne infections, including ones that are not yet recognized. Potential disadvantages are cost and less effectiveness than FFP.

Cryoprecipitate consists of the cryoproteins recovered from FFP when it is rapidly frozen and then allowed to thaw at 2° to 6° C. These cryoproteins include fibrinogen, factor VIII, von Willebrand factor, factor XIII, and fibronectin. About 40% of the components in FFP are recovered. The cryoproteins are suspended in a small amount of plasma that contains ABO isoagglutinin at the concentration found in normal plasma. A pool of 10 units of cryoprecipitate (each derived from one unit of FFP) contains an amount of fibrinogen equivalent to four units of FFP but in one fourth to one fifth the volume. Consequently, a cryoprecipitate pool permits more rapid replacement of fibrinogen than FFP but has the disadvantage of more donor exposures. After the cryoprecipitate is removed from FFP, the residual product is known as cryopoor plasma. Once frozen, cryopoor plasma has the same shelf life as FFP.

Transfusion of Red Cells

INDICATIONS FOR ALLOGENEIC TRANSFUSION

Acute Blood Loss

The decision whether to use red cells depends on the etiology and duration of the anemia, the rate of change of the anemia, and assessment of the patient's ability to compensate for the diminished capacity to carry oxygen that results from the decrease in red cell mass. Management of acute anemia caused by bleeding or operative blood loss will differ from management of chronic anemia to which the patient has adapted. However, the question underlying any red cell transfusion is whether there is sufficient oxygen delivery to tissues for current needs.

Compensatory mechanisms for acute blood loss include adrenergic response, leading to constriction of venous beds, which improves venous return; increased stroke volume, tachycardia, or both; and increased peripheral resistance, which eventually redistributes blood flow to essential organs. Also contributing to the maintenance of intravascular volume is the shifting of fluid to the intravascular space; this shifting occurs relatively rapidly from the extravascular space and more slowly from the intracellular to the extravascular space.[37]

A decrease in blood volume has distinct effects on oxygen delivery, depending on the volume of blood lost and the functioning of the compensatory cardiovascular responses. Restoration of intravascular volume, usually with crystalloid, ensures adequate perfusion of peripheral tissue and is the first treatment goal for a patient with acute blood loss. Whether red cell transfusion is required depends on the extent of blood loss and the presence of comorbid conditions that may limit host response to the blood loss. The American College of Surgeons has correlated blood loss with clinical findings. Loss of up to 15% of total blood volume (class I hemorrhage) usually has little effect; this amount is the maximum permitted in normal blood donation. A class II hemorrhage (15% to 30% loss) results in tachycardia, decreased pulse pressure, and, possibly, restlessness. A class III hemorrhage (30% to 40% loss) leads to obvious signs of hypovolemia; mental status often remains normal. Red cell transfusion is usually indicated when blood loss exceeds 30% in a patient without other significant comorbid conditions. However, the presence of serious cardiac, peripheral vascular, or pulmonary disease can lower this threshold. For example, anemic patients with significant coronary artery disease are more likely to have serious postoperative myocardial complications.

The threshold for red cell transfusion has been evaluated in two randomized, controlled trials. In one study of transfusion after coronary artery bypass, patients who received transfusions for hemoglobin levels below 8 g/dl did no worse than control patients who received transfusions for hemoglobin levels below 9 g/dl.[38] The other trial compared outcomes in critical care patients who received transfusions when their hemoglobin level fell below either 7 g/dl or 10 g/dl.[39] Enrollment in this study was limited to patients who were euvolemic at entry and whose hemoglobin levels were from 7 to 9 g/dl; patients who had undergone routine cardiac procedures or who were actively bleeding upon entry to the intensive care unit were excluded. There was no statistical difference in 30-day mortality for these two groups. However, in the subgroups of patients younger than 55 years and patients whose illness was less severe, as defined by standardized clinical criteria, Kaplan-Meier survival estimates were significantly better in the patients who were not transfused

Table 4 Indications for Platelet Transfusion

Platelet Count (μl)	*Indication for Transfusion*
< 10,000	All patients, even if asymptomatic
< 20,000	Coagulation disorder or minor bleeding
< 50,000–100,000	Major bleeding or surgical procedure

unless hemoglobin levels dropped below 7 g/dl. These results are provocative, but they must be interpreted cautiously. They do suggest that more restrictive transfusion policies may be safely adopted for selected patients. However, the enrollment criteria may have biased the findings, and this calls into question the applicability of these findings to other settings.

Chronic Anemia

In the chronically anemic patient, an increase in red cell 2,3-diphosphoglycerate leads to a shift in the oxygen dissociation curve and improved delivery of oxygen to tissues. This adaptation augments the mechanisms for improved oxygen delivery described above. Indications for transfusion depend on clinical assessment of the adequacy of oxygen delivery and are also guided by the etiology of the anemia. If the anemia can be reversed with iron, folic acid, or vitamin B_{12}, transfusion therapy is indicated only in patients with clinical findings that cannot be tolerated while endogenous red cell mass is being regenerated. Patients with chronic renal disease are typically deficient in erythropoietin. Replacement therapy with exogenous erythropoietin [*see 5:III Anemia: Production Defects*] often obviates transfusion. Patients with anemia that is a result of chronic disease such as rheumatoid arthritis, malignancy, or AIDS may also respond to erythropoietin.[40,41]

Relatively little is known about transfusion thresholds in specific medical illnesses. An observational trial has addressed the effect of anemia on the 30-day mortality of elderly patients hospitalized with acute myocardial infarction. Mortality was reduced in those patients who were transfused to a hematocrit of 30% to 33%, but transfusion had little or no effect on patients who presented with a hematocrit already in the 30% to 33% range.[42]

INDICATIONS FOR AUTOLOGOUS TRANSFUSION

Whether the criteria for autologous transfusion should be the same as that for allogeneic transfusion remains unresolved. Although the risk associated with autologous blood is less than that associated with allogeneic blood, it is not zero. Errors in labeling, storage, and processing can still occur. For these reasons, many argue that uniform standards based on oxygen delivery should apply, regardless of the blood source. Others, citing the reduced risk, advocate returning most or all of the predeposited units to the patient. There is no clinical evidence that either transfusion policy is associated with better or worse patient outcomes.[43]

Intraoperative and postoperative blood salvage can also help limit allogeneic blood use. Blood salvage is employed in procedures associated with the shedding of large volumes of blood; it involves returning concentrated red cells to the patient after those cells have been washed. Preoperative isovolemic hemodilution (PIH) is a process in which blood collected immediately before surgery is returned as needed postoperatively. This strategy has been shown to be a cost-effective alternative to preoperative autologous donation in patients undergoing radical prostatectomy.[44] PIH would be particularly useful if an oxygen-carrying blood substitute were available to replace the autologous blood that is removed. Until it is clear that the cardiovascular risks associated with acute hemodilution do not outweigh the risks associated with allogeneic blood, this approach should be considered with caution.

Transfusion of Platelets

In general, the decision to transfuse platelets rests on the answers to two questions: (1) Is the thrombocytopenia the result of underproduction or increased consumption of platelets? and (2) Do the existing platelets function normally?

INDICATIONS FOR TRANSFUSION

Low Platelet Count

Thrombocytopenia can result from decreased production caused by marrow hypoplasia or from increased consumption caused by conditions such as idiopathic thrombocytopenic purpura (ITP). In a patient with ITP, surviving platelets are larger and younger and function better than would be expected given the platelet count; platelet transfusion is largely avoided or minimized for such a patient. In contrast, with hypoplasia, platelet function is more severely impaired, and the risk of bleeding is relatively higher. Thus, the decision to transfuse patients who have hypoproliferative thrombocytopenia is generally based on their platelet count and is initiated prophylactically when the count drops below a certain threshold. Published consensus guidelines provide an excellent summary of all aspects of platelet therapy.[45]

Studies have shown that the prevalence of bleeding increases significantly below a threshold of about 10,000 platelets/μl in otherwise asymptomatic patients.[46] The desire to avoid allogeneic donor exposure, cost concerns, and increasing platelet demand have encouraged the use of transfusion policies similar to the policy proposed by Wandt and colleagues [*see Table 4*].[45,46]

Nonfunctioning Platelets

Platelet function is the second criterion for the transfusion of platelets. Transfusion is appropriate in a bleeding patient whose platelet count is adequate but whose platelets are nonfunctional as a result of medications such as aspirin or nonsteroidal anti-inflammatory drugs or as a result of bypass surgery. In a bleeding patient, if platelet dysfunction is from inherited or acquired defects, transfusion is indicated to provide a minimum number of normal platelets. Platelet function is abnormal in uremic patients, and definitive treatment requires correction of the uremia. Some studies suggest that interventions that increase von Willebrand factor levels, such as desmopressin (1-desamino-8-D-arginine vasopressin [DDAVP]) conjugated estrogen, or cryoprecipitate, may favorably influence platelet function in uremia.[47] In vitro evidence suggests that DDAVP may improve platelet dysfunction caused by glycoprotein IIb/IIIa (GPIIb/IIIa) inhibitors (e.g., eptifibatide, abciximab, tirofiban) or aspirin.[48]

CONTRAINDICATIONS TO PLATELET TRANSFUSION

Proper investigation of the causes of thrombocytopenia will identify clinical situations in which platelets should be withheld because they contribute to evolution of the illness. These disorders include thrombotic microangiopathies such as TTP, hemolytic-uremic syndrome, and HELLP syndrome (hemolysis, elevated liver enzymes, and a low platelet count). Posttransfu-

sion purpura is usually unresponsive to platelet transfusion but may respond to plasma exchange or intravenous immunoglobulin (IVIg). Platelet transfusions will not help patients with autoimmune thrombocytopenia (e.g., ITP), but they also will not harm them.

RESPONSE TO PLATELET TRANSFUSIONS

Both platelet and host factors influence the response to platelet transfusions. Length of in vitro storage, storage temperature, adequacy of oxygenation, and extent of pretransfusion manipulation all influence in vivo survival. Important host factors that influence survival are temperature, splenomegaly, ABO compatibility, and immune status.

A transfusion of appropriately stored fresh platelets—whether pooled concentrates or SDPs—should contain about 6,000/μl to 10,000/μl platelets per unit (5.5×3^{10} platelets). Thus, in an unsensitized 75 kg (165 lb) recipient, each unit should yield an increment of about 60,000 platelets/μl. When needed, a posttransfusion count is usually obtained after 1 hour but can be obtained as early as 10 minutes after transfusion. A patient is considered refractory to platelet transfusions when the 1-hour posttransfusion increment is less than 10,000 platelets/μl after the patient is given 3.3×10^{11} platelets.

PLATELET TRANSFUSIONS IN REFRACTORY CASES

Platelets have platelet-specific antigens, human leukocyte antigens (HLA), and blood group antigens. Immune response to any of these can contribute to platelet unresponsiveness. Platelet surfaces have only class I HLA antigens, of which only HLA-A and HLA-B are clinically important. Polymorphic antigens are found in association with each of the major platelet proteins: HPA1a/2a (formerly called $Pl^{A1/A2}$) and Pen on glycoprotein IIIa, Bak system on glycoprotein IIb, and Br and Ko on glycoproteins Ia and Ib. Each of these antigen groups is associated with isoimmune neonatal thrombocytopenia. The prevalence of antibodies to platelet-specific antigens is increased for patients sensitized to HLA antibodies; therefore, antibodies to both sets of epitopes may contribute to refractoriness in patients who fail to respond to HLA-matched platelets.[49]

Treating a patient refractory to platelet transfusions involves addressing nonimmune causes (e.g., fever, sepsis, bleeding, and disseminated intravascular coagulation [DIC]) and providing recently collected ABO-compatible products. If these strategies fail, minimization of the effects of HLA antibodies or platelet antigens through HLA typing, platelet crossmatching, or both is indicated.[50] Selecting platelets matched at the HLA-A and HLA-B loci may improve responsiveness in about half of patients with positive HLA antibody screens. Unless contraindicated because of transplant considerations, an empirical trial of donations from family members may also be helpful.

In one study undertaken to determine the best method of treating refractory cases, platelet selection by crossmatching was compared with selection by HLA criteria. Selection by crossmatching was equivalent to HLA selection and yielded better results.[51] Another study found that crossmatched platelets provided equivalent platelet increments that were independent of the grade of HLA match.[52] Although these results are promising, the effectiveness of selection either by HLA and crossmatching or by crossmatching alone is often limited by nonimmune host factors. Additionally, these techniques are not yet routinely available.

Modifying the effects of alloimmunization is difficult. IVIg can improve platelet increments but not platelet survival. An analysis of IVIg therapy found that about 50% of alloimmunized patients appeared to benefit from such therapy.[50] Plasma exchange is of limited value because it is difficult to remove IgG antibodies. In some patients, the HLA antibodies responsible for refractoriness may regress over time, so it is important to periodically retest for them. If the HLA antibody screen becomes negative, a trial of non–HLA-matched platelets is warranted.

All in all, the best strategy is prevention, which can be achieved by avoiding unnecessary transfusions and using only leukocyte-depleted products. A randomized, prospective trial of how best to prevent alloimmunization of newly diagnosed patients with acute myeloid leukemia showed equivalent rates of alloimmunization and platelet refractoriness for filtered platelet concentrates, filtered SDPs, and ultraviolet B–irradiated platelets.[53] However, leukocyte reduction did not prevent secondary immune responses in patients already sensitized through either pregnancy or transfusion.[54]

Transfusion of Fresh Frozen Plasma, Plasma Derivatives, and Recombinant Products

FRESH FROZEN PLASMA

Despite a paucity of indications for FFP use, roughly two million units are transfused annually[55] [*see Table 3*]. FFP is most appropriate for replacing the multiple coagulation deficiencies that result from massive transfusion, liver disease, warfarin toxicity, or acute or chronic DIC. In addition, it can be used to treat thrombotic microangiopathies and specific factor deficiencies when factor concentrates are not available. After one blood volume exchange using only red cells, plasma components are diluted to about 40% of their original concentration; after two blood volume exchanges, plasma components are diluted to 15%. Prothrombin time (PT) and partial thromboplastin time (PTT) become prolonged when coagulation components are lower than 30%, but abnormal bleeding from dilution usually does not occur until these values are less than 17% of normal. Microvascular bleeding associated with a PT and PTT greater than 1.5 times normal is an indication for FFP.[56] Whether FFP replacement is needed when PT and PTT are over 1.5 times normal but not associated with bleeding is less clear-cut; paracentesis and thoracentesis did not cause increased bleeding in patients with PT and PTT that were up to twice normal values.[57]

The FFP dose depends on whether or not a consumptive process is being treated in addition to hemodilution. For hemodilution, 15 ml/kg will usually be sufficient. However, if consumption is occurring, the dose is best guided by the effect of treatment on PT and PTT. If fibrinogen is lower than 80 mg/dl, cryoprecipitate may be required to rapidly increase fibrinogen. However, four units of FFP can be used in most cases to provide the same amount of fibrinogen as one pool of cryoprecipitate. Urgent reversal of the effects of warfarin can usually be accomplished with about 5 to 10 ml/kg of FFP.

Factor XI concentrates, which still have some thrombogenic potential, are available but not yet licensed in the United States.[58] Therefore, FFP is the treatment for factor XI deficiency. FFP is not used to replace antithrombin III, because a purified concentrate is available.[59]

Thrombotic microangiopathies [*see 5:XIII Hemorrhagic Disorders*] are treated with either FFP transfusions or, more often, plasma exchange with either FFP or cryopoor plasma.[60] Studies suggest that cryopoor plasma may be an alternative to FFP in the

treatment of TTP.[61] The dose of either product is usually equal to a plasma volume exchange of 1.0 to 1.5, which is carried out daily until clinical improvement occurs.

FACTOR VIIA

Recombinant activated factor VII was approved in 1999 for the treatment of bleeding episodes in patients with hemophilia A or B who have antibodies (inhibitors) to factor VIII or IX, respectively. Factor VIIa is also the treatment of choice for the rare patient with factor VII deficiency, whether acquired—as, for example, a consequence of liver disease—or inherited. It is not approved for this purpose, however. In addition, factor VIIa is useful in activation of the coagulation tissue factor pathway. For patients with inhibitors, factor VIIa is given at a dosage of 90 μg/kg as a slow I.V. push over 2 to 5 minutes; the dosage is repeated every 2 hours, as needed. For factor VII deficiency, the dosage is 20 to 30 μg/kg given as a slow I.V. push over 10 minutes; given in this manner, factor VIIa treatment will reduce the PT to normal within 20 minutes after administration. Depending on the clinical setting, the PT will become prolonged again 3 to 4 hours after treatment.

Off-label use of this product to treat uncontrolled hemorrhage in patients who do not have a preexisting bleeding disorder and who are unresponsive to FFP is becoming more common.[62] However, it is important not to use factor VIIa in patients with DIC, because VIIa may exacerbate the DIC. The high cost of this product and its potential for contributing to DIC should limit its use to carefully selected patients for whom other alternatives are not available.

FACTOR VIII CONCENTRATES

The introduction of plasma-derived factor VIII concentrates in the 1960s brought a significant improvement in the treatment of hemophilia A. Unfortunately, these concentrates were derived from large donor pools, and contamination of the factor with HBV, HVC, and, especially, HIV resulted in the widespread transmission of these infections in the hemophilia community. Since 1980, new methods of heat sterilization, solvent/detergent treatment, and immunoaffinity purification have yielded an array of factor concentrates that are highly purified and unable to transmit these infections. The efficacy of these viral-inactivation methods has been validated by using reverse transcription and PCR studies to measure HCV RNA in factor VIII concentrates; HCV RNA was present in 100% of products before treatment but was undetectable after treatment. Besides reducing the risk of infection, use of high-purity factor VIII concentrates may be associated with better preservation of patients' cell-mediated immunity. The 1980s also saw the advent of recombinant factor VIII concentrates.

The factor VIII preparation Humate-P is also rich in von Willebrand factor and is approved for the treatment of von Willebrand disease. This product has the major advantage of being free of the risks of infection associated with cryoprecipitate. If Humate-P is not available, the factor VIII preparations Alphanate or Koate-DVI may be used, but they are not approved for this purpose and their efficacy is uncertain.

The advances in safety and purity of factor VIII concentrates, especially in the case of the recombinant products, have increased the cost per unit fivefold to 10-fold. Recombinant products are used primarily for newly diagnosed patients with hemophilia who have not been exposed to plasma products. Work is just beginning on modifying these recombinant products to make them more effective by reducing immunogenicity and prolonging circulation time.[63]

The possibility that a nonhuman source of factor VIII would be useful in the treatment of patients with acquired factor VIII inhibitors led to the development of a highly purified porcine factor VIII concentrate. This was shown to be effective for patients whose anti–factor VIII antibody does not cross-react with the porcine product.[64] About one third of patients develop antibodies to the porcine product, which limits its usefulness for repeat treatments.

FACTOR IX CONCENTRATES

Factor IX complex concentrates contain about equal amounts of the vitamin K–dependent factors II, VII, IX, and X. These preparations are available in several degrees of purity, but all have the disadvantage of being thrombogenic when used for extended periods or in patients with liver disease. Highly purified factor IX, prepared by immunoaffinity chromatography, is free of this complication and is the product of choice in treating factor IX deficiency.[65] Activated prothrombin complex concentrates (Autoplex-T and FEIBA) have been used to bypass the need for factor VIII in selected patients with hemophilia A and acquired inhibitors. This provides an alternative for patients who do not respond to porcine factor VIII [*see 5:XIII Hemorrhagic Disorders*].

Transfusion of Granulocytes

Studies have shown granulocyte transfusion to be effective in the treatment of neutropenic patients. Transfusion of granulocytes in doses in the range of 8.3×10^{10} can be obtained by apheresis of donors who have been pretreated with granulocyte colony-stimulating factor (G-CSF) and a single dose of dexamethasone. Granulocyte transfusions at these dose levels have been shown to produce measurable, sustained increments in neutrophils, even into the normal range. The indications and clinical benefits of granulocyte transfusion at these higher doses are still being determined. Randomized trials are required to fully define the clinical efficacy of granulocyte transfusions. After collection, granulocytes must be stored at room temperature and irradiated to prevent transfusion-associated graft versus host disease. Crossmatching should be done to ensure compatibility.[66]

Transfusion of Immune Globulin

Many human immune globulin preparations are available. Immune serum globulin, administered intramuscularly, is used to treat chronic immunodeficiency disease and for prevention or alleviation of measles. Hepatitis A can now be prevented by vaccination [*see 4:VII Acute Viral Hepatitis*]. Alternatively, a traveler who will spend less than 3 months in an endemic area can receive 0.02 ml/kg of immune serum globulin. Hepatitis B immune globulin is used for postexposure prophylaxis against HBV infection [*see CE:V Adult Preventive Health Care*]. It is prepared from plasma with high titers of antibody to hepatitis B surface antigen. Rh_0(D) immune globulin is used to prevent the development of anti-Rh_0 (anti-D) antibodies in Rh-negative women who have just given birth, undergone amniocentesis, or aborted, if the biologic father is thought to be Rh positive.

Intravenous administration of human immune globulin promptly elevates circulating IgG levels and is preferable to intramuscular administration. Several preparations of IVIg are

available to treat chronic immunodeficiency disease[67] [*see 6:VIII Deficiencies in Immunoglobulins and Cell-Mediated Immunity*]. The intravenous dosage for such deficiency syndromes is 0.2 g/kg/mo but can be raised to 0.3 g/kg/mo or the agent can be given more often if needed.

The most common side effects of IVIg therapy—headache, nausea, and fever—usually respond to symptomatic treatment and reduction of the infusion rate. Rarer and potentially more severe side effects are anaphylactic reactions, hemolysis from anti-A and anti-B antibodies, and acute renal failure. Renal failure has been attributed to osmotic nephrosis caused by the high sucrose concentration in many IgG preparations.[68,69] In one study, aseptic meningitis was the most common of the serious side effects, with a frequency of 11% (95% confidence interval, 4% to 23%); patients with a history of migraine had a significantly higher incidence of aseptic meningitis.[70]Aseptic meningitis usually occurs within 24 hours after administration and does not respond to a reduction of the infusion rate. Patients may be required to stay in the hospital for symptomatic treatment; if further treatment is needed, changing the lot or preparation of IVIg may alleviate this side effect. Current manufacturing practices eliminate HCV from IVIg preparations.

Transfusion of Stem Cells

Stem cell transplantation, initially pioneered for use in leukemia, is used to treat a number of life-threatening, malignant, hereditary, and immunologic disorders [*see 12:XV Chronic Lymphoid Leukemias and Plasma Cell Disorders*].

Complications of Transfusions

HEMOLYTIC TRANSFUSION REACTIONS

Hemolytic transfusion reactions are classified as immediate or delayed, depending on their pathophysiology. Immediate hemolytic reactions are the result of a preexisting antibody in the recipient that was not detected during pretransfusion testing. Delayed hemolytic reactions are the result of an anamnestic response to an antigen to which the recipient is already sensitized. The renewed antigenic stimulation in a person already primed by previous antigenic exposure results in recrudescence of antibody to levels that can cause hemolysis. This is in contrast to an immune response during primary sensitization, which seldom causes hemolysis, because antibody levels develop at a much slower rate.

Patients with sickle cell disease are more likely than others to become alloimmunized and to have delayed hemolytic transfusion reactions, which often occur in association with recrudescence of an occlusive pain crisis. These reactions are occasionally associated with severe hemolysis involving autologous, as well as allogeneic, red cells. The cause of these episodes is unknown but has been attributed to so-called bystander hemolysis associated with abnormal function of CD59 (MIRL, membrane inhibitor of reactive lysis), transfusion-associated marrow suppression, or both.[71]

Diagnosis of Hemolytic Reactions

The pathophysiologic differences between immediate and delayed hemolytic transfusion reactions account for some of their differences in clinical findings. Fever is a common sign associated with both immediate and delayed hemolytic transfusion reactions.

Clinical evidence of hemolysis is likely to be more severe in immediate hemolytic reactions and may include back pain, pain along the vein into which the blood is being transfused, changes in vital signs, evidence of acute renal failure, and signs of developing DIC. These findings are probably caused by immune complexes activating the complement and kinin systems, by the direct effects of red cell stroma on kidney function, and possibly by the release of inflammatory cytokines such as interleukin-1β (IL-1β), IL-6, and tumor necrosis factor (TNF).[72]

In delayed hemolytic reactions, hemolysis with hemoglobinemia and hematuria (sometimes associated with renal failure) also occurs, but it is less common and generally less severe. In many delayed hemolytic transfusion reactions, the only clinical findings may be a newly positive Coombs test result, the appearance of a new antibody against red cell antigens that are not present on the recipient's red cells, or both. When hemolysis is absent, these reactions are sometimes called delayed serologic transfusion reactions. At the Mayo Clinic, two surveys sought to identify the relative incidence of both kinds of delayed transfusion reactions. The most recent survey, covering the period from 1993 to 1998, revealed a relative increase in delayed serologic transfusion reactions and an associated decrease in delayed hemolytic reactions, with overall increases in the incidence of these reactions. The earlier survey, which covered the period from 1980 to 1992, revealed an association between delayed transfusion reactions and the presence of antibodies to Jk^a and Fy^a or antibodies with multiple specificity; this association was not found in the later survey. These changes probably result from improved systems for identifying clinically significant nonhemolytic antibodies.[73]

In some cases, antiglobulin testing may yield positive results after all the transfused cells have been cleared, often with only complement being detected on the red cells. This finding has been attributed to autoimmune hemolysis after the delayed transfusion reaction.

Treatment of Hemolytic Reactions

As soon as a hemolytic transfusion reaction is suspected, the transfusion should be immediately discontinued. The diagnosis can be confirmed or excluded by sending the remaining blood product, together with a freshly drawn posttransfusion specimen, to the blood bank. The blood bank rechecks all records, confirms the patient's type and antibody screen, checks for evidence of hemoglobin in the plasma, and rechecks the crossmatch and antiglobulin test results. These tests will confirm or disprove the diagnosis and identify the antibody causing the immediate hemolytic reaction, when present. Until these studies have been completed, any further blood products can be given only with the approval of the blood bank's medical director.

The side effects of an acute hemolytic transfusion reaction can be managed by supporting renal blood flow with furosemide and supporting tubular urine flow with mannitol; treating shock, if required, with pressors; and giving platelets and FFP as needed to control coagulopathy if DIC develops. Intravenous steroids may be useful. Until the antibody causing the immune hemolysis is identified, only type O red cells and AB plasma should be used.

Managing delayed transfusion reactions is simpler because of the slower tempo at which these reactions develop. The diagnosis requires identifying a new antibody against red cell antigens and searching for clinical evidence of hemolysis. Treatment requires replacement with the appropriate antigen-negative blood

products. Acute renal failure and DIC are unlikely but would be managed as described for immediate hemolytic reactions. The severe, atypical delayed transfusion reactions sometimes found in patients with sickle cell disease may require steroids and transfusion support.

Prevention of Hemolytic Reactions

Prevention of immediate and delayed hemolytic transfusion reactions depends on recognizing their respective proximate causes. Immediate hemolytic reactions are usually caused by technical errors made during the procurement or processing of blood specimens, during pretransfusion testing, or during product infusion. In a review of transfusion-related deaths reported to the Food and Drug Administration between 1990 and 1998, approximately 50% were caused by clerical errors that led to transfusion of ABO-incompatible blood, a rate virtually unchanged since reporting began in 1976.[74] Prevention of immediate transfusion reactions is best accomplished by following protocols for obtaining specimens from patients in adequate time before transfusion and checking to see that blood products are appropriate for the intended recipient.

Delayed transfusion reactions are the result of an anamnestic response of antibodies from a previous transfusion (or pregnancy) that are not present in detectable levels at the time the specimen is crossmatched. A careful transfusion history can best prevent delayed hemolytic reactions. Many patients will know whether there were difficulties involving blood obtained for transfusion. If a patient has a history of difficulty with crossmatches, the blood bank can obtain the details from the institution responsible for the previous transfusion. A proper transfusion history can uncover patients likely to have antibodies that the blood bank would not detect. For example, antibodies to Jk^a and Fy^a are characteristically hard to identify because they are quick to rise on stimulation and fall equally rapidly, making later detection difficult.

FEBRILE TRANSFUSION REACTIONS

Nonhemolytic febrile transfusion reactions occur in 1% to 2% of all transfusions and are more likely to occur after platelet transfusions. Until recently, febrile transfusion reactions were attributed to recipient antibody reactions against HLA antigens on donor leukocytes in the transfused product. It is now believed that cytokines produced during storage may also contribute to these reactions.[75] This conclusion is based on observations that platelet products associated with transfusion reactions have higher levels of inflammatory cytokines such as IL-1β, TNF, IL-6, and IL-8 in the supernatant than are found in platelets that do not cause febrile transfusion reactions.[75]

Diagnosis of Febrile Reactions

Febrile reactions are characterized by the development of fever during transfusion or within 5 hours after transfusion. These reactions may be limited to an increase in body temperature of 1° to 2° F but are often associated with chills and rigors.

The differential diagnosis for a patient undergoing a nonhemolytic febrile transfusion reaction should always include unrecognized sepsis. When febrile reaction is suspected, immediate management consists of discontinuing the transfusion, obtaining appropriate cultures, and returning the product to the blood bank. The blood bank obtains cultures from the product and verifies that no errors have occurred in its preparation. The probability that a febrile transfusion reaction has occurred is influenced by the type of product, the number of white cells contained therein, and the transfusion history of the recipient. Febrile reactions to products that have few or no white cells, such as deglycerolized red cells or FFP, are unusual. Unmodified whole blood and red cells contain between 1.3×10^9 and 3×10^9 white cells and are much more likely to cause febrile reactions. In the case of platelets, reactions can be from cytokines made during in vitro storage or from bacterial contamination.

Treatment of Febrile Transfusion Reactions

Febrile transfusion reactions are usually self-limited and respond to symptomatic management with antipyretics. However, symptoms may be of sufficient magnitude to require the use of 50 to 75 mg of meperidine by intravenous bolus. To prevent further occurrences, leukocyte-depleted products are indicated for patients who have had two or more febrile transfusion reactions.

Prevention of Febrile Reactions

Newer designs of filters for leukocyte reduction should decrease the white cell content to below the threshold for febrile transfusion reactions. Because inflammatory cytokines may be involved in febrile transfusion reactions, methods are being implemented to accomplish leukocyte reduction either during or after collection but before storage. In a study comparing products that underwent leukocyte reduction either before storage or at the bedside, significantly fewer febrile reactions occurred in patients receiving prestorage leukocyte-depleted products; there was no difference in the number of allergic reactions.[76]

Prestorage leukocyte reduction is particularly important for platelets because platelets are stored at room temperature and accumulate significantly more cytokines than do red cells, which are refrigerated. Febrile transfusion reactions are also more likely with older products. In one study, platelets that were used after they were in storage for 3 days or less were found to cause significantly fewer febrile transfusion reactions than platelets that were used after longer storage periods.[77] Unfortunately, testing for infectious diseases often takes 2 to 3 days, during which time the product cannot be used. It is therefore impractical to rely on younger products to reduce the risk of febrile transfusion reactions. Other benefits of leukocyte reduction are prevention of HLA alloimmunization; prevention of transmission of leukocyte-bound viruses such as cytomegalovirus (CMV), Epstein-Barr virus, HTLV-I, and HTLV-II; and, possibly, reduction of immune modulation.[78]

Whether these advantages justify leukocyte reduction for all blood products remains an unsettled issue because it is unclear whether the benefits justify the associated increased costs. Managing patients who continue to have febrile reactions after receiving leukocyte-depleted products is a clinical problem for which there are no clear solutions. In addition to premedication with steroids, use of HLA-matched products for patients demonstrated to have HLA antibodies may be helpful. Occasionally, use of washed products is beneficial.

TRANSFUSION-RELATED ACUTE LUNG INJURY

Transfusion-related acute lung injury (TRALI) usually presents as bilateral pulmonary infiltrates within 4 hours after transfusion.[79] The clinical and radiographic picture is that of normal-pressure acute respiratory distress syndrome (ARDS); therefore, the differential diagnosis is sufficiently broad to make the possible causal role of transfusion often go unnoticed. Current evidence suggests that TRALI is associated with the interaction of

antibodies (HLA class 1 or class 2 antibodies, antimonocyte antibodies, or antigranulocyte antibodies), with the corresponding antigens on monocytes or granulocytes. A recent study found such associations in 14 of 16 TRALI patients.[80,81] These interactions cause endothelial injury, alveolar exudation, and the associated clinical findings of ARDS.

A second form of TRALI that involves two clinical events has been proposed.[82-84] In this two-event model, which is based on clinical findings and rat lung studies, the first step is the priming of neutrophils by mediators that arise in certain clinical settings (e.g., recent surgery, massive transfusion, cytokine therapy, or infection).[82] The primed neutrophils adhere to pulmonary endothelium and are activated by a second event, such as exposure to biologically active lipids from blood products.

Diagnosing TRALI depends on excluding cardiac and other causes of ARDS. Demonstration of antileukocyte antibodies helps confirm the diagnosis, but their absence does not exclude TRALI in the appropriate clinical setting. Early diagnosis is important because most patients will improve within 24 hours after conservative treatment is initiated. Unfortunately, a minority of patients develop TRALI associated with severe pulmonary edema and fluid filling the trachea, for which no effective therapy exists.

ALLERGIC TRANSFUSION REACTIONS

Allergic transfusion reactions are more common than febrile nonhemolytic transfusion reactions, occurring in 3% to 4% of transfusions. Allergic transfusion reactions usually present as pruritus and urticaria. A small percentage of patients have anaphylactoid symptoms, including wheezing, bronchospasm, and, occasionally, true anaphylaxis.[85] These reactions had been attributed to an immune response to plasma proteins. However, a 1999 study suggested that they may instead be provoked by increased levels of RANTES (regulated on activation, normal T cell expressed and secreted), an inflammatory chemokine that is stored in platelet alpha granules and accumulates during storage.[86] This is an intriguing hypothesis, because RANTES is known to affect eosinophil and basophil function.

In most cases, symptoms of allergic reactions are local and do not require discontinuance of the transfusion if they are controlled with antihistamines. There is, however, no means as yet to identify the rare patient who will progress to anaphylaxis.[85] It is known that IgA deficiency is associated with an increased likelihood of anaphylaxis, but many patients who are IgA deficient never have any difficulty.

For most patients with urticaria, which seldom progresses to anaphylaxis, management is symptomatic. However, patients known to be IgA deficient should receive cells that have been washed to remove plasma. When plasma products are required, they should be administered in a facility equipped to manage anaphylactic reactions. Using IgA-deficient plasma can minimize the risk, but such plasma is difficult to obtain and may require drawing from a rare donor pool, testing family members, or both.

ATYPICAL TRANSFUSION REACTIONS

Occasionally, patients have reactions that do not fit the categories already defined but clearly seem related to blood transfusion. These reactions have mainly consisted of severe hypotension after platelet infusions. No allergic features are present. The reactions are associated with blood-product infusions through a negatively charged leukocyte reduction filter, and they often occur in patients who are receiving angiotensin-converting enzyme (ACE) inhibitors. A recent study suggests that such reactions may be caused by excessive accumulation of des-Arg9-bradykinin. This metabolite of bradykinin is known to be vasoactive and to be metabolized by ACE.[87] Clinical observations suggest that atypical hypotensive reactions are more likely to occur in patients receiving ACE inhibitors during plasma exchange, hemodialysis, low-density lipoprotein apheresis, IgG-affinity column apheresis, and desensitization immunotherapy. These findings have led to the suggestion that ACE inhibitors should be withheld for 24 hours before any of these procedures are initiated. Such reactions are sufficiently rare that it may be adequate to limit this restriction to patients who have already experienced one of these reactions.[88]

TRANSFUSION-ASSOCIATED GRAFT VERSUS HOST DISEASE

The diagnosis of transfusion-associated graft versus host disease (TA-GVHD) should be considered in any patient who presents after transfusion with fever, skin rash, and diarrhea and has pancytopenia and abnormal results on liver function tests.[89] Signs and symptoms in neonates are similar to those in adults, but fever and rash develop later: in adults, fever occurs after a median of 10 days after transfusion; in neonates, fever occurs after 28 days, with rash appearing 1 to 2 days later.[90] TA-GVHD is a much-feared consequence of transfusion therapy because mortality approaches 100%. It results from transfusing immunocompetent lymphocytes into a recipient who is unable to reject the allogeneic cells. Reaction of the transfused lymphocytes with host antigens leads to the multiple manifestations of TA-GVHD.

TA-GVHD is best prevented by identifying potentially susceptible recipients. Patients who are at significant risk for TA-GVHD include premature infants receiving large doses of allogeneic lymphocytes, patients with congenital defects in cellular immunity or immunity resulting from illness or chemotherapy, and patients who are unable to reject infused cells because of shared antigens with the allogeneic lymphocytes. Patients undergoing autologous or allogeneic bone marrow transplantation are particularly at risk. Many case reports document the association of Hodgkin disease with TA-GVHD, which occurs presumably as a result of acquired defects in T cell immunity. The intensive chemotherapy that is used to treat leukemia, high-grade lymphomas, and solid tumors may also set the stage for TA-GVHD. However, no cases have been identified in AIDS patients. One hypothesis explaining this surprising finding is that the HIV-mediated injury to $CD4^+$ T cells blocks the development of TA-GVHD.[91]

Patients whose risk for TA-GVHD is a result of receiving transfusions from a homozygous donor of a shared haplotype are the hardest to identify a priori. Donor lymphocytes are not rejected by the recipient but do respond to the nonshared recipient haplotype. This mechanism probably accounts for the majority of cases of TA-GVHD. The chances of receiving haplotype-homozygous blood from an unrelated donor vary with different populations. In Japan, the risk for adults may be as high as 1 in 874; it is estimated to be 1 in 102 in neonates because of the use of fresh whole blood from family members.[92] In France, the risk is estimated to be 1 in 16,835. In the United States, the risk for the white population is thought to be about 1 in 7,147. Risk increases if first-degree relatives are donors.

Once patients at risk are identified [*see Table 5*], pretreatment of all cellular transfused products with gamma radiation is indicated. On the basis of in vitro studies, the current recommended

Table 5 Patients for Whom Irradiated Blood Products Are Recommended

- Fetuses and neonates
- Patients with congenital immunodeficiency
- Allogeneic and autologous bone marrow transplantation patients
- Recipients of some solid-organ transplants*
- Patients with hematologic malignancies†
- Patients with nonhematologic malignancies, especially if undergoing intensive chemotherapy‡
- Recipients who may share haplotypes with donor§

*No consensus, but most agree that heart, liver, and lung recipients should receive irradiated products, whereas recipients of renal allografts do not require irradiated blood.
†Patients with low-grade lymphomas and leukemias in remission may not require irradiated products. Applying restriction to all lymphoma and leukemia patients avoids mistakes.
‡Except for immunosuppression from intensive chemotherapy, there is no consensus.
§Donors in this group include directed donors and first- and second-degree relatives.

dose is 2,500 cGy, which does not affect red cell function or platelet survival if administered immediately before transfusion.[93] However, irradiated red cells stored for 42 days show significant increases in plasma potassium and hemoglobin and a small but significant decrease in cell survival. Consequently, recommended storage after irradiation is only 28 days; most institutions prefer to irradiate immediately before product release when possible. Platelets have normal storage survival 5 days after irradiation and can be irradiated at regional centers before distribution. Leukocyte reduction may provide some protection against TA-GVHD, which is related to the dose of lymphocytes. However, filtration alone is not preventive and must never be used as a substitute for gamma irradiation. Because of the risk associated with a one-way HLA match, blood-bank standards require that family members' blood and blood of directed donors be irradiated.

Treatment of TA-GVHD remains ineffective. Prevention by providing irradiated blood products to all recipients may become the most practical solution to this complication.[94]

BACTERIAL AND PROTOZOAN INFECTIONS

Platelets are associated with the majority of cases of transfusion-related sepsis because the platelets are stored at room temperature.[95] Controlling this problem requires improved disinfection of skin, better detection of subclinical infection, and development of methods for storage at lower temperatures or postcollection sterilization. If sepsis is suspected in patients who have been given red cells, the possibility of *Yersinia enterocolitica* infection should be considered.[96] This organism can grow in the cold, iron-rich environment provided by stored red cells. When such infections occur, the blood is almost always at least 2 weeks old; this period corresponds to the time needed for the usually small initial inoculum to reach clinically significant amounts. Malaria infections have been almost completely eliminated by predonation screening. *Trypanosoma cruzi* can cause a chronic parasitic infection; the incidence of blood-borne transmission has increased to the point that pretransfusion testing for it may soon be needed. Spirochetes cannot be transmitted by products that have been stored longer than 80 hours and are no longer considered a clinically significant source of blood-borne infection.

CYTOMEGALOVIRUS INFECTION

CMV is a common blood-borne infection of no clinical consequence to healthy, immunocompetent recipients, but it can be a severe problem for patients with either acquired or congenital immunodeficiency [*see Table 6*]. Judged on the basis of screening for antibody to CMV, more than 40% of healthy donors may have the potential to transmit CMV.

There are two approaches to preventing CMV transmission. The first is to use CMV antibody–negative products. The second, more practical approach is to use leukocyte-reduced products, because CMV is transmitted only by leukocytes. On the basis of a prospective, randomized study of more than 500 transplant patients, products that have undergone leukocyte reduction to the current standard of fewer than 0.5×10^6 leukocytes per milliliter are considered to be as effective as seronegative products in preventing CMV infection.[97] It is unclear which product provides the best protection against transfusion-associated CMV infection.[98] Direct comparisons between seroconversion rates after transfusion of prestorage leukocyte-depleted products and seroconversion rates after transfusion of CMV-negative products are required to settle this issue.

IMMUNE MODULATION AS A RESULT OF TRANSFUSION

Evidence that transfusions result in modulation of host immunity has come from studies of transplantation, cancer recurrence, and posttransfusion infection rates. The effect was first observed in cadaver-kidney transplantation; patient survival was shown to increase with increased transfusions. Although this benefit became less important with the introduction of cyclosporine, Opelz and colleagues[99] found increased cadaver-graft survival in transfused recipients whose immunosuppression regimen included cyclosporine.

The hypothesis that immune modulation is related to infused white cells has been supported by studies in animal models and by clinical observations of tumor recurrence and posttransfusion infection rates. Bordin and colleagues[100] have shown in a rabbit model that the number of pulmonary metastases is increased by allogeneic blood transfusions but not by blood from syngeneic littermates. This effect of allogeneic blood is abrogated by prestorage leukocyte reduction but not by poststorage reduction. Randomized clinical studies of posttransfusion infection and cancer recurrence have produced conflicting results.[101] The conflicting data concerning the magnitude and clinical relevance of transfusion-induced immunomodulation need to be resolved. If leukocyte reduction is shown to reduce posttransfusion infections and cancer recurrence, the argument for universal leukocyte reduction of cellular blood products, which is already strong, would become irrefutable. Until this matter is settled, the possible immunomodulatory effect of blood transfusion is another reason to avoid allogeneic blood transfusion whenever possible.

Table 6 Patients for Whom Cytomegalovirus-Negative Blood Products Are Recommended

- Neonates, especially if weight is less than 1,200 g
- Pregnant women, as a means of preventing primary intrauterine infections
- Recipients of solid-organ transplants, especially when the recipient and the organ donor are both CMV negative
- Patients with severe combined immunodeficiency

Apheresis

Apheresis therapy is the converse of transfusion therapy; it entails treating disease by removing plasma, specific antibodies, or cells. It has been tried in a broad spectrum of diseases [*see Table 7*]. Therapeutic apheresis has real risks and may provide little benefit. It is usually an acute intervention that is only transiently effective, unless the underlying problem is being treated effectively. Consequently, it is important to identify criteria for both starting and stopping such treatment. Indications for apheresis that are approved by the American Association of Blood Banks and the American Society for Apheresis have been summarized.[102]

INDICATIONS FOR APHERESIS THERAPY

Neurologic Diseases

Neurologic diseases whose pathogenesis may be antibody mediated are now the most common indications for plasma exchange. Myasthenia gravis occurs when antibodies to acetylcholine receptors cause abnormal neuromuscular transmission. Reductions in these antibody titers from plasma exchange are associated with clinical improvement. A randomized trial compared the use of plasma exchange with the use of IVIg therapy in the treatment of myasthenia gravis; the investigators noted a trend toward better results with plasma exchange, but this trend was not statistically significant.[103] Similar findings were reported from much larger studies of Guillain-Barré syndrome, which is thought to be caused by antibodies to myelin. Two large series comparing plasma exchange with current best therapy showed faster improvement with the addition of plasma exchange. Randomized comparisons of plasma exchange and IVIg in the treatment of Guillain-Barré syndrome have shown these approaches to be equivalent; no additional benefit from using both therapies was shown.[104,105]

Chronic inflammatory demyelinating polyneuropathy (CIDP) is an autoimmune disorder that causes proximal and distal weakness; it has a progressive or relapsing course and is sometimes associated with monoclonal gammopathies. CIDP responds to plasma exchange, except in patients with distal weakness and associated IgM monoclonal gammopathies[106,107]; such patients respond poorly to all modalities of therapy.[108] IVIg therapy and plasma exchange have been shown to be comparably effective in CIDP.[109]

The use of plasma exchange in multiple sclerosis remains controversial. Meta-analysis of six controlled trials of plasma exchange provided some evidence of benefit, but the authors concluded that the subgroups of patients likely to benefit needed further definition.[110] A randomized study of plasma exchange in patients with acute inflammatory demyelinating central nervous system disease showed a significant benefit from the therapy. However, patients continued to have problems with relapse.[111]

Hematologic Diseases

Leukapheresis The hematologic diseases that require apheresis are those associated with obstruction of vascular flow by cells or the blockage of flow by proteins as a result of increased viscosity or cryoprecipitation; antibody-mediated diseases that lead to destruction of the formed elements of the blood; and thrombotic microangiopathies.

Leukostasis is a function of cell number and cell type. Myeloblasts are more likely to cause stasis than are an equivalent number of lymphocytes in a patient with chronic lymphocytic leukemia. Unless pulmonary or cerebral leukostasis is severe enough to cause progression in clinical findings, hydroxyurea is the treatment of choice. It will usually decrease the cell count sufficiently within 24 hours. However, when clinical findings demand improvement within 4 to 8 hours, leukapheresis in addition to hydroxyurea is usually needed.

Red cell exchange Red cell exchange has been used to treat acute chest crises, stroke, and priapism, and it is sometimes used to prepare patients with sickle cell disease for surgery. In these patients, the indications for red cell exchange, versus simple transfusion, are poorly defined. For example, in an analysis of causes and outcomes in acute chest syndrome, simple transfusions were used instead of red cell exchange in about two thirds of patients.[112] Many believe that red cell exchanges should be reserved for patients with progressive pulmonary disease or for those who fail to respond to transfusions. Using Rh and Kell antigen–compatible red cells reduces the incidence of alloimmunization from 7% to 1%[113] and should be standard practice. Red cell exchange leads to less iron accumulation than transfusion therapy, which is an advantage in the treatment of patients with sickle cell disease who require long-term therapy, such as those who have a history of stroke.[114]

Plasma exchange The concentration of paraprotein influences plasma protein viscosity, as does its heavy-chain class. IgM is the largest plasma protein and is nearly 100% intravascular; it is most likely to cause hyperviscosity. IgA and IgG3 are more likely to aggregate and are associated with hyperviscosity more often than other IgG subclasses. As in leukostasis, the choice between plasma exchange and chemotherapy is guided mainly by the clinical symptoms and their rate of progression.

Table 7 Indications for Plasma Exchange

Indications Based on Randomized Trials	*Indications Based on Consensus and on Case Reports*	*Possible Indications*
	Myasthenia gravis	
	Hyperviscosity	Pemphigus vulgaris
Guillain-Barré syndrome	Hemolytic-uremic syndrome	Goodpasture syndrome
Chronic inflammatory polyneuropathy	Persistent HELLP syndrome	Autoimmune hemolytic anemia
Peripheral neuropathy associated with MGUS	Posttransfusion purpura	Antibody to coagulation factors
Thrombotic thrombocytopenic purpura	Cryoglobulinemia	Idiopathic thrombocytopenic purpura
	Vasculitis	Cold agglutinin disease
	Familial hypercholesterolemia	

HELLP—hemolysis, elevated liver enzymes, and low platelet count MGUS—monoclonal gammopathy of unknown significance

Plasma exchange can lower viscosity within hours, whereas most chemotherapy requires days. Acute-onset renal failure caused by myeloma proteins can be improved by lowering the plasma concentration of paraprotein, but more data are needed for this to be considered an established indication.[115]

Despite the role that antibody and immune complexes play in hematologic cytopenia, there are no well-controlled studies supporting the use of plasma exchange. The available case reports usually describe the role of plasma exchange as being that of backup after failure of more established therapies. FFP or cryopoor plasma is used in replacement therapy for thrombotic microangiopathies. Case reports suggest that patients with severe preeclampsia, HELLP syndrome, or both may benefit from plasma exchange with FFP replacement if they fail to improve after delivery.[116]

Antibody-Mediated Renal, Muscular, and Cutaneous Diseases

Despite promising reports from case studies, controlled trials of patients with pemphigus vulgaris,[117] polymyositis, dermatomyositis,[118] and Goodpasture syndrome[119] have raised doubts concerning the value of plasma exchange. However, plasma exchange does appear to be valuable in stopping pulmonary hemorrhage in Goodpasture syndrome.

Immune Complex Diseases

The only indication for plasma exchange in rheumatoid arthritis and systemic lupus erythematosus is severe vasculitis that does not respond to other therapies.

Metabolic Diseases

Plasma exchange and selective removal of low-density lipoproteins (LDLs) have both been used to treat familial hypercholesterolemia [*see 9:II Diagnosis and Treatment of Dyslipidemia*]. Selective removal of LDLs can be accomplished by immunoadsorption, heparin precipitation, or dextran sulfate cellulose absorption, whereas plasma exchange causes significant reduction of both LDLs and HDLs.

COMPLICATIONS OF PLASMA EXCHANGE

The complications associated with plasma exchange are best divided into problems related to apheresis machines and problems related to venous access, type of replacement fluids, and anticoagulant. Apheresis machines accomplish cell and plasma separation by either centrifugation or membrane filtration. All systems monitor air and access pressure, allowing air emboli to be eliminated and access problems to be promptly recognized. Excess transmembrane pressure may cause red cell hemolysis, which leads to increased hemoglobin in the separated plasma. The majority of complications associated with plasma exchange result from the replacement fluid and anticoagulant used. Plasma removed by exchange is commonly replaced with 5% albumin, which carries no risk of infection and does not increase the citrate return but does dilute coagulation factors, causing mild coagulopathy for 24 to 48 hours. On an every-other-day treatment schedule, coagulation abnormalities are usually not clinically significant, but they may become significant if the patient is on a daily treatment schedule. Using FFP prevents dilutional coagulopathy but increases risks of blood-borne infection and allergic reactions. Peripheral venous access is often inadequate to maintain the required flow rates of 45 to 80 ml/min, necessitating central venous access with a large, double-lumen catheter; life-threatening or fatal complications from central catheter placement have been reported.[120] Catheter malfunction should always be considered when a patient shows clinical evidence of hypovolemia, shock, or both while undergoing plasma exchange. The majority of complications, however, are side effects of the citrate anticoagulant. These can include paresthesias, abdominal cramps, and, in rare instances, cardiac arrhythmias or seizures. Citrate toxicity is usually managed easily by slowing the return rate and providing extra calcium, either orally or sometimes intravenously. Patients with renal failure who receive large amounts of citrate may develop a profound metabolic alkalosis.[121]

Future Prospects for Transfusion Therapy

The evolution of transfusion practice has been a steady progression from whole blood to fractionated products designed for specific therapies. The search for a practical replacement for red cells that would allow stable storage, provide adequate oxygen delivery, and be free of significant toxins has been long and filled with substantial obstacles. Hemoglobin-based substitutes are most promising, but they still have problems with purification, adequate oxygen unloading, and potential toxicity. Chemically modified bovine hemoglobin is currently in phase-3 clinical trials and is probably closest to licensure. In the case of coagulation components, recombinant products are beginning to provide highly specific treatment of clinical problems that are poorly managed by current therapy. Bioengineering holds the promise of improving the effectiveness of current recombinant products.[62] Recombinant factor VIIa provides specific therapy for deficiency states and an alternative approach for patients with high-titer factor VIII inhibitors. Recombinant antithrombin III provides specific replacement for patients with congenital deficiency. The off-label use of these products is increasing and brings with it the need for careful studies of cost-effectiveness, which are only beginning to emerge.

A major change in transfusion practice may evolve from the availability of cytokines that can modify endogenous production. Erythropoietin has changed the treatment of anemia associated with chronic renal disease; as a result, many dialysis patients no longer require transfusions. Erythropoietin can also facilitate patients' self-banking their blood for anticipated surgical needs. In some cases, erythropoietin use is accepted by Jehovah's Witness patients, thereby allowing such patients to undergo surgical procedures that would otherwise not be possible. The availability of myeloid growth factors has contributed substantially to the development of methods for collection, and it is possible that mobilizing leukocytes with growth factors can increase the effectiveness of granulocyte transfusions. Thrombopoietin may in time be used to enhance platelet apheresis collections. The immunomodulatory effects of blood transfusion are in the early stages of description. It may be that a better understanding of the mechanisms underlying immunomodulation will permit using these effects to therapeutic advantage, such as to induce tolerance to a transplanted organ or to downregulate antibody production.

Even in an era of accelerating change, certain aspects of transfusion medicine will remain constant. The blood donor remains a kingpin who cannot be replaced by recombinant methodology. Transfusion practice has improved in safety, but there will always be residual risks. Each transfusion will always require careful assessment of whether the risks to the recipient of the transfusion will exceed the risks of going without. Further decre-

ment of these risks will require new systemic approaches for improving the collection of specimens, selecting patients for transfusion of specific products, and continuing education of clinicians who make transfusion decisions.[122]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Williams AE, Thomson RA, Schreiber GB, et al: Estimates of infectious disease risk factors in US blood donors. JAMA 277:967, 1997

2. Thomas MJG, Gillon J, Desmond MJ: Consensus Conference on Autologous Transfusion: preoperative autologous donation. Transfusion 36:633, 1996

3. Starkey JM, MacPherson JL, Bolgiano DC, et al: Markers for transfusion-transmitted disease in different groups of blood donors. JAMA 262:3452, 1989

4. Strauss RG, Burmeister LF, Johnson K, et al: Randomized trial assessing the feasibility and safety of biologic parents as RBC donors for their preterm infants. Transfusion 40:450, 2000

5. Dodd RY, Notari IV, Stramer SL: Current prevalence and incidence of infectious disease markers and estimated window-period risk in the American Red Cross blood donor population. Transfusion 42:975, 2002

6. Stramer SL: US NAT yield, where are we after 2 years? Transfus Med 12:243, 2002

7. Hening H, Puchta I, Lum J, et al: Frequency and load of hepatitis B virus DNA in first-time blood donors with antibodies to hepatitis B core antigen. Blood 100:2637, 2002

8. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. MMWR Morb Mortal Wkly Rep 47(RR-19):1, 1998

9. Akahane Y, Kojuma M, Sugai Y, et al: Hepatitis C virus infection in spouses of patients with type C chronic liver disease. Ann Intern Med 120:748, 1994

10. Fried MW: Diagnostic testing for hepatitis C: practical considerations. Am J Med 107:31S, 1999

11. Epstein JS: Clarification of the use of unlicensed anti-HCV supplemental test results in regard to donor notification. FDA memorandum, August 19, 1993. U.S. Food and Drug Administration, Rockville, Maryland
http://www.fda.gov/cber/bldmem/081993.pdf

12. Lemaire JM, Courouce AM, Defer C, et al: HCV RNA in blood donors with isolated reactivities by third-generation PCR. Transfusion 40:867, 2000

13. Bower WA, Nainan OV, Han X, et al: Duration of viremia in hepatitis A virus infection. J Infect Dis 182:12, 2000

14. Lefrére JJ, Sender A, Mercier B, et al: High rate of GB virus type C/HGV transmission from mother to infant: possible implications for the prevalence of infection in blood donors. Transfusion 40:602, 2000

15. Sentjens R, Basaras M, Simmonds P, et al: HGV/GB virus C transmission by blood components in patients undergoing open-heart surgery. Transfusion 43:1558, 2003

16. Lackritz EM, Satten GA, Aberle-Grasse J, et al: Estimated risk of transmission of the human immunodeficiency virus by screened blood in the United States. N Engl J Med 333:1712, 1995

17. Jackson JB, Hanson MR, Johnson GM, et al: Long-term follow-up of blood donors with indeterminate human immunodeficiency virus type 1 results on Western blot. Transfusion 35:98, 1995

18. Busch MP, Kleinman SG, Williams AE, et al: Frequency of human immunodeficiency virus (HIV) infection among contemporary anti-HIV-1 and anti-HIV-1/2 supplemental test–indeterminate blood donors. Transfusion 36:37, 1996

19. Sayre KR, Dodd RY, Tegtmeier G, et al: False-positive human immunodeficiency virus type 1 Western blot tests in noninfected blood donors. Transfusion 36:45, 1996

20. Murphy EL, Glynn SA, Fridey J, et al: Increased prevalence of infectious diseases and other adverse outcomes in human T lymphotropic virus types I- and II-infected donors. J Infect Dis 176:1468, 1997

21. Marsh BJ: Infectious complications of human T cell leukemia/lymphoma virus type I infections. Clin Infect Dis 23:138, 1996

22. HTLV-associated myelopathy in a cohort of HTLV-I and HTLV-II infected blood donors. The REDS investigators. Neurology 48:315, 1997

23. Busch MP, Switzer WM, Murphy EL, et al: Absence of evidence of infection with divergent primate T-lymphotropic viruses in United States blood donors who have seroindeterminate HTLV test results. Transfusion 40:443, 2000

24. False-positive serologic tests for human T-cell lymphotropic virus type I among blood donors following influenza vaccination, 1992. MMWR Morb Mortal Wkly Rep 42:173, 1993

25. Dodd RY: Emerging infections, transfusion safety, and epidemiology. N Engl J Med 349:1205, 2003

26. Pealer LN, Marfin AA, Peterson LR, et al: Transmission of West Nile Virus through blood transfusion in the United States in 2002. N Engl J Med 349:1236, 2003

27. Llewelyn CA, Hewitt PE, Knight RS, et al: Possible transmission of variant Creutzfeldt-Jakob disease by blood transfusion. Lancet 363:417, 2004

28. Cooper JD, Bird SM: Predicting the incidence of variant Creutzfeldt-Jakob disease from UK dietary exposure to bovine spongiform encephalopathy for the 1940–1969 and post-1969 birth cohorts. Int J Epidemiol 32:784, 2003

29. Yamamoto FI, Clausen H, White T, et al: Molecular genetic basis of the histo-blood group ABO system. Nature 345:229, 1990

30. Reid ME: Applications of DNA-based assays in blood group antigen and antibody identification. Transfusion 43:1748, 2003

31. Avent ND, Reid ME: The Rh blood group system: a review. Blood 95:375, 2000

32. Reid ME, Rios M, Powell VI, et al: DNA from blood samples can be used to genotype patients who have recently received a transfusion. Transfusion 40:48, 2000

33. Marrosszeky S, McDonald J, Sutherland H, et al: Suitability of preadmission blood samples for pretransfusion testing in elective surgery. Transfusion 37:910, 1997

34. Seftel MD, Growe GH, Petraszko T: Universal prestorage leukoreduction in Canada decreases platelet alloimmunization and refractoriness. Blood 103:333, 2004

35. Chambers LA, Herman JH: Considerations in the selection of a platelet component: apheresis versus whole blood derived. Transfus Med Rev 13:331, 1999

36. Heal JM, Rowe JM, McMican A, et al: The role of ABO matching in platelet transfusion. Eur J Haematol 50:110, 1993

37. Practice guidelines for blood component therapy: a report by the American Society of Anesthesiologists Task Force on Blood Component Therapy. Anesthesiology 84:732, 1996

38. Bracey AW, Radovancevic R, Riggs SA, et al: Lowering the hemoglobin threshold for transfusion in coronary artery bypass procedures: effect on patient outcome. Transfusion 39:1070, 1999

39. A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. Transfusion Requirements in Critical Care Investigators, Canadian Critical Care Trials Group. N Engl J Med 340:409, 1999

40. Cazzola M, Mercuriali F, Brugnara C: Use of recombinant human erythropoietin outside the setting of uremia. Blood 89:4248, 1997

41. Samol J, Littlewodd TJ: Efficacy of rHuEPO in cancer-related anemia. Br J Haemol 121:3, 2002

42. Wu W, Rathore MPH, Wang Y, et al: Blood transfusion in elderly patients with acute myocardial infarction. N Engl J Med 17:1230, 2001

43. Churchill WH, McGurk S, Chapman RH, et al: The Collaborative Hospital Transfusion Study: variations in the use of autologous blood account for hospital differences in red cell use during primary hip and knee surgery. Transfusion 38:530, 1998

44. Monk TG, Goodnough LT, Birkmeyer JD, et al: Acute normovolemic hemodilution is a cost-effective alternative to preoperative autologous blood donation by patients undergoing radical retropubic prostatectomy. Transfusion 35:559, 1995

45. Guidelines for the use of platelet transfusions. British Committee for Standards in Hematology, Blood Transfusion Task Force. Br J Haematol 122:10, 2003

46. Wandt H, Frank M, Ehniger G, et al: Safety and cost effectiveness of a 10×10^9/L trigger for prophylactic platelet transfusions compared with the traditional 20×10^9/L trigger: a prospective comparative trial in 105 patients with acute myeloid leukemia. Blood 91:3601, 1998

47. Shemin D, Elnour M, Amarantes B, et al: Oral estrogens decrease bleeding time and improve clinical bleeding in patients with renal failure. Am J Med 89:436, 1990

48. Reiter RA, Mayr F, Blazicek H, et al: Desmopressin antagonizes the in vitro platelet dysfunction induced by GPIIb/IIIa inhibitors and aspirin. Blood 102:4594, 2003

49. Kickler T, Kennedy SD, Braine HG: Alloimmunization to platelet-specific antigens on glycoproteins IIb-IIIa and Ib/IX in multiply transfused thrombocytopenic patients. Transfusion 30:622, 1990

50. Delaflor-Weiss E, Mintz PD: The evaluation and management of platelet refractoriness and alloimmunization. Transf Med Rev 14:180, 2000

51. Moroff G, Garratty G, Heal JM, et al: Selection of platelets for refractory patients by HLA matching and prospective crossmatching. Transfusion 32:633, 1992

52. Friedberg RC, Donnelly SF, Mintz PD: Independent roles for platelet crossmatching and HLA in the selection of platelets for alloimmunized patients. Transfusion 34:215, 1994

53. Leukocyte reduction and ultraviolet irradiation of platelets to prevent alloimmunization and refractoriness to platelet transfusions. The Trial to Reduce Alloimmunization to Platelets Study Group. N Engl J Med 337:1861, 1997

54. Novotny VM, van Doorn R, Witvliet MD, et al: Occurrence of allogeneic HLA and non-HLA antibodies after transfusion of prestorage filtered platelets and red blood cells: a prospective study. Blood 85:1736, 1995

55. Sullivan MT, McCullough J, Schreiber GB, et al: Blood collection and transfusion in the United States in 1997. Transfusion 42:1253, 2002

56. Practice parameters for the use of fresh-frozen plasma, cryoprecipitate, and platelets. Administration Practice Guidelines Development Task Force of the College of American Pathologists. JAMA 271:777, 1994

57. McVay PA, Toy PT: Lack of increased bleeding after paracentesis and thoracentesis in patients with mild coagulation abnormalities. Transfusion 31:164, 1991

58. Mannucci PM, Bauer KA, Santagostino E, et al: Activation of the coagulation cascade after infusion of a factor XI concentrate in congenitally deficient patients. Blood 84:1314, 1994

59. Lechner K, Kyrle PA: Antithrombin III concentrates—are they clinically useful? Thromb Haemost 73:340, 1995

60. George JN: How I treat patients with thrombotic thrombocytopenic purpura–hemolytic uremia syndrome. Blood 96:1223, 2000

61. Rock G, Sutton DM, Nair RC: Cryosupernatant as replacement fluid for plasma exchange in thrombotic thrombocytopenic purpura. Members of the Canadian Apheresis Group. Br J Haematol 94:383, 1996

62. O'Connell NM, Perry DJ, Hodgson AJ, et al: Recombinant FVIIA in the management of uncontrolled hemorrhage. Transfusion 43:1711, 2003

63. Saenko EL, Ananyeva M, Shima M, et al: The future of recombinant coagulation

factors. J Thromb Haemost 1:922, 2002

64. Brettler DB, Forsberg AD, Levine PH, et al: The use of porcine factor VIII concentrate (Hyate:C) in the treatment of patients with inhibitor antibodies to factor VIII: a multicenter US experience. Arch Intern Med 149:1381, 1989

65. Thompson AR: Factor IX concentrates for clinical use. Semin Thromb Hemost 19:25, 1993

66. Price TH: The current prospects for neutrophil transfusions for the treatment of granulocytopenic infected patients. Transfus Med Rev 14:2, 2000

67. Brugnara C, Churchill WH: Plasma component therapy. Thrombosis and Hemorrhage, 2nd ed. Loscalzo J, Schafer AI, Eds. Williams & Wilkins, Baltimore, 1998, p 1135

68. Ahsan N, Weigand LA, Abendroth CS, et al: Acute renal failure following immunoglobulin therapy. Am J Nephrol 16:532, 1996

69. Renal insufficiency and failure associated with immune globulin intravenous therapy —United States, 1985–1998. MMWR Morb Mortal Wkly Rep 48:518, 1999

70. Sekul EA, Cupler EJ, Dalakas MC: Aseptic meningitis associated with high-dose intravenous immunoglobulin therapy: frequency and risk factors. Ann Intern Med 121:259, 1994

71. Garratty G: Severe reactions associated with transfusion of patients with sickle cell disease. Transfusion 37:357, 1997

72. Davenport RD, Kunkel SL: Cytokine roles in hemolytic and nonhemolytic transfusion reactions. Transfus Med Rev 8:157, 1994

73. Pineda AA, Vamvakas EC, Gorden LD, et al: Trends in the incidence of delayed hemolytic and delayed serologic transfusion reactions. Transfusion 39:1097, 1999

74. Lee J-H: Transfusion fatalities reported to the FDA (1990–1998). Bacterial Contamination of Platelets Workshop. Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, Rockville, Maryland, September 24, 1999
http://www.fda.gov/cber/minutes/bact092499.pdf

75. Heddle NM, Klama L, Singer J: The role of plasma from platelet concentrates in transfusion reactions. N Engl J Med 331:625, 1994

76. Federowicz I, Barrett BB, Andersen JW, et al: Characterization of reactions after transfusion of cellular blood components that are white cell reduced before storage. Transfusion 36:21, 1996

77. Kelley DL, Mangini J, Lopez-Plaza I, et al: The utility of ≤3-day-old whole-blood platelets in reducing the incidence of febrile nonhemolytic transfusion reactions. Transfusion 40:439, 2000

78. Dzik S, Aubuchon J, Jeffries L, et al: Leukocyte reduction of blood components: public policy and new technology. Transfus Med Rev 14:34, 2000

79. Popovsky MA, Chaplin HC Jr, Moore SB: Transfusion-related acute lung injury: a neglected, serious complication of hemotherapy. Transfusion 32:589, 1992

80. Kopko PM, Paglieroni TG, Popovsky MA, et al: TRALI: correlation of antigen-antibody and monocyte activation in donor-recipient pairs. Transfusion 43:177, 2003

81. Kao GS, Wood IG, Dorfman DM, et al: Investigations into the role of anti-HLA class II antibodies in TRALI. Transfusion 43:185, 2003

82. Silliman CC, Paterson AJ, Dickey WO, et al: The association of biologically active lipids with the development of transfusion-related acute lung injury: a retrospective study. Transfusion 37:719, 1997

83. Silliman CC, Boshkov LK, Mehdizadehkashi Z, et al: Transfusion-related acute lung injury: epidemiology and a prospective analysis of etiologic factors. Blood 101:454, 2003

84. Silliman CC, Voelkel NF, Allard JD, et al: Plasma and lipids from stored packed red blood cells cause acute lung injury in an animal model. J Clin Invest 101:1458, 1998

85. Sandler SG, Mallory D, Malamut D, et al: Hemagglutination assays for the diagnosis and prevention of IgA anaphylactic transfusion reactions. Transfus Med Rev 9:1, 1995

86. Klüter H, Bubel S, Kirchner H, et al: Febrile and allergic transfusion reactions after the transfusion of white cell–poor platelet preparations. Transfusion 39:1179, 1999

87. Cyr M, Hume HA, Champagne M, et al: Anomaly of the des-Arg9-bradykinin metabolism associated with severe hypotensive reactions during blood transfusions: a preliminary study. Transfusion 39:1084, 1999

88. Owen HG, Brecher ME: Atypical reactions associated with use of angiotensin-converting enzyme inhibitors and apheresis. Transfusion 34:891, 1994

89. Ohto H, Anderson KC: Survey of transfusion-associated graft-versus-host disease in immunocompetent recipients. Transfus Med Rev 10:31, 1996

90. Ohto H, Anderson KC: Posttransfusion graft-versus-host disease in Japanese newborns. Transfusion 36:117, 1996

91. Ammann AJ: Hypothesis: absence of graft-versus-host disease in AIDS is a consequence of HIV-1 infection of CD4+ T cells. J Acquir Immune Defic Syndr 6:1224, 1993

92. Ohto H, Yasuda H, Noguchi M: Risk of transfusion-associated graft-versus-host disease as a result of directed donation from relatives (letter). Transfusion 32:691, 1992

93. Morof FG, Luban NLC: The irradiation of blood and blood components to prevent graft-versus-host disease: technical issues and guidelines. Transfus Med Rev 11:15, 1997

94. Nollet KE, Holland PV: Toward a coalition against transfusion-associated GVHD (editorial). Transfusion 43:1655, 2003

95. McDonald CP, Hartley S, Orchard K, et al: Fatal Clostridium perfringens sepsis from a pooled platelet transfusion. Transfusion 37:259, 1997

96. Red blood cell transfusions contaminated with Yersinia enterocolitica: United States, 1991–1996, and initiation of a national study to detect bacteria-associated transfusion reaction. MMWR Morb Mortal Wkly Rep 46:553, 1997

97. Bowden RA, Slichter SJ, Sayers M, et al: A comparison of filtered leukocyte-reduced and cytomegalovirus (CMV) seronegative blood products for the prevention of transfusion-associated CMV infection after marrow transplant. Blood 86:3598, 1995

98. Nichols WG, Price TH, Gooley T, et al: Transfusion-transmission cytomegalovirus infection after receipt of leukoreduced blood products. Blood 101:4195, 2003

99. Opelz G, Vanrenterghem Y, Kirste G, et al: Prospective evaluation of pretransplant blood transfusions in cadaver kidney recipients. Transplantation 63:964, 1997

100. Bordin JO, Bardossy L, Blajchman MA: Growth enhancement of established tumors by allogeneic blood transfusion in experimental animals and its amelioration by leukodepletion: the importance of the timing of the leukodepletion. Blood 84:344, 1998

101. Vamvakas EC, Blajchman MA: Deleterious clinical effects of transfusion-associated immunomodulation: fact or fiction? Blood 97:1180, 2001

102. Smith JW, Weinstein R, Hillyer C: Therapeutic apheresis: a summary of current indication categories endorsed by the AABB and the American Society for Apheresis. Transfusion 43:820, 2003

103. Gajdos P, Chevret S, Clair B, et al: Clinical trial of plasma exchange and high-dose intravenous immunoglobulin in myasthenia gravis. Myasthenia Gravis Clinical Study Group. Ann Neurol 41:789, 1997

104. Randomised trial of plasma exchange, intravenous immunoglobulin, and combined treatments in Guillain-Barré syndrome. Plasma Exchange/Sandoglobulin Guillain-Barré Syndrome Trial Group. Lancet 349:225, 1997

105. van der Meché FG, Schmitz PI: A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barré syndrome. The Dutch Guillain-Barré Study Group. N Engl J Med 326:1123, 1992

106. Dyck PJ, Low PA, Windebank AJ, et al: Plasma exchange in polyneuropathy associated with monoclonal gammopathy of undetermined significance. N Engl J Med 325:1482, 1991

107. Hahn AF, Bolton CF, Pillay N, et al: Plasma-exchange therapy in chronic inflammatory demyelinating polyneuropathy: a double-blind, sham-controlled, cross-over study. Brain 119:1055, 1996

108. Katz JS, Saperstein DS, Gronseth G, et al: Distal acquired demyelinating symmetric neuropathy. Neurology 54:615, 2000

109. Dyck PJ, Litchy WJ, Kratz KM, et al: A plasma exchange versus immune globulin infusion in chronic inflammatory demyelinating polyradiculoneuropathy. Ann Neurol 36:838, 1994

110. Vamvakas EC, Pineda AA, Weinshenker BG: Meta-analysis of clinical studies of the efficacy of plasma exchange in the treatment of chronic progressive multiple sclerosis. J Clin Apheresis 10:163, 1995

111. Weinshenker BG, O'Brien PC, Petterson TM, et al: A randomized trial of plasma exchange in acute central nervous system inflammatory demyelinating disease. Ann Neurol 46:878, 1999

112. Vichinsky EP, Neumayr LD, Earles AN, et al: Causes and outcomes of the acute chest syndrome in sickle cell disease. N Engl J Med 342:1855, 2000

113. Vichinsky EP, Earles AN, Johnson RA, et al: Alloimmunization in sickle cell anemia and transfusion of racially unmatched blood. N Engl J Med 322:1617, 1990

114. Hilliard LM, Williams BF, Lounsbury AE, et al: Erythrocytapheresis limits iron accumulation in chronically transfused sickle cell patients. Am J Hematol 59:28, 1998

115. Zucchelli P, Pasquali S, Cagnoli L, et al: Controlled plasma exchange trial in acute renal failure due to multiple myeloma. Kidney Int 33:1175, 1988

116. Martin JN, Files FC, Blake PG: Plasma exchange for preeclampsia: I. Postpartum use for persistently severe preeclampsia-eclampsia with HELLP syndrome. Am J Obstet Gynecol 162:126, 1990

117. Guillaume JC, Roujeau JC, Morel P, et al: Controlled study of plasma exchange in pemphigus. Arch Dermatol 124:1659, 1988

118. Miller FW, Leitman SF, Cronin ME, et al: Controlled trial of plasma exchange and leukapheresis in polymyositis and dermatomyositis. N Engl J Med 326:1380, 1992

119. Johnson JP, Moore J Jr, Austin HA III, et al: Therapy of anti-glomerular basement membrane antibody disease: analysis of prognostic significance of clinical, pathologic and treatment factors. Medicine (Baltimore) 64:219, 1985

120. Rizvi MA, Vesely JN, Chandler GL, et al: Complications of plasma exchange in 71 consecutive patients treated for clinically suspected thrombotic thrombocytopenic purpura–hemolytic-uremic syndrome. Transfusion 40:869, 2000

121. Pearl RG, Rosenthal MM: Metabolic alkalosis due to plasmapheresis. Am J Med 79:391, 1985

122. Dzik WH: Emily Cooley Lecture 2002: Transfusion safety in the hospital. Transfusion 43:1190, 2003

Acknowledgment

The author would like to thank Yoriko Saito, M.D., and Siobhan McGurk for their many suggestions and contributions to this chapter.

XI HEMATOPOIETIC CELL TRANSPLANTATION

Frederick R. Appelbaum, m.d.

Because hematopoietic stem cells can be collected from the peripheral blood, bone marrow, and umbilical cord blood, the term bone marrow transplantation is being replaced by the more inclusive term hematopoietic cell transplantation. With transplantation, an abnormal but nonmalignant lymphohematopoietic system can be replaced with a healthy one, making transplantation an effective therapy for a variety of nonmalignant diseases (e.g., severe combined immunodeficiency disease, Wiskott-Aldrich syndrome, aplastic anemia, thalassemia, sickle cell anemia, and Gaucher disease). In addition, hematopoietic cell transplantation is used to treat a variety of malignancies because it allows administration of higher and potentially more effective doses of chemotherapy and radiotherapy that would otherwise cause unacceptable myelosuppression. Allogeneic transplantation also confers its own antitumor effects beyond those of chemoradiotherapy. Worldwide, an estimated 48,000 patients underwent hematopoietic cell transplantation in 2002.[1]

The Hematopoietic Stem Cell

Three features of the lymphohematopoietic system make transplantation feasible: its regeneration capacity, the homing of stem cells to sites that promote survival and proliferation, and the ability of stem cells to survive, with little damage, the freezing and thawing process entailed in cryopreservation. In mice, a single hematopoietic stem cell can reconstitute a lethally irradiated recipient.[2] In humans, transplantation of considerably less than 10% of a donor's marrow regularly results in complete and sustained replacement of the recipient's entire lymphohematopoietic system, including red cells, platelets, granulocytes, T cells, and B cells, as well as pulmonary alveolar macrophages, Kupffer cells of the liver, osteoblasts, Langerhans cells of the skin, and microglial cells of the brain. Studies suggest that transplanted hematopoietic cells rarely differentiate into hepatic, epithelial, or mucosal cells.[3,4]

The mechanisms of homing are not entirely understood, but a remarkably high percentage of transplanted primitive hematopoietic cells are retained in the marrow shortly after intravenous injection. Cell adhesion molecules on marrow endothelial cells, termed selectins, which bind to carbohydrate-based ligands on early hematopoietic cells, may be responsible.

The human hematopoietic stem cell expresses distinctive surface antigens. One of these, the CD34 antigen, is expressed on only 1% to 5% of normal adult bone marrow cells, but when marrow is cultured in vitro, virtually all colonies derive from the CD34+ population. Successful transplantation in humans can be carried out using only positively selected CD34+ cells. Over 90% of CD34+ cells also express CD38, but the 10% that are CD34+ and CD38– are the population best able to support long-term hematopoiesis in vitro and are thus considered a more primitive population. The most primitive subset of these cells stain poorly with Rh123, a mitochondrial dye. These cells also lack known markers of B cell or T cell lineage and are therefore said to be lineage negative.

Types of Hematopoietic Stem Cell Transplantation

Hematopoietic cell transplantation can be categorized according to the relation between the donor and the recipient and according to the anatomic source of the stem cell. Hematopoietic stem cells for transplantation may derive from bone marrow, peripheral blood, or umbilical cord blood and may be harvested from a syngeneic, allogeneic, or autologous donor.

SYNGENEIC TRANSPLANTATION

Identical twins are the best possible donors of stem cells. When syngeneic donors are used, neither graft rejection nor graft versus host disease (GVHD) will develop in the recipient. Syngenicity is easily established by DNA typing, using one of two techniques—either restriction fragment length polymorphisms (RFLPs) or variable nucleotide tandem repeats. Only about one in 100 patients undergoing transplantation will have an identical twin.

ALLOGENEIC TRANSPLANTATION

Allogeneic transplantation, which involves a related or unrelated donor, is more complicated than syngeneic or autologous transplantation (see below) because of immunologic differences between donor and host. With allogeneic hematopoietic cell transplantation, in which the immune system of the patient is provided by the graft, the clinical concerns are not only with the prevention of graft rejection by host cells surviving the pretransplant preparative regimen but also with the prevention of donor cells from causing immune-mediated injury to the patient (i.e., GVHD).

Immunologic reactivity between donor and host is largely mediated by immunocompetent cells that react with human leukocyte antigens (HLAs), which are encoded by genes of the major histocompatibility complex. HLA molecules display both exogenous peptides (e.g., from an infecting virus) and endogenous peptides, presenting them to T cells, an important step in the initiation of an immune response. If two persons do not share the same HLA antigens, T cells taken from one person will react vigorously to the mismatched HLA molecules on the surface of the cells from the other. These are reactions against so-called major HLA determinants. Even when two persons who are not identical twins have identical HLA types, the endogenous peptides presented by the HLA antigens will differ, triggering a response against so-called minor HLA determinants.

The genes encoding HLA class I and class II antigens are tightly linked and tend to be inherited together as haplotypes with low recombination frequencies. For any given patient, there is a 25% probability that any one sibling has inherited the same paternal and maternal haplotype, making the siblings identical with regard to HLA genotype. Given that the average number of children per family in the United States is slightly more than two, the average chance that a patient has an HLA-matched sibling is approximately 35%. The formula for calculating the probability that a patient has an HLA-identical sibling is $1 – (0.75)^n$, where n equals the number of siblings.

Allogeneic transplantation has been performed using HLA-identical sibling donors, other matched and mismatched family-member donors, and matched unrelated donors. The best re-

sults have been achieved with HLA-identical sibling donors. With transplantation from family-member donors who are identical for one haplotype but mismatched for a single locus (i.e., HLA-A, HLA-B, or HLA-D) on the other haplotype, the survival rate is nearly equal to that with HLA-identical donors, although there is a higher incidence of GVHD.[5] Transplants using family-member donors mismatched for two or more loci have worse results—a higher incidence of GVHD and graft rejection and a lower probability of survival.[5]

Because of the highly polymorphic nature of HLA antigens, the probability that two unrelated persons will match is extremely low. Matched unrelated donor transplantation was first performed in the late 1970s. The broader application of this technique was made feasible by the creation of large donor registries in the late 1980s. Since then, the number of unrelated-donor transplantations has rapidly increased [*see Figure 1*]. Currently, more than seven million healthy persons have volunteered to serve as marrow donors in the United States alone, making the odds of finding an unrelated donor matched for HLA-A, HLA-B, and HLA-D approximately 50%. On average, it takes about 3 months from the time a search is initiated to identify a donor and begin transplantation. In 2002, approximately 2,100 unrelated-donor transplantations were performed in the United States.[6]

Results of early transplantations suggested that GVHD was more common and graft rejection more frequent with unrelated donors than with related donors matched for HLA compatibility.[7] The difference in incidence of GVHD may be partially explained by the increased disparity in minor HLA determinants in unrelated individuals. Undetected disparities in HLA types between supposedly matched unrelated donor-recipient pairs may also have contributed to the increased incidence of GVHD. Before 1998, HLA typing for HLA-A, HLA-B, and HLA-C was conducted using serologic methods. Since then, studies using automated direct sequencing of these genes have demonstrated allele-level mismatching in 30% of pairs previously defined as HLA identical.[8,9] Such mismatching in class I antigens has been shown to be associated with increased GVHD, whereas mismatches at class II have been shown to be associated with an increased risk of graft rejection. Currently, outcomes of transplantations using fully matched donors approach those seen with matched sibling transplants in many disease categories.

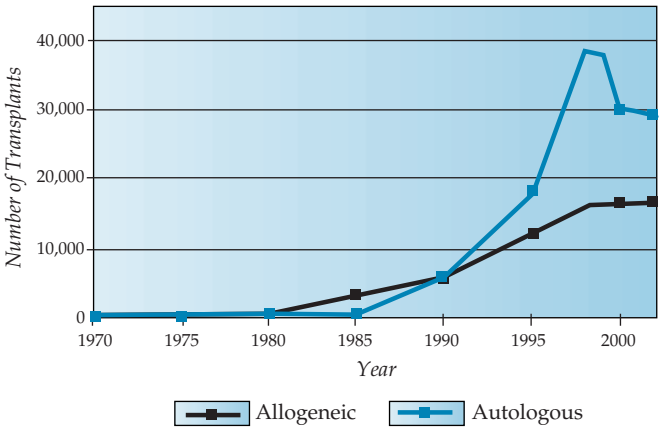


Figure 1 **Depicted are the estimated total numbers of allogeneic and autologous hematopoietic stem cell transplantations performed worldwide from 1970 to 2002, according to estimates of the International Bone Marrow Transplant Registry. The drop in autologous transplantations in 1999 was due to decreased transplantation for breast cancer, whereas the flattening in growth of allogeneic transplantation is the result of fewer transplantations for chronic myelogenous leukemia since the introduction of imatinib mesylate.**

AUTOLOGOUS TRANSPLANTATION

Compared with allogeneic transplantation, autologous transplantation has the advantage of avoiding GVHD and associated complications; disadvantages are that the autologous cells lack the antitumor effect of an infusion of allogeneic leukocytes (the so-called graft versus tumor effect) and may contain viable tumor cells. Removal of tumor cells by use of antibodies to tumor-specific antigens together with complement, toxins, or immunomagnetic beads is very efficient, reducing the number of tumor cells 1,000-fold to 10,000-fold.[10] Other methods of purifying stem cells that are currently under investigation are antibody adherence and flow techniques that select normal hematopoietic stem cells while leaving tumor cells behind; in vitro treatment of the autologous cells with selective chemotherapeutic agents; and in vitro culturing to selectively grow hematopoietic cells. Gene-marking studies have demonstrated that remaining tumor cells can contribute to relapse[11]; however, it remains unknown which, if any, methods of cell purification can prevent relapse. Further, many of the techniques result in delayed hematologic and immunologic recovery after transplantation. Several retrospective analyses suggest that in vitro marrow treatment may be effective in acute myeloid leukemia (AML) and B cell non-Hodgkin lymphoma, but sufficiently large prospective, controlled studies have not been published.[10]

PERIPHERAL BLOOD CELL TRANSPLANTATION

Hematopoietic stem cells normally circulate in the peripheral blood, albeit at very low numbers. Early experiments in animal models showed that at least 10 times more mononuclear cells are needed to rescue animals from lethal total body irradiation when the cells are collected from peripheral blood rather than from marrow of untreated animals. Initial attempts to use peripheral blood stem cells as a source of hematopoietic grafts were complicated by the large number of collections (phereses) required—often seven or more—and by slow engraftment. Subsequently, it was shown that a marked increase in the number of hematopoietic progenitors in the blood, measured either as hematopoietic colony-forming units or as CD34+ cells, occurs during recovery from previous chemotherapy or shortly after exposure to hematopoietic growth factors.[12,13] This led to studies of the use of peripheral blood stem cells as a substitute for marrow. These studies were initially conducted in the autologous setting because peripheral blood stem cell collections contain a large number of T cells, which could induce GVHD if the collections were used for allogeneic transplantation. In the autologous setting, cells sufficient in number to achieve engraftment can usually be collected with one to three leukaphereses of 4 hours' duration after treatment of the patient with granulocyte colony-stimulating factor (G-CSF) or granulocyte-macrophage colony-stimulating factor (GM-CSF). The exact mechanism by which myeloid growth factors cause the remarkable increase in number of peripheral blood stem cells is unclear; however, recent murine studies suggest that myeloid growth factors activate neutrophils to release serine proteases, and serine proteases in turn proteolytically cleave vascular adhesion molecules in the marrow, releasing hematopoietic stem cells.[14] If more than 2.5 million CD34+ cells/kg are collected and subsequently used for

autologous transplantation, recovery to 500 granulocytes/mm^3 within 12 days after transplantation and recovery to 20,000 platelets/mm^3 within 14 days after transplantation almost always occur.[15] This rate of recovery is significantly faster than the rate with autologous marrow stem cells [*see Figure 2*]. Although it is not yet known whether peripheral blood is more likely or less likely than marrow to be contaminated with transplantable tumor cells, mobilized peripheral blood has almost entirely replaced marrow as the source of stem cells for autologous transplantation.

Peripheral blood stem cells have also been used for allogeneic transplantation. Initial studies with G-CSF–mobilized peripheral blood stem cells from HLA-identical matched donors suggested that they engraft more rapidly, and the incidence of acute GVHD did not appear to be greater than would be expected with marrow, despite the transplantation of at least 10 times more mature T cells.[16] Randomized trials have confirmed that the use of peripheral blood stem cells accelerates engraftment without increasing the incidence of acute GVHD.[17-19] Although the incidence of chronic GVHD may be somewhat higher with peripheral blood stem cells, the incidence of tumor recurrence appears to be less and overall disease-free survival tends to be higher with use of peripheral blood stem cells, particularly in patients with higher-risk leukemias.[17-19]

UMBILICAL CORD BLOOD TRANSPLANTATION

Umbilical cord blood is rich in CD34+ cells, and these cells can thus serve as an alternative source of stem cells. The first human umbilical cord blood transplantations were performed in the late 1980s for patients with Fanconi anemia. In a series of 44 children treated with cord blood from siblings, the speed of myeloid engraftment was similar to that seen with marrow transplantation, but platelet recovery was slower.[1] The incidence of GVHD was 6%, which was low; this probably reflected both the young ages of the recipients and the fact that umbilical cord blood is relatively devoid of mature T cells. Subsequently, several studies were undertaken that entailed banking unrelated umbilical cord blood and using it for subsequent transplantation.[19] A summary of the first 562 unrelated cord blood transplants facilitated by the New York Blood Center's program reported engraftment in approximately 80% of patients, but the time to engraftment was significantly prolonged—24 days for neutrophils and 72 days for platelets.[20] A close relation was found between the number of nucleated cord blood cells infused and the incidence and speed of engraftment. The overall incidence of severe acute GVHD was 23%.

Currently, umbilical cord blood is used as the source of stem cells for between 15% and 20% of pediatric allogeneic transplantations. Because the rates of graft failure and transplant-related death increase as the cell dose per kilogram of cord blood drops, umbilical cord blood is rarely used for adult transplantation. Advantages of umbilical cord blood as a source of stem cells include an apparent ability to use donors with greater HLA disparity, a seemingly lower risk of GVHD, and the rapid availability of stored unrelated units. Disadvantages include slower engraftment, a higher incidence of graft failure, slower immune reconstitution, and a trend toward higher disease-recurrence rates. Studies comparing the outcomes of matched unrelated donors with those of unrelated cord blood transplants have, in some cases, shown equivalent survival, but others have shown a trend toward improved survival with the use of unrelated marrow donors.[21]

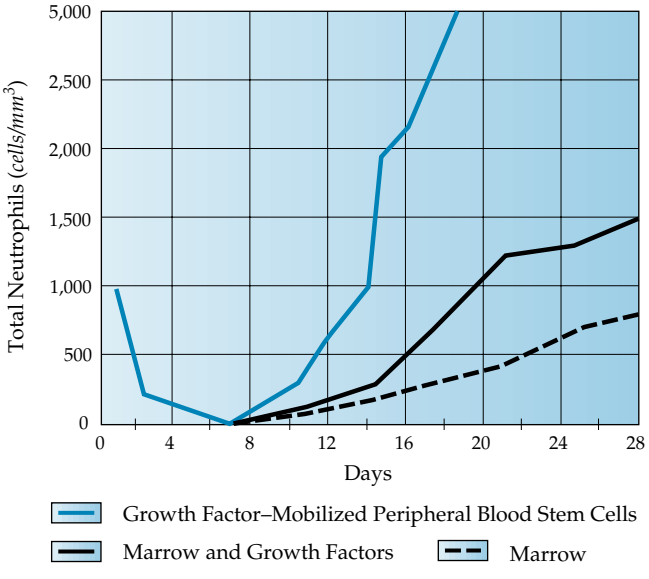


Figure 2 **Shown are the typical patterns of myeloid recovery after hematopoietic stem cell transplantation using marrow alone, marrow plus posttransplant myeloid growth factors, and growth factor–mobilized peripheral blood stem cells.**

Transplantation Procedure

PREPARATIVE REGIMENS

A preparative regimen is administered before hematopoietic stem cells are transplanted. The purpose of this regimen is to eliminate the abnormal or malignant cells causing disease and to suppress the immune system sufficiently to avoid graft rejection. The appropriate regimen for any particular patient is determined by the disease to be treated, the age and health of the patient, and the source of the stem cells to be grafted.

At one extreme, patients undergoing transplantation for the treatment of severe combined immunodeficiency disease with stem cells from an HLA-matched sibling require no preparative regimen, because there are no abnormal cells to eliminate (the disease being caused by a lack of normal cells) and because the patients' immune systems are already sufficiently suppressed to avoid graft rejection. Patients with aplastic anemia lack a normal hematopoietic system but are sufficiently immunocompetent to reject allogeneic marrow if no immunosuppression is given. In this setting, treatment with high-dose cyclophosphamide plus antithymocyte globulin is sufficiently immunosuppressive to ensure engraftment as long as the donor is an HLA-identical sibling. When the transplant is from an unrelated donor, greater immunosuppression is required; thus, low-dose total body irradiation is often added to the treatment. When transplantation is used to treat diseases characterized by an abnormal but nonmalignant population of cells, such as thalassemia and sickle cell anemia, the preparative regimen must eliminate the abnormal population and suppress the patient's immune system. To accomplish this, high-dose busulfan is often added to cyclophosphamide in the preparative regimen.

In developing preparative regimens for transplantation to treat malignancies, most investigators have focused on the use of agents that are highly active against the malignancy being treated and whose dominant dose-limiting toxicity in the nontransplant setting is myelosuppression. Thus, the therapies most commonly used are alkylating agents (e.g., cyclophosphamide,

busulfan, thiotepa, melphalan, carmustine), etoposide, cytarabine, and total body irradiation.

High-dose preparative regimens are typically used when allogeneic transplantation is performed to treat malignancy; however, the allogeneic graft-versus-tumor effect suppresses tumor growth independent of the preparative regimen, which has led investigators to ask whether less intensive, nonmyeloablative regimens may be effective and less toxic. Allogeneic engraftment has been achieved with lower-dose regimens that combine fludarabine with busulfan, cyclophosphamide, or total body irradiation. In one study, for example, patients with hematologic malignancies were treated with fludarabine and 200 cGy of total body irradiation, with encouraging results.[22] Complete responses were achieved in a variety of malignancies, particularly in patients with more indolent hematologic malignancies, with much less toxicity than seen with standard transplant approaches.[22] The decreased toxicity associated with this approach has permitted its study in patients 70 years of age and older. The appropriate role of nonmyeloablative transplants in the treatment of specific malignancies is still being defined.[22,23]

STEM CELL COLLECTION AND INFUSION

Marrow is usually obtained from the donor's anterior and posterior iliac crests with the donor under spinal or general anesthesia. A total marrow volume equivalent to 10 to 15 ml/kg is withdrawn; withdrawal of marrow volume is limited to 3 to 5 ml at each aspiration site to avoid excessive dilution with peripheral blood circulating within the iliac crest. The marrow is filtered through 0.3 mm and 0.2 mm screens to remove bone spicules and fat. The marrow may require further in vitro treatment to remove other cells, such as donor red cells to prevent a hemolytic transfusion reaction in the setting of ABO incompatibility; donor T cells to prevent GVHD; and tumor cells from autologous marrow (see above). The risks associated with marrow donation are small; in one series, there were six serious but nonfatal complications out of 1,220 consecutive marrow donations.

Peripheral blood stem cells are usually collected by use of continuous-flow apheresis from donors previously treated with hematopoietic growth factor alone or, in the case of autologous transplantation, after chemotherapy plus treatment with growth factors. Attempts are made to collect a minimum of 5×10^6 CD34+ cells/kg; there is consistent rapid engraftment with this dose.[15]

Marrow and peripheral blood stem cell infusions are usually well tolerated, though patients sometimes develop fever, cough, or mild shortness of breath. Slowing the infusion usually alleviates these symptoms.

ENGRAFTMENT

The rate of engraftment depends on the source of the stem cells, the choice of prophylaxis against GVHD, and whether hematopoietic growth factors are used. The most rapid engraftment is seen with peripheral blood stem cells; in this setting, engraftment is typically achieved by day 12. In marrow or umbilical cord blood transplantation, the granulocyte count usually reaches 100 cells/mm^3 by about day 16 and 500 cells/mm^3 by day 22. The rate of myeloid recovery can be accelerated by 4 to 6 days with the use of G-CSF or GM-CSF after marrow transplantation, but posttransplant growth factors have less of an effect when mobilized peripheral blood is the source of transplanted stem cells [*see Figure 2*].[24] The use of methotrexate after allogeneic transplantation delays recovery by an average of 4 days. Platelet recovery generally occurs shortly after granulocyte recovery.

Complications of Transplantation

EARLY DIRECT TOXICITIES OF THE PREPARATIVE REGIMEN

Pretransplant preparative regimens are associated with a substantial array of toxicities, which vary considerably depending on the specific regimen used. For example, nausea, vomiting, and mild skin erythema develop immediately in almost all patients after the standard cyclophosphamide–total body irradiation regimen. Occasionally, hemorrhagic cystitis is seen despite bladder irrigation or therapy with a sulfhydryl compound (MESNA); in rare instances (fewer than 2% of cases), acute hemorrhagic carditis develops. Oral mucositis inevitably develops about 5 to 7 days after transplantation, usually requiring narcotic analgesia. Patient-controlled analgesia provides the greatest patient satisfaction and, surprisingly, results in a lower cumulative dose of narcotics. Within 10 days after transplantation, complete alopecia and profound granulocytopenia have developed in most patients.

Veno-occlusive disease of the liver (also referred to as sinusoidal obstruction syndrome) is a serious complication of high-dose chemoradiotherapy; it develops in approximately 10% to 20% of patients.[25] Veno-occlusive disease of the liver, characterized by the development of ascites, tender hepatomegaly, jaundice, and fluid retention, may occur at any time during the first month after transplantation; the peak incidence occurs at around day 16. Histologic features of veno-occlusive disease of the liver include concentric narrowing or fibrous obliteration of terminal hepatic venules and sublobular veins and necrosis of zone 3 hepatocytes [*see Figure 3*]. Predisposing factors are pretransplant hepatitis of any cause and the use of more intensive conditioning

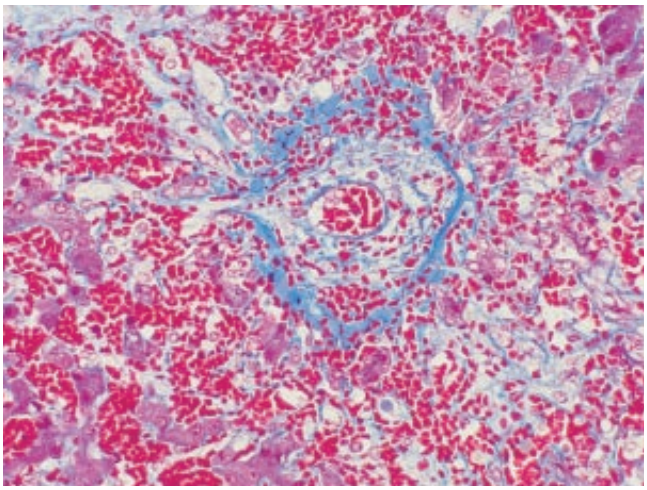


Figure 3 **Photomicrograph of a liver biopsy stained with Trichrome shows the typical changes of veno-occlusive disease of the liver. A sublobular vein is outlined by dense, blue connective tissue in the outer adventitial layer. There is marked narrowing of the lumen of the vein by a widened and edematous subendothelial zone containing trapped red cells and loose extracellular matrix. The hepatocyte cords surrounding the vein are necrotic, and the intervening sinusoids are hemorrhagic. Deposition of coagulants in the sinusoids and in the subendothelial zone of the vein obstructs outflow of blood from the liver, producing sinusoidal hypertension and hepatomegaly.**

regimens. Although the precise sequence of events leading to the clinical presentation of veno-occlusive disease is unknown, direct cytotoxic injury to hepatic venular and sinusoidal endothelium occurs early on, with subsequent deposition of fibrin and the development of a local hypercoagulable state. Direct cytotoxic injury to zone 3 hepatocytes is a contributing factor. Approximately 30% of patients who develop veno-occlusive disease of the liver die as a result of the disease, with progressive hepatic failure leading to a terminal hepatorenal syndrome. Antithrombotic and thrombolytic agents, including prostaglandin E_1 and tissue plasminogen activator, with or without heparin, have been evaluated as treatment. These therapies have been associated with significant toxicities, and randomized trials demonstrating efficacy are lacking. It has been reported that defibrotide, a polydeoxyribonucleotide, may benefit patients with veno-occlusive disease.[26]

Most pneumonias that occur after transplantation are caused by microbial agents, but idiopathic interstitial pneumonia occurs in up to 5% of patients.[27] Most experts consider idiopathic interstitial pneumonia to be a direct toxicity of intensive chemotherapy, but evidence for a role of soluble cytokines is growing. Biopsies reveal some cases to be characterized by diffuse alveolar damage, whereas other cases have a more clearly interstitial component. Treatment with high-dose glucocorticoids is often attempted, but randomized trials evaluating their efficacy have not been performed.

LATE DIRECT TOXICITIES OF THE PREPARATIVE REGIMEN

Direct complications of chemoradiotherapy seen late after transplantation include a decreased growth rate in children and a delay in the development of secondary sex characteristics. Most children will have a deficiency in growth factor and should undergo replacement therapy. Ovarian failure develops in most postpubertal women. Azoospermia develops in most men. Aseptic osteonecrosis occurs in as many as 10% of transplant patients, particularly in those with chronic GVHD necessitating corticosteroid treatment. Similarly, cataracts develop in 10% to 20% of patients; the risk is higher in patients who take steroids to treat chronic GVHD.[28]

Patients treated with high-dose chemoradiotherapy and hematopoietic cell transplantation are at increased risk for the development of secondary malignancies.[29,30] The risk is highest in patients receiving T cell–depleted marrow and those who receive multiple cycles of highly immunosuppressive drugs after transplantation to treat GVHD; in such cases, a high incidence of Epstein-Barr virus–associated lymphoproliferative disease is seen. A smaller increase is seen in the incidence of solid tumors after transplantation, with a 2.9% 10-year cumulative rate. The actuarial incidence of myelodysplasia after autologous transplantation for non-Hodgkin lymphoma and Hodgkin disease may be as high as 10%.[31]

GRAFT FAILURE

Although complete and sustained engraftment is the general rule after transplantation, marrow function does not return in some cases; and in other cases, after temporary engraftment, marrow function is lost. Graft failure after autologous transplantation can result from marrow damage before harvesting, during ex vivo treatment, during storage, or after exposure to myelotoxic agents after transplantation.[32] Infections with cytomegalovirus (CMV) or human herpesvirus type 6 may also result in poor marrow function. Graft failure after allogeneic transplantation may be the result of immunologically mediated graft rejection and is more common after conditioning regimens that are less immunosuppressive, in recipients of T cell–depleted marrow, and in recipients of HLA-mismatched marrow.

The treatment of graft failure begins with removal of all potentially myelosuppressive agents. A reasonable second step is to attempt a short trial of a myeloid growth factor (GM-CSF or G-CSF); 40% to 50% of patients respond.[33] Identification of persistent host lymphocytes in peripheral blood or marrow of the patient suggests immunologic rejection. These patients should receive further immunosuppression before a second transplant is performed. Several studies have reported successful second transplants after a regimen of cyclophosphamide and antithymocyte globulin or a regimen of anti-CD3 antibody and high-dose steroids.

GRAFT VERSUS HOST DISEASE

When allogeneic T cells that are transferred with the graft or that develop from it react with targets of the genetically different host, GVHD results.[34] GVHD that develops within the first 3 months after transplantation is categorized as acute and is characterized by an erythematous maculopapular skin rash that, unlike many rashes, often appears on the palms and soles [*see 2:VI Cutaneous Adverse Drug Reactions*]. Acute GVHD is also characterized by persistent anorexia or diarrhea, or both, and by liver disease, evidenced by increased levels of bilirubin, alanine and aspartate aminotransferases, and alkaline phosphatase. Also characteristic of acute GVHD is epithelial damage to the skin, liver, and intestines [*see Figure 4*]. Skin, liver, and endoscopic intestinal biopsies are the usual methods of establishing a diagnosis, and they may reveal damaged epidermis and hair follicles, segmental disruption in small bile ducts, and intestinal mucosal ulceration caused by destruction of intestinal crypts. Clinical

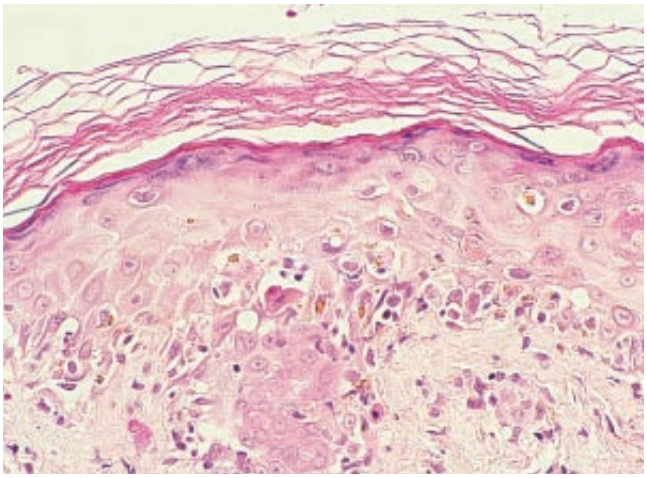


Figure 4 **This photomicrograph of a skin biopsy stained with hematoxylin and eosin demonstrates the features of acute graft versus host disease of the skin. Both intercellular edema (spongiosis) and intracellular edema (reticular degeneration) of the lower epidermis are evident. Mononuclear cells are scattered throughout the epidermal area along with many bodies that have undergone apoptosis, including dead epidermal cells with hypereosinophilic cytoplasm and dense basophilic pyknotic nuclei. The inflammatory process has produced incontinence of melanin pigment into the papillary dermis, as well as coarse intraepidermal blocks of melanin, leading to hyperpigmentation. Because the apoptosis and basal layer damage are extensive, the epidermis will become grossly scaly and will possibly slough.**

staging of acute GVHD is determined by the extent of involvement of skin, liver, and gut [*see Table 1*]. The incidence of acute GVHD increases in older patients, in recipients of mismatched marrow, and in patients unable to receive full doses of the drugs used to prevent GVHD.[35]

Two general approaches are used to prevent acute GVHD: use of immunosuppressive agents during the early posttransplant period and removal of T cells from the transplanted cell population. Methotrexate alone and cyclosporine alone are equally effective as prophylaxis, but their use in combination is more effective. Prednisone, FK 506 (tacrolimus), rapamycin, and mycophenolate mofetil have also been used in various combinations.[36] Removal of T cells from the allogeneic marrow is effective in preventing acute GVHD, but in most circumstances, it has been associated with an increased incidence of graft rejection and leukemic relapse. Accordingly, several potential therapies are now under study, including partial T cell depletion and complete T cell depletion followed by the reintroduction of a fraction of the T cells. Once acute GVHD develops, it can be treated with glucocorticoids, antithymocyte globulin, and monoclonal antibodies targeted against T cells or their receptors.

GVHD that develops or persists 3 months or more after transplantation is termed chronic GVHD. Chronic GVHD has features in common with collagen vascular diseases, including a malar rash, sclerodermatous changes, sicca syndrome, arthritis, obliterative bronchiolitis, and, in some cases, bile duct degeneration and cholestasis. Chronic GVHD develops in 20% to 40% of patients, more often in older patients and in those who previously had acute GVHD.[37] Prednisone, cyclosporine, or the two in combination is the usual treatment[38]; in some studies, azathioprine or thalidomide was useful.[39] In most patients, chronic GVHD eventually resolves and immunosuppressive therapy can be withdrawn, but 1 to 3 years of treatment may be required. Patients with chronic GVHD who are on immunosuppressive therapy are susceptible to bacterial infections and should receive prophylactic antibiotics.

INFECTIOUS DISEASES

During the first 2 to 3 weeks after transplantation, all patients are severely granulocytopenic. To reduce the risk of disseminated bacterial infections, posttransplantation patients are usually placed on broad-spectrum antibiotics once they become granulocytopenic. In addition, the prophylactic administration of fluconazole reduces the incidence of *Candida albicans* infection.[40] The treatment of patients who become febrile despite prophylactic antibiotic and antifungal therapy is a difficult challenge; in such cases, therapy is guided by individual aspects of the patient's condition and by the institution's experience. For example, if fever develops in a patient who received prophylactic treatment with levofloxacin, the subsequent choice of antibiotic might be guided by whether an intra-abdominal source of infection is suspected, in which case meropenem or imipenem is appropriate therapy; when intra-abdominal infection is not a concern, ceftazidime might be selected.[41] If fever persists for more than 72 hours, amphotericin B or voriconazole is often added to the treatment regimen.[42] Granulocyte transfusions can be effective in treating specific infections, particularly now that donors can be treated with G-CSF before donation to greatly increase the number of granulocytes that can be collected and transfused.[42] There is no established role, however, for prophylactic granulocyte transfusions. Laminar airflow isolation can reduce the incidence of infection but has no impact on survival in transplant patients treated for malignancy. With current methods of supportive care, the risk of death from an infectious cause during the period of granulocytopenia is less than 5%.

In the past, CMV infection frequently occurred after transplantation, particularly in recipients of allogeneic marrow. It has been shown, however, that primary CMV infection can be prevented in CMV-seronegative patients by the use of CMV-seronegative blood products. In CMV-seropositive patients, treatment with ganciclovir as soon as virus excretion is evident can diminish the incidence of CMV-associated disease and death, but in some patients, CMV disease develops before or at the same time as viral excretion is noted. Ganciclovir prophylaxis beginning at the time of engraftment can prevent the development of CMV infection in most patients, but ganciclovir causes significant marrow suppression in at least 10% of patients.[43] At most centers, after transplantation, peripheral blood is monitored for the development of CMV antigenemia, and prophylaxis with ganciclovir is initiated only if and when patients test positive for the presence of CMV antigen. Foscarnet is effective for patients who develop CMV antigenemia or infection despite ganciclovir therapy or for patients who cannot tolerate ganciclovir.

Herpes simplex virus infection, when not prevented, contributes to the severity of early oral mucositis and esophagitis. However, the prophylactic use of acyclovir at a dosage of 250 mg/m^2 I.V. every 8 hours can prevent herpes simplex virus reactivation in almost all seropositive patients.

Pneumocystis carinii once caused pneumonia in 5% to 10% of patients after transplantation, but now this complication can be prevented in virtually all patients by first treating the patient with oral trimethoprim-sulfamethoxazole for 1 week before transplantation and then resuming treatment 2 days a week once engraftment occurs and continuing it for as long as immunosuppressive therapy is given. Desensitization should be attempted in patients with allergic reactions to trimethoprim-sulfamethoxazole. Dapsone (50 mg p.o., b.i.d.), although not as effective as trimethoprim-sulfamethoxazole, can serve as a substitute but must be avoided in patients with glucose-6-phosphate dehydrogenase deficiency.

Table 1 Clinical Staging of Acute Graft versus Host Disease*

Stage	*Skin Changes*	*Liver*	*Gut*
I	Maculopapular rash < 25% body surface	Bilirubin 2–3 mg/dl	Diarrhea 500–1,000 ml/day
II	Maculopapular rash 25%–30% body surface	Bilirubin 3–6 mg/dl	Diarrhea 1,000–1,500 ml/day
III	Generalized erythroderma	Bilirubin 6–15 mg/dl	Diarrhea > 1,500 ml/day
IV	Desquamation and bullae	Bilirubin > 15 mg/dl	Pain and ileus

*Overall severity ranges from mild skin involvement (stage I) to severe multiorgan involvement, usually with a fatal outcome (stage IV).

Table 2 Disease-Free Survival after Hematopoietic Stem Cell Transplantation

Disease	*Five-Year Disease-Free Survival (%)*
Chronic myeloid leukemia	
Chronic phase	60–70
Accelerated phase	30–40
Blastic phase	15–20
Acute myeloid leukemia	
First remission	40–70
Second remission	30
Chronic lymphocytic leukemia	50
Acute lymphocytic leukemia	30–50
Multiple myeloma	35
Non-Hodgkin lymphoma	
After first relapse, chemosensitive tumors	40–50
Hodgkin disease	
First treatment after standard treatment	40–70
Advanced disease	15–30
Myelodysplastic syndromes	45
Severe aplastic anemia	> 90
Thalassemia major	70–90
Fanconi anemia	50–70
Sickle cell disease	50–90
Severe combined immunodeficiency disease	90
Breast cancer (stage IV disease)	10–30

More than 3 months after transplantation, patients are still at risk for varicella-zoster virus infections and, if they have chronic GVHD, for recurrent bacterial infections. Varicella-zoster virus infection usually occurs initially as localized disease (i.e., herpes zoster, or shingles), but it can disseminate; disseminated infection is often fatal if left untreated. Thus, patients with localized varicella-zoster virus infection should be treated with acyclovir to prevent dissemination. Many centers now routinely place all allogeneic transplant recipients on prophylactic acyclovir therapy for the first year after transplantation. In an effort to reduce late bacterial infections, many centers place patients with chronic GVHD receiving immunosuppression on daily trimethoprim-sulfamethoxazole therapy.

Hematopoietic Stem Cell Transplantation for Specific Diseases

TREATMENT OF IMMUNODEFICIENCY STATES

The widest experience in treating immunodeficiency with hematopoietic stem cell transplantation has been in the treatment of severe combined immunodeficiency disease.[44] When current techniques of supportive care are used, the expected outcome of transplantation from an HLA-identical donor is excellent, with a better than 90% probability of long-term survival [*see Table 2*]. In patients without matched donors, transplantation from a haplotype-mismatched parent results in engraftment and survival longer than 2 years in 50% to 70% of patients. The experience in the treatment of Wiskott-Aldrich syndrome and other immunodeficiency states is limited. Cures have been noted in more than half of patients, with the best results seen in patients who undergo transplantation when they are younger than 5 years.

TREATMENT OF NONMALIGNANT DISEASES OF HEMATOPOIESIS

Aplastic Anemia

Transplantation from matched siblings after a preparative regimen of high-dose cyclophosphamide and antithymocyte globulin, together with the use of methotrexate and cyclosporine for GVHD prophylaxis, is a very effective regimen for patients with aplastic anemia. Current results suggest a cure rate greater than 90% [*see Table 2*].[45] Results with mismatched or unrelated matched donors are somewhat worse; therefore, patients with aplastic anemia who are without sibling donors are often given a trial of immunosuppressive therapy before transplantation.

Thalassemia

Marrow transplantation from an HLA-identical sibling after a preparative regimen of busulfan and cyclophosphamide can cure from 70% to 90% of patients with thalassemia major [*see Table 2*]. The best results have been obtained in patients who undergo transplantation before they develop hepatomegaly or portal fibrosis and who have been given adequate iron chelation therapy. In one study of 121 such patients, the probabilities of survival and disease-free survival 5 years after transplantation were 95% and 90%, respectively.[46] Prolonged survival can also be achieved with aggressive chelation therapy, but transplantation remains the only curative treatment. Fewer than 30% of patients with thalassemia have an HLA-identical sibling. Selection of alternative donors of hematopoietic stem cells (i.e., unrelated persons or HLA-nonidentical family members) has been aided by the establishment of worldwide donor registries, by improvements in the methods of controlling GVHD, and by prevention of fungal and cytomegalovirus infection.

Sickle Cell Anemia

Experience in transplantation for sickle cell disease is small but growing. In a European study of 100 patients with sickle cell disease who received transplants from HLA-matched siblings, the survival rate at 4 years was 88%, and disease-free survival was 80%.[47] In a study of 59 patients in the United States, similar rates were reported—93% and 84%, respectively.[48]

Other Nonmalignant Diseases

Hematopoietic stem cell transplantation has been used successfully to treat a variety of other nonmalignant but nonetheless fatal diseases. Included in this group are congenital disorders of white cells, including Kostmann syndrome, chronic granulomatous disease, neutrophil actin defects, leukocyte adhesion deficiency, and Chédiak-Higashi syndrome. Congenital anemias, including Fanconi anemia and Blackfan-Diamond anemia, are likewise treatable with hematopoietic stem cell transplantation.[49,50] Osteopetrosis is a rare inherited disorder caused by an inability of the osteoclast to resorb bone. Because the osteoclast is a specialized macrophage derived from the marrow, it follows that osteopetrosis can be treated with marrow transplantation.

A final category of treatable nonmalignant diseases are storage diseases caused by enzymatic deficiencies, including Maroteaux-Lamy syndrome, metachromatic leukodystrophy, Gaucher disease, Hurler syndrome, and Hunter syndrome [*see 5:VII Nonmalignant Disorders of Leukocytes*]. Transplantation for these disorders has not been universally successful, but treatment early in the disease course, before irreversible end-organ damage occurs, increases the chance for a successful outcome. Studies have recently been done on the use of transplantation in the treatment of severe autoimmune disorders. These studies are based on the demonstration that transplantation can cure autoimmune diseases in some animal models and on the observation that some patients with coexistent hematologic malignancies and autoimmune disorders have been cured of both with transplantation.[51]

TREATMENT OF MALIGNANT DISEASES

Acute Myeloid Leukemia

Allogeneic marrow transplantation cures 15% to 20% of patients with AML in whom induction therapy fails and, indeed, is the only therapy that can cure such patients.[52] Thus, all patients 55 years of age or younger with newly diagnosed AML should have their HLA type determined, as should their families, soon after diagnosis to enable transplantation for those in whom induction therapy fails. Allogeneic transplantation can cure approximately 30% of patients in second remission and 35% of patients in untreated first relapse—situations that are clear indications for the procedure, because these results are superior to those achieved without transplantation.[53,54] The best results with allogeneic transplantation for AML are obtained in patients undergoing transplantation in first remission, in whom a cure rate of 40% to 70% is reported [*see Table 2*].

In 14 prospective studies, the cure rate with marrow transplantation for patients with HLA-matched siblings ranged from 40% to 64%, whereas the cure rate with chemotherapy for patients without HLA-matched siblings ranged from 19% to 24%, suggesting that allogeneic transplantation is the preferred form of postremission therapy for younger patients with AML and a matched sibling.[55-57] However, important advances have been made in both chemotherapy and transplantation since many of these studies were conducted. Further, it remains untested whether transplantation in first remission is superior to a regimen of initial chemotherapy followed by transplantation as salvage therapy.

In several phase II studies, autologous transplantation for patients with AML in first or second remission yielded results not dissimilar to those achieved with allogeneic transplantation. In general, relapse rates after autologous transplantation have been substantially higher than those seen with allogeneic transplantation, but the mortality associated with transplant-related complications has been lower. Several large trials compared allogeneic transplantation, autologous transplantation, and aggressive chemotherapy as postremission therapy for patients with AML in first remission. The European Organization for Research and Treatment of Cancer compared these therapies in 333 patients with AML in first remission and found the rate of disease-free survival at 4 years to be 54% with allogeneic transplantation, 49% with autologous transplantation, and 30% with continued chemotherapy.[58] In a recent update of a similarly designed American Intergroup study, the estimated rate of survival at 5 years was 52% with allogeneic transplantation, 42% with autologous transplantation, and 39% with chemotherapy.[59] In a study from the United Kingdom Medical Research Council, patients with AML in first remission who had been treated with three cycles of postremission therapy were randomized either to receive no further therapy or to undergo autologous transplantation in first remission.[60] The group who underwent transplantation had a lower relapse rate, and the rates of disease-free survival and overall survival were improved. The benefits of allogeneic transplantation over autologous transplantation or chemotherapy are most apparent for patients with high-risk AML, as determined by cytogenetic risk status and other prognostic factors.[59]

Acute Lymphocytic Leukemia

As with AML, allogeneic transplantation can cure 15% to 20% of patients with acute lymphocytic leukemia (ALL) in whom induction therapy fails or in whom chemotherapy-resistant disease develops; thus, these patients are candidates for the procedure. The results of transplantation for patients in second remission are better, with cure rates of 30% to 50% [*see Table 2*]. However, further intensive chemotherapy also can cure some patients who suffered an initial relapse. This is particularly true for children who experience a relapse more than 18 months after initial induction chemotherapy. In a study comparing the use of allogeneic transplantation in 255 children with the use of chemotherapy in an equal number of children, the rates of disease-free survival at 5 years were found to be 40% in transplant patients and 17% in chemotherapy patients. The relative benefits of transplantation were similar for children with short and long initial remissions. Thus, allogeneic transplantation can be recommended for all patients with ALL in second complete remission who have appropriate donors.[61]

Allogeneic transplantation for ALL in first remission results in long-term disease-free survival in 40% to 70% of adult patients. In a retrospective study comparing these results with those achieved with chemotherapy, no clear advantage could be found for either approach.[62] In the largest prospective, randomized study published to date (involving 572 patients), the 10-year survival rate for those undergoing allogeneic transplantation was 46%; for those undergoing autologous transplantation, 34%; and for those receiving continued chemotherapy, 31%.[63] In standard-risk patients, there was no difference in outcome between the three approaches (i.e., 10-year survival of 49% with allogeneic transplantation, 49% with autologous transplantation, and 40% with chemotherapy), whereas for high-risk patients, allogeneic transplantation provided the best results (44% versus 10% versus 11%). Because children with ALL, in general, respond well to chemotherapy, there is no role for transplantation at first remission except for those with very high risk disease (e.g., Philadelphia chromosome–positive ALL).[64]

Myelodysplastic Syndromes

The myelodysplastic syndromes are generally considered to be incurable except with marrow transplantation. In some patients, the myelodysplastic syndromes have a relatively indolent course, and transplantation can be safely withheld until the disease progresses. However, once significant granulocytopenia (fewer than 1,000 cells/mm^3) or thrombocytopenia (fewer than 40,000 cells/mm^3) develops or the proportion of blast cells in the marrow exceeds 5%, transplantation should be seriously considered, because without transplantation, the expected survival time is short. When an HLA-matched sibling is available to

serve as a donor, the chance of long-term survival with transplantation is roughly 55%, with better results being obtained in younger patients and in those who receive transplants earlier in the course of their disease.[65] Similar results have been reported with matched unrelated donor transplants.[66] No role has been established for autologous transplantation in the myelodysplastic syndromes, but this is a current focus of research.

Chronic Myeloid Leukemia

Allogeneic and syngeneic marrow transplantation are the only forms of therapy known to cure chronic myeloid leukemia (CML). Five-year disease-free survival rates are 15% to 20% for patients who undergo transplantation in blast crisis, 30% to 40% for patients who undergo transplantation during the accelerated phase, and 60% to 70% for patients who undergo transplantation during the chronic phase [*see Table 2*].

Time from diagnosis influences the outcome of transplantation during the chronic phase. The best results are obtained in patients who receive transplants within 1 year of diagnosis; progressively worse results are seen the longer the procedure is delayed.[61] Previous exposure to busulfan is an adverse risk factor for transplantation.[67] A growing number of patients between 55 and 65 years of age with CML have undergone transplantation, with results not significantly worse than those seen in younger patients.[68] Although the initial experience with the use of unrelated-donor transplantation in CML was substantially worse than the experience with matched-sibling transplantation, more recent results at some centers have demonstrated a 70% probability of disease-free survival at 3 years.[69]

The overall role of hematopoietic cell transplantation in CML has become more complex with the introduction of imatinib mesylate, a very effective, relatively nontoxic oral agent used for treatment of CML.[70] Imatinib does not result in complete, molecular-level remissions in most patients, and therefore, many investigators would agree that early allogeneic transplantation remains the treatment of choice for younger patients with matched donors. For older patients or those without appropriate donor matches, an initial trial of imatinib is generally preferred.

Chronic Lymphocytic Leukemia

Use of marrow transplantation in chronic lymphocytic leukemia (CLL) has received only limited attention, probably because of the indolent nature of the disease and its propensity to occur in older patients. Of the small number of patients receiving allogeneic transplantation, many have had complete remissions, and approximately half have remained disease free.[71,72] However, the transplant-related mortality in this group of patients has been substantial. The number of patients treated with autologous transplantation is limited.[71,73] Complete remissions have been achieved, some of which appear to be sustained. Further study is needed to determine whether transplantation significantly improves long-term outcome in CLL.

Non-Hodgkin Lymphoma

Patients with disseminated intermediate or high-grade non-Hodgkin lymphoma in whom conventional therapy fails can seldom be cured without transplantation. High-dose therapy followed by autologous or allogeneic marrow transplantation can cure a substantial number of such patients. A number of studies have documented cure rates of 40% to 50% in patients who receive transplants after an initial relapse and whose tumors remain sensitive to chemotherapy [*see Table 2*].[74] In one randomized study of 216 patients, the 5-year disease-free survival rate for patients who underwent autologous transplantation for chemosensitive disease was 46%, compared with 12% for patients in the chemotherapy group ($P = 0.001$). Cure rates decrease substantially once the disease becomes resistant to conventional-dose chemotherapy.[74] A poor performance status and large tumor bulk are additional adverse risk factors. As in other diseases, patients who receive transplants of allogeneic marrow have a lower relapse rate but a higher risk of nonrelapse mortality than patients who receive transplants of autologous marrow.[75]

For most categories of intermediate- and high-grade non-Hodgkin lymphoma, the outcomes for allogeneic and autologous transplantation appear roughly similar, though an advantage has been suggested for the use of allogeneic transplantation in patients with lymphoblastic lymphoma. The role of transplantation for patients in first remission is unsettled. Of the randomized studies that have thus far been performed, some have found a significant benefit, some have found no benefit, and others have found a benefit only for the subgroup of patients with intermediate-to-high risk disease or high-risk disease.[76,77]

Autologous marrow transplantation has also been studied in patients with low-grade disease. The probabilities of survival at 3 years after transplantation have averaged 83% for patients who undergo transplantation in first remission, 65% for patients in first relapse or second remission, and 50% for patients with more advanced disease. Whether these results are superior to what could be achieved with therapies that do not involve transplantation has not been demonstrated in prospective, randomized trials. Further, late relapses have been seen after autologous transplantation, raising questions about the possibility of cure, but studies of allogeneic transplantation in patients with low-grade disease have shown long-term disease-free survival in some of these patients.

Hodgkin Disease

The results of transplantation for Hodgkin disease mimic those for non-Hodgkin lymphoma. A substantial number of patients in whom first-line chemotherapy for Hodgkin disease has failed can be cured with salvage transplantation. Results of treatment for recurrent Hodgkin disease have been shown to be better when transplantation is performed in patients who have chemotherapy-sensitive disease with minimal tumor bulk and a good performance status. In this setting, cure rates of 40% to 70% have been reported [*see Table 2*]. As with non-Hodgkin lymphoma, lower relapse rates but higher nonrelapse mortality are seen with the use of allogeneic stem cells than with the use of autologous stem cells.[78] There is currently no established role for transplantation for patients in first remission, though pilot studies in patients with high-risk disease are being performed.

Multiple Myeloma

Allogeneic marrow transplantation has been used to treat patients with multiple myeloma. For patients with multiple myeloma in whom first-line therapy has failed, allogeneic transplantation has achieved overall survival rates averaging 35% at 5 years[*see Table 2*]. An important finding is that there appears to be a plateau in the rate of disease-free survival, suggesting that although some of these patients were cured, transplant-associated complications were substantial and occurred more frequently than in most other hematologic malignancies.[79] With improvements in supportive care measures, the outcome of allogeneic transplantations for multiple myeloma has improved.[80]

Autologous transplantation has also been studied. There is less evidence that this approach can lead to long-term cure, at least with current techniques. However, when employed before patients have developed resistant disease, autologous transplantation can result in a substantial reduction in tumor burden and, in many cases, at least temporary complete remission. In two large randomized studies, the addition of high-dose chemotherapy and autologous transplantation to the initial treatment regimen led to significantly longer disease-free survival and greater overall survival than that which occurred with conventional chemotherapy alone.[81,82] A strategy of autologous transplantation followed by nonmyeloablative allogeneic transplantation has been applied with very encouraging early results.

Other Hematologic Malignancies

Long-term survival has been documented after allogeneic marrow transplantation in patients with hairy-cell leukemia, myelofibrosis, and various myeloproliferative syndromes, but the number of patients reported in any one disease category is small.

Breast Cancer

In women with stage IV breast cancer, high-dose chemotherapy followed by autologous transplantation results in a higher rate of complete remission than standard-dose chemotherapy. Studies of a large number of patients have reported disease-free survival at 5 years to be 10% to 30% [*see Table 2*].[83] In patients who underwent transplantation for stage IV disease, the highest rates of progression-free survival (32%) were seen in patients who received transplants after they had achieved a complete response to conventional chemotherapy; lower rates were seen in patients with partially responding disease (13%) or progressive disease (7%). Although these results appear to be superior to those achieved with standard-dose chemotherapy, longer follow-up and the completion of randomized trials are needed to determine whether patients who achieved complete response are cured and whether this percentage of patients is truly higher than that seen with standard-dose chemotherapy. On the basis of these initial high response rates, high-dose chemotherapy followed by autologous transplantation has been increasingly studied in patients with earlier-stage disease. Pooled registry results show a 42% 3-year disease-free survival for patients who received transplants for inflammatory breast cancer and a 74% 3-year disease-free survival for women with high-risk (> 10 nodes) stage II disease. On the basis of these results, randomized trials were initiated to test whether autologous transplantation would lead to improved cure rates for women with high-risk disease. Two such trials have been published: the Netherlands trial and an Eastern Cooperative Oncology Group study. In the Netherlands trial, high-dose alkylating therapy followed by transplantation improved relapse-free survival rates in patients with stage II or III disease and 10 or more positive nodes.[84] In an Eastern Cooperative Oncology Group study, the use of intensive chemotherapy reduced the risk of relapse of disease but did not improve disease-free survival rates.[85]

Testicular Cancer

Although standard-dose chemotherapy for testicular cancer is very effective, conventional regimens fail in 30% to 40% of cases. High-dose chemotherapy with autologous marrow support has resulted in a 2-year disease-free survival of approximately 20% in patients with advanced recurrent disease—a rate seemingly better than that achieved with conventional approaches.[86]

Other Solid Tumors

The utility of high-dose chemotherapy with autologous stem cell support for several other solid tumors, including ovarian cancer, small cell lung cancer, neuroblastoma, and pediatric sarcomas, is being studied. As in virtually all other situations, best results occur in patients with limited tumor bulk in whom the tumor remains sensitive to conventional-dose chemotherapy. Randomized trials evaluating the usefulness of transplantation in these settings have not been reported.

Treatment of Posttransplant Relapse

Patients with malignancies who experience relapse after autologous transplantation occasionally respond to further conventional-dose chemotherapy, particularly when the interval from transplantation to relapse is long. There are more options available to the patient who experiences relapse after allogeneic transplantation. Patients with CML frequently respond to interferon therapy, and other patients occasionally respond to withdrawal of immunosuppression. Patients who experience relapse after allogeneic transplantation sometimes respond to nonirradiated donor lymphocyte infusions. In a summary of 258 patients reported by a European registry, complete responses were seen in 75% of patients with CML, 38% with myelodysplasia, 24% with AML, and 15% with myeloma.[87] Responses were seldom seen in ALL. The major complications of posttransplant donor lymphocyte infusions have been GVHD and myelosuppression, both of which can be severe or fatal. Starting the transfusion with a low cell dose and then gradually increasing the dose can lessen the risk of severe toxicity. A second hematopoietic cell transplant can occasionally be effective, particularly in younger patients and in patients who experience a longer interval from first transplant to relapse and who do not have advanced disease.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. International Bone Marrow Transplant Registry (IBMTR) and the Autologous Blood and Marrow Transplant registry (ABMTR) IBMTR/ABMTR Statistical Center, Health Policy Institute of the Medical College of Wisconsin, Milwaukee, 2003 http://www.ibmtr.org

2. Spangrude GJ, Heimfeld S, Weissman IL: Purification and characterization of mouse hematopoietic stem cells. Science 241:58, 1988

3. Korbling M, Katz RL, Khanna A, et al: Hepatocytes and epithelial cells of donor origin in recipients of peripheral-blood stem cells. N Engl J Med 346:738, 2002

4. Tran SD, Pillemer SR, Dutra A, et al: Differentiation of human bone marrow-derived cells into buccal epithelial cells in vivo: a molecular analytical study. Lancet 361:1084, 2003

5. Anasetti C, Amos D, Beatty PG, et al: Effect of HLA compatibility on engraftment of bone marrow transplants in patients with leukemia or lymphoma. N Engl J Med 320:197, 1989

6. National Marrow Donor Program, Minneapolis, Minnesota, 2003 http://www.nmdp.org

7. Kernan NA, Bartsch G, Ash RC, et al: Analysis of 462 transplantations from unrelated donors facilitated by the National Marrow Donor Program. N Engl J Med 328:593, 1993

8. Petersdorf EW, Gooley TA, Anasetti C, et al: Optimizing outcome after unrelated marrow transplantation by comprehensive matching of HLA class I and II alleles in the donor and recipient. Blood 92:3515, 1998

9. Petersdorf EW, Hansen JA, Martin PJ, et al: Major-histocompatibility-complex class I alleles and antigens in hematopoietic-cell transplantation. N Engl J Med 345:1794, 2001

10. Gribben JG, Freedman AS, Neuberg D, et al: Immunologic purging of marrow assessed by PCR before autologous bone marrow transplantation for B-cell lymphoma. N Engl J Med 325:1525, 1991

11. Brenner MK, Rill DR, Moen RC, et al: Gene-marking to trace origin of relapse after autologous bone-marrow transplantation. Lancet 341:85, 1993

12. Richman CM, Weiner RS, Yankee RA: Increase in circulating stem cells following

chemotherapy in man. Blood 47:1031, 1976

13. Socinski MA, Cannistra SA, Elias A, et al: Granulocyte-macrophage colony stimulating factor expands the circulating haemopoietic progenitor cell compartment in man. Lancet 1:1194, 1988

14. Levesque JP, Takamatsu Y, Nilsson SK, et al: Vascular cell adhesion molecule-1 (CD106) is cleaved by neutrophil proteases in the bone marrow following hematopoietic progenitor cell mobilization by granulocyte colony-stimulating factor. Blood 98:1289, 2001

15. Bensinger WI, Longin K, Appelbaum F, et al: Peripheral blood stem cells (PBSCs) collected after recombinant granulocyte colony stimulating factor (rhG-CSF): an analysis of factors correlating with the tempo of engraftment after transplantation. Br J Haematol 87:825, 1994

16. Bensinger WI, Weaver CH, Appelbaum FR, et al: Transplantation of allogeneic peripheral blood stem cells mobilized by recombinant human granulocyte colony-stimulating factor. Blood 85:1655, 1995

17. Bensinger WI, Martin P, Storer B, et al: A prospective, randomized trial of transplantation of marrow vs. peripheral blood stem cells from HLA-identical siblings in patients treated for hematologic malignancies. N Engl J Med (in press)

18. Couban S, Simpson DR, Barnett MJ, et al: A randomized multicenter comparison of bone marrow and peripheral blood in recipients of matched sibling allogeneic transplants for myeloid malignancies. Blood 100:1525, 2002

19. Schmitz N, Beksac M, Hasenclever D, et al: Transplantation of mobilized peripheral blood cells to HLA-identical siblings with standard-risk leukemia. Blood 100:761, 2002

20. Rubinstein P, Carrier C, Scaradavou A, et al: Outcomes among 562 recipients of placental-blood transplants from unrelated donors. N Engl J Med 339:1565, 1998

21. Rocha V, Cornish J, Sievers EL, et al: Comparison of outcomes of unrelated bone marrow and umbilical cord blood transplants in children with acute leukemia. Blood 97:2962, 2001

22. McSweeney PA, Niederwieser D, Shizuru JA, et al: Hematopoietic cell transplantation in older patients with hematologic malignancies: replacing high-dose cytotoxic therapy with graft-versus-tumor effects. Blood 97:3390, 2001

23. Khouri IF, Keating M, Korbling M, et al: Transplant-lite: induction of graft-versus-malignancy using fludarabine-based nonablative chemotherapy and allogeneic blood progenitor-cell transplantation as treatment for lymphoid malignancies. J Clin Oncol 16:2817, 1998

24. Nemunaitis J, Rabinowe SN, Singer JW, et al: Recombinant granulocyte-macrophage colony-stimulating factor after autologous bone marrow transplantation for lymphoid cancer. N Engl J Med 324:1773, 1991

25. Bearman SI: The syndrome of hepatic veno-occlusive disease after marrow transplantation. Blood 85:3005, 1995

26. Richardson PG, Elias AD, Krishnan A, et al: Treatment of severe veno-occlusive disease with defibrotide: compassionate use results in response without significant toxicity in a high-risk population. Blood 92:737, 1998

27. Crawford SW, Hackman RC: Clinical course of idiopathic pneumonia after marrow transplantation. Am Rev Respir Dis 147:1393, 1993

28. Flowers MD, Deeg HJ: Delayed complications after hematopoietic cell transplantation. Hematopoietic Cell Transplantation, 3rd ed. Blume KG, Forman SJ, Appelbaum FR, Eds. Blackwell Publishers, Malden, Massachusetts, 2004, p 944

29. Deeg HJ, Socie G, Schoch G, et al: Malignancies after marrow transplantation for aplastic anemia and Fanconi's anemia: a joint Seattle and Paris analysis of results in 700 patients. Blood 87:386, 1996

30. Curtis RE, Rowlings PA, Deeg HJ, et al: Solid cancers after bone marrow transplantation. N Engl J Med 336:897, 1997

31. Stone RM, Neuberg D, Soiffer R, et al: Myelodysplastic syndrome as a late complication following autologous bone marrow transplantation for non-Hodgkin's lymphoma. J Clin Oncol 12:2535, 1994

32. Rabinowe SN, Neuberg D, Bierman PJ, et al: Long-term follow-up of a phase III study of recombinant human granulocyte-macrophage colony-stimulating factor after autologous bone marrow transplantation for lymphoid malignancies. Blood 81:1903, 1993

33. Nemunaitis J, Singer JW, Buckner CD, et al: Use of recombinant human granulocyte-macrophage colony-stimulating factor in graft failure after bone marrow transplantation. Blood 76:245, 1990

34. Ferrara JLM, Deeg HJ: Graft-versus-host disease. N Engl J Med 324:667, 1991

35. Nash RA, Pepe MS, Storb R, et al: Acute graft-versus-host disease: analysis of risk factors after allogeneic marrow transplantation and prophylaxis with cyclosporine and methotrexate. Blood 80:1838, 1992

36. Chao NJ: Pharmacology and the use of immunosuppressive agents after hematopoietic cell transplantation. Hematopoietic Cell Transplantation, 3rd ed. Blume KG, Forman SJ, Appelbaum FR, Eds. Blackwell Publishers, Malden, Massachusetts, 2004, p 209

37. Atkinson K, Horowitz MM, Gale RP, et al: Risk factors for chronic graft-versus-host disease after HLA-identical sibling bone marrow transplantation. Blood 75:2459, 1990

38. Koc S, Leisenring W, Flowers ME, et al: Therapy for chronic graft-versus-host disease: a randomized trial comparing cyclosporine plus prednisone versus prednisone alone. Blood 100: 48, 2002

39. Vogelsang GB, Farmer ER, Hess AD, et al: Thalidomide for the treatment of chronic graft-versus-host disease. N Engl J Med 326:1055, 1992

40. Goodman JL, Winston DJ, Greenfield RA, et al: A controlled trial of fluconazole to prevent fungal infections in patients undergoing bone marrow transplantation. N Engl J Med 326:845, 1992

41. Pizzo PA: Management of fever in patients with cancer and treatment-induced neutropenia. N Engl J Med 328:1323, 1993

42. Bensinger WI, Price TH, Dale DC, et al: The effects of daily recombinant human granulocyte colony-stimulating factor administration on normal granulocyte donors undergoing leukapheresis. Blood 81:1883, 1993

43. Goodrich JM, Bowden RA, Fisher L, et al: Ganciclovir prophylaxis to prevent cytomegalovirus disease after allogeneic marrow transplant. Ann Intern Med 118:173, 1993

44. Buckley RH, Schiff SE, Schiff RI, et al: Hematopoietic stem-cell transplantation for the treatment of severe combined immunodeficiency. N Engl J Med 340:508, 1999

45. Storb R, Leisenring W, Anasetti C, et al: Long-term follow-up of allogeneic marrow transplants in patients with aplastic anemia conditioned by cyclophosphamide combined with antithymocyte globulin (letter). Blood 89:3890, 1997

46. Lucarelli G, Galimberti M, Giardini C, et al: Bone marrow transplantation in thalassemia. Ann N Y Acad Sci 30:850, 1998

47. Vermylen C, Cornu G, Ferster A, et al: Haematopoietic stem cell transplantation for sickle cell anaemia: the first 50 patients transplanted in Belgium. Bone Marrow Transplant 22:1, 1998

48. Walters MC, Storb R, Patience M, et al: Impact of bone marrow transplantation for symptomatic sickle cell disease: an interim report. Blood 95:1918, 2000

49. Guardiola P, Pasquini R, Dokal I, et al: Outcome of 69 allogeneic stem cell transplantations for Fanconi anemia using HLA-matched unrelated donors: a study on behalf of the European Group for Blood and Marrow Transplantation. Blood 95:422, 2000

50. MacMillan ML, Auerbach AD, Davies SM, et al: Haematopoietic cell transplantation in patients with Fanconi anaemia using alternate donors: results of a total body irradiation dose escalation trial. Br J Haematol 109:121, 2000

51. Nash RA, Bowen JD, McSweeney PA, et al: High-dose immunosuppressive therapy and autologous peripheral blood stem cell transplantation. Blood 102:2364, 2003

52. Biggs JC, Horowitz MM, Gale RP, et al: Bone marrow transplants may cure patients with acute leukemia never achieving remission with chemotherapy. Blood 80:1090, 1992

53. Clift RA, Buckner CD, Appelbaum FR, et al: Allogeneic marrow transplantation during untreated first relapse of acute myeloid leukemia. J Clin Oncol 10:1723, 1992

54. Gale RP, Horowitz MM, Rees JKH, et al: Chemotherapy versus transplants for acute myelogenous leukemia in second remission. Leukemia 10:13, 1996

55. Nesbit ME Jr, Buckley JD, Feig SA, et al: Chemotherapy for induction of remission of childhood acute myeloid leukemia followed by marrow transplantation or multiagent chemotherapy: a report from the children's cancer group. J Clin Oncol 12:127, 1994

56. Archimbaud E, Thomas X, Michallet M, et al: Prospective genetically randomized comparison between intensive postinduction chemotherapy and bone marrow transplantation in adults with newly diagnosed acute myeloid leukemia. J Clin Oncol 12:262, 1994

57. Hewlett J, Kopecky JK, Head D, et al: A prospective evaluation of the roles of allogeneic marrow transplantation and low-dose monthly maintenance chemotherapy in the treatment of adult acute myelogenous leukemia (AML): a Southwest Oncology Group study. Leukemia 9:562, 1995

58. Zittoun RA, Mandelli F, Wellemze R, et al: Autologous or allogeneic bone marrow transplantation compared with intensive chemotherapy in acute myelogenous leukemia. N Engl J Med 332:217, 1995

59. Slovak ML, Kopecky KJ, Cassileth PA, et al: Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group study. Blood 96:4075, 2000

60. Burnett AK, Goldstone AH, Stevens RM, et al: Randomised comparison of addition of autologous bone-marrow transplantation to intensive chemotherapy for acute myeloid leukaemia in first remission: results of MRC AML 10 trial. Lancet 351:700, 1998

61. Barrett AJ, Horowitz MM, Pollock BH, et al: Bone marrow transplants from HLA-identical siblings as compared with chemotherapy for children with acute lymphoblastic leukemia in a second remission. N Engl J Med 19:1253, 1994

62. Horowitz MM, Messerer D, Hoelzer D, et al: Chemotherapy compared with bone marrow transplantation for adults with acute lymphoblastic leukemia in first remission. Ann Intern Med 115:13, 1991

63. Thiebaut A, Vernant JP, Degos L, et al: Adult acute lymphocytic leukemia study testing chemotherapy and autologous and allogeneic transplantation. A follow-up report of the French protocol LALA 87 (review). Hematol Oncol Clin North Am 14:1353, 2000

64. Arico M, Valsecchi MG, Camitta B, et al: Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. N Engl J Med 342:998, 2000

65. Appelbaum FR, Anderson JE: Allogeneic bone marrow transplantation for myelodysplastic syndrome: outcomes analysis according to IPSS score. Leukemia 12(suppl 1): S25, 1998

66. Deeg HJ, Storer B, Slattery JT, et al: Conditioning with targeted busulfan and cyclophosphamide for hemopoietic stem cell transplantation from related and unrelated donors in patients with myelodysplastic syndrome. Blood 100:1201, 2002

67. Goldman JM, Szydlo R, Horowitz MM, et al: Choice of pretransplant treatment and timing of transplants for chronic myelogenous leukemia in chronic phase. Blood 82:2235, 1993

68. Clift RA, Appelbaum FR, Thomas ED: Treatment of chronic myeloid leukemia by marrow transplantation (editorial). Blood 82:1954, 1993

69. Hansen JA, Gooley TA, Martin PJ, et al: Unrelated donor marrow transplants for patients with chronic myeloid leukemia in chronic phase. N Engl J Med 338:962, 1998

70. O'Brien SG, Guilhot F, Larson RA, et al: Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 348:994, 2003

71. Michallet M, Archimbaud E, Bandini G, et al: HLA-identical sibling bone marrow transplantation in younger patients with chronic lymphocytic leukemia: European

Group for Blood and Marrow Transplantation and the International Bone Marrow Transplant Registry. Ann Intern Med 124:311, 1996

72. Doney KC, Chauncey T, Appelbaum FR: Allogeneic related donor hematopoietic stem cell transplantation for treatment of chronic lymphocytic leukemia. Bone Marrow Transplant 29:817, 2002

73. Rabinowe SN, Soiffer RJ, Gribben JG, et al: Autologous and allogeneic bone marrow transplantation for poor prognosis patients with B-cell chronic lymphocytic leukemia. Blood 82:1366, 1993

74. Philip T, Guglielmi C, Hagenbeek A, et al: Autologous bone marrow transplantation as compared with salvage chemotherapy in relapses of chemotherapy-sensitive non-Hodgkin's lymphoma. N Engl J Med 333:1540, 1995

75. Peniket AJ, Ruiz de Elvira MC, Taghipour G, et al: An EBMT registry matched study of allogeneic stem cell transplants for lymphoma: allogeneic transplantation is associated with a lower relapse rate but a higher procedure-related mortality rate than autologous transplantation. Bone Marrow Transplantation 31:667, 2003

76. Haioun C, Lepage E, Gisselbrecht C, et al: Benefit of autologous bone marrow transplantation over sequential chemotherapy in poor-risk aggressive non-Hodgkin's lymphoma: updated results of the prospective study LNH87-2. J Clin Oncol 15:1131, 1997

77. Gianni AM, Bregni M, Siena S, et al: High-dose chemotherapy and autologous bone marrow transplantation compared with MACOP-B in aggressive B-cell lymphoma. N Engl J Med 336:1290, 1997

78. Anderson JE, Litzow MR, Appelbaum FR, et al: Allogeneic, syngeneic, and autologous marrow transplantation for Hodgkin's disease: the 21-year Seattle experience. J Clin Oncol 11:2342, 1993

79. Gahrton G, Tura S, Ljungman P, et al: Allogeneic bone marrow transplantation in multiple myeloma. N Engl J Med 325:1267, 1991

80. Gahrton G, Svensson H, Cavo M, et al: Progress in allogenic bone marrow and peripheral blood stem cell transplantation for multiple myeloma: a comparison between transplants performed 1983-93 and 1994-98 at European Group for Blood and Marrow Transplantation centers. Br J Haematol 113:209, 2001

81. Attal M, Harousseau JL, Stoppa AM, et al: A prospective, randomized trial of autologous bone marrow transplantation and chemotherapy in multiple myeloma. N Engl J Med 335:91, 1996

82. Child JA, Morgan GJ, Davies FE, et al: High-dose chemotherapy with hematopoietic stem-cell rescue for multiple myeloma. N Engl J Med 348:1875, 2003

83. Antman K, Rowlings P, Vaughn W, et al: High-dose chemotherapy with autologous hematopoietic stem cell support for breast cancer in North America. J Clin Oncol 15:1870, 1997

84. Rodenhuis S, Bontenbal M, Beex LVAM, et al: High-dose chemotherapy with hematopoietic stem-cell rescue for high-risk breast cancer. N Engl J Med 349:7, 2003

85. Tallman MS, Gray R, Robert NJ, et al: Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem-cell transplantation in high-risk breast cancer. N Engl J Med 349:17, 2003

86. Broun ER, Nichols CR, Kneebone P, et al: Long-term outcome of patients with relapsed and refractory germ cell tumors treated with high-dose chemotherapy and autologous bone marrow rescue. Ann Intern Med 117:124, 1992

87. Kolb HJ, Schattenberg A, Goldman JM, et al: Graft-versus-leukemia effect of donor lymphocyte transfusions in marrow grafted patients. Blood 86:2041, 1995

Acknowledgments

Figures 1 and 2 Marcia Kammerer.

Figures 3 and 4 Courtesy of the Fred Hutchinson Cancer Research Center.

This work was supported in part by grants CA-18029, CA-47748, and CA-26386 from the National Institutes of Health, U.S. Department of Health and Human Services.

LAWRENCE L. K. LEUNG, M.D.

Hemostasis, the process of blood clot formation, is a coordinated series of responses to vessel injury. It requires complex interactions between platelets, the clotting cascade, blood flow and shear, endothelial cells, and fibrinolysis.

Platelet Plug Formation

Platelets are activated at the site of vascular injury to form a plug to stop bleeding. Physiologic platelet stimuli include adenosine diphosphate (ADP), epinephrine, thrombin, and collagen. ADP and epinephrine are relatively weak platelet stimulators; thrombin and collagen are strong agonists. Thrombin activation is mediated by G protein–coupled protease-activated receptors (PAR),[1] specifically PAR-1 and PAR-4. Thrombin cleaves the external domain of the PAR to initiate transmembrane signaling [*see Figure 1*].[2] Platelet responses to ADP require the coordinated activation of two G–protein–coupled receptors, P2Y1 and P2Y12, which lead to activation of phospholipase C and suppression of cyclic adenosine monophosphate (cAMP) formation, respectively. Antiplatelet drugs such as ticlopidine and clopidogrel block activation of P2Y12.[3] There are also specific receptors for epinephrine, thromboxane A_2, and collagen.

Platelet activation involves four distinct processes: adhesion (deposition of platelets on subendothelial matrix); aggregation (cohesion of platelets); secretion (release of platelet granule proteins); and procoagulant activity (enhancement of thrombin generation) [*see Figure 2*].

ADHESION

Platelet adhesion is primarily mediated by the binding of platelet surface receptor glycoprotein (GP) Ib-IX-V complex to the adhesive protein von Willebrand factor (vWF) in the subendothelial matrix.[4] Deficiency of GPIb-IX-V complex or vWF leads to two congenital bleeding disorders, Bernard-Soulier disease and von Willebrand disease, respectively [*see 5:XIII Hemorrhagic Disorders*]. Other adhesive interactions (e.g., binding of platelet collagen receptor GPIa-IIa to collagen fibrils in the matrix) also contribute to platelet adhesion.[5]

AGGREGATION

Platelet aggregation involves binding of fibrinogen to the platelet fibrinogen receptor (i.e., the GPIIb-IIIa complex). GPIIb-IIIa (also termed aIIbb3) is a member of a superfamily of adhesive protein receptors, called integrins, which are found in many different cell types. It is the most abundant receptor on the platelet surface. GPIIb-IIIa does not bind fibrinogen on nonstimulated platelets. After platelet stimulation, GPIIb-IIIa undergoes a conformational change and is converted from a low-affinity fibrinogen receptor to a high-affinity receptor in a process termed inside-out signaling. Fibrinogen, a divalent molecule, serves to bridge the activated platelets [*see Figure 3*]. The cytosolic portion of the activated GPIIb-IIIa complex binds to the platelet cytoskeleton and can mediate platelet spreading and clot retraction (in a process termed outside-in signaling).[6] Congenital deficiency of GPIIb-IIIa or fibrinogen leads to Glanzmann thrombasthenia and afibrinogenemia. The GPIIb-IIIa–fibrinogen pathway is the final common course for platelet aggregation. Blockade of this pathway is the basis of an important class of antiplatelet drugs.

PROTEIN SECRETION

After stimulation, platelet granules release ADP and serotonin, which stimulate and recruit additional platelets; adhesive proteins such as fibronectin and thrombospondin, which reinforce and stabilize platelet aggregates; factor V, a component of the clotting cascade; thromboxane, which stimulates vasoconstriction; and growth factors such as platelet-derived growth factor (PDGF), which stimulate proliferation of smooth muscle cells and mediate

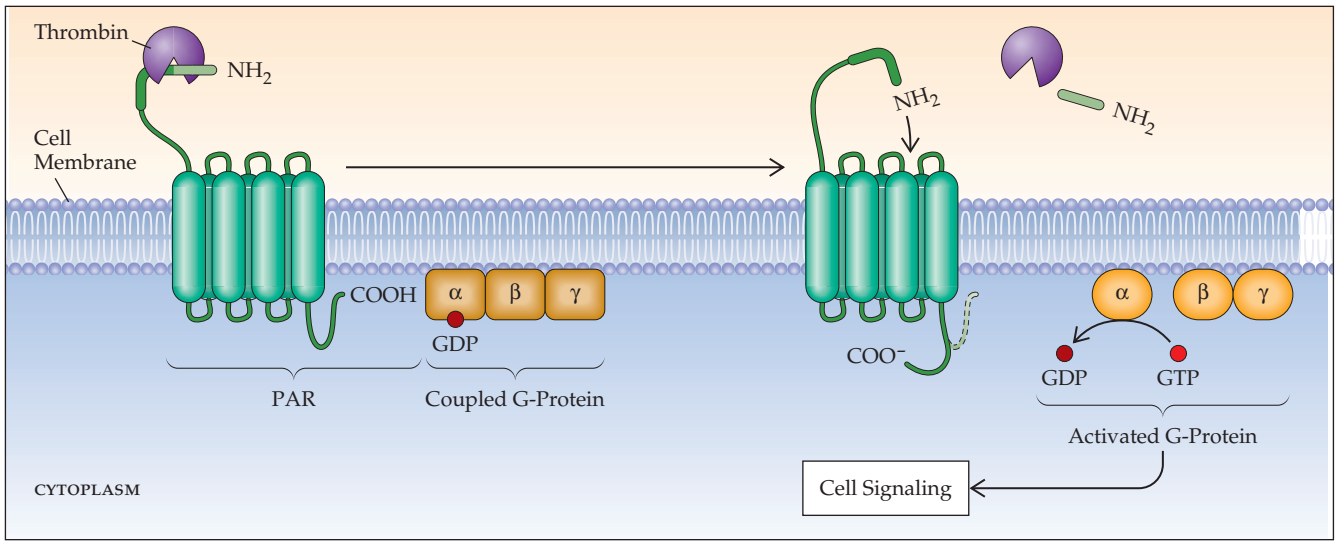


Figure 1 **Thrombin activation is mediated by G protein–coupled protease-activated receptor (PAR). Thrombin cleaves the NH_2-terminal exodomain of the PAR, exposing a new NH_2 terminus, which then serves as a tethered ligand to bind intramolecularly to the body of the receptor to initiate transmembrane signaling.**

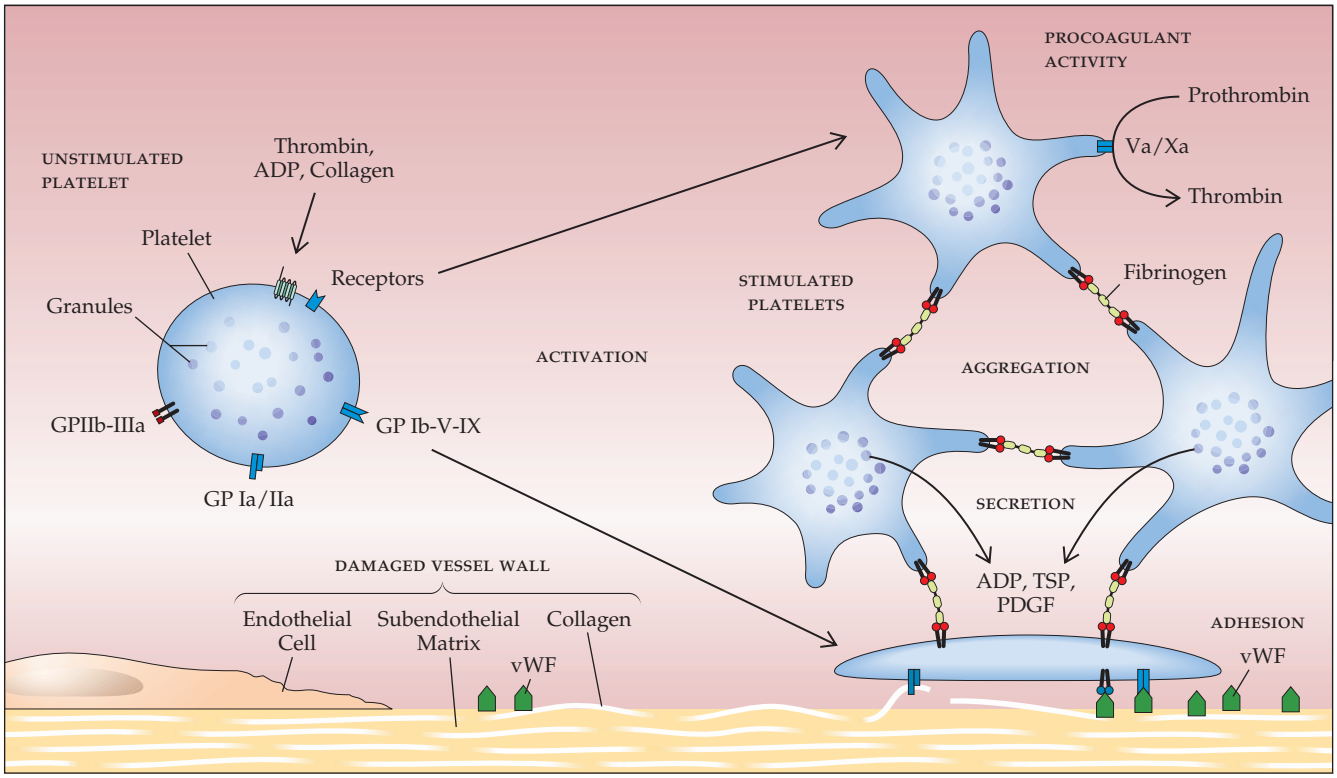


Figure 2 **After platelets are activated, they undergo significant morphologic changes, producing elongated pseudopods. They also become extremely adhesive. The functional response of activated platelets involves four distinct processes: adhesion (deposition of platelets on subendothelial matrix); aggregation (cohesion of platelets); secretion (release of platelet granule proteins); and procoagulant activity (enhancement of thrombin generation).**

tissue repair. PDGF may also contribute to the development of atherosclerosis and reocclusion after coronary angioplasty.

PROCOAGULATION

Platelet procoagulation involves the assembly of the enzyme complexes of the clotting cascade on the platelet surface. It is an important example of the close interrelationship between platelet activation and the activation of the clotting cascade.

Clotting Cascade

The central feature of the clotting cascade is the sequential activation of a series of proenzymes (zymogens) to enzymes, ultimately generating fibrin and reinforcing the platelet plug. Another key feature, amplification, ensures rapid response for effective hemostasis but demands tight regulation to prevent untoward thrombosis.

The clotting cascade is usually depicted as comprising intrinsic and extrinsic pathways [*see Figure 4*]. The intrinsic pathway is initiated by the exposure of blood to a negatively charged surface (e.g., glass), whereas the extrinsic pathway is activated by tissue factor or thromboplastin. Both pathways converge on the activation of factor X, which then activates prothrombin (factor II) to thrombin, the final enzyme of the clotting cascade.

Although this classic view of the clotting cascade has been useful in the interpretation of clotting times, it is not completely accurate. Patients who are severely deficient in factor XII—as well as many patients deficient in factor XII—do not bleed clinically, which indicates that the initiation part of the intrinsic pathway (the contact phase) is not important in vivo. It is now established that generation or exposure of tissue factor at the wound site is the primary physiologic event that initiates clotting [*see Figure 4*].[7] Tissue factor functions as a cofactor that is absolutely required by factor VII/factor VIIa to initiate clotting. Factor VIIa activates factor X directly and indirectly via the activation of factor IX. This dual pathway of factor X activation is necessary apparently because of the limited amount of tissue factor generated in vivo and the presence of the tissue factor pathway inhibitor (see below), which, when complexed with factor Xa, inhibits the tissue factor/factor VIIa complex.

All of the procoagulants are synthesized in the liver except vWF, which is synthesized in megakaryocytes and endothelial cells. The vitamin K–dependent procoagulants are prothrombin, factor VII, factor IX, and factor X; the vitamin K–dependent anticoagulants are protein C and protein S. For these factors, the formation of α-carboxyglutamic acid residues by vitamin K–dependent carboxylation of glutamic acid residues endows them with calcium-binding properties and the ability to interact with phospholipid membrane surfaces, which are required for biologic activity.[8]

INTERACTION BETWEEN ACTIVATED PLATELETS AND THE CLOTTING CASCADE

There is an extremely close interaction between the clotting cascade and activated platelet surface in vivo. When platelets are activated, anionic lipids become exposed on the platelet surface, and factor V (stored in platelet granules) is released and bound on the anionic lipids. The factor V on the platelet surface is acti-

vated to factor Va and acts as an assembly site for the binding of factor Xa (enzyme) and prothrombin (substrate) known as the prothrombinase complex. At the assembly site, thrombin generation by the prothrombinase complex is approximately 300,000 times more efficient than thrombin generation by fluid-phase factor Xa and prothrombin alone, and the platelet plug keeps the thrombin localized. Factor Xa bound on factor Va is also relatively protected from inhibition by circulating inhibitors such as antithrombin III (AT-III) (see below). Similar enzyme complex assembly applies to the activation of factor X by factor VIIIa (cofactor) and factor IXa (the intrinsic tenase). The result of these processes is efficient amplification and localization of clotting.

Control Mechanisms

Coagulation is modulated by a number of mechanisms: dilution of procoagulants in flowing blood; removal of activated factors through the reticuloendothelial system, especially in the liver; and control by natural antithrombotic pathways. At least seven separate and distinct control systems modulate each phase of hemostasis and protect against thrombosis, vascular inflammation, and tissue damage [*see Table 1*]. Antithrombin III, protein C, protein S, and tissue factor pathway inhibitor (TFPI) collectively regulate the clotting cascade; prostacyclin and nitric oxide modulate vascular and platelet reactivity; ecto-ADPase inhibits platelet recruitment; and fibrinolysis removes the fibrin clot.

ANTITHROMBIN III–HEPARAN SULFATE SYSTEM

Antithrombin III is a circulating plasma protease inhibitor. It inhibits thrombin and factor Xa, the two key enzymes in the clotting cascade. AT-III also inhibits activated factor XII and factor XI. In the absence of the glycosaminoglycan heparin, AT-III inhibits thrombin and factor Xa relatively slowly (complete inhibition requires a few minutes). When present, heparin binds to a discrete binding site on AT-III that causes a conformational change in AT-III, which then inhibits thrombin instantaneously and irreversibly. This augmentation of the inhibition of thrombin and factor Xa is the basis for the therapeutic use of heparin as an anticoagulant. Heparan sulfate proteoglycans on the luminal surface of endothelial cells appear to activate AT-III in a manner similar to that of heparin [*see Figure 5*].[9]

Thus, the endothelial surface is normally coated with a layer of AT-III that is already activated by the endogenous heparan sulfate. Because 1 ml of blood can be exposed to as much as 5,000 cm^2 of endothelial surface, the AT-III–heparan sulfate system is poised to rapidly inactivate any thrombin in the general circulation.

PROTEIN C AND PROTEIN S–THROMBOMODULIN SYSTEM

Thrombomodulin is an integral membrane protein found on the luminal surface of the vascular endothelium in the microcirculation. The binding of thrombin to thrombomodulin results in a remarkable switch in thrombin's substrate specificities: it no longer clots fibrinogen or activates platelets [*see Figure 6*]. On the other hand, it acquires the ability to activate protein C in plasma.[10] A distinct endothelial receptor for protein C has been found that enhances the activation of protein C by the thrombin-thrombomodulin complex.[11] Activated protein C degrades factor Va and factor VIIIa, the two cofactors responsible for the assembly of the prothrombinase and intrinsic tenase complex in the clotting cascade. Protein S serves as a cofactor for activated protein C. Deficiencies of AT-III, protein C, and protein S are important causes of a hypercoagulable state.

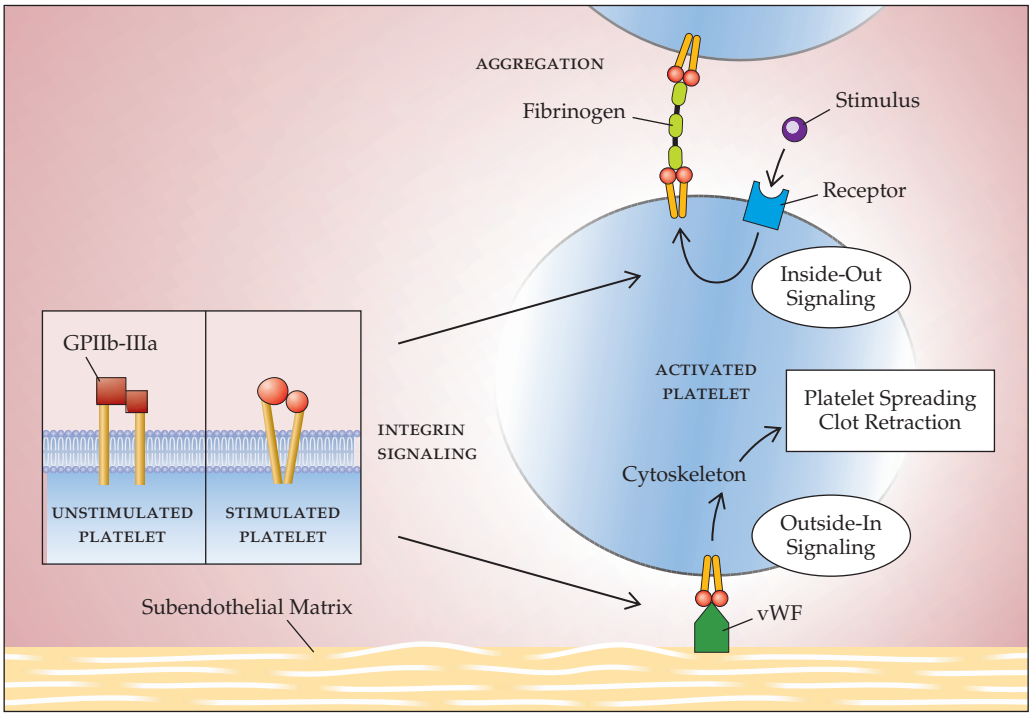


Figure 3 **Platelet aggregation involves binding of the divalent molecule fibrinogen to the platelet fibrinogen receptor (the GPIIb-IIIa complex). After platelet stimulation, GPIIb-IIIa is converted from a low-affinity fibrinogen receptor to a high-affinity receptor (inside-out signaling). The cytosolic portion of the activated GPIIb-IIIa complex can mediate platelet spreading and clot retraction (outside-in signaling).**

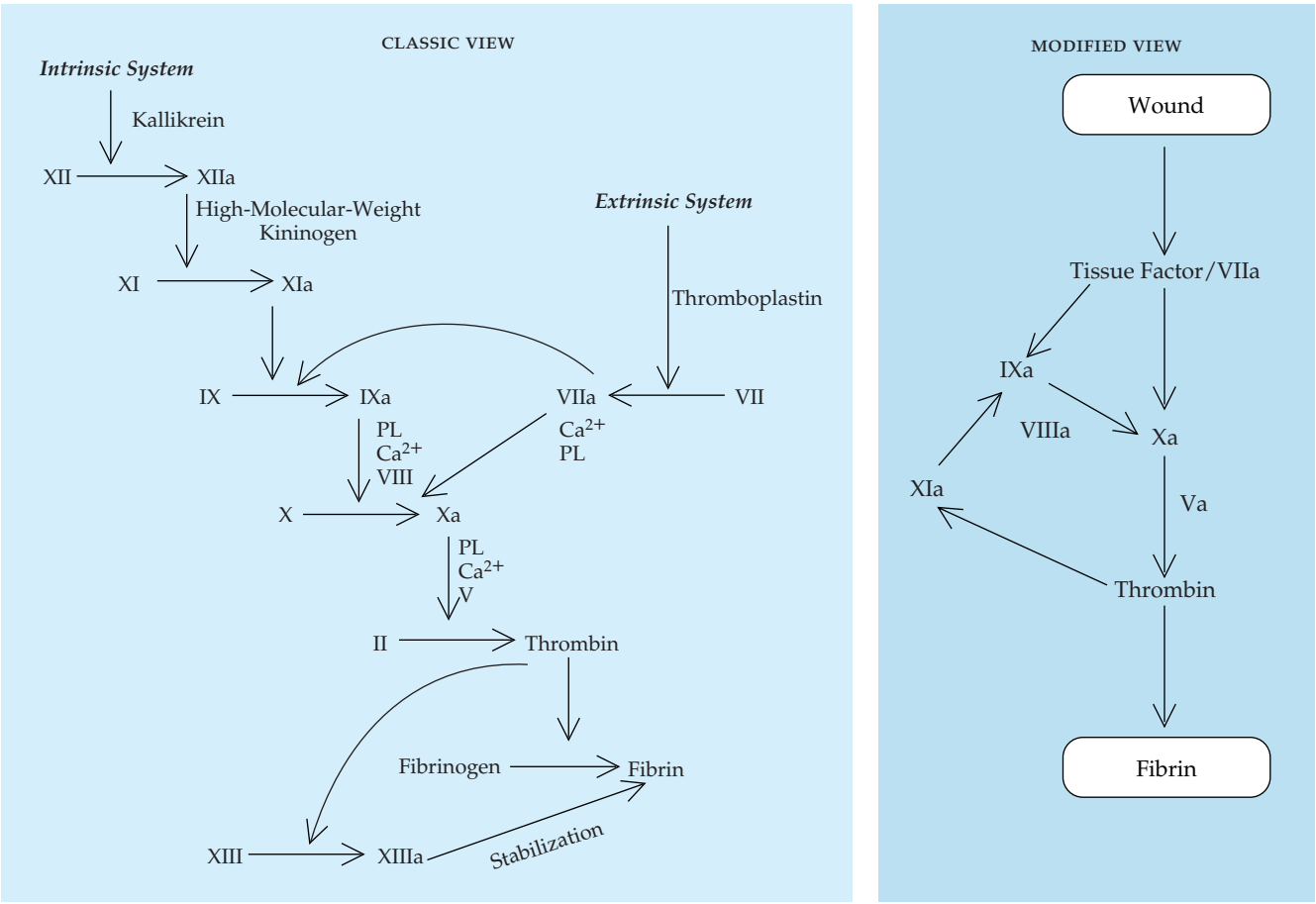


Figure 4 **In the classic view of the clotting cascade (*left*), the intrinsic pathway is initiated by the exposure of blood to a negatively charged surface (e.g., glass) and the extrinsic pathway is activated by tissue factor or thromboplastin. In the modified view (*right*), generation or exposure of tissue factor at the wound site is the primary physiologic event that initiates clotting.**

Protein C and protein S both show some structural similarity to the vitamin K–dependent clotting factors (prothrombin, factor VII, factor IX, and factor X). Protein S circulates in two forms: a free form, in which it is active as an anticoagulant, and a bound, inactive form, in which it is complexed to C4b-binding protein of the complement system. C4b-binding protein acts as an acute-phase reactant. The resultant increase in inflammatory states reduces the activity of free protein S, enhancing the likelihood of thrombosis.

TISSUE FACTOR PATHWAY INHIBITOR

Tissue factor pathway inhibitor is a circulating plasma protease inhibitor that is synthesized by the microvascular endothelium. Unlike AT-III, TFPI has a very low plasma concentration. TFPI inhibits factor Xa. The TFPI/factor Xa complex becomes an effective inhibitor of tissue factor/factor VIIa, thus mediating feedback inhibition of tissue factor/factor VIIa [*see Figure 7*]. Animal studies have shown that depletion of the endogenous TFPI sensitizes the animals to the development of disseminated intravascular coagulation induced by tissue factor or endotoxin.[12]

TFPI is primarily synthesized by the microvascular endothelium. Approximately 20% of TFPI circulates in plasma associated with lipoproteins; the majority remains associated with the endothelial surface, apparently bound to the cell-surface glycosaminoglycans. The plasma level of TFPI is greatly increased after intravenous administration of heparin. This release of endothelial TFPI may contribute to the antithrombotic efficacy of heparin and low-molecular-weight heparin. Recombinant TFPI is now in early clinical trials.[13]

PROSTACYCLIN

Upon cell perturbation, the fatty acid arachidonic acid is released from cell membrane phospholipids by phospholipase A_2. The enzyme prostaglandin endoperoxide H synthase-1 (PGHS-1) converts arachidonic acid into prostaglandin endoperoxides

Table 1 Natural Antithrombotic Mechanisms of Endothelial Cells

Regulation of clotting cascade	Tissue factor pathway inhibitor Antithrombin III Protein C/Protein S
Modulation of vessel and platelet reactivity	Prostacyclin Nitric oxide
Inhibition of platelet recruitment	Ecto-ADPase (CD39)
Removal of fibrin clot	Fibrinolysis

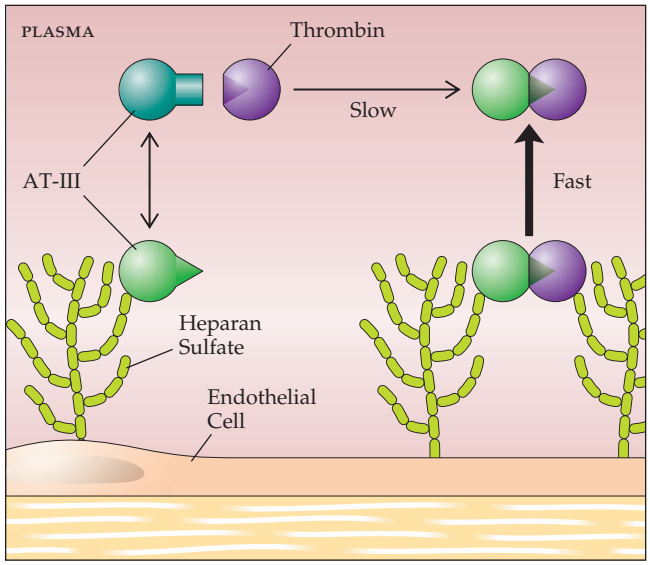


Figure 5 **In the absence of heparan sulfate (HS), antithrombin III (AT-III) inhibits thrombin slowly. When HS is present, it binds to a specific site on AT-III that causes a conformational change in AT-III, allowing it to reach the active site of thrombin and inhibit the enzyme instantaneously. HS also binds to a specific site on thrombin, positioning it for optimal inhibition by AT-III.**

and finally to thromboxane A_2 (TXA_2) in platelets and prostacyclin (PGI_2) in endothelial cells. TXA_2 and PGI_2 have opposite functions. TXA_2 is a potent stimulator of platelet aggregation and causes vasoconstriction, whereas PGI_2 inhibits platelet aggregation and induces vasodilatation. PGI_2 functions by activating adenylate cyclase, which leads to an increase in intracellular cAMP [*see Figure 8*].

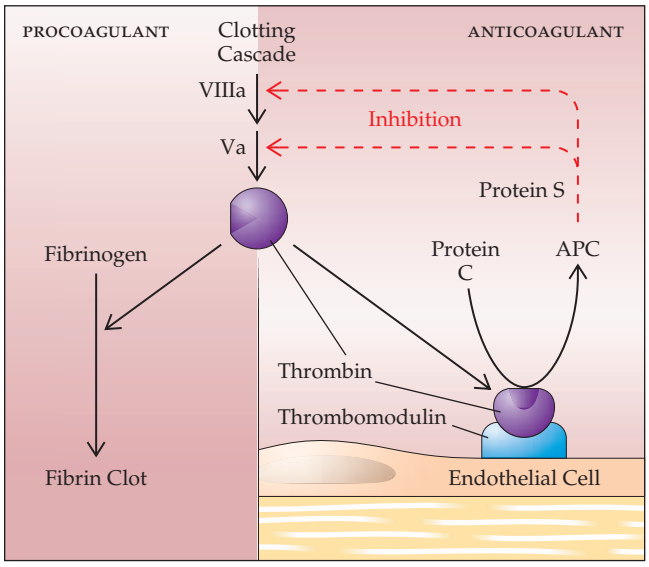


Figure 6 **The protein C/protein S pathway is complementary to the AT-III pathway. When thrombin binds to thrombomodulin, thrombin undergoes a conformational change and no longer clots fibrinogen or activates platelets. However, it acquires the ability to activate protein C in plasma. Protein S serves as a cofactor for activated protein C. Activated protein C degrades activated factors V and VIII, the two cofactors in the clotting cascade.**

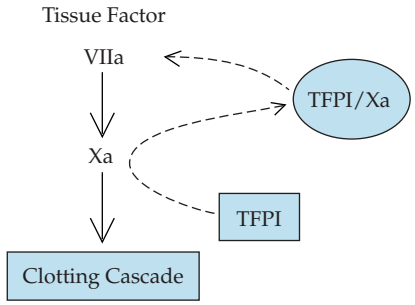


Figure 7 **Tissue factor pathway inhibitor (TFPI) binds to and inhibits factor Xa. After binding to factor Xa, TFPI undergoes a conformational change. The TFPI/factor Xa complex then mediates feedback inhibition of tissue factor/factor VIIa.**

Cyclooxygenase-1 and Cyclooxygenase-2

Cyclooxygenase-1 (COX-1) is the constitutive isoform of PGHS. Cyclooxygenase-2 (COX-2) is an inducible isoform of PGHS. COX-2 is undetectable in most tissues. However, it can be rapidly induced in response to growth factors, endotoxins, and cytokines in endothelial cells and monocytes (although not in platelets).[14] Recent evidence indicates that endothelial COX-2 is a major source of PGI_2 under physiologic conditions in humans, perhaps because of continual COX-2 induction by hemodynamic shear in the circulation.[15] Aspirin acetylates and irreversibly inhibits both COX-1 and COX-2. Other nonsteroidal anti-inflammatory drugs (NSAIDs) also inhibit COX-1 and COX-2, although not permanently. Selective COX-2 inhibitors are now available as a new generation of NSAIDs.[16]

Because aspirin irreversibly inhibits COX-1 and because platelets cannot make new COX-1, brief exposure to aspirin will permanently inhibit TXA_2 production for the life span of affected platelets.

NITRIC OXIDE

Nitric oxide (NO) is formed from L-arginine in endothelial cells. NO stimulates guanylate cyclase, leading to an increase in cyclic guanosine monophosphate (cGMP) in target cells; causes vasodilatation; and inhibits platelet adhesion and aggregation [*see Figure 8*].[17] NO is rapidly destroyed by hemoglobin and thus functions as a local (i.e., paracrine) hormone. Intravenous infusion of an arginine analogue that blocks NO production leads to an immediate and substantial rise in blood pressure. This phenomenon suggests that NO is released continually and basally to regulate vascular tone (in contrast to the production of PGI_2, which is more stimulus-responsive). There is significant synergism between NO and PGI_2. Formation of NO is catalyzed by NO synthases, which exist in different isoforms in various tissues. In addition to regulating vascular events, NO has a wide range of biologic effects (e.g., neurotransmittal function in the central nervous system).

ECTO-ADPase (CD39)

CD39 is an integral membrane protein found on the endothelial cell surface. It is an active enzyme that rapidly hydrolyzes ADP to AMP, thus functioning as a cell-bound ecto-ADPase. It limits the recruitment of additional platelets into the growing platelet plug by removing ADP released from the dense granules of activated platelets and from damaged erythrocytes and endothelial cells.[18]

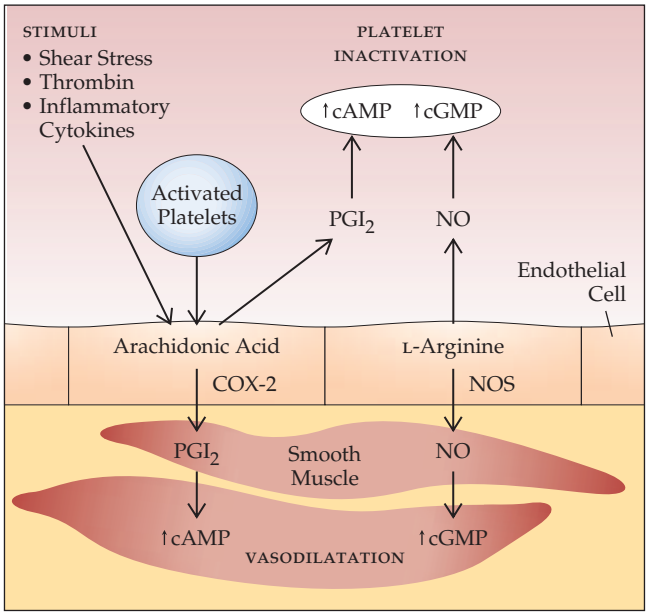


Figure 8 **Significant synergism exists between nitric oxide (NO) and prostacyclin (PGI_2), leading to platelet inactivation and vasodilatation. The enzyme prostaglandin endoperoxide H synthase–1 (PGHS-1) converts arachidonic acid into PGI_2 in endothelial cells. PGI_2 activates adenylate cyclase, which leads to an increase in intracellular cyclic adenosine monophosphate (cAMP), inhibiting platelet aggregation and inducing vasodilatation. NO, formed from L-arginine, stimulates production of cyclic guanosine monophosphate (cGMP). Cyclooxygenase-2 (COX-2) is the induced isoform of PGHS; its formation presumably results from hemodynamic shear in the circulation. NO formation is catalyzed by NO synthases (NOS).**

FIBRINOLYSIS

Tissue plasminogen activator (t-PA) is released from perturbed endothelial cells near the site of vascular injury. t-PA converts plasminogen to plasmin. Like the AT-III interaction with thrombin, which is accelerated in the presence of endothelial cell surface heparan sulfate, generation of plasmin takes place optimally on a surface (in this case, the fibrin clot). Both t-PA and plasminogen bind to fibrin (via recognition of lysine residues), which facilitates plasmin generation and localized fibrinolysis [*see Figure 9*].

Plasmin cleaves the polymerized fibrin strand at multiple sites, releasing fibrin degradation products. One of the major fibrin degradation products is D-dimer, which consists of two D domains from adjacent fibrin monomers that have been cross-linked by activated factor XIII [*see Figure 10*]. Plasmin has a broad substrate specificity and, in addition to fibrin, cleaves fibrinogen and a variety of plasma proteins and clotting factors. Plasmin bound on the fibrin clot is protected from inactivation, whereas plasmin released into the circulation is rapidly inactivated by plasma α_2-antiplasmin. Thus, localized fibrinolysis is achieved, but nonspecific plasmin degradation of plasma proteins is prevented. In rare cases, patients have bleeding problems caused by a congenital deficiency in α_2-antiplasmin.

Urokinase is the second physiologic plasminogen activator. It is present in high concentration in the urine. Although t-PA is largely responsible for initiating intravascular fibrinolysis, urokinase is the major activator of fibrinolysis in the extravascular compartment. Urokinase is secreted by many cell types in the form of prourokinase, also termed single-chain urokinase-type plasminogen activator (scu-PA). Prourokinase is converted to urokinase by plasmin. Urokinase lacks fibrin specificity in converting plasminogen to plasmin, whereas prourokinase displays such specificity.

The major physiologic inhibitor of t-PA and urokinase plasminogen activator (u-PA) is plasminogen activator inhibitor-1 (PAI-1).[19] Substantial amounts of PAI-1 are found in platelets. PAI-1 is also released from endothelial cells. PAI-1 deficiency is associated with bleeding diathesis, usually related to trauma or surgery.[20] A second inhibitor, PAI-2, is normally secreted by monocytes. During pregnancy, PAI-2 levels are greatly increased because of synthesis by the placenta. The biologic importance of PAI-2 remains to be established.

Thrombin-Activatable Fibrinolysis Inhibitor

Plasma carboxypeptidase is a newly recognized thrombin-activatable fibrinolysis inhibitor (TAFI) [*see Figure 11*].[21,22] TAFI is the second known physiologic substrate for the thrombin-thrombomodulin complex. One may envisage that after the initial fibrin clot is formed by thrombin at the site of a vascular wound, thrombin binds to thrombomodulin on the nearby intact endothelial surface. The thrombomodulin-bound thrombin leads to the generation of activated protein C, which dampens the clotting cascade and prevents excessive thrombin generation. At the same time, the thrombomodulin-bound thrombin activates TAFI, thus slowing down the lysis of the existing clot. In hemophilia, the decreased generation of thrombin may lead to suboptimal activation of TAFI and result in premature clot lysis, which contributes to the delayed bleeding observed in these patients.[22] Whether excessive TAFI activity leads to thrombosis is unknown at present.

Overview of Blood Coagulation

The clotting cascade is initiated by the exposure of tissue factor at a vascular wound, which leads to the generation of thrombin and the deposition of a fibrin clot [*see Figure 12*]. Simultaneously, the damaged endothelium releases t-PA, which converts plasminogen to plasmin, which then lyses the clot.

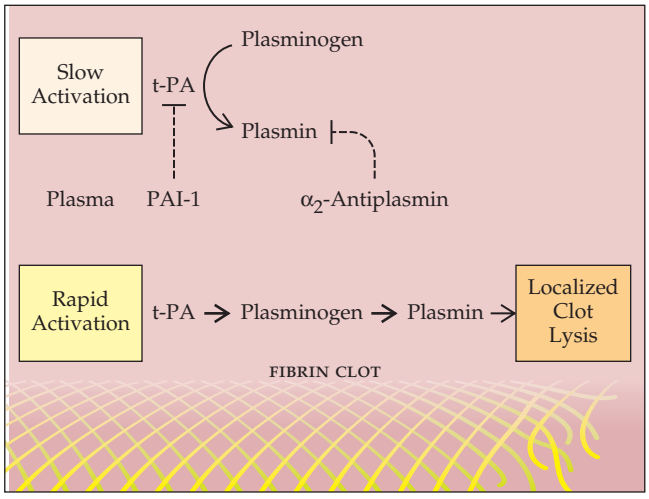


Figure 9 **Tissue-type plasminogen activator (t-PA), released from perturbed endothelial cells near an injured blood vessel, converts plasminogen to plasmin. Free plasmin is rapidly inactivated by plasma α_2-antiplasmin; plasmin bound to the fibrin clot is protected from inactivation.**

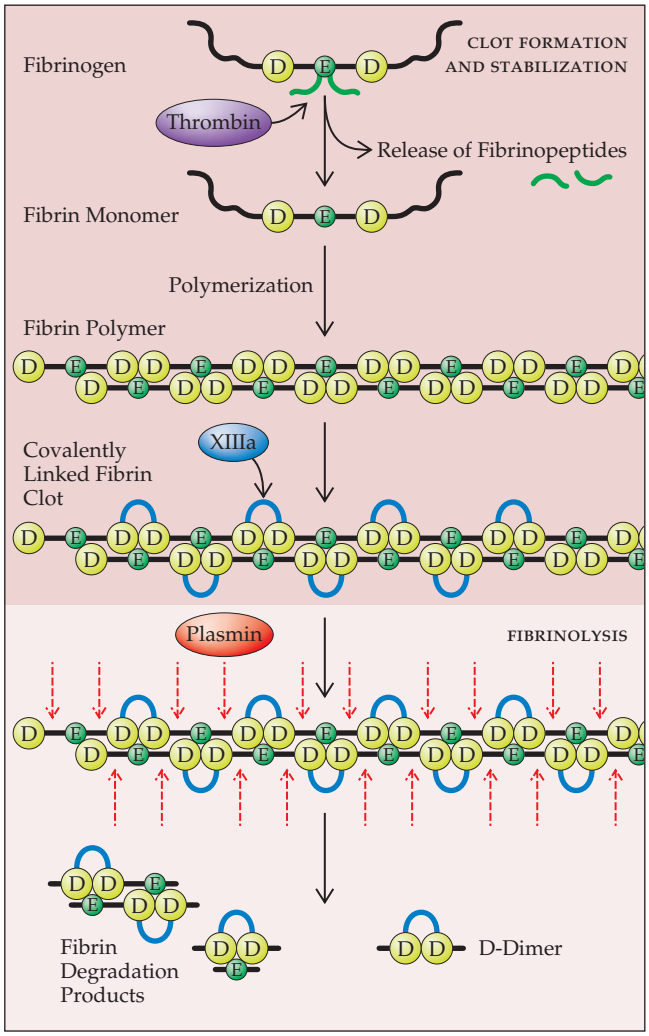


Figure 10 **The transformation of fibrinogen to fibrin is initiated by thrombin cleavage of fibrinopeptides A and B from the E domains of fibrinogen to form fibrin monomer. The cleavage apparently changes the overall negative charge of the E domain to a positive charge. This change in charge permits the spontaneous polymerization of fibrin monomers, because the positively charged E domain assembles with the negatively charged D domains of other monomers. The polymer is initially joined by hydrogen bonds. Thrombin activates factor XIII, which catalyzes the formation of covalent bonds between adjacent D domains in the fibrin polymer. Plasmin cleaves the polymerized fibrin strand at multiple sites and releases fibrin degradation products, including D-dimer.**

Both pathways are regulated: TF/factor VIIa is regulated by the TFPI/factor Xa complex, and thrombin is regulated by protein C and protein S. Similarly, the activity of t-PA is regulated by PAI-1. Thrombin and plasmin are under the control of their respective inhibitors, AT-III and α_2-antiplasmin. When these two pathways work in coordinated symmetry, a clot is laid down to stop bleeding, and clot lysis and tissue remodeling follow. Diminished thrombin generation (as in factor VIII deficiency) or enhanced plasmin production (as in α_2-antiplasmin deficiency) causes hemorrhage [*see 5:XIII Hemorrhagic Disorders*]. Conversely, excessive production of thrombin (as in AT-III or protein C deficiency) leads to thrombosis [*see 5:XIV Thrombotic Disorders*].

Heterogeneity of Endothelial Cells and Vascular Bed–Specific Hemostasis

Although the endothelium is generally considered to be a distinct, homogeneous organ system, there are significant differences between arterial, venous, and capillary endothelial cells in terms of morphology and disease susceptibility. Recent studies have shown distinct sets of proteins that mark the arterial and venous endothelial cells from the earliest stages of angiogenesis. Ephrin-B2, an Eph family transmembrane ligand, marks arterial but not venous endothelial cells. Conversely, Eph-B4, a receptor tyrosine kinase for ephrin-B2, marks veins but not arteries.[23]

It is also likely that endothelia from different vascular beds are not identical.[24] For example, the high endothelium in the postcapillary venules of lymph nodes and Peyer patches regulates the circulation of lymphocytes from blood to lymphatics and peripheral tissues. Specific adhesive protein receptors and matrix proteins are highly expressed in these high endothelial venules. The specialized endothelium representing the blood-brain barrier is another example.

These differences between arterial and venous endothelial cells and the vascular bed–specific endothelium may partly account for their different susceptibilities to thrombosis. For example, whereas AT-III and protein C deficiencies are usually associated with deep vein thrombosis of the lower extremities, thrombosis of portal and hepatic veins is frequently associated with myeloproliferative diseases.[25] In both conditions, the underlying defect is a systemic hypercoagulable state, and yet there is a clear predisposition of thrombosis to specific vascular beds. Thus, clinical thrombosis is attributable to an imbalance between systemic prothrombotic stimuli and local antithrombotic mechanisms [*see 5:XIV Thrombotic Disorders*].

Platelet Production and Thrombopoietin

Platelets are derived from megakaryocytes, which arise from pluripotent myeloid stem cells. Platelet production is controlled

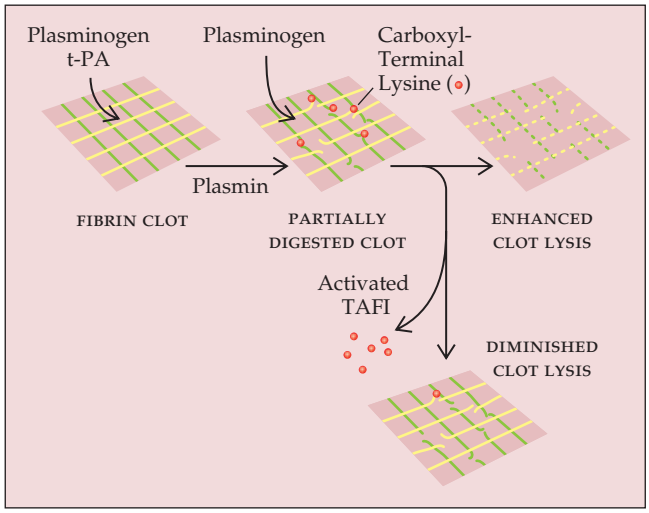


Figure 11 **Plasma carboxypeptidase is a thrombin-activatable fibrinolysis inhibitor (TAFI). When fibrin is degraded by plasmin, new carboxyl-terminal lysines are exposed in the partially digested clot. These lysines provide additional sites for plasminogen incorporation and activation in the clot, setting up a positive feedback loop in clot lysis. Thrombin activates carboxypeptidase-B in plasma, which removes the exposed carboxyl-terminal lysines and prevents further plasminogen incorporation into the clot.**

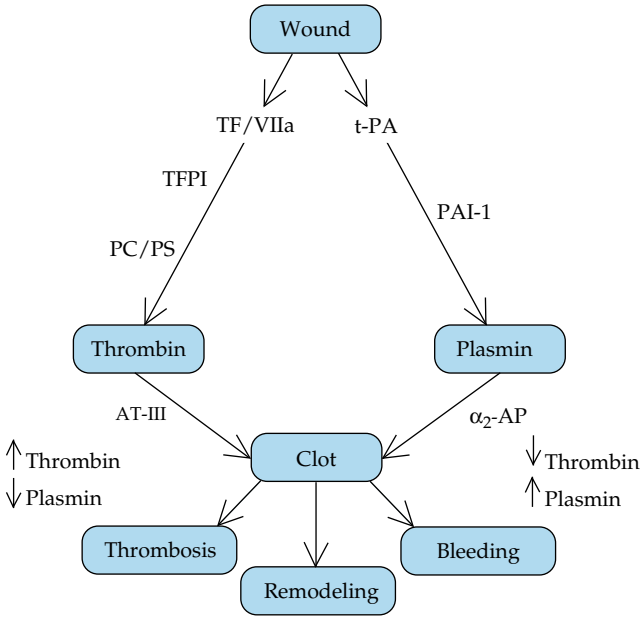


Figure 12 **Exposure of tissue factor at a vascular wound initiates the clotting cascade. Generation of thrombin and deposition of a fibrin clot occur simultaneously with release of t-PA from the damaged epithelium and conversion of plasminogen to plasmin. Plasmin then lyses the clot. When these two pathways work in coordinated symmetry, a clot is laid down to stop bleeding, and clot lysis and remodeling follow. (α_2-AP—α_2-antiplasmin; AT-III—antithrombin III; PAI-1—plasminogen activator inhibitor–1; PC/PS—protein C/protein S; TF—tissue factor; TFPI—tissue factor pathway inhibitor; t-PA—tissue-type plasminogen activator)**

by a thrombopoietin that is involved in the final maturation of the megakaryocyte. Thrombopoietin has multiple actions in megakaryocyte development.[26] It shares some structural homology with erythropoietin and is produced principally by the liver. It increases the size and number of megakaryocytes, stimulates the expression of platelet-specific markers, and is a potent megakaryocyte colony-stimulating factor. Although thrombopoietin is clearly a key factor, stem cell factor (also called kit ligand), interleukin-3 (IL-3), IL-6, and IL-11 all play contributory roles in controlling megakaryocytopoiesis.

Megakaryocytes undergo endomitosis, in which nuclear divisions occur without cell division and are followed by nuclear fusion, to yield a cell with a chromosomal content of 8n, 16n, or 32n. The megakaryocyte cytoplasm then changes into a series of thin, cylindrical strands that eventually fragment into small pieces of megakaryocytes, called proplatelets, that are released into the circulation. Megakaryocyte volume correlates with ploidy and cytoplasmic maturity; the largest megakaryocytes produce the greatest number of platelets. Large platelets called megathrombocytes are seen in the peripheral blood in thrombocytopenic states, especially in idiopathic thrombocytopenic purpura [*see 5:XIII Hemorrhagic Disorders*]. These megathrombocytes probably are young proplatelets and account for the increase in mean platelet volume that occurs during response to or recovery from acute thrombocytopenia.

Platelets entering the circulation survive about 8.5 to 10 days and have a half-life of about 4 days. Approximately 30% to 40% of the platelets are present in a splenic pool that can freely exchange with the circulation. When the need for platelets arises, production can increase sevenfold to eightfold. Because there is no marrow pool of platelets waiting to be released, meeting increased requirements for platelets may require a few days. Platelets have receptors for thrombopoietin and remove it from plasma, and the platelet mass functions as a major thrombopoietin regulator.[27] In states of megakaryocyte hypoplasia and thrombocytopenia, little thrombopoietin is metabolized and the plasma thrombopoietin level rises, leading to increased production of megakaryocytes and platelets. In the setting of thrombocytosis, thrombopoietin metabolism increases, lowering the plasma thrombopoietin level and decreasing platelet production.

Coagulation Tests and Their Use

TESTS OF COAGULATION CASCADE

Most coagulation tests measure the time required for fibrinogen from plasma to form fibrin strands, which can be detected by either optical or electrical devices. Prolongation may represent a low factor concentration, inactive factor or factors, or the presence of inhibitors.

Partial Thromboplastin Time

The partial thromboplastin time (PTT), sometimes termed the activated PTT (aPTT), tests the intrinsic coagulation system. A negatively charged surface (e.g., kaolin or silica), followed by cephalin, is added to whole plasma to activate factors XII and XI. The PTT is most sensitive to abnormalities and deficiencies in the sequence of the coagulation cascade before factor X activation. The PTT is also quite sensitive to the action of heparin. It is used to monitor and adjust anticoagulant therapy with regular heparin but not with low-molecular-weight heparins.

Prothrombin Time

The prothrombin time (PT) is a test of the extrinsic system. It detects deficiencies in fibrinogen, factor II (prothrombin), factor V, factor VII, and factor X. Tissue factor is added to whole plasma, leading to fibrin formation, normally in 9 to 12 seconds. Results are usually reported using the international normalized ratio (INR). The INR is calculated by using the following equation:

$$\text{INR} = (\text{Log patient PT} \,/\, \text{Log control PT})^{C}$$

where C represents the international sensitivity index (ISI). In this way, the thromboplastin used in an individual laboratory, with its specific ISI, is calibrated against a standard reference thromboplastin, and the PT is reported as an INR.[28] The presence of a lupus anticoagulant may also interfere with the PT.[29]

Dilute Russell Viper Venom Time

Russell viper venom contains an enzyme that activates factor X; therefore, the dilute Russell viper venom time (DRVVT) measures the common pathway of the clotting cascade. It is sensitive to the presence of a lupuslike anticoagulant that inhibits the phospholipid-dependent prothrombinase complex.

Thrombin Time

The thrombin time (TT) is used to test abnormalities of the conversion of fibrinogen to fibrin. It can be prolonged because of hypofibrinogenemia, abnormal fibrinogen (dysfibrinogen), or the presence of inhibitors (e.g., fibrin degradation products) that interfere with fibrin polymerization. The clinical factors com-

monly associated with prolonged TT are severe liver disease, disseminated intravascular coagulation, and heparin therapy.

Reptilase Time

Reptilase is a thrombinlike enzyme that converts fibrinogen to fibrin. The reptilase time (RT) is prolonged under conditions similar to those for prolonged TT, with one significant difference: reptilase is not inhibited by the AT-III–heparin complex. Therefore, RT is not prolonged by heparin. A long thrombin time and normal RT suggest a heparin effect.

Fibrinopeptide A

Thrombin activates fibrinogen by splitting off two peptides, fibrinopeptide A (FPA) and FPB, from the Aα and Bβ chain of fibrinogen and converting fibrinogen to fibrin monomer. Measurement of FPA in the blood can be used as an index of thrombin activity in vivo. Because the clotting cascade can be activated during the blood-sample collection, however, precautions are required in the measurement and interpretation of FPA levels.

Fibrinogen

The fibrinogen level in plasma can be measured either antigenically or more commonly by clotting assays. The results are reported in mg/dl.

D-Dimer and Fibrin-Fibrinogen Degradation Products

Fibrinogen degradation products (FDP) and fibrin-fibrinogen split products (FSP) result from plasmin degradation of fibrinogen and fibrin clot [*see Figure 9*]. D-dimer is released by the plasmin-mediated degradation of fully polymerized fibrin. Plasmin cleavage of fibrinogen or soluble fibrin monomer does not yield the D-dimer. Thus, elevated D-dimer is a specific measure of intravascular fibrin deposition and plasmin degradation characteristic of disseminated intravascular coagulation. The D-dimer test has largely replaced the FSP test.

Factor XIII

Factor XIII is the only clotting factor whose activity is not assessed in PT or PTT because the end point for both tests is the formation of fibrin polymers, irrespective of whether these polymers are cross-linked covalently by activated factor XIII. Factor XIII deficiency may be suspected in an infant who has significant bleeding after circumcision or, more rarely, in an adult patient who has unexplained bleeding.

Plasminogen and α_2-Antiplasmin

The activation of the plasminogen-plasmin system can be inferred from the findings of a long TT, a low plasma fibrinogen level, and an elevated D-dimer level. Another crude test used to measure plasminogen-plasmin activation is the euglobulin lysis time. The sensitivity and specificity of this test is not well defined, however. During extensive thrombosis and fibrinolysis, both plasminogen and α_2-antiplasmin (the physiologic inhibitor of plasmin) are consumed. The direct measurement of plasma levels of plasminogen and α_2-antiplasmin is sometimes useful to assess the extent of fibrinolysis and the requirement for replenishment of these plasma proteins using fresh frozen plasma.

TESTS OF PLATELETS AND OF PLATELET FUNCTION

Peripheral Blood Smear Evaluation

This examination provides quick, definitive information to confirm or question a platelet count. Normally, there are eight to 12 platelets per high-power field (1,000 × magnification), corresponding to a normal platelet count of 150,000 to 300,000/ml. The smear also shows platelet granularity and whether megathrombocytes are present.

Bleeding Time

This test primarily measures platelet function. A spring-loaded device is used to make a standard skin incision on the forearm. A prolonged bleeding time with platelets greater than 100,000/ml suggests impaired function. The bleeding time is difficult to standardize, and a normal bleeding time does not predict the safety of surgical procedures or accurately predict hemorrhage.[30] It should not be used as a general screening test in a preoperative setting. Although once used in the screening of patients for von Willebrand disease or certain platelet function disorders, for these purposes bleeding time has been largely replaced by the platelet function–100 assay (PFA-100).

Platelet Function Assay–100

PFA-100 is a newly developed automated test for platelet function. Citrated whole blood is aspirated through a capillary tube under high shear onto a membrane coated with collagen and epinephrine or collagen and ADP in which a central aperture is made. The time it takes for blood flow through the membrane to stop is denoted as closure time and is a measure of platelet function. The closure time is prolonged in patients with von Willebrand disease or other platelet functional defects.[31] PFA-100 should be considered the first-line test for platelet function disorders.

Platelet Aggregometry

Platelet aggregometers are photometric devices for recording the transmission of light through a suspension of platelets. When platelets aggregate, light passes through the suspension more readily. To test aggregation, dilute concentrations of platelet agonists (e.g., ADP, epinephrine, collagen, and ristocetin) are added to citrated platelet-rich plasma. With the weak agonists, such as ADP and epinephrine, the initial primary wave of aggregation is followed by a secondary wave. The secondary wave reflects the induction of the platelet release reaction, in which platelet granule contents are released to augment further platelet aggregation. A suboptimal secondary wave is seen with platelet storage pool defects in which either platelet granule content is diminished or its release activity is impaired. The latter is commonly associated with aspirin intake or uremia-related thrombocytopathy. Patients with von Willebrand disease will have a suboptimal platelet aggregation response to ristocetin but a normal response to the other agonists. Platelet aggregation testing is labor intensive and expensive and should be performed only in clinical coagulation laboratories that do this test regularly.

TESTS OF INHIBITORS OF HEMOSTASIS

Mixing Studies

A prolonged clotting time (e.g., PTT of 60 seconds [normal, 28 to 30 seconds]) can be caused by either a clotting factor deficiency or an inhibitor. An inhibitor is generally an antibody directed against a specific clotting factor or against a phospholipid-protein complex, the so-called lupus anticoagulant [*see 5:XIV Thrombotic Disorders*]. In a mixing study, one volume of a patient's plasma is mixed with an equal volume of normal plasma. The resulting mixture will provide at least 50% of a deficient factor and

correct the abnormality. If the problem is caused by an inhibitor, the resulting plasma mixture still has a prolonged clotting time. A mixing study should always be done when a prolonged clotting time is noted.

Antithrombin III

Bioassays and immunoassays are available for assessing AT-III activity. A functional assay is preferable to an antigenic assay.

Protein C and Protein S

Functional and immunologic methods are available. Because protein C and protein S are vitamin K dependent, their measurement can be problematic in patients taking warfarin. It is best to measure protein C or protein S when the patient has been off warfarin for 3 to 4 weeks.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Coughlin SR: How thrombin 'talks' to cells: molecular mechanisms and roles in vivo. Arterioscler Thromb Vasc Biol 18:514, 1998

2. Kahn ML, Zheng YW, Huang W, et al: A dual thrombin receptor system for platelet activation. Nature 394:690, 1998

3. Woulfe D, Yang J, Brass L: ADP and platelets: the end of the beginning (commentary). J Clin Invest 107:1503, 2001

4. Clemetson KJ: Platelet GPIb-V-IX complex. Thromb Haemost 78:266, 1997

5. Sixma JJ, Zanten HV, Huizinga EG, et al: Platelet adhesion to collagen: an update. Thromb Haemost 78:434, 1997

6. Shattil SS, Kashiwagi H, Pampori N: Integrin signaling: the platelet paradigm. Blood 91:2645, 1998

7. Rapaport SI, Rao LVM: The tissue factor pathway: how it has become a "prima ballerina". Thromb Haemost 74:7, 1995

8. Furie B, Bouchard BA, Furie BC: Vitamin K-dependent biosynthesis of γ-carboxyglutamic acid. Blood 93:1798, 1999

9. Marcum JA, McKenney JB, Rosenberg RD: Acceleration of thrombin-antithrombin complex formation in rat hindquarters via heparinlike molecules bound to the endothelium. J Clin Invest 74:341, 1986

10. Esmon CT: The roles of protein C and thrombomodulin in the regulation of blood coagulation. J Biol Chem 264:4743, 1989

11. Esmon CT, Ding W, Yasuhiro K, et al: The protein C pathway: new insights. Thromb Haemost 78:70, 1997

12. Morten S, Bendz B: Tissue factor pathway inhibitor: clinical deficiency states. Thromb Haemost 78:467, 1997

13. Bajaj MS, Bajaj SP: Tissue factor pathway inhibitor: potential therapeutic applications. Thromb Haemost 78:471, 1997

14. Smith WL, Garavito RM, DeWitt DL: Prostaglandin endoperoxide H synthases (cyclooxygenases)-1 and -2. J Biol Chem 271:33157, 1996

15. McAdam BF, Mardini IA, Kapoor S, et al: Systemic biosynthesis of prostacyclin by cyclooxygenase (COX)-2: the human pharmacology of a selective inhibitor of COX-2. Proc Natl Acad Sci USA 96:272, 1999

16. Hawkey CJ: COX-2 inhibitors. Lancet 353:307, 1999

17. Moncada S, Higgs A: The L-arginine-nitric oxide pathway. N Engl J Med 329:2002, 1993

18. Marcus AJ, Broekman JM, Drosopoulos J, et al: Thromboregulation by endothelial cells (brief reviews). Arterioscler Thromb Vasc Biol 21:178, 2001

19. Van Meijer M, Pannekoek H: Structure of plasminogen activator inhibitor 1 (PAI-1) and its function in fibrinolysis: an update. Fibrinolysis 9:263, 1995

20. Fay WP, Parker AC, Condrey LR, et al: Human plasminogen activator inhibitor-1 (PAI-1) deficiency: characterization of a large kindred with a null mutation in the PAI-1 gene. Blood 90:204, 1997

21. Bajzar L, Morser J, Nesheim M: TAFI, or plasma procarboxypeptidase B, couples coagulation and fibrinolytic cascades through the thrombin-thrombomodulin complex. J Biol Chem 271:16603, 1996

22. Broze GJ, Higuchi DA: Coagulation-dependent inhibition of fibrinolysis: role of carboxypeptidase-U and the premature lysis of clots from hemophilic plasma. Blood 88:3815, 1996

23. Wang HU, Chen ZF, Anderson DJ: Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4. Cell 93:741, 1998

24. Garlanda C, Dejana E: Heterogeneity of endothelial cells: specific markers. Arterioscler Thromb Vasc Biol 17:1193, 1997

25. Dilawari JB, Bambery P, Chawla Y, et al: Hepatic outflow obstruction (Budd-Chiari syndrome): experience with 177 patients and a review of the literature. Medicine (Baltimore) 73:21, 1994

26. Kaushansky K: Thrombopoietin. N Engl J Med 339:746, 1998

27. Kuter DJ, Rosenberg RD: The reciprocal relationship of thrombopoietin (c-Mpl ligand) to changes in the platelet mass during busulfan-induced thrombocytopenia in the rabbit. Blood 85:2720, 1995

28. Hirsh J, Dalen JE, Anderson DR, et al: Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. Chest 114(suppl):445S, 1998

29. Moll S, Ortel TL: Monitoring warfarin therapy in patients with lupus anticoagulants. Ann Intern Med 127:177, 1997

30. The bleeding time (editorial). Lancet 337:1447, 1991

31. Fressinaud E, Veyradier A, Truchaud F, et al: Screening for von Willebrand disease with a new analyzer using high shear stress: a study of 60 cases. Blood 91:1325, 1998

Acknowledgments

Figures 1, 2, 3, 5, 6, and 8 through 11 Seward Hung.

Figures 4, 7, and 12 Marcia Kammerer.

XIII PLATELET AND VASCULAR DISORDERS

Lawrence L. K. Leung, m.d.

Bleeding or bruising that is spontaneous or excessive after tissue injury may be caused by abnormal platelet number or function, abnormal vascular integrity, coagulation defects, fibrinolysis, or a combination of these abnormalities. This chapter addresses hemorrhagic disorders associated with quantitative or qualitative platelet abnormalities and disorders associated with blood vessels. Hemorrhagic disorders associated with abnormalities in coagulation (e.g., von Willebrand disease and hemophilia) are covered elsewhere [*see 5:XV Coagulation Disorders*].

Approach to the Patient with a Bleeding Disorder

A bleeding disorder may be suspected when a patient complains of excessive bruising or bleeding that often occurs secondary to trauma. The clinical evaluation of a patient with a suspected bleeding disorder begins with a careful history. Assessment of the presenting complaint may suggest where in the hemostatic process a defect is located and whether the defect is inherited or acquired—information that contributes to a rational approach to laboratory evaluation.

PATIENT HISTORY

Bleeding History

Patients suspected of having a bleeding disorder should be questioned about past bleeding problems, bleeding outcomes after surgeries and tooth extractions, character of menses, and dietary habits that might predispose to deficiencies of vitamin K, vitamin B_{12}, and folic acid. The patient should also be questioned about sexual activity, anemia, transfusions, recurrent infections, connective tissue diseases, malignancies, liver and kidney diseases, immunocompromised states, and drug use [*see* Medication History, *below*].

Many healthy people consider their bleeding and bruising to be excessive, whereas patients with underlying von Willebrand disease, the most common hereditary bleeding disorder, often fail to identify their bleeding symptoms.[1] Given the variability in patients' perceptions of bleeding, as well as the lack of a uniform clinical measure of bleeding severity, a dialogue between the patient and the physician is essential for the evaluation of a bleeding disorder. It is therefore necessary to ask for specific information from patients about bleeding and bruising: (1) If the patient is easily bruised, what size are the bruises? (2) If the patient has had surgery, were blood transfusions needed? (3) If the patient had a wisdom tooth extracted, were return visits required for packing, suturing, or transfusion? The response to trauma is an excellent screening test. A history of surgical procedures, tooth extractions, or significant injury without abnormal bleeding is good evidence against the presence of an inherited hemorrhagic disorder.

The type of bleeding is informative and may suggest the underlying disorder [*see Table 1*]. Active bleeding can be caused by a localized anatomic lesion or an underlying bleeding diathesis. Mucosal bleeding, with recurrent epistaxis, gum bleeding, ecchymoses, and menorrhagia, is suggestive of von Willebrand disease or other platelet disorders. Deep-tissue bleeding (e.g., hemarthrosis and painful muscle hematomas) is more commonly seen in hemophilia and clotting factor deficiencies. Patients with clotting factor deficiencies may have delayed bleeding, presumably because the initial platelet thrombus provides immediate hemostasis but is not properly stabilized by the fibrin clot.

Medication History

A careful history of medication use is a critical aspect of the diagnostic evaluation. The patient should be questioned about use of recreational drugs, prescribed medications, over-the-counter medications, and herbal products. Aspirin use is of particular importance. Aspirin can partially impair platelet function and trigger bleeding symptoms in a patient with mild underlying von Willebrand disease. Because several hundred drug formulations contain aspirin (often with no indication of aspirin content in the product name), identification of aspirin as the cause of a hemorrhagic disorder can be difficult.

LABORATORY EVALUATION

The laboratory evaluation begins with general screening tests, such as platelet count, bleeding time (BT), prothrombin time (PT), activated partial thromboplastin time (aPTT), and thrombin time (TT). These tests are supplemented with specific tests that define platelet or clotting factor abnormalities. Specific tests include examination of the peripheral blood smear; platelet aggregation in response to adenosine diphosphate (ADP), epinephrine, collagen, and ristocetin; platelet release assays; coagulation factor assays; and assessment of factor XIII activity via clot solubility testing [*see 5:XII Hemostasis and Its Regulation*].

In many patients with a bleeding disorder, the likely diagnosis will be suggested from the history and physical examination; the diagnosis can then be confirmed with the appropriate specific tests. When the diagnosis is not immediately apparent, three initial tests should be performed: platelet count, PT, and aPTT. The pattern of results provided by these tests suggests a diagnosis that can then be confirmed with specific testing [*see Table 2*].

Both the PT and aPTT provide a global assessment of the clotting cascade: the PT measures the extrinsic pathway, and the aPTT measures the intrinsic pathway [*see 5:XII Hemostasis and Its Regulation*]. Prolongations of both the PT and the aPTT suggest a clotting defect in the final common portion of the cascade that involves either factor X, factor V, prothrombin, or fibrinogen.

A prolonged PT with a normal aPTT is most commonly seen in a patient taking warfarin; in the absence of warfarin, these test results will indicate either factor VII deficiency or, more rarely, an inhibitor against factor VII.

A prolonged aPTT with a normal PT has a broader differential diagnosis. This combination of test results may denote the presence of an inhibitor against a clotting factor or a deficiency of one of the clotting factors in the intrinsic pathway. It is important to perform a repeat aPTT with equal volumes of the patient's plasma and normal plasma (a mixing study). If the normal plasma does not correct the prolonged aPTT, an inhibitor exists (e.g., a lupuslike anticoagulant or an inhibitor directed against a specific clotting factor). If the normal plasma corrects the prolonged aPTT, the patient has a clotting factor deficiency involving factor XII, factor XI, factor VIII, factor IX, or, more rarely, prekallikrein

Table 1 Clinical Manifestations of Hemorrhagic Disorders

Manifestation	*Bleeding Disorder*	
	Platelet Defect	*Clotting Factor Deficiency*
Site of bleeding	Skin, mucous membranes (gingivae, nares, genitourinary tract)	Deep in soft tissues (joints, muscles)
Petechiae	Present	Absent
Ecchymoses	Small, superficial	Large, palpable
Hemarthroses, muscle hematomas	Rare	Common
Bleeding after minor cuts	Common	Rare
Bleeding after surgery	Immediate, mild	Delayed, severe

or high-molecular-weight kininogen. Because the clinical presentations of these clotting factor deficiencies are quite different (e.g., factor VIII and factor IX deficiencies are X linked, frequently with a positive family history), correlation with the clinical setting should be sought and the specific clotting factor levels subsequently determined [*see 5:XV Coagulation Disorders*]

Thrombocytopenia

Thrombocytopenia, a decreased platelet count, is a common clinical finding that may be caused by decreased platelet production or accelerated platelet removal. Accelerated platelet removal may result from immunologic mechanisms, nonimmunologic mechanisms, or sequestration of platelets in the spleen [*see Table 1*]. Thrombocytopenia can range from a transient, isolated finding to a severe, life-threatening condition.

DIAGNOSTIC EVALUATION

Clinical Manifestations

Patients with thrombocytopenia may be asymptomatic; in these patients, the finding of a low platelet count may be first detected on a routine complete blood count. The most common symptomatic presentation of thrombocytopenia is bleeding—characteristically, mucosal and cutaneous. The hallmark of thrombocytopenia is nonpalpable petechiae, which reflect bleeding probably from capillaries or postcapillary venules [*see Table 1*]. Petechiae usually are only a few millimeters in diameter and occur at sites of increased intravascular pressure, such as over the lower extremities and on the oral mucosa, and at sites constricted by certain types of clothing, such as brassiere straps. Purpura, more extensive subcutaneous bleeding, may occur with a confluence of petechial lesions. Palpable purpura indicates an additional component of vascular inflammation and suggests underlying systemic vasculitis, such as cryoglobulinemia. Thrombocytopenia also leads to mucosal bleeding; deep-tissue bleeding is less common.

Laboratory Tests

Blood count and peripheral smear Laboratory examination should start with the complete blood count and examination of the peripheral smear. The importance of examination of the peripheral smear for estimation of platelet numbers, morphology, and the presence or absence of platelet clumping, as well as evaluation of associated white and red blood cell changes, cannot be overemphasized.

Normally, there are eight to 12 platelets per high-power field (×1,000 magnification), corresponding to a normal platelet count of 15,000 to 30,000/μl. There is no clearly demarcated level of platelets above which patients can be considered safe from bleeding. In general, a platelet count greater than 20,000/μl is considered safe; platelet counts of 10,000/μl or below may be tolerated in nonsurgical patients [*see 5:X Transfusion Therapy*]. Patients with idiopathic thrombocytopenic purpura bleed less at a given platelet level than patients with aplastic anemia [*see* Idiopathic Thrombocytopenic Purpura, *below*]. Presumably, the larger, younger platelets are more effective in hemostasis. The risk of intracranial hemorrhage usually directs therapy.

Elderly patients and patients with coexistent illnesses bleed more than young patients and patients with thrombocytopenia alone. An associated disorder, such as liver dysfunction or connective tissue disease, increases the risk of serious bleeding.

In the initial laboratory evaluation, the complete blood count will establish whether the thrombocytopenia is a single disorder or is associated with anemia or leukopenia, which suggests a production defect as the underlying cause [*see* Platelet Production Defects, *below*]. If thrombocytopenia is an isolated finding, the physician should confirm the platelet count by repeating the complete blood count. A falsely low platelet count can be the result of in vitro platelet clumping caused by the presence of cold-dependent or ethylenediamenetetraacetic acid–dependent agglutinins. Examination of the blood smear and a repeat platelet count in a citrated or heparin-anticoagulated blood sample will resolve this problem.[2]

The peripheral smear may reveal morphologic abnormalities in platelets and indicate the presence of polychromatophilia, neutropenia, lymphopenia, spherocytosis, blastomycosis, or fragmented microangiopathic erythrocytes. The mean platelet volume, as determined by automated blood cell counters, may provide an additional clue to the cause of the thrombocytopenia. Low platelet volumes (< 6.4 femtoliters) suggest poor production, whereas larger volumes suggest rapid platelet regeneration or dysplastic platelet production.

Bone marrow aspirate and biopsy When accelerated platelet removal appears to be the cause of the patient's thrombocytopenia, a rapid differential diagnosis should be made [*see Table 3*]. A bone marrow aspirate and biopsy will be very helpful in narrowing the diagnosis. Usually, thrombocytopenia with an abundance of normal megakaryocytes in the marrow is the result of accelerated platelet removal.[3] Normally, platelets survive for 10 days and have a half-life of about 4 days; in accelerated-removal states, such as idiopathic thrombocytopenic purpura, the

Table 2 Typical Results of Tests for Hemostatic Function in Bleeding Disorders

Disorder	*PC*	*PT*	*aPTT*
Thrombocytopenia	Low	Normal	Normal
Platelet function abnormalities	Normal	Normal or low	Normal
Vascular purpuras	Normal	Normal	Normal
von Willebrand disease	Normal	Normal	Long
Hemophilia A	Normal	Normal	Long
Disseminated intravascular coagulation	Low	Long	Long

aPTT—activated partial thromboplastin time PC—platelet count
PT—prothrombin time

platelet half-life may be as short as 30 to 60 minutes. The platelet count will then reflect the balance between accelerated platelet removal and compensatory megakaryopoiesis.

Platelet survival studies are not generally available and are not usually necessary to determine whether accelerated platelet removal is occurring. Infusion of random-donor platelets can be used as a diagnostic and therapeutic procedure. When accelerated platelet removal is responsible for the thrombocytopenia, transfusion with six platelet packs only slightly elevates the platelet count, which then returns to baseline values in less than 24 hours. This therapeutic test becomes unreliable, however, if the patient has been previously alloimmunized by blood or platelet transfusions or by multiple pregnancies.

PLATELET PRODUCTION DEFECTS

Inadequate Platelet Production Due to Stem Cell Destructon

Disorders that injure stem cells or prevent their proliferation frequently cause thrombocytopenia. These disorders affect multiple hematopoietic cell lines, and the resulting thrombocytopenia is accompanied by varying degrees of anemia and leukopenia.

Diagnosis Diagnosis of a platelet production defect is readily established by examination of a bone marrow aspirate and biopsy. The finding of a hypoplastic marrow in which the total cellularity is reduced, along with a decrease in megakaryocytes, implies aplastic or hypoplastic anemia. The first presumption of a cause in these cases is drug toxicity. A marrow that is fibrosed or infiltrated with leukemic or other malignant cells represents the syndrome of pancytopenia from infiltrated marrow.

A marrow aspirate and biopsy sample showing normal cellularity and normal maturation of the erythroid and myeloid precursors, with decreased numbers of apparently normal megakaryocytes, suggest that the patient has ingested a drug, such as ethanol, that specifically affects the megakaryocytic progenitor cells.[4] Ethanol also produces ineffective megakaryopoiesis. In vitamin B_{12} deficiency and folate deficiency, all three marrow cell lines are affected. The marrow smear shows many large hyperlobated megakaryocytes. Some myeloproliferative disorders are characterized by ineffective megakaryopoiesis with bizarre binucleate megakaryocytes.

TREATMENT

If a drug is the suspected cause of the thrombocytopenia, it should be discontinued. Specific replacement is required for deficiencies of vitamin B_{12} and folate. When the thrombocytopenia is causing significant bleeding, platelet transfusion will be required until the situation resolves [*see 5:X Transfusion Therapy*].

Interleukin-11 (IL-11), which plays a contributory role in megakaryopoiesis, has been approved for secondary prophylaxis against thrombocytopenia after chemotherapy[5]; however, it has limited efficacy and is associated with moderate toxicity.[6] Two forms of recombinant thrombopoietin—one full length and one with a truncated form—have undergone extensive clinical trials. Both types are potent stimulators of megakaryocyte growth and platelet production and are effective in reducing the thrombocy-

Table 3 Causes of Thrombocytopenia

Type	*Disorder*	*Cause*
Platelet production defect	Marrow aplasia or hypoplasia, pancytopenia	Radiation, cytotoxic drugs, idiopathic
	Marrow infiltration, pancytopenia	Cancer (leukemia, lymphoma), fibrosis
	Selective impairment of platelet production	Drugs (ethanol, gold, trimethoprim-sulfamethoxazole, sulfonamides, thiazides, phenylbutazone); infections (childhood rubella, HIV)
	Ineffective megakaryopoiesis	Vitamin B_{12} deficiency, folic acid deficiency, myelodysplastic syndrome, alcohol abuse
Accelerated platelet removal	Immune destruction	Autoantibodies (idiopathic thrombocytopenic purpura, systemic lupus erythematosus, lymphoproliferative disease); proven drug antibodies (quinidine, quinine, heparin, GPIIb-IIIa antagonists); infections (infectious mononucleosis, HIV, gram-negative septicemia, malaria); suspected drug antibodies (thiazide diuretics, acetaminophen, cimetidine, aminosalicylic acid); posttransfusion purpura
	Nonimmunologic removal	Disseminated intravascular coagulation, preeclampsia, vasculitis, thrombotic thrombocytopenic purpura, hemolytic-uremic syndrome, HELLP syndrome, severe bleeding, platelet washout after massive transfusion, giant hemangioma, gram-negative septicemia
	Hypersplenism	Enlarged spleen from various causes

HELLP—hemolysis, elevated liver enzymes, low platelet count

topenia after nonmyeloablative chemotherapy. They have, however, elicited antibody formation; even more worrisome, use of the truncated form has led to the development of functionally neutralizing antibodies that cross-react with the endogenous thrombopoietin.[6] Thrombopoietin mimetics are now in clinical trials.

Inadequate Platelet Production Due to Low Thrombopoietin Level

Moderate thrombocytopenia, generally in the 50,000 to 100,000/μl range, is commonly seen in patients with cirrhosis, which has been conventionally ascribed to platelet sequestration caused by hypersplenism. In addition, there is evidence that low-grade disseminated intravascular coagulation (DIC) occurs in cases of severe liver disease. Impaired clearance of fibrin degradation products may interfere with platelet function and fibrin polymerization. In one study, patients with advanced cirrhosis and thrombocytopenia were found to have low-normal serum levels of thrombopoietin (TPO). Serum TPO levels increased rapidly after orthotopic liver transplantation, and normalization of thrombocytopenia occurred within 14 days after transplantation, irrespective of the change in spleen size.[7] The data indicate that the liver is a major site of TPO production, and decreased hepatic TPO production accounts for a significant part of the thrombocytopenia in liver disease.

ACCELERATED PLATELET REMOVAL DUE TO IMMUNE DESTRUCTION

Idiopathic Thrombocytopenic Purpura

The estimated incidence of idiopathic thrombocytopenic purpura (ITP) is 50 to100 new cases per million persons per year, equally distributed between children and adults.[8,9] ITP typically appears in young women. Predisposing diseases and contributing factors may include infectious mononucleosis and other acute viral illnesses, Graves disease, and Hashimoto thyroiditis,[10] as well as antiphospholipid antibody syndrome.[11] For ITP patients who have antiphospholipid antibody, the outcomes, courses, and response to therapy do not differ from those of other ITP patients.

Pathophysiology ITP is an autoimmune disorder characterized by rapid platelet destruction that is caused by the presence of antibodies against the patient's own platelets. These autoantibodies bind to specific proteins on the platelet surface, and the antibody-coated platelets are removed by the reticuloendothelial system, especially in the spleen. The immunoglobulin on the platelet membrane is usually IgG (most commonly, IgG1). In some patients, only IgG2, IgG3, or IgG4 is present on the platelet surface, suggesting oligoclonality.[12] Immunoglobulin on the platelet membrane is frequently directed against the platelet glycoprotein (GP) IIb-IIIa, the receptor complex that mediates fibrin-ogen binding and platelet aggregation; fortunately, most of these antibodies are not capable of functionally neutralizing the GPIIb-IIIa complex. Less frequently, the immunoglobulin is directed against the GPIb complex.[13] Thrombopoietin levels are normal in ITP, indicating a normal or increased megakaryocyte mass (in contrast to a high thrombopoietin level in aplastic anemia).[14] The marrow may respond to the thrombocytopenia by increasing platelet production. In many cases, however, the marrow response is suboptimal, probably because the antiplatelet antibodies also react with megakaryocyte cell surface antigens. The platelets produced in ITP are usually large and functional, which may account for the clinical observation that most patients with ITP do not have significant clinical bleeding.

Clinical features The onset of ITP is usually insidious. History and physical examination are usually negative except for the presence of petechiae, most commonly in the lower extremities. Clinical bleeding is usually mild, consisting of purpura, epistaxis, gingival bleeding, and menorrhagia. Blood blisters (wet purpura) in the mouth indicate the presence of severe thrombocytopenia. Retinal hemorrhages are uncommon. The spleen is usually not palpable. The presence of a palpable spleen raises the possibility of systemic lupus erythematosus (SLE), lymphoma, infectious mononucleosis, or hypersplenism from underlying chronic liver disease.

Laboratory evaluation The peripheral smear is usually normal; the few platelets that are present are large and well granulated. The presence of hypochromia suggests iron deficiency from chronic blood loss; spherocytes raise the possibility of associated autoimmune hemolysis (Evans syndrome); and red blood cell fragments (schistocytes) suggest DIC, thrombotic thrombocytopenic purpura (TTP), or hemolytic-uremic syndrome (HUS). The marrow shows abundant megakaryocytes; erythroid and myeloid precursors remain normal. Results of tests for SLE are negative. Platelet-associated IgG (PA-IgG) levels are elevated; however, because platelets normally contain IgG in their α-granules, PA-IgG does not distinguish between antiplatelet antibodies, immune complexes deposited on platelet surfaces, and antibodies released from the platelet granules and bound on its surface. Therefore, tests for PA-IgG are not useful in the diagnosis of ITP, unless the tests are performed by special research laboratories that measure platelet antigen-specific antibodies.[15]

Differential diagnosis The differential diagnosis of ITP includes a falsely low platelet count resulting from ethylenediamenetetraacetic acid (EDTA)–dependent or cold-dependent agglutinins that cause in vitro platelet clumping (diagnosed by reexamination of the platelet count in a citrated or heparin-anticoagulated blood sample); the gestational thrombocytopenia of pregnancy (usually a mild problem that is not associated with increased bleeding risk [*see* Idiopathic Thrombocytopenic Purpura in Pregnancy, *below*]); myelodysplastic syndrome (usually associated with anemia and leukopenia); and underlying lymphoproliferative disease.

Course and prognosis ITP is a relatively benign disorder that has a mortality of approximately 1% to 5%; most deaths in adult cases result from intracranial bleeding. Acute ITP is usually confined to children and young adults and is frequently preceded by a viral illness. Permanent spontaneous remission occurs in less than 3 months. Chronic ITP, the usual adult variety, refers to disease that persists for more than 3 months. Although spontaneous remissions and relapses do occur in chronic ITP, long-term spontaneous remissions are uncommon. On the other hand, the long-term prognosis of ITP is benign, even in refractory cases, when these patients are managed properly.[16]

Treatment The treatment of ITP depends on the age of the patient; disease severity; whether petechiae are present alone or with moderate or severe mucosal or central nervous system bleeding; and whether the patient is pregnant.[17]

The American Society of Hematology has released an evidence-based practice guideline for the management of ITP,[15] which can be summarized as follows:

1. Patients with platelet counts above 50,000/μl do not routinely require treatment.
2. Treatment is indicated in patients with platelet counts below 20,000 to 30,000/μl and in patients with platelet counts below 50,000/μl who have significant mucosal bleeding or risk factors for bleeding (e.g., hypertension, peptic ulcer disease, or a vigorous lifestyle).
3. Patients with platelet counts below 20,000/μl need not be hospitalized if they are asymptomatic or if they have only mild purpura.

Patients with asymptomatic mild or moderate thrombocytopenia (i.e., platelet count > 50,000/μl) do not require active therapy. They may be followed and simply alerted to report any mucosal bleeding or crops of new petechiae. Avoidance of aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) is strongly advised. The role of *Helicobacter pylori* eradication in the management of ITP is controversial. It may have a limited value in improving the thrombocytopenia in young patients who have evidence of *H. pylori* infection and have relatively mild thrombocytopenia (i.e., 30,000 to 70,000/μl) of short duration (< 2 years). *H. pylori* eradication is not useful in patients with chronic severe ITP.[18]

For patients with moderate mucosal bleeding, therapy is begun with prednisone at a dosage of 60 to 100 mg/day in divided doses. Corticosteroids interfere with the macrophage attack on platelets and eventually reduce the amount of antiplatelet antibody produced by splenic and marrow lymphoid cells. A study showed that a 4-day course of high-dose dexamethasone (40 mg daily) achieved a high remission rate (85%) in newly diagnosed ITP patients; of those patients who responded to therapy, 50% had a sustained response.[19] This remarkably high rate of long-term response after a single course of dexamethasone will require confirmation and longer follow-up.

Unless bleeding is severe, the patient need not be hospitalized. Heavy physical activity, particularly any activity that involves the Valsalva maneuver, should be avoided so as not to increase intracranial pressure. The avoidance of aspirin and other NSAIDs should be emphasized. If required, red blood cell transfusions can be given; however, it is rarely necessary to transfuse platelets in such cases.

The platelet count usually rises several days to 2 to 3 weeks after the start of therapy. When the platelet count reaches normal levels, the prednisone dose can be tapered over a 3- to 4-week period. Although complete long-term remissions with prednisone alone have been reported, sustained complete response after therapy occurs in fewer than 10% of patients.

Splenectomy, usually performed by laparoscopy, is indicated if platelet counts remain below 30,000/μl after 4 to 6 weeks of steroid therapy or when the platelet count begins to fall again after the tapering of steroid. The procedure produces long-standing remission in about 65% of patients with ITP. It is best to administer intravenous immune globulin (IVIg) a few days before splenectomy so that the patient will have a platelet count of at least 30,000 to 50,000/μl at the time of surgery. Generally, a full course of IVIg therapy (1 g/kg/day for 2 days or 0.4 g/kg/day for 5 days) is not required when given as preparation for splenectomy. IVIg will produce a transient increase in the platelet count in the majority of patients, but it is a very expensive therapy. The platelet count usually begins to rise on the first postoperative day, often overshooting normal values by the second week. Pneumococcal, *Haemophilus influenzae*, and meningococcal vaccines should be administered 1 to 2 weeks before surgery.

If the patient is elderly or frail and hence may not survive splenectomy, the disease may be controlled by administration of the minimum amount of corticosteroids required to raise the platelet count to 30,000 to 50,000/μl, a level above which severe bleeding rarely occurs. Because patients with ITP who are classified as therapeutic failures generally do well clinically, the role of such potentially dangerous agents as cyclophosphamide and azathioprine in the management of such cases should be evaluated on a case-by-case basis.

Severe mucosal or CNS bleeding is a true medical emergency requiring hospitalization. Red cells are transfused as required, and prednisone is administered immediately, beginning with a 100 mg dose and then continuing at a level of 25 mg every 6 hours. A full course of IVIg should be administered, and transfusion with 8 to 10 units of random-donor platelets should be carried out when the infusion of the first dose of IVIg, usually given over approximately 60 minutes, is complete. The platelet transfusion after the infusion of IVIg produces a greater and more durable increase in the platelet count. Side effects include generalized aches, headache, flushing, fever, and chills. When severe uterine bleeding occurs, a single 25 mg dose of conjugated estrogen can be administered intravenously to control the hemorrhage. It should be emphasized that the benefit of IVIg is usually transitory and lasts only a few days. Plans for splenectomy should follow this emergency therapy.

The mechanism of action of IVIg is not completely understood. It may produce reticuloendothelial blockade by blocking the IgG-Fc sites on the monocyte-macrophages. Highly specific anti-idiotype antibodies may also block the binding of platelet autoantibodies to the platelet GPIIb-IIIa antigen.[20] Studies indicate that the catabolic rate of IgG is mediated by a new receptor for the Fc component of IgG, termed FcRn (neonatal Fc receptor, so named because it was initially identified in neonatal intestinal epithelium), on the vascular endothelial cells. Normally, IgG, but not IgM, that enters the cell through the process of pinocytosis is protected from catabolic breakdown by binding to the FcRn. After the administration of high-dose IVIg, this receptor is presumably saturated, permitting the degradation of the pathologic antibody to occur in proportion to its concentration in plasma.[21]

Refractory Idiopathic Thrombocytopenic Purpura

About 40% of ITP patients are characterized as refractory; they either remain severely thrombocytopenic after splenectomy and corticosteroid therapy or go into remission but later experience a relapse. Approximately 25% to 40% of patients will have a relapse 5 to 10 years after an initially successful splenectomy.[16] Because serious hemorrhage is uncommon with platelet counts above 30,000/μl, it is often prudent to accept an incomplete response and not proceed to more toxic forms of management. Immunosuppressive agents are generally the mainstay of therapy at this stage. However, it should be emphasized that there are no large randomized studies to address this difficult problem and that generally these patients should be referred to a hematologist.

There are several major treatment alternatives for refractory patients. Rituximab, a chimeric anti-CD20 monoclonal antibody, when administered at 375 mg/m^2 I.V. once weekly for 4 weeks, produces a lasting and substantial response in approximately one third of patients with chronic refractory ITP; however, long-

term follow-up is limited.[22] The majority of the responses occur within 8 weeks after the first infusion. The therapy is generally well tolerated, with most of the side effects (i.e., fever, chills, mild hypotension, and bronchospasm) being infusion related and occurring during or after the first infusion. Rituximab produces a profound and prolonged peripheral B cell depletion in all patients, which can last for more than a year, but serious infection is rare. Azathioprine (100 to 150 mg/day orally) or, alternatively, cyclophosphamide (100 to 150 mg/day orally) plus prednisone (40 to 60 mg/day orally) can be given, but this therapy requires weekly monitoring of complete blood count and platelet count. Prednisone may be tapered and azathioprine or cyclophosphamide adjusted to avoid severe leukopenia. A frequent mistake is to discontinue the therapeutic trial prematurely. Both azathioprine and cyclophosphamide are myelosuppressive and should be given in sufficient dosages to cause a mild leukopenia, with a white blood cell count of approximately 3,000/μl, and both have been associated with development of myelodysplastic syndrome and acute myeloid leukemia. After 1 month, alternate-day prednisone therapy should be considered to avoid steroid side effects. Because of the concern of long-term marrow toxicity associated with azathioprine and cyclophosphamide, rituximab should be considered the first-line therapy in refractory ITP patients, if treatment is indicated.

Another alternative is antibody therapy with intermittent courses of IVIg at the dosage schedules described (see above). The cost of this therapy and the usual short-lived response make it an unattractive choice. Anti-D antibody has been used successfully in $Rh^+(D^+)$ patients with ITP; in the presumed mechanism of action, the antibody-coated red blood cells block Fc receptors on macrophages and prevent the accelerated removal of platelets. Other therapeutic options include vincristine, vinblastine,[23] danazol,[24] high-dose dexamethasone, cyclosporine, interferon alfa, and plasmapheresis.

In the refractory splenectomized patient, it is important to check for the continued presence of Howell-Jolly bodies and the possibility of an accessory spleen. The disappearance of Howell-Jolly bodies suggests the presence of a remaining accessory spleen or a regenerated spleen.

Patients with clinically significant thrombocytopenic bleeding can also benefit from fibrinolysis inhibitor ε-aminocaproic acid (EACA). EACA can be given at 2 to 3 g orally four times daily until hemostasis is achieved.

HIV-Related Idiopathic Thrombocytopenic Purpura

HIV-1–related ITP appears to have a pathophysiology that is somewhat different from that of non–HIV-associated ITP, in that the antigenic specificity for the antiplatelet antibody is different. Two major antigenic determinants have been identified—a linear peptide in the platelet membrane GPIIIa and a cleavage product of talin, a platelet cytoskeletal protein, that can be generated by HIV-1 protease.[25,26] In patients with HIV infection, platelets also contain increased amounts of IgG, IgM, complement, and immune complexes. Platelet survival is moderately short, and platelet production is impaired, especially at the later stages of the disease.[27]

The use of immunosuppressive agents in HIV-infected patients is hazardous. If the drop in the platelet count is modest, no therapy is needed. When the thrombocytopenia is severe, a short course of prednisone can be administered, followed by splenectomy.

Acute thrombocytopenic hemorrhage in HIV-associated ITP may be managed with high-dose IVIg, similar to the management of other ITPs. Chronic HIV-associated ITP may respond to oral zidovudine (AZT) or other antiviral therapies [*see 7:XXXIII HIV and AIDS*]. Anti-D antibody, dapsone, and interferon have also been used with some success.[28-30] Patients who refuse splenectomy or who are thought to be poor surgical candidates may respond to low-dose splenic irradiation.[31]

Idiopathic Thrombocytopenic Purpura in Pregnancy

Mild thrombocytopenia, generally in the range of 110,000 to 150,000/μl and seldom below 70,000/μl, occurs in 5% of healthy pregnant women. When thrombocytopenia is observed for the first time during pregnancy, the differential diagnosis must include preeclampsia [*see Table 3*]. If other diagnoses can be excluded, the diagnosis is gestational thrombocytopenia (incidental thrombocytopenia of pregnancy); it requires no management.[32] If the diagnosis of ITP is made, the the patients is considered to be at high risk for complications. The platelet counts should be monitored regularly and closely, especially in the third trimester, because in many pregnant women with ITP, thrombocytopenia progressively worsens over the course of the pregnancy. The therapeutic choices are limited because splenectomy may cause spontaneous abortion and immunosuppressive agents may damage the developing fetus; therefore, therapy is usually limited to corticosteroids or IVIg. Because corticosteroids increase the risk of preeclampsia and gestational diabetes, IVIg is the drug of choice. Generally, no treatment is required until the platelet count has fallen to 20,000 to 30,000/μl or there is clinical bleeding. Typically, a single dose of IVIg (1 g/kg I.V. over 6 hours) will raise the platelet count to above 50,000/μl in the majority of patients, which will last for 3 to 4 weeks. Repeated doses can be given if necessary. In cases of severe thrombocytopenic hemorrhage, however, all of the available therapies should be used to protect the life and well-being of the mother.

Because the antiplatelet autoantibody in ITP has broad specificity and is almost always an IgG, it can cross the placenta and produce thrombocytopenia in the fetus. During a vaginal delivery, the pressure applied to the head of a thrombocytopenic fetus may induce an intracranial hemorrhage. Concern about this occurrence led many experts in the past to recommend early cesarean sections in women with a history of ITP or active disease. No data exist, however, to support this recommendation, and a much more conservative approach is now generally accepted. Most pregnant women with ITP undergo vaginal deliveries; cesarean sections are performed only for obstetric indications.

There is no correlation between maternal platelet count and the infant's platelet count. A mother with a history of ITP who has a normal platelet count can deliver a thrombocytopenic neonate (~10% incidence). Alternatively, a thrombocytopenic mother can have an infant with a normal platelet count. Measurement of maternal antiplatelet antibody is of no clinical utility. The best predictor of thrombocytopenia in a neonate is the mother's previous experience of giving birth to an infant with neonatal thrombocytopenia.[33] Neonatal severe thrombocytopenia—defined as a platelet count at birth that is less than 20,000/μl—is uncommon (1% to 5% of births), and severe bleeding complications are rare (< 1%).[34,35] The occurrence of neonatal severe thrombocytopenia is also unpredictable. The risk of intracranial hemorrhages in these infants is low (< 1%), and it cannot be reduced by cesarean section.[36] Percutaneous umbilical blood sampling is generally not recommended. Many infants who are born to mothers with ITP will have a decrease in platelet count after delivery, with a nadir on day 2; the infant's platelet

count should be monitored daily for several days.[32] Maternal ITP is not a contraindication to breast-feeding.

Thrombocytopenic Purpura with Lymphomas and Systemic Lupus Erythematosus

Patients with SLE, Hodgkin disease, or non-Hodgkin lymphoma can present with a clinical picture identical to that seen in ITP. The diagnostic approach and therapy are the same in these cases as they are in ITP. Splenomegaly with splenic sequestration, marrow infiltration with malignant cells, and recent antineoplastic or immunosuppressive therapy should be excluded. Patients with SLE or lymphoma may have Evans syndrome, in which ITP is associated with autoimmune hemolytic anemia. The management of Evans syndrome is the same as that of ITP and autoimmune hemolytic anemia.

Posttransfusion Purpura

Posttransfusion purpura (PTP) is characterized by acute onset of severe thrombocytopenia, often with a platelet count below 10,000/μl, accompanied by clinical bleeding. It may occur from 2 to 10 days after a transfusion of packed red blood cells or platelet-containing components. Almost all of the affected patients are multiparous women. Such disorders as septic thrombocytopenia, DIC, and heparin-induced thrombocytopenia must be considered in the differential diagnosis. The thrombocytopenia usually lasts for about 4 weeks. Because platelet transfusions are usually futile and sometimes precipitate severe systemic responses, they should be avoided if possible.

The pathophysiology of PTP is not completely understood. In most cases, the patient has been exposed to platelet alloantigens during pregnancy or as a result of a transfusion. Most patients with this disorder have antibodies to the human platelet antigen–1 (HPA-1), a polymorphic epitope present on platelet surface GPIIIa. The HPA-1 has two isoforms, HPA-1a and HPA-1b (previously PLA-1 and PLA-2). In the United States, approximately 98% of the white population, 99% of the African-American population, and 99% of the Asian-American population are homozygous for HPA-1. Patients in whom PTP develops are usually HPA-1a negative and HPA-1b positive. The patient has been sensitized to the HPA-1a antigen, most frequently during pregnancy, and reexposure to HPA-1a platelets during red cell transfusion leads to an anamnestic response and the destruction of the foreign platelets. It is puzzling that alloantibody directed against an antigen present on foreign platelets results in destruction of the patient's autologous platelets, which do not express the HPA-1a antigen. There is evidence suggesting that the HPA-1a antigen becomes soluble and attaches to the HPA-1a–negative platelets. Alternatively, exposure to foreign platelets may induce the formation of a true autoantibody against the endogenous platelets. The HPA-1a/HPA-1b polymorphism accounts for 80% to 90% of PTP. However, the presence of an alloantibody is necessary but insufficient for the development of PTP. Some patients with anti–HPA-1a antibodies become refractory to platelet transfusions but do not have PTP.[37] In addition, the incidence of PTP is far less common than might be predicted by the 1% to 2% of the general population who are homozygous for HPA-1b.

Confirmation of the diagnosis of PTP requires serologic studies demonstrating the presence of anti–HPA-1a antibody and a homozygous HPA-1b genotype. Several rapid platelet genotyping techniques based on the polymerase chain reaction have been developed. Homozygous deficiency of platelet CD36 (glycoprotein IV) occurs in 3% to 5% of Asians and Africans, and alloantibody against CD36 has also been found to be associated with PTP.[38] There are no controlled clinical trials evaluating therapy for PTP because of the limited number of cases. IVIg, used at doses similar to those used in the treatment of ITP, is efficacious in about 80% of cases. Plasmapheresis is also efficacious, but it is more cumbersome than IVIg administration. Use of high doses of corticosteroids is not consistently effective.[39] Transfusion of HPA-1a–negative platelets may provide some transient benefit in life-threatening bleeding situations.[40]

Drug-Induced Immune Platelet Destruction

Drug-induced immune platelet destruction is indistinguishable from ITP. The bone marrow shows abundant megakaryocytes, and special laboratories can detect the presence of antidrug antibodies.

Quinidine and quinine purpura The pathogenic antibodies in cases of quinidine and quinine purpura develop as early as 12 days after exposure to the offending agent. In most cases, drug-dependent antibodies to platelet surface GPIb-IX have been identified in patients' sera.[41] The antibodies are drug dependent because they bind to the platelets only in the presence of quinine or quinidine. Presumably, the binding of the drugs to these platelet surface glycoproteins induces new antigenic sites on the proteins that are recognized by the antibodies.

The agent (quinidine or quinine) should be withdrawn in such cases. Neither corticosteroid therapy nor emergency splenectomy is of documented benefit in purpura induced by these agents. Plasmapheresis to remove the drug and antibodies would appear to be a logical treatment, but there are no systematic studies of its effectiveness. Transfused platelets are removed as rapidly as the recipient's own platelets. Treatment with prednisone and IVIg in a dose similar to that used in ITP is recommended. Platelet transfusion after IVIg infusion may be given to control life-threatening bleeding.

A quinine-induced thrombocytopenia that is closely followed by the development of HUS has been recognized. Quinine-dependent antibodies to platelets, as well as to endothelial cells, have been found in patients' sera.[42] Even the small amount of quinine in tonic water seems to be sufficient to trigger recurrent bouts of the syndrome. Other drugs that may occasionally produce drug-dependent thrombocytopenia include dipyridamole and trimethoprim-sulfamethoxazole.[43]

Heparin-induced thrombocytopenia Heparin-induced thrombocytopenia (HIT) is a frequent cause of drug-induced thrombocytopenia in hospitalized patients. Despite the presence of modest to moderate thrombocytopenia, HIT is rarely associated with bleeding but is associated with significant and sometimes fatal thrombosis [*see 5:XIV Thrombotic Disorders*].

Gold-induced thrombocytopenia Gold salt therapy for rheumatoid arthritis produces thrombocytopenia, which is sometimes severe, in 1% to 3% of patients. There are drug-induced autoantibodies that target platelet membrane GPV, but the presence of gold is not required for their reactivity.[44] Most patients respond to therapy with 60 mg of prednisone daily. IVIg is also efficacious.

Cocaine-associated thrombocytopenia An ITP-like syndrome has been reported in intravenous cocaine users. They

have been shown to respond to an approach similar to that employed in patients with ITP.[45]

Thrombocytopenia caused by GPIIb-IIIa receptor antagonists Three parenteral GPIIb-IIIa antagonists—abciximab, eptifibatide , and tirofiban—have been approved for use in the treatment of acute coronary artery syndrome and as adjunctive therapy in coronary angioplasty. In contrast to the low platelet counts in other types of drug-induced thrombocytopenia, patients who have low platelet counts resulting from GPIIb-IIIa receptor antagonists can develop acute, often profound thrombocytopenia within a few hours after drug administration. In patients receiving abciximab, thrombocytopenia occurs in about 1% after the first exposure. After a second exposure, the incidence of thrombocytopenia rises to 4%.[46] The incidence of drug-induced thrombocytopenia associated with eptifibatide and tirofiban is probably also about 1% after first exposure.[47]

The abrupt development of severe thrombocytopenia in patients who have never been exposed to these drugs initially suggested that platelets were being destroyed by a nonimmune mechanism. However, accumulating evidence indicates that drug-dependent antibodies, which occur naturally, are the underlying cause. Preexisting anti–GPIIb-IIIa autoantibodies are present in these patients, and after the administration of the anti–GPIIb-IIIa antagonist, the binding of the drug to GPIIb-IIIa induces conformational changes in GPIIb-IIIa such that new epitopes are exposed that are recognized by the autoantibodies. These actions would explain the acute onset of profound thrombocytopenia.[47]

When thrombocytopenia develops (i.e., when platelet counts drop below 100,000/μl), the GPIIb-IIIa antagonist and any other potentially offending medications (e.g., heparin) should be discontinued immediately. Depending on the platelet count, it may not be advisable to discontinue antiplatelet agents such as aspirin or clopidogrel, because in such cases, patients are at high risk for acute coronary artery or stent thrombosis. If the platelet count drops below 10,000/μl, strong consideration should be given to platelet transfusion. In general, only one single-platelet transfusion is sufficient. There are anecdotal reports of acute coronary thrombosis associated with platelet transfusion in this setting when the platelet count climbs over 50,000/μl and the patient is off all antiplatelet agents. Thus, antiplatelet agents may need to be reinstituted. Because eptifibatide and tirofiban have very short half-lives and are cleared from the circulation within hours, the duration of thrombocytopenia is short, once the offending drugs have been discontinued. However, because abciximab has a much longer half-life—with inhibition of platelet function reported up to 1 week after drug discontinuance—thrombocytopenia can persist for 5 to 7 days. Platelet counts should be obtained in all patients before, as well as within 2 to 4 hours after, the initiation of an intravenous GPIIb-IIIa antagonist. It should be noted that a subgroup of patients develop delayed thrombocytopenia 5 to 8 days after abciximab administration. On the basis of limited published experience, it appears to be safe to administer eptifibatide or tirofiban to patients who are sensitive to abciximab, and vice versa.[47]

Thrombocytopenia caused by metabolites of naproxen and acetaminophen Five patients have experienced thrombocytopenia after taking naproxen and acetaminophen. In each case, antibodies that reacted with normal platelets in the presence of a known drug metabolite of naproxen or acetaminophen were identified.[48] Therefore, the sensitizing agents are drug metabolites that formed in vivo.

ACCELERATED REMOVAL OF PLATELETS BY NONIMMUNOLOGIC MECHANISMS

There are several nonimmunologic causes for thrombocytopenia. Blood vessel wall injury with increased thrombin generation and increased platelet activation and consumption occurs in several of these conditions.

Thrombotic Thrombocytopenic Purpura and Adult Hemolytic-Uremic Syndrome

TTP and HUS encompass a group of clinical syndromes characterized by widespread platelet-fibrin thrombi deposition in the small arteries and arterioles and capillaries. Thrombotic microangiopathy is a distinct feature of both TTP and HUS; however, the underlying pathogenetic processes in TTP and HUS may differ [*see* Pathogenesis, *below*]. Familial TTP/HUS is rare and usually occurs in the immediate postnatal period or infancy, although there are reported cases of delayed onset until the second to third decade of life. More frequently, TTP is either idiopathic or secondary to a variety of conditions [*see* Etiology, *below*].

Etiology TTP/HUS occurs spontaneously and is also associated with pregnancy, cancer, bone marrow transplantation, autoimmune diseases, and various drugs. In pregnancy, it resembles severe preeclampsia. In the postpartum period, the CNS manifestations may initially be confused with postpartum depression, with tragic results. Cases have been reported after a normal delivery and with abruptio placentae and preeclampsia.

Several drugs appear to cause TTP/HUS. These include chemotherapeutic drugs (e.g., mitomycin C, bleomycin, and cisplatin), immunosuppressive agents (e.g., cyclosporine and FK506), the antiplatelet agent ticlopidine, oral contraceptives, and quinine. Anecdotal cases of TTP/HUS associated with clopidogrel, which is related to ticlopidine, have also been reported.[49]

Pathogenesis There have been significant recent advances in the understanding of TTP, showing that the proper processing of von Willebrand factor (vWF) multimers plays a key role in its pathogenesis. vWF is an abundant plasma protein that mediates platelet adhesion to the subendothelium and serves as a carrier molecule for factor VIII [*see 5:XII Hemostasis and Its Regulation*]. vWF is synthesized by both megakaryocytes and endothelial cells. Monomers of vWF (280,000 daltons) are cross-linked by disulfide bonds to form vWF multimers, which are released into the circulation by endothelial cells and are stored within platelet α-granules and the Weibel-Palade bodies in endothelial cells. The stored vWF multimers can be released upon platelet or endothelial stimulation. These released vWF multimers are larger than plasma vWF multimers and are referred to as ultra-large vWF (ULvWF) multimers, with a molecular size up to 20 million daltons. Functionally, these are the most reactive vWF multimers. In 1982, ULvWF multimers were found in the plasma of patients with chronic relapsing TTP, giving rise to the hypothesis that TTP may result from the deficiency of a vWF-cleaving protease (depolymerase), which causes ULvWF multimers to circulate, contributing to the development of thrombosis.[50] This hypothesis has been proved largely correct with the recent identification of the vWF-cleaving protease and the demonstration that deficiency of the vWF-cleaving protease activity is associated with TTP.

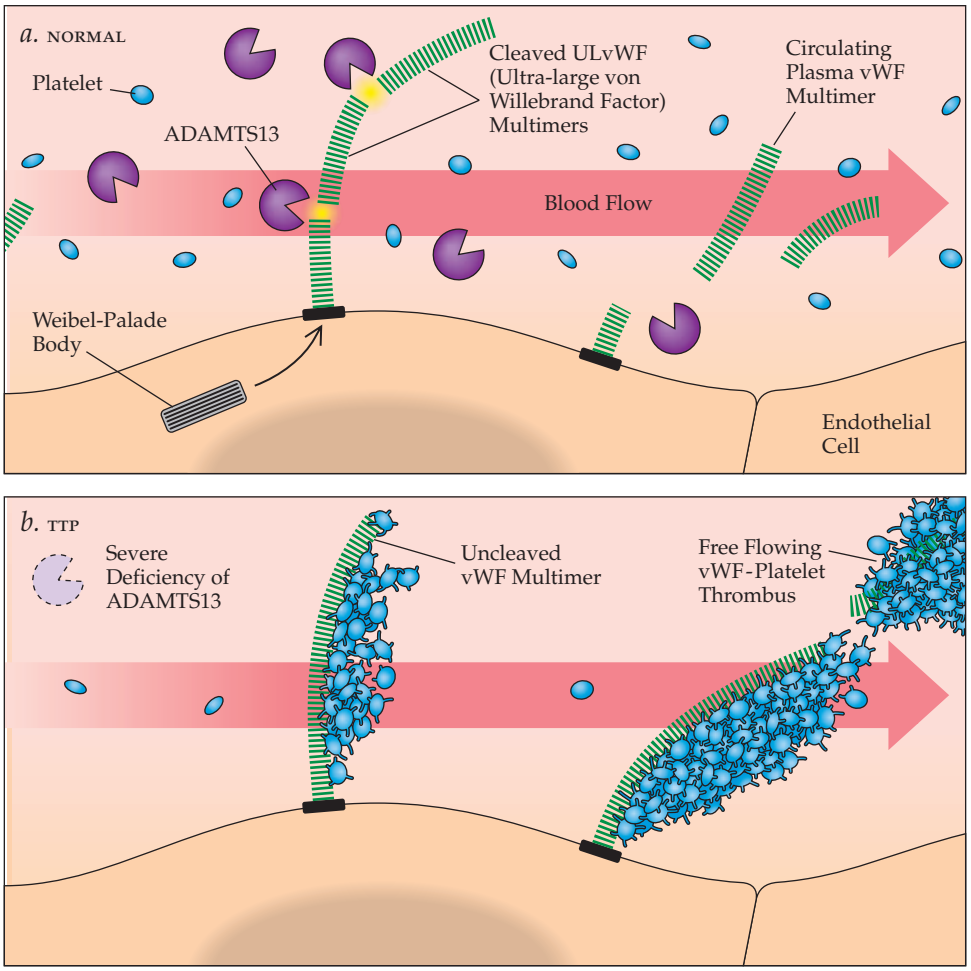


Figure 1 **ADAMTS13 activity in normal and thrombotic thrombocytopenia purpura plasma. (*a*) In normal persons, ADAMTS13 enzyme molecules from the plasma attach to and then cleave the unusually large von Willebrand factor (ULvWF) multimers that are secreted in long strings from stimulated endothelial cells. (*b*) In patients with thrombotic thrombocytopenic purpura (TTP), a deficiency of ADAMTS13 prevents the cleavage of ULvWF multimers secreted by endothelial cells. Platelets carried by flowing blood adhere to the uncleaved ULvWF multimers, resulting in the development of platelet thrombi.**

The vWF-cleaving protease has been identified as ADAMTS13 (a disintegrin-like and metalloprotease with thrombospondin type 1 motif 13).[51] It is a novel metalloprotease that cleaves vWF monomer at a specific site (842Tyr-843Met) in the A2 domain. Current data indicate that ULvWF multimers are secreted from stimulated endothelial cells as a long "string" anchored on the endothelial cell surface. Plasma ADAMTS13 may attach, under flowing conditions within the blood, to the cell surface–bound ULvWF multimers (via the A3 domain in the vWF monomer) and cleave them into the vWF multimers that are normally found in plasma.[52] Partial unfolding of the ULvWF multimers by shear stress forces in the blood presumably enhances the enzymatic cleavage process. Patients with familial TTP have hereditary deficiency of ADAMTS13,[53] whereas patients with acquired idiopathic TTP have antibodies that inhibit the ADAMTS13 activity.[54,55] In either case, the persistence of ULvWF multimers on the stimulated surface of the endothelial cell leads to the adhesion and subsequent aggregation of platelets, which in turn lead to the formation of platelet thrombi [*see Figure 1*]. (Presumably, platelets do not bind to the smaller plasma vWF multimers, because the platelet binding sites are not exposed in these vWF multimers). In addition to causing ischemic injury at the site of thrombi formation, it is likely that platelet thrombi resulting from aggregation of ULvWF multimers will break up and embolize downstream, resulting in further ischemic tissue damage.

The gene encoding ADAMTS13 is located on chromosome 9q34. More than 50 mutations in this gene have been identified in patients with familial TTP, most of which result in greatly reduced ADAMTS13 secretion in vitro.[51] In many cases of acquired idiopathic TTP, an IgG autoantibody against ADAMTS13 is produced transiently, leading to severe deficiency of ADAMTS13 activity. ULvWF multimers are detectable in the plasma in some patients during the acute episodes but not after recovery.[50]

Although the role of ADAMTS13 in the pathogenesis of TTP has been established, it appears that the majority of patients with HUS do not have severe ADAMTS13 deficiency, strongly suggesting that the pathogenesis of HUS is different.

ADAMTS13 as a screening assay The clinical utility of measuring ADAMTS13 is not established. In part, this is be-

cause there is no gold standard for its measurement; most of the current assays have long turnaround times and are not readily available. Furthermore, the sensitivity and specificity of ADAMTS13 deficiency for the diagnosis of TTP remains unclear. Decreased vWF-cleaving activity is found in many clinical conditions that are not associated with TTP, including cirrhosis, chronic renal insufficiency, ITP, DIC, SLE, leukemia, pregnancy, and the postoperative state; it is also associated with advancing age.[56,57] In a prospective study involving 37 patients, severe deficiency in the ADAMTS13 level (< 5%) was found in 80% of patients with idiopathic TTP but in none of the patients with TTP associated with hematopoietic stem cell transplantation, cancer, drugs, or pregnancy.[58] Thus, acquired TTP may be considered as either idiopathic or secondary; the former is generally associated with severe ADAMTS13 deficiency, whereas the latter is not. Among the patients with idiopathic TTP and severe ADAMTS13 deficiency, 44% had inhibitors. Other studies found an incidence of inhibitors in idiopathic TTP of 65% to 95%; however, part of the variation in study results may have to do with patient selection.[59,60]

The reason why an inhibitor is not detectable in a substantial number of the idiopathic TTP patients is unclear. It is possible that the current assay is not sufficiently sensitive; alternatively, the assay may involve a nonneutralizing antibody that binds to ADAMTS13 and accelerates its clearance. New assays using a recombinant vWF fragment as the substrate for the ADAMTS13 protease are in development and should help clarify some of these issues.[61,62]

Clinical features and diagnosis The five major manifestations (pentad) of TTP are (1) severe microangiopathic hemolytic anemia associated with a very high serum lactic dehydrogenase (LDH) level and a blood smear showing the characteristic schistocytes and helmet cells; (2) moderate to severe thrombocytopenia with increased marrow megakaryocytes, which indicates intravascular platelet activation and consumption; (3) fever, which is occasionally quite high; (4) CNS signs and symptoms that can be quite mild initially with transient agitation, headache, and disorientation but that can sometimes progress explosively to hemiparesis, aphasia, seizures, focal deficits, coma, and death; and (5) renal disease, which is usually mild and produces moderate elevations of serum creatinine and urinary protein levels. It should be emphasized that many patients do not present with all these signs and symptoms. Patients with familial TTP/HUS typically exhibit a chronic relapsing course.

The adult form of HUS has features similar to those of TTP, although the pathophysiology may not be identical [*see* Pathogenesis, *above*]. Common features of TTP and HUS include microangiopathic hemolytic anemia, thrombocytopenia, and the presence of platelet fibrin thrombi in the small vessels. Renal involvement is uniformly severe in HUS, whereas CNS disease is less prominent than in TTP. There is a distinct form of HUS that occurs in children after gastrointestinal infection with *Escherichia coli*, usually serotype 0157:H7. These patients present with bloody diarrhea and hemorrhagic colitis. *E. coli* 0157:H7 or other strains elaborate verotoxins (also called Shiga toxins) that bind to specific receptors on the endothelial surface, causing cell damage and even cell death.[63] Verotoxin-1 (VT-1) can induce the upregulation of various prothrombotic and proinflammatory adhesive molecules on endothelial cells.[64] The microvascular endothelial cells are particularly susceptible because they have a high expression of VT-1 receptors, which may explain the propensity for thrombosis in the microcirculation. Antibiotic treatment of children with *E. coli* 0157:H7 infection increases rather than decreases the risk of HUS, presumably because it causes the release of verotoxins from injured bacteria in the intestine, making the toxins more available for absorption. Thus, routine treatment with antibiotics is not recommended.[65]

Whereas a severe deficiency of ADAMTS13 (< 5%) may be specific for TTP,[66] patients with severe ADAMTS13 deficiency may have prolonged asymptomatic periods. It is becoming clear that loss of ADAMTS13 activity, with an associated increase in circulating ULvWF multimers, is necessary but insufficient to cause an acute clinical TTP episode. The current data support the hypothesis that severe ADAMTS13 deficiency, be it from familial or acquired cause, predisposes the patient to thrombosis, and a second vascular inflammatory stimulus, such as infection, surgery, or pregnancy, causes the endothelium to increase its release of the stored UlvWF multimers, which, in the setting of grossly impaired processing, gives rise to ULvWF platelet thrombi in the microcirculation and clinical thrombosis.

Differential diagnosis Both TTP and HUS must be differentiated from SLE and from Evans syndrome. Microangiopathic hemolysis, neutrophilic leukocytosis, and a negative direct Coombs test (direct antiglobulin test) strongly suggest TTP or HUS. Coagulation tests usually reveal no significant abnormalities (i.e., no evidence of DIC); serum LDH is usually elevated. A marrow biopsy is generally not required but may show the characteristic, but not pathognomonic, platelet-fibrin hyaline thrombi in small arteries and arterioles.

Treatment Prompt institution of plasma exchange with fresh frozen plasma is the treatment of choice for TTP/HUS. In a large randomized trial by the Canadian Apheresis Group, intensive plasma exchange was more effective than plasma infusion in terms of patient survival (78% versus 63%).[67] In that study, 1.5 times the calculated plasma volume was removed and replaced with fresh frozen plasma during each of the first 3 days of therapy; subsequently, one single-volume exchange a day was performed for a minimum of 7 days. Some investigators obtained good results with a daily single-volume exchange instead of a 1.5-volume exchange.[68] It is reasonable to start with a daily single-volume exchange if the patient is clinically relatively stable, with moderate thrombocytopenia and no significant neurologic impairment. However, if the clinical situation worsens, more intensive double-volume plasma exchange (5,000 to 6,000 ml/day, or approximately 80 ml/kg/day) is indicated. Because vWF multimers are present in cryoprecipitate, cryosupernatant (i.e., fresh frozen plasma from which cryoprecipitate has been removed) can be substituted as replacement fluid when a patient is not responding to routine plasma exchange. One uncontrolled study showed increased benefit from this preparation as compared with fresh frozen plasma.[69] Once therapeutic benefit has been achieved (as measured by restoration of normal CNS function, by rising platelet counts, and by falling LDH levels), the intensity and frequency of plasma exchange can be reduced to single-volume exchanges, first three times weekly and then twice weekly.

Although the importance of prompt plasma exchange has been established, the use of corticosteroids,[70] aspirin, and dipyridamole has not been tested in prospective clinical trials. With the observation of autoantibody against ADAMTS13 as a significant cause of acquired idiopathic TTP, rituximab (375 mg/m^2

I.V. once weekly for 4 weeks) has been tried and reported to be effective, although the overall experience is still limited.[71-73] Because pheresis tends to lower the platelet count in a patient who is already thrombocytopenic, the problem of platelet transfusion arises. Some investigators have observed that platelet infusion may lead to exacerbation of TTP,[74] whereas others use platelet transfusions as required.

In the previously described prospective study of patients with severe deficiency in ADAMTS13 activity (see above), plasma exchange proved to be a useful therapy; among the patients with idiopathic TTP and severe ADAMTS13 deficiency without detectable inhibitors, the majority responded to plasma exchange with complete remission and a rise in the ADAMTS13 level.[58] However, in patients with mild ADAMTS13 deficiency and high inhibitor levels, plasma exchange was not effective in reducing the inhibitor titer or in increasing the ADAMTS13 activity. Nevertheless, some of these patients had a favorable clinical response, including resolution of thrombocytopenia and cessation of hemodialysis. Among the patients whose TTP was not idiopathic, response to plasma exchange was variable. Of note, mortality in patients with idiopathic TTP with severe ADAMTS13 deficiency has been shown to be 15% to 20%, whereas mortality in patients with nonidiopathic TTP has been much higher, at 55% to 60%.[58,60] Many of the patients in the latter group have had serious underlying disease and comorbidities, such as hematopoietic stem cell transplantation, which likely has contributed to the high mortality.

Management of acute TTP should therefore depend on the clinical manifestations and course of the disease.[75] In all cases, plasma exchange is first-line therapy. Patients who respond promptly and completely to plasma exchange—which most likely will be those patients with idiopathic TTP and very low or nondetectable levels of inhibitors—may not need any further treatment. For patients who show a suboptimal response—such as an initial rise in platelet counts or a recurrence in thrombocytopenia when the plasma exchange treatments are decreased—glucocorticoid is indicated. For patients who experience a more aggressive course (e.g., those with severe neurologic abnormalities or those who do not respond to the initial plasma exchange with or without steroid therapy), more intensive immunosuppressive therapies, such as rituximab, should be considered.

Microangiopathy may persist for weeks or months after all other evidence of disease has subsided. In a large follow-up study of TTP patients, about one third of patients who entered remission had a relapse over a 10-year period.[76] The risk of recurrence is largely restricted to patients with severe ADAMTS13 deficiency—primarily, patients with idiopathic TTP. Relapse seldom occurs in patients who have TTP in association with hematopoietic stem cell transplantation or drugs. Conflicting data have been reported regarding patients who have had relapses after TTP associated with pregnancy.

Most experts treat adult HUS in a manner similar to that for TTP. However, the response to plasma exchange appears to be less favorable in HUS than in TTP, which may be consistent with the recent finding that the ADAMTS13 level is generally not diminished in HUS.

Thrombocytopenia Induced by Infection

Severe viral, bacterial, fungal, and parasitic infections can produce DIC and, consequently, thrombocytopenia [*see 5:XIV Thrombotic Disorders*]; however, mechanisms other than DIC may also cause infection-associated thrombocytopenia.

Viral infections Viral infections such as dengue fever and congenital rubella can directly damage the megakaryocytes. Varicella can cause a form of thrombocytopenia that has the characteristics of an immune reaction: increased numbers of megakaryocytes, no evidence of DIC, and the presence of PA-IgG. Usually, no therapy is required. The acute thrombocytopenia in infectious mononucleosis is probably immune mediated, as shown by the increase in marrow megakaryocytes and the favorable response to corticosteroids.

Bacterial septicemia Patients who have severe gram-negative septicemia and platelet counts lower than 50,000/μl usually have evidence of DIC. However, many patients who have both gram-negative and gram-positive septicemia and platelet counts between 50,000 and 150,000/μl have no signs of DIC. The PA-IgG levels are often elevated, which may represent immune complexes deposited on the platelet surface rather than antiplatelet autoantibodies. The key to controlling the thrombocytopenia is establishing appropriate therapy for the infection. If DIC is present, it should be managed with careful control of hypotension and blood volume.

Protozoan infection Thrombocytopenia is common in malaria, although DIC is rare. Platelet survival is short, and elevated PA-IgG has been found to be elevated. The IgG antibody appears to bind to malarial antigens adsorbed to the platelet surface.[77]

Thrombocytopenia during Pregnancy and Peripartum Period

Mild thrombocytopenia, with platelet counts generally in the range of 110,000 to 150,000/μl and seldom below 70,000/μl, occurs in 5% to 8% of pregnant women (gestational thrombocytopenia). It has no clinical significance, but it must be distinguished from ITP, pregnancy-associated TTP, and preeclampsia.

In addition to having hypertension, proteinuria, and evidence of pathologic changes in the kidneys, liver, CNS, and placenta, approximately 15% of patients with preeclampsia have moderate thrombocytopenia. Only a minority of patients with preeclampsia and thrombocytopenia demonstrate laboratory evidence of DIC. The megakaryocyte number is increased, and platelet survival is somewhat shortened. Some patients with preeclampsia and thrombocytopenia also have microangiopathic hemolysis, which suggests that damaged vessels containing fibrin strands are destroying red blood cells and platelets. Intense vasospasm that causes endothelial damage and leads to platelet activation, adherence, and destruction may also play a role. The clinical picture may be indistinguishable from TTP, in which case it should be managed as TTP. Otherwise, management consists of prenatal care for preeclampsia and efforts to detect thrombocytopenia as early as possible.

HELLP syndrome The HELLP syndrome refers to a disorder that occurs during pregnancy and is characterized by hemolysis, elevated levels of liver enzymes, and a low platelet count. It probably represents an extremely severe form of preeclampsia. At some point between the 23rd and 39th week of pregnancy, affected patients present with thrombocytopenia marked by a platelet count of less than 100,000/μl, microangiopathic hemolysis, abnormal liver function test results, and, occasionally, hypertension.[78] The results of the standard coagulation tests for DIC are normal, although there may be some elevation in the level of

fibrin degradation products. Patients with the HELLP syndrome are often severely ill, with circulatory, respiratory, and renal failure; postpartum hemorrhage; intrahepatic hemorrhage; and seizures.

The differential diagnosis includes acute fatty liver of pregnancy and TTP/HUS. In acute liver of pregnancy, patients have a prolonged PT and aPTT and low fibrinogen levels; hypertension and proteinuria are usually absent. In TTP/HUS, the liver enzymes are normal or only mildly elevated.

HELLP is treated by terminating the pregnancy, usually by delivery, and by providing meticulous supportive care. In a large series of patients with HELLP, the nadir of thrombocytopenia occurred 1 to 2 days after delivery.[79] It may also develop for the first time within 24 to 48 hours post partum.[80] Treatment of HELLP remains controversial, but corticosteroids appear to be beneficial.[81] Persistent thrombocytopenia with microangiopathy or the presence of organ failure suggests postpartum TTP/HUS; in such patients, plasma exchange therapy should be considered.

Platelet Washout and Vascular Bed Abnormalities

Patients who have brisk bleeding during surgery and who require massive transfusions (e.g., 10 units of packed red cells and multiple units of fresh frozen plasma) frequently develop nonimmune thrombocytopenia. If the platelet level falls below 100,000/μl and the patient is undergoing surgery or another hemostatic challenge, platelets should be administered. Platelets may also be removed by an abnormal vascular bed. In giant hemangiomas, there is sluggish blood flow through improperly endothelialized channels. These surfaces may produce low-grade DIC.

Platelet Sequestration

Another major mechanism of thrombocytopenia is platelet sequestration. Platelet counts of 40,000 to 80,000/μl are common in patients with marked splenomegaly. Clinically significant hemorrhage rarely occurs unless a coexistent hemorrhagic disorder is present. Management is directed toward the primary disease. Splenectomy is rarely necessary.

Platelet Function Disorders

The clue to the existence of a platelet function defect is the finding of clinical hemorrhage in the presence of a prolonged bleeding time and a platelet count higher than 100,000/μl. Petechiae are rare. Platelet morphology and tests of platelet function may be abnormal [*see Table 4*].

HEREDITARY ABNORMALITIES OF PLATELET FUNCTION

Platelet Membrane Disorders

Bernard-Soulier syndrome is a rare autosomal recessive disease that is characterized by giant platelets, a prolonged bleeding time, moderate thrombocytopenia, and risk of fatal hemorrhage. The defect, which is an absence of the platelet GPIb-IX-V complex (the major vWF binding site of the platelet), causes impaired platelet adhesion to wound surfaces. Ristocetin-induced platelet agglutination is abnormal and is not corrected by the addition of normal plasma containing vWF. Acute hemorrhage is treated by platelet transfusions.

Glanzmann thrombasthenia is a rare autosomal recessive disorder in which platelet morphology and the platelet count are normal but the bleeding time is prolonged. Because the critically important GPIIb-IIIa complex that forms the platelet binding site for fibrinogen is absent, the platelets do not undergo aggregation after stimulation by ADP, thrombin, or collagen. Ristocetin-induced agglutination, however, is normal. Treatment consists of platelet transfusions when necessary.

Platelet Granule Disorders

Patients with the gray platelet syndrome, a rare disorder, have mucosal bleeding, ecchymoses, and petechiae. Moderate thrombocytopenia is present, and the bleeding time is prolonged. The platelets are larger than normal and appear agranular because of the absence of α-granules. Because the α-granule contents are severely reduced, platelet adhesion and platelet-supported coagulation are deficient. Platelet aggregation with collagen is abnormal. Bleeding episodes should be treated by infusion of normal platelets.

Another rare disorder, the dense granule deficiency syndrome, is characterized by mucosal bleeding associated with a normal platelet count, normal platelet morphology, and variable prolongation of the bleeding time. Platelet aggregation with ADP and collagen are abnormal. The decrease in the dense granular contents of ADP impairs ADP-mediated events. Hemorrhage is treated by platelet transfusion.

1-Desamino-8-D-arginine vasopressin (DDAVP, or desmopressin) is an alternative therapy for patients with primary platelet disorders that require surgery.

ACQUIRED ABNORMALITIES OF PLATELET FUNCTION

Myeloproliferative Diseases and Associated Platelet Abnormalities

Platelet function abnormalities occur in the myeloproliferative diseases: chronic myeloid leukemia, polycythemia vera, essential thrombocythemia, and acute leukemia. The platelet count in chronic myeloproliferative disorders is often very high, but the bleeding time may be prolonged, and clinical bleeding may appear as mucosal hemorrhage and hematomas. The abnormality resembles an acquired storage-pool defect. Megakaryocytes often are abnormal with separated nuclei; the peripheral blood platelets are large and may be degranulated. Management of acute hemorrhage consists of transfusion of normal platelets to bring the level of normal platelets up to 50,000/μl. Aspirin and other NSAIDs should be avoided.

Uremia and Associated Platelet Abnormalities

A prolonged bleeding time associated with clinical bleeding despite a normal platelet count has been well documented in uremia. Uremic platelet dysfunction is presumably caused, in part, by several dialyzable uremic toxins, including phenolic acids and guanidinosuccinic acid.[82] DDAVP (0.3 μg/kg in 50 ml of saline over a 30-minute period) is effective in controlling uremic bleeding for about 4 to 6 hours. DDAVP infusion produces an increase in plasma vWF activity, particularly among the larger multimers of vWF, which may enhance platelet adhesion.

The hematocrit should be maintained above 30% in bleeding uremic patients because the bleeding time is prolonged when the hematocrit falls below 26%.[83] Bleeding may also be controlled by the use of conjugated estrogens. Conjugated estrogen (Premarin) given orally (50 mg/day) or intravenously (0.6 mg/kg/day) for 4 to 5 days shortens the bleeding time by approximately 50% for about 2 weeks.[84] The advantage of conjugated estrogens over DDAVP is the longer duration of their benefi-

Table 4 Classification of Platelet Function Disorders

Type	*Characteristic*	*Cause*
Congenital	Membrane abnormalities	Bernard-Soulier disease (GPIb-IX-V defect, impaired adhesion); Glanzmann thrombasthenia (GPIIb-IIIa defect, impaired aggregation)
	Granule abnormalities	Gray platelet syndrome (absent or impaired α-granule release, impaired aggregation); dense granule deficiency (absent or impaired dense granule release, impaired aggregation)
	Deficiency of a plasma factor	von Willebrand disease (deficiency or abnormality of von Willebrand factor, impaired adhesion); afibrinogenemia (deficiency of fibrinogen, impaired aggregation)
Acquired	Production of abnormal platelets	Myeloproliferative disease (essential thrombocytopenia, chronic myelogenous leukemia, polycythemia vera, myelofibrosis, acute myelogenous leukemia); myelodysplasia
	Dysfunction of normal platelets	Systemic disease (uremia, liver disease, paraproteinemias, disseminated intravascular coagulation); drugs (aspirin and other nonsteroidal anti-inflammatory drugs, ticlopidine, clopidogrel, GPIIb-IIIa antagonists, dextran, antibiotics [penicillin, carbenicillin, moxalactam], psychotropic drugs)

cial effect on platelet function, but they have a more delayed onset of action. The two drugs can be used concomitantly.

Patients with end-stage renal disease have complex hemostatic disorders. Despite decreased platelet function (caused by uremic toxins present in the circulating blood), thrombosis of vascular access shunt commonly occurs in patients with end-stage renal failure who are on hemodialysis. Hemostatic parameters suggestive of a hypercoagulable state, such as increased plasma fibrinogen and factor VIII levels, have been described.[82]

Effects of Macroglobulinemia and Other Dysproteinemias on Platelet Function

The presence of high concentrations of viscous proteins produces complicated effects on the entire hemostatic mechanism. The proteins appear to coat platelets and interfere with adhesion and perhaps with aggregation. Management is directed at the primary disease, but if hyperviscosity and bleeding are significant, prompt plasmapheresis may be required to lower the level of abnormal protein and to correct the bleeding disorder.

Drug-Induced Platelet Disorders

Aspirin and other nonsteroidal anti-inflammatory drugs In normal persons, ingestion of 0.6 g of aspirin prolongs the template bleeding time by 2 to 3 minutes. The platelets are irreversibly affected. Thromboxane A_2 (TXA_2) is a potent inducer of platelet release and aggregation [*see 5:XII Hemostasis and Its Regulation*]. Aspirin acetylates and irreversibly inhibits cyclooxygenase-1 (COX-1) and blocks the subsequent generation of thromboxane. Some apparently normal persons display marked sensitivity to the action of aspirin, so that their bleeding times are very much prolonged and they have clinically significant bleeding, particularly during or after surgery or trauma. These patients may have a mild form of von Willebrand disease or storage-pool disease, and their mild bleeding diathesis becomes exacerbated by aspirin's antiplatelet effect.

Uremic patients are especially sensitive to bleeding induced by aspirin. A small dose of aspirin does not prolong the bleeding time of normal persons, but in uremic patients, it produces a significant prolongation, often as much as 15 minutes. The combination of alcohol and aspirin is also dangerous because of aspirin's ability to prolong the bleeding time.

Aspirin-induced bleeding is diagnosed by determining the existence of an acquired platelet function defect (a platelet count above 100,000/μl, abnormal platelet aggregation test results, and no prior bleeding history) and finding evidence of aspirin ingestion. Because approximately 300 compounds on the market contain aspirin, a negative history should be supplemented either by determining a serum salicylate level or by detecting an abnormal collagen aggregation pattern that reverts to normal in 7 days (the typical pattern of aspirin ingestion).

If bleeding is significant, it can be managed by platelet transfusion. Because inhibition of platelet COX-1 by aspirin is irreversible, the hemostatic compromise may last for 4 to 5 days after the aspirin has been discontinued. If the patient needs analgesia, acetaminophen or codeine can be used because neither affects platelet function.

Alcohol In addition to producing thrombocytopenia by suppressing platelet production, alcohol consumption can cause platelet function defects.[85] In vitro studies have shown that alcohol impairs platelet aggregation and TXA_2 release. Platelet function returns to normal after 2 to 3 weeks of abstinence.

Antibiotics Carbenicillin and ticarcillin can inhibit platelet aggregation and contribute to a bleeding disorder, as can massive doses of penicillin. Massive doses of penicillin impair collagen-induced and ristocetin-induced platelet aggregation. Moxalactam, a third-generation cephalosporin, also causes a platelet function disorder. The clinical situation is most important when an acquired platelet function defect develops in a pancytopenic patient being treated for septicemia. Changing the antibiotics usually corrects this problem.

Miscellaneous agents A wide variety of other agents can modify platelet function [*see Table 5*].[86]

Thrombocytosis and Thrombocythemia

DIAGNOSIS

A platelet count higher than 500,000/μl is referred to as reactive thrombocytosis. In reactive thrombocytosis, tests of platelet function (including platelet aggregation studies) are generally normal, and patients do not experience an increased incidence of hemorrhage or thromboembolism even when the platelet count exceeds 1 million/μl.

Elevated platelet counts (often, 1 million to 3 million/μl or more) also occur in chronic myeloid leukemia, agnogenic myeloid metaplasia with myelofibrosis, polycythemia vera, and essential thrombocythemia. In the diagnosis of essential

Table 5 Selected Platelet-Modifying Agents[98]

- Anesthetics
  - General
    - Halothane
  - Local
    - Butacaine
    - Cocaine
    - Cyclaine
    - Dibucaine
    - Procaine
    - Tetracaine
- Antibiotics (β-lactam)
  - Cephalosporins
    - Cefazolin
    - Cefotaxime
    - Cefoxitin
    - Cephalothin
    - Moxalactam
  - Penicillins
    - Ampicillin
    - Apalcillin
    - Azlocillin
    - Carbenicillin
    - Methicillin
    - Mezlocillin
    - Nafcillin
    - Penicillin G
    - Piperacillin
    - Sulbenicillin
    - Temocillin
    - Ticarcillin
- Antibiotics (other)
  - Nitrofurantoin
- Anticoagulants
  - Heparin
- Antihistamines
  - Chlorpheniramine
  - Diphenhydramine
  - Mepyramine
- Cardiovascular drugs
  - Diltiazem
  - Isosorbide dinitrate
  - Nifedipine
  - Nitroglycerin
  - Nitroprusside
  - Propranolol
  - Quinidine
  - Verapamil
- Drugs that increase platelet cAMP concentration
  - Dipyridamole*
  - Iloprost
  - Prostacyclin
- Fibrinolytic agents
- Foods and food additives
  - Ajoene
  - Chinese black tree fungus
  - Cloves
  - Cumin
  - Ethanol
  - Omega-3 fatty acids
  - Onion extract
  - Turmeric
- Glycoprotein IIb-IIIa antagonists
  - Abciximab
  - Eptifibatide
  - Lamifiban
  - Tirofiban
- Narcotics
  - Heroin
- Nonsteroidal anti-inflammatory drugs
  - Aspirin*
  - Diflunisal
  - Ibuprofen
  - Indomethacin
  - Meclofenamic acid
  - Mefenamic acid
  - Naproxen
  - Phenylbutazone
  - Piroxicam
  - Sulfinpyrazone*
  - Sulindac
  - Tolmetin
- Oncologic drugs
  - BCNU
  - Daunorubicin
  - Mithramycin
- Plasma expanders
  - Dextrans
  - Hydroxyethyl starch
- Psychotropic drugs
  - Phenothiazines
    - Chlorpromazine
    - Promethazine
    - Trifluoperazine
  - Tricyclic antidepressants
    - Amitriptyline
    - Imipramine
    - Nortriptyline
- Miscellaneous agents
  - Clofibrate
  - Clopidogrel
  - Ketanserin
  - Radiographic contrast agents
    - Conray-60
    - Renografin-76
    - Renovist II
  - Ticlopidine*

*Used as a therapeutic antithrombotic agent.
BCNU— bischloronitrosourea (carmustine) cAMP—cyclic adenosine monophosphate

thrombocythemia, the platelet count is higher than 600,000/μl and other causes of thrombocytosis (e.g., another myeloproliferative disorder or reactive thrombocytosis) have been excluded. A gain-of-function mutation in tyrosine Janus kinase–2 (JAK2) has been found in many patients with myeloproliferative disorders; it is found in about 80% of patients with polycythemia vera and in about 25% to 50% of patients with essential thrombocythemia.[87-89] In myeloproliferative disorders, test results of platelet function are frequently abnormal [*see* Platelet Function Disorders, *above*]. Some patients with myeloproliferative disorders seem to show an enhanced propensity for hemorrhage and thromboembolism. Neither platelet number nor measurements of platelet function predict the degree of thrombosis or hemorrhage.

Clinically, the hemorrhagic signs include mucosal, particularly gastrointestinal, bleeding; hematomas; and ecchymoses. There may be splenic vein thrombosis, portal or mesenteric vein thrombosis, and recurrent deep vein thrombosis with or without pulmonary embolism. Arterial thrombosis is less common.

TREATMENT

Patients with essential thrombocythemia and polycythemia vera may have debilitating erythromelalgia (burning and itching of the fingers and toes) that can progress to ischemic acrocyanosis.[90] This symptom complex appears to be caused by occlusion and inflammation of arterioles by platelet aggregates. Aspirin or indomethacin produces relief within hours. Aspirin given at a dosage of 325 mg daily can produce lasting benefit.

Hemorrhage and thrombosis are uncommon events even with platelet counts of 1 million/μl. In a patient with essential thrombocythemia who has clinically significant hemorrhage or thrombosis, good control of the platelet count can be achieved with oral hydroxyurea (15 mg/kg/day), with adjustments in the dosage as needed to lower the platelet count. Hydroxyurea therapy requires careful monitoring of the blood count; thus far, hydroxyurea therapy does not appear to increase the risk of a second malignant disorder. Newer therapies for thrombocythemia include the use of anagrelide, a powerful platelet-lowering agent.[91] A recent prospective, randomized trial comparing hydroxyurea with anagrelide showed that although the two agents were equally efficacious in long-term control of the platelet count, anagrelide was associated with an increased risk of arterial thrombosis (mostly transient ischemic attacks) and serious hemorrhage.[92] Thus, hydroxyurea (0.5 to 2.0 g daily to maintain the platelet count at less than 400,000/μl) plus low-dose aspirin (81 mg daily) should be the first-line therapy for patients with essential thrombocythemia who are considered at high risk for vascular events.

Vascular Purpuras

Vascular purpuras are a heterogeneous group of disorders [*see Table 6*] that are characterized by cutaneous hemorrhage and are occasionally associated with mucosal bleeding. The leakage occurs from terminal arterioles, capillaries, and postcapillary venules. The results of tests of platelet number and function and tests of procoagulant function are normal.

HEREDITARY HEMORRHAGIC TELANGIECTASIA

Hereditary hemorrhagic telangiectasia (HHT) is transmitted as an autosomal dominant disease and has an estimated incidence of one in 5,000 to 8,000 persons.[93] Recent linkage analyses have identified at least three HHT loci, including the genes for endoglin and activinlike receptor kinase. Both proteins are expressed on vascular endothelial cells and may function as receptors for transforming growth factor–β (TGF-β). TGF-β plays a complex role in coordinating responses between endothelial cells and the extracellular matrix. A mutation in the gene for either endoglin or activinlike receptor kinase results in a 50% reduction in the normal quantity of protein on endothelial cells (haploinsufficiency) and leads to the development of abnormal blood vessels and arteriovenous malformations (AVMs).[94]

HHT generally does not present at birth but manifests itself with age. Recurrent epistaxis is typically the earliest sign of disease, commonly occurring in childhood. Pulmonary AVMs, occurring in about 30% of HHT patients, become apparent after puberty and may present as dyspnea, chest pain, and hemoptysis. Physical examination may reveal chest bruits and digital clubbing. Mucocutaneous telangiectasias, which occur in the majority of patients and at characteristic sites (e.g., lips, oral cavity, fingers, and nose), typically become noticeable by the third decade of life and increase in size and number as the patient ages. The diagnosis of HHT should be based on four criteria: (1) spontaneous and recurrent epistaxis, (2) multiple mucocutaneous telangiectasias, (3) evidence of visceral telangiectasias and AVMs (e.g., gastrointestinal tract, pulmonary, hepatic, or cerebral AVMs), and (4) a positive family history of HHT.

Coagulation test results are generally normal. The pulmonary AVMs may be associated with hypoxemia and secondary polycythemia. The diagnosis can be confirmed by pulmonary angiography. If the shunts are large and clinically significant, they can be treated by balloon embolotherapy.[95] Paradoxical embolus with stroke can occur in patients with HHT who have pulmonary arteriovenous shunts and malformations. It has been advocated that asymptomatic HHT patients be screened for AVMs and be treated prophylactically with embolotherapy if AVMs are found.[96] Management of recurrent epistaxis should be conservative and often involves devising methods for obtaining nasal tamponade. Cauterization should be avoided because damage to nasal mucosa may result in vascular regrowth. Gastrointestinal bleeding is managed by the use of iron preparations when possible.

Table 6 Vascular Purpuras

Congenital disorders	Hereditary hemorrhagic telangiectasia Hereditary connective tissue disorders Ehlers-Danlos syndrome Marfan syndrome Pseudoxanthoma elasticum
Acquired disorders	Scurvy Amyloidosis Glucocorticoid excess Immunoglobulin disorders Waldenström macroglobulinemia Cryoglobulinemia Hepatitis B Henoch-Schönlein purpura Embolism (i.e., microvascular damage from septic and bland emboli from heart valves, fat embolism, and cholesterol embolism) Senile purpura

ACQUIRED DISORDERS OF BLOOD VESSELS THAT CAUSE BLEEDING

Scurvy

Vitamin C is required for the normal metabolism of collagen, folate, and perhaps iron. The patient with scurvy suffers primarily from impaired collagen synthesis. The lack of proper collagen support for the microvasculature leads to perifollicular hemorrhages, bleeding gums, and even deep tissue hematomas. Presumably, similar collagen defects lead to the so-called corkscrew hair and hyperkeratosis associated with this disorder.[97] The characteristic clinical picture in a malnourished person suggests the diagnosis. Plasma or buffy coat levels of ascorbic acid are low, and other vitamin deficiencies are usually present. Effective therapy consists of 1 g of ascorbic acid daily in divided doses.

Glucocorticoid Excess

Glucocorticoid excess, whether from endogenous or exogenous causes, produces cutaneous hemorrhages, probably because of glucocorticoid-induced catabolism of protein in vascular supportive tissues.

Amyloidosis

Amyloidosis can present as subcutaneous ecchymoses that have a predilection for the neck and upper chest. Biopsy of the site shows the amyloid, which by its infiltration may weaken the vessel walls or interfere with surface activation of platelets, procoagulants, or both. In patients with primary systemic amyloidosis, especially when accompanied by a huge amyloid spleen, the amyloid can in some instances adsorb enough factor X to cause profound factor X deficiency and clinical bleeding. Infusions of fresh frozen plasma are usually ineffective. Recombinant factor VIIa is effective, but it is extremely expensive and provides only temporary benefit. Splenectomy may provide long-term benefit and should be considered in a patient with recurrent serious clinical bleeding.

Immunoglobulin Disorders

The purpuric lesions in patients with immunoglubulin disorders (e.g., cryoglobulinemia, Waldenström macroglobulinemia, Henoch-Schönlein purpura, and multiple myeloma) may be raised (palpable purpura). On biopsy, these lesions may show mast cell degranulation and, when stained appropriately, immune complex deposition. Presumably, the immune complexes provide the chemotactic stimulation that leads to the congregation of neutrophils. Damage to the microvasculature is caused by the complement attack complex and by the release of the contents of the neutrophil granules. This inflammatory component produces the palpable purpura.

Damage to the Microvasculature Due to Emboli

DIC and TTP can cause localized vaso-occlusions leading to microvascular damage and leakage of red blood cells. Similar damage can be caused by emboli that arise from infected heart valves. Fat embolism may complicate fractures of the long bones and pelvis. The syndrome consists of fever; confusion; and petechiae, purpura, or both over the neck, chest, face, and axillae. Cholesterol embolism can also cause petechiae, usually over the lower extremities. It typically occurs in a patient with severe ath-

erosclerosis who has recently undergone an invasive procedure involving the abdominal aorta or renal arteries. Biopsy of the purpura shows cholesterol crystals when an appropriate stain is used.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Ginsburg D: Molecular genetics of von Willebrand disease. Thromb Haemost 82:585, 1999

2. George JN, el-Harake MA, Raskob GE: Chronic idiopathic thrombocytopenic purpura. N Engl J Med 331:1207, 1994

3. Hirsh J, Dalen JE, Anderson DR, et al: Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. Chest 114:445S, 1998

4. Gewirtz AM, Hoffman R: Transitory hypomegakaryocytic thrombocytopenia: aetiological association with ethanol abuse and implications regarding regulation of human megakaryocytopoiesis. Br J Haematol 62:333, 1986

5. Tepler I, Elias L, Smith JW II, et al: A randomized placebo-controlled trial of recombinant human interleukin-11 in cancer patients with severe thrombocytopenia due to chemotherapy. Blood 87:3607, 1996

6. Kuter DJ, Begley CG: Recombinant human thrombopoietin: basic biology and evaluation of clinical studies. Blood 100:3457, 2002

7. Peck-Radosavljevic M, Wichlas M, Zacherl J, et al: Thrombopoietin induces rapid resolution of thrombocytopenia after orthotopic liver transplantation through increased platelet production. Blood 95:795, 2000

8. Frederiksen H, Schmidt K: The incidence of idiopathic thrombocytopenic purpura in adults increases with age. Blood 94:909, 1999

9. Cines DB, McMillan R: Management of adult idiopathic thrombocytopenic purpura. Annu Rev Med 56:425, 2005

10. Hofbauer LC, Heufelder AE: Coagulation disorders in thyroid diseases. Eur J Endocrinol 136:1, 1997

11. Stasi R, Stipa E, Masi M, et al: Prevalence and clinical significance of elevated antiphospholipid antibodies in patients with idiopathic thrombocytopenic purpura. Blood 84:4203, 1994

12. Chan H, Moore JC, Finch CN, et al: The IgG subclasses of platelet-associated autoantibodies directed against platelet glycoproteins IIb/IIIa in patients with idiopathic thrombocytopenic purpura. Br J Haematol 122:818, 2003

13. He R, Reid DM, Jones CE, et al: Spectrum of Ig classes, specificities, and titers of serum antiglycoproteins in chronic idiopathic thrombocytopenic purpura. Blood 83:1024, 1994

14. Emmons RV, Reid DM, Cohen RL, et al: Human thrombopoietin levels are high when thrombocytopenia is due to megakaryocyte deficiency and low when due to increased platelet destruction. Blood 87:4068, 1996

15. George JN, Woolf SH, Raskob GE, et al: Idiopathic thrombocytopenic purpura: a practice guideline developed by explicit methods for the American Society of Hematology. Blood 88:3, 1996

16. Stasi R, Stipa E, Masi M, et al: Long term observation of 208 adults with chronic idiopathic thrombocytopenic purpura. Am J Med 98:436, 1995

17. Cortelazzo S, Finazzi G, Buelli M, et al: High risk of severe bleeding in aged patients with chronic idiopathic thrombocytopenic purpura. Blood 77:31, 1991

18. Stasi R, Rossi Z, Stipa E, et al: *Helicobacter pylori* eradication in the management of patients with idiopathic thrombocytopenic purpura. Am J Med 118:414, 2005

19. Cheng Y, Wong RS, Soo YO, et al: Initial management of immune thrombocytopenic purpura with high-dose dexamethasone. N Engl J Med 349:831, 2003

20. Berchtold P, Dale GL, Tani P, et al: Inhibition of autoantibody binding to platelet glycoprotein IIb/IIIa by anti-idiotypic antibodies in intravenous gammaglobulin. Blood 74:2414, 1989

21. Yu Z, Lennon VA: Mechanism of intravenous immune globulin therapy in antibody-mediated autoimmune diseases. N Engl J Med 340:227, 1999

22. Cooper N, Stasi R, Cunningham-Rundles S, et al: The efficacy and safety of B-cell depletion with anti-CD20 monoclonal antibody in adults with chronic immune thrombocytopenic purpura. Br J Haematol 125:232, 2004

23. Facon T, Caulier MT, Wattel E, et al: A randomized trial comparing vinblastine in slow infusion and by bolus I.V. injection in idiopathic thrombocytopenic purpura: a report on 42 patients. Br J Haematol 86:678, 1994

24. Laveder F, Marcolongo R, Zamboni S: Thrombocytopenic purpura following treatment with danazol. Br J Haematol 90:970, 1995

25. Nardi MA, Liu L-X, Karpatkin S: GPIIIa-(49-66) is a major pathophysiologically relevant antigenic determinant for anti-platelet GPIIIa of HIV-1-related immunologic thrombocytopenia. Proc Natl Acad Sci USA 94:7589, 1997

26. Koefoed K, Ditzel HJ: Identification of talin head domain as an immunodominant epitope of the antiplatelet antibody response in patients with HIV-1–associated thrombocytopenia. Blood 104:4054, 2004

27. Najean Y, Rabin JD: The mechanism of thrombocytopenia in patients with HIV infection. J Lab Clin Med 123:415, 1994

28. Scaradavou A, Woo B, Woloski BMR, et al: Intravenous anti-D treatment of immune thrombocytopenic purpura: experience in 272 patients. Blood 89:2689, 1997

29. Durand JM, Lefevre P, Hovette P, et al: Dapsone for thrombocytopenic purpura related to human immunodeficiency virus infection. Am J Med 90:675, 1991

30. Marroni M, Gresele P, Landonio G, et al: Interferon-alpha is effective in the treatment of HIV-1–related, severe, zidovudine-resistant thrombocytopenia. Ann Intern Med 121:423, 1994

31. Needleman SW, Sorace J, Poussin-Rosillo H, et al: Low-dose splenic irradiation in the treatment of autoimmune thrombocytopenia in HIV-infected patients. Ann Intern Med 116:310, 1992

32. Kelton JG: Idiopathic thrombocytopenic purpura complicating pregnancy. Blood Reviews 16:43, 2002

33. Christiaens GC, Nieuwenhuis HK, Bussel JB: Comparison of platelet counts in first and second newborns of mothers with immune thrombocytopenic purpura. Obstet Gynecol 90:546, 1997

34. Burrows RF, Kelton JG: Fetal thrombocytopenia and its relation to maternal thrombocytopenia. N Engl J Med 329:1463, 1993

35. Webert KE, Mittal R, Sigouin C, et al: A retrospective 11-year analysis of obstetric patients with idiopathic thrombocytopenic purpura. Blood 102:4306, 2003

36. Payne SD, Resnik R, Moore TR, et al: Maternal characteristics and risk of severe neonatal thrombocytopenia and intracranial hemorrhage in pregnancies complicated by autoimmune thrombocytopenia. Am J Obstet Gynecol 177:149, 1997

37. Pappalardo PA, Secord AR, Quitevis P, et al: Platelet transfusion refractoriness associated with HPA-1a (Pl (A1)) alloantibody without coexistent HLA antibodies successfully treated with antigen-negative platelet transfusions. Transfusion 41:984, 2001

38. Curtis BR, Ali S, Glazier AM, et al: Isoimmunization against CD36 (glycoprotein IV): description of four cases of neonatal isoimmune thrombocytopenia and brief review of the literature. Transfusion 42:1173, 2002

39. McCrae KR, Herman JH: Posttransfusion purpura: two unusual cases and a literature review. Am J Hematol 52:205, 1996

40. Loren AW, Abrams CS: Efficacy of HPA-1a (PIA1)-negative platelets in a patient with post-transfusion purpura. Am J Hematol 76:258, 2004

41. Burgess JK, Lopez JA, Berndt MC, et al: Quinine-dependent antibodies bind a restricted set of epitopes on the glycoprotein Ib-IX complex: characterization of the epitopes. Blood 92:2366, 1998

42. Glynne P, Salama A, Chaudhry A, et al: Quinine-induced immune thrombocytopenic purpura followed by hemolytic uremic syndrome. Am J Kidney Dis 33:133, 1999

43. Kaufman DW, Kelly JP, Johannes CB, et al: Acute thrombocytopenic purpura in relation to the use of drugs. Blood 82:2714, 1993

44. Garner SF, Campbell K, Metcalfe P, et al: Glycoprotein V: the predominant target antigen in gold-induced autoimmune thrombocytopenia. Blood 100:344, 2002

45. Burday MJ, Martin SE: Cocaine-associated thrombocytopenia. Am J Med 91:656, 1992

46. Tcheng JE, Kereiakes DJ, Lincoff AM, et al: Abciximab readministration: results of the ReoPro Readministration Registry. Circulation 104:870, 2001

47. Aster RH: Immune thrombocytopenia caused by glycoprotein IIb/IIIa inhibitors. Chest 127:53S, 2005

48. Bougie D, Aster R: Immune thrombocytopenia resulting from sensitivity to metabolites of naproxen and acetaminophen. Blood 97:3846, 2001

49. Bennett CL, Connors JM, Carwile JM, et al: Thrombotic thrombocytopenic purpura associated with clopidogrel. N Engl J Med 342:1773, 2000

50. Moake JL: Thrombotic microangiopathies. N Engl J Med 347:589, 2002

51. Levy G, Motto DG, Ginsburg D: ADAMTS13 turns 3. Blood 106:11, 2005

52. Dong JF, Moake JL, Bernardo A, et al: ADAMTS-13 metalloprotease interacts with the endothelial cell-derived ultra-large von Willebrand factor. J Biol Chem 278:29633, 2003

53. Levy GG, Nichols WC, Ian EC, et al: Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. Nature 413:488, 2001

54. Furlan M, Robles R, Galbusera M, et al: von Willebrand factor–cleaving protease in thrombotic thrombocytopenic purpura and the hemoloytic-uremic syndrome. N Engl J Med 339:1578, 1998

55. Tsai HM, Lian EC, et al: Antibodies to von Willebrand factor–cleaving protease in acute thrombotic thrombocytopenic purpura. N Engl J Med 339:1585, 1998

56. Mannucci PM, Canciani MT, Forza I, et al: Changes in health and disease of the metalloprotease that cleaves von Willebrand factor. Blood 98:2730, 2001

57. Moore JC, Hayward CP, Warkentin TE, et al: Decreased von Willebrand factor protease activity associated with thrombocytopenic disorders. Blood 98:1842, 2001

58. Zheng XL, Kaufman RM, Goodnough LT, et al: Effect of plasma exchange on plasma ADAMTS13 metalloprotease activity, inhibitor level, and clinical outcome in patients with idiopathic and nonidiopathic thrombotic thrombocytopenic purpura. Blood 103:4043, 2004

59. Veyradier A, Obert B, Houllier A, et al: Specific von Willebrand factor–cleaving protease in thrombotic microangiopathies: a study of 111 cases. Blood 98:1765, 2001

60. Vesley SK, George JN, Lammle B, et al: ADAMTS13 activity in thrombotic thrombocytopenic purpura–hemolytic uremic syndrome: relation to presenting features and clinical outcomes in a prospective cohort of 142 patients. Blood 102:60, 2003

61. Kokame K, Matsumoto M, Fujimura Y, et al: VWF73, a region from D1596 to R1668 of von Willebrand factor, provides a minimal substrate for ADAMTS-13. Blood 103:607, 2004

62. Whitelock JL, Nolasco L, Bernardo A, et al: ADAMTS-13 activity in plasma is rapidly measured by a new ELISA method that uses recombinant VWF-A2 domain as substrate. J Thromb Haemost 2:485, 2004

63. Boyce TG, Swerdlow DL, Griffin PM: *Escherichia coli* 0157:H7 and the hemolytic-uremic syndrome. N Engl J Med 333:364, 1995

64. Morigi M, Galbusera M, Binda E, et al: Verotoxin-1–induced up-regulation of adhesive molecules renders microvascular endothelial cells thrombogenic at high shear stress. Blood 98:1828, 2001

65. Wong CS, Jelacic S, Habeeb RL, et al: The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N Engl J Med 342:1930, 2000

66. Bianchi V, Robles R, Alberio L, et al: Von Willebrand factor–cleaving protease (ADAMTS13) in thrombocytopenic disorders: a severely deficient activity is specific for thrombotic thrombocytopenic purpura. Blood 100:710, 2002

67. Rock G, Shumak KH, Buskard NA, et al: Comparison of plasma exchange with plasma infusion in the treatment of thrombotic thrombocytopenic purpura. Canadian Apheresis Study Group. N Engl J Med 325:393, 1991

68. George JN, Gilcher RO, Smith JW, et al: Thrombotic thrombocytopenic purpura–hemolytic uremic syndrome: diagnosis and management. J Clin Apheresis 13:120, 1998

69. Rock G, Shumak KH, Sutton DM, et al: Cryosupernatant as replacement fluid for plasma exchange in thrombotic thrombocytopenic purpura. Members of the Canadian Apheresis Group. Br J Haematol 94:383, 1996

70. Bell WR, Braine HG, Ness PM, et al: Improved survival in thrombotic thrombocytopenic purpura–hemolytic uremic syndrome: clinical experience in 108 patients. N Engl J Med 325:398, 1991

71. Zheng X, Pallera AM, Goodnough LT, et al: Remission of chronic thrombotic thrombocytopenic purpura after treatment with cyclophosphamide and rituximab. Ann Intern Med 138:105, 2003

72. Yomtovian R, Niklinski W, Silver B, et al: Rituximab for chronic recurring thrombotic thrombocytopenic purpura: a case report and review of the literature. Br J Haematol 124:787, 2004

73. Fakhouri F, Vernant JP, Veyradier A, et al: Efficiency of curative and prophylactic treatment with rituximab in ADAMTS13 deficient–thrombotic thrombocytopenic purpura: a study of 11 cases. Blood 2005

74. Harkness DR, Byrnes JJ, Lian EC-Y: Hazard of platelet transfusion in thrombotic thrombocytopenic purpura. JAMA 246:1931, 1981

75. Sadler JE, Moake JL, Miyata T, et al: Recent advances in thrombotic thrombocytopenic purpura. Hematology 407, 2004

76. Shumak KH, Rock G, Nair RC, et al: Late relapses in patients successfully treated for thrombotic thrombocytopenic purpura. Ann Intern Med 122:569, 1995

77. Mohanty D, Marwaha N, Ghosh K, et al: Functional and ultrastructural changes of platelets in malarial infection. Trans R Soc Trop Med Hyg 82:369, 1988

78. Van Dam PA, Renier M, Baekelandt M, et al: Disseminated intravascular coagulation and the syndrome of hemolysis, elevated liver enzymes, and low platelets in severe preeclampsia. Obstet Gynecol 73:97, 1989

79. Martin JN Jr, Blake PG, Perry KG Jr, et al: The natural history of HELLP syndrome: patterns of disease progression and regression. Am J Obstet Gynecol 164:1500, 1991

80. Sibai BM, Ramadan MK, Usta I, et al: Maternal morbidity and mortality in 442 pregnancies with hemolysis, elevated liver enzymes, and low platelets (HELLP syndrome). Am J Obstet Gynecol 169:1000, 1993

81. Sibai BM: Diagnosis, controversies, and management of the syndrome of hemolysis, elevated liver enzymes, and low platelet count. Obstet Gynecol 103:981, 2004

82. Boccardo P, Remuzzi G, Galbusera M: Platelet dysfunction in renal failure. Semin Thromb Hemost 30:579, 2004

83. Weigert AL, Schafer AI: Uremic bleeding: pathogenesis and therapy. Am J Med Sci 316:94, 1998

84. Mannucci PM: Hemostatic drugs. N Engl J Med 339:245, 1998

85. Lacoste L, Hung J, Lam JY: Acute and delayed antithrombotic effects of alcohol in humans. Am J Cardiol 87:82, 2001

86. George JN, Shattil SJ: The clinical importance of acquired abnormalities of platelet function. N Engl J Med 324:27, 1991

87. Baxter EJ, Scott LM, Campbell PJ, et al: Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. Lancet 365:1954, 2005

88. James C, Ugo V, Le Couedic JP, et al: A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 434:1144, 2005

89. Kralovics R, Passamonti F, Buser AS, et al: A gain-of-function mutation of JAK2 in myeloproliferative disorders. N Engl J Med 352:1779, 2005

90. Layzer RB: Hot feet: erythromelalgia and related disorders. J Child Neurol 16:199, 2001

91. Anagrelide, a therapy for thrombocythemic states: experience in 577 patients. Anagrelide Study Group. Am J Med 92:69, 1992

92. Harrison CN, Campbell PJ, Buck G, et al: Hydroxyurea compared with anagrelide in high-risk essential thrombocytopenia. N Engl J Med 353:33, 2005

93. Begbie ME, Wallace GM, Shovlin CL: Hereditary haemorrhagic telangiectasia (Osler-Weber-Rendu syndrome): a view from the 21st century. Postgrad Med J 79:18, 2003

94. Abdalla SA, Pece-Barbara N, Vera S, et al: Analysis of ALK-1 and endoglin in newborns from families with hereditary telangiectasia type 2. Hum Mol Genet 9:1227, 2000

95. White RI Jr, Lynch-Nyhan A, Terry P, et al: Pulmonary arteriovenous malformation: techniques and long-term outcome of embolotherapy. Radiology 169:663, 1988

96. Shovlin CL: Molecular defects in rare bleeding disorders: hereditary haemorrhagic telangiectasia. Thromb Haemost 78:145, 1997

97. Hirschmann JV, Raugi GJ: Adult scurvy. J Am Acad Dermatol 41:895, 1999

98. George J: Hemostasis and thrombosis. Hematology: Basic Principles and Practice, 3rd ed. Hoffman R, Benz EJ Jr, Shattil SJ, Eds. Churchill Livingstone, Philadelphia, 1999, p 1928

Acknowledgment

Figure 1 Seward Hung.

XIV THROMBOTIC DISORDERS

Lawrence L. K. Leung, m.d.

Thrombosis is more than excessive blood clotting; it also involves vascular inflammation. The classic triad of Virchow identifies three major elements in the pathophysiology of thrombosis: endothelial injury, a decrease in blood flow, and an imbalance between procoagulant and anticoagulant factors.

Endothelial cells can be activated or injured by a variety of stimuli, including mechanical trauma, endotoxins and cytokines, proteases, inflammatory mediators, immune complex deposition, oxygen radicals, and hypoxia. Each of these stimuli affects multiple facets of endothelial cell function, ultimately changing the cell from its natural antithrombotic state to a prothrombotic one.

The vascular endothelium, in its unique location in the vessel wall, is capable of sensing and responding to the different mechanical forces in the blood circulation. The shear stress caused by the friction from blood flow seems to be particularly important in modulating endothelial functions. In areas of linear flow, the blood moves in ordered laminar patterns in a regular pulsatile fashion. Such a steady, laminar blood flow apparently promotes an antithrombotic endothelial phenotype. In areas of disrupted flow, such as at vascular bifurcations or stenoses, the endothelium may be exposed to significant changes in shear gradients, and the cells may become activated and prothrombotic.

Imbalance between procoagulant and anticoagulant factors can be hereditary or acquired [*see Table 1*]. Some clotting factors, such as factor VIII and fibrinogen, are acute-phase reactants: their plasma levels increase significantly with acute inflammation, possibly conferring a transient prothrombotic state. Some hereditary deficiencies of anticoagulant proteins, such as factor V Leiden and antithrombin (AT), are associated with recurrent thrombosis; these are among the best understood clinical hypercoagulable states.

Although the hypercoagulable state is systemic, thrombosis occurs locally (e.g., in the lower extremities). The clinical outcome likely reflects a complex interaction between the systemic prothrombotic predisposition and local hemostatic control mechanisms specific to the vascular bed. Specific and distinctive gene transcript expression patterns in different vascular beds have now been demonstrated.[1]

Assessment of Patients with Thrombotic Disorders

PRIMARY CLINICAL ISSUES IN THROMBOTIC DISORDERS

Important questions in the assessment of thrombosis include the following: (1) How likely is it that the thrombosis is caused by an underlying hypercoagulable state? (2) How extensive a workup is indicated? (3) When should the workup be done? Answers to these questions come from a consideration of the patient's age at the time of the first thrombosis; presence or absence of a provoking factor; family history and past medical history of response to situations associated with high risk of thrombosis; and the site, type, and severity of the thrombosis.

Age of Onset of First Thrombosis

In a retrospective study involving 150 families with an inherited predisposition to recurrent thrombosis (thrombophilia), the mean age at the time of the first thrombosis was 35 to 40 years. However, the first episode of thrombosis can occur as early as the second decade of life if the patient has more than one hereditary risk factor.[2]

Presence or Absence of a Provoking Factor

Common triggers of thrombosis are surgery, trauma, pregnancy, malignancy, prolonged immobilization, and infection. Malignancy or infection can be clinically overt or subclinical. Those circumstances can provoke thrombosis even in persons with a normal coagulation system, and they often uncover a thrombophilia that had been clinically silent.[2] However, sometimes no provoking factors can be identified. Such a spontaneous, idiopathic thrombosis, especially when it occurs in a young person, strongly suggests an underlying hereditary hypercoagulable state.

If thrombosis develops in a patient who has had previous pregnancies or surgeries (especially orthopedic procedures)

Table 1 Inherited and Acquired Hypercoagulable States

- *Inherited*
  - Resistance to activated protein C/factor V Leiden
  - Prothrombin gene mutation 20210A
  - Antithrombin III deficiency
  - Protein C deficiency
  - Protein S deficiency
  - Hyperhomocysteinemia
- *Acquired*
  - Antiphospholipid antibody syndrome
  - Hypercoagulable state associated with physiologic or thrombogenic stimuli:
    - Advancing age
    - Oral contraceptives
    - Pregnancy
    - Surgery
    - Trauma
  - Hypercoagulable state associated with other clinical conditions:
    - Malignancy—Trousseau syndrome
    - Heparin-induced thrombocytopenia with thrombosis
    - Nephrotic syndrome
    - Hyperviscosity (polycythemia vera, Waldenström macroglobinemia, multiple myeloma)
    - Myeloproliferative disorders (polycythemia vera, essential thrombocythemia)
    - Paroxysmal nocturnal hemoglobinuria
    - Sickle cell anemia
- *Rare or not well established*
  - Dysfibrinogenemia
  - Hypoplasminogenemia, dysplasminogenemia
  - Abnormal thrombomodulin
  - Factor XII deficiency
  - Elevated factor VII, factor VIII, fibrinogen, lipoprotein(a), plasminogen activator inhibitor–1

Table 2 Screening Tests for Patients with Suspected Hypercoagulable State

Underlying State	*Laboratory Evaluation*
Venous thrombosis	Resistance to activated protein C Factor V Leiden (genetic test) Clotting assay (unnecessary if the genetic test for factor V Leiden is positive) Prothrombin mutation 20210A (genetic test) Antithrombin III (functional assay) Protein C (functional assay) Protein S Functional assay Antigenic assay for free protein S
Arterial thrombosis	Antibodies associated with heparin-induced thrombocytopenia* Chronic disseminated intravascular coagulation (Trousseau syndrome)* Lipoprotein(a)
Venous thrombosis and/or arterial thrombosis	Plasma homocysteine Fasting level Level after methionine loading (if thrombophilia is strongly suspected) Antiphospholipid antibody Clotting assays for lupuslike anticoagulant ELISA for anticardiolipin antibodies IgG and IgM Dysfibrinogenemia (if thrombophilia is strongly suspected) Functional assay for fibrinogen level Thrombin time, reptilase time

*In appropriate clinical settings.
ELISA—enzyme-linked immunosorbent assay

without any thrombotic complications, an acquired hypercoagulable state should be considered. Likely conditions in such cases are antiphospholipid antibody syndrome or Trousseau syndrome (see below).

Family History

Objectively documented venous thromboembolism before 50 years of age in a first-degree family member strongly suggests a hereditary thrombotic disorder. However, a negative family history does not exclude a hereditary condition. Clinical thrombosis is frequently the culmination of multiple thrombogenic risk factors, only one of which may be hereditary and irreversible. In patients with symptomatic thrombosis and well-documented hereditary hypercoagulable states, it is not uncommon to find other family members with the same deficiency but no clinical thrombosis.

Recurrent Thrombosis

A patient who experiences recurrent thrombosis likely has a hypercoagulable state (hereditary or acquired). However, if a patient who initially presented with deep vein thrombosis (DVT) in a lower extremity returns with symptoms involving the same leg, the problem may be postphlebitic syndrome rather than recurrent thrombosis. Acute exacerbation of the postphlebitic syndrome, with its increased leg edema and pain, can be difficult to distinguish from recurrent acute DVT. As an anticipatory measure, it is sometimes useful to obtain a repeat compression ultrasound study of the lower extremity after resolution of an acute episode of DVT; the repeat scan can provide a baseline for future comparison.

Site of Thrombosis

Most commonly, thromboses involve the deep veins of the lower extremities. Thrombosis at an atypical site, such as the hepatic, mesenteric, or cerebral veins (or skin necrosis after warfarin administration), increases the likelihood of an underlying hypercoagulable state. Spontaneous axillary vein thrombosis may also indicate the presence of an underlying hypercoagulable state, but this association is controversial.[2-4]

Recurrent thrombosis at arterial sites has a differential diagnosis—and therefore a workup—that is quite different from that for recurrent venous thrombosis [*see Table 2*]. Most of the common hereditary hypercoagulable states (e.g., AT deficiency or factor V Leiden) are associated with venous thromboses, such as DVT in the lower extremities. They are seldom associated with arterial thromboses, such as transient ischemic attack, stroke, digital ischemia, and myocardial infarction. A few hypercoagulable states, such as the antiphospholipid syndrome and hyperhomocysteinemia, are associated with both types of thrombosis.

The Hypercoagulable Workup

Extent of the workup On the basis of the above clinical considerations, one may estimate the likelihood of an underlying thrombophilia in a given patient with thrombosis [*see Table 3*]. Because studies of cost-effectiveness and outcomes are not available, it is difficult to list strict practice guidelines regarding the extent of the hypercoagulable workup. In general, however, if the likelihood of an underlying hypercoagulable state is high, an extensive workup is warranted.

A limited workup is appropriate for mild to moderate DVT of the lower extremities with an obvious provoking factor. For example, in a young woman who experiences DVT in the superficial femoral vein while on an oral contraceptive, evaluation of factor V Leiden, prothrombin mutation 20210A, AT, protein C, protein S, homocysteine, anticardiolipin antibodies, and lupus anticoagulant may be sufficient. On the other hand, an acquired hypercoagulable state should be considered in an elderly patient with a spontaneous DVT and no history of previous thrombosis. Diagnostic possibilities in such cases would include antiphospholipid antibody syndrome, acquired AT deficiency (if the patient has evidence of nephrotic syndrome), or Trousseau syndrome.

When the clinical history strongly suggests thrombophilia—as in a patient with recurrent thrombosis or thrombosis at atypical sites—one may argue that a workup for an underlying hypercoagulable state is unnecessary because the result will not alter the management of the case. However, the identification of any underlying risk factors will improve the understanding of the disease for both the patient and the treating physician; and it will guide the counseling of the patient, especially regarding the need for screening of related family members. In the case of hyperhomocysteinemia and elevated lipoprotein(a) levels, identifi-

Table 3 Clinical Features That Suggest Thrombophilia

Age at onset of first thrombosis < 50 yr
No identifiable risk factor
Positive family history
Recurrent thrombosis
Atypical site of thrombosis

cation of risk factors will permit the use of specific therapies (see below).

Timing of the workup The clinician needs to know not only what tests to order but when to order them. In acute thrombosis, many inhibitors of the clotting cascade (e.g., AT and protein C) are consumed. Immediately after the episode, their plasma levels may be decreased, even in patients who do not have a hereditary deficiency. Heparin therapy can reduce antithrombin levels up to 20%, whereas warfarin treatment reduces the levels of protein C and protein S. Usually it is best to postpone measurement of these inhibitors until the acute thrombotic episode is completely resolved, preferably 4 weeks after termination of oral anticoagulation therapy. Tests for specific genotypes (e.g., factor V Leiden) can be performed at any time, however.

Frequency and Relative Risk of Venous Thromboembolism

The frequency of various hypercoagulable states in unselected patients who present with venous thrombosis ranges from 1% to 25% [*see Table 4*]. It should be recognized that these thrombophilias do not confer equivalent thrombotic risk. Factor V Leiden and prothrombin mutation 20210A, the two most prevalent risk factors, confer only a modest increase in relative risk of thrombosis, approximately threefold to sevenfold above normal. Moderate hyperhomocysteinemia, another common risk factor, also carries a modest increase in risk. Heterozygous deficiencies of AT, protein C, and protein S are generally considered more significant risk factors than factor V Leiden. AT deficiency and the antiphospholipid antibody syndrome are probably the greatest risk factors. The recurrence rate in patients with antiphospholipid antibody syndrome is as high as 50% to 70% in some studies.

Patients with symptomatic thrombosis frequently have more than one risk factor, which may have a synergistic effect in increasing the thrombosis risk. For example, women with factor V Leiden who use oral contraceptives have a risk of venous thromboembolism that is 35-fold higher than that in the general population.

Table 4 Frequency and Relative Risk of Venous Thrombosis in Selected Hypercoagulable States*

Condition	*Relative Risk†*	*Frequency‡*
Antithrombin III deficiency	High	1%–2%
Protein C deficiency	High	3%–4%
Protein S deficiency	High	2%–3%
Factor V Leiden	Modest	20%–25%
Prothrombin mutation 20210A	Modest	10%
Hyperhomocysteinemia	Modest	10%
Oral contraceptive use	Modest	NA

*The incidence of venous thromboembolism in the normal population is estimated to be 0.008% a year (0.03% a year in patients who take oral contraceptives).
†Modest risk is defined as an approximate 2.5-fold to fivefold increase in thromboembolism, on the basis of data from the Leiden Thrombophilia Study[146]; high risk is defined as an approximate threefold to fourfold increase over that for factor V Leiden.[2]
‡In unselected patients with venous thromboembolism.

Hereditary Hypercoagulable States

ANTITHROMBIN DEFICIENCY

Epidemiology and Etiology

The frequency of symptomatic inherited AT deficiency in the general population has been estimated to be approximately 1 per 2,000 people.[5] The deficiency is transmitted in an autosomal dominant pattern. Homozygous AT deficiency has not been reported, presumably because the condition is incompatible with normal fetal development.

There are two types of inherited AT deficiency. Type I is quantitative, as measured by antigenic and functional assays. A large number of molecular mutations have been characterized in type I AT deficiency, including partial gene deletions and single-nucleotide substitutions that cause nonsense or missense mutations leading to premature stop signals in the protein-translation process.

Type II deficiency is qualitative; plasma levels of AT antigen are normal. The underlying defect is generally a single nucleotide change that causes missense mutations, giving rise to a dysfunctional protein. Many of these proteins have decreased affinity for heparin binding.

In rare cases, AT deficiency is acquired. This condition may occur after administration of intravenous heparin for more than 3 days or after asparaginase therapy. It may also develop in patients with disseminated intravascular coagulation (DIC), severe liver disease, or the nephrotic syndrome.

Pathophysiology

AT inactivates factor Xa and thrombin by forming a stable stoichiometric complex with each of them. AT is present in sufficient amounts in plasma to inactivate all the thrombin formed in a given plasma volume, but it does so slowly unless it is activated by endothelial cell surface heparan sulfate or by administered heparin [*see 5:XIII Hemorrhagic Disorders*]. Patients with hereditary AT deficiency have evidence of continuous factor X activation and thrombin generation (as supported by elevated plasma levels of prothrombin fragment F1.2) even when they are clinically asymptomatic.

Clinical Presentation

Patients with AT deficiency show an increased incidence of venous thrombosis, usually triggered by a prothrombotic stimulus such as surgery, infection, immobilization, or trauma. This association suggests that the superimposition of a prothrombotic stimulus on an underlying subclinical hypercoagulable state leads to clinical thrombosis.

Typical clinical presentations are DVT of the legs, pulmonary embolism, and occasionally mesenteric vein thrombosis. There is no convincing evidence to suggest that AT deficiency increases the risk of arterial thrombosis.[6]

Affected patients usually have a family history of recurrent thromboses, generally beginning in youth and often associated with surgery or trauma. Pregnancy and the use of oral contraceptives also increase the risk of thromboses in AT-deficient patients. The tendency to thrombosis increases with advancing age: by age 50, only 10% of AT-deficient patients are free of symptoms.

Diagnosis

The AT level should be determined by a functional assay rather than an antigenic assay, so that both type I and type II de-

ficiency can be evaluated. Patients with AT deficiency have a surprisingly modest reduction in the protein: values measured by both bioassay and immunoassay range from 25% to 60% of normal in type I disease.

Treatment

Study of a large AT-deficient kindred indicates that long-term anticoagulant prophylaxis is not warranted in asymptomatic carriers of this deficiency.[7] Asymptomatic carriers should receive prophylactic anticoagulation only in situations known to increase the risk of thrombogenesis, such as abdominal surgery.[7] Once such patients have experienced a thrombotic event, however, they probably require lifelong warfarin therapy. Warfarin is the mainstay of long-term therapy for patients with AT deficiency and recurrent thromboembolism.

Acute episodes of thrombosis must be treated with heparin. Because AT deficiency may render heparin relatively ineffective, the physician should be alert to heparin resistance. In patients receiving unfractionated heparin, resistance is manifested by minimal prolongation of the partial thromboplastin time (PTT) despite the administration of therapeutic doses. If low-molecular-weight heparin (LMWH) is used, as is commonly the case, the level of anti–factor Xa (anti-FXa) should be checked to ensure that a therapeutic anticoagulant effect is achieved [*see 5:XII Hemostasis and Its Regulation*].

If heparin resistance occurs despite increased doses of heparin, heparin plus purified AT concentrates or fresh frozen plasma should be given. AT has a half-life of about 60 hours. These preparations can be used to carry an AT–deficient patient through surgery or delivery and should bring the AT level up to nearly 100%, depending on the patient's baseline AT level. The AT level should be checked and the infusion repeated at 24-hour intervals to maintain a normal AT level for 5 to 7 days after delivery or surgery.

Pregnancy in an AT–deficient patient is difficult to manage. Because warfarin may cause fetal malformations and neonatal hemorrhage, patients should be treated with full-dose unfractionated heparin or LMWH; those receiving LMWH should be switched to unfractionated heparin 1 to 2 weeks before delivery so that rapid reversal of anticoagulation, if necessary, can be more easily attained. If a therapeutic effect cannot be achieved (as measured by the PTT with unfractionated heparin or the anti-FXa level with LMWH), an AT infusion can be given.[8] Generally, this is not necessary. Anticoagulation should be promptly reinstituted after delivery.

PROTEIN C AND PROTEIN S DEFICIENCY

Pathophysiology and Clinical Presentation

Deficiency or defect in protein C or protein S results in a loss of ability to inactivate excess factor VIIIa and factor Va, the two major cofactors that regulate amplification of the clotting cascade. Protein C levels are low in patients with DIC and liver disease, probably because the activation of hemostasis consumes this factor [*see 5:XIII Hemorrhagic Disorders*].

Homozygous protein C deficiency causes lethal thrombosis in infants. Heterozygous protein C deficiency probably occurs with a prevalence of 1 per 200 to 300 in the general population. Clinical expression of heterozygous protein C deficiency varies: many persons with heterozygous deficiency, as well as persons with low-normal protein C levels from other causes, do not experience thrombosis,[9] whereas other patients with heterozygous deficiency exhibit a definite tendency toward venous thrombosis even though their protein C levels are 40% to 50% of normal. This phenotypic variability suggests multiple gene interactions and supports the hypothesis that clinical thrombosis in such patients may result from a combination of protein C deficiency and one or more other prothrombotic mutations.[10] Cerebral venous thrombosis presumably accounts for cases of cerebral hemorrhagic infarction that occur in young adults with protein C deficiency.

Deficiency of protein S also leads to venous thrombosis, including mesenteric vein thrombosis. Pregnancy and the use of oral contraceptives lower the protein S level, which may account for some cases of thromboembolism that occur under such circumstances.[11] Acquired protein S deficiency also occurs in patients with the nephrotic syndrome, who lose protein S in urine.[12] Case reports have associated protein S deficiency with warfarin-induced skin necrosis.[13]

Diagnosis

Functional and antigenic assays for protein C and protein S are now available in most coagulation laboratories. Functional assays are preferable for diagnosis. It is important to measure free protein S because some patients who have low free protein S levels have normal or borderline total protein S levels. Coagulation assays for protein C and protein S can give falsely low values in patients with factor V Leiden.[14]

Treatment

Warfarin is the treatment of choice for preventing thrombosis, even though it lowers protein C levels still further. Because the half-life of protein C is only 6 to 7 hours, much shorter than that of prothrombin and factor X, a period of enhanced hypercoagulability follows initiation of warfarin therapy in patients with protein C deficiency. Heparin should be given along with warfarin during the initiation of anticoagulation; it can be withdrawn afterward. Warfarin-induced skin necrosis is a rare complication of anticoagulation therapy.

FACTOR V LEIDEN

Epidemiology and Etiology

Factor V Leiden is a mutated form of factor V (first identified by researchers in Leiden, The Netherlands) that, once activated, is relatively resistant to the anticoagulant effects of activated protein C (APC). The defect is transmitted as an autosomal dominant trait. Approximately 5% of the general white population is heterozygous for factor V Leiden; the defect is almost absent in other ethnic groups.[15] Factor V Leiden is now considered to be the most common hereditary hypercoagulable state. Its prevalence in patients with thrombophilia is as high as 20% to 50%.[16] In a large cohort study of unselected patients with a first episode of symptomatic DVT, factor V Leiden was found in 16% of patients.[17] In women who have thrombosis while taking oral contraceptives, the frequency of factor V Leiden is about 23%.[18] The relative risk of DVT in a factor V Leiden homozygote (estimated incidence, 0.5% to 1% a year) is approximately 80-fold higher than in a normal person.[19] The risk of thrombosis in persons who are heterozygous for factor V Leiden is estimated to be fourfold to eightfold higher than that in normal persons; the relative risk increases to more than 30-fold when factor V Leiden is combined with oral contraceptive use. The absolute risk of thrombosis, however, is low.[20] Association of factor V Leiden

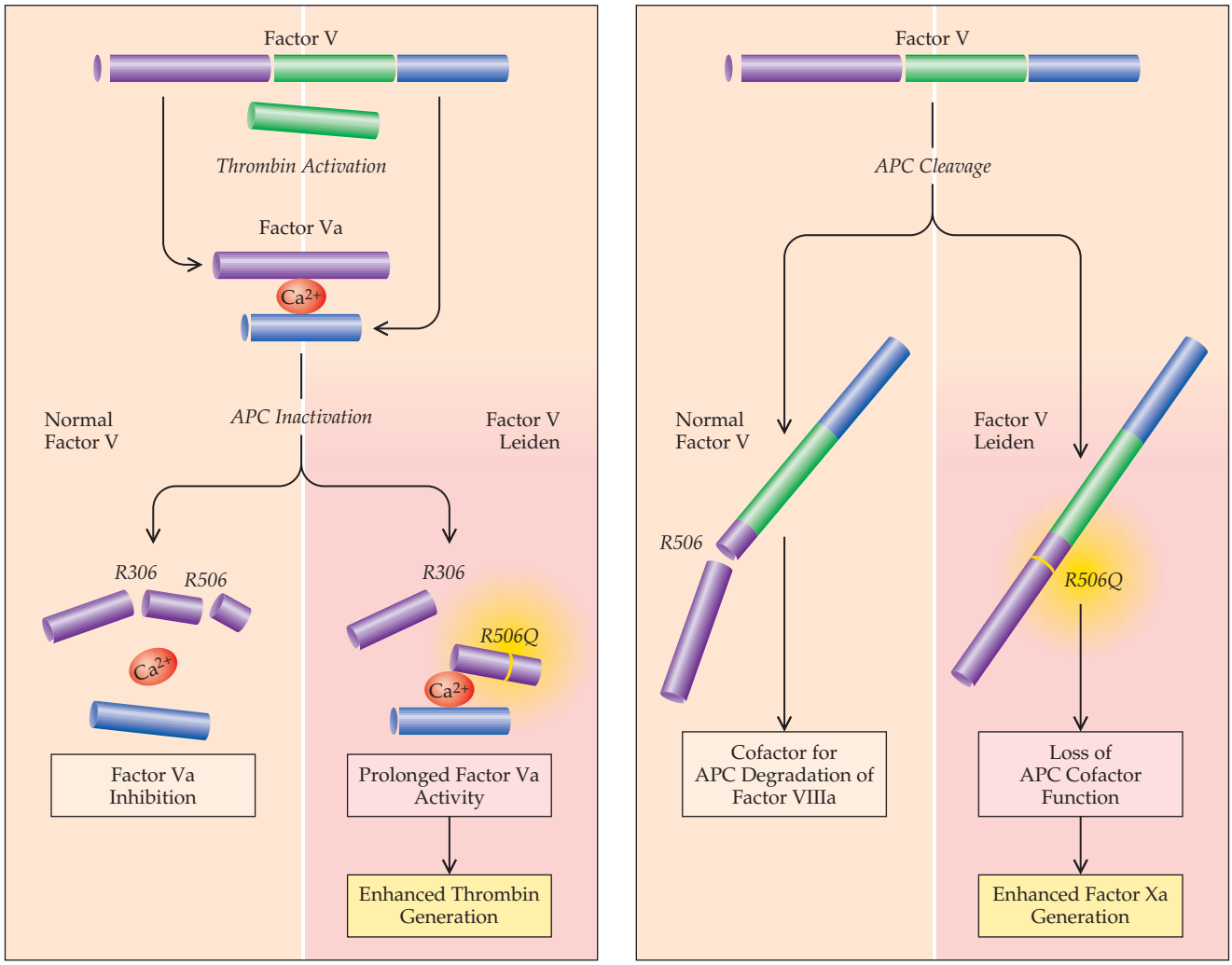


Figure 1 **Degradation of thrombin-activated factor V Leiden by activated protein C (APC) is significantly slower than that of normal activated factor V (factor Va), which leads to enhanced thrombin generation (left). Recent evidence suggests that normal factor V, together with protein S, serves as a cofactor of APC in the inhibition of factor VIIIa (right). This APC cofactor function of factor V requires the cleavage of factor V by APC at arginine 506; therefore, factor V Leiden has a poor cofactor function.**

with deficiencies of protein C, protein S, or AT has been reported in some families.[21-23] Overall, although factor V Leiden is highly prevalent, it is a relatively weak risk factor for thrombosis.

Approximately 5% of cases associated with inherited resistance to APC are attributable to other mutations and defects.[24-26] Conditions such as factor VIII elevation, pregnancy, oral contraceptive use, and lupus anticoagulant may result in APC resistance.[27] APC resistance that is not caused by factor V Leiden may be a risk factor for stroke[28,29] and venous thrombosis.[30,31] The overall risk of venous thrombosis from APC resistance is similar to or less than that posed by factor V Leiden.[32]

Pathophysiology

Resistance to the anticoagulant effects of APC is caused by a specific mutation in factor V (factor V Leiden or factor V R506Q) that results from a single-nucleotide substitution that leads to the replacement of arginine with glutamine at position 506.[32] Arginine 506 is located at one of the two major APC cleavage sites of activated factor V. Activated factor V Leiden expresses normal procoagulant activity, but its degradation by APC is approximately 10 times slower than that of normal activated factor V (factor Va). This slowing leads to increased thrombin generation.[33] In addition, recent evidence suggests that factor V (but not factor Va), together with protein S, serves as a cofactor of APC in the inhibition of the factor VIIIa/factor IXa complex and that factor V Leiden has a poor APC cofactor function [*see Figure 1*].

Clinical Presentation

Clinical manifestations of factor V Leiden are similar to those of deficiencies of AT, protein C, and protein S—mainly, venous thrombosis. However, the first thrombotic manifestation in factor V Leiden often occurs later than in the other hereditary thrombophilic states. Approximately 25% of apparently healthy men older than 60 years who experience a first episode of venous thrombosis have factor V Leiden.[34] There are conflicting data on whether factor V Leiden is associated with an increased risk of recurrent deep vein thrombosis. Several studies reported a slightly enhanced recurrence risk (twofold to fourfold), but more recent studies have shown that the risk of recurrence is similar to that in persons without the mutation.[17,35,36]

Diagnosis

Factor V Leiden can be identified rapidly and precisely with simple DNA-based tests. These tests allow the diagnosis to be made in patients receiving anticoagulation therapy with warfarin and in those who have coexisting antiphospholipid antibodies. Because factor V Leiden is not the sole cause of APC resistance, it may be worthwhile to pursue the diagnosis with an APC-resistance test in selected cases.

Treatment

Management of factor V Leiden is similar to that of AT, protein C, and protein S deficiencies. Patients with a first episode of venous thrombosis should receive anticoagulation therapy for 6 months. Thereafter, they should be given prophylactic anticoagulation therapy in situations known to provoke thrombosis. Long-term anticoagulation should be considered in patients with recurrent thrombosis.[37]

Young women known to be factor V Leiden carriers should avoid the use of oral contraceptives, which increases the relative risk of thrombosis (although the risk remains low in terms of absolute incidence). The optimal treatment of carriers during pregnancy has not been established. The rate of venous thromboembolism is low, about 2% without thrombosis prophylaxis.[20] My practice is not to use thrombosis prophylaxis routinely during pregnancy, but I will consider postpartum prophylaxis for 6 weeks, especially when the family history of thrombosis is strong. Routine screening of family members of patients with factor V Leiden is not cost-effective.[38]

PROTHROMBIN GENE MUTATION 20210A

A G-to-A mutation at nucleotide position 20210 in the 3' untranslated region of the prothrombin gene is associated with an increased incidence of venous thrombosis. The prevalence of the mutation in healthy persons is about 2.3%. Like factor V Leiden, this mutation is very rare in Asians and Africans. Unlike factor V Leiden, it is more common in southern Europeans than in northern Europeans.[39] The relative risk of thrombosis in persons with this mutation is 2.8, which is similar to the relative risk in those with factor V Leiden.[40] The mutation can be found in up to 18% of patients with thrombosis and family histories of thrombosis. The most common presentation is DVT of the lower extremities. Prospective studies have not shown an increased risk of recurrent DVT in patients with this mutation.[41] However, carriers who are heterozygous for both factor V Leiden and the prothrombin mutation have a higher risk of recurrent thrombosis.[35] The combination of oral-contraceptive use and the prothrombin gene mutation is associated with an increased incidence of cerebral vein thrombosis in young women.[42]

HYPERHOMOCYSTEINEMIA

Homocysteine is a highly reactive amino acid that is normally found in blood at levels of 5 to 15 mmol/L. Normally, homocysteine is derived from methionine by a transmethylation process and is remethylated to methionine or converted to cysteine [*see Figure 2*]. Metabolism of homocysteine requires betaine, cobalamin (vitamin B_{12}), folate, and pyridoxine (vitamin B_6).

Homocysteine can promote oxidation of low-density lipoprotein (LDL) cholesterol and presumably is toxic to vascular endothelium.[43,44] It may also inhibit thrombomodulin expression and protein C activation and suppress endothelial heparan sulfate expression; both of these effects lead to hypercoagulability.[45,46] Homocysteine also enhances the binding of lipoprotein(a) (an atherogenic lipoprotein) to fibrin, which may provide a link between hyperhomocysteinemia, thrombosis, and premature atherosclerosis [*see* Lipoprotein(a), *below*].[47] The vascular damage caused by high homocysteine levels leads to arterial and venous thrombosis and, perhaps, accelerated atherosclerosis.

Epidemiology and Etiology

Hyperhomocysteinemia can be divided into three classes: severe (homocysteine plasma concentration > 100 mmol/L), moderate (25 to 100 mmol/L), or mild (16 to 24 mmol/L). Severe hyperhomocysteinemia is usually caused by a homozygous deficiency of the enzyme cystathionine β-synthase. Affected persons have severe mental retardation, ectopic lens, skeletal abnormalities, and severe early-onset arterial and venous thrombotic disease.[48]

Mild or moderate hyperhomocysteinemia results from either hereditary or acquired defects in the homocysteine metabolic pathway. Heterozygous deficiency in cystathionine β-synthase is quite common in the general population, with a frequency of 0.3% to 1.4%.[48] A defect in the remethylation pathway is commonly caused by a thermolabile mutant of the methylenetetrahydrofolate reductase (MTHFR) enzyme whose activity is approximately 50% of normal; the homozygous state has a prevalence of 5% in the general population.[49] However, the homozygous form of the MTHFR thermolabile enzyme isoform is not clinically relevant in patients whose diet includes adequate folate.

Common causes of acquired hyperhomocysteinemia are deficiencies of dietary cobalamin, folate, or pyridoxine. A prospective study found that mild hyperhomocysteinemia is quite common in the elderly, despite normal serum vitamin concentrations.[50] Acquired hyperhomocysteinemia is also common in patients with end-stage renal disease.

Mild to moderate hyperhomocysteinemia is associated with cerebrovascular disease, coronary artery disease, and peripheral vascular disease in persons younger than 55 years and with carotid artery stenosis in the elderly.[51,52] It is found in 10% of patients with a first episode of DVT.[53] In a prospective study, a graded relationship was found between elevated plasma homocysteine levels and mortality in patients with coronary artery disease.[54]

Clinical Presentation

Severe hyperhomocysteinemia should be suspected in patients with the characteristic phenotype (see above). Mild to moderate hyperhomocysteinemia should be suspected in cases of arterial and venous thrombotic disease—including cerebrovascular disease, peripheral arterial disease, and DVT—especially in young persons.

Diagnosis

Plasma homocysteine exists in free and protein-bound forms and is generally measured and reported as total plasma homocysteine (normal range, 5 to 15 mmol/L). Diagnosis of hyperhomocysteinemia is usually made by measuring plasma homocysteine levels after an overnight fast. Because as many as 40% of patients with hyperhomocysteinemia may have a normal fasting level, a methionine-loading test should be considered when indicated.[55] However, methionine is not generally available in most pharmacies. Plasma folate and vitamin B_{12} levels should also be measured to exclude hyperhomocysteinemia caused by folate or B_{12} deficiencies.

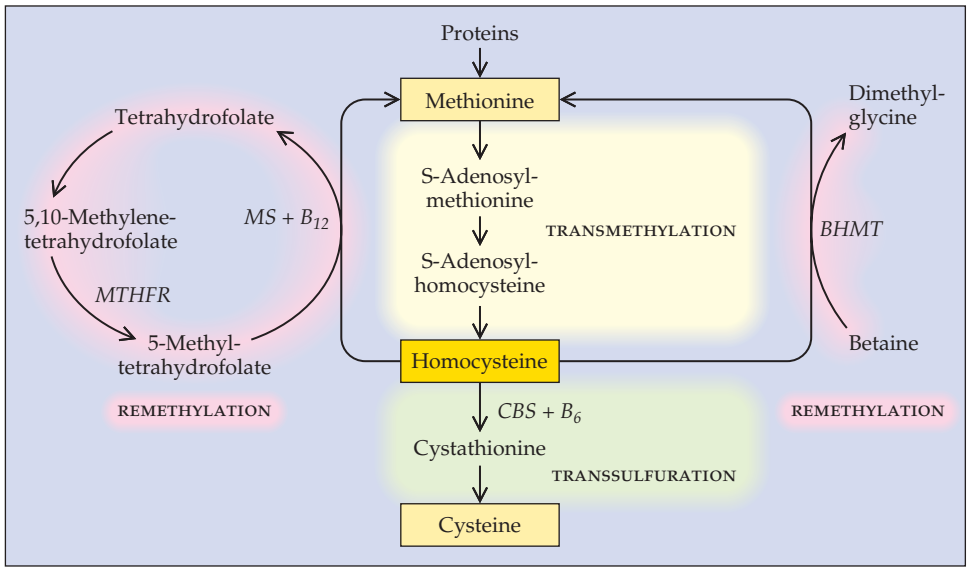


Figure 2 **Homocysteine's intracellular metabolism occurs through remethylation to methionine or transsulfuration to cysteine.[142] Elevation in plasma homocysteine levels can result from hereditary deficiency in cystathionine β-synthase (CBS); a thermolabile mutant of methylene-tetrahydrofolate reductase (MTHFR); or low dietary levels of cobalamin (vitamin B_{12}), folate, or pyridoxine (vitamin B_6), which are essential cofactors in the metabolic process. (BHMT—betaine-homocysteine methyltransferase)**

Treatment

Daily use of oral pyridoxine (250 mg) and folic acid (5 mg) brings elevated homocysteine levels down to normal in most cases.[56] Patients who have vitamin B_{12} deficiency should be given B_{12} supplements. Repeat measurement of plasma homocysteine levels (generally done 1 month after starting supplementation) may be prudent to ensure that the treatment with pyridoxine and folate is working. Betaine (3 g p.o., b.i.d.) is sometimes effective in patients with hyperhomocysteinemia that is resistant to pyridoxine and folate. It is currently unknown whether correction of hyperhomocysteinemia by these measures leads to clinical benefit.

LIPOPROTEIN(A)

Lipoprotein(a) [Lp(a)] is an independent risk factor for coronary artery thrombosis.[57] A prospective case-control study associated elevated plasma Lp(a) levels with an approximately threefold increase in risk of coronary artery disease in men.[58] The association between high Lp(a) and ischemic stroke in young adults is controversial.[59,60] Distributions of Lp(a) are skewed in the general population—especially among whites, in whom the median is 3.7 mg/dl but the mean is 6.9 mg/dl.[61] The 95th percentile for plasma Lp(a) is estimated to be in the 25 to 30 mg/dl range.

The Lp(a) class of lipoproteins is formed by the assembly of LDL particles and apoprotein(a), a protein that has some structural similarities to plasminogen (specifically, in the kringle domains) and competes with plasminogen for the endothelial cell binding site, thereby displacing plasminogen and downregulating plasmin generation at the endothelial cell surface.[62] High plasma concentrations of Lp(a) may therefore suppress the endothelial fibrinolytic response. Lp(a) is found in the intima of human atherosclerotic vessels, and transgenic mice expressing human Lp(a) develop extensive atherosclerosis.[63] Measurement of Lp(a) levels can be done in commercial laboratories and should be considered in young patients with arterial thrombosis. Elevated LDL cholesterol levels appear to elicit or exacerbate the risk factors associated with high Lp(a), and therefore, diet, exercise, and standard pharmacologic approaches should be used in patients with high LDL cholesterol levels.[64] In small studies, niacin at high doses (2 to 4 g p.o. daily) and tamoxifen (20 mg daily) have lowered elevated Lp(a) levels by 30% to 40%.[65,66] High doses of niacin are frequently associated with facial flushing and headaches. These unpleasant side effects can be ameliorated by starting niacin at a low dose (e.g., 300 mg daily) and then increasing the dose incrementally over time or through the use of extended-release niacin. Liver function should be checked periodically.

DYSFIBRINOGENEMIA

Approximately 300 abnormal fibrinogens (dysfibrinogens) have been reported, and about 85 structural defects have been identified in dysfibrinogenemia. These are most commonly characterized by functional defects of fibrinopeptide A release and fibrin polymerization and less commonly by defective plasminogen binding and activation. About half of the fibrinogen mutations are not associated with any clinical symptoms. Mild bleeding or recurrent thrombosis occurs in about equal numbers in the remaining mutations.[67] In rare cases, patients experience both bleeding and thrombosis. Acquired dysfibrinogenemia may complicate hepatocellular carcinoma or chronic liver disease. Evaluation in a general laboratory usually shows a discrepancy between antigenic and functional levels of fibrinogen, because most patients with dysfibrinogens have suboptimal clotting function, with prolonged thrombin time (TT) and reptilase time (RT). The abnormal fibrinogens form fibrin clots that are resistant to clot lysis. Precise identification of the structural defect requires substantial effort in a research laboratory. Management of recurrent thrombosis caused by dysfibrinogenemia is the same as that in other patients with thrombophilia.

DYSPLASMINOGEN AND ABNORMAL FIBRINOLYSIS

In rare cases, abnormal plasminogens (dysplasminogens), which are defective in their activation to plasmin, are associated

with thrombosis. Patients with such a disorder have a low plasma plasminogen level on functional assays.[68] Increased levels of plasminogen activator inhibitor–1 (PAI-1) and decreased plasma fibrinolytic activity have been reported in patients with preeclampsia.[69] Acquired impairment of fibrinolytic activity may be associated with postoperative thrombosis.[70] However, more studies are required to establish the role of abnormal fibrinolysis in recurrent clinical thrombosis.[71] Antigenic assays for tissue plasminogen activator (t-PA) and PAI-1 are available in some commercial laboratories, but specific functional assays are available only in research laboratories.

ELEVATED FIBRINOGEN, FACTOR VII, AND FACTOR VIII LEVELS

A high plasma fibrinogen level is an independent risk factor for coronary artery disease.[72] An elevated factor VII level has also been associated with the development of heart disease.[73] A factor VIII level above the 90th percentile of normal is associated with an approximately fivefold increased risk of a first episode of DVT[74,75]; it also increases the risk of recurrence.[76] Additional studies are required to establish the clinical utility of measuring these parameters in patients with thrombosis.

Acquired Hypercoagulable States

ANTIPHOSPHOLIPID ANTIBODY SYNDROME

The antiphospholipid antibody syndrome is caused by autoantibodies to proteins associated with negatively charged phospholipids. The terms antiphospholipid and anticardiolipin are used synonymously. Antiphospholipid antibodies also include lupus anticoagulant, which is an inhibitor that was first identified in patients with systemic lupus erythematosus. Many patients who have this inhibitor do not have lupus, and it is sometimes called lupuslike anticoagulant.

Epidemiology

Antiphospholipid antibody syndrome occurs secondary to systemic lupus erythematosus and, less commonly, to rheumatoid arthritis, temporal arteritis, and other connective tissue disorders. It is also associated with HIV-1 and hepatitis C infections, lymphoproliferative diseases, and certain drugs (e.g., phenothiazine and procainamide). When no risk factor can be identified, the syndrome is regarded as primary. In a large cohort study of 1,000 patients with antiphospholipid antibody syndrome, 53% of patients were classified as having primary antiphospholipid antibody syndrome, and 47% had secondary antiphospholipid antibody syndrome.[77]

Pathophysiology

The two most common protein targets for the antiphospholipid antibodies appear to be β_2-glycoprotein I (β_2-GPI) and prothrombin. β_2-GPI is a plasma protein that binds anionic phospholipids with high affinity. It has weak anticoagulant function in vitro. β_2-GPI can induce cardiolipin from its usual bilaminar form to a hexagonal form that is highly immunogenic.[78] The anticardiolipin antibody enzyme-linked immunosorbent assay (ELISA) usually detects antibodies directed against the cardiolipin/β_2-GPI complex. Lupus anticoagulant antibodies have been purified that specifically react with prothrombin but not thrombin. These purified antibodies can enhance the binding of prothrombin to the cultured endothelial cell surface.[79] Other protein-phospholipid targets may also be involved. Antiphosphatidylethanolamine antibodies are found in many patients with antiphospholipid antibody syndrome, and some of these antibodies inhibit activated protein C function.[80] Antibodies to heparin and heparan sulfate, which inhibit the heparin-dependent neutralization of thrombin by AT, have been found.[81] On the basis of this heterogeneity of antiphospholipid antibodies, it seems likely that multiple mechanisms are involved in the pathogenesis of thrombosis in this syndrome.

Clinical Presentation

Thrombotic events occur in approximately 30% of patients with antiphospholipid antibodies (overall incidence, 2.5 events per 100 patient-years).[82] In the cohort study cited above, 37% of patients presented with venous thrombosis, 27% with arterial thrombosis, 15% with both venous and arterial thrombosis, and 12% with fetal loss only[77] [*see 15:IV Systemic Lupus Erythematosus*].

Diagnosis

The diagnosis of antiphospholipid antibody syndrome should be considered in any patient who presents with an idiopathic arterial or venous thrombosis or in a woman with a history of recurrent miscarriages. The diagnosis is confirmed by the presence of anticardiolipin antibodies on ELISA (see above) or lupus anticoagulant on clotting assays.

The general criteria for the diagnosis of lupus anticoagulant are (1) prolongation of at least one phospholipid-dependent clotting assay; (2) proof, by mixing studies, that the prolongation is caused by an inhibitor and not a clotting factor deficiency; and (3) confirmation that the inhibitor is phospholipid dependent [*see Table 5*]. The clotting tests commonly used are activated PTT (aPTT) and dilute Russell viper venom time (RVVT). The reagents in aPTT are variably sensitive to the lupus anticoagulant and are influenced by concentrations of some plasma clotting factors (e.g., factor VIII). Therefore, an aPTT reagent that is

Table 5 Proposed Clinical and Laboratory Criteria for the Antiphospholipid Antibody Syndrome[93]

Clinical features	Pregnancy morbidity (any of the following): More than one unexplained fetal death at greater than 10 wk Delivery at less than 34 wk, with severe pregnancy-induced hypertension Three or more pregnancy losses at less than 10 wk Thrombosis: Venous (superficial thrombophlebitis; deep vein thrombosis; pulmonary embolism; cerebral and retinal vein thrombosis; renal, splanchnic, and mesenteric vein thrombosis) Arterial (ischemic cerebral infarction, transient cerebral ischemia, amaurosis fugax, migraine, carotid and vertebrobasilar artery thrombosis, aortic arch syndrome, peripheral arterial thrombosis and embolism, renal and mesenteric artery thrombosis, livedo reticularis)
Laboratory features	Lupuslike anticoagulant: Activated partial thromboplastin time, dilute Russell viper venom time, kaolin clotting time, tissue thromboplastin inhibition test Anticardiolipin antibodies (either of the following): IgG anticardiolipin antibodies (> 20 GPL) IgM anticardiolipin antibodies (> 20 MPL) Thrombocytopenia (platelet count < 100,000/μl)

GPL—IgG phospholipid units MPL—IgM phospholipid units

Table 6 Classification of Antiphospholipid Antibodies[147]

Autoimmune causes
- Primary (do not fulfill criteria for systemic lupus erythematosus)
- Secondary (fulfill criteria for systemic lupus erythematosus or other connective tissue diseases)
- Drug-induced (e.g., phenothiazines, quinidine, quinine, synthetic penicillins, hydralazine)

Alloimmune causes
- Infections (viral, bacterial, fungal)
- Malignancies (e.g., hairy-cell leukemia, lymphoproliferative disease)

sensitive to the lupus anticoagulant should be used in the screening test. Dilute RVVT is much more sensitive than aPTT but is a manual test and not as well standardized. Other tests, such as kaolin clotting time and the tissue thromboplastin inhibition test, are useful when available. The presence of an inhibitor necessitates a mixing study to demonstrate lack of correction with normal plasma. Correction of the prolongation by addition of phospholipid in the form of platelet lysates or as hexagonal-phase phospholipid will confirm the diagnosis. Clotting factor assays can be carried out in equivocal cases. A lupus anticoagulant will cause functional deficiency of several phospholipid-dependent clotting factors, not just one particular factor.

Anticardiolipin antibodies are reported as IgG (in IgG phospholipid [GPL] units) and IgM (in IgM phospholipid [MPL] units). The prevalence of elevated anticardiolipin IgG and IgM antibodies in normal populations is approximately 5%; with repeated testing, the prevalence is less than 2%.[83] High titers of anticardiolipin IgG antibodies (> 33 GPL) are associated with an approximately fivefold increase in overall thrombotic risk.[82,84] The importance of low titers of IgG antibodies (< 20 GPL), isolated IgM antibodies, and IgA antibodies has not been established.[85,86] Both functional and antigenic assays should be ordered in the evaluation of a patient, because these two assays do not completely overlap. In one study of antiphospholipid antibody syndrome, 88% of patients had anticardiolipin antibodies (IgG, IgM, or both) and 54% of patients had lupus anticoagulant. Lupus anticoagulant was typically found in association with anticardiolipin antibodies, but it occurred in isolation in about 12% of patients.[77] Certain infections and drug exposures may lead to a transient appearance of antiphospholipid antibodies, which disappear after the resolution of infection or discontinuance of the drug [*see Table 6*]. Therefore, laboratory tests should be repeated at least once (6 weeks after the first tests) to confirm the diagnosis. Conversely, approximately 20% of patients with low titers of anticardiolipin IgG antibodies will have higher titers upon repeat testing. Retesting is also warranted in patients with new or recurrent thrombosis.[85]

Treatment

The current therapeutic recommendations for antiphospholipid antibody syndrome are mostly based on observational studies that support an association between antiphospholipid antibodies and thrombosis, particularly recurrent thrombosis.[82-86] In the acute treatment of DVT in patients with antiphospholipid antibody syndrome, monitoring the effect of unfractionated heparin can be problematic because lupus anticoagulant prolongs the aPTT. The use of LMWH circumvents this problem because LMWH does not require dose titration and monitoring. The patient should be treated with LMWH and warfarin in the usual fashion, with an overlap of at least 5 days before discontinuing LMWH.

Retrospective analysis shows that patients with the antiphospholipid antibody syndrome and a history of thrombosis have a high rate of recurrent thrombosis (in the range of 50% to 70%) if they are not given prolonged warfarin therapy.[87,88] The site of the first thrombotic event (i.e., arterial or venous) tends to predict the site of the recurrent event.[88] High-intensity warfarin therapy, to an international normalized ratio (INR) of 3.0 to 3.5, had been advocated for these patients. This recommendation was based primarily on a retrospective analysis,[87] however, and two prospective clinical trials have now demonstrated that most of these patients can be adequately treated with conventional levels of anticoagulation (i.e., an INR of 2.0 to 3.0).[89] On the other hand, there are clearly some patients with antiphospholipid antibody syndrome who experience recurrent thrombosis with conventional anticoagulation and hence require a higher level of treatment.

The optimal duration of oral anticoagulation therapy for antiphospholipid antibody syndrome has not been fully established. Recurrent venous thrombosis or ischemic stroke usually justifies long-term warfarin. In a patient with a first episode of DVT who is found to have antiphospholipid antibodies, warfarin therapy is indicated for at least 6 months and perhaps for life.[90] The severity of the specific thrombotic episode, the coexistence of any reversible thrombotic risk factors, and the risks of long-term oral anticoagulation therapy should also be considered. It should be noted that the lupus anticoagulant may occasionally increase the prothrombin time (PT) and, in turn, the INR, thus posing a problem for the monitoring of warfarin therapy.[91] When PT and INR increase, use of a lupus anticoagulant–insensitive thromboplastin reagent is helpful.

In asymptomatic patients with anticardiolipin antibodies or lupus anticoagulant but no history of thrombosis, anticoagulation is not required.

Pregnancy Loss in Antiphospholipid Antibody Syndrome

Several prospective studies confirm an association between recurrent miscarriages and antiphospholipid antibodies. The antibodies presumably cause pregnancy loss by promoting placental thrombosis.[92] Antiphospholipid antibodies should be measured in patients with a history of unexplained second- or third-trimester loss, fetal demise, early-onset severe preeclampsia, and intrauterine growth retardation.[93] In contrast, antiphospholipid antibodies are not associated with sporadic early pregnancy loss,[94] which is frequently the result of genetic abnormalities in the fetus. The relationship of antiphospholipid antibodies with infertility is uncertain at present.

The management of pregnant women with antiphospholipid antibody syndrome is difficult because of the syndrome's association with thrombosis and the increased risk of bleeding with antithrombotic therapy. In a prospective, randomized, placebo-controlled trial, a combination of prednisone and aspirin was demonstrated to be ineffective in promoting live birth; in fact, it increased the risk of prematurity.[95] On the other hand, two prospective trials have demonstrated that heparin and low-dose aspirin (81 mg a day) provide a significantly better pregnancy outcome than low-dose aspirin alone, with viable infants being delivered in 70% to 80% of cases.[96,97] Furthermore, low-dose heparin (given initially as 5,000 units subcutaneously twice daily

and adjusted to maintain the aPTT within the upper limits of the normal range) seems to be as effective as higher-dose heparin combined with low-dose aspirin.[98] Treatment should begin as soon as pregnancy is confirmed. LMWH is preferable to unfractionated heparin for long-term use because LMWH can be given once or twice daily and may reduce the risk of osteopenia and heparin-induced thrombocytopenia (see below). Enoxaparin (40 mg once daily) and aspirin (100 mg daily) has been given from week 12 of gestation until 6 weeks postpartum with good results.[99]

HEPARIN-INDUCED THROMBOCYTOPENIA AND THROMBOSIS

Epidemiology

Heparin-induced thrombocytopenia (HIT) is a relatively common antibody-mediated drug reaction, occurring in about 1% of patients receiving porcine heparin and 5% of patients receiving bovine heparin.[100] The incidence of HIT is much lower in patients treated with LMWH. In a subset of patients, HIT progresses to a potentially fatal disorder characterized by venous and arterial thrombosis. Interestingly, both the frequency of HIT antibody formation and the clinical manifestations of HIT vary considerably in different patient populations. The incidence of HIT antibody formation is much higher after cardiac surgery than after orthopedic surgery (50% versus 15%); however, the incidence of clinically significant postoperative HIT appears to be lower in cardiac surgical patients than in orthopedic patients, in whom the incidence is 5%.[101,102]

Pathophysiology

The pathogenesis of HIT is attributable to the presence of an IgG antibody that recognizes a complex of heparin and platelet factor 4 (PF4) [*see Figure 3*].[103] PF4 is a cationic protein found in platelet α-granules; when released from the granules, PF4 binds to the negatively charged heparin molecule with high affinity. The IgG antibody binds to the PF4-heparin complex on platelet membranes, forming a ternary complex that in turn binds to the platelet membrane FcγRII receptor. This binding activates the platelets, leading to further release of PF4 and formation of PF4-heparin complex. The immune complex–coated platelets are cleared rapidly by the reticuloendothelial system, giving rise to thrombocytopenia. The thrombotic complications in HIT are caused by activation of platelets by the immune complex, which leads to the formation of platelet microparticles and enhanced thrombin generation.[104] PF4 also binds to heparinlike sulfated glycosaminoglycans (e.g., heparan sulfate) on the endothelial cell surface. In vitro evidence indicates that the antibody in HIT is able to bind to endothelial cells. The cells may then become activated, giving rise to thrombosis.[105] Given that only a subset of patients who form HIT antibodies experience clinical HIT,[102] the induction of HIT antibodies and the development of thrombocytopenia and subsequent thrombosis should be regarded as a continuum. Concomitant thrombotic risk factors probably play a major role in determining the clinical progression and manifestations of HIT.

Clinical Presentation

HIT typically develops 5 to 10 days after the initiation of heparin therapy. However, in patients who received heparin within the previous 100 days and are being retreated, the onset can be rapid—within hours after starting heparin.[106] Conversely, onset of HIT may not occur until as long as 19 days after heparin therapy is stopped.[107] This delayed-onset HIT appears to be associated with a higher titer of IgG antibodies against the PF4-heparin complex.

HIT is generally defined as a platelet count below 150×10^9/L or a drop in the platelet count by more than 50% from the postoperative peak at 5 to 14 days after heparin is started. The mean platelet count in HIT is about 60×10^9/L. Severe thrombocytopenia, with platelet counts below 20×10^9/L, occurs in fewer than 10% of patients; in 10% to 15% of patients, despite the 50% drop from peak levels, the platelet count nadir is above 150×10^9/L.[102]

The risk of HIT-associated thrombosis was once thought to be quite small; however, it is now recognized that thrombosis occurs in about one third to one half of patients with HIT, with venous thrombosis occurring more frequently than arterial thrombosis.[104] Thrombosis can occur at any platelet count, even at a very low one.

DVT, with or without pulmonary embolism, is the most common event leading to the diagnosis of HIT. The disorder may be further complicated by limb gangrene, especially in the setting of warfarin treatment without concomitant alternative anticoagulant coverage (see below). Cerebral vein thrombosis and adrenal hemorrhagic necrosis are uncommon but well-documented complications of HIT, and early diagnosis and urgent therapy can be lifesaving. Arterial thrombosis may present as limb ischemia, stroke, myocardial infarction, or, less commonly, mesenteric thrombosis and renal arterial thrombosis. Some patients have laboratory findings that support a diagnosis of DIC. Heparin-induced skin lesions may occur at heparin injection sites and range from painful erythematous papules to extensive dermal necrosis.[108]

Unlike antibodies induced by quinine, quinidine, or sulfonamides, which can persist for years, heparin-induced antibodies appear to be quite transient. They fall to undetectable levels at a median of 50 to 85 days, depending on the assay performed.[106]

Diagnosis

The diagnosis of HIT is supported by the finding of heparin-induced platelet aggregation in the presence of the patient's serum. However, the sensitivity of this test can be as low as 50%.[109] Sometimes, the patient's serum can cause spontaneous aggregation of donor platelets in the absence of heparin, most likely caused by the presence of immune complexes unrelated to HIT, which makes proper interpretation of the test result impossible. Heparin-induced platelet serotonin release using washed platelets has high sensitivity and specificity for HIT but is available only in a few specialized laboratories. ELISAs to detect antibodies that are reactive to the PF4-heparin complex are commercially available and have become the most commonly used test for HIT. These ELISAs have a higher sensitivity than the platelet aggregation assay and can be more easily performed in a general clinical diagnostic laboratory. However, false positive results (i.e., positive tests in the absence of HIT or thrombocytopenia) occur in 10% to 15% of medical patients and in more than 20% of patients receiving heparin for peripheral vascular surgery. A seroconversion rate as high as 50% has been reported in patients undergoing cardiopulmonary bypass surgery, limiting its usefulness in that situation.[110] Because HIT can be complicated by serious thrombotic problems, however, diagnosis of HIT should be based primarily on appropriate clinical findings, and management should be started while laboratory confirmation is awaited.

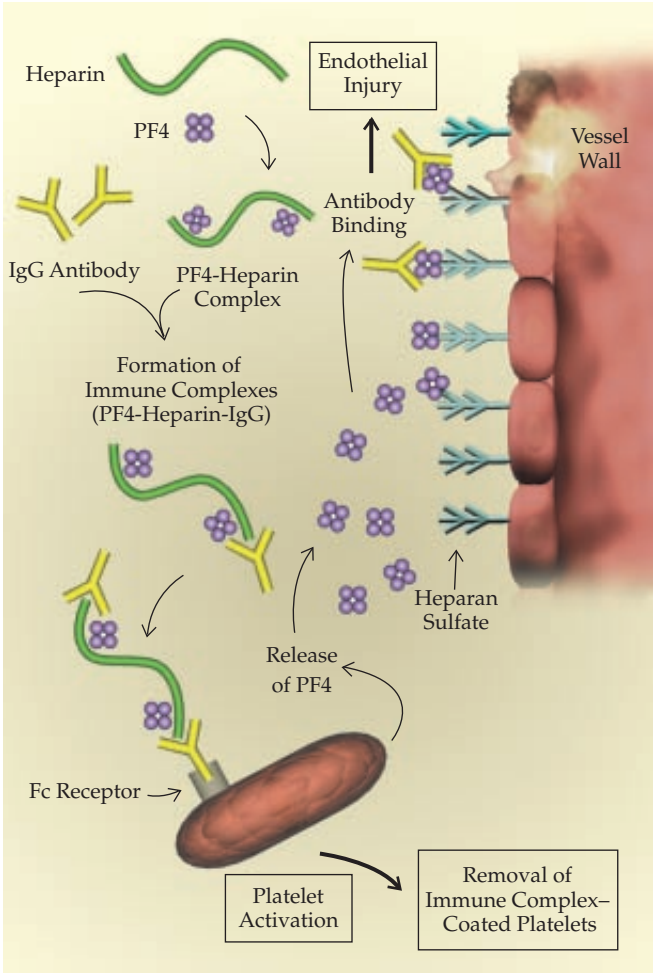


Figure 3 **In a proposed explanation for heparin-induced thrombocytopenia, IgG antibodies recognize platelet factor 4 (PF4)–heparin complexes. The resulting PF4-heparin-IgG immune complexes bind to Fc receptors on circulating platelets. Fc-mediated platelet activation releases PF4 from α-granules in platelets, establishing a cycle of platelet activation and formation of prothrombotic platelet microparticles. Removal of immune complex–coated platelets by the reticuloendothelial system results in thrombocytopenia. PF4 also binds to heparan sulfate on the surface of endothelial cells, leading to immune-mediated injury, thrombosis, and disseminated intravascular coagulation.[145]**

Treatment

Management of HIT consists of stopping heparin immediately and starting alternative anticoagulation therapy. It is important to discontinue all types of heparin: there are anecdotal reports of HIT caused by trace amounts of heparin used in heparin flushes of intravascular lines. Even if the patient has mild thrombocytopenia alone without any evidence of thrombosis, it is advisable to discontinue heparin and treat the underlying hypercoagulable state with an alternative anticoagulant. This aggressive approach is supported by a retrospective cohort study in which thrombosis developed within 30 days of heparin cessation in approximately half of patients who initially had no clinical symptoms from their HIT.[100]

HIT may develop in patients who were receiving heparin for preexisting thrombosis and are in the process of being switched over to warfarin. If the patient has been on warfarin for 4 or 5 days and the INR has reached an adequate therapeutic range, the clinician may rely on warfarin alone and monitor the patient carefully. However, if the patient has been on warfarin for less than 4 days or has evidence of a thrombotic complication from the HIT, an alternative anticoagulant should be used in addition to warfarin.

Alternative anticoagulant agents include lepirudin, argatroban, and fondaparinux. Although LMWH is much less immunogenic than unfractionated heparin in causing HIT,[104] it cannot be used as a safe substitute when a patient develops HIT caused by unfractionated heparin. LMWH and unfractionated heparin have extensive cross-reactivity (> 90%) in terms of antibody recognition. LMWH is not an appropriate choice in patients with HIT.

Hirudin, a 65-amino-acid protein originally extracted from the salivary gland of the medicinal leech (*Hirudo medicinalis*), is a potent direct thrombin inhibitor. Hirudin binds directly to thrombin's active site, independently of AT. It is not neutralized by PF4. Hirudin's anticoagulant function is monitored by aPTT.

Lepirudin is a recombinant form of hirudin that is approved for treatment of HIT. In two prospective clinical trials, use of lepirudin reduced serious thrombotic complications to about 20% (compared with a rate of about 40% in historical control subjects).[111,112] Lepirudin was given by intravenous bolus (0.4 mg/kg), followed by continuous infusion at 0.15 mg/kg/hr for 2 to 10 days as indicated. The dose was adjusted to maintain a target aPTT of 1.5 to 3 times normal. Patients experienced an increase of minor bleeding (from puncture sites, epistaxis, and hematuria) but no intracranial bleeding. Of note, 40% of patients developed antihirudin antibodies. These are not neutralizing antibodies and may actually enhance the drug's potency, perhaps by delaying its clearance—another reason to monitor aPTT levels. Judging from cardiology intervention trials, bleeding risk would be substantially increased with concomitant use of thrombolytics; therefore, it is not advisable to use the agents in combination. There is no effective antidote for lepirudin. It has a half-life of approximately 1.3 hours and is cleared by the kidneys. Thus, in patients with renal insufficiency, the dose of lepirudin needs to be adjusted carefully on the basis of creatinine clearance values.

Argatroban, a synthetic direct thrombin inhibitor, is approved for prophylaxis or treatment of thrombosis in patients with HIT. It is given by continuous intravenous infusion of 2 mg/kg/min to maintain an aPTT of 1.5 to 3 times baseline (not to exceed 100 seconds or 10 mg/kg/min). In one study, argatroban reduced the serious thrombotic complications of HIT by about 50%, as compared with historical controls, with a major bleeding rate of about 7%.[113] Its half-life is only 40 to 50 minutes. In contrast to lepirudin, argatroban is cleared by the liver and therefore can be used more easily in patients with renal insufficiency. Like lepirudin, argatroban does not have a specific antidote, and bleeding complications need to be watched for carefully.

A third alternative anticoagulant is fondaparinux, a synthetic pentasaccharide that activates AT, leading to thrombin inhibition. Fondaparinux has been approved for prophylaxis against DVT in orthopedic surgery. Because of its small size and reduced negative charge, fondaparinux does not form a complex with PF4 and therefore does not react with the antibody directed against the heparin-PF4 complex in HIT. Case reports describe the successful use of fondaparinux in HIT at a fixed dose of 7.5 mg administered subcutaneously once daily.

Patients who have HIT without associated thrombotic com-

plications should be treated with one of the alternative anticoagulants until the platelet count has returned to normal, which generally takes 5 to 7 days. It is my practice to discontinue the alternative anticoagulant at that time. However, some experienced clinicians choose to continue empirical anticoagulation for up to 1 month, on the rationale that HIT represents an intensive hypercoagulable state. In patients who require prolonged anticoagulation—whether because they have other indications for anticoagulation or because they have had thrombotic complications from HIT—warfarin is used for long-term treatment. Warfarin should be started only after the patient has received adequate anticoagulation therapy with one of the alternative anticoagulants. The two agents should be used concurrently for at least 5 days before the alternative anticoagulant is discontinued. Both lepirudin and argatroban may increase the PT and INR, thus interfering with warfarin dose adjustments. The PT should be rechecked 6 hours after the discontinuance of lepirudin or argatroban to ensure that an INR of 2 to 3 has been achieved. For patients who have had HIT-associated thrombosis, warfarin therapy for 3 months should be adequate.

Some patients with serologically confirmed HIT may, at some point in the future, require surgery that involves cardiopulmonary bypass. The use of recombinant hirudin or argatroban in that situation has been described anecdotally.[112,114,115] Given the transient nature of the heparin-induced antibodies, subsequent reuse of heparin is theoretically reasonable, and indeed, a limited number of patients have been given heparin again after the disappearance of heparin-induced antibodies without any significant clinical sequelae.[106,116] Nevertheless, the use of heparin in this situation should be restricted to patients with a compelling indication for it, such as cardiac or vascular surgery, and only after the absence of detectable heparin-dependent antibodies has been confirmed by a sensitive assay, such as a PF4-heparin ELISA. Also, because reexposure to heparin may elicit a recurrence of heparin-dependent antibodies, heparin should be used only during the procedure itself, and an alternative anticoagulant should be started postoperatively for prophylaxis against recurrence of HIT. The anti–heparin-PF4 antibody should be checked postoperatively.

TROUSSEAU SYNDROME

Epidemiology

Some patients with cancer—especially those with occult solid tumors of the pancreas, ovary, liver, brain, colon, lung, or breast—may experience spontaneous venous thrombosis of the upper and lower extremities, or Trousseau syndrome. These patients also have an increased propensity toward recurrent arterial thrombosis and thromboembolism.[117] In a prospective study of patients who presented with idiopathic symptomatic DVT, cancer was diagnosed in approximately 8% of the patients during a 2-year follow-up (odds ratio, 2.3). In patients with recurrent thrombosis, the incidence of cancer was even higher (17%; odds ratio, 4.3).[118]

Pathophysiology

The underlying cause of Trousseau syndrome is a chronic, compensated form of DIC [*see 5:XIII Hemorrhagic Disorders*]. The activated procoagulants generated in DIC enhance thrombosis. Immunochemical staining of sections from tumors commonly associated with Trousseau syndrome often reveals tissue factor on the tumor surface.[119,120] A cancer procoagulant has been purified from some adenocarcinomas and leukemic cells; this procoagulant, identified as a cysteine protease, can activate the clotting cascade by directly activating factor X.[121,122] Interaction of the tumor cells with monocytes, platelets, and endothelial cells may also generate inflammatory cytokines and induce endothelial and monocytic procoagulant activities, further exacerbating the thrombosis. Tumor cells may secrete soluble mucins—complex polysaccharides that can activate leukocytes, leading to platelet-leukocyte microthrombi and thrombin generation.[123] Heparin blocks the tumor mucin activation of leukocytes, which may partially explain heparin's efficacy over warfarin in the treatment of this condition.

Clinical Presentation

Venous thrombosis in Trousseau syndrome usually manifests as migratory superficial thrombophlebitis or DVT of the lower extremities. The recurrent arterial thrombosis in these patients arises from a nonbacterial thrombotic endocarditis in which sterile fibrin is deposited in the mitral valve. The fibrin clot may embolize to cause digital ischemia, transient ischemic attacks, and stroke.

Diagnosis

Coagulation studies in Trousseau syndrome show evidence of chronic, low-grade DIC (i.e., slightly low or even high fibrinogen and platelet levels and high levels of D-dimer). The PT and PTT are generally not prolonged. Overt DIC in such patients is uncommon.

How aggressively should one pursue the diagnosis of an underlying cancer in patients with idiopathic DVT? Research has not yet demonstrated the benefit and cost-effectiveness of an extensive screening approach.[124] In a large cohort study, cancer diagnoses after a primary thrombotic event were highest during the first 6 months of follow-up and declined rapidly to normal levels after the first year. Moreover, 40% of the patients who were diagnosed with cancer in the first year had distant metastases at the time of the diagnosis. It is unclear whether an earlier diagnosis after the thrombotic event would have changed the outcome in these patients. The researchers concluded that an aggressive search for a hidden underlying cancer in such patients is not warranted.[125]

At present, it is prudent to perform a careful history, physical examination, chest x-ray, routine blood counts, and chemistries. Some experts have also recommended multiple tests for fecal occult blood, prostate-specific antigen tests in men, and mammography and pelvic ultrasonography in women.[126,127] Careful follow-up examination and tests should be done as indicated by the initial evaluation.[128]

Treatment

The key to management of Trousseau syndrome is diagnosis and treatment of the underlying tumor. Unfortunately, tumors often present explosively in patients with Trousseau syndrome and may not respond to the usual therapies. In a prospective study, cancer patients had an approximately threefold increase in the rate of recurrent thrombosis and twofold increase in the rate of major bleeding during warfarin treatment of DVT. These complications occurred mostly during the first few months of anticoagulation and did not reflect underanticoagulation or over-anticoagulation but correlated with the extent and severity of the underlying cancer.[129] The likely explanation for the increased thrombosis recurrence in these patients is relative warfarin resistance, whereas the

increased bleeding may be related to bleeding at the primary tumor site. A prospective trial found that the risk of recurrent venous thromboembolism was 50% lower in cancer patients who received long-term treatment with the LMWH dalteparin than in those who received oral anticoagulation; there was no significant difference in the risk of major bleeding.[130] If a patient is receiving chemotherapy for the underlying cancer, an exacerbation of the DIC associated with tumor lysis should be anticipated. An increase in the dose of heparin may be required.

THROMBOTIC REACTIONS TO ESTROGENS

Oral contraceptives increase the risk of thromboembolic disease approximately fourfold.[131] Epidemiologic studies indicate that contraceptives containing a third-generation progestin (e.g., desogestrel) carry a twofold greater risk of thrombosis than those with a second-generation drug (levonorgestrel).[132] For that reason, the preferred choice for first-time users of oral contraceptives is a compound containing a second-generation drug (e.g., Alesse, Levlite, Levora, Nordette, Triphasil, Trivora). The risk of venous thromboembolism disappears when the drugs are discontinued.

In postmenopausal women, estrogen replacement increases the risk of venous thromboembolism about threefold. The absolute risk is low, however—it is estimated to be approximately 3.2 per 10,000 patient-years. Therefore, estrogen replacement is not contraindicated in patients who require hormonal treatment to control severe postmenopausal symptoms. However, patients with a previous history of DVT or pulmonary embolism are at increased risk for recurrence and therefore should avoid hormone replacement therapy if possible.[133]

The association between estrogen use and thromboembolic disease remains unexplained.[134] Estrogen treatment is known to produce changes in the plasma levels of many proteins involved in coagulation and fibrinolysis, including decreases in protein S and AT and increases in plasminogen. The changes are generally quite modest, however, and are not thought to account for the increased risk of thrombosis. On the other hand, in women who are heterozygous for factor V Leiden, oral contraceptive use increases the thrombotic risk synergistically to about 50 times normal.[18] There is also a moderate synergistic increase in thrombotic risk (15-fold) with the combination of factor V Leiden and hormone replacement therapy.[135]

Table 7 General Guidelines for Management of Patients with Venous Thromboembolism

Recurrence Risk	*Management*
High Recurrent idiopathic thrombosis One life-threatening thrombosis One spontaneous thrombosis at an unusual site (e.g., mesenteric or cerebral thrombosis) One spontaneous thrombosis associated with antiphospholipid antibody syndrome One thrombosis with two permanent risk factors One thrombosis with Trousseau syndrome	Lifelong oral anticoagulation therapy: INR 2.0 to 3.0
Medium One thrombosis with one permanent risk factor (except Trousseau syndrome) Idiopathic thrombosis with no identifiable risk factor	6 mo of oral anticoagulation therapy after first episode of thrombosis; vigorous prophylaxis in high-risk situations
Low One thrombosis with reversible risk factor	6 wk–3 mo of oral anticoagulation therapy after first episode of thrombosis; vigorous prophylaxis in high-risk situations

INR—international normalized ratio

Management of Venous Thromboembolism

The acute management of an initial episode of DVT or pulmonary embolism in patients with proven or presumed underlying risk factors for thrombosis is the same as that in other patients: heparin and then warfarin [*see 1:XVIII Venous Thromboembolism*].

The optimal intensity and duration of warfarin treatment in DVT have been the subject of many large clinical trials over the past decade. As regards the intensity of oral anticoagulation, an INR of 2 to 3 has proved optimal, with a low rate of thrombosis recurrence and a rate of major bleeding of about 3% a year.[136,137] In comparison, treatment to an INR of 1.5 to 1.9 is less effective in reducing recurrent thrombosis (although it is better than placebo) and provides no significant reduction in bleeding risk.[136,137]

Warfarin therapy usually should be continued for 3 to 6 months. In a prospective study of oral anticoagulation therapy in patients with a first episode of venous thromboembolism, 6 weeks of therapy was adequate for patients with temporary, reversible risk factors for thrombosis (e.g., surgery, trauma, temporary immobilization, or use of oral contraceptives). On the other hand, 6 months of oral anticoagulation therapy was clearly superior for patients with idiopathic venous thromboembolism (who are presumed to have intrinsic risk factors).[138] The recurrence rate was quite high, however—approximately 12% at 2 years. On the basis of this evidence, at least 6 months of oral anticoagulation therapy is indicated in a patient with a first episode of idiopathic DVT or pulmonary embolism.

Further management should depend on results of the hypercoagulable workup (see above). A risk assessment for other predisposing factors will also be relevant.

Unfortunately, even when warfarin treatment is continued for 12 months after a first episode of idiopathic DVT, this does not reduce the risk of recurrent thrombosis once warfarin treatment is discontinued.[139] Generally, there is a rapid rebound phase of recurrent DVT of about 10% during the first 6 to 12 months after warfarin therapy. This suggests that there is a subset of patients (10% to 20%) who have a stronger tendency toward thrombosis and thus experience recurrences fairly soon after discontinuance of oral anticoagulation.

Persistent elevation of D-dimer levels may help identify patients who are more likely to have recurrent thrombosis. In a prospective study, D-dimer levels that remained elevated 1 month after the discontinuance of warfarin were associated with a higher recurrence risk (approximately threefold to eightfold), whereas normal D-dimer levels had a high negative predictive value for recurrence.[140] Thus, patients with elevated D-dimer levels merit more vigilant monitoring and consideration of long-term anticoagulation.

Because the optimal duration of long-term oral anticoagulation therapy for patients with thromboembolism remains undefined, this question is best addressed individually, on the basis of an estimation of the risk of recurrence [*see Table 7*].[141,142] Before undertaking long-term anticoagulation therapy in a high-risk

patient, the clinician must take the patient's risk of bleeding into account.

Of note, none of the published studies shows a significant difference in mortality between patients who receive long-term therapy and those who receive short-term therapy. There is also no evidence to suggest that prophylactic anticoagulation therapy improves overall survival. In historical studies, families with deficiencies of AT and protein C show no higher mortality than the general population displays.[143,144] Current clinical trials are studying the use of full-dose oral anticoagulation therapy for an extended period of time followed by an indefinite period of low-dose anticoagulation therapy. The results of these studies will help define the optimal long-term treatment for patients who are at high risk for recurrence of thromboembolism.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Chi JT, Chang HY, Haraldsen G, et al: Endothelial cell diversity revealed by global expression profiling. Proc Natl Acad Sci USA 100:10623, 2003

2. Martinelli I, Mannucci PM, De Stefano V, et al: Different risks of thrombosis in four coagulation defects associated with inherited thrombophilia: a study of 150 families. Blood 92:2353, 1998

3. Martinelli I, Cattaneo M, Panzeri D, et al: Risk factors for deep venous thrombosis of the upper extremities. Ann Intern Med 126:707, 1997

4. Prandoni P, Polistena P, Bernardi E, et al: Upper-extremity deep vein thrombosis: risk factors, diagnosis, and complications. Arch Intern Med 157:57, 1997

5. Abilgaard U: Antithrombin and related inhibitors of coagulation. Recent Advances in Blood Coagulation. Poller L, Ed. Churchill Livingstone, Edinburgh, Scotland, 1981, p 151

6. De Stefano V, Leone G, Mastrangelo S, et al: Clinical manifestations and management of inherited thrombophilia: retrospective analysis and follow-up after diagnosis of 238 patients with congenital deficiency of antithrombin III, protein C, protein S. Thromb Haemost 72:352, 1994

7. Demers C, Ginsberg JS, Hirsh J, et al: Thrombosis in antithrombin-deficient persons: report of a large kindred and literature review. Ann Intern Med 116:754, 1992

8. Menache D, O'Malley JP, Schorr JB, et al: Evaluation of the safety, recovery, half-life, and clinical efficacy of antithrombin III (human) in patients with hereditary antithrombin III deficiency. Blood 75:33, 1990

9. Miletich J, Sherman L, Broze G Jr: Absence of thrombosis in subjects with heterozygous protein C deficiency. N Engl J Med 317:991, 1987

10. Miletich JP, Prescott SM, White R, et al: Inherited predisposition to thrombosis. Cell 72:477, 1993

11. Boerger LM, Morris PC, Thurnau GR, et al: Oral contraceptives and gender affect protein S status. Blood 69:692, 1987

12. Vigano-D'Angelo S, D'Angelo A, Kaufman CE Jr, et al: Protein S deficiency occurs in the nephrotic syndrome. Ann Intern Med 107:42, 1987

13. Sallah S, Abdallah JM, Gagnon GA: Recurrent warfarin-induced skin necrosis in kindreds with protein-S deficiency. Haemostasis 28:25, 1998

14. Faioni EM, Franchi F, Asti D, et al: Resistance to activated protein C in nine thrombophilic families: interference in a protein S functional assay. Thromb Haemost 70:1067, 1993

15. Rees DC, Cox M, Clegg JB: World distribution of factor V Leiden. Lancet 346:1133, 1995

16. Svensson PJ, Dahlback B: Resistance to activated protein C as a basis for venous thrombosis. N Engl J Med 330:517, 1994

17. Simioni P, Prandoni P, Lensing AWA, et al: The risk of recurrent venous thromboembolism in patients with an Arg506 Gln mutation in the gene for factor V (factor V Leiden). N Engl J Med 336:399, 1997

18. Vandenbroucke JP, Koster T, Briet E, et al: Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. Lancet 344:1453, 1994

19. Rosendaal FR, Koster T, Vandenbroucke JP, et al: High-risk of thrombosis in patients homozygous for factor V Leiden (APC-resistance). Blood 85:1504, 1995

20. Middeldorp S, Henkens CMA, Koopman MMW, et al: The incidence of venous thromboembolism in family members of patients with factor V Leiden mutation and venous thrombosis. Ann Intern Med 128:15, 1998

21. Koeleman BPC, Reitsma PH, Allaart RC, et al: Activated protein C resistance as an additional risk factor for thrombosis in protein C–deficient families. Blood 84:1031, 1994

22. Zoller B, Berntsdotter A, Garcia de Frutos P, et al: Resistance to activated protein C as an additional genetic risk factor in hereditary deficiency of protein S. Blood 85:3518, 1995

23. van Boven HH, Reitsma PH, Rosendaal FR, et al: Factor V Leiden (FV R506Q) in families with inherited antithrombin deficiency. Thromb Haemost 75:417, 1996

24. Williamson D, Brown K, Luddington R, et al: Factor V Cambridge: a new mutation (Arg306ΔThr) associated with resistance to activated protein C. Blood 91:1140, 1998

25. Chan WP, Lee CK, Kwong YL, et al: A novel mutation of Arg306 of factor V gene in Hong Kong Chinese. Blood 91:1135, 1998

26. Zoller B, Svensson PJ, He X, et al: Identification of the same factor V gene mutation in 47 out of 50 thrombosis-prone families with inherited resistance to activated protein C. J Clin Invest 94:2521, 1994

27. Bertina RM: Laboratory diagnosis of resistance to activated protein C (APC-resistance). Thromb Haemost 78:478, 1997

28. van der Bom JG, Bots ML, Haverkate F, et al: Reduced response to protein C is associated with increased risk for cerebral vascular disease. Ann Intern Med 125:265, 1996

29. Fisher M, Fernandez JA, Ameriso SF, et al: Activated protein C resistance in ischemic stroke not due to factor V arginine506Δglutamine mutation. Stroke 27:1163, 1996

30. Rodeghiero F, Tosetto A: Activated protein C resistance and factor V Leiden mutation are independent risk factors for venous thromboembolism. Ann Intern Med 130:643, 1999

31. de Visser MCH, Rosendaal FR, Bertina RM: A reduced sensitivity to activated protein C in the absence of factor V Leiden increases the risk of venous thrombosis. Blood 93:1271, 1999

32. Bertina RM, Koeleman BPC, Koster T, et al: Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369:64, 1994

33. Kalafatis M, Bertina RM, Rand MD, et al: Characterization of the molecular defect in factor V R506Q. J Biol Chem 270:4053, 1995

34. Ridker PM, Hennekens CH, Lindpaintner K, et al: Mutation in the gene coding for coagulation factor V and the risk for myocardial infarction, stroke, and venous thrombosis in apparently healthy men. N Engl J Med 332:912, 1995

35. De Stefano V, Martinelli I, Mannucci PM, et al: The risk of recurrent deep venous thrombosis among heterozygous carriers of both factor V Leiden and the G20210A prothrombin mutation. N Engl J Med 341:801, 1999

36. Eichinger S, Weltermann A, Mannhalter C, et al: The risk of recurrent venous thromboembolism in heterozygous carriers of factor V Leiden and a first spontaneous venous thromboembolism. Arch Intern Med 162:2357, 2002

37. Dahlback B: Resistance to activated protein C caused by the factor V R506Q mutation is a common risk factor for venous thrombosis. Thromb Haemost 78:483, 1997

38. Middeldorp S, Meinardi JR, Koopman MM, et al: A prospective study of asymptomatic carriers of the factor V Leiden mutation to determine the incidence of venous thromboembolism. Ann Intern Med 135:322, 2001

39. Rosendaal FR, Doggen CJ, Zivelin A, et al: Geographic distribution of the 20210 G to A prothrombin variant. Thromb Haemost 79:706, 1998

40. Poort SR, Rosendaal FR, Reitsma PH, et al: A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 88:3698, 1996

41. The risk of recurrent venous thromboembolism in carriers and non-carriers of the G1691A allele in the coagulation factor V gene and the G20210A allele in the prothrombin gene. DURAC Trial Study Group. Duration of Anticoagulation. Thromb Haemost 81:684, 1999

42. Martinelli I, Sacchi E, Landi G, et al: High risk of cerebral-vein thrombosis in carriers of a prothrombin-gene mutation and in users of oral contraceptives. N Engl J Med 338:1793, 1998

43. Harker LA, Harlan JM, Ross R: Effect of sulfinpyrazone on homocysteine-induced endothelial injury and arteriosclerosis in baboons. Circ Res 53:731, 1983

44. Starkebaum G, Harlan JM: Endothelial cell injury due to copper-catalyzed hydrogen peroxide generation from homocysteine. J Clin Invest 77:1370, 1986

45. Lentz SR, Sadler JE: Inhibition of thrombomodulin surface expression and protein C activation by the thrombogenic agent homocysteine. J Clin Invest 88:1906, 1991

46. Nishinaga M, Ozawa T, Shimada K: Homocysteine, a thrombogenic agent, suppresses anticoagulant heparan sulfate expression in cultured porcine aortic endothelial cells. J Clin Invest 92:1281, 1993

47. Harpel PC, Chang VT, Borth W: Homocysteine and other sulfhydryl compounds enhance the binding of lipoprotein(a) to fibrin: a potential biochemical link between thrombosis, atherogenesis, and sulfhydryl compound metabolism. Proc Natl Acad Sci USA 89:10193, 1992

48. Mudd SH, Levy HL, Skovby F: Disorders of transulfuration. The Metabolic Basis of Inherited Disease. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill, New York, 1989, p 693

49. Malinow MR: Homocyst(e)ine and arterial occlusive diseases. J Intern Med 236:603, 1994

50. Naurath HJ, Joosten E, Riezier R, et al: Effects of vitamin B_{12}, folate, and vitamin B_6 supplements in elderly people with normal serum vitamin concentrations. Lancet 346:85, 1995

51. Clarke R, Daly L, Robinson K, et al: Hyperhomocysteinemia: an independent risk factor for vascular disease. N Engl J Med 324:1149, 1991

52. Selhub J, Jacques PF, Bostom AG, et al: Association between plasma homocysteine concentrations and extracranial carotid-artery stenosis. N Engl J Med 332:286, 1995

53. Den Heijer M, Koster T, Blom HJ, et al: Hyperhomocysteinemia as a risk factor for deep-vein thrombosis. N Engl J Med 334:759, 1996

54. Nygard O, Nordrehaug JE, Refsum H, et al: Plasma homocysteine levels and mortality in patients with coronary artery disease. N Engl J Med 337:230, 1997

55. Bostom AG, Jacques PF, Nadeau MR, et al: Post-methionine load hyperhomocysteinemia in persons with normal fasting total plasma homocysteine: initial results from The NHLBI Family Heart Study. Arteriosclerosis 116:147, 1995

56. Franken DG, Boers GHJ, Blom HJ, et al: Treatment of mild hyperhomocysteinemia in vascular disease patients. Arterioscler Thromb Vasc Biol 14:465, 1994

57. Bostom A, Cupples A, Jenner J, et al: Elevated plasma lipoprotein(a) and coronary heart disease in men aged 55 years and younger. JAMA 276:544, 1996

58. Wild SH, Fortmann SP, Marcovina SM: A prospective case-control study of lipoprotein(a) levels and Apo(a) size and risk of coronary heart disease in Stanford Five-City Project participants. Arterioscler Thromb Vasc Biol 17:239, 1997

59. Nagayama M, Shinohara Y, Nagayama T: Lipoprotein(a) and ischemic cerebrovascular disease in young adults. Stroke 25:74, 1994

60. Ridker PM, Stampfer MJ, Hennekens CH: Plasma concentration of lipoprotein(a) and the risk of future stroke. JAMA 273:1269, 1995

61. Marcovina SM, Albers JJ, Jacobs DR Jr, et al: Lipoprotein(a) concentrations and apolipoprotein(a) phenotypes in Caucasians and African-Americans. The CARDIA study. Arterioscler Thromb 13:1037, 1993

62. Hajjar KA, Gavish D, Breslow JL, et al: Lipoprotein (a) modulation of endothelial cell surface fibrinolysis and its potential role in atherosclerosis. Nature 339:303, 1989

63. Callow MJ, Verstuyft J, Tangirala R, et al: Atherogenesis in transgenic mice with human apolipoprotein B and lipoprotein(a). J Clin Invest 96:1639, 1995

64. Maher VM, Brown BG, Marcovina SM, et al: Effects of lowering elevated LDL cholesterol on the cardiovascular risk of lipoprotein(a). JAMA 274:1771, 1995

65. Gurakar A, Hoeg JM, Kostner G, et al: Levels of lipoprotein(a) decline with neomycin and niacin treatment. Atherosclerosis 57:293, 1985

66. Shewmon DA, Stock JL, Rosen CJ, et al: Tamoxifen and estrogen lower circulating lipoprotein(a) concentrations in healthy postmenopausal women. Arterioscler Thromb 14:1586, 1994

67. Martinez J: Congenital dysfibrinogenemia. Curr Opin Hematol 4:357, 1997

68. Aoki N, Moroi M, Sakata Y, et al: Abnormal plasminogen: a hereditary molecular abnormality found in a patient with recurrent thrombosis. J Clin Invest 61:1186, 1978

69. Estelles A, Gilabert J, Aznar J, et al: Changes in the plasma levels of type 1 and type 2 plasminogen activators in normal pregnancy and in patients with severe preeclampsia. Blood 74:1332, 1989

70. Prins MH, Hirsh J: A critical review of the evidence supporting a relationship between impaired fibrinolytic activity and venous thromboembolism. Arch Intern Med 151:1721, 1991

71. Wiman B: Plasminogen activator inhibitor-1 (PAI-1) in plasma: its role in thrombotic disease. Thromb Haemost 74:71, 1995

72. Meade TW: Fibrinogen in ischemic heart disease. Eur Heart J 16(suppl):31, 1995

73. Meade TW, Mellows S, Brozovic M, et al: Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. Lancet 2:533, 1986

74. Koster T, Blann AD, Briet E, et al: Role of clotting factor VIII in effect of von Willebrand factor on occurrence of deep-vein thrombosis. Lancet 345:152, 1995

75. O'Donnell J, Tuddenham EGD, Manning R, et al: High prevalence of elevated factor VIII levels in patients referred for thrombophilia screening: role of increased synthesis and relationship to the acute phase reaction. Thromb Haemost 77:825, 1997

76. Kyrle PA, Minar E, Hirschl M, et al: High plasma levels of factor VIII and the risk of recurrent venous thromboembolism. N Engl J Med 343:457, 2000

77. Antiphospholipid syndrome: clinical and immunologic manifestations and patterns of disease expression in a cohort of 1,000 patients. Euro-Phospholipid Project Group. Arthritis Rheum 46:1019, 2002

78. Rauch J, Janoff AS: Role of antibodies in understanding the interactions between anti-phospholipid antibodies and phospholipids. Phospholipid-Binding Antibodies. N Harris, T Exner, GRV Hughes, et al, Eds. CRC Press, Boca Raton, Florida, 1991, p 108

79. Rao LVM, Hoang AD, Rapaport SI: Mechanisms and effects of the binding of lupus anticoagulant IgG and prothrombin to surface phospholipid. Blood 88:4173, 1996

80. Smirnov MD, Triplett DT, Comp PC, et al: On the role of phosphatidylethanolamine in the inhibition of activated protein C activity by antiphospholipid antibodies. J Clin Invest 95:309, 1995

81. Shibata S, Harpel PC, Gharavi A, et al: Autoantibodies to heparin from patients with antiphospholipid antibody syndrome inhibit formation of antithrombin III–thrombin complexes. Blood 83:2532, 1994

82. Finazzi G, Brancaccio V, Ciavarella N, et al: Natural history and risk factors for thrombosis in 360 patients with antiphospholipid antibodies: a four-year prospective study from the Italian registry. Am J Med 100:530, 1996

83. Vila P, Hernandez MC, Lopez-Fernadez MF, et al: Prevalence, follow-up and clinical significance of the anticardiolipin antibodies in normal subjects. Thromb Haemost 72:209, 1994

84. Ginsburg KS, Liang MH, Newcomer L, et al: Anticardiolipin antibodies and the risk for ischemic stroke and venous thrombosis. Ann Intern Med 117:997, 1992

85. Silver RM, Porter TF, van Leeuween I, et al: Anticardiolipin antibodies: clinical consequences of "low titers." Obstet Gynecol 87:494, 1996

86. Selva-O'Callaghan A, Ordi-Ros J, Monegal-Ferran F, et al: IgA anticardiolipin antibodies: relation with other antiphospholipid antibodies and clinical significance. Thromb Haemost 79:282, 1998

87. Khamashta MA, Cuardrado MJ, Mujic F, et al: The management of thrombosis in the antiphospholipid-antibody syndrome. N Engl J Med 332:993, 1995

88. Rosove MH, Brewer MC: Antiphospholipid thrombosis: clinical course after the first thrombotic event in 70 patients. Ann Intern Med 117:303, 1992

89. Crowther MA, Ginsberg JS, Julian J, et al: A comparison of two intensities of warfarin for the prevention of recurrent thrombosis in patients with the antiphospholipid antibody syndrome. N Engl J Med 349:1133, 2003

90. Kearon C, Gent M, Hirsh J, et al: A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med 340:901, 1999

91. Moll S, Ortel TL: Monitoring warfarin therapy in patients with lupus anticoagulants. Ann Intern Med 127:177, 1997

92. Rand JH, Wu X-X, Andree HAM, et al: Pregnancy loss in the antiphospholipid-antibody syndrome—a possible thrombogenic mechanism. N Engl J Med 337:154, 1997

93. Kutteh WH, Rote NS, Silver R: Antiphospholipid antibodies and reproduction: the antiphospholipid antibody syndrome. Am J Reprod Immunol 41:133, 1999

94. Infante-Rivard C, David M, Gauthier R, et al: Lupus anticoagulants, anticardiolipin antibodies, and fetal loss: a case-controlled study. N Engl J Med 325:1063, 1991

95. Laskin CA, Bombardier C, Hannah ME, et al: Prednisone and aspirin in women with autoantibodies and unexplained recurrent fetal loss. N Engl J Med 337:148, 1997

96. Kutteh WH: Antiphospholipid antibody–associated recurrent pregnancy loss: treatment with heparin and low-dose aspirin is superior to low-dose aspirin alone. Am J Obstet Gynecol 174:1584, 1996

97. Rai R, Cohen H, Dave M, et al: Randomized, controlled trial of aspirin and aspirin plus heparin in pregnant women with recurrent miscarriage asssociated with phospholipid antibodies. BMJ 314:253, 1997

98. Kutteh WH, Ermel LD: A clinical trial for the treatment of antiphospholipid antibody–associated recurrent pregnancy loss with lower dose heparin and aspirin. Am J Reprod Immunol 35:402, 1996

99. Eldor A: Thrombophilia, thrombosis and pregnancy. Thromb Haemost 86:104, 2001

100. Warkentin TE: Heparin-induced thrombocytopenia: a ten-year retrospective. Ann Rev Med 50:129, 1999

101. Bauer TL, Arepally G, Konkle BA, et al: Prevalence of heparin-associated antibodies without thrombosis in patients undergoing cardiopulmonary bypass surgery. Circulation 95:1242, 1997

102. Warkentin TE: Heparin-induced thrombocytopenia: a clinicopathologic syndrome. Thromb Haemost 82:439, 1999

103. Amiral J, Bridey F, Dreyfus M, et al: Platelet factor 4 complexed to heparin is the target for antibodies generated in heparin-induced thrombocytopenia. Thromb Haemost 68:95, 1992

104. Warkentin TE, Levine MN, Hirsh J, et al: Heparin-induced thrombocytopenia in patients treated with low-molecular-weight heparin or unfractionated heparin. N Engl J Med 332:1330, 1995

105. Visentin GP, Ford SE, Scott JP, et al: Antibodies from patients with heparin-induced thrombocytopenia/thrombosis are specific for platelet factor 4 complexed with heparin or bound to endothelial cells. J Clin Invest 93:81, 1994

106. Warkentin TE, Kelton JG: Temporal aspects of heparin-induced thrombocytopenia. N Engl J Med 344:1286, 2001

107. Warkentin TE, Kelton JG: Delayed-onset heparin-induced thrombocytopenia and thrombosis. Ann Intern Med 135:502, 2001

108. Sallah S, Thomas DP, Roberts HR: Warfarin and heparin-induced skin necrosis and the purple toe syndrome: infrequent complications of anticoagulant treatment. Thromb Haemost 78:785, 1997

109. Greinacher A, Amiral J, Dummel V, et al: Laboratory diagnosis of heparin-associated thrombocytopenia and comparison of platelet aggregation test, and platelet factor 4/heparin enzyme linked immunosorbent assay. Transfusion 34:381, 1994

110. Warkentin TE, Greinacher A: Heparin-induced thrombocytopenia and cardiac surgery. Ann Thorac Surg 76:2121, 2003

111. Greinacher A, Volpol H, Potzsch B: Recombinant hirudin in the treatment of patients with heparin-induced thrombocytopenia (HIT). Blood 88(suppl):281a, 1996

112. Schiele F, Vuillemenot A, Kramarz P, et al: Use of recombinant hirudin as antithrombotic treatment in patients with heparin-induced thrombocytopenia. Am J Hematol 50:20, 1995

113. Lewis BE, Wallis DE, Berkowitz SD, et al: Argatroban anticoagulant therapy in patients with heparin-induced thrombocytopenia. Circulation 103:1838, 2001

114. Matsuo T, Yamada T, Yamanashi T, et al: Anticoagulant therapy with MD805 of a hemodialysis patient with heparin-induced thrombocytopenia. Thromb Res 58:663, 1990

115. Gillis S, Merin G, Zahger D, et al: Danaparoid for cardiopulmonary bypass in patients with previous heparin-induced thrombocytopenia. Br J Haematol 98:657, 1997

116. Potzsch B, Klovekorn WP, Madlener K: Use of heparin during cardiopulmonary bypass in patients with a history of heparin-induced thrombocytopenia. N Engl J Med 343:515, 2000

117. Callander N, Rapaport SI: Trousseau's syndrome. West J Med 158:364, 1993

118. Prandoni P, Lensing AWA, Buller HR, et al: Deep-vein thrombosis and the incidence of subsequent symptomatic cancer. N Engl J Med 327:1128, 1992

119. Callander NS, Varki N, Rao LVM: Immunohistochemical identification of tissue factor in solid tumors. Cancer 70:1194, 1992

120. Contrino J, Hair G, Kreutzer DL, et al: In situ detection of tissue factor in vascular endothelial cells: correlation with malignant phenotype of human breast disease. Nat Med 2:209, 1996

121. Falanga A, Gordon SG: Isolation and characterization of cancer procoagulant: a cysteine proteinase from malignant tissue. Biochemistry 24:5558, 1985

122. Falanga A, Consonni R, Marchetti M, et al: Cancer procoagulant and tissue factor are differentially modulated by all-*trans*-retinoic acid in acute promyelocytic leukemic cells. Blood 92:143, 1998

123. Wahrenbrock M, Borsig L, Le D, et al: Selectin-mucin interactions as a probable molecular explanation for the association of Trousseau syndrome with mucinous adenocarcinomas. J Clin Invest 112:853, 2003

124. Prins MH, Hettiarachchi RJK, Lensing AWA, et al: Newly diagnosed malignancy in patients with venous thromboembolism: search or wait and see? Thromb Haemost 78:121, 1997

125. Sorensen HT, Mellemkjaer L, Steffensen FH, et al: The risk of a diagnosis of cancer after primary deep venous thrombosis or pulmonary embolism. N Engl J Med 338:1169, 1998

126. Silverstein RL, Nachman RL: Cancer and clotting: Trousseau's warning. N Engl J Med 327:1163, 1992

127. Buller H, Ten Cate JW: Primary venous thromboembolism and cancer screening. N Engl J Med 338:1221, 1998

128. Cornuz J, Pearson SD, Creager MA, et al: Importance of findings on the initial evaluation for cancer in patients with symptomatic idiopathic deep venous thrombosis. Ann Intern Med 125:785, 1996

129. Prandoni P, Lensing AW, Picciolo A, et al: Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. Blood 100:3484, 2002

130. Low-molecular-weight heparin versus a coumarin for the prevention of recurrent venous thromboembolism in patients with cancer. Randomized Comparison of Low-Molecular-Weight Heparin versus Oral Anticoagulant Therapy for the Prevention of Recurrent Venous Thromboembolism in Patients with Cancer (CLOT) Investigators. N Engl J Med 349:146, 2003

131. Vandenbroucke JP, Koster T, Briet E, et al: Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. Lancet 344:1453, 1994

132. Helmerhorst FM, Bloemenkamp KWM, Rosendaal FR, et al: Oral contraceptives and thrombotic disease: risk of venous thromboembolism. Thromb Haemost 78:327, 1997

133. Hoibraaten E, Qvigstad E, Arnesen H, et al: Increased risk of recurrent venous thromboembolism during hormone replacement therapy—results of the randomized, double-blind, placebo-controlled estrogen in venous thromboembolism trial (EVTET). Thromb Haemost 84:961, 2000

134. Vandenbroucke JP, Rosing J, Bloemenkamp KW, et al: Oral contraceptives and the risk of venous thrombosis. N Engl J Med 344:1527, 2001

135. Rosendaal FR, Vessey M, Rumley A, et al: Hormonal replacement therapy, prothrombotic mutations and the risk of venous thrombosis. Br J Haematol 116:851, 2002

136. Ridker PM, Goldhaber SZ, Danielson E, et al: Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. N Engl J Med 348:1425, 2003

137. Kearon C, Ginsberg JS, Kovacs MJ, et al: Comparison of low-intensity warfarin therapy with conventional-intensity warfarin therapy for long-term prevention of recurrent venous thromboembolism. N Engl J Med 349:631, 2003

138. Schulman S, Rhedin AS, Lindmarker P, et al: A comparison of six weeks with six months of oral anticoagulant therapy after a first episode of venous thromboembolism. N Engl J Med 332:1661, 1995

139. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. Warfarin Optimal Duration Italian Trial Investigators. N Engl J Med 345:165, 2001

140. Palareti G, Legnani C, Cosmi B, et al: Predictive value of D-dimer test for recurrent venous thromboembolism after anticoagulation withdrawal in subjects with a previous idiopathic event and in carriers of congenital thrombophilia. Circulation 108:313, 2003

141. Bauer KA: Management of patients with hereditary defects predisposing to thrombosis including pregnant women. Thromb Haemost 74:94, 1995

142. De Stefano V, Finazzi G, Mannucci PM: Inherited thrombophilia: pathogenesis, clinical syndromes, and management. Blood 87:3531, 1996

143. Rosendaal FR, Heijboer H, Briet E, et al: Mortality in hereditary antithrombin deficiency: 1830 to 1989. Lancet 337:260, 1991

144. Allaart CF, Rosendaal FR, Noteboom WM, et al: Survival in families with hereditary protein C deficiency, 1820 to 1993. BMJ 311:910, 1995

145. Aster RH: Heparin-induced thrombocytopenia and thrombosis. N Engl J Med 332:1374, 1995

146. Koster T, Rosendaal FR, de Ronde H, et al: Venous thrombosis due to poor anticoagulant response to activated protein C: Leiden Thrombophilia Study. Lancet 342:1503, 1993

147. Triplett DA: Protean clinical presentation of antiphospholipid-protein antibodies (APA). Thromb Haemost 74:329, 1995

Acknowledgments

Figures 1 and 2 Seward Hung.

Figure 3 Dr. Rajeev Doshi.

XV COAGULATION DISORDERS

Lawrence L. K. Leung, m.d.

Bleeding or bruising that is spontaneous or excessive after tissue injury may be caused by coagulation defects, fibrinolysis, abnormal platelet number or function, abnormal vascular integrity, or a combination of these abnormalities. This chapter addresses hemorrhagic disorders associated with abnormalities in coagulation. Hemorrhagic disorders associated with quantitative or qualitative platelet abnormalities and disorders associated with blood vessels are discussed elsewhere [*see 5:XIII Platelet and Vascular Disorders*].

Disorders of coagulation may be inherited or acquired. Congenital coagulation disorders are rare and are most frequently caused by defects in single coagulation proteins, with the two X-linked disorders—factor VIII and factor IX deficiencies—accounting for the majority of defects. Acquired coagulation disorders are more common than the inherited disorders and are more complex in their pathogenesis. The most common acquired hemorrhagic disorders are vitamin K deficiency, drug-induced hemorrhage, and disseminated intravascular coagulation (DIC).

Hereditary Coagulation Disorders

The coagulation disorders appear clinically as either spontaneous hemorrhage or excessive hemorrhage after trauma or surgery. The patient history usually indicates whether the disorder is congenital or acquired. The hereditary disorders are characterized by their appearance early in life and by the presence of a single abnormality that can account for the entire clinical picture.

von Willebrand disease

Pathophysiology

von Willebrand disease (vWD), the most common hereditary bleeding disorder, is caused by a deficient or defective plasma von Willebrand factor (vWF). The gene encoding vWF is on chromosome 12. vWF has specific domains for binding clotting factor VIII, heparin, collagen, platelet GPIb, and platelet GPIIb-IIIa. These domains relate directly to the following functions of vWF: (1) its action as a carrier molecule for factor VIII:C, in which it protects the clotting factor from proteolysis and substantially prolongs its plasma half-life; (2) its promotion of primary platelet adhesion at high wall shear rates by linking platelets via their GPIb-IX-V receptor to subendothelial tissues at the wound site; and (3) its support of platelet aggregation by linking platelets via their GPIIb-IIIa receptors.[1] The vWF circulates as multimers that range in size from 0.5 million daltons (the dimer) to 20 million daltons. Even larger noncirculating multimers are present in endothelial cells, where they are stored in the Weibel-Palade bodies. The ultralarge vWF multimers are normally processed by the ADAMTS13 metalloprotease into smaller multimers as they are released from the endothelial cells. The vWF is released either into the circulation or abluminally, where it attaches to subendothelial collagen. Platelet α-granules also contain vWF, which is released when platelets are activated. The vWF multimers that are 12 million daltons or larger are the most effective in supporting platelet adhesion.

Laboratory Evaluation

The many variant forms of vWD differ in their clinical manifestations, laboratory abnormalities, and required therapies. Because vWF is a carrier protein for factor VIII, the activated partial thromboplastin time (aPTT) is prolonged when the vWF level is low. The platelet count is usually, but not invariably, normal. Bleeding time is generally prolonged but not sufficiently reliable to be used for diagnosis. An automated platelet function test utilizing a platelet function analyzer (PFA-100) has been shown to be a better screening test for vWD than the bleeding time.[2,3] In this assay, citrated whole blood is aspirated through a capillary tube under high shear rates onto a membrane coated with collagen in which a central aperture is made. Platelets are activated by either adenosine diphosphate (ADP) or epinephrine. The closure time is a measure of platelet-vWF interaction.

The diagnosis of vWD requires the determination of factor VIII and vWF levels. There are two caveats: (1) laboratory testing is notoriously variable and (2) the patient's blood group affects the vWF level—that is, patients with blood group O have lower vWF levels than those with blood group A, B, or AB, by as much as 30%.[4] The vWF level is measured by either immunologic or functional methods. The former is reported as a percentage of normal vWF antigen. Because vWF circulates in physiologically important multimeric forms, it is sometimes helpful to determine the multimeric composition of the vWF in the patient's plasma. This is especially useful in identifying type 2 vWD (see below). The functional level of vWF is tested by the ristocetin-induced platelet aggregation test. Ristocetin is added to a patient's platelet-rich plasma, where it causes vWF to bind to platelets via the GPIb-IX-V receptor, leading to platelet activation and aggregation. In some laboratories, formalin-fixed platelets are used and, after the addition of ristocetin, agglutination of fixed platelets is measured.

Clinical Variants

The classification scheme for variants of vWD comprises three major groups: type 1 is a partial quantitative deficiency of vWF, type 2 is a qualitative abnormality of vWF, and type 3 is a severe and virtually total quantitative deficiency of vWF [*see Table 1*].[5]

Type 1 Type 1 vWD is the most common form of vWD, accounting for 75% of cases. It is generally an autosomal dominant trait that usually appears in the heterozygous form. In many cases, a mutation in the vWF protein occurs such that the mutant provWF monomers form dimers normally with the wild-type provWF monomers, but the resulting dimers are trapped in the endoplasmic reticulum and cannot be secreted.[6] Patients with classic type 1 vWD have a lifelong history of mild to moderate bleeding, typically from mucosal surfaces. They may be unaware of a bleeding disorder until they undergo surgery or experience trauma, when bleeding may be severe. vWF antigen, factor VIII, and the ristocetin cofactor levels are all decreased.

Type 2 Type 2 vWD is characterized by qualitative abnormalities of vWF and a variable decrease in vWF antigen, factor VIII, and ristocetin cofactor. In type 2A, the largest multimers are absent; in type 2B, multimers bind excessively to platelets be-

Table 1 Classification and Differentiation of von Willebrand Disease

	Type 1	*Type 2A*	*Type 2B*	*Type 2M*	*Type 2N*	*Type 3*	*Pseudo–von Willebrand Disease*
Inheritance	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal dominant	Autosomal recessive	Autosomal dominant
Incidence	~75%	~20%	~5%	Rare	Rare	Uncommon	Uncommon
Cause	Deficiency of normal vWF	Abnormal vWF	Abnormal vWF	Abnormal vWF	Abnormal vWF	Severe deficiency of vWF	Abnormal platelet membrane
Template bleeding time	N or ↑	↑	↑	↑	N or ↑	↑↑	N or ↑
Factor VIII assay	↓	N or ↓	N or ↓	N or ↓	↓↓	↓↓	N or ↓
vWF antigen	↓	Variable	Variable	Variable	N	↓↓	Variable
Ristocetin cofactor (RIPA)	↓	↓	↑	↓	N	↓↓	↑
Plasma vWF multimer analysis	N	Only low-molecular-weight forms present	Only low- and intermediate-molecular-weight forms present	N	N	Variable	Only low- and intermediate-molecular-weight forms present

N—normal ↓—decreased ↑—increased vWF—von Willebrand factor

cause of a gain-of-function mutation; in type 2M, the abnormal vWF does not bind to GPIb-IX-V; and in type 2N, the binding site of vWF for factor VIII is mutated. All type 2 vWD variants, except for the rare type 2M, are characterized by a loss of the high-molecular-weight vWF multimers in the VWF polymer analysis.

Type 3 The rare homozygous or double heterozygous form (type 3 vWD) is characterized by severe hemorrhage, a long aPTT, and factor VIII levels of less than 5%.

Pseudo–von Willebrand disease A platelet form of vWD, which is termed pseudo–von Willebrand disease (pseudo–vWD), has been described in which an abnormal GPIb is present on platelets, causing excessive binding of normal plasma vWF to unstimulated platelets.

The mean level of vWF antigen is 100 IU/dl, but the population distribution of vWF levels is very broad, with the 95% values encompassing 50 IU/dl to 200 IU/dl. The reasons for this broad distribution in vWF levels are not completely understood, but it makes the commonly used threshold (vWF level at 2 SD [standard deviation] below the mean) inadequate for diagnosis of type 1 vWD. This problem with diagnosis is compounded by the fact that mild bleeding symptoms are extremely common in the general population. A recent survey estimates that 25% of men and 46% of women would give a positive history of bleeding symptoms, such as frequent epistaxis, easy bruising, and postpartum bleeding.[7] This suggests that type 1 vWD may be overdiagnosed, and it has been proposed that a more stringent diagnostic criterion be used—namely, limiting the diagnosis of type 1 vWD to patients with a vWF antigen level of less than 20 IU/dl.[6] Patients with modestly reduced vWF antigen levels (i.e., between 30 and 50 IU/dl) usually do not have identifiable vWF gene mutations and rarely cosegregate with a family history of bleeding. Patients with modestly reduced vWF levels (and no history of family bleeding) may have a modestly increased risk of bleeding.

Treatment

Mild or moderate types 1 and 2 1-Desamino-8-D-arginine vasopressin (DDAVP or desmopressin) is effective in the management of traumatic bleeding and before surgery in some patients with mild or moderate type 1 and type 2A vWD. The intravenous administration of DDAVP at a dosage of 0.3 mg/kg over a 15- to 30-minute period causes the release of large amounts of vWF from endothelial cell stores. The peak response usually occurs in 30 to 60 minutes and persists for up to 4 to 6 hours. Repeated DDAVP administrations over a 24-hour period are ineffective; tachyphylaxis follows depletion of the endothelial vWF stores. A DDAVP nasal spray (300 μg) can be used in the ambulatory treatment of patients with vWD, both for the management of bleeding episodes and as preparation for minor surgery.[8] The side effects of intravenous DDAVP are generally mild, including significant water retention and, rarely, thrombosis. Myocardial infarction has been reported. Because of the variability of response to DDAVP, a patient should be given a trial infusion of DDAVP before undergoing a planned procedure to determine whether the patient has an adequate response. Fibrinolysis inhibitor ε-aminocaproic acid (EACA), 3 g four times daily orally for 3 to 7 days, is also useful for dental procedures and minor bleeding events. Aspirin must be avoided.

Moderate and severe types 2 and 3 Patients with severe types 2A and 2B and with type 3 vWD generally require replacement therapy with Humate-P—a pasteurized intermediate-purity factor VIII concentrate that has a substantial amount of large vWF multimers—or with cryoprecipitate infusion containing vWF, factor VIII, and fibrinogen. Cryoprecipitate is generally not recommended. Transfusion of normal platelets can also be attempted on the grounds that platelet vWF can be hemostatically effective.[9]

Treatment during pregnancy Treatment is generally not needed during pregnancy in women with vWD. The plasma vWF level rises during the second and third trimesters but falls

rapidly after delivery. Late hemorrhage may occur 2 to 3 weeks post partum.[10] DDAVP is not used before delivery because of the concern that it may initiate contractions. Patients with type 2B vWD may have worsening thrombocytopenia during pregnancy because of the increase of abnormal vWF in plasma.[11]

HEMOPHILIA A

Hemophilia A affects one in 10,000 males and is characterized by a deficient or defective clotting factor VIII. The factor VIII gene, which is located on chromosome X at Xq28, is among the largest known human genes, spanning 186 kb and containing 26 exons. It encodes a protein of about 300,000 daltons, which circulates in plasma at very low concentrations and is normally bound to and protected by vWF. The primary source of factor VIII production is likely the liver, because hemophilia A can be corrected by liver transplantation.

Because the gene for factor VIII coagulant activity is carried on the X chromosome, the disease is manifested in hemizygous males. All of the daughters of a hemophiliac male will be carriers, whereas half of the sons of a mother who carries the hemophilia trait will be hemophiliacs and half of her daughters will be carriers. Families appear to be affected to varying degrees, depending on the specific nature of the genetic defect.

The clinical severity of hemophilia A correlates well with the measured levels of factor VIII coagulant activity. In general, factor VIII levels below 1% are associated with severe hemorrhagic symptoms; levels between 1% and 5%, with moderate hemophilia; and levels between 5% and 25%, with mild hemophilia [*see Table 2*].

Approximately one third of hemophilia A cases represent new mutations and have a negative family history. More than 300 abnormal factor VIII genes have been found. The abnormalities, which include point mutations, gene insertions, and gene deletions, result in either deficient factor VIII production or the generation of a functionally defective factor VIII. An inversion within intron 22 of the factor VIII gene, which results in a truncated and unstable factor VIII protein, is found in approximately 45% of all severely affected hemophilia A patients (factor VIII levels below 1%).[12]

Diagnosis

Diagnosis is made on the basis of the clinical picture, family history (positive in two thirds of cases), and the factor VIII coagulant activity level. In most cases, the type of bleeding history and a classic family history rule out vWD (which, unlike hemophilia A, is autosomally transmitted). Accurate DNA analysis for the common intron 22 inversion is now available in DNA testing laboratories. This test provides molecular diagnosis in approximately 45% of patients with severe hemophilia. However, it should not be ordered in patients with mild or moderate hemophilia.

Table 2 Correlation of Factor VIII Coagulant Activity Level with Bleeding Patterns in Hemophilia

Plasma Factor VIII Level	*Bleeding Pattern*
< 1%	Severe, presentation in first year of life, bleeding with circumcision, spontaneous hemarthrosis and deep-tissue bleeding
1%–5%	Moderate, presentation in childhood, bleeding after trauma, spontaneous hemarthrosis rare
5%–25%	Mild; may be present in childhood; bleeding after trauma, surgery, or dental extraction
25%–50%	May be undetected, may present in adulthood with bleeding after major trauma or surgery

Treatment

General principles The psychosocial aspects of hemophilia are complex. A child is often absent from school, is prone to crippling deformities, and runs a risk of drug addiction because of severe pain. Parents are understandably deeply concerned and sometimes troubled by guilt. Treatment should address these issues as well as the specific coagulation problem.

Factor VIII replacement Factor VIII concentrates are effective in controlling spontaneous and traumatic hemorrhage. Currently available factor VIII products derived from plasma have been purified to varying degrees (e.g., Humate-P [intermediate purity], Koate-HP [high purity], and Monoclate [ultrapurity]) and have undergone viral inactivation. There are two forms of full-length recombinant factor VIII (Recombinate and Kogenate), and they are safe and efficacious.[13,14] A second-generation, B-domain–deleted recombinant factor VIII (ReFacto) has also been developed and has been found to be effective and well tolerated.[15] The new recombinant factor VIII has the advantage of considerably higher specific activity, and the final formulation is stable without added human serum albumin, thus further reducing the potential risk of transmission of human infectious agents.

Dental prophylaxis is critically important to reduce the need for dental surgery. Aspirin must be avoided. Revaccination against hepatitis B virus also should be considered.

Genetic counseling should be part of the management program. Because of the difficult life severe hemophiliacs lead, a woman may opt to terminate pregnancy if she is certain of her carrier status or if she knows that her fetus is affected. There are several strategies for detecting carriers. In women who are carriers, factor VIII levels are typically about half of normal, whereas vWF levels are normal. The ratio of factor VIII to vWF for carriers is thus 0.5[16]; however, the error rate for this test is 10% to 17%. A more accurate genetic diagnosis for carriers can be made by a linkage approach. This approach is based on restriction fragment length polymorphisms (RFLPs) within the factor VIII gene. Analysis of the affected male will establish the pattern for the X chromosome carrying the hemophilia allele, without knowledge of the precise mutation. There are a large number of intragenic polymorphisms that allow the two copies of factor VIII genes in a female potential carrier to be distinguished, identifying her carrier state with high accuracy.

These molecular probes for RFLPs are now being used to determine the status of the fetus. Tissue can be obtained by amniocentesis or chorionic villus sampling.

Management of acute hemorrhage Deep tissue bleeding, hemarthrosis, and hematuria are the common forms of clinical bleeding in hemophilia A. Acute threats to life are posed by retroperitoneal hemorrhage; bleeding of the mouth, tongue, or neck that impairs the airway; and intracranial hemorrhage. Both ultrasonography and computed tomography can be used to identify retroperitoneal and intramuscular hematomas.

Principles of replacement therapy A plasma procoagulant level of 100% means that there is one unit of procoagulant per milliliter of plasma. Most persons have 40 ml of plasma per kilogram of body weight. Thus, from a determination of a patient's plasma volume and procoagulant level, the required amount of factor VIII replacement can be calculated. For example, in the case of a 60 kg boy who has an uncomplicated hemarthrosis of the knee and a baseline factor VIII of less than 1%, raising the factor VIII level to about 25% (0.25 U/ml) for 2 to 3 days should suffice. This patient has a plasma volume of 60 kg × 40 ml/kg, or 2,400 ml; he will need 0.25 U/ml × 2,400 ml, or 600 U of factor VIII, as an initial bolus. Another method of estimation is based on the following effect: the infusion of 1 U of factor VIII per kilogram increases factor VIII levels by 2%. Thus, dividing the desired level of factor VIII increase by 2 will give the number of U/kg required. In the example cited, 25% of factor VIII will require 12.5 U/kg, or 750 U, of factor VIII replacement.

The biologic half-life of factor VIII is approximately 12 hours; the dose can be repeated every 12 to 24 hours as long as needed to control the hemorrhage. In patients with hemarthrosis, the factor VIII level should be maintained for 2 to 3 days.

Elective surgery and dental extraction Dental work should be performed by a dentist who is experienced in the treatment of hemophiliacs. Before dental extraction, factor VIII is administered to raise the level to approximately 50%. The fibrinolytic inhibitor EACA is started the night before surgery at a loading dose of 3 g orally and continued at 2 to 3 g three or four times daily for 7 to 10 days after the dental work has been completed. Usually, further administration of factor VIII is not required.

Before elective surgery, the factor VIII level should be raised to 50% to 100% (0.5 to 1.0 U/ml) and then maintained above 50% for the next 10 to 14 days. Maintaining a higher concentration of factor VIII does not reduce the frequency of hemorrhage.[17]

DDAVP can be used to treat acute traumatic hemorrhage in patients with mild to moderate hemophilia and even to prepare such patients for minor surgery. DDAVP, which causes the release of vWF from endothelial cell stores, cannot be used repeatedly over many days, because such stores become depleted. DDAVP is infused at a dosage of 0.3 μg/kg in 50 ml of saline over 15 to 30 minutes and produces a prompt increase in factor VIII. The biologic half-life of the released factor VIII is 11 to 12 hours.

Management of an inhibitor Inhibitors tend to occur in more severely affected patients, who tend to receive the greatest number of factor VIII concentrates. In a recent single-center study of 431 patients over 3 decades, approximately 10% of patients with severe hemophilia A had an inhibitor (about a third were children younger than 10 years).[18] Not all inhibitors produce clinical problems. Assays for factor VIII inhibitors should be performed at regular intervals in all patients who have severe hemophilia.

Hemorrhage in a patient with an inhibitor can be life threatening. In a patient who has an inhibitor titer of less than 5 Bethesda units and who is not a vigorous antibody responder, a large amount of factor VIII concentrate should be administered in an attempt to overwhelm the antibody. Alternative therapies are porcine factor VIII (Hyate:C), prothrombin complex concentrates (e.g., Konyne and Proplex) to circumvent the factor VIII deficiency,[19,20] and activated prothrombin complex concentrates, such as Autoplex-T and FEIBA.

Recombinant activated factor VII (rFVIIa) has been found to be safe and efficacious in 70% to 85% of more than 1,500 bleeding episodes in hemophilia patients with inhibitors.[21,22] Recombinant factor VIIa may compete against the normal plasma unactivated factor VII for tissue factor binding and thus enhance thrombin generation at the bleeding site.[23] In addition, high-dose rFVIIa may bind to activated platelets and activate factors IX and X on the platelet surface in the absence of tissue factor.[24]

High-dose intravenous IgG has been used to treat nonhemophiliacs with acquired factor VIII inhibitors, but it is usually not efficacious in hemophiliacs with inhibitors (alloantibodies).

OTHER HEREDITARY HEMORRHAGIC DISORDERS

Factor IX Deficiency (Hemophilia B)

Factor IX deficiency (hemophilia B, or Christmas disease) is an X-linked disorder that is clinically indistinguishable from hemophilia A. The factor IX gene is on the X chromosome and produces a clotting factor that, like other vitamin K–dependent factors, has a region rich in γ-carboxylated glutamic acids. Calcium ion bridges link this region to the activated platelet cell surface, where factor IXa interacts with factor VIIIa to form a membrane-associated complex that efficiently converts factor X to factor Xa (intrinsic tenase) [*see 5:XII Hemostasis and Its Regulation*]. A large number of insertions, rearrangements, and deletions have been detected in the factor IX gene, and the hemophilia B syndrome is very heterogeneous.[25]

Diagnosis Diagnosis of hemophilia B requires a factor IX assay. The management principles are the same as those for hemophilia A. A plasma-derived pasteurized factor IX concentrate preparation (Mononine) displays excellent specific activity and a desirable biologic half-life of 18 to 34 hours. Recombinant factor IX is also commercially available.

Treatment The level of factor IX that is needed to control hemostasis in patients with hemophilia B is somewhat lower than the level of factor VIII required for the treatment of hemophilia A—about 15% to 20% for the former and 30% to 50% for the latter. Factor IX is a smaller molecule than factor VIII and is distributed in the albumin space. In making replacement calculations, it is assumed that administration of 1 U/kg of factor IX will increase the plasma level by 1%. Factor IX has a biphasic half-life, and plasma levels of this factor can be maintained by infusing the concentrate every 24 hours. Molecular biology techniques can now detect the factor IX deficiency carrier state and permit accurate genetic counseling. Sustained correction of a bleeding disorder in hemophilia B mice has been demonstrated by the gene therapy approach,[26] and clinical trials of factor IX in hemophilia B patients have been initiated.[27]

Factor XI Deficiency

Patients with factor XI deficiency frequently come to medical attention when a prolonged aPTT is detected in preoperative screening. It is most frequently observed in Ashkenazi Jews, although sporadic cases have been described in people of different ethnic origins. Factor XI deficiency is inherited as an autosomal recessive trait, and heterozygous deficiency is not associated with any clinical bleeding. Homozygous or compound heterozygous deficient patients generally have factor XI levels of less than 15%, and most bleeding manifestations in these patients are related to trauma or surgery, especially at sites of high

fibrinolytic activity (e.g., the urinary tract, tonsils, and tooth sockets).[28] Factor XI plays a supportive role in the clotting cascade. It is activated by thrombin and then functions in a positive feedback manner to augment thrombin generation and clot stabilization [*see 5:XII Hemostasis and Its Regulation*]. Thus, factor XI is primarily required in situations in which there is a significant hemostatic challenge; this explains the mild bleeding diathesis in factor XI deficiency.

For patients with severe factor XI deficiency (< 15%) who require surgery, fresh frozen plasma should be used to replenish the plasma level to more than 50%. EACA given orally at a dosage of 3 g three or four times daily is also effective for minor surgical or dental procedures. In a recent retrospective study of 62 women with severe factor XI deficiency, about 70% of the women did not have any postpartum hemorrhage. Of the 30% who did have postpartum hemorrhage, some had a history of recurrent clinical bleeding. Postpartum hemorrhage had no relationship with the particular abnormal factor XI genotype or with the level of factor XI.[29]

Fibrinolytic Abnormalities

Two uncommon congenital hemorrhagic disorders have been ascribed to abnormalities of fibrinolysis. Deficiency of α_2-antiplasmin, the major plasmin inhibitor, has led to uncontrolled plasmin activity with consequent hemorrhage. Enhanced fibrinolytic activity with occasional clinical bleeding has also been linked to deficiency of plasminogen activator inhibitor–1 (PAI-1), the physiologic inhibitor of tissue plasminogen activator (t-PA) and urokinase.[30] Treatment of both types of fibrinolytic abnormalities consists of the antifibrinolytic agents, tranexamic acid, or EACA, all of which block the binding of plasminogen and plasmin to fibrin.

Acquired Hemorrhagic Disorders

In addition to the hereditary coagulation disorders, several acquired disorders have been identified that can lead to generalized hemorrhage.

VITAMIN K DEFICIENCY

A vitamin K–dependent carboxylase in the liver synthesizes γ-carboxyglutamic acid, which is required for the biologic function of prothrombin and factors VII, IX, and X. In the absence of vitamin K, an abnormal prothrombin that lacks γ-carboxyglutamic residues is synthesized. Specific immunoassays performed in patients with vitamin K deficiency reveal a sharp decrease in normal prothrombin levels and a concomitant increased level of the abnormal des-γ-carboxyprothrombin. The same molecular derangement occurs with factors VII, IX, and X.[31]

Clinical Features and Diagnosis

Deficiency of vitamin K, which decreases levels of prothrombin and factors VII, IX, and X, occurs in cases of severe malnutrition, intestinal malabsorption, and obstructive jaundice. In patients with obstructive jaundice, bile salts, which are necessary for the emulsification and absorption of the fat-soluble vitamins (vitamins A, D, E, and K), cannot enter the intestine. Long-term ingestion of oral antibiotics suppresses vitamin K production by intestinal organisms. The effect is especially marked in patients who, because of their illness, are unable to consume a full, nourishing diet. Mucosal bleeding and ecchymoses occur if the procoagulant levels fall below 10% to 15% of normal.

Treatment

Therapy with vitamin K_1 (phytonadione), 10 to 25 mg/day orally for 2 or 3 days—or parenteral vitamin K_1 in cases of obstructive jaundice—usually reverses the abnormality in about 6 to 24 hours. If there is severe bleeding, fresh frozen plasma (approximately 3 units) restores procoagulant levels rapidly [*see* Principles of Replacement Therapy, *above*].

DRUG-INDUCED HEMORRHAGE

Warfarin-Induced Hemorrhage

Warfarin overdose or potentiation of its action by other drugs can cause very severe bleeding. The prothrombin time (PT) is prolonged, and mucosal bleeding, gastrointestinal bleeding, or ecchymosis is the usual pattern. If hemorrhage is significant, treatment to restore procoagulant levels to 30% of normal must be started with fresh frozen plasma. If there is no urgency, oral vitamin K_1 may be given. Generally, 1 to 2.5 mg of vitamin K_1 will be sufficient to return anticoagulation (defined as the international normalized ratio [INR]) to therapeutic levels after 16 hours. High doses of vitamin K_1 (10 mg or more) should be avoided because they may cause warfarin resistance for up to a week. Surreptitious warfarin use can be identified by a serum warfarin assay, which is available at special laboratories. Factitious or accidental ingestion of some of the long-acting vitamin K antagonists that are used as rodenticides (superwarfarins) may lead to prolonged bleeding symptoms. The synthesis of vitamin K–dependent clotting factors can be impaired for months after the initial exposure. Repeated administration of fresh frozen plasma, supplemented by massive doses of oral vitamin K_1 (100 to 150 mg/day), may be required to control bleeding symptoms.

Heparin-Induced Hemorrhage

Heparin overdose may not be obvious. It causes subcutaneous hemorrhages and deep tissue hematomas. The aPTT, PT, and thrombin time (TT) are vastly prolonged, but the reptilase time (RT) is normal. Intravenous protamine administration at a dosage of 1 mg/100 U of administered heparin terminates the overdose response. Because the half-life of protamine is shorter than that of heparin, a heparin rebound may occur, necessitating a second administration of protamine. Low-molecular-weight heparin (LMWH) preparations cause as much bleeding as standard unfractionated heparin. The ability of protamine to reverse the actions of LMWH is incomplete. Protamine (1 mg/100 U of anti–factor Xa) can be tried; if protamine treatment is unsuccessful, recombinant factor VIIa should be considered.

Hemorrhage Caused by Thrombolytic Therapy

Thrombolytic therapy is now used for acute myocardial infarction and for some cases of pulmonary embolism. The complications of thrombolytic therapy are essentially all hemorrhagic. In general, bleeding has been confined to relatively trivial oozing at vascular invasion sites, but subdural hematomas, cerebral infarction, and intracranial bleeding have also occurred. The thrombolytic agents, even those designed to be relatively fibrin specific, occasionally cause a significant systemic lytic state, with low levels of fibrinogen, factor V, and factor VIII. Furthermore, the generation of fibrinogen degradation products in turn interferes with the formation of a firm clot and with platelet function.

If thrombolytic therapy is suspected as the cause of bleeding in a particular patient, blood should be drawn quickly for an aPTT, a TT, an RT, and a fibrinogen level. If thrombolytic thera-

Table 3 Causes of Disseminated Intravascular Coagulation (DIC)

Events that initiate DIC
Septicemia
Cancer procoagulants (Trousseau syndrome)
Acute promyelocytic leukemia
Crush injury, complicated surgery
Severe intracranial hemorrhage
Retained conception products, abruptio placentae, amniotic fluid embolism
Eclampsia, preeclampsia
Major ABO blood mismatch, hemolytic transfusion reaction
Burn injuries
Heatstroke
Malignant hypertension
Extensive pump-oxygenation (repair of aortic aneurysm)
Giant hemangioma (Kasabach-Merritt syndrome)
Severe vasculitis
Events that complicate and propagate DIC
Shock
Complement pathway activation

py is the cause, the aPTT is prolonged, the fibrinogen level is usually below 50 mg/dl, and the TT and RT are both prolonged (as a result of the fibrin degradation products and decreased plasma fibrinogen).

The disorder is treated with cryoprecipitate (to raise the fibrinogen level to approximately 100 mg/dl), fresh frozen plasma, and platelet concentrates. If these measures do not stop the bleeding, the use of a specific antifibrinolytic agent such as EACA should be considered. EACA is given as a 5 g bolus I.V. over 30 to 60 minutes and then in a dosage of 1 g/hr by continuous I.V. infusion.[32]

DYSPROTEINEMIAS

The abnormal proteins associated with myeloma and macroglobulinemia can interfere with platelet function and cause clinical bleeding. These proteins can cause abnormalities in the coagulation tests as well. Both IgG and IgA myeloma proteins can cause prolonged TTs by interfering with the fibrin polymerization process. Less commonly, they may interact with specific clotting factors. Management is directed at the primary disease. Generally, these paraproteins do not cause clinically significant bleeding. If bleeding occurs, plasmapheresis rapidly corrects the defects by abruptly lowering the level of abnormal protein.

DISSEMINATED INTRAVASCULAR COAGULATION

Pathophysiology

Many different circumstances can cause DIC [*see Table 3*]. In each case, massive activation of the clotting cascade overwhelms the natural antithrombotic mechanisms, giving rise to uncontrolled thrombin generation. This condition results in thromboses in the arterial and venous beds, leading to ischemic infarction and necrosis that intensify the damage, release tissue factor, and further activate the clotting cascade. Massive coagulation depletes clotting factors and platelets, giving rise to consumption coagulopathy and bleeding. Tissue damage and the deposition of fibrin result in the release and activation of plasminogen activators and the generation of plasmin in amounts that overwhelm its inhibitor, α_2-antiplasmin. Plasmin degrades fibrinogen, prothrombin, and factors V and VIII and produces fibrin-fibrinogen degradation products. These substances interfere with normal fibrin polymerization and impair platelet function by binding to the platelet surface GPIIb-IIIa fibrinogen receptor. These fibrin-fibrinogen degradation products thus function as circulating anticoagulant and antiplatelet agents, exacerbating the consumption coagulopathy, and play a significant role in the bleeding diathesis [*see Figure 1*].

Endotoxin released during gram-negative septicemia enhances the expression of tissue factor, thereby accelerating procoagulant activation while suppressing thrombomodulin expression. These actions downregulate the protein C/protein S system, further promoting the tendency to DIC.[33] Experimental-endotoxemia models also showed marked suppression of fibrinolysis activity caused by a sustained increase in plasma PAI-1.[34] In patients with solitary or multiple hemangiomas associated with thrombocytopenia (Kasabach-Merritt syndrome), DIC is presumably initiated by prolonged contact of abnormal endothelial surface with blood in areas of vascular stasis. Platelets and fibrinogen are consumed in these hemangiomas, where fibrinolysis appears to be enhanced,[35] and such consumption can lead to hemorrhage. Certain snakebites can also produce DIC; several mechanisms have been identified. For example, Russell viper venom contains a protease that directly activates factor X and can produce almost instantaneous defibrination.

Clinical Consequences

The consequences of DIC depend on its cause and the rapidity with which the initiating event is propagated. If the activation occurs slowly, an excess of procoagulants is produced, predisposing to thrombosis. At the same time, as long as the liver can compensate for the consumption of clotting factors and the

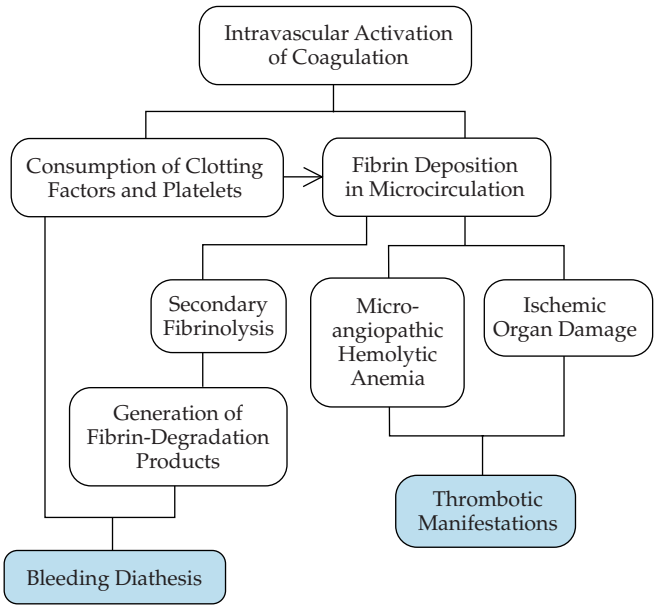


Figure 1 **In compensated disseminated intravascular coagulation (DIC), such as that which occurs in Trousseau syndrome, thrombotic manifestations predominate in the clinical presentation. In decompensated DIC, however, fibrin-fibrinogen degradation products exacerbate the consumption coagulopathy and play a significant role in the bleeding diathesis.**

bone marrow maintains an adequate platelet output, the bleeding diathesis will not be clinically apparent. The clinical situation consists of primarily thrombotic manifestations, which can be both venous thrombosis and arterial thrombosis [*see 5:XIV Thrombotic Disorders*]. Venous thromboses commonly involve deep vein thrombosis in the extremities or superficial migratory thrombophlebitis. Patients can also experience arterial thrombosis, leading to digital ischemia, renal infarction, or stroke. Arterial ischemia can in part be the result of emboli that originate from fibrin clots in the mitral valve, a condition termed nonbacterial thrombotic endocarditis, or marantic endocarditis. This condition is sometimes known as compensated, or chronic, DIC and accounts for Trousseau syndrome[36] (a chronic DIC caused by an underlying malignancy, most frequently pancreatic or other gastrointestinal cancer). The cancer cells may produce either tissue factor or another procoagulant that activates the clotting system.

If the reaction is brisk and explosive, the clinical picture is dominated by intravascular coagulation; depletion of platelets, fibrinogen, prothrombin, and factors V and VIII; and the production, by plasmin action, of fibrin degradation products, which further interfere with hemostasis. The clinical consequence is a profound systemic bleeding diathesis, with blood oozing from wound sites, intravenous lines, and catheters, as well as bleeding into deep tissues. The intravascular fibrin strands produce microangiopathic hemolytic anemia.

Diagnosis

Microangiopathic red blood cells on smear and a moderate to severe thrombocytopenia suggest the diagnosis. A number of laboratory abnormalities are present in DIC, depending on the stage of the DIC. Because of clotting factor depletion, the aPTT and PT are prolonged and the fibrinogen level is low. Because fibrin degradation products interfere with fibrin polymerization, the TT and RT are also prolonged. The level of fibrin degradation products, as measured by the D-dimer level, is elevated. Plasma plasminogen, protein C, and α_2-antiplasmin levels are also low because of consumption; however, these measurements are generally not required. In the case of compensated DIC, most of these parameters can be normal except for the elevation of the D-dimer level, which indicates the presence of intravascular cross-linked fibrin deposition and fibrinolysis. Sometimes, the fibrinogen level can even be high because fibrinogen is an acute-phase reactant. When the DIC becomes decompensated, consumption coagulation predominates and the other laboratory abnormalities listed are present (see above). Repetition at regular intervals of specific coagulation tests (see above), especially the platelet count, fibrinogen level, and D-dimer level, is critical. These tests provide a kinetic parameter that greatly aids in the assessment of the severity of the DIC and the choice of appropriate management.

Treatment

Currently, management must be directed at the primary disease to switch off the initiating event. This approach may involve chemotherapeutic treatment of an underlying tumor, administration of antibiotics and surgical drainage of an abscess, or emptying the uterus when complications of pregnancy have been the inciting cause. Hemodynamic support is essential. The use of antifibrinolytic agents such as EACA or aprotinin is contraindicated. Despite its bleeding complications, DIC is a severe hypercoagulable state, and these agents block the fibrinolytic system and may exacerbate its thrombotic complications. The administration of blood products, such as platelets, fresh frozen plasma, or cryoprecipitate, may add fuel to the fire and worsen the consumption coagulopathy. However, if clinical bleeding becomes significant, it is prudent to give vigorous blood product support.

The use of heparin in cases of acute DIC has not been established. Although heparin, by activating antithrombin (AT), is effective in inhibiting thrombin and therefore should be efficacious in the treatment of DIC, its use is generally limited to situations of chronic or compensated DIC. Heparin, given subcutaneously, is effective in the treatment of venous thrombosis in patients with Trousseau syndrome. In the case of decompensated DIC, in which bleeding is the major clinical manifestation, heparin may significantly exacerbate the bleeding and is therefore generally not indicated. The use of high-dose AT infusion has been advocated in this situation, but its efficacy has not been established by randomized studies.[37,38]

Recombinant human activated protein C (APC, or drotrecogin alfa [activated]) has been shown to significantly reduce mortality in patients with severe sepsis (mortality was 24.7% in patients given APC versus 30.8% in patients given placebo).[39] Although it is associated with a slightly increased risk of bleeding, APC appears to be an effective agent in the treatment of severe DIC in patients with sepsis, even for patients with normal protein C levels.[40] In large randomized trials, neither recombinant tissue factor pathway inhibitor (TFPI) nor AT concentrate reduced mortality in septic patients.[41,42] In cases of DIC associated with solitary or multiple hemangiomas, the hemangiomas can be excised when they are localized, and they occasionally show a good response to local irradiation. Attempts to control DIC associated with hemangiomas by the administration of heparin, corticosteroids, aspirin, and estrogens have not been successful. The key to successful management of DIC associated with certain snakebites is identification of the type of snake and prompt administration of appropriate antivenin.

ACQUIRED HEMOPHILIA AND OTHER DISORDERS OF CIRCULATING INHIBITORS

In addition to the hemorrhage caused by the circulating alloantibody inhibitors in severe hemophilias A and B, clinical hemorrhage is occasionally caused by circulating inhibitors directed against specific clotting factors, which seem to appear spontaneously. Because acquired autoantibody to factor VIII, which gives rise to the clinical picture of acquired hemophilia, is the most common of these circulating inhibitors, it will be described here in some detail, but many of the same principles apply to other inhibitors.

Autoantibodies to factor VIII are usually IgGa and, frequently, IgG4 and thus do not fix complement. They are usually directed against the functionally important A2 and C2 domains[43] on factor VIII. About half of the patients with an acquired factor VIII inhibitor have no identifiable associated disorder, but many disease states have been identified in the other patients, including autoimmune disorders such as systemic lupus erythematosus, lymphoproliferative disorders, plasma cell malignancies, drug reactions (e.g., reaction to penicillin), the postpartum state, and skin disorders.[44]

Diagnosis

Patients with an acquired factor VIII inhibitor commonly present with new-onset mucosal hemorrhages, hematomas, and ecchymoses but have a negative bleeding history. Typically, the

clinical picture of acquired hemophilia caused by factor VIII inhibitor occurs in an elderly patient or in a young woman during pregnancy or in the postpartum period. The laboratory hallmark of an acquired inhibitor to a clotting factor is a prolonged clotting time that is not corrected by mixing equal parts of the patient's plasma with normal plasma. In the case of factor VIII inhibitor, the PTT is prolonged and the PT and TT are normal. The antibody binds to factor VIII with complex kinetics such that the inhibitory effect becomes apparent only after prolonged incubation. Therefore, if an acquired factor VIII inhibitor is suspected, mixing studies should be performed after 5-minute and 60-minute incubations. The diagnosis can be confirmed by demonstration of a very low factor VIII level when other clotting factor levels are normal. Determination of the titer of the factor VIII inhibitor (expressed in Bethesda units [BU] per milliliter, with 1 BU/ml indicating a sufficient number of inhibitors to cause the complete inhibition of factor VIII in 1 ml of blood) is useful in choosing the appropriate therapy.

Treatment

The hemorrhage caused by circulating inhibitors may be clinically life threatening. Attempts at factor replacement are usually not successful, because the inhibitor inactivates the exogenous factor VIII. Occasionally, if the inhibitor has a low titer (e.g., < 2 to 3 BU/ml), massive factor VIII replacement can overwhelm the inhibitor. However, this treatment may trigger a significant anamnestic response resulting in increased levels of antibody, which complicates further management. Immunosuppressive therapy with a combination of cyclophosphamide (given either as a monthly intravenous pulse therapy or orally on a daily basis) and prednisone has been successful in most cases.[45,46] The inhibitor usually becomes undetectable after three or four monthly cycles of chemotherapy. In the case of severe or life-threatening hemorrhage in which there is insufficient time to reduce the level of inhibitor, porcine factor VIII can be administered, because the antibody usually displays low cross-reactivity.

Another alternative therapy for acute bleeding is the administration of procoagulant complexes, which may bypass the inhibitor block by providing large amounts of factor X and factor VII.[47] Still other therapeutic options include plasmapheresis and high-dose intravenous IgG, although the response rate for intravenous immune globulin (IVIg) appears to be quite low.[48] Recombinant activated factor VII (90 μg/kg given as an I.V. bolus every 2 to 3 hours) has been used successfully in patients with this condition. There is growing evidence that rituximab, given intravenously at 375 mg/m^2 once weekly for 4 weeks, is effective.[49,50] In patients with a very high titer inhibitor (>100 BU/ml), a combination of rituximab and cyclophosphamide may be required.

ACQUIRED VON WILLEBRAND DISEASE

Diagnosis

Patients with acquired von Willebrand disease, who are generally in their 50s and 60s and do not have a personal or family history of a bleeding disorder, present with mucocutaneous-type bleeding.[51] The workup is the same as that for vWD. The acquired form of the disease frequently occurs in the setting of underlying lymphoproliferative, myeloproliferative, or cardiovascular disease. A study showed that acquired vWD is quite common in patients with severe aortic stenosis. vWF abnormalities are directly related to the severity of aortic stenosis and improve after valve replacement.[52] Acquired vWD is also occasionally associated with angiodysplasia in patients with recurrent gastrointestinal bleeding. Frequently, a small monoclonal gammopathy is found on serum protein electrophoresis. The plasma antibody to vWF is functional in a minority of cases, as demonstrated by inhibition of vWF in a functional assay by mixing studies.[53] However, most cases involve nonneutralizing antibodies to vWF, which can be demonstrated by enzyme-linked immunosorbent assay (ELISA). Presumably, the antibody binds to vWF and causes its rapid clearance, leading to a low plasma vWF level. Nonimmune mechanisms (e.g., adsorption of vWF onto tumor cells) have also been described. Multimeric analysis of plasma vWF typically shows a decrease in the high-molecular-weight multimers, resembling type 2A vWD.

Treatment

DDAVP is useful in correcting the bleeding diathesis in about one third of cases of acquired vWD. High-dose intravenous IgG (1 g/kg I.V. daily for 1 to 2 days) generally garners a good temporary response, with an increase in the vWF level and a shortening of the aPTT, lasting from a few days to 2 weeks. If the patient has a defined lymphoproliferative, myeloproliferative, or autoimmune disease, the underlying disease should be treated. However, the response of acquired vWD to immunosuppressive therapy with cyclophosphamide and prednisone is generally not as favorable as the response in the case of acquired factor VIII inhibitor.

HEMORRHAGE CAUSED BY SEVERE LIVER DISEASE

Patients with severe liver disease may suffer life-threatening hemorrhages. The most frequent types are esophageal and gastrointestinal hemorrhages related to varices, gastritis, or peptic ulcer. There may also be bleeding from biopsy sites and during and after surgery. Mucosal and soft tissue bleeding may occur, but generally, this is not the dominant bleeding problem.

The coagulopathy of severe liver disease is complex and not well delineated. Because the liver is the major site of synthesis for all the clotting factors, decreased levels of multiple clotting factors are observed, including fibrinogen, prothrombin, factor V, and factor VII; factor VIII is excepted, presumably because it is an acute-phase reactant. An increased level of abnormal fibrinogen with reduced clotting capability is also observed in patients with cirrhosis.[54] In addition, there is reduced clearance of activated clotting factors by the liver. DIC appears to occur commonly in patients with cirrhosis[55] (presumably because of triggering of the clotting cascade by hepatic tissue damage), but its precise role in both acute fulminant hepatitis and chronic liver disease has not been firmly established. Moderate thrombocytopenia is common, resulting from a combination of decreased platelet production (from relative deficiency of thrombopoietin [TPO], because the liver is the major site of TPO synthesis) and increased platelet destruction from hypersplenism. Platelet function is generally maintained. There is also evidence of hyperfibrinolysis, but its contribution to the overall hemostatic defect is uncertain. The liver also synthesizes most of the natural anticoagulant proteins. AT, protein C, and protein S levels are decreased. The best screening tests for this disorder include the PT, aPTT, platelet count, fibrinogen level, and D-dimer level. Specific assays that may guide therapy include factor V, factor VII, and AT. Replacement for active bleeding is accomplished by administering fresh frozen plasma, cryoprecipitates, and platelets as required. Prothrombin-complex concentrates are not recommend-

ed, because they do not replenish all the deficient clotting factors and may exacerbate the DIC. In general, although the multiple hemostatic defects contribute to the bleeding diathesis in severe liver disease, hemodynamic and anatomic factors are the primary determinants in this situation.

PRIMARY FIBRINOLYSIS

Cases of generalized primary fibrinolysis are rare. Many of the early reports of primary fibrinolysis probably represented secondary fibrinolysis associated with DIC. Postprostatectomy hematuria may constitute a true example of hemorrhage caused by localized fibrinolysis. The high concentration of urokinase in the urine in this condition causes plasminogen to be converted to plasmin with resulting clot lysis. If other causes of persistent postoperative hematuria can be ruled out, the condition can be treated with oral or intravenous EACA. Local instillation of EACA by urethral catheter is also effective.

BLEEDING AFTER CARDIOPULMONARY BYPASS

A mild thrombocytopenia (approximately 100,000/μl) commonly occurs in patients after cardiopulmonary bypass surgery.[56] A significant acquired platelet function disorder develops in some patients, probably caused by contact between the platelets and the oxygenator apparatus, which in turn leads to partial platelet degranulation.[57] In addition to the release of platelet granule contents, activation of fibrinolysis may occur together with modest clotting factor depletion.[58] The hemorrhage in such cases generally responds to platelet transfusions. The use of DDAVP in this setting has been reported to reduce postoperative blood loss; however, a meta-analysis of 17 clinical trials showed only a modest beneficial effect.[59]

The bovine protease inhibitor aprotinin has been shown to reduce bleeding and transfusion requirements in patients undergoing cardiopulmonary bypass.[60] Aprotinin inhibits plasmin and may also attenuate the systemic inflammatory response by inhibition of the proinflammatory mediator kallikrein.[61] It reduces plasmin-mediated proteolysis of platelet membrane proteins and preserves platelet function.[62] Randomized clinical trials showed that the two antifibrinolytic agents EACA and tranexamic acid are equally efficacious as aprotinin in this setting.[63,64] Aprotonin should be reserved for patients who are likely to require blood transfusion, especially those undergoing second operations and those with preexisting hemostatic defects. Preoperative testing of hemostasis appears not to be useful.

During bypass surgery, patients are sometimes exposed to topical thrombin (fibrin glue), which is used for local hemostasis control. Generally, bovine thrombin and trace amounts of other clotting factors to which patients may develop antibodies are used in these preparations. The antibodies against bovine thrombin cause a prolongation of the TT but are innocuous in themselves. However, potentially serious complications arise when the antibodies cross-react with human thrombin. Some patients develop antibodies against bovine factor V that cross-react with human factor V and may lead to clinical bleeding.[65,66] A review of reported cases found that bovine thrombin-associated factor V antibodies developed in 40% to 66% of cardiac surgery patients and in 20% of neurosurgery patients, and clinical bleeding complications occurred in about one third of these cases.[67] Mixing studies utilizing the patient's plasma and normal plasma will reveal the presence of the inhibitors, and the measurement of the appropriate factor levels will allow the correct diagnosis to be made. Sometimes, plasmapheresis is required to control the acute bleeding.

EVALUATION OF POSTOPERATIVE BLEEDING

Serious hemorrhage during or after surgery is a complicated clinical problem requiring rapid diagnosis and prompt intervention. The first question is whether the bleeding has a local anatomic cause (e.g., unligated vessel) or is the result of a systemic hemostatic failure. If the patient is bleeding only in the operative area, it would suggest a local anatomic cause, such as an unligated bleeding vessel. The patient's bleeding history, especially with the results of prior surgical procedures, is extremely useful, but the available history may be inadequate or incomplete. A revealing clue to a systemic malfunction is bleeding at multiple sites, particularly areas other than that of the surgical wound. Bleeding around a catheter, from venipuncture sites, and from venous cutdowns is highly indicative of a hemorrhagic disorder.

Rapid assessment of the total clinical setting is imperative. The following questions should be addressed:

- Does the patient have underlying renal, hepatic, or malignant disease?
- Has the surgery required pump bypass techniques or the induction of hypothermia, or has the patient been in shock or been hypothermic?
- How many units of blood and blood products have been given and over what period of time?
- Were baseline screening procoagulant tests obtained before surgery, and is the patient's frozen plasma still available?

The differential diagnosis of postoperative hemorrhage should include a number of bleeding disorders [*see Table 4*].

Prompt resolution of postoperative bleeding requires a panel of coagulation tests—including aPTT, PT, TT, fibrinogen assay, and D-dimer—a platelet count, and a well-stained blood smear for evaluation of platelet and red cell morphology. This battery of tests should be performed immediately. More specialized studies can be obtained if there is evidence of a specific disorder.

Table 4 Differential Diagnosis of Postoperative Hemorrhage

Dilutional thrombocytopenia caused by massive transfusion
Acquired platelet function defect after cardiopulmonary bypass
Inadequate heparin neutralization
Disseminated intravascular coagulation
Coagulopathy caused by shock liver
Acquired antithrombin and anti–factor V inhibitors after exposure to fibrin glue
Heparin-induced thrombocytopenia
Thrombocytopenia caused by GPIIb-IIIa inhibitors (e.g., abciximab)
Hyperfibrinolysis after prostate surgery
Undiagnosed von Willebrand disease or hemophilia
Thrombocytopenia caused by posttransfusion purpura
False abnormalities in coagulation test results

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Recombinant factor VIIa has not been approved by the FDA for uses described in this chapter.

References

1. Clemetson KJ: Platelet GPIb-V-IX complex. Thromb Haemost 78:266, 1997

2. Fressinaud E, Veyradier A, Truchaud F, et al: Screening for von Willebrand disease with a new analyzer using high shear stress: a study of 60 cases. Blood 91:1325, 1998

3. Posan E, McBane RD, Grill DE, et al: Comparison of PFA-100 testing and bleeding time for detecting platelet hypofunction and von Willebrand disease in clinical practice. Thromb Haemost 90:482, 2003

4. Triplett DA: Laboratory diagnosis of von Willebrand's disease. Mayo Clin Proc 66:832, 1991

5. Sadler JE, Gralnick HR: A new classification for von Willebrand's disease. Blood 84:676, 1994

6. Sadler JE: New concepts in von Willebrand disease. Annu Rev Med 56:173, 2005

7. Sadler JE: Von Willebrand disease type I: a diagnosis in search of a disease. Blood 101:2089, 2003

8. Rose EH, Aledort LM: Nasal spray desmopressin (DDAVP) for mild hemophilia A and von Willebrand disease. Ann Intern Med 114:563, 1991

9. Castillo R, Monteagudo J, Escolar G, et al: Hemostatic effects of normal platelet transfusion in severe von Willebrand's disease patients. Blood 77:1901, 1991

10. Ito M, Yoshimura K, Toyoda N, et al: Pregnancy and delivery in patients with von Willebrand's disease. J Obstet Gynaecol Res 23:37, 1997

11. Rick ME, Williams SB, Sacher RA, et al: Thrombocytopenia associated with pregnancy in a patient with type IIb von Willebrand's disease. Blood 69:786, 1987

12. Antonarakis SE, Rossiter JP, Young M, et al: Factor VIII inversions in severe hemophilia A: results of an international consortium study. Blood 86:2206, 1995

13. Bray GL, Gomperts ED, Courter S, et al: The Recombinate Study Group: a multicenter study of recombinant factor VIII (Recombinate): safety, efficacy, and inhibitor risk in previously untreated patients with hemophilia A. Blood 83:2428, 1994

14. Seremetis S, Lusher JM, Abildgaard CF, et al: Human recombinant DNA-derived antihemophilic factor (factor VIII) in the treatment of hemophilia A: conclusions of a 5-year study of home therapy. The KOGENATE Study Group. Haemophilia 5:9, 1999

15. Courter SG, Bedrosian CL: Clinical evaluation of B-domain deleted recombinant factor VIII in previously untreated patients. Semin Hematol 38(suppl 4):52, 2001

16. Green PP, Mannucci PM, Briet E, et al: Carrier detection in hemophilia A: a cooperative international study: II. The efficacy of a universal discriminant. Blood 67:1560, 1986

17. Rochat C, McFadyen ML, Schwyzer R, et al: Continuous infusion of intermediate-purity factor VIII in haemophilia A patients undergoing elective surgery. Haemophilia 5:181, 1999

18. Yee TT, Pasi KJ, Lilley PA, et al: Factor VIII inhibitors in haemophiliacs: a single-center experience over 34 years, 1964–97. Br J Haematol 104:909, 1999

19. Hough RE, Hampton KK, Preston FE, et al: Recombinant VIIa concentrate in the management of bleeding following prothrombin complex concentrate-related myocardial infarction in patients with haemophilia and inhibitors. Br J Haematol 111:974, 2000

20. Lusher JM: Controlled clinical trials with prothrombin complex concentrates. Prog Clin Biol Res 150:277, 1984

21. Hedner U, Glazer S, Falch J: Recombinant activated factor VII in the treatment of bleeding episodes in patients with inherited and acquired bleeding disorders. Transfusion Med Rev 7:78, 1993

22. Negrier C, Lienhart A: Overall experience with NovoSeven. Blood Coagul Fibrinolysis 11 (suppl 1):S19, 2000

23. van't Veer C, Golden NJ, Mann KG: Inhibition of thrombin generation by the zymogen factor VII: implications for the treatment of hemophilia A by factor VIIa. Blood 95:1330, 2000

24. Monroe DM, Hoffman M, Allen GA, et al: The factor VII–platelet interplay: effectiveness of recombinant factor VIIa in the treatment of bleeding in severe thrombocytopathia. Semin Thromb Hemost 26:373, 2000

25. Roberts HR, Eberst ME: Current management of hemophilia B. Hematol Oncol Clin North Am 7:1269, 1993

26. Wang L, Takabe K, Bidlingmaier SM, et al: Sustained correction of bleeding disorder in hemophilia B mice by gene therapy. Proc Natl Acad Sci USA 30:3906, 1999

27. White GC II: Gene therapy in hemophilia: clinical trials update. Thromb Haemost 86:172, 2001

28. Salomon O, Seligsohn U: New observation on factor XI deficiency. Haemophilia 10(suppl 4):184, 2004

29. Salomon O, Steinberg DM, Tamarin I, et al: Plasma replacement therapy during labor is not mandatory for women with severe factor XI deficiency. Blood Coagul Fibrinolysis 16:37, 2005

30. Fay WP, Parker AC, Condrey LR, et al: Human plasminogen activator inhibitor-1 (PAI-I) deficiency: characterization of a large kindred with a null mutation in the PAI-1 gene. Blood 90:204, 1997

31. Furie B, Furie BC: Molecular and cellular biology of blood coagulation. N Engl J Med 326:800, 1992

32. Sane DC, Califf RM, Topol EJ, et al: Bleeding during thrombolytic therapy for acute myocardial infarction: mechanisms and management. Ann Intern Med 111:1010, 1989

33. Esmon CT, Fukudome K, Mather T, et al: Inflammation, sepsis, and coagulation. Haematologica 84:254, 1999

34. Levi M: Disseminated intravascular coagulation: what's new? Crit Care Clin 21:449, 2005

35. Hall GW: Kasabach-Merritt syndrome: pathogenesis and management. Br J Haematol 112:851, 2001

36. Rickles FR, Levine MN, Edwards RL: Hemostatic alterations in cancer patients. Cancer Metastasis Rev 11:237, 1992

37. Jochum M: Influence of high-dose antithrombin concentrate therapy on the release of cellular proteases, cytokines, and soluble adhesion molecules in acute inflammation. Semin Hematol 32:14, 1995

38. Fourrier F, Chopin C, Huart JJ, et al: Double-blind, placebo-controlled trial of antithrombin III concentrates in septic shock with disseminated intravascular coagulation. Chest 104:882, 1993

39. Bernard GR, Vincent JL, Laterre PF, et al: Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 344:699, 2001

40. Dhainaut JF, Yan SB, Joyce DE, et al: Treatment effects of drotrecogin alfa (activated) in patients with severe sepsis with or without overt disseminated intravasular coagulation. J Thromb Haemost 2:1924, 2004

41. Abraham E, Reinhart K, Opal S, et al: Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: a randomized controlled trial. JAMA 290:238, 2003

42. Warren BL, Eid A, Singer P, et al: Caring for the critically ill patient: high-dose antithrombin III in severe sepsis: a randomized controlled trial. JAMA 286:1869, 2001

43. Scandella DH, Nakai H, Felch M, et al: In hemophilia A and autoantibody inhibitor patients: the factor VII A2 domain and light chain are most immunogenic. Thromb Res 101:377, 2001

44. Ludlam CA, Morrison AE, Kessler C: Treatment of acquired hemophilia. Sem Hematol 31:16, 1994

45. Lian C-Y, Larcada AF, Chiu AY-Z: Combination immunosuppressive therapy after factor VIII infusion for acquired factor VIII inhibitor. Ann Intern Med 110:774, 1989

46. Shaffer LG, Phillips MD: Successful treatment of acquired hemophilia with oral immunosuppressive therapy. Ann Intern Med 127:206, 1997

47. Morrison AE, Ludlam CA, Kessler C: Use of porcine factor VIII in the treatment of patients with acquired hemophilia. Blood 81:1513, 1993

48. Crenier L, Ducobu J, des Grottes JM, et al: Low response to high-dose intravenous immunoglobulin in the treatment of acquired factor VIII inhibitor. Br J Haematol 95:750, 1996

49. Wiestner A, Cho HJ, Asch AS, et al: Rituximab in the treatment of acquired factor VIII inhibitors. Blood 100:3426, 2002

50. Stasi R, Brunetti M, Stipa E, et al: Selective B-cell depletion with rituximab for the treatment of patients with acquired hemophilia. Blood 103:4424, 2004

51. Federici AB, Rand JH, Bucciarelli P: Acquired von Willebrand syndrome: data from an international registry. Thromb Haemost 84:345, 2000

52. Vincentelli A, Susen S, Le Tourneau T, et al: Acquired von Willebrand syndrome in aortic stenosis. N Engl J Med 349:343, 2003

53. Mohri H, Motomura S, Kanamori H, et al: Clinical significance of inhibitors in acquired von Willebrand syndrome. Blood 91:3623, 1998

54. Francis JL, Armstrong DJ: Acquired dysfibrinogenemia in liver disease. J Clin Pathol 35:667, 1982

55. Stein SF, Harker LA: Kinetic and functional studies of platelets, fibrinogen, and plasminogen in patients with hepatic cirrhosis. J Lab Clin Med 99:217, 1986

56. Pouplard C, May MA, Regina S, et al: Changes in platelet count after cardiac surgery can effectively predict the development of pathogenic heparin-dependent antibodies. Br J Haematol 128:837, 2005

57. Kestin AS, Valeri CR, Khuri SF, et al: The platelet function defect of cardiopulmonary bypass. Blood 82:107, 1993

58. Bolan CD, Alving BM: Pharmacologic agents in the management of bleeding disorders. Transfusion 30:541, 1990

59. Cattaneo M, Harris AS, Stromberg U, et al: The effect of desmopressin on reducing blood loss in cardiac surgery: a meta-analysis of double-blind, placebo-controlled trials. Thromb Haemost 74:1064, 1995

60. Lemmer JH, Dilling WE, Morton JR, et al: Aprotinin for primary coronary artery bypass grafting: a multicenter trial of three dose regimens. Ann Thorac Surg 62:1659, 1996

61. Khan MM, Gikakis N, Miyamoto S, et al: Aprotinin inhibits thrombin formation and monocyte tissue factor in simulated cardiopulmonary bypass. Ann Thorac Surg 68:473, 1999

62. Day JR, Punjabi PP, Randi AM, et al: Clinical inhibition of the seven-transmembrane thrombin receptor (PAR1) by intravenous aprotinin during cardiothoracic surgery. Circulation 110:2597, 2004

63. Greilich PE, Brouse CF, Whitten CW, et al: Antifibrinolytic therapy during cardiopulmonary bypass reduces proinflammatory cytokine levels: a randomized, double-blind, placebo-controlled study of epsilon-aminocaproic acid and aprotinin. J Thorac Cardiovasc Surg 126:1498, 2003

64. Hekmat K, Zimmermann T, Kampe S, et al: Impact of tranexamic acid vs. aprotinin on blood loss and transfusion requirements after cardiopulmonary bypass: a prospective, randomised, double-blind trial. Curr Med Res Opin 20:121, 2004

65. Zehnder JL, Leung LL: Development of antibodies to thrombin and factor V with recurrent bleeding in a patient exposed to topical bovine thrombin. Blood 76:2011, 1990

66. Berruyer M, Amiral J, French P, et al: Immunization by bovine thrombin used with fibrin glue during cardiovascular operations. J Thorac Cardiovasc Surg 105:892, 1993

67. Streiff MB, Ness PM: Acquired FV inhibitors: a needless iatrogenic complication of bovine thrombin exposure. Transfusion 42:18, 2002

I ORGANS AND CELLS OF THE IMMUNE SYSTEM

John R. David, m.d.
Cox Terhorst, ph.d.

The Characteristics of the Immune System

The immune system mediates the individual's relationship with the microbial environment. Immunity involves innate, or natural, responses and highly specific acquired, or adaptive, responses. The essential difference between the two types of immunity lies in the means by which microorganisms are recognized. In innate immunity, glycolipids and macromolecules with repeat patterns that are unique to infectious organisms are recognized by cell surface receptors on macrophages, dendritic cells, natural killer (NK) and NK T (NKT) cells, as well as by the complement system. In acquired immunity, lymphocytes use very specific antigen receptors to recognize infectious agents and other antigens, either directly or when processed by antigen-presenting cells (APCs), such as dendritic cells. Thus, an interplay exists between innate and acquired immunity at the level of the APC. Once an otherwise healthy person has had an infection with bacteria or with a virus, the immune system recognizes that pathogen and prevents its recurrence. In addition, the immune system has the remarkable capacity to discriminate between antigens, even if their structures are closely related. Lymphocytes can also react to self-antigens, causing autoimmunity.

The immune response needs to be able to distinguish between self and nonself. Otherwise, T cells and antibodies would constantly be attacking autologous cells, tissue components, or even commensal bacteria. In the 1950s, Sir Frank Macfarlane Burnet first proposed that in the prenatal state, the interaction of self-antigens with antigen-specific lymphocytes leads to the elimination of self-reactive lymphocytes and hence to immunologic tolerance.[1] When immunologic tolerance breaks down, the antibodies and sensitized (antigen-reactive) cells that are directed against self-antigens cause autoimmune diseases [*see 6:IX Immunologic Tolerance and Autoimmunity*].

Lymphocytes

There are two major groups of lymphocytes, the T cells (also called thymus-derived lymphocytes, or T lymphocytes) and the B cells (also called bone marrow–derived lymphocytes, or B lymphocytes). T cells and B cells make up 80% to 95% of the peripheral blood lymphocytes.

T cells and B cells have a vast power of antigen recognition. Two unique features underlie this power: (1) a B cell family of variable genes, which combined can recognize an almost infinite number of antigens; and (2) a T cell family of variable genes with only a slightly more limited capacity. Neither T cells nor B cells constitute a homogeneous population of cells; each group comprises a number of subgroups that can be differentiated by the constant region of their receptors, by specific sets of developmentally expressed surface markers, by their location in lymphoid organs, and by their function. The binding of combinations of monoclonal antibodies to surface receptors is currently the most specific technique used to identify the major subsets of these cells.

T cells

Mature T cells express either αβ T cell receptors (TCR-αβ) or γδ TCR (TCR-γδ) in a complex with the CD3 proteins. CD4 is expressed on 50% to 65% of peripheral T αβ cells, and CD8 is expressed on 25% to 35% of peripheral T αβ cells. Usually, T γδ have no CD4 or CD8. Although CD4 and CD8 are expressed together on cortical thymocytes, only one or the other is expressed on the complementary subsets of mature thymocytes and peripheral T αβ cells ($CD4^+$ and $CD8^+$ T cells) [*see Figure 1*]. $CD4^+$ T cells recognize antigen when the antigen is presented in association with major histocompatibility complex (MHC) class II molecules (HLA-D and HLA-DR) or in association with CD1d, the latter being NKT cells. $CD8^+$ T cells recognize antigen in the context of MHC class I molecules (HLA-A, HLA-B, and HLA-C) [*see 6:III Immune Response Mechanisms*]. The context of antigen recognition by T γδ is unknown.

$CD4^+$ helper T (T_H) cells can be further differentiated into T_{H1} and T_{H2} cells on the basis of the cytokines they produce.[2] T_{H1} cells secrete interleukin-2 (IL-2) and interferon gamma (IFN-γ), which are important for cell-mediated immunity. T_{H2} cells secrete IL-4, IL-5, IL-6, IL-10, and IL-13, which are critical for antibody production. The cytokines that are produced by each of these cell types also influence the other cell type. For example, the IFN-γ produced by T_{H1} cells can inhibit the function of T_{H2} cells, whereas IL-10, which is secreted by T_{H2} cells, monocytes, macrophages, and B cells, can inhibit the function of T_{H1} cells. In addition to having a helper function, T_{H1} cells can induce inflammatory cascades leading to autoimmunity, as occurs in inflammatory bowel diseases and rheumatoid arthritis.

CD4 on the surface of helper T cells plays an important role in HIV infection. In early infection, the virus uses CD4 as a coreceptor, together with CCR5, which itself is a receptor for several chemokines, including RANTES (regulated on action, normal T cell expressed and secreted), macrophage inflammatory protein–1α (MIP-1α), and MIP-1β. One percent of whites are homozygous for a defect in the CCR5 receptor and are resistant to HIV infection. The chemokine receptor CCR4, which is a receptor for stromal cell–derived factor–1 (SDF-1), is involved in late HIV infection.[3-5]

B cells and plasma cells

B cells are precursors of the immunoglobulin-producing cells (plasma cells) of the immune system and are identified by the presence of immunoglobulin on their surface. These surface membrane immunoglobulin–positive ($SmIg^+$) cells constitute 5% to 15% of the peripheral blood lymphocytes. The majority of B cells have both IgM and IgD on their surface; about one quarter of all B cells have only IgM or IgD on their surface. One percent of B cells exhibit IgG or IgA.

On the surface of B cells is the complement receptor 2 (CR2 or CD21), which binds C3d/C3dg and Epstein-Barr virus. FcγRIIb (CD32) is the main Fc receptor on B cells, which is involved in B

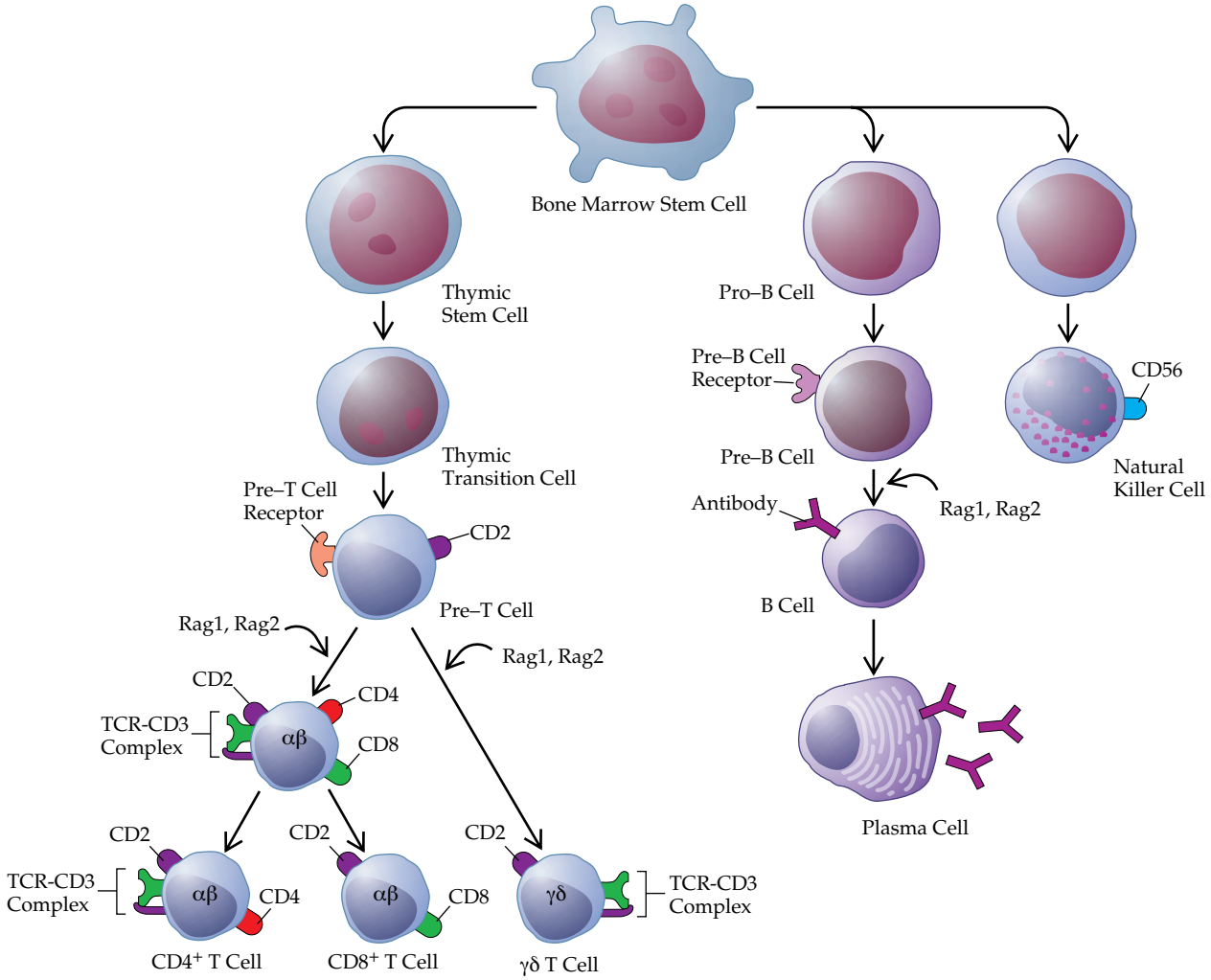


Figure 1 **Lineages of cells of the immune system and of blood cells all begin with the stem cell. Stem cells that differentiate to generate B cells reside in the bone marrow, and those that produce T cells migrate from the bone marrow to the thymus. T cell maturation involves the progressive expression of selected cell surface markers and the activation of various genes, including α, β, γ, and δ genes that code for the chains that make up the αβ and γδ T cell receptors (TCRs). Individuals lacking the Rag1 or Rag2 enzyme do not progress past the pre–T or pre–B cell stage and therefore have no lymphocytes. Positive and negative selection of T cells occurs at the so-called double-positive ($CD4^+$,$CD8^+$) stage. Natural killer cells develop both in the bone marrow and in the thymus.**

lymphocyte activation. A common marker that is used to identify B cells is CD19, which forms a larger complex with CD21 and CD81 (target for antiproliferative antigen–1 [TAPA-1]).[6]

B1 cells, a subset of B cells, develop early and have a very long life. B1 cell progenitors are found in fetal liver and in embryonic omentum but not in adult bone marrow. B1 cells that express CD5 on their surface are referred to as B1a, and B1 cells that lack CD5 are called B1b. B1 cells are frequently associated with autoantibody production. They also produce substantial amounts of IL-10.[7]

Under the influence of antigen, T cells, and accessory cells, B cells differentiate into plasma cells, the mature antibody-producing cells [*see Figure 1*]. Plasma cells are larger than lymphocytes and are characterized by an eccentric round nucleus with coarse heterochromatin arranged in a cartwheel pattern. Plasma cells have a highly basophilic cytoplasm and a well-developed endoplasmic reticulum, often organized in parallel concentric layers. Plasma cells may be distended with granular material, which consists of the antibody they are producing [*see Figure 2*]. Sometimes, one or more of the endoplasmic cisterns are distended by large inclusions called Russell bodies. These bodies are aggregates of incompletely formed immunoglobulin molecules. Plasma cells no longer bear surface immunoglobulin. They are also end cells, which means they do not divide. The immature precursors of plasma cells, the plasmablasts, are difficult to distinguish from lymphoblasts and large lymphocytes. Plasma cells are not normally found in the peripheral blood.

NATURAL KILLER CELLS

Natural killer cells are large granular lymphocytes that lack the TCR-CD3 complex characteristic of T cells or the SmIg characteristic of B cells. A bone marrow–derived stem cell is the precursor of T, B, and NK cells [*see Figure 1*]. In vitro, NK cells can kill a variety of tumor cells and virus-infected cells in a nonspecific manner; that is, they do not require previous sensitization or the presence of antibody to be cytotoxic. The granules contain pore-forming proteins that can mediate cell lysis. NK cells express killing inhibitory receptors (KIR) that recognize class 1 MHC molecules. Thus, NK cell functions are inhibited by cells that express MHC class 1 but are activated by cells lacking MHC class 1. Human NK cells express a second group of inhibitory re-

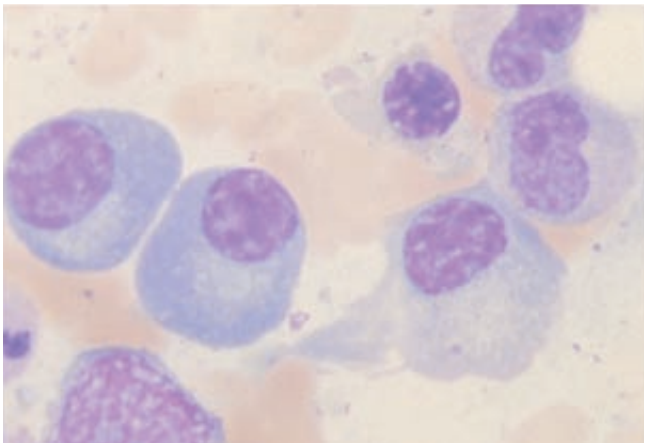


Figure 2 **Plasma cells are the antibody-producing cells of the immune system. They differentiate from B cells; are 6 to 20 μm in diameter; and have an eccentric nucleus, a highly basophilic cytoplasm, and a prominent, clear juxtanuclear area that contains the Golgi apparatus and the diplosome.**

ceptors, which comprises two subunits: a variable subunit NKG and the invariant cell surface structure CD94. Its ligand is unknown. IL-12 stimulates NK cells to proliferate and to produce IFN-γ, which is important for a number of immune reactions.[8,9]

Monocytes, Macrophages, and Dendritic Cells

Monocytes belong to the mononuclear phagocytic system, previously called the reticuloendothelial system. They are large mononuclear cells that constitute 3% to 8% of the peripheral blood leukocytes. Their cytoplasm is much more abundant than that of the lymphocytes. Usually, their nucleus is eccentric and either oval or kidney shaped [*see Figure 3*]. Lysosomes filled with degradative enzymes appear as small vacuoles in the cytoplasm. Monocytes originate from promonocytes, which are rapidly dividing precursors in the bone marrow. When the mature cells enter the peripheral blood, they are called monocytes; when they leave the blood and infiltrate tissues, they undergo additional changes and are then known as macrophages. Other cells derived from this lineage include Kupffer cells, alveolar macrophages, microglia, and osteoclasts.

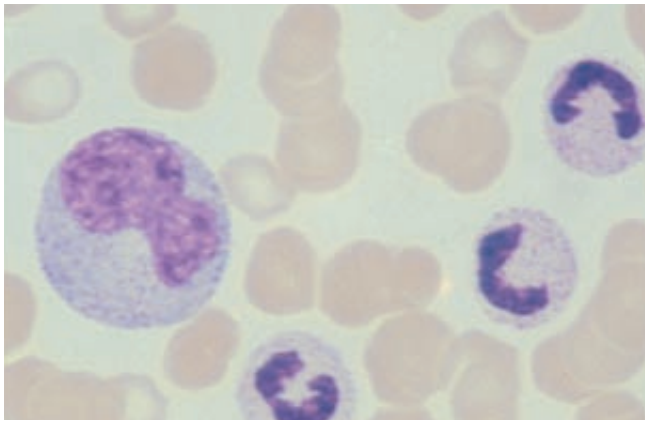


Figure 3 **A monocyte (large cell at left), which can reach 17 μm in diameter, has abundant basophilic cytoplasm and a large eccentric nucleus.**

Macrophages contain pattern recognition receptors (Toll-like receptors [TLRs] 1 through 10)[10] and scavenger receptors. Furthermore, receptors for antibody and complement enhance their ability to phagocytose organisms that are coated with these substances. The antibody receptors recognize the Fc portion of IgG1, IgG3, and IgE. There are two complement receptors, CR1 and CR3. CR1 has a high affinity for the complement component C3b and a lower affinity for iC3b and C4b. CR3, also called macrophage antigen–1 (MAC-1), interacts with iC3b as well as with certain carbohydrate molecules, including carbohydrate-containing antigens of the protozoan *Leishmania*. Through these receptors, macrophages act as effector cells, attacking microorganisms and neoplastic cells and removing foreign material.

Of equal importance, macrophages present processed antigen to lymphocytes and thus play a major role in the induction of acquired immune responses. A small amount of MHC class II antigen is present on monocytes, and its expression is greatly increased when macrophages are activated. Macrophages can be activated by a number of cytokines, including IFN-γ, granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage-activating factor (MAF), and migration inhibitory factor (MIF) [*see 6:III Immune Response Mechanisms*]. Cytokines such as IL-4 and transforming growth factor–β (TGF-β) antagonize this activation.

Macrophages themselves produce a large number of soluble substances that are important in the immune response and in the process of inflammation. These substances include enzymes such as plasminogen activating factor and elastase; growth factors such as GM-CSF; cytokines such as IL-1, IL-6, IL-10, IL-12, and tumor necrosis factor–α (TNF-α); factors that are critical for combating microorganisms, such as oxygen metabolites and nitric oxide; complement components for the classical and the alternative pathway; MIPs; and factors that promote tissue repair, such as fibroblast growth factor (FGF).

Dendritic cells are the professional APCs that engage specific responses by T cells.[11-13] They are present where antigens and microorganisms have first contact with the body—for example, in the skin (Langerhans cells) and the gastrointestinal and respiratory tracts. The function of the dendritic cells is discussed [*see* Lymph Nodes and Spleen, *below*].

Lymphoid Organs and Lymphocyte Traffic

The immune system consists of a number of lymphoid organs, including the thymus, lymph nodes, spleen, and tonsils; aggregates of lymphoid tissue in nonlymphoid organs, such as Peyer patches in the gut; clusters of lymphoid cells dispersed throughout the connective and epithelial tissues of the body, as well as throughout the bone marrow and blood; and a variety of individual cells that travel from the various lymphoid organs to the rest of the body. The lymphocytes are derived from precursors in the bone marrow: the T cells develop in the thymus and are then exported to the periphery, whereas the B and NK cells develop in the bone marrow and then go out to the periphery. Of the nonlymphoid hematopoietic cells, the monocytes, macrophages, and dendritic cells are key elements of innate and acquired immunity, whereas granulocytes (e.g., neutrophils, eosinophils, basophils, and mast cells) and platelets play important accessory roles in the immune system.

THE THYMUS

The thymus, which originates from the third and fourth pharyngeal pouches of the embryo, lies in the anterior mediastinum

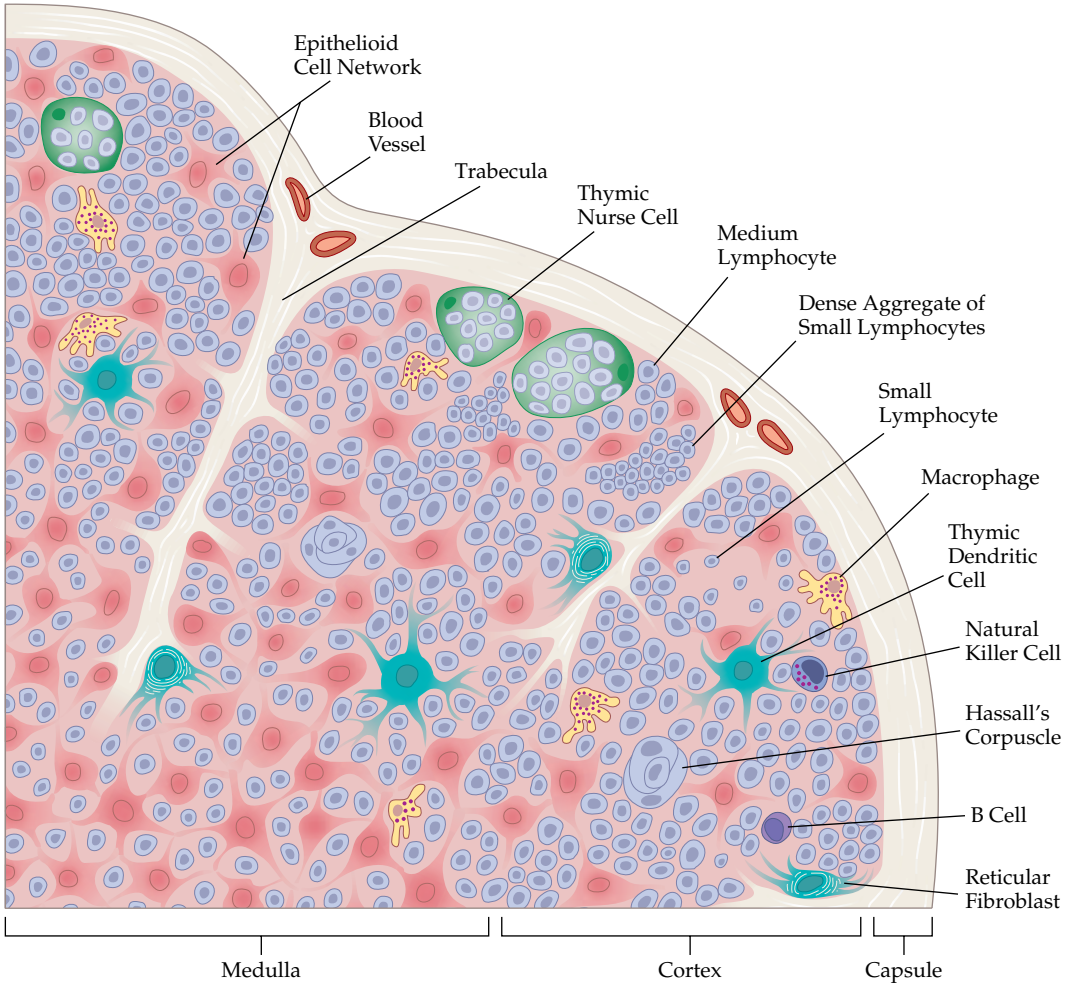


Figure 4 **Many lobules make up the thymus gland. Most of the lymphocytes in the cortex are immature, rapidly dividing cells that can readily be destroyed by cortisone. During maturation, they move to the medulla, where they become immunocompetent and resistant to cortisone. From there, they migrate to the secondary lymphoid organs, including the lymph nodes and the spleen. Cell division and maturation are influenced by the epithelial cells; dense aggregates of these cells form bodies known as Hassall corpuscles.**

and consists of many lobules, each containing a cortex and a medulla [*see Figure 4*]. Bone marrow–derived T cell precursors enter the thymus, congregating first in the subcapsular area. They develop into cells expressing the TCR-αβ–CD3 complex and subsequently acquire the potential to react with different peptides bound to MHC. These cortical $CD4^+$,$CD8^+$ thymocytes undergo either negative or positive selection, which involves complex mechanisms [*see* Development of the T Cell Repertoire, *below*]. Only a small percentage of positively selected thymocytes, $CD4^+$ (MCH class II recognizing) or $CD8^+$ (MHC class 1 recognizing), migrate to the medulla and then move into the peripheral lymphoid system. It is unclear whether all γδ T cells differentiate in the thymus.

Mature T cells emigrate through the wall of the postcapillary venules of the medulla to enter the bloodstream and subsequently home to the lymphoid system's peripheral organs. Once there, the lymphocytes leave the bloodstream, again through the postcapillary venules, and travel further into the T cell regions of the peripheral lymphoid organs. These include the inner cortex of the lymph nodes; the periarterial sheaths of the spleen; and the intranodular areas in Peyer patches, the tonsils, and the appendix. Some of the T cells in the intestinal mucosa (intraepithelial lymphocytes) are thought to differentiate outside the thymus.

Persons who are born without a thymus have lymphocytopenia, with a marked depletion or an absence of T cells. The T cell zones of the peripheral lymphoid system are also devoid of lymphocytes. There is marked impairment of cell-mediated immunity, and antibody responses that require cooperation from T cells (except the IgM response) are severely impaired [*see 6:VIII Deficiencies in Immunoglobulins and Cell-Mediated Immunity*]. The thymus involutes with age, which may explain the development of immune system deficiencies in elderly persons.

LYMPH NODES AND SPLEEN

The major sites of initial T cell activation are in the lymph nodes and in the spleen, where blood-borne lymphocytes and lymph-borne antigen, soluble mediators, and cells converge [*see Figure 5*]. Lymph and cells enter the node in the subcapsular region, through the afferent lymphatics; percolate into the subcapsular sinus; and leave through the efferent lymphatics in the hilum. In the spleen, the lymphocytes are concentrated in the

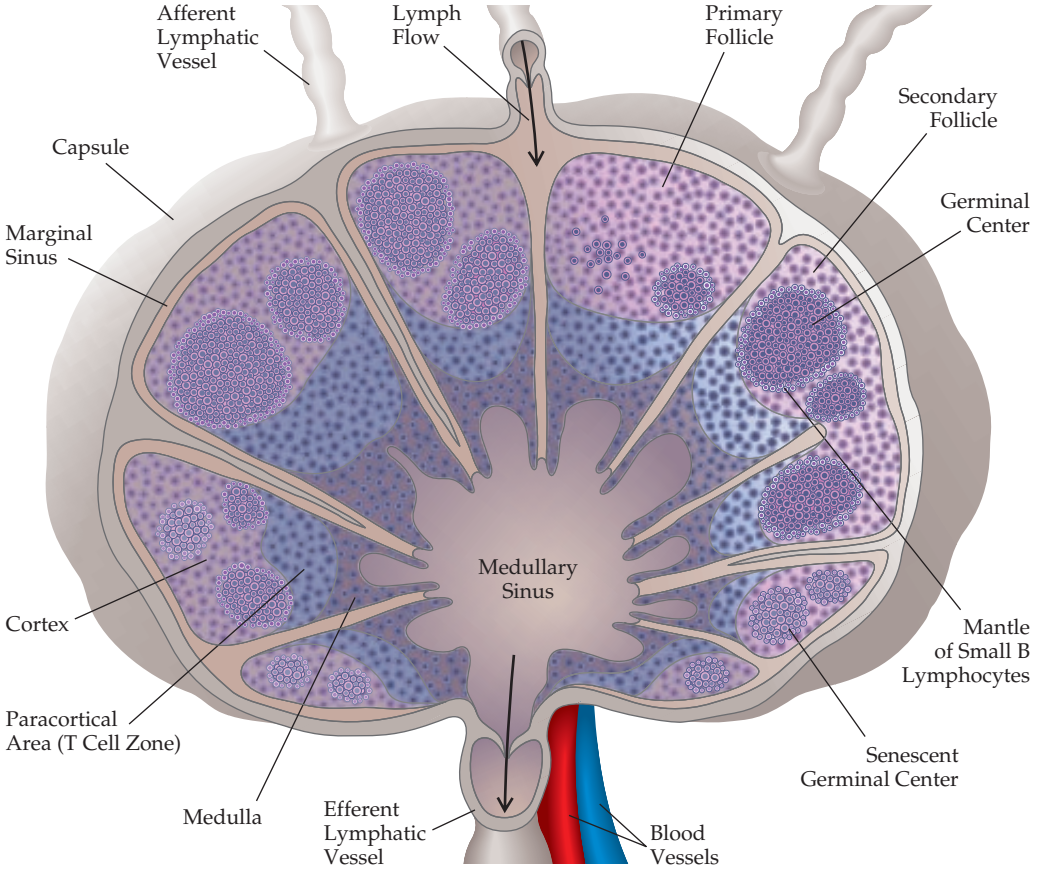


Figure 5 **Lymph containing lymphocytes, antigen, and soluble mediators drains from surrounding tissues and enters the lymph nodes via afferent lymphatic vessels. The lymph node consists of a cortex and a medulla. The outermost area of the cortex has B cells organized into lymphoid follicles, and its deep, or paracortical, area consists mainly of T cells and dendritic cells. After B cells encounter antigens for which they have receptors, along with the appropriate T cells, central areas of marked B cell proliferation called germinal centers develop in the lymphoid follicles. As these reactions die out, the germinal centers become senescent. In the medulla are medullary cords, which are strings of macrophages and plasma cells.**

white pulp, which consists of follicles with germinal centers that surround the central arterioles [*see Figure 6*].

Lymph Nodes

The infrastructure of the lymph node is an extensive reticular network where APCs and T cells meet and interact. For example, bone marrow–derived dendritic cells in the skin pick up antigen and travel through the lymphatics to the draining lymph node. The cells then migrate through the floor of the subcapsular sinus of the lymph node to the interfollicular regions, where they settle in the reticular network of the paracortex as interdigitating dendritic cells (IDCs). T cells from the blood migrate through specialized postcapillary venules, known as high endothelial venules (HEVs), and migrate along the same reticular network, where they come in contact with the numerous antigen-presenting IDCs.[14]

The germinal centers contain B cells, which are derived from stem cells and differentiate in the bone marrow [*see Figure 1*]. The lymphoid follicles contain follicular dendritic cells (FDCs), which are not derived from the bone marrow. FDCs, which are found only in lymphoid follicles, express complement receptors CR1, CR2, CR3, and Fc receptors. These receptors enable the FDCs to present antigen to activated B cells in the form of antigen-antibody-complement complexes. FDCs can retain these complexes for a long time.

Active germinal centers are surrounded by a mantle of B cells (follicular mantle cells) that express IgD on their surface and can mature into plasma cells that produce IgM antibodies only. The B cells in the germinal center (centrocytes) undergo class switching to produce the other isotypes, such as IgG, IgA, and IgE. The B cells with high-affinity antibody on their surfaces are thought to be selected by binding to the antigen-retaining FDCs. The B cells that are not selected die by apoptosis (programmed cell death). The most mature B cell, the memory cell, is also found in the germinal center and can develop into plasma cells producing all the antibody isotypes.

Bone marrow–derived dendritic cells, found in the paracortical areas of the lymph node, are professional APCs that play a crucial role in initiating T cell–dependent immune responses.[12-14] These cells are also found as immature cells in nonlymphoid organs, especially in the epidermis, where they are called Langerhans cells. Through their many receptors, particularly TLR1 through TLR10 and scavenger receptors, immature dendritic cells efficiently internalize and process antigens and present antigen through their MHC. There are two subsets of dendritic cells, myeloid and plasmacytoid, which produce distinct cytokines. Some cytokines (e.g.,

IL-1β and TNF-α), promote the migration of the dendritic cells via afferent lymphatics to the lymph nodes. At this location, they mature, lose their ability to phagocytose, and express critical costimulatory molecules CD80 (B7-1) and CD86 (B7-2); these costimulatory molecules enhance their ability to present antigen to T cells. Immature dendritic cells are also found in lymph nodes, where these cells can phagocytize antigen entering via the afferent lymphatics and then mature into APCs.

MUCOSAL AND SKIN IMMUNE RESPONSES

Lymphocytes are also found in various other locations. Gut-associated lymphoid tissue includes Peyer patches and the appendix. These gut-associated lymphoid tissues contain regions with concentrations of T cells or B cells similar to those found in germinal centers. Specialized epithelial cells termed M cells are thought to have a unique ability to take up and present antigen to the adjacent lymphocytes.[15] M cells are found close to Peyer patches. Other lymphocytes in the intestine are the lamina propria lymphocytes (LPLs), found in the villi, and the intraepithelial lymphocytes (IELs), found between epithelial cells. Migration and adherence of LPLs are in part dictated by integrins and selectins, which are surface molecules on tissues and cells that mediate cell interactions and homing. Mucosa-associated lymphoid cells are also found in the respiratory tract and genitourinary tract. The specialized immune system of the skin contains Langerhans cells in the epidermis (specialized myeloid dendritic cells) and a higher concentration of γδ T cells than elsewhere.[16]

LYMPHOCYTE CIRCULATION

There are three major types of lymphocyte circulation: (1) the seeding of the stem cells from the fetal liver or bone marrow to the primary lymphoid organs and the subsequent differentiation and distribution of these cells to the peripheral lymphoid system, (2) the recirculation of lymphocytes from blood to lymph to blood, and (3) the distribution of effector cells to particular parts of the body. Lymphocytes circulate continuously from blood to tissues and back to the blood. However, the trafficking of naive T cells (CD45RA) is different from that of activated effector or memory cells (CD45RO). Naive T cells recirculate through the secondary lymphoid tissues, such as the lymph nodes, spleen,

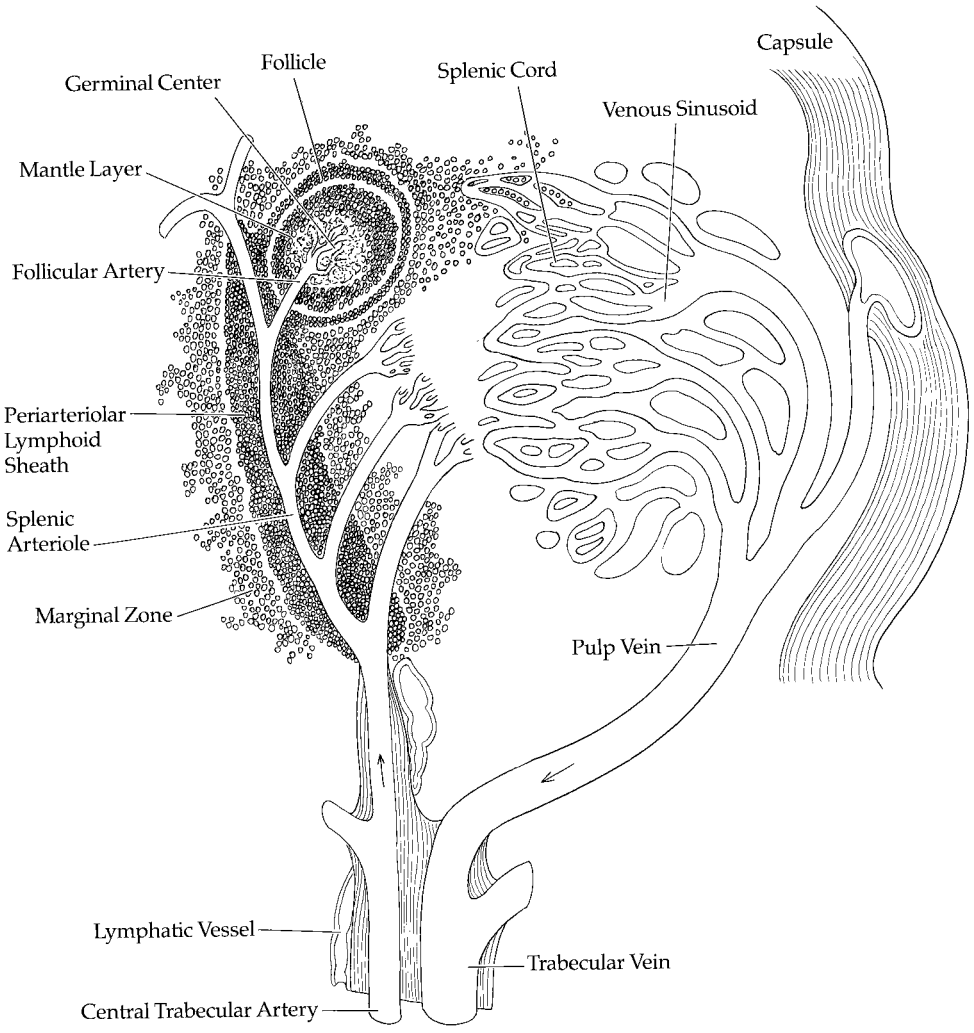


Figure 6 **In the spleen, the lymphoid cells around the splenic arteriole form the periarteriolar lymphoid sheath, which is predominantly a thymus-dependent area. The follicle with a germinal center contains B cells. The T cells are found mainly in the central region of the periarteriolar lymphoid sheaths, whereas the B cells in the germinal centers concentrate more toward the periphery of the sheaths. Lymphocytes enter and leave the periarteriolar lymphoid sheaths via the capillaries of the central arterioles in the marginal zone.**

tonsils, and Peyer patches, to special microenvironments where they encounter antigen, cytokines, and other cells leading to their activation. In contrast, activated effector or memory cells can also traffic to extralymphoid sites in various tissues, such as skin and intestinal lamina propria.[17]

The homing of lymphocytes to the vascular endothelium and their passage through it are controlled by the expression of various receptors on the lymphocyte surface and counterreceptors on the vascular endothelium. To stop the flow of cells in the blood and lymph vessels, initial primary adhesion occurs between lymphocyte receptors on the cells' microvilli, such as L-selectin, and the counterreceptor on the endothelium, such as peripheral lymph node addressin (PNAd). Other cell receptors allow attachment to endothelial E-selectin and P-selectin. Subsequently, the cells can attach and roll using integrin-Ig surface receptors such as α4β7 and α4β1, which bind to endothelial mucosal addressin cell adhesion molecule–1 (MAdCAM-1) and vascular cell adhesion molecule–1 (VCAM-1), respectively. These interactions can lead to stable arrest involving a receptor that triggers adhesion through intracellular signaling by a guanosine triphosphate (GTP) binding protein. Cooperation between receptor interactions is essential because the initial interaction with L-selectin may be too weak to induce the LFA-1/ICAM-1 (leukocyte-function–associated antigen–1/intercellular adhesion molecule–1) stable interaction and therefore requires the α4β7/MAdCAM-1 interaction. In contrast, when tissues display high levels of receptor L-selectin, contact and rolling mediated by L-selectin may be sufficient to allow LFA-1–mediated stable arrest. The lymphocytes can then pass through the vascular endothelium, a process called diapedesis. A similar process is involved with the diapedesis of other leukocytes.[17]

Variations on these mechanisms are thought to channel subsets of lymphocytes to the various microenvironments in the lymph nodes, such as the germinal centers, the paracortical areas, and the T zone, where cells, antigen, and soluble factors lead to particular immunologic responses. These microenvironments are further regulated by various cytokines, such as TGF-β1, which can upregulate integrins and mediate B cell binding to APCs. Cytokines such as TNF-α regulate lymphocyte adhesion receptors, and other cytokines influence lymphocyte activity as they traverse the tissues. Some chemokines, such as RANTES, also activate the expression of adhesion molecules on the surface of effector T cells. Other chemokines, such as thymus-expressed chemokine (TECK), recruit T cells bearing the TECK receptors (CCR9) to their homing organ, in this case the gut.[18,19]

The Hallmarks of the Immune System: Specificity and Memory

The immune responses are controlled by three large gene families: (1) the genes coding for the variable elements of the immunoglobulins, (2) the genes coding for the TCRs, and (3) the genes coding for the MHC antigens [*see 6:II Antigens, Antibodies, and T Cell Receptors*]. In each person, there are an enormous number of genes coding for the variable elements of the immunoglobulins and TCRs, allowing specific recognition of millions of antigens. However, the extreme variability of the MHC applies to the population as a whole; any individual will have only a few variations.

The ability of T cells and antibodies produced by B cells to discriminate between antigens is governed by two independent sets of variable region genes for T and B cells, each composed of rearranging V (variable), D (diversity), and J (joining) DNA segments. Rearrangement of the DNA sequences of these genes occurs as T cells and B cells mature, giving rise to TCRs and immunoglobulins. Additional somatic mutation occurs in B cells during further maturation, which expands the repertoire. This system can discriminate between billions of antigens [*see 6:II Antigens, Antibodies, and T Cell Receptors*].

Each lymphocyte has a surface receptor that recognizes a single antigenic determinant, or epitope. B cell receptors recognize native antigens. After B cells interact with an antigen, they proliferate and differentiate into plasma cells for the production of antibodies, and some become memory cells.

Before being recognized by T cells, an antigen is taken up by an APC (e.g., a dendritic cell or a macrophage), which breaks up the antigen into small peptide fragments. In the APC, certain peptide fragments or epitopes are taken up by MHC class II molecules and carried to the surface of the APC. Thus, TCRs do not recognize native antigen, only processed parts of it [*see 6:II Antigens, Antibodies, and T Cell Receptors*].

The requirement that the antigen be presented in association with an MHC class I or II molecule is referred to as MHC restriction. The nature of the MHC explains why some individuals may not respond to certain antigens. For example, although TCRs recognize epitopes that are bound to MHC molecules, some antigen peptides may not fit into the groove of the particular MHC molecule of an individual. Thus, the appropriate T cell type will not react to that epitope, and the individual will not be able to mount an immune response against it.

DEVELOPMENT OF THE T CELL REPERTOIRE

Of the two types of T cells, αβ and γδ, only the development of the cells that express TCR-αβ is well understood. After gene rearrangement occurs, the TCR-αβ is expressed on the surface of the immature cortical thymocyte together with the CD3 proteins [*see Figure 1*]. At about the same time as gene rearrangement occurs, CD4, CD8, and CD2 are expressed.[6] Single-positive (either $CD4^+$ or $CD8^+$) mature thymocytes bearing high levels of TCR-αβ are selected from the pool of $CD4^+$,$CD8^+$ thymocytes by processes termed positive selection and negative selection.

Positive Selection

Positive selection is controlled by epithelial cells of the thymic cortex and dedicated APCs, such as macrophages, dendritic cells, and interdigitating cells. Many of these stromal cells are located in the corticomedullary junction. Because T cells can react to antigens only in association with self-MHC, only T cells with a TCR that can bind to self-MHC are selected. When these cells react with self-MHC on thymic stromal cells, $CD4^+$,$CD8^+$ TCR-αβ cells that bind to MHC class II molecules become $CD4^+$ and $CD8^-$, downregulating the CD8 and upregulating the TCR-CD3 complex. Conversely, the $CD4^+$,$CD8^+$ cells that bind to MHC class I molecules downregulate the CD4 and become $CD4^-$,$CD8^+$ TCR-αβ cells. In this manner, self-MHC–restricted $CD4^+$ TCR-αβ cells and $CD8^+$ TCR-αβ cells are selected. Only thymocytes whose TCR has a moderate affinity for MHC-peptide are positively selected. CD4 or CD8 cells contribute to the avidity of the interaction as they themselves bind to MHC class II and class I, respectively.

Negative Selection

Most T cells do not interact with self-MHC and therefore undergo apoptosis because of the lack of any interaction. The active process of negative selection eliminates T cells that have a TCR with a high avidity for MHC-peptide. If they were not removed, such T cells could cause serious autoimmune disease. There are

many self-antigens on thymic epithelial cells, and studies show that certain self-antigens can be presented by various APCs in the thymus, such as macrophages and dendritic cells. It is possible that negative selection for some self-antigens occurs when the T cells move to the peripheral lymphoid system after leaving the thymus. This high-affinity interaction between the self-antigen presented by the MHC and the TCR on immature T cells triggers several processes leading to cell death (e.g., apoptosis) (see below).

POSITIVE AND NEGATIVE SELECTION OF B CELLS

Positive selection of precursor B cells takes place in the bone marrow. The pre–B cells bear on their surface the IgM heavy chain (μ) in association with the so-called surrogate light chain (lambda 5 or λ5). If the latter receptor is not on the surface, the cell is eliminated by apoptosis. Pre–B cells develop into immature B cells, which carry IgM on their surface. Those cells develop into mature B cells unless they are eliminated by antigen recognition and negative selection in the lymph nodes. If the affinity is high, elimination involves apoptosis, whereas if the affinity is low, it involves anergy (the cell is present but does not develop).

APOPTOSIS

Removal of autoreactive T and B cells in the periphery also uses apoptosis. Apoptotic signals acting on the cell membrane lead to programmed cell death with nuclear chromatin condensation and cleaving, and within minutes to hours, the cell is destroyed and cleared by macrophages. The main receptors on lymphoid cells that trigger apoptosis are members of the 18-member family of TNF receptors.[20] Examples include the Fas receptor (CD95/APO-1), which is triggered by Fas ligand; and the TNF receptor–1 (TNFR-1), triggered by TNF-α and lymphotoxin-α. Of note, the TNFRs also trigger pathways that lead to activation of nuclear factor κB (NF-κB), which protects against apoptosis.[21] The signals for apoptosis eventually act on a family of cysteine proteases similar to the IL-1b converting enzyme (ICE), the prototype that acts on the cytokine precursor IL-1b, converting it to the active cytokine. ICE-related proteases activate other proteases, which then act on a number of substrates; this, in turn, leads to apoptosis.

AMPLIFICATION AND THE AMNESTIC RESPONSE

Characteristic of the immune response is its ability to increase the number of antigen-specific lymphocytes after an antigenic stimulus occurs; because of memory T and B cells, subsequent exposure to the same antigen results in a faster and a greater response (i.e., the anamnestic response). The basis for this enhanced response is the proliferation of antigen-specific lymphocytes and the production of memory cells after lymphocytes interact with an antigen. These responses are mediated through production of cytokines by lymphocytes and other cells [*see 6:III Immune Response Mechanisms*]. Immune response mechanisms are also amplified through the release of mediator substances from antibody-coated mast cells and basophils, the activation of complement proteins, and the expression of integrin molecules on cells. Altered vascular permeability, the expression of receptors for leukocytes on endothelial cells, and the release of chemotactic factors through these secondary mechanisms attract a host of other cell types to the reaction. These cells greatly contribute to the resulting inflammation by aiding the phagocytic process and the disposal of foreign antigens.

After T cells interact with an antigen, they proliferate and differentiate into effector cells, some helper T cells, and other cytotoxic T cells (CTLs); all of these have a memory cell component. An independent lineage of $CD4^+$ T cells are regulatory T cells (Treg), which control immune responses by producing inhibitory cytokines and by direct cell-to-cell contact.[22,23] Antigen-specific Treg cells may also develop as an independent lineage in the thymus.[24] The memory aspects of Treg cells are not known.

The memory B cells enhance the immune response to previously encountered antigen as part of the anamnestic response. These memory B cells undergo somatic mutation in the variable regions of their immunoglobulin genes. When this somatic mutation occurs in the lymph node germinal centers that contain antigens bound to follicular dendritic cells, it leads to the selection of memory B cells that have high-affinity receptors for the antigens.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Burnet M: The Clonal Selection Theory of Acquired Immunity. Cambridge University Press, London, 1959

2. Mosmann TR, Coffman RL: Heterogeneity of cytokine secretion patterns and functions of helper T cells. Adv Immunol 46:111, 1989

3. Feng Y, Broder CC, Kennedy PE, et al: HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein–coupled receptor. Science 272:872, 1996

4. Choe H, Farzan M, Sun Y, et al: The β-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell 85:1135, 1996

5. Doranz BJ, Rucker J, Yi Y, et al: A dual-tropic primary HIV-1 isolate that uses fusin and the β-chemokine receptors CKR-5, CKR-3, and CKR-2 as fusion cofactors. Cell 85:1149, 1996

6. CD Index. PROW (Protein Reviews on the Web). National Center for Biotechnology Information, 2002
http://www.ncbi.nlm.nih.gov/prow/guide/45277084.htm

7. Qin XF, Schwers S, Yu W, et al: Secondary V(D)J recombination in B-1 cells. Nature 397:355, 1999

8. Natarajan K, Dimasi N, Wang J, et al: Structure and function of natural killer cell receptors: multiple molecular solutions to self, nonself discrimination. Annu Rev Immunol 20:853, 2002

9. Vilches C, Parham P: KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. Annu Rev Immunol 20:217, 2002

10. Janeway CA Jr, Medzhitov R: Innate immune recognition. Annu Rev Immunol 20:197, 2002

11. Mellman I, Steinman RM: Dendritic cells: specialized and regulated antigen processing machines. Cell 106:255, 2001

12. Rissoan MC, Soumelis V, Kadowaki N, et al: Reciprocal control of T helper cell and dendritic cell differentiation. Science 283:1183, 1999

13. Siegal FP, Kadowaki N, Shodell M, et al: The nature of the principal type 1 interferon-producing cells in human blood. Science 284:1835, 1999

14. Liu YJ, Kanzler H, Soumelis V, et al: Dendritic cell lineage, plasticity and cross-regulation. Nat Immunol 2:585, 2001

15. Neutra MR, Mantis NJ, Kraehenbuhl JP: Collaboration of epithelial cells with organized mucosal lymphoid tissues. Nat Immunol 2:1004, 2001

16. Randolph GJ: Is maturation required for Langerhans cell migration? J Exp Med 196:413, 2002

17. Steeber DA, Tedder TF: Adhesion molecule cascades direct lymphocyte recirculation and leukocyte migration during inflammation. Immunol Res 22:299, 2000

18. Rossi D, Zlotnik A: The biology of chemokines and their receptors. Annu Rev Immunol 18:217, 2002

19. Sallusto F, Mackay CR, Lanzavecchia A: The role of chemokine receptors in primary, effector, and memory immune responses. Annu Rev Immunol 18:593, 2000

20. Gupta S: A decision between life and death during TNF-alpha–induced signaling. J Clin Immunol 22:185, 2002

21. Beg AA, Baltimore D: An essential role for NF-κB in preventing TNF-α–induced cell death. Science 274:782, 1996

22. Shimizu J, Yamazaki S, Takahashi T, et al: Stimulation of CD25(+)CD4(+) regulatory T cells through GITR breaks immunological self-tolerance. Nat Immunol 3:135, 2002

23. Sakaguchi S: Regulatory T cells: key controllers of immunologic self-tolerance. Cell 101:455, 2000

24. Jordan MS, Boesteanu A, Reed AJ, et al: Thymic selection of $CD4^+CD25^+$ regulatory T cells induced by an agonist self-peptide. Nat Immunol 2:301, 2001

Acknowledgments

Figures 1, 4, and 5 Seward Hung.

Figure 6 Carol Donner.

II INNATE IMMUNITY

M. Kathryn Liszewski, b.a.
Wayne M. Yokoyama, m.d.
John P. Atkinson, m.d., f.a.c.p.

The body uses two forms of immunity to recognize and respond to infectious microorganisms: innate immunity, which is nonspecific, and adaptive immunity, which is highly specific. The innate immune system is ancient, being present in all multicellular organisms. Adaptive immunity developed from the innate system, beginning about 400 million years ago in primitive fishes such as the lamprey. In vertebrates, the two systems cooperate to produce an immune response. Genetically, the innate system is hardwired and inflexible. In contrast, the adaptive system splices DNA to make new genes that can code for an almost infinite array of immunoglobulin and T cell receptors.

The innate immune system is particularly active at the interface between the environment and those surfaces of the body that are lined with epithelial cells—namely, the skin and the gastrointestinal, genitourinary, and sinopulmonary tracts. Intact physical barriers are, of course, critically important for preventing infections.

The epithelial cells' innate immune response to disturbance by a pathogen has only recently begun to be deciphered. For example, Toll-like receptors—the mammalian counterpart of Toll receptors in fruit flies—were not discovered until 1997. Engagement of these epithelial cell receptors results in the activation of genes that encode inflammatory cytokines, such as tumor necrosis factor (TNF), interferons, and interleukin-1 (IL-1) and IL-12, as well as genes that encode antimicrobial peptides ("Nature's antibiotics").

In addition to the epithelial barrier itself, the fluids in these tracts contain mucus, natural antibodies (IgG and IgA), a complement system, and lectins. Mucus itself is a protective film that traps organisms and debris and also contains antibacterial substances. The complement system in secretions is present at about 10% to 20% of the concentration found in plasma. The lectins in these secretions bind sugars on pathogens and thereby activate the lectin pathway of complement activation. It is probably safe to say that for every major human pathogen, there is a lectin that recognizes a sugar structure on the pathogen's surface. Granulocytes undergo margination in small blood vessels throughout much of these barrier tissues and are available for rapid recruitment to a site of possible infection. Monocytes/macrophages are also present in secretions and in most tissues, where they phagocytose unwanted microbes.

From an evolutionary standpoint, the goal of the immune system is for the individual to survive to reproductive age. This concept suggests that a so-called hyperreactive immune system in early life may provide a survival advantage, even if it subsequently predisposes to allergies or autoimmunity. A second concept is that innate immunity provides guidance to the adaptive immune system regarding which infectious disease antigens it should respond to and what strategy to employ for the most efficient destruction of pathogens. In other words, it tells the adaptive immune system when and how to respond to a pathogen. A third point is that, in contrast to normally sterile areas of the body, such as the bloodstream or spinal fluid, epithelial surfaces are inhabited by large numbers of commensal organisms, against which an immune response would be unnecessary and even potentially harmful. Consequently, at these barriers, the immune system must not only discriminate between self and nonself but, faced with a sea of nonself microorganisms, must single out the dangerous ones. Failure to meet this substantial challenge is not uncommon and results in allergies (reaction to a foreign but benign substance) and autoimmunity (reaction to self).

The innate immune system also accounts for many aspects of the sepsis syndrome. Most house officers vividly remember their first patient with bacteremia or septicemia.[1] Neither textbooks nor attending physicians could provide sound explanations for the signs and symptoms in these patients. Instead, the explanation offered would usually involve "too many bugs in the blood." Of course, microbes in the bloodstream are the primary cause, but the innate immune system's reaction to those organisms is what produces the clinical syndrome. The simplified paradigm is that the bacteria and their products become bound—either directly or through soluble proteins (i.e., lectins, natural antibodies, complement proteins)—to receptors on endothelial and epithelial cells, and this binding signals cytokine and chemokine release (creating a so-called cytokine storm). Most of the clinical signs and symptoms of sepsis result from the effects of excessive quantities of these mediators on vascular endothelium and epithelial cells.

Epithelial Cell Receptors and Their Signaling Pathways

Toll and pattern-recognition receptors

The breakthrough in the field of epithelial cell receptors took place when Janeway and colleagues recognized that the Toll (German slang for "way out" or "crazy") receptor of the fruit fly had intracytoplasmic signaling motifs with homology to those found in mammalian IL-1 and TNF receptors.[2-6] Sequencing of the fruit fly genome led to the discovery of 10 to 20 of these receptors, as well as related soluble binding proteins. These proteins directly recognize key structural components on microbial pathogens—for example, lipopolysaccharides in the outer membrane of gram-negative organisms or peptidoglycans in the cell wall of gram-positive organisms [*see Table 1*]. These structural components are not easily modified in a substantial fashion by the organism. In many cases, soluble binding proteins make first

Table 1 Toll-like Receptors and Their Ligands [4]

Receptor	*Ligands*
TLR2	Lipoproteins, lipoarabinomannan, LPS from *Leptospira* and *Porphyromonas gingivalis*, peptidoglycan from gram-positive bacteria, zymosan from yeast, GPI anchor from *Treponema cruzi*
TLR4	LPS from gram-negative bacteria (except *Leptospira* and *P. gingivalis*), plant products (Taxol), viral products (RSV), host-derived products (HSP60, fibronectin fragments)
TLR5	Flagellin
TLR9	CpG DNA

GPI—glycosylphosphatidylinositol HSP60—heat shock protein 60 LPS—lipopolysaccharide RSV—respiratory syncytial virus TLR—Toll-like receptor

Table 2 Pattern-Recognition Receptors[36]

Pattern-Recognition Receptor (PRR) Type	*PRR*	*Protein/Domain Family*	*Ligands*	*Function*
Secreted PRRs	MBL	C-type lectin	Terminal mannose residents	Activation of the lectin pathway of complement
	CRP, SAP	Pentraxins	Phosphorylcholine on microbial membranes	Opsonization; activation of classical complement pathway
	LBP	Lipid-transfer protein family	LPS	LPS recognition
Cell-surface PRRs	CD14	Leucine-rich repeats, C-type lectin	LPS, peptidoglycan	Coreceptor for TLRs
	Macrophage mannose receptor	C-type lectin	Terminal mannose residues	Phagocytosis
	Macrophage scavenger receptor	Scavenger receptor, cysteine-rich domain	LDL, anionic polymers	Phagocytosis, LPS clearance, and lipid homeostasis
	MARCO	Scavenger receptor, cysteine-rich domain	Bacterial cell walls	Phagocytosis
Intracellular PRRs	PKR	dsRNA-binding domain, protein kinase domain	dsRNA	Activation of NF-κB and MAP kinases; inhibition of translation and induction of apoptosis in virally infected and stressed cells
	Nod	Leucine-rich repeats, nucleotide-binding domain, CARD domain	Ligands for most Nod proteins are unknown; Nodl and Nod2 recognize LPS	Activation of NF-κB and MAP kinases; some family members may be involved in the induction of apoptosis; the exact function is unknown

CARD—caspase-recruitment domain CRP—C-reactive protein dsRNA—double-stranded RNA LBP—lipopolysaccharide-binding protein LDL—low-density lipoprotein LPS—lipopolysaccharide MAP—mitogen-activated protein MARCO—macrophage receptor with collagenous structure MBL—mannan-binding lectin NF-kB—nuclear factor-kB Nod—nucleotide-binding oligomerization domain PKR—dsRNA-activated protein kinase SAP—serum amyloid protein TLRs—Toll-like receptors

contact with the organism, which then facilitates engagement of the receptor. The subsequent cellular signaling involves nuclear factor–κB (NF-κB) and its intracellular signaling cohorts; these cohorts are analogous to those that mammalian cells use to set off an inflammatory response. Once activated, NF-κB travels to the nucleus, where it triggers the transcription of genes that prepare the host cell to do battle with the infecting organism. There is specificity to the response in the sense that the cytokines and the defensins elaborated are tailored to the class of microbe. For example, the cell elaborates different antibacterial defensins for a gram-negative organism than for a gram-positive organism, and it elaborates antifungal peptides distinct from those liberated in an antibacterial response.

Toll-like receptors and pattern recognition receptors (PRRs) [*see Table 2*] were eventually found on human epithelial cells and monocytes/macrophages. Amazingly, this same general scheme (i.e., the Toll-like receptor and its signaling pathway) is used by tomato plants to ward off the mosaic virus.[7]

Innate Immune Cells

NATURAL KILLER CELLS

Natural killer (NK) cells constitute the third major population of lymphocytes, after T cells and B cells.[8] They were initially identified because they spontaneously (i.e., naturally) kill certain tumor cells, a process that does not require prior exposure to the target. Like T cells and B cells, NK cells are involved in host immune defense. They are more closely related to T cells than to B cells in that they share effector functions, including the same killing mechanism and the capacity to produce cytokines. Unlike other lymphocytes, NK cells are components of innate immunity—they respond early against infections (and possibly tumors) and do not require gene rearrangement for maturation and function. Much remains to be clarified regarding the nature of NK cell receptors and their ligands. NK cells are negative for the T cell receptor (CD3) and B cell receptor (membrane immunoglobulin). Most NK cells in human peripheral blood are $CD56^+$; this feature can be used for their identification, because expression of this adhesion-promoting molecule is restricted to NK cells and a small population of T cells. NK cells also express the transmembrane form of the low-affinity receptor for IgG (CD16 or FcγRIIIA) that is absent on mature T cells. When CD16 on the NK cell binds the Fc portion of IgG that is coating a target, this receptor activates release of cytoplasmic granules containing molecules that form pores in the target cell membrane and others that mediate apoptosis, resulting in antibody-dependent cellular cytotoxicity. Natural killing is mediated by the same mechanism, although CD16 is not required. Identification of the receptors that initiate natural killing is a topic of active research.

Of note, NK cells are better able to kill cells that lack major histocompatibility complex (MHC) or human leukocyte antigen (HLA) class I molecules, such as may result from tumorigenesis or viral infection. Although decreased expression of MHC class I molecules may allow targets to evade destruction by MHC class I–restricted cytotoxic T cells, it makes the targets more susceptible to killing by NK cells. This finding led to the so-called missing-self hypothesis, which holds that NK cells survey tissues for

MHC class I molecules, which are normally expressed on most nucleated cells in the body.[9] If MHC class I expression is decreased or absent, the NK cells are released from the negative influence of MHC class I and kill the target. This process may provide a fail-safe mechanism to protect the body from disease processes that evade acquired, specific T cell immunity.

Ongoing studies indicate that NK cells express a multitude of inhibitory receptors that guide their capacity to kill tumor and virus-infected cells.[10] These receptors, termed killer immunoglobulin-like receptors and CD94/NKG2 heterodimers, bind to HLA molecules on their targets. Subsequently, specific tyrosine residues are phosphorylated within immunoreceptor tyrosine-based inhibitory motifs (ITIMs) in the cytoplasmic domains. This results in the recruitment and activation of cytoplasmic phosphatases that dephosphorylate molecules in the activation cascade, hence inhibiting NK cell stimulation. NK cells also express related molecules lacking ITIMs. These molecules are prime candidates for activation receptors that bind as yet uncharacterized molecules on the surface of the target cell.

Early studies of NK cells focused on their role in tumor surveillance (i.e., the eradication of cancers before they become clinically apparent). NK cells can be expanded in tissue culture by exposure to high concentrations of the lymphokine IL-2, which results in production of lymphokine-activated killer (LAK) cells. Adoptive transfer of LAK cells into patients with radiation- and chemotherapy-resistant tumors can lead to remissions.[11] However, high concentrations of IL-2 have to be simultaneously administered, increasing the risk of potentially serious complications such as pulmonary edema and capillary leak syndromes.

Subsequent studies have shown that NK cells also play a critical role in early innate immune responses to viral infections. Persons who lack NK cells suffer recurrent, severe systemic viral infections, particularly from herpesviruses.[12] Depletion of NK cells has also been described in patients with advanced HIV infection and AIDS. This depletion, which apparently results from infection of the NK cell itself by herpesvirus 6 and HIV,[13] may partially account for these patients' susceptibility to opportunistic infections such as those from herpesvirus and cytomegalovirus.

NK cell lymphomas and leukemias are rarely found in Western populations[14] but are more common in Asian populations. NK cell lymphomas often present as nasal tumors, are associated with Epstein-Barr virus infections, and may occur in conjunction with autoimmune syndromes.

Natural killing by peripheral blood NK cells is altered in a variety of conditions. However, the significance of such findings is unclear. On the other hand, as more reagents become available to definitively detect NK cells in clinical specimens, the apparent role of NK cells in disease pathogenesis will become clearer, as suggested by studies indicating marked expansion of NK cells in synovial fluid of patients with early rheumatoid arthritis.[14]

γδ T CELLS AND CD1–RESTRICTED T CELLS

T cells were traditionally thought to respond only to peptide antigens presented in the context of self-MHC. However, CD1-restricted T cells respond to a variety of nonpeptide antigens, especially glycolipids. These cells appear to be more closely related to NK cells than to traditional T cells, because they have a very restricted T cell receptor repertoire and often express molecules usually found on NK cells; hence their designation as NKT cells. NKT cells may have arisen to combat pathogens that feature lipid antigens, such as mycobacteria. Increasing evidence points to populations of T cells as being specifically designed to present complex lipids and related antigens for an immune response.

γδ T cells are found in large numbers in the mucous membranes of many tissues, including the skin, small intestine, female reproductive tract, and lung. These locations suggest that they are a first line of defense against invading microorganisms. These cells appear to mediate cellular immune functions without requiring antigen processing. Thus, these cells are probably important in triggering and controlling the local immune response to pathogens such as *Mycobacterium tuberculosis* and *Listeria monocytogenes*. In so doing, they shape the local inflammatory reaction to the invading pathogen, which includes playing a role in maintaining host tissue integrity. There is still much to learn about how these NKT cell and γδ T cell populations function in the immune response. One hypothesis is that they are the T cell equivalents of natural antibodies and lectins, insofar as they are programmed to provide an innate immune response to certain organisms.

MACROPHAGES

Macrophages are present in body secretions and in tissue sites bordering epithelial tissues. In these locations, they respond to pathogens with phagocytosis and the release of cytokines. They accomplish this by expressing many types of receptors, including PRR, Toll, lectin, scavenger, Fcγ, and complement.[15] In general, these receptors can be divided into two groups. Opsonic receptors recognize microbes coated with immunoglobulins and complement. In contrast, the nonopsonic receptors directly recognize a structure on the surface of the infecting organism. Most of the PRR, Toll, and lectin receptors are nonopsonic. The principles that govern the function of the Toll receptors and PRR (see above) also apply to macrophages, including the signaling pathways and cellular response of cytokine production. Of course, a major difference is that macrophages can phagocytose and kill an organism through oxidative and other microbicidal means not available to epithelial cells. The secreted cytokines, like those from epithelial cells, are designed to activate nearby cells for an inflammatory/immune response. The macrophage lectin receptors recognize sugar moieties (e.g., mannose, galactose, β-glucan, galactose, *N*-acetylglucosamine, and other oligosaccharides). The fine structure of these moieties differs sufficiently from that of host sugars that receptors recognize them but not host sugars. One complement receptor, CR3, like several of the lectin receptors, has multiple specificities: it recognizes C3bi (an opsonic receptor), as well as β-D-glucose, some mannoses, and *N*-acetylglucosamine–containing sugars. In most cases, the recognition of microbes by macrophages leads to phagocytosis and an immune response. However, some organisms use these receptors to gain entrance to intracellular compartments, which can be favorable locations.

DENDRITIC CELLS

The immune system has developed a system for capturing, processing, and then displaying antigens to lymphocytes. Macrophages can phagocytose and present antigens, but they do so much less efficiently than the so-called professional antigen-presenting cells (APCs, or dendritic cells). Most antigens enter via epithelial barriers, where they encounter a cell type bearing long dendritic processes. In the skin, these dendritic cells are called Langerhans cells. The capture of antigens by receptors on APCs (most such receptors are similar to those on macrophages, as described above) occurs by phagocytosis and pinocytosis. The APCs are relatively inefficient (i.e., immature)

in their baseline state, but in an environment rich in inflammatory cytokines from epithelial cells or macrophages (as part of the innate immune response), these immature dendritic cells gradually turn into exceptionally efficient APCs. After antigen capture, the immature or maturing APCs migrate to regional lymph nodes, where they encounter T cells; this encounter initiates an adaptive immune response.

Cytokines of the Innate Immune System

In response to invading microbes, epithelial cells and macrophages secrete cytokines that mediate immune and inflammatory reactions. Cytokines communicate to the cells that produce these signals (autocrine action) and to nearby cells (paracrine action) through cytokine receptors. Through this process, cytokines prepare the local environment to engage microbes and to prepare immunocompetent cells for an immune response. Cytokines commonly released by macrophages and epithelial cells include TNF, IL-1, IL-6, IL-12, IL-15, interferon alfa, and interferon beta. Some of these cytokines of innate immunity also provide second signals for activation of B and T cells [*see 6:III Adaptive Immunity: Antigens, Antibodies, and T Cell and B Cell Receptors*].

The Complement System

The complement system lies at the interface between innate and acquired immunity. As a key component of innate immunity, it promotes the inflammatory response and attacks and destroys foreign substances.[16-19] In this process, it is a facilitator and instructor of the adaptive immune response against the foreign antigen. Further, it provides one of the main effector mechanisms for antibody-mediated host defense [*see Table 3*].

Table 3 Roles of the Complement System in Immunity

First line of defense (innate immunity)
Mediates an inflammatory response
Modifies membranes of invading microbes
Instructor of adaptive immunity
Antigen identification, processing, transportation, and retention
Cellular activation and induction of costimulatory molecules, cytokines, and other mediators of immune response (lowers threshold for B cell signaling)
Effector arm of humoral immunity
Promotes microbial killing via opsonization and lysis
Removes antigens by immune complex formation

From its initial description more than 100 years ago as a heat-labile lytic substance that "complemented" antibodies in destroying bacteria, we now know that the complement system consists of three initiating arms, as well as many regulators and receptors [*see Figure 1*]. The early-reaction sequences behave as biologic cascades in which, by limited proteolysis, one component activates the next. This produces a rapid and robust amplification of the activation process. The primary goal is to attach clusters of C3b on a target (opsonization) so that phagocytes bearing C3 receptors can adhere to and then ingest the material. To counteract the potential for self-injury, almost half of the complement proteins serve as regulators or inhibitors. Despite this tight control, complement is an important contributor to tissue injury in ischemia-reperfusion injury and disease states, especially those featuring autoantibodies and immune complexes. Inherited deficiencies of components in the activating cascade predispose to infectious diseases, primarily of a pyogenic type; surprisingly, such deficiencies also predispose to autoimmunity, especially systemic lupus erythematosus (SLE). Deficiencies of regulators lead to excessive activation and tissue injury. Complement determinations facilitate the diagnosis and management of a number of illnesses. Although no agents for inhibiting complement activation are as yet commercially available, two promising candidates are in clinical trials, and more are in the pipeline.

COMPLEMENT NOMENCLATURE

The early part of the complement system is divided into three branches: the antibody-initiated classical pathway (CP), the antibody-independent (i.e., innate) alternative pathway (AP), and the more recently described lectin pathway (LP). Although each branch is triggered differently, all share the common goal of depositing clusters of C3b on a target. This deposition results in the assembly of a common lytic mechanism, called the membrane attack complex (MAC) or C5b-9.

The CP proteins are identified by numbers (C1, C4, C2, and C3), and the AP proteins are designated by capital letters (factors B, D, and P) [*see Table 4*]. Each MAC consists of one each of C5b, C6, C7, and C8, and many C9s. In the early steps of the cascades, activated proteases cleave the next component in line to liberate a small fragment (these fragments are designated by a lower case "a" [e.g., C3a]), while the large fragment attaches to the target, after which it is designated "b" (e.g., C3b). Limited proteolytic cleavage produces further degradative fragments that are designated by lowercase letters; for example, C3b is cleaved to C3bi, and C3bi is in turn cleaved to C3c and C3dg [*see Figure 2*]. Complement proteins expressed on cell membranes are regulators or receptors for activated components. They may be identified according to so-called cluster of differentiation (CD) designations or according to function. For example, complement receptor 1 (CR1) is also called CD35 and the C3b/C4b or immune adherence receptor.

THE ALTERNATIVE PATHWAY

The AP is an ancient pathway of innate immunity. Unlike the CP, the AP does not require antibody for initiation. Rather, the natural breakdown (i.e., low-grade turnover) of plasma C3 via spontaneous cleavage of a highly reactive thioester bond allows C3 to attach to any nearby host or foreign surface. Regulatory proteins on host cells protect cells by inactivating such fragments. However, foreign membranes usually do not possess such inhibitors, so amplification (the feedback loop of AP) becomes engaged. Target-bound C3b binds to plasma component factor B. The latter undergoes proteolytic cleavage, mediated by plasma protein serine protease factor D, to produce Bb + Ba. The AP C3 convertase thus formed, C3bBb, is stabilized by properdin, which increases the half-life of the enzyme complex. As the stabilized convertase cleaves more C3 to C3b, a feedback loop becomes engaged for autoamplification. Through this mechanism, the AP can deposit several million C3bs on a bacterial surface in a few minutes.[20]

Membrane Attack Complex

As a further assault against a pathogen, the AP assembles the MAC. In this case, the C5 convertase (C3bBbC3bP) cleaves C5 to C5b. This promotes assembly of C6 + C7 + C8 and multiple C9s

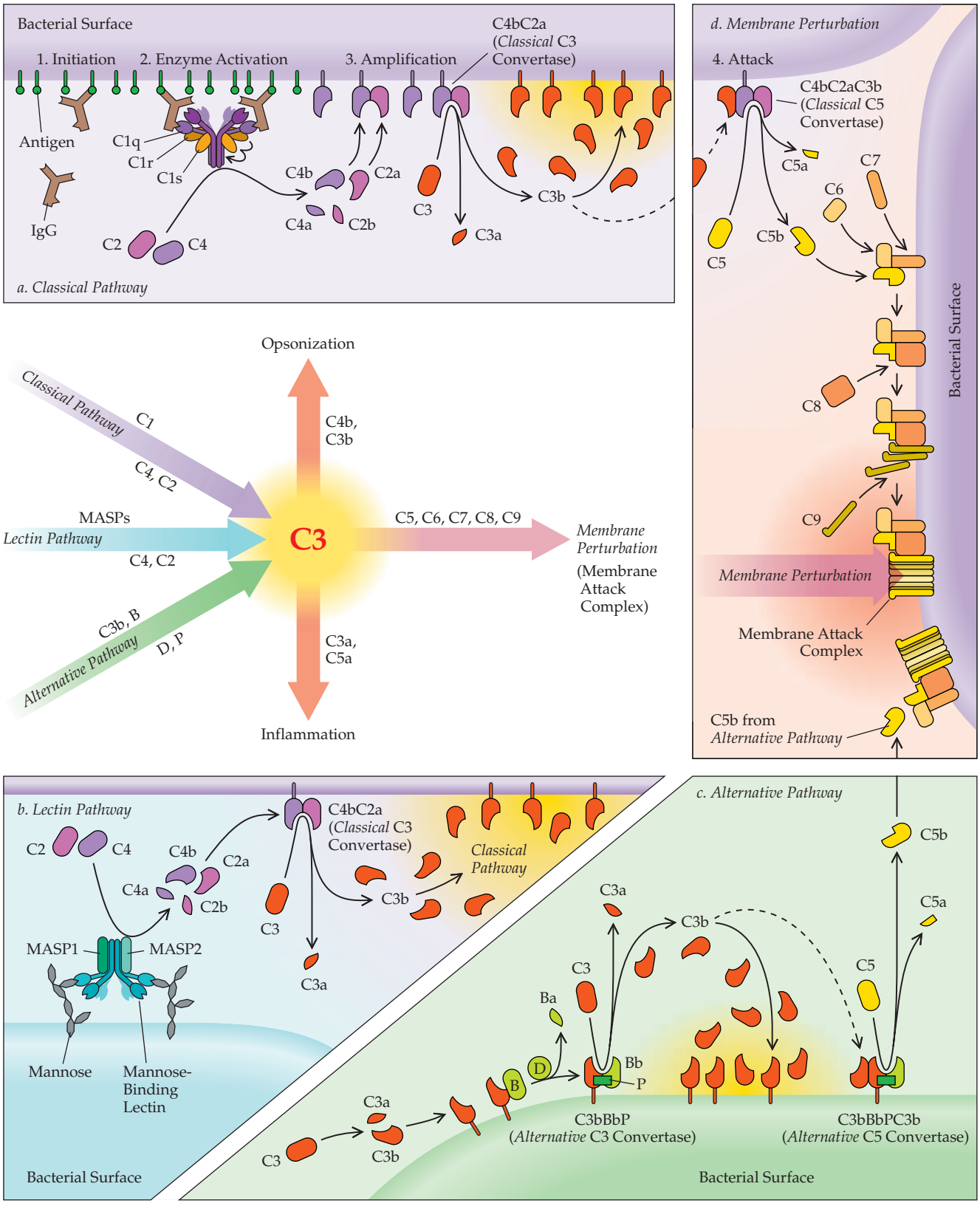


Figure 1 **Function of the complement system. The early part of the complement system comprises three branches: the classical pathway (*a*), the lectin pathway (*b*), and the alternative pathway (*c*). These pathways converge to produce membrane pertubation (*d*) on bacterial surfaces. The most important function of complement is opsonization—the coating of a pathogen with clusters of complement activation fragments. These fragments, in turn, facilitate interactions with complement receptors and in some cases (e.g., certain gram-negative bacteria and viruses) lead to lysis. The second critical function of complement is to activate cells so as to promote inflammatory and immune responses. Through opsonization and cell activation, complement serves as nature's adjuvant to prepare, facilitate, and instruct the host for an adaptive immune response.[34] (MASPs—mannan-binding lectin–associated serine proteases)**

Table 4 Plasma Components of the Complement Cascades

Pathway	*Component*	*Serum Concentration (μg/ml)*	*Function*
Classical pathway (CP)	C1	50	Triggers CP
	C1q	Part of C1 complex	Binds Fc portion of IgG/IgM
	C1r	Part of C1 complex	Protease; cleaves C1s
	C1s	Part of C1 complex	Protease; cleaves C4 and C2
	C4	200–500	Opsonin,* C4a release
	C2	20	Protease; cleaves C3/C5
Alternative pathway (AP)	MBL	150 (wide range)	Binds sugars to trigger LP
	MASP-1	5	Protease
	MASP-2	5	Protease; cleaves C4 and C2
	Factor B	150–250	Protease; cleaves C3/C5
	Factor D	1-2	Protease; cleaves factor B
	Properdin	25	Stabilizes AP convertases
Central protein	C3	550–1,200	Opsonin,* C3a release
Terminal pathway	C5	70	MAC component (C5b), C5a release
	C6	60	MAC component
	C7	60	MAC component
	C8	60	MAC component for pore formation
	C9	60	MAC component for pore formation

*C4b and C3b; also forms part of the C3 and C5 convertases.
LP—lectin pathway MAC—membrane attack complex MASP—MBL-associated serine protease MBL—mannan-binding lectin

to allow perforation (i.e., channel or pore formation) of the foreign membrane. MAC assembly occurs through protein-protein interactions (i.e., no proteolysis is involved after C5 cleavage) to form the lytic complex. Cleavage of C5 also releases the anaphylatoxin C5a, which has potent inflammatory and chemotactic properties.

CLASSICAL PATHWAY

Four concepts underlie the function of the CP: initiation, enzyme activation, amplification, and attack.

Initiation

The CP is triggered by an interaction between the C1q subcomponent [*see Figure 3*] of the C1 complex after it attaches to the Fc portion of antibody (an immune complex). IgM and IgG subclasses 1, 2, and 3 activate the CP, whereas IgA, IgD, IgE, and IgG4 do not. Additionally, C-reactive protein is an acute-phase reactant that activates the CP.[21]

Enzyme Activation

After C1 binds to antibody, the C1r subcomponent undergoes an autoactivation process, activating C1s by cleavage. C4 and C2 are then cleaved by C1s. Large proteolytic fragments derived from C4 (C4b) and C2 (C2a) (an exception to the nomenclature whereby smaller fragments are designated with an "a") attach to the target surface, as well as to antibodies, to assemble the C3 convertase, C4b2a. This enzyme complex rapidly converts C3 to C3b. C4, a structural cousin of C3, also possesses the remarkable post-translational thioester modification. Activated C4b (produced by cleavage of the parent protein) has the transient ability to "glue" covalently onto a nearby surface via a hydroxyl group (i.e., to form an ester linkage).

Amplification

Each activated C1 produces large numbers of C4b and C2a. Most of the C4b serve as opsonins to coat nearby surfaces, whereas others form convertases that rapidly and robustly amplify the system via C3b. Clusters of C3b deposited on a surface serve as ligands for complement receptors.

Attack

Whereas some C3bs serve as ligands for complement receptors, others form C5 convertases (i.e., C4b2a3b). These convertases cleave C5 to initiate the MAC.

LECTIN PATHWAY

Lectins are carbohydrate-binding proteins that were initially identified by their capability of agglutinating red blood cells. They are now recognized as important players in innate immu-

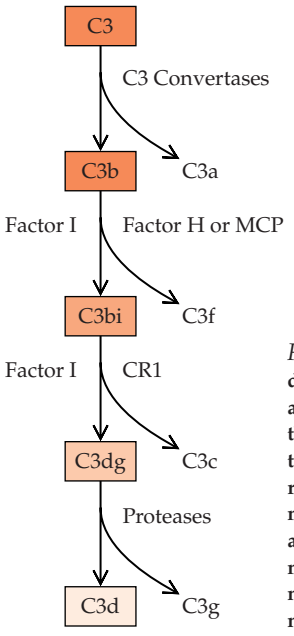


Figure 2 **C3 activation and degradation. The boxed activation fragments are those that are covalently attached to the target. The other fragments are released into the surrounding milieu. C3a is a potent anaphylatoxin. C3f may be a marrow-mobilizing factor for neutrophils. C3c and C3g have no known function.**

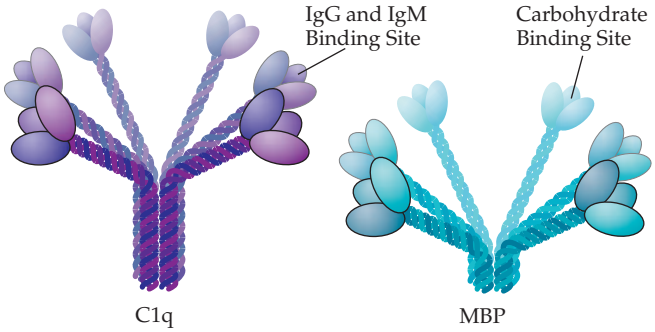


Figure 3 **Similarity of structure between C1q and mannan-binding lectin (MBL, also called mannose-binding protein [MBP]).[19] Ficins, surfactins, and other members of the collectin family of proteins share this general structure. Many activate the lectin pathway upon interaction with their ligands.**

nity. The activation scheme of the lectin pathway is similar to that of the CP, except that lectins substitute for antibodies, and associated proteases replace the C1r/C1s subcomponent of C1. In particular, mannan-binding lectin (MBL, also called mannose-binding protein [MBP]) is a plasma protein that preferentially binds to repeating mannoses and other sugars on pathogens [*see Figure 3*]. The resulting MBL-associated serine proteases (MASPs) cleave C4 and C2. MBL levels are elevated as part of the acute-phase response.[22,23]

RECEPTORS AND REGULATORS OF COMPLEMENT

Many of the effects of the complement system (e.g., immune adherence, phagocytosis, cell signaling) result from the interaction of receptors and activation fragments [*see Tables 5 and 6*].[24] For example, complement receptors mediate clearance of immune complexes bearing complement proteins.[25] Complement receptor 1 (CR1; CD35) on erythrocytes binds C3b/C4b-coated immune complexes for processing and transport to the liver and spleen. In these organs, immune complexes are transferred from erythrocytes to tissue macrophages, and the erythrocytes then return to the circulation. CR1 on granulocytes and monocytes promotes immune complex adherence and phagocytosis, whereas CR1 on B cells, tissue macrophages, and follicular-dendritic cells (FDCs) facilitates trapping and processing of immune complexes in lymphoid organs. Other receptors for C3b fragments help to localize immune complexes to FDC-rich and B cell–rich areas of the spleen and lymph nodes, where they guide the adaptive immune response. Receptors for anaphylatoxins exert vasomodulatory and chemotactic effects after the binding of C3a and C5a. Finally, receptors in the lectin group enhance phagocytosis.

Complement regulatory proteins [*see Table 7*] provide the checks and balances to the system.[24] Unregulated, the system would fire to exhaustion, a point well illustrated by inborn errors of several regulatory proteins. The system is designed so that the initial activation is unencumbered on foreign surfaces, yet rapidly inhibited on self-tissue.

Regulation of the early phase of the CP is provided by C1 inhibitor (C1-Inh), which prevents excessive plasma C1 activation. The C3/C5 convertases are regulated by a family of proteins that include two membrane proteins, decay accelerating factor (DAF; CD55) and membrane cofactor protein (MCP; CD46), as well as the serum inhibitors C4b-binding protein and factor H. These proteins function by disassembling the convertases (decay-accelerating activity), by facilitating proteolytic inactivation of C4b or C3b, or by both processes [*see Figure 4*]. Inactivation occurs in collaboration with the plasma serine protease factor I and is called cofactor activity. The MAC also has regulators in plasma and on cells. Vitronectin (S-protein) binds and inactivates MAC liberated into the fluid phase. On host cells, the glycolipid-anchored CD59 binds C8 and C9 to prevent proper MAC insertion.

COMPLEMENT PARTICIPATION IN ADAPTIVE IMMUNITY

In addition to complement's long-appreciated role as a mediator of antibody-directed events, complement activation is instrumental for generating a humoral immune response.[26] More than 25 years ago, it was found that mice depleted of C3 by cobra venom factor had markedly impaired primary antibody responses. Similar results were associated with deficiencies or defects in C4, C2, C3, CR1, or CR2. The major findings of these studies were variable but often included a low primary IgM response; a failure to class switch; and, on a second antigenic challenge, a lack of recall (memory). The defective response could be overcome with larger doses of antigen.

These results have given rise to two hypotheses, which are not mutually exclusive.[26] One hypothesis suggests that the binding of C3b to an antigen enhances the immune response by promoting transmittal of the antigen to a local lymph node or spleen for uptake, processing, and retention by APCs (especially FDCs). The second hypothesis suggests that the binding of CR2 with C3d complexed on antigen enhances signaling through the B cell antigen receptor. Reconstitution studies are helping to define mechanisms and determine the contributions of CR2 of B cells, relative to the contributions of FDCs. In such studies, B cell deficiency of CD21 (CR2) most closely resembled the knockout, although a contribution of FDCs to long-term memory was apparent.[26]

Table 5 Complement Receptors for C3 and C4

Name (CD Number)	*Primary Ligand*	*Location*	*Function*
CR1 (CD35)	C3b/C4b	Peripheral blood cells (except for platelets), FDC, B cells, podocytes	Immune adherence, phagocytosis, antigen localization
CR2 (CD21)	C3dg/C3d	B cells, FDC	Coreceptor for B cell signaling, antigen localization
CR3 (CD11b/CD18), CR4 (CD11c/CD18)	C3bi	Myeloid lineage	Phagocytosis, adherence

FDC—follicular dendritic cells

Table 6 Complement Receptors for Anaphylatoxins

Receptor	*Ligand*	*Distribution*	*Function*
C3aR	C3a	Myeloid lineage, including mast cells, smooth muscle, epithelial/endothelial cells	Cell activation including granule exocytosis, upregulation of adhesins, chemotaxis, cytoskeletal effects
C5aR	C5a	Similar to C3aR	Similar to C3a but with more chemotactic effects

COMPLEMENT AND DISEASE

The complement system is a double-edged sword. Once unsheathed, it can attack in a robotic fashion. The pathophysiology of many inflammatory diseases involves the synthesis of autoantibodies and the presence of excessive quantities of immune complexes. If the host produces an antibody that reacts to a self-antigen (e.g., on an erythrocyte), the complement cascade becomes activated. Thus, just as complement can destroy a microbe, it may lyse an erythrocyte, opsonize a platelet, or disrupt a basement membrane. If immune complexes lodge in blood vessel walls in a particular tissue, they may activate complement to produce synovitis, vasculitis, dermatitis, or glomerulonephritis. Similarly, a powerful complement barrage may result from ischemia-reperfusion injury as the alternative pathway elicits C3b deposition on the damaged tissue, which is regarded as foreign.

Complement component deficiencies, although rare, predispose to autoimmune diseases (e.g., SLE) and bacterial infections [*see Table 8*].[17,18,27,28] Deficiencies of complement regulatory proteins allow excessive activation of complement cascades [*see Table 7*]. These conditions are usually inherited as autosomal recessive traits, with the exception of deficiencies of C1-Inh (which is an autosomal dominant trait) and properdin (which is X-linked). The effects of these conditions are predictable, because the affected person experiences a loss of function of the deficient protein and of all the proteins that would ordinarily follow in the cascade. Deficiencies of early components (e.g., C1q, C1r/C1s, C4, and C2) predispose to SLE, whereas deficiency of C3, MBL, or MAC components leads to recurrent bacterial infections.

C1-Inh deficiency causes hereditary angioedema, whose symptoms range in severity from a minor inconvenience to life-threatening laryngeal edema [*see 6:XIII Urticaria, Angioedema, and Anaphylaxis*].[29] Deficiency of factor H may lead to uncontrolled AP activation on, and damage to, endothelial cells, resulting in hemolytic uremic syndrome or glomerulonephritis.[30]

Acquired deficiencies of complement also predispose to illness. C1-Inh deficiency may occur as a result of excessive utilization of C1-Inh (usually because of a malignancy) or inactivation of C1-Inh by an autoantibody. An acquired deficiency of C3 may occur as a result of production of an autoantibody that binds and stabilizes the C3 convertase. In the AP, this antibody is called the C3 nephritic factor, whereas in the CP it is termed the C4 nephritic factor. Most patients with C3 nephritic factors are chil-

Table 7 Complement Regulatory Proteins

Regulation Site	*Protein*	*Tissue Distribution*	*Function*	*Disease Associated with Deficiency*
Initiation of complement cascade	C1 inhibitor	Plasma	Inactivates C1r, C1s, and MASPs; a SERPIN	Hereditary angioedema
Convertases	Factor I	Plasma	Cleaves C3b and C4b; requires a cofactor protein	Infection (secondary to low C3 levels)
	Membrane cofactor protein	Most cells	Cofactor for cleavage of C4b and C3b	Hemolytic-uremic syndrome
	Decay-accelerating factor	Most cells	Destabilizes C3 and C5 convertases	None*
	C4b-binding protein	Plasma	Cofactor for cleavage of C4b; decays CP C3 and C5 convertases	None*
	Factor H	Plasma	Cofactor for cleavage of C3b; destabilizes AP C3 and C5 convertases	Hemolytic-uremic syndrome; glomerulonephritis
	Complement receptor type 1 (CR1)	Blood cells	Receptor for C3b and C4b; cofactor activity for C3b and C4b and decays C3 and C5 convertases	None†
MAC	S protein	Plasma	Blocks fluid-phase MAC	Unknown
	CD59	Most cells	Blocks MAC on host cells	Paroxysmal nocturnal hemoglobinuria
Other	Anaphylatoxin inactivator	Plasma	Inactivates C3a, C4a, and C5a	Hives

*Too few cases of complete deficiency described to establish an association.
†A complete deficiency has not been reported.
AP—alternative pathway CP—classical pathway MAC—membrane attack complex MASPs—mannan-binding lectin–associated serine proteases SERPIN—serine protease inhibitor

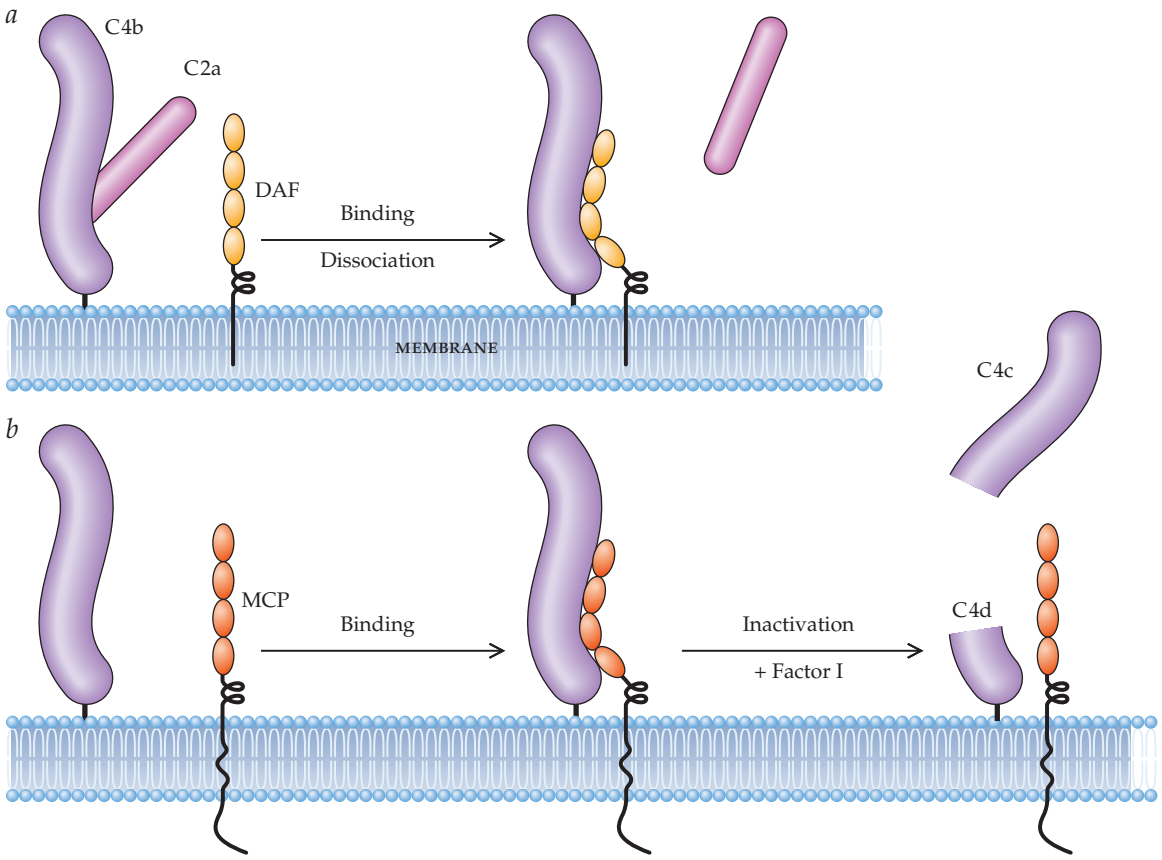


Figure 4 **The membrane proteins decay-accelerating factor (DAF) and membrane cofactor protein (MCP) regulate the C3 and C5 convertases.[35] These proteins function by disassembling the convertases (decay-accelerating activity), by facilitating proteolytic inactivation, or by both processes. In the classical pathway, the components of C3 convertase are the proteases C4b and C2a. Decay-accelerating activity (*a*) occurs when DAF binds C4b, displacing the C2a. Proteolytic inactivation (*b*) occurs when MCP, in concert with the serine protease factor I, cleaves C4b; this prevents C4b from interacting with newly formed C2a. The residual bound C4d has no known biologic activity. In the classical pathway, C5 convertase (which consists of a C4bC2a with an attached C3b) is similarly inactivated by decay-accelerating activity. Although C3 and C5 convertases in the alternative pathway have a different structure, they are disassembled in an identical fashion by DAF and MCP.**

dren; they may present with a combination of glomerulonephritis, partial lipodystrophy, and frequent infections with encapsulated bacteria.

An acquired hematopoietic stem cell disorder produces paroxysmal nocturnal hemoglobinuria [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*]. A mutation in a stem cell prevents expression of an enzyme needed to produce so-called greasy foot (i.e., glycolipid) proteins. As a result, blood cells are deficient in proteins that have this cellular anchor. In particular, deficiencies of CD59 and DAF predispose erythrocytes to complement-mediated hemolytic anemia.

COMPLEMENT MEASUREMENT

Complement levels can be assessed using either antigenic or functional assays [*see Table 9*]. The former are easier to perform and are most commonly employed for measuring C3 and C4 levels. The total hemolytic complement (THC or CH_{50}) assay measures activation of the entire CP by assessing the ability of the test serum to lyse sheep erythrocytes optimally sensitized with rabbit antibody. Interpretation of the results is rather straightforward [*see Table 10*]. Decreased C4 and C3 levels almost always indicate CP activation, whereas AP activation is indicated by normal levels of C4 but low levels of C3 (and factor B, if measured). All nine components of the CP (C1 through C9) are needed to obtain a normal result on the CH_{50} assay. A CH_{50} of 200 means the tested serum lysed 50% of the antibody-coated sheep erythrocytes when assayed at a dilution of 1:200. A similar assay, AH_{50}, measures the total alternative pathway, with the target for lysis being

Table 8 Clinical Manifestations of Complement Deficiency in the Activation Pathways

Pathway Involved	*Deficient Component*	*Clinical Syndrome*
Classical pathway	C1q	SLE, infections*
	C1r/C1s	SLE, infections*
	C4	SLE, infections*
	C2	SLE, infections*
Lectin pathway	MBL	Infections
Central component	C3	Severe infections,* glomerulonephritis, SLE
Membrane perturbation	C5, C6, C7, C8, or C9	*Neisseria* infections
Alternative pathway	Properdin, factor D	*Neisseria* infections

*Typically, with commonly encountered pyogenic organisms.
MBL—mannan-binding lectin SLE—systemic lupus erythematosus

Table 9 Assays for Complement Activation in Human Disease

Method	*Use*	*Comments*
CH_{50} or THC	Screen for a component deficiency or activation of classical pathway	Functional assay; requires appropriate sample handling
AP_{50}	Screen for component deficiency or activation of alternative pathway	Functional assay; requires appropriate sample handling
Antigenic (ELISA, immunodiffusion, nephelometry)	Standard method for C3, C4, factor B, C1-inhibitor, and MBL determinations	Widely available, easy to perform, reliable, inexpensive
Antigenic or hemolytic assay of individual components	To further define a suspected deficiency	Samples usually sent to laboratories specializing in complement assays
Activation fragments C3a, C5a, Bb C1-INH; C1r/C1s C5b-9 (neoantigen)	Additional tests to detect complement turnover	More sensitive than static levels; sample collection technique important; often available through commercial laboratories; expensive
C1-inhibitor function	When clinical picture is consistent with HAE but C1-inhibitor levels by antigenic assay are normal or elevated	15% of HAE patients have normal or elevated levels of a nonfunctional protein
Immunofluorescence	Demonstration of complement activation fragments in tissue	C1q, C4, and C3 are most commonly studied in kidney and skin biopsy specimens
Antiglobulin testing (nongamma Coombs)	Demonstration of C3 fragments on erythrocytes	Usual fragment detected is C3d

AP_{50}—alternative pathway equivalent of THC or CH_{50} ELISA—enzyme-linked immunosorbent assay HAE—hereditary angioedema MBL—mannan-binding lectin THC—total hemolytic assay for classical pathway (also called CH_{50})

unsensitized rabbit red blood cells. Less widely used tests include measurement of the anaphylatoxins C5a and C3a or activation fragments, such as C3d and Bb. Tests showing increased levels of these substances have the advantage of reflecting ongoing activation and are more sensitive. In addition, specialized laboratories can determine the functional and antigenic levels of each of the complement components and regulators.

EVOLVING THERAPEUTIC STRATEGIES

Currently, no therapeutic agent is commercially available to block the deleterious effects of pathologic complement activation.[31-33] Two plasma-based therapies are undergoing clinical trials. A humanized monoclonal antibody to C5 adopts the straightforward strategy of blocking the function of a single component; specifically, it blocks cleavage of C5 by C5 convertases. A significant advantage of this approach is that recombinantly produced humanized monoclonal antibodies are already an established therapy in clinical medicine (e.g., anti-TNF). Recombinant monoclonal antibodies have a relatively long half-life and, if humanized, are usually nonimmunogenic. However, it is unclear how much tissue damage in clinical syndromes results from the activation of C5 as compared with that caused by C3. An advantage of this therapeutic approach is that activation up to and including C3 remains intact.

A second compound in clinical trials is a solubilized version of CR1 (sCR1), which degrades C3b/C4b and decays C3/C5 convertases. Proof of principle has been established in animal models, as it has been for the monoclonal antibody to C5. Poten-

Table 10 Interpretation of Complement Assays

Test Result			*Interpretation*
THC (units/ml)	*C4 (mg/dl)*	*C3 (mg/dl)*	
150–250	16–40	100–180	Normal range
250	40	200	Acute-phase response
100	10	80	CP activation
100	30	50	AP activation
< 10 or 0	30	140	Inherited deficiency or in vitro activation*
50	< 8	100	Partial C4 deficiency or fluid-phase activation†

*In vitro activation is more common than an inherited deficiency state. The lack of activity (< 10 THC) in the setting of normal C4 and C3 antigenic levels suggests (1) an improperly handled sample, (2) cold activation (such as by cryoglobulins) after collection of the sample, or (3) homozygous component deficiency (most commonly C2 with a lupus presentation or, if a *Neisseria* infection is present, of an AP or membrane-attack-complex component).

†Detectable THC excludes a complete deficiency of C4. A partial C4 deficiency, such as of C4A, could give this result. Some types of immune complexes, especially cryoglobulins, and a deficiency of the C1-inhibitor (hereditary angioedema) also give this pattern. In these cases, measurement of C2 is often helpful: a low value suggests activation, whereas a normal value suggests an inherited, partial C4 deficiency. Also, C4A and C4B alleles can be assessed by commercial laboratories.

AP—alternative pathway CP—classical pathway THC—total hemolytic complement (also called CH_{50})

tial barriers to the use of sCR1 in clinical practice include cost, lack of selectivity, the need for intravenous administration, and the potential for infectious complications.

The search for a complement therapeutic has led to two unanticipated observations. First, complement activation contributes to tissue damage in ischemia-reperfusion injury syndromes such as myocardial infarctions and stroke; second, C5b–C9 and C5a bear responsibility for more of the tissue damage during complement activation than predicted.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Cross AS, Opal SM: A new paradigm for the treatment of sepsis: is it time to consider combination therapy? Ann Intern Med 138:502, 2003

2. Janeway CAJ: Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb Symp Quant Biol 54:1, 1989

3. Medzhitov R, Preston-Hurlburt P, Janeway CAJ: A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature 388:394, 1997

4. Akira S, Takeda K, Kaisho T: Toll-like receptors: critical proteins linking innate and acquired immunity. Nat Immunol 2:675, 2001

5. Revillard J-P: Innate immunity. Eur J Dermatol 12:224, 2002

6. Bendelac A, Fearon DT: Innate immunity: receptors and effectors of innate immunity. Curr Opin Immunol 12:11, 2000

7. Ellis J, Dodds P, Pryor T: Structure, function and evolution of plant disease resistance genes. Curr Opin Plant Biol 3:278, 2000

8. Yokoyama W: Natural killer cells. Fundamental Immunology. Paul WE, Ed. Lippincott-Raven, New York, 1999, p 575

9. Ljunggren HG, Karre K: In search of the 'missing self': MHC molecules and NK cell recognition. Immunol Today 11:237, 1990

10. Rosenberg SA, Lotze MT, Muul LM, et al: A progress report on the treatment of 157 patients with advanced cancer using lymphokine-activated killer cells and interleukin-2 or high-dose interleukin-2 alone. N Engl J Med 316:889, 1987

11. Biron CA, Byron KS, Sullivan JL: Severe herpesvirus infections in an adolescent without natural killer cells (comment). N Engl J Med 320:1731, 1989

12. Lusso P, Malnati MS, Garzino-Demo A, et al: Infection of natural killer cells by human herpesvirus 6. Nature 362:458, 1993

13. Jaffe ES, Lindskog S, Ahren B: Report of the workshop on nasal and related extranodal angiocentric T/natural killer cell lymphomas: definitions, differential diagnosis, and epidemiology. Am J Surg Pathol 20:103, 1996

14. Tak PP, Kummer JA, Hack CE, et al: Granzyme-positive cytotoxic cells are specifically increased in early rheumatoid synovial tissue. Arthritis Rheum 37:1735, 1994

15. Linehan SA, Martinez-Pomares L, Gordon S: Macrophage lectins in host defence. Microbes Infect 2:279, 2000

16. Volanakis JE: Overview of the complement system. The Human Complement System in Health and Disease, 10th ed. Volanakis JE, Frank MM, Eds. Marcel Dekker, New York, 1998, p 97

17. Walport MJ: Complement. First of two parts. N Engl J Med 344:1058, 2001

18. Walport MJ: Complement. Second of two parts. N Engl J Med 344:1140, 2001

19. Frank MM, Atkinson JP: Complement system. Samter's Immunologic Diseases, 6th ed. Austen KF, Frank MM, Atkinson JP, et al, Eds. Lippincott Williams & Wilkins, Baltimore, 2001, p 281

20. Ollert MW, Kadlec JV, David K, et al: Antibody-mediated complement activation on nucleated cells: a quantitative analysis of the individual reaction steps. J Immunol 153:2213, 1994

21. DuClos TW: Function of C-reactive protein. Ann Med 32:274, 2000

22. Turner MW: Mannose-binding lectin: the pluripotent molecule of the innate immune system. Immunol Today 17:532, 1996

23. Fujita T: Evolution of the lectin-complement pathway and its role in innate immunity. Nat Rev Immunol 2:346, 2002

24. Liszewski MK, Farries TC, Lublin DM, et al: Control of the complement system. Adv Immunol 61:201, 1996

25. Hebert LA: The clearance of immune complexes from the circulation of man and other primates. Am J Kidney Dis 17:353, 1991

26. Nielsen CH, Leslie RG: Complement's participation in acquired immunity. J Leukoc Biol 72:249, 2002

27. Barilla-LaBarca ML, Atkinson JP: Rheumatic syndromes associated with complement deficiency. Curr Opin Rheumatol 15:55, 2003

28. Pickering MC, Botto M, Taylor PR, et al: Systemic lupus erythematosus, complement deficiency, and apoptosis. Adv Immunol 76:227, 2000

29. Davis AE III: C1 inhibitor gene and hereditary angioedema. The Human Complement System in Health and Disease. Volanakis JE, Frank MM, Eds. Marcel Dekker, New York, 1998, p 455

30. Zipfel PF: Hemolytic uremic syndrome: how do factor H mutants mediate endothelial damage? Trends Immunol 22:345, 2001

31. Liszewski MK, Atkinson JP: Complement inhibition. Biological Therapy in Rheumatology. Smolen J, Lipsky P, Eds. Martin Dunitz, London, 2002, p 453

32. Kirshfink M: Targeting complement in therapy. Immunol Rev 180:177, 2001

33. Klickstein LB, Moore FD Jr, Atkinson JP: Therapeutic inhibition of complement activation with emphasis on drugs in clinical trials. Therapeutic Immunology. Austen KF, Burakoff SJ, Strom TB, et al, Eds. Blackwell Science, Malden, MA, 2000, p 287

34. Liszewski MK, Atkinson JP: The complement system. Fundamental Immunology, 3rd ed. Paul WE, Ed. Raven Press, New York, 1993, p 917

35. Liszewski MK, Farries TC, Lublin DM, et al: Control of the complement system. Adv Immunol 61:201, 1996

36. Medzhitov R: Toll-like receptors and innate immunity. Nat Rev Immunol 1:135, 2001

Acknowledgment

Figures 1 through 4 Seward Hung.

III ADAPTIVE IMMUNITY: ANTIGENS, ANTIBODIES, AND T CELL AND B CELL RECEPTORS

Cox Terhorst, ph.d.
John R. David, m.d.

Antigens

An antigen is any substance capable of generating an immune response; that is, antigens react with T cells and B cells to induce the formation of antibodies and to sensitize lymphocytes, and those antibodies and sensitized cells then react with the antigens. The first antigens to be studied were microorganisms and foreign proteins, and it remains true that foreign proteins are almost universally antigenic. In addition, polysaccharides can induce antibody formation when coupled to proteins, and certain purified polysaccharides are themselves effective antigens. One such example is purified pneumococcal polysaccharide, which can be used as a vaccine against the particular strain of pneumococcus from which the polysaccharide is obtained. Although most antigens are macromolecules, some small molecules can also be antigenic. Antibodies to DNA or RNA occur in many patients with rheumatic diseases, especially systemic lupus erythematosus.

antigen recognition

Antigens are recognized not only by antibodies but also by antigen-specific B cell receptors (BCRs) [*see* B cell receptors, *below*] and T cell receptors (TCRs) [*see* T Cell Receptors, *below*], which are located on the surfaces of B cells and T cells, respectively. In general, TCRs and antibodies recognize different antigenic determinants.

Two distinct types of TCRs exist: TCR-αβ and TCR-γδ. Each has different mechanisms for recognizing antigens [*see 6:IV Cell-Cell Interactions, Cytokines, and Chemokines in Immune Response Mechanisms*]. For example, T cells bearing TCR-αβ recognize antigens that have been processed by antigen-presenting cells (APCs) to become peptide fragments bound to major histocompatibility complex (MHC) class I or class II molecules on the surface of the APCs. In contrast, T cells bearing TCR-γδ appear not to require antigen presentation by MHC molecules. Helper T cells recognize only peptide fragments bound to MHC class II molecules, whereas cytotoxic T cells recognize processed viral antigens presented by both MHC class I and MHC class II molecules on the surface of virus-infected cells. Pure lipids from mycobacteria can be presented as antigen to TCR-αβ by CD1 molecules rather than by MHC class I or class II molecules.

adjuvants

Adjuvants are substances capable of increasing the immunogenicity of antigens and are critical in the production of vaccines. Many microbial products have been used as adjuvants,[1] including substances from *Mycobacterium tuberculosis,* bacillus Calmette-Guérin (BCG), *Corynebacterium parvum, Brucella abortus,* and *Bordetella pertussis,* as well as toxoids from *Vibrio cholerae* and *Clostridium tetani.* Adjuvants have also been derived from vaccinia virus and other poxviruses, BCG, and *Salmonella* by transfecting the organism with genes for an antigen of interest. Others have come from lipopolysaccharide derivatives, such as monophosphoryl lipid A. Freund complete adjuvant, which consists of dead mycobacteria in oil emulsified with an antigen in aqueous solution, is now infrequently used because it generally causes a strong local inflammatory reaction. Other adjuvants are extracts from the soap bark tree (*Quillaja saponaria*), polymers (e.g., inulin), peptide complexes, and a number of cytokines. The aluminum salt alum is approved for general use as an adjuvant in humans. The complex of antigen and complement fragments, particularly those derived from C3, probably serves as the physiologic adjuvant.

Antibodies

Antibodies are a heterogeneous group of serum proteins called immunoglobulins. Immunoglobulins are secreted by differentiated B cells called plasma cells. According to Sir Francis MacFarlane Burnett's clonal selection theory, a single plasma cell produces only one specific antibody. This theory is aptly illustrated by the disease multiple myeloma, in which malignant proliferation of plasma cells occurs in bone marrow. The multiple myeloma plasma cell produces an abnormal monoclonal immunoglobulin called a myeloma protein. Its homogeneity is clearly visible on electrophoresis as a single dense band. Certain antibodies that are produced in response to highly homogeneous antigens, such as streptococcal polysaccharide, may also be relatively homogeneous.

Immunoglobulin molecules consist of two identical heavy chains and two identical light chains [*see Figure 1a*]. Each light chain is attached to a heavy chain by disulfide (S—S) bonds, and the heavy chains are also attached to each other by one or more S—S bridges. Amino acid sequences in both heavy and light chains are divided into regions that are either constant or variable [*see Figure 1b*]; in addition, each variable region contains sequences that are hypervariable.

classification of immunoglobulins

There are five classes of immunoglobulins, and each class contains a specific heavy chain: IgG contains two γ chains; IgM, two μ chains; IgA, two α chains; IgD, two δ chains; and IgE, two ε chains. There are also two types of light chains, κ and δ, which can be differentiated antigenically. One form of IgM, the secreted form, is a pentamer. The monomeric form of IgM is expressed on the extracellular surface of B cells. IgA exists as a monomer or a dimer. The polymeric forms of IgM and IgA have an additional J (joining) chain that facilitates polymerization.

IgG IgG is the major immunoglobulin in the serum, where it exists as a monomer. IgG has a half-life in the blood of approximately 23 days. It is the main antibody raised after antigenic challenge. There are four subclasses of IgG—IgG1, IgG2, IgG3, and IgG4—each different in structure and biologic properties. For example, only IgG1 and IgG3 bind the first component of

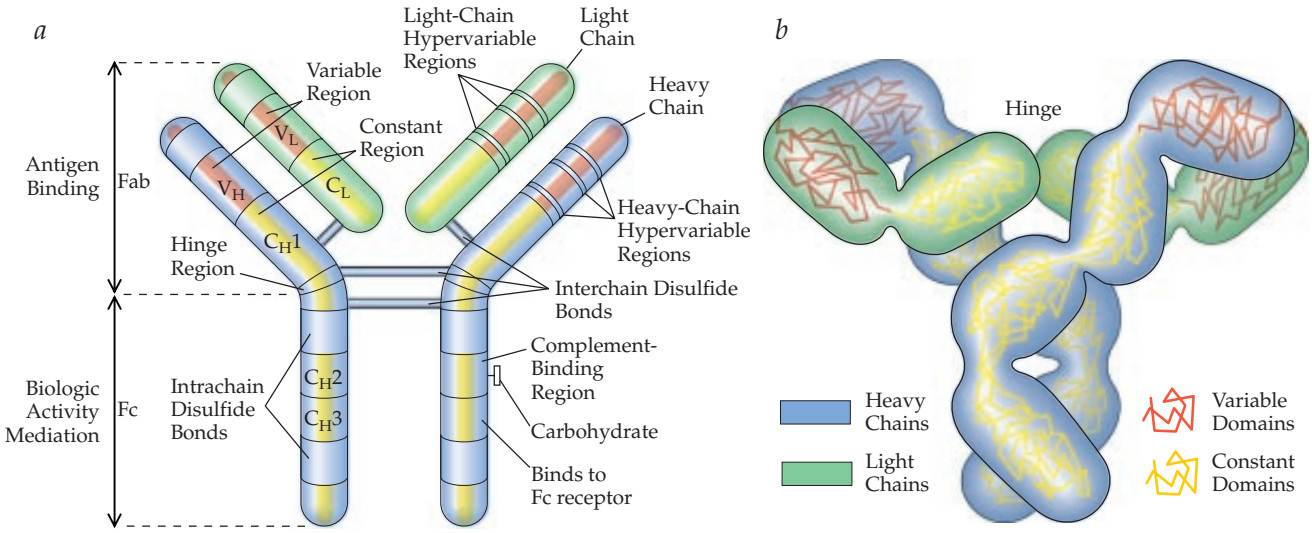


Figure 1 **(*a*) The immunoglobulin molecule is a Y-shaped protein made up of four polypeptide chains. Two heavy chains (blue) are joined to two light chains (green) by disulfide bonds. Blue squares represent intrachain S—S bonds; blue bars indicate interchain S—S bonds. The heavy chains extend from the stem of the Y into the arm; the two light chains are confined to the arms. Each polypeptide has regions whose amino acid sequences are constant (white and yellow) and variable (red). The variable regions also contain hypervariable regions. All antibodies of a given type have the same constant regions, but the variable regions differ from one clone of a B cell to another. The heavy- and light-chain variable regions fold to create an antigen-binding site. (*b*) Schematic model of the domain structure of an antibody molecule. The domains have a characteristic folding pattern, which is also seen in the T cell receptor and proteins of the major histocompatibility complex.[17]**

complement and adhere to monocytes. The antibodies that coat microorganisms and render them more susceptible to phagocytosis (i.e., opsonization) are of the IgG class. IgG antibodies can also neutralize viruses and toxins such as diphtheria toxin. Human antibodies to polysaccharides are mainly of the IgG class, but lesser amounts of IgM and IgA are also produced. Although the fetus does not produce this class of immunoglobulin, IgG readily crosses the placenta; therefore, IgG antibodies found in the newborn are from the mother.

Clinically, IgG has been used successfully for reconstituting the immunity of patients with primary immune deficiencies, such as agammaglobulinemia, and for preventing hemolytic disease in the newborn. Women with Rh-negative blood who bear a fetus with Rh1-positive red blood cells are sensitized at the first delivery by Rh1-positive red blood cells from the fetus. The mother then produces anti-Rh1 IgG antibodies that will cross the placenta during subsequent pregnancies; these antibodies react with fetal red blood cells, causing hemolytic disease. Erythroblastosis fetalis can be prevented by injecting IgG rich in anti–Rh-positive antibodies (RhoGAM) into an Rh1-negative mother at the time of delivery or abortion. These antibodies presumably combine with any fetal Rh1-positive red blood cells present and prevent them from immunizing the mother.

IgG can be split into three fragments by the proteolytic enzyme papain. Two of the fragments are similar and are called Fab; the third is called Fc [*see Figure 1a*]. The Fc portion is responsible for the biologic activity of the various immunoglobulins; among other things, the Fc portion controls the ability of immunoglobulins to bind to cells, fix complement, and traverse the placenta. Another proteolytic enzyme, pepsin, splits the IgG molecule behind the S—S bonds that bridge the heavy chains, leaving one large fragment, $F(ab')_2$, which is able to bind and precipitate antigen because of its bivalency and capacity to form a lattice.

IgA IgA is the predominant immunoglobulin in secretions, where it is usually found as a dimer and is released as such by local plasma cells. Monomeric IgA constitutes 15% of the serum immunoglobulins. In the serum, it has a half-life of 5 to 6 days. There are two subclasses of IgA: IgA1 and IgA2. The IgA dimer combines with the secretory piece, which is a polypeptide chain produced by local epithelial cells. In this form, IgA is quite resistant to proteolytic digestion. Unlike serum IgA, IgA combined with the secretory piece can undergo active transport across the mucosal epithelium by endocytosis [*see Figure 2*].

IgA is present in saliva, tears, and colostrum. It also occurs in the respiratory and gastrointestinal tracts, in the vagina, and in the prostate. The increased levels of antibodies to dietary antigens that are found in persons with IgA deficiency suggest that the IgA class of immunoglobulin normally limits the absorption of such antigens.

IgA may play an important role in local immunity by neutralizing viruses and by combining with viruses and bacteria, thereby preventing their adherence to mucosal surfaces. Although IgA does not bind to the first component of complement, as IgG and IgM do, it can lead to activation of the alternative complement pathway [*see 6:II Innate Immunity*]. One of the complement components generated by this pathway, C3b, aids in the opsonization of bacteria, enhancing their uptake and killing by phagocytes.

IgM Most B cells have monomeric IgM on their surface. However, IgM exists primarily as a pentamer and is found mainly in the serum, where it makes up 10% of the immunoglobulins. In the immune response, IgM is the first immunoglobulin raised in response to antigen stimulation. Cells that produce IgM or their precursors do not become memory cells, so that a second challenge with antigen produces no more IgM antibody than the first stimulus. Because the IgM response to antigen is short-lived, the presence of specific IgM in the serum may be helpful in establishing the diagnosis of acute infection with a particular pathogen. The fetus makes IgM anti-

bodies to certain microorganisms, which can be helpful in the diagnosis of fetal toxoplasmosis, rubella, or syphilis. Not all fetuses infected by these organisms produce such antibodies, however.

As a pentamer, IgM is highly efficient at fixing complement. Molecule for molecule, IgM is 20 times as effective as IgG in agglutinating bacteria and red blood cells and 1,000 times as active in bactericidal reactions. Isohemagglutinins, such as anti-A and anti-B, are of the IgM class. Waldenström macroglobulinemia is a disease characterized by the monoclonal production of IgM.

IgD IgD, which is a monomer, occurs in the serum in trace amounts. It is found in relatively high concentrations in umbilical cord blood. Most of the B cells of umbilical cord blood have IgD on their surface, and most B cells in adults have both IgM and IgD on their surface. Plasma cells that produce IgD have been found in the tonsils and adenoids, although they are very rare in other lymphoid tissues. The function of IgD is unknown.

IgE IgE is present in trace amounts in the serum, constituting only 0.004% of the immunoglobulins. Plasma cells that produce IgE are found in the tonsils and adenoids and on the mucosa of the respiratory and GI tracts.

Distinct receptors for IgE are found on the surface of mast cells, B cells, T cells, macrophages, and eosinophils. IgE binds to its receptor on these cells by its Fc portion; heating the antibody destroys its cell-binding ability. Formerly referred to as reagin, IgE plays a primary role in immediate hypersensitivity—namely, the immune reaction in hay fever, extrinsic asthma, wheal-and-flare reactions, and anaphylaxis. IgE binds tightly to mast cells and basophils. When these IgE-coated cells interact with specific antigens, termed allergens, they release potent mediators of immediate hypersensitivity, including histamine, slow-reacting substance of anaphylaxis (SRS-A), and an eosinophilic chemotactic factor [*see 6:X Allergic Response*]. Levels of IgE are higher than normal in persons with atopic dermatitis, as are the levels of IgE antibody specific for allergens to which the individual is susceptible. In patients with allergies, specific IgE antibodies are detected by means of a radioimmunoassay called the radioallergosorbent test (RAST).

The exact function of IgE is unknown. Certainly, the manifestations of immediate hypersensitivity, such as hay fever and extrinsic asthma, do not appear to serve any useful purpose for the person affected or for the species in general. Therefore, the observation that IgE levels are sometimes elevated in persons living in the tropics, and especially in those afflicted with helminthic parasites, was greeted by immunologists as a possible indication that IgE plays a protective role against parasites. The mediators released could affect the parasite either directly or by producing an increase in vascular permeability and the release of eosinophilic chemotactic factor, which may lead to the accumulation of other necessary antibodies (e.g., IgG) and cells to attack the parasite. In this context, it is of interest that eosinophils can mediate IgG-dependent damage to schistosomula (the larval form of the parasite *Schistosoma mansoni*). In addition, parasite-specific IgE immune complexes can induce a macrophage-mediated cytotoxic response to schistosomulum organisms.

Antigenic Differences of Immunoglobulins

Immunoglobulins have three types of serologic, or antigenic, determinants: isotypic, allotypic, and idiotypic.

Isotypic determinants Isotypic determinants distinguish between the constant regions of the various classes and subclasses of heavy chains and light chains; they represent the different constant-region genes. For example, there are four IgG heavy-chain isotypes: γ_1, γ_2, γ_3, and γ_4, representing the subclasses IgG1, IgG2, IgG3, and IgG4, respectively. There is only one κ light-chain isotype and one λ light-chain isotype.

Allotypic determinants Allotypic determinants distinguish between immunoglobulins of a particular isotype; they represent different alleles of immunoglobulin genes and therefore are genetically determined according to mendelian laws in a man-

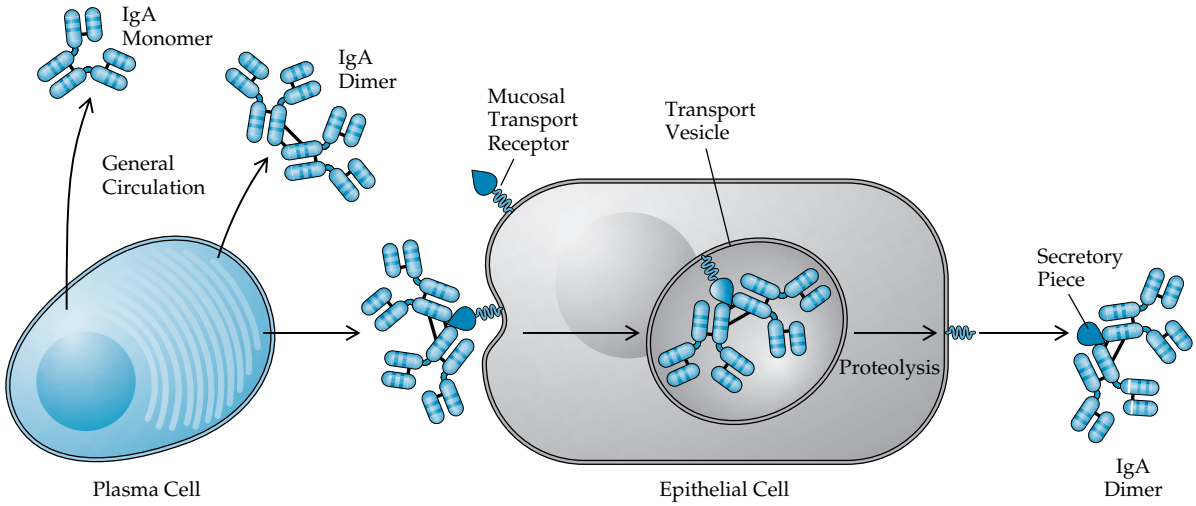


Figure 2 **Plasma cells secrete IgA molecules into the general circulation as either monomers or dimers. The circulating dimer can combine with a mucosal transport receptor on the surface of an epithelial cell. As the antibody-receptor complex is transported through the cell, the receptor is cleaved. The portion of the receptor that remains attached to the antibody dimer is called the secretory piece. The secretory piece is joined to the constant region of IgA by disulfide bonds. The mucosal transport receptor contains five immunoglobulin-like domains and is anchored in the membrane by a proteolytically labile segment.[18]**

ner similar to the way that the ABO blood groups are determined. The γ heavy chains have more than 20 different allotypic markers, which are collectively termed Gm. In addition, κ light chains contain a set of at least three allotypic markers, collectively called Km. There are no known allotypic markers for the μ, δ, and ε heavy chains or for the λ light chain.

Idiotypic determinants An idiotope is defined as a single antigenic determinant on the hypervariable region of an antibody. An idiotype is the antigenic character of the variable region of an antibody. Idiotypic determinants distinguish one immunoglobulin from another of the same allotype.

Genetic Source of Antibody Diversity

The carboxyl-terminal halves of all κ light chains have almost identical amino acid sequences; this portion of the molecule is therefore called the constant, or C, domain. The amino-terminal half has a variable sequence of amino acids and is known as the variable, or V, domain [*see Figure 1*]. The first 110 amino acids of the amino-terminal portion of the λ light chain and the heavy chain are also variable. The remaining 75% of the heavy chain is constant and contains three homologous regions.

Within the variable regions, three areas—referred to as the hypervariable, or complementarity-determining, regions—show great variation; these areas correspond to the antigen-binding site of the antibody. X-ray analysis has shown that immunoglobulin molecules are built up from compact globular units connected by short segments of more or less linear polypeptide chains [*see Figure 1*]. As expected, the hypervariable regions are located at the interface between immunoglobulin and antigen.

The most intriguing aspect of the genetic control of immunoglobulin synthesis is the diversity of the product: plasma cells can make antibodies that react with an indefinite number of different antigenic sites. How can DNA code for such a large number of antigens, many of which have only recently (on the evolutionary time scale) come into existence?

VDJ RECOMBINATION

In all cells, DNA for the κ light chain codes for more than 300 variable (V) regions, five joining (J) regions, and one constant (C) region. The V and J regions are separated from the C region by an intervening stretch of DNA. Thus, in the so-called germline configuration, DNA encodes the information for at least 1,500 different combinations of V and J regions; in other words, at least 1,500 different κ light chains are possible.

The emergence of individual plasma cell lines is the result of somatic recombination in the DNA and RNA splicing [*see Figure 3*]. As the pre–B cell differentiates into a plasma cell, rearrangements and deletions in the DNA bring one of the V genes, chosen at random, adjacent to one of the J genes. This V-J unit and the remaining J regions are separated from the C gene by a short length of DNA. In the next step, the DNA is transcribed to nuclear RNA, and the stage is set for a second event.

This event begins when an enzyme cleaves the nuclear RNA to produce messenger RNA (mRNA). In a process called RNA splicing, the segment that separates the V-J unit from the C region is removed, along with any superfluous J segments. The remaining V-J-C segment is now translated into one of the 1,500 κ light-chain proteins. Actually, the number of possible proteins is higher because the joining of any V region to a J region can involve one of a variety of base pairs at the recombination site.

Variability in the heavy chain makes an important contribution to the specificity of an antibody and also results from somatic recombination and RNA splicing. The germline configuration of the DNA carries instructions for several hundred different heavy-chain V genes, six J genes, 10 to 20 diversity (D) genes, and nine C genes (these C genes code for the heavy-chain classes: IgM, IgD, IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, and IgE). DNA deletion, transcription to nuclear RNA, and RNA splicing produce the final V-D-J-C sequence in the mRNA that is translated by ribosomes to a heavy-chain protein. This assembly process produces more than 18,000 possible varieties of heavy-chain proteins (antibody specificity does not vary with class, so the nine C genes do not enter into the calculation).

The combination of more than 1,500 light-chain varieties with the 18,000 heavy-chain varieties yields more than 27 million different kinds of antibodies with different antigen-binding sites. In addition, somatic hypermutation occurs, particularly during affinity maturation, and the rate of somatic hypermutation is relatively high (one base pair per 1,000 cell divisions). Therefore, the potential number of specific antibodies in a single person is indefinite.

Although some of the mechanisms of V-D-J rearrangement are unique to B cells (and T cells, as synthesis of TCRs occurs by a similar mechanism [*see* T Cell Receptors, *below*]), the general mechanisms of DNA repair are also engaged.[2] Two genes involved in V-D-J rearrangement in B cells are the recombination-activating genes *rag1* and *rag2*.[3-5] Disruption of the *rag1* or *rag2* gene causes a block in B cell development before the transition from pro–B cell to pre–B cell, as found in patients with severe combined immunodeficiency syndrome or Omenn syndrome.[6] Disruption of the surface IgM gene, the J region of the heavy chain, or the J region of the κ light chain leads to a similar block in B cell development.

ALLELIC EXCLUSION

In an individual B cell, only one of the chromosomes undergoes complete V-D-J recombination leading to expression of heavy and light chains. A mechanism exists that prevents the other chromosome from being rearranged and therefore expressed in the same cell. This is called allelic exclusion. It prevents a B cell from expressing two entirely different immunoglobulins or BCRs. A similar mechanism operates during synthesis of TCRs in T cells [*see* T Cell Receptors, *below*].

Primary and Secondary Antibody Responses

When antigen is first introduced into the body, a primary response occurs that is characterized by a lag phase that lasts several days, during which no antibody is detected. Increasing amounts of IgM antibody appear in the serum, usually reaching a peak level after 7 days. After 6 to 7 days, IgG antibody is also detected. The IgM titer begins to wane before the maximal IgG titer is reached, about 10 to 14 days after the antigen is introduced. Antibody titers then decrease, and very little antibody is detected 4 to 5 weeks after a single dose of antigen.

If the antigen is encountered a second time, a secondary response (also called an anamnestic or booster response) occurs because of the existence of memory B cells. Both IgM and IgG titers rise exponentially, without the lag phase seen in the primary response. Whereas the peak IgM level during the secondary response may be the same as, or slightly higher than, the peak IgM level during the primary response, the IgG peak level dur-

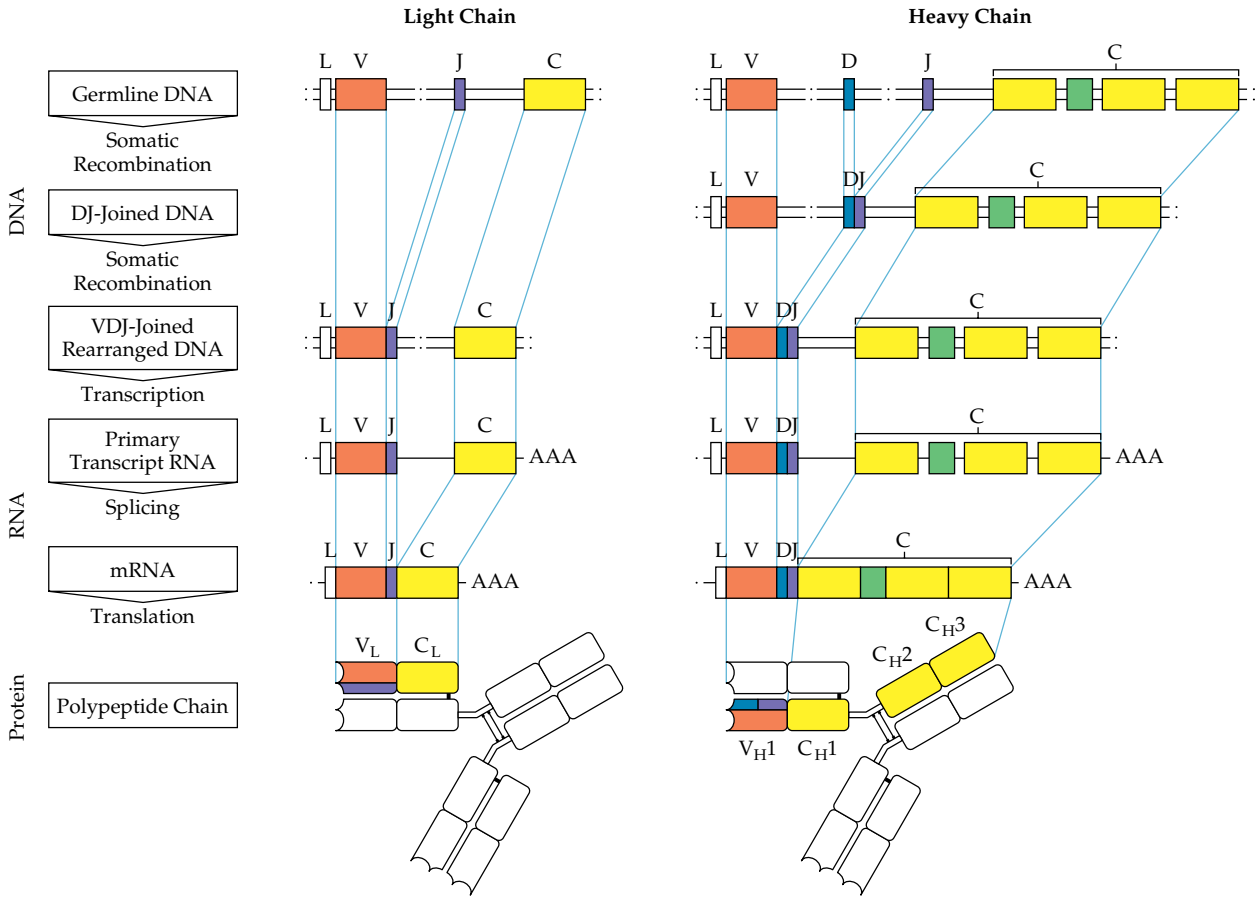


Figure 3 **Variable-region genes are constructed from gene segments. Light-chain variable genes are constructed from two segments (center panel). A variable (V) and a joining (J) gene segment in the genomic DNA are joined to form a complete light-chain variable-gene–region gene. The constant region is encoded in a separate exon and is joined to the variable-region gene by RNA splicing of the light-chain message to remove the L to V and the J to C introns. Immunoglobulin chains are extracellular proteins, and the V gene segment is preceded by an exon encoding a leader peptide (L), which directs the protein into the cell's secretory pathways and is then cleaved. Heavy-chain variable regions are constructed from three gene segments (right panel). First the diversity (D) and J gene segments join, then the V gene segment joins to the combined DJ sequence, all at the genomic DNA level. The heavy-chain constant-region sequences are encoded in several exons: note the separate exon encoding the hinge domain (purple). The constant-region exons together with the L sequence are spliced to the variable-domain sequence during processing of the heavy-chain gene RNA transcript. Posttranslational alterations remove the L sequence and attach carbohydrate moieties.[19]**

ing the secondary response is much greater and lasts longer than the peak level during the primary response. This variation in response is an apt illustration of immunologic memory and is caused by a proliferation of antigen-specific B cells and helper T cells during the primary response. The characteristics of the primary and secondary responses explain the need for booster injections in immunization programs.

AFFINITY MATURATION BY HYPERMUTATION

The binding properties of antibodies change with time by a process termed affinity maturation, which involves somatic hypermutation and selection. After the first stimulation, the antibodies have progressively greater affinity for the antigen as antigen exposure progresses, and increasingly stable antigen-antibody complexes are formed. In addition, the antibody becomes less specific, and cross-reactions with related antigens increase. The lessening specificity reflects the fact that cross-reactions were previously too weak to detect; they become apparent as antibody develops greater affinity for antigen.

IMMUNOGLOBULIN CLASS SWITCH RECOMBINATION

The genes that code for the IgM and IgD heavy chains (the μ and δ genes, respectively) play a critical role in the primary immune response. Whereas IgM antibodies are unable to act in many tissues of the human body, IgG, IgA, and IgE serve functions in the peripheral immune system. Class switching means that the same variable region can be transferred from the heavy chain of IgM to one of the other antibody heavy chains. In the switch from IgM to IgG production that constitutes the booster response, constant-region genes are deleted before the DNA is transcribed to RNA.[6,7] If the cell switches to production of IgG3, for example, the genes for μ and δ are deleted [*see Figure 4*]. After transcription, RNA splicing produces an mRNA with the sequence V-D-J-$C_{\gamma 3}$, which is translated into protein.

MECHANISMS INVOLVED IN SOMATIC HYPERMUTATION AND CLASS SWITCH RECOMBINATION

An exciting recent breakthrough explains the mechanism underlying both somatic hypermutation (SHM) and class switch

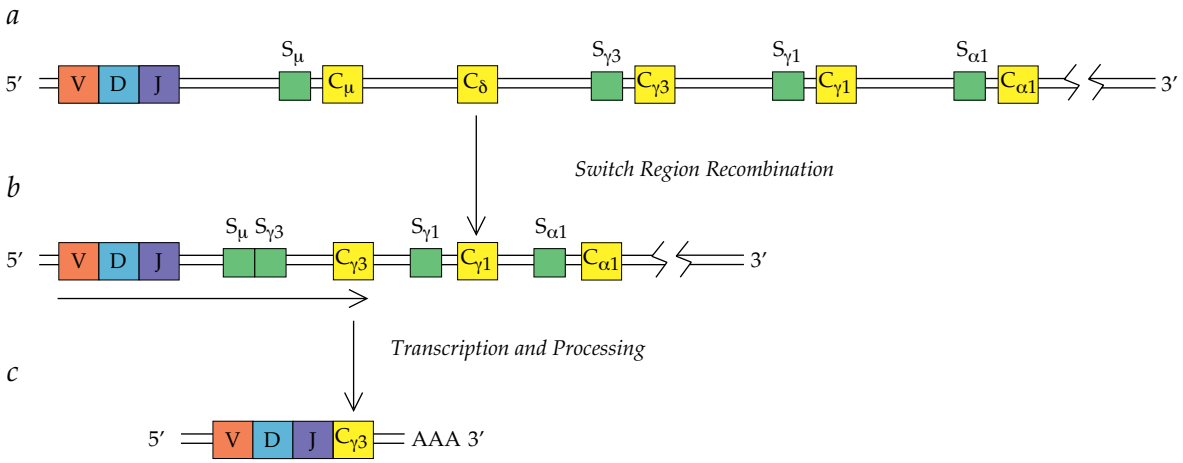


Figure 4 **In the booster response, a plasma cell switches from IgM production to IgG production, a process called class switching. In this hypothetical model of heavy-chain class switching to $C_{\gamma 3}$, the heavy-chain C-region exon clusters (yellow) and the switch regions (green) are indicated (*a*). The switch region is a stretch of DNA that directs the deletion events. In this model, recombination of the switch regions S_{μ} and $S_{\gamma 3}$ and the deletion of the intervening DNA occur first to produce a DNA sequence in which the gene for $C_{\gamma 3}$ has been brought into close proximity to the V-D-J segment (*b*). Further processing and transcription of this DNA yields the messenger RNA (mRNA) encoding IgG3 (*c*).**

recombination (CSR). This was the discovery of activation-induced cytosine deaminase (AID), an enzyme necessary for both processes. SHM and CSR occur only in the B cells in germinal centers of lymph nodes that have been stimulated by antigen. Further, AID is expressed only in antigen-stimulated B cells.

AID converts a cytosine to uracil in the variable region of the antibody gene. The cell regards this as an error, because uracil does not belong in DNA, and the correction process can introduce a variety of mutations. If a glycosylase removes the uracil, then during the subsequent replication or repair process, low-fidelity or error-prone DNA polymerase may fill in the gap with a different base or may fill it in and extend it by strand displacement. If a glycosylase does not remove the uracil, then in the subsequent replication process, a high-fidelity DNA polymerase may recognize the uracil as thymine and pair it with an adenine [*see Figure 5*].

Enhancers in the intron DNA also play a role in determining the location of the hypermutation. The AID enzyme plays a similar role in CSR, determining which heavy chain the VDJ will be switched to from the μ chain. The processes are not exactly parallel, however. For instance, SHM occurs in the G2 phase of the cell cycle and involves homologous recombination and a κ intronic enhancer element, whereas CSR occurs in the G1 phase and involves nonhomologous end joining and a 3′ immunoglobulin heavy (IgH) gene enhancer element. Thus, the same unique mechanism is employed in the development of two quite different properties of antibodies, with their high specificity and their function depending on SHM and CSR, respectively.

Immunoglobulin Receptors

B cell receptors

In addition to being secreted, immunoglobulins are also expressed on the surface of B cells, where they act as antigen receptors.[8] These cell surface membrane immunoglobulins (SmIgs) differ from secreted immunoglobulins in that they have a transmembrane domain and are monomeric. The first SmIg that a B cell expresses is IgM; at a later stage of B cell development, IgD is coexpressed. SmIgs do not travel to the cell surface by themselves; the process requires formation of a complex consisting of the immunoglobulin and two polypeptide chains called Igα and Igβ, which takes place in the endoplasmic reticulum [*see Figure 6a*]. Binding of the resultant BCR with antigen drives the B cell to maturation. Stimulation by the helper T cell activates the B cell, causing it to differentiate into a plasma or memory cell that produces secretory antibody specific for the antigen encountered. Igα and Igβ are not expressed after terminal differentiation. The mature plasma cell ceases to express SmIgs, although it may retain the SmIg mRNA.

BCRs play important roles in regulation of the immune response. B cell responses to antigen can become anergic, thus providing a control mechanism for B cell responses and antibody production. A precursor of the BCR expressed on the surface of pre–B cells is thought to control allelic exclusion. In addition, BCRs interplay with Fc receptors.

Fc receptors

Fc receptors bind the Fc portion of an immunoglobulin; they are expressed on a multitude of cells, including mast cells, macrophages, eosinophils, and tumor cells. Fc receptors are composed of a family of molecules. The Fc receptor for IgE (FcεRI) is the model for all Fc receptors and consists of three polypeptide chains, designated α, β, and γ. FcεRI mediates signal transduction in the mast cell when IgE binds to the receptor. FcεRIα is the binding site for the Fc portion of IgE. FcεRIβ is a transmembrane molecule that connects FcεRIα with FcεRIγ, the chain responsible for recruiting signal transduction molecules.[9,10] FcRIIβ1, another Fc receptor that is expressed on B cells, provides a negative feedback signal to the BCR, which leads to the termination of humoral immune responses.[11]

Engineered Antibodies

Lymphocyte hybridoma

The development of the lymphocyte hybridoma, a product of cell fusion, has had a revolutionary impact on immunology and clinical medicine. B lymphocyte hybridomas are the means by

which extraordinarily high titers of very specific, pure antibodies can be produced for experimental and clinical purposes. The B lymphocyte hybridoma, as developed by Köhler and Milstein,[12] is the product of the fusion of a mouse myeloma cell and a lymphocyte from the spleen of a mouse immunized with a specific antigen. The hybrids can be cloned and selected for specific antibody production.

HUMAN MONOCLONAL ANTIBODIES

Several methods of generating human monoclonal antibodies exist. One method entails taking the complementary DNA coding for a mouse monoclonal antibody and systematically replacing the mouse sequences with human sequences. Another method of humanizing mouse monoclonal antibodies entails making a transgenic mouse containing large segments of human DNA, including several variable regions and all the human constant regions.[13] Because this mouse still has its own immunoglobulin regions, it is bred with a mouse in which there has been a targeted disruption of the mouse J region of the immunoglobulin heavy-chain and κ light-chain genes. The progeny of this breeding can then be injected with any antigen, and they will produce humanized monoclonal antibodies to it. Because the transgenic mouse contains a limited number of human variable regions, the potency of these antibodies relies on the natural somatic mutation and affinity maturation that occurs in the mouse. A third method of generating human monoclonal antibodies entails constructing expression libraries of human variable regions either in bacteria or in bacteriophages. In theory, all the variable regions can be cloned. The antibody can be expressed on the surface of the bacterium or bacteriophage and selected by affinity to the antigen of interest.

Monoclonal antibodies are being used in many therapies. For instance, humanized monoclonal antibodies to tumor necrosis factor–α (TNF-α) have been used successfully in the treatment of Crohn disease and rheumatoid arthritis. In other applications, monoclonal antibodies are used to remove T cells and tumor cells before bone marrow transplantation and during acute transplant rejection. Other potential uses of monoclonal antibodies are the production of anti-IgE antibodies and anti–hormone receptors to prevent allergy and modulate endocrine abnormalities, respectively. A monoclonal anti-IgE, omalizumab, was recently approved by the FDA for treatment of allergic asthma.

T Cell Receptors

Unlike the immunoglobulin receptors on B cells, which recognize free antigens, TCRs recognize antigens only in conjunction with autologous MHC antigens, which are expressed on the surface of professional APCs (e.g., dendritic cells, macrophages, and B cells). The $CD4^+$ helper T cells (as well as the few $CD4^+$ cytotoxic T cells) require MHC class II molecules, and the $CD8^+$ cytotoxic T cells require MHC class I molecules [*see 6:V Adaptive Immunity: Histocompatibility Antigens and Immune Response Genes*]. This phenomenon is referred to as MHC restriction. The ability of T cells to recognize these self-MHC molecules is determined during development in the thymus before the lymphocytes are exposed to antigen [*see 6:I Organs and Cells of the Immune System*].

The molecules that make up the TCR have been identified, and the genes that encode these molecules have been isolated and cloned. The TCR is composed of six distinct polypeptides, known as the TCR-CD3 complex.[14-16] From 85% to 95% of normal peripheral blood lymphocytes carry TCR-αβ; only 5% to 15% carry TCR-γδ. The antigen-recognizing portion of the TCR-αβ complex consists of two glycosylated polypeptide chains, termed TCR-α and TCR-β, that are linked by disulfide bonds to form a heterodimer. The corresponding polypeptide chains in the TCR-γδ complex consist of TCR-γ and TCR-δ. The TCR-α, TCR-β, TCR-γ, and TCR-δ chains each contain variable and constant portions that are analogous to those of immunoglobulin molecules.

TCR-αβ and TCR-γδ heterodimers are closely associated with the CD3 proteins CD3γ, CD3δ, CD3ε, and CD3ζ [*see Figure 6b*]. The CD3 proteins are present on all peripheral blood T cells and on 90% of thymocytes. Expressions of the CD3 and the TCR complexes on the cell surface are mutually dependent: neither complex is observed on the surface of the T cell without the other. Structural and functional data suggest that the activities of the TCR are distributed among the subunits of the TCR-CD3 complex: the TCR polypeptides (α, β, γ, and δ) bind to antigen and

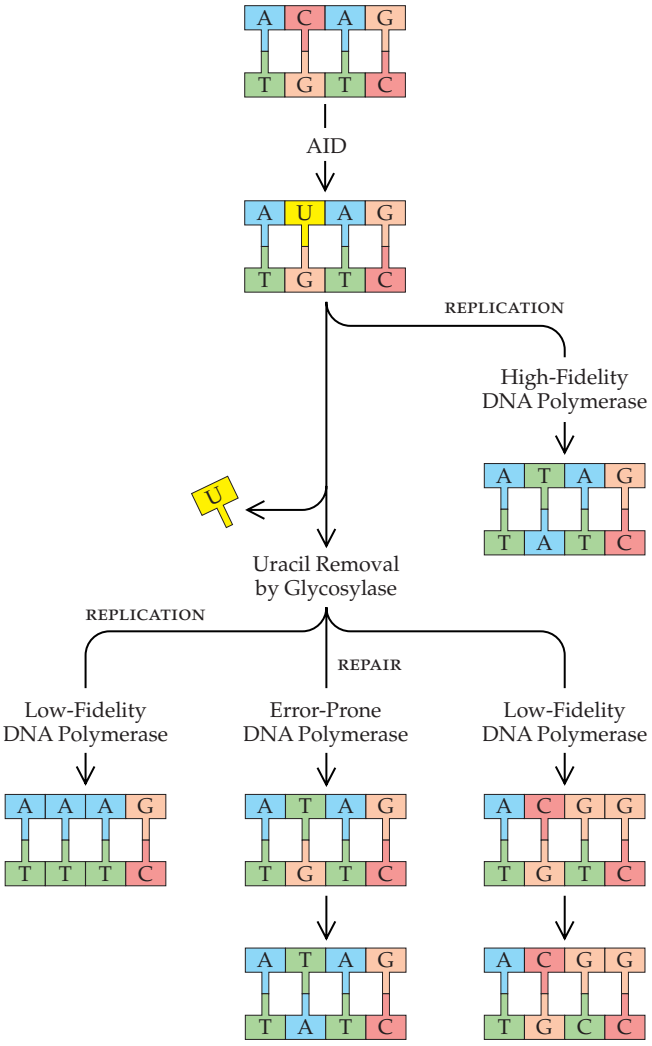


Figure 5 **Antibody diversity is promoted by the enzyme activation-induced cytosine deaminase (AID), which generates mutations by converting cytosine (C) to uracil (U) in the variable region of the antibody gene. During replication, high-fidelity DNA polymerase will recognize the uracil as a thymine (T) and pair it with an adenine (A). Alternatively, the uracil may be removed by a glycosylase. Subsequently, mutations may occur during replication, as a low-fidelity DNA polymerase fills the gap more or less at random, or during repair, as an error-prone DNA polymerase preferentially fills the gap with thymine or as a low-fidelity DNA polymerase fills in the gap and extends it by synthesizing bases by strand displacement, which mostly occurs opposite adenine and thymine.[20]**

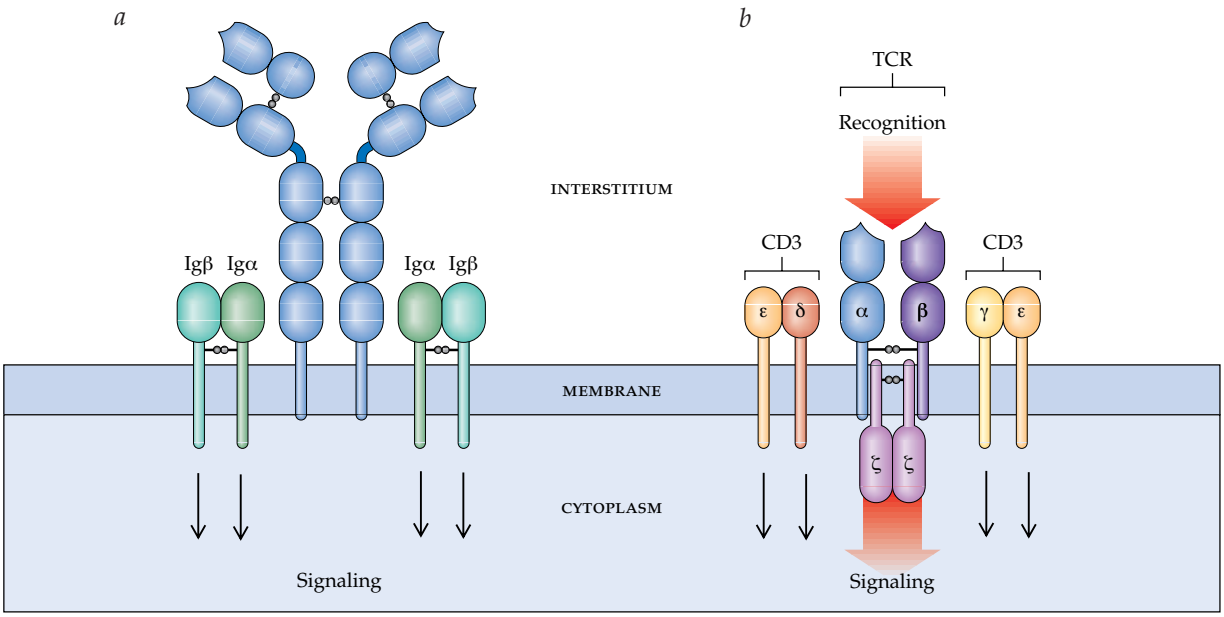


Figure 6 **(*a*) Cell surface membrane immunoglobulins form a complex with the proteins Igα and Igβ. Igα and Igβ are linked by disulfide bonds, but the exact stoichiometry is unknown. The exact ratio of these two proteins to each immunoglobulin molecule is also unknown. (*b*) The TCR-CD3 complex is shown. A T cell receptor for antigens is composed of six distinct polypeptide chains. Two of the chains, α and β, are the disulfide-bonded chains of the heterodimer TCR that binds to antigen. The four other chains—γ, δ, and two ε chains—are collectively called CD3. CD3 associates with TCR and transports it to the T cell surface. When antigen binds to the TCR, CD3, along with a homodimer of ζ chains, sends a signal to the nucleus, via intracellular signaling pathways. Specific genes are then transcribed, and cytokines, chemokines, and other immunodulatory molecules are produced that mediate the antigen-specific immune response.**

MHC gene products, and the CD3 proteins transduce the binding signal to the cytoplasm of the T cell, which results in activation of T cell functions.

By the use of mice containing spontaneous and engineered mutations of various genes of TCRs, the development pathway of the T cell has been confirmed. The organization of the genes that encode the human TCR-α, -β, -γ, and -δ chains is analogous to that of the immunoglobulin heavy-chain genes: there are V, D, and J segments, which are flanked by recognition sequences that mediate site-specific recombination [*see* Antibodies, *above*]. Thus, the diversity of TCRs is generated by many of the same mechanisms that are used by B cells for the production of immunoglobulins. T cells and B cells may in fact use the same recombination enzyme, or recombinase.

The genomic sequences that encode the TCR-β chain contain two very similar constant-region genes, $C\beta_1$ and $C\beta_2$, each of which is associated with a cluster of six or seven J genes and a single D gene. There are at least 70 Vβ genes that are associated with the two Cβ genes. These variable regions are distinct from the immunoglobulin variable regions. Rearrangement of the β-chain gene segments can lead to the production of approximately 3,600 different β chains.

The TCR-α genes are arranged differently. A single Cα gene is preceded by a very large stretch of DNA containing at least 50 distinct J genes. A Dα gene segment has not been demonstrated directly. Some Vα genes are organized as families of related genes. Rearrangement of the α-chain gene segments can account for approximately 2,500 different polypeptides. No somatic hypermutation has been detected in TCRs. Thus, 10^7 TCR-αβs can be formed. The genes for CD3γ, CD3δ, CD3ε, and CD3ζ are transcribed in all T cells; however, these genes do not undergo rearrangement.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Furlong ST, David JR: Adjuvants. Therapeutic Immunology. Austen KF, Burakoff SJ, Rosen FS, Eds. Blackwell Scientific Publications, Cambridge, 1996

2. Frank KM, Sekiguchi JM, Seidl KJ, et al: Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. Nature 396:173, 1998

3. Oettinger MA, Schatz DG, Gorka C, et al: RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science 248:1517, 1990

4. Hofker MH, Walter MA, Cox DW: Complete physical map of the human immunoglobulin heavy chain constant region gene complex. Proc Natl Acad Sci USA 86:5567, 1989

5. Seidl KJ, Manis JP, Bottaro A, et al: Position-dependent inhibition of class-switch recombination by PGK-neor cassettes inserted into the immunoglobulin heavy chain constant region locus. Proc Natl Acad Sci USA 96:3000, 1999

6. Villa A, Santagata S, Bozzi F, et al: Partial V(D)J recombination activity leads to Omenn syndrome. Cell 93:885, 1998

7. Rolink A, Melchers F, Andersson J: The SCID but not the RAG-2 gene product is required for S mu-S epsilon heavy chain class switching. Immunity 5:319, 1996

8. Kurosaki T: Genetic analysis of B cell antigen receptor signaling. Annu Rev Immunol 17:555, 1999

9. Kinet JP: The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. Annu Rev Immunol 17:931, 1999

10. Metzger H: It's spring, and thoughts turn to...allergies. Cell 97:287, 1999

11. Scharenberg AM, Kinet JP: PtdIns-3,4,5-P3: a regulatory nexus between tyrosine kinases and sustained calcium signals. Cell 94:5, 1998

12. Köhler G, Milstein C: Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256:495, 1975

13. Lonberg N, Taylor LD, Harding FA, et al: Antigen-specific human antibodies from mice comprising four distinct genetic modifications. Nature 368:856, 1994

14. Terhorst C, Spits H, Staal F: T lymphocyte signal transduction. Molecular Immunology. Hames BD, Glover DM, Eds. IRL Press, Washington, DC, 1995, p 237

15. Garcia KC, Teyton L, Wilson IA: Structural basis of T cell recognition. Annu Rev Immunol 17:369, 1999

16. Davis MM, Boniface JJ, Reich Z, et al: Ligand recognition by alpha beta T cell receptors. Annu Rev Immunol 16:523, 1998

17. Tonegawa S: The molecules of the immune system. Sci Am 253:122, 1985

18. Underdown BJ, Schiff JM: Immunoglobulin A: strategic defense initiative at the mucosal surface. Annu Rev Immunol 4:389, 1986

19. Janeway CA Jr, Travers P: Immunobiology: The Immune System in Health and Disease, 3rd ed. Current Biology Ltd, London, 1997

20. Gearhart PJ: Roots of antibody diversity. Nature 419:29, 2002

Acknowledgments

Figures 1 through 6 Seward Hung.

IV CELL-CELL INTERACTIONS, CYTOKINES, AND CHEMOKINES IN IMMUNE RESPONSE MECHANISMS

John R. David, m.d.
Cox Terhorst, ph.d.

The immune response is defined by three principles: discrimination between self and nonself, specificity, and memory [*see 6:I Organs and Cells of the Immune System*]. This chapter will discuss the manifestation of those principles through antigen processing and presentation, the T cell response to antigen, interactions between T cells and B cells, and the actions of cytokines and chemokines. The cellular and humoral mechanisms of innate immunity are described in detail elsewhere [*see 6:II Innate Immunity*].

Two principal arenas of the immune response are sites of pathogenic invasion and the lymph nodes that drain these sites. The immune response begins with exposure of epithelial cells, macrophages, and dendritic cells to a pathogen. In the lymph nodes, these cells (antigen-processing cells [APCs]) concentrate and process antigens and present them to T and B cells.

The critical first step of the T cell immune response to a specific antigen is the recognition and binding of processed antigen on the surface of an APC with a T cell receptor (TCR) on the surface of a helper ($CD4^+$) T cell [*see 6:I Organs and Cells of the Immune System*]. This event is relayed to the helper T cell nucleus by a cascade of cytoplasmic signaling molecules. In the nucleus, activation of specific transcription factors stimulates expression of the genes that encode soluble factors—cytokines and chemokines—that mediate the immune response. This response has two aspects: humoral and cell mediated. In the humoral response, cytokines secreted by a specific form of activated T cells induce antigen-stimulated B cells to differentiate into antibody-secreting plasma cells. In the cell-mediated response, cytokines from $CD4^+$ T cells induce $CD8^+$ T cells to differentiate into cytolytic effectors and also can activate macrophages, another effector cell. B cells recognize nonprocessed antigens in solution or antigens that are attached to the surface of follicular dendritic cells—a cell type that is particularly adept at antigen processing, presentation, and retention. Both T cells and B cells are induced to expand their population and control the initial infection and to produce memory cells for long-term acquired immunity.

Antigen Processing and Presentation

THE MAJOR HISTOCOMPATIBILITY COMPLEX

The major histocompatibility complex (MHC) is a membrane glycoprotein complex that binds antigenic peptide in the cytoplasm of an APC and transports it to the cell surface for interaction with T cells.[1] An extensive polymorphism exists in the MHC gene (i.e., there are many alleles per locus); however, each person expresses only a small number of different MHC molecules. To ensure an adequate immune response against a wide range of nonself antigens, each MHC molecule must be able to bind a large number of different peptides.

Two main classes of MHC have been identified. The two classes of molecules have a similar structure: two immunoglobulin-like domains and a binding site for processed antigens (peptides) [*see 6:V Adaptive Immunity: Histocompatibility Antigens and Immune Response Genes*]. Whereas MHC class I molecules bind only smaller peptides of defined lengths (eight to 11 amino acids), MHC class II molecules bind longer peptides with no apparent restriction on peptide length. An interesting finding is that certain peptides bind only to specific alleles of either MHC class I or MHC class II molecules. Thus, persons who lack one of those alleles would not develop an immune response to its associated peptide.

The differences in peptide binding between MHC class I and MHC class II molecules result from small structural dissimilarities within the relatively fixed framework of the peptide-binding site and probably also from fundamental differences in the mechanism of peptide processing, which takes place in the endoplasmic reticulum (ER) for MHC class I molecules and in the endosomes and lysosomes for MHC class II molecules.

The MHC Class I Pathway

Antigen processing with MHC class I molecules essentially entails three steps: (1) generation of antigenic peptides in the cytosol of the APC, (2) transport of the peptides into the ER, and (3) assembly of the peptide–MHC class I complexes. The completed complexes then migrate through the Golgi apparatus to the cell surface and insert in the plasma membrane for presentation on the extracellular surface [*see Figure 1a*].

The MHC Class II Pathway

Assembly of antigenic peptide–MHC class II molecule complexes in the APC requires four steps: (1) uptake of exogenous antigens by vesicles (endosomes, lysosomes, and, possibly, undefined endosomal subcompartments), (2) proteolytic degradation of proteins in the endosome, (3) assembly of MHC class II molecules in the ER and migration of these molecules through the Golgi apparatus to the endosomes, and (4) assembly of the peptide–MHC class II complexes in the late endosome. After the peptides are bound to MHC class II molecules, the endosome containing these complexes migrates to the cell surface and inserts in the plasma membrane [*see Figure 1b*]. Of note is that partially unfolded antigenic proteins can bind to MHC class II molecules before undergoing proteolytic degradation. This may explain why the length of peptides bound to MHC class II molecules is highly variable.

ALTERNATIVE ANTIGEN-PRESENTING COMPLEXES

A third class of MHC includes CD1 molecules, which are related to MHC class I molecules but which bind lipid and glycolipid antigens for presentation to T cells. For example, CD1

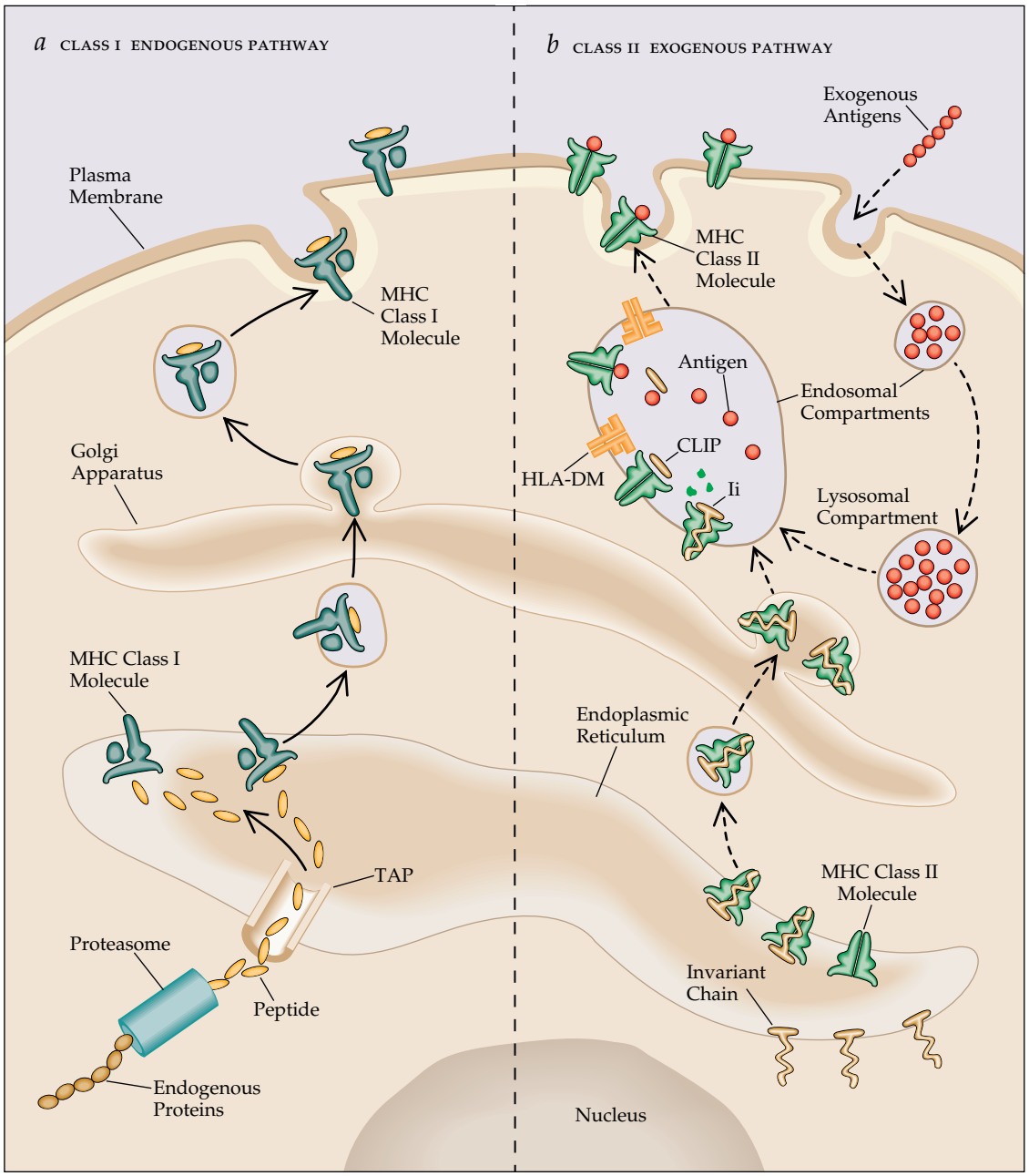


Figure 1 **The pathways for formation of antigenic peptide–MHC molecule complexes on antigen-presenting cells. (*a*) In the MHC class I molecule pathway, endogenous proteins are broken down by proteasomes into smaller peptides. In the endoplasmic reticulum (ER), an antigenic peptide binds to the peptide binding site in an MHC class I molecule. The peptide-MHC complex then migrates through the Golgi apparatus to the cell surface. (*b*) In the MHC class II molecule pathway, the α and β subunits of the MHC class II molecule bind to the invariant chain (Ii) in the ER. Ii is partially degraded in an endosomal compartment. The portion of Ii that occupies the antigenic peptide binding site on the MHC class II molecule (called CLIP) is removed with the help of HLA-DM, freeing the molecule for binding processed antigen. Once antigenic peptide has bound to an MHC class II molecule, the complex migrates to the cell surface.**

molecules present glycolipids from *Mycobacterium tuberculosis* to a restricted group of T cells.[2] This ability to recognize non-protein microbial antigens suggests that T cells recognize a broader range of antigens than was once thought. A subset of $CD1^+$ T cells often reacts to self-antigens; these cells have been implicated in such autoimmune diseases as type 1 diabetes mellitus and systemic lupus erythematosus. It has been suggested that they are involved at the early innate phase of these immune responses.[3] Another molecule related to MHC class I, MR1, is found on a subset of T cells that are preferentially located in the gut lamina propria and are called mucosal-associated invariant T (MAIT) cells. MAIT cells are probably involved in the host response at the site of pathogen entry and may regulate intestinal B cell activity. MAIT cells are absent in humans and mice that have B cell deficiency, which suggests that the selection or expansion of this cell population requires B cells.[4]

Superantigens

Superantigens constitute a class of immunostimulatory proteins derived from microbial agents (e.g., viral proteins and the staphylococcal toxins that cause toxic-shock syndrome and food poisoning). Superantigens bind to MHC class II molecules outside the conventional antigen-binding site and stimulate massive T cell activation[5] [*see 7:XXX Sepsis*]. Different MHC class II alleles have distinct binding constants for superantigens; thus, superantigens can activate distinct segments of the T cell repertoire.

PROFESSIONAL ANTIGEN-PRESENTING CELLS

Whereas MHC class I molecules are expressed on the surface of all eukaryotic cells, MHC class II molecules have a restricted tissue distribution. In fact, certain cells—including B cells, dendritic cells, macrophages, Langerhans cells, and endothelial cells—are termed professional APCs because they present antigenic peptide with MHC class II molecules more efficiently than other APCs.[6] This efficiency is primarily attributed to their ability to process endocytosed antigens. Professional APCs also interact with T cells more efficiently because they have several cell surface markers that bind to costimulatory molecules on the surface of T cells (see below). The most potent APCs are dendritic cells, which present antigen only on maturation and migration to the lymph nodes. This maturation process is triggered by the uptake of antigen.

T Cell Responses to Antigen

T cell recognition of antigen proceeds in two distinct stages. The first step, which is nonspecific, is the adhesion of a T cell to an APC. The second step, which is specific, is an interaction between the TCR and a compatible antigen-MHC complex on the APC. This process highlights two fundamental properties of T cells: their broad scope of action (i.e., their ability to migrate throughout the body and adhere to many types of cells) and their great specificity for particular antigens.

The diversity of the variable regions in TCRs facilitates highly specific responses to antigens [*see 6:III Adaptive Immunity: Antigens, Antibodies, and T Cell and B Cell Receptors*]. The prototypical T cells, αβ T cells, have a TCR made up of α and β chains, which are expressed in association with CD3. A subset of peripheral T cells, γδ T cells—so called because their TCR is made up of γ and δ chains—may recognize antigen that has not been processed.

APCs process many peptide antigens simultaneously and thus express on their cell surfaces a large number of different antigen-MHC complexes. A given αβ T cell clone can recognize only a small number of these complexes. T cells screen APCs for compatible antigen-MHC complexes by adhering to the APC. Adhesion is aborted if the TCRs do not recognize specific antigen; adhesion is intensified when a TCR makes contact with a compatible antigen-MHC complex. On adhesion, TCR-CD3 complexes aggregate on the surface of the T cell and bind to the antigen-MHC complexes that have aggregated on the surface of the APC. In the absence of antigen recognition, TCR-CD3 complexes are unable to cluster and detachment occurs immediately.

Interactions between a helper T cell clone and an APC or between a cytotoxic T cell and its target cell follow the same general pattern. Initially, only a very small number of antigen-MHC complexes engage with specific TCRs in an area of cell-cell contact established by adhesion. Subsequently, TCR-CD3 complexes and MHC molecules bearing the correct antigenic peptide migrate into the contact region. This establishes a high local density of TCRs, which promotes antigen binding and T cell activation. The existence of these clustered TCRs has been demonstrated experimentally, through the use of tagged monoclonal antibodies.[7] In addition, visualization of clusters of TCRs using fluorescence-tagged antigen-MHC complexes has permitted incisive studies of T cell responses to infections with Epstein-Barr virus or HIV.[8,9]

COSTIMULATORY MOLECULES

The recognition of antigen by the TCR and subsequent activation of the T cell is regulated by other T cell surface molecules [*see Figure 2 and Table 1*].[10] Costimulatory molecules are involved in both adhesion and T cell signaling and play a major role in the coordination and kinetic regulation of T cell activation. Clinically, these costimulatory molecules are important because aberrant activation of the TCR can initiate disastrous immune responses, such as those seen in autoimmune diseases.

The costimulatory molecules CD28 and cytotoxic T cell–associated antigen 4 (CTLA-4) initiate a signaling pathway that is different from, and often independent of, the pathway mediated by the TCR-CD3 complex. CD28 is expressed on the surface of essentially all $CD4^+$ and most $CD8^+$ T cells. CD28 binds to B7 on APCs, leading to T cell activation and proliferation. Although CTLA-4 is homologous to CD28, it has opposite effects: binding of CTLA-4 to B7 delivers an inhibitory signal that leads to downregulation of T cell proliferation. Of note is that research entailing manipulation of CD28–CTLA-4 interactions with their natural ligands suggests that these costimulatory molecules have a potential role in the treatment of such diseases as arthritis, multiple sclerosis, and asthma and in protection against HIV infection.[11] For example, a soluble construct of CTLA-4 (called CTLA-4–Ig) is able to inhibit T cell activation when administered early in the immune response. Several new members of the B7 and CD28–CTLA-4 families have recently been discovered, and these may also be important for regulating the responses of previously activated T cells.[12]

The SLAM (signaling lymphocyte activation molecule) family of receptors serves as costimulatory molecules on T cells (and other immune cells). In T cells and natural killer (NK) cells, SLAM-associated protein (SAP) regulates signaling of the SLAM family of receptors. The *SAP* (or *SH2D1A*) gene is defective or absent in patients with X-linked lymphoproliferative syndrome.[13,14]

Fas (also called APO-1 and CD95) has been implicated in both positive and negative signaling events after binding with its ligand (CD95L) on cytolytic effector cells. Fas is a member of the tumor necrosis factor receptor (TNFR) family, and binding of Fas on T cells with its ligand typically causes the programmed death of the target cell (i.e., apoptosis; see below). However, Fas can also function as a costimulatory molecule for TCR-CD3 activation. Thus, a single molecule can have different signaling outcomes at different stages of T cell development.

T Cell Signaling after Antigen Recognition

In response to antigen recognition, resting T cells undergo a complex series of events known as T cell activation.[15,16] The recognition and binding of TCRs to antigen-MHC complexes and of costimulatory molecules to their appropriate ligands—

all of which takes place on the cell surface—is followed by an intracellular cascade of biochemical events (i.e., signal transduction) that ultimately reaches specific genes in the nucleus. The resulting production of cytokines, chemokines, and other immunomodulatory molecules leads to cell proliferation, differentiation, and expression of unique effector functions.

Most of the biochemical events that occur immediately after engagement of the TCR with the antigen-MHC complex have been defined [*see Figure 3*]. Transduction of signals starts with the CD3 elements of the TCR-CD3 complex and proceeds to the nucleus via different pathways. One pathway involves the calcium-dependent enzyme calcineurin; other coupled pathways involve the enzymes Ras and Rac and serine-threonine kinase. These pathways converge on the activation of transcription factors that control the expression of genes mediating T cell effector function (e.g., the gene for the cytokine interleukin-2 [IL-2]) [*see Figure 3*]. Thus, the exquisitely T cell clone–specific TCR connects, via a number of intermediate molecules, with universal signal transduction pathways.

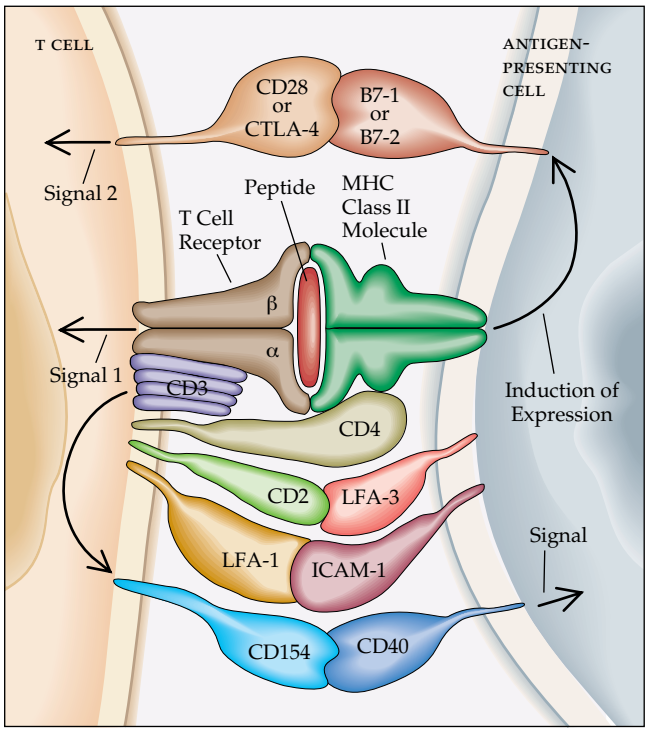
Drugs are being developed to interfere with nodal points of these signal transduction pathways. However, cyclosporine, which interferes with the calcineurin pathway, remains the most successful agent of this type.

B Cell Responses to Antigen

When a B cell receptor (BCR) binds soluble antigen, one of two events takes place: apoptosis or proliferation and further maturation of B cells. The signals for these processes are generated intracellularly by Igα and Igβ. These proteins are associated with BCRs in the way that CD3 proteins are associated with TCRs. Igα and Igβ recruit signal transduction molecules in a manner similar to that for activating T cells and involving some of the same proteins.

Relatively little is known about the signaling that takes place in B cells when they have presented an antigen to a helper T cell that recognizes that antigen. T cell–B cell interaction can also lead to either apoptosis or proliferation of B cells. With T cell help, however, proliferation results in the generation of several different classes of B cells.

a



b

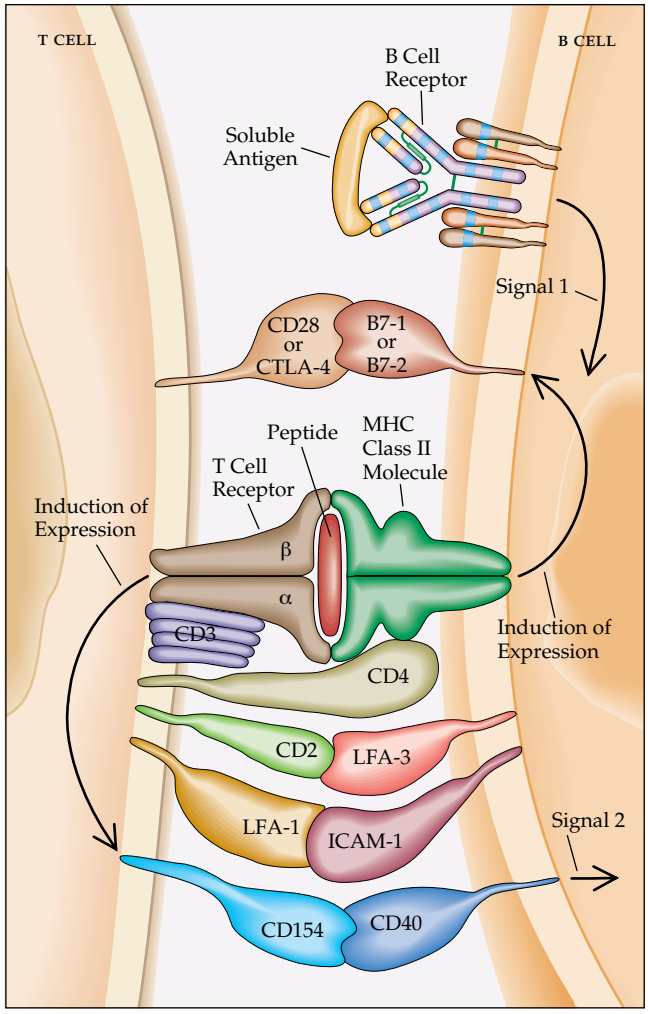


Figure 2 **(*a*) Two signals are necessary for activation of an antigen-specific T cell by an antigen-presenting cell (APC). Signal 1 is initiated by the interaction between the antigen bound to a class II major histocompatibility complex (MHC) and the T cell receptor (TCR) and its coreceptor (in this example, CD4). The costimulatory molecules B7-1 and B7-2 are transiently expressed on the surface of so-called professional APCs and are presumed to be inducible by signaling from the antigen-MHC complex. Thus, signal 2 is initiated by the binding of CD28 on the T cell to B7-1 or B7-2. CTLA-4, which is homologous to CD28, is upregulated after T cell activation. (*b*) Activation of B cells can occur with either helper T cells or soluble antigen. Binding of soluble antigen to the B cell receptor, a cell surface immunoglobulin associated with Igα and Igβ, can lead to B cell proliferation or apoptosis. Alternatively, B cells can process antigen for presentation to TCRs on T cells. Binding of antigen-MHC complex to the TCR induces expression of the CD40 ligand (CD154) on the T cell surface, which in turn induces expression of CD40 on the B cell surface. This results in B cell proliferation and is indispensable for immunoglobulin class switching and probably for somatic mutation. CD154 signaling also plays an important role in the maturation of dendritic cells. (ICAM-1—intercellular adhesion molecule–1; LFA—leukocyte-function–associated antigen)**

B CELL SUBSETS

Follicular Mantle Cells

Follicular mantle cells derive their name from their location, surrounding the active germinal centers in lymph nodes. These cells express IgD but not CD38 on their surface (i.e., they are $IgD^+,CD38^-$) and can mature into plasma cells that produce IgM antibodies only. These cells produce IgM antibodies that display extensive heterogeneity, but the cells do not undergo somatic mutation to generate larger numbers of heterogeneous antibody molecules.

Centrocytes

Centrocytes are found in the germinal center of lymph nodes. The process leading to somatic mutation occurs in the centrocytes, which are $IgD^-,CD38^+$ B cells that have undergone class switching to produce IgG. The maturation step from $IgD^+,CD38^-$ B cells to $IgD^-,CD38^+$ B cells depends on cell-cell contact with antigen-specific T cells and cytokines. IL-4 in particular is of great importance for the events that lead to class switching.

Memory B Cells

The most mature B cell, the memory cell, does not express IgD or CD38 but is distinguished by the presence of other cell surface markers and by its location within the germinal center of lymph nodes. Memory cells can also develop into plasma cells producing IgM, IgG, IgA, and IgE. In general, the generation of plasma cells from memory B cells is independent of T cell help.

T Cell–B Cell Interactions

An important pathway for B cell activation involves CD40, a B cell surface receptor belonging to the TNFR family. The natural ligand for CD40, CD40L (also called CD154), is a glycoprotein related to TNF. Transient CD40L expression on the surface of T cells is induced by the binding of TCR to antigen-MHC complex and by binding of B7 to CD28 or CTLA-4. The binding of CD40 to CD40L stimulates B cell immunoglobulin class switching. This was best demonstrated by elucidation of the molecular defect in a genetic immunodeficiency termed hyper-IgM syndrome [*see 6:VIII Deficiencies in Immunoglobulins and Cell-Mediated Immunity*]. Patients with the X-linked form of this syndrome have a mutation in the gene encoding CD40L that results in defective antibody class switching. T cell–independent B cell responses and responses induced by anti-CD40 are unaffected by this mutation. Thus, T cell help for B cell activation is primarily directed through the CD40-CD40L costimulatory pathway. The observation that persons with hyper-IgM syndrome are susceptible to opportunistic infections such as those seen in AIDS—that is, infections from pathogens normally dealt with by T cells—underscores the role of CD40L costimulation in normal T cell activation. Absence of CD40L has a more dramatic effect on immunoglobulin class switching than does an absence of T cells. Thus, it seems that CD40L may be expressed on cells other than T cells that could induce immunoglobulin class switching in B cells through the ligation of CD40.

Cytokines

CYTOKINES AND T CELL SUBSETS

Cytokines, a diverse group of proteins produced by a number of different cell types, are critical in the regulation of immune responses [*see Table 2*]. They are also important in the differentiation of cell systems. In general, cytokines are synthesized in response to stimulation of cells. In some cases, the active cytokine is released from an inactive precursor by proteolysis. Occasionally, cytokines are stored in cells.

After secretion, cytokines usually act locally by binding to receptors on cell surfaces. Cytokines may act on many different cells, and different cytokines may have similar activities. Cytokine receptors are often composed of the same protein chains. For instance, the cytokine receptors IL-2R, IL-4R, IL-7R, IL-9R, IL-13R, and IL-15R have the γ chain in common. Most cytokines form a network that regulates the activation of cells and the production of other cytokines.[17] The binding of cytokines to receptors on cell surfaces activates intracellular signaling mech-

Table 1 Costimulatory Molecules on T Cells

Molecule	*Ligand*	*Function*
CD4	MHC molecules on the surface of APCs	Helps generate $CD4^+$ T cells in the thymus; directly influences TCR-initiated signaling
CD8	MHC molecules on the surface of APCs	Helps generate $CD8^+$ T cells in the thymus; directly influences TCR-initiated signaling
CD2	CD58 on the surface of APCs	Promotes T cell activation and proliferation
CD11/CD18	ICAM-1 (CD54), the complement component iC3b, extracellular matrix proteins	Promotes T cell activation and proliferation
CD28	B7-1 (CD80), B7-2 (CD86) on the surface of APCs	Promotes T cell activation and proliferation
CTLA-4	B7-1 (CD80), B7-2 (CD86) on the surface of APCs	Inhibits T cell activation and proliferation
SLAM*	SLAM receptors on the surface of APCs	Promotes activation and proliferation of T cells and other immune cells
CD27	CD70 on activated T and B cells	T cell proliferation
Fas (APO-1; CD95)	CD95L on cytolytic effector cells	Causes apoptosis or promotes TCR-CD3 activation

*Also found on natural killer cells.
APC—antigen-presenting cell CTLA-4—cytotoxic T lymphocyte–associated antigen–4 ICAM-1—intercellular adhesion molecule–1 SLAM—signaling lymphocyte activation molecule TCR—T cell receptor

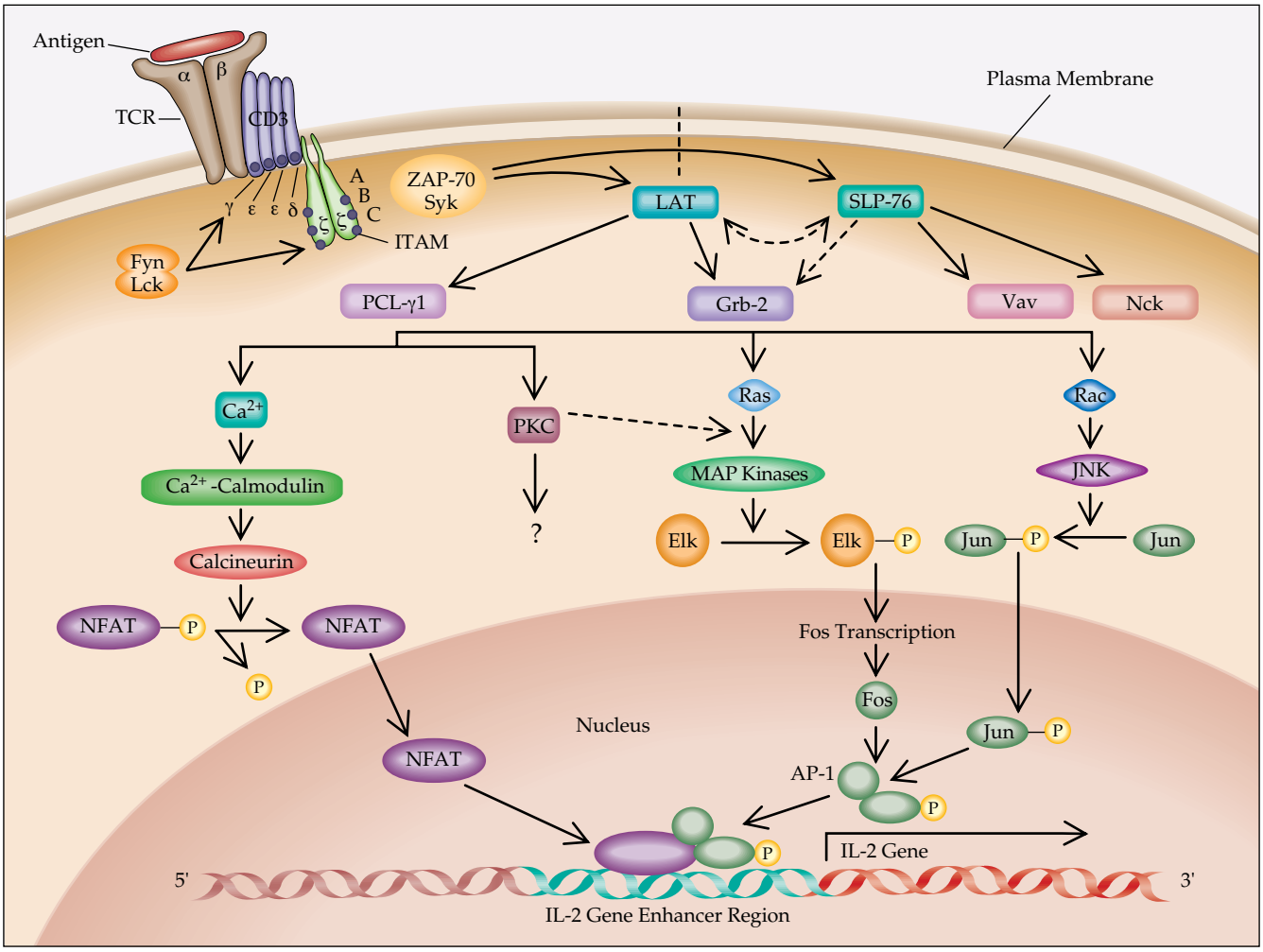


Figure 3 **The downstream signaling pathways induced after TCR stimulation are shown. Phosphorylation of the cytoplasmic tails of the CD3-γ, CD3-δ, CD3-ε, and CD3-ξ by the Src kinase Fyn or Lck recruits the kinase Syk or ZAP-70, which relays the signal of TCR binding through tyrosine phosphorylation of two adapter proteins, LAT and SLP-76. Nck is an adapter protein associated with SLP-76. These adapter proteins act on components of classic signal transduction pathways, including phospholipase C-γ1 (PLC-γ1), growth factor receptor–bound protein–2 (Grb-2, a protein-linking receptor tyrosine kinase), the p85 subunit of phosphatidylinositol-3–kinase (PI-3 kinase), and possibly Vav (a guanosine triphosphate/guanosine diphosphate [GTP/GDP] exchange factor for Rho-family GTPases, including Ras and Rac). Activation of PLC-γ causes the release of diacylglycerol, which in turn activates protein kinase C (PKC). PKC is involved in initiating the cascade of other serine-threonine kinases, including Raf-1, mitogen-activated protein (MAP) kinase, and MAP kinase kinase (MEK). The central signal transduction molecule Ras appears to be involved in late events after PKC activation. The Ras and Rac pathways and the serine-threonine kinase pathways are coupled; they activate early genes, such as *jun* and *fos*. The Ca^{2+} generated by PLC-γ1 activates another downstream cascade, the calcineurin pathway. Calcineurin, a serine-threonine phosphatase, is a calcium- and calmodulin-dependent enzyme involved in induction of transcription factor NFAT (nuclear factor for activated T cells). PI-3 kinase is a ubiquitous enzyme in the mitogenic signaling and apoptotic pathways of both receptor and nonreceptor protein tyrosine kinases. These various signaling pathways converge by delivering a distinct set of transcription factors, including Jun, Elk, NFAT, Fos, and AP-1 (all DNA-binding proteins), to the promoter region of the *IL-2* gene, stimulating expression of the gene and production of IL-2.**

anisms, which lead to expression of particular genes (e.g., genes for the bound cytokine, or other cytokines). The principal intracellular cytokine signal transducers are two families of transcription factors: Jak (Janus kinase) protein tyrosine kinases and STAT (signal transducers and activators of transcription).

Helper T cells

Cytokines play a major role in T cell development and regulation. Indeed, the two helper T cell subsets, Th1 and Th2, are defined by the cytokines they produce [*see 6:X Allergic Response*]. Th1 and Th2 cells play key roles in determining the balance between host resistance and immunopathology.[18] Th1 responses can help eradicate infectious agents, but a Th1-dominated response that is poorly effective or too prolonged can result in host damage. Th2 responses are primarily involved in allergic reactions, antibody production, and antibody class switching. They can limit potentially harmful Th1-mediated responses and may be part of the suppressor mechanism for exaggerated or inappropriate Th1 responses.[19]

Each of the two helper T cell subsets inhibits the development and function of the other. Interferon gamma (IFN-γ) produced by Th1 cells inhibits the development and function of

Table 2 Selected Cytokines

Cytokine	*Sources*	*Principal Functions*	*Comments*
Interferon-γ (IFN-γ)	T cells, NK cells	Primary macrophage-activating factor (MAF)	Used to treat chronic granulomatous disease, drug-resistant leishmaniasis
Interleukin-1	Monocytes, macrophages, other cells	MAF, endogenous pyrogen	IL-1 inhibitors counteract rheumatoid arthritis and other inflammatory processes
Interleukin-2	Th1 cells	Activation of lymphocytes, NK cells; MAF	Used in cancer therapy
Interleukin-4	Th1 cells, other cells	Activation of lymphocytes, monocytes; B cell class switching	—
Interleukin-5	Th2 cells, activated mast cells	Eosinophil recruitment and activation	—
Interleukin-6	Th1 cells, macrophages	Activates T cells and macrophages, promotes inflammation	—
Interleukin-8	T cells, macrophages	Activates neutrophils	—
Interleukin-10	T cells, B cells, macrophages, activated mast cells, keratinocytes	Suppresses lymphocyte responses by downregulating macrophage cytokines	—
Interleukin-12	Macrophages	Stimulates development of Th1 cells, stimulates production of IFN-γ by Th1 and NK cells	Possible use as vaccine adjuvant
Interleukin-18	Macrophages	Stimulates production of IFN-γ	—
Lymphotoxin	Th1 cells	Lyses tumor cells, activates neutrophils, increases vascular adhesion and extravasation of leukocytes	—
Migration inhibitory factor	Macrophages, T cells	Counterregulates glucocorticoid action	—
Transforming growth factor–β	Platelets, lymphocytes, activated macrophages, placenta cells, others	Anti-inflammatory; inhibits activation of macrophages and maturation of cytotoxic T cells	—
Tumor necrosis factor–α	Macrophages, activated T cells, NK cells, mast cells	Enhances protective inflammatory response	Antibodies to TNF-α are used in the treatment of Crohn disease and rheumatoid arthritis

NK—natural killer Th1—type 1 helper T cells Th2—type 2 helper T cells

Th2 cells, whereas IL-4 and IL-10 produced by Th2 cells inhibit the development and function of Th1 cells. IL-4 acts partly by downregulating the expression of the IL-12 receptor, IL-12Rβ, which is upregulated by IFN-γ [*see Figure 4*]. IL-12 enhances cell function because it is a potent growth factor for NK cells, which also produce IFN-γ. Presumably, the reason Th1 and Th2 cells inhibit each other is that both subsets also induce inflammation, which must be regulated [*see* Inflammatory Cytokines *and* Anti-inflammatory Cytokines, *below*].

Certain microorganisms may affect the relative balance of Th1 and Th2. For example, the parasite *Schistosoma* drives a strong Th2 immune response. A carbohydrate found on the surface of *Schistosoma,* lacto-*N*-fucopentaose-III (LNFP-III), induces expansion of B_1 cells and secretion of large amounts of IL-10. Acting through macrophages, IL-10 both blocks development of Th1 cells and favors Th2 responses.[20,21] A related carbohydrate, lacto-*N*-neotetraose (LNnT)—which is found on other helminths, *Helicobacter pylori*, *Neisseria meningitidis,* and other microorganisms—has a mode of action similar to that of LNFP-III.[22] LNFP-III is also found on the surface of cancer cells, and LNnT is found in human milk, which suggests that in these circumstances, as well, the two carbohydrates are being used to shut off the host's protective cell-mediated immune response. Logically, such carbohydrates may have therapeutic value for the inhibition of Th1-initiated autoimmune diseases.

MEMORY T CELLS

On reexposure to an antigen, memory cells mediate a faster and stronger immune response than naive T cells. Memory cells secrete a full range of cytokines and may show the same selection of cytokines as Th1 and Th2 cells.[10] The requirements for proliferation and cytokine production in memory cells are not as strict as those for production in naive T cells, but optimum responses require costimulation.[23] Memory T cells selectively migrate (home) to specific nonlymphoid tissues such as gut, skin, and lung. Those that home to the gut have specialized integrins (adhesion molecules) on their surface that mediate this migration [*see 6:I Organs and Cells of the Immune System*]. In the absence of antigenic stimulation, memory cells appear to persist as nondividing cells. A reencounter with an antigen can expand the population to a stable, higher level; competition from another antigen can decrease the population.[23]

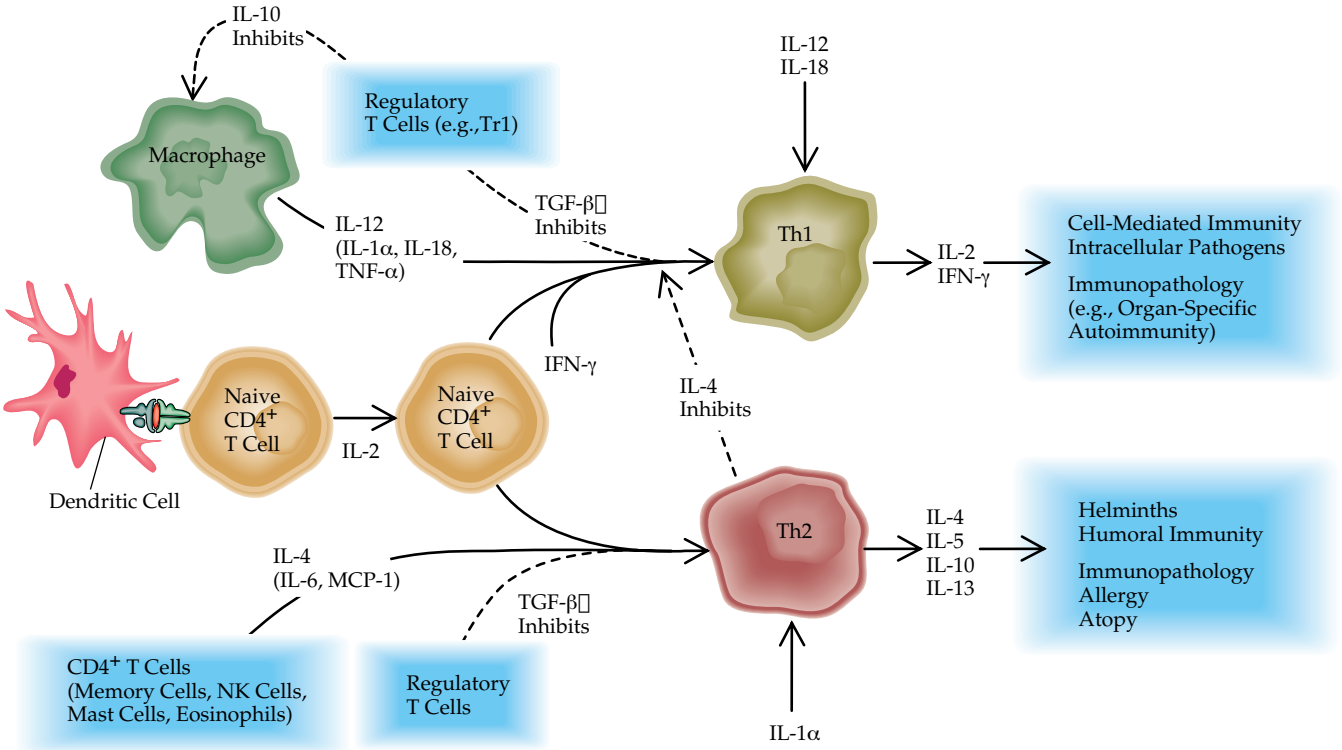


Figure 4 **Regulation of helper T cell responses. In response to IL-12 and cofactors such as IL-18 and IL-1a, naive $CD4^+$ T cells can develop into type 1 helper T (Th1) cells responsible for cell-mediated immunity; differentiation is dependent on interferon gamma (IFN-γ). Th1 cells produce IFN-γ and IL-2. Th1 responses are directly antagonized by IL-4 and indirectly by IL-10, as it inhibits production of IL-12 and IL-18 by macrophages. Th2 cells, which are responsible for inducing antibody production by B cells and allergic responses, depend on IL-4 for differentiation from naive $CD4^+$ T cells. Th2 cells produce IL-4, IL-5, IL-10, and IL-13. TGF-β can inhibit both Th1 and Th2 development. (MCP—monocyte chemoattractant protein; TGF—transforming growth factor)**

REGULATORY T CELLS

Natural regulatory T (Treg) cells are a population of T cells that primarily regulate responses to self antigens by other T cells and, therefore, contribute to the maintenance of self-tolerance. For example, removal of certain Treg cells from the periphery of normal mice leads to spontaneous development of various autoimmune and inflammatory diseases (e.g., autoimmune thyroiditis, gastritis, type 1 diabetes mellitus, and inflammatory bowel disease). Treg cells function via secreted or membrane-bound transforming growth factor–β (TGF-β) and by secretion of IL-10.[24]

Most Treg cells are $CD4^+$ and constitutively express CD25 on their surface. $CD4^+$,$CD25^+$ T cells are produced by the normal thymus as a functionally mature T cell subpopulation with a broad T cell repertoire; they can recognize both self and nonself antigens.[25] The *Foxp*-3 gene serves as a marker of this natural regulatory population. However, induction of Treg cells also takes place in the periphery (adaptive Treg cells).[26] Chronic activation of human $CD4^+$ T cells in the presence of IL-10 gives rise to a subset of antigen-specific $CD4^+$ T cell clones with low proliferative capacity that produce high levels of IL-10, low levels of IL-2, and no IL-4. These T cells, called type 1 regulatory T (Tr1) cells, suppress the proliferation of $CD4^+$ T cells in response to antigen, which suppresses antigen-specific immune responses and actively downregulates a pathologic immune response.[27]

INFLAMMATORY CYTOKINES

Interleukin-12 and Interleukin-18

IL-12 is a critical cytokine that stimulates the differentiation of naive helper T cells into Th1 cells, and it stimulates NK cells and Th1 cells to produce IFN-γ. IL-12 also enhances the cytolytic function of cytolytic T cells and NK cells. IL-12 is produced by activated macrophages and dendritic cells, in response to a variety of microorganisms. It has been used experimentally as an adjuvant in vaccines aimed at stimulating Th1-induced cellular immunity. As with most cytokines, binding of IL-12 with its receptor generates signaling through the Jak and STAT pathways.

IL-18, which is produced by macrophages, also stimulates the production of IFN-γ but appears to be less necessary than IL-12; whereas IL-12–deficient mice are susceptible to *Leishmania major* infection, IL-18–deficient mice combat such infection normally, although their initial immune response is slow.[28] Similarly, IL-12 is critical for immunity to cytomegalovirus, whereas IL-18 is not.[29]

Tumor Necrosis Factor–α

TNF-α, a major inflammatory cytokine, is one of the most abundant substances produced by macrophages after stimulation with IFN-γ, migration inhibitory factor (MIF), or bacterial lipopolysaccharide (LPS). TNF-α is also produced by activated T cells, NK cells, and mast cells.

At low concentrations, TNF-α enhances the protective inflammatory response, activating and enhancing the function of various leukocytes, including neutrophils, macrophages, and eosinophils. This can further stimulate macrophages to produce cytokines, including TNF-α itself, IL-1, IL-6, MIF, and a variety of chemotactic cytokines [*see* Chemokines, *below*]. TNF-α enhances expression of MHC class I molecules, potentiates cytotoxic T cell–induced cell lysis, functions as an endogenous pyrogen (i.e., induces fever, by direct and indirect actions on the brain), activates the clotting system and the production of acute-phase proteins by the liver, and can cause immunodeficiency through suppression of the bone marrow. When present for prolonged periods, TNF-α causes cachexia.

TNF-α plays a primary role in the host response to gram-negative bacteria. The LPS of these bacteria causes the release of MIF, which in turn enhances TNF-α production by macrophages. At low concentrations of LPS, TNF-α mediates a protective response. At high concentrations of LPS, however, TNF-α mediates disseminated intravascular coagulation—part of what is known as the Shwartzman reaction—and can cause death from shock.

Protective immunity to certain intracellular organisms, such as *Leishmania,* is enhanced by TNF-α, and TNF-α also has potent antiviral activity. However, many of the symptoms of malaria, especially of cerebral malaria, and some symptoms of HIV infection may be mediated by TNF-α. Antibodies to TNF-α have been approved for the treatment of Crohn disease and rheumatoid arthritis [*see 4:IV Inflammatory Bowel Diseases and 15:II Rheumatoid Arthritis*].[30]

Lymphotoxin

Produced exclusively by Th1 cells, lymphotoxin has many of the same biologic properties as TNF-α and utilizes the same cell receptor as TNF-α; it is also referred to as TNF-β. Like TNF-α, lymphotoxin lyses tumor cells but not normal cells, activates neutrophils, and increases vascular adhesion and extravasation of leukocytes. In addition, lymphotoxin plays a role in the development of lymphoid tissue.

Interleukin-1 and Interleukin-6

IL-1 is produced mainly by monocytes and macrophages but also by other cells, such as epithelial and endothelial cells. It is an endogenous pyrogen, and many of its functions are similar to those of TNF-α. It induces the production of additional IL-1 and of IL-6 from macrophages and induces glucocorticoid synthesis and the release of prostaglandin, collagenase, and acute-phase proteins. IL-1 increases the expression of surface molecules on endothelial cells, leading to adhesion of leukocytes and coagulation, and stimulates the production of macrophage chemokines that in turn activate neutrophils. IL-1 differs from TNF-α in that it does not produce necrosis of tumors or tissue injury, increase expression of MHC, or, by itself, mediate the Shwartzman reaction.

Macrophages produce an IL-1 receptor antagonist (IL-1ra) that, along with the IL-1 receptors shed from activated cells, inhibits IL-1 and thus acts as a regulator. Such natural inhibitors to IL-1 are now in clinical use to counteract certain inflammatory processes, especially in rheumatoid arthritis.

IL-6 is induced by IL-1 and by TNF-α from macrophages and, in turn, inhibits macrophage production of IL-1 and TNF-α. Like IL-1 and TNF-α, IL-6 is an endogenous pyrogen. IL-6 acts on hepatic cells to produce acute-phase proteins, such as fibrinogen, α_2-macroglobulin, and serum amyloid A protein. This cytokine can also inhibit macrophage activation.

Interferon Gamma and Other Macrophage-Activating Factors

IFN-γ, produced by T cells and NK cells, is the primary macrophage-activating factor (MAF). MAFs play an important role in cell-mediated immunity because activated macrophages produce many cytokines and chemokines intimately involved in inflammation, including TNF-α, IL-1, IL-6, and MIF. Other MAFs include granulocyte-macrophage colony-stimulating factor (GM-CSF) and MIF. IL-1 and TNF-α have weak MAF activity. IL-12 stimulates NK cells to produce greater amounts of IFN-γ, enhancing IFN-γ–dependent reactions. Both on its own and by enhancing the effects of TNF-α, IFN-γ causes the expression of adhesion molecules on the surface of vascular endothelial cells, leading to T cell adhesion and extravasation.

The inflammatory effects of IFN-γ are countered by TGF-β and IL-10, which inhibit macrophage activation. IFN-γ has been used successfully to treat chronic granulomatous disease and drug-resistant visceral leishmaniasis.[31]

Migration Inhibitory Factor

MIF, the first T cell cytokine to be discovered, derives its name from the fact that it inhibits the random migration of macrophages in vitro. MIF acts as an endogenous hormone that counterregulates glucocorticoid action.[32] Macrophages and T cells release MIF in response to glucocorticoids and other inflammatory stimuli. MIF then overrides the immunosuppressive effects of steroids on macrophage and T cell cytokine production.[32,33]

The gene for MIF is expressed in many different tissues.[34] Large quantities of the gene are found in macrophages and pituitary cells. Indeed, LPS stimulates the release of MIF by the pituitary. When given to mice, MIF greatly enhances the lethality of LPS; conversely, anti-MIF antibodies completely reverse the lethality of LPS.[35] MIF upregulates the receptor for LPS on macrophages (Toll-like receptor 4). Recombinant MIF activates macrophages to kill *Leishmania* and stimulates macrophages to produce TNF-α and nitric oxide. Mice lacking the gene for MIF show enhanced resistance to the lethal effects of high doses of LPS and *Staphylococcus aureus* enterotoxin B, as well as to *Pseudomonas aeruginosa* and *Escherichia coli*, but they are susceptible to *Leishmania, Salmonella,* and *Cysticercosis*.[36-39] MIF has been shown to play a pathogenic role in several experimental models of inflammation and autoimmunity, including glomerulonephritis, arthritis, inflammatory bowel disease, atherogenesis, and acute lung injury; it also inhibits p53 and enhances carcinogenesis. Mechanisms of MIF action include the activation of the extracellular signal-regulated kinase–1 and –2 (ERK-1 and -2), leading to activation of phospholipase A_2, cyclooxygenase-2, and prostaglandin E_2; enhancement of Toll-like receptor–4; and inhibition of p53 and apoptosis.

Anti-MIF therapies are under development. The goal is to increase the immunosuppressive and anti-inflammatory properties of endogenously released glucocorticoids, thereby reducing the need for steroid therapy in a variety of autoimmune and inflammatory conditions.[32] Anti-MIF therapy should also have a role in treating some gram-negative infections and preventing septic shock.

Interleukin-5

IL-5 mainly affects eosinophil recruitment and activation. IL-5 is produced by Th2 cells and activated mast cells and stimulates the growth and differentiation of eosinophils. In addition to IL-5, other substances involved in the activation of eosino-

phils are TNF-α and an eosinophil cytotoxicity–enhancing factor derived from monocytes. Activated eosinophils produce tissue damage in allergic states and kill helminthic parasites.

Other Interleukins

DNA sequence information from the Human Genome Project has led to the identification of a number of new cytokines: IL-19 through IL-24. Their role in the immune response needs further exploration. Several of these cytokines appear to have some properties similar to those of IL-10, IL-12, and IL-15. Hematopoietic cytokines and growth factors are discussed elsewhere [*see V:I Approach to Hematologic Disorders*].

ANTI-INFLAMMATORY CYTOKINES

Interleukin-4 and Interleukin-13

IL-4 stimulates the expression of an adhesion molecule on endothelial cells, leading to the binding of eosinophils, lymphocytes, neutrophils, and monocytes and their subsequent extravasation. However, IL-4 also acts as an anti-inflammatory cytokine, inhibiting activated macrophages and diminishing the production of TNF-α and nitric oxide. IL-13 has the ability to take over some of the functions of IL-4. Both IL-4 and IL-13 induce IgE synthesis in B cells and differentiation of T cells to Th2 cells and can suppress inflammatory processes induced by Th1 cells. These cytokines also act on macrophages to suppress the inflammatory response. Activated mast cells and basophils produce additional IL-4. Of interest is that mutated IL-4 can inhibit IgE synthesis by IL-4 and IL-13 and may prove useful in treating some allergic states.

Transforming Growth Factor–β

TGF-β is produced by a variety of cells, including platelets, lymphocytes, activated macrophages, and placenta cells. It is an important anti-inflammatory cytokine because it inhibits the activation of macrophages and the maturation of cytotoxic T cells and thus controls the effects of many cytokines.

Interleukin-10

IL-10 is an important regulatory cytokine. It is produced by $CD4^+$ and $CD8^+$ T cells, B cells, macrophages, activated mast cells, and keratinocytes. Although usually associated with activity of Th2 cells, IL-10 can also be produced by Th1 cells. IL-10 suppresses lymphocyte responses by downregulating macrophage cytokines—including IL-1, TNF-α, IL-6, IL-8, GM-CSF, and granulocyte colony-stimulating factor (G-CSF)—and inhibiting nitric oxide production.

Chemokines

Chemokines are a superfamily of low-molecular-weight chemotactic cytokines that mediate the directional migration of leukocytes during normal and inflammatory processes.[40] They play an important role in attracting granulocytes into sites of inflammation. There are four distinct families of chemokines, distinguished on the basis of the position of their first two conserved cysteine residues: CXC (the first two cysteines are separated by one amino acid), CC, C, and CX3C. The receptors for chemokines are all integral membrane G-protein–coupled receptors, which constitute one of the largest classes of signaling molecules.

CXC chemokines predominantly activate neutrophils.[41] This family includes IL-8; β-thromboglobulin (β-TG); the growth-related gene products gro-α, gro-β, and gro-γ; and platelet factor 4. They are usually produced by monocytes, but some are produced by other cells, including T cells, endothelial cells, and platelets. IL-8 induces expression of neutrophil-binding integrins on endothelial cells, resulting in the rapid accumulation of neutrophils in tissues. The chemokine gro also stimulates neutrophil accumulation, as well as the release of lysosomal enzymes that contribute to the local inflammatory response. Platelet factor 4 and β-TG are released from aggregated platelets and stimulate fibroblasts, which are required for repair at sites of hemorrhage or thrombosis.

The CC chemokines activate T cells, monocytes, and eosinophils. This family includes RANTES (regulated on activation, normal T cell expressed and secreted), macrophage chemotactic and activating factor (MCAF), macrophage inflammatory protein–1α (MIP-1α), and MIP-1β. CC chemokines are produced by activated T cells and monocytes. RANTES is a potent attractant for memory T cells (but not for naive T cells) and also attracts monocytes. MCAF acts exclusively on monocytes, attracting them, activating them, and regulating the expression of integrins on their surface. MIP-1α and MIP-1β attract only monocytes. The CC chemokines eotaxin, eotaxin-2, and monocyte chemoattractant protein–4 (MCP-4) predominantly activate eosinophils.[41]

In addition to their role in inflammation, chemokines are important in the hemostasis of lymphocytes moving through the lymphatic system; in the location of T cells, B cells, and dendritic cells in the lymph node; in Th1 and Th2 cell development; and in angiogenesis, angiostasis, and metastasis of tumor. In fetal mice lacking the CXC chemokine receptor–4, the heart and cerebellum do not develop properly, indicating that chemokines also play a part in nonlymphoid organ development.[42,43]

Chemokine receptors play an important role as coreceptors for HIV. The virus first interacts with CD4 on T cells but requires a coreceptor to penetrate the cell membrane. The CC chemokine receptor–5 (CCR5), which mediates activation of T cells and macrophages, is the major coreceptor for some HIV-1 strains. The natural ligands for CCR5 include RANTES, MIP-1α, and MIP-1β.[44] The CXC chemokine receptor–4 appears to be important in late-stage HIV infection.[45]

Effector Mechanisms in Cell-Mediated Immunity

Cell-mediated immunity encompasses the killing of invading microorganisms, such as bacteria, viruses, fungi, and parasites; the destruction of tumor cells; the rejection of tissue grafts; and injury to tissues in various disease states, including autoimmunity. Cell-mediated immune reactions can also be induced by contact with antigens, such as those found in poison ivy and numerous drugs. Drugs are more likely to provoke cell-mediated reactions when applied topically than when given systemically.

Most cell-mediated immune reactions involve initial interaction between sensitized T cells and antigens on presenting cells. This reaction can trigger several effector pathways, including activation of cytotoxic T cells, stimulation of T cell production of cytokines that activate macrophages and promote the proliferation of NK cells, and production of antibodies involved in antibody-dependent cell-mediated cytotoxicity by NK cells and other cell types. Although cell-mediated immune reactions other than antibody-dependent cell-mediated cytotoxicity do

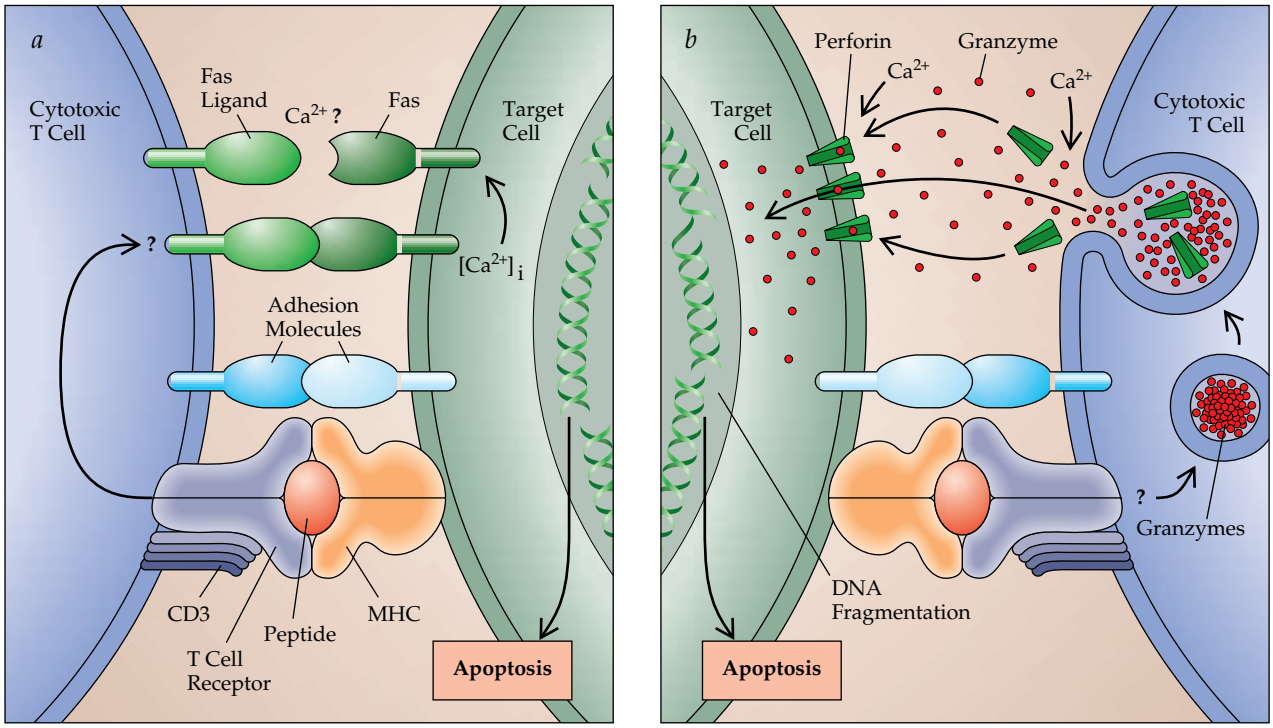


Figure 5 **Cytotoxic cells recognize surface markers on cells that are to be destroyed. (*a*) Apoptosis is triggered by the cytotoxic T cell through nonsecretory Fas–Fas ligand interaction. (*b*) Apoptosis is triggered by the cytotoxic T cell by means of secretory mechanisms initiated by perforin and granzymes.**

not require the presence of antibody or complement, they can be modified by these humoral factors. Subsequent events require cooperation between different subsets of T cells; the reactions involved are controlled by various cytokines.

The mechanisms of cell-mediated immunity involving T cell–macrophage interactions can be both protective (leading to the killing of invading microorganisms) and harmful (leading to inflammation and tissue destruction). Sometimes, the two go hand in hand; in tuberculosis, for example, both the killing of tubercle bacilli and the development of cavities in the lungs are consequences of T cell–macrophage interactions. In addition to the acquired cell-mediated immunity discussed above, innate immunity also involves the mounting of an immune response by cells directly stimulated by microorganisms. Macrophages, other granulocytic cells, and NK cells are involved in innate cell-mediated immune responses [*see 6:II Innate Immunity*].

CYTOTOXIC T CELLS

Cytotoxic T cells are antigen-specific effector cells that are important in resisting infectious agents, especially viruses that are present in cells other than macrophages; in killing tumors; and in allograft rejection. Most cytotoxic T cells are $CD8^+$ T cells that recognize antigen presented by MHC class I molecules, although a considerable number of $CD4^+$ T cells have the capability to kill target cells. Killing by a cytotoxic T cell begins with adhesion to the target cell (which requires magnesium ions), followed by the delivery of cytotoxic chemicals to the target cell (which requires calcium ions). The cytotoxic T cell then dissociates from the target cell; death proceeds in the absence of the cytotoxic T cell, which recycles to attack another target cell. If the cytotoxic T cell adheres to a cell that does not carry the targeted antigenic peptide–MHC molecule combination, no cytotoxic chemicals are released and the cells dissociate more rapidly.

Cytotoxic T cells develop granules that contain cytotoxic molecules, including perforins (proteins that produce holes or pores in a cell's surface membrane), serine proteases (granzyme A and granzyme B), and serine esterases.[46] Of these, perforins are the most important, as has been shown in mice in which the gene encoding perforin has been deleted. A second killing mechanism involves the Fas ligand on the cytotoxic T cell and Fas on the target cell. Binding of these leads to apoptosis of the target cell. This is the only mechanism of killing available to mice that lack perforin, and it is used preferentially—but not exclusively—in $CD4^+$ cytotoxic T cells [*see Figure 5*].

Viral infection results in the production of a large number of virus-specific cytotoxic T cells. This is most dramatically shown during the initial responses to B cells infected with Epstein-Barr virus. Cytotoxic T cell clones specific for some antigen-MHC complexes are extremely abundant, constituting approximately 50% of all cytotoxic T cells.[9] When the cytotoxic T cell response diminishes, these abundant T cell clones are probably removed through apoptotic mechanisms.

ACTIVATED MACROPHAGES

Macrophages are usually activated by Th1 cells that have been stimulated by antigens or microorganisms. Those Th1 cells then express CD40 ligand (CD40L) and produce macrophage-activating cytokines, especially IFN-γ and MIF. These cytokines, in combination with CD40L interacting with the CD40 on the macrophage surfaces, induces intracellular signaling transcription pathways in the macrophage. These pathways result in activation of transcription factors leading to production of various proteins and surface markers that characterize the activated macrophage. Activated macrophages produce reactive oxygen intermediates, including nitric oxide, that are involved in the destruction of microorganisms or foreign cells.

Bacterial killing also involves phagolysosomal fusion, which mobilizes enzymes such as cathepsins.

Activated macrophages produce many of the inflammatory cytokines (e.g., IL-12, TNF-α, IL-1, and MIF) and chemokines (e.g., MIP-1) that are involved in immunity to microorganisms and foreign antigens and in enhancing the process of activation itself. Activated macrophages also express MHC class II and costimulatory molecules, which further amplify the process. In addition, activated macrophages produce the cytokines IL-10 and TGF-β; these counteract the activation, damping down the inflammatory process and acting as feedback regulators [*see 6:I Organs and Cells of the Immune System*].

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Davis MM, Boniface JJ, Reich Z, et al: Ligand recognition by alpha beta T cell receptors. Annu Rev Immunol 16:523, 1998

2. Kronenberg M, Gapin L: The unconventional lifestyle of NKT cells. Nat Rev Immunol 2:557, 2002

3. Park SH, Chiu YH, Jayawardena J, et al: Innate and adaptive functions of the CD1 pathway of antigen presentation. Semin Immunol 10:391, 1998

4. Treiner E, Duban L, Bahram S, et al: Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature 422:164, 2003

5. Jardetzky TS, Brown JH, Gorga JC, et al: Three-dimensional structure of a human class II histocompatibility molecule complexed with superantigen. Nature 368:711, 1994

6. Banchereau J, Steinman RM: Dendritic cells and the control of immunity. Nature 392: 245, 1998

7. Viola A, Schroeder S, Sakakibara Y, et al: T lymphocyte costimulation mediated by reorganization of membrane microdomains. Science 283:680, 1999

8. Altman JD, Moss PAH, Goulder PJR, et al: Phenotypic analysis of antigen-specific T lymphocytes. Science 274:94, 1996

9. McMichael AJ, O'Callaghan CA: A new look at T cells. J Exp Med 187:1367, 1998

10. Van Parijs L, Abbas AK: Homeostasis and self-tolerance in the immune system: turning lymphocytes off. Science 280:243, 1998

11. Ward SG: The complexities of CD28 and CTLA-4 signalling: PI3K and beyond. Arch Immunol Ther Exp (Warsz) 47:69, 1999

12. Sharpe AH, Freeman GJ: The B7-CD28 superfamily. Nat Rev Immunol 2:116, 2002

13. Sayos J, Wu C, Morra M, et al: The X-linked lymphoproliferative disease gene product SAP regulates signals induced through the co-receptor SLAM. Nature 395:462, 1998

14. Engel P, Eck MJ, Terhorst C: The *SAP* and *SLAM* gene families in immune responses and X-linked lymphoproliferative disease. Nat Rev Immunol 3:813, 2003

15. Bolen JB, Brugge JS: Leukocyte protein tyrosine kinases: potential targets for drug discovery. Annu Rev Immunol 15:371, 1997

16. Rao A, Luo C, Hogan PG: Transcription factors of the NFAT family: regulation and function. Annu Rev Immunol 15:707, 1997

17. O'Garra A: Cytokines induce the development of functionally heterogeneous T helper cell subsets. Immunity 8:275, 1998

18. Sher A, Gazinelli RT, Jankovic D, et al: Cytokines as determinants of disease and disease interactions. Braz J Med Biol Res 31:85, 1998

19. Granger DN: Cell adhesion and migration: II. Leukocyte-endothelial cell adhesion in the digestive system. Am J Physiol 275:G982, 1997

20. Li S, Paulsson KM, Sjogren HO, et al: Peptide-bound major histocompatibility complex class I molecules with tapasin before dissociation from transporter associated with peptide processing. J Biol Chem 274:8649, 1999

21. Faveeuw C, Mallevaey T, Paschinger K, et al: Schistosome N-glycans containing core alpha 3-fucose and core beta 2-xylose epitopes are strong inducers of Th2 responses in mice. Eur J Immunol 33:1271, 2003

22. Terrazas LI, Walsh KL, Piskorska D, et al: The schistosome oligosaccharide lacto-N-neotetraose expands Gr1(+) cells that secrete anti-inflammatory cytokines and inhibit proliferation of naive CD4(+) cells: a potential mechanism for immune polarization in helminth infections. J Immunol 167: 5294, 2001

23. Dutton RW, Bradley LM, Swain SL: T cell memory. Annu Rev Immunol 16:201, 1998

24. Wood KJ, Sakaguchi S: Regulatory T cells in transplantation tolerance. Nat Rev Immunol 3:199, 2003

25. Jordan MS, Boesteanu A, Reed AJ, et al: Thymic selection of $CD4^+CD25^+$ regulatory T cells induced by an agonist self-peptide. Nat Immunol 2:301, 2001

26. Hori S, Nomura T, Sakaguchi S: Control of regulatory T cell development by the transcription factor Foxp3. Science 299:1057, 2003

27. Groux H, O'Garra A, Bigler M, et al: A CD4+ T-cell subset inhibits antigen-specific T-cell responses and prevents colitis. Nature 389:737, 1997

28. Pien GC, Satoskar AR, Takeda K, et al: Cutting edge: selective IL-18 requirements for induction of compartmental IFN-gamma responses during viral infection. J Immunol 165:4787, 2000

29. Monteforte GM, Takeda K, Rodriguez-Sosa M, et al: Genetically resistant mice lacking IL-18 gene develop Th1 response and control cutaneous *Leishmania* major infection. J Immunol 164:5890, 2000

30. Baert FJ, Rutgeerts PR: Anti-TNF strategies in Crohn's disease: mechanisms, clinical effects, indications. Int J Colorectal Dis 14:47, 1999

31. Locksley RM, Fowell DJ, Shinkai K, et al: Development of CD4+ effector T cells and susceptibility to infectious diseases. Adv Exp Med Biol 452:45, 1998

32. Bucala R: Neuroimmunomodulation by macrophage migration inhibitory factor (MIF). Ann NY Acad Sci 840:74, 1998

33. Calandra T, Roger T: Macrophage migration inhibitory factor: a regulator of innate immunity. Nat Rev Immunol 3:791, 2003

34. Lanahan A, Williams JB, Sanders LK, et al: Growth factor-induced delayed early response genes. Mol Cell Biol 12:3919, 1992

35. Bernhagen J, Calandra T, Mitchell RA, et al: MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia. Nature 365:756, 1993

36. Bozza M, Satoskar AR, Lin G, et al: Targeted disruption of migration inhibitory factor gene reveals its critical role in sepsis. J Exp Med 189:341, 1999

37. Satoskar AR, Stamm LM, Zhang XM, et al: Mice lacking natural killer (NK) cells develop an efficient Th-1 response and control cutaneous *L. major* infection. J Immunol 161:6747, 1999

38. Koebernick H, Grode L, David JR, et al: Macrophage migration inhibitory factor (MIF) plays a pivotal role in immunity against *Salmonella typhimurium*. Proc Natl Acad Sci USA 99:13681, 2002

39. Rodríguez-Sosa M, Rosas LE, David JR, et al: Macrophage migration inhibitory factor plays a critical role in mediating protection against the helminth parasite *Taenia crassiceps*. Infect Immun 71:1247, 2003

40. Baggiolini M, Dewald B, Moser B: Human chemokines: an update. Annu Rev Immunol 15:675, 1997

41. Petering H, Gotze O, Kimmig D, et al: The biologic role of interleukin-8: functional analysis and expression of CXCR1 and CXCR2 on human eosinophils. Blood 93:694, 1999

42. Rossi D, Zlotnik A: The biology of chemokines and their receptors. Annu Rev Immunol 18:217, 2000

43. Proudfoot AE: Chemokine receptors: multifaceted therapeutic targets. Nat Rev Immunol 2:106, 2002

44. Olbrich H, Proudfoot AE, Opperman M: Chemokine-induced phosphorylation of CC chemokine receptor 5 (CCR5). J Leukoc Biol 65:281, 1999

45. Choe H: Chemokine receptors in HIV-1 and SIV infection. Arch Pharm Res 21:634, 1998

46. Darmon AJ, Bleackley RC: Proteases and cell-mediated cytotoxicity. Crit Rev Immunol 18:255, 1998

Acknowledgments

Figures 1 through 4 Dimitry Schidlovsky.

Figure 5 Seward Hung.

V ADAPTIVE IMMUNITY: HISTOCOMPATIBILITY ANTIGENS AND IMMUNE RESPONSE GENES

Edgar L. Milford, m.d.
Charles B. Carpenter, m.d.

The major histocompatibility complex (MHC) was first appreciated in mice as a set of proteins, encoded by closely linked genes on chromosome 17, that serve as the major targets for rejection of skin grafts. Humans were subsequently shown also to have MHC antigens, which are homologous to those found in the mouse but are encoded in the human leukocyte antigen (*HLA*) region on the short arm of chromosome 6 [*see Figure 1*]. Initially, human MHC antigens could be defined only by use of sera from multiparous women who had mounted humoral immune responses against the paternally derived MHC antigens in their fetuses. The development of DNA-based methods for genotyping of individuals has permitted more extensive study of these extraordinarily polymorphic molecules. This chapter reviews the genetics and structure of the MHC, its function in the immune response, and its association with disease.

Structure and Antigens of the Major Histocompatibility Complex

There are two structural types of MHC molecules, called class I and class II. The molecules of both classes are active in antigen recognition and help focus immune defenses during invasions from the microbial world. They are also engaged in the communication that occurs between cells during the immune response. MHC molecules act by binding peptide fragments of antigens that have been processed in specialized antigen-presenting cells. Clonally determined antigen receptors on T cells then recognize and bind to specific peptide-MHC complexes, setting into motion the appropriate immune response. Segments of MHC molecules show sequence homologies with immunoglobulins, T cell antigen receptors, and T cell interaction molecules such as CD4 and CD8, which suggests that all these molecules share a common evolutionary ancestry.

The sequence and structure of MHC molecules have been extensively elucidated, and it has been determined that the polymorphic, or antigenic, portions of MHC molecules are quite small. In fact, the polymorphic portions frequently comprise only one to four amino acid substitutions encoded in regions of DNA nucleotide sequence hypervariability. A specific configuration in an MHC molecule resulting from particular substitutions of amino acids is called an epitope.

MHC class I antigens

MHC class I antigens consist of two polypeptide chains held together noncovalently. One chain is heavy (44 kd) and glycosylated, and it determines antigen specificity. The extracellular portion of this class I heavy chain is divided into three domains, des-

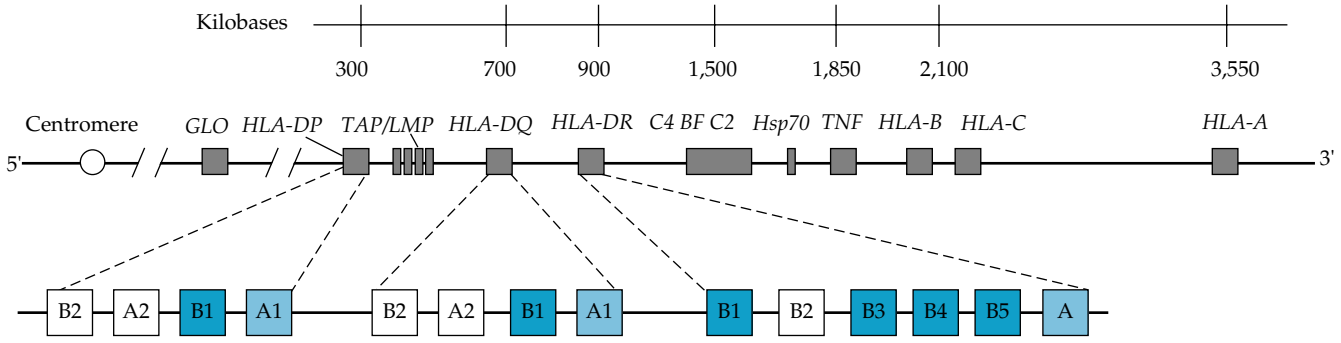


Figure 1 **The best-characterized loci of the human major histocompatibility complex (MHC), located in the *HLA* region of the short arm of chromosome 6, are depicted. Distances are shown in recombination units (centimorgans), as determined by crossover frequencies in family studies, and in kilobases, as determined by sequence analysis of fragments produced by DNAses having defined cleavage sites. MHC class II molecules are encoded in the *HLA-DP*, *HLA-DQ*, and *HLA-DR* genes, and MHC class I molecules are encoded by *HLA-B*, *HLA-C*, and *HLA-A* genes. A cluster of closely linked complement genes—*C4*, *BF*, and *C2*—lies in the center of the region. There are two structural genes for *C4*, interspersed with two genes for the adrenal enzyme 21-hydroxylase. Next is the heat shock protein gene, *Hsp70*, followed by the tumor necrosis factor (TNF) genes, *A* and *B*. The orientation of the complement cluster and the *TNF* cluster has not been established, but an expanded view of this area could be depicted as –(*C4-210HA-C4B-210HB-BF-C2*)–(*HSP70*)–(*TNFA-TNFB*). *GLO* is a marker gene for the enzyme glyoxylase. An expansion of the class II region is in the lower portion of the figure. Each class II molecule is a heterodimer of an α and a β chain, which are encoded in the *A* and *B* genes, respectively. Pseudogenes, which are not expressed on the cell surface, are shown in white boxes. *HLA-DP* and *HLA-DQ* have one expressed heterodimer, A1B1; *HLA-DR* has only one A chain but nine genes for B chains (four are shown in the figure). The principal expressed heterodimers for *HLA-DR* are AB1, AB3, AB4, and AB5. In the region between *HLA-DP* and *HLA-DQ* lie the closely linked *TAP1*, *TAP2*, *LMP2*, and *LMP7* genes. The *TAP* genes encode peptide transporters, whereas the *LMP* genes encode proteosomes that fragment proteins into peptides. This cytoplasmic system is believed to be responsible for production and delivery of peptides to MHC class I molecules before their movement to the cell surface.**

ignated α_1, α_2, and α_3. The other chain is a small (11.5 kd) protein known as β_2-microglobulin [*see Figure 2*]. Class I heavy chains are the gene products of three MHC loci, designated *HLA-A, HLA-B,* and *HLA-C* [*see Table 1*]. There are many alleles for each locus; therefore, considerable polymorphism exists. β_2-Microglobulin is encoded by a gene on chromosome 15. Both the β_2-microglobulin and the α_3 domain of the heavy glycosylated chain of MHC class I antigens demonstrate considerable structural similarity to the constant region of the heavy chain of IgG (C_H3).

MHC class I molecules have been crystallized, and their structure has been determined by x-ray diffraction to a resolution of 3.5 angstroms (Å).[1] Two of the heavy-chain domains, α_1 and α_2, are located at the membrane-distal portion of the heavy chain and form a groove along the top surface of the molecule. The sides of the groove are composed of α helices from the α_1 and α_2 domains, and the base is composed of eight antiparallel β-pleated sheets from these domains. The hypervariable (antigenic) regions are found mostly along the sides of the groove, but there is also variability in the β-pleated sheet region. The rest of the molecule shows minimal variability in relation to other molecules of the same *HLA* locus. In the crystals studied, the groove, which faces away from the cell membrane and is approximately 25 Å long and 10 Å wide, contains material representing processed antigen (i.e., peptide fragments). When peptides eluted from purified class I molecules are sequenced, they show patterns of amino acids, called motifs, that bind to particular sets of HLA class I molecules.[2] These findings helped confirm the hypothesis that MHC molecules bind and present processed antigens to responding T cells and that the T cell receptor (TCR) recognizes foreign antigen as a peptide in the context of self-antigen; that is, it binds to a surface composed of both MHC and a bound peptide.

MHC class I antigens can be expressed on all cell types except erythrocytes and trophoblasts and can be detected by staining with labeled antibodies. Striated muscle cells and liver parenchymal cells are normally negative for class I antigens (i.e., they lack class I molecules or express only a low density of class I molecules), but in inflammatory states, these cells may become strongly positive for class I antigens.

MHC CLASS II ANTIGENS

Some antibodies, elicited by immunizations with histoincompatible cells, react with a limited variety of cells, most notably B cells, monocytes, dendritic cells, and activated T cells. Normally, these cells are the only ones found to bear MHC class II antigens. As is the case with class I antigens, however, inflammatory states cause many tissues to express class II antigens.

Each MHC class II antigen consists of two membrane-inserted glycosylated polypeptides, designated α (34 kd) and β (28 kd), which are bound together noncovalently [*see Figure 2*]. The extracellular portion of the α chain is divided into two domains, designated α1 and α2; the extracellular portion of the β chain is also divided into two domains, β_1 and β_2. Class II antigens are encoded by the *HLA-D* region, which is divided into at least three subregions: *HLA-DP, HLA-DQ,* and *HLA-DR* [*see Figure 1*].

Crystallographic studies indicate that MHC class II molecules have a structure similar to that of MHC class I molecules, with the α_1 and β_1 domains forming a groove in which β-pleated sheets form the base and α helices form the sides.[3] As in MHC class I molecules, the hypervariable (antigenic) regions of MHC class II molecules are located primarily along the groove, which again indicates a molecular basis for TCR recognition of foreign antigen together with self-MHC.

Class II MHC antigens can be identified by the use of sera from multiparous women that react predominantly with B cells. A serum is first exposed to platelets from a pool of many persons, because platelets contain MHC class I, but not MHC class II, antigens and thus will absorb antibodies to class I antigens, leaving antibodies to class II antigens in the serum. The naming of genes from the HLA-D region is now based on knowledge of the biochemistry of expressed antigens and on a growing database of DNA nucleotide sequences. The gene encoding the HLA-DR α chain, for example, is called *DRA*. Similarly, the closely linked genes encoding the β chains have been named *DRB1* (encoding the β chains for DR1 through DR18), *DRB3* (encoding the β chain for DR52), *DRB4* (encoding the β chain for DR53), and *DRB5* (encoding the β chain for DR51). Because *DRB2* expresses no protein product, it is called a pseudogene. Each of the HLA-DR β chains associates with the common nonpolymorphic HLA-DRA α chain to form functional class II HLA-DR molecules. HLA-DRA α chains are always the same; the difference in *HLA-DR* antigenic alleles is accounted for by variations in the genes encoding the HLA-DR β chains. The *HLA-DQ* locus contains the genes *DQA1, DQB1, DQA2,* and *DQB2*. *DQA2* and *DQB2* are pseudogenes, whereas the products of *DQA1* and *DQB1*—that is, the α and β chains of HLA-DQ—are both polymorphic. *HLA-DP* gene organization is similar to that of *HLA-DQ* [*see Figure 1*].

NOMENCLATURE OF HLA ANTIGENS

The nomenclature of the *HLA* system is coordinated through the World Health Organization Nomenclature Committee for Factors of the HLA System.[4] The prefix for the gene name is *HLA*, followed by a hyphen, then a locus name (e.g., *DRB1, DQA2, C*). A specific allele (DNA sequence variant at a locus) is denoted by appending an asterisk to the gene name followed by a unique alphanumeric identifier for that allele. The alphanumeric identifier is composed of up to nine characters: AACCSSXXN, where AA is an integer that refers (when possible) to the serologic family of which the allele is a member; CC is an integer defining the nucleotide coding variant resulting in a unique peptide product; SS is an integer defining synonymous variants (different DNA sequence but same amino acid sequence) of a coding variant; XX is an integer defining variants outside the coding region. An N is appended to the identifier if the allele is a null or nonexpressed variant. Thus, *HLA-DRB1*030502* encodes a DR molecule that is serologically in the DR3 group and has a different amino acid sequence from *DRB1*0301, DRB1*0302, DRB1*0303,* or *DRB1*0304*. It also has a different nucleotide sequence from *DRB1*030501*, but has the same amino acid sequence, making it a synonymous variant.

FREQUENCY OF DIFFERENT HLA ALLELES

Two terms, haplotype and linkage disequilibrium, describe important associations between MHC genes. Haplotype refers to the set of closely linked genes on any one chromosome. Every person has two haplotypes of the MHC, one from each parent. Each haplotype has a particular set of antigens determined by the *HLA-A, HLA-B, HLA-C, HLA-DR*, and other loci.

The second term, linkage disequilibrium, refers to the observation that in a population, some HLA antigens coincide within a single haplotype much more frequently than expected. If discrete genes were distributed independently throughout the population, the frequency at which any two linked antigens encoded at different loci would occur within a haplotype is the product of their frequencies in the population. However, in whites, the

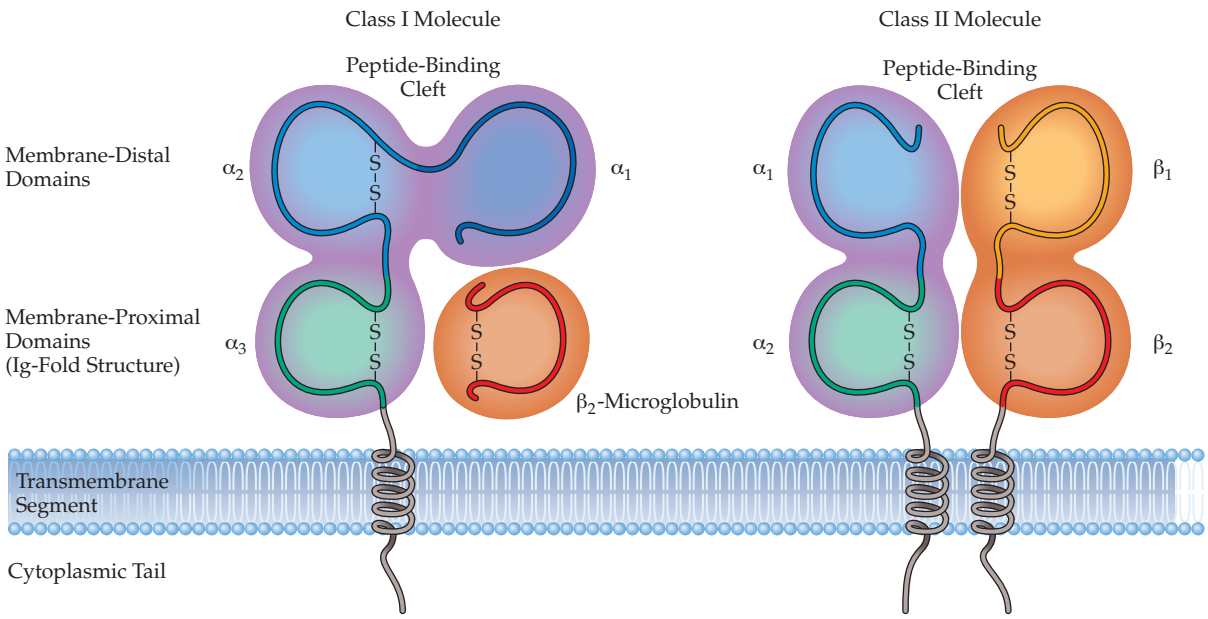


Figure 2 **MHC molecules are of two structural types with very similar peptide-binding sites on the membrane-distal surface. (*a*) MHC class I molecules consist of heavy chains made up of three polypeptide domains (α_1, α_2, α_3) and a noncovalently associated light chain, β_2-microglobulin. (*b*) MHC class II molecules are heterodimers of α and β chains with a very similar overall structure and peptide-binding surface.[20]**

HLA-A1 antigen and the HLA-B8 antigen are associated six to 21 times more often than would be predicted from their gene frequencies. Such linkage disequilibrium may occur because not enough evolutionary time has elapsed for the genes governing the antigens to be evenly distributed or because such an association results in a selective advantage to the individual. Recombination, or crossover, takes place during meiosis and occurs about 1.5% of the time between MHC class I and MHC class II loci. Over many generations, recombination leads to an equilibrium of linked alleles in a population unless selective pressures favor survival of certain haplotypes. A hypothetical example of such selection would be the survival of persons bearing HLA haplotypes that confer resistance to epidemics, such as smallpox and plague. Racial differences are reflected in marked variations in the frequencies of certain HLA antigens and haplotypes. Fewer, though often striking, examples of such differences are also observed in various ethnic groups.[5]

Role of MHC in Immune Response

THE MIXED LYMPHOCYTE REACTION

When lymphocytes from one person are cultured with those from another, the cells are stimulated to divide. This division, which can be measured from the rate of uptake of ^{3}H-thymidine into the cells, is called the mixed lymphocyte reaction (MLR). By preventing the division of one of the sets of cells by treatment with mitomycin or irradiation, it is possible to study the antigens on the membrane of the treated cells that stimulate this proliferative response. In humans, *HLA-DR* antigenic determinants are mainly responsible for evoking a primary MLR. HLA-DQ antigens play a lesser role, and HLA-DP antigens do not appear to be involved in the primary MLR. However, responding lymphocytes that have been primed by previous exposure to HLA-DQ or HLA-DP antigens proliferate vigorously when reexposed to the same antigen in a secondary MLR. The primary MLR is driven by the very high precursor frequency of naive cells having affinity to HLA-DRB1, not by primed memory cells.

ANTIGEN PROCESSING AND PRESENTATION

The breakdown of protein molecules into peptide fragments is an important part of the process by which antigens are presented to T cells and other immune effector cells. MHC molecules come to the cell surface with peptides already bound. Proteins are first degraded internally, and the peptide fragments are bound to MHC class I and MHC class II molecules within the cell. Class I molecules are expressed on virtually all tissues. Virally infected cells are recognized principally by class

Table 1 Antigens of the HLA System

	Antigen	*Number of Antigenic Specificities**	*Number of DNA Variants†*
Class I	HLA-A	22	290
	HLA-B	48	555
	HLA-C	9	140
Class II	HLA-DR (1-18)	17	356
	HLA-D	24 (cellular)	356
	HLA-DR (51-53)	3	76
	HLA-DQ	7	25 (DQA1)
			56 (DPB1)
	HLA-DP	6 (cellular)	20 (DPA1)
			106 (DPB1)

*Antigenic specificities are defined by reactivity with HLA-specific sera from multiparous women (serologic specificity) or by proliferative response of cells (cellular specificity).
†Variants at each locus; most variants result in a different expressed amino acid sequence.

Table 2 Cell-Mediated Lympholysis in a Mixed Culture

*Lymphocyte Antigens Matched between Responder and Stimulator Cells**		*Mixed Lymphocyte Culture Reaction*	*Cell-Mediated Lympholysis*
Class II	*Class I*		
Yes	Yes	No	No
No	No	Yes	Yes
No	Yes	Yes	No†
Yes	No	No	No
{ Yes / No }	{ No / Yes } ‡	Yes	Yes

*The stimulator cells that induce proliferation of the responder T cells in the mixed lymphocyte culture reaction also serve as the targets for the cytotoxic cells that develop from the responder population (as measured by the cell-mediated lympholysis assay).
†Low numbers of cytotoxic T cells may develop against class II antigens.
‡Stimulator cells from two individuals are mixed with responder cells from a single individual.

I–restricted T cells, usually those with a cytotoxic function. In contrast, class II–directed T cells are restricted to antigen-presenting cells of the immune system (i.e., B cells, macrophages, dendritic cells, or Langerhans cells) that are principally concerned with defense against external infectious agents. Because class II–positive cells also carry class I molecules, they may act as antigen-presenting cells for both exogenous and endogenous proteins [*see 6:IV Cell-Cell Interactions, Cytokines, and Chemokines in Immune Response Mechanisms*].

Exogenous and endogenous antigens reach the cell surface by different pathways. Exogenous proteins are taken up into endosomes or lysosomes, where they are catabolized. Peptides from exogenous proteins are generally bound to MHC class II molecules, and the class II–peptide complexes are then brought to the surface for presentation to T cells. Peptides from endogenous proteins (e.g., secretory proteins or products of viral infection) appear to be complexed in the endoplasmic reticulum to MHC class I molecules. Genes called *LMP*, which are also located in the MHC region, encode proteins that are responsible for breaking down proteins into small peptides (eight to 10 amino acids long); closely linked *TAP* genes encode chaperones that transport peptides across intracellular membranes [*see Figure 1*].[6-8] This system delivers peptides of intracytoplasmic origin to newly formed class I molecules. As noted, certain peptide sequence motifs are known to be characteristic of peptides eluted from purified molecules of a given MHC allele.[9,10] These findings indicate that the allelic sequence differences at the margins of the peptide-binding groove determine which peptide sequences will bind. Class I–bound peptides are usually nine amino acids long, with residues at particular locations that have similar charge or hydrophobicity (e.g., at positions 1, 3, and 9) for different groups of *HLA* alleles. In addition, a number of synthetic peptides representing immunogenic portions of infectious agents or other foreign proteins align on similar common motifs. Peptides eluted from purified HLA-DR class II molecules are variable in length, up to 25 residues, and have a minimal length of 13 to 14 amino acids. The motifs for DR1 represent a positively charged residue at position 1, a hydrogen bond donor at position 6, and hydrophobic residue at position 10.[11] Prediction of binding affinity for a given HLA sequence is becoming common practice for development of peptide vaccines and studies of the specific immune response to protein antigens.[10,12]

CLONAL SELECTION OF T CELLS

The corecognition of MHC and peptide fragments of an antigen bound in the groove of the class I or class II molecules appears to require that the binding surface of the TCR and the binding surface formed by MHC plus peptide be attached at multiple points [*see 6:III Adaptive Immunity: Antigens, Antibodies, and T Cell and B Cell Receptors*]. Each T cell clone is specific for a self-MHC–peptide complex and generally does not have sufficient affinity for MHC or peptide to bind well to either component alone. There is extensive evidence that the development of the T cell repertoire in the thymus begins during the fetal period and continues well into adult life as new precursor cells from the bone marrow mature in the thymus. In this process, many potential clones are destroyed and others are selected to mature. The selected T cell clones then leave to populate the rest of the body. The MHC of the host plays the major role in selection: T cell clones that are strongly autoreactive to self-MHC molecules are eliminated, leaving clones with weak affinity to self-MHC to survive. Because the surviving clones have a large variety of T cell receptor rearrangements [*see 6:III Adaptive Immunity: Antigens, Antibodies, and T Cell and B Cell Receptors*], the individual retains the necessary repertoire of T cell clones that can recognize self-MHC plus peptide. The successful crystallization of a complex consisting of a human TCR, its viral peptide, and the HLA-A2 molecule that binds it has revealed the configuration and extent of the binding surface between the TCR and the MHC-peptide surface. The axis of the TCR is diagonal to that of the MHC helices, so that it covers a large portion of both α helices and the peptide between them. Although the extensive MHC polymorphisms increase the likelihood that a particular peptide fragment will be bound so that it can be recognized by T cells, a given individual has a small repertoire of such MHC binding sites compared with the rich combinatorial possibilities in the TCR gene complex. The inheritance of multiple *HLA* loci from two parents, however, increases the potential for recognizing a greater number of different self-MHC–peptide complexes and therefore increases the likelihood that at least some persons will survive a given infection.

The alloresponse, which is the immune response mounted against another individual's cells, is a special case. Except for direct activation of T cell subsets with bacterial superantigens (e.g., staphylococcal exotoxins), the in vitro proliferation of T cells in the MLR is the most vigorous antigen-specific response known because it does not require the priming that is needed to induce proliferation to microbial antigens. Transplantation is, of course, a technological artifact and would not have been encountered during evolution; only pregnancy has the potential for exposing the cells of one person to those of another having different *HLA* haplotypes. The allobarrier could have made pregnancy difficult or impossible except for the presence of several imperfectly defined mechanisms at the placental level that protect the fetus from rejection. The existence of such mechanisms suggests that the need for MHC polymorphisms is most important and requires special protection at the maternal-fetal interface.

Alloreactive T cells are known to either indirectly perceive allo-MHC peptides presented on self-MHC molecules or directly recognize intact allo-MHC molecules that hold a self-peptide.[13] Because a number of peptides derived from endogenous proteins occupy MHC binding sites at all times, such self-peptides need not be polymorphic or unique to an individual. The functional significance of the indirect, as well as the direct, pathways in transplantation has been established. It has been shown in an-

imal models that immunization with synthetic allopeptides alone can cause accelerated graft rejection,[14,15] whereas administration of such peptides by the oral or intrathymic route can increase tolerance for alloantigens. Also, priming to allopeptides presented by self-MHC molecules is a feature of rejection activity in human transplant recipients.[16]

GENERATION OF CYTOTOXIC T CELLS

The MLR leading to the generation of cytotoxic T cells requires two distinct types of responding T cells. The process begins with the stimulating cell—a B cell, dendritic cell, or monocyte—which has both MHC class I and MHC class II molecules on its surface. The class II molecule stimulates subsets of responding T cells to proliferate and become helper T cells. This subset is marked by the CD4 antigen. The class I molecule sensitizes a second subset of T cells, which become cytotoxic T cells if stimulated by the proliferating helper T cells. One of these stimulatory signals is mediated by the lymphokine interleukin-2 (IL-2). This second T cell subset is marked by the CD8 antigen. Cytotoxic T cells that develop against cells that differ only in their class II antigens bear the CD4 marker. The two stimuli—the one that induces helper T cell proliferation and the one that sensitizes T cells to become cytotoxic—can be delivered by different cells [*see Table 2*]. This type of cell interaction and cooperation is thought to mirror in vivo events that lead to graft rejection by cytotoxic T cells, showing why it is desirable to have both class I antigen and class II antigen compatibility between donor and recipient cells.

It was formerly thought that CD4+ T cells were simply helper lymphocytes and that CD8+ T cells were either cytotoxic or suppressor lymphocytes, but these functional divisions do not appear to be clear-cut. Ongoing molecular studies indicate that the CD4 surface molecule is closely associated with the TCR and guides interaction between T cells and antigen-presenting cells by binding to a nonpolymorphic region of MHC class II molecules [*see 6:III Adaptive Immunity: Antigens, Antibodies, and T Cell and B Cell Receptors*]. Similarly, the CD8 molecule binds to MHC class I molecules on antigen-presenting cells. CD4 and CD8 molecules also increase the strength with which the TCR complex binds to the antigen-MHC complex. In addition, these surface molecules participate in signaling activation of the adherent T cell.

Immune Response Genes

As previously mentioned, many lines of evidence indicate that MHC class II molecules are the expressed products of immune response (Ir) genes; in other words, immune responsiveness can be a direct function of antigen presentation. If an antigen fragment is not bound to a class II molecule, a person's immune system is unable to recognize it. Certain diseases in animals—including virally induced forms of leukemia, mammary tumors, and lymphocytic choriomeningitis—have been linked to polymorphism of MHC class II genes. However, the ability of specific HLA antigens to confer susceptibility to clinically important infectious agents has rarely been suggested (see below). It is likely that evolution has resulted in selection of MHC alleles that are capable of binding at least some portions of antigenic molecules on infectious agents. In addition, the duplication of class II genes with expression of HLA-DR, HLA-DQ, and HLA-DP sets of molecules increases the likelihood that a response can be initiated in a given case. In particular, polymorphisms on both α and β chains of HLA-DQ and HLA-DP provide considerable variation in binding configurations, especially

Table 3 Diseases Showing Positive HLA Antigen Association(s)[22]

Type	*Disease*	*Serologic HLA Antigen*	*Relative Risk**
Rheumatologic	Ankylosing spondylitis	B27	90.0
	Reiter syndrome	B27	37.0
	Acute anterior uveitis	B27	8.2
	Reactive arthritis	B27	18.0
	Psoriatic arthritis	B27	10.7
		B38	9.1
	Juvenile rheumatoid arthritis	B27	3.9
	Juvenile rheumatoid arthritis (pauciarticular)	DR5	3.3
	Rheumatoid arthritis	DR4/Dw4	6.0
	Sjögren syndrome	Dw3	10.0
	Systemic lupus erythematosus	DR3	2.6
Gastrointestinal	Gluten-sensitive enteropathy	DR3	12.0
	Chronic active hepatitis	DR3	6.8
	Ulcerative colitis	B5	3.8
	IgA deficiency	DR3	13.0
Hematologic	Idiopathic hemochromatosis	A3	6.7
		B14	2.7
		A3, B14	90.0
	Pernicious anemia	DR5	5.0
	Hodgkin disease	DP3	2.0
Dermatologic	Dermatitis herpetiformis	DR3	17.3
	Psoriasis vulgaris	Cw3	7.5
	Psoriasis vulgaris (Japanese)	Cw6	8.5
	Pemphigus vulgaris (Jewish)	DR4	24.0
		A26	4.8
	Behçet disease (white)	B5	3.8
	Behçet disease (Japanese)	B51	12.4
Endocrine	Diabetes mellitus, type 1	DR4	6–7
		DR3	5.0
		DR2	0.25
		BfF1+	15.0
	Graves disease	B8	2.5
		DR3	3.7
	Graves disease (Japanese)	B35	4.4
		DR3	3.7
	Addison disease	Dw3	10.5
	Subacute thyroiditis	B35	13.7
	Hashimoto thyroiditis	DR5	3.0
	Congenital adrenal hyperplasia	B47	15.4
Neurologic	Myasthenia gravis	B8	3.0
	Multiple sclerosis	DR2/Dw2	6.0
	Narcolepsy	DR2, DQ6	130.0
Psychiatric	Bipolar disorder	B16	2.3
	Schizophrenia	A28	2.3
Renal	Idiopathic membranous nephropathy	DR3	5.7
	Goodpasture syndrome	DR2	16.0
	Minimal change disease	DR7	4.2
	IgA nephropathy (French, Japanese)	DR4	3.1
	Gold/penicillamine nephropathy	DR3	14.0
	Polycystic kidney disease	B5	2.6
Infectious	Tuberculoid leprosy (Asians)	B8	6.8
	Paralytic polio	B16	4.3
	Low versus high response vaccinia	Cw3	12.7
	Falciparum malaria, severe	B53	0.4–0.5

*Relative risk = $\frac{\text{(\% antigen-positive patients) (\% antigen-negative control subjects)}}{\text{(\% antigen-negative patients) (\% antigen-positive control subjects)}}$

when αβ dimers are composed of chains inherited from both parents; for example, $\alpha_{mother}/\beta_{father}$ may provide a peptide-binding molecule not present in either parent. There are also many non-MHC influences on immune responsiveness; none of these have yet been well characterized clinically.

Studies in humans have also suggested the ability of the MHC to suppress immunologic responses to environmental agents, such as streptococcal infection, schistosomiasis, and leprosy, as well as antigens from cedar pollen and hepatitis B vaccine. For example, the in vitro IgE response to cedar pollen antigen is suppressed by T cells of persons bearing HLA-DQ3, but the mechanisms of such T cell–mediated suppression are ill defined.[17,18]

COMPLEMENT FACTOR GENES

Several complement proteins are encoded by genes that are linked to the MHC. These proteins include C2 and factor B (Bf), which are closely linked and also similar in structure, suggesting gene duplication. In addition, two loci for C4 (*C4A* and *C4B*) are closely linked to C2 and Bf. The C2 deficiency associated with systemic lupus erythematosus is associated with the *HLA-A25, B18* haplotype. Indeed, researchers have found extended haplotypes in which the same *HLA-B, HLA-DR, HLA-DQ*, and complement types are found in apparently unrelated persons with the same disease. These circumstances could result from a mutation occurring in a common ancestor. Alternatively, there may be selective pressures to keep in close proximity genes that produce proteins that act together.

NONIMMUNOLOGIC FUNCTIONS OF MHC GENES

MHC genes are possibly also important in a variety of nonimmunologic cell-cell interactions. In 1976, a study showed that when a male mouse was presented with two females in estrus that were genetically identical except in portions of the MHC, the male would most often choose to mate with the female of an MHC type different from his own.[19] Further experiments showed that the male discriminated between MHC types by sense of smell. The advantage most apparent in this example of opposites attracting is that the heterozygosity of genes in the region that encodes for MHC ensures a wider range of immune defenses for the hybrid progeny of such matings. There is no evidence that humans can sense HLA antigens, however.

Disease and the Major Histocompatibility Complex

HLA-ASSOCIATED DISEASE

Many diseases have been associated with certain MHC antigens [*see Table 3*]. Such associations per se show only that the MHC molecules or some other genes closely linked in the HLA region have an influence on initiation or expression of disease. A relative risk of 5, for example, means only that there is a fivefold increase in the likelihood of disease in a person with a particular HLA antigen, compared with someone who does not have that antigen. It indicates nothing about the frequency of the disease itself, which may be rare or common. One explanation for such associations is that the disease in question is related to a deficiency in the immune response to a particular causative organism. There is increasing evidence, however, that organ-specific HLA-associated diseases—such as type 1 diabetes mellitus, multiple sclerosis, Graves disease, the glomerulonephritides, celiac disease, ankylosing spondylitis, and rheumatoid arthritis—have a major component of autoimmunity.

In animal models in which appropriate breeding studies have been done, it has been demonstrated that autoimmune states depend on five to 15 randomly segregating genes, one of which is in the MHC. Polygenic etiology of human autoimmunity is very likely, and the HLA components may be useful targets for intervention, particularly in diseases in which HLA presentation of an immunogenic self-peptide is a key event. Also, with the development of inflammation, de novo expression of HLA class II molecules on tissue cells may provide the immune stimulus for perpetuation of the autoimmune process. For example, patients with thyroiditis show aberrant expression of HLA-DR on thyroid cells, providing a possible mechanism by which thyroid antigen could be presented to T cells.

There has been some progress in discerning which diseases may be directly related to immunogenic peptide presentation. Analysis of the sequences of genes encoding MHC class II molecules from patients with type 1 diabetes mellitus suggests that inheritance of particular HLA alleles is important in determining susceptibility to this disease, involving a T cell–mediated autoimmune response to pancreatic islet cell antigens. Resistance to type 1 diabetes mellitus is strongly associated with the presence of aspartate at position 57 of the *HLA-DQB* chain. In persons with the *HLA-DR2* haplotype, for example, the relative risk for the disease drops to 0.2 [*see Table 3*]. *HLA-DR2* is in linkage disequilibrium with *HLA-DQB* alleles, such as *DQB1*0602*, encoding aspartate at position 57. In contrast, when aspartate is not present at position 57, particularly in persons with the *HLA-DR3* or *HLA-DR4* haplotype, there is an increased risk of type 1 diabetes mellitus. Amino acid residue 57 on the HLA-DQB chain would lie toward one end of the groove; aspartate at that position may influence binding of a peptide to this class II molecule, causing reduction of helper T cell responses or activation of suppressor T cell responses to pancreatic islet cell antigens. Many studies in certain ethnic groups have shown that the greatest susceptibility to type 1 diabetes mellitus is related to *HLA-DQ*. The *DQA/DQB* heterodimer *DQA1*0301/DQB1*0201* is associated with the highest risk. What is of interest here is that this heterodimer is uncommon, occurring mostly in persons who have inherited the *DQA* gene from one parent and the *DQB* from the other. Whereas *DQA1*0301* and *DQB1*0201*, usually found with *DR4* and *DR3* haplotypes, respectively, separately increase the risk for type 1 diabetes mellitus, together they provide the highest risk of disease. As noted previously, the formation of a heterodimer from the products of genes inherited from both parents does occur with the HLA-DQ molecule. The hypothesis is that this "new" peptide-binding site will be most effective in the presentation of pancreatic islet cell autoantigen. Definition of the binding motifs of this site may provide a clue to the antigen. There are additional and independent effects of *HLA-DR*—particularly the *DR4* alleles, some of which are associated with enhancement and others with suppression of the risk for diabetes. Amino acid differences in the hypervariable regions of MHC class II molecules have also been associated with such autoimmune disorders as pemphigus vulgaris and rheumatoid arthritis.

The association of narcolepsy with *HLA-DR2* (*DRB1*1501*) is more than 90%, but the highest association is with *HLA-DQA1*0102/DQB1*0602*. The HLA effect is dominant, not recessive, and there is no indication of an immunologic defect in affected persons. An abnormality in a peptide neurotransmitter or its receptor has been postulated, but the relation to the *HLA-D*–region genes remains elusive.

About 80% to 90% of celiac disease is associated with *HLA-DQA1*0501/DQB1*0201*. The peptide-binding groove of this

molecule is known to bind a peptide of wheat protein gliadin, which is a potentiating if not etiologic factor in this disease.

Although an HLA molecule may determine specificity to a particular autoantigen, it is possible that genes controlling other factors (e.g., the production of antigen receptors, specific subsets of regulatory cells, or helper and suppressor molecules) are responsible for a general tendency toward an abnormal immune response. Additional study of the peculiar role of the HLA system in autoimmunity may well reveal mechanisms of autoimmune disease that are currently unknown.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Bjorkman PJ, Saper MA, Samraoui B, et al: Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329:506, 1987

2. Buchsbaum S, Barnea E, Dassau L, et al: Large-scale analysis of HLA peptides presented by HLA-Cw4. Immunogenetics 55:172, 2003

3. Brown JH, Jardetzky TS, Gorga JC, et al: Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364:33, 1993

4. Marsh SG, Albert ED, Bodmer WF, et al: Nomenclature for factors of the HLA system, 2002. Tissue Antigens 60:407, 2002

5. Mori M, Beatty PG, Graves M, et al: HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. Transplantation 64:1017, 1997

6. Grandea AG 3rd, Lehner PJ, Cresswell P, et al: Regulation of MHC class I heterodimer stability and interaction with TAP by tapasin. Immunogenetics 46:477, 1997

7. Powis SH, Tonks S, Mockridge I, et al: Alleles and haplotypes of the MHC-encoded ABC transporters TAP1 and TAP2. Immunogenetics 37:373, 1993

8. Szafer F, Oksenberg JR, Steinman L: New allelic polymorphisms in TAP genes. Immunogenetics 39:374, 1994

9. Griffin TA, Yuan J, Friede T, et al: Naturally occurring A pocket polymorphism in HLA-B*2703 increases the dependence on an accessory anchor residue at P1 for optimal binding of nonamer peptides. J Immunol 159:4887, 1997

10. Sathiamurthy M, Hickman HD, Cavett JW, et al: Population of the HLA ligand database. Tissue Antigens 61:12, 2003

11. Murthy VL, Stern LJ: The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding. Structure 5:1385, 1997

12. Borras-Cuesta F, Golvano J, Garcia-Granero M, et al: Specific and general HLA-DR binding motifs: comparison of algorithms. Hum Immunol 61:266, 2000

13. Gallon L, Watschinger B, Murphy B, et al: The indirect pathway of allorecognition: the occurrence of self-restricted T cell recognition of allo-MHC peptides early in acute renal allograft rejection and its inhibition by conventional immunosuppression. Transplantation 59:612, 1995

14. Vella JP, Magee C, Vos L, et al: Cellular and humoral mechanisms of vascularized allograft rejection induced by indirect recognition of donor MHC allopeptides. Transplantation 67:1523, 1999

15. Vella JP, Vos L, Carpenter CB, et al: Role of indirect allorecognition in experimental late acute rejection. Transplantation 64:1823, 1997

16. Ciubotariu R, Tsang ML, Steinman RM, et al: Dendritic cells crossprime allo-specific self-restricted CD4(+) T cells after coculture with dead allogeneic cells. Hum Immunol 63:517, 2002

17. Ikagawa S, Matsushita S, Chen YZ, et al: Single amino acid substitutions on a Japanese cedar pollen allergen (Cry j 1)–derived peptide induced alterations in human T cell responses and T cell receptor antagonism. J Allergy Clin Immunol 97:53, 1996

18. Honda K, Matsushita S, Yasuda N, et al: Segregation analysis of IgE responses to *Cryptomeria japonica* pollen antigen in vivo. Eur Arch Otorhinolaryngol 251(suppl 1):S68, 1994

19. Yamazaki K, Boyse EA, Mike V, et al: Control of mating preferences in mice by genes in the major histocompatibility complex. J Exp Med 144:1324, 1976

20. Abbas AK, Lichtman AH, Pober JS: The major histocompatibility complex. Cellular and Molecular Immunology, 2nd ed. WB Saunders Co, Philadelphia, 1994, p 96

Acknowledgments

Figure 1 Laura Brown.

Figure 2 Seward Hung.

VII IMMUNOGENETICS OF DISEASE

Edgar L. Milford, m.d.
Charles B. Carpenter, m.d.

Differences in genetic makeup from individual to individual have long been recognized to have physiologic consequences in both health and disease. The recent ability to do high-throughput sequencing of genes has revealed that many genes have variants that are present in a significant proportion of the population. Inherited variants of specific genes, either alone or in combination with other genes, may confer a differential risk of disease or of rejection of transplanted tissue.

Genetic Polymorphism

The fundamental basis of genetic polymorphism in a population is variation of the nucleotide sequence of DNA at homologous locations in the genome. These differences in sequence can result from mutations involving a single nucleotide or from deletions or insertions of variable numbers of contiguous nucleotides. Each of these variants presumably occurred in a single ancestor in the distant past. Most new mutations are extinguished through random genetic drift and never become established in the population at any significant frequency. When the gene frequency of a mutation becomes established at more than 1% to 2%, it is often given the more dignified appellation of allele.

Allelic variants can occur anywhere in the genome. Some are found within coding regions of genes, and others are located in introns or gene regulatory regions. However, still others are found in areas that are not closely linked to any known expressed gene.

Equilibrium, Disequilibrium, Genotypes, and Haplotypes

There can be multiple polymorphic nucleotide positions in or near an expressed gene on the same chromosome. In such cases, it is desirable to know whether specific variants at each of the polymorphic positions are independent of the variants at the other positions. If examination of a population shows that the variants at the different positions occur independently of one another, the system is said to be in Hardy-Weinberg equilibrium.[1] If certain variants at one of the positions are statistically associated with specific variants at another of the linked positions, the system is said to exhibit linkage disequilibrium.[1]

Hardy-Weinberg equilibrium can be reestablished over many generations through recombination events. The closer the polymorphic loci are to each other on the chromosome, the less likelihood there is of a recombination and the more likely it is for the specific alleles at the two linked loci to be inherited en bloc as a haplotype. For example, if there are two polymorphic positions within a gene, each of which has two alleles, a given individual will have up to four definable alleles. These alleles are inherited as two parental haplotypes, each of which carries one allele from each of the two loci. Most methods used to type individuals cannot organize the genotype into haplotypes without additional information. The common assays simply define the genotype at each of the two polymorphic positions. Extensive population studies permit sophisticated maximum-likelihood estimates of haplotype frequencies within the population.[2] These studies, combined with confirmatory cloning and sequencing studies of individual DNA strands, often reveal that some theoretically possible haplotypes never occur, whereas others can be assumed when a specific allele is present (because of linkage disequilibrium) [*see Figure 1*]. The ability to deduce haplotypes provides a much higher degree of specificity to the analysis of genetic polymorphism, because the haplotype more accurately defines a larger inherited region of DNA.

Types of Genetic Polymorphism

Single-nucleotide polymorphisms (SNPs) are allelic variants that have been generated as the result of conversion of one nucleotide to another at a homologous position. When present within a coding region (exon) of a gene, the expressed product may or may not have a single amino acid difference, depending on the resulting codon change. In some cases, the change can lead to either a nonsense codon or a stop codon, which halts the transcription process and results in the production of a truncated peptide. SNPs that are located in regulatory regions of an expressed gene can alter the transcription efficiency of that gene but not the protein sequence [*see Figure 2*].

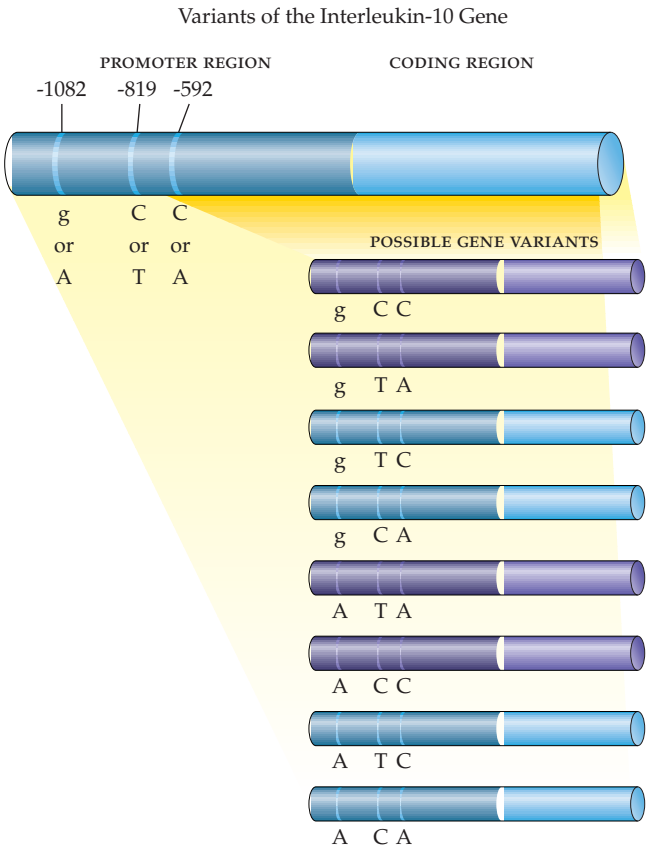


Figure 1 **Single-nucleotide variants occur at positions –1082, –819, and –592 in the promoter region of the interleukin-10 (IL-10) gene. Although eight variants are theoretically possible, only three of these potential IL-10 variants (in purple) are actually observed in large population studies. This is a consequence of strong linkage disequilibrium between the variants at those three positions.**

Deletion or insertion mutants have also been found in functional genes, sometimes at frequencies that merit their inclusion as alleles. Again, the consequence of a deletion depends on the precise location of the deletion; whether it produces a nonsense frameshift; and whether it alters the function of the expressed product. Angiotensin-converting enzyme (ACE) represents a gene that has a deletion variant in which a 278-base-pair segment of intron 16 is excised. This deletion variant is associated with increased ACE levels.

Another class of allelic variance in association with a particular gene is short tandem repeat (STR) polymorphism. Short sequences of two to four base pairs at a given location can be duplicated back-to-back a specific number of times and inherited as a genetic variant. Because such variation would usually result in a nonsense codon, these STRs are almost always located in noncoding regions. The interferon gamma (IFN-γ) gene has such an STR within intron 1, in which the (CA) dinucleotide motif is repeated a variable number of times. The allele with $(CA)_{12}$—that is, with 12 repeats of the CA motif—is associated with high IFN-γ production [*see Figure 3*].

METHODS OF DETECTION OF GENETIC POLYMORPHISM

DNA-based genotyping methods are rapid, accurate, and economical. SNPs can easily be detected, with a high degree of specificity and sensitivity. The assays depend on amplification of the polymorphic locus in question to produce sensitivity in the setting of a background of sample genomic DNA. Specificity is ensured by using tailored oligonucleotides that are complementary to the DNA sequence of the allele one wants to detect.

One strategy for typing is to use polymerase chain reaction to amplify a segment of DNA that includes the polymorphic position and a moderate amount of flanking DNA on both the 3′ and 5′ sides of the polymorphic position. This is done with primers that are complementary to conserved sequences in either side of the desired segment to be amplified. This yields an amplicon of known size that contains inherited alleles and is present in an amount that can be tested for the presence of specific alleles without significant interference from genomic DNA. The amplicon can then be probed, using a set of fluoresceinated or radiolabeled oligonucleotides, each of which is complementary to the DNA sequence of one of the possible alleles. This method is often referred to as site-specific oligonucleotide probe (SSOP) testing.

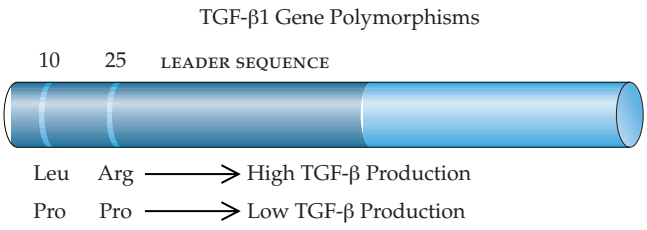


Figure 2 **Single-nucleotide polymorphisms have been identified in the gene for transforming growth factor–β1 (TGF-β1). Each polymorphism involves two alleles in the leader sequence of the gene. These biallelic nucleotide substitutions produce codon changes that result in alternative amino acids. Leu10 is in linkage disequilibrium with Arg25, and Pro10 is in linkage disequilibrium with Pro25. The Leu10Arg25 variant is associated with high TGF-β1 production, whereas the Pro10Pro25 variant is associated with lower production. This may be the consequence of different efficiency of posttranslational modification for the two variants, which differ only in the leader amino acid sequence.**

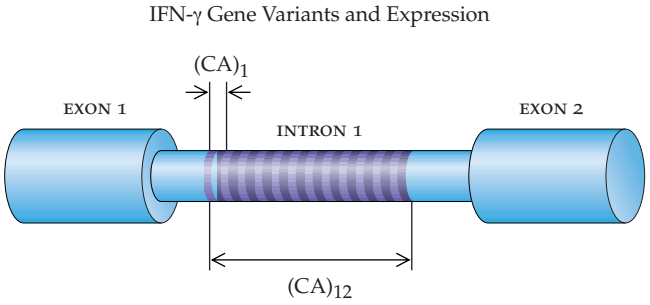


Figure 3 **Illustration of a short tandem repeat (STR) polymorphism within intron 1 of the interferon gamma (IFN-γ) gene. STR polymorphisms in this intron differ according to the number of repetitions of the cytosine-arginine (CA) motif. The allelic variant with 12 tandem repeats [(CA_{12})] is associated with higher IFN-γ production.**

Another strategy for SNP typing, which does not require two steps, is called site-specific priming (SSP). This method takes advantage of the fact that the 3′ terminal base of a primer is where DNA synthesis commences during each cycle of PCR. For synthesis to proceed, the 3′ base must be closely bonded to its complementary base on the template DNA. Therefore, the terminal 3′ base of the primer can be used to render the PCR reaction itself exquisitely sensitive to the identity of the base that is on the template. For detection of SNPs, one can craft a set of PCR primers that are complementary to the alleles to be detected, with the terminal 3′ base of one of the primers located at the polymorphic position. The second PCR primer is usually complementary to a conserved segment of DNA and positioned to yield a product of a convenient size. If an allele is present, use of the appropriate set of primers will produce an amplicon. The amplicon can be separated from genomic DNA by simple agarose gel electrophoresis and identified by ethidium bromide staining under ultraviolet light, and the expected size can be confirmed.

Both SSOP and SSP can be modified to detect deletion or insertion variants. With SSP, using primers that flank the deletions or insertions, amplicons of characteristic sizes are produced. SSOP can confirm the presence or absence of the deletions or insertions through the use of probes that include the junctions of the deleted or inserted regions.

RELEVANCE OF GENETIC POLYMORPHISM IN HUMANS

Historically, polymorphisms in several genetic systems have been recognized as a barrier to transfusion and transplantation. The ABO blood group antigens were among the earliest genetically determined glycoproteins that exhibited mendelian inheritance and had biologic relevance in humans.[3] Mismatch for the ABO antigens is a risk factor not only for transfusion reactions but also for solid-organ transplantation because of the prominent expression of these antigens on the vascular endothelium.

The major histocompatibility complex (MHC)—so called because of its prominent role in rejection of allogeneic tissue—is a primary barrier to transplantation of solid organs, tissue, and hematopoietic stem cells. This closely linked cluster of highly polymorphic genes, grouped on the short arm of chromosome 6, encodes cell surface molecules (human leukocyte antigens [HLA]). The normal role of the MHC is presentation of endogenous and exogenous peptide antigen fragments to T cells,

thereby initiating an immune response against the molecule (or pathogenic organism) from which the peptide was derived.[4] The extreme variability of molecular structure in the MHC antigens permits a wide range of different peptides to be presented by autologous human antigen-presenting cells, although some persons may have a specific repertoire of MHC antigens that do not present certain antigens effectively. The focused immunogenicity of MHC molecules and the variability of these molecules from person to person render them prominent targets for the immune response in the context of solid-organ and bone marrow transplantation. In cases in which live allogeneic cells are the target of the immune response, the apparent target is the nonself MHC molecule itself. Freedom from rejection and, in the case of bone marrow transplantation, graft versus host disease (GVHD) is improved with HLA matching of donor and recipient.

Innate and Adaptive Responses

It has become abundantly clear that the selective (adaptive) immunologic response, which is important in organ transplantation, tissue transplantation, and defense against certain microorganisms, is closely associated with the innate cellular and humoral pathways of nonspecific tissue injury, inflammation, hypoxia, and healing. Macrophages, for example, play a central role in the response to hypoxia, trauma, bacterial invasion, and inflammation caused by exogenous toxins, but they are also important in the processing and presentation of antigen to the specific immune system. Natural killer (NK) cells, which constitute approximately 10% of human mononuclear cells, are thought to be important mediators of innate immunity. Their cytolytic activity is regulated by inhibitory receptors, called killer immunoglobulin-like receptors (KIRs).[5] Class I MHC molecules are ligands for the KIRs—in particular, genetically determined epitopes on HLA-B and HLA-C molecules that have limited polymorphism.[5] In bone marrow transplantation, recipients who present the appropriate class I ligands to donor NK cells will downregulate the NK response. This is thought to decrease both GVHD and graft versus tumor activity.

Humans also have innate humoral immunity against a number of glycoprotein antigens. This so-called natural antibody is thought to have protective effects against a wide range of bacterial products. At the same time, the humoral immune system is able to mount a robust adaptive response to an astonishingly broad spectrum of specific antigens, if challenged to do so. The genes responsible for the adaptive immune response are highly polymorphic, but they are found only in specific subsets of T cells with antigen receptor genes that are rearranged during thymic development and in B cells that undergo somatic mutation in response to antigenic challenge. Specific germline variant alleles of the T cell receptor for antigen (before somatic mutation) are also associated with differential susceptibility to a number of immunologically mediated conditions, including renal allograft rejection and several rheumatic diseases, such as rheumatoid arthritis.[6]

Other Polymorphic Genes Involved in Organ and Tissue Injury

Variants of genes can influence organ and tissue physiology, directly induce diseases, or render the person more susceptible or resistant to a pathologic state. Variants that directly induce a profound disease state are usually rare in the population, because the disease may cause death before the person can reproduce. Variants or mutations that cause severe early disease are not discussed in this subsection. Polymorphic variants of loci that have a more subtle effect on disease susceptibility are more likely to become established in the population at frequencies of 1% or more (i.e., to become alleles). Several patterns can be appreciated with these alleles. Variant alleles may exhibit a gene-dose effect, with heterozygotes having an intermediate influence, between that of the normal genotype (the so-called wild type) and the homozygous variant genotype. In other cases, a variant allele appears to have a dominant influence; presumably, these variants are able to achieve significant frequency in the population because the condition they produce does not substantially decrease reproduction. The disease phenotype that is a measurable physiologic consequence of a particular genotype may be a downstream effect that depends on multiple influences, including the genotype in question, interaction with other genes, and environmental exposure.

Loci that encode cytokines, chemokine receptors, costimulatory molecules, and components of physiologically important pathways such as the angiotensin system are all concrete examples in which genetic polymorphism influences pathophysiology. These examples can be used to highlight some ways in which determination of individual genotype can assist in assessing risk of disease.

Cytokines and chemokines are secreted proteins and glycoproteins that act as important signaling devices in both the innate and the adaptive responses. They serve variously as chemoattractants and as inducers or suppressors of leukocyte, endothelial cell, platelet, fibroblast, and myocyte function. They have a particularly notable effect on cells that bear the appropriate receptors. Cytokines and chemokines often represent a common pathway that links the classical immune pathway and other pathways of tissue injury and repair, such as those involved in ischemia, trauma, and toxic damage.[7]

Costimulatory molecules such as CTLA-4 are expressed on the cell membranes of T cells and serve as ligands for complementary molecules on antigen-presenting cells [*see 6:IX Immunologic Tolerance and Autoimmunity*]. The engagement of costimulatory molecules with their ligands can augment or suppress the magnitude of the immune response induced by the recognition of antigen via the T cell receptor.[8,9] Soluble CTLA-4 has been used to block antigen-dependent T cell activation by competitive blockade of normal cell membrane–bound interaction.

Functional Consequences of Specific Genetic Variants

CYTOKINE POLYMORPHISMS

Variants in the genes that govern the production of cytokines such as interleukin-10 (IL-10), tumor necrosis factor–α (TNF-α), and transforming growth factor–β (TGF-β) can help determine whether a person has high or low levels of these cytokines.[10] The cytokine network is thought to play an important role both in rejection of allografts and in tolerance,[11] and a number of clinical effects of these polymorphisms in cytokine genes have now been described [*see Tables 1 and 2*].

TGF-β has two well-studied dimorphic positions within the leader sequence of the gene [*see Figure 2*]. These polymorphisms are in linkage disequilibrium; only two variants of the TGF-β gene have been described, rather than the four theoretically possible combinatorial variants. TGF-β is considered to be

Table 1 Cytokine Genetic Polymorphisms and Their Pathophysiologic Effects[9,31-59]

Locus	*Position*	*Genotype*	*Pathophysiologic Effect*
CTLA4	Microsatellite	Allele 3 and allele 4	Increased rejection, liver/kidney transplants
IFN-γ	Microsatellite	Allele 2 (12 CA repeat)	Increased production
IFN-γ	Microsatellite	Allele 2 (12 CA repeat)	Increased production
IFN-γ	Microsatellite	Allele 2 (12 CA repeat)	Increased acute rejection, kidney transplants
IFN-γ	Microsatellite	Allele 3 homozygotes	Increased GVHD, bone marrow transplant patients
IFN-γ	T+874A	T allele	Increased production
IL-10	-1082A	A allele	Low producer
IL-10	-1082A	A/A homozygotes	Increased frequency in Wegener granulomatosus
IL-10	-1064	Low producer	Increased GVHD, bone marrow transplant patients
IL-10	-1082, -819, -592	Low producer	Increased rejection, pediatric heart transplant patients
IL-10	-1064	High producer	Increased graft survival, renal transplants
IL-10	-1082A	High producer	Increased rejection, if high TNF genotype
IL-10	-1082A	High producer	Increased rejection episodes, renal transplants
IL-10	-1082A	Recipient high; donor low	Increased rejection, renal transplants
IL-4	-590T	Recipient and donor low	Decreased rejection, renal transplants
IL-6	-174C	G allele	Increased acute GVHD, bone marrow transplants
TGF-β1	Arg25Pro	A/A homozygotes	Increased production
TGF-β1	Arg25Pro	Arg	Increased in patients with fibrotic lung disease
TGF-β1	Arg25Pro	A/A homozygotes	Progression of renal insufficiency, heart transplant patients
TGF-β1	Arg25Pro	A/A homozygotes	Decreased gingival hyperplasia with cyclosporine
TGF-β1	Arg25Pro	A/A homozygotes	Increased coronary vasculopathy, heart transplants
TGF-β1	Arg25Pro	Arg	No correlation with renal transplant rejection
TGF-β1	Leu10Pro	Leu	Progression of renal insufficiency, heart transplant patients
TGF-β1	Leu10Pro	L/L homozygotes	Decreased renal dysfunction, heart transplant patients
TGF-β1	Leu10Pro	Pro	Association with dilated cardiomyopathy
TGF-β1	Leu10Pro	Pro/Pro	Increased gingival hyperplasia with cyclosporine
TNF-α	-308A	High A/A or A/G	Increased rejection and creatinine, renal transplants
TNF-α	-308A	High A/A or A/G	Increased GVHD and mortality, bone marrow transplant patients
TNF-α	-308A	High A/A or A/G	Increased rejection, renal transplants
TNF-α	-308A	Low producer	Decreased acute rejection, pediatric heart transplants
TNF-α	-308A	A allele	Sixfold to sevenfold higher production
TNF-α	-308A	A allele	Risk factor renal transplants, if HLA-DR mismatch
TNF-α	-308A	A allele	Increased rejection, pediatric heart transplants
TNF-α	-308A	A allele	Increased frequency in primary sclerosing cholangitis
TNF-α	-308A	A allele	Increased mortality, heart transplant patients
TNF-α	-308A	A allele	Increased rejection, renal transplants
TNF-α	-308A	A allele	Increased hepatitis C recurrence after liver transplants
TNF-α	-308A	A allele	Decreased plasma TNF levels
TNF-α	-308A	A/A homozygotes	Increased rejection, liver transplants
TNF-α	-308A	A/A homozygotes	Increased acute rejection, liver transplants
TNF-α	Microsatellite	High producer	Increased rejection, cardiac transplants (low IL-10 subset)
TNF-α	Microsatellite	High producer	Increased acute GVHD, bone marrow transplant patients
TNF-α	Microsatellite	a9	Increased in rejection, renal transplants, in patients with HLA-B35
TNF-α	Microsatellite	d3/d3 homozygotes	Increased GVHD grade III/IV
TNF-α	Microsatellite	d4	Decreased in rejection, renal transplants, in patients with HLA-B44
TNF-α	NcoI	Low recipient	Increased infection, renal transplants
TNF-α	-308A	A allele	Nonischemic cardiac dysfunction

Note: Some of these gene variants appear to have paradoxical effects, depending on the investigator and the assay system.
GVHD—graft versus host disease IFN—interferon IL—interleukin TGF—transforming growth factor TNF—tumor necrosis factor

a major mediator of fibrosis in kidney and lung allografts.[12,13] Specific variants of the TGF gene that result in high production of TGF-β (so-called high-producer genotypes) are associated with poor outcome in lung transplants: 98% of patients with chronic rejection are homozygous for the high-producer TGF-β genotype represented by Leu at position 10 and Arg at position 25. Moreover, fibrosis develops in the lung grafts of 93% of those with homozygous high-producer TGF-β genotype but only in 7% of those with heterozygous (high/low) producer genotype.[12] TGF-β also mediates the gingival hypertrophy induced by the immunosuppressive agent cyclosporine. Increased gingival hypertrophy has been reported in patients with the low-producer TGF-β genotype, represented by Pro at both position 10 and position 25. Because the two variants differ only in the leader amino acid sequence, the different production levels may be the consequence of different efficiency of posttranslational modification.

ANGIOTENSIN SYSTEM POLYMORPHISMS

The renin-angiotensin system is a metabolic-hormonal pathway that plays a critical role in blood pressure homeostasis and salt and water balance. In the renin-angiotensin pathway, the

Table 2 Renin-Angiotensin System and Chemokine Polymorphism and Pathophysiology[26,60-72]

Locus	*Position*	*Genotype*	*Pathophysiologic Effect*
ACE	Deletion variant	I/I or I/D	Favorable function trend, pediatric renal transplants
ACE	Deletion variant	D	Favorable renal function, bone marrow transplant patients
ACE	Deletion variant	D/D homozygotes	Favorable renal function, renal transplants
ACE	Deletion variant	D	Increased cardiac allograft vascular disease
ACE	Deletion variant	D/D homozygotes	Increased risk of renal failure, pediatric kidney transplants
ACE	Deletion variant	D/D homozygotes	Inreased risk of renal failure, high-risk renal transplants
ACE	Deletion variant	D	Increased frequency in systemic lupus erythematosus
ACE	Function	Function	ACE inhibitor suppresses IL-12, IFN-γ
ACE	Deletion variant	D/D homozygotes	Increased ACE levels
ACE	Deletion variant	D/D homozygotes	Increased blood pressure with salt administration
ACE	Deletion	D	Increased risk of renal dysfunction, renal transplants
AGT	Met235Thr	Met vs Thr	No function association, pediatric renal transplants
AGT	G → A,-6	A/A	Increased risk of renal dysfunction, renal transplants
AT1	A1166C	A vs C	No function association, pediatric renal transplants
AT1	Function	Function	Activation of NF-κB via IL-12 and IFN-γ
AT1	A1166C	C	Increased blood pressure, renal transplant patients
CCR2	V64I	I	Less acute rejection, renal transplants
CCR2	Expression	Expression	High expression in renal transplants during rejection
CCR5	A59029G	A/A homozygotes	Less acute rejection, renal transplants
CCR5	Expression	Expression	High expression in renal transplants during rejection
CXCR3	Expression	Expression	High expression in heart biopsy infiltrates

Note: Some of the gene variants appear to have paradoxical effects, depending on the investigator and the assay system.
ACE—angiotensin-converting enzyme IFN—interferon IL—interleukin

prohormone angiotensinogen (AGT) is converted to angiotensin I by renin. Angiotensin-converting enzyme then catalyzes the conversion of angiotension I to angiotensin II [*see Figure 4*]. Angiotensin II is one of the most potent vasoconstrictive human hormones. In addition, angiotensin II has indirect inflammatory and fibrotic effects, which are distinct from its physiologic vasoconstrictive role. These indirect effects appear to be mediated by cytokines. Angiotensin II promotes the secretion of a number of inflammatory cytokines, including TGF-β, platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), IL-6, IL-12, TNF-α, and IFN-γ.[14-17] There are two receptors for angiotensin II, type 1 (AT1) and type 2 (AT2). AT1 receptors mediate the major vasoconstrictive activity of angiotensin II but also appear to be involved in angiotensin II–dependent augmentation of immune activation and stimulation of TGF-β production. The AT2 receptors are implicated in remodeling; may promote angiotensin II–dependent apoptosis; and have some functions that oppose the AT1 receptor, including vasodilation and increased production of nitric oxide.[18,19]

Several of the genes encoding members of the renin-angiotensin pathway exhibit polymorphisms that influence function. Genomic variants of the genes encoding AGT, ACE, AT1, and AT2 have been described.[20] There is evidence that the AGT(A/A) and ACE(D/D) variants result in increased angiotensin II activity; in turn, the angiotensin II can interact differentially with receptors of different genotypes and influence ultimate pathophysiology. A deletion variant of the ACE gene (D14091-14378) and a single-nucleotide polymorphic variant of the AGT gene (G → A, -6) are correlated with increased peripheral ACE activity and AGT levels, respectively.[21,22] Both genotypes confer increased susceptibility to hypertension, and ACE(D14091-14378) also worsens ischemic heart disease and progression of intrinsic renal insufficiency [*see Table 2*].[23] An analysis of ACE polymorphism in diabetes revealed that the ACE(D) allele is highly associated with diabetic nephropathy.[24]

Of the different classes of white blood cells, T cells contain the highest level of ACE, approximately 28-fold more than

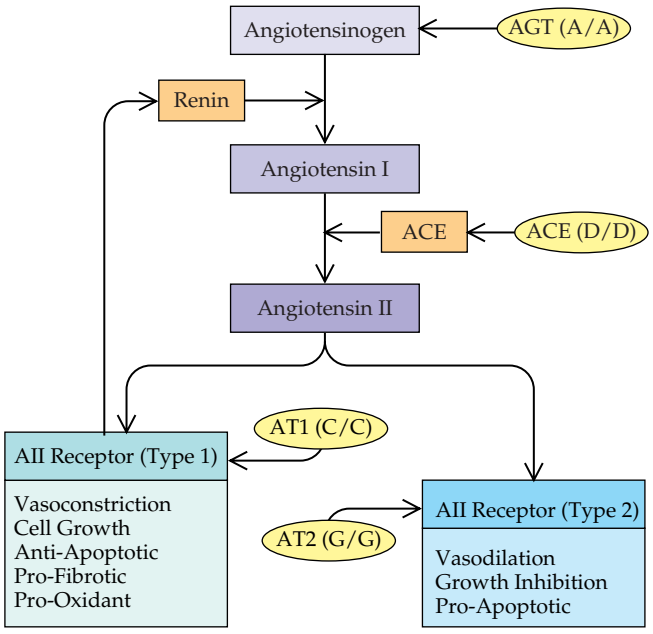


Figure 4 **The renin-angiotensin system is illustrated, along with proven variants of genes responsible for its components. The variant genes indicated are thought to result in a quantitative increase in function in the system. The final hormone, angiotensin II (AII), has a variety of vasoactive, inflammatory, or anti-inflammatory effects, which appear to be dependent on the receptor that is engaged. (ACE—angiotensin-converting enzyme; AGT—angiotensinogen; AT1—angiotensin type 1 receptor; AT2—angiotensin type 2 receptor)**

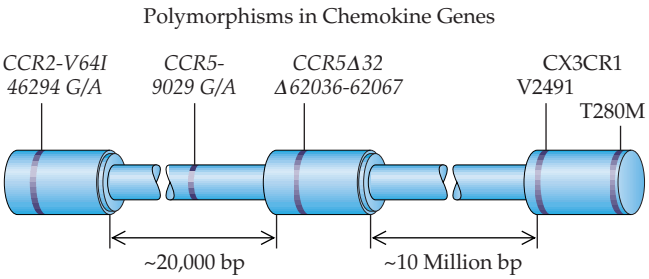


Figure 5 **Locations of chemokine receptor genes and variant positions associated with those genes are shown. The CCR2, CCR5, and CX3CR1 genes are located on the same chromosome. CX3CR1 has two variable positions within the coding region. The location of the deletion variant of CCR5 and the location of the G-to-A variant of CCR5 in the 5′ promoter region of the gene are illustrated. The CCR2 gene has a G-to-A variant located within the coding region.**

monocytes. Indeed, immunologically competent T cells appear to be the major cell type expressing ACE in blood.[25] The ACE expression can vary up to 100-fold during the differentiation of T cells. Monocytes express angiotensin II, the final product of the angiotensin synthetic pathway. Monocyte angiotensin II appears to mediate recruitment of inflammatory cells during renal damage through the synthesis of monocyte chemoattractant protein-1.[26]

A variant of the AT2 gene (A → G, 1332) has been associated with congenital anomalies of the kidney and urinary tract. These developmental abnormalities are preceded by delayed apoptosis of undifferentiated mesenchymal cells surrounding the urinary tract during key ontogenic events.[27] In kidney transplant recipients, specific variants of the ACE and AGT genes are correlated with poor clinical outcomes. Renal transplant patients who have either the ACE(D14091-14378) or the AGT homozygous (G → A, -6)/(G → A, -6) variant have poorer renal transplant function at 3 years, as well as more rapid progression of transplant failure, defined as an increase of serum creatinine levels over time. Diastolic blood pressure in these patients was also significantly higher as a function of the AT1(A → C, 1166) C gene dose. The pathophysiologic reasons for the association between specific angiotensin system gene polymorphisms and renal transplant outcomes are not well understood. Further work is needed to reveal the degree to which this association is a function of hypertensive organ damage or modulation of the immunologic response mediated by angiotensin.

CHEMOKINE RECEPTOR POLYMORPHISMS

Chemokines are molecules with a variety of functions, some of which influence the recruitment of inflammatory cells to sites of injury. Three of the genes encoding chemokine receptors are located on one chromosome. CCR2 and CCR5 are located within 20,000 base pairs of each other; CX3CR1 is located 10 million base pairs away from CCR5 [*see Figure 5*].

The leukocyte chemokine receptor CCR5 is expressed on monocytes, as well as on helper T cells involved in augmentation of the immune response (T_{H1} subset)[28] [*see 6:X Allergic Response*]. CCR5 is a coreceptor for entry of HIV-1 into macrophages. CCR5 binds the inflammatory chemokines RANTES (regulated on activation, normal T cell expressed and repeated), macrophage inflammatory protein (MIP)-1α, and MIP-1β, whereas CCR2 and CX3CR1 are receptors for the chemokines monocyte chemoattractant protein (MCP)-1 and fractalkine, respectively. Antagonists of CCR5, such as met-RANTES, prolong renal allograft survival in MHC-incompatible mice. Furthermore, prolonged heart transplant survival is achieved if the recipient is a homozygous CCR2 or CCR5 knockout. In humans, a 32-base-pair deletion of the CCR5 gene (CCR5 Δ32) renders the gene nonfunctional. There is also a polymorphic single-nucleotide variant of CCR5, CCR5-9029(G → A), which is located in the promoter region of the gene. The G variant is associated with defective transcription. Gene variants of some of these chemokine receptors have been associated with different rates of HIV disease progression.[29] The CCR5 Δ32 variant is associated with lower incidence and severity of asthma and rheumatoid arthritis.[30] Patients with homozygosity for CX3CR1-V2491I(G → A), CX3CR1-T280M(C → T), and CCR5-9029(G → A) have higher HIV progression rates. In contrast, patients with CCR2V64I(G → A) and CCR5(Δ32) exhibit slower progression of HIV, presumably because of reduced binding of the virus to target cells.

In renal transplant patients, the A/A homozygous genotype of the CCR59029(A → G) polymorphic locus is associated with significantly lower incidence of acute rejection episodes in the first posttransplant year. Although this could be explained by a protective effect of A/A homozygosity, it might instead be from a dominant detrimental effect of the G variant, given that both A/G heterozygotes and G/G homozygotes have been found to have similarly high rejection frequencies, which were twice that of patients with A/A genotype.[14]

Practical Applications of Genotyping Polymorphisms

The genetic polymorphisms discussed in this subsection represent but examples of the many inherited variants of physiologically important genes that influence susceptibility to disease. These variants can act alone, in conjunction with variants at other loci, or through interaction with environmental factors to increase or decrease disease incidence or severity. The cytokine genes, chemokine genes, and genes of the renin-angiotensin system are important modulators of the immune response and, in the case of the renin-angiotensin axis, of hypertension and vascular disease. Knowledge of a patient's genotype may assist physicians in assessing prior risk of a pathophysiologic outcome and in tailoring therapy. Clinical trials may, in some cases, be better interpreted by knowledge of participant genotypes, because certain genotypes may impart differential incidence of disease and responsiveness to pharmacologic agents.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Kaessmann H, Zollner S, Gustafsson AC, et al: Extensive linkage disequilibrium in small human populations in Eurasia. Am J Hum Genet 70:673, 2002

2. Mori M, Beatty PG, Graves M, et al: HLA gene and haplotype frequencies in the North American population: the National Marrow Donor Program Donor Registry. Transplantation 64: 1017, 1997

3. Rydberg L: ABO-incompatibility in solid organ transplantation. Transfus Med 11:325, 2001

4. Rudolph MG, Wilson IA: The specificity of TCR/pMHC interaction. Curr Opin Immunol 14:52, 2002

5. Vilches C, Parham P: KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. Annu Rev Immunol 20:217, 2002

6. VanderBorght A, Geusens P, Vandevyver C, et al: Skewed T-cell receptor variable gene usage in the synovium of early and chronic rheumatoid arthritis patients and per-

sistence of clonally expanded T cells in a chronic patient. Rheumatology (Oxford) 39:1189, 2000

7. Goddard S, Williams A, Morland C, et al: Differential expression of chemokines and chemokine receptors shapes the inflammatory response in rejecting human liver transplants. Transplantation 72:1957, 2001

8. Onodera K, Chandraker A, Schaub M, et al: CD28-B7 T cell costimulatory blockade by CTLA4Ig in sensitized rat recipients: induction of transplantation tolerance in association with depressed cell-mediated and humoral immune responses. J Immunol 159:1711, 1997

9. Slavcheva E, Albanis E, Jiao Q, et al: Cytotoxic T-lymphocyte antigen 4 gene polymorphisms and susceptibility to acute allograft rejection. Transplantation 72:935, 2001

10. Rudwaleit M, Siegert S, Yin Z, et al: Low T cell production of TNFalpha and IFNgamma in ankylosing spondylitis: its relation to HLA-B27 and influence of the TNF-308 gene polymorphism. Ann Rheum Dis 60:36, 2001

11. Suthanthiran M: The importance of genetic polymorphisms in renal transplantation. Curr Opin Urol 10:71, 2000

12. El-Gamel A, Awad MR, Hasleton PS, et al: Transforming growth factor-beta (TGF-beta1) genotype and lung allograft fibrosis. J Heart Lung Transplant 18:517, 1999

13. Freedman BI, Yu H, Spray BJ, et al: Genetic linkage analysis of growth factor loci and end-stage renal disease in African Americans. Kidney Int 51:819, 1997

14. Klahr S, Morrissey J: Angiotensin II and gene expression in the kidney. Am J Kidney Dis 31:171, 1998

15. Guo G, Morrissey J, McCracken R, et al: Contributions of angiotensin II and tumor necrosis factor-alpha to the development of renal fibrosis. Am J Physiol Renal Physiol 280:F777, 2001

16. Moriyama T, Fujibayashi M, Fujiwara Y, et al: Angiotensin II stimulates interleukin-6 release from cultured mouse mesangial cells. J Am Soc Nephrol 6:95:101, 1995

17. Khalil A, Tullus K, Bakhiet M, et al: Angiotensin II type 1 receptor antagonist (losartan) down-regulates transforming growth factor-beta in experimental acute pyelonephritis. J Urol 164:186, 2000

18. Cigola E, Kajstura J, Li B, et al: Angiotensin II activates programmed myocyte cell death in vitro. Exp Cell Res 231:363, 1997

19. Yamada T, Horiuchi M, Dzau VJ: Angiotensin II type 2 receptor mediates programmed cell death. Proc Natl Acad Sci U S A 93:156, 1996

20. Abdi R, Tran TB, Sahagun-Ruiz A, et al: Chemokine receptor polymorphism and risk of acute rejection in human renal transplantation. J Am Soc Nephrol 13:754, 2002

21. Ruprecht B, Schurmann M, Ziegenhagen MW, et al: [Corrected normal values for serum ACE by genotyping the deletion-/insertion-polymorphism of the ACE gene.] Pneumologie 55:326, 2001

22. Bloem LJ, Manatunga AK, Tewksbury DA, et al: The serum angiotensinogen concentration and variants of the angiotensinogen gene in white and black children. J Clin Invest 95:948, 1995

23. Petrovic D, Zorc M, Kanic V, et al: Interaction between gene polymorphisms of renin-angiotensin system and metabolic risk factors in premature myocardial infarction. Angiology 52:247, 2001

24. Solini A, Dalla Vestra M, Saller A, et al: The angiotensin-converting enzyme DD genotype is associated with glomerulopathy lesions in type 2 diabetes. Diabetes 51:251, 2002

25. Costerousse O, Allegrini J, Lopez M, et al: Angiotensin I–converting enzyme in human circulating mononuclear cells: genetic polymorphism of expression in T-lymphocytes. Biochem J 290:33, 1993

26. Ruiz-Ortega M, Lorenzo O, Egido J: Angiotensin III increases MCP-1 and activates NF-kappaB and AP-1 in cultured mesangial and mononuclear cells. Kidney Int 57:2285, 2000

27. Pope JC 4th, Brock JW 3rd, Adams MC, et al: How they begin and how they end: classic and new theories for the development and deterioration of congenital anomalies of the kidney and urinary tract, CAKUT. J Am Soc Nephrol 10:2018, 1999

28. Reynes J, Portales P, Segondy M, et al: CD4 T cell surface CCR5 density as a host factor in HIV-1 disease progression. AIDS 15:1627, 2001

29. Tang J, Shelton B, Makhatadze NJ, et al: Distribution of chemokine receptor CCR2 and CCR5 genotypes and their relative contribution to human immunodeficiency virus type 1 (HIV-1) seroconversion, early HIV-1 RNA concentration in plasma, and later disease progression. J Virol 76:662, 2002

30. Mitchell TJ, Walley AJ, Pease JE, et al: Delta 32 deletion of CCR5 gene and association with asthma or atopy. Lancet 356:1491, 2000

31. Pravica V, Perrey C, Stevens A, et al: A single nucleotide polymorphism in the first intron of the human IFN-gamma gene: absolute correlation with a polymorphic CA microsatellite marker of high IFN-gamma production. Hum Immunol 61:863, 2000

32. Asderakis A, Sankaran D, Dyer P, et al: Association of polymorphisms in the human interferon-gamma and interleukin-10 gene with acute and chronic kidney transplant outcome: the cytokine effect on transplantation. Transplantation 71:674, 2001

33. Cavet J, Dickinson AM, Norden J, et al: Interferon-gamma and interleukin-6 gene polymorphisms associate with graft-versus-host disease in HLA-matched sibling bone marrow transplantation. Blood 98:1594, 2001

34. Pravica VP, Borreiro LF, Hutchinson IV: Genetic regulation of interferon-gamma production. Biochem Soc Trans 25:176S, 1997

35. Agarwal P, Oldenburg MC, Czarneski JE, et al: Comparison study for identifying promoter allelic polymorphism in interleukin 10 and tumor necrosis factor alpha genes. Diagn Mol Pathol 9:158, 2000

36. Murakozy G, Gaede KI, Ruprecht B, et al: Gene polymorphisms of immunoregulatory cytokines and angiotensin-converting enzyme in Wegener's granulomatosis. J Mol Med 79:665, 2001

37. Cavet J, Middleton PG, Segall M, et al: Recipient tumor necrosis factor-alpha and interleukin-10 gene polymorphisms associate with early mortality and acute graft-versus-host disease severity in HLA-matched sibling bone marrow transplants. Blood 94:3941, 1999

38. Awad MR, Webber S, Boyle G, et al: The effect of cytokine gene polymorphisms on pediatric heart allograft outcome. J Heart Lung Transplant 20:625, 2001

39. Pelletier R, Pravica V, Perrey C, et al: Evidence for a genetic predisposition towards acute rejection after kidney and simultaneous kidney-pancreas transplantation. Transplantation 70:674, 2000

40. Sankaran D, Asderakis A, Ashraf S, et al: Cytokine gene polymorphisms predict acute graft rejection following renal transplantation. Kidney Int 56:281, 1999

41. Poole KL, Gibbs PJ, Evans PR, et al: Influence of patient and donor cytokine genotypes on renal allograft rejection: evidence from a single centre study. Transpl Immunol 8:259, 2001

42. Awad MR, El-Gamel A, Hasleton P, et al: Genotypic variation in the transforming growth factor-beta1 gene: association with transforming growth factor-beta1 production, fibrotic lung disease, and graft fibrosis after lung transplantation. Transplantation 66:1014, 1998

43. Shahbazi M, Fryer AA, Pravica V, et al: Vascular endothelial growth factor gene polymorphisms are associated with acute renal allograft rejection. J Am Soc Nephrol 13:260, 2002

44. Linden GJ, Haworth SE, Maxwell AP, et al: The influence of transforming growth factor-beta1 gene polymorphisms on the severity of gingival overgrowth associated with concomitant use of cyclosporin A and a calcium channel blocker. J Periodontol 72:808, 2001

45. Densem CG, Hutchinson IV, Cooper A, et al: Polymorphism of the transforming growth factor-beta 1 gene correlates with the development of coronary vasculopathy following cardiac transplantation. J Heart Lung Transplant 19:551, 2000

46. Baan CC, Balk AH, Holweg CT, et al: Renal failure after clinical heart transplantation is associated with the TGF-beta 1 codon 10 gene polymorphism. J Heart Lung Transplant 19:866, 2000

47. Holweg CT, Baan CC, Niesters HG, et al: TGF-beta1 gene polymorphisms in patients with end-stage heart failure. J Heart Lung Transplant 20:979, 2001

48. Poli F, Boschiero L, Giannoni F, et al: Tumour necrosis factor-alpha gene polymorphism: implications in kidney transplantation. Cytokine 12:1778, 2000

49. Hahn AB, Kasten-Jolly JC, Constantino DM, et al: TNF-alpha, IL-6, IFN-gamma, and IL-10 gene expression polymorphisms and the IL-4 receptor alpha-chain variant Q576R: effects on renal allograft outcome. Transplantation 72:660, 2001

50. Bathgate AJ, Pravica V, Perrey C, et al: Polymorphisms in tumour necrosis factor alpha, interleukin-10 and transforming growth factor beta1 genes and end-stage liver disease. Eur J Gastroenterol Hepatol 12:1329, 2000

51. Azzawi M, Hasleton PS, Turner DM, et al: Tumor necrosis factor-alpha gene polymorphism and death due to acute cellular rejection in a subgroup of heart transplant recipients. Hum Immunol 62:140, 2001

52. Rosen HR, Lentz JJ, Rose SL, et al: Donor polymorphism of tumor necrosis factor gene: relationship with variable severity of hepatitis C recurrence after liver transplantation. Transplantation 68:1898, 1999

53. Abdallah AN, Cucchi-Mouillot P, Biteau N, et al: Analysis of the polymorphism of the tumour necrosis factor (TNF) gene and promoter and of circulating TNF-alpha levels in heart-transplant patients suffering or not suffering from severe rejection. Eur J Immunogenet 26:249, 1999

54. Bathgate AJ, Pravica V, Perrey C, et al: The effect of polymorphisms in tumor necrosis factor-alpha, interleukin-10, and transforming growth factor-beta1 genes in acute hepatic allograft rejection. Transplantation 69:1514, 2000

55. Turner D, Grant SC, Yonan N, et al: Cytokine gene polymorphism and heart transplant rejection. Transplantation 64:776, 1997

56. Asano H, Kobayashi T, Uchida K, et al: Significance of tumor necrosis factor microsatellite polymorphism in renal transplantation. Tissue Antigens 50:484, 1997

57. Middleton PG, Taylor PR, Jackson G, et al: Cytokine gene polymorphisms associating with severe acute graft-versus-host disease in HLA-identical sibling transplants. Blood 92:3943, 1998

58. Sahoo S, Kang S, Supran S, et al: Tumor necrosis factor genetic polymorphisms correlate with infections after renal transplantation. Transplantation 69:880, 2000

59. Densem CG, Hutchinson IV, Yonan N, et al: Tumour necrosis factor alpha gene polymorphism: a predisposing factor to non-ischaemic myocardial dysfunction? Heart 87:153, 2002

60. Filler G, Yang F, Martin A, et al: Renin angiotensin system gene polymorphisms in pediatric renal transplant recipients. Pediatr Transplant 5:166, 2001

61. Juckett MB, Cohen EP, Keever-Taylor CA, et al: Loss of renal function following bone marrow transplantation: an analysis of angiotensin converting enzyme D/I polymorphism and other clinical risk factors. Bone Marrow Transplant 27:451, 2001

62. Viklicky O, Hubacek JA, Pitha J, et al: ACE gene polymorphism and long-term renal graft function. Clin Biochem 34:87, 2001

63. Pethig K, Heublein B, Hoffmann A, et al: ACE-gene polymorphism is associated with the development of allograft vascular disease in heart transplant recipients. J Heart Lung Transplant 19:1175, 2000

64. Barocci S, Ginevri F, Valente U, et al: Correlation between angiotensin-converting enzyme gene insertion/deletion polymorphism and kidney graft long-term outcome in pediatric recipients: a single-center analysis. Transplantation 67:534, 1999

65. Broekroelofs J, Stegeman CA, Navis G, et al: Risk factors for long-term renal survival after renal transplantation: a role for angiotensin-converting enzyme (insertion/deletion) polymorphism? J Am Soc Nephrol 9:2075, 1998

66. Pullmann R Jr, Lukac J, Skerenova M, et al: Association between systemic lupus erythematosus and insertion/deletion polymorphism of the angiotensin converting en-

zyme (ACE) gene. Clin Exp Rheumatol 17:593, 1999

67. Constantinescu CS, Goodman DB, Ventura ES: Captopril and lisinopril suppress production of interleukin-12 by human peripheral blood mononuclear cells. Immunol Lett 62:25, 1998

68. Poch E, Gonzalez D, Giner V, et al: Molecular basis of salt sensitivity in human hypertension: evaluation of renin-angiotensin-aldosterone system gene polymorphisms. Hypertension 38:1204, 2001

69. Abdi R, Tran TB, Zee R, et al: Angiotensin gene polymorphism as a determinant of posttransplantation renal dysfunction and hypertension. Transplantation 72:726, 2001

70. Kranzhofer R, Browatzki M, Schmidt J, et al: Angiotensin II activates the proinflammatory transcription factor nuclear factor-kappaB in human monocytes. Biochem Biophys Res Commun 257:826, 1999

71. Segerer S, Cui Y, Eitner F, et al: Expression of chemokines and chemokine receptors during human renal transplant rejection. Am J Kidney Dis 37:518, 2001

72. Melter M, Exeni A, Reinders ME, et al: Expression of the chemokine receptor CXCR3 and its ligand IP-10 during human cardiac allograft rejection. Circulation 104:2558, 2001

Acknowledgment

Figures 1 through 5 Seward Hung.

VIII DEFICIENCIES IN IMMUNOGLOBULINS AND CELL-MEDIATED IMMUNITY

FRED S. ROSEN, M.D.

Immunoglobulin Deficiency Syndromes

Insufficient production of one or more kinds of antibodies characterizes the immunoglobulin deficiency syndromes [*see Table 1*].[1,2] Patients with these deficiencies are subject to recurrent pyogenic infections, such as otitis media, sinusitis, and pneumonia. Repeated episodes of pneumonia can lead to chronic obstructive pulmonary disease. For many of these deficiencies, the genetic basis has now been defined. The primary care physician's role in these disorders is to suspect the diagnosis under the appropriate clinical circumstances—often, unusual susceptibility to certain infections in a patient with a family history of the same—and to order the preliminary laboratory studies. Definitive diagnosis and management is typically the responsibility of the immunologist. Control of the infections to which these patients are susceptible is principally managed by the intravenous administration of large doses of γ-globulin.

X-LINKED AGAMMAGLOBULINEMIA

X-linked agammaglobulinemia, also known as congenital agammaglobulinemia or Bruton disease, was the first immunodeficiency disorder to be described, in 1952.

Genetics and Pathogenesis

The gene responsible for X-linked agammaglobulinemia is located on the long arm of the X chromosome (Xq21.33–q22).[3-5] This gene, termed *btk*, is a member of the *src* family of oncogenes and encodes a unique tyrosine kinase.[4-8] It probably plays a critical role in the maturation of B cells: pre–B cells are present in normal numbers in the bone marrow of males with X-linked agammaglobulinemia, but they do not develop into mature B cells.[2] Because the genes governing the structure of immunoglobulins are on autosomal chromosomes, the mechanism of the disorder must also involve a defect in a regulatory gene.

In patients with X-linked agammaglobulinemia, the lymphoid organs are characterized by a lack of germinal follicles, B cells, and plasma cells. On bone marrow studies, pre–B cells (which contain immunoglobulin μ heavy chains in their cytoplasm and therefore can be identified by immunofluorescence staining with antiserum to the μ chain) are present in normal numbers.

Diagnosis

Clinical manifestations Because infants are born with IgG from their mother in their blood, boys who have X-linked agammaglobulinemia do not start to show the effects of the disorder

Table 1 Primary Specific Immunodeficiencies Involving Antibodies

Designation	*Usual Phenotypic Expression*		*Presumed Level of Basic Cellular Defect*	*Known or Presumed Pathogenetic Mechanism*	*Inheritance*
	Antibody Deficiencies	*Cellular Abnormalities*			
X-linked agammaglobulinemia	All immunoglobulins	↓ B cells	Pre–B cells	Mutations in the gene for Bruton's X-linked tyrosinase (*btk*)	X-linked
Common variable immunodeficiency	All immunoglobulins	Faulty B cell maturation	Immaturity of B cells	↓ Helper T cell function Intrinsic B cell defect Underproduction of B cells Autoantibodies to B cells	Unknown
Selective IgA deficiency	IgA	↓ IgA plasma cells ± ↑ IgA^+ B cells	Terminal differentiation of IgA^+ B cells impaired	Unknown	Usually unknown (autosomal recessive more common than autosomal dominant); frequent in families of patients with common variable immunodeficiency
Ig deficiencies, with increased IgM	IgG, IgA, and IgE	↓ IgG and IgA plasma cells ↑ IgM and IgD plasma cells ± ↑ IgM^+ B cells	Failure of immunoglobulin class switching	X-linked form: mutations in the gene for the CD40 ligand Autosomal recessive form: activation-induced cytidine deaminase	X-linked, autosomal recessive, or unknown
Selective deficiency of IgG subclasses	One or more IgG isotypes	↓ Plasma cells ± ↓ T cells	Unknown	Unknown	Unknown
κ-Chain deficiency	IgG(κ)	↓ $κ^+$ B cells	Unknown	Point mutation at 2p11	Autosomal recessive
Transient hypogammaglobulinemia of infancy	IgG and IgA	↓ Plasma cells B cells normal	Impaired terminal differentiation of B cells	↓ Helper T cells	Frequent in heterozygous individuals in families with various severe combined immunodeficiencies

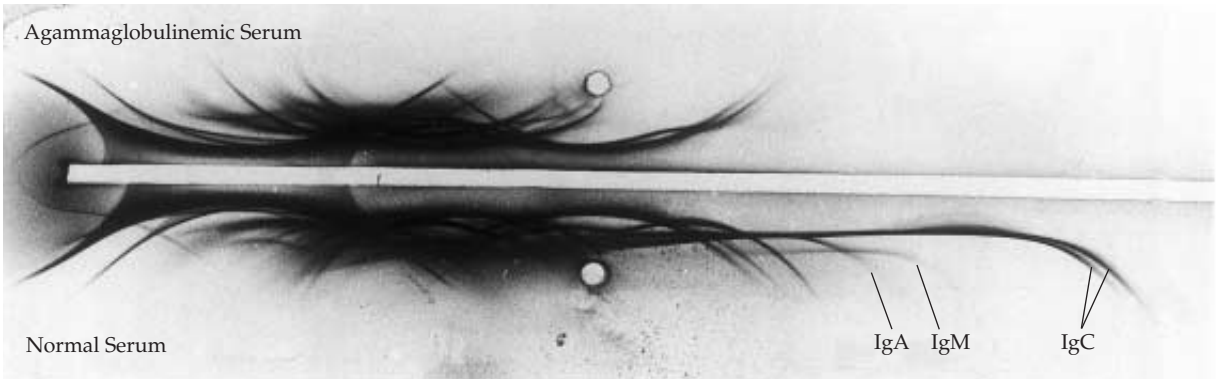


Figure 1 **When an immunoelectrophoretic pattern of agammaglobulinemic serum is compared with 6a normal serum pattern, the absence of IgA, IgM, and IgG—characteristic of the disorder—is clearly demonstrated.**

until 6 to 15 months of age. They then demonstrate unusual susceptibility to infections by pyogenic organisms (e.g., otitis media, sinusitis, and pneumonia from *Haemophilus influenzae*, pneumococci, streptococci, staphylococci, and meningococci). Those infections are more frequent and more severe in boys with X-linked agammaglobulinemia than in normal children, and recurrent infection by the same organism is common. Frequently, the infections are slow to respond to antibiotics. Recurrent pulmonary infections often lead to bronchiectasis and pulmonary insufficiency. Affected males have normal resistance to the common viral diseases, fungi, and most gram-negative organisms, but some have developed polio after receiving oral polio vaccine. About one third of patients have symptoms that resemble rheumatoid arthritis, including swollen and painful joints. A severe late complication is a fatal syndrome similar to dermatomyositis but with central neurologic involvement, as well. This syndrome is gradual in onset, usually starting in the second or third decade of life. In several patients with this syndrome, echoviruses have been cultured from the blood, stool, and cerebrospinal fluid.[9]

Laboratory testing Diagnosis begins with measuring the serum level of each class of immunoglobulin [*see Figure 1*]. Patients with X-linked agammaglobulinemia usually have less than 100 mg/dl of IgG (normal levels are 614 to 1,295 mg/dl), and they have levels of IgA, IgM, IgD, and IgE that are extremely low or undetectable. Such findings should prompt referral of the patient to an immunologist.

In patients with X-linked agammaglobulinemia, analysis of white blood cells by flow cytometry reveals a lack of B cells. These patients are unable to mount an antibody response to antigen challenge, such as routine diphtheria-pertussis-tetanus (DPT) or *H. influenzae* vaccination, and they cannot neutralize the toxin in a Schick test (intradermal injection of diphtheria toxin). In contrast, cell-mediated immune functions, such as delayed hypersensitivity–mediated skin reactions and graft rejection, are essentially normal, and the T cells respond in vitro to phytohemagglutinin and produce lymphokines normally.

Screening All subsequent male offspring of the mother or maternal aunts of a patient with X-linked agammaglobulinemia should be screened for mutations of the *btk* gene. Because the defect is limited to B cells, female carriers of the gene can be detected by analysis of X-chromosome inactivation in B cells.[10,11] In female carriers, pre–B cells in which the X chromosome bearing the normal gene has been inactivated will not develop into B cells; therefore, all mature B cells will bear an active X chromosome containing only the normal gene.

Treatment

Preparations of 5% or 10% γ-globulin solution are now used as replacement therapy for agammaglobulinemia. Parents can be reassured that these preparations pose no risk of transmitting HIV or other viral infection. Intravenous administration of these preparations is well tolerated; large doses can be given without discomfort or pain. Infants do not require permanent intravenous access.

Dosages of γ-globulin are adjusted according to the patient's health. The minimal effective dosage of intravenous γ-globulin is 300 mg/kg a month; however, higher doses, such as 500 mg/kg a month, are usually optimal.[12] Dividing the monthly dosage of γ-globulin and administering it at 1-week or 2-week intervals is preferable, because it maintains higher immunoglobulin levels. The γ-globulin is infused at a rate of 3 ml/min or slower. Side effects may include headache, shaking chills, flank pain, fever, and hypotension. These can be ameliorated by giving an antihistamine or methylprednisolone before the infusion.

Bacterial infections in patients with X-linked agammaglobulinemia require vigorous antibiotic treatment. Antibiotics should be given in prolonged courses (e.g., 2 weeks) at full doses.

Prognosis The prognosis is very good for patients whose condition is diagnosed and treated early. A recent study of 31 patients with X-linked agammaglobulinemia found that early and prolonged γ-globulin replacement therapy is effective in preventing bacterial infections and pulmonary insufficiency. Viral infections still developed, however, and one patient died of enteroviral meningoencephalopathy.[13]

COMMON VARIABLE IMMUNODEFICIENCY

Common variable immunodeficiency (CVID) is so called because it accounts for over 50% of cases of immunodeficiency and because patients present with variable clinical manifestations and somewhat inconsistent laboratory findings; disease course varies, as well.

Etiology and Pathogenesis

The cause of CVID is unknown. CVID does not appear to be genetically transmitted—apparently the germ cells are not in-

volved—although some family clusters have been seen. CVID affects males and females equally.

A variety of pathogenetic mechanisms underlie CVID.[2] These include (1) B cells that do not respond to stimulatory signals from T cells, (2) B cells that can synthesize but cannot secrete immunoglobulins, (3) the absence of helper T cells (required for normal B cell function), and (4) the presence of autoantibodies to B cells. In a few cases of CVID, B cells cannot be detected. All patients show markedly low serum levels of all immunoglobulins.

Diagnosis

Clinical manifestations Onset of CVID can occur at any age, but it usually occurs after puberty. Patients have the same heightened vulnerability to infections as those with X-linked agammaglobulinemia; also, there is chronic involvement of the sinuses and respiratory tract.

CVID is associated with several autoimmune diseases, such as rheumatoid arthritis, idiopathic thrombocytopenia, hemolytic anemia, neutropenia, and, predominantly, pernicious anemia. Infectious diarrhea and malabsorption syndrome are common. CVID is also associated with severe malabsorption syndrome caused by gluten-sensitive enteropathy. It is unclear whether CVID is a cause or an effect of these disorders. Chronic lung disease that produces bronchiectasis is common in CVID; this condition should be differentiated from cystic fibrosis, chronic allergy, and α_1-antitrypsin deficiency. In contrast to X-linked agammaglobulinemia, CVID is often marked by considerable enlargement of regional lymph nodes and splenomegaly.

Laboratory tests IgG levels in patients with CVID are generally lower than 250 mg/dl, and other immunoglobulins are also markedly decreased. B cells are usually present, but they do not mature normally into plasma cells, which synthesize and secrete immunoglobulins. Tests of cell-mediated immunity also demonstrate defects.

Lymphoid hyperplasia may occur in the gut of patients with CVID. This can be visualized by barium contrast x-ray of the upper GI tract, which is indicated in CVID patients with GI symptoms.

Treatment

Treatment of CVID is essentially the same as that of X-linked agammaglobulinemia: replacement γ-globulin therapy and vigorous use of antibiotics during acute infections. Diarrhea in these patients is frequently caused by *Giardia lamblia* infection, which can be rapidly controlled with quinacrine hydrochloride or metronidazole.[11] Special care must be taken if steroids are used as therapy for the associated autoimmune diseases, because these agents may heighten susceptibility to infection.

Prognosis

Patients with CVID can have a normal life span. Women with the disease can carry a normal pregnancy to term and have normal babies. Although those babies will lack maternal IgG and the passive immunity it confers in the first months of life, they do well without treatment with γ-globulin.

SELECTIVE IMMUNOGLOBULIN DEFICIENCIES

Selective IgA Deficiency

Epidemiology Selective IgA deficiency is one of the most common immunodeficiencies in whites, occurring in one in 600 to 800 persons in this population. It does not occur in Africans and almost never occurs in Asians.

Genetics and pathogenesis The genetics of IgA deficiency are unclear. Data on inheritance are conflicting, with some suggesting autosomal dominant inheritance and others suggesting autosomal recessive inheritance.

A few patients lacking serum IgA have secretory IgA, and some patients have monomeric IgM in their secretions. B cells bearing surface IgA are present, indicating that the defect is probably in the terminal differentiation of IgA-secreting cells. In vitro, IgA-bearing cells can be stimulated by mitogens to produce IgA.[14]

Diagnosis Many patients with IgA deficiency are surprisingly healthy. Nevertheless, IgA deficiency is associated with many clinical syndromes. Patients most often come to medical attention because of recurrent sinus and pulmonary infection by bacteria and viruses. These patients also show a higher incidence of autoimmune, GI, allergic, connective tissue, and malignant diseases. Some patients with IgA deficiency produce antibodies to bovine proteins, suggesting that IgA in the gut normally helps prevent absorption of foreign antigens. IgA deficiency is found in about 70% of patients with ataxia-telangiectasia (see below).

The serum IgA level is less than 5 mg/dl (normal, 60 to 309 mg/dl). Other immunoglobulin levels are normal. Although patients with IgA deficiency usually also have defects in T cell function, most of these patients have normal cell-mediated immunity.

Treatment There is currently no satisfactory means of supplying adequate levels of IgA. Sinus and pulmonary infections in IgA-deficient patients are treated by standard means.

Complications In extremely rare instances, patients with IgA deficiency produce IgE antibodies to IgA and will have anaphylactic reactions when given immunoglobulin.[15] Immunoglobulin replacement therapy should be avoided in such patients; blood transfusion can also precipitate an anaphylactic reaction. Patients who require blood should receive red cells from an IgA-deficient donor because anaphylactic reactions may occur even if the red blood cells are washed three times.

Immunoglobulin Deficiency with Elevated IgM

The combination of markedly elevated IgM levels and deficiency of other immunoglobulins is termed the hyper-IgM syndrome. The IgM in these patients is heterogeneous; thus, it is polyclonal and does not emerge from malignant cells.

In 70% of hyper-IgM cases, the syndrome is X-linked; in the remainder, it is autosomal recessive and affects both males and females. The X-linked form of the hyper-IgM syndrome results from a genetic defect in the CD40 ligand, which is found on the surface of activated T cells.[16-18] Normally, this ligand interacts with the CD40 molecule on the B cell surface, inducing isotype switching. The autosomal recessive form of the hyper-IgM syndrome results from a genetic defect in an enzyme called activation-induced cytidine deaminase (AID).[19] This enzyme is involved in RNA editing, but its precise role in immunoglobulin class switching is unknown.

Diagnosis Patients with hyper-IgM syndrome show increased susceptibility to infection similar to that seen in X-linked

agammaglobulinemia (see above). Immunoglobulin assays show an elevated level of IgM (350 to 1,000 mg/dl); the IgD level may also be elevated. IgA is usually undetectable, and the IgG level is normally less than 100 mg/dl. Many plasma cells, as well as lymphocytoid and plasmacytoid cells structurally similar to those of Waldenström macroglobulinemia, are seen in the gut, lymphoid organs, and blood. These plasma cells stain with fluorescein-labeled antibodies to IgM. In the X-linked form of hyper-IgM syndrome, lymph nodes are small and contain no germinal centers. In AID deficiency, lymph nodes are enlarged and contain germinal centers. Lymph node biopsy is not usually obtained for clinical diagnosis, however.

Treatment Treatment for hyper-IgM syndrome is the same as that for X-linked agammaglobulinemia (see above).

Selective Deficiencies of IgM or the Subclasses of IgG

Selective IgM deficiency is rare. This deficiency may precede the onset of CVID. Patients with selective deficiencies of the IgG subclasses have a decrease in total IgG, the degree of which depends on the subclass involved. The decrease is most profound in the case of IgG1 deficiency because almost three quarters of IgG molecules belong to this subclass. Some patients with IgG deficiency are unable to mount an antibody response to certain antigens. Patients with IgG2 deficiency are especially prone to infection by bacteria with a large amount of surface polysaccharide, such as pneumococci and *H. influenzae*. The diagnosis is confirmed by quantitation of the IgG subclasses and administration of a polysaccharide-antigen vaccine (typically, pneumococcal vaccine); patients with IgG deficiency will fail to produce antibodies in response to vaccination. Patients with selective deficiencies of the IgG subclasses respond to intravenously administered γ-globulin.

Deficiencies of Cell-Mediated Immunity

Extreme susceptibility to opportunistic infection is the most important clinical feature of deficiencies of cell-mediated immunity, or T cell deficiencies. Such deficiencies, which manifest as impairment in delayed hypersensitivity, may be inherited or may be secondary to another disorder [*see Table 2*]. The acquired immunodeficiency syndrome is discussed elsewhere [*see 7:XXXIII HIV and AIDS*].

In general, patients with T cell deficiencies have more frequent and more severe infections than do patients who have pure B cell deficiencies [*see Table 3*].[2] Patients with deficiencies of cell-mediated immunity cannot cope with a number of ordinarily innocuous organisms, such as *Candida albicans* and *Pneumocystis carinii*, and are especially susceptible to enteric bacteria, viruses, and fungi. Live attenuated vaccines are dangerous in these patients: vaccination for smallpox or administration of bacillus Calmette-Guérin (BCG) has led to rapid death.

Determining the defects of cell-mediated immunity requires testing in a specialized immunology laboratory. An extensive array of tests is available at such laboratories [*see Table 4*]. The choice of tests and the order in which they are performed depend on the particular case.

Table 2 Conditions Associated with Impaired Delayed Hypersensitivity

Primary deficiencies of cell-mediated immunity [*see Table 3*]
Chromosomal abnormalities: Bloom syndrome, Down syndrome, Fanconi syndrome
Infections: HIV (AIDS), lepromatous leprosy, Epstein-Barr virus (X-linked lymphoproliferative syndrome), chronic mucocutaneous candidiasis, secondary syphilis, and many other viral and parasitic diseases
Neoplasms: thymoma, Hodgkin disease and other lymphomas, any advanced malignant disease
Connective tissue diseases: systemic lupus erythematosus, advanced rheumatoid arthritis
Physical agents: burns, x-irradiation
Other conditions: sarcoidosis, malnutrition, aging, inflammatory bowel disease, intestinal lymphangiectasia
Iatrogenic causes: chemotherapy, postsurgery, x-irradiation therapy

CONGENITAL THYMIC HYPOPLASIA

Pathogenesis

Congenital thymic hypoplasia (DiGeorge syndrome) results from the lack of normal development of the third and fourth brachial, or pharyngeal, pouches, which leads to abnormality in the great vessels and to the absence of the thymus and the parathyroids. Congenital thymic hypoplasia is not inherited; rather, it is thought to result from an intrauterine accident occurring before the eighth week of pregnancy. The absence of the thymus leads to deficiency in cell-mediated immunity.

Diagnosis

Clinical manifestations Patients with congenital thymic hypoplasia have distinctive facial features, including low-set ears, a shortened philtrum, and ocular hypertelorism. Hypocalcemia from associated parathyroid deficiency is a universal finding and often results in neonatal tetany. There can be a right-sided aortic arch or tetralogy of Fallot or many other cardiac malformations.

Laboratory tests The T cell defect in children with congenital thymic hypoplasia varies from mild to profound. Severely affected children do not exhibit delayed hypersensitivity reactions; their lymphocytes do not respond to mitogens or antigens in vitro, nor do they produce lymphokines. The lymph nodes lack paracortical lymphocytes. Plasma cells are present, however, and immunoglobulin levels are normal. Although patients with congenital thymic hypoplasia produce specific antibodies when they are immunized with various antigens, the antibody response is not quite normal, because secondary responses are lacking.

As the patient ages, T cell function improves; and usually by the time the child is 5 years of age, skin testing reveals no abnormality in cell-mediated immunity. However, the abnormal T cell phenotype—as indicated by a higher than normal ratio of $CD4^+$ to $CD8^+$ T cells—persists for life. Karyotyping reveals microdeletions at chromosome 22q11 in approximately 90% of patients.

Treatment

Thymus transplantation should be undertaken in those infants with congenital thymic hypoplasia who experience frequent infections. Transplantation of fetal thymus results in rapid acquisition of normal T cell function, which is thought to be secondary to production of a thymic hormone secreted by the thymic epithelium. Rejection appears not to be a problem.

Table 3 Classification of Primary Specific Immunodeficiencies Involving Cell-Mediated Immunity

Designation	*Usual Phenotypic Expression*		*Presumed Level of Basic Cellular Defect*	*Known or Presumed Pathogenetic Mechanism*	*Inheritance*	*Main Associated Features*
	Functional Deficiencies	*Cellular Abnormalities*				
Congenital thymic hypoplasia (DiGeorge syndrome)	CMI, impaired antibody	↓ T cells	Thymocytes	Embryopathy of third and fourth pharyngeal pouch areas	Usually not familial	Hypoparathyroidism Abnormal facies Cardiovascular abnormalities
Severe combined immunodeficiency	CMI, antibody	- T cells, + B cells	LSC	Mutation in γ chain of IL-2R, IL-4R, IL-7R, IL-11R, IL-15R, *JAK3*, or IL-7 receptor α chain	X-linked or autosomal recessive	—
		- T cells, - B cells		Mutation in *RAG1* or *RAG2*	Autosomal recessive	
Adenosine deaminase (ADA) deficiency	CMI, antibody	↓ T cells, ± B cells	LSC or early T cells	Metabolic effects of ADA deficiency	Autosomal recessive	—
Purine nucleoside phosphorylase (PNP) deficiency	CMI ± antibody	↓ T cells	T cells	Metabolic effects of PNP deficiency	Autosomal recessive	Hypoplastic anemia
Reticular dysgenesis	CMI, antibody, phagocytes	↓ T cells, ↓ B cells, ↓ phagocytes	HSC	Unknown	Autosomal recessive	Neutropenia
Wiskott-Aldrich syndrome	Antibody to certain antigens (mainly polysaccharides), CMI (progressive)	↓ T cells, ↑ B cells (progressive)	HSC	Mutations in *WASP* gene	X-linked	Thrombocytopenia Eczema Lymphoreticular cancers
Immunodeficiency with ataxia-telangiectasia	CMI, antibody (partial)	↓ T cells, ↓ plasma cells (mainly those cells producing IgA, IgE, ± IgG)	Defective checkpoints in T and B cell division	Mutations in *ATM* gene	Autosomal recessive	Cerebellar ataxia Telangiectasia Chromosomal abnormalities Raised serum α-fetoprotein levels
MHC class II deficiency	CMI ± antibody	None	T cells, B cells, and antigen-presenting cells	Defects of promoter proteins	Autosomal recessive	Intestinal malabsorption
CD3 deficiency	CMI	None	T cells	Mutations in CD3-ε or CD3-γ	Autosomal recessive	—
CD8 deficiency	CMI	↓ $CD8^+$ T cells, normal number of $CD4^+$ cells	Early T cells	Mutations in *ZAP* genes	Autosomal recessive	—

CMI—cell-mediated immunity LSC—lymphocytic stem cell HSC—hematopoietic stem cell

SEVERE COMBINED IMMUNODEFICIENCY

Severe combined immunodeficiency disease (SCID) is characterized by marked depletion of cells that mediate both humoral and cellular immunity—B cells and T cells, respectively. SCID is fatal if left untreated.

Several variants of SCID have been identified. They are designated as T^-B^- or T^-B^+, depending on whether B cells are normal or increased (B^+) or absent (B^-). In addition to the extent of B cell involvement, the variants also differ in the site of the basic cellular defect, the pathogenetic mechanism, and the mode of inheritance [*see Table 3*].

Genetics and Pathogenesis

T^-B^+ SCID may be transmitted as either an X-linked or an autosomal recessive trait. The specific genetic defect responsible for the X-linked form of T^-B^+ SCID results from mutations in the γ chain of the interleukin-2 receptor (IL-2R),[20] whose gene is localized to the long arm of the X chromosome at Xq13.[21] This γ chain is also found in the receptors for IL-4, IL-7, IL-11, and IL-15.[22] Engagement of the IL-7 receptor by IL-7 is required for T cell maturation, so precursor T cells in these patients do not mature.

When any of those receptors, or IL-2R, are engaged by its ligands, a cytoplasmic tyrosine kinase (Janus-family tyrosine kinase, or JAK3) bound to the γ chain is activated. The gene encoding JAK3 is on an autosome, not the X chromosome. Thus, autosomal recessive T^-B^+ SCID is caused by mutations in the *JAK3* gene.[23,24]

T^-B^- SCID is inherited in an autosomal recessive manner. About half of the cases are caused by a deficiency in the enzyme adenosine deaminase (ADA),[25] and another large fraction results from mutations in the recombination-activating genes *RAG-1* and *RAG-2*.[26] These recombinase enzymes are required for the gene rearrangements that occur before T cell receptor or immunoglobulin synthesis. Other patients with autosomal recessive T^-B^- SCID lack the enzyme purine nucleoside phosphorylase (PNP).[27]

ADA deficiency leads to an accumulation of adenosine, adenosine triphosphate (ATP), and deoxy-ATP (dATP). It has

Table 4 Laboratory Tests Used to Determine Deficiencies of Cell-Mediated Immunity

Skin test: 24- to 48-hr reaction to *Candida*, *Trichophyton*, PPD
Response to nonspecific mitogens: phytohemagglutinin, concanavalin A, pokeweed mitogen
Response to specific mitogens: diphtheria, tetanus, *Candida*
Response to alloantigens: mixed lymphocyte reaction
When responses to alloantigens and nonspecific and specific mitogens are negative: repeat tests while stimulating cells with IL-2
Enumerate T cells with monoclonal antibody to CD3, with or without a cell sorter
Enumerate T cell subsets with monoclonal antibody to CD4 for helper T cells and with monoclonal antibody to CD8
Enumerate T cells positive for Ia (class II) antigens (which measures the number of activated T cells)
Quantitate IL-2 receptors with monoclonal antibody TAC
Quantitate IL-2 and interferon-gamma synthesis
Enumerate NK cells with monoclonal antibodies Leu-7 and Leu-11
Assay NK cell activity using cell line K-562
Assay cytotoxic T cell activity using cell lines of cloned T cells
Enumerate monocytes with monoclonal antibody Mo-1
Assay for IL-1 production by stimulated monocytes
Determine serum level of anti–T cell antibodies
Determine if antibody to HIV is present
HLA typing
Assay erythrocytes for adenosine deaminase and purine nucleoside phosphorylase activity
Detect thymus shadow on x-ray

Note: All patients with defects in cell-mediated immunity should receive all tests listed, except the last three, for optimal examination. The last three tests are for patients suspected of having severe combined immunodeficiency or congenital thymic hypoplasia. HLA typing is needed for prospective recipients of bone marrow transplants.
IL-1—interleukin-1 IL-2—interleukin-2 PPD—purified protein derivative of tuberculin NK—natural killer

been shown that dATP poisons ribonucleotide reductase, an enzyme required for DNA synthesis. Thus, lymphocytes lacking ADA cannot divide until the dATP overload is decreased or removed. In a similar manner, lymphocytes that lack PNP accumulate guanosine, guanosine triphosphate (GTP), and deoxy-GTP, causing metabolic abnormalities that resemble those seen in ADA deficiency. SCID caused by ADA or PNP deficiency can be diagnosed prenatally by amniocentesis because fibroblasts in the amniotic fluid also show the enzymatic defect.

CD8 deficiency is a rare form of SCID that results from mutations in the *ZAP-70* gene.[28,29] ZAP-70 is a tyrosine kinase that binds to the CD3 chain and is involved in signal transduction from the T cell receptor (the TCR-CD3 complex). $CD8^+$ T cells fail to mature, and mature $CD4^+$ T cells fail to function as a result of the mutations in *ZAP-70*.

Another variant of SCID is reticular dysgenesis, a severe combined immunodeficiency with a generalized granulocyte deficiency. Newborns with this disease lack granulocytes in the blood and bone marrow and die of infection within the first few days of life.

Diagnosis

Clinical manifestations Chronic pulmonary infections, diarrhea, moniliasis, and failure to thrive are the most common manifestations of SCID. The lymph nodes are small to absent despite chronic infections, which usually begin at 3 to 6 months of age.

Laboratory tests Complete blood counts show a low number of lymphocytes. There is absence of a thymic shadow on chest x-ray. (Autopsy in fatal cases has revealed an embryonic thymus that resembles the thymus at 6 weeks of gestation, before invasion with lymphocytes.) Tests for cutaneous delayed hypersensitivity and contact sensitization and in vitro assays of blood lymphocytes are negative, demonstrating the absence of T cells, a phytohemagglutinin response, and lymphokine production. Antibody levels are usually low, although occasionally the IgM level is normal; and sometimes, a myeloma component is seen.

Treatment

Hundreds of cases of SCID have been successfully treated by transplantation of bone marrow cells.[30] By 3 to 8 months after receiving the bone marrow, these patients show normal delayed hypersensitivity and T cell function and are no longer abnormally susceptible to infection.

Immunologic reconstitution with bone marrow cells should be attempted only in specialized centers where comprehensive histocompatibility typing and intensive 24-hour care can be given. If the donor and the recipient are not exceedingly well matched, fatal graft versus host disease (GVHD) will ensue. Even an HLA-mismatched blood transfusion can produce fatal GVHD in such patients: the patient is immunocompromised and thus cannot reject the injected cells, but the histoincompatible cells that have been administered recognize the patient's cells as foreign and react against them.

The manifestations of GVHD include fever, diarrhea, depression of the bone marrow, splenomegaly, and an erythematous rash on the face, trunk, and extremities. The reaction eventually leads to death. GVHD can be avoided by irradiating the blood before transfusion.

It is possible to establish grafts of half-matched (haploidentical) parental marrow in infants with SCID. GVHD can be avoided in those cases if the parental marrow is depleted of T cells before transplantation by passage over lectin columns or by treatment with anti–T cell monoclonal antibody plus complement.

Patients with ADA deficiency have also been treated successfully with infusions of purified adenosine deaminase modified with polyethylene glycol. The ADA gene has been cloned and inserted into a retroviral vector.[31] In a few ADA-deficient children, this vector has been transfected into peripheral blood lymphocytes, which were then reinfused. This gene therapy procedure has corrected the immunodeficiency in these patients, although it must be repeated periodically.[32] Successful gene therapy has also been carried out in X-linked SCID by transducing a Maloney virus vector bearing the gene for the common γ chain into bone marrow cells. Sustained responses have been reported in four of these patients: T cell number and function normalized in these patients, as did B cell function, and infusions of γ globulin were no longer required.[33]

WISKOTT-ALDRICH SYNDROME

An X-linked recessive disease, Wiskott-Aldrich syndrome (WAS), results from a mutation that has been mapped to the Xp11.3–p11.22 region of the X chromosome. The *WAS* gene has been cloned.[34,35]

The lymphoid system of a patient with WAS appears anatomically intact at birth. Starting in the first months of life, however, there is a decrease in T cells in the paracortical areas of the lymph nodes and a polyclonal expansion of B cells. The T cells in these patients respond poorly to mitogens. The protein encoded by the *WAS* gene appears to be involved in signal transduction that

leads to reorganization of the cytoskeleton when lymphocytes are stimulated, which results in defective collaboration between T cells and B cells. Lymphocytes have a markedly abnormal appearance when visualized by scanning electron microscopy. Platelets are abnormally small and few in number.[36] Certain missense mutations in the *WAS* gene lead to a mild disease called X-linked thrombocytopenia.[37]

Diagnosis

Clinical manifestations WAS is characterized by eczema, easy bruising, increased susceptibility to infection (both pyogenic and opportunistic), and bloody diarrhea. These manifestations appear in the first months of life. An increased incidence of hematopoietic malignancies is seen, starting in the second or third decade of life.

Laboratory testing Patients with WAS have normal levels of IgG, high levels of IgE and IgA, and low levels of IgM. Severe thrombocytopenia is universal. Tests of cell-mediated immunity [*see Table 4*] show a variety of abnormalities: WAS patients lack isohemagglutinins and are unable to make antibodies to polysaccharides. They respond to some protein antigens but not to others; in addition, they may exhibit anergy and may not display positive results to skin tests for the usual bacterial or fungal antigens.

Treatment

WAS patients have been treated with marrow transplantation after receiving irradiation or busulfan and antilymphocyte serum to destroy residual lymphocytes; they have then shown normal immune and platelet functions. In WAS patients who do not receive a bone marrow transplant and who experience severe bleeding from thrombocytopenia, splenectomy may be considerably beneficial.[38]

IMMUNOLOGIC DEFICIENCY WITH ATAXIA-TELANGIECTASIA

Ataxia-telangiectasia (A-T) is a disease associated with defects in cell-mediated immunity and with immunoglobulin deficiencies. It is inherited as an autosomal recessive trait. The gene for A-T (*ATM* for A-T mutated) maps to the chromosomal region 11q22.3.[39,40] Normally, the gene appears to function in repair of breaks in double-stranded DNA. Patients with A-T have a disorder of the cell-cycle checkpoint pathway that results in an extreme hypersensitivity to ionizing radiation. Consequently, frequent chromosomal breaks, inversions, and translocations are observed. Postmortem examination may disclose abnormalities in the thymus, which is small and deficient in lymphocytes. There also may be an abnormality in lymph node structure.

Diagnosis

Clinical manifestations A-T presents as a progressive neurologic disease that begins in early childhood. It is characterized by cerebellar ataxia, starting at 18 months of age, followed by increasing tremor and deterioration of mental function. By 5 years of age, progressive telangiectasia is seen in the vessels of the bulbar conjunctiva and is later visible on the skin. The immune deficiencies in these patients leads to recurrent sinus and bronchial infections and subsequent bronchiectasis. An unusually high incidence of lymphoid malignant disorders has been reported in patients with A-T.[41]

Laboratory testing About 70% of patients with A-T have a severe deficiency in IgA. On tests of cell-mediated immunity [*see Table 4*], some A-T patients are anergic and fail to show delayed hypersensitivity responses to common microbial antigens. They may also have abnormal in vitro cell-mediated immune responses and may tolerate allografts.

Treatment and Prognosis

No satisfactory treatment for A-T is currently available. Persons with A-T who survive into their second decade may fail to mature sexually. A-T patients usually die of lymphoid malignancies or other causes by the end of their second decade.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Rosen FS, Wedgwood RJ, Eibl M, et al: Primary immunodeficiency diseases: report of a WHO Scientific Group. Clin Exp Immunol 109(suppl 1):1, 1997

2. Rosen FS, Cooper MD, Wedgwood RJP: The primary immunodeficiencies. N Engl J Med 333:431, 1995

3. Kwan SP, Terwilliger J, Parmley R, et al: Identification of a closely linked DNA marker, DXS178, to further refine the X-linked agammaglobulinemia locus. Genomics 6:238, 1990

4. Vetrie D, Vorechovsky I, Sideras P, et al: The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature 361:226, 1993

5. Tsukada S, Saffran DC, Rawlings DJ, et al: Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell 72:279, 1993

6. Hagemann TL, Chen Y, Rosen FS, et al: Genomic organization of the Btk gene and exon scanning for mutations with X-linked agammaglobulinemia. Hum Mol Genet 3:1743, 1994

7. Zhu Q, Zhang M, Winkelstein J, et al: Unique mutations of Bruton's tyrosine kinase in fourteen unrelated X-linked agammaglobulinemia families. Hum Mol Genet 3:1899, 1994

8. Conley ME, Fitch-Hilgenberg ME, Cleveland GL, et al: Screening of genomic DNA to identify mutations in the gene for tyrosine kinase. Hum Mol Genet 3:1751, 1994

9. Misbah SA: Chronic enteroviral meningoencephalitis in agammaglobulinemia: case report and literature review. J Clin Immunol 12:266, 1992

10. Fearon ER, Winkelstein JA, Civin CI, et al: Carrier detection in X-linked agammaglobulinemia by analysis of X-chromosome inactivation. N Engl J Med 316:427, 1987

11. Conley ME, Brown P, Pickard AR, et al: Expression of the gene defect in X-linked agammaglobulinemia. N Engl J Med 315:564, 1986

12. Buckley RH, Schiff RI: The use of intravenous immune globulin in immunodeficiency diseases. N Engl J Med 325:110, 1991

13. Quartier P, Debre M, De Blic J, et al: Early and prolonged intravenous immunoglobulin replacement therapy in childhood agammaglobulinemia: a retrospective survey of 31 patients. J Pediatr 134:589, 1999

14. Conley ME, Cooper MD: Immature IgA B cells in IgA-deficient patients. N Engl J Med 305:495, 1981

15. Burks AW, Sampson HA, Buckley RH: Anaphylactic reactions after gamma globulin administration in patients with hypogammaglobulinemia: detection of IgE antibodies to IgA. N Engl J Med 314:560, 1986

16. Fuleihan R, Ramesh N, Loh R, et al: Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM. Proc Natl Acad Sci USA 90:2170, 1993

17. Korthauer U, Graf D, Mages HW, et al: Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. Nature 361:539, 1993

18. Mayer L, Kwan S-P, Thompson C, et al: Evidence for a defect in "switch" T cells in patients with immunodeficiency and hyperimmunoglobulin M. N Engl J Med 314:409, 1986

19. Revy P, Muto T, Levy Y, et al: Activation-induced deaminase (AID) deficiency causes the autosomal recessive form of the hyper-IgM syndrome (HIGM2). Cell 102:565, 2000

20. Noguchi M, Yi H, Rosenblatt HM, et al: Interleukin 2 receptor γ chain mutation results in X-linked severe combined immunodeficiency in humans. Cell 73:147, 1993

21. Puck JM, Conley ME, Bailey LC: Refinements of linkage of human severe combined immunodeficiency (SCDX1) to polymorphic markers in Xq13. Am J Hum Genet 53:176, 1993

22. Puel A, Ziegler SF, Buckley RH, Leonard WJ: Defective IL7R expression in T(-)B(+)NK(+) severe combined immunodeficiency. Nat Genet 20:394, 1998

23. Macchi P, Villa A, Gillani S, et al: Mutations of Jak 3 gene in patients with autosomal recessive combined immune deficiency (SCID). Nature 377:65, 1995

24. Russell SM, Tayebi N, Nakajima H, et al: Mutation of Jak3 in a patient with SCID: essential role of Jak3 in lymphoid development. Science 270:797, 1995

25. Hirschhorn R: Adenosine deaminase deficiency. Immunodefic Rev 2:175, 1990

26. Schwarz K, Gauss GH, Ludwig L, et al: RAG mutations in human B cell-negative SCID. Science 274:97, 1996

27. Markert ML: Purine nucleoside phophorylase deficiency. Immunodef Rev 3:45, 1991

28. Arpaia E, Shahar M, Dadi H, et al: Defective T cell receptor signaling and CD8+ thymocyte selection in humans lacking Zap-70 kinase. Cell 76:947, 1994

29. Chan AC, Kadlecek TA, Elder ME, et al: ZAP-70 deficiency in autosomal recessive form of severe combined immunodeficiency. Science 264:1599, 1994

30. Buckley RH, Schiff RI, Schiff SE, et al: Human severe combined immunodeficiency: genetic, phenotypic, and functional diversity in one hundred eight infants. J Pediatrics 130:378, 1997

31. Williams DA, Lemischka IR, Nathan DG, et al: Introduction of new genetic material into pluripotent hematopoietic stem cells of the mouse. Nature 310:476, 1984

32. Blaese RN, Culver KW, Miller AD, et al: T-lymphocyte-directed gene therapy for ADA deficiency SCID: initial trial results after 4 years. Science 270:470, 1995

33. Hacein-Bey-Abina S, Le Deist F, Carlier F, et al: Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N Engl J Med 346:1185, 2002

34. Derry JMJ, Ochs HD, Francke U: Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. Cell 78:635, 1994

35. Kwan S-P, Hagemann T, Radke BE, et al: Identification of mutations in the gene responsible for the Wiskott-Aldrich syndrome and characterization of a polymorphic dinucleotide repeat at the DXS 6940 locus adjacent to the disease gene. Proc Natl Acad Sci USA 92:4706, 1995

36. Remold-O'Donnell E, Rosen FS, Kenney DM: Defects in Wiskott-Aldrich syndrome blood cells. Blood 87:2621, 1996

37. Villa A, Notarangelo L, Macchi P, et al: X-linked thrombocytopenia and Wiskott-Aldrich syndrome are allelic diseases with mutations in the WASP gene. Nat Genet 9:414, 1995

38. Mullen CA, Anderson KD, Blaese RM: Splenectomy and/or bone marrow transplantation in the management of Wiskott-Aldrich syndrome: long term follow-up of 62 cases. Blood 82:2961, 1993

39. Gatti RA, Berkel I, Boder E, et al: Localization of an ataxia-telangiectasia gene to chromosome 11q22-23. Nature 336:577, 1988

40. Savitsky K, Barshira A, Gilad S, et al: A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268:1749, 1995

41. Swift M: Genetic aspects of ataxia-telangiectasia. Immunodefic Rev 2:67, 1990

IX IMMUNOLOGIC TOLERANCE AND AUTOIMMUNITY

Paul Anderson, m.d., ph.d.

A central concept of immunology is that autoimmune reactions are injurious to the host. Around 1900, Paul Ehrlich postulated that the immune system acquires a state of tolerance to self-antigens; as a corollary to that, he proposed that the breakdown of tolerance would lead to self-destruction, a condition he described as "horror autotoxicus."[1] Subsequent work by mid–20th-century researchers such as Ray Owen,[2] Macfarlane Burnet,[3] and Peter Medawar[4] established the basic mechanism for the development of immunologic tolerance. In recent years, many important advances have been made in our understanding of tolerance at the molecular and cellular levels. These advances are beginning to transform the clinical management of autoimmune diseases and may lead to therapies that prevent rejection of transplanted organs.

Tolerance

Tolerance is defined as a state of immunologic unresponsiveness to antigens, whether self or foreign. Antigens are recognized by specific receptors expressed on the surface of T cells and B cells. Binding of an antigen to the receptor can either activate or inhibit these immune effector cells. The molecular and cellular factors that determine whether receptor ligation induces immunity or tolerance are beginning to be unraveled.

Mechanisms of Tolerance

Tolerance results from one of three inhibitory influences on T and B cells: (1) clonal deletion, in which antigenic recognition leads to the activation-induced death of specific lymphocytes; (2) clonal anergy, in which lymphocytes are not killed but are rendered unresponsive to the recognized antigen; and (3) T cell–mediated suppression, in which regulatory T cells actively inhibit an immune response to an antigen. Several factors help determine which of those responses will occur.

Immature lymphocytes are more susceptible to induction of tolerance than are mature lymphocytes. Tolerance can be induced in immature lymphocytes either centrally or in the periphery. Central tolerance is acquired when immature lymphocytes encounter antigens in the organs that generate these cells: the thymus (T cells) and the bone marrow (B cells).

T cells recognize antigens that have been processed into peptides and presented in a complex with major histocompatibility complex (MHC) molecules (self-MHC–peptide complexes). Consequently, immature T cells must be screened for their ability to recognize self-MHC. This screening takes place in the thymus gland. T cells bearing receptors that recognize self-MHC are subjected to the processes of positive and negative selection [*see Figure 1*].[5] Positive selection occurs when T cells bearing receptors with a moderate affinity for self-MHC–peptide complexes receive survival and maturation signals after receptor ligation. Once these cells mature, they are exported to the periphery. Negative selection occurs when T cells bearing receptors with a high affinity for self-MHC–peptide complexes undergo activation-induced death. The thymus gland is capable of presenting many self-antigens that are normally expressed outside of the thymus or during restricted developmental stages.[6,7] This allows the elimination of most T cells bearing high-affinity receptors for self-MHC–peptide complexes and plays a major role in preventing autoimmunity in peripheral organs. The promiscuous expression of peripheral antigens in thymic epithelial cells is regulated by the autoimmune regulator (AIRE). This transcriptional modulator is mutated in persons with autoimmune polyglandular syndrome type 1 (APS-1), which is characterized by mucocutaneous candidiasis in association with autoimmune tissue damage that variably targets the parathyroid, adrenal glands, ovaries, and other tissues.[6-8] The severity of this syndrome highlights the critical importance of central tolerance to immune homeostasis.

Because positive selection allows the maturation of T cells bearing receptors capable of low-affinity interactions with self-MHC–peptide complexes, potentially self-reactive T cells are normally found in peripheral lymphoid organs. Peripheral tolerance prevents these cells from inducing autoimmune disease.

Peripheral tolerance is achieved in one of three ways.[9] Perhaps the most common mechanism is the failure of T cells bearing low-affinity receptors to recognize self-antigen in the periphery. In this situation, the potentially self-reactive T cell is not activated and remains functionally naive. These cells are functional, however, as is shown by the fact that they can be activated by immunization with self-antigen delivered in the presence of immune adjuvants (e.g., complete Freund adjuvant, which contains microbial products that strongly activate the immune system at many levels). Failure to respond to self-antigen may simply re-

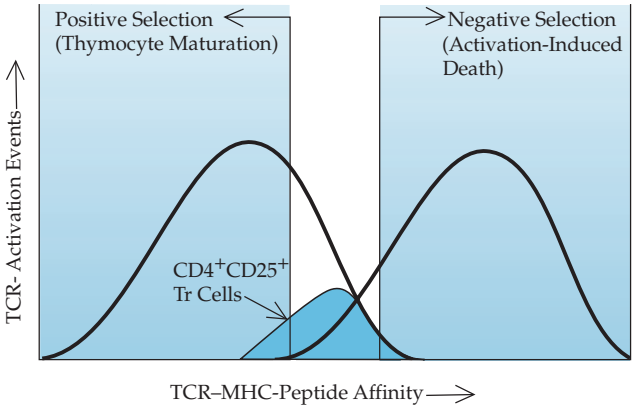


Figure 1 **In the thymus, tolerance is induced through positive and negative selection of immature T cells. The fate of a particular T cell depends on the affinity of its receptor (TCR) for complexes of major histocompatibility complex (MHC) and self-peptides. After ligation, T cells whose receptors have low affinity for self-MCH–peptide complexes receive survival and maturation signals and are exported to the periphery (positive selection); T cells with high affinity undergo activation-induced death (negative selection).**

$CD4^+$ regulatory T cells (Tr) that express CD25 have intermediate affinity for self-MCH–peptide complexes. This subpopulation of T cells matures in the thymus gland; suppression of their activation takes place in the periphery.

flect a receptor-binding affinity that is below the threshold for T cell activation.

T cells bearing receptors with high affinity for a self-antigen can also remain in an unactivated state if that self-antigen is sequestered from immune effector cells. An example of an antigen that is sequestered from the immune system is myelin basic protein. Because T cells do not normally circulate through the central nervous system, potentially self-reactive cells can persist in an unactivated state in the periphery. Similarly, pancreatic islet cells are normally sequestered from the immune system. In transgenic mice, recombinant proteins expressed on pancreatic islet cells are ignored by high-affinity T cells specific for the recombinant protein. This appears to result from the failure of naive T cells to contact islet cells in the absence of inflammation. In contrast, T cells do become activated in an antigen-specific manner in transgenic mice that express the same recombinant protein in hepatocytes. It therefore appears that circulating lymphocytes contact different tissues in different ways.

A second mechanism of peripheral tolerance involves the elimination of self-reactive T cells by apoptosis. This process is analogous to clonal deletion in the thymus (i.e., negative selection). An example of peripheral deletion is the ability of superantigens (bacterial proteins that bridge selected T cell receptors and selected MHC molecules in an antigen-nonspecific manner) to induce the activation and subsequent death of T cells[10] [*see 7:XXX Sepsis*]. Whether peripheral deletion plays an important role in tolerance to self-antigens is not known, however.

A third mechanism of peripheral tolerance involves the acquisition of anergy after ligation of the T cell receptor complex.[11] This antigen-nonresponsive state can be induced in several distinct ways. The most extensively characterized mechanism of anergy induction occurs when the T cell receptor is ligated in the absence of costimulation. In the classic studies of Schwartz and colleagues, T cell clones that were activated by MHC-peptide complexes incorporated into artificial lipid bilayers were rendered nonresponsive to subsequent challenge with peptide-pulsed antigen-presenting cells (APCs).[12] It was subsequently shown that once a T cell has bound with an antigen, the cell requires a so-called second signal delivered by one or more costimulatory molecules to be primed for an immune response. T cells express several surface molecules that can transmit this second signal. These costimulatory receptors are engaged by ligands expressed on the surface of APCs. T cells that are activated in the absence of costimulation acquire defects in the transcriptional control pathways for the production of interleukin-2 (IL-2), an important T cell autocrine growth factor.[13] In vitro anergy can often be overcome by supplying exogenous IL-2 to anergic T cells.

Costimulatory signals can be delivered to T cells by soluble factors or cell surface molecules expressed on APCs. The most potent costimulatory signals are delivered when CD28,[14] CD154,[15] or both are ligated on the surface of T cells [*see Figure 2*]. Blockade of costimulatory signals by monoclonal antibodies or recombinant receptor antagonists confers potent immunosuppression and allows the acceptance of skin, cardiac, and pancreatic allografts in rodents.[16] Simultaneous blockade of the CD28 and CD154 pathways is significantly more immunosuppressive than blockade of a single costimulatory pathway. The ligand for CD154 is CD40, a protein expressed on the surface of activated B cells, dendritic cells, and macrophages.[15] The ligands for CD28 (B7-1, B7-2, and related proteins[14]) are expressed on the surface of APCs, such as dendritic cells, monocytes, and B cells. Their expression is induced when APCs are activated in the course of microbial infection. This property heightens the immune response in the setting of perceived danger (i.e., microbial infection). B7-1 and B7-2 have overlapping immunostimulatory roles: mice lacking either protein are only partially deficient in generating an immune response to foreign antigen.[14] Additional costimulatory molecules that are involved in fine tuning the immune response include B7 homologues expressed on APCs (e.g., B7-H1, B7-H2, B7-H3, and B7-DC) that bind to ligands expressed on T cells (ICOS, PD-1, and possibly others).[14]

Ligation of CD28 induces the expression of CTLA-4 (cytotoxic T lymphocyte–associated protein 4), a structurally related protein that turns off activated T cells.[17,18] By this mechanism, the activated T cell initiates a program that will ensure its elimination at the conclusion of the immune response. Compared with CD28, CTLA-4 has a higher affinity for B7-1 and B7-2.[14] The importance of the negative regulatory influence of CTLA-4 is dramatically observed in CTLA-4–null mice. These animals develop a fatal lymphoproliferative syndrome from the uncontrolled activation of self-reactive T cells.[19]

Given the central importance of CD28-B7 interactions in T cell activation and the ability of costimulatory blockade to prevent allograft rejection, it might seem paradoxical that NOD mice (a strain that develops spontaneous diabetes) lacking either CD28 alone or both B7-1 and B7-2 have more severe diabetes.[20] The reason for this appears to be that CD28-B7 interactions are required for the maturation of self-reactive regulatory T cells (Tregs). Tregs, which are generated within the thymus gland, form a distinct class of regulatory T cells that play a major role in ensuring tolerance to self-antigens in the periphery. Just as APS-1 provides a clinical demonstration of the importance of central tolerance to normal immune function, the immune dysregulation, polyendrocrinopathy, enteropathy, and X-linked syndrome (IPEX) dramatically demonstrates the importance of Tregs in the maintenance of self-tolerance.[21] IPEX patients have mutations in the FOXP3 transcription factor that is essential for the maturation and function of Tregs. These patients exhibit hyperactivation of T cells that are reactive with self-antigens, resulting in polyendocrinopathy, inflammatory bowel disease, and allergy.

In rodents, FOXP3-dependent Tregs comprise a subset of peripheral blood $CD4^+$ T cells that express CD25, a subunit of the IL-2 receptor.[22] The selective removal of $CD4^+$ and $CD25^+$ T cells from BALB/c mice results in the development of T cell–mediated autoimmune thyroiditis, gastritis, and diabetes. The $CD4^+$ and $CD25^+$ Treg cells that mature in the thymus gland bear receptors that have an intermediate affinity for self-MHC–peptide complexes [*see Figure 1*]. In the periphery, antigen exposure confers the ability to suppress the activation of $CD4^+$ and $CD25^+$ T cells in an antigen-independent, cell contact–dependent manner. Although these cells secrete IL-10, a potent anti-inflammatory cytokine, their suppressive activity is cytokine independent. $CD4^+$ and $CD25^+$ T cells can suppress graft versus host disease in allotransplants, and they can prevent autoimmune disease in several different animal models. Consequently, these cells probably play an essential role in maintaining peripheral tolerance to self-antigens.

Although T cells play a dominant role in the maintenance of immune tolerance, non–T cells are also important in this process. Natural killer T (NKT) cells are specialized effectors of innate immunity that are activated by endogenous or microbe-derived lipids bound to CD1.[23] Activated NKT cells express

large amounts of interferon gamma and IL-4, allowing them to have profound effects on the immune response. Results in animal models have implicated NKT cells in the suppression of autoimmune disease.[24] Several autoimmune mouse strains (e.g., NOD, MRL-lpr/lpr, and SJL/J) have reduced the numbers of NKT cells as compared with nonautoimmune strains. In these models, adoptive transfer of NKT cells can ameliorate disease. Moreover, activation of NKT cells by the natural product α-galactosylceramide prevents autoimmune disease in several murine models. Although the role of NKT cells in human autoimmune disease remains to be determined, these results show that components of the innate immune response can have profound effects on discrimination of self from nonself [*see 6:II Innate Immunity*].

Autoimmunity

Despite the multiple and redundant mechanisms that exist to ensure immunologic tolerance to self, autoimmune phenomena are relatively common. In some cases, autoimmune responses accompany a normal immune response to a microbial pathogen. Thus, the appearance of rheumatoid factor (anti-immunoglobulin antibodies) in the serum of patients with bacterial endocarditis is relatively common. In general, these autoantibodies are not pathogenic. Their appearance probably results from antigen-nonspecific activation of T cells and B cells bearing low-affinity receptors for self-antigens that are normally held in check by mechanisms of peripheral tolerance. The ability of bacterial products (e.g., lipopolysaccharide) to function as immune adjuvants appears to overcome these repressive influences.

Molecular Mimicry

Clinical observations have established a link between certain microbial infections and specific autoimmune syndromes. Examples include streptococcal infection and rheumatic fever, *Borrelia burgdorferi* infection and Lyme arthritis, *Trypanosoma cruzi* infection and Chagas disease, and B4 coxsackievirus infection and type 1 diabetes mellitus.[25] These associations suggest that the immune response to a specific microbial peptide may be redirected toward a similar self-peptide, a phenomenon known as molecular mimicry. Although this is an appealing hypothesis, definitive evidence for molecular mimicry has yet to be demonstrated in any of these diseases. The strongest evidence, to date, for molecular mimicry comes from the molecular analysis of the immune response to the tick-borne spirochete *B. burgdorferi*. About 10% of infected patients develop persistent synovitis, despite the eradication of the spirochete by antibiotic therapy.[26] Most patients with treatment-resistant Lyme arthritis have the HLA-DRB1*04041 or HLA-DRB1*0101 major histocompatibility alleles, implicating antigen presentation in disease pathogenesis. These same HLA alleles confer an increased relative risk for rheumatoid arthritis and, perhaps, for synovial disease in general. Treatment-resistant Lyme arthritis is associated with a cellular and humoral immune response to the *B. burgdorferi* outer-surface protein OspA, and a computer algorithm predicted that an immunodominant OspA epitope (OspA165-173) should bind to HLA-DRB1*0401. This prediction was confirmed experimentally, and indeed, most patients with treatment-resistant Lyme arthritis have T cells that recognize this immunodominant epitope. The computer algorithm also predicted that a peptide epitope encoded by leukocyte function–associated antigen–1α (LFA-1α) would bind to HLA-DRB1*0401. This binding was confirmed experimentally, and most patients with treatment-resistant Lyme arthritis have T cells that weakly respond to the LFA-1α peptide epitope. The LFA-1α peptide does not bind to HLA-DRB1*0101, however, indicating that this mechanism cannot explain all cases of treatment-resistant Lyme arthritis. Moreover, it remains to be proved that T cells reactive with LFA-1α are necessary and sufficient for the onset of treatment-resistant Lyme arthritis.[27]

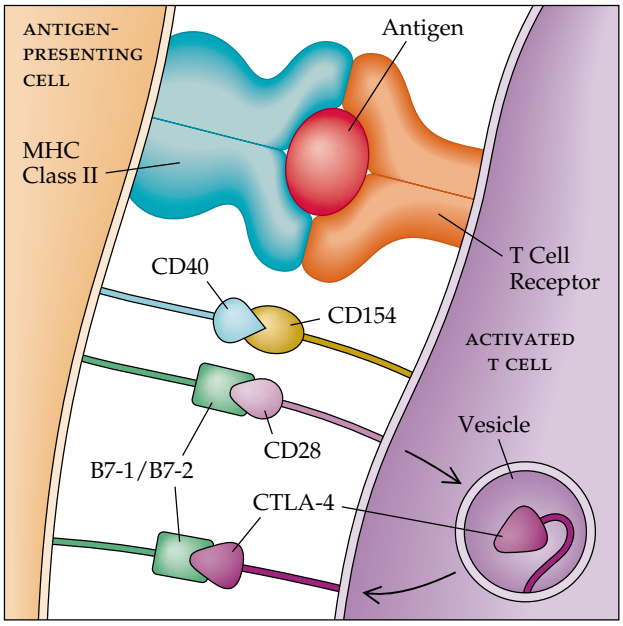


Figure 2 **Activation of T cells begins when the T cell receptor binds with a complex of an MHC molecule and a peptide expressed on the surface of an antigen-presenting cell (APC). Activation is completed by a second signal generated by the ligation of costimulatory molecules expressed on the cell surface of the APC. B7-1/B7-2 interacts with CD28 on the T cell, and CD40 interacts with CD154.**

In unactivated T cells, CTLA-4 (a relative of CD28) is a component of intracellular vesicles. After CD28 ligation, CTLA-4 moves to the cell surface and binds with B7-1/B7-2, generating negative signals that turn off the immune response.

Identification of Autoantigens

A common feature of autoimmune diseases is the appearance of autoantibodies in the serum. In some cases, these autoantibodies are directly pathogenic: the clinical syndrome is produced when the antibody binds to its target antigen. The molecular pathogenesis of these autoimmune conditions can be determined with some precision. Unfortunately, the molecular defects that allow the bypass of tolerance to the disease-inducing autoantigen are less well understood.

Pathogenic Autoantibodies

Myasthenia gravis Nearly all patients with myasthenia gravis have autoantibodies to acetylcholine receptors (ACRs) on skeletal muscle [*see 11:III Diseases of Muscle and the Neuromuscular Junction*]. However, the degree of neuromuscular blockade seen in this disease does not always parallel the serum levels of anti-ACR antibodies. The antibodies are polyclonal and bind to several distinct epitopes on the ACR. Although these antibodies can directly inhibit ACR function, they can also promote the endo-

cytosis and accelerated degradation of the ACR or activate complement-mediated destruction of the postsynaptic surface.[28] The supposition that these anti-ACR antibodies have direct pathogenic effects is supported by the fact that injection of ACR antibodies can induce myasthenic weakness in animals and that plasmapheresis is an effective treatment in some patients. Some patients with myasthenia gravis have a coincident thymoma, and thymectomy can be an effective treatment in such patients, suggesting that defects in thymic selection of maturing T cells may play a role in the autoimmune response to the ACR. This could involve impaired negative selection of $CD4^+$ helper T cells reactive with ACR-derived peptides or impaired generation of Tregs specific for ACR-derived peptide epitopes.

Pemphigus Pemphigus vulgaris and bullous pemphigoid are autoimmune skin diseases characterized by the presence of serum autoantibodies that react with adhesion molecules found at the dermoepidermal basement membrane zone [*see 2:IX Vesiculobullous Diseases*]. In pemphigus vulgaris, a common target antigen is desmoglein 3, a desmosomal adhesion molecule. In bullous pemphigoid, common target antigens are two major hemidesmosomal proteins of 180 kd and 230 kd. Several lines of evidence have implicated autoantibodies targeting these proteins in the pathogenesis of pemphigus. First, autoantibodies are consistently present in patients with pemphigus, levels of those antibodies correlate with disease activity, and the removal of the antibodies by plasmapheresis results in improvement of symptoms. Second, serum from patients with pemphigus vulgaris causes pemphigus-like lesions in mice. Third, newborns of mothers with pemphigus have transient disease resulting from transplacental transmission of maternal antibody.

Autoimmune endocrinopathies Autoantibodies reactive with hormone receptors can contribute to endocrine disorders. High levels of antibody reactive with the peripheral insulin receptor can result in insulin-dependent diabetes mellitus. Paradoxically, low levels of antibody may stimulate the insulin receptor by mimicking insulin, resulting in hypoglycemia.

Autoimmune disease of the thyroid is associated with antibodies directed toward three antigens: microsomal thyroid peroxidase, thyroglobulin, and the thyroid receptor for thyroid-stimulating hormone (TSH). Antibodies to the TSH receptor may mimic the action of TSH, thereby resulting in Graves disease [*see 3:I Thyroid*]. Another apparent autoimmune disease of the thyroid, Hashimoto thyroiditis, is associated with antibodies to the TSH receptor, but the pathogenic role of the antibodies in this disease is unclear. Less commonly, antibodies to the TSH receptor may block the action of TSH and cause hypothyroidism. A pathogenic role for the other two classes of antithyroid autoantibodies has not been established.

Antiphospholipid syndrome The antiphospholipid syndrome (APS) consists of recurrent thrombosis, fetal loss, and thrombocytopenia in association with antibodies to cardiolipin or other negatively charged phospholipids, along with abnormalities of certain clotting tests caused by an inhibitor referred to as the lupus anticoagulant [*see 5:XIV Thrombotic Disorders*]. APS can be primary or secondary; secondary APS is usually associated with systemic lupus erythematosus (SLE) or its variants. The antiphospholipid antibodies do not bind to phospholipids alone but to a complex of phospholipids and the plasma proteins β_2-glycoprotein I and prothrombin. These antibodies induce the expression of adhesion molecules on endothelial cells that promote the binding of monocytes and platelets as the first step in a thrombotic cascade.

Nonpathogenic Autoantibodies

Autoantibodies reactive with intracellular targets can serve as markers of specific autoimmune diseases. For example, antibodies reactive with citrullinated peptides are specific markers of rheumatoid arthritis, antibodies reactive with the mitochondrial enzyme 2-oxo acid dehydrogenase are specific markers of primary biliary cirrhosis, and antibodies reactive with the Smith small nuclear ribonucleoprotein (snRNP) complex are specific markers of SLE. Although these autoantibodies are unlikely to be pathogenic, their presence is highly correlated with specific autoimmune diseases. An understanding of the process that promotes the disease-specific bypass of tolerance to a selected antigen is likely to shed light on the pathogenic mechanism underlying individual autoimmune syndromes. An important insight into the mechanism by which tolerance is abrogated in an antigen-specific manner came with the realization that the targets of many autoantibodies found in the serum of patients with autoimmune disease are proteins that are modified in cells undergoing apoptotic cell death.[29] During apoptosis, myriad intracellular proteins, nucleic acids, and lipids are subjected to enzymatic and nonenzymatic modification. These modifications include protease cleavage, phosphorylation, transglutamination, ubiquitination, citrullination, and isoaspartylation.[30] It has been proposed that these modifications create neo-epitopes to which the immune system has not been tolerized.

Although proteins that are modified during apoptosis are preferred targets of the autoantibodies found in the serum of patients with autoimmune disease, it is clear that apoptosis per se is not sufficient to break tolerance to these self proteins. Apoptosis is a ubiquitous process, yet most persons do not develop autoimmune disease. Apparently, the necessary additional element is delay in the execution of the apoptotic program or the clearance of the apoptotic cell. This phenomenon has been demonstrated in mice that lack the first component of complement. C1q functions as an opsonin that binds to apoptotic cells and promotes their clearance by professional phagocytes (neutrophils and macrophages). In the absence of C1q, the clearance of apoptotic corpses is delayed. Delayed clearance of apoptotic cells somehow increases their immunogenicity.[31]

A similar phenomenon occurs when the execution of the apoptotic program is delayed. For example, influenza virus–induced apoptosis in macrophages has been shown to increase the immunogenicity of viral proteins.[32] This phenomenon requires the phagocytosis of infected macrophages by dendritic cells. By a process of cross-priming, the dendritic cell can then present antigens derived from the infected macrophage in a highly efficient manner. Because influenza virus encodes several genes that function to inhibit apoptosis (e.g., *NS1*), virus-induced apoptosis requires many hours to complete. During this delay, the virus replicates within the infected cell, and the virus-infected cell expresses stress-response proteins (heat shock proteins [HSPs]), including HSP70 and HSP90. These HSPs function as natural adjuvants that can deliver peptides to class I MHC molecules expressed by APCs.[33] The generation of modified peptides and the induction of HSPs may account for the increased immunogenicity of apoptotic cells and the generation of autoantibodies reactive with proteins that are modified during apoptotic cell death.

In this model, the autoantibodies that serve as markers of specific autoimmune diseases are generated when the target cell undergoes delayed or aberrant apoptosis. This implies that the primary insult to the target tissue is produced by a stimulus that induces aberrant cell death and modification of the specific autoantigen. Such a process may be initiated by specific environmental factors (e.g., viruses, toxins, or ultraviolet radiation).

The autoantibodies that directly and indirectly contribute to the pathogenesis of autoimmune disease are produced by differentiated B cells. A reduced activation threshold for B cells has been implicated in the pathogenesis of SLE.[34] Moreover, increased activation of B cells may contribute to the predisposition to lymphoma observed in some autoimmune diseases.[34] Finally, preliminary results suggesting therapeutic efficacy of B cell–depleting monoclonal antibodies in patients with rheumatoid arthritis and SLE indicate that B cells play an important role in the pathogenesis of autoimmune disease.[35,36]

GENETICS OF AUTOIMMUNITY

Systemic autoimmunity is a multigenic trait that is significantly influenced by environmental factors. For example, the concordance rate for SLE in monozygotic twins is only 30%, indicating that both genetic and environmental factors contribute to disease onset. The specific genes that promote autoimmunity can be identified in two ways. Most of the genes currently known to promote autoimmunity have been discovered using case-control association methodologies.[37] These studies have linked the expression of specific HLA haplotypes to specific autoimmune diseases. In a similar fashion, case-control studies have linked defects in both classical pathway complement components (C1q, C2, and C4) and Fc receptor alleles to the development of SLE. In families in which two or more members have SLE, genetic linkage analysis has been applied in an attempt to identify disease-susceptibility loci. These studies have identified several chromosomal loci with significant linkage to human SLE.[37] It is likely that future studies will identify a cohort of genes that, alone or in combination, contribute to the autoimmune diathesis.

Studies of transgenic mice that either lack or overexpress specific genes have identified three groups of genes that encode distinct classes of proteins that modify susceptibility to autoimmune disease. Absence of these genes results in autoimmunity. The first group of genes encode proteins involved in the initiation or execution of apoptotic cell death. These proteins include dedicated death receptors and their ligands. Specific members of this family (e.g., Fas and tumor necrosis factor type I [TNF RI]) are required for the clonal elimination of activated T cells after an immune response to microbial infection. Mice lacking either Fas or its ligand develop lymphadenopathy and splenomegaly from the accumulation of previously activated T cells. In some strains of transgenic mice (e.g., MRL), but not in others (e.g., BALB/c), failure to eliminate activated T cells results in an autoimmune disease that resembles SLE. Thus, the absence of Fas or Fas ligand (FasL) promotes the phenotypic expression of an autoimmune diathesis that is intrinsic to the MRL strain (a genetic phenomenon known as epistasis). Although defects in Fas or FasL are not linked to autoimmunity in patients with SLE, mutations in either Fas or FasL produce the autoimmune lymphoproliferative syndrome (ALPS), an autosomal dominant condition characterized by lymphadenopathy, splenomegaly, and autoantibody production.[38] ALPS is also caused by mutations in caspase-10, a component of the effector arm of the apoptotic death program. Thus, ALPS is an autoimmune disease that results from defective execution of an apoptotic program in activated T cells. The importance of the apoptotic program in determining susceptibility to autoimmune disease is further demonstrated by the autoimmune syndromes observed in mice lacking BIM, a pro-apoptotic protein, or overexpressing BCL-2, an anti-apoptotic protein.

The second group of genes encode proteins involved in the recognition, clearance, or elimination of apoptotic cells. These include proteins that serve as opsonins to promote the phagocytosis of apoptotic cells (e.g., C1q, IgM, SAP/CRP), as well as phagocyte receptors that promote the recognition and ingestion of apoptotic cells (e.g., Mer).

The third group of genes encode proteins that set the threshold for lymphocyte activation. Increased activation of T cells or B cells is likely to disrupt normal mechanisms of tolerance, resulting in autoimmunity. Examples include costimulatory lymphocyte surface molecules (e.g., CD22, PD-1), kinases and phosphatases involved in lymphocyte activation (e.g., Lyn, Cbl-b), and transcription factors (e.g., Foxo3a) that promote lymphocyte activation. These animal studies reveal that genes involved in the regulation of apoptosis or lymphocyte activation are proven modifiers of autoimmune disease in mice and are candidates for the modulation of autoimmunity in humans.

Genome-wide linkage mapping has identified mutations in NOD2 as an etiologic factor in familial Crohn disease, an autoimmune inflammatory process that targets the intestinal mucosa.[39] NOD2 is a cytosolic protein that recognizes muramyl dipeptide (MDP), a metabolite of bacterial peptidoglycan.[40] Recognition of MDP promotes the oligomerization of NOD2, which results in activation of NF-κB and caspase-1—signaling events that lead to the secretion of inflammatory cytokines. Selected patients with Crohn disease possess mutant NOD2 that is unable to promote MDP-mediated activation of NF-κB. The mechanism by which ineffective recognition of bacterial products leads to intestinal inflammation remains to be determined.

ORGAN-SPECIFIC VERSUS SYSTEMIC AUTOIMMUNITY

For many years, organ-specific immunity was thought to result from the activation of lymphocytes bearing receptors specific for a tissue-restricted antigen. In rare cases of molecular mimicry, this may be the case. However, it now appears that the target of an autoimmune attack can shift from one tissue to another in response to defined or undefined genetic modifiers. For example, persons with APS-1 develop various combinations of autoimmune thyroiditis, parathyroid disease, and type 1 diabetes mellitus.[41] Although the factors that determine which tissues become targets of autoimmune attack have not been identified, the fact that different tissues are affected in different persons suggests that unique, tissue-specific autoantigens may not be the primary triggers of disease.

The concept that organ-specific autoimmunity need not be driven by a tissue-specific autoantigen is supported by observations made in two different animal models of autoimmunity. In NOD mice whose MHC locus is replaced with that of another strain, autoimmune thyroiditis develops instead of diabetes.[20] This result suggests that the NOD strain harbors an autoimmune diathesis that can manifest itself as different types of organ-specific autoimmunity. In support of this concept, NOD mice lacking the costimulatory molecule B7-2 develop autoimmune peripheral neuropathy, rather than diabetes.[20] Although the mechanism by which individual tissues are selected for im-

mune attack is not known, these results strongly suggest that factors other than tissue-restricted autoantigens can be the primary determinant of organ-specific autoimmune disease.

Another instructive example of organ-specific autoimmunity that arises in the absence of a defined, tissue-restricted autoantigenic trigger is the inflammatory arthritis that develops in the F1 progeny of K/B×NOD mice.[42] The K/B strain expresses a transgenic T cell receptor that recognizes a self-peptide derived from glucose-6-phosphate isomerase (GPI) presented in the context of Ag7, a class II MHC molecule from the NOD strain. In K/B×NOD mice, T cells bearing the transgene provide help for B cells encoding immunoglobulins that bind to GPI. GPI is an enzyme expressed in all cells, yet anti-GPI antibodies somehow provoke a symmetrical, inflammatory arthritis involving diarthrodial joints in these mice. Although the mechanism by which anti-GPI antibodies provoke arthritis is not fully understood, this model illustrates the potential for an immune response that is directed at a ubiquitous antigen to trigger organ-specific autoimmunity.

One way in which organ-specific autoimmunity can be induced in the absence of a tissue-specific autoantigen is by the pathologic overexpression of inflammatory cytokines. Thus, overexpression of tumor necrosis factor–α (TNF-α) in transgenic mice is sufficient to induce a symmetrical polyarthritis that resembles rheumatoid arthritis.[43] This appears to result from the ability of TNF-α to initiate an inflammatory cytokine cascade within the cells that make up the synovium. The importance of TNF-α in the pathogenesis of rheumatoid arthritis has been dramatically validated by the clinical efficacy of TNF blockers such as infliximab and etanercept[44] [*see 15:II Rheumatoid Arthritis*]. In an analogous fashion, BAFF/Blys, a TNF-α–related protein that promotes the survival and differentiation of B cells, has been proposed to participate in the induction of SLE-like autoimmune syndromes.[45,46] Transgenic mice engineered to overexpress BAFF/Blys develop hypergammaglobulinemia and autoimmune symptoms because of the survival of autoreactive B cells that would normally be deleted from the B cell repertoire. These observations suggest that neutralization of TNF family members may play an important role in the treatment of selected autoimmune diseases.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Ehrlich P, Morgenroth J: On haemolysis: third communication. Berlin Klin Wochenschr 37:453, 1900

2. Owen R: Immunogenetic consequences of vascular anastomosis between bovine twins. Science 102:400, 1945

3. Burnet F: Clonal selection theory: a modification of Jerne's theory of antibody production using the concept of clonal selection. Aust J Sci 20:67, 1957

4. Billingham R, Brent L, Medawar P: Actively acquired tolerance of foreign cells. Nature 172:603, 1953

5. Nossal GJ: Negative selection of lymphocytes. Cell 76:229, 1994

6. Kyewski B, Derbinski J: Self-representation in the thymus: an extended view. Nat Rev Immunol 4:688, 2004

7. Mathis D, Benoist C: Back to central tolerance. Immunity 20:509, 2004

8. Peterson P, Pitkanen J, Sillanpaa N, et al: Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED): a model disease to study molecular aspects of endocrine autoimmunity. Clin Exp Immunol 135:348, 2004

9. Fazekas de St. Groth B: DCs and peripheral T cell tolerance. Semin Immunol 13:311, 2001

10. Sundberg E, Li Y, Mariuzza R: So many ways of getting in the way: diversity in the molecular architecture of superantigen-dependent T-cell signaling complexes. Curr Opin Immunol 14:36, 2002

11. Macian F, Im S, Garcia-Cozar F, et al: T-cell anergy. Curr Opin Immunol 16:209, 2004

12. Quill H, Schwartz R: Stimulation of normal inducer T cell clones with antigen presented by purified Ia molecules in planar lipid membranes. J Immunol 138:3704, 1987

13. Nelson BH: IL-2, regulatory T cells, and tolerance. J Immunol 172:3983, 2004

14. Sharpe AH, Freeman GJ: The B7-CD28 superfamily. Nat Rev Immunol 2:116, 2002

15. Quezada SA, Jarvinen LZ, Lind EF, et al: CD40/CD154 interactions at the interface of tolerance and immunity. Annu Rev Immunol 22:307, 2004

16. Sun Y, Subudhi SK, Fu YX: Co-stimulation agonists as a new immunotherapy for autoimmune diseases. Trends Mol Med 9:483, 2003

17. Chikuma S, Bluestone JA: CTLA-4 and tolerance: the biochemical point of view. Immunol Res 28:241, 2003

18. Greenwald RJ, Latchman YE, Sharpe AH: Negative co-receptors on lymphocytes. Curr Opin Immunol 14:391, 2002

19. Tivol E, Borriello F, Schweitzer A, et al: Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. Immunity 3:541, 1995

20. Lesage S, Goodnow C: Organ-specific autoimmune disease: a deficiency of tolerogenic stimulation. J Exp Med 194:F31, 2001

21. Torgerson T, Ochs H: Immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome and immune dysregulation. Curr Opin Immunol 2:481, 2002

22. Sakaguchi S: Naturally arising $CD4^+$ regulatory T cells for immunologic self-tolerance and negative control of immune responses. Annu Rev Immunol 22:531, 2004

23. Brigl M, Brenner MB: CD1: antigen presentation and T cell function. Annu Rev Immunol 22:817, 2004

24. Van Kaer L: α-Galactosylceramide therapy for autoimmune diseases: prospects and obstacles. Nat Rev Immunol 5:31, 2004

25. Rose N, Mackay I: Molecular mimicry: a critical look at exemplary instances in human diseases. Cell Mol Life Sci 57:542, 2000

26. Steere AC, Glickstein L: Elucidation of Lyme arthritis. Nat Rev Immunol 4:143, 2004

27. Benoist C, Mathis D: Autoimmunity provoked by infection: How good is the case for T cell epitope mimicry? Nature Immunol 2:797, 2001

28. Hughes BW, Moro De Casillas ML, Kaminski HJ: Pathophysiology of myasthenia gravis. Semin Neurol 24:21, 2004

29. Hall JC, Casciola-Rosen L, Rosen A: Altered structure of autoantigens during apoptosis. Rheum Dis Clin North Am 30:455, 2004

30. Utz P, Gensler T, Anderson P: Death, autoantigen modifications, and tolerance. Arthritis Res 2:101, 2000

31. Manderson A, Botto M, Walport M: The role of complement in the development of systemic lupus erythematosus. Annu Rev Immunol 22:431, 2004

32. Albert M, Sauter B, Bhardwaj N: Dendritic cells acquire antigen from apoptotic cells and induce class I-restricted CTLs. Nature 392:86, 1998

33. Albert ML: Death-defying immunity: do apoptotic cells influence antigen processing and presentation? Nat Rev Immunol 4:223, 2004

34. Criscione LG, Pisetsky DS: B lymphocytes and systemic lupus erythematosus. Curr Rheumatol Rep 5: 264, 2003

35. Looney RJ, Anolik JH, Campbell D, et al: B cell depletion as a novel treatment for systemic lupus erythematosus: a phase I/II dose-escalation trial of rituzimab. Arthritis Rheum 50:2580, 2004

36. Kotzin BL: The role of B cells in the pathogenesis of rheumatoid arthritis. J Rheumatol Suppl 73:14, 2005

37. Raman K, Mohan C: Genetic underpinnings of autoimmunity—lessons from studies in arthritis, diabetes, lupus and multiple sclerosis. Curr Opin Immunol 15:651, 2003

38. Rieux-Laucat F, Fischer A, Deist FL: Cell-death signaling and human disease. Curr Opin Immunol 15:325, 2003

39. Beutler B: Autoimmunity and apoptosis: the Crohn's connection. Immunity 15:5, 2001

40. Inohara N, Nunez G: NODs: intracellular proteins involved in inflammation and apoptosis. Nat Rev Immunol 3:371, 2003

41. Ruan Q, She J: Autoimmune polyglandular syndrome type 1 and the autoimmune regulator. Clin Lab Med 24:305, 2004

42. Monach PA, Benoist C, Mathis D: The role of antibodies in mouse models of rheumatoid arthritis, and relevance to human disease. Adv Immunol 82:217, 2004

43. Sfikakis P, Kollias G: Tumor necrosis factor biology in experimental and clinical arthritis. Curr Opin Rheumatol 15:380, 2003

44. Vilcek J, Feldmann M: Historical review: cytokines as therapeutics and targets of therapeutics. Trends Pharmacol Sci 25:201, 2004

45. Ramanujam M, Davidson A: The current status of targeting BAFF/BLyS for autoimmune diseases. Arthritis Res Ther 6:197, 2004

46. Cancro M: The BLyS family of ligands and receptors: an archetype for niche-specific regulation. Immunol Rev 202:237, 2004

Acknowledgment

Figure 2 Seward Hung.

X ALLERGIC RESPONSE

Pamela J. Daffern, m.d.
Lawrence B. Schwartz, m.d., ph.d.

Definition of Allergic Response

The word anaphylaxis was coined in 1902 by Charles Richet, in order to contrast the condition with prophylaxis. Richet described anaphylaxis as "the peculiar attribute which certain poisons possess of increasing instead of diminishing the sensitivity of an organism to their action.…"[1] One hundred years later, we understand anaphylaxis as the extreme of a spectrum of events mediated by immunoglobulin E (IgE). Persons with IgE-mediated disorders have a genetic propensity to form IgE antibodies against otherwise innocuous environmental antigens (allergens); this propensity is termed atopy (from the Greek *atopos*, meaning "out of place"). In atopic persons, IgE mediates a wide range of reactions, including dermatitis, rhinitis, asthma, urticaria, angioedema, and anaphylaxis.

Confusion arises over the misapplication of the term allergy to describe any untoward reaction to food or medications or to perceived environmental exposures. This confusion is further complicated by the fact that both IgE-mediated and non–IgE-mediated forms of rhinitis, asthma, and atopic dermatitis occur, often in the same person.

In the nonatopic person, exposure to allergen results in immunologic tolerance or neglect, whereas in atopic persons, exposure results in sensitization. On reexposure to the allergen, atopic persons mount an immunologically mediated inflammatory response in the target organ. Other environmental factors—such as tobacco smoke, air pollution, respiratory virus infection, and lack of exposure to certain microbes in childhood—may also promote an allergic inflammatory response.

Epidemiology of Atopic Disorders

Up to 30% of the United States population may be affected by allergic rhinoconjunctivitis, asthma, or atopic dermatitis. This high incidence of atopic disease may reflect societal factors. Fetal development takes place in an intrauterine environment that favors atopic sensitization[2]; the maternal immune system suppresses cell-mediated immune responses in order to prevent rejection of the fetus. Thus, the neonate may enter the world with T cells that are already primed by common environmental and food allergens that have crossed the placenta. It has been proposed that microbial exposure and infections during infancy shift the immune response away from the allergic pattern to a protective immune response.[3] Specifically, after macrophages or dendritic cells ingest microbes, T cells produce cytokines that promote non-IgE responses by B cells. Therefore, the increasing prevalence of atopic disorders in countries that have adopted a Western lifestyle, including overuse of antibiotics, has been attributed to a lack of microbial antigen stimulation.

Humoral and Cellular Mechanisms of Allergic Inflammation Associated with Immediate Hypersensitivity

antigen-presenting cells and sensitization

All persons encounter environmental antigens that are capable of inducing an allergic response. Soluble antigens, such as allergens, undergo endocytosis by professional antigen-presenting cells (APCs), which include dendritic cells, such as epidermal Langerhans cells; macrophages; and B cells.[4] However, only dendritic cells and Langerhans cells are able to prime naive T cells and thus are responsible for the sensitization phase.[5,6] Once primary sensitization has been achieved, monocytes and B cells amplify the process. B cells bind allergen through immunoglobulin receptors specific for the allergen, as opposed to nonspecific endocytic pathways used by other APCs. The internalization of antigen results in two processes. The first is general activation of the APC: this includes upregulation of major histocompatibility complex (MHC) and accessory molecules. The second process is fusion of the endocytic vesicle with lysosomes, which results in the formation of specialized antigen-processing vesicles in which antigens are hydrolyzed into protein fragments. The linear peptides that result are incorporated into the antigen-binding groove of a class II human lymphocyte antigen (HLA) molecule during its transport to the cell surface.

In general, APCs will co-express a heterogeneous assortment of allergen-derived peptides and HLA class II molecules on their surface. The efficiency with which processed allergen peptides bind to the HLA class II molecules presumably depends on variations in the HLA loci; these variations are genetically determined. The binding efficiency in turn influences the predisposition of the person to develop allergy to or tolerance of a particular antigen. The APC loaded with processed antigen/HLA class II complexes presents this complex to $CD4^+$,$CD8^-$ helper T cells. The genetically determined binding efficiency of an HLA-derived molecule to an antigen also may influence how T cells develop when exposed to that complex.[7] In addition, the quantity of interleukin-12 (IL-12) produced by APCs also influences the type of T cell response.[5]

t cells and mediation of allergic inflammation

The helper T cell response is influenced not only by APCs but also by the age of the person and by the amount, type, duration, and route of allergen exposure.[7,8] Also, the cytokine milieu during lymphocyte differentiation determines the type of effector function of the helper T cell [*see Table 1*].

For example, bacterial DNA sequences have immunostimulatory regions containing deoxycytidine-phosphate-deoxyguanosine (CpG) repeats. CpG repeats are recognized as foreign by pattern recognition receptors called Toll-like receptor-9 (TLR-9) on APCs.[9,10] These CpG repeats stimulate macrophages and dendritic cells to secrete inflammatory cytokines, including IL-12 and IL-18. These cytokines then induce T cells and natural killer (NK) cells to produce interferon gamma (IFN-γ), a cytokine known to promote nonallergic, protective responses.

Table 1 Cytokines Involved in IgE-Mediated Allergic Inflammation

Cytokine	*Source*	*Function*
IL-3	T_{H2} cells,* mast cells, basophils, eosinophils	Promotes granulocyte and macrophage maturation; eosinophil activation and survival
IL-4	T_{H2} cells,* mast cells, basophils	Promotes differentiation of T_{H0} to T_{H2} cells; antagonizes differentiation of T_{H0} to T_{H1} cells; IgE isotype switching
IL-5	T_{H2} cells,* mast cells, eosinophils	Promotes eosinophil development, activation, and survival
IL-13	T_{H2} cells,* mast cells, basophils	IgE isotype switching, eosinophil activation
GM-CSF	T_{H2} cells and activated macrophages,* endothelial and epithelial cells	Promotes granulocyte and macrophage maturation, eosinophil activation, and survival
TNF-α	Monocytes/macrophages,* mast cells	Promotes chemotaxis and activation of leukocytes and vascular endothelium

*Major source.
GM-CSF—granulocyte-macrophage colony-stimulating factor IL—interleukin TNF—tumor necrosis factor

This pattern of response by helper T cells is termed a T_{H1} response, because it is associated with differentiation of naive helper T (T_{H0}) cells into mature T_{H1} cells. Similarly, the helper T cells of persons without atopy respond to presentation of potentially allergenic peptides by ignoring them or by producing IFN-γ and directing the production of allergen-specific IgG1 and IgG4 antibodies.[11]

In contrast, helper T cells of atopic persons respond to processed aeroallergens by forming IL-4, IL-5, and IL-13 and by directing the production of allergen-specific IgE antibodies. This type of helper T cell response is termed a T_{H2} response. IL-4 and IL-13 share a number of functions, because both cytokines signal through the IL-4Rα/IL-13Rα heterodimer.[12] However, only IL-4 is able to induce the differentiation of T_{H0} cells to T_{H2} cells and to antagonize the differentiation of T_{H0} cells to T_{H1} cells, resulting in IgE-mediated allergic inflammation. In contrast, both IL-12 and IFN-γ induce the differentiation to T_{H1} cells; T_{H2} cell differentiation is inhibited by IFN-γ. Differentiation to T_{H1} cells results in cell-mediated immunity and inflammation.[13] Therefore, the differentiation of T_{H0} cells to either T_{H1} cells or T_{H2} cells appears to be the crucial event that determines which type of immune response will follow.

GENETICS AND THE DEVELOPMENT OF ATOPY

Research has begun to identify specific genetic variants that contribute to the development of the atopic state. For example, a mutation of the IL-12R beta$_2$-chain gene has recently been shown to impair signaling through IL-12. Because IL-12 is a potent inducer of IFN-γ production and because IFN-γ downregulates IgE production (see above), this mutation results in increased IgE production in atopic persons.[14] Polymorphisms in the gene for STAT-6, a transcription factor selectively regulated by IL-4 and IL-13 (cytokines that upregulate IgE production), have also been described.[15] These genetic variations in STAT-6 also appear to be associated with a predisposition to atopy. Finally, an asthma gene (*ADAM-33*) associated with bronchial hyperresponsiveness but not atopy was recently defined by genetic-linkage analysis of affected sibling pairs. The *ADAM-33* gene product, a membrane metalloprotease, may function to modulate the response to cytokines in the lung by solubilizing cytokine membrane receptors, but its precise role still needs to be determined.[16] Like other allergies, however, asthma involves environmental factors. For example, the predisposition to asthma is modified by the presence of allergens and endotoxins.[17]

IgE SYNTHESIS

Once allergen is processed by APCs and presented to T_{H2} cells, a specific sequence of events must follow for IgE production by B cells to occur [*see Figure 1*]. The switch from IgM or IgG production to IgE production by B cells occurs in the genome and requires two signals.[18] The first signal is delivered through the IL-4Rα chain by either IL-4 or IL-13.[12] Signaling through these cytokine receptors initiates transcription from the germline promoter site of the constant portion of the heavy chain of IgE. The IgE heavy-chain gene is located downstream of the IgG and IgM heavy-chain genes and replaces IgG or IgM on the immunoglobulin molecule. The second signal is delivered through activation of the cluster differentiation 40 (CD40) receptor on B cells.[19] Signaling through CD40 activates the recombinases necessary to remove the upstream IgG or IgM heavy-chain constant region and replace it with the corresponding region of IgE. This process switches the type of antibody being produced without altering its antigenic specificity. Stimulation of B cells through CD40 also stimulates growth, differentiation, and survival of these cells.[20]

The ligand for CD40 (CD154, CD40L) is expressed not only on T cells but also on mast cells and basophils. Importantly, all of these cells also secrete IL-4, IL-13, or both, and therefore could potentially play a role in directing B cell production of IgE. However, it seems likely that T cells are responsible for initiating the switch to antigen-specific IgE production. Mast cells and basophils may then amplify deviation of immune responses toward IgE production after the primary IgE sensitization has occurred.[19] It seems likely that binding to CD40 on B cells by mast cells and basophils would enhance polyclonal (i.e., not antigen-dependent or specific) IgE production by B cells, because mast cells and basophils are not dedicated APCs. IgE antibody secreted by B cells circulates briefly, having a serum half-life of 2 to 3 days, before binding to IgE receptors.

IgE RECEPTORS AND REGULATION OF IgE

Receptors for IgE (FcεR) are expressed on various cells.[21] The high-affinity receptor for IgE, FcεRI, has two forms that differ by the presence or absence of a beta chain. The beta chain is present in the high-affinity receptor found on mast cells and basophils.[19] The presence of the beta chain amplifies the cellular signaling that occurs when IgE bound to FcεRI is cross-linked by allergen. Its presence also increases the amount of IgE receptor on the surfaces of mast cells and basophils by up to sixfold.[22]

Levels of FcεRI on the surface of basophils have been shown to correlate with serum IgE levels in various IgE-associated diseases.[23,24] The high-affinity receptor lacking the beta chain is also expressed on monocytes, Langerhans cells, dendritic cells (i.e., APCs other than B cells), activated eosinophils, and epithelial cells.

The low-affinity IgE receptor, FcεRII, bears structural homology to C-type lectins, but not to FcεRI. (Lectin receptors recognize pathogens and also function as adhesion receptors and signaling molecules.) FcεRII, also known as CD23, is expressed on B and T cells, monocytes, eosinophils, and platelets.[25] CD23 expression is increased by IL-4 and IL-13, and increased CD23 expression would presumably facilitate allergen uptake and presentation to T cells by APCs.[26] Furthermore, B cells from allergic asthmatic patients exposed to allergen have increased CD23 expression.[27] Whether allergic inflammation is initiated when IgE is bound to FcεRII is not clear. However, the solubilized form of CD23 may play a regulatory role in IgE synthesis.[26,28]

When a sensitized individual is exposed to allergen, the allergen binds to IgE receptors on mast cells and basophils. If multivalent, the allergen will cross-link a critical number of cell-bound IgE receptors, leading to cellular activation, secretion of mediators, and production of the symptoms characteristic of early-phase allergic responses.[28]

Clearly, treatment that interferes with IgE activation of mast cells and basophils may be beneficial. Omalizumab, a recombinant, humanized monoclonal antibody directed against the Fcε portion of IgE, has recently been developed.[29] Important features of this anti-IgE molecule are (1) it does not bind IgE already attached to FcεRI, and therefore does not cause anaphylaxis; (2) it does not activate complement; and (3) it has a much longer half-life than IgE. In phase III trials, omalizumab was administered by subcutaneous injections given every 2 or 4 weeks to patients with allergic rhinitis or with allergic asthma of varying severity.[30,31] All studies showed dramatic reductions in free IgE levels that were dependent on omalizumab dose as well as baseline IgE levels.[32] As levels of serum IgE decreased, so did surface expression of FcεRI on basophils. Moreover, the posttreatment level of free IgE directly correlated with reduced symptom scores, reduced use of rescue medication, and improved quality of life. For asthma, significant reductions in asthma exacerbations, in hospitalizations for asthma, and in the dose of inhaled or oral steroids were also found. A recent phase II trial of omalizumab in peanut-sensitive children showed a decreased sensitivity to oral peanut challenges.

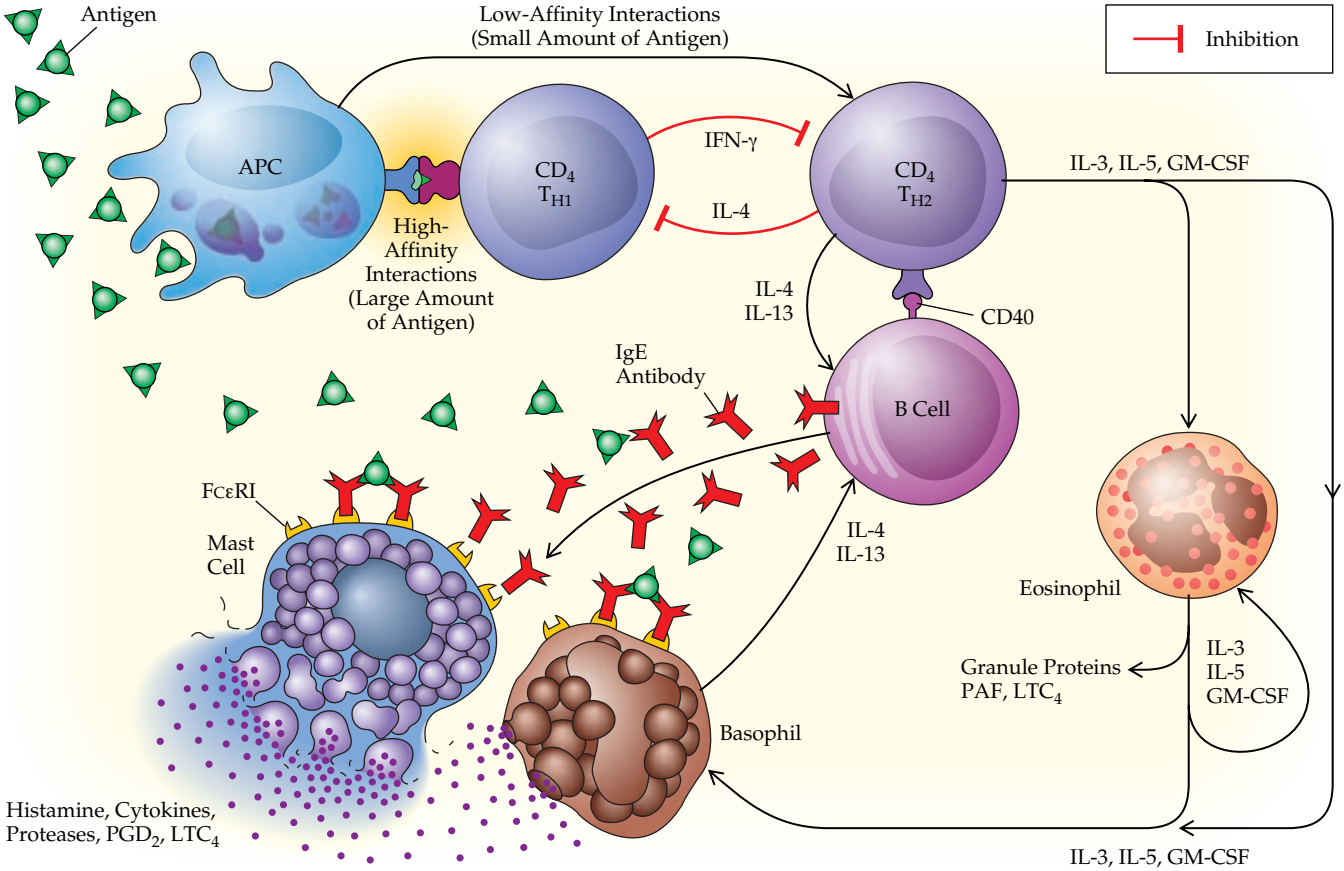


Figure 1 **Inflammatory mechanisms in allergic inflammation. Antigen is taken up by antigen-presenting cells (APCs), processed, and then presented to $CD4^+$ helper T cells (T_H). The strength of interactions between APCs and helper T cells and the quantity of antigen present determine the type of T cell response. Production of interferon gamma (IFN-γ) during T_{H1} responses downregulates T_{H2} responses, whereas interleukin-4 (IL-4) production by T_{H2} inhibits T_{H1} responses. IL-4 is also critical for switching B cell antibody production to IgE. Signaling through cluster differentiation 40 (CD40) on the B cell is also required for IgE production. Other T_{H2} cytokines, such as IL-3, IL-5, and granulocyte-macrophage colony-stimulating factor (GM-CSF) lead to eosinophil (Eos) and basophil (Baso) production and activation. IgE binds to high-affinity receptors on basophils and mast cells (FcεRI); cross-linking by allergen initiates mediator release. (CpG—deoxycytidine-phosphate-deoxyguanosine; LTC_4—leukotriene C_4; PAF—platelet-activating factor; PGD_2—prostaglandin D_2)**

Eosinophils share a common origin with basophils: a single bone-marrow–derived myeloid progenitor cell has the capacity to give rise to a mixed colony of eosinophils and basophils or to pure colonies of either cell type.[33] A common origin for eosinophils and basophils is further supported by the presence of Charcot-Leyden crystal (CLC) protein and major basic protein (MBP) in both cell types. Eosinophil development is uniquely dependent on the presence of IL-5, a cytokine whose chief source is the T_{H2} helper cell.[34] Along with IL-5, other T cell cytokines—IL-3 and granulocyte-macrophage colony-stimulating factor (GM-CSF)—promote maturation, activation, and prolonged survival of eosinophils.[33] However, only IL-5 potently stimulates the bone marrow to produce eosinophils. In vitro, a low dose of IL-3 favors the development of basophils from progenitors, whereas a high dose of IL-3 favors the development of eosinophils. In contrast, other cytokines may inhibit the growth of eosinophil progenitors. Transforming growth factor–β (TGF-β) contributes to eosinophil apoptosis in vitro and influences progenitor development toward the basophil pathway.[35] IFN-α inhibits progenitor cells in vitro and has been used for treatment of certain patients with eosinophilia refractory to treatment with prednisone.[36]

Eosinophils dwell primarily in tissue. Circulating eosinophils have a short half-life and represent only about 1% of the total number of eosinophils in the body. Epithelial surfaces of mucosal tissues that are exposed to the external environment are heavily inhabited by eosinophils, whereas other tissues are normally devoid of eosinophils.[33] The epithelial tissues of the respiratory tract produce GM-CSF, which is capable of prolonging eosinophil survival in vitro for up to 14 days.

Cell Surface Receptors

Two overlapping populations of circulating eosinophils are thought to represent differing states of eosinophil activation.[37] Nonallergic individuals have greater numbers of eosinophils of normal density and fewer numbers of low-density activated eosinophils. The reverse is true for patients with disorders leading to eosinophilia. This heterogeneity suggests that priming of eosinophils by various cytokines may lead to changes in expression of surface receptors and mediator release [*see Table 2*]. For example, both high-affinity receptors (FcεRI) and low-affinity receptors (FcεRII) for IgE have been found on peripheral blood eosinophils from patients with hypereosinophilic syndrome. However, eosinophils derived from normal donors or from patients with allergy fail to stain with a panel of monoclonal antibodies directed against IgE receptors.[38] Similar differences have been observed for IgG receptors (FcγR) on eosinophils. Freshly isolated eosinophils express FcγRIIb, a low-affinity IgG receptor that may inhibit mediator release when cross-linked.[34] Both FcγRI and FcγRIII can be induced on eosinophils in vitro when these cells are cultured with IFN-γ, which, in contrast to FcγRIIb, may result in activation. Sera from patients with hay fever contain allergen-specific IgG1 and IgG3, which cause eosinophils to degranulate in vitro in an allergen-dependent manner.[39] Surface receptors for IgA are also present on eosinophils and provide a potent stimulus for release of granule proteins in vitro. The presence of secretory IgA (sIgA) together with eosinophils at mucosal surfaces suggests that IgA-dependent activation also occurs in vivo.[40]

Receptors for complement (C3a and C5a); the lipid mediators platelet-activating factor (PAF), leukotriene C_4 (LTC_4), and LTB_4; and numerous cytokines and chemokines bind to and activate eosinophils.[33,34] Chemokines of the C-C family play an important chemotactic role for eosinophils. Chemokines of this large family have adjacent cysteine residues (C-C) and have the same receptors. A particular C-C chemokine receptor, CCR3, is found abundantly on eosinophils but not on neutrophils.[41] CCR3 binds at least four chemokines that play crucial roles in the homing of eosinophils to epithelial tissues and that activate eosinophils to release mediators. Another mechanism, which leads to preferential accumulation of eosinophils rather than neutrophils at sites of allergic inflammation, relates to differences in expression of surface adhesion molecules. Eosinophils and neutrophils share several selectins and integrins that initiate the rolling of circulating cells along the endothelium, as well as the subsequent firm adhesion, diapedesis, and transmigration of these cells through the vessel wall. However, eosinophils—but not neutrophils—express an integrin, very late antigen (VLA)-4, whose ligand on endothelial cells (VCAM-1) is upregulated by IL-4 and IL-13, cytokines that are present during T_{H2} responses; consequently, these cytokines promote adherence of eosinophils, but not neutrophils, to endothelium.[42]

Table 2 Receptors on Eosinophils

IgE receptors	Lipid-mediator receptors
FcεRI (high affinity)	Leukotriene (LT) receptors
FcεRII (low affinity)	LTC_4
IgA receptor	LTB_4
Complement receptors	Platelet-activating factor (PAF)
C3a	Chemokine receptors
C5a	CCR3
	Others

Mediators

An array of inflammatory mediators are produced when eosinophils are activated. Preformed mediators are stored in granules and rapidly released once eosinophils are activated. Major basic protein (MBP) is the principal constituent of the granule proteins.[43] Other granule proteins include eosinophil peroxidase (EPO), eosinophil-derived neurotoxin (EDN), and eosinophil cationic protein (ECP). MBP, ECP, and EPO have been shown to damage parasites in vitro; in patients with eosinophil-associated diseases, these proteins are present in high concentrations that can cause toxicity to autologous cells and tissues. Unfortunately, MBP and EPO cause ciliostasis and detachment of respiratory epithelial cells in vitro, and they may contribute to epithelial damage and inflammation in allergic respiratory disorders.[43] However, in one study, treatment of asthmatic patients with anti–IL-5 monoclonal antibody resulted in the selective elimination of eosinophils from the airway, but airway hyperreactivity or the airway response to inhaled allergen were not affected. This leaves open the question of the precise role that eosinophils play in the pathogenesis of atopic asthma.[44] Proteases present in the eosinophils may contribute to airway damage by degrading collagen.[45]

Lipid mediators are rapidly generated by eosinophils after appropriate stimulation. PAF production may lead to activation of platelets, neutrophils, and smooth muscle cells, and thereby induce bronchoconstriction and amplify inflammation. The major eicosanoid product of eosinophils is LTC_4, from which LTD_4 and LTE_4 are derived. These sulfidopeptides are extremely potent at contracting airway smooth muscle, stimulating mucus produc-

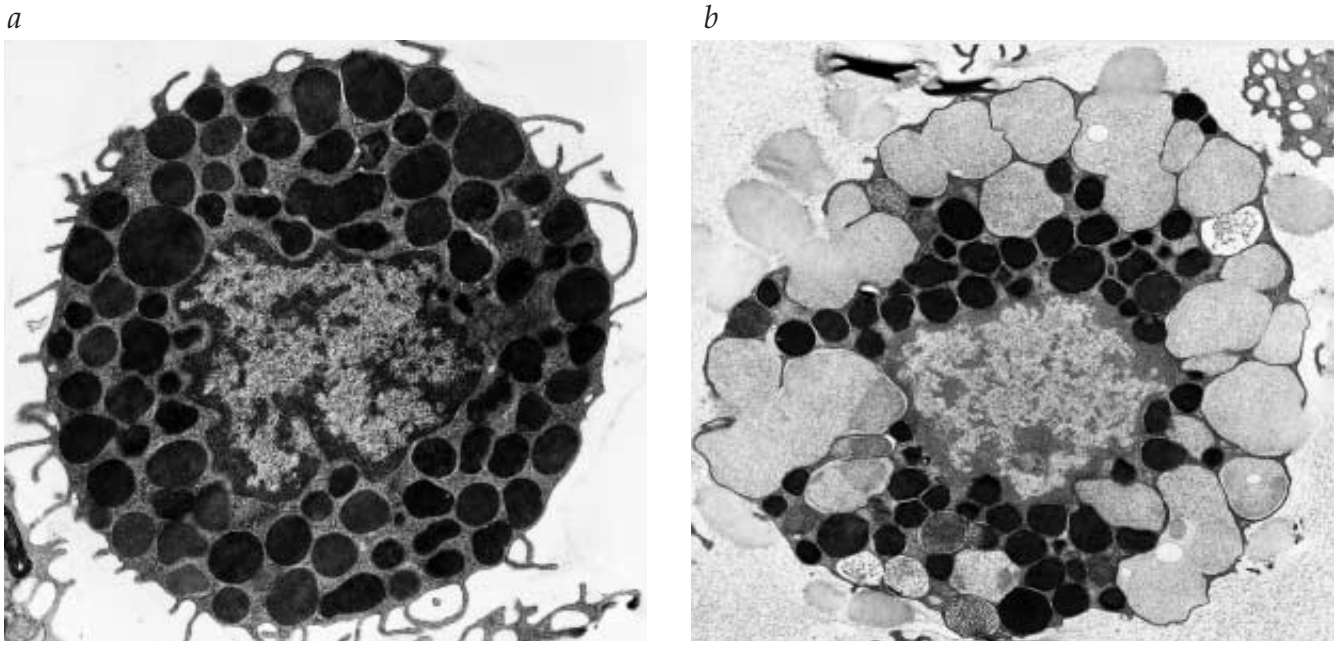


Figure 2 **(*a*) Before introduction of antigen, a sensitized mast cell contains many osmotic granules. (*b*) Sixty seconds after treatment with antigen, the peripheral granules have enlarged, neighboring granules have fused, and expulsion of granules from the mast cell has begun.**

tion, causing capillary leakage, and promoting chemotaxis of eosinophils.

Numerous cytokines have been identified as potential eosinophil products. Some may function in an autocrine or paracrine manner to activate or prime eosinophils. Others enhance eosinophil development and survival. In addition, eosinophils produce cytokines that regulate immune responses. However, eosinophils elaborate a considerably smaller quantity of cytokines than do lymphocytes. Therefore, the importance of the eosinophil-derived cytokines to allergic inflammation is unclear. Some cytokines that have been demonstrated in vitro have been confirmed in vivo by identifying the protein product in eosinophils infiltrating affected tissues. For example, eosinophils from nasal polyp tissue stain for TGF-β1 and could contribute to the structural pathology.[46] Exposure of allergic patients to allergen revealed eosinophils in nasal mucosal tissues that stain for IL-5 protein; however, much larger contributions of IL-5 are anticipated from T cells in the same tissue.[47]

MAST CELLS AND BASOPHILS AS EFFECTORS OF THE ALLERGIC RESPONSE

Microscopy of mast cells and basophils reveals intensely staining metachromatic granules [*see Figure 2*]. Other common features shared by these cells include the presence of high-affinity receptors for IgE, the release of histamine after cross-linking of the FcεRI by allergen, and common intracellular signaling pathways.[48] There are also numerous differences between the two cell types. Basophils generally complete their maturation in the bone marrow, circulate in the blood, and then are recruited to sites of inflammation.[49] Mast cells that complete their maturation in the bone marrow appear to remain there, whereas those found in peripheral tissues develop from progenitor cells that seed these tissues. Mature mast cells in peripheral tissues may reside there for many months, retaining antigen-specific IgE for periods that exceed the lifespan of IgE in the circulation. Mast cells are strategically distributed in tissues or at mucosal surfaces that interface with the external environment; they are also in proximity to blood vessels and nerves.[50]

All mast cells contain tryptase in their granules; its release is characteristic of mast cell degranulation. However, several additional features further distinguish two types of mast cells [*see Figure 3*].[51] Mast cells of the T type (MC_T cells) are normally the predominant type of mast cell found in the mucosa of the small intestine and in the alveolar wall and epithelium of the respiratory tract. MC_T cells are identified morphologically by a scroll-rich granule structure; they contain tryptase but not chymase, cathepsin G, or mast cell carboxypeptidase. The numbers of MC_T cells in respiratory epithelium are increased in allergic airway inflammation, making them more accessible to inhaled allergens. In a study in asthmatics, increased mast cells predominantly of the MC_{TC} type were localized to the airway smooth muscles but were not present in control subjects or in patients with eosinophilic bronchitis.[52]

In contrast, the MC_{TC} type of mast cell is the dominant type of mast cell in the dermis, conjunctiva, blood vessel walls, and small-intestinal submucosa. Morphologically, MC_{TC} cells display a lattice/grating, scroll-poor granule structure. In addition to tryptase, TC-type mast cells contain chymase, cathepsin G, and mast cell carboxypeptidase. The development of both mast cell types requires stem cell factor (SCF), the ligand for the Kit (tyrosine kinase) receptor. Factors that regulate the recruitment, development, or survival of one mast cell type over the other are not known. Lineage-committing growth factors such as GM-CSF may divert hematopoietic progenitor cells that are capable of forming mast cells when exposed to SCF alone to non–mast cell lineages.[53]

Mediators

Mast cells and basophils form histamine by decarboxylation of histidine. They then store the histamine in their granules. Degranulation releases the histamine, which then interacts with histamine receptors on various tissues. Histamine induces smooth-

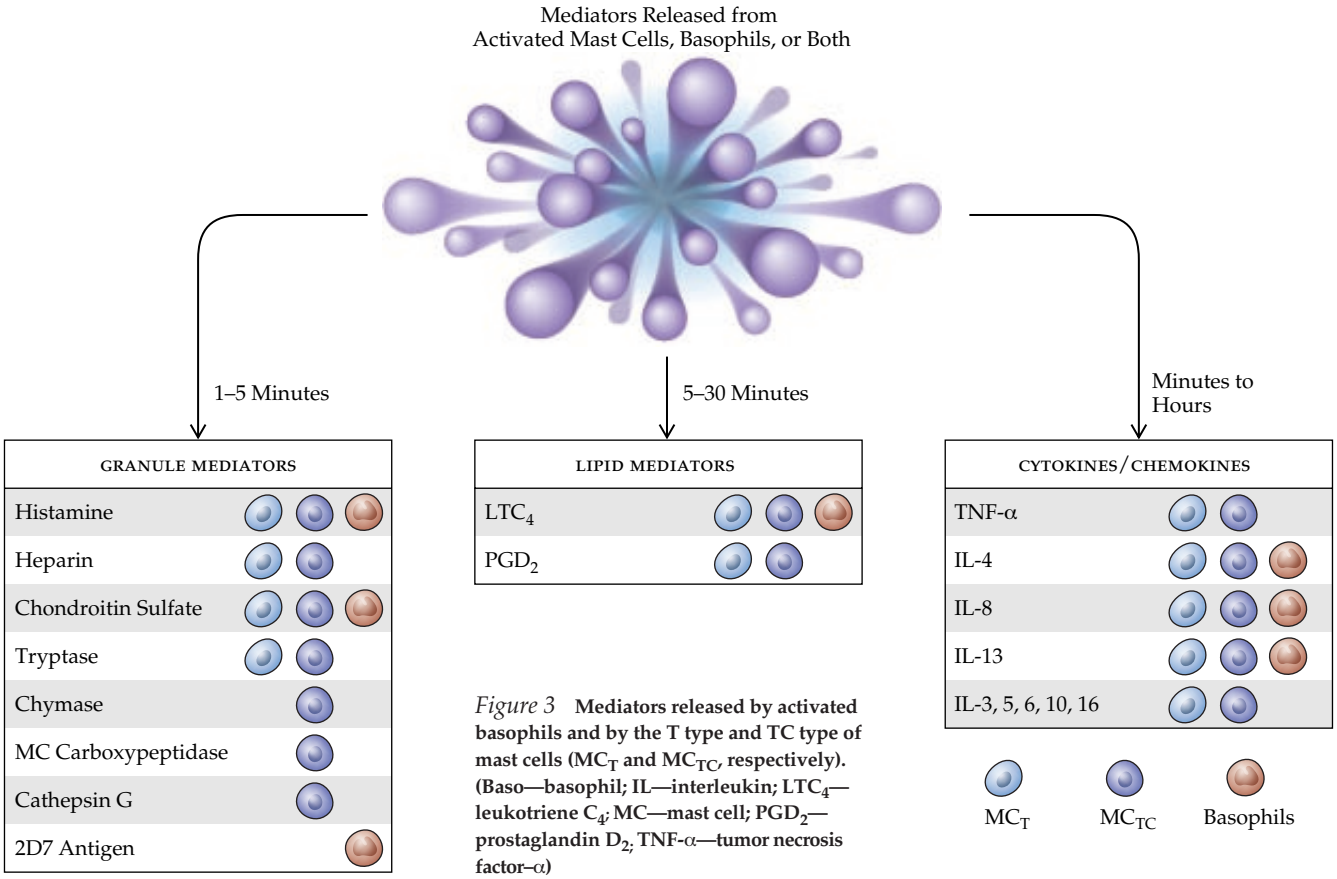


Figure 3 **Mediators released by activated basophils and by the T type and TC type of mast cells (MC_T and MC_{TC}, respectively). (Baso—basophil; IL—interleukin; LTC_4—leukotriene C_4; MC—mast cell; PGD_2—prostaglandin D_2; TNF-α—tumor necrosis factor–α)**

muscle contraction, increases mucous secretion in the airway, and stimulates nerve fibers. In addition, it enhances vascular permeability and dilates blood vessels, which results in hypotension if a critical number of cells degranulate. Chondroitin sulfates are proteoglycans that are present in the granules of both basophils and mast cells; heparin proteoglycan is stored exclusively in all mast-cell secretory granules. Both chondroitin sulfate and heparin proteoglycans play a role in packaging of histamine, proteases, and carboxypeptidases in the granules.[54] Heparin is also involved in the processing of chymase and tryptase to catalytically active enzymes. Neutralization of the acidic granule pH during degranulation facilitates the dissociation of histamine from the protease-proteoglycan macromolecular complex.[55] Consequently, histamine appears in the serum within minutes of induction of systemic anaphylaxis by allergen-dependent cross-linking of IgE on mast cells and basophils. Not surprisingly, peak plasma levels of histamine occur 5 minutes after insect-sting–induced anaphylaxis begins and decline to baseline within 20 minutes. Because they are relatively transient, these elevations in histamine levels in plasma are difficult to utilize for the clinical determination of anaphylaxis as a cause of hypotension. However, tryptase diffuses into, and is removed from, the circulation more slowly than histamine. Tryptase levels peak in the circulation 15 minutes to 2 hours after mast-cell degranulation and decline with a half-life of about 2 hours. Peak levels during insect-sting–induced anaphylaxis correlate closely to the drop in mean arterial blood pressure, which is an important measure of clinical severity. For that reason, serum or plasma tryptase levels have recently been recognized as a clinically useful marker for the diagnosis of systemic anaphylaxis.[55]

Prostaglandin D_2 (PGD_2) is a newly synthesized cyclooxygenase product of arachidonic acid produced by MC_T and MC_{TC} cells, but not by basophils. PGD_2 causes airway smooth muscle to contract, blood vessels to dilate, and platelets to remain unaggregated. In one study of patients with systemic mastocytosis and recurrent episodes of cardiovascular collapse that did not respond to antihistamines, therapeutic success was achieved with cyclooxgenase inhibition that diminished PGD_2 production.[56] LTC_4 is produced by both mast cells and basophils, as well as by eosinophils, and it is a potent mediator of airway smooth muscle contraction and mucus secretion. Effects of LTC_4 are blocked by 5-lipoxygenase inhibition and by leukotriene receptor antagonists.

Mast cells secrete a diverse array of cytokines, including TNF-α, GM-CSF, SCF, and interleukins 3, 4, 5, 6, 10, 13, and 16.[50] TNF-α can reside preformed in mast cell granules and is also synthesized and secreted after mast cell activation. TNF-α causes chemotaxis and activation of many leukocytes, as well as activation of vascular endothelium. IL-4 and IL-13 are central to T_{H2} differentiation, IgE isotype switching, and induction of the adhesion receptors VCAM-1 on endothelium and VLA-4 on eosinophils. Basophils do not synthesize TNF-α and generally produce fewer cytokines than do mast cells. However, activated basophils synthesize more IL-4 and IL-13 on a per-cell basis than any other cell type. In tissues with allergic inflammation that are challenged with allergen, basophils appear to be the predominant source of antigen-specific production of IL-4 and IL-13.[49]

As with eosinophils, a subpopulation of low-density (so-called hypodense) basophils can be detected in peripheral blood samples. This subpopulation is more sensitive to the effects of

glucocorticoids than the higher-density basophils. However, functional differences in the hypodense basophils have not been characterized, as they have for eosinophils. Recently, basophil-specific markers have been developed.[57] The monoclonal antibodies named 2D7 and BB1 detect basophil-specific antigens in secretory granules and should prove useful for more precise assessment of basophil involvement in human allergic diseases. For example, substantial numbers of basophils can now be detected in skin and respiratory tissues during the late-phase response to an allergen challenge, and these cells account for a major portion of the IL-4–containing cells in such tissues. Basophils appear to be similar to eosinophils in expression of numerous cytokines and chemokine receptors, including CCR3. Exposure of basophils to most CC chemokines leads to histamine release.

Pamela J. Daffern, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

Lawrence B. Schwartz, M.D., Ph.D., has a licensing agreement with the Pharmacia Corporation for a tryptase assay kit.

References

1. Richet C: Anaphylaxis. The University Press, Liverpool, 1913

2. Prescott SL, Macaubas C, Holt BJ, et al: Transplacental priming of the human immune system to environmental allergens: universal skewing of initial T cell responses toward the Th2 cytokine profile. J Immunol 160:4730, 1998

3. Bjorksten B: The intrauterine and postnatal environments. J Allergy Clin Immunol 104:1119, 1999

4. Klein J, Sato A: The HLA system: first of two parts. N Engl J Med 343:702, 2000

5. Langenkamp A, Messi M, Lanzavecchia A, et al: Kinetics of dendritic cell activation: impact on priming of Th1, Th2 and nonpolarized T cells. Nat Immunol 1:311, 2000

6. Palucka K, Banchereau J: Dendritic cells: a link between innate and adaptive immunity. J Clin Immunol 19:12, 1999

7. Klein J, Sato A: The HLA system: second of two parts. N Engl J Med 343:782, 2000

8. Rogers PR, Croft M: Peptide dose, affinity, and time of differentiation can contribute to the Th1/Th2 cytokine balance. J Immunol 163:1205, 1999

9. Wild JS, Sur S: CpG oligonucleotide modulation of allergic inflammation. Allergy 56:365, 2001

10. Tighe H, Corr M, Roman M, et al: Gene vaccination: plasmid DNA is more than just a blueprint. Immunol Today 19:89, 1998

11. Kay AB: Allergy and allergic diseases: first of two parts. N Engl J Med 344:30, 2001

12. Wills-Karp M, Luyimbazi J, Xu X, et al: Interleukin-13: central mediator of allergic asthma. Science 282:2258, 1998

13. Till S, Durham S, Dickason R, et al: IL-13 production by allergen-stimulated T cells is increased in allergic disease and associated with IL-5 but not IFN-gamma expression. Immunology 91:53, 1997

14. Kondo N, Matsui E, Kaneko H, et al: Reduced interferon-gamma production and mutations of the interleukin-12 receptor beta (2) chain gene in atopic subjects. Int Arch Allergy Immunol 124:117, 2001

15. Tamura K, Arakawa H, Suzuki M, et al: Novel dinucleotide repeat polymorphism in the first exon of the STAT-6 gene is associated with allergic diseases. Clin Exp Allergy 31:1509, 2001

16. Van Eerdewegh P, Little RD, Dupuis J, et al: Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426, 2002

17. Gehring U, Bischof W, Fahlbusch B, et al: House dust endotoxin and allergic sensitization in children. Am J Respir Crit Care Med 166:939, 2002

18. Busse WW, Lemanske RF Jr: Asthma. N Engl J Med 344:350, 2001

19. Bacharier LB, Geha RS: Molecular mechanisms of IgE regulation. J Allergy Clin Immunol 105:S547, 2000

20. Doyle IS, Hollmann CA, Crispe IN, et al: Specific blockade by CD54 and MHC II of CD40-mediated signaling for B cell proliferation and survival. Exp Cell Res 265:312, 2001

21. Kinet JP: The high-affinity IgE receptor (Fc epsilon RI): from physiology to pathology. Annu Rev Immunol 17:931, 1999

22. Donnadieu E, Jouvin MH, Kinet JP: A second amplifier function for the allergy-associated Fc(epsilon)RI-beta subunit. Immunity 12:515, 2000

23. Saini SS, MacGlashan DWJ, Sterbinsky SA, et al: Down-regulation of human basophil IgE and Fc epsilon RI alpha surface densities and mediator release by anti-IgE infusions is reversible in vitro and in vivo. J Immunol 162:5624, 1999

24. Saini SS, Richardson JJ, Wofsy C, et al: Expression and modulation of Fc epsilon RI-alpha and Fc epsilon RIbeta in human blood basophils. J Allergy Clin Immunol 107:832, 2001

25. Squire CM, Studer EJ, Lees A, et al: Antigen presentation is enhanced by targeting antigen to the Fc epsilon RII by antigen-anti-Fc epsilon RII conjugates. J Immunol 152:4388, 1994

26. Kisselgof AB, Oettgen HC: The expression of murine B cell CD23, in vivo, is regulated by its ligand, IgE. Int Immunol 10:1377, 1998

27. Bonnefoy JY, Lecoanet-Henchoz S, Aubry JP, et al: CD23 and B-cell activation. Curr Opin Immunol 7:355, 1995

28. Pearlman DS: Pathophysiology of the inflammatory response. J Allergy Clin Immunol 104:S132, 1999

29. Presta LG, Lahr SJ, Shields RL, et al: Humanization of an antibody directed against IgE. J Immunol 151:2623, 1993

30. Milgrom H, Fick RB Jr, Su JQ, et al: Treatment of allergic asthma with monoclonal anti-IgE antibody. rhuMAb-E25 Study Group. N Engl J Med 341:1966, 1999

31. Adelroth E, Rak S, Haahtela T, et al: Recombinant humanized mAB-E25, an anti-IgE mAb, in birch pollen-induced seasonal allergic rhinitis. J Allergy Clin Immunol 106:253, 2000

32. Johansson SG, Haahtela T, O'Byrne PM: Omalizumab and the immune system: an overview of preclinical and clinical data. Ann Allergy Asthma Immunol 89:132, 2002

33. Gleich GJ: Mechanisms of eosinophil-associated inflammation. J Allergy Clin Immunol 105:651, 2000

34. Rothenberg ME: Eosinophilia. N Engl J Med 338:1592, 1998

35. Atsuta J, Fujisawa T, Iguchi K, et al: Inhibitory effect of transforming growth factor beta 1 on cytokine-enhanced eosinophil survival and degranulation. Int Arch Allergy Immunol 108:31, 1995

36. Gratzl S, Palca A, Schmitz M, et al: Treatment with IFN-alpha in corticosteroid-unresponsive asthma. J Allergy Clin Immunol 105:1035, 2000

37. Fukuda T, Dunnette SL, Reed CE, et al: Increased numbers of hypodense eosinophils in the blood of patients with bronchial asthma. Am Rev Respir Dis 132:981, 1985

38. Smith SJ, Ying S, Meng Q, et al: Blood eosinophils from atopic donors express messenger RNA for the alpha, beta, and gamma subunits of the high-affinity IgE receptor (Fc epsilon RI) and intracellular, but not cell surface, alpha subunit protein. J Allergy Clin Immunol 105:309, 2000

39. Kaneko M, Swanson MC, Gleich GJ, et al: Allergen-specific IgG1 and IgG3 through Fc gamma RII induce eosinophil degranulation. J Clin Invest 95:2813, 1995

40. Abu-Ghazaleh RI, Fugisawa T, Mestecky J, et al: IgA-induced eosinophil degranulation. J Immunol 142:2393, 1989

41. Heath H, Qin S, Rao P, et al: Chemokine receptor usage by human eosinophils: the importance of CCR3 demonstrated using an antagonistic monoclonal antibody. J Clin Invest 99:178, 1997

42. Schleimer RP, Sterbinsky SA, Kaiser J, et al: IL-4 induces adherence of human eosinophils and basophils but not neutrophils to endothelium: association with expression of VCAM-1. J Immunol 148:1086, 1992

43. Gleich GJ, Adolphson MS, Leiferman KM: Annual Reviews of Medicine: Selected Topics in Clinical Sciences. Creger WP, Coggins CH, Hancock EW, Eds. Annual Reviews, Inc., Palo Alto, California, 1993, p 85

44. Leckie MJ, ten Brinke A, Khan J, et al: Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. Lancet 356:2144, 2000

45. Mallya SK, Hall JE, Lee HM, et al: Interaction of matrix metalloproteinases with serine protease inhibitors: new potential roles for matrix metalloproteinase inhibitors. Ann NY Acad Sci 732:303, 1994

46. Bachert C, Gevaert P, Holtappels G, et al: Nasal polyposis: from cytokines to growth. Am J Rhinol 14:279, 2000

47. Lee CH, Lee KS, Rhee CS, et al: Distribution of RANTES and interleukin-5 in allergic nasal mucosa and nasal polyps. Ann Otol Rhinol Laryngol 108:594, 1999

48. Holgate ST: The role of mast cells and basophils in inflammation. Clin Exp Allergy 30:28, 2000

49. Bochner BS: Systemic activation of basophils and eosinophils: markers and consequences. J Allergy Clin Immunol 106:S292, 2000

50. Williams CM, Galli SJ: The diverse potential effector and immunoregulatory roles of mast cells in allergic disease. J Allergy Clin Immunol 105:847, 2000

51. Gurish MF, Austen KF: The diverse roles of mast cells. J Exp Med 194:71, 2001

52. Brightling CE, Bradding P, Symon FA, et al: Mast-cell infiltration of airway smooth muscle in asthma. N Engl J Med 346:1699, 2002

53. Mekori YA, Metcalfe DD: Mast cell–T cell interactions. J Allergy Clin Immunol 104:517, 1999

54. Galli SJ: Mast cells and basophils. Curr Opin Hematol 7:32, 2000

55. Schwartz LB, Irani AM: Serum tryptase and the laboratory diagnosis of systemic mastocytosis. Hematol Oncol Clin North Am 14:641, 2000

56. Roberts LJ 2nd, Sweetman BJ, Lewis RA, et al: Increased production of prostaglandin D2 in patients with systemic mastocytosis. N Engl J Med 303:1400, 1980

57. Kepley CL, Craig SS, Schwartz LB: Identification and partial characterization of a unique marker for human basophils. J Immunol 154:6548, 1995

Acknowledgment

Figures 1 and 3 Seward Hung.

XI DIAGNOSTIC AND THERAPEUTIC PRINCIPLES IN ALLERGY

Mitchell H. Grayson, m.d.
Phillip E. Korenblat, m.d., f.a.c.p.

By definition, allergy is an untoward physiologic event mediated by immune mechanisms, usually involving the interaction of an allergen with the allergic antibody, IgE. Common illnesses mediated in this manner include allergic asthma and rhinitis, Hymenoptera hypersensitivity, and certain other causes of anaphylaxis. In addition, a significant proportion of drug, food, and skin reactions are allergic in origin.

Allergic diseases in general, and asthma in particular, have been increasing in prevalence in high-income societies.[1] Although there are undoubtedly many reasons for this increase, one is described in the so-called hygiene hypothesis, which posits that greater exposure to infectious agents (and bacterial endotoxins in particular) early in life reduces the likelihood of subsequent allergy.[2] This hypothesis, which requires further proof of causality, acknowledges an etiologic role for both genetic and environmental factors in allergy: a child with a hereditary predisposition to atopy is more likely to develop clinical allergy if raised in a relatively aseptic environment.

History

In allergic illnesses, the importance of a careful and thorough medical history cannot be overstated. The clinician must dissect the allergic reaction to understand the nature of the event and identify the antigen that was responsible for the reaction. Formal diagnosis of allergy has three elements: characterization of the allergic reaction, correlation with antigen exposure, and demonstration of IgE specific for the suspected allergen. The history is essential for the first two elements, and for practical purposes, the history can sometimes obviate the third element.

The history should begin with a review of the patient's symptoms and their temporal pattern. If the presenting symptoms include wheezing, the clinician should remember the time-honored statement that all that wheezes is not asthma. Furthermore, all that is asthma is not allergy [*see 14:II Asthma*].

The presenting symptoms must match the set of features that characterize the suspected allergic illness. For example, patients with perennial allergic rhinitis typically present with sneezing, rhinorrhea, nasal itching, and nasal congestion. Postnasal drainage is not the only symptom of this disease, so postnasal drainage alone—even with evidence of antigen exposure and the presence of specific IgE antibodies—typically would not support the diagnosis of allergic rhinitis.

A central aspect of the history is to establish a link between the time and site of exposure to the presumed allergen and the development of allergic symptoms. Seasonal allergic events are often so characteristic that the diagnosis can be made solely on the basis of the presenting symptoms and their correlation with environmental exposure to the allergen; laboratory evidence may not be needed for the diagnosis. Similarly, symptoms that develop immediately after exposure to animals or their dander often do not need additional supporting evidence for diagnosis.

In the United States, the presence of airborne pollen may vary both temporally and geographically.[3] In general, early spring is characterized by the presence of tree pollen, and late spring is accompanied by grass pollen. Ragweed and other weed pollens are prevalent in the fall, usually until the first hard frost. Mold spores can be found indoors year-round, except possibly in very dry areas. Outdoor mold spores peak during the summer and fall months, and they diminish when snow covers the landscape [*see Table 1*].

Illnesses such as asthma or rhinitis that occur on a perennial basis, if allergic, should correlate with environmental exposure to a perennial allergen (e.g., dust mites, indoor mold spores, animal dander, or cockroach antigen). Such exposure most often takes place in the household, but the possibility of exposure to allergens in the workplace should not be forgotten. It should be noted that in some warm climates (e.g., that of the southern United States), the pollen season may be nearly year-round and, thus, may be a cause of perennial symptoms.

Allergic reactions to ingested substances typically include skin eruptions, abdominal discomfort, or respiratory symptoms. Severe and life-endangering reactions involving the cardiovascular or respiratory system, or both, may also occur. The list of ingested substances said to cause allergic reactions is seemingly endless. However, foods (particularly peanuts, tree nuts, shellfish, and seeded fruits) and medications are the most common triggers of this type of allergic reaction [*see 6:XVI Food Allergies*]. Again, the history is essential to establishing a particular substance as the probable cause of an allergic reaction.

The family history is important. Allergic predisposition is genetically mediated, so patients with allergies often report that family members have similar problems. However, in a patient who has both a personal and a family history of angioedema, the disorder may be inherited but not allergic: hereditary angioedema results from the absence of the C1 esterase inhibitor.

Physical Examination

The physical examination of a patient with a suspected allergic illness requires an in-depth focus on the involved organ system or systems. In atopic dermatitis, the skin findings may include patches of lesions that are pruritic, erythematous, papular,

Table 1 Inhaled Aeroallergens That Cause Rhinitis, Conjunctivitis, and Asthma

- Pollens (tree, grass, and weed pollens)
- Dust mites (*Dermatophagoides* species)
- Animal proteins (cat, dog, horse, guinea pig, gerbil, mouse, and rat proteins)
- Fungal spores (*Alternaria*, *Aspergillus*, *Penicillium*, and *Cladosporium* species)
- High-molecular-weight proteins (e.g., as derived from insects, insect venoms, and latex)
- Low-molecular-weight inorganic and organic chemicals (e.g., toluene diisocyanate and plicatic acid)

scaling, crusting, vesicular, or lichenified—qualities that may occur alone or in combination. Lesions are usually characterized by periodic exacerbations, and it is important to examine these lesions for pyogenic infections.

The distribution of allergic dermatitis lesions varies with the age of the patient. In infants, the dermatitis begins to appear by the sixth to eighth week of life. At this age, the eruptions ordinarily involve the scalp, face (especially the cheeks), ears, and extensor surfaces of the extremities. The trunk, buttocks, and anogenital regions may also be involved. The dermatitis may continue into childhood. Alternatively, allergic dermatitis may first develop at about 2 years of age. Dermatitis in childhood is often found in the antecubital and popliteal fossa, on the neck, and at the flexor and extensor areas of the wrist. In adolescents and adults, the lesions frequently involve the neck and the flexural areas but may occur anywhere on the skin.[4]

Typical urticarial lesions are pruritic, transient (individual lesions resolve within 24 hours), erythematous, and raised; they comprise a wheal with a surrounding erythematous flare. Urticaria can be confused with skin lesions of vasculitis. The presence of hemorrhage or a lesion that lasts longer than 24 hours should raise the specter of urticarial vasculitis. Skin biopsy may be required to differentiate urticarialike lesions.

The hallmarks of allergic rhinoconjunctivitis are bilateral erythema and edema of the conjunctiva, watery ocular discharge, and, often, mild periorbital edema.[5] Allergic shiners (bluish discoloration just below the eye orbits) may be observed. Patients with allergic rhinitis may also have an extra fold in the lower eyelids (Dennie-Morgan lines). On the exterior portion of the nose, a crease may be present as a result of continued upward rubbing of the tip of the nose (the so-called allergic salute). Examination of the nasal cavity often reveals watery secretions and edematous, bluish nasal turbinates that partially occlude the nasal passages [*see 6:XII Allergic Rhinitis, Conjunctivitis, and Sinusitis*]. Translucent nasal polyps may be observed, but these are not necessarily a hallmark of allergy; they can be seen in both allergic and nonallergic patients.

The chest examination often may reveal no abnormalities. However, a methodical examination is warranted. The clinician should observe specifically for cyanosis and the use of accessory muscles for respiration. In addition, auscultation is indicated for a prolonged expiratory respiratory phase or for inspiratory and expiratory wheezing. If wheezing is present, it is important to confirm that the sounds emanate from the lungs and not the trachea. All too often, extrathoracic obstruction is missed on the physical examination.

Although cardiovascular findings are not commonly associated with allergic diseases, it is important to remember that hypotension, tachyarrhythmia, and—particularly if the patient is using a beta-adrenergic blocking medication—bradycardia may be seen in cases of anaphylaxis.

Table 2 Time before Skin Testing to Stop Antihistamines*

Antihistamine (Trade Name)	*Days*
Azelastine	7
Cetirizine (Zyrtec)	7
Chlorpheniramine	3
Desloratadine (Clarinex)	7
Diphenhydramine	3
Fexofenadine (Allegra)	7
Loratadine (Claritin)	7

*Note: Other medications (e.g., tricyclic antidepressants) may also have antihistaminic activity.

Assays of IgE

Because allergic diseases result from the interaction of an allergen with specific IgE, analysis for specific IgE in a patient with clinical allergy is a major diagnostic consideration. Specific IgE can be identified both by in vivo methods (skin testing) and in vitro methods (e.g., radioallergosorbent testing [RAST]).[6]

SKIN TESTING

Epicutaneous Testing

The most rapid and sensitive test for allergy is skin testing. This in vivo method depends on mast cell–bound or basophil-bound IgE specific for the allergen being tested. Because a positive test requires degranulation of mast cells or basophils and subsequent histamine release, antihistamines will interfere with the outcome. In general, patients should discontinue antihistamines 1 week before skin testing, although certain antihistamines can be discontinued 3 days beforehand [*see Table 2*]. Corticosteroids do not inhibit this immediate-phase response, and hence, their use is not a contraindication for skin testing.

Skin testing should be performed by a qualified allergist. Initial testing is performed by pricking the epidermis with a small amount of the specific allergen. In patients with IgE specific for the allergen, a wheal-and-flare response will develop at the site within 20 minutes. The areas of edema and erythema are then measured. The results are often reported as wheal size over flare size (both in millimeters) or, alternatively, identified on an arbitrary scale of 1 to 4+, correlating to the size of the wheal, the flare, or both. Histamine is used as a positive control, and because some patients develop hives in response to any strong pressure on the skin (dermatographism), saline is used as a negative control.

Intradermal Testing

If the results of epicutaneous skin testing are negative but the patient's symptoms strongly suggest an allergic etiology, intradermal testing can be performed. This involves injecting 0.02 ml of a dilute allergen solution (usually a 1:100 or 1:1,000 dilution of the concentrated extract) into the dermis. As with epicutaneous testing, the skin is observed for the development of a wheal and flare within 20 minutes. Grading of the results is similar to that for epicutaneous testing.

Intradermal testing has a higher sensitivity but a lower specificity than epicutaneous testing. This means that intradermal testing produces more false positives but fewer false negatives than epicutaneous testing. Although the relevance of isolated positive intradermal tests for aeroallergens is debated, intradermal testing is crucial for the evaluation of drug and insect allergy.

Compared with epicutaneous testing, intradermal testing exposes the body to a significant antigen load and, therefore, poses a higher risk of a systemic reaction. For that reason, intradermal testing is contraindicated in patients who have not had a prior negative result on epicutaneous testing. It is not surprising that five of the six skin-testing fatalities reported from 1945 to 1987 occurred in patients who underwent intradermal testing with-

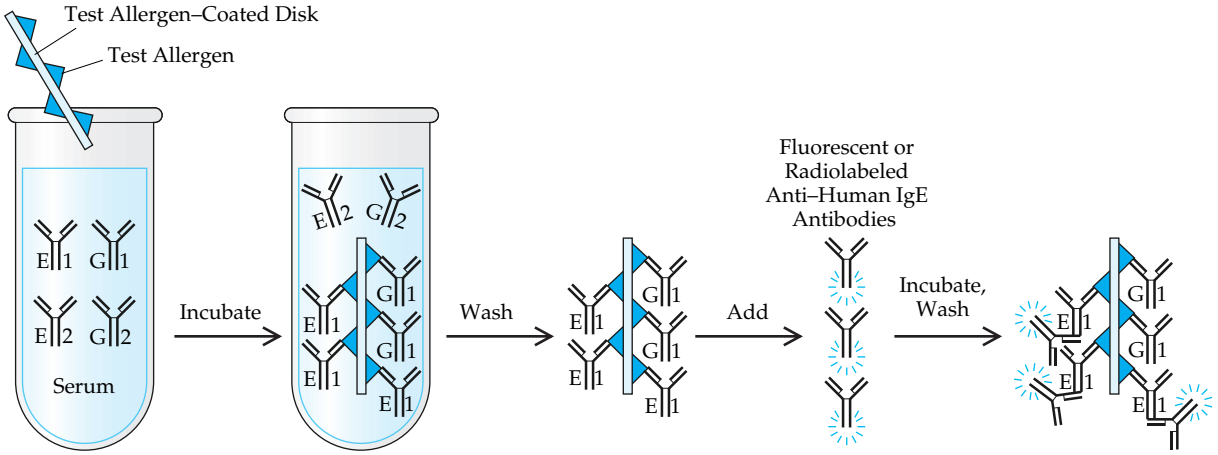


Figure 1 **The radioallergosorbent test (RAST). A solid-phase disk coated with the test allergen is incubated with the patient's serum. IgE and IgG antibodies to the test allergen (E1 and G1, respectively) will bind with the allergens on the disk, whereas IgE and IgG antibodies to other allergens (E2 and G2, respectively) will remain free in the serum. After the free antibodies have been washed away, the disk is incubated with antibodies against human IgE that have been labeled with a radioactive or a fluorescent tracer. The tagged anti–human IgE antibodies will then bind IgE attached to the disk; a second washing then removes any unbound tagged antibody. The level of radioactivity or fluorescence is then proportional to the amount of specific IgE against the antigen (E1). IgG against the antigen (G1) does not react with the tagged antibody and therefore is not counted in this test.**

out previous epicutaneous testing.[7] Food allergens should never be used for intradermal testing, because they are associated with a high rate of false positive irritant responses. Furthermore, some foods (e.g., peanuts and shellfish) are such potent antigens that they could provoke severe systemic reactions if injected intradermally.

Inaccurate or incorrect skin-testing results can occur for a variety of reasons. For example, the use of low-potency extracts can lead to false negative results, as can certain patient factors, such as (1) age (wheals are small in infants, increase until age 50, and then decline), (2) race (whites produce smaller wheals than African Americans[8]), and (3) antihistamine use (including drugs with antihistaminic properties, such as tricyclic antidepressants). In addition, skin-testing results depend on vascular leak; medications such as adrenergic agents can inhibit this response, leading to false negatives. False positives most often result because of irritant reactions, dermatographism, or a nonspecific reaction from a nearby strong reaction (a so-called bystander reaction).

RADIOALLERGOSORBENT TESTING

RAST and other in vitro tests measure the concentration of nonspecific and allergen-specific IgE in the patient's serum. Because these tests do not depend on IgE-mediated histamine release for their interpretation, they are not adversely affected by the use of antihistamines and other medications (except for anti-IgE, omalizumab [see below]). Although there are circumstances in which high levels of nonspecific IgE can be found, determination of nonspecific IgE is generally of little value, because IgE concentrations vary substantially and there is significant overlap between patients with atopic disease and patients with nonatopic disease. However, the determination of allergen-specific IgE can be useful, especially in patients in whom skin testing cannot be performed (e.g., because of skin disease or inability to stop using antihistamines).

RAST is the most common method of determining allergen-specific IgE in the serum [*see Figure 1*]. This test involves adding the patient's serum to a solid phase (usually a disk) coated with the allergen to be tested. Antibodies in the patient's serum that are specific for the allergen will bind to the solid phase. After the disk is washed, to remove the unbound antibodies to other allergens, antibodies against human IgE that have been tagged with a radioactive isotope are added. The disk is then washed again, to remove unbound tagged anti-IgE. The level of radiation that is present after washing the disk is directly proportional to the quantity of allergen-specific IgE in the patient's serum. Comparing these values with known standards allows for the determination of allergen-specific IgE. Gaining in popularity is the CAP-RAST system, which is a test for specific IgE that incorporates a solid phase consisting of an encapsulated hydrophilic carrier (in the shape of a cup or "CAP") to which antigen is covalently bonded. This allows for better allergen attachment and much more accurate quantification of specific IgE than can be obtained with traditional RAST testing. Furthermore, CAP-RAST testing is usually performed with fluorescently labeled anti-IgE, as opposed to the radiolabeled anti-IgE in traditional RAST testing. Although the CAP-RAST method is different from traditional RAST testing, some laboratories may refer to both of these modalities as RAST tests.

Although RAST results generally correlate with allergic sensitivity, RAST is more likely than skin testing to produce false positive results. As such, the sensitivity of RAST is lower than that of skin testing. Therefore, skin testing is still the preferred method of identifying the allergens to which a person is sensitive.

INTERPRETATION OF IgE TEST RESULTS

Regardless of the modality used to test for IgE, all results must be correlated with the clinical findings. Only tests whose results fit with the patient's symptoms should be considered relevant for explaining those specific symptoms. In other words, a positive result is useful for therapeutic intervention only if the patient has symptoms when exposed to the allergen, and a negative test is useful only if the patient has no symptoms on exposure to the allergen. An example would be a patient who has a positive skin test to a tree pollen yet has no symptoms in the spring but instead has symptoms in the fall, in a region devoid of tree pollen at that time of the year. Even if tests for weeds and

molds were negative, this seasonality of symptoms would still suggest that a fall pollen or untested mold spore is to blame for the symptoms rather than trees, as the testing would suggest. A positive skin test in the absence of exposure or symptoms, however, does not mean that the patient will not develop symptoms to the antigen at some time in the future. In general, the clinician should use the clinical history to guide all testing modalities, rather than using the testing to try to identify unknown triggers.

Treatment

ENVIRONMENTAL CONTROL

The most effective therapeutic intervention for atopic disease is complete removal of the offending allergen or allergens from the patient's environment [*see Table 3*]. For example, environmental control for a patient who is allergic to dust mites would include encasing the pillows and mattress in dust-mite–proof covers, washing all bedding in hot (> 130° F) water weekly, and lowering the ambient humidity in the house to below 45%. Some authorities also recommend the removal of bedroom carpet as an additional control for dust-mite exposure; this recommendation is controversial, however. For pet-allergic patients, the pet should be removed from the household or, at a minimum, should be kept out of the bedroom at all times. Pollen-sensitive individuals will benefit from staying in air-conditioned environments during the time of year when the offending pollen is prevalent.

Table 3 Environmental Control for Allergy Management

General measures
Eliminate irritants, especially cigarette smoke, from home
Keep relative humidity at 45% or less by using air conditioners and dehumidifiers
Specific measures
Pollens: use air conditioner and keep windows of house and car closed; during peak pollen season, avoid outside activities
Molds: outdoor molds can be excluded by keeping windows closed; use exhaust fan in bathroom and kitchen to keep humidity at 50% or less
Dust mites: cover mattresses, box spring, and pillows with impermeable cases; all bedding should be washed in hot water (> 130° F) once a week; if possible, remove carpet; keep the humidity at 45% or less
Feathers: replace feather pillow with Dacron (washable) pillow and wash regularly
Pets: remove the pet from the home; if the patient does not agree to remove the pet, the pet should not be permitted in the bedroom; in addition, to decrease antigen shedding, the pet should be bathed twice weekly

PHARMACOLOGIC AGENTS

Although environmental control measures constitute the primary treatment for atopic disease, such interventions are sometimes impossible to carry out or do not completely eliminate the allergen and, therefore, do not fully resolve the disease. This is the point at which pharmacotherapy should be added. The medications used in allergic disease are targeted to various components of the allergic cascade [*see Figure 2*]. These medications include antihistamines and decongestants, anti-IgE, long-acting and short-acting bronchodilators, corticosteroids (both topical and systemic), leukotriene receptor antagonists, and theophylline. Although cromolyn and nedocromil sodium have an established tradition of use and favorable safety profiles, their minimal efficacy does not justify their inclusion on this list.

Antihistamines and Decongestants

Antihistamines block the action of histamine at its receptor.[9] Although there are at least four histamine receptors, most allergic symptoms have been attributed to the H_1 receptor. Symptoms mediated by histamine include pruritus, nasal itching, conjunctivitis, and the wheal-and-flare response.

H_1 receptor antagonists can be divided into two broad categories on the basis of their ability to cross the blood-brain barrier and cause sedation. The classic antihistamines, which cause more sedation, include over-the-counter drugs such as diphenhydramine and chlorpheniramine, as well as prescription medications such as cyproheptadine and hydroxyzine. These medications are potent antihistamines, but their usefulness is limited by their central nervous system side effects. Of particular significance is that CNS effects have been shown to last beyond the sedative effects of these medications, leading to decreased reaction time. Therefore, the recommended choice for long-term therapy is a second-generation or third-generation (active metabolite) antihistamine, which will produce minimal sedation. Examples of such agents include over-the-counter loratadine and the prescription drugs cetirizine (Zyrtec), desloratadine (Clarinex), and fexofenadine (Allegra). Also available as a nasal spray is azelastine (Astelin).

Antihistamines do not have a significant effect on nasal congestion. For intermittent congestion, a systemic decongestant may be used. Since phenylpropanolamine (PPA) was taken off the market because of its association with increased frequency of strokes, pseudoephedrine has been the only systemic decongestant available in the United States. Nevertheless, although decongestants provide some relief of the sensation of nasal fullness, they do not alter the underlying etiology.

Anti-IgE Therapy

The allergic response requires the presence of IgE. A newly approved therapeutic modality in the United States is omalizumab, a humanized anti-IgE monoclonal antibody that is administered subcutaneously on a biweekly to monthly schedule and can reduce serum unbound IgE to undetectable levels. This molecule consists of the hypervariable region from a mouse antibody against human IgE that is genetically grafted onto a human IgG molecule—hence the term humanized. Clinically, omalizumab has been shown to significantly reduce symptom scores in patients with allergic rhinitis, as well as to relieve symptoms and modestly improve airway function in patients with moderate to severe asthma.[10-12] This medication has the ability to block the allergic cascade at its initiation and has not been associated with significant side effects.

The appropriate end points for omalizumab therapy remain uncertain. Standard measures of allergic sensitivity are not useful in patients who are taking omalizumab: skin testing will produce negative results, and total IgE will be elevated because IgE is bound to the anti-IgE medication in circulation. Current data support continued administration for a minimum of 12 weeks. Patients with low pulmonary functions, patients who have had emergency department visits within the preceding year, and patients on high-dose inhaled corticosteroids are the most likely to respond to therapy with anti-IgE.[13] The relatively high cost of this medication in the United States may limit its usefulness.

Bronchodilators

Both short-acting and long-acting bronchodilators are available for treatment of asthma. Short-acting bronchodilators relieve bronchoconstriction but have no effect on the underlying inflammatory process, whereas long-acting bronchodilators not only provide symptomatic relief of bronchoconstriction but also may have slight anti-inflammatory properties. Unfortunately, over time there is loss of potency of bronchodilators when they are used alone (subsensitivity). This loss of potency does not occur, however, when bronchodilators are combined with an inhaled corticosteroid. Given the lack of sufficient anti-inflammatory activity and the subsensitivity of bronchodilators, the recommended use of these medications is in combination with an inhaled corticosteroid rather than as monotherapy.[14,15]

Corticosteroids

Corticosteroids inhibit the production of inflammatory cytokines and chemokines, thus reducing the inflammation and cellular recruitment to sites of disease. These medications play a major role in the treatment of allergic disease. They may be given locally (topically) or systemically (orally).

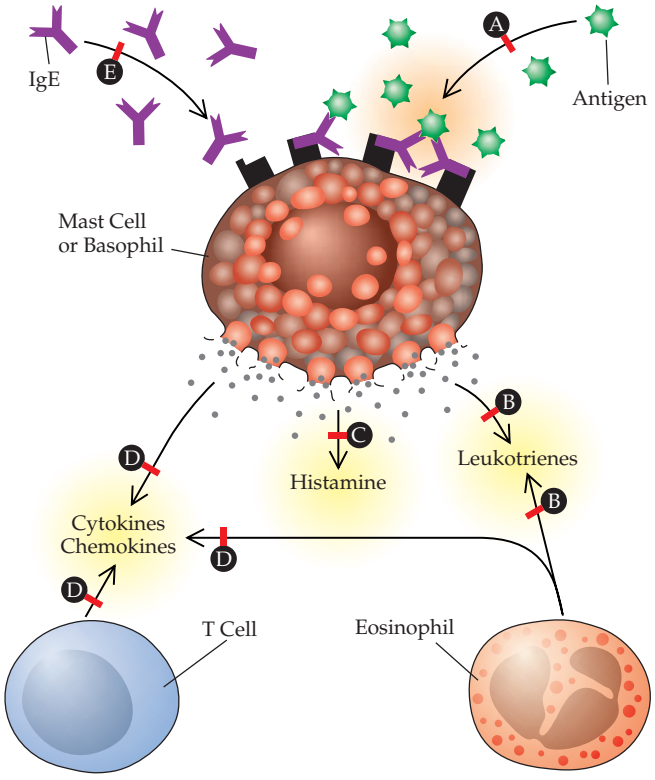


Figure 2 **Mechanisms of action of medications used in allergic diseases. Environmental control (A) minimizes exposure to the antigen to which the patient has specific IgE. If the antigen is present, it binds with specific IgE; cross-linking of the antigen-bound IgE on the surface of a mast cell or basophil starts the allergic cascade, with release of leukotrienes, histamine, cytokines, chemokines, and other mediators such as prostaglandins and proteases (not shown). Leukotriene antagonists (B) block the action of leukotrienes; antihistamines compete with histamine at H_1 receptors; corticosteroids (D) inhibit the production of inflammatory cytokines and chemokines. Anti-IgE therapy (e.g., omalizumab) (E) works by directly reducing the amount of IgE in the body. It is unclear at which sites immunotherapy and cromolyn and nedocromil sodium exert their anti-inflammatory action.**

Topical corticosteroids Topical corticosteroids are capable of potent anti-inflammatory effects and are a mainstay of allergic therapy. These medications, which are inhaled for asthma or taken intranasally for rhinitis, have the ability to abrogate the inflammatory response and interfere with multiple aspects of the allergic cascade. However, unlike antihistamines, which provide rapid relief (within 1 to 2 hours), topical corticosteroids may require 3 to 5 days of therapy before full relief is realized. In rhinitis, steroids can relieve the congestion and raise the threshold for the development of symptoms to allergen exposure.[16,17] Significant systemic effects are uncommon with inhaled or intranasal corticosteroids given at the usual recommended doses.

Oral corticosteroids Oral corticosteroids are potent at resolving and preventing most allergic disease. Unfortunately, the usefulness of chronic systemic corticosteroid use is limited by the potentially devastating side effects of these agents, which include weight gain, abnormal fat deposition, adrenal suppression, cataracts, type 2 diabetes mellitus, and osteoporosis.

In general, oral corticosteroids are prescribed only for short bursts and do not require a taper, because therapy for less than 2 weeks is not associated with adrenal suppression. Longer courses are reserved for patients whose condition has been refractory to all other standard therapies.

Leukotriene Antagonists

Leukotriene antagonists (either receptor antagonists or 5-lipoxygenase inhibitors) also have anti-inflammatory properties.[18] Leukotrienes are found at sites of allergic inflammation. Although corticosteroids affect many other inflammatory pathways, they do not seem to have a clinically significant impact on the generation and release of leukotrienes. These molecules are capable of inducing further inflammation by causing the release of additional mediators, as well as the recruitment of inflammatory cells to sites of allergic disease. Consequently, leukotriene antagonists are used in the treatment of both asthma and allergic rhinitis.[19]

Theophylline

Although generally not viewed as a major therapeutic option because of their narrow therapeutic window and significant side effects, methylxanthines still have some usefulness in asthma care. Recent asthma treatment guidelines suggest theophylline (at a serum concentration of 5 to 15 μg/ml) to be an alternative treatment in mild and moderate persistent asthma.[20] The addition of low-dose theophylline (5 to 10 μg/ml) to inhaled corticosteroids has shown benefit in asthma, with a lower risk of side effects than with higher theophylline doses.[21] As a result, methylxanthines remain worthy of consideration as part of the therapeutic regimen, especially in patients for whom cost is an issue.

IMMUNOTHERAPY

Immunotherapy, or allergy shots, involves injecting increasing doses of the offending antigen or antigens in an attempt to attenuate the specific allergic response. Clinical trials have shown that immunotherapy is successful in treating allergic rhinitis with or without associated asthma.[22,23] An immunotherapy extract is prepared on the basis of skin-testing results. The patient then receives increasing doses subcutaneously on a weekly or twice-weekly schedule for about 5 months. After this so-called build-up phase, the patient is maintained on a stable dose that is administered weekly to monthly for several years. Usual-

ly, patients achieve maximal benefit after being on the maintenance dose for 1 year. The duration of treatment is still under investigation; therefore, the discontinuance of immunotherapy must be determined on an individual basis.

Immunotherapy is usually reserved for those patients in whom environmental and pharmacologic interventions have been less than fully successful. The only patients for whom immunotherapy is almost always indicated are those who have systemic symptoms from venom (Hymenoptera) allergy. Immunotherapy is often indicated for allergic rhinitis or asthma that is clearly associated with sensitivity to specific allergens.[24] It is also used in children, because some data suggest that early treatment of allergic rhinitis with immunotherapy may prevent the subsequent development of asthma.[25] Because of the small but real risk of anaphylaxis, immunotherapy should be given only in a medical office or in another carefully screened location, where personnel and supplies are readily available to treat reactions. Similarly, patients with an FEV_1 (forced expiratory volume in 1 second) of less than 70% of predicted or those having an asthma exacerbation should not be given immunotherapy because of the risk of developing even worse bronchospasm. Currently, there is no role for immunotherapy in the treatment of food allergies.

Standard immunotherapy (also known as conventional high-dose immunotherapy) should not be confused with other invalidated and inappropriate methods of immunotherapy. These techniques, which should be avoided, include skin-titration testing and treatment (the Rinkel method), subcutaneous provocation and neutralization, and sublingual provocation.[6]

Mitchell H. Grayson, M.D., is a current recipient of grant/research support and/or member of the speakers' bureaus for Genentech, Merck & Co., Inc., and Novartis.

Phillip E. Korenblat, M.D., F.A.C.P., is a current recipient of grant/research support, a consultant, and/or a member of the speakers' bureaus for the following companies: AstraZeneca, Aventis, Genentech, GlaxoSmithKline, Merck & Co., Inc., Novartis, and Schering.

References

1. Woolcock AJ, Peat JK: Evidence for the increase in asthma worldwide. Ciba Found Symp 206:122,1997

2. Martinez FD: The coming-of-age of the hygiene hypothesis. Respir Res 2:129, 2001

3. Lewis WH, Vinay P, Zenger VE: Airborne and Allergic Pollen of North America. Johns Hopkins University Press, Baltimore, 1983

4. Korenblat PE, Wedner HJ: Allergy, theory and practice, 2nd ed. WB Saunders Co, Philadelphia, 1992, p 210

5. Naclerio R, Solomon W: Rhinitis and inhalant allergens. JAMA 278:1842, 1997

6. Practice parameters for allergy diagnostic testing. Joint Task Force on Practice Parameters for the Diagnosis and Treatment of Asthma. The American Academy of Allergy, Asthma and Immunology and the American College of Allergy, Asthma and Immunology. Ann Allergy Asthma Immunol 75:543, 1995

7. Lockey RF, Benedict LM, Turkeltaub PC, et al: Fatalities from immunotherapy and skin testing. J Allergy Clin Immunol 79:660, 1987

8. Van Niekerk CH, Prinsloo AE: Effect of skin pigmentation on the response to intradermal histamine. Int Arch Allergy Appl Immunol 76:73, 1985

9. Day J: Pros and cons of the use of antihistamines in managing allergic rhinitis. J Allergy Clin Immunol 103:S395, 1999

10. Busse WW: Anti-immunoglobulin E (omalizumab) therapy in allergic asthma. Am J Respir Crit Care Med 164:S12, 2001

11. Casale TB: Anti-immunoglobulin E (omalizumab) therapy in seasonal allergic rhinitis. Am J Respir Crit Care Med 164:S18, 2001

12. Casale TB, Condemi J, LaForce C, et al: Effect of omalizumab on symptoms of seasonal allergic rhinitis: a randomized controlled trial. JAMA 286:2956, 2001

13. Bousquet J, Wenzel S, Holgate S, et al: Predicting response to omalizumab, an anti-IgE antibody, in patients with allergic asthma. Chest 125:1378, 2004

14. Sears MR: Asthma treatment: inhaled beta-agonists. Can Respir J 5(suppl A):54A, 1998

15. Taylor DR, Sears MR, Cockcroft DW: The beta-agonist controversy. Med Clin North Am 80:719, 1996

16. Corren J: Intranasal corticosteroids for allergic rhinitis: how do different agents compare? J Allergy Clin Immunol 104:S144, 1999

17. O'Byrne PM: Inhaled corticosteroids in asthma: importance of early intervention. Inhaled Glucocorticoids in Asthma: Mechanisms and Clinical Actions. Schleimer RP, Busse WW, O'Byrne P, Eds. Marcel Dekker, New York, 1996, p 493

18. O'Byrne PM, Israel E, Drazen JM: Antileukotrienes in the treatment of asthma. Ann Intern Med 127:472, 1997

19. Grayson MH, Bochner BS: New concepts in the pathogenesis and treatment of allergic asthma. Mt Sinai J Med 65:246, 1998

20. National asthma education and prevention program expert panel report 2: guidelines for the diagnosis and management of asthma (NIH Publication No. 97-4051). National Institutes of Health, Bethesda, Maryland, 2002
http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm

21. Evans DJ, Taylor DA, Zetterstrom O, et al: A comparison of low-dose inhaled budesonide plus theophylline and high-dose inhaled budesonide for moderate asthma. N Engl J Med 337:1412, 1997

22. Dykewicz MS, Fineman S, Skoner DP, et al: Diagnosis and management of rhinitis: complete guidelines of the Joint Task Force on Practice Parameters in Allergy, Asthma and Immunology. American Academy of Allergy, Asthma, and Immunology. Ann Allergy Asthma Immunol 81:478, 1998

23. Abramson MJ, Puy RM, Weiner JJ: Is allergen immunotherapy effective in asthma? A meta-analysis of randomized controlled trials. Am J Respir Crit Care Med 151:969, 1995

24. Bousquet J, Lockey R, Malling HJ: Allergen immunotherapy: therapeutic vaccines for allergic diseases. A WHO position paper. J Allergy Clin Immunol 102:558, 1998

25. Jacobsen L: Preventive aspects of immunotherapy: prevention for children at risk of developing asthma. Ann Allergy Asthma Immunol 87(suppl):43, 2001

Acknowledgments

Figure 1 Tom Moore.

Figure 2 Seward Hung.

XII ALLERGIC RHINITIS, CONJUNCTIVITIS, AND SINUSITIS

Raymond G. Slavin, m.d.

Allergic rhinitis, conjunctivitis, and sinusitis are closely related disorders. Allergic rhinitis and conjunctivitis share the same causes and pathophysiology; sinusitis typically occurs as a complication of allergic rhinitis.

Allergic Rhinitis

Allergic rhinitis is an allergic inflammatory response in the nose. It can be classified as seasonal or perennial, depending on the allergens triggering the reaction.

EPIDEMIOLOGY

Allergic rhinitis is the most common atopic disorder in the United States. It affects about 24 million Americans—an estimated 8% of the population—with an equal distribution between males and females.[1] The prevalence of allergic rhinitis varies by age: 32% of patients are 17 years of age or younger, 43% are 18 to 44 years of age, 17% are 45 to 64, and only 8% are 65 years of age or older. The costs of treating allergic rhinitis (and indirect costs of the disorder, such as lowered productivity and time lost from work or school) are substantial. The total direct health care cost of treating allergic rhinitis is estimated at $3.4 billion.[1]

ETIOLOGY AND PATHOPHYSIOLOGY

The airborne allergens responsible for allergic rhinitis can be divided into seasonal (trees, grass, weeds, and mold) and nonseasonal or perennial (house dust mites, pets, insects).[2] These aeroallergens land on the nasal mucosa, are processed by antigen-presenting cells, and are then presented to helper T cells. In genetically predisposed persons, this interaction promotes the generation and release of cytokines that induce B cells to produce antigen-specific IgE. The IgE attaches to receptors on mast cells and basophils, and the patient is thereby sensitized. On subsequent exposure, the allergen bridges IgE molecules, resulting in release of mediators, most notably histamine.[3] Histamine causes increased epithelial permeability, vasodilatation, and stimulation of a parasympathetic reflex. As a result, acetylcholine is released, resulting in marked hypersecretion of mucus and increased blood flow. Activation of centers in the central nervous system results in sneezing.

DIAGNOSIS

Clinical Manifestations

Symptoms of allergic rhinitis may include paroxysms of sneezing, nasal congestion, clear rhinorrhea, and itching of the nose and palate. Distinct temporal patterns of symptom production may aid diagnosis. For example, seasonal allergic rhinitis symptoms typically appear during a specific time of the year when aeroallergens are abundant in the outside air. Symptoms of rhinitis that occur whenever the patient is exposed to a pet with fur suggest IgE-mediated sensitivity to that species.[4] Allergic rhinitis may result in fatigue and significant disability.

Physical Examination

The patient with allergic rhinitis may appear uncomfortable, exhibiting mouth breathing. Children in particular may have so-called allergic shiners (dark rings under the eyes). Allergic shiners develop because the edematous nasal tissue compresses the veins that drain the eyes, leading to pooling of blood under the orbits. On the bridge of the nose, a so-called allergic crease may be present—a result of continued upward rubbing of the tip of the nose (the so-called allergic salute). On nasal examination, the mucosa typically appears pale and swollen, with a bluish-gray appearance when the mucosal edema is severe. Many patients have a normal examination, although they often may be sneezing and have rhinorrhea with mucosal edema. The other physical findings tend to be present in the more severely affected patients.

Laboratory Testing

Although a careful history is the most important step toward the diagnosis of allergic disease, skin testing may be useful in pinpointing the offending allergen [*see 6:XI Diagnostic and Therapeutic Principles in Allergy*]. The simplicity, ease and rapidity of performance, low cost, and high sensitivity of skin tests make them preferable to in vitro testing.[5]

DIFFERENTIAL DIAGNOSIS

The two nasal conditions most commonly confused with allergic rhinitis are infectious rhinitis and perennial nonallergic rhinitis (vasomotor rhinitis). Infectious rhinitis is characterized by constitutional symptoms and purulent rhinorrhea. A nasal smear shows a preponderance of neutrophils, whereas in allergic rhinitis, eosinophils predominate. Perennial nonallergic rhinitis is more frequent in women and is precipitated by such nonspecific factors as changes in temperature, humidity, and barometric pressure; strong odors; alcohol; and cigarette smoke. Nasal congestion frequently shifts from side to side and is often alleviated by exercise.[6]

TREATMENT

Therapy for allergic rhinitis comprises three elements: first, minimizing contact with the allergen (environmental control); second, pharmacotherapy; and third, immunotherapy, which is reserved for selected patients. Together, these treatments ensure an excellent prognosis for allergic rhinitis.

Environmental Control

Reducing or completely avoiding the offending allergen is a vital part of allergy management [*see 6:XI Diagnostic and Therapeutic Principles in Allergy*]. In the case of seasonal allergies, keeping the doors and windows closed and the air conditioning on will reduce the aeroallergen burden manyfold.[7] Measures to avoid house dust mites should focus on the patient's bedroom and include encasing the mattress, box spring, and pillows in occlusive covers; weekly washing of bedding at 130° F or hotter, dehumidification to less than 50%; and removal of reservoirs, such as carpeting. Removal of pets is the optimal approach for pet-sensitive patients. If the patient will not part with the pet,

weekly washing of the animal will reduce airborne levels of its allergen.[8] Also, patients with allergic rhinitis appear to be more sensitive to nonspecific irritants, such as cigarette smoke.[9]

Pharmacotherapy

Oral antihistamines are effective in reducing itching, sneezing, and rhinorrhea from allergic rhinitis. A major limitation of the first-generation (classic) antihistamines has been sedation. The second-generation antihistamines—cetirizine (Zyrtec), fexofenadine (Allegra), and loratadine (Claritin) and its metabolite desloratadine (Clarinex)—produce significantly less sedation. In patients with nasal congestion, an antihistamine-decongestant combination can be used.[10] An intranasal antihistamine spray (Astelin) has also proved to be efficacious. In severe cases, a short course of oral corticosteroids may be needed.

The most effective medications for controlling symptoms of allergic rhinitis are nasally inhaled corticosteroids.[11] They include beclomethasone (Beconase), budesonide (Rhinocort), flunisolide (Flonase), mometasone (Nasonex), and triamcinolone (Nasacort). These agents are generally not associated with significant systemic side effects. Local side effects (e.g., nasal irritation and a burning sensation) are minimized if patients are instructed to direct the spray toward the ear and away from the septum.

Leukotriene receptor antagonists have been approved for use in allergic rhinitis and can be considered as a component of combination therapy, particularly if there is associated asthma.

Omalizumab (Xolair) is a recombinant, humanized, monoclonal anti-IgE antibody for treatment of moderate to severe asthma. It has been shown to significantly reduce serum IgE and to have beneficial effects on allergic rhinitis.[12] However, its cost and the present indication by the Food and Drug Administration only for asthma preclude its routine use in allergic rhinitis.

Immunotherapy

Allergen immunotherapy is highly effective in controlling symptoms of allergic rhinitis. It should be considered in patients with severe symptoms that cannot be controlled by other treatment modalities and in those with comorbid conditions such as asthma. Immunotherapy may prevent worsening of asthma or possibly prevent its development.[13] The effectiveness of symptomatic medications, particularly intranasal corticosteroids, has made immunotherapy less necessary.

COMPLICATIONS

There is good evidence that poorly managed allergic rhinitis can result in otitis media[14] and sinusitis.[15] Rhinitis and asthma frequently coexist.[16] More than that, rhinitis appears to be a risk factor for development of asthma,[17,18] and treatment of rhinitis can improve coexisting asthma.[19] Prevention of asthma is an especially important goal in patients with a family history of asthma or atopic disease and early sensitization to aeroallergens.[20]

Allergic Conjunctivitis

Allergic conjunctivitis is the ocular counterpart of allergic rhinitis, and the two often occur together. Approximately 70% of patients with allergic conjunctivitis have an associated atopic disease, such as allergic rhinitis, asthma, or atopic dermatitis.

EPIDEMIOLOGY

Seasonal and perennial allergic conjunctivitis are the most prevalent forms of ocular allergy. Most reports agree that allergic conjunctivitis affects up to 20% of the world's population.[21]

ETIOLOGY AND PATHOGENESIS

Allergic conjunctivitis is triggered by the same aeroallergens and results from the same pathophysiologic processes as allergic rhinitis [*see* Allergic Rhinitis, Etiology and Pathophysiology, *above*].

DIAGNOSIS

Clinical Manifestations

Patients with allergic conjunctivitis present with itching of the eyes, accompanied by tearing and a burning sensation. The reaction is usually bilateral, although unilateral conjunctivitis may occur in a patient who has had direct hand-to-eye contact with an allergen such as dog or cat dander.

The periocular tissues are usually swollen and reddened. The conjunctiva is injected, with mild to moderate chemosis, and there is a ropy mucous discharge in the tear film.

Laboratory Tests

Although examination of the ocular discharge in allergic conjunctivitis typically reveals large numbers of eosinophils, this test is almost never done. Instead, allergic conjunctivitis is generally diagnosed clinically. As with allergic rhinitis, skin testing may be performed to identify the offending allergen or allergens [*see* Allergic Rhinitis, Laboratory Testing, *above*].

DIFFERENTIAL DIAGNOSIS

The eye condition that is most likely to be confused with allergic conjunctivitis is infectious conjunctivitis (viral or bacterial). Patients with infectious conjunctivitis complain of matting of the eyelids, with a clear to mucopurulent ocular discharge. The conjunctiva is deeply red, and although a burning sensation is common, itching is not as profound as in allergic conjunctivitis. Vernal conjunctivitis, which may also be confused with allergic conjunctivitis, is a severe, bilateral recurrent condition of the eye often occurring in the spring. It is marked by intense pruritus and a typical cobblestone appearance of the upper eyelid.

TREATMENT

Because allergic conjunctivitis results from the same allergens as allergic rhinitis, environmental control measures and immunotherapy are also the same [*see* Allergic Rhinitis, Treatment, *above*].

Drug treatment for allergic conjunctivitis typically begins with a topical over-the-counter antihistamine-decongestant combination such as antazoline-naphazoline (Vasocon-A) or pheniramine-naphazoline (Naphcon-A).[21] The next line of therapy would include a selective H_1 receptor antihistamine, a category that includes ketotifen (Zaditor), epinastine (Elestat), levocabastine (Livostin), azelastine (Optivar), and olopatadine (Patanol). Ketotifen, epinastine, and olopatadine also have mast cell-stabilizing properties. A meta-analysis confirms the benefit of topical mast cell stabilizers and antihistamines over placebo for the treatment of allergic conjunctivitis.[22] An additional therapeutic option is a nonsteroidal anti-inflammatory agent such as ketorolac (Acular).[21] For the most severe cases of allergic conjunctivitis, the clinician may consider giving corticosteroid eyedrops—loteprednol etabonate (Lotemax) or rimexolone (Vexol)—for 2 to 3 weeks. Long-term use of these agents has been

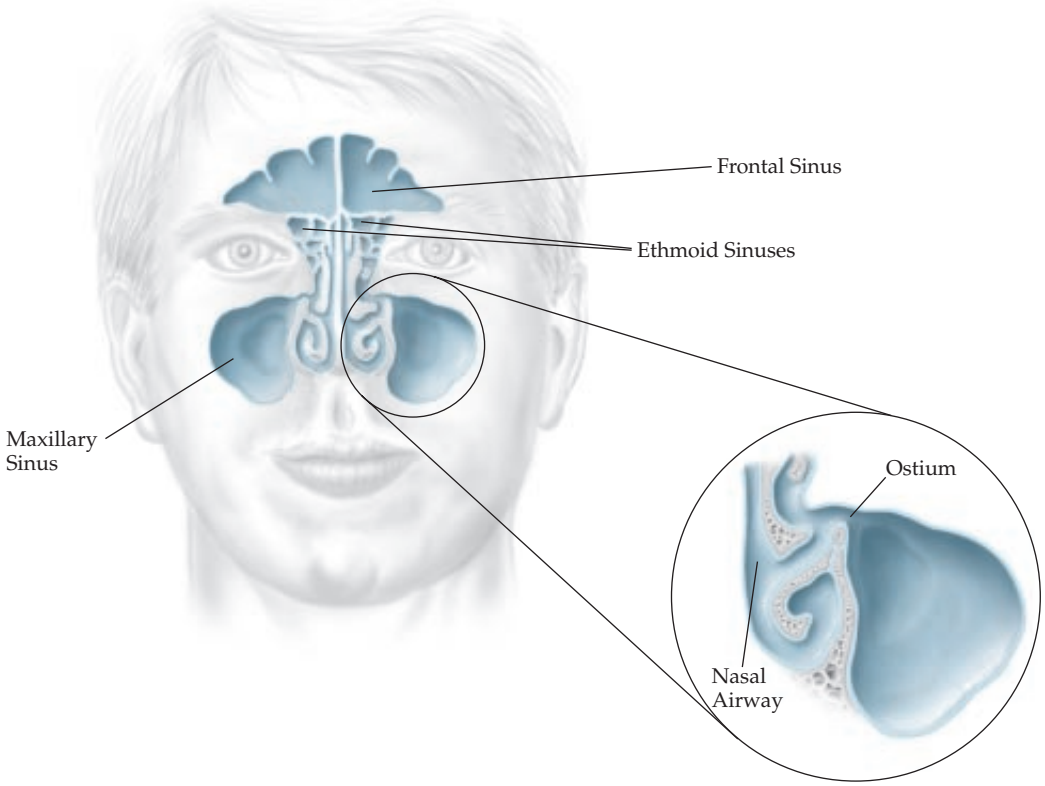


Figure 1 **The paranasal sinuses drain into the nasal passages via narrow ostia. The ostium through which the maxillary sinus drains is on the superior medial wall of the sinus, and hence the maxillary sinuses drain against the force of gravity. Edema of the nasal mucosa from allergic rhinitis can obstruct the ostia, and the resulting accumulation of mucus within the sinuses promotes bacterial infection.**

associated with the development of glaucoma, cataracts, and secondary infection and hence should be managed by an ophthalmologist.

Sinusitis

DEFINITION AND CLASSIFICATION

It has been suggested that the term rhinosinusitis may be more accurate than the term sinusitis, for the following reasons: (1) rhinitis typically precedes sinusitis, (2) sinusitis without rhinitis is rare, (3) the mucosa of the nose and sinuses are contiguous, and (4) symptoms of nasal discharge are prominent in sinusitis.[23]

Rhinosinusitis is classified as acute, recurrent acute, subacute, and chronic. Acute sinusitis is defined as inflammation of the sinuses for less than 4 weeks. Subacute sinusitis, lasting from 4 to 8 weeks, is the development and manifestation of minimal to moderate signs of sinus inflammation without an overt upper respiratory tract infection (URI) or abrupt onset of symptoms. Chronic sinusitis is defined as persistent sinus inflammation for more than 8 weeks. An operational definition of chronic sinusitis is persistent inflammation, documented with imaging techniques, continuing for at least 4 weeks after initiation of appropriate medical therapy in the absence of an intervening acute episode.

EPIDEMIOLOGY

Rhinosinusitis is the most frequently reported chronic disease in the United States, affecting 16% of the adult population. Chronic rhinosinusitis accounts for 11.6 million physician office visits a year, and the overall direct cost in the United States is estimated to be $4.3 billion annually.[24]

In one study of patients with rhinosinusitis, a 36-item health survey showed significant worsening in several domains, including bodily pain, general health, vitality, and social functioning. Comparison with other chronic diseases (e.g., chronic obstructive pulmonary disease, heart failure, angina, and back pain) revealed significantly worse bodily pain and social functioning in patients with sinusitis.[25]

PATHOGENESIS

The paranasal sinuses are composed of the ethmoid, frontal, maxillary, and sphenoid sinuses [*see Figure 1*]. Microorganisms, pollutants, irritants, and other foreign particles that escape the filtering apparatus of the nose are trapped in the mucus of the sinuses. The steady beating of the cilia that line the sinuses moves mucus out of the sinuses and into the nasal passages via the drainage ostia. This ongoing clearance of the sinuses is important for maintaining health.

The key factors that predispose an individual to rhinosinusitis are local [*see Table 1*]. The most common of these are viral URIs and allergic rhinitis. Edema of the nasal mucosa, which is characteristic of acute infectious or allergic rhinitis, results in obstruction of the ostia, decreased ciliary action in the paranasal sinuses, and increased mucus volume and viscosity. The subsequent accumulation of mucus in the sinus provides an environment for secondary bacterial infection and the conversion of mucus to mucopus.

Cultures from both adults and children with acute sinusitis grow predominantly aerobic organisms, with the heaviest yield being *Streptococcus pneumoniae, Haemophilus influenzae,* and *Moraxella* (formerly *Branhamella) catarrhalis.* Although the role of viruses and bacteria in causing acute infectious sinusitis is well established, the role of microbial infection in chronic sinus disease is much less clear. It was once believed that anaerobic organisms were responsible for instances of chronic sinusitis, but aerobes have now been implicated as the major cause. A noninfectious form of chronic rhinosinusitis, sometimes referred to as chronic hyperplastic eosinophilic rhinosinusitis, is marked by a preponderance of eosinophils and mixed mononuclear cells and by a paucity of neutrophils. It is often associated with nasal polyps, asthma, and aspirin sensitivity.[26]

DIAGNOSIS

Clinical Manifestations

Acute sinusitis The most important clinical clue to the diagnosis of acute sinusitis is the failure of symptoms to resolve after a typical cold. The previously clear nasal discharge becomes yellow or green. Fever persists and chills may develop. Pain is often felt in the cheek, or it may be referred to the forehead. The discomfort is often worse on bending over or straining. If the ostium of the maxillary sinus is blocked, pain may be severe and felt in the teeth.

On physical examination, thick, purulent, green or deep-yellow secretions are seen in the nose on the side of the diseased sinus. Because the maxillary sinus is most frequently involved, purulent secretions will be seen most often in the middle meatus, which is the drainage site of the maxillary sinus [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*]. The middle meatus may be hidden by the middle turbinate, so it may be necessary to shrink the turbinate with a topical decongestant. Once this is accomplished, the nose, particularly the middle meatus, can be examined thoroughly not only for pus but also for underlying problems, such as nasal septal deviation, spurs, and polyps. Frequently, a streak of pus is visible along the lateral wall of the oropharynx. When the diagnosis of sinusitis is in doubt, referring the patient to an otolaryngologist for fiberoptic nasopharyngoscopy can be helpful, because this technique affords a better opportunity for visualization of the drainage ostia of infected sinuses.

Table 1 Factors Predisposing to Sinusitis

Upper respiratory infection	Allergic rhinitis Anatomic variants Septal deviation Haller cells (infraorbital ethmoid cells) Hypertrophied adenoids Nasal polyps, chronic mucosal thickening Nasal or sinus tumors
Foreign bodies	Cigarette smoke Swimming and diving; barotraumas Rhinitis medicamentosa Cocaine abuse Nasal intubation Periapical abscess in a protruding tooth Dental extraction or injections

Chronic sinusitis If mucopus is not evacuated, acute sinusitis may enter a subacute or chronic phase. Chronic maxillary sinusitis may exist alone, but it is usually associated with chronic ethmoid and frontal sinusitis. The lack of pain or systemic symptoms makes chronic sinusitis difficult to diagnose on history alone. A patient may complain of dull pressure in the face or head. Chronic sinusitis generally presents as persistent, sometimes unilateral nasal stuffiness, hyposmia, purulent nasal and postnasal secretions, sore throat, fetid breath, and malaise. The secretions often pool in the hypopharynx at night, and the patient complains of increasing postnasal drainage with resultant cough and, sometimes, wheezing. On physical examination, a patient with chronic sinusitis may display an edematous and hyperemic nasal mucosa bathed in mucopus. Nasal polyps may accompany chronic sinusitis.

Nasal Smear and Sinus Culture

Nasal culture does not give an adequate picture of the organisms responsible for sinusitis. Microscopic examination of nasal secretions, however, may be of great diagnostic value. In instances of sinusitis, one sees sheets of polymorphonucleur neutrophils and bacteria. This is unlike viral URIs, in which polymorphonuclear neutrophils are scanty, or allergic rhinitis, in which a high percentage of eosinophils may be seen. Antral puncture provides a true specimen of the microbiology of the sinus cavity and is generally performed by an otolaryngologist when it is important to determine the pathogen (e.g., if fungal infection is suspected).[27]

Radiology

Two imaging modalities are used for the diagnosis of sinusitis: plain x-rays and computed tomography. In adults, plain films of the sinuses that show mucosal thickening greater than 8 mm, an air-fluid level, or opacification have been shown to correlate with positive bacterial cultures on antral punctures. In children older than 1 year, abnormal findings on maxillary sinus radiographs are generally related to inflammation of the upper airway. Crying has not been shown to be a cause of abnormalities on sinus radiographs in these children.[28]

The diagnostic value of plain films is controversial. Some authorities advise against plain radiographic studies, particularly for diagnosing chronic sinusitis. Conventional radiographs can depict changes of acute sinusitis in maxillary, ethmoid, frontal, and sphenoid sinuses but cannot delineate the status of individual ethmoid air cells or the osteomeatal complex, nor can they accurately show the extent of inflammatory disease in affected patients. For these reasons, CT is the radiographic modality of choice for examining the paranasal sinuses. Coronal CT scans demonstrate the osteomeatal complex and detect subtle disease that is not shown on plain films. The cost of CT scans used to be prohibitive, but through improved technology and the use of limited slices, the price has been reduced to the point where it is quite close to that of plain films in most centers. A limited four-slice coronal CT scan of the sinuses provides much more information than plain films do, and compared with full CT, four-slice coronal CT provides the increased information at a much reduced radiation dose and cost.[29]

Transillumination and ultrasonography are used in the diagnosis of sinusitis. Both are subject to great error, however, and cannot be recommended at the present time.

Ancillary Laboratory Tests

Other laboratory tests may have to be considered in some cases of treatment-resistant sinusitis. Underlying allergy can be determined by appropriate skin testing after a careful history has identified likely allergens. Immunologic testing may be indicated, because patients with refractory sinusitis may have immune dysfunction.[30] Associated immunodeficiency is diagnosed by serum immunoglobulin levels and by antibody responses to specific antigens such as pneumococci, diphtheria, and tetanus. Other considerations in medically resistant sinusitis include cystic fibrosis, fungal infection, and anatomic abnormalities.

DIFFERENTIAL DIAGNOSIS

The condition most often misdiagnosed as rhinosinusitis is a viral URI, which is the most important predisposing cause of acute rhinosinusitis. Rhinosinusitis is probably present if the URI symptoms do not resolve in 3 to 6 days; if the secretions, particularly postnasal secretions, turn yellow or green and persist throughout the day; and if the patient notes fullness of the head and discomfort in the face and teeth.

TREATMENT

Concern has been raised about the overdiagnosis of rhinosinusitis and unnecessary treatment with antibiotics of uncomplicated viral upper respiratory infection. More strict criteria for the use of antibiotics are symptoms for 10 to 14 days or severe symptoms, such as fever with purulent nasal discharge, facial pain or tenderness, and periorbital swelling.[31,32]

The antibiotic of choice for treatment of acute sinusitis is ampicillin or amoxicillin. An appropriate dosage of amoxicillin for acute sinusitis in the adult is 875 mg twice a day for 10 to 14 days. In patients with penicillin sensitivity, trimethoprim-sulfamethoxazole (one double-strength tablet twice a day) is an adequate alternative. More and more cases of β-lactamase–producing organisms are being reported. In penicillin-resistant sinusitis, recommended antibiotics include amoxicillin with clavulanic acid (Augmentin), the quinolones (e.g., levofloxacin [Levaquin]), and telithromycin (Ketek). Antibiotic treatment for chronic sinusitis should be continued for at least 2 weeks. If the patient reports feeling better by the last day of the regimen but still has purulent nasal discharge, the antibiotic can be continued for another 5 to 7 days.

Ancillary treatments for sinusitis, including oral decongestants and mucus thinners, have been advocated, but there are no controlled studies showing their effectiveness. The addition of intranasal corticosteroids may be modestly beneficial in the treatment of patients with recurrent acute or chronic rhinosinusitis.[33]

In some cases of chronic resistant sinusitis, surgical treatment must be considered. A wide array of surgical procedures are available, but functional endoscopic sinus surgery (FESS) has emerged as the technique of choice.[34]

COMPLICATIONS

Complications of sinusitis have decreased in incidence since the introduction of antibiotics. The complications most commonly encountered are cellulitis, abscess, and cavernous sinus thrombosis (all involving the orbit); epidural or subdural abscess; mucocele formation; and osteomyelitis.[35] It is evident in both children[36] and adults[37] not only that there is an association between sinusitis and asthma but also that sinusitis is an important trigger for asthma. In a patient who has both sinusitis and asthma, the asthma will be difficult to manage until the sinusitis is brought under control by either medical or surgical means.

PROGNOSIS

The prognosis for patients with sinusitis should be excellent if the diagnosis is made accurately and promptly and an appropriate antibiotic is administered for a sufficient period of time. Consultation with a specialist should be sought in the following situations:

- If there is a need to clarify the allergic or immunologic basis for sinusitis.
- If sinusitis is refractory to the usual antibiotic treatment.
- If sinusitis is recurrent.
- If sinusitis is associated with unusual opportunistic infections.
- If sinusitis significantly affects performance and quality of life.

Consultation is also appropriate when concomitant conditions are present that complicate assessment or treatment, including chronic otitis media, bronchial asthma, nasal polyps, recurrent pneumonia, immunodeficiencies, aspirin sensitivity, allergic fungal disease, granulomas, and multiple antibiotic sensitivities.

The author has received grants for clinical research from Genentech, Inc., and for educational activities from AstraZeneca Pharmaceuticals LP and has served as a consultant for Dey, Inc., and Aventis.

References

1. Law AW, Reed SD, Sundy JS, et al: Direct costs of allergic rhinitis in the United States: Estimates from the 1996 medical expenditure panel survey. J Allergy Clin Immunol 111:296, 2003

2. Badhwar AK, Druce HM: Allergic rhinitis. Med Clin North Am 76:789, 1992

3. Naclerio RM: Allergic rhinitis. N Engl J Med 325:860, 1991

4. Norman PS: Allergic rhinitis. J Allergy Clin Immunol 75:531, 1985

5. Dykewicz MS, Fineman S: Executive Summary of Joint Task Force Practice Parameters on diagnosis and management of rhinitis. Ann Allergy Asthma Immunol 81:463, 1998

6. Zeiger RS: Allergic and non-allergic rhinitis: classification and pathogenesis (pts 1 and 2). Am J Rhinol 3:21, 113, 1989

7. Nelson HS, Hirsch SR, Ohman JL, et al: Recommendations for the use of residential air-cleaning devices in the treatment of allergic respiratory disease. J Allergy Clin Immunol 82:661, 1988

8. Patel NJ, Bush RK: Role of environmental allergens in rhinitis. Immunol Allergy Clin North Am 20:323, 2000

9. Effect of passive smoking on respiratory symptoms, bronchial responsiveness, lung function, and total serum IgE in the European Community Respiratory Health Survey: a cross-sectional study. European Community Respiratory Health Survey. Lancet 358:2103, 2001

10. Corren J: Allergic rhinitis: treating the adult. J Allergy Clin Immunol 105:S610, 2000

11. Kaszuba SM, Baroody FM, deTineo M, et al: Superiority of an intranasal corticosteroid compared with an oral antihistamine in the as-needed treatment of seasonal allergic rhinitis. Arch Intern Med 161:2581, 2001

12. Casale TB, Condemi J, LaForce C, et al: Effect of omalizumab on symptoms of seasonal allergic rhinitis: a randomized controlled trial. JAMA 286:2956, 2001

13. Durham SR, Walker JM, Varga EM, et al: Long term clinical efficacy of grass pollen immunotherapy. N Engl J Med 341:468, 1999

14. Rachelefsky GS: National guidelines needed to manage rhinitis and prevent complications. Ann Allergy Asthma Immunol 82:296, 1999

15. Skoner DP: Complications of allergic rhinitis. J Allergy Clin Immunol 105:S605, 2000

16. Fox RW, Lockey RF: The impact of rhinosinusitis on asthma. Curr Allergy Asthma Rep 3:513, 2003

17. Greisner WA 3rd, Settipane RJ, Settipane GA: The course of asthma parallels that of allergic rhinitis: a 23-year follow-up study of college students. Allergy Asthma Proc 21:371, 2000

18. Leynaert B, Neukirch C, Kony S, et al: Association between asthma and rhinitis according to atopic sensitization in a population-based study. J Allergy Clin Immunol 113:86, 2004

19. Bousquet J, van Cauwenberge P, Khaltaev N: Allergic rhinitis and its impact on

asthma: area workshop report. J Allergy Clin Immunol 108:147, 2001

20. What are the candidate groups for pharmacotherapeutic intervention to prevent asthma? ETAC Study Group. Pediatr Allergy Immunol 11(suppl 13):41, 2000

21. Bielory L: Allergic and immunologic disorders of the eye. J Allergy Clin Immunol 106:1019, 2000

22. Owen CG, Shah A, Henshaw K, et al: Topical treatments for seasonal allergic conjunctivitis: systematic review and meta-analysis of efficacy and effectiveness. Br J Gen Pract 54:451, 2004

23. Kaliner MA, Osguthorpe JD, Fireman P: Sinusitis: bench to bedside. J Allergy Clin Immunol 99:S829, 1997

24. Anand VK: Epidemiology and economic impact of rhinosinusitis. Ann Otol Rhinol Laryngol Suppl 193:3, 2004

25. Gliklich RE, Metson R: The health impact of chronic sinusitis in patients seeking otolaryngologic care. Otolaryngol Head Neck Surg 113:104, 1995

26. Hamilos D: Chronic sinusitis. J Allergy Clin Immunol 106:213, 2000

27. Orlandi RR: Biopsy and specimen collection in chronic rhinosinusitis. Ann Otol Rhinol Laryngol Suppl 193:24, 2004

28. Kovatch AL, Wald ER, Ledesma-Medina J, et al: Maxillary sinus radiographs in children with nonrespiratory complaints. Pediatrics 73:306, 1984

29. Wippold FJ II, Levitt RG, Evens RG, et al: Limited coronal CT: an alternative screening examination for sinonasal inflammatory disease. Allergy Proc 16:165, 1995

30. Chee L, Graham SM, Carothers DG, et al: Immune dysfunction in refractory sinusitis in a tertiary care setting. Laryngoscope 111:233, 2001

31. Clinical Practice Guidelines: management of sinusitis. American Academy of Pediatrics. Subcommittee on the Management of Sinusitis and Committee on Quality Improvement. Pediatrics 108:798, 2001

32. Snow V, Mottur-Pilson C, Hickner JM, et al: Principles of appropriate antibiotic use for acute rhinosinusitis in adults. Ann Intern Med 134:495, 2001

33. Lund V, Black S, Laszlo ZS, et al: Budesonide aqueous nasal spray is effective as monotherapy in stable patients with chronic rhinosinusitis. J Allergy Clin Immunol 109:290, 2002

34. Kennedy DW: Functional endoscopic sinus surgery technique. Arch Otolaryngol 111:643, 1985

35. Sheffield RW, Cassisi NJ, Karlan MS: Complications of sinusitis: what to watch for. Postgrad Med 63:93, 1978

36. Rachelefsky GS, Katz RM, Siegal SC: Chronic sinus disease with associated reactive airway disease in children. Pediatrics 73:526, 1984

37. Slavin RG: Relationship of nasal disease and sinusitis to bronchial asthma. Ann Allergy 49:76, 1982

Acknowledgment

Figure 1 Alice Y. Chen.

XIII URTICARIA, ANGIOEDEMA, AND ANAPHYLAXIS

VINCENT S. BELTRANI, M.D.

Definition

Urticaria (hives), angioedema, and anaphylaxis are the prototypical manifestations of mast cell activation. The common denominator in these conditions is the release of potent inflammatory mediators from activated mast cells[1] [*see 6:X Allergic Response*]. Urticaria and angioedema are effected primarily by activation of cutaneous mast cells, which are preferentially located around capillaries, lymphatics, appendages, and nerves in the skin. Massive activation of mast cells in the intestinal tract, respiratory tract, and central nervous system produces the multisystemic, potentially catastrophic symptom complex of anaphylaxis.

Although the three conditions have common features and may occur in various combinations, they are more easily understood when discussed individually.

Urticaria

Urticaria is a cutaneous eruption that consists of erythematous, pruritic wheals. Although urticaria is typically transient, it can be persistent, with lesions occurring for weeks to months.

EPIDEMIOLOGY

Urticaria is a common problem, with 15% to 23% of the general population experiencing at least one episode in their lifetime.[2,3] The precise prevalence of urticaria may never be known. Many patients experience transient episodes of hives and do not report them to a health care provider because of their readily identifiable cause, inconsequential nature, and spontaneous resolution. If the papular urticaria that develops after an arthropod bite is included in the spectrum of urticarial lesions, urticaria must be considered a virtually universal human experience.

ETIOLOGY

Injecting histamine into the skin will produce the so-called triple response of Lewis—a prototypical hive. This response comprises the following: (1) erythema, the clinical manifestation of vasodilatation; (2) a wheal, the result of vascular leakage; and (3) pruritus, caused by activation of dendritic itch receptors on nonmyelinated C fibers (neurons) in the epidermis [*see Table 1*]. A fourth feature of intradermally injected histamine is its spontaneous dissipation within 1 hour. Urticaria lasting longer than 1 hour is not caused solely by histamine. Multiple inflammatory mediators have been identified in the effluent of urticarial lesions, and some of these vasoactive mediators (e.g., prostaglandin D_2 and platelet-activating factor) have produced urticaria—with and without pruritus—lasting longer than 1 hour when injected subcutaneously.

PATHOGENESIS

Because histamine plays a leading role in the pathogenesis of urticaria, tracing the source and mechanism of histamine release is the key to understanding urticarial lesions. Most of the body's histamine is stored in tissue mast cells; much smaller amounts are present in basophils and CNS neurons. Mast cells at different anatomic locations, and even at a single site, can vary substantially in mediator content, sensitivity to agents that induce activation, quantity of mediator released, and response to pharmacologic agents.[1] Agents having the ability to initiate the release of mediators from mast cells are called secretagogues. There is debate regarding whether the number of cutaneous mast cells in patients with persistent urticaria increases[4] or remains unchanged.[5] However, it has been generally agreed that these cells have a lower threshold for mediator release. Thus, a more appropriate label for chronic or idiopathic urticaria would be twitchy mast cell syndrome.

A practical categorization of urticaria is a three-part classification based on the etiology and mechanism of mast cell degranulation [*see Table 2*]. The first category comprises cases with an identifiable cause; the second, idiopathic cases; and the third, mastocytosis. Mastocytosis encompasses a wide spectrum of clinical conditions, characterized by a localized or diffuse increase in mast cells in the skin or internal organs. Most cases of mastocytosis are transient, which suggests that this disorder represents a hyperplastic response to abnormal stimuli rather than a true neoplastic process.

DIAGNOSIS

The diagnosis of urticaria is made almost exclusively from an appropriate, complete history. The history should include questions about substances or circumstances that may trigger the urticaria; the clinical features of the urticaria, including its duration, the presence and degree of pruritus, and whether the urticaria is localized or generalized; underlying illnesses; any previous diagnostic procedures or therapy for the condition; and family history of urticaria. A personal or family history of atopy should also be noted. Although the occurrence of urticaria with identifiable triggers is increased in atopic patients, whether the incidence of atopic disease is higher in patients with idiopathic urticaria remains debatable.[6,7]

Many patients presenting with IgE-induced urticaria can identify the cause of their generalized, very pruritic, explosive hives. By merely avoiding that trigger, they remain symptom

Table 1 Pathologic Changes in Urticaria and Their Associated Mediators

Symptom	*Pathologic Event*	*Mediators*
Wheal	Vascular permeability	Histamine (H_1) Prostaglandin D_2 Platelet activating factor Bradykinin Leukotrienes C4, D4, E4
Flare	Vasodilatation	Histamine (H_1) Prostaglandin D_2 Platelet activating factor Bradykinin Leukotrienes C4, D4, E4
Pruritus	Sensory nerve stimulation	Histamine (H_1)

Table 2 Classification of Urticarial Lesions

- Identifiable cause
  - Immunologic
  - Nonimmunologic (e.g., cyclooxygenase pathway, opiates)
  - Physical urticaria
    - Aquagenic urticaria
    - Cholinergic urticaria
    - Cold urticaria
    - Contact urticaria (e.g., jellyfish, nettles)
    - Delayed pressure
    - Dermatographism
    - Solar urticaria
    - Vibratory urticaria
- Nonidentifiable cause (idiopathic)
  - Persistent — occurring almost daily
  - Episodic — recurrent, with days of no hives between episodes
  - Associated with an underlying disease
    - Anaphylaxis (IgE-induced)
    - Anaphylactoid (non–IgE-induced)
    - Bullous pemphigoid
    - Erythema multiforme
    - Leukocytoclastic vasculitis
    - Serum sickness (via immune complexes)
    - Systemic lupus erythematosus
    - Viral syndrome (via immune complexes)
- Mastocytosis
  - Mastocytoma
    - Urticaria pigmentosa
    - Diffuse cutaneous mastocytosis
    - Telangiectasia macularis eruptiva perstans (TMEP)
  - Systemic mastocytosis

free. These patients are at greater risk of developing fatal anaphylactic reactions, with or without urticaria.

Common Triggers of Urticaria

Before concluding that urticaria is idiopathic, the clinician must complete a systematic review of possible mast cell secretagogues. These include immunologic and nonimmunologic activators, as well as some whose mechanism is unknown [*see Table 3*].

Drugs Drugs are probably the most easily recognized of the identifiable causes of urticaria because the symptoms usually appear within 36 hours of administration of the drug.[8] The penicillins, aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs), and sulfa drugs are most commonly involved, but virtually any drug may elicit urticaria. When urticaria develops within 1 to 2 weeks after initiation of therapy with a drug that is known to cause urticaria, that drug must be suspected.

Drugs can cause urticaria via immunologic and nonimmunologic mechanisms. The best understood mechanism involves drug-specific IgE antibodies. These IgE-induced reactions typically arise within 2 weeks after a drug is started, are not dose-related, occur from seconds to minutes after administration of the drug, and may herald an anaphylactic episode [*see Figure 1*].[8] Non-IgE reactions to drugs (e.g., aspirin or other NSAIDs, opioids, and vancomycin) can occur on first exposure or from hours to days after ingestion, are dose related, and may herald an anaphylactoid reaction.[9]

Although urticarial reactions to aspirin and other NSAIDs are rarely induced through IgE, they occur most frequently in atopic persons.[10] Angioedema, with or without urticaria, is the most common symptom of NSAID hypersensitivity. Respiratory symptoms are not more likely to occur in patients who develop an urticarial reaction from an NSAID. Some cyclooxygenase-2 (COX-2) inhibitors, especially rofecoxib, are relatively safe in patients who experience urticaria or angioedema from standard NSAIDs.[11] Patients whose history suggests a non-IgE drug reaction should not undergo routine skin testing and radioallergosorbent testing (RAST) [*see 6:XI Diagnostic and Therapeutic Principles in Allergy*].

Drug-specific IgE antibodies can be detected by skin testing, but penicillin is the only antibiotic for which reliable skin-test reagents have been developed. Standardized antigens for penicillin include penicilloyl-poly-L-lysine (penicilloyl polylysine), which is considered the major determinant and is commercially available, and several investigational minor determinants. With these reagents, numerous studies have documented the presence of penicillin-specific IgE antibodies in patients who have experienced penicillin-induced urticaria. In contrast, IgE antibodies to other antibiotics have not been demonstrated routinely in patients who have experienced an antibiotic-induced urticaria. Although it is possible that these reactions are not IgE mediated, it is more likely that the IgE antibodies have not been detected because the patients had antibodies directed against a drug metabolite not used in the testing.[12] Consequently, the diagnostic test for confirming drugs (other than the penicillins) as an identifiable cause of an individual's urticaria is to carefully rechallenge the patient, under direct medical su-

Table 3 Mast Cell Secretagogues

Immunologic activators (act on receptors)	IgE antigens (e.g., foods, drugs, latex) IgG directed against IgE (autoimmunity) Anti-FcεRI (IgE receptor) antibodies Lectins (e.g., strawberries, conconavalin A) Neuropeptides (e.g., substance P, somatostatin) Complement activators (C3a, C5a) Radiocontrast media Blood products Cytokines IL-1, IL-3, IL-6 Granulocyte-macrophage colony-stimulating factor Histamine-releasing factors (HRF) c-*kit* ligand
Nonimmunologic activators	Ionophores (opiates, adrenocorticotropic hormone [ACTH], compound 48/80) Arachidonic acid metabolic pathway inhibition (nonsteroidal anti-inflammatory drugs) Direct effect on cell Opiates (e.g., morphine, codeine) Radiocontrast media Peptides Jellyfish, lobster, eosinophil major basic protein (EMB), polymyxin B, defensins Irradiation Dextran Physical contact (pressure, light, water)
Mechanism unknown	Alcohol, amphetamine, bradykinin, ciprofloxacin, papaverine, rifampin, thiamine, thiopental, tolazoline, vancomycin

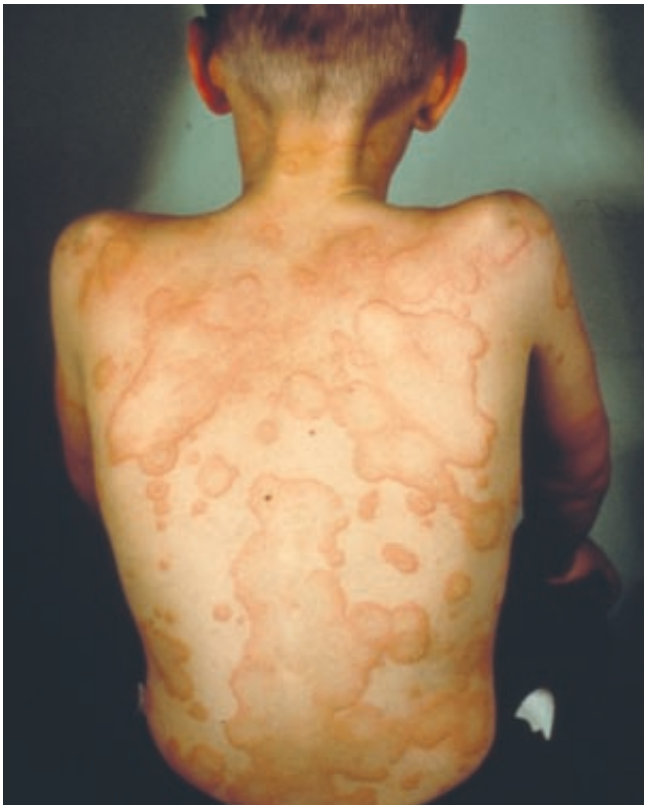


Figure 1 **Generalized, symmetrical, very pruritic urticaria appeared within 10 minutes of an intramuscular injection of penicillin G in this boy. The reaction is polymorphic, with papular urticaria evolving to larger, evanescent urticarial plaques that appear annular because of central clearing. The patient in this photograph subsequently demonstrated a positive prick test to the penicillin allergen penicilloyl polylysine.**

pervision, several weeks after the original episode has resolved. In practice, this is rarely done unless the drug in question is absolutely required to treat a disease.

Foods Foods and additives are the second most easily recognized IgE-induced trigger. Symptoms usually appear within 1 hour after ingestion, and in 80% of these cases, GI symptoms (e.g., cramps, diarrhea, and nausea and vomiting) also occur. Respiratory or, less frequently, cardiovascular symptoms may also accompany or precede cutaneous reactions.[13]

Foods are a common cause of urticaria.[14,15] Although studies of different populations of patients with urticaria provide estimates of the prevalence of food-induced urticaria, prevalence in the general population is unknown. Determining prevalence is complicated by the fact that even in patients with histories of adverse reactions to foods, only about 60% or fewer have reproducible reaction to foods.[16]

Urticarial reactions to foods may result from exposure by ingestion, injection, contact, or inhalation. Eggs, peanuts, milk, nuts, soy, wheat, fish, and shellfish are the foods most often implicated in allergic reactions, but IgE-mediated reactions to numerous other foods and to contaminating substances in foods, such as molds or antibiotics, have been reported.[17] Certain foods, such as egg white, strawberries, and shellfish, have been shown to contain substances that liberate histamine directly through a nonimmunologic mechanism.[18] Urticaria can also result from the ingestion of foods that contain large amounts of histamine, either naturally or as a result of spoilage (e.g., scombroidosis) [*see 8:I Management of Poisoning and Drug Overdose*].

Some children who experience urticaria after exposure to certain foods such as milk, eggs, soy, or wheat early in life may later tolerate these foods without difficulty. Loss of sensitivity to foods such as peanuts,[19] nuts, or fish may occur less frequently.[20]

The diagnostic tools available to determine whether foods play a role in the production of urticaria in a patient include the history, physical examination, skin testing or RAST, diet and symptom diaries, elimination diets, and food challenges. Although slightly less sensitive than skin-prick test, RASTs for specific IgE antibodies are more widely available; they require only a serum sample and are performed by commercial laboratories, and therefore, they are practical in most primary care practices. As with skin tests, a negative result is very reliable in ruling out an IgE-mediated reaction to a particular food, but a positive result has low specificity.[21] Many patients have positive skin tests and RASTs to several members of a botanical or animal species, indicating immunologic cross-reactivity, but very few patients have symptomatic intrabotanical or intraspecies cross-reactivity. The practice of avoiding all foods within a botanical family when one member is suspected of provoking allergic symptoms generally appears to be unwarranted.[21]

Occupational and hobby exposures Contact urticarial reactions are seen in certain occupational situations, such as health care workers (latex induced) and food handlers (shellfish). Atopic individuals are at a higher risk of developing these immediate-type contact reactions, which may present as pruritus, urticaria, or anaphylaxis.

Latex or natural-rubber latex hypersensitivity is a fairly common identifiable IgE-induced cause of urticaria [*see 2:V Contact Dermatitis and Related Disorders*]. These patients may experience localized urticaria at the contact area, generalized urticaria with angioedema, or urticaria with systemic involvement (including anaphylaxis). The diagnosis is made from a history of exposure and confirmed by a skin-prick test or RAST.

Systemic illness Urticaria occurs in a variety of autoimmune and infectious diseases. Urticaria is rarely the sole symptom of an underlying disease, however. If the history and physical examination do not suggest an underlying problem, routine laboratory testing is not indicated.

Generalized, urticarial lesions that persist for longer than 24 hours or that burn or sting more than they itch may be a manifestation of rheumatoid arthritis, systemic lupus erythematosus, or other rheumatic disease. Lesions associated with rheumatic illness usually do not blanch on diascopy and may leave ecchymosis and eventually hyperpigmentation. Patients who present in this manner should be assessed for rheumatic disease [*see 15:I Introduction to Rheumatic Diseases*].

Approximately 5% to 10% of patients with chronic urticaria have been reported to have antithyroid antibodies but are clinically and biochemically euthyroid.[22,23] For that reason, autoimmune thyroid serologic studies have been recommended for patients with chronic urticaria.[24] There is only anecdotal evidence that treating these patients with exogenous thyroid hormone leads to significant improvement of their urticaria, however.[25,26]

Changes in mast cell reactivity apparently can be part of the immune response to infection. Urticaria reportedly can be a feature of streptococcal pharyngitis, otitis media, infectious mononucleosis, and hepatitis (a slightly higher incidence of

hepatitis C antibodies has been reported in patients with urticaria, but whether there is a causal relationship is questionable). Pathogens reportedly associated with urticaria include coxsackievirus, *Mycoplasma*, fungi, and *Candida*. A causative relationship of urticaria with *Helicobacter pylori* has not been confirmed.[27,28] Extensive searches for occult focal infections (e.g., sinusitis) as the cause of urticaria are consistently unsuccessful.[29]

A number of parasitic infestations produce transient urticaria.[30] The urticaria in these patients usually appears from the second to the sixth week of infestation. Random examinations of stool for ova and parasites rarely, if ever, prove positive in patients with urticaria who do not have typical symptoms of parasitic infestation.

Psychological factors Emotional stress can influence mast cell and IgE activity, resulting in the release of vasoactive mediators and exacerbations of chronic urticaria.[31] However, there is no good evidence that psychological factors by themselves can cause urticaria, so urticaria without an identifiable cause should not be dismissed as a psychosomatic illness.

Neoplasms Lymphomas and carcinomas may promote urticaria (paraneoplastic syndrome), but urticaria in patients with neoplasms is usually coincidental. In most cases, the malignancy is known; current evidence does not warrant routinely subjecting patients with unexplained urticaria to an exhaustive evaluation for an occult neoplasm.

Chronic urticaria may occur as part of Schnitzler syndrome, which also includes a monoclonal IgM gammopathy, intermittent fever, joint or bone pain, lymphadenopathy, leukocytosis, and an elevated erythrocyte sedimentation rate (ESR). In 15% of cases, Schnitzler syndrome evolves to lymphoplasmocytic malignancy.[32]

Genetic factors Several of the physical urticarias can be familial. Examples include urticaria induced by cold, heat, light, water, and vibration (see below), as well as urticaria associated with erythropoietic protoporphyria. The Muckle-Wells syndrome is a form of familial urticaria associated with deafness and amyloidosis.[33]

Localized Urticaria

Papular urticaria, some of the physical urticarias, and contact urticaria are the entities to consider when urticaria is restricted to a limited area of the body. Dermatographism, cold urticaria, delayed pressure urticaria, solar urticaria, and aquagenic urticaria are localized wheals produced by specific physical stimuli (i.e., stroking of the skin, cold, sustained pressure, ultraviolet light, and water, respectively).

Papular urticaria Papular urticaria consists of 4 to 8 mm wheals or firm papules, often in grouped clusters and especially on areas of exposed skin. Papular urticaria that is very pruritic, persists longer than typical hives, and is located on exposed parts of the body is often caused by insect bites (fleas, bedbugs, scabies, and other mites). The pattern of the eruption corresponds to the biting habits of the offending insect (e.g., mosquito bites often comprise three quasilinear lesions—referred to as breakfast, lunch, and supper), and the seasonal occurrence corresponds to the peak prevalence of that insect. IgE and IgG antibodies against mosquito antigens have been detected in human sera,[34] but there have been no reported cases of anaphylaxis or death associated with hypersensitivity to mosquitoes. Arthropod bites are the only known cause of bulla on papular urticaria. Papular urticaria persists for 2 to 10 days and may leave postinflammatory hyperpigmentation. Occasionally, healed lesions may recrudesce when fresh crops appear.[35]

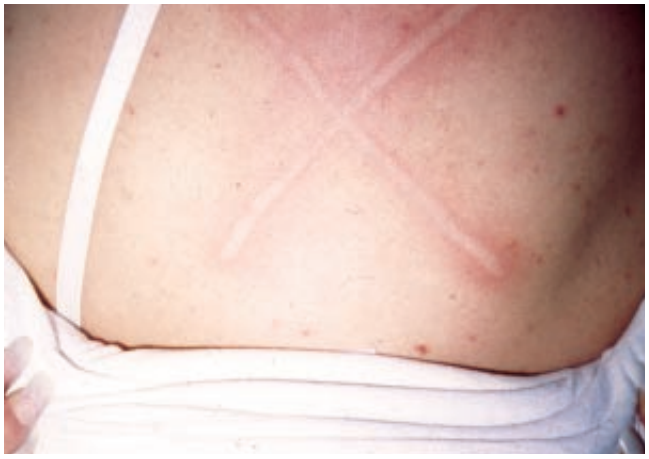


Figure 2 **Dermatographism elicited by gentle stroking of the skin of the back.**

Dermatographism Firm stroking of the skin may elicit a wheal and erythema in 5% of a healthy population, but only in a minority of these persons does it also cause any pruritus (so-called symptomatic dermatographism) [*see Figure 2*]. The etiology of dermatographism is uncertain, but passive transfer tests are sometimes positive. Dermatographism (Darier sign) is a common finding in patients with idiopathic urticaria and may be associated with other conditions. For example, dermatographism can be elicited in more than 90% of patients with mastocytosis.[36] Confirmation of mastocytosis always requires biopsy, however. The elicitation of symptomatic dermatographism in patients with urticaria supports the use of both H_1 and H_2 receptor antagonists, which may more effectively reduce wheal size and duration of urticaria.[37]

Delayed pressure urticaria Urticaria that results from localized, continuous (4 to 6 hours) pressure is seen most often in patients with persistent urticaria without an identifiable cause.[38] It may be associated with systemic complaints such as myalgias, arthralgias, and fever. It responds best to aspirin or NSAIDs and poorly to antihistamines.

Cold urticaria The lesions of cold urticaria develop 5 to 30 minutes after exposure to cold and can be caused by wind, bathing, contact, or eating cold foods or drinking cold liquids.[39] Although the urticaria may appear during the period of exposure, more often it develops upon rewarming of the skin. The urticaria usually lasts approximately 30 minutes and resolves spontaneously.

Cold urticaria is often idiopathic and acquired. Patients with these lesions usually have a positive response to an ice-cube–challenge test.[40] Rare forms of acquired cold urticaria include delayed, localized, and reflex cold urticaria. In delayed cold urticaria, lesions develop several hours after exposure; localized cold urticaria lesions occur only at sites of injections or bites; and reflex cold urticaria lesions present as widespread whealing in response to a fall in core body temperature.

Much rarer than acquired forms of cold urticaria is familial cold urticaria. In this autosomal dominant disorder, lesions appear 30 minutes after exposure to generalized cooling, rather than to local application of cold, and may persist for up to 48 hours.

Solar urticaria Solar urticaria is a rare idiopathic disorder in which erythema heralds a pruritic wheal that appears within 5 minutes after exposure to a specific wavelength of light and dissipates within 15 minutes to 3 hours after onset [*see Figure 3*].[41,42] Solar urticaria is usually provoked by light in the visible spectrum, although the specific wavelength that leads to mast cell degranulation may vary from patient to patient. The severity of the reaction depends on the duration of the exposure, the intensity of the irradiation, and the light spectrum.[43] These reactions are believed to result from the development of an antigenic photoproduct, which then triggers an IgE-mediated response. Patients should usually be referred to an allergist or dermatologist for provocative testing.

Aquagenic urticaria Urticaria that appears 2 to 30 minutes after water immersion, regardless of its temperature or source (seawater or tap water) has been reported in a few patients.[44] These pruritic, follicular, cholinergic-like wheals can be reproduced by applying wet compresses to the patient's back for at least 30 minutes. It is believed that aquagenic urticaria occurs when sensitized mast cells are activated by a water-soluble antigen that diffuses through the epidermis, causing the release of acetylcholine and histamine.

Vibratory urticaria Urticaria that follows massage and vigorous toweling has been described in a single family.[45]

Contact urticaria Immediate contact reactions can appear on normal or eczematous skin within minutes to an hour after exposure. The reaction then will disappear within a few hours. Itching, tingling, or burning accompanied by erythema are the mildest form of contact reactions. They are often caused by cosmetics, fruits, and vegetables. Generalized urticaria after a local contact is a rare phenomenon, but it can occur with some allergens.[46]

Contact reactions may have either immunologic or nonimmunologic mechanisms. Immunologic mechanisms require prior sensitization to the causative agent. The respiratory and gastrointestinal tracts are typically the routes of sensitization, but sensitization to natural latex and some foods may occur through the skin. The substances causing immunologic immediate contact reactions are usually proteins. Foods most commonly involved are fish, shellfish, and wheat flour. Most cases of protein contact dermatitis develop after the person has handled food products for a protracted period. Symptoms usually appear within 30 minutes after direct cutaneous contact with the offending agent.[47] Specific IgE antibodies against the causative allergen can be found by skin testing or RAST.

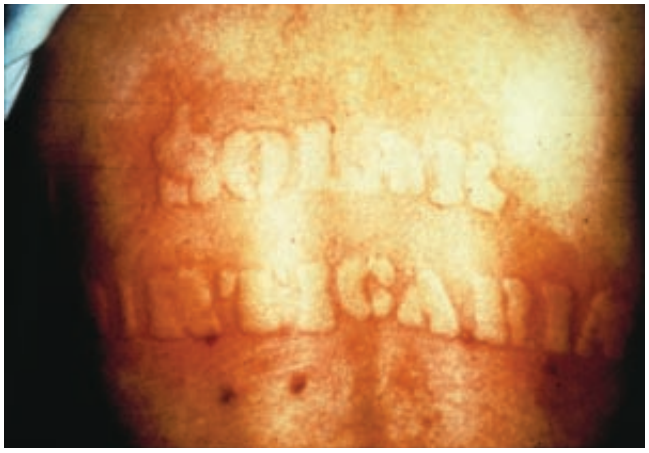


Figure 3 **Solar urticaria.**

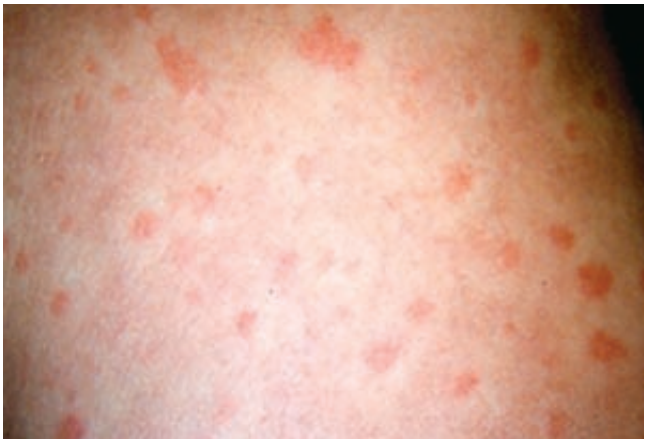


Figure 4 **Lesions of generalized urticaria tend to be symmetrical and sometimes have a halo of pallor surrounding the wheal.**

Most immediate contact reactions are nonimmunologic and occur without previous sensitization. These reactions remain localized. The pathophysiology of nonimmunologic immediate contact reactions has not been established, but it may involve direct influence on dermal vessel walls or a non-IgE release of inflammatory mediators. A list of chemicals that cause occupational allergic contact dermatitis can be found on the Internet at http://www.haz-map.com/allergic.htm.

Generalized Urticaria

The clinical features and natural history of generalized urticaria are as varied and unpredictable as the etiology. Generalized lesions tend to be numerous and symmetrical. Characteristically, they are intensely pruritic, especially at onset. Except for cholinergic papular urticaria, little information about the etiology can be obtained from the morphology. Individual lesions fade completely within 24 hours. Occasionally a halo of pallor surrounds the wheal [*see Figure 4*].

Cholinergic urticaria The lesions of cholinergic urticaria are highly distinctive, consisting of 2 to 3 mm scattered papular wheals surrounded by large, erythematous flares. These lesions are extremely pruritic, and they may affect the entire body but often spare the palms, soles, and axilla.[48] Precipitating stimuli include exercise, warm temperature, ingestion of hot or spicy foods, and possibly emotional stress. The condition often remits within several years but can last for more than 30 years. The diagnosis can be made by provocation with exercise or a hot bath. Cholinergic urticaria can be aborted by the prompt application of cold water or ice to the skin, and a refractory period of up to 24 hours can be induced by a hot bath.

Table 4 H_1 Antihistamines Available for Treatment of Urticaria

Chemical Group	*Agents*	*Antihistaminic Activity*	*Sedation*	*Anticholinergic Activity*	*Cost*
Ethanolamine derivatives	Diphenhydramine	+	++/+++	++	$
	Clemastine	+	++	+	$
Ethylenediamine derivatives	Tripelennamine (PBZ)	++	+	+	$
Piperidine derivatives	Azatadine	+	+	+	$$
	Cyproheptadine	+/++	+/++	+	$
	Fexofenadine (Allegra)*	++/+++	0	0	$$$
	Loratadine (Claritin)*	+/++	0	0	$$$$
Piperazine derivatives	Hydroxyzine	+++	++	+	$
	Cetirizine (Zyrtec)	+++	+/0	0/+	$$$$
Propylamine derivatives	Acrivastine	++	0/+	0	$$$
	Brompheniramine	+	+	+	$
	Chlorpheniramine	+/++	+	+	$
	Dexchlorpheniramine	+/++	+	+	$$
Phenothiazine derivatives	Promethazine	++/+++	+++	++	$
	Trimeprazine	++	+/++	++	$
Tricyclic antidepressants	Doxepin	+++++	+++	+++	$
	Amitriptyline	+++++	+++/++++	+++	$

*Considered second-generation antihistamines, which are nonsedating (Zyrtec less so) and have other anti-inflammatory properties besides being antihistaminic.

Physical Examination

Recognition of urticaria does not usually present a problem. Unfortunately, except for the contact and physical urticarias, the examination does not facilitate identification of the cause. Episodes of angioedema occur in half the patients presenting with persistent urticaria. The individual swellings of angioedema always last longer than an individual hive and are almost always nonpruritic.

Laboratory Evaluation

The use of laboratory tests in patients with urticaria should be directed toward confirmation of diagnoses suggested by the history and physical examination. Routine laboratory testing should not be performed, because it has consistently proved disappointing for the identification of an etiology. A skin biopsy is indicated if the diagnosis of urticaria is in question. A biopsy should be performed on any urticaria that lasts more than 24 hours, is only mildly pruritic or nonpruritic, is associated with vesicles or bullae, or does not respond to appropriate therapy. The subtleties of the histologic variances demand interpretation by a dermatopathologist.

TREATMENT

Eliminating or avoiding the triggers of mast cell activation is the basis of treatment for urticaria. However, this strategy may be impractical in patients with persistent idiopathic urticaria, which usually has multiple triggers. Any underlying disease should be treated. Idiopathic urticaria is managed symptomatically. Fortunately, the hyperreactive state in patients with idiopathic urticaria eventually resolves spontaneously.

H_1 Receptor Antagonists

When used appropriately, antihistamines can offer significant relief to most patients with urticaria. The more the skin lesions resemble the triple response of Lewis, the better they respond to antihistamine treatment. Urticarial vasculitis and delayed pressure urticaria are resistant to antihistamines. Antihistamines compete with histamine for H_1 receptor sites on effector cells and thereby prevent, but do not reverse, responses mediated by histamine alone. There are eight recognized chemical groups of H_1 receptor antihistamines; all effectively compete for H_1 receptor sites [*see Table 4*]. Among these groups are tricyclic antidepressants, which also have potent antihistaminic activity.

The choice of antihistamine is based on its effectiveness, frequency of administration, and side-effect profile. The dose of the agent selected should be increased to tolerance; if adequate relief is not achieved at the maximal tolerated dose, a drug from another group can be added. Patients do not all respond in the same way to agents from each group. Most of the so-called first-generation (sedative) antihistamines are virtually equivalent in effectiveness, with the major differences being the degree of sedation or anticholinergic effects. Activation of H_1 receptors in the brain is responsible for alertness; inhibiting these sites with antihistamines results in sedation. Second-generation (nonsedating) antihistamines tend not to cause drowsiness, because they cross the blood-brain barrier poorly.[49] Many patients find the itching and urticaria to be most troublesome in the evening and at night, so a useful strategy is to combine sedating antihistamines given at bedtime with nonsedating antihistamines during the day. This combination is effective, promotes compliance, and is economical. Tachyphylaxis has not been noted with H_1 receptor antagonists.

Because other mediators besides histamine are involved in urticaria, antihistamines are not a panacea. Also, none of the antihistamines have the ability to displace histamine from the H_1 receptor site, so the best clinical results are attained when the antihistamines occupy those receptors before the arrival of histamine; hence, round-the-clock dosing is necessary for patients with persistent symptoms.

An effective cocktail for persistent urticaria is fexofenadine (180 mg) or loratidine (10 mg) in the early morning and cetirizine (10 to 20 mg) in the early evening. If this is insufficient, the tricyclic antidepressant doxepin, 10 to 50 mg, can be added at bedtime. (A single dose of doxepin suppresses the histamine-induced wheal and flare for 4 to 6 days.[50]) This cocktail controls symptoms in more than three quarters of patients with persistent urticaria. Prednisone, 0.5 to 1.0 mg/kg/day, should be used only for patients with refractory idiopathic urticaria or with urticarial vasculitis. The goal of treatment should not be to attain a hive-free status but, rather, to minimize compromise of the patient's quality of life from both the disease and its treatment.

In urticaria with an identifiable cause, antihistamines are discontinued once the substance is gone from the body. In persistent urticaria, antihistamines can be sequentially discontinued when patients have been completely free of hives for at least 96 hours. At that point, the morning dose of nonsedating antihistamines can be discontinued. If the patient is still symptom free after another 96 hours, the doxepin dosage can begin to be reduced and, lastly, the cetirizine can be discontinued.

H_2 Receptor Antagonists

Human skin has H_2 receptors as well as H_1 receptors. H_2 receptors are present on the cutaneous arterioles, and their activation can result in vasodilatation (noted as flushing). For that reason, combining H_2 antagonists with H_1 antagonists can be helpful in patients who have prominent flushing, dermatographism, or angioedema.[51] The available evidence does not justify the routine addition of H_2 antagonists to H_1 antagonists in patients with persistent urticaria or urticarial vasculitis.

Beta Agonists

Beta agonists increase intracellular levels of cyclic adenosine monophosphate (cAMP), thereby reducing mediator release by mast cells and promoting vasoconstriction of cutaneous vasculature. Any explosive, generalized urticaria demands the subcutaneous administration of 0.2 ml of aqueous epinephrine 1:1000 (which has combined alpha-agonist and beta-agonist properties), in addition to H_1 antagonists and H_2 antagonists (e.g., doxepin, 10 mg). This is the treatment of choice for anaphylaxis (see below).

Oral beta agonists have been tried for chronic urticaria and angioedema in conjunction with H_1 antagonists and H_2 antagonists. Terbutaline (2.5 to 5.0 mg q.i.d.) deserves a trial in patients not responding to standard treatment. Some studies have demonstrated efficacy, and others have found none.[52,53]

Corticosteroids

Because corticosteroids do not inhibit cutaneous mast cell degranulation, they have no effect on acute urticaria. However, these agents are often used in patients with persistent urticaria whose symptoms are disabling and unresponsive to maximum standard therapy.[54] In these cases, steroids are given in a pulse dose to break the cycle of a resistant episode. The recommended starting dosage of prednisone for persistent urticaria is 0.5 to 1.0 mg/kg/day. This dosage should not be reduced until the patient shows definite clinical improvement.

A protocol for steroid therapy for patients with persistent urticaria has been recommended by the Parameters of Care Committee of the American Academy of Allergy, Asthma and Immunology.[55] Daily steroids are recommended only during the first 1 or 2 weeks for patients with persistent urticaria who have had no relief for a protracted period. The goal is then to utilize an alternate-day regimen with a gradually decreasing dosage over a period of months. Patients should be started on a daily dose of prednisone, 0.5 to 1.0 mg/kg (while continuing the maximum antihistamine regimen). If the symptoms become tolerable, the prednisone dose is decreased by 5 mg every 1 to 3 days until 25 mg a day is reached. The patient's progress is then reassessed every 1 to 2 weeks. Once the patient's condition stabilizes, the dose is decreased by 2.5 to 5 mg every 2 to 3 weeks. When the lowest dose is reached, alternate-day therapy may be tried. Usually, the alternate-day dose is 1.5 times the daily dose. Should some rebound occur on the off day, the alternate-day treatment can be given in divided doses (at 8 A.M. and at 5 P.M.). Once a maintenance dose is reached, the dose of prednisone should be reduced by 1 mg every 1 to 2 weeks.

Other Agents

There are recent reports of success using the anabolic steroid stanozolol for chronic urticaria,[56] aquagenic urticaria,[57] familial cold urticaria,[58] and cholinergic urticaria. Nifedipine, 20 mg three times daily, has been reported effective for chronic urticaria.[59] This treatment deserves further evaluation. Patients with chronic urticaria are advised to avoid aspirin and all NSAIDs, yet there are anecdotal reports of patients with urticaria who benefit from these drugs. Indomethacin has been used successfully in the management of urticarial vasculitis.[60]

Cyclosporine has proved effective in some cases of chronic idiopathic urticaria refractory to antihistamines, as well as in urticarial vasculitis and solar urticaria.[61] Doses used are 2.5 to 6 mg/kg daily. Higher doses can cause elevation in the blood urea nitrogen (BUN) and serum creatinine levels, but these have returned to normal on discontinuance of the drug.

Leukotriene antagonists have been combined with antihistamines for the management of allergic rhinitis and have been noted to be more effective than the antihistamine alone. Therefore, many allergists have tried this combination for urticaria, with some anecdotal success. There is nothing in the literature to support its use, however, and in my experience, the use of a leukotriene antagonist with an antihistamine offers no advantage for persistent urticaria without angioedema.

PROGNOSIS

Except for IgE-induced urticaria, which may progress to fatal anaphylaxis, the prognosis for the other urticarias is benign, although prolonged episodes of these disorders can be extremely bothersome. To date, there is no evidence that the natural history of any of the urticarial syndromes, whether induced by an identifiable cause or idiopathic, is influenced by treatment. Almost all cases of persistent urticaria eventually resolve, however; even the majority of cases of IgE-induced urticarias (especially those without anaphylaxis) are rarely permanent. Chronic urticaria tends to last longer in elderly persons than in younger ones. Studies of chronic (persistent) idiopathic urticaria have found that with or without treatment, 50% of cases will resolve within 6 to 12 months of onset; 20%, within 12 to 36 months; and another 20%, within 36 to 60 months. Less than 2% of cases persist for 25 years or longer. Over 50% of patients will have at least one recurrence.[62] Interestingly, although anaphylactic or anaphylactoid reactions have been noted in patients with identifiable causes of urticaria, there have been no reports of these reactions ever occurring in patients with persistent urticaria without an identifiable

cause. More than 50% of patients with idiopathic urticaria can be made comfortable with appropriate antihistamine therapy. Immunosuppression with corticosteroid dependence occurs in fewer than 5% of patients.

Angioedema

Angioedema is an episodic, asymmetrical, nonpitting swelling of loose tissue (usually skin) [*see Figure 5*]. It is usually nonerythematous and nonpruritic, and it may be painless. Angioedema rarely lasts less than 2 hours, and it frequently persists for 24 hours or longer. It may occur together with urticaria. Angioedema involving the face can be disfiguring during its course. Laryngeal swelling from angioedema may compromise the airway, leading to stridor and even asphyxiation. Gastrointestinal involvement can cause crampy abdominal pain, followed by watery diarrhea. Most cases of angioedema are a reaction to a food or a drug, but some episodes have no identifiable trigger. There are both hereditary and acquired forms of angioedema.

EPIDEMIOLOGY

It is estimated that approximately 10% of the population will experience at least one episode of angioedema.[63] Angioedema occurs episodically in 50% of patients with urticaria. Of patients who have angioedema as their primary disorder, approximately 20% will also experience episodes of urticaria.[64]

ETIOLOGY

Angioedema can be induced by a variety of mechanisms, including IgE, inhibition of the cyclooxygenase pathway of arachidonic acid metabolism, activation of the kinin-forming system, and activation of complement. In some patients, none of these mechanisms can be identified; these cases are labeled idiopathic.

IgE

IgE-induced angioedema resembles IgE allergy and is typically provoked by foods or drugs. It tends to occur in atopic persons and can be confirmed by prick skin testing or RAST.

Cyclooxygenase Inhibition

There is increasing evidence that the inhibition of the enzyme cyclooxygenase causes the de novo release of leukotrienes, an inflammatory mediator derived from arachidonic acid, in response to injury. Of particular interest in the skin is leukotriene B_4, which can induce neutrophil chemotaxis and increase vascular permeability.[65] Aspirin and other NSAIDs directly inhibit the ability of cyclooxygenase to decrease the formation of prostaglandins and thromboxanes, but not leukotrienes, from arachidonic acid. Angioedema (with or without urticaria) may occur in 100% of patients with hypersensitivity to aspirin or other NSAIDs.[11] Interestingly, not all patients who are hypersensitive to aspirin react to other NSAIDs,[66] and in one study, only 3% of patients sensitive to both aspirin and other NSAIDs reacted to the COX-2 inhibitor rofecoxib.[67]

Activation of the Kinin-Forming System

Bradykinin increases vascular permeability. Angiotensin-converting enzyme (ACE) inhibitors inhibit the kininase enzymes required for degradation of bradykinin, and the resulting elevation in bradykinin levels may lead to angioedema.[68] Angioedema has been reported in approximately 0.1% to 0.5% of patients who take ACE inhibitors.[69] However, because these agents are so widely used, ACE inhibitor–induced angioedema is relatively common.

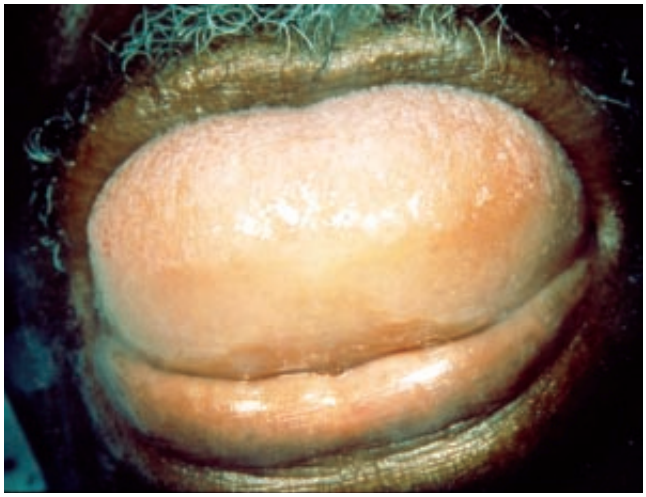


Figure 5 **Angioedema of the tongue is evident in this photograph of a 54-year-old man. This episode, the patient's fifth, was unresponsive to epinephrine, antihistamines, and prednisone; his sixth episode required intubation for 92 hours, after which the angioedema resolved spontaneously.**

Angiotensin II receptor blockers (ARBs), such as losartan and valsartan, do not increase bradykinin levels. Nevertheless, rare instances of angioedema have been reported with the use of ARBs.[70]

Complement Activation

Increased susceptibility to angioedema can result from either an inherited defect in C1-esterase inhibitor (C1-INH) activity or an acquired deficiency of C1-INH. The inherited form of the disease, known as hereditary angioedema, is rare. There are two principal types of hereditary angioedema: type 1, which accounts for 80% to 85% of cases and is caused by decreased production of C1-INH, and type 2, in which normal or elevated amounts of functionally deficient C1-INH are produced.[71] A third, very rare form of hereditary angioedema that may be X-linked has recently been described in women.[72]

Acquired angioedema results from increased metabolism or destruction of C1-INH. Two types of acquired angioedema have been described. Type 1, which is caused by excessive activation of complement and subsequent consumption of C1-INH, typically occurs in patients with rheumatologic disorders and B cell lymphoproliferative diseases. Patients with type 2 produce autoantibodies against C1-INH, leading to its inactivation.[73,74]

PATHOGENESIS

Angioedema is consistently described as a variant of urticaria in which the subcutaneous tissues, rather than the dermis, are mainly involved. However, unlike urticaria, which seems to be mediated primarily by histamine, angioedema seems to be mediated primarily by bradykinin and leukotrienes. Anecdotal evidence indicates that although urticaria can be elicited with a histamine prick or intradermal injection, the injection of histamine deeper in the dermis does not produce angioedema. On the other hand, there are patients whose angioedema will dissipate with the administration of antihistamines (especially the combination of H_1 and H_2 receptor antagonists).[75] These obser-

vations suggest that several vasoactive mediators are capable of producing angioedema.

Unfortunately, angioedema is almost never biopsied, so there are no documented pathologic descriptions of the disorder. The histopathology is always included with urticaria, and its morphology seems to be assumed. Teleologically, the vasodilatation and vascular leakage occur deeper in the skin, and the specific cellular infiltrate, if any, remains uncertain.

DIAGNOSIS

Diagnosis of angioedema is usually straightforward. Cellulitis, edematous states, trauma (stings), and fasciitis occasionally are considerations in the differential diagnosis. Insights into causes and mechanisms of induction are derived primarily from the history.

The history in a patient with angioedema—especially one who has had repeated episodes—should include the following questions: (1) Is the angioedema always, sometimes, or never associated with urticaria? (2) Is the swelling pruritic? (3) Are there accompanying gastrointestinal symptoms (e.g., pain, nausea and vomiting, or diarrhea)? (4) Is the patient taking any medications? (5) Can the patient identify any apparent triggers for the angioedema?

Patients with IgE-induced angioedema are most likely to present with concomitant urticaria. This form of angioedema may be pruritic and may progress to an anaphylactic reaction. Typically, IgE-induced angioedema occurs within 30 minutes after contact with the IgE antigen. It is most likely to occur in atopic patients. Gastrointestinal symptoms may occur but are uncommon. IgE-induced angioedema often occurs as a drug reaction, with β-lactam antibiotics being the most common trigger.

Cyclooxygenase inhibitors (i.e., aspirin or other NSAIDs) are more apt to cause nonpruritic angioedema. NSAID-induced angioedema is occasionally accompanied by urticaria.

Angioedema induced by ACE inhibitors is nonpruritic and rarely occurs with urticaria. No sex predominance has been noted in patients without gastrointestinal tract involvement, but all patients with GI involvement have been women.[76]

Complement-activated angioedema is never pruritic and is not accompanied by urticaria. In 20% to 25% of patients with hereditary angioedema, there is no family history of the disease (these cases may represent new mutations).[77] Therefore, a positive family history of hereditary angioedema is not a prerequisite for the consideration of this disorder in the differential diagnosis when typical symptoms are present. Symptoms of hereditary angioedema are usually mild or nonexistent during childhood, typically first manifesting during the second decade of life. Acquired angioedema usually develops during or after the fourth decade of life.

Hereditary and acquired angioedema have similar clinical presentations. Episodes can occur without provocation, but some episodes may be associated with trauma, medical procedures, emotional stress, menstruation, oral contraceptive use, infections, or the use of medications such as ACE inhibitors.[71] Manifestations include marked edema of the skin and lining of hollow visceral organs. GI tract involvement results in varying degrees of intestinal obstruction, with severe abdominal pain, nausea, and vomiting. Despite the absence of fever and leukocytosis, these cases are often mistaken for an acute abdomen, which occasionally leads to unnecessary surgical exploration of the abdomen. Typically, the attacks last about 2 to 5 days before resolving spontaneously.

Laboratory Tests

IgE-induced drug reactions are readily identifiable with skin-prick tests or RAST. In complement-activated angioedema, a low level of the complement component C4 is a constant finding and therefore represents a sensitive screening test. A normal level, especially during an attack, rules out both hereditary and acquired angioedema. In patients with suspected complement-activated angioedema, confirmation of the diagnosis can be obtained by measuring antigenic levels of C1-INH, which are low in 85% of patients, or functional levels, which are low in 100% of patients. Hereditary forms of complement-activated angioedema can be distinguished from acquired forms by measurement of C1q complement—levels of which are normal in hereditary forms but decreased in acquired forms.

TREATMENT

Discontinuance of the causative agent is an obvious initial step in angioedema. Emergency measures are necessary to secure the airway if there is airway obstruction by a swollen tongue, uvula, or epiglottis. Monitoring the airway in these patients until the angioedema resolves is imperative. Subcutaneous epinephrine should be given and is helpful in most types of angioedema, except those associated with low levels of C1-INH. Aerosolized epinephrine sprayed on the swollen mucous membrane may at times be helpful.

Antihistamines (both H_1 and H_2 receptor antagonists) are indicated for IgE-induced angioedema (see above). Idiopathic angioedema has been split into those presentations that respond to antihistamine therapy and those that do not.[78] Doxepin (see above) should be given to all patients with idiopathic angioedema, but results are often disappointing if this agent is administered without epinephrine. Leukotriene inhibitors counteract the vasodilation produced by leukotrienes and can reduce the edema.

Intramuscular or intravenous glucocorticoids (prednisone, 0.5 to 1 mg/kg/day, or methylprednisolone, 0.4 to 0.8 mg/kg/day) can be used as adjunctive treatment. However, the anti-inflammatory action of these agents does not affect the underlying cause of the inflammation, and they require hours to take effect. Injectable C1-INH concentrate has been developed and is effective in treating patients with hereditary angioedema,[79] but it is difficult to obtain.

To prevent future episodes of angioedema, patients should avoid identified triggers. ACE inhibitors are contraindicated in patients with idiopathic or C1-INH deficiency, and ARBs should be used only with extreme caution. Patients with idiopathic angioedema should undergo an annual general medical evaluation to identify any underlying occult disease.

Anaphylaxis

Anaphylaxis is an explosive, massive activation of mast cells, with release of their inflammatory mediators in the skin, respiratory tract, and circulatory system resulting in urticaria, wheezing, and hypotension.

The term anaphylaxis has been restricted to IgE-mediated mast cell and basophil activation. Anaphylactoid reactions, although similar in presentation, result from non–IgE-dependent mechanisms and are less likely to have a fatal outcome.[80] For practical purposes, however, it does not matter whether the patient is having true anaphylaxis or an anaphylactoid reaction, because the clinical manifestations and the treatment of these two types of reactions are identical.

EPIDEMIOLOGY

The authors of all epidemiologic reports regarding anaphylaxis believe the incidence to be underestimated because of failure to report or recognize every episode. A Dutch study estimated that only 4% to 8% of anaphylactic reactions were reported.[81] From the combined results of reported series, several significant conclusions can be drawn: First, the occurrence of atopy in anaphylaxis patients can be as high as 53%.[82] Second, the incidence of females predisposed to anaphylactic episodes can be as high as 61%.[83] Third, when the cause of anaphylaxis is found, food and drugs head the list, with peanuts and shellfish being the most common offending foods and NSAIDs and antibiotics being the most common drug offenders.[84] Fourth, cutaneous symptoms are by far the most common manifestation.[85] Fifth, the risk of anaphylaxis in hospitalized patients is reported to be 196 per million population, with the risk being highest in women and in persons younger than 30 years.[86]

ETIOLOGY

A number of substances are known to cause anaphylactic and anaphylactoid reactions [*see Table 5*]. IgE-mediated anaphylaxis is caused by agents that act as haptens (e.g., β-lactam antibiotics) or by complete antigens (e.g., venoms, foods, allergen extracts). Anaphylatoxins (C3a and C5a) often mediate reactions to human plasma and blood products. The nonimmunologic mast cell activators include radiocontrast media, opiates, and some muscle relaxants. Other anaphylactoid-inducing agents include those agents that modulate arachidonic acid metabolism (i.e., aspirin and other NSAIDs). In a number of cases, the mechanism that leads to anaphylactic or anaphylactoid reactions is unknown (i.e., idiopathic, exercise, and cold urticaria or cholinergic urticaria with anaphylaxis; mastocytosis; and some drug-induced reactions).[87] Patients with idiopathic persistent urticaria or episodic urticaria do not experience anaphylaxis.

PATHOGENESIS

Any of the mast cell secretagogues [*see Table 3*] have the potential to induce an anaphylactic or anaphylactoid reaction. Activation of the mast cell through the FcεRI receptor by an antigen releases the greatest amount of histamine. The physiologic responses to the release of inflammatory mediators include smooth-muscle spasm in the bronchi and GI tract, vasodilatation, increased vascular permeability, and stimulation of nociceptor nerve endings.

DIAGNOSIS

The classic symptoms of anaphylaxis include flushing, urticaria, angioedema, pruritus, bronchospasm, and abdominal cramping with nausea, vomiting, and diarrhea. Hypotension and shock can result from intravascular volume loss, vasodilatation, and myocardial dysfunction. Symptoms usually begin within 5 to 30 minutes after the causative agent is introduced into the body and within 2 hours after it is ingested. The shorter the latent period, the more ominous the prognosis. In rare cases, symptoms can be delayed in onset for several hours. These are called late reactions. The biphasic reaction, which includes both immediate and late reactions, tends not to be recognized and therefore is more likely to result in a fatal outcome. Least common is the protracted reaction, in which the immediate reaction persists for hours.

Table 5 Estimated Incidence or Prevalence of Acute Anaphylactic Reactions[91]

Cause	*Incidence/Prevalence*
General cause	1 per 2,700 hospitalized patients
Insect sting	0.4%–0.8% of United States population
Radiocontrast medium	1 per 1,000–14,000 procedures
Penicillin (fatal outcome)	1.0–7.5 per million treatments
General anesthesia	1 per 300 inductions
Hemodialysis	1 per 1,000–5,000 sessions
Immunotherapy (severe reactions)	0.1 per million injections

Table 6 Grading System for Anaphylaxis

Group	*Clinical Manifestations*
I	Pruritus, flushing, urticaria, or angioedema
II	Pruritus, flushing, urticaria, or angioedema Nausea, dyspnea, tachycardia, or hypotension
III	Pruritus, flushing, urticaria, or angioedema Nausea, dyspnea, tachycardia, or hypotension Bronchospasm and shock
IV	Respiratory arrest Cardiac arrest Other manifestations may be present

Clinical Manifestations

At the onset of anaphylaxis, patients often initially experience a sense of impending doom, accompanied by generalized pruritus and flushing. Almost all patients with anaphylaxis present with cutaneous manifestations that include pruritus, flushing, urticaria, or angioedema.

Anaphylaxis is graded by its clinical presentation [*see Table 6*]. Cases with signs and symptoms limited to the skin are designated as group I. Group II comprises cutaneous manifestations plus nausea, dyspnea, tachycardia, or hypotension; group III includes all the manifestations of groups I and II plus bronchospasm and true shock. Group IV consists of respiratory arrest, cardiac arrest, or both, with or without other manifestations.

Physical Examination

Cutaneous involvement Flushing, urticaria, and angioedema have been reported in 88% to 100% of patients experiencing anaphylaxis. Pruritus, especially of the scalp, soft palate, palms, soles, and anogenital areas, usually heralds an impending anaphylactic or anaphylactoid reaction or may be the only cutaneous signs of the episode. Conjunctival pruritus, injection, and edema are not unusual.

Respiratory involvement Nasal congestion (occurring in up to 56% of patients), rhinorrhea (16%), laryngeal edema, dyspnea (47%), bronchospasm (24% to 47%), cough, and hoarseness may all be part of the anaphylaxis syndrome.

Cardiovascular involvement Tachycardia and hypoten-

sion are common cardiovascular manifestations of anaphylaxis. Uncommon findings include bradycardia (6%), angina (6%), syncope, palpitations, and cardiac arrest (2% to 14%).

Gastrointestinal involvement GI symptoms, including nausea, vomiting, diarrhea, abdominal cramps, and bloating occur in 30% of patients with anaphylaxis.

Neurologic involvement Dizziness or syncope (33%), headache (up to 15%), and seizures (1.5%) may be among the presenting symptoms of anaphylaxis.

Laboratory Evaluation

Anaphylaxis is a clinical diagnosis. Laboratory studies are rarely helpful. Postmortem testing may help clarify the diagnosis in cases of so-called sudden death or in patients who are dead on arrival at the emergency department.

If a patient is seen shortly after an episode, plasma histamine, urinary histamine, or serum tryptase may be helpful in confirming the diagnosis. Plasma histamine levels rise within 10 minutes, but they fall again within 1 hour. Serum β-tryptase levels peak by 1 hour and may remain elevated for as long as 5 hours. However, a negative histamine and tryptase study does not completely rule out the diagnosis of anaphylaxis. Skin testing and RAST for the causative agent (e.g., food, Hymenoptera venom, latex, or drug), if indicated, should be performed 4 to 6 weeks after the episode for greatest sensitivity.

TREATMENT IN THE FIELD

The essential steps in the treatment of anaphylaxis are (1) prevention, (2) recognition, (3) prompt therapy, and (4) early transport to an emergency care facility.

Prevention

Prevention depends on recognition of persons at risk. Use of oral rather than parenteral medications should always be considered in patients at high risk for anaphylaxis. This includes patients with atopy or those with a possible history of allergic reactions to drugs. If drugs are administered parenterally, such patients should remain in a medically supervised area for at least 30 minutes afterward. Patients with known food or drug allergies must read labels to avoid the foods or drugs to which they are allergic. Severely food-allergic patients must be especially careful when dining out and may wish to avoid eating in restaurants altogether. Patients with a history of anaphylactic reaction to Hymenoptera venom should be given information on avoiding future stings and should be referred to an allergist for consideration of venom immunotherapy [*see 6:V Allergic Reactions to Hymenoptera*]. Patients with a history of anaphylaxis should always carry an epinephrine autoinjector (Epi-Pen).

Recognition

Immediate diagnosis of a developing reaction is imperative. Because of the risk of respiratory and cardiovascular collapse, the patient's airway, breathing, and circulation (the so-called emergency ABCs) must be rapidly assessed.

Prompt Initiation of Therapy

Anaphylaxis can rarely be overtreated. Treatment must be expeditious and appropriate. A protocol and supplies for prompt treatment should be in place at every medical office or facility. A protocol for diagnosis and management of anaphylaxis has been developed by the Joint Task Force on Practice Parameters.[88] The supplies should include oxygen, aqueous epinephrine, injectable antihistamines, intravenous or intramuscular glucocorticoids, oropharyngeal airways, and I.V. fluids. If the clinical assessment even suggests an anaphylactic reaction, it is best to call 911 and initiate therapy.

Whenever possible, decrease the absorption of the antigen. With insect bites and stings on an extremity, for example, apply a tourniquet above the injection site to block venous return and remove the insect stinger. Inject epinephrine (1:1000) locally.

Give supplemental oxygen, 6 to 8 L/min, and administer epinephrine (1:1000) subcutaneously or intramuscularly. The epinephrine dose is 0.2 to 0.5 mg in adults and 0.01 mg/kg in children. If the patient is in cardiopulmonary arrest, epinephrine (1:10,000) should be administered intravenously, in a dose of 0.1 to 1.0 mg for adults and 0.001 to 0.002 mg in children. Patients and their caregivers should recognize that more than one dose of epinephrine may be required.[89]

Intravenous H_1 antihistamines (e.g., diphenhydramine, 50 mg) and H_2 antihistamines (e.g., ranitidine, 50 mg, or cimetidine, 300 mg) should be given. If the patient can swallow, H_2 antihistamines can be given orally. Bronchospasm may be treated with aerosolized beta-adrenergic agonists (albuterol). Severe bronchospasm may require endotracheal intubation or cricothyrotomy. Respiratory failure can occur with or without upper airway compromise. Persistent hypoperfusion and ischemia may lead to myocardial infarction, cerebral ischemia, or renal failure.

Once the acute reaction is under control, systemic corticosteroids (e.g., hydrocortisone sodium phosphate, 100 mg every 2 to 4 hours) may be administered. The patient can be transferred to the emergency department.

PROGNOSIS

Most patients experience only a single episode of anaphylaxis,[82] but some patients have three or more episodes.[90] Death from anaphylaxis is uncommon. Complications are also unusual; most patients recover completely. However, respiratory failure from severe bronchospasm or laryngeal edema can cause hypoxia, which if prolonged could lead to brain injury. Hypotension and hypoxia may lead to cardiac ischemia or arrhythmias.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Galli SJ: New concepts about the mast cell. N Engl J Med 328:257, 1993

2. Doutre M: Physiopathology of urticaria. Eur J Dermatol 9:601, 1999

3. Longley J, Duffy TP, Kohn S: The mast cell and mast cell disease. J Am Acad Dermatol 32:545, 1995

4. Haas N, Toppe E, Henz BM: Microscopic morphology of different types of urticaria. Arch Dermatol 134:41, 1998

5. Bedard PM, Brunet C, Pelletier G, et al: Increased compound 48/80 induced local histamine release from nonlesional skin of patients with chronic urticaria. J Allergy Clin Immunol 78:1121, 1986

6. Kaplan AP: Urticaria and angioedema. Allergy: Principles & Practice, 5th ed. Middleton E, Reed CE, Ellis EF, et al, Eds. Mosby, St Louis, 1998, p 1104

7. Plumb J, Norlin C, Young PC: Exposures and outcomes of children with urticaria seen in a pediatric practice-based research network: a case-control study. Arch Pediatr Adolesc Med 155:1017, 2000

8. Shipley D, Ormerod AD: Drug-induced urticaria: recognition and treatment. Am J Clin Dermatol 2:151, 2001

9. Bircher AJ: Drug-induced urticaria and angioedema caused by non-IgE mediated pathomechanisms. Eur J Dermatol 9:657, 1999

10. Sanchez-Borges M, Capriles-Hulett A: Atopy is a risk factor for non-steroidal anti-inflammatory drug sensitivity. Ann Allergy Asthma Immunol 84:101, 2000

11. Sanchez Borges M, Capriles-Hulett A, Caballero-Fonseca F, et al: Tolerability to new COX-2 inhibitors in NSAID-sensitive patients with cutaneous reactions. Ann Allergy

Asthma Immunol 87:201, 2001

12. Gruchalla RS, Beltrani VS: Drug-induced cutaneous reactions. Allergic Skin Disease. Leung DYM, Greaves MW, Eds. Marcel Dekker, New York, 2000, p 318

13. Sampson HA: Food allergy. Part 2: diagnosis and management. J Allergy Clin Immunol 103:981, 1999

14. Sehgal VN, Rege VL: An interrogative study of 158 urticaria patients. Ann Allergy 31:279, 1973

15. Eggesbo M, Halvorsen R, Tambs K, et al: Prevalence of parentally perceived adverse reaction to food in young children. Pediatr Allergy Immunol 10:122, 1999

16. Bock SA, Sampson HA, Atkins FM, et al: Double-blind, placebo-controlled food challenge as an office procedure: a manual. J Allergy Clin Immunol 82:986, 1988

17. Rockwell WJ: Reactions to molds in foods. Food allergy: a practical approach to diagnosis and management. Chiaramonte LT, Schneider AT, Lifshitz F, Eds. Marcel Dekker, New York, 1988, p 153

18. Anderson JA: Milestones marking the knowledge of adverse reactions to food in the decade of the 1980s. Ann Allergy 72:143, 1994

19. Spergel JM, Beausoleil JL, Pawlowski NA: Resolution of childhood peanut allergy. Ann Allergy Asthma Immunol 85:435, 2000

20. Bock SA: Natural history of severe reactions to foods in young children. J Pediatr 107:676, 1985

21. Sicherer SH, Sampson HA: Food hypersensitivity and atopic dermatitis: pathophysiology, epidemiology, diagnosis, and management. J Allergy Clin Immunol 104:S114, 1999

22. Leznoff A, Sussman GL: Syndrome of idiopathic chronic urticaria and angioedema with thyroid auto-immunity: a study of 90 patients. J Allergy Clin Immunol 84:66, 1989

23. Turktas I, Gokcora N, Demirsoy S, et al: The association of chronic urticaria and angioedema with autoimmune thyroididitis. Int J Dermatol 36:187, 1997

24. Zauli D, Delonardi G, Foderaro S, et al: Thyroid autoimmunity in chronic urticaria. Allergy Asthma Proc 22:93, 2001

25. Rumbyrt JS, Katz JL, Schocket AL: Resolution of chronic urticaria in patients with thyroid autoimmunity. J Allergy Clin Immunol 96:901, 1995

26. Heymann WR: Chronic urticaria and angioedema associated with thyroid autoimmunity: review and therapeutic implications. J Am Acad Dermatol 40:229, 1999

27. Liutu M, Kalimo K, Uksila J, et al: Etiologic aspects of chronic urticaria. Int J Dermatol 37:515, 1998

28. Schnyder B, Helbling A, Pichler WJ: Chronic idiopathic urticaria: natural course and association with *Helicobacter pylori* infection. Int Arch Allergy Immunol 119:60, 1999

29. Nelson H: Routine sinus roentgenograms and chronic urticaria. JAMA 251:1680, 1984

30. deGentile L, Grandiere-Perez L, Chabasse D: Urticaria and parasites. Allergy Immunol (Paris) 31:288, 1999

31. Picardi A, Abeni D: Stressful life events and skin diseases: disentangling evidence from myth. Psychother Psychosom 70:118, 2001

32. Lipsker D, Veran Y, Grunenberger F, et al: The Schnitzler syndrome: four new cases and review of the literature. Medicine (Baltimore) 80:37, 2001

33. Muckle TJ: The 'Muckle-Wells' syndrome. Br J Dermatol 100:87, 1979

34. Demain JG, Taylor TM: Reactions to stinging and biting arthropods. Cutaneous Allergy. Charlesworth EN, Ed. Blackwell Science Publications, Cambridge, England, 1996, p 299

35. Stibich AS, Schwartz RA: Papular urticaria. Cutis 68:89, 2001

36. Tharp MD, Longley BJ Jr: Mastocytosis. Dermatol Clin 19:679, 2001

37. Kobza-Black A: Management of urticaria. Clin Exp Dermatol 27:328, 2002

38. Sibbald RB: Physical urticaria. Dermatol Clinics North Am 4:57, 1984

39. Wanderer AA: Cold urticaria syndromes: historical background, diagnostic classification, clinical and laboratory characteristics, pathogenesis, and management. J Allergy Clin Immunol 85:965, 1990

40. Neittaanmaki H: Cold urticaria: clinical findings in 220 patients. J Am Acad Dermatol 13:636, 1985

41. Monfrecola G, Masturzo E, Riccardo AM, et al: Solar urticaria: a report of 57 cases. Am J Contact Dermat 11:89, 2000

42. Uetsu N, Miyauchi-Hashimoto H, Okamoto H, et al. The clinical and photobiological characteristics of solar urticaria in 40 patients. Br J Dermatol 142:32, 2000.

43. Ryckaert S, Roelandts R: Solar urticaria. Arch Dermatol 134:71, 1998

44. Luong KV, Nguyen LT: Aquagenic urticaria: report of a case and review of the literature. Ann Allergy Asthma Immunol 80:483, 1998

45. Paterson R, Mellies CJ, Blankenship ML, et al: Vibratory angioedema: a hereditary type of physical hypersensitivity. J Allergy Clin Immunol 50:174, 1972

46. Lahti A: Immediate contact reactions. Curr Probl Dermatol 22:17, 1995

47. Hjorth N, Ree-Peterson J: Occupational protein contact dermatitis in foodhandlers. Contact Derm 2:28, 1976

48. Hirschmann JV, Lawlor F, et al: Cholinergic urticaria. Arch Dermatol 123:462, 1987

49. Lee EE, Maibach HI: Treatment of urticaria: an evidence-based evaluation of antihistamines. Am J Clin Dermatol 2:27, 2001

50. Goldsobel AB, Rohr AS, Siefel SC, et al: Efficacy of doxepin in the treatment of chronic idiopathic urticaria. J Allergy Clin Immunol 78:867, 1986

51. Mansfield LE, Smith JA, Nelson HS: Greater inhibition of dermographia with combination of H1 and H2 antagonists. Ann Allergy 50:264, 1983

52. Kennes B, De Maubeuge J, Delespesse G: Treatment of chronic urticaria with a beta2-adrenergic stimulant. Clin Allergy 7:35, 1977

53. Spangler DL, Vanderpool GE, Carroll MS, et al: Terbutaline in the treatment of chronic urticaria. Ann Allergy 45:246, 1980

54. Kaplan AP: Clinical practice: chronic urticaria and angioedema. N Engl J Med 346:175, 2002

55. The diagnosis and management of urticaria: a practice parameter. Part I: acute urticaria/angioedema; part II: chronic urticaria/angioedema. Joint Task Force on Practice Parameters. Ann Allergy Asthma Immunol 85:521, 2000

56. Parsad D, Pandhi R, Juneja A: Stanozolol in chronic urticaria: a double-blind, placebo controlled trial. J Dermatol 28:299, 2001

57. Fearfield LA, Gazzard B, Bunker CB: Aquagenic urticaria in HIV virus infection treated with stanozolol. Br J Dermatol 137:620, 1997

58. Omerud AD, Smart L, Reid TM, Milford-Ward A: Familial cold urticaria: investigation of a family and response to stanozolol. Arch Dermatol 129:34, 1993

59. Bressler RB, Sowell K, Huston DP: Therapy of chronic idiopathic urticaria with nifedipine: demonstration of beneficial effect in a double blinded, placebo controlled cross-over trial. J Allergy Clin Immunol 83:756, 1989

60. Millins JL, Randle HW, Solley GO, et al: The therapeutic response of urticarial vasculitis to indomethacin. J Am Acad Dermatol 3:349, 1980

61. Grattan CE, O'Donnell BF, Francis DM, et al: Randomized double-blind study of cyclosporin in chronic `idiopathic' urticaria. Br J Dermatol 143:365, 2000

62. Beltrani VS: An overview of chronic urticaria. Clin Rev Allergy Immunol 23:147, 2002

63. Hedner T, Samuelsson O, Lunde H, et al: Angioedema in relation to treatment with angiotensin converting enzyme inhibitors. Br Med J 304:941, 1992

64. Champion RH: Urticaria and angio-edema: a review of 554 patients. Br J Dermatol 81:588, 1969

65. Henig NR, Henderson WR Jr: Anti-leukotriene agents in the treatment of asthma. Current Review of Allergic Diseases. Kaliner MA, Ed. Current Medicine, Philadelphia, 2000, p 71

66. Quiralte J, Bianco C, Castillo R, et al: Intolerance to nonsteroidal anti-inflammatory drugs: results of controlled drug challenges in 98 patients. J Allergy Clin Immunol 98:678, 1996

67. Kelkar PS, Butterfield JH, Teaford HG: Urticaria and angioedema from cyclooxygenase-2 inhibitors. J Rheumatol 28:2553, 2001

68. Agostini A, Cicardi M, Cugno M, et al: Angioedema due to angiotensin-converting enzyme inhibitors. Immunopharmacology 15:21, 1999

69. Hedner T, Samuelsson O, Lunde H, et al: Angioedema in relation to treatment with angiotensin converting enzyme inhibitors. BMJ 304:941, 1992

70. Rodgers JE, Patterson JH: Angiotensin II receptor blockers: clinical relevance and therapeutic role. Am J Health Syst Pharm 58:671, 2001

71. Nzeako UC, Frigas E, Tremaine WJ: Hereditary angioedema: a broad review for clinicians. Arch Intern Med 161:2417, 2001

72. Bork K, Barnstedt SE, Koch P, et al: Hereditary angioedema with normal C1-inhibitor activity in women. Lancet 356:213, 2000

73. Laurent J, Guinnepain MT: Angioedema associated with C1 inhibitor deficiency. Clin Rev Allergy Immunol 17:513, 1999

74. Jackson J, Sims RB, Whelan A, et al: An IgG autoantibody which inactivates C1-inhibitor. Nature 323:722, 1986

75. Black AK, Greaves MW: Antihistamines in urticaria and angioedema. Clin Allergy Immunol 17:249, 2002

76. Chase MP, Fiarman GS, Scholz FJ, et al: Angioedema of the small bowel due to an angiotensin-converting enzyme inhibitor. J Clin Gastroenterol 31:254, 2000

77. Agostini A, Ciccardi M: Hereditary and acquired C-1 inhibitor deficiency: biological and clinical characteristics in 235 patients. Medicine (Baltimore) 71:206, 1992

78. Cicardi M, Bergamaschini L, Zingale LC, et al: Idiopathic nonhistaminergic angioedema. Am J Med 106:650, 1999

79. Bork K, Barnstedt SE: Treatment of 193 episodes of laryngeal edema with C1 inhibitor concentrate in patients with hereditary angioedema. Arch Intern Med 161:714, 2001

80. Luskin AT, Luskin SS: Anaphylaxis and anaphylactoid reactions: diagnosis and management. Am J Ther 3:515, 1996

81. Van der Klauw MM, Stricker BHCH, Herings RMC, et al: A population-based case-cohort study of drug-induced anaphylaxis. Br J Clin Pharmacol 35:400, 1993

82. Yokum MW, Butterfield J, Klein J, et al: Epidemiology of anaphylaxis in Olmsted County, a population-based study. J Allergy Clin Immunol 104:452, 1999

83. Yocum MW, Khan DA: Assessment of patients who have experienced anaphylaxis: a 3-year survey. Mayo Clin Proc 69:16, 1994

84. Kemp SF, Lockey RF, Wolf BL, Lieberman P: Anaphylaxis: a review of 266 cases. Arch Intern Med 155:1749, 1995

85. Perez C, Tejedor MA, Hoz A, et al: Anaphylaxis: a review of 182 patients (abstr). J Allergy Clin Immunol 95:368, 1995

86. Kaufman DW: An epidemiologic study of severe anaphylactic and anaphylactoid reactions among hospital patients: methods and overall risks—abstract from report from the International Collaborative Study of Severe Anaphylaxis. Epidemiology 9:141, 1998

87. Boxer M, Greenberger PA, Patterson R: Clinical summary and course of idiopathic anaphylaxis in 73 patients. Arch Intern Med 147:26, 1987

88. The diagnosis and management of anaphylaxis. Joint Task Force on Practice Parameters, American Academy of Allergy, Asthma and Immunology, American College of Allergy, Asthma and Immunology, and the Joint Council of Allergy, Asthma and Immunology. J Allergy Clin Immunol 101:S465, 1998

89. Korenblat P, Lundie MJ, Dankner RE, et al: A retrospective study of epinephrine administration for anaphylaxis: How many doses are needed? Allergy Asthma Proc 20:383, 1999

90. Weiler JM: Anaphylaxis in the general population: a frequent and occasionally fatal disorder that is under-recognized. J Allergy Clin Immunol 104:271, 1999

91. Sim TC: Anaphylaxis. How to manage and prevent this medical emergency. Postgrad Med 92:277, 1992

XIV DRUG ALLERGIES

Mark S. Dykewicz, m.d.
Heather C. Gray, m.d.

Adverse drug reactions are the most common iatrogenic illnesses, occurring in 1% to 15% of drug courses. Most adverse drug reactions result from nonimmunologic or unknown mechanisms (e.g., toxic overdose, toxic side effects, intolerance). Approximately 6% to 10% of drug reactions are caused by proven or suspected immunologic mechanisms mediated by specific antibodies, sensitized T cells, or both.[1] Immunologic reactions develop in patients who have become sensitized by previous exposure or continuous exposure to the same or an antigenically related drug. Although identification of antibodies or sensitized T cells directed against the drug helps confirm the diagnosis of immunologic drug reactions, diagnosis usually is based on clinical presentation.[2] This chapter provides an overview of pathogenesis, discusses recognition of both common and uncommon patterns of allergic drug reactions, and explains the application of diagnostic tests and management techniques.

Pathogenesis

Hypersensitivity drug reactions can be influenced by intrinsic properties of the drug, its administration, and the host. Drug factors that increase risk include a higher molecular weight, the ability of the drug or its reactive metabolites to readily bind to self-proteins to form antigenic hapten-protein conjugates, higher dose, parenteral (as opposed to oral) administration, and repeated exposure to the drug. Host factors that increase risk include adult age, female gender, concurrent infections, and HIV infection.[3]

Although β-lactam antibiotics covalently bind to self-protein carriers to form antigenic conjugates, many non–β-lactam antibiotics require enzyme systems, such as cytochrome P-450, to form reactive products that then bind with self-protein carriers. Consequently, testing with a parent drug may not identify sensitivity to reactive intermediate products. Rashes from sulfonamides are more likely in patients who are slow acetylators, because these patients preferentially metabolize sulfa drugs by alternative oxidative pathways that produce highly reactive metabolites that bind to self-protein carriers.

Classification of Drug Reactions

Classifying drug reactions on the basis of either the temporal relation between drug exposure and adverse manifestations or the presumptive immunologic mechanism may aid in evaluation and management. Temporally, drug reactions are classified as immediate, accelerated, or delayed. Immediate reactions occur within 1 hour and include anaphylaxis. Accelerated reactions, such as urticaria and angioedema, occur within 72 hours of administration. Delayed reactions, which occur 72 hours or more after administration, include urticarial and nonurticarial skin rashes; serum sickness–like reactions; fever; and a variety of cardiopulmonary, hematologic, hepatic, renal, and vasculitic effects.

The Gell and Coombs classification system defines four basic immunologic mechanisms for drug reactions [*see Figure 1*]. However, some clinical presentations may involve several mechanisms, and not all immunologic mechanisms conform to the Gell and Coombs classification. Type I (anaphylactic) reactions occur when drug antigen cross-links adjacent IgE antibodies that are bound to the surfaces of mast cells or basophils, with consequent cell activation and release of mediators such as histamine, tryptase, and leukotrienes. Common causes of type I reactions include antibiotics, vaccines, allergen extracts, and

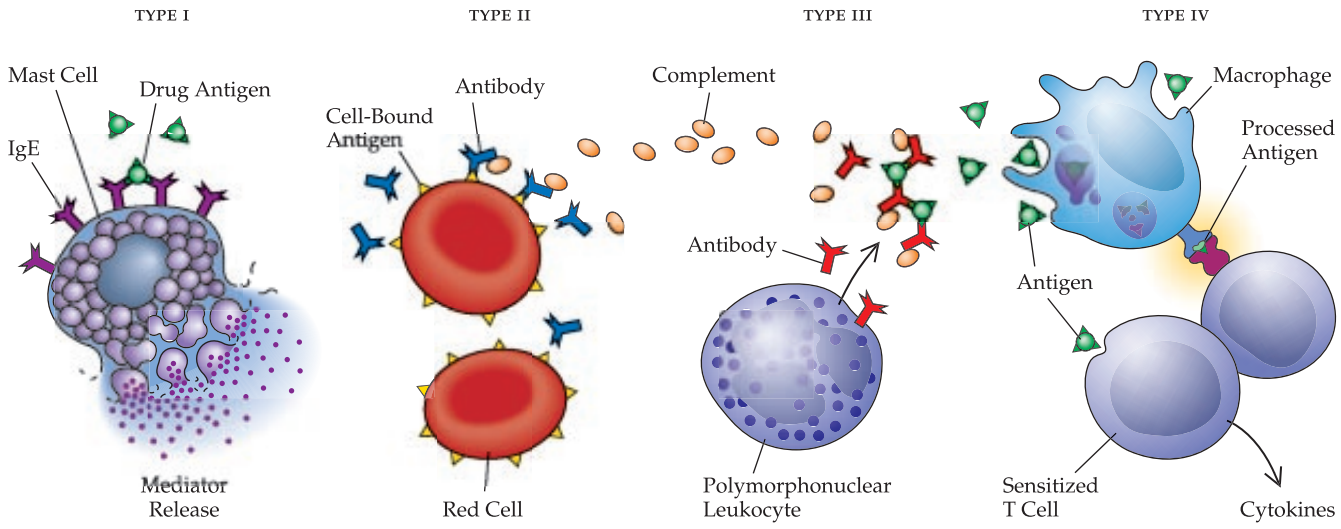


Figure 1 **The Gell and Coombs system defines four basic immunologic mechanisms for drug reactions.[25] Type I reactions (anaphylaxis) result from IgE antibodies binding to drug antigen and cross-linkage of adjacent IgE molecules, leading to mediator release. Type II (cytotoxic) reactions occur when IgG or IgM antibodies recognize drug antigen associated with cell membranes, causing complement activation. Type III (immune complex) reactions involve the formation of antigen-antibody complexes. Type IV (delayed hypersensitivity) reactions are mediated by sensitized lymphocytes.**

proteins (e.g., antisera, insulin). Reactions occur within seconds to minutes of exposure and range from full anaphylaxis to any component thereof, including pruritus, flushing, angioedema, urticaria, bronchospasm, laryngeal edema, rhinoconjunctivitis, hypotension, tachycardia, nausea, vomiting, diarrhea, and abdominal or uterine cramping. In contrast to anaphylaxis, syncope or vasovagal reactions are typically characterized by blanching rather than flushing and by bradycardia rather than tachycardia.

Type II (cytotoxicity) reactions result in cell destruction mediated by an interaction between IgG or IgM antibodies, complement, and a drug antigen associated with cell membranes. Clinical sequelae include immune hemolytic anemia, thrombocytopenia, and granulocytopenia. Heparin-induced thrombocytopenia is mediated by antibodies directed against antigen complexes of heparin and platelet factor 4 on the surface of platelets [*see 5:XIV Thrombotic Disorders*].

Type III (immune complex) reactions develop when a drug combines with antibodies to form immune complexes, the deposition of which causes tissue damage. Serum sickness is a type III reaction and may manifest as skin lesions (e.g., urticaria, angioedema, maculopapular or morbilliform rash, palpable purpura), arthralgias and arthritis, lymphadenopathy, fever, nephritis, and hepatitis. Serum sickness usually occurs after 1 to 4 weeks of drug use but may occur sooner in patients with previous exposure. Drug-induced lupus syndromes (e.g., from procainamide, hydralazine, or phenytoin) are also type II reactions. Patients rarely have renal involvement or anti–double-stranded DNA antibodies but usually have antihistone antibodies.

Type IV (delayed hypersensitivity) reactions are mediated by sensitized $CD4^+$ T cells. Allergic contact dermatitis is a classic example and typically develops 24 to 72 hours after topical exposure.

Diagnosis

To identify drug reactions, the physician needs to be familiar with general principles of drug reactions [*see Table 1*] and with the individual drugs the patient has taken. Diagnosis depends largely on the nature of the reaction and its timing in relation to drug use. For example, adverse immunologic reactions to drugs usually occur in the early weeks of drug exposure and become less common with continued drug administration. To confirm the clinical diagnosis, skin testing or drug challenges may be valuable in selected cases.

CLINICAL PRESENTATION

Dermatologic Reactions

Drug reactions involving the skin range from maculopapular and morbilliform rashes to urticaria, angioedema, erythema multiforme, erythema nodosum, bullous eruptions, and exfoliations. Drug eruptions usually occur within days of exposure (but may occur after drug cessation) and are symmetrical, truncal, and accompanied by pruritus and fever and occasionally eosinophilia. Palm and sole involvement suggests a viral exanthem rather than a drug reaction. In contrast to allergic contact dermatitis that is mediated by $CD4^+$ T cells, nonurticarial systemic drug rashes are more likely mediated by $CD8^+$ T cells.

Photosensitive drug rashes occur by two mechanisms: phototoxic or photoallergic. Phototoxic reactions are nonimmunologic, generally appear as sunburn 4 to 8 hours after light exposure, and often occur with the first exposure to the drug (e.g., tetracycline). Photoallergic reactions are typically eczematous rashes that occur after days or months of exposure (e.g., some sulfa reactions). Alteration of the drug by ultraviolet light enables conjugation of the drug to self-proteins and T cell–mediated immune responses. Neither type of photosensitive reaction is predictive of other types of adverse reactions to a drug.

Table 1 General Considerations in the Clinical Evaluation of Drug Reactions[4]

Identify drugs that have a history of causing problems in the patient; determine whether there are cross-reacting agents, and avoid them.
If the patient has a late reaction, such as a drug rash, take a careful history of all drugs used in the past month, because it is possible that the causative drug has been discontinued.
Drugs administered with impunity for prolonged periods (e.g., months to years) are rarely responsible for adverse immunologic reactions. Reactions are more likely to result from drugs introduced more recently.
Have a high index of suspicion for drug reactions whenever a patient experiences adverse clinical manifestations. Bear in mind that drug reactions can cause internal organ involvement (e.g., nephritis, hepatitis, isolated lymphadenopathy), often in the absence of eosinophilia.
If an immunologic drug reaction is suspected, stop all nonessential drugs and substitute non–cross-reactive drugs if possible.

Fever

Drug fever may occur by itself or be associated with other allergic manifestations, such as a rash. The fever stems from the release of pyrogens by phagocytic cells that have engulfed drug-IgG immune complexes or from cytokine release and other incompletely established processes associated with specific T cell activation. Drug fever usually occurs 7 to 10 days into a treatment course, with prompt defervescence within 48 hours of cessation of the responsible agent.

Systemic Manifestations

Drug reactions can result in systemic involvement, such as interstitial nephritis, nephrotic syndrome, hepatic reactions, myocarditis, and vasculitis. Lung involvement may present as part of a syndrome consisting of malaise, nonproductive cough, chest discomfort, and migratory infiltrates, with or without peripheral eosinophilia (Löffler syndrome). Long-term treatment with penicillin, sulfonamides, or phenytoin can result in generalized lymphadenopathy. Aseptic meningitis has been reported from nonsteroidal anti-inflammatory drugs (NSAIDs), radiocontrast media, and other agents.[3-5]

TESTING

Skin Testing

Drug reactions can be identified through immediate-type skin testing only when the process is IgE mediated. Non–IgE-mediated reactions, such as nonurticarial skin rashes, cannot be identified in this manner. Skin testing is reliable for protein agents and small-molecular-weight drugs whose allergenic metabolites have been identified and made available for skin testing (e.g., penicillin). Skin testing requires knowledgeable personnel and the use of appropriate concentrations (i.e., high

enough to provoke a reaction but low enough to avoid causing a systemic response).

In Vitro Testing

There are a limited number of radioallergosorbent tests (RASTs) available for detection of IgE antibodies against drugs, including β-lactam antibiotics and anesthetic agents. In vitro tests for drug allergy are generally less sensitive than skin tests but may be useful in certain cases in which skin testing is not possible (e.g., in patients with severe, generalized eczema or in those who must take medications that can suppress skin-test responses).[6]

Drug Challenges

When the probability of a true allergy is remote, a graded challenge can be used to confirm the clinical diagnosis of drug reaction. In graded challenges, the patient is given a test dose at a dose lower than would cause a serious reaction, followed by escalation of doses in large increments and observation between doses.

Reactions to Specific Drugs

PENICILLIN

Penicillin, a β-lactam antibiotic, is among the most common causes of immunologic drug reactions. Most deaths from penicillin reactions occur in patients with no history of penicillin allergy**.** Nonimmunologic rashes are frequently seen with ampicillin or amoxicillin in patients who have concomitant viral infections, chronic lymphocytic leukemia, or hyperuricemia and in patients taking allopurinol. These rashes are typically nonpruritic and are not associated with an increased risk for future intolerance of penicillin antibiotics.

Most immunologic reactions to penicillins are directed against β-lactam core determinants. Less than 5% of penicillin is metabolized to the minor determinants, which include benzyl penicillin G, penicilloates, and benzylpenicilloylamine. IgE antibodies to the minor determinants are usually responsible for severe immediate-type reactions to penicillin. The benzylpenicilloyl [BPO] moiety, the so-called major determinant, makes up 95% of penicillin metabolites but is less commonly responsible for severe immediate reactions.

Patients who have suffered IgE-mediated penicillin reactions tend to lose their sensitivity over time if penicillin is avoided. By 5 years after an immediate reaction, 50% of patients have negative skin tests. However, there is controversy about whether patients who have lost their sensitivity to penicillin may be more likely than others to be sensitized with subsequent penicillin exposure.[3,7,8] Skin testing with a major determinant preparation and penicillin G identifies at least 90% to 93% of patients at risk for immediate reaction to penicillin. The negative predictive value of penicillin skin testing is significantly increased by the addition of a minor determinant mixture, but that is not currently commercially available. Penicillin skin testing is usually reliable in identifying patients at risk for immediate reactions to the semisynthetic penicillins,[3,9] but reactions to unique semisynthetic side-chain determinants may occur in special-risk populations, such as cystic fibrosis patients.[10] Skin testing with the side-chain determinants is not commonly done.

Not everyone with a history of a reaction to penicillin should be skin tested, but it is important to skin-test patients with a history of anaphylaxis or urticaria associated with penicillin use before they are given penicillin again. Patients who have had maculopapular or morbilliform skin rashes from penicillin are not at a higher risk for immediate-type reactions, but skin testing may be considered because studies have demonstrated that patient histories can be unreliable. Penicillin should not be readministered to patients with a history of penicillin-induced Stevens-Johnson syndrome, toxic epidermal necrolysis (TEN), other exfoliative dermatitis, or bullous skin lesions; therefore, skin testing is not indicated in these cases. Patients with a family history of penicillin allergy but no personal history do not have an increased risk of allergy to penicillin and, therefore, do not require skin testing. Desensitization is not required in a patient with a history of an immediate reaction and a negative skin test, but a small test dose may be given as an additional precaution if the previous reaction was life threatening. If there is a compelling indication for a penicillin antibiotic (e.g., neurosyphilis), a rapid desensitization protocol should be used [*see Table 2*].[4,11-13]

Table 2 Intravenous Desensitization Protocol for β-Lactam Antibiotics[26]

Dose	*β-Lactam Stock Concentration (mg/ml)*	*Cumulative Dose Given (ml)*
1	0.1	0.1
2		0.2
3		0.4
4		0.8
5	1.0	0.15
6		0.30
7		0.60
8		1.00
9	10	0.2
10		0.4
11		0.8
12	100	0.15
13		0.3
14		0.6
15		1.00
16	1000	0.2
17		0.4

Note: observe patient for approximately 15–40 min after each interval dose and for 30 min after final dose.

Cephalosporins and penicillin have similar bicyclic β-lactam structures and amide side chains [*see Figure 2*]. The degree of immunologic cross-reactivity between these β lactams is controversial, but patients with penicillin allergy are more likely than the general population to have a reaction to another β-lactam drug. Most experts agree that cephalosporins should be avoided in patients who have a history of an immediate-type reaction to penicillin. Immunologic reactions to cephalosporins are more

often related to side chains of the cephalosporin antibiotics than to β-lactam core determinants. Cephalosporin skin testing is experimental and has uncertain negative predictive value. There is a lower incidence of immediate-type reactions to third-generation cephalosporins than to the first- and second-generation compounds. Because of its 3-methylthiotetrazole side chain, cefoperazone can cause a nonimmunologic disulfiram-like reaction if taken after ingestion of alcohol. Carbapenems (imipenem) and carbacephems (loracarbef) contain bicyclic β-lactam rings that cause significant cross-reactivity with penicillin. Reactions to monobactams (aztreonam) are typically directed against side-chain determinants rather than the monocyclic β-lactam nucleus, but cross-reactivity occurs with ceftazidime, which shares an identical side chain with penicillin.[13-17]

SULFA

Drug exanthems from sulfonamide antibiotics are more common than immediate-type reactions, and there is significant cross-reactivity among sulfa compounds. Adverse effects from sulfa drugs occur in 2% to 10% of patients who do not have AIDS, whereas 50% of AIDS patients suffer ill effects.[3] The incidence of reactions may be related to the degree of immunodeficiency from HIV infection. The sulfapyridine moiety of sulfasalazine is responsible for most skin rashes from that agent. A 1-month graded challenge protocol is usually successful at inducing tolerance to sulfasalazine in patients who require this agent for treatment of inflammatory bowel disease.[18] Desensitization to sulfonamides may be considered to permit the use of these agents for *Pneumocystis carinii* pneumonia (PCP) prophylaxis in AIDS, toxoplasmosis, and other infections for which there are no good alternatives. A number of protocols have been published. Patients with acute PCP may require rapid desensitization protocols that permit therapeutic use of the medication within 6 to 8 hours[19]; those patients receiving a sulfonamide prophylactically can be desensitized with slowly increasing doses. Desensitization protocols should not be attempted in patients with a history of severe drug reactions, such as Stevens-Johnson syndrome or TEN.

VANCOMYCIN

Vancomycin infusions are commonly associated with the red man syndrome (hypotension, flushing, erythema, pruritus, urticaria, and pain or muscle spasms of the chest and back). The syndrome is caused by non–IgE-mediated histamine release that is more likely with rapid infusion rates (> 10 mg/min). Tolerance of readministration is promoted by reduction of the infusion rate and pretreatment with H_1 (but not H_2) antihistamines. Rarer IgE-mediated reactions to vancomycin can be identified by skin tests.[2-4]

ACE INHIBITORS

Nonimmunologic adverse effects of angiotensin-converting enzyme (ACE) inhibitors are thought to stem from an accumulation of bradykinin and other vasoactive peptides. The most frequently documented adverse reactions include cough (10% to 25%), rhinitis, and angioedema (0.1% to 0.2%). The onset of cough can occur from 1 day to up to 12 months after starting these drugs and may be associated with increased bronchial reactivity to methacholine. Cough usually resolves within several weeks after drug cessation but may persist for more than a month. In about 60% of cases, angioedema occurs within 2 weeks after patients start the drug, but it can occur after

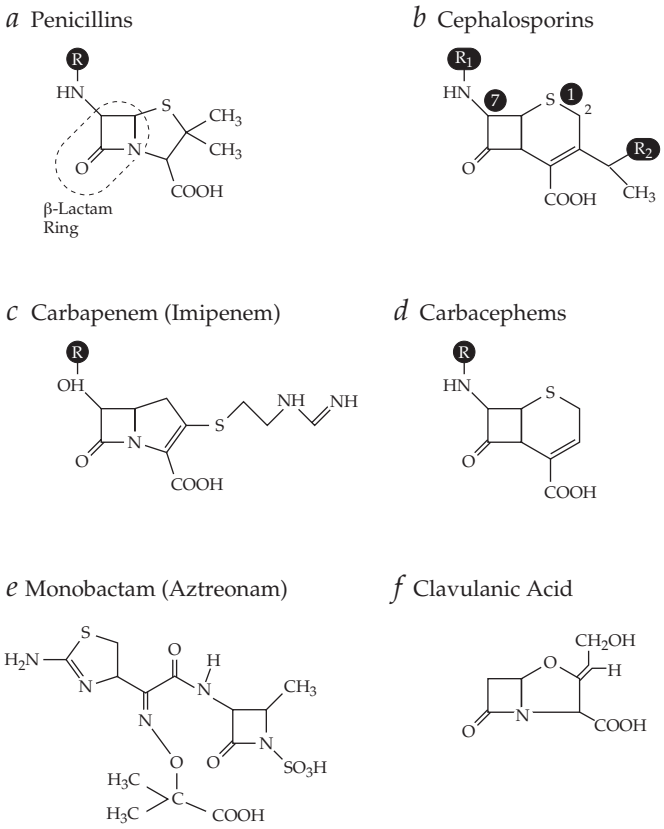


Figure 2 **Structure of β-lactam antibiotics. Substitutions at the R position of 6-aminopenicillanic acid create penicillin derivatives. Substitutions at positions 1, R1, R2, and C7 of 7-aminocephalosporanic acid create cephalosporin derivatives.**

months. Angioedema usually involves the face and oropharyngeal tissue. It can result in life-threatening upper airway obstruction. Visceral angioedema can cause abdominal pain. Cough and angioedema do not usually occur in the same patient. Skin testing is of no value. Intolerance to one ACE inhibitor usually predicts intolerance to all drugs of this class. Patients with idiopathic angioedema and urticaria are susceptible to more severe and frequent episodes when given ACE inhibitors. The angiotensin II receptor blockers are generally well tolerated in patients with idiopathic angioedema and urticaria or with ACE inhibitor–related angioedema.[2,4,20]

ASPIRIN AND NSAIDS

Aspirin and other NSAIDs may induce a variety of reactions, ranging from bronchospasm, rhinorrhea, urticaria, and angioedema to anaphylaxis. Aspirin-sensitive respiratory reactions are likely caused by derangement of arachidonic acid metabolism with increased leukotriene production. Mast cell activation may also occur. Patients with respiratory sensitivity to aspirin generally develop dose-dependent reactions to aspirin or structurally distinct NSAIDs that are significant inhibitors of cyclooxygenase-1 (COX-1) but often tolerate agents that have less effect on COX-1 (e.g., salsalate, acetaminophen, sodium or magnesium salicylate). Selective COX-2 inhibitors (e.g., celecoxib, rofecoxib) are well tolerated in these patients.[21] Patients may present with the so-called aspirin triad of concomitant asthma, nasal polyps, and aspirin sensitivity. Between 30% and 40% of patients with polyps and sinusitis and between 8% and 19% of

adults with asthma have positive bronchial responses to aspirin. Urticarial reactions from aspirin typically occur in a patient subset different from that in which respiratory reactions occur. Some patients with skin reactions to aspirin (often those with a history of idiopathic urticaria and angioedema) have cross-reactivity with NSAIDs, whereas other skin reactors and those who develop anaphylaxis have only specific sensitivity to aspirin or a particular NSAID. However, skin testing is not helpful. With appropriate precautions, oral desensitization with aspirin by experienced clinicians can be performed over several days. It is usually effective in patients with respiratory sensitivity but not in those with skin reactions.[4,5,22,23]

ANESTHETIC AGENTS

Adverse reactions to local anesthetic agents are rarely IgE mediated. More commonly, such reactions are toxic responses from inadvertent intravenous administration, overdose, rapid absorption, or anxiety. Symptoms often involve the central nervous system or the cardiovascular system and include hypotension, convulsions, and cardiorespiratory failure. Concurrent administration of epinephrine may be responsible for shakiness and tachycardia. Allergic contact dermatitis and some large local reactions do occur through delayed-type immunologic responses. Local anesthetics are either benzoid acid esters (type I [e.g., procaine, benzocaine]) or nonesters and amides (type II [e.g., lidocaine, bupivacaine, mepivacaine]). There is no cross-reactivity between the two classes, but type I agents cross-react with each other. Management of suspected local anesthetic allergy includes subcutaneous test dosing with the local anesthetic without epinephrine.

Histamine release has been implicated in some reactions from anesthesia-induction agents and muscle relaxants, but the responsible mechanisms for many reactions are not established. During anesthesia, generalized reactions may be caused by muscle relaxants (e.g., succinylcholine, alcuronium, pancuronium), induction agents (e.g., thiopental), opiates, or antibiotics. Narcotics stimulate mast cells directly without an IgE mechanism.

RADIOGRAPHIC CONTRAST MEDIA

Radiographic contrast media cause non–IgE-mediated anaphylactoid reactions that involve direct mast cell and perhaps complement activation; therefore, immediate-type skin testing and test dosing are not helpful. Shellfish allergy results from IgE-mediated reactions to shellfish proteins, and therefore, it is not predictive of risk for contrast reactions. A previous anaphylactoid reaction to contrast at any time in a patient's history is predictive of persistently increased risk for a repeat anaphylactoid reaction, even though the patient may have tolerated contrast without a reaction in the interim. Asthma and allergies are also associated with increased risk.[24]

The use of nonionic contrast media and medication pretreatment can reduce the risk of reaction. One commonly used pretreatment regimen consists of corticosteroids (prednisone, 50 mg, given 13 hours, 7 hours, and 1 hour before contrast administration), H_1 antihistamines (diphenhydramine, 50 mg orally, 1 hour before administration), and oral adrenergic agents (ephedrine, 25 mg, or albuterol, 4 mg, given orally, 1 hour before administration). H_2 receptor blockers are sometimes added. Despite an adequate pretreatment regimen, reactions can still occur. Corticosteroids administered only 1 to 2 hours before contrast do not reliably prevent reactions.

Management

DESENSITIZATION

If the probability of a drug allergy is high and drug administration is essential, one may consider desensitization, in which the drug is administered in increasing doses in small increments. Because of the risk of adverse reactions, only experienced physicians should perform desensitization. Once desensitization is achieved, the drug must be continued; otherwise, desensitization will be lost, and the patient will require repeat desensitization before readministration. Pretreatment with antihistamines and corticosteroids is not reliable for preventing IgE-mediated anaphylaxis but can be useful when an anaphylactoid reaction is of concern (as with radiographic contrast media). In extreme circumstances, when continued administration of a drug is essential but the patient is experiencing a reaction such as a late drug rash or interstitial nephritis, continued drug administration may be tolerated if corticosteroids and antihistamines are given to suppress the immunologic reaction. However, in such cases there is the risk of progression to an exfoliative skin rash, mucocutaneous disorders (e.g., Stevens-Johnson syndrome, TEN), nephritis, hepatitis, or serum sickness.

MANAGEMENT OF ACUTE REACTIONS

Treatment of acute adverse immunologic drug reactions includes stopping all nonessential suspect drugs and, if necessary, substituting new drugs that should not have cross-reactivity with any of the suspect drugs. Epinephrine, antihistamines, and corticosteroids are the mainstays of treatment of anaphylaxis, and other resuscitative measures may be required [*see 6:XIII Urticaria, Angioedema, and Anaphylaxis*]. Mild maculopapular rashes may respond to antihistamines alone, but progressing rashes or rashes associated with fever, nausea, or arthralgias should also be treated with systemic corticosteroids. For prolonged, severe reactions, several weeks of prednisone may be required.[2,3,13]

Mark S. Dykewicz, M.D., has received grants for clinical research or educational activities or served as an advisor or consultant to Aventis Pharmaceuticals, Inc.; GlaxoSmithKline; Schering-Plough Corporation; and Wallace Pharmaceuticals during the past year. Heather C. Gray, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Gruchalla RS: Drug allergy. J Allergy Clin Immunol 111(2 suppl):548, 2003

2. Dykewicz MS: Drug allergy. Comprehensive Therapy 22:353, 1996

3. DeSwarte RD: Drug allergy. Allergic Diseases: Diagnosis and Management, 5th ed. Patterson R, Ed. Lippincott, Philadelphia, 1993, p 395

4. Dykewicz MS: Drug allergy. An Expert Guide to Allergy and Immunology. Slavin RG, Reisman RE, Eds. American College of Physicians, Philadelphia, 1999, p 127

5. DeShazo RD, Kemp SF: Allergic reactions to drugs and biologic agents. JAMA 278:1895, 1997

6. Primeau MN, Adkinson NF Jr: Recent advances in the diagnosis of drug allergy. Curr Opin Allergy Clin Immunol 1:337, 2001

7. Parker PJ, Parrinello JT, Condemi JJ, et al: Penicillin resensitization among hospitalized patients. J Allergy Clin Immunol 88:213, 1991

8. Solensky R, Earl HS, Gruchalla RS: Lack of penicillin resensitization in patients with a history of penicillin allergy after receiving repeated penicillin courses. Arch Intern Med 162:822, 2002

9. Sogn DD, Evans R III, Shepherd GM, et al: Results of the National Institute of Allergy and Infectious Diseases collaborative clinical trial to test the predictive value of skin testing with major and minor penicillin determinant derivatives in hospitalized patients. Arch Intern Med 152:1025, 1992

10. Ramesh S: Antibiotic hypersensitivity in patients with CF. Clin Rev Allergy Immunol 23:123, 2002

11. Sullivan T, Yecies L, Shatz G, et al: Desensitization of patients allergic to penicillin using orally administered beta-lactam antibiotics. J Allergy Clin Immunol 69:275, 1982

12. Wickern GM, Nish WA, Bitner AS, et al: Allergy to beta-lactams: a survey of current practices. J Allergy Clin Immunol 94:725, 1994

13. Executive summary of disease management of drug hypersensitivity: a practice parameter. Joint Task Force on Practice Parameters, and the Joint Council of Allergy, Asthma and Immunology. Ann Allergy Asthma Immunol 83:665, 1999

14. Kelkar PS, Li JT: Current concepts: cephalosporin allergy. N Engl J Med 345:804, 2001

15. Anne S, Reisman RE: Risk of administering cephalosporin antibiotics to patients with histories of penicillin allergy. Ann Allergy Asthma Immunol 74:167, 1995

16. Adkinson NF Jr: Immunogenicity and cross-allergenicity of aztreonam. Am J Med 88:12S, 1990

17. Saxon A, Adelman DC, Patel A, et al: Imipenem cross-reactivity with penicillin in humans. J Allergy Clin Immunol 82:213, 1988

18. Purdy BH, Philips DM, Summers RW: Desensitization for sulfasalazine skin rash. Ann Intern Med 100:512, 1984

19. Gluckstein D, Ruskin J: Rapid oral desensitization to trimethoprim-sulfamethoxazole (TMP-SMZ): use in prophylaxis for *Pneumocystis carinii* pneumonia in patients with AIDS who were previously intolerant to TMP-SMZ. Clin Infect Dis 20:849, 1995

20. Simon SR, Black HR, Moser M, et al: Cough and ACE inhibitors. Arch Intern Med 152:1698, 1992

21. Pacor ML, Di Lorenzo G, Biasi D, et al: Safety of rofecoxib in subjects with a history of adverse cutaneous reactions to aspirin and/or non-steroidal anti-inflammatory drugs. Clin Exp Allergy 32:397, 2002

22. Stevenson DD, Hankammer MA, Mathison DA, et al: Aspirin desensitization treatment of aspirin-sensitive patients with rhinosinusitis-asthma: long-term outcomes. J Allergy Clin Immunol 98:751, 1996

23. Namazy JA, Simon RA: Sensitivity to nonsteroidal anti-inflammatory drugs. Ann Allergy Asthma Immunol 89:542, 2002

24. Maddox TG: Adverse reactions to contrast material: recognition, prevention, and treatment. Am Fam Physician 66:1229, 2002

25. Grammer LC: Drug allergy. Atlas of Allergies. Slavin RG, Fireman P, Eds. Mosby-Wolfe Publishers, Barcelona, Spain, 1996, p 286

26. Sullivan TJ: Drug allergy. Allergy: Principles and Practice, 4th ed. Middleton E, Reed CE, Ellis EF, et al, Eds. CV Mosby, St Louis, 1993, p 1726

Acknowledgment

Figures 1 and 2 Seward Hung.

XV ALLERGIC REACTIONS TO HYMENOPTERA

David B. K. Golden, m.d., f.a.c.p.

Allergic reactions to insect venom primarily occur as a result of stings by insects of the order Hymenoptera. Allergic swelling can occur at the site of the insect sting, but only rarely does anaphylaxis result. Nonallergic reactions to insect venom have also been reported; these include nephropathy, central and peripheral neurologic syndromes, idiopathic thrombocytopenic purpura, and rhabdomyolysis. These are toxic reactions and are not IgE mediated. Allergic reactions to stings manifest themselves as either late-phase local inflammation (i.e., severe prolonged swelling) or systemic responses (e.g., anaphylaxis).

Epidemiology

Allergic reactions to Hymenoptera stings have been reported in persons of all ages. The reactions may be preceded by a number of uneventful stings. Systemic allergic reactions are reported in up to 1% of children and 3% of adults, although allergic antibodies to Hymenoptera venoms can be detected in 17% to 26% of adults.[1] The frequency of large local allergic reactions is uncertain but is estimated to be about 10% in adults. Fatal allergic reactions to insect stings may occur at any age but are most common in adults older than 45 years.[2] Half of those persons in whom a fatal reaction occurred had no previous history of allergy to insect stings; the other half had previous reactions but failed to take adequate preventive measures. In the United States, at least 40 deaths occur each year as the result of insect stings; other sting fatalities may go unrecognized. In many cases of unexplained sudden death, postmortem blood samples show the presence of both Hymenoptera venom–specific IgE antibodies and elevated serum tryptase levels, indicating the true cause of the fatal reactions.[3]

For 50 years, whole body extracts were used as standard treatment for immunotherapy; such use was based on a lack of knowledge of the natural history of anaphylaxis.[4] We now recognize that the risk of an anaphylactic reaction to a sting varies in accordance with the history of previous stings and is correlated with the results of venom skin testing or radioallergosorbent testing (RAST). The risk declines gradually with time [*see Table 1*]. In high-risk patients, the risk of reaction is 50% to 70%; other persons with a history of insect-sting allergy are at much lower risk. Most affected children have only cutaneous systemic reactions, with no respiratory or vascular symptoms, and have less than a 10% risk of a systemic reaction to a subsequent sting.[5] The risk is also less than 10% in adults or children who have experienced only large local reactions to stings. Furthermore, the allergy is self-limited in many cases. The risk of reaction falls from 50% initially to 33% after 3 to 5 years; the risk is 20% to 25% if more than 10 years have passed since the reaction.[6] However, in some individuals, the risk of anaphylaxis persists for decades, even with no intervening stings.

Etiology

Hymenoptera allergy is directed against the allergenic proteins in the venoms of the stinging insects. Three families of Hymenoptera are important causes of allergy. The bees (i.e., honeybees and bumblebees) and vespids (i.e., yellow jackets, hornets, and wasps) are the best known [*see Figures 1 and 2*]. Imported fire ants (*Solenopsis* species) are a rapidly increasing public health hazard in the Southeast and South Central United States, especially on the Gulf Coast [*see Figure 3*].[7] Honeybee stings are more common in agricultural areas. Yellow jackets are the most frequent culprits in the northern areas of North America and Europe, whereas paper wasps (*Polistes* species) are more commonly implicated along the Gulf Coast in the United States and the Mediterranean Coast in Europe.

Knowledge of the behavior of these insects can be helpful in evaluating the history of affected patients. Honeybees are relatively docile and rarely sting or swarm unless provoked. Stings usually occur as a result of garden exposures or from going barefoot outdoors. The barbed stinger of the honeybee remains in the skin, causing the death of the honeybee. Africanized honeybees (killer bees) are more aggressive and are now present in the southern United States.[8] Although an Africanized honeybee is no different from a domestic honeybee with regard to anatomy or venom, Africanized honeybees have a tendency to swarm with little provocation and to sting in large numbers. A large number of stings can cause massive envenomation; the resulting toxic reactions have been fatal to livestock and humans. Bumblebees sting infrequently, but a few cases of systemic reactions have been reported.

Yellow jackets usually nest underground or in the cracks of buildings or wooden ties or logs used in residential landscaping, whereas hornets generally build their nests in shrubs and trees. Paper wasps build an open nest with visible cells; nests are often found on the eaves or windowsills of a home or in the railings of decks or fences. Yellow jackets are scavengers; they are commonly found around food at picnics and in orchards, trash cans, and dumpsters. They are highly aggressive and will sting quite readily. Wasps are less aggressive but will sting readily when disturbed. The vespid stinger usually has finer barbs than the stinger of the bee and does not commonly remain in the skin.

Table 1 Risk of Systemic Reactions and Clinical Recommendations Based on Reaction to Previous Stings and Results of the Venom Skin Test or RAST

Reaction to Previous Sting	*Skin Test or RAST*	*Risk of Systemic Reaction*	*Clinical Recommendation*
None	Positive	10%–20%	Avoidance
Large local	Positive	5%–10%	Avoidance
Cutaneous systemic	Positive, child Positive, adult	1%–10% 10%–20%	Avoidance Venom immunotherapy
Anaphylaxis	Positive Negative	30%–70% 5%–10%	Venom immunotherapy Repeat skin test/RAST

RAST—radioallergosorbent testing

Figure 1 **The honeybee (*Apis mellifera*).**

Figure 2 **The European hornet (*Vespa crabro germana*) was introduced into the United States in the mid-19th century. In the United States, its habitat includes most of the eastern United States, Louisiana, and the Dakotas. Although it is a woodland species, its nests can be found in barns, attics, hollow walls, birdhouses, and abandoned beehives.**

Fire ants have stingers, and it is the sting rather than the bite that causes the allergic reaction. Fire ants are widespread in the southeastern United States; in many areas, stings are very frequent, with up to 50% of the population being stung each year. Fire ants build nests in the shape of large mounds; these nests are common in residential and coastal areas. In most cases of fire-ant stings, multiple ants each administer multiple stings, which cause minimal pain. The unique lesions form sterile pustules that can become infected if excoriated [*see Figure 4*].

The allergic sensitivity is directed against proteins in the venoms (but not in the saliva or bodies) of the stinging insects.[9] Honeybee venom contains unique allergens, whereas the vespid venoms cross-react extensively with one another and contain essentially the same allergens. The venom of *Polistes* wasps is less cross-reactive than that of the other vespids. Only 50% of patients who are allergic to yellow-jacket venom experience reactions to wasp venom. The allergenic proteins in fire-ant venom are unique.

Pathogenesis

The pathogenesis of Hymenoptera allergy is the same as that of other forms of anaphylaxis. An initial encounter with a sting in genetically susceptible individuals causes the production of IgE antibodies to the venom allergens. The IgE antibodies become affixed to tissue mast cells and circulating basophils, which thus become armed for response to a later encounter with the same allergen. A subsequent sting can cause cross-linking of these allergic antibodies, leading to the release of mediators (e.g., histamine, leukotrienes, and cytokines) that cause the clinical manifestations of the allergic reaction. There is an association between conditions involving abnormal mast cell number or function and insect-sting anaphylaxis. The allergic reactions to stings are more severe in patients with elevated baseline serum tryptase levels or mastocytosis, and there is a higher incidence of treatment failures and relapse after treatment in these patients.[10-12] Whole body extracts of fire ants are used for diagnostic testing and are effective for immunotherapy, whereas whole body extracts of the other Hymenoptera insects have proved not to contain venom allergens.

Diagnosis

CLINICAL MANIFESTATIONS

Allergic reactions to insect stings may cause local allergic inflammation or the full spectrum of manifestations of anaphylaxis. Large local reactions are late-phase allergic reactions. Progressive swelling begins 6 to 12 hours after the sting, reaching peak size in 24 to 48 hours and resolving in 5 to 10 days. Large local reactions are usually defined as being greater than 6 in. in diameter; they can be massive in size and cause considerable pain. On the extremities, inflammatory lymphangitic streaks occur toward the axillary or inguinal nodes; these streaks are mistaken for signs of infection. Systemic reactions most commonly cause cutaneous signs and symptoms, including generalized flushing, pruritus, urticaria, and angioedema. Other typical manifestations are respiratory (e.g., throat or chest tightness, dyspnea, wheezing) or circulatory (e.g., light-headedness or unconsciousness). Less common signs of anaphylaxis include gastrointesti-

Figure 3 **Red imported fire ant (*Solenopsis invicta*).**

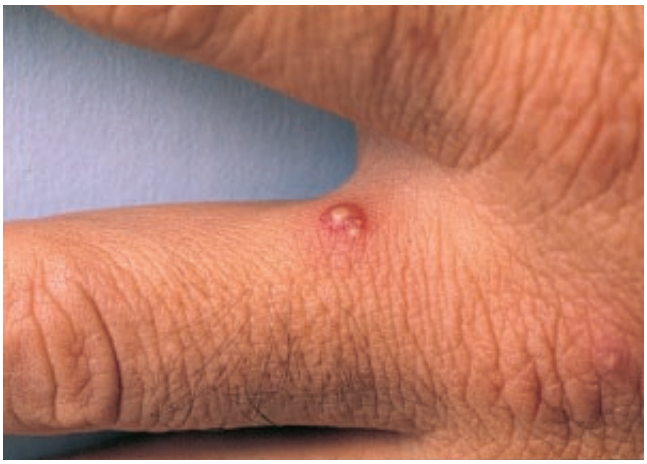


Figure 4 **Appearance of a pustule resulting from the sting of a fire ant. This photograph was taken 24 hours after the patient was stung.**

nal or uterine cramps, cardiac arrhythmias (e.g., tachycardia or, occasionally, bradycardia), and coronary vasospasm.

In children, systemic reactions to stings usually cause only cutaneous symptoms (e.g., urticaria or angioedema). Respiratory symptoms are less common, and circulatory manifestations are infrequent. Systemic reactions usually follow a predictable and individual pattern in each patient, with worsening of the reaction occurring in less than 10% of cases.[13] Affected individuals commonly do not seek medical attention and usually fail to report having sting reactions unless they are asked.

The diagnosis of insect-sting allergy rests on a history of allergic reactivity, because venom-specific IgE antibodies can be detected in many normal individuals. The positive venom skin test provides confirmation of the allergic nature of the sting reaction and helps define allergenic specificity. The history is most important and should be reviewed in detail with respect to the nature, number, and timing of stings in the past; the time course of the reaction; and all associated symptoms and treatments. The family history, atopic history, and general medical history are also of interest. In addition, it is helpful to know of any medications the patient took before the reaction occurred, as well as any medications the patient is currently using.

PHYSICAL EXAMINATION

It is most important to measure the vital signs, including airflow, when there is dyspnea and to document cutaneous signs. Some patients have symptoms, such as dizziness and dyspnea, that do not correspond to the objective signs (e.g., blood pressure, peak expiratory flow rate) and may be the result of anxiety, panic, and hyperventilation. Any history suggestive of systemic allergic reaction must be taken seriously.

DIAGNOSTIC TESTS

The diagnosis of insect-venom allergy can be confirmed by skin tests or serologic tests using Hymenoptera venoms. Both methods are useful, and they are often complementary in the diagnostic evaluation of affected patients. Both methods require specific experience and training to prevent false interpretations. Diagnostic tests are not usually performed in the absence of a history of a systemic allergic reaction. This is because a positive test occurs in 20% to 30% of adults and is associated with a relatively low 17% chance of a systemic reaction to a future sting.[1]

Intradermal skin tests using serial dilutions of the five Hymenoptera venom protein extracts is the recommended procedure. In the case of fire-ant sensitivity, whole body extracts of imported fire ants give reasonable diagnostic sensitivity and specificity. For Hymenoptera venom testing, intradermal tests are performed with venom concentrations ranging from 0.001 to 1.0 μg/ml to find the minimal concentration that yields a positive result, as compared with a negative diluent control (e.g., human serum albumin saline) and a positive histamine control. Puncture tests with a venom concentration of 0.01 μg/ml may be used initially for patients with a history of very severe reactions.

The diagnosis of insect-sting allergy by detection of allergen-specific IgE antibodies in serum (typically by RAST) is a method of high potential but variable performance.[14] An elevation in the level of venom-specific IgE is certainly diagnostic; but the test is often qualitative and poorly standardized, and it yields negative results in 15% to 20% of patients whose skin-test results are positive.

In the majority of patients who have a definite history of insect-sting reactions, skin-test results are clearly positive; however, in many others, the results are clearly negative. Negative skin-test results in a patient with a history of insect-sting reactions may represent the loss of sensitivity, but it is important to test for venom-specific IgE antibodies in the serum (e.g., by use of RAST).[15] If necessary, the venom skin test may be repeated after several months. A few cases of sting anaphylaxis are non–IgE mediated and may be related to subclinical mastocytosis or simply toxic mast cell hyperreleasability.[12] It is important to note that the degree of sensitivity as detected by skin testing or RAST does not correlate reliably with the degree of sting reaction. The strongest sensitivity to skin tests often occurs in patients who have had only large local reactions, and some patients who have had near-fatal anaphylactic shock show only weak sensitivity on skin testing or RAST. Because of cross-reactivity, skin tests are positive to all three of the common vespid venoms (i.e., yellow jacket, yellow hornet, and white-faced hornet) in 95% of patients allergic to yellow-jacket venom. More than half of patients sensitive to yellow-jacket venom also have positive reactions to testing for sensitivity to *Polistes* wasp venom. It is possible to determine whether the patient has a specific or a cross-reactive sensitivity to wasp venom using a RAST-inhibition test in specialized laboratories.[16]

Differential Diagnosis

Although the diagnosis of insect-sting allergy is relatively straightforward, the history and diagnostic tests can be misleading in some cases. Local swelling may be the result of nonallergic inflammation, but infection is very uncommon and would likely occur many days after the sting. Local cutaneous signs should not be mistaken for systemic eruption. Symptoms of dyspnea, chest discomfort, and dizziness can be the result of hyperventilation associated with anxiety. Patients with asthma who receive a sting may have asthmatic symptoms that are difficult to distinguish from an allergic reaction. Approximately 1% of patients with a history of allergic reactions to insect stings have an underlying abnormality in the release of mediators by mast cells or basophils, as demonstrated by elevated baseline serum tryptase levels. Some of these patients have a form of mastocytosis.[10-12]

Treatment

ACUTE TREATMENT

The treatment of the acute systemic allergic reaction to insect stings is the same as that of other causes of anaphylaxis. The treatment of choice is epinephrine by intramuscular injection.[17,18] The recommended dose is 0.3 to 0.5 mg (0.3 to 0.5 ml of a solution of 1:1,000 weight in volume [w/v]) for adults and 0.01 mg/kg for children. Delay in the use of epinephrine has contributed to fatal reactions. Some persons in anaphylactic shock are resistant to epinephrine. Patients taking beta-blocker medications can also be resistant to the effects of epinephrine. In some cases, anaphylaxis is prolonged or recurrent (biphasic) for 6 to 24 hours and may require intensive medical care.[19] All patients with anaphylaxis should receive full emergency medical attention and remain under observation for 6 hours or longer. Corticosteroids have no role in the treatment of acute anaphylaxis; they are administered to prevent late-phase reactions, but there is no evidence that steroids prevent biphasic anaphylaxis.

Large local reactions may require a burst of corticosteroid, which is most effective if started within 2 hours of the sting. After an initial dose of 40 to 60 mg, the dose is tapered over 3 to 5 days.

PREVENTIVE TREATMENT

General Measures

Patients who are discharged from emergency care after suffering anaphylaxis must receive information on the risk of future reactions. They should also be advised to receive allergy consultation, and they should be given information about prevention. When outdoors, the affected person should avoid bushes and gardens, as well as food and drink that are most likely to attract insects. Drinks, especially in cans, bottles, and straws, can be an unsuspected source of a sting to the tongue or throat. Prescription of an epinephrine autoinjector (e.g., EpiPen and EpiPen Jr., Dey, Napa, California) should be considered in any patient who has experienced a systemic allergic reaction. Some patients may need to use epinephrine immediately after receiving a sting (until they can be immunized); however, most patients can wait for the signs of a developing reaction before using epinephrine. Delay in treatment is reasonable because the majority of persons with a history of mild to moderate systemic reactions do not react to a challenge sting. The age at which a patient should be prescribed an adult-strength autoinjector, rather than a pediatric autoinjector, is uncertain; use of the adult-strength injector may be considered when the child attains a weight of 25 kg.[20] All patients should understand that use of an epinephrine kit is not a substitute for emergency medical attention.

Venom Immunotherapy

Patient selection Current indications for venom immunotherapy are a history of previous systemic allergic reaction to a sting and a positive venom skin test.[21-23] The patients at highest risk are those with a recent history of anaphylaxis and positive skin-test results; in such patients, the risk of a systemic reaction to a subsequent sting is approximately 50%. Children and adults with a history of large local reactions are at low risk for a systemic reaction (i.e., < 10%),[24] as are children whose systemic reactions are limited to cutaneous signs and symptoms.[5] In these low-risk persons, venom immunotherapy is not required, but some patients will still request treatment for reasons related to fear of reaction or frequent exposure. Children with moderate or severe systemic reactions have a relatively high risk of recurrence even 10 to 20 years after allergic reaction to an insect sting.[25] There are some cases of progressively worsening reactions in adults, so all adults with systemic reactions are advised to undergo venom immunotherapy. There is no test that accurately predicts which patients will progress to more severe reactions and which will not.

Initial therapy Initial venom immunotherapy can be completed with a regimen of eight weekly injections or a traditional regimen lasting for 4 to 6 months.[21-23] Rush immunotherapy, typically administered over a period of 2 to 3 days, has been reported to be as safe and effective as the usual regimens.[26,27] The recommended maintenance dose is 100 μg of each of the venoms for which a positive result was seen on skin testing. Standard therapy is 85% to 98% effective in completely preventing systemic allergic reactions, but some patients require higher doses for full protection.[28] The same dose has been recommended for children 3 years of age and older, even though their immune response to venom immunotherapy is twice that of adults.

Adverse reactions Venom immunotherapy causes reactions no more frequently than inhalant allergen immunotherapy.[29] Systemic symptoms occur in 10% to 15% of patients during the initial weeks of treatment, regardless of the regimen used. Most reactions are mild, and fewer than half of the reactions require epinephrine. In unusual cases, there can be repeated problems with systemic reactions to injections. Large local reactions are common but are not predictive of systemic reactions to subsequent injections. All patients must achieve the full 100 μg dose to have optimal clinical protection.

Maintenance and monitoring After reaching the full dose, the same dose is repeated at 4-week intervals for at least 1 year. The dosing interval may then be increased to once every 6 to 8 weeks over several years of treatment. Therapeutic efficacy can be confirmed serologically, but use of only some assays for venom-specific IgG antibodies has correlated strongly with clinical protection.[30] Venom skin testing or RAST is repeated periodically—usually every 2 to 3 years—to determine whether there has been a significant decline in venom-specific IgE.[31] The results of skin testing generally remain unchanged during the first 2 to 3 years but show a significant decline after 4 to 6 years.[6] Fewer than 20% of patients have negative skin-test results after 5 years, but 50% to 60% have negative results after 7 to 10 years.[32]

Duration The package inserts for the commercial venom immunotherapy products available in the United States recommend indefinite immunotherapy. However, the published practice parameters reflect more recent experience and recommendations.[23,31] In most patients, skin-test results and RAST results remain positive after 5 to 10 years of treatment. Studies of several hundred adults and children show that even when skin tests remain positive, venom immunotherapy can usually be stopped after 5 years.[31] Observation of patients for 5 to 10 years after completing a 5- to 8-year course of venom treatment has shown a 5% to 10% risk of systemic symptoms after any sting but only a 2% risk of a reaction requiring epinephrine treatment.[33] Patients who have a higher frequency of relapse include those receiving honeybee-venom therapy, those with a history of very severe

pretreatment sting reactions, and those who have had a systemic reaction to a sting or an injection during the period of venom immunotherapy.[22] Several studies have shown that 5 years of therapy is superior to 3 years for suppression of the IgE response and for longer-lasting remission.[34,35] Some patients prefer to continue venom treatment for their continued sense of security. Children who have had a 3- to 5-year course of venom immunotherapy show persistent tolerance even 10 to 20 years after discontinuing treatment.[25]

David B. K. Golden, M.D., F.A.C.P., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Golden DK, Marsh DG, Kagey-Sobotka A, et al: Epidemiology of insect venom sensitivity. JAMA 262:240, 1989

2. Barnard JH: Studies of 400 Hymenoptera sting deaths in the United States. J Allergy Clin Immunol 52:259, 1973

3. Schwartz HJ, Sutheimer C, Gauerke B, et al: Venom-specific IgE antibodies in postmortem sera from victims of sudden unexpected death. J Allergy Clin Immunol 73:189, 1984

4. Golden DB, Langlois J, Valentine MD, et al: Treatment failures with whole-body extract therapy of insect sting allergy. JAMA 246:2460, 1981

5. Valentine MD, Schuberth KC, Kagey-Sobotka A, et al: The value of immunotherapy with venom in children with allergy to insect stings. N Engl J Med 323:1601, 1990

6. Reisman RE: Natural history of insect sting allergy: relationship of severity of symptoms of initial sting anaphylaxis to re-sting reactions. J Allergy Clin Immunol 90:335, 1992

7. Stafford CT: Hypersensitivity to fire ant venom. Ann Allergy Asthma Immunol 77:87, 1996

8. Schumacher MJ, Egen NB: Significance of Africanized bees for public health: a review. Arch Intern Med 155:2038, 1995

9. King TP, Spangfort MD: Structure and biology of stinging insect venom allergens. Int Arch Allergy Immunol 123:99, 2000

10. Ludolph-Hauser D, Rueff F, Fries C, et al: Constitutively raised serum concentrations of mast-cell tryptase and severe anaphylactic reactions to Hymenoptera stings. Lancet 357:361, 2001

11. Fricker M, Helbling A, Schwartz L, et al: Hymenoptera sting anaphylaxis and urticaria pigmentosa: clinical findings and results of venom immunotherapy in ten patients. J Allergy Clin Immunol 100:11, 1997

12. Oude Elberink J, de Monchy J, Kors JW, et al: Fatal anaphylaxis after a yellow jacket sting, despite venom immunotherapy, in two patients with mastocytosis. J Allergy Clin Immunol 99:153, 1997

13. van der Linden PG, Hack CE, Struyvenberg A, et al: Insect-sting challenge in 324 subjects with a previous anaphylactic reaction: current criteria for insect-venom hypersensitivity do not predict the occurrence and the severity of anaphylaxis. J Allergy Clin Immunol 94:151, 1994

14. Hamilton RG: Responsibility for quality IgE antibody results rests ultimately with the referring physician. Ann Allergy Asthma Immunol 86:353, 2001

15. Golden DB, Kagey-Sobotka A, Norman PS, et al: Insect sting allergy with negative venom skin test responses. J Allergy Clin Immunol 107:897, 2001

16. Hamilton RH, Wisenauer JA, Golden DB, et al: Selection of Hymenoptera venoms for immunotherapy on the basis of patient's IgE antibody cross-reactivity. J Allergy Clin Immunol 92:651, 1993

17. Simons FE, Gu X, Simons KJ: Epinephrine absorption in adults: intramuscular versus subcutaneous injection. J Allergy Clin Immunol 108:871, 2001

18. Nicklas RA, Bernstein IL, Li JT, et al: The diagnosis and management of anaphylaxis. Joint Task Force on Practice Parameters. J Allergy Clin Immunol 101:S465, 1998

19. Lockey RF, Turkeltaub PC, Baird-Warren IA, et al: The Hymenoptera venom study I, 1979–1982: demographic and history-sting data. J Allergy Clin Immunol 82:370, 1988

20. Simons FE, Peterson S, Black CD: Epinephrine dispensing for the out-of-hospital treatment of anaphylaxis in infants and children: a population-based study. Ann Allergy Asthma Immunol 86:622, 2001

21. Golden DB, Schwartz HJ: Guidelines for venom immunotherapy. J Allergy Clin Immunol 77:727, 1986

22. Muller U, Mosbech H: Position paper: immunotherapy with Hymenoptera venoms. Allergy 48:37, 1993

23. Moffitt JE, Golden DB, Reisman RE, et al: Stinging insect hypersensitivity: a practice parameter update. J Allergy Clin Immunol 114:869, 2004

24. Mauriello PM, Barde SH, Georgitis JW, et al: Natural history of large local reactions from stinging insects. J Allergy Clin Immunol 74:494, 1984

25. Golden DBK, Kagey-Sobotka A, Norman PS, et al: Outcomes of allergy to insect stings in children with and without venom immunotherapy. N Engl J Med 351:668, 2004

26. Sturm G, Kranke B, Rudolph C, et al: Rush Hymenoptera venom immunotherapy: a safe and practical protocol for high-risk patients. J Allergy Clin Immunol 110:928, 2002

27. Tankersley MS, Walker RL, Butler WK, et al: Safety and efficacy of an imported fire ant rush immunotherapy protocol with and without prophylactic treatment. J Allergy Clin Immunol 109:556, 2002

28. Rueff F, Wenderoth A, Przybilla B: Patients still reacting to a sting challenge while receiving conventional Hymenoptera venom immunotherapy are protected by increased venom doses. J Allergy Clin Immunol 108:1027, 2001

29. Lockey RF, Turkeltaub PC, Olive ES, et al: The Hymenoptera venom study: III. Safety of venom immunotherapy. J Allergy Clin Immunol 86:775, 1990

30. Golden DB, Lawrence ID, Hamilton RH, et al: Clinical correlation of the venom-specific IgG antibody level during maintenance venom immunotherapy. J Allergy Clin Immunol 90:386, 1992

31. Graft DF, Golden DK, Resiman RE, et al: The discontinuation of Hymenoptera venom immunotherapy. Report from the Committee on Insects. J Allergy Clin Immunol 101:573, 1998

32. Golden DB, Kwiterovich KA, Kagey-Sobotka A, et al: Discontinuing venom immunotherapy: extended observations. J Allergy Clin Immunol 101:298, 1998

33. Golden DB, Kagey-Sobotka A, Lichtenstein LM: Survey of patients after discontinuing venom immunotherapy. J Allergy Clin Immunol 105:385, 2000

34. Keating MU, Kagey-Sobotka A, Hamilton RG, et al: Clinical and immunologic follow-up of patients who stop venom immunotherapy. J Allergy Clin Immunol 88:339, 1991

35. Lerch E, Muller U: Long-term protection after stopping venom immunotherapy: results of re-stings in 200 patients. J Allergy Clin Immunol 101:606, 1998

Acknowledgments

The photographs in Figures 1 and 2 are by Stephen B. Bambara; the photographs in Figures 3 and 4 are by James Baker, Ph.D. All photographs are courtesy of the North Carolina State University Department of Entomology.

XVI FOOD ALLERGIES

A. Wesley Burks, m.d.

Definition

Food hypersensitivity (allergy) and food intolerance constitute the category of adverse reactions to food.[1,2] An adverse food reaction is a clinically abnormal response to an ingested food or food additive. Both food hypersensitivity and food intolerance have often been overdiagnosed, and both terms have been applied incorrectly to all adverse reactions to foods.

food intolerance

Food intolerance is a general term describing an abnormal physiologic response to an ingested food or food additive. Such reactions are apparently nonimmunologic in nature and have many possible causes, including toxic contaminants (e.g., histamine in scombroid fish poisoning or toxins secreted by *Salmonella*, *Shigella*, and *Campylobacter*), pharmacologic properties of the food (e.g., caffeine in coffee, tyramine in aged cheeses), characteristics of the host such as metabolic traits (e.g., lactase deficiency), and idiosyncratic responses.

food hypersensitivity

Food hypersensitivity is an immunologic reaction resulting from the ingestion of a food or food additive. This reaction can develop after ingestion of a small amount of the substance, and it is unrelated to any physiologic effect of the food or food additive. To most physicians, the term is synonymous with reactions that involve the immunoglobulin E (IgE) mechanism, of which anaphylaxis is the classic example. Although IgE-mediated (type I) hypersensitivity accounts for the majority of well-characterized allergic reactions to food, non–IgE-mediated immune mechanisms are believed to be responsible for a variety of hypersensitivity disorders. This chapter examines adverse food reactions that are IgE mediated, those that are non–IgE mediated, and those that have characteristics of both.

Epidemiology

The true incidence of adverse food reactions is still unknown. Up to 15% of the general population believe that they may be allergic to some food. However, the best available studies suggest that the actual prevalence of food allergy is 1.5% to 2% of the adult population.[3] The incidence of adverse food reactions in young children is estimated to be between 6% and 8%. Several well-controlled studies have shown that the vast majority of allergic reactions occur in the first year of life.

Pathophysiology

IgE-mediated reactions

A variety of hypersensitivity responses to an ingested food antigen may result from the lack of development of oral tolerance or a breakdown of oral tolerance in the gastrointestinal tract of a person who is genetically predisposed to such hypersensitivity. Either a failure to develop oral tolerance or a breakdown in oral tolerance results in excessive production of food-specific IgE antibodies. These food-specific antibodies bind high-affinity FcεI receptors on mast cells and basophils and low-affinity FcεII receptors on macrophages, monocytes, lymphocytes, eosinophils, and platelets.[2] After antigen-presenting cells process the food allergen and present the antigen to specific antibodies on mast cells or basophils, those cells release mediators such as histamine, prostaglandins, and leukotrienes [*see 6:X Allergic Response*]. These mediators promote vasodilation, smooth muscle contraction, and mucus secretion, resulting in the symptoms of immediate hypersensitivity. The activated mast cells also may release various cytokines that play a part in the IgE-mediated late-phase response. With repeated ingestion of a specific food allergen, mononuclear cells are stimulated to secrete histamine-releasing factors (HRF). The so-called spontaneous generation of HRF by activated mononuclear cells in vitro has been associated with increased cutaneous irritability in children with atopic dermatitis. A rise in plasma histamine levels has been associated with IgE-mediated allergic symptoms after blinded food challenges. In IgE-mediated GI reactions, endoscopic observation has revealed local vasodilation, edema, mucus secretion, and petechial hemorrhaging. Increased stool and serum levels of prostaglandin E_2 (PGE_2) and prostaglandin F_2 (PGF_2) have been seen after food challenges that cause diarrhea.

non–IgE-mediated reactions

Although a variety of non–IgE-mediated immune mechanisms for allergic reactions to food have been proposed, the scientific evidence supporting these mechanisms is limited. Type III (antigen-antibody complex–mediated) hypersensitivity reactions have been examined in several studies. Whereas IgE–food antigen complexes are seen more commonly in patients with food hypersensitivity, there is little support for disease mediated by food antigen–immune complexes. Type IV (cell-mediated) hypersensitivity has been suggested as the mechanism for several disorders in which the clinical symptoms do not appear until several hours after the ingestion of the suspected food. This type of immune response may contribute to some adverse food reactions (e.g., enterocolitis), but except in gluten-sensitive enteropathy, significant supporting evidence of a specific cell-mediated hypersensitivity disorder is lacking.

Diagnosis

IgE-mediated hypersensitivity

Gastrointestinal Reactions

Food-induced, IgE-mediated GI allergy may manifest as a variety of syndromes, including the oral allergy syndrome, immediate GI hypersensitivity, and a small subgroup of cases of allergic eosinophilic gastroenteritis.[4]

Oral allergy syndrome The oral allergy syndrome is considered a form of contact urticaria that is confined almost exclusively to the oropharynx and rarely involves other target organs [*see Table 1*]. The symptoms include rapid onset of pruritus and angioedema of the lips, tongue, palate, and throat. The symptoms generally resolve quite rapidly. This syndrome is most commonly associated with the ingestion of fresh fruits and vegetables. Interestingly, patients with allergic rhinitis as-

Table 1 Oral Allergy Syndrome[27]

Oral manifestations	Burning Swelling Itching Erythema Immediate onset of symptoms
Age of onset	Beyond infancy Typically younger than 5 yr
Proteins implicated	Heat-labile fresh fruit and vegetable allergens Pollen and latex cross-reactivity
Pathology	IgE antibodies
Treatment	Avoidance Cooking
Natural history	Unknown

sociated with certain airborne pollens (especially ragweed and birch pollens) are frequently afflicted with this syndrome. Patients with ragweed allergy may experience these symptoms after contact with certain melons (e.g., watermelons, cantaloupe, honeydew) and bananas. Patients with birch sensitivity often have symptoms after the ingestion of raw potatoes, carrots, celery, apples, and hazelnuts. The diagnosis of this syndrome is based on a suggestive history and positive prick skin tests with the implicated fresh fruits or vegetables, although the sensitivity of these tests may be limited in this disorder [*see* Prick Skin Tests, *below*].[5]

Immediate GI hypersensitivity Immediate GI hypersensitivity may accompany allergic manifestations in other target organs [*see Table 2*].[6,7] The GI symptoms vary but may include nausea, abdominal pain, abdominal cramping, vomiting, and diarrhea. In children with atopic dermatitis and food allergy, the frequent ingestion of a food allergen appears to induce partial desensitization of GI mast cells, resulting in less pronounced symptoms.

The diagnosis of immediate GI sensitivity is based on a suggestive clinical history, positive prick skin tests, resolution of symptoms after complete elimination of the suspected food allergen for up to 2 weeks, and positive results on oral food challenges. After patients have avoided a particular food for 10 to 14 days, it is not unusual for them to experience vomiting during a challenge, even though they were previously able to eat the food without vomiting.

Respiratory and Skin Reactions

Respiratory and ocular symptoms are common manifestations of IgE-mediated reactions to foods.[2,7] Symptoms may include periocular erythema, pruritus, and tearing; nasal congestion, pruritus, sneezing, and rhinorrhea; and coughing, voice changes, and wheezing. Isolated naso-ocular symptoms are an uncommon manifestation of food hypersensitivity reactions.

The skin is a frequent target organ in IgE-mediated food hypersensitivity reactions. The ingestion of food allergens can either trigger immediate cutaneous symptoms or aggravate chronic cutaneous symptoms. Acute urticaria and angioedema are probably the most common cutaneous manifestations of food hypersensitivity, generally appearing within minutes after ingestion of the food allergen. The foods that commonly cause these reactions in children include eggs, milk, peanuts, and tree nuts. In adults, this list includes fish, shellfish, tree nuts, and peanuts.

Atopic dermatitis is a chronic skin disorder that generally begins in early infancy and is characterized by typical distribution, extreme pruritus, and a chronically relapsing course; it is associated with asthma and allergic rhinitis [*see 2:IV Eczematous Disorders, Atopic Dermatitis, and Ichthyoses*].[8] Approximately one third of young children with atopic dermatitis have one or more food allergies. The immediate allergic reaction to foods in these patients is IgE mediated. However, their skin disease may be worsened by a mixed IgE-mediated and non–IgE-mediated reaction. Most young children with atopic dermatitis who have food allergy have reactions to milk, eggs, or peanuts.

MIXED IgE-MEDIATED AND NON–IgE-MEDIATED HYPERSENSITIVITY

Allergic Eosinophilic Gastroenteropathy

Allergic eosinophilic gastroenteropathy is a disorder characterized by infiltration of the gastric or intestinal walls with eosinophils; absence of vasculitis; and, frequently, peripheral blood eosinophilia [*see Table 3*].[4,9] Patients with this syndrome frequently have postprandial nausea and vomiting, abdominal pain, diarrhea, and, occasionally, steatorrhea; young infants experience failure to thrive, and adults have weight loss. There appears to be a subset of patients with allergic eosinophilic gastroenteritis who have symptoms secondary to food. These patients generally have the mucosal form of this disease, which is characterized by IgE-staining cells in jejunal tissue, elevated IgE in duodenal fluids, atopic disease, elevated serum IgE concentrations, positive prick skin tests to a variety of foods and inhalants, peripheral blood eosinophilia, iron deficiency anemia, and hypoalbuminemia.

The diagnosis of allergic eosinophilic gastroenteropathy is based on an appropriate history and a GI biopsy demonstrating a characteristic eosinophilic infiltration. Biopsies may need to be performed in multiple sites (up to eight) to effectively exclude eosinophilic gastroenteritis, because the eosinophilic infiltrates may be quite patchy. Patients with the mucosal form of the disease may have atopic symptoms, including food allergy, elevated serum IgE concentrations, and peripheral eosinophil-

Table 2 Immediate Gastrointestinal Hypersensitivity[27]

Manifestations	Nausea, abdominal pain and vomiting within 1–2 hr Diarrhea within 2–6 hr Frequently associated with atopic disease Food-specific IgE antibodies Radiographic: gastric hypotonia and pylorospasm
Age at onset	Infancy, childhood
Proteins implicated	Milk, eggs, peanuts, soy, cereal, fish
Pathology	IgE mediated
Treatment	Protein elimination
Natural history	80% of cases resolve after protein-elimination diet (except in the case of peanut and fish allergy)

Table 3 Allergic Eosinophilic Gastroenterocolitis[27]

Manifestations	Abdominal pain Anorexia Early satiety Failure to thrive Gastric outlet obstruction Gastric or colonic bleeding ±70% of cases atopic Elevated IgE ±Food-specific IgE 50% of cases with peripheral eosinophilia Radiographic Antral obstruction, gastroesophageal reflux, bowel wall edema
Age at onset	Birth to adolescence
Proteins implicated	Cow's milk, eggs, fish, soy, cereals Less than 50% of patients have skin-test specificity
Pathology	Marked eosinophilic infiltration of mucosa and submucosa in gastric antrum, esophagus, duodenum, and colon
Treatment	50% of patients respond to dietary elimination of documented allergen Excellent response to hydrolyzed protein formula in patients younger than 2 yr Excellent response to L-amino acid formula Responsive to steroids
Natural history	Disorder is typically prolonged

ia; they may also have positive skin tests or positive radioallergosorbent tests (RASTs). Other laboratory results consistent with this disease include findings of Charcot-Leyden crystals in the stool; anemia; and hypoalbuminemia. Such patients may also have abnormal results on D-xylose testing. An elimination diet for up to 12 weeks may be necessary before complete resolution of symptoms and normalization of intestinal histology.

NON–IgE-MEDIATED HYPERSENSITIVITY

Dietary Protein Enterocolitis

Enterocolitis from dietary protein (also known as protein intolerance) occurs most commonly in young infants between 1 week and 3 months of age [*see Table 4*]. The typical symptoms are isolated to the GI tract and consist of recurrent vomiting, diarrhea, or both. The symptoms can be severe enough to cause dehydration. Cow's milk and soy protein (particularly in infant formulas) are most often responsible for this syndrome, although egg sensitivity has been reported in older patients. The stools in affected children will often contain occult blood, polymorphonuclear neutrophils (PMNs), and eosinophils and are frequently positive for reducing substances (indicating malabsorbed sugars). Prick skin tests for the putative food protein are characteristically negative. Jejunal biopsies classically reveal flattened villi, edema, and increased numbers of lymphocytes, eosinophils, and mast cells. A food challenge with the responsible protein generally results in vomiting, diarrhea, or both within minutes to several hours; occasionally, patients experience shock.[4,10] It is not uncommon to find children who are sensitive to both cow's milk and soy protein. This sensitivity also tends to be lost by 18 to 24 months of age. Elimination of the offending allergen generally will result in improvement or resolution of the symptoms within 72 hours, although secondary disaccharidase deficiency may persist longer. Oral food challenges, which should be done in a medical setting because they can induce severe vomiting, diarrhea, dehydration, or hypotension, consist of administering 0.6 g/kg body weight of the suspected food allergen.

Dietary Protein Proctitis

Patients with dietary protein proctitis generally present in the first few months of life. This disorder is often secondary to cow's milk or soy protein hypersensitivity.[11] Most infants with this disorder do not appear ill and have normally formed stools; they generally come to medical attention because of blood (gross or occult) in their stools. GI lesions are confined to the small bowel. Grossly, the lesions range from patchy mucosal injection to severe friability with small aphthoid ulcerations and bleeding. Microscopically, they are characterized by mucosal edema with eosinophils in the epithelium and lumina propria. If lesions are severe, with crypt destruction, PMNs are also prominent.[12] It is believed that colitis induced by cow's milk or soy protein resolves by 6 months to 2 years after allergen avoidance; however, this belief is not strongly supported by well-controlled studies. Elimination of the offending food allergen leads to resolution of hematochezia within 72 hours, but the mucosal lesions may take up to 1 month to disappear.

Celiac Disease

Celiac disease is an extensive enteropathy leading to malabsorption. Total villous atrophy and an extensive cellular infiltrate are associated with sensitivity to gliadin, the alcohol-soluble portion of gluten found in wheat, oats, rye, and barley. The overall incidence of celiac disease is thought to be one in 4,000 population, but there is wide regional variation; for example, the incidence in Ireland has been reported to be as high as one in 500 population. There is apparently a genetic predisposition to this disease, given that approximately 90% of patients are HLA-B8 positive and nearly 80% have the HLA-DW3 antigen. Pa-

Table 4 Dietary Protein Enterocolitis[27]

Manifestations	Diarrhea with bleeding Anemia Emesis Abdominal distention Failure to thrive Hypotension Fecal leukocytes Normal IgE Food challenge: vomiting in 3–4 hr; diarrhea in 5–8 hr
Age at onset	1 day to 1 yr
Implicated proteins	Cow's milk, soy, rice, poultry, fish
Pathology	Patchy villous injury and colitis
Treatment	80% or more of patients respond to hydrolyzed casein formula; symptoms clear in 3–10 days Up to 20% of patients require L-amino acid formula or temporary intravenous therapy
Natural history	With treatment, 50% of cases resolve by 18 mo and 90% by 36 mo; in soy allergy, illness is often more persistent

tients often present with diarrhea or frank steatorrhea, abdominal distention and flatulence, and weight loss; occasionally, they have nausea and vomiting. Other extraintestinal symptoms and oral ulcers secondary to malabsorption are not common.

HISTORY AND PHYSICAL EXAMINATION

As with virtually all medical disorders, the diagnostic approach to a patient with a suspected adverse food reaction begins with the history and physical examination. Depending on the information derived from these initial steps, various laboratory studies may be helpful [*see Table 5*].[13-15]

In cases of suspected adverse food reactions, the value of the history depends largely on the patient's recollection of symptoms and on the examiner's ability to differentiate disorders provoked by food hypersensitivity from disorders with other etiologies. The history may be directly useful in diagnosing food allergy involving acute events (e.g., anaphylaxis after eating fish) but is not always reliable: in many series, less than 50% of reported allergic reactions to food could be substantiated by a double-blind, placebo-controlled food challenge (DBPCFC).[3,7] In chronic disorders such as atopic dermatitis, the history is typically an unreliable indicator of the offending allergen.

Several items of information are important in establishing that an allergic reaction to food occurred: (1) the type of food suspected to have provoked the reaction; (2) the quantity of the food ingested; (3) the length of time between ingestion and onset of symptoms; (4) the specific symptoms provoked; (5) whether similar symptoms developed on other occasions when the food was eaten; (6) whether other factors (e.g., exercise) were involved in the episode; and (7) the length of time since the last reaction.

Although any food may cause an allergic reaction, only a few foods account for 90% of such reactions. In children, these foods are eggs, milk, peanuts, soy, and wheat (and in Scandinavian countries, fish).

Diet Diary

A diet diary is often a useful adjunct to the medical history. Patients are asked to keep a chronologic record of all foods ingested over a specified period of time and to record any symptoms they experience during this period. The diary can be reviewed at a subsequent visit to determine whether there is a relationship between the foods ingested and the symptoms experienced. Uncommonly, this method will reveal an unrecognized association between a food and a patient's symptoms. Nevertheless, a diet diary is superior to the history in that it involves the collection of information on a prospective basis and is less dependent on the patient's or parent's memory.

Elimination Diet

Elimination diets are often used both in diagnosis and in management of adverse food reactions. If a certain food or foods are suspected of provoking the reaction, they are completely eliminated from the diet. The success of an elimination diet depends on several factors, including the correct identification of the allergen or allergens involved, the ability of the patient to maintain a diet completely free of all forms of the possible offending allergen, and the assumption that other factors will not provoke similar symptoms during the study period. The likelihood of all of these conditions being met is often slim. For example, in a young infant who is reacting to cow's milk formula, resolution of symptoms after substitution of soy formula or casein hydrolysate (e.g., Alimentum, Nutramigen) is highly suggestive of cow's milk allergy but also could reflect lactose intolerance. Avoidance of suspected food allergens before a blinded challenge is recommended so that reactions may be heightened. Elimination diets are rarely diagnostic of food allergy, particularly in chronic disorders such as atopic dermatitis or asthma.

Table 5 Methods Used in the Evaluation of Allergic Reactions to Food

Medical history
Diet diary
Elimination diet
Prick skin testing
Radioallergosorbent tests
Open or single-blind challenge
Double-blind, placebo-controlled food challenge (optimal)

LABORATORY TESTS

Allergy Skin Tests

Prick skin tests Prick skin tests are highly reproducible[16] and are often utilized to screen patients with suspected IgE-mediated food allergies. The glycerinated food extracts (1:10 or 1:20 dilution)[7] and appropriate positive (histamine) and negative (saline) controls are applied by either the prick or puncture technique. A test that elicits a wheal (not including erythema) at least 3 mm greater than the negative control is considered positive; any smaller result is considered negative. Appropriate and good-quality food extracts must be utilized for results to be reliable.

A negative skin test confirms the absence of an IgE-mediated reaction (overall negative predictive accuracy is greater than 95%). However, skin testing with commercial reagents often fails to detect IgE-mediated sensitivity to certain fruits and vegetables (e.g., apples, oranges, bananas, pears, melons, potatoes, carrots, and celery), presumably because of the labile nature of the responsible allergens in these foods. In such cases, it may be necessary to use the so-called prick-by-prick method, in which the device used for introducing the allergen into the skin is first pricked into the food. In addition, false negative results are particularly common in very young children, possibly because of lower skin reactivity: children younger than 1 year may have IgE-mediated food allergy without a positive skin test, and children younger than 2 years may have smaller wheals.

A positive skin test to a food is not definitive; it merely indicates the possibility that the patient has symptomatic reactivity to that specific food (overall, the positive predictive accuracy is less than 50%). However, a positive skin test to a food that provokes a severe anaphylactic reaction when eaten by itself may be considered diagnostic. Atopy patch tests for food allergy have been developed, but there is as yet insufficient evidence to support their adoption in clinical practice.

Intradermal skin tests An intradermal skin test is more sensitive than a prick skin test but is much less specific than a DBPCFC.[17] In one study, no patient who had a negative prick skin test but a positive intradermal skin test to a specific food had a positive DBPCFC to that food.[17] In addition, intradermal skin testing is more likely to induce a systemic reaction than is

prick skin testing. For those reasons, intradermal skin tests have no role in the diagnosis of food allergy.

In Vitro Assays

RASTs and similar in vitro assays (including enzyme-linked immunosorbent assays [ELISAs]) are often used to screen for IgE-mediated food allergies. Although RASTs are generally considered slightly less sensitive than skin tests, one study comparing skin tests and Phadebas RAST (Pharmacia & Upjohn) with DBPCFCs found prick skin tests and Phadebas RAST to have similar sensitivity and specificity when a Phadebas score of 3 or greater was considered positive.[18] In this study, lowering the cutoff point for a positive result to a score of 2 brought a slight improvement in sensitivity at the expense of a significant decrease in specificity. In general, in vitro measurements of serum food-specific IgE performed in high-quality laboratories provide information similar to prick skin tests. The newest generation of in vitro studies for specific IgE includes the capsulated hydrolic carrier polymer–fluoroenzymeimmunoassay (RAST-FEIA). For patients with suspected food allergy, there are now accepted levels of food-specific IgE concentrations on CAP-FEIA testing that can predict a patient's being allergic to that food with greater than 95% certainty.[15] CAP-FEIA is best used for patients with allergic reactions to milk, eggs, and peanuts (and, possibly, wheat, soy, and fish).

Double-Blind Placebo-Controlled Food Challenge

The DBPCFC has been considered the gold standard for the diagnosis of food allergy.[4] This test has been used successfully by many investigators in both children and adults for the past several years to examine a wide variety of food-related complaints. The selection of foods to be tested in the oral challenge is based on the history or prick skin test (RAST) results.

A DBPCFC is the best means of controlling for the variability of chronic disorders (e.g., chronic urticaria, atopic dermatitis), any potential temporal effects, and acute exacerbations secondary to reducing or discontinuing medications. In particular, psychogenic factors and observer bias are eliminated. Rarely, a false negative DBPCFC may occur when the challenge material a patient receives is not of sufficient quantity to provoke the reaction or when the lyophilization of the food antigen has altered the relevant allergenic epitopes (as may occur with fish). Nevertheless, at present, the DBPCFC has proved to be the most accurate means of diagnosing food allergy.

PRACTICAL APPROACH TO DIAGNOSING FOOD ALLERGY

The diagnosis of food allergy remains a clinical exercise that utilizes a careful history, selective prick skin tests or RASTs (if an IgE-mediated disorder is suspected), appropriate exclusion diet, and blinded provocation. Other diagnostic tests that do not appear to be of significant value include assessment of food-specific IgG or IgG4 antibody levels, assessment of food antigen–antibody complexes, measures of lymphocyte activation (e.g., ^{3}H uptake, interleukin-2 production, presence of leukocyte inhibitory factor), and sublingual or intracutaneous provocation. Blinded challenges may not be necessary in suspected GI disorders, which often can be diagnosed on the basis of prechallenge and postchallenge laboratory values and biopsy results.

An exclusion diet that eliminates all foods suspected by history or prick skin testing (or, in IgE-mediated disorders, RAST) should be conducted for at least 1 to 2 weeks. Some patients with GI disorders may need to have the exclusion diet extended for up to 12 weeks after appropriate biopsies. If no improvement is noted after institution of the diet, food allergy is unlikely. In patients with some chronic diseases, such as atopic dermatitis or chronic asthma, other precipitating factors may make it difficult to discriminate the effects of the food allergen from other provocative factors.

Open or single-blind challenges in a clinic setting may be helpful to test for allergy to specific foods. Such challenges are less cumbersome and time-consuming than DBPCFCs. It is important that the clinician make an unequivocal diagnosis of food allergy; a presumptive diagnosis of food allergy based on a patient's history and prick skin tests or RAST results is no longer acceptable. There are exceptions to this, such as the patient who experiences severe anaphylaxis after the isolated ingestion of a specific food. Because of reliance on presumptive diagnosis, over one quarter of the United States population have altered their eating habits on the basis of misconceptions about food allergy.

Treatment

The only proven therapy for food allergy is the strict elimination of that food from the patient's diet. In infants, breast-feeding avoids contact with potentially allergenic foods if the mother avoids those foods. In infants with a family history of atopy, moreover, exclusive breast-feeding for at least the first 4 months of life appears to lessen the likelihood of atopic dermatitis.[19] Elimination diets should be supervised because they may lead to malnutrition or eating disorders, especially if they involve the elimination of a large number of foods or are utilized for extended periods of time. Studies have shown that symptomatic food sensitivity generally is lost over time, except for sensitivity to peanuts, tree nuts, and seafood.

Symptomatic food sensitivity is usually very specific; patients rarely react to more than one member of a botanical family or animal species. Consequently, clinicians should confirm that patients are not unnecessarily limiting their diet for fear of allergic reactions. Risk factors for more severe anaphylactic reactions include the following: (1) a history of a previous anaphylactic reaction; (2) a history of asthma, especially if poorly controlled; (3) allergy to peanuts, nuts, fish, or shellfish; (4) current treatment with beta blockers or angiotensin-converting enzyme inhibitors; and, possibly (5) female sex.

PHARMACOLOGIC THERAPY

Several medications have been used in an attempt to protect patients with food hypersensitivity, including oral cromolyn, H_1 and H_2 antihistamines, ketotifen, corticosteroids, and prostaglandin synthetase inhibitors. Some of these medications may modify food allergy symptoms, but overall they have minimal efficacy or unacceptable side effects.

Epinephrine

The importance of prompt administration of epinephrine when symptoms of systemic reactions to foods develop cannot be overemphasized [*see 6:XIII Urticaria, Angioedema, and Anaphylaxis*]. Patients with a history of anaphylaxis should always carry an epinephrine autoinjector (Epi-Pen [0.3 mg] or Epi-Pen, Jr. [0.15 mg]). For anaphylaxis, epinephrine is given in a dose of 0.01 mg/kg. The route of administration can be intramuscular or subcutaneous, but most recent studies suggest that intramuscular administration is better.

Table 6 Resources for Patients with Food Allergy

The Food Allergy and Anaphylaxis Network
10400 Easton Place, Suite 107
Fairfax, VA 22030-5647
800-929-4040
http://www.foodallergy.org

National Allergy and Asthma Network/Mothers of Asthmatics
2751 Prosperity Avenue
Suite 150
Fairfax, VA 22031
800-878-4403 / 703-385-4403
http://www.aanma.org

Asthma and Allergy Foundation of America
1125 15th Street, NW, Suite 502
Washington, DC 20005
800-727-8462
http://www.aafa.org

IMMUNOTHERAPY

Blinded, placebo-controlled studies of rush immunotherapy for the treatment of peanut hypersensitivity demonstrated efficacy in a small number of patients.[20] The adverse reaction rates were significant, and such reactions preclude general clinical application at this time. Except for patients who are at risk of life-threatening reactions to minuscule amounts of peanut, there is no use for immunotherapy in food allergies. Newer types of immunotherapy for preventing food-induced anaphylaxis that are being developed include the following: (1) humanized anti-IgE monoclonal antibody therapy, (2) plasmid-DNA immunotherapy, (3) peptide fragments (so-called overlapping peptides), (4) cytokine-modulated immunotherapy, (5) immunostimulatory sequence-modulated immunotherapy, (6) bacterial-encapsulated allergen immunotherapy, and (7) recombinant protein immunotherapy.[21] A recent study with anti-IgE in peanut-allergic patients[22] demonstrated that this medication may eventually be helpful for preventive treatment of food-induced anaphylaxis.

PATIENT EDUCATION

Patient education and support are essential for patients with food allergy. In particular, adults and older children who are prone to anaphylaxis, as well as parents of pediatric patients, must be informed in a direct but sympathetic way that these reactions are potentially fatal.

When eating away from home, food-sensitive persons should feel comfortable requesting information about the contents of prepared foods. For school-aged children, the American Academy of Pediatrics Committee of School Health has recommended that schools be equipped to treat anaphylaxis in allergic students. Children older than 7 years can usually be taught to inject themselves with epinephrine. The physician must be willing to explain these issues to school personnel and, with the parents, help instruct these personnel in how to deal with them. In the home, the family should eliminate the incriminated allergen or, if this is not practical, place warning stickers on foods with the offending antigens. A variety of groups can help provide patients with support, advocacy, and education about food allergy [*see Table 6*].

Prognosis

Young children who are diagnosed with anaphylaxis to foods such as milk, eggs, wheat, or soybeans often outgrow this clinical sensitivity after several years.[9,23,24] Children who develop food sensitivity after 3 years of age are less likely to lose their food reactions over a period of several years. Patients who have very mild reactions (i.e., skin symptoms only) to peanuts early in life (i.e., during the first 12 to 24 months) may outgrow their symptoms.[25,26] Allergies to foods such as tree nuts, fish, and seafood are generally not outgrown regardless of the age at which they develop. These persons appear likely to retain their allergic sensitivity for a lifetime. Consequently, new strategies are being evaluated to desensitize patients to these foods.

The author owns stock, stock options, and/or bonds in SEER Inc.

References

1. Anderson J, Sogn D: Adverse reactions to foods. American Academy of Allergy and Immunology Committee on Adverse Reactions to Foods and the National Institute of Allergy and Infectious Disease. National Institute of Allergy and Infectious Diseases, Washington, DC, 1984

2. Sampson HA, Burks AW: Mechanisms of food allergy. Annu Rev Nutr 16:161, 1996

3. Sampson HA: Food allergy. J Allergy Clin Immunol 111(2 suppl):S540, 2003

4. Sampson HA, Anderson JA: Summary and recommendations: classification of gastrointestinal manifestations due to immunologic reactions to foods in infants and young children. J Pediatr Gastroenterol Nutr 30(suppl):S87, 2000

5. Bock SA: Natural history of severe reactions to foods in young children. J Pediatr 107:676, 1985

6. Bock SA, Atkins FM: Patterns of food hypersensitivity during sixteen years of double-blind, placebo-controlled food challenges. J Pediatr 117:561, 1990

7. Sampson HA: Adverse reactions to foods. Allergy: Principles and Practice, 5th ed. Middleton E Jr, Reed CE, Ellis EF, et al, Eds. Mosby-Year Book, St Louis, 1998, p 1162

8. Eigenmann PA, Sicherer SH, Borkowski TA, et al: Prevalence of IgE-mediated food allergy among children with atopic dermatitis. Pediatrics 101:E8, 1998

9. Bock SA: Prospective appraisal of complaints of adverse reactions to foods in children during the first 3 years of life. Pediatrics 79:683, 1987

10. Goldman AS, Anderson DW, Sellers WA, et al: Milk allergy: I. Oral challenge with milk and isolated milk proteins in allergic children. Pediatrics 32:425, 1963

11. Crowe SE, Perdue MH: Gastrointestinal food hypersensitivity: basic mechanisms of pathophysiology. Gastroenterology 103:1075, 1992

12. Jenkins HR, Pincott JR, Soothill JF, et al: Food allergy: the major cause of infantile colitis. Arch Dis Child 59:326, 1984

13. Burks AW, Sampson HA: Diagnostic approaches to the patient with suspected food allergies. J Pediatr 121:S64, 1992

14. Schwartz HJ: Asthma and food additives. Food Allergy: Adverse Reactions to Foods and Food Additives, 2nd ed. Metcalfe DD, Sampson HA, Simon RA, Eds. Blackwell Science, Oxford, 1997, p 411

15. Sampson HA: Utility of food-specific IgE concentrations in predicting symptomatic food allergy. J Allergy Clin Immunol 107:891, 2001

16. Bock SA, Lee WY, Remigio L, et al: Appraisal of skin tests with food extracts for diagnosis of food hypersensitivity. Clin Allergy 8:559, 1978

17. Bock SA, Buckley J, Holst A, et al: Proper use of skin tests with food extracts in diagnosis of hypersensitivity to food in children. Clin Allergy 7:375, 1977

18. Sampson HA, Albergo R: Comparison of results of skin tests, RAST, and double-blind, placebo-controlled food challenges in children with atopic dermatitis. J Allergy Clin Immunol 74:26, 1984

19. Schoetzau A, Filipiak-Pittroff B, Franke K, et al: Effect of exclusive breast-feeding and early solid food avoidance on the incidence of atopic dermatitis in high-risk infants at 1 year of age. Pediatr Allergy Immunol 13:234, 2002

20. Oppenheimer JJ, Nelson HS, Bock SA, et al: Treatment of peanut allergy with rush immunotherapy. J Allergy Clin Immunol 90:256, 1992

21. Burks AW, Bannon GA, Lehrer SB: Classic specific immunotherapy and new perspectives in specific immunotherapy for food allergy. Allergy 67:121, 2001

22. Leung DY, Sampson HA, Yunginger JW, et al: Effect of anti-IgE therapy in patients with peanut allergy. N Engl J Med 348:986, 2003

23. Bock SA: The natural history of food sensitivity. J Allergy Clin Immunol 69:173, 1982

24. Bock SA, Atkins FM: Patterns of food hypersensitivity during sixteen years of double-blind, placebo-controlled food challenges. J Pediatr 117:561, 1990

25. Skolnick HS, Conover-Walker MK, Koerner CB, et al: The natural history of peanut allergy. J Allergy Clin Immunol 107:367, 2001

26. Hourihane JO, Roberts SA, Warner JO: Resolution of peanut allergy: case-control study. BMJ 316:1271, 1998

27. Zeiger RS: Dietary aspects of food allergy prevention in infants and children. J Pediatr Gastroenterol Nutr 30(suppl):S77, 2000

I INFECTIONS DUE TO GRAM-POSITIVE COCCI

Dennis L. Stevens, ph.d., m.d., f.a.c.p.

Pneumococci

Pneumococci are gram-positive, lancet-shaped diplococci that may also grow in short chains. When cultured on blood agar plates, they form small, colorless mucoid colonies with characteristic central depressions. The colonies are surrounded by greenish discoloration of the blood agar, termed α-hemolysis. Laboratories have traditionally identified pneumococci on the basis of sensitivity to surface-active agents such as bile or optochin; rapid speciation can now be accomplished with a latex agglutination test or a DNA probe.

Although there is only one species of *Streptococcus pneumoniae,* there are at least 84 distinct serotypes, which are differentiated by the composition of the polysaccharide polymer that forms their outer capsule. Each capsular type is chemically and antigenically unique. The pneumococcal polysaccharide capsule is crucial to virulence. The capsule allows the bacteria to resist phagocytosis by leukocytes unless the organisms have been opsonized by antibody or serum complement components. Antibodies to capsular polysaccharide are essential for recovery from untreated pneumococcal pneumonia. Although such antibodies provide long-lasting immunity, the immunity is strictly type specific. Certain pneumococcal serotypes, such as the heavily encapsulated type 3 pneumococcus, are particularly virulent. In fact, only 23 serotypes account for about 80% of bacteremic pneumococcal infections in the United States; capsular polysaccharides from these 23 serotypes are incorporated into the polyvalent pneumococcal vaccine that has been available since 1983.

Although the polysaccharide capsule is the critical factor in determining the virulence of pneumococci, several proteins also contribute to the pathogenesis of pneumococcal infections. Surface protein A, neuraminidase, pneumolysin, and IgA protease appear to be the most important of these proteins.[1]

epidemiology of pneumococcal infections

The nasopharynx is the natural habitat of the pneumococcus; humans are the only known hosts. The nasopharyngeal carrier rate varies widely, from a low of 5% to 10% to a high of up to 60% in closed populations during the winter.[2] Pneumococci are fastidious, nonsporulating bacteria. Because they are rapidly killed by drying or by extremes in temperature, person-to-person spread by droplets requires close contact.

The pneumococcus remains the most important cause of bacterial pneumonia; it may account for as many as 500,000 cases each year in the United States. Like other respiratory tract infections, pneumococcal pneumonia is more common in winter. Cigarette smoking is the strongest independent risk factor for invasive pneumococcal disease in immunocompetent, nonelderly adults.[3] Extremes of age and chronic disease or immunosuppression are also risk factors [*see Table 1*].[4] Diseases that increase risk include cirrhosis, sickle cell anemia, multiple myeloma, chronic lung disease, and cancer. Organ transplantation recipients are also highly susceptible.

The pneumococcus accounts for up to 40% of community-acquired pneumonias, causing or contributing to 40,000 deaths annually. The overall case-fatality rate of this pneumonia is 5% to 12%[5]; patients who are very young, elderly, alcoholic, leukopenic, infected with HIV, or infected with certain virulent pneumococcus serotypes (notably, type 3) are at the highest risk. Pneumococcal bacteremia develops in about 50,000 Americans annually, with a case-fatality rate of 20%—a rate that has not changed over the past 40 years.[6] Pneumococcal otitis media is one of the most common childhood illnesses; the estimated incidence in the United States is seven million cases a year. Pneumococci cause about 3,000 cases of meningitis each year, with a case-fatality rate of about 21%.[7]

Epidemics of pneumococcal disease were common in the preantibiotic era, but sporadic cases are now the rule. Outbreaks have occurred, however, in shelters for the homeless, in jails, and in child care centers. Pneumococcal infection is typically community acquired, but it can occur nosocomially, particularly in chronically ill immunosuppressed patients; because of underlying diseases, the case-fatality rate of nosocomial pneumococcal bacteremia is 40%.[8] The incidence of pneumococcal bacteremia is increased about 40-fold in patients with AIDS.

pathogenesis of pneumococcal infections

Pneumococci produce IgA protease, which cleaves the secretory immunoglobulin IgA1. IgA protease may be partly responsible for the high carrier rate and the respiratory tract pathogenicity of these bacteria. The organisms do not provoke an inflammatory response or cause clinical illness in the nasopharynx. Host defense mechanisms prevent penetration into more vulnerable areas, such as the paranasal sinuses, the middle ear, and pulmonary alveoli. Examples of these defenses include the cough, gag, and sneeze reflexes; the viscous mucus that lines the respiratory epithelium and traps bacteria; and ciliary action, which expels trapped particles. Pneumococcal infection often follows influenza[9] or other viral infections of the upper respiratory tract, which impair host defenses by increasing the volume and decreasing the viscosity of secretions. Other factors that predispose to pneumococcal infection include dementia, seizure disorders, alcoholism, stupor, and other conditions that increase

Table 1 Incidence of Pneumococcal Disease According to Age and Underlying Disease

Risk Factor	*Cases of Pneumococcal Disease per 100,000 Population*
Age	
< 1 yr	21
1 yr	51
> 65 yr	45
Chronic medical conditions	176–483
Severe immunosuppression	562–2,031

the likelihood that oropharyngeal contents will be aspirated into the lungs. Air pollution, cigarette smoking, heart failure, chronic obstructive pulmonary disease (COPD), and HIV infection increase the risk of pneumococcal infection, but diabetes does not.

Pneumococcal pneumonia begins with aspiration of small quantities of oropharyngeal secretions containing pneumococci. Normally, alveolar macrophages ingest and kill such bacteria, but in the absence of type-specific immunity, strains of pneumococci are protected from ingestion. The presence of pneumococci in the alveoli provokes a vigorous inflammatory response. The earliest manifestations are vasodilatation, increased vascular permeability, and exudation of edema fluid—the classic pathologic stage of congestion. Pneumococci can survive and even proliferate in the edema fluid. The organisms appear to float to adjacent alveoli, producing lobar consolidation. Within hours to days, polymorphonuclear leukocytes arrive and progressively pack the alveoli to produce the traditional stages of red and gray hepatization. However, only a minority of the pneumococci are phagocytosed and killed by the leukocytes in the first days of infection. Capsular polysaccharide antigen is often present in the serum and urine during these early stages.

After 5 to 7 days, type-specific anticapsular antibody appears. This results in a more efficient and enhanced opsonization that involves anticapsular antibody and the classical complement pathway. As pneumococci are ingested and killed by the polymorphonuclear leukocytes, the patient experiences a clinical crisis, marked by an abrupt fall in temperature and an increase in well-being; resolution commences. Because pneumococci only rarely produce significant tissue necrosis, healing is usually complete, and residual fibrosis is minimal.

CLINICAL PNEUMOCOCCAL INFECTIONS

Pneumonia

Clinical presentation Sir William Osler's classic description of pneumococcal pneumonia was recorded in 1892,[10] only 10 years after the discovery of the pneumococcus, but it remains lucid and accurate today:

> Abruptly, or preceded by a day or two of indisposition, the patient has a severe chill, lasting from ten to thirty minutes. In no acute disease is an initial chill so constant or so severe. The fever rises quickly. There is pain in the side, often of an agonizing character. A short, dry painful cough soon develops and the respirations are increased in frequency. When seen on the second or third day the patient presents an appearance which may be quite pathognomonic. He lies flat in bed, often on the affected side; the face is flushed, particularly the cheeks; the breathing is hurried; the alae nasi dilate with every inspiration; the eyes are bright, the expression is anxious, and there is a frequent short cough which makes the patient wince and hold his side. The expectoration is blood-tinged and extremely tenacious. The temperature rises rapidly to 104° or 105°. The pulse is full and bounding and the pulse-respiration ratio much disturbed. Examination of the lung shows the physical signs of consolidation—blowing breathing and fine rales. After persisting for from seven to ten days the crisis occurs, and with a fall in the temperature the patient passes from a condition of extreme distress and anxiety to one of comparative comfort.
>
> Pneumonic sputum is viscid, tenacious, and blood-tinged. The gummous viscidity, together with the red blood–corpuscles in various stages of alteration, give pathognomonic characters to the sputa, unknown in any other disease. The rusty tinge becomes more marked as the disease progresses, and so tenacious is the expectoration that it has to be wiped from the lips of the patient, and a spitcup, half full, may be inverted without spilling. Microscopically, the sputum contains red blood–corpuscles in all stages of degeneration, alveolar epithelium, diplococci and other micro-organisms, cell-moulds of the alveoli, and, in some cases, small fibrinous casts of the bronchioles. The latter are sometimes plainly visible to the naked eye.

It should be emphasized, however, that not all patients with pneumococcal pneumonia present with a viral upper respiratory tract infection that is followed by the abrupt onset of fever, chills, pleurisy, dyspnea, and cough productive of purulent or blood-tinged sputum. If the patient has used antipyretics, fever can be modest or absent. The initial rigor may also be absent, but patients may occasionally have recurrent chills. In elderly or debilitated patients, confusion or stupor may be the presenting feature, far overshadowing pulmonary symptoms. In contrast, patients with chronic pulmonary disease may have rapidly progressive respiratory failure, which is disproportionately more severe than fever, sputum production, or other manifestations of infection. In rare cases, asplenic patients present with shock and hemorrhagic skin lesions that reflect disseminated intravascular coagulation. Because modest hyperbilirubinemia is common in pneumococcal pneumonia, right lower lobe pneumonia may masquerade as acute cholecystitis because of fever, jaundice, and right upper quadrant discomfort.

Laboratory tests The key to the diagnosis of pneumococcal pneumonia is the Gram stain of a sputum smear, which typically reveals many polymorphonuclear leukocytes and abundant lancet-shaped gram-positive diplococci. If the patient cannot spontaneously produce an adequate sputum specimen, chest physiotherapy or nasotracheal suctioning should be considered. Sputum specimens should be promptly cultured on blood agar plates, preferably in 5% CO_2 incubators. Because pneumococci are fastidious and fragile, the sputum culture may be negative in patients with clearly positive Gram stains. It is important to obtain blood cultures in all patients with suspected pneumococcal pneumonia. In 25% to 30% of these patients, the blood culture will be positive—a finding that confirms the diagnosis even if sputum cultures are negative. An immunochromatographic test can be used to detect pneumococcal polysaccharide in the urine of patients with pneumonia, but pneumococcal nasopharyngeal colonization can produce false positive results.[11]

Imaging tests The classic physical and radiographic findings of lobar consolidation may be absent in patients with pneumococcal pneumonia. In fact, a bronchopneumonic pattern is radiographically more common than lobar consolidation. Dehydration may minimize pulmonary findings, and underlying chronic lung disease may predispose to patchy areas of pulmonary infiltration. Pleural effusions are relatively common and can occasionally obscure the underlying pulmonary parenchymal involvement.

Differential diagnosis Other bacterial and nonbacterial pneumonias must be considered in the differential diagnosis of pneumococcal pneumonia [*see 7:XX Pneumonia and Other Pulmonary Infections*]. Less commonly, pulmonary edema, pulmonary emboli, atelectasis, or lung tumors are mistaken for pneumonia.

Complications and prognostic indicators The complications associated with pneumococcal pneumonia have diminished

markedly in the antibiotic era. Intrathoracic complications include pleurisy with sterile pleural effusion (common) and empyema (uncommon). Lung abscess is rare; if abscess or empyema occurs, it is likely to be caused by the heavily encapsulated type 3 pneumococcus or by a concomitant anaerobic infection. Purulent pericarditis is even rarer than abscess or empyema. Radiographic abnormalities often resolve slowly. About one third of patients have persistent consolidation at 1 month; although consolidation should resolve in all patients by 8 to 10 weeks, volume loss, pleural disease, and interstitial changes can persist for up to 4 months.

Bacteremia is by far the most common extrathoracic complication. Approximately 90% of patients with pneumococcal bacteremia have pneumonia, but patients occasionally present with pneumococcal bacteremias without an identifiable septic focus. Recurrent pneumococcal bacteremia may develop in patients with underlying diseases. Bacteremia is an adverse prognostic sign. The reported fatality rate is 11% to 36%, but it is probably higher in the elderly and in patients with severe underlying diseases. Metastatic infections such as meningitis, septic arthritis, peritonitis, and endocarditis are relatively uncommon; they appear to be more common and to occur at younger ages in African Americans than in whites.[12]

Several factors are associated with worse outcomes in pneumococcal pneumonia [*see Table 2*]. Interestingly, a lack of febrile response and a normal or low white blood cell count are readily measurable factors that are associated with worse outcome. Thus, although white blood cell counts of 25,000 to 30,000 with a left shift may be alarming, they indicate a favorable host response to infection.[13]

Table 2 Factors Associated with Adverse Outcomes in Pneumococcal Pneumonia

Infection with type 3 pneumococci
Bacteremia
Hypothermia
Neutropenia
Multilobe infection
Extrapulmonary infection (meningitis, empyema)
Hypotension

Pneumococcal Infections in Persons Infected with HIV

The pneumococcus is the leading cause of invasive bacterial respiratory tract infections in HIV-positive persons. The clinical features, causative serotypes, antimicrobial resistance patterns, and mortalities of pneumococcal infection in HIV-positive patients are similar to those in HIV-negative patients. Severe infections, unusual extrapulmonary manifestations, and late relapses are more common in these patients, however. Pneumococcal vaccine should be administered as soon as possible after the diagnosis of HIV infection; despite the decreased efficacy of vaccine in patients with advanced AIDS, vaccination is a cost-effective preventive intervention at all stages of HIV infection.[14]

Other Pneumococcal Infections

Upper respiratory tract infections Pneumococci spread from the nasopharynx to the upper respiratory tract to produce acute otitis media, especially in children. Acute mastoiditis, once a frequent sequela of acute otitis media, is now unusual. Pneumococci are a cause of acute purulent sinusitis in all age groups [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*].

Meningitis Pneumococci can reach the central nervous system either by bacteremic spread from a pulmonary focus or by direct extension from otitis or sinusitis. Patients with skull fractures and cerebrospinal fluid rhinorrhea or otorrhea are particularly prone to recurrent attacks of pneumococcal meningitis [*see 7:XXXVI Bacterial Infections of the Central Nervous System*].

Septic arthritis and osteomyelitis Pneumococci are a relatively common cause of acute septic arthritis, which results from bacteremic seeding; prosthetic joints may be involved [*see 7:XV Septic Arthritis*]. Pneumococcal osteomyelitis is uncommon.

Cardiac infections Antibiotics have greatly reduced the incidence of pneumococcal pericarditis, which can result from direct extension of pneumonia or empyema or from hematogenous seeding. Pneumococcal endocarditis has also become uncommon in the antibiotic era.[15]

Postsplenectomy infections Overwhelming postsplenectomy pneumococcal infection is an uncommon but important syndrome. In addition to patients who have undergone splenectomy, other patients at risk include those with sickle cell disease or other hemoglobinopathies that produce functional asplenia, as well as those with congenital asplenia. The syndrome is often marked by acute onset of fever, hemorrhagic skin lesions suggestive of disseminated intravascular coagulation or purpura fulminans, and shock. Hypoglycemia may be present. If therapy is not administered, death often occurs in less than 24 hours. Even when patients receive penicillin and cardiovascular support, mortality exceeds 50%. A similar syndrome can occur in healthy adults but is rare.[16]

Other infections Primary pneumococcal peritonitis can occur in patients who have cirrhosis, nephrotic syndrome, systemic lupus erythematosus, or other host defects [*see 7:XXI Peritonitis and Intra-Abdominal Abscesses*]. In rare cases, pneumococci can infect the liver, the gallbladder, or pelvic organs; soft tissue infections and cellulitis caused by *S. pneumoniae* are also unusual.

TREATMENT OF PNEUMOCOCCAL INFECTIONS

For many years, all isolates of *S. pneumoniae* were penicillin sensitive; until 1965, most isolates in the United States were sensitive to less than 0.04 μg/ml penicillin. Subsequently, resistance to penicillin has become progressively more prevalent. Although pneumococci do not display plasmid-mediated penicillinase production, they can develop chromosomal mutations that confer resistance to penicillin by altering the affinity of the penicillin-binding proteins in their cell walls. Gradual remodeling of three or four of the penicillin-binding proteins in parallel produces a stepwise increase in the level of resistance. The DNA sequences responsible for resistance probably originated in other streptococcal species and were transferred to pneumococci by heterologous recombination.[16]

Pneumococci for which the minimum inhibitory concentration (MIC) of penicillin is less than 0.1 μg/ml are considered penicillin sensitive; those in which the MIC of penicillin is 0.1 to 1.0 μg/ml are considered intermediately resistant; and those for which the MIC is greater than 1.0 μg/ml are considered resistant. About 24% of pneumococcal organisms isolated in the

United States are penicillin-nonsusceptible *S. pneumoniae* (PNSP), but in some areas, more than one third of *S. pneumoniae* isolates are resistant to penicillin.[17] The prevalence of PNSP increases in patients who have been using antibiotics; it is highest in children younger than 6 years and in adults older than 65 years. Resistant strains can be spread from person to person, especially in closed population groups, such as those in day care centers, jails, and nursing homes. Infection with PNSP is associated with an increased likelihood of an adverse outcome.[18]

Pneumococci that are resistant to penicillin are often resistant to other antimicrobial drugs.[17] First- and second-generation cephalosporins are generally ineffective against these organisms, but third-generation cephalosporins (particularly ceftriaxone and cefotaxime) and carbapenems are usually active. Erythromycin and the other macrolides are generally ineffective, as are clindamycin and trimethoprim-sulfamethoxazole (TMP-SMX); chloramphenicol has variable efficacy. Rifampin and vancomycin are active against virtually all isolates, but vancomycin tolerance has been identified and may become a concern in the future.[19] Whereas many pneumococci have become resistant to the older fluoroquinolones, such as ciprofloxacin,[20] newer agents in this class, such as levofloxacin, sparfloxacin, gatifloxacin, and moxifloxacin, are generally active against penicillin-resistant pneumococci. Levofloxacin resistance has been documented, however, especially in elderly nursing home patients, in patients with COPD, and in patients who had prior exposure to fluoroquinolones.[20] Linezolid[21] is active against all pneumococci, as are ketolides, the still-investigational glycylcyclines, and daptomycin. Because of the increasing problem of PNSP, organisms isolated from clinical sources should be screened for penicillin resistance with a 1 μg oxacillin disk. Isolates with oxacillin inhibition zones of less than 19 mm should be studied further to determine the MICs of penicillin and other antimicrobial agents that are likely to be used in treatment.

Guidelines for treatment of patients with PNSP are being formulated.[22] Pneumonia caused by PNSP strains can still be treated with penicillin G, although dosages must be in the range of 10 to 20 million units a day for average-size adults [*see Table 3*]. Ceftriaxone and cefotaxime are generally effective, but treatment failures in patients with meningitis have been reported. Until the results of susceptibility testing are available, it may be advisable to add vancomycin to the regimen of cefotaxime or ceftriaxone for patients with life-threatening pneumococcal infections such as meningitis. Vancomycin is an effective alternative for patients in whom treatment fails and for those who cannot tolerate cephalosporins; however, because of the limited CSF penetration of I.V. vancomycin in patients with pneumococcal meningitis, concurrent therapy with intrathecal vancomycin or I.V. rifampin may be advisable. The role of imipenem or meropenem in treating patients with PNSP is being explored. The newer fluoroquinolones (levofloxacin, sparfloxacin, gatifloxacin, and moxifloxacin) may be useful in adults, though none have indications in children. Linezolid has been approved for the treatment of pneumonia caused by PNSP, though no studies have been completed in patients with meningitis. To limit the spread of PNSP, physicians should avoid the inappropriate use of antibiotics and encourage the appropriate use of the pneumococcal vaccine.

Penicillin remains the drug of choice for susceptible pneumococci; cephalosporins are also active against such organisms but must be used with caution in patients allergic to penicillin. Other drugs that can be used to treat susceptible pneumococci include erythromycin and the new macrolides, ketolides, vancomycin, and the newer fluoroquinolones. Clindamycin has been a useful drug for treating a variety of pneumococcal infections, though inducible clindamycin resistance has been recently reported. TMP-SMX is a particularly good choice for patients with sinusitis or otitis; chloramphenicol may be useful for treating meningitis caused by susceptible pneumococci in patients who cannot tolerate penicillin, third-generation cephalosporins, or meropenem. Patients with pneumococcal pneumonia can be switched from intravenous to oral antibiotics once they are clinically stable, even if bacteremia was present.[23]

Table 3 Antibiotic Treatment for Penicillin-Resistant *S. pneumoniae*

Drug	*Dosage*	*Comments*
Penicillin G	10–20 million units/day	Despite resistance, high-dose penicillin is effective for pneumonia
Cephalosporins		Relative efficacy high, but cephalosporin resistance occurs*
Ceftriaxone	1 g I.V. q.d.	
Cefotaxime	3 g I.V. q. 6 hr	
Chloramphenicol	1 g I.V. q. 6 hr	Variable efficacy; used sparingly because of toxicity
Rifampin*	600 mg I.V. q. 24 hr	Used as part of combination therapy for meningitis only
Vancomycin	1 g I.V. q. 12 hr	Intermediate efficacy; not justified for pneumonia but can be considered for meningitis*
Fluoroquinolones		High efficacy for pneumonia; resistance to fluoroquinolones has been described
Levofloxacin	500 mg I.V. q.d.	
Gatifloxacin	400 mg I.V. q.d.	
Moxifloxacin	400 mg I.V. q.d.	
Linezolid	600 mg I.V. q. 12 hr	High efficacy for pneumonia; not studied for meningitis
Telithromycin	800 mg p.o., q.d.	High efficacy; macrolide-resistant pneumococci are susceptible to telithromycin and other ketolides
Meropenem	1 g I.V. q. 8 hr	Could be used for cephalosporin-resistant pneumococci, but its effectiveness has not been studied systematically

*Triple-drug therapy with vancomycin, rifampin, and either ceftriaxone or cefotaxime may be necessary for pneumococcal meningitis if the minimum inhibitory concentration of either cephalosporin is > 4 mg/ml. Despite this resistance, vancomycin and cefotaxime demonstrate synergy.

PREVENTION OF PNEUMOCOCCAL INFECTIONS

A vaccine containing 50 μg of purified capsular polysaccharide from 14 pneumococcal types was approved by the Food and Drug Administration in 1977, and an expanded vaccine containing 25 μg of polysaccharide from 23 pneumococcal types was released in 1983. These 23 types account for the great majority of serious pneumococcal infections in the United States, including 88% of invasive infections with penicillin-resistant strains.[17] Because polysaccharide vaccines are not effective in children younger than 2 years, a protein-polysaccharide conjugate vaccine containing approximately 2 μg of polysaccharides from the seven most important strains conjugated to diphtheria toxoid was introduced for pediatric use in 2000.[24]

The 23-valent pneumococcal vaccination produces serum antibody titers that are protective in healthy adults; although most elderly people mount an adequate serologic response, some produce low antibody titers or functionally deficient antibodies. Patients with Hodgkin disease respond better to vaccination before staging laparotomy and splenectomy; chemotherapy markedly impairs the response to vaccination and accelerates the decline in antibody titers. Dialysis and transplant patients may respond to vaccination suboptimally,[25] as do many patients with myeloma, lymphoma, leukemia, AIDS, and low serum vitamin B_{12} levels.

The pneumococcal vaccine is safe.[26] Mild erythema and pain at the injection site may occur in up to one third of vaccine recipients, but fever, severe local reactions, and other serious side effects occur in fewer than 1%. Despite its immunogenicity and safety, the pneumococcal vaccine remains controversial because its efficacy has been difficult to demonstrate in certain population groups. In general, the evidence for efficacy has been stronger for the prevention of invasive pneumococcal disease than for nonbacteremic pneumonias[27]; the evidence is also stronger for the protection of low-risk persons than for the protection of elderly or debilitated patients.

Perhaps as a result of conflicting data, pneumococcal vaccine has not been well accepted, and only 35% of vaccine candidates have been immunized. Given the expense of large-scale field trials, the controversy is unlikely to be resolved in the near future, and recommendations for vaccine use must be formulated from the current information. Because of its safety and modest cost, the pneumococcal vaccine is cost-effective, even in the elderly.[28] The vaccine is recommended for healthy adults older than 65 years and for patients with chronic cardiopulmonary disease, functional or anatomic asplenia (including sickle cell disease), Hodgkin disease, multiple myeloma, cirrhosis, alcoholism, renal failure, CSF leaks, immunosuppression, or HIV infection. Members of certain vulnerable population groups, such as Native Americans, are also candidates for the vaccine. When possible, the vaccine should be administered 2 weeks before elective splenectomy or chemotherapy [*see CE:V Adult Preventive Health Care*].

The pediatric conjugate vaccine is safe and effective[29] and has reduced by 60% the rate of invasive infection caused by vaccine and vaccine-related serotypes in children younger than 5 years. Currently, it is recommended for all infants, as well as for children up to 60 months of age who are at increased risk for pneumococcal infection; vulnerable children older than 5 years should continue to receive the adult 23-valent vaccine.[22]

Although antibody titers wane,[30] routine revaccination is not recommended. Apart from local reactions, however, revaccination after 5 or more years is safe[31] and should be considered for adults who are at the highest risk for pneumococcal infection. The safety of the pneumococcal vaccine during pregnancy has not been evaluated; women who are at high risk for pneumococcal infection should be vaccinated before they become pregnant. The pneumococcal and influenza vaccines may be administered simultaneously at different sites when both are indicated.

Because of vaccine failures, additional strategies have been devised to protect vulnerable patients from pneumococcal infection. Long-term prophylaxis with orally administered low-dose penicillin protects children with sickle cell anemia from acquiring pneumococcal septicemia. It has also been recommended for young children with anatomic or functional asplenia. For asplenic adults, many physicians prefer vaccination with pneumococcal, meningococcal, and *Haemophilus influenzae* type b vaccines; these vaccines can be administered simultaneously. Vulnerable patients should be instructed to seek medical care at the first sign of infection and should be provided with ampicillin for self-administration if medical attention is not immediately available.

Additional work to enhance the efficacy of pneumococcal vaccines is needed. The development of a conjugated polysaccharide-protein vaccine for adults is particularly promising.

Streptococci

The streptococci, a large and diverse group of organisms widely distributed in nature, are part of the normal human flora. They vary in their pathogenic potential from many harmless species to a few very significant pathogens, such as *S. pneumoniae* and *S. pyogenes*.

Streptococci are gram-positive, round to ovoid cocci that can appear in pairs but more characteristically grow in chains of varying length. They are fastidious organisms requiring many nutrients. Although most are facultative anaerobes, some are obligate anaerobes. All species are nonmotile and nonsporulating and lack the enzyme catalase.

When grown on blood agar plates, most streptococci form small (1 to 2 mm), round, nonpigmented colonies. Streptococci are often characterized by changes in the appearance of the blood agar surrounding their colonies. Three types of reaction may occur: α-, β-, and γ-hemolysis. In α-hemolysis, green discoloration results from reduction of red blood cell hemoglobin and not from true hemolysis. In β-hemolysis, lysis of the red blood cells produces clearing of the blood agar. In the γ reaction, there is no change in the agar. Hemolysis alone, however, cannot be used for classification, because biologically dissimilar species can cause identical hemolytic reactions.

A far superior classification is the serologic system. In 1933, Rebecca Lancefield demonstrated that an antigenic carbohydrate could be extracted from the cell wall of streptococci. On the basis of chemical composition and immunologic reactivity of this carbohydrate, hemolytic streptococci can be divided into 18 groups (A through H and K through T). The species within each group tend to be biologically similar and to have a similar human pathogenic potential [*see Table 4*]. There are also separately classified nongroupable and anaerobic streptococci.

GROUP A STREPTOCOCCI

The group A streptococcus, *S. pyogenes*, ranks as one of the most important human pathogens. This organism can be recognized in the laboratory by the generous zone of β-hemolysis produced by most isolates. Up to 4% of group A streptococci, however, may be nonhemolytic while still retaining their full pathogenic potential. Because many other streptococci can produce β-hemolysis, a confirmatory test is necessary to identify β-hemolytic

streptococci as group A. The traditional screening test is the bacitracin test; almost all group A streptococci are inhibited by very low concentrations of bacitracin. Definitive speciation of group A streptococci can now be accomplished rapidly through identification of the group A carbohydrate by immunologic techniques.

The group A streptococcus is structurally complex. The capsule consists of hyaluronic acid, which is similar to the hyaluronic acid of human connective tissue and is therefore nonantigenic. The cell wall is composed of group-specific carbohydrates, structural proteins, and mucopeptide. The group-specific carbohydrate is responsible for the Lancefield serologic grouping; in group A streptococci, this carbohydrate is a polymer of rhamnose and *N*-acetyl glucosamine. Three structural proteins are part of the cell wall. The most important is the M protein, which is a crucial marker of virulence because it impedes phagocytosis by polymorphonuclear leukocytes. Anti-M antibodies are protective because they opsonize the bacteria, thus promoting phagocytosis and killing. Among group A streptococci, there are more than 120 antigenically distinct M proteins and hence more than 120 M types. M typing of strains is useful for epidemiologic purposes. Clearly, a polyvalent M protein vaccine containing the most common M types would be necessary to provide protection in communities.

The third component of the group A cell wall is a mucopeptide composed of repeating units of *N*-acetyl glucosamine and *N*-acetyl muramic acid. This element can produce carditis in laboratory rabbits, but its role in human rheumatic fever is unknown. C5 peptidase is a cell-associated protein that cleaves the C5 component of complement, rendering it inactive. Streptolysin S is a cell-associated hemolysin that lyses red blood cells on contact. The final structural element is the lipid-protein protoplast membrane. This membrane has certain antigens that cross-react with antigens in human cardiac muscle. The importance of this phenomenon in the pathogenesis of rheumatic fever is uncertain.

Epidemiology of Group A Streptococcal Infections

The natural reservoirs of group A streptococci are the human pharynx and skin. Group A streptococci are transmitted by

Table 4 Medically Important Streptococci and Enterococci

Streptococcus *Group*	*Representative Species*	*Hemolysis*	*Characteristics*	*Human Habitat*	*Human Diseases*
A	*S. pyogenes*	β (rarely γ)	Bacitracin sensitive; produces many extracellular enzymes and toxins	Nasopharynx Skin Rectum	Many diseases (*S. pyogenes* is the major human pathogen)
B	*S. agalactiae*	β (α, γ)	May be bacitracin sensitive; hydrolyzes sodium hippurate	Nasopharynx Genitourinary tract	Neonatal sepsis and meningitis Puerperal infections Urinary tract infections Endocarditis
C	*S. equi*	β (α, γ)	May be bacitracin sensitive; produces extracellular enzymes similar to those made by group A	Nasopharynx Genitourinary tract GI tract Skin	Skin and wound infections Bacteremias Endocarditis Pharyngitis
G	*S. canis*	β (α, γ)	Similar to group C	Similar to group C	Similar to group C
D	*S. bovis* *S. equinus*	β, γ (α)	Heat resistant; bile resistant; NaCl sensitive; penicillin sensitive	GI tract Genitourinary tract	Urinary tract infections Bacteremia Endocarditis
F	*S. anginosus*	β (α, γ)	Tiny colonies on blood agar; growth enhanced by 10% CO_2	Oropharynx GI tract Genitourinary tract	Sinusitis Meningitis Brain abscess Pneumonia Endocarditis
H	*S. sanguis*	α (β, γ)	Bile resistant	Oropharynx GI tract	
K	*S. salivarius*	γ (α)	Can grow at 45° C	Oropharynx	
Nongroupable	Viridans streptococci, including *S. mitis*, *S. mutans*, and others	α	No group-specific carbohydrate antigen yet recognized	Oropharynx	Endocarditis
Anaerobic streptococci	Many, including the peptostreptococci	γ (β, α)	Obligate anaerobes; tiny colonies on blood agar; no group-specific carbohydrate	Oropharynx GI tract Genitourinary tract	Sinusitis Pneumonia Lung abscess Empyema Brain abscess Soft tissue infections Bone and joint infections
Enterococci	*Enterococcus faecalis* *E. faecium*	β, γ (α)	Heat resistant; bile resistant; 6.5% NaCl resistant; penicillin resistant	GI tract Genitourinary tract	Endocarditis Peritonitis Urinary tract infections Wound infections

droplets, either from asymptomatic nasopharyngeal carriers or from persons with symptomatic pharyngitis.[32] Both the carrier rate and the incidence of pharyngitis are highest in late winter and early spring, in temperate climates, and in school-aged children. The epidemiology of streptococcal pyoderma is quite different. It is most common in young children and has its peak in late summer and early fall. Group A streptococci can colonize normal skin and spread from person to person by direct contact. Nosocomial transmission can occur, resulting most often in postoperative wound infections or postpartum sepsis. On occasion, streptococcal infections can occur in epidemic fashion as a result of contaminated foods.

Several different M types of group A streptococci may be present in a community over the course of a year; thus, individuals (predominantly children from 5 to 10 years of age) may have three to four episodes of pharyngitis per year. In contrast, although 5% of adults may carry group A streptococci in their throats, symptomatic infection is uncommon, perhaps because of the acquistion of type-specific immunity from frequent pharyngeal infections earlier in life. The epidemiology of specific streptococcal infections varies widely, both geographically and temporally. Currently, scarlet fever and rheumatic fever are uncommon in the Western world, where they were very common as late as the 1940s. Interestingly, in northern climates, streptococcal pharyngitis has not decreased in prevalence over the past 100 years, even in developed countries. Around 1980, there was a dramatic increase in the incidence of severe invasive group A streptococcal infections such as necrotizing fasciitis, bacteremia, and the streptococcal toxic-shock syndrome (TSS); this higher incidence has persisted to the present time.[33]

Pathogenesis of Group A Streptococcal Infections

The ability of streptococci to adhere to mucosal and epidermal surfaces is important for the first step of pathogenesis, colonization. In the throat, colonization appears to be related to the elaboration of fibronectin-binding proteins (protein F). On the skin, group A streptococci adhere to CD44 on the surface of keratinocytes, largely through the bacteria's hyaluronic acid capsule. Adherence through these ligand pairs is sufficient to cause pharyngitis and impetigo, respectively. Although invasion of cell lines of mucosal and keratinocyte origin can occur in vitro, the role of cellular invasion in the pathogenesis of pharyngitis, impetigo, cellulitis, and erysipelas is unclear, because cellular and tissue destruction are either uncommon or are not sufficient to be clinically recognized. Still, an intracellular location could in part explain prolonged carriage and, in some cases, the failure of penicillin to eradicate infection. In necrotizing fasciitis, pneumonia, myositis, and other infections associated with bacteremia, direct invasion probably occurs. The organisms elaborate enzymes such as streptolysin O and S and nicotinamide adenine dinucleotide glycohydrolase (NADase), which can account for their characteristic ability to produce inflammation and tissue damage and to spread rapidly in tissues. Streptokinase catalyzes the conversion of plasminogen to plasmin and thus promotes fibrinolysis. Four distinct streptococcal deoxyribonuclease (DNase) enzymes have been identified; most streptococcal strains produce DNase B. Group A streptococci also produce hyaluronidase, which is known as spreading factor because of its ability to digest the hyaluronic acid of connective tissue. Except for streptolysin S, these enzymes are all antigenic. Although antibodies directed against these enzymes do not protect the patient against recurrent streptococcal infections, the tests for such antibodies can be used to obtain accurate serodiagnosis.

The second most common way in which group A streptococci cause disease is by elaboration of exotoxins, which are responsible for scarlet fever and streptococcal TSS. Three antigenically distinct pyrogenic exotoxins (formerly called erythrogenic toxins)—types A, B, and C—have been recognized for several decades, yet in the past 3 years, 12 additional pyrogenic exotoxins have been discovered. Elaboration of type A and type C pyrogenic exotoxin depends on lysogeny of the streptococci by a bacteriophage; on occasion, certain group C or group G streptococci or even staphylococci can produce a pyrogenic toxin. From the site of the streptococcal infection, which is usually in the pharynx, the toxin enters the circulation, producing the characteristic rash of scarlet fever through an unknown mechanism. In addition, these toxins function as superantigens that simultaneously attach to the major histocompatibility complex of antigen-presenting cells and to V beta regions of the T cell receptor, resulting in generation of inflammatory cytokines (tumor necrosis factor–α [TNF-α], interleukin-1 [IL-1], and IL-6), as well as the lymphokines TNF-β, gamma interferon, and IL-2. These cytokines play important roles in the pathogenesis of shock and organ failure. Antibodies to the toxins prevent development of a rash but do not protect against the underlying infection.

Finally, an immune response to antecedent infection with group A streptococci can produce nonsuppurative syndromes. Acute rheumatic fever (ARF) and acute glomerulonephritis begin 1 to 3 weeks after a group A streptococcal infection, usually when the infection is no longer active. Despite their common origins as poststreptococcal disorders, ARF and acute glomerulonephritis differ significantly in their epidemiologic and pathogenetic features.

Acute Group A Streptococcal Diseases

Upper respiratory tract infection Group A streptococci are the most important cause of bacterial pharyngitis and tonsillitis.[34] They also cause sinusitis, otitis, mastoiditis, cervical lymphadenitis, peritonsillar abscesses, and retropharyngeal abscesses. Rarely, streptococcal pharyngitis has been associated with streptococcal TSS [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*].

Pneumonia Streptococcal pneumonia is uncommon, accounting for less than 5% of all bacterial pneumonias. Although it is seen most frequently as a sequela of influenza, it can occur as a primary infection and may even produce epidemics in closed groups. Streptococcal pneumonia presents acutely as fever, chills, and productive cough. Its distinguishing features are pleuritic chest pain and rapidly progressive empyemas, both of which occur in approximately 60% of patients.

Lymphadenitis and lymphangitis Group A streptococci are frequently responsible for suppurative lymphadenitis, particularly in patients with streptococcal infections of the respiratory tract or skin. In patients with a variety of soft tissue infections, red streaks that extend proximally are suggestive of group A streptococcal infection and are harbingers of lymphadenitis and subsequent bacteremia.

Wound infections Group A streptococcal wound infections are characterized by early onset, often within 24 hours after surgery or injury. Patients present acutely with fever and sys-

temic toxicity. Frequently, the infected wound appears relatively benign. A modest amount of thin serosanguineous fluid is often present in the wound. Gram stain of a smear of the fluid reveals gram-positive cocci in chains but a relative absence of acute inflammatory cells. If the infection is caused by strains of group A streptococci that produce pyrogenic exotoxins, a diffuse generalized erythema may develop (surgical scarlet fever).

Bone and joint infections Group A streptococci are relatively common causes of infections of bones and joints, usually from bacteremic seeding.

Postpartum infections The incidence of puerperal fever (childbed fever) diminished markedly from the 1850s, when it was common, to the 1990s, but it has increased in frequency over the past decade. The onset is abrupt, most often within 24 to 48 hours after delivery. Fever, chills, systemic toxicity, abdominal or pelvic pain, and a serosanguineous and odorless vaginal discharge are common; without treatment, endometritis may rapidly progress to pelvic peritonitis and bacteremia. These infections have also been associated with streptococcal TSS. Group A streptococci are the classic cause of puerperal fever, but other organisms, including group B streptococci and other streptococci and gram-negative bacilli, can cause infections with similar clinical features.

Bacteremia The incidence of group A streptococcal bacteremia appears to be rising. It often develops from a primary infection of skin, soft tissue, or a wound.[35] The respiratory tract is the second most common source, but bacteremia after uncomplicated streptococcal pharyngitis is uncommon. In 18% of cases, no primary focus of infection can be identified. About 80% of patients have underlying problems, ranging from diabetes or malignancies to drug abuse. Osteomyelitis, septic arthritis, meningitis, and endocarditis[36] are among the potential complications; overall mortality is 15%. Bacteremia is present in 60% of patients with necrotizing fasciitis and streptococcal TSS. Although meningitis can result from group A streptococcal bacteremia, this rare infection is more often a sequela of neurosurgical conditions.[37]

Skin and soft tissue infections Group A streptococci are important causes of cutaneous infections, which can range from very superficial infections of the epidermis (e.g., impetigo) to deeper subcutaneous infections with systemic symptoms (e.g., erysipelas or cellulitis)[38] to life-threatening processes (e.g., necrotizing fasciitis and gangrenous myositis) that require surgery and antibiotic therapy[39] [*see 2:VII Fungal, Bacterial, and Viral Infections of the Skin*].

Necrotizing fasciitis The widespread publicity that has been given to necrotizing fasciitis (previously known as streptococcal gangrene) has fueled popular concern about invasive group A streptococcal infections. An estimated 10,000 to 15,000 cases of invasive group A streptococcal infections occur each year in the United States; of these, necrotizing fasciitis occurs in 5% to 10%, with a case-fatality rate of about 30%. Most group A streptococci that cause invasive disease produce pyrogenic (formerly erythrogenic) toxin, but the genetic heterogenicity of causative strains does not support a clonal basis for the resurgence of invasive streptococcal infections.

Group A streptococcal necrotizing fasciitis is usually community acquired and sporadic in nature.[39] It occurs in all age groups. Many patients have predisposing conditions, which may include chickenpox, trauma, diabetes, or alcoholism.[40,41] In roughly 50% of patients, necrotizing fasciitis begins at the site of cutaneous penetrating trauma such as burns, insect bites, sliver injuries, abrasions, chickenpox vesicles, or surgical incisions. In the remaining cases, necrotizing fasciitis begins at the site of nonpenetrating deep trauma such as hematoma, ankle sprain, tendon rupture, or muscle tear. Most likely, group A streptococci translocate from the pharynx to the site of injury via the bloodstream. Because this type of infection begins deep within fascia and muscle, cutaneous changes are initially less prominent than pain, swelling, and systemic toxicity. Without aggressive therapy, however, the necrosis spreads to the skin and deep tissues. Computed tomography or magnetic resonance imaging can help define the depth and extent of infection, but in patients with preexisting trauma, infection may be difficult to distinguish from trauma alone. In addition to antibiotics and meticulous metabolic and circulatory support, prompt and aggressive surgical debridement is required; amputation may be necessary. Bacteremia and toxic streptococcal syndrome often complicate necrotizing fasciitis and are adverse prognostic features.

Diseases Caused by Toxin Production

Streptococcal toxic-shock syndrome Since the first description of streptococcal TSS in 1987, the incidence of this disorder has remained relatively constant.[42] Like staphylococcal TSS [*see* Staphylococci, *below*], streptococcal TSS appears to be mediated by a toxin, which in the case of the streptococcal form is probably a pyrogenic exotoxin. Like the staphylococcal toxin, the streptococcal toxin enters the circulation and functions as a superantigen, stimulating the release of host proteins that appear to mediate the shock syndrome.

The primary focus of infection is most often a soft tissue infection; respiratory infections are the next most common focus. Hypotension occurs in all patients and is often severe. Clinical features can include a generalized erythematous rash (10% of cases) that may undergo desquamation; acute respiratory distress syndrome (60% of cases); renal failure (80% of cases); and soft tissue necrosis, such as necrotizing fasciitis or myositis. Laboratory evidence of multiorgan involvement typically can be found and characteristically includes evidence of renal impairment, hepatic abnormalities, and disseminated intravascular coagulation, though clinical evidence of coagulopathy is rarely present.[43]

Management issues are complex and are reviewed in detail elsewhere,[43] but aggressive fluid replacement is crucial; with it, hypotension resolves in 50% of patients. Patients with refractory hypotension may require replacement of albumin and correction of hypocalcemia. Brisk hemolysis can occur, particularly in the 60% of patients who have bacteremia, and may necessitate transfusion. Surgical consultation and aggressive surgical debridement of devitalized tissue are also crucial. Dialysis is necessary in patients who experience renal failure; ventilatory support is necessary in acute respiratory distress syndrome. Mortality in most streptococcal TSS series ranges from 30% to 70%. Intravenous immunoglobulin looked promising in initial case reports and in one observational study, but a subsequent double-blind study in Scandinavia did not demonstrate significant reduction in mortality or attenuation of necrotizing fasciitis, though the trial was cut short because of low enrollment.[44] Earlier diagnosis, aggressive fluid resuscitation, and appropriate antibiotics may be responsible for the somewhat lower mortalities reported in more recent series.[43]

Scarlet fever The incidence of scarlet fever has declined sharply in the antibiotic era. The initial symptoms are fever and sore throat. Within 1 to 5 days, the characteristic fine, red, sandpaperlike eruption appears, often beginning on the chest and rapidly spreading to other parts of the body. Although the tongue and buccal mucosa are classically involved, the perioral area may be spared, thus accounting for the typical circumoral pallor. The rash is caused by hyperemia and capillary damage produced by pyrogenic toxin. In areas of trauma, such as the antecubital fossae, punctate hemorrhages (Pastia sign) may occur. Nausea and vomiting may be present, and fever and prostration may be severe. Desquamation of skin and mucous membranes is prominent during healing; one characteristic feature is the strawberry tongue. Therapy is the same as that for the underlying streptococcal infection.

Delayed Diseases Caused by the Host Immune Response

Group A streptococci can produce disease by a mechanism that is rarely a feature of other infections—a host immune response that subsequently produces tissue damage. ARF and acute glomerulonephritis are the major syndromes produced in this fashion; both are characterized by a latent period between the streptococcal infection and its inflammatory consequences. Erythema nodosum is a third poststreptococcal syndrome, but it is less specific and less serious than ARF and acute glomerulonephritis.

Acute rheumatic fever ARF is strictly a sequela of streptococcal pharyngitis. Although many strains of group A streptococci can provoke ARF, the syndrome most often follows pharyngeal infection by certain M protein types (especially M5, M18, and M3), which are heavily encapsulated, mucoid organisms. In about half the cases, the antecedent pharyngitis is clinically silent, but all patients with ARF have serologic evidence of a recent group A streptococcal infection. The risk of ARF after untreated streptococcal pharyngitis is less than 3%; the latent period between the pharyngitis and the onset of ARF averages 18 days.

Although common in developing nations, ARF is now quite uncommon in industrialized nations. However, the occurrence of several outbreaks in the United States in the 1980s demonstrates that ARF can still be an important problem in affluent societies. The disease occurs most often in children from 5 to 15 years of age; adults may present with atypical features, including synovitis without carditis.

Despite intensive study, the pathogenesis of ARF is unclear. The most widely held theory proposes that a genetically susceptible host develops an autoimmune response to epitopes in the organism that are cross-reactive with epitopes in tissues of the heart, joints, skin, or CNS.[45] Alternatively, streptococcal toxins or immune complexes may create alterations in tissue antigens that in turn provoke an autoimmune response that damages host tissues.

There are five major clinical manifestations of ARF: carditis, polyarthritis, chorea, subcutaneous nodules, and erythema marginatum. Carditis develops in about 60% of patients; it involves the endocardium, myocardium, and pericardium. The typical manifestations of rheumatic carditis are sinus tachycardia (sometimes with first-degree heart block), mitral regurgitation, a pericardial friction rub, and cardiomegaly; congestive heart failure indicates severe carditis. Although most cases of carditis resolve within 3 months, patients with moderate to severe carditis or recurrent ARF are at risk for the late manifestation of mitral valve or aortic valve scarring [*see 1:XI Valvular Heart Disease*]. Polyarthritis develops in about 70% of patients with ARF. It is characteristically a migratory arthritis involving the large joints of the extremities; it resolves without sequelae in days to weeks.[46] On rare occasions, adults may develop a persistent arthropathy of the hands and feet. Chorea, subcutaneous nodules, and erythema marginatum are all self-limited; each occurs in fewer than 10% of children with ARF and only very rarely in adults.

Table 5 Revised Jones Criteria for the Diagnosis of Acute Rheumatic Fever[47]

Major manifestations	Carditis Polyarthritis Chorea Erythema marginatum Subcutaneous nodules
Minor manifestations	Clinical Previous rheumatic fever or rheumatic heart disease Arthralgia Fever Laboratory Elevated acute-phase reactants Erythrocyte sedimentation rate C-reactive protein level Prolonged PR interval
Supporting evidence of streptococcal infection	Increased titer of antistreptococcal antibodies (e.g., anti–streptolysin O) Positive throat culture for group A *Streptococcus*

Note: The presence of two major manifestations or of one major and two minor manifestations indicates a high probability of acute rheumatic fever, if there is supporting evidence of group A streptococcal infection.

The minor manifestations of ARF are fever, arthralgias, and inflammation. The inflammation in such cases is evidenced by elevated erythrocyte sedimentation rates and C-reactive protein levels.

The diagnosis of ARF is made on the basis of clinical features. The classic Jones criteria include the presence of either two major manifestations or one major and two minor manifestations, as well as laboratory evidence of a recent streptococcal infection (e.g., a positive throat culture or rising antistreptococcal antibody levels) [*see Table 5*].[47]

Treatment of ARF focuses on eradication of streptococcal pharyngitis and reduction of inflammation [*see Table 6*]. Most clinicians recommend a course of penicillin or, as an alternative, other antistreptococcal antibiotics, even if the throat culture is negative at the time ARF is diagnosed. Anti-inflammatory therapy includes aspirin and bed rest until inflammatory symptoms resolve; corticosteroids may have a role for patients with severe carditis.

ARF can be prevented by adequate treatment of streptococcal pharyngitis, even if antibiotics are delayed for up to 9 days after the onset of pharyngitis. Patients with a history of ARF are particularly vulnerable to recurrent attacks. Consequently, they should receive continuous prophylaxis for at least 5 years with daily oral penicillin (250 mg twice a day) or monthly injections of 1.2 million units of benzathine penicillin G. Oral sulfadiazine or erythromycin may be administered to patients allergic to penicillin. Prophylaxis can be discontinued when young patients who are at low risk for recurrence reach adulthood or when small children are no longer in the household.

Table 6 Drug Treatment for Acute Rheumatic Fever[93]

Drug	*Dosage*	*Comments*
Penicillin	Acute treatment: Benzathine penicillin G, 1.2 million units I.M. Penicillin V,* 500 mg p.o., b.i.d.–t.i.d. Prophylaxis: Penicillin VK, 250 mg p.o., b.i.d. Benzathine penicillin G, 1.2 million units I.M. q. 4 wk	Acute treatment is given for patients with acute rheumatic fever who have a positive throat culture for group A *Streptococcus;* benzathine penicillin can be given if compliance is uncertain; prophylaxis for patients with history of acute rheumatic fever is begun immediately after the acute episode and continued until the patient reaches young adulthood
Sulfadiazine	1 g p.o., q.d	For prophylaxis in patients allergic to penicillin
Erythromycin	Acute treatment: erythromycin ethylsuccinate, 250 mg q.i.d. Prophylaxis: 250 mg p.o., b.i.d.	Acute treatment or prophylaxis for patients allergic to penicillin
Aspirin	90–100 mg/kg/day for 2 wk, then 60–70 mg/kg/day for 6 wk	Anti-inflammatory agents do not alter the development of chronic heart disease but reduce symptoms; blood salicylate levels of 20 mg/100 ml are desirable
Prednisone	40–60 mg q.d. for 2 wk, then taper off over 3 wk	Prednisone or a nonsteroidal anti-inflammatory drug can be used when symptoms are not controlled with aspirin alone

*Phenoxymethyl penicillin.

Acute glomerulonephritis Acute poststreptococcal glomerulonephritis appears to result from the deposition of circulating antigen-antibody complexes and complement in renal glomeruli [*see 10:V Glomerular Diseases*]. Only group A streptococci of certain M protein types can produce glomerulonephritis; it is unknown why only approximately 15 of the more than 80 M types are nephritogenic. Acute glomerulonephritis can follow either streptococcal pharyngitis or pyoderma but is more common after skin infections.

The incidence of acute glomerulonephritis after infection with nephritogenic streptococci varies but may reach 10% to 15% in certain epidemic situations. The prognosis is generally good, especially in children, and recurrences are uncommon; therefore, penicillin prophylaxis is not indicated. In fact, it is not certain that penicillin therapy for streptococcal pharyngitis or pyoderma affects the incidence of subsequent acute glomerulonephritis.

Erythema nodosum The painful pretibial nodules of erythema nodosum can develop after streptococcal infections of the skin, pharynx, or other sites. However, this nonsuppurative sequela may be associated with other infectious processes, such as tuberculosis or systemic mycoses, and with noninfectious processes, such as inflammatory bowel disease or hypersensitivity reactions.

Diagnosis of Group A Streptococcal Infections

The diagnosis of group A streptococcal infections depends on recognizing the clinical syndromes and culturing the bacterium from appropriate specimens. Streptococcal pharyngitis can be diagnosed rapidly by detecting streptococcal antigens on throat swab specimens, but results are dependent on the type of rapid test and the expertise of the laboratory technician [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*]. Rapid tests have incomplete sensitivity, so if a rapid test is negative, a throat culture should be obtained. Serologic tests may be extremely useful, albeit retrospectively, in documenting recent streptococcal infection. Many of the enzymes and toxins produced by group A streptococci are antigenic, and a variety of antibody tests are available. The most widely used is the anti–streptolysin O titer, which is elevated after most respiratory tract infections; anti-DNase B titers are also elevated after most group A streptococcal infections [*see Table 7*].

Treatment of Group A Streptococcal Infections

Despite the widespread use of penicillin, group A streptococci continue to be uniformly sensitive to this agent, which remains the drug of choice. The dose schedule and duration vary enormously. The standard therapy for pharyngitis is 10 days of low-dose oral penicillin (or a single I.M. dose of benzathine penicillin G), but a large trial suggested that 5 days of therapy with various

Table 7 Laboratory Tests for Streptococcal Pharyngitis[94]

Test	*Sensitivity (%)*	*Specificity (%)*	*Comments*
Rapid tests	70–90	95	Results are operator dependent; only qualified personnel should perform the test
Anti-streptolysin O titer	80	90	Elevation may also occur after infection with group C or G streptococci; a single titer > 200 Todd units/ml is significant; a rise in titer of > 2 dilutions between acute and convalescent sera is significant; responses are good after throat infections and poor after skin infections
Anti-DNase B titers	80	95	Consistent titer rises after throat or skin infection
Anti-hyaluronidase titers	80	95	Consistent titer rises after throat or skin infection

DNase—deoxyribonuclease

β-lactams or macrolides may be as effective.[48] Similarly, although oral penicillin is usually administered four times a day, a meta-analysis found that twice-daily dosing is as effective for children.[49] In contrast, more invasive streptococcal infections require much more intensive treatment, such as prolonged high-dose I.V. therapy for osteomyelitis or endocarditis. In patients who are allergic to penicillins, cephalosporins are effective but must be used with caution because of the risk of allergic cross-reactivity. Erythromycin, vancomycin, and clindamycin are generally excellent alternatives, although occasionally, clinical isolates are resistant to these agents. Erythromycin-resistant strains have become prevalent, especially in Japan and Finland and, more recently, in the eastern United States. Reducing the use of macrolide antibiotics can control this problem. Many strains of group A streptococci are now resistant to tetracyclines and TMP-SMX. The newer fluoroquinolones levofloxacin, sparfloxacin, gatifloxacin, and moxifloxacin are active against group A streptococci, but these agents are not indicated in children. In patients with necrotizing fasciitis and streptococcal TSS, clindamycin's ability to suppress streptococcal toxins may make it preferable to penicillin.

INFECTIONS CAUSED BY NON–GROUP A STREPTOCOCCI

Streptococci belonging to all the Lancefield groups can cause disease ranging in severity from urinary tract infections to meningitis, bacteremia, and endocarditis [*see Table 4*]. None of the non–group A organisms, however, have been implicated in rheumatic fever.

Groups C and G Streptococci

Groups C and G streptococci have many similarities to group A streptococci, although they are much less commonly implicated as pathogens. Most strains of groups C and G are β-hemolytic, and some are bacitracin sensitive. Bacteria in both groups can produce many of the same enzymes and toxins produced by group A streptococci, including streptolysin O, streptokinase, NADase, DNase, hyaluronidase, and pyrogenic toxin. The group-specific carbohydrate of group C is similar to that of group A, and groups A, C, and G all have cross-reacting cell membrane antigens.

First recognized as animal pathogens, groups C and G streptococci may be part of the resident flora of the pharynx, skin, genitourinary tract, and gastrointestinal tract.[50] Group C streptococci can cause pharyngitis in adults. Groups C and G streptococci can cause skin and wound infections, puerperal sepsis, bacteremia, endocarditis, pericarditis, septic arthritis, meningitis, scarlatiniform eruptions, myositis, and TSS.[51] In rare instances, groups C and G streptococci cause poststreptococcal glomerulonephritis. Groups C and G streptococci are sensitive to penicillin, which is the drug of choice, and to erythromycin, vancomycin, and other antibiotics.

Group B Streptococci

Group B streptococci are important human pathogens.[52] Most group B streptococci are β-hemolytic; up to 20% are sensitive to low-dose bacitracin. They can be presumptively separated from group A streptococci on the basis of their ability to hydrolyze sodium hippurate, but definitive identification depends on immunologic identification of the group B carbohydrate.

Group B streptococcal infections are the leading cause of bacterial disease and death in newborns; maternal infections can also occur in the peripartum period.[52] The organisms reside in the vagina and infect neonates during passage through the birth canal; premature labor, prolonged membrane rupture, and intrapartum fever are risk factors for neonatal infection. The intrapartum administration of antibiotics to women at high risk can prevent neonatal infection; screening during the third trimester of pregnancy and treatment with penicillin have substantially reduced the incidence of perinatal and peripartum group B streptococcal infections.[53]

Group B streptococci can cause a broad range of infections in nonpregnant adults. Persons of any age can be affected, but elderly patients are most vulnerable.[54] Other risk factors include genitourinary disorders, diabetes, cancer, HIV infection, and peripheral vascular disease. In addition to bacteremia, infections caused by group B streptococci include urinary tract infections, septic arthritis, endocarditis, meningitis, pulmonary infections, and infections of the skin, soft tissue, and wounds. Necrotizing fasciitis and streptococcal TSS[55] have also been reported.

Although group B streptococci are sensitive to clinically achievable levels of penicillin, the MICs are somewhat higher than those for group A streptococci. The great majority of group B streptococci are sensitive to ampicillin, cephalosporins, vancomycin, erythromycin, and clindamycin but are resistant to tetracycline.

Group D Streptococci

Group D streptococci (*S. bovis* and *S. equinus*) can be identified in the laboratory by their ability to resist heat and bile but not 6.5% NaCl broth. Both species reside in the human GI and genitourinary tracts. *S. equinus* rarely causes human disease, but *S. bovis* is a relatively common cause of endocarditis, which may be more severe than the subacute bacterial endocarditis caused by viridans streptococci.[56] *S. bovis* bacteremia is significantly linked to carcinoma of the colon; therefore, the GI tract of any patient with this infection should be aggressively evaluated. *S. bovis* is sensitive to penicillin and other antibiotics and responds well to therapy with penicillin alone.

Nongroupable Streptococci

Up to 30% of streptococcal isolates are not groupable under the currently available Lancefield antisera classification. These organisms are capable of producing a great variety of pyogenic infections, among which subacute bacterial endocarditis stands out as the most important. Nongroupable streptococci can also cause bacteremia and pneumonia, especially in immunosuppressed patients[57]; meningitis is rare.[58] Members of the *S. milleri* group have a propensity to produce abscesses, particularly in the brain or liver.[59] Classically, subacute endocarditis is caused by viridans streptococci—nongroupable, α-hemolytic bacteria that are part of the normal flora of the oropharynx [*see 7:XVIII Infective Endocarditis*]. Many species of viridans streptococci can cause endocarditis; *S. sanguis*, *S. mutans*, *S. mitis*, and members of the *S. milleri* group are commonly responsible. Infections caused by members of the *S. milleri* group may be more destructive than those caused by the other species. About 10% of viridans streptococci are nutritionally variant, requiring supplementary pyridoxal to grow on agar plates. Up to one third of these organisms require serum penicillin levels of greater than 0.1 μg/ml to inhibit growth. Although this level is higher than the levels that kill most other streptococci, this concentration can be easily achieved clinically. Synergistic penicillin-aminoglycoside therapy has been recommended for endocarditis caused by these organisms. Penicillin-tolerant and penicillin-resistant strains of viridans streptococci have been

increasingly recognized; vancomycin or penicillin-aminoglycoside therapy has been recommended for endocarditis caused by penicillin-resistant streptococci.

Anaerobic Streptococci

Various anaerobic and microaerophilic streptococci reside in the oropharynx, the GI tract, and the genitourinary tract. These organisms cause a great variety of infections, including aspiration pneumonia, lung abscess, empyema, sinusitis, brain abscess, bone and joint infections, and skin and wound infections. Anaerobic streptococci often participate in these processes, along with other anaerobic or aerobic bacteria. Foul-smelling pus is characteristic of these infections, and gas may be present in soft tissues. The anaerobic streptococci are penicillin sensitive.

Enterococci

Although they were long considered to be streptococci, the enterococci have been reclassified into their own genus, *Enterococcus*. Of the infections caused by enterococci, 80% to 85% are caused by *E. faecalis;* most of the others are caused by *E. faecium,* but *E. durans, E. avium,* and other species can also cause disease in humans. Morphologically indistinguishable from streptococci and immunologically similar to members of group D streptococci, the enterococci are metabolically unique in their ability to resist heat, bile, and 6.5% NaCl broth. Unlike streptococci, enterococci are uniformly penicillin resistant. Both major enterococcal species reside in the GI and genitourinary tracts, and both can cause urinary, abdominal, and disseminated infections, including endocarditis on preexisting valvular lesions.

PATHOGENESIS OF ENTEROCOCCAL INFECTIONS

The primary reason enterococci have emerged as major human pathogens is that these organisms are resistant to many antibiotics. Enterococci are intrinsically resistant to penicillin because they have unique penicillin-binding proteins that permit cell wall synthesis to proceed even in the presence of β-lactam antibiotics. In addition, enterococci that produce β-lactamase have been recognized; although these strains are still relatively uncommon, they may pose problems in the future. Enterococci can also demonstrate low-level resistance to aminoglycosides, but synergistic therapy with a penicillin and an aminoglycoside has been successful against most species. However, high-level resistance to the aminoglycosides streptomycin and gentamicin has emerged during the past 20 years. More recently, vancomycin-resistant enterococci (VRE) have been recognized as nosocomial pathogens.[60] VRE colonization of the GI tract and skin often precedes overt infections. Colonization and infection are usually hospital-acquired; they are most common in patients who have prolonged hospitalizations or serious underlying diseases or who have undergone previous therapy with multiple antibiotics. Antibiotics that are active against anaerobes are particularly likely to promote intestinal colonization.[61] VRE are often carried by medical personnel and can spread from patient to patient. Because they often occur in very ill patients and are so difficult to treat, VRE infections pose a major threat. The mortality associated with VRE bacteremia approaches 45%.[59]

CLINICAL ENTEROCOCCAL INFECTIONS

Enterococci, alone or with other enteric organisms, are relatively common causes of urinary tract infections, wound infections, and peritonitis and intra-abdominal abscesses. Enterococci have become an increasingly prominent cause of bacteremia,[62] which usually originates from a focus in the urinary tract or abdomen; the incidence of nosocomial bacteremias caused by these organisms is also increasing, particularly in patients who have received cephalosporins or other broad-spectrum antibiotics. Enterococcal endocarditis may affect normal, diseased, or prosthetic valves and may pursue an acute destructive course. Enterococcal meningitis is much less common but may be very difficult to treat. Enterococcal infections are most common in persons with underlying genitourinary or GI disease, in the elderly, and in debilitated persons. Enterococci have become an important cause of disease in hospitalized patients; they are now the second most common nosocomial pathogen in the United States, occurring less commonly than *Escherichia coli* but more commonly than *Pseudomonas aeruginosa* and *Staphylococcus aureus*.

TREATMENT OF ENTEROCOCCAL INFECTIONS

Changing resistance patterns will necessitate changes in antibiotic therapy for patients with enterococcal infections [*see Table 8*]. Traditionally, ampicillin has been the drug of choice for enterococcal urinary tract infections, largely because of the high concentrations of drug excreted in the urine. In penicillin-allergic patients, nitrofurantoin, a fluoroquinolone, or TMP-SMX may be effective for the treatment of simple urinary tract infections; however, because in vitro susceptibility testing of enterococci is often misleading and because resistance to these agents may rapidly emerge, patients should be continually evaluated through assessment of their clinical response and by use of follow-up cultures.

Deep tissue and bloodstream infections have traditionally been treated with ampicillin and gentamicin, with vancomycin, or with vancomycin and gentamicin. Although these agents are still appropriate for most enterococcal infections, all clinically significant isolates should be subjected to testing for β-lactamase production, high-level aminoglycoside resistance, and vancomycin resistance to determine whether an alternative therapy is necessary. Infections caused by enterococci that produce β-lactamase can be treated with an antimicrobial agent that combines a penicillin with a β-lactamase inhibitor. Infections caused by strains that are highly resistant to aminoglycosides can be treated with vancomycin. Infections caused by vancomycin-resistant strains are the most difficult to treat; linezolid has been particularly helpful, and quinupristin-dalfopristin is also effective.[63] Doxycycline, chloramphenicol, and, for urinary tract infections, nitrofurantoin have also been used. In all cases, removal of infected indwelling catheters and foreign bodies and drainage of septic foci are essential. It is also essential to limit the spread of VRE by intensive surveillance, infection-control measures,[64] and prudent use of vancomycin and other antibiotics.

Staphylococci

Staphylococci are nonsporulating, nonmotile gram-positive cocci that have an average diameter of 1 μm. Microscopically, staphylococci tend to be larger and rounder than streptococci. Because cell division occurs on three planes, these organisms are typically found in grapelike clusters and tetrads, as well as in pairs and sometimes in short chains. When grown on blood agar, staphylococci form small (1 to 2 mm), smooth, round colonies that are often pigmented and may be surrounded by a zone of β-hemolysis.

Table 8 Antibiotic Treatment for Enterococcal Infections[95]

Drug	*Dose*	*Comments*
Ampicillin + gentamicin	Ampicillin: 2 g I.V. q. 4 hr Gentamicin: 1 mg/kg I.V. q. 8 hr	Indicated in patients with endocarditis or meningitis*
Vancomycin + gentamicin	Vancomycin: 1 g I.V. q. 12 hr Gentamicin: 1 mg/kg I.V. q. 8 hr	For penicillin-allergic patients*
Ampicillin-sulbactam	2 g I.V. q. 4 hr	For infections caused by enterococci that produce β-lactamase; these appear to be rare
Vancomycin	1 g I.V. q. 12 hr	For infection caused by strains resistant to aminoglycosides
Linezolid	600 mg q. 12 hr I.V. or p.o.	For infection caused by vancomycin-resistant strains
Quinupristin-dalfopristin	7.5 mg/kg I.V. q. 8 hr	For infection caused by vancomycin-resistant strains; not effective against *E. faecalis*
Doxycycline	100–200 mg I.V. or p.o. q. 12 hr	For infection caused by vancomycin-resistant strains; experience very limited
Chloramphenicol	500 mg I.V. or p.o. q. 6 hr	For infection caused by strains resistant to vancomycin, aminoglycosides, and penicillin
Nitrofurantoin	100 mg p.o. q. 12 hr	For urinary tract infections caused by vancomycin-resistant strains

*If high-level resistance to gentamicin (500–2,000 μl) is detected, perform susceptibility testing to streptomycin; substitute streptomycin if sensitive.

Staphylococci are very hardy organisms and can withstand much more physical and chemical stress than pneumococci and streptococci. For example, staphylococci resist drying, withstand 10% NaCl broth, and will survive and even replicate at temperatures between 10° and 45° C. Because staphylococci are facultative anaerobes, they will grow in the presence or absence of oxygen. Staphylococci are catalase positive.

Of the species of staphylococci, *S. aureus* is by far the most important human pathogen. *S. aureus* can be tentatively identified in the laboratory on the basis of its production of a golden-yellow pigment, which may not be apparent until after 24 hours of growth. Because some strains are nonpigmented, definitive identification of *S. aureus* depends on its ability to produce the enzyme coagulase and to ferment mannitol. The most important coagulase-negative organism is *S. epidermidis,* which is universally present as part of the normal skin flora but can produce disease in certain circumstances, such as infection of an indwelling prosthesis. Another important coagulase-negative species is *S. saprophyticus,* which causes urinary tract infections.

STAPHYLOCOCCUS AUREUS

The cellular structure of *S. aureus* is complex.[65] Most strains have polysaccharide microcapsules.

The cell wall of *S. aureus* is structurally similar to that of group A streptococci: both have a carbohydrate antigen, a protein component, and a mucopeptide. The carbohydrate antigen is a teichoic acid, which in *S. aureus* is a polymer of *N*-acetylglucosamine and polyribitol phosphate. Antibodies to teichoic acid can be detected in normal human serum, and elevated antibody titers are present in patients with deep-seated staphylococcal infections. Teichoic acid has no established role in virulence, and antibodies to this carbohydrate are not protective. The protein component of the cell wall includes protein A, which reacts with IgG of normal human serum. Protein A interacts with the Fc component rather than the Fab component of IgG and hence is not a true antigen. Protein A may be antiphagocytic, but its role in virulence has not been clearly established. The cell wall mucopeptide of staphylococci is structurally similar to the mucopeptide of other gram-positive bacteria.

Epidemiology of Staphylococcal Infections

S. aureus is an extremely widespread organism; it can be cultured from environmental surfaces, clothing, bedding, and the air, although transmission of staphylococci by airborne droplets or fomites does not appear to be a major route of infection. Humans—either asymptomatic carriers or persons with staphylococcal lesions—are the reservoir of infection. Nasal carriage is common; rates of nasal carriage range from 15% to more than 50% and are increased in hospitalized patients, dialysis patients, intravenous drug users, and insulin-injecting diabetic patients. Other relatively common sites of colonization include the rectum, the perineum, and the pharynx. *S. aureus* can often be cultured from normal skin, but skin colonization is probably brief, with repeated repopulation occurring from the nose. It is unclear why some persons are prolonged carriers, others are transient carriers, and others resist colonization and why some persons carry many organisms, whereas others have low colony counts. Person-to-person transmission is more likely from nasal carriers of large numbers of organisms and from persons with active staphylococcal infections of the skin or respiratory tract. Nasal carriage increases the risk of bacteremia and of skin and wound infections, including recurrent furunculosis[66]; upper respiratory tract viral infections appear to increase bacterial shedding by *S. aureus* nasal carriers. Periodic application of mupirocin ointment in nasal carriers reduces the rate of carriage and lowers the risk of skin infections.

Minor staphylococcal infections are extremely common. It is the rare person who has not at some time experienced a staphylococcal furuncle (boil), paronychia, or hordeolum (sty). Serious staphylococcal infections generally require a predisposing insult to host defenses. Most often, this takes the form of skin disease, trauma, or a viral infection of the respiratory tract, especially influenza or measles. Other predisposing factors include foreign bodies, liver disease, neoplasia, diabetes, renal failure, defects in leukocyte or immunoglobulin function, narcotics addiction, and broad-spectrum antibiotic therapy.

Unlike the incidence of infections from other gram-positive cocci, the incidence of serious staphylococcal infection increased sharply after the introduction of antibiotics. Much of this increase can be attributed to the development of antibiotic resis-

tance. When penicillin was first introduced, fewer than 10% of staphylococci were penicillin resistant; this percentage has increased steadily and now includes the great majority of staphylococci. During the late 1950s, staphylococcal infection reached epidemic proportions, and hospital-acquired infections were a particularly grave problem. With the introduction of methicillin in 1959 and other penicillinase-resistant antibiotics in the 1960s, the incidence of nosocomial staphylococcal infection temporarily declined. Unfortunately, however, methicillin-resistant staphylococci have slowly but progressively emerged as major nosocomial pathogens in many parts of the world. In addition, in recent years, community-acquired methicillin-resistant *S. aureus* (MRSA) infections have increased in prevalence in many regions of the United States, Japan, and Southeast Asia. Vancomycin-intermediate and vancomycin-resistant strains have appeared in both Japan and the United States.[67]

Pathogenesis of Staphylococcal Infections

The earliest tissue response in staphylococcal infection is acute inflammation with a vigorous exudation of polymorphonuclear leukocytes. Vascular thrombosis and tissue necrosis quickly lead to abscess formation. As a result of the development of a fibrin meshwork and, later, fibroblast proliferation, these abscesses become walled-off zones of loculated infection and tissue destruction, with dying leukocytes and viable bacteria at the center. Fibrosis and scarring are often prominent in healing.

Most strains of *S. aureus* produce a variety of extracellular products, including both enzymes and toxins, that may account for the tendency to produce burrowing, destructive, localized infections. The enzyme coagulase causes plasma to clot, thus promoting the fibrin meshwork that contributes to abscess formation. Staphylococci can also produce lipase, protease, hyaluronidase, and DNase, which can add to tissue damage. Another important enzyme is penicillinase. Because penicillinase has no role in pathogenicity, staphylococci that produce penicillinase are no more virulent than non–penicillinase-producing strains. Nevertheless, this enzyme is clinically and epidemiologically important because it hydrolyzes the β-lactam ring of penicillin, thereby inactivating the molecule. The production of penicillinase is controlled by plasmids, or episomes, which are extrachromosomal DNA molecules that replicate during cell division. Unlike the R factors of gram-negative bacilli, however, the plasmids responsible for penicillinase production do not usually mediate resistance to multiple antibiotics.

Of even greater interest are the nonenzymatic toxins produced by *S. aureus*. α-Toxin is a cytotoxin that produces pores in cell membranes, thereby altering their permeability and resulting in cell damage or death. α-Toxin damages red and white blood cells and activates platelets. Injection of α-toxin into animals can produce dermal necrosis and contraction of vascular smooth muscle, leading to tissue ischemia. Another potential virulence factor is leukocidin, which consists of two leukotoxic proteins that are capable of disrupting lysosomal membranes. Occasionally, strains of staphylococci produce exfoliatin, which causes the epidermolysis characteristic of staphylococcal scalded skin syndrome (SSSS). Some strains of staphylococci can also produce one of four antigenically distinct enterotoxins that cause the vomiting and diarrhea characteristic of staphylococcal food poisoning. In rare instances, staphylococci produce an erythrogenic toxin that causes scarlet fever. Finally, a staphylococcal exotoxin, toxic-shock syndrome toxin–1 (TSST-1), appears to be responsible for menstrually related staphylococcal TSS. Interestingly, staphylococcal TSS associated with nasal packing and other surgical wound infections is mediated by one of the enterotoxins, most commonly enterotoxin-B.

Host resistance and immunity to staphylococci are poorly understood. The importance of the granulocyte is supported by the susceptibility to staphylococcal infections seen in patients with neutropenia or various disorders of neutrophil function, such as chronic granulomatous disease, Chédiak-Higashi syndrome, and various disorders of chemotaxis, such as the lazy leukocyte syndrome. The most important factors predisposing to staphylococcal infections are not immunologic defects but mechanical defects. Minute skin abrasions, for example, probably provide the portal of entry in staphylococcal skin infections and in many cases of staphylococcal bacteremia. I.V. drug abuse accounts for many cases of staphylococcal bacteremia and endocarditis. Indwelling venous catheters are particularly important in nosocomial infections; plastic catheters become coated with fibrinogen and fibrin, which interact with adhesins on the bacterial cell surface and bind staphylococci to the catheter.

Clinical Staphylococcal Infections

Skin and soft tissue infections The most frequent manifestations of staphylococcal disease are skin and soft tissue infections. These range from processes that can cause great discomfort but are rarely hazardous, such as impetigo, folliculitis, furuncles, carbuncles, and paronychia, to much more serious deep tissue infections, such as cellulitis and wound sepsis [*see 2:VII Fungal, Bacterial, and Viral Infections of the Skin*].

Staphylococcal skin infections are epidemiologically significant in the transmission of serious staphylococcal disease. Even localized infections can give rise to bacteremia and endocarditis, and furuncles of the nose and face may occasionally produce CNS infection such as cavernous vein thrombophlebitis by spreading along venous channels that lack valves.

Bone and joint infections Staphylococci are among the principal causes of septic arthritis and osteomyelitis of the long bones and of vertebral bodies and disk spaces [*see 7:XV Septic Arthritis and 7:XVI Osteomyelitis*]. These infections may result from hematogenous seeding of proximal and distal long bones and joint spaces or contiguous spread of infection; infections of orthopedic prostheses are particularly serious.[68] In addition to causing acute osteomyelitis, staphylococci have a propensity for causing chronic bone infections, usually of the metaphysis of long bones, in which active localized infection can persist for many decades without undergoing dissemination. *S. aureus* is also the leading cause of septic bursitis.

Respiratory tract infections Although staphylococci reside in the nasopharynx, they do not cause pharyngitis; they are rarely involved in acute otitis, sinusitis, or mastoiditis, but they can participate in chronic infections of these regions [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*].

S. aureus accounts for fewer than 10% of all cases of bacterial pneumonia, but it is nevertheless an important pathogen of the lower respiratory tract. Staphylococcal pneumonia can occur either by spread from the upper airways or by hematogenous seeding. Except in infants, airborne staphylococcal pneumonia is rare as a primary infection. It usually occurs after viral respiratory tract infections, especially influenza. Thus, clinical illness often follows a biphasic course: viral symptoms lasting 4 or 5 days precede an abrupt deterioration as staphylococcal pneumonia pro-

gresses. Staphylococcal pneumonia also occurs as a nosocomial infection, particularly in patients on ventilators and in elderly, debilitated patients[69]; in such patients, mortality is 32%. *S. aureus* can cause pneumonia in HIV-infected persons; these infections are often community acquired and have a mortality of 38%, even with appropriate antibiotic therapy.

Patients with staphylococcal pneumonia are acutely ill; purulent sputum production is the rule, and polymorphonuclear leukocytes and staphylococci can be identified on Gram stain of the sputum. Radiographs show a characteristic patchy bronchopneumonic pattern, often with multiple areas of consolidation. Because these organisms tend to produce tissue necrosis, cavitation is relatively common, and lung abscesses and empyemas may develop; fibrosis and scarring often occur during healing. Hematogenous staphylococcal pneumonia is a complication of staphylococcal bacteremia, especially when endocarditis of the tricuspid valve leads to septic embolization. Sputum production tends to occur later and is less prominent than in airborne pneumonias. Typical radiographic features include multiple small areas of infiltration, which, although sometimes evanescent, often progress to cavitation. Staphylococcal pneumonia is a life-threatening disease requiring high-dose parenteral antibiotic therapy, often for 2 to 4 weeks. Patients with abscess formation and empyema generally require therapy for 3 to 4 weeks; empyemas must be drained.

Bloodstream infections Most patients with *S. aureus* bacteremia are acutely ill; more than 50% have temperatures in excess of 40° C (104° F), and most experience chills and exhibit systemic toxicity. Serious underlying diseases are present in 75% to 85% of patients. The overall mortality is about 23%.[70] Up to 80% of patients with staphylococcal bacteremia acquire the infection in the hospital, often from infected I.V. catheters, skin and wound infections, or pulmonary tract infections.[71] I.V. drug abuse is responsible for more than 50% of the community-acquired cases. Patients with community-acquired bacteremias are more likely to have endocarditis and secondary metastatic infections than patients with nosocomial infections, who are more likely to have an evident portal of entry and severe underlying diseases. Bacteremic seeding can lead to secondary staphylococcal infection of the lungs, bones and joints, and the genitourinary tract. Evidence of CNS involvement suggests endocarditis and is an adverse prognostic sign.

Patients with staphylococcal bacteremia have generally been treated with parenteral antibiotics for 4 to 6 weeks. This recommendation was made on the basis of the results of a study conducted from 1940 to 1954 in which endocarditis developed in 64% of 55 patients with staphylococcal bacteremia.[72] Subsequent investigations suggested that patients with a removable or treatable primary focus of infection (most often, infected I.V. devices) who have no clinical evidence of endocarditis can be treated safely with only 10 to 21 days of I.V. antibiotics.[73] Other studies, however, reported development of endocarditis in 22% to 38% of patients with a primary focus of bacteremia. Therefore, the possibility of endocarditis should not be ignored, even if a primary focus is present. Moreover, major complications, including shock, acute respiratory distress syndrome, and metastatic infection, can occur even in the absence of endocarditis.

Controlled trials are necessary to determine the safety and efficacy of short-term therapy for staphylococcal bacteremia. It may therefore be prudent to treat patients with staphylococcal bacteremia as though they have endocarditis, unless all of the following features are present: clear evidence of only a transient bacteremia; a removable primary focus of infection; a benign clinical course; an absence of echocardiographically demonstrable valvular abnormalities and metastatic infection; and intact host defenses. Negative titers of teichoic acid antibodies would bolster the decision to shorten the I.V. antibiotic therapy, which should, in any case, be continued for at least 10 to 15 days. A regimen of I.V. antibiotics followed by oral antibiotic therapy has been suggested, but more studies are needed before this regimen can be widely recommended. Clearly, the physician must individualize therapy by balancing the morbidity and expense of 4 to 6 weeks of antibiotic therapy against the risks of undertreating endocarditis, which is a life-threatening infection.[74] Transesophageal echocardiography can establish a diagnosis of *S. aureus* endocarditis[75] [*see 7:XVIII Infective Endocarditis*].

Because of the high mortality associated with staphylococcal bacteremia and endocarditis, combination therapies utilizing nafcillin or vancomycin with gentamicin or rifampin are being studied. Thus far, combination therapy appears to reduce the duration of bacteremia, but it does not change the long-term mortality. Combination treatment may also be useful in patients who fail to respond to conventional single-drug treatment. I.V. drug abusers with uncomplicated right-sided endocarditis may respond well to abbreviated therapy.

Because of the propensity of staphylococci to form abscesses, it is particularly important to evaluate bacteremic patients for a loculated infection that may require surgical drainage. Management of staphylococcal bacteremia in patients with Hickman catheters almost always requires catheter removal in addition to antibiotic therapy. Staphylococcal endocarditis appears to be increasing in frequency; its clinical features and therapy are reviewed elsewhere[76] [*see 7:XVIII Infective Endocarditis*].

CNS infection Staphylococcal infections of the CNS present most often as aseptic meningitis in patients with staphylococcal bacteremia, and especially those with endocarditis; the CSF typically reveals modest pleocytosis, and cultures and Gram stains are negative. *S. aureus* is an uncommon cause of purulent meningitis. Direct extension of infection from traumatic or neurosurgical wounds or from osteomyelitis of the mastoids or other cranial bones is a more frequent cause of staphylococcal meningitis than hematogenous seeding. Nafcillin has been shown to penetrate the CSF and should be administered in dosages of approximately 2 g I.V. every 4 hours for the average-sized adult with normal renal function. When nafcillin cannot be used because the infecting organisms are methicillin-resistant[77] or because the patient is allergic to penicillin, therapy must be individualized. Options such as high-dose I.V. vancomycin and rifampin should be considered. In patients who do not respond favorably, there may be a role for intrathecal vancomycin or bacitracin; or for I.V. TMP-SMX, erythromycin, or chloramphenicol.

Another uncommon infection of the CNS in which *S. aureus* is a major pathogen is spinal epidural abscess. Patients with this disease often have underlying vertebral osteomyelitis. These patients present with fever and back pain; radicular pain, weakness, and paralysis evolve as cord compression occurs. Immediate surgical decompression is mandatory.

GI tract infections Staphylococci can also affect the GI tract. Staphylococcal food poisoning occurs when food handlers who have contaminated superficial wounds or who are shedding in-

fected nasal droplets inoculate foods with enterotoxin-producing strains of *S. aureus*. If the contaminated food is not refrigerated, the organisms produce enterotoxin within 4 to 6 hours. The symptoms begin abruptly within 1 to 6 hours after the ingestion of preformed enterotoxin. They consist of salivation, nausea, vomiting, abdominal pain, diarrhea, and prostration. Nausea and vomiting are the most prominent features, because of the enterotoxin's ability to affect the vomiting center of the brain. The symptoms usually subside within 24 hours; there is no specific therapy other than fluid and electrolyte repletion, although I.V. hydration may be necessary in cases of severe disease. In the 1950s and 1960s, GI superinfection or enterocolitis occurred in patients receiving tetracycline or chloramphenicol treatment. Recently, similar cases have occurred, caused by methicillin-resistant strains.

Genitourinary infections *S. aureus* can occasionally produce genitourinary infections. From 7% to 27% of patients with staphylococcal bacteremia have positive urine cultures. Most of these patients have no gross renal infection, but in some, hematogenous pyelonephritis, renal cortical abscesses, renal carbuncles, perinephric abscesses, or prostatic abscesses may develop. These processes should be considered in all patients with positive urine cultures and in patients with staphylococcal bacteremia and persistent fever, flank pain, or abnormal urine sediments. I.V. pyelography, nephrotomography, renal ultrasonography, and CT may be diagnostically helpful. Prostatic abscesses may require repeated rectal examination or ultrasound for diagnosis. *S. aureus*, either alone or with other genitourinary pathogens, can also cause nosocomial urinary tract infections in patients with indwelling urethral catheters.

Other staphylococcal infections *S. aureus* can produce infection of sebaceous glands of the eyelid (sty), purulent conjunctivitis, and orbital cellulitis. It can cause parotitis, which occurs most often in patients who are elderly, debilitated, and dehydrated. Decades ago, parotitis was most common in patients with renal failure, whereas today, it is most common in patients with Sjögren syndrome. Mastitis is most frequent in nursing mothers in the early postpartum period but may occur in neonates or in nonlactating women. Infection of muscle is rare in temperate climates but common in the tropics. Tropical pyomyositis presents as fever and multiple muscle abscesses. The diagnosis should be suspected in recent immigrants from tropical areas but should also be considered in patients who have not traveled and in patients with AIDS. CT may be helpful in detecting muscle abscesses. Purulent pericarditis can result either from direct extension of a pneumonia or empyema or from hematogenous seeding. Staphylococcal bacteremia can produce metastatic abscesses of any organ, including the liver and spleen. Patients who have metastatic abscesses may have persistent fever; in such patients, diagnosis may be difficult if localizing symptoms are absent. *S. aureus* can cause peritonitis in patients on continuous ambulatory peritoneal dialysis.

Treatment of Staphylococcal Infections

It is extremely important to establish a microbiologic diagnosis in staphylococcal infections and to utilize cultures and sensitivities to direct therapy. Choice of empirical antibiotics and of the route of administration depend on the seriousness of the infection, the trends in staphylococcal resistance patterns in the specific geographical area, and the risk factors for methicillin resistance in a given patient.

In addition to paronychia and sties, minor staphylococcal infections of the skin, such as folliculitis and furunculosis, generally respond well to the topical application of warm soaks. Larger focal infections such as carbuncles may have to be incised and drained. Antibiotics should be added when fever or systemic symptoms are present, when lesions are large or numerous, when lesions fail to respond to local therapy, when patients have underlying medical problems such as valvular heart disease or a cardiac prosthesis, or when the nose or face is involved. Cloxacillin or dicloxacillin in an oral dosage of 250 to 500 mg every 6 hours is generally sufficient. Cephalexin, erythromycin, or the newer macrolides are excellent alternatives, as is clindamycin in an oral dosage of 150 to 300 mg every 6 hours [*see Table 9*].

For serious staphylococcal infections, such as cellulitis, deep wound sepsis, pneumonia, septic arthritis, osteomyelitis, bacteremia, endocarditis, or meningitis, parenteral antibiotics are mandatory. Oxacillin or nafcillin in a dosage of 1 g every 4 hours may suffice for cellulitis or pneumonia [*see 7:XIV Chemotherapy of Infection*], but a dosage of 2 g every 4 hours should be used for osteomyelitis, endocarditis, or CNS infections in the average-sized adult with normal renal function. Parenteral cephalosporins such as cephalothin and cefuroxime are excellent alternatives unless CNS infection is present. For patients who are allergic to penicillin, erythromycin and clindamycin have produced good results in the treatment of cellulitis, pneumonia, and osteomyelitis; vancomycin is preferred for the treatment of endocarditis in these patients because it is a bactericidal antibiotic.

Although only about 10% of all *S. aureus* strains are penicillin sensitive, penicillin G remains the drug of choice for sensitive organisms. Care must be taken in sensitivity testing, however, because penicillinase is an inducible enzyme. It is best to reserve penicillin G for staphylococci that are shown to be penicillinase negative or for staphylococci that have extremely large zones of inhibition by penicillin on Kirby-Bauer sensitivity testing. Methicillin resistance may be overlooked occasionally on routine sensitivity testing because it is more readily expressed at lower temperatures (30° C), at higher salt concentrations, or after 48 hours of incubation.

Synergism between penicillins and aminoglycosides against *S. aureus* has been demonstrated in vitro and in experimental staphylococcal infections. However, a clinical benefit of combination therapy for *S. aureus* endocarditis has not yet been demonstrated. Most staphylococci are extremely sensitive to rifampin, which may be combined with other antistaphylococcal drugs in certain difficult clinical situations. However, staphylococci rapidly develop resistance to rifampin when the drug is used alone, and resistance may even arise during combination therapy. The utility of ciprofloxacin is limited by the emergence of resistance. Only time will tell if the newer fluoroquinolones, such as levofloxacin and sparfloxacin, will be more successful; combinations of rifampin and a fluoroquinolone are being investigated.

MRSA Since its first appearance in 1970, MRSA has been a growing problem. These organisms produce a cell-wall penicillin-binding protein with a low affinity for β-lactam antibiotics. Initially, MRSA was a nosocomial pathogen in university hospitals. Although hospitals are still the most common setting, these organisms have also become important problems in long-term care facilities and are increasingly being acquired in the community.[78] MRSA has proved to be difficult to control; infected patients should be subjected to strict barrier precautions.[79] Both pa-

Table 9 Antibiotic Treatment for Staphylococcal Infections

Route	*Drug*	*Dosage*	*Comments*
Oral	Cloxacillin or dicloxacillin	250–500 mg q. 6 hr	Excellent efficacy
	Erythromycin	250 mg q. 6 hr	Alternative for penicillin-allergic patients with cellulitis, pneumonia, or osteomyelitis
	Clindamycin	150–300 mg q. 6 hr	Alternative for penicillin-allergic patients with cellulitis, pneumonia, or osteomyelitis
	Rifampin *plus* ciprofloxacin	600 mg q.d. 250 mg b.i.d.	May be useful in right-sided endocarditis; utilized because of noncompliance in I.V. drug abusers
Oral or intravenous	Linezolid	600 mg q. 12 hr	For MRSA; may be more suitable than other anti-MRSA agents for staphylococcal TSS; for vancomycin-resistant infection
Intravenous	Vancomycin	1 g q. 12 hr	Alternative for penicillin-allergic patients with endocarditis and for MRSA
	Oxacillin or nafcillin	For cellulitis or pneumonia: 1 g q. 4 hr For osteomyelitis, endocarditis, CNS infections: 2 g q. 4 hr	Favored choice for non-MRSA strains
	Cephalosporins Cephalothin Cefuroxime	 2 g I.V. q. 6 hr 1,500 mg I.V. q. 8 hr	Alternative agent unless CNS infection is present
	Penicillin G	2–4 million units q. 4–6 hr	First choice for sensitive organisms
	Daptomycin	4 mg/kg/day	For MRSA, except in endocarditis; for vancomycin-resistant infection; dose-related myopathy
	Quinupristin-dalfopristin	7.5 mg/kg q. 8 hr	For MRSA; can cause phlebitis and myopathy; requires central line placement

CNS—central nervous system MRSA—methicillin-resistant *Staphylococcus aureus* TSS—toxic-shock syndrome

tients and hospital personnel may become asymptomatic carriers. The carrier state is typically prolonged, often exceeding 3 years; antibiotics are usually ineffective. Topical mupirocin ointment can reduce the MRSA carrier rate,[80] but because recolonization is common, mupirocin is not recommended for extended use in long-term care facilities.

The virulence and clinical manifestations of MRSA are no different from those of methicillin-susceptible *S. aureus*[81]; compared to methicillin-sensitive strains, however, a higher percentage of MRSA strains possess the toxins TSST-1, enterotoxins, and the Panton-Valentine leukocidin. MRSA strains are also resistant to all β-lactam antibiotics, erythromycin, and chloramphenicol; even if the organisms appear to be sensitive to cephalosporins in disk diffusion testing, one should not rely on these agents. In the past, vancomycin has been the only alternative for treating MRSA. Interestingly, vancomycin is less effective than nafcillin for strains sensitive to both agents. Linezolid, daptomycin, teicoplanin, and quinupristin-dalfopristin are newer antibiotics with activity against MRSA. Daptomycin is approved for skin and soft tissue infection, but because it is rapidly bactericidal, it needs to be studied as a treatment for endocarditis caused by MRSA. Linezolid has been approved for skin and soft tissue infections, as well as for pneumonia caused by MRSA; evidence is accumulating that, like clindamycin, it is a potent suppressor of staphylococcal toxin production. Because of this characteristic, it may be a more suitable agent to use in patients with staphylococcal TSS. Quinupristin-dalfopristin often causes phlebitis and myopathy. Thus, its use requires placement of a central line.

Vancomycin resistance With the increased use of vancomycin, strains of *S. aureus* with reduced susceptibility to vancomycin have begun to appear.[67] Although these organisms are still very uncommon in the United States, vigorous control measures are required to prevent them from joining VRE as major nosocomial pathogens.[82] Daptomycin and linezolid have excellent activity against vancomycin-intermediate and vancomycin-resistant *Staphylococcus.*

Adjunctive measures In addition to vigorous antibiotic therapy, other measures are often necessary to treat staphylococcal infections. In particular, it is important to remove indwelling venous catheters or other foreign bodies that may be a portal of entry for staphylococcal bacteremia or a nidus for persistent infection. Patients must be evaluated for the presence of staphylococcal abscesses, which often must be drained. Finally, endocarditis and metastatic infection are a concern in all patients with staphylococcal bacteremia.

Prevention Epidemiologic control of staphylococcal infection requires ongoing surveillance and the reporting of infections. Contact precautions should be followed in the management of patients with active infections of the skin or wounds. Respiratory precautions may be useful in the management of patients with staphylococcal pneumonia, although droplet spread is much less important than transmission by direct contact. The treatment of nasal or rectal carriers can be frustrating, particularly if the carriers are hospital personnel or persons suffering from recurrent furunculosis. Topical treatment with germicidal soaps,

povidone-iodine, or antibiotic ointments has been advocated, but long-term results have been disappointing. Orally administered antibiotics, including rifampin, TMP-SMX, and ciprofloxacin, have also failed to live up to initially promising findings. Bacterial interference, which attempts to replace epidemiologically virulent strains of staphylococci with strains that have been deliberately colonized and are less virulent, has generally been abandoned, in part because infection has been caused by these supposedly less virulent strains.

Attempts to develop staphylococcal vaccines are continuing.[83] In a clinical trial in dialysis patients, vaccination with a staphylococcal surface carbohydrate conjugated to *Pseudomonas* exotoxin A significantly reduced the incidence of bacteremia, though protective antibodies lasted only 8 months.[84]

Intranasal therapy with mupirocin ointment may help control the staphylococcal carrier state. A naturally occurring antibiotic produced by *P. fluorescens,* mupirocin inhibits *S. aureus* (including penicillin- and methicillin-resistant strains) and other aerobic gram-positive cocci by binding reversibly to bacterial isoleucyl transfer RNA synthetase; no cross-resistance has been observed between mupirocin and other antibiotics. A placebo-controlled trial of mupirocin in 34 patients who were *S. aureus* carriers found that a monthly course of nasal mupirocin reduced the incidence of nasal colonization and skin infections for at least 1 year. Mupirocin was administered intranasally twice daily for 5 days; side effects were minimal.[85] Additional studies have confirmed the benefits of mupirocin. However, because resistance to mupirocin and recolonization after therapy can occur, routine use of mupirocin should be avoided.[80] The FDA has approved mupirocin for the topical treatment of impetigo.

Staphylococcal Toxic-Shock Syndrome

Staphylococcal TSS was first reported in 1978; by 1990, more than 3,300 cases had been reported in the United States, 90% of which occurred during menstruation in women who were using tampons. The incidence of staphylococcal TSS declined precipitously after superabsorbent tampons were withdrawn from the market. Currently, fewer than 100 cases occur each year, and nonmenstrual cases occur more often than those associated with menstruation. Most cases are now nosocomially acquired, often as a result of postoperative staphylococcal wound infections, particularly those associated with nasal packing after rhinoplasty. Staphylococcal TSS caused by toxin-producing strains of MRSA has been reported from the United States and Japan.

TSS is a multisystem disease with diverse clinical manifestations [*see Table 10*]. Leukocytosis and thrombocytopenia (< 100,000 platelets/mm^3) are common findings. Urinalysis may show mild pyuria and, occasionally, microscopic hematuria. Blood urea nitrogen and creatinine levels are elevated in more than 50% of patients. Serum bilirubin and hepatic enzyme levels are elevated in about half of patients. Serum creatine kinase levels are high in more than one third of patients, and myoglobinuria has developed in some patients. Elevated serum amylase levels are also found, but such elevations may be related to azotemia rather than to clinically evident pancreatitis. Unexplained marked hypocalcemia is often observed. The drop in serum calcium level is out of proportion to the degree of hypoalbuminuria noted in some patients and may be caused by elevated serum calcitonin levels. Blood cultures show no growth in almost all cases. Group A streptococci can produce a severe form of TSS that resembles staphylococcal TSS; in streptococcal TSS, however, bacteremia is common, necrotizing fasciitis may be present, and mortality is much higher (30% to 70%) [*see* Streptococcal Toxic-Shock Syndrome, *above*].

Treatment The management of staphylococcal TSS calls for immediate correction of hypotension and shock with vigorous fluid replacement, attention to the site of *S. aureus* colonization or infection (e.g., removal of tampon and drainage of any abscess), and systemic antimicrobial therapy with an antistaphylococcal agent. Albumin replacement may be needed to counter capillary leak; in addition, correction of hypocalcemia, use of renal dialysis, and ventilator support may be necessary. Because of the emergence of staphylococcal TSS caused by MRSA, empirical antibiotic therapy should be with vancomycin, daptomycin, or linezolid [*see Table 11*]. Protein synthesis inhibitors such as clindamycin and linezolid have been shown to suppress TSST-1 production, so linezolid is a reasonable choice. In contrast, cell-wall–active agents cause release of TSST-1. A retrospective analysis of 45 patients suggested that glucocorticoids may assist in recovery,[86] but more data are needed before glucocorticoids can be recommended for all patients with staphylococcal TSS. Most patients recover in 1 to 2 weeks; mortality is about 5%.

Table 10 Clinical Manifestations of Staphylococcal Toxic-Shock Syndrome

High fever
Scarlatiniform eruption
Hypotension and shock
Desquamation, particularly of palms and soles, during convalescence
MANIFESTATIONS OF SPECIFIC ORGAN INVOLVEMENT
Mucous membranes: hyperemia of conjunctivae, pharynx, and vagina
Gastrointestinal tract: vomiting and diarrhea at onset
Muscle: severe myalgias (elevated serum creatine kinase level; rhabdomyolysis in some patients)
Central nervous system: toxic encephalopathy (disorientation, delirium, or obtundation in absence of high fever or shock)
Kidney: azotemia; pyuria in absence of demonstrable urinary tract infection
Liver: elevations of serum bilirubin and aspartate aminotransferase
Blood: thrombocytopenia

COAGULASE-NEGATIVE STAPHYLOCOCCI

S. epidermidis and other coagulase-negative staphylococcal species are organisms of low pathogenic potential that are universally present as part of the normal skin flora. Because of their ubiquity, coagulase-negative staphylococci are frequently isolated from clinical specimens, including blood cultures. Most often, they are skin contaminants rather than true pathogens; unfortunately, neither clonal variability[87] nor molecular typing[88] can reliably distinguish between contaminants and pathogens. In certain circumstances, however, these organisms can cause significant disease. In the great majority of cases, coagulase-negative staphylococci are opportunistic pathogens, producing infection only in patients with significant underlying medical problems.[89] They most often produce infections in patients with indwelling I.V. catheters or implanted prosthetic devices. Many strains produce an exopolysaccharide that functions as a capsule.[90] This slime layer is an adhesive that binds the organism to implanted

Table 11 Antibiotic Treatment for Staphylococcal Toxic-Shock Syndrome*

Drug	*Dosage*	*Comments*
Vancomycin	1 g I.V. q. 12 hr	Consider if patient has risk factors for MRSA
Daptomycin	4 mg/kg/day	Rapidly bactericidal
Linezolid	600 mg I.V. q. 12 hr	Suppresses toxin synthesis; has good activity against MRSA
Clindamycin	600–800 mg I.V. q. 8 hr	Suppresses toxin synthesis; inducible clindamycin resistance has been described in community-acquired MRSA
Nafcillin	2 g I.V. q. 4 hr	For non-MRSA infection

*No comprehensive studies of antibiotic treatment in staphylococcal toxic shock syndrome have been done. In general, clindamycin and linezolid are favored, because of their ability to suppress toxin synthesis; conversely, cell-wall–active agents cause release of toxin.

devices; it also inhibits the action of antibiotics, further contributing to pathogenicity.

Coagulase-negative staphylococci are now the leading cause of nosocomial bacteremia, which most often results from the use of I.V. catheters, especially centrally placed catheters. Patients with this infection may have signs of phlebitis at the needle or catheter site and may present with persistent low-grade fevers or high-spiking temperature elevations. The I.V. needle or catheter should be removed and a course of parenteral antibiotics administered. Metastatic infection is uncommon but can be serious. These organisms can also be an important cause of bacteremia in immunosuppressed patients. The mortality in patients with coagulase-negative staphylococcal bacteremia approaches 40%, in part because these patients have such serious underlying diseases; mortality attributable to the infection itself is about 13%.

An even more serious problem occurs when *S. epidermidis* or other coagulase-negative staphylococci infect indwelling ventriculoatrial shunts. Patients with this condition may present with bacteremia, meningitis, or both. In addition to high-dose antibiotic therapy, management must often include shunt removal or revision. Vascular grafts and joint prostheses may also become infected. These infections are generally indolent and are more likely to produce pain and joint dysfunction than fever and local inflammatory signs. Diagnosis may be difficult because coagulase-negative staphylococci can be recovered from joint aspirates, either as pathogens or contaminants. These organisms can sometimes cause indolent wound infections or osteomyelitis. Here, too, diagnosis may be difficult, but visualization of the staphylococci on Gram stain, their consistent isolation on culture, and the absence of other pathogens should suggest their etiologic role.

The most serious therapeutic problem caused by coagulase-negative staphylococci is prosthetic valve endocarditis. The disease typically has a subacute course, but eradication of the organisms is often very difficult because of antibiotic resistance. Medical therapy with high-dose antibiotics should be attempted, but valve replacement is necessary if infection persists or if significant valve dysfunction occurs. Mortality is high, approaching 50% with or without surgical therapy. Coagulase-negative staphylococci may also cause endocarditis on native valves.[89]

Antibiotic therapy for deep-seated coagulase-negative staphylococcal infections is difficult. Fifty percent of strains are resistant to methicillin and other semisynthetic penicillins. Whereas most strains appear to be sensitive to cephalosporins on disk diffusion testing, these results correlate poorly with actual bactericidal activity. Vancomycin, gentamicin, and rifampin are bactericidal against most coagulase-negative staphylococci. Vancomycin or rifampin resistance is rare but may emerge during therapy. Vancomycin is generally the drug of choice, but for serious infections, the addition of rifampin or gentamicin should be strongly considered. Linezolid and quinupristin-dalfopristin have good activity against methicillin-resistant coagulase-negative staphylococci, and clinical trials have proved their efficacy in vivo.[91] Although these agents are good alternatives, additional studies are clearly needed.

S. saprophyticus is a coagulase-negative member of the normal skin flora that can be distinguished from *S. epidermidis* by its resistance to the antibiotic novobiocin. Previously dismissed as a nonpathogen, *S. saprophyticus* is now clearly recognized to be an important cause of urinary tract infections. *S. saprophyticus* is second only to *E. coli* as a cause of cystitis in sexually active young women, accounting for 10% to 15% of such cases.[92] *S. saprophyticus* is much less likely to produce infections in men; however, *S. saprophyticus* commonly infects elderly men who have undergone urinary tract instrumentation. The clinical features of *S. saprophyticus* infections resemble those of other urinary tract infections; bacteremia and systemic toxicity are rare, even when upper urinary tract infection occurs. *S. saprophyticus* is sensitive to most antimicrobial agents that are used to treat urinary tract infections.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Gillespie SH, Balakrishnan I: Pathogenesis of pneumococcal infection. J Med Microbiol 49:1057, 2000

2. Obaro SK, Monteil MA, Henderson DC: The pneumococcal problem. BMJ 312:1521, 1996

3. Nuroti JP, Butler JC, Farley MM, et al: Cigarette smoking and invasive pneumococcal disease. N Engl J Med 342:681, 2000

4. Kyaw MH, Christie P, Clarke SC, et al: Invasive pneumococcal disease in Scotland, 1999–2001: use of record linkage to explore associations between patients and disease in relation to future vaccination policy. Clin Infect Dis 37:1283, 2003

5. Feikin DR, Schuchat A, Kolczak M, et al: Mortality from invasive pneumococcal pneumonia in the era of antibiotic resistance, 1995–1997. Am J Public Health 90:223, 2000

6. Mufson MA, Stanek RJ: Bacteremic pneumococcal pneumonia in one American city: a 20-year longitudinal study, 1978–1997. Am J Med 107(1A):34S, 1999

7. Fiore AE, Moroney JF, Farley MM, et al: Clinical outcomes of meningitis caused by *Streptococcus pneumoniae* in the era of antibiotic resistance. Clin Infect Dis 30:71, 2000

8. Rubins JB, Cheung S, Carson P, et al: Identification of clinical risk factors for nosocomial pneumococcal bacteremia. Clin Infect Dis 29:178, 1999

9. O'Brien KL, Walters MI, Sellman J, et al: Severe pneumococcal pneumonia in previously healthy children: the role of preceding influenza infection. Clin Infect Dis 30:784, 2000

10. Osler W: The Principles and Practice of Medicine: Designed for the Use of Practitioners and Students of Medicine. D Appleton & Co, New York, 1892, p 517

11. Dowell SF, Garman RL, Liu G, et al: Evaluation of binax NOW, an assay for the detection of pneumococcal antigen in urine samples, performed among pediatric patients. Clin Infect Dis 32:824, 2001

12. Harrison LH, Dwyer DM, Billmann L, et al: Invasive pneumococcal infection in Baltimore, MD: implications for immunization policy. Arch Intern Med 160:89, 2000

13. Stevens DL, Rout RK, Rosen H, et al: Differential white blood cell count: underutilized or overrated? Curr Opin Infect Dis 13:205, 2000

14. Breiman RF, Keller DW, Phelan MA, et al: Evaluation of effectiveness of the 23-valent pneumococcal capsular polysaccharide vaccine for HIV-infected patients. Arch Intern Med 160:2633, 2000

15. Siegel M, Timpone J: Penicillin-resistant *Streptococcus pneumoniae* endocarditis: a case report and review. Clin Infect Dis 32:972, 2001

16. Tomasz A: New faces of an old pathogen: emergence and spread of multidrug-resistant *Streptococcus pneumoniae*. Am J Med 107(1A):55S, 1999

17. Whitney CG, Farley M, Hadler J, et al: Increasing prevalence of multidrug-resistant *Streptococcus pneumoniae* in the United States. N Engl J Med 343:1917, 2000

18. Metlay JP, Hofmann J, Cetron MS, et al: Impact of penicillin susceptibility on medical outcomes for adult patients with bacteremic pneumococcal pneumonia. Clin Infect Dis 30:520, 2000

19. Normark BH, Novak R, Ortqvist A, et al: Clinical isolates of *Streptococcus pneumoniae* that exhibit tolerance of vancomycin. Clin Infect Dis 32:552, 2001

20. Ho PL, Tse WS, Tsang KWT, et al: Risk factors for acquisition of levofloxacin-resistant *Streptococcus pneumoniae*: a case-control study. Clin Infect Dis 32:701, 2001

21. Plouffe JF: Emerging therapies for serious gram-positive bacterial infections: a focus on linezolid. Clin Infect Dis 31(suppl 4):S144, 2000

22. Ball P: Therapy for pneumococcal infection at the millennium: doubts and certainties. Am J Med 107(1A):77S, 1999

23. Ramirez JA, Bordon J: Early switch from intravenous to oral antibiotics in hospitalized patients with bacteremic community-acquired *Streptococcus pneumoniae* pneumonia. Arch Intern Med 161:848, 2001

24. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Preventing pneumococcal disease among infants and young children. MMWR Morb Mortal Wkly Rep 49:1, 2000

25. Blumberg EA, Brozena SC, Stutman P, et al: Immunogenicity of pneumococcal vaccine in heart transplant recipients. Clin Infect Dis 32:307, 2001

26. Recommendations of the Advisory Committee on Immunization Practices (ACIP): Prevention of pneumococcal disease. MMWR Morb Mortal Wkly Rep 46:1, 1997

27. Jackson LA, Neuzil KM, Yu O, et al: Effectiveness of pneumococcal polysaccharide vaccine in older adults. N Engl J Med 348:1747, 2003

28. Ament A, Baltussen R, Duru G, et al: Cost-effectiveness of pneumococcal vaccination of older people: a study of 5 Western European countries. Clin Infect Dis 31:444, 2000

29. Whitney CG, Farley MM, Hadler J, et al: Decline in invasive pneumococcal disease after the introduction of protein-conjugate vaccine. N Engl J Med 348:1737, 2003

30. Mufson MA: Antibody response of pneumococcal vaccine: need for booster dosing? Int J Antimicrob Agents 14:107, 2000

31. Jackson LA, Benson P, Sneller V-P, et al: Safety of revaccination with pneumococcal polysaccharide vaccine. JAMA 281:243, 1999

32. Carapetis JR, Currie BJ, Kaplan EL: Epidemiology and prevention of group A streptococcal infections: acute respiratory tract infections, skin infections, and their sequelae at the close of the twentieth century. Clin Infect Dis 28:205, 1999

33. O'Brien KL, Beall B, Barrett NL, et al: Epidemiology of invasive group A *Streptococcus* disease in the United States, 1995–1999 Clin Infect Dis 35:268, 2002

34. Bisno AL, Gerber MA, Gwaltney JM, et al: Diagnosis and management of group A streptococcal pharyngitis: a practice guideline. Clin Infect Dis 25:574, 1997

35. De Quirós JCLB, Moreno S, Cercenado E, et al: Group A streptococcal bacteremia: a 10-year prospective study. Medicine (Baltimore) 76:238, 1997

36. Baddour LM: Infective endocarditis caused by β-hemolytic streptococci. The Infectious Diseases Society of America's Emerging Infections Network. Clin Infect Dis 26:66, 1998

37. Sommer R, Rohner P, Garbino J, et al: Group A β-hemolytic *Streptococcus* meningitis: clinical and microbiological features of nine cases. Clin Infect Dis 29:929, 1999

38. Dupuy A, Benchikhi H, Roujeau J-C, et al: Risk factors for erysipelas of the leg (cellulitis): case-control study. BMJ 318:1591, 1999

39. Bisno AL, Stevens DL: Streptococcal infections of skin and soft tissues. N Engl J Med 334:240, 1996

40. Kaul R, McGeer A, Low DE, et al: Population-based surveillance for group A streptococcal necrotizing fasciitis: clinical features, prognostic indicators, and microbiologic analysis of seventy-seven cases. Am J Med 103:18, 1997

41. Zurawski CA, Bardsley M, Beall B, et al: Invasive group A streptococcal disease in metropolitan Atlanta: a population-based assessment. Clin Infect Dis 27:150, 1998

42. Stevens DL: Streptococcal toxic shock syndrome. Clin Microbiol Infect 8:133, 2002

43. Stevens DL: Group A streptococcal sepsis. Curr Infect Dis Rep 5:379, 2003

44. Darenberg OJ, Ihendyane N, Sjolin J, et al: Intravenous immunoglobulin G therapy in streptococcal toxic shock syndrome: a European randomized, double-blind, placebo-controlled trial. Clin Infect Dis 37:333, 2003

45. Stollerman GH: Rheumatic fever. Lancet 349:935, 1997

46. Williamson L, Bowness P, Mowat A, et al: Lesson of the week: difficulties in diagnosing acute rheumatic fever—arthritis may be short lived and carditis silent. BMJ 320:362, 2000

47. Guidelines for the diagnosis of rheumatic fever. Jones Criteria, 1992 update. Special Writing Group of the Committee on Rheumatic Fever, Endocarditis, and Kawasaki Disease of the Council on Cardiovascular Disease in the Young of the American Heart Association. JAMA 268:2069, 1992

48. Adam D, Scholz H, Helmerking M: Short-course antibiotic treatment of 4782 culture-proven cases of group A streptococcal tonsillopharyngitis and incidence of post-streptococcal sequelae. J Infect Dis 182:509, 2000

49. Lan AJ, Colford JM Jr: The impact of dosing frequency on the efficacy of 10-day penicillin or amoxicillin therapy for streptococcal tonsillopharyngitis: a meta-analysis. Pediatrics 105:E19, 2000

50. Eriksson BKG: Anal colonization of group G β-hemolytic streptococci in relapsing erysipelas of the lower extremity. Clin Infect Dis 29:1319, 1999

51. Hirose Y, Yagi K, Honda H, et al: Toxic shock-like syndrome caused by non–group A β-hemolytic streptococci. Arch Intern Med 157:1891, 1997

52. Krohn MA, Hillier SL, Baker CJ: Maternal peripartum complications associated with vaginal group B streptococci colonization. J Infect Dis 179:1410, 1999

53. Schrag SJ, Zywicki S, Farley MM, et al: Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. N Engl J Med 342:15, 2000

54. Henning KJ, Hall EL, Dwyer DM, et al: Invasive group B streptococcal disease in Maryland nursing home residents. J Infect Dis 183:1138, 2001

55. Gardam MA, Low DE, Saginur R, et al: Group B streptococcal necrotizing fasciitis and streptococcal toxic shock-like syndrome in adults. Arch Intern Med 158:1704, 1998

56. Kupferwasser I, Darius H, Muller AM, et al: Clinical and morphological characteristics in *Streptococcus bovis* endocarditis: a comparison with other causative microorganisms in 177 cases. Heart 80:276, 1998

57. Shenep JL: Viridans-group streptococcal infections in immunocompromised hosts. Int J Antimicrob Agents 14:129, 2000

58. Cabellos C, Viladrich PF, Corredoira J, et al: Steptococcal meningitis in adult patients: current epidemiology and clinical spectrum. Clin Infect Dis 28:1104, 1999

59. Bert F, Bariou-Lancelin M, Lambert-Zechovsky N: Clinical significance of bacteremia involving the "*Streptococcus milleri*" group: 51 cases and review. Clin Infect Dis 27:385, 1998

60. Murray BE: Vancomycin-resistant enterococcal infections. N Engl J Med 342:710, 2000

61. Donskey CJ, Chowdhry TK, Hecker MT, et al: Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. N Engl J Med 343:1925, 2000

62. Caballero-Granado FJ, Becerril B, Cuberos L, et al: Attributable mortality rate and duration of hospital stay associated with enterococcal bacteremia. Clin Infect Dis 32:587, 2001

63. Winston DJ, Emmanouilides C, Kroeber A, et al: Quinupristin/dalfopristin therapy for infections due to vancomycin-resistant *Enterococcus faecium*. Clin Infect Dis 30:790, 2000

64. Tenorio AL, Badri SM, Sahgal NB, et al: Effectiveness of gloves in the prevention of hand carriage of vancomycin-resistant *Enterococcus* species by health care workers after patient care. Clin Infect Dis 32:826, 2001

65. Lowy FD: *Staphylococcus aureus* infections. N Engl J Med 339:520, 1998

66. von Eiff C, Becker K, Machka K, et al: Nasal carriage as a source of *Staphylococcus aureus* bacteremia. N Engl J Med 344:11, 2001

67. Fridkin SK: Vancomycin-intermediate and -resistant *Staphylococcus aureus:* what the infectious disease specialist needs to know. Clin Infect Dis 32:108, 2001

68. Murdoch DR, Roberts SA, Fowler VG Jr, et al: Infection of orthopedic prostheses after *Staphylococcus aureus* bacteremia. Clin Infect Dis 32:647, 2001

69. Gonzalez C, Rubio M, Romero-Vivas J, et al: Bacteremic pneumonia due to *Staphylococcus aureus:* a comparison of disease caused by methicillin-resistant and methicillin-susceptible organisms. Clin Infect Dis 29:1171, 1999

70. Mylotte MJ, Tayara A: *Staphylococcus aureus* bacteremia: predictors of 30-day mortality in a large cohort. Clin Infect Dis 31:1170, 2000

71. Jensen AG, Wachmann CH, Poulsen KB, et al: Risk factors for hospital-acquired *Staphylococcus aureus* bacteremia. Arch Intern Med 159:1437, 1999

72. Wilson R, Hamburger M: Fifteen years' experience with *Staphylococcus* septicemia in a large city hospital: analysis of fifty-five cases in the Cincinnati General Hospital, 1940 to 1954. Am J Med 22:437, 1957

73. Nolan CM, Beaty HN: *Staphylococcus aureus* bacteremia: current clinical patterns. Am J Med 60:495, 1976

74. Rader BL, Wandall DA, Frimodt-Moller N, et al: Clinical features of *Staphylococcus aureus* endocarditis: a 10-year experience in Denmark. Arch Intern Med 159:462, 1999

75. Fowler VG Jr, Sanders LL, Kong LK, et al: Infective endocarditis due to *Staphylococcus aureus:* 59 prospectively identified cases with follow-up. Clin Infect Dis 28:106, 1999

76. Tornos P, Almirante B, Mirabet S, et al: Infective endocarditis due to *Staphylococcus aureus:* deleterious effect of anticoagulant therapy. Arch Intern Med 159:473, 1999

77. Lu C-H, Chang W-N: Adults with meningitis caused by oxacillin-resistant *Staphylococcus aureus*. Clin Infect Dis 31:723, 2000

78. Gorak EJ, Yamada SM, Brown JD: Community-acquired methicillin-resistant *Staphylococcus aureus* in hospitalized adults and children without known risk factors. Clin Infect Dis 29:797, 1999

79. Kotilainen P, Routamaa M, Peltonen R, et al: Eradication of methicillin-resistant *Staphylococcus aureus* from a health center ward and associated nursing home. Arch Intern Med 161:859, 2001

80. Laupland KB, Conly JM: Treatment of *Staphylococcus aureus* colonization and prophylaxis for infection with topical intranasal mupirocin: an evidence-based review. Clin Infect Dis 37:933, 2003

81. Stevens DL, Herr D, Lampiris H, et al: Linezolid versus vancomycin for the treatment of methicillin-resistant *Staphylococcus aureus* infections. Clin Infect Dis 34:1481, 2002

82. Hiramatsu K: Vancomycin-resistant *Staphylococcus aureus:* a new model of antibiotic resistance. Lancet Infect Dis 1:147, 2001

83. McKenney D, Poullot KL, Wang Y, et al: Broadly protective vaccine for *Staphylococcus aureus* based on an in vivo–expressed antigen. Science 284:1523, 1999

84. Shinefield H, Black S, Fattom A, et al: Use of a *Staphylococcus aureus* conjugate vaccine in patients receiving hemodialysis. N Engl J Med 346:491, 2002

85. Raz R, Miron D, Colodner R, et al: A 1-year trial of nasal mupirocin in the prevention of recurrent staphylococcal nasal colonization and skin infection. Arch Intern Med 156:1109, 1996

86. Todd JK, Ressman M, Caston SA, et al: Corticosteroid therapy for patients with toxic shock syndrome. JAMA 252:3399, 1985

87. Van Eldere J, Peetermans WE, Struelens M, et al: Polyclonal staphylococcal endocarditis caused by genetic variability. Clin Infect Dis 31:24, 2000

88. Seo SK, Venkataraman L, DeGirolami PC, et al: Molecular typing of coagulase-negative staphylococci from blood cultures does not correlate with clinical criteria for true bacteremia. Am J Med 109:697, 2000

89. Proctor RA: Editorial response: coagulase-negative staphylococcal infections: a diagnostic and therapeutic challenge. Clin Infect Dis 31:31, 2000

90. Fey PD, Ulphani JS, Gotz F, et al: Characterization of the relationship between polysaccharide intercellular adhesion and hemagglutination in *Staphylococcus epidermidis*. J Infect Dis 179:1561, 1999

91. Clark NM, Hershberger E, Zervosc MJ, et al: Antimicrobial resistance among gram-positive organisms in the intensive care unit. Curr Opin Crit Care 9:403, 2003

92. Ronald A: The etiology of urinary tract infection: traditional and emerging pathogens. Am J Med 113(suppl 1A):14S, 2002

93. Moellering RC: *Enterococcus* species, *Streptococcus bovis,* and *Leuconostoc* species. Principles and Practice of Infectious Diseases, 5th ed. Mandell GL, Bennett JE, Dolin R, Eds. Churchill Livingstone, Philadelphia, 2000, p 2147

94. Martin DR: Laboratory evaluation of streptococci. Streptococcal Infections: Clinical Aspects, Microbiology and Molecular Pathogenesis. Stevens DL, Kaplan EL, Eds. Oxford University Press, New York, 2000, p 266

95. Bisno AL: Nonsuppurative poststreptococcal sequelae: rheumatic fever and glomerulonephritis. Principles and Practice of Infectious Diseases, 5th ed. Mandell GL, Bennett JE, Dolin R, Eds. Churchill Livingstone, Philadelphia, 2000, p 2117

II TUBERCULOSIS

Michael K. Leonard, Jr., m.d.
Henry M. Blumberg, m.d., f.a.c.p.

Tuberculosis is a bacterial disease caused by *Mycobacterium tuberculosis,* a relatively slow-growing, aerobic, acid-fast bacillus. Classically, tuberculosis is a pulmonary disease, but disseminated and extrapulmonary manifestations may also occur, especially in immunocompromised persons. Tuberculosis is transmitted person to person and is usually contracted by inhalation of *M. tuberculosis* droplet nuclei generated by an infectious person.

After *M. tuberculosis* enters the body, the host's cell-mediated immunity may contain the organism but not eradicate all the bacilli, resulting in latent *M. tuberculosis* infection (LTBI). *M. tuberculosis* can remain dormant and persist (e.g., within macrophages); persons with dormant *M. tuberculosis* are at risk for reactivation and development of active tuberculosis. Treatment of LTBI can markedly reduce the risk of progression to active disease.[1,2]

If host defenses are unable to contain the infection, the bacillary load increases markedly and LTBI progresses to active tuberculosis. Persons with tuberculosis (also called tuberculosis disease or active tuberculosis) generally are symptomatic and may be infectious if they have pulmonary or laryngeal disease. Active tuberculosis is a life-threatening condition that requires treatment with a multidrug regimen for a minimum of 6 months.[3]

Epidemiology

Tuberculosis has emerged as an enormous global public health epidemic. Worldwide, it is the second leading infectious cause of death, after HIV infection.[4] The World Health Organization (WHO) has estimated that every year, about 9 million persons develop active tuberculosis, and more than 2 million persons die from the disease.[5] Most of these deaths occur in resource-poor countries, where about 95% of the cases are found. Most cases of tuberculosis (5 to 6 million a year) occur in persons 15 to 49 years of age. Sub-Saharan Africa has the highest incidence (≥ 300 per 100,000 population annually), in part because of high rates of HIV coinfection [*see Figure 1*].[4] For example, rates of HIV coinfection in patients with tuberculosis reportedly exceed 60% in Botswana, South Africa, Zambia, and Zimbabwe. The most populous countries of Asia have the largest numbers of cases: India, China, Indonesia, Bangladesh, and Pakistan together account for more than half the global burden, and 80% of new cases occur in 22 high-burden countries. In general, tuberculosis is declining in western and central Europe, North and South America, and the Middle East. By contrast, there have been striking increases in countries of the former Soviet Union and in sub-Saharan Africa, because of the HIV epidemic.[4,5] It has been estimated that about two billion persons, or one third of the world's population, are infected with *M. tuberculosis* and thus are at risk for progression to active disease. There are major concerns that without increased attention to the disease and the development of new tools for treatment and control (e.g., an effective vaccine; new therapeutic agents and shorter treatment regimens; and improved diagnostics, including those for LTBI), the global tuberculosis epidemic will continue to worsen.

In the United States and western Europe, tuberculosis was a leading cause of death until early in the 20th century. The incidence of tuberculosis in the United States began to decline with improved living conditions and public health measures, even before the availability of effective chemotherapy. After the introduction of effective therapy in the mid-20th century, the incidence of tuberculosis decreased even further. Between 1985 and 1992, however, the United States experienced a resurgence of tuberculosis because of underfunding of tuberculosis control efforts, the decline of the public health infrastructure, and the emergence of the HIV epidemic. With increased attention and funding, the number of tuberculosis cases in the United States declined from a peak of 26,673 in 1992 to 14,511 in 2004 (a decline in rate from 10.5 to 4.9 cases per 100,000 population) [*see Figure 2*].[6] The implementation of directly observed therapy (DOT) and improved infection control activities in hospitals and other institutional settings have contributed greatly to this decline.[6]

Most tuberculosis cases in the United States now occur in foreign-born persons and in nonwhites.[6] In 2004, tuberculosis case rates in African Americans born in the United States were more than eightfold higher than those in native whites. In the United States, rates are also much higher in Hispanics and Asians (especially foreign born) than in whites. The 2004 tuberculosis rate in foreign-born persons (22.5 cases per 100,000 population) was 8.7 times greater than that in United States–born persons (2.6 cases per 100,000 population).[6] In 2003, the top five countries of birth of foreign-born patients with tuberculosis were Mexico (25.6%), the Philippines (11.6%), Vietnam (8.4%), India (7.7%), and China (4.8%). Molecular typing studies have suggested that in foreign-born persons in the United States, most tuberculosis cases result from the reactivation of LTBI, whereas in persons born in the United States, many cases (perhaps a third or more) result from recent transmission.[7,8] Foreign-born persons may also be more likely to experience extrapulmonary tuberculosis.

Tuberculosis is not evenly distributed within the population. The disease is much more common in the economically disadvantaged, including the homeless and indigent inner-city residents.[9] Tuberculosis is 200 times more likely to occur in HIV-positive persons than in HIV-negative persons.[10] Persons with HIV coinfection are more likely to have extrapulmonary or disseminated tuberculosis, frequently along with pulmonary disease.

Other population groups that are at increased risk or that have a disproportionately high incidence of disease include immigrants (this is especially true during the first 5 years of arriving in the United States); substance abusers, including injection drug users and alcohol abusers; homeless persons; residents in certain institutional settings, such as correctional facilities and long-term care facilities; persons who are taking immunosuppressive drugs; and persons who have certain malignancies, diabetes mellitus, renal failure, or other debilitating conditions.[1,11] Travelers to countries where tuberculosis is endemic are likely to be at somewhat increased risk for developing tuberculosis.[12]

During 2004, drug resistance in initial isolates of *M. tuberculosis* from persons with no previous history of treatment of tuberculosis was more common in foreign-born patients than in United States–born patients. Such isolates included strains of *M. tuberculosis* resistant to at least isoniazid and rifampin (multidrug-resistant tuberculosis [MDR-TB]). The rate of MDR-TB was higher in foreign-born than in United States–born persons (1.4% versus 0.6%), reflecting likely exposure to tuberculosis in countries where rates

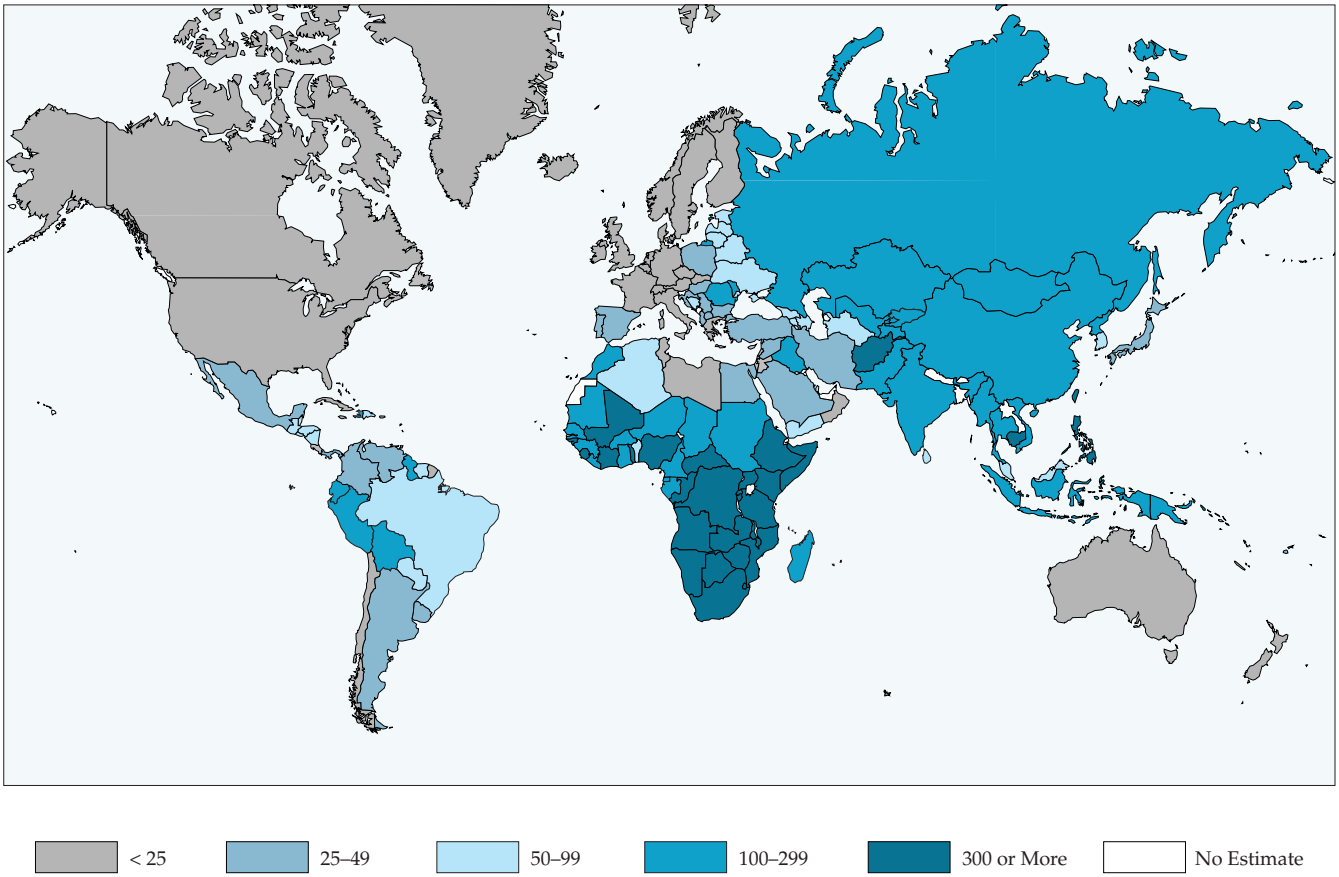


Figure 1 **Incidence of global tuberculosis as estimated by the World Health Organization, 2002.[1]**

of MDR-TB are higher than in the United States. Rates of MDR-TB in the United States have decreased since the early 1990s.[6] The decrease occurred in large part because of the dramatic reduction of MDR-tuberculosis cases in New York City: in the early 1990s, MDR-TB accounted for nearly a fifth of all tuberculosis cases in New York City.[13,14] A key element of this decrease has been enhanced tuberculosis control through an improved public health structure and greater attention to treatment of tuberculosis, including greater use of DOT.

Etiology and Genetics

Tuberculosis is spread person to person via airborne droplet nuclei. These particles, which are 1 to 5 μm in diameter and contain *M. tuberculosis,* are generated by persons with pulmonary or laryngeal tuberculosis when they cough, sneeze, speak, or sing. Most secondary cases of tuberculosis occur in household members or other close contacts of the index case. Prolonged exposure to the index case increases the risk of becoming infected, although on occasion, transmission can occur after brief exposures. Infectivity is greatest in patients whose sputum smear is AFB positive; this group may include those with cavitary disease or tuberculosis of the larynx. Coughing further enhances shedding. Persons with tuberculosis who are AFB smear–negative (and culture positive) are thought to be less infectious than AFB smear–positive patients but may still transmit tuberculosis.[15]

Some patients may have an increased susceptibility to tuberculosis that is genetically determined. For example, concordance for tuberculosis is significantly higher in monozygous twins (65% to 85%) than in dizygous twins (25% to 35%).[16] African Americans and Native Americans may be more susceptible than whites to *M. tuberculosis* infection.[17] Other studies have suggested that patients carrying mutations in the receptors for interferon gamma (IFN-γ) and interleukin-12 (IL-12) are at increased risk for severe atypical mycobacterial and disseminated bacillus Calmette-Guérin (BCG) infections.[18] Several associations have also been made with variants of genes thought to be important in the pathogenesis of tuberculosis, including *NRAMP1* and genes that code for the vitamin D receptor (*VDR*), IL-10, tumor necrosis factor–α (TNF-α), and IL-1. Four polymorphic-derived deletions or point mutations of the *NRAMP1* gene have been associated with increased susceptibility to tuberculosis in Gambia and in other populations in Japan, Guinea, and Korea.[18] Associations between genetic polymorphisms and tuberculosis susceptibility differ according to ethnic origin,[19] but the extent to which genetic polymorphisms contribute to the global burden of disease has not been fully elucidated, in part because of the great difficulty of separating lifelong environmental influences from genetic predisposition.[4]

Pathogenesis

The pathogenesis of tuberculosis is unique among infectious diseases because of the highly variable but sometimes long latency period between infection and clinical illness. Although a single tubercle bacillus theoretically can cause infection, it must first bypass the upper airway defense mechanisms and lodge in the distal pulmonary alveoli. Infectious droplet nuclei are inhaled

and lodge in the alveoli in the distal airways, whereas larger particles are usually trapped in the upper respiratory tract. *M. tuberculosis* is taken up by alveolar macrophages, and this can result in infection with the organism. After exposure to someone with infectious tuberculosis, the exposed person experiences one of four potential outcomes [*see Figure 3*]: (1) no infection (as measured by a negative tuberculin skin test); (2) infection, with rapid progression to active disease (primary tuberculosis); (3) LTBI, in which immune mechanisms prevent the progression to active disease; and (4) LTBI followed by subsequent reactivation and development of active tuberculosis months to years later.[20,21]

The immune response mounted against *M. tuberculosis* is multifaceted and complex. Effective innate immune responses to *M. tuberculosis* are clearly important, given that a significant proportion of persons exposed to *M. tuberculosis* do not become infected after exposure. For example, contact investigations have shown that at most only 30% to 50% of persons with heavy exposure to someone with tuberculosis will become infected, as demonstrated by conversion on tuberculin skin testing. If infection does occur, *M. tuberculosis* multiplies within alveolar macrophages and subsequently disseminates through blood and lymphatic pathways to areas of high oxygen tension. Hence, the lung apices are a common repository. Other frequently infected areas include the renal cortex, the vertebral column, and the metaphyseal ends of long bones. After 6 to 8 weeks, an adaptive cell-mediated immunity is well established, and results of tuberculin skin testing become positive.

M. tuberculosis is an obligate aerobe. Consequently, it grows most successfully in those human tissues having the highest oxygen tension, such as the lung apices. It is a slow-growing organism, with a generation time estimated to be 12 to 18 hours. As a result, tuberculous lesions in humans typically evolve from a subacute to a long-term stage, and laboratory isolation of the organism usually requires weeks rather than a day or two, as is the case for most bacteria.

The cell walls of *Mycobacterium* species have a high lipid content because of the presence of mycolic acids. Therefore, mycobacteria are impermeable to, and undetectable with, the usual bacteriologic stains, such as Gram stain. Mycobacteria, including *M. tuberculosis,* are acid-fast bacilli (AFB); the lipoid capsule of the acid-fast organism takes up carbol-fuchsin and resists decolorization with a dilute acid rinse. Finally, although tubercle bacilli require special enriched media to grow in the laboratory, these organisms are resistant in vivo to physical stress.

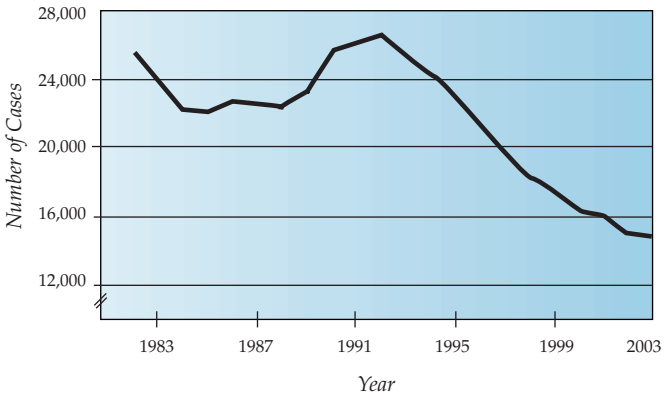


Figure 2 **Number of tuberculosis cases reported in the United States, 1982–2003.[4]**

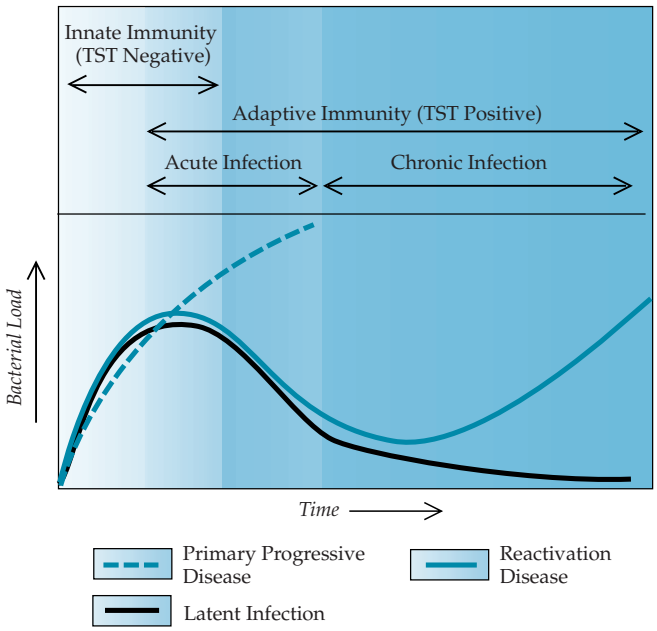


Figure 3 **Natural history of *Mycobacteria tuberculosis* infection.[96] The innate immune system provides the first line of defense against *M. tuberculosis,* and often prevents infection. Results of tuberculin skin testing (TST) are negative at this stage. Subsequent control of tuberculosis is provided by the adaptive immune system; this results in positive TST results. In 90% of cases, host defenses kill the great majority of organisms, but some *M. tuberculosis* bacilli persist within macrophages, resulting in latent infection. Reactivation of disease occurs in 5% to 10% of patients, sometimes years later. Less than 5% of patients experience primary progressive disease.**

A wide range of immune components are involved in an effective response against *M. tuberculosis.* These components include T cells ($CD4^+$ and $CD8^+$, which are activated in response to *M. tuberculosis* infection), cytokines (including IFN-γ, IL-12, TNF-α, and IL-6), and macrophages.[21] The macrophage is felt to play a key role in the control of *M. tuberculosis* infection; the organism can multiply within resting macrophages, but it can be inhibited or killed when the macrophage is activated.[21]

Cytokines produced by T cells contribute to the immune response in a multitude of ways, such as by activating macrophages, the host cell in which *M. tuberculosis* primarily resides. $CD4^+$ and $CD8^+$ T cells can also be cytotoxic for infected cells. $CD4^+$ T cells play a key role in the immune response, as demonstrated by a marked increase in susceptibility to tuberculosis in HIV-infected persons whose $CD4^+$ T cells are depleted. The predominance of T helper type 1 (Th1) cell response is associated with protection and control of tuberculosis infection, whereas Th2 responses predominate in patients who are unable to contain the infection and who develop active disease. Th1 responses are markedly impaired in HIV-infected persons, especially those with low $CD4^+$ T cell counts and advanced disease. IL-12 is an important cytokine in controlling *M. tuberculosis.* It is produced by activated macrophages and drives development of a Th1 response, which stimulates $CD4^+$ T cells to release IFN-γ. IFN-γ alone is insufficient to control *M. tuberculosis.* IFN-γ is, however, a crucial element in the control of tuberculosis, and it also stimulates the macrophage to release TNF-α, which is important in granuloma formation and control of the extent of infection. The importance of TNF-α has been demonstrated by the marked in-

crease in risk of progression to active tuberculosis, including extrapulmonary and disseminated disease, in patients with LTBI who were treated with the anti-TNF monoclonal antibody infliximab for rheumatologic and immunologic diseases.[22]

After exposure to and infection with *M. tuberculosis,* most persons develop LTBI [*see Figure 3*]. This chronic infection stimulates formation of granulomas in the lungs or other tissues; these granulomas consist of lymphocytes ($CD4^+$ and $CD8^+$ T cells, as well as B cells) that surround macrophages, some of which contain *M. tuberculosis,* as well as other cells such as fibroblasts. The development of a granuloma serves to limit the spread of the infection by walling off the organisms from the rest of the lung or other organ tissue. After the development of cell-mediated immunity, host defenses are able to respond to *M. tuberculosis;* the great majority of the organisms are killed, and the mycobacterial load is greatly reduced [*see Figure 3*]. However, host defenses are not able to entirely eradicate all organisms; some *M. tuberculosis* persist within macrophages, and thus the possibility remains for reactivation of disease. The mechanism of resistance by *M. tuberculosis,* persistence in macrophages over many years, and reactivation is poorly understood.

As a result of host defenses, most patients experience complete healing of these initial tuberculous lesions. In patients whose primary lesions heal, the chest radiograph may be normal or it may show focal calcifications. The primary lower lobe lesion and its draining node may be recognized radiographically as the Ghon complex; apical calcifications (Simon foci) may be present. Although inactive, these lesions contain small numbers of dormant but viable tubercle bacilli, and breakdown of the lesions can lead to reactivation of infection.

In about 5% to 10% of immunocompetent persons infected with *M. tuberculosis,* LTBI progresses to active disease.[23] The risk of reactivation is greatest within the first 2 years of the initial infection, but there is a subsequent lifetime risk of up to 5% for reactivation, which can occur many decades after initial infection.[24] Immune compromise increases the risk of progression to active disease. HIV infection is the greatest single risk factor for progression to active disease in adults. Progression from LTBI to active disease takes place at a rate of 10% a year in HIV-infected persons[2]; those with low $CD4^+$ T cell counts may be incapable of controlling infection and may develop active disease rapidly after exposure and infection. Other medical conditions that predispose to the development of active disease include diabetes mellitus, renal failure, certain malignancies, cancer chemotherapy, therapy with corticosteroids or other immunosuppressive drugs (including TNF-α inhibitors such as infliximab, etanercept, and adalimumab), transplantation, and malnutrition. Tuberculosis may also develop in patients without these underlying risk factors, however, for reasons that are not well understood. In some patients, including those who are infected with HIV and those living in regions where tuberculosis is hyperendemic, exogenous reinfection can occur.[25] In the United States, however, molecular epidemiologic investigations indicate that most recurrences result from relapse of disease rather than reinfection with a different strain of *M. tuberculosis*.[26]

Forms of Tuberculosis

PULMONARY TUBERCULOSIS

In the United States, about 80% of tuberculosis cases are pulmonary disease.[27] Pulmonary tuberculosis can be divided into primary tuberculosis (i.e., developing soon after infection) and secondary tuberculosis (i.e., developing after a variable period of LTBI). Secondary disease is also known as postprimary or reactivation tuberculosis.

Primary Disease

Primary tuberculosis is frequently localized to the middle and lower lung zones and is accompanied by hilar or paratracheal lymphadenopathy. In some cases, the lesion heals spontaneously and may later be evident on chest x-ray as a small calcified nodule (Ghon lesion). Primary disease was once most common in young children, but it has been seen with increasing frequency in adults who are debilitated or immunosuppressed, especially from HIV infection. Primary tuberculosis typically manifests as one of four broad syndromes: a syndrome resembling atypical pneumonia; tuberculous pleuritis with pleural effusion; direct progression to upper lobe disease; and progression to extrapulmonary disease. Uncommon manifestations of primary tuberculosis include erythema nodosum and other hypersensitivity reactions, such as reactive arthritis (Poncet disease).

The most common form of primary tuberculosis is a syndrome that is similar to atypical pneumonia, with fever and nonproductive cough. The chest radiograph may show unilateral, lower lobe, patchy parenchymal infiltrates; paratracheal or hilar adenopathy; or both. Although patients with this form of tuberculosis should receive full antituberculous chemotherapy, the symptoms may resolve even without therapy. Resolution without therapy would not be expected in most immunocompromised persons, however.

Tuberculous pleuritis with pleural effusion results from penetration of bacilli into the pleural space from an adjacent subpleural focus. This can occur early in the course of infection and may represent a hypersensitivity response to only a few organisms in the pleural space.[28] In immunocompetent patients, this form of tuberculous pleuritis may go unnoticed, and the process may resolve spontaneously. Other patients, however, including both immunocompetent and immunosuppressed persons, may have an acute illness with high fever, cough, and pleuritic chest pain; if the effusion is large, dyspnea may also occur. Chest radiography often reveals unilateral pleural effusion, generally without identifiable parenchymal lesions. The tuberculin skin test is strongly positive in the majority of immunocompetent persons with tuberculous pleurisy, but it is positive in only about 40% of HIV-infected patients who have the syndrome.

Direct progression of primary tuberculosis to upper lobe disease is relatively rare. Progression of primary infection to extrapulmonary tuberculosis (also known as progressive primary tuberculosis) was once most common in young children, who presented with cervical adenitis, miliary tuberculosis, or tuberculous meningitis; currently, it is most often observed in persons with HIV infection.

Secondary Disease

Reactivation pulmonary disease is the most common clinical form of tuberculosis. Classic symptoms include cough, fever, and sweats. Symptoms usually begin insidiously and progress over a period of several weeks or even months before diagnosis. Cough may be nonproductive, or it may gradually become productive. Dyspnea is relatively uncommon in the absence of underlying chronic lung disease. Systemic symptoms, which are often prominent, include fever, anorexia, weight loss, night sweats, and malaise. Fever is reported in 37% to 80% of patients with tuber-

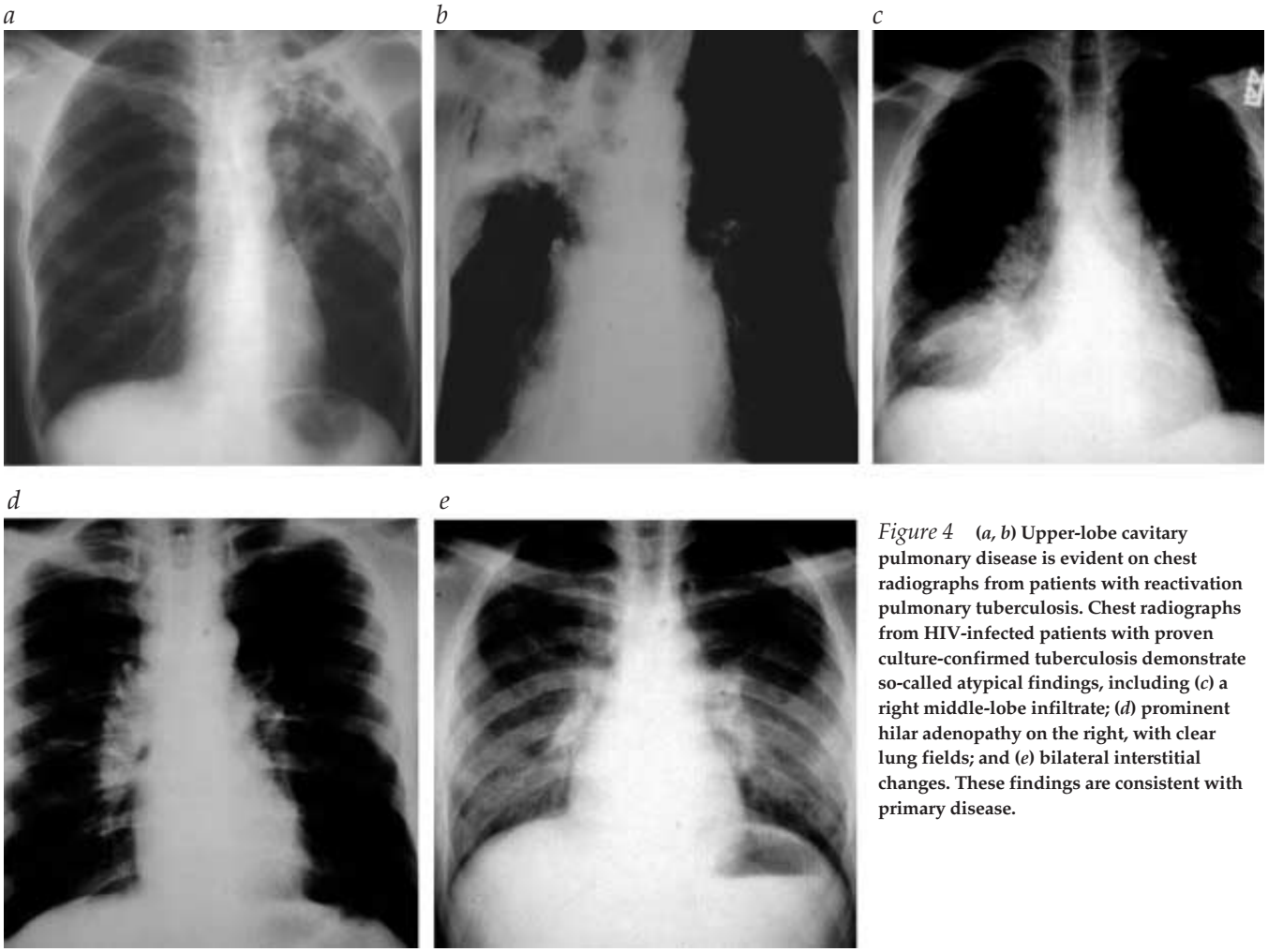


Figure 4 **(*a*, *b*) Upper-lobe cavitary pulmonary disease is evident on chest radiographs from patients with reactivation pulmonary tuberculosis. Chest radiographs from HIV-infected patients with proven culture-confirmed tuberculosis demonstrate so-called atypical findings, including (*c*) a right middle-lobe infiltrate; (*d*) prominent hilar adenopathy on the right, with clear lung fields; and (*e*) bilateral interstitial changes. These findings are consistent with primary disease.**

culosis[28]; low-grade fevers are typical, but some patients have high temperatures and even chills. However, in some patients with pulmonary tuberculosis, these classic symptoms may be absent, making diagnosis more difficult.[29] In addition, in patients with advanced disease who present with respiratory failure, tuberculosis may not be considered in the differential diagnosis, and as a result, the diagnosis of pulmonary tuberculosis may be delayed.

Hemoptysis from endobronchial erosion may occur in tuberculosis; it is usually minor but denotes advanced disease. Massive hemoptysis resulting from the erosion of a pulmonary artery by an advancing cavity (Rasmussen aneurysm) is a terminal event that was occasionally seen in the prechemotherapy era but is currently rare. Hemoptysis may also occur in patients with inactive disease (e.g., after completion of therapy) who develop *Aspergillus* superinfection of a residual cavity (aspergilloma).

Physical Examination Findings

Physical examination is typically of limited usefulness in differentiating tuberculosis from other pulmonary infections. Some patients with tuberculosis have no abnormalities that are detectable by chest examination, whereas others have rales in the involved areas. Coarse rhonchi may evolve as secretions increase. Bronchial breath sounds may be present in areas of consolidation.

Imaging Studies

A chest radiograph is an important tool in the diagnosis of pulmonary tuberculosis. Typical features include unilateral or bilateral infiltration; cavitation is common in patients with reactivation disease [*see Figure 4*]. The most frequent sites of involvement in reactivation disease are, in decreasing order, the apical and posterior segments of the right upper lobe, the apical-posterior segment of the left upper lobe, and the superior segments of the lower lobes. Lower-zone disease is seen at presentation in fewer than 15% of HIV-seronegative adults; it is much more common in HIV-infected persons (e.g., as part of primary disease) and is somewhat more common in patients with diabetes mellitus and in patients with prominent peribronchial and endobronchial involvement. Chest radiographs appear normal in about 10% of persons with tuberculosis and HIV coinfection,[30] whereas normal chest radiographs are extremely rare in HIV-seronegative persons with tuberculosis.

Computed tomography is more sensitive than chest radiography. CT scans may show nodular or branching linear centrilobular lesions in very early disease and in patients with tuberculosis and an apparently normal chest radiograph.

EXTRAPULMONARY TUBERCULOSIS

Tuberculosis may affect any organ system. Extrapulmonary tuberculosis results from hematogenous dissemination of tuber-

cle bacilli with incomplete immunologic control of the disease, either during primary infection or as a result of reactivation from a site of latent infection.

In order of frequency, extrapulmonary tuberculosis involves the lymph nodes, the pleura, the genitourinary tract, bone and joints, the meninges, and the peritoneum. Extrapulmonary tuberculosis, including miliary (disseminated) disease, is being seen more frequently because of its increased prevalence in HIV-infected persons. It is not unusual for HIV-infected patients with tuberculosis to have concomitant pulmonary and extrapulmonary disease. Very young children, immunocompromised patients, and perhaps patients born in foreign countries are also at increased risk for extrapulmonary disease, as are patients with LTBI who have been treated with TNF-α inhibitors such as infliximab, etanercept, and adalimumab.[22]

Tuberculous Lymphadenitis

Tuberculosis of the lymph nodes is the most common form of extrapulmonary tuberculosis, accounting for up to 40% of extrapulmonary cases. It is frequently encountered in HIV-infected persons. Tuberculous lymphadenitis is also seen in young children; women, especially nonwhites, also appear to be at increased risk. The cervical nodes (posterior and anterior) and supraclavicular lymph nodes are most commonly affected. In addition, mediastinal lymphadenitis may appear with early or primary disease, because these nodes drain the lung. HIV-seronegative patients are often afebrile and present with slowly enlarging, painless mass lesions. Patients with HIV infection or AIDS may be febrile. The diagnosis can be established by fine-needle aspiration or lymph node biopsy. Therapeutic lymph node excision is not indicated except in unusual circumstances. For large lymph nodes that are fluctuant and that appear to be about to drain spontaneously, aspiration or incision and drainage appear to be beneficial, although this approach has not been examined systematically.[3]

Pleural Tuberculosis

Pleural tuberculosis generally results from primary progressive *M. tuberculosis* infection; less commonly, it may occur as part of the reactivation of LTBI. As noted, pleural disease typically occurs when a few organisms from the lungs gain access to the pleural space and, in the presence of cell-mediated immunity, cause a hypersensitivity response. Physical findings are those of a pleural effusion: dullness to percussion and absence of breath sounds. A chest radiograph usually demonstrates a unilateral pleural effusion. Thoracentesis is required for diagnosis; findings include an exudative effusion with a protein concentration that is greater than 50% of the serum level, a normal to low glucose level, and the presence of white blood cells, most of which are lymphocytes and mononuclear cells rather than neutrophils. A pleural biopsy can greatly increase the likelihood of obtaining a positive culture (increasing the yield to greater than 80%) compared with pleural fluid culture alone.

Tuberculous empyema has become much less common than it was in the prechemotherapy era. It results from the rupture of a lung cavity into the pleural space or from a bronchopleural fistula. The rupture of a cavity results in the discharge of a large number of organisms into the pleural space; in addition, parenchymal disease is often present on chest radiograph, and a pyopneumothorax with an air-fluid level may also be visible. The effusion is purulent and thick and contains large numbers of white blood cells, predominantly lymphocytes. Treatment consists of drainage (often requiring a surgical procedure) and antituberculous chemotherapy. Surgery, when needed, should be undertaken by experienced thoracic surgeons.[3]

Genitourinary Tuberculosis

Genitourinary tuberculosis accounts for about 15% of extrapulmonary tuberculosis cases and may involve any part of the genitourinary tract. It usually results from hematogenous seeding after primary infection. Historically, it has occurred years after primary infection. Local symptoms predominate: dysuria, hematuria, and frequent urination are common, and flank pain may also be noted.[28] Renal tuberculosis often has an insidious onset and subtle symptoms; consequently, advanced destruction of the kidneys may have taken place by the time the diagnosis is established. The urine sediment is abnormal in about 90% of patients with renal tuberculosis; urinary findings include pyuria, hematuria, or both. Imaging studies may reveal structural abnormalities; calcification, cavitation, and ureteral strictures and fibrosis suggest tuberculosis, whereas calyceal dilatation, cortical scarring, and papillary necrosis are nonspecific. A CT scan is as useful as an intravenous pyelogram for visualizing advanced urinary tract tuberculosis; ultrasonographic studies are less accurate. The finding of so-called sterile pyuria (i.e., acidic urine that contains white blood cells but in which no bacterial organisms are isolated on routine urine culture) should prompt culture of the urine for mycobacteria. AFB smears should be performed on urine, although the yield is low compared with that of AFB urine culture; nucleic acid amplification (NAA) testing (see below) of urine may provide a more rapid diagnosis than AFB culture but is not a substitute for a culture.[31]

Genital tuberculosis is more common in women than men. In women, genital involvement may occur without renal tuberculosis; pelvic pain, menstrual irregularities, and infertility are possible presenting complaints. Occasionally, ovarian masses from tuberculosis may be mistaken for ovarian tumors. The differential diagnosis may also include pelvic inflammatory disease. The physical examination may be normal or may reveal an adnexal mass. Endometrial curettage, cervical biopsy, and laparoscopic exploration are all useful as diagnostic procedures. Surgery is often needed for diagnosis of tubo-ovarian abscesses or pelvic peritonitis.

Male genital tuberculosis can result from hematogenous seeding or can spread from infected urine after reactivation in the upper urinary tract. About 50% of male patients with genital tuberculosis also have renal tuberculosis; this is a higher proportion than is found in women with genital tuberculosis. Male patients may present with slowly enlarging mass lesions of the epididymis, prostate, or seminal vesicles. Genital tuberculosis in both men and women responds well to chemotherapy.

Musculoskeletal Tuberculosis

Tuberculosis can affect any bone or joint, but involvement of the spine (Pott disease) is the most common type of skeletal tuberculous disease, accounting for up to 50% of cases.[32] The thoracic spine is the most common site of spinal tuberculosis. Upper thoracic vertebral body involvement is more common in children, whereas lower thoracic and upper lumbar disease is more common in adults. Often, two or more vertebral bodies are involved; vertebral body involvement can lead to disease of an adjacent intervertebral disk and paraspinous abscesses. With advanced disease, collapse of vertebral bodies may result in kyphosis (gibbus) or even paraplegia.[33]

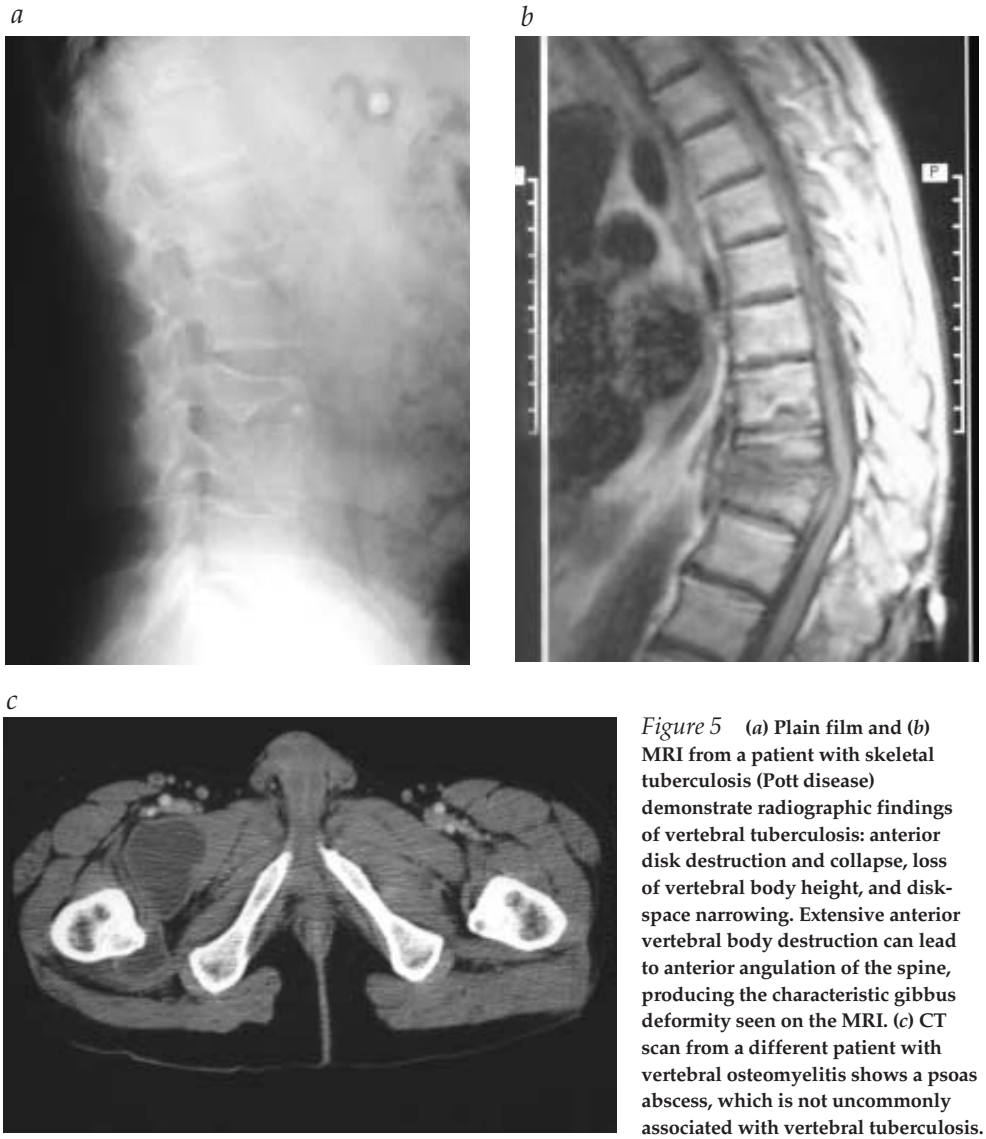


Figure 5 **(*a*) Plain film and (*b*) MRI from a patient with skeletal tuberculosis (Pott disease) demonstrate radiographic findings of vertebral tuberculosis: anterior disk destruction and collapse, loss of vertebral body height, and disk-space narrowing. Extensive anterior vertebral body destruction can lead to anterior angulation of the spine, producing the characteristic gibbus deformity seen on the MRI. (*c*) CT scan from a different patient with vertebral osteomyelitis shows a psoas abscess, which is not uncommonly associated with vertebral tuberculosis.**

The usual presenting symptom of skeletal tuberculosis is pain. Patients with joint involvement can have swelling of the joint and limitation of motion. Tuberculous joint involvement can sometimes become apparent after trauma to the joint. Because of the subtle nature of the symptoms, especially initially, diagnosis of skeletal and joint disease can be delayed for long periods.

Radiographically, tuberculosis of bones appears as an array of destructive osteolytic lesions with relatively little reactive bone formation. CT and MRI are useful imaging tests [*see Figure 5*]. Definitive diagnosis requires biopsy and culture of affected bone or, in patients with joint involvement, arthrocentesis with culture of synovial fluid. Tuberculous arthritis is characteristically a chronic, slowly progressive, destructive monoarticular process. The synovial fluid has a high protein level, a low glucose level, and a poor mucin clot. The white blood cell count is variable but is typically in the range of 10,000 to 20,000/μl; neutrophils often predominate. AFB smears are rarely positive because of the paucibacillary nature of this form of tuberculosis, but culture and biopsies of the synovium are helpful.

Tuberculous rheumatism (Poncet disease) is a rare form of acute polyarthritis that results from a hypersensitivity reaction rather than direct synovial infection. Prosthetic joint tuberculosis and tubercular tenosynovitis are uncommon. For skeletal tuberculosis, tumor is the major consideration in the differential diagnosis, and fungal and pyogenic infections are additional considerations.

Randomized trials in patients with spinal tuberculosis demonstrated no additional benefit of surgical debridement or radical operation (i.e., resection of the spinal focus and bone grafting) in combination with chemotherapy compared with chemotherapy alone.[34] Myelopathy with or without functional impairment most often responds to chemotherapy. However, in some circumstances, surgery appears to be beneficial and may be indicated. Indications for surgery include the failure to respond to chemotherapy in conjunction with evidence of ongoing infection; the need for spinal cord decompression in patients with persistent or recurrent neurologic deficits; or instability of the spine.[3]

Tuberculous Pericarditis

Although infrequent, tuberculous pericarditis is a very serious form of tuberculosis. Infection of the pericardium can result from either hematogenous dissemination of bacilli or contiguous spread from lung or mediastinal nodes. Pathologically, the disease progresses from inflammation to effusion and eventually to fibrous organization. Symptoms are nonspecific and initially in-

clude the insidious onset of fever, weight loss, and night sweats. Subsequently, cardiopulmonary symptoms occur; these include cough, dyspnea, orthopnea, ankle swelling, and chest pain. Physical examination may disclose a pericardial rub or a pulsus paradoxus. A chest x-ray may show a pericardial effusion.

The diagnosis of tuberculous pericarditis depends on direct examination of pericardial fluid or tissue. The pericardial fluid is turbid or hemorrhagic. White blood cell counts typically range from 5,000 to 10,000/μl; lymphocytes are predominant. High protein levels and low glucose levels are typical. AFB smears or cultures of pericardial fluid are positive in about half of cases, but pericardial biopsy with culture has a higher diagnostic yield. The major differential diagnosis includes idiopathic, bacterial, or viral pericarditis and neoplasm.

If left untreated, tuberculous pericarditis has a high mortality, and constriction eventually occurs in many survivors**.** Surgery is indicated if clinical tamponade progresses or recurs despite repeated pericardiocentesis [*see 1:XIII Diseases of the Pericardium, Cardiac Tumors, and Cardiac Trauma*]. However, in the absence of tamponade, medical therapy is generally sufficient. Antituberculous chemotherapy with four drugs should be started immediately and supplemented initially with corticosteroids; corticosteroids should be tapered over a 12-week period [*see* Treatment, *below*]. Corticosteroids have been useful in reducing mortality from tuberculous pericarditis and enhancing clinical response to therapy, but they do not appear to reduce progression to constriction or the need for pericardiectomy.

Central Nervous System Disease

Tuberculous meningitis is a particularly devastating manifestation of tuberculosis, with high mortality (about 40%) and morbidity.[35] Children younger than 5 years of age and HIV-infected persons are at increased risk for tuberculous meningitis. The clinical manifestations, laboratory findings, and outcomes are similar in patients with and without HIV infection.[35,36] Tuberculous meningitis may result from hematogenous seeding of the meninges, or it can be caused by the breakdown of an old submeningeal granuloma with rupture into the subarachnoid space.

Clinical manifestations of tuberculous meningitis result both from the presence of *M. tuberculosis* and from the inflammatory host immune response.[4] Clinical manifestations may include headache, fever, altered mental status, cranial nerve findings, and nuchal rigidity. The intense inflammatory reaction is most prominent at the base of the brain and can have three effects: direct compression of neural tissues, especially cranial nerves; vasculitis, leading to areas of infarction; and obstruction of the free flow of cerebrospinal fluid, leading to cerebral edema, hydrocephalus, or subarachnoid block. CT or MRI may demonstrate basal meningeal enhancement and hydrocephalus. Up to 50% of patients with tuberculous meningitis have abnormalities on chest radiograph indicating old healed tuberculosis or current pulmonary disease or miliary disease. Lumbar puncture is an essential diagnostic test and should be carried out if meningeal signs are present. The CSF opening pressure is usually elevated but on occasion may be normal. CSF examination generally reveals an elevated white blood cell count (often in the range of 100 to 1,000/μl), typically with a lymphocyte predominance, although in early disease, neutrophils can predominate; CSF protein is elevated and CSF glucose level is usually decreased. Acid-fast smears of CSF are insensitive; they are positive in only about 10% of patients with culture-confirmed tuberculous meningitis.[35] CSF cultures may be eventually positive in up to 75% of cases. NAA tests (see below) may be positive but are insensitive in diagnosing tuberculous meningitis; a negative NAA test on CSF does not rule out tuberculous meningitis.[37] The principal considerations in the differential diagnosis include cryptococcal meningitis and other less common forms of fungal meningitis (e.g., histoplasmosis, blastomycosis, cocciodiomycosis), viral meningitis or encephalitis (e.g., from herpes simplex virus, enteroviruses, West Nile virus, and other arboviruses), and, if antibiotics have been given, partially treated bacterial meningitis [*see 7:XXXVI Bacterial Infections of the Central Nervous System*]. Noninfectious diseases in the differential diagnosis include carcinomatous meningitis, neurosarcoidosis, and CNS vasculitis.

Initiation of empirical therapy is crucial for patients with presumed tuberculous meningitis. Corticosteroids are indicated as adjunctive therapy (see below). Without therapy, tuberculous meningitis is universally fatal. The prognosis is worse in young children and in patients who present with altered mental status.[38]

Less common forms of CNS tuberculosis include radiculomyelitis and other infections of the spinal cord or epidural space[39]; and cerebral tuberculomas,[40] which typically present as slowly enlarging mass lesions. Tuberculomas may be initially misdiagnosed as brain tumors before surgical exploration and brain biopsy confirm the proper diagnosis. Tuberculomas may also develop during the course of therapy for tuberculous meningitis; their appearance does not necessarily represent treatment failure.

Abdominal Tuberculosis

Tuberculosis can involve the peritoneum or any intra-abdominal organ. The clinical manifestations depend on the area of involvement. Peritoneal disease is the most common type of abdominal tuberculosis. In the gut, tuberculosis may occur in any location from the mouth to the anus but is most common in the terminal ileum and cecum; other portions of the colon and rectum are less frequently involved.[28] Peritoneal tuberculosis may be secondary to hematogenous spread or, in women, to genital tuberculosis. Ileocecal and anorectal tuberculosis likely arise from the ingestion of tubercle bacilli in association with pulmonary disease. In peritoneal tuberculosis, the onset may be insidious; pain is a frequent presenting manifestation, often accompanied by abdominal swelling and increasing girth, fever, weight loss, and anorexia. Active pulmonary tuberculosis is often not present in patients with peritoneal tuberculosis. Because peritoneal tuberculosis may occur in patients with preexisting disorders, including hepatic cirrhosis with ascites, the symptoms of tuberculosis may be obscured.[28] The findings of ascites, abdominal tenderness, and fever should prompt an evaluation for infection, which should include paracentesis. Examination of peritoneal fluid generally shows an elevation in the white blood cell count with a lymphocytic predominance, elevated protein levels, and decreased glucose levels. AFB smears and cultures of peritoneal fluid are often negative unless extremely large volumes of fluid are examined. Laparoscopy with biopsy is recommended if tuberculosis is suspected; it has a much higher diagnostic yield. The differential diagnosis includes carcinomatosis, lymphoma, and cirrhosis. Tuberculous enteritis often involves the ileocecal region and may mimic Crohn disease or a malignancy. CT scanning and barium examinations are helpful, but colonoscopy and biopsy are required for diagnosis.

Miliary (Disseminated) Tuberculosis

Miliary, or disseminated, tuberculosis is defined as involvement of many organs simultaneously. It can occur as a result of

primary progressive disease or reactivation of latent infection.[4] Miliary tuberculosis is both a radiologic and pathologic term used to describe the hematogenous dissemination of *M. tuberculosis*. Radiologically, the term miliary refers to the pattern often seen on chest radiography, which is described as resembling millet seeds (e.g., a small reticulonodular pattern rather than an infiltrate). Not all patients with disseminated disease have pulmonary involvement, however.

The epidemiology of miliary or disseminated tuberculosis has changed dramatically over time. The incidence of miliary disease decreased markedly after the introduction of effective chemotherapy for tuberculosis. However, the advent of HIV/AIDS brought an increase in the number of cases. Miliary tuberculosis was once a disease primarily of children but now occurs primarily in HIV-infected persons, especially those with low $CD4^+$ T cell counts; it is also occasionally seen in elderly persons or other immunocompromised patients.

The onset of miliary tuberculosis is usually subacute. Symptoms often progress over a period of 1 to 4 months before diagnosis. Fever, anorexia, and weight loss occur in most patients. Respiratory symptoms occur in about half of patients with miliary disease, but hemoptysis is quite rare. Many other symptoms may appear; headache is particularly important because it may reflect coexisting tuberculous meningitis. Variant presentations, which account for a small percentage of cases, include cryptic miliary tuberculosis, in which patients have normal chest x-rays and exhibit problems typical of fever of undetermined origin [*see 7:XXIV Hyperthermia, Fever, and Fever of Undetermined Origin*]. An uncommon occurrence is fulminating miliary disease, which can have a sepsislike picture that includes respiratory failure, acute respiratory distress syndrome, disseminated intravascular coagulation, and multiorgan failure; often, tuberculosis is not considered early in this diagnosis.[41,42]

The physical examination in patients with miliary tuberculosis usually yields nonspecific results. Various pulmonary findings are seen in up to 50% of observed cases, hepatomegaly is seen in 30%, and splenomegaly or lymphadenopathy in 15%. Choroidal tubercles are less common, but they are diagnostically useful if present. Laboratory findings are often nonspecific. The complete blood count may be normal, but dramatic abnormalities have been well documented in this disorder and may range from pancytopenia to leukemoid reactions. Abnormal liver function test results, especially elevation of the alkaline phosphatase level, occur in 30% of cases. Hyponatremia is less common but, if present, should raise the possibility of inappropriate antidiuretic hormone secretion or adrenal insufficiency. AFB sputum smears are positive in only a minority (no more than 30%) of patients who have a miliary pattern on chest radiography. The use of fiberoptic bronchoscopy, bronchial brushings, and transbronchial biopsy to collect specimens may enhance the accuracy of bacteriologic studies, providing confirmation of the diagnosis in patients with miliary disease who have abnormal chest x-rays but negative sputum smears. Liver biopsy is especially helpful, revealing granulomas and providing positive cultures in about 60% of patients. Bone marrow biopsy is positive in about one third of all patients with miliary tuberculosis and has an even higher yield in those with hematologic abnormalities. The differential diagnosis in miliary tuberculosis includes histoplasmosis and other mycotic infections, sarcoidosis and other connective-tissue diseases, and malignant disorders.

Disseminated or miliary disease is fatal without chemotherapy; even with combined chemotherapy, mortality may be as high as 20%.[43,44] Adverse prognostic features include the presence of meningitis, extremes of age (i.e., old age and early childhood), delay in presentation, and the presence of underlying diseases. Clinical improvement with treatment is often very slow, and fever can persist for 1 to 3 weeks.

Other Forms of Extrapulmonary Tuberculosis

Other, less common forms of extrapulmonary tuberculosis include infection of the eye, skin (lupus vulgaris), upper respiratory tract (especially the larynx), pancreas, ear, and adrenal gland. Adrenal disease (often occurring with miliary or disseminated disease) is particularly important to consider and often is a manifestation of advanced disease presenting as signs of adrenal insufficiency. Onset of adrenal tuberculosis is usually insidious but can be acute; the disease should be considered in all patients with active or remote tuberculosis who are doing poorly, particularly if hypotension, hyponatremia, or hyperkalemia is present. Congenital tuberculosis is rare but can result from transplacental spread of *M. tuberculosis* to the fetus or from ingestion of contaminated amniotic fluid. Affected infants have disseminated disease, including involvement of the liver, spleen, lymph nodes, and other organs.

TUBERCULOSIS IN HIV-INFECTED PATIENTS

The interaction between HIV and *M. tuberculosis* is synergistic, each increasing the pathogenicity of the other.[45] HIV infection increases the susceptibility to developing active disease after infection with *M. tuberculosis,* and immune activation by *M. tuberculosis* increases HIV plasma viremia and appears to increase the rate of HIV disease progression and mortality.[46] HIV infection may increase the rate of tuberculosis after treatment completion or cure, in part because of an increased risk of reinfection, especially in highly endemic areas.[47] Dramatic point-source outbreaks of tuberculosis have been reported to occur where HIV-infected persons congregate, both in the United States and in other countries.[48,49] Many of these outbreaks have occurred in health care settings; in the United States, subsequent outbreaks have been prevented by implementation of effective tuberculosis infection control measures.[23,50] Such outbreaks will likely continue to occur in resource-poor areas where such measures have not been implemented. In the United States, HIV-related tuberculosis outbreaks have also been reported in other institutional settings, such as correctional facilities and homeless shelters.[51,52]

Tuberculosis can occur at any stage of HIV infection, but the clinical presentation is affected by the level of immunosuppression. Because *M. tuberculosis* is more virulent than opportunistic pathogens encountered in persons with HIV/AIDS, tuberculosis may be seen at higher $CD4^+$ T cell counts (i.e., > 200 cells/μl) than are generally seen with other opportunistic infections. When tuberculosis occurs early in the course of HIV infection, before severe immunosuppression has developed, the clinical and radiographic features resemble tuberculosis in patients who are HIV seronegative. In patients with more advanced HIV disease and lower $CD4^+$ T cell counts, *M. tuberculosis* tends to produce disease that is more widespread and severe than conventional tuberculosis and that has so-called atypical features [*see Figure 4*]. With progressive immunodeficiency, extrapulmonary involvement becomes increasingly common.

Pulmonary involvement remains common at all stages of HIV disease; however, the radiographic pattern is very different in persons with advanced immunodeficiency, in whom the most

common abnormalities are intrathoracic adenopathy, focal lower or middle lobe infiltrates, and diffuse miliary or nodular infiltrates. This pattern is consistent with a primary tuberculosis–like pattern. Overall, sputum AFB smears are less likely to be positive in patients with pulmonary disease and HIV coinfection than in non–HIV-infected patients, and HIV-infected patients are less likely to have cavitary disease. In one study, 8% of HIV-infected patients with pulmonary tuberculosis had normal chest radiographs.[30] Those with advanced HIV/AIDS frequently have concomitant pulmonary and extrapulmonary or disseminated tuberculosis. Up to 60% of HIV-infected patients with low $CD4^+$ T cell counts (< 200/μl) who develop tuberculosis have involvement of one or more extrapulmonary sites, including diffuse lymphadenitis, disseminated pleural and pericardial disease, or multiorgan involvement. Mycobacteremia and meningitis are also common in patients with advanced HIV infection.

Not surprisingly, given the atypical features of pulmonary tuberculosis in HIV-infected patients, particularly those with low $CD4^+$ T cell counts, delayed or missed diagnoses have been commonly reported. Thus, a high index of suspicion is crucial in making an appropriate diagnosis. The keys to the diagnosis of HIV-related tuberculosis are knowledge of the epidemiology of tuberculosis, recognition of the ways that immunodeficiency changes the clinical presentation, and an assiduous effort to obtain specimens for mycobacterial smear and culture.[53]

Targeted Testing for LTBI

The decrease in tuberculosis cases in the United States has brought a renewed focus on the treatment of LTBI as an important tuberculosis control strategy.[54] Targeted tuberculin testing for LTBI is a critical component of this strategy. Such testing identifies persons who are at high risk for developing tuberculosis and consequently would benefit from treatment of LTBI. (This type of treatment was previously called preventive therapy or chemoprophylaxis.) In immunocompetent persons, the lifetime risk of progression from latent infection with *M. tuberculosis* to active disease ranges from 5% to 10%; by contrast, in persons infected with HIV, the annual risk of disease progression is 10%.

HIV/AIDS is clearly the greatest risk factor for progression to active tuberculosis after infection with *M. tuberculosis*. Other risk factors include infection within the past 2 years (e.g., as indicated by a history of contact with a person known to have tuberculosis), injection drug use, silicosis, and certain other medical conditions and circumstances (e.g., diabetes mellitus, renal failure, certain malignancies, gastrectomy or jejunoileal bypass, solid-organ transplantation, or the use of immunosuppressive drugs; identification of LTBI is of particular importance in patients who are to be treated with TNF-α inhibitors such as infliximab, etanercept, or adalimumab).[22] The risk for progression is also higher in immigrants who have arrived in the United States within the past 5 years from areas where there is a high incidence of tuberculosis; in racial or ethnic minorities; in children 4 years of age or younger; and in children and adolescents who are exposed to adults at high risk.

TUBERCULIN SKIN TESTING

Until recently, the tuberculin skin test was the only available diagnostic test for LTBI, and it remains the most commonly used test. Tuberculin skin testing has a number of important limitations (see below), and new, improved diagnostic tests for LTBI (including those that can distinguish between infection with *M. tuberculosis* and *M. bovis* BCG [BCG vaccination]) are urgently needed.[55] It is hoped that tuberculin skin testing will be replaced with improved diagnostic tests in the coming years.

Tuberculin skin testing should be performed only on persons at increased risk for tuberculosis.[1] False positive results are common when the test is used in populations with a low prevalence of tuberculosis.[23,28] False positives also occur in persons vaccinated with BCG vaccine or in persons who are sensitized to environmental mycobacteria. False negative reactions are common in immunosuppressed persons and in those with overwhelming tuberculosis disease. In addition, testing is inconvenient, in that patients must return 48 to 72 hours after placement to have the result read.

The Mantoux method should be used for tuberculin skin testing; there is no role for multiple-puncture tests, such as the Tine test.[1] The standard test material used in the Mantoux is intermediate-strength (5 tuberculin units) purified protein derivative (PPD). There are two commercially available PPD preparations in the United States, Aplisol and Tubersol; these have similar sensitivity and specificity in at-risk populations, but Tubersol is slightly more specific and therefore may be of benefit when testing low-risk populations (e.g., certain groups of health care workers who are required to undergo regular tuberculin testing).[56] In addition, switching from Tubersol to Aplisol has been associated with false positive results.[57]

In reading a tuberculin skin test, the diameter of induration rather than of erythema should be determined and recorded. The criterion for a positive test (i.e., induration of ≥ 5 mm, ≥ 10 mm, or ≥ 15 mm) varies according to the population group to which the patient belongs, and the choice of criterion is influenced by the patient's likelihood of becoming infected with *M. tuberculosis* and the risk of developing active disease if infected[28] [*see Table 1*]. Anergy testing along with tuberculin testing is not routinely recommended, especially in HIV-infected patients.[1,28] In addition to its use as a diagnostic test for LTBI, a positive test may provide additional support for the diagnosis of active tuberculosis in culture-negative cases when there is a high clinical index of suspicion.

All patients with a positive tuberculin skin test must be evaluated for evidence of active disease with a chest radiograph. In addition, sputum specimens should be tested if symptoms suggestive of tuberculosis are present or abnormalities are found on the chest radiograph.

Repeated tuberculin skin testing will not cause a truly tuberculin-negative person (i.e., one who has not been infected with *M. tuberculosis* or sensitized to other mycobacteria) to become tuberculin positive.[58] In some persons with LTBI, the ability to react to tuberculin skin testing diminishes over time; administration of a tuberculin skin test to such persons can restore reactivity, thereby boosting the response to future tests.[59] Boosting is believed to result from recall of waned, cell-mediated immunity; it is common in persons older than 55 years and in those born outside the United States who have been vaccinated with BCG. Two-step testing is done to avoid interpreting the boost as a recent conversion and new infection in persons who will be undergoing serial testing. If the reaction to the first tuberculin skin test is negative, the test is repeated in 1 to 3 weeks. Two-step testing should be performed when initially testing persons who have not had a test in the previous 12 months and who will be subject to regular testing in the future, such as health care workers and employees and residents of group settings.

NEW DIAGNOSTIC TESTS FOR LTBI

Because of the limitations of the tuberculin skin test, new diagnostic tests for latent tuberculosis infection are needed.[55] A number of tests are under development. Two assays that utilize peripheral blood are commercially available: a whole-blood IFN-γ release assay (QuantiFERON-TB Gold, Cellestis Ltd, Victoria, Australia), which was approved by the Food and Drug Administration in December 2004, and an enzyme-linked immunospot assay (T SPOT-TB, Oxford Immunotec, Oxford, England), which is approved in Europe.[54,60,61] The Centers for Disease Control and Prevention has published guidelines for the use of the first-generation QuantiFERON-TB assay,[62] which is no longer available, and will be publishing guidelines for use of the second-generation assay, QuantiFERON-TB Gold, which uses tuberculosis-specific antigens, in selected patient populations. The advantages of T cell–based IFN-γ assays are that testing can be accomplished with a single patient visit, the test assesses responses to multiple antigens simultaneously, and the test does not boost anamnestic immune responses. Limitations of the currently available tests include the need to draw blood and to process the sample within 12 hours after collection.

It is hoped that the newer-generation tests that use *M. tuberculosis*–specific antigens will have improved utility, which should lead to wider availability and use of these tests.[60] However, prospective studies are needed to determine whether IFN-γ responses are predictive of high risk of progression to active tuberculosis, to gauge the utility of such tests in specialized subgroups of patients (including children and HIV-infected persons), and to ascertain whether treatment of LTBI on the basis of the results of IFN-γ responses will reduce the tuberculosis burden in low-incidence areas such as the United States.[54]

Table 1 Criteria for a Positive Tuberculin Skin Test by Risk Group[97]

Reaction size (induration) ≥ 5 mm, plus any of the following risk factors:
HIV infection
Recent contact with a patient with infectious TB
Fibrotic changes on chest x-ray consistent with prior TB
Organ transplantation, treatment with tumor necrosis factor–α inhibitors (e.g., infliximab, etanercept, adalimumab), or other immunosuppression (treatment with ≥ 15 mg/day of prednisone or an equivalent dose of another corticosteroid for 1 mo or longer*)
Reaction size (induration) ≥ 10 mm, plus any of the following risk factors:
Recent immigration (within the past 5 yr) from a country with a high prevalence of TB
Injection-drug use
Residence or employment† in a high-risk congregate setting: prison or jail; nursing home or other long-term facility for the elderly; hospital or other health care facility; residential facility for AIDS patients; homeless shelter
Employment in a mycobacteriology laboratory
High-risk clinical conditions: silicosis; diabetes mellitus; chronic renal failure; some hematologic disorders (e.g., leukemias and lymphomas); other specific malignancies (e.g., carcinoma of the head or neck, lung carcinoma); weight loss of ≥ 10% ideal body weight; gastrectomy; jejunoileal bypass
Age < 4 yr or, in an infant, child, or adolescent, exposure to a high-risk adult
Reaction size (induration) ≥ 15 mm

*Risk of TB in patients treated with corticosteroids increases with higher dose and longer duration.

†For persons who are otherwise at low risk and are tested at the start of employment, induration of ≥ 15 mm is considered positive.

TB—tuberculosis

Diagnosis

The keys to the diagnosis of tuberculosis are a high index of suspicion and familiarity with the range of clinical presentations that may occur, including atypical ones in HIV-infected patients that often reflect primary disease rather than reactivation of latent tuberculosis infection.[63]

Unfortunately, delay in diagnosis is common; such delays can increase the risk of a poor outcome and lead to further transmission of tuberculosis, including the precipitation of outbreaks in health care and institutional settings.[6,64-66]

CHEST RADIOGRAPHY

Patients presenting with clinical manifestations suggestive of tuberculosis with pulmonary involvement should have a chest radiograph performed. In immunocompetent patients, the chest radiograph may show upper-lobe disease, often with cavitation [*see* Pulmonary Tuberculosis, *above*]. HIV-infected patients, especially those with advanced disease and low $CD4^+$ T cell counts, are less likely to have cavitation visible on chest radiographs, regardless of the duration of symptoms. The longer the time since the onset of symptoms, the more likely it is that cavitation will be present. In HIV-infected patients, especially those with advanced disease who have a low $CD4^+$ T cell count, there is a greater likelihood of atypical findings on chest x-ray that reflect primary disease; such findings may include lower-zone infiltrates or hilar or mediastinal adenopathy [*see* Tuberculosis in HIV-Infected Patients, *above*].

Although the findings on chest radiography may be suggestive of tuberculosis, definitive diagnosis requires identification of *M. tuberculosis* through culture. In addition, a positive culture for *M. tuberculosis* is a prerequisite for susceptibility testing.

MICROBIOLOGY STUDIES

A number of different diagnostic tests for tuberculosis are available. However, AFB smear and culture are critical in the evaluation of a patient with suspected tuberculosis.

Acid-Fast Bacteria Smear Microscopy

The Kinyoun and the Ziehl-Neelsen basic fuchsin stains are the traditional methods used for visualizing mycobacteria in clinical specimens. In the United States and other developed countries, an auramine-rhodamine stain with fluorescent microscopy is used because it is more sensitive and less time consuming than the carbol-fuschin stain (e.g., the Ziehl-Neelson stain). Throughout most of the world, AFB smear microscopy is the major diagnostic tool for tuberculosis. In the United States and other developed countries, AFB culture, which is more sensitive than an AFB smear, is also used in conjunction with smear microscopy. AFB smear microscopy has a sensitivity of only about 50% to 60% in culture-confirmed cases, in part because a positive smear requires a sputum sample containing 5,000 to 10,000 AFB/μl, whereas a positive AFB sputum culture requires only 10 to 100 AFB/μl.[28] Another limitation is that smear microscopy cannot distinguish *M. tuberculosis* from other mycobacteria.

A presumptive diagnosis of tuberculosis can be made in the setting of a positive AFB smear and clinical manifestations con-

sistent with the disease. Culture confirmation is necessary for a definitive diagnosis; NAA tests, which can be performed directly on clinical specimens, can also provide confirmation of *M. tuberculosis* on AFB smear–positive respiratory specimens.

Mycobacterial Culture

A definitive diagnosis of tuberculosis generally depends on the isolation and identification of *M. tuberculosis* from a clinical specimen; most often this is a sputum sample from a patient with pulmonary disease. Conventional culture of mycobacteria with solid media requires incubation for 3 to 6 weeks. Use of broth-based media can result in recovery of a positive culture 10 to 14 days sooner than with solid media.[67] Broth-based media are also preferable because they are more sensitive than solid media, although neither type of medium recovers all isolates.[67,68] For these reasons, it is recommended that a broth system be used for primary mycobacterial culture but that a solid medium also be inoculated.

DNA probes can be used to rapidly identify colonies of *M. tuberculosis* complex (i.e., *M. tuberculosis, M. bovis, M. africanum,* and *M. microti*) and have replaced biochemical tests in most laboratories. Commercially available probes can identify the *M. tuberculosis* complex but do not differentiate *M. tuberculosis* from other members of the complex; they can also identify *M. avium* complex, *M. kansasii*, and *M. gordonae*. Probes provide species identification within a few hours with nearly 100% accuracy if sufficient growth is tested. However, a positive culture is required before the probe can be used for species identification. High-performance liquid chromatography can also be used to determine the mycobacterial species, but it is usually available only at large public health and reference laboratories.

Susceptibility Testing

Initial isolates from all patients should be tested for drug susceptibility to identify an effective antituberculous regimen.[28] In addition, drug susceptibility tests should be repeated if the patient continues to produce culture-positive sputum after 3 months of adequate therapy or if the patient does not respond clinically to therapy. It has been proposed that susceptibility test results for the first-line antituberculous drugs be reported within an average of 28 days of receipt of the specimen in the laboratory; this requires the use of broth-based media for identification and susceptibility testing. According to the National Committee for Clinical Laboratory Standards Subcommittee for Antimycobacterial Susceptibility Testing, isolates of *M. tuberculosis* should be tested for susceptibility to isoniazid (at two concentrations), rifampin, ethambutol, and pyrazinamide (the latter three at one concentration each).[69] Such testing provides comprehensive information regarding the initial four-drug therapy recommended for treatment of most tuberculosis patients in the United States.

If resistance to rifampin or to any two first-line drugs is found, the isolate should be tested for susceptibility to second-line drugs (e.g., capreomycin, ethionamide, kanamycin, ofloxacin, *para*-aminosalicylic acid, rifabutin, and streptomycin) and for susceptibility to ethambutol at a higher concentration than was used initially.[69,70] Second-line drug testing can be done only on solid media, preferably using the agar proportion method. This process is slow; it can take up to 2 months for results to become available.

Nucleic Acid Amplification Tests

NAA techniques can be used to identify mycobacterial DNA or RNA of *M. tuberculosis* in clinical specimens (e.g., sputum or other respiratory specimens) and provide immediate confirmation that a patient has tuberculosis. NAA testing has been most commonly used to confirm the diagnosis of tuberculosis in patients who have positive AFB smears on sputum or other respiratory specimens. Two commercially available tests are approved by the FDA for use on respiratory specimens.[68,71] The tests are rapid, taking less than 6 hours, and are performed directly on clinical specimens. In sputum and respiratory specimens that are AFB smear–positive, NAA tests have sensitivities and specificities greater than 95%; in AFB smear–negative specimens, specificity remains above 95% but sensitivity is much lower, often less than 50%.[24] The FDA has approved NAA tests in conjunction with cultures of respiratory specimens from patients who have not been treated for tuberculosis.[71] NAA tests have been performed on nonrespiratory specimens, although they are not FDA approved for this use. The performance of NAA tests on nonrespiratory specimens has varied; the sensitivity appears to be less than that for respiratory specimens.[68] NAA tests are particularly useful when the positive predictive value of a positive AFB smear of sputum for *M. tuberculosis* is low. This is the case in settings where recovery of nontuberculous mycobacteria is common, such as with HIV-infected patients, especially those with advanced disease. Currently, NAA tests cannot replace conventional methods for the diagnosis and management of tuberculosis. In hospitalized patients, the AFB smear is used to assess infectivity and the need to institute isolation precautions against airborne infection. Culture must be performed to recover the isolate for susceptibility testing. Thus, NAA testing is a supplement to the traditional diagnostic tests, and it results in additional expense—predominantly, laboratory costs for reagents and technicians' time. The increase in laboratory expenditures, however, may be offset by savings elsewhere in the hospital or public health department. Hospitals can benefit from the ability to discharge patients from airborne-infection isolation rooms despite being AFB positive, when negative results on NAA testing show that they do not have tuberculosis. Negative NAA tests also can help avoid unnecessary therapy for tuberculosis and may shorten hospital stays. In public health departments, positive NAA test results can facilitate contact investigations of persons who have tuberculosis.

Additional Diagnostic Tests

Other diagnostic tests may be useful in facilitating the diagnosis of tuberculosis. Sputum induction by ultrasonic nebulization of hypertonic saline may be useful for patients who are not able to expectorate sputum. The yield of an induced sputum test appears to be as good as that of specimens obtained by fiberoptic bronchoscopy, and yield of repeated inductions may be superior.[72,73] Bronchoscopy with bronchoalveolar lavage or biopsy is sometimes performed as a diagnostic test, especially when sputum cannot be obtained or in patients with radiographic abnormalities suggestive of other diagnoses (e.g., bronchogenic carcinoma). It is essential that specimens be submitted to the microbiology laboratory for AFB smear and culture to establish a diagnosis in such cases.

Molecular typing (so-called DNA fingerprinting) of *M. tuberculosis* isolates has proved very useful in furthering the understanding of the epidemiology of tuberculosis. Molecular typing has led to an increase in knowledge regarding the transmission dynamics of *M. tuberculosis*. It has also proved useful in the evaluation of patients with a second episode of tuberculosis, in that it enables one to differentiate relapses from reinfection with a new

strain. Molecular typing is also helpful in evaluating outbreaks and in identifying laboratory cross contamination.[8]

Extrapulmonary Tuberculosis

To establish the diagnosis of extrapulmonary tuberculosis, appropriate specimens should be obtained for AFB staining, mycobacterial culture, and drug susceptibility testing.[27,30] Depending on the clinical circumstances, specimens may include pleural fluid; pericardial or peritoneal fluid; pleural, pericardial, and peritoneal biopsy specimens; lymph node tissue; bone marrow; bone; blood; urine; brain tissue; or cerebrospinal fluid. Blood from patients with HIV infection should be sent for AFB culture when extrapulmonary or disseminated tuberculosis is suspected. Tissue specimens should also be examined microscopically, after routine and AFB staining, but the absence of AFB and of granulomas or even failure to culture *M. tuberculosis* does not necessarily exclude the diagnosis of tuberculosis. In some cases, a presumptive diagnosis of tuberculosis is made on the basis of epidemiologic findings (e.g., close contact of an active case), consistent clinical and radiologic findings, and a positive tuberculin skin test.

Treatment

PRINCIPLES OF TUBERCULOSIS TREATMENT

The goals of antituberculosis therapy are to ensure a cure without relapse, to prevent death, to stop transmission of *M. tuberculosis*, and to prevent the emergence of drug-resistant disease.[4] Therapy is initiated with a multidrug regimen to kill tubercle bacilli rapidly, to minimize or prevent the development of drug-resistant *M. tuberculosis* strains, and to eliminate persistent organisms from host tissue to prevent relapse. Active tuberculosis should never be treated with a single drug, because of the risk of emergence of resistance, and a single drug should never be added to a failing regimen.

The initial therapy for tuberculosis generally consists of a four-drug regimen (isoniazid, rifampin, pyrazinamide, and ethambutol) [*see Table 2*]. Detailed discussion of the pharmacokinetics, pharmacodynamics, and available preparations of these drugs is beyond the scope of this chapter, but reviews of these subjects have been published.[3,74]

Tuberculosis requires prolonged treatment. The minimum length of therapy for the treatment of drug-susceptible tuberculosis is 6 to 9 months with a rifampin-based regimen (so-called short-course therapy). Longer courses of therapy are required for drug-resistant tuberculosis, especially multidrug-resistant disease (i.e., disease caused by *M. tuberculosis* that is resistant to at least isoniazid and rifampin).Treatment of tuberculosis has two phases: initiation (also known as the bactericidal or intensive phase) and continuation (also known as the subsequent sterilizing phase). These phases reflect the current understanding of the pathophysiology of tuberculosis. Three separate subpopulations of *M. tuberculosis* are thought to exist in the host with tuberculosis.[74] The first and largest of the subpopulations consists of rapidly growing extracellular organisms that mainly reside in well-oxygenated cavities (abscesses) containing 10^7 to 10^8 organisms. The second subpopulation resides within poorly oxygenated, closed, solid, caseous lesions (e.g., noncaseating granulomas) containing 10^4 to 10^5 organisms. These organisms are considered semidormant and undergo only intermittent bursts of metabolic activity. The third subpopulation consists of a small number of organisms (fewer than 10^4 to 10^5) believed to be semidormant within acidic environments—both intracellular (e.g., in macrophages) or extracellular within areas of active inflammation and recent necrosis.

Initiation of tuberculosis treatment is usually with a four-drug regimen consisting of isoniazid (also called INH), rifampin, pyrazinamide (PZA), and ethambutol. Isoniazid and rifampin are the two most important antituberculous drugs and are the cornerstones of therapy. PZA is an important first-line drug that is a necessary component for so-called short-course therapy (i.e., 6 to 9 months). Of these agents, isoniazid is the most potent for killing the rapidly multiplying *M. tuberculosis* bacilli (i.e., those in the first subpopulation) during the initial part of therapy; that is, it has early bactericidal activity. Rifampin and ethambutol have less early bactericidal activity than isoniazid but considerably more than PZA, which has weak early bactericidal activity during the first 2 weeks of treatment. The use of drugs that have potent early bactericidal activity reduces the chance of the development of resistance.

The rapidly dividing population of bacilli (i.e., the first subpopulation) is eliminated early in effective therapy, and after 2 months of treatment, about 80% of patients are culture negative. The remaining (i.e., the second and third) subpopulations of *M. tuberculosis* account for treatment failures and relapses and are the reason that prolonged therapy is required for eradication. For the continuation phase of therapy, antituberculous drugs are chosen on the basis of their sterilizing activity, which is defined by a drug's ability to kill bacilli mainly in the second and third subpopulations.The use of drugs that have good sterilizing activity is essential for short-course therapy (e.g., 6-month regimens). Rifampin and PZA have the greatest sterilizing activity, followed by isoniazid and streptomycin. The sterilizing activity of rifampin persists throughout the course of therapy; however, the sterilizing activity of PZA is mainly seen during the initial 2 months of therapy.

Directly Observed Therapy

Successful treatment of tuberculosis depends not only on the correct choice of antimycobacterial drugs but also on the provision of those drugs within a clinical and social framework based on individual patient circumstances.[3] Furthermore, the treatment of tuberculosis is much different from the treatment of other diseases because of the public health implications of the disease. Whether care is provided by a physician in private practice or through a public health program, the provider has the dual responsibility of selecting an appropriate regimen and ensuring that treatment is completed.[3] For that reason, DOT is recommended for all patients diagnosed with active tuberculosis, in that it helps maximize completion rates [*see Figure 6*], decrease risk for emergence of resistance, and enhance tuberculosis control.[75,76] DOT is generally provided by public health agencies.

INITIATION OF THERAPY

The decision to initiate combination antituberculosis chemotherapy (e.g., a four-drug regimen) should be based on epidemiologic information; clinical, pathologic, and radiographic findings; and the results of microscopic examination of AFB-stained sputum smears (or other diagnostic specimens, as appropriate) and cultures for mycobacteria.[3] Given that *M. tuberculosis* is a relatively slow-growing organism and that cultures can take up to 4 to 5 weeks to become positive, empirical therapy with an appropriate multidrug regimen needs to be initiated when there is

Table 2 Recommended Doses and Adverse Effects of Antituberculosis Medications for Adults*

Rank	*Drug (Route)*	*Daily Dose (Maximum Daily Dose)*	*Twice-Weekly Dose (Maximum Dose)†*	*Thrice-Weekly Dose (Maximum Dose)†*	*Adverse Effects*
First-line drugs	Isoniazid (p.o., I.M., I.V.)	5 mg/kg (300 mg)	15 mg/kg (900 mg)	15 mg/kg (900 mg)	Liver enzyme elevation, hepatitis, peripheral neuropathy, central nervous system effects, rash
	Rifampin (p.o., I.V.)	10 mg/kg (600 mg)‡	10 mg/kg (600 mg)‡	10 mg/kg (600 mg)‡	Orange discoloration of secretions and urine, GI upset, hepatitis, immune-mediated toxicity (e.g., thrombocytopenia, renal failure), flulike symptoms, many drug interactions, rash
	Rifabutin (p.o.)	5 mg/kg (300 mg)‡	5 mg/kg (300 mg)‡	5 mg/kg (300 mg)‡	Similar to rifampin; fewer drug interactions
	Pyrazinamide (p.o.)	40–55 kg: 1,000 mg 56–75 kg: 1,500 mg 76–90 kg: 2,000 mg	40–55 kg: 2,000 mg 56–75 kg: 3,000 mg 76–90 kg: 4,000 mg	40–55 kg: 1,500 mg 56–75 kg: 2,500 mg 76–90 kg: 3,000 mg	GI upset; hepatitis; hyperuricemia; arthralgias
	Ethambutol (p.o.)	40–55 kg: 800 mg 56–75 kg: 1,200 mg 76–90 kg: 1,600 mg	40–55 kg: 2,000 mg 56–75 kg: 2,800 mg 76–90 kg: 4,000 mg	40–55 kg: 1,200 mg 56–75 kg: 2,000 mg 76–90 kg: 2,400 mg	Optic neuritis
Second-line drugs	Cycloserine (p.o.)	10–15 mg/kg in two doses (1 g in two doses)§	No data to support intermittent use	No data to support intermittent use	Psychosis, seizures, depression
	Ethionamide (p.o.)	15–20 mg/kg (1 g) h.s., with main meal, or in two divided doses	No data to support intermittent use	No data to support intermittent use	GI upset, hepatotoxicity, hypothyroidism, metallic taste, bloating
	Streptomycin (I.V., I.M.)	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	Ototoxicity (hearing loss, vestibular dysfunction), nephrotoxicity
	Amikacin-kanamycin (I.V., I.M.)	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	Ototoxicity (hearing loss, vestibular dysfunction), nephrotoxicity
	Capreomycin (I.V., I.M.)	15 mg/kg (1 g); 10 mg/kg in persons > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	15 mg/kg (1 g); 10 mg/kg in patients > 59 yr (750 mg)‖	Ototoxicity (hearing loss, vestibular dysfunction), nephrotoxicity, hypokalemia, hypomagnesemia, eosinophilia
	para-Aminosalicylic acid (PAS) (p.o., I.V.)	8–12 g in two or three doses	No data to support intermittent use	No data to support intermittent use	GI upset, hypersensitivity, hepatotoxicity
	Levofloxacin (p.o., I.V.)	500–1,000 mg	No data to support intermittent use	No data to support intermittent use	GI upset, dizziness, cartilage damage at high doses
	Moxifloxacin (p.o., I.V.)	400 mg	No data to support intermittent use	No data to support intermittent use	GI upset, dizziness, cartilage damage at high doses
	Gatifloxacin (p.o., I.V.)	400 mg	No data to support intermittent use	No data to support intermittent use	GI upset, dizziness, cartilage damage at high doses

*See Table 3 for recommended regimens.
†Must be administered by directly observed therapy only.
‡Dose adjustment may be necessary with concomitant use of protease inhibitors or nonnucleoside reverse transcriptase inhibitors.
§Serum concentration measurements of cycloserine are often useful for optimizing doses in individual patients; the goal is a peak concentration of 20–35 mg/dl.
‖The usual dose is 750–1,000 mg, given as a single dose 5–7 days/wk; reduced to 2–3 days/wk after the first 2–4 mo or after culture conversion, depending on the efficacy of other drugs in the regimen.
GI—gastrointestinal

high clinical suspicion of active disease. Therapy should be started before culture confirmation and in some cases before AFB smear microscopy results are known. NAA tests [*see* Nucleic Acid Amplification Tests, *above*] may be useful in selected cases, because of their ability to provide an immediate definitive diagnosis (e.g., confirmation of AFB smear–positive respiratory specimens). The threshold for initiating empirical therapy should be low for patients with potentially life-threatening forms of tuberculosis that can progress rapidly, such as tuberculous meningitis, pericarditis, or miliary disease.

TREATMENT REGIMENS

Guidelines published by the American Thoracic Society (ATS), the CDC, and the Infectious Diseases Society of America (IDSA) outline recommended treatment regimens for use in the United States and other industrialized countries [*see Table 3*].

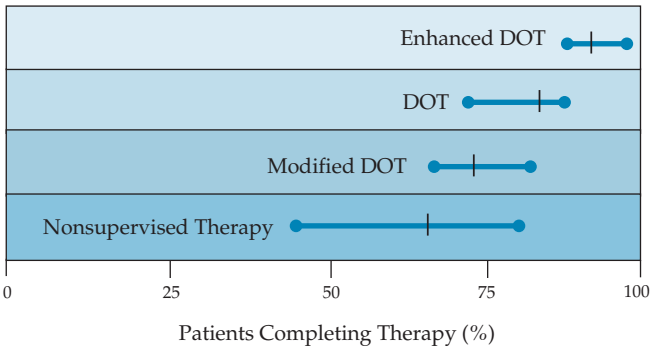


Figure 6 **Impact of directly observed therapy (DOT) on completion rates of therapy for pulmonary tuberculosis.[74] Median completion rates were 61.4% for nonsupervised therapy, 78.6% for modified DOT (i.e., initial inpatient DOT followed by a variety of outpatient strategies, including self-administered medication), 86.3% for DOT, and 91.0% for enhanced DOT (i.e., DOT with multiple incentives and enablers).**

Recommendations are graded and evidence based, using the IDSA–United States Public Health Service rating system.

Drug-Susceptible Pulmonary Disease

When testing confirms drug susceptibility in patients who have been started on an empirical four-drug regimen (e.g., isoniazid, rifampin, PZA, and ethambutol) for pulmonary disease, treatment can be modified accordingly [*see Table 3 and Figure 7*]. PZA can be discontinued after 2 months of therapy (i.e., at the end of the initiation phase). Ethambutol can also be discontinued after 2 months of therapy or as soon as drug susceptibility is confirmed. Isoniazid and rifampin are maintained in the continuation phase (4 more months) for a minimum of 6 months of therapy. Patients at high risk for relapse include those with cavitary pulmonary tuberculosis who remain culture positive after 2 months of therapy.[77] Such patients should have the continuation phase of therapy extended 3 additional months (to 7 months in the continuation phase and a total of 9 months of therapy).

In addition to the total duration of therapy, the number of completed doses is important for treatment success. Doses should be counted and tracked to ensure the proper amount of therapy is delivered. Lack of adherence to antituberculosis therapy is the most common cause of treatment failure, relapse, and the emergence of drug resistance. DOT has been proved to improve completion rates and outcome, and it should be considered the standard of care [*see Figure 6*].[3] Drug-susceptible disease can be successfully treated with antituberculosis therapy administered on an intermittent basis (e.g., twice or thrice weekly), especially in the continuation phase; this tactic facilitates supervision of therapy, thereby helping to improve outcomes. Intermittent therapy (e.g., therapy administered twice or thrice weekly) should be given by DOT only and only to patients with drug-susceptible disease.

HIV-Infected Patients

Because tuberculosis may be the disease that brings an HIV-infected person into the health care system for the first time, all patients diagnosed with tuberculosis should be offered—and strongly encouraged to undergo—HIV serologic testing.[3] The treatment of tuberculosis in patients with HIV coinfection is similar to that in HIV-seronegative patients, with two exceptions. First, HIV-infected patients should not be treated with once-weekly isoniazid-rifapentine in the continuation phase; this regimen is reserved for highly selected HIV-seronegative patients without cavitary disease. In HIV-infected patients, the risk of relapse with this regimen is increased to an unacceptable degree; when relapse occurs, it is often with organisms that have acquired rifamycin resistance.[78] Second, HIV-infected patients with

Table 3 Treatment Guidelines for Drug-Susceptible Pulmonary Tuberculosis in Adults*[3]

	Initial Phase		Continuation Phase			Total Dosage Range (Minimal Duration)	Rating/Evidence†	
Regimen	Drugs	Interval	Regimen	Drugs	Interval		HIV-Negative Patients	HIV-Positive Patients
1	Isoniazid + Rifampin + Pyrazinamide + Ethambutol	Daily or 5 days/wk‡ for 8 wk	1a	Isoniazid + rifampin	Daily or 5 days/wk‡ for 18 wk	182–130 (26 wk)	A/I	A/II
			1b	Isoniazid + rifampin	Twice weekly for 18 wk	92–76 (26 wk)	A/I	A/II§
			1c	Isoniazid + rifapentine	Once weekly for 18 wk	74–58 (26 wk)	B/I	E/I
2	Isoniazid + Rifampin + Pyrazinamide + Ethambutol	Daily or 5 days/wk‡ for 2 wk, then twice weekly for 6 wk	2a	Isoniazid + rifampin	Twice weekly for 18 wk	62–58 (26 wk)	A/II	B/II§
			2b	Isoniazid + rifapentine	Once weekly for 18 wk	44–40 (26 wk)	B/I	B/II
3	Isoniazid + Rifampin + Pyrazinamide + Ethambutol	Three times weekly for 8 wk	3a	Isoniazid + rifampin	Three times weekly for 18 wk	78 (26 wk)	B/I	B/II
4	Isoniazid + Rifampin + Ethambutol	Daily or 5 days/wk‡ for 8 wk	4a	Isoniazid + rifampin	Daily or 5 days/wk‡ for 31 wk	273–195 (39 wk)	C/I	C/II
			4b	Isoniazid + rifampin	Twice weekly for 31 wk	118–102 (39 wk)	C/I	C/II

*From the American Thoracic Society, the Centers for Disease Control and Prevention, and the Infectious Diseases Society of America [*see Table 2 for dosages*].
†Rating levels: A, preferred regimen; B, acceptable alternative; C, offer when A and B cannot be given; D, generally should not be given; E, should never be given. Evidence levels: I=randomized clinical trial; II=data from clinical trials that were not randomized or were conducted in other populations; III=expert opinion.
‡5 days/wk administration is always given by directly observed therapy; rating for these regimens is A/III.
§Not recommended for HIV-positive patients with CD4+ T cell counts < 100 cells/μl.

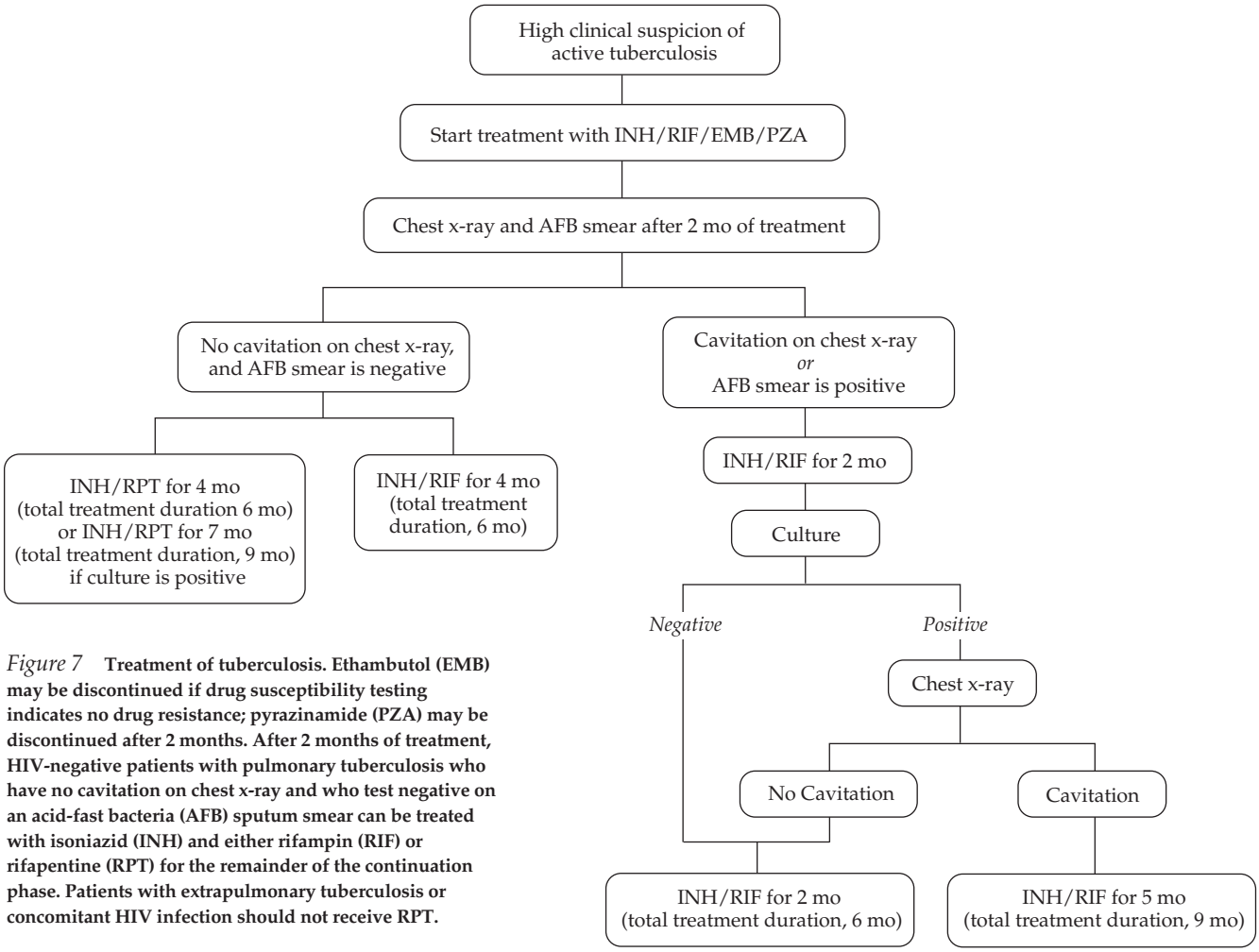


Figure 7 **Treatment of tuberculosis. Ethambutol (EMB) may be discontinued if drug susceptibility testing indicates no drug resistance; pyrazinamide (PZA) may be discontinued after 2 months. After 2 months of treatment, HIV-negative patients with pulmonary tuberculosis who have no cavitation on chest x-ray and who test negative on an acid-fast bacteria (AFB) sputum smear can be treated with isoniazid (INH) and either rifampin (RIF) or rifapentine (RPT) for the remainder of the continuation phase. Patients with extrapulmonary tuberculosis or concomitant HIV infection should not receive RPT.**

$CD4^+$ T cell counts of less than 100/μl should not receive twice-weekly intermittent regimens (e.g., isoniazid-rifampin or isoniazid-rifabutin), because acquired rifamycin resistance has also been reported in this setting.[3,79] Instead, HIV-infected patients with low $CD4^+$ T cell counts should receive daily or thrice-weekly therapy.[3] HIV-infected patients with drug-susceptible pulmonary tuberculosis can generally be treated for 6 months [*see Table 3*]. For HIV-infected patients with tuberculosis who are slow to respond to therapy or who have a suboptimal response (e.g., whose cultures are positive after 2 months of therapy), prolongation of the continuation phase to 7 months (for a total of 9 months of treatment) is suggested.[3]

In the United States, most patients with HIV-related tuberculosis have advanced immunosuppression and high plasma HIV RNA levels at the time of diagnosis.[80] Thus, they meet the criteria for antiretroviral therapy.[81] In addition, the use of antiretroviral therapy during the treatment of tuberculosis in persons with HIV infection may improve tuberculosis treatment outcomes.[3,82] However, strict adherence to antiretroviral therapy is necessary to promote a sustained virologic response; moreover, the use of antiretroviral therapy in HIV-infected patients with tuberculosis is complicated by overlapping toxicity profiles of some antituberculosis and antiretroviral drugs, as well as by complex drug-drug interactions and the occurrence of paradoxical or immune reconstitution reactions.

Paradoxical or immune reconstitution reactions are characterized by exacerbation of symptoms and signs or by radiographic manifestations of tuberculosis. These reactions are more common in HIV-infected patients who are started on antiretroviral therapy early in the course of antituberculosis therapy.[83] Therefore, although there are no data to indicate the best time to start antiretroviral therapy, some experts have recommended delaying its initiation, if possible, until after the patient has received 1 to 2 months of antituberculosis therapy.[3]

The use of antiretroviral therapy during tuberculosis treatment is complex for both the patient and the physician. Thus, there needs to be close coordination of care between the physicians treating each disease. The interaction between rifampin (and other rifamycins) and antiretroviral agents, especially the protease inhibitors, is a major concern and a challenge in treating tuberculosis in HIV-infected patients. Because rifamycins induce the hepatic cytochrome P-450 3A enzyme system, their use leads to reductions in the serum levels of a variety of drugs—in some cases to nontherapeutic ranges. A long list of clinically significant drug-drug interactions involving the rifamycins have been reported, including those with protease inhibitors and nonnucleoside reverse transcriptase inhibitors; there are generally no significant drug-drug interactions with nucleoside reverse transcriptase inhibitors (NRTIs) [*see 7:XXXIII HIV and AIDS*]. Rifampin is the most potent cytochrome P-450 inducer, followed by rifapen-

tine and rifabutin. Conversely, the protease inhibitors are cytochrome P-450 inhibitors that raise rifabutin levels to potentially toxic concentrations and necessitate dose modifications. Rifampin cannot be given with most protease inhibitors, because its use results in low serum levels of these drugs. Rifabutin lowers drug serum levels to a lesser degree than rifampin, and therefore it can be used with certain protease inhibitors.

Possible options in the treatment of tuberculosis in HIV-infected patients include the following: (1) use of a rifampin-based regimen, which can be given to patients receiving antiretroviral therapy with NRTIs and efavirenz; (2) the substitution of rifabutin for rifampin in a multidrug regimen when the patient is receiving antiretroviral therapy with a protease inhibitor; (3) the use of rifampin in a multidrug regimen when antiretroviral therapy cannot be given; and (4) the use of a non–rifamcyin-based regimen in patients receiving antiretroviral drugs, including protease inhibitors. Despite the potential for avoiding drug-drug interactions, however, regimens that do not include a rifamycin are not recommended for patients with HIV infection, because worse outcomes have been reported.[80]

It must be emphasized that recommendations on the use of antiretroviral therapies in HIV-infected patients with tuberculosis continue to evolve. The latest recommendations and information, including acceptable antiretroviral regimens and necessary dose adjustments, are available from CDC on the Internet (www.cdc.gov/nchstp/tb/TB_HIV_Drugs/TOC.htm).[84]

Extrapulmonary Tuberculosis

The basic principles that underlie the treatment of pulmonary tuberculosis also apply to extrapulmonary forms of the disease. A 6-month course of therapy is recommended for treating tuberculosis involving any site except the meninges; for the meninges, a 9- to 12-month regimen is recommended. Prolongation of therapy also should be considered for patients with tuberculosis that is slow to respond, regardless of the site.

The addition of corticosteroids is recommended for patients with tuberculous pericarditis and meningitis, because it improves outcome and decreases mortality.[3,85,86] Evidence-based guidelines for the treatment of extrapulmonary tuberculosis and the adjunctive use of corticosteroids have been developed [*see Table 4*]. Corticosteroids should be given for tuberculous pericarditis during the first 11 weeks of antituberculosis therapy. Corticosteroids do not reduce the risk of the development of constrictive pericarditis, however. For patients with tuberculous meningitis, adjunctive dexamethasone is recommended.

Tuberculosis in Pregnant Patients

Tuberculosis that is discovered during pregnancy should be treated without delay. Because tuberculosis can spread to the fetus, treatment in pregnant women should be initiated whenever the probability of maternal disease is moderate to high.

The initial treatment regimen in pregnant patients consists of isoniazid, rifampin, and ethambutol. Consideration should be given to including PZA in the initial regimen as well: PZA has not been widely used in the United States to treat pregnant women with tuberculosis, but it is recommended for use in this setting by the WHO and the International Union Against Tuberculosis and Lung Disease, as well as some public health agencies in the United States.[3,87] PZA is recommended for use in all HIV-infected pregnant patients and in pregnant patients who are thought to be at high risk for drug-resistant tuberculosis (pending susceptibility test results). If PZA is not included in the regimen, the minimum duration of treatment is 9 months. Supplemental pyridoxine (vitamin B_6), 25 to 50 mg a day, is indicated for all pregnant women taking isoniazid, to prevent peripheral neuropathy. Aminoglycosides and fluoroquinolones should be avoided in pregnancy because of potential adverse effects on the fetus.

Drug-Resistant Tuberculosis

Treatment of drug-resistant tuberculosis, especially MDR-TB, is quite challenging and should be done by, or in close consultation with, an expert in this subject. Recommendations for the treatment of drug-resistant tuberculosis have been developed [*see Table 5*]. Treatment of isolated isoniazid resistance can be accomplished with a 6-month regimen of daily rifampin, PZA, and ethambutol. Treatment of isolated rifampin resistance requires a minimum of 12 months of therapy with a regimen such as isoniazid, PZA, ethambutol, and a fluoroquinolone. Treatment of MDR-TB (resistance to both isoniazid and rifampin) requires 18 to 24 months, depending on the full resistance pattern, and is associated with significantly higher morbidity and mortality than drug-susceptible disease. Treatment of MDR-TB requires the use of second-line drugs, which have less in vitro activity and significantly more toxicity than first-line drugs [*see Table 2*].

Common errors that lead to the emergence of drug resistance include the addition of a single drug to a failing regimen, failure to identify preexisting or acquired drug resistance, initiation of

Table 4 Evidence-Based Guidelines for Duration of Therapy for Drug-Susceptible Extrapulmonary Tuberculosis and Adjunctive Use of Corticosteroids*[3,85]

Site	*Duration of Antimicrobial Therapy (Rating)*	*Adjunctive Corticosteroids (Rating)*	*Corticosteroid Regimens*
Lymph node	6 mo (A/I)	Not recommended (D/III)	Pericarditis: prednisone, 60 mg/day, wk 1–4; 30 mg/day, wk 5–8; 15 mg/day, wk 9–10; 5 mg/day, wk 11
Bone and joint	6–9 mo (A/I)	Not recommended (D/III)	
Pleural disease	6 mo (A/II)	Not recommended (D/III)	
Pericarditis	6 mo (A/II)	Strongly recommended (A/I)	
CNS, including meningitis	9–12 mo (B/II)	Strongly recommended (A/I)	Meningitis: dexamethasone for 6 wk; in children < 25 kg, 8 mg/day × 3 wk; in children > 25 kg and in adults, 12 mg/day × 3 wk; in all patients, dose is tapered over the next 3 wk
Disseminated disease	6 mo (A/II)	Not recommended (D/III)	
Genitourinary	6 mo (A/II)	Not recommended (D/III)	
Peritoneal	6 mo (A/II)	Not recommended (D/III)	

*Preferred duration of therapy for extrapulmonary tuberculosis caused by drug-resistant organisms is unknown.
CNS—central nervous system

Table 5 Potential Regimens for the Treatment of Drug-Resistant Tuberculosis[3]

Pattern of Drug Resistance	*Suggested Regimen (Alternative Choice)*	*Duration of Treatment (mo)*	*Comments*
Isoniazid (± streptomycin)	Rifampin, pyrazinamide, ethambutol; addition of a fluoroquinolone* may strengthen regimen for patients with extensive disease	6	Use at least three other drugs; include rifampin, pyrazinamide, and ethambutol or streptomycin
Rifampin	Isoniazid, ethambutol, and a fluoroquinolone,* plus pyrazinamide for the first 2 mo; an injectable agent† may be included for the first 2–3 mo for patients with extensive disease	12–18	Isoniazid, pyrazinamide, and streptomycin for 9 mo is an alternative regimen; however, prolonged therapy with an injectable agent may not be feasible or desirable, and an all-oral regimen should be as effective; some experts continue pyrazinamide throughout the course of therapy
Isoniazid and rifampin (± streptomycin)	A fluoroquinolone,* pyrazinamide, ethambutol, and an injectable agent,† ± an alternative agent‡	18–24	Extended treatment is needed to lessen the risk of relapse; in patients with extensive disease, the addition of an alternative agent may be prudent to lessen the risk of failure and additional acquired drug resistance; consider resectional surgery as adjunct to chemotherapy
Isoniazid, rifampin (± streptomycin), and ethambutol or pyrazinamide	A fluoroquinolone* (ethambutol or pyrazinamide, if active), injectable agent,† and two alternative agents‡	24	Use the first-line agents to which the strain is susceptible; add two or more alternative agents in patients with extensive disease; consider resectional surgery as adjunct to chemotherapy

Note: treatment of drug-resistant tuberculosis should be carried out by, or in consultation with, a physician with expertise and experience in treating drug-resistant disease [*see Table 2 for dosages*].
*For example, levofloxacin, moxifloxacin, gatifloxacin.
†Injectable agents may include aminoglycosides (streptomycin, amikacin, or kanamycin) or the polypeptide capreomycin.
‡Alternative agents are ethionamide, cycloserine, *para*-aminosalicylic acid, clarithromycin, amoxicillin-clavulanate, and linezolid.

an inadequate primary regimen, failure to identify and address noncompliance, and use of monotherapy for active disease (in cases in which therapy for latent tuberculosis was prescribed but active disease was present).[88] In patients suspected of having MDR-TB (e.g., those who failed to complete an earlier regimen or who followed their therapy erratically, those with recent exposure to an MDR-TB case, or those from extremely high-risk areas), the physician should consider initiating therapy with extended empirical regimens, pending culture results. This is especially the case for patients with extensive pulmonary disease or extrapulmonary disease such as tuberculous meningitis or miliary disease. Documented MDR-TB requires treatment with at least four drugs (and more if possible) to which the organisms are susceptible, such as three oral drugs and one injectable agent [*see Table 5*].

The role of surgery in MDR-TB has not been examined in randomized studies but is thought by some experts to be beneficial in selected cases. In one case series, surgical resection in conjunction with fluoroquinolone therapy was associated with improved microbiologic and clinical outcomes in 205 patients with MDR-TB.[89] Surgery should be deferred until the patient has completed several months of intensive chemotherapy, and it should be performed by an experienced surgeon.[3]

MONITORING RESPONSE TO TREATMENT

For patients undergoing treatment of pulmonary tuberculosis, a sputum specimen for AFB smear and culture should be obtained at least monthly until two consecutive specimens are culture negative.[3] It is essential to obtain an AFB sputum smear and culture after 2 months of therapy, because of their value in predicting risk of relapse. Drug susceptibility tests should be repeated on *M. tuberculosis* isolates from patients whose cultures are positive after 3 months of treatment.

In patients with pulmonary tuberculosis, a repeat chest radiograph should be taken after 2 months of therapy; more frequent chest radiographs are not indicated. However, a chest radiograph taken at the completion of therapy can be useful for providing a baseline against which subsequent films can be compared.

Bacteriologic monitoring is more difficult in patients with extrapulmonary disease. In these cases, the response to treatment often must be assessed clinically because it is not feasible to obtain follow-up cultures.

All patients undergoing treatment of tuberculosis should be seen on a monthly basis; at each visit, they should undergo a clinical evaluation to identify possible adverse effects of the antituberculosis medications and to assess adherence. Baseline liver function tests, creatinine level, and platelet count should be obtained on all patients. In patients taking first-line antituberculosis drugs, ATS/CDC/IDSA guidelines do not recommend monthly liver or renal function tests during treatment unless baseline abnormalities were present or there are clinical reasons to obtain the tests.[3] Patients taking ethambutol should be questioned monthly regarding visual disturbances; monthly testing of visual acuity and color vision is recommended for those treated with dosages higher than 20 mg/kg/day or for those who require ethambutol for more than 2 months.

CONTACT INVESTIGATION AND REPORTING OF TUBERCULOSIS CASES

Physicians are required by law to report tuberculosis cases to their local public health agency. In some hospitals, this is handled by the infection control department, but the physician should ensure that the case has been reported expeditiously, so that the local health department can contact the patient while he or she is still hospitalized. This will help to ensure that the patient is not lost to follow-up after discharge. Discharge planning should be carried out in a collaborative fashion with the involvement of the public health department, which should have the resources and the ability to provide DOT to patients with tuberculosis in the outpatient setting. The local public health agency is

responsible for conducting a contact investigation to identify others who have been exposed to an infectious patient (e.g., in the home, at work, and in other social settings). This can lead to the identification of newly infected contacts, for whom treatment of latent tuberculosis is a priority, and it can lead to other potential cases. In addition, when the patient is a child, the contact investigation can lead to the identification of the source case. Priority for contact investigations should be given to instances in which infants or HIV-infected persons (or other highly immunocompromised persons) have been exposed, given the rapid progression from infection to active disease in these settings.

LTBI

Therapy for latent tuberculosis can markedly reduce the risk of progression to active disease and is recommended for patients who are at increased risk for disease progression. Patients with LTBI (i.e., those who test positive on tuberculin skin testing or other diagnostic tests but whose chest radiograph is negative and who have no signs or symptoms of tuberculosis) who are at increased risk for progression to active disease should be encouraged to undergo therapy.

The CDC and ATS have issued guidelines on the treatment of latent tuberculosis [*see Table 6*]. The preferred regimen is a 9-month course of isoniazid; 6 months of isoniazid is an alternative in HIV-seronegative adults. The recommendation for this duration of therapy comes from reanalysis of data from older trials.[90] Rifampin taken for 4 months is an alternative for the treatment of LTBI and is recommended for adults suspected of harboring an isoniazid-resistant strain of *M. tuberculosis*.

A 2-month regimen of rifampin plus PZA for the treatment of LTBI is not recommended, because of its unacceptably high rate of hepatotoxicity in these patients. A CDC survey suggests that the risk of death with this regimen is nearly 1 in 1,000 persons, and the rate of hospitalization is 3 in 1,000 persons.[91] Rifampin and PZA, however, remain an important component of a multidrug regimen for patients with active tuberculosis. Isoniazid may also be given twice weekly in DOT to facilitate adherence in institutional settings or when resources are available.

Patients receiving therapy for LTBI should have an initial clinical evaluation, followed by follow-up evaluations at least monthly; no more than 1 month's supply of medication should be dispensed at a time. The monthly clinical evaluation should include questioning about side effects and a brief clinical assessment for signs of hepatitis. Although adverse reactions to isoniazid are not common, they can be serious. Hepatotoxicity is the most important side effect. However, at a tuberculosis clinic in Seattle, hepatotoxicity occurred in only 0.15% of patients who completed isoniazid monotherapy for LTBI—a rate much lower than those reported in earlier studies.[92] The rate of isoniazid-related hepatotoxicity has been estimated to be 1 to 3 per 1,000 patients. Age is a risk factor: isoniazid-induced hepatotoxicity is rare in patients younger than 20 years, but the rate increases with advancing age. Risk may also be higher in patients with underlying liver disease (including hepatitis C), in those with a history of heavy alcohol consumption, and in the postpartum period (especially for Hispanic women). Asymptomatic, and generally transient, elevations of the aminotransferase level can occur in 10% to 20% of patients taking isoniazid for LTBI. The risk of fatal hepatitis from isoniazid is currently reported to range from 0 to 0.3 (median, 0.04) per 1,000 patients.[91,92] Death has been associated with continued administration of isoniazid despite onset of hepatitis symptoms. The drug should be discontinued when levels of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) exceed five times normal in asymptomatic patients or three times normal in those with symptoms.

Patients should be advised to discontinue isoniazid at the onset of symptoms that are consistent with hepatitis (e.g., nausea, loss of appetite, and dull midabdominal pain) and to immediately seek medical evaluation should these symptoms occur. Liver function tests should be obtained for any patient who develops symptoms suggestive of hepatitis. We recommend baseline liver function tests for all adult patients at the start of therapy for LTBI. However, ATS/CDC guidelines recommend baseline laboratory testing only for patients whose initial evaluation suggests a liver disorder and for those at increased risk for hepatotoxicity, including HIV-infected patients, pregnant women,

Table 6 CDC Guidelines for the Treatment of Latent Tuberculosis in Adults[97]

Drug	*Dosage and Duration (Maximum Dose)*	*Rating/Evidence Level*		*Comments*
		HIV-Negative Patients	*HIV-Positive Patients**	
Isoniazid	5 mg/kg (300 mg) daily for 9 mo	A/II	A/II	Preferred for adults and children; indicated for HIV-infected patients and those with fibrotic lesions on chest x-ray; in HIV-infected patients, may be given concurrently with antiretroviral treatment; DOT must be used with twice-weekly dosing
	900 mg twice weekly for 9 mo	B/II	B/II	
Isoniazid	5 mg/kg (300 mg) daily for 6 mo	B/II	C/I	Alternative for HIV-negative patients; DOT must be used with twice-weekly dosing
	900 mg twice weekly for 6 mo	B/II	C/I	
Rifampin	10 mg/kg (600 mg) daily for 4 mo	B/II	B/III	Alternative regimen; for contacts of patients with isoniazid-resistant, rifampin-susceptible TB; HIV-infected patients taking protease inhibitors or certain NNRTIs cannot take rifampin and must instead use rifabutin; the combination of rifampin and pyrazinamide is not recommended (D/II) for treatment of latent TB infection because of high risk of hepatotoxicity

*Current data on interactions with HIV-related drugs are available at http://www.aidsinfo.nih.gov/guidelines.
CDC—Centers for Disease Control and Prevention DOT—directly observed therapy NNRTI—nonnucleoside reverse transcriptase inhibitor

women in the immediate postpartum period (i.e., within 3 months after delivery), patients with a history of chronic liver disease (e.g., hepatitis B, hepatitis C, alcoholic hepatitis, or cirrhosis), regular users of alcohol, and patients at risk for chronic liver disease.[1] Baseline laboratory tests should also be obtained in patients who are taking other potentially hepatotoxic medications for chronic medical conditions. Active hepatitis and end-stage liver disease are relative contraindications to the use of isoniazid for treatment of latent tuberculosis. Routine laboratory monitoring (e.g., monthly AST or ALT measurement) during treatment of latent tuberculosis is recommended for persons whose baseline liver function test results are abnormal and for other patients at risk for hepatic disease (see above).

Peripheral neuropathy is also a side effect of isoniazid. It is relatively uncommon, but the risk is increased in persons with a nutritional deficiency, as well as those with diabetes mellitus, HIV infection, renal failure, or alcoholism, and in women who are pregnant or breast-feeding. Pyridoxine (25 to 50 mg a day) is recommended for patients with these risk factors to help prevent neuropathy. Some clinicians routinely give supplemental pyridoxine to all patients taking isoniazid.

VACCINES

BCG vaccination involves use of a live attenuated strain of *M. bovis*. The primary benefit of BCG administration appears to be in preventing disseminated tuberculosis and tuberculous meningitis in young children; variable efficacy has been reported in adults. BCG has had little effect on the global epidemiology of tuberculosis. BCG is not recommended for use in the United States but is widely used outside of the United States, especially in developing countries. Interestingly, in the tropics, administration of BCG has been associated with a decreased risk of leprosy.[93] The vaccine can produce a positive tuberculin test in recipients, and because of the low incidence of new tuberculous infections in the United States, case finding and treatment of latent tuberculous infection are considered more efficient and effective strategies. Interpretation of a tuberculin skin test reaction is not changed for patients who have received BCG,[1] because tuberculin sensitivity tends to wane considerably after BCG vaccination, and BCG is often given in areas where tuberculosis is endemic. Given that many BCG-vaccinated persons come from areas with a high prevalence of tuberculosis, it is important that those who have significant reactions to the tuberculin skin test be evaluated for the presence of disease and managed accordingly. Appropriate follow-up includes a careful medical history, a chest x-ray to rule out disease, and evaluation for treatment of latent tuberculosis. Newer diagnostic tests are needed for distinguishing between infection with *M. tuberculosis* and immunization with BCG.

Hospital-Based Prevention

Tuberculosis infection control efforts, utilizing a hierarchy of measures recommended by the CDC, have proved effective in preventing nosocomial transmission of tuberculosis.[22,93,94] Atop this hierarchy are administrative controls, which include high suspicion for tuberculosis, careful screening of patients, precautions against airborne infection for patients suspected of having tuberculosis, and prompt diagnosis and initiation of effective therapy. Engineering controls and respiratory protection constitute the second and third tiers of the hierarchy of control measures. Guidelines for implementing a tuberculosis infection control program in health care facilities have been published.[95]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Targeted tuberculin testing and treatment of latent tuberculosis infection. American Thoracic Society, Centers for Disease Control and Prevention. Am J Respir Crit Care Med 161:S221, 2000

2. Pape JW, Jean SS, Ho JL, et al: Effect of isoniazid prophylaxis on incidence of active tuberculosis and progression of HIV infection. Lancet 342:268, 1993

3. Treatment of tuberculosis. American Thoracic Society, Centers for Disease Control and Prevention, Infectious Diseases Society of America. MMWR Recomm Rep 52(RR-11):1, 2003
http://www.cdc.gov/mmwr/PDF/rr/rr5211.pdf

4. Frieden TR, Sterling TR, Munsiff SS, et al: Tuberculosis. Lancet 362:887, 2003

5. Global tuberculosis control: surveillance, planning, financing. WHO Report 2004 (WHO/HTM/TB/2004.331). World Health Organization. WHO, Geneva, 2004

6. Trends in tuberculosis—United States, 2004. MMWR Morb Mortal Wkly Rep 54:245, 2005

7. Geng E, Kreiswirth B, Driver C, et al: Changes in the transmission of tuberculosis in New York City from 1990 to 1999. N Engl J Med 346:1453, 2002

8. Barnes PF, Cave MD: Molecular epidemiology of tuberculosis. N Engl J Med 349:1149, 2003

9. Sotir MJ, Parrott P, Metchock B, et al: Tuberculosis in the inner city: impact of a continuing epidemic in the 1990s. Clin Infect Dis 29:1138, 1999

10. Havlir DV, Barnes PF: Tuberculosis in patients with human immunodeficiency virus infection. N Engl J Med 340:367, 1999

11. Cobelens FGJ, van Deutekom H, Draayer-Jansen IW, et al: Risk of infection with *Mycobacterium tuberculosis* in travelers to areas of high tuberculosis endemicity. Lancet 356:461, 2000

12. Frieden TR, Sterling T, Mendez Pablos, et al: The emergence of drug-resistant tuberculosis in New York City. N Engl J Med 328:521, 1993

13. Frieden TR, Sherman LF, Maw KL, et al: A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA 276:1229, 1996

14. Behr MA, Warren SA, Salamon H, et al: Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet 353:444, 1999

15. Comstock GW: Tuberculosis in twins: a re-analysis of the Prophit survey. Am Rev Respir Dis 117:621, 1978

16. Stead WW, Senner JW, Reddick WT, et al: Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. N Engl J Med 322:422, 1990

17. Goldfeld AE: Genetic susceptibility to pulmonary tuberculosis in Cambodia. Tuberculosis 84:76, 2004

18. Delgado JC, Baena A, Thim S, et al: Ethnic-specific genetic associations with pulmonary tuberculosis. Infect Dis 186:1463, 2002

19. Boom WH, Canaday DH, Fulton SA, et al: Human immunity to *M. tuberculosis:* T cell subsets and antigen processing. Tuberculosis 83:98, 2003

20. Flynn J: Immunology of tuberculosis and implications in vaccine development. Tuberculosis 84:93, 2004

21. Keane J, Gershon S, Wise RP, et al: Tuberculosis associated with infliximab, a tumor necrosis factor alpha–neutralizing agent. N Engl J Med 345:1098, 2001

22. Tuberculosis in the Workplace. Field MJ, Ed. National Academy Press, Washington, DC, 2001

23. Small PM, Fujiwara PI: Management of tuberculosis in the United States. N Engl J Med 345:189, 2001

24. Pape JW, Jean SS, Ho JL, et al: Effect of isoniazid prophylaxis on incidence of active tuberculosis and progression of HIV infection. Lancet 342:268, 1993

25. Lambert ML, Hasker E, Van Deun A, et al: Recurrence in tuberculosis: relapse or reinfection? Lancet Infect Dis 3:282, 2003

26. Jasmer RM, Bozeman L, Schwartzman K, et al: Recurrent tuberculosis in the United States and Canada: relapse or reinfection? Tuberculosis Trials Consortium. Am J Respir Crit Care Med 170:1360, 2004

27. Reported Tuberculosis in the United States, 2003. Centers for Disease Control and Prevention. US Department of Health and Human Services, CDC, Atlanta, September 2004
http://www.cdc.gov/nchstp/tb/surv/surv2003/default.htm

28. Diagnostic standards and classification of tuberculosis in adults and children. American Thoracic Society, Centers for Disease Control and Prevention. Am J Respir Crit Care Med 161:1376, 2000
http://www.thoracic.org/adobe/statements/tbadult1-20.pdf.

29. Miller LG, Asch SM, Yu EI, et al: A population-based survey of tuberculosis symptoms: how atypical are atypical presentations? Clin Infect Dis 30:293, 2000

30. Perlman DC, el-Sadr WM, Nelson ET, et al: Variation of chest radiographic patterns in pulmonary tuberculosis by degree of human immunodeficiency virus–related immunosuppression. Clin Infect Dis 25:242, 1997

31. Moussa OM, Eraky I, El-Far MA, et al: Rapid diagnosis of genitourinary tuberculosis by polymerase chain reaction and non-radioactive DNA hybridization. J Urol 164:584, 2000

32. Colmenero JD, Jimenez-Mejias ME, Reguera JM, et al: Tuberculous vertebral osteomyelitis in the new millennium: still a diagnostic and therapeutic challenge. Eur J Clin Microbiol Infect Dis 23:477, 2004

33. Colmenero JD, Jimenez-Mejias ME, Sanchez-Lora FJ, et al: Pyogenic, tuberculous, and brucellar vertebral osteomyelitis: a descriptive and comparative study of 219 cases. Ann Rheum Dis 56:709, 1997

34. Five-year assessment of controlled trials of short-course chemotherapy regimens of 6, 9 or 18 months' duration for spinal tuberculosis in patients ambulatory from the start or undergoing radical surgery. Medical Research Council Working Party on Tuberculosis of the Spine. Int Orthop 23:73, 1999

35. Porkert MT, Sotir M, Moore PP, et al: Tuberculous meningitis at a large inner-city medical center. Am J Med Sci 313:325, 1997

36. Verdon R, Chevret S, Laissy JP, et al: Tuberculous meningitis in adults: review of 48 cases. Clin Infect Dis 22:982, 1996

37. Pai M, Flores LL, Pai N, et al: Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. Lancet Infect Dis 3:633, 2003

38. Humphries MJ: Factors of prognostic significance in Chinese children with tuberculous meningitis. Tubercle 71:161, 1990

39. Hernandez-Albujar S, Arribas JR, Royo A, et al: Tuberculosis radiculomyelitis complicating tuberculosis meningitis: case report and review. Clin Infect Dis 30:915, 2000

40. Farrar DJ, Flanigan TP, Gordon NM, et al: Tuberculous brain abscess in a patient with HIV infection: case report and review. Am J Med 102:297, 1997

41. Heffner JE, Strange C, Sahn SA: The impact of respiratory failure on the diagnosis of tuberculosis. Arch Intern Med 148:1103, 1998

42. Zahar JR, Azoulay E, Klement E, et al: Delayed treatment contributes to mortality in ICU patients with severe active pulmonary tuberculosis and acute respiratory failure. Intensive Care Med 27:513, 2001

43. Mert A, Bilir M, Tabak F, et al: Miliary tuberculosis: clinical manifestations, diagnosis and outcome in 38 adults. Respirology 6:217, 2001

44. Maartens G, Willcox PA, Benatar SR: Miliary tuberculosis: rapid diagnosis, hematologic abnormalities, and outcome in 109 treated adults. Am J Med 89:291, 1990

45. de Jong BC, Israelski DM, Corbett EL, et al: Clinical management of tuberculosis in the context of HIV infection. Annu Rev Med 55:283, 2004

46. Collins KR, Quinones-Mateu ME, Toossi Z, et al: Impact of tuberculosis on HIV-1 replication, diversity, and disease progression. AIDS Rev 4:165, 2002

47. Sonnenberg P, Murray J, Glynn JR, et al: HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. Lancet 358:1687, 2001

48. Jarvis WR: Nosocomial transmission of multidrug-resistant *Mycobacterium tuberculosis*. Am J Infect Control 23:146, 1995

49. Campos PE, Suarez PG, Sanchez J, et al: Multidrug-resistant *Mycobacterium tuberculosis* in HIV-infected persons, Peru. Emerg Infect Dis 9:1571, 2003

50. Blumberg HM: Tuberculosis infection control in healthcare settings. Practical Handbook for Healthcare Epidemiologists, 2nd ed. Lautenbach E, Woeltje K, Eds. Slack Inc, Thorofare, New Jersey, 2004

51. McLaughlin SI, Spradling P, Drociuk D, et al: Extensive transmission of *Mycobacterium tuberculosis* among congregated, HIV-infected prison inmates in South Carolina, United States. Int J Tuberc Lung Dis 7:665, 2003

52. McElroy PD, Southwick KL, Fortenberry ER, et al: Outbreak of tuberculosis among homeless persons coinfected with human immunodeficiency virus. Clin Infect Dis 36:1305, 2003

53. Burman WJ, Jones BE: Clinical and radiographic features of HIV-related tuberculosis. Semin Respir Infect 18:263, 2003

54. Blumberg HM, Leonard MK Jr, Jasmer RM: Update on the treatment of tuberculosis and latent tuberculosis infection. JAMA 293:2776, 2005

55. Ending Neglect: The Elimination of Tuberculosis in the United States. Geiter L, Ed. Committee on the Elimination of Tuberculosis in the United States, Division of Health Promotion and Disease Prevention, Institute of Medicine. National Academy Press, Washington, DC, 2000

56. Villarino ME, Burman W, Wang Y-C, et al: Comparable specificity of 2 commercial tuberculin reagents in persons at low risk for tuberculous infection. JAMA 281:169, 1999

57. Blumberg HM, White N, Parrott P, et al: False-positive tuberculin skin test results among health care workers. JAMA 283:2793, 2000

58. Lee E, Holzman RS: Evolution and current use of the tuberculin test. Clin Infect Dis 34:365, 2002

59. Menzies D: Interpretation of repeated tuberculin tests: boosting, conversion, and reversion. Am J Respir Crit Care Med 159:15, 1999

60. Mazurek GH, Villarino ME: Guidelines for using the QuantiFERON-TB test for diagnosing latent *Mycobacterium tuberculosis* infection. Centers for Disease Control and Prevention. MMWR Recomm Rep 52(RR-2):15, 2003
http://www.cdc.gov/nchstp/tb/pubs/tbfactsheets/250103.htm

61. Barnes PF: Diagnosing latent tuberculosis infection: turning glitter to gold. Am J Respir Crit Care Med 170:5, 2004

62. Ewer K, Deeks J, Alvarez L, et al: Comparison of T-cell–based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. Lancet 361:1168, 2003

63. Bennett CL, Schwartz DN, Parada JP, et al: Delays in tuberculosis isolation and suspicion among persons hospitalized with HIV-related pneumonia. Chest 117:110, 2000

64. Rao VK, Iademarco EP, Fraser VJ, et al: Delays in the suspicion and treatment of tuberculosis among hospitalized patients. Ann Intern Med 130:404, 1999

65. Chin DP, Crane CM, Diul MY, et al: Spread of *Mycobacterium tuberculosis* in a community implementing recommended elements of tuberculosis control. JAMA 283:2968, 2000

66. Tuberculosis outbreak in a community hospital—District of Columbia, 2002. MMWR Morb Mortal Wkly Rep 53:214, 2004

67. Anargyros P, Astill DS, Lim IS: Comparison of improved BACTEC and Lowenstein-Jensen media for culture of mycobacteria from clinical specimens. J Clin Microbiol 28:1288, 1990

68. Woods GL: The mycobacteriology laboratory and new diagnostic techniques. Infect Dis Clin North Am 16:127, 2002

69. Woods GL: Susceptibility testing for mycobacteria. Clin Infect Dis 31:1209, 2000

70. Susceptibility testing of mycobacteria, *Nocardia*, and other aerobic actinomycetes. National Committee for Clinical Laboratory Standards. National Committee for Clinical Laboratory Standards, Wayne, PA, 2000

71. Update: nucleic acid amplification tests for tuberculosis. MMWR Morb Mortal Wkly Rep 49:593, 2000

72. Conde MB, Soares LM, Mello CQ, et al: Comparison of sputum induction with fiberoptic bronchoscopy in the diagnosis of tuberculosis. Am J Respir Crit Care Med 162:2238, 2000

73. Al Zahrani K, Al Jahdali H, Poirier L, et al: Yield of smear, culture and amplification tests from repeated sputum induction for the diagnosis of pulmonary tuberculosis. Int J Tuberc Lung Dis 5:1, 2001

74. Nuermberger E, Grosset J: Pharmacokinetic and pharmacodynamic issues in the treatment of mycobacterial infections. Eur J Clin Microbiol Infect Dis 23:243, 2004

75. Chaulk CP, Kazdanjian VA: Directly observed therapy for the completion of tuberculosis: consensus statement of the public health tuberculosis guidelines panel. JAMA 279:943, 1998

76. Weis SE, Slocum PC, Blais FX, et al: The effect of directly observed therapy on the rates of drug resistance and relapse in tuberculosis. N Engl J Med 330:1179, 1994

77. Benator D, Bhattacharya M, Bozeman L, et al: Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial. Lancet 360:528, 2002

78. Vernon A, Burman W, Benator D, et al: Acquired rifamycin monoresistance in patients with HIV-related tuberculosis treated with once-weekly rifapentine and isoniazid. Tuberculosis Trials Consortium. Lancet 353:1843, 1999

79. Acquired rifamycin resistance in persons with advanced HIV disease being treated for active tuberculosis with intermittent rifamycin-based regimens. MMWR Morb Mortal Wkly Rep 51:214, 2002

80. Burman WJ, Jones BE: Treatment of HIV-related tuberculosis in the era of effective antiretroviral therapy. Am J Respir Crit Care Med 164:7, 2001

81. Yeni PG, Hammer SM, Hirsch MS, et al: Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society–USA Panel. JAMA 292:251, 2004

82. Dean GL, Edwards SG, Ives NJ, et al: Treatment of tuberculosis in HIV-infected persons in the era of highly active antiretroviral therapy. AIDS 16:75, 2002

83. Narita M, Ashkin D, Hollender ES, et al: Paradoxical worsening of tuberculosis following antiretroviral therapy in patients with AIDS. Am J Respir Crit Care Med 158:157, 1998

84. Updated guidelines for the use of rifamycins for the treatment of tuberculosis among HIV-infected patients taking protease inhibitors or non-nucleoside reverse transcriptsase inhibitors. Centers for Disease Control and Prevention, Atlanta, January 20, 2004
http://www.cdc.gov/nchstp/tb/TB_HIV_Drugs/TOC.htm

85. Thwaites GE, Nguyen DB, Nguyen HD, et al: Dexamethasone for the treatment of tuberculous meningitis in adolescents and adults. N Engl J Med 351:1741, 2004

86. Quagliarello V: Adjunctive steroids for tuberculous meningitis: more evidence, more questions. N Engl J Med 351:1792, 2004

87. Davidson PT: Managing tuberculosis during pregnancy. Lancet 346:199, 1995

88. Mahmoudi A, Iseman MD: Pitfalls in the care of patients with tuberculosis: common errors and their association with the acquisition of drug resistance. JAMA 270:65, 1993

89. Chan ED, Laurel V, Strand MJ, et al: Treatment and outcome analysis of 205 patients with multidrug-resistant tuberculosis. Am J Respir Crit Care Med 169:1103, 2004

90. Comstock GW: How much isoniazid is needed for prevention of tuberculosis among immunocompetent adults? Int J Tuberc Lung Dis 3:847, 1999

91. Update: Adverse event data and revised American Thoracic Society/CDC recommendations against the use of rifampin and pyrazinamide for treatment of latent tuberculosis infection—United States, 2003. Centers for Disease Control and Prevention. MMWR Morb Mortal Wkly Rep 52:735, 2003

92. Nolan CM, Goldberg SV, Buskin SE**:** Hepatotoxicity associated with isoniazid preventive therapy: a 7-year survey from a public health tuberculosis clinic. JAMA 281:1014, 2003

93. Randomized controlled trial of single BCG, repeated BCG, or combined BCG and killed *Mycobacterium leprae* vaccine for prevention of leprosy and tuberculosis in Malawi. Karonga Prevention Trial Group. Lancet 348:17, 1996

94. Blumberg HM, Watkins DL, Berschling JD, et al: Preventing the nosocomial transmission of tuberculosis. Ann Intern Med 122:658, 1995

95. Guidelines for preventing the transmission of *Mycobacterium tuberculosis* in healthcare settings, 1994. Centers for Disease Control and Prevention. MMWR Recomm Rep 43(RR-13):1, 1994

96. Boom WH, Canaday DH, Fulton SA, et al: Human immunity to *M. tuberculosis*: T cell subsets and antigen processing. Tuberculosis 83:98, 2003

97. Targeted tuberculin testing and treatment of latent tuberculosis infection. American Thoracic Society. MMWR Recomm Rep 49(RR-6):1, 2000

III INFECTIONS DUE TO NEISSERIA

Jeanne M. Marrazzo, m.d., m.p.h.

Infections caused by *Neisseria* species are among the most frequently encountered and potentially dangerous diseases. The two major species of concern are *N. meningitidis* and *N. gonorrhoeae.* Both are gram-negative cocci that reside primarily in polymorphonuclear white blood cells and tend to cluster in pairs [*see Figure 1*]. No nonhuman reservoir exists for either organism. Although there is some overlap in the clinical syndromes these two species elicit, they are commonly known for distinctly different presentations.

Infections Caused by *Neisseria meningitidis*

The most clinically relevant classification scheme for *N. meningitidis* utilizes the organism's capsular polysaccharides and includes at least 13 serogroups. The most common of these are A, B, C, X, Y, L, and W135.[1] In addition to serogroup, the organism can be classified by class 1 outer-membrane proteins (OMP) (serosubtype), class 2 or 3 OMP (serotype), and lipooligosaccharide (immunotype).

Invasive disease caused by *N. meningitidis* is a significant source of morbidity and mortality worldwide. Important trends include increased recognition of the Y serogroup's role in outbreaks, consideration of college freshmen as candidates for vaccination, and the relative increase of *N. meningitidis* as a pathogen in children as the success of immunization against *Haemophilus influenzae* type b is increasingly realized.

epidemiology and transmission

N. meningitidis is most commonly found in the setting of asymptomatic nasopharyngeal carriage. The prevalence of nasopharyngeal colonization in the United States is estimated at 5% to 10%.[2] Duration of colonization varies from transient (days) to as long as 2 years, with a median of 9 to 10 months. Development of humoral immunity to the colonizing strain of the organism is common and may last at least 4 to 6 months after exposure.

Transmission results from direct transfer of respiratory or oral secretions and requires prolonged intimate contact. Consequently, habitation in very close quarters (e.g., military barracks, correctional facilities, and college dormitories) can facilitate colonization, the prevalence of which can reach 20% to 40% during outbreaks.[3] Facilitation of transmission within close quarters is also reflected by seasonal patterns in invasive infection. Although disease occurs year-round, the incidence peaks in late winter and early spring. Transmission has also been documented through prolonged exposure to an infected person on an airplane and in laboratory technicians working with isolates on open laboratory benches.[4,5]

Despite longstanding availability of antibiotics and meningococcal vaccine, most areas of the world continue to have stable rates of endemic disease. After *Streptococcus pneumoniae*, *N. meningitidis* is the most common cause of bacterial meningitis worldwide, causing an estimated 25% of cases.[6] In the United States, laboratory-based surveillance for invasive meningococcal disease from 1992 to 1996 revealed an average annual incidence of 1.1 cases per 100,000 population (approximately 2,454 cases a year).[2,6,7] Incidence was highest in infants younger than 1 year, with a rate of 15.9/100,000 for children 4 to 5 months of age. The serogroup associated with the most cases of invasive disease was serogroup C (35%); however, significant geographic variation was seen, and the incidence of disease from serogroup Y increased during the study period.[7] For reasons that are unclear, rates of invasive meningococcal disease in the United States are higher in blacks than in nonblacks.

The risk of invasive meningococcal infection in college students has been summarized by the Centers for Disease Control and Prevention (CDC),[8] which initiated surveillance in this group in 1998. Whereas the incidence in undergraduates overall was lower than that in persons 18 to 23 years of age who were not enrolled in college, rates were relatively high in the approximately 590,000 freshmen who lived in dormitories (4.6/100,000).

Sporadic outbreaks of invasive meningococcal disease also continue to occur in sub-Saharan Africa, in an area extending from Senegal in the west to Ethiopia in the east, known as the meningitis belt. Outbreaks caused primarily by serogroup A occur there during the dry season, from December to June. However, serogroups B and C have caused large epidemics elsewhere, including the United States.[7]

pathogenesis and immunity

The cascade of events from exposure to colonization and from invasion to specific manifestations of clinical disease is complex. Most persons who develop invasive meningococcal disease do so after recent colonization.[1] Concomitant upper respiratory viral infection, cigarette smoking, underlying chronic illness, and preceding *Mycoplasma* infection may facilitate meningococcal infection and colonization. Colonization requires attachment to nonciliated columnar mucosal cells of the nasopharynx. Meningococci secrete proteases that cleave IgA, which may disable a major mucosal defense mechanism; how this mechanism contributes to pathogenesis is unclear, however. Endocytosis and intracellular transport of meningococci-laden vacuoles mediate the organisms' passage to the submucosa.

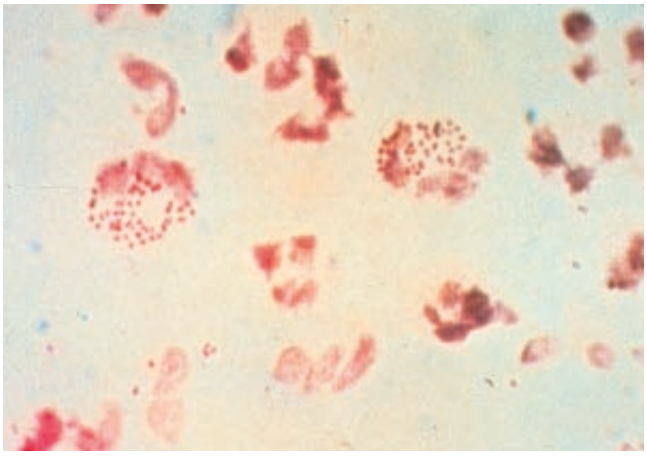


Figure 1 **Gram stain of urethral secretions showing abundant *N. gonorrhoeae* as gram-negative diplococci within polymorphonuclear cells.**

Immunity to meningococcus involves interactions at the nasopharyngeal mucosa, innate immune mechanisms, and acquired antibody.[9,10] The importance of antibody is highlighted by several findings. Natural immunity is acquired after, and boosted by, meningococcal colonization of the nasopharynx. Immunity correlates with bactericidal antibody levels in serum, and the age-specific attack rate of meningococcal disease is reciprocally related to the presence of serogroup-specific antibodies.[9] Further, hypogammaglobulinemia is a risk factor for invasive disease. In military recruits, risk of disease is highest in the absence of bactericidal activity against the prevalent pathogenic strain.[11] Specific antibody directly binds the meningococcus and activates complement-mediated phagocytosis.

Both pathways of the complement system—classic and alternative—appear critical in controlling meningococcal infection.[9] Persons with deficiencies in the terminal complement system (C5, C6, C7, C8, or C9) develop antibody to meningococci, but the rates of meningococcal infection in these individuals are 1,400-fold to 10,000-fold higher than in the general population.[12] Approximately 50% of affected persons have at least one episode of meningococcal infection, and 20% to 25% have more than one episode. Of note, infections in persons with complement deficiency often involve unusual serogroups and are less severe. Persons with deficiencies of properdin (an alternative-pathway component) or factor D are also at increased risk for meningococcal disease, with a case-fatality rate of over 50%. Recurrent meningococcal disease should prompt screening for complement deficiencies,[13] which involves testing the serum for hemolytic whole-complement activity.

Coagulopathy and microvascular thrombosis are hallmarks of meningococcal sepsis. The most visible manifestation of these processes is purpura fulminans [*see Figure 2*]. Dysfunction of the activation pathway of protein C appears to play a key role in these thrombotic events. When activated by its binding to thrombomodulin and the endothelial protein C receptor, protein C normally functions to keep thrombin's procoagulant properties in check. In children with purpuric lesions from meningococcal sepsis, endothelial protein C activation is impaired.[14]

CLINICAL PRESENTATIONS

Meningococcal Bacteremia

Meningococcal bacteremia occurs across a spectrum of clinical presentations, ranging from an acute fulminant disease that is fatal within hours to asymptomatic infection. The presence of *N. meningitidis* in the blood is usually associated with severe illness. However, bacteremia may occasionally be found in persons who appear healthy or who have only mild systemic symptoms (usually fever and, sometimes, upper respiratory symptoms or rash resembling a viral exanthem).

Symptoms associated with so-called benign bacteremia usually resolve before the infection is identified by isolation of meningococcus from blood cultures. This syndrome differs from the so-called chronic meningococcemia associated with low-grade fever, rash, and polyarticular arthritis that can be confused with disseminated gonococcal infection (DGI).

The rash of chronic meningococcemia usually takes the form of a nonspecific maculopapular eruption, but it may be petechial. Unlike persons who experience recurrent meningococcal meningitis (see below), patients with chronic meningococcemia appear immunologically normal and are usually infected with typical serogroups. *N. meningitidis* should be considered in the evaluation of any patient with a chronic arthritis-dermatitis syndrome.

The most notorious presentations of meningococcal disease take two forms: meningococcemia and meningococcal sepsis. The two forms may occur simultaneously.

The term meningococcemia usually denotes bacteremia accompanied by signs of sepsis and frequently by the classic purpuric or petechial rash typical of this syndrome, which occurs in 75% of patients with this disease [*see Figure 2*]. Although the rash of purpura fulminans is obvious, patients presenting early in the course of disease may have more subtle abnormalities of the skin and mucous membranes. These abnormalities include palpebral and conjunctival petechiae and lesions in areas subject to physical pressure, such as the waist and soles of the feet. Because these findings can herald the development of full-blown meningococcemia, they should be carefully sought in persons in whom this illness is a consideration; so that a thorough examination can be performed, the patient should be examined without clothes. Affected patients may also complain of intense, diffuse myalgias in the initial period of their illness. Evolution of individual petechiae to coalescence and eventual frank ecchymosis proceeds apace with the progression of the thrombocytopenia of disseminated intravascular coagulation (DIC).

Meningococcal sepsis is characterized by a rapid progression from general, nonspecific complaints that may resemble a viral illness to hypotension, multiorgan failure, and DIC. Despite the availability of antibiotics and advances in critical care, meningococcal sepsis still carries a mortality of up to 40%.[2] Concomitant adrenal hemorrhage, known as the Waterhouse-Friderichsen syndrome, may also occur. Myocarditis has been noted on histopathologic examination in more than 70% of patients with fatal meningococcemia.[15,16] Meningococcemia may also precipitate pericarditis, sometimes to the point of tamponade. The formation of arterial thrombi can result in peripheral gangrene.

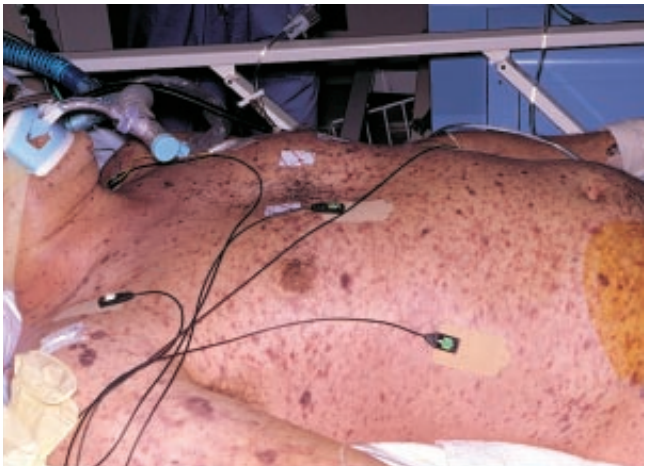
Meningitis

Acute meningococcal meningitis results from hematogenous dissemination of the organism and is usually accompanied by classic signs of bacterial meningeal inflammation: fever, headache, and nuchal rigidity. Other findings typical of meningitis may occur, including photophobia, nausea, vomiting, focal neurologic signs, seizures, and progression to obtundation and coma.[17] From 50% to 75% of patients have a petechial rash suggestive of meningococcemia.[18] The cerebrospinal fluid is typically purulent, with numerous polymorphonuclear neutrophils (PMNs), low glucose (< 50 mg/dl, present in 75% of cases), and an elevated protein concentration. In contrast, the uncommon syndrome of recurrent meningococcal meningitis produces less severe clinical and CSF findings.

Bacteremic Pneumonia

Bacteremic pneumonia from *N. meningitidis* constituted 3% of cases of invasive meningococcal disease reported by Schuchat[6] in 1997 and may be more commonly associated with serogroup Y isolates. However, meningococcal pneumonia probably occurs more frequently than is indicated by detection through positive blood cultures, and this disease has occurred in outbreaks. In a well-described outbreak involving 68 military recruits, most of the individuals had fever, rales, and

a



b

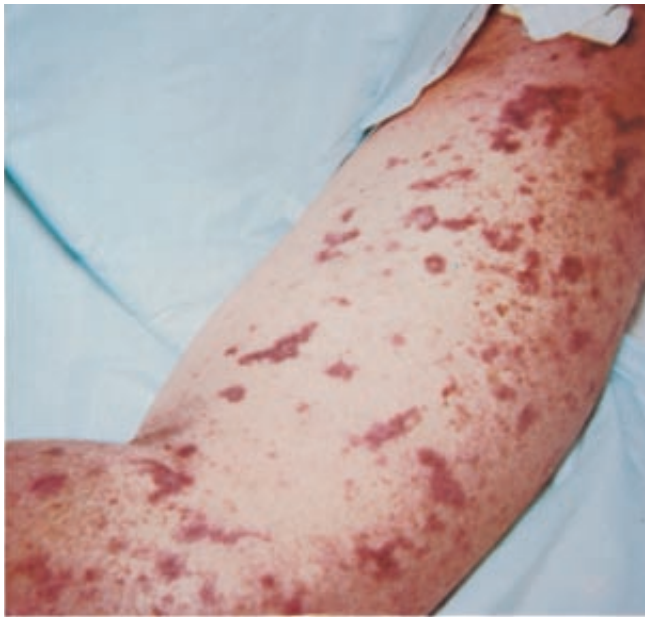


Figure 2 **(*a*) The rash of purpura fulminans, seen here with petechiae as well as larger, coalescent hemorrhagic lesions, in a patient with meningococcal sepsis and meningitis caused by serogroup B. (*b*) Closer view of petechiae, some of which have coalesced into purpuric and intracutaneous hemorrhagic lesions.**

pharyngitis.[19] Although disease was multilobar in 40% of the patients, no deaths occurred. Diagnosis by sputum culture is problematic, given the potential confounding effect of upper airway colonization. Instead, most patients are diagnosed either at bronchoscopy or, if they have systemic infection, with positive cultures at nonpulmonary sites.

Other Meningococcal Infections

Uncommonly, *N. meningitidis* can cause syndromes for which *N. gonorrhoeae* is well known, including urethritis in men and cervicitis in women. One proposed mechanism of meningococcal genital infection is acquisition through orogenital sex, which may transmit nasopharyngeal colonizers to genital sites.[20] These syndromes respond well to the antibiotics typically used for gonococcal genital infections (see below).[21] Sex partners should be treated, but no further chemoprophylaxis (e.g., treatment of household contacts) is indicated.

Cases of *N. meningitidis* infecting numerous other body sites have been reported, including endocarditis, cellulitis, conjunctivitis, otitis media, epiglottitis, and arthritis.[2,22,23]

DIAGNOSIS

N. meningitidis is a fastidious aerobic organism that grows best on chocolate agar and is distinguished from *N. gonorrhoeae* by its ability to ferment both maltose and glucose. The bacteria can be isolated from blood in approximately 75% of persons with invasive meningococcal disease and from the CSF in approximately 46% to 94% of persons with meningitis.[17] Sepsis occurs in an estimated 5% to 20% of persons with positive blood cultures. Gram stain of CSF remains a very useful means of detecting the meningococcus; however, its sensitivity is probably no higher than 50% to 70% and, like that of culture, is reduced by recent antibiotic use. Detection assays for polysaccharide antigens using countercurrent immunoelectrophoresis or latex agglutination are frequently applied to CSF. Their advantages include rapid turnaround time, high specificity, and ability to individuate serogroups. Their major disadvantage is relatively low sensitivity (meaning that false negatives occur), particularly if serogroup B is responsible; in addition, they are not reliable when used to detect antigen in urine or serum. The polymerase chain reaction (PCR) has been successfully applied to the detection of *N. meningitidis*, especially in the United Kingdom, but is not commercially available in the United States.[24] Of note, the isolation of meningococci from upper respiratory secretions does not itself indicate meningococcal disease.

COMPLICATIONS AND PROGNOSIS

Meningococcal sepsis confers a mortality of up to 40%, and 11% to 19% of survivors suffer sequelae, including hearing loss, neurologic disability, and amputation because of peripheral gangrene.[25] The case-fatality rate for all invasive meningococcal disease is estimated to be 11%, and it is significantly higher in the presence of bacteremia (17%) than with meningitis alone (3%). Mortality from *N. meningitidis* is most profound in children: in developed countries, invasive meningococcal disease is the leading cause of death in this age group. Among 295 adolescents and young adults with invasive meningococcal disease in Maryland during the 1990s, 22.5% of those 15 through 24 years of age died.[26] Clearly, this disease continues to present a major challenge to physicians, communities, and the public health system.

DIFFERENTIAL DIAGNOSIS

In its earliest stage, the presentation of invasive meningococcal disease can be nonspecific, resembling a typical viral illness with fever and myalgias. Appearance of a rash should prompt consideration of common viral etiologies, including the enteroviruses. The presence of severe systemic illness should bring to mind Rocky Mountain spotted fever (RMSF), vasculitides (polyarteritis nodosa, Churg-Strauss syndrome, and anaphylactoid [Henoch-Schönlein] purpura), and toxic shock syndromes associated with staphylococcal, streptococcal, and, less commonly, clostridial infections. Less common infections that can present in similar fashion include epidemic typhus, infections caused by *Streptobacillus moniliformis* and *Spirillum minus* (rat-bite fever), gonococcemia, septicemia caused by *H. influen-*

Table 1 Chemoprophylaxis for Meningococcal Disease[1]

Drug	*Dosage*	*Relative Efficacy*	*Comments*
Rifampin	Children < 1 mo: 5 mg/kg p.o., q. 12 hr for 2 days Children ≥ 1 mo: 10 mg/kg p.o., q. 12 hr for 2 days Adults: 600 mg q. 12 hr for 2 days	First choice	Not recommended for pregnant women, because it is teratogenic in animals; because reliability of oral contraceptives may be affected by rifampin, alternative contraceptive measures should be considered during its administration
Ciprofloxacin	Adults: 500 mg p.o. once	Alternative	Not generally recommended for persons < 18 yr or for pregnant and lactating women, because it causes cartilage damage in immature laboratory animals; however, ciprofloxacin has been used extensively in children with cystic fibrosis without reported adverse outcomes; CDC recommends it for chemoprophylaxis in children when no acceptable alternative is available
Ceftriaxone	Children < 15 yr: 125 mg I.M. once Adults: 250 mg I.M. once	Alternative	Indicated for pregnant women

zae type b, typhoid fever, and acute *S. aureus* endocarditis. Key pieces of information can help narrow the differential diagnosis. Tick-borne diseases such as RMSF usually occur within specific geographic confines and are generally associated with outdoor exposure to animal or insect vectors during the temperate months, principally in summer. Travel history is important in consideration of typhoid fever, as is information regarding the host's general immune status.

TREATMENT

Emergency Management

Infected patients should be isolated and droplet precautions observed until effective antimicrobial therapy has been given for at least 24 hours.[27] Antibiotic therapy should be started as soon as possible. Earlier initiation of antibiotics has been demonstrated to favorably affect outcome in some, though not all, studies,[28] but the disease is frequently so severe that reasonable measures to impact its early course should be undertaken.

Penicillin has constituted the mainstay of antibiotic therapy of meningococcal disease for several decades. Although β-lactamase–producing strains with high-level resistance (minimum inhibitory concentration [MIC] ≥ 250 μg/ml) exist, and strains with altered penicillin-binding proteins and intermediate resistance (MIC, 0.1 to 1.0 μg/ml) have been isolated clinically, treatment failures with penicillin have not been reported.[29] Similarly, isolates with high-level resistance to chloramphenicol have been reported outside the United States. Given these data, meningococcal isolates from blood and CSF should be routinely evaluated for penicillin susceptibility at a reference laboratory.

Antibiotic Therapy

Until the presence of meningococcus is confirmed, the patient should receive empirical treatment for bacterial meningitis [*see 7:XXXVI Bacterial Infections of the CSF*]. Once *N. meningitidis* is identified in the CSF or blood, monotherapy with penicillin G (300,000 U/kg/day I.V., up to 24 million U/day) or ceftriaxone (50 mg/kg/day, up to 2 g) may be used. For persons who cannot tolerate penicillins or cephalosporins, chloramphenicol is an option; however, hematologic toxicity remains a concern. These antibiotics provide adequate CSF penetration, especially in the presence of meningeal inflammation. Treatment for 10 to 14 days is commonly recommended.[29] Although fluoroquinolones provide excellent activity against *N. meningitidis* and achieve very good CSF levels, their role in the treatment of invasive meningococcal disease requires further study.

Adjunctive Therapy

Adjunctive therapy of meningococcal infection with corticosteroids has been a subject of intense debate. Steroid therapy has not been shown to improve outcomes associated with meningococcal disease and is therefore not recommended. Another strategy under pursuit is repletion of activated protein C, which is severely depleted in severe meningococcemia.[30] One small open-label study showed that compared with predicted outcomes, there were reductions in morbidity and mortality from severe meningococcemia in patients treated with activated protein C.[31] In a large, randomized, placebo-controlled study of patients with sepsis, recombinant protein C reduced mortality but also increased bleeding events.[32] None of the patients in this study had sepsis from *N. meningitidis*. However, protein C depletion appears to be more severe in meningococcal disease than in related conditions that commonly cause sepsis, suggesting that further study is warranted. Recombinant protein C is commercially available, and its use in patients with sepsis from *N. meningitidis* should be considered.

PREVENTIVE THERAPY

Although outbreaks account for only 2% to 3% of all cases of meningococcal disease in the United States, prevention of the spread of disease carries a high priority. The risk of invasive disease in family members of persons with invasive meningococcal disease is increased by a factor of 400 to 800. Further, case clusters cause great alarm in the community. Consequently, assessment of the need for prophylaxis and coordination of its administration are critical steps in the management of invasive meningococcal disease. Assistance in carrying out these steps can be provided by the public health agencies to which cases of meningococcal disease must be reported (see below).

Prophylaxis is recommended for close contacts of infected persons [*see Table 1*]. Close contacts are defined as household members, day care center contacts, and anyone directly exposed to the patient's oral secretions (e.g., by kissing, by mouth-to-mouth resuscitation, during endotracheal intubation, or during endotracheal tube management by health care workers not wearing masks).[33] The likelihood of contracting invasive

disease from close contact is highest in the first few days after exposure; thus, prophylaxis should ideally be administered within 24 hours after identification of the index case and is unlikely to be of value if given beyond 14 days after onset of illness in the index case.[34]

IMMUNIZATION

Vaccine

A quadrivalent polysaccharide vaccine for protection against *N. meningitidis* serogroups A, C, Y, and W-135 (Menomune) is currently available.[35] It is administered subcutaneously as a single 0.5 ml dose, induces protective antibody within 7 to 10 days, and is generally well tolerated. No serious side effects have been reported. The vaccine may be administered to pregnant and lactating women, because no adverse events associated with immunization during pregnancy have been reported. An important consideration is that the vaccine does not provide protection against serogroup B infection, which causes over half of cases in infants younger than one year. Vaccine efficacy against serogroups A and C is estimated at 85% to 100% in older children and adults.[36] Although protective immunity to serogroup A can be conferred in infants older than 3 months, immunity to serogroup C is difficult to attain in children younger than 1 year. No data are available on efficacy of the Y and W-135 polysaccharides in older children and adults, but these polysaccharides are safe and immunogenic. For all serogroups, limited data suggest that the duration of immunoprotection is probably no more than 3 years. Given this, as well as the difficulty of inducing an adequate immune response in infants, routine vaccination of infants is not recommended.

Several new meningococcal conjugate vaccines are under study for serogroups A, C, Y, and W-135, and they are likely to become available in the United States in the next few years.[37] Unlike polysaccharide vaccines, they induce a stronger, longer-lasting immune response that can be boosted by subsequent doses. Serogroup C conjugate vaccines were introduced in the United Kingdom in 1999.

Recommendations for Vaccine Use

Vaccination is recommended for persons at increased risk, for prospective travelers, and for the control of outbreaks.

Persons at increased risk Persons at increased inherent risk include military recruits and persons with terminal complement pathway deficiencies or functional or anatomic asplenia. Persons exposed routinely to *N. meningitidis* through occupational exposure (e.g., clinical or research laboratory personnel) should also consider vaccination.

Prospective travelers Vaccination is recommended for travelers to areas endemic for invasive meningococcal disease, including parts of sub-Saharan Africa during peak periods of disease incidence (generally the dry season, from December to June). In addition, a large international outbreak in pilgrims to the Hajj in Saudi Arabia in 2001 prompted recommendations that travelers to this site also be immunized.[38] Updated recommendations for this and other travel-related immunizations can be obtained at www.cdc.gov/travel.

Control of outbreaks Vaccination may be considered as a means of controlling outbreaks caused by serogroups covered by the vaccine. The Advisory Committee on Immunization Practices (ACIP) recommends that mass vaccination of persons 2 years of age or older be considered when three cases of serogroup C meningococcal disease occur within a 3-month period in a community or organization (e.g., a school), with an incidence of 10 cases per 100,000 population or greater.[39]

Immunization of college freshmen, particularly those living in dormitories, deserves special mention. Although the rate of invasive infection in this group exceeds that of any age group other than children younger than 2 years, it is still below the threshold recommended for initiating meningococcal vaccination campaigns. Thus, ACIP recommends that health care providers and colleges inform students and their parents about the vaccine's availability and potential benefits.[8] Many colleges recommend meningococcal vaccination to incoming freshmen, and some offer the vaccine through their student health service.

Requirements for Reporting Meningococcal Disease

Infection with *N. meningitidis* is reportable by law to most local and state health departments. Public health agencies will assist clinicians in the identification and treatment of exposed contacts; in the case of outbreaks, these agencies may institute other control measures. Physicians should not assume that clinical laboratories will execute reporting. Contact information for reporting communicable diseases can be found through state health departments, Web sites for which can be found at http://www.cdc.gov/mmwr/international/relres.html.

Infections Caused by *Neisseria gonorrhoeae*

Known primarily as a cause of sexually transmitted infections, *N. gonorrhoeae* remains an important cause of cervicitis, urethritis, proctitis, and pelvic inflammatory disease (PID) [*see 7:XXII Sexually Transmitted Diseases*]. No vaccine is available.

EPIDEMIOLOGY

An estimated 600,000 new cases of gonococcal infection occur in the United States each year.[40] These cases are a mix of symptomatic infections, which occur mostly in men, and asymptomatic infections detected through routine testing, largely in women.

The incidence of gonorrhea declined steadily in the United States from 1978 through 1997, but rates increased in 1998 and have not declined since then. In particular, the incidence in adolescents, especially those in large cities, remains high; in addition, sustained outbreaks have occurred recently in men who have sex with men. In the United States in 2000, the highest reported rates of gonorrhea were in women 15 to 19 years of age (715.6 per 100,000 population) and in men 20 to 24 years of age (589.7 per 100,000 population).[41] In 42 states, the incidence of gonorrhea in women remains above the objective of 19 new cases per 100,000 population from Healthy People 2010 (http://www.health.gov/healthypeople/document/html/objectives/25-02.htm), a set of health objectives for the United States developed through the Office of Disease Prevention and Health Promotion of the United States Department of Health and Human Services.

Rates of gonococcal infection have increased among men who report having sex with other men.[42] In a surveillance project at six sexually transmitted disease clinics in five U.S. cities, positivity of urethral gonorrhea in this population was 21% for those who were HIV positive and 12% for those who were HIV

negative. These findings have spurred some government agencies to recommend routine screening at least annually for gonorrhea in this population.[43]

Typing methods for the gonococcus are generally less clinically useful than for meningococcus and are used primarily to study gonococcal epidemiology. Most widely used are auxotyping, which classifies the organism on the basis of nutritional requirements, and protein I-serotyping, which is based on the stable antigenic diversity of its largest surface protein and further classifiable into different serovars by coagglutination assays. These two methods are combined into auxotype/serovar (A/S) classes for nomenclature of many strains (e.g., AHA/IA-1 denotes a strain that requires arginine, hypoxanthine, and uracil for growth and exhibits protein IA with a type 1 coagglutination pattern).[21]

PATHOGENESIS

The pathogenesis of *N. gonorrhoeae* has been closely studied in an experimental model using male human volunteers.[44] Like *N. meningitidis*, the gonococcus possesses a protease that may be important in cleaving IgA at the mucosal surface. Columnar or cuboidal epithelium is the main target for attachment, which is mediated primarily by pili that protrude from the cell surface and by outer membrane proteins termed Opa proteins.[45] Because they require iron for growth, gonococci possess transferrin receptors; they also contain lipo-oligosaccharide (LOS). The toxicity of LOS may be especially important in incapacitating the ciliary function of cells lining the fallopian tubes.[45] Within 24 to 48 hours after attachment, the organism's penetration into submucosa elicits an intense neutrophilic inflammatory response. Submucosal microabscesses form, purulent exudate collects, and the affected epithelium sloughs, resulting in ready detection of intracellular gram-negative diplococci in neutrophils on Gram stain.

The role of systemic antibody in gonococcal infection is unclear, given that individuals may be infected multiple times. Gonococcal strains can differ in the clinical manifestations they induce. Some strains, such as AHU/IA-1, AHU/IA-2 and CU, have been associated with the uncommon finding of asymptomatic urethral infection in men and a higher incidence of DGI.[46]

CLINICAL PRESENTATIONS

N. gonorrhoeae is known primarily for its propensity to cause superficial mucosal infections that are transmitted through sexual contact. In men, these infections most commonly involve the urethra. After an incubation period of 1 to 14 days, 95% of men infected at this site experience dysuria and urethral discharge. The discharge is typically purulent but can be serous or seropurulent, especially early in the course of disease. Mild edema and erythema of the urethral meatus may be seen. If left untreated, infection usually resolves over the course of several weeks, and persistent asymptomatic carriage is thought to be unlikely. In some men with untreated urethral infection, infection may spread to cause epididymitis, prostatitis, seminal vesiculitis, and infections of the Cowper and Tyson glands.[21] The urethra may also be infected in women, but isolated urethral infection is unusual except in women who have undergone a hysterectomy. Urethral infection in women is less frequently symptomatic.

Endocervical Infection

In women, the most common site of infection is the endocervical canal, where *N. gonorrhoeae* may cause mucopurulent cervicitis (MPC). MPC presents as either mucopurulent endocervical discharge or easily induced endocervical bleeding. Cervical ectopy, if present, may appear edematous. However, at least half of women with gonococcal infection of the cervix have neither signs of MPC nor gonococci detected on Gram stain of endocervical secretions. If symptoms develop, they are nonspecific and typical of most lower genital tract infections: abnormal, increased, or malodorous vaginal discharge; bleeding between menses; menorrhagia; pelvic pain; or pain with intercourse. If cervical infection is left untreated, *N. gonorrhoeae* may ascend to infect the upper genital tract—including the endometrium, fallopian tubes, ovaries, or adnexa (PID)—or the perihepatic space (Fitz-Hugh-Curtis syndrome). PID is estimated to occur in 10% to 20% of infected women. Finally, abscesses of Bartholin glands are not uncommon.[47]

Infection at Other Mucosal Sites

Both men and women can be infected at common mucosal sites, including the rectum, pharynx, and conjunctiva. Approximately 35% to 50% of women with endocervical infection are also infected at the rectum, usually without local symptoms. Receptive anal sex is not a prerequisite for rectal infection in women; rather, these infections may result from perianal inoculation with infected cervicovaginal secretions. Men who practice receptive anal sex with other men are also at risk of rectal infection. The presentation of rectal gonococcal infection ranges from asymptomatic colonization detected at routine screening to overt proctitis. Even in symptomatic patients, the range of manifestations is wide, including mild perianal pruritus, painless mucopurulent rectal discharge, mild rectal bleeding, severe rectal pain, tenesmus, and constipation.

Gonococcal infection of the pharynx is rarely symptomatic. Acquired by receptive oral sex (by either fellatio or cunnilingus but more efficiently by fellatio), pharyngeal infection is usually detected through routine screening. In persons with gonococci detected at nonpharyngeal sites, the prevalence of pharyngeal infection ranges from 3% to 7% in heterosexual men, 10% to 20% in heterosexual women, and 10% to 25% in men who have sex with other men.[21]

Gonococcal conjunctivitis occurs uncommonly in adults. It usually results from autoinoculation, particularly in laboratory and medical personnel.

Disseminated Gonococcal Infection

The term disseminated gonococcal infection refers to gonococcal infections that have spread beyond the genitourinary tract. The most common presentation is the acute arthritis-dermatitis syndrome, which is estimated to occur in 0.5% to 3.0% of persons with untreated mucosal gonococcal infection.[48,49] This syndrome may comprise the triad of tenosynovitis, dermatitis, and polyarthralgias without purulent arthritis, or it may appear as purulent arthritis alone. DGI should be strongly considered in any young, sexually active person with acute, nontraumatic oligoarthritis or tenosynovitis.

The arthritis of DGI can affect joints of any size and is typically asymmetrical.[50] With tenosynovitis, major tendon sheaths and their insertions are often tender and inflamed; the clinician should palpate these sites if this diagnosis is at all entertained. The classic rash of DGI usually consists of relatively few (< 20) tender, necrotic pustules on an erythematous base that often resolve within several days if left untreated.

Diagnosis of DGI is made more often in women than in men.

Predisposing factors include recent menstruation, recent pregnancy, and terminal complement deficiency. The gonococcal strains that cause DGI are often those associated with asymptomatic genital disease and with resistance to complement-mediated bactericidal activity of normal human serum. The organism is recovered by culture from normally sterile sites, including blood or joint fluid, in less than 50% of persons with DGI; however, it can be cultured from mucosal sites or from sexual partners in more than 80% of cases. Amplified nucleic acid assays, such as PCR, have increased the yield of *N. gonorrhoeae* detection in joint fluid, but a subset of those affected still have sterile joint fluid in the presence of urogenital gonococci, suggesting that immunomodulatory responses are important in the pathogenesis of DGI.[51,52]

More invasive infection with *N. gonorrhoeae* is uncommon. Endocarditis and meningitis have been reported, however.[53,54]

DIAGNOSIS

N. gonorrhoeae may be diagnosed presumptively by direct visualization on a Gram stain smear of secretions from a compatible clinical site and, specifically, by growth in culture or detection by antigen detection, nonamplified DNA probe, and a nucleic acid amplification test (NAAT). In men with urethral discharge from *N. gonorrhoeae*, the organism is seen on Gram stain in 95% [*see Figure 1*]. Although reasonably specific for detection of endocervical *N. gonorrhoeae* (> 95%), Gram stain is considerably less sensitive (50% to 70%) and thus not recommended as the sole means of diagnosis of endocervical infection. Gram stain of anoscopically obtained rectal secretions has a sensitivity of 70% to 80% and thus may assist in making a specific diagnosis of gonococcal proctitis.[21] Because nongonococcal *Neisseria* species colonize the pharynx, Gram stain of the pharynx is not specific to infection with *N. gonorrhoeae* and thus is not recommended.

Traditional bacterial culture remains the mainstay of microbiologic diagnosis for normally sterile specimens in which invasive disease is suspected, such as blood and joint fluid, and is commonly used for the diagnosis of cervical, urethral, pharyngeal, and rectal infections. Cultures are obtained from these sites with a sterile Dacron swab that is then swept across the surface of a plate containing chocolate agar supplemented with glucose, vancomycin, colistin, and nystatin (Thayer-Martin media) and held in an environment with a high level of CO_2. The organism grows best under aerobic conditions at 35° to 37° C; however, it can also grow under anaerobic conditions. Unlike *N. meningitidis*, it does not ferment lactose. Relative to an expanded diagnostic standard that incorporates results of NAAT, culture for *N. gonorrhoeae* has an estimated sensitivity of 90% and specificity of greater than 99%.[55]

Other tests commonly used to diagnose gonococcal infection are a nonamplified DNA probe and several types of NAAT. The DNA probe is a nucleic acid hybridization test with a sensitivity of approximately 85% and specificity of 98%.[56] Available NAATs include PCR, ligase chain reaction (LCR), transcription-mediated assay (TMA), and hybrid capture tests. In general, the performance of these tests exceeds that of the nonamplification techniques, enhancing sensitivity while maintaining excellent specificity.[56] Further, NAATs have the major advantage of performing well on noninvasive specimens: urine in men and, in women, urine and vaginal swabs. Vaginal swabs may be collected either by patients or by clinicians, which provides opportunities for novel screening strategies. The NAATs in general have sensitivities for detection of *N. gonorrhoeae* of 95% to 99%, with a specificity greater than 99% for cervical and urethral specimens. However, none is currently recommended for use on specimens other than urine, cervical, or urethral samples.[56]

TREATMENT

Recommended treatment of gonorrhea varies according to the site of infection and the likelihood of antibiotic resistance [*see Table 2*]. Currently recommended regimens for the treatment of gonorrhea are available online at www.cdc.gov/std/treatment.[43]

The gonococcus has multiple means of acquiring resistance to antibiotics. Plasmid-mediated mechanisms confer resistance to penicillin by encoding altered penicillin-binding proteins (PBPs). Resistance to tetracyclines is mediated by chromosomal mechanisms. Resistance to fluoroquinolones is conferred by production of an altered DNA gyrase to which these antibiotics are unable to bind and hence are rendered ineffectual. The Gonococcal Isolate Surveillance Project (GISP) annually updates important trends in gonococcal resistance patterns (http://www.cdc.gov/ncidod/dastlr/gcdir). Because these patterns can emerge and progress surprisingly rapidly, physicians should be aware of them. Although some problems begin in relative geographic isolation, they often mark the start of significant nationwide trends.[57-59] For example, recommendations for the empirical use of single-dose fluoroquinolone therapy were prominent in the CDC's 1998 Sexually Transmitted Disease Treatment Guidelines; the 2002 document emphasizes that because up to 14% of gonococcal isolates in Hawaii exhibit resistance to fluoroquinolones, these drugs should not be used in this area.[60,61] Patients in whom physicians should consider the possibility of quinolone-resistant *N. gonorrhoeae* (QRNG) include those who (1) have had failures with fluoroquinolone therapy, (2) have traveled to Hawaii or Southeast Asia (where resistance is endemic) or have sexual partners who may have acquired a gonococcal infection there, and (3) reside in California, where recent data indicate an increasing prevalence of QRNG. Because active surveillance is critical, physicians who encounter documented or suspected cases of QRNG should report this to their local health department. In 2000, 25% of GISP isolates were resistant to penicillin, tetracycline, or both. To date, no isolates resistant to cephalosporins have been detected in GISP. Because persons with gonococcal infection are at risk for other STDs, particularly chlamydial infection, the CDC recommends that treatment regimens for gonorrhea be partnered with an antibiotic effective against *C. trachomatis* as well. Although azithromycin is active against the gonococcus, a dose of 2 g orally is required to effect acceptable cure rates. This is double the dose required to treat chlamydial infection, and such a high dose frequently causes gastrointestinal side effects. Similarly, ciprofloxacin remains effective for gonococcal infections not caused by QRNG, but it is not effective against chlamydial infection. Both infections can be treated with a 1-week course of levofloxacin or ofloxacin, however.

COMPLICATIONS AND PROGNOSIS

The best-known complications of *N. gonorrhoeae* infection include PID and neonatal conjunctivitis (ophthalmia neonatorum). PID in which gonococci play an etiologic role may be more purulent and severe than PID caused by *C. trachomatis*. If left untreated, 10% to 40% of women with gonococcal cervical infection will develop PID.[62] Ophthalmia neonatorum caused by *N. gonorrhoeae* can be severe, resulting in perforation of the

Table 2 Treatment of Gonorrhea[1]

Disease	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Uncomplicated infection of the cervix, urethra, or rectum*	Cefixime	400 mg p.o. once	First choice	Quinolones not recommended in areas where resistance is a concern; spectinomycin is used for patients intolerant of cephalosporins and quinolones, it is unreliable against pharyngeal infection, and posttreatment cultures are recommended
	Ciprofloxacin	500 mg p.o. once	Alternative first choice	
	Ofloxacin	400 mg p.o. once	Alternative first choice	
	Levofloxacin	250 mg p.o. once	Alternative first choice	
	Ceftriaxone	125 mg I.M. once	Second choice	
	Spectinomycin	2 g I.M. once	Second choice	
Conjunctivitis (not ophthalmia neonatorum)	Ceftriaxone	1 g I.M. once	—	—
Disseminated gonococcal infection (DGI)	*Parenteral*			Parenteral therapy is given for 24–48 hr after improvement, then followed by oral therapy for 1 wk total antibiotic therapy
	Ceftriaxone	1 g I.M. or I.V. q. 24 hr	First choice	
	Cefotaxime	1 g I.V. q. 8 hr	Alternative	
	Ceftizoxime	1 g I.V. q. 8 hr	Alternative	
	Ciprofloxacin	400 mg I.V. q. 12 hr	Alternative	
	Ofloxacin	400 mg I.V. q. 12 hr	Alternative	
	Levofloxacin	250 mg I.V. q.d.	Alternative	
	Spectinomycin	2 g I.M. q. 12 hr	Alternative	
	Oral			
	Cefixime	400 mg p.o., b.i.d.		
	Ciprofloxacin	500 mg p.o., b.i.d.		
	Ofloxacin	400 mg p.o., b.i.d.		
	Levofloxacin	500 mg p.o., q.d.		
Meningitis	Ceftriaxone	1–2 g I.V. q. 12 hr for 10–14 days	—	—
Endocarditis	Ceftriaxone	1–2 g I.V. q. 12 hr for ≥ 4 wk	—	—
Ophthalmia neonatorum	Ceftriaxone	25–50 mg/kg I.V. or I.M. in a single dose, not to exceed 125 mg	—	Topical antibiotic therapy alone is inadequate

*Treatment of gonorrhea in the adult should always be accompanied by treatment of chlamydial infection (azithromycin, 1 g orally in a single dose, or doxycycline, 100 mg orally b.i.d. for 7 days), and patients should abstain from sex during treatment. Test of cure is not routinely recommended.

globe and blindness. Moreover, accruing evidence suggests that gonococcal infection may profoundly impact global morbidity through its role in facilitating transmission and acquisition of HIV. In men infected with HIV, the presence of gonococcal urethritis was shown to increase the quantity of HIV shed in semen by approximately eightfold.[63,64] Remarkably, treatment with routine antibiotics aimed at *N. gonorrhoeae* resulted in significant reduction of the associated HIV shedding.[64] Similarly, cervical inflammation in women, commonly caused by gonococcal infection, is associated with increased shedding of HIV; treatment of cervicitis reduces this shedding.[65,66] These observations suggest that the inflammation associated with gonococcal infection significantly increases the likelihood that HIV may be more efficiently transmitted through unprotected sex. In persons who are not infected with HIV, the same inflammatory cells elicited by gonococcal infection provide a ready target for HIV infection; thus, risk of HIV acquisition is very likely increased in this setting.[67]

Management of Sexual Partners and Chemoprophylaxis

Persons infected with *N. gonorrhoeae* should be interviewed and counseled about the importance of treatment for their sexual partners. The CDC recommends that sexual partners with whom patients have had sex within the past 60 days be tested and treated for both gonococcal infection and chlamydial infection.[43] If the patient's last sexual contact was more than 60 days before being interviewed, the last partner should be evaluated and treated. For infants, most states require by law that, at birth, an agent be administered to prevent gonococcal ophthalmia neonatorum. Options include single applications of one of several agents, including silver nitrate 1% aqueous solution, erythromycin 0.5% ophthalmic ointment, and tetracycline 1% ophthalmic ointment.

Other *Neisseria* Species

Several typically nonpathogenic *Neisseria* species are found as saprophytes in the upper respiratory tract: the nonchromogens *N. lactamica*, *N. mucosa*, and *N. sicca* and the chromogens *N. flavescens* and *N. subflava*. All of these species can occasionally cause disease.[68,69] Meningitis that is clinically indistinguishable from that caused by *N. meningitidis* has been attributed to each of these species, most frequently to *N. subflava* and *N. mucosa*. Endocarditis has also been attributed to nonpathogenic species, particularly *N. sicca*, *N. mucosa*, and *N. subflava*. Odontogenic infections or bite wounds may also harbor these species.[70] Infections with nonpathogenic *Neisseria* species can be treated effectively with penicillin, but occasionally, strains are resistant. The antimicrobial susceptibility of the strain causing the infection should be used to guide therapy.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Apicella MA: *Neisseria meningitidis*. Principles and Practice of Infectious Diseases, 5th ed. Mandell GL, Bennett JE, Dolin R, Eds. Churchill Livingstone, Philadelphia, 2000, p 2228

2. Rosenstein NE, Perkins BA, Stephens DS, et al: Meningococcal disease. N Engl J Med 344:1378, 2001

3. Tappero JW, Reporter R, Wenger JD, et al: Meningococcal disease in Los Angeles County, California, and among men in the county jails. N Engl J Med 335:833, 1996

4. Exposure to patients with meningococcal disease on aircrafts—United States, 1999–2001. MMWR Morb Mortal Wkly Rep 50:485, 2001

5. Laboratory-acquired meningococcal disease—United States, 2000. MMWR Morb Mortal Wkly Rep 51:141, 2002

6. Schuchat A, Robinson K, Wenger JD, et al: Bacterial meningitis in the United States in 1995. Active Surveillance Team. N Engl J Med 337:970, 1997

7. Rosenstein NE, Perkins BA, Stephens DS, et al: The changing epidemiology of meningococcal disease in the United States, 1992–1996. J Infect Dis 180:1894, 1999

8. Meningococcal disease and college students. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 49:13, 2000

9. Pollard AJ, Frasch C: Development of natural immunity to *Neisseria meningitidis*. Vaccine 19:1327, 2001

10. Kolb-Maurer A, Unkmeir A, Kammerer U, et al: Interaction of *Neisseria meningitidis* with human dendritic cells. Infect Immun 69:6912, 2001

11. Goldschneider I, Gotschlich EC, Artenstein MS: Human immunity to the meningococcus: I. The role of humoral antibodies. J Exp Med 129:1307, 1969

12. Figueroa JE, Densen P: Infectious diseases associated with complement deficiencies. Clin Microbiol Rev 4:359, 1991

13. Fijen CA, Kuijper EJ, te Bulte MT, et al: Assessment of complement deficiency in patients with meningococcal disease in The Netherlands. Clin Infect Dis 28:98, 1999

14. Faust SN, Levin M, Harrison OB, et al: Dysfunction of endothelial protein C activation in severe meningococcal sepsis. N Engl J Med 345:408, 2001

15. Garcia NS, Castelo JS, Ramos V, et al: Frequency of myocarditis in cases of fatal meningococcal infection in children: observations on 31 cases studied at autopsy. Rev Soc Bras Med Trop 32:517, 1999

16. Hardman JM, Earle KM: Myocarditis in 200 fatal meningococcal infections. Arch Pathol 87:318, 1969

17. Kaplan SL: Clinical presentations, diagnosis, and prognostic factors of bacterial meningitis. Infect Dis Clin North Am 13:579, 1999

18. Acute meningococcal meningitis: analysis of features of the disease according to the age of 255 patients. Copenhagen Meningitis Study Group. J Infect 34:227, 1997

19. Koppes GM, Ellenbogen C, Gebhart RJ: Group Y meningococcal disease in United States Air Force recruits. Am J Med 62:661, 1977

20. Tayal SC, Rashid S, Muttu KM, et al: Meningococcal carriage: prevalence and sex-related risk factors. J Infect 34:101, 1997

21. Hook EW 3rd, Handsfield HH: Gonococcal infections in the adult. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mardh P-A, Lemon SM, et al, Eds. McGraw-Hill Book Co, New York, 1999, p 451

22. Joao I, Marques P, Namora J, et al: *Neisseria meningitidis* native valve endocarditis: a case report. Rev Port Cardiol 20:877, 2001

23. Lin VH, Parekh RS, McQuillan MA, et al: Meningococcal endocarditis presenting as cellulitis. Clin Infect Dis 21:1023, 1995

24. Taha MK: Molecular detection and characterization of *Neisseria meningitidis*. Expert Rev Mol Diagn 2:143, 2002

25. Kirsch EA, Barton RP, Kitchen L, et al: Pathophysiology, treatment and outcome of meningococcemia: a review and recent experience. Pediatr Infect Dis J 15:967, 1996

26. Harrison LH, Pass MA, Mendelsohn AB, et al: Invasive meningococcal disease in adolescents and young adults. JAMA 286:694, 2001

27. Guideline for isolation precautions in hospitals. Part II. Recommendations for isolation precautions in hospitals. Hospital Infection Control Practices Advisory Committee. Am J Infect Control 24:32, 1996

28. Strang JR, Pugh EJ: Meningococcal infections: reducing the case fatality rate by giving penicillin before admission to hospital. BMJ 305:141, 1992

29. Quagliarello VJ, Scheld WM: Treatment of bacterial meningitis. N Engl J Med 336:708, 1997

30. Smith OP, White B, Vaughan D, et al: Use of protein-C concentrate, heparin, and haemodiafiltration in meningococcus-induced purpura fulminans. Lancet 350:1590, 1997

31. White B, Livingstone W, Murphy C, et al: An open-label study of the role of adjuvant hemostatic support with protein C replacement therapy in purpura fulminans-associated meningococcemia. Blood 96:3719, 2000

32. Bernard GR, Vincent JL, Laterre PF, et al: Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 344:699, 2001

33. Prevention and control of meningococcal disease. MMWR Morb Mortal Wkly Rep 49:1, 2000

34. Peltola H: Prophylaxis of bacterial meningitis. Infect Dis Clin North Am 13:685, 1999

35. Rosenstein NE, Fischer M, Tappero JW: Meningococcal vaccines. Infect Dis Clin North Am 15:155, 2001

36. Rosenstein N, Levine O, Taylor JP, et al: Efficacy of meningococcal vaccine and barriers to vaccination. JAMA 279:435, 1998

37. Jodar L, Feavers IM, Salisbury D, et al: Development of vaccines against meningococcal disease. Lancet 359:1499, 2002

38. Update: assessment of risk for meningococcal disease associated with the Hajj 2001. MMWR Morb Mortal Wkly Rep 50:221, 2001

39. Control and prevention of serogroup C meningococcal disease: evaluation and management of suspected outbreaks: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 46:13, 1997

40. Fox KK, Whittington WL, Levine WC, et al: Gonorrhea in the United States, 1981–1996: demographic and geographic trends. Sex Transm Dis 25:386, 1998

41. Sexually Transmitted Disease Surveillance, 2000. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, 2001

42. Fox KK, del Rio C, Holmes KK, et al: Gonorrhea in the HIV era: a reversal in trends among men who have sex with men. Am J Public Health 91:959, 2001

43. Sexually transmitted disease treatment guidelines. MMWR Morb Mortal Wkly Rep 51(RR-6):1, 2002

44. Cohen MS, Cannon JG: Human experimentation with *Neisseria gonorrhoeae*: progress and goals. J Infect Dis 179 (suppl 2):S375, 1999

45. Sparling PF: Biology of *Neisseria gonorrhoeae*. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mardh P-A, Lemon SM, et al, Eds. McGraw Hill Book Co, New York, 1999, p 433

46. Whittington WL, Holmes KK: Unique gonococcal phenotype associated with asymptomatic infection in men and with erroneous diagnosis of nongonococcal urethritis. J Infect Dis 181:1044, 2000

47. Hoosen AA, Nteta C, Moodley J, et al: Sexually transmitted diseases including HIV infection in women with Bartholin's gland abscesses. Genitourin Med 71:155, 1995

48. Rompalo AM, Hook EW 3rd, Roberts PL, et al: The acute arthritis-dermatitis syndrome: the changing importance of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. Arch Intern Med 147:281, 1987

49. O'Brien JP, Goldenberg DL, Rice PA: Disseminated gonococcal infection: a prospective analysis of 49 patients and a review of pathophysiology and immune mechanisms. Medicine (Baltimore) 62:395, 1983

50. Rice PA, Handsfield HH: Arthritis associated with sexually transmitted diseases. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mardh P-A, Lemon SM, et al, Eds. McGraw Hill Book Co, New York, 1999, p 921

51. Liebling MR, Arkfeld DG, Michelini GA, et al: Identification of *Neisseria gonorrhoeae* in synovial fluid using the polymerase chain reaction. Arthritis Rheum 37:702, 1994

52. Muralidhar B, Rumore PM, Steinman CR: Use of the polymerase chain reaction to study arthritis due to *Neisseria gonorrhoeae*. Arthritis Rheum 37:710, 1994

53. Lin MF, Lau YJ, Hu BS, et al: Gonococcal meningitis and intra-abdominal abscess in the presence of a ventriculoperitoneal shunt. Scand J Infect Dis 32:567, 2000

54. Mofredj A, Baraka D, Madec Y, et al: Disseminated gonococcal infection and meningitis. Am J Med 109:71, 2000

55. Black CM, Morse SA: The use of molecular techniques for the diagnosis and epidemiologic study of sexually transmitted infections. Curr Infect Dis Rep 2:31, 2000

56. Koumans EH, Johnson RE, Knapp JS, et al: Laboratory testing for *Neisseria gonorrhoeae* by recently introduced nonculture tests: a performance review with clinical and public health considerations. Clin Infect Dis 27:1171, 1998

57. Ehret JM, Judson FN: Quinolone-resistant *Neisseria gonorrhoeae*: the beginning of the end? Report of quinolone-resistant isolates and surveillance in the southwestern United States, 1989 to 1997. Sex Transm Dis 25:522, 1998

58. Bal C: Increasing antimicrobial resistance in STDs and the need for surveillance: *Neisseria gonorrhoeae* as a model. FEMS Immunol Med Microbiol 24:447, 1999

59. National sentinel surveillance of antimicrobial resistance in *Neisseria gonorrhoeae*. Commun Dis Rep CDR Wkly 10:195, 2000

60. Trees DL, Sandul AL, Neal SW, et al: Molecular epidemiology of *Neisseria gonorrhoeae* exhibiting decreased susceptibility and resistance to ciprofloxacin in Hawaii, 1991–1999. Sex Transm Dis 28:309, 2001

61. Fox KK, Knapp JS, Holmes KK, et al: Antimicrobial resistance in *Neisseria gonorrhoeae* in the United States, 1988–1994: the emergence of decreased susceptibility to the fluoroquinolones. J Infect Dis 175:1396, 1997

62. Platt R, Rice PA, McCormack WM: Risk of acquiring gonorrhea and prevalence of abnormal adnexal findings among women recently exposed to gonorrhea. JAMA 250:3205, 1983

63. Moss GB, Overbaugh J, Welch M, et al: Human immunodeficiency virus DNA in urethral secretions in men: association with gonococcal urethritis and CD4 cell depletion. J Infect Dis 172:1469, 1995

64. Cohen MS, Hoffman IF, Royce RA, et al: Reduction of concentration of HIV-1 in semen after treatment of urethritis: implications for prevention of sexual transmission of HIV- 1. AIDSCAP Malawi Research Group. Lancet 349:1868, 1997

65. Kreiss J, Willerford DM, Hensel M, et al: Association between cervical inflammation and cervical shedding of human immunodeficiency virus DNA. J Infect Dis 170:1597, 1994

66. McClelland RS, Wang CC, Mandaliya K, et al: Treatment of cervicitis is associated with decreased cervical shedding of HIV-1. AIDS 15:105, 2001

67. Weir SS, Feldblum PJ, Roddy RE, et al: Gonorrhea as a risk factor for HIV acquisition. AIDS 8:1605, 1994

68. Tronel H, Chaudemanche H, Pechier N, et al: Endocarditis due to *Neisseria mucosa* after tongue piercing. Clin Microbiol Infect 7:275, 2001

69. Baraldes MA, Domingo P, Barrio JL, et al: Meningitis due to *Neisseria subflava*: case report and review. Clin Infect Dis 30:615, 2000

70. Capitini CM, Herrero IA, Patel R, et al: Wound infection with *Neisseria weaveri* and a novel subspecies of *Pasteurella multocida* in a child who sustained a tiger bite. Clin Infect Dis 34:E74, 2002

IV INFECTIONS DUE TO GRAM-POSITIVE BACILLI

FREDERICK S. SOUTHWICK, M.D.

Diphtheria

Diphtheria derives its name from the Greek word *diphtheria*, meaning leather hide, which is the character of the pharyngeal membrane that is a hallmark of this disease. Once the cause of major epidemics in Europe and the United States, diphtheria was nearly eradicated by medical discoveries in the late 1800s and early 1900s. Conditions in both developing and developed countries have led to several recent outbreaks and have raised concerns that diphtheria may again become a serious public health problem.

MICROBIOLOGY

Diphtheria is caused by *Corynebacterium diphtheriae*, a gram-positive rod with club-shaped swellings at each end. It is nonmotile and is capable of growing on blood agar plates, causing a narrow band of hemolysis. Most strains produce a highly lethal exotoxin, whose production requires the presence of a bacteriophage that carries a specific determinant for toxin production. The organism produces characteristic metachromatic granules that stain bluish purple with methylene blue. Their snapping division results in angular and palisade arrangements of the cells on smear that frequently take on the appearance of Chinese lettering.[1]

ETIOLOGY AND EPIDEMIOLOGY

Humans are the only natural host for *C. diphtheriae*. The organism is most commonly spread by upper respiratory tract droplets. Persons incubating the disease, those convalescing from the infection, and healthy carriers can spread the disease to others through close contact. Asymptomatic carriage can persist for years within a population. For example, in Russia the clonal strain responsible for a large diphtheria outbreak was carried asymptomatically in the population for at least 5 years before the epidemic.[2] Spread is more common during the colder months of the year, when people tend to be crowded indoors.[3] Persons with *C. diphtheriae* skin lesions can also serve as reservoirs for the organism, and contamination of the environment tends to be greater from skin infection than from upper respiratory tract infection.[4] Because the organism can survive as long as 6 months in fomites and dust, these objects and particles can serve as vehicles for transmission.

Diphtheria toxoid immunization prevents the serious complications of diphtheria, alleviating the clinical manifestations of the disease by blocking the toxin's ability to enter cells. Immunization reduces local colonization of the nasopharynx with toxin-producing strains by reducing their survival advantage. In the United States, with the advent of widespread vaccine administration in the 1940s, the carrier state has dropped to very low levels, and the incidence of diphtheria has steadily declined. Before the initiation of the vaccination program, as many as 125,000 cases and 10,000 deaths were reported annually in the United States. Subsequently, the incidence declined to zero to five cases a year from 1980 to 1990. History indicates that diphtheria outbreaks occur in cycles that may include quiescent periods of up to 100 years. Therefore, the downturn in incidence may represent a normal cycle rather than herd immunity. Occasional outbreaks have been observed in Texas, Washington, and South Dakota.[5] Two outbreaks were associated with urban alcoholics who practiced poor hygiene and lived in crowded environments; a third occurred in a Native-American community. Day care centers can also serve as a site for the spread of diphtheria. Outbreaks in Russia (15,211 cases) and the Ukraine (2,987 cases) are thought to have resulted from the decreased immunization of infants and children, as well as from waning immunity to diphtheria in older people.[6] The resurgence of diphtheria in developed countries raises concern that reductions in the immunization of young children and the failure to revaccinate older adults could lead to a worldwide increase in the incidence of this very serious disease. The finding that only 30% of persons older than 70 years in the United States have protective antibody levels to diphtheria further emphasizes the potential for resurgence.[7]

PATHOGENESIS

C. diphtheriae attaches to mucosal surfaces, particularly in the nasopharynx. Ocular and genital mucosae are less often infected. The skin has become an increasingly common site of infection, particularly in the United States. The diphtheria bacillus rarely invades living tissue but generally remains in superficial layers of the mucosa and skin. The major manifestations of this often serious disease result from production of a potent exotoxin. When iron concentrations are low, diphtheria bacilli possessing the corynebacteriophage containing the *tox* gene produce high concentrations of the toxin. This protein has a molecular weight of 62,000 and consists of two major fragments, designated A and B. Fragment B binds to a specific host cell membrane receptor, resulting in endocytosis of the entire molecule. Fragment B then forms a membrane channel that allows fragment A to enter the cell cytoplasm, block protein synthesis, and induce cell death within hours [*see Figure 1*].[8] Antitoxin antibody can neutralize toxin adsorbed to cells or in the extracellular fluid. However, once the toxin penetrates cells, its toxic effects are irreversible.

In the host, the cytotoxic effects of the toxin are most marked in regions where bacterial growth is heavy and toxin concentrations are highest. Tissue necrosis is associated with an inflammatory response, leading to the formation of an adherent membrane that is greenish-gray to black and consists of fibrin, necrotic tissue, lymphocytes, polymorphonuclear leukocytes, erythrocytes, and bacterial colonies. All cells are susceptible to the lethal effects of the toxin; however, the heart, kidneys, and nervous system are injured most often.

DIAGNOSIS

Clinical Manifestations

The incubation period for respiratory diphtheria is generally 2 to 4 days but can be as long as 7 days. Infection is usually associated with low-grade fever (approximately 38° C [100.4° F]); early in the infection, systemic complaints often are minimal. Symptoms depend on the location and duration of the infection before treatment. Concentrations of bacteria increase over time, resulting in the release of increasing amounts of the cytotoxic exotoxin.

Early recognition and treatment are therefore critical for reducing the complications associated with toxin dissemination.

Respiratory tract infection Pharyngeal infection, the most common form of diphtheria, presents as a sore throat and malaise. Symptoms may be mild in vaccinated patients, whereas unvaccinated patients tend to have more severe disease. Initially, diphtheria pharyngitis results in the development of a patchy white exudate that can be readily removed and is indistinguishable from group A streptococcal or viral pharyngitis. However, as the toxin begins to cause cell necrosis, a thin membrane begins to form that progressively thickens and spreads over the tonsils, posterior pharynx, and uvula [*see Figure 2*]. Initially, the membrane is white and smooth but later becomes gray, with patches of green and black necrosis. Particularly in children, the membrane can spread from the posterior wall up into the nose or down to the larynx or even the tracheobronchial tree.[9] As the membrane spreads, it can interfere with normal airflow and lead to suffocation. Such extensive spread is generally associated with increased release of exotoxin, resulting in myocardial and neurologic complications and greatly increasing mortality. Laryngeal involvement results in hoarseness and may serve as a warning of impending respiratory compromise.[10] Less often, the membrane may be limited to the nasopharynx, causing a serosanguineous nasal discharge. Such limited involvement is less likely to be associated with generalized toxicity. Patients who have no membrane, but only pharyngeal erythema, almost always have milder, uncomplicated disease. Other clinical manifestations include tachycardia, cervical adenopathy, leukocytosis, and proteinuria. Development of the classic bull neck, the result of massive adenopathy, is now uncommon.

Cutaneous diphtheria Although outbreaks of cutaneous diphtheria were originally described primarily in tropical areas, outbreaks over the past 3 decades have been reported in the Pacific Northwest, Midwest, and southern United States. Cutaneous infection may result in greater environmental contamination than respiratory diphtheria and may carry a greater risk of spread to others. Unlike pharyngeal diphtheria, which usually develops in the winter months, cutaneous infections tend to peak in the late summer and early fall. Most cases are reported among residents of Seattle's Skid Road district, who practice poor hygiene and often have preexisting sores.[11] Skin infections have also been described in young schoolchildren. The classic punched-out ulcerative lesion often found in the tropics is rare in temperate climates. When the lesion does develop, it generally begins as a pustule that progresses to an ulcer with a gray-brown membrane at its base. More commonly, diphtheria superinfects preexisting skin lesions, including traumatic breaks in the skin, insect bites, ecthyma, and impetigo. In association with *C. diphtheriae*, other skin pathogens, particularly *Staphylococcus* and *Streptococcus*, are found on culture. These infections tend to be indolent and are rarely associated with signs of intoxication, probably because cutaneous infections induce high levels of antitoxin antibody.

Other sites of infection Less often, *C. diphtheriae* infects the conjunctiva, eye, ear, and vagina. In Seattle, ocular infections in Skid Road residents sometimes accompanied skin involvement.

Complications

Systemic complications are caused by release of the diphtheria exotoxin, which predominantly damages the heart and nervous system.

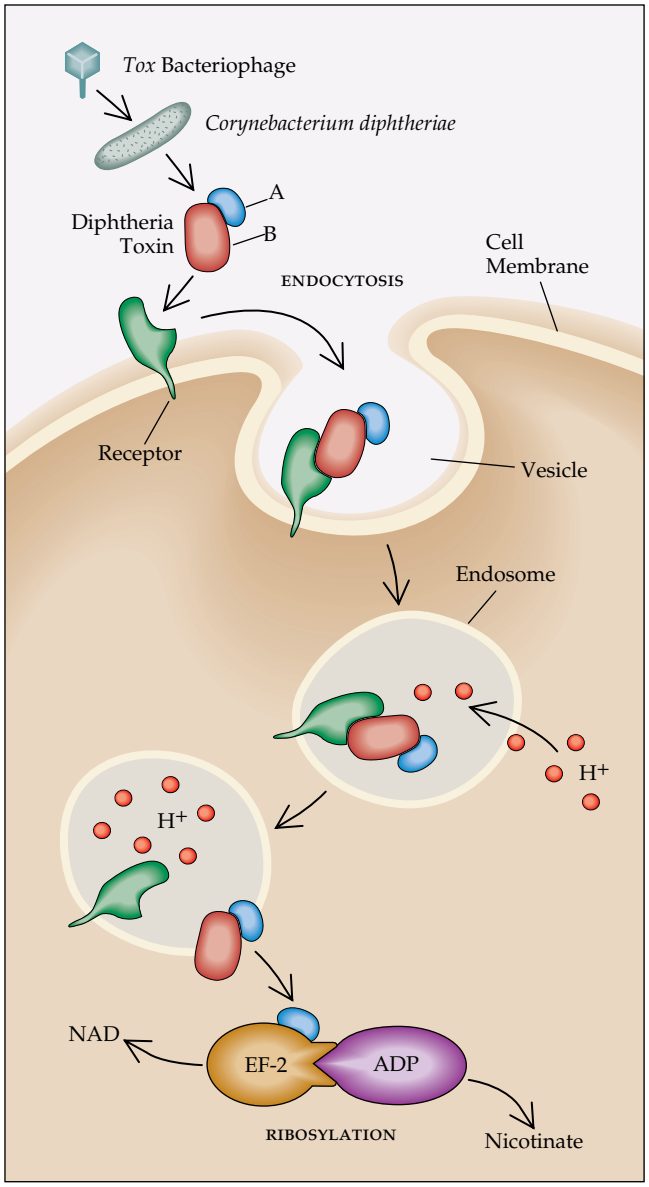


Figure 1 **Mechanism of action of diphtheria toxin. Diphtheria toxin production is encoded by a bacteriophage carrying the *tox* gene that gains entry into *Corynebacterium diphtheriae*. The toxin consists of an A fragment and a B fragment. The B fragment binds to the receptor (heparin-binding epidermal growth factor precursor) on the cell surface, and the whole molecule is then taken into the cell by endocytosis. In the closed environment of the endosome, acidification occurs (H^+). The low pH level causes the B region to unfold and form a membrane channel, allowing the A domain to pass through the membrane into the cytoplasm. The disulfide bond linking the A and B regions is reduced, and the A subunit is then freed to bind to ADP-ribosylate elongation factor–2 (EF-2). Ribosylation interferes with the ability of EF-2 to add amino acids to a peptide chain, blocking protein synthesis and causing cell death. (A—A fragment; B—B fragment; NAD—nicotinamide adenine dinucleotide; ADP—adenosine diphosphate)**

Myocarditis Subtle evidence of myocarditis is found in up to two thirds of patients with respiratory diphtheria.[12] Clinically

significant cardiac dysfunction is observed in 10% to 20% of patients. The severity of cardiac compromise correlates with the extent and severity of respiratory tract involvement. Cardiac toxicity generally occurs within 1 to 2 weeks after onset of the illness, often when pharyngeal symptoms are improving. The electrocardiogram generally reflects the severity of myocardial involvement and should be followed closely in all cases. ST segment and T wave changes and first-degree heart block are found in less severe disease, whereas left bundle branch block and AV block are associated with high mortality. In addition to damaging the Purkinje system, the toxin causes necrosis of cardiac muscle cells that can result in acute heart failure and circulatory collapse. A poor outcome is more likely in patients with extensive pharyngeal membrane and an aspartate aminotransferase level above 80 IU/L.[13] Recovery in more severe cases of myocarditis can result in normalization of the ECG; however, patients usually sustain permanent injury to the myocardium.

Neurologic toxicity Neurologic complications occur in approximately 10% of respiratory cases. As with myocardial involvement, the likelihood of neurologic involvement correlates with the severity of the respiratory infection. Symptoms develop 10 to 28 days after the onset of respiratory complaints. Two types of neuropathy are seen. The first, cranial nerve involvement, is generally limited to the glossopharyngeal and vagus nerves, resulting in difficulty swallowing, aspiration, nasal regurgitation, and loss of the gag reflex. Less often, oculomotor and facial nerves become impaired. The second type of neuropathy tends to occur somewhat later in the course of the illness and resembles Guillain-Barré syndrome. These patients have quadriparesis associated with hyporeflexia. Degeneration of myelin sheaths and axon cylinders is observed in biopsy specimens. After treatment of the infection, slow but complete neurologic recovery ensues.[14]

Physical Examination and Laboratory Tests

Prompt recognition and treatment of respiratory diphtheria are critical for preventing complications and mortality. In the early stages, diphtheria pharyngitis mimics group A streptococcal pharyngitis and mononucleosis. As diphtheria progresses, unlike in pharyngitis and mononucleosis, the exudate changes, becoming darker and forming a membrane that cannot be removed without causing bleeding. Neurologic abnormalities, such as ninth and 10th nerve deficits or ECG changes, should also alert the clinician to the possibility of diphtheria. The microbiology laboratory staff must be notified of the possibility of *C. diphtheriae* because normal throat flora usually overgrow on blood agar plates. Nonnutritive, moist, reducing transport medium is helpful in preventing the overgrowth of competitors, and samples need to be inoculated on Löffler and tellurite media for proper identification. Assays for diphtheria toxin production also need to be performed.

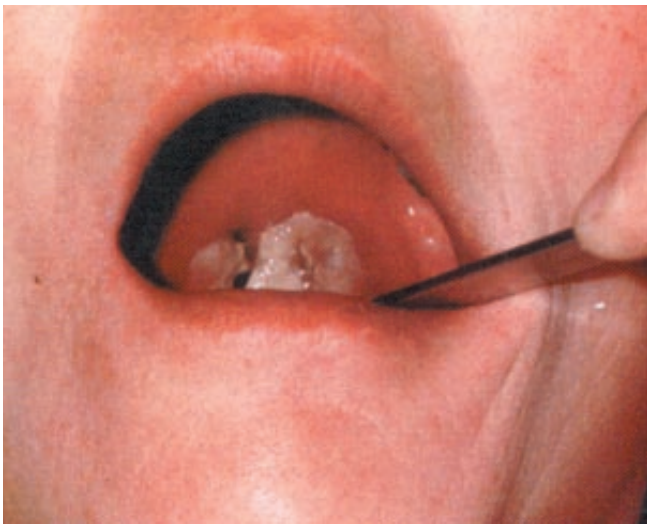


Figure 2 **Diphtheritic membrane extending across the uvula in a 47-year-old woman. Neck edema also has developed.[10]**

TREATMENT

All patients with a clinical diagnosis of respiratory diphtheria should be hospitalized and isolated, because the course of illness is unpredictable. Rapid administration of antiserum is of primary importance. The sooner diphtheria antitoxin is given, the more favorable the outcome. Antitoxin is most effective if given within 4 days after the onset of illness. The antibody blocks entry of toxin into cells and therefore is effective only in neutralizing toxin in the extracellular space before entry into the cytoplasm. Dosage of antitoxin is adjusted to the severity of disease. Patients with extensive involvement of the tonsils and pharynx or larynx, who are expected to have higher concentrations of toxin, should be given at least half the treatment dose of antitoxin by intravenous infusion over 60 minutes. If extensive disease has been present for 3 or more days or if a bull neck has developed, administration of 80,000 to 120,000 units is recommended. For milder disease of shorter duration (48 hours or less), 20,000 to 40,000 units may be used. Antitoxin is given once; repeated doses provide no added benefit. Because antiserum is derived from horse serum, approximately 10% of patients have an allergic reaction. If skin or eye testing demonstrates hypersensitivity, desensitization should be attempted.

Antibiotic Therapy

Antibiotic therapy should be initiated as soon as possible and serves three purposes: (1) it shuts off toxin production; (2) it eradicates other potential pharyngeal pathogens, including group A streptococci; and (3) it eliminates the carrier state, preventing the spread of *C. diphtheriae* to other nonimmunized persons. Erythromycin is considered the treatment of choice. Penicillin is also effective; it is given intramuscularly until the patient is able to swallow, then orally (as penicillin V) for the remainder of the 2-week course [*see Table 1*]. A study of Vietnamese children found penicillin to be more effective than erythromycin.[15] *C. diphtheriae* is generally susceptible to clindamycin and rifampin. However, there has been less experience with the use of these antimicrobial agents.

Other Measures

Extensive pharyngeal and laryngeal membrane formation can lead to upper airway obstruction. Therefore, patients need to be closely monitored, and if signs of obstruction are detected, prompt intubation or tracheostomy must be performed. Sedatives may obscure the development of respiratory difficulties and should be avoided. Patients with myocarditis need cardiac monitoring and should initially be kept at bed rest. Treatment of heart failure with digoxin may result in further impairment of electrical conduction and lead to heart block. Experience with cardiac pacemakers is limited, but pacemakers would be expected to reduce mortality from complete heart block.

Table 1 Antibiotic Treatment of Infections Caused by Gram-Positive Bacilli

Pathogen (Disease)	*Drug*	*Dosage*	*Comment*
Corynebacterium diphtheriae (diphtheria)	Erythromycin	500 mg I.V. or p.o., q.i.d. × 2 wk	First choice
	Penicillin	Penicillin G, 600,000 U I.M. b.i.d., then penicillin V p.o. 250 mg q.i.d. × 2 wk	First choice; in children, penicillin may be more effective than erythromycin
C. urealyticum	Vancomycin	1 g I.V. q. 12 hr × 2–3 wk	Use antibiotic sensitivity testing to guide therapy
C. jeikeium	Vancomycin	1 g I.V. q. 12 hr × 2–3 wk	First choice; use antibiotic sensitivity testing to guide therapy
	Penicillin G + gentamicin	20 million U/day I.V. divided q. 4–6 hr 5 mg/kg/day I.V. divided q. 8 hr	Alternative
C. ulcerans	Erythromycin	500 mg I.V. or p.o., q.i.d. × 2 wk	—
Rhodococcus equi	Vancomycin ± rifampin ± erythromycin, imipenem, amikacin, or ciprofloxacin	1 g I.V. q. 12 hr × 6 wk 600 mg p.o., q.d. 1 g I.V. q. 6 hr 0.5–1 g I.V. q. 6 hr 15 mg/kg I.V. q.d., or divided q. 8–12 hr 500–750 mg p.o. or 400 mg I.V. q. 12 hr	Prolonged therapy often required, relapse common; combination therapy recommended; no one regimen has proved to be more effective
Listeria monocytogenes	Ampicillin ± gentamicin	2 g I.V. q. 6 hr × 3–6 wk 5 mg I.V. q.d. or divided q. 8 hr	Clinical efficacy of gentamicin has not been proved
	Trimethoprim-sulfamethoxazole (TMP-SMX)	20 mg TMP, 100 mg SMX/kg/day I.V., divided q. 6–8 hr × 3–6 wk	Use in penicillin-allergic patients
Nocardia	Sulfisoxazole or sulfadiazine	1.5–2 g I.V. q. 6 hr	May not be available on hospital formulary; follow by oral sulfonamide × 10–11 mo
	TMP-SMX	20 mg TMP, 100 mg SMX/kg/day I.V., divided q. 6–8 hr × 1–2 mo	Follow by oral sulfonamide × 10–11 mo
	Minocycline	100 mg I.V. q. 12 hr	Alternative drug; may cause vertigo
	Imipenem ± amikacin	0.5–1 g I.V. q. 6 hr 15 mg/kg I.V. q.d. or divided q. 8–12 hr	Alternative

Isolation and Treatment of the Carrier State

When diphtheria is suspected, the patient should be isolated until two cultures from the infected site are negative. Cultures should be obtained from all persons who have been in close contact with the patient to determine whether they are pharyngeal carriers. All carriers need to be treated with erythromycin or penicillin for 14 days, and eradication of the carrier state must be documented by follow-up cultures.

PREVENTION

Active immunization using formalin-detoxified diphtheria toxin effectively prevents diphtheria. Preschool children (6 weeks to 7 years of age) should be immunized with three 0.5 ml intramuscular injections of diphtheria-tetanus–acellular pertussis (DTaP) vaccine spaced 4 to 8 weeks apart. A fourth dose should be given 6 to 12 months later. Immunity to the toxin is not lifelong, and if primary immunization was completed before 4 years of age, a booster is recommended at the time of school entry. Subsequently, booster injections need to be given every 10 years to maintain protective immunity. A single dose of vaccine is sufficient for most age groups, with the exception of persons 30 to 49 years of age, who may require three doses to generate protective antibody titers.[16] Surveys indicate that high percentages of adults in the United States and Europe fail to demonstrate a significant immune response to diphtheria toxoid. It has been estimated that epidemic diphtheria is favored when more than 70% of the population lacks protective immunity, a condition now present in many developed countries. A toxoid booster inoculation every 10 years is strongly recommended for all adults.

Nondiphtheria *Corynebacterium* and *Rhodococcus*

In addition to the species *diphtheriae*, the genus *Corynebacterium* contains a large number of species that for decades were considered to be culture contaminants and constituents of the normal human flora. These organisms were previously termed diphtheroids. As the number of immunocompromised hosts and persons with prosthetic devices has increased, the role of nondiphtheria corynebacteria as true pathogens has become evident[17] [*see Table 2*]. *C. urealyticum* and *C. jeikeium* are two nondiphtheria strains that are particularly important nosocomial pathogens.

C. UREALYTICUM

C. urealyticum is a slow-growing, urease-positive organism that is widely distributed on the skin of hospitalized patients. *C.*

Table 1 (*continued*)

Pathogen (Disease)	*Drug*	*Dosage*	*Comment*
Bacillus anthracis (anthrax)	Ciprofloxacin or doxycycline + one or two additional antibiotics: rifampin vancomycin penicillin G ampicillin chloramphenicol imipenem clindamycin clarithromycin	400 mg I.V. q. 12 hr followed by 500 mg p.o., b.i.d. × 60 days 100 mg I.V. q. 12 hr followed by 100 mg p.o., b.i.d. × 60 days 600 mg p.o., q.d. 1 g I.V. q. 12 hr 20 million U/day I.V. divided q. 4–6 hr 2 g I.V. q. 4 hr 1 g I.V. q. 6 hr 500–1,000 mg I.V. q. 6 hr 600 mg I.V. q. 8 hr 500 mg p.o., b.i.d.	Prolonged treatment necessary because spores may persist and later germinate; no one combination regimen preferred; some experts favor the addition of clindamycin, which theoretically may block toxin production; chloramphenicol associated with granulocytopenia
B. cereus	Vancomycin	1 g I.V. q. 12 hr	First-line therapy; duration depends on the type of infection
	Imipenem	500–1,000 mg I.V. q. 6 hr	Alternative
	Clindamycin	600 mg I.V. q. 8 hr	Alternative
	Ciprofloxacin	400 mg I.V. q. 12 hr	Alternative
Other *Bacillus* species	Vancomycin ± gentamicin	1 g I.V. q. 12 hr 5 mg/kg/day I.V. or divided q. 8 hr	First-line therapy; duration depends on the type of infection
	Clindamycin ± gentamicin	600 mg I.V. q. 8 hr 5 mg/kg/day I.V. or divided q. 8 hr	Alternative
	Imipenem	500–1,000 mg I.V. q. 6 hr	Alternative
Erysipelothrix (erysipeloid)	Penicillin G	Benzathine form; single 600,000 U I.M. dose	First-line therapy
	Erythromycin	250–500 mg p.o., q.i.d. × 10 days	Alternative
Erysipelothrix (endocarditis)	Penicillin G	20 million U/day I.V. divided q. 4–6 hr × 4–6 wk	First-line therapy; recommended for endocarditis
	Ceftriaxone	1 g I.V. q.d. × 4–6 wk	For penicillin-allergic patient
	Cefazolin	1.5–2 g I.V. q. 8 hr × 4–6 wk	For penicillin-allergic patient

urealyticum is one cause of alkaline-encrusted cystitis, a urinary tract infection that is severe and difficult to treat.[18] The urea-splitting activity of the organism leads to alkaline urine and the formation of struvite stones. Factors that predispose an individual to this infection include previous urinary tract infections, urologic instrumentation, immunosuppression, underlying inflammation of the bladder, and bladder neoplasia. Less commonly, *C. urealyticum* can cause septicemia, endocarditis, osteomyelitis, and pneumonia. This organism is often resistant to most antibiotics; therefore, vancomycin is recommended for initial treatment pending sensitivity testing [*see Table 1*].

C. JEIKEIUM

C. jeikeium frequently colonizes the skin of hospitalized patients. Patients at highest risk for colonization and subsequent infection with *C. jeikeium* include those receiving broad-spectrum antibiotics and those requiring prolonged hospitalization. In patients with neoplastic disease, other risk factors include prolonged neutropenia and breaks in the integument.[19] In most patients, infection presents as bacteremia. The incidence is highest in neutropenic patients and those who have undergone cardiac surgery.[17] Bacteremia is most often associated with colonization of an intravenous catheter.[20] Prosthetic endocarditis and native valve endocarditis have been reported, as have rare cases of extravascular infections, including cutaneous lesions, pneumonia, peritonitis, prosthetic knee infection, and ventriculostomy infection. Most isolates of this organism tend to be multiply resistant and frequently are sensitive only to vancomycin[21] [*see Table 1*]. Contaminated catheter lines can often be sterilized with antibiotics alone.[20] In patients with prosthetic valve infection, however, the foreign material often has to be removed to control the infection.

RHODOCOCCUS EQUI

Rhodococcus equi, also known as *Corynebacterium equi*, primarily infects immunocompromised hosts with defects in cell-mediated immunity, particularly patients with AIDS[22] and those with solid-organ transplants.[23] Cases in normal hosts have also been reported.[24,25] *R. equi* is found in the soil and at particularly high concentrations in horse manure. Infection is generally acquired through the lungs.

The primary manifestation in most cases is cavitary lung disease resembling tuberculosis, nocardiosis, or fungal infection. Lung consolidation without cavitation is also seen. Bronchoscopy,

Table 2 Clinical Syndromes Caused by Nondiphtheria *Corynebacterium* Species

Species	*Clinical Syndrome*
C. ulcerans	Pharyngitis, skin ulcer, diphtheria
C. pseudotuberculosis (C. ovis)	Suppurative lymphadenitis
C. (Arcanobacterium) haemolyticum	Pharyngitis, scarlatiniform rash
C. pseudodiphtheriticum (C. hofmannii)	Endocarditis, pneumonia, tracheobronchitis, lymphadenitis
C. urealyticum (formerly group D2)	Alkaline-encrusted cystitis, urinary tract infections
C. jeikeium (group JK)	Nosocomial septicemia, wound infection, endocarditis
Rhodococcus equi (C. equi)	Necrotizing pneumonia in patients with AIDS

thoracentesis, or surgery may be required to make the diagnosis, although blood cultures are frequently positive.[24] *R. equi* organisms may be mistaken for contaminating diphtheroids, or because they are modified acid-fast positive, the infection may be misdiagnosed as tuberculosis. Extrapulmonary infections may also occur.

Macrolides and rifampin act synergistically in combination, and regimens containing these two agents are often recommended. These antibiotics achieve high intracellular levels, an important characteristic for clearing *R. equi*, because this pathogen primarily multiplies in cells. The organism is also sensitive to vancomyin and aminoglycosides, and vancomycin is often included in the initial regimen [*see Table 1*]. Cephalosporins should be avoided because of the frequent development of resistance, and multidrug resistance is becoming more common.[26] Short courses of therapy are associated with relapse; therefore, therapy needs to be continued for many weeks. Mortality in AIDS patients is approximately 15%; however, half of the survivors are never completely cured of *R. equi* infection.[22]

Listeriosis

Listeria monocytogenes, a food-borne pathogen, is the cause of listeriosis, a serious and often fatal infection.

MICROBIOLOGY

L. monocytogenes is an aerobic and facultatively anaerobic, non–spore-forming, gram-positive rod. As opposed to corynebacteria, this organism is motile, possessing one to five flagella. *Listeria* organisms grow at a wide range of temperatures (3° to 42° C), which explains its ability to contaminate refrigerated foods. This bacterium can also grow at acidic concentrations of a pH of 5 or higher and salt concentrations of as high as 10% to 12%. *Listeria* organisms can be readily cultured on blood agar plates, where they cause slight zones of β-hemolysis. *Listeria* organisms on occasion can appear somewhat coccoid and, therefore, may be mistaken for diphtheroids or *Streptococcus pneumoniae* on Gram stains of cerebrospinal fluid. There are at least 11 serotypes of *L. monocytogenes*. Serotypes 1b and 4b are most commonly associated with listeriosis.

ETIOLOGY AND EPIDEMIOLOGY

L. monocytogenes is found in soil, dust fertilizer, sewage, stream water, plants, processed foods, and the intestinal tract of many mammals. Investigations of multiple outbreaks indicate that both sporadic and common-source outbreaks of listeriosis are the result of food contamination. Outbreaks have been linked to raw vegetables, Mexican-style cheese, milk, undercooked chicken, and foods purchased in delicatessens.[27] Prepared refrigerated foods stored for prolonged periods and requiring no further high-temperature heating are most likely to be contaminated because *Listeria* organisms can readily multiply on refrigerated foods.

The overall incidence of listeriosis is low: 0.7 cases per 100,000 population. This infection more often occurs in persons older than 70 years (2.1 cases per 100,000); pregnant women (12 cases per 100,000); patients with defects in cell-mediated immunity, including renal transplant recipients; patients receiving high doses of corticosteroids; and patients with AIDS (100 cases per 100,000).[28] At a large referral hospital, *Listeria* organisms were the third most common cause of community-acquired bacterial meningitis in adults (12% of cases).[29] Despite its relatively low incidence, listeriosis concerns public health officials because this disease is associated with a high fatality rate (23%), unlike infections from other food-borne pathogens, such as *Salmonella* organisms, which are rarely fatal. Given the increasing numbers of elderly and immunocompromised patients, the incidence of listeriosis is likely to increase.

PATHOGENESIS

L. monocytogenes has an unusual life cycle[30] [*see Figure 3a*]. Several proteins (internalins[31]) on the surface of the bacterium allow attachment and subsequent ingestion of *Listeria* organisms by host cells. Once internalized, the bacterium is surrounded by host cell membrane, forming a phagolysosome, a closed space that is generally toxic for pathogens. *Listeria* organisms evade destruction by producing the exotoxin listeriolysin O, which lyses the confining membranes. All pathogenic strains of *Listeria* organisms produce listeriolysin O, and their escape into cytoplasm of the host cell is required for pathogenesis. Once in the growth-permissive cytoplasm, the bacteria proliferate, with doubling times of about 1 hour. The *Listeria* surface protein ActA possesses binding sequences to attract actin regulatory proteins that stimulate actin filament assembly.[30,32] About 2 hours after entry into the cytoplasm, actin filaments polarize at one end of the bacteria and provide the force for movement through the cytoplasm [*see Figure 3b*]. Many of the bacteria migrate to the periphery of the cytoplasm, where they push against the host cell's outer membrane to form elongated protrusions or filopods that can be ingested by adjacent cells. Once a bacterium enters the adjacent cell, the life cycle begins anew. *Listeria* organisms, therefore, can spread from cell to cell without directly contacting the extracellular environment. A number of other virulence factors, in addition to ActA and internalins, have been identified, and their contributions to *Listeria* pathogenesis are being defined.

The *Listeria* organism's intracellular lifestyle explains many of this pathogen's unique clinical characteristics.[30] Although the association between contaminated foods and the *Listeria* organism has been well documented, evidence of gastrointestinal disease has been absent in most cases. The *Listeria* organism's capacity to enter the gastrointestinal tract without causing erosive lesions is explained by the ability of this pathogen to stimulate phagocytosis by gastrointestinal cells and macrophages. Subsequently, the *Listeria* pathogen commandeers host cell actin regulatory proteins to spread from cell to cell and eventually enter the bloodstream either in monocytes or as free organisms after cell lysis.

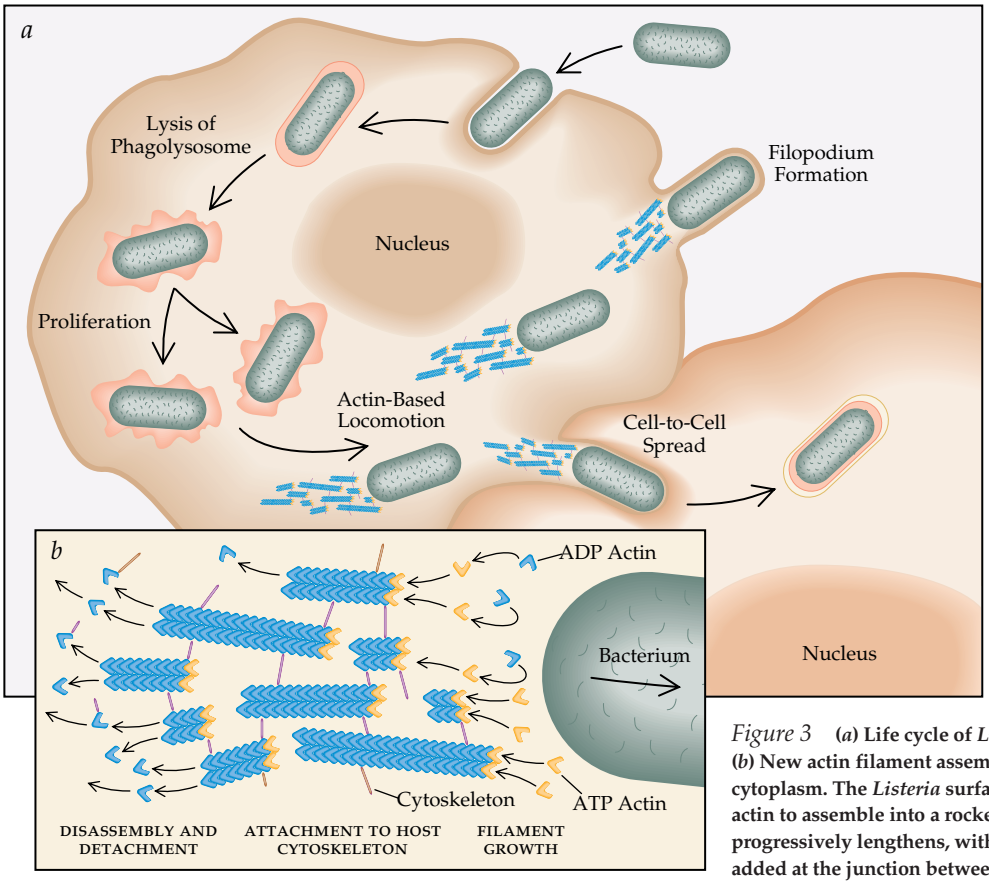


Figure 3 **(*a*) Life cycle of *Listeria monocytogenes* in host cells. (*b*) New actin filament assembly drives the bacterium through the cytoplasm. The *Listeria* surface protein ActA induces host cell actin to assemble into a rocket tail. The actin filament tail progressively lengthens, with new host cell actin monomers being added at the junction between the bacterium and the actin filament tail. The older regions of the actin tail attach to the host cell's cytoskeleton, providing a purchase so that the forces of actin filament lengthening can be applied to the bacterium to drive it through the cytoplasm. Bacteria are able to move at rapid speeds (0.02 to 1.4 µm/sec).[30]**

The ability of the *Listeria* organism to avoid the extracellular environment also explains the increased incidence of listeriosis in immunocompromised patients, neonates, and pregnant women. Increased risk of listeriosis has not been associated with deficiencies of immunoglobulins or complement. However, clinical conditions and therapies (particularly corticosteroids) that lead to deficiencies in cell-mediated immunity, the primary defense for controlling intracellular pathogens, increase the risk for listeriosis. For example, treatment with fludarabine and prednisone in patients with chronic lymphocytic leukemia markedly lowers the $CD4^+$ T cell counts and increases the incidence of listeriosis.[33] Patients with AIDS are most likely to contract *Listeria* infection when their $CD4^+$ T cell counts fall below $40/mm^3$.[34]

Clinical Manifestations

Listeriosis varies in its clinical presentation; primary manifestations most often are sepsis, meningitis, or both. Other extravascular infections are also reported, but they are surprisingly rare.

Infection during pregnancy Nearly one third of individuals who contract listeriosis are pregnant women, with infection occurring most frequently in the third trimester. The symptoms tend to be relatively mild, consisting of a flulike illness with chills, fever, and muscle aches. Back pain, a less frequent complaint, suggests a urinary tract infection; however, urinalysis and urine culture results are normal. Blood cultures, although not always obtained, are positive. Symptoms usually resolve spontaneously without therapy.[35]

Neonatal listeriosis The *Listeria* organism can cross the placental barrier, probably as a result of cell-to-cell spread mediated by host cell actin. The organism may cause amnionitis and precipitate premature labor, leading to septic abortion. Transplacental transmission can also cause the unique clinical syndrome of granulomatosis infantiseptica. The organism can disseminate in utero, forming abscesses and granulomas involving the fetal liver, spleen, lungs, kidneys, brain, and skin.[36] Mortality is high, ranging from 35% to 55%.

Chorioamnionitis is the most common early manifestation of perinatal *Listeria* infection. A Gram stain of the meconium frequently reveals gram-positive bacilli, suggesting the diagnosis. In addition to causing congenital infection, the *Listeria* pathogen can cause meningitis in neonates 7 to 28 days of age. Listerial meningitis is the third most common form of meningitis in neonates, accounting for 5% to 15% of cases.

Adult meningitis and meningoencephalitis Meningitis is the most common manifestation of listeriosis. The *Listeria* pathogen has a predilection for the central nervous system, particularly the meninges. Although the clinical presentation of listerial meningitis is similar to that of other forms of bacterial meningitis (i.e., *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis*), several characteristics distinguish *Listeria* infections. Meningeal signs develop less frequently in patients with listerial meningitis than in patients with other forms of bacterial meningitis.[37] However, tremor and grand mal or focal motor seizures are observed with a higher frequency, suggesting more

extensive invasion of the CNS.[30] The CSF response may reflect the intracellular nature of *Listeria* organisms. Compared with the CSF cell counts in other forms of bacterial meningitis, both the total number of white blood cells and the percentage of neutrophils tend to be reduced in patients with listerial meningitis. Because *Listeria* primarily grows in cells, it is less commonly found in Gram stains of CSF (positive in 5% to 33% of cultures versus 80% in other forms of bacterial meningitis).[38]

Meningoencephalitis, a direct invasion of the cerebral cortex, can also result from *Listeria* infection, but it is not a recognized complication of other forms of bacterial meningitis. The ability of the *Listeria* pathogen to cross the meninges and blood-brain barrier is also likely to be the result of endothelial cell or macrophage phagocytosis of the organisms and utilization of the host cell contractile system to migrate to and grow in the brain. *Listeria* organisms most commonly invade the brain stem, causing a syndrome that has been called rhomboencephalitis.[39] The disease is usually biphasic; CNS manifestations are preceded by 7 to 10 days of malaise, fatigue, headache, nausea or vomiting, and fever. These symptoms are followed by the development of multiple cranial nerve deficits, particularly of the sixth and seventh nerves. Brain stem damage can also cause hemiparesis, ataxia, and respiratory dysfunction, often followed by respiratory arrest. In cases of meningoencephalitis, CSF monocyte counts may reach 80% to 90%. In some cases, the CSF may contain no cells or only a few cells and have normal protein and glucose levels. Magnetic resonance imaging reveals areas of increased resonance and is the best way to visualize the brain stem. Computed tomography with contrast generally shows areas of increased uptake with or without ring enhancement.

The diagnosis of both listerial meningitis and meningoencephalitis frequently is delayed. A monocytic response in the CSF and a negative CSF Gram stain can lead the clinician to confuse listerial meningitis and meningoencephalitis with herpes and other forms of viral encephalitis, viral meningitis, tuberculous meningitis, Lyme disease, syphilis, cryptococcal meningitis, Wegener granulomatosis, or CNS sarcoidosis. Particularly in the immunocompromised host, the possibility of *Listeria* infection must always be considered as a possible cause of CNS infection.

Bacteremia Bacteremia without meningitis or focal infections occurs in 5% to 30% of adult cases. There are no distinctive clinical features except for peripheral monocytosis, which is present in a small percentage of patients. Diagnosis is based on blood culture findings.

Miscellaneous infections Like many other pathogens that are able to enter the bloodstream, the *Listeria* organism can cause focal infections at many extravascular sites, including bones, normal and prosthetic joints, eyes,[40] spinal cord, pleura, peritoneum, and liver.[41] Isolated brain abscesses have also been reported.[37] In rare cases, the *Listeria* organism causes endocarditis, myocarditis, and mycotic aneurysms. Although *Listeria* enters via the gastrointestinal tract, symptomatic gastroenteritis is uncommon. However an outbreak of febrile gastroenteritis associated with contaminated delicatessen precooked turkey was reported in Los Angeles.[42]

Laboratory Tests

L. monocytogenes can be readily cultured, although a diphtheroid-like organism discovered in blood or CSF cultures is frequently misinterpreted as a contaminant. Diphtheroids can be rapidly differentiated from *Listeria* organisms by using a microscopic bacterial motility test. Decreasing motility is indicative of *Listeria* organisms. Gram stains of cerebrospinal or meconium fluid showing gram-positive rods strongly suggest *Listeria* infection. In an immunocompromised host, treatment of listeriosis should be initiated pending the laboratory staff's final diagnosis.

Table 3 Dietary Recommendations for Preventing Food-Borne Listeriosis

General recommendations
Thoroughly cook raw food from animal sources
Wash raw vegetables thoroughly
Keep uncooked meats separate from vegetables and cooked foods
Avoid consumption of unpasteurized milk or foods made with raw milk
Wash hands, knives, and cutting boards after handling uncooked foods
Additional recommendations for high-risk persons*
Avoid soft cheeses (e.g., Mexican style, feta, Brie, Camembert, and blue veined); hard cheeses, cream cheese, cottage cheese, and yogurt can be eaten
Leftovers or ready-to-eat foods (e.g., hot dogs) should be reheated until they are steaming hot
Pregnant women and immunosuppressed persons should consider avoiding foods from delicatessen counters or thoroughly cooking cold cuts before eating

*High-risk persons include immunocompromised persons, pregnant women, neonates, and the elderly.

PREVENTION AND TREATMENT

Preventive Measures

It is not possible to eliminate the large reservoir of *Listeria* organisms found throughout the environment. Physicians can help prevent disease by instructing patients at risk on how to minimize the multiplication of *Listeria* organisms in foodstuffs and how to kill the organisms on potentially contaminated foods. Patients need to avoid unsterilized dairy products, undercooked meats, and prepared foods that have been refrigerated but not resterilized by high-temperature reheating. More detailed preventive instructions are provided by the Centers for Disease Control and Prevention (CDC) [*see Table 3*].

Antibiotic Therapy

No clinical trials comparing various antibiotic regimens have been published. Bacteriostatic drugs, such as chloramphenicol and tetracycline, are associated with high failure rates in patients with listeriosis and cannot be recommended. Ampicillin is generally recommended as the treatment of choice, and the *Listeria* pathogen is generally sensitive[43] [*see Table 1*]. In immunosuppressed patients, relapse has been reported after 2 weeks of penicillin therapy. The poor response to bacteriostatic drugs and the slow response to ampicillin therapy probably result from the *Listeria* organism's ability to survive and grow in cells. The intracellular level of ampicillin may not be sufficient for complete sterilization. Immunosuppression reduces the host's ability to clear infected cells. The exact duration of antibiotic treatment required to prevent relapse is not known; however, 3 to 6 weeks of therapy is probably prudent for immunosuppressed patients. Antibi-

otics that penetrate cells poorly, such as aminoglycosides, may be synergistic in vitro but are unlikely to prove efficacious in the living host. Although some experts have recommended that an aminoglycoside be added to ampicillin, the *Listeria* organism grows in cells in the presence of extracellular gentamicin concentrations of 10 to 20 µg/ml. Furthermore, addition of gentamicin to ampicillin therapy has failed to improve outcome in a mouse infection model.[44] Therefore, aminoglycosides are unlikely to work in patients with listeriosis and certainly should be avoided in kidney transplant recipients and other patients with renal dysfunction. On the other hand, trimethoprim-sulfamethoxazole, a drug combination that readily enters cells and is bactericidal for *Listeria* organisms, may prove to be the most effective agent for treating *Listeria* infection. Trimethoprim-sulfamethoxazole has proved to be effective in patients with listeriosis and penicillin hypersensitivity [*see Table 1*].[39] Tissue culture and animal studies suggest that levofloxacin may also be effective; however, treatment of human *Listeria* infection with this quinolone has not been reported.[45]

Listeria infection in the CNS is associated with a high mortality despite appropriate antibiotic treatment. In patients with meningoencephalitis, mortality ranges from 36% to 51%. Survivors often have permanent neurologic sequelae. Mortality in patients with meningitis is somewhat lower (26%), but it is higher in patients with seizures and in those older than 65 years.[37] Early recognition and rapid institution of antibiotics are critical for improving outcome.[39]

Nocardiosis

MICROBIOLOGY

Nocardia species are thin, aerobic, gram-positive bacilli that form branching filaments that tend to fragment into coccobacilli. Gram stain is often taken up, variably resulting in an irregular, beaded appearance in exudates [*see Figure 4*]. *Nocardia asteroides* is the most common *Nocardia* species to cause human disease in the United States (80% to 90% of cases). *N. brasiliensis*, *N. otitidiscaviarum*, and *N. farcinica*[46] are rarer human pathogens. All four organisms are commonly found in soil. In addition to being gram positive, they are modified acid-fast positive. This characteristic helps differentiate *Nocardia* species from the *Actinomyces* organism, a gram-positive, anaerobic pathogen that also forms branching filaments. The *Nocardia* organism grows slowly on blood agar plates and Sabouraud dextrose agar. In sputum specimens, other organisms often overgrow, obscuring *Nocardia* colonies. Colonies generally have an orange pigment and appear heaped up and folded. They can also be white or pink. Colony characteristics suggestive of the *Nocardia* organism may not develop for 2 to 4 weeks. Samples should be obtained when the patient is not taking antibiotics. Modified Thayer-Martin medium and buffered charcoal–yeast extract agar can be used to enhance recovery of *Nocardia* species.

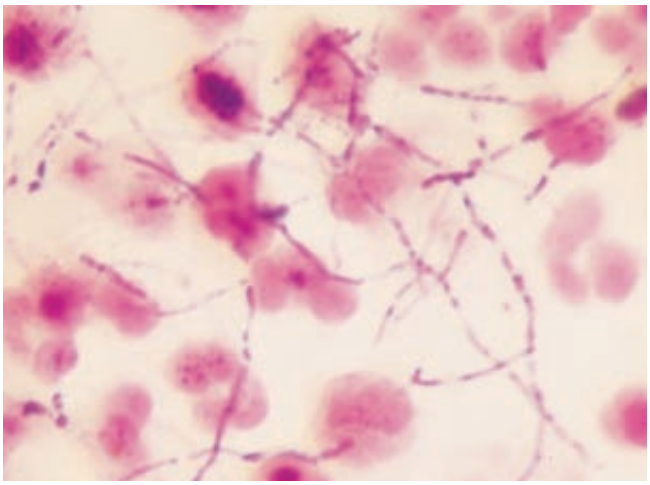


Figure 4 **Gram stain of joint fluid containing *Nocardia asteroides*.**

ETIOLOGY AND EPIDEMIOLOGY

Nocardiosis is relatively rare, infecting approximately 1,000 persons a year in the United States. Often found in soil, *Nocardia* organisms most often infect immunocompromised hosts and primarily cause pulmonary infections, brain abscesses, and skin infections.

PATHOGENESIS

In most cases, the *Nocardia* organism gains entry to the host through the respiratory tract. Inhaled bacteria elicit a neutrophil response that inhibits but does not kill the organism. *Nocardia* organisms are phagocytosed by neutrophils and incorporated into phagolysosomes. In this closed membrane space, the organism is able to survive for prolonged periods. Like *Mycobacterium tuberculosis*, pathogenic *Nocardia* species produce superoxide dismutase, a product that inactivates the toxic oxygen by-products of the neutrophils and macrophages. In addition, both *M. tuberculosis* and *Nocardia* species synthesize a second product: a mycolic acid that inhibits the fusion of lysosomes with the phagolysosomal compartment. This inhibitory activity prevents the release of toxic proteases and other antibacterial products that would otherwise reach the intracellular bacteria. The host defenses utilized to protect against nocardiosis are multifactorial and include neutrophils, macrophages, and cell-mediated and humoral immunity. Patients at risk for nocardiosis include those with chronic granulomatous disease (which compromises the ability of neutrophils to produce toxic oxygen by-products) and those with dysgammaglobulinemia. The highest percentage of cases occur in patients with impaired cell-mediated immunity, including renal and cardiac transplant patients,[47] other patients on high-dose corticosteroids, patients with Cushing disease, cancer patients,[48] and those with AIDS who have $CD4^+$ T cell counts below 250 cells/mm^3.[49] Patients with chronic pulmonary disorders, particularly alveolar proteinosis, are also at increased risk. Approximately one third of those who acquire nocardiosis have no identifiable predisposing condition.

DIAGNOSIS

Clinical Manifestations

Nocardiosis has no pathognomonic characteristics, and delays in diagnosis are common. This infection must always be considered in the immunocompromised host.

Pulmonary nocardiosis Pulmonary infection is the most common manifestation of nocardiosis, occurring in approximately two thirds of cases. In most cases, pulmonary disease is subacute in onset, mimicking pulmonary tuberculosis. Patients complain of productive cough, pleuritic chest pain, dyspnea, fever, anorexia, and weight loss. Occasionally, hemoptysis de-

velops, particularly in patients with large cavitary lung lesions; if left untreated, the disease tends to run a chronic waxing-and-waning course. Chest x-ray findings are variable and include the following, in order of most to least frequent: pulmonary nodules or mass lesions, areas of consolidation, cavitary lesions, interstitial infiltrates, and pleural effusions. In addition, CT may demonstrate areas of low attenuation in consolidations, multiple nodules, and chest wall extension of the infection. AIDS patients are more likely to have multiple pulmonary nodules, cavitary lung lesions, and upper lobe infiltrates.[49] On occasion, the infiltrate spontaneously resolves, particularly in patients with normal immune function. A brain abscess may develop later as a consequence of transient dissemination of the organism.

CNS infection In approximately one third of patients with nocardiosis, the CNS becomes infected. Brain abscess is the most common CNS infection. The lesions are often multilocular and can occur in any region of the brain.[50] CT or MRI with contrast demonstrates ring-enhancing lesions, as observed with other causes of pyogenic brain abscess. In AIDS patients, brain abscess is often accompanied by an abnormal chest x-ray, which suggests the diagnosis.[51] With other patients, abnormalities on chest x-ray are not always present. However, when abnormalities are detected, the combined findings of a lung nodule on chest x-ray and ring-enhancing CNS lesions are often mistaken for lung carcinoma with metastasis. Patients treated with surgical drainage of their abscesses have a higher survival rate than those who are not.[50] Nocardial meningitis is a less common CNS manifestation and is often associated with brain abscess (40% of meningitis cases). Patients with nocardial meningitis typically have subacute to chronic meningitis characterized by fever, stiff neck, and headache. CSF analysis demonstrates a predominance of polymorphonuclear neutrophils, a low CSF glucose level, and an elevated protein level. CSF cultures are often negative, particularly in the first 3 days of the disease, and patients fail to fully respond to empirical antibiotic therapy. Appropriate treatment is often delayed, and mortality is high (50% to 60%).

Cutaneous infection Cutaneous involvement is uncommon and is generally caused by *N. brasiliensis*. A break in the skin caused by trauma, an insect bite, a thorn bush scratch,[52] and even a cat scratch can result in local invasion by *Nocardia* organisms. A pustule or a moderately erythematous, nonfluctuant nodule develops at the site of inoculation. Regional adenopathy is generally found. The presence of multiple subcutaneous nodules indicates dissemination of the organism and more often occurs in the immunocompromised patient. In the tropical regions of South and Central America, ulcerations and large tumorlike lesions called mycetomas occur on the lower legs and are caused by *N. asteroides*.

Other infections Dissemination occurs in approximately 40% of pulmonary *Nocardia* infections and can result in localized infection in any organ. Septic arthritis, osteomyelitis, endophthalmitis, sinusitis, peritonitis, and purulent pericarditis have all been reported.

Laboratory Tests

Invasive procedures are generally required to obtain infected tissue samples. The histopathology of biopsy specimens usually reveals an acute inflammatory response with a predominance of neutrophils. Tissue necrosis with minimal fibrosis often results in the formation of multilocular abscesses with minimal capsular formation. Gram stain or Brown-Brenn stains often reveal gram-positive beaded branching forms. *Nocardia* species can also be visualized by using a modified acid-fast stain. The organism is not well seen after hematoxylin-eosin or periodic acid–Schiff stain. Culture is the definitive way to prove the diagnosis. For sputum cultures, selective media may be required to prevent the overgrowth of more rapidly growing mouth flora. For the diagnosis of meningitis, large volumes of CSF should be obtained for culture.

Nocardia organisms are slow growing and are difficult to identify on routine culture. When a potential case is encountered, it is important for the clinician to alert microbiology and pathology laboratory staffs that *Nocardia* is a possible pathogen, so that cultures can be incubated for a longer period.

TREATMENT

Sulfonamides alone and trimethoprim-sulfamethoxazole remain the treatments of choice. High intravenous doses of these agents are required: sulfadiazine (1.5 to 2 g every 6 hours) or trimethoprim-sulfamethoxazole (20 mg/kg/day of trimethoprim and 100 mg/kg/day of sulfamethoxazole given in three divided doses) to maintain serum sulfonamide levels in the 12 to 15 mg/dl range. Once substantial improvement is documented, oral treatment can be substituted after 1 to 2 months of intravenous therapy. For sulfa-allergic patients, possible alternatives need to be determined on the basis of sensitivity testing. Minocycline, imipenem, amoxicillin–potassium clavulanate, and amikacin alone or in various combinations have been successful in individual patients [*see Table 1*].[53] Because of the intracellular nature of nocardiosis and the organism's slow growth rate, 12 months of antibiotic therapy is required to prevent relapse. In cases of abscess formation in the brain, subcutaneous tissue, or other organs (except the lung), surgical drainage is also required for cure.

PROGNOSIS

The overall mortality from nocardiosis is approximately 25%. Otherwise healthy persons with pulmonary nocardiosis have a better prognosis (15% mortality). Fatality rates are higher in patients with bacteremia,[54] patients with acute infection (symptomatic for less than 3 weeks), patients receiving corticosteroids or cytotoxic agents, patients with disseminated disease involving two or more noncontiguous organs, and patients with meningitis.

Anthrax

Bacillus anthracis causes infections primarily in animals, particularly herbivores. However, contact with animals or animal products can produce *B. anthracis* infections in humans. Although *B. anthracis* was once the cause of severe epidemics, our understanding of this pathogen's epidemiology and the vaccination of domestic animals have resulted in a marked reduction of naturally acquired anthrax in the United States. In developing countries, however, outbreaks associated with exposure to animals and animal products continue to be reported. In October 2001, anthrax spores were used in a bioterrorist attack in the United States.[55] This attack emphasized the need for all health care personnel to be familiar with the clinical manifestations, diagnostic approach, treatment, and prevention of anthrax. The identification of a single case of *B. anthracis* is now a cause for alarm.[56]

MICROBIOLOGY

B. anthracis is an aerobic, gram-positive rod that forms endospores. The spores are highly resistant to adverse conditions and are able to survive at extreme temperatures, at high pH and salinity levels, and in disinfectants. The organism can be readily cultured on standard blood and nutrient agar plates. For contaminated specimens (e.g., stool), selective media or decontamination methods can be used that take advantage of the spores' ability to resist heat, ethanol, and various antibiotics. The spores germinate when they are exposed to an environment rich in amino acids, nucleosides, and glucose, such as the blood or tissues of a mammalian host. Once germination occurs, the organism multiplies rapidly. On blood agar plates, vegetative bacteria form gray-white colonies and are nonhemolytic. A fluorescent antibody stain can be used to identify the organism.

ETIOLOGY AND EPIDEMIOLOGY

Anthrax is primarily a disease of herbivores (cattle, sheep, horses, goats, and swine). Humans become infected as a result of contact with infected animals (agricultural exposure) or through exposure of infected animal products (industrial exposure). The October 2001 bioterrorist attack through the United States mail system has emphasized the danger of anthrax spores as a biologic weapon. Accidental laboratory-related infections have also been reported. *B. anthracis*, like all *Bacillus* species, is a saprophyte that grows in the soil, and animals generally contract the infection through contact with soil. Because domestic animals in the United States are vaccinated against anthrax, agricultural exposure is rare, and the diagnosis of anthrax in the United States or other developed countries should immediately raise the possibility of a bioterrorist attack.

Most cases of anthrax in the United States have occurred as a result of contact with animal products imported from Asia, the Middle East, and Africa. Wool, goat hair, and animal hides are the most common sources of infection. Persons who work in the early stages of processing these materials are exposed to the highest inoculum of spores and are most likely to contract disease. Processed materials have also caused human disease. Cases have been traced to shaving-brush bristles, wool coats, yarn, goat-skin bongo drums, and heroin preparations. The largest recent outbreak of anthrax occurred in Sverdlovsk (now Yekaterinburg), Russia, in 1979, resulting in approximately 96 inhalation cases and 64 deaths. Accidental aerosol release of anthrax spores from a germ-warfare facility is suspected, and polymerase chain reaction (PCR) analysis of tissue samples from 11 victims has identified multiple strains of *B. anthracis* in each sample, consistent with infection by a manufactured preparation of bacterial spores.[57] In October 2001, *B. anthracis* spores were sent in at least five letters to Florida, Washington, D.C., and New York City, resulting in 22 cases of anthrax (11 cases of confirmed inhalational anthrax and 11 cases of cutaneous disease [seven confirmed and four suspected]). Several cases developed as a consequence of cross-contamination of mail, and a number of postal workers were infected by spores aerosolized during mail processing.[55]

PATHOGENESIS

Spores gain entry into the epidermis through abrasions in the skin and can be inhaled into the lungs as 1 to 5 µm particles. Once in the host, the spores germinate, multiply, and produce toxins that cause tissue edema and necrosis. In the lungs, spores are ingested by macrophages, where many are lysed and destroyed. However, surviving spores are transported to the mediastinal lymph nodes, where they germinate, multiply rapidly, and quickly enter the vascular system, causing bacteremia. Extrapolation from monkey experiments indicates that inhalation of 2,500 to 55,000 spores is required to cause fatal disease in 50% of humans. However other experimental data suggest that the inhalation of as few as 1 to 3 spores may be sufficient to cause disease, and two cases of fatal inhalation anthrax in New York City and Connecticut suggest that in some individuals, fatal doses may be quite low.[56] Coating of anthrax spores to prevent their aggregation improves their ability to infect the lung (such spores are termed weaponized). The spores used in the United States attack were weaponized, which explains the efficiency of infection. Another major epidemiologic concern is the duration of risk for contracting disease after the inhalation of spores. In Sverdlovsk, cases occurred up to 43 days after exposure,[58] and in monkey experiments, viable spores were found in mediastinal lymph nodes 100 days after the spores were inhaled.[59]

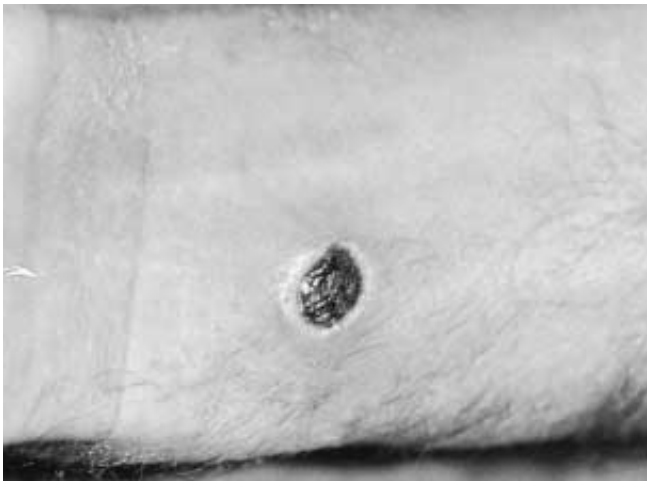


Figure 5 **Typical dark, necrotic, painless pruritic skin lesion of anthrax on the wrist of a shepherd from Morocco.**

Three toxin components are synthesized on the bacterial surface and account for the major pathologic consequences of infection: protective antigen, lethal toxin, and edema toxin. Protective antigen binds host cell receptors and transports either lethal or edema toxin in the cells, which can cause cell swelling and death. Lethal toxin has a protease activity that cleaves specific kinases, which in turn may induce cell lysis.[60,61]

DIAGNOSIS

Clinical Manifestations

Skin infection Skin disease is the most common manifestation of anthrax, and half of the victims (11 of 22) in the United States bioterrorist attack presented with cutaneous disease. In the absence of exposure to animals or animal products, the diagnosis of cutaneous anthrax should immediately raise the possibility of a bioterrorist attack. One to 7 days after inoculation of spores into the skin, a small papule develops and progresses to a vesicle over the ensuing few days. Erythema and nonpitting edema often surround the vesicle. Initially, the vesicular fluid is serous and contains large numbers of organisms. Once the vesicle ruptures, a black eschar becomes evident at the base of the ulcer [*see Figure 5*]. Anthrax derives its name, which is taken from

the Greek word for coal, from these characteristic black lesions. Despite the erythema and swelling, lesions are not painful but may be mildly pruritic. Lymphangitis, lymphadenopathy, fever, and malaise may accompany the skin infection. After 1 to 2 weeks, the skin lesion dries and a permanent scar is formed. Lesions occur primarily on exposed regions of the body. The arms are the most frequent site of infection; the face and neck are also commonly involved. Lesions are usually single but may occur at multiple sites as a result of simultaneous inoculations.[62]

Respiratory infection (Woolsorters' disease) Unlike cutaneous anthrax, which is rarely fatal, inhalational anthrax is usually a fulminant disease with a high mortality. The index case in the United States bioterrorist attack presented as typical inhalational anthrax, followed quickly by sepsis and meningitis.[63] The pulmonary form of disease is usually biphasic in its presentation.[64] From 1 to 5 days after inhalation of spores, the patient has symptoms suggestive of a viral syndrome: nonproductive cough, malaise, fatigue, myalgia, and mild fever. Occasionally, the sensation of chest heaviness is reported. Other manifestations reported in the United States bioterrorist cases included sweats, often drenching; nausea and vomiting; abdominal pain; headache; confusion; and sore throat.[65] Rhonchi may be heard on pulmonary auscultation. Within 2 to 4 days, symptoms temporarily resolve but are rapidly followed by the second, more severe stage of the disease. This stage involves the sudden onset of severe respiratory distress with dyspnea, cyanosis, and diffuse diaphoresis, accompanied by fever, tachycardia, and tachypnea. On pulmonary auscultation, moist, crepitant rales are evident, and findings are consistent with pleural effusion. Chest x-ray often demonstrates a widened mediastinum. Infiltrates or consolidation may also be seen, and pleural effusions are commonly found.[65] Death often occurs within 24 hours and may be preceded by septic shock. A Russian patient was described as dying suddenly, in midsentence. Recent experience in the United States suggests that early aggressive antibiotic therapy may modify the outcome, and death rates were reduced from above 85% to below 50%.[56] Autopsy usually reveals hemorrhagic necrosis of the thoracic lymph nodes, drainage from the lungs, and hemorrhagic mediastinitis. A high index of suspicion is critical because any delay in diagnosis and treatment greatly increases the likelihood of a fatal outcome.

Gastrointestinal infection Gastrointestinal anthrax has not been reported in the United States. It occurs primarily in developing countries, usually after ingestion of contaminated meat. The incubation period is usually 3 to 5 days. Patients initially have nausea, vomiting, anorexia, and fever. Acute abdominal pain, hematemesis, and bloody diarrhea follow rapidly. Findings on examination suggest an acute surgical abdomen, and there is moderate leukocytosis with immature band forms. Rapid progression to toxemia and shock leads to death within 2 to 5 days after the initial onset of symptoms.

An oropharyngeal form of anthrax has also been described. Inflammatory lesions that resemble the cutaneous lesions develop on the posterior pharynx, hard palate, or tonsils. Tissue necrosis and edema are accompanied by sore throat, dysphagia, fever, regional lymphadenopathy, and toxemia.

Meningitis Anthrax meningitis can result from bacteremia precipitated by cutaneous, respiratory, or gastrointestinal infection. This complication is relatively rare, occurring in fewer than 5% of patients. In the index case in the United States bioterrorist attack, the patient presented to the emergency department with confusion, and gram-positive rods were identified in his CSF.[63] The onset of meningeal symptoms usually occurs simultaneously with the primary lesion or within several days after its onset. Meningitis is hemorrhagic and rapidly fatal (within 6 days).

Patient History and Laboratory Tests

A careful epidemiologic history is the single most important means of suggesting the diagnosis. A history of contact with herbivores or products from these animals, particularly if the products come from outside the United States, should raise the possibility of anthrax. The sudden appearance of several cases of severe acute febrile illness with a fulminant fatal outcome should raise the possibility of a bioterrorist attack. Certain occupational groups are at higher risk of being exposed to anthrax spores disseminated in the mail: post-office workers, members of the news media, and politicians and their staffs. Therefore, the occupational history can provide a critical clue for recognizing early inhalational and cutaneous anthrax.

The physical appearance of the skin lesions is characteristic, and Gram stains and cultures of the ulcer base are frequently positive. A history of exposure to dust from a contaminated animal product can usually be obtained in the prodromal phase of the respiratory illness, when symptoms are mild. In the absence of a thorough history, the disease is initially mistaken for a viral respiratory illness or bronchitis. Patients with gastrointestinal disease have a history of eating undercooked, often spoiled, meat.

Gram stain of the peripheral blood may reveal gram-positive bacilli, and in cases of meningitis, the CSF Gram stain is often positive. Blood cultures are positive in most cases of inhalational anthrax, and specimens for culture should be drawn immediately. The microbiology laboratory must be alerted to the possibility of anthrax, or the organism may be identified only as a *Bacillus* species and the pathogen misinterpreted as a contaminant. In the United States, the public health Laboratory Response Network, consisting of 81 clinical laboratories, has been established to specifically identify bioweapon pathogens, and all suspected samples should be referred to one of these laboratories for confirmatory diagnosis.[56]

In addition to microbiologic-study results, a chest x-ray showing a widened mediastinum, infiltrates, and pleural effusions suggests inhalational anthrax. A chest CT scan is also helpful in this disease, revealing hyperdense hilar and mediastinal nodes, mediastinal edema, and infiltrates and pleural effusions. Thoracentesis may reveal hemorrhagic pleural fluid.

A variety of rapid-assay kits are available to detect *B. anthracis* spores on environmental surfaces. However, none of these kits has been independently evaluated or endorsed by the CDC or the Food and Drug Administration, and many false positive results were reported during the United States bioterrorist attack.[56]

Enzyme-linked immunosorbent assays are available that measure antibody titers against lethal and edema toxins. A fourfold rise in titers over 4 weeks or a single titer of 1:32 is considered positive. This assay is not helpful during the acute illness. In patients receiving antibiotic prophylaxis, the antibody response may be blunted.

TREATMENT AND PROGNOSIS

Antibiotic therapy should be immediately initiated in all patients deemed at high risk who have fever or a systemic illness

consistent with inhalational anthrax. Any delay in therapy increases the risk of a fatal outcome. First-line antibiotics are intravenous ciprofloxacin or doxycycline combined with one or two other antibiotics with activity against the pathogen [*see Table 1*]. Because anthrax strains may have constitutive and inducible β-lactamases, monotherapy with penicillin or ampicillin is not recommended. When meningitis is suspected, doxycycline should not be used, because of its poor CNS penetration. Once the patient is stable, oral antibiotics can be given, with ciprofloxacin (500 mg twice daily) or doxycycline (100 mg twice daily) being the treatment of choice. Because of the risk of delayed germination of spores in the host, therapy should be continued for 60 days. Oral ciprofloxacin or doxycycline for 60 days is recommended for cutaneous disease. Excision of skin lesions is contraindicated because of the increased risk of precipitating bacteremia. However, after appropriate antibiotic therapy, excision and skin grafting may be necessary.[66]

Before antibiotics became available, cutaneous disease was associated with a mortality of 10% to 20%. With appropriate antibiotic treatment, fewer than 1% of patients die. Despite appropriate antibiotics and respiratory support, inhalational anthrax in the past was almost always fatal. However, experiences in Russia and the United States demonstrated that with early systemic antibiotic therapy, mortality can be reduced to approximately 50%. Gastrointestinal disease also is associated with a high mortality (25% to 100%).

PREVENTION

Postexposure prophylaxis is critical for the prevention of secondary cases. The selection of patients for prophylaxis depends on the environmental setting and the conditions of the spore release. Nasal swab cultures are insensitive and should not be used to determine whether an individual should receive prophylactic antibiotics. Nasal swab cultures are recommended only as an epidemiologic tool to determine the extent of exposure in a specific area or building. In cases of suspected exposure to *B. anthracis*, the preventive regimens of choice are oral fluoroquinolones (e.g., ciprofloxacin, 500 mg twice daily; levofloxacin, 500 mg a day; or ofloxacin, 400 mg twice daily) or, if fluoroquinolones are contraindicated, doxycycline (100 mg twice daily). Prophylaxis should be continued until exposure is excluded. If exposure is confirmed, prophylaxis should be continued for 60 days. In persons thought to be heavily exposed, prophylaxis for 100 days may be considered.[67]

A killed vaccine derived from a component of the anthrax exotoxin is available and is recommended for all industrial workers at risk for exposure to contaminated animal products. As a result of increased concerns about biologic warfare and bioterrorism, military personnel are now vaccinated. Postexposure vaccination, although not approved by the FDA, has been shown to provide additional protection in animal studies and is recommended as an option by the CDC. In such cases, the vaccine is considered an investigational new drug and should be administered with informed consent.[68]

Anecdotal reports in the lay press have associated anthrax vaccine with a high incidence of serious reactions. Nevertheless, although mild localized reactions occur in up to 30% of anthrax vaccine recipients, to date, with the exception of two cases of optic neuritis,[69] and one of delayed anaphylaxis,[70] serious adverse reactions have not been reported.[71]

For decontamination after anthrax exposure, exposed skin should be washed extensively with soap and water. Because of the potential danger of secondary aerosolization, decontamination of exposed environments is recommended. Decontamination is technically difficult and requires expert analysis that takes into account the nature of the exposure and the environmental conditions.[56]

Infections and Disorders Due to Other *Bacillus* Species

Bacillus organisms are found in soil, dust, decaying organic matter, and water. Some species are part of the normal flora, particularly in patients who have had prolonged hospitalizations. Despite their widespread distribution, these organisms rarely cause infection and are more often isolated as a contaminant. Risk factors associated with serious *Bacillus* infections include use of intravascular catheters; intravenous drug abuse; sickle cell disease; and immunosuppression caused by malignancy, neutropenia, corticosteroid therapy, or AIDS.[72] Because these organisms are frequently resistant to third-generation cephalosporins, prolonged antibiotic treatment regimens that include these agents may select out for *Bacillus* organisms.

B. cereus is the most frequent *Bacillus* species to cause invasive infection, followed by *B. subtilis* [*see Table 4*].

DIAGNOSIS

Clinical Manifestations

With the exception of *B. cereus* eye infections, *Bacillus* infections are rare and cannot be clinically distinguished from those caused by other pyogenic organisms. Pneumonia can develop in the compromised host, and necrotizing pneumonia caused by *B. cereus* has been reported in patients with acute leukemia and hepatic malignancy. Fatal pseudomembranous tracheobronchitis and pneumonia have been reported in a patient with aplastic anemia.[73]

Bacterial endocarditis caused by *Bacillus* species is a well-recognized complication of intravenous drug abuse. The tricuspid valve is most often involved, and the course of illness tends to be indolent. *Bacillus* species are among many pathogens that infect prosthetic valves.[74] Indwelling catheters may become colonized with *Bacillus* organisms, resulting in positive blood culture results.

High-grade bacteremia can be complicated by fatal meningoencephalitis in the immunocompromised host.[75] Necrotizing fasciitis can be caused by *B. cereus*.

In rare cases, meningitis may result from *Bacillus* bacteremia, but more often, it has followed inadvertent inoculation of organ-

Table 4 Clinical Syndromes Caused by *Bacillus* Species

Species	*Clinical Syndrome*
B. alvei	Meningitis, pneumonia, empyema, bacteremia
B. cereus	Ophthalmitis, bacteremia, pneumonia, osteomyelitis, endocarditis, soft tissue infection
B. sphaericus	Peritonitis, pleuritis, lung infection, meningitis, bacteremia
B. subtilis	Meningitis, otitis, mastoiditis, urinary tract infection, bacteremia, endocarditis, ventricular shunt infection

isms into the CSF during spinal anesthesia.[76] Because of their hardy growth characteristics, *Bacillus* species are common laboratory contaminants. Pseudobacteremia caused by contamination of alcohol swabs and rubber stoppers on blood culture bottles and pseudomeningitis caused by *Bacillus* contamination of commercial culture media have been reported.[77]

Motor vehicle accidents in which injuries result from direct contact with the road may result in *B. cereus* soft tissue and bone infections that necessitate extensive surgical debridement and amputation for cure. Close-range gunshot wounds and injection of contaminated heroin have also been complicated by *B. cereus* soft tissue infection.[78,79]

Ocular infections *B. cereus* is a primary pathogen for ocular infections. This species is one of the most common agents associated with posttraumatic endophthalmitis. When *B. cereus* is the causative agent, an intraocular foreign body is often present. Injuries caused by metal fragments and injuries associated with contamination by soil and dust increase the risk of infection by *B. cereus*. The onset of infection is rapid and leads to destruction of the vitreous and retinal tissue, causing loss of vision 12 to 48 hours after inoculation. Patients frequently become systemically ill, with fever and leukocytosis. Endophthalmitis and panophthalmitis can also be related to intravenous drug abuse. Early diagnosis and treatment are critical for preventing permanent structural changes and blindness.[80]

Food poisoning *B. cereus* produces two enterotoxins, diarrheal toxin and emetic toxin, that are responsible for two food-poisoning syndromes. The emetic form is associated primarily with the ingestion of contaminated fried rice.[81] From 1 to 6 hours after ingestion, the person experiences vomiting and symptoms identical to those of staphylococcal food poisoning. A case of fulminant fatal hepatic failure and rhabdomyolysis associated with ingestion of *B. cereus* emetic toxin has also been reported.[82] The diarrheal form has a longer incubation period (10 to 12 hours), and manifestations are related to lower rather than upper gastrointestinal symptoms. Symptoms include abdominal pain, profuse watery diarrhea, tenesmus, and nausea. This syndrome is usually self-limited and lasts 12 to 24 hours. Outbreaks are generally associated with ingestion of contaminated meat, vegetables, and mayonnaise.[83]

TREATMENT

B. cereus is resistant to penicillin and other β-lactam antibiotics, including cephalosporins; it is sensitive to vancomycin, gentamicin, imipenem, and ciprofloxacin. Clindamycin, erythromycin, chloramphenicol, and tetracycline have also been shown to be active. Other *Bacillus* species are susceptible to penicillins and cephalosporins. Pending speciation and susceptibility testing, vancomycin or clindamycin with or without gentamicin is the empirical treatment of choice when infection with *Bacillus* species is suspected [*see Table 1*].

If *Bacillus* bacteremia develops in immunocompromised patients with long-term indwelling catheters, antibiotic therapy must be instituted and the catheter removed to prevent recurrence of the infection.[84] In addition to antibiotic treatment, the deep-seated soft tissue infection associated with necrotizing fasciitis requires aggressive surgical debridement.[85]

For intraocular infections, systemic therapy is usually supplemented with intravitreous clindamycin and gentamicin administered by an ophthalmologist. Intravitreous dexamethasone and early vitrectomy have been recommended for the management of sight-threatening endophthalmitis by *B. cereus*.[80]

For patients with food poisoning, antibiotics are not required, because the disease is self-limited. Supportive care may include intravenous fluids if the patient becomes severely dehydrated.

Infections Due to the *Erysipelothrix* Organism

Human infections by the *Erysipelothrix* organism are rare and are always caused by *E. rhusiopathiae* (formerly *E. insidiosa*). This aerobic gram-positive bacillus grows on routine nonselective media and is nonhemolytic, nonmotile, and catalase negative. Because of the cell wall's high lipid content (about 30%), *E. rhusiopathiae* is resistant to desiccation and may tolerate salting, pickling, and smoking. The organism is capable of growing over a broad temperature range (4° to 42° C) and is widespread in nature. *E. rhusiopathiae* can infect mammals, birds, fish, shellfish, and insects. Human infection results from handling dead infected animal parts. Infection can also result from cat bites.[86] Most cases of skin infection occur during the summer and early fall. Infection is almost always the result of occupational exposure of slaughterhouse workers, butchers, fishers, farmers, and veterinarians. The organism is usually traumatically inoculated. Bacilli remain extracellular and are often located deep in the skin near capillaries, where the organism can gain entry into the bloodstream.[87]

DIAGNOSIS

Clinical Manifestations

Erysipeloid Between 2 and 7 days after skin inoculation, a purplish-red, nonvesicular area arises with a sharply defined, raised, serpiginous border. Lesions most often develop on the face and hands. Proximal regions of the hand and fingers are involved, whereas the distal phalanges are usually spared. Lesions spread peripherally at a slow pace (1 cm/day). Over time, the central part of the lesion begins to heal, resulting in a pale center surrounded by a fiery-red outer border. Lesions itch or burn and are rarely associated with lymphangitis or lymphadenitis. Fever and systemic complaints are rare.

Endocarditis Endocarditis is rare but may occur on both deformed and normal heart valves. The onset is acute or subacute and most often involves the aortic valve. In approximately 40% of cases, a skin lesion is noted just before or at the time of diagnosis, suggesting the possibility of *E. rhusiopathiae*.[88] Often, the skin lesion has healed by the time endocarditis becomes apparent. Mental-status changes associated with multiple cerebral hemorrhages may also accompany this form of endocarditis.[89]

Patient History and Laboratory Tests

An appropriate epidemiologic history will suggest the diagnosis. Morphologically, the skin lesions resemble erysipelas (caused by group A streptococci). However, the rate of progression of *Erysipelothrix* infection is considerably slower, and unlike erysipelas, this skin lesion is not associated with tenderness or lymphadenopathy. Injection and culturing of nonbacteriostatic normal saline is rarely successful. Biopsy of a full-thickness skin specimen from the advancing border of the lesion and culturing in glucose-containing broth result in the highest yields. PCR assays for erysipelas have proved to be useful in swine and have been applied to humans.[90] Diagnosis of endocarditis depends on isolation of the organism from blood cultures.

TREATMENT

The treatment of choice for erysipeloid is penicillin; a single injection of 600,000 units of penicillin G benzathine generally is curative. For penicillin-allergic patients, oral erythromycin (250 to 500 mg every 6 hours) is effective. This skin infection is usually self-limited and lasts about 3 weeks; however, antibiotic treatment hastens the healing. Bacterial endocarditis is best treated with intravenous penicillin G. For penicillin-allergic patients, intravenous cefazolin or ceftriaxone is recommended [*see Table 1*].[90,91] Most strains are resistant to vancomycin, which therefore should not be used for empirical therapy for endocarditis if *E. rhusiopathiae* is a possibility. Despite appropriate therapy, the mortality associated with endocarditis is 30% to 40%.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Krech T, Hollis DG: *Corynebacterium* and related organisms. Manual of Clinical Microbiology, 5th ed. Balows A, Hausler WJ, Herrmann KL, et al, Eds. American Society for Microbiology, Washington, DC, 1991, p 280

2. Skogen V, Cherkasova VV, Maksimova N, et al: Molecular characterization of *Corynebacterium diphtheriae* isolates, Russia, 1957–1987. Emerg Infect Dis 8:516, 2002

3. Bisgard KM, Hardy IR, Popovic T, et al: Respiratory diphtheria in the United States, 1980 through 1995. Am J Public Health 88:787, 1998

4. Koopman JS, Campbell J: The role of cutaneous diphtheria infections in a diphtheria epidemic. J Infect Dis 131:239, 1975

5. Golaz A, Lance-Parker S, Welty T, et al: Epidemiology of diphtheria in South Dakota. S D J Med 53:281, 2000

6. Control of epidemic diphtheria in the newly independent states of the former Soviet Union, 1990-1998. J Infect Dis 181(suppl 1):S1, 2000

7. McQuillan GM, Kruszon-Moran D, Deforest A, et al: Serologic immunity to diphtheria and tetanus in the United States. Ann Intern Med 136:660, 2002

8. Holmes RK: Biology and molecular epidemiology of diphtheria toxin and the *tox* gene. J Infect Dis 181(suppl 1):S156, 2000

9. Pancharoen C, Mekmullica J, Thisyakorn U, et al: Clinical features of diphtheria in Thai children: a historic perspective. Southeast Asian J Trop Med Public Health 33:352, 2002

10. Kadirova R, Kartoglu HU, Strebel PM: Clinical characteristics and management of 676 hospitalized diphtheria cases, Kyrgyz Republic, 1995. J Infect Dis 181(suppl 1):S110, 2000

11. Mofred A, Guérin JM, Falfoul-Borsali N, et al: Cutaneous diphtheria. Rev Med Interne 15:515, 1994

12. Loukoushkina EF, Boko PV, Kolbasova EV, et al: The clinical picture of diphtheritic carditis in children. Eur J Pediatr 157:528, 1998

13. Havaldar PV, Sankpal MN, Doddannavar RP: Diphtheritic myocarditis: clinical and laboratory parameters of prognosis and fatal outcome. Ann Trop Paediatr 20:209, 2000

14. Logina I, Donaghy M: Diphtheritic polyneuropathy: a clinical study and comparison with Guillain-Barré syndrome. J Neurol Neurosurg Psychiatry 67:433, 1999

15. Kneen R, Pham NG, Solomon T, et al: Penicillin vs. erythromycin in the treatment of diphtheria. Clin Infect Dis 27:845, 1998

16. Sutter RW, Hardy IR, Kozlova IA, et al: Immunogenicity of tetanus-diphtheria toxoids (Td) among Ukrainian adults: implications for diphtheria control in the newly independent states of the former Soviet Union. J Infect Dis 181(suppl 1):S197, 2000

17. Coyle MB, Lipsky BA: Coryneform bacteria in infectious diseases: clinical and laboratory aspects. Clin Microbiol Rev 3:227, 1990

18. Giannakopoulos S, Alivizatos G, Deliveliotis C, et al.: Encrusted cystitis and pyelitis. Eur Urol 39:446, 2001

19. Zinner SH: Changing epidemiology of infections in patients with neutropenia and cancer: emphasis on gram-positive and resistant bacteria. Clin Infect Dis 29:490, 1999

20. Wang CC, Mattson D, Wald A: *Corynebacterium jeikeium* bacteremia in bone marrow transplant patients with Hickman catheters. Bone Marrow Transplant 27:445, 2001

21. Traub WH, Geipel U, Leonhard B, et al: Antibiotic susceptibility testing (agar disk diffusion and agar dilution) of clinical isolates of *Corynebacterium jeikeium*. Chemotherapy 44:230, 1998

22. Capdevila JA, Bujan S, Gavalda J, et al: *Rhodococcus equi* pneumonia in patients infected with the human immunodeficiency virus: report of two cases and review of the literature. Scand J Infect Dis 29:535, 1997

23. Munoz P, Burillo A, Palomo J, et al: *Rhodococcus equi* infection in transplant recipients: case report and review of the literature. Transplantation 65:449, 1998

24. Verville TD, Huycke MM, Greenfield RA, et al: *Rhodococcus equi* infections of humans: 12 cases and a review of the literature. Medicine (Baltimore) 73:119, 1994

25. Weinstock DM, Brown AE: *Rhodococcus equi*: an emerging pathogen. Clin Infect Dis 34:1379, 2002

26. Hsueh PR, Hung CC, Teng LJ, et al: Report of invasive *Rhodococcus equi* infections in Taiwan, with an emphasis on the emergence of multidrug-resistant strains. Clin Infect Dis 27:370, 1998

27. Schlech III WF: Foodborne listeriosis. Clin Infect Dis 31:770, 2000

28. Crum NF: Update on *Listeria monocytogenes* infection. Curr Gastroenterol Rep 4:287, 2002

29. Hussein AS, Shafran SD: Acute bacterial meningitis in adults: a 12-year review. Medicine (Baltimore) 79:360, 2000

30. Southwick FS, Purich DL: Intracellular pathogenesis of listeriosis. N Engl J Med 334:770, 1996

31. Braun L, Cossart P, Interactions between *Listeria monocytogenes* and host mammalian cells. Microbes Infect 2:803, 2000

32. Chakraborty T: Molecular and cell biological aspects of infection by *Listeria monocytogenes*. Immunobiology 201:155, 1999

33. Anaissie E, Kontoyiannis DP, Kantarjian H, et al: Listeriosis in patients with chronic lymphocytic leukemia who were treated with fludarabine and prednisone. Ann Intern Med 117:466, 1992

34. Juardo RL, Farley MM, Pereira E, et al: Increased risk of meningitis and bacteremia due to *Listeria monocytogenes* in patients with human immunodeficiency virus infection. Clin Infect Dis 17:224, 1993

35. Mylonakis E, Paliou M, Hohmann EL, et al: Listeriosis during pregnancy: a case series and review of 222 cases. Medicine (Baltimore) 81:260, 2002

36. Nolla-Salas J, Bosch J, Gasser I, et al: Perinatal listeriosis: a population-based multicenter study in Barcelona, Spain (1990–1996). Am J Perinatol 15:461, 1998

37. Mylonakis E, Hohmann EL, Calderwood SB: Central nervous system infection with *Listeria monocytogenes*: 33 years' experience at a general hospital and review of 776 episodes from the literature. Medicine (Baltimore) 77:313, 1998

38. Aouaj Y, Spanjaard L, van Leeuwen N, et al: *Listeria monocytogenes* meningitis: serotype distribution and patient characteristics in The Netherlands, 1976-95. Epidemiol Infect 128:405, 2002

39. Armstrong RW, Fung PC: Brainstem encephalitis (rhombencephalitis) due to *Listeria monocytogenes*: case report and review. Clin Infect Dis 16:689, 1993

40. Lohmann CP, Gabel VP, Heep M, et al: *Listeria monocytogenes*-induced endogenous endophthalmitis in an otherwise healthy individual: rapid PCR-diagnosis as the basis for effective treatment. Eur J Ophthalmol 9:53, 1999

41. Vargas V, Aleman C, de Torres I, et al: *Listeria monocytogenes*-associated acute hepatitis in a liver transplant recipient. Liver 18:213, 1998

42. Frye DM, Zweig R, Sturgeon J, et al: An outbreak of febrile gastroenteritis associated with delicatessen meat contaminated with *Listeria monocytogenes*.Clin Infect Dis 35:943, 2002

43. Temple ME, Nahata M C: Treatment of listeriosis. Ann Pharmacother 34:656, 2000

44. Hof H, Nichterlein T, Kretschmar M: Management of listeriosis. Clin Microbiol Rev 10:345, 1997

45. Nichterlein T, Bornitz F, Kretschmar M, et al: Successful treatment of murine listeriosis and salmonellosis with levofloxacin. J Chemother 10:313, 1998

46. Torres OH, Domingo P, Pericas R, et al: Infection caused by *Nocardia farcinica*: case report and review. Eur J Clin Microbiol Infect Dis 19:205, 2000

47. Torres HA, Reddy BT, Raad II, et al: Nocardiosis in cancer patients. Medicine (Baltimore) 81:388, 2002

48. Roberts SA, Franklin JC, Mijch A, et al: *Nocardia* infection in heart-lung transplant recipients at Alfred Hospital, Melbourne, Australia, 1989–1998. Clin Infect Dis 31:968, 2000

49. Marquez-Diaz F, Soto-Ramirez LE, Sifuentes-Osornio J: Nocardiasis in patients with HIV infection. AIDS Patient Care STDS 12:825, 1998

50. Lee GY, Daniel RT, Brophy BP, et al: Surgical treatment of nocardial brain abscesses. Neurosurgery 51:668, 2002

51. LeBlang SD, Whiteman MLH, Post MJD, et al: CNS *Nocardia* in AIDS patients: CT and MRI with pathologic correlation. J Comput Assist Tomogr 19:15, 1995

52. Angelika J, Hans-Jurgen G, Uwe-Frithjof H: Primary cutaneous nocardiosis in a husband and wife. J Am Acad Dermatol 41:338, 1999

53. Arduino RC, Johnson PC, Miranda AG: Nocardiosis in renal transplant recipients undergoing immunosuppression with cyclosporine. Clin Infect Dis 16:505, 1993

54. Kontoyiannis DP, Ruoff K, Hooper DC: *Nocardia* bacteremia: report of four cases and review of the literature. Medicine (Baltimore) 77:255, 1998

55. Update: investigation of bioterrorism-related inhalational anthrax—Connecticut, 2001. MMWR Morb Mortal Wkly Rep 50:1049, 2001

56. Inglesby TV, O'Toole T, Henderson DA, et al: Anthrax as a biological weapon, 2002: updated recommendations for management. JAMA 287:2236, 2002

57. Jackson PJ, Hugh-Jones ME, Adair DM, et al: PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: the presence of multiple *Bacillus anthracis* strains in different victims. Proc Natl Acad Sci USA 95:1224, 1998

58. Meselson M, Guillemin J, Hugh-Jones M, et al: The Sverdlovsk anthrax outbreak of 1979. Science 266:1202, 1994

59. Henderson DW, Peacock S, Beltons FC: Observations on the propyhylaxis of experimental pulmonary anthrax in the monkey.J Hyg 54:28, 1956

60. Little SF, Ivins BE: Molecular pathogenesis of *Bacillus anthracis* infection. Microbes Infect 1:131, 1999

61. Mourez M, Lacy DB, Cunningham K, et al: 2001: a year of major advances in anthrax toxin research. Trends Microbiol 10:287, 2002

62. Smego RA Jr, Gebrian B, Desmangels G: Cutaneous manifestations of anthrax in rural Haiti. Clin Infect Dis 26:97, 1998

63. Bush LM, Abrams BH, Beall A, et al: Index case of fatal inhalational anthrax due to bioterrorism in the United States. N Engl J Med 345:1607, 2001

64. Shafazand S, Doyle R, Ruoss S, et al: Inhalational anthrax: epidemiology, diagnosis, and management. Chest 116:1369, 1999

65. Jernigan JA, Stephens DS, Ashford DA, et al: Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis 7:933, 2001

66. Aslan G, Terzioglu A: Surgical management of cutaneous anthrax. Ann Plast Surg 41:468, 1998

67. Bell DM, Kozarsky PE, Stephens DS: Clinical issues in the prophylaxis, diagnosis, and treatment of anthrax. Emerg Infect Dis 8:222, 2002

68. Notice to Readers: Additional options for preventive treatment for persons exposed to inhalational anthrax. MMWR Morb Mortal Wkly Rep 50(50):1142, 2001

69. Kerrison JB, Lounsbury D, Thirkill CE, et al: Optic neuritis after anthrax vaccination. Ophthalmology 109:99, 2002

70. Swanson-Biearman B, Krenzelok EP: Delayed life-threatening reaction to anthrax vaccine. J Toxicol Clin Toxicol 39:81, 2001

71. Lange JL, Lesikar SE, Rubertone MV, et al: Comprehensive systematic surveillance for adverse effects of Anthrax Vaccine Adsorbed, US Armed Forces, 1998–2000.Vaccine 21:1620, 2003

72. Sliman R, Rehm S, Shlaes DM: Serious infections caused by *Bacillus* species. Medicine (Baltimore) 66:218, 1987

73. Strauss R, Mueller A, Wehler M, et al: Pseudomembranous tracheobronchitis due to *Bacillus cereus*. Clin Infect Dis 33:E39, 2001

74. Castedo E, Castro A, Martin P, et al: *Bacillus cereus* prosthetic valve endocarditis. Ann Thorac Surg 68:2351, 1999

75. Marley EF, Saini NK, Venkatraman C, et al: Fatal *Bacillus cereus* meningoencephalitis in an adult with acute myelogenous leukemia. South Med J 88:969, 1995

76. Barrie D, Wilson JA, Hoffman PN, et al: *Bacillus cereus* meningitis in two neurosurgical patients: an investigation into the source of the organism. J Infect 25:291, 1992

77. Loeb M, Wilcox L, Thornley D, et al: *Bacillus* species pseudobacteremia following hospital construction. Can J Infect Control 10:37, 1995

78. Krause A, Freeman R, Sisson PR, et al: Infection with *Bacillus cereus* after close-range gunshot injuries. J Trauma 41:546, 1996

79. Dancer SJ, McNair D, Finn P, et al: *Bacillus cereus* cellulitis from contaminated heroin. J Med Microbiol 51:278, 2002

80. Hemady R, Zaltas M, Paton B, et al: *Bacillus*-induced endophthalmitis: new series of 10 cases and review of the literature. Br J Ophthalmol 74:26, 1990

81. *Bacillus cereus* food poisoning associated with fried rice at two child day care centers: Virginia, 1993: Centers for Disease Control and Prevention. JAMA 271:1074, 1994

82. Mahler H, Pasi A, Kramer JM, et al: Fulminant liver failure in association with the emetic toxin of *Bacillus cereus* (comments). N Engl J Med 336:1142, 1997

83. Gaulin C, Viger YB, Fillion L: An outbreak of *Bacillus cereus* implicating a part-time banquet caterer. Can J Public Health 93:353, 2002

84. Cotton DJ, Gill VJ, Marshall DJ, et al: Clinical features and therapeutic interventions in 17 cases of *Bacillus* bacteremia in an immunosuppressed patient population. J Clin Microbiol 25:672, 1987

85. Meredith FT, Fowler VG, Gautier M, et al: *Bacillus cereus* necrotizing cellulitis mimicking clostridial myonecrosis: case report and review of the literature. Scand J Infect Dis 29:528, 1997

86. Talan DA, Citron DM, Abrahamian FM, et al: Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85, 1999

87. Brooke CJ, Riley TV: *Erysipelothrix rhusiopathiae*: bacteriology, epidemiology and clinical manifestations of an occupational pathogen. J Med Microbiol 48:789, 1999

88. Gorby GL, Peacock JE Jr: *Erysipelothrix rhusiopathiae* endocarditis: microbiologic, epidemiologic, and clinical features of an occupational disease. Rev Infect Dis 10:317, 1988

89. Ko SB, Kim DE, Kwon HM, et al: A case of multiple brain infarctions associated with *Erysipelothrix rhusiopathiae* endocarditis. Arch Neurol 60:434, 2003

90. Venditti M, Gelfusa V, Tarasi A, et al: Antimicrobial susceptibilities of *Erysipelothrix rhusiopathiae*. Antimicrob Agents Chemother 34:2038, 1990

91. Fidalgo SG, Longbottom CJ, Rjley TV: Susceptibility of *Erysipelothrix rhusiopathiae* to antimicrobial agents and home disinfectants. Pathology 34:462, 2002

Acknowledgments

Figure 1 R. J. Collier, M.D., Harvard Medical School. Adapted by Dimitry Schidlovsky.

Figure 3 Dimitry Schidlovsky.

Figure 4 Courtesy of Professor Jean Hilarie Saurat, Geneva University Hospital, Geneva, Switzerland.

V ANAEROBIC INFECTIONS

Anthony W. Chow, m.d., f.a.c.p, f.r.c.p.c.

General Considerations

anaerobic bacteria and oxygen tolerance

Obligate anaerobes require reduced oxygen tension (< 10%) for growth; they do not survive as surface growths on solid medium in ambient air (i.e., 20% oxygen). In contrast, facultative bacteria can grow in air with or without 10% carbon dioxide, whereas microaerophilic or capnophilic bacteria grow poorly in ambient air, but they grow better in air with reduced oxygen tension (i.e., 10% oxygen and 10% carbon dioxide). Obligate anaerobes vary greatly in their sensitivity to oxygen. Extremely oxygen-sensitive anaerobes, such as spirochetes and some *Clostridium* species, cannot tolerate even 0.5% oxygen. As a general rule, clinical isolates commonly recovered from anaerobic infections are relatively aerotolerant, able to survive in 2% to 8% oxygen.

obligate anaerobes in the normal flora

Quantitatively, obligate anaerobes are the predominant normal flora on mucocutaneous surfaces—especially the oral cavity, skin, and gastrointestinal and genital tracts—where they outnumber facultative bacteria by a factor of 10 to 10^3 [*see Table 1*]. The major genera of obligate anaerobes and their distribution in the normal flora vary according to body site [*see Table 2*]. These indigenous bacteria have unique ecologic niches at different body sites. For example, in the oral cavity, *Actinomyces viscosus* (along with the aerobes *Streptococcus sanguis, S. mutans,* and *S. mitis*) preferentially colonizes the tooth surface; in contrast, *Veillonella parvula* and *S. salivarius* have a predilection for the tongue and buccal mucosa.[1] *Bacteroides vulgatus, B. thetaiotaomicron, B. fragilis*, and *B. distasonis* are primarily indigenous in the colon[2]; *Prevotella bivia* and *P. disiens* are primarily resident in the female genital tract.[3] In addition, this ecosystem is readily influenced by a variety of physiologic and other host factors, such as age, pregnancy, menses, diet, underlying disease, hospitalization, and antimicrobial therapy [*see Table 3*].

pathogenesis of anaerobic infections

Host Conditions

Anaerobic infections characteristically are polymicrobial (mixed) and include both anaerobic and facultative organisms. The organisms tend to be acquired endogenously. The particular mix of pathogens reflects the combined influence of the complex commensal flora at a specific body site and the unique microbiota of the underlying conditions. Because these organisms are generally of low pathogenicity, anaerobic or mixed infections generally develop as a consequence of either structural alterations in the normal mucosal barrier or tissue ischemia with lowered oxidation-reduction potential. Knowledge of the anatomic location of the primary source of infection and the underlying condition of the host, therefore, is essential in predicting the probable organisms causing anaerobic and mixed infections associated with the indigenous microflora.

Table 1 The Predominant Normal Flora at Various Body Sites

Site	Anaerobes		Aerobes	
	Concentration	*Predominant Genera*	*Concentration*	*Predominant Genera*
Skin	10^4–10^5/cm^2	*Propionibacterium*	10^2–10^3/cm^2	*Staphylococcus* *Micrococcus* Diphtheroids
Oropharynx	10^6–10^{11}/ml	*Peptostreptococcus* *Veillonella* *Actinomyces* *Prevotella* *Porphyromonas* *Fusobacterium*	10^4–10^6/ml	*Streptococcus*
Stomach and upper small bowel	10^1–10^4/ml	*Peptostreptococcus* *Veillonella*	10^1–10^4/ml	*Streptococcus* *Lactobacillus*
Lower small bowel	10^4–10^7/ml	*Bacteroides* *Bifidobacterium*	10^4–10^7/ml	Coliforms
Colon	10^{11}–10^{12}/ml	*Bacteroides* *Bifidobacterium* *Eubacterium* *Clostridium* *Peptostreptococcus* *Veillonella*	10^8–10^9/ml	*Escherichia* *Enterococcus* *Lactobacillus*
Female genitalia (vagina and endocervix)	10^8–10^{10}/g	*Peptostreptococcus* *Lactobacillus* *Bacteroides* *Prevotella*	10^7–10^9/g	*Lactobacillus* *Streptococcus* *Staphylococcus*

Table 2 Classification of Anaerobic Bacteria and Their Distribution in Normal Flora and in Infection

Anaerobic Bacteria	*Genera*	*Normal Flora*				*Predominant Species from Anaerobic Infections*
		Skin	*Oropharynx*	*Intestine*	*Genitalia*	
Sporulating bacilli	*Clostridium*	0	±	2	±	*C. perfringens, C. difficile, C. ramosum, C. septicum, C. novyi*
Nonsporulating bacilli Gram negative	*Bacteroides*	0	2	2	1	*B. fragilis group (B. fragilis, B. vulgatus, B. thetaiotamicron, B. distasonis)*
	Prevotella	0	2	1	1	*P. melaninogenica, P. intermedia, P. bivia, P. disiens*
	Porphyromonas	0	2	1	1	*P. asaccharolytica, P. gingivalis*
	Fusobacterium	0	2	±	1	*F. nucleatum, F. necrophorum, F. varium, F. mortiferum*
Gram positive	*Actinomyces*	0	1	±	±	*A. israelii, A. naeslundii, A. viscosus*
	Bifidobacterium	0	±	2	±	*B. eriksonii, B. breve*
	Eubacterium	±	1	2	±	*E. lentum, E. limosum*
	Lactobacillus	0	±	1	2	*L. acidophilus, L. casei, L. plantarum*
	Propionibacterium and *Arachnia*	2	±	±	1	*P. acnes, P. granulosum, A. propionica*
Cocci						
Gram negative	*Veillonella*	0	2	1	1	*V. parvula*
Gram positive	*Peptostreptococcus*	1	2	2	1	*P. asaccharolyticus, P. prevotii, P. magnus, P. variabilis, P. anaerobius, P. micros, P. productus*
Spirochetes	*Treponema*	0	1	±	±	*T. vincentii, T. denticola*

0, absent or rare; ±, irregularly present; 1, usually present; 2, predominant

Microbial Synergy

Two thirds of anaerobic infections involve both anaerobes and facultative bacteria. The infectivity of obligate anaerobes in these instances is often facilitated by the coexistence of facultative organisms. Such examples of microbial synergy are particularly well demonstrated in periodontal infection and in various animal models of intra-abdominal and subcutaneous abscesses.[4,5] For example, microbial synergy is common between *Bacteroides* species and aerobic bacteria or anaerobic cocci and between most *Peptostreptococcus* species and *Pseudomonas aeruginosa* or *Staphylococcus aureus.* Anaerobes may require symbiotic facultative bacteria for providing necessary growth factors, lowering the oxidation-reduction potential of the environment, or impairing local host defenses. Conversely, the presence of obligate anaerobes may benefit coexisting facultative bacteria by growth enhancement,[6] protection from phagocytosis (e.g., succinic acid production by *Bacteroides* species),[7] or protection from β-lactam antibiotics (e.g., β-lactamase production).[8] Infective synergy between anaerobes and facultative bacteria is best demonstrated within tissues in which bacterial clearance is normally slow (e.g., subcutaneous abscesses or fibrin clot in intraperitoneal infection) or is hampered by underlying disease. An understanding of the dynamic interactions between different components of a complex flora in mixed infections has important therapeutic implications. Microorganisms in mixed infections may respond to antimicrobial agents differently than do those in monomicrobial infections, and it may not be necessary to eradicate every bacterial species in mixed infection to achieve a cure.

Virulence Factors

A number of microbial virulence factors are considered important in the pathogenesis of anaerobic infections [*see Table 4*].

Extracellular or membrane-bound enzymes Obligate anaerobes possess a number of extracellular or membrane-bound

Table 3 Effect of Host Conditions on the Indigenous Microflora at Various Body Sites

Site	*Host Condition*	*Change in Normal Flora*
Gingiva	Dental caries and periodontal disease	Increased motile anaerobic gram-negative bacilli and spirochetes
Oropharynx	Hospitalization, antibiotics, or serious illness	Increased facultative gram-negative bacilli
Upper small bowel	Achlorhydria, vagotomy, pyloroplasty	Increased *E. coli, B. fragilis,* and *Bifidobacterium*
Small bowel	Regional enteritis, decreased motility, or stasis secondary to blind loop, obstruction, diverticuli, irradiation, etc. Disrupted anatomic continuity after bowel resection or bypass surgery	Colonic flora
Large bowel	Colonic resection with ileostomy	Decreased anaerobes and some facultative bacteria
Vagina	Parturition, hysterectomy, or irradiation	Increased *E. coli* and *B. fragilis*

Table 4 Microbial Virulence Factors Important in Mixed Anaerobic Infections

Microbial Factor	*Pathogenic Effect*
Histolytic enzymes (e.g., collagenase, fibrinolysin, hemolysins, hyaluronidase, protease, lipase, ribonuclease, deoxyribonuclease; *Bacteroides fragilis, Porphyromonas gingivalis, Prevotella melaninogenica*)	Tissue destruction
Hemolysins (*Clostridium perfringens*)	Hemolysis
Enterotoxins (*C. perfringens, C. difficile*)	Alteration of cell function with diarrhea or cell death
Neurotoxins (*C. tetani, C. botulinum*)	Blockade of neuromuscular junctions with either spasticity (tetanus) or paralysis (botulism)
Endotoxin (gram-negative anaerobes; *B. fragilis, Fusobacterium nucleatum, P. intermedia, Veillonella parvula*)	Direct toxicity; Hageman factor, complement activation
Heparinase (*P. gingivalis, Bacteroides* species, *Fusobacterium* species)	Promotion of coagulation and tissue ischemia
Capsular polysaccharide (*B. fragilis, P. melaninogenica, Peptostreptococcus* species)	Inhibition of phagocytosis; abscess formation
IgA protease (*P. melaninogenica, P. gingivalis, P. intermedia*)	Impairment of secretory and mucosal immunity
Succinic acid (*B. fragilis*)	Inhibition of phagocytosis
β-Lactamase (*B. fragilis, P. melaninogenica, P. gingivalis, F. nucleatum*)	Antibiotic resistance
Oxygen-scavenging enzymes (e.g., superoxide dismutase, catalase, peroxidase; *P. gingivalis, B. fragilis, C. perfringens*)	Oxygen tolerance
Surface ligands and charge (*P. gingivalis, P. intermedia, F. nucleatum, Actinomyces naeslundii*)	Adherence and bacterial interaggregation
Bacteriocin and metabolites (e.g., fatty acids, H_2S, NH_3; *Propionibacterium acnes, B. fragilis, F. necrophorum*)	Inhibition of other normal flora

enzymes that promote tissue destruction. These include lipases, proteases, nucleases, and heparinases.[9] Membrane-bound enzymes, such as superoxide dismutase[10] and β-lactamases,[8] may be important for protecting virulent organisms from the toxic effects of oxygen and β-lactam antibiotics, respectively. Catalase may serve a function similar to that of superoxide dismutase. Organisms lacking these enzymes are susceptible to killing by toxic oxygen radicals and common antibiotics in the environment.

Capsular polysaccharide *B. fragilis, Prevotella melaninogenica,* and a number of *Peptostreptococcus* species are encapsulated and exhibit increased virulence in abscess formation and systemic invasion.[11] Interestingly, organisms that are normally nonencapsulated and unable to induce abscesses by themselves during experimental infection may become heavily encapsulated after participating in a mixed infection with other aerobic and anaerobic bacteria.[5] These heavily encapsulated strains are able to induce abscesses thereafter when inoculated alone. This phenomenon may help explain how nonpathogenic organisms that are part of the normal host flora can become pathogens. The capsular materials of *B. fragilis* and *P. melaninogenica* have been extracted and purified. These large-molecular-weight polysaccharides have been demonstrated to inhibit phagocytosis in vitro and to promote abscess formation in several animal models.[12] In addition, several oral anaerobes (e.g., *P. melaninogenica, P. gingivalis,* and *P. intermedia*) can secrete IgA proteases that may impair secretory and local mucosal immunity.[13]

Lipopolysaccharide Like their aerobic counterparts, anaerobic gram-negative bacteria possess lipopolysaccharides (LPS) in their outer cell membrane. However, the structure and biologic activity of LPS from several anaerobic bacteria are distinctly different from those of the classic LPS of Enterobacteriaceae. For example, LPS of *B. fragilis* and *P. intermedia* lack 2-keto-3-deoxyoctanoic acid and L-glycero-D-mannoheptose, and they have little endotoxic potency.[14] The LPS of *F. nucleatum* and *V. parvula,* on the other hand, have biochemical and biologic properties similar to those of classic endotoxin.

Fatty acids and other metabolites *B. fragilis* and other anaerobes produce various short-chain fatty acids that are deleterious to mammalian cell function. Infections with *B. fragilis* are associated with the production of high concentrations of succinic acid, which impairs the generation of the respiratory burst and profoundly reduces phagocytic killing and chemotactic responses of neutrophils.[15] This effect is most evident at the low pH and the low Eh (redox potential) conditions present in abscesses and mixed infections. Succinic acid production may represent an important virulence mechanism by *Bacteroides* species in the pathogenesis of synergistic mixed infections.

Toxins Several *Clostridium* species produce potent exotoxins. The most important of these is *C. perfringens* α-toxin, which is a lecithinase that exhibits hemolytic, necrotizing, and lethal properties.[16] α-Toxin disrupts membranes containing phospholipid-lecithin complexes, including human cell and mitochondrial membranes, and has direct myocardial depressant properties. A second clostridial exotoxin, β-toxin, is a potent cytotoxin that has cytolytic activity, particularly against endothelial cells. Toxin production at a site of injury allows rapid invasion and destruction of healthy tissues. The paucity of leukocytes in the exudate of clostridial myonecrosis may reflect the presence of these cytotoxins.

C. difficile, the major cause of antibiotic-associated diarrhea and enterocolitis, produces two high-molecular-weight enterotoxins, toxin A (enterotoxin) and toxin B (cytotoxin).[17] Both toxins inactivate Rho proteins, a family of small guanosine triphosphate–binding proteins that regulate actin cytoskeleton and various signal transduction processes. Toxin A causes intestinal fluid secretion, mucosal injury, and inflammation. Toxin B has no demonstrable effect on cell permeability and fluid secretion, but like toxin A, it disrupts tight junctions in human epithelial cell monolayers. Toxin B is 10 times more potent on a

Table 5 Anaerobic Bacteria Associated with Specific Infections[57-86]

Site	*Infections Likely to Involve Anaerobes*	*Anaerobes Recovered (Percentage of Infections)*	*Predominant Isolates*
Head and neck	Periapical abscess	90	Anaerobes: *Peptostreptococcus* species, *Prevotella melanogenica, P. intermedia, Porphyromonas gingivalis, P. asaccharolyticus, Actinomyces* species, *Fusobacterium nucleatum, F. necrophorum* Aerobes: *Streptococcus* species
	Periodontal infection	100	
	Fascial space infections	100	
	Peritonsillar abscess	84	
	Chronic sinusitis	53	
	Chronic otitis media	56	
Intracranial	Brain abscess (nontraumatic)	89	Anaerobes: *Peptostreptococcus* species, *Bacteroides fragilis, F. nucleatum, A. israelii* Aerobes: *Streptococcus milleri*, other *Streptococcus* species
	Subdural empyema	~50	
Pleuropulmonary	Aspiration pneumonia	87	Anaerobes: *Peptostreptococcus* species, *B. fragilis* group, *F. nucleatum*, pigmented and nonpigmented *Prevotella* species, *Clostridium* species, *Eubacterium* species, *Actinomyces* species, *Lactobacillus* species, *Veillonella parvula* Aerobes: *S. pneumoniae, Haemophilus influenzae, Staphylococcus aureus, Pseudomonas aeruginosa, Enterobacteriaceae* species
	Lung abscess	93	
	Empyema	76	
	Necrotizing pneumonia	94	
	Bronchiectasis	27	
	Hospital-acquired pneumonia	35	
Intra-abdominal	Intra-abdominal sepsis	83	Anaerobes: *B. fragilis*, other members of *B. fragilis* group, *Peptostreptococcus* species, *Clostridium* species, *Fusobacterium* species, *Biliophila wadsworthia* Aerobes: *Escherichia coli, P. aeruginosa, Klebsiella* species, *Enterococcus* species, *S. aureus*
	Hepatic abscess	25	
	Appendiceal abscess	92	
Female genital tract	Vulvovaginal abscess	75	Anaerobes: *Prevotella bivia, P. disiens, Bacteroides* species, *Peptostreptococcus asaccharolyticus, P. anaerobius, Actinomyces* species Aerobes: *Gardnerella vaginalis, E. coli*, group B streptococci, coagulase-negative staphylococci, *Neisseria gonorrhaeae, Chlamydia trachomatis*
	Salpingitis and pelvic peritonitis	25	
	Tubo-ovarian abscess	92	
	Posthysterectomy wound infections	67	
	Septic abortion and postpartum endometritis	73	
Skin, soft tissue, and bone	Crepitant cellulitis	75	Anaerobes: *Peptostreptococcus* species, *Bacteroides fragilis, P. melaninogenica, F. nucleatum, C. perfringens* Aerobes: *Proteus mirabilis, E. coli, Enterococcus* species, *S. aureus, P. aeruginosa, S. anginosus, S. aureus, Eikenella corrodens*
	Synergistic necrotizing cellulitis	89	
	Necrotizing fasciitis	47	
	Myonecrosis	100	
	Infected pressure ulcer	63	
	Pilonidal abscess	73	
	Perirectal abscess	77	
	Diabetic foot ulcer	63	
	Breast abscess	79	
	Bite wound infections	53	
Blood	Primary bacteremia	4	*B. fragilis* group, *F. necrophorum, Peptostreptococcus* species, *C. perfringens, C. septicum*

molar basis than toxin A in mediating damage to human colonic mucosa.[18]

Tetanus toxin (tetanospasmin) is produced in a tetanus-infected wound and is transported intra-axonally along motor nerves to the spinal cord. Here, the toxin alters normal control of the reflex arc by suppressing the inhibitory neurotransmitter γ-aminobutyric acid (GABA), producing severe muscle spasms. Like tetanospasmin, botulinum toxin also binds irreversibly to presynaptic nerve endings of cranial and peripheral nerves. Once bound, botulinum toxin prevents the release of the neurotransmitter acetylcholine, producing flaccid paralyses.

Specific Anaerobic Infections

Anaerobic bacteria can cause infections throughout the body. These infections can be conveniently divided into three categories on the basis of unique clinical, microbiologic, and epidemiologic features. These include infections caused by (1) *Bacteroides* and other mixed anaerobes, (2) *Actinomyces* species, and (3) *Clostridium* species.

INFECTIONS CAUSED BY *BACTEROIDES* AND OTHER MIXED ANAEROBES

Epidemiology

Anaerobic infections caused by *Bacteroides* and mixed anaerobes may involve any organ and may occur in persons of all ages. Obligate anaerobes are particularly prevalent in infections of the head and neck, lung and pleural space, intra-abdominal organs, the female genital tract, and necrotic skin and soft tissues. Predisposition to these infections is increased by local ischemia or tissue necrosis, such as from trauma, bites, surgical manipulation, irradiation, or neoplasm.

Etiology

The predominant obligate anaerobes commonly isolated from anaerobic or mixed infections at different anatomic sites include *Bacteroides, Prevotella, Porphyromonas, Fusobacterium, Peptostreptococcus, Actinomyces*, and *Clostridium* [*see Table 5*]. All members of the genus *Bacteroides* are thin, pleomorphic, gram-negative bacilli that are nonmotile and nonsporulating. The *B. fragilis* group dominates in the colonic microflora, whereas *Porphyromonas* and

Prevotella species reside primarily in the oropharynx. *B. fragilis* [*see Figure 1a*], the encapsulated member of the *B. fragilis* group, is by far the most common in anaerobic infections. Other members of this group include *B. ovatus*, *B. thetaiotaomicron*, *B. distasonis*, and *B. vulgatus.*

Prevotella melaninogenica (formerly *B. melaninogenicus*) is recognized in the laboratory by the production of a dark pigment. The most prevalent *Porphyromonas* species isolated from oral and periodontal infections are *P. gingivalis* and *P. asaccharolyticus*.

Fusobacterium species are gram-negative bacilli with typically pointed ends. They generally reside in the oropharynx or the gastrointestinal tract. Among these, *F. nucleatum* [*see Figure 1b*] is the most important species in head and neck and pleuropulmonary infections.

Peptostreptococcus species are small gram-positive cocci that are common in mixed infections at all body sites but particularly the oral cavity and the female genital tract. *Veillonella* species are anaerobic gram-negative cocci that are occasionally isolated from mixed infections, but their pathogenicity is uncertain.

Among the nonsporulating anaerobic gram-positive bacilli, *Propionibacterium acnes* is occasionally isolated in blood cultures but almost always as a skin contaminant. Another nonsporulating anaerobic gram-positive bacillus with low pathogenicity is *Lactobacillus,* a dominant member of the normal vaginal flora. *Lactobacillus* species occasionally can cause serious infections, including bacteremia.[19] Clinically important *Actinomyces* and *Clostridium* species are discussed separately (see below).

Clinical Presentations

Head and neck infections Orofacial anaerobic infections are commonly odontogenous in origin and include periapical abscesses, gingival and periodontal infections, and orofacial fascial space infections.[1] The clinical manifestations of these infections are largely dictated by the anatomic location and the extent and routes of spread. An aggressive form of gingivitis is Vincent angina, or trench mouth. This is a fulminant form of necrotizing, ulcerative gingivitis associated with severe pain, tissue destruction, foul breath, and a putrid discharge. Masticator-space infection generally originates from a molar or premolar tooth of the mandible and is characterized by pain and swelling at the angle of the jaw and by severe trismus. Extension of infection into the sublingual and submandibular spaces bilaterally may cause swelling of the base of the tongue and potential airway obstruction (Ludwig angina).[20] Extension of infection into the posterior compartment of the lateral pharyngeal space may be complicated by thrombophlebitis of the jugular vein (Lemierre syndrome).

Obligate anaerobes are also commonly present in chronic sinusitis, otitis media and mastoiditis, and tonsillar and peritonsillar abscesses. Acute sinusitis is seldom caused by anaerobic organisms unless it is associated with a dental infection. *Fusobacterium* and anaerobic gram-positive cocci are also commonly isolated in chronic or recurrent maxillary sinusitis. *Fusobacterium necrophorum* and *P. melaninogenica* are most frequently recovered from tonsillar and peritonsillar abscesses.

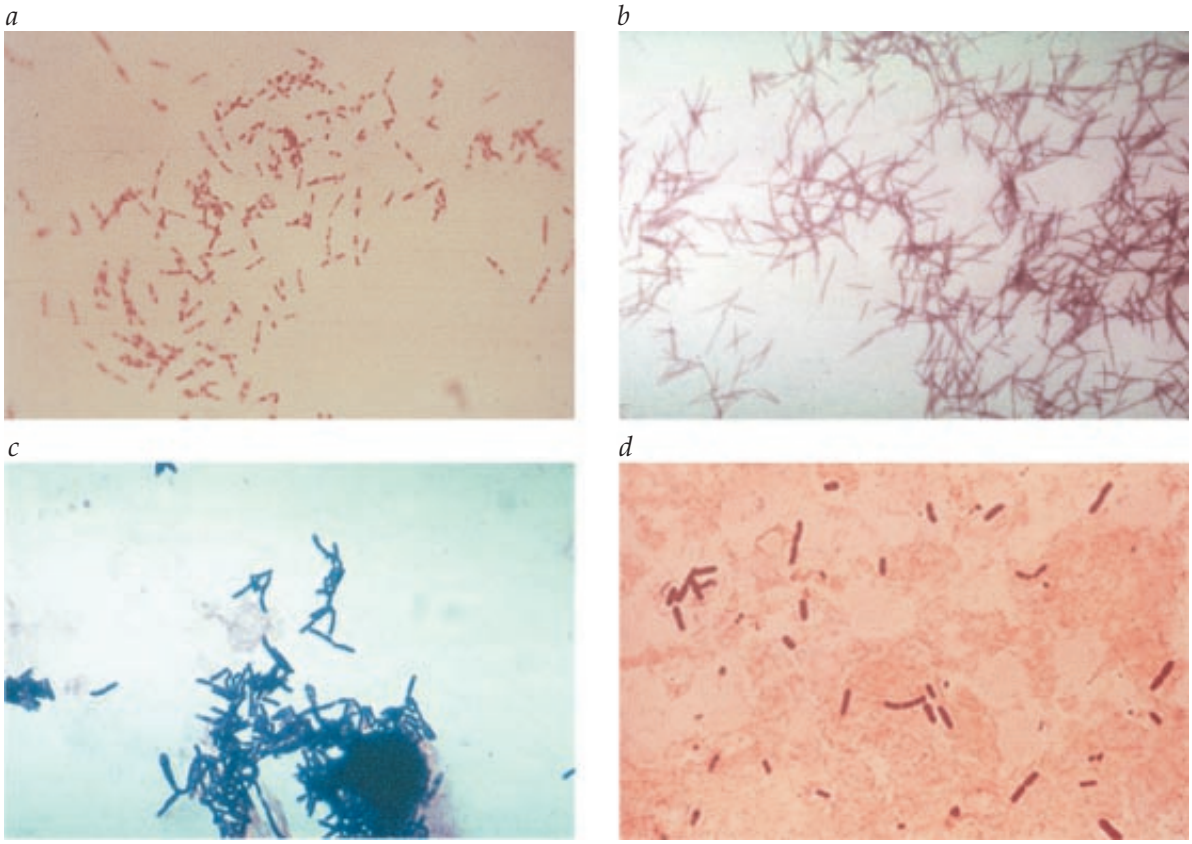


Figure 1 **Gram stains of representative obligate anaerobes commonly isolated in anaerobic infections (magnification: × 1,000). (*a*) *Bacteroides fragilis* from a pelvic abscess. (*b*) *Fusobacterium nucleatum* from an anaerobic pleuropulmonary infection. (*c*) *Clostridium perfringens* from myonecrosis. (*d*) *Actinomyces israelii* from cervicofacial actinomycosis.**

Intracranial infections Anaerobic bacteria are frequent pathogens in intracranial infections, particularly brain abscess caused by hematogenous dissemination from chronic and suppurative pulmonary foci; they are also frequently isolated in patients with cyanotic congenital heart disease. Contiguous spread from chronic otitis media, mastoiditis, or sinusitis may result in subdural empyema, epidural abscess, or suppurative thrombophlebitis of cortical vessels or venous sinuses. Purulent meningitis seldom involves anaerobic bacteria except in the newborn. Cerebral abscesses of sinus or dental origin are more probably caused by *S. milleri,* either alone or in mixed culture with other oropharyngeal aerobes and anaerobes. Otogenic cerebral abscesses, on the other hand, frequently involve *B. fragilis*, *Proteus* species, and streptococci.

Pleuropulmonary infections Anaerobic pleuropulmonary infections include aspiration pneumonitis, putrid lung abscess, necrotizing pneumonia, and empyema. Pneumonitis is usually the initial lesion; related symptoms in the early phases may be indistinguishable from symptoms of acute bacterial pneumonia of other causes. If the initial lesion remains untreated, however, pulmonary abscess may ensue after 8 to 14 days. Approximately one half of patients with lung abscess develop putrid-smelling expectorations. The subsequent clinical course depends largely on the nature of the underlying pulmonary pathologic condition. About 10% of patients with anaerobic infections of the lung parenchyma develop empyema. Necrotizing pneumonia is characterized by multiple small cavities within a pulmonary segment or lobe. The course is often fulminant, with rapid extension into adjacent segments. Anaerobic pleuropulmonary infections are typically polymicrobial. Predominant anaerobic isolates include *Peptostreptococcus* species, *F. nucleatum,* and the saccharolytic black-pigmented anaerobic gram-negative bacilli *(P. melaninogenica* and *P. intermedia)*. Aerobic and microaerophilic streptococci (e.g., *S. intermedius*) are also frequently isolated. Aerobes such as *S. aureus*, *Escherichia coli, Klebsiella pneumoniae,* and *P. aeruginosa* are more likely to be isolated along with anaerobic bacteria in hospital-acquired infections than in community-acquired aspiration pneumonia.

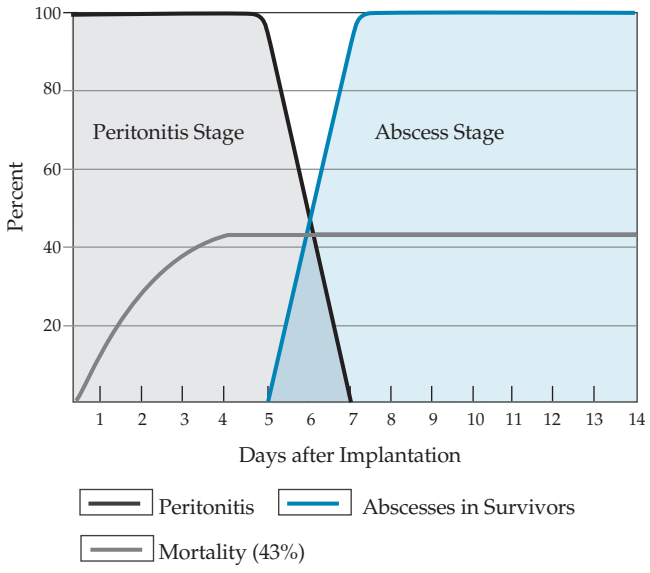


Figure 2 **Biphasic disease model of mixed infection in intra-abdominal sepsis.[25]**

Intra-abdominal infections Intra-abdominal sepsis most commonly results from bacterial contamination of intraperitoneal or retroperitoneal spaces after intestinal perforation. The initial event is peritonitis, either generalized or localized, with subsequent abscess formation. Common predisposing conditions include penetrating trauma, perforated appendicitis or diverticulitis, inflammatory bowel disease, intestinal malignancy with strangulation or obstruction, and anastomotic leak after intestinal surgery. Although a multiplicity of anaerobic and facultative bacteria may be isolated in intra-abdominal infections—particularly *B. fragilis, Peptostreptococcus* species, *Clostridium* species, Enterobacteriaceae, and *Enterococcus faecalis*—it is not always clear which components are the primary pathogens and which are merely symbionts or commensals. Animal studies of experimental peritonitis simulating intestinal perforation suggest that such infections follow a biphasic process[21] [*see Figure 2*]. Early peritonitis and bacteremia are related to aerobic coliform bacteria, whereas late abscesses are caused by anaerobes, often in synergy with facultative bacteria. The therapeutic implications of these studies are clear: both microbial components of intra-abdominal sepsis should receive appropriate antimicrobial attention.

Female genital tract infections Mixed aerobes and anaerobes are particularly important in closed-space infections such as vulvovaginal, adnexal, or tubo-ovarian abscesses and in postsurgical and postpartum infections. Other common gynecologic and obstetric infections involving mixed aerobic and anaerobic flora include acute and chronic salpingitis, infections associated with contraceptive intrauterine devices, postpartum or post–cesarean section wound infections, endometritis, and amnionitis. The most common anaerobes found include *Prevotella* species (especially *P. bivia, P. disiens,* and *P. melaninogenica*), *Peptostreptococcus* species, and *Actinomyces* species. The most common facultative pathogens are Enterobacteriaceae, especially *E. coli*, and aerobic or microaerophilic streptococci. *B. fragilis* is not a common organism in the normal vagina, but its prevalence is increased in posthysterectomy and post–cesarean section infections and in pelvic infections associated with malignancy and immunosuppressive therapy. Bacterial vaginosis also appears to be a polymicrobial infection involving both aerobes and anaerobes.[22] In addition to *Gardnerella vaginalis,* high concentrations of *Prevotella*, *Peptostreptococcus*, and *Mobiluncus* (a motile anaerobic curved gram-negative bacillus) can be regularly isolated from vaginal secretions.

Necrotic skin and soft tissue infections Necrotic wound and soft tissue infections are especially likely to develop in areas that are regularly exposed to fecal or oral contamination and that have been injured by trauma, ischemia, or surgery. Obligate anaerobes are particularly prevalent in infected pressure ulcers, diabetic foot infections, human bites, and infected pilonidal cysts. Clinical manifestations include crepitant cellulitis, synergistic necrotizing cellulitis or gangrene, myonecrosis, and necrotizing fasciitis. The range of bacterial isolates in such infections is enormous. Anaerobes, including *Bacteroides*, *Peptostreptococcus*, and *Clostridium* species, are almost universally present in mixed cultures.

Anaerobic bacteremia and endocarditis Polymicrobial bacteremia is particularly prevalent in infections of the gastrointestinal tract and the female genital tract, as well as infections of den-

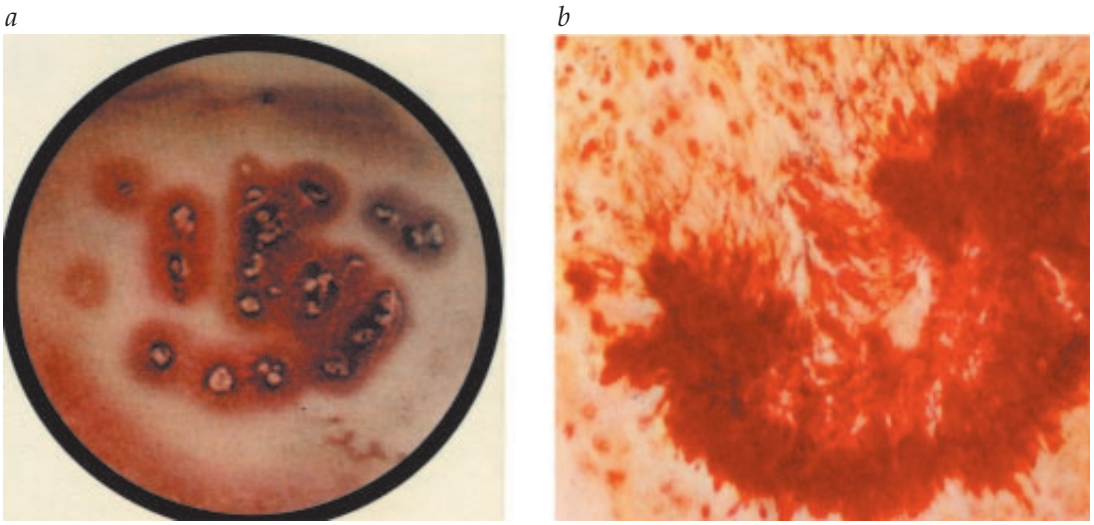


Figure 3 **Sulfur granules are a characteristic feature of *Actinomyces* infection. (*a*) Gross appearance in exudates. (*b*) Histologic appearance in infected tissue.**

tal origin. Anaerobic bacteremia originating from the female genital tract and odontogenous sources tends to be transient and self-limited. In contrast, anaerobic bacteremia originating from the gastrointestinal tract or from necrotic soft tissue tends to be recurrent and persistent in the absence of surgical drainage. The clinical manifestations of anaerobic bacteremia and the specific organisms involved depend to a large extent on the portal of entry and the nature of the underlying disease. For example, *B. fragilis* is most common in bacteremia of gastrointestinal and necrotic soft tissue origin.[23] Bacteroidaceae bacteremia of the female genital tract and of odontogenous origin rarely involves *B. fragilis;* more commonly, *Peptostreptococcus* is isolated. *Fusobacterium*, when isolated, is usually oropulmonary or pelvic in origin. Several clinical features are particularly distinctive in anaerobic bacteremia. Excessive jaundice with hyperbilirubinemia has been noted in 10% to 40% of cases.[24] Suppurative thrombophlebitis may be present in 5% to 12% of cases, primarily involving the pelvic, hepatic, mesenteric, and portal veins. Although anaerobic bacteria can cause endocarditis, such infections appear to be exceedingly rare.

INFECTIONS CAUSED BY *ACTINOMYCES*

Epidemiology

Actinomycosis is a relatively rare condition but has a worldwide distribution with no predilection for age, race, season, or occupation. However, a male-to-female predominance of 1.5:1 to 3:1 has been reported in many series. Predisposing factors include dental caries and extractions, gingivitis and gingival trauma, diabetes mellitus, immunosuppression, malnutrition, and local tissue damage caused by neoplastic disease or irradiation. In children, the development of actinomycosis should arouse suspicion of an underlying immunodeficiency state, particularly chronic granulomatous disease.

Etiology

Actinomyces species are nonsporulating strict or facultative anaerobes with a variable cellular morphology that ranges from diphtheroidal forms to coccoid filaments [*see Figure 1c*]. They are normal constituents of the oral, gastrointestinal, and genital flora. The term actinomycosis literally translates as "ray fungus," which reflects the organism's characteristic filamentous, fungus-like appearance in infected tissues. However, *Actinomyces* species are true bacteria with filaments that are much narrower than fungal hyphae. They require an enriched culture medium, such as brain-heart infusion broth, for growth; cultures should be observed for at least 14 to 21 days to allow adequate detection. Human actinomycosis is primarily caused by *A. israelii*. Other species known to cause human disease include *A. odontolyticus, A. naeslundii,* and *A. graevenitzii*. *A. bovis* causes the disease known as lumpy jaw in cattle but does not cause human disease. Actinomycosis is typically a polymicrobial infection; *A. actinomycetemcomitans* and *Haemophilus aphrophilus* are the most common coisolates. However, the significance of these coexisting bacteria in the pathogenesis of actinomycosis remains unclear.

Clinical Presentation

Actinomycosis is a chronic disease characterized by abscess formation, draining sinus tracts, fistulas, and tissue fibrosis. It can mimic a malignancy or granulomatous disease. A hallmark of actinomycosis is the tendency to spread without regard for anatomic barriers, including the fascial planes or the lymphatics; the development of multiple sinus tracts is also characteristic. Pain is generally an uncommon feature, particularly in chronic cases. Another characteristic is the presence of sulfur granules within infected tissue. The granules are 100 to 1,000 µm in diameter and are hard in consistency. They are often visible to the naked eye or by microscopy with low magnification [*see Figure 3a*]. The granules are composed of an internal tangle of mycelial fragments and a rosette of peripheral clubs [*see Figure 3b*]. Filaments within a granule are often visible on Gram or methenamine-silver stain, though more calcified granules may be difficult to identify. Cervicofacial involvement is the most common manifestation, accounting for 50% of all cases; thoracic, abdominal, pelvic, and disseminated infections occur less frequently.

Cervicofacial actinomycosis Fistulization from the perimandibular region is the most easily recognized manifestation of cervicofacial actinomycosis. Characteristic lesions usually develop slowly, over weeks to months, with adherence to overlying

skin giving it a bluish or reddish appearance. This is often mistaken for cellulitis but, in fact, more likely represents venous congestion. Over time, sinus tracts invariably form on the skin surface or oral mucosae, eventually erupting to express a thick yellow or serous exudate, which yields the characteristic sulfur granules. A characteristic inflammatory, cicatricial scarring eventually results. Less commonly, actinomycosis may present as an acute suppurative infection with a rapidly progressive, fluctuant, and pyogenic mass. At this stage, the patient may experience pain and trismus that appear disproportionate to the local (visible) inflammation.

Thoracic actinomycosis Pulmonary actinomycosis usually results from aspiration of organisms from the oropharynx. The disease has an insidious onset and a subacute course. Typical symptoms include cough, hemoptysis, chest wall discomfort, fever, and weight loss. Pulmonary osteoarthropathy is less common. The radiographic findings are variable and include patchy infiltrates, mass lesions, or cavitation. The infection may extend directly into the pleural space, ribs, and chest wall to produce empyema, osteomyelitis, and draining fistulous tracts. Less frequently, thoracic actinomycosis can extend to the mediastinum and present as pericarditis. The differential diagnosis includes carcinoma, tuberculosis, nocardiosis, and systemic mycosis. Because *A. israelii* is part of the normal oral flora, culturing the organism from sputum or bronchoscopic washings per se is not diagnostic of infection. Definitive diagnosis requires percutaneous needle aspiration, bronchoscopic biopsy, or open lung biopsy.

Abdominal actinomycosis Gastrointestinal actinomycosis generally originates from damaged intestinal mucosa. Any portion of the intestinal tract may be involved, but ileocecal infection is most frequent. The disease can spread to the omentum, mesenteric lymph nodes, and intra-abdominal viscera and can produce fistulous invasions of the abdominal wall or perineum. Presenting manifestations include pain and fever, palpable mass lesions, and draining sinus tracts. Carcinoma, tuberculosis, and Crohn disease are prominent considerations in the differential diagnosis. Unless draining sinus tracts are present, surgery is required for definitive diagnosis. If the bladder is involved, sulfur granules may be present in the urine. A careful search for the characteristic histopathology may be needed to establish the diagnosis, even when excisional biopsy is performed. It is important to confirm the diagnosis, because actinomycosis requires postsurgical administration of antibiotics to prevent recurrent disease.

Pelvic actinomycosis Pelvic actinomycosis is often associated with the use of intrauterine devices (IUDs) for contraception.[25] The pathogenesis of pelvic actinomycosis is not certain, but the disease likely results from upward spread of organisms from the perineum via the intestinal tract. Clinically, pelvic actinomycosis may present as endometritis, salpingo-oophoritis, or tubo-ovarian abscess; bladder invasion or systemic infection arising from a vaginal focus is rare. Optimal management of asymptomatic women with chronic *Actinomyces* colonization is uncertain. Removal of IUDs has been suggested, but the role of antibiotics is unclear. Pelvic inflammatory disease is the major consideration in the differential diagnosis. Surgery is generally required to establish the diagnosis. Actinomycosis of the male genitourinary tract is rare, but infections of the prostate have been reported.

Disseminated infection Disseminated actinomycosis most often occurs by hematogenous spread from a pulmonary focus. Any body site can be involved, including soft tissues, bone, the brain, and visceral organs.

INFECTIONS CAUSED BY CLOSTRIDIA

Clostridia are sporulating gram-positive bacilli; all species are obligate anaerobes, but some (e.g., *C. perfringens*) are relatively aerotolerant. Clinically important *Clostridium* species can be categorized into three major groups: histotoxic species (*C. perfringens*, *C. novyi*, *C. septicum*, *C. bifermentans*, and *C. sordellii*), enterotoxigenic species (*C. perfringens* and *C. difficile*), and neurotoxic species (*C. tetani* and *C. botulinum*).

Diseases from Histotoxic Clostridia

Most invasive infections caused by histotoxic clostridia originate from the gastrointestinal tract and are precipitated by trauma or underlying intestinal disorders. *C. perfringens* is the species that most commonly causes human disease [*see Figure 1d*], but isolation of *C. perfringens* may simply represent contamination of a wound surface. The spectrum of disease produced by *C. perfringens* and other histotoxic clostridia is broad, ranging in severity from relatively benign and localized conditions (e.g., a Welch abscess within a wound cavity) to fulminant infections associated with sepsis and high mortality (e.g., clostridial myonecrosis or parturient endometritis). *C. septicum* infections are particularly likely to occur in patients with underlying malignancies. *C. tertium* bacteremia occurs especially in neutropenic patients.

Soft tissue infections Clostridial crepitant cellulitis is a moderately serious gas-forming infection of the skin and subcutaneous tissues that does not involve muscle or produce a toxemic state, as does clostridial gas gangrene (clostridial myonecrosis; see below). It develops primarily in devitalized tissue in inadequately debrided wounds. The infection extends gradually through tissue planes and is accompanied by the formation of large quantities of gas, which is easily palpated as crepitus (more than is usually evident in cases of clostridial myonecrosis). A thin, dark, gray-brown, foul exudate is produced. There is relatively little local pain or change in the overlying skin. Diagnosis is usually based on a Gram stain revealing plump, gram-positive bacilli (without spores) and variable numbers of polymorphonuclear leukocytes. Of note, crepitant cellulitis is more commonly caused by non–spore-forming anaerobes than by clostridia [*see* Necrotic Skin and Soft Tissue Infections, *above*]. A more serious complication is necrotizing fasciitis caused by *C. sordellii* with toxic-shock manifestations in black tar heroin users.[26]

Clostridial myonecrosis Clostridial myonecrosis (i.e., gas gangrene or clostridial myositis) is a fulminant and rapidly progressive infection caused by *C. perfringens*. It initially involves injured or devitalized muscle and then aggressively invades contiguous normal muscle. Predisposing conditions include penetrating war wounds, surgery on the biliary tract or colon, infarcted bowel from an incarcerated hernia, arterial disease, and intramuscular injection of epinephrine. Spontaneous nontraumatic cases have also been reported. The incubation period is short, ranging from 8 to 72 hours. Onset is acute. Local pain is the earliest symptom, followed by pallor, apprehension, marked tachycardia, moderate fever, and lethargy. Hypotension, shock, and oliguria quickly supervene. There is extensive intravascular hemolysis, evidenced by the appearance of so-called port-wine

urine. The overlying skin is swollen and exquisitely tender and becomes dark yellow or bronze. In an involved extremity, the skin is pale and cold because of ischemia. Crepitus is present but not prominent. A thin, brownish discharge with a foul odor is evident; tense blebs containing dark, thin fluid often develop. The overall mortality in gas gangrene is 15% to 30%. Gas gangrene of the abdominal wall has a mortality of 50%.

Uterine infection and septic abortion Almost all cases of clostridial uterine infection are caused by *C. perfringens,* which is present in the genital tract in 5% of healthy women. Infection occurs in the setting of incomplete abortion, premature rupture of membranes, or operative termination of pregnancy. The organisms may enter and remain confined to the myometrium, or they may spread hematogenously, as occurs in the septicemic form of clostridial uterine infection. The incubation period is short, usually 12 to 72 hours, and is followed by vaginal bleeding, low-grade fever, and lower abdominal pain. The onset of systemic symptoms, such as vomiting, diarrhea, marked tachycardia, high fever, and chills, is abrupt. A foul-smelling vaginal discharge is present, the uterus and adnexa are tender, and perforation of the uterus may lead to pelvic peritonitis. Jaundice and hemoglobinemia may result from massive intravascular hemolysis. Hypotension, shock, and renal shutdown are common in such cases. Mortality in patients with intravascular hemolysis from postabortal *C. perfringens* uterine infection is approximately 50%. A variant of this infection is fulminant endometritis caused by *C. sordellii* in parturient young women, which is marked by profound leukocytosis and shock and which produces significant mortality.[27]

Diseases from Enterotoxigenic Clostridia

Food poisoning *C. perfringens* ranks second only to *S. aureus* in causing outbreaks of food poisoning in the United States.[28] Most strains of *C. perfringens* that are associated with food poisoning are heat-resistant type A organisms. The vehicle is usually a meat product that has been stewed or boiled and then left for a few hours before serving or rewarming. The initial cooking kills contaminating vegetative forms but not heat-resistant spores of *C. perfringens*. During cooling, the spores germinate and multiply in the anaerobic environment. Large numbers of viable organisms are then ingested. These organisms multiply and sporulate in the small intestine. Sporulation is associated with the elaboration of an enterotoxin that produces ileal fluid accumulation and diarrhea. Abdominal cramps and diarrhea develop about 8 to 12 hours after ingestion of contaminated food. Nausea occurs occasionally, but vomiting does not occur. Fever, headache, and systemic symptoms are absent. The illness is mild and runs its course in 24 hours.

Enteritis necroticans In developing countries, intestinal infection with *C. perfringens* type C can produce a β-toxin that causes a severe hemorrhagic, inflammatory, or ischemic necrosis of the jejunum known as enteritis necroticans or pigbel.[29] The infection primarily affects chronically ill persons who consume pig intestines (chitterlings).

***C. difficile*–associated diarrhea and colitis** *C. difficile* accounts for about 25% of all cases of diarrhea in patients receiving antibiotics. *C. difficile* is acquired in the gastrointestinal tract early in infancy, often in neonatal nurseries. Up to 78% of neonates harbor the organism, but the carrier rate declines to less than 50% in older children and to about 5% in healthy adults. Despite the high carrier rate, infants and young children are highly resistant to the effects of *C. difficile* toxins, possibly because their intestinal mucosa is relatively immature and lacks toxin receptors. The carrier rate is much higher in debilitated adults, and the organism is transmitted person-to-person in health care facilities. About 20% of all hospitalized patients acquire the organism, but two thirds remain asymptomatic.

At least 300,000 cases of *C. difficile*–associated diarrhea (CDAD) occur in the United States every year.[30] Patients receiving tube feedings are at particular risk. CDAD was first observed in patients receiving clindamycin, but many other antibiotics can cause diarrhea. The risk is highest from clindamycin, the cephalosporins, and ampicillin and lowest for vancomycin, metronidazole, and the aminoglycosides. There is also an increased risk of CDAD associated with the use of proton pump inhibitors in hospital inpatients.[31] In rare instances, CDAD may develop in patients who have not been hospitalized or exposed to antibiotics. For example, cancer chemotherapy predisposes to *C. difficile* infection even in the absence of antibiotic therapy.

The clinical presentation of CDAD is quite variable. The symptoms range from mild diarrhea to severe colitis with fever, leukocytosis, abdominal cramps, and bloody diarrhea. The onset may be within the first few days of antibiotic therapy or as late as 6 weeks after antibiotics have been discontinued. Endoscopy reveals pseudomembrane formation in about half of cases. Toxic megacolon may occur in severe cases, particularly in patients who have received antimotility agents. Intestinal perforation may occur but is uncommon.

Diseases from Neurotoxic Clostridia

Tetanus All of the clinical features of tetanus are caused by a potent neurotoxin—tetanospasmin—produced by *C. tetani*. The toxin travels to the spinal cord and suppresses the inhibitory neurotransmitter GABA in the neuromuscular junction, resulting in severe muscle spasms.

Tetanus is now rare in industrialized countries because of widespread immunization. In the United States, fewer than 50 cases are reported annually, mostly in inadequately immunized older adults.[32] In developing countries, tetanus is still a major problem, causing the deaths of an estimated one million persons annually, half of whom are newborns. *C. tetani* is commonly found in the soil, in the intestines of domestic animals, and occasionally in human feces. The organism is commonly introduced by a laceration or puncture wound that is usually sustained outdoors. Tetanus may also occur in association with pregnancy (postpartum and postabortion tetanus), injection of illicit narcotics, surgery (postoperative tetanus), burns, vaccination, intramuscular injections, chronic skin ulcers, dog bites, and umbilical stump infection in newborns (neonatal tetanus). In 10% to 20% of patients with tetanus, there is no history of injury or evidence of an infected lesion.

The disease is characterized by generalized rigidity and intermittent, intense muscle spasms. The incubation period ranges from 1 to 55 days, but onset of symptoms occurs within 14 days after the initial injury in over 80% of patients. The usual presenting symptoms are restlessness; pain caused by muscle spasm; and stiffness of the back, neck, thighs, and abdomen. When muscle spasms occur, the characteristic clinical features are determined by the relative strengths of the opposing muscles: the greater strength of the masseter over the opposing digastricus and mylohyoid results in trismus; the greater strength of the ex-

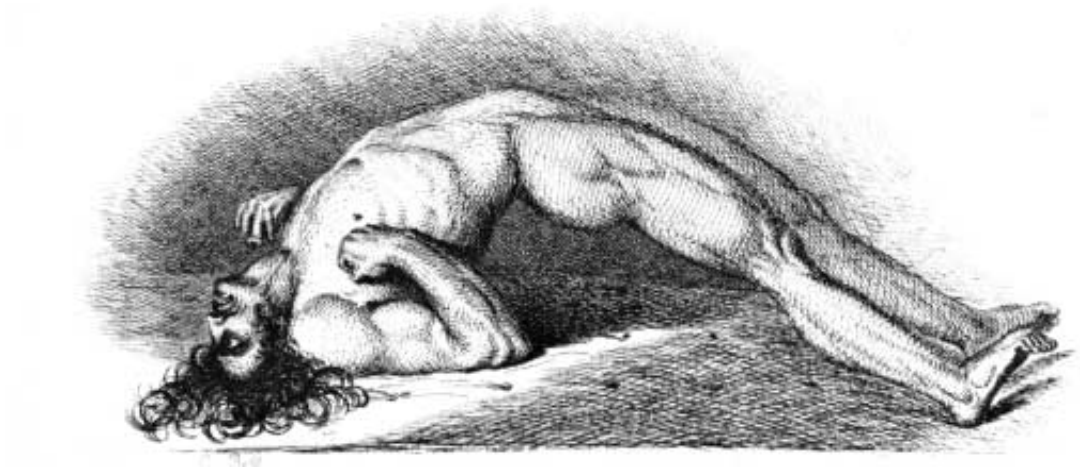


Figure 4 **Spasms in a wounded soldier with tetanus are illustrated in this drawing by Scottish surgeon and anatomist Sir Charles Bell in his book *The Anatomy and Philosophy of Expression*, published in 1832. The classic signs of tetanus—risus sardonicus, trismus, and opisthotonos—are shown.**

tensor groups over the flexors in the lower extremities produces characteristic extension at the hips and knees; and the greater strength of the biceps results in flexion of the forearms. This combination of flexion of the upper extremities and extension of the lower extremities is termed opisthotonos [*see Figure 4*].

Difficulty in opening the mouth (trismus, or lockjaw) is the first symptom in more than 50% of patients. Dysphagia, caused by spasm of pharyngeal muscles, may be an early symptom. Deep tendon reflexes are hyperactive, but the plantar responses are flexor. As the process progresses, violent spasms of the paraspinal, abdominal, and limb musculature occur, but the patient remains conscious. Trismus and stiffness of the facial muscles produce risus sardonicus, a characteristic sneering expression. Sudden stimuli (e.g., bright light or noise) can precipitate tonic seizure accompanied by diaphragmatic, intercostal, glottal, or laryngeal spasm; such spasms can result in hypoxia and respiratory arrest. Fever may be caused by the marked muscular rigidity and spasms alone. Severe sympathetic hyperactivity, evidenced by labile hypertension or hypotension, tachypnea, tachycardia, arrhythmias, profuse sweating, and marked intermittent vasoconstriction, may occur singly or in varying combinations. After the manifestations of tetanus have peaked, they persist at the same level for about a week and then gradually diminish over several weeks. Residual stiffness may persist for several more weeks. In patients who have survived moderate to severe tetanus, respiratory assistance has been required for 2 to 4 weeks; the average hospital stay has been 5 to 8 weeks. The major complications of tetanus are respiratory arrest secondary to tetanic spasms, pneumonia secondary to aspiration, pulmonary emboli, cardiac problems related to sympathetic overactivity or to cardiomyopathy, and fractures of thoracic vertebrae caused by violent spasms.

Botulism About 100 cases of botulism are reported in the United States each year, but a surge in cases associated with black tar heroin and the threat of international bioterrorism have renewed interest in a disease that was first recognized in the early 17th century.[33]

As with tetanus, all the clinical manifestations of botulism are caused by a potent neurotoxin produced by *C. botulinum.* The organism has been isolated from soil everywhere in the world. The spores are very hardy, resisting dryness and extremes of temperature. Spores that are introduced into food, wounds, or the human intestinal tract can germinate and elaborate botulinum toxin. Although seven antigenically distinct forms of the toxin (serotypes A through G) have been identified, just three types (A, B, and E) account for nearly all human disease. Botulism occurs in three main forms: food-borne botulism, wound botulism, and infant botulism.

Food-borne botulism results from the ingestion of home-processed foods that are improperly cooked or refrigerated. The vegetative organisms produce the toxin, which is heat labile and can be destroyed by boiling for 10 minutes or by heating to 80° C for 30 minutes. However, if contaminated food is not heated sufficiently, the toxin resists gastric acid and intestinal trypsin and is absorbed from the intestinal tract. About half the cases of food-borne botulism are caused by type A toxin; the rest are divided evenly between types B and E.

Wound botulism typically follows severe trauma, such as a crush injury involving an extremity. However, heroin use has become the leading predisposing factor in the United States; most cases have occurred in California after subcutaneous injection of black tar heroin from Mexico. About 80% of wound botulism is caused by type A toxin; most of the rest is caused by type B toxin.

Infant botulism is caused by the ingestion of *C. botulinum* spores rather than preformed toxin. The spores then germinate, colonizing the intestinal tract with toxin-producing organisms. Although the source of the spores eludes detection in most cases, honey has been implicated in about 15% of cases. Toxin type A and type B each accounts for about half the cases of infant botulism.

The symptoms of food-borne botulism usually begin 18 to 36 hours after ingestion of the toxin; the incubation period for wound botulism is typically longer. Although food-borne botulism may be heralded by gastrointestinal symptoms such as nausea, vomiting, abdominal cramps, and diarrhea (earlier stage) or constipation (later stage), neurologic manifestations soon predominate. In wound botulism, gastrointestinal symptoms are absent and the wound may appear surprisingly benign. Cranial nerve symptoms such as blurred vision and diplopia are usually the earliest neurologic complaints, followed by dysphagia, dysarthria, and dry mouth. Symmetrical motor paralysis ensues, characteristically progressing in a descending fashion that be-

gins with the arms and then involves the respiratory muscles and lower body. Autonomic dysfunction can produce constipation, urinary retention, and orthostatic hypotension. Sensory deficits are absent, and mentation is normal. Respiratory arrest occurs in severe cases; mechanical ventilation and respiratory support may be required for weeks to months before full recovery. Infant botulism presents as lethargy, constipation, poor feeding, and floppiness, typically in the second month of life.

NEWLY RECOGNIZED ANAEROBES

Molecular tools such as 16S ribosomal DNA sequencing are continuing to identify new genera and species of clinically relevant anaerobic bacteria.[34,35] For example, the anaerobic gram-negative bacillus *Bilaphila wadsworthia* is now known to be an important pathogen and is frequently isolated from gangrenous appendicitis.[36] Newly described anaerobic cocci, gram-positive non–spore-forming rods, and clostridia have also been isolated from various infections.[35,37]

Diagnosis of Anaerobic Infections

CLINICAL CLUES TO ANAEROBIC INFECTIONS

Apart from actinomycosis and clostridial myonecrosis, infections involving obligate anaerobes are generally indistinguishable from infections caused by other pathogens. Clinical manifestations are largely determined by the organ system involved and by the extent and chronicity of the infection. The two most helpful clinical clues are the presence of local tissue ischemia or necrosis and the proximity of infection to mucosal surfaces where obligate anaerobes normally reside. A putrid, foul-smelling discharge is virtually diagnostic of infection involving anaerobes, although the absence of foul odor does not rule out this possibility. Similarly, crepitus or black discoloration of affected tissue is only suggestive evidence.

MICROBIOLOGIC DIAGNOSIS OF ANAEROBIC INFECTIONS

A well-performed Gram stain of appropriately collected clinical material is a very useful diagnostic tool. Anaerobic infections are typically polymicrobial, and the characteristic cellular morphology of certain anaerobic pathogens may be recognized by an accomplished microscopist [*see Figure 1*]. The finding of so-called sterile pus by conventional culture methods in the face of a positive Gram stain should be considered presumptive evidence of an anaerobic infection. In the final analysis, however, the accurate diagnosis of anaerobic infection depends on the ability of the laboratory to isolate these fastidious organisms from clinical material likely to yield meaningful bacteriologic data.

Specimen Collection and Transport

One of the major handicaps in the recovery of anaerobic bacteria is improper specimen collection and transport. Care must be taken to avoid specimens that may be contaminated by commensal flora of mucocutaneous surfaces where anaerobes normally reside (e.g., throat swabs, expectorated sputum, voided urine, bronchoscopic and nasotracheal aspirates, vaginal secretions, feces, colostomy effluent, or superficial wound swabs). Blood and other body fluids that are normally sterile and aseptically obtained should be routinely cultured for anaerobic bacteria. Other clinical materials likely to yield meaningful bacteriologic data for anaerobic infections include specimens from tissue biopsy or curettage or from deep wounds during surgery.

Proper specimen transport to preclude aeration is critical for microbiologic confirmation of an anaerobic infection. Many fastidious organisms are extremely oxygen sensitive and cannot withstand even a brief moment of exposure to air. Furthermore, in mixed infections, the presence of facultative organisms that grow faster than anaerobes frequently precludes recovery of the latter. Several commercially available systems for anaerobic transport of clinical specimens have been evaluated and have been shown to provide excellent recovery of fastidious anaerobes.[38] If commercial anaerobic transport vials are not available, specimens should be collected with a sterile needle and syringe. Air in the syringe is carefully expelled. The needle is capped to minimize aeration, and the specimen should be promptly delivered to the clinical laboratory. If swabs are to be used, they should be prepared, stored, and transported in gas-filled containers under anaerobic conditions. Immediate processing of specimens by the laboratory also improves recovery, but in practice, this is often not feasible.

RADIOLOGIC AND IMAGING STUDIES

Noninvasive tests such as computed tomography, magnetic resonance imaging, ultrasonography, and gallium or indium scanning are most useful for localization of suppurative infections in the central nervous system and in intra-abdominal and pelvic organs. The sensitivity and specificity of these tests in the detection of abscess and the differentiation from tumor, hematoma, and other noninflammatory space-occupying lesions in various sites remain to be determined by careful prospective study. In general, it may be said that a positive scan is highly suggestive, particularly when supported by the clinical picture; however, a negative scan is much less useful in ruling out infection.

SPECIFIC ANAEROBIC INFECTIONS

Actinomycosis

The diagnosis of actinomycosis depends on identification of the organism by smear or culture and by characteristic histopathology from tissue biopsy. The identification of sulfur granules establishes the diagnosis of actinomycosis [*see Figure 3*]; however, sulfur granules may constitute no more than 1% of total tissues in a given lesion and, hence, are easily missed by routine tissue staining. Similar granules may be seen with other microorganisms, notably *Nocardia brasiliensis* and *Streptomyces madurae* (both of which can cause mycetoma), as well as *S. aureus* (a cause of botryomycosis). However, these other granules do not have peripheral clubs, which appear to be specific to *Actinomyces* species. Not all *Actinomyces* species form sulfur granules (e.g., *A. odontolyticus* does not), and a peripheral fringe of clubs may be absent in certain instances, such as in a tonsillar crypt infection or in pelvic actinomycosis associated with an IUD. Additionally, *Actinomyces* species can be morphologically differentiated from *Nocardia;* moreover, *Nocardia* is acid fast in modified acid-fast stains, whereas *Actinomyces* is not.

Clostridial Myonecrosis versus Crepitant Cellulitis

Surgical exploration is necessary to distinguish myonecrosis from anaerobic cellulitis. In gas gangrene, the involved muscle looks cooked and lacks contractility, whereas in clostridial cellulitis, the muscle is visibly healthy. A presumptive diagnosis is based on a typical Gram stain of the wound drainage or aspirate that reveals many clostridia but few leukocytes. The presence of gas in subcutaneous tissue is not pathognomonic of clostridial

infection. Patients with diabetes mellitus are particularly prone to crepitant cellulitis caused by enteric bacteria or *Bacteroides* species. Perineal phlegmons, which result from extension of perirectal abscesses caused by mixed anaerobic and facultative organisms, may also involve subcutaneous gas formation. Crepitus from trapped air after traumatic injury can usually be distinguished from anaerobic cellulitis or myonecrosis by the fact that the former does not spread.

C. difficile–Associated Diarrhea

The diagnosis of CDAD is established by demonstrating the toxins of *C. difficile* in stool specimens by immunoassays. Toxins A and B can be detected using specific antibodies. In approximately 5% to 20% of patients, more than one stool specimen is required to detect *C. difficile* toxin. Consequently, when enzyme-linked immunosorbent assay results are negative but clinical suspicion is high, tests should be repeated using the tissue culture cytotoxicity assay. Some clinical laboratories utilize screening tests that detect the presence of *C. difficile* in fecal specimens, either by culture or by detecting the presence of glutamate dehydrogenase, a metabolic enzyme expressed at high levels by all strains of *C. difficile*, both toxigenic and nontoxigenic.[39] Although these rapid screening tests may be cost-effective in some instances where large volumes of fecal specimens are processed, they are more suitable for excluding rather than establishing the diagnosis of CDAD because of their high negative (approximately 98%) but low positive (approximately 60%) predictive values.

Tetanus

The diagnosis of fully developed tetanus presents little difficulty. Acute strychnine poisoning is the only disease that resembles tetanus. A greater problem in the differential diagnosis occurs earlier in the course of the illness, when trismus is the principal manifestation. Trismus may occur in patients with intraoral disease, especially dental or jaw infections, and is occasionally seen in patients with trichinosis. Hepatic encephalopathy is sometimes associated with prominent muscle stiffness and rigidity. However, the associated liver disease is usually obvious. Furthermore, a sudden stimulus, such as jarring a bed rail, is likely to cause spasms in a patient with tetanus but not in a patient with hepatic encephalopathy. Trismus may also develop as an acute reaction to phenothiazines (the so-called grimacing syndrome). Unlike trismus from tetanus, the masseter muscle spasm in this drug reaction is painful and intermittent, and to some degree it can be overcome voluntarily. This drug reaction is readily reversed with intravenous diphenhydramine.

Botulism

Because botulism is uncommon, the diagnosis may not be entertained despite characteristic clinical findings. Clustering of cases in a family or community and a history of eating home-canned or spoiled foods may be important clues. The differential diagnosis includes Guillain-Barré syndrome, Eaton-Lambert syndrome, myasthenia gravis, cerebrovascular accidents, tick paralysis, and chemical intoxication. In patients with botulism, results of complete blood counts, blood chemistries, CNS imaging studies, and cerebrospinal fluid analysis are all normal. Rapid repetitive electromyography, however, is highly suggestive of botulism if it demonstrates a pattern of facilitation. A positive diagnosis can be established by demonstrating botulinum toxin in serum or stool specimens; the toxin may also be detected in food samples.

Management of Anaerobic Infections

Successful treatment of anaerobic infections requires rational antibiotic selection in conjunction with judicial surgical resection and drainage. The choice of antibiotics should be guided by culture results and antibiotic susceptibility data.

Several methods for antimicrobial susceptibility testing of obligate anaerobes have been validated by the National Committee for Clinical Laboratory Standards.[40] Both agar dilution and microbroth testing methods are appropriate, whereas the E-test, which utilizes a predefined gradient of antibiotic concentrations on a plastic strip, offers a more expensive but practical and fairly accurate alternative for susceptibility testing of individual anaerobic isolates. In light of the growing concern of emerging antibiotic resistance among anaerobic bacteria, the need for more regular susceptibility testing of clinical isolates of anaerobic bacteria has become evident.[41] Susceptibility testing is particularly important in clinical settings where there has been a suboptimal response to empirical antibiotic regimens. Certainly, antimicrobial susceptibility testing should be routinely performed on organisms that are frequently resistant to antibiotics commonly used as empirical therapy, such as members of the *B. fragilis* group; pigmented *Prevotella*, including *P. bivia* and *P. disiens;* and certain *Fusobacterium* species. In the absence of specific culture or susceptibility data, initial antibiotic therapy must be chosen empirically and directed against the pathogens most likely to be present in a particular clinical setting [*see Table 5*], in accordance with predicted in vitro susceptibility patterns [*see Table 6*].

PREDICTED ANTIMICROBIAL SUSCEPTIBILITY

Although penicillin G has been considered the agent of choice for a number of mixed infections at various sites above the diaphragm (particularly oropulmonary and head and neck infections), β-lactamase production and treatment failure have been increasingly reported.[42] β-lactamase production is increasingly recognized in oral isolates of *P. intermedia, F. nucleatum,* and *Peptostreptococcus micros*.[42,43] Among the cephalosporins, only cefoxitin, cefotetan, and ceftizoxime have an enhanced antianaerobic spectrum. These agents appear to have comparable activity against *B. fragilis,* with resistance rates ranging from 10% to 20%; none are as active as clindamycin or metronidazole. Among the penems, imipenem-cilastatin, meropenem, and erzapenem are the most broadly active.[44] The monobactam aztreonam is inactive against anaerobes, as well as gram-positive aerobes.

All strains of *B. fragilis* produce β-lactamases and are resistant to penicillin, but extended-spectrum penicillins in combination with β-lactamase inhibitors (e.g., ampicillin-sulbactam, ticarcillin-clavulanate, and piperacillin-tazobactam) are active against most strains. Increasing resistance of *B. fragilis* to cefoxitin and clindamycin has also been reported, and they are no longer the agents of choice in intra-abdominal infections.[45] Cefotetan, ceftizoxime, piperacillin-tazobactam, imipenem, and meropenem remain active. Antibiotic susceptibilities of the non-*fragilis* species of the *B. fragilis* group are more variable than those of *B. fragilis*. Only metronidazole, imipenem, and chloramphenicol are predictably active against nearly all isolates.[46]

Erythromycin and ketolides are relatively inactive against *Fusobacterium* species and most *B. fragilis* strains. Similarly, the first- or second-generation quinolones (e.g., norfloxacin, ciprofloxacin, enoxacin, ofloxacin, and levofloxacin) are relatively inactive as single-agent therapy for anaerobic or mixed infections. However, the third-generation quinolones moxifloxacin and gatifloxacin have good in vitro activity against most anaerobes, including

Table 6 Predicted in Vitro Susceptibility of Clinically Important Anaerobes to Major Classes of Antimicrobial Agents

Antibiotic	*Above Diaphragm*		*Above or Below Diaphragm*		*Below Diaphragm*	
	Fusobacterium *Species*	Porphyromonas *and* Prevotella *Species*	Peptostreptococcus *Species*	Actinomyces *Species*	Bacteroides fragilis *Group*	Clostridium *Species*
Penicillin	S	S-R	S	S	R	S*
Ampicillin-sulbactam†	S	S	S	S	S	S
Piperacillin, ticarcillin	S	S	S	S	S-R	S
Piperacillin-tazobactam	S	S	S	S	S	S
Imipenem, meropenem	S	S	S	S	S	S
Cefazolin	S	S-R	S	S-R	R	S
Cefoxitin	S	S-R	S	S-R	S	S-R
Cefotetan	S	S	S	S-R	S-R	S
Ceftizoxime	S	S	S	S-R	S	S-R
Cefoperazone, cefotaxime	S	S	S	S-R	S-R	S
Ceftriaxone, ceftazidime	S	S-R	S	S-R	S-R	S
Clindamycin	S*	S	S*	S	S*	S-R
Macrolides	R	S-R	S	S-R	R	S
Metronidazole	S	S	S-R	S	S	S
Ciprofloxacin, levofloxacin	R	R	R	R	R	R
Moxifloxacin, gatifloxacin, gemifloxacin	S	S	S	S	S	S
Tetracycline	S	S-R	S-R	S	S-R	S-R

S—> 80% of strains sensitive S-R—30%–80% of strains sensitive R—< 30% of strains sensitive
*Emerging resistance noted.
†Similar combinations currently available, including amoxicillin-clavulanate and ticarcillin-clavulanate, are comparably active.

B. fragilis,[47] whereas gemifloxacin is less active.[48]

Metronidazole has excellent activity against *B. fragilis*, *Fusobacterium* species, and *Clostridium perfringens*. *Peptostreptococcus* and *Bacteroides* species other than *B. fragilis* are only moderately sensitive, whereas nonsporulating gram-positive bacilli are relatively resistant. Metronidazole lacks activity against aerobic bacteria and should not be used as a single agent for empirical therapy, because most infections involving anaerobic bacteria are in fact mixed infections. On the other hand, metronidazole is the only agent with consistent bactericidal activity against *B. fragilis*. Metronidazole crosses the blood-brain barrier well, so it is particularly useful for treating anaerobic brain abscess or infective endocarditis.

Tetracycline and its analogues can no longer be recommended for the empirical treatment of anaerobic infections because of the substantial resistance acquired by *B. fragilis* and virtually all classes of other anaerobic bacteria. Tetracycline remains useful in the treatment of actinomycosis, however. Trimethoprim-sulfamethoxazole has only limited activity against anaerobic bacteria. Vancomycin is effective against some gram-positive anaerobes (particularly *C. difficile*), but it has no activity against gram-negative anaerobes. Aminoglycosides are uniformly inactive against obligate anaerobes.

EMPIRICAL ANTIMICROBIAL THERAPY

The recommended regimens for empirical therapy for various anaerobic or mixed infections vary according to the site of infection [*see Table 7*]. In general, therapy should be directed at both the aerobic and anaerobic components of the suspected microflora. Monotherapy with a broad-spectrum single agent (e.g., ceftizoxime, cefotetan, ampicillin-sulbactam, ticarcillin-clavulanate, piperacillin-tazobactam, imipenem-cilastatin, meropenem, or ertapenem) may be used to minimize toxicity and reduce cost. Parenteral administration, relatively high dosages, and prolonged duration of treatment (3 to 6 weeks) are usually required because of the extent of tissue necrosis and the tendency for relapse with these infections. There are insufficient published data to evaluate the efficacy of empirical therapy with the newer

Empirical Antimicrobial Regimens for Suspected Anaerobic or Mixed Infections

Site or Type of Infection	*Treatment of Choice*	*Alternative Regimens*
Teeth and periodontium, lungs, pleura, human-bite wounds	Ampicillin-sulbactam, 1–2 g I.V. q. 6 hr Penicillin G, 4 million U I.V. q. 4 hr *plus* metronidazole, 500 mg I.V. q. 6 hr	Clindamycin, 600 mg I.V. q. 6 hr Cefotaxime, 2 g I.V. q. 6 hr Ceftizoxime, 1–2 g I.V. q. 8 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr
Ears, nose, sinus, mastoids	Ciprofloxacin, 400 mg I.V. q.12 hr *plus* metronidazole, 500 mg I.V. q. 6 hr *plus* penicillin G, 4 million U I.V. q. 4 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr	Piperacillin-tazobactam, 4/0.5 g I.V. q. 6 hr Imipenem, 0.5–1 g I.V. q. 6 hr Moxifloxacin, 400 mg I.V. q. 24 hr Gatifloxacin, 400 mg I.V.
Brain abscess, subdural empyema		
Odontogenic	Penicillin G, 4 million U I.V. q. 4 hr *plus* metronidazole, 500 mg I.V. q. 6 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr	Cefotaxime, 2 g I.V. q. 6 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr Imipenem, 1 g I.V. q. 6 hr
Otogenic or rhinogenic	Ciprofloxacin, 400 mg I.V. q.12 hr. *plus* metronidazole, 500 mg I.V. q. 6 hr *plus* penicillin G, 4 million U I.V. q. 4 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr	Piperacillin-tazobactam 4/0.5 g I.V. q. 6 hr Imipenem, 0.5–1 g I.V. q. 6 hr
Intra-abdominal, pelvic, or necrotic soft tissue infections	Ampicillin-sulbactam, 1.5–3 g I.V. q. 6 hr Ticarcillin-clavulanate, 3.1 g I.V. q. 6 hr Ertapenem, 1 g I.V. q. 24 hr Ciprofloxacin, 400 mg I.V. q. 12 hr *plus* metronidazole, 500 mg I.V. q. 6 hr *plus* penicillin G, 4 million U I.V. q. 4 hr	Piperacillin-tazobactam, 4/0.5 g I.V. q. 6 hr Imipenem, 0.5–1 g I.V. q. 6 hr Third- or fourth-generation cephalosporin, I.V. (cefotaxime, 2 g q. 6 hr; ceftriaxone, 1–2 g q. 12–24 hr; ceftizoxime, 1–2 g q. 8–12 hr; ceftazidime, 2 g q. 12 hr; cefepime, 1–2 g q. 12 hr) *plus* metronidazole, 500 mg ceftazidime q. 6 hr
Actinomycosis	Penicillin G, 3–4 million U I.V. q. 4 hr Amoxicillin, 500 mg I.V. q. 8 hr	Clindamycin, 600 mg I.V. q. 6 hr Doxycycline, 100 mg I.V. q. 12–24 hr
Clostridial myonecrosis	Penicillin G, 4 million U I.V. q. 4 hr *plus* clindamycin, 600 mg I.V. q. 6 hr *with* or *without* gentamicin, 2–2.5 mg/kg I.V. q. 8 hr or ciprofloxacin, 400 mg I.V. q. 12 hr	Piperacillin-tazobactam, 4/0.5 g I.V. q. 6 hr Imipenem, 0.5–1 g I.V. q. 6 hr Third- or fourth-generation cephalosporin I.V. (cefotaxime, 2 g q. 6 hr; ceftriaxone, 1–2 g q. 12–24 hr; ceftizoxime, 1–2 g q. 8–12 hr; ceftazidime, 2 g q. 12 hr; cefepime, 1–2 g q. 12 hr) *plus* metronidazole, 500 mg I.V. q. 6 hr
C. difficile–associated enterocolitis	Metronidazole, 500 mg p.o., q. 8 hr	Vancomycin, 125 mg p.o., q. 6 hr Bacitracin, 250,000 U p.o., q. 6 hr
Tetanus	Metronidazole, 500 mg I.V. q. 6 hr Penicillin G, 3 million U I.V. q. 6 hr	First- or second-generation cephalosporin I.V. (cefazolin, 0.5–1 g q. 6–8 hr; cefuroxime, 750 mg q. 8 hr; ceftriaxone, 1 g q. 12–24 hr)
Botulism	Penicillin G, 3 million U I.V. q. 6 hr Metronidazole, 500 mg I.V. q. 6 hr	First- or second-generation cephalosporin I.V. (cefazolin, 0.5–1 g q. 6–8 hr; cefuroxime, 750 mg q. 8 hr; ceftriaxone, 1 g q. 12–24 hr)

quinolones (e.g., moxifloxacin and gatifloxacin) against severe anaerobic or mixed infections, despite their favorable in vitro susceptibility profiles.

SURGICAL DRAINAGE

Surgical drainage of abscesses and resection of necrotic tissue may be the decisive therapeutic modality for most suppurative anaerobic infections. However, several exceptions are noteworthy. In lung abscess, nonsurgical treatment alone is often effective, perhaps because of spontaneous drainage and expectoration of abscess contents through the tracheobronchial tree. Certain cerebral abscesses, even when well encapsulated, may also respond to antibiotics alone. A similar favorable experience has been noted with hepatic and tubo-ovarian abscesses. Although abscesses do not always require drainage, it is not clear what factors reliably predict a favorable response to antibiotics alone. When necrotic tissue is present, however, surgical excision is invariably required.

HYPERBARIC OXYGEN AND OTHER ADJUNCTIVE MEASURES

Hyperbaric oxygen therapy, consisting of 100% oxygen at 2 to 3 atm administered for approximately 2 hours, has been used as an adjunct to surgical debridement and antibiotics in the treatment of clostridial gas gangrene and other necrotizing soft tissue infections. Despite enthusiasm for hyperbaric oxygen therapy, there have been no controlled trials of its use in humans, and its precise role in the management of these infections remains to be defined. Nevertheless, hyperbaric oxygen therapy has been recommended for selected cases of necrotic soft tissue infections and for recalcitrant anaerobic osteomyelitis of the maxilla or mandible. Other adjunctive measures useful for the management of anaerobic infections include topical wound irrigation with 3% hydrogen peroxide solution to control foul-smelling discharge and administration of fibrinolytic agents (e.g., trypsin) to prevent intra-abdominal abscess formation postoperatively.[49]

Actinomycosis

High-dose penicillin remains the drug of choice for actinomycosis. For severe infections, a 4- to 6-week course of intravenous penicillin G should be followed by oral penicillin V (2 to 4 g/day in four divided doses) for 6 to 12 months. For mild infections, a 2-month course of oral penicillin V is generally appropriate. Oral amoxicillin (500 mg three times daily) is equally effective. Acceptable alternatives to penicillin include the tetracyclines, erythromycin, and clindamycin. Agents generally deemed to have poor activity against *Actinomyces* species include oral cephalexin; oxacillin and dicloxacillin; the first- and second-generation fluoroquinolones; metronidazole; aminoglycosides; and aztreonam.[50] For patients with penicillin allergy, tetracycline probably offers the best alternative, especially in milder disease. Therapy does not need to be directed against other commensal flora that are recovered along with *Actinomyces* species, because antibiotic regimens effective against *Actinomyces* alone are usually curative. Surgical intervention may be necessary in more complicated cases.

Clostridial Myonecrosis

Aggressive surgical debridement is mandatory for suspected gas gangrene to determine the extent of infection and to eliminate all necrotic tissue. The procedure includes multiple incisions for drainage; fasciotomy for decompression of muscle compartments; and excision of necrotic muscle. If the process is extensive and if irreversible changes have occurred in an extremity, amputation becomes necessary. Early use of hyperbaric oxygen therapy may limit the zone of frank muscle necrosis, reducing the extent of the surgical debridement. Simple drainage, with fasciotomy where indicated, is usually sufficient for treating clostridial cellulitis. Antibiotic therapy is required to treat bacteremia and prevent the spread of infection. The combination of penicillin G and clindamycin is recommended, because the latter rapidly inhibits toxin synthesis. When a Gram stain indicates a polymicrobial infection, monotherapy with broad-spectrum agents, such as piperacillin-tazobactam, imipenem, or meropenem, is indicated. Alternatively, an aminoglycoside (e.g., gentamicin), a newer cephalosporin (e.g., cefotaxime), or a fluoroquinolone (e.g., ciprofloxacin) may be added initially.

C. difficile–Associated Diarrhea

Symptomatic patients should receive oral metronidazole or vancomycin. Either agent, given at a dosage of 500 mg three times daily for 10 days, produces a clinical cure in 94% of patients. Metronidazole is considerably less expensive, but some *C. difficile* strains are resistant to the drug. Because vancomycin is active against all isolates and is not absorbed in the upper gastrointestinal tract, it may be preferable for severely ill patients. Orally administered bacitracin alleviates symptoms as effectively as metronidazole or vancomycin but is less effective in eradicating *C. difficile* and its toxins from the stools. For patients with CDAD who cannot tolerate oral therapy, intravenous therapy with vancomycin plus metronidazole or intravenous therapy with metronidazole alone has been recommended.

Regardless of the treatment regimen, symptoms recur in about 20% of patients.[51] Relapse is less likely to occur in patients who mount a serum antibody response to toxin A during the initial episode. Patients with recurrent symptoms should be retreated with oral metronidazole or vancomycin. Various regimens, including prolonged and tapering courses of vancomycin, have been suggested.[52] Cholestyramine resin, which binds the toxin in the intestine, may be useful for some patients. It is administered in conjunction with oral vancomycin for 1 to 2 weeks (4 g three or four times daily). Administration of the yeast *Saccharomyces boulardii* may improve the results of antibiotic therapy in some patients who experience relapses of *C. difficile* colitis.[53] Attempts to reconstitute the normal colonic flora by the administration of donated stool directly through a colonoscope have also been tried, with some success.[51] Prevention of antibiotic-associated diarrhea by the coadministration of *Lactobacillus* GG was evaluated in one randomized clinical trial but was unsuccessful during 21 days of follow-up.[54]

Tetanus

Treatment of tetanus is aimed at controlling muscle spasms, managing dysautonomia, neutralizing circulating toxin, eliminating the continuing source of the toxin, and preventing respiratory complications. Patients with tetanus should be monitored in an intensive care unit. External stimuli (e.g., noise or bright lights) that may precipitate muscle spasms must be kept to a minimum. A urinary catheter is required. Because of the danger of precipitating pharyngeal or laryngeal spasms by oral feedings, fluids and electrolytes are initially administered intravenously. Later, when the danger of aspiration is reduced, a nasojejunal tube may be used, provided that a cuffed endotracheal tube or tracheostomy tube is in place. Antitoxin treatment with human tetanus immune globulin (500 to 1,000 units) is administered intramuscularly; it neutralizes circulating toxin only.

The use of muscle relaxants is essential. Intravenous diazepam (40 to more than 200 mg a day, titrated according to need) is the drug of choice because it acts rapidly as a muscle relaxant and produces a sedative effect without inducing depression. If severe spasms cannot be controlled by diazepam, assisted ventilation and neuromuscular blockade may be necessary. Beta blockers, such as propranolol and labetalol, are used to control sympathetic overactivity.

The value of antimicrobial agents in the treatment of tetanus is doubtful. The only beneficial effect would be to clear the wound of vegetative cells of *C. tetani* that could produce additional toxin. Penicillin has been the traditional drug of choice, but metronidazole is now preferred. If feasible, wound debridement is carried out to eliminate the site of toxin elaboration. Surgical wound care should be performed only after the initial doses of tetanus immune globulin and antibiotics have been administered and after muscle spasms have been controlled. The wound should be thoroughly irrigated and left open. Because clinical tetanus does not establish natural immunity, the patient should be immunized with the first dose of adsorbed tetanus toxoid before discharge.

Botulism

Antitoxin should be administered promptly to neutralize any circulating botulinum toxin before it binds to cholinergic synapses. The currently recommended dose is one vial of intravenous trivalent (types A, B, and E) antitoxin, which is of equine origin and available from the Centers for Disease Control and Prevention (404-639-2206 during business hours; 404-639-2888 at other times). Hypersensitivity reactions can occur in up to 7% of recipients; such reactions may be severe. A human botulism immune globulin is being developed for the treatment of infant botulism. Wound botulism should be managed with surgical debridement and intravenous penicillin. Patients with severe botulism require mechanical ventilation and metabolic support.

Table 8 Recommendations for Tetanus Prophylaxis in Routine Wound Management

History of Immunization with Adsorbed Tetanus Toxoid	*Clean, Minor Wounds*		*All Other Wounds**	
	Td†	*TIG*	*Td*†	*TIG*
Unknown or fewer than three doses	Yes	No	Yes	Yes
Three or more doses‡	No§	No	No‖	No

*Such as, but not limited to, wounds contaminated with dirt, feces, soil, saliva, etc.; puncture wounds; avulsions; and wounds resulting from missiles, crushing, burns, and frostbite.
†For children younger than 7 yr, DTP (DT, if pertussis vaccine is contraindicated) is preferred to tetanus toxoid alone. For persons 7 yr of age or older, Td is preferred to tetanus toxoid alone.
‡If only three doses of fluid toxoid have been received, then a fourth dose of toxoid, preferably an adsorbed toxoid, should be given.
§Yes, if more than 10 yr since last dose.
‖Yes, if more than 5 yr since last dose.
DT—diphtheria and tetanus toxoids adsorbed (dose of diphtheria toxoid is higher than that in Td, dose of tetanus toxoid is the same)
DTP—diphtheria and tetanus toxoids and pertussis vaccine adsorbed Td—tetanus and diphtheria toxoids adsorbed TIG—tetanus immune globulin

Prevention

SURGICAL PROPHYLAXIS FOR POSTOPERATIVE WOUND INFECTIONS

Anaerobic infections can be prevented by avoiding conditions that predispose to tissue invasion by commensal microflora. In traumatic wounds, the most effective prophylaxis is thorough debridement and cleansing of the wound, elimination of foreign bodies and dead space, and the reestablishing of good circulation. Preoperative mechanical cleansing of the bowel with a low-residue or liquid diet followed by cathartics, enemas, and luminal antibiotics can reduce the incidence of postoperative wound infections after colorectal surgery. Parenteral perioperative antibiotics are used in gastrointestinal and gynecologic surgery when there is heavy contamination with normal microflora at the operative site; examples of such surgeries include elective colorectal surgery, cesarean section after premature rupture of membranes, vaginal hysterectomy in a premenopausal woman, and radical pelvic or head and neck surgery for malignancy. Several studies have shown significant reduction in the frequency of postoperative infections, from about 20% to 30% down to 4% to 8%. Cefoxitin is the agent of choice for prophylaxis of postsurgical intra-abdominal wound infections.[55] A first-generation cephalosporin, such as cefazolin, is as effective as some second- or third-generation cephalosporins. In surgery on contaminated or so-called dirty sites, early treatment rather than prophylaxis is essential for reducing the incidence of postoperative morbidity.

INFECTION CONTROL MEASURES FOR *C. DIFFICILE*–ASSOCIATED DIARRHEA

Enteric precautions, strict adherence to hand washing, and restrictions on the use of antibiotics are necessary to control the intrahospital spread of CDAD. The use of environmental disinfectants containing hypochlorite can reduce the incidence of *C. difficile* acquisition and the subsequent development of disease. However, antibiotic therapy for asymptomatic carriers is not beneficial.

WOUND CARE AND ACTIVE IMMUNIZATION FOR TETANUS

About two thirds of tetanus cases in the United States occur after puncture wounds, lacerations, and other penetrating trauma. Prompt and thorough wound debridement is of paramount importance in preventing trauma-induced tetanus. Prophylactic antibiotics do not reliably prevent the development of tetanus. The United States Public Health Service has issued specific recommendations for wound management and tetanus prophylaxis[56] [*see Table 8*]. Active immunization should be promoted for the general public. Infants should receive diphtheria-pertussis-tetanus (DPT) vaccine at 2 months of age. Two additional doses are given at 4 and 6 months of age. A booster injection of DPT is given at 18 months of age and again 4 years later. At 16 years of age, a booster dose of combined adult-type tetanus and diphtheria toxoids is administered. Thereafter, immunity is maintained by booster injections every decade. About 70% of Americans have protective levels of tetanus antibodies (> 0.15 IU/ml). Nearly 90% of children are protected, but the rate declines rapidly after 40 years of age, decreasing to less than 30% after 70 years of age. Adults who have not been immunized should receive two doses of alum-precipitated tetanus toxoid intramuscularly, 1 month apart, followed by a booster dose after 1 year. Thereafter, immunity is maintained by booster injections every decade.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Chow AW: Infections of the oral cavity, neck and head. Mandell GL, Bennett JE, Dolin R, Eds. Principles and Practice of Infectious Diseases, 6th ed. Churchill Livingstone, New York, 2005, p 787

2. Poxton IR, Brown R, Sawyerr A, et al: The mucosal anaerobic gram-negative bacteria of the human colon. Clin Infect Dis 25(suppl 2):S111, 1997

3. Hillier SL, Lau RJ: Vaginal microflora in postmenopausal women who have not received estrogen replacement therapy. Clin Infect Dis 25(suppl 2):S123, 1997

4. Goldstein EJ: Intra-abdominal anaerobic infections: bacteriology and therapeutic potential of newer antimicrobial carbapenem, fluoroquinolone, and desfluoroquinolone therapeutic agents. Clin Infect Dis 35:S106, 2002

5. Brook I: The role of encapsulated anaerobic bacteria in synergistic infections. FEMS Microbiol Rev 13:65, 1994

6. Brook I: Failure of penicillin to eradicate group A beta-hemolytic streptococci tonsillitis: causes and management. J Otolaryngol 30:324, 2001

7. Rotstein OD, Nasmith PE, Grinstein S: pH-dependent impairment of the neutrophil respiratory burst by the *Bacteroides* byproduct succinate. Clin Invest Med 11:259, 1988

8. Brook I: Beta-lactamase-producing bacteria in mixed infections. Clin Microbiol Infect 10:777, 2004

9. Hofstad T: Pathogenicity of anaerobic gram-negative rods: possible mechanisms. Rev Infect Dis 6:189, 1984

10. Bergan T: Pathogenicity of anaerobic bacteria. Scand J Gastroenterol Suppl 91:1, 1984

11. Stingele F, Corthesy B, Kusy N, et al: Zwitterionic polysaccharides stimulate T cells with no preferential V beta usage and promote anergy, resulting in protection against experimental abscess formation. J Immunol 172:1483, 2004

12. Gonzalez D, Tzianabos AO, Genco CA, et al: Immunization with *Porphyromonas gingivalis* capsular polysaccharide prevents *P. gingivalis*–elicited oral bone loss in a murine model. Infect Immun 71:2283, 2003

13. Hofstad T: Virulence factors in anaerobic bacteria. Eur J Clin Microbiol Infect Dis 11:1044, 1992

14. Poxton IR, Edmond DM: Biological activity of *Bacteroides* lipopolysaccharide: reappraisal. Clin Infect Dis 20(suppl 2):S149, 1995

15. Stehle HW, Leblebicioglu B, Walters JD: Short-chain carboxylic acids produced by gram-negative anaerobic bacteria can accelerate or delay polymorphonuclear leukocyte

apoptosis in vitro. J Periodontol 72:1059, 2001

16. Stevens DL, Bryant AE: The role of clostridial toxins in the pathogenesis of gas gangrene. Clin Infect Dis 35:S93, 2002

17. Blake JE, Mitsikosta F, Metcalfe MA: Immunological detection and cytotoxic properties of toxins from toxin A–positive, toxin B–positive *Clostridium difficile* variants. J Med Microbiol 53:197, 2004

18. Riegler M, Sedivy R, Pothoulakis C, et al: *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. J Clin Invest 95:2004, 1995

19. Salminen MK, Rautelin H, Tynkkynen S, et al: *Lactobacillus* bacteremia, clinical significance, and patient outcome, with special focus on probiotic *L. rhamnosus* GG. Clin Infect Dis 38:62, 2004

20. Srirompotong S, Art-Smart T: Ludwig's angina: a clinical review. Eur Arch Otorhinolaryngol 260:401, 2003

21. Weinstein WM, Onderdonk AB, Bartlett JG, et al: Experimental intra-abdominal abscesses in rats: development of an experimental model. Infect Immun 10:1250, 1974

22. Brook I: Microbiology and management of polymicrobial female genital tract infections in adolescents. J Pediatr Adolesc Gynecol 15:217, 2002

23. Salonen JH, Eerola E, Meurman O: Clinical significance and outcome of anaerobic bacteremia. Clin Infect Dis 26:1413, 1998

24. Goldstein EJ: Anaerobic bacteremia. Clin Infect Dis 23(suppl 1):S97, 1996

25. Fiorino AS: Intrauterine contraceptive device-associated actinomycotic abscess and *Actinomyces* detection on cervical smear. Obstet Gynecol 87:142, 1996

26. Kimura AC, Higa JI, Levin RM, et al: Outbreak of necrotizing fasciitis due to *Clostridium sordellii* among black-tar heroin users. Clin Infect Dis 38:87, 2004

27. Sinave C, Le Templier G, Blouin D, et al: Toxic shock syndrome due to *Clostridium sordellii*: a dramatic postpartum and postabortion disease. Clin Infect Dis 35:1441, 2002

28. Olsen SJ, MacKinnon LC, Goulding JS, et al: Surveillance for foodborne-disease outbreaks—United States, 1993–1997. MMWR CDC Surveill Summ 49:1, 2000

29. Petrillo TM, Beck-Sague CM, Songer JG, et al: Enteritis necroticans (pigbel) in a diabetic child. N Engl J Med 342:1250, 2000

30. Mylonakis E, Ryan ET, Calderwood SB: *Clostridium difficile*–associated diarrhea: a review. Arch Intern Med 161:525, 2001

31. Dial S, Alrasadi K, Manoukian C, et al: Risk of *Clostridium difficile* diarrhea among hospital inpatients prescribed proton pump inhibitors: cohort and case-control studies. CMAJ 171:33, 2004

32. Pascual FB, McGinley EL, Zanardi LR, et al: Tetanus surveillance—United States, 1998–2000. MMWR Surveill Summ 52:1, 2003

33. Sobel J, Tucker N, Sulka A, et al: Foodborne botulism in the United States, 1990–2000. Emerg Infect Dis 10:1606, 2004

34. Jousimies-Somer H, Summanen P: Recent taxonomic changes and terminology update of clinically significant anaerobic gram-negative bacteria (excluding spirochetes). Clin Infect Dis 35: S17, 2002

35. Drancourt M, Berger P, Raoult D: Systematic 16S rRNA gene sequencing of atypical clinical isolates identified 27 new bacterial species associated with humans. J Clin Microbiol 42:2197, 2004

36. Rautio M, Saxen H, Siitonen A, et al: Bacteriology of histopathologically defined appendicitis in children. Pediatr Infect Dis J 19:1078, 2000

37. Finegold SM, Jousimies-Somer H: Recently described clinically important anaerobic bacteria: medical aspects. Clin Infect Dis 25(suppl 2):S88, 1997

38. Citron DM, Warren YA, Hudspeth MK, et al: Survival of aerobic and anaerobic bacteria in purulent clinical specimens maintained in the Copan Venturi Transystem and Becton Dickinson Port-a-Cul transport systems. J Clin Microbiol 38:892, 2000

39. Zheng L, Keller SF, Lyerly DM, et al: Multicenter evaluation of a new screening test that detects *Clostridium difficile* in fecal specimens. J Clin Microbiol 42:3837, 2004

40. Hecht DW: Evolution of anaerobe susceptibility testing in the United States. Clin Infect Dis 35:S28, 2002

41. Hecht DW: Prevalence of antibiotic resistance in anaerobic bacteria: worrisome developments. Clin Infect Dis 39:92, 2004

42. Brook I: Antibiotic resistance of oral anaerobic bacteria and their effect on the management of upper respiratory tract and head and neck infections. Semin Respir Infect 17:195, 2002

43. Chan Y, Chan CH: Antibiotic resistance of pathogenic bacteria from odontogenic infections in Taiwan. J Microbiol Immunol Infect 36:105, 2003

44. Goldstein EJ, Citron DM, Merriam CV, et al: Comparative in vitro activities of ertapenem (MK-0826) against 469 less frequently identified anaerobes isolated from human infections. Antimicrob Agents Chemother 46:1136, 2002

45. Solomkin JS, Mazuski JE, Baron EJ, et al: Guidelines for the selection of anti-infective agents for complicated intra-abdominal infections. Clin Infect Dis 37:997, 2003

46. Snydman DR, Jacobus NV, McDermott LA, et al: National survey on the susceptibility of *Bacteroides Fragilis* Group: report and analysis of trends for 1997–2000. Clin Infect Dis 35:S126, 2002

47. Edmiston CE, Krepel CJ, Seabrook GR, et al: In vitro activities of moxifloxacin against 900 aerobic and anaerobic surgical isolates from patients with intra-abdominal and diabetic foot infections. Antimicrob Agents Chemother 48:1012, 2004

48. Goldstein EJ: Review of the in vitro activity of gemifloxacin against gram-positive and gram-negative anaerobic pathogens. J Antimicrob Chemother 45(suppl 1):55, 2000

49. Rotstein OD: Interactions between leukocytes and anaerobic bacteria in polymicrobial surgical infections. Clin Infect Dis 16(suppl 4):S1904, 1993

50. Tanaka-Bandoh K, Watanabe K, Kato N, et al: Susceptibilities of *Actinomyces* species and *Propionibacterium propionicus* to antimicrobial agents. Clin Infect Dis 25:S262, 1997

51. Persky SE, Brandt LJ: Treatment of recurrent *Clostridium difficile*–associated diarrhea by administration of donated stool directly through a colonoscope. Am J Gastroenterol 95:3283, 2000

52. McFarland LV, Elmer GW, Surawicz CM: Breaking the cycle: treatment strategies for 163 cases of recurrent *Clostridium difficile* disease. Am J Gastroenterol 97:1769, 2002

53. McFarland LV, Surawicz CM, Greenberg RN, et al: A randomized placebo-controlled trial of *Saccharomyces boulardii* in combination with standard antibiotics for *Clostridium difficile* disease. JAMA 271:1913, 1994

54. Thomas MR, Litin SC, Osmon DR, et al: Lack of effect of *Lactobacillus* GG on antibiotic-associated diarrhea: a randomized, placebo-controlled trial. Mayo Clin Proc 76:883, 2001

55. de Lalla F: Surgical prophylaxis in practice. J Hosp Infect 50(suppl A):S9, 2002

56. Recommended adult immunization schedule—United States, 2002–2003. MMWR Morb Mortal Wkly Rep 51:904, 2002

57. Dahlen G: Microbiology and treatment of dental abscesses and periodontal-endodontic lesions. Periodontol 2000 28:206, 2002

58. Booth V, Downes J, Van den BJ, et al: Gram-positive anaerobic bacilli in human periodontal disease. J Periodontal Res 39:213, 2004

59. Swift JQ, Gulden WS: Antibiotic therapy—managing odontogenic infections. Dent Clin North Am 46:623, 2002

60. Schraff S, McGinn JD, Derkay CS: Peritonsillar abscess in children: a 10-year review of diagnosis and management. Int J Pediatr Otorhinolaryngol 57:213, 2001

61. Finegold SM, Flynn MJ, Rose FV, et al: Bacteriologic findings associated with chronic bacterial maxillary sinusitis in adults. Clin Infect Dis 35:428, 2002

62. Brook I, Yocum P, Shah K: Aerobic and anaerobic bacteriology of concurrent chronic otitis media with effusion and chronic sinusitis in children. Arch Otolaryngol Head Neck Surg 126:174, 2000

63. Le Moal G, Landron C, Grollier G, et al: Characteristics of brain abscess with isolation of anaerobic bacteria. Scand J Infect Dis 35:318, 2003

64. Bernardini GL: Diagnosis and management of brain abscess and subdural empyema. Curr Neurol Neurosci Rep 4:448, 2004

65. El Solh AA, Pietrantoni C, Bhat A, et al: Microbiology of severe aspiration pneumonia in institutionalized elderly. Am J Respir Crit Care Med 167:1650, 2003

66. Mansharamani NG, Koziel H: Chronic lung sepsis: lung abscess, bronchiectasis, and empyema. Curr Opin Pulm Med 9:181, 2003

67. Verma P: Laboratory diagnosis of anaerobic pleuropulmonary infections. Semin Respir Infect 15:114, 2000

68. Mwandumba HC, Beeching NJ: Pyogenic lung infections: factors for predicting clinical outcome of lung abscess and thoracic empyema. Curr Opin Pulm Med 6:234, 2000

69. Morrissey BM, Evans SJ: Severe bronchiectasis. Clin Rev Allergy Immunol 25:233, 2003

70. Robert R, Grollier G, Frat JP, et al: Colonization of lower respiratory tract with anaerobic bacteria in mechanically ventilated patients. Intensive Care Med 29:1062, 2003

71. Rahimian J, Wilson T, Oram V, et al: Pyogenic liver abscess: recent trends in etiology and mortality. Clin Infect Dis 39:1654, 2004

72. Hebb JK, Cohen CR, Astete SG, et al: Detection of novel organisms associated with salpingitis, by use of 16S rDNA polymerase chain reaction. J Infect Dis 190:2109, 2004

73. Aimakhu CO, Olayemi O, Odukogbe AA: Microbiological pattern of the contents of pelvic abscess at the University College Hospital, Ibadan, Nigeria. J Obstet Gynaecol 23:67, 2003

74. Canas AM, Holloran-Schwartz B, Myles T: Tuboovarian abscess 12 years after total abdominal hysterectomy. Obstet Gynecol 104:1039, 2004

75. French L, Smaill F: Antibiotic regimens for endometritis after delivery. Cochrane Database Syst Rev (4):CD001067, 2004

76. Feingold DS: Gangrenous and crepitant cellulitis. J Am Acad Dermatol 6:289, 1982

77. Pizzo LJ: Synergistic necrotizing cellulitis of the head and neck. Am J Otolaryngol 3:452, 1982

78. Elliott D, Kufera JA, Myers RA: The microbiology of necrotizing soft tissue infections. Am J Surg 179:361, 2000

79. Livesley NJ, Chow AW: Infected pressure ulcers in elderly individuals. Clin Infect Dis 35:1390, 2002

80. Marrie TJ, Aylward D, Kerr E, et al: Bacteriology of pilonidal cyst abscesses. J Clin Pathol 31:909, 1978

81. Garcea G, Lloyd T, Jacobs M, et al: Role of microbiological investigations in the management of non-perineal cutaneous abscesses. Postgrad Med J 79:519, 2003

82. Gerding DN: Foot infections in diabetic patients: the role of anaerobes. Clin Infect Dis 20(suppl 2):S283, 1995

83. Edmiston CE Jr, Walker AP, Krepel CJ, et al: The nonpuerperal breast infection: aerobic and anaerobic microbial recovery from acute and chronic disease. J Infect Dis 162:695, 1990

84. Talan DA, Abrahamian FM, Moran GJ, et al: Clinical presentation and bacteriologic analysis of infected human bites in patients presenting to emergency departments. Clin Infect Dis 37:1481, 2003

85. Chow AW: Infections caused by *Bacteroides* and other mixed nonsporulating anaerobes. Internal Medicine, 5th ed, Stein JH, Ed. Lippincott Williams & Wilkins, Philadelphia, 1998, p 1613

86. Bartlett JG: Anaerobic bacteria. Infectious Diseases, 3rd ed. Gorbach SL, Bartlett JG, Blacklow NR, Eds. Lippincott Williams & Wilkins, Philadelphia, 2004, p 1768

Acknowledgment

Figure 3 Reproduced with permission from Finegold SM, Sutter VL: Anaerobic Infections, 4th ed. Scope Publications, Kalamazoo, Michigan, 1982

VI SYPHILIS AND THE NONVENEREAL TREPONEMATOSES

Michael Augenbraun, m.d.

Syphilis

Syphilis is an infectious disease with complex acute and chronic manifestations that is transmitted primarily through sexual contact. The disease has been recognized for many centuries, although its origin remains unknown.

epidemiology

Transmissible syphilis in the United States has become increasingly concentrated in a few geographic areas, particularly in the southeastern part of the country [*see Figure 1*]. It disproportionately affects minority populations and occurs more frequently in men than in women. Recent outbreaks have been reported in homosexual men. Sexual transmission rates from infected to uninfected persons may range from 30% to 60%.[1-3]

The incidence of syphilis was declining from the 1940s through the 1980s, but a sharp and surprising increase occurred in the late 1980s. By 1990, the Centers for Disease Control and Prevention (CDC) reported the largest number of primary and secondary cases of syphilis in 40 years.[4] This increase was strongly linked to the epidemic use of crack cocaine. With the increased incidence of primary syphilis came an increase in the numbers of cases of congenital syphilis.[5]

Just as surprising as the increase in the number of syphilis cases in the late 1980s has been the rapid decline in the number of syphilis cases in the latter half of the 1990s.[6] The reasons for this decline are unclear, but by 1999 there had been an 87% decline from the peak number of cases in 1990. This was the lowest reported rate since the collection of reliable statistics began [*see Figure 2*].

As a result of the low incidence rates of primary and secondary syphilis, its geographic confinement to a limited number of areas, the continued efficacy of penicillin therapy, and the reliability of serologic diagnostic tests, the CDC has initiated a program intended to eliminate syphilis in the United States.[7] The stated goal is to reduce primary and secondary cases to 1,000 or fewer in a given year and to increase the number of syphilis-free counties to 90% by the year 2005. The CDC hopes to achieve this goal through the use of enhanced surveillance, greater cooperation with community groups in the remaining areas of high morbidity, and rapid outbreak response.

etiology

Syphilis is transmitted primarily through sexual contact between infected and uninfected partners. Passage through the placenta in an infected pregnant woman may result in congenital disease. Although the disease can also be transmitted through nonsexual contact with infectious lesions, laboratory accidents, and the administration of contaminated blood products, these methods of transmission are relatively rare. Humans are the only known host of *Treponema pallidum*.

pathogenesis and disease course

T. pallidum is a bacterium belonging to the order Spirochaetales. This order also includes similar familiar bacteria genera, such as *Leptospira* and *Borrelia*. All of these organisms are slender, helically coiled, gram-negative bacteria. Treponemal species other than *T. pallidum* may live as commensals in the oral cavity or the genital tract (e.g., *T. denticola* and *T. oralis*).

T. pallidum probably gains entry to the subcutaneous tissues through small abrasions in overlying skin and mucous membranes. The organism does not possess well-recognized virulence factors, such as lipopolysaccharide. Once penetration of the epithelium occurs, organisms replicate locally at the site of inoculation and spread to regional lymph nodes. Depending on factors such as inoculation load and history of previous syphilis infection, the incubation period from exposure to clinical disease can range from 3 weeks to 3 months.

The characteristic first lesion of primary syphilis [*see Table 1*] is the chancre [*see Figure 3*]. It is a typically indurated and nontender ulcerative lesion, often accompanied by a nontender, nonsuppurative regional lymphadenopathy. In some patients, the chancre is inconspicuous, and in others it can be more than 1 to 2 cm in diameter. Although not common, multiple chancres may occur, particularly in HIV-infected persons. In the absence of effective therapy, the chancre resolves within 2 to 8 weeks. Healing may take longer in immunocompromised persons.

Spirochetemia develops soon after the appearance of the primary chancre, but clinical evidence of dissemination is rarely seen at the same time. Signs and symptoms of dissemination and secondary syphilis include fever, malaise, diffuse lymphadenopathy, patchy alopecia, headache, and classic hyperpigmented macular-papular rash on the palms and soles [*see Figure 4*]. Condylomata lata and mucous patches may appear in a variety of locations, such as the genitals and the gluteal and nasolabial folds [*see Figure 5*]. Other anatomic sites, such as the eyes, gastrointestinal mucosa, liver, and bones, may also be involved.

The clinical manifestations of secondary syphilis [*see Table 1*] resolve after weeks to months even without therapy, and the disease enters the latent stage. During the early latency period, the signs and symptoms of secondary syphilis may recrudesce. This period generally lasts 1 year after infection but can extend to 4 or 5 years.[8]

Latent syphilis is, by definition, recognized only by the presence of a reactive serologic test. During this period, the organisms evade any meaningful host immunologic responses. Although infected persons mount both significant cell-mediated and humoral responses to infection, neither response is adequate for control. Many reasons have been postulated for this phenomenon, including the lack of immunogenic proteins and polysaccharides in the outer membrane of *T. pallidum,* as well as the organism's ability to cloak itself with host proteins.[9]

Once syphilis enters the latent phase, the course of disease is variable. Although the majority of infected persons will never suffer undue consequences, data from the Oslo[8] study suggests that years later, tertiary disease [*see Table 1*] will develop in approximately one third of these individuals, manifesting as neurosyphilis, cardiovascular syphilis, or gummatous (benign) syphilis. The pathology associated with tertiary syphilis is quite varied, and there is substantial overlap between clinical entities. Chronologically, meningeal and meningovascular

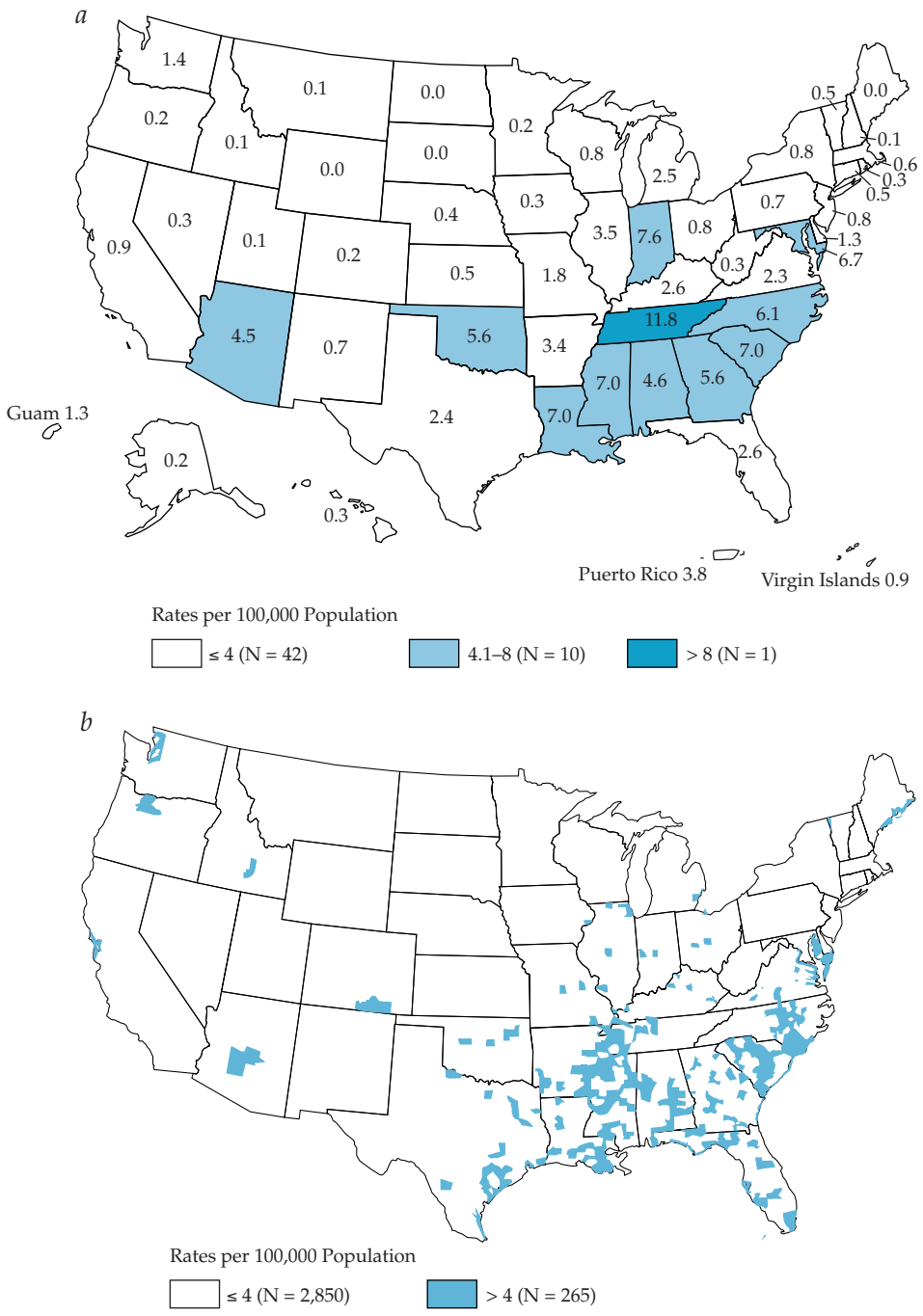


Figure 1 **Rates of primary and secondary syphilis by (*a*) state and (*b*) county.**

symptoms (i.e., stroke) may be the earliest manifestations of symptomatic neurosyphilis. In later stages, parenchymatous disease with associated cognitive and long-tract signs becomes prominent. As in cardiovascular and gummatous syphilis, the lesion in CNS syphilis is endarteritis.

DIAGNOSIS

Depending on the stage of disease, the diagnosis of syphilis may be made by visualizing the organisms, by serologic studies, by pathology, or by clinical presentation. None of these methods is entirely sensitive or specific.

Clinical Manifestations

Primary syphilis In primary syphilis, the chancre [*see* Pathogenesis, *above*] typically occurs at the site of inoculation and thus may be seen anywhere in the genital region (i.e., penis, scrotum, inner thigh, buttocks, vaginal labia, cervix, or anus). Lesions may be single or multiple. Often, they occur where tissue proximates infected sites—so-called kissing lesions.

Typically, the chancre is indurated with a raised but clean margin. It is nontender. It may have a gray-white exudative covering, particularly if the lesion is secondarily infected. More commonly, it is clean without bleeding. There is associated

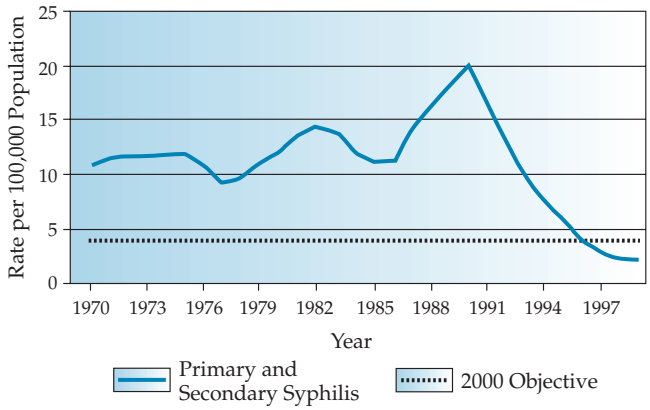


Figure 2 **Reported rates of primary syphilis in the United States over 3 decades. The highest rate occurred in the late 1980s and early 1990s, after which there has been a steep decline in the number of syphilis cases.**

nontender and nonsuppurative regional lymphadenopathy. Cervical or rectal chancres may result in deep pelvic noninguinal adenopathy.

Secondary syphilis The widely varied clinical manifestations of secondary syphilis reflect the capacity of *T. pallidum* to disseminate virtually anywhere in the body within weeks of the appearance (and resolution) of the primary chancre. With spirochetemia, which develops after the first chancre [*see* Pathogenesis, *above*], comes a host of systemic symptoms that are relatively nonspecific, including headache, malaise, fatigue, arthralgias, myalgias, and diffuse painless lymphadenopathy. The finding of enlarged epitrochlear lymph nodes in the absence of any upper extremity pathology is considered to be highly suspicious for secondary syphilis.

The rash of secondary syphilis is one of the most common signs of the disease. The classic finding is maculopapular hyperpigmented lesions 3 to 10 mm in diameter, often with fine peripheral scaling on the palms and soles. There are so few clinical entities that cause these lesions that the presence of the lesions is nearly pathognomonic. Rash will be seen in 75% to 100% of patients at some point in the course of infection. However, the rash of secondary syphilis does not always fit this classic pattern. It may present in any of a variety of ways—macular, maculopapular, papular, pustular, or plaques of varying size—on virtually any site of the body. Raised, moist, nontender plaques known as condylomata lata may form in intertriginous areas, such as in the gluteal fold, under the breasts, and in the inguinal area. Similar mucous patches may form on

Table 1 Clinical Manifestations of Syphilis in Adults

Stage	*Manifestations*
Primary	Chancre and inguinal lymphadenopathy
Secondary	Rash, diffuse lymphadenopathy, fever, alopecia, condyloma latum, and mucous patches
Early latent	None (secondary syphilis may recur)
Late latent or latent syphilis of unknown duration	None
Tertiary	
Neurosyphilis	
Asymptomatic	None
Meningitis	Headache, cranial nerve palsies, delirium, seizures, and findings of raised intracranial pressure
Meningovascular	Hemiparesis, hemiplegia, aphasia, and seizures
Parenchymatous	
General paresis	Memory loss, personality changes, cognitive dysfunction, confusion, and seizures
Tabes dorsalis	Lightning pains, ataxia, pupillary abnormalities, and impaired proprioception
Cardiovascular	Aortic aneurysm, aortic regurgitation, chest wall mass, and hoarseness
Gummatous	Chronic inflammation with focal destructive lesions in nearly any tissue or organ of the body

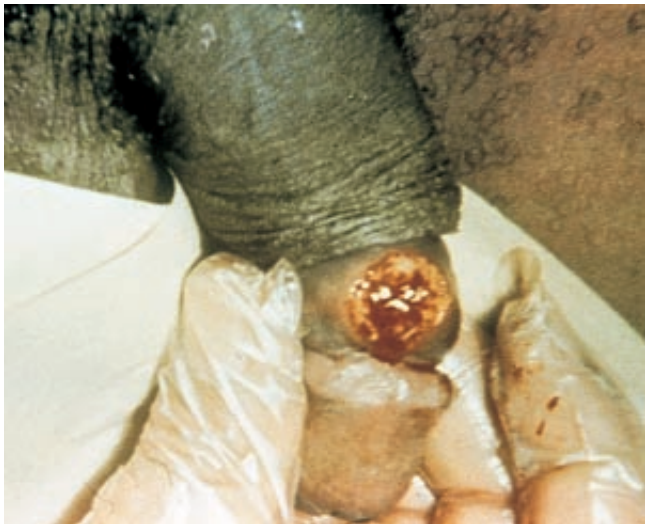


Figure 3 **Indurated, nontender ulcerative chancres are the first lesions of early syphilis.**

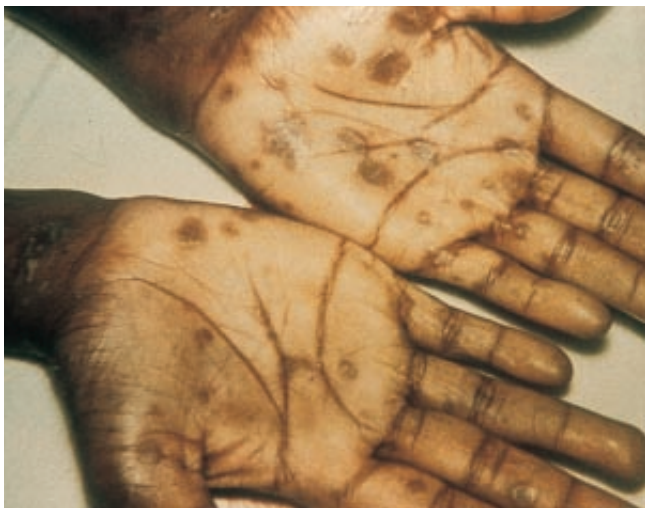


Figure 4 **A hyperpigmented maculopapular rash on the palms and soles is a classic sign of spirochetemia.**

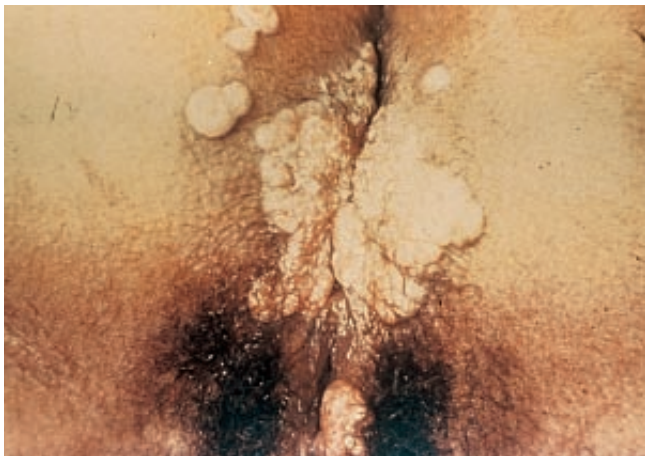


Figure 5 **Highly infectious condylomata lata appear on genitalia and gluteal folds.**

mucosal surfaces, such as the nares, mouth, vagina, and anus. These lesions can be confused with chancres but are not, strictly speaking, ulcers. They teem with spirochetes and are highly infectious. Secondary syphilis can also involve hair follicles, leading to patchy alopecia of scalp hair, as well as thinning of the eyebrows, eyelashes, and beard.

Skeletal manifestations of secondary syphilis can include osteitis, arthritis, and bursitis.[10] Syphilitic hepatitis with hepatomegaly and an elevated alkaline phosphatase level may occur.[11] Various forms of nephropathy, including glomerulonephritis and nephrotic syndrome, have also been noted in secondary syphilis.[12] Gastritis with attendant abdominal pain, nausea, and vomiting can result from infiltration of the gastric mucosa by spirochetes.[13] Both anterior and posterior uveitis can occur and may be either asymptomatic or associated with altered vision.[14, 15] Some of these cases come to light after failure to respond or after worsening of symptoms with steroids.

Involvement of the CNS occurs early in syphilis.[16] At least 40% of lumbar puncture samples from patients with secondary disease (more in some series) have been found to have CSF abnormalities. *T. pallidum* may actually be recovered from up to 30% of these samples. Whereas symptomatic parenchymatous neurosyphilis is often associated with late-stage syphilis, meningeal involvement may be seen during or shortly after the nonneurologic symptoms of secondary syphilis. This condition is characterized by findings typical of a bacterial meningitis, including headache, stiff neck, photophobia, nausea, and vomiting. Cranial neuropathies associated with ocular and otic deficits, as well as facial nerve palsies, can occur. In some instances, papilledema associated with hydrocephalus has been reported, as have delirium, seizures, aphasia, and hemiplegia. Most patients no longer demonstrate the rash of secondary syphilis at the time these symptoms develop.

Latent syphilis By definition, latent syphilis is not associated with any clinical finding. Early latency delineates that period after the resolution of secondary disease during which the signs and symptoms of secondary disease may recur. Although this latency period may last as long as 5 years, the majority of cases of latent syphilis occur within the first to second year, followed by tertiary syphilis.

Neurosyphilis It is presumed that in the majority of patients with primary or secondary syphilis who experience invasion of the CNS by spirochetes, resolution of this aspect of infection occurs with or without the administration of systemic antibiotics. A small number of these patients have continuing CNS involvement with asymptomatic CSF changes and are at high risk for progressive symptomatic neurologic disease for a period of years. It is difficult to ascertain which individuals may be at risk, and anyone with presumably latent syphilis may harbor CSF changes suggesting continued neurosyphilis.

Late symptomatic neurosyphilis may present as one of several clinical syndromes that occur within the following conditions (the syndromes often overlap):

1. Meningovascular neurosyphilis. *T. pallidum* tends to cause endarteritis, which can compromise vascular supply to portions of the CNS and cause infarctions that clinically resemble any other cerebrovascular accident. These events can occur anywhere along the neuroaxis, including the cerebrum, the cerebellum, the brain stem, and the spinal cord. Signs and symptoms therefore correspond to the portion of the CNS involved. The time elapsed from primary syphilis to meningovascular syphilis is usually 5 to 12 years but may be as long as several decades.
2. Parenchymatous neurosyphilis. Direct involvement of the CNS parenchyma results in a wide range of clinical syndromes characterized pathologically by fibrosis and atrophy. General paresis is a meningoencephalitic syndrome that may be similar in presentation to many psychiatric and neurologic disorders. Patients exhibit emotional lability, delusions, paranoia, and memory loss and may progress to dementia, delirium, and seizures. They may have difficulty with speech, and their pupils may become constricted and unresponsive to light and painful stimuli (Argyll Robertson pupil). They may lose the capacity for facial expressions and are noted to have trembling of the lips and tongue. In tabes dorsalis, there is a pronounced involvement of the dorsal roots and the posterior columns of the spinal cord. Patients complain of shooting pains in the back and lower extremities, where they also notice paresthesias. These paroxysms of pain can be referred to other locations, including the abdomen and the throat. Patients lose proprioception and vibratory sense and develop a wide-based gait. Optic atrophy may occur. Knee and ankle reflexes are diminished or lost entirely, plantar responses are flexor, and ataxia is usually present. Symptomatic parenchymatous disease develops 15 to 25 years after primary infection. The long-term outlook in general paresis without treatment is grim. Death may occur within months or, more likely, 4 to 5 years after the onset of symptoms. Untreated tabes dorsalis leads to eventual incapacitation, although it has been known to remit spontaneously.

Cardiovascular syphilis After the introduction of antibiotics, cardiovascular syphilis became extremely rare. *T. pallidum* can cause endarteritis of the aortic vasa vasorum, leading to progressive medial necrosis and loss of elastic tissue, dilatation, and, eventually, aneurysm. The proximal thoracic aorta and the aortic arch are most commonly involved, whereas involvement of the abdominal aorta is extremely rare. Symptoms usually develop when the aneurysm begins to encroach upon or erode adjacent structures, such as the chest wall, the superior vena cava, the recurrent laryngeal nerve, the trachea, and the

mainstem bronchi. These aneurysms rarely dissect. In 30% of cases, aortic root dilatation leads to aortic regurgitation in which there is no evidence of aortic stenosis. Uncommonly, the coronary artery ostia may be involved, causing ischemic heart disease. Cardiovascular syphilis is usually preceded by a latent period of 15 to 30 years.

Gummatous (benign) syphilis As with neurosyphilis and cardiovascular syphilis, the incidence of late gummatous syphilis has declined dramatically since the introduction of effective antimicrobial agents. The hallmark of this disease are gummas, which are indolent, destructive granulomatous lesions of soft tissue and bone [*see Figure 6*]. They can lead to dramatic scarring and disfigurement. They are primarily immunologically mediated responses. Although almost any part of the body may be involved, gummas most frequently occur in soft tissue and skeletal structures. Gummas do not develop until decades after infection. They appear slowly, often as a nodule with minimal attendant inflammation. Central coagulative necrosis develops, with softening and involution of the lesion. Eventually, fibrosis occurs, with deep scarring. Involvement of the nasal septum can lead to collapse of the nasal structures. Perforation of the hard palate may occur. Skeletal lesions can lead to fractures.

Laboratory Tests

Primary syphilis Clinical appearance of a genital ulcer is not adequate to establish its etiology. Dark-field microscopy of lesion exudate is a reliable diagnostic method. Incident light through a microscope fitted with polarizing lenses allows the viewer to identify the corkscrew morphology of treponemes as white against a black background [*see Figure 7*]. Exudate for dark-field microscopy can be collected from the base of a chancre that has been cleaned with saline and to which gentle pressure has been applied. Examination must be performed promptly because any drying of the specimen reduces the sensitivity of the test. Therapy, intentional or not, with antibiotics active against *T. pallidum* will rapidly reduce the yield of a dark-field specimen. Nonpathogenic treponemes in the oral cavity reduce the specificity of dark-field microscopy on lesions from this area. A direct fluorescent antibody test (DFA-TP) can also be performed on lesion exudate when the immediate collection of a specimen is not possible. Nucleic acid amplification tests have been developed for *T. pallidum,* but these tests should be considered research tools.

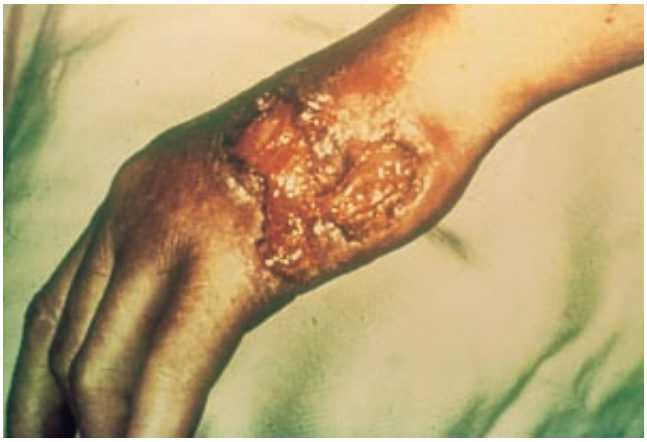


Figure 6 **Granulomatous lesions called gummas are particularly destructive to soft tissue and bone.**

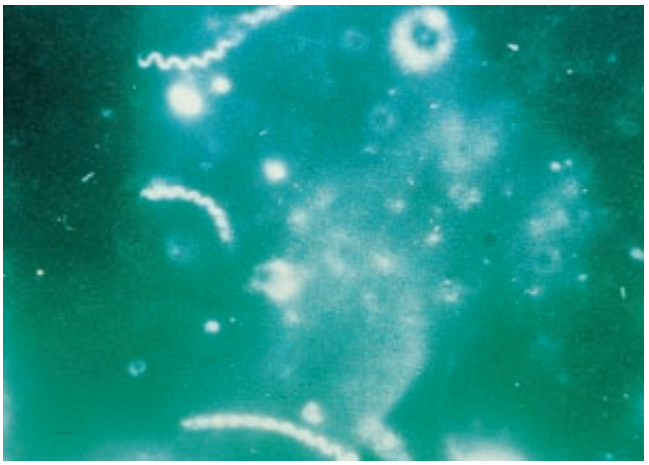


Figure 7 **In dark-field microscopy, treponemes appear white against a black background.**

Besides dark-field microscopy, the most commonly used tests for primary syphilis are serologic tests. Syphilis serology testing requires the use of two tests performed in sequence—one for screening and the other for confirmation. Screening tests are characterized by high sensitivity but variable specificity. They make use of standardized amounts of cardiolipin, cholesterol, and lecithin, to which *T. pallidum*–directed immunoglobulins (IgG or IgM) have an affinity. These tests include the Venereal Disease Research Laboratory (VDRL) test, rapid plasma reagin (RPR) test, unheated serum reagin (USR) test, and toluidine red unheated serum test (TRUST). All are flocculation tests in which visible (macroscopic, microscopic, or both) clumping occurs when test reagents are added to serum from a patient with syphilis. Results may be reported qualitatively or quantitatively. For quantitative results, testing is performed on sequentially diluted serum. The final dilution in which a reaction occurs is reported. In primary syphilis, the sensitivity of any of these nontreponemal serologic tests (NTSTs) ranges only from 78% to 86%. They may also be erroneously reactive in association with a number of other clinical factors, including older age, pregnancy, a history of intravenous drug use, and collagen vascular diseases.[17]

Treponemal-specific tests such as the fluorescent treponemal antibody absorption (FTA-ABS) and the *T. pallidum* particle agglutination (TP-PA) tests use fluorescent antibody and hemagglutinating antibody, respectively, to detect patients' antibodies to *T. pallidum*. In these cases, a suspension of *T. pallidum* serves as the antigenic target. These tests are not routinely quantified. Depending on the strength of the reaction, they may be reported as reactive, nonreactive, or inconclusive. In primary syphilis, these tests can become reactive before the NTSTs. False positive results occur in a small percentage of those tested, so these tests should not by themselves be used for screening. They are best used to confirm a reactive NTST.

Secondary syphilis Dark-field microscopy may be useful for examining exudate from condylomata lata and mucous patches, but ordinarily, this test has limited utility in the diagnosis of secondary syphilis. Serologic tests remain the standard for diagnosing syphilis at this stage. Because these tests quanti-

tatively reflect the degree of immune response associated with infection, the highest NTST titers are generally seen during secondary syphilis. The likelihood of a false negative test in secondary syphilis is remote. An exception to this occurs in the event of the so-called prozone phenomenon, in which extremely high concentrations of reaginic antibodies block the flocculation process. This effect may be expected in no more than 2% of cases and can be overcome by performing repeated tests on diluted samples.

CSF findings in meningitis associated with secondary syphilis are nonspecific, except with use of the CSF-VDRL. A positive finding is invariably indicative of neurosyphilis unless there has been significant contamination of the CSF with blood. Sensitivity of the CSF-VDRL ranges only from 30% to 70%, so a negative test does not rule out neurosyphilis.[18] Other abnormalities of the CSF that may occur with or without a reactive CSF-VDRL include a mononuclear pleocytosis with counts of 10 to 500 cells/mm^3 and an elevated protein level of 46 to 200 mg/dl.

Latent syphilis As disease activity wanes and immunologic mechanisms exert some control, the height of titer in the NTST declines. Titers of between 1:1 and 1:16 are not uncommon. Higher titers suggest but do not confirm the presence of disease activity. The treponemal serologic tests are also reactive.

Neurosyphilis The laboratory diagnosis of neurosyphilis depends on the analysis of serologic tests and the CSF examination. The treponemal serologic test is almost always positive in neurosyphilis but it is not a specific marker for disease, because it is generally reactive for the life of the infected individual, whether treated or not. The NTSTs, generally regarded as sensitive but nonspecific in the evaluation of primary syphilis, are nonreactive in 25% to 30% of cases of neurosyphilis.[19]

CSF evaluation remains the gold standard by which neurosyphilis is diagnosed. Because lumbar puncture is associated with some morbidity, efforts have been made to define subgroups of patients with latent syphilis (i.e., asymptomatic with serologic evidence of disease) who may be at risk for asymptomatic neurosyphilis and therefore at a higher risk for progression to symptomatic late neurosyphilis. Current guidelines suggest that patients with latent syphilis in the setting of HIV and patients with latent syphilis whose titers rise or fail to decline after 2 to 3 years of therapy are at risk.[20] Some experts recommend CSF evaluation in latent syphilis if the serologic titer is at least 1:32.

CSF analysis is warranted in any patient with serologic evidence of syphilis and neurologic, ophthalmologic, otic, or psychiatric disease. Tests conducted on the CSF should include a quantitative CSF-VDRL, a cell count with differential, and a protein assay. The VDRL is the only nontreponemal test that is standardized for CSF. In 25% to 30% of neurosyphilis cases, this test may be nonreactive. It is otherwise extremely specific and by itself may be diagnostic of neurosyphilis. The CSF cell count can range from 5 to 200 cells/mm^3, with a lymphocytic predominance. The CSF protein concentration is also usually increased, although levels should not exceed 200 g/dl. Lymphocytic pleocytosis and increased CSF protein concentrations are nonspecific for neurosyphilis and may be seen in a wide range of pathologic processes. The diagnosis of neurosyphilis in which the results of only these tests are abnormal rests on these findings and clinical suspicion. In some instances when tabes dorsalis has run its course, the CSF is completely normal.

Cardiovascular syphilis As with other forms of tertiary syphilis, the nontreponemal serologic test for late cardiovascular syphilis may be nonreactive, whereas the treponemal-specific test is almost always reactive. Although a chest x-ray may be normal, it can also show a mediastinal mass consistent with the thoracic aortic aneurysm. Calcification of the aneurysm is often seen on radiograph but is not specific for syphilis.

Gummatous syphilis Clinical suspicion is the most important tool in the diagnosis of late gummatous syphilis. Nontreponemal serologic tests are nonreactive in a sizable number of cases. The treponemal serologic tests should always be reactive. Interestingly, spirochetes are not often demonstrated in biopsied tissue. Dark-field microscopy is of little use. More sensitive tests based on nucleic acid amplification may detect treponemal DNA in these lesions.[21]

DIFFERENTIAL DIAGNOSIS

Primary Syphilis

Any clinical condition likely to cause a genital ulcer should be considered in the differential diagnosis when primary syphilis is suspected. The most common cause of genital ulcer disease in the United States is herpes simplex virus type 2 (HSV-2). The ulcers of HSV-2 are usually preceded by prodromal symptoms of burning and tingling, and they pass through a series of predictable stages, including papules, vesicles, pustules, and ulcers, before healing. The ulcers are tender and shallow, with an erythematous base. Lesions are often multiple and can coalesce to form large irregularly shaped ulcers. The course of illness varies, depending on whether it is a primary infection or a recurrence of established infection. The former may last as long as 3 weeks, whereas the latter usually resolves after 10 days. Diagnosis may be made on clinical grounds or with viral culture. Regional lymphadenopathy is common. Worldwide, the other common infectious cause of genital ulcer is chancroid, caused by *Haemophilus ducreyi* [*see 7:XXII Sexually Transmitted Diseases*]. Chancroid is seen sporadically in the United States, usually in conjunction with focal urban outbreaks, but is a common cause of genital ulcer disease in developing countries. Chancroid lesions are painful and indurated, with ragged margins. Gram stain of exudate may reveal gram-variable rods that characteristically line up together in patterns referred to as a school of fish or railroad tracks. Special nutrient media with antibiotic supplements should be used for isolation of *H. ducreyi.*

Secondary Syphilis

It would be nearly impossible to list all the conditions to which secondary syphilis bears some resemblance. A rash of any description, whose etiology may be construed as allergic, infectious, or immune mediated may potentially be syphilitic. Thus, syphilis testing for any patient with a rash of unknown etiology is warranted. The finding of a palmar rash or a rash on the soles of the feet—classic for syphilis—may be seen in a limited number of conditions, including Rocky Mountain spotted fever, atypical measles, and meningococcemia. A rash accompanying any of these alternative diagnoses would ordinarily be associated with more severe systemic illness than is seen with secondary syphilis. Syphilis needs to be considered in the differential diagnosis for aseptic meningitis. More typical causes for this problem may include enteroviruses, HSV-1, and medication-associated aseptic meningitis.

Latent Syphilis

In 1% to 2% of the general population, nontreponemal serologic tests such as the RPR, VDRL, and USR may be falsely reactive. In these circumstances, the treponemal serologic tests, such as the FTA-ABS and TP-PA, are nonreactive. An increased risk of this finding is commonly associated with older age, pregnancy, collagen vascular disease, and intravenous drug use.

Neurosyphilis

Although the pattern of classic meningovascular or parenchymatous neurosyphilis may be recognizable, the clinically varied manifestations of CNS syphilis require the clinician to consider it with nearly any presentation. A wide array of stroke syndromes caused by thrombotic or hemorrhagic mechanisms should be included in the differential diagnosis. Neurosyphilis should be considered in a relatively young person with a history of sexually transmitted diseases who otherwise has no risk factors for a cerebrovascular accident (e.g., atherosclerosis or hypertension) and suffers a stroke. Diabetic neuropathy has been said to mimic tabes dorsalis. Neurosyphilis should be considered in the differential diagnosis of any patient presenting with a dementia.

Cardiovascular Syphilis

Cardiovascular pathology leading to ischemia is more likely to result from atherosclerotic disease than from syphilis. Likewise, aortic aneurysm formation secondary to long-standing hypertension would be more common than syphilitic aneurysms. Aortic regurgitation often follows aortic stenosis. Besides syphilitic aortitis, pure valvular regurgitation without stenosis may be seen in patients with congenital valvular disease (e.g., Marfan syndrome), healed infective endocarditis, or other cardiac trauma.

Gummatous Syphilis

The lesions of late benign syphilis may be easily confused with other granulomatous processes, such as tuberculosis, sarcoid, fungal infection, and, occasionally, neoplastic disease.

TREATMENT

Penicillin remains the antimicrobial agent of choice in all stages of syphilis [*see Table 2*]. The studies that established this therapeutic standard were conducted decades ago. They used varying formulations of penicillin not currently available and would not meet present standards for randomization, blinding, sample size, and other aspects of good study design. These studies provide a weak basis for conclusions about optimal management today.

Primary and Secondary Syphilis

Therapy for primary syphilis should serve two overlapping goals: (1) to reduce and eliminate the capacity of the infected person to transmit disease and (2) to achieve the microbiologic eradication of *T. pallidum* from the host to resolve clinical manifestations and avoid long-term sequelae.

Sexual transmission of syphilis occurs almost exclusively during the early stages of disease, when contact with infectious lesions is most common. Penicillin eliminates infectivity relatively easily. Early studies of penicillin demonstrated the ability of the drug to rapidly convert dark-field–positive lesions to dark-field–negative lesions and hasten their resolution.

All the clinical manifestations of primary and secondary syphilis will resolve without therapy. However, the resolution of clinical manifestations does not ensure the eradication of *T. pallidum*. In such circumstances, patients are assumed to remain infected and to have entered the latent stage of disease. Studies of therapeutic efficacy rely on changes in serologic markers that may or may not reflect microbiologic activity of disease. There is general agreement today that a fourfold (two dilution) decrease in the NTST titer should occur as early as 3 months and as late as 12 or even 24 months after successful therapy. This notion is based on clinical experience and data from a study and subsequent reanalysis of the data.[22,23] The criteria for failure and the need for retreatment in both of these studies were poorly defined. Although it has thus been easy to create general categories for the patterns of response expected after therapy, few studies since the introduction of penicillin have followed patients long enough to establish the risk of disease progression.

Penicillin therapy For the past 3 decades, benzathine penicillin G (2.4 million units given once I.M.) has been the standard regimen for the treatment of primary syphilis and is currently recommended by both the CDC and the World Health Organization (WHO).[20,24] Studies have demonstrated the efficacy of benzathine penicillin but have also suggested a 3% to 10% failure rate by the serologic criteria mentioned above.[22] Where comparative data exist, this single injection appears to be nearly as effective, if not as effective, as a variety of more complicated regimens.[25] The ease and cost-effectiveness of a single injection clearly favors its use over multiple-injection–based therapies or therapies requiring hospitalization for intravenous treatment.

Nonpenicillin therapy It is advantageous to have alternative therapies to penicillin not only for the possibility of resistant *T. pallidum* but also to treat patients who are allergic to penicillin [*see* Complications, *below*]. There have been few comparative trials that have examined the role of nonpenicillin therapies. Tetracyclines have been used for decades as an alternative to penicillin, and the CDC currently recommends doxycycline (100 mg orally twice a day for 14 days) as second-line therapy in primary syphilis.[17] Retreatment rates after tetracycline therapy are higher than those after penicillin therapy.[22] Clinical experience supports the use of tetracyclines as alternatives to penicillin for cases of early stage syphilis in which penicillin cannot be used.

Erythromycin congeners, primarily the potentially hepatotoxic estolate form, have also been studied and appear to be as effective as tetracyclines in the treatment of early stage syphilis.[22] However, oral absorption of erythromycin is erratic, and the difficulty of achieving adequate serum levels with the widely used base form makes this an unreliable choice for primary syphilis treatment. Questions regarding erythromycin resistance have also been raised. A high-dose amoxicillin-probenecid combination is another possible alternative, although compliance with such a complex oral regimen makes it difficult to recommend.

The role of first-generation cephalosporins for the treatment of primary syphilis has also been examined.[26,27] Some efficacy was demonstrated for these agents, but at least one study suggested an unacceptably high failure rate.[28] Given the unclear role of these agents in the treatment of penicillin-allergic patients,

Table 2 Treatment of Syphilis

Stage	*Drug*	*Dose*	*Relative Efficacy*	*Cost/Day ($)*	*Comment*
Primary, secondary, and early latent syphilis	Benzathine penicillin G	2.4 million U I.M. once	Drug of choice	0.75–0.99	—
	Doxycycline	100 mg p.o. twice a day for 14 days	—	0.05–0.12	Therapy for nonpregnant patients who are allergic to penicillin
	Tetracycline	500 mg p.o. four times a day for 14 days	—	0.10–0.24	
	Ceftriaxone	1 g I.M. or I.V. daily for 8–10 days	—	60.00–69.99	Optimal dose has not been defined
	Azithromycin	2 g p.o. once	—	10.00–19.99	—
Late latent or latent syphilis of unknown duration	Benzathine penicillin G	2.4 million U I.M. once a wk for 3 wk (total 7.2 million U)	Drug of choice	0.75–0.99	May be associated with gastrointestinal upset
	Doxycycline	100 mg p.o. twice a day for 28 days	—	0.05–0.12	Therapy for nonpregnant patients who are allergic to penicillin
	Tetracycline	500 mg p.o. four times a day for 28 days	—	0.10–0.24	
Tertiary syphilis (gummatous and cardiovascular)	Benzathine penicillin G	2.4 million U I.M. once a wk for 3 wk	Drug of choice	0.75–0.99	—
	Doxycycline	100 mg p.o. twice a day for 28 days	—	0.05–0.12	Therapy for nonpregnant patients who are allergic to penicillin
	Tetracycline	500 mg p.o. four times a day for 28 days	—	0.10–0.24	—

*Monthly cost.

interest in them waned. The third-generation cephalosporin ceftriaxone is a suitable alternative agent for syphilis treatment. In vitro and animal-model data suggest that ceftriaxone has good antitreponemal activity.[29-31] Favorable pharmacokinetics and ceftriaxone's ability to penetrate the blood-brain barrier also make it an attractive choice for therapy. In small groups of patients with primary syphilis, ceftriaxone appeared to be effective.[32,33] Despite its perceived efficacy, ceftriaxone probably should not be given as a single dose for the treatment of primary syphilis, because it would not remain at treponemicidal levels for an acceptable period of time. The proper dosage and duration of therapy are matters of conjecture.

The azalide antibiotic azithromycin has many properties that suggest its usefulness as therapy for primary syphilis. It is active against *T. pallidum* in vitro and has been effective in experimental models of syphilis.[34,35] Although plasma concentration of azithromycin may be nearly unmeasurable, high levels are achievable in tissue.[36] Its extremely long half-life makes it ideal therapy for *T. pallidum*, an organism with a prolonged doubling time. In a number of small studies, azithromycin in varying doses appeared effective for incubating, primary, and secondary syphilis.[37,38]

Latent Syphilis

The successful treatment of latent syphilis is made problematic by a lack of clear and readily measurable clinical end points. By definition, there is no chancre or rash, the resolution of which would suggest response. The objective of therapy is the prevention of long-term sequelae. Except for waiting perhaps decades to see if late-stage disease develops, changes in serologic titers are the only indication of efficacy. Unfortunately, by its very nature, late latent syphilis elicits little in the way of an inflammatory response. Thus, nontreponemal serologic test results are often very low, and their decline after therapy may be proportionately slow. Seroreversal may never be achieved.

Because early latent syphilis occurs shortly after secondary syphilis appears to resolve, it should respond appropriately to similar single-dose and multiple-dose regimens. In most clinical settings, patients with latent-stage disease are unable to reliably give any history suggestive of recent primary disease and are usually treated as having late latent-stage disease.

Studies have suggested that penicillin, which is effective in primary disease, would be effective for latent disease as well. A variety of penicillin regimens were compared and patients followed over a wide range of periods,[39] and benzathine penicillin G provided equivalent duration of in vivo drug levels with fewer injections. Although there are no data directly supporting the current recommendation of three weekly injections of 2.4 million units of benzathine penicillin G (total, 7.2 million units), this therapy appears reasonable on the basis of theoretical assumptions regarding the slow doubling time of *T. pallidum* and the need for extended duration of therapy. The recommended regimen has been used effectively for decades and only rarely results in progression of disease.

There are virtually no useful data on the treatment of latent syphilis with nonpenicillin therapies. Recommendations of 28 days of tetracycline, chloramphenicol, or erythromycin are largely without clinical data to support their effectiveness.

Late Symptomatic Syphilis (Excluding Neurosyphilis)

Incidences of gummatous syphilis and cardiovascular syphilis have declined so dramatically in the penicillin era that it has been difficult to collect enough cases to conduct a study of

Table 2 (continued)

Stage	Drug	Dose	Relative Efficacy	Cost/Day ($)	Comment
Neurosyphilis	Aqueous crystalline penicillin G	18–24 million U/day, given as 3–4 million U I.V. every 4 hr for 10–14 days	—	10.00–19.99	—
	Procaine penicillin G and probenecid	Penicillin, 2–4 million U/day I.M. for 10–14 days; plus probenecid, 500 mg p.o. four times a day for 10–14 days	—	Penicillin, 2.00–2.99; probenecid, 10.00–19.99*	Can be used if compliance is good
	Ceftriaxone	2 g I.M. or I.V. daily for 10–14 days	—	60.00–69.99	Limited supporting data
Congenital syphilis	Aqueous crystalline penicillin G	100,000–150,000 U/kg/day, given as 50,000 U/kg/dose I.V. every 12 hr during the first 7 days of life and every 8 hr thereafter for a total of 10 days	—	10.00–19.99	—
	Procaine penicillin G	50,000 U/kg/day I.M. once a day for 10 days	—	2.00–2.99	—
Nonvenereal treponematoses (yaws, endemic syphilis, and pinta)	Benzathine penicillin G	Adults, 1.2 million U I.M. once; children, 600,000 U I.M. once	—	0.75–0.99	—
	Doxycycline	100 mg p.o. twice a day for 14 days	—	0.05–0.12	—
	Tetracycline	500 mg p.o. four times a day for 14 days	—	0.10–0.24	—

newer therapeutic modalities. A few studies conducted on small groups of patients soon after the introduction of penicillin demonstrated that penicillin therapy probably halts the progression of gummatous disease, although existing damage to skin, mucous membranes, soft tissue, and skeletal structures is likely irreparable. The therapeutic efficacy of penicillin for cardiovascular syphilis has never been clearly established[40] because of the difficulty of establishing a diagnosis and determining the efficacy of therapy in individuals with already advanced disease (i.e., cardiac structural damage). Despite suggestions that penicillin appeared to have some efficacy in preventing the progression of otherwise uncomplicated aortitis, carefully controlled studies comparing various therapies have never been conducted. In over 4 decades, there have been no studies examining the newer, long-acting preparations of penicillin, nor have there been studies of alternatives to penicillin.

Given the relative safety and ease of use of penicillin therapy for the treatment of cardiovascular syphilis, the WHO and the CDC recommendations appear reasonable, provided that patients are adequately evaluated for other complications of syphilis, such as neurosyphilis, and that adequate clinical and serologic follow-up can be performed.

Neurosyphilis

Frank symptomatic neurosyphilis, like cardiovascular syphilis and gummatous syphilis, has to a certain extent disappeared from clinicians' experience because of the success of therapies for primary and latent-stage disease. Although there is a significant body of work describing and essentially confirming the efficacy of penicillin in the treatment of neurosyphilis, this literature, like much of the syphilis-treatment literature, is compromised by poor posttreatment follow-up and poorly defined outcomes. Nonetheless, penicillin appears to be effective against neurosyphilis, in many instances improving or halting the inexorable progression of neurologic symptoms.

In the 1970s, the CDC recommended either 7.2 million units of benzathine penicillin G I.M. over 3 weeks or 9 million units of aqueous procaine penicillin G over 15 days as therapy for neurosyphilis. The former regimen was also recommended as treatment for late latent syphilis and probably offered some assurance that adequate therapy would be provided for patients with unappreciated asymptomatic neurosyphilis. Serious questions about the efficacy of these regimens, particularly of 7.2 million units of benzathine penicillin G for neurosyphilis, emerged almost simultaneously with their recommendation. Reports detailed the inability of the regimen to achieve acceptable spirocheticidal levels in CSF,[41-45] as well as treatment failures.[46-48] The response of patients with reported treatment failures to increased doses of penicillin has led the CDC to recommend prolonged courses of aqueous or procaine penicillin G for the treatment of neurosyphilis.

Despite the new CDC recommendations for therapy, there are no data from randomized, controlled trials that demonstrate superior efficacy of the new recommendations over previous recommendations. Some experts still contend, also in the absence of supporting data, that lower-dose regimens may be appropriate for asymptomatic neurosyphilis.

The efficacy of high-dose intravenous therapy for neurosyphilis in certain circumstances is also questionable. It is probable that advanced parenchymal disease (e.g., paresis or tabes dorsalis) will not improve after even the most aggressive therapy and may in fact progress. In patients with neurosyphilis and concomitant HIV infection, disease progression

may be seen after therapy.[49] Controversy also exists regarding the efficacy of procaine penicillin G.[49]

Ceftriaxone is an attractive alternative to penicillin in the treatment of neurosyphilis.[32,33,50] It has a long half-life, penetrates the CNS well, and has been shown in vivo and in vitro to have activity against *T. pallidum*. Although there are few well-designed trials supporting the clinical efficacy of ceftriaxone for the treatment of neurosyphilis, infectious disease specialists in the United States commonly use it as an alternative to penicillin.[51]

High-dose amoxicillin (2 g orally three times a day) combined with probenecid has been proposed as an alternative outpatient regimen for neurosyphilis because it has been documented to achieve treponemicidal levels of β-lactam antibiotic in the CSF.[52-54] No randomized, controlled studies have been done to evaluate this regimen for this purpose.

Both chloramphenicol and doxycycline have been examined in individual cases and in small groups of patients with neurosyphilis.[55,56] Limited data suggest that these drugs may have efficacy in the treatment of neurosyphilis, but there is not enough evidence to make final recommendations.

Oral steroids are a frequent adjunct in the therapy of sensorineural deafness complicating neurosyphilis,[57,58] but no controlled trials have been properly conducted to evaluate this practice.

Serologic Follow-up

In the treatment of syphilis, the lack of a microbe that can be cultured, the problem of latency, and the uncertainty of therapeutic efficacy combine to foster a unique clinical reliance on quantitative changes in serologic markers to determine response to and adequacy of therapy. Few, if any, infectious diseases require similar follow-up after treatment.

Current standards of care require patients with primary syphilis to return for clinical and serologic evaluation at 3, 6, and 12 months after therapy.[20] Patients with latent syphilis are followed less frequently, perhaps at 6-month intervals for up to 24 months. HIV-coinfected patients are followed more regularly. Recurrence of disease or lack of resolution of signs or symptoms of disease after treatment may be considered evidence of treatment failure. As a measure of disease activity, a two-tube, or fourfold, dilution increase in the NTST during follow-up is also widely accepted as a marker of treatment failure. In all of these instances, distinguishing treatment failure from reinfection may be difficult. At the opposite end of the therapeutic spectrum, cure may reasonably be assumed to have occurred if the NTST becomes nonreactive after therapy. Unfortunately, this development may occur slowly if at all.

A two-tube, or fourfold, dilution decrease in the NTST after treatment has become accepted as the measure of success. This finding will occur more rapidly after treatment for primary syphilis. Cumulative data and clinical experience over several decades indicate that in immunocompetent patients treated with standard regimens of benzathine penicillin G or tetracycline, the NTST should decline by two or three dilutions within 6 and 12 months of treatment, respectively. The proportion of patients expected to be seronegative within 12 to 24 months is less certain, but the weight of evidence suggests that at least 90% of patients with primary or secondary syphilis should serorevert within 1 or 2 years, respectively.

The treponemal-specific serologic tests revert to negative within 3 years in up to 25% of successfully treated patients with first-episode primary syphilis, so that a negative treponemal antibody test result does not always exclude a past history of syphilis.[20,59] However, seroreversal of the treponemal antibody test results occurs rarely, if ever, after treatment of syphilis at the secondary stage or beyond or after treatment of a second episode of primary syphilis. Some data suggest a higher prevalence of nonreactive treponemal antibody tests in HIV-infected patients after treatment of syphilis.[60] It is possible that HIV-infected persons serorevert more often than non–HIV-infected persons, but the issue is unresolved.

Serologic response after treatment in latent disease may be even more problematic than in primary disease. Long-term studies are not available. Another confounding issue is variation in the definition of late syphilis, which determines the dose or duration of therapy of latent infection. Most early writers defined late infection as syphilis of 4 years' duration, on the basis of observations that relapses of secondary syphilis in untreated individuals ceased after this interval. The usual cutoff in the United States, by contrast, is 1 year, after which sexual transmission is rare and classic tertiary syphilis begins to appear.[61] In 1986, the WHO defined late latent syphilis as infection of 2 years' duration.[24] All these definitions are arbitrary, and there are insufficient data for choosing among them on the basis of documented differences in the clinical or serologic response to various treatment regimens. Moreover, the duration of infection cannot be determined accurately in most asymptomatic seropositive patients, further confounding interpretation of the serologic response to treatment in most published studies.

Older data and the accumulated experience of recent years indicate that NTST results decline more slowly after treatment of late syphilis than after treatment of early syphilis and that these tests often remain reactive indefinitely, usually at low titer. Serologic titers in late latent syphilis and in tertiary syphilis decline slowly after treatment, and nonreactive tests will develop in only a minority of patients after upwards of 5 years.[25] The serologic response to the treatment of symptomatic late-stage syphilis may also result in only gradual declines in the NTST. Whether this reflects failure to achieve biologic cure or is merely a serologic phenomenon (so-called immunologic memory) is unresolved. For clinical application, a reasonable guideline is that failure of the NTST to decline by two or more dilutions within 12 months of treatment of late syphilis and any subsequent rise in titer are indications for comprehensive reevaluation and retreatment.

CSF abnormalities in neurosyphilis are also slow to respond after therapy. When reactive, the CSF-VDRL may remain so for years. Other parameters of inflammation, such as cell count and protein level, should resolve more rapidly. It appears to be a reasonable and acceptable practice to repeat CSF examination in patients treated for neurosyphilis at 3-month to 6-month intervals after therapy to ensure an appropriate response. Although the CDC suggests retreatment if CSF abnormalities persist after 2 years, there is no evidence that this condition is clearly indicative of treatment failure or that retreatment will result in an improved response. It is unclear whether serologic response can be expected to differ in all clinical settings, and it would be prudent to consider more frequent serologic follow-up for the HIV-infected patient treated for any stage of syphilis.

Despite considerable effort to develop guidelines for expected serologic response after treatment, it must be remembered that microbiologic response and serologic response are not synonymous. *T. pallidum* has been isolated by rabbit inoculation of lymph nodes of both patients and rabbits and has been identi-

fied histologically or immunochemically in aqueous humor after penicillin therapy that resulted in clinical and serologic resolution. Relapse (especially of neurosyphilis) has been reported after apparently adequate penicillin treatment of HIV-infected patients with primary syphilis.

Management of Sexual Partners and Incubating Syphilis

Syphilis treatment is not complete until treatment of sexual partners of the index case has been considered. The CDC recommends that sexual contacts exposed to an index case of primary, secondary, or early latent syphilis receive presumptive therapy within 90 days regardless of the serologic test results. For persons exposed more than 90 days before, therapy should be based on the serologic test results. The need for therapy for contacts of patients with latent or late-stage syphilis should be based on clinical evaluation.

High-risk sexual activity (as indicated by the recent acquisition of such sexually transmitted organisms as *Chlamydia trachomatis* and *Neisseria gonorrhoeae*) carries with it the risk of simultaneous acquisition of syphilis. Because the incubation period of syphilis may be much longer than that of diseases caused by these other organisms, neither clinical nor serologic evidence of syphilis may be present at the time of presentation for clinical care. It would therefore be advantageous for therapies directed toward *C. trachomatis* or *N. gonorrhoeae* to also adequately treat syphilis. Treatment with standard doses of ceftriaxone for gonorrhea will probably abort incubating syphilis. The same cannot be said for single-dose therapy with quinolone antibiotics that do not have activity against *T. pallidum.* Activity of single-dose cefixime against *T. pallidum* is uncertain. Spectinomycin is not active in this instance.[62] Although data are lacking to support a 7-day course of doxycycline, this therapy may in fact be effective. Single-dose azithromycin may also be useful for this purpose.[38]

COMPLICATIONS

Jarisch-Herxheimer Reaction

The Jarisch-Herxheimer reaction is an acute systemic syndrome characterized by fever, chills, headaches, myalgias, tachycardia, tachypnea, and nausea that occurs within the first few hours after the treatment of syphilis—usually secondary syphilis. It may result from the release of pyrogenic endotoxins from treponemes adversely affected by therapy. Treatment with anti-inflammatory agents during the first 24 to 48 hours is warranted. Some clinicians advocate the use of preemptive steroid therapy to mitigate any reaction in which additional end-organ damage (e.g., ocular or cardiovascular) may be problematic, but this practice is not supported by available data. In pregnancy, a Jarisch-Herxheimer reaction may precipitate premature labor or fetal distress.[63] Pregnant women should be advised to seek obstetric evaluation if they note contractions or decreased fetal movements after syphilis treatment.

Penicillin Allergy and Desensitization

In the United States, 5% to 10% of the population may have a history of penicillin allergy, and life-threatening complications may result in 2% to 13% of related allergic events.[64] Despite the relatively high risk of adverse events, there are certain clinical situations in which nonpenicillin therapies for syphilis are not acceptable and penicillin must be used (i.e., congenital syphilis, syphilis in pregnancy, and neurosyphilis).

Even where a history of penicillin allergy exists, the drug may not necessarily be contraindicated. A history of allergy may be inaccurate. Additionally, with the passage of time after an allergic event, many patients may stop expressing penicillin-specific IgE. Penicillin skin testing is a well-established means of securing the patient's true allergy status. The penicilloyl hapten moiety that is the largest product of penicillin metabolism is referred to as the major determinant and, conjugated to polylysine, is commercially available for skin testing. The remaining metabolites are referred to as the minor determinants and are available only at certain academic centers. Even in this setting, the quality criteria for their production have not been standardized. Major determinant skin testing, along with that for benzathine penicillin G (considered a minor determinant), should identify 90% to 97% of patients with penicillin allergy. Nearly all allergic patients will be identified if the other minor determinants are also used.[65] Patients who have a positive response to any of the major or minor determinants should undergo either oral or intravenous desensitization.[66]

Desensitization is usually performed only in a hospital setting. Syphilis treatment regimens may be initiated immediately after the successful completion of desensitization. Unless penicillin therapy is continued indefinitely after desensitization, patients should be considered potentially allergic.

SPECIAL SITUATIONS

Syphilis in Pregnancy

Routine screening and confirmatory serologic tests should reveal syphilis in pregnant women. In pregnancy, quantitative nontreponemal tests reflective of previously treated syphilis may not be specifically indicative of syphilis. An increase in titer in this setting should be considered nonspecific.

The treatment of syphilis in pregnancy must not only cure disease in the adult and prevent sexual transmission but also prevent fetal infection via transplacental passage of *T. pallidum* [*see* Congenital Syphilis, *below*]. These goals can be achieved with therapeutic regimens recommended for the nonpregnant individual. Studies have been conducted in small groups of pregnant women with syphilis using 2.4 to 4.8 million units of various forms of penicillin, often over the course of at least 1 week.[67-73] In each case, the incidence of congenital syphilis was exceedingly low. It was observed that where failures occurred, mothers had been treated late in pregnancy. More studies support the contention that treatment late in pregnancy carries a greater risk for congenital disease than treatment earlier in pregnancy.[74,75] No other factors appear to conclusively increase the risk of congenital syphilis after appropriate therapy of the mother, provided that currently recommended doses are used and that the serologic and clinical markers of disease do not suggest relapse or reinfection. Follow-up in this manner is no more difficult than would be expected after the treatment of the nonpregnant patient with syphilis.

Congenital Syphilis

Diagnosis The CDC recommends that, in screening for congenital syphilis, the mother's serum be tested rather than cord blood. Reactive standard serologic tests in the infant may reflect passive transfer of antibody. Although there is interest in the use of tests for IgM antibodies in the infant, such tests may be insensitive, and nonreactive tests should be interpreted with caution. Demonstration of the presence of *T. pallidum* by direct

microscopic examination in the umbilical cord, placenta, nasal discharge, or skin lesions is considered definitive for diagnosis.

Any rise in the incidence of secondary syphilis is invariably accompanied by a rise in the incidence of congenital syphilis. In most cases, the newborn is asymptomatic. Symptomatic disease may occur early or later after birth. In the perinatal period, involvement of the mucocutaneous tissues may be seen, characterized by rhinitis and a desquamative, vesicular, or bullous rash of the skin, particularly on the palms and on the soles of the feet. Hepatic disease can also occur and may result in significant morbidity and mortality. Hepatomegaly, splenomegaly, and hematologic abnormalities occur in this setting. Osteochondritis and perichondritis lead to the classic findings of the so-called saddle-nose deformity and saber shins. After 6 to 12 months, the child enters a latent phase of disease not unlike that seen in adults. Late-stage disease in children may manifest as interstitial keratitis, eighth-nerve deafness, arthropathy, and dental malformations.

Treatment Except where adequate treatment and follow-up of the mother with syphilis can be documented more than 1 month before delivery, some therapeutic intervention will be necessary for the infant. No well-controlled study has determined the appropriate dose of penicillin for the infant suspected of having congenital syphilis, but for those with symptomatic disease at birth, 50,000 U/kg of I.V. aqueous crystalline penicillin G every 12 hours for 10 days or 50,000 U/kg of I.M. procaine penicillin G every day for 10 days appears to be satisfactory. These doses are somewhat arbitrary but are based on doses that may be used to treat bacterial meningitis and would also safely provide at least the standard spirocheticidal levels in the CSF. Because of the difficulty of ruling out active disease in the newborn, many experts recommend the same course of treatment for infants whose mothers' treatment histories are negative or unclear or who received nonpenicillin therapy during pregnancy or within 4 weeks of delivery.

For many years, a single dose of benzathine penicillin for the newborn was considered adequate therapy for congenital syphilis once neurosyphilis had been ruled out. However, reports of treatment failures have cast doubt on this assumption.[76] Benzathine penicillin continues to be recommended as an option in a variety of settings, particularly when the infant is asymptomatic and neurosyphilis has been excluded, but questions exist about the adequacy of maternal therapy and whether follow-up of the child can be assured.[20,77] There are no data to support the contention, largely based on clinical experience, that in low-risk situations these therapies result in nearly universal prevention of congenital syphilis. It is doubtful that most infants receiving this therapy have syphilis at all.

There are no clinical trials examining the role of other antibiotics for the treatment of congenital syphilis. Nonpenicillin therapies recommended for syphilis in adults cannot automatically be recommended for syphilis in infants. Tetracyclines are contraindicated in pregnancy and in infants because of their toxicity to deciduous teeth and long bones and because of concern that they may cause maternal hepatotoxicity, particularly when there is preexisting renal dysfunction.[78] Case reports of symptomatic congenital syphilis in instances in which women received erythromycin during pregnancy as therapy for syphilis suggest that this drug is not a suitable alternative to penicillin.[79-81] These clinical observations are supported by studies demonstrating that placental transfer of erythromycin may not be consistent.[82,83] Issues of compliance and gastrointestinal disturbance in pregnant women treated with erythromycin, although not studied, raise additional questions about this drug's role in pregnancy.

As in adults, serologic follow-up after treatment in infants is essential to ensure the adequacy of response. The ideal frequency of such follow-up is poorly defined, and recommendations range from every month to every 3 months during the first 6 months, with continued follow-up through 12 to 24 months, depending on response.

Syphilis in Children

Penicillin therapy Children with primary, secondary, or early latent syphilis should be treated with benzathine penicillin G (50,000 U/kg I.M., to 2.4 million units in a single dose). Children with late latent syphilis or syphilis of unknown duration should also receive benzathine penicillin G (50,000 U/kg I.M., to 2.4 million units administered as three doses at 1-week intervals). In both cases, a CSF examination should be performed to rule out neurosyphilis.

Alternative therapy The American Academy of Pediatrics recommends that tetracyclines be considered as an alternative to penicillin for both primary and latent disease after 9 years of age.[84] Erythromycin may be considered in early aquired syphilis even in children younger than 9 years (500 mg orally four times a day for 14 days). Follow-up must be ensured. Newer therapies, such as ceftriaxone and azithromycin, may also be effective. There are no studies to support any of these recommendations. Early interest in chloramphenicol as therapy for syphilis has waned because of the drug's well-known side effects.

Syphilis and HIV

Risk behaviors that contribute to the transmission of syphilis are often the same as those that contribute to the transmission of HIV, and therefore, coinfection is common. Each infection has been demonstrated to have some effect on the other. Substantial evidence suggests that early stage syphilis enhances the transmission of HIV.[85,86] Presumably, the inflammatory milieu of syphilitic lesions can serve either as a portal of entry of HIV or as the route through which HIV may be transmitted to an uninfected partner.

In coinfected persons, the deleterious effect of HIV on the immune system can alter the natural history, diagnosis, and management of syphilis. Since the earliest years of the HIV epidemic, case studies have reported on individuals who failed to respond to conventional doses of penicillin therapy and who often progressed rapidly from primary syphilis to neurosyphilis.[87,88] Other important findings have included seronegative secondary syphilis,[89] a likelihood of higher-titer NTSTs by stage,[90] and an increased incidence of biologic false positive NTSTs.[21,91]

A study prospectively followed both HIV-seronegative and HIV-seropositive patients with primary syphilis who received either conventional penicillin therapy or an enhanced regimen of high-dose amoxicillin and probenecid. No neurologic complications and no differences in response to therapy were noted between the groups.[52,92] Except for a tendency for HIV-infected persons to present with multiple chancres, as opposed to a single chancre in non–HIV-infected individuals, no significant differences were noted in presentation or rapidity to resolution of primary syphilis findings. In another study, the CSF white

blood cell count, protein, and VDRL were slower to resolve after therapy in HIV-infected individuals than in non–HIV-infected individuals.[93] Clinical failures were not noted.

The propensity for *T. pallidum* to invade the CNS early during infection and the likely importance of immune mechanisms for its control suggest possible mechanisms for the risk of neurosyphilis in coinfected persons. This issue is further complicated by the likelihood for HIV-infected persons to have CSF abnormalities even in the absence of discernible CNS pathology.[94] Many experts recommend that HIV-infected persons with latent syphilis or primary syphilis undergo a CSF evaluation and that those with CSF abnormalities be treated accordingly.

Nonvenereal Treponematoses

The nonvenereal treponematoses—yaws, endemic syphilis, and pinta—are a group of infections distributed throughout tropical and semitropical areas of the world. They are primarily noted to cause a variety of skin and skeletal lesions. There is little biologic difference between the treponemes that cause these conditions.[95] Their distinguishing nomenclature is based primarily on the clinical disease that characterizes each. Each is linked to poor hygiene and person-to-person transmission and disproportionately affects children. As a result of mass treatment programs conducted in the 1950s and the 1960s, initiated primarily by the WHO in conjunction with local authorities, the incidences of all of the nonvenereal treponematoses are today quite low.

YAWS

Epidemiology

Yaws is a disease primarily affecting individuals who live in rural areas of the tropics. It is seen in parts of South America, Africa, Southeast Asia, and Oceania. In the Western Hemisphere, yaws occurs sporadically in small regions of Colombia, Haiti, Suriname, and Guyana. Propagation depends on a hot, humid climate. Persons younger than 15 years appear to be most at risk for acquisition of this infection. It is known by a number of names in different parts of the world, including frambesia, pian, bouba, and parangi.

Pathogenesis

T. pallidum subspecies *pertenue* is associated with yaws. Transmission occurs through skin-to-skin contact, where abrasions or other defects allow spirochetes to enter and cause infection. The incubation period of yaws is like that of syphilis. Within 9 days to 3 months, a granulomatous papule forms in the area of inoculation. This papule gradually enlarges and becomes either hyperkeratotic or frankly ulcerative, often with the development of an overlying amber crust of bacterial superinfection. As with syphilis, even in the absence of specific treatment, these primary lesions resolve.

Diagnosis

The diagnosis of yaws should be suspected in anyone with skin or bone lesions who has spent time in a tropical region.

Clinical manifestations Systemic symptoms of fever, arthralgias, and regional lymphadenopathy occur in conjunction with the development of a diffuse rash similar in appearance to the primary lesion. Osteitis and periosteitis can occur involving the fingers (polydactylitis), tibia (saber shin), and the paranasal maxillae. Secondary lesions may resolve and then relapse over the course of many years. Gummatous lesions with significant dermal scarring and destruction of underlying bone occur years after infection. Whether CNS involvement occurs is debatable.

Laboratory tests Dark-field microscopy of both primary and secondary lesions may yield spirochetes. Nonspecific treponemal serologic tests (e.g., VDRL or RPR) and treponemal-specific serologic tests (e.g., FTA-ABS or TP-PA) should be reactive in cases of yaws, but these tests cannot distinguish between yaws and other treponematoses.

Differential Diagnosis

The differential diagnosis includes syphilis, eczema, psoriasis, scabies, pyoderma, tuberculosis, leprosy, and ecthyma, as well as other processes.

Treatment

Penicillin remains the treatment of choice. For primary disease, doses of benzathine penicillin, 1.2 million units I.M. once in adults and 600,000 units I.M. once in children, should be adequate. Tetracyclines and chloramphenicol are acceptable alternatives. Close contacts of the patient should also be treated.

ENDEMIC SYPHILIS

Epidemiology

Unlike yaws, endemic syphilis, or bejel, is epidemiologically restricted to the arid regions of the Middle East and North Africa. Like the other treponematoses, endemic syphilis is transmitted by close personal contact and affects primarily children living in rural areas.

Pathogenesis

Endemic syphilis is caused by the spirochete *T. pallidum* subspecies *endemicum*. The primary lesion of endemic syphilis usually occurs at a site of inoculation on the oral mucosa and is often unappreciated. This has led to the suspicion that contaminated utensils and drinking vessels may be involved in the spread of the disease.

As in the case of yaws, gummatous late-stage lesions result in destruction and deformation of the mucosa and bony structures, particularly of the mouth and nasal structures. Long bones are sometimes involved. Hyperkeratosis of the soles of the feet occurs. CNS disease is uncommon.

Diagnosis

The diagnosis of endemic syphilis should be considered in any person from an endemic area who has skin or bone disease.

Clinical manifestations Secondary lesions, which may be the first manifestation of infection to be noticed, occur on and around the oral mucosa as well as on other moist, nonkeratinized epithelium and in the intertriginous areas. In rare cases, lesions are disseminated across the surface of the body. Diffuse lymphadenopathy and osteitis are also seen in this circumstance.

Laboratory tests Dark-field microscopy of primary and secondary lesions will be positive for spirochetes, and serologic testing will also be reactive.

Differential Diagnosis

The differential diagnosis of endemic syphilis includes many of the conditions considered in the differential diagnosis of yaws.

Treatment

As in the case of yaws, a single I.M. injection of 1.2 million units of benzathine penicillin is effective to eradicate disease.

PINTA

Epidemiology

Pinta (Spanish for blemish) occurs only in the Western Hemisphere and only in rural areas of southern Mexico and parts of Colombia. All age groups are susceptible to infection.

Pathogenesis

The organism that causes pinta, *T. carateum*, is transmitted through contact of broken skin with infectious lesions.

Within a few weeks of inoculation, a small papule develops at the infection site. The initial papule is followed rapidly by the appearance of other small, pruritic, erythematous papules around the site of the lesion.

Diagnosis

Pinta should be suspected in anyone with characteristic dermal lesions who has been in an endemic area.

Clinical manifestations Papules may coalesce to form squamous plaques. They may then persist for some time before resolving and leaving areas of hypopigmentation. Within 3 to 12 months, a disseminated form of skin lesion known as a pintid appears, which is similar to the initial lesion. These also may coalesce. Diffuse lymphadenopathy is common. Eventually, longstanding lesions leave atrophic and hypopigmented patches, particularly in the area of the wrists, elbows, and ankles.

Laboratory tests Except in the late stages of disease, lesions should yield dark-field microscopic results positive for spirochetes. Standard serologic tests for syphilis are usually reactive except during the earliest phases of disease.

Differential Diagnosis

Conditions with which pinta may be confused include tinea versicolor, vitiligo, and chloasma.

Treatment

The same penicillin doses used for yaws and endemic syphilis should be effective for pinta. Tetracyclines and chloramphenicol are acceptable alternatives. Lesions may be slow to heal. The hypopigmented, or achromic, lesions of late-stage pinta may never improve despite therapy.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Schroeter A, Turner RH, Lucas JB, et al: Therapy for incubating syphilis: effectiveness of gonorrhea treatment. JAMA 218:711, 1971

2. Schober P, Gabriel G, White P, et al: How infectious is syphilis? Br J Vener Dis 59:217, 1983

3. von Wessouwetz J: The incidence of infection in contacts of early syphilis. J Vener Dis Infect 28:132, 1948

4. Surveillence for primary and secondary syphilis. MMWR Morb Mortal Wkly Rep 42(SS3):13, 1993

5. Surveillence for geographic and secular trends in congenital syphilis. MMWR Morb Mortal Wkly Rep 42(SS6):59, 1993

6. Division of STD Prevention: Sexually Transmitted Disease Surveillence 1998. U.S. Department of Health and Human Services. Centers for Disease Control and Prevention, Atlanta, September 1999

7. The National Plan to Eliminate Syphilis from the United States. National Center for HIV, STD and TB Prevention. Division of STD Prevention. Centers for Disease Control and Prevention, Atlanta, October 1999

8. Clark E, Danbolt N: The Oslo Study of the natural course of untreated syphilis. Med Clin North Am 48:613, 1964

9. Musher D: The biology of *Treponema pallidum*. Sexually Transmitted Diseases. Homes K, Mardh P, Sparling P, et al, Eds. McGraw-Hill, New York, 1999

10. Tight R, Wagner J: Skeletal involvement in secondary syphilis detected by bone scanning. JAMA 235:2326, 1976

11. Campisi D, Whitcomb C: Liver disease in early syphilis. Arch Intern Med 139:365, 1978

12. Bhorade M, Carag HB, Lee HJ, et al: Nephropathy of secondary syphilis: a clinical and pathological spectrum. JAMA 216:1159, 1971

13. Winters H, Notar-Francesco V, Bromberg K, et al: Gastric syphilis: five recent cases and a review of the literature. Ann Intern Med 116:314, 1992

14. Tait I: Uveitis due to secondary syphilis. Br J Vener Dis 59:397, 1983

15. Ormerod LD, Puklin JE, Sobel JD: Syphilitic posterior uveitis: correlative findings and significance. Clin Infect Dis 32:1661, 2001

16. Lukehart S, Hook E, Baker-Zander S, et al: Invasion of the central nervous system by *Treponema pallidum*: implications for diagnosis and treatment. Ann Intern Med 109:855, 1988

17. Augenbraun M, Dehovitz J, Feldman J, et al: Biological false-positive syphilis test results for women infected with human immunodeficiency virus. Clin Infect Dis 19:1040, 1994

18. Hart G: Syphilis tests in diagnostic and therapeutic decision making. Ann Intern Med 104:368, 1986

19. Larsen S, Steiner B, Rudolph A: Laboratory diagnosis and interpretation of tests for syphilis. Clin Micro Rev 8:1, 1995

20. 1998 guidelines for treatment of sexually transmitted diseases. MMWR Morb Mortal Wkly Rep 47(RR-1):28, 1997

21. Horowitz H, Valsamis M, Wicher V, et al: Brief report: cerebral syphilitic gumma confirmed by polymerase chain reaction in a man with human immunodeficiency virus infection. N Engl J Med 331:1488, 1994

22. Schroeter A, Lucas J, Price E, et al: Treatment for early syphilis and reactivity of serologic tests. JAMA 221:471, 1972

23. Brown S, Zaidi A, Larsen S, et al: Serologic response to syphilis treatment: a new analysis of old data. JAMA 253:1296, 1985

24. Management of Patients with Sexually Transmitted Diseases. World Health Organization Technical Report Series, World Health Organization, Geneva, 1991, p 810

25. Idsoe O, Guthrie T, Wilcox W: Penicillin in the treatment of syphilis. Bull WHO 47(suppl):1, 1972

26. Duncan W, Knox J: Cephalexin therapy for infectious syphilis. Arch Dermatol 110:77, 1974

27. Nicolis G, Loucopoulos A: Cephalothin in the treatment of syphilis. Br J Vener Dis 50:270, 1974

28. Duncan W, Knox J: Cephalexin therapy for infectious syphilis. Arch Dermatol 110:77, 1974

29. Korting HC, Walther D, Riethmuller, et al: Comparative in vitro susceptibility of *Treponema pallidum* to ceftizoxime, ceftriaxone and penicillin G. Chemotherapy 32:352, 1986

30. Johnson R, Bey R, Wolgamot S: Comparison of the activities of ceftriaxone and penicillin G against experimentally induced syphilis in rabbits. Antimicrob Agents Chemother 21:984, 1982

31. Korting H, Walther D, Riethmuller U, et al: Ceftriaxone given repeatedly cures manifest syphilis in the rabbit. Chemotherapy 33:376, 1987

32. Hook EW 3rd, Roddy RE, Handsfield HH: Ceftriaxone therapy for incubating and early syphilis. J Infect Dis 158:881, 1988

33. Moorthy T, Lee C, Lim K, et al: Ceftriaxone for treatment of primary syphilis in men: a preliminary study. Sex Transm Dis 14:116, 1987

34. Stamm L, Parrish E: In-vitro activity of azithromycin and CP-63,956 against *Treponema pallidum*. J Antimicrob Chemother 25(suppl A):11, 1990

35. Lukehart S, Fohn M, Baker-Zander S: Efficacy of azithromycin for therapy of active syphilis in the rabbit model. J Antimicrob Chemother 25(suppl A):91, 1990

36. Drew R, Gallis H: Azithromycin-spectrum of activity, pharmacokinetics and clinical applications. Pharmacotherapy 12:161, 1992

37. Verdon MS, Handsfield H, Johnson R: Pilot study of azithromycin for treatment of primary and secondary syphilis. Clin Infect Dis 19:486, 1994

38. Hook E, Stephans J, Ennis D: Azithromycin compared with penicillin G benzathine for the treatment of incubating syphilis. Ann Intern Med 131:434, 1999

39. Hatos G: Evaluation of 460 cases of treated syphilis. Med J Aust 2:415, 1972

40. St. John R: Treatment of cardiovascular syphilis. J Am Vener Dis Assoc 3:148, 1976

41. Yoder F: Penicillin treatment of neurosyphilis. JAMA 232:270, 1975

42. Mohr J, Griffiths W, Jackson R, et al: Neurosyphilis and penicillin levels in cerebrospinal fluid. JAMA 236:2208, 1976

43. Dunlop E, Al-Egaily S, Houang E: Penicillin levels in blood and CSF achieved by treatment of syphilis. JAMA 241:2538, 1979

44. Ducas J, Robson HG: Cerebrospinal fluid penicillin levels during therapy for latent syphilis. JAMA 246:2583, 1981

45. Goh B, Smith G, Smarasinghe L, et al: Penicillin concentrations in serum and cerebrospinal fluid after intramuscular injection of aqueous procaine penicillin 0.6 MU with and without probenecid. Br J Vener Dis 60:371, 1984

46. Hooshmand H, Escobar M, Kopf S: Neurosyphilis. JAMA 219:726, 1972

47. Greene B, Miller N, Bynum T: Failure of penicillin G benzathine in the treatment of neurosyphilis. Arch Intern Med 140:1117, 1980

48. Jorgenson J, Tikjob G, Weismann K: Neurosyphilis after treatment of latent syphilis with benzathine penicillin. Genitourin Med 62:129, 1986

49. Gordon S, Eaton M, George R, et al: The response of symptomatic neurosyphilis to high-dose intravenous penicillin G in patients with human immunodeficiency virus infection. N Engl J Med 331:1469, 1994

50. Marra CM, Boutin P, McArthur JC, et al: A pilot study evaluating ceftriaxone and penicillin G as treatment agents for neurosyphilis in human immunodeficiency virus–infected individuals. Clin Infect Dis 30:540, 2000

51. Augenbraun M, Workowski K: Ceftriaxone therapy for syphilis: report from the emerging infections network. Clin Infect Dis 29:1337, 1999

52. Rolfs R, Joesoef R, Hendershot EF, et al: A randomized trial of enhanced therapy for early syphilis in patients with and without human immunodeficiency virus infection. The Syphilis and HIV Study Group. N Engl J Med 337:307, 1997

53. Faber W, Bos J, Rietra P, et al: Treponemicidal levels of amoxicillin in cerebrospinal fluid after oral administration. Sex Transm Dis 10:148, 1983

54. Morrison R, Harrison S, Tramont E: Oral amoxycillin, an alternative treatment for neurosyphilis. Genitourin Med 61:359, 1985

55. Romanowski B, Starreveld E, Jarema A: Treatment of neurosyphilis with chloramphenicol: a case report. Br J Vener Dis 59:225, 1983

56. Yim C, Flynn N, Fitzgerald F: Penetration of oral doxycycline into the cerebrospinal fluid of patients with latent or neurosyphilis. Antimicrob Agents Chemother 28:347, 1985

57. Rothenberg R, Becker G, Wiet R: Syphilitic hearing loss [editorial]. South Med J 72:118, 1979

58. Steckelberg J, McDonald T: Otologic involvement in late syphilis. Laryngoscope 94:753, 1984

59. Romanowski B, Sutherland R, Fick G, et al: Serologic response to treatment of infectious syphilis. Ann Intern Med 114:1005, 1991

60. Haas J, Bolan G, Larsen S, et al: Sensitivity of treponemal tests for detecting prior treated syphilis during human immunodeficiency virus infection. J Infect Dis 162:862, 1990

61. Zenker P, Rolfs R: Treatment of syphilis, 1989. Rev Infect Dis 12(suppl 6):S590, 1990

62. Peterman TA, Zaidi AA, Lieb S, et al: Incubating syphilis in patients treated for gonorrhea: a comparison of treatment regimens. J Infect Dis 170:689, 1994

63. Klein V, Cox S, Mitchell M, Wendel G: The Jarisch-Herxheimer reaction complicating syphilotherapy in pregnancy. Obstet Gynecol 75:375, 1990

64. Sullivan T, Yecies L, Shatz G, et al: Desensitization of patients allergic to penicillin using orally administered β-lactam antibiotics. J Allergy Clin Immunol 69:275, 1982

65. Lin R: A perspective on penicillin allergy. Arch Intern Med 152:930, 1992

66. Wendel G, Stark B, Jamison R, et al: Penicillin allergy and desensitization in serious infections during pregnancy. N Engl J Med 312:1229, 1985

67. Ingraham N: Prevention and treatment of prenatal syphilis. Am J Med 5:693, 1948

68. Cross J, McCain J, Heyman A: The use of crystalline penicillin G in the treatment of syphilis in pregnancy. Am J Obstet Gynecol 57:461, 1949

69. Bundeson H, Rodriques J, Aron H: Outcome of pregnancies of women treated with aqueous penicillin for early infectious syphilis. Arch Dermatol Syph 62:230, 1950

70. Wammock V, Carrozzino O, Ingraham N, et al: Penicillin therapy of the syphilitic pregnant woman. Am J Obstet Gynecol 59: 806, 1950

71. Smith C, O'Brien J, Simpson W, et al: Treatment of early syphilis with N, N' Dibenzylenediamine penicillin G. Am J Syph 38:136, 1954

72. Nelson N, Struve V: Prevention of congenital syphilis by the treatment of syphilis in pregnancy. JAMA 161:869, 1956

73. Jackson F, Vanderstoep E, Knox J, et al: Use of aqueous benzathine penicillin G in the treatment of syphilis in pregnant women. Am J Obstet Gynecol 83:1389, 1962

74. Taber L, Huber T: Congenital syphilis. Prog Clin Biol Res 3:183, 1975

75. Mascola L, Pelosi R, Blount J, et al: Congenital syphilis: Why is it still occurring? JAMA 252:1719, 1984

76. Beck-Sague C, Alexander E: Failure of benzathine penicillin G treatment in early congenital syphilis. Pediatr Infect Dis 6:1061, 1987

77. Zenker P, Berman S: Congenital syphilis: trends and recommendations for evaluation and management. Pediatr Infect Dis J 110:516, 1991

78. Wendel G: Gestational and congenital syphilis. Clin Perinatol 15:287, 1988

79. Hashiasaki P, Wertzberger G, Conrad G, et al: Erythromycin failure in the treatment of syphilis in a pregnant woman. Sex Transm Dis 10:36, 1983

80. Fenton L, Light I: Congenital syphilis after maternal therapy with erythromycin. Obstet Gynecol 47:492, 1976

81. Pavithran K: Failure of erythromycin to prevent prenatal syphilis and cure syphilis. Ind J Sex Transm Dis 7:30, 1986

82. Kiefer L, Rubin A, McCoy J, et al: The placental transfer of erythromycin. Am J Obstet Gynecol 69:174, 1955

83. Philipson A, Sabath L, Charles D: Transplacental passage of erythromycin and clindamycin. N Engl J Med 288:1219, 1973

84. American Academy of Pediatrics: Syphilis. Report of the Committee on Infectious Diseases. Red Book, 25th ed. Pickering LK, Ed. American Academy of Pediatrics, Elk Grove, Illinois, 2000, p 547

85. Kreiss J, Coombs R, Plummer F, et al: Isolation of human immunodeficiency virus from genital ulcers in Nairobi prostitutes. J Infect Dis 160:380, 1989

86. Telzak E, Chaisson M, Bevie P, et al: HIV-1 seroconversion in patients with and without genital ulcer disease: a prospective study. Ann Intern Med 119:1181, 1993

87. Berry C, Hooten T, Collier A, et al: Neurologic relapse after benzathine penicillin therapy for secondary syphilis in a patient with HIV infection. N Engl J Med 316:1587, 1987

88. Johns D, Tierney M, Felsenstein D: Alteration in the natural history of neurosyphilis by concurrent infection with the human immunodeficiency virus. N Engl J Med 316:1569, 1987

89. Hicks C, Benson P, Lupton G: Seronegative secondary syphilis in a patient infected with human immunodeficiency virus with Kaposi's sarcoma. Ann Intern Med 107:492, 1987

90. Gourevitch M, Selwyn P, Davenny D, et al: Effects of HIV infection on the serologic manifestation and response to treatment of syphilis in intravenous drug users. Ann Intern Med 118:350, 1993

91. Rompalo A, Cannon R, Quinn T, et al: Association of biologic false positive reactions for syphilis with human immunodeficiency virus infection. J Infect Dis 165:1124, 1992

92. Rompalo AM, Joesoef MR, O'Donnell JA, et al: Clinical manifestations of early syphilis by HIV status and gender: results of the syphilis and HIV study. Sex Transm Dis 28:158, 2001

93. Marra C, Longstreith W, Maxwell C, et al: Resolution of serum and cerebrospinal fluid abnormalities after treatment of neurosyphilis. Sex Transm Dis 23:184, 1996

94. Hollander H: Cerebrospinal fluid normalities and abnormalities in individuals infected with human immunodeficiency virus. J Infect Dis 158:855, 1988

95. Noordhoek G, Hermans P, Paul A, et al: *Treponema pallidum* subspecies *pallidum* (Nichols) and *Treponema pallidum* subspecies *pertenue* (CDC 2575) differ in at least one nucleotide: comparison of two homologous antigens. Microb Pathogenesis 6:29, 1989

VII LYME DISEASE AND OTHER SPIROCHETAL ZOONOSES

David C. Tompkins, m.d.
Benjamin J. Luft, m.d., f.a.c.p.

Lyme Disease

Lyme disease is a vector-borne zoonosis caused by *Borrelia burgdorferi,* which is a thin, spiral, motile, extracellular bacterium belonging to the family Spirochaetaceae. The hallmark of infection is erythema migrans, an annular erythematous skin lesion that usually appears at the site of the tick bite.

HISTORY

In 1976, Steere and colleagues noted an association between erythema migrans and a cluster of patients with knee arthritis in Old Lyme, Connecticut; they called this syndrome Lyme arthritis.[1] In the subsequent decade, these investigators defined the multisystem nature of the disease and modified its name to Lyme disease.

Isolation of the Lyme disease pathogen was not accomplished until 1981, when Burgdorfer and colleagues demonstrated a new spirochete in *Ixodes* ticks collected on Shelter Island, New York.[2] In 1982, spirochetes were identified in the midgut of the adult form of the deer tick *I. dammini,* and were named *B. burgdorferi*. Conclusive evidence that *B. burgdorferi* causes Lyme disease came in 1984, when *B. burgdorferi* was cultured from the blood of patients with erythema migrans, from the rash itself, and from the cerebrospinal fluid of a patient with meningoencephalitis and a history of erythema migrans.

EPIDEMIOLOGY

In the United States, surveillance for Lyme disease was begun by the Centers for Disease Control and Prevention (CDC) in 1982. Since that time, the number of reported cases has increased dramatically; nevertheless, as many as 90% of Lyme disease cases may be going unreported. Lyme disease has been reported in 49 of the 50 states, but most cases occur in the Northeast, the Midwest, and northern California. Nine states account for more than 90% of the nationally reported cases, with Connecticut leading the group. The other states are Rhode Island, New York, Pennsylvania, Delaware, New Jersey, Maryland, Massachusetts, and Wisconsin.[3,4] The rising incidence of Lyme disease in the United States may be explained by multiple factors, including an increase in the numbers of ixodid ticks, the expansion of residential areas into previously rural woodlands (habitats favored by ixodid ticks and their hosts), an exploding deer population, and increased recognition.

The Lyme disease pathogen, *B. burgdorferi*, is maintained in and transmitted by ticks of the *I. ricinus* complex, including *I. scapularis* in the northeast and north central United States [*see Figure 1*], *I. pacificus* on the West Coast of the United States, *I. ricinus* in Europe, and *I. persulcatus* in Asia.[5] In Europe, three genospecies of the *B. burgdorferi sensu lato* complex are pathogenic, including *B. burgdorferi sensu stricto*, *B. garinii*, and *B. afzelii*. *B. burgdorferi* is the only pathogenic species in North America. The varying relative distribution of these genospecies from region to region throughout Europe and Asia may account for the relative variability of disease syndromes associated with Lyme disease. In the United States, most patients have symptomatic illness,[6,7] whereas in Europe most patients are asymptomatic.

In field studies in Connecticut and New York, *B. burgdorferi* has been found in 10% to 50% of nymphal and adult *I. scapularis* ticks.[8,9] Although *B. burgdorferi* has been demonstrated in mosquitoes and deer flies, only ticks of the *I. ricinus* complex seem to be important in the transmission of the spirochete to humans.[10] An enzootic cycle of infection is maintained through passage of *B. burgdorferi* back and forth between ticks and their hosts. Infected nymphal ticks transmit *B. burgdorferi* to mice, which serve as a reservoir from which uninfected larvae may acquire infecting organisms. In this manner, a high rate of infection can be maintained in the tick population when the organism, ticks, mice, and deer are all present in the environment.

In temperate climatic zones, the seasonal variation of onset of Lyme disease is explained by the ecology of the predominant tick vectors. The ixodid tick has a three-stage life cycle (larva, nymph, and adult) that spans 2 years. Larvae hatch from fertilized eggs in late spring and feed once for 2 or more days in midsummer. Preferred hosts include a broad range of small mammals. The next spring they molt into nymphs and feed again for 3 or 4 days, with the same host range. After this second blood meal, the nymphs molt into adults. Adult *I. scapularis* organisms have a narrower host range, with a preference for deer. Mating occurs on deer, and the female deposits her eggs and the cycle begins anew.[11] During their 2-year life cycle, ticks typically feed once during each of the three stages, usually the late summer for larval ticks, the following spring for nymphs, and autumn for the adults. *I. scapularis* nymphs appear to be the most important vector for transmission of *B. burgdorferi*. According to laboratory studies, a minimum of 36 to 48 hours of attachment of the tick is required for transmission. In the United States, most cases involving *B. burgdorferi* occur between May and August, which corresponds with increased outdoor human and nymphal tick activity.

Figure 1 ***Ixodes scapularis*, also known as deer tick or black-legged tick, is the vector of Lyme disease. Adult ticks are approximately 2.5 mm in size—about the size of a sesame seed.**

The risk of infection in a given area depends largely on the density of the tick population, as well as on their feeding habits and animal hosts, which have evolved differently in different locations. In the northeastern and north central United States, *I. scapularis* ticks are abundant, and a highly efficient cycle of *B. burgdorferi* transmission occurs between immature larval and nymphal *I. scapularis* ticks and white-footed mice. This results in high rates of infection in nymphal ticks and a high frequency of Lyme disease in humans during the late spring and summer months.[12,13] The proliferation of deer, which are the preferred host of the adult tick, was a major factor in the emergence of epidemic Lyme disease in the northeastern United States during the late 20th century.

I. scapularis and other ticks in the *I. ricinus* complex may transmit multiple pathogens. *I. scapularis* is also a vector for *Anaplasma phagocytophila,* which causes human granulocytic ehrlichiosis [*see 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*] and *Babesia microti,* which causes babesiosis [*see 7:XXXIV Protozoan Infections*]. The proportion of *I. scapularis* or *I. ricinus* ticks coinfected with both *B. burgdorferi* and *A. phagocytophila* is generally low, ranging from less than 1% to 6% in six geographic areas. A higher prevalence of tick coinfection (26%) has been reported in Westchester County, New York. The proportion of *Ixodes* ticks coinfected with *B. burgdorferi* and *Babesia microti* has ranged from 2% in New Jersey to 19% on Nantucket Island, Massachusetts. In patients with a confirmed tick-borne infection, coinfection rates as high as 39% have been reported. The most commonly recognized coinfection in most of the eastern United States is Lyme disease and babesiosis, accounting for approximately 80% of coinfections.[14]

CLINICAL MANIFESTATIONS

Lyme disease is a progressive infectious disease with a wide array of clinical manifestations. In general, three stages of the illness can be distinguished: early localized disease, early disseminated disease, and persisting late disease.

Infection begins locally in the skin after a feeding tick inoculates *B. burgdorferi*. In most persons, the initial sign of infection is the development of erythema migrans.[15] Even at this early phase of infection, the clinical expression of the disease is highly variable. Some persons are relatively asymptomatic, whereas others experience fever, arthralgias, myalgias, conjunctivitis, meningismus, or multifocal erythema migrans, and still others develop more dramatic signs of infection, including acute meningitis, myocarditis with or without conduction block, hepatitis, myositis, or frank arthritis. Up to 50% of infected persons will progress to symptomatic late disease if not treated during the acute phase of the infection. In the chronic phase of the illness, localized inflammatory processes may occur in one or more organ systems, particularly the nervous system and the musculoskeletal system.

Erythema Migrans

The most common manifestation of early localized Lyme disease is erythema migrans, which occurs in up to 85% of patients and develops 3 to 30 days (typically within the first 7 to 10 days) after the bite. The lesion generally appears at the site of a tick bite and is frequently located around the knees, in the axilla, or in the groin.[16]

Erythema migrans usually begins as a red macule or papule, which expands over the course of days to weeks, presumably as the spirochetes spread centrifugally through the skin [*see Figure 2*]. Secondary cutaneous lesions may develop from hematogenous dissemination of spirochetes. Local symptoms include pruritus, tenderness, or paresthesias but are generally rare or absent in secondary lesions. Erythema migrans may also appear as a target lesion with variable degrees of central clearing and occasionally with vesicular or necrotic areas in the center.

In an observational cohort study in 10 endemic states, 118 patients with microbiologically confirmed erythema migrans presented a median of 3 days after symptom onset. Early erythema migrans commonly had homogeneous or central redness rather than peripheral erythema with partial central clearing. The most common associated symptoms were low-grade fever, headache, neck stiffness, arthralgia, myalgia, or fatigue.[17] Subsequent episodes of erythema migrans have been reported in patients who received appropriate antimicrobial therapy for an initial episode, whereas primary failure of antibiotics is rare, occurring in only about 0.14% of patients.[18]

Lymphocytoma

Borrelial lymphocytoma is an uncommon early manifestation of Lyme disease (occurring in approximately 5% of cases), occurring more often in children than in adults.[19] It is a tumorlike nodule that typically appears on the ear lobe, nipple, or scrotum and is characterized by a dense lymphocytic infiltrate in the dermis or subcutaneous tissue.[20] Borrelial lymphocytoma is usually caused by *B. garinii* and *B. afzelii* and is seen more frequently in Europe than in the United States. The lymphocytoma may occur with other manifestations of infection, such as meningitis, choroiditis, or arthritis. Histopathologically, lymphocytoma may be difficult to differentiate from lymphoma. IgG or IgM antibodies against *B. burgdorferi* are found in the serum of 80% of all patients with borrelial lymphocytoma. Direct detection of *B. burgdorferi* or specific DNA in lesional skin by culture or polymerase chain reaction is helpful to the diagnosis. In a study from Slovenia, 36 cases of borrelial lymphocytoma were detected during the period 1986 to 1990; patients were treated with antibiotic therapy, and all had complete recovery within an average of 3 weeks.[21]

Acrodermatitis Chronica Atrophicans

In European patients, especially elderly women with *B. afzelii* infection, a chronic, slowly progressive skin condition called acrodermatitis chronica atrophicans may develop on sun-ex-

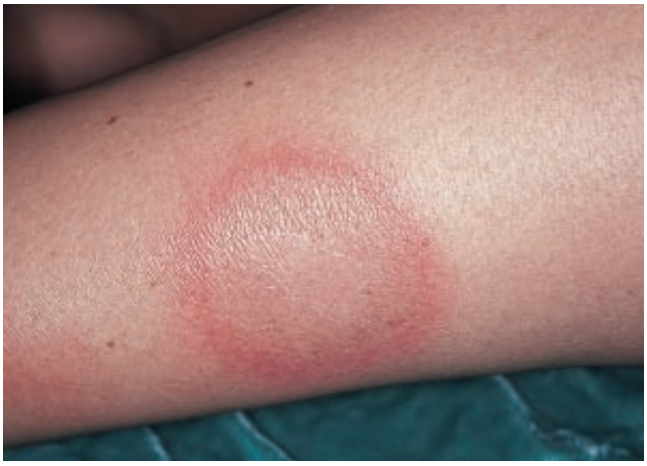


Figure 2 **An erythema migrans lesion has enlarged over several days and now has a red border with clearing in the center.**

posed acral surfaces. The organism has been cultured from such lesions as long as 10 years after the onset of the disease.[22] These lesions may be preceded by erythema migrans and may represent a late or chronic stage of infection. Early lesions have erythematous nodules or plaques with central clearing and involve the extensor areas of the extremities or joints. Later lesions become atrophic and poikilodermatous, resembling scleroderma or lichen sclerosus et atrophicus.

Carditis

Within several weeks after the onset of Lyme disease, about 4% to 10% of untreated patients develop acute cardiac involvement—most commonly, fluctuating degrees of atrioventricular block; occasionally, acute myopericarditis or mild left ventricular dysfunction; and rarely, cardiomegaly or fatal pancarditis.[23] Diffuse T wave changes, ST segment depression, and arrhythmias also are frequently observed. Patients may complain of dizziness, palpitations, dyspnea, chest pain, or syncope.[24,25] Although temporary pacing is frequently required for complete heart block, permanent pacing is rarely needed.[26] Recovery from the acute cardiac manifestations of Lyme disease is usually complete.

In Europe, *B. burgdorferi* has been isolated from endomyocardial biopsy samples from several patients with chronic dilated cardiomyopathy.[27,28] However, this complication has not been observed in the United States.[29] A study from the Netherlands noted improvement in left ventricular ejection fraction in eight of nine *B. burgdorferi*–seropositive patients with idiopathic dilated cardiomyopathy who were treated with antibiotics.[30] Further studies are warranted to clarify the role of *B. burgdorferi* in acute and chronic heart failure.

Musculoskeletal Features

In the United States, arthritis is the predominant manifestation of disseminated *B. burgdorferi* infection, with about 60% of untreated patients developing joint manifestations, usually weeks to years after the initial infection. Months after the onset of illness, patients begin to have intermittent episodes of joint swelling and pain—primarily in the large joints and, occasionally, in the temporomandibular joint.[23] Musculoskeletal features include arthralgia; intermittent episodes of migratory arthritis, usually monoarthritis or asymmetrical oligoarthritis; and chronic arthritis, usually of the knees.[31] Patients with Lyme arthritis usually have higher *Borrelia*-specific antibody titers than patients with any other manifestations of the illness, including late neurologic manifestations.[32] In an observational cohort study of 15 patients with Lyme arthritis, fibromyalgia, or both, symptoms of Lyme arthritis resolved with antibiotic therapy, whereas symptoms of fibromyalgia persisted in 14 of 15 patients treated with antibiotic therapy.[33]

Although most patients respond favorably to antibiotic therapy, about 10% of adults and fewer than 5% of children with arthritis associated with Lyme disease develop inflammatory joint disease that persists for longer than 1 year, which may eventually lead to joint destruction. In about 10% of patients, particularly those with *HLA-DRB1*0401* or related alleles, knee arthritis persists for months or even years.[32] Autoimmunity may develop within the inflammatory milieu of affected joints in these patients because of molecular mimicry between an immunodominant T cell epitope of the outer surface protein A (OspA) of *B. burgdorferi* and human lymphocyte function–associated antigen–1, an adhesion molecule that is highly expressed on T cells in synovium.[34]

Neurologic Involvement

Lyme disease is associated with both acute and chronic neurologic abnormalities, affecting both the central and peripheral nervous systems. All the neurologic syndromes associated with *B. burgdorferi* infection can occur without previous erythema migrans.[35] Clinical data support the hypothesis that *B. burgdorferi* invades the nervous system early in the course of the infection. There is a high frequency of nonspecific complaints that may be referable to central nervous system involvement in patients with erythema migrans. In one series of 314 patients with erythema migrans, 64% had headache and 48% complained of a stiff neck.[36] Additional evidence of early invasion of the nervous system is that within the first 3 months after infection, approximately 12% to 15% of patients experience acute meningitis, cranial neuritis, or painful radiculitis, alone or in combination.[37]

The distinguishing features of meningitis in Lyme disease are evident on CSF analysis: a mild CSF pleocytosis largely consisting of polymorphonuclear leukocytes or mononuclear cells, a modest elevation of the CSF protein level, and a normal CSF glucose level. Meningoencephalitis may be a prominent feature, manifesting as difficulty with memory and concentration and emotional lability.[38] MRI abnormalities have been observed in more severely affected patients. In the absence of local antibody production, the diagnosis of chronic *B. burgdorferi* CNS infection is questionable, although this infection has been demonstrated to occur in rare cases.

Cranial neuropathies may occur with or without meningitis. Any cranial nerve may be affected by Lyme disease, but the seventh nerve (unilaterally and bilaterally) is by far the most frequently affected, with involvement in up to 10% of patients.[38] Lyme disease may be responsible for approximately 25% of new-onset Bell palsy in an endemic area, with the palsy sometimes developing before positive serology for Lyme disease.[39] Bell palsy in Lyme disease is presumably a peripheral neuropathy, given that no antibody has been found in the CSF of some patients who have been tested.

Lyme disease can cause a painful radiculitis that is manifested by neuropathic symptoms such as numbness, tingling, and burning. This radiculoneuropathy may affect the limbs or the trunk. Fifty percent of patients with this radiculitis have associated cranial nerve palsies. The peripheral nerve damage in Lyme disease is usually an axonopathy rather than a demyelinating syndrome.[40]

Acute or subacute myelitis can occur and is associated with spastic paraparesis and CSF pleocytosis.[41] In rare cases, mononeuritis and a Guillain-Barré–like syndrome have been reported. Acute painful radiculoneuritis is the most striking early-stage neurologic syndrome. In the United States, a subtle form of Lyme encephalopathy, which seems to represent CNS infection, has been reported; it manifests predominantly with cognitive abnormalities. Less common neurologic manifestations include sudden sensorineural hearing loss, cerebellitis, intracranial aneurysm, and myelitis.

DIAGNOSIS

In patients in the United States, the diagnosis of Lyme disease is usually made on the basis of the characteristic clinical findings; a history of tick exposure in an area where the disease is endemic, and, except in patients with erythema migrans, an antibody response to *B. burgdorferi* by enzyme-linked immunosorbent assay (ELISA) and Western blot (immunoblot) testing, interpreted according to the criteria of the CDC and the Association of State

and Territorial Public Health Laboratory Directors.[42,43] In endemic areas, the presence of the characteristic rash is often appropriate to trigger initiation of antibiotic therapy without serologic confirmation. Also, antibiotic treatment of early localized disease may blunt the antibody response.

Microscopy

B. burgdorferi, a loosely coiled spirochete that is approximately 0.2 µm wide and 10 to 30 µm long, is readily visible in skin-biopsy specimens observed by phase-contrast or dark-field microscopy. The organism can also be detected with acridine orange, Giemsa, or silver (Warthin-Starry or a modified Dieterle) stains or by fluorescent antibody techniques. In addition to being identified in skin, spirochetes have been observed in other tissues, such as the myocardium, synovium, and the nervous system, but the yields have been very poor because of low numbers of spirochetes. The routine use of skin biopsies of erythema migrans lesions for diagnosis of *B. burgdorferi* infection is limited by the need for special media and by the protracted periods needed for culture growth. Demonstration of *B. burgdorferi* from skin biopsies of erythema migrans by culture or PCR is generally not indicated, except perhaps in cases of reinfection when serology may not be helpful.

Culture

In patients presenting with erythema migrans, culture of the involved skin in Barbour-Stoenner-Kelly medium is virtually 100% specific and reasonably sensitive (57% to 86%) for *B. burgdorferi*.[44] The yield from plasma samples is lower, and only occasionally have CSF samples in patients with meningitis yielded culture growth.

In patients with late-stage infection who develop arthritis, PCR testing is greatly superior to culture in the detection of *B. burgdorferi* in joint fluid.[45] *B. burgdorferi* has not been isolated from the CSF of patients with chronic neuroborreliosis, and *B. burgdorferi* DNA has been detected in CSF samples in only a small number of such patients. The Lyme urine antigen test has provided grossly unreliable results and, therefore, should not be used to support the diagnosis of Lyme disease.[46]

Serology

Within 3 to 4 weeks after the onset of borrelial infection, an increase in the IgM response to one or more spirochetal antigens can be detected in most patients; the IgM response usually peaks after 6 to 8 weeks and then gradually declines. Humoral responses of IgM to other antigens gradually develop as the disease progresses. Although there are differences in antigenic responses between patients from North America and patients from Europe, antibodies against one or more of the major protein antigens (i.e., OspC [23 kd], 31 kd, 34 kd, 60 kd, or 66 kd) develop as the infection continues. Specific IgG and IgA responses gradually increase during the second and third months of infection and, once established, may remain detectable for years.

Some epitopes of antigenic components of *B. burgdorferi*, such as proteins with molecular masses of 41 kd and 60 kd, are common to *Treponema pallidum*, oral treponemes, and even *Escherichia coli*.[47] Such cross-reactivity may result in a significant titer on ELISA screening (see below). False positive reactions have also been reported when sera from patients with juvenile rheumatoid arthritis, rheumatoid arthritis, systemic lupus erythematosus, infectious mononucleosis, or subacute bacterial endocarditis were analyzed for antibodies to *B. burgdorferi*.

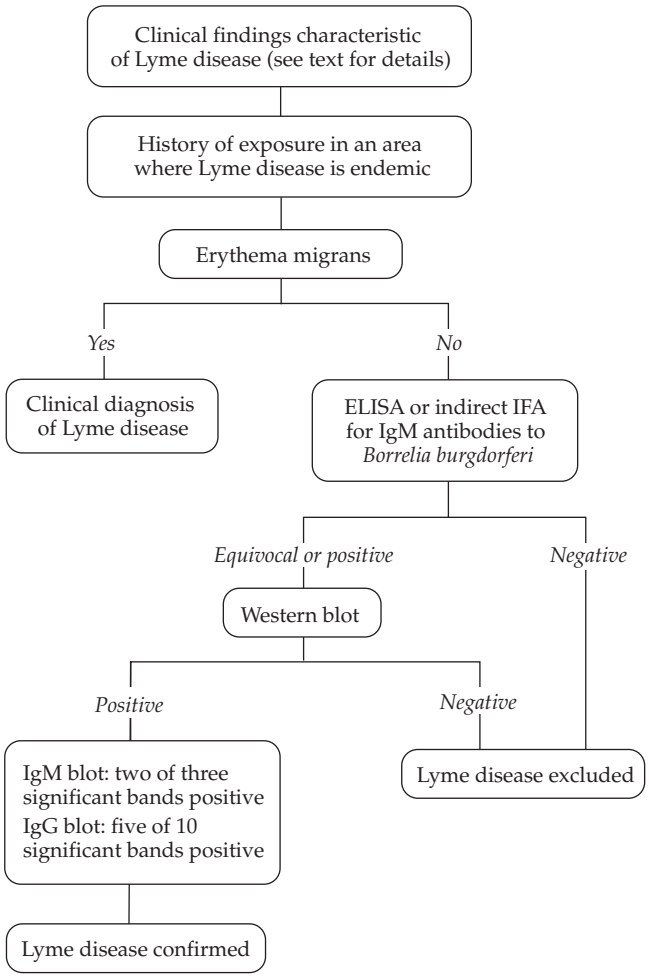


Figure 3 **Diagnosis of Lyme disease. (ELISA—enzyme-linked immunosorbent assay; IFA—immunofluorescence assay)**

Serodiagnostic tests are insensitive during the first several weeks of infection. In the United States, only about 20% to 30% of patients have positive responses, usually of the IgM isotype, during this period.[48] By convalescence, 2 to 4 weeks later, about 70% to 80% of treated patients have seroreactivity. After 1 month, the majority of patients with active infection have IgG antibody responses. In persons who have been ill for longer than 1 month, a positive IgM test result by itself is likely to be false positive; therefore, such a response should not be used to support the diagnosis in this setting.[49] In patients with acute neuroborreliosis, especially those with meningitis, the intrathecal production of IgM, IgG, or IgA antibody against *B. burgdorferi* may often be demonstrated by antibody-capture enzyme immunoassay, but this test is less often positive in patients with chronic neuroborreliosis.[50]

Two-step serologic testing Immunofluorescence assays (IFAs) and ELISAs are the two most commonly used methods for the detection of antibodies to *B. burgdorferi*. ELISAs are more sensitive than IFAs and offer the advantage of easier screening of large numbers of samples. Although *B. burgdorferi* antibody tests are potentially useful and constantly improving, they have limited sensitivity (primarily in early disease) and specificity, and they have not yet been standardized.[51] These limitations have led

to erroneous diagnoses and may have contributed to fundamental misunderstandings of Lyme disease. Therefore, a straightforward two-step serologic approach has been proposed by the CDC [*see Figure 3*]: a positive or equivocal first test, usually an ELISA or indirect IFA, is followed by an immunoblot test on the same serum sample, which can detect IgM and IgG antibodies to individual *B. burgdorferi* antigens.[42,52]

The following two criteria are used for interpretation of the immunoblot test:

- The IgM blot is positive if two of the three following bands are present: 23 kd (OspC), 39 kd, and 41 kd.
- The IgG blot is positive if five of the 10 following bands are present: 18 kd, OspC, 28 kd, 30 kd, 39 kd, 41 kd, 45 kd, 58 kd, 66 kd, and 93 kd.

If the immunoblot test result is negative, the reactive ELISA or IFA result was very probably false positive. Neither ELISA nor immunoblot testing permits detection of fourfold rises in antibody titer (seroconversion).

Some patients lack diagnostic serum levels of specific antibodies but have neurologic involvement and, as a result, have diagnostic levels of antibody in their CSF, because *B. burgdorferi* organisms reaching this immunologically privileged site remain viable and induce a local immune response.[53,54] Intrathecal production of antibody against *B. burgdorferi* may be demonstrated by using the following formula:

$$\frac{\text{Anti–}B.\ burgdorferi\ \text{titer in CSF} \times \text{serum IgG concentration}}{\text{Anti–}B.\ burgdorferi\ \text{titer in serum} \times \text{CSF IgG concentration}}$$

If the ratio is greater than 1, localized production of anti-*Borrelia* antibodies has occurred.

After antibiotic treatment, antibody titers fall slowly, but IgG and even IgM responses may persist for many years after treatment. Thus, even an IgM response cannot be interpreted as a demonstration of recent infection or reinfection unless the appropriate clinical characteristics are present.

New serologic tests An ELISA has been developed on the basis of a conserved immunodominant portion (C6) of the *B. burgdorferi* variable surface antigen (VlsE).[55] This assay can be used in vaccinated and unvaccinated patients. The C6 Lyme ELISA is important because it detects both IgM and IgG antibodies in patients with Lyme disease but not in uninfected vaccine recipients.[56] This assay, when used together with an assay for another immunodominant antigen from OspC, may have the same level of specificity and sensitivity as the two-step approach (see above).[57]

TREATMENT

The primary goals of therapy for Lyme disease are the control of inflammation and the eradication of the infection. Lyme disease is most responsive to antibiotics early in the course of the disease: erythema migrans typically resolves promptly and later-stage disease is prevented.[58,59] Early localized infection that is limited to a single skin lesion, with mild or no systemic symptoms, is uniformly responsive to short-course oral therapy with a number of agents [*see Table 1*]. Of the antibiotics studied to date, the most effective agents for this stage of disease have been amoxicillin, 500 mg three times a day; doxycycline, 100 mg twice daily; and cefuroxime axetil, 500 mg twice a day. Each of these agents is taken for 14 to 21 days.[60] An advantage of doxycycline is its efficacy against human granulocytic ehrlichiosis, a possible coinfection. Amoxicillin should be used in children and pregnant women. About 10% of patients with early-stage Lyme disease experience a Jarisch-Herxheimer reaction (higher fever, redder rash, or greater pain) during the first 24 hours of antibiotic therapy. Patients should be warned of the reaction; if it occurs, the symptoms may be treated with anti-inflammatory agents

Table 1 Antibiotic Therapy for Lyme Disease

Disease Stage or Manifestation	*Agents and Dosage**	*Comments*
Early localized infection with a single skin lesion and mild or no systemic symptoms	Amoxicillin, 500 mg p.o., t.i.d. × 14–21 days Doxycycline, 100 mg p.o., b.i.d. × 14–21 days Cefuroxime axetil, 500 mg p.o., b.i.d. × 14–21 days	Doxycycline is also effective for human granulocytic ehrlichiosis, a possible coinfection; amoxicillin should be used in children and pregnant women; Jarisch-Herxheimer reactions occur during the first 24 hr of antibiotic therapy in about 10% of patients
Carditis with AV nodal block (PR interval > 0.3 sec)	Ceftriaxone, 2 g/day × 14–21 days Penicillin G, 20 million units in four divided doses a day × 14–21 days	Oral regimens are for first-degree AV block < 0.3 sec; for higher-degree AV block (> 0.3 sec), intravenous administration should be used for at least part of the treatment course
Facial palsy with normal CSF findings	Amoxicillin, 500 mg p.o., t.i.d. × 21–30 days Doxycycline, 100 mg p.o., b.i.d. × 21–30 days Cefuroxime axetil, 500 mg p.o., b.i.d. × 21–30 days	Most experts prefer a 30-day course of treatment to reduce the likelihood of late neurologic relapses
Meningitis and other neurologic disorders	Ceftriaxone, 2 g/day × 14–28 days Penicillin G, 20 million units in four divided doses a day × 14–28 days	Clearing of inflammatory CSF findings may lag behind bacteriologic cure
Arthritis	Amoxicillin, 500 mg p.o., t.i.d. ≥ 2 mo Doxycycline, 100 mg p.o., b.i.d. ≥ 2 mo Cefuroxime axetil, 500 mg p.o., b.i.d. ≥ 2 mo Ceftriaxone, 2 g/day ≥ 1 mo Penicillin G, 20 million units in four divided doses a day ≥ 1 mo	Arthritis may persist despite appropriate antibiotic treatment; if PCR tests of joint fluid are negative, patients with persistent arthritis may be treated with anti-inflammatory agents or arthroscopic synovectomy

*All treatment is with single agents.
AV—atrioventricular CSF—cerebrospinal fluid PCR—polymerase chain reaction

such as aspirin. Although carditis resolves spontaneously, patients who have atrioventricular nodal block with a PR interval greater than 0.3 second should receive an intravenous regimen as at least part of the antibiotic course (e.g., ceftriaxone, 2 g/day, or penicillin G, 20 million units in four divided doses a day) for 14 to 21 days. Cardiac monitoring is recommended, but the insertion of a permanent pacemaker is not necessary.[61]

Isolated facial palsies resolve completely or almost completely in nearly all patients. Patients who have facial palsy should undergo a careful neurologic evaluation, including a lumbar puncture and CSF examination. If facial palsy is the only clinical abnormality and CSF findings are normal, current practice is to administer oral antibiotics for 21 to 30 days. Most experts prefer a 30-day course of treatment, however, because of the late neurologic relapses that occasionally occur after shorter courses of therapy. Patients with evidence of active neuroborreliosis should receive a 2- to 4-week course of intravenous ceftriaxone or penicillin G.

Arthritis does not always respond to antibiotic therapy.[62] About 10% of patients in the United States have persistent joint inflammation for months or even several years after 2 months or more of oral antibiotic therapy or 1 month or more of intravenous antibiotic therapy.[62] Patients who have persistent arthritis despite appropriate treatment and negative PCR test results of joint fluid may be treated with anti-inflammatory agents or arthroscopic synovectomy.

B. burgdorferi has not been linked statistically to congenital anomalies, and no increased risk of an adverse outcome of pregnancy has been associated with asymptomatic seropositivity or history of previous Lyme disease.[63] It is appropriate to maintain a lower threshold for institution of aggressive antibiotic therapy for suspected Lyme disease during pregnancy, but women should be reassured that no cases of fetal Lyme disease have occurred with currently recommended antibiotic regimens.

Despite receiving appropriate treatment for Lyme disease, a small percentage of patients continue to have subjective symptoms—primarily, musculoskeletal pain, neurocognitive difficulties, or fatigue—that may last for years. This disabling syndrome, which is sometimes called chronic Lyme disease, is similar to chronic fatigue syndrome or fibromyalgia.[64] In a large study, however, pain and fatigue were no more common in Lyme disease patients than in age-matched control subjects who had not had the disease.[65] In a study of patients with post–Lyme disease syndrome who received either intravenous ceftriaxone for 30 days, followed by oral doxycycline for 60 days, or intravenous and oral placebo preparations for the same duration, there were no significant differences between the two study arms in the percentage of patients who said that their symptoms had improved, gotten worse, or stayed the same.[66] There is no evidence that treatment of asymptomatic seropositive patients is beneficial.

PREVENTION

Primary prevention strategies will help reduce the number of Lyme disease cases, and some strategies may also prevent other tick-borne illnesses, including babesiosis and human granulocytic ehrlichiosis in the United States and tick-borne encephalitis in Europe. The first line of defense is avoidance of tick-infested habitats when possible; use of personal protective measures (e.g., repellents and protective clothing) in tick-infested habitats and checking for and removing attached ticks; and modifications of landscapes in or near residential areas. Tick control (burning or removing vegetation, acaricide use, and deer elimination) reduces *I. scapularis* populations by up to 94%, and acaricide application to wildlife decreases nymphal *I. scapularis* populations by up to 83%. Transmission of borrelial infection requires a period of 24 to 72 hours of tick attachment. Therefore, removal of ticks within 24 hours will usually prevent infection. If an engorged tick is found, a single 200 mg dose of doxycycline administered within 72 hours is effective at preventing the development of Lyme disease.[67] A vaccine for Lyme disease (LYMErix), consisting of recombinant OspA and approved by the Food and Drug Administration for persons 15 to 70 years of age, was introduced in the United States in 1998 but was withdrawn from the market in 2002 because of low sales.

Leptospirosis

Leptospirosis is a worldwide zoonosis acquired by contact with infected animals or by exposure to contaminated soil or freshwater. A variety of animals can be chronically infected and shed viable organisms that can persist in the environment, leading to human infection. The importance of environmental and occupational exposure is reflected by several of the synonyms for this disease: rice-field fever, cane-cutter fever, swamp fever, and mud fever. Clinically, leptospirosis can range from asymptomatic infection to severe multisystem disease with significant mortality. Leptospirosis cases are often undiagnosed because of the protean manifestations of infection, lack of awareness by clinicians, and limited diagnostic tools.[68]

MICROBIOLOGY

The genus *Leptospira* consists of tightly coiled spirochetes that are thin (0.1 μm) and vary in length from 6 to 20 μm [*see Figure 4*]. *Leptospira* can be cultured on artificial media, but initial cultures may take weeks to grow.[69]

In the past, the genus *Leptospira* was divided into two species, the pathogenic *L. interrogans* and the nonpathogenic *L. biflexa,* on the basis of serologic reactivity to the lipopolysaccharide O antigen. A large number of serovars have been identified within each of these species. The classification of the genus *Leptospira* has undergone extensive reorganization with the application of DNA-relatedness techniques. Thirteen named species and four unnamed genomospecies are now recognized.[70] The complete genetic sequence of *L. interrogans* serovar *lai* has been determined.[71]

EPIDEMIOLOGY

A wide variety of mammals can become infected with *Leptospira* and serve as reservoir hosts. Particularly important in human infection are rodents, livestock, and pets.[72,73] Animals are often infected early in life and develop persistent infection of the proximal renal tubules. Infected animals are frequently asymptomatic and shed viable spirochetes in the urine for extended periods, contaminating water and soil. *Leptospira* can persist in a warm, moist environment for several weeks—a fact that correlates with the higher incidence of human infection in tropical areas, particularly during the rainy season. In temperate regions, where survival of the organism in the environment is limited by temperature, the incidence of leptospirosis is seasonal, peaking in the summer and early fall. In the United States, the greatest number of cases have been reported in Hawaii, but cases occur throughout the country. Studies have demonstrated leptospiruria in 41 of 500 healthy dogs in Kansas and 35% of Texas cattle.[74,75]

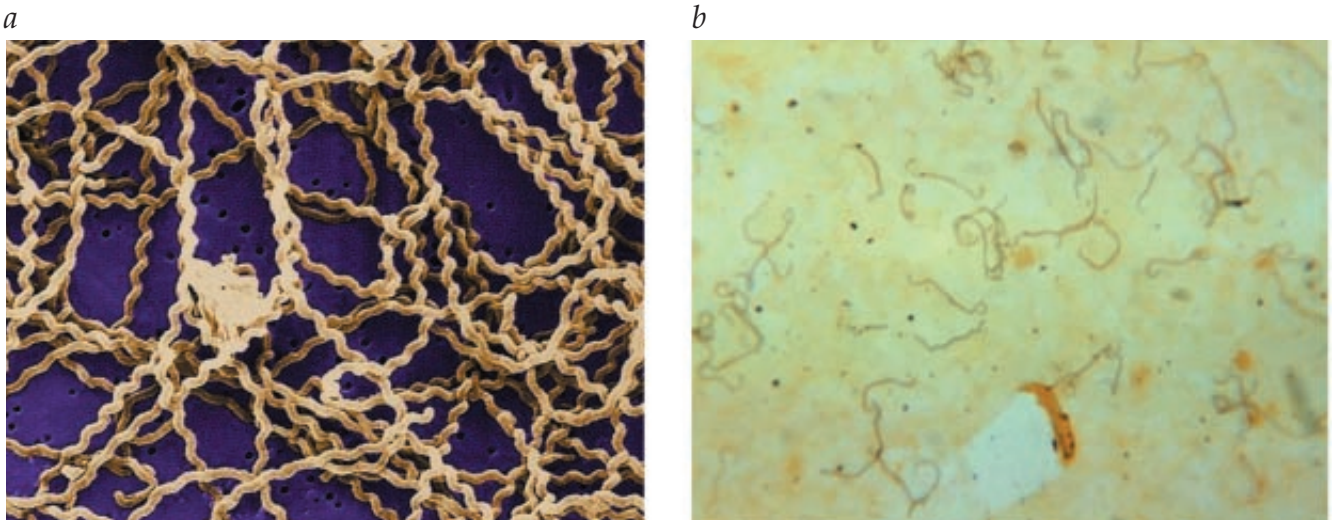


Figure 4 **(*a*) A scanning electron micrograph reveals numerous corkscrew-shaped *Leptospira* spirochetes atop a 0.1 µm polycarbonate filter. (*b*) *Leptospira* spirochetes are visible on a photomicrograph of a liver smear, using a silver-staining technique, taken from a patient with a fatal case of leptospirosis.**

Human infection can result from direct contact with urine or tissues of infected animals; such exposure occurs in veterinarians, dairy workers, hunters, and animal handlers. The more common exposure is to contaminated water and soil.[76] Outbreaks of leptospirosis have been reported related to ecotourism and adventure sporting events.[77,78]

PATHOGENESIS

Leptospires enter the body through minor cuts and abrasions, mucous membranes, and conjunctiva and by inhalation of infected aerosols. Infection spreads throughout the body via the bloodstream. Motility and the ability to migrate through tissues are felt to be important to the pathogenesis of *Leptospira*, allowing the spirochete to establish initial infection and to disseminate to sites of end-organ damage.[71] A variety of hemolysins and phospholipases and a sphingomyelinase have been identified and may have a role in the ability of this organism to move through tissues. The mechanism by which *Leptospira* causes tissue damage is not fully understood, but a systemic vasculitis may facilitate migration of the organism into a variety of tissues, accounting for the broad spectrum of clinical illness.[79] In addition, the host immune response is felt to contribute to tissue damage and the clinical severity of disease.[80] In animals with experimental infections that mimic the more severe icteric Weil disease and hemorrhagic syndromes, the livers and kidneys demonstrate large numbers of leptospires and associated tissue inflammation. Several leptospiral proteins have been identified with homology to host proteins important in hemostasis, which may activate hemolytic pathways and contribute to the hemorrhagic complications seen in severe disease.[81]

DIAGNOSIS

Clinical Manifestations

Infection with *Leptospira* can result in a spectrum of clinical manifestations. Asymptomatic infection is not uncommon, and subclinical or very mild disease that does not lead to medical attention has been reported in several studies.[82,83] When symptomatic, leptospirosis has classically been described as a biphasic illness, with a self-limited septicemic phase followed (although not invariably) by an immune phase. Symptoms of infection develop after an incubation period of 2 to 20 days (mean, 10 days). Illness begins abruptly with high fevers, chills, rigors, headache, myalgias, abdominal pain, nausea, vomiting, diarrhea, and cough. Conjunctival suffusion and muscle tenderness, particularly of the calf and lumbar areas, have been cited as distinctive examination findings. Less common physical findings include lymphadenopathy, splenomegaly, and hepatomegaly.[84] The acute illness lasts 5 to 7 days. *Leptospira* can be recovered from the blood and cerebrospinal fluid during the first week of illness, although symptoms of meningitis are not prominent.

The septicemic phase may be followed by a period of improvement in symptoms and absence of fever lasting several days. The onset of the immune phase of illness coincides with the development of specific antibody to *Leptospira* and the clearance of organisms from the blood and CSF. Leptospires remain detectable in the kidney, urine (leptospiruria), and aqueous humor for several weeks. The immune phase of leptospirosis is often more severe than the septicemic phase and is potentially fatal. The immune phase lasts 4 to 30 days and is characterized by aseptic meningitis, uveitis, iritis, rash, hemorrhagic pneumonitis, and hepatic and renal involvement.[71,84-88] Mortality in patients with severe disease results from multiorgan failure and pulmonary hemorrhage.

The septicemic and immune phases of leptospirosis are illustrated by the meningeal findings. Early in infection, patients often have severe retro-orbital headaches and photophobia. CSF studies reveal a neutrophilic pleocytosis, with counts ranging from 10 to 1,000 cells/µl. *Leptospira* can often be demonstrated in the CSF by culture and PCR. During the immune phase of illness, approximately 25% of patients will develop an aseptic meningitis characterized by headache, vomiting, and signs of meningeal irritation. CSF analysis reveals a lymphocytic pleocytosis, and *Leptospira* can no longer be isolated.

Acute renal failure is reported in 16% to 40% of cases and is associated with tubular necrosis and interstitial nephritis. Jaundice develops in approximately 40% of patients and does not seem to result from hepatocellular damage but, rather, from cholestasis.[84] Serum bilirubin levels may be very elevated (usually to less than 20 mg/dl), with more moderate elevations in

transaminase concentrations (less than 200 IU/L) and mild elevation of the alkaline phosphatase level.[89] Bilirubin levels may take days to weeks to normalize. Pulmonary involvement occurs in up to 70% of cases, and symptoms may range from cough, dyspnea, and hemoptysis to respiratory failure. Severe hemorrhagic pneumonitis and acute pulmonary distress syndrome may occur in the absence of hepatic and renal failure.[90] Radiographs reveal patchy alveolar infiltrates that may progress to large areas of consolidation that are thought to represent pulmonary hemorrhage. Significant pulmonary involvement is a poor prognostic sign and is associated with increased mortality.[84]

Weil disease is the most severe form of *Leptospira* infection, with a mortality of up to 40%. It may develop during the immune phase of a biphasic illness or progress directly from the acute phase without the characteristic brief improvement in symptoms to fulminant illness. Weil disease is characterized by high fevers and the rapid onset of liver failure, acute renal failure, hemorrhagic pneumonitis, cardiac arrhythmias, and circulatory collapse.[71,91]

General Laboratory Studies

During the initial septicemic phase of illness, results of routine laboratory tests are not specific; the neutrophil count may be normal or elevated. During the immune phase of illness, the laboratory values are consistent with the specific end-organ dysfunction. Liver dysfunction is often characterized by markedly elevated serum bilirubin levels with less pronounced increases in serum transaminase and alkaline phosphatase levels. Renal function can deteriorate rapidly, with evidence of interstitial nephritis on biopsy. Renal injury may be compounded by associated hypovolemia. In patients with evidence of meningitis, CSF assays reveal a neutrophilic pleocytosis early and, later, a lymphocytic pleocytosis (usually below 500 cells/μl); modest elevation in protein levels (50 to 100 mg/ml); and a normal glucose level.

Specific Laboratory Studies

Leptospirosis can be diagnosed in three ways: by the direct demonstration of organisms in blood, urine, or tissues; by culture; or serologically.

Microscopy Dark-field microscopy of blood or urine and silver staining of infected tissues have been used to directly visualize leptospires, but both techniques have limited sensitivity and specificity.[92] Approximately 10^4 leptospires/ml are necessary for one cell/field to be visible by dark-field microscopy.[93] Immunostaining has been shown to enhance the ability to directly demonstrate leptospires in tissue; however, the reagents are not commercially available.[94,95]

Culture *Leptospira* grows slowly in the laboratory, which can limit the utility of culture for diagnosis. *Leptospira* can be isolated from a variety of body fluids using commercially available semisolid, albumin-polysorbate media. Antibiotic use may limit the sensitivity of culture, and the timing of specimen collection is important. During the first 7 to 10 days of illness, the spirochete can be recovered from blood, particularly during periods of fever, and from CSF. The culture media should be inoculated with several drops of the patient's blood or CSF at the bedside. Alternatively, blood specimens can be collected in heparin or sodium oxalate (citrate anticoagulation is inhibitory); these should be inoculated within 24 hours. Recovery of leptospires from the urine begins after the first week of illness, and as with blood and CSF, inoculation of media should be done promptly (in less than 1 hour after collection). The cultures are incubated at 30° C and are not reported as negative until after a minimum of 6 to 8 weeks. Identification at the species level is available from a few reference laboratories.

Serology The mainstay for the diagnosis of leptospirosis has been by serology, using a microscopic agglutination test (MAT).[96] The end point is the highest dilution of serum in which 50% agglutination occurs. The MAT uses live organisms, which are coincubated with the patient's serum and examined by dark-field microscopy for agglutination. A fourfold rise in paired titers or a single titer of greater than 1:800 in a patient with an appropriate history supports the diagnosis. Many patients will have a negative test result during the acute illness, and seroconversion may be delayed as long as 30 days after the onset of clinical illness. The MAT is technically complex to perform and to interpret. In addition, live cultures must be maintained for all the serovars required as antigens.

Because of the complexity of the MAT, rapid screening tests for leptospiral antibody have been developed. IgM antibody becomes detectable during the first week of illness and is more sensitive than MAT when the first specimen is taken early in the illness. Commercially available IgM dipstick assays have been shown to perform well.[97,98]

DNA amplification An attractive method to confirm a diagnosis of leptospirosis is PCR, which is more sensitive than culture and is particularly useful in early infection, before significant antibody titers develop. *Leptospira* DNA has been amplified from blood, urine, and various tissues.[94,99] However, the PCR test is currently limited to research laboratories.

TREATMENT

Antibiotic therapy does not appear to have a clear advantage over placebo for treatment of leptospirosis but does tend to reduce mortality, duration of fever, length of hospitalization, and extent of leptospiruria.[100] Most practitioners treat patients with severe disease with parenteral penicillin (1.5 million units every 6 hours), ampicillin (1 g every 6 hours), or ceftriaxone (1 g daily), each for 7 days. Mild infections can be treated with oral doxycycline (100 mg every 12 hours), ampicillin, or amoxicillin. Jarisch-Herxheimer reactions have been reported after treatment with penicillin and ampicillin.[101,102]

PREVENTION

Prevention of leptospirosis can be achieved by limiting high-risk exposure with appropriate protective measures. These include not swimming in potentially contaminated freshwater, wearing rubber gloves and goggles when handling animals, and not walking barefoot. Weekly doxycycline prophylaxis has been shown to be effective at preventing symptomatic infection in persons at high risk for exposure.[103]

Relapsing Fever

Relapsing fever, caused by spirochetes of the *Borrelia* genus, is a febrile illness characterized by recurrent episodes of fever and septicemia separated by afebrile periods. Two forms of the disease are recognized: louse-borne and tick-borne.

Figure 5 ***Ornithodorus*, or soft tick, is the vector of the *Borrelia* species that causes tick-borne relapsing fever. Adult ticks are approximately 2.5 mm in size—about the size of a sesame seed.**

EPIDEMIOLOGY

Louse-borne relapsing fever (LBRF) is caused by infection with *B. recurrentis*. This spirochete is transmitted from person to person by the human body louse (*Pediculus humanus*), and humans are the only known reservoir for this organism.[104] LBRF usually occurs in epidemics that are associated with social catastrophes, such as war and famine, which foster the conditions (e.g., crowding and poor hygiene) that allow lice to flourish and spread from person to person. In addition, LBRF is endemic in areas of central and eastern Africa and Peru.[105] Humans become infected by crushing the lice, which releases the spirochetes and permits them to penetrate the skin or mucous membranes.

Tick-borne relapsing fever (TBRF) is caused by at least 15 different species of *Borrelia*, each of which is associated with a distinct member of the soft ticks of the genus *Ornithodoros*. In contrast to *Dermacentor* and *Ixodes* ticks, which are teardrop shaped with mouth parts visible dorsally, *Ornithodoros* ticks have an oval body shape and mouth parts located on the ventral surface (not visible dorsally) [*see Figure 5*]. Adult *Ornithodoros* ticks are about the size of a sesame seed. Animal reservoirs for these *Borrelia* species include rodents and small animals, such as squirrels, rabbits, chipmunks, owls, and lizards. *Ornithodoros* ticks are obligate blood feeders that typically inhabit caves, decaying wood, rodent burrows, and animal shelters. Their bite often goes undetected because they feed rapidly (over 5 to 20 minutes) at night and have a painless bite.[106] Infection of humans occurs when the tick releases saliva or excrement during feeding. TBRF has been reported worldwide, with the exception of Antarctica, Australia, and certain areas in the southwestern Pacific. Most cases in the United States occur west of the Mississippi River, particularly in the mountain and high-desert areas.[107] Human infection occurs with activity that brings the person into the tick's environment (e.g., camping and cave exploration).

PATHOGENESIS

After entering the bloodstream, *Borrelia* spirochetes multiply and can reach levels of 10^5 to 10^8 organisms/μl. These periods of spirochetemia are associated with fever and systemic symptoms. With the development of a specific antibody response, the spirochetes are cleared from the blood, becoming sequestered in internal organs, and patients become asymptomatic. In response to immune pressure, the *Borrelia* organisms undergo modifications of their outer-surface proteins that allow the spirochetes to reemerge into the circulation; reemergence results in recurrence of symptoms.[108] This process of antigenic variation in response to specific antibody formation can occur for a number of cycles and is responsible for the relapsing course of clinical disease. The number of febrile relapses is higher for tick-borne infection than for louse-borne infection.

CLINICAL MANIFESTATIONS

The incubation period of relapsing fever can range from 4 to 18 days and averages about 1 week. In both TBRF and LBRF, illness begins with the abrupt onset of fevers (usually above 39° C [102.2° F]), rigors, myalgias, arthralgias, and severe headache.[107,109] Physical findings include conjunctival suffusion, petechiae, abdominal tenderness with hepatosplenomegaly, and altered sensorium. A truncal rash that can be petechial, macular, or papular is common during the primary febrile episode. Almost a third of patients develop neurologic complications, including meningitis, cranial nerve palsies, coma, and seizures.

The primary febrile episode is unremitting until it terminates abruptly after 3 to 6 days. After a period of well-being lasting 7 to 10 days, symptoms recur. Each relapse is usually shorter in duration and less severe in intensity. LBRF often is associated with a single relapse, whereas TBRF often is associated with multiple symptomatic relapses.

Mortality from relapsing fever ranges from 2% to 40% and is higher with LBRF. Death is often from complications of myocarditis, with associated arrhythmias, hepatic failure, and cerebral hemorrhage.

DIAGNOSIS

The febrile periods of illness are associated with high-level spirochetemia, and dark-field microscopy or Giemsa or Wright staining of blood smears can directly demonstrate *Borrelia* [*see Figure 6*]. Detection of spirochetes is improved with lysed thick smears or acridine orange staining.[110,111] Spirochetes are rarely seen on blood smears obtained when the patient is afebrile. Culture is difficult and not readily available. Serology is of limited value, in part because of the antigenic variability of the organisms. Currently available assays target antigens common to other spirochetes and bacteria and do not have a high sensitivity. Serologic tests for both syphilis and Lyme disease may be positive.[112]

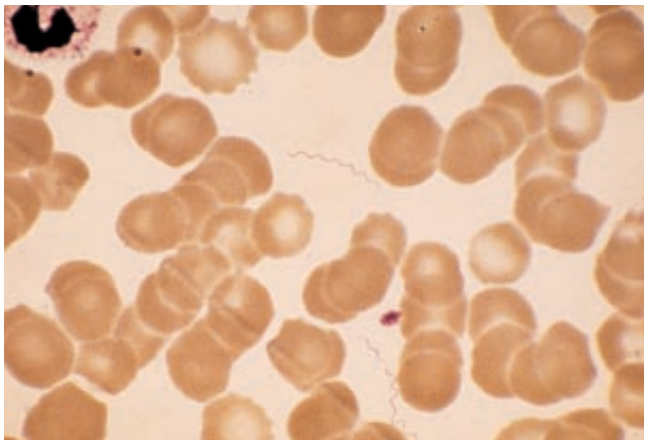


Figure 6 **In this thin Wright stain of peripheral blood from a patient in the febrile stage of relapsing fever, several spirochetes can be seen.**

TREATMENT

The *Borrelia* species that cause relapsing fever are susceptible in vitro to penicillins, tetracyclines, macrolides, cephalosporins, and chloramphenicol.[113] LBRF can be successfully treated with a single oral dose of 500 mg of tetracycline. Young children and pregnant women may be treated with a single dose of erythromycin (500 mg).[114] TBRF requires a 5- to 10-day course of antibiotics; shorter courses of treatment are associated with a higher rate of treatment failures. Tetracycline and erythromycin are felt to be equally effective. In patients with meningitis or encephalitis, a 14-day course of parenteral penicillin G, ceftriaxone, or cefotaxime is generally used. Jarisch-Herxheimer reactions are common within the first several hours after initiating antibiotic therapy and may be severe, with rigors, high fever, and hypotension. For this reason, it is recommended that all patients be kept under observation for several hours after the first dose of antibiotics.

Rat-Bite Fever

Rat-bite fever is a systemic febrile illness that results from infection with *Streptobacillus moniliformis* in North America and Europe and with *Spirillum minus* in Asia. These bacteria are transmitted to humans by the bite of a rat or other small rodent. Illnesses from the two organisms are similar, but each has unique clinical features [*see Table 2*].

Table 2 Characteristics of Rat-Bite Fever

	Streptobacillus moniliformis *Infection*	Spirillum minus *Infection*
Incubation period	10 days (1–22 days)	1–4 wk
Bite site	Healed by the time symptoms appear	Swollen, painful; may ulcerate
Onset	Chills, fever, joint symptoms	Fever without joint symptoms
Joint symptoms	Asymmetrical polyarthritis common; occasional frank septic arthritis	Usually absent
Course	Relapsing course uncommon	Relapsing course common

STREPTOBACILLUS MONILIFORMIS INFECTION

S. moniliformis is a pleomorphic gram-negative bacillus that is frequently part of the nasopharyngeal flora of rats and other small rodents.[115] Infection typically is transmitted by the bite or scratch of rats (also mice and squirrels) or by carnivores that prey on rodents. Despite the term rat-bite fever, approximately 30% of rat-bite fever cases occur in patients who have no history of being bitten by rats. Transmission may occur from handling rats at home and the workplace (pet shops, laboratories). Consequently, rat-bite fever should be in the differential diagnosis of unexplained febrile illness or sepsis in patients with any history of rat exposure. Illness can also result from oral ingestion of the organism—for example, eating food that is contaminated with rodent droppings (Haverhill fever).

Clinical Features

The usual incubation period after a rodent bite is less than 10 days (range, 1 to 22 days). Onset of illness is abrupt, with fever, chills, headache, vomiting, and severe migratory arthralgias and myalgias.[116] In contrast to *S. minus* infection, ulceration at the initial bite site and regional lymphadenopathy are usually absent with *S. moniliformis* infection. Within several days after the onset of the fever, a nonpruritic, maculopapular, petechial, vesicular or pustular rash develops on the extremities, with involvement of the palms and soles. An asymmetrical polyarthritis occurs in a significant number of patients; frank septic arthritis may develop. The large joints are most commonly involved (i.e., knees, ankles, wrists, shoulders, and hips).[117]

Fever associated with *S. moniliformis* infection tends to subside after several days, even without specific therapy. The other clinical manifestations usually resolve over the next several weeks. Rare, potentially fatal complications include cardiac involvement (e.g., endocarditis, myocarditis, and pericarditis), meningitis, pneumonia, and abscesses in a variety of solid organs. Two cases of fulminant sepsis and death in previously healthy persons have been reported.[118]

Diagnosis

S. moniliformis can be readily grown using enriched media. The mainstay of diagnosis is culture. In addition, pleomorphic gram-negative bacilli can be demonstrated on Gram stains of blood, joint fluid, and abscess aspirates. Antibody detection by ELISA and PCR of bacterial 16S ribosomal RNA in tissue samples have been described, but these tests are currently still limited to research centers.[119]

SPIRILLUM MINUS INFECTION

S. minus, a short, thick, gram-negative, tightly coiled spirochete, is the cause of rat-bite fever in Asia. Approximately 25% of rats in endemic areas are carriers of *S. minus,* and the major route of transmission of infection to humans is through the occurrence of rat bites.[120]

Clinical Features

The original rat bite heals promptly, but 1 to 4 weeks later, the site of the bite becomes swollen, indurated, and painful, and it may ulcerate. In contrast to *S. moniliformis* infection, *S. minus* infection usually includes regional lymphadenitis. Headache, fever, chills, and malaise accompany the formation of the ulcer.[121] A sparse, dusky-red maculopapular rash appears on the trunk and extremities in many cases. The severe arthritis and myalgias seen in *S. moniliformis* infection are rare in disease caused by *S. minus*.

Symptoms of *S. minus* infection subside after a few days but recur several days later. Without specific antibiotic therapy, fevers lasting 3 to 4 days recur at regular intervals between afebrile periods lasting 3 to 9 days. This relapsing course can persist for several months; in rare instances, fever relapses have occurred for a year or more. More serious complications include endocarditis, myocarditis, meningitis, conjunctivitis, hepatitis, and pleural effusions. Mortality without antibiotic therapy ranges from 6% to 10%.

Diagnosis

Leukocytosis is often present with *S. minus,* and up to 50% of patients will have a false positive test for syphilis. *S. minus* cannot be grown on artificial media. Initial diagnosis relies on visualization of spirochetes in blood, exudates, or lymph node tissue on culture staining or dark-field microscopy. Intraperitoneal inoculation of mice has been used to establish the diagnosis.[121,122] No specific serologic test is available.

TREATMENT

Both *S. moniliformis* and *S. minus* infections are effectively treated with penicillin given for 10 to 14 days.[123] Initial therapy for severely ill patients should be parenteral, but once the patient becomes stable, the course may be completed with oral penicillin or ampicillin. A Jarisch-Herxheimer reaction has been reported with initial therapy for *S. minus* infection. Oral tetracycline is appropriate for penicillin-allergic patients.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Steere AC, Malawista SE, Snydman DR, et al: Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. Arthritis Rheum 20:7, 1977

2. Burgdorfer W, Barbour AG, Hayes S, et al: Lyme disease: a tick-borne spirochetosis. Science 216:1317, 1982

3. Lyme disease—United States, 1999. MMWR Morb Mortal Wkly Rep 50:181, 2001

4. Melski JW: Lyme borreliosis. Semin Cutan Med Surg 19:10, 2000

5. Nadelman RB, Wormser GP: Lyme borreliosis. Lancet 352:557, 1998

6. Lastavica CC, Wilson ML, Berardi VP, et al: Rapid emergence of a focal epidemic of Lyme disease in coastal Massachusetts. N Engl J Med 320:133, 1989

7. Steere AC, Taylor E, Wilson ML, et al: Longitudinal assessment of the clinical and epidemiologic features of Lyme disease in a defined population. J Infect Dis 154:295, 1986

8. Magnarelli LA, Anderson JF, Apperson CS, et al: Spirochetes in ticks and antibodies to *Borrelia burgdorferi* in white-tailed deer from Connecticut, New York State, and North Carolina. J Wildl Dis 22:178, 1976

9. Bosler EM, Coleman JL, Benach JL, et al: Natural distribution of the *Ixodes dammini* spirochete. Science 220:321, 1983

10. Magnarelli LA, Anderson JF: Ticks and biting insects infected with the etiologic agent of Lyme disease, *Borrelia burgdorferi*. J Clin Microbiol 26:14, 1988

11. Spielman A, Wilson ML, Levine JF, et al: Ecology of *Ixodes dammini*–borne human babesiosis and Lyme disease. Ann Rev Entomol 30:439, 1985

12. Spielman A: The emergence of Lyme disease and human babesiosis in a changing environment. Ann N Y Acad Sci 740:146, 1994

13. Orloski KA, Hayes EB, Campbell GL, et al: Surveillance for Lyme disease—United States, 1992–98. MMWR CDC Surveill Summ 49(SS-3):1, 2000

14. Krause PJ, Telford SR III, Spielman A, et al: Concurrent Lyme disease and babesiosis: evidence for increased severity and duration of illness. JAMA 275:1657, 1996

15. Luft BJ, Dattwyler RJ: Lyme borreliosis. Curr Clin Top Infect Dis 11:56, 1989

16. Barbour AG, Hayes SF: Biology of *Borrelia* species. Microbiol Rev 50:381, 1986

17. Hassler D, Maiwald M: [Reinfection with *Borrelia burgdorferi* in an immunocompetent patient]. Dtsch Med Wochenschr 119:338, 1994

18. Maraspin V, Lotric-Furlan S, Strle F: Development of erythema migrans in spite of treatment with antibiotics after a tick bite. Wien Klin Wochenschr 114:616, 2002

19. Hovmark A, Asbrink E, Olsson I: The spirochetal etiology of lymphadenosis benigna cutis solitaria. Acta Dermato-Venereologica 66:479, 1986

20. Van Dam AP, Kuiper H, Vos K, et al: Different genospecies of *Borrelia burgdorferi* are associated with distinct clinical manifestations of Lyme borreliosis. Clin Infect Dis 123:603, 1991

21. Strle F, Pleterski-Rigler D, Stanek G, et al: Solitary borrelial lymphocytoma. Infection 20:201, 1992

22. Asbrink E, Hovmark A: Successful cultivation of spirochetes from skin lesions of patients with erythema chronicum migrans Afzelius and acrodermatitis chronica atrophicans. Acta Pathol Microbiol Immunol Scand [B] 93:161, 1985

23. Steere AC: Lyme disease. N Engl J Med 321:586, 1989

24. van der Linde MR: Lyme carditis: clinical characteristics of 105 cases. Scand J Infect Dis 77:81, 1991

25. van der Linde MR, Ballmer PE: Lyme carditis. Aspects of Lyme Borreliosis. Weber K, Burgdorfer W, Schierz G, Eds. Springer-Verlag, Berlin, 1993, p 131

26. McAlister HF, Klementowicz PT, Andrews C, et al: Lyme carditis: an important cause of reversible heart block. Ann Intern Med 110:339, 1989

27. Stanek G, Klein J, Bittner R, et al: Isolation of *Borrelia burgdorferi* from the myocardium of a patient with longstanding cardiomyopathy. N Engl J Med 322:249, 1990

28. Lardieri G, Salvi A, Camerini F, et al: Isolation of *Borrelia burgdorferi* from myocardium. Lancet 342:490, 1993

29. Sonnesyn SW, Diehl SC, Johnson RC, et al: A prospective study of the seroprevalence of *Borrelia burgdorferi* infection in patients with severe heart failure. Am J Cardiol 76:97, 1995

30. Gasser R, Dusleag J, Reisinger E, et al: Reversal by ceftriaxone of dilated cardiomyopathy *Borrelia burgdorferi* infection. Lancet 339:1174, 1992

31. Steere AC, Schoen RT, Taylor E: The clinical evolution of Lyme arthritis. Ann Intern Med 107:725, 1987

32. Akin E, McHugh GL, Flavell RA, et al: The immunoglobin (IgG) antibody response to OspA and OspB correlates with severe and prolonged Lyme arthritis and the IgG response to P35 correlates with mild and brief arthritis. Infect Immun 67:173, 1999

33. Dinerman H, Steere AC: Lyme disease associated with fibromyalgia. Ann Intern Med 117:281, 1992

34. Steere AC, Gross D, Meyer AL, et al: Autoimmune mechanisms in antibiotic treatment-resistant lyme arthritis. J Autoimmun 16:263, 2001

35. Reik L, Burgdorfer W, Donaldson JO: Neurologic abnormalities in Lyme disease without erythema chronicum migrans. Am J Med 12:475, 1986

36. Steere AC, Hutchinson GJ, Craft J, et al: The early clinical manifestations of Lyme disease. Ann Intern Med 99:76, 1983

37. Pachner AR, Steere A: The triad of neurologic manifestations of Lyme disease: meningitis, cranial neuritis, and radiculoneuritis. Neurology 35:47, 1985

38. Oschmann P, Dorndorf W, Hornig C, et al: Stages and syndromes of neuroborreliosis. J Neurol 245:262, 1998

39. Halperin JJ, Golightly M: Lyme borreliosis in Bell's palsy. Long Island Neuroborreliosis Collaborative Study Group. Neurology 42:1268, 1992

40. Hansen K, Lebech AM: The clinical and epidemiologic profile of Lyme neuroborreliosis in Denmark 1985–1990. Brain 115:399, 1992

41. Logigian EL, Kaplan RF, Steere AC: Chronic neurologic manifestations of Lyme disease. N Engl J Med 323:1438, 1990

42. Case definitions for infectious conditions under public health surveillance. MMWR Recomm Rep 46(RR-10):1, 1997

43. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. MMWR Morb Mortal Wkly Rep 44:590, 1995

44. Berger BW, Johnson RC, Kodner C, et al: Cultivation of *Borrelia burgdorferi* from erythema migrans lesions and perilesional skin. J Clin Microbiol 30:359, 1992

45. Nocton JJ, Dressler F, Rutledge BJ, et al: Detection of *Borrelia burgdorferi* DNA by polymerase chain reaction in synovial fluid from patients with Lyme arthritis. N Engl J Med 330:229, 1994

46. Klempner MS, Schmid C, Hu L, et al: Intralaboratory reliability of serologic and urine testing for Lyme disease. Am J Med 110:217, 2001

47. Magnarelli LA, Miller JN, Anderson JF, et al: Cross-reactivity of non-specific treponemal antibody in serologic tests for Lyme disease. J Clin Microbiol 28:1276, 1990

48. Engstrom SM, Shoop E, Johnson RC: Immunoblot interpretation criteria for serodiagnosis of early Lyme disease. J Clin Microbiol 33:419, 1995

49. Dressler F, Whalen JA, Reinhardt BN, et al: Western blotting in the serodiagnosis of Lyme disease. J Infect Dis 167:392, 1993

50. Steere AC, Berardi VP, Weeks KE, et al: Evaluation of the intrathecal antibody response to *Borrelia burgdorferi* as a diagnostic test for Lyme neuroborreliosis. J Infect Dis 161:1203, 1990

51. Jespersen DJ, Smith TF, Rosenblatt JE, et al: Comparison of the Borrelia DotBlot G, MarDx, and VIDAS enzyme immunoassays for detecting immunoglobulin G antibodies to *Borrelia burgdorferi* in human serum. J Clin Microbiol 40:4782, 2002

52. Norman GL, Antig JM, Bigaignon G, et al: Serodiagnosis of Lyme borreliosis by *Borrelia burgdorferi sensu stricto, B garinii*, and *B afzelii* western blots (immunoblots). J Clin Microbiol 34:1732, 1996

53. Dattwyler RJ, Volkman DJ, Halperin JJ, et al: Specific immune responses in Lyme borreliosis: characterization of T cell and B cell responses to *Borrelia burgdorferi*. Ann N Y Acad Sci 539:93, 1988

54. Stiernstedt G, Gustafsson R, Karlsson M, et al: Clinical manifestations and diagnosis of neuroborreliosis. Ann N Y Acad Sci 539:46, 1988

55. Philipp MT, Marques AR, Fawcett PT, et al: C6 test as an indicator of therapy outcome for patients with localized or disseminated lyme borreliosis. J Clin Microbiol 41:4955, 2003

56. Marques AR, Martin DS, Philipp MT: Evaluation of the C6 peptide enzyme-linked immunosorbent assay for individuals vaccinated with the recombinant OspA vaccine. J Clin Microbiol 40:2591, 2002

57. Bacon RM, Biggerstaff BJ, Schriefer ME, et al: Serodiagnosis of Lyme disease by kinetic enzyme-linked immunosorbent assay using recombinant V1sE1 or peptide antigens of *Borrelia burgdorferi* compared with 2-tiered testing using whole-cell lysates. J Infect Dis 187:1187, 2003

58. Steere AC, Hutchinson GJ, Rahn DW, et al: Treatment of the early manifestations of Lyme disease. Ann Intern Med 99:22, 1983

59. Dattwyler RJ, Volkman DJ, Connaty SM, et al: Amoxycillin plus probenecid versus doxycycline for treatment of erythema migrans borreliosis. Lancet 336:1404, 1990

60. Massarotti EM, Luger SW, Rahn DW, et al: Treatment of early Lyme disease. Am J Med 92:396, 1992

61. van der Linde MR: Lyme carditis: clinical characteristics of 105 cases. Scand J Infect Dis 77:81, 1991

62. Steere AC, Levin RE, Molloy PJ, et al: Treatment of Lyme arthritis. Arthritis Rheum 37:878, 1994

63. Strobino BA, Williams CL, Abid S, et al: Lyme disease and pregnancy outcome: a prospective study of two thousand prenatal patients. Am J Obstet Gynecol 169:367, 1993

64. Dinerman H, Steere AC: Lyme disease associated with fibromyalgia. Ann Intern Med 117:281, 1992

65. Seltzer EG, Gerber MA, Cartter ML, et al: Long-term outcomes of persons with Lyme disease. JAMA 283:609, 2000

66. Klempner MS, Hu LT, Evans J, et al: Two controlled trials of antibiotic treatment in

patients with persistent symptoms and a history of Lyme disease. N Engl J Med 345:85, 2001

67. Nadelman RB, Nowakowski J, Fish D, et al: Prophylaxis with single-dose doxycycline for the prevention of Lyme disease after an *Ixodes scapularis* tick bite. N Engl J Med 345:79, 2001

68. Plank R, Dean D: Overview of the epidemiology, microbiology, and pathogenesis of *Leptospira* spp. in humans. Microb Infect 2:1265, 2000

69. Levett PN: *Leptospira* and *Leptonema.* Manual of Clinical Microbiology, 8th ed. Murray PR, Baron EJ, Jorgensen JH, et al, Eds. American Society for Microbiology Press, Washington, DC, 2003 p 929

70. Brenner DJ, Kaufmann AF, Sulzer KR: Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and four new *Leptospira* genomospecies. Int J Syst Bacteriol 49:839, 1999

71. Bharti AR, Nally JE, Ricaldi JN, et al: Leptospirosis: a zoonotic disease of global importance. Lancet Infect Dis 3:757, 2003

72. Richardson DJ, Gauthier JL: A serosurvey of leptospirosis in Connecticut peridomestic wildlife. Vector Borne Zoonotic Dis 3:187, 2003

73. Boutilier P, Carr A, Schulman RL: Leptospirosis in dogs: a serologic survey and case series 1996 to 2001. Vet Ther 4:387, 2003

74. Harkin KR, Roshto YM, Sullivan JT, et al: Comparison of polymerase chain reaction assay, bacteriologic culture and serologic testing in assessment of prevalence of urinary shedding of leptospires in dogs. J Am Vet Med Assoc 222:1230, 2003

75. Talpada MD, Garvey N, Sprowls R, et al: Prevalence of leptospiral infection in Texas cattle: implications for transmission to humans. Vector Borne Zoonotic Dis 3:141, 2003

76. Farr RW: Leptospirosis. Clin Infect Dis 21:1, 1995

77. Boland M, Sayers G, Coleman T, et al: A cluster of leptospirosis cases in canoeists following a competition on the river Liffey. Epidemiol Infect 132:195, 2004

78. Sejvar J, Bancroft E, Winthrop K, et al: Leptospirosis in "Eco-Challenge" athletes, Malaysian Borneo, 2000. Emerg Infect Dis 9:702, 2003

79. Arean VM: The pathologic anatomy and pathogenesis of fatal human leptospirosis (Weil's disease). Am J Pathol 40:393, 1962

80. Abdulkader RC, Daher EF, Camargo ED, et al: Leptospirosis severity may be associated with the intensity of humoral immune response. Revis Inst Med Trop Sao Paulo 44:79, 2002

81. Ren SX, Fu G, Jiang XG, et al: Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. Nature 422:888, 2003

82. Bovet P, Yersin C, Merien F, et al: Factors associated with clinical leptospirosis: a population based case control study in the Seychelles (Indian Ocean). Int J Epidemiol 28:583, 1999

83. Ashford DA, Kaiser RM, Spiegel RA, et al: Asymptomatic infection and risk factors for leptospirosis in Nicaragua. Am J Trop Med Hyg 63:249, 2000

84. Katz AR, Ansdell VE, Effler PV, et al: Assessment of the clinical presentation and treatment of 353 cases of laboratory-confirmed leptospirosis in Hawaii, 1974–1998. Clin Infect Dis 33:1834, 2001

85. Carvalho CR, Bethlem EP: Pulmonary complications of leptospirosis. Clin Chest Med 23:469, 2002

86. Sitprija V, Losuwanrak K, Kanjanabuch T: Leptospiral nephropathy. Semin Nephrol 23:42, 2003

87. Bertherat E, Renaut A, Nabias R, et al: Leptospirosis and Ebola virus infection in five gold-panning villages in northeastern Gabon. Am J Trop Med Hyg 60:610, 1999

88. Faine S, Adler B, Bolin C, et al: *Leptospira* and Leptospirosis, 2nd ed. MedSci, Melbourne, 1999

89. Vinetz JM, Glass GE, Flexner CE, et al: Sporadic urban leptospirosis. Ann Intern Med 125:794, 1996

90. Leptosporosis associated with outbreak of acute febrile illness and pulmonary haemorrhage, Nicaragua, 1995. The Epidemic Working Group. Lancet 347:535, 1996

91. Feigin RD, Anderson DC: Human leptospirosis. CRC Crit Rev Clin Lab Sci 5:413 1975

92. Vijayachari P, Sugunan AP, Umapathi T, et al: Evaluation of darkground microscopy as a rapid diagnostic procedure in leptospirosis. Indian J Med Res 114:54, 2001

93. Turner LH: Leptospirosis III. Maintenance, isolation and demonstration of leptospires. Trans R Soc Trop Med Hyg 64:623, 1970

94. Brown PD, Carrington DG, Gravekamp C, et al: Direct detection of leptospiral material in human postmortem samples. Res Microbiol 154:581, 2003

95. de Brito T, Morais CF, Yasuda PH, et al: Cardiovascular involvement in human and experimental leptospirosis: pathologic findings and immunohistochemical detection of leptospiral antigen. Ann Trop Med Parasitol 81:207, 1987

96. Bajani MD, Ashford DA, Bragg SL, et al: Evaluation of four commercially available rapid serologic tests for diagnosis of leptospirosis. J Clin Microbiol 41:803, 2003

97. Levett PM, Branch SL, Whittington CI, et al: Two methods for the rapid serological diagnosis of acute leptospirosis. Clin Diagn Lab Immunol 8:349, 2001

98. Smits HL, Ananyina YV, Chereshsky A, et al: International multicenter evaluation of the clinical utility of a dipstick assay for detection of *Leptospira*-specific immunoglobulin M antibodies in human serum specimens. J Clin Micro 37:2904, 1999

99. Lester JW, LeFebvre RB: Detection of *Leptospira interrogans*. Methods Mol Biol 216:193, 2003

100. Guidugli F, Castro AA, Atallah AN: Antibiotics for treating leptospirosis. Cochrane Database Syst Rev 2:CD001306, 2000

101. Edwards CN, Levett PN: Prevention and treatment of leptospirosis. Expert Rev Anti Infect Ther 2:293, 2004

102. Faucher JF, Hoen B, Estavoyer JM: The management of leptospirosis. Exp Opin Pharm 5:819, 2004

103. Guidugli F, Castro AA, Atallah AN: Antibiotics for preventing leptospirosis. Cochrane Database Syst Rev 4:CD001305, 2000

104. Bryceson AD, Parry EH, Perine PL, et al: Louse-borne relapsing fever. Q J Med 39:129, 1970

105. Borgnolo G, Hailu B, Ciancarelli A, et al: Louse-borne relapsing fever: a clinical and an epidemiologic study of 389 patients in Asella Hospital, Ethiopia. Trop Geogr Med 45:66, 1993

106. Sonenshine D: Biology of Ticks. Oxford University Press, New York, 1991

107. Dworkin MS, Anderson DE, Schwan TG, et al: Tick-borne relapsing fever in the northwestern United States and southwestern Canada. Clin Infect Dis 26:122, 1998

108. Barbour AG: Antigenic variation of a relapsing fever *Borrelia* species. Annu Rev Microbiol 44:155, 1990

109. Southern PM, Sanford JP: Relapsing fever: a clinical and microbiological review. Medicine (Baltimore) 48:129, 1969

110. Sciotto CG, Lauer BA, White WL, et al: Detection of *Borrelia* in acridine orange stained blood smears by fluorescence microscopy. Arch Pathol Lab Med 107:384, 1983

111. Burgdorfer W: The diagnosis of relapsing fever. The Biology of Parasitic Spirochetes. Johnson RC, Ed. Academic Press, New York, 1976, p 225

112. Relapsing fever. MMWR Morb Mortal Wkly Rep 22:242, 1973

113. Butler T: Relapsing fever: new lessons about antibiotic action. Ann Intern Med 102:397, 1985

114. Butler T, Jones PK, Wallace CK: *Borrelia recurrentis* infection: single-dose antibiotic regimens and management of the Jarisch-Herxheimer reaction. J Infect Dis 137:573, 1978

115. Koopman JP, van den Brink ME, Vennix P, et al: Isolation of *Streptobacillus moniliformis* from the middle ear of rats. Lab Anim 25:35, 1991

116. Ojukwu IC, Christy C: Rat-bite fever in children: case report and review. Scand J Infect Dis 34:474, 2002

117. Thong BY, Barkham TM: Suppurative polyarthritis following a rat bite. Ann Rheum Dis 62:805, 2003

118. Fatal Rat-Bite Fever—Florida and Washington, 2003. MMWR Morb Mortal Wkly Rep 53:1198 2005

119. Wallet F, Savage C, Loiez C, et al: Molecular diagnosis of arthritis due to *Streptobacillus moniliformis*. Diag Microbiol Infect Dis 47:623, 2003

120. McHugh TP, Bartlett RL, Raymond JI: Rat bite fever: report of a fatal case. Ann Emerg Med 14:1116, 1985

121. Taber LH, Feigin RD: Spirochetal infections. Pediatr Clin North Am 26:377, 1979

122. Dow GR, Rankin RJ, Saunder BW: Rat-bite fever. N Z Med J 105:133, 1992

123. Roughgarden JW: Antimicrobial therapy of rat bite fever. Arch Intern Med 116:39, 1965

Acknowledgments

Figure 1 CDC/ Jim Gathany.

Figure 4a CDC/Janice Carr.

VIII INFECTIONS DUE TO ESCHERICHIA COLI AND OTHER ENTERIC GRAM-NEGATIVE BACILLI

Michael S. Donnenberg, m.d.

This chapter describes infections caused by *Escherichia coli* and related members of the family Enterobacteriaceae, excluding genera that principally cause enteric infections. Infections caused by *Salmonella, Shigella*, and *Yersinia* are described elsewhere [*see 7:IX Infections Due to the Enteric Pathogens Campylobacter, Salmonella, Shigella, Yersinia, Vibrio, and Helicobacter*]. The family Enterobacteriaceae comprises facultative anaerobic gram-negative bacilli that ferment sugars. These organisms are often motile because of the presence of peritrichous flagella (i.e., flagella that are distributed around the entire cell). Many Enterobacteriaceae species reside principally in the gastrointestinal tract of vertebrates, though some are found primarily in the environment. They are oxidase negative and catalase positive and are capable of reducing nitrate to nitrite. They cause a variety of diseases, including diarrhea, urinary tract infections (UTIs), and nosocomial infections.

Escherichia coli Infections

E. coli, the most common facultative anaerobe in the human intestine, is distinguished from other members of the family Enterobacteriaceae primarily on the basis of *E. coli's* ability to ferment lactose and to produce indole and its inability to hydrolyze urea. Several virulence factors are shared by most members of the species; these include the ability to produce a highly reactive lipopolysaccharide in the cell wall, the ability to produce type 1 mannose-binding fimbriae, and, in many strains, the ability to produce an antiphagocytic capsule and to sequester iron.

E. coli and *Salmonella* diverged from a common ancestor about 100 million years ago.[1] Ample time for diversifying selection and a prodigious capacity for genetic exchange have fostered a tremendous degree of genetic diversity in *E. coli* strains. Thus, it is not surprising that this species has the ability to cause a diverse array of infectious diseases. In fact, the organisms that are commonly referred to as *Shigella* are actually members of the species *E. coli*,[2] but because of historical and clinical considerations, they are often discussed separately [*see 7:IX Infections Due to the Enteric Pathogens Campylobacter, Salmonella, Shigella, Yersinia, Vibrio, and Helicobacter*]. The ability of *E. coli* to cause a variety of clinical syndromes by a plethora of mechanisms is entirely dependent upon unique virulence attributes that are encoded by distinct sets of virulence genes. Each group of *E. coli* that causes a particular clinical syndrome by a recognized pathogenic mechanism may be referred to as a pathotype [*see Table 1*]. With the advent of genomic sequencing, it has become evident that different strains of *E. coli,* which are so alike in most characteristics that they could easily be regarded as identical in clinical microbiology laboratories, are remarkably different in genetic content. For example, a strain isolated from a woman with pyelonephritis, a strain isolated from a child with hemorrhagic colitis, and a laboratory strain originally isolated from an asymptomatic volunteer share only 39% of their genes; in contrast, 47% of the total number of genes in the three strains is unique to one strain only.[3,4] It is not known what roles the overwhelming majority of these genes play in the biology of the organism or in the pathogenesis of infections.

Because each pathotype of *E. coli* produces disease that has more or less distinct epidemiologic, pathogenetic, and clinical features, the discussion of specific pathotypes is organized on the basis of whether the infections caused by those pathotypes occur within or outside of the GI tract.

DIARRHEA

Six pathotypes of *E. coli* cause diarrhea [*see Table 1*].[5] Of these, the epidemiologic, pathogenetic, and clinical features of enteroinvasive *E. coli* are the same as those of *Shigella,* and they are described elsewhere [*see 7:IX Infections Due to the Enteric Pathogens Campylobacter, Salmonella, Shigella, Yersinia, Vibrio, and Helicobacter*].

Diarrhea Due to Enterotoxigenic E. coli

Epidemiology Enterotoxigenic *E. coli* (ETEC) is a common cause of watery diarrhea in children in developing nations and in people of all ages who visit these countries. ETEC infections are spread through ingestion of contaminated food and water in regions where sanitation is inadequate. A relatively large inoculum is required to produce illness; therefore, person-to-person transmission is not significant for this organism.

Pathogenesis ETEC produces either a heat-labile toxin (LT), a heat-stable toxin (ST), or both. LT has the same three-dimensional structure, receptor, and mechanism of action as cholera toxin [*see 7:IX Infections Due to the Enteric Pathogens Campylobacter, Salmonella, Shigella, Yersinia, Vibrio, and Helicobacter*]. Like cholera toxin, LT toxin induces a chain of events that leads to an elevation in the intracellular concentration of cyclic adenosine monophosphate (cAMP) and the opening of the cystic fibrosis transmembrane conductance regulator [*see Figure 1*]. The efflux of chloride ions into the intestinal lumen is accompanied by sodium ions and water, and diarrhea ensues. ST is a small disulfide-rich peptide that resembles the endogenous hormone guanylin. Like its homologue, ST causes increases in the levels of cyclic guanosine monophosphate (cGMP), which also lead to electrolyte efflux and fluid efflux through the cystic fibrosis transmembrane conductance regulator. For these toxins to reach their targets in sufficient quantity and for a sufficient duration to induce symptoms, the organisms must adhere to the intestinal mucosa.[6] This binding is accomplished through an antigenically and morphologically diverse array of pilus and nonpilus adhesins known as colonization factor antigens.[5]

Diagnosis After ingestion of food or water that is contaminated with large numbers of ETEC organisms, there is an incubation period of 0 to 2 days before the onset of symptoms. The disease begins abruptly with the onset of watery diarrhea without blood or mucus. Nausea, vomiting, abdominal cramps, and fever are rarely prominent. The disease is self-limited, lasting 3 to 5 days in travelers.[5] The endemic disease is sometimes more severe, occasionally inducing cholera-like purging in children residing in developing countries. Immunity appears to be strain specific and may reflect protective responses against particular

Table 1 Clinical, Epidemiologic, Pathogenetic, and Therapeutic Aspects of Infection with Various Pathotypes of *E. coli*

Pathotype	*Clinical Features*	*Epidemiologic Features*	*Virulence Factors*	*Management*
Enterotoxigenic	Watery diarrhea	Childhood diarrhea in developing countries; traveler's diarrhea	Pili, heat-labile and heat-stable enterotoxins	Fluid replacement; fluoroquinolone or rifaximin can be used alone or in combination with loperamide
Enteropathogenic	Watery diarrhea, vomiting	Infants in developing countries	Bundle-forming pilus, attaching and effacing effect	Fluid replacement
Enterohemorrhagic	Watery diarrhea, hemorrhagic colitis, hemolytic-uremic syndrome	Food-borne and water-borne outbreaks in developed countries	Shiga toxins, attaching and effacing effect	Fluid replacement, supportive care; antibiotics and antimotility agents are contraindicated
Enteroaggregative	Diarrhea with mucus	Childhood diarrhea, traveler's diarrhea	Pili, cytotoxins	Fluid replacement; antibiotic treatment for patients with AIDS?
Diffuse-adhering	Diarrhea	Older children?	Unknown	Fluid replacement
Extraintestinal	UTI, neonatal meningitis, nosocomial infections	Sexually active women, neonates, hospitalized patients	Fimbriae of types 1, P, and others; hemolysin; capsule; iron-acquisition systems	Antibiotics for symptomatic patients and in selected patients with asymptomatic UTI; antibiotic therapy guided by susceptibility testing for other infections

UTI—urinary tract infection

colonization-factor antigens and toxins. The diagnosis of ETEC infection is often suspected on the basis of epidemiology and the clinical presentation but is rarely confirmed. Microbiologic diagnosis requires the use of bioassays to identify LT and ST toxins or the use of DNA probing techniques or polymerase chain reaction to detect the genes that encode these toxins; these tests are performed on *E. coli* organisms recovered from the stool.

Treatment As with all diarrheal disease, the initial management of ETEC infection involves ensuring adequate fluid repletion [*see 4:III Diarrheal Diseases*]. Several well-conducted clinical trials have demonstrated that traveler's diarrhea that is partly caused by ETEC responds rapidly to treatment with any of several regimens containing antibiotics, with or without antimotility agents.[7,8] Ciprofloxacin, taken every 12 hours for 3 days in combination with loperamide, is a particularly effective regimen.[9] Rifaximin, a nonabsorbable agent, is approved by the Food and Drug Administration for traveler's diarrhea caused by noninvasive strains of *E. coli*. Studies have found rifaximin to be comparable in efficacy to ciprofloxacin.[10] Azithromycin is another effective alternative.[11] Rifaximin can be used as a single agent or in combination with loperamide. ETEC organisms are generally susceptible to fluoroquinolones; however, should resistance to fluoroquinolones increase, rifaximin may prove to be an important alternative to ciprofloxacin for the treatment of traveler's diarrhea. For adults traveling to countries where ETEC is endemic, these medications can be provided before they travel, for prompt use if symptoms occur. Travelers should also be counseled to seek medical care if diarrhea is accompanied by blood or fever or if it persists despite treatment. Travelers can reduce the risk of acquiring ETEC by fastidiously avoiding all food that is not served steaming hot, fruit that they have not peeled, and beverages that are not bottled [*see CE:VII Health Advice for International Travelers*]. Bismuth subsalicylate tablets, taken four times a day, also provide some protection, but difficulty with compliance limits the use of this therapy.[7] The prospects for developing an ETEC vaccine are limited by the antigenic heterogeneity of the organisms.

Diarrhea Due to Enteropathogenic E. coli

Enteropathogenic *E. coli* (EPEC) strains are characterized by the ability of these organisms to bind intimately to the apical surface of enterocytes, where they destroy microvilli and induce the formation of cellular pedestals, which then embrace the bacteria [*see Figure 2*]. This histopathologic appearance is known as the attaching and effacing effect.[12] EPEC strains are distinguished from enterohemorrhagic *E. coli* (EHEC) strains by the inability of the former to produce Shiga toxin (see below). Although rare outbreaks involving contaminated food or water may strike individuals of any age,[13-15] most EPEC infections occur in infants in developing countries.[16] However, atypical EPEC strains lacking the bundle-forming pilus found in typical strains have recently been identified as a cause of diarrhea in children in developed countries as well.[17,18]

Diarrhea Due to Enterohemorrhagic and Other Shiga Toxin–Producing E. coli

Epidemiology The Shiga toxin–producing *E. coli* strains constitute a heterogeneous group of organisms, of which EHEC is a subset. Like EPEC organisms, EHEC organisms induce the attaching and effacing effect on epithelial cells. Of the many Shiga toxin–producing *E. coli* strains that have been described in the literature, EHEC of serotype O157:H7 is the most important, having caused both the largest number of outbreaks and the outbreaks involving the greatest number of patients.[19-21] There are approximately 0.9 cases of EHEC O157:H7 infections per 100,000 persons annually in the United States.[22] The reservoir for EHEC is infected cattle, but the organism can be found in a variety of other ruminants. The disease often appears in outbreaks associated with the consumption of contaminated food. Undercooked

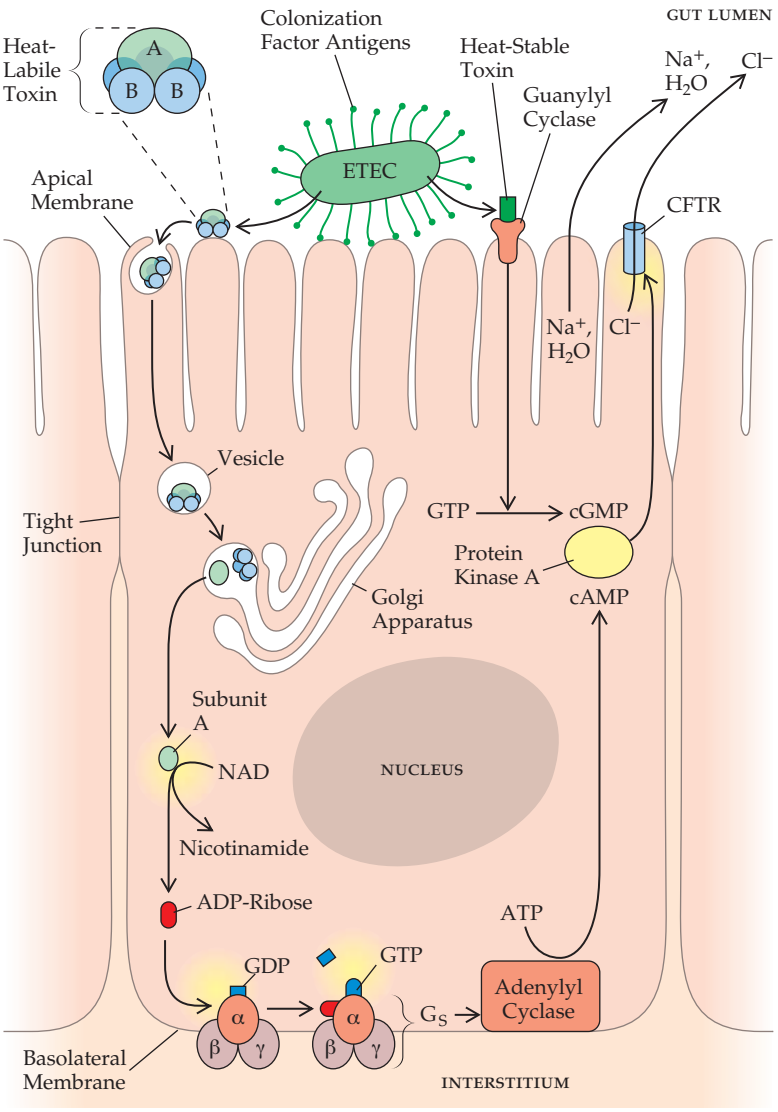


Figure 1 **The pathogenesis of enterotoxigenic *Escherichia coli* (ETEC) infections. ETEC organisms adhere to intestinal epithelial cells by one of several adhesins known as colonization factor antigens. No damage is caused to the epithelial cells. ETEC produces either a heat-labile toxin (LT), a heat-stable toxin (ST), or both. LT is composed of a pentamer of receptor-binding B subunits that are noncovalently attached to a single enzymatically active A subunit. After binding its glycolipid receptor and entering the enterocyte, LT is transported retrograde in membrane-bound vesicles to the Golgi apparatus. The A subunit reaches its target in the basolateral cell membrane, which is the alpha subunit of the stimulatory heterotrimeric G protein receptor (G_A), and catalyzes the adenosine diphosphate (ADP) ribosylation of the target. The modified alpha subunit is locked in its guanosine triphosphate (GTP)–bound active form, which forces it to constitutively activate adenylyl cyclase. The elevated levels of cyclic adenosine monophosphate (cAMP) that result lead to activation of protein kinase A (PKA) and to phosphorylation and opening of the cystic fibrosis transmembrane conductance regulator (CFTR). ST activates a distinct but convergent pathway. A disulfide-rich peptide, ST resembles the endogenous peptide hormone guanylin. Like its homologue, ST binds to and activates membrane-bound guanylyl cyclase–C, causing elevations in the levels of cyclic guanosine monophosphate (cGMP). As is the case with LT, this leads to electrolyte and fluid efflux through the CFTR.[5,29] (Cl^-—chloride; GDP—guanosine diphosphate; GTP—guanosine triphosphate; NAD—nicotinamide-adenine dinucleotide).**

ground beef has been associated with many outbreaks and is the leading risk factor for sporadic cases, but a variety of other foods and beverages, including other beef products, lettuce, sprouts, fruit, fruit juices, and milk, have been implicated. The disease has a low inoculum and can spread from person to person, particularly in day care centers or within the families of young children. An outbreak of EHEC infection resulting from airborne dispersal of bacteria after a country fair emphasizes the highly infectious nature of this pathogen.[23] It can also be spread through contamination of the water supply or through contamination of swimming pools, lakes, or water parks, and it can be contracted directly from infected animals at farms and petting zoos.[19,24-26]

Pathogenesis Animal models have demonstrated that the diarrhea caused by EHEC infection depends on the attaching and effacing effect.[27] However, the severe systemic complications of EHEC infection are caused by the expression of Shiga toxins and are independent of the attaching and effacing effect. Shiga toxins are encoded by bacteriophages related to the classic lambda phage; these bacteriophages stably infect the bacteria.[28] The three-dimensional structures of these toxins resemble those of LT toxin and cholera toxin, but the receptors, enzymatic activities, and targets are entirely different. Shiga toxins catalyze the depurination of ribosomal RNA, leading to a cessation of protein synthesis and to cell death [*see Figure 3*].[29] Microvascular endothelial cells, such as those of the kidney, are particularly sensitive to the toxin.

Diagnosis The classic presentation of EHEC infection begins with watery diarrhea and severe cramps and progresses to bloody diarrhea. Fever is low grade or absent. However, symptomatic disease may range from heme-negative watery diarrhea to frank hematochesia. The severe cramps, the absence of fever, or the presence of blood can lead clinicians to confuse EHEC infections with a variety of noninfectious illnesses, including intussusception, inflammatory bowel disease, and bowel ischemia; this in turn can lead to unnecessary diagnostic procedures or surgical intervention.[30] The dreaded complication of EHEC infection is the hemolytic-uremic syndrome (HUS), which sometimes manifests itself as thrombotic thrombocytopenic purpura. HUS ensues in approximately 5% to 10% of those patients with EHEC O157:H7 infection several days after the onset of diarrhea. It carries a high risk of death or permanent renal impairment. Various studies have demonstrated that children, the elderly, those who have consumed antimotility agents, and those with high white cell counts have an increased risk of HUS.[31,32] Antibiotic therapy

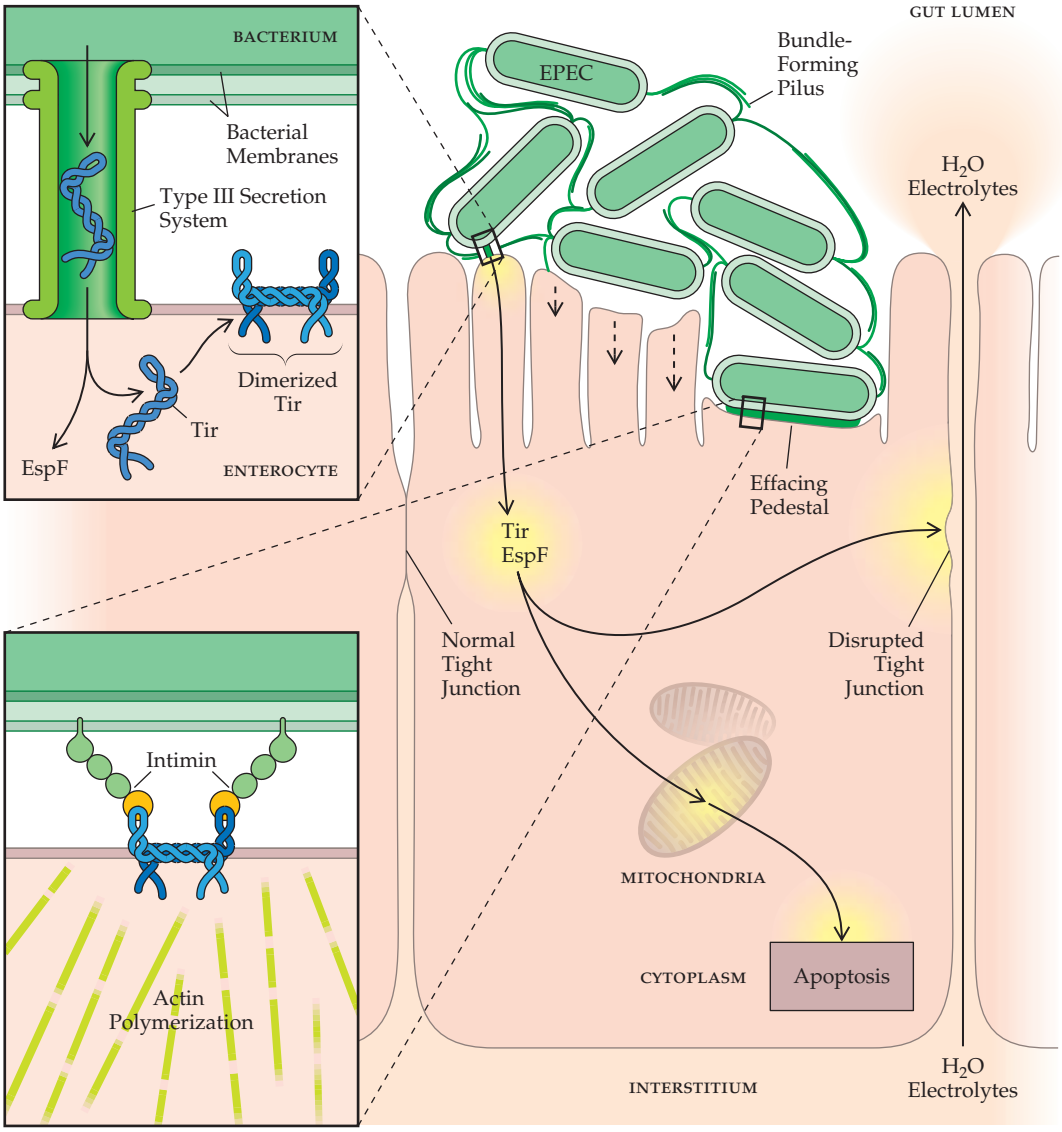


Figure 2 **The pathogenesis of enteropathogenic *E. coli* (EPEC) infections. Typical EPEC strains produce a plasmid-encoded bundle-forming pilus that is required for reversible aggregation of the bacteria and for full virulence.[63] The bacteria are able to inject into host cells a number of proteins, including Tir and EspF, through a specialized chromosome-encoded type III secretion system. Tir forms a dimer in the host cell membrane, where it serves as a receptor for the bacterial outer membrane protein adhesin intimin, a protein of multiple domains that is also essential for virulence. Binding of intimin triggers Tir to activate the cellular machinery for actin polymerization, leading to the formation of the characteristic attaching and effacing pedestals to which the organisms adhere. EspF causes loss of intestinal barrier function through disruption of tight junctions and induces host cell death through apoptosis.[12]**

may also be a risk factor for HUS. Studies on animal models have shown that many antibiotics induce the production of Shiga toxins[33]; clinical studies have reported an association between antibiotic therapy and an increased risk of HUS,[34] although some studies have found no such association.[35] It is possible that some antibiotics increase the risk of HUS, whereas others decrease the risk. However, prospective, randomized studies will be required to test this hypothesis.

HUS is characterized by microangiopathic hemolytic anemia, which results from Shiga toxin–induced damage to endothelial cells, leading to activation of coagulation in the microvasculature. The kidney is particularly susceptible, but ischemic necrosis of the bowel, brain, eye, or virtually any organ can occur.

The diagnosis of EHEC infection is extremely important, both for the patient (to prevent unnecessary interventions and provide appropriate care) and for public health, because any patient with EHEC could represent the index case of a large outbreak—the detection and interruption of which depend critically upon timely diagnosis and reporting. Fortunately, unlike most strains of *E. coli*, O157:H7 ferments sorbitol slowly, enabling the identification of these organisms on specific indicator plates. However, not all microbiology laboratories offer this test, and many of those that perform it do so only on request. Therefore, the recognition of sporadic cases and outbreaks of EHEC infection depends largely on the acumen of the astute clinician, who must suspect the diagnosis and request the test.

The diagnosis of *E. coli* infection caused by Shiga toxin–producing organisms other than O157:H7 is more difficult, because these organisms do not have a distinguishing phenotype that can be assayed in most laboratories. Colonies cultured from the stool of patients with suspected non-O157:H7 Shiga toxin–producing *E. coli* infection should be sent to referral laboratories, state health laboratories, or the Centers for Disease Control and Prevention for assays to detect Shiga toxins or the genes that encode them.

Treatment The treatment of EHEC infection is supportive and includes oral rehydration, observation, and, if necessary, hospital admission for the management of complications. Until proved safe and effective, antibiotics should be regarded as contraindicated in patients with known or suspected EHEC infection. Antimotility agents are also contraindicated in these patients, because their use is associated with a possible risk of HUS[31] and with a theoretical risk of increasing the duration of toxin exposure.

The risk of EHEC infection can be reduced by routine adherence to good hygienic practice in food preparation. Ground beef should always be cooked to an internal temperature of 68.3° C (well done or until the juices run clear).[21] A broad array of efforts

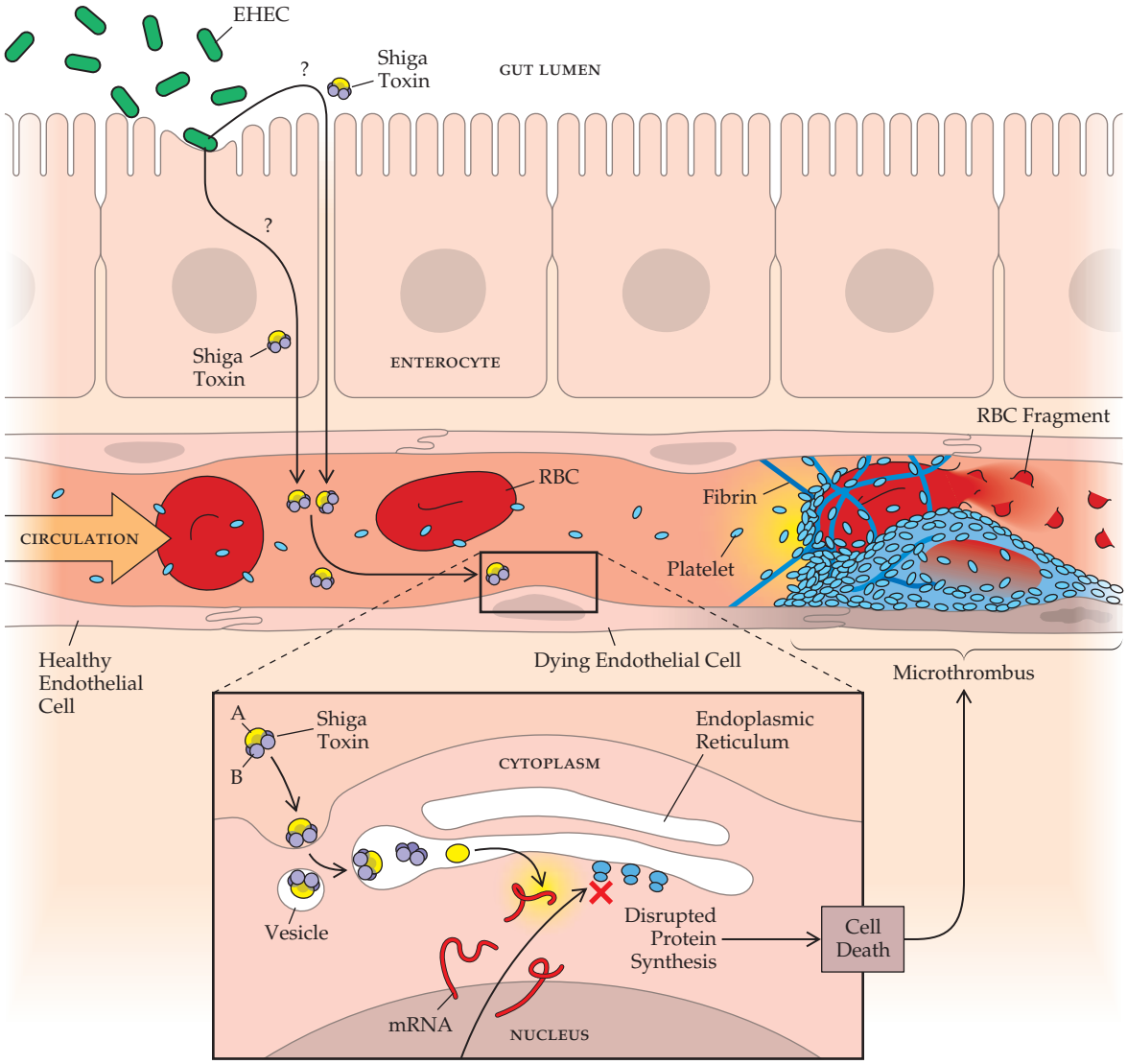


Figure 3 **The pathogenesis of enterohemorrhagic *E. coli* (EHEC) infections. Like EPEC, EHEC strains express intimin and Tir and employ the type III secretion system, which cause the attaching and effacing effect on cells. However, EHEC also produces Shiga toxins, which are encoded by bacteriophages that stably infect the bacteria. Shiga toxins are similar in quaternary structure to the heat-labile toxin of ETEC; they are composed of a pentamer of receptor-binding B subunits that are noncovalently bound to a single enzymatic A subunit. Shiga toxins are able to cross intestinal epithelial monolayers by an unknown mechanism, after which they presumably are spread systemically through the bloodstream. Endothelial cells appear to be particularly important target cells for Shiga toxins. After binding its glycolipid receptor and entering the host cell, the toxin is transported retrograde in membrane-bound vesicles to the endoplasmic reticulum. The A subunit induces depuration of a specific adenine residue in ribosomal RNA; this leads to cessation of protein synthesis and death of the cell. The altered surface of the intoxicated endothelial cell serves as a nidus for activation of the coagulation cascade, which leads to the formation of microthrombi and causes distal ischemic necrosis, platelet consumption, and red cell fragmentation—the hallmarks of the hemolytic-uremic syndrome.[12]**

to reduce EHEC infections is under development. These include measures to reduce colonization in cattle; vaccine development; and improved food-safety measures.

Diarrhea Due to Enteroaggregative E. coli *and Diffuse-Adhering* E. coli

Enteroaggregative *E. coli* (EAEC) and diffuse-adhering *E. coli* (DAEC) are distinguished by characteristic patterns of adherence to tissue culture cells on in vitro assays. It is likely that these pathotypes contain a heterogeneous mixture of strains that do not entirely share pathogenetic or clinical features.[5,36,37] Our understanding of the epidemiology, pathogenesis, and management of these infections is rudimentary, but several interesting features have been defined.[38]

Epidemiology EAEC organisms are associated with acute and chronic diarrhea in children in developing and developed countries[17,39,40]; they have also been isolated from patients with AIDS.[41,42] In addition, recent studies have found that EAEC is a frequent cause of travelers's diarrhea, rivaling the incidence of ETEC.[43] Interestingly, carriage of EAEC, even in the absence of overt symptoms, is associated with evidence of growth retardation in children.[44] DAEC has been associated with diarrhea in older children in developing countries.[45,46]

Pathogenesis EAEC strains produce a variety of pili and toxins, but further study is needed to clarify the role that these factors play in causing disease. The pathogenetic mechanisms of DAEC remain obscure.

Diagnosis Diarrhea caused by EAEC may contain mucus or blood and may be accompanied by cramps. In some cases, particularly in patients with AIDS, the diarrhea can be protracted. The clinical features of DAEC infection have not been well described. The diagnosis of these infections requires tissue culture assays that are performed only in the research setting.

Treatment Case reports and studies in small series of patients have suggested that AIDS patients who are infected with EAEC may respond to antibiotic treatment.[47]

EXTRAINTESTINAL INFECTIONS

E. coli can infect a variety of extraintestinal sites, including the urinary tract, the meninges (in neonates), the lungs, the peritoneum, the gallbladder, and the biliary tree; *E. coli* infections can also occur in association with intravascular and prosthetic devices [*see Table 1*]. *E. coli* can also be involved in polymicrobial infections such as intra-abdominal abscesses and skin and soft tissue infections. Bacteremia can complicate extraintestinal *E. coli* infections. Although the virulence factors responsible for all of these diseases have not been well studied, it is clear that the strains that cause extraintestinal infections are not a random sample of *E. coli* from the host intestine. On the other hand, these strains lack most of the virulence factors associated with those that cause diarrhea; rather, the strains that cause extraintestinal infections are more likely to possess the genes for hemolysin to produce an antiphagocytic capsule and a variety of fimbriae, and to sequester iron than are strains isolated from the GI tract. The similarity in the repertoire of putative virulence factors in strains isolated from UTIs and other extraintestinal infections has led to the conclusion that extraintestinal pathogenic *E. coli* (ExPEC) should be considered a single pathotype.[48]

E. coli is by far the leading cause of UTIs in otherwise healthy persons, and it is an important cause of UTIs in those patients whose urinary tracts are compromised by anatomic abnormalities, foreign bodies, or host immune defects. The bacteria that cause UTIs are more likely than fecal *E. coli* isolates to possess the genes for a variety of factors under study as potential virulence determinants. Of these factors, P fimbriae and hemolysin are prominent.[49] Interestingly, however, the ubiquitous type 1 fimbriae that are produced by virtually all strains of *E. coli* and related organisms play a critical role in the pathogenesis of *E. coli*–associated UTI. However, there is much more to the pathogenesis of UTI than the production of type 1 fimbriae, because the presence of these organelles does not distinguish uropathogenic *E. coli* from other strains. The genome of a prototype strain of extraintestinal pathogenic *E. coli* confirms that these strains possess many genes that are absent from *E. coli* K-12.[4] More detailed information on the pathogenesis, clinical features, and treatment of UTI is presented elsewhere [*see 7:XXIII Infections of the Urinary Tract*].

The diagnosis of these infections is confirmed by isolating the organism in culture from appropriate clinical specimens. Treatment is guided by susceptibility testing. The rising prevalence of resistance to trimethoprim-sulfamethoxazole in *E. coli* isolates from the community and the alarming rise in the prevalence of strains that produce extended-spectrum β-lactamases are narrowing therapeutic options.[50-52]

Proteus Infections

Named after a character in Homer's *Odyssey* who could change shape, organisms belonging to the genus *Proteus* are noted for their ability to take two forms: (1) typical bacillary swimmer cells, which express a variety of surface fimbriae as well as flagella, and (2) highly elongated swarm cells, which express hundreds of flagella and few other surface structures. Swarm cells provide a challenge to the clinical microbiologist because the organisms frequently swarm over culture plates, which interferes with the isolation of individual colonies. By far, the most commonly isolated *Proteus* species is *P. mirabilis*.

Pathogenesis

Proteus rapidly hydrolyzes urea to form carbon dioxide and ammonium hydroxide. In the urinary tract, this reaction results in an increase in pH. This reaction also alters the solubility of polyvalent ions and leads to the formation of struvite calculi. These stones can then obstruct urinary catheters, leading to bacterial persistence and urosepsis. In a murine model of ascending UTI, a urease-negative mutant of *P. mirabilis* was found to be severely limited in its ability to colonize the urinary tract, and this mutant strain failed to form stones.[53] Studies in animal models have confirmed the ability of *P. mirabilis* to produce several fimbriae that play a role in infection.[54]

EPIDEMIOLOGY AND ETIOLOGY

Although *P. mirabilis* occasionally causes UTI in otherwise healthy persons, it is more commonly isolated from patients with abnormalities of the urinary tract, particularly those with indwelling urethral catheters and those with stones. Notably, *Proteus* has a predilection for causing upper urinary tract infections.[55] Whereas *P. mirabilis* rarely causes infections outside the urinary tract, other members of the genus and closely related organisms such as *Morganella morganii* may be isolated from skin

infections and soft tissue infections, especially skin ulcers in patients with diabetes mellitus.

Diagnosis

The diagnosis of a urinary tract infection caused by *Proteus* organisms can be suspected in patients with compatible symptoms and signs who have complicated urinary tracts, including those patients with indwelling urinary catheters and functional or anatomic abnormalities. Evidence of upper urinary tract involvement (e.g., fever, flank pain, and sepsis), catheter obstruction, or known renal calculi strongly supports the diagnosis, which can be confirmed by culture.

Treatment

Most strains of *P. mirabilis* are susceptible to many antibiotics, but some *Proteus* species are highly resistant. In the treatment of UTIs caused by *Proteus* organisms, the underlying urinary abnormality and the clinical appearance of the patient must be taken into account. Treatment of asymptomatic bacteriuria in patients with urinary calculi or anatomic or functional abnormalities provides no long-term benefit to the patient and merely increases the potential for antimicrobial resistance [*see 7:XXIII Infections of the Urinary Tract*].

Klebsiella Infections

Klebsiella pneumoniae and, to a lesser extent, *K. oxytoca* are responsible for the vast majority of human *Klebsiella* infections. These organisms are found in the environment and in the GI tract.

EPIDEMIOLOGY AND ETIOLOGY

Although *Klebsiella* species occasionally cause UTI in otherwise healthy persons, the majority of *Klebsiella* infections occur in compromised hosts. *K. pneumoniae* can cause community-acquired pneumonia, classically in patients with alcoholism, with dramatic findings such as currant jelly sputum and a bulging fissure sign (Friedländer pneumonia). In practice, pulmonary infections caused by *Klebsiella* are difficult to distinguish on clinical grounds from those caused by other organisms.[56] Most *Klebsiella* infections occur in hospitals, where these bacteria cause UTI, pneumonia, wound infections, biliary infections, and primary bacteremia.[57] *Klebsiella* are often second only to *E. coli* as a cause of nosocomial gram-negative bacteremia.[57,58] These infections can occur as epidemics or sporadically.[59] Particularly problematic are outbreaks of nosocomial infections caused by strains of *Klebsiella* that produce extended-spectrum β-lactamases.[51,52] These organisms are frequently resistant to multiple antibiotics, owing to the presence of plasmids that encode multiple antibi-otic-resistance genes.

Pathogenesis

The primary virulence factor for *Klebsiella* is a luxuriant polysaccharide capsule, which has antiphagocytic properties. *Klebsiella* organisms also produce type 1 fimbriae; like ExPEC organisms, they also produce other fimbriae and have iron-acquisition systems and a reactive lipopolysaccharide.[60]

Diagnosis

The clinical features of community-acquired pneumonia caused by *K. pneumoniae* are not sufficiently distinctive to differentiate patients with infections caused by this organism from those who have infections caused by other agents. Similarly, nosocomial infections and UTIs caused by *Klebsiella* cannot be diagnosed on the basis of clinical features.

In the microbiology laboratory, a presumptive identification of *Klebsiella* infection is often made when a highly mucoid, lactose-fermenting, ampicillin-resistant, gram-negative bacillus is isolated and is confirmed by further testing.

Treatment

Treatment of *Klebsiella* infections requires consultation with a clinical microbiology laboratory or specialists in infectious diseases to ensure selection of an antibiotic to which the organism is susceptible. *Klebisella* isolates that do not produce extended-spectrum β-lactamases are usually resistant to ampicillin but may be sensitive to cephalosporins, trimethoprim-sulfamethoxazole, aminoglycosides, combinations of penicillins and β-lactamase inhibitors, and fluoroquinolones, any of which may be used to treat infections caused by susceptible organisms.

Enterobacter and *Serratia* Infections

Epidemiology and Etiology

Enterobacter and *Serratia* are closely related to *Klebsiella;* like *Klebsiella,* they are principally opportunistic pathogens that cause a variety of nosocomial infections. In addition, *S. marcescens* has been isolated from the blood or other clinical specimens of patients who use injection drugs.[61] Although several species of *Enterobacter*—including *E. cloacae* and *E. aerogenes,* as well as the closely related species *Pantoea agglomerans*—cause infections, *S. marcescens* is the only *Serratia* species that has been isolated from patients with appreciable frequency.

Pathogenesis

Enterobacter and *Serratia* organisms share with *Klebsiella* organisms the ability to produce antiphagocytic capsules. One important clinical difference is that *Enterobacter* species have an inducible chromosomal cephalosporinase. This enzyme renders the organisms resistant not only to first-generation cephalosporins but also to many second- and third-generation cephalosporins after high-level induction of cephalosporinase in the presence of these antibiotics. Such resistance may not be detected in the microbiology laboratory on initial testing and may become apparent only when the organism is again isolated from patients who fail to respond to therapy.

Diagnosis

As with infections caused by *Klebsiella,* infections caused by *Enterobacter* and *Serratia* do not have distinguishing clinical features and are diagnosed by culture of the organism from a normally sterile site or from a nonsterile site in a patient with associated signs or symptoms.

Treatment

Second- and third-generation cephalosporins should be used cautiously in patients with serious *Enterobacter* infections, even if the isolate appears to be susceptible on initial testing.[62] When the organism is susceptible to alternative agents such as extended-spectrum penicillins, carbapenems, or fluoroquinolones, their use may be advisable.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Whittam TS: Genetic variation and evolutionary processes in natural populations of *Escherichia coli*. *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology. Neidhardt FC, Ed. ASM Press, Washington, DC, 1996, p 2708

2. Pupo GM, Karaolis DK, Lan R, et al: Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infect Immun 65: 2685, 1997

3. Perna NT, Plunkett G 3rd, Burland V, et al: Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature 409:529, 2001

4. Welch RA, Burland V, Plunkett C 3rd, et al: Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci USA 99:17020, 2002

5. Nataro JP, Kaper JB: Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 11:142, 1998

6. Zafriri D, Oron Y, Eisenstein BI, et al: Growth advantage and enhanced toxicity of *Escherichia coli* adherent to tissue culture cells due to restricted diffusion of products secreted by the cells. J Clin Invest 79:1210, 1987

7. DuPont HL, Ericsson CD: Prevention and treatment of traveler's diarrhea. N Engl J Med 328:1821, 1993

8. Guerrant RL, Van Gilder T, Steiner TS, et al: Practice guidelines for the management of infectious diarrhea. Clin Infect Dis 32:331, 2001

9. Petruccelli BP, Murphy GS, Sanchez JL, et al: Treatment of traveler's diarrhea with ciprofloxacin and loperamide. J Infect Dis 165:557, 1992

10. DuPont HL, Jiang ZD, Ericsson CD, et al: Rifaximin versus ciprofloxacin for the treatment of traveler's diarrhea: a randomized, double-blind clinical trial. Clin Infect Dis 33:1807, 2001

11. Adachi JA, Ericsson CD, Jiang ZD, et al: Azithromycin found to be comparable to levofloxacin for the treatment of travelers with acute diarrhea acquired in Mexico. Clin Infect Dis 37:1165, 2003

12. Donnenberg MS, Whittam TS: Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. J Clin Invest 107:539, 2001

13. Schroeder SA, Caldwell JR, Vernon TM, et al: A waterborne outbreak of gastroenteritis in adults associated with *Escherichia coli*. Lancet 1:737, 1968

14. Viljanen MK, Peltola T, Junnila SY, et al: Outbreak of diarrhoea due to *Escherichia coli* O111:B4 in schoolchildren and adults: association of Vi antigen-like reactivity. Lancet 336:831, 1990

15. Hedberg CW, Savarino SJ, Besser JM, et al: An outbreak of foodborne illness caused by *Escherichia coli* O39:NM, an agent not fitting into the existing scheme for classifying diarrheogenic *E. coli*. J Infect Dis 176:1625, 1997

16. Donnenberg MS: Enteropathogenic *Escherichia coli*. Infections of the Gastrointestinal Tract. Blaser MJ, Smith PD, Ravdin JI, et al, Eds. Raven Press, New York, 1995, p 709

17. Robins-Browne RM, Bordun AM, Tauschek M, et al: *Escherichia coli* and community-acquired gastroenteritis, Melbourne, Australia. Emerg Infect Dis 10:1797, 2004

18. Cohen MB, Nataro JP, Bernstein DI, et al: Prevalence of diarrheagenic *Escherichia coli* in acute childhood enteritis: a prospective controlled study. J Pediatr 146:54, 2005

19. Tarr PI: *Escherichia coli* O157:H7: clinical, diagnostic, and epidemiological aspects of human infection. Clin Infect Dis 20:1, 1995

20. Michino H, Araki K, Minami S, et al: Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am J Epidemiol 150:787, 1999

21. Bell BP, Goldoft M, Griffin PM, et al: A multistate outbreak of *Escherichia coli* O157:H7–associated bloody diarrhea and hemolytic uremic syndrome from hamburgers: the Washington experience. JAMA 272:1349, 1994

22. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—10 sites, United States, 2004. MMWR Morb Mortal Wkly Rep 54:352, 2005

23. Varma JK, Greene KD, Reller ME, et al: An outbreak of *Escherichia coli* 0157 infection following exposure to a contaminated building. JAMA 290:2709, 2003

24. Outbreaks of *Escherichia coli* O157:H7 infections among children associated with farm visits—Pennsylvania and Washington, 2000. MMWR Morb Mortal Wkly Rep 50:293, 2001

25. Hilborn ED, Mermin JH, Mshar PA, et al: A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. Arch Intern Med 159:1758, 1999

26. Boyce TG, Swerdlow DL, Griffin PM: *Escherichia coli* O157:H7 and the hemolytic-uremic syndrome. N Engl J Med 333:364, 1995

27. Tzipori S, Gunzer F, Donnenberg MS, et al: The role of the *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohemorrhagic *Escherichia coli* infection. Infect Immun 63:3621, 1995

28. O'Brien AD, Newland JW, Miller SF, et al: Shiga-like toxin–converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. Science 226:694, 1984

29. Sears CL, Kaper JB: Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion. Microbiol Rev 60:167, 1996

30. Griffin PM, Ostroff SM, Tauxe RV, et al: Illnesses associated with *Escherichia coli* O157:H7 infections: a broad clinical spectrum. Ann Intern Med 109:705, 1988

31. Bell BP, Griffin PM, Lozano P, et al: Predictors of hemolytic uremic syndrome in children during a large outbreak of *Escherichia coli* O157:H7 infections. Pediatrics 100:E12, 1997

32. Siegler RL, Pavia AT, Christofferson RD, et al: A 20-year population-based study of postdiarrheal hemolytic uremic syndrome in Utah. Pediatrics 94:35, 1994

33. Zhang X, McDaniel AD, Wolf LE, et al: Quinolone antibiotics induce Shiga toxin–encoding bacteriophages, toxin production, and death in mice. J Infect Dis 181:664, 2000

34. Wong CS, Jelacic S, Habeeb RL, et al: The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N Engl J Med 342:1930, 2000

35. Safdar N, Said A, Gangnon RE, et al: Risk of hemolytic uremic syndrome after antibiotic treatment of *Escherichia coli* 0157:H7 enteritis: a meta-analysis. JAMA 288:996, 2002

36. Okeke IN, Lamikanra A, Czeczulin J, et al: Heterogeneous virulence of enteroaggregative *Escherichia coli* strains isolated from children in Southwest Nigeria. J Infect Dis 181:252, 2000

37. Czeczulin JR, Whittam TS, Henderson IR, et al: Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. Infect Immun 67:2692, 1999

38. Nataro JP: Enteroaggregative *Escherichia coli* pathogenesis. Curr Opin Gastroenterol 21:4, 2005

39. Bhan MK, Raj P, Levine MM, et al: Enteroaggregative *Escherichia coli* associated with persistent diarrhea in a cohort of rural children in India. J Infect Dis 159:1061, 1989

40. Huppertz HI, Rutkowski S, Aleksic S, et al: Acute and chronic diarrhoea and abdominal colic associated with enteroaggregative *Escherichia coli* in young children living in western Europe. Lancet 349:1660, 1997

41. Polotsky Y, Nataro JP, Kotler D, et al: HEp-2 cell adherence patterns, serotyping, and DNA analysis of *Escherichia coli* isolates from eight patients with AIDS and chronic diarrhea. J Clin Microbiol 35:1952, 1997

42. Wanke CA, Mayer H, Weber R, et al: Enteroaggregative *Escherichia coli* as a potential cause of diarrheal disease in adults infected with human immunodeficiency virus. J Infect Dis 178:185, 1998

43. Adachi JA, Jiang ZD, Mathewson JJ, et al: Enteroaggregative *Escherichia coli* as a major etiologic agent in traveler's diarrhea in 3 regions of the world. Clin Infect Dis 32:1706, 2001

44. Steiner TS, Lima AA, Nataro JP, et al: Enteroaggregative *Escherichia coli* produce intestinal inflammation and growth impairment and cause interleukin-8 release from intestinal epithelial cells. J Infect Dis 177:88, 1998

45. Giron JA, Jones T, Millan-Velasco F, et al: Diffuse-adhering *Escherichia coli* (DAEC) as a putative cause of diarrhea in Mayan children in Mexico. J Infect Dis 163:507, 1991

46. Gunzburg ST, Chang BJ, Elliott SJ, et al: Diffuse and enteroaggregative patterns of adherence of enteric *Escherichia coli* isolated from aboriginal children from the Kimberley region of Western Australia. J Infect Dis 167:755, 1993

47. Wanke CA, Gerrior J, Blais V, et al: Successful treatment of diarrheal disease associated with enteroaggregative *Escherichia coli* in adults infected with human immunodeficiency virus. J Infect Dis 178:1369, 1998

48. Russo TA, Johnson JR: Proposal for a new inclusive designation for extraintestinal pathogenic isolates of *Escherichia coli*: ExPEC. J Infect Dis 181:1753, 2000

49. Stamm WE, Hooton TM, Johnson JR, et al: Urinary tract infections: from pathogenesis to treatment. J Infect Dis 159:400, 1989

50. Talan DA, Stamm WE, Hooton TM, et al: Comparison of ciprofloxacin (7 days) and trimethoprim-sulfamethoxazole (14 days) for acute uncomplicated pyelonephritis in women: a randomized trial. JAMA 283:1583, 2000

51. Wiener J, Quinn JP, Bradford PA, et al: Multiple antibiotic-resistant *Klebsiella* and *Escherichia coli* in nursing homes. JAMA 281:517, 1999

52. Lautenbach E, Patel JB, Bilker WB, et al: Extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: risk factors for infection and impact of resistance on outcomes. Clin Infect Dis 32:1162, 2001

53. Johnson DE, Russell RG, Lockatell CV, et al: Contribution of *Proteus mirabilis* urease to persistence, urolithiasis, and acute pyelonephritis in a mouse model of ascending urinary tract infection. Infect Immun 61:2748, 1993

54. Mobley HLT: Virulence of *Proteus mirabilis*. Urinary Tract Infections: Molecular Pathogenesis and Clinical Management. Mobley HLT, Warren JW, Eds. ASM Press, Washington, DC, 1996, p 245

55. Fairley KF, Carson NE, Gutch RC, et al: Site of infection in acute urinary-tract infection in general practice. Lancet 2:615, 1971

56. Sahly H, Podschun R, Ullmann U: *Klebsiella* infections in the immunocompromised host. Adv Exp Med Biol 479:237, 2000

57. Carpenter JL: *Klebsiella* pulmonary infections: occurrence at one medical center and review. Rev Infect Dis 12:672, 1990

58. Wisplinghoff H, Bischoff T, Tallent SM, et al: Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin Infect Dis 39:309, 2004

59. Geerdes HF, Ziegler D, Lode H, et al: Septicemia in 980 patients at a university hospital in Berlin: prospective studies during 4 selected years between 1979 and 1989. Clin Infect Dis 15:991, 1992

60. Jarvis WR, Munn VP, Highsmith AK, et al: The epidemiology of nosocomial infections caused by *Klebsiella pneumoniae*. Infect Control 6:68, 1985

61. Mills J, Drew D: *Serratia marcescens* endocarditis: a regional illness associated with intravenous drug abuse. Ann Intern Med 84:29, 1976

62. Chow JW, Fine MJ, Shlaes DM, et al: *Enterobacter* bacteremia: clinical features and emergence of antibiotic resistance during therapy. Ann Intern Med 115:585, 1991

63. Bieber D, Ramer SW, Wu CY, et al: Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. Science 280:2114, 1998

Acknowledgment

Figures 1 through 3 Seward Hung.

IX INFECTIONS DUE TO THE ENTERIC PATHOGENS CAMPYLOBACTER, SALMONELLA, SHIGELLA, YERSINIA, VIBRIO, AND HELICOBACTER

MARCIA B. GOLDBERG, M.D.

In the United States, diarrhea is the third most common medical complaint, with an annual incidence of 1.5 to 5.0 illnesses per person. Rates of occurrence are highest in children, followed by the elderly. Worldwide, diarrhea is the second most common cause of death for all age groups and the leading cause of death in children.

The most common bacterial causes of diarrhea in the United States are gram-negative bacteria. *Campylobacter* species are the most common infectious organisms (46%), followed by *Salmonella* (28%), *Shigella* (17%), and *Escherichia coli* O157 (5%) [*see Figure 1*].[1]

Campylobacter Infections

EPIDEMIOLOGY

Campylobacter organisms are the most common cause of bacterial gastroenteritis in the United States, with 2.1 to 2.4 million cases estimated to occur each year.[2] This disease is most common in children. The incidence of *Campylobacter* enteritis is as much as 39-fold higher in HIV-infected persons than in the general population[3]; recurrent infection and infection with antibiotic-resistant strains are also more common.[4,5]

Campylobacter infection is acquired by ingestion of contaminated foodstuffs in 80% of cases. Exposure to infected dogs, unpasteurized milk, contaminated unchlorinated water, and infected persons has also been reported to spread the organism. Improper handling of uncooked chicken and consumption of undercooked chicken are the most common food-related causes. The infectious inoculum is only approximately 800 organisms[6]; therefore, secondary infection occurs in as many as two thirds of household contacts. Neonates born to infected mothers are at extremely high risk for infection, which can be life threatening.

C. jejuni is the most common cause of *Campylobacter* enteritis. Other species of *Campylobacter*, particularly *C. coli* and *C. upsaliensis* (for which dogs are the normal host), are common causes of enteritis in AIDS patients. *C. fetus* is the predominant cause of systemic infection. Two thirds of patients with systemic *Campylobacter* infection are males. Many are in their middle 50s; more than 25% are farm workers, butchers, or abattoir workers; and most have a major underlying illness, such as alcoholism, cirrhosis, diabetes, lymphoproliferative disorder, or valvular or atherosclerotic heart disease.

PATHOGENESIS

Campylobacter enteritis is thought to involve bacterial invasion of intestinal epithelial cells,[7] as does *Salmonella* or *Shigella* enteritis. Intestinal lesions are similar to those of granulomatous or idiopathic ulcerative colitis. The pathogenesis of systemic *Campylobacter* infection is poorly understood. Genetic techniques have been developed that will allow molecular characterization of the pathogenesis of *Campylobacter* infection.[8,9] *C. jejuni* infection may be complicated by Guillain-Barré syndrome. In such cases, antibodies to ganglioside GM1 and other peripheral nerve gangliosides are frequently detected. The lipopolysaccharide from *C. je-*

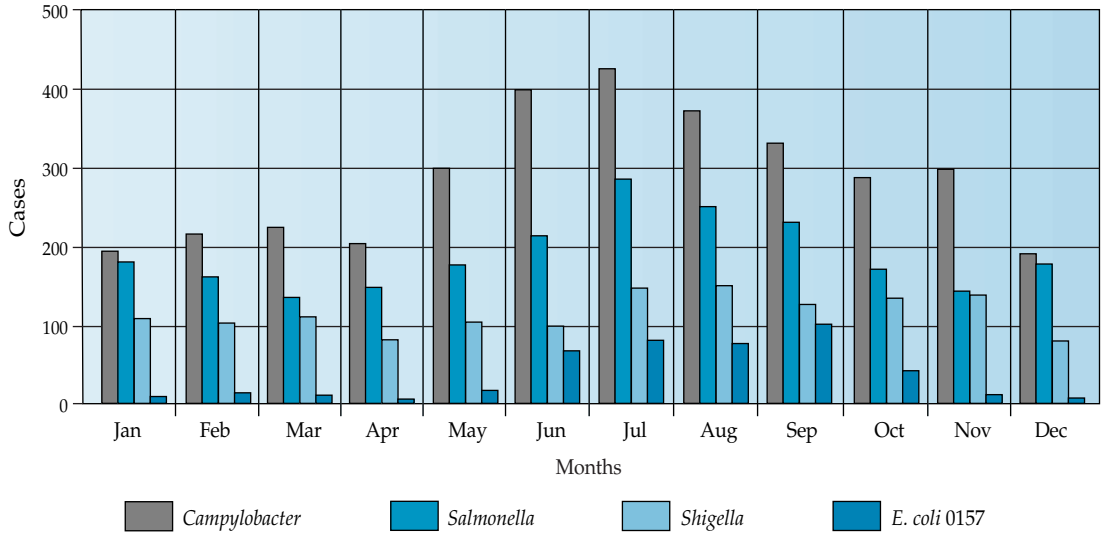


Figure 1 **Cases of gastroenteritis caused by *Campylobacter, Salmonella, Shigella,* and *E. coli* O157 reported to the Collaborating Sites Foodborne Disease Active Surveillance Network (CDC/USDA/FDA), 1996.[1]**

juni isolates displays molecular mimicry with several gangliosides, which suggests that the neurologic syndrome is caused by autoimmune recognition of gangliosides induced by exposure to the *C. jejuni* lipopolysaccharide.

DIAGNOSIS

Clinical Manifestations

The incubation period for *Campylobacter* enteritis is usually 2 to 4 days. Illness typically begins with crampy periumbilical pain and fever, which is followed by diarrhea with profuse, watery, foul-smelling stools. Abdominal pain can range from mild to severe, mimicking acute appendicitis, bowel perforation, or intussusception. Stool is grossly bloody or melanotic in 30% of patients and guaiac positive in 60%.

Patients with *Campylobacter* bacteremia present with symptoms common to many bacteremias: fever, malaise, headache, chills, night sweats, anorexia, and abdominal pain. Lethargy and confusion may occur without focal neurologic findings. A nonproductive cough may be present in the absence of focal pulmonary findings. Endocarditis may occur in 10% of patients. Meningitis is uncommon, although neonates are at extremely high risk.

Physical Examination

Physical findings in both *Campylobacter* enteritis and systemic *Campylobacter* infection are nonspecific.

Laboratory Tests

In *Campylobacter* enteritis, stool examination may reveal the diagnosis in as many as two thirds of cases. In fresh stool, *Campylobacter* organisms are small, motile, comma- or corkscrew-shaped; in Gram stains, they may be gull winged, curved, gram-negative organisms. Leukocytes are present in the stool in about 75% of cases.

Campylobacter organisms do not grow on standard media used in the detection of enteric pathogens but, rather, require specific media and specific environmental conditions.[10] If a stool sample cannot be delivered to a laboratory quickly, storage of the sample in an airtight container at 4° C or inoculation of transport medium (e.g., Cary-Blair) can improve the recovery of organisms.

Diagnosis of systemic *Campylobacter* infection is based on the isolation of the organism in blood cultures. *Campylobacter* organisms may grow slowly, so cultures should be incubated for 2 weeks or longer. In both systemic and intestinal infection, leukocytosis is common.

DIFFERENTIAL DIAGNOSIS

In the setting of acute onset of watery diarrhea, fever, and abdominal pain, *Campylobacter* infection should be considered. The differential diagnosis of diarrheal disease with polymorphonuclear leukocytes in the stool is enteritis caused by *Campylobacter*, *Salmonella*, *Shigella*, or *Yersinia enterocolitica* infection or by idiopathic ulcerative colitis. *Salmonella* and *Yersinia* infections are less likely when there is detectable blood in the stool. Systemic *Campylobacter* infection cannot be easily distinguished from other bacteremias.

TREATMENT

Campylobacter enteritis is usually self-limited, making specific therapy unnecessary. However, it may be prudent to administer antibiotics to patients with moderately severe disease (i.e., those with high fever, bloody diarrhea, or bacteremia or those who pass more than eight stools a day), immunosuppressed patients, pregnant women, or patients with symptoms that worsen or persist more than 7 days after diagnosis. Administration of antibiotics before laboratory confirmation of infection has been shown to reduce the median duration of symptoms from 10 days to 5 days.[11] However, empirical institution of antibiotics for diarrheal syndromes suggestive of enterohemorrhagic *E. coli* may increase the risk of hemolytic-uremic syndrome.[12]

Erythromycin is the treatment of choice for *Campylobacter* enteritis [*see Table 1*]. The newer macrolides, including azithromycin, and fluoroquinolones have also been used successfully [*see Table 1*]. However, resistance to fluoroquinolones is increasing rapidly worldwide; in 1998, approximately 10% of isolates in Minnesota were resistant to fluoroquinolones.[13] Moreover, strains resistant to both fluoroquinolones and macrolides have been isolated in Thailand. Fluid replacement should be administered as needed. On the basis of data for *Shigella* and *Salmonella* enteritis, antimotility agents such as Lomotil (diphenoxylate hydrochloride with atropine sulfate) should probably be avoided.

For systemic *Campylobacter* infection, imipenem or gentamicin, or a combination of the two, is the treatment of choice. Chloramphenicol, other aminoglycosides, and tetracyclines are also effective. Most isolates are resistant to penicillin, cephalosporins, vancomycin, and rifampin.

COMPLICATIONS

Postinfectious arthritis occurs in approximately 1% of patients, Guillain-Barré syndrome (or its variant, Miller Fisher syndrome) occurs in approximately 0.1%,[14] and hemolytic-uremic syndrome occurs rarely. Postinfectious arthritis presents as a sterile monoarticular or migratory polyarticular arthritis—particularly involving the knee—that begins 7 to 10 days after the onset of diarrhea and may persist for months or become chronic; 60% of patients carry the HLA-B27 haplotype, which is associated with an increased prevalence of ankylosing spondylitis and related spondyloarthropathies. In patients with Guillain-Barré syndrome, up to 40% have evidence of recent *Campylobacter* infection,[14] 20% are left with some long-term disability, and 5% die of the syndrome or its complications. Both postinfectious arthritis and Guillain-Barré syndrome are thought to be caused by autoimmune processes.

PROGNOSIS

Recurrence or relapse of *Campylobacter* infection may occur. Severe, persistent, or bacteremia-associated enteritis may occur in AIDS patients.

Salmonella Infections

EPIDEMIOLOGY

Disease caused by *Salmonella* organisms can be divided into two categories: typhoidal and nontyphoidal. *S. typhi* and *S. paratyphi* are the pathogens in typhoidal disease. Nontyphoidal disease results from all other serovars that infect humans. In the United States, nontyphoidal *Salmonella* organisms, most commonly *S. typhimurium* and *S. enteriditis*, are the second most frequently isolated cause of bacterial diarrhea (approximately 1.4 million cases annually).[15]

More than 95% of nontyphoidal *Salmonella* infections are food-borne; the remainder are nosocomial infections or are ac-

quired from pets (reptiles and birds); from infected persons; or from contaminated water, drugs, or solutions. Outbreaks have been linked to eggs, cheese, fresh fruits and vegetables, juice, dry cereal, and ice cream premix.[16] *Salmonella* organisms can be passed transovarially from chicken to egg, and most *S. enteriditis* cases are traced to undercooked eggs.[16] *Salmonella* organisms are commonly carried by farm animals, and antibiotic resistance, particularly of *S. typhimurium*, has been linked to the use of antibiotics in animal feed.[17]

In the United States, *S. typhi* and *S. paratyphi* cause fewer than 1,000 cases of infection annually, more than 70% of which are acquired abroad.[15] Humans are the only host of these pathogens, and spread is person-to-person or via contaminated foodstuffs or water.

The inoculum for *Salmonella* infection is approximately 10^5 to 10^9 organisms but is lower in infants, in persons with pernicious anemia, and in persons taking antacids or H_2 receptor blockers. Additional risk factors are old age, alteration of the endogenous bowel flora (e.g., by recent antimicrobial therapy), HIV infection, therapeutic immunosuppression, alteration of the reticuloendothelial system (e.g., malaria or *Bartonella* infection), sickle-cell disease, splenectomy, diabetes, malignancy, and rheumatologic disorders, including lupus. HIV-infected persons have a 20- to 100-fold increased risk of *Salmonella* infection and a significantly increased risk of severe invasive disease.

Classification of *Salmonella* organisms is confusing because disease may be caused by more than 2,500 serovars, all of which belong to a single species, designated *S. enterica* and *S. choleraesuis*. However, in clinical settings, serovars are commonly referred to as species (e.g., *S. enterica* serovar *enteriditis* is called *S. enteriditis)*, as will be done herein.

PATHOGENESIS

Salmonella organisms are internalized by intestinal M cells, which are specialized epithelial cells that overlie Peyer patches and are enriched in the ileum. Organisms are transported to the lymphoid tissue of the Peyer patches, from which they may enter the systemic circulation. *Salmonella* organisms have specialized mechanisms for survival inside of macrophages. Recruitment of inflammatory cells, including macrophages, into the Peyer patches may lead to their enlargement and necrosis after several weeks of infection. *Salmonella* organisms are also able to induce internalization into and transit across nonspecialized intestinal epithelial cells into deeper tissues.

Salmonella serovars vary in their pathogenesis. *S. anatum*, *S. derby*, and *S. newport* are usually limited to the intestine. *S. choleraesuis* rapidly enters the bloodstream and causes little damage in the intestine. *S. typhi* bacteremia often leads to seeding of the liver and biliary tree.

GASTROENTERITIS

Diagnosis

Clinical manifestations The incubation period for gastroenteritis caused by *Salmonella* organisms typically ranges from 6 to 72 hours. Symptoms of the disorder include nausea, vomiting, fever, diarrhea, and abdominal cramps. Stools are typically loose, of moderate volume, and do not contain blood. In rare cases, the presentation may mimic appendicitis or inflammatory bowel disease. Diarrhea is generally self-limited, lasting 3 to 7 days.

Physical examination Physical findings are nonspecific. Fever with temperatures to 39° C (102° F) for 1 to 2 days, mild abdominal tenderness, and hyperactive bowel sounds may be present. A more prolonged or more hectic fever pattern suggests bacteremia, metastatic foci, or both.

Laboratory tests Microscopic examination of stool reveals leukocytes and, rarely, red blood cells. *Salmonella* organisms are readily cultured from stool on selective media used routinely in clinical microbiology laboratories. Isolation of *Salmonella* organisms from otherwise sterile body fluids is also accomplished using routine media. The mean duration of fecal carriage after resolution of diarrhea for nontyphoidal *Salmonella* organisms is 1 month in adults and 7 weeks in young children.[18]

Differential Diagnosis

The presentation of *Salmonella* gastroenteritis is similar to that of other febrile diarrheal syndromes, including infection caused by *Campylobacter* organisms, infection caused by *Shigella* organisms, infection caused by *Y. enterocolitica*, and inflammatory bowel disease. Frankly bloody diarrhea is more suggestive of *Shigella* or enterohemorrhagic *E. coli* infection than of *Salmonella* infection.

Treatment

Antibiotics are not recommended for patients with uncomplicated gastroenteritis, because the illness is generally self-limited, and antibiotic therapy may prolong intestinal carriage.[19]

Antibiotics should be administered to patients who are severely ill or at risk for extraintestinal spread of infection, including infants; persons older than 50 years; and patients with cardiac valvular or mural abnormalities, patients with prosthetic vascular grafts, or patients who are receiving immunosuppressive therapy. Effective therapeutic agents include fluoroquinolones, trimethoprim-sulfamethoxazole, ampicillin, and third-generation cephalosporins. Because resistance to trimethoprim-sulfamethoxazole or ampicillin is common, treatment with a fluoroquinolone or a third-generation cephalosporin is appropriate when susceptibilities of the isolate are not known. Duration of therapy should be only 48 to 72 hours or until the patient becomes afebrile; longer therapy may increase the likelihood of long-term carriage.

Complications

Infants are at high risk for central nervous system infection. In some patients, *Salmonella* organisms invade the intestinal epithelium and enter the bloodstream, causing septicemia. Septicemia may in turn cause metastatic infection of other organs, including endovascular sites, the hepatobiliary tract, the spleen, bone, and, less commonly, the brain.

CARRIER STATE

The carrier state, defined as carriage of *Salmonella* organisms in the stool for more than 1 year after initial infection, occurs in 0.2% to 0.6% of persons infected with nontyphoidal *Salmonella* and 3% of persons infected with *S. typhi*.[20] The site of chronic infection is the biliary tree. The presence of stones or chronic scarring makes eradication more difficult. Long-term carriage may occur more rarely in the urinary tract, particularly in the setting of obstructive uropathy, tuberculosis, or schistosomiasis.

Table 1 Selected Therapies for Enteric Pathogens

Pathogen	*Drug (Trade Name)*	*Dose*	*Relative Efficacy*	*Cost ($)*	*Comment*
Campylobacter	Erythromycin	12.5 mg/kg p.o. every 6 hr for 7 days	Drug of choice	1.00–1.49/day	*Campylobacter* enteritis strains resistant to erythromycin and related drugs have been isolated in Thailand
	Azithromycin (Zithromax)	500 mg p.o. every day for 3 days	Has been used successfully	10.00–19.99/day	*Campylobacter* enteritis strains resistant to azithromycin and related drugs have been isolated in Thailand
	Ciprofloxacin (Cipro)	500 mg p.o. twice a day for 3 days	Has been used successfully	7.00–7.99/day	*Campylobacter* enteritis strains resistant to ciprofloxacin and related drugs are increasing
Salmonella	Ciprofloxacin (Cipro)	500 mg p.o. twice a day for 10 days	Drug of choice for isolates resistant to ampicillin and trimethoprim-sulfamethoxazole	7.00–7.99/day	Antimicrobial therapy for *Salmonella* enteric fever; must confirm organism susceptibility to nalidixic acid and ciprofloxacin; if nalidixic acid–resistant, use higher doses of ciprofloxacin (10 mg/kg twice a day)
	Ceftriaxone (Rocephin)	1–2 g I.M. or I.V. daily for 10–14 days	Drug of choice for isolates resistant to ampicillin and trimethoprim-sulfamethoxazole	60.00–69.99/day	Antimicrobial therapy for *Salmonella* enteric fever
	Amoxicillin	1 g p.o. every 6 hr for 10–14 days	For susceptible isolates only	0.25–0.49/day	
	Trimethoprim-sulfamethoxazole (Bactrim)	1 double-strength tablet p.o. twice a day for 10–14 days	For susceptible isolates only	0.10–0.24/day	
	Dexamethasone	3 mg/kg I.V. once, then 1 mg/kg I.V. every 6 hr for 48 hr	Adjunct to antimicrobial therapy when disease is accompanied by shock, coma, stupor, obtundation, or delirium	80.00–89.99/day	Glucocorticosteroid therapy for *Salmonella* enteric fever; continuation beyond 48 hr may increase the risk of relapse
	Ampicillin or amoxicillin	2–4 g/day in divided doses for 4–6 wk	Treatment of choice	0.75–0.99/day or 0.25–0.49/day	Antibiotic therapy for *Salmonella* carrier state
	Trimethoprim-sulfamethoxazole (Bactrim)	1 double-strength tablet p.o. twice a day for 4–6 wk	For penicillin-allergic patients or ampicillin-resistant isolates	0.10–0.24/day	
	Ciprofloxacin (Cipro)	500 mg p.o. twice a day for 4–6 wk	For penicillin-allergic patients or ampicillin-resistant isolates	7.00–7.99/day	
Shigella	Trimethoprim-sulfamethoxazole	160/800 mg p.o. twice a day for 3–5 days	Treatment of choice	0.10–0.24/day	Antibiotic therapy for *Shigella* enteritis acquired in the United States
	Norfloxacin (Noroxin)	400 mg p.o. twice a day for 3–5 days	Alternative therapy or for trimethoprim-resistant isolates	4.00–4.99/day	Antibiotic therapy for *Shigella* enteritis acquired in the United States
			Treatment of choice		Antibiotic therapy for *Shigella* enteritis acquired in Southeast Asia, Africa, the Middle East, Japan, South America, Spain, or Canada; trimethoprim resistance is common
	Ciprofloxacin (Cipro)	500 mg p.o. twice a day for 3–5 days	Alternative therapy or for trimethoprim-resistant isolates	7.00–7.99/day	Antibiotic therapy for *Shigella* enteritis acquired in the United States
			Treatment of choice		Antibiotic therapy for *Shigella* enteritis acquired in Southeast Asia, Africa, the Middle East, Japan, South America, Spain, or Canada; trimethoprim resistance is common
	Ofloxacin (Floxin)	300 mg p.o. twice a day for 3–5 days	Alternative therapy or for trimethoprim-resistant isolates	7.00–7.99/day	Antibiotic therapy for *Shigella* enteritis acquired in the United States
			Treatment of choice		Antibiotic therapy for *Shigella* enteritis acquired in Southeast Asia, Africa, the Middle East, Japan, South America, Spain, or Canada; trimethoprim resistance is common

Table 1 (*continued*)

Pathogen	*Drug (Trade Name)*	*Dose*	*Relative Efficacy*	*Cost ($)*	*Comment*
Yersinia	Gentamicin	5 mg/kg/day in divided doses	Suggested therapy	3.00–3.99/day	Antibiotic therapy for *Y. enterocolitica* septicemia; may be used alone or in combination with other antibiotics listed as therapy for *Y. enterocolitica* in this table
	Doxycycline	100 mg I.V. every 12 hr	Suggested therapy	—	
	Trimethoprim-sulfamethoxazole	8–10 mg trimethoprim/kg/day I.V. in divided doses every 6 hr	Suggested therapy	10.00–19.99/day	
	Ciprofloxacin (Cipro)	200–400 mg I.V. every 12 hr	Suggested therapy	60.00–69.99/day	
	Ampicillin	100–200 mg/kg/day I.V. in divided doses	Suggested therapy	7.00–7.99/day	Antibiotic therapy for *Y. pseudotuberculosis* septicemia
	Streptomycin	20–30 mg/kg/day I.M.	Suggested therapy	3.00–3.99/day	
	Tetracycline	20–30 mg/kg/day p.o. or I.V. in divided doses	Suggested therapy	—	
Vibrio	Tetracycline	10–12.5 mg/kg p.o. four times a day for 2 days	Treatment of choice	0.10–0.24/day	Antibiotic therapy for *V. cholerae*; eradicates the organism
		1 g p.o. once	Alternative therapy	—	Antibiotic therapy for *V. cholerae*; controls symptoms but may result in asymptomatic bacteriologic relapse
	Ciprofloxacin (Cipro)	1 g p.o. once	For tetracycline-resistant strains	—	Antibiotic therapy for *V. cholerae*
	Erythromycin	250 mg p.o. four times a day for 3 days	For tetracycline-resistant strains	1.00–1.49/day	
	Trimethoprim-sulfamethoxazole	1 double-strength tablet twice a day for 3 days	For tetracycline-resistant strains	0.10–0.24/day	
Helicobacter	Proton-pump inhibitor (omeprazole [Prilosec] or lansoprazole [Prevacid]) *or* Ranitidine bismuth citrate (Tritec) *plus* Two of the following: Amoxicillin Clarithromycin (Biaxin) Metronidazole (Flagyl)	Omeprazole 20 mg or lansoprazole 30 mg p.o. twice a day for 14 days 400 mg p.o. twice a day for 14 days 1 g p.o. twice a day 500 mg p.o. twice a day 500 mg p.o. twice a day for 14 days	Suggested therapy	100.00–149.99/mo 90.00–99.99/mo 100.00–149.99/mo 0.25–0.49/day 7.00–7.99/day 0.50–0.74/day	Triple-therapy combination[50]; for resistant *H. pylori*, ranitidine bismuth citrate may be slightly better than proton-pump inhibitor; metronidazole may cause nausea, metallic taste, headaches, or disulfiram-like reaction (avoid alcohol)
	Proton-pump inhibitor (omeprazole [Prilosec] or lansoprazole [Prevacid]) *plus* Tetracyline *plus* Bismuth subsalicylate (Pepto-Bismol) *plus* Metronidazole (Flagyl)	Omeprazole 20 mg or lansoprazole 30 mg p.o. twice a day for 14 days 500 mg p.o. four times a day for 14 days 525 mg p.o. four times a day for 14 days 500 mg p.o. three times a day for 14 days	Suggested therapy	100.00–149.99/mo 90.00–99.99/mo 0.10–0.24/day 3.00–3.99/day 0.50–0.74/day	Quadruple-therapy combination[50]; metronidazole may cause nausea, metallic taste, headaches, or disulfiram-like reaction (avoid alcohol)
	Proton-pump inhibitor (omeprazole [Prilosec] or lansoprazole [Prevacid]) *plus* Tetracyline *plus* Bismuth subsalicylate (Pepto-Bismol) *plus* Furazolidone (Furoxone)	Omeprazole 20 mg or lansoprazole 30 mg p.o. twice a day for 14 days 500 mg p.o. four times a day for 14 days 525 mg p.o. four times a day for 14 days 500 mg p.o. three times a day for 14 days	For patients who have failed combination therapies listed above	100.00–149.99/mo 90.00–99.99/mo 0.10–0.24/day 3.00–3.99/day —	Quadruple-salvage-therapy combination[50]; furazolidone, a monoamine oxidase inhibitor, may cause nausea, vomiting, headache, tachycardia, and elevated blood pressure; must consider potential interactions with other medications and foods

Diagnosis

Clinical manifestations The carrier state is generally asymptomatic. If the carrier state is accompanied by obstructive uropathy or gallstones, symptoms associated with that syndrome may be present.

Physical examination The physical examination of a patient in the carrier state is unremarkable.

Laboratory tests The carrier state is documented by culture of *Salmonella* organisms utilizing routine techniques.

Treatment

Ampicillin or amoxicillin [*see Table 1*] will eradicate the *Salmonella* organisms in 80% of persons in the carrier state. Alternative therapy is trimethoprim-sulfamethoxazole or ciprofloxacin [*see Table 1*]. Long-term gallbladder carriage can be effectively eliminated by cholecystectomy accompanied by 10 to 14 days of any of the regimens recommended for treatment of bacteremia. However, this approach should be reserved for individuals in whom eradication is required for public health reasons (e.g., food handlers and health care workers).

BACTEREMIA AND VASCULAR INFECTION

Nontyphoidal *Salmonella* organisms have a propensity to colonize sites of vascular abnormality, such as prosthetic vascular grafts, atherosclerotic plaques, and aneurysms.[21] The presence of high-grade bacteremia (more than 50% of three or more blood cultures are positive) is suggestive of endovascular infection.

Diagnosis

Clinical manifestations Patients with endovascular infection typically have high fevers that persist even after several days of therapy.

Physical examination The physical examination may reveal tenderness at the site of the infection, but it may also yield normal results.

Laboratory tests High-grade bacteremia (i.e., more than 50% of three or more blood cultures are positive) is suggestive of endovascular infection. Echocardiography and other imaging studies should be conducted to search for endovascular lesions.

Treatment

Life-threatening bacteremia, endovascular infection, or focal metastatic infection should be empirically treated with both a fluoroquinolone and a third-generation cephalosporin until antibiotic susceptibilities are known. For documented or suspected endovascular infection, 6 weeks of I.V. therapy with a β-lactam antibiotic, such as ampicillin or a cephalosporin, should be administered. For low-grade bacteremia, 7 to 14 days of I.V. therapy is adequate. Endovascular infection may require surgical intervention. However, in cases in which surgical resection of infected grafts is not feasible, lifelong oral suppressive therapy has been successful.[22]

In patients with AIDS, a first episode of *Salmonella* bacteremia should be treated with 7 to 14 days of I.V. antibiotics, followed by 4 weeks of an oral fluoroquinolone. AIDS patients who experience relapses of *Salmonella* bacteremia should be treated with an oral fluoroquinolone or trimethoprim-sulfamethoxazole for long-term suppression.

ENTERIC FEVER

Diagnosis

Clinical manifestations The incubation period for organisms that cause enteric fever ranges from 5 to 21 days. Initially, diarrhea or constipation may develop, followed by fever and then by headache, malaise, anorexia, myalgias, arthralgia, cough, and sore throat.[23] Approximately 1 week later, systemic toxemia, neuropsychiatric manifestations (psychosis and confusion), coryza, cough, sore throat, chest pain, and nausea and vomiting with or without abdominal pain may occur. Symptoms resolve spontaneously after 2 to 4 weeks.

Physical examination Results of the physical examination vary depending on the phase of the illness. During the toxemic

Table 2 Probability of Positive Cultures from Various Sites during Enteric Fever[24]

Phase of Illness (Time since Initial Exposure)	*Disease Manifestations*	*Cultures*				
		Blood	*Stool*	*Urine*	*Bone Marrow*	*Rose Spots*
Incubation period (0–1 wk)	Diarrhea in 10%–20%; constipation in some	Negative	Transiently positive	Negative	Negative	Negative
Active invasion (1–2 wk)	Fever, headache, malaise, anorexia, myalgia, arthralgia, cough, and sore throat	80%–90% positive	Negative	Negative	Negative	Negative
Established disease (2–4 wk)	Systemic toxemia, neuropsychiatric manifestations (psychosis and confusion), coryza, cough, sore throat, chest pain, nausea, vomiting, and abdominal pain	80%–90% positive	80% positive	25% positive	90% positive	60% positive
Convalescent period (4–5 wk)	—	Negative, except with ongoing disease or relapse	50% positive	10% positive	Decreasingly positive	Decreasingly positive
Late focal complications (> 5 wk)	Cholecystitis, osteomyelitis, soft tissue abscess	Negative, except with ongoing disease or relapse	Decreasingly positive	Decreasingly positive	—	—

Table 3 Vaccination for Enteric Fever

Vaccine	*Dose*	*Relative Efficacy*	*Cost ($)*	*Comment*
Ty21a (Vivotif Berna)	One enteric-coated capsule 1 hr before a meal every day for 4 days	43%–96%; not effective if administered with concomitant antimicrobial or antimalarial therapy	18.92	Live-attenuated vaccine; not recommended for immunosuppressed or children younger than 6 yr; boost every 5 yr; few side effects
Vi capsular polysaccharide [ViCPS] (Typhim Vi)	25 μg in 0.5 ml I.M. once	55% in one study[51]	30.21	Side effects: fever (1%), headache (1.5%–3%), local erythema or induration (7%); not recommended for children younger than 2 yr; boost every 2 yr
Heat-killed whole organism *Salmonella typhi*	0.5 ml I.M. for 2 doses administered more than 4 wk apart	51%–77%	9.39	Side effects: fever (17%–29%), severe headache (10%), pain at injection site (35%–60%), rare severe reactions; boost (0.1 ml) every 3 yr; dose for children 6 mo to 10 yr of age is 0.25 ml

phase of the illness, the patient's condition is acute. In 50% of patients, splenomegaly and hepatomegaly are present. In 30% of patients, rose spots (2 to 4 mm, slightly raised, discrete, irregular, blanching pink macules) appear on the anterior chest in crops of 5 to 15. The spots last 3 to 4 days and fade without a scar.

Laboratory tests The probability of culturing the organism from various sites varies during the course of illness [*see Table 2*].[24] Cultures of bone marrow and punch biopsies of rose spots may be positive when other cultures are negative, particularly during or after antibiotic therapy. Leukopenia, anemia, and moderately elevated liver function test results and muscle enzyme levels are common. Leukocytosis, transient thrombocytopenia, and clotting abnormalities are less common.

Differential Diagnosis

Enteric fever may be confused with other systemic febrile illnesses.

Treatment

Enteric fever can be successfully treated with ampicillin or amoxicillin, third-generation cephalosporins, fluoroquinolones, or trimethoprim-sulfamethoxazole [*see Table 1*]. However, given the prevalence of organisms resistant to ampicillin and third-generation cephalosporins among isolates from the Indian subcontinent, Southeast Asia, and Africa, therapy should consist of a fluoroquinolone or trimethoprim-sulfamethoxazole until antibiotic susceptibilities are known.[25] In cases of severe disease (associated shock, coma, stupor, obtundation, or delirium), dexamethasone [*see Table 1*] has been shown to improve survival,[26] although continuation of steroids beyond 48 hours may increase the rate of relapse.[27]

Several enteric fever vaccines are available [*see Table 3*]. Vaccination is recommended for individuals traveling to developing countries who will have prolonged exposure to contaminated food and drink. Vaccination is an adjunct and not a replacement for careful eating and drinking.

Complications

Nonmetastatic complications, which include pneumonia (5% of cases), intestinal hemorrhage (3% to 20% of cases), intestinal perforation (2% to 3% of cases), acute cholecystitis (2% of cases), and myocarditis (1% to 2% of cases), often occur 2 to 4 weeks after the initial onset of symptoms. Less commonly, endocarditis, pericarditis, orchitis, and splenic or liver abscesses may occur. In the setting of intestinal perforation, repeat blood cultures should be drawn, and antibiotics should be broadened to cover anaerobic and aerobic bowel flora.

Shigella Infections

EPIDEMIOLOGY

Shigella organisms are estimated to cause 450,000 cases of diarrhea annually in the United States,[15] and they are an important cause of diarrhea and death worldwide.[28] *Shigella* includes four species: *S. dysenteriae, S. sonnei, S. boydii,* and *S. flexneri*. In industrialized countries, *S. sonnei* is currently most common, and *S. flexneri* accounts for essentially all other cases. In developing countries, *S. dysenteriae* is also common. Humans are the only natural host of *Shigella* organisms, and they are spread by fecal-oral contact or, in 20% of cases, through contaminated food or water.

The infectious inoculum is as few as 10 to 100 organisms.[29] Consequently, outbreaks spread readily and recurrences are common in day care centers, mental institutions, and other crowded settings. Disease is most common among young children, and secondary infection rates in families are as high as 20%.

PATHOGENESIS

Initially, *Shigella* organisms colonize the proximal small bowel, where secretion of an enterotoxin probably causes the initial symptoms. The organisms then pass into the colon, where they invade and spread through the epithelial layer. *Shigella* organisms induce the release of proinflammatory cytokines. Bacterial spread, in conjunction with the intense acute inflammatory response, leads to the formation of ulcerations and microabscesses in the colonic epithelium. *Shigella* organisms rarely gain access to the bloodstream or infect deeper tissues. The toxin Stx (formerly called Shiga toxin) is present only in *S. dysenteriae* and contributes to the more severe diarrhea that can accompany *S. dysenteriae* infection. Related enterotoxins are produced by members of all *Shigella* species.

DIAGNOSIS

Clinical Manifestations

The incubation period for *Shigella* organisms is typically 3 days, with a range of 1 to 7 days. Initial symptoms usually include fever, abdominal cramps, and watery diarrhea, which are followed by abdominal cramps, tenesmus, rectal urgency, and

small-volume diarrhea. Diarrhea during this later stage is frequent (eight to 10 episodes a day), and stools may be bloody with mucus. If left untreated, disease is generally self-limited, lasting 7 days or less.

Physical Examination

Physical findings are nonspecific. Patients may appear toxic and have a high temperature (up to 41° C [106° F]). Abdominal tenderness, particularly in the left lower quadrant, and hyperactive bowel sounds are common.

Laboratory Tests

Stool cultures are usually positive during the acute illness. Later, *Shigella* organisms may be isolated by direct culture of material from rectal ulcerations. Direct microscopic examination of stool stained with methylene blue (or another stain) is extremely important because the presence of abundant leukocytes in the proper clinical setting is strongly suggestive of shigellosis. In *Salmonella* enteritis, *Campylobacter* enteritis, and idiopathic ulcerative colitis, leukocytes may also be present in the stool. The white blood cell count may be elevated with an increase in the percentage of immature forms, and metabolic abnormalities may be present.

DIFFERENTIAL DIAGNOSIS

In the setting of high fever, tenesmus, rectal urgency, and diarrhea with blood- and mucus-containing stools, *Shigella* infection should be suspected. However, *Shigella* infection can resemble any febrile diarrheal syndrome, including those caused by *Campylobacter*, *Y. enterocolitica*, and *Salmonella* organisms and by inflammatory bowel disease.

TREATMENT

For patients with significant dehydration, particularly young children and the elderly, oral and, if necessary, intravenous rehydration should be administered. Antibiotics are not essential, but administration of an antibiotic to which the organism is susceptible has been shown to shorten the course of clinical illness and the period of fecal excretion.[30,31] Given the ease of person-to-person *Shigella* transmission, many experts recommend that all individuals with documented infection be treated for public health reasons. For *Shigella* infection acquired in the United States with an unknown antibiotic susceptibility pattern, the treatment of choice is trimethoprim-sulfamethoxazole [*see Table 1*] because ampicillin resistance is common. The alternative therapy or therapy for trimethoprim-resistant isolates is norfloxacin, ciprofloxacin, or ofloxacin [*see Table 1*]. For *Shigella* acquired in Southeast Asia, Africa, the Middle East, Japan, South America, Spain, or Canada, the treatment of choice is a fluoroquinolone, because resistance to trimethoprim-sulfamethoxazole is common [*see Table 1*]. Azithromycin has also been used successfully. Cephalosporins have limited efficacy.

COMPLICATIONS

Complications are generally rare but include intestinal obstruction in the setting of severe colonic disease (2.5% of cases), bacteremia (4% of cases), colonic perforation (1.7% of fatal cases), and toxic megacolon (3% of *S. dysenteriae* infections). Proctitis and rectal prolapse (in children) also occur as complications. Metabolic disturbances are relatively common, although severe dehydration is uncommon because stool volume is generally low. Seizures occur in approximately 5% of infected children, generally in the setting of high fever and metabolic abnormalities.[32]

Reactive arthritis may occur 1 to 2 weeks after diarrhea, either alone or accompanied by conjunctivitis and urethritis (Reiter syndrome); 70% of these patients have the HLA-B27 haplotype. Although most commonly caused by enterohemorrhagic *E. coli* infection, hemolytic-uremic syndrome may occur after *Shigella* infection. Hemolytic-uremic syndrome is associated with infection caused by *S. dysenteriae* and is thought to be mediated by Stx.

PROGNOSIS

Prognosis is generally excellent. Settings that predispose to recurrence of disease are those with suboptimal hygiene, such as day care centers or crowded living conditions.

Yersinia Infections

EPIDEMIOLOGY

All three *Yersinia* species, *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*, are pathogenic for humans. *Y. enterocolitica* and *Y. pseudotuberculosis* are widely distributed in nature and cause a variety of zoonoses. In the United States, *Y. enterocolitica* is estimated to cause approximately 100,000 cases of diarrhea annually,[15] predominantly in children. *Y. pseudotuberculosis* is a sporadic cause of purulent mesenteric adenitis.

Transmission of *Y. enterocolitica* is usually by ingestion of contaminated foodstuffs, including meat, milk and other dairy products, mussels, tofu, and oysters. Person-to-person transmission has also been reported, with a significant rate of secondary cases within families.[33] Transfusion of *Y. enterocolitica*–contaminated red blood cells has been a cause of severe transfusion reactions with high mortality.[34] Among the approximately 60 serogroups of *Y. enterocolitica*, three (O:3, O:8, and O:9) cause most human disease. Individuals with underlying cirrhosis or hemochromatosis, particularly if being treated with desferrioxamine, are at increased risk for bacteremia and metastatic disease.

PATHOGENESIS

Y. enterocolitica and *Y. pseudotuberculosis* transit the stomach, invade the mucosa of the terminal ileum at Peyer patches, and colonize mesenteric lymph nodes. *Yersinia* organisms are able to both replicate intracellularly and bind to macrophages and inhibit phagocytosis. The inflammatory response that ensues in the terminal ileum may be clinically mistaken for appendicitis or Crohn ileitis.

DIAGNOSIS

Clinical Manifestations

In older children and adults, *Y. enterocolitica* causes a febrile diarrheal syndrome in which the prominent symptom is colicky abdominal pain, predominantly localized to the right lower quadrant. In younger children, the illness may include vomiting and bloody diarrhea. Exudative pharyngitis has also been reported.[35] Septicemia, which occurs in individuals with iron overload or other chronic diseases, is often complicated by metastatic abscesses formation. *Y. pseudotuberculosis* causes mesenteric adenitis, which mimics acute appendicitis, with fever and right lower quadrant pain.

Physical Examination

On physical examination, patients with *Yersinia* infection are typically found to have fever and right lower quadrant tenderness.

Laboratory Tests

Leukocytosis is common. Microscopic examination of the stool will reveal fecal leukocytes. Whereas isolation of *Yersinia* organisms from normally sterile body fluids is straightforward, isolation from stool is difficult because the organisms appear similar to and grow more slowly than other Enterobacteriacae organisms; selective media for *Yersinia* organisms have not been developed. *Yersinia* serology, which turns positive 1 to 2 weeks after infection, may be useful in the setting of one of the autoimmune syndromes that may complicate infection. *Y. enterocolitica* is carried in the stool for an average of 1 month after the resolution of symptoms.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of diarrheal disease with polymorphonuclear leukocytes in the stool is enteritis caused by infection with *Campylobacter*, *Salmonella*, *Shigella*, or *Y. enterocolitica* or caused by idiopathic ulcerative colitis. Mesenteric adenitis caused by *Yersinia* infection may mimic appendicitis.

TREATMENT

Enteritis and mesenteric adenitis caused by *Y. enterocolitica* or *Y. pseudotuberculosis* are generally self-limited, and no clear role for antibiotic therapy has been established. Although direct proof is lacking, in principle, antiperistaltic agents should be avoided.

Septicemia caused by either *Y. enterocolitica* or *Y. pseudotuberculosis* should be aggressively treated with antibiotics because mortality is high. The suggested agents for treatment of *Y. enterocolitica* septicemia are gentamicin, doxycycline, trimethoprim-sulfamethoxazole, or a fluoroquinolone, alone or in combination [*see Table 1*].[36] *Y. enterocolitica* organisms produce a β-lactamase and are therefore resistant to penicillins and cephalosporins. The suggested agents for treatment of *Y. pseudotuberculosis* septicemia are ampicillin, streptomycin, and tetracycline [*see Table 1*]. For either *Y. enterocolitica* or *Y. pseudotuberculosis*, enteric precautions should be adopted.

COMPLICATIONS

Either reactive polyarthritis or erythema nodosum complicates 1% to 5% of *Y. enterocolitica* infections in adults in the United States and 10% to 30% of such infections in Scandinavia. On occasion, *Y. pseudotuberculosis* infections in adults are also complicated by reactive polyarthritis or erythema nodosum. These disorders may be the presenting complaint. Therefore, in individuals with such complaints, *Yersinia* infection should be considered and serology and cultures should be performed. These complications of *Yersinia* infections rarely occur in young children.

Reactive polyarthritis develops 2 days to 1 month after the onset of gastrointestinal symptoms. It persists for more than a month in two thirds of patients and for more than 4 months in one third of patients. Knees, ankles, toes, fingers, and wrists may be involved. Synovial fluid usually contains fewer than 25,000 white blood cells/ml, with a preponderance of polymorphonuclear leukocytes and negative cultures. The disorder may be particularly severe in individuals with the HLA-B27 haplotype, presenting as full-blown ankylosing spondylitis or as Reiter syndrome.

Erythema nodosum develops 2 to 20 days after the onset of gastrointestinal symptoms. It typically resolves spontaneously within a month and is twice as common in women than in men. Skin lesions are usually on the legs and trunk.

PROGNOSIS

The prognosis for *Y. enterocolitica* enteritis is excellent. In contrast, *Y. enterocolitica* septicemia is fatal in 50% of cases and *Y. pseudotuberculosis* septicemia is fatal in 75% of cases, despite appropriate antimicrobial therapy.

Vibrio Infections

CHOLERA

Epidemiology

The natural reservoir of *V. cholerae* is aquatic invertebrates in brackish or marine environments. Certain strains have become pathogenic for humans and have caused seven pandemics throughout history, the most recent of which began in Asia in 1961. There are 139 serotypes of *V. cholerae*, defined by the O surface antigen. Initially, the seventh pandemic was caused by an O1 strain of biotype El Tor, but in 1992, a new strain emerged from the pandemic strain that carried a new O antigen, which was designated O139. The emergence of the new strain was concurrent with rapid spread of cholera through Latin America and Africa. Currently, the continuing pandemic is caused by two strains, O139 and an O1 strain that differs from the initial O1 strain and likely emerged from the O139 strain.[37]

Spread of cholera is primarily through ingestion of fecally contaminated water and food, especially, but not exclusively, seafood. Only about 50 cases of cholera are estimated to occur in the United States annually,[15] predominantly in persons who acquire it while traveling in endemic areas. The rate of secondary infection in households is very high, suggesting that person-to-person transmission is occurring. Because *V. cholerae* is sensitive to gastric acid, the infectious inoculum in normal hosts is 10^{10} organisms, but in persons taking antacids or who have achlorhydria, the inoculum may be 10^6 organisms or fewer. Natural immunity to cholera is long lasting, with specificity to the O1 antigen. Thus, in endemic areas, disease is most common in children, but nonimmune adults traveling in the region are susceptible. Furthermore, alteration of the O antigen, as seen with the shift from O1 to O139 in 1992, leads to widespread susceptibility in adults.

Pathogenesis

V. cholerae colonize the proximal small intestine and secrete one major and at least two minor enterotoxins. The major toxin, cholera toxin (CTX), binds the ganglioside GM1 on the intestinal epithelium and activates adenylate cyclase, which in turn leads to the accumulation of cyclic adenosine monophosphate (cAMP). Increased levels of cAMP cause chloride secretion and inhibit sodium absorption, resulting in the characteristic profuse isotonic watery diarrhea.

V. cholerae does not invade intestinal tissue, and there is no tissue inflammatory response. Absorptive processes unaffected by cAMP, including glucose-facilitated salt and water absorption, are unaltered, allowing oral rehydration with glucose-electrolyte solutions.

Diagnosis

Clinical manifestations The hallmark of cholera is massive watery diarrhea consisting of thin, gray-brown, mucus-containing fluid (so-called rice-water stools). The diarrhea begins a few days after initial exposure to *V. cholerae* and is soon followed by copious vomiting without retching. Notably absent are fever (which occurs in only 5% of cases), abdominal pain, tenesmus, and strain. Muscle cramps may occur as a result of electrolyte shifts. Dehydration ranges from mild to severe. A rare variant in the presentation is so-called cholera sicca, in which abdominal ileus and distention occur, and diarrhea is absent. If left untreated, symptoms of cholera resolve in 1 to 7 days.

Physical examination Patients appear anxious and restless, with sunken eyes, dry mucous membranes, poor skin turgor, hyperactive bowel sounds, hypotension, and tachycardia. Obtundation may occur with severe dehydration. As much as 1 L/hr of fluid may be excreted in stools.

Laboratory tests Microscopic examination of Gram stains of stool reveals the small (1 to 3 μm long), comma-shaped, gram-negative organisms. *V. cholerae* grow well on a variety of selective media; TCBS (thiosulfate-citrate-bile salt-sucrose) agar is commonly used.

In patients with moderate-to-severe dehydration, packed cell volume, serum specific gravity, and total protein levels are elevated. In patients with severe disease, prerenal azotemia, metabolic acidosis with an elevated anion gap, normal to low serum potassium levels, normal to slightly low serum sodium and chloride levels , and leukocytosis are seen.

Differential Diagnosis

In patients with severe dehydration, cholera resembles little else. In patients with mild to moderate dehydration, the presentation is similar to that of enterotoxigenic *E. coli* (ETEC) or rotavirus.

Treatment

The cornerstone of management is prompt replacement of lost fluid and electrolytes. Replacement fluid and electrolytes should be administered intravenously for patients with severe dehydration, patients with moderate dehydration who do not tolerate oral intake, and patients who are purging large amounts of fluid (10 to 20 ml/kg/hr). Initially, replacement fluid should consist of lactated Ringer solution or a solution containing 5 g NaCl, 4 g $NaHCO_3$, and 1 g KCl for each liter of sterile distilled water. The goal is to return the patient to normal hydration status within 4 hours, which, in severely dehydrated persons, may require administration of fluids at 50 to 100 ml/kg/hr. Once rehydration has been accomplished, as determined by hemodynamic status and urine production, normal hydration status should be maintained by replacement of continuing losses of fluid. For those patients with mild disease, rehydration can be accomplished by oral administration of a glucose-electrolyte solution that consists of 20 g glucose, 3.5 g NaCl, 2.5 g $NaHCO_3$, and 1.5 g KCl for each liter of water.

Antibiotics are an adjunct to rehydration therapy; they can reduce the duration of symptoms and infection, as well as the overall fluid requirements. Tetracycline is the drug of choice [*see Table 1*]. Tetracycline-resistant isolates have been identified in Latin America, Bangladesh, and Tanzania; appropriate therapy for these isolates is ciprofloxacin, erythromycin, or trimethoprim-sulfamethoxazole. The only cholera vaccine (Wyeth Laboratories) commercially available in the United States provides protection in only 30% to 50% of cases and has a high rate of side effects.[38] In general, the vaccine should be administered only if required for entry into a foreign country.

Complications and Prognosis

Acute renal failure has complicated cholera in patients with severe dehydration who did not receive proper rehydration. In patients with severe dehydration who receive no treatment, mortality is about 50%.

OTHER INFECTIONS CAUSED BY *VIBRIO* SPECIES

In the United States, other *Vibrio* species are estimated to cause approximately 8,000 cases of disease annually.[15] These species include *V. parahaemolyticus*, *V. vulnificus*, and nontoxigenic strains of *V. cholerae*.

V. parahaemolyticus

V. parahaemolyticus is a common cause of diarrhea in the Chesapeake Bay area, the Gulf Coast region, and the Pacific Northwest. The organism is ubiquitous in coastal waters and is typically acquired by ingestion of raw or undercooked shellfish. Clinical illness most commonly consists of explosive watery diarrhea, crampy abdominal pain, and low-grade fever. Wound infections and septicemia also occur. The organism can be cultured from stool on selective media such as TCBS and from sterile body fluids in routine blood culture media or on blood agar plates. The illness is self-limited, lasting less than 7 days. Antibiotic treatment has not been shown to shorten the course of the illness, although if administered, tetracycline is probably the best choice because of its documented efficacy against *V. cholerae*.

V. vulnificus

V. vulnificus, an important cause of serious disease, can be isolated from waters of the eastern and western coasts of the United States, as far north as Cape Cod and the northern Pacific coast,[39] particularly during the summer months. In compromised persons, particularly those with chronic liver disease or iron-overload states but also other immunosuppressed persons, *V. vulnificus* can cause an overwhelming sepsis, accompanied by metastatic cutaneous lesions in the form of hemorrhagic bullae or vesicles that evolve into necrotic ulcers. This syndrome with bacteremia is fatal in over 50% of cases. More than 90% of patients have consumed raw oysters within a week of the onset of illness. Therapy consists of adequate debridement and parenteral antibiotics. Tetracycline is probably the drug of choice; either cefotaxime or ciprofloxacin is an alternative. Animal studies suggest that the combination of tetracycline and cefotaxime may be better than either drug alone.[40]

In both immunocompromised and healthy hosts, *V. vulnificus* can cause a second syndrome—a rapidly progressive cellulitis, fasciitis, or myositis that follows infection of a superficial wound, typically after cleaning shellfish. In addition to antimicrobials, incision and drainage with debridement are often necessary.

Non-O1 V. cholerae

Strains of *V. cholerae* (non-O1 *V. cholerae*) that do not produce cholera toxin can cause an acute febrile gastroenteritis. Non-O1 *V. cholerae* are distributed worldwide in saltwater and freshwa-

ter. Exposure is generally by ingestion of raw shellfish, particularly oysters. Diarrhea can range from mild to severe. Usually, no treatment is necessary, but in severe disease, rehydration and tetracycline are recommended.

Helicobacter Infections

EPIDEMIOLOGY

Helicobacter pylori is directly linked to the development of duodenal ulcer, gastric ulcer disease, and gastritis and to the long-term complications of gastric adenocarcinoma and gastric mucosa–associated lymphoid tissue (MALT) lymphoma. The organism is believed to be the cause of peptic ulcer disease in 90% of individuals with duodenal ulcers and in 70% to 90% of individuals with gastric ulcers. Cure of the infection significantly decreases the risk of long-term malignant complications.[41] Whether *H. pylori* also plays a role in peptic ulcer disease in the setting of nonsteroidal anti-inflammatory drug usage is less clear.

H. pylori infection is the most common long-term bacterial infection worldwide, with higher prevalence in developing countries than in industrialized countries. In the United States, the prevalence is 10% for persons 18 to 30 years of age, 30% to 45% for persons 30 to 60 years of age, and 50% for persons older than 60 years.[42,43] The prevalence is severalfold higher in African Americans and Hispanics than in non-Hispanic whites,[44] which is likely a reflection of the impact of socioeconomic status and living conditions on transmission as well as possibly an indication of a minor role of hereditary susceptibility.

Evidence suggests that transmission occurs from person to person by the fecal-oral or oral-oral route.[45] Humans appear to be the major reservoir, although the organism has also been isolated from other primates and domestic cats. A second *Helicobacter species*, *H. heilmanii*, has been observed in the stomachs of humans,[46] but whether it has a role in the pathogenesis of gastric disease is unknown.

PATHOGENESIS

H. pylori organisms colonize the gastric mucus and adhere to gastric epithelial cells. They survive the acidic environment of the stomach by surrounding themselves with pH-buffering ammonia that is produced from hydrolysis of urea by a bacterial urease. *H. pylori* induces autocrine stimulation and proinflammatory cytokine release, causing alterations in acid secretion and acute and chronic inflammation that, in turn, potentiate the disease process and contribute to the development of the malignant complications of infection. The exact mechanisms are not fully understood, although a cluster of genes known as the *cag* pathogenicity island appears to be involved.

DIAGNOSIS

Clinical Manifestations

Acute *H. pylori* infection produces symptoms of gastroenteritis in up to 60% of patients, including nausea, vomiting, epigastric pain, and fever. Symptoms last 3 to 14 days. A period of hypochlorhydria begins a few weeks after clinical symptoms resolve and lasts 2 to 8 months. Hypochlorhydria is associated with gastritis, which causes the symptoms that frequently bring patients to medical attention.

Physical Examination

During the acute gastroenteritis syndrome, physical findings are nonspecific. Epigastric tenderness and fever may be present. Long-term findings are those of gastritis or peptic ulcer disease.

Laboratory Tests

Three of the more useful diagnostic tests in the initial diagnosis of *H. pylori* infection are (1) the carbon-13 (^{13}C) urea breath test (UBT), (2) the stool antigen test (HpSA), and (3) the enzyme-linked immunosorbent assay (ELISA) serology test [*see Table 4*]. The UBT and HpSA are noninvasive. The UBT does not pose a significant radiation exposure risk and is therefore safe for pregnant women and children. Because proton-pump inhibitors reduce the load of *H. pylori*, they should be stopped 2 weeks before either the UBT or the HpSA is performed.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of the acute gastroenteritis syndrome caused by *H. pylori* includes other causes of acute small intestinal infection, including toxin-induced gastroenteritis caused by *Clostridium perfringens, Bacillus cereus, Staphylococcus aureus*, ETEC, giardiasis, and viral gastroenteritis.

TREATMENT

Given the potential serious long-term complications of *H. pylori* infection, any individual who has documented infection should be treated. Recommended treatment regimens all consist of combinations of antibiotics and antisecretory agents or a bismuth formulation [*see Table 1*]. For duodenal ulcer disease caused by antibiotic-susceptible *H. pylori*, success rates range from 95% to 99%. Success rates may be lower in patients without duodenal ulcer disease. Success rates have been shown to

Table 4 Sensitivity and Specificity of Tests for *Helicobacter pylori* Infection[52,53]

Test	*Sensitivity (%)*		*Specificity (%)*	
	*Before Treatment**	*After Treatment†*	*Before Treatment**	*After Treatment†*
UBT	95.3	88.5	97.7	99.4
HpSA	94.3	92.3	91.8	96.2
ELISA	> 90	Not useful	> 90	Not useful

*N = 501.
†N = 133.
ELISA—enzyme-linked immunosorbent assay HpSA—*H. pylori* stool antigen UBT—13C-urea breath test

be reduced by treatment with two agents rather than three or four; shortening of the course of therapy from 14 days to 10 or 7 days; or lowering of the dose of metronidazole from 500 mg to 250 mg. The combination of clarithromycin and metronidazole may be more successful than the combination of clarithromycin and ampicillin, although rigorous comparisons have not yet been performed.

Cure should be confirmed in all treated patients. The optimal tests for confirmation are the UBT and the HpSA [*see Table 4*] because serum antibody levels will remain elevated for at least 9 months. To ensure that any residual organisms have had the chance to multiply to detectable levels, UBT or HpSA should be delayed until at least 4 weeks after the course of therapy is completed, and proton-pump inhibitors should be stopped 2 weeks before the test. H_2 receptor antagonists can be continued until the day of testing because they have no effect on *H. pylori*.

Resistance to clarithromycin currently occurs in 11% of isolates in the United States and is an increasing problem. Resistance to metronidazole has also been reported; however, the combination of 500 mg of metronidazole with ampicillin or clarithromycin and a proton-pump inhibitor is effective even in the setting of metronidazole resistance.

Although information about the antibiotic susceptibility of the infecting strain is always useful, isolation of *H. pylori* is not trivial, requiring endoscopy and biopsy. Therefore, it is reasonable to treat suspected *H. pylori* infection empirically and obtain antibiotic susceptibility tests only if the initial therapy fails. In general, if a 14-day course of combination therapy fails, the possibility of resistance should be considered. Moreover, if the failed course of therapy included metronidazole and clarithromycin, antibiotic susceptibilities should be determined. Particular antibiotic combinations should not be repeated unless in vitro susceptibility is demonstrated.

COMPLICATIONS

H. pylori infection is probably the cause of the majority of MALT lymphomas. More than 90% of individuals with MALT lymphoma have evidence of *H. pylori* infection, and eradication of the infection during the premalignant phase leads to cure of the lymphomatous process.[47,48]

H. pylori gastritis can progress to chronic atrophic gastritis, which is thought to be a precursor for gastric adenocarcinoma. A prospective study of 1,526 patients in Japan who had duodenal ulcers, gastric ulcers, gastric hyperplasia, or nonulcer dyspepsia showed that gastric cancer developed in patients infected with *H. pylori*, but not in uninfected patients.[49] Treatment of *H. pylori* infection may reduce the risk of adenocarcinoma. However, gastric adenocarcinoma will develop in only a minority of patients with *H. pylori* gastritis, and some gastric adenocarcinomas occur in patients without evidence of prior *H. pylori* infection.

PROGNOSIS

In follow-up to a course of combination therapy, infection is again detected in a small but significant percentage of cases. Presence of *H. pylori* organisms usually represents recurrence rather than reinfection because most isolates are identical to the original isolate. If standard combination therapies fail, antibiotic susceptibility testing of the isolate and salvage therapy should be considered [*see Table 1*].

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Altekruse SF, Stern NJ, Fields PI, et al: *Campylobacter jejuni* —an emerging foodborne pathogen. Emerging Infect Dis 5:28, 1999

2. Tauxe RV: Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations. *Campylobacter,* 2nd ed. Nachamkin I, Blaser MJ, Eds. American Society for Microbiology, Washington, DC, 2000, p 9

3. Sorvillo FJ, Lieb LE, Waterman SH: Incidence of campylobacteriosis among patients with AIDS in Los Angeles County. J Acquir Immune Defic Syndr 4:598, 1991

4. Perlman DM, Ampel NM, Schifman RB, et al: Persistent *Campylobacter jejuni* infections in patients infected with the human immunodeficiency virus (HIV). Ann Intern Med 108:540, 1988

5. Angulo FJ, Swerdlow DL: Bacterial enteric infections in persons infected with human immunodeficiency virus. Clin Infect Dis 21(suppl 1):S84, 1995

6. Black RE, Levine MM, Clements ML, et al: Experimental *Campylobacter jejuni* infection in humans. J Infect Dis 157:472, 1988

7. Rivera-Amill V, Kim BJ, Seshu J, et al: Secretion of the virulence-associated *Campylobacter* invasion antigens from *Campylobacter jejuni* requires a stimulatory signal. J Infect Dis 183:1607, 2001

8. Hendrixson DR, Akerley BJ, DiRita VJ: Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility. Mol Microbiol 40:214, 2001

9. Colegio OR, Griffin TJ 4th, Grindley ND, et al: In vitro transposition system for efficient generation of random mutants of *Campylobacter jejuni*. J Bacteriol 183:2384, 2001

10. Allos BM: *Campylobacter jejuni* infections: update on emerging issues and trends. Clin Infect Dis 32:1201, 2001

11. Salazar-Lindo E, Sack RB, Chea-Woo E, et al: Early treatment with erythromycin of *Campylobacter jejuni*–associated dysentery in children. J Pediatr 109:355, 1986

12. Wong C, Jelacic S, Habeeb R, et al: The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N Engl J Med 342:1930, 2000

13. Smith KE, Besser JM, Hedberg CW, et al: Quinolone-resistant *Campylobacter jejuni* infections in Minnesota, 1992–1998. N Engl J Med 340:1525, 1999

14. Allos BM: Association between *Campylobacter* infection and Guillain-Barre syndrome. J Infect Dis 176:S125, 1997

15. Mead PS, Slutsker L, Dietz V, et al: Food-related illness and death in the United States. Emerging Infect Dis 5:607, 1999

16. Hohmann EL: Nontyphoidal salmonellosis. Clin Infect Dis 32:263, 2001

17. Rabsch W, Tschape H, Baumler AJ: Non-typhoidal salmonellosis: emerging problems. Microbes Infect 3:237, 2001

18. Buchwald D, Blaser M: A review of human salmonellosis: II. Duration of excretion following infection with non-typhi *Salmonella*. Rev Infect Dis 6:345, 1984

19. Neill MA, Opal SM, Heelan J, et al: Failure of ciprofloxacin to eradicate convalescent fecal excretion after acute salmonellosis: experience during an outbreak in health care workers. Ann Intern Med 114:195, 1991

20. Pegues DA, Hohmann EL, Miller SI: *Salmonella* including *S. typhi.* Infections of the Gastrointestinal Tract. Blaser MJ, Smith PD, Ravdin JI, et al, Eds. Raven Press, New York, 1995, p 785

21. Benenson S, Raveh D, Schlesinger Y, et al: The risk of vascular infection in adult patients with nontyphi *Salmonella* bacteremia. Am J Med 110:60, 2001

22. Donabedian H: *Salmonella* aortitis [letter]. Clin Infect Dis 22:739, 1996

23. Butler T, Islam A, Kabir I, et al: Patterns of morbidity and mortality in typhoid fever dependent on age and gender: review of 552 hospitalized patients with diarrhea. Rev Infect Dis 13:85, 1991

24. Goldberg MB, Rubin RH: The spectrum of *Salmonella* infection. Infect Dis Clin North Am 2:571, 1988

25. Rowe B, Ward LR, Threlfall EJ: Multidrug-resistant *Salmonella typhi*: a worldwide epidemic. Clin Infect Dis 24(suppl 1):S106, 1997

26. Punjabi NH, Pulungsih SP, Woodward TE: Treatment of severe typhoid fever in children with high dose dexamethasone. Pediatr Infect Dis J 7:598, 1988

27. Cooles P: Adjuvant steroids and relapse of typhoid fever. J Trop Med Hyg 89:229, 1986

28. Kotloff KL, Winickoff JP, Ivanoff B, et al: Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull WHO 77:651, 1999

29. DuPont HL, Levine MM, Hornick RB, et al: Inoculum size in shigellosis and implications for expected mode of transmission. J Infect Dis 159:1126, 1989

30. Islam MR, Alam AN, Hussain MS, et al: Effect of antimicrobial (nalidixic acid) therapy in shigellosis and predictive values of outcome variables in patients susceptible or resistant to it. Trop Med Hyg 98:212, 1995

31. Haltalin KC, Nelson JD, Ring R 3rd, et al: Double-blind treatment study of shigellosis comparing ampicillin, sulfadiazine, and placebo. J Pediatr 70:970, 1967

32. Khan WA, Dhar U, Salam MA, et al: Central nervous system manifestations of childhood shigellosis: prevalence, risk factors, and outcome. Pediatrics 103:488, 1999

33. Lee LA, Taylor J, Carter GP, et al: *Yersinia enterocolitica* O:3: an emerging cause of pediatric gastroenteritis in the United States. J Infect Dis 163:660, 1991

34. From the Centers for Disease Control and Prevention: Red blood cell transfusions contaminated with *Yersinia enterocolitica*—United States, 1991–1996, and initiation of a national study to detect bacteria-associated transfusion reactions. JAMA 278:196, 1997

35. Adamkiewicz TV, Berkovitch M, Krishnan C, et al: Infection due to *Yersinia enterocolitica* in a series of patients with β-thalassemia: incidence and predisposing factors. Clin Infect Dis 27:1362, 1998

36. Crowe M, Ashford K, Ispahani P: Clinical features and antibiotic treatment of septic arthritis and osteomyelitis due to *Yersinia enterocolitica*. J Med Microbiol 45:302, 1996

37. Reeves PR, Lan R: Cholera in the 1990s. Br Med Bull 54:611, 1998

38. Ryan ET, Calderwood SB: Cholera vaccines. Clin Infect Dis 31:561, 2000

39. Strom MS, Paranjpye RN: Epidemiology and pathogenesis of *Vibrio vulnificus*. Microb Infect 2:177, 2000

40. Chuang YC, Ko WC, Wang ST, et al: Minocycline and cefotaxime in the treatment of experimental murine V*ibrio vulnificus* infection. Antimicrob Agents Chemother 42:1319, 1998

41. Cohen H: Peptic ulcer and *Helicobacter pylori*. Gastroenterol Clin North Am 29:775, 2000

42. Pounder RE, Ng D: The prevalence of *Helicobacter pylori* infection in different countries. Aliment Pharmacol Ther 9(suppl 2):33, 1995

43. Dooley CP, Cohen H, Fitzgibbons PL, et al: Prevalence of *Helicobacter pylori* infection and histologic gastritis in asymptomatic persons. N Engl J Med 321:1562, 1989

44. Staat MA, Kruszon-Moran D, McQuillan GM, et al: A population-based serologic survey of *Helicobacter pylori* infection in children and adolescents in the United States. J Infect Dis 174:1120, 1996

45. Perez-Perez GI, Taylor DN, Bodhidatta L, et al: Seroprevalence of *Helicobacter pylori* infections in Thailand. J Infect Dis 161:1237, 1990

46. Svec A, Kordas P, Pavlis Z, et al: High prevalence of *Helicobacter heilmannii*–associated gastritis in a small, predominantly rural area: further evidence in support of a zoonosis? Scand J Gastroenterol 35:925, 2000

47. Wotherspoon AC, Doglioni C, Diss TC, et al: Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori*. Lancet 342:575, 1993

48. Roggero E, Zucca E, Pinotti G, et al: Eradicaton of *Helicobacter pylori* infection in primary low-grade gastric lymphoma of mucosa-associated lymphoid tissue. Ann Intern Med 122:767, 1995

49. Uemura N, Okamoto S, Yamamoto S, et al: *Helicobacter pylori* infection and the development of gastric cancer. N Engl J Med 345:784, 2001

50. Shiotani A, Nurgalieva ZZ, Yamaoka Y, et al: *Helicobacter pylori*. Med Clin North Am 84:1125, 2000

51. Klugman KP, Koornhof HJ, Robbins JB, et al: Immunogenicity, efficacy, and serological correlate of protection of *Salmonella typhi* Vi capsular polysaccharide vaccine three years after immunization. Vaccine 14:435, 1996

52. Vaira D, Malfertheiner P, Megraud F, et al: Diagnosis of *Helicobacter pylori* infection using a novel, noninvasive antigen based assay in a European multicentre study. Lancet 354:30, 1999

53. Vaira D, Malfertheiner P, Megraud F, et al: Noninvasive antigen-based assay for assessing *Helicobacter pylori* eradication: a European multicenter study. The European *Helicobacter pylori* HpSA Study Group. Am J Gastroenterol 95:925, 2000

X INFECTIONS DUE TO HAEMOPHILUS, MORAXELLA, LEGIONELLA, BORDETELLA, AND PSEUDOMONAS

Shawn J. Skerrett, m.d.

Haemophilus, Moraxella, Legionella, Bordetella, and *Pseudomonas* are gram-negative bacteria that are important respiratory pathogens. All can also cause other types of infections. Vaccines have markedly reduced the incidence of disease from *Haemophilus* and *Bordetella,* but these organisms remain sources of morbidity and potentially life-threatening infection.

Haemophilus influenzae Infections

In 1892, the German bacteriologist Richard Pfeiffer made the sensational claim that he had identified the etiologic agent of influenza.[1] He had repeatedly observed a bacillus in purulent sputum obtained from influenza victims and found that this organism could be cultured on media supplemented with blood. Pfeiffer's blood-loving bacillus was named *Haemophilus influenzae*.[1] Although the causative link with influenza proved false, *H. influenzae* was subsequently recognized as the leading cause of meningitis in young children and an important agent of respiratory infections.

H. influenzae occurs in encapsulated and unencapsulated forms.[2] Encapsulated strains express one of six antigenically distinct capsular polysaccharides, designated *a* through *f*. The type b strain was responsible for most invasive *Haemophilus* infections before the widespread use of conjugate vaccines began in the 1980s. The other capsular groups are rarely implicated as pathogens. Unencapsulated strains do not react with antisera directed at the polysaccharide antigens and thus are nontypeable. These strains are common commensals but can cause mucosal and invasive infections.

epidemiology

H. influenzae is strictly a human parasite: no nonhuman host is known, and *H. influenzae* is not found free in nature. Essentially all persons are colonized by one or more strains of *Haemophilus* by 1 year of age.[2] The primary reservoir for *H. influenzae* is the human upper respiratory tract, although the organism can also be carried on the conjunctivae and in the genital tract. Population studies have found that 40% to 80% of persons harbor unencapsulated strains of *H. influenzae* in the nasopharynx.[2,3] Type b strains were recovered from nasopharyngeal swabs of 2% to 4% of children before the widespread use of the conjugate vaccines, but now such strains are found in less than 1% of children in vaccinated populations. The specific strains colonizing a person's pharynx change over time, and multiple strains can be found concurrently.[2,3] Colonization of the normally sterile tracheobronchial tree with nontypeable strains of *H. influenzae* occurs commonly in people with chronic airway disease, such as cystic fibrosis, bronchiectasis, and chronic obstructive pulmonary disease (COPD), and in asymptomatic smokers.[2,4] In very young infants, nasopharyngeal colonization with *H. influenzae* is associated with an increased risk of otitis media.[2] Persistent colonization of the lower respiratory tract with *H. influenzae* is associated with increased airway inflammation[5] and may contribute to the progression of COPD.[4]

H. influenzae can be transmitted from person to person, leading to secondary cases of infection in households, day care centers, and institutions.[2] The risk of invasive *H. influenzae* infection is age related. The incidence is highest in children younger than 5 years, with a second peak late in life.[6] The distribution and burden of invasive *H. influenzae* infection have been dramatically altered by the use of conjugate vaccines directed against the type b capsular polysaccharide.[7] In unvaccinated populations, the incidence of infection peaks between 6 and 18 months of age, but in vaccinated populations, the risk is highest in the first 6 months of life, before the primary immunization series is completed. In the prevaccine era, there were approximately 20,000 cases a year of invasive *H. influenzae* infection in children younger than 5 years in the United States. Meningitis accounted for 60% of these cases, and 95% of the infections were caused by type b strains.[7] Currently, fewer than 400 cases of invasive *H. influenzae* infection in children younger than 5 years are reported to the Centers for Disease Control and Prevention (CDC) each year, and the minority of these are type b infections.[6] Conjugate vaccine usage has also reduced the overall incidence of invasive *H. influenzae* infection in older children and adults, although an increase in non–serotype b infections has been observed.[8] Most cases of invasive *H. influenzae* in the United States now occur in adults, with pneumonia being the most common type of infection.[6] Certain population groups remain at increased risk, including Native Americans, Alaskan Natives, and African Americans.[7] An increased risk of invasive *H. influenzae* type b infection is also observed in association with complement deficiencies, hypogammaglobulinemia, and splenic dysfunction. The incidence of *H. influenzae* pneumonia is increased in the settings of recent influenza, HIV infection, alcoholism, advanced age, and underlying COPD.

pathogenesis and immunity

All *Haemophilus* organisms are small, pleomorphic, nonmotile, gram-negative coccobacilli. They are facultative anaerobes with fastidious nutritional requirements. Aerobic growth of *H. influenzae* requires X and V factors, both of which are contained in red blood cells. X factor (hematin) is not a single nutrient but, rather, refers to several diffusable pigments that supply protoporphyrins. V factor is nicotinamide adenine dinucleotide, a coenzyme that is released by lysis of red cells. All species of *Haemophilus* grow well on chocolate agar, which contains red cells that have been disrupted by heating.

Unencapsulated strains of *H. influenzae* selectively colonize and locally invade the respiratory tract. *H. influenzae* binds to airway mucins but evades mechanical clearance by inhibiting ciliary beat frequency and damaging ciliated epithelium. The ciliotoxic properties of *H. influenzae* have been attributed to the lipid A, peptidoglycan, and lipoprotein components of the bacterial cell

wall. Most strains of *H. influenzae* also produce a protease that cleaves both serum and secretory IgA, thus inactivating a major component of mucosal host defense.[2,3] *H. influenzae* attaches to respiratory epithelial cells with the aid of several adhesins, including pili, high-molecular-weight hemagglutinins, and other surface proteins.[3] *H. influenzae* preferentially binds to nonciliated cells and to epithelium damaged by recent viral infection or exposure to toxins such as cigarette smoke.[2,3] *H. influenzae* not only can attach to the surface of respiratory epithelium but can enter and survive in epithelial cells and mucosal macrophages and penetrate intercellular junctions to invade the submucosal interstitium.[2,4]

H. influenzae type b is less efficient at colonization than unencapsulated strains, but it has a much greater capacity to invade and survive within the bloodstream.

The magnitude and duration of bacteremia is a critical determinant in the development of *H. influenzae* meningitis, which results mainly from hematogenous seeding rather than direct spread from contiguous infection.[2] The type b capsule facilitates invasion by shielding the organism from complement- and antibody-mediated phagocytosis.[2] The lipo-oligosaccharide in the *H. influenzae* outer membrane is another important factor in the pathophysiology of invasive *H. influenzae* infection: its lipid A component has typical endotoxin activity and contributes to local and systemic inflammatory manifestations of infection, including septic shock.[3]

Antibodies constitute the principal defense against *H. influenzae* type b. These antibodies are directed against the bacteria's capsule, which is composed of repeating units of polyribosyl ribitol phosphate (PRP). Anticapsular antibodies facilitate ingestion of *H. influenzae* by phagocytes and mediate the complement-dependent bactericidal activity of immune serum. Antibodies specific for the PRP component of *H. influenzae* appear by 4 years of age and coincide with the age-related decline in susceptibility to invasive disease.[2] Host resistance to nontypeable strains of *H. influenzae* is less well understood but also involves humoral immunity.[2-4]

CLINICAL SYNDROMES

In the prevaccine era, *H. influenzae* type b was responsible for nearly all invasive *H. influenzae* infections in young children and approximately 50% of invasive infections in adults.[9-11] Meningitis accounted for more than half of the childhood cases of invasive *H. influenzae* type b infection, followed by pneumonia, epiglottitis, primary bacteremia, cellulitis, and musculoskeletal infections.[12] In adults, pneumonia has been the most common manifestation of invasive *H. influenzae* infection, followed by primary bacteremia, epiglottitis, meningitis, obstetric infections, and septic arthritis.[9-11] Widespread use of conjugate vaccines has dramatically reduced the overall incidence of invasive type b infections, but infections caused by nontypeable *H. influenzae* remain important in both children and adults. Unencapsulated strains of *H. influenzae* are principally associated with otitis media, sinusitis, bronchitis, and pneumonia but can also infect other sites.

Meningitis

H. influenzae type b is the leading cause of meningitis in children younger than 5 years in unvaccinated populations but has become rare in immunized communities.[7,12] In adults, *H. influenzae* (predominantly type b) accounts for approximately 4% of cases of community-acquired and nosocomial meningitis.[9,13,14] Adult cases often are associated with contiguous infection of the paranasal sinuses, mastoids, or middle ear or with antecedent head trauma.[13-15] Childhood cases, in contrast, result from bacteremic seeding and often follow an upper respiratory infection.

The clinical manifestations of *H. influenzae* meningitis in both children and adults are typical of acute bacterial meningitis.[9,15] Most patients present with fever, headache, and lethargy. The cerebrospinal fluid is usually cloudy, with a neutrophilic pleocytosis, increased protein concentration, and reduced glucose levels. The diagnosis is made by detecting *H. influenzae* in blood or CSF (see below). The clinical course may be fulminant, leading to death within 24 hours. The overall mortality of *H. influenzae* meningitis in young children is approximately 5%, but 20% to 30% of survivors suffer residual neurologic sequelae. Morbidity is reduced in children treated with dexamethasone.[16] The mortality in adult cases is 10% to 30%.

Epiglottitis

Epiglottitis is a life-threatening infection that is most prevalent in children 2 to 7 years of age but also occurs in older children and adults. *H. influenzae* type b is the most common cause of epiglottitis in all age groups, and the incidence of epiglottitis has declined since the availability of conjugate vaccines.[2,17]

The usual presenting symptoms of epiglottitis are sore throat, odynophagia, dyspnea, and fever; respiratory symptoms are more common in children than adults.[17] The onset of epiglottitis is often abrupt, and the progression can be very rapid. On examination, patients are anxious and may assume an airway-protective posture, sitting erect with the neck extended and the chin protruding forward. Drooling and a muffled voice are useful clues. Stridor, tachypnea, and use of accessory muscles to breathe are common findings in children. The diagnosis can be confirmed by laryngoscopy, with visualization of the cherry-red swollen epiglottis in children or the diffuse supraglottic edema more typical of adult cases. Laryngoscopy must be undertaken with care because the procedure may precipitate acute airway obstruction. A lateral neck x-ray also can be diagnostic, but the supine position should be avoided and patients with suspected epiglottitis should not travel to the radiology suite unless accompanied by a physician prepared to secure an airway. Blood cultures are positive in many cases.

An artificial airway (endotracheal intubation or tracheostomy) should be established emergently in all children with epiglottitis and in adults with evidence of airway compromise. All patients with epiglottitis should be admitted to the hospital for observation and treatment with parenteral antibiotics. Death can result from acute airway obstruction or from septic shock. The mortality is less than 5%.

Pneumonia

H. influenzae is the second or third most common cause of community-acquired pneumonia in adults, identified in 1% to 12% of hospitalized cases.[18] Most of the isolates in these cases are nontypeable strains.[19] In contrast, community-acquired *H. influenzae* pneumonia in children is caused mainly by type b and now is rare in vaccinated populations.[12]

In adults, the risk factors for *H. influenzae* pneumonia include COPD, advanced age, alcoholism, and HIV infection.[9,19,20] *H. influenzae* has also been identified as an important cause of nosocomial pneumonia in the first 5 days of hospitalization, before the indigenous flora of the upper respiratory tract have been replaced.[21]

H. influenzae pneumonia presents as a typical acute pneumonia, with fever, cough productive of purulent sputum, chest pain, and dyspnea.[9,19] Crackles and diminished breath sounds usually

are evident on lung auscultation. The chest x-ray may reveal a bronchopneumonic pattern or focal consolidation, often multilobar. Pleural effusions are common. The diagnosis of *H. influenzae* pneumonia is made most often from the sputum Gram stain and culture, although the organism is frequently missed on sputum smears.[18] Blood cultures are positive in 20% to 30% of cases. Complications of *H. influenzae* pneumonia include empyema, purulent pericarditis, and metastatic infection. The reported mortality of *H. influenzae* pneumonia is 5% to 20%, but it is higher in patients with bacteremia or suppurative complications.

Tracheobronchitis

H. influenzae can cause tracheobronchitis in patients with underlying COPD and, more rarely, in immunocompromised patients.[4,19] The airways of patients with chronic lung disease are often colonized with nontypeable strains of *H. influenzae*, and the acquisition of new strains is associated with acute exacerbations of chronic bronchitis.[4,22] The clinical presentation is characterized by an increase in cough, purulent sputum production, and dyspnea. Fever and leukocytosis may also be present, but the chest film is clear. The sputum Gram stain shows abundant leukocytes, with a predominance of gram-negative coccobacilli.

Otitis Media and Sinusitis

Nontypeable strains of *H. influenzae* are isolated from 15% to 30% of purulent middle-ear effusions in both children and adults.[23] *H. influenzae* is now the leading cause of persistent otitis media in children immunized with the pneumococcal conjugate vaccine.[24]

As with all cases of otitis media, infants typically present with fever and irritability and older children and adults complain primarily of ear pain. The clinical diagnosis is made by otoscopy. A bacteriologic diagnosis requires tympanocentesis, but this procedure is warranted only for evaluation of treatment failures. Nontypeable strains of *H. influenzae* can also be isolated from maxillary aspirates in approximately 25% of children and adults with acute sinusitis.[25] In addition to experiencing common cold symptoms such as rhinorrhea and cough, patients with sinusitis often have unilateral pain, purulent nasal discharge, and reduced transillumination. Confirmatory sinus x-rays and diagnostic aspiration are usually unnecessary.[25] Empirical management includes topical or systemic decongestants, or both, and antibiotics that are effective against *S. pneumoniae, H. influenzae,* and *Moraxella (Branhamella) catarrhalis.*

Extrapulmonary Infections

Cellulitis *H. influenzae* type b is an important cause of cellulitis in unvaccinated children younger than 5 years and is rarely implicated in skin infections of older persons.[9,12,26] The face is the most common site in young children. A raised, warm, tender area of erythema on the cheek or periorbital area is typical, sometimes with a distinctive bluish hue. Underlying sinusitis is common with periorbital and orbital infection. Blood cultures are often positive.

Bone and joint infections In the prevaccine era, *H. influenzae* type b caused approximately 10% of cases of septic arthritis and hematogenous osteomyelitis in infants and young children, but this is now uncommon.[27] Most cases of *H. influenzae* septic arthritis in children are monoarticular, involving large weight-bearing joints. A respiratory source of infection is usually evident. The diagnosis is established by arthrocentesis.

Rare cases of *H. influenzae* pyarthrosis have been described in adults.[9,28] About half of these cases are polyarticular, and most are in patients with alcoholism, immunodeficiency, joint disease, or trauma. Extra-articular sites of infection are apparent in 50% of patients. As in children, most adult cases of *H. influenzae* septic arthritis are caused by type b strains and accompanied by bacteremia.

Conjunctivitis Nontypeable strains of *H. influenzae* have long been appreciated as common causes of purulent conjunctivitis. These infections are contagious, and outbreaks can occur, particularly in day care centers.[2] *H. influenzae* conjunctivitis is easily diagnosed from stains and cultures of conjunctival swabs.

Robert Koch was the first to identify small gram-negative rods in conjunctival exudates collected in Egypt in 1883.[1] This organism was successfully cultured 3 years later by John Weeks, an American ophthalmologist. The Koch-Weeks bacillus, formerly named *Haemophilus aegyptius,* is now known to be a variant of *H. influenzae*. A particularly virulent strain of this organism has been responsible for outbreaks and sporadic cases of Brazilian purpuric fever in young children. This is a fulminant illness in which resolution of an episode of purulent conjunctivitis is followed by high fever, vomiting, abdominal pain, petechiae, purpura, peripheral necrosis, and vascular collapse.[2]

Bacteremia *H. influenzae* b bacteremia without localizing infection can occur in children younger than 5 years and in older persons with defective humoral immunity.[9,12] These patients can deteriorate rapidly, and in rare cases, the bacteremia resembles meningococcemia with purpura fulminans or the Waterhouse-Friderichsen syndrome.[26] Nontypeable strains of *H. influenzae* have also been implicated in neonatal and puerperal sepsis.[9,10,26]

Other infections *H. influenzae* is a rare cause of genitourinary infections, including salpingitis, endometritis, chorioamnionitis, epididymitis, and prostatitis.[10,26] Other rare manifestations of *H. influenzae* infection include appendicitis, chlolecystitis, peritonitis, and endocarditis.[9,26]

MICROBIOLOGIC DIAGNOSIS

The diagnosis of invasive *H. influenzae* infection is generally made from positive cultures of blood, CSF, joint fluid, or other normally sterile sites. CSF and joint fluid Gram stains reveal bacteria in 70% to 80% of cases of meningitis and septic arthritis, respectively. However, the morphology of the organisms is often misleading: plump gram-negative rods, filamentous organisms, gram-negative diplococci, and underdecolorized gram-positive cocci have been described.[9,15,26] Similarly, sputum Gram stains of patients with tracheobronchitis or pneumonia usually show abundant leukocytes and the characteristic gram-negative coccobacilli, but misinterpretations are common.[18] The small gram-negative rods can be overlooked in the pink background on the slide or misinterpreted as gram-positive cocci if the specimen is poorly decolorized.

Type b capsular antigen can be detected by latex agglutination or enzyme-linked immunosorbent assay (ELISA) in CSF, serum, or concentrated urine in up to 90% of invasive infections. Antigen detection is particularly attractive in patients who have already received antibiotic therapy at the time of presentation, but the results rarely influence outcome, because cultures or Gram stains are diagnostic in nearly all antigen-positive cases.[29]

ANTIBIOTIC TREATMENT

The initial antibiotic of choice for meningitis and other invasive *H. influenzae* infections is a third-generation cephalosporin such as ceftriaxone or cefotaxime. These agents are highly active against all isolates of *H. influenzae*, penetrate the CSF, and eradicate nasopharyngeal carriage of *H. influenzae* type b. The pediatric dose of ceftriaxone is 50 mg/kg every 12 hours and that for cefotaxime is 50 mg/kg every 6 hours. For adults, the maximum dose of ceftriaxone is 2 g every 12 hours, and that for cefotaxime is 2 g every 6 hours. The preferred alternative treatment for *H. influenzae* meningitis is chloramphenicol (75 to 100 mg/kg/day in six divided doses). Children with *H. influenzae* meningitis also should receive dexamethasone (0.6 mg/kg/day in four divided doses). Other antibiotics that are highly active against nearly all strains of *H. influenzae* include carbapenems, aztreonam, β-lactam/β-lactamase inhibitor combinations, fluoroquinolones, azithromycin, and tetracycline[30,31] [*see Table 1*]. Ampicillin is highly effective against sensitive strains, but 35% to 40% of North American isolates of *H. influenzae* are resistant to ampicillin. Most ampicillin resistance is mediated by β-lactamase production, but conventional susceptibility testing is recommended for invasive isolates to detect all resistant strains. Intravenous antibiotic therapy is recommended for meningitis and endocarditis and for the initial treatment of other invasive infections. Less severe *H. influenzae* infections can be managed with oral antibiotics. Most *H. influenzae* infections can be cured with 7 to 10 days of therapy, but treatment should continue for 3 to 6 weeks in cases of osteomyelitis or endocarditis.

Table 1 Antibiotics Effective against *Haemophilus influenzae* and *Moraxella catarrhalis*[30,31]

Antibiotic	H. influenzae *Susceptible (%)*	M. catarrhalis *Susceptible (%)*
Ceftriaxone	100	100
Cefpodoxime	100	99
Cefdinir	98	100
Cefuroxime	98–99	99
Ampicillin	68–72	< 10
Amoxicillin-clavulanate	> 99	> 99
Fluoroquinolones	> 99	100
Azithromycin	> 99	100
Clarithromycin	60–90	100
Telithromycin	98	100
Tetracycline	> 98	> 99
Chloramphenicol	> 99	100
Trimethoprim-sulfamethoxazole	79–82	> 97

PREVENTION

Immunization

The introduction of conjugate vaccines in the late 1980s has resulted in a greater than 97% reduction in the incidence of invasive *H. influenzae* type b infections in children younger than 5 years.[7,12] Currently, four conjugate vaccines are licensed in the United States.[32] All are composed of the PRP component of the *H. influenzae* type b capsule in covalent linkage with a carrier protein and elicit protective antibodies directed at PRP. Conjugate vaccines have succeeded the polysaccharide vaccines introduced in the 1970s; those earlier polysaccharide vaccines did not contain a carrier protein and were less effective.[12]

The vaccines are given in a primary series starting at 2 months of age, with the schedule depending on the product. A booster dose is given at 12 to 15 months of age. Immunocompromised children may require additional boosters.[32] Local reactions are experienced by about 25% of persons, but systemic reactions are rare. Adults at high risk for invasive *H. influenzae* infection as a result of sickle cell disease, splenectomy, immunoglobulin or complement deficiency, B cell malignancy, or HIV infection should be considered for vaccination, but specific guidelines have not been developed.

Secondary Prevention

Chemoprophylaxis with rifampin (20 mg/kg/day; maximum, 600 mg a day for 4 days) is more than 95% effective in eliminating nasopharyngeal carriage of *H. influenzae* and is recommended to prevent secondary cases of invasive infection in children at risk. Rifampin prophylaxis should be administered to all adults and children in households with at least one member (other than the index case) who is younger than 4 years and is unimmunized or incompletely immunized.[32] Treatment should also be given to all contacts in households with at least one member younger than 1 year who has completed the primary series but has not yet received a booster dose. Chemoprophylaxis of children and supervisory adults in child care and nursery school settings is recommended if at least two cases of invasive disease have occurred in the attendees within 60 days and if incompletely immunized children are in attendance. Finally, rifampin should be given to victims of invasive disease unless they were treated with ceftriaxone or cefotaxime, which eradicate the carrier state.

Other Haemophilus *Infections*

There are now 15 recognized species of the genus *Haemophilus*. *H. parainfluenzae, H. aphrophilus,* and *H. paraphrophilus* can be found among the flora of the oral cavity and upper respiratory tract. These organisms are primarily commensals, but in rare cases, they cause endocarditis and respiratory tract infection. Endocarditis caused by *Haemophilus* species usually develops on a diseased valve, is typically subacute in onset, and is often complicated by major systemic emboli. Combination antibiotic therapy is generally curative, but valve replacement is frequently necessary. Non-*influenzae* species of *Haemophilus* have antimicrobial susceptibilities that are similar to *H. influenzae,* but they are less likely to be ampicillin resistant.[33]

***Moraxella* (*Branhamella*) *catarrhalis* Infections**

M. catarrhalis was first described by the German microbiologist Seifert in 1882. Originally called *Micrococcus catarrhalis,* the organism was renamed *Neisseria catarrhalis* in 1920, then assigned to the new genus *Branhamella* in 1970, and subsequently reclassified as *Moraxella (Branhamella) catarrhalis*. Although recognized as a cause of tracheobronchitis and pneumonia by early investigators, *M. catarrhalis* was dismissed as a harmless commensal for much of the 20th century, until regaining stature as a potential pathogen over the past 25 years.[34]

EPIDEMIOLOGY

The distribution of *M. catarrhalis* is limited to the human respiratory tract, except for occasional isolates from the conjunctiva and genital tract. Nearly 80% of children harbor *M. catarrhalis* in the nasopharynx by 2 years of age, but colonization declines

with age. Fewer than 6% of healthy adults carry *M. catarrhalis*, although the organism can be isolated from the sputum of 5% to 32% of patients with chronic bronchitis or bronchiectasis.[4,35,36] Colonization is a dynamic process, with individual strains persisting for only a few months. Carriage of *M. catarrhalis* also exhibits seasonal variation, with colonization and infection being more prevalent in the winter months.[36]

M. catarrhalis is the third most common cause of otitis media, sinusitis, and acute exacerbations of chronic bronchitis, after *H. influenzae* and *S. pneumoniae*.[25,35] The majority of lower respiratory tract infections attributed to *M. catarrhalis* in adults occur in patients with underlying COPD.[35]

PATHOGENESIS

M. catarrhalis is a nonmotile, gram-negative diplococcus that is morphologically indistinguishable from *Neisseria* species. *M. catarrhalis* grows well aerobically on blood agar and chocolate agar. The characteristic convex, opaque colonies can be pushed across the agar surface in the manner of a hockey puck. Other species of *Moraxella* are predominantly gram-negative coccobacilli that can be found among the normal flora of humans and other mammals, but they rarely cause disease.

The virulence of *M. catarrhalis* is related to its capacity to colonize the respiratory mucosa. *M. catarrhalis* expresses a number of potential adhesins, but the mechanisms underlying the respiratory tropism of this organism are poorly understood.[4,35] Similarly, the factors that influence the transition from asymptomatic colonization to infection have not been deciphered. In the case of otitis media and sinusitis, it is likely that nasopharyngeal bacteria will gain access to adjacent spaces after mucosal injury. The expression of particular virulence factors also may be important in pathogenesis. For example, strains of *M. catarrhalis* isolated from symptomatic persons are more likely to exhibit resistance to the bactericidal activity of serum complement than strains harbored by asymptomatic carriers. Recovery from *M. catarrhalis* infection is accompanied by strain-specific antibody responses,[35,36] but the key elements of host defense against this organism are unknown.

CLINICAL SYNDROMES

Tracheobronchitis

Acute exacerbations of chronic bronchitis are the most common manifestations of *M. catarrhalis* infection in adults.[4,33,34] As with exacerbations caused by other organisms, dyspnea and cough productive of purulent sputum are the cardinal symptoms; fever also may be present. *M. catarrhalis* can be implicated by the predominance of gram-negative diplococci on Gram stains of purulent sputum, with isolation of *M. catarrhalis* in culture. In approximately 30% of cases involving *M. catarrhalis,* additional pathogens are present in the sputum.

Pneumonia

Most cases of pneumonia caused by *M. catarrhalis* occur in the elderly, particularly those with underlying COPD.[35] The clinical features are those of a mild, acute pneumonia: fever, cough productive of purulent sputum, dyspnea, and chest pain are the common symptoms. Focal evidence of consolidation may be evident on physical examination. The chest x-ray usually reveals patchy alveolar or interstitial opacities that are often multilobar. Pleural effusions are uncommon. The diagnosis is usually made from the sputum Gram stain and culture. Bacteremia, empyema, and other suppurative complications are rare.

Otitis Media and Sinusitis

M. catarrhalis can be isolated from middle-ear effusions in 15% to 20% of cases of acute otitis media in children, usually in pure culture.[35] Polymerase chain reaction detects *M. catarrhalis* DNA in an even higher proportion of middle-ear exudates. Similarly, maxillary aspirates yield *M. catarrhalis* in about 20% of children with acute sinusitis.[25,35] However, less than 10% of episodes of acute sinusitis in adults can be attributed to *M. catarrhalis*.[25]

Other Infection

Nosocomial outbreaks of *M. catarrhalis* infection, predominantly involving the respiratory tract and occurring mainly in patients with underlying cardiopulmonary disease, have been reported in children and adults.[35] The mode of transmission has not been clear. *M. catarrhalis* has also been reported as a rare cause of invasive infections, including meningitis, endocarditis, septic arthritis, and bacteremia.[35] Underlying lung disease and neutropenia have been identified as risk factors for *M. catarrhalis* bacteremia, but invasive infections originating in the upper respiratory tract have occurred in previously healthy individuals. *M. catarrhalis* can also cause conjunctivitis and rare urogenital infections.[35]

ANTIBIOTIC TREATMENT

More than 90% of clinical isolates of *M. catarrhalis* produce β-lactamases that render ampicillin and amoxicillin ineffective.[30,31] Penicillin–β-lactamase–inhibitor combinations, second- and third-generation cephalosporins, fluoroquinolones, macrolides, and tetracyclines are active against nearly all strains of *M. catarrhalis* [*see Table 1*].

Legionellosis

In the summer of 1976, a mysterious outbreak of pneumonia occurred among participants in the American Legion's bicentennial convention in Philadelphia, killing 34 of the 221 people afflicted.[37] The disorder became known as Legionnaires disease. Its cause remained obscure until Joseph McDade and his colleagues at the CDC isolated a novel bacterium from the spleens of guinea pigs that had been injected with lung tissue harvested from victims of the epidemic.[38] The new agent was named *Legionella pneumophila*. Studies with stored sera identified this organism as the cause of Pontiac fever—so named after an outbreak of a nonpneumonic, influenzalike illness that occurred in Pontiac, Michigan, in 1968—as well as other cases of epidemic and sporadic respiratory disease dating back to 1943.[39] *L. pneumophila* was the first of several previously unknown agents to emerge as important human pathogens in the late 20th century. In the years since the Philadelphia outbreak of legionellosis, many other *Legionella* species have been identified and implicated in a growing spectrum of human disease.

The family Legionellaceae now includes more than 40 species of the genus *Legionella,* with over 60 distinct serotypes.[40] Almost all of these organisms have been isolated from the environment, and approximately half of them have been associated with human disease. *L. pneumophila* remains the most important human pathogen in this family, accounting for more than 80% of legionellosis cases diagnosed in the United States.[40,41] Other species of *Legionella* produce a similar spectrum of disease and may be underappreciated as causes of human infection.[40]

EPIDEMIOLOGY

The Legionellaceae are widely distributed in freshwater habitats throughout the world.[39,40] They thrive in warm water en-

riched with iron, often in biofilms with other organisms that supply essential nutrients, such as blue-green algae or macrophytes. Legionellae can also replicate in free-living protozoa, and amebae may be their natural hosts. Man-made aquatic environments that often harbor legionellae include hot-water systems, cooling towers, air conditioners, industrial coolants, whirlpool baths, and humidifiers. The factors that promote the growth of legionellae in water tanks and potable water supplies include a temperature between 30° and 60° C, stagnation, decayed or rusted pipes and fixtures, and infrequent or ineffective decontamination. Water systems harboring legionellae can be disinfected by superheating (> 70° C) and flushing, hyperchlorination, or metallic ionization.[40,41]

Legionellosis is acquired mainly by the inhalation of aerosolized bacteria from an environmental source,[39-41] typically a man-made device. Examples of proven disseminators include showerheads, water faucets, cooling towers, evaporative condensers, respiratory therapy equipment, produce misters, whirlpool baths, and decorative fountains. Only a few cases of legionellosis have been linked to natural aquatic sources such as hot springs. Potting soil also has been identified as a source of infection. Transient contamination of the oropharynx followed by microaspiration may occur occasionally, particularly in patients with nasogastric tubes; stable colonization of the upper respiratory tract by legionellae has not been described. Wounds can be infected by direct inoculation with contaminated water, and oral ingestion has been postulated as a potential source of rare intestinal infections. There is no evidence for person-to-person transmission.

The epidemiologic patterns of legionellosis include common-source epidemics, sporadic cases, and endemic disease in institutions with contaminated water supplies. The infection is more common in the summer months, when seasonal conditions promote the growth of legionellae in the environment, and the use of air conditioners and other cooling devices facilitates the dissemination of airborne bacteria.

Legionella species have been reported to cause from 1% to 15% of community-acquired pneumonias leading to hospitalization,[18] but they consistently rank among the leading causes of severe community-acquired pneumonia leading to respiratory failure.[42] In hospitals with contaminated water supplies, legionellae are important agents of nosocomial pneumonias.

The risk factors for *Legionella* pneumonia include cigarette smoking; alcoholism; recent travel; chronic lung, heart, or kidney disease; diabetes mellitus; cancer; and the use of immunosuppressive medications (especially corticosteroids).[41] Transplant recipients are at particularly high risk and have been the sentinel victims in several nosocomial outbreaks. *Legionella* infections are uncommon but unusually severe in HIV-infected patients.

The epidemiologic features of pneumonic and nonpneumonic legionellosis differ in several respects.[39,43] During epidemics, the attack rate of *Legionella* pneumonia varies from 0.2% to 7%, whereas nonpneumonic legionellosis afflicts over 65% of the people exposed. The incubation period of Legionnaires disease varies from 2 to 12 days, but Pontiac fever usually strikes within 48 hours of exposure. One or more risk factors can be identified in most victims of *Legionella* pneumonia. In contrast, nonpneumonic legionellosis readily affects the normal host. Common sources have resulted in cases of both Legionnaires disease and Pontiac fever, suggesting that bacterial inoculum and host factors are important in determining the clinical course of infection. It is possible that Pontiac fever results primarily from the inhalation of bacterial endotoxin.[43]

PATHOGENESIS AND PATHOLOGY

The Legionellaceae are small, pleomorphic, flagellated gram-negative bacilli that often appear coccobacillary in tissue.[39,40] They are enveloped by a laminated outer membrane with a unique lipid structure that differs from the lipopolysaccharide of other gram-negative bacteria. Legionellae are facultative intracellular parasites: they can replicate independently or in other cells. They are strictly aerobic but grow best in a reduced-oxygen environment. Buffered charcoal–yeast extract agar supplemented with L-cysteine, ferric iron, and α-ketoglutarate is the most suitable artificial medium.

After inhalation into the lungs, *L. pneumophila* enters alveolar macrophages via coiling or conventional phagocytosis and replicates in a specialized vacuole.[40,44] The formation of this protected intracellular compartment is directed by the *dot* (defective organelle trafficking) and *icm* (intracellular multiplication) genes of *L. pneumophila*. These genes encode for a type IV secretion system that modifies the endocytic pathway of the host cell and delays fusion of the phagosome with lysosomes. As nutrients become scarce, the production of cytotoxins by the stressed bacteria leads to rupture of the macrophage, and the cycle begins anew. The secretion of proteases by *L. pneumophila* contributes to tissue injury.[44] Spreading infection leads to the recruitment of blood-borne phagocytes, but *L. pneumophila* is relatively resistant to killing by neutrophils and multiplies within monocytes. Humoral immunity appears to play little role in host defense against this infection, because *L. pneumophila* is resistant to the lytic effects of antibody and complement, and opsonization by specific antibody promotes uptake of the organism by phagocytic cells without stimulating bacterial killing. Ultimately, the infection is probably contained by suppression of intracellular replication by interferon-gamma–mediated activation of macrophages and the destruction of infected cells by cytotoxic lymphocytes.[45]

Impairments in specific host defenses underlie the increased risk of legionellosis in certain groups. For example, the alveolar macrophages of smokers are greater in number and more susceptible to parasitism by *L. pneumophila* than the macrophages of nonsmokers.[46] Smokers' macrophages exhibit greater expression of the CD11b/CD18 integrin complex that mediates uptake of *L. pneumophila,* and they also contain more iron, an essential bacterial nutrient, than the alveolar macrophages of nonsmokers. Toll-like receptors (TLRs) stimulate the immune response to *Legionella,* and single-nucleotide polymorphisms that affect the functions of TLR4 and TLR5 have been associated with susceptibility to Legionnaires disease.[47,48] Conditions associated with impaired cellular immunity, such as T cell malignancies, HIV infection, and the use of immunosuppressive medications, result in deficient macrophage activation and failure to contain intracellular infection.

Pathologically, pneumonic legionellosis is characterized by confluent multilobular or lobar consolidation, with no predilection for particular segments.[49] Involvement is bilateral in approximately half the cases. Small pleural effusions are common, but empyema is unusual. The most common histologic pattern is an acute fibrinopurulent alveolitis, with little or no involvement of the conducting airways. There is a dense intra-alveolar infiltrate of macrophages and polymorphonuclear leukocytes, accompanied by an abundant fibrinous exudate and proteinaceous debris. The alveolar septae are congested, and there may be a minor vasculitic component. The bacteria are found in the alveolar exudate and are located mainly within macrophages.

A second histologic pattern of Legionnaires disease is diffuse alveolar damage. This is characterized by an intra-alveolar, fi-

brinoserous exudate with hyaline membrane formation and evidence of epithelial injury. Few organisms can be identified in this setting, but bacterial invasion of the interstitium may be observed. These features may coexist with the fibrinopurulent pattern and are predominantly found in immunocompromised patients. Fibrosis may occur with resolution of the infection. Bronchiolitis obliterans organizing pneumonia is an unusual complication.

CLINICAL SYNDROMES

Two distinct clinical syndromes of legionellosis have been recognized: the pneumonic form known as Legionnaires disease, and the nonpneumonic, self-limited syndrome called Pontiac fever. The spectrum of infection also includes asymptomatic seroconversion and rare focal infections of extrapulmonary organs.

Legionnaires Disease

The classic picture of Legionnaires disease is that of a chronically ill, middle-aged smoker with rapidly progressive pneumonia associated with nonpurulent sputum, diarrhea, confusion, hyponatremia, and abnormal liver function tests. The clinical features of Legionnaires disease are protean, however, so the disorder cannot reliably be distinguished from other pneumonias on clinical grounds alone[39,50,51] [*see Table 2*].

Legionnaires disease typically begins with a brief prodrome of profound weakness, malaise, myalgias, and headache. High fever and shaking chills accompany the development of a cough that is productive in approximately half the cases. Sputum is generally nonpurulent. Chest pain is common, and progressive dyspnea is the rule. Watery diarrhea, nausea, and vomiting are often reported.

Physical examination usually reveals a distressed, tachypneic patient. High fever is a hallmark of legionellosis, and temperatures exceeding 39° C (102.2° F) are found in most patients, even those who are immunosuppressed.[50] Relative bradycardia (a heart rate less than 100 beats a minute despite a temperature of 40° C [104° F] or higher) may be present. Evidence of consolidation is typically evident on lung auscultation. An encephalopathy ranging from mild confusion to obtundation is common, and focal neurologic deficits have been described.

Routine laboratory tests also typically show an array of abnormalities[39,50,51] [*see Table 2*]. The complete blood count typically demonstrates leukocytosis with a left shift, although a white blood cell count above 15,000/mm^3 is unusual. Hyponatremia and abnormal liver enzymes are reported in approximately half the cases. Hypophosphatemia, azotemia, an elevated creatine kinase level, and an increased serum amylase level may also be found. Urinalysis often reveals proteinuria, hematuria, or both.

The chest x-ray in *Legionella* pneumonia may demonstrate poorly marginated, rounded nodules; patchy alveolar infiltrates; or consolidation in a segmental or lobar distribution.[50,52] The infiltrates are unilobar at presentation in most cases, but radiologic progression to multilobar disease is common. Small pleural effusions are frequently evident, but empyema is unusual. Cavitation is an uncommon finding that is most often seen in immunocompromised patients.

Extrapulmonary manifestations of Legionnaires disease can be striking and may reflect the effects of bacterial toxins, host-derived mediators, or metastatic infection.[41] Neurologic deficits are usually metabolic in origin, but focal bacterial encephalitis has been described. The pathogenesis of the watery diarrhea that often accompanies *Legionella* pneumonia is obscure, but intestinal abscesses and peritonitis have been reported. Acute renal failure may complicate legionellosis, as a result of acute tubular necrosis, rhabdomyolysis, or glomerulonephritis. Pyelonephritis is rare. Overt dissemination of *Legionella* infection from the lungs to other organs occurs most commonly in immunocompromised persons.

Table 2 Manifestations of Legionnaires Disease[80]

Manifestations	*Percentage of Cases*
Symptoms and signs	
Fever	67–100
Chills	42–77
Cough	65–92
Sputum production	41–75
Hemoptysis	0–41
Chest pain	14–52
Dyspnea	14–59
Headache	17–56
Myalgias	20–43
Altered mental status	21–43
Diarrhea	21–47
Nausea, vomiting	9–37
Abdominal pain	5–20
Laboratory abnormalities	
Leukocytosis	53–78
Hyponatremia	17–55
Elevated liver enzymes	14–88
Hypophosphatemia	25–51
Azotemia	15–60
Proteinuria	25–56
Hematuria	0–50

Unusual sites of *Legionella* infection include the skin (cellulitis and wound infection), the heart (myocarditis, pericarditis, and prosthetic valve endocarditis), and the abdomen (pancreatitis, colitis, and appendicitis).[41] Infections at these sites have been reported in the absence of pneumonia. In some instances, the infection appears to have resulted from direct inoculation of tissue or a prosthetic device with contaminated water. Other cases may result from bacteremic seeding from the lungs.

Pontiac Fever

The most common nonpneumonic form of legionellosis is Pontiac fever.[39] This syndrome is characterized by a prodrome of malaise, diffuse myalgias, and headache, followed hours later by fever and chills. A mild, nonproductive cough and sore throat may develop, but respiratory tract symptoms are not prominent. Nausea, vomiting, and diarrhea are common, as are neurologic disturbances such as dizziness, neck pain or stiffness, confusion, irritability, nightmares, and ataxia. The physical examination is unremarkable except for the presence of fever, and laboratory tests reveal only leukocytosis. The chest film is clear. The diagnosis is primarily made serologically and supported by the isolation of *Legionella* from an environmental source. Evidence of infection can be detected by PCR or, in some cases, urinary antigen testing. The acute illness resolves spontaneously in 2 to 5 days, but full recovery may take weeks.

MICROBIOLOGIC DIAGNOSIS

The diagnosis of legionellosis requires the isolation of the organism in culture, detection of microbial antigens or nucleic

Table 3 Diagnostic Tests for Legionellosis[18,53]

Test	*Specimen*	*Sensitivity (%)*	*Specificity (%)*	*Comment*
Culture	Sputum, BALF	11–80	100	All species; requires selective media
Direct fluorescent antibody	Sputum, BALF	22–75	> 90	*Legionella pneumophila* only
Urinary antigen	Urine	48–90	> 95	*L. pneumophila* serogroup 1 only
Serology				
Fourfold change*	Blood	60–80	>90	*L. pneumophila* only
Single titer		10–27	> 85	
PCR	Sputum, BALF, serum, urine	10–100	> 95	All species; not widely available

*Between acute and convalescent samples.
BALF—bronchoalveolar lavage fluid PCR—polymerase chain reaction

acids in body fluids, or demonstration of serologic evidence of infection [*see Table 3*]. However, an early clue to the diagnosis of Legionnaires disease can be obtained from nonspecific stains of sputum, bronchoalveolar lavage fluid, or lung tissue.[18] Although legionellae are poorly seen on routine Gram stain, visualization of these small, pleomorphic gram-negative bacilli is improved if basic fuchsin is used as the counterstain in place of safranin O. The intracellular organisms can also be seen on Gimenez, Dieterle, or Wright-Giemsa stains. *L. micdadei* is unique among the Legionellaceae in appearing acid fast in clinical specimens.

Culture of sputum, bronchoalveolar lavage fluid, or lung tissue is the preferred means of confirming the diagnosis.[18,41] Specialized media are required for the isolation of *Legionella* (see above), so cultures for these pathogens must be specifically requested of the laboratory. All *Legionella* species can be isolated after 3 to 5 days of incubation, and the recovery of organisms from any body fluid or tissue is diagnostic of infection. Sputum submitted from patients with suspected legionellosis should not be discarded even if epithelial cell contamination occurs, because poor-quality specimens may still yield the organism. Experienced microbiology laboratories can recover legionellae from pretreatment sputum in most cases of legionellosis, but not all laboratories are skilled in this regard, and many patients do not produce sputum.

Antigen detection provides an opportunity for more rapid diagnosis.[18,53] The direct fluorescent antibody (DFA) test can identify organisms in sputum, bronchoalveolar lavage fluid, or lung tissue, even after a few days of antibiotic therapy. However, the interpretation of DFA preparations requires considerable expertise, and the sensitivity of this test is limited to specific antigens. The most commonly used monoclonal antibody preparation detects only *L. pneumophila* serogroups 1 through 12, and commercially available pooled antisera are limited to a few species. Urinary antigen detection is currently the most helpful test for rapid diagnosis of legionellosis. Antigens appear in the urine within 3 days of the onset of illness and may persist for months, although most patients stop excreting antigen within 6 weeks. The sensitivity of this test is higher in cases of severe pneumonia than in cases of milder disease.[54] The major limitation of urinary antigen testing is that currently available tests reliably detect only *L. pneumophila* serogroup 1.

DNA amplification techniques are very promising for the rapid diagnosis of legionellosis.[18,53] Preliminary experience suggests that PCR can detect any species of *Legionella* in sputum, throat swabs, and bronchoalveolar lavage fluid with high sensitivity and specificity. *Legionella* DNA has also been detected in the urine and blood of patients with legionellosis. These tests hold great promise but require additional evaluation and refinement to establish their role in clinical practice.

Serologic evidence of infection can be obtained by indirect immunofluorescence, microagglutination, or ELISA.[18] A fourfold increase in titer between acute and convalescent samples is very specific for infection with *L. pneumophila* serogroup 1, but tests for other serogroups and species are more subject to cross-reactivity and have not been standardized. Approximately half of patients with Legionnaires disease will seroconvert within 4 weeks from the onset of illness, and 75% will do so within 9 weeks. A small number of patients take as long as 14 weeks to demonstrate an antibody response, and no serologic response will be detected in up to 25% of cases. The major limitation of seroconversion is the delay in diagnostic information. A single titer of 1:256 or higher

Table 4 Treatment of Legionellosis

Antibiotic Class	*Drug*	*Dosage*	*Comment*
Fluoroquinolones	Levofloxacin	500 mg p.o. or I.V. q. 24 hr for 10 days	May be preferred in immunocompromised patients because the agents are bactericidal
	Gatifloxacin	400 mg p.o. or I.V. q. 24 hr for 10 days	
	Moxifloxacin	400 mg p.o. q. 24 hr for 10 days	
	Ofloxacin	400 mg p.o. or I.V. q. 24 hr for 10 days	
Macrolides	Azithromycin	500 mg p.o. or I.V. q. 24 hr for 5–10 days	Most active of its class against *Legionella*
	Clarithromycin	500 mg p.o. or I.V. q. 24 hr for 10 days	—
	Erythromycin	500 mg p.o. or 1 g I.V. q. 6 hr for 10–14 days	Less well tolerated than other macrolides
Tetracyclines	Doxycycline	100 mg p.o. or I.V. q. 12 hr for 10–14 days	May be less effective than fluoroquinolones or macrolides

suggests acute infection with *L. pneumophila* but is found in less than 30% of patients during the acute phase of illness.

TREATMENT AND PROGNOSIS

There are no controlled trials to guide the treatment of legionellosis, but retrospective observations, anecdotal experience, animal studies, and in vitro testing support the efficacy of fluoroquinolones, macrolides, tetracyclines, and rifampin in the treatment of Legionnaires disease [*see Table 4*].[41,42,55] Fluoroquinolones may be preferred in immunocompromised patients because they are bactericidal (in contrast to macrolides and tetracyclines), exhibit the most activity in experimental models, and may induce a more rapid clinical response than macrolides.[55,56] All of the fluoroquinolones appear to be effective, but ciprofloxacin is less active than other agents in this class. Azithromycin is the most active macrolide against *Legionella*. Both azithromycin and clarithromycin are more effective and better tolerated than erythromycin. Rifampin is highly active against *Legionella,* and anecdotal evidence supports its use in combination with erythromycin in severely ill or immunocompromised patients. Of the tetracyclines, doxycycline and minocycline are the most active agents, but these drugs appear to be less effective than fluoroquinolones or macrolides. Trimethoprim-sulfamethoxazole also may be effective, but clinical experience is limited. All of these agents share the ability to penetrate the intracellular environment of macrophages, where legionellae replicate. β-Lactam antibiotics and aminoglycosides generally are ineffective in legionellosis.

The optimum duration of therapy is unknown. Immunocompromised persons and patients with unusually severe disease may warrant treatment for 2 to 3 weeks, but experience with newer macrolides and fluoroquinolones suggests that shorter courses are adequate.[41,55]

The prognosis of *Legionella* pneumonia is influenced by the status of host defenses and the timely initiation of effective antibiotic treatment.[41,42] With appropriate treatment, the mortality of Legionnaires disease varies from approximately 5% in healthy persons to 25% in immunocompromised patients.

Pertussis (Whooping Cough)

Whooping cough was first described in 16th-century France.[57] The name pertussis was coined by the English physician Thomas Sydenham in 1679, in reference to the intensive cough that is the hallmark of this disease. *Bordetella pertussis,* the causative agent, was first isolated by Bordet and Gengou in 1906. Other members of the genus that are associated with human disease include *B. parapertussis* and *B. bronchiseptica.*[58] *B. parapertussis* lacks pertussis toxin but causes a mild form of whooping cough. *B. bronchiseptica* is a zoonotic pathogen that can cause human respiratory illness; in immunocompromised patients, manifestations of disease may be severe.[58]

EPIDEMIOLOGY

B. pertussis is exclusively a human pathogen: no carrier state, animal host, or natural reservoir is known. The infection is transmitted from person to person by aerosol droplets and is very contagious, spreading quickly among close contacts. Pertussis is endemic year-round, with the peak incidence in summer and fall.[58] Epidemics occur at intervals of 2 to 5 years.[58] The attack rate is highest in infants younger than 1 year and is higher in females than males.[58] In the prevaccine era, pertussis was the leading cause of death in children younger than 14 years in the United States. The introduction of the killed whole cell vaccine in the 1940s led to a marked reduction in the prevalence of pertussis, which fell steadily until reaching its nadir in 1981. Since 1982, there has been a resurgence of reported cases of pertussis in the United States. During this time, the incidence in children younger than 5 years has remained relatively constant, but an increasing proportion of cases has been recognized in adolescents and young adults.[58] More than 60% of the 11,647 cases of pertussis reported to the CDC in 2003 occurred in persons older than 10 years.[6] Most early childhood cases of pertussis now occur in children who either are too young to have received vaccine or have been inadequately vaccinated.[6,58,59] Cases in adolescents and adults probably reflect the waning of immunity 5 to 10 years after vaccination. It is likely that pertussis is markedly underreported in older children and adults, because the characteristic whooping cough is usually absent in these age groups. Mildly ill or asymptomatic adults with unrecognized pertussis serve as important reservoirs for the infection of more susceptible children.

Although effective vaccines have dramatically reduced the importance of pertussis in developed countries, the disease remains a leading cause of morbidity and mortality in unvaccinated populations. An estimated 50 million cases of pertussis occur worldwide each year, leading to 300,000 deaths.[60,61]

Table 5 Virulence Factors of *Bordetella pertussis*[60]

Virulence Factor	*Effects*
Filamentous hemagglutinin	Attachment to ciliated epithelium; uptake by leukocytes via complement receptor 3
Agglutinogens, fimbriae	Attachment to epithelium
Lipopolysaccharide	Endotoxin activity
Tracheal cytotoxin	Ciliostasis, epithelial injury
Adenylate cyclase toxin	Induces increased cyclic adenosine monophosphate; impairs leukocyte function
Dermonecrotic toxin	Vascular smooth muscle contraction, ischemic necrosis
Pertussis toxin	Inhibits G protein–coupled signaling; lymphocytosis, hyperinsulinemia, encephalopathy
Type III secretion system	Delivers proteins into cytoplasm of target cells

PATHOGENESIS

Bordetellae are small, aerobic, gram-negative coccobacilli. They have fastidious growth requirements (see below). Pertussis is primarily a disease of the conducting airways, characterized histologically by generalized inflammation of the bronchi and bronchioles with a mucopurulent luminal exudate.[57] The bacilli grow on the surface of ciliated epithelium, where they thrive through the expression of surface components and secreted toxins that suppress mechanical and cellular host defenses [*see Table 5*].

Pertussis toxin is the most intriguing factor produced by *B. pertussis.* This heterodimeric protein catalyzes adenosine diphosphate (ADP)–ribosylation of guanine nucleotide–binding (G) proteins, thereby inhibiting signal transduction through G-coupled transmembrane receptors.[58,60] How this action is linked to

the pathophysiology of pertussis is not well understood. Pertussis toxin causes lymphocytosis, hyperinsulinemia, and possibly encephalopathy, but its role in the respiratory manifestations of whooping cough remains enigmatic.[58,60] Antibodies directed against pertussis toxin are protective against disease, yet *B. parapertussis,* which does not produce the toxin, can cause a similar (albeit milder) respiratory illness.

Relatively little is known of the essential elements of host defense against pertussis, but cellular and humoral immunity are both involved.[58,60] Protective maternal antibodies are not transferred across the placenta, contributing to the susceptibility of infants. The resolution of whooping cough is accompanied by the disappearance of *B. pertussis* from the airway epithelium.

CLINICAL MANIFESTATIONS

Classic pertussis is a three-stage illness lasting 4 to 8 weeks.[57,58,60] After an incubation period of 6 to 20 days, the clinical manifestations begin with the catarrhal stage. This phase lasts 1 to 2 weeks and resembles the common cold, with rhinorrhea, conjunctivitis, lacrimation, low-grade fever, and mild nonproductive cough. The distinctive paroxysmal stage follows for 2 to 4 weeks and is characterized by violent spasms of coughing. In a typical episode, a single expiration is punctuated by five to 20 explosive coughs in rapid succession, followed by a gasping, forceful inhalation through a narrowed glottis that produces the characteristic whoop. Paroxysms often are accompanied by cyanosis and followed by emesis. Infants have less pronounced whoops but may become apneic during coughing fits. The violent coughing can result in epistaxis, conjunctival hemorrhage, petechiae, rib fractures, pneumothorax, incontinence, hernias, and rectal prolapse. Paroxysms occur 10 to 30 times a day, often begin without warning, and can be triggered by activities such as eating, drinking, sneezing, yawning, or exposure to airborne irritants. The paroxysms are exhausting, and patients avoid oral intake, leading to dehydration and weight loss.

Physical examination during the spasmodic stage often is unremarkable. Eyelid edema may be present, and auscultation of the chest may reveal coarse rales and rhonchi. Fever is unusual in the absence of complications. The blood count at this time is characterized by marked lymphocytosis: the total WBC count often exceeds 30,000/mm^3, of which more than 60% are lymphocytes. Hypoglycemia is occasionally observed. Chest radiographs are usually normal but may demonstrate hyperinflation, atelectasis, or perihilar infiltrates.[57,62] Complications that can develop during the paroxysmal phase include pneumonia, otitis media, seizures, and encephalopathy. Pneumonia, caused either by *B. pertussis* or by a secondary invader, develops in 10% to 15% of young children and is responsible for most deaths.[59] The convalescent third stage of pertussis, during which the frequency and severity of the cough gradually diminish, usually lasts 1 to 3 weeks but may extend for months.

Adults and adolescents with partial immunity have a less severe illness and typically present with a persistent cough of several weeks' duration. The cough is paroxysmal and often productive of purulent or nonpurulent sputum, but whoops are unusual. As in children, violent coughing in adults with pertussis may result in posttussive vomiting, syncope, urinary incontinence, and rib fractures. About half of adult patients report a preceding catarrhal illness. Lymphocytosis is usually absent. Numerous recent studies suggest that 10% to 32% of adolescents and adults with cough lasting longer than 1 to 2 weeks have serologic evidence of pertussis.[58,63] Annual serologic screening programs have found that undiagnosed pertussis is common in adolescents and adults.

A presumptive diagnosis of pertussis can be made in susceptible children with the characteristic whooping cough and lymphocytosis. However, the typical whoop is not uniformly present, and similar illnesses can be caused by *B. parapertussis*, respiratory viruses (particularly adenovirus, parainfluenza, and respiratory syncytial virus), *Mycoplasma pneumoniae,* and *Chlamydia pneumoniae*.[58]

MICROBIOLOGIC DIAGNOSIS

Microbiologic confirmation of pertussis is best made during the catarrhal or early paroxysmal stages, when *B. pertussis* can be detected in the nasopharynx by culture, immunofluorescence, or PCR.[58,60,64] Cultures of nasopharyngeal aspirates have higher yields than those of nasopharyngeal swabs; if the latter are used, calcium alginate and Dacron are the preferred materials. Throat specimens are unsatisfactory because the oropharynx lacks the ciliated epithelium to which *B. pertussis* adheres. The nasopharyngeal sample should be directly streaked on suitably enriched media, such as charcoal-blood agar or Bordet-Gengou agar. If a delay in plating is anticipated, a charcoal-containing transport medium (e.g., Regan-Lowe) should be used.[64] Typical colonies of tiny gram-negative bacilli appear after 3 to 7 days of culture, and identification can be confirmed by agglutination or immunofluorescence. Culture yield is reduced by antibiotic therapy or prior immunization and falls with the duration of illness. Cultures are usually negative during the paroxysmal phase and in adults with persistent cough.

B. pertussis can also be detected in respiratory secretions by DFA staining in 30% to 65% of cases during the catarrhal stage.[58,60,64] Polyclonal antibody preparations have been plagued by variable sensitivity and poor specificity, but a monoclonal reagent has been found to be more satisfactory.[64] Nonetheless, immunofluorescence is mainly used as a screening test for outbreak investigation.

PCR testing of nasopharyngeal specimens has been found to be more sensitive than culture or immunofluorescence for the diagnosis of pertussis, with specificity greater than 85%.[58,60,64] This very promising technique has become increasingly available, but the methods and reagents for the diagnosis of pertussis by PCR have not yet been standardized.

Detection of serum antibody responses to *B. pertussis* by ELISA remains an important if imperfect diagnostic tool. Serologic tests are most useful for the diagnosis of pertussis in unvaccinated children, in whom the sensitivity of paired specimens or a single elevated titer more than 5 weeks after the onset of symptoms is 50% to 90%.[58,64] Multiple antigens are often used in antibody assays to enhance sensitivity, but pertussis toxin is the only antigen unique to *B. pertussis*.[58,64] The interpretation of serologic results in vaccinated persons is more difficult. The meaning of a single elevated antibody titer should be defined in relation to age-matched, population-specific controls.[63,64] Paired specimens have limited utility because the anamnestic response to infection is usually so rapid that no significant rise in antibody concentrations occurs between acute and convalescent sera.[63,64]

TREATMENT

Antibiotic therapy with erythromycin speeds elimination of *B. pertussis* from the nasopharynx and reduces transmission of infection.[58,60] Antibiotic treatment may also ameliorate the severity of disease, even if it is started in the paroxysmal phase, but this

is less clear. The estolate form of erythromycin, 40 mg/kg/day (maximum, 2 g/day) divided in two to four doses, is preferred and should be administered for 7 to 14 days.[65] There is limited evidence that the newer macrolides, clarithromycin (15 to 20 mg/kg/day in two divided doses, to a maximum dosage of 1 g/day for 7 days) and azithromycin dihydrate (15 to 20 mg/kg/day in one dose, to a maximum dosage of 600 mg/day for 5 days), are also effective.[65] Alternative agents that speed elimination of *B. pertussis* include trimethoprim-sulfamethoxazole, tetracycline, and chloramphenicol.[60] Fluoroquinolones are active against *B. pertussis* in vitro, but their efficacy in the treatment of pertussis has not been demonstrated.

The value of antibiotic therapy in adults with persistent cough and suspected pertussis is unknown.

Infants with pertussis should be hospitalized for monitoring, airway suctioning, oxygen supplementation, hydration, and nutritional support. Treatment of severe cases with pertussis immune globulin has shown promise in animal studies and human trials.[66] Systemic corticosteroids may be helpful in infants with life-threatening disease.[60] Inhaled corticosteroids and bronchodilators are of uncertain value in ameliorating symptoms.

PREVENTION

Pertussis can be prevented by quarantine of infected individuals, prophylactic antibiotic treatment of exposed infants, and vaccination. Patients with active pertussis in the catarrhal and early paroxysmal stages should be isolated for 5 to 7 days after starting antibiotics. Prophylactic treatment of household contacts with a full course of erythromycin can reduce the spread of disease and is recommended for young children who have not completed primary immunization. In neonates, however, erythromycin prophylaxis is associated with an increased risk of hypertrophic pyloric stenosis.[67]

Immunization with inactivated whole-cell vaccine was introduced in 1948, and acellular vaccines composed of inactivated pertussis toxin and one or more additional components were licensed in 1991. Current recommendations call for primary immunization with an acellular vaccine, which is combined with diphtheria and tetanus toxoids (DTaP), at ages 2, 4, and 6 months, with boosters at 15 to 18 months and at 4 to 6 years.[68] Pertussis vaccines provide approximately 80% protection from clinical pertussis, and immunized persons who become ill after exposure experience less severe disease.[58,68] Protection declines with time.[58] Historically, boosters have not been given to older children and adults because of the unacceptable toxicity of the whole-cell vaccine; reactions to whole-cell pertussis vaccine (DTP) in children include local pain and swelling, fever, anorexia, irritability, vomiting, hypotonic hyporesponsiveness, seizures, and anaphylaxis. All of these reactions are much less common with the acellular vaccines.[58,68] The acellular vaccine is immunogenic and well tolerated by adolescents and adults.[69,70] The routine use of vaccine in adolescents and adults, who constitute the major reservoir for pertussis, has the potential to reduce the overall burden of pertussis and the transmission of the disease to children; however, recommendations for boosters in these age groups have not been established.

Pseudomonas aeruginosa Infections

P. aeruginosa is an important opportunistic pathogen, with a prominent role in cystic fibrosis. The organism was first isolated from a surgical wound in 1882 by the French pharmacist Carle Gessard.[71] The species designations in both the original name for the organism, *Bacillus pyocyaneus,* and its modern name, *P. aeruginosa,* refer to the blue-green pigments produced by these bacteria, which may serve as clinical markers for infection.[72] More than a dozen species of *Pseudomonas* are now recognized, but only *P. aeruginosa* is an important human pathogen.

EPIDEMIOLOGY

P. aeruginosa is widely distributed in nature, particularly in water, soil, and vegetation. Nutritional versatility, suppression of competitors, and resistance to antibiotics contribute to the diverse habitat of this organism.[72] *P. aeruginosa* is not part of the normal flora of most humans, but it can be cultured from the stool, throat, nose, or moist areas of the skin of up to 20% of healthy persons.[71,72]

P. aeruginosa is mainly an opportunistic and nosocomial pathogen. It has a marked predilection for colonizing damaged epithelium, such as burned skin, bronchiectatic airways, and sites of indwelling appliances such as tracheostomy tubes, endotracheal tubes, urinary catheters, intravascular devices, peritoneal dialysis catheters, and body piercings.[71,73] Patients with cystic fibrosis are especially prone to airway colonization. The risk of colonization at any site increases with exposure to broad-spectrum antibiotics and duration of hospitalization.[72] Colonization is an important antecedent to invasive infection, particularly in neutropenic patients.[71,72] Additional risk factors for invasive disease include cytotoxic chemotherapy, treatment with corticosteroids, and HIV infection.[71,74] Persons with diabetes mellitus are at risk for invasive external otitis.

Outbreaks of *P. aeruginosa* infection can result from common-source exposures. Community outbreaks of folliculitis have been associated with swimming pools, whirlpool baths, and hot tubs. Ocular infections have resulted from contaminated cosmetics and contact lens solution.[71] In hospitals, epidemics of *P. aeruginosa* infection have been linked to contaminated sinks, mops, cleaning solutions, therapy pools, respiratory therapy equipment, endoscopic devices, flowers, vegetables, and water pitchers.[71]

PATHOGENESIS

P. aeruginosa is a motile gram-negative bacillus that is typically thin and either straight or slightly curved. It does not ferment sugars and grows strictly aerobically. *P. aeruginosa* has simple nutritional requirements and tolerates a wide range of environmental conditions. It grows well on most laboratory media and emits a sweet, grapelike odor. Most strains produce one or more diffusable pigments, including pyocyanin, pyoverdin, pyorubin, and pyomelanin.[71]

P. aeruginosa is a formidable pathogen that is well equipped by an exceptionally large genome to colonize and invade epithelial surfaces, undermine host defenses, and induce systemic toxicity. *P. aeruginosa* attaches to tracheobronchial mucus and replicates in a biofilm on the mucosal surface, using quorum sensing to coordinate environmental adaptation and virulence factor production.[74,75] Although *P. aeruginosa* adheres poorly to normal epithelium, it readily attaches to damaged epithelial cells via pili and other adhesins.[74,75] The persistence of some strains of *P. aeruginosa* on epithelial surfaces is aided by the abundant secretion of tenacious mucopolysaccharide, which facilitates adherence to host cells, inhibits ciliary activity, and blocks ingestion by phagocytes. The virulence of *P. aeruginosa* is further aided by a type III secretion system that permits injection of exotoxins S, T, U, and Y directly into host cells, thus promoting cytotoxicity and

inflammation.[74,75] Invasion across epithelial surfaces by *P. aeruginosa* is made possible by proteolytic enzymes that damage epithelial cells, loosen tight junctions, and digest extracellular matrix.[74,75] These enzymes also interfere with host defenses by cleaving immunoglobulins, complement, and cytokines. Secreted toxins such as exotoxin A, leukocidin, hemolysins, and pyocyanin cause additional tissue damage, hemolysis, and cytotoxicity to macrophages, neutrophils, and lymphocytes. Vascular injury by exotoxin A can result in local or distant hemorrhagic necrosis. Damage to epithelial and endothelial barriers permits the entry of bacteria, bacterial products, and endogenous proinflammatory mediators into the systemic circulation, leading to systemic toxicity and shock.

The striking association of *P. aeruginosa* with cystic fibrosis (CF) is an area of special interest.[76] Most patients with CF are initially colonized with an environmental, nonmucoid strain of *P. aeruginosa*. These bacteria eventually undergo a phenotypic shift in the CF airway to express an unusual lipopolysaccharide structure and hypersecrete mucoid exopolysaccharide (alginate).[76] The factors controlling these adaptations and the mechanisms underlying the persistence of *P. aeruginosa* in the CF airway remain incompletely understood. *P. aeruginosa* binds somewhat more avidly to CF epithelium than to normal epithelial cells.[75,76] However, the uptake of *P. aeruginosa* by epithelial cells is defective in CF. The cystic fibrosis transmembrane conductance regulator (CFTR) may serve as a receptor for the internalization of surface bacteria, and shedding of these infected cells may be a normal defense mechanism against *P. aeruginosa* that is defective in CF.[75,76] Also, human β defensin 1, an antibacterial peptide secreted by airway epithelial cells, is inactive in the high salt concentrations that some investigators have found in CF airways.[75,76] Chronic inflammation precipitated by *P. aeruginosa* infection contributes to the airway injury and progressive loss of pulmonary function that characterize CF.

CLINICAL SYNDROMES

Pneumonia

Nosocomial pneumonia *P. aeruginosa* is one of the most common causes of hospital-acquired pneumonia and is implicated in approximately 25% of cases complicating mechanical ventilation.[21,77] Risks for nosocomial *P. aeruginosa* pneumonia include prolonged hospitalization, mechanical ventilation for 7 days or longer, and exposure to broad-spectrum antibiotics.[77] Tracheal colonization with *P. aeruginosa* usually precedes infection.

The clinical features of ventilator-associated pneumonia are nonspecific: fever, purulent tracheal secretions, leukocytosis, and new or changing infiltrates on chest films.[21,77] The radiographic abnormalities are multifocal and bilateral in most cases of *P. aeruginosa* pneumonia and are more severe than those observed with other nosocomial pathogens.[21] Cavitation is evident on plain films or chest tomography in about 25% of cases.[78]

The diagnosis of hospital-acquired pneumonia in critically ill patients is challenging because tracheal colonization with potential pathogens cannot be distinguished easily from infection on clinical grounds.[21,77] The diagnosis of *P. aeruginosa* pneumonia, particularly in intubated patients, is most accurately made from quantitative cultures of bronchoscopic specimens. Yields of colony-forming units (CFU) of 1,000/ml from a protected specimen brush or 10,000/ml from a bronchoalveolar lavage are indicative of infection. Blood cultures are positive in less than 10% of cases.

Hospital-acquired pneumonia caused by *P. aeruginosa* is associated with a worse outcome than infections caused by other pathogens. In well-documented cases of *P. aeruginosa* ventilator-associated pneumonia, the reported crude mortality ranges from 45% to 85% and the attributable mortality is about 40%.[77,79] In fatal cases, death usually results from septic shock or multiple-organ failure.[77] Relapses are relatively common after successful treatment.[77,79]

Community-acquired pneumonia *P. aeruginosa* is a rare cause of community-acquired pneumonia but is disproportionately represented among the most severe cases.[80,81] Most patients with *P. aeruginosa* pneumonia have underlying conditions that place them at particular risk for colonization of the lower airways with *P. aeruginosa* (e.g., cystic fibrosis, bronchiectasis, or tracheostomy); have important defects in host defenses caused by malignancy, aplastic anemia, cytotoxic chemotherapy, or advanced HIV infection; or have been recently hospitalized.[80,81] The few cases of *P. aeruginosa* pneumonia that have been reported in healthy persons have occurred predominantly in smokers.[80]

Community-acquired *P. aeruginosa* pneumonia is typically acute in onset and rapidly progressive.[72,80] Fever, cough, chest pain, and progressive dyspnea are the usual symptoms. The cough is generally productive of purulent sputum that may be streaked with blood. On examination, patients are apprehensive, often confused, and acutely ill, with fever, tachypnea, tachycardia, hypotension, and cyanosis. The chest x-ray typically reveals multifocal, nodular alveolar infiltrates, often bilateral. Small areas of parenchymal cavitation may be evident, but large effusions and empyema are unusual. Routine laboratory findings are nonspecific and may include leukocytosis or leukopenia, elevated transaminase levels, and azotemia. Sputum Gram stains demonstrate gram-negative bacilli in nearly all cases, and sputum cultures grow *P. aeruginosa* in more than 80% of cases.[72,80] Blood cultures often are positive. The clinical course is usually rapidly progressive despite therapy, with most patients experiencing respiratory failure. Mortality exceeds 25%.[80,81]

Pneumonia in HIV disease *P. aeruginosa* infections of the lower respiratory tract are unusually common in patients with advanced HIV disease, and the incidence may be increasing.[74,82] These infections occur in the setting of very low $CD4^+$ T cell counts (median, < 25 cells/mm^3). Additional risk factors in some cases include neutropenia, recent antibiotic therapy, and treatment with corticosteroids. Most infections are community acquired. The clinical course can be fulminant, but a subacute presentation also has been described, characterized by productive cough, fever, and dyspnea for days to weeks. The radiographic features are similar to those of other cases of community-acquired *P. aeruginosa* pneumonia. *P. aeruginosa* is readily identified in purulent sputum in the majority of cases; blood cultures are positive in 10% to 30% of cases.

Most patients with *P. aeruginosa* pneumonia complicating HIV infection respond to antibiotic treatment, but relapses are common.

Cystic Fibrosis

More than 80% of patients with CF develop chronic airway infection with mucoid strains of *P. aeruginosa*. This infection is associated with chronic airway inflammation, progressive loss of pulmonary function, and eventual death from respiratory failure.[74,76] Intensive and prolonged treatment of initial infection with aerosolized tobramycin or colistin may delay chronic infec-

tion and preserve pulmonary function.[76] Once chronic infection with mucoid *P. aeruginosa* is established, it is virtually impossible to eradicate.

Patients with CF experience pulmonary exacerbations that are associated with an increased density of *P. aeruginosa* in sputum. These episodes are characterized symptomatically by increased coughing, more purulent or bloody sputum, chest congestion, and dyspnea. Physical signs include tachypnea, use of accessory muscles for breathing, fever, and weight loss. A decline in airflow is evident on pulmonary function testing, and the chest x-ray may demonstrate new or changing infiltrates.[76]

Treatment of pulmonary exacerbations in CF includes chest physiotherapy, bronchodilators, mucolytics, and combination antibiotic therapy guided by in vitro susceptibility testing of sputum isolates.[76] Intermittent maintenance therapy with inhaled tobramycin (300 mg twice daily during alternate months) can suppress *P. aeruginosa,* improve airflow, and reduce the need for hospitalization.[83] Macrolides such as azithromycin also may improve lung function in CF patients colonized with *P. aeruginosa*, by mechanisms that are unclear.[84]

Bacteremia

P. aeruginosa bacteremia occurs predominantly in hospitalized patients with severe underlying disease.[74,85-87] The most common sources of infection are the lower respiratory tract, the urinary tract, and the skin; intravascular catheters are implicated in 7% to 16% of cases. No primary source of infection can be identified in up to 50% of episodes. The overall mortality of *P. aeruginosa* bacteremia is approximately 20%. An increased risk of death is observed in patients with pneumonia, severe sepsis, fatal underlying disease, and delayed administration of effective antibiotic therapy.

Endocarditis

P. aeruginosa is a rare cause of infective endocarditis. The majority of reported cases have occurred in injection drug users.[74,88] Occasional episodes have complicated cardiac surgery or involved seeding of a prosthetic device. Most cases have been localized to the right side of the heart and have developed on normal valves.

Patients with right-sided infection often present with a prolonged febrile illness and may complain of cough or chest pain associated with radiographic evidence of septic pulmonary emboli.[88] Patients with aortic or mitral valve involvement follow a more acute clinical course that may include evidence of congestive heart failure, myocardial abscess, or systemic embolization.[88] The diagnosis of *P. aeruginosa* endocarditis is made from blood cultures, cardiac auscultation, and echocardiography.

Most right-sided infections can be cured by 6 weeks of high-dose parenteral treatment with an extended-spectrum penicillin and an aminoglycoside. Left-sided infections usually require valve replacement surgery in conjunction with medical therapy.[74,88,89] The overall survival in left-sided infections is approximately 50%.

Urinary Tract Infections

P. aeruginosa is a common cause of urinary tract infection in hospitals and chronic care facilities. These infections usually follow instrumentation, surgery, or catheterization of the urinary tract, often in patients recently treated with antibiotics.[72,74] Symptomatic infections can be treated with single-agent antibiotic therapy. Catheter-associated infections are unlikely to be eradicated unless the catheter is removed.

Ear Infections

Swimmer's ear *P. aeruginosa* is commonly isolated from cases of acute diffuse external otitis (swimmer's ear), which is associated with freshwater swimming in warm, humid climates.[74] This condition is characterized by an inflamed, draining, pruritic, and sometimes painful external auditory canal. Local cleansing followed by topical treatment with an otic solution containing 2% acetic acid or an antibiotic in combination with corticosteroids speeds recovery. Recurrences are common.

Malignant (necrotizing) otitis externa *P. aeruginosa* is the pathogen in nearly all cases of malignant external otitis, a chronic, invasive infection of the external auditory canal that predominantly afflicts elderly persons with diabetes mellitus.[74,90] Rarely, cases have been reported in AIDS patients and immunocompromised children. The infection spreads to the temporal bone and mastoid, often involving the adjacent cranial nerves as they exit the skull. Severe otalgia is the presenting symptom, and most patients have purulent otorrhea with an intact tympanic membrane. Facial nerve paresis is evident in 30% to 40% of cases; other cranial neuropathies are less common. Fever and other systemic signs of infection are usually absent. The erythrocyte sedimentation rate (ESR) is nearly always markedly elevated. Technetium-99m bone scanning is very sensitive but not specific for this condition. CT and magnetic resonance imaging are useful for defining the extent of disease, with MRI being more sensitive for delineating the soft tissue component.[74,90] The diagnosis can be made from the clinical presentation, imaging studies, and cultures of diseased bone or granulation tissue in the external canal.

Local debridement and antibiotic therapy for 4 to 8 weeks are successful in most cases. An oral fluoroquinolone (e.g., ciprofloxacin, 750 mg twice daily) is usually effective. Parenteral therapy with an antipseudomonal β-lactam with or without an aminoglycoside may be required for recalcitrant cases or those caused by resistant strains.[74,90] The ESR, serial gallium-67 scans, or both can be used as markers of disease activity and guides to duration of therapy.

Eye Infections

P. aeruginosa is an important cause of keratitis, particularly in contact lens wearers.[74] This infection develops as a complication of corneal injury, which permits bacterial attachment to the damaged epithelium and subsequent invasion.

The clinical manifestations begin with a small, painful ulcer that spreads with gray discoloration and edema of the cornea, anterior chamber reaction, and mucopurulent exudate. The microbiologic diagnosis can be made by Gram stain and culture of purulent material in the base of the ulcer.

Emergent treatment is with frequent (every 15 to 30 minutes) topical applications of an ophthalmic solution containing high concentrations of an aminoglycoside or fluoroquinolone. *P. aeruginosa* is also occasionally implicated in posttraumatic endophthalmitis and nosocomial conjunctivitis.

Central Nervous System Infections

Meningitis caused by *P. aeruginosa* occurs most often as a complication of neurosurgery.[74] Rare cases result from *P. aeruginosa* bacteremia, particularly in debilitated patients with cancer or severe burns.

Fever and depressed consciousness are the cardinal manifestations. The onset often is insidious in postoperative cases. The

diagnosis is made from culture and Gram stain of the cerebrospinal fluid.

Ceftazidime is the preferred treatment, often given in combination with an intravenous aminoglycoside. Alternative agents that achieve adequate levels in the cerebrospinal fluid include meropenem, aztreonam, and ciprofloxacin. Intrathecal therapy with aminoglycosides may be required for refractory cases or highly resistant organisms.[74]

Bone and Joint Infections

P. aeruginosa is associated with a distinctive spectrum of acute and chronic osteochondral infections that result from hematogenous seeding, direct inoculation, or contiguous spread from adjacent tissues.[74] Hematogenous infections are most commonly seen in injection drug users and typically involve the sternoclavicular joints, symphysis pubis, sacroiliac joints, or cervical vertebrae. Lumbosacral vertebral osteomyelitis may complicate instrumentation of the urinary tract via contamination of the venous plexus shared by the pelvis and the spine. *P. aeruginosa* is also the most common cause of osteochondritis after puncture wounds of the foot. This syndrome is most often seen in children and results from direct inoculation of bacteria colonizing the moist insoles of their sneakers. Chronic osteomyelitis caused by *P. aeruginosa* can develop as a complication of orthopedic surgery, sternotomy, or contiguous infection of pressure sores or ischemic skin ulcers.

The cardinal symptom of acute osteochondral *P. aeruginosa* infection is pain lasting days to weeks. Examination is notable for local tenderness and diminished mobility, frequently with warmth and erythema. Fever and leukocytosis are often present, and the ESR is almost always elevated. The clinical presentation of chronic osteomyelitis is subtler than that of acute infection. Pain, nonunion, and draining ulcers are common, often without systemic manifestations.

Imaging studies are important in the evaluation of *P. aeruginosa* bone and joint infections.[91] Plain x-rays or CT scans can reveal evidence of bony destruction or periosteal reaction but may be negative early in the clinical course. MRI is more sensitive than radiography, particularly for nonosseous changes. Technetium-99m bone scans are invariably positive in osteomyelitis but may be normal when the infection is confined to the joint. Confirmation of *P. aeruginosa* as the pathogen is made from joint aspiration or bone biopsy. Blood cultures are negative in most cases.

Treatment of *P. aeruginosa* osteochondral infections usually requires 4 to 6 weeks of antibiotic therapy. An antipseudomonal β-lactam in combination with an aminoglycoside is generally recommended, but single-agent therapy, particularly with a fluoroquinolone, may be effective.[74,91] Surgical debridement is necessary when bone necrosis is evident, but most infections confined to a joint can be treated with antibiotics alone. In the case of *P. aeruginosa* osteochondritis of the foot in children, 1 week of antibiotic therapy after debridement is sufficient for cure. Chronic infections are difficult to eradicate, and suppressive therapy may be required.

Skin Infections

P. aeruginosa causes a variety of cutaneous infections that can affect normal skin after submersion in contaminated water.

Folliculitis Outbreaks of folliculitis, or swimmer's itch, have been associated with the use of hot tubs, whirlpools, therapy baths, and swimming pools.[74] The presenting feature is a diffuse, pruritic eruption that may be follicular, maculopapular, vesicular, or pustular. The rash develops 8 hours to 5 days after exposure and usually involves skin that had been covered by a swimsuit. Mild headache and other complaints may be present, but fever is unusual. *P. aeruginosa* can be identified from stains and cultures of pustular lesions. Topical drying agents may ease symptoms, but the rash resolves within 7 days without specific therapy.

Hot-foot syndrome Another waterborne infection is *Pseudomonas* hot-foot syndrome, which is characterized by the development of tender, warm, erythematous 1- to 2-cm nodules on a background of plantar or palmar erythema.[92] These lesions appeared in children 10 to 40 hours after being in a contaminated wading pool and resolved without antibiotic treatment within 7 to 14 days.

Green-nail syndrome This syndrome results from discoloration of the nail plate by pyocyanin and occurs in persons who frequently submerge their hands. Local cleansing and debridement, including draining of any associated paronychia, are generally successful.[74,93]

Toe web infection *P. aeruginosa* is a frequent cause of toe web infections. These occur most often in tropical climates and may complicate minor trauma or tinea pedis. The affected toe webs are macerated, thickened, and painfully fissured; a purulent exudate is often present. Examination under a Wood lamp can reveal green-white fluorescence from the elaboration of pyoverdin. Toe web infections often can be managed with local measures, but systemic antibiotics may be needed in severe cases.[74,93]

Burn wound sepsis In years past, *P. aeruginosa* was the most common agent of burn wound sepsis, which was the leading cause of death after thermal injury. With the development of effective topical antimicrobials, wound excision, and other advances in burn wound care, burn wound sepsis has become far less common. Fungi are now the dominant pathogens, but *P. aeruginosa* remains an important consideration. Suspicion of wound sepsis should be aroused by the appearance of dark-brown, black, or violaceous discoloration of the wound. Hemorrhage, purulent drainage, or green pigment also may be present. Blood cultures may be positive, but the diagnosis of invasive infection is made from wound biopsies demonstrating microbial invasion of normal tissue. Treatment of *P. aeruginosa* infection is with systemic antibiotics, topical mafenide acetate cream, and surgical debridement.[94] *P. aeruginosa* can also cause posttraumatic cellulitis, deep abscesses, necrotizing fasciitis, and other locally invasive pyodermas, particularly in immunocompromised persons.[74]

Ecthyma gangrenosum Ecthyma gangrenosum is a distinctive skin infection that results in most cases from *P. aeruginosa* bacteremia, usually in the setting of neutropenia, cancer, burns, AIDS, or other severely debilitating conditions.[74,93] The lesions may be discrete or multiple and begin as painless macules or nodules that may become vesicular or bullous. The lesions undergo central hemorrhagic necrosis and ulceration over a period of 12 to 24 hours, with a surrounding rim of tender erythema. Ecthyma gangrenosum is characterized histologically by bacterial invasion of dermal veins. Blood cultures are usually positive, and *P. aeruginosa* can be demonstrated in biopsies or scrapings from the base of

the ulcer. Treatment requires prompt administration of intravenous antibiotics and management of the underlying condition.

ANTIBIOTIC TREATMENT

The treatment of *P. aeruginosa* infections is particularly challenging because of the remarkable resistance of this organism to most antibiotics and its propensity to develop further resistance during the course of therapy. The intrinsic antimicrobial insensitivity of *P. aeruginosa* arises from several complementary mechanisms.[74,95] First, a relatively impermeable outer membrane limits the access of antibiotics to their binding sites. Second, an active efflux pump can expel agents that penetrate the outer barrier. Third, expression of a chromosomally encoded β-lactamase is induced on exposure to β-lactam antibiotics. Furthermore, mutations in *P. aeruginosa* can result in the rapid selection of resistant organisms during the course of treatment. Fourth, *P. aeruginosa* can acquire plasmids bearing additional resistance genes, such as those encoding for extended-spectrum β-lactamases, carbapenemases, and enzymes that inactivate aminoglycosides.[74,95] Finally, the growth of *P. aeruginosa* in biofilms induces the production of periplasmic glucans that bind and sequester antibiotics.[96]

Antimicrobial agents that are effective against most strains of *P. aeruginosa* include β-lactams, fluoroquinolones, and aminoglycosides [*see 7:XIV Chemotherapy of Infection*]. In the United States, the β-lactams showing the greatest activity against *P. aeruginosa* are meropenem and piperacillin-tazobactam; the most effective fluoroquinolone remains ciprofloxacin, and the most active aminoglycoside is amikacin.[97] Local resistance patterns vary, however, so it is important to determine the antibiotic susceptibilities of individual strains by in vitro testing. Hospital isolates exhibit greater resistance than outpatient isolates, and organisms cultured in intensive care units are the most resistant. Multidrug-resistant strains are increasing in prevalence; these may be susceptible to antibiotic combinations or polymyxins.[74,97]

Treatment of suspected *P. aeruginosa* infection should begin with high doses of an antipseudomonal β-lactam in combination with an aminoglycoside or fluoroquinolone. The use of two agents increases the likelihood of effective initial therapy, which is associated with reduced mortality.[85,87] Whether combination therapy should be continued after sensitivities are known is unclear.[98] On the one hand, the combination of an aminoglycoside with a β-lactam antibiotic results in synergistic killing of *P. aeruginosa* and improves outcome in some animal models of infection.[80,99] Prospective observational studies have found combination therapy with β-lactams and aminoglycosides to be associated with improved survival in neutropenic or bacteremic patients with *P. aeruginosa* infections.[99,100] On the other hand, aminoglycosides are relatively toxic, and other antibiotic combinations are less predictably synergistic against *P. aeruginosa*.[80] Recent retrospective clinical studies have not identified any survival advantage of dual-drug therapy over treatment with a single effective agent.[77,85-87] Combination therapy also has not been shown to suppress the emergence of resistant strains. Despite this uncertainty, it may be prudent to use combination therapy whenever possible because of the high mortality and refractory nature of *P. aeruginosa* infections. An exception is urinary tract infection, which can be treated effectively with a single agent. Most *P. aeruginosa* infections can be cured with 2 to 3 weeks of antibiotic therapy, but longer courses are required for endocarditis and osteomyelitis.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Turk DC, May JR: *Haemophilus influenzae*: its clinical importance. English Universities Press, London, 1967

2. Murphy TF: *Haemophilus* infections. Principles and Practice of Infectious Diseases, 6th ed. Mandell GL, Bennett JE, Dolin R, Eds. Elsevier Churchill Livingstone, Philadelphia, 2005, p 2661

3. Rao VK, Krasan GP, Hendrixson DR, et al: Molecular determinants of the pathogenesis of disease due to nontypable *Haemophilus influenzae*. FEMS Microbiol Rev 23:99, 1999

4. Murphy TF: *Haemophilus influenzae* in chronic bronchitis. Semin Respir Infect 15:41, 2000

5. Bresser P, Out TA, van Alphen L, et al: Airway inflammation in nonobstructive and obstructive chronic bronchitis with chronic *Haemophilus influenzae* airway infection: comparison with noninfected patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 162:947, 2000

6. Hopkins RS, Jajosky RA, Hall PA, et al: Summary of notifiable diseases—United States, 2003. MMWR Morb Mortal Wkly Rep 52:1, 2005

7. Bisgard KM, Kao A, Leake J, et al: *Haemophilus influenzae* invasive disease in the United States, 1994–1995: near disappearance of a vaccine-preventable childhood disease. Emerg Infect Dis 4:229, 1998

8. Perdue DG, Bulkow LR, Gellin BG, et al: Invasive *Haemophilus influenzae* disease in Alaskan residents aged 10 years and older before and after infant vaccination programs. JAMA 283:3089, 2000

9. Hirschmann JV, Everett ED: *Haemophilus influenzae* infections in adults: report of nine cases and a review of the literature. Medicine (Baltimore) 58:80, 1979

10. Invasive *Haemophilus influenzae* disease in adults: a prospective, population-based surveillance. CDC Meningitis Surveillance Group. Ann Intern Med 116:806, 1992

11. Sarangi J, Cartwright K, Stuart J, et al: Invasive *Haemophilus influenzae* disease in adults. Epidemiol Infect 124:441, 2000

12. Peltola H: Worldwide *Haemophilus influenzae* type b disease at the beginning of the 21st century: global analysis of the disease burden 25 years after the use of the polysaccharide vaccine and a decade after the advent of conjugates. Clin Microbiol Rev 13:302, 2000

13. Durand ML, Calderwood SB, Weber DJ, et al: Acute bacterial meningitis in adults: a review of 493 episodes. N Engl J Med 328:21, 1993

14. Hussein AS, Shafran SD: Acute bacterial meningitis in adults: a 12-year review. Medicine (Baltimore) 79:360, 2000

15. Spagnuolo PJ, Ellner JJ, Lerner PI, et al: *Haemophilus influenzae* meningitis: the spectrum of disease in adults. Medicine (Baltimore) 61:74, 1982

16. Tunkel AR, Hartman BJ, Kaplan SL, et al: Practice guidelines for bacterial meningitis. Clin Infect Dis 39:1267, 2004

17. Kucera CM, Silverstein MD, Jacobson RM, et al: Epiglottitis in adults and children in Olmsted County, Minnesota, 1976 through 1990. Mayo Clin Proc 71:1155, 1996

18. Skerrett SJ: Diagnostic testing for community-acquired pneumonia. Clin Chest Med 20:531, 1999

19. Musher DM, Kubitschek KR, Crennan J, et al: Pneumonia and acute febrile tracheobronchitis due to *Haemophilus influenzae*. Ann Intern Med 99:444, 1983

20. *Haemophilus influenzae* pneumonia in human immunodefiency virus–infected patients. Grupo Andaluz para el Estudio de las Enfermedades Infecciosas. Clin Infect Dis 30:461, 2000

21. Niederman MS, Craven DE, Bonten MJ, et al: Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. Am J Respir Crit Care Med 171:388, 2005

22. Sethi S, Evans N, Grant BJ, et al: New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med 347:465, 2002

23. Klein JO: Otitis media. Clin Infect Dis 19:823, 1994

24. Casey JR, Pichichero ME: Changes in frequency and pathogens causing acute otitis media in 1995–2003. Pediatr Infect Dis J 23:824, 2004

25. Sande MA, Gwaltney J: Acute community-acquired bacterial sinusitis: continuing challenges and current management. Clin Infect Dis 39:S151, 2004

26. Kostman JR, Sherry BL, Fligner CL, et al: Invasive *Haemophilus influenzae* infections in older children and adults in Seattle. Clin Infect Dis 17:389, 1993

27. Sonnen GM, Henry NK: Pediatric bone and joint infections: diagnosis and antimicrobial management. Pediatr Clin North Am 43:933, 1996

28. Borenstein DG, Simon GL: *Hemophilus influenzae* septic arthritis in adults: a report of four cases and a review of the literature. Medicine (Baltimore) 65:191, 1986

29. Perkins MD, Mirrett S, Reller LB: Rapid bacterial antigen detection is not clinically useful. J Clin Microbiol 33:1486, 1995

30. Thornsberry C, Sahm DF, Kelly LJ, et al: Regional trends in antimicrobial resistance among clinical isolates of *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* in the United States: results from the TRUST surveillance program, 1999–2000. Clin Infect Dis 34:S4, 2002

31. Johnson DM, Sader HS, Fritsche TR, et al: Susceptibility trends of *Haemophilus influenzae* and *Moraxella catarrhalis* against orally administered antimicrobial agents: five-year report from the SENTRY Antimicrobial Surveillance Program. Diagn Microbiol Infect Dis 47:373, 2003

32. *Haemophilus influenzae* Infection. 2003 Red Book: Report of the Committee on Infectious Diseases, 26th ed. Pickering LK, Ed. American Academy of Pediatrics, Elk Grove Village, Illinois, 2003, p 262

33. *Haemophilus* endocarditis: report of 42 cases in adults and review. Haemophilus Endocarditis Study Group. Clin Infect Dis 24:1087, 1997

34. Berk SL: From *Micrococcus* to *Moraxella*: the reemergence of *Branhamella catarrhalis*. Arch Intern Med 150:2254, 1990

35. Verduin CAM, Hol C, Fleer A, et al: *Moraxella catarrhalis*: from emerging to established pathogen. Clin Microbiol Rev 15:125, 2002

36. Murphy TF, Brauer AL, Grant BJ, et al: *Moraxella catarrhalis* in chronic obstructive pulmonary disease: burden of disease and immune response. Am J Respir Crit Care Med 172:195, 2005

37. Fraser DW, Tsai TR, Orenstein W, et al: Legionnaires' disease: description of an epidemic of pneumonia. N Engl J Med 297:1189, 1977

38. McDade JE, Shepard CC, Fraser DW, et al: Legionnaires' disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N Engl J Med 297:1197, 1977

39. Winn WC Jr: Legionnaires disease: historical perspective. Clin Microbiol Rev 1:60, 1988

40. Fields BS, Benson RF, Besser RE: *Legionella* and Legionnaires' disease: 25 years of investigation. Clin Microbiol Rev 15:506, 2002

41. Stout JE, Yu VL: Legionellosis. N Engl J Med 337:682, 1997

42. Roig J, Rello J: Legionnaires' disease: a rational approach to therapy. J Antimicrob Chemother 51:1119, 2003

43. Fields BS, Haupt T, Davis JP, et al: Pontiac fever due to *Legionella micdadei* from a whirlpool spa: possible role of bacterial endotoxin. J Infect Dis 184:1289, 2001

44. Swanson MS, Hammer BK: *Legionella pneumophila* pathogenesis: a fateful journey from amoebae to macrophages. Annu Rev Microbiol 54:567, 2000

45. Neild AL, Roy CR: Immunity to vacuolar pathogens: what can we learn from *Legionella*? Cell Microbiol 6:1011, 2004

46. Skerrett SJ: Host defenses against respiratory infection. Med Clin North Am 78:941, 1994

47. Hawn TR, Verbon A, Lettinga KD, et al: A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. J Exp Med 198:1563, 2003

48. Hawn TR, Verbon A, Janer M, et al: Toll-like receptor 4 polymorphisms are associated with resistance to Legionnaires' disease. Proc Natl Acad Sci USA 102:2487, 2005

49. Winn WC, Myerowitz RL: The pathology of the legionella pneumonias: a review of 74 cases and the literature. Hum Pathol 12:401, 1981

50. Kirby BD, Snyder KM, Meyer RD, et al: Legionnaires' disease: report of sixty-five nosocomially acquired cases of review of the literature. Medicine (Baltimore) 59:188, 1980

51. Sopena N, Sabria-Leal M, Pedro-Botet ML, et al: Comparative study of the clinical presentation of *Legionella* pneumonia and other community-acquired pneumonias. Chest 113:1195, 1998

52. The radiologic manifestations of Legionnaire's disease. The Ohio Community-Based Pneumonia Incidence Study Group. Chest 117:398, 2000

53. Den Boer JW, Yzerman EPF: Diagnosis of *Legionella* infection in Legionnaires' disease. Eur J Clin Microbiol Infect Dis 23:871, 2004

54. Blazquez RM, Espinosa FJ, Martinez-Toldos CM, et al: Sensitivity of urinary antigen test in relation to clinical severity in a large outbreak of *Legionella* pneumonia in Spain. Eur J Clin Microbiol Infect Dis 24:488, 2005

55. Edelstein PH: Antimicrobial chemotherapy for Legionnaires disease: time for a change. Ann Intern Med 129:328, 1998

56. Sabria M, Pedro-Botet ML, Gomez J, et al: Fluoroquinolones vs macrolides in the treatment of Legionnaires disease. Legionnaires Disease Therapy Group. Chest 128:1401, 2005

57. Lapin JH: Whooping cough. Charles C. Thomas, Springfield, Illinois, 1943

58. Matoo S, Cherry JD: Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. Clin Microbiol Rev 18:326, 2005

59. Guris D, Strebel PM, Bardenheier B, et al: Changing epidemiology of pertussis in the United States: increasing reported incidence among adolescents and adults, 1990–1996. Clin Infect Dis 28:1230, 1999

60. Kerr JR, Matthews RC: *Bordetella pertussis* infection: pathogenesis, diagnosis, management, and the role of protective immunity. Eur J Clin Microbiol Infect Dis 19:77, 2000

61. Crowcroft NS, Stein C, Duclos P, et al: How best to estimate the global burdern of pertussis? Lancet Infect Dis 3:413, 2003

62. Gordon M, Davies HD, Gold R: Clinical and microbiologic features of children presenting with pertussis to a Canadian pediatric hospital during an eleven-year period. Pediatr Infect Dis J 13:617, 1994

63. Senzilet LD, Halperin SA, Spika JS, et al: Pertussis is a frequent cause of prolonged cough illness in adults and adolescents. Clin Infect Dis 32:1691, 2001

64. Muller FM, Hoppe JE, Wirsing von Konig CH: Laboratory diagnosis of pertussis: state of the art in 1997. J Clin Microbiol 35:2435, 1997

65. Pertussis. 2003 Red Book: Report of the Committee on Infectious Diseases, 26th ed. Pickering LK, Ed. American Academy of Pediatrics, Elk Grove Village, Illinois, 2003, p 472

66. Bruss JB, Malley R, Halperin S, et al: Treatment of severe pertussis: a study of the safety and pharmacology of intravenous pertussis immunoglobulin. Pediatr Infect Dis J 18:505, 1999

67. Honein MA, Paulozzi LJ, Himelright IM, et al: Infantile hypertrophic pyloric stenosis after pertussis prophylaxis with erythromycin: a case review and cohort study. Lancet 354:2101, 1999

68. Pertussis vaccination: use of acellular pertussis vaccines among infants and young children. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 46(RR-7):1, 1997

69. Ward JI, Cherry JD, Chang SJ, et al: Efficacy of an acellular pertussis vaccine among adolescents and adults. APERT Study Group. N Engl J Med 353:1555, 2005

70. Pichichero ME, Rennels MB, Edwards KM, et al: Combined tetanus, diphtheria, and 5-component pertussis vaccine for use in adolescents and adults. JAMA 293:3003, 2005

71. Morrison AJ Jr, Wenzel RP: Epidemiology of infections due to *Pseudomonas aeruginosa*. Rev Infect Dis 6 (suppl 3):S627, 1984

72. Bodey GP, Bolivar R, Fainstein V, et al: Infections caused by *Pseudomonas aeruginosa*. Rev Infect Dis 5:279, 1983

73. Keene WE, Markum AC, Samadpour M: Outbreak of *Pseudomonas aeruginosa* infections caused by commercial piercing of upper ear cartilage. JAMA 291:981, 2004

74. Pier GB, Ramphal R: *Pseudomonas aeruginosa*. Principles and Practice of Infectious Diseases, 6th ed. Mandell GL, Bennett JE, Dolin R, Eds. Elsevier Churchill Livingstone, Philadelphia, 2005, p 2587

75. Sadikot RT, Blackwell TS, Christman JW, et al: Pathogen-host interactions in *Pseudomonas aeruginosa* pneumonia. Am J Respir Crit Care Med 171:1209, 2005

76. Gibson RL, Burns JL, Ramsey BW: Pathophysiology and management of pulmonary infections in cystic fibrosis. Am J Respir Crit Care Med 168:918, 2003

77. Chastre J, Fagon JY: Ventilator-associated pneumonia. Am J Respir Crit Care Med 165:867, 2002

78. Winer-Muram HT, Jennings SG, Wunderink RG, et al: Ventilator-associated *Pseudomonas aeruginosa* pneumonia: radiographic findings. Radiology 195:247, 1995

79. Chastre J, Wolff M, Fagon JY, et al: Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults. JAMA 290:2588, 2003

80. Hatchette TF, Gupta R, Marrie TJ: *Pseudomonas aeruginosa* community-acquired pneumonia in previously healthy adults: case report and review of the literature. Clin Infect Dis 31:1349, 2000

81. Arancibia F, Bauer TT, Ewig S, et al: Community-acquired pneumonia due to Gram negative bacteria and *Pseudomonas aeruginosa*: incidence, risk, and prognosis. Arch Intern Med 162:1849, 2002

82. Afessa B, Green B: Bacterial pneumonia in hospitalized patients with HIV infection: the Pulmonary Complications, ICU Support, and Prognostic Factors of Hospitalized Patients with HIV (PIP) Study. Chest 117:1017, 2000

83. Intermittent administration of inhaled tobramycin in patients with cystic fibrosis. Cystic Fibrosis Inhaled Tobramycin Study Group. N Engl J Med 340:23, 1999

84. Southern KW, Barker PM: Azithromycin for cystic fibrosis. Eur Respir J 24: 834, 2004

85. Vidal F, Mensa J, Almela M, et al: Epidemiology and outcome of *Pseudomonas aeruginosa* bacteremia, with special emphasis on the influence of antibiotic treatment: analysis of 189 episodes. Arch Intern Med 156:2121, 1996

86. Chatzinikolaou I, Abi-Said D, Bodey GP, et al: Recent experience with *Pseudomonas aeruginosa* bacteremia in patients with cancer: retrospective analysis of 245 episodes. Arch Intern Med 160:501, 2000

87. Micek ST, Lloyd AE, Ritchie DS, et al: *Pseudomonas aeruginosa* bloodstream infection: importance of appropriate initial antimicrobial treatment. Antimicrob Agents Chemother 49:1306, 2005

88. Komshian SV, Tablan OC, Palutke W, et al: Characteristics of left-sided endocarditis due to *Pseudomonas aeruginosa* in the Detroit Medical Center. Rev Infect Dis 12:693, 1990

89. Gavin PJ, Suseno MT, Cook FV, et al: Left-sided endocarditis caused by *Pseudomonas aeruginosa*: successful treatment with meropenem and tobramycin. Diagn Microbiol Infect Dis 47:427, 2003

90. Grandis JR, Bransteter IV BF, Yu VL: The changing face of malignant (necrotising) external otitis: clinical, radiological, and anatomic correlations. Lancet Infect Dis 4:34, 2004

91. Lew DP, Waldvogel FA: Osteomyelitis. Lancet 364:369, 2004

92. Fiorillo L, Zucker M, Sawyer D, et al: The *Pseudomonas* hot-foot syndrome. N Engl J Med 345:335, 2001

93. Agger WA, Mardan A: *Pseudomonas aeruginosa* infections of intact skin. Clin Infect Dis 20:302, 1995

94. Tredget EE, Shankowsky HA, Rennie R, et al: *Pseudomonas* infections in the thermally injured patient. Burns 30:3, 2004

95. Livermore DM: Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? Clin Infect Dis 34:634, 2002

96. Mah TF, Pitts B, Pellock B, et al: A genetic basis for *Pseudomonas aeruginosa* biofilm antibiotic resistance. Nature 426:306, 2003

97. Karlowsky JA, Jones ME, Thornsberry C, et al: Stable antimicrobial susceptibility rates for clinical isolates of *Pseudomonas aeruginosa* from the 2001–2003 tracking resistance in the United States today surveillance studies. Clin Infect Dis 40(suppl 2):S89, 2005

98. Safdar N, Hendelsman J, Maki DG: Does combination antimicrobial therapy reduce mortality in gram-negative bacteremia? A meta-analysis. Lancet Infect Dis 4:519, 2004

99. Hilf M, Yu VL, Sharp J, et al: Antibiotic therapy for *Pseudomonas aeruginosa* bacteremia: outcome correlations in a prospective study of 200 patients. Am J Med 87:540, 1989

100. Leibovici L, Paul M, Poznanski O, et al: Monotherapy versus beta-lactam–aminoglycoside combination treatment for gram-negative bacteremia: a prospective, observational study. Antimicrob Agents Chemother 41:1127, 1997

XI INFECTIONS DUE TO BRUCELLA, FRANCISELLA, YERSINIA PESTIS, AND BARTONELLA

W. CONRAD LILES, M.D, PH.D.

Brucellosis, tularemia, (bubonic) plague, and bartonellosis are zoonoses (i.e., infectious diseases that can be transmitted from animals to humans) caused by gram-negative bacilli. The causative agents of these diseases are *Brucella* species, *Francisella tularensis, Yersinia pestis,* and *Bartonella* species, respectively. Infected arthropods, such as ticks or fleas, can serve as vectors for the transmission of tularemia, plague, and bartonellosis. In general, the diagnosis of these zoonoses requires the physician to consider the clinical presentation (which is not always distinctive) in light of the epidemiology of these diseases [*see Table 1*]. Treatment is principally with appropriate empirical antimicrobials [*see Table 2*].

Brucellosis

Brucellosis continues to be a major zoonosis worldwide.[1] The animal reservoirs for the disease include a variety of domesticated animals: cattle, goats, sheep, pigs, and dogs. Human brucellosis (variously known as undulant fever, Malta fever, and Mediterranean fever) is rare in the United States but remains a significant health problem in developing countries. In humans, brucellosis is a disease of protean manifestations, ranging from an indolent febrile syndrome to fulminant endocarditis. Like plague and tularemia, brucellosis is recognized as a potential agent of biological terrorism.[2]

EPIDEMIOLOGY

Virtually all cases of brucellosis arise from direct or indirect exposure to animals.[1,3-5] In animals, brucellosis is a chronic disease that persists lifelong and causes infectious abortion and sterility. *B. abortus* infects predominantly cattle; *B. melitensis,* goats and sheep; and *B. suis,* domestic and wild swine. *B. canis*, the least common cause of human brucellosis, is mostly found in kennel-raised dogs. Infected animals shed large numbers of *Brucella* organisms in their milk, urine, and afterbirth. Hence, brucellosis is considered an occupational hazard for farmers, ranchers, veterinarians, abattoir workers, and laboratory personnel.[3,4,6]

Brucellosis is distributed worldwide, but its incidence varies markedly, depending largely on the degree of its control in domestic animals and the pasteurization of milk. The disease is relatively common in southern Europe, the Middle East, the

Table 1 Diagnosis of Diseases Caused by *Brucella*, *Francisella*, *Yersinia*, and *Bartonella* Species

	Brucella *(Brucellosis)*	Francisella tularensis *(Tularemia)*	Yersinia pestis *(Plague)*	Bartonella *(Oroya Fever, Bacteremia, Cat-Scratch Disease)*
Animal reservoir	Cattle, goats, sheep, swine, dogs	Widespread: mammals, birds, fish, amphibians, arthropods	Rodents (rats, ground squirrels, prairie dogs); other animals possible	Cats, phlebotomine sandflies (Andes Mts.)
Transmission	Direct contact through skin lesions, airborne, ingestion of contaminated dairy products	Direct contact, airborne, arthropod bites	Flea bites, animal bites or scratches, airborne	Oroya fever: sandfly bites Bacteremia: body lice Cat-scratch disease: cat bites or scratches; possibly, cat flea bites
Incubation period (range)	10–14 days (5 days to several months)	3–7 days (1–21 days)	Bubonic: 2–7 days Pneumonic: 2–3 days (1–14 days)	Oroya fever: 3–12 wk Bacteremia: days to months Cat-scratch disease: 3–10 days
Clinical manifestations	Protean: fever and tachycardia, flulike syndrome, liver or spleen enlargement, lymphadenopathy, rashes or skin lesions	Flulike illness, ulcer at site of skin inoculation, regional lymphadenopathy, rash, GI symptoms	Flulike illness, buboes (inguinal/axillary lymphadenopathy) Septicemic: GI symptoms	Oroya fever: flulike illness, lymphadenopathy, hemolytic anemia, thrombocytopenia Bacteremia: fever, malaise, anemia, thrombocytopenia, splenomegaly, cardiac murmur, bacillary angiomatosis (in HIV) Cat-scratch disease: inoculation-site lesion, regional lymphadenopathy, low-grade fever
Laboratory tests	WBCs normal or decreased; RBCs and platelets typically decreased; ESR variable; culture of body fluid and tissue; serology	Screening tests often unremarkable; mild to moderate elevation of AST and ALT; serology (late)	WBCs 15,000–25,000/µl, left shift; fibrin split products ↑; DIC possible; AST, ALT, and bilirubin elevated; culture of body fluid and tissue; direct immunofluorescence	Oroya fever: blood culture Bacteremia: blood culture Cat-scratch disease: diagnosis usually clinical

ALT—alanine aminotransferase AST—aspartate aminotransferase ESR—erythrocyte sedimentation rate RBC—red blood cell WBC—white blood cell

Table 2 Treatment of Diseases Caused by *Brucella*, *Francisella*, *Yersinia*, and *Bartonella* Species

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Cost*	*Comments*
Brucella *(Brucellosis)*	Doxycycline *plus* Gentamicin *or* Streptomycin	100 mg p.o., b.i.d., for 6 wk 3–5 mg/kg/day I.V. in divided doses for 2–3 wk 1 g I.M. q.d. for 2–3 wk	First choice	\$0.05—0.12/day 120 mg I.V. q. 8 hr: \$5.00–5.99/day 500 mg I.M. b.i.d.: \$3.00–3.99/day	Relapse rate ~ 6%
	Doxycycline *plus* Rifampin	100 mg p.o., b.i.d., for 6 wk 600–900 mg p.o., b.i.d., for 6 wk	Alternative choice	\$0.05–0.12/day 600 mg p.o. once daily: \$60.00–69.99/mo.	Relapse rate ~ 15%
	Ciprofloxacin	500 mg p.o., b.i.d., for 30 days, with rifampin (see above)	Adjunctive therapy	—	Adjunctive therapy
Francisella tularensis *(Tularemia)*	Streptomycin	500 mg–1 g I.M. q. 12 hr for 7–14 days	First choice	500 mg I.M. b.i.d.: \$3.00–3.99/day	—
	Gentamicin	3–5 mg/kg/day I.V. in divided doses q. 8 hr for 7-14 days	Equally effective	120 mg I.V. q. 8 hr: \$5.00–5.99/day	—
	Tetracycline	500 mg p.o., q.i.d., for 21–28 days	Alternative choice	500 mg p.o., q.i.d.: \$0.10—0.24/day	Relapse can occur if therapy < 21 days
	Doxycycline	100 mg p.o., b.i.d., for 21–28 days	Alternative choice	—	Relapse can occur if therapy < 21 days
	Chloramphenicol	1 g I.V., p.o.* q. 6 hr for 21–28 days	Alternative choice	1 g I.V. q. 6 hr: \$150.00–199.99/day	Relapse can occur if therapy < 21 days; consider for treatment if signs or symptoms of meningitis are present
	Ciprofloxacin	500 mg p.o., b.i.d., for 14–21 days	Alternative choice	500 mg p.o., b.i.d.: \$7.00–7.99/day	Limited data
Yersinia pestis *(Bubonic Plague)*	Streptomycin	30 mg/kg/day I.M. in divided doses q.12 hr, for 10 days or at least 3 days after clinical recovery; or for 5 days, followed by gentamicin for 5–10 days (see below)	First choice	500 mg I.M. b.i.d.: \$3.00–3.99/day	Continue antibiotic therapy for 10 days, or at least 3 days after clinical recovery
	Gentamicin	2 mg/kg I.V. loading dose, then 1.7 mg/kg I.V. q. 8 hr	Equally effective	120 mg I.V. q. 8 hr: \$5.00–5.99/day	Continue antibiotic therapy for 10 days, or at least 3 days after clinical recovery
	Tetracycline	500 mg I.V./p.o., q.i.d.	Alternative choice	500 mg p.o., q.i.d.: \$0.10–0.24/day	Continue antibiotic therapy for 10 days, or at least 3 days after clinical recovery
	Doxycycline	100 mg I.V./p.o., b.i.d.	Alternative choice	100 mg p.o., b.i.d.: \$0.05–0.12/day	Continue antibiotic therapy for 10 days, or at least 3 days after clinical recovery
	Chloramphenicol	500 mg–1 g I.V./p.o.* q.i.d.	Alternative choice	1 g I.V. q. 6 hr: \$150.00–199.99/day	Continue antibiotic therapy for 10 days, or at least 3 days after clinical recovery
	Tetracycline	500 mg p.o., q.i.d.	Chemoprophylaxis	\$0.10–0.24/day	Chemoprophylaxis for contacts of patients with pneumonic plague
	Doxycycline	100 mg p.o., b.i.d.	Chemoprophylaxis	\$0.05–0.12/day	Chemoprophylaxis for contacts of patients with pneumonic plague
	Streptomycin	20 mg/kg/day I.M. in two divided doses	Chemoprophylaxis	500 mg I.M., b.i.d.: \$3.00–3.99/day	Chemoprophylaxis for contacts younger than 8 yr of age of patients with pneumonic plague
	Trimethoprim-sulfamethoxazole	40 mg/kg p.o., b.i.d.	Chemoprophylaxis	800/160 mg p.o., b.i.d.: \$0.10–0.24/day	Chemoprophylaxis for contacts younger than 8 yr of age of patients with pneumonic plague

Table 2 (*continued*)

Table 2 Treatment of Diseases Caused by *Brucella*, *Francisella*, *Yersinia*, and *Bartonella* Species

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Cost*	*Comments*
Bartonella (*Bartonellosis*)	OROYA FEVER				
	Chloramphenicol	500 mg p.o.*/I.V. q.i.d. for ≥ 1 wk	First choice	1 g I.V. q. 6 hr: $150.00–199.99/day	—
	Tetracycline	500 mg p.o./I.V. q.i.d. for ≥ 2 wk	Alternative choice	$0.10–0.24/day	—
	Doxycycline	100 mg p.o./I.V. q.i.d. for ≥ 2 wk	Alternative choice	$0.05–0.12/day	—
	Ampicillin	500 mg p.o./I.V. q.i.d. for ≥ 2 wk	Alternative choice	500 mg p.o., q.i.d.: $0.75–0.99/day	—
	URBAN TRENCH FEVER WITH BACTEREMIA				
	Erythromycin	500 mg p.o., q.i.d., for 14 days	—	—	Azithromycin and clarithromycin are probably equally effective
	Doxycycline	100 mg p.o., b.i.d., for 14 days	—	$0.05—0.12/day	—
	URBAN TRENCH FEVER WITH ENDOCARDITIS				
	Gentamicin	2 mg/kg I.V. loading dose, then 1.7 mg/kg I.V. q. 8 hr	—	120 mg I.V. q. 8 hr: $5.00–5.99/day	In conjunction with valve replacement
	Ciprofloxacin *plus* Rifampin	Ciprofloxacin, 500 mg p.o., b.i.d., plus rifampin, 600 mg p.o., q.d, for ≥ 4–6 mo	—	—	Limited data; monitor for surgical valve replacement; levofloxacin can be substituted for ciprofloxacin
	Ciprofloxacin *plus* a macrolide	Ciprofloxacin, 500 mg p.o., b.i.d., plus an oral macrolide (erythromycin, clarithromycin, or azithromycin) for ≥ 4–6 mo	—	—	Limited data; monitor for surgical valve replacement; levofloxacin can be substituted for ciprofloxacin
	CAT-SCRATCH DISEASE				
	Azithromycin	500 mg p.o. once, then 250 mg p.o., q.d., for 4 days	—	250 mg p.o., q.d.: $6.00–6.99/day	Azithromycin is only agent shown to be effective in placebo-controlled, double-blind clinical trial; alternative choices include rifampin, gentamicin, trimethoprim-sulfamethoxazole, doxycycline, ciprofloxacin, and levofloxacin
	BACILLARY ANGIOMATOSIS				
	Erythromycin	500 mg p.o., q.i.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	First choice	250 mg (base) p.o., q.i.d.: $1.00–1.49/day	—
	Doxycycline	100 mg p.o., b.i.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	First choice	$0.05–0.12/day	—
	Clarithromycin	500 mg p.o., b.i.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	Alternative choice	500 mg p.o., b.i.d.: $7.00–7.99/day	—
	Azithromycin	250 mg p.o., q.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	Alternative choice	250 mg p.o., q.d.: $6.00–6.99/day	—
	Ciprofloxacin	500–750 mg p.o., b.i.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	Alternative choice	500 mg p.o., b.i.d.: $7.00–7.99/day	—
	Doxycycline *plus* Rifampin	Doxycycline: 100 mg I.V./po., b.i.d., plus rifampin, 300 mg p.o., b.i.d., for 2 mo (4 mo for osteomyelitis or peliosis hepatis)	Alternative choice	$0.05–0.12/day	For severe or relapsed cases

*Note: Chloramphenicol is not distributed as an oral drug in the United States.

Indian subcontinent, parts of Africa, Mexico, and Central and South America.[1,3-5]

In the United States, the incidence of brucellosis has declined to less than 100 cases annually, coincident with public health programs and the virtual elimination of *B. abortus* infection in cattle. Currently, most cases of brucellosis acquired in the United States are caused by *B. melitensis* and can be linked to ingestion of unpasteurized goat-milk products from foreign countries, such as Mexico, and exposure to *Brucella* cultures in microbiology laboratory personnel.[4,6-8] Transmission of brucellosis by bone marrow transplantation was recently reported.[9]

ETIOLOGY

Brucellosis is transmitted to humans by direct contact with infected animals or their body fluids (inoculation occurs through cuts or abraded skin); via inhalation of aerosolized

organisms; or by ingestion of unpasteurized dairy products. Because of the low density of organisms in muscle, consumption of meat products has rarely been implicated in infection.[1,3-5]

PATHOGENESIS

Brucella organisms are small, nonmotile, non–spore-forming, gram-negative coccobacilli. The organisms are relatively hardy and can survive in dairy products or refrigerated meats for weeks to months. Although they grow aerobically, some species require supplemental carbon dioxide for primary isolation. In culture, *Brucella* strains are relatively slow growing; isolation from clinical specimens can require incubation for 30 days or more.[1,4]

Infection can occur via the gastrointestinal tract, broken skin, oral mucosa, conjunctiva, or respiratory tract. Once inoculated, bacteria travel to regional lymph nodes and multiply. They then enter the bloodstream and localize in cells of the reticuloendothelial system (particularly the liver, spleen, bone marrow, and lymph nodes) and kidney. At tissue sites of infection, noncaseating granulomas characteristically develop. Bacteremia can lead to metastatic infection at various sites, including bone, joints, meninges, and cardiac valves.[1,3-5]

DIAGNOSIS

Clinical Manifestations

The incubation period for brucellosis is usually 10 to 14 days, but it ranges from 5 days to several months. Onset of illness can be abrupt or insidious. The presentation is nonspecific; fever, chills, headache, myalgias, anorexia, malaise, and fatigue are common. Fever and tachycardia are present in approximately 90% of cases, hepatomegaly in 25% to 30%, splenomegaly in about 40%, and lymphadenopathy (most commonly cervical) in 10% to 40%. Cutaneous manifestations, which occur in approximately 5% of patients, include a variety of nonspecific maculopapular rashes, erythema nodosum, petechiae, and purpura. Other symptoms and signs will depend on specific organ involvement in individual cases.

Brucellosis can follow an acute, subacute, or chronic course. Untreated disease may persist for months to years and may cause a syndrome of relapsing fevers. Because the symptoms and signs of brucellosis may be nonspecific, it is important to obtain a detailed history that includes occupation, travel, exposure to animals, or ingestion of high-risk foods, such as unpasteurized dairy products[1,3-5] [*see Table 1*].

Laboratory Tests

The leukocyte count is typically normal or decreased. Anemia and thrombocytopenia are relatively common. The erythrocyte sedimentation rate is variable and rarely helpful.

Definitive diagnosis of brucellosis depends on isolation of the organism from blood, bone marrow, urine, cerebrospinal fluid, or other body fluids or tissues. The sensitivity of a single blood culture ranges from 15% to 70%, depending on the method used and the length of incubation. Because *Brucella* organisms grow slowly, cultures should be maintained for at least 4 weeks to increase recovery. The use of lysis centrifugation culture systems has been reported to hasten recovery. Culture of bone marrow is reportedly more sensitive than conventional blood culture.

Presumptive diagnosis of brucellosis can be made by serologic studies showing specific brucella IgG antibodies or the presence of IgM antibody to *Brucella*.[3-5,10] It should be noted that the standard brucella serum agglutination test does not detect antibodies to *B. canis*. In suspected *B. canis* infection, serologic tests specific for that species should be ordered. Molecular diagnosis using polymerase chain reaction has been developed but is currently not available for routine clinical use.[11]

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of brucellosis includes enteric fever, malaria, infectious mononucleosis (Epstein-Barr virus [EBV] infection), atypical mononucleosis (cytomegalovirus infection), Q fever, miliary tuberculosis, subacute bacterial endocarditis, HIV infection, and visceral leishmaniasis (kala-azar).[4]

TREATMENT

Brucellosis requires prolonged treatment with a tetracycline, typically supplemented with a second antibiotic: an aminoglycoside, rifampin, or perhaps a fluoroquinolone[1,3-5] [*see Table 2*]. Significant resistance to trimethoprim-sulfamethoxazole has limited the use of this agent.[5] Recent studies from Spain and the Middle East suggest that the best initial regimen is doxycycline (100 mg p.o., b.i.d.) for 6 weeks, plus either gentamicin (3 to 5 mg/kg a day I.V. in divided doses) or streptomycin (1 g/day I.M.) for the first 2 to 3 weeks. Relapse occurs in approximately 6% of patients treated with this regimen.[12] Doxycycline can also be combined with 6 weeks of rifampin (600 to 900 mg p.o., b.i.d.), although relapse has been reported in approximately 15% of patients treated with this regimen.[13] Fluoroquinolones have demonstrated good activity against *Brucella* in vitro, but clinical results have been disappointing; therefore, those agents should be regarded as adjunctive rather than primary therapy.[13] Although corticosteroid therapy has been recommended for management of brucellosis involving the central nervous system, no controlled studies have demonstrated its efficacy. No satisfactory vaccines for human brucellosis are currently available.[1]

COMPLICATIONS

Brucellosis can cause complications in almost any organ system. Endocarditis occurs in less than 2% of cases but accounts for the majority of brucellosis-related deaths. Infected aneurysms of the brain, aorta, and other vessels can occur. Direct invasion of the central nervous system occurs in only about 5% of cases. Meningitis is the most frequent CNS complication. Bone and joint involvement is relatively common, with a reported occurrence of 20% to 60%. Sacroiliitis and spondylitis are the most common musculoskeletal complications.[14] The GI tract is involved in as many as 70% of patients. Inflammation of Peyer patches can cause ileitis or colitis. Brucellosis can cause granulomatous hepatitis, pyogenic hepatic abscesses, or cholecystitis. Although renal involvement in brucellosis is unusual, orchitis has been reported in 20% of men with brucellosis. Respiratory tract involvement in brucellosis ranges from flulike symptoms to pneumonia with or without pleural effusions. A variety of ocular complications have been reported, including uveitis and endophthalmitis.

PROGNOSIS

A relapse rate of 5% to 25% can be expected in patients treated for brucellosis, depending on the antibiotic regimen employed (see above). Relapse is especially common in spondylitis.[14] Mortality is highest in cases of endocarditis. Cardiac valve replace-

ment surgery, in addition to prolonged antimicrobial therapy, is often required for successful resolution of endocarditis.[15]

Tularemia

Infection by *Francisella tularensis* causes clinical tularemia. This disease is a zoonosis of substantial complexity whose natural hosts or vectors include more than 100 species of wild mammals; several species of birds, fish, and amphibians; and more than 50 species of blood-sucking arthropods (including ticks, fleas, and deerflies). Tularemia has been recognized widely throughout the Northern Hemisphere, where it was formerly known by a variety of names, such as rabbit fever, deerfly fever, market men's disease, wild-hare disease, Ohara disease (Japan), and water-rat trappers' disease (Russia). *F. tularensis* could potentially be used in biological warfare or terrorism.[16]

EPIDEMIOLOGY

Tularemia is most commonly found in the temperate zones of the Northern Hemisphere. Five subspecies of *F. tularensis* have been recognized on the basis of biologic properties and geographic location. Type A biovar (*F. tularensis* biovar *tularensis*), the most virulent, is found only in North America, where it causes 70% to 90% of all cases of tularemia. This biovar is transmitted by infected rabbits and ticks, and untreated infection can be fatal. The less virulent type B biovar (*F. tularensis* biovar *palearctica*) is present in temperate zones worldwide. It is transmitted by rodents or aquatic animals and can cause mild or subclinical disease. *F. tularensis* biovar *palearctica mediasiatica* and *F. tularensis* biovar *palearctica japonica* are found in central Asia and Japan, respectively. *F. tularensis* biovar *novicida* rarely causes disease in humans or animals.

In the United States, human tularemia has been reported from every state except Hawaii.[17,18] In 1993, the last year in which tularemia was designated a notifiable disease by the Centers for Disease Control and Prevention (CDC), 132 cases were reported in the United States[19] [*see Figure 1*]. The disease clusters in the lower midwestern states, especially Arkansas, Oklahoma, and Missouri.[17,20]

The most important reservoirs of *F. tularensis* are rabbits, hares, and hard ticks.[18] The three major North American tick vectors of tularemia are *Dermacentor variabilis* (dog tick), which is widely distributed throughout North America; *Amblyomma americanum* (Lone star tick) in the southeastern and south central United States; and *D. andersoni* in the western United States. Infected *Ixodes* ticks can also transmit tularemia.[18,21,22] Aside from hard ticks, other reported arthropod vectors for transmission of tularemia to humans include deerflies, horseflies, and mosquitoes.[17,20,21]

Hunters, trappers, game wardens, veterinarians, meat handlers, laboratory workers, and pet owners are considered populations at increased risk for tularemia. However, as many as 40% of patients with tularemia relate no history of contact with a potential animal reservoir or arthropod vector.

ETIOLOGY

Humans generally acquire tularemia by direct contact with infected wild mammals or via bites from infected arthropods. Domestic animals, such as pet cats and dogs that have preyed on infected wild animals, can serve as sources of tularemia transmission to humans via bites, scratches, or licks.[23,24] Transmission has also resulted from inhalation of aerosolized organisms or ingestion of contaminated meat or water.[17,18,25] No documented human-to-human transmission of tularemia has been reported.[17,18,25,26]

The etiology of human tularemia shows some seasonal variation. During the summer, most cases occur as a result of arthropod bites. During the fall and winter, hunters and trap-

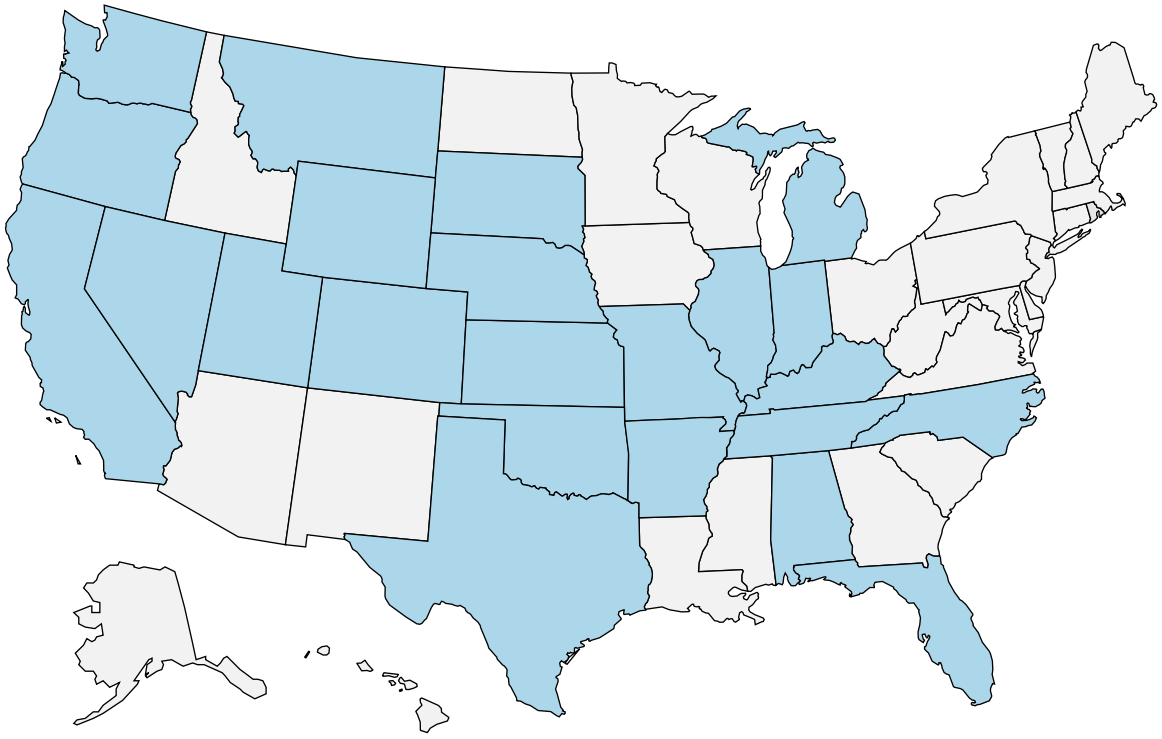


Figure 1 **States in blue are those reporting cases of tularemia in 1993.**

pers may acquire tularemia while skinning and dressing wild game.

PATHOGENESIS

F. tularensis is a small, pleomorphic, nonmotile, intracellular, gram-negative coccobacillus that grows fastidiously as an obligate aerobe. In culture, it requires cysteine-enriched blood agar for growth. In the environment, it can remain viable at low temperatures for 3 to 4 months in mud, water, or decayed animal carcasses.

F. tularensis is highly infectious. As few as 10 to 50 organisms can cause clinical tularemia after inhalation or intradermal injection, and 10^8 organisms can cause disease if ingested. The skin and mucous membranes are the most common portals of entry for *F. tularensis.* After inoculation, the organism multiplies locally to produce an erythematous, tender papule within 2 to 5 days, which subsequently becomes an ulcer with a black base (chancriform lesion). Bacteria then usually spread to regional lymph nodes, causing lymphadenopathy. Sepsis syndrome can occur with bacteremia or in cases of widely disseminated disease.

Histopathologically, tularemia is characteristically associated with mononuclear cell infiltration and granuloma formation. Those features can resemble the pathology of tuberculosis.

DIAGNOSIS

Clinical Manifestations

The usual incubation period of tularemia is 3 to 7 days (range, 1 to 21 days). Classically, tularemia presents as a severe illness with a sudden onset of fever, chills, headache, malaise, fatigue, and generalized myalgias and arthralgias. However, the severity of illness, especially in tick-borne disease, is highly variable, ranging from mild, afebrile, self-limited disease to rare cases of septic shock. Typically, an ulcer with surrounding inflammation develops at the site of inoculation. The inoculation ulcer may persist for weeks to months in untreated cases.

Most patients (70% to 90%) acquire infection by inoculation of the skin. Regional lymphadenopathy often develops in the inguinal/femoral nodes in adults and the cervical nodes in children. A generalized rash, usually maculopapular but occasionally pustular, is seen in approximately 20% of patients. Erythema nodosum has been reported infrequently.

The clinical manifestations of tularemia have been categorized into six recognized syndromes: (1) ulceroglandular, (2) glandular, (3) oculoglandular, (4) oropharyngeal, (5) pleuropulmonary, and (6) typhoidal.

Ulceroglandular tularemia and glandular tularemia are the most common forms of clinical tularemia, accounting for 75% to 85% of cases. Both result from inoculation of the skin. Ulceroglandular disease is characterized by an ulcerated lesion at the site of inoculation and regional lymphadenopathy. The glandular and ulceroglandular syndromes are similar except that the skin lesion is absent in glandular tularemia.

Oculoglandular tularemia is caused by accidental inoculation of the eye with contaminated fingers. Conjunctival inflammation is extremely painful, and numerous ulcers and a yellow exudate are often apparent. Conjunctival involvement is typically associated with preauricular lymphadenopathy (Parinaud complex).

Oropharyngeal tularemia can follow oral inoculation. This syndrome is characterized by an exudative or membranous pharyngitis and cervical lymphadenopathy. A subcategory of oropharyngeal tularemia that can develop after massive ingestion of *F. tularensis* is gastrointestinal tularemia. It is characterized by fever, intestinal ulceration, mesenteric lymphadenopathy, diarrhea, abdominal pain, nausea, vomiting, and gastrointestinal bleeding.

Pleuropulmonary tularemia can be a primary infection, following inhalation of aerosolized organisms, or a secondary manifestation resulting from hematogenous spread of infection to the lungs. The latter occurs in about 10% to 15% of cases of ulceroglandular tularemia. Mediastinal lymphadenopathy may evolve during the course of pleuropulmonary tularemia.

Typhoidal tularemia is the most difficult presentation to diagnose and is associated with the highest mortality.[17,18,25-27] Neither skin lesions nor lymphadenopathy are present; fever and malaise may be the only clinical manifestations. Secondary pleuropulmonary involvement occurs in approximately 50% of cases of typhoidal tularemia.

Laboratory Tests

Laboratory screening tests are often unremarkable in cases of tularemia. The leukocyte count may be slightly elevated or normal. Serum levels of hepatic aminotransferases tend to be mildly to moderately elevated. With pleuropulmonary involvement, chest films show parenchymal infiltrates (typically patchy, in one or multiple lobes), and pleural effusions are common.[17,18,25,26]

Definitive laboratory diagnosis of tularemia is generally based on detection of antibodies to *F. tularensis.* Serum agglutinins are usually detectable by day 10 to 14 of illness and peak approximately 2 to 6 weeks later. Early, appropriate antimicrobial therapy may blunt the rise in antibody titer. A fourfold increase in titer between acute-phase and convalescent-phase specimens is considered diagnostic, but a single specimen with a titer of 1:160 or greater is highly suggestive. A specific enzyme-linked immunosorbent assay (ELISA) for *F. tularensis* antibodies is now available and may eventually replace the serum agglutinins assay. PCR technology has been employed for molecular diagnosis but is not widely available for clinical use.[28-30]

Airborne transmission of *F. tularensis* is possible, so attempts to isolate and culture the organism should be avoided in most clinical laboratories. Biosafety level 2 is recommended for clinical laboratory work with material where contamination with *F. tularensis* is suspected. Biosafety level 3 is required for culture of the organism in large quantities.[31]

DIFFERENTIAL DIAGNOSIS

The skin lesion of ulceroglandular tularemia may resemble that seen in sporotrichosis, ecthyma from *Staphylococcus aureus* or *Streptococcus pyogenes*, skin infection with atypical mycobacteria (e.g., *Mycobacterium marinum*), syphilis, anthrax, rat-bite fever, or rickettsial disease. The lymphadenopathy of tularemia may be similar to that of plague, lymphogranuloma venereum, cat-scratch disease, or necrotizing lymphadenitis. Oropharyngeal tularemia can mimic the exudative pharyngitis caused by *S. pyogenes, Arcanobacterium haemolyticus, Corynebacterium diphtheriae,* and EBV. Pleuropulmonary tularemia is often similar to any of the atypical pneumonias, including those caused by respiratory viruses, *Mycoplasma pneumoniae, Legionella pneumophila, Chlamydia pneumoniae, C. psittaci, Coxiella burnetii, Histoplasma capsulatum,* and *Coccidioides immitis*.[26] Typhoidal tularemia may be confused with enteric fever (typhoid or paratyphoid fever), rickettsial infections, brucellosis, EBV, primary toxoplasmosis, miliary tuberculosis, sarcoidosis, or lymphoma.

TREATMENT

Because of the lack of rapid diagnostic laboratory tests for tularemia, initial treatment is often empirical and based on clinical suspicion. Although streptomycin (500 mg to 1 g I.M. every 12 hours) is the conventional drug of choice, evidence indicates that gentamicin (3 to 5 mg/kg daily I.V. in divided doses every 8 hours) is equally effective. With either agent, treatment should be given for 10 to 14 days in cases of severe clinical illness.[32] Tetracycline (or chloramphenicol) can be used, but relapse may occur if the duration of therapy is less than 21 days.[33] Ciprofloxacin was recently reported to be effective for treatment, but clinical experience is limited.[34-36]

Only partial immunity develops after infection with *F. tularensis*, and cases of reinfection have been documented. Chemoprophylaxis is not recommended for persons potentially exposed to tularemia. A live attenuated tularemia vaccine has been developed. It may be useful for the partial protection of laboratory workers but is not commercially available.[37,38]

COMPLICATIONS

Tularemia can be complicated by bacteremia, meningitis (usually marked by a lymphocytic pleocytosis in the CSF), rhabdomyolysis, acute renal failure, endocarditis, osteomyelitis, pericarditis, and septic shock.

PROGNOSIS

Typhoidal and ulceroglandular tularemia have the worst prognosis, with approximately 5% mortality reported in untreated cases. Prognosis is excellent for patients treated with appropriate antibiotics, with significant clinical improvement generally occurring within 48 hours of starting therapy.[17,18,25]

Plague

Yersinia infections primarily affect rodents, pigs, and birds; humans are accidental hosts. Nevertheless, *Yersinia pestis* (formerly *Pasteurella pestis*) has made its mark in human history as the causative agent of plague—also known as bubonic plague, black plague, and the Black Death. Plague was initially described in biblical times; in the Middle Ages, epidemic bubonic plague killed an estimated one fourth of Europe's population.[39] This disease of antiquity has persisted to the present. Although human plague is now relatively infrequent worldwide, aerosolized *Y. pestis* is well recognized for its potential use as a biological weapon.[40]

EPIDEMIOLOGY

Plague currently exists in widely scattered foci in Asia, Africa, and the Americas. In the 1990s, countries with the greatest number of cases included Tanzania, Madagascar, Democratic Republic of Congo, Vietnam, Peru, India, Myanmar, Zimbabwe, Mozambique, Uganda, and China [*see Figure 2*]. In the United States, about 10 cases of plague are reported annually, primarily from the Four Corners area, where Utah, Colorado, New Mexico, and Arizona meet. In that region, plague is endemic in prairie dog colonies.[41,42]

ETIOLOGY

Plague is a zoonotic infection that affects predominately urban and sylvatic rodents (e.g., rats and ground squirrels, respectively). It is transmitted to its natural animal reservoirs by flea bites or ingestion of contaminated animal tissues. Humans represent accidental hosts who usually acquire infection via bites from infected fleas.[39] Bites or scratches from an infected animal can also transmit plague to humans. In endemic areas, risk factors for acquiring plague include direct contact with

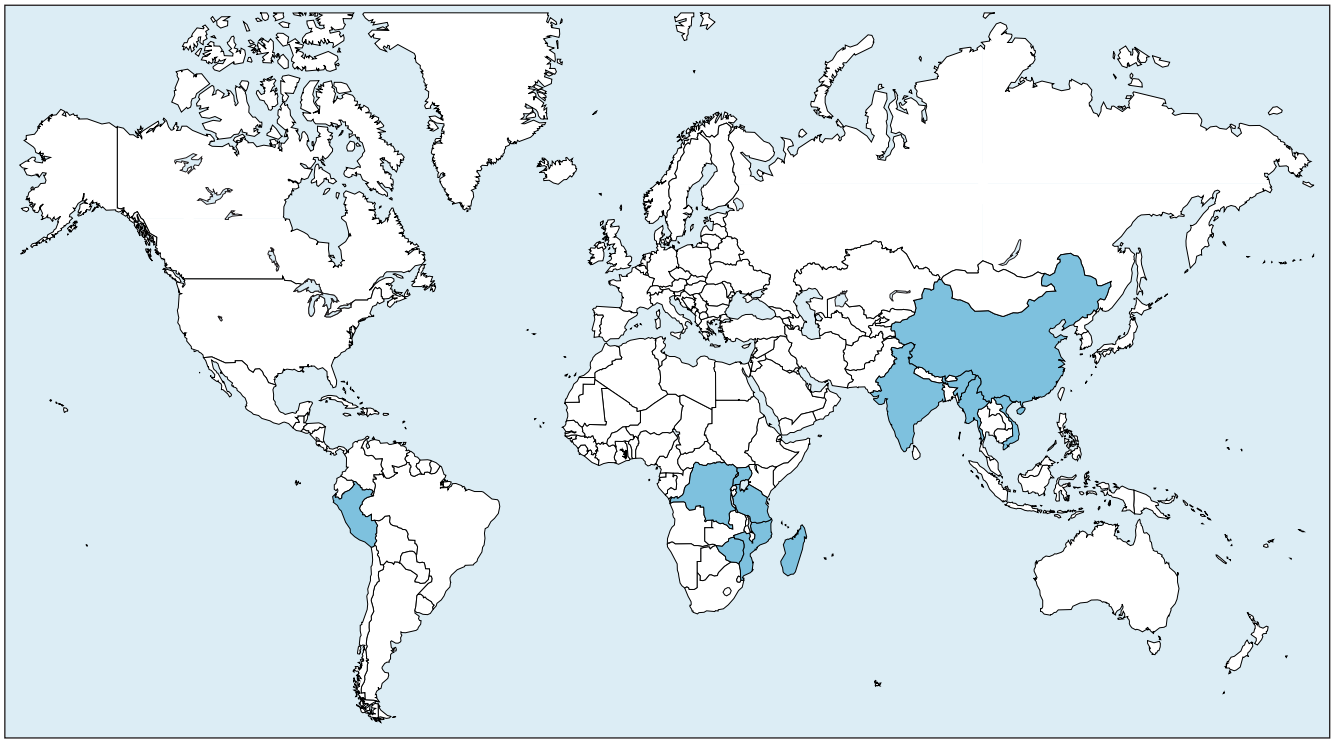


Figure 2 **Shaded countries reported more than 100 cases of plague from 1990–1995.**

rodents or carnivores that prey on them (e.g., cats, dogs), the presence of wild rodents near the place of residence, and pet dogs or cats with fleas.[43]

Plague may also result from inhalation of aerosolized bacilli. Airborne transmission can occur via the cough of an infected patient with pulmonary involvement or from close contact with an infected animal or cadaver.

PATHOGENESIS

Y. pestis is a pleomorphic, non–spore-forming, gram-negative, bipolar-staining coccobacillary member of the family Enterobacteriaceae. It is a nonmotile, facultative anaerobe that is able to grow aerobically on routine microbiologic media. The organism can survive at room temperature in dried blood or the environment for weeks to months.

Y. pestis is one of the most virulent bacteria known. It is usually inoculated through the skin or mucous membranes, where it invades cutaneous lymphatics. Monocytes and macrophages, which can phagocytize *Y. pestis* without killing it, may play a role in spread of infection from the inoculation site. Plague can involve almost any tissue or organ and cause massive destruction if left untreated. Initially, infected lymph nodes are tender and edematous. Later, hemorrhagic necrosis develops in affected nodes, and bacteria enter the bloodstream, where they disseminate to other sites (e.g., liver, spleen, lungs, and central nervous system). Infarcts and hemorrhagic, necrotic nodules are characteristic histopathologic findings at affected tissue sites.[39]

DIAGNOSIS

Clinical Manifestations

The usual incubation periods for bubonic plague and pneumonic plague are 2 to 7 days and 2 to 3 days, respectively (range, 1 to 14 days). Plague generally presents as an acute illness characterized by the abrupt onset of fever, chills, headache, gastrointestinal symptoms, and local pain, followed within hours by the development of a painful, swollen mass of lymph nodes (buboes) in the groin or axilla. Skin overlying buboes is usually red-purple in color. Buboes are initially tense and hard but rapidly become fluctuant. Spontaneous rupture and drainage can occur. Buboes do not develop in patients with septicemic plague; instead, these patients have gastrointestinal signs and symptoms: nausea, vomiting, diarrhea, and abdominal pain. Severe pharyngitis, severe diarrhea, and cough with or without hemoptysis may also be early clinical manifestations of plague[39,44] [*see Table 1*].

Laboratory Tests

The leukocyte count is typically elevated in the range of 15,000 to 25,000 cells/μl, with a shift to the left; leukemoid reactions (> 50,000 cells/μl) can occur. The platelet count may be normal or mildly depressed or may be very low if disseminated intravascular coagulation (DIC) is present. The level of fibrin split products is frequently elevated, even in patients without frank DIC. Serum levels of hepatic aminotransferases and bilirubin are often increased.

Definitive diagnosis of plague is usually based on isolation of *Y. pestis* from cultures of blood, other body fluids, or tissues. Aspiration of buboes may readily yield material for diagnosis. The organism can be easily grown on blood and MacConkey agar, as well as in infusion broth. A direct immunofluorescence test has been developed for rapid identification of *Y. pestis* in clinical specimens, and a passive hemagglutination test is available for presumptive diagnosis based on serology. Both of those tests are available from the CDC Division of Vector-Borne Infectious Diseases, Fort Collins, Colorado (970-221-6400).

DIFFERENTIAL DIAGNOSIS

Clinical suspicion based on knowledge of plague epidemiology is essential for timely diagnosis and institution of appropriate antimicrobial therapy. Plague should be suspected in febrile individuals who have been exposed to rodents or other mammals in regions of the world where plague is endemic. Illness may resemble that caused by a number of other serious infectious diseases, especially tularemia. Other causes of acute, painful lymphadenitis include staphylococcal, streptococcal, and *Pasteurella* infection. Plague pneumonia can be confused with other causes of atypical pneumonia or the hantavirus pulmonary syndrome. The combination of plague-associated abdominal symptoms and DIC can resemble an acute surgical abdomen or *Capnocytophaga canimorsus* sepsis.[44]

TREATMENT

Streptomycin (30 mg/kg I.M. in divided doses every 12 hours) remains the drug of choice for the treatment of plague. Timely administration of streptomycin reduces mortality to approximately 5%. Gentamicin, which is more widely available than streptomycin, also appears to be effective. To prevent relapses, antibiotic treatment should be continued for 10 days or for at least 3 days after defervescence and clinical recovery. Most patients improve rapidly and defervesce within 72 hours of initiation of antimicrobial therapy, although buboes can persist for weeks. Some authorities advocate switching from streptomycin to a different antibiotic for completion of therapy after 5 days of treatment in order to minimize the risk of ototoxicity and nephrotoxicity.

Tetracycline or doxycycline can also be used for plague. Chloramphenicol is the preferred agent for the treatment of plague meningitis, pleuritis, endophthalmitis, and myocarditis because of superior penetration into those tissues. Penicillins, cephalosporins, and macrolides are suboptimal agents for treatment of plague. Trimethoprim-sulfamethoxazole and fluoroquinolones appear to be active against plague in vitro and in animal models, but clinical experience with these agents for the treatment of plague is limited[39,44,45] [*see Table 2*].

All cases of suspected or documented plague should be reported to the state and city or county health departments, the CDC, and the World Health Organization. Plague patients with cough or other signs or symptoms of pneumonia should be placed in strict respiratory isolation for at least 48 hours after initiation of appropriate antibiotic therapy.

Contacts of patients with pneumonic plague or suspected septicemic plague with pulmonary involvement should receive chemoprophylaxis with either tetracycline (500 mg p.o., q.i.d.) or doxycycline (100 mg p.o., b.i.d.). Streptomycin (20 mg/kg/day I.M. in two divided doses) or trimethoprim-sulfamethoxazole (40 mg/kg p.o., b.i.d.) can be used for children younger than 8 years.

Plague vaccine (Cutter Laboratories) is a killed, formalin/phenol-fixed whole-bacteria vaccine that provides only partial protection against infection. Vaccinated individuals who are exposed to plague should still receive chemoprophylaxis.[46]

COMPLICATIONS

Septic shock, DIC, acute respiratory distress syndrome (ARDS), and meningitis are recognized complications of plague, especially in cases in which antimicrobial therapy is inadequate or delayed. Plague meningitis, a rare complication that typically occurs more than 1 week after inadequate antimicrobial therapy, is characterized by fever, headache, menigismus, and neutrophilic pleocytosis of the cerebrospinal fluid.

PROGNOSIS

Reported mortality of untreated plague ranges from 40% to 90%. Mortality is greatest in pneumonic plague. The overall mortality for cases of plague in the United States during the past 20 years has been approximately 15%.[42]

Bartonella Infections

Bartonella (formerly *Rochalimaea*) organisms are small, gram-negative bacilli that are extremely fastidious. They require specialized culture methods for isolation and specific staining techniques for detection in clinical specimens.[47-49] Infections by these organisms can occur in both immunocompromised and immunocompetent individuals.[47,49] Four *Bartonella* species have been documented to be pathogenic in humans: *B. bacilliformis, B. quintana, B. henselae,* and *B. elizabethae*.[47,49] The spectrum of diseases caused by *Bartonella* can be divided into five categories: (1) Oroya fever (classic bartonellosis), (2) trench fever, (3) bacteremia and endocarditis, (4) cat-scratch disease, and (5) bacillary angiomatosis.

OROYA FEVER (CLASSIC BARTONELLOSIS)

Oroya fever (chronic bartonellosis; Carrión disease) is caused by infection with *B. bacilliformis* transmitted to humans by the bite of infected phlebotomine sandflies. The disease is geographically restricted to valleys of the Andes Mountains at elevations between 800 m (2,624 ft) and 2,500 m (8,200 ft) above sea level, which are found in Peru, Ecuador, and Colombia.

Diagnosis

Oroya fever, caused by primary bacteremia, develops 3 to 12 weeks after inoculation. The onset of illness can be either insidious or abrupt. Clinical manifestations include fever, chills, headache, myalgias and arthralgias, abdominal pain, generalized lymphadenopathy, progressive hemolytic anemia, and thrombocytopenia.[47,49,50]

Definitive diagnosis of acute bartonellosis is based on the isolation of *B. bacilliformis* from blood cultures. The organism can be grown on Colombia agar with 5% defibrinated human blood or semisolid nutrient agar with 10% rabbit hemoglobin aerobically at 28º C. Colonies usually appear in 7 to 10 days. The presence of intraerythrocytic bacilli in a Giemsa-stained thin blood smear provides rapid confirmation of acute bartonellosis. Serologic tests have been developed in research laboratories but are not widely available.[47,49]

Treatment

Standard treatment consists of chloramphenicol (2 g/day p.o. or I.V.) for at least 1 week. Alternative agents include tetracycline (or doxycycline) or ampicillin.[51] Without treatment, mortality is high. Survivors are temporarily at risk for secondary salmonellosis or toxoplasmosis. Asymptomatic bacteremia is estimated to persist in approximately 15% of survivors.

In untreated cases, resolution of primary infection may be followed weeks to months later by the onset of verruca peruana. This manifestation of chronic bartonellosis consists of recurrent crops of nodular skin lesions, which may ulcerate and bleed. Mucosal and visceral lesions may also occur. Histopathologic examination of active lesions demonstrates neovascular proliferation and occasional bacilli in the interstitium.[47,49]

TRENCH FEVER

Trench fever is caused by infection with *B. quintana,* which is transmitted to humans via feces of infected human body lice (*Pediculus humanus*).[52] Epidemics of the disease occurred in troops engaged in trench warfare in World War I and in World War II troops deployed in crowded and unsanitary conditions.

Diagnosis

The incubation period of classic trench fever was 3 to 38 days, and illness was typically characterized by the abrupt onset of fever, associated with headache, dizziness, conjunctival injection, myalgias, arthralgias, hepatosplenomegaly, a transient maculopapular rash, leukocytosis, and albuminuria. The duration of acute illness was generally 3 to 5 days, and relapses were not uncommon. Less commonly, unremitting fever would persist for 2 to 6 weeks.[47,49]

Treatment

Much of the clinical experience with trench fever predates the antibiotic era. It is not known whether one regimen is superior to another.

BACTEREMIA AND ENDOCARDITIS

Recently, a syndrome termed urban trench fever was recognized in urban homeless persons (usually with a history of alcoholism and malnutrition) and in HIV-infected individuals.[47,49,53,54] In the urban homeless population, the syndrome is caused by *B. quintana* bacteremia and is often associated with apparently culture-negative endocarditis.[47,49,53-56] In the HIV-infected population, the bacteremia can be caused by either *B. quintana* or *B. henselae,* and concomitant bacillary angiomatosis (see below) is often present.[47,49,57,58] As with classic trench fever, the human body louse is the vector for transmission of *B. quintana* (the animal reservoir remains unknown); cats transmit *B. henselae.* Clinical manifestations of urban trench fever include fever, malaise, anemia, thrombocytopenia, splenomegaly, and cardiac murmurs.[47,49,53-56]

Diagnosis

Diagnosis is based on the isolation of *B. henselae* or *B. quintana* from blood, which is best accomplished using lysis-centrifugation blood culture. Growth of those organisms is slow, and incubation for more than 30 days may be required. Acridine orange staining can be used to enhance detection of *Bartonella* organisms in culture. An antibody titer of 1:1,600 or higher against either *B. henselae* or *B. quintana* also strongly suggests bartonellosis as the cause of endocarditis in a patient with valvular vegetations.[47,49]

Treatment

For bacteremia without endocarditis, a 14-day regimen of either a macrolide or a tetracycline (e.g., doxycyline, 100 mg p.o., b.i.d.) is recommended. No definitive treatment regimen has been established for endocarditis. Surgical valve replacement

plus aminoglycoside therapy has been used. An alternative approach has been to give prolonged (4 to 6 months or more) combination therapy with a fluoroquinolone plus either rifampin or a macrolide. Patients treated with that regimen should be followed carefully for clinical deterioration that would necessitate valve replacement[47,49,54-56] [*see Table 2*].

CAT-SCRATCH DISEASE

Although *Afipia felis* was first identified as the causative agent of cat-scratch disease , it is now clear that *B. henselae* causes the overwhelming majority of cases.[47,49,59] The domestic cat serves as the major reservoir of *B. henselae*, which is transmitted to humans by a scratch or bite—usually from a kitten or feral cat—or possibly via infected cat fleas.[47,49] The incubation period is generally 3 to 10 days.

Typically, a primary cutaneous papule or pustule develops at the site of inoculation. Regional lymphadenopathy follows in approximately 90% of cases and persists for 10 to 120 days. Lymphadenopathy is most common at axillary, cervical, and submandibular sites. Less commonly, epitrochlear, inguinal, femoral, or supraclavicular nodes may be involved. Approximately 10% of involved nodes suppurate spontaneously. Low-grade fever occurs in approximately 50% of patients and lasts for several days. Less frequent symptoms and signs include malaise, fatigue, headache, and rash.[47,49]

Relatively common presentations of atypical cat-scratch disease are Parinaud oculoglandular syndrome—a self-limited syndrome of granulomatous conjunctivitis and preauricular lymphadenitis—and encephalitis, which occurs primarily in children.[47,49,60,61] About one half of patients with cat-scratch disease encephalitis experience seizures. Although full recovery from cat-scratch disease encephalitis may require weeks to months, the overall prognosis is good, with few patients experiencing permanent neurologic sequelae.[47,49,61] Other atypical manifestations of cat-scratch disease include self-limited granulomatous hepatitis or splenitis, atypical pneumonitis, osteitis, and neuroretinitis.[47,49,62-66]

Most cases of cat-scratch disease are diagnosed presumptively, on the basis of the clinical presentation and a history of exposure to cats. The differential diagnosis of typical cat-scratch disease includes tularemia, mycobacterial infection, plague, brucellosis, mononucleosis, syphilis, lymphogranuloma venereum, sporotrichosis, histoplasmosis, toxoplasmosis, and lymphoma.

Serologic confirmation of cat-scratch disease is usually not required. When clinically warranted, either immunofluorescence antibody testing or ELISA, available through commercial clinical diagnostic laboratories, can be used to detect antibodies against *B. henselae.* Those serologic studies have essentially supplanted the cat-scratch disease skin test in clinical practice. In biopsy specimens obtained from lymph nodes, skin, or conjunctiva, cat-scratch disease can be confirmed by the detection of small, pleomorphic bacilli in sections prepared with a Warthin-Starry or Brown-Hopps stain.[47-49] PCR technology has been used to detect *B. henselae* in clinical specimens but is currently not widely available for routine clinical diagnosis.[47-49,67]

In immunocompetent patients, both typical and atypical CSD generally resolve spontaneously without antimicrobial therapy. Most *B. henselae* isolates are susceptible in vitro to a wide variety of antimicrobials, including β-lactams, tetracyclines, macrolides, aminoglycosides, fluoroquinolones, vancomycin, and rifampin. However, clinical response does not correlate well with susceptibility testing for *B. henselae*.[47,49,68] Although a variety of agents have been reported to be effective in anecdotal reports and case series (including rifampin, gentamicin, trimethoprim-sulfamethoxazole, doxycycline, ciprofloxacin, and ofloxacin), only azithromycin (500 mg p.o. for 1 day, then 250 mg p.o. daily for 4 days) has been shown to accelerate resolution of typical cat-scratch disease lymphadenopathy in a placebo-controlled, double-blind clinical trial.[69]

BACILLARY ANGIOMATOSIS

Bacillary angiomatosis (epithelioid angiomatosis, bacillary epithelioid angiomatosis) is a form of neovascular proliferation associated with infection by either *B. henselae* or *B. quintana.* It was first described as a disease of the skin and regional lymph nodes of HIV-infected persons. Subsequently, bacillary angiomatosis was documented in a large variety of tissues, including liver, spleen, bone, brain, respiratory tract, gastrointestinal tract, and uterine cervix. It has also been found, albeit rarely, in other immunocompromised hosts and in apparently immunocompetent patients. In HIV-infected persons, it is a relatively late opportunistic infection, usually occurring after the $CD4^+$ lymphocyte count has declined to less than 100 cells/μl.[47,49,57,58]

Cutaneous disease is the most frequently recognized manifestation of bacillary angiomatosis. Lesions generally appear in crops and can have verrucous, papular, or pedunculated features. They often have an erythematous base and vascular appearance but may occasionally be dry, scaly, hyperkeratotic, or plaquelike. Painful subcutaneous nodules and osseous lesions can develop. Regional lymphadenopathy is common, but involved nodes rarely suppurate or drain.[47,49]

Peliosis hepatis, consisting of venous lakes in the hepatic parenchyma, is a relatively common manifestation of bacillary angiomatosis. Splenitis can also occur. Hepatosplenomegaly is often present, and an increased serum alkaline phosphatase level is a common laboratory finding. Peliotic spaces in the liver and spleen appear as hypodense lesions on abdominal computed tomography. Thrombocytopenia and pancytopenia may develop, as well as concomitant bacteremia or endocarditis, in patients with bacillary peliosis.[47,49,57,58]

The lesions of bacillary angiomatosis may be indistinguishable clinically from Kaposi sarcoma, and biopsy is usually necessary to distinguish between the two disorders and guide management. Both hematoxylin-eosin staining, which differentiates bacillary angiomatosis from Kaposi sarcoma, and Warthin-Starry staining, which detects *Bartonella* bacilli, should be performed on biopsy specimens. Lesions of bacillary angiomatosis are characterized by lobular proliferation of endothelial-lined capillaries; the endothelial cells are protuberant, unlike the spindle-shaped endothelial cells in Kaposi sarcoma. Recovery of *Bartonella* organisms by culture from lesions of bacillary angiomatosis is difficult.[47-49]

The recommended first-line therapy for bacillary angiomatosis in HIV-infected patients is either erythromycin (500 mg p.o., q.i.d.) or doxycycline (100 mg p.o., b.i.d.). Alternative agents include clarithromycin (500 mg p.o., b.i.d.), azithromycin (250 mg p.o. daily), and ciprofloxacin (500 to 750 mg p.o., b.i.d.). For severe cases, consideration should be given to combination therapy with doxycycline (100 mg p.o. or I.V., b.i.d.) plus rifampin (300 mg p.o., b.i.d.). Bacillary angiomatosis usually responds rapidly to either erythromycin or doxycycline therapy. A Jarisch-Herxheimer reaction may occur with the initiation of antimicrobial therapy. Patients who have only cutaneous lesions

should receive at least 2 months of treatment; in those with osteomyelitis or peliosis hepatis, therapy should be continued for a minimum of 4 months. Relapses can occur after discontinuance of therapy. HIV-infected patients who experience relapses should be treated lifelong with either a tetracycline or a macrolide.[47,49]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Corbel MJ: Brucellosis: an overview. Emerg Infect Dis 3:213, 1997

2. Leggiadro RJ: The threat of biological terrorism: a public health and infection control reality. Infect Control Hosp Epidemiol 21:53, 2000

3. Solera J, Martinez-Alfaro E, Espinosa A: Recognition and optimum treatment of brucellosis. Drugs 53:245, 1997

4. Sanford JP: *Brucella* pneumonia. Semin Respir Infect 12:24, 1997

5. Memish Z, Mah MW, Al Mahmoud S, et al: *Brucella* bacteraemia: clinical and laboratory observations in 160 patients. J Infect 40:59, 2000

6. Fiori PL, Mastrandrea S, Rappelli P, et al: *Brucella abortus* infection acquired in microbiology laboratories. J Clin Microbiol 38:2005, 2000

7. Summary of notifiable diseases, United States – 1999. MMWR Morb Mortal Wkly Rep 48:1, 2001

8. Doyle TJ, Bryan RT: Infectious disease morbidity in the US region bordering Mexico, 1990-1998. J Infect Dis 182:1502, 2000

9. Ertem M, Kurekci AE, Aysev D, et al: Brucellosis transmitted by bone marrow transplantation. Bone Marrow Transplant 26:225, 2000

10. Gad El-Rab MO, Kambal AM: Evaluation of a *Brucella* enzyme immunoassay test (ELISA) in comparison with bacteriological culture and agglutination. J Infect 36:197, 1998

11. Morata P, Queipo-Ortuno MI, Reguera JM, et al: Posttreatment follow-up of brucellosis by PCR assay. J Clin Microbiol 37:4163, 1999

12. Solera J, Espinosa A, Martinez-Alfaro E, et al: Treatment of human brucellosis with doxycycline and gentamicin. Antimicrob Agents Chemother 41:80, 1997

13. Agalar C, Usubuntun S, Turkyilmaz R: Ciprofloxacin and rifampicin versus doxycycline and rifampicin in the treatment of brucellosis. Eur J Clin Microbiol Infect Dis 18:535, 1999

14. Solera J, Lozano E, Martinez-Alfaro E, et al: Brucellar spondylitis: review of 35 cases and literature survey. Clin Infect Dis 29:1440, 1999

15. Uddin MJ, Sanyal SC, Mustafa AS, et al: The role of aggressive medical therapy along with early surgical intervention in the cure of *Brucella* endocarditis. Ann Thorac Cardiovasc Surg 4:209, 1998

16. Franz DR, Jahrling PB, Friedlander AM, et al: Clinical recognition and management of patients exposed to biological warfare agents. JAMA 278:399, 1997

17. Boyce JM: Recent trends in the epidemiology of tularemia in the United States. J Infect Dis 131:197, 1975

18. Spach DH, Liles WC, Campbell GL, et al: Tick-borne diseases in the United States. N Engl J Med 329:936, 1993

19. Summary of notifiable diseases, United States, 1993. MMWR Morb Mortal Wkly Rep 42:i, 1, 1994

20. Taylor JP, Istre GR, McChesney TC, et al: Epidemiologic characteristics of human tularemia in the southwest-central states, 1981-1987. Am J Epidemiol 133:1032, 1991

21. Guerrant RL, Humphries MK Jr, Butler JE, et al: Tickborne oculoglandular tularemia: case report and review of seasonal and vectorial associations in 106 cases. Arch Intern Med 136:811, 1976

22. Markowitz LE, Hynes NA, de la Cruz P, et al: Tick-borne tularemia: an outbreak of lymphadenopathy in children. JAMA 254:2922, 1985

23. Liles WC, Burger RJ: Tularemia from domestic cats. West J Med 158:619, 1993

24. Capellan J, Fong IW: Tularemia from a cat bite: case report and review of feline-associated tularemia. Clin Infect Dis 16:472, 1993

25. Evans ME, Gregory DW, Schaffner W, et al: Tularemia: a 30-year experience with 88 cases. Medicine (Baltimore) 64:251, 1985

26. Gill V, Cunha BA: Tularemia pneumonia. Semin Respir Infect 12:61, 1997

27. Penn RL, Kinasewitz GT: Factors associated with a poor outcome in tularemia. Arch Intern Med 147:265, 1987

28. Fulop M, Leslie D, Titball R: A rapid, highly sensitive method for the detection of *Francisella tularensis* in clinical samples using the polymerase chain reaction. Am J Trop Med Hyg 54:364, 1996

29. Johansson A, Berglund L, Eriksson U, et al: Comparative analysis of PCR versus culture for diagnosis of ulceroglandular tularemia. J Clin Microbiol 38:22, 2000

30. Junhui Z, Ruifu Y, Jianchun L, et al: Detection of Francisella tularensis by the polymerase chain reaction. J Med Microbiol 45:477, 1996

31. Pike RM: Laboratory-associated infections: summary and analysis of 3921 cases. Health Lab Sci 13:105, 1976

32. Mason WL, Eigelsbach HT, Little SF, et al: Treatment of tularemia, including pulmonary tularemia, with gentamicin. Am Rev Respir Dis 121:39, 1980

33. Risi GF, Pombo DJ: Relapse of tularemia after aminoglycoside therapy: case report and discussion of therapeutic options. Clin Infect Dis 20:174, 1995

34. Limaye AP, Hooper CJ: Treatment of tularemia with fluoroquinolones: two cases and review. Clin Infect Dis 29:922, 1999

35. Johansson A, Berglund L, Gothefors L, et al: Ciprofloxacin for treatment of tularemia in children. Pediatr Infect Dis J 19:449, 2000

36. Ikaheimo I, Syrjala H, Karhukorpi J, et al: In vitro antibiotic susceptibility of *Francisella tularensis* isolated from humans and animals. J Antimicrob Chemother 46:287, 2000

37. Burke DS: Immunization against tularemia: analysis of the effectiveness of live *Francisella tularensis* vaccine in prevention of laboratory-acquired tularemia. J Infect Dis 135:55, 1977

38. Waag DM, Sandstrom G, England MJ, et al: Immunogenicity of a new lot of *Francisella tularensis* live vaccine strain in human volunteers. FEMS Immunol Med Microbiol 13:205, 1996

39. Perry RD, Fetherston JD: *Yersinia pestis* – etiologic agent of plague. Clin Microbiol Rev 10:35, 1997

40. Inglesby TV, Dennis DT, Henderson DA, et al: Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. JAMA 283:2281, 2000

41. Fatal human plague – Arizona and Colorado, 1996. MMWR Morb Mortal Wkly Rep 46:617, 1997

42. Summary of notifiable diseases, United States – 1999. MMWR Morb Mortal Wkly Rep 48:1, 2001

43. Gage KL, Dennis DT, Orloski KA, et al: Cases of cat-associated human plague in the Western US, 1997-1998. Clin Infect Dis 30:893, 2000

44. Cleri DJ, Vernaleo JR, Lombardi LJ, et al: Plague pneumonia disease caused by *Yersinia pestis*. Semin Respir Infect 12:12, 1997

45. Wong JD, Barash JR, Sandfort RF, et al: Susceptibilities of *Yersinia pestis* strains to 12 antimicrobial agents. Antimicrob Agents Chemother 44:1995, 2000

46. Prevention of plague: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 45:1, 1996

47. Spach DH, Koehler JE: *Bartonella*-associated infections. Infect Dis Clin North Am 12:137, 1998

48. Doern GV: Detection of selected fastidious bacteria. Clin Infect Dis 30:166, 2000

49. Maguina C, Gotuzzo E: Bartonellosis: new and old. Infect Dis Clin North Am 14:1, 2000

50. Kosek M, Lavarello R, Gilman RH, et al: Natural history of infection with *Bartonella bacilliformis* in a nonendemic population. J Infect Dis 182:865, 2000

51. Sobraques M, Maurin M, Birtles RJ, et al: In vitro susceptibilities of four *Bartonella bacilliformis* strains to 30 antibiotic compounds. Antimicrob Agents Chemother 43:2090, 1999

52. Raoult D, Roux V: The body louse as a vector of reemerging human diseases. Clin Infect Dis 29:888, 1999

53. Brouqui P, Lascola B, Roux V, et al: Chronic *Bartonella quintana* bacteremia in homeless patients. N Engl J Med 340:184, 1999

54. Ohl ME, Spach DH: *Bartonella quintana* and urban trench fever. Clin Infect Dis 31:131, 2000

55. Baorto E, Payne RM, Slater LN, et al: Culture-negative endocarditis caused by *Bartonella henselae*. J Pediatr 132:1051, 1998

56. Brouqui P, Raoult D: Endocarditis due to rare and fastidious bacteria. Clin Microbiol Rev 14:177, 2001

57. Koehler JE, Sanchez MA, Garrido CS, et al: Molecular epidemiology of bartonella infections in patients with bacillary angiomatosis-peliosis. N Engl J Med 337:1876, 1997

58. Gasquet S, Maurin M, Brouqui P, et al: Bacillary angiomatosis in immunocompromised patients. AIDS 12:1793, 1998

59. Giladi M, Avidor B, Kletter Y, et al: Cat scratch disease: the rare role of *Afipia felis*. J Clin Microbiol 36:2499, 1998

60. Grando D, Sullivan LJ, Flexman JP, et al: *Bartonella henselae* associated with Parinaud's oculoglandular syndrome. Clin Infect Dis 28:1156, 1999

61. Armengol CE, Hendley JO: Cat-scratch disease encephalopathy: a cause of status epilepticus in school-age children. J Pediatr 134:635, 1999

62. Jacobs RF, Schutze GE: *Bartonella henselae* as a cause of prolonged fever and fever of unknown origin in children. Clin Infect Dis 26:80, 1998

63. Ormerod LD, Skolnick KA, Menosky MM, et al: Retinal and choroidal manifestations of cat-scratch disease. Ophthalmology 105:1024, 1998

64. Ormerod LD, Dailey JP: Ocular manifestations of cat-scratch disease. Curr Opin Ophthalmol 10:209, 1999

65. Reed JB, Scales DK, Wong MT, et al: *Bartonella henselae* neuroretinitis in cat scratch disease: diagnosis, management, and sequelae. Ophthalmology 105:459, 1998

66. Solley WA, Martin DF, Newman NJ, et al: Cat scratch disease: posterior segment manifestations. Ophthalmology 106:1546, 1999

67. Jensen WA, Fall MZ, Rooney J, et al: Rapid identification and differentiation of *Bartonella* species using a single-step PCR assay. J Clin Microbiol 38:1717, 2000

68. Rolain JM, Maurin M, Raoult D: Bactericidal effect of antibiotics on *Bartonella* and *Brucella* spp.: clinical implications. J Antimicrob Chemother 46:811, 2000

69. Bass JW, Freitas BC, Freitas AD, et al: Prospective randomized double blind placebo-controlled evaluation of azithromycin for treatment of cat-scratch disease. Pediatr Infect Dis J 17:447, 1998

XII MYCOPLASMA INFECTIONS

R. Doug Hardy, m.d.

Mycoplasmas are the smallest known free-living organisms. Their size—150 to 350 nm—is closer to that of viruses than of bacteria. Unlike viruses, however, mycoplasmas are able to grow in cell-free media and possess both RNA and DNA. Notably, they lack a cell wall and are bound by a cell membrane. The absence of a rigid cell wall explains many of the biologic properties of mycoplasmas, including resistance to β-lactam antibiotics and marked pleomorphism among individual cells.

Mycoplasmas are prokaryotes of the class Mollicutes. The entire genomes of many *Mycoplasma* species have been sequenced and have been found to be among the smallest of prokaryotic genomes. The characterization of *M. genitalium,* which has only 580,070 base pairs and 468 predicted proteins, has helped define the minimal set of genes necessary for cellular life.[1,2] The absence of genes related to the synthesis of amino acids, fatty acid metabolism, and cholesterol necessitates a parasitic or saprophytic dependence on the host for exogenous nutrients, such as nucleic acid precursors, amino acids, fatty acids, and sterols.

At least 13 *Mycoplasma* species, two *Acholeplasma* species, and one *Ureaplasma* species have been isolated from humans with varying frequency; most of these species are thought to be normal inhabitants of oral and urogenital mucous membranes.[3] Only three species, *M. pneumoniae, M. hominis,* and *U. urealyticum,* have been shown conclusively to be pathogenic in humans. *M. pneumoniae* is the species that has been most clearly demonstrated to cause disease in humans. The respiratory tract is its primary site of involvement. *M. hominis* and *U. urealyticum* are associated with a variety of genitourinary tract disorders and neonatal infections. Evidence also implicates a fourth species, *M. genitalium,* as a cause of disease in humans. Rarely, other *Mycoplasma* species may cause disease in immunocompromised persons.

Mycoplasma pneumoniae

EPIDEMIOLOGY

M. pneumoniae is a major cause of community-acquired respiratory illness in children and adults, both in terms of clinical severity and numbers affected. *M. pneumoniae* is often grouped with *Chlamydophila pneumoniae* and *Legionella* species as being among the most important bacterial causes of so-called atypical community-acquired pneumonia. However, it is now known that atypical pneumonia caused by these pathogens cannot be reliably differentiated from pneumonia caused by typical community-acquired pathogens, such as *Streptococcus pneumoniae,* on the basis of clinical manifestations, general laboratory tests, or radiographic findings.[4,5] Hence, many experts believe that the term atypical is inappropriate and misleading. It is important to note, however, that the etiologic diagnosis of infection by these pathogens requires advanced methods of detection—methods that go beyond Gram stain or routine culture. *Mycoplasma* pneumonia has also been referred to as Eaton agent pneumonia, primary atypical pneumonia, and walking pneumonia.

M. pneumoniae was first isolated in the early 1940s by Eaton from a patient with primary atypical pneumonia; it was proved to be a cause of clinical disease in humans in the 1960s.[5,6] *M. pneumoniae* is the only *Mycoplasma* species that has clearly been established to be a cause of human respiratory tract disease, although *M. genitalium* has been isolated from the respiratory tract of patients with mixed infections involving *M. pneumoniae,* an organism with which *M. genitalium* can be confused.[7] *M. pneumoniae* is known to cause many acute respiratory syndromes in humans, including pharyngitis, tracheobronchitis, reactive airway disease (wheezing), and community-acquired pneumonia. It is likely that the incidence of upper respiratory tract illness is 10 to 20 times that of pneumonia. Infection is spread from one person to another by respiratory droplets expectorated during coughing; infection results in clinically apparent disease in the majority of cases.[5] The incubation period for *M. pneumoniae* is 2 to 4 weeks; hence, the course of infection in a specific population (e.g., within a family or an institution) may last several weeks [*see Figure 1*].[8] Intrafamily attack rates may be as high as 84% in children and 41% in adults. Outbreaks of *M. pneumoniae*–related illness often occur in institutional settings such as military bases, boarding

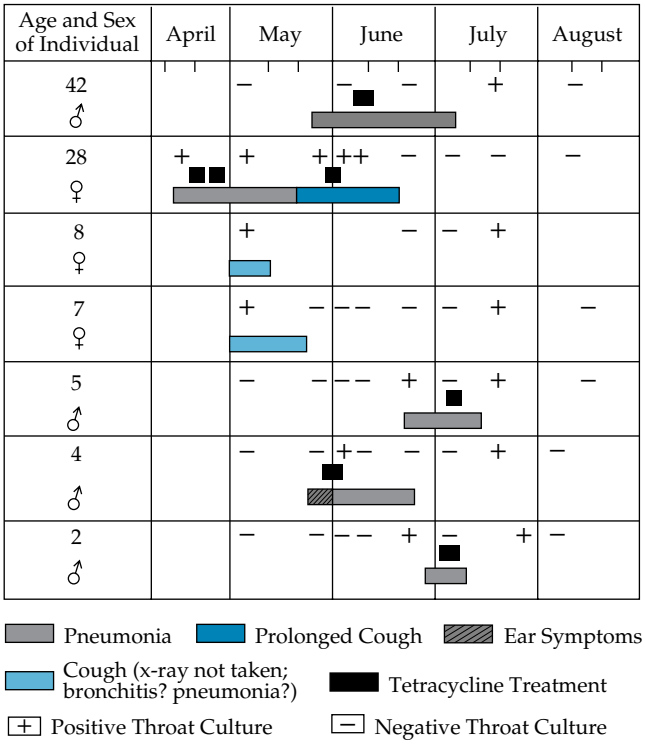


Figure 1 **During a 5-month period, a *Mycoplasma pneumoniae* infection spread through seven members of a family. Positive throat cultures were obtained in all seven individuals, but clinically documented pneumonia was observed in only five. Pneumonia was treated with tetracycline. In the other two infected individuals, coughs developed during the period in which their throat cultures were positive, suggesting the possibility of *Mycoplasma*-induced bronchitis or pneumonia; however, radiologic confirmation of the diagnosis was not obtained. A prolonged cough followed the episode of pneumonia in one patient, and ear symptoms were observed in another. Positive throat cultures were obtained in six of seven patients after cessation of therapy and disappearance of symptoms.**

schools, and summer camps. Infections tend to be endemic, with sporadic epidemics every 4 to 7 years, without seasonal preponderance.[5]

The importance of *M. pneumoniae* as a cause of community-acquired pneumonia has become more evident with the development of advanced techniques for identifying this pathogen. For community-acquired pneumonia in adults, *M. pneumoniae* is the most frequently detected atypical organism. Analysis of 13 studies of community-acquired pneumonia published since 1995 (which included 6,207 ambulatory and hospitalized adults) showed that the overall prevalence of *M. pneumoniae* was 22.7%; by comparison, the prevalence of *C. pneumoniae* was 11.7%, and that of *Legionella* species was 4.6%.[9] It must be acknowledged that these studies represent varying populations and used varying methods to establish the etiologies of community-acquired pneumonia.[9]

In community-acquired pneumonia, *M. pneumoniae* often occurs as part of a mixed infection. In a 1-year prospective study of 346 consecutive adults admitted with community-acquired pneumonia, 64% of those with *M. pneumoniae* infections were coinfected with other agents—most often with *S. pneumoniae*, followed by *C. pneumoniae* and *Legionella* species.[10] In a study of *C. pneumoniae* community-acquired pneumonia in adults, 35.5% had a mixed infection; *M. pneumoniae* was the second most common copathogen, after *S. pneumoniae*.[11] Review of epidemiologic studies revealed the presence of at least one other pathogen in 33% to 64% of *M. pneumoniae* pneumonia.[9,11] *S. pneumoniae* with *M. pneumoniae* or *C. pneumoniae* appears to be the most common combinations of agents in mixed pneumonia, and these combinations have been given the most attention in clinical investigations.[9,12,13] Further study is required to define the clinical significance of *M. pneumoniae* coinfections, to determine whether infection with *M. pneumoniae* predisposes patients to invasion by other respiratory pathogens, and to establish whether the etiologic agents of mixed infections have an additive or synergistic clinical impact.

PATHOGENESIS, PATHOLOGY, AND IMMUNOLOGY

Classically, mycoplasmas act as extracellular parasites. The pathogenicity of *M. pneumoniae* depends on its extracellular attachment and the initiation of injury to the membrane of the host cell.[14] *M. pneumoniae* penetrates mucociliary secretions by means of gliding motility and adheres firmly to the surface of respiratory epithelial cells.[15] *M. pneumoniae* attaches to ciliated respiratory epithelial cells at the base of the cilia by means of a complex terminal organelle at one end of the elongated organism. This cytodherence is mediated by interactive adhesins and accessory proteins clustered at the tip organelle; it is not limited to epithelial cells or to human-derived tissues.[16,17] Multiple unlinked, incomplete copies of the genes that encode these adhesins constitute approximately 5% of the *Mycoplasma* genome; it has been proposed that recombination between gene copies generates antigenically variant proteins that avoid host immune recognition.[17] One or more sialated glycoproteins act as one type of receptor on the surface of the epithelium.[18,19]

M. pneumoniae causes physiologic and cytolytic injury to the host cells (e.g., ciliated respiratory epithelium), in part by the production of hydrogen peroxide.[20,21] Inhibition of host catalase by *M. pneumoniae*–derived hydrogen peroxide and superoxide anion, followed by oxidation of host membrane lipids and proteins, may then result in cumulative local cytotoxic effects.[22,23] Other mechanisms of physiologic and cytolytic injury are also under investigation. Lung biopsies from patients with *M. pneumoniae* community-acquired pneumonia reveal an inflammatory process involving the trachea, bronchioles, and peribronchial tissue with a monocytic infiltrate coinciding with a luminal exudate of polymorphonuclear leukocytes.[24,25]

Although *M. pneumoniae* is classically thought of as an extracellular organism, it (as well as some other mycoplasmas, including *M. penetrans, M. fermentans* var. *incognitos*, and *M. gallisepticum*) can live and replicate intracellularly in human cells.[26-29] Residence of *M. pneumoniae* in an intracellular reservoir, perhaps protected from the host's immunologic response and from antibiotics, might explain the organism's apparent ability to establish chronic infection, such as occurs in asthmatic patients.[30,31]

Mycoplasmas have been shown to activate natural killer cells, macrophages, and complement; to stimulate the proliferation of B cells and T cells; and to induce expression of major histocompatibility complex class I and II molecules.[32] Because mycoplasmas lack a cell wall, they also lack cell wall–derived stimulators of the innate immune system, such as lipopolysaccharide, lipoteichoic acid, and murein (peptidoglycan) fragments. However, lipoproteins from various *Mycoplasma* species appear to have potent inflammatory properties. Three lipoproteins/lipopeptides of *M. fermentans* origin—macrophage-activating lipopeptide–2 (MALP-2), P48, and M161Ag (identical to MALP-404)—have been shown to have the ability to modulate the host immune system through a Toll-like receptor 2 (TLR2) and TLR6 pathway and through β_2 integrin.[33-35] MALP-2 is a 2-kd lipopeptide from the cell membrane of *M. fermentans* that has been shown to induce cytokines and chemokines in human monocytes.[36] In *M. pneumoniae,* the genes for more than 30 different lipoproteins have been detected.[1] Activation of TLR2 on macrophages by *M. pneumoniae*–derived lipoproteins with resultant production of tumor necrosis factor–α and nitric oxide has been reported.[37]

Studies in mice with severe combined immunodeficiency support the hypotheses that innate immunity provides most of the defense against *Mycoplasma* infection in the lungs, that humoral immunity provides protection against dissemination of *Mycoplasma* infection, and that cellular immunity may play a role in exacerbating mycoplasmal lung disease.[38] Also supportive are the findings that patients with T cell deficiencies do not have more severe pneumonia from *M. pneumoniae* and that patients with humoral immunodeficiencies do not have more severe lung disease from *M. pneumoniae* than immunocompetent patients in the early stages of infection but do have more disseminated disease (e.g., arthritis, meningitis, and osteomyelitis).[39-41]

M. pneumoniae infection in humans provokes a specific immune response that results in the production of secretory IgA and circulating IgG antibodies. Naturally acquired humoral immunity is associated with partial protection against reinfection; in particular, the immunity that follows more severe *M. pneumoniae* infections (e.g., pneumonia) is more protective and longer lasting than that which follows mild infection.[42] Bona fide second attacks of *M. pneumoniae* pneumonia have been reported occasionally, however, both in immunocompetent and in immunodeficient patients. The ability of *M. pneumoniae* and other mycoplasmas to trigger autoimmune responses is an area of active research.[43]

The role of cytokines and chemokines in *M. pneumoniae* infection has been investigated in vitro, in animal models, and in humans.[44-47] Evidence indicates that various cytokines (including inflammatory, anti-inflammatory, T helper type 1, and T helper type 2 cytokines) as well as chemokines have pivotal roles in modulating *M. pneumoniae* disease. However, a concerted theory

on the pathogenic and relative importance of these various modulators is lacking.

DIAGNOSIS

Acute Clinical Manifestations

M. pneumoniae infections have an incubation period of 2 to 4 weeks in naturally occurring cases, but the incubation period is shorter in experimental infection.[5] Approximately 20% of *M. pneumoniae* infections are asymptomatic.[48] Clinically apparent acute *M. pneumoniae* infections generally manifest as pharyngitis, tracheobronchitis, reactive airway disease (wheezing), or a nonspecific upper respiratory tract syndrome. Pneumonia, however, is the best-described clinical manifestation of *M. pneumoniae* infection and occurs in 3% to 13% of infected persons.[48]

The onset of pneumonia is usually gradual (occurring over several days) but may be more abrupt.[49] Although *Mycoplasma* pneumonia may begin with a sore throat, the most common presenting symptom is cough. The cough is typically nonproductive, but some patients produce purulent sputum; in rare instances, the sputum is blood streaked.[48] Headache, malaise, chills, and a temperature of 37.8° to 40.0° C (100° to 104° F) are present in the majority of patients. In a smaller percentage, nasal symptoms occur. Symptoms and signs of pneumonia are not useful for differentiating *M. pneumoniae* pneumonia from other types of community-acquired pneumonia.[4,48,50,51]

Symptoms related to the ears or tympanic membrane findings on physical examination may be noted in approximately 6% of patients with *M. pneumoniae* pneumonia; 1% may have bullous myringitis. Because direct cultures rarely if ever grow *M. pneumoniae,* there is little evidence to support the old belief that this organism is an important cause of otitis media, with or without bullous myringitis.[52]

Wheezes or rales are present in approximately 80% of patients with *Mycoplasma* pneumonia; bronchial breath sounds are noted less frequently. In many patients, however, pneumonia can be diagnosed only on chest x-ray. Pleural rubs are rare, as are pleural effusions of sufficient magnitude to be detected on physical examination.[48]

Symptoms usually resolve within 2 to 3 weeks after the onset of illness. In about 20% of patients, symptoms and radiologic abnormalities persist for at least a month. Although *M. pneumoniae* pneumonia is generally self-limited, appropriate antimicrobial therapy can significantly shorten the duration of clinical illness.[53] In some patients, long-term recurrent wheezing may follow the resolution of acute pneumonia.[54,55] Relapse and recurrence of pneumonia are infrequently reported.

Pulmonary complications of *M. pneumoniae* pneumonia may occur and, on rare occasions, result in death. Fulminant disease may be more common in young, previously healthy adults. Multilobar involvement, lung abscess, massive pleural effusion, acute respiratory distress syndrome, and diffuse interstitial fibrosis have been described. At particular risk for these complications are patients with Down syndrome, sickle cell disease, and other immunodeficient states.[56-59] *M. pneumoniae* does not seem to play an important role in lower respiratory tract infections in either children or adults with HIV infection.[60]

Chronic Infection and Wheezing

In many patients, *M. pneumoniae* can be cultured from the respiratory tract for up to several months after clinical and radiologic resolution of acute pneumonia.[7] There have been no controlled studies of such chronic infections using methods more sensitive than culture, such as polymerase chain reaction (PCR). *M. pneumoniae* apparently can be a long-term respiratory tract pathogen associated with recurrent wheezing, and it may contribute to the severity of chronic asthma.[54,55]

Extrapulmonary Manifestations

A variety of extrapulmonary manifestations may develop during the course of acute *M. pneumoniae* respiratory tract infection. The most important of these are neurologic, dermatologic, cardiac, rheumatologic, and hematologic in nature.[3,56,61] Overall, these manifestations are uncommon, given the frequency of *Mycoplasma* infection. The mechanisms of these complications appear to include disseminated infection (especially in patients with humoral immunodeficiencies); autoimmune phenomena; and, possibly, toxin production.[62] Notably, respiratory disease is absent in many patients with extrapulmonary *M. pneumoniae* disease.[63]

Neurologic disease is reportedly the most common nonpulmonary manifestation of *M. pneumoniae* infection.[63-67] A wide spectrum of neurologic manifestations has been reported. The most common ones are meningoencephalitis, encephalitis, polyradiculopathy (including Guillain-Barré syndrome), and aseptic meningitis; less common neurologic manifestations include cranial neuropathy, acute psychosis, cerebellar ataxia, acute demyelinating encephalomyelitis, cerebrovascular thromboembolic events, and transverse myelitis. In a prospective 5-year study of acute encephalitis in children, *M. pneumoniae* was determined to be the likely cause in 11 of 159 patients (6.9%). Findings on clinical examination, cerebrospinal fluid analysis, electroencephalography, and neuroimaging in patients with *M. pneumoniae* encephalitis are indistinguishable from those in patients with viral encephalitis. Long-term sequelae, such as seizure disorders, hemiparesis, expressive dysphasia, dysarthria, and truncal ataxia, have been described in 48% to 64% of patients after *M. pneumoniae* encephalitis. Preliminary evidence indicates that direct invasion of the central nervous system by *M. pneumoniae* may be the pathogenic mechanism in encephalitis cases in which prodromal symptoms are of short duration (< 5 days), whereas immunologic phenomena may be responsible for cases with a more prolonged prodrome.[63,65-67] The roles of antimicrobial therapy and immune-modulating therapy in the treatment of *M. pneumoniae* neurologic disease are unknown.

Skin eruptions described with *M. pneumoniae* infection include erythematous (macular or maculopapular), vesicular, bullous, petechial, and urticarial rashes; in one study, 17% of patients with *M. pneumoniae* pneumonia had an exanthem.[68] Erythema multiforme major (Stevens-Johnson syndrome) is the most clinically significant skin eruption associated with *M. pneumoniae* infection; it appears to occur more commonly with *M. pneumoniae* than with other infectious agents.[69]

Hematologic manifestations of *Mycoplasma* infection include hemolytic anemia, aplastic anemia, disseminated intravascular coagulation, hypercoagulopathy, and perhaps thrombotic thrombocytopenic purpura.[70-73] Cold agglutinin–induced hemolytic anemia is an uncommon manifestation of *Mycoplasma* pneumonia. Although up to 83% of patients with *M. pneumoniae* infection may have reticulocytosis and a positive direct Coombs test, clinically significant hemolysis is rare.[70] When anemia does occur, it begins suddenly in the second or third week of illness, and the agglutinin titer is high (> 1:512). The process is usually self-limited, lasting a few weeks.

Myocardial disease, pericardial disease, rhabdomyolysis, arthralgias, and arthritis (septic and reactive) have been attributed to *M. pneumoniae*.[3,74-76] Septic arthritis has been described most commonly in hypogammaglobulinemic patients.[41] In addition, hepatitis and pancreatitis have been reported.

General Laboratory Tests and Radiologic Studies

As with clinical findings, general laboratory tests and chest radiography have not been found to be useful for differentiating *Mycoplasma* pneumonia from other types of community-acquired pneumonia.[4,48,50,77] The frequent occurrence of community-acquired pneumonia of mixed etiology may be partly responsible for this finding.[11,78] Measurement of cold agglutinin titers is no longer recommended for the diagnosis of *Mycoplasma* pneumonia, because they are nonspecific and because assays specific for *M. pneumoniae* are now available.

The radiologic abnormalities of *Mycoplasma* pneumonia are varied. The most common pattern on chest radiography is that of peribronchial pneumonia with thickened bronchial markings; streaks of interstitial infiltration; and areas of subsegmental atelectasis, often involving a single lower lobe. Other frequent patterns include platelike atelectasis, nodular infiltration, and hilar adenopathy.[48] Segmental or lobar consolidation is not uncommon. Clinically evident pleural effusions are uncommon in *Mycoplasma* pneumonia, but lateral decubitus views reveal that up to 20% of patients have pleural effusions.

Specific Laboratory Testing

Laboratory diagnosis of an acute *M. pneumoniae* infection can be established (1) by the isolation of *M. pneumoniae* from respiratory tract secretions (oropharyngeal, nasopharyngeal, or pulmonary) with PCR or culture (which requires special media) or (2) by the use of specific serologic tests, such as complement fixation, indirect immunofluorescence, and enzyme immunoassays (EIAs) for specific IgM and IgG antibodies in paired (acute and convalescent) serum samples [*see Table 1*]. Of these methods, EIAs are the most widely used and the most adaptable to the clinical laboratory setting.[79] Serum samples for *M. pneumoniae* serology taken only during the acute phase of illness may not be indicative of infection, because antibodies to *M. pneumoniae* may not develop for 2 weeks or more; therefore, it is important to test both acute and convalescent serum samples for accurate diagnosis. IgM antibodies against *M. pneumoniae* may not be produced during reinfection in older patients (i.e., persons older than 40 years with preexisting anti–*M. pneumoniae* IgG antibodies).[80] In addition, specific IgM can persist for up to a year after acute *M. pneumoniae* infection and so may indicate recent infection rather than acute infection.[79]

M. pneumoniae culture is not recommended for routine diagnosis, because the organism may take weeks to grow and is often difficult to isolate from clinical specimens, so sensitivity is low. PCR allows a rapid and specific diagnosis to be made early in the course of clinical illness. The combination of respiratory tract PCR testing and EIA has been recommended as the most sensitive and rapid approach to the diagnosis of *M. pneumoniae* infection.[81] In children, PCR assays of nasopharyngeal or oropharyngeal samples appear equally effective for the diagnosis of serologically confirmed *M. pneumoniae* pneumonia; however, testing of specimens from both sites is optimal.[82]

TREATMENT

Because the clinical manifestations of *Mycoplasma* pneumonia are not distinctive and laboratory diagnosis is often made retrospectively, treatment is usually empirical, as it is with community-acquired pneumonia in general. The possibility of *M. pneumoniae* infection deserves particular consideration in a patient with community-acquired pneumonia who has failed to respond to treatment with a penicillin or a cephalosporin, because β-lactam antibiotics are ineffective against mycoplasmas. Clinical trials ranging from observational reports to randomized, double-blind, placebo-controlled studies have demonstrated that appropriate antimicrobial therapy significantly decreases the duration of fever, cough, malaise, hospitalization, and radiologic abnormalities in *Mycoplasma* pneumonia.[53] In one study of military trainees with serologically confirmed *M. pneumoniae* pneumonia, treatment with oral erythromycin for 7 days reduced mean duration of fever from 4.2 days to 2.4 days; the duration of hospitalization was reduced from 14.1 days to 7.0 days; and the period in which radiologic abnormalities were present was reduced from 14.8 days to 7.2 days.[83]

Treatment options for acute *M. pneumoniae* infection include macrolides, ketolides, tetracyclines, and most fluoroquinolones (ciprofloxacin and ofloxacin are not recommended because of their high minimum inhibitory concentrations and poor performance in animal studies) [*see Table 2*]. Antimicrobial resistance has not been clinically important in *M. pneumoniae* infections in North America. The optimal antibiotic choice, dosage, and duration are not clear; however, 14 days of therapy is generally recommended.

Even after therapy with appropriate antibiotics, *M. pneumoniae* can still be cultured from respiratory tract secretions.[8,84] The persistence of *M. pneumoniae* in the nasopharynx of children after clarithromycin therapy (given for an acute exacerbation of wheezing accompanying *M. pneumoniae* infection) has been associated with recurrent wheezing.[85] In animal models of *Mycoplasma* pneumonia, treatment with antimycoplasmal agents such as ketolides, quinolones, clarithromycin, and azithromycin also has not eradicated this organism from the respiratory tract.[55] The efficacy of combination antimicrobial therapy for the eradication of *M. pneumoniae* has not been investigated.

Table 1 Accuracy of Diagnostic Tests for *Mycoplasma pneumoniae* Infection[109,110]

Test	*Sensitivity (%)*	*Specificity (%)*	*Comment*
Culture*	≤ 60	100	Not recommended for clinical diagnosis
Polymerase chain reaction (PCR)*	65–90	90–100	Combination of PCR and enzyme immunoassays (EIAs) is optimal for diagnosis
Serologic studies	55–100	55–100	EIAs for *M. pneumoniae* IgM and IgG in paired (acute and convalescent) serum samples is the recommended serologic method

*Using respiratory tract secretions.

Table 2 Antimicrobial Treatment of *Mycoplasma* Infections

Pathogen	*Infection*	*Representative Drugs**
Mycoplasma pneumoniae	Respiratory tract infection	Macrolides: azithromycin, clarithromycin, erythromycin Ketolides: telithromycin Tetracyclines: doxycycline Fluoroquinolones†: levofloxacin, gatifloxacin, moxifloxacin, gemifloxacin
Ureaplasma urealyticum	Nongonococcal urethritis	Doxycycline‡ Azithromycin Alternatives: erythromycin base, levofloxacin
M. genitalium	Nongonococcal urethritis, pelvic inflammatory disease	Azithromycin
M. hominis	Pelvic inflammatory disease; persistent postabortion or postpartum fever; acute pyelonephritis	Doxycycline Clindamycin

*The optimal antibiotic choice for *Mycoplasma* infection is not clear, and a variety of dosages and treatment durations are used; however, the typical duration of *M. pneumoniae* treatment is 14 days.
†Ciprofloxacin and ofloxacin are not recommended.
‡*U. urealyticum* strains that are resistant to tetracyclines are usually sensitive to macrolides or quinolones.

In both animal and human studies, antimycoplasmal therapy significantly alleviated chronic respiratory disease caused by *M. pneumoniae*.[86,87] A randomized, double-blind, placebo-controlled trial in patients with stable chronic asthma showed that 6 weeks of treatment with clarithromycin produced significant improvement on respiratory function testing in those patients who tested positive for *M. pneumoniae* by PCR but not in those who tested negative; control subjects who received placebo also showed no improvement.[87] Additional clinical studies are needed to strengthen this observation.

Mycoplasma hominis, Mycoplasma genitalium, *and* Ureaplasma urealyticum

Because mycoplasmas and ureaplasmas often colonize the lower genitourinary tract of healthy adults (especially those who are sexually active), positive cultures for these organisms do not necessarily constitute proof of infection. Despite this difficulty, there is evidence that *M. hominis, M. genitalium,* and *U. urealyticum* can cause several clinical syndromes.[3,88] The significance of isolating these organisms in a variety of other syndromes is investigational or unknown.

NONGONOCOCCAL URETHRITIS

About 40% of men who experience an initial episode of nongonococcal urethritis have infection caused by *C. trachomatis* [*see 7:XIII Diseases Due to Chlamydia*].[89] In many cases of *Chlamydia*-negative nongonococcal urethritis (negative culture and negative serology), *U. urealyticum* may be the causative agent, as judged by (1) the presence of larger numbers of *U. urealyticum* organisms in *Chlamydia*-negative cases than in *Chlamydia*-positive cases; (2) the production of urethritis in human volunteers and nonhuman primates by intraurethral inoculation of *U. urealyticum* clinical isolates; and (3) the differential response to sulfisoxazole therapy (in one study, all 13 patients with *Chlamydia*-positive, *Ureaplasma*-negative nongonococcal urethritis showed a response, compared with only 14 of 30 *Chlamydia*-negative, *Ureaplasma*-positive patients).[89] *C. trachomatis* is susceptible to sulfonamides, but *U. urealyticum* is not. Such evidence suggests that *U. urealyticum* is the cause of at least some initial episodes of *Chlamydia*-negative nongonococcal urethritis in men—perhaps as many as 15% to 25% of these episodes.[90] Other studies have implicated *U. urealyticum* as a cause of acute urethral syndrome in some women; this organism may also cause chronic voiding symptoms in women, which may be mistaken for interstitial cystitis.[91] It is possible that factors such as serotype, strain-specific virulence determinants, or host factors will explain why positive cultures for *U. urealyticum* are not better correlated with evidence of clinical infection. Alternatively, disease may develop only upon initial exposure to ureaplasmas. *U. urealyticum* also has been implicated in urethroprostatitis and epididymitis.[3]

M. genitalium also appears to cause acute and possibly chronic nongonococcal urethritis.[88,92] One group of investigators found *M. genitalium* DNA in *C. trachomatis*–negative samples from 22% of heterosexual men with nongonococcal urethritis but from only 4% of asymptomatic control subjects.[93] Other studies have confirmed the disproportionate detection rates of *M. genitalium* in patients with nongonococcal urethritis. *M. genitalium* does not have a known role in prostatitis or epididymitis.[88] *M. hominis* does not appear to have a primary etiologic role in nongonococcal urethritis.

Treatment

Doxycycline (100 mg orally two times daily for 7 days) or azithromycin (1 g orally as a single dose) is the recommended treatment for nongonococcal urethritis. Erythromycin base (500 mg orally four times daily for 7 days) or levofloxacin (500 mg once daily for 7 days) are alternatives. These treatment regimens apply to cases associated with *U. urealyticum* as well as *M. genitalium*.[94] Tetracycline resistance has increased in recent decades and has been reported in as many as one third of clinical *U. urealyticum* isolates.[95] Resistant strains are known to cause persistent urethritis that often does not respond to treatment with tetracyclines; however, they are usually sensitive to macrolides or quinolones.[96] Sexual contacts of an index case should be treated at the same time as the index case. Other causes of treatment failure include poor compliance with medications, reinfection, disease caused by *Trichomonas vaginalis* or herpes simplex virus, prostatitis, and noninfectious etiologies.

For *M. genitalium,* the treatment of choice seems to be azithromycin. Treatment failures have been reported with other macrolides and with quinolones.[88,92]

UPPER URINARY TRACT INFECTION

M. hominis causes approximately 5% of acute pyelonephritis cases, judging from culture and serologic data.[97] However, the same association cannot be made for *U. urealyticum*.[3] Treatment of acute pyelonephritis from *M. hominis* is the same as that for pelvic inflammatory disease (PID) from this organism (see below).

U. urealyticum has a limited role in the production of urinary calculi. The frequency with which *U. urealyticum* reaches the kidney, the predisposing factors that allow this to occur, and the relative frequency of renal calculi induced by this organism as compared with other organisms is not known.[3]

PELVIC INFLAMMATORY DISEASE

M. hominis may cause some episodes of PID. In most of these cases, the organism occurs as part of a polymicrobial infection, but isolation of *M. hominis* from laparoscopic cultures of fallopian tubes in women with acute salpingitis has been reported, and the organism may be responsible for a few cases of PID on its own.[98] The prevalence of *M. hominis* involvement in PID may vary with global geographic location. It remains unclear how large a role *M. hominis* plays in the development of acute salpingitis and its sequelae. Some data also exist for an association of *M. genitalium* with PID.[88] In contrast, *U. urealyticum* is not considered to be a cause of PID.[3]

Treatment

M. hominis is resistant to the macrolides. Doxycycline is generally the drug of choice for *M. hominis* infections, although resistance has been reported.[99] Clindamycin is also generally active against *M. hominis*. Quinolones and ketolides have been found to be active in vitro against *M. hominis*; however, clinical experience is lacking.[96,100]

POSTABORTION AND POSTPARTUM FEVER

Positive blood cultures and concurrent seroconversion have implicated *M. hominis* as the primary pathogen in approximately 10% of women who have fever after abortion.[3] *M. hominis* is also responsible for 5% to 10% of fevers that arise more than 24 hours after vaginal delivery.[101] Women with vaginal colonization by *M. hominis* who have low or absent antibody titers against *Mycoplasma* may be predisposed to postpartum fever, which is probably associated with endometritis. These infections are usually self-limited; if symptoms persist, however, specific antimicrobial therapy should be given as for *M. hominis* PID (see above).

OTHER CONDITIONS ASSOCIATED WITH MYCOPLASMAS

M. hominis rarely causes brain abscess, wound infection, poststernotomy mediastinitis,[102] neonatal meningitis, and other nongenitourinary infections.[103,104] These infections are more common in immunocompromised or hypogammaglobulinemic persons. *U. urealyticum* and *M. hominis* can cause septic arthritis in immunodeficient patients,[3] and *U. urealyticum* likely causes neonatal pneumonitis and contributes to neonatal chronic lung disease (including bronchopulmonary dysplasia).[105] It is unclear whether *U. urealyticum* and *M. hominis* can cause male and female infertility,[106] spontaneous abortion,[107] premature labor and low birth weight,[108] and chorioamnionitis.[105]

The author participates in the speakers' bureau for Sanofi-Aventis.

Antimicrobial drugs discussed in this chapter have not been approved by the FDA for use in chronic respiratory Mycoplasma hominis *infection,* M. genitalium *infection, or* M. hominis *infection; azithromycin and levofloxacin have not been approved for use in* Ureaplasma *infection. The use of polymerase chain reaction for the diagnosis of* Mycoplasma *infection is investigational and is not approved by the FDA.*

References

1. Himmelreich R, Plagens H, Hilbert H, et al: Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. Nucleic Acids Res 25:701, 1997

2. Mushegian AR, Koonin EV: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci USA 93:10268, 1996

3. Taylor-Robinson D: Infections due to species of *Mycoplasma* and *Ureaplasma*: an update. Clin Infect Dis 23:671, 1996

4. Halm EA, Teirstein AS: Clinical practice: management of community-acquired pneumonia. N Engl J Med 347:2039, 2002

5. Denny FW, Clyde WA Jr, Glezen WP: *Mycoplasma pneumoniae* disease: clinical spectrum, pathophysiology, epidemiology, and control. J Infect Dis 123:74, 1971

6. Eaton MD, Meiklejohn G, Van Herick W: Studies on the etiology of primary atypical pneumonia: a filterable agent transmissible to cotton rats, hamsters, and chick embryos. J Exp Med 79:649, 1944

7. Baseman JB, Dallo SF, Tully JG, et al: Isolation and characterization of *Mycoplasma genitalium* strains from the human respiratory tract. J Clin Microbiol 26:2266, 1988

8. Foy HM, Grayston JT, Kenny GE, et al: Epidemiology of *Mycoplasma pneumoniae* infection in families. JAMA 197:859, 1966

9. Gleason PP: The emerging role of atypical pathogens in community-acquired pneumonia. Pharmacotherapy 22:2S, 2002

10. Lieberman D, Schlaeffer F, Boldur I, et al: Multiple pathogens in adult patients admitted with community-acquired pneumonia: a one year prospective study of 346 consecutive patients. Thorax 51:179, 1996

11. Miyashita N, Fukano H, Okimoto N, et al: Clinical presentation of community-acquired *Chlamydia pneumoniae* pneumonia in adults. Chest 121:1776, 2002

12. File TM Jr, Tan JS, Plouffe JF: The role of atypical pathogens: *Mycoplasma pneumoniae, Chlamydia pneumoniae,* and *Legionella pneumophila* in respiratory infection. Infect Dis Clin North Am 12:569 vii, 1998

13. Esposito S, Bosis S, Cavagna R, et al: Characteristics of *Streptococcus pneumoniae* and atypical bacterial infections in children 2–5 years of age with community-acquired pneumonia. Clin Infect Dis 35:1345, 2002

14. Collier AM, Clyde WA: Relationships between *Mycoplasma pneumoniae* and human respiratory epithelium. Infect Immun 3:694, 1971

15. Krause DC: *Mycoplasma pneumoniae* cytadherence: unravelling the tie that binds. Mol Microbiol 20:247, 1996

16. Baseman JB, Reddy SP, Dallo SF: Interplay between mycoplasma surface proteins, airway cells, and the protean manifestations of mycoplasma-mediated human infections. Am J Respir Crit Care Med 154:S137, 1996

17. Krause DC: *Mycoplasma pneumoniae* cytadherence: organization and assembly of the attachment organelle. Trends Microbiol 6:15, 1998

18. Kahane I, Banai M, Razin S, et al: Attachment of mycoplasmas to host cell membranes. Rev Infect Dis 4 Suppl:S185, 1982

19. Roberts DD, Olson LD, Barile MF, et al: Sialic acid-dependent adhesion of *Mycoplasma pneumoniae* to purified glycoproteins. J Biol Chem 264:9289, 1989

20. Somerson NL, Walls BE, Chanock RM: Hemolysin of *Mycoplasma pneumoniae*: tentative identification as a peroxide. Science 150:226, 1965

21. Cherry JD, Taylor-Robinson D: Peroxide production by mycoplasmas in chicken tracheal organ cultures. Nature 228:1099, 1970

22. Almagor M, Yatziv S, Kahane I: Inhibition of host cell catalase by *Mycoplasma pneumoniae*: a possible mechanism for cell injury. Infect Immun 41:251, 1983

23. Kahane I: Oxidative damage induced by mycoplasmas. Molecular and Diagnostic Procedures in Mycoplasmology. Razin S, Tully JG, Eds. Academic Press, New York, 1995, p 415

24. Murray HW, Masur H, Senterfit LB, et al: The protean manifestations of *Mycoplasma pneumoniae* infection in adults. Am J Med 58:229, 1975

25. Rollins S, Colby T, Clayton F: Open lung biopsy in *Mycoplasma pneumoniae* pneumonia. Arch Pathol Lab Med 110:34, 1986

26. Andreev J, Borovsky Z, Rosenshine I, et al: Invasion of HeLa cells by *Mycoplasma penetrans* and the induction of tyrosine phosphorylation of a 145-kDa host cell protein. FEMS Microbiol Lett 132:189, 1995

27. Taylor-Robinson D, Davies HA, Sarathchandra P, et al: Intracellular location of mycoplasmas in cultured cells demonstrated by immunocytochemistry and electron microscopy. Int J Exp Pathol 72:705, 1991

28. Dallo SF, Baseman JB: Intracellular DNA replication and long-term survival of pathogenic mycoplasmas. Microb Pathog 29:301, 2000

29. Winner F, Rosengarten R, Citti C: In vitro cell invasion of *Mycoplasma gallisepticum*. Infect Immun 68:4238, 2000

30. Hardy RD, Jafri HS, Olsen K, et al: *Mycoplasma pneumoniae* induces chronic respiratory infection, airway hyperreactivity, and pulmonary inflammation: a murine model of infection-associated chronic reactive airway disease. Infect Immun 70:649, 2002

31. Martin RJ, Kraft M, Chu HW, et al: A link between chronic asthma and chronic infection. J Allergy Clin Immunol 107:595, 2001

32. Yang J, Hooper WC, Phillips DJ, et al: Regulation of proinflammatory cytokines in human lung epithelial cells infected with *Mycoplasma pneumoniae*. Infect Immun 70:3649, 2002

33. Takeuchi O, Kaufmann A, Grote K, et al: Cutting edge: preferentially the R-stereoisomer of the mycoplasmal lipopeptide macrophage-activating lipopeptide-2 activates immune cells through a Toll-like receptor 2- and MyD88-dependent signaling pathway. J Immunol 164:554, 2000

34. Luhrmann A, Deiters U, Skokowa J, et al: In vivo effects of a synthetic 2-kilodalton macrophage-activating lipopeptide of *Mycoplasma fermentans* after pulmonary application. Infect Immun 70:3785, 2002

35. Seya T, Matsumoto M: A lipoprotein family from *Mycoplasma fermentans* confers host immune activation through Toll-like receptor 2. Int J Biochem Cell Biol 34:901, 2002

36. Kaufmann A, Muhlradt PF, Gemsa D, et al: Induction of cytokines and chemokines in human monocytes by *Mycoplasma fermentans*–derived lipoprotein MALP-2. Infect Immun 67:6303, 1999

37. Chiba H, Pattanajitvilai S, Mitsuzawa H, et al: Pulmonary surfactant proteins A and D recognize lipid ligands on *Mycoplasma pneumoniae* and markedly augment the innate immune response to the organism. Chest 123:426S, 2003

38. Cartner SC, Lindsey JR, Gibbs-Erwin J, et al: Roles of innate and adaptive immunity in respiratory mycoplasmosis. Infect Immun 66:3485, 1998

39. Foy HM: Infections caused by *Mycoplasma pneumoniae* and possible carrier state in different populations of patients. Clin Infect Dis 17(suppl 1):S37, 1993

40. Lo SC: Mycoplasmas and AIDS. Mycoplasmas: Molecular Biology and Pathogenesis. Maniloff J, Ed. American Society for Microbiology, Washington DC, 1992, p 525

41. Roifman CM, Rao CP, Lederman HM, et al: Increased susceptibility to *Mycoplasma* infection in patients with hypogammaglobulinemia. Am J Med 80:590, 1986

42. Foy HM, Kenny GE, Cooney MK, et al: Naturally acquired immunity to pneumonia due to *Mycoplasma pneumoniae*. J Infect Dis 147:967, 1983

43. Baseman JB, Tully JG: Mycoplasmas: sophisticated, reemerging, and burdened by their notoriety. Emerg Infect Dis 3:21, 1997

44. Yang J, Hooper WC, Phillips DJ, et al: Cytokines in *Mycoplasma pneumoniae* infections. Cytokine Growth Factor Rev 15:157, 2004

45. Rios AM, Mejias A, Chavez-Bueno S, et al: The impact of cethromycin (ABT-773) therapy on microbiologic, histologic, immunologic and respiratory indices in a murine model of *Mycoplasma pneumoniae* lower respiratory infection. Antimicrob Agents Chemother 48:2897, 2004

46. Fonseca-Aten M, Rios AM, Mejias A, et al: *Mycoplasma pneumoniae* induces host-dependent pulmonary inflammation and airway obstruction in mice. Am J Respir Cell Mol Biol 32:201, 2005

47. Hardy RD, Atkinson TP, Cassell GH: Immune responses to *Mycoplasma*. Mucosal Immunology, 3rd ed. Mestecky J, Lamm EL, Bienenstock J, et al, Eds. Elsevier, New York, 2005

48. Clyde WA Jr: Clinical overview of typical *Mycoplasma pneumoniae* infections. Clin Infect Dis 17(suppl 1):S32, 1993

49. Broome CV, LaVenture M, Kaye HS, et al: An explosive outbreak of *Mycoplasma pneumoniae* infection in a summer camp. Pediatrics 66:884, 1980

50. Niederman MS, Mandell LA, Anzueto A, et al: Guidelines for the management of adults with community-acquired pneumonia: diagnosis, assessment of severity, antimicrobial therapy, and prevention. Am J Respir Crit Care Med 163:1730, 2001

51. Gupta SK, Sarosi GA: The role of atypical pathogens in community-acquired pneumonia. Med Clin North Am 85:1349, 2001

52. Roberts DB: The etiology of bullous myringitis and the role of mycoplasmas in ear disease: a review. Pediatrics 65:761, 1980

53. McCracken GH Jr: Current status of antibiotic treatment for *Mycoplasma pneumoniae* infections. Pediatr Infect Dis 5:167, 1986

54. Hansbro PM, Beagley KW, Horvat JC, et al: Role of atypical bacterial infection of the lung in predisposition/protection of asthma. Pharmacol Ther 101:193, 2004

55. Fonseca-Aten M, Rios AM, Hardy RD: The role of atypical agents in respiratory tract disease. The Microbe-Host Interface in Respiratory Tract Infections: Current Research and Hot Topics. Kimpen JLL, Ramilo O, Eds. Horizon Bioscience, Wymondham, Norfolk, UK, 2005, p 133

56. Broughton RA: Infections due to *Mycoplasma pneumoniae* in childhood. Pediatr Infect Dis 5:71, 1986

57. Yoshida T, Asato Y, Kukita I, et al: A 7-year-old boy with mycoplasmal infection requiring extracorporeal membrane oxygenation. Eur J Pediatr 162:44, 2003

58. Orlicek SL, Walker MS, Kuhls TL: Severe *Mycoplasma pneumonia* in young children with Down syndrome. Clin Pediatr (Phila) 31:409, 1992

59. Shulman ST, Bartlett J, Clyde WA Jr, et al: The unusual severity of mycoplasmal pneumonia in children with sickle-cell disease. N Engl J Med 287:164, 1972

60. Tarp B, Jensen JS, Ostergaard L, et al: Search for agents causing atypical pneumonia in HIV-positive patients by inhibitor-controlled PCR assays. Eur Respir J 13:175, 1999

61. Ali NJ, Sillis M, Andrews BE, et al: The clinical spectrum and diagnosis of *Mycoplasma pneumoniae* infection. Mycoplasma Q J Med New Series 58:241, 1986

62. Blasco Patino F, Perez Maestu R, Lopez de Letona JM: [Mechanisms of disease in *Mycoplasma pneumoniae* infection: clinical manifestations and complications.] Rev Clin Esp 204:365, 2004

63. Bitnun A, Ford-Jones EL, Petric M, et al: Acute childhood encephalitis and *Mycoplasma pneumoniae*. Clin Infect Dis 32:1674, 2001

64. Smith R, Eviatar L: Neurologic manifestations of *Mycoplasma pneumoniae* infections: diverse spectrum of diseases: a report of six cases and review of the literature. Clin Pediatr (Phila) 39:195, 2000

65. Pfausler B, Engelhardt K, Kampfl A, et al: Post-infectious central and peripheral nervous system diseases complicating *Mycoplasma pneumoniae* infection: report of three cases and review of the literature. Eur J Neurol 9:93, 2002

66. Narita M, Yamada S: Two distinct patterns of central nervous system complications due to *Mycoplasma pneumoniae* infection. Clin Infect Dis 33:916, 2001

67. Bitnun A, Ford-Jones E, Blaser S, et al: *Mycoplasma pneumoniae* encephalitis. Semin Pediatr Infect Dis 14:96, 2003

68. Foy HM, Kenny GE, McMahan R, et al: *Mycoplasma pneumoniae* pneumonia in an urban area: five years of surveillance. JAMA 214:1666, 1970

69. Tay YK, Huff JC, Weston WL: *Mycoplasma pneumoniae* infection is associated with Stevens-Johnson syndrome, not erythema multiforme (von Hebra). J Am Acad Dermatol 35:757, 1996

70. Cherry JD: Anemia and mucocutaneous lesions due to *Mycoplasma pneumoniae* infections. Clin Infect Dis 17(suppl 1):S47, 1993

71. Chryssanthopoulos C, Eboriadou M, Monti K, et al: Fatal disseminated intravascular coagulation caused by *Mycoplasma pneumoniae*. Pediatr Infect Dis J 20:634, 2001

72. Hirshberg SJ, Charles RS, Ettinger JB: Pediatric priapism associated with *Mycoplasma pneumoniae*. Urology 47:745, 1996

73. Stephan JL, Galambrun C, Pozzetto B, et al: Aplastic anemia after *Mycoplasma pneumoniae* infection: a report of two cases. J Pediatr Hematol Oncol 21:299, 1999

74. Kenney RT, Li JS, Clyde WA Jr, et al: Mycoplasmal pericarditis: evidence of invasive disease. Clin Infect Dis 17(suppl 1):S58, 1993

75. Szymanski M, Petric M, Saunders FE, et al: *Mycoplasma pneumoniae* pericarditis demonstrated by polymerase chain reaction and electron microscopy. Clin Infect Dis 34:E16, 2002

76. Berger RP, Wadowksy RM: Rhabdomyolysis associated with infection by *Mycoplasma pneumoniae*: a case report. Pediatrics 105:433, 2000

77. Waites KB, Talkington DF: *Mycoplasma pneumoniae* and its role as a human pathogen. Clin Microbiol Rev 17:697, 2004

78. Miyashita N, Fukano H, Yoshida K, et al: Is it possible to distinguish between atypical pneumonia and bacterial pneumonia? Evaluation of the guidelines for community-acquired pneumonia in Japan. Respir Med 98:952, 2004

79. Talkington DF, Shott S, Fallon MT, et al: Analysis of eight commercial enzyme immunoassay tests for detection of antibodies to *Mycoplasma pneumoniae* in human serum. Clin Diagn Lab Immunol 11:862, 2004

80. Petitjean J, Vabret A, Gouarin S, et al: Evaluation of four commercial immunoglobulin G (IgG)- and IgM-specific enzyme immunoassays for diagnosis of *Mycoplasma pneumoniae* infections. J Clin Microbiol 40:165, 2002

81. Ferwerda A, Moll HA, de Groot R: Respiratory tract infections by *Mycoplasma pneumoniae* in children: a review of diagnostic and therapeutic measures. Eur J Pediatr 160:483, 2001

82. Michelow IC, Olsen K, Lozano J, et al: Diagnostic utility and clinical significance of naso- and oropharyngeal samples used in a PCR assay to diagnose *Mycoplasma pneumoniae* infection in children with community-acquired pneumonia. J Clin Microbiol 42:3339, 2004

83. Shames JM, George RB, Holliday WB, et al: Comparison of antibiotics in the treatment of mycoplasmal pneumonia. Arch Intern Med 125:680, 1970

84. Smith CB, Friedewald WT, Chanock RM: Shedding of *Mycoplasma pneumoniae* after tetracycline and erythromycin therapy. N Engl J Med 276:1172, 1967

85. Esposito S, Blasi F, Droghetti R, et al: Double-blind, randomized, placebo controlled trial of clarithromycin therapy in children with recurrent acute wheezing (abstr G-1542). 43rd Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, Sept 14–17, 2003

86. Fonseca-Aten M, Rios AM, Mejias A, et al: The effect of telithromycin in a murine model of chronic pulmonary mycoplasmosis (abstr g-015). 104th American Society for Microbiology General Meeting, New Orleans, May 23–27, 2004

87. Kraft M, Cassell GH, Pak J, et al: *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* in asthma: effect of clarithromycin. Chest 121:1782, 2002

88. Taylor-Robinson D: *Mycoplasma genitalium*: an update. Int J STD AIDS 13:145, 2002

89. Bowie WR, Wang S, Alexander ER, et al: Etiology of nongonococcal urethritis: evidence for *Chlamydia trachomatis* and *Ureaplasma urealyticum*. J Clin Invest 59:735, 1977

90. Stamm WE, Hicks CB, Martin DH, et al: Azithromycin for empirical treatment of the nongonococcal urethritis syndrome in men. JAMA 274:545, 1995

91. Potts JM, Ward AM, Rackley RR: Association of chronic urinary symptoms in women and *Ureaplasma urealyticum*. Urology 55:486, 2000

92. Taylor-Robinson D, Gilroy CB, Thomas BJ, et al: *Mycoplasma genitalium* in chronic non-gonococcal urethritis. Int J STD AIDS 15:21, 2004

93. Totten PA, Schwartz MA, Sjostrom KE, et al: Association of *Mycoplasma genitalium* with nongonococcal urethritis in heterosexual men. J Infect Dis 183:269, 2001

94. Sexually transmitted diseases treatment guidelines 2002. Centers for Disease Control and Prevention. MMWR Recomm Rep 51(RR-6):1, 2002

95. Chua KB, Ngeow YF, Ng KB, et al: *Ureaplasma urealyticum* and *Mycoplasma hominis* isolation from cervical secretions of pregnant women and nasopharyngeal secretions of their babies at delivery. Singapore Med J 39:300, 1998

96. Waites KB, Crabb DM, Bing X, et al: In vitro susceptibilities to and bactericidal activities of garenoxacin (BMS-284756) and other antimicrobial agents against human mycoplasmas and ureaplasmas. Antimicrob Agents Chemother 47:161, 2003

97. Thomsen AC: Occurrence and pathogenicity of *Mycoplasma hominis* in the upper urinary tract: a review. Sex Transm Dis 10(4 suppl):323, 1983

98. Stacey CM, Munday PE, Taylor-Robinson D, et al: A longitudinal study of pelvic inflammatory disease. Br J Obstet Gynaecol 99:994, 1992

99. McCormack WM: Susceptibility of mycoplasmas to antimicrobial agents: clinical

implications. Clin Infect Dis 17(suppl 1):S200, 1993

100. Waites KB, Crabb DM, Duffy LB: In vitro activities of ABT-773 and other antimicrobials against human mycoplasmas. Antimicrob Agents Chemother 47:39, 2003

101. Moller BR: The role of mycoplasmas in the upper genital tract of women. Sex Transm Dis 10:281, 1983

102. Sielaff TD, Everett JE, Shumway SJ, et al: *Mycoplasma hominis* infections occurring in cardiovascular surgical patients. Ann Thorac Surg 61:99, 1996

103. Madoff S, Hooper DC: Nongenitourinary infections caused by *Mycoplasma hominis* in adults. Rev Infect Dis 10:602, 1988

104. McMahon DK, Dummer JS, Pasculle AW, et al: Extragenital *Mycoplasma hominis* infections in adults. Am J Med 89:275, 1990

105. Hardy RD, Ramilo O: Mycoplasmal infections. Infectious Diseases of the Fetus and Newborn Infant, 6th ed. Remington JS, Klein JO, Eds. Elsevier (in press)

106. Stray-Pedersen B: Repercussion of mycoplasmas on male and female sterility. Acta Eur Fertil 16:101, 1985

107. Donders GG, Van Bulck B, Caudron J, et al: Relationship of bacterial vaginosis and mycoplasmas to the risk of spontaneous abortion. Am J Obstet Gynecol 183:431, 2000

108. Berman SM, Harrison HR, Boyce WT, et al: Low birth weight, prematurity, and postpartum endometritis: association with prenatal cervical *Mycoplasma hominis* and *Chlamydia trachomatis* infections. JAMA 257:1189, 1987

109. Waites KB, Talkington DF: *Mycoplasma pneumoniae* and its role as a human pathogen. Clin Microbiol Rev 17:697, 2004

110. Templeton KE, Scheltinga SA, Graffelman AW, et al: Comparison and evaluation of real-time PCR, real-time nucleic acid sequence-based amplification, conventional PCR, and serology for diagnosis of *Mycoplasma pneumoniae*. J Clin Microbiol 41:4366, 2003

XIII DISEASES DUE TO CHLAMYDIA

WALTER E. STAMM, M.D.

The chlamydiae are obligate intracellular bacteria that produce a wide variety of infections in many mammalian and avian species.[1] Three species of *Chlamydia* infect humans: *C. trachomatis*, *C. psittaci*, and *C. pneumoniae* [*see Table 1*]. *C. trachomatis* is exclusively a human pathogen and is transmitted from person to person via sexual contact, perinatal transmission, or close contact in households. *C. trachomatis* causes trachoma in arid developing parts of the world and is a major cause of sexually transmitted and perinatal infections worldwide. In the United States, *C. trachomatis* infection is the most common reportable infectious disease.[2] *C. psittaci*, in contrast, is more widely distributed in nature, producing genital, conjunctival, intestinal, or respiratory infections in many avian and mammalian species.[3] Humans are occasionally infected by avian strains after contact with infected birds, incurring a pneumonitis or a systemic infection termed psittacosis.

C. pneumoniae, the third chlamydial species that infects humans, was identified and characterized in the 1980s.[4] It is a fastidious organism that produces upper respiratory tract infection and pneumonitis in both children and adults.[5] Recent studies have also linked *C. pneumoniae* infection to atherosclerotic cardiovascular disease and perhaps to other illnesses, such as asthma and sarcoidosis.[6] Transmission is believed to occur via aerosol droplets; no animal reservoir has been identified.

All chlamydiae share a unique life cycle characteristic of the genus [*see Figure 1*].[7] The organism is transmitted in an extracellular nonreplicating form known as the elementary body. The elementary body adheres to and is phagocytosed into a host epithelial cell. Once inside the epithelial cell, the elementary body transforms into the intracellular replicative form of the organism, the reticulate body. Reticulate bodies divide by binary fission within membrane-bound vacuoles called inclusions. After approximately 36 hours, reticulate bodies condense into elementary bodies, the inclusion ruptures, and the elementary bodies disperse to infect adjacent epithelial cells or to be transmitted to other hosts. A unique feature of the chlamydial inclusion is its ability to resist lysosomal fusion; the mechanism of resistance is as yet unknown. The intracellular milieu also provides the organism with nutrients and serves as a refuge from host immune defense mechanisms.

Although chlamydiae were initially thought to be large viruses, they in fact possess both DNA and RNA, have a cell wall and ribosomes similar to those of gram-negative bacteria, and are inhibited by a variety of antimicrobials (see below).[7] *C. trachomatis*, *C. psittaci,* and *C. pneumoniae* all share a genus-specific lipopolysaccharide antigen. The commonly available complement fixation test for *Chlamydia* measures antibodies against this antigen. The three species can be differentiated serologically with the microimmunofluorescence test, which is directed against epitopes in the major outer membrane protein (MOMP).

Characteristically, chlamydial species produce chronic, persistent, and often asymptomatic infections of the epithelial lining of the eye, respiratory tract, and urogenital tract. Infection of epithelial cells induces secretion of cytokines and initiation of an innate immune response. Induction and persistence of a chronic inflammatory response may eventually result in fibrosis, scarring, and other chronic sequelae of infection.[8]

In cell cultures, chlamydia species are susceptible to many antimicrobials, including tetracycline, doxycycline, erythromycin, azithromycin, rifampin, clindamycin, ofloxacin, and levofloxacin. Although antibiotic-resistant strains of *Chlamydia* have been described, clinical antimicrobial resistance has yet to become a frequent problem in treating chlamydial infections.[9]

Diseases Due to *C. trachomatis*

Unlike *C. pneumoniae,* which appears to have only a single serotype, and *C. psittaci,* whose number of serotypes is unknown, *C. trachomatis* has at least 18 distinct serotypes (serovars).[10] These serotypes confer tissue tropism and disease specificity: serovars A, B, Ba, and C are associated with trachoma, whereas serovars D through K are associated with sexually transmitted and perinatally acquired infections. Serovars L1, L2, and L3 are more invasive than the other serovars, spread to lymphatic tissue, and grow readily in macrophages; they produce the clinical syndromes of lymphogranuloma venereum and hemorrhagic proctocolitis. In addition to these more commonly reported syndromes, *C. trachomatis* has been reported as an infrequent cause of other infections, including endocarditis, peritonitis, pleuritis, and periappendicitis.[11]

LABORATORY DIAGNOSIS

Laboratory testing for *C. trachomatis* has evolved considerably over the past decade. Four types of confirmatory procedures are now available: (1) direct microscopic examination of swabs or tissue scrapings utilizing fluorescent antibody staining (DFA); (2) cell culture isolation of the organism; (3) detection of chlamydial antigens or genes in specimens by immunologic means or nucleic acid amplification testing; and (4) serologic testing for antibodies to *C. trachomatis* [*see Table 2*].

Except in cases of inclusion conjunctivitis, DFA has largely been abandoned for diagnosis of chlamydial infection. Even in inclusion conjunctivitis, nucleic acid amplification testing of conjunctival smears is probably a better choice because of its higher sensitivity. Cell culture techniques for isolation of *Chlamydia* are not widely available and have numerous drawbacks: culture has exacting requirements for specimen transport and is technically demanding and expensive; moreover, its sensitivity is only 60% to 80% of that of newer diagnostic tests.[12] For those reasons, nonculture alternatives utilizing antigen or gene detection are the diagnostic methods of choice in most cases. The least expensive and most widely used of the nonculture tests are enzyme-linked immunoassays that detect chlamydial antigens on urethral or endocervical swabs. These tests, however, have limited sensitivity and specificity and cannot be used to test vaginal swabs or urine for *Chlamydia*.

The newer nucleic acid amplification tests, such as polymerase chain reaction (PCR), ligase chain reaction (LCR), and transcription-mediated amplification (TMA), are the most sensitive and specific of the currently available tests.[13] Uniquely, these tests have high accuracy even when used on specimens that contain only small numbers of organisms (e.g., first-void urine or vaginal swabs). The ability to use such specimens is of particular

Table 1 Comparative Features of *Chlamydia* species

Species	*C. trachomatis*	*C. psittaci*	*C. pneumoniae*
Natural hosts	Humans	Birds, mammals	Humans
Mode of transmission	Sexual; close personal contact; maternal/infant	Zoonotic	Respiratory droplets
Typical diseases	STDs, LGV, trachoma	Pneumonia, psittacosis	Upper respiratory infections, pneumonia, atherosclerotic vascular disease
Risk groups	STDs: sexually active adolescents and young adults Trachoma: young children in arid, unhygienic, crowded conditions	Bird fanciers, pet-shop workers, veterinarians, animal and poultry workers	Young children, older adults
Number of serotypes	18	Unknown	1

LGV—lymphogranuloma venereum STDs—sexually transmitted diseases

importance because their ease of collection facilitates patient compliance and community-based screening programs.

Serologic tests are of little value in the diagnosis of most chlamydial ocular or genital infections.[14] They are useful only in invasive syndromes such as pelvic inflammatory disease, epididymitis, lymphogranuloma venereum, and infant pneumonia, which are associated with significant increases in both microimmunofluorescence and complement-fixing antibodies.

SEXUALLY TRANSMITTED DISEASES

C. trachomatis is the most common bacterial cause of sexually transmitted disease (STD) in the United States, being responsible for an estimated four million cases a year.[2] The spectrum of illness attributable to these infections parallels that of gonococcal infection. In men, the most common syndromes are nongonococcal urethritis and acute epididymitis. In women, mucopurulent cervicitis, urethritis, bartholinitis, acute salpingitis, and perihepatitis are the syndromes most commonly seen. Inclusion conjunctivitis, proctitis, and Reiter syndrome affect both sexes.

Sexually transmitted *C. trachomatis* infection has a widespread distribution and a high incidence and prevalence in adolescents and young adults in the United States and Europe.[15] The age of peak incidence is the late teens and early 20s. Prevalence has been reported at 3% to 8% in general medical clinics and urban high schools, over 10% in asymptomatic military personnel undergoing routine physical examination, and as high as 15% to 20% in men and women attending STD clinics.[16,17] The highest prevalences of infection have been observed in persons who are single, have multiple sexual partners, are not using barrier contraception, report genital symptoms, have an infected sexual partner, or are attending a high-risk clinic such as an STD clinic. As many as 90% of infections in many settings may be asymptomatic and are thus discoverable only by screening. Recurrent chlamydial infections—often from untreated sexual partners—are common in high-risk groups.[18]

C. trachomatis initially infects the columnar epithelium of the genital tract and induces an inflammatory response that may persist for months or years. Serious sequelae, such as scarring of the fallopian tubes and damage to the upper genital tract, occur most often with repeated or persistent infections.[19] The mechanism through which repeated infection induces inflammation and subsequent complications is not clear. The *C. trachomatis* 60 kd heat shock protein has been implicated in these deleterious inflammatory responses and may induce antibodies that cross-react with human heat-shock protein. Other chlamydial proteins also may be important in this process.[20] Host genetic susceptibility may also play a role: persons with particular HLA haplotypes appear to be more susceptible to scarring with either genital or ocular infection.[21]

Nongonococcal Urethritis

Diagnosis Nongonococcal urethritis (NGU), the most common chlamydial urogenital infection in men, typically presents with burning pain on urination, urethral discharge, or urethral itching. A discharge from the urethra may be apparent on observation or may be visible only after stripping the urethra. The discharge is generally clear or mucoid, but it may be mucopurulent or purulent.[22] However, up to half of the men with *C. trachomatis* infection of the urethra will have no clinical manifestations of urethritis. Many of these patients will nevertheless have an increased number of leukocytes on Gram stain of a urethral smear. A presumptive diagnosis of NGU can be made on the basis of a leukocytic urethral exudate (≥ 4 polymorphonuclear leukocytes [PMNs] per 1,000× oil-immersion field) in the absence of concurrent gonococcal infection by Gram stain or culture [*see Table 2*]. About 40% of such cases of presumptive NGU are caused by *Chlamydia*; the remainder are caused by *Ureaplasma urealyticum, Mycoplasma genitalium,* and other microbes.[23] The specific diagnosis of chlamydial infection can be confirmed by a nucleic acid amplification assay such as PCR or LCR, an antigen detection test, or culture.

Treatment Men with presumed or confirmed NGU should be treated with oral doxycycline, 100 mg twice a day for 7 days, or azithromycin, 1 g orally as a single dose. The two regimens appear equally effective.[24] The azithromycin regimen offers the advantage of single-dose therapy but is considerably more expensive. Female partners of men with NGU should be examined and treated for presumed concomitant chlamydial infection.

Epididymitis

In 1% to 2% of men with chlamydial urethritis, the infection ascends in the genital tract to cause acute epididymitis. *C. trachomatis* is the major cause of acute epididymitis in heterosexual men younger than 35 years; other causes include *Neisseria gonorrhoeae* and, less commonly, urinary pathogens such as *Escherichia coli* or *Pseudomonas aeruginosa*.[25] Urinary pathogens are a more common cause of epididymitis in homosexual men who practice rectal intercourse and in men older than 35 years who have had urologic instrumentation or surgery.

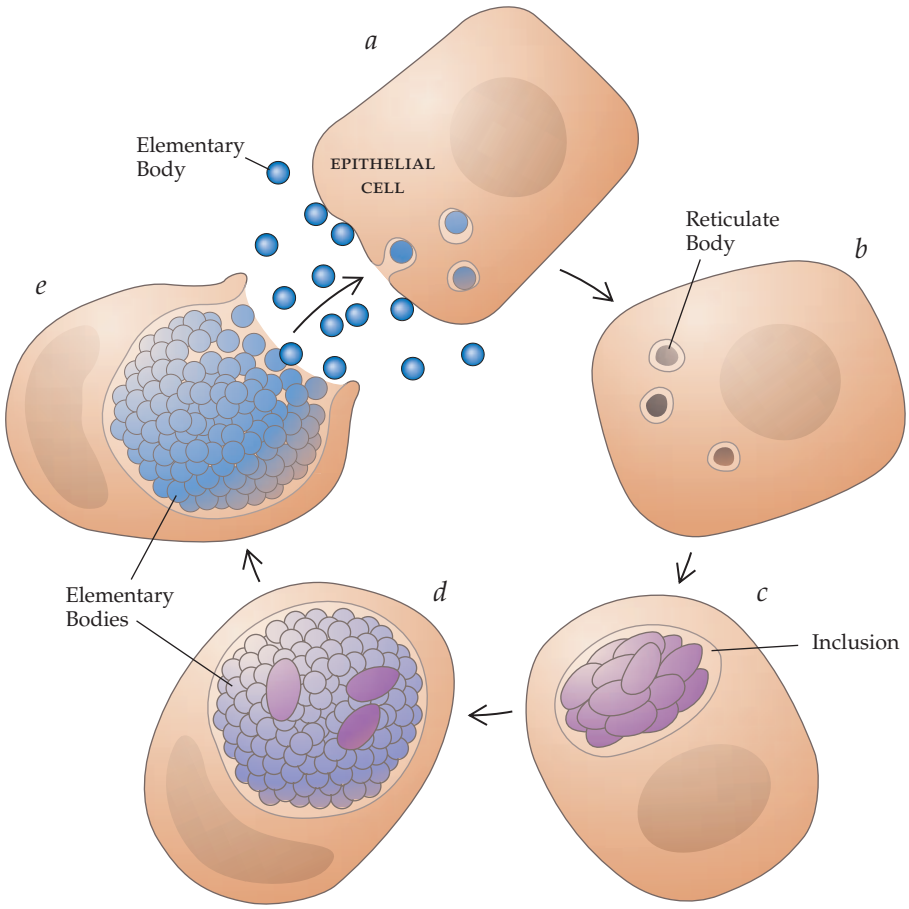


Figure 1 **Life cycle of *Chlamydia*. (*a*) Chlamydiae are transmitted in an extracellular nonreplicating form known as the elementary body. The elementary body adheres to and is phagocytosed into a host epithelial cell. (*b*) Once inside the epithelial cell, the elementary body transforms into the intracellular replicative form of the organism, the reticulate body. (*c*) Reticulate bodies divide by binary fission within membrane-bound vacuoles called inclusions. (*d*) The reticulate bodies then reorganize into elementary bodies; multiplication stops. (*e*) After 35 to 40 hours, the inclusion ruptures, and elementary bodies are released to infect adjacent epithelial cells or to be transmitted to other hosts.**

Diagnosis Epididymitis caused by *C. trachomatis* typically presents as unilateral scrotal pain, fever, and epididymal tenderness and swelling. Testicular torsion should be excluded by a radionuclide scan or Doppler flow study if the diagnosis is in doubt.

Treatment In some patients, epididymitis may be mild enough to be treated on an outpatient basis with oral antibiotics. In other cases, hospitalization is required for management of pain and initiation of parenteral antibiotics. Initial empirical therapy with ofloxacin, 300 mg orally or intravenously twice daily, is recommended until the etiologic agent is identified [*see Table 2*].[26]

Mucopurulent Cervicitis

Endocervical infection with *C. trachomatis* is considered the female counterpart of male urethritis.[27]

Diagnosis *C. trachomatis* cervical infection is often clinically silent, although a careful speculum examination will reveal mucopurulent cervicitis in some cases. When clinical manifestations are present, they are nonspecific—for example, vaginal discharge, vaginal bleeding, lower abdominal pain, or dysuria. On examination, findings include a yellow or green mucopurulent discharge from the endocervical os, the presence of cervical ectropion, edema in the area of ectropion, and easily induced mucosal bleeding.[27] A Gram stain of the endocervical exudate typically shows large numbers of PMNs. Women with suspected mucopurulent cervicitis should be evaluated for both chlamydial and gonococcal infection [*see Table 2*].

Treatment Chlamydial mucopurulent cervicitis is best treated with a single dose of 1 g of azithromycin. This regimen is effective and safe. Although more expensive than doxycycline, azithromycin is cost-effective because adolescents often fail to complete a 7-day course of doxycycline.[28] Unlike doxycycline, azithromycin appears to be safe during pregnancy and is used by many practitioners for treatment of chlamydial mucopurulent cervicitis or cervical infection in pregnant women.[29] Sexual partners of women found to have chlamydial cervical infection or mucopurulent cervicitis should be examined and empirically treated for chlamydial infection.

Table 2 Clinical Characteristics of Common *C. trachomatis* Infections in Adults

	Infection	Symptoms and Signs	Presumptive Diagnosis	Definitive Diagnosis	Treatment
Men	Nongonococcal urethritis	Urethral discharge, dysuria	Urethral leukocytosis; no gonococci seen	Urine or urethral NAAT	Azithromycin, 1 g p.o. (single dose) *or* Doxycyline, 100 mg p.o. b.i.d., for 7 days
	Epididymitis	Unilateral epididymal tenderness, swelling; pain; fever; presence of NGU	Urethral leukocytosis; pyuria on urinalysis	Urine or urethral NAAT; urine culture	Outpatient: Ofloxacin, 300 mg b.i.d. for 10 days *or* Ceftriaxone, 1 g I.M., plus doxycycline, 100 mg p.o., b.i.d., for 7 days
	Proctitis	Rectal pain, discharge, bleeding; history of receptive anal intercourse	≥ 1PMNs/OIF on rectal Gram stain; no gonococci seen	Rectal culture or DFA (NAAT untested)	Doxycycline, 100 mg p.o., b.i.d., for 7 days
	Conjunctivitis	Ocular pain, redness, discharge; simultaneous genital infection	Gram stain of conjunctival swab negative for bacterial pathogens; PMNs on smear	DFA or NAAT on conjunctival swab	Azithromycin, 1 g p.o. (single dose) *or* Doxycyline, 100 mg p.o., b.i.d., for 7 days
Women	Cervicitis	Mucopurulent cervical discharge; cervical bleeding; cervical edema	≥ 20 PMNs/OIF on cervical Gram stain	Urine or cervical NAAT	Azithromycin 1 g p.o. (single dose) *or* Doxycycline, 100 mg p.o., b.i.d., for 7 days
	Urethritis	Dysuria, frequency; no hematuria	Pyuria on UA; negative urine Gram stain and culture	Urine, cervical, or urethral NAAT	Azithromycin, 1 g, p.o. (single dose) *or* Doxycycline, 100 mg p.o., b.i.d., for 7 days
	Salpingitis	Lower abdominal pain, adnexal pain, cervical motion tenderness	Evidence of mucopurulent cervicitis	Urine or cervical NAAT	Outpatient: Ofloxacin, 400 mg b.i.d., p.o., plus metronidazole, 500 mg p.o., b.i.d., for 14 days *or* Ceftriaxone, 250 g I.M. plus doxycycline, 100 mg p.o., b.i.d., for 14 days
	Conjunctivitis	Ocular pain, redness, discharge; simultaneous genital infection	Gram stain of conjunctival swab negative for bacterial pathogens; PMNs on smear	DFA or NAAT on conjunctival swab	Azithromycin, 1 g p.o. (single dose) *or* Doxycycline, 100 mg p.o., b.i.d., for 7 days

DFA—direct fluorescent antibody NAAT—nucleic acid amplification test NGU—nongonococcal urethritis OIF—oil-immersion field PMN—polymorphonuclear neutrophil UA—urinalysis

Acute Urethral Syndrome in Women

In women presenting with dysuria, frequency, and pyuria, *C. trachomatis* is the pathogen that is most commonly identified if urine cultures fail to find *E. coli*, *Staphylococcus saprophyticus*, or other uropathogens.[30] Chlamydial infection should be considered to be the cause of so-called dysuria-pyuria syndrome when the urine culture does not demonstrate expected urinary bacterial pathogens, symptoms have been present for more than 7 days, the patient has a new sexual partner, or concomitant mucopurulent cervicitis is present on examination.[30,31]

Pelvic Inflammatory Disease

C. trachomatis is thought to account for about half of all cases of pelvic inflammatory disease (PID) in the United States.[32] Ascending intraluminal spread of *C. trachomatis* from the lower genital tract can produce endometritis, endosalpingitis, and pelvic peritonitis. On examination, evidence of mucopurulent cervicitis is generally found in women who have laparoscopically verified salpingitis due to *C. trachomatis*.

Diagnosis Chlamydial salpingitis often produces fewer clinical symptoms and signs than gonococcal or anaerobic salpingitis.[33] However, patients often present with lower abdominal pain, adnexal tenderness, vaginal discharge or bleeding, and uterine tenderness. They may or may not have fever.

Treatment Empirical treatment of PID must provide antimicrobial coverage for the major pathogens, including *C. trachomatis*, *N. gonorrhoeae*, and vaginal anaerobes [*see Table 2*]. Screening of high-risk young women for chlamydial cervical infection followed by treatment has been shown to prevent subsequent PID in a prospective cohort study.[34] Thus, chlamydial screening of all sexually active adolescent girls and women younger than 25 years with new sexual partners is strongly recommended.[35]

Complications Infertility and ectopic pregnancy are important sequelae of chlamydial tubal infection. Antecedent *C. trachomatis* PID has been linked to infertility from fallopian tube scarring. In more recent studies, demonstration of persistent, slowly replicating *C. trachomatis* in tubal tissue suggests that these patients may have chronic infection.

Perihepatitis, or Fitzhugh-Curtis syndrome, develops in a subset of women with chlamydial salpingitis. These patients present with right upper quadrant pain, fever, and often adnexal tenderness. The diagnosis can be established with an endocervical nucleic acid amplification test for chlamydia or by demonstration of high titers of antibody to *C. trachomatis*.

Acute Proctitis

Men who practice receptive anal intercourse may develop acute proctitis from *C. trachomatis* strains of serovars D through K or, rarely, L1 through L3.[36] Since 1998, chlamydial proctitis has been resurgent in homosexual and bisexual men in the United States.[37] These infections occasionally develop in heterosexual women. The severity of disease ranges from asymptomatic (which generally occurs in patients infected with serovars D through K) to severe (generally, serovars L1 through L3).

Diagnosis Patients with severe infection typically present with rectal pain, a rectal mucosal discharge, tenesmus, and rectal bleeding. Characteristically, a rectal Gram stain shows one or more PMNs per 1000× oil-immersion field and anoscopy demonstrates a mucopurulent discharge and easily induced mucosal bleeding.[36] The specific diagnosis can be made by rectal culture for *Chlamydia* or by a chlamydial DFA. The sensitivity and specificity of the newer nucleic acid amplification tests for rectal specimens has not been extensively studied, and these tests are not approved for this purpose.

Treatment Treatment regimens for chlamydial proctitis have not been extensively studied. Nevertheless, a 7-day regimen of doxycycline (100 mg b.i.d.) is recommended for cases caused by serovars D through K, and a 21-day course is recommended for cases caused by serovars L1 through L3 [*see Table 2*].

Reiter Syndrome

An unusual complication of chlamydial urethritis, Reiter syndrome (more recently termed reactive arthritis) consists of conjunctivitis, urethritis (or cervicitis in women), oligoarthritis, and characteristic lesions of the skin and mucous membranes [*see 15:III Seronegative Spondyloarthropathies*]. Although the pathogenesis of Reiter syndrome is obscure, more than 80% of affected patients have the HLA B27 haplotype, so a genetic predisposition is clearly involved.[38] Recent studies have found chlamydial antigens and genes in the involved joints, suggesting that viable *Chlamydia* organisms migrate from the urethra to the synovial tissue in these cases.

Lymphogranuloma Venereum

Lymphogranuloma venereum (LGV) is a sexually transmitted infection caused by *C. trachomatis* strains of the L1, L2, and L3 serovars. Typically, the disease begins with a transient primary genital lesion, followed by multilocular suppurative regional lymphadenopathy and, sometimes, hemorrhagic proctitis with associated regional lymphadenitis.[39] The acute phase of the disease is generally associated with fever, leukocytosis, and in some cases proctitis. Possible late complications include genital elephantiasis, rectal strictures, and fistulas of the penis, urethra, and rectum.

Epidemiology LGV occurs primarily in the second and third decades of life and is approximately four times more common in men than in women. Worldwide, the incidence is declining, but the disease is still endemic in parts of Asia, Africa, South America, and the Caribbean. The disease is rare in the United States, with only 186 cases reported in 1995. In the United States, many cases involve travelers or military personnel returning from endemic areas; most microbiologically confirmed cases have been in homosexual or bisexual men.[40]

Clinical manifestations The initial lesion of LGV in heterosexual persons is often a papule, vesicle, or ulcer on the penis or, in women, on the labia. This lesion typically heals without scarring. LGV strains of *C. trachomatis* have occasionally been recovered from these genital ulcers, as well as from the urethra of men or the endocervix of women who present with inguinal adenopathy. It is thus not entirely clear whether the most important primary site of infection is the skin lesion, the urethra, or the endocervix. Alternatively, primary anal or rectal infection may develop in men or women after receptive rectal intercourse.[41] Initial inoculation of the oral mucosa can also occur, leading to primary oral or pharyngeal lesions that may go unnoticed until adenopathy develops.

From the initial site of primary infection (urogenital, anal, rectal, or oral), the organism spreads to adjacent regional lymphatics. Typically, penile, vulvar, or anal infections spread to the inguinal and femoral lymph nodes. Rectal infection produces hypogastric and deep iliac lymphadenitis, whereas upper vaginal or cervical infection may result in obturator and iliac lymphadenitis. Heterosexual men with urogenital infection commonly display the so-called inguinal syndrome, which consists of painful inguinal lymphadenopathy developing approximately 2 to 6 weeks after the presumed exposure. Adenopathy generally is unilateral, with palpable enlargement of the iliac and femoral nodes on the same side as the enlarged inguinal nodes. Although the nodes may be discrete at first, they increasingly become matted, fluctuant, and suppurative. The overlying skin becomes fixed and inflamed and eventually develops multiple draining fistulas. Enlargement of lymph nodes both above and below the inguinal ligament may produce the so-called sign of the groove. On biopsy, the infected nodes have small abscesses surrounded by histiocytes. Spontaneous healing may occur after several months, but scarring and granulomatous masses often persist.

In the United States, LGV proctitis has increasingly been recognized in homosexual men. These patients present with rectal pain and mucopurulent bloody rectal discharge.[42] They often complain of tenesmus and such systemic signs and symptoms as fever, malaise, and weakness. Sigmoidoscopy reveals ulcerative proctocolitis with a purulent exudate and mucosal bleeding.

During the active development of regional lymphadenopathy, many patients experience extensive constitutional symptoms, including fever, chills, headache, meningismus, myalgia, and arthralgias. Complications may include arthritis, aseptic meningitis, encephalitis, hepatitis, and arthritis. If left untreated, rectal infections may eventually progress to perirectal abscesses, anal fistula, and fistulas involving the rectum, the vagina, the bladder, and the pelvic musculature. Rectal strictures are a late complication, as is elephantiasis due to associated lymphatic obstruction.[26]

Laboratory tests LGV is most readily diagnosed serologically. Both the LGV complement fixation test and the microimmunofluorescence test become strongly positive soon after the onset of lymphadenopathy. Alternatively, demonstration of *C. trachomatis* by culture or nucleic acid amplification tests from urethral or cervical specimens or from pus aspirated from buboes can confirm the diagnosis.

Differential diagnosis LGV must be differentiated from other sexually transmitted conditions that produce genital ulcers

and associated adenopathy. These include genital herpes simplex virus (HSV) infection, syphilis, chancroid, and granuloma inguinale. The clinical presentation, epidemiologic circumstances, and specific laboratory testing can usually differentiate these conditions unambiguously.

Treatment Doxycycline, 100 mg orally twice daily for 21 days, is generally recommended for treatment of LGV, despite the absence of trials demonstrating its efficacy. Alternative agents include erythromycin and sulfonamides.

PERINATAL INFECTIONS

Depending on the population tested, 5% to 25% of pregnant women have *C. trachomatis* infection of the cervix. A high proportion of infants born to these infected women will acquire *C. trachomatis* infection during passage through the birth canal.[35] If not identified and treated, the infection may persist for months or years.[43]

Clinically apparent neonatal inclusion conjunctivitis develops in approximately 50% to 60% of infants with perinatally acquired *C. trachomatis* infection.[44] In addition, *C. trachomatis* can often be isolated from the rectum, vagina, and nasopharynx of these infants.[45] Approximately 10% of perinatally infected infants experience a distinctive afebrile pneumonitis-like syndrome.[46]

Neonatal chlamydial conjunctivitis typically develops 5 to 14 days after birth and generally presents as a mucopurulent ocular discharge. Other causes of neonatal conjunctivitis—such as *N. gonorrhoeae*, *Haemophilus influenzae*, *Streptococcus pneumoniae,* and HSV—should also be considered. The diagnosis can be readily confirmed by DFA of a conjunctival smear or by PCR or LCR testing of the discharge. Because infants tend to be infected at multiple sites, systemic therapy is indicated, generally with erythromycin, 50 mg/kg/day orally in four divided doses for 10 to 14 days.

To prevent neonatal infection, all pregnant women should be screened for *C. trachomatis* infection in the third trimester, preferably with a nucleic acid amplification test. Those women found to be infected should be treated with a single 1 g oral dose of azithromycin; amoxicillin, 500 mg orally three times a day for 7 days; or erythromycin base, 500 mg orally four times a day for 7 days.[47]

ADULT INCLUSION CONJUNCTIVITIS

Inclusion conjunctivitis caused by *C. trachomatis* is occasionally seen in adults.

Diagnosis

Adult chlamydial conjunctivitis usually arises from inadvertent inoculation of the conjunctiva in patients with sexually acquired chlamydial infection. These patients typically present with an acute unilateral follicular conjunctivitis. Associated preauricular adenopathy may be evident. Such cases must be distinguished from acute conjunctivitis due to other organisms, such as adenovirus, HSV, or bacterial pathogens. If not treated, the disease may persist for weeks to months, but it typically resolves without scarring or visual impairment. Approaches to diagnosis are the same as those used for neonatal conjunctivitis (see above).

Treatment

Systemic treatment is indicated to cure both the ocular and the genital infection. Treatment regimens are the same as for genital infection (see above). Sexual partners should be evaluated and treated to prevent reinfection.

TRACHOMA

C. trachomatis—specifically, serovars A, B, Ba, and C—is the source of trachoma, the major preventable cause of blindness throughout the world.

Epidemiology

Trachoma remains prevalent in arid portions of the developing world, especially North Africa and the Middle East. Trachoma was formerly endemic in Native Americans of the southwestern United States, but it has largely been eliminated from that population. Transmission of *C. trachomatis* in endemic areas is believed to occur via eye to hand to eye contact, as well as via contaminated towels and other fomites.[48] Flies and other insects also may play a role in transmission. Most infections take place in early childhood, particularly in areas marked by crowding, poor standards of cleanliness, and the lack of clean water. The overall incidence of trachoma, as well as its severity, has been reduced dramatically over the past 35 years, primarily because of improved sanitary conditions and greater availability of clean water.

Pathogenesis

Severe, chronic, blinding trachoma is generally associated with repeated reinfection or persistent infection by trachoma-causing strains of *C. trachomatis*.[49] The immune response to the *C. trachomatis* 60 kd heat shock protein appears to be strongly linked to the conjunctival scarring and pannus formation that causes blindness in trachoma. Host genetic factors are likely also to be important, because specific HLA types are more commonly found in persons with progressive trachoma.[50]

Diagnosis

Clinical manifestations Initially, trachoma presents as a conjunctivitis characterized by multiple lymphoid follicles. Infection generally begins in children between 1 and 4 years of age. Reinfection is common, as is asymptomatic infection.[51] With repeated episodes, the cornea becomes involved, with inflammatory infiltrates and pannus formation (superficial vascularization). Conjunctival scarring eventually develops and causes the eyelids to turn inward and abrade the eyeball.[52] Eventually, corneal epithelial ulceration occurs. Destruction of the lacrimal glands and ducts produces a dry eye.

Laboratory tests Trachoma can be diagnosed on clinical grounds in a patient from an endemic area with follicular conjunctivitis. The diagnosis can be confirmed with culture, DFA, or nucleic acid amplification testing of secretions.

Treatment

Antimicrobial treatment should be provided in the early phases of trachoma, when its impact is potentially greatest. Oral tetracycline (in older children and adults) or erythromycin for 4 to 8 weeks has been recommended. Azithromycin (20 mg/kg orally in a single dose) can be used but may be prohibitively expensive in the developing world. Topical antimicrobials can be used, but their effectiveness is uncertain.

Public health measures to control endemic trachoma have focused on improving hygienic conditions and making clean water available. More recently, mass treatment of villages with single-

dose azithromycin or topical antimicrobials has been utilized. Personal cleanliness and a reduction in flies in a village may also be important.

Disease Due to *C. psittaci*

PSITTACOSIS

C. psittaci infects a wide variety of avian species, including parrots, parakeets, pigeons, finches, chickens, pheasants, and turkeys. Humans become infected through exposure to infected birds. Human infection is subacute, characterized by pneumonitis and a variety of systemic manifestations.[53]

Epidemiology

Parrots, parakeets, and budgerigars (so-called psittacine birds) are the most common sources of human psittacosis. However, human cases have also been attributed to pigeons, ducks, turkeys, chickens, and a variety of other bird species. Psittacosis is an occupational disease of pet-shop employees, pigeon fanciers, taxidermists, veterinarians, poultry workers, and others who work around birds. The actual incidence is unknown: 50 to 100 cases are reported each year in the United States,[54] but it is likely that many cases, especially milder ones, go undiagnosed and thus unreported.

Pathogenesis

C. psittaci can be isolated from the nasal secretions, excreta, tissues, and feathers of infected birds. The birds may be ill, but many show only minor evidence of infection; complete absence of symptoms is uncommon, however.

C. psittaci enters humans via the upper respiratory tract and spreads throughout the body via the bloodstream, localizing primarily in pulmonary alveolar macrophages and the endothelial cells of the liver and spleen.[55] A lymphocytic inflammatory response ensues in the lungs and at other sites of infection. Little is known of the pathogenesis of psittacosis at the cellular or molecular level.

Diagnosis

Clinical manifestations Psittacosis may vary in severity from a mild flulike illness to a fatal disease.[56] The incubation period is typically 7 to 14 days. Often, the illness begins abruptly, with shaking chills and fever. In other cases, onset is more gradual. Among the most common symptoms are a severe headache and a nonproductive cough. Systemic illness characterized by fever, malaise, myalgias, and chills is also common. Features that may help differentiate psittacosis from other pneumonic processes include the absence of signs of consolidation, absence of a pleural effusion, a relative bradycardia, absence of neutrophilia, splenomegaly, abnormal liver function tests, and a rash resembling the rose spots of typhoid fever (Horder spots). The course may be relatively benign and short-lived or severe and prolonged, with complications such as myocarditis, pericarditis, meningitis, or encephalitis.[57]

A history of a recent exposure to birds is most helpful. Other diseases that must be considered include influenza, other viral pneumonias, *M. pneumoniae*, Q fever, Legionnaires disease, and other bacterial or fungal pneumonias. With patients in whom pneumonia is not prominent, other systemic febrile illnesses such as brucellosis, leptospirosis, mononucleosis, hepatitis, or typhoid fever must be considered.

Laboratory tests In psittacosis, the white blood cell count is most commonly normal or decreased and the erythrocyte sedimentation rate is not elevated. Liver function tests may be abnormal. Changes on chest x-ray are typically nonspecific, with patchy infiltrates that may be lobar, wedge shaped, segmental, or nodular. The diagnosis of psittacosis is best confirmed by serologic studies, because *C. psittaci* is difficult to isolate from blood or infected secretions. In addition, the organism is hazardous to work with in the laboratory, and most clinical laboratories do not offer culture diagnosis. A fourfold rise in titer of complement-fixing antibody in acute and convalescent sera from a patient with a compatible clinical syndrome confirms the diagnosis.

Treatment

Tetracycline generally produces a rapid and dramatic response when initiated during the early phase of psittacosis.[58] Generally, 2 g daily given in four divided doses is sufficient. Treatment should be continued for at least 14 days after defervescence to prevent relapse of infection. An alternative to tetracycline is erythromycin. Although as many as 20% of patients died in the preantibiotic era, fewer than 5% of patients do not respond to therapy currently.

Diseases Due to *C. pneumoniae*

C. pneumoniae is a recently described species that, as its name indicates, is associated primarily with respiratory tract infections. Initially, *C. pneumoniae* strains were thought to be variants of *C. psittaci,* but subsequent studies have demonstrated that *C. pneumoniae* is a distinct species, on the basis of differences in small subunit ribosomal RNA sequence, morphology, and antigenic properties.[59] The organisms are fastidious and cannot be easily isolated from respiratory or other clinical specimens. *C. pneumoniae* from clinical specimens grows most effectively in HL cells and Hep 2 cells.[60]

EPIDEMIOLOGY

Because of the difficulty in isolating the organism, most epidemiologic studies have depended on microimmunofluorescence serologic studies. Such studies have indicated that *C. pneumoniae* infection is extremely prevalent, with approximately 40% to 50% of adults demonstrating seropositivity in almost all countries examined worldwide. Infections are uncommon until late childhood; the peak period of incidence appears to be between 10 and 20 years of age.[61] New infections or reinfections are acquired throughout life, however, with the seroprevalence continuing to increase throughout adult life. Serologic studies suggest that infections are more common in men than in women. Transmission is thought to occur from person to person via the respiratory route, much like *Mycoplasma* or respiratory viral infections. Most transmission occurs in schools or households. Well-described outbreaks of *C. pneumoniae* infection have also occurred in settings such as military barracks or school dormitories.[62]

PATHOGENESIS

Little is yet known about the molecular pathogenesis of *C. pneumoniae* infection. However, the organism is thought to initially infect the upper respiratory tract epithelium. In many individuals, long-lived, asymptomatic infection persists at these sites. Recent studies have demonstrated that after infection of

respiratory tract epithelial cells and inflammatory cells, *C. pneumoniae* is likely transmitted throughout the body via macrophages in the bloodstream.[63] There is clear evidence that replication of the organism occurs in vascular endothelium and synovial membranes. As with *C. trachomatis*, the outer membrane protein of *C. pneumoniae* may induce host immune responses that cross-react with human proteins, resulting in autoimmune inflammatory damage to tissues.

DIAGNOSIS

Clinical Manifestations

C. pneumoniae, like other chlamydial species, can produce asymptomatic infection of the respiratory tract and endovascular tissues. *C. pneumoniae* was originally discovered as a cause of upper and lower respiratory tract infection, and the organism may produce a pneumonic illness somewhat like that produced by *Mycoplasma,* although few studies have utilized culture to confirm the presence of *C. pneumoniae* in infected tissues. Characteristically, patients with *C. pneumoniae* pneumonia have prominent antecedent upper respiratory infections and a mild illness with fever, nonproductive cough, and small segmental infiltrates on chest x-ray. Pleuritis and pleural effusion are uncommon. It is likely that *C. pneumoniae* also causes other respiratory infections, including bronchitis, pharyngitis, sinusitis, and otitis media, but this is less extensively studied.

In addition to respiratory diseases, *C. pneumoniae* has recently been associated in serologic studies with a wide range of other conditions, including myocarditis, pericarditis, aseptic meningitis, erythema nodosum, sarcoidosis, asthma, chronic fatigue syndrome, multiple sclerosis, and Alzheimer disease. At present, the validity of these associations remains uncertain; their significance must be demonstrated in further studies.

There is more convincing evidence linking *C. pneumoniae* infection with atherosclerotic vascular disease. More than 30 epidemiologic studies have clearly demonstrated a strong association between serologic evidence for *C. pneumoniae* infection and atherosclerotic disease of the coronary and other arteries.[64] These associations remain after adjustment for potential confounding variables. In addition, *C. pneumoniae* has been identified in atherosclerotic plaques by a variety of techniques, including electron microscopy, DNA hybridization, immunocytochemistry, and PCR.[65] The organism has also been cultured from atherosclerotic plaques, demonstrating the presence of viable *C. pneumoniae* in the vessel wall.[66] Studies in cell-culture systems and in animal models support the hypothesis that *C. pneumoniae* can infect vascular endothelial cells, including smooth muscle cells and macrophages. Animal models also support the contention that *C. pneumoniae* infection of the upper respiratory tract is followed by widespread dissemination of the organism to atheromatous lesions in vessels.[67] In animals, antimicrobial treatment appears to slow the progression of atherosclerotic lesions associated with *C. pneumoniae* infection. Human trials are now ongoing to ascertain whether treatment with antimicrobials can reduce the risk of atherosclerotic heart disease and other vascular diseases associated with *C. pneumoniae* infection.

Laboratory Tests

Confirmation of *C. pneumoniae* respiratory infection is difficult because cell culture techniques are not widely available and in any case are not very sensitive. PCR and DNA probes have been utilized in research laboratories but are not available through most clinical laboratories. Microimmunofluorescence studies showing an increase in antibody to *C. pneumoniae* between acute and convalescent sera provide a specific diagnosis, but this technique is available in only a small number of laboratories. Complement-fixing antibody can also be measured, but that does not distinguish *C. pneumoniae* infection from *C. psittaci* or *C. trachomatis* infection. Accurate diagnosis of *C. pneumoniae* infection likely will await development and implementation of a more accurate and convenient PCR technology.

TREATMENT

There have been no controlled trials to ascertain the best antimicrobial regimen for *C. pneumoniae* respiratory infections. In vitro, *C. pneumoniae* is inhibited by erythromycin, tetracycline, doxycycline, azithromycin, clarithromycin, and some fluoroquinolones, such as levofloxacin. Most experts recommend tetracycline (2 g daily in four divided doses) or doxycycline (100 mg b.i.d.) as initial therapy. Treatment should be given for at least 2 to 3 weeks. Many cases can be managed in the outpatient setting, although severe disease (especially in the elderly) may occasionally require hospitalization and ventilatory support.

The impact of antimicrobial therapy on either treatment or prevention of atherosclerotic cardiovascular disease is unknown. Antibiotics are not currently recommended for this indication.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Schachter J, Stamm WE: *Chlamydia.* Manual of Clinical Microbiology, 7th ed. Murray PR, Baron EJ, Pfaller MA, et al, Eds. ASM Press, Washington, DC, 1999, p 795

2. Ten leading nationally notifiable infectious diseases—United States, 1995. MMWR Morb Mortal Wkly Rep 45:883, 1996

3. VanBuuren CE, Dorrestein GM, vanDijk JE: *C. psittaci* in birds: a review of the pathogenesis and histopathological features. Vet Q 16:38, 1994

4. Grayston JT, Kuo CC, Campbell LA, et al: *Chlamydiae pneumoniae* sp nov. for *Chlamydia* sp strain TWAR. Int J Syst Bacteriol 39:88, 1989

5. Grayston JT: Infections caused by *Chlamydia pneumoniae* strain TWAR. Clin Infect Dis 15:757, 1992

6. Grayston JT: Background and current knowledge of *Chlamydiae pneumoniae* and atherosclerosis. J Infect Dis 181(S3):S402, 2000

7. Schachter J: Biology of *Chlamydia trachomatis*. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 391

8. Brunham RC: Human Immunity to Chlamydiae. *Chlamydia* Intracellular Biology, Pathogenesis, and Immunity. Stephens RS, Ed. ASM Press, Washington, DC, 1999, p 211

9. Stamm WE: Potential for antimicrobial resistance in *Chlamydia pneumoniae*. J Infect Dis 181(S3):S456, 2000

10. Lampe MF, Suchland RJ, Stamm WE: Nucleotide sequence of the variable domains within the major outer membrane protein gene from serovariants of *Chlamydia trachomatis*. Infect Immun 61:213, 1993

11. Jones RB: *Chlamydia trachomatis* (trachoma, perinatal infections, lymphogranuloma venereum, and other genital infections). Principles and Practice of Infectious Diseases, 4th ed. Mandell GL, Bennet JE, Dolin R, Eds. Churchill Livingstone, New York, 1995, p 1679

12. Stamm WE: *Chlamydia trachomatis*—the persistent pathogen: Thomas Parran Award lecture. Sex Transm Dis 28:684, 2001

13. Stamm WE: *Chlamydia trachomatis* infections: progress and problems. J Infect Dis 179:S380, 1999

14. Campbell LA, Marrazzo JM, Stamm WE, et al: Chlamydiae. Laboratory Diagnosis of Bacterial Infections. Cimolai N, Ed. Marcel Dekker, New York, 2001, p 795

15. *Chlamydia trachomatis* genital infections—United States, 1995. MMWR Morb Mortal Wkly Rep 46:193, 1997

16. Stamm WE: *Chlamydia trachomatis* infections of the adult. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 407

17. Gaydos CA, Howell MR, Pare B, et al: *Chlamydia trachomatis* infections in female military recruits. N Engl J Med 339:739, 1998

18. Xu F, Schillinger JA, Markowitz LE, et al: Repeat *Chlamydia trachomatis* infection in women: analysis through a surveillance case registry in Washington State, 1993–98. Am J Epidemiol 152:1164, 2000

19. Hillis SE, Owens LM, Marchbanks PA, et al: Recurrent chlamydial infections increase the risks of hospitalization for ectopic pregnancy and pelvic inflammatory disease. Am J Obstet Gynecol 176:103, 1997

20. LaVerda D, Albanese LN, Ruther PE, et al: Seroreactivity of *Chlamydia trachomatis* Hsp10 correlates with severity of human genital tract disease. Infect Immun 68:303, 2000

21. Gaur LK, Peeling RW, Cheang M, et al: Association of *Chlamydia trachomatis* heat-shock protein 60 antibody and HLA class II DQ alleles. J Infect Dis 180:234, 1999

22. Stamm WE, Koutsky LA, Benedetti JK, et al: *Chlamydia trachomatis* urethral infections in men: prevalence, risk factors, and clinical manifestations. Ann Intern Med 100:47, 1984

23. Horner P, Thomas B, Gilroy CB, et al: Role of *Mycoplasma genitalium* and *Ureaplasma urealyticum* in acute and chronic nongonococcal urethritis. Clin Infect Dis 32:995, 2001

24. Martin DH, Mroczkowski TF, Dalu ZA, et al: A controlled trial of a single dose of azithromycin for the treatment of chlamydial urethritis and cervicitis. N Engl J Med 327:921, 1992

25. Berger RE, Kessler D, Holmes KK: Etiology and manifestations of epididymitis in young men: correlations with sexual orientation. J Infect Dis 155:1341, 1987

26. 1998 guidelines for treatment of sexually transmitted diseases. MMWR Morb Mortal Wkly Rep 47:1, 1998

27. Holmes KK, Stamm WE: Lower genital tract infection syndromes in women. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 761

28. Haddix AC, Hillis SD, Kassler WJ: The cost-effectiveness of azithromycin for *Chlamydia trachomatis* infections in women. Sex Transm Dis 22:274, 1995

29. Bush MR, Rosa C: Azithromycin and erythromycin in the treatment of cervical chlamydial infection during pregnancy. Obstet Gynecol 84:61, 1994

30. Stamm WE, Wagner KF, Amsel R, et al. Causes of the acute urethral syndrome in women. N Engl J Med 303:409, 1980

31. Stamm WE, Stapleton AE: Approach to the patient with urinary tract infections. Infectious Diseases, 2nd ed. Gorbach SL, Bartlett JG, Blacklow NR, Eds. WB Saunders Co, Philadelphia, 1997, p 943

32. Cates W Jr, Rolfs RT JR, Aral SO: Sexually transmitted diseases, pelvic inflammatory disease, and infertility: an epidemiologic update. Epidemiol Rev 12:199, 1990

33. Cates W Jr, Joesoef MR, Goldman MB: Atypical pelvic inflammatory disease: can we identify clinical predictors? Am J Obstet Gynecol 169:341, 1993

34. Scholes D, Stergachis A, Heidrich FE, et al: Prevention of pelvic inflammatory disease by screening for cervical chlamydia infection. N Engl J Med 334:1362 1996

35. Nelson HD, Helfand M: Screening for chlamydial infection. Am J Prev Med 20(3S):95, 2001

36. Boisvert JF, Koutsky LA, Suchland RJ, et al: Clinical features of *Chlamydia trachomatis* rectal infection by serovar among homosexually active men. Sex Transm Dis 26:392, 1999

37. Resurgent bacterial STDs among men who have sex with men. King County, Washington 1997–1999. MMWR Morb Mortal Wkly Rep 48:773, 1999

38. Hughes RA, Keat AC: Reiter's syndrome and reactive arthritis: a current view. Semin Arthritis Rheum 24:190, 1994

39. Perrine PL, Stamm WE: Lymphogranuloma venereum. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 425

40. Division of STD Prevention, Sexually Transmitted Disease Surveillance, 1995. U. S. Department of Health and Human Services, Public Health Service. Atlanta Centers for Disease Control and Prevention, 1996

41. Hopsu-Havu VK, Sonck CE: Infiltrative, ulcerative and fistular lesions of the penis due to lymphogranuloma venereum. Br J Vener Dis 49:193, 1973

42. Schachter J, Smith DE, Dawson CR, et al: Lymphogranuloma venereum. I. comparison of the Frei test, complement fixation test, and isolation of the agent. J Infect Dis 120:372, 1969

43. Bell TA, Stamm WE, Wang SP, et al: Chronic *Chlamydia trachomatis* infections in infants. JAMA 267:400, 1992

44. Hammerschlag MR, Anderka M, Semine DZ, et al: Prospective study of maternal and infantile infection with *Chlamydia trachomatis*. Pediatrics 64:142, 1979

45. Schachter J, Grossman M, Sweet RL, et al: Prospective study of perinatal transmission of *Chlamydia trachomatis*. JAMA 255:3374, 1986

46. Beem MD, Saxon EM: Respiratory-tract colonization and a distinctive pneumonia syndrome in infants infected with *Chlamydia trachomatis*. N Engl J Med 296:306, 1977

47. Cohen I, Veille JC, Calkins BM: Improved pregnancy outcome following successful treatment of chlamydial infection. JAMA 263:3160, 1990

48. Taylor HR: A trachoma perspective. Ophthalmic Epidemiol 8:69, 2001

49. Grayston JT, Wang SP, Yeh LJ, et al: Importance of reinfection in the pathogenesis of trachoma. Rev Infect Dis 7:717, 1985

50. Mabey D, Fraser-Hurt N: Trachoma. BMJ 323:218, 2001

51. West SK, Munoz B, Mkocha H, et al: Progression of active trachoma to scarring in a cohort of Tanzanian children. Ophthalmic Epidemiol 8:137, 2001

52. Bailey RL, Arullendran P, Whittle HC, et al: Randomized controlled trial of single-dose azithromycin in treatment of trachoma. Lancet 342:453, 1993

53. Yung AP, Grayson ML: Psittacosis—a review of 135 cases. Med J Aust 148:228, 1988.

54. Compendium of psittacosis (chlamydiosis) control, 1997. MMWR Morb Mortal Wkly Rep 46(RR-13):1, 1997

55. MacFarlane JT, MacRae AD: Psittacosis. Br Med Bull 39:163, 1983.

56. Mason AB, Jenkins P: Acute renal failure in fulminant psittacosis. Respir Med 88:239, 1994

57. Shapiro DS, Kenney SC, Johnson M, et al: *Chlamydia psittaci* endocarditis diagnosed by blood culture. N Engl J Med 325:1192, 1992

58. Grayston JT, Campbell LA, Kuo CC, et al: A new respiratory tract pathogen: *Chlamydia pneumoniae* strain TWAR. J Infect Dis 161:618, 1990

59. Chi EY, Kuo CC, Grayston JT: Unique ultrastructure in the elementary body of *Chlamydia* sp. strain TWAR. J Bacteriol 169:3797, 1987

60. Saikku P: The epidemiology and significance of *Chlamydia pneumoniae*. J Infect Dis 25:27, 1992

61. Falsey AR, Walsh EE: Transmission of *Chlamydia pneumoniae*. J Infect Dis 168:493, 1993

62. Saikku P, Wang SP, Klumola M, et al: An epidemic of mild pneumonia due to an unusual strain of *Chlamydia psittaci*. J Infect Dis 151:832, 1985

63. Boman J, Gaydos CA: Polymerase chain reaction detection of *Chlamydia pneumoniae* in circulating white blood cells. J Infect Dis 181(suppl 3):S452, 2000

64. Danesh J, Collins R, Peto R: Chronic infections and coronary heart disease: Is there a link? Lancet 350:430, 1997

65. Taylor-Robinson D, Thomas BJ: *Chlamydia pneumoniae* in atherosclerotic tissue. J Infect Dis 181 (suppl 3):S437, 2000

66. Maass M, Bartels C, Engel PM, et al: Endovascular presence of viable *Chlamydia pneumoniae* is a common phenomenon in coronary artery disease. J Am Coll Cardiol 31:827, 1998

67. Moazed TC, Kuo CC, Grayston JT, et al: Evidence of systemic dissemination of *Chlamydia pneumoniae* via macrophages in the mouse. J Infect Dis 177:1322, 1998

Acknowledgment

Figure 1 Seward Hung.

XIV ANTIMICROBIAL THERAPY

Timothy H. Dellit, m.d.
Harvey B. Simon, m.d., f.a.c.p.
Thomas M. Hooton, m.d., f.a.c.p.

Overview of Antimicrobial Therapy

The essential feature of effective antimicrobial agents is the ability to inhibit the growth of microorganisms at concentrations tolerated by the host. Antimicrobial agents generally target anatomic structures or biosynthetic pathways unique to microorganisms.

The appropriate choice of an antimicrobial for an infection depends on five considerations: (1) the infecting organism and its antimicrobial susceptibilities; (2) the type of infection (e.g., abscess, bacteremia, meningitis, or urinary tract infection); (3) host factors (e.g., neutropenia, immune deficiencies, concurrent illnesses, age, drug allergies, and renal function); (4) factors associated with antimicrobial agents (e.g., dosage, routes of administration, drug interactions, serum levels and tissue penetration, potential toxicities, and cost); and (5) public health considerations. The widespread use of antibiotics has led to selection for highly resistant organisms that subsequently pose a risk for the patient and the community.

identification of the infecting organism

Prompt identification of the causative organism is essential for the selection of appropriate antimicrobial drugs. A Gram stain of infected body fluids can provide early clues to the etiologic agent. Other relatively simple laboratory procedures, such as immunodiagnostic tests, may provide rapid identification of infecting organisms. If it is not possible to perform a Gram stain, initial decisions are based on the clinical features and the usual bacteriology of the illness, pending results of appropriate cultures. Culture results must be interpreted with full recognition of the indigenous flora on various mucosal surfaces and on the skin. Culture identification of the infecting organism and knowledge of the local antibiograms can offer clues to the likely antimicrobial susceptibilities [*see Table 1*]. However, even within a single institution's antibiogram, patterns of susceptibilities are influenced by patient characteristics (e.g., previous antimicrobial exposure) and location within the hospital (e.g., intensive care unit or general ward). National trends in antimicrobial resistance are tracked by the Centers for Disease Control and Prevention (CDC) (http://www.cdc.gov/drugresistance/community).

determination of bacterial susceptibility to specific drugs

Determination of the in vitro susceptibility of isolated bacteria is usually performed to ensure that the empirical antimicrobial regimen is optimal. Timely availability of susceptibility data allows tailoring, when appropriate, of the antimicrobial regimen. However, the susceptibility of certain species (e.g., group A streptococci) to the drug of choice (e.g., penicillin G) has been so uniform that routine in vitro testing is unnecessary. Moreover, in vitro susceptibility data may not accurately predict the therapeutic outcome. For example, *Enterobacter* infections treated with third-generation cephalosporins, such as ceftazidime, may lead to induction of AmpC β-lactamases and clinical failure despite initial in vitro susceptibility.[1] Similarly, *Klebsiella* species and *Escherichia coli* may harbor extended-spectrum β-lactamases (ESBLs) and still appear susceptible to cephalosporins in vitro unless the clinical microbiology laboratory actively screens for their presence. Cephalosporin treatment of such an apparently susceptible organism with an unrecognized ESBL, however, is associated with a high rate of failure.[2]

Disk diffusion is the classic method of testing antimicrobial susceptibility. When necessary, organisms can be tested by tube dilution techniques to determine the minimal inhibitory concentration (MIC) and the minimal bactericidal concentration (MBC) of various antibiotics. Newer techniques include the gradient diffusion method and various nucleic acid–based tests that can detect genes conferring antimicrobial resistance in bacteria.[3]

the site of infection and ancillary therapy

The efficacy of antimicrobial therapy depends on drug delivery to the site of infection. Transport across the blood-brain barrier varies considerably among antibiotics.[4] Intracellular parasites, such as rickettsiae and mycobacteria, require therapy with agents that can reach the intracellular milieu of the organism. Necrotic lesions, such as sequestra in bone or large abscesses, hinder penetration by antimicrobial agents. Drainage, debridement, or both, combined with antimicrobials, are necessary. Antimicrobial therapy for infection in or around a foreign body or prosthesis often is ineffective unless these materials are removed or replaced. Infections behind obstructing lesions, such as pneumonia that occurs behind a blocked bronchus or cholecystitis caused by biliary obstruction, will not respond to antibiotics until the obstruction is relieved.

factors affecting dosage and route of administration

Many antimicrobial agents are absorbed sufficiently well via the oral route to provide effective serum levels in patients with normal gastrointestinal function.[5] For some antimicrobials (e.g., fluoroquinolones, metronidazole, and linezolid), the pharmacokinetics of the oral agents are almost the same as those of the parenteral formulations, and oral therapy, when appropriate, is much less expensive. The enteral absorption of some antimicrobial agents is impeded by food and some medications (e.g., antacids); these antibiotics, such as tetracyclines, should be administered at least 1 hour before meals or several hours after meals. Certain antibiotics, such as neomycin, are essentially nonabsorbable and are used primarily for their effects on bowel flora. Intramuscular administration is used in a small number of conditions (e.g., benzathine penicillin for syphilis or ceftriaxone for gonorrhea), but most antimicrobial agents cannot be given intramuscularly because of local pain or necrosis at the injection site. Intravenous administration must be used in the treatment of major and life-threatening infections such as septic shock, meningitis, and endocarditis. Patients who are clinically stable after initial intravenous treatment may be ready for discharge from the hospital and may be able to receive parenteral antibiotics on an outpatient basis. Newer antimicrobial agents with long half-lives can be administered with improved intravenous catheters and delivery devices in simplified regimens. Alternatively, the enhanced bioavailability of some antimicrobial agents has allowed many patients to complete a course of oral antimicrobial therapy after they are discharged. Such pharmacologic advances have led to substantial economic benefits, enhanced patient comfort, and good therapeutic results with few complications.[6]

Concurrent Illnesses

Patients with immunosuppressive illnesses are vulnerable to opportunistic pathogens. These patients may require broader antimicrobial coverage and combination therapy for ordinary pathogens. The same is true to a lesser extent for patients with chronic debilitating illness. Patients with renal insufficiency or liver disease may be unusually susceptible to direct drug toxicity [*see* Direct Drug Toxicity, *below*].

Many antimicrobial agents (e.g., penicillins, cephalosporins, aminoglycosides, vancomycin, and fluoroquinolones) are excreted primarily by the kidneys. Guidelines for dosage adjustment in patients who have severe renal failure and are undergoing dialysis or peritoneal dialysis are provided elsewhere [*see 10:IX Pharmacologic Approach to Renal Insufficiency, Appendix A*]. Antibiotics with nonrenal metabolism may be preferable in azotemic patients, provided that the infecting organism is susceptible. A number of antibiotics, including nafcillin, piperacillin, doxycycline, metronidazole, and rifampin, are excreted by the liver.

Pregnancy

The administration of antimicrobial agents during pregnancy and in the postpartum period poses several problems. Foremost is the question of safety, both for the mother and for the fetus or neonate. Although most antibiotics cross the placenta and enter maternal milk, the concentrations to which the fetus or neonate is exposed vary widely. Because the immature liver may lack the enzymes required to metabolize certain drugs, pharmacokinetics and toxicities in the fetus are often very different from those in older children and adults. Teratogenicity is a major concern when any drug is administered during pregnancy. Finally, it may be necessary to alter the dosage schedules of drugs that appear to be safe to use during pregnancy. Increases in maternal blood volume, glomerular filtration rate (GFR), and hepatic metabolic activity often reduce the maternal serum levels of antimicrobials by 10% to 50%, especially late in pregnancy and in the early postpartum period. In some women, delayed gastric emptying may reduce the absorption of antibiotics that have been administered orally during pregnancy.

Even though 25% to 40% of women receive antibiotics during pregnancy, data regarding safety and efficacy in this setting are often scarce. Some general recommendations have been proposed, but they are intended only as a guide [*see Table 2*]; in all cases, therapy must be individualized, and both the indications for antibiotics and the possible risks to mother and fetus must be considered. Individual decisions are also required for lactating mothers. Although most antimicrobial agents appear safe for

Table 1 Antimicrobial Susceptibilities of Clinical Isolates at a Representative Hospital

Organism	*No. of Isolates Tested*	*Strains Susceptible to Various Antimicrobials (%)*										
		Pen	*Oxa*	*Cef*	*Ctx*	*Clinda*	*Erythro*	*Levo*	*Tetra*	*TMP-SMX*	*Vanco*	
GRAM-POSITIVE												
Staphylococcus aureus												
Methicillin-susceptible	1,250	14	100	100	—[†]	96	67	95	95	98	100	
Methicillin-resistant	1,107	0	0	0	—[†]	44	7	20	92	90	100	
Coagulase-negative staphylococci	159	6	21	21	—[†]	56	22	43	79	53	100	
Streptococcus pneumoniae	210	75*	—[†]	—[†]	98	91	82	99	88	—[†]	100	
		Amp	*Amp-Sulb*	*Azt*	*Cef*	*Ceftaz*	*Gent*	*Imi*	*Levo*	*Pip-Tazo*	*Tetra*	*TMP-SMX*
GRAM-NEGATIVE												
Acinetobacter species	656	—[†]	75	1	—[†]	19	43	100	—[†]	22	28[‖]	—[†]
Citrobacter freundii complex	87	5	48	75	8	74	87	100	73	77	57	79
Enterobacter cloacae	248	5	34	84	2	83	94	100	92	80	52	80
Escherichia coli	1,479	57	79	98	90	99	96	100	93	97	69	77
Klebsiella pneumoniae	430	3	90	99	93	99	98	100	95	92	68	92
Proteus mirabilis	179	95	98	100	90	100	98	100	89	100	2	86
Pseudomonas aeruginosa	793	—[†]	—[†]	63	—[†]	91	88	82	53[‡]	92	—[†]	—[†]
Serratia marcescens	209	1	3	94	1	96	78	100	74	82	3	76
Stenotrophomonas maltophilia	112	—	—[†]	—[†]	—[†]	—[†]	—[†]	—[†]	98[§]	—[†]	99[‖]	94

Note: This listing represents the results of standard disk susceptibility tests performed on clinical isolates (from both hospitalized and ambulatory patients) as reported in a representative hospital's antibiogram in 2002. Drugs selected for testing often represent a class of antimicrobial agents (e.g., oxacillin for the penicillinase-resistant penicillins such as oxacillin and nafcillin) and vary according to the specific organism.

**S. pneumoniae* treated with penicillin: 10% resistant and 15% intermediate.

[†]Insufficient data, or drug was not tested.

[‡]Ciprofloxacin, a more active fluoroquinolone for *P. aeruginosa*, was tested.

[§]Moxifloxacin, a more active fluoroquinolone for *S. maltophilia*, was tested.

[‖]Minocycline, a more active tetracycline for *Acinetobacter species* and *S. maltophilia*, was tested.

Amp—ampicillin Amp-Sulb—ampicillin-sulbactam Azt—aztreonam Cef—cefazolin Ceftaz—ceftazidime Clinda—clindamycin Ctx—ceftriaxone Erythro—erythromycin Gent—gentamicin Imi—imipenem Levo—levofloxacin Oxa—oxacillin Pen—penicillin Pip-Tazo—piperacillin-tazobactam Tetra—tetracycline TMP-SMX—trimethoprim-sulfamethoxazole Vanco—vancomycin

Table 2 Antibiotics in Pregnancy

*Drug**	*Major Toxic Potential*		*Excreted in Mother's Milk*
	Maternal	*Fetal*	
Class B: Presumed safe based on animal studies			
Azithromycin	GI intolerance	None known	Yes (in animals)
Aztreonam	Allergies	None known	Trace
Cephalosporins	Allergies	None known	Trace
Clindamycin	Allergies, colitis	None known	Trace
Ertapenem	Allergies	None known	Trace
Erythromycin base	Allergies, GI intolerance	None known	Yes
Ethambutol	Optic neuritis	None known	Unknown
Penicillins	Allergies	None known	Trace
Meropenem	Allergies	None known	Trace
Metronidazole (avoid in first trimester)	Hypersensitivity, alcohol intolerance, neuropathy	None known (teratogenic in animals but appears safe in human experience)	Yes
Nitrofurantoin	Pulmonary toxicity, hepatotoxicity	None known	Yes
Quinupristin-dalfopristin	Hepatotoxicity, arthralgias	None known	Unknown
Class C: Uncertain safety; no human studies and animal studies show an adverse effect			
Clarithromycin	GI intolerance	Teratogenic in animals at high doses	Yes (in animals)
Chloramphenicol (avoid at term)	Blood dyscrasias	Gray syndrome	Yes
Fluoroquinolones	GI intolerance	Arthropathies in immature animals	Trace
Imipenem	Allergies, nephrotoxicity	None known	Trace
Isoniazid	Allergies, hepatotoxicity	Neuropathy, seizures	Yes
Linezolid	GI intolerance, marrow suppression	None known	Unknown
Rifampin	Hypersensitivity, hepatotoxicity	None known	Yes
Sulfonamides and trimethoprim-sulfamethoxazole (avoid at term)	Allergies, crystalluria	Kernicterus (at term), hemolysis (G6PD deficiency)	Yes
Vancomycin	Nephrotoxicity	None known	Yes
Class D: Unsafe; evidence of risk that may in certain circumstances be justified			
Aminoglycosides	Nephrotoxicity and ototoxicity	Ototoxicity	Yes
Erythromycin estolate	Hepatotoxicity	None known	Yes
Tetracyclines	Hepatotoxicity, renal failure	Tooth discoloration and dysplasia, impaired bone growth	Yes

*Antibiotics are listed according to FDA pregnancy class. No antibiotics are designated class A with safety established in human studies.
G6PD—glucose-6-phosphate dehydrogenase GI—gastrointestinal

breast-fed infants, chloramphenicol, fluoroquinolones, and tetracyclines should be avoided.[7]

Advanced Age

Physiologic changes that occur with age can alter the pharmacokinetics of antimicrobial agents. For example, decreased gastric acidity and intestinal motility can impair drug absorption; increased body fat and decreased serum albumin levels can alter drug distribution; and decreased hepatic blood flow and enzymatic action can delay drug metabolism. Although these factors have not consistently affected antibiotic levels in the elderly, the decrease in GFR that occurs with age can lead to the accumulation of drugs excreted by the kidney. The high therapeutic index of the penicillins and cephalosporins obviates major changes in dosage schedules in elderly patients who have normal serum creatinine levels. However, in the case of aminoglycosides and vancomycin, decreased dosage schedules are often required. Ideally, drug levels should be measured and renal function should be monitored when these agents are given. The dosage of amantadine and rimantadine should also be reduced in elderly patients.

COMPLICATIONS OF ANTIMICROBIAL THERAPY

There are three general types of complications with antimicrobial agents: hypersensitivity reactions (which are not dose related), direct drug toxicity (which usually is dose related and manifests itself in a single organ or, occasionally, in several organs), and microbial superinfection. In rare situations, antimicrobial therapy can lead to a worse clinical outcome. For example, antibiotic therapy has been associated with the development of hemolytic-uremic syndrome in children with gastroenteritis caused by *E. coli* 0157:H7.[8]

Hypersensitivity Reactions

A history of allergies should be taken before antimicrobial therapy is initiated in any patient.[9] More information is available regarding allergies to the penicillins than allergies to other agents, but skin eruptions, drug fever, and even anaphylaxis may be produced by many antibiotics. Allergic reactions occur in 1% to 10% of patients who receive penicillin. Fatal anaphylactic reactions are much less frequent [*see* Allergic Reactions to β-Lactam Antibiotics, *below*].

Direct Drug Toxicity

Although antimicrobial agents can damage virtually all human organ systems, the potential for toxicity varies widely from drug to drug.[10] The principal antibiotics that are directly toxic to the kidney are aminoglycosides and polymyxins. Patients with preexisting renal insufficiency are at increased risk for toxic reactions to various antibiotics, including nephrotoxicity, coagulopathies and other hematologic toxicities, seizures, and ototoxicity and other neurotoxicities.

Hematologic toxicity Chloramphenicol can produce bone marrow suppression, which is usually mild and reversible, but rarely presents as an irreversible fatal aplastic anemia. Chloramphenicol, sulfonamides, nitrofurantoin, and primaquine can cause hemolytic anemia in patients who have deficiencies of erythrocyte glucose-6-phosphate dehydrogenase (G6PD). Hemolytic anemia, thrombocytopenia, and leukopenia that involve an immune mechanism can be caused by penicillins, cephalosporins, tetracyclines, and rifampin, but these reactions are uncommon. Macrolides and trimethoprim-sulfamethoxazole have been associated with agranulocytosis. Trimethoprim can produce anemia, leukopenia, and thrombocytopenia from folate deficiency, which is reversible with folinic acid. Linezolid can also produce myelosuppression, which appears to be reversible with discontinuance of the drug. Patients receiving linezolid for more than 2 weeks are at greatest risk, and the complete blood count with differential should be monitored weekly in such patients. Neutropenia can occur during therapy with penicillins, cephalosporins, or vancomycin. It may be severe but is self-limited, with recovery occurring 1 to 7 days after the antibiotic is withdrawn. Penicillins inhibit platelet aggregation by adenosine diphosphate, which may account for the bleeding that occurs in some patients receiving these antibiotics in high doses. Various cephalosporins may produce coagulopathies by prolonging the prothrombin time. The methylthiotetrazole side chain present in cephalosporins such as cefotetan appears to be responsible.

Ototoxicity Antibiotics may produce a wide range of toxic effects on the central and peripheral nervous systems. Ototoxicity, either vestibular or auditory, can be produced by any of the aminoglycosides; neuromuscular blockade is much less common. Minocycline has occasionally been reported to produce significant vestibular reactions. Vancomycin can cause auditory neurotoxicity. Intravenous administration of penicillin and other β-lactams may produce seizures, especially when administered in very high doses or when given to azotemic patients or to patients with underlying epilepsy.

Neurotoxicity Peripheral neuropathy can occur as a complication of therapy with nitrofurantoin, particularly when renal failure is present. The neuropathy that occurs with isoniazid can be prevented by the daily administration of 50 mg of pyridoxine. Tetracycline may in rare instances produce reversible benign intracranial hypertension with headache and papilledema. Nalidixic acid may also produce intracranial hypertension and seizures in children. Metronidazole can sometimes cause ataxia, encephalopathy, seizures, or peripheral neuropathies. Ofloxacin has been reported to cause seizures, and mania has been attributed to clarithromycin. Optic neuritis, usually manifested by decreased visual acuity and decreased perception of the color green, may occur as a side effect of ethambutol.

Nephrotoxicity The principal antimicrobials that produce adverse effects on the liver are those used in the treatment of tuberculosis: isoniazid, rifampin, aminosalicylic acid, and pyrazinamide. Trovafloxacin was restricted for use in seriously ill patients because of hepatotoxicity, but other fluoroquinolones have not been implicated. The tetracyclines can occasionally cause fatty liver with hepatotoxicity; this is most likely to occur in patients receiving 2 g or more daily by the intravenous route. Patients receiving high-dose β-lactam antibiotics may develop hepatitis or cholestasis, presumably as a result of hypersensitivity reactions. Nitrofurantoin may cause chronic active hepatitis in some patients. Erythromycins and sulfonamides have been associated with acute hepatitis.

GI complications GI reactions to antibiotics result either from direct irritation by the drug, the occurrence of which is usually dose related, or from bacterial overgrowth.[11] Irritative GI side effects are usually produced when antibiotics are administered orally rather than parenterally. The predominant site of irritation varies from drug to drug; for example, erythromycin more commonly produces gastric irritation with epigastric distress and nausea, whereas tetracyclines may produce diarrhea and upper GI symptoms. Some qualitative and quantitative changes in the intestinal flora occur after antibiotic administration. These intestinal changes may contribute to flatulence and other lower GI symptoms, including antibiotic-associated diarrhea, which are quite common when broad-spectrum antibiotics are administered orally. An important subset of antibiotic-associated diarrhea involves the selective overgrowth of *Clostridium difficile* and resultant pseudomembranous colitis [*see 7:V Anaerobic Infections and 4:III Diarrheal Diseases*].

Other toxicities Antibiotics may cause various other toxicities. Erythromycin and other macrolides can cause prolongation of the QTc interval and polymorphic ventricular tachycardia. In rare instances, this toxicity occurs in the absence of predisposing factors, but it is more likely to occur in patients with significant heart disease and in patients taking terfenadine, astemizole, or cisapride. Several fluoroquinolones, particularly grepafloxacin, sparfloxacin, and gatifloxacin, can have similar effects on cardiac conduction.[12] All fluoroquinolones can cause tendinitis. Trimethoprim-sulfamethoxazole can cause hyperkalemia, particularly in azotemic patients. Sulfonamides, fluoroquinolones, and tetracyclines can produce photosensitivity.[13]

Microbial Superinfection

Antimicrobial therapy reduces the number of susceptible organisms from the normal flora of the skin, oropharynx, genitourinary tract, and gastrointestinal tract and exerts selective pressures that favor survival of drug-resistant organisms. Such resistant organisms can occasionally establish a superinfection (i.e., a new infection caused by a different pathogen from the one being treated) either at the site of the original infection or at remote sites.

PUBLIC HEALTH CONSIDERATIONS

Antimicrobial Resistance

The extensive use of antimicrobial agents, especially in intensive care units[14] and other health care facilities, strongly favors the selection of resistant microbial species, particularly bacterial strains harboring plasmids that confer transmissible resistance.[15-17] The widespread use of antibiotics in animal husbandry and agri-

culture compounds the problem; roughly 50% of the 25,000 tons of antibiotics that are sold annually in the United States are used in agriculture and aquaculture.[18] Infections from highly resistant strains of *Enterococcus, Streptococcus pneumoniae, Staphylococcus, Neisseria gonorrhoeae, Salmonella, Serratia, Klebsiella, Acinetobacter, Enterobacter, Pseudomonas,* and *Mycobacterium* have become important problems. Infections from resistant strains can spread rapidly—first within an institution, then throughout a community, and eventually even globally.[19] Although antimicrobial resistance is a worldwide problem, control depends on local measures, beginning with the judicious prescription of antibiotics by individual practitioners in combination with sound infection control practices. Through the Campaign to Prevent Antimicrobial Resistance in Healthcare Settings (http://www.cdc.gov/drugresistance/healthcare), the CDC offers resources for infection control. There is national concern that the relatively small number of new antimicrobial agents in development will not keep pace with the emergence of drug resistance.

Specific Antimicrobial Agents

The simultaneous use of multiple antibiotics in a shotgun fashion should be avoided because of the problems of drug toxicity and hypersensitivity reactions, microbial superinfections, and antagonisms between certain agents. Most bacterial infections can be treated satisfactorily with a single antimicrobial agent. There are a limited number of situations, however, in which the simultaneous administration of different antimicrobial agents is warranted: (1) when synergism occurs between two antimicrobials against a specific infecting agent; (2) to prevent the emergence of resistance to one or more drugs; (3) for treatment of polymicrobial infections for which one antibiotic is not sufficient; and (4) for initial empirical treatment of life-threatening infections before isolation of the etiologic agent.

Although the choice of antimicrobial drugs must always be individualized, there are useful guidelines that can be followed [*see Table 3*].[20,21]

β-LACTAM ANTIBIOTICS

Penicillins

The penicillins are bactericidal antibiotics that impair synthesis of the bacterial cell wall peptidoglycan by attaching to penicillin-binding proteins located in the cell membrane. These enzymes are responsible for linking individual elements of the bacterial cell wall together. Penicillins and other β-lactam antibiotics have different affinities for the various penicillin-binding proteins.[22]

The penicillins may be classified into subgroups on the basis of their structure, β-lactamase susceptibility, and spectrum of action. Dosages of these agents vary according to the type and severity of infection. The penicillins are generally well tolerated. Hypersensitivity and GI reactions are the most common side effects, though granulocytopenia, hemolytic anemia, bleeding, interstitial nephritis, hepatitis, and seizures may occur. The adverse effects of penicillin are generally shared by all its derivatives. However, broad-spectrum penicillins are more likely to cause pseudomembranous colitis, and penicillins with α-carboxyl substitutions, such as ticarcillin, are more likely to impair platelet function. Patients allergic to one penicillin compound are likely to be allergic to other penicillins.

Penicillin G and penicillin V Penicillin G, the first antibiotic to be used for systemic infections, is still the drug of choice for various infections. Resistance to penicillin has not been observed in group A streptococci and *Treponema pallidum*, and penicillin G is highly effective in subacute bacterial endocarditis caused by susceptible bacteria. Penicillin V (phenoxymethyl penicillin) has the same spectrum of activity as penicillin G but is more acid-stable and is thus better absorbed from the GI tract. The narrow spectra of penicillin G and V, however, limit their use to a small number of indications, and resistance to these agents is increasing, particularly with *S. pneumoniae* strains. Although all forms of penicillin introduced since the release of penicillin G are prescribed by weight, penicillin G is still prescribed for parenteral administration by unitage. For interconversion, 1 mg of penicillin G equals approximately 1,600 units. Procaine penicillin G and benzathine penicillin G are slowly absorbed, allowing dosing every 12 to 24 hours and weekly, respectively, for treatment of highly susceptible pathogens.

Penicillinase-resistant penicillins Penicillinase-resistant penicillins were developed for their activity against β-lactamase–producing *S. aureus*. Nafcillin and oxacillin are usually administered parenterally. Oxacillin is also available in an oral formulation, but cloxacillin and, especially, dicloxacillin are preferred for oral administration because of superior GI absorption. The emergence of methicillin-resistant *S. aureus* (MRSA) continues to be a major public health issue.

Penicillinase-susceptible broad-spectrum penicillins Ampicillin has a broader spectrum of activity than penicillin G. Its spectrum encompasses not only most pneumococci, meningococci, gonococci, and various streptococci but also some gram-negative bacilli. Like penicillin G, however, ampicillin is readily cleaved by β-lactamase and is useless in the treatment of infections caused by *S. aureus* or other organisms producing this enzyme. Plasmids conferring ampicillin resistance have appeared in *S. typhi*, *Haemophilus influenzae,* and *N. gonorrhoeae*. Increasing resistance has appeared in strains of *E. coli, S. pneumoniae, N. gonorrhea,* and nontyphoidal *Salmonella*.

Amoxicillin has a spectrum of activity that is identical to that of ampicillin, but amoxicillin is more efficiently absorbed from the GI tract, and effective concentrations are present in the circulation for twice as long.

Extended-spectrum carboxypenicillins The change from an amino to a carboxyl substituent on the side chain converts ampicillin to carbenicillin. Carbenicillin has an extended spectrum of activity against gram-negative bacilli, including *Pseudomonas, Proteus* species other than *P. mirabilis*, and some strains of *Enterobacte*r. β-Lactamase–producing *S. aureus* is resistant to carboxypenicillins. The indanyl carbenicillin ester is acid stable and is suitable for oral administration in the treatment of urinary tract infections caused by susceptible gram-negative bacilli, including *Pseudomonas,* but serum levels are too low for use in other infections. Ticarcillin is the other carboxypenicillin; it is administered intravenously to treat serious infections caused by susceptible organisms. Ticarcillin has an extended spectrum of activity against gram-negative bacilli, including *P. aeruginosa*, but it is ineffective against *Klebsiella*. Because of the widespread use of these antibiotics, however, many strains of *P. aeruginosa* are now resistant to the carboxypenicillins.[23]

Extended-spectrum ureidopenicillins Like the carboxypenicillins, the ureidopenicillins (i.e., azlocillin, mezlocillin, and

Table 3 Antimicrobial Drugs of Choice for Various Infections in Adults[20]

	Causative Organism	Drug of Choice*	Alternative Drugs*
Gram-Positive Cocci	*Staphylococcus aureus*		
	Methicillin-susceptible	Penicillinase-resistant penicillin[1]	A cephalosporin,[2,3] clindamycin, vancomycin, amoxicillin-clavulanate, ampicillin-sulbactam, ticarcillin-clavulanate, piperacillin-tazobactam, a carbapenem,[4] a fluoroquinolone[5]
	Methicillin-resistant[6]	Vancomycin, with or without rifampin or gentamicin	Linezolid; quinupristin-dalfopristin; daptomycin; clindamycin[7]; trimethoprim-sulfamethoxazole (TMP-SMX), minocycline,[8] or a fluoroquinolone[5] with or without rifampin
	Coagulase-negative staphylococci[9]	Vancomycin, with or without rifampin or gentamicin	Linezolid; quinupristin-dalfopristin; daptomycin; a fluoroquinolone[5] with or without rifampin
	Anaerobic streptococcus (*Peptostreptococcus*)	Penicillin G[10]	Clindamycin, a cephalosporin,[2,3] vancomycin
	Streptococcus bovis	Penicillin G[10]	A cephalosporin,[2,3] vancomycin
	Groups A, C, and G streptococci	Penicillin G[10] or penicillin V	A cephalosporin,[2,3] a macrolide,[11] clindamycin, vancomycin
	Group B streptococcus	Penicillin G,[10] ampicillin, or amoxicillin	A cephalosporin,[2,3] macrolide,[11] vancomycin
	S. pneumoniae		
	Penicillin susceptible	Penicillin G[10] or penicillin V, ampicillin, or amoxicillin	A cephalosporin,[2,3] a macrolide,[11] a fluoroquinolone,[5] a carbapenem,[4] vancomycin
	Non–penicillin susceptible	Vancomycin, ceftriaxone or cefotaxime, a fluoroquinolone,[5] a carbapenem[4]	Linezolid, quinupristin-dalfopristin
	Viridans streptococci	Penicillin G,[10] with or without gentamicin	A cephalosporin,[2,3] vancomycin
	Enterococcus		
	Penicillin-susceptible endocarditis or other serious infection	Penicillin G[10] or ampicillin + gentamicin[12] or streptomycin	Vancomycin with gentamicin[12] or streptomycin
	Uncomplicated urinary tract infection	Ampicillin or amoxicillin	Nitrofurantoin, a fluoroquinolone,[5] fosfomycin
	Vancomycin-resistant	Linezolid or quinupristin-dalfopristin	Chloramphenicol, doxycycline[8]
Gram-Positive Bacilli	*Bacillus anthracis*	Penicillin G[10]	Ciprofloxacin,[5] a tetracycline,[8] erythromycin
	B. cereus, B. subtilis	Vancomycin	Clindamycin, a carbapenem[4]
	Clostridium difficile	Metronidazole[13]	Vancomycin (oral)
	C. perfringens	Penicillin G[10] or clindamycin	Metronidazole,[13] a carbapenem,[4] chloramphenicol
	C. tetani	Metronidazole[13]	Penicillin G,[10] a tetracycline[8]
	Corynebacterium diphtheriae	Erythromycin	Penicillin G[10]
	Corynebacterium, JK group	Vancomycin	Penicillin G[10] with gentamicin, erythromycin
	Erysipelothrix rhusiopathiae	Penicillin G[10]	A cephalosporin,[2,3] a fluoroquinolone[5]
	Listeria monocytogenes	Ampicillin, with or without gentamicin	TMP-SMX
	Propionibacterium	Penicillin G[10]	Clindamycin
	Rhodococcus equi	Imipenem,[4] aminoglycosides, erythromycin, or vancomycin, with or without rifampin	Ciprofloxacin,[5] TMP-SMX, tetracycline,[8] clindamycin
Gram-Negative Cocci	*Moraxella catarrhalis*	Second- or third-generation cephalosporin[2,3] or a fluoroquinolone[5]	TMP-SMX, amoxicillin-clavulanate, a macrolide, a tetracycline[8]
	Neisseria gonorrhoeae	Ceftriaxone, cefpodoxime, or a fluoroquinolone[5]	Spectinomycin
	N. meningitidis		
	Meningitis, bacteremia	Penicillin G[10]	Ceftriaxone, cefotaxime, TMP-SMX, a fluoroquinolone,[5] chloramphenicol
	Carrier state	Rifampin	Minocycline,[8] ciprofloxacin[5]
Enteric Gram-Negative Bacilli	*Bacteroides fragilis*	Metronidazole	Clindamycin, cefoxitin, cefotetan, cefmetazole, a carbapenem,[4] ampicillin-sulbactam, amoxicillin-clavulanate, ticarcillin-clavulanate, piperacillin-tazobactam, chloramphenicol
	Campylobacter jejuni	Erythromycin or azithromycin	A fluoroquinolone,[5] a tetracycline,[8] gentamicin
	Citrobacter	A carbapenem[4]	A fluoroquinolone,[5] amikacin, TMP-SMX, a third- or fourth-generation cephalosporin[2,3]
	Enterobacter	A carbapenem[4]	A third- or fourth-generation cephalosporin,[2,3] a fluoroquinolone,[5] ticarcillin-clavulanate, piperacillin-tazobactam, aztreonam, gentamicin
	Escherichia coli		
	Serious infection	A third- or fourth-generation cephalosporin[2,3]	A carbapenem,[4] ampicillin-sulbactam, amoxicillin-clavulanate, ticarcillin-clavulanate, piperacillin-tazobactam, a fluoroquinolone,[5] an aminoglycoside, aztreonam
	Uncomplicated cystitis	TMP-SMX or a fluoroquinolone[5]	Ampicillin or amoxicillin
	Helicobacter pylori	Omeprazole + amoxicillin + clarithromycin; or tetracycline[8] + metronidazole + bismuth subsalicylate	Tetracycline[8] + clarithromycin + bismuth subsalicylate; amoxicillin + metronidazole[13] + bismuth subsalicylate; amoxicillin + clarithromycin

Note: all superscript numbers refer to footnotes that follow table.

(*continued*)

Table 3 (*continued*)

	Causative Organism	Drug of Choice*	Alternative Drugs*
Enteric Gram-Negative Bacilli (continued)	*Klebsiella*	A third- or fourth-generation cephalosporin[2,3]	A carbapenem,[4] ampicillin-sulbactam, amoxicillin-clavulanate, ticarcillin-clavulanate, piperacillin-tazobactam, a fluoroquinolone,[5] TMP-SMX, an aminoglycoside, aztreonam
	Proteus mirabilis	Ampicillin or amoxicillin	A cephalosporin,[2,3] a fluoroquinolone,[5] TMP-SMX, a carboxypenicillin or ureidopenicillin,[14] a carbapenem,[4] an aminoglycoside, aztreonam
	Non-*mirabilis*, including *P. vulgaris*, *Morganella morganii*, and *Providencia rettgeri*	A third- or fourth-generation cephalosporin[2,3]	A carbapenem,[4] ampicillin-sulbactam, amoxicillin-clavulanate, ticarcillin-clavulanate, piperacillin-tazobactam, a fluoroquinolone,[5] an aminoglycoside, aztreonam
	Providencia stuartii	A third- or fourth-generation cephalosporin[2,3]	A carbapenem,[4] ticarcillin-clavulanate, piperacillin-tazobactam, a fluoroquinolone,[5] TMP-SMX, an aminoglycoside, aztreonam
	Salmonella typhi	Ceftriaxone or a fluoroquinolone[5]	Chloramphenicol, TMP-SMX, ampicillin
	Other *Salmonella* species	Ceftriaxone or cefotaxime or a fluoroquinolone[5]	Amoxicillin, TMP-SMX, chloramphenicol
	Serratia	A carbapenem[4]	A third- or fourth-generation cephalosporin,[2,3] ticarcillin-clavulanate, piperacillin-tazobactam, a fluoroquinolone,[5] TMP-SMX, an aminoglycoside, aztreonam
	Shigella	A fluoroquinolone[5]	TMP-SMX, ampicillin, ceftriaxone, azithromycin
	Yersinia enterocolitica	TMP-SMX	A fluoroquinolone,[5] cefotaxime, ceftriaxone, an aminoglycoside
Other Gram-Negative Bacilli	*Acinetobacter*	Imipenem or meropenem[4]	An aminoglycoside, sulbactam (available as ampicillin-sulbactam)
	Aeromonas hydrophila	TMP-SMX	A fluoroquinolone,[5] an aminoglycoside, a carbapenem[4]
	Bartonella henselae (cat-scratch disease)	Azithromycin	Ciprofloxacin,[5] TMP-SMX, gentamicin, rifampin, erythromycin
	B. henselae (bacillary angiomatosis)	Erythromycin	Doxycycline,[8] azithromycin
	Bordetella pertussis (whooping cough)	Erythromycin	Azithromycin or clarithromycin, TMP-SMX
	Brucella	A tetracycline,[8] with rifampin	A tetracycline[8] with gentamicin or streptomycin; chloramphenicol, with or without streptomycin; TMP-SMX with or without gentamicin; ciprofloxacin[5] with rifampin
	Burkholderia cepacia	TMP-SMX	Imipenem or meropenem,[4] ceftazidime, chloramphenicol
	B. pseudomallei (melioidosis)	Imipenem or meropenem[4]	Ceftazidime, amoxicillin-clavulanate, chloramphenicol + doxycycline[8] + TMP-SMX
	Calymmatobacterium granulomatis (granuloma inguinale)	TMP-SMX	A tetracycline[8] or fluoroquinolone[5] with or without gentamicin
	Capnocytophaga canimorsus	Penicillin G[10]	Cefotaxime or ceftriaxone, a carbapenem,[4] a fluoroquinolone,[5] clindamycin
	Eikenella corrodens	Ampicillin	A macrolide, a tetracycline,[8] amoxicillin-clavulanate, ampicillin-sulbactam, ceftriaxone
	Francisella tularensis (tularemia)	Streptomycin	Gentamicin, a tetracycline,[8] chloramphenicol, ciprofloxacin[5]
	Fusobacterium	Penicillin G[10]	Metronidazole, clindamycin, cefoxitin, chloramphenicol
	Gardnerella vaginalis (bacterial vaginosis)	Metronidazole (oral)[13]	Intravaginal metronidazole, intravaginal or oral clindamycin
	Haemophilus ducreyi (chancroid)	Azithromycin or ceftriaxone	Ciprofloxacin[5] or erythromycin
	H. influenzae Bronchitis, otitis media	TMP-SMX	Ampicillin or amoxicillin, amoxicillin-clavulanate, a macrolide, a fluoroquinolone,[5] cefuroxime, cefotaxime, ceftriaxone
	Meningitis, epiglottitis, life-threatening infections	Cefotaxime or ceftriaxone	Chloramphenicol, meropenem
	Legionella species	Fluoroquinolone[5] or azithromycin, with or without rifampin	Erythromycin, doxycycline,[8] TMP-SMX
	Leptotricia buccalis	Penicillin G[10]	A tetracycline,[8] clindamycin, erythromycin
	Pasteurella multocida	Penicillin G[10]	A tetracycline,[8] a cephalosporin,[2,3] amoxicillin-clavulanate, ampicillin-sulbactam
	Pseudomonas aeruginosa Urinary tract infections	Ciprofloxacin[5]	Levofloxacin,[5] a carboxypenicillin or ureidopenicillin,[14] ceftazidime, cefepime, imipenem or meropenem,[4] an aminoglycoside, aztreonam
	Other infections	Ceftazidime, cefepime, imipenem or meropenem,[4] a carboxypenicillin or ureidopenicillin[14] with or without tobramycin or gentamicin	Ciprofloxacin[5], amikacin
	Stenotrophomonas maltophila	TMP-SMX	Minocycline,[8] ticarcillin-clavulanate, a fluoroquinolone[5]
	Streptobacillus moniliformis (rat-bite fever)	Penicillin G[10]	A tetracycline,[8] streptomycin
	Vibrio cholerae	A tetracycline[8]	A fluoroquinolone,[5] TMP-SMX
	V. vulnificus	A tetracycline[8]	Cefotaxime
	Yersinia pestis (plague)	Streptomycin with or without a tetracycline[8]	A tetracycline,[8] chloramphenicol, gentamicin, TMP-SMX

Note: all superscript numbers refer to footnotes that follow table.

(*continued*)

Table 3 (*continued*)

	Causative Organism	*Drug of Choice**	*Alternative Drugs**
Acid-Fast Bacilli	*Mycobacterium avium* complex	Clarithromycin or azithromycin with ethambutol, with or without rifabutin	Amikacin, ciprofloxacin[5]
	Mycobacterium fortuitum/chelonae complex	Amikacin with clarithromycin	Cefoxitin, rifampin, a sulfonamide, doxycycline,[8] ethambutol, linezolid
	Mycobacterium kansasii	Isoniazid with rifampin, with or without ethambutol or streptomycin	Clarithromycin or azithromycin, ethionamide, cycloserine
	Mycobacterium leprae	Dapsone with rifampin, with or without clofazimine	Minocycline,[8] ofloxacin or sparfloxacin, clarithromycin
	Mycobacterium marinum	Minocycline[8]	TMP-SMX, rifampin, clarithromycin, doxycycline,[8] ethambutol
	Mycobacterium tuberculosis	Isoniazid with rifampin, pyrazinamide, and ethambutol	Streptomycin, a fluoroquinolone,[5] cycloserine, ethionamide, para-aminosalicylic acid, capreomycin, kanamycin, amikacin
Actino-mycetes	*Actinomyces israelii*	Penicillin G[10]	A tetracycline,[8] erythromycin, clindamycin
	Nocardia	TMP-SMX	Minocycline,[8] sulfisoxazole, imipenem or meropenem,[4] amikacin, cycloserine, linezolid
Chlamydia	*Chlamydia psittaci* (psittacosis)	A tetracycline[8]	Chloramphenicol
	C. trachomatis		
	Inclusion conjunctivitis	Erythromycin	A sulfonamide (topical plus oral)
	Lymphogranuloma venereum	A tetracycline[8]	Erythromycin
	Pneumonia	Erythromycin	A sulfonamide
	Trachoma	Azithromycin	A tetracycline[8] (topical plus oral), a sulfonamide (topical plus oral)
	Urethritis or pelvic inflammatory disease	Doxycycline[8] or azithromycin	Erythromycin, ofloxacin, amoxicillin
	C. pneumoniae	A macrolide or tetracycline[8]	A fluoroquinolone[5]
Ehrlichia	*Ehrlichia chaffeensis*	Doxycycline[8]	Chloramphenicol
	E. ewingii	Doxycycline[8]	—
	E. phagocytophila	Doxycycline[8]	Rifampin
Myco-plasma	*Mycoplasma pneumoniae*	A macrolide or tetracycline[8]	A fluoroquinolone[5]
	Ureaplasma urealyticum	Erythromycin	A tetracycline,[8] clarithromycin, azithromycin, ofloxacin
Rickettsia	Various rickettsial organisms Rocky Mountain spotted fever, epidemic and endemic (murine) typhus, rickettsial pox, Q fever, scrub typhus	Doxycycline[8]	Chloramphenicol, a fluoroquinolone,[5] rifampin
Spirochetes	*Borrelia burgdorferi* (Lyme disease)	Doxycycline,[8] amoxicillin, ceftriaxone	Penicillin G,[10] azithromycin, clarithromycin
	B. recurrentis (relapsing fever)	A tetracycline[8]	Penicillin G[10]
	Leptospira	Penicillin G[10]	A tetracycline[8]
	Treponema pallidum (syphilis)	Penicillin G[10]	A tetracycline,[8] ceftriaxone

*These are general recommendations; susceptibility data should be used to guide therapy. Empirical therapy should be based on knowledge of local susceptibility patterns when possible.

Note: all superscript numbers in table refer to footnotes below.

1. For severe infections, I.V. nafcillin or oxacillin is preferred. For mild infections, oral cloxacillin or dicloxacillin may be used. High doses of penicillin G, ampicillin, amoxicillin, carbenicillin, or ticarcillin do not overcome the clinical resistance of penicillinase-producing staphylococci to these drugs. The combinations of ampicillin with sulbactam, amoxicillin with clavulanate, ticarcillin with clavulanate, and piperacillin with tazobactam have activity against β-lactamase–producing *S. aureus*.

2. Cephalosporins are sometimes used as alternatives to penicillin in patients with suspected penicillin allergy but should not be used in patients with severe hypersensitivity reactions such as anaphylaxis or urticaria.

3. For treatment of staphylococcal or nonenterococcal streptococcal infections, a first-generation cephalosporin such as cefazolin (I.V.) or cephalexin (p.o.) may be used. Cefoxitin, cefotetan, and cefmetazole have increased anaerobic activity, including against *B. fragilis*. Ceftazidime and cefepime are active against *P. aeruginosa*. (See text for further details.)

4. Imipenem tends to be more active against gram-positive cocci, whereas meropenem is more active against gram-negative bacilli. Ertapenem has decreased activity against *Acinetobacter*, *Pseudomona*s, and *E. faecalis*. None of the carbapenems are active against methicillin-resistant *S. aureus*, *E. faecium*, or *Stenotrophomonas*.

5. Levofloxacin, gatifloxacin, gemifloxacin, and moxifloxacin have increased activity against *S. pneumoniae*, including penicillin-resistant strains. They also have good activity against *S. aureus*, but resistance among methicillin-resistant strains is common. Ciprofloxacin is most active against *Pseudomonas*. For *M. tuberculosis*, levofloxacin, ofloxacin, ciprofloxacin, gatifloxacin, or moxifloxacin could be used. Fluoroquinolones are not recommended for children or pregnant women. (See text for further details.)

6. These strains are also resistant to other β-lactams, including cephalosporins and carbapenems.

7. Isolate should be examined for inducible resistance to clindamycin before this agent is used.

8. Tetracycline should not be used in pregnant women or children younger than 9 years.

9. In vitro susceptibility testing with cephalosporins or penicillins may be misleading because of heteroresistance and because these antibiotics may be bacteriostatic only. For serious infections, vancomycin is preferred.

10. Crystalline penicillin G is administered parenterally for serious infections. For less severe infections caused by pneumococci, group A streptococci, or T. *pallidum*, procaine penicillin is administered I.M. once or twice daily. For mild infections caused by streptococci and pneumococci, oral penicillin V is preferable to oral penicillin G. Benzathine penicillin G is given I.M. (once monthly for the prophylaxis of rheumatic fever, a single injection for the treatment of group A streptococcal pharyngitis) when patient's compliance with oral medication is questionable and for treatment of syphilis, in one to three doses at weekly intervals, depending on the stage of the disease. Penicillin-resistance is increasing in *S. pneumoniae* strains.

11. Macrolide resistance is increasing among strains of *S. pneumoniae* and group A streptococci.

12. Various aminoglycosides have been used in synergistic combination with penicillin or vancomycin. Because of the appearance of enterococcal strains resistant to the synergistic action with streptomycin (but not gentamicin), gentamicin is preferred for use in the combination.

13. Antibiotics may be administered orally for antibiotic-associated pseudomembranous enterocolitis. Vancomycin and metronidazole are equally effective, but metronidazole is much less expensive.

14. The carboxypenicillins are carbenicillin and ticarcillin; the ureidopenicillins are mezlocillin, azlocillin, and piperacillin. When one of these drugs is used for a severe infection, an aminoglycoside is often recommended, as well.

piperacillin) are ampicillin derivatives. They are active against most organisms that are susceptible to ampicillin, including many pneumococci, most streptococci, *N. meningitidis,* and most *E. coli* and *P. mirabilis* strains. They have extended activity against *P. aeruginosa,* although the prevalence of resistance is rising. β-Lactamase–producing staphylococci and *H. influenzae* are resistant to the ureidopenicillins.

The carboxypenicillins and ureidopenicillins also differ from ampicillin in their increased effectiveness against many anaerobes, including about 50% of *Bacteroides fragilis* strains. However, the major role of these drugs depends on their spectrum of activity against resistant gram-negative bacilli, which is somewhat broader for the ureidopenicillins than it is for the carboxypenicillins. Despite these in vitro differences, clinical studies have not demonstrated that the ureidopenicillins are superior to the carboxypenicillins in treating organisms susceptible to both groups. The newer cephalosporins and carbapenems, however, appear preferable for treating infections caused by resistant *Klebsiella* and *Serratia* strains. Because of the concern about emerging resistance, these drugs should generally be used with an aminoglycoside to treat seriously ill patients, especially patients with neutropenia.

The serum levels, tissue distribution, half-life, and recommended dosage ranges of piperacillin, mezlocillin, and azlocillin are similar to those of ticarcillin. Unlike ticarcillin and carbenicillin, which are solely excreted by the kidneys, the ureidopenicillins are excreted in the bile and the urine and do not accumulate in patients with renal failure. The toxicities of these drugs are similar, except that the ureidopenicillins are less likely to impair platelet function or cause hypokalemia.

Penicillin–β-lactamase inhibitor combinations The major mechanism of resistance to the penicillins is bacterial production of β-lactamase enzymes that hydrolyze the β-lactam ring, rendering the molecule inactive. Clavulanate, sulbactam, and tazobactam are β-lactam compounds that have little intrinsic antibacterial activity except for sulbactam in *A. baumannii* infections.[24] They do, however, bind irreversibly to the β-lactamase enzymes that are produced by many bacteria, thus inactivating these enzymes and rendering the organisms susceptible to β-lactamase–susceptible penicillins.

Clavulanate is combined with amoxicillin in an oral formulation. Ampicillin-sulbactam, ticarcillin-clavulanate, and piperacillin-tazobactam are parenteral formulations. The combinations are active against ampicillin-susceptible organisms and various ampicillin-resistant organisms, including β-lactamase–producing strains of *Moraxella catarrhalis, H. influenzae, E. coli, K. pneumoniae,* some *Proteus* species, and *S. aureus* (except methicillin-resistant strains). The combinations are also active against many anaerobes, including *B. fragilis*. However, *P. aeruginosa* and many strains of *Serratia* and *Enterobacter* are resistant to amoxicillin-clavulanate and ampicillin-sulbactam. These agents are not active against bacteria resistant to the penicillin derivative when the resistance is not mediated by β-lactamase production. For example, penicillin-resistant strains of *S. pneumoniae* have altered penicillin-binding proteins and are not affected by the addition of a β-lactamase inhibitor.

Amoxicillin-clavulanate therapy has been used successfully to treat upper and lower respiratory tract infections, urinary tract infections, and human and animal bites. Twice-daily administration is effective. The ampicillin-sulbactam combination has been used successfully to treat gynecologic and intra-abdominal infections, as well as infections of the upper and lower respiratory tracts, urinary tract, skin and soft tissues, and bones and joints. Ticarcillin-clavulanate has been successful in the treatment of pulmonary, urinary tract, bone, soft tissue, and bloodstream infections. Piperacillin-tazobactam is very similar to ticarcillin-clavulanate in its very broad spectrum of activity, pharmacokinetics, toxicities, and expense. The recommended dosage of 3 g of piperacillin and 375 mg of tazobactam every 6 hours for adults with normal renal function is lower than the recommended dosage for piperacillin alone (18 g/day) and may not be sufficient for some serious infections caused by *P. aeruginosa*. An increased dose of 4 g of piperacillin and 500 mg of tazobactam is approved for nosocomial pneumonia when there is concern for infection from *Pseudomonas* species.

Cephalosporins

The cephalosporins and closely related cephamycins (e.g., cefoxitin and cefotetan) comprise a large family of β-lactam antibiotics. Like the penicillins, the cephalosporins are bactericidal antibiotics that inhibit bacterial cell wall synthesis and have a low intrinsic toxicity. The adverse effects of the cephalosporins are mainly hypersensitivity reactions, local pain (with intramuscular use), and thrombophlebitis (with intravenous use). Less common toxicities include GI symptoms, elevated liver enzyme levels, and renal impairment. Third-generation and fourth-generation cephalosporins may cause seizures, including nonconvulsive status epilepticus, in patients with renal failure.[25] Pseudomembranous colitis may also develop. Because the cephalosporins share immunologic cross-reactivity, patients who are allergic to one cephalosporin are likely to be allergic to others. There is also a possibility of cross-reactivity in penicillin-allergic patients.[26]

The cephalosporins are grouped into generations on the basis of their antibacterial spectrum[27] [*see Table 4*]. In general, activity against gram-positive cocci diminishes from the first generation to the third generation, whereas the spectrum of activity against gram-negative organisms increases. The single fourth-generation agent, cefepime, has an extended spectrum of activity against both gram-positive and gram-negative organisms. Agents in each group exhibit pharmacologic differences in serum levels and half-lives, leading to substantial variation in dosing. Most cephalosporins are excreted primarily by the kidneys and require dose reduction in the presence of renal failure, with the notable exceptions of ceftriaxone and cefoperazone [*see 10:IX Pharmacologic Approach to Renal Insufficiency, Appendix A*]. Cephalosporins cross the placenta and penetrate the pericardium and joints. The third- and fourth-generation cephalosporins have good cerebrospinal fluid penetration and are effective in treating meningitis, whereas none of the first-generation drugs and only one of the second-generation agents (cefuroxime) are effective in treating meningitis.

First-generation cephalosporins The first-generation cephalosporins are active against many gram-positive bacteria, including penicillin-susceptible pneumococci, penicillinase-producing *S. aureus,* and most streptococci. However, they are ineffective against MRSA, coagulase-negative staphylococci, penicillin-resistant pneumococci, enterococci, and *Listeria*. Although the first-generation cephalosporins are active against many strains of *E. coli, K. pneumoniae,* and *P. mirabilis,* they are ineffective against many other gram-negative species because these organisms produce β-lactamases.

Table 4 Properties of Cephalosporin Antibiotics

	Specific Agent	Trade Names	Comments*
First Generation	Oral		
	Cefadroxil	Duricef, Ultracef	Longer half-life
	Cephalexin	Keflex	Most experience with this agent
	Cephradine	Anspor, Velosef	Properties are similar to those of cephalexin
	Parenteral		
	Cefazolin	Ancef, Kefzol	Longer half-life; well tolerated when given I.M.
	Cephapirin	Cefadyl	Properties are similar to those of other first-generation cephalosporins
	Cephradine	Anspor, Velosef	Properties are similar to those of other first-generation cephalosporins
Second Generation	Oral		
	Cefaclor	Ceclor	Moderately active against *Haemophilus influenzae*
	Cefuroxime axetil	Ceftin	Active against *H. influenzae*
	Cefprozil	Cefzil	Active against *H. influenzae*
	Cefditoren pivoxil	Spectracef	Active against *H. influenzae*
	Loracarbef	Lorabid	A carbacephem with properties and spectrum similar to cefuroxime
	Parenteral		
	Cefamandole	Mandol	Active against *H. influenzae*; may cause bleeding
	Cefmetazole	Zefazone	Spectrum and half-life similar to cefoxitin; may cause bleeding
	Cefonicid	Monocid	Spectrum similar to that of cefamandole
	Ceforanide	Precef	Spectrum similar to that of cefamandole
	Cefotetan	Cefotan	Spectrum similar to that of cefoxitin; longer half-life than cefoxitin; may cause bleeding
	Cefoxitin	Mefoxin	Active against anaerobes, including *Bacteroides fragilis*
	Cefuroxime	Zinacef, Kefurox	Active against *H. influenzae*; only second-generation drug approved for meningitis (selected pathogens)
Third and Fourth Generation	Oral		
	Cefixime	Suprax	More active against gram-negative bacilli, gonococci, *Moraxella catarrhalis*, and *H. influenzae* than other oral cephalosporins but much less active against *Staphylococcus aureus*; not active against *Pseudomonas*; currently not available in the United States
	Cefdinir	Omnicef	Similar to cefpodoxime
	Cefpodoxime	Vantin	Similar to cefixime but more active against *S. aureus*
	Ceftibuten	Cedax	Similar to cefixime except poor activity against pneumococci and staphylococci
	Parenteral		
	Cefepime	Maxipime	Fourth-generation cephalosporin active against most gram-positive cocci (except enterococci and methicillin-resistant staphylococci), *Neisseria*, *Haemophilus*, enteric gram-negative bacilli, and *Pseudomonas*
	Cefoperazone	Cefobid	Increased activity against *P. aeruginosa* but less against Enterobacteriaceae; may cause bleeding
	Cefotaxime	Claforan	More active against gram-positive cocci
	Ceftazidime	Fortaz, Tazidime, Tazicef	More active against *Pseudomonas*
	Ceftizoxime	Cefizox	Properties are similar to those of cefotaxime
	Ceftriaxone	Rocephin	Longer half-life; less active against *Pseudomonas*, *B. fragilis*

Note: Detailed information about the various cephalosporins is covered in the text.
*Agents are being compared with other members of the same generation of cephalosporins.

Second-generation cephalosporins The second-generation cephalosporins consist of two groups: the true cephalosporins (such as cefuroxime) and the cephamycins (such as cefoxitin, cefotetan, and cefmetazole). Because they are not as vulnerable to the β-lactamases that are produced by many of the gram-negative bacteria, the second-generation cephalosporins have enhanced activity against gram-negative bacilli, including many strains *of E. coli, Klebsiella, Serratia,* and *Proteus* that are resistant to the first-generation cephalosporins. However, second-generation cephalosporins are not effective against *Pseudomonas* or *Enterobacter* species. The second-generation cephalosporins retain some activity against gram-positive organisms that are susceptible to first-generation cephalosporins. The cephamycins have enhanced anaerobic activity, including activity against *B. fragilis,* and are effective in intra-abdominal and pelvic infections; however, these agents have largely been supplanted by combination therapy with pencillin and a β-lactamase inhibitor and by the carbapenems.

Third- and fourth-generation cephalosporins The third-generation cephalosporins differ from other cephalosporins in some important respects. Their enhanced ability to resist hydrolysis by the β-lactamases of many gram-negative bacilli gives them an expanded antibacterial spectrum. Pharmacologically, these drugs also have an important advantage: unlike older cephalosporins, they achieve therapeutic levels in the cerebrospinal fluid and can be used to treat meningitis. These advantages, however, come at an increased cost, in that there is an emerging resistance to these agents and that their effect on the normal flora causes increased colonization with vancomycin-resistant *Enterococcus* (VRE).

Although third-generation agents are less active than the older cephalosporins against many gram-positive cocci,[28] they are active against most penicillin-nonsusceptible pneumococci. Like other cephalosporins, the third-generation agents are inactive against enterococci, MRSA, and *Listeria*; however, they have enhanced potency against many gram-negative bacilli, including *E.*

coli, Klebsiella, Proteus, Serratia, and *Citrobacter* organisms. They are also very active against penicillinase-producing and non–penicillinase-producing strains of *H. influenzae* and gonococci. The third-generation cephalosporins are active against most *Salmonella* species and have been clinically effective in the treatment of typhoid fever and other *Salmonella* infections. Resistant strains are now beginning to emerge in the United States, however.[29] Although most *Enterobacter* species initially appear susceptible to third-generation cephalosporins, they may rapidly develop resistance during therapy as a result of inducible cephalosporinases.[1] Hence, it may be prudent to avoid using a cephalosporin as the sole therapy for these organisms. Similarly, *Klebsiella* species may appear susceptible by disk-diffusion testing but may harbor an ESBL, which renders therapy with a third-generation cephalosporin ineffective.[2] Increased zone size around the combined ceftazidime-clavulanate disk relative to the zone size of ceftazidime alone is indicative of the presence of ESBLs.

Although third-generation cephalosporins exhibit activity against *B. fragilis* and other anaerobes, the cephamycins cefoxitin, cefotetan, and cefmetazole are more active. Ceftazidime has the greatest activity against *P. aeruginosa*. Cefoperazone also demonstrates good activity, but the other third-generation cephalosporins have been disappointing. Other nosocomial gram-negative pathogens that are resistant to these drugs include *Acinetobacter* and *Stenotrophomonas*.

Cefepime is a fourth-generation cephalosporin with broad antimicrobial activity against both aerobic gram-positive bacteria (e.g., penicillin-nonsusceptible *S. pneumoniae*) and methicillin-susceptible *S. aureus*; it is also effective against gram-negative bacteria, including *H. influenzae, Neisseria,* and Enterobacteriaceae. Its activity against *Pseudomonas* is similar to that of ceftazidime. Compared to the third-generation cephalosporins, cefepime is a weaker inducer of AmpC β-lactamase and is more stable in the presence of this enzyme.[30] The clinical significance of this finding remains to be determined.

Carbapenems

Imipenem and meropenem were the first carbapenems available for clinical use in the United States; the third, ertapenem, received Food and Drug Administration approval in 2002. Like other β-lactam antibiotics, they are bactericidal and act by inhibiting bacterial cell wall synthesis.[31] Three properties account for the extraordinarily broad antibacterial spectrum of the carbapenems: the ability to penetrate the cell membrane of gram-negative bacteria; high affinity for critical penicillin-binding proteins; and resistance to hydrolysis by β-lactamases.

Imipenem is extensively hydrolyzed in the renal tubule, which results in low urinary levels of the drug and the production of a nephrotoxic metabolite. Cilastatin prevents this degradation by inhibiting the brush-border enzyme dehydropeptidase-1; cilastatin and imipenem are administered simultaneously in equal doses. Meropenem and ertapenem are not susceptible to degradation by dehydropeptidase and can therefore be administered without cilastatin. Ertapenem has a longer half-life than the other carbapenems, allowing for a single daily dose; imipenem, on the other hand, requires dosing every 6 hours, and meropenem, every 8 hours.[32] The carbapenems are primarily excreted in the urine; the dosage should be reduced in azotemic patients.

The carbapenems have broader antibacterial spectra than any other β-lactam antibiotics; they are effective against most clinically important gram-positive and gram-negative bacteria, including anaerobes. Whereas imipenem tends to be more active against gram-positive cocci, meropenem appears to be more active against gram-negative bacilli. However, neither drug is active against methicillin-resistant staphylococci, *E. faecium,* or *Stenotrophomonas*. Ertapenem has a similarly broad spectrum of coverage; however, compared to the other carbapenems, ertapenem has decreased activity against *Pseudomonas, Acinetobacter,* and *E. faecalis*.

Carbapenems are extraordinarily active against gram-negative bacteria. Virtually all Enterobacteriaceae are susceptible. *Haemophilus* and *Neisseria* species are also susceptible to carbapenems but at drug concentrations somewhat higher than those of third-generation cephalosporins. *Acinetobacter,* which is resistant to most other β-lactam antibiotics, is usually susceptible to imipenem and meropenem, although resistance appears to be increasing. All three carbapenems retain activity against most bacterial strains expressing the inducible AmpC β-lactamase enzyme (e.g., *Serratia, Citrobacter,* and *Enterobacter*), as well as organisms containing ESBL; in fact, carbapenems are the drugs of choice for infections caused by such organisms. Gram-negative anaerobes, including *B. fragilis,* are susceptible. *P. aeruginosa* is generally susceptible to imipenem and meropenem, but resistance has emerged with increased carbapenem use.

The safety of carbapenems seems comparable to that of other β-lactam antibiotics. Nausea and vomiting, local pain at injection sites, and hypersensitivity are the most common reactions. Seizures, although unusual, are a potential concern with imipenem. They have been observed in 0.9% of patients who have received the drug; risk factors for seizures include excessive dosage, preexisting CNS lesions, epilepsy, and renal insufficiency. Meropenem is less likely to provoke seizures. Transient elevations of liver enzymes and leukopenia can occur in patients who are given carbapenems. Antibiotic-associated pseudomembranous colitis has occurred.

Because the structure of the carbapenems resembles that of the penicillins and cephalosporins, there is potential for cross-reactivity in patients allergic to other β-lactam antibiotics. Clinical experience in this situation is limited, but it appears prudent to avoid carbapenems in patients who experience anaphylactic reactions to β-lactam drugs and to use carbapenems with caution in patients who have milder allergies to penicillins or cephalosporins.[33]

Carbapenems have been used successfully in patients with pneumonia, intra-abdominal infections, complicated skin and soft tissue infections, complicated urinary tract infections, endocarditis, bacteremia, osteomyelitis, and febrile neutropenia. The broad spectrum and apparent low toxicity of carbapenems are impressive, but they should be used with restraint and selectivity.

Monobactams

The monobactams are monocyclic β-lactam antibiotics that lack the thiazolidine ring found in penicillins and the dihydrothiazine ring found in cephalosporins.[31] Aztreonam is currently the only available monobactam. The antibacterial activity of aztreonam depends on its ability to penetrate the outer membrane of gram-negative bacilli, as well as on its high affinity for penicillin-binding protein–3 and its resistance to hydrolysis by the β-lactamases of gram-negative bacilli.

The antibacterial activity of aztreonam is restricted to aerobic or facultative aerobic gram-negative bacteria. Most strains of *H. influenzae,* gonococci, and meningococci are susceptible to aztreonam, as are most enteric gram-negative bacilli, including *E. coli,*

Klebsiella, Proteus, and *Enterobacter*. Most strains of *P. aeruginosa* are also susceptible, but somewhat higher concentrations of aztreonam are required to kill these bacteria. *Acinetobacter, S. maltophila,* and *B. cepacia* are generally resistant to aztreonam, as are all gram-positive bacteria and anaerobes.

Excellent serum levels are achieved after intramuscular or intravenous administration, and the drug is widely distributed in body tissues and fluids, including the CSF. The serum half-life of aztreonam is about 90 minutes, and glomerular filtration is the major means by which the drug is eliminated. The dosage of aztreonam must be reduced in azotemic patients. The drug is efficiently removed by hemodialysis but not by peritoneal dialysis.

Aztreonam is well tolerated. Its toxicities resemble those of other β-lactam antibiotics; these toxicities include occasional instances of local reactions at the site of injection, rash, diarrhea, nausea, and vomiting. Neither nephrotoxicity nor ototoxicity has been reported. Aztreonam does not cross-react with serum antibodies of penicillin-allergic and cephalosporin-allergic patients, and the drug has been well tolerated in penicillin-allergic patients.

Aztreonam has been used successfully in the treatment of a wide variety of infections caused by gram-negative bacteria, including pneumonias, skin and soft tissue infections, bone and joint infections, urinary tract infections, and bacteremias. Although aztreonam achieves therapeutic concentrations in the CSF, experience in the treatment of meningitis has been very limited.

Because aztreonam combines the spectrum of activity of the aminoglycosides with the low toxicity of the β-lactam antibiotics, it is an attractive alternative for many gram-negative infections. Synergy can be demonstrated between aztreonam and aminoglycosides for some gram-negative bacilli; however, it is not synergistic with other β-lactam agents. More clinical experience is needed to determine the circumstances in which aztreonam should be substituted for or used as a supplement to aminoglycosides.

ALLERGIC REACTIONS TO β-LACTAM ANTIBIOTICS

Hypersensitivity is the most common adverse reaction to β-lactam antibiotics [*see 2:IV Cutaneous Adverse Drug Reactions and 6: XIV Drug Allergies*]. Most often, it is a delayed reaction characterized by maculopapular eruptions, fever, or both. Eosinophilia may also be present. Much less common but much more serious is immediate hypersensitivity, mediated by IgE. Manifestations may include early-onset urticaria, laryngeal edema, or anaphylaxis. Immune complexes can produce serum sickness in some patients. Hypersensitivity to β-lactam antibiotics can also cause hemolytic anemia or allergic interstitial nephritis.

Establishing Drug Allergy

As many as 10% of all patients report a history of penicillin allergy, but only 10% to 20% of those reporting allergy are actually at risk for immediate hypersensitivity reactions.[34] A careful history is the best way to establish true drug allergy. A patient's recollection of rash, urticaria, arthralgias, wheezing, or anaphylaxis confirms the diagnosis of hypersensitivity, whereas treatment failure, diarrhea, vaginitis or other superinfections, or vague symptoms do not. Skin tests have a limited role in predicting true penicillin allergy, but they can help exclude IgE-mediated (anaphylactic) hypersensitivity.[35,36] To document severe allergy, both benzyl penicilloylpolylysine (a major-determinant antigen) and the so-called minor-determinant antigens should be injected; only the former is available commercially, but both are available to most experienced allergists. A wheal-and-flare reaction signifies IgE-mediated allergies; negative reactions to both major and minor determinants make anaphylaxis unlikely, but small test doses of penicillin should be administered for additional safety. Facilities and equipment necessary for the treatment of anaphylaxis should always be on hand when skin testing is performed. Patients with positive skin-test responses have a 41% to 67% chance of exhibiting significant penicillin allergy if they are rechallenged with the drug. The risk to patients with negative skin-test responses is only 1% to 4%, and no life-threatening reactions have been reported.

Desensitization Technique

Patients with penicillin allergy, whether documented by history or by skin testing, should receive other antibiotics for infections. In rare cases in which there is no acceptable alternative to a penicillin (e.g., syphilis in a pregnant woman), desensitization can be attempted. Desensitization is carried out with the equipment and medications for the treatment of anaphylaxis at the bedside: a laryngoscope, oxygen, an endotracheal tube, epinephrine (1:1,000 weight in volume), sodium bicarbonate solution, and a running intravenous infusion. Penicillin doses are administered in graded increases in the forearm at a site low enough for a tourniquet to be applied proximally should a reaction occur. The initial dose of 1 unit of penicillin is applied by scratch test. If there is no wheal-and-flare reaction within 15 minutes, a dose of 2 units is injected intradermally. If no local or systemic reactions occur after another 15 minutes, a dose of 5 units is injected intradermally. If no reactions occur, successive doses that approximately double each previous dose are injected intradermally at 15-minute intervals. When the amount injected becomes large, the penicillin dose is administered subcutaneously. When a dose of 100,000 units has been injected without reaction, penicillin can be given intravenously. If an immediate local or systemic reaction occurs, it can be controlled. Use of an alternative drug is then advisable. Oral desensitization regimens are also available. Neither corticosteroids nor antihistamines will prevent anaphylaxis in an individual who is highly sensitive to penicillin.

Penicillin Allergy and Hypersensitivity to Other Drugs

Patients who are allergic to one penicillin should be considered allergic to all penicillins. However, this generalization may not be true for some children in whom a rash has developed after taking ampicillin or amoxicillin, particularly in the setting of infectious mononucleosis. Patients who are allergic to penicillin have an 8% likelihood of having an allergic reaction to a cephalosporin antibiotic[26]; cephalosporins are best avoided in patients who have IgE-mediated (anaphylactic) hypersensitivity to penicillin. Although there is much less experience with carbapenems (e.g., imipenem, meropenem, and ertapenem), the same guidelines apply for the use of these drugs in patients who are allergic to other β-lactams. However, monobactams (e.g., aztreonam) have not provoked cross-reactive hypersensitivity reactions. There are no reliable skin tests for cephalosporin, carbapenem, or monobactam allergies.

AMINOGLYCOSIDES

The aminoglycosides are bactericidal drugs that inhibit protein synthesis by binding irreversibly to the 30S ribosomal subunit of susceptible bacteria.[37] They exhibit concentration-dependent activity, as well as a postantibiotic effect. Because oxygen is

required to transport aminoglycosides across the outer bacterial membrane, these agents are ineffective against anaerobes and may function poorly in the acidic, anaerobic milieu of abscesses. Although various aminoglycosides display activity against a wide range of microorganisms, they are used chiefly to treat infections caused by aerobic gram-negative bacilli, including *Pseudomonas*. Aminoglycosides are also used in combination with cell wall–active antibiotics (e.g., penicillins and vancomycin) for the synergistic treatment of deep tissue infections caused by enterococci and coagulase-negative staphylococci. In addition, streptomycin is still used to treat tuberculosis, tularemia, and plague.[38]

The aminoglycosides are not absorbed from the GI tract and must be administered intravenously or intramuscularly. The drugs penetrate pleural, ascitic, and synovial fluids in the presence of inflammation, but they diffuse poorly into other body fluids, such as the CSF, respiratory tract secretions, and the aqueous humor.[39] The aminoglycosides are excreted by glomerular filtration, and their dosages must be reduced in the presence of azotemia. Blood levels may be monitored to ensure proper dosing. Peak levels may be lower than anticipated in febrile patients, patients with an expanded extracellular volume, and patients with major burns. The major toxicities of the aminoglycosides include renal damage and ototoxicity, which may be vestibular or auditory.[40] Other adverse reactions include rash, nausea and vomiting, and neuromuscular blockade, which is rare but may occur in patients with myasthenia gravis or in those receiving succinylcholine or similar drugs. Endotoxinlike reactions to gentamicin have been reported.

Specific Aminoglycosides

Gentamicin, netilmicin, and tobramycin have similar spectra of activity, except tobramycin is more active against *P. aeruginosa*. Amikacin is used principally for gentamicin-resistant gram-negative bacilli. Streptomycin is occasionally used in the multidrug treatment of tuberculosis and in synergistic therapy of gentamicin-resistant enterococci.[38] Spectinomycin is used only in the treatment of gonorrhea in patients who are allergic to cephalosporins and fluoroquinolones.[41]

Once-Daily Aminoglycoside Dosing

In patients with normal renal function, aminoglycosides are conventionally administered in divided doses at 8- to 12-hour intervals. To decrease toxicity and cost, once-daily dosing regimens have been widely adopted.[42] In most protocols, the total doses are equivalent in the single-dose and divided-dose regimens. Meta-analyses of such trials concluded that in patients with normal renal function, once-daily dosing is as effective as divided dosing and has a lower risk of toxicity.[43] Although most trials evaluated immunocompetent adults, similar trends were noted for children and for patients with febrile neutropenia. Once-daily dosing has not been studied adequately in pregnant women or in patients with renal dysfunction, burns, ascites, or endocarditis.

The Changing Role of Aminoglycosides

The aminoglycosides are extremely active antibiotics that are clinically effective against many serious infections caused by gram-negative bacilli, and these agents are inexpensive. These assets, however, must be weighed against the potential of aminoglycosides to produce renal and otovestibular toxicity. New efforts to improve the toxic-to-therapeutic ratio of aminoglycosides include once-daily dosage schedules and reevaluations of the recommended therapeutic ranges. To determine the future role of aminoglycosides, the cost-effectiveness of these agents needs to be compared directly with that of the new β-lactams and the fluoroquinolones.

POLYMYXINS

Polymyxins are cationic polypeptides that disrupt the bacterial cell membrane through a detergentlike mechanism. With the development of less toxic agents, such as extended-spectrum penicillins and cephalosporins, parenteral polymyxin use was largely abandoned, except for the treatment of multidrug-resistant pulmonary infections in patients with cystic fibrosis. More recently, however, the emergence of multidrug-resistant gram-negative bacteria, such as *P. aeruginosa* and *Acinetobacter baumannii*, and the lack of new antimicrobial agents has led to the revived use of the polymyxins, particularly colistin (polymyxin E) and polymyxin B.[44] Colistin is bactericidal against gram-negative bacilli, including strains of *Acinetobacter*, *P. aeruginosa*, *Klebsiella*, *Enterobacter*, *E. coli*, *Citrobacter*, *Morganella*, *H. influenzae*, and some strains of *Stenotrophomonas maltophilia*. It is not active against *P. mallei*, *B. cepacia*, *Proteus*, *Providencia*, or *Serratia*. Colistin sulfate is available as an oral formulation for treatment of bowel decontamination; colistimethate sodium can be administered intravenously, intramuscularly, or by nebulization. The recommended intravenous dose is 2.5 to 5 mg/kg/day divided into two to four equal doses; however, dosing must be reduced if renal insufficiency is present. In the largest study of colistin toxicity, which examined 317 courses of therapy, rates of renal insufficiency and neurotoxicity were 20.2% and 7.3%, respectively.[45] Respiratory insufficiency and apnea were seen in 2.1% of patients.

Polymyxin B has been used extensively in topical otic and ophthalmic solutions, but it has more limited parenteral experience than colistin. Originally, colistimethate sodium was thought to be less toxic than polymyxin B; however, if the drugs are administered at comparable doses, their toxicities may be similar.

Given the concern for potential toxicity, the use of colistin and polymyxin B should be viewed as a treatment of last resort in patientswho have serious infections caused by multidrug-resistant gram-negative pathogens for which no other therapeutic options exist.

TETRACYCLINES

Tetracyclines inhibit bacterial protein synthesis by reversibly binding to the 30S ribosomal subunit. Originally, they were widely used because of their broad spectrum of activity against both gram-positive and gram-negative bacteria; they are used less extensively now because of the availability of the more effective bactericidal penicillins, cephalosporins, and fluoroquinolones. The emergence of resistance among gram-negative bacilli, group A streptococci, and pneumococci and the increased risk of superinfection caused by drug-resistant organisms have also tended to limit their use.

Tetracyclines are the drugs of choice in the treatment of rickettsial diseases, such as Rocky Mountain spotted fever, ehrlichiosis, and Q fever; they also have activity against *Francisella, Brucella,* and spirochetes.[46] Tetracyclines are useful in the treatment of pneumonia caused by *Mycoplasma pneumoniae* or *Chlamydia pneumoniae*[47]; urogenital infections caused by *C. trachomatis*; and other chlamydial diseases, such as psittacosis, trachoma, and lymphogranuloma venereum. They may also be used in the

treatment of urinary tract infections caused by susceptible organisms. In the penicillin-allergic patient, doxycycline is an alternative therapy for syphilis, leptospirosis, and cat and dog bites infected by *Pasteurella*. As antimicrobial resistance continues to emerge, a new use has been found for minocycline in the treatment of MRSA. Doxycycline is active against some strains of VRE.[48]

Doxycycline has emerged as the favored tetracycline. It is available in oral and intravenous formulations, has a long half-life that allows administration in one or two daily doses of 100 mg, and does not accumulate in the presence of renal failure. Minocycline is similarly available in oral and intravenous dosing of 100 mg twice daily, whereas the parent tetracycline requires dosing of 250 to 500 mg four times daily. Concomitant ingestion of milk and antacids impairs the absorption of all tetracyclines. The tetracyclines should not be given to children younger than 9 years or to pregnant women, because permanent discoloration of teeth may result. Other potential side effects include phototoxicity, hepatotoxicity, esophageal ulceration, and, rarely, pseudotumor cerebri. In addition, minocycline has been associated with vestibular reactions and skin and mucous membrane discoloration.

GLYCYLCYCLINES

Tigecycline received approval by the Food and Drug Administration in 2005 as the first member of the glycylcycline family. Tigecycline is a semisynthetic derivative of minocycline. Its mechanism of action is similar to that of minocycline in that tigecycline inhibits protein synthesis by binding to the 30S ribosomal subunit.[49] Modification of position 9 in the tetracycline ring, however, allows tigecycline to overcome the two major forms of tetracycline resistance—namely, ribosomal protection and drug efflux. Tigecycline has a broad spectrum of bacteriostatic activity against gram-positive, gram-negative, atypical, and anaerobic bacteria. It is active against multidrug-resistant bacteria, such as penicillin-resistant *S. pneumoniae*, MRSA, VRE, *E. coli* containing ESBL, and *K. pneumoniae*. It is also active against some strains of *Acinetobacter*; however, it is not active against *Pseudomonas* or *Proteus*. Tigecycline is currently indicated in the treatment of complicated intra-abdominal infections and complicated skin and soft tissue infections.

Tigecycline is available only as an intravenous formulation; it is administered at a loading dose of 100 mg, followed by 50 mg every 12 hours. Excretion is primarily through the biliary system. Side effects are similar to those of the tetracyclines, with GI disturbances such as nausea and emesis being the most common.

MACROLIDES

The macrolides are composed of 14, 15, or 16 carbon atoms joined together in a complex circular molecule that is linked to various side chains.[50] In the United States, erythromycin, clarithromycin, azithromycin, and dirithromycin are available. Macrolides inhibit bacterial protein synthesis by reversibly binding to the 50S ribosomal subunit of susceptible microorganisms.

Erythromycin

Erythromycin is active against such gram-positive bacteria as penicillin-susceptible *S. pneumoniae, C. diphtheriae,* and, historically, group A streptococci. In the late 1980s, macrolide-resistant group A streptococci began to emerge worldwide. Resistance in Finland reached 19% overall and 42% in one region; however, these levels decreased to 6% to 8% after a reduction in the use of macrolides in outpatient therapy.[51] Similarly, a clonal outbreak of erythromycin-resistant group A streptococci was recognized in Pittsburgh in 1998.[52] Currently, 5% of group A streptococci are macrolide resistant in the United States.[53] Most pneumococci that are not susceptible to penicillin are also resistant to erythromycin. Erythromycin remains active against *M. pneumoniae, C. trachomatis,* and some gram-negative bacilli, including *Legionella pneumophila, Campylobacter,* and *Bordetella pertussis*. *Neisseria* and *T. pallidum* are also susceptible to erythromycin.

Erythromycin is excreted to a large extent in the bile and only to a minor degree in the urine. The dosage need not be altered in the presence of renal insufficiency. Erythromycin penetrates pleural and peritoneal fluids and crosses the placenta; this allows it to be used to treat syphilis in pregnant women who are allergic to penicillin.

Because erythromycin is active against *Legionella, Mycoplasma,* and *Chlamydia* species, it is an effective treatment for patients with atypical pneumonia [*see 7:XX Pneumonia and Other Pulmonary Infections*]. Erythromycin is also effective for *Campylobacter* gastroenteritis, for the treatment of diphtheria, and for chemoprophylaxis in pertussis carriers. In patients who cannot tolerate penicillins and cephalosporins, erythromycin is an effective alternative for the treatment of streptococcal pharyngitis, though resistance is emerging, and treatment of syphilis. Other uses of erythromycin include granuloma inguinale and chancroid, prophylaxis for elective bowel surgery (for which it is administered orally with neomycin), and acne (for which it is administered topically or orally).

Erythromycin is administered orally at a dosage of 250 mg to 1 g four times daily. Therapeutic serum levels can be achieved with any of the oral erythromycin preparations. Intravenous preparations of erythromycin are available for the treatment of severe infections, but prolonged therapy is difficult because of the frequent occurrence of thrombophlebitis at infusion sites. GI intolerance is the most common side effect. Other adverse reactions shared by the macrolides include rash, stomatitis, pseudomembranous colitis, pancreatitis, and ototoxicity. Because of the occasional occurrence of cholestatic hepatitis after administration of the estolate form, erythromycin base or stearate is preferred for adults.

Clarithromycin

Clarithromycin is a semisynthetic 14-member macrolide that is available only in oral formulation. It is acid stable and is well absorbed from the GI tract irrespective of food ingestion.[50] Like erythromycin, clarithromycin achieves wide tissue penetration. Because of its longer half-life, however, clarithromycin may be administered once or twice a day, rather than four times a day, in 250 to 500 mg doses. In addition, clarithromycin produces fewer GI side effects than erythromycin; some people report a metallic taste. Clarithromycin and other macrolides can produce ventricular arrhythmias when administered with cisapride, astemizole, or disopyramide.

Clarithromycin is highly active against organisms that are sensitive to erythromycin, including streptococci, pneumococci, staphylococci, *Legionella, Campylobacter, Mycoplasma,* and *C. pneumoniae*. Clarithromycin also exhibits excellent activity against *M. catarrhalis* and *H. influenzae*; it is therefore an attractive agent for the treatment of respiratory tract infections, including sinusitis, pharyngitis, bronchitis, and pneumonia caused by susceptible organisms. Clarithromycin has been used successfully in patients who have Legionnaires disease, but azithromycin is the

most active macrolide against *Legionella,* the etiologic agent of this community-acquired pneumonia [*see* Azithromycin, *below*]. The drug has assumed an important role in the multidrug regimens used to treat disseminated *M. avium* complex infections and is active against other nontuberculous mycobacteria, such as *M. chelonae* and *M. fortuitum*. It is also used as a component of some *H. pylori* regimens.

Azithromycin

Azithromycin, a 15-member macrolide, is active against the same broad range of organisms as those that clarithromycin inhibits.[50] An intravenous preparation is available. Like the other macrolides, azithromycin is well absorbed from the GI tract. However, azithromycin differs from clarithromycin in that the bioavailability of azithromycin is decreased by food. As a result, azithromycin should be taken 1 hour before or 2 hours after meals. Azithromycin clears rapidly from serum and moves promptly into interstitial and intracellular tissue compartments. Tissue levels are extraordinarily prolonged, with an average terminal half-life of 68 hours. Therefore, tissue levels of azithromycin can be expected to remain in the therapeutic range from 4 to 7 days after a 5-day treatment course. These unique pharmacokinetics support the current program of administering azithromycin once daily for 5 days. Plasma and tissue levels are considerably lower, however, than those for clarithromycin. Like clarithromycin, azithromycin appears to be well tolerated and to have fewer GI side effects than erythromycin.

Azithromycin has been effective in the treatment of pharyngitis, sinusitis, bronchitis and pneumonia, and skin and soft tissue infections. Azithromycin has also assumed an important role in the treatment of Legionnaires disease and other atypical and community-acquired pneumonias [*see 7:XX Pneumonia and Other Pulmonary Infections*]. In addition to these indications, azithromycin has been used successfully as a single 1 g dose for urethritis and cervicitis caused by *C. trachomatis*. This provides a significant advantage because no other single-dose regimen now exists for these infections. Another important use for azithromycin is in the prophylaxis and treatment of *M. avium* complex infections in patients with AIDS.

Dirithromycin

Dirithromycin is an oral macrolide antibiotic with an antibacterial spectrum similar to that of erythromycin. It can be administered once daily but has no other advantages over erythromycin. The drugs produce similar GI side effects, and dirithromycin is substantially more expensive.

KETOLIDES

Ketolides, which are derived from erythromycin, inhibit bacterial protein synthesis by binding to the 50S ribosomal subunit, which is similar to the action of macrolides.[54] Telithromycin, the first member of the ketolide class of antimicrobial agents, received FDA approval in 2004. Telithromycin is predominantly active against respiratory pathogens, such as *S. pneumoniae* (including penicillin-resistant and macrolide-resistant strains), *H. influenzae, M. catarrhalis, M. pneumoniae, C. pneumoniae,* and *Legionella*; it also has activity against *S. pyogenes* and methicillin-susceptible *S. aureus*. However, telithromycin is inactive against Enterobacteriaceae. Telithromycin has been effective in the management of acute exacerbations of chronic bronchitis, acute bacterial sinusitis, and community-acquired pneumonia. Because telithromycin maintains activity against macrolide-resistant *S. pneumoniae*, it has potential as a fluoroquinolone-sparing agent in situations in which macrolide resistance is a concern.

Telithromycin is well absorbed from the GI tract, and its bioavailability is not affected by food ingestion. Because of its long half-life, the standard dose is 800 mg once a day for 5 to 7 days. Telithromycin is well tolerated; its most adverse side effects are diarrhea, nausea, and headache.

CLINDAMYCIN

Clindamycin inhibits bacterial protein synthesis by binding to the bacterial 50S ribosomal subunit at a site that overlaps with chloramphenicol and macrolides; the use of clindamycin with these drugs results in antagonism.[55] Clindamycin is rapidly absorbed and penetrates most tissues well, with the exception of CSF and brain. It may be administered orally (150 to 450 mg four times a day) or intravenously (600 to 900 mg every 8 hours), depending on the severity of infection. Because clindamycin is excreted primarily by the liver, no dose adjustment is necessary in patients with renal insufficiency; however, in patients with moderate to severe hepatic disease, clindamycin should be administered in half the usual dose. Clindamycin is active against most anaerobes, although resistance is seen in 10% to 20% of infections with *B. fragilis,* peptostreptococci, and clostridia other than *C. perfringens*. Clindamycin also has activity against most aerobic gram-positive bacteria, including *S. aureus, S. pneumoniae,* and group A and other streptococci except *Enterococcus*. It is also active against *B. anthracis,*[56] but clinical experience is very limited.

Clindamycin is indicated in the treatment of serious infections caused by susceptible anaerobes, particularly those originating in the GI and female genital tracts.[57] Clindamycin has also been very effective in the treatment of aspiration pneumonia and lung abscess. Because of its ability to rapidly reduce toxin production, clindamycin may be very helpful in the management of streptococcal toxic-shock syndrome and necrotizing infections caused by *C. perfringens* and *S. pyogenes,* possibly in conjunction with penicillin. Clindamycin has also been used to treat protozoan infections, such as toxoplasmosis and babesiosis.[47] Because clindamycin does not readily penetrate the CNS, even when there is marked meningeal inflammation, it should not be used to treat meningitis. For acne therapy, clindamycin should be used only topically.

Nausea, vomiting, and diarrhea are the most common side effects associated with clindamycin, but hypersensitivity reactions, neutropenia, and thrombocytopenia may also occur infrequently. Although *C. difficile* colitis has been reported with most antimicrobials, the incidence with clindamycin may approach 10% and should limit its indiscriminate use.

METRONIDAZOLE

Metronidazole was originally noted for its effectiveness against certain protozoa such as *Trichomonas, Giardia,* and *Entamoeba* and was later recognized for its anaerobic activity. Metronidazole undergoes a reductive process within the bacterial cell, a process culminating in the damage of bacterial DNA by the activated metabolites.[55]

Metronidazole is bactericidal against almost all anaerobic gram-negative bacilli and against most *Clostridium* species. Although true anaerobic streptococci are generally susceptible to metronidazole, microaerophilic streptococci and *Actinomyces* and *Propionibacterium* species are often resistant. It has supplanted penicillin as the drug of choice for *B. fragilis* and *C. tetani*. Although metronidazole and oral vancomycin are equally effective

in treating *C. difficile* colitis, metronidazole has become the preferred agent because of its lower cost and because of concern for the emergence of VRE. Metronidazole is also effective both orally and vaginally in the treatment of bacterial vaginosis.

When metronidazole is administered orally, it is well absorbed and is widely distributed in body tissues, including those of the CNS. For serious anaerobic infections, the drug is administered intravenously with a loading dose of 15 mg/kg, followed by 7.5 mg/kg every 6 hours until the patient is well enough to take an oral dosage. This is roughly equivalent to 500 mg every 6 to 8 hours. The dosage need not be reduced in azotemic patients, but it should be reduced in patients with hepatic insufficiency.

Side effects of metronidazole include dry mouth associated with a metallic taste and nausea. Concurrent use of alcohol may cause a reaction similar to that produced when alcohol is drunk after disulfiram ingestion. Neurologic symptoms, including peripheral neuropathy and encephalopathic reactions, and neutropenia are uncommon. Pancreatitis has been reported. Metronidazole is mutagenic for bacteria. Metronidazole is carcinogenic for rats and mice; carcinogenicity in humans has not been demonstrated but remains a concern.

Metronidazole is effective in the treatment of a variety of infections caused by anaerobes, including CNS infections, bone and joint infections, abdominal and pelvic sepsis, and endocarditis. Failures have been reported in the treatment of pleuropulmonary infection, which may reflect the polymicrobial nature of these infections.

CHLORAMPHENICOL

Although chloramphenicol remains a useful antibiotic, its role has decreased as less toxic alternatives have become available.[48,55] Because of the rare occurrence of aplastic anemia, clinical use of chloramphenicol should be limited to serious infections for which alternative antibiotics may be less effective. Such infections include those caused by VRE, *B. fragilis,* and *Salmonella,* as well as meningitis caused by *H. influenzae, N. meningitidis,* or *S. pneumoniae* in patients who are allergic to penicillin and cephalosporins. However, other agents are usually preferable to chloramphenicol for such infections. For example, the newer cephalosporins and fluoroquinolones should be considered for salmonellosis; second-generation and third-generation cephalosporins should be considered for ampicillin-resistant strains of *H. influenzae*; linezolid and quinupristin-dalfopristin should be considered for VRE; and combinations of a penicillin and a β-lactamase inhibitor, carbapenems, cefmetazole, cefoxitin, cefotetan, clindamycin, or metronidazole can be used to treat infections caused by *B. fragilis*.

Chloramphenicol may be administered orally or intravenously. Chloramphenicol diffuses rapidly into most tissues, CSF, ascitic fluid, and aqueous humor. The drug is lipid soluble and achieves levels in the brain up to nine times higher than in the serum. Chloramphenicol is inactivated in the liver by conjugation with glucuronic acid; blood levels may increase in patients with marked cirrhosis and jaundice.

One of the more serious and potentially fatal side effects of chloramphenicol is aplastic anemia, which can occur weeks to months after the completion of therapy. A separate dose-related suppression of the bone marrow with neutropenia, anemia, and thrombocytopenia is common during treatment and is usually reversible. Another potentially fatal adverse reaction is the gray-baby syndrome in neonates, which is caused by a diminished ability to conjugate chloramphenicol and to excrete the active form in the urine. Other adverse effects include rash, stomatitis, GI intolerance, and neurotoxic reactions.

In view of the risk of aplastic anemia and neonatal toxicity, the following guidelines should be employed in the use of chloramphenicol: (1) it should be used in the treatment of only those infections for which it is clearly indicated and for which less toxic alternatives are not feasible; (2) repeated courses should be avoided; (3) the complete blood count should be checked two or three times weekly; and (4) the patient should be observed closely for evidence of sore throat, which may be a marker of granulocytopenia.

VANCOMYCIN

Vancomycin is a glycopeptide that impairs cell wall synthesis of gram-positive bacteria. Its spectrum of activity includes staphylococci, streptococci, pneumococci, enterococci, clostridia, *Corynebacterium* species, and other gram-positive bacteria.[57] It is generally bactericidal but is bacteriostatic against some strains of enterococci, coagulase-negative staphylococci, and corynebacteria.

Vancomycin is poorly absorbed when administered orally. The oral route is employed only for the treatment of staphylococcal enterocolitis and *C. difficile* colitis at a dosage of 125 to 250 mg every 6 hours. In adults with severe staphylococcal infection or viridans streptococcal or enterococcal endocarditis who have normal renal function and who are highly allergic to penicillin, vancomycin is administered intravenously at a dosage of 500 mg every 6 hours or 1 g every 12 hours. To avoid hypotension and histaminelike reactions, the drug should always be infused slowly over a period of 1 hour. Pretreatment with antihistamines can help avert the so-called red-man syndrome; in rare cases, desensitization may be necessary. Other adverse effects may include thrombophlebitis at the infusion site, chills and fever, ototoxicity, and possibly nephrotoxicity.

Vancomycin is the drug of choice in the treatment of infections caused by MRSA. Concomitant administration of an aminoglycoside is often necessary when vancomycin is used in the treatment of enterococcal endocarditis. Vancomycin can be very useful in therapy for prosthetic valve endocarditis caused by coagulase-negative staphylococci; it is frequently administered in combination with rifampin and gentamicin in this setting.

Vancomycin does not penetrate normal meninges but does enter the CSF when the meninges are inflamed. Most of the drug is eliminated through the kidneys; in the presence of renal failure, the dose must be reduced. Although serum vancomycin levels are often monitored, measurements are not necessary in routine cases. However, determination of serum levels is an important guide to dosage when the drug must be administered in the presence of impaired renal function; in this setting, peak serum levels of 20 to 40 mg/ml and trough levels of 10 mg/ml or less are considered ideal. Particular attention to ototoxicity and nephrotoxicity is required when vancomycin is administered with an aminoglycoside.

The epidemiology of vancomycin use provides a cautionary tale for this drug and for other antimicrobials.[58] Although vancomycin is a valuable and effective drug, it is often used inappropriately. This has resulted in substantial financial costs and the emergence of vancomycin-resistant organisms, including enterococci[59] and staphylococci[60,61]; infections with these organisms are very difficult to treat [*see 7:I Infections Due to Gram-Positive Cocci*].

STREPTOGRAMINS

Quinupristin and dalfopristin are two structurally distinct

streptogramins that bind to separate sites on the bacterial 50S ribosomal subunit and act synergistically to inhibit protein synthesis.[62] The drugs are marketed together in a 30-to-70 ratio as Synercid. Although quinupristin-dalfopristin is active against a variety of gram-positive bacteria, including *S. pneumoniae* and group A streptococci, its major use is in the treatment of serious infections caused by vancomycin-resistant *E. faecium*. *E. faecalis,* however, is intrinsically resistant to quinupristin-dalfopristin as a result of drug efflux. Quinupristin-dalfopristin may also be useful in occasional vancomycin-intolerant patients with severe infections caused by MRSA or coagulase-negative staphylococci. The combination has activity against *S. aureus* that is intermediately resistant to vancomycin and against vancomycin-resistant *S. aureus*. Although quinupristin-dalfopristin was first marketed in 1999, resistance is already emerging.

The usual dose of quinupristin-dalfopristin is 7.5 mg/kg given intravenously over 1 hour every 8 hours. Because quinupristin-dalfopristin is associated with a high incidence of phlebitis, a central line should be used for intravenous delivery. Other adverse effects include arthralgias and myalgias, which may be severe, and elevated bilirubin levels. The drug is metabolized by the liver, so no dose reduction is required in azotemic patients. Careful attention to drug interactions is important because quinupristin-dalfopristin inhibits the metabolism of drugs such as cyclosporine through the hepatic cytochrome P450 CYP3A4 enzyme.

OXAZOLIDINONES

In 2000, linezolid became the first member of the oxazolidinone class to be approved for clinical use in the United States. Linezolid is a synthetic antibiotic that inhibits protein synthesis by binding to a site on the bacterial 23S ribosomal RNA of the 50S subunit, thus preventing the formation of the 70S initiation complex that is required for ribosomal function.[62,63] It has a unique mechanism of action, and no cross-resistance with other antimicrobials has been reported.

Linezolid is active against nearly all aerobic gram-positive cocci at concentrations of 4 mg/ml or less, including penicillin-resistant pneumococci, MRSA, and VRE; however, resistant strains have been isolated.[64] The drug is bacteriostatic against staphylococci and enterococci, but it is bactericidal against most streptococci. Linezolid is also active against *L. monocytogenes, M. catarrhalis, H. influenzae, N. gonorrhoeae, B. pertussis, P. multocida,* and *Nocardia* species. *C. difficile, C. perfringens,* and *Bacteroides* species are susceptible, but enteric gram-negative bacilli and *Pseudomonas* species are not.

Intravenous and oral preparations of linezolid are available; the usual dosage is 600 mg every 12 hours regardless of the choice of route. The oral form is absorbed rapidly and completely and has 100% bioavailability that is not affected by meals. Linezolid is widely distributed in well-perfused tissues. Nonrenal mechanisms account for 65% of the drug's clearance. Patients with mild to moderate renal or hepatic insufficiency do not require dose reduction. Linezolid is removed by hemodialysis.

Linezolid is fairly well tolerated. Nausea, vomiting, and headaches are the most common side effects, but reversible bone marrow suppression, including thrombocytopenia, leukopenia, and anemia, can also occur, and patients require careful monitoring during prolonged treatment courses as a precaution against these complications. Because linezolid is a reversible inhibitor of monoamine oxidase, patients taking linezolid may experience an exaggerated hypertensive response to sympathomimetic agents such as pseudoephedrine. In addition to avoiding decongestants, patients taking linezolid should avoid foods or beverages with high tyramine content; these include aged cheeses, air-dried meats, tap beer, red wine, soy sauce, and sauerkraut. Concomitant use of adrenergic or serotoninergic antidepressants should be avoided.[65]

Linezolid has been used successfully in the treatment of multidrug-resistant gram-positive bacterial infections, including VRE and MRSA,[62,66] but clinical experience with deep-seated infections such as endocarditis and osteomyelitis is limited. Although resistance is uncommon, it can develop during therapy. As a result, it may be wise to reserve this unique antibiotic for serious infections caused by MRSA, VRE, or coagulase-negative staphylococci that do not respond to vancomycin. It also has a role in decreasing hospital stay for some patients with resistant gram-positive infections,[67] although these patients require close outpatient monitoring.

DAPTOMYCIN

Daptomycin is a cyclic lipopeptide naturally produced by *Streptomyces roseosporus*. It has a unique mechanism of action: it binds to bacterial membranes and causes a rapid depolarization of membrane potential, which leads to the inhibition of protein, DNA, and RNA synthesis, resulting in bacterial death.[68] It is active against most gram-positive aerobic bacteria, including VRE, MRSA**,** glycopeptide intermediately susceptible *S. aureus,* coagulase-negative staphylococci, and penicillin-resistant *S. pneumoniae*. It also has activity against *S. pyogenes, S. agalactiae, S. dysgalactiae,* and *C. jeikeium*. It is currently approved only for complicated skin and soft tissue infections.[69] Daptomycin achieves suboptimal levels in the lungs and should not be used for pulmonary infections.

Daptomycin is bactericidal, and synergy with gentamicin has been demonstrated against staphylococci and enterococci. Although there is limited experience with this new drug, no mechanism of resistance has been identified. In addition, there appears to be no cross-resistance with other antibiotic classes.

Daptomycin is available only in intravenous formulation. The main adverse event seen in early studies of daptomycin administered every 12 hours was a reversible myopathy that affected skeletal muscles. The current once-daily intravenous dose of 4 mg/kg appears to improve the efficacy-toxicity ratio by decreasing the incidence of myopathy. Because daptomycin is primarily excreted by the kidney, the dose should be 4 mg/kg every 48 hours in patients who have a creatinine clearance rate of less than 30 ml/min, including patients on hemodialysis or continuous ambulatory peritoneal dialysis. The role of daptomycin in therapeutics is currently limited, given its narrow indications and the availability of alternative agents, including vancomycin and linezolid.

SULFONAMIDES AND TRIMETHOPRIM

Sulfonamides

Sulfonamides were the first class of antimicrobial agents to be discovered. They inhibit dihydropteroate synthetase in the bacterial folic acid pathway. Although their clinical role has diminished, they are still useful in certain situations. Because of their efficacy and low cost, sulfonamides can be useful in treating uncomplicated urinary tract infections caused by *E. coli,* although resistance is high. Sulfonamides are also useful for *Nocardia* and *Toxoplasma* infections. Topical sulfonamides are still used in a

few situations. Sulfacetamide eyedrops are sometimes employed to treat superficial ocular infections. Topical silver-sulfadiazine cream is administered to burn surfaces to suppress bacterial growth and to prevent subsequent invasive infection. Both the silver ion and the sulfadiazine components of the compound probably contribute to the antibacterial activity.

Trimethoprim

Trimethoprim inhibits dihydrofolate reductase in the bacterial folic acid pathway. Trimethoprim is well absorbed from the GI tract and is widely distributed in most tissues, including the prostate. Most of the drug is excreted unchanged in the urine. Its antibacterial spectrum encompasses many aerobic gram-negative bacilli, but it is not active against *P. aeruginosa*. Trimethoprim is generally well tolerated. Side effects include skin rash (less common than with sulfonamides) and megaloblastic marrow changes. Trimethoprim inhibits renal potassium excretion, and reversible hyperkalemia has been observed in AIDS patients receiving high-dose trimethoprim-sulfamethoxazole.

Trimethoprim is approved only for the treatment of uncomplicated urinary tract infections, for which it has been shown to be as effective as trimethoprim-sulfamethoxazole in 3- and 7-day regimens. The oral dosage is 100 mg twice daily. As a result of the widespread use of trimethoprim in many parts of the world, resistance in uropathogenic *E. coli* has emerged; this has impaired its usefulness in the treatment of urinary tract infections.

TRIMETHOPRIM-SULFAMETHOXAZOLE

Use of the trimethoprim-sulfamethoxazole combination extended the list of clinical situations in which sulfonamides appear to be of value[46]: urinary tract infections, prostatitis, acute otitis media, sinusitis or bronchitis caused by susceptible strains of *H. influenzae* and *S. pneumoniae,* and certain infections caused by MRSA. Trimethoprim-sulfamethoxazole can be used to prevent or treat traveler's diarrhea. It is the drug of choice for the prevention[70] and treatment of *P. jirovecii* pneumonia and nocardiosis. With widespread use, trimethoprim-sulfamethoxazole–resistant strains of *E. coli* are appearing. In areas where the prevalence of *E. coli* resistance is at least 20%, an alternative agent such as a fluoroquinolone should be considered for urinary tract infections.[71,72]

The trimethoprim-sulfamethoxazole synergistic combination is available in oral or intravenous preparations in a 1-to-5 ratio (80 or 160 mg of trimethoprim and 400 or 800 mg of sulfamethoxazole). Both drugs are excreted primarily by the kidneys. The oral dosage of trimethoprim-sulfamethoxazole for the treatment of most infections, including urinary tract infection, in an adult is two single-strength tablets (or one double-strength tablet) every 12 hours. For serious systemic infections, the intravenous dosage is 8 to 10 mg/kg (based on the trimethoprim component) in two to four equal doses every 6 to 12 hours. For *Pneumocystis* pneumonia, the dosage is 15 to 20 mg/kg (based on the trimethoprim component) in equally divided doses every 6 hours. Reduction in dosage is necessary if renal function is impaired.

Adverse reactions to trimethoprim-sulfamethoxazole are similar to those caused by sulfonamides alone and include hypersensitivity reactions, photosensitivity, and nausea and vomiting. Hepatitis, pancreatitis, aseptic meningitis, and megaloblastic anemia occur infrequently. Although hypersensitivity reactions are more common in HIV-infected patients, desensitization protocols have enabled many of these patients with prior reactions to tolerate this medication.[73]

FLUOROQUINOLONES

Since the introduction of ciprofloxacin more than a decade ago, this class of antimicrobial agents has continued to evolve and now includes agents that differ from each other significantly with regard to activity as well as pharmacokinetic and pharmacodynamic properties [*see Table 5*]. The first-generation fluoroquinolones (e.g., ciprofloxacin and ofloxacin) are primarily active against gram-negative and a few gram-positive organisms. The second-generation fluoroquinolone levofloxacin has improved activity against gram-positive bacteria and atypical bacteria but is less potent against certain gram-negative bacteria, such as *P. aeruginosa*. The newer fluoroquinolones (i.e., gatifloxacin, gemifloxacin, and moxifloxacin) have enhanced coverage of gram-positive, atypical, and anaerobic organisms, compared with that of the first-generation and second-generation fluoroquinolones.[74] In particular, these new agents appear to be more active against *S. pneumoniae* and may have decreased potential for the development of resistance.[75]

The fluoroquinolones are bactericidal compounds that inhibit DNA synthesis and introduce double-strand DNA breaks by targeting DNA gyrase and topoisomerase IV.[76] In gram-negative bacteria such as *E. coli, Pseudomonas,* and *N. gonorrhoeae,* the primary target is DNA gyrase and the secondary target is topoisomerase IV. In gram-positive bacteria such as *S. aureus* and *S. pneumoniae,* the primary and secondary targets are reversed. These differences in primary and secondary targets are particularly evident in the earlier fluoroquinolones (e.g., ciprofloxacin and levofloxacin). The newer fluoroquinolones (e.g., gatifloxacin, gemifloxacin, and moxifloxacin) bind equivalently to the two targets. This may imply that concurrent mutations in the two target enzymes would be required for resistance to develop, although this is controversial.

Fluoroquinolones demonstrate concentration-dependent killing. The ratios of the peak concentration to MIC and of the area

Table 5 Selected Properties of Fluoroquinolone Antibiotics

Generic Name	*Trade Name*	*Comments**
Ciprofloxacin	Cipro	Once-daily (p.o.) or twice-daily dosing (p.o., I.V.); most active fluoroquinolone against *Pseudomonas aeruginosa*
Gatifloxacin	Tequin	Once-daily dosing (p.o., I.V.); enhanced activity against *Streptococcus pneumoniae*, atypicals, anaerobes
Gemifloxacin	Factive	Once-daily dosing (p.o.); enhanced activity against *S. pneumoniae*, atypicals, anaerobes
Levofloxacin	Levaquin	Once-daily dosing (p.o., I.V.); moderately active against *S. pneumoniae* and *P. aeruginosa*
Lomefloxacin	Maxaquin	Once-daily dosing (p.o.); less active against *S. pneumoniae*, streptococci, atypicals; phototoxicity
Moxifloxacin	Avelox	Once-daily dosing (p.o., I.V.); enhanced activity against *S. pneumoniae*, atypicals, anaerobes; hepatic elimination
Ofloxacin	Floxin	Twice-daily dosing (p.o., I.V.); same spectrum as levofloxacin

*See text for details. Only properties that distinguish among the fluoroquinolones are listed here.

under the curve to MIC appear to best correlate with clinical efficacy. On the basis of these ratios, ciprofloxacin is the most active fluoroquinolone against *Pseudomonas*, and moxifloxacin demonstrates the most favorable parameters for *S. pneumoniae*. Because the newer fluoroquinolones bind equally to DNA gyrase and topoisomerase IV and because they have enhanced pharmacokinetic and pharmacodynamic parameters for *S. pneumoniae,* it has been argued that they are the preferred fluoroquinolones for community-acquired pneumonia; it is argued that because of their specific properties, the newer fluoroquinolones would prevent the emergence of resistance and maintain the antimicrobial activity of the class.[75,77] When antimicrobial resistance does develop, there tends to be cross-resistance to other fluoroquinolones. Such resistance is usually mediated chromosomally, but plasmid-mediated resistance raises the possibility of transferable resistance.

The fluoroquinolones are broad-spectrum antimicrobials.[78] Most enteric gram-negative bacilli, including *E. coli, Proteus, Klebsiella,* and *Enterobacter,* are highly susceptible. *Pseudomonas* susceptibility has decreased in recent years as fluoroquinolone use has increased. Common GI pathogens such as *Salmonella, Shigella,* and *Campylobacter* species have typically been susceptible, although *Campylobacter* resistance is increasing. Other gram-negative organisms that are killed by low concentrations of the fluoroquinolones are *H. influenzae, P. multocida, M. catarrhalis,* and *Y. enterocolitica*. *Acinetobacter* and *Serratia* are somewhat less susceptible. *B. cepacia* and *S. maltophilia* are fluoroquinolone resistant. Fluoroquinolone-resistant strains of *N. gonorrhoeae* have emerged in the Far East, Hawaii, and mainland United States.[79] Ciprofloxacin is the drug of choice for *B. anthracis,* though other fluoroquinolones are also active in vitro.[56] Among gram-positive cocci, methicillin-susceptible strains of *S. aureus* and coagulase-negative staphylococci are usually susceptible to fluoroquinolones, but resistance has developed, particularly in MRSA, when fluoroquinolones have been used as monotherapy. Activity against enterococci is marginal. Ciprofloxacin and ofloxacin are only moderately active against *S. pneumoniae*. Subsequent generations of fluoroquinolones have increased pneumococcal activity, even against non–penicillin-susceptible pneumococci. Intracellular pathogens such as *Chlamydia, Mycoplasma, Legionella,* and *M. tuberculosis* are susceptible to fluoroquinolones. Gatifloxacin, gemifloxacin, and moxifloxacin are active against anaerobes, though *C. difficile* is resistant.

The fluoroquinolones are rapidly absorbed from the GI tract and have a nearly 100% bioavailability. Penetration into body fluids and tissues is generally excellent, and therapeutic concentrations are readily achieved in bile, stool, urine, prostate, respiratory tract, bone, and muscle. The fluoroquinolones appear to penetrate the CSF in the presence of meningeal inflammation,[80] but experience in treating meningitis is scant. Although serum protein binding is modest, the fluoroquinolones have long serum half-lives, which range from 3 to 4 hours for ciprofloxacin to 12 hours for moxifloxacin. Most fluoroquinolones are eliminated by glomerular filtration and tubular secretion, and their dosages should be reduced in the presence of moderately severe renal failure. Moxifloxacin, however, is excreted chiefly by the liver and achieves low urine levels. As a result, moxifloxacin should not be used in the treatment of urinary tract infections.

The fluoroquinolones appear to be very well tolerated, with mild GI side effects (nausea, vomiting, or anorexia), CNS side effects (light-headedness, dizziness, somnolence, or insomnia), or rash occurring in fewer than 10% of treated patients.[78] Gemifloxacin use for more than 7 days may be associated with an increased risk of rash, particularly in women younger than 40 years. Sparfloxacin contains a halide at position 8 and is associated with significantly more photosensitivity reactions than the other fluoroquinolones. Sparfloxacin and grepafloxacin were withdrawn because they were shown to prolong the QTc interval, but this adverse event has also been reported with levofloxacin, gatifloxacin, and moxifloxacin. Risk factors include underlying cardiac disease, advanced age, hypokalemia, hypomagnesemia, and the concomitant use of other agents that may prolong the QTc interval, such as antiarrhythmics, macrolides, and certain antihistamines. Tendinitis and tendon rupture occur very rarely. Less common side effects include allergic interstitial nephritis, pseudomembranous colitis, and neutropenia. Hepatic toxicity, including fulminant liver failure, has led to the withdrawal of trovafloxacin. Because fluoroquinolones have caused arthropathy in young animals, these drugs should be avoided in children and in women who are pregnant or nursing.

The fluoroquinolones have been useful clinically in a variety of infections, including urinary tract, genital, prostatic, GI, respiratory tract, soft tissue, and bone infections.[78] The fluoroquinolones are effective in preventing gram-negative bacteremia when administered prophylactically to neutropenic patients, but they do not reduce gram-positive bacteremias, febrile episodes, or infection-related mortality. Oral ciprofloxacin in combination with amoxicillin-clavulanate has been used successfully in patients at low risk of complications who have fever and neutropenia after chemotherapy.[81] Use of fluoroquinolones in uncomplicated urinary tract infections has increased in response to the concern about rising resistance to trimethoprim-sulfamethoxazole among uropathogens.[71] Fluoroquinolones should be considered as first-line therapy for pyelonephritis and complicated urinary tract infections. The fluoroquinolones, administered in various regimens ranging from a single dose to 5 days of therapy, are effective in preventing and treating traveler's diarrhea and shigellosis. They have been highly effective in the treatment of typhoid fever. However, prolongation of the carrier state limits their role in nontyphoidal *Salmonella* enteritis, and the development of resistance limits their role in *Campylobacter* enteritis. The fluoroquinolones may be useful in the empirical treatment of severe community-acquired gastroenteritis, particularly if treatment is started early. The role of fluoroquinolones in the intensive care unit, particularly as monotherapy, is limited by the widespread resistance to these agents that has developed among many gram-negative bacteria responsible for nosocomial infections (e.g., *Acinetobacter* and *Pseudomonas*).

Because of their extraordinarily broad antimicrobial activity, their favorable pharmacokinetics, and their low toxicity, the fluoroquinolones are extremely valuable drugs. Like all antimicrobial agents, however, fluoroquinolones should be used judiciously, especially in view of the emerging resistance that accompanies the increased use of these drugs.[82,83]

NITROFURANTOIN

Nitrofurantoin has been commercially available for over 50 years and continues to have an important therapeutic role. It is readily absorbed from the GI tract and rapidly excreted by the kidneys. Therapeutic utility results from the high urinary concentrations achieved. Because antimicrobial levels are not attained in the blood, this drug should be employed only in the treatment of uncomplicated mild to moderate cystitis or for the prevention of cystitis.[84] The spectrum of antibacterial activity in-

cludes *E. coli*, enterococci (including VRE), and some strains of *Klebsiella* and *Enterobacter*. *Proteus* and *Pseudomonas* species are usually resistant. Nitrofurantoin is administered orally at a dosage of 50 mg four times daily or 100 mg twice daily. It should not be administered when there is significant impairment of renal function. Nausea and vomiting are common side effects, but their incidence is decreased by use of the microcrystalline formulation and by taking the drug with food. Less frequent adverse reactions include rash, hypersensitivity pneumonitis, peripheral neuropathy, hepatitis, and hemolytic anemia in association with G6PD deficiency.

FOSFOMYCIN

Fosfomycin is a broad-spectrum antibiotic that inhibits cell wall synthesis and is active against *E. coli* and many other common urinary tract pathogens.[85] However, it has poor in vitro activity against *S. saprophyticus*. Although fosfomycin has been used parenterally in Europe for many years, the drug is approved in the United States only for the single-dose oral treatment of uncomplicated urinary tract infections in women. A 3 g sachet dose is generally effective and well tolerated, with diarrhea being the most common side effect.

RIFAMYCINS

Since its development in the 1960s, rifampin has emerged as a major antituberculosis drug. Rifampin is also active against a variety of bacteria, including coagulase-negative staphylococci and *S. aureus,* but resistance develops rapidly as a result of a single point mutation in bacterial DNA–dependent RNA polymerase when used as monotherapy.[86] When rifampin is combined with a second drug, however, resistance is less likely. It may be combined with vancomycin to treat serious infections with coagulase-negative staphylococci, such as prosthetic valve endocarditis. The combination of rifampin and a fluoroquinolone has been effective in the oral treatment of *Staphylococcus*-infected orthopedic prostheses,[87] *S. aureus* right-sided endocarditis in injection drug users,[88] and osteomyelitis in the diabetic foot.[89] It is also effective in eradicating nasopharyngeal carriage of *N. meningitidis*. More study is needed to determine the optimal role for rifampin in antimicrobial therapy.

Rifaximin, a poorly absorbed (< 0.4%) rifamycin, was approved by the FDA in 2004 for the treatment of traveler's diarrhea caused by noninvasive *E. coli*[90]; in this setting, rifaximin has been demonstrated to have an efficacy similar to that of ciprofloxacin.[91] Because of its poor bioavailability, rifaximin should not be used when there is concern for invasive disease, as evidenced by fever or bloody stools. Rifaximin is well tolerated at oral dosages of either 200 mg three times daily or 400 mg twice daily for 3 days; both dosings achieve very high stool concentrations. Rifaximin is currently being evaluated for its efficacy in the management of small bowel overgrowth syndromes and hepatic encephalopathy.

TOPICAL ANTIMICROBIAL AGENTS

Topical antimicrobial agents have been used for prophylaxis against cutaneous infections, for the treatment of minor wounds and infections, and for the eradication of *S. aureus* nasal carriage. These agents provide high drug concentrations to the desired site with minimal toxicity.[92]

Mupirocin inhibits protein synthesis by preventing the incorporation of isoleucine into the growing protein by binding to isoleucyl transfer-RNA synthetase. It is predominantly active against aerobic gram-positive cocci, including *S. aureus, S. epidermidis,* and β-hemolytic streptococci. Intranasal mupirocin is highly effective for the short-term elimination of *S. aureus*[93]; it has been shown to decrease *S. aureus* infections in dialysis patients,[94] as well as *S. aureus* infections in postsurgery wounds in colonized patients.[95] Although targeted populations may benefit from mupirocin, frequent recolonization and the development of resistance warrant caution against widespread use.

Bacitracin is active against a variety of gram-positive and gram-negative organisms; neomycin and polymyxin target gram-negative organisms, and polymyxin is bactericidal for *P. aeruginosa*.[92] Silver sulfadiazine is most commonly used for the prevention of wound infections in burn patients because of its broad gram-positive and gram-negative activity, including activity against *S. aureus* and *P. aeruginosa.*

Antimicrobial Chemoprophylaxis

The term antimicrobial chemoprophylaxis refers to the use of antimicrobial agents in the prevention of infection either before or very shortly after the introduction of pathogenic organisms—for example, after the occurrence of a compound fracture but before the appearance of clinical infection. Prophylaxis is most effective when a specific drug is selected for its activity against a particular organism, such as postexposure anthrax prophylaxis with ciprofloxacin, doxycycline, or amoxicillin.[56] When prophylaxis is aimed at preventing infection by all possible organisms through the use of broad-spectrum antimicrobials, it merely increases the selection pressure for the emergence of resistant organisms in any infection that may follow. Thus, prophylaxis is ineffective in the prevention of complicating bacterial or mycotic infections in patients with viral respiratory tract infections. Most uses of prophylaxis fall into three general categories: prevention of infection after exposure to a specific pathogen, prevention of specific types of infection in highly susceptible individuals, and prevention of postoperative infectious complications. In many instances, prophylactic use of antimicrobial agents is widely practiced, but convincing data validating the efficacy of this approach are not available.[96]

ANTIMICROBIAL PROPHYLAXIS FOR SURGICAL PROCEDURES

The use of antimicrobial prophylaxis in surgical patients involves a risk-to-benefit appraisal that varies depending on the nature of the operative procedure. To help prevent wound infections in patients undergoing elective surgery, antibiotics should be administered within 2 hours before the incision is made.[97] If prophylactic antibiotics are to be effective, administration should be timed so that therapeutic levels are attained at surgery; in addition, a low-spectrum antimicrobial should be used to inhibit the emergence of resistant organisms. Antibiotics should usually be stopped 24 hours after the procedure. The indications for prophylaxis with different operations have been reviewed, but in many instances, the available data are insufficient to make recommendations.[98] For clean elective surgical procedures such as mastectomy and thyroidectomy in which no tissue (other than the skin) carrying indigenous flora is penetrated, the risks of routine antibiotic prophylaxis outweigh the possible benefits.

When cardiovascular prostheses are employed, vancomycin or a first-generation cephalosporin (e.g., cefazolin) should be administered within 2 hours before surgery and continued for 3 to 5 days. Similar regimens may be employed for major vascular surgery. Because of the grave consequences of infection in a

prosthetic joint, vancomycin or a first-generation cephalosporin is used prophylactically when a total hip replacement is performed. In the repair of open fractures, which are commonly contaminated, prophylactic treatment with a cephalosporin for 5 to 7 days is warranted. Controlled clinical studies have indicated that the administration of oral antibiotics (e.g., enteric-coated erythromycin plus neomycin, or tetracycline plus neomycin) just before colon surgery significantly reduces the incidence of infectious complications. The erythromycin-neomycin combination, 1 g of each administered orally at 1:00 P.M., 2:00 P.M., and 11:00 P.M. on the day before surgery, has been found to reduce the number of aerobic and anaerobic organisms remaining in the colon at the time of surgery. Although the additive benefit of parenteral antimicrobial drugs in colorectal surgery has not been clearly established, a survey showed that most surgeons do combine parenteral cefoxitin or cefotetan with the previously established oral regimens.[99]

Timothy H. Dellit, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Harvey B. Simon, M.D., F.A.C.P, has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Thomas M. Hooton, M.D., F.A.C.P., has received grants for clinical research from MedImmune, Inc., and has served as an advisor for Procter & Gamble Co. and Bayer.

References

1. Kaye KS, Cosgrove S, Harris A, et al: Risk factors for emergence of resistance to broad-spectrum cephalosporins among Enterobacter spp. Antimicrob Agents Chemother 45:2628, 2001

2. Patterson DL, Ko WC, von Gottberg A, et al: Outcome of cephalosporin treatment for serious infections due to apparently susceptible organisms producing extended-spectrum beta-lactamases: implications for the clinical microbiology laboratory. J Clin Microbiol 39:2206, 2001

3. Cockerill FR III: Conventional and genetic laboratory tests used to guide antimicrobial therapy. Mayo Clin Proc 73:1007, 1998

4. Lutsar I, McCraken GH Jr, Friedland IR: Antibiotic pharmacodynamics in cerebrospinal fluid. Clin Infect Dis 27:1117, 1998

5. Sensakovic JW, Smith LG: Oral antibiotic treatment of infectious diseases. Med Clin North Am 85:115, 2001

6. Gilbert DN, Dworkin RJ, Raber SR, et al: Outpatient parenteral antimicrobial-drug therapy. N Engl J Med 337:829, 1997

7. Ito S: Drug therapy for breast-feeding women. N Engl J Med 343:118, 2000

8. Wong CS, Jelacic S, Habeeb RL, et al: The risk of hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. N Engl J Med 342:1930, 2000

9. Lee CE, Zembower TR, Fotis MA, et al: The incidence of antimicrobial allergies in hospitalized patients: implications regarding prescribing patterns and emerging bacterial resistance. Arch Intern Med 160:2819, 2000

10. Cunha BA: Antibiotic side effects. Med Clin North Am 85:149, 2001

11. Hogenauer C, Hammer HF, Krejs GJ, et al: Mechanisms and management of antibiotic-associated diarrhea. Clin Infect Dis 27:702, 1998

12. Bertino JS Jr, Owens RC Jr, Carnes TD, et al: Gatifloxacin-associated corrected QT interval prolongation, torsades de pointes, and ventricular fibrillation in patients with known risk factors. Clin Infect Dis 34:861, 2002

13. Vassileva SG, Mateev G, Parish LC: Antimicrobial photosensitive reactions. Arch Intern Med 158:1993, 1998

14. Kollef MH, Fraser VJ: Antibiotic resistance in the intensive care unit. Ann Intern Med 134:298, 2001

15. Virk A, Steckelberg JM: Clinical aspects of antimicrobial resistance. Mayo Clin Proc 75:200, 2000

16. Moellering RC Jr: Antibiotic resistance: lessons for the future. Clin Infect Dis 27(suppl 1):S135, 1998

17. Tenover FC: Development and spread of bacterial resistance to antimicrobial agents: an overview. Clin Infect Dis 33(suppl 3):S108, 2001

18. Khachatourians GG: Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria. CMAJ 159:1129, 1998

19. Williams RJ: Globalization of antimicrobial resistance: epidemiological challenges. Clin Infect Dis 33(suppl 3):S116, 2001

20. The choice of antimicrobial drugs. Med Lett Drugs Ther 43:69, 2001

21. Henry NK, Hoecker JL, Rhodes KH: Antimicrobial therapy for infants and children: guidelines for the inpatient and outpatient practice of pediatric infectious diseases. Mayo Clin Proc 75:86, 2000

22. Wright AJ: The penicillins. Mayo Clin Proc 74:290, 1999

23. Giamarellou H, Antoniadou A: Antipseudomonal antibiotics. Med Clin North Am 85:19, 2001

24. Corbella X, Ariza J, Ardanuy C, et al: Efficacy of sulbactam alone and in combination with ampicillin in nosocomial infections caused by multiresistant *Acinetobacter baumannii*. J Antimicrob Chemother 42:793, 1998

25. Martinez-Rodríguez JE, Barriga FJ, Santamaria J, et al: Nonconvulsive status epilepticus associated with cephalosporins in patients with renal failure. Am J Med 111:115, 2001

26. Kelkar PS, Li JT: Cephalosporin allergy. N Engl J Med 345:804, 2001

27. Marshall WF, Blair JE: The cephalosporins. Mayo Clin Proc 74:187, 1999

28. Guglielmo BJ, Luber AD, Paletta D Jr, et al: Ceftriaxone therapy for staphylococcal osteomyelitis: a review. Clin Infect Dis 30:205, 2000

29. Dunne EF, Fey PD, Kludt P, et al: Emergence of domestically acquired ceftriaxone-resistant *Salmonella* infections associated with AmpC beta-lactamase. JAMA 284:3151, 2000

30. Sanders CC: Cefepime: the next generation? Clin Infect Dis 17:369, 1993

31. Hellinger WC, Brewer NS: Carbapenems and monobactams: imipenem, meropenem, and aztreonam. Mayo Clin Proc 74:420, 1999

32. Livermore DM, Sefton AM, Scott GM: Properties and potential of ertapenem. J Antimicrob Chemother 52:331, 2003

33. McConnell SA, Penzak SR, Warmack TS, et al: Incidence of imipenem hypersensitivity reactions in febrile neutropenic bone marrow transplant patients with a history of penicillin allergy. Clin Infect Dis 31:1512, 2000

34. Salkind AR, Cuddy PG, Foxworth JW: Is this patient allergic to penicillin? An evidence-based analysis of the likelihood of penicillin allergy. JAMA 285:2498, 2001

35. Harris AD, Sauberman L, Kabbash L, et al: Penicillin skin testing: a way to optimize antibiotic utilization. Am J Med 107:166, 1999

36. Dodek P, Phillips P: Questionable history of immediate-type hypersensitivity to penicillin in staphylococcal endocarditis: treatment based on skin-test results versus empirical alternative treatment—a decision analysis. Clin Infect Dis 29:1251, 1999

37. Edson RS, Terrell CL: The aminoglycosides. Mayo Clin Proc 74:519, 1999

38. Maurin M, Raoult D: Use of aminoglycosides in treatment of infections due to intracellular bacteria. Antimicrob Agents Chemother 45:2977, 2001

39. Lacy MK, Nicolau DP, Nightingale CH, et al: The pharmacodynamics of aminoglycosides. Clin Infect Dis 27:23, 1998

40. Forge A, Schacht J: Aminoglycoside antibiotics. Audiol Neurootol 5:3, 2000

41. Lind I: Antimicrobial resistance in *Neisseria gonorrhoeae*. Clin Infect Dis 24(suppl 1):S93, 1997

42. Chuck SK, Raber SR, Rodvoid KA, et al: National survey of extended-interval aminoglycoside dosing. Clin Infect Dis 30:433, 2000

43. Bailey TC, Little JR, Littenberg B, et al: A meta-analysis of extended-interval dosing versus multiple daily dosing of aminoglycosides. Clin Infect Dis 24:786, 1997

44. Falagas ME, Kasiakou SK: Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. Clin Infect Dis 40:1333, 2005

45. Koch-Weser J, Sidel VM, Federman EB, et al: Adverse effects of sodium colistimethate: manifestations and specific reaction rates during 317 courses of therapy. Ann Intern Med 72:857, 1970

46. Smilack JD: The tetracyclines. Mayo Clin Proc 74:727, 1999

47. Ailani RK, Agastya G, Ailani RK, et al: Doxycycline is a cost-effective therapy for hospitalized patients with community-acquired pneumonia. Arch Intern Med 159:266, 1999

48. Klein NC, Cunha BA: New uses of older antibiotics. Med Clin North Am 85:125, 2001

49. Rubinstein E, Vaughan D: Tigecycline: a novel glycylcycline. Drugs 65:1317, 2005

50. Alvarez-Elcoro S, Enzler MJ: The macrolides: erythromycin, clarithromycin, and azithromycin. Mayo Clin Proc 74:613, 1999

51. Seppala H, Klaukka T, Vuopio-Varkila J, et al: The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. N Engl J Med 337:441, 1997

52. Martin JM, Green M, Barbadora KA, et al: Erythromycin-resistant group A streptococci in schoolchildren in Pittsburgh. N Engl J Med 346:1200, 2002

53. Waites K, Brown S: Antimicrobial resistance among isolates of respiratory tract infection pathogens from the southern United States: data from the PROTEKT US surveillance program 2000/2001. South Med J 96:974, 2003

54. Lonks JR, Goldmann DA. Telithromycin: a ketolide antibiotic for treatment of respiratory tract infections. Clin Infect Dis 40:1657, 2005

55. Kasten MJ: Clindamycin, metronidazole, and chloramphenicol. Mayo Clin Proc 74:825, 1999

56. Swartz MN: Recognition and management of anthrax: an update. N Engl J Med 345:1621, 2001

57. Wilhelm MP, Estes L: Vancomycin. Mayo Clin Proc 74:928, 1999

58. Jarvis WR: Epidemiology, appropriateness, and cost of vancomycin use. Clin Infect Dis 26:1200, 1998

59. Murray BE: Vancomycin-resistant enterococcal infections. N Engl J Med 342:710, 2000

60. Smith TL, Pearson ML, Wilcox KR, et al: Emergence of vancomycin resistance in *Staphylococcus aureus*. N Engl J Med 340:493, 1999

61. Chang S, Siever DM, Hageman JC, et al: Infection with vancomycin-resistant *Staphylococcus aureus* containing the vanA resistance gene. N Engl J Med 348:1342, 2003

62. Eliopoulos GM: Quinupristin-dalfopristin and linezolid: evidence and opinion. Clin Infect Dis 36:473, 2003

63. Diekema DJ, Jones RN: Oxazolidinone antibiotics. Lancet 358:77, 2001

64. Tsiodras S, Gold HS, Sakoulas G, et al: Linezolid resistance in a clinical isolate of *Staphylococcus aureus*. Lancet 358:207, 2001

65. Bernard L, Stern R, Lew D, et al: Serotonin syndrome after concomitant treatment with linezolid and citalopram. Clin Infect Dis 36:1197, 2003

66. Stevens DL, Herr D, Lampiris H, et al: Linezolid versus vancomycin for the treatment of methicillin-resistant *Staphylococcus aureus* infections. Clin Infect Dis 34:1481, 2002

67. Li JZ, Willke RJ, Rittenhouse BE, et al: Effect of linezolid versus vancomycin on length of hospital stay in patients with complicated skin and soft tissue infections caused by known or suspected methicillin-resistant staphylococci: results from a randomized clinical trial. Surg Infect 4:57, 2003

68. Tally FP, DeBruin MF: Development of daptomycin for gram-positive infections. J Antimicrob Chemother 46:523, 2000

69. Carpenter CF, Chambers HF: Daptomycin: another novel agent for treating infections due to drug-resistant gram-positive pathogens. Clin Infect Dis 38:994, 2004

70. El-Sadr WM, Luskin-Hawk R, Yurik TM, et al: A randomized trial of daily and thrice-weekly trimethoprim-sulfamethoxazole for the prevention of *Pneumocystis carinii* pneumonia in human immunodeficiency virus–infected patients. Clin Infect Dis 29:775, 1999

71. Warren JW, Abrutyn E, Hebel JR, et al: Guidelines for antimicrobial treatment of uncomplicated acute bacterial cystitis and acute pyelonephritis in women. Clin Infect Dis 29:745, 1999

72. Gupta K, Sahm DF, Mayfield D, et al: Antimicrobial resistance among uropathogens that cause community-acquired urinary tract infections in women: a nationwide analysis. Clin Infect Dis 33:89, 2001

73. Demoly P, Messaad D, Sahla H, et al: A new protocol for TMP-SMX desensitization in HIV-infected patients is safe and effective. J Allergy Clin Immunol 102:1033, 1998

74. Saravolatz LD, Leggett J: Gatifloxacin, gemifloxacin, and moxifloxacin: the role of three newer fluoroquinolones. Clin Infect Dis 37:1210, 2003

75. Scheld WM: Maintaining fluoroquinolone class efficacy: review of influencing factors. Emerg Infect Dis 9:1, 2003

76. Hooper DC: Mechanisms of action of antimicrobials: focus on fluoroquinolones. Clin Infect Dis 32(suppl 1):S9, 2001

77. Talan DA: Clinical perspectives on new antimicrobials: focus on fluoroquinolones. Clin Infect Dis 32(suppl 1):S64, 2001

78. Hooper DC: New uses for new and old quinolones and the challenge of resistance. Clin Infect Dis 30:243, 2000

79. Ohye R, Lee V, Whiticar P, et al: Fluoroquinolone resistance in *Neisseria gonorrhoeae*, Hawaii, 1999, and decreased susceptibility to azithromycin in *N. gonorrhoeae*, Missouri, 1999. JAMA 284:1917, 2000

80. Lipman J, Allworth A, Wallis SC: Cerebrospinal fluid penetration of high doses of intravenous ciprofloxacin in meningitis. Clin Infect Dis 31:1131, 2000

81. Freifeld A, Marchigiani D, Walsh T, et al: A double-blind comparison of empirical oral and intravenous antibiotic therapy for low-risk febrile patients with neutropenia during cancer chemotherapy. N Engl J Med 341:305, 1999

82. Chen DK, McGeer A, de Azavedo JC, et al: Decreased susceptibility of *Streptococcus pneumoniae* to fluoroquinolones in Canada. N Engl J Med 341:233, 1999

83. Ho P-L, Que T-L, Tsang D N-C, et al: Emergence of fluoroquinolone resistance among multiply resistant strains of *Streptococcus pneumoniae* in Hong Kong. Antimicrob Agents Chemother 43:1310, 1999

84. Brumfitt W, Hamilton-Miller JMT: Efficacy and safety profile of long-term nitrofurantoin in urinary infections: 18 years' experience. J Antimicrob Chemother 42:363, 1998

85. Nicolle LE: Urinary tract infection: traditional pharmacologic therapies. Am J Med 113(1A):35S, 2002

86. Zavasky DM, Sande MA: Reconsideration of rifampin: a unique drug for unique infection. JAMA 279:157, 1998

87. Zimmerli W, Widmer AF, Blatter M, et al: Role of rifampin for treatment of orthopedic implant-related staphylococcal infection: a randomized controlled trial. JAMA 279:1537, 1998

88. Dworkin RJ, Lee BL, Sande MA, et al: Treatment of right-sided *Staphylococcus aureus* endocarditis in intravenous drug users with ciprofloxacin and rifampicin. Lancet 2:1071, 1989

89. Senneville E, Yazdanpanah Y, Cazaubiel M, et al: Rifampicin-ofloxacin oral regimen for the treatment of mild to moderate diabetic foot osteomyelitis. J Antimicrob Chemother 48:927, 2001

90. DuPont HL, Jiang ZD, Ericsson CD, et al: Rifaximin versus ciprofloxacin for the treatment of traveler's diarrhea: a randomized, double-blind clinical trial. Clin Infect Dis 33:1807, 2001

91. Huang DB, DuPont HL: Rifaximin: a novel antimicrobial for enteric infections. J Infect 50:97, 2005

92. Spann CT, Tutrone WD, Weinberg JM, et al: Topical antibacterial agents for wound care: a primer. Dermatol Surg 29:620, 2003

93. Laupland KB, Conly JM: Treatment of *Staphylococcus aureus* colonization and prophylaxis for infection with topical intranasal mupirocin: an evidence-based review. Clin Infect Dis 37:933, 2003

94. Tacconelli E, Carmeli Y, Aizer A, et al: Mupirocin prophylaxis to prevent *Staphylococcus aureus* infection in patients undergoing dialysis: a meta-analysis. Clin Infect Dis 37:1629, 2003

95. Perl TM, Cullen JJ, Wenzel RP, et al: Intranasal mupirocin to prevent postoperative *Staphylococcus aureus* infections. N Engl J Med 346:1871, 2002

96. Osmon DR: Antimicrobial prophylaxis in adults. Mayo Clin Proc 75:98, 2000

97. Classen DC, Evans RS, Pestotnik SL, et al: The timing of prophylactic administration of antibiotics and the risk of surgical-wound infection. N Engl J Med 326:281, 1992

98. Antimicrobial prophylaxis in surgery. Med Lett Drugs Ther 41:75, 1999

99. Nichols, RL, Smith JW, Garbiea RY, et al: Current practices of preoperative bowel preparation among North American colorectal surgeons. Clin Infect Dis 24:609, 1997

XV SEPTIC ARTHRITIS

Brian F. Mandell, m.d., ph.d

Bacterial infections account for less than 20% of all cases of acute monoarticular and oligoarticular arthritis. Crystal-induced arthritis is approximately four times more common. Because septic arthritis represents a potential threat to life and limb, the possibility of infection dictates the sequence and pace of the diagnostic evaluation. This chapter reviews the clinical presentation, diagnosis, microbiology, and treatment of joint infections. Lyme disease, which can manifest as monoarticular or oligoarticular arthritis of the large joints, is discussed elsewhere [*see 7:VII Leptospirosis, Relapsing Fever, Rat-Bite Fever, and Lyme Disease*].

Epidemiology

Case series of patients with septic arthritis include increasing numbers of immunosuppressed and elderly patients with significant comorbidity [*see* Clinical Subgroups at Risk for Septic Arthritis, *below*]. Large retrospective studies from Europe indicate a consistent male-to-female ratio of approximately 1:1, a mean age of approximately 55 years, and polyarticular involvement in up to 15% of cases.[1-3] These data are comparable to those reported from the United States,[4,5] although some clinicians' experiences may include younger patients with disseminated gonococcal infection (DGI) or HIV infection or older patients with prosthetic joints.[6] Patients with damaged joints are at increased risk. Infection complicates 0.05% to 0.48% of arthroscopies[7]; a higher risk is associated with reconstructive or repeat procedures. In rare cases, infection complicates arthrocentesis, intra-articular injection therapy, or direct traumatic penetration of the joint.

Pathogenesis

Presumably, most bacteria reach the joint via the bloodstream and the vascular synovial membrane (and, in children, the epiphyseal plate). Joint invasion may be facilitated by the absence of a developed basement membrane in the normal synovium. Previously damaged joints seem to be at increased risk for infection,[1,6] which may be attributable to unique properties of *Staphylococcus,* the neovascularization of inflamed synovial membranes, or the increased expression of adhesion molecules on activated synovial endothelial cells. Some specific *Staphylococcus, Streptococcus,* and *Neisseriae* strains exhibit tropism to joints.[8] In rare instances, joints are directly infected by traumatic inoculation; this circumstance may be more common in infections involving small finger joints.

Synovial reaction and biochemical changes to cartilage, attributable to mediators released from inflammatory cells and bacteria, occur within hours after experimental joint infection.[9] Experimental models of bacterial arthritis have shown that even if antibiotic therapy is begun within 24 hours of infection, proteoglycans are already at least transiently depleted from cartilage. Loss of cartilage and erosion of subchondral bone may begin within days of infection and evolve into destruction of the joint. Polymorphonuclear neutrophils (PMNs), reacting to chemotactic factors, adhere to endothelial adhesion molecules and infiltrate the synovium, where they participate in the subsequent recruitment of mononuclear cells.

Some studies suggest that the degree of damage is determined by several factors, including the type of infecting organism and the accumulation of PMNs and their release of proteolytic enzymes and perhaps oxygen by-products into the joint space. *N. gonorrhoeae* induces only limited PMN activation in vitro, which may partially explain why minimal joint destruction is observed with this infection, even in studies of rabbits in which the bacteria is directly inoculated into knee joints. Preliminary animal studies suggest that potent anti-inflammatory therapy with corticosteroids, in conjunction with appropriate antimicrobial therapy, may reduce this cartilage destruction.[10,11] These observations have not yet been translated into clinical practice. Consistent with animal data, several studies reported that a delay in the initiation of treatment was a major determinant of poorer functional outcome.[1]

Diagnosis of Septic Arthritis

HISTORY

A thorough history remains a key element in the diagnosis of septic arthritis. Pertinent features include acute onset of joint pain or a significant change in the pattern of chronic joint pain; a history of penetrating joint trauma; a history of prodromal extra-articular symptoms suggestive of bacteremia; any comorbid immunosuppression, including diabetes mellitus, intravenous drug use, or prior intravenous catheterization; the presence of sexually transmitted diseases (STDs); and geographic location (e.g., in the case of Lyme disease). A complaint of sudden onset of articular or periarticular pain should never be ignored, especially if the pain is present both when the joint is at rest and when it is in motion. Because joint infections occur infrequently, patients and physicians may easily but incorrectly attribute the pain to trauma or overuse syndromes, particularly in the case of shoulder pain, or to crystal-induced arthritis if lower extremity or wrist synovitis is present.

Surprisingly, fever is not uniformly present in adults[1-4,12] or children[13] with septic arthritis. Fever may be present in fewer than 60% of patients with nongonococcal septic arthritis. Rigors may be present in fewer than 10% of patients.[12] High spiking fevers (> 39° C [102.2° F]) and rigors can also occur in patients with crystal-induced arthritis.[14] Thus, systemic features are neither sensitive enough nor specific enough to warrant making or excluding the diagnosis of septic arthritis without examination of the synovial fluid.

PHYSICAL EXAMINATION

Detailed physical examination is warranted in all cases of possible septic arthritis, as in any patient with a possible systemic infection or inflammatory disorder. The pattern of joint involvement and the presence of tendinitis or enthesitis should be noted, as should any inflammation of the eyes, skin, or mucosae. The latter finding suggests the possibility of a reactive arthritis. Pain with passive motion of a joint or palpation of the

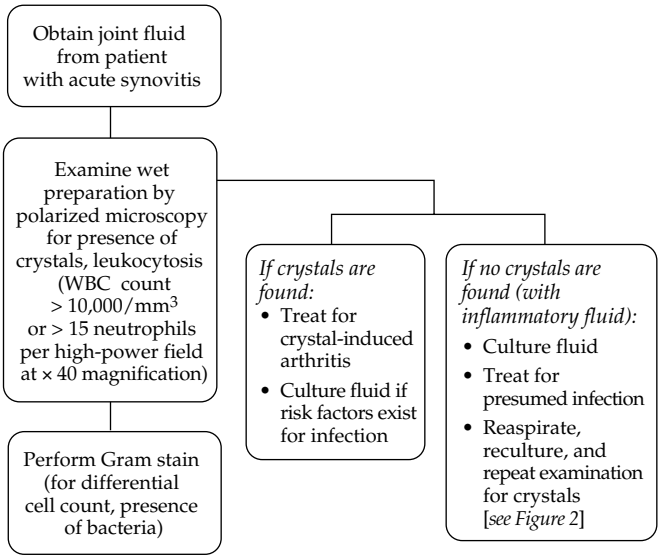


Figure 1 **Algorithm for rapid evaluation of patients with acute monoarticular arthritis on the basis of synovial fluid analysis.**

joint capsule, in the absence of trauma, suggests synovitis. Mild to moderate swelling of an infected hip or shoulder, however, is usually not detectable by examination. Septic bursitis, most commonly affecting the prepatella or olecranon bursae,[15] must be distinguished from arthritis by clinical examination. Unnecessary arthrocentesis can be avoided by appropriate recognition of a periarticular bursitis that needs to be drained. Septic olecranon and prepatellar bursitis, which are the most common sites for septic bursitis, are often associated with a peculiar periarticular, often pitting, edema. A complete joint examination should be undertaken, with specific attention paid to the sternoclavicular, sacroiliac, and midfoot joints. In the absence of risk factors for osteomyelitis or stress fractures, exquisite tenderness of the midfoot should raise particular suspicion for gouty arthritis. Infection of fibrocartilagenous joints seems to be more common in intravenous drug users and in patients with indwelling lines. Psoriatic arthritis may produce dactylitis or mimic a septic joint; thus, even trivial psoriatic skin involvement should be noted. Potential portals of infection (e.g., prior I.V. access sites) should be sought out meticulously.

LABORATORY TESTS

Synovial fluid leukocyte count, polarized microscopy, Gram stain, and culture are the most important initial laboratory investigations in the evaluation of suspected septic arthritis. Acutely or painfully swollen joints should be aspirated [*see Figures 1 and 2*] and the synovial fluid analyzed. Synovial fluid analysis is the diagnostic test of choice to distinguish between crystal-induced arthritis and infection-induced arthritis [*see* Differential Diagnosis of Acute Arthritis, *below*]. Testing for peripheral blood leukocytosis[12,14] and neutrophilia, synovial fluid glucose measurement, mucin clot testing, and assessment for elevations in acute-phase reactants are of insufficient sensitivity and specificity to confirm or exclude the diagnosis of septic arthritis. Synovial fluid leukocytosis and neutrophilia are invariably present in septic arthritis, but their presence does not allow one to distinguish between infection-induced and crystal-induced arthritis [*see 15:IX Crystal-Induced Joint Disease*].[16] Most infected joint fluids have a striking degree of leukocytosis. Although in approximately 25% of cases fluid from infected joints may have a white blood cell (WBC) count of less than 20,000/mm^3, on differential counts, more than 85% of cells are almost always neutrophils.[3,16] Infected bursal fluid may have only a few thousand WBCs. A rapid approximation of the WBC count can be obtained by microscopic examination of a drop of

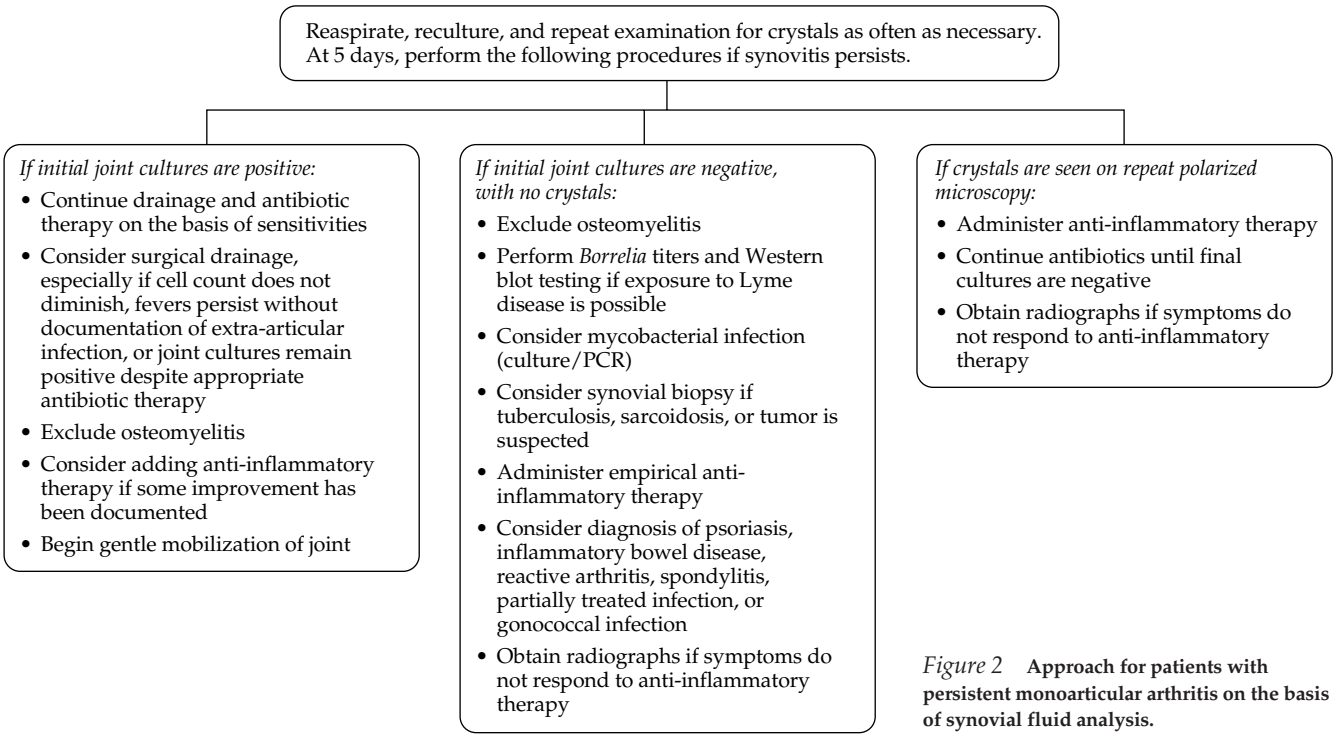


Figure 2 **Approach for patients with persistent monoarticular arthritis on the basis of synovial fluid analysis.**

Table 1 Likely Causes of Septic Arthritis on the Basis of Gram Stain and Patient Demographics

Findings on Gram Stain	*Patient Characteristics*	*Organism(s) of Concern*
No bacteria	Healthy and young	Gonococci, staphylococci
No bacteria	Presence of rheumatoid arthritis	Staphylococci
No bacteria	Significant immunosuppression, intravenous drug use, or recent hospitalization with possible gram-negative infection	Staphylococci, streptococci, *Pseudomonas* (fungal infection much less likely)
No bacteria	Intermittent synovitis of large joint, in an area endemic for Lyme disease	*Borrelia*
No bacteria or gram-negative bacilli	Recent cat or dog bite	*Pasteurella multocida*
Gram-positive bacilli	No specific features	Staphylococci, streptococci
Gram-negative diplococci	No specific features	Gonococci; consider meningococcemia
Gram-negative bacilli	No specific features	Treat for possible *Pseudomonas* infection
Gram-negative bacilli	Presence of systemic lupus erythematosus or sickle cell anemia	Include coverage for *Salmonella*, as well as for *Pseudomonas*
No bacteria	Patient has a prosthetic joint	*Staphylococcus epidermidis, S. aureas* (possibly methicillin-resistant strains)
No bacteria	Exposure to freshwater or saltwater, especially with any injury; chronic joint swelling	*Mycobacterium marinum*

synovial fluid (magnification: × 40), with each WBC estimated to represent 500 WBCs/mm^3.

Gram stain and culture of synovial fluid should be performed in all cases of suspected septic arthritis [*see Table 1*]. A Gram stain will permit evaluation of the percentage of PMNs, but it is not sensitive (< 60%) for the presence of bacteria.[2,3,16] Gonococci are observed even less frequently. Antibiotic-containing culture plates should not be utilized, and samples must be delivered promptly to the laboratory. In the absence of prior antibiotic therapy, synovial fluid culture remains the gold standard for the diagnosis of nongonococcal joint infection. Synovial biopsy may have a slightly greater yield of positive cultures, but it is infrequently performed. There is insufficient evidence to recommend routine biopsy.

Blood cultures should be obtained, as should cultures of any other potential extra-articular source of infection. If gonococcal infection is suspected, workup should include rectal, cervical, urethral, and pharyngeal cultures—any of which may be positive, even in the absence of local symptoms.[17] Cultures and Gram stain of any suspicious skin lesions should be obtained, although these are usually sterile in cases of DGI.

A wet preparation of synovial fluid should be examined by polarized microscopy for the presence of crystals. In rare cases, crystal arthropathy and septic arthritis may coexist; thus, the presence of crystals does not rule out septic arthritis.[5,18] Crystals can be found in synovial fluids from asymptomatic patients with a history of gout. If crystals are observed, patients are generally treated for crystalline arthritis alone [*see 15:IX Crystal-Induced Joint Disease*], unless there is real suspicion of coexistent infection. Hence, synovial fluid cultures should be obtained even when crystals are observed, especially if there is any suspicion of current or remote infection or if the use of therapeutic intra-articular corticosteroid injections is anticipated. Some clinicians order a bacterial culture for all inflammatory synovial fluids, even if crystals are observed. Local laboratory policy may dictate synovial fluid culture technique, but several studies suggest that the diagnostic yield may be improved with direct inoculation of fluid into blood culture vials or isolator tubes.

A potential advance in diagnostic testing has been the development of polymerase chain reaction (PCR) techniques for detection of bacterial DNA within synovial fluid and tissue.[19] PCR has been used to diagnose the presence of *Chlamydia* and *Yersinia* organisms and the presence of *Borrelia burgdorferi, N. gonorrhoeae,* and *Ureaplasma urealyticum* DNA in synovial fluids. Caution must be used, however, when interpreting the results of PCR analyses. Contamination with even minute amounts of DNA can cause false positive results. The test also does not distinguish between the presence of live and dead organisms. Once PCR becomes better standardized and clinical experience is accumulated, PCR may be useful in diagnosing gonococcal, mycobacterial, and partially treated bacterial infections, which are not rapidly diagnosed by routine culture methods. The serum uric acid level should absolutely not be used to diagnose or exclude gouty arthritis; the level may be low, normal, or high at the time of a gout attack. Autoimmune serologies (e.g., antinuclear antibody and rheumatoid factor) are of no value in the initial management of patients with acute monoarticular arthritis. The erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP) level may be helpful in monitoring response to treatment in some patients, particularly those with polyarticular or axial joint infections (in whom it can be difficult to assess reduction in local inflammation by physical examination). Although these approaches have not been tested in this specific setting, CRP has been shown to be of value in monitoring patients with bacterial endocarditis.[20] The ESR may be normal in patients with septic arthritis, but it is usually elevated.[1,2,12]

DIAGNOSTIC IMAGING

In the absence of trauma, radiographs are of limited utility in the early diagnosis of acute synovitis. Radiographs are useful when there is clinical suspicion of chronic osteomyelitis, osteonecrosis, or pathologic or insufficiency fracture. Sequential radiographs are useful in evaluating a patient for the development of osteomyelitis. Ultrasound imaging is occasionally useful in documenting hip effusions in children or effusions in other joints in children or adults, especially joints having abnormal anatomy caused by prior mechanical or inflammatory trauma. Patients with suspected sternoclavicular or sacroiliac joint infections should have a computed tomography or magnetic resonance imaging scan performed to evaluate for mediastinal or pelvic extension. Because of its superior soft tissue imaging properties, MRI is useful for the evaluation of periarticular abscesses and soft tissue infections (e.g., psoas abscess mimicking septic arthritis of the hip and subdeltoid bursitis communicat-

ing with the shoulder joint). MRI can detect abnormal marrow signal secondary to osteomyelitis or avascular necrosis but may not help in distinguishing between septic and noninfectious inflammatory synovitis.[21]

Radionuclide scans and other imaging procedures are occasionally useful in localizing and defining the extent of infection. Radionuclide bone scans are sensitive early indicators of changes in periarticular bone metabolism. In a study comparing the use of a technetium-99 (^{99m}Tc) scan with the clinical gold standard of synovial fluid culture in patients with septic arthritis, ^{99m}Tc scanning had a sensitivity of 100%.[22] Radionuclide scans are not specific enough to establish a diagnosis. Scanning with ^{99m}Tc is helpful in differentiating bone from adjacent soft tissue inflammation.[23] Gallium-67 (^{67}Ga) localizes to areas of inflammation and increased bone metabolism. In cases of septic arthritis related to intravenous drug use, ^{67}Ga scans showed positive results several days before ^{99m}Tc scans did. The ^{67}Ga scan has been largely replaced by the indium-111 (^{111}In) leukocyte scan. Although less sensitive than ^{99m}Tc, ^{111}In is thought to be more specific for joint infection than either ^{99m}Tc or ^{67}Ga, because leukocytes do not normally migrate to areas of increased bone metabolism associated with aseptic processes. No radionuclide scan will reliably distinguish between infection-induced and crystal-induced arthritis.

In the vast majority of patients with septic arthritis, especially those from whom synovial fluid can be obtained, radionuclide imaging studies are unnecessary.

Differential Diagnosis of Acute Arthritis

In the patient with acute monoarticular or oligoarticular arthritis, the clinician must promptly distinguish between crystal-induced arthritis and infection-induced arthritis, because the initial treatments of these two conditions are strikingly different. The diagnostic test of choice is synovial fluid analysis. Acutely inflamed or swollen, painful joints, in the absence of obvious trauma, should be aspirated. A single drop of fluid (even the small amount in the hub of the needle of an apparently unsuccessful arthrocentesis) may be sufficient to establish a diagnosis of crystal-induced arthritis. Wet preparation can provide an initial cell count estimate and a polarized microscopic evaluation for crystals. The fluid sample should then be air-dried and a Gram stain performed to both screen for bacteria and calculate the percentage of PMNs. The inability to demonstrate eosinophils by use of this stain technique is rarely of importance. Reactive and enteropathic arthritis are diagnoses that should generally be made only after infection has been excluded by culture of synovial fluid.

Specific Infectious Agents

Causative pathogens are documented by culture in approximately two thirds of presumed septic arthritis cases in most series, and identification rates as high as 93% have been reported. A portal of entry can be identified in half of the patients with septic arthritis; approximately 25% of cases are iatrogenic. In most series, 75% of cases have a possible extra-articular source of infection (e.g., cutaneous, respiratory, or genitourinary).

Determining the causative organism has implications for initial and adjuvant therapies, especially if the organism is resistant to particular therapeutic agents.

Traditionally, bacterial arthritis is divided into nongonococcal and gonococcal types. The distinction is clinically useful because gonococcal infections have a better prognosis than nongonococcal infections. Viral arthritis, which tends to have a self-limited course (except in some cases induced by parvovirus and hepatitis B or C), will not be discussed in this chapter.

NEISSERIA GONORRHOEAE

In some practice settings, DGI is a relatively common cause of septic arthritis and tenosynovitis in healthy, sexually active patients. In a review of 41 cases of gonococcal arthritis, 83% of patients were female, and the mean age was 23 years.[17] Knee synovitis was present in 54% of the patients, followed in frequency by hand and wrist synovitis. Involvement of the hip, but not the spine, is common. Dermatitis (usually sparse peripheral necrotic pustules) and migratory polyarthralgias/polyarthritis were present in 39% and 66% of the patients, respectively. Along with tenosynovitis, these findings constitute the classic triad of DGI. Genitourinary involvement was noted in 63% of the patients but is often asymptomatic in women. Absence of pelvic inflammatory disease does not exclude a diagnosis of DGI. Rectal and pharyngeal colonization are commonly asymptomatic; these sites, as well as blood, should be routinely cultured in all patients suspected of having DGI. Female genital infection may have occurred long before systemic dissemination. The frequency of positive cultures in one study of DGI was as follows: urogenital, 86%; synovial fluid, 44%; rectal, 39%; blood, 13%; and pharyngeal, 7%.[17]

Arthritis and vasculitic skin lesions may reflect local metastatic infection or a sterile inflammation induced by immune complexes. This dual pathophysiology may partially account for both the low frequency of positive joint cultures and the favorable joint outcome. Data compiled by the Centers for Disease Control and Prevention indicate that up to 30% of gonococcal isolates obtained in STD clinics in the United States in 1997 were resistant to penicillin or tetracycline.[24] Resistant strains are capable of systemic dissemination. There have been only rare reports of resistance to ceftriaxone; therefore, initial therapy should include parenteral therapy with ceftriaxone or ciprofloxacin. Fluoroquinolone resistance is now rarely documented.[24] Appropriate regimens, once the infection has been identified, include ceftriaxone (1 to 2 g intramuscularly or intravenously once daily), imipenem (0.5 g intravenously every 6 hours), or cefotaxime (1 g intravenously every 8 hours until 24 to 48 hours after resolution of symptoms). If DGI is clinically suspected but no organism is cultured, empirical therapy with a broad-spectrum cephalosporin is generally continued for approximately 1 week until the lack of any response is documented. Subsequent outpatient therapy should be continued with ciprofloxacin (500 mg orally b.i.d.) or ampicillin (500 to 1,000 mg orally q.i.d.), if the identified pathogen shows sensitivity to these agents. Recognized infection should always prompt an evaluation for other STDs, including syphilis and HIV. Empirical treatment for *Chlamydia trachomatis* infection (doxycycline, 100 mg orally b.i.d. for 7 days, or azithromycin, 1 g orally in a single dose) should also be given because this infection is frequently asymptomatic and can result in infertility if untreated; both partners should be treated whenever possible. Dosages may need to be modified in the presence of severe renal insufficiency. Once the diagnosis is confirmed and clinical response exhibited, addition of an NSAID is a reasonable adjunctive therapy because, as noted above, the synovitis may be induced by immune complexes as well as direct infection.

A dermatitis-arthritis syndrome, similar to DGI, can also be associated with *Haemophilus influenzae, N. meningitidis,* and *Streptobacillus moniliformis* infections, as well as with endovascular infections. Disseminated *Neisseria* infections, which may be recurrent, have been associated with the presence of terminal complement deficiencies.

STAPHYLOCOCCUS

Gram-positive bacteria remain the most common cause of septic arthritis, accounting for 70% to 80% of cases. *S. aureus* accounts for more than half of the cases of culture-positive septic arthritis in studies at university hospitals and for even higher percentages of certain patient subgroups: 70% to 80% of patients with polyarticular septic arthritis; more than 80% of infected patients with rheumatoid arthritis (RA)[25]; and 82% of infected hemodialysis patients.[26] Staphylococcal arthritis was particularly frequent in a series of patients with endocarditis related to intravenous drug abuse.[27] The predominance of *S. aureus* in septic arthritis has remained unchanged over the past 40 years,[1] and there is now an increasing prevalence of methicillin-resistant *S. aureus* (MRSA) and *S. epidermidis* in patients with prosthetic joints. *S. aureus* should always be considered when selecting antibiotics in the initial treatment of presumed septic arthritis.

Gram stain cannot be relied on to differentiate between *Staphylococcus* and *Streptococcus,* because in biologic smears, *Staphylococcus* may not exhibit the clusters seen when grown in vitro. Bacteremia should be assumed in cases of staphylococcal arthritis, although it cannot always be documented. Endovascular infection with possible metastatic seeding to other areas should be considered in patients with persistent fever, leukocytosis, thrombocytosis, or an elevated ESR or CRP level.

Suspected staphylococcal joint infection should be treated initially with vancomycin (1 g intravenously every 12 hours, if renal function is normal) until methicillin resistance can be excluded. An appropriate penicillin or a cephalosporin can often be substituted for completion of a 4- to 6-week treatment course once the sensitivity to these agents is identified.

STREPTOCOCCUS

Non–group A, β-hemolytic streptococci are the second most common cause of septic arthritis, accounting for 10% to 21% of culture-positive cases.[1-3] The number of reported group B (and to a lesser extent, groups C and G) streptococcal infections has been increasing.[28] Group B streptococcal infection may be particularly virulent in diabetic patients; it involves axial joints (sacroiliac, sternoclavicular, and manubriosternal) and may be associated with poor functional outcome. Other manifestations of group B streptococcal sepsis include myositis, fasciitis, and endophthalmitis. Some investigators have expressed concern over the development of tolerance to penicillin in these usually penicillin-sensitive organisms and have suggested adding an aminoglycoside to therapy or using alternative antibiotics until sensitivity and bactericidal tests are available. *S. pneumoniae* may account for up to 10% of the cases of septic arthritis. In one study, 11% of patients with septic arthritis requiring transfer to the intensive care unit had pneumococcal infection.[5] The incidence of penicillin-resistant streptococci, including *S. pneumoniae*, is rising. Polyarticular pneumococcal infections occur, which is not surprising given the high incidence of bacteremia in patients with pneumococcal pneumonia. Elderly adults may be particularly susceptible, and fever or an extra-articular source of infection may not be evident.[29] Antibiotic therapy must be tailored with knowledge of local resistance patterns. *Enterococci* can occasionally infect native or prosthetic joints.[30]

Diagnosis is made on the basis of a Gram stain or preliminary culture showing gram-positive cocci; for initial therapy, vancomycin (1 g every 12 hours, with normal renal function) is a reasonable choice. Definitive therapy should be determined on the basis of culture results.

GRAM-NEGATIVE ORGANISMS

Gram-negative organisms account for 9% to 20% of septic arthritis cases reported at teaching hospitals and occur most frequently in immunosuppressed patients.[1-4] Although certain pathogens have been associated with specific patient subgroups (e.g., *Salmonella* in patients with systemic lupus erythematosus or sickle cell anemia, *Pseudomonas aeruginosa* in intravenous drug abusers), joint infections with these organisms can occur in immunocompetent patients. In sickle cell patients, *Salmonella* more commonly causes osteomyelitis than septic arthritis and may infect multiple sites.[31]

Comorbid medical conditions, prior antibiotic use, and extra-articular infections (particularly urinary tract infections and decubitus ulcers) predispose to gram-negative septic arthritis. Reviews of septic arthritis in intravenous drug abusers over the past decade report a lower percentage of gram-negative pathogens than were reported in earlier series, in which organisms such as *Enterobacter, Klebsiella, Serratia*, and *P. aeruginosa* accounted for more than 80% of infections.[27] Pathogens and sensitivities associated with intravenous drug abuse are specific to geographic regions, and choice of antimicrobial therapy for this patient subgroup must take these factors into account. This same population is also at risk for HIV and viral hepatitis.

Ampicillin is bactericidal for *Salmonella;* however, an increasing number of *Salmonella* infections are resistant to this drug. Chloramphenicol, trimethoprim-sulfamethoxazole, and especially quinolones are used as alternative agents.[31] For other gram-negative infections, cefotaxime (2 g intravenously every 6 hours) or imipenem (0.5 g intravenously every 6 hours) can be used with or without an aminoglycoside until culture results arrive. If pseudomonal infection is suspected, it is reasonable to initiate dual drug therapy, including an aminoglycoside.

H. influenzae joint infections in children are becoming far less common since the widespread institution of immunization programs.

ANAEROBES

Anaerobic joint infections are rare, accounting for at most 7% of cases in reported series; however, they should be considered in patients with deep-tissue or abdominal abscesses. Intra-articular emphysema has been described with anaerobic infection (*Clostridium* species) but is not specific to anaerobes.[32] Empirical treatment for anaerobic septic arthritis is generally not undertaken unless there is a coincident infection with a high likelihood of anaerobic involvement. In those cases, treatment should be based on the suspected organisms and the portal of infection (oral, abdominal, or traumatic).

UREAPLASMA UREALYTICUM

U. urealyticum septic arthritis has been reported in hypogammaglobulinemic patients. These patients, who are often already receiving γ-globulin replacement therapy, may develop a destructive monoarthritis, oligoarthritis, or, in rare instances, polyarthritis[33] with fever and subcutaneous nodules. Positive cul-

tures from affected synovial fluid or tissue are notoriously difficult to obtain. PCR now offers the potential for rapid, definitive diagnosis. Eradication of the infection may require several months of tetracycline therapy. Although rapid recognition and treatment have resulted in complete remission,[34] not all patients respond. The microbiology laboratory should be alerted to the possibility of this infection before fluid samples are taken for culture.

This destructive *U. urealyticum* septic arthritis should be distinguished from the rheumatoid-like, but generally nonerosive, sterile polyarthritis associated with hypogammaglobulinemia. Initial treatment of acute-onset arthritis in this setting should include both a tetracycline (e.g., doxycycline, 100 mg every 12 hours) and aggressive intravenous immunoglobulin replacement until *Ureaplasma* infection is excluded.

PASTEURELLA MULTOCIDA

Pasteurella multocida, a pleomorphic gram-negative coccobacillus parasitic to many cats and dogs, is an infrequent cause of septic arthritis.[35] Infection generally follows inoculation from an animal bite, although in one third of cases no trauma is documented. The organism is almost always sensitive to a penicillin; aminoglycosides are generally ineffective. Initial antibiotic coverage after a recent animal bite should include a penicillin or a third-generation cephalosporin. Ampicillin-sulbactam (1.5 g intravenously every 6 hours) and ceftriaxone (2 g intravenously q.d.) are reasonable initial therapies pending final culture results.

MYCOBACTERIUM MARINUM

Mycobacterium marinum is an atypical mycobacterium acquired through exposure to freshwater, saltwater, swimming pools, or fish tanks. *M. marinum* usually causes cutaneous and subcutaneous infection, but it can cause subcutaneous nodules, tenosynovitis, and septic arthritis.[36,37] Monoarticular, often proliferative, involvement of the metacarpophalangeal and proximal interphalangeal joints is most frequently reported. The arthritis may be only slightly painful with no systemic symptoms. Delays in diagnosis are common (mean, 8 months). Optimal growth of the organism from synovial fluid or tissue requires special attention in the microbiology laboratory. *M. marinum* infection may elicit a positive purified protein derivative (PPD) reaction. Treatment usually consists of prolonged chemotherapy and surgical debridement. There are no evidence-based guidelines to aid in determining the need for initial surgical debridement, although this procedure is frequently utilized, especially when the course of infection is protracted. A reasonable initial therapy is a combination of rifampin (600 mg orally) and ethambutol (15 mg/kg/day; maximum, 1,600 daily), although treatment with trimethoprim-sulfamethoxazole (160/800 mg orally b.i.d.) or ciprofloxacin (500 to 750 mg orally b.i.d.) may also be effective. Although there are no data demonstrating the superiority of combination therapy, there has been anecdotal treatment failure with doxycycline monotherapy. The organisms are usually sensitive to amikacin and clarithromycin.

MYCOBACTERIUM TUBERCULOSIS

Arthritis is an uncommon extrapulmonary complication of tuberculosis[38,39]; osteomyelitis of the low thoracic and lumbar spine occurs more frequently.[40] Recognition of tuberculous peripheral joint and bursa infections are often significantly delayed, partly because of the indolent nature of the inflammation, the rarity of the disease, and the frequent absence of obvious pulmonary disease or a positive PPD test.[38] Although joint motion is frequently maintained, radiographs may show destruction of bone and loss of cartilage. Synovial fluid leukocytosis is modest to marked, usually with a predominance of neutrophils. Acid-fast staining of fluid is usually negative; cultures are usually positive. Limited data suggest that biopsy of synovium may yield a higher percentage of positive cultures than do cultures of fluid. Because *M. tuberculosis* is a rare cause of peripheral arthritis, it is not cost-effective to obtain mycobacterial cultures initially from all patients with acute synovitis. Microbiologic diagnosis takes several weeks; therefore, molecular diagnostic tests for *M. tuberculosis* DNA should be requested only if there is a significant clinical concern about this infection. No large studies have been undertaken to compare diagnostic approaches or treatment regimens for tuberculous arthritis. As with therapy for pulmonary tuberculosis, there is a potential for developing drug resistance; thus, initial therapy should include a combination of at least two drugs, including rifampin (600 mg/day), isoniazid (300 mg/day), and perhaps ethambutol (15 mg/kg/day) for the first few months of a planned 18- to 24-month course of treatment.

TREPONEMA PALLIDUM

The arthritis associated with later stages of syphilis is characteristically a chronic, proliferative, symmetrical polyarthritis involving small and large joints.[41] An acute monoarthritis is not typical of syphilis. Synovial fluid usually contains a predominance of mononuclear cells. Treatment should be based on the stage of disease (usually secondary or tertiary) and the presence or absence of HIV infection [*see 7:VI Syphilis and the Nonvenereal Treponematoses*]. Penicillin remains the antibiotic of choice, although ceftriaxone or doxycycline may be used in some patients without HIV infection.

FUNGAL INFECTIONS

Fungal arthritis is rare and is seemingly more common in immunosuppressed patients. The arthritis is usually subacute or chronic. Osteomyelitis is more common than arthritis in *Cryptococcus, Coccidioides,* and blastomycosis infections.[42,43] Arthritis associated with histoplasmosis may occur with erythema nodosum and may clinically mimic sarcoidosis. Arthritis caused by aspergillosis infection has been reported after systemic infection and iatrogenic inoculation at the time of arthrocentesis. Sporotrichosis arthritis can occur after cutaneous inoculation or, in rare cases, through inhalation.[44] Surgical debridement may be helpful; there are limited data, however, to support initial surgery in addition to antifungal therapy. Except in cases of local inoculation (e.g., rose thorn sporotrichosis), the fungal infection is generally assumed to be systemic, and intravenous amphotericin B should be administered. Intra-articular amphotericin B and oral regimens with azole antibiotics have been utilized. These alternative agents and routes of administration should be considered on an individual basis; consultation with an infectious disease specialist should take place early in the treatment program.

Clinical Subgroups at Risk for Septic Arthritis

Several factors predispose certain groups of patients to septic arthritis. Local risk factors include the presence of inflammatory arthritis, prior intra-articular corticosteroid injection, and the presence of implanted synthetic material.[6] Other factors in-

clude advanced age, a remote focus of infection, diabetes mellitus, indwelling lines, intravenous drug use, and immunosuppression.

ELDERLY PATIENTS

Age older than 60 years has been repeatedly found to be a risk factor for septic arthritis, perhaps in part because of comorbid conditions.[1,6] Prosthetic joints, which also pose an increased risk of infection, are more common in older patients.

The clinical presentation of septic arthritis tends to be more insidious in elderly patients than in younger patients. The absence of fever is common.[29] Older patients are particularly at risk for septic arthritis as a postoperative complication, often after a bacteremic episode. In a retrospective study of 21 patients older than 60 years with septic arthritis, mortality was 21%.[2]

RHEUMATOID ARTHRITIS PATIENTS

Patients with RA have consistently accounted for a disproportionate percentage of patients with septic arthritis; in some series, 50% of septic arthritis patients had RA. These patients may be predisposed to septic arthritis because of poor clearance of bacteria from abnormal joints or because of phagocytic defects acquired secondary to drugs or disease. Patients with longstanding, severe RA are more likely to develop septic arthritis than patients with less severe disease. The majority of RA patients with septic arthritis present with insidious worsening of joint symptoms, and the diagnosis of infection is often delayed.[25] Although there have been reports of so-called pseudoseptic rheumatoid joints characterized by highly inflammatory, sterile synovial fluids that improve with antirheumatic therapy, any isolated, acutely inflamed joint in a patient with RA should be considered infected until proved otherwise.

With septic arthritis, the outcome is significantly worse for RA patients than nonrheumatoid patients, perhaps because of delays in diagnosis. Reported mortality for RA patients with monoarticular infection has ranged from 15% to 22%. When both monoarticular and polyarticular infections are included, mortality reaches 44% in RA patients, compared with 26% in unselected septic arthritis patients. For patients with a combination of polyarticular septic arthritis and RA, mortality is 50%. Joint outcome is also poor; only 35% of RA patients achieve full functional recovery of the infected joint, compared with 70% of nonrheumatoid patients. A retrospective literature review of 181 RA patients reported recurrence of infection in nonprosthetic joints in 20% of patients.[25]

INTRAVENOUS DRUG ABUSERS

Intravenous drug abusers are at risk for repetitive transient bacteremia and direct bacterial inoculation of soft tissue. Bacteremia is thought to be the most common mechanism of infection. Although medium to large joints (such as the wrist and knee) are most commonly involved, fibrocartilaginous joints (sternoclavicular, sacroiliac) and the spine are frequently infected. Gram-positive organisms (especially *S. aureus* and MRSA) are the most common pathogens,[27] followed by gram-negative bacteria (*Enterobacter, Serratia,* and *P. aeruginosa*).

Not all studies of septic arthritis in intravenous drug abusers have specified whether the patients were also HIV positive. Two Spanish studies, one with 25 patients and one with 16 patients, reported no differences in clinical characteristics, infecting organisms, or outcome between intravenous drug abusers with HIV and those without HIV.[45]

HEMODIALYSIS PATIENTS

Hemodialysis patients are at increased risk for septic arthritis because of recurrent vascular infections, intravenous catheterizations, repeated skin trauma, and immune deficits such as decreased clearance of transient bacteremia. Hemodialysis patients are at particularly high risk for *S. aureus* colonization, with reported nasal carriage rates over 40%.

A 10-year review of hemodialysis patients in three tertiary care hospitals identified 11 episodes of septic arthritis, with 82% attributable to *S. aureus*.[26] Almost all of the episodes involved joints above the diaphragm, including one sternoclavicular and one acromioclavicular joint infection.

HIV PATIENTS

Two articular syndromes unique to HIV-infected patients have been reported that can mimic acute septic arthritis. The first, which has been referred to as AIDS-associated arthritis, is characterized by the development over several days of extreme pain and disability involving the knees or ankles, noninflammatory sterile synovial fluid, and an excellent response to nonsteroidal anti-inflammatory drugs (NSAIDs). The reported duration of symptoms is 1 to 6 weeks. The second syndrome, the painful articular syndrome, is characterized by an explosive onset of pain in the shoulders and elbows. The pain resembles gout in its intensity and often requires intravenous narcotics for analgesia. Synovial fluid is noninflammatory, and symptoms last a few hours to several days. The etiology of these syndromes remains unknown.

Despite the multiple immune deficits associated with HIV infection, there have been relatively few reported cases of septic arthritis in patients with HIV. *S. aureus* and *S. pneumoniae* are the most commonly reported organisms.[46] With an increasing degree of $CD4^+$ T cell depletion, infection with *Candida albicans, Cryptococcus neoformans, Histoplasma capsulatum, Nocardia asteroides,* and *Sporothrix schenckii* can occur. HIV patients are also at increased risk for pyomyositis, most commonly occurring secondary to *S. aureus*.[46]

ORGAN TRANSPLANTATION PATIENTS

A review of the University of Pennsylvania renal transplant experience reported the occurrence of septic arthritis in 0.8% of the patient population.[47] All cases occurred within 18 months of transplantation, the period of most intense immunosuppression with high-dose steroids. Infection was caused by bacteria in 66% of cases (half of which were caused by *S. aureus*), mycobacteria in 17%, and disseminated cytomegalovirus infection in 17%. Only three cases of fungal arthritis have been reported in renal transplant patients, but this incidence is likely underreported. With the current ubiquitous use of cyclosporine, gouty arthritis (occasionally polyarticular with atypical joint involvement) is extremely common and may mimic septic arthritis.[48]

INDIVIDUALS WITH PROSTHETIC JOINTS

Prosthetic joint infections, in the absence of wound drainage, are frequently not recognized early and may result in a nonfunctional joint with chronic osteomyelitis. Pain is the predominant or only symptom.[49] Effusions may be small (knee) or impossible to detect clinically (hip or shoulder). Fever and leukocytosis are frequently absent. In studies, *S. epidermidis* was responsible for approximately 40% of both early postoperative and late infections.[49] Although the ESR is usually elevated, it may be normal.[50,51] One study demonstrated a sensitivity of only

Table 2 Initial Empirical Antibiotic Therapy for Acute Septic Arthritis in Adults*

Findings on Gram Stain	*Drug of Choice*	*Alternative Drug*	*Comments*
Gram-positive cocci (small) in pairs and chains	Vancomycin, 1.0 g I.V. every 12 hr†	Cefotaxime, 2.0 g I.V. every 6–8 hr	Smear suggests streptococci or pneumococci; if only single cocci are observed or if cocci are large, treat as for *Staphylococcus aureus* (see next entry)
Gram-positive cocci (large) singly or in small groups	Vancomycin, 1.0 g I.V. every 12 hr	Nafcillin, 2.0 g I.V. every 4 hr, or oxacillin, 2.0 g I.V. every 4 hr‡	Assume that penicillinase-producing *S. aureus* and, possibly, methicillin-resistant strains are present
Gram-negative cocci	Ceftriaxone, 2.0 g I.V. every 24 hr	Imipenem, 0.5 g I.V. every 6 hr	Usually *Neisseria* species; treat with third-generation cephalosporin or imipenem in documented or suspected cases of infections caused by penicillin-resistant *Neisseria* species
Gram-negative bacilli	Cefotaxime, 2.0 g I.V. every 6 hr§	Imipenem, 0.5 g I.V. every 6 hr	Antibiotic susceptibility testing mandatory; fluoroquinolones may be of value
No organisms seen			
Healthy, young patient	Ceftriaxone, 2.0 g I.V. every 24 hr with or without cephalothin, 1 g I.V. every 8 hr; use vancomycin if methicillin-resistant strains are a possibility	Imipenem, 0.5 g I.V. every 6 hr	Probably gonococcal disease, but initial therapy should also treat gram-positive cocci
Patient with underlying disease (e.g., neoplasm, rheumatoid arthritis, intravenous drug use) or immunosuppression	Vancomycin, 1.0 g I.V. every 12 hr, and ciprofloxacin, 400 mg I.V. every 12 hr	Imipenem, 0.5 g I.V. every 6 hr	Broad-spectrum coverage for staphylococci and many types of gram-negative bacilli§

*These recommendations are for nonallergic patients with normal renal function. *Final antibiotic choice should be determined by culture result.* Refer to Table 1 and to text for other special situations.
†Sensitivity of *S. pneumoniae* to penicillin should not be assumed. In geographic areas where penicillin-resistant pneumococcus is not present, a penicillin can be used.
‡It is imperative to document that staphylococci are sensitive to nafcillin or to oxacillin.
§In neutropenic patients or in patients at risk for *Pseudomonas* infection, consider using an aminoglycoside and either ceftazidime or piperacillin.

60% when an ESR threshold greater than 30 mm/hr was used.[50] The CRP assay may be a more sensitive test.[51] Three-phase bone scans are also surprisingly insensitive.[51] If loosening is noted on standard radiographs, bone scanning is of little additional diagnostic value. Arthrocentesis or intraoperative tissue sampling with culture, before antibiotic therapy, remains the gold standard for diagnosis. The value of fluid Gram stain is questionable.[51] Knee replacement can be repeated in some patients as a two-step procedure. Total debridement, followed by 4 to 6 weeks of systemic antibiotic therapy before attempted repeat arthroplasty, is frequently successful.[52]

Treatment of Septic Arthritis

DRAINAGE

Treatment of septic arthritis consists of drainage, parenteral antibiotics (guided by Gram stain, if possible [*see Table 1*]), and initial (not prolonged) joint immobilization for pain control. The controversy surrounding medical versus surgical drainage remains unresolved. A retrospective analysis by Broy and Schmid included pooled data from 80 studies.[53] A slightly larger number of patients who were treated medically (needle drainage) rather than surgically (arthrotomy or arthroscopy) had a good joint outcome (66% versus 57% [$P < 0.05$]) with restoration of normal joint function and minimal residual pain. Perhaps because of increased systemic comorbidity in the medically treated cohort—an inherent problem in nonrandomized studies—medically treated patients had a higher mortality (10% versus 3%; $P < 0.05$). These nonrandomized studies do not permit generation of firm therapeutic guidelines.

Clinicians generally agree that surgical drainage is indicated in the following situations: septic arthritis of the hip or of joints that are difficult to aspirate or monitor for adequate drainage, extensive spread of infection to the soft tissues, and inadequate clinical response to appropriate antibiotics after 5 to 7 days. There is an apparent consensus among rheumatologists that surgical drainage is rarely necessary for patients with gonococcal septic arthritis in whom the response to appropriate antibiotics is characteristically rapid (less than a week) and complete. No prospective or controlled studies support or refute the need to openly drain or lavage an uncomplicated, acutely infected, nongonococcal joint at the initiation of therapy; nonetheless, this orthopedic practice is common in many parts of the country. Surgical drainage should be considered when a previously damaged (e.g., rheumatoid) joint becomes infected, because the data indicate a particularly poor outcome for these patients with medical therapy.[25] The data are insufficient to mandate immediate surgical drainage of infected joints in immunosuppressed patients. In general practice, arthroscopic lavage seems to be supplanting open joint drainage; it permits earlier physical therapy and mobilization of the involved joint.[54]

MEDICAL THERAPY

The choice of antibiotics [*see Table 2*] should be guided by patient characteristics and the local microbiologic sensitivity pattern. Intra-articular antibiotics are not required[55] and may cause a chemical synovitis. Parenteral antibiotics are generally administered for approximately 4 to 6 weeks. Acetaminophen and NSAIDs should be avoided until the diagnosis is solid (i.e., through microbiologic identification or by documented defervescence and clinical improvement with antibiotics alone). High-dose NSAIDs can transiently blunt the inflammatory response and may suggest a response to coadministered antibiotics, which might then be construed as evidence for a culture-negative (e.g., gonococcal) infection. There are no data demon-

strating that NSAIDs prolong the course of infection, and some animal data suggest that anti-inflammatory therapy may be beneficial.[10,11] Narcotics may be used initially for pain control. Joint immobilization for a short period (1 to 3 days) is appropriate to help with pain control, but daily range-of-motion exercises should be started as soon as possible to hasten return of joint function.[56] Once a definitive diagnosis is made, and if there are no contraindications, NSAID therapy may be initiated to facilitate joint mobilization.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Weston VC, Jones AC, Bradbury N, et al: Clinical features and outcome of septic arthritis in a single UK health district 1982–1991. Ann Rheum Dis 58:214, 1999

2. Shirtliff ME, Mader JT: Acute septic arthritis. Clin Microbiol Rev 15:527, 2002

3. Gupta MN, Sturrock RD, Field M: A prospective 2-year study of 75 patients with adult onset septic arthritis. Rheumatology 40:24, 2001

4. Rosenthal J, Bole GG, Robinson WD: Acute nongonococcal infectious arthritis: evaluation of risk factors, therapy, and outcome. Arthritis Rheum 23:889, 1980

5. Martens PB, Ho G Jr: Septic arthritis in adults: clinical features, outcome, and intensive care requirements. J Intensive Care Med 10:246, 1995

6. Kanndorp CJE, Krijnen P, Moens HJB, et al: The outcome of bacterial arthritis: a prospective community-based study. Arthritis Rheum 40:884, 1997

7. McAllister DR, Parker RD, Cooper AE, et al: Outcomes of postoperative septic arthritis after anterior cruciate ligament reconstruction. Am J Sports Med 27:562, 1999

8. Nair SP, Williams RJ, Henderson B: Advances in our understanding of the bone and joint pathology caused by *Staphylococcus aureus* infection. Rheumatology 39:821, 2000

9. Bremell T, Abdelnour A, Tarkowski A: Histopathological and serological progression of experimental *Staphylococcus aureus* arthritis. Infect Immunol 60:2976, 1992

10. Wysenbeek AJ, Volchek J, Amit M, et al: Treatment of staphylococcal septic arthritis in rabbits by systemic antibiotics and intra-articular corticosteroids. Ann Rheumatic Dis 57:687, 1998

11. Stricker SJ, Lozman PR, Makowski AL, et al: Chondroprotective effect of betamethasone in lapine pyogenic arthritis. J Pediatr Orthop 16:231, 1996

12. Gupta MN, Sturrock RD, Field M: Prospective comparative study of patients with culture proven and high suspicion of adult onset septic arthritis. Ann Rheum Dis 62:327, 2003

13. Yagupsky P, Bar-Ziv Y, Howard CB, et al: Epidemiology, etiology, and clinical features of septic arthritis in children younger than 24 months. Arch Pediatr Adolesc Med 149:537, 1995

14. Ho G Jr, DeNuccio M: Gout and pseudogout in hospitalized patients. Arch Intern Med 153:2787, 1993

15. Garcia-Porrua C, Gonzalez-Gay MA, Ibanez D, et al: The clinical spectrum of severe septic bursitis in northwestern Spain: a 10-year study. J Rheumatol 26:663, 1999

16. Krey PR, Bailen DA: Synovial fluid leukocytosis: a study of extremes. Am J Med 67:436, 1979

17. Wise CM, Morris CR, Wasilauskas BL, et al: Gonococcal arthritis in an era of increasing penicillin resistance: presentations and outcomes in 41 recent cases (1985–1991). Arch Intern Med 154:2690, 1994

18. Yu KH, Luo SF, Liou LB, et al: Concomitant septic and gouty arthritis: an analysis of 30 cases. Rheumatology 42:1062, 2003

19. Liebling MR, Arkfeld DG, Michelini GA, et al: Identification of *Neisseria gonorrhoeae* in synovial fluid using the polymerase chain reaction. Arthritis Rheum 37:702, 1994

20. Olaison L, Hogevik H, Alestig K: Fever, C-reactive protein, and other acute-phase reactants during treatment of infective endocarditis. Arch Intern Med 157:885, 1997

21. Graif M, Schweitzer ME, Deely D, et al: The septic versus nonseptic inflamed joint: MRI characteristics. Skeletal Radiol 28:616, 1999

22. Mudun A, Unal S, Aktay R, et al: Tc-99m nanocolloid and Tc-99m MDP three-phase bone imaging in osteomyelitis and septic arthritis: a comparative study. Clin Nucl Med 20:772, 1995

23. Aliabadi P, Nikpoor N: Imaging osteomyelitis. Arthritis Rheum 37:617, 1994

24. Fox KK, Knapp JS, Holmes KK, et al: Antimicrobial resistance in *Neisseria gonorrhoeae* in the United States, 1988–1994: the emergence of decreased susceptibility to the fluoroquinolones. J Infect Dis 175:1396, 1997

25. Gardner GC, Weisman MH: Pyarthrosis in patients with rheumatoid arthritis: a report of 13 cases and a review of the literature from the past 40 years. Am J Med 88:503, 1990

26. Slaughter S, Dworkin RJ, Gilbert DN, et al: *Staphylococcus aureus* septic arthritis in patients on hemodialysis treatment. West J Med 163:128, 1995

27. Sapico FL, Liquete JA, Sarma RJ: Bone and joint infections in patients with infective endocarditis: review of a 4-year experience. Clin Infect Dis 22:783, 1996

28. Schattner A, Vosti KL: Bacterial arthritis due to beta-hemolytic streptococci of serogroups A, B, C, F, and G. Medicine (Baltimore) 77:122, 1998

29. Ross JJ, Saltzman CL, Carling P, et al: Pneumococcal septic arthritis: review of 190 cases. Clin Infect Dis 36:319, 2003

30. Raymond NJ, Henry J, Workowski KA: Enterococcal arthritis: case report and review. Clin Infect Dis 21:516, 1995

31. Anand AJ, Glatt AE: Salmonella osteomyelitis and arthritis in sickle cell disease. Semin Arthritis Rheum 24:211, 1994

32. Roy R, Matei D, Kaufman LD: Emphysematous septic arthritis: case report and review of the literature. J Rheumatol 22:1776, 1995

33. Lapusan DL, Duffy L, Hamilos DL: Disseminated septic arthritis in a patient with common variable hypogammaglobulinemia. Ann Allergy Asthma Immunol 86:368, 2001

34. Franz A, Webster AD, Furr PM, et al: Mycoplasmal arthritis in patients with primary immunoglobulin deficiency: clinical features and outcome in 18 patients. Br J Rheumatol 36:661, 1997

35. Kumar A, Kannampuzha P: Septic arthritis due to *Pasteurella multocida*. South Med J 85:329, 1992

36. Alloway JA, Evangelisti SM, Sartin JS: *Mycobacterium marinum* arthritis. Semin Arthritis Rheum 24:382, 1995

37. Harth M, Ralph ED, Faraawi R: Septic arthritis due to *Mycobacterium marinum*. J Rheumatol 21:957, 1994

38. Malaviya AN, Kotwal PP: Arthritis associated with tuberculosis. Best Pract Res Clin Rheumatol 17:319, 2003

39. Wallace R, Cohen AS: Tuberculosis arthritis: a report of two cases with review of biopsy and synovial fluid findings. Am J Med 61:277, 1976

40. Kotevoglu N, Tasbasi I: Diagnosing tuberculous spondylitis: patients with back pain referred to a rheumatology outpatient department. Rheumatol Int 24:9, 2003

41. Reginato AJ: Syphilitic arthritis and osteitis. Rheum Dis Clin North Am 19:379, 1993

42. Holley K, Muldoon M, Tasker S: *Coccidioides immitis* osteomyelitis: a case series review. Orthopedics 25:827, 2002

43. Abril A, Campbell MD, Cotton VR Jr, et al: Polyarticular blastomycotic arthritis. J Rheumatol 25:1019, 1998

44. Perez-Gomez A, Prieto A, Torresano M, et al: Role of the new azoles in the treatment of fungal osteoarticular infections. Semin Arthritis Rheum 27:226, 1998

45. Rivera J, Lopez-Longo J, Sanchez-Atrio A: Septic arthritis in patients with acquired immunodeficiency syndrome with human immunodeficiency virus infection. J Rheumatol 19:1960, 1992

46. Vassilopoulos D, Chalasani P, Jurado RL, et al: Musculoskeletal infections in patients with human immunodeficiency virus infection. Medicine (Baltimore) 76:284, 1997

47. Bomalaski JS, Williamson PK, Goldstein CS: Infectious arthritis in renal transplant patients. Arthritis Rheum 29:227, 1986

48. George T, Mandell BF: Gout in the transplant patient. J Clin Rheum 1:328, 1995

49. Lidgren L, Knutson K, Stefansdottir A: Infection and arthritis: infection of prosthetic joints. Best Pract Res Clin Rheumatol 17:209, 2003

50. Levitsky KA, Hozack WJ, Balderston RA, et al: Evaluation of the painful prosthetic joint: relative value of bone scan, sedimentation rate, and joint aspiration. J Arthroplasty 6:237, 1991

51. Spangehl MJ, Masri BA, O'Connell JX, et al: Prospective analysis of preoperative and intraoperative investigations for the diagnosis of infection at the sites of 202 revision total hip arthroplasties. J Bone Joint Surg 81A:672, 1999

52. Hirakawa K, Stulberg BN, Wilde AH, et al: Results of 2-stage reimplantation for infected total knee arthroplasty. J Arthroplasty 13:22, 1998

53. Broy SB, Schmid FR: A comparison of medical drainage (needle aspiration) and surgical drainage (arthrotomy or arthroscopy) in the initial treatment of infected joints. Clin Rheum Dis 12:501, 1986

54. Wirtz DC, Marth M, Miltner O, et al: Septic arthritis of the knee in adults: treatment by arthroscopy or arthrotomy. Int Orthop 25:239, 2001

55. Hamed KA, Tam JY, Prober CG: Pharmacokinetic optimisation of the treatment of septic arthritis. Clin Pharmacokinet 31:156, 1996

56. Pfeiffenberger J, Meiss L: Septic conditions of the shoulder: an updating of treatment strategies. Arch Orthop Trauma Surg 115:325, 1996

Acknowledgment

The author thanks Dr. Mathilde Pioro for her assistance in the preparation of an initial draft of the manuscript.

XVI OSTEOMYELITIS

Layne O. Gentry, m.d., f.a.c.p.

Osteomyelitis is an infection of bone characterized by progressive inflammatory destruction of bone, bone necrosis, and new bone formation. Osteomyelitis poses a challenge because it is often difficult to diagnose and treat. Even when appropriate antibiotic therapy is instituted promptly, osteomyelitis can cause serious morbidity. Chronic cases are often refractory to treatment. In diabetic patients, amputation to remove the site of infection is often necessary. The etiology of osteomyelitis has evolved considerably over the past 2 decades. Twenty years ago, most cases of osteomyelitis were caused by susceptible strains of *Staphylococcus aureus* and were cured with surgical debridement and 4 to 6 weeks of methicillin therapy. Although *S. aureus* is still a common cause of osteomyelitis, an increasing number of infections are caused by gram-negative organisms, such as *Pseudomonas aeruginosa*, or are polymicrobial. Coagulase-negative staphylococci, once regarded as contaminants, are now considered significant pathogens. Increasingly, osteomyelitis is a complication of reconstructive orthopedic surgery; it is also a complication of median sternotomy after cardiothoracic surgery. The growing incidence of diabetes in the population is also responsible for increasing numbers of osteomyelitis diabetes foot infections. These changes, together with improvements in diagnosis and therapy, indicate a need to reexamine traditional management of osteomyelitis.

Classification

Osteomyelitis is classified on the basis of pathogenesis, duration of disease, extent or type of bone involvement, anatomy of the bone infection, and host characteristics. Osteomyelitis can present as either an acute condition or a chronic infection. Acute osteomyelitis is usually defined as the first clinical episode, complete with the signs, symptoms, and radiographic and histologic findings associated with bone infection. Acute infections often occur soon after an episode of bacteremia and are usually accompanied by fever and bone pain. Chronic osteomyelitis is defined as bone infection that has failed to resolve after one or more treatment attempts. Retained foci of dead bone, or sequestra, are characteristic of chronic osteomyelitis.

Waldvogel and coworkers described three types of osteomyelitis: hematogenous osteomyelitis, osteomyelitis secondary to contiguous focus of disease, and osteomyelitis associated with vascular insufficiency.[1] In hematogenous osteomyelitis, infection is introduced through the bloodstream and usually affects the metaphyses of long bones. Hematogenous osteomyelitis is diagnosed in about 20% of cases of osteomyelitis; it is the most common type of osteomyelitis in children. In osteomyelitis secondary to a contiguous focus of disease, the bone becomes infected from an external source (e.g., penetrating trauma or open fracture) or from the spread of infection from adjacent soft tissue (e.g., abscess or skin infection). Contiguous-focus disease is usually seen in older persons and accounts for 50% of cases of osteomyelitis. Osteomyelitis associated with vascular insufficiency, which is responsible for about 30% of cases, is usually seen in patients with diabetic neuropathy.

Clinical Syndromes

hematogenous osteomyelitis

Hematogenous osteomyelitis is usually seen in children between 1 and 15 years of age and adults older than 50 years or in persons who abuse I.V. drugs. In children, infection usually occurs as a single focus in the metaphyseal area of long bones (particularly the tibia and femur). In older persons, infection of the vertebral bodies is more common. When osteomyelitis is seen in young adults, it is often associated with I.V. drug abuse or sickle cell anemia. Several predisposing factors are associated with the development of hematogenous osteomyelitis. Children may be predisposed to infection by minor trauma that causes a small hematoma, vessel obstruction, and bone necrosis. Thus, hematogenous osteomyelitis may be seen in teenagers who play contact sports. About one third of patients have a history of trauma to the site of osteomyelitis. Some children with local trauma may have a history of infection at a distant site. In adults, predisposing factors include advanced age, immunodeficiency, chronic bacteremia, I.V. drug abuse, long-term indwelling catheters, and sickle cell anemia.

Etiology

Most cases of hematogenous osteomyelitis are monomicrobial. Although *S. aureus* causes 60% to 90% of cases of hematogenous osteomyelitis, certain organisms tend to cause infections in certain age groups. In newborns, group B streptococci and gram-negative bacilli are common. In children, streptococci and *Haemophilus influenzae* are often seen. However, evidence from a retrospective study in Canada showed that vaccination of infants and children succeeded in eliminating *H. influenzae* type b as an infective agent in hematogenous osteomyelitis.[2] Polymicrobial hematogenous osteomyelitis is usually caused by *S. aureus* and a streptococcus.

The incidence of hematogenous osteomyelitis caused by gram-negative organisms is about 25%. *Escherichia coli* is the most common gram-negative organism isolated from hematogenous bone infections. *P. aeruginosa*, which has a predilection for the cervical vertebrae, and *Serratia* species are common pathogens in I.V. drug abusers. Furthermore, *P. aeruginosa* can be isolated from patients with long-term indwelling catheters.[3] *Salmonella* species are frequently associated with osteomyelitis in patients with sickle cell anemia[4] and may also be seen in patients with AIDS and other hemoglobinopathies.

Tuberculous infection of bone may be seen in patients from areas endemic for tuberculosis and in patients with AIDS. Fungal osteomyelitis, usually caused by opportunistic organisms, is found in patients who are severely immunodeficient or neutropenic or who have indwelling venous catheters. Because of their increasing frequency in all types of infection, *Candida* species may be seen more often in bone infections. Recent case reports reviewed 58 cases of candidal vertebral osteomyelitis[5] and 11 cases of deep sternal wounds caused by *C. albicans* after coronary artery bypass grafting.[6] Treatment of these infections is difficult, and therapy needs to be individualized against the specific organism.

Clinical Features

Children with acute hematogenous osteomyelitis usually pre-

sent with fever and localized pain. Leukocytosis may be present. A decreased range of motion and signs of local infection may be seen in the affected area. Drainage is not usually seen. Blood cultures are positive in more than half of patients.[3] Although in most children symptoms are present for 3 weeks or less, some children may present with vague symptoms of 1 to 3 months' duration.

Vertebral osteomyelitis is a form of hematogenous osteomyelitis most commonly seen in older persons and I.V. drug abusers. Most patients present with a constant dull pain that progresses slowly. Percussion over the affected area usually causes intense pain. Neurologic signs are generally absent but, when present, may indicate an epidural abscess. Fewer than half of patients are febrile. Vertebral infection typically involves the vertebral body rather than the spinous or transverse processes; often, two adjacent vertebrae and the disk space between them are affected. The lumbar region is most frequently involved in pyogenic hematogenous osteomyelitis. Thoracic vertebrae are often infected in spinal tuberculosis (Pott disease), and the cervical spine is often the site of infection in patients who abuse I.V. drugs.[7]

In about 10% of patients with acute hematogenous osteomyelitis, the disease progresses to a chronic condition. These patients have recurrent episodes of clinical exacerbations that are usually more indolent than those seen in the acute episode. Patients who are more susceptible to chronic osteomyelitis include those for whom therapy was delayed and those with compound fractures.

Despite the serious nature of hematogenous osteomyelitis, the prognosis is good for most patients. In a retrospective review of 69 children with acute hematogenous osteomyelitis, major sequelae were seen in only 3% of patients and minor problems in 2%. A favorable long-term outcome was associated with early hospitalization and initiation of appropriate antibiotic therapy.[8] Patients with severe underlying conditions have a worse prognosis.

Special Presentations of Hematogenous Osteomyelitis

Brodie abscess In patients with Brodie abscess, the infected portion of bone is completely replaced by pus and forms an intraosseous abscess. The infection, which is contained within a sclerotic membrane, may become quiescent. However, the risk of recurrence is high, and infection may spread. Patients are usually afebrile and may present only with local pain or swelling. Brodie abscess may be misdiagnosed as a bone tumor; the diagnosis of Brodie abscess is confirmed by histologic examination and culture. *S. aureus* is usually the etiologic agent, although *S. epidermidis* is occasionally implicated. Although seemingly benign, Brodie abscess should be aggressively treated. Pus may accumulate and spread through the tissues, causing a draining sinus; if near the joint surface of an intracapsular bone, infection can cause pyogenic arthritis.

Osteomyelitis in patients with sickle cell anemia There are two main differences between hematogenous osteomyelitis in children with sickle cell anemia and that in children who are otherwise healthy. First, infection in children with sickle cell anemia usually localizes in the diaphysis rather than the metaphysis. Second, *Salmonella* species frequently cause infection in this group, although staphylococci are also common.[4] Differentiating between osteomyelitis and thrombotic crisis in patients with sickle cell anemia may be difficult. In both situations, patients may present with bone pain and fever, and the erythrocyte sedimentation rate (ESR) and leukocyte count may be elevated. A bone biopsy is usually required for diagnosis of osteomyelitis.

Table 1 Common Clinical Associations in Osteomyelitis

Organism	*Common Clinical Associations*
Staphylococcus aureus	Found in 50%–70% of cases
Coagulase-negative staphylococci	Infections of prosthetic devices
Gram-negative bacilli	Decubitus ulcers, vascular insufficiency
Pseudomonas aeruginosa	Puncture wounds
Streptococci and anaerobes	Diabetic foot lesions, decubitus ulcers, bite wounds
Pasteurella multocida	Cat bites
Actinomyces	Periodontal infection or sinusitis
Eikenella corrodens	Mandibular osteomyelitis

Multiple sites of bone may be infected. Antibiotics for empirical therapy should be effective against both *Salmonella* and staphylococci. A regimen of a semisynthetic penicillin and an aminoglycoside is recommended [*see* Treatment Overview, *below*].

Hematogenous osteomyelitis in I.V. drug abusers In I.V. drug abusers, hematogenous osteomyelitis is associated with subtle clinical signs and symptoms. Patients may present with localized pain, but fever is usually absent. Although vertebral osteomyelitis is common, infection of the pubis and the clavicle is also seen. Culture of the infected site usually yields *S. aureus* or *S. epidermidis*, although *P. aeruginosa* is often seen. Serial radiographs may be necessary, and surgery may be required to confirm the microbiologic diagnosis.

OSTEOMYELITIS SECONDARY TO A CONTIGUOUS FOCUS OF INFECTION

Osteomyelitis in adults almost always derives from a contiguous source of infection. These bone infections are often complex and heterogeneous. Organisms may be inoculated directly into the bone after an open fracture, a penetrating wound, or a surgical procedure; or they may spread from an adjacent soft tissue infection. Because techniques for joint replacement have improved, the number of artificial joints inserted has increased and so has the number of infections associated with prosthetic joints.

Etiology

The cause of osteomyelitis depends, in part, on the route of entry, the phenotypic characteristics of organisms, and the epidemiologic background of the patient. Thus, the bacteriology of osteomyelitis in a diabetic patient who has had multiple hospital admissions and has been treated with multiple antibiotics over several years differs from that in a patient with a community-acquired infection. Organisms with common clinical associations are listed [*see Table 1*].

Because of virulence factors and adherence characteristics, *S. aureus* is the most common single pathogen in contiguous-focus osteomyelitis and accounts for 50% to 70% of cases.[9] In contrast to hematogenous osteomyelitis, contiguous-focus disease is often polymicrobial; multiple organisms may be isolated from 30% to 50% of such patients. The incidence of osteomyelitis caused by gram-negative bacilli is increasing. Gram-negative pathogens are important in patients who have undergone multiple hospital procedures, have had an extended hospital stay, are in intensive care units, or have had an open fracture. *P. aeruginosa* is often responsible for osteomyelitis associated with comminuted fractures and puncture wounds to the heel.[10] *S. epidermidis* is another common

pathogen in contiguous-focus osteomyelitis, particularly in patients who have infected orthopedic prostheses.

Osteomyelitis caused by anaerobic bacteria usually results from contiguous spread of a polymicrobial infection. Anaerobes should be suspected when osteomyelitis is associated with a human bite or is contiguous to a dental infection, intra-abdominal abscess, decubitus ulcer, or otorhinolaryngologic infection.

Clinical Features

In the initial stages of contiguous-focus osteomyelitis, patients may have pain, fever, swelling, and erythema. However, during recurrent episodes of chronic infection, fever subsides, and pain and drainage from a sinus tract or ulcer are often seen. In patients with generalized vascular insufficiency, the disease starts insidiously in an area of traumatized skin. The patient may present with an ingrown toenail, a perforating foot ulcer, cellulitis, or a superficial or deep wound infection. Fever and systemic signs of infection are not usually present.

Osteomyelitis after replacement of the hip joint may occur soon after surgery or later. Often evident within the first few days or weeks after surgery, acute contiguous infections result directly from infected skin, subcutaneous tissue, or muscle. Fever, pain, erythema, edema, and purulent drainage are often present when early infections are caused by pyogenic organisms such as *S. aureus*, streptococci, or enteric gram-negative bacilli. When early infections are caused by less pathogenic organisms, such as *S. epidermidis* or diphtheroids, the disease presents more insidiously. Chronic contiguous infections are usually diagnosed 6 to 24 months after surgery. The patient usually presents with persistent pain. Most infections are probably introduced during surgery but remain quiescent for a long time.

Sternal osteomyelitis, a serious complication of median sternotomy after cardiothoracic surgery, occurs in 1% to 5% of patients. Patients have fever, a slightly erythematous wound, and persistent pain. Frequent pathogens include *S. epidermidis*, *S. aureus*, gram-negative bacilli, and, more recently, *Candida* species. The presence of sternal wires, used to approximate the sternum, is a major risk factor for chronic recurrent osteomyelitis. Surgical removal of the wires is essential for recovery. The risk of death may be high in complicated cases.[10]

OSTEOMYELITIS ASSOCIATED WITH VASCULAR INSUFFICIENCY

Osteomyelitis secondary to vascular insufficiency occurs most frequently in older patients with diabetes mellitus or severe vascular impairment. In these patients, osteomyelitis usually develops by contiguous spread of infection from soft tissue to underlying bone; often, it occurs in the small bones of the feet in patients in whom soft tissue breaks down over weight-bearing or pressure-bearing areas. Bone infections develop in about 25% of diabetic patients with superficial mild to moderate foot infections; however, of those patients with serious foot infections, over 50% will have osteomyelitis.[11] Extensive debridement is necessary, and about two thirds of cases require bone resection or partial amputation.[12]

Complex foot lesions in diabetic patients result from a combination of neuropathy, atherosclerotic peripheral vascular disease, and repetitive trauma to the area. Atherosclerosis of the tibial and peroneal arteries of the lower leg is common in diabetic patients.[13] Limb ischemia, combined with poor collateral circulation, impairs wound healing in foot ulcers and allows for the contiguous spread of infection to bone. In addition, this anoxic environment contributes to the development of gangrenous changes and anaerobic infections. Furthermore, peripheral vascular disease may compromise the efficacy of antibiotic therapy by preventing the accumulation of adequate drug levels in the infected tissues.

Osteomyelitis should be considered in all diabetic patients with deep or chronic foot ulcers or infections. The Group Health Cooperative of Puget Sound conducted a study to determine the incidence of foot ulcers and the risk of developing serious complications among 8,905 patients with type 1 or type 2 diabetes. Over a period of 3 years, 5.8% developed a foot ulcer, and of those, 15% developed osteomyelitis.[14] Distinguishing between true bone infection and noninfectious destructive neuropathic bone changes in patients with diabetes is challenging and requires diagnostic imaging tests. Although *S. aureus* is the most common pathogen isolated from patients with osteomyelitis associated with vascular insufficiency, multiple organisms, including both anaerobes and aerobes, may be present, especially in hospitalized patients. Mixed gram-positive and gram-negative infections are often seen in patients with chronic or previously treated infections.

Pathogenesis

Pathogens gain access to bone in humans by three methods: direct inoculation, extension from contiguous sites, and hematogenous dissemination. Each type of infection may be further complicated by the presence of a prosthesis.

The sequelae associated with the hematogenous dissemination of bacteria to the bone vary with the age of the patient.[3] In infants, infection can spread via transphyseal vessels to the epiphysis and then to the joint space. Concurrent septic arthritis and epiphyseal destruction may be seen. In children older than 1 year, the growth plate is avascular, and osteomyelitis usually begins in the metaphyseal region of the long bones. Capillary loops do not anastomose in this region, and bacterial emboli can settle in this area and cause a microinfarction. Microscopic metaphyseal hemorrhage and necrosis, caused by insignificant trauma, provide a favorable environment for infection. Once infection is established, it can spread to the periosteum. If pus reaches the subperiosteal space, the periosteum becomes elevated. Osteoblasts beneath the detached periosteum gradually produce new bone.

The pathogenesis of hematogenous osteomyelitis in adults differs from that in children.[8] After growth ceases and the epiphyses close, organisms causing osteomyelitis no longer have a predilection for the metaphyseal area of long bones. The vertebrae, sternoclavicular and sacroiliac joints, and symphysis pubis are often affected. Infection may spread from the subchondral bone to the joint space. Sites of previous injury or trauma often provide the point of seeding of infectious organisms.

Unlike hematogenous osteomyelitis, contiguous-focus disease may invade any traumatized bone. Local inflammatory response or injury can devitalize bone and tissue. Microbes can multiply in these areas of dead bone. Osteomyelitis often complicates decubitus ulcers or cutaneous ulcerations in diabetic patients. Undetected trauma in diabetic patients with neuropathy provides a portal of entry for bacteria. Moreover, compromised blood flow impairs wound healing and allows unchecked bacterial proliferation.

Histologic findings in acute osteomyelitis include microbes, neutrophils, and congested or thrombosed blood vessels [*see Figure 1*]. The local inflammatory reaction contributes to further bone necrosis, which is a distinguishing feature of chronic osteomyelitis. As necrosis progresses, new bone gradually surrounds the infected area of dead bone (sequestrum). The forma-

tion of a functionally inert, nonresorbable substratum—either a sequestrum or a foreign body—upon which bacteria can attach and multiply marks the transition from acute to chronic disease.

Diagnosis

The diagnosis of osteomyelitis should be considered in patients with fever and localized skeletal pain, positive blood cultures, persistently draining wounds, and radiologic findings that suggest a localized inflammatory process. The diagnosis can then be confirmed by the histologic examination and bacteriologic culture of bone specimens obtained either by surgery or by multiple percutaneous needle biopsies. A detailed history with emphasis on previous trauma or surgery is important. Although no laboratory test specifically indicates the diagnosis of osteomyelitis, a complete blood cell count, ESR, and C-reactive protein test may be useful. The white blood cell count is usually elevated only in early stages of the disease. The ESR may be helpful in diagnosing acute hematogenous osteomyelitis in children but is less useful in newborns and children with sickle cell anemia. C-reactive protein is an acute-phase reactant, and although the C-reactive protein test is relatively insensitive and nonspecific, it may be more specific for infection than the ESR.

RADIOLOGIC AND NUCLEAR FINDINGS

There is a plethora of imaging techniques for diagnosing osteomyelitis [*see Table 2*]. The accuracy of imaging studies is affected by the extent of inflammation, the duration and site of infection, the vascularity of the site, and associated pathologic conditions. Osteomyelitis cannot be absolutely confirmed or excluded on the basis of results from any one test.

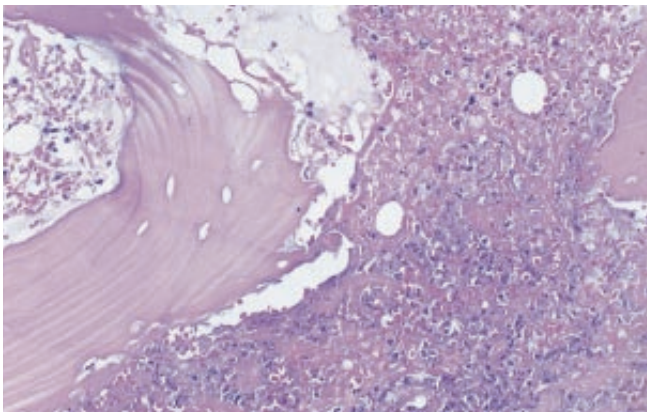
Radiograph

Radiographs, which are relatively inexpensive, should be the initial diagnostic technique for almost all patients suspected of having osteomyelitis. However, findings on conventional radiographs are often normal until at least 2 weeks after the onset of symptoms. Furthermore, radiographic changes can be subtle and nonspecific. In children with hematogenous osteomyelitis, the earliest changes are deep soft tissue swelling, which may be difficult to detect; periosteal elevation; and radiolucent areas of bone destruction, which are not usually detected until 50% to 75% of bone density has been lost.[15] In contiguous-focus disease, the initial radiologic manifestation is periosteal reaction. Radiographs may show the presence of sequestra, sclerosis, and significant bone lysis in patients with chronic disease.

Bone Scan

The technetium-99m (^{99m}Tc) methylene diphosphonate three-phase bone scan is often used to identify osteomyelitis, because it is more sensitive than conventional radiography and can confirm a suspected diagnosis of osteomyelitis within 48 hours after onset of symptoms. The bone scan should be the next imaging study performed in patients with normal radiographs. Classic findings on bone scan are increased blood flow, pooling, and reactive new bone formation. Although sensitivity is good, specificity is poor [*see Table 2*]. Cellulitis and septic arthritis cannot be differentiated from osteomyelitis on bone scan. False positive results may be seen in patients with previous bone injury, tumor, or infarction. Furthermore, bone scans may remain positive for 6 weeks to 6 months after therapy because of bone metabolism and remodeling. In children with uncomplicated hematogenous osteomyelitis, the bone scan has a high sensitivity and specificity; a positive bone scan strongly suggests osteomyelitis, and a negative scan in children older than 3 years excludes the diagnosis.[16] Sensitivity, however, is lower in neonates and in patients with sickle cell anemia.

a



b

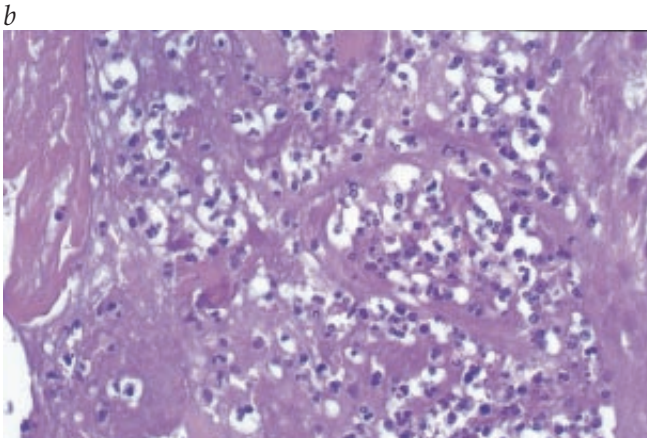


Figure 1 **(*a*) Photomicrograph of a specimen from bone biopsy shows acute osteomyelitis (magnification: × 40). Note missing osteoclastic cells (signifying bone death) and intense inflammatory reaction adjacent to bone fragment. (*b*) Higher magnification (× 100) shows intense polymorphonuclear infiltration at same site.**

Gallium Scan

Gallium scans may be useful in infants and in children in whom bone scans are negative. The low specificity of gallium imaging may be improved if it is combined with bone scans. When interpreted in combination with a bone scan, the gallium scan can in some cases be used to differentiate acute cellulitis from acute osteomyelitis. Disadvantages associated with gallium scans include higher doses of radiation and a 1- to 2-day waiting period for results.

Leukocyte Scan

Indium-111 (^{111}In)–labeled and ^{99m}Tc-labeled leukocyte scans, which are more specific but less sensitive than bone scans, have been used to evaluate patients with osteomyelitis. Leukocyte scans are not useful in patients with chronic, indolent infections. In a patient who is suspected of having osteomyelitis and has recently had a fracture, a leukocyte scan may be the study of choice because results from a bone scan will be equivocal. The cumbersome and expensive technique of leukocyte scanning re-

Table 2 Radiographic and Radionuclide Techniques for Diagnosing Osteomyelitis

Test	*Sensitivity*	*Specificity*	*Indications/Comments*
Radiograph	Good after 3 weeks	Poor	Initial test for most patients
^{99m}Tc bone scan	Good in first 2 weeks	Poor	Use in patients with suspected osteomyelitis who have a negative result on radiograph
Gallium	Moderate after 2 weeks	Poor	Use in infants and children with a negative result on bone scan
Indium	Moderate	Moderate	May be used in patients with a recent fracture; may be useful in diagnosis of infected joint prosthesis 2–4 mo after implantation
CT	Excellent	Moderate	Useful in osteomyelitis of the skull, pelvis, and sternoclavicular junction
MRI	Excellent	Excellent	Reliably distinguishes tumor or infarction from osteomyelitis; useful in surgical planning for diabetes; good for vertebral disease; very expensive
PET	Excellent	Moderate	May be useful in chronic osteomyelitis in patients with negative results on other tests

CT—computed tomography MRI—magnetic resonance imaging PET—positron emission tomography ^{99m}Tc—technetium-99m

quires in vitro cell isolation and radiolabeling, which is time-consuming and potentially biohazardous. However, leukocyte scans, especially in combination with bone scans, may be sensitive and specific in detecting osteomyelitis in diabetic patients.[17-19] In a study of 52 diabetic patients with chronic foot ulcers, both specificity and accuracy were significantly higher with leukocyte scans than with bone scans.[17] The combined use of leukocyte scans and bone scans has effectively differentiated osteomyelitis from soft tissue infections in patients with diabetes; the sensitivity of the combination has ranged from 93% to 100%, and the specificity has ranged from 80% to 83%. ^{111}In-labeled and ^{99m}Tc-labeled leukocyte scanning have been shown to be useful in the diagnosis of prosthetic joints 2 to 6 months after implantation. Biopsy confirmation is necessary for diagnosis.

Computed Tomography and Magnetic Resonance Imaging

Both CT and MRI have excellent sensitivity and resolution and allow simultaneous evaluation of bone and surrounding soft tissues. CT reliably detects sequestra and devitalized bone and is useful for evaluating patients with osteomyelitis of the skull. However, CT is prone to image degradation and is less specific than other tests. MRI, with its excellent specificity, is helpful in distinguishing bone tumor or infarction from osteomyelitis.[20] Furthermore, MRI is particularly reliable in distinguishing normal areas from abnormal areas when surgery is being planned for diabetic patients with osteomyelitis.[21] In addition, MRI is the technique of choice for detecting and assessing the site and extent of infection in the spine. However, the expense of MRI precludes its use on a routine basis. A limited MRI scan with specialized views and minimal use of gadolinium contrast can be used to control costs.[22] Metallic prostheses exclude the use of MRI and distort image reflection on CT, thereby obscuring early changes of infection.

Positron Emission Tomography

Positron emission tomography (PET) has been used as an imaging tool for infections of the bone and osteosynthetic material.[23] PET with fluorodeoxyglucose (FDG) was shown to be superior to bone scans in diagnosing chronic osteomyelitis of the central skeleton.[24] Because of its excellent sensitivity, FDG PET may be an effective technique for excluding chronic osteomyelitis when a bone scan is negative.[25]

CULTURES

The gold standard for diagnosis of osteomyelitis is aerobic and anaerobic culture of a biopsy specimen obtained under direct vision during surgery. Alternatively, culture of multiple specimens obtained by needle biopsy under ultrasound or radiographic guidance has been a reliable, cost-effective means of diagnosing osteomyelitis at our center. Combining surgical debridement with the obtaining of culture specimens is also cost-effective. Blood cultures should also be obtained, especially from patients suspected of having hematogenous osteomyelitis. Results of swab cultures of the sinus tract are usually inaccurate. Studies have shown that the results of cultures of sinus tract and bone biopsy specimens match in only 44% to 57% of patients with chronic osteomyelitis.[26,27]

Treatment Overview

Optimal management of the patient with osteomyelitis entails microbiologic confirmation of the diagnosis by biopsy and complete surgical debridement, followed by institution of antibiotic therapy that is based on culture and sensitivity results [*see Table 3*]. When debridement is inadequate, therapy often fails, despite appropriate antimicrobial therapy; therefore, "a subtle balance between medical and surgical therapy is necessary if a potentially curative outcome is to be achieved."[28]

SURGICAL THERAPY

The type of surgical therapy depends on the extent of infection. Surgical management of acute infection requires debridement of dead tissue, whereas chronic osteomyelitis requires surgical debridement of all devitalized bone and soft tissue and removal of foreign bodies. Adequate surgical management is essential for successful treatment. Inadequate surgical debridement, regardless of antibiotic therapy, is the most common cause of treatment failure. Sequestered dead bone can serve as a nidus for persistent infection. In some cases, a two-staged debridement protocol may be necessary to sterilize the wound.[29]

Surgical therapy entails filling the dead space created by debridement and reestablishing blood supply to the poorly perfused area. Skin, bone, or muscle grafts are used to cover the

Table 3 Choices for Antibiotic Treatment of Bacterial Osteomyelitis

Organism	*Antibiotic*	*Dosage*	
		Children	*Adults*
Staphylococcus aureus			
Methicillin-sensitive	Nafcillin or oxacillin	150 mg/kg I.V., q. 4 hr	2 g I.V., q. 6 hr
	Cefazolin	—	2 g I.V., q. 8 hr
	Cephalexin	100 mg/kg p.o., q. 6 hr	—
Methicillin-resistant or *S. epidermidis*	Vancomycin*	40 mg/kg I.V., q. 6 hr	1 g I.V., q. 12 hr
Streptococci	Ampicillin	150 mg/kg I.V., q. 6 hr	2 g I.V., q. 6 hr
	Ceftriaxone	50 mg/kg I.V., q. 24 hr	2 g I.V., q. 24 hr
Enterococci	Ampicillin + gentamicin†	—	2 g I.V., q. 6 hr 1 mg/kg q. 8 hr
Enterobacteriaceae	Ciprofloxacin	—	750 mg p.o., q. 12 hr
	Ceftriaxone	50 mg/kg I.V., q. 24 hr	2 g I.V., q. 24 hr
Pseudomonas aeruginosa	Ceftazidime	—	2 g I.V., q. 12 hr
	Ciprofloxacin	—	750 mg p.o., q. 12 hr
	Semisynthetic penicillin‡		
	Piperacillin	—	3 g I.V., q. 6 hr
	with		
	Aminoglycoside‡		
	Amikacin	—	15 mg/kg q.d.
	Gentamicin	—	5 mg/kg q.d.
Mixed infections (diabetic foot)	Ticarcillin-clavulanate§	—	3.1 g I.V., q. 6 hr
	Clindamycin + ciprofloxacin	—	0.9 g I.V., q. 8 hr 750 mg p.o., q. 12 hr
	Imipenem-cilastatin‖	—	1 g I.V., q. 6 hr

*Adding rifampin may increase efficacy.
† Use combination for the first 2 wk, followed by ampicillin alone.
‡ Use a semisynthetic penicillin in combination with an aminoglycoside.
§ Use in limb-threatening infections.
‖ Use in life-threatening infections.

wound. Cancellous bone grafts can be used to fill the dead space. Revascularization procedures include the use of local pedicle muscle flaps and myocutaneous flaps. In patients with very large bone defects, the Ilizarov technique has been used to reconstruct large defects in bone and soft tissue. In this procedure, major segmental resections of infected bone are performed, followed by realignment of opposing ends of now noninfected bone so that new bone growth fills in the defect. Each salvage procedure has associated morbidity and expense.

ANTIBIOTIC THERAPY

Oral and I.V. Antibiotic Therapy

The standard treatment for chronic osteomyelitis is I.V. antimicrobial therapy targeting the causative organism. Because of rising hospital costs, alternatives to an extended hospital stay for parenteral treatment of osteomyelitis in otherwise healthy patients are being sought. The economic benefits of oral therapy are obvious: decreased hospital stay, reduced pharmacy and supply costs, no need for surgical insertion of a catheter, and decreased catheter-related complications. For patients who require a complete regimen of I.V. antibiotics, outpatient therapy should be considered. Patients should be educated in the care and use of the catheter while they are still in the hospital. Furthermore, drug toxicity and clinical response should be closely monitored by the physician. Administration of I.V. antibiotics during clinic visits or by a home health agency can substantially reduce costs over in-hospital care.

Alternative Delivery of Antibiotics

Antibiotic-impregnated acrylic beads have been used for local treatment of bone infections; however, they have not been approved for use in the United States. Although the beads deliver a high concentration of antibiotic to the area of infection, bactericidal levels of antibiotic are present locally for only 2 to 4 weeks. Thus, the beads should be used in conjunction with systemic antibiotics. Both biodegradable and nonbiodegradable beads have been used, and both require surgical placement. Nonbiodegradable beads, such as the polymethylmethacrylate beads, must be surgically removed after 2 to 4 weeks. Biodegradable antibiotic beads do not require surgical removal and provide local bactericidal concentrations for extended periods.[30,31]

Choice of Antimicrobial Agents

Staphylococcal infections Methicillin-resistant *S. aureus* infections represent nearly 60% of nosocomial *S. aureus* isolates detected in hospitals and reported to the CDC.[32] For infections caused by susceptible *S. aureus*, a β-lactamase–resistant semisynthetic penicillin is the drug of choice because most strains of *S. aureus* are resistant to penicillin. At our center, only 45% of *S. aureus* organisms are methicillin susceptible, but more than 70% of coagulase-negative staphylococci are resistant to methicillin. The

problem of methicillin-resistant *S. aureus* is worsening, especially in tertiary medical centers. Although vancomycin is the preferred agent for treatment of methicillin-resistant staphylococcal infections, its use as monotherapy has come into question. At our center, the use of vancomycin alone to treat osteomyelitis caused by methicillin-resistant *S. aureus* has failed in the standard 4- to 6-week treatment regimen, and rifampin (300 mg p.o. twice daily) is now routinely added to the regimen, unless there is a contraindication to its use. Similar regimens are used to treat methicillin-resistant *S. epidermidis* infections, especially sternal bone infections after median sternotomy. Unfortunately, because failure of vancomycin alone has been observed with increasing frequency, objective clinical trials to determine efficacy are not likely to be performed, for ethical reasons.

In selected cases, after 2 weeks of parenteral therapy with vancomycin, patients can be switched to oral therapy if the appropriate requirements are met. Those requirements include good response to I.V. therapy, strict patient compliance, microbiologic diagnosis, isolated pathogen highly susceptible to the proposed antibiotic, and a high level of bioavailability of the oral agent (serum bactericidal titers may be useful).

Although oral therapy with ciprofloxacin has been effective in treating staphylococcal osteomyelitis,[33,34] reports of variable efficacy and the development of resistance indicate that ciprofloxacin should not be used as monotherapy in osteomyelitis caused by staphylococci.[35] However, long-term treatment with the combination of ciprofloxacin and rifampin has been successful for implant-related staphylococcal infections in patients who have stable implants and symptoms of short duration.[36] A recent study of 17 diabetic patients with mild to moderate foot lesions associated with 20 osteomyelitic bones received rifampin plus ofloxacin for 6 months; 88.2% of the patients were cured, and cure was maintained by 76.5% at the end of a 22-month follow-up.[37] Quinolones are contraindicated in children and pregnant women.

Enterococcal infections *Enterococcus faecalis* is an increasingly troublesome pathogen, especially in patients with infected prostheses and patients with diabetes who have osteomyelitis in the diabetic foot. Fortunately, *E. faecalis* does not usually cause osteomyelitis of long bones in the absence of prostheses. Because of the serious nature of a possible associated bacteremia in acute infections, aggressive therapy is necessary until acute symptoms resolve and blood cultures become negative.

E. faecalis is becoming a problem in postoperative orthopedic patients and in diabetic patients. The isolation of vancomycin-resistant *Enterococcus* species in tertiary medical centers has increased over the past few years. In a recent study, 89 patients with osteomyelitis were treated with linezolid and were evaluated for clinical efficacy, safety, and tolerability.

The clinical cure rate of 22 evaluable patients was 81.8%. The authors concluded that linezolid (administered either intravenously or by mouth) was successful in treating osteomyelitis caused by resistant gram-positive organisms.[38] However, hematologic complications have been found to be a limiting factor in long-term use of linezolid.[39] More experience is necessary before the ultimate use of linezolid in these patients is determined.

Gram-negative infections Treatment of gram-negative osteomyelitis requires an accurate identification of the organism and antimicrobial susceptibility. Oral ciprofloxacin is an effective and inexpensive choice for treating bone infections caused by Enterobacteriaceae.[22,33,34] Furthermore, studies have shown that ciprofloxacin and ofloxacin have been effective in treating osteomyelitis caused by multiresistant gram-negative bacteria.[40] Because of reports of the development of resistance, however, patients should be carefully monitored when either agent is used as monotherapy in pseudomonal osteomyelitis. Ceftazidime and cefepime are other potent antipseudomonal agents, although neutropenia has been associated with cefepime.[41] In cases of resistant strains of *Pseudomonas*, a regimen of an extended-spectrum penicillin and an aminoglycoside may be used; however, efficacy is not definitive in cases of osteomyelitis, and the toxicity of an extended regimen of I.V. aminoglycosides is a serious concern.

Treatment of Clinical Syndromes

HEMATOGENOUS OSTEOMYELITIS

In children with acute hematogenous osteomyelitis, antibiotic therapy alone may be sufficient. When the disease is recognized early and treated promptly, cure can be anticipated. Surgery should be considered for patients who do not respond to antibiotic therapy. For compliant children with a documented microbiologic infection who respond initially to parenteral therapy and who have no complications, the course of therapy may be completed with oral agents.[42] Cure rates of 95% have been obtained with this regimen. The suggested duration of I.V. therapy ranges from 4 to 14 days, and oral therapy is continued for 14 to 26 days.[43] Although some investigators have suggested that monitoring blood concentrations of antibiotics may be unnecessary,[42] most agree that adequate gastrointestinal absorption of the antibiotic should be confirmed by repeated measurements of serum bactericidal activity.[44] Close outpatient follow-up is essential.

In adults with hematogenous osteomyelitis, surgical debridement and drainage of soft tissue abscesses are often necessary. Antibiotic therapy is usually maintained for 4 to 6 weeks. I.V. administration is used most often; however, oral ciprofloxacin is used to treat gram-negative infections.

OSTEOMYELITIS SECONDARY TO A CONTIGUOUS FOCUS OF INFECTION

The diagnosis of osteomyelitis secondary to a contiguous focus of infection is based on a biopsy of the bone and not of the contiguous lesion. Once bone infection is confirmed by biopsy, treatment should proceed according to general principles of osteomyelitis management [*see* Treatment Overview, *above*].

OSTEOMYELITIS ASSOCIATED WITH VASCULAR INSUFFICIENCY

Prompt, aggressive surgical and antibiotic therapy and proper wound care are the best means of preserving a functional limb and optimizing outcome in patients with osteomyelitis associated with vascular insufficiency. Hyperbaric oxygen therapy, by increasing oxygen levels in ischemic tissues, has improved wound healing and reduced the amputation rate in high-risk patients. Revascularization of lesions in the lower-extremity vasculature by angioplasty and bypass grafting may help oxygenate tissue with critical ischemia, thereby preventing limb amputations. Prognosis for wound healing improves when the tissue oxygen level is greater than 40 mm Hg and when the popliteal pulse is palpable.[11]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Waldvogel FA, Medoff G, Swartz MM: Osteomyelitis: a review of clinical features, therapeutic considerations, and unusual aspects. N Engl J Med 282:198, 1970

2. Howard AW, Viskontas D, Sabhagh C: Reductions in osteomyelitis and septic arthritis related to *Haemophilus influenzae* type B vaccination. J Pediatr Orthop 19:705, 1999

3. Lew DP, Waldvogel FA: Osteomyelitis. N Engl J Med 336:999, 1997

4. Sadat-Ali M: The status of acute osteomyelitis in sickle cell disease: a 15-year review. Int Surg 83:84, 1998

5. Miller DJ, Mejicano GC: Vertebral osteomyelitis due to *Candida* species: case report and literature review. Clin Infect Dis 333:523, 2001

6. Malani PM, McNeil SA, Bradley SF, et al: *Candida albicans* sternal wound infections: a chronic and recurrent complication of median sternotomy. Clin Infect Dis 35:1316, 2002

7. Endress C, Guyot DR, Fata J, et al: Cervical osteomyelitis due to I.V. heroin use: radiologic findings in 14 patients. Am J Roentgenol 155:333, 1990

8. Christiansen P, Frederiksen B, Glazowski J, et al: Epidemiologic, bacteriologic, and long-term follow-up data of children with acute hematogenous osteomyelitis and septic arthritis: a ten-year review. J Pediatr Orthop B 8:302, 1999

9. Gentry LO: Osteomyelitis and other infections of bones and joints. The Staphylococci in Human Disease. Crossley KB, Archer GL, Eds. Churchill Livingstone, New York, 1997, p 455

10. Laughlin TJ, Armstrong DG, Caporusso J, et al: Soft tissue and bone infections from puncture wounds in children. West J Med 166:126, 1997

11. Lipsky BA: Osteomyelitis of the foot in diabetic patients. Clin Infect Dis 25:1318, 1997

12. Ramsey SD, Newton K, Blough D, et al: Incidence, outcomes, and cost of foot ulcers in patients with diabetes. Diabetes Care 22:382, 1999

13. Hill SL, Holtzman GI, Buse R: The effects of peripheral vascular disease with osteomyelitis in the diabetic foot. Am J Surg 177:282, 1999

14. Ramsey SD, Newton K, Blough D, et al: Incidence, outcomes, and cost of foot ulcers in patients with diabetes. Diab Care 22:382, 1999

15. Jaramillo D, Treves ST, Kasser JR, et al: Osteomyelitis and septic arthritis in children: appropriate use of imaging to guide treatment. AJR Am J Roentgenol 165:399, 1995

16. Sutter CW, Shelton DK: Three-phase bone scan in osteomyelitis and other musculoskeletal disorders. Am Fam Physician 54:1639, 1996

17. Harvey J, Cohen MM: Technetium-99-labeled leukocytes in diagnosing diabetic osteomyelitis in the foot. J Foot Ankle Surg 36:209, 1997

18. Johnson JE, Kennedy EJ, Shereff MJ, et al: Prospective study of bone, indium-111 labeled white blood cell, and gallium-67 scanning for the evaluation of osteomyelitis in the diabetic foot. Foot Ankle Int 17:16, 1996

19. Crerand S, Dolan M, Laing P, et al: Diagnosis of osteomyelitis in neuropathic foot ulcers. J Bone Joint Surg Br 78:51, 1996

20. Umans H, Haramati N, Flusser G: The diagnostic role of gadolinium enhanced MRI in distinguishing between acute medullary bone infarct and osteomyelitis. Magn Reson Imaging 18:255, 2000

21. Sammak B, Abd El Bagi M, Al Shahed M, et al: Osteomyelitis: a review of currently used imaging techniques. Eur Radiol 9:894, 2000

22. Gentry LO, Clint B: Adult osteomyelitis. Current Practice of Medicine. Bone RC, Ed. Churchill Livingstone, New York, 1996, p 15.6

23. Robiller FC, Stumpe KD, Kossmann T, et al: Chronic osteomyelitis of the femur: value of PET imaging. Eur Radiol 10:855, 2000

24. Guhlmann A, Brecht-Krauss D, Suger G, et al: Fluorine-18-FDG PET and technetium-99m antigranulocyte antibody scintigraphy in chronic osteomyelitis. J Nucl Med 39:2145, 1998

25. Zhuang H, Duarte PS, Pourdehand M, et al: Exclusion of chronic osteomyelitis with F-18 fluorodeoxyglucose positron emission tomographic imaging. Clin Nucl Med 25:281, 2000

26. Patzakis MJ, Wilkins J, Kumar J, et al: Comparison of the results of bacterial cultures from multiple sites in chronic osteomyelitis of long bones: a prospective study. J Bone Joint Surg Am 76:664, 1994

27. Mackowiak PA, Jones SR, Smith JW: Diagnostic value of sinus-tract cultures in chronic osteomyelitis. JAMA 239:2772, 1978

28. Snyder RJ, Cohen MM, Sun C, et al: Osteomyelitis in the diabetic patient: diagnosis and treatment: part 2. Medical, surgical, and other alternatives. Ostomy Wound Manage 47:24, 2001

29. Patzakis MJ, Greene N, Holtom P, et al: Culture results in open wound treatment with muscle transfer for tibial osteomyelitis. Clin Orthop 360:66, 1999

30. Calhoun JH, Mader JT: Treatment of osteomyelitis with a biodegradable antibiotic implant. Clin Orthop 341:206, 1997

31. Kanellakopoulou K, Giamarellos-Bourboulis EJ: Carrier systems for the local delivery of antibiotics in bone infections. Drugs 59:1223, 2000

32. Kaye K, Lynfield R, Kreiswirth B: Methicillin-resistant *Staphylococcus aureus*. Paper presented at the International Conference on Emerging Infectious Diseases, February 29 to March 3, 2004, Atlanta, Georgia

33. Gentry LO, Rodriguez GG: Oral ciprofloxacin compared with parenteral antibiotics in the treatment of osteomyelitis. Antimicrob Agents Chemother 34:40, 1990

34. Lew DP, Waldvogel FA: Use of quinolones in osteomyelitis and infected orthopaedic prosthesis. Drugs 58(suppl 2):85, 1999

35. Trucksis M, Hooper DC, Wolfson JS: Emerging resistance to fluoroquinolones in staphylococci: an alert editorial. Ann Intern Med 114:424, 1991

36. Zimmerli W, Widmer AF, Blatter M, et al: Role of rifampin for treatment of orthopedic implant-related staphylococcal infections: a randomized controlled trial. Foreign-Body Infection (FBI) Study Group. JAMA 279:1537, 1998

37. Senneville E, Yazdanpanah Y, Cazubiel M, et al: Rifampin-ofloxacin oral regimen for the treatment of mild to moderate diabetic foot osteomyelitis. J Antimicrob Chemother 48:927, 2001

38. Rayner CR, Baddour LM, Birmingham MC, et al: Linezolid in the treatment of osteomyelitis: results of compassionate use experience. Infection 32:8, 2004

39. Kuter DJ, Tillotson GS: Hematologic effects of antimicrobials: focus on the oxazolidinone linezolid. Pharmacotherapy 21:1010, 2001

40. Galanakis N, Giamarellou H, Moussas T, et al: Chronic osteomyelitis caused by multi-resistant gram-negative bacteria: evaluation of treatment with newer quinolones after prolonged follow-up. J Antimicrob Chemother 39:241, 1997

41. Wong BB, Ko GJ: Neutropenia in patients receiving long-term cefepime therapy for osteomyelitis. Am J Health Syst Pharm 60:2229, 2003

42. Tetzlaff TR, McCracken GH Jr, Nelson JD: Oral antibiotic therapy for skeletal infection of children: II. Therapy of osteomyelitis and suppurative arthritis. J Pediatr 92:485, 1978

43. Peltola H, Unkila-Kallio L, Kallio MJ: Simplified treatment of acute staphylococcal osteomyelitis of childhood. Finnish Study Group. Pediatrics 99:846, 1997

44. Nelson JD: Toward simple but safe management of osteomyelitis. Pediatrics 99:883, 1997

XVII INFECTIONS DUE TO RICKETTSIA, EHRLICHIA, AND COXIELLA

DANIEL J. SEXTON, M.D., F.A.C.P.

Taxonomy

Like other gram-negative bacteria, *Rickettsia* and *Ehrlichia* are members of the alpha group of purple bacteria. In the past, the family Rickettsiaceae was divided into three tribes: Ehrlichiae, Rickettsiae, and Wolbachia. The tribe Ehrlichiae was divided into three genera: *Ehrlichia*, *Cowdria*, and *Neorickettsia*. A new taxonomy of existing organisms has been proposed [*see Figure 1*] that is based on genetic analysis and divides the order Rickettsiales into two families, Rickettsiaceae and Anaplasmataceae; however, this taxonomy is still controversial and is likely to change in the future. The family Rickettsiaceae contains the genera *Orientia* and *Rickettsia*. The genus *Rickettsia* is divided into the typhus and spotted-fever groups, which in turn contain numerous species. For example, over 15 different spotted-fever-group rickettsial species have been named. The family Anaplasmataceae is divided into a complex number of *Anaplasma* and *Ehrlichia* genera. Some organisms previously considered to be *Ehrlichia* have been renamed *Anaplasma*.

Although the term ehrlichiosis is generally used by clinicians to describe tick-borne diseases caused by organisms in the genera *Ehrlichia* and *Anaplasma*, some organisms that were previously considered to be (and are still widely considered to be) *Ehrlichia* have been assigned to the genus *Anaplasma*. For example, the agent of human granulocytic ehrlichiosis (HGE) has been renamed *A. phagocytophilum* (some sources list this as *A. phagocytophilia* or *A. phagocytophila*). Although this chapter uses the old term, HGE, some authors now call this disease human granulocytic anaplasmosis. Moreover, some authors now use the term anaplasmosis instead of ehrlichiosis to describe infection by agents in the family Anaplasmataceae. To make taxonomic matters even more complicated, a surprising number of new spotted-fever-group rickettsial species have been discovered over the past 10 years. Additionally, previously recognized spotted-fever-group *Rickettsia* species (e.g., *R. parkeri*) that were thought not to cause human disease were recently confirmed to be human pathogens.[1]

Biology and Ecology

Rickettsia are gram-negative bacteria that are difficult to see in tissue without the use of special stains. *Rickettsia* can usually be visualized using Gimenez or special immunoperoxidase stains or can be detected in tissue by the use of direct fluorescent antibody staining techniques.

Rickettsia organisms cannot be propagated on cell-free media, but they can be grown in tissue cultures or in the yolk sacs of developing chick embryos. Members of the spotted-fever group, such as *R. rickettsii*, grow in both the nucleus and the cytoplasm of host cells, whereas typhus-group *Rickettsia* organisms grow only in the cytoplasm of infected host cells. Rickettsial cell walls contain lipopolysaccharides and outer-membrane proteins (rOmps). These rOmps are capable of eliciting a protective immune response in animals. Their exact function in arthropods and mammalian cells is uncertain.

Ehrlichia and *Anaplasma* organisms may infect monocytes (in human monocytic ehrlichiosis [HME]) or granulocytes (in HGE).

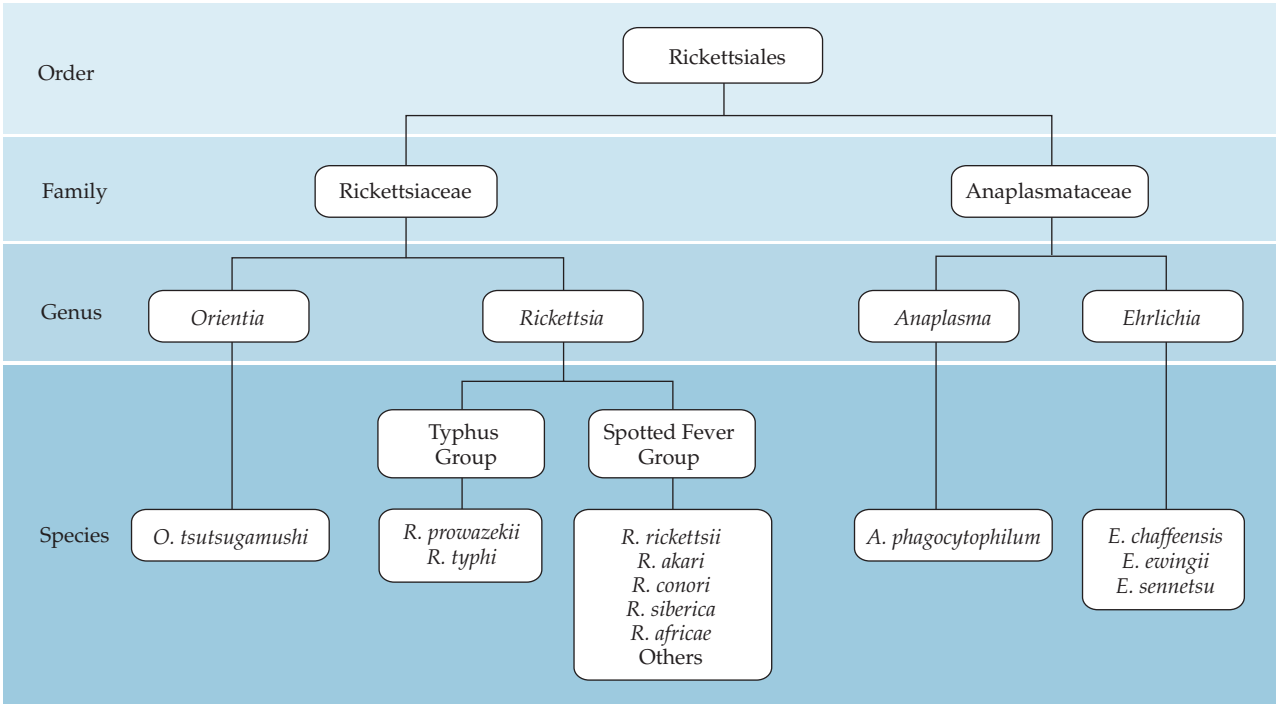


Figure 1 **Taxonomy of the rickettsias.**

Ehrlichia organisms replicate in the phagosomes of the host cell. Like *Rickettsia* organisms, *Ehrlichia* organisms have distinct ribosomes and are surrounded by an outer and inner membrane. *Ehrlichia* organisms can be detected by light microscopy when they exist as single organisms and can be stained with Gram, Giemsa, Wright, or silver stains. *Ehrlichia* organisms cannot be grown on cell-free media and thus far have been cultured only in eukaryotic cells and in yolk sacs.

Most *Rickettsia* and *Ehrlichia* organisms cycle between arthropod vectors and small mammals, infecting humans only as an incidental host. A notable exception to this pattern is *R. prowazekii*, which is primarily maintained in a human-louse cycle.

Pathophysiology

After inoculation into humans, *Rickettsia* organisms proliferate intracellularly and then spread throughout the body via the bloodstream or lymphatics. Spotted-fever-group *Rickettsia* species, such as *R. rickettsii*, have a specific and characteristic tropism for endothelial cells. After attachment to the host endothelial cell plasma membrane via a cholesterol-containing receptor, these organisms enter the cell via pinocytosis and then escape from phagosomes into the cytosol. *R. rickettsii* divides by binary fission and then spreads cell to cell via filopodia derived from host cell membranes.

The hallmark of most rickettsial diseases is cell injury or death associated with an intense vasculitis. Widespread *Rickettsia*-induced vasculitis in humans leads to minute focal areas of hemorrhage and edema (as a result of increased vascular permeability), along with activation of the humoral immune response, inflammation, and coagulation. Vasculitis from rickettsial infection may result in hypotension, shock, widespread organ dysfunction, and death. Although the histopathologic and clinical features of rickettsial infection have been well characterized, the precise cellular mechanism by which most *Rickettsia* organisms produce cell damage to small blood vessels is unknown.

Individual *Ehrlichia* and *Anaplasma* species infect hosts ranging from humans to dogs, sheep, and horses. Four species are known to infect humans: *E. chaffeensis* (the organism responsible for human monocytic ehrlichiosis), *A. phagocytophilum* (the agent of HGE), *E. sennetsu*, and *E. ewingii*.

Ehrlichia organisms enter host cells by phagocytosis and then grow and divide while in the host cell phagosome. After a few days, tightly packed clusters of organisms are observable as intracellular inclusions. Later, additional growth and replication of *Ehrlichia* organisms result in large, mature inclusions called morulae. In vitro growth of *Ehrlichia* and *Anaplasma* in tissue culture or yolk cell systems takes 8 to 36 days. Only a few isolates of *E. chaffeensis* and *A. phagocytophilum* have been cultured from humans, because in vitro cultivation of these organisms from clinical specimens is extremely difficult.

There is no serologic cross-reactivity between *Ehrlichia* and *Rickettsia*, although some cross-reactivity exists between *E. chaffeensis, E. ewingii*, and *A. phagocytophilum*. In animals, both experimentally induced and naturally acquired *Ehrlichia* infections may result in antibody responses, but these responses may not be protective. The relationship between humoral antibody responses and immunity in human ehrlichiosis is still not understood.

How *Ehrlichia* produces human disease is also poorly understood. *Ehrlichia* does not cause vasculitis, nor do human tissues infected with *Ehrlichia* demonstrate cell necrosis, abscess formation, or a severe inflammatory response. Patients infected with *Ehrlichia* may develop lymphadenopathy and perivascular lymphohistiocytic infiltrates without apparent endothelial injury or thrombosis. Some patients with *A. phagocytophilum* infection contract secondary opportunistic fungal infections. There is also no understanding of the mechanism for the thrombocytopenia that occurs, predictably and early, in most forms of ehrlichiosis. Studies in animal models have shown that decreased hematopoietic production, immune-mediated platelet destruction, and splenic sequestration are unlikely to be the primary mechanism for this thrombocytopenia.[2]

Rocky Mountain Spotted Fever

EPIDEMIOLOGY

Rocky Mountain spotted fever (RMSF) is an acute tick-borne illness caused by *R. rickettsii*. RMSF occurs widely in the Western Hemisphere, from southern Canada to Mexico. RMSF also occurs in Central America and has been described in Colombia and Brazil. The primary vectors of RMSF are the Rocky Mountain wood tick in the western United States and the American dog tick in the eastern United States. Other tick vectors are responsible for transmission in Mexico, Central America, and South America. Ticks represent both the reservoir and the vector of RMSF. *R. rickettsii* normally cycles between ticks and small mammals such as moles and rodents; humans are only an incidental host.

Although RMSF occurs sporadically in most of the continental United States in rural, suburban, and urban locations, it is most often reported from the southeastern and south central United States. Cases of RMSF have been reported from Central Park in New York City and from suburban areas such as Long Island, New York. Over 90% of all cases of RMSF occur from early spring to early autumn.

Up to one third of patients with proven RMSF cannot recall a history of tick bite or recent tick contact. Rickettsial transmission in such cases presumably occurs through inapparent or unrecognized tick bite or through contact with infected tick tissues.

DIAGNOSIS

Clinical Manifestations

RMSF typically begins abruptly. The incubation period after tick bite or tick exposure ranges from 2 to 12 days. The early phases of illness are usually marked by nonspecific symptoms such as fever, headache, malaise and myalgias, and nausea and vomiting. Some patients, especially children, have prominent abdominal pain; this pain may be severe enough to mimic an acute surgical abdomen. Rarely, gastrointestinal involvement in the early phases of RMSF leads to an erroneous diagnosis, such as acute appendicitis, cholecystitis, or even bowel obstruction. Hepatic involvement in severe cases can result in jaundice, which may be striking.

The skin rash that gives RMSF its name may begin as a macular or maculopapular eruption and evolve into a generalized petechial rash [*see Figure 2*]. Generalized ecchymosis and even gangrene of the digits, genitals, ears, and nose may rarely occur in severe cases.[3] The rash typically develops on the third to the fifth day of illness. Appearance of the rash may be delayed, however; and in a small percentage of patients, the rash does not develop at all. Delay or absence of the rash greatly complicates clinical diagnosis. In one study, only 14% of RMSF patients had a rash on

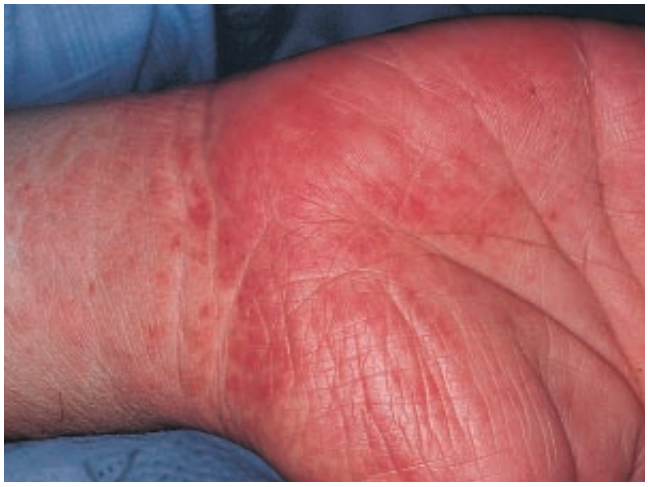


Figure 2 **Purpuric macules on the palm and wrist of a patient with Rocky Mountain spotted fever.**

the first day of illness and fewer than 50% developed a rash in the first 72 hours of illness.[4] Absence of rash does not correspond to milder disease; a small percentage of patients with so-called spotless RMSF have fatal illness.[5] Moreover, the characteristic skin rash of RMSF may be overlooked in dark-skinned patients.

Patients with RMSF may occasionally present with a wide variety of neurologic symptoms ranging from confusion to seizures and encephalopathy. Some patients with RMSF have focal neurologic signs that lead to misdiagnoses. Neurologic symptoms may persist after recovery of acute illness in some patients.[6] Other atypical and uncommon presenting features of RMSF include periorbital edema (especially in children) and cough.

The diagnosis of RMSF is notoriously difficult, even for experienced physicians in highly endemic areas. In the first few days of illness, particularly when the characteristic rash is absent, RMSF may mimic an array of nonspecific viral diseases. If early RMSF is mistaken for a bacterial illness and if β-lactams or other antibiotics ineffective against *R. rickettsii* are given, the subsequent appearance of a rash may be tragically mistaken for a drug eruption and lifesaving therapy may be omitted or delayed. RMSF has been confused with measles, staphylococcal bacteremia, hepatitis, leptospirosis, meningococcemia, and infectious mononucleosis.[7] The clinical features of RMSF and ehrlichiosis are often similar [*see Table 1*]. Even experienced physicians cannot always distinguish the two diseases, although the presence of severe leukopenia and the absence of rash favor the diagnosis of ehrlichiosis.[8]

It is axiomatic that the diagnosis of RMSF must be based on the clinical features and an appropriate epidemiologic setting rather than on any single laboratory test. There is no completely reliable diagnostic test for RMSF in the early phases of illness; thus, therapy should always begin before laboratory confirmation is obtained.

Laboratory Studies

The white blood cell count remains normal in most patients with RMSF. In some cases, however, the WBC may be low or elevated. Thrombocytopenia generally occurs in severe cases and is a helpful diagnostic finding. The low platelet count may be accompanied by reduced fibrinogen concentrations and elevated levels of fibrin split products. Other characteristic laboratory abnormalities in patients with RMSF include elevated blood levels of aminotransferases, bilirubin, and creatinine.

Diagnostic proof of RMSF can be obtained by direct immunofluorescent or immunoenzyme staining of skin biopsy samples or, in the convalescent phase of the disease, by the detection of characteristic antibodies. These antibodies typically do not appear before day 8 to 10 of illness; thus, serologic testing has no role in the initial diagnosis of acutely ill patients with suspected RMSF.

Culture of *Rickettsia* from blood samples can be done only in specialized laboratories and the delay in obtaining a result makes positive cultures useful only in retrospect. Polymerase chain reaction technology has been applied to the diagnosis of RMSF, but such testing is neither sensitive nor widely available. Because some patients die before generating an antibody increase against *R. rickettsii* antigen, the diagnosis in fatal cases may be completely missed unless immunohistochemical staining is performed on tissues obtained at autopsy.[9]

TREATMENT

Tetracyclines and chloramphenicol are the only antibiotics known to be effective against *R. rickettsii*. Doxycycline is the preferred agent in all patients except pregnant women, for whom chloramphenicol remains the agent of choice. Therapy should be initiated as early as possible in the course of RMSF. Therapy started more than 5 days after onset is much less likely to be effective than therapy given earlier.[10]

The usual dosage of doxycycline is 200 mg a day in two divided doses. The same dosage can be administered intravenously or orally. Intravenous therapy is preferred for patients who are seriously ill or who are experiencing nausea or vomiting. Children with RMSF should be given doxycycline at a dosage of 2.5 to 3 mg/kg/day in two divided doses. Chloramphenicol given at an initial loading dose of 50 mg/kg, followed by 50 mg/kg/day in four divided doses, is an appropriate alternative for patients who cannot tolerate tetracyclines or who are pregnant [*see Table 2*].

Unfounded fears about the risk of tetracycline therapy in children with suspected rickettsial disease may unnecessarily delay the institution of potentially lifesaving therapy.[11] Because RMSF can normally be cured with a short course of doxycycline therapy, concerns about staining of the teeth should not prohibit the use of doxycycline in young children with suspected RMSF; indeed, such therapy may be lifesaving, and it is considerably safer than therapy with chloramphenicol. In addition to antimicrobial therapy, treatment should include measures to correct associated complications, such as hypotension, heart failure, and electrolyte disturbances. Glucocorticoid therapy has never been shown in a controlled study to be useful in the treatment of RMSF, and it is not recommended.

PREVENTION

The prevention of RMSF relies on avoidance of tick-infested areas or the use of protective clothing and tick repellent while in such areas. Prompt removal of ticks after attachment is particularly important. Ticks should be removed using tweezers, tissue, or cloth to protect the fingers, because crushed tick tissues can be infectious.

PROGNOSIS

In consequence of the difficulty in making a conclusive diagnosis of RMSF in the early phases of illness, delays in the initiation of therapy are common. Such delays are often associated

Table 1 Diagnosis of Diseases Caused by *Rickettsia*, *Ehrlichia*, and *Coxiella* Species

Disease (Pathogen)	*Rocky Mountain Spotted Fever* (Rickettsia rickettsii)	*Rickettsialpox* (Rickettsia akari)	*TIBOLA*	*Murine Typhus* (Rickettsia typhi)	*Epidemic Typhus* (Rickettsia prowazekii)	*Scrub Typhus* (Orientia tsutsugamushi)	*Ehrlichial Diseases*	*Q Fever* (Coxiella burnetii)
Animal Reservoir	Rocky Mountain wood tick (western U.S.), American dog tick (eastern U.S.)	House mouse via bloodsucking mite	Tick (*Dermacentor marginatus*)	Rat, cat, or mouse flea	Human body louse (squirrel lice and fleas in sylvatic form)	Mite (Pacific rim)	White-tailed deer, rodents	Cattle, goat, sheep, cats, dogs, ticks
Transmission	Tick bite	Mite bite	Tick bite	Flea bite	Louse or flea bite	Mite bite	Tick bite	Airborne, direct contact, infected milk
Incubation Period (Range)	2–12 days	2–12 days	3–10 days	8–16 days	7–10 days	7–10 days	3–12 days	29 days (range, 10–39 days)
Clinical Manifestations	Fever, headache, malaise, myalgias, nausea/vomiting; abdominal pain; rash on days 3–5 in most but not all cases; jaundice	Eschar at bite site, then severe headache, fever, chills, myalgias, papulovesicular rash	Localized lesion at bite site; cervical or regional lymphadenopathy; fever and rash in < 10% of cases; alopecia may occur and persist at site of local lesion	Fever, chills, headache, myalgias, GI symptoms; rash at end of first week beginning on trunk, spreading to extremities; cough, dyspnea in severe illness	Acute severe infection, fever, headache, myalgias, cough, abdominal pain, GI symptoms, jaundice; red macular/maculopapular rash on trunk spreading to extremities; neurologic symptoms may occur	Necrotic eschar at bite site; fever, headache, anorexia, malaise, chills; macular/maculopapular rash in half of patients, usually sparing the face; generalized lymphadenopathy; respiratory symptoms; neurologic symptoms may occur	Headache, fever, anorexia, malaise, cough, nausea and vomiting; rash uncommon	Fever, headache, myalgias, anorexia, malaise, chills; chest pain, cough, rales, pneumonia; hepatitis; chronic disease in immunocompromised patients; may cause endocarditis
Laboratory Test Results	Thrombocytopenia, elevated plasma aminotransferase, elevated bilirubin, elevated serum creatinine, positive serology, positive PCR; culture difficult	Thrombocytopenia, mild elevations in plasma aminotransferases; positive serology	Thrombocytopenia, positive serology, positive PCR	Thrombocytopenia; positive serology	Elevated aminotransferases, thrombocytopenia; positive serology	Serology not reliable in early phases; antibodies present 10–20 days after onset	Serology not reliable in early phases, antibodies present 10–20 days after onset; positive PCR; culture difficult	Elevated antibody titer, positive PCR

PCR—polymerase chain reaction TIBOLA—tick-bite–associated lymphadenopathy

with fatal outcome if they exceed 5 days. A number of host factors have been associated with severe or fatal RMSF, including increasing age, male gender, and the presence of glucose-6-phospate dehydrogenase (G6PD) deficiency. Black race and alcohol use have also been associated with more severe disease and a higher fatality, but it is difficult to exclude the role of delay in seeking or receiving antimicrobial therapy in these patients. The biologic causes for the worse outcomes seen in patients older than 40 years and in male patients remain poorly understood. Patients with abnormal renal function at the time of presentation have a worse prognosis than those with normal renal function.[12]

Rickettsialpox

Rickettsialpox is an uncommon mite-borne rickettsial disease caused by *R. akari*. The disease's name is derived from its clinical resemblance to chickenpox.

EPIDEMIOLOGY

R. akari is transmitted to the house mouse by a blood-sucking mite (*Liponyssoides sanguineus*). The mouse serves as a reservoir for the disease, and as with RMSF, humans are only an incidental host. When mouse populations are reduced by vermin-eradication programs, *L. sanguineus* will bite humans and transmit rickettsialpox. Initially, rickettsialpox was recognized only in urban areas such as New York City, Pittsburgh, Cleveland, and Boston. However, cases have now been described in rural areas and in areas outside the previously recognized geographic range for the disease, including North Carolina, the Ukraine, Croatia, Turkey, and the Yucatan region of Mexico.[13]

DIAGNOSIS

Rickettsialpox is usually recognized by the appearance of one or more characteristic skin lesions (eschars) at the site of the bite, followed soon thereafter by the abrupt onset of headache, fever, chills, and myalgias and a papulovesicular

Table 2 Treatment of Diseases Caused by *Rickettsia*, *Ehrlichia*, and *Coxiella* Species

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Cost*[*]	*Comments*
Rocky Mountain spotted fever, rickettsialpox, murine typhus, infections due to other spotted-fever-group *Rickettsia* species	Doxycycline	100 mg q. 12 hr I.V. or p.o for 5–7 days	First choice	Oral, $6–8	I.V. preferred for adult and pediatric patients who are seriously ill or who have nausea or vomiting
	Chloramphenicol	50 mg/kg initial loading dose, then 50 mg/kg/day in 4 divided doses	—	—	For pregnant women and those who cannot tolerate tetracycline
Epidemic typhus	Doxycycline	Single dose of 200 mg; repeat weekly for prophylaxis	First choice	$1	—
	Chloramphenicol	50 mg/kg initial loading dose, then 50/mg/kg day in four divided doses for 5–7 days	—	—	For pregnant women and those who cannot tolerate tetracycline
Scrub typhus	Doxycycline plus rifampin	Doxycycline, 100 mg q. 12 hr I.V. or p.o. for 3–7 days; rifampin, 600 mg/day for 5–7 days	First choice	Oral doxycycline, $3–8; rifampin, $20–28	This regimen may not be effective in geographic areas where strains with reduced susceptibility to tetracyclines have been documented (see text for details)
	Azithromycin or rifampin	Azithromycin, 500 mg/day for 5–7 days; rifampin, 600 mg/day for 5–7 days	—	Azithromycin, $75–105; rifampin, $20–28	For use in areas with strains with reduced susceptibility to tetracycline; for pregnant women
Ehrlichia, Anaplasma phagocytophilia infection	Doxycycline	100 mg q. 12 hr I.V. or p.o. for 5–7 days	First choice	Oral, $6–8	—
	Chloramphenicol	50 mg/kg initial loading dose, then 50/mg/kg day in four divided doses for 5–7 days	—	—	Efficacy unproven; for pregnant women and those who cannot tolerate tetracyclines
Q fever (acute disease)[†]	Tetracycline	500 mg, p.o., q. 6 hr for 14 days	Equally effective	$15	—
	Doxycycline	100 mg, p.o., q. 12 hr for 14 days	Equally effective	$15	—
	Ciprofloxacin	500 mg, p.o., q. 12 hr for 14 days	Equally effective	$131	—

[*]Rounded to nearest dollar. Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.
[†]For recommendations on treatment of chronic Q fever, see text.

rash. Most patients with rickettsialpox are only mildly ill and recover without treatment after 10 to 14 days of illness. Headache is characteristically severe. The skin rash of rickettsialpox can be distinguished from that of chickenpox by the fact that it typically begins as a maculopapular eruption that quickly evolves into a papulovesicular eruption. Most patients with vesicular lesions also have a primary eschar at the site of their mite bite. Lesions in rickettsialpox are often scant, characteristically scab, and resolve without scarring. Unlike the rash of chickenpox, the rash of rickettsialpox does not appear in crops or clusters, nor do the vesicular lesions of rickettsialpox have the characteristic dew-drop appearance of chickenpox lesions. Certain other rickettsial diseases that produce eschars and vesicular rashes can mimic rickettsialpox [*see* Other Spotted-Fever-Group Rickettsial Infections, *below*].

The diagnosis of rickettsialpox is typically confirmed by serologic testing. However, biopsy of the eschar or the rash can be diagnostic if special fluorescent antibody reagents are available to demonstrate *R. akari* in tissue.

TREATMENT

Treatment for rickettsialpox is identical to that for RMSF (see above).

Other Spotted-Fever-Group Rickettsial Infections

Other tick-borne spotted-fever-group rickettsial infections include Mediterranean spotted fever, African tick-bite fever, Queensland tick typhus, Siberian tick typhus, Japanese spotted fever, Flinders Island spotted fever, tick-borne lymphadenopathy (TIBOLA), and *R. parkeri* infection. All of these illnesses respond to treatment with tetracyclines or chloramphenicol. Most are milder in course and severity than RMSF, although Mediterranean spotted fever may, in rare cases, be fatal.

In addition to rickettsialpox, six spotted-fever-group rickettsioses are characterized by localized eschars (Mediterranean spotted fever, African tick-bite fever, *R. parkeri* infection, Queensland tick typhus, and Siberian tick typhus). TIBOLA is not usually associated with a generalized rash; rather, this unusual rick-

ettsiosis is characterized by a localized area of scalp alopecia. The alopecia typically follows the development of a crusted lesion at the site of a resolving tick bite. Localized hair loss may be prolonged. Cervical or regional lymphadenopathy is common in TIBOLA. Only a minority of TIBOLA patients have fever, and rash occurs in fewer than 10% of patients.[14]

Infection with a newly recognized rickettsial agent, *R. mongolotimonae*, has been seen in southern France. In the few cases reported to date, clinical manifestations have typically been similar to those of Mediterranean spotted fever.[15]

A recent report of human infection with *R. parkeri* suggested that some cases of apparent RMSF in the United States, especially those in which an eschar was present at the site of a tick bite, may in fact be infection with the spotted-fever-group agent *R. parkeri*.[16]

Murine Typhus

EPIDEMIOLOGY, ETIOLOGY, AND PATHOGENESIS

Murine typhus occurs worldwide. It is caused by *Rickettsia typhi,* which is transmitted to humans by the bite of an infected flea. Although house mice, domestic cats, and shrews may be occasional hosts for infected fleas, the principal reservoir of *R. typhi* is the rat. In addition to serving as hosts for infected fleas, rats may become rickettsemic after contact with such fleas and then transmit organisms to other, simultaneously feeding fleas. When such fleas happen to bite humans, murine typhus may result. Murine typhus is uncommon in the United States, although a relatively large number of cases have been reported in southern Texas during the past 20 years. A total of 47 cases of murine typhus were reported in Hawaii during 2003, which is a remarkable increase from previous numbers, and serologic evidence of *R. typhi* infection has been demonstrated in a variety of rodent species in Hawaii.[17]

DIAGNOSIS

Clinical Manifestations

Onset of murine typhus is typically abrupt, occurring after an incubation period ranging from 8 to 16 days. Early manifestations of illness are nonspecific, typically comprising fever, chills, headache, and myalgias. In addition, many patients have gastrointestinal complaints, including nausea, vomiting, abdominal pain, and diarrhea. Rash typically occurs near the end of the first week of illness. The rash is classically described as beginning on the trunk and spreading to the extremities. Rash may be faint or difficult to see, and in some patients, it does not occur at all. Murine typhus is usually a mild illness; even in untreated cases, deaths are rare. However, severe disease with mental confusion, encephalopathy, renal dysfunction, and severe thrombocytopenia can occur. Patients with severe illness may develop cough and dyspnea. As in other rickettsial diseases, the presence of G6PD deficiency and advanced age appear to be associated with more severe or fatal disease.

Laboratory Studies

Thrombocytopenia is present in most cases of murine typhus and is an important laboratory clue to the diagnosis. As in other rickettsial diseases, serology is the most reliable and widely available diagnostic method. Convalescent antibodies typically appear after 8 to 10 days of illness. In severely ill patients with pulmonary manifestations, chest x-rays may show interstitial infiltrates or changes suggestive of pneumonia or pulmonary edema.

TREATMENT

As in other rickettsial diseases, the only effective agents for treatment of murine typhus are tetracyclines and chloramphenicol, with doxycycline the preferred choice in most patients. The dosages of doxycycline in the treatment of murine typhus are the same as those in the treatment of RMSF [*see Table 2*].

Epidemic (Louse-Borne) Typhus

EPIDEMIOLOGY, ETIOLOGY, AND PATHOGENESIS

Louse-borne, or epidemic, typhus, caused by *R. prowazekii*, killed millions of people in Eastern Europe and Russia during the periodic wars that devastated this geographic area during the past 2 centuries. Fortunately, epidemic typhus is now a rare disease. However, sporadic outbreaks still occur in parts of Russia, Algeria, and Burundi and in the Andes Mountains of South America.[18] War, famine, and human cruelty can result in epidemics; for example, more than 45,000 cases of epidemic typhus occurred in Burundi after the 1993 civil war in that country.[19]

Until 25 years ago, epidemic typhus was thought to involve an exclusive cycle between the body louse and infected humans. However, a sylvatic cycle of infection with *R. prowazekii* is now known to be endemic in the eastern United States. This sylvatic cycle of infection involves flying squirrels (*Glaucomys volans*) and their ectoparasites, but secondary transmission in humans has been recognized when lice that normally infest squirrels seek a host in humans.[20] A case report of epidemic typhus in a patient from the southwestern United States (an area outside the known geographic range of the flying squirrel) suggests the existence of additional vectors or wild-animal reservoirs for *R. prowazekii*.[21]

The principal vector for epidemic typhus is the human body louse (*Pediculus humanus corporis*), although occasionally, the head louse (*P. humanus capitis*) can also transmit infection [*see 2:VIII Parasitic Infestations*]. Both the squirrel flea (*Orchopeas howardi*) and the squirrel louse (*Neohaematopinus sciuropteri*) also act as vectors for *R. prowazekii* in the sylvatic cycle of infection. Infected fleas and lice are capable of producing a rickettsemia in flying squirrels and thus allowing propagation of infection to additional generations of ectoparasites.

In its classic epidemic form, typhus is transmitted by body lice. While taking a blood meal from humans, body lice defecate and regurgitate infective gastrointestinal contents; these highly infective substances are then inoculated into the skin when the person scratches the pruritic feeding site. Lice feces remain infectious for as long as 100 days. Thus, human-to-human transmission of *R. prowazekii* can occur via the sharing of clothes or via transfer of infected lice feces from one human to another.

Transmission of the sylvatic form of epidemic typhus to humans occurs only when humans have direct contact with infected squirrels or when squirrels nesting in the attics of homes are removed or killed, leaving the lice that infested their nests to seek an alternative (human) host.

PATHOPHYSIOLOGY

After entering the human body, *R. prowazekii* spreads via the bloodstream and lymphatics to produce a generalized vasculitis. The precise mechanism by which *R. prowazekii* produces cellular injury is poorly understood, but the net effect of this infection is

widespread endothelial injury. This rickettsial vasculitis may produce diffuse myocarditis, along with macroscopic and microscopic damage to muscles and neural tissue, the spleen, kidneys, and other internal organs. Central nervous system involvement may result in so-called typhus nodules, which comprise perivascular infiltrates consisting of lymphocytes, macrophages, and plasma cells.

DIAGNOSIS

Clinical Manifestations

R. prowazekii infection produces two distinct clinical syndromes. The most common is an acute, severe infection that occurs 7 to 10 days after exposure to infected lice and may result in death. A second, recrudescent form, called Brill-Zinsser disease, may occur from 1 to 5 decades after a primary infection.

Patients with acute epidemic typhus infection typically become abruptly ill with fever, headache, and myalgias. Other nonspecific symptoms, such as cough, abdominal pain, nausea, and diarrhea, are common. Skin rash in patients with epidemic typhus classically begins several days after the onset of symptoms as a red macular or maculopapular eruption on the trunk that then spreads to the extremities. Although skin rash is classically described as sparing the palms and soles, exceptions to this rule occur. In severe cases of epidemic typhus, skin rash may become petechial; rarely, gangrene of the extremities has been described.

In humans, the clinical features of the sylvatic form of typhus are similar to those of the epidemic form. In one series, however, only half of the patients with the sylvatic form of *R. prowazekii* infection had a skin rash.

Patients with epidemic typhus often have neurologic symptoms that may range from mild confusion and drowsiness to coma, seizures, and focal neurologic findings. As in other rickettsial diseases, jaundice and myocarditis may occur in severe cases.

Recrudescent *R. prowazekii* infection (Brill-Zinsser disease) is generally a much milder illness than acute epidemic typhus. The onset of Brill-Zinsser disease is often abrupt, with chills, fever, and headache. Skin rash typically begins 4 to 6 days after the onset of symptoms, and it is often scant or evanescent. Because patients with Brill-Zinsser disease are often elderly or have other chronic medical conditions, their symptoms (e.g., confusion, dyspnea, and lethargy) may be incorrectly attributed to preexisting or coexisting cardiac, cerebrovascular, or pulmonary disease.

Laboratory Studies

Both forms of *R. prowazekii* infection are best diagnosed by serologic testing. The indirect immunofluorescent antibody (IFA) test and an immunoblot technique are reliable serologic methods. A fourfold rise in antibody titers 10 to 21 days after onset of symptoms is considered diagnostic in both forms of *R. prowazekii* infection, although only an IgG antibody response is elicited in Brill-Zinsser disease.

Common accompanying laboratory findings in acute epidemic typhus include increased plasma aminotransferase levels and thrombocytopenia. Patients with severe involvement may have pulmonary infiltrates and other laboratory evidence of myocarditis.

TREATMENT

Tetracyclines and chloramphenicol are effective treatments for both acute and recrudescent *R. prowazekii* infections, but the drug of choice is doxycycline. In one study, 35 of 37 patients with epidemic typhus were cured with a single 200 mg dose of doxycycline.[22] However, the two remaining patients had a relapse 6 and 7 days after initial response. Patients with *R. prowazekii* infection characteristically improve within 48 hours of initiation of antirickettsial therapy. In severe cases, supportive care with vasopressors, intravenous fluids, oxygen, and even dialysis may be necessary.

PREVENTION

Efficacy of prophylaxis for epidemic typhus is directly related to the efficacy of prophylactic measures against lice infestation. Because humans who have contacts with other lice-infested humans can secondarily acquire lice even if they have good hygiene, all louse-infested persons and workers in close contact with such persons should use long-acting topical insecticides such as malathion, lindane, or a pyrethroid (permethrin). Fabrics and clothing treated with permethrin remain toxic to lice even after 20 washings.[23]

In epidemic settings, the use of chloramphenicol or tetracycline for prophylaxis of *R. prowazekii* infection is highly effective. Even a single 200 mg dose of doxycycline taken once weekly by travelers or health care workers in areas where epidemic typhus is present has been shown to protect against infection. Prophylaxis should be continued for 1 week after leaving an endemic or epidemic area.[24]

Inactivated vaccines have been shown to confer protection against experimental *R. prowazekii* infections.[24] Such vaccines are not currently commercially available nor are they likely to become available, in view of the effectiveness of both prophylactic antibiotics and other preventive measures.

PROGNOSIS

In the preantibiotic era, higher mortality from epidemic typhus was associated with older age and male gender. With prompt institution of antimicrobial therapy, however, mortality is now rare.

Scrub Typhus

Scrub typhus is a mite-borne disease, caused by *Orientia tsutsugamushi*, that was first described by Chinese physicians in the third century. Widespread outbreaks of scrub typhus in Allied soldiers in the Pacific theater of World War II led to investigations by military physicians that provided much of our current understanding of the clinical features, diagnosis, and prevention of this highly geographically focal disease.

EPIDEMIOLOGY

O. tsutsugamushi is distributed throughout the Pacific rim and is endemic in Korea, China, Taiwan, Japan, Pakistan, India, Thailand, Malaysia, and northern Australia. Most cases of scrub typhus occur in rural areas, but cases may also occur in suburban areas, such as those around Bangkok, where the seroprevalence in the general population may be as high as 20%.[25]

The reservoir and vector of scrub typhus are trombiculid mites of the genus *Leptotrombidium*. Larval mites (known as chiggers) maintain the infection in successive generations via transovarial transmission. Because the geographic distribution of mites is often highly focal, areas as small as a few hundred square meters of scrub vegetation may contain enormous numbers of infected chiggers. The risk of disease transmission by these chiggers may be extremely high in humans who enter these focal ar-

eas (also called mite islands). Because of the ease of air travel and the long incubation period of scrub typhus (up to 2 weeks), tourists to endemic areas may fall ill after returning home to regions where the illness is not familiar to physicians. Numerous cases of scrub typhus have been described in tourists returning to the United States, Europe, and Canada from endemic regions.

BIOLOGY AND PATHOPHYSIOLOGY

O. tsutsugamushi is distinct serologically and epidemiologically from typhus-group *Rickettsia*. However, this organism shares many of the microbiologic characteristics of typhus and spotted-fever-group *Rickettsia*, such as inability to be propagated on cell-free media. Unlike other *Rickettsia* species, *O. tsutsugamushi* has a trilaminar outer membrane. *O. tsutsugamushi* is also unique in that it is released from infected cells by a peculiar budding process involving the plasma membrane of host cells. Once released from the infected cell, organisms are in turn phagocytosed by adjacent cells and eventually disseminate widely throughout the body.

There are three strains of *O. tsutsugamushi*: Karp, Gilliam, and Kato. Infection with one strain does not preclude infection with a different strain.

DIAGNOSIS

Clinical Manifestations

Onset of scrub typhus typically occurs 7 to 10 days after the person is bitten by an infected mite. The illness may begin gradually or abruptly. In either case, headache, anorexia, malaise, chills, and fever are prominent early symptoms. Approximately one half of patients with scrub typhus develop a characteristic macular or maculopapular rash. In severe cases, this rash may become petechial. Rash from scrub typhus typically spares the face.

A localized necrotic skin lesion or eschar is a hallmark of scrub typhus [*see Figure 3*]. Eschars typically occur at the site of the infected chigger bite and may appear before the onset of systemic symptoms. Some patients have multiple eschars. In various case series, eschars have occurred in as few as 40% and as many as 88% of patients with scrub typhus.[26]

Generalized lymphadenopathy occurs in most patients with scrub typhus; some patients also have splenomegaly. Respiratory involvement occurs more often in scrub typhus than in other rickettsial diseases, and cough may be present in as many as one half of patients. As in other rickettsial diseases producing vasculitis, scrub typhus may be marked by neurologic involvement, which may manifest as aseptic meningitis, seizures, or focal neurologic abnormalities. In some patients, neurologic abnormalities are the dominant symptoms.

DIAGNOSIS

The initial diagnosis of scrub typhus must be based on clinical and epidemiologic features, because serologic testing is not reliably positive in the early phases of illness. Convalescent antibodies, which are best detected by an IFA technique, occur in the majority of patients between 10 and 20 days after onset of illness. Because more than one strain of *O. tsutsugamushi* is capable of causing scrub typhus, a battery of antigens should be used to detect convalescent antibodies. In addition to IFA testing, an enzyme-linked immunosorbent assay (ELISA) and a dot-blot immunoassay have been developed to diagnose scrub typhus.

Although not commonly done, biopsy of the generalized rash or the eschar can help establish the diagnosis of scrub typhus.

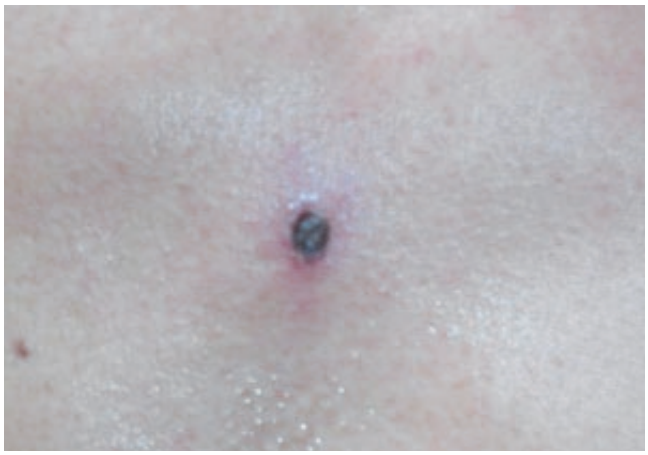


Figure 3 **Eschar at the site of a mite bite in a patient with scrub typhus.**

Examination of the biopsy sample will reveal vasculitis with a perivascular collection of lymphocytes and macrophages.

Culture of *O. tsutsugamushi* can be performed in specialized centers with the necessary laboratory facilities and diagnostic reagents; a PCR-based test has also been developed. However, such tests are rarely available in regions of the world where scrub typhus is endemic.

DIFFERENTIAL DIAGNOSIS

Scrub typhus is one of the classic causes of tropical fever in the Pacific Rim, where it is well known to be easily confused with malaria, dengue, or typhoid fever. Coinfection with scrub typhus and leptospirosis has been described in agricultural workers in Thailand. One of these workers, who was treated with penicillin only, died of respiratory failure attributed to untreated scrub typhus infection.[27]

TREATMENT

The treatment of scrub typhus is the same as that of most other rickettsial diseases [*see Table 2*]. Doxycycline is typically the agent of choice. Regimens of doxycycline as short as 1 day have been advocated for the therapy of scrub typhus, but a small number of patients treated with this short course suffer relapse.[28,29] To prevent relapse, most experts recommend a 3- to 7-day course of therapy with doxycycline.

In the mid-1990s, strains of *O. tsutsugamushi* with reduced susceptibility to tetracycline were reported in Thailand.[30] Azithromycin may be effective against strains with reduced susceptibility to tetracycline and, therefore, may be selected for therapy in areas where such strains are known or suspected to exist. However, there is little clinical experience, as well as few in vitro susceptibility data, to support this choice. Azithromycin may also be used to treat pregnant woman infected with typhus.

A randomized trial in northern Thailand compared the efficacy of doxycycline alone with the combination of doxycycline and rifampin for scrub typhus. The median duration of fever was significantly shorter in patients treated with the combined regimen.[31]

PREVENTION

Although there is no vaccine available to prevent the transmission of scrub typhus, several studies have demonstrated that doxycycline is highly effective for prophylaxis when used by nonimmune persons living and working in areas where scrub

typhus is highly endemic. A weekly dose of 200 mg of doxycycline has been shown to be reasonably effective in preventing transmission.

Human Monocytic Ehrlichiosis

EPIDEMIOLOGY

The principal vector of HME is the Lone Star tick (*Amblyomma americanum*) The causative agent of HME, *E. chaffeensis*, was first isolated from a soldier at Fort Chaffee, Arkansas, in 1990. Since then, HME has been recognized as endemic throughout the southeastern and south central United States. In addition, a few cases of HME have been recognized in New England and the Pacific Northwest, and isolated cases have been reported in Europe, Africa, and Mexico.[32]

In some locations, HME is probably more common than RMSF. In a prospective study of 35 consecutive patients from North Carolina who presented to outpatient facilities with fever and a recent history of tick bite, 26% had HME, 17% had RMSF, and one patient was coinfected with *R. rickettsii* and *E. chaffeensis*. Most of these patients received outpatient treatment with doxycycline, and all of them recovered quickly.[33]

Estimates of the incidence of HME have varied widely. A prospective study of febrile patients hospitalized in southeast Georgia in the United States showed that the prevalence of HME was 5.7 cases per 100,000 population.[34] A remarkably higher incidence was described in a golf-oriented retirement community in Tennessee, in which the annual incidence of disease was 660 per 100,000.[35] A serosurvey in the same community revealed that 12.5% of the residents had serologic evidence of past infection. Tick bites, exposure to wildlife, and golfing were associated with an increased risk of infection.

White-tailed deer are thought to be the principal animal reservoirs for *E. chaffeensis* infection. A study from Georgia found serologic evidence of *E. chaffeensis* infection in 27 of 35 deer and isolated *E. chaffeensis* from five of them, confirming that deer are naturally infected in endemic areas.[36] Although the role of wild canids and domestic dogs in the epidemiology of HME is uncertain, in one study 15 of 21 coyotes trapped in Oklahoma were found to be infected with *E. chaffeensis*.[37]

DIAGNOSIS

As with other tick-borne diseases, the diagnosis of HME is usually based on the recognition and synthesis of characteristic clinical and laboratory features in patients who reside in a geographic location where ehrlichiosis is known to occur during a time of year when tick exposure is likely or known. In other words, a history of tick bite or tick exposure during the spring or summer months in a resident of an endemic area, coupled with the presence of leukopenia or thrombocytopenia and abnormal liver function test results, provides strong circumstantial evidence for the diagnosis of HME. Even the presence of some of these findings is sufficient justification for the initiation of therapy.

Clinical Manifestations

After an incubation period of 5 to 14 days, patients with HME typically develop fever along with malaise and headache. Chills occur in approximately two thirds of patients, and gastrointestinal symptoms, including nausea and vomiting, may occur in up to one half of patients. Cough may also be a prominent symptom of HME, leading to diagnostic confusion with a host of respiratory illnesses. Although HME usually presents as an acute illness, subacute infection from *E. chaffeensis* has been described. In one study, for example, six of 41 patients with HME diagnosed at a medical center in Missouri during a 4-year period had protracted fever, ranging in duration from 17 to 51 days.[38] In addition, rare cases of subclinical or self-limiting infection with *E. chaffeensis* have been described.

Skin rash is uncommon in patients with HME, but when present, it may be macular, maculopapular, or petechial. Although skin rash was reported in 36% of cases in one case series of 211 patients with HME, skin rash has been less common in the experience of many clinicians working in HME-endemic regions.[39] The clinical features of HME are highly variable. Some patients present only with headache, anorexia, and malaise, whereas other patients have prominent neurologic symptoms that may include mental-status changes and stiff neck.

Laboratory Studies

The most common laboratory abnormalities seen in patients with HME are leukopenia (often accompanied by a left shift), thrombocytopenia, and elevated levels of aminotransferases (transaminases), lactate dehydrogenase, and alkaline phosphatase. Anemia and an elevated plasma creatinine concentration also may be seen. Later in the course of illness or during recovery, a striking atypical lymphocytosis may occur.

Abnormalities of the cerebrospinal fluid are common in patients in whom a lumbar puncture is performed because of neurologic symptoms. Lymphocytic pleocytosis and elevated CSF protein levels were found in 21 of 38 patients with *E. chaffeensis* infection in one study.[40]

The detection of morulae in lymphocytes in smears of the peripheral blood or buffy coat can occasionally be useful and even diagnostic. Unfortunately, morulae are seen in only a small minority of patients with HME. Thus, such testing has a low sensitivity, even though the finding of morulae is highly specific.

HME can be confirmed serologically with an IFA test using *E. chaffeensis* as the test antigen. IFA tests can be obtained through all state health departments. The current case definition of HME used by the Centers for Disease Control and Prevention requires at least a fourfold rise or fall in IFA titer against *E. chaffeensis* between the acute stage and convalescent stage, with a minimum titer of 1:64.[41] An important limitation of this test is that antibodies first become detectable 2 to 3 weeks after the onset of the illness. Thus, as with RMSF, serology is useful only in confirming infection, not in the decision to initiate therapy

Culture of *Ehrlichia* is extremely difficult. Only a few isolates of *E. chaffeensis* have been made from humans, and in such cases, detection required over 30 days of cultivation.[42] PCR-based testing is available for the diagnosis of HME, but such testing is performed only in special laboratories and remains mostly a research tool.

DIFFERENTIAL DIAGNOSIS

Ehrlichiosis may be easily confused with RMSF, a wide number of common viral illnesses (e.g., mononucleosis), thrombotic thrombocytopenic purpura, hematologic malignancy, cholangitis, the early phases of hepatitis A infection, and community-acquired pneumonia.

TREATMENT

The treatment of HME is the same as that of HGE (see below).

PROGNOSIS

Estimated mortality for patients with HME has ranged from 2% to 5%. The available mortality data are limited and are based on small case series, which may overestimate the actual risk.

Human Granulocytic Ehrlichiosis (Anaplasmosis)

EPIDEMIOLOGY

First described in 1994 in patients from the north central United States, HGE is now known to occur in Wisconsin, Minnesota, Connecticut, New York, Massachusetts, California, and Florida, as well as in western Europe.[43] A population-based surveillance study from northwestern Wisconsin reported an incidence of 9.5 HGE cases per 100,000 population in an area where the incidence of Lyme disease was 57 cases per 100,000 population.[44] Epidemiologic serosurveys have demonstrated that 3% to 15% of asymptomatic persons living in endemic areas may have antibodies to the HGE agent.[45]

A. phagocytophilum, the causative agent of HGE, is primarily transmitted by *Ixodes scapularis*, the tick that is also the vector of Lyme disease and babesiosis. *I. pacificus*, the black-legged tick, is the primary vector of HGE in the western United States, and *I. ricinis* is the presumed vector in Europe.

DIAGNOSIS

The diagnosis of HGE must sometimes be made on a circumstantial basis, by synthesis of the history and the clinical and epidemiologic features of an individual case. Clinicians should consider HGE as a possible diagnosis in any patient with a nonspecific febrile illness who becomes ill in the spring or summer months, who inhabits or has visited an endemic region, or who has a history of recent tick bite or tick exposure. Laboratory studies can support, and sometimes confirm, the diagnosis. However, the absence of such findings—in particular, the failure to detect morulae in leukocytes—should not dissuade clinicians when the overall picture suggests HGE.

Clinical Manifestations

The incubation period (5 to 14 days) and clinical features of HGE are similar to those of HME. Most patients with HGE have nonspecific symptoms such as malaise, myalgia, headache, nausea, vomiting, arthralgias, and cough. In a study of 18 adults with HGE, symptoms appeared an average of 5.5 days after a tick bite was noted.[46]

Clinically, HGE can range from mild to severe. Fatal HGE has been documented: a retrospective case study of 41 patients with laboratory-diagnosed HGE infection found a case-fatality rate of 4.9%.[47] Many of the patients who died had secondary opportunistic infections, such as fungal pneumonia. However, other studies have estimated that the mortality for HGE may actually be less than 1%.[43]

Laboratory Studies

The diagnosis of HGE can be confirmed by finding characteristic intraleukocytic morulae in the peripheral blood [*see Figure* 4] or buffy coat, by serology (using IFA testing), or by PCR testing.

The frequency of detecting morulae in the peripheral smears of patients with HGE has varied in different case series. In one report, typical morulae were found in the peripheral smear in 28 of 35 patients with laboratory-confirmed HGE infection.[47] The percentage of infected neutrophils ranged from 1% to 44% (median, 5%). In another study of patients with HGE, from the upper Midwest and New York, morulae were visible in neutrophils in 86 of 141 patients (61%).[48]

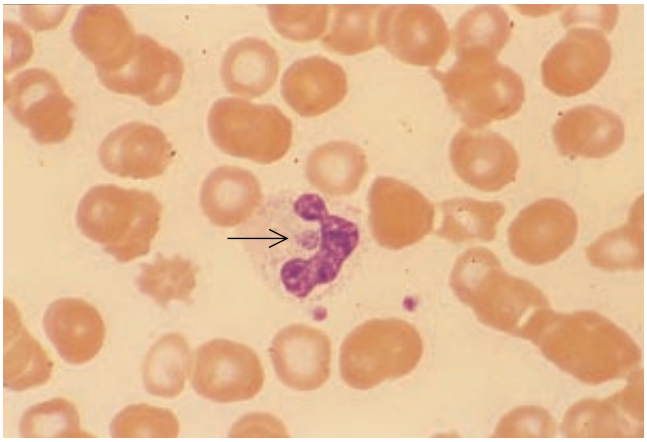


Figure 4 **In a peripheral blood smear from a patient with human granulocytic ehrlichiosis, a typical round ehrlichial morula is seen at the center of the neutrophil, adjacent to the nuclear lobes (arrow). Two platelets can be seen below and to the right of the neutrophil, for comparison. Wright-Giemsa stain was used; original magnification:** x **370.**

The most characteristic laboratory abnormality in patients with HGE is thrombocytopenia. In addition, patients with HGE typically manifest relative and absolute lymphopenia during the early phases of their infection, and significant increases in the band neutrophil counts occur during the first week of illness.[48] It is important to emphasize that automated blood counts are unable to distinguish between band and normal neutrophils or to detect morulae. Therefore, a manual differential WBC should be ordered whenever HGE is suspected.

Diagnosis of Coinfection with Borrelia burgdorferi

Because the agent of Lyme disease and HGE are transmitted by the same vector, it is not surprising that a number of reports have described patients who were coinfected with *B. burgdorferi* and *A. phagocytophilum*.[49] In studies of *I. scapularis* ticks from different locales, 2.2% to 26% of ticks were infected with both pathogens.[50] The diagnosis of HGE can easily be missed in such patients because the typical rash of early Lyme disease (erythema migrans) may mislead the clinician into failing to consider the possibility of coinfection with *A. phagocytophilum*. Such coinfection should be suspected when patients with presumed Lyme disease also have some or many of the following findings: leukopenia, thrombocytopenia, cough, high fever, or abnormal liver enzyme test results. Similarly, when rash is present in a patient with known or presumed HGE, the possibility of coinfection with *B. burgdorferi* should be suspected.

TREATMENT

As with HME, treatment should be initiated in all patients suspected of having HGE. The drug of choice is doxycycline, given orally or intravenously at a dosage of 200 mg/day in two divided doses. Intravenous therapy is preferred for patients who are seriously ill or experiencing nausea or vomiting. Children should be given doxycycline at a dosage of 2.5 to 3 mg/kg/day in two divided doses. There is no consensus on the use of chloramphenicol; many experts advise against its use for any form of ehrlichiosis. Patients who have intolerance or allergy to tetracy-

clines can be treated with rifampin for 7 to 10 days, but such patients require careful follow-up and monitoring, because there is only anecdotal information on the efficacy of this therapy in patients with HGE (and even less information on its efficacy in HME).[43]

At present, there are no guidelines for the treatment of ehrlichiosis in pregnancy. However, a report describing the successful treatment of two cases of HGE using rifampin suggests that other *Ehrlichia* species (e.g., *E. chaffeensis*) may be susceptible to rifampin as well.[51]

In one study, *A. phagocytophilum* was susceptible in vitro to fluoroquinolones and rifampin. However, the most active fluoroquinolone was trovafloxacin, which is not in general use because of concerns about hepatic toxicity.[52]

Even without treatment, most patients with ehrlichiosis probably make a full recovery. There is no evidence that untreated ehrlichiosis produces a chronic illness such as that which occurs with Lyme disease. However, infection may not confer long-lasting immunity. One report describes a patient who experienced two episodes of HGE spaced 2 years apart. At the onset of the second infection, the antibody titer to *A. phagocytophilum* had fallen from 1,280 to 80.[53]

Q Fever

EPIDEMIOLOGY AND ETIOLOGY

Q fever (the Q stands for "query") was first described in 1935 by Edward Derrick, after he investigated an outbreak of febrile illness involving abattoir workers in Queensland, Australia. Q fever is now known to be a worldwide zoonosis with highly variable clinical features. The causative organism, *Coxiella burnetii,* is a strictly intracellular pathogen that replicates and persists in cells in phagolysosomes. Microbial products, including acid phosphatase, help the organism resist the acidic and presumably harsh vacuolar environment.[54] In addition, vegetative *C. burnetii* cells form endogenous sporelike structures that resist extreme environmental conditions; spore formation has been observed in infected cardiac valves.[55] Because of these microbiologic characteristics, *C. burnetii* is capable of surviving prolonged drying in dust and excreta and can remain viable for months in water and milk. *C. burnetii* can be grown in cultured mammalian cells and in small animals but is not able to replicate on cell-free media.

The most common reservoirs for *C. burnetii* are cattle, goats, and sheep. Animals infected with *C. burnetii* are rarely symptomatic, but they shed organisms in their milk, urine, feces, and placentas. Infected placental tissue, postpartum discharges, and feces are presumed to be the principal sources of transmission to other animals and to humans. Domestic cats and dogs may also acquire *C. burnetii* and become sources for zoonotic transmission to humans.[56] Numerous tick species are also naturally infected with *C. burnetii*. Such infected ticks may transmit *C. burnetii* to other generations of ticks transovarially and via tick bites to wild rodents. Livestock sometimes acquire *C. burnetii* infections from infected ticks, but most often, they become infected by inhalation of contaminated dust.

Dairy and slaughterhouse workers are at increased risk for acquiring Q fever. However, sporadic cases of Q fever may occur in humans who acquire infection via infectious aerosols that were generated at relatively long distances from the site of acquisition. Infected milk may also account for some outbreaks of the disease in humans. Human infection from tick bites is extremely rare. *C. burnetii* is highly infectious in the laboratory, and outbreaks have occurred in researchers. Transmission between humans is rare, but transmission during delivery (to an obstetrician)[57] and sexual transmission[58] have been described. Exposure to wild rabbits, parturient cats, and products of feline parturition may lead to Q fever pneumonia in humans, sometimes by an indirect route (e.g., from contaminated clothing). Such cases illustrate an important clinical point: Q fever may occur in urban dwellers without obvious or direct contact with animals.

Q fever is endemic throughout the six populated continents, but it is particularly common in the Mediterranean and Persian Gulf regions. The diagnosis of Q fever should be considered in travelers and military personnel who return to the United States from endemic areas. In the United States and Canada, Q fever still occurs sporadically, particularly in areas where cattle, sheep, and goats are raised. Because of the difficulty in diagnosis and because spontaneous improvement occurs in many patients who do not receive effective antimicrobial therapy, the true prevalence of Q fever in the United States is unknown.

DIAGNOSIS

Clinical Manifestations

Q fever is usually classified into acute and chronic forms. Asymptomatic infection is also common. Acute Q fever may manifest as a self-limiting febrile illness, an influenzalike lower respiratory tract infection, hepatitis, or pneumonia. A small number of well-documented cases of meningoencephalitis from *C. burnetii* infection have also been described.[59] Patients with chronic *C. burnetii* infection may develop granulomatous hepatitis with prolonged fever, myocarditis, pericarditis, or endocarditis.

The incubation period of Q fever ranges from 10 to 39 days, with an average duration of 20 days.[60] The initial manifestations of acute *C. burnetii* infection are systemic and nonspecific—headache, chills, fever, myalgias, anorexia, and malaise. Skin rash does not occur in patients with Q fever. High fevers and headache often persist, and after 4 or 5 days of fever, patients typically manifest pneumonia, cough, chest pain, and inspiratory rales. In most cases, chest radiographs show focal areas of pneumonitis. Radiologic changes may be more marked than symptoms or physical findings. A minority of patients with Q fever pneumonia develop pleural effusions, hemoptysis, and even respiratory failure.[61] Pulmonary involvement occurs in about 50% of patients, but the incidence of this complication may vary with geographic location. In certain parts of the world, hepatitis may be more common than pneumonia in patients with acute Q fever.[62] Hepatitis in patients with acute Q fever usually lasts for 1 to 2 weeks and follows a benign, self-limited course. Complications are rare. However, granulomatous hepatitis with hepatosplenomegaly, jaundice, and abnormal liver function test results can also persist along with fever for up to 3 or 4 weeks.

In approximately 2% to 11% of infected persons, a chronic form of Q fever will develop insidiously a few months to as long as 20 years after the acute illness. Risk factors for chronic Q fever include underlying valvular heart disease and an immunocompromised state.[63] Infective endocarditis caused by *C. burnetii* is a common feature of the chronic syndrome.[64] The disease may affect healthy valves, previously damaged native valves, or prosthetic valves. Q fever endocarditis is sometimes insidious in onset and is often accompanied by granulomatous hepatitis. Other forms of endovascular infection with *C. burnetii* also occur, including infections of aneurysms and grafts.

Laboratory Studies

The diagnosis of acute Q fever is usually established by demonstration of a fourfold or greater rise in complement-fixing antibody titer against *C. burnetii* phase II antigen.[65,66] IFA techniques and ELISA offer greater sensitivity than complement-fixation methods. IFA techniques for the early detection of specific IgM antibody are the serodiagnostic method of choice. The diagnosis of chronic Q fever or Q fever endocarditis is established by detecting elevated titers (i.e., > 1:200) of IgG or IgA antibodies against *C. burnetii* phase I antigen or by a ratio of anti–phase I antibody to anti–phase II antibody of 1 or greater.

C. burnetii can be isolated from blood, sputum, or urine by intraperitoneal inoculation in guinea pigs, inoculation into chick embryos, or inoculation of cultured human fetal diploid fibroblasts. Because this organism is so highly contagious in the laboratory, attempts at isolation should be made only in a special biologic containment facility.

Echocardiography may not be diagnostic in Q fever endocarditis. A study that examined the heart valves removed from 28 patients with Q fever endocarditis showed that infected valves were usually fibrotic and calcified but that most had only slight inflammation. Only two of 16 patients with native Q fever endocarditis and two of 10 patients with bioprosthetic Q fever endocarditis had macroscopic vegetations, yet *C. burnetii* was isolated by culture in 64% of these patients and identified by PCR in 75%.[67]

DIFFERENTIAL DIAGNOSIS

Early in its course, Q fever resembles a variety of acute febrile illnesses, including an array of viral respiratory infections, viral hepatitis, and infectious mononucleosis. In patients with a history of contact with livestock, other zoonoses (e.g., brucellosis and leptospirosis) should be considered along with Q fever. In patients who present with pneumonia, infection with *Mycoplasma pneumoniae*, *Chlamydia pneumoniae* and *C. psittaci*, *Legionella pneumophila*, *Histoplasma capsulatum,* and the agents causing viral pneumonia should also be considered in the differential diagnosis. When hepatitis is present and noncaseating granulomas are found on liver biopsy, a host of causes of granulomatous hepatitis (e.g., tuberculosis, sarcoidosis, brucellosis, and histoplasmosis) must be considered. Q fever should always be considered when the clinical features of endocarditis are present but blood cultures are negative.

TREATMENT AND PREVENTION

Effective treatment for acute Q fever is tetracycline, 500 mg orally every 6 hours; doxycycline, 100 mg orally every 12 hours; or ciprofloxacin, 500 mg orally every 12 hours. Most patients treated early in the course of the infection recover rapidly; however, acute Q fever is usually self-limited.

If left untreated, Q fever endocarditis is fatal. Valve replacement is often necessary, in addition to prolonged antimicrobial therapy. Prolonged antibiotic therapy is necessary both for those who have valve surgery and for those in whom medical therapy alone is used. A study of cardiac valves in patients with Q fever endocarditis found that histologic, microbiologic, or molecular detection of *C. burnetii* was possible in more than 80% of patients who had been treated for less than 1 year; cultures and PCR tests were still positive in 22% and 33% of patients, respectively, who were treated for more than 1 year before valve excision.[67]

Medical regimens employing different antimicrobial agents, either alone or in combination, have been tried with variable success in Q fever endocarditis.[68,69] Some experts recommend treatment with doxycycline plus rifampin or with doxycycline plus a fluoroquinolone for at least 3 years.[70] Chloroquine has been also used as adjunctive therapy because of its ability to block intracellular vacuole acidification and, hence, growth of *C. burnetii*.

Patients with Q fever need not be isolated, because secondary cases do not occur. Killed vaccines made from *C. burnetii* grown in chick embryo cultures are immunogenic and can provide protection for persons at high risk, such as dairy and slaughterhouse workers, woolsorters, tanners, and laboratory workers. However, vaccines against *C. burnetii* are not commercially available for human use in the United States.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Paddock CD, Sumner JW, Comer JA et al: Rickettsia parkeri: a newly recognized cause of spotted fever rickettsiosis in the United States. Clin Infect Dis 38:805, 2004

2. Borjesson DL, Simon SI, Tablin F, et al: Thrombocytopenia in a mouse model of human granulocytic ehrlichiosis. J Infect Dis 184:1475, 2001

3. Kirkland KB, Marcom K, Sexton DJ, et al: Rocky Mountain spotted fever complicated by gangrene: report of six cases and review of the literature. Clin Infect Dis 16:629, 1993

4. Kirkland KB, Wilkinson WE, Sexton DJ: Therapeutic delay and mortality in cases of Rocky Mountain spotted fever. Clin Infect Dis 20:1118, 1995

5. Sexton DJ, Corey GR: Rocky Mountain "spotless" and "almost spotless" fever: a wolf in sheep's clothing. Clin Infect Dis 15:439, 1992

6. Archibald LK, Sexton DJ: Long-term sequelae of Rocky Mountain spotted fever. Clin Infect Dis 20:1122, 1995

7. Kirk J, Fine DJ, Sexton DJ, et al: Rocky Mountain spotted fever: a clinical review based upon 48 confirmed cases 1943–1986. Medicine (Baltimore) 69:35, 1990

8. Carpenter C, Gandhi TK, Kong LK, et al: The incidence of ehrlichial and rickettsial infection in patients with unexplained fever and recent history of tick bite in central North Carolina. J Infect Dis 180:900, 1990

9. Paddock CD, Greer PW, Ferebee TL, et al: Hidden mortality attributable to Rocky Mountain spotted fever: immunohistochemical detection of fatal, serologically unconfirmed disease. J Infect Dis 179:1469, 1999

10. Kirkland KB, Wilkinson WE, Sexton DJ: Therapeutic delay and mortality in cases of Rocky Mountain spotted fever. Clin Infect Dis 20:1118, 1995

11. Purvis JJ, Edwards MS: Doxycycline use for rickettsial diseases in pediatric patients. Pediatr Infect Dis J 19:871, 2000

12. Conlon PJ, Procop G, Fowler V, et al: Predictors of prognosis and acute renal failure in patients with Rocky Mountain spotted fever. Am J Med 101:621, 1996

13. Krussell A, Comer JA, Sexton DJ: Rickettsialpox in North Carolina: a case report. Emerg Infect Dis 8:727, 2002

14. Raoult D, Roux V: Rickettsioses as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10:694, 1997

15. Fournier P-E, Tissot-Dupont H, Gallais H, et al: Rickettsia mongolotimonae: a rare pathogen in France. Emerg Infect Dis 6:290, 2000

16. Paddock CD, Sumner JW, Comer JA, et al: Rickettsia parkeri: a newly recognized cause of spotted fever rickettsiosis in the United States. Clin Infect Dis 38:805, 2004

17. Murine typhus—Hawaii 2002. MMWR Morb Mortal Wkly Rep 52(50):1224, 2002

18. Walker DH, Barbour AG, Oliver JH, et al: Emerging bacterial zoonotic and vector-borne diseases: ecological and epidemiological factors. JAMA 275:463, 1996

19. Raoult D, Ndihokubwayo JB, Tissot-Dupont H, et al: Outbreak of epidemic typhus associated with trench fever in Burundi. Lancet 357:353, 1998

20. Duma RJ, Soneshine DE, Bozeman FM, et al: Epidemic typhus in the United States associated with flying squirrels. JAMA 245:2318, 1981

21. Massung RF, Davis LE, Slater K, et al: Epidemic typhus meningitis in the southwestern United States. Clin Infect Dis 32:979, 2001

22. Huys J, Kayihigi J, Freyens P, et al: Single-dose treatment of epidemic typhus with doxycycline. Chemotherapy 18:314, 1973

23. Sholdt LL, Rogers EJ Jr, Gerberg EJ, et al: Effectiveness of permethrin-treated military uniform fabric against human body lice. Mil Med 154:90, 1989

24. Kazar J, Brezina R: Control of rickettsial diseases. Eur J Epidemiol 7:282, 1991

25. Strickman D, Tanskul P, Eamsila C, et al: Prevalence of antibodies to rickettsiae in the human population of suburban Bangkok. Am J Trop Med Hyg 51:149, 1994

26. Sheehy TW, Hazlett D, Turk RE: Scrub typhus: a comparison of chloramphenicol and tetracycline in its treatment. Arch Intern Med 132:77, 1973

27. Watt G, Jjongsakul K, Suttinont C: Possible scrub typhus coinfections in Thai agricultural workers hospitalized with leptospirosis. Am J Trop Med Hyg 68:89, 2003

28. Song JH, Lee C, Chang WH, et al: Short-course doxycycline treatment versus conventional tetracycline therapy for scrub typhus: a multicenter randomized trial. Clin Infect Dis 21:50, 1995

29. Brown GW, Saunders JP, Singh S, et al: Single dose doxycycline therapy for scrub typhus. Trans R Soc Trop Med Hyg 72:412, 1978

30. Strickman D, Sheer T, Salata K, et al: In vitro effectiveness of azithromycin against doxycycline-resistant and -susceptible strains of Rickettsia tsutsugamushi, etiologic agent of scrub typhus. Antimicrob Agents Chemother 39:2406, 1995

31. Watt G, Kantipong P, Jongsakul K, et al: Doxycycline and rifampicin for mild scrub-typhus infections in northern Thailand: a randomised trial. Lancet 356:1057, 2000

32. Gongora-Biachi RA, Zavala-Velazquez J, Castro-Sansores CJ, et al: First case of human ehrlichiosis in Mexico. Emerg Infect Dis 5:481, 1999

33. Carpenter C, Gandhi TK, Kong LK, et al: The incidence of ehrlichial and rickettsial infection in patients with unexplained fever and recent history of tick bite in central North Carolina. J Infect Dis 180:900, 1999

34. Fishbein DB, Kemp A, Dawson JE, et al: Human ehrlichiosis: prospective active surveillance in febrile hospitalized patients. J Infect Dis 160:803, 1989

35. Standaert SM, Dawson JE, Schaffner W, et al: Ehrlichiosis in a golf-oriented retirement community. N Engl J Med 333:420, 1995

36. Lockhart JM, Davidson WR, Stallknecht DE, et al: Isolation of Ehrlichia chaffeensis from wild white-tailed deer (Odocoileus virginianus) confirms their role as natural reservoir hosts. J Clin Microbiol 35:1681, 1997

37. Kocan AA, Levesque GC, Whitworth LC, et al: Naturally occurring Ehrlichia chaffeensis infection in coyotes from Oklahoma. Emerg Infect Dis 6:477, 2000

38. Roland WF, McDonald G, Caldwell CW, et al: Ehrlichiosis a cause of prolonged fever. Clin Infect Dis 20:821, 1995

39. Fishbein DB, Dawson JE, Robinson LE: Human ehrlichiosis in the United States, 1985 to 1990. Ann Intern Med 120:736, 1994

40. Ratnasamy N, Everett ED, Roland WE, et al: Central nervous system manifestations of human ehrlichiosis. Clin Infect Dis 23:314, 1996

41. Dawson JE, Fishbein DG, Eng TR, et al: Diagnosis of human ehrlichiosis with the indirect fluorescent antibody test: kinetics and specificity. J Infect Dis 162:91, 1990

42. Dumler JS, Bakken JS: Ehrlichial diseases of humans: emerging tick-borne infections. Clin Infect Dis 20:1102, 1995

43. Bakken JS, Dumler JS: Human granulocytic ehrlichiosis. Clin Infect Dis 31:554, 2000

44. Belongia EA, Gale CM, Reed KD, et al: Population-based incidence of human granulocytic ehrlichiosis in northwestern Wisconsin, 1997–1999. J Infect Dis 184:1470, 2001

45. Walder G, Tiwald G, Dierich MP, et al: Serological evidence for human granulocytic ehrlichiosis in Western Austria. Eur J Clin Microbiol Infect Dis 22:543, 2003

46. Aguero-Rosenfeld ME, Horowitz HW, Wormser GP, et al: Human granulocytic ehrlichiosis: a case series from a medical center in New York state. Ann Intern Med 125:904, 1996

47. Bakken JS, Krueth J, Wilson-Noldskog C, et al: Clinical and laboratory characteristics of human granulocytic ehrlichiosis. JAMA 275:199, 1996

48. Bakken JS, Aguero-Rosenfeld ME, Tilden RL, et al: Serial measurements of hematologic counts during the active phase of human granulocytic ehrlichiosis. Clin Infect Dis 32:862, 2001

49. Nadelman RB, Horowitz HW, Hsieh TC, et al: Simultaneous human granulocytic ehrlichiosis and Lyme borreliosis. N Engl J Med 337:27, 1997

50. Schwartz I, Fish D, Daniels TJ: Prevalence of the rickettsial agent of human granulocytic ehrlichiosis in ticks from a hyperendemic focus of Lyme disease (letter). N Engl J Med 337:49, 1997

51. Buitrago MI, Ijdo JW, Rinaudo P, et al: Human granulocytic ehrlichiosis during pregnancy treated successfully with rifampin. Clin Infect Dis 27:213, 1998

52. Klein MB, Nelson CM, Goodman JL: Antibiotic susceptibility of the newly cultivated agent of human granulocytic ehrlichiosis: promising activity of quinolones and rifamycins. Antimicrob Agents Chemother 41:76, 1997

53. Horowitz HW, Aguero-Rosenfeld ME, Dumler JS, et al: Reinfection with the agent of human granulocytic ehrlichiosis. Ann Intern Med 129:461, 1998

54. Baca OG, Li YP, Kumar H: Survival of the Q fever agent Coxiella burnetii in the phagolysosome. Trends Microbiol 2:476, 1994

55. McCaul TF, Dare AJ, Gannon JP, et al: In vivo endogenous spore formation by Coxiella burnetii in Q fever endocarditis. J Clin Pathol 47:978, 1994

56. Marrie TJ, Raoult D: Q fever: a review and issues for the next century. Int J Antimicrob Agents 8:145, 1997

57. Stein A, Raoult D: Q fever during pregnancy: a public health problem in southern France. Clin Infect Dis 27:592, 1998

58. Milazzo A, Hall R, Storm PA, et al: Sexually transmitted Q fever. Clin Infect Dis 33:399, 2001

59. Bernit E, Pouget J, Jambon F, et al: Neurological involvement in acute Q fever: a report of 29 cases and review of the literature. Arch Intern Med 162:693, 2002

60. Spelman DW: Q fever: a study of 111 consecutive cases. Med J Aust 1:547, 1982

61. Sampere M, Font B, Font J, et al: Q fever in adults: review of 66 clinical cases. Eur J Clin Microbiol Infect Dis 22:108, 2003

62. Tissot-Dupont H, Raoult D, Brouqui P, et al: Epidemiologic features and clinical presentation of acute Q fever in hospitalized patients: 323 French cases. Am J Med 93:427, 1992

63. Brouqui P, Tissot-Dupont H, Drancourt M, et al: Chronic Q fever: ninety-two cases from France, including 27 cases without endocarditis. Arch Intern Med 153:642, 1993

64. Tobin MJ, Cahill N, Gearty G, et al: Q fever endocarditis. Am J Med 72:396, 1982

65. Soriano F, Camacho MT, Ponte C, et al: Serological differentiation between acute (late control) and endocarditis Q fever. J Clin Pathol 46:411, 1993

66. Fournier P-E, Marrie TJ, Raoult D: Diagnosis of Q fever. J Clin Microbiol 36:1823, 1998

67. Lepidi H, Houpikian P, Liang Z, et al: Cardiac valves in patients with Q fever endocarditis: microbiological, molecular, and histologic studies. J Infect Dis 187:1097, 2003

68. Tobin MJ, Cahill N, Gearty G, et al: Q fever endocarditis. Am J Med 72:396, 1982

69. Stein A, Raoult D: Detection of Coxiella burnetii by DNA amplification using polymerase chain reaction. J Clin Microbiol 30:2462, 1992

70. Levy PY, Drancourt M, Etienne J, et al: Comparison of different antibiotic regimens for therapy of 32 cases of Q fever endocarditis. Antimicrob Agents Chemother 35:533, 1991

Acknowledgments

Figure 2 Slide courtesy of Dr. Mark Lebwohl, Mount Sinai Medical Center, New York.

Figure 4 Slide courtesy of Dr. P. Joanne Cornbleet, Stanford University School of Medicine, Stanford, California.

XVIII INFECTIVE ENDOCARDITIS

David T. Durack, m.b., d.phil., f.a.c.p.
Adolf W. Karchmer, m.d., f.a.c.p.

Infective endocarditis is a relatively rare but important disease, presenting in a wide variety of forms and manifestations; all are associated with significant morbidity and mortality. The diagnosis and management of infective endocarditis involves many clinical specialties, including internal medicine, cardiology, infectious diseases, microbiology, radiology, cardiovascular surgery, neurology, nephrology, and dentistry.

Definitions and Terminology

Infective endocarditis is a localized microbial infection of cardiac valves or mural endocardium caused by bacteria or fungi. The primary lesion is a vegetation, which is an infected platelet-fibrin thrombus located inside the heart. Native valve endocarditis is an infection of either normal or abnormal natural heart valves, whereas prosthetic valve endocarditis (PVE) involves implanted artificial valves. Nosocomial endocarditis occurs as a complication of medical treatment. On the basis of its clinical course, endocarditis may be classified as either acute or subacute. The acute form, often called acute bacterial endocarditis (ABE), is caused by invasive pathogens and tends to be rapidly progressive, usually leading to hospital admission less than a week after clinical onset and causing death in less than 4 weeks unless successfully treated. The subacute form, often called subacute bacterial endocarditis (SBE), is usually caused by low-grade pathogens or commensal organisms, with symptoms often present for weeks to months before diagnosis. Nonbacterial thrombotic endocarditis (NBTE) and marantic endocarditis refer to sterile vegetations in the heart, which can develop at the same locations as the vegetations of infective endocarditis. Infective endaortitis and infective endarteritis are analogous conditions that are localized on the endothelial surface of the aorta or large arteries, respectively.

Epidemiology

The overall incidence of infective endocarditis ranges from 1.7 to 6.2 cases per 100,000 person-years.[1-3] Although the overall incidence has remained relatively stable over the past 5 decades, the relative frequency of the disease in particular subgroups has changed, as have many other aspects of its epidemiology. The median age of patients with endocarditis has increased; 50% of patients are older than 55 years.[1,4,5]

Among the major predisposing conditions, chronic rheumatic heart disease has become relatively uncommon in developed countries, whereas degenerative valvular diseases, such as calcified aortic stenosis, calcified mitral valve annulus, and mitral valve prolapse have become more important.[6] Cases of endocarditis associated with parenteral drug abuse and prosthetic cardiac valves have become more common, as has nosocomial infective endocarditis. Infections involving implanted intravascular devices other than valves are increasing in frequency.

Etiology

Gram-positive cocci, comprising various species of streptococci and enterococci, as well as *Staphylococcus aureus*, are the leading cause of community-acquired native valve endocarditis[1-3,6] [*see Table 1*]. SBE is usually caused by relatively avirulent bacteria, the most common species being streptococci from the normal oral or gastrointestinal flora. These organisms lack sufficient invasiveness to infect normal heart valves or endocardium, but they can infect deformed heart valves and some congenital cardiac lesions. The leading examples are the α-hemolytic streptococci, a heterogeneous collection of species that are loosely grouped together under the term viridans streptococci because they cause incomplete, greenish hemolysis when grown on blood agar.

Many of the streptococci that cause endocarditis can be categorized according to Lancefield serogrouping. About 20% are group D (mainly enterococci and *Streptococcus bovis*), about 15% are group H (*S. sanguis* and others), and about 15% belong to other serogroups, including B, C, G, and K. About 5% are anaerobic streptococci, and the remaining 40% to 45% are nongroupable viridans streptococci.

Enterococcus faecalis, E. faecium, and the nonenterococcal group D streptococcus *S. bovis*, all of which originate from the GI and genitourinary tracts, are important causes of SBE. The portal of entry for *S. bovis* is often a malignant or premalignant lesion in the colon.[7] Consequently, evaluation for possible colonic lesions is mandated for patients who have developed *S. bovis* SBE.

Coagulase-negative staphylococci (many of which are species other than *S. epidermidis*), and fastidious gram-negative HACEK organisms (*Haemophilus, Actinobacillus actinomycetemcomitans, Cardiobacterium hominis, Eikenella*, *Kingella*) are other important causes of infective endocarditis. Species that less commonly cause infective endocarditis are *Brucella*, *Legionella, Coxiella burnetii, Chlamydia psittaci*, and *Corynebacterium*.[8,9] *Bartonella*, recognized as a cause of infective endocarditis only in the mid-1990s, may account for up to 3% of cases in some regions.[10] *B. quintana* endocarditis occurs more commonly in the homeless, and endocarditis caused by *B. henselae* may be associated with cats.[10-12] Corynebacteria (diphtheroids) usually involve prosthetic heart valves. Endocarditis caused by *Coxiella burnetii*, a notably indolent infection, occurs in the setting of preexisting valvular disease[13] [*see 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*].

ABE is commonly caused by primary pathogens, microorganisms that are more invasive than most of the species causing SBE. These invasive organisms can infect both normal and abnormal heart valves and mural endocardium; can rapidly destroy cardiac valves; can more readily invade perivalvular tissues to form abscesses; and can establish metastatic suppurative foci at distant sites in the body such as the brain or spleen. *S. aureus* is the most frequent and most important cause of ABE; others include *S. pneumoniae*, group A streptococci, *Neisseria gonorrhoeae, Salmonella* species, other members of the family Enterobacteriaceae, and *Pseudomonas aeruginosa*. *Salmonella* is the most frequent cause of infective aortitis in the elderly; these patients may have a continuous *Salmonella* bacteremia in the absence of a cardiac murmur.

Nosocomial endocarditis is a complication of nosocomial bacteremias or fungemias, often associated with intravascular and other indwelling devices, genitourinary tract manipulation, or wound infections.[14-18] The primary pathogens are *S. aureus*, enterococci, and coagulase-negative staphylococci, most of which are methicillin-resistant *S. epidermidis*. In parenteral drug abusers, *S. aureus* (often methicillin resistant) causes 50% or more of cases of infective endocarditis[19] [*see Table 1*].

Fungal endocarditis can be caused by yeasts or molds.[20,21] The leading causative agents are yeasts of the genera *Candida* and *Torulopsis;* endocarditis from these yeasts arises as a complication of parenteral drug abuse, of prosthetic valves, or of intravascular devices, such as pacemakers. Endocarditis caused by mycelial or dimorphic fungi is relatively rare but can occur in parenteral drug abusers, patients with prosthetic valves, or immunocompromised patients. *Aspergillus* species cause most of these cases; the remainder are caused by *Histoplasma*, *Penicillium*, and a wide variety of other fungi.

Table 1 Causes of Native Valve Endocarditis[2,5,6,18,49,109]

Microorganism	*Representative Percentage of Cases*		
	Community Acquired	*Nosocomial*	*I.V. Drug Abusers*
Staphylococcus aureus	25	55	55
Coagulase-negative staphylococci	3	5	2
Streptococci	35	5	10
Enterococci	10	20	15
Gram-negative bacilli	3	5	8
HACEK	3	rare	rare
Pneumococci	1	rare	rare
Yeasts and fungi	1	5	4
Polymicrobial and miscellaneous	5	3	4
Culture-negative	15	2	2

Note: Reported percentages of cases vary widely between many published series. HACEK—*Haemophilus* species, *Actinobacillus actinomycetemcomitans*, *Cardiobacterium hominis*, *Eikenella* species, and *Kingella* species

Pathogenesis

HOST FACTORS

Two host factors strongly predispose to development of infective endocarditis: (1) a damaged or otherwise abnormal endocardial surface and (2) high-velocity, turbulent blood flow across a defective valve or a congenital defect. Both factors favor platelet deposition. When blood is driven from a high-pressure area into a low-pressure sink, the dynamics of pressure gradients and turbulent flow favor deposition of platelets on nearby endothelium, especially if that endothelium is already abnormal as a result of previous disease such as rheumatic fever, congenital abnormalities, degenerative conditions, or infective endocarditis. Subsequently, a form of localized thrombosis can occur, resulting in the formation of a sterile platelet-fibrin aggregate known as nonbacterial thrombotic endocarditis.[22] During transient bacteremias, passing microorganisms can adhere to this nidus, which provides them with both nutrition and protection from host defenses, especially leukocytes. The platelet-fibrin layers form an effective physical barrier between the embedded bacteria and leukocytes from the blood. This situation permits luxuriant growth; the organisms often attain vast numbers and form dense colonies within the platelet-fibrin aggregate. The bacteria themselves may promote further thrombosis by elaborating extracellular products that cause platelet aggregation or by eliciting procoagulant tissue factor from the endothelial substrate and local monocytes.[23,24] In this way, the newly infected thrombotic lesion grows to form a macroscopic vegetation; these vegetations constitute the prototypical pathologic lesion of infective endocarditis.[22]

Local host defenses that could inhibit or kill bacteria in the vegetation include leukocytes, antibodies and complement, and platelet-derived bactericidal proteins (thrombocidins).[25,26] These antimicrobial host defenses may succeed in curing some early cases of endocardial infection, but apparently they seldom if ever succeed in eradicating the organisms once a vegetation is fully established.

PATHOGEN FACTORS

The streptococci that most frequently cause SBE—the viridans streptococci (including *S. mutans, S. mitior*, and *S. sanguis* [group H]), and *S. bovis* (group D)—readily adhere to platelet-fibrin thrombi, owing to the presence of dextran and other "sticky" molecules (adhesins) on their surfaces.[27] Streptococcal species that do not produce dextran may also cause endocarditis, but they do so less frequently. Virulent organisms, such as those associated with ABE, may adhere to either normal or abnormal endocardial surfaces. Fibronectin receptors and clumping factor, which occur on the surface of *S. aureus*, appear to facilitate the adherence of these organisms to cardiac valves.[27-29] Bacterial species that are resistant to platelet-derived microbicidal proteins are more likely to cause endocarditis than susceptible strains.

Clinical Presentations

SUBACUTE BACTERIAL ENDOCARDITIS

The constitutional symptoms of SBE usually begin insidiously and often persist for weeks to months. Fevers, sweats, weakness, myalgias, arthralgias, malaise, anorexia, and easy fatigability are prominent. Fewer than 5% of patients are afebrile, and such patients are often elderly, markedly malnourished, or azotemic. Chills and chilly sensations are common, but frank rigors are unusual.

Because of greater awareness of this disease and more frequent use of blood cultures in the evaluation of febrile illnesses, the diagnosis of SBE is now made earlier than in the past; consequently, many of the classic features of longer-standing SBE (e.g., splenomegaly, clubbing, and Osler nodes) are seldom present. Fever and other nonspecific symptoms in the presence of a predisposing cardiac lesion may be the only clinical manifestations of SBE in some patients. The presenting complaints may arise from organs or sites other than the heart, which may confuse the diagnosis. For example, if the patient has meningitis, cerebral emboli, or glomerulonephritis, the physician's attention may be focused on the central nervous system or kidneys as the primary site of illness. Although it was once taught that the symptoms of endocarditis in the elderly are milder than in young patients, current data suggest that the clinical features are fairly similar across age groups.[30]

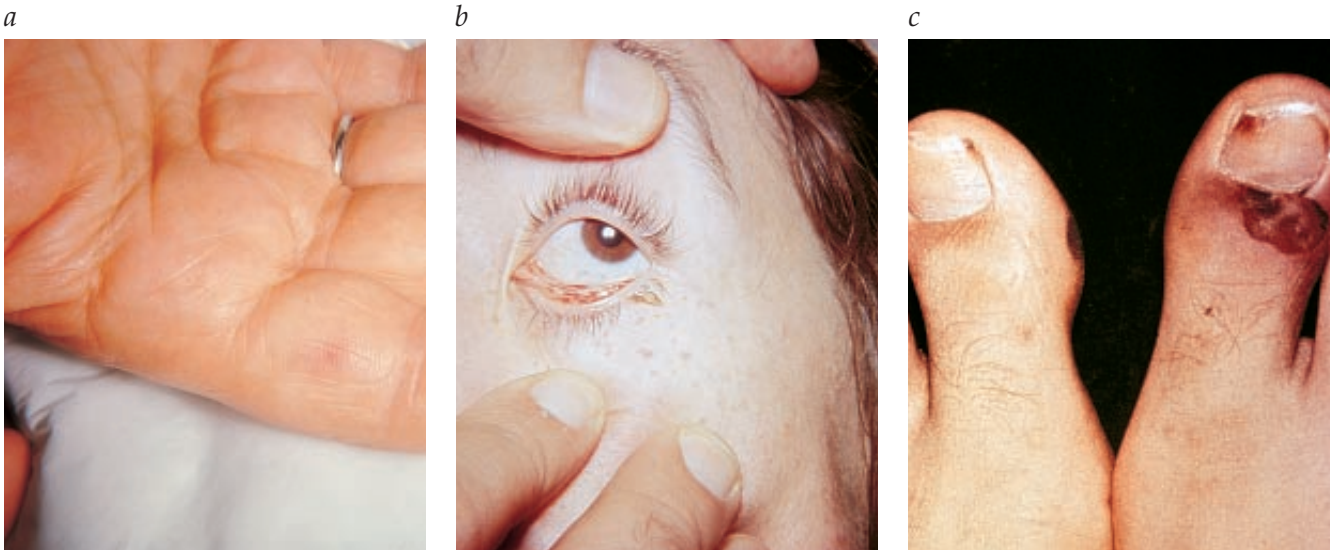


Figure 1 **Findings on physical examination of patients with endocarditis. (*a*) Erythematous, palpable, nontender lesion at the base of the first finger consistent with a Janeway lesion. (*b*) Conjunctival petechiae. (*c*) Pustulonecrotic septic embolic lesions at the base of the nail of the right great toe and at the medial aspect of the left great toe at the level of the distal interphalangeal joint.**

ACUTE BACTERIAL ENDOCARDITIS

The onset of ABE is usually abrupt, and rigors are common.[31,32] Fevers reach 39.4° to 40.6° C (102.9° to 105.1° F) and are often remittent. Cutaneous manifestations, particularly petechiae and small peripheral infarcts, may be prominent, especially in ABE from *S. aureus*. Occasionally, the clinical features in a patient with acute *S. aureus* endocarditis mimic those of acute meningococcemia, with similar skin lesions, including petechiae, purpura, and focal gangrene; similar hematologic changes, including disseminated intravascular coagulation; and neurologic findings of nuchal rigidity and CSF pleocytosis. Pustular petechiae or purulent purpura strongly suggest *S. aureus* endocarditis rather than meningococcemia [*see Figure 1*]. A Gram stain and culture of material from a skin lesion sometimes can reveal the etiologic agent and thus direct antibiotic therapy.

Emboli are common in ABE. Metastatic infections in the spleen, bones (particularly the vertebrae), joints, kidneys, brain, eye (endophthalmitis), and lungs may arise from either septic embolization or sustained bacteremia. These metastatic infections may cause organ-specific symptoms or persistent fever in spite of antimicrobial therapy and may require drainage or surgical intervention. Osler nodes may occur, but less often than in SBE; Janeway lesions occur in 5% to 10% of patients who have *S. aureus* endocarditis [*see* Cutaneous Manifestations, *below*].

The appearance of a new cardiac murmur, particularly one characteristic of valvular insufficiency, strongly suggests valvular destruction and thus helps confirm a diagnosis of ABE. Valvular damage can lead to severe heart failure, necessitating valve replacement surgery. A few patients with ABE have no detectable cardiac murmur.

CARDIAC FINDINGS

The cardiac manifestations of infective endocarditis reflect any underlying valvular or congenital lesions, with superimposed findings from endocarditis itself. Murmurs are present in more than 90% of patients. Changes in the intensity of a systolic murmur may be associated with the development of anemia, high fever, or tachycardia and are often of little significance. However, the appearance of a new murmur indicating valvular regurgitation is a key diagnostic event, which also may have implications for surgical intervention. A new aortic diastolic murmur suggests dilatation of the aortic annulus or eversion, rupture, or fenestration of an aortic leaflet. The sudden onset of a loud mitral pansystolic murmur suggests rupture of a chorda tendineae or fenestration of a mitral valve leaflet.

Heart failure resulting from valvular dysfunction may be progressive, becoming severe in some patients. Extension of infection into the annulus may result in cardiac rhythm disturbances, particularly if the infection progresses through the right coronary and noncoronary leaflet portion of the aortic annulus into the membranous septum and the area of the atrioventricular node.[33] Occasionally, annular infection extends to the pericardium and causes pericarditis [*see* Diagnostic Features of Cardiac Complications, *below*].

CUTANEOUS MANIFESTATIONS

Petechiae commonly occur in the conjunctivae [*see Figure 1*], in the oropharynx, and on the skin; they are particularly common on the lower extremities. Petechiae may continue to appear for some time despite appropriate antibiotic treatment. Linear subungual "splinter" hemorrhages located in the base or in the middle of the nail bed are a feature of SBE, whereas splinter hemorrhages in the distal nail bed are more often the result of trauma. Osler nodes—tender, purplish subcutaneous nodules that develop in the pulp of the fingers and disappear within several days—occur in about 5% of patients with endocarditis. Osler nodes may be caused by small emboli or may result from an immunologically mediated small-vessel vasculitis. Small, flat, nonpainful erythematous or hemorrhagic areas on the palms or soles, called Janeway lesions, are common in ABE and also may occur in SBE [*see Figure 1*].

MUSCULOSKELETAL FEATURES

Myalgias, arthralgias, arthritis, or low back pain occurs in 40% to 50% of patients with SBE; in about half of these patients, such symptoms represent either initial or prominent manifestations of the disease. Painful, warm, red, tender joints may be noted, but joint effusions are rare. Immunologic mechanisms

cause synovial inflammation, articular symptoms, and abnormal laboratory test results. Rheumatoid factor is present in up to 50% of patients with endocarditis of more than 6 weeks' duration; antinuclear antibody is found in some patients; and circulating immune complexes occur in 75% to 90% of patients. Clubbing of the fingers is now seen in less than 15% of patients.

OCULAR FINDINGS

Petechial hemorrhages, flame-shaped hemorrhages, Roth spots, and cotton-wool exudates may be seen in the retinas of patients with endocarditis. Roth spots, which are oval white areas surrounded by a zone of bright-red hemorrhage, are noted in 3% to 5% of patients. Such ocular findings are not pathognomonic of endocarditis; they may be observed in patients with other disorders, such as severe anemia or collagen vascular diseases.

EMBOLIC PHENOMENA

Significant arterial emboli occur in 30% to 50% of patients with endocarditis.[34,35] Symptoms and signs include stroke; monocular blindness with occlusion of the central retinal artery; acute abdominal pain, ileus, and melena from mesenteric arterial occlusion; and pain and gangrene in the extremities. Emboli to the CNS are common and especially important, because they adversely affect survival rates and often result in permanent disability.[35] Coronary emboli, which are found in as many as 50% of endocarditis patients who undergo autopsy, are often asymptomatic but occasionally result in frank myocardial infarction.

Factors associated with increased risk of embolization include vegetations of 10 mm or more in size as seen on echocardiography; vegetations on the mitral valve, particularly the anterior leaflet; vegetations that increase in size despite appropriate antibiotic therapy; and infection by *S. aureus*.[34-36] The incidence of arterial emboli decreases about 10-fold during the initial 2 weeks of antimicrobial therapy.[34] Nevertheless, emboli occasionally occur late, after microbiologic cure has been achieved, and do not necessarily indicate that antimicrobial treatment has failed.

Pulmonary emboli are a common and important complication of right-sided endocarditis, frequently causing pulmonary infarcts or focal pneumonitis. Either may evolve into lung abscess, empyema, or pyopneumothorax, especially when invasive pathogens such as *S. aureus* are involved.

SPLENIC ABNORMALITIES

Splenomegaly occurs in 15% to 30% of patients with endocarditis. Splenic infarcts occur in up to 40% of patients; they may occur with or without splenomegaly. Splenic infarcts may cause sharp left upper quadrant pain, but they are commonly asymptomatic. Splenic abscesses, which develop in about 5% of patients, may manifest as left shoulder, left upper quadrant, or pleuritic left chest pain; fever that persists during antibiotic therapy; or a relapse of bacteremia. Splenic lesions are best imaged by computed tomography, magnetic resonance techniques, and, to a lesser degree, ultrasonography; however, the differentiation of abscesses from infarcts by imaging is difficult.

RENAL MANIFESTATIONS

Microscopic hematuria is observed in about 50% of infective endocarditis patients. Embolic renal infarction may cause flank pain and hematuria, but it rarely causes renal failure. Some degree of acute renal failure occurs in up to one third of patients and is associated with a worse prognosis,[37] but it usually resolves completely if therapy for endocarditis is successful. Diffuse membranoproliferative glomerulonephritis results from immune complex glomerulonephritis, with deposition of IgG, IgM, and complement in a granular or nodular fashion in the glomerular basement membrane or on the basement membrane in subepithelial or subendothelial locations. Bacterial antigen may be identified in these glomerular deposits. In such patients, serum complement levels usually are reduced. Diffuse membranoproliferative glomerulonephritis may cause renal failure, which generally resolves after treatment of the infection. Focal embolic glomerulonephritis, originally thought to result from small bacterial emboli, is now recognized as an anatomic variation of immune complex disease. It occasionally causes renal failure.

MYCOTIC ANEURYSMS

Mycotic aneurysms are arterial aneurysms that develop in association with infections, especially infective endocarditis. Mycotic aneurysms occur in 2% to 8% of infective endocarditis patients and can form in any artery; they are particularly important when they involve cerebral arteries[38,39] [*see* Neurologic Manifestations, *below*]. They may become symptomatic early, during the active phase of endocarditis, or late, after valvular infection has been eradicated.[40] The clinical manifestations, which arise from enlargement or rupture of the aneurysm, include headache, pain, a pulsatile mass, persistent fever despite appropriate antibiotic treatment, focal signs from pressure on adjacent structures, the sudden development of an expanding hematoma, or signs of major blood loss. Small aneurysms (less than 5 mm in size) may resolve during antibiotic therapy. For larger aneurysms, to prevent possible rupture, therapeutic embolization, clipping, or excision is usually indicated when such interventions are feasible without undue risk.

NEUROLOGIC MANIFESTATIONS

Neurologic complications develop in 25% to 40% of patients with endocarditis, causing major morbidity and increased mortality.[41-43] Altered mental status at presentation is associated with higher mortality at 6 months.[44] Cerebral embolism may be an initial manifestation of endocarditis. Strokes caused by cerebral emboli, commonly to the middle cerebral artery or one of its branches, are the most frequent major neurologic complication, occurring in about 15% of patients. Some patients have multiple small embolic infarcts, which may manifest as an altered level of consciousness, seizures, fluctuating focal neurologic signs, or a combination of these symptoms. An intracerebral hemorrhage in a patient with infective endocarditis can be secondary to an embolic stroke or to rupture of a mycotic aneurysm.[42] Cerebral mycotic aneurysms occur in 1% to 5% of patients with infective endocarditis and most commonly affect the distal branches of the middle cerebral arteries.[39] Multiple aneurysms occur in some patients. In patients with symptomatic intracranial aneurysms, the overall mortality is high—above 50%. Patients with mycotic aneurysms may present with headache, focal signs, or, if the aneurysm ruptures, manifestations of acute intracerebral or subarachnoid hemorrhage. A slowly leaking aneurysm may cause mild meningeal irritation. In such cases, the cerebrospinal fluid, although sterile, may contain erythrocytes, leukocytes, and an increased concentration of protein.

In the search for an intracerebral aneurysm, contrast-enhanced CT may provide localizing information by detecting intracerebral bleeding. Magnetic resonance angiography is insufficiently sensitive to reliably detect aneurysms that are 5 mm or

less; therefore, cerebral angiography is the optimal diagnostic test.[45,46] Cerebral angiography for the detection of aneurysms has been advised for patients with focal neurologic signs, especially in the setting of ABE, and for patients with persistent unexplained headache or meningeal irritation, particularly if anticoagulant therapy is planned.

Treatment of a mycotic aneurysm that has been detected by angiography depends on its location and surgical accessibility, the presence or absence of hemorrhage, and changes in size that may occur during antimicrobial therapy. Resolution or healing of mycotic aneurysms during treatment of endocarditis has been demonstrated by angiography.[47] However, a leaking aneurysm, an aneurysm that is large or progressively enlarging, or one that persists after antibiotic therapy should be removed surgically, provided it is accessible. These complex cases should be managed with the input from specialists in infectious disease, neurology, radiology, and neurosurgery.

Brain abscesses are uncommon in SBE. In patients with acute *S. aureus* endocarditis, septic emboli may give rise to multiple intracerebral foci of inflammation or to small abscesses.

Toxic encephalopathy and seizures, which are usually triggered by emboli or strokes, also may complicate active endocarditis. A CSF pleocytosis with polymorphonuclear leukocytes predominating is observed in some patients with ABE caused by pyogenic organisms, especially *S. aureus*. Patients with SBE may have findings of aseptic inflammation in the CSF.

ENDOCARDITIS ASSOCIATED WITH PARENTERAL DRUG ABUSE

The annual incidence of endocarditis in injecting drug users (IDUs) is 0.2% to 2.0%. At the time of their initial attack of endocarditis, 70% to 80% of IDUs have no history or findings of preexisting valvular heart disease. In IDUs, the tricuspid valve is infected more frequently (55%) than the aortic valve (35%) or mitral valve (30%). Multiple episodes of endocarditis are common in IDUs.

S. aureus is responsible for approximately 55% of endocarditis in IDUs [*see Table 1*].[19,48] Although polymicrobial endocarditis, such as simultaneous infection by *P. aeruginosa* and *S. aureus*, is extremely rare in non-IDUs with native valve endocarditis, up to 5% of cases in IDUs involve multiple organisms. In IDU-associated endocarditis, enterococci, streptococci, and *Candida* primarily infect the aortic and mitral valves. *S. aureus* and *P. aeruginosa* infect valves on both the right and the left side of the heart. *S. aureus*, however, accounts for almost 80% of right-sided endocarditis in IDUs. Methicillin-resistant *S. aureus* (MRSA) is often encountered in these cases, possibly associated with repeated self-administration of antibiotics. Blood cultures generally reveal the causative organism. Microbiologic evaluation of surgically removed arterial emboli may be required to identify the cause in a few cases, such as those caused by fungi.

Although many of the manifestations of endocarditis in IDUs are similar to those of ABE in non-IDUs, there are some differences, because of the high frequency of tricuspid valve involvement, the spectrum of infecting organisms, and occasional pulmonic valve infection. High fevers, chills, rigors, malaise, cough, and, especially, pleuritic chest pain are common presenting complaints in right-sided endocarditis in IDUs. Septic pulmonary emboli occur in about 75% of cases, particularly in patients with *S. aureus* infection, and cause sputum production, hemoptysis, and initial radiologic findings that may suggest pneumonia. Cavitation of embolic pulmonary lesions is quite common. Significant cardiac murmurs are heard in most patients at some time during their illness but may not be present initially. The murmur of tricuspid regurgitation, a short ejection systolic murmur that is louder on inspiration, may be difficult to detect. Hemodynamically, significant tricuspid insufficiency is manifested by V waves in the jugular vein and a pulsating liver.

Because *S. aureus* and other pyogenic bacterial species are the predominant causes of infective endocarditis in IDUs, metastatic infections are a frequent complication. Neurologic manifestations and peripheral emboli are common; the latter may occlude major vessels and require surgical management.

PROSTHETIC VALVE ENDOCARDITIS

The incidence of PVE is 1% to 2% at 1 year and approximately 0.5% per year thereafter, resulting in a cumulative incidence of 4% to 5% during the first 5 years after valve implantation.[49-51] Infection may be introduced at the time of valve placement or from transient bacteremia at any time thereafter. The overall risks of infection are similar for mechanical and porcine bioprosthetic valves and for aortic and mitral valve prostheses.[49-51] The leading cause of PVE during the first year after surgery is methicillin-resistant coagulase-negative staphylococci, predominantly *S. epidermidis* [*see Table 2*]. Coagulase-negative staphylococci continue to cause cases of PVE that occur a year or more after surgery; however, these staphylococci are often species other than *S. epidermidis*, and only 20% to 30% are methicillin resistant. *S. aureus*, streptococci, enterococci, and fastidious gram-negative coccobacilli, the leading organisms associated with native valve endocarditis, cause about three quarters of PVE cases after 1 year following valve replacement.[49-51] Coagulase-negative staphylococci are responsible for at least 35% of cases of PVE and, therefore, should not be dismissed as contaminants (which they usually are in other settings) if isolated from the blood of a patient who has a prosthetic valve.

Infection of prosthetic valves is often associated with invasion of perivalvular tissues—resulting in valve-ring abscessses and valvular dysfunction—and occasionally with myocardial abscesses.[41,49,52,53] Necrosis of the annulus from invasive infection can cause partial dehiscence of the prosthesis, resulting in hemodynamically significant paravalvular regurgitation. These pathologic changes can occur with either porcine or mechanical prostheses, particularly when the valves are in the aortic position or when infection occurs during the first postoperative

Table 2 Etiology of Prosthetic Valve Endocarditis[49,90,110,111]

Organism	*Time of Onset after Valve Implantation and Percentage of Cases*		
	≤ 2 Months	*2–12 Months*	*> 12 Months*
Coagulase-negative staphylococci	54	56	15
Staphylococcus aureus	8	9	13
Gram-negative bacilli	12	3	1
Streptococci	rare	3	34
Enterococci	rare	6	11
Corynebacteria	8	rare	1
HACEK	rare	3	14
Fungi	6	6	3
Miscellaneous	6	6	1
Culture-negative	5	9	5

year.[51,54] Occasionally, vegetations may partially obstruct the valve orifice or restrict valve movement, causing functional stenosis. Such changes are more likely with prostheses in the mitral position. When infection is restricted to the leaflets of a porcine bioprosthetic valve, leaflet destruction, obstructing vegetations, and the delayed onset of leaflet stiffness may cause clinically significant valvular dysfunction.[54]

The dominant clinical feature of PVE that occurs during the first 60 days after surgery (early PVE) is fever, whether or not there is a regurgitant murmur associated with the prosthetic valve. Prosthetic valve dysfunction with resulting heart failure is seen in some patients. Petechiae occur in about half of patients with early PVE, but Roth spots, Osler nodes, and Janeway lesions are not common. Emboli are common in early PVE; emboli that occlude large peripheral arteries suggest fungal endocarditis. Because blood cultures are often negative in patients with fungal endocarditis, this diagnosis must often be made on the basis of histologic examination, culture, and sometimes molecular testing of surgical or autopsy specimens or vegetation recovered at embolectomy.

The clinical features of PVE that occurs more than 60 days after surgery (late PVE) are similar to those of acute or subacute endocarditis on native valves, depending on the infecting organism.

Diagnosis

DUKE CRITERIA

Although the diagnosis of endocarditis may be readily evident in patients with the classic syndrome of fever, a murmur associated with valvular dysfunction, typical peripheral signs, and bacteremia, the diagnosis is less evident in most patients.[55] A diagnostic approach known as the Duke criteria has been designed by the Duke Endocarditis Service[56] [*see Tables 3 and 4*]. With these criteria, the diagnosis of endocarditis can be established definitively by either pathologic or clinical criteria. The pathologic criteria include direct evidence gleaned from surgery or autopsy, and the clinical criteria are derived from microbiologic data (culture, serologic or molecular testing), echocardiography, physical examination, and other laboratory findings [*see Table 4*]. In retrospective evaluations of pathologically proven cases of native valve endocarditis and PVE, these clinical criteria have been found to be both sensitive[56-58] and highly specific.[57,59,60] Misdiagnoses rarely resulted when these clinical criteria were

Table 3 Duke Criteria for the Diagnosis of Infective Endocarditis[52,72,112,113]

Definite

Pathologic criteria

1. Microorganisms: demonstrated by culture, Gram stain, histologically, or by a validated molecular test in a vegetation, in a vegetation that has embolized, or in an intracardiac abscess specimen

or

2. Pathologic lesions: vegetation or intracardiac abscess confirmed by histologic examination showing active endocarditis

Clinical criteria*

1. Two major criteria

or

2. One major and three minor criteria

or

3. Five minor criteria

Possible

Findings consistent with infective endocarditis that fall short of "definite" but are not "rejected"

Rejected

1. Firm alternative diagnosis for manifestations of infective endocarditis

or

2. Resolution of endocarditis syndrome with antibiotic therapy for ≤ 4 days

or

3. No pathologic evidence of infective endocarditis at surgery or autopsy, with antibiotic therapy for ≤ 4 days

*See Table 4 for definitions of major and minor criteria.

Table 4 Definitions of Terms Used in Duke Criteria for the Diagnosis of Infective Endocarditis[56]

Major criteria

1. Positive blood cultures for the following:

Typical microorganism consistent with diagnosis from two separate blood cultures

a. Viridans streptococci, *Streptococcus bovis*, or HACEK organisms

or

b. Community-acquired *Staphylococcus aureus* or enterococci, in the absence of a primary focus

or

Persistently positive blood cultures, defined as microorganisms consistent with diagnosis from the following:

a. At least two blood samples drawn > 12 hr apart

or

b. Three of three or a majority when more than three blood cultures are drawn, with first and last samples drawn at least 1 hr apart

2. Evidence of endocardial involvement

Positive echocardiogram for IE

a. Oscillating intracardiac mass on valve or supporting structures, in the path of regurgitant jets, or on implanted material, in the absence of an alternative anatomic explanation

or

b. Abscess

or

c. New partial dehiscence of prosthetic valve

or

New valvular regurgitation (changing of preexisting murmur not sufficient)

Minor criteria

1. Predisposition: predisposing heart condition or I.V. drug use
2. Fever: temperature ≥ 38° C (100.4° F)
3. Vascular phenomena: major arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial hemorrhage, conjunctival hemorrhages, Janeway lesions
4. Immunologic phenomena: glomerulonephritis, Osler nodes, Roth spots, rheumatoid factor
5. Microbiologic evidence: positive blood culture but not meeting a major criterion (see above)* or serologic evidence of active infection with organism consistent with IE
6. Echocardiogram consistent with infective endocarditis but not meeting a major criterion (see above)

*Excludes single positive cultures for coagulase-negative staphylococci, diphtheroids, and organisms that do not commonly cause endocarditis.
IE—infective endocarditis

used to reassess previously pathologically confirmed cases. The specificity and negative predictive values of these criteria have been reported as 99% and 92%, respectively.[59,60] Use of these clinical criteria very rarely results in the rejection of cases considered to be endocarditis on independent expert evaluation.[61]

ECHOCARDIOGRAPHY

The Duke criteria utilize echocardiography for defining the anatomic features of endocarditis. Both the sensitivity and specificity of echocardiography are high when experienced echocardiographers apply specific criteria. In patients with pathologically proven native valve endocarditis, vegetations can be detected in 60% to 75% by use of transthoracic echocardiography (TTE) and in 87% to 94% with transesophageal echocardiography (TEE); the specificity of both techniques is high.[62-64] TEE is notably superior to TTE in the evaluation of patients with suspected PVE[65] [*see* Prosthetic Valve Endocarditis, *above*]. Notwithstanding the increased sensitivity of TEE compared with TTE, if the decision is made to treat all cases identified by the Duke criteria as definite endocarditis and most cases identified as possible endocarditis, then evaluation by TEE rarely alters the treatment decision. Occasional exceptions to this observation are PVE cases that were missed by TTE. Echocardiography is the preferred technique for identification of perivalvular infection and other intracardiac complications of endocarditis. For detection of abscesses, a transesophageal study is significantly more sensitive (76% to 87%) than transthoracic imaging (18% to 28%), with equal specificity.[53,65,66]

All patients in whom endocarditis is seriously suspected should undergo echocardiography. However, echocardiography should not be regarded as a general screening test for patients with a low prior probability of endocarditis (e.g., most patients with acute febrile illnesses). Given its high sensitivity, a negative transesophageal echocardiogram is good evidence against endocarditis in patients at low or intermediate risk of this infection. Conversely, if the prior probability of endocarditis is high, a negative echocardiogram does not fully exclude the diagnosis; the false negative rate for transesophageal echocardiography is 6% to 13%.[67] A repeat transesophageal study in patients with endocarditis reduces the false negative results to approximately 5%.[67] Finally, echocardiography cannot reliably distinguish an infected vegetation from a sterilized vegetation, nonbacterial thrombotic vegetations of marantic endocarditis, or intracardiac thrombi, as seen in the antiphospholipid antibody syndrome.

Transthoracic echocardiography has limited usefulness in the diagnosis of PVE because the prosthesis itself produces echoes that often obscure vegetations and abscesses. Transesophageal two-dimensional and Doppler echocardiography more effectively assess prosthetic valves and perivalvular tissues, especially when a mitral valve prosthesis is present.[68,69] In PVE, the transesophageal technique identified vegetations in 82% of patients, compared with a 36% identification rate with transthoracic echocardiography.[53] Similarly, detection of paravalvular abscesses in patients with PVE is markedly increased by use of transesophageal rather than transthoracic echocardiography.[68]

BLOOD CULTURE

For patients suspected of having endocarditis, identification of the microbial etiologic agent and determination of its antimicrobial susceptibility are of paramount importance for both diagnosis and treatment. In most patients with SBE, blood cultures drawn before initiation of antibiotic therapy will all be positive, reflecting the sustained bacteremia associated with an infected endothelial surface. To optimize the value of blood cultures, best-practice guidelines recommend careful antiseptic skin preparation to minimize contamination of the specimen with skin flora: cleaning, followed by application of 70% isopropyl alcohol, which is allowed to dry, followed by application of an iodophor or chlorhexidine, which is also allowed to dry.[70] Blood should be drawn using three separate venipunctures, at three different sites, that are taken over several hours if the patient presents with a subacute syndrome or that are taken over a few minutes if acute endocarditis is suspected. The ideal volume for each specimen is 12 to 20 ml, which should be divided equally into two culture bottles, yielding six bottles in all. Obtaining a large volume of blood for culture maximizes the yield,[70] but smaller volumes must be accepted from infants and small children. Use of both aerobic and anaerobic broth media also helps maximize yield. If the patient has recently received any antibiotic treatment, use of a medium that contains an antibiotic removal device such as resin will increase the yield by 5% to 15%.[70] Blood cultures should be incubated for 5 days, because modern improved culture media will yield growth for the vast majority of relevant microorganisms within this period. This is true even for most of the so-called fastidious or slow-growing bacteria such as the HACEK group and the nutritionally variant streptococci (*Abiotrophia* species).

Culture-Negative Endocarditis

Blood cultures are negative in 5% to 20% of patients with infective endocarditis[9,71] [*see Tables 1 and 2*]. In about half of these cases, blood cultures are negative as a consequence of previous antimicrobial therapy, even though the vegetation is still infected. Recovery of the causative organism from the blood of these previously treated patients may often be accomplished by repeating blood cultures several days after antibiotics have been discontinued, but some patients remain persistently blood culture negative. Blood cultures may also be negative in some cases of right-sided endocarditis caused by relatively noninvasive organisms. Currently, however, most cases of right-sided endocarditis occur in IDUs and are caused by pyogenic bacteria, such as *S. aureus* and *P. aeruginosa*, which are readily isolated from the blood. Culture of bone marrow or arterial blood does not provide materially more information than can be obtained from culture of venous blood.

Several additional factors can thwart isolation of the infecting agent from blood of patients with infective endocarditis. *Bartonella* species may require more than 5 days for isolation, and certain mycelial fungi such as *H. capsulatum* and *Aspergillus* species are difficult to isolate even with special techniques such as lysis centrifugation. Finally, isolation of *C. burnetii* and *C. psittaci* requires techniques beyond the capabilities of most laboratories.[13,72,73] Endocarditis caused by *Bartonella* species, *Legionella* species, *Brucella* species, *C. psittaci*, and *C. burnetii* can be presumptively diagnosed with serologic tests. These tests should be performed in the evaluation of apparently culture-negative endocarditis, particularly when negative cultures cannot be attributed to prior antimicrobial therapy.[73] Direct cultures on special media, histopathologic examination with special stains, and molecular techniques to recover DNA or 16S ribosomal RNA all can be used to determine etiology from examination of vegetations that have been removed from valves or, having embolized, from peripheral arteries.[73-76]

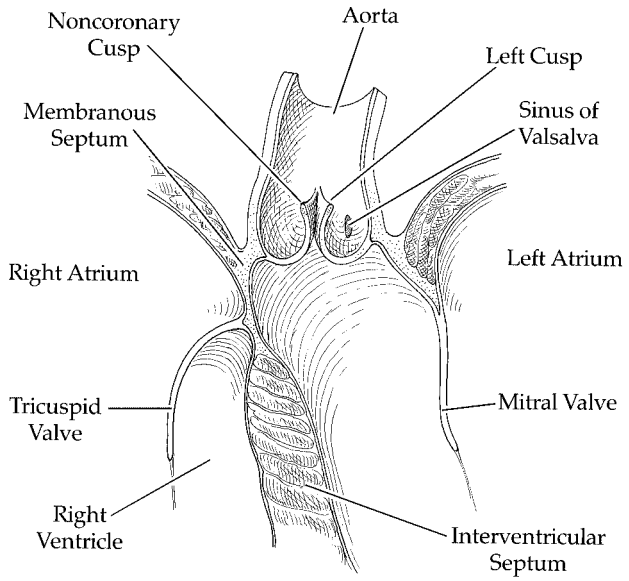


Figure 2 **Anatomic relations between the noncoronary cusp and left cusp of the aortic valve, interventricular septum, membranous septum, tricuspid valve, and mitral valve are shown schematically at the level of the aortic root.**

Bacteremia in Patients with Prosthetic Heart Valves

As in native valve endocarditis, the diagnosis of PVE is usually based on history, physical examination, echocardiography, and the detection of bacteremia. However, only a subgroup of patients with a prosthetic valve who experience bacteremia subsequently develop PVE. Of bacteremic patients with a prosthetic valve whose initial positive blood culture is not sentinel evidence of active PVE itself, only 15% to 20% will develop PVE caused by that blood culture isolate.[17]

Transient bacteremia with gram-negative bacilli from extracardiac sources usually does not result in colonization of prosthetic valves; however, recrudescent or sustained bacteremia that occurs after the extracardiac focus of infection has been eradicated suggests PVE, as does persistent gram-negative bacillary bacteremia with no identifiable extracardiac source, even if a new regurgitant murmur or other signs of endocarditis are lacking. Persistent or high-grade coagulase-negative staphylococcal bacteremia in a patient with a prosthetic valve strongly suggests PVE. Similarly, PVE is likely when coagulase-negative staphylococci that have been isolated sporadically from multiple blood cultures are shown by molecular techniques to belong to a single clone. Blood cultures are negative in about 6% of PVE cases. Negative cultures are usually the result of previous antibiotic therapy or reflect unique characteristics of the infecting organisms.

ADJUNCTIVE LABORATORY TESTS

Results of many laboratory tests are likely to be abnormal in patients with endocarditis because of the systemic impact of the infection and because of injury to various organs. Such tests include the complete blood count, urinalysis, blood urea nitrogen level, creatinine concentration, rheumatoid factor, quantitative immunoglobulins, complement levels, erythrocyte sedimentation rate (ESR), and C-reactive protein level. These adjunctive tests may offer clues to the diagnosis and may be useful for monitoring the progress of treatment for endocarditis, but they usually are not specific and are not critical in establishing a diagnosis; hence, they are not among the Duke diagnostic criteria.

DIAGNOSTIC FEATURES OF CARDIAC COMPLICATIONS

Heart failure is the most frequent cardiac complication of infective endocarditis and may result from a variety of factors. Preexisting valvular disease can be worsened by the effects of infective endocarditis, which can include tears, perforations, and obstruction of valves and rupture of chordae tendineae. These complications can also affect previously normal valves. Damage to the aortic valve by infection can cause rapidly progressive and severe hemodynamic impairment, more so than comparable damage to the mitral valve. Coronary artery embolism can cause silent or overt myocardial infarction, which can contribute to heart failure. A mycotic aneurysm of a sinus of Valsalva or an aortic annulus abscess may rupture through the membranous septum into the right atrium or ventricle [*see Figure 2*]. Flow through the resulting fistula causes a sudden rise in the jugular venous pressure and a continuous or to-and-fro murmur and thrill along the left sternal border. The lungs remain relatively clear. Valvular damage and subsequent dysfunction, as well as intracardiac fistula formation, can be accurately defined with two-dimensional and Doppler echocardiography from a transthoracic or transesophageal approach.

The development of new conduction abnormalities may signal the extension of infection into the septum, affecting its conduction tissues.[33,52] PR interval prolongation, left bundle branch block, or right bundle branch block with left anterior hemiblock suggests the extension of infection from the aortic valve. The proximity of the weakest area of the aortic valve annulus to the membranous septum and the conduction system accounts for the development of these conduction abnormalities [*see Figure 3*]. Similarly, extension of infection from the mitral annulus, which is close to the bundle of His and to the atrioventricular node, may also produce conduction defects, but such extension occurs less frequently than extension from the aortic valve. In the absence of digitalis toxicity or a recent inferior myocardial infarction, the development of nonparoxysmal junctional tachycardia, a Wenckebach block, or complete heart block with a narrow QRS complex serves as a clue to the spread of infection from the mitral annulus into the AV node and proximal bundle of His. Ventricular premature beats in patients with endocarditis who do not have electrolyte abnormalities or digitalis toxicity may reflect myocarditis, myocardial abscesses, or coronary arterial emboli.

The presence of an annular abscess may be suggested by the recent onset of aortic regurgitation, persistent fever during appropriate antimicrobial therapy, and the development of pericarditis.[77,78] Pericarditis caused by extension of a valve ring abscess into the epicardium is more often hemorrhagic or fibrinous than purulent. Occasionally, pericarditis in the course of infective endocarditis is the result of transmural myocardial infarction secondary to coronary emboli. Transesophageal echocardiography is the most sensitive and preferred noninvasive test for detecting valve-ring abscesses in both native valve endocarditis and PVE.

Differential Diagnosis

The possibility of infective endocarditis should be considered in any patient with a heart murmur and fever. The physician must be particularly alert for atypical cases in which the clinical

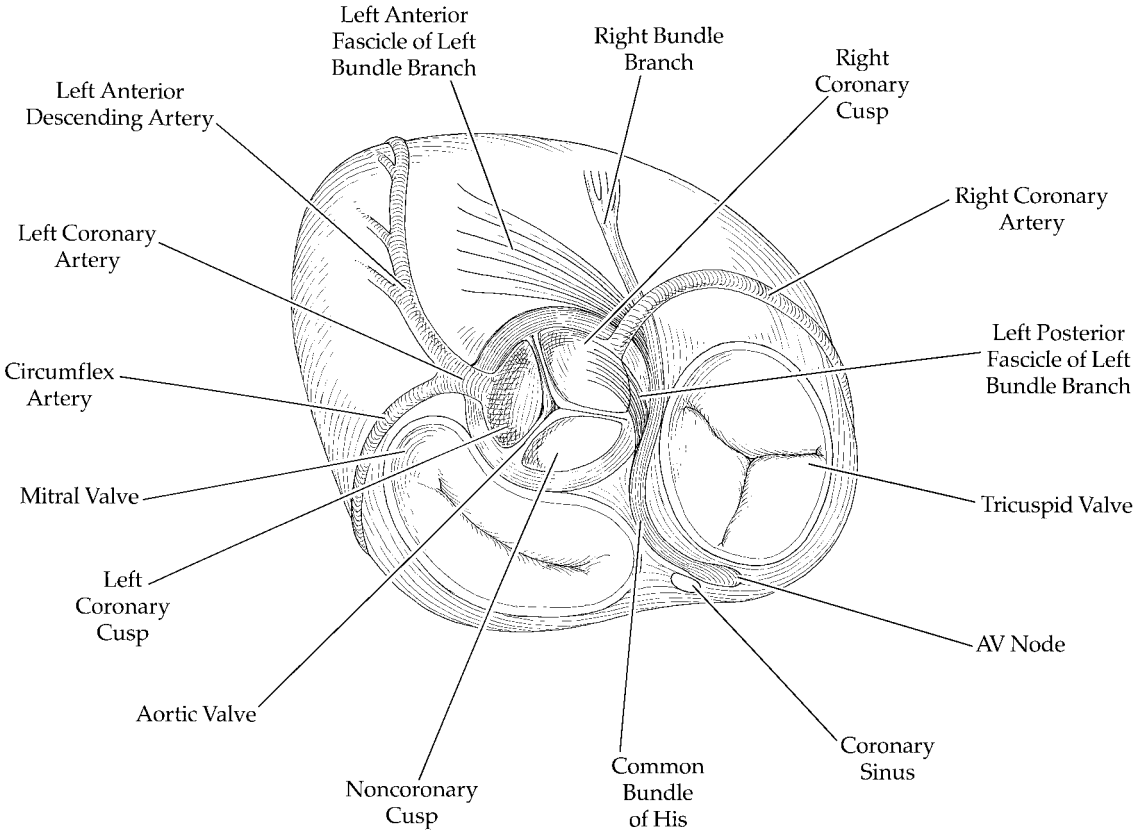


Figure 3 **The close relation of the three cardiac valves and the cardiac conduction system, as seen in this superior schematic view, accounts for the appearance of conduction defects in endocarditis. (AV—atrioventricular)**

findings reflect complications of endocarditis affecting organs other than the heart.

Infective endocarditis can cause fever of undetermined origin (FUO), so the differential diagnosis includes the many other infections that may cause FUO. These include tuberculosis, salmonellosis, and various intra-abdominal and genitourinary infections [*see 7:XXIV Hyperthermia, Fever, and Fever of Undetermined Origin*].

A variety of noninfectious illnesses can mimic infective endocarditis, including immune-mediated diseases and rheumatologic conditions such as juvenile rheumatoid arthritis and polymyalgia rheumatica. Acute rheumatic fever can cause fever, cardiac murmurs, and heart failure, but acute rheumatic fever can be distinguished from infective endocarditis on clinical grounds and by negative blood cultures, raised anti-streptolysin O antibody titer, and response to salicylates [*see 7:I Infections Due to Gram-Positive Cocci*]. Marantic endocarditis, which can give rise to multiple embolic episodes and fever, is usually associated with an underlying neoplasm or chronic wasting disease. In polyarteritis nodosa, the presence of fever, anemia, and renal involvement may suggest SBE, and the findings on biopsy of a lesion in a large artery may even resemble those of a mycotic aneurysm. In both systemic lupus erythematosus and antiphospholipid antibody syndrome, the manifestations of fever, nonbacterial thrombotic vegetations, systemic emboli, and spontaneous thrombotic events can simulate infective endocarditis.

A cardiac myxoma, usually in the left atrium, may mimic infective endocarditis in both clinical and laboratory features, including low-grade fever, weight loss, arthralgias, cutaneous lesions, clubbing of the fingers, emboli to major arteries, and auscultatory findings suggesting mitral stenosis and regurgitation. Cardiac myxoma syndrome may further simulate infective endocarditis by giving rise to cerebral aneurysms at the sites of myxomatous emboli. Negative blood cultures and echocardiography can help establish the correct diagnosis.

Neoplasms may mimic infective endocarditis by inducing marantic endocarditis or by their hemodynamic effects. For example, richly vascular tumors may be associated with fever, anemia, and hyperdynamic circulation with flow murmurs. A left renal tumor mass may be mistaken for splenomegaly. Carcinoid tumors occasionally mimic endocarditis when they produce endocardial and valvular fibrosis that lead to tricuspid insufficiency and pulmonary stenosis.

Treatment

Two major modalities are used to treat endocarditis: (1) antibiotic therapy and (2) surgical debridement of vegetations and infected perivalvular tissue, with valve repair or replacement as needed.

ANTIMICROBIAL TREATMENT

Effective antimicrobial treatment requires identification of the etiologic agent and determination of its antimicrobial susceptibility. Therefore, in the evaluation of a patient with subacute or indolent disease, it is usually best to delay antibiotic therapy until the results of blood cultures are obtained. If recently administered antibiotics have rendered the initial cultures negative, this

delay provides an opportunity to obtain additional blood cultures after the antibiotics and their effects have dissipated. However, if the infection is fulminant or if there is valvular dysfunction that may require urgent surgical intervention, empirical antibiotic therapy must be initiated promptly after blood culture specimens have been obtained. Bactericidal antibiotics are used parenterally in high doses. With the exception of PVE caused by staphylococci, antimicrobial therapy for PVE caused by a specific organism utilizes the same drugs recommended for native valve endocarditis. However, therapy is usually administered over a longer period, typically 6 weeks. Patients must be evaluated frequently to assess the efficacy of antimicrobial therapy and the development of complications of therapy or infection.

Streptococci

Viridans streptococci, once mostly penicillin sensitive, have demonstrated increasing resistance to penicillin and the cephalosporins over the past 20 years.[79] Accordingly, in the planning of endocarditis treatment, all streptococci must be evaluated for susceptibility to penicillin by determining the minimum inhibitory concentration (MIC). Various regimens provide effective treatment for endocarditis caused by those streptococci that are fully penicillin sensitive (MIC < 0.2 μg/ml) [*see Table 5*].[80] One regimen employs parenteral penicillin alone in high doses for 4 weeks. A second regimen utilizes the synergism achieved against most strains of nonenterococcal streptococci by the combination of penicillin and gentamicin. This synergism allows effective treatment with only 2 weeks of combination therapy.[80,81] The short-duration regimen should be considered only for selected cases of native valve endocarditis with favorable prognostic features: streptococcal infections that are not complicated by hypotension, renal failure, thrombocytopenia, mycotic aneurysms, or heart failure caused by valvular dysfunction. The etiologic species should be highly susceptible to penicillin (MIC < 0.2 μg/ml) and should not be a nutritionally variant strain—that is, dependent upon pyridoxal or cysteine for growth. PVE should not be treated with the 2-week regimen. Ceftriaxone, given in a single daily intravenous dose for 4 weeks, is now often recommended for treatment of endocarditis caused by penicillin-sensitive streptococci because this regimen is easily adapted for outpatient treatment and has little toxicity.[49,80] Although short-duration combination therapy for penicillin-susceptible streptococcal endocarditis using single daily doses of both ceftriaxone and an aminoglycoside has been successful, experience with this regimen is limited, and the regimen is not recommended for general use. *S. bovis* is highly penicillin sensitive and can be treated with regimens recommended for other penicillin-sensitive streptococci. When combination penicillin-gentamicin therapy is used to treat endocarditis caused by nonenterococcal streptococci, particularly when short-duration therapy is planned, the streptococcus should be screened for high-level resistance to gentamicin. Although rare in these organisms, high-level resistance would preclude bactericidal synergy and thus indicate the need for an alternative regimen.

Endocarditis caused by relatively penicillin-resistant (MIC = 0.2 to 0.5 μg/ml) viridans or other nonenterococcal streptococci is treated with a higher dose of penicillin G, combined with gentamicin. If the strain is even more resistant to penicillin (MIC > 0.5 μg/ml), the infection is treated with one of the standard regimens for enterococcal endocarditis [*see Table 5*]. Nutritionally variant streptococci (previously called *S. adjacens* or *S. defectivus;* now named *Abiotrophia* species) are often relatively penicillin resistant, so endocarditis caused by these organisms should be treated with the standard regimen for enterococcal endocarditis.[49,80,82] Endocarditis caused by pneumococci or group A streptococci is treated with intravenous penicillin G in a dosage of 20 million units daily for 4 weeks.[49,80] Pneumococci that are found to be the cause of endocarditis must be tested for susceptibility to penicillin. Vancomycin is the preferred treatment for endocarditis caused by penicillin-resistant strains with an MIC greater than 1.0 μg/ml. When pneumococcal endocarditis is complicated by concurrent meningitis, the treatment regimen must ensure adequate penetration of antibiotic into cerebrospinal fluid.[83] The β-hemolytic streptococci belonging to groups B, C, and G have slightly reduced susceptibility to penicillin; for endocarditis, they should be treated as if they were relatively penicillin resistant, with MICs of 0.2 to 0.5 μg/ml.[80,84]

Enterococci

Enterococci are relatively resistant to penicillin, ampicillin, and vancomycin and are fully resistant to cephalosporins. Antibacterial synergism is essential for optimal antimicrobial treatment of enterococcal endocarditis. To achieve this, the enterococcus must simultaneously be exposed to a cell wall–active antibiotic such as penicillin, ampicillin, or vancomycin, at a concentration at or above the organism's MIC, and to an aminoglycoside that will exert a lethal effect.[49,80,85] The ability of some enterococci to grow in the presence of gentamicin at concentrations of 500 μg/ml or higher indicates high-level aminoglycoside resistance; against such organisms, combination therapy will fail to exert a lethal effect regardless of the cell wall–active antimicrobial agent employed. High-level resistance to gentamicin is the consequence of aminoglycoside-modifying enzymes.

Previously, synergistic bactericidal therapy could be reliably anticipated when gentamicin was combined with penicillin, ampicillin, or vancomycin. This provided regimens for treatment of enterococcal endocarditis[49,80] [*see Table 5*].

Currently, antimicrobial resistance in enterococci presents a complex problem that must be carefully considered in the selection of therapy for enterococcal endocarditis.[85] The causative strain must be screened for high-level resistance to gentamicin. Use of gentamicin in combination therapy in the face of high-level resistance exposes the patient to potential toxicity and is without therapeutic benefit. Furthermore, resistance to cell wall–active agents has become increasingly prevalent in enterococci. Intrinsic resistance to penicillin and ampicillin (MIC ≥ 32 μg/ml) is prevalent in *E. faecium*. Penicillin and ampicillin resistance caused by β-lactamase production, which is not detectable with MIC tests but requires screening with the chromogenic cephalosporin nitrocefin, is occasionally seen in *E. faecalis*. Finally, vancomycin resistance (MIC ≥ 16 μg/ml) is being encountered increasingly in *E. faecalis* and *E. faecium*. If an enterococcus is resistant to a cell wall–active agent, that agent cannot participate in the synergistic killing of the strain. Vancomycin is a suitable cell wall–active agent for combination therapy when organisms have intrinsic or β-lactamase–mediated resistance to penicillin or ampicillin; ampicillin-sulbactam is suitable when enterococci have resistance that is mediated by β-lactamase. Enterococci that are resistant to vancomycin may be susceptible to penicillin and ampicillin but more often are resistant to these antibiotics as well. For a few of these enterococci, teicoplanin, a glycopeptide antibiotic that has not been approved for use in the United States but is available elsewhere, remains an effective cell wall–active antimicrobial.[80,85] Endocarditis caused by van-

Table 5 Antimicrobial Therapy for Endocarditis in Adults[49,80,114]

Infectious Agent	*Drug*	*Dosage and Route of Administration*	*Duration of Therapy* (wk)*
Penicillin-susceptible viridans and other non-enterococcal streptococci (minimum inhibitory concentration [MIC] < 0.2 μg/ml)	*Preferred Regimen*		
	Penicillin G	12–18 million units I.V. daily (in divided doses q. 4 hr)	4
	or		
	Penicillin G	12–18 million units I.V. daily (in divided doses q. 4 hr)	2
	plus		
	gentamicin	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	2
	or		
	Ceftriaxone	2 g I.V. daily (as a single dose)	4
	Alternative Regimen‡		
	Vancomycin	30 mg/kg I.V. daily (in divided doses q. 12 hr)	4
Relatively penicillin-resistant streptococci			
	Preferred Regimen		
MIC 0.2–0.5 μg/ml	Penicillin G	20–30 million units I.V. daily (in divided doses q. 4 hr)	4
	plus		
	gentamicin§	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	2
MIC > 0.5 μg/ml	Penicillin G plus gentamicin	Dosages same as in previous regimen	4
	Alternative Regimen‡		
	Vancomycin§	30 mg/kg I.V. daily (in divided doses q. 12 hr)	4
Enterococci‖	*Preferred Regimen*		
	Penicillin G	20–30 million units I.V. daily (in divided doses q. 4 hr)	4–6
	plus		
	gentamicin	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	4–6
	or		
	Ampicillin	12 g I.V. daily (in divided doses q. 4 hr)	4–6
	plus		
	gentamicin	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	4–6
	Alternative Regimen‡		
	Vancomycin	30 mg/kg I.V. daily (in divided doses q. 12 hr)	4–6
	plus		
	gentamicin	3–5 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	4–6
Staphylococci (methicillin susceptible) in the absence of prosthetic material	*Preferred Regimen*		
	Nafcillin or oxacillin	12 g I.V. daily (in divided doses q. 4 hr)	4–6
	plus		
	gentamicin		
	(optional; see text)	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	3–5 days
	Alternative Regimens‡		
	Cefazolin	12 g I.V. daily (in divided doses q. 4 hr)	4–6
	plus		
	gentamicin		
	(optional; see text)	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)†	3–5 days
	or		
	Vancomycin	30 mg/kg I.V. daily (in divided doses q. 12 hr)	4–6
Staphylococci (methicillin resistant) in the absence of prosthetic material	Vancomycin	30 mg/kg I.V. daily (in divided doses q. 12 hr)	4–6
Staphylococci (methicillin susceptible) in the presence of prosthetic material	Nafcillin or oxacillin	12 g I.V. daily (in divided doses q. 4 hr)	6–8
	plus		
	rifampin	300 mg p.o., q. 8 hr	6–8
	plus gentamicin#	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)	2
Staphylococci (methicillin resistant) in the presence of prosthetic material	Vancomycin	30 mg/kg I.V. daily (in divided doses q. 12 hr)	6–8
	plus		
	rifampin	300 mg p.o., q. 8 hr	6–8
	plus gentamicin#	3 mg/kg I.M. or I.V. daily (in divided doses q. 8 hr)	2
HACEK organisms	Ceftriaxone**	2 g I.V. daily (as a single dose)	4

*Treatment programs are longer for prosthetic valve endocarditis (6–8 wk).
†Peak serum gentamicin concentration should be approximately 3 μg/ml and trough concentrations < 1 μg/ml. Some authorities prefer a gentamicin dosage of 1.5 mg/kg q. 8 hr, which results in peak concentrations of approximately 5 μg/ml.
‡Alternative regimen is for use in patients with a history of penicillin hypersensitivity.
§Some investigators recommend omission of gentamicin if the MIC is > 0.2 but < 0.5 μg/ml; the role of gentamicin in combination with vancomycin has not been fully established.
‖Enterococci have become increasingly antibiotic resistant (see text); clinical isolates must be fully evaluated for microbial resistance to select optimal therapy.
#Administer during the initial 2 weeks (see text).
**Cefotaxime or another third-generation cephalosporin at a comparable dosage could be used.

comycin-resistant enterococci that are not susceptible to ampicillin requires special, individualized regimens, chosen with the assistance of an infectious disease consultant. All enterococci that cause endocarditis must be tested for susceptibility to antimicrobials that have therapeutic potential. Thereafter, a bactericidal synergistic combination of a cell wall–active agent (e.g., penicillin, ampicillin, vancomycin, ampicillin-sulbactam, or teicoplanin) plus gentamicin can be selected. Single daily dosing of gentamicin should not be used in treating enterococcal endocarditis. If a synergistic combination is not possible because of high-level resistance to gentamicin, prolonged therapy (8 to 12 weeks) with high doses of an effective cell wall–active agent should be administered. To prevent or minimize possible toxicity of gentamicin or vancomycin, it is important to monitor serum levels, renal function, and otologic symptoms every 3 to 5 days throughout therapy.[49,80]

Bactericidal synergistic therapy for enterococcal endocarditis is associated with cure rates approaching 85%. Treatment with an effective cell wall–active agent alone results in cure rates of 40% to 50% at best.[80,86] If bactericidal synergistic therapy is not available, patients with enterococcal endocarditis in whom treatment with a cell wall–active agent alone has failed should undergo excision of the infected valve while suppressive antibiotic therapy is continued.[49,80,85,86]

Staphylococci

Staphylococcus aureus The treatment of choice for *S. aureus* native valve endocarditis is a penicillinase-resistant penicillin such as nafcillin or oxacillin[49,80] [*see Table 5*]. Vancomycin provides an alternative treatment for patients who are hypersensitive to penicillins. Vancomycin, although active against both methicillin-sensitive *S. aureus* (MSSA) and MRSA, is not the first choice, because it is somewhat less effective than the penicillinase-resistant penicillins.[87] However, vancomycin should be added initially because of the increasing frequency of MRSA strains, not only in nosocomial infections but also in community-acquired infections. If the blood culture yields MSSA, the vancomycin should be discontinued; if MRSA is yielded, then vancomycin alone is continued[80] [*see Table 5*]. Penicillin G, 20 to 30 million units daily, should be substituted for the few strains of *S. aureus* that prove to be penicillin sensitive, but only after susceptibility has been confirmed by tube dilution testing and the isolate has been demonstrated not to produce penicillinase. Gentamicin is synergistic with penicillinase-resistant penicillins against *S. aureus* in vitro, and combination therapy may reduce the duration of bacteremia and fever in patients. However, the in vivo benefit is marginal; combination therapy does not improve overall cure rates. If the option to add gentamicin is chosen, it should be stopped after about 3 days to minimize the risk of nephrotoxicity.

Rare strains of *S. aureus* that are partially or even fully resistant to glycopeptides such as vancomycin have been detected. In the unlikely event of endocarditis being caused by such a strain, alternative drugs to be considered include linezolid, quinupristin-dalfopristin, or teicoplanin and daptomycin.

Recommended regimens for treatment of endocarditis in IDUs are usually the same as regimens for treatment in other patients. However, in IDUs who have MSSA endocarditis restricted to the tricuspid valve, without complicating metastatic infection, successful treatment has been accomplished with a penicillinase-resistant penicillin plus an aminoglycoside (e.g., nafcillin plus gentamicin or tobramycin) given for only 2 weeks.[88] Such patients who do not respond promptly and completely to this short-term regimen should receive longer courses of treatment.

Coagulase-negative staphylococci Endocarditis caused by coagulase-negative staphylococci is usually engrafted on a prosthetic valve. More than 80% of the coagulase-negative staphylococci that cause PVE within 1 year after valve implantation are resistant to methicillin and to all β-lactam antibiotics, including cephalosporins.[49,80] Only when the onset of this form of PVE occurs more than a year after surgery does the frequency of β-lactam resistance in the coagulase-negative staphylococci fall below 30%.[89,90] Moreover, the β-lactam resistance may not be detected by standard antimicrobial susceptibility testing, particularly when automated instruments or microsystems are used. Most of these coagulase-negative staphylococci are sensitive to vancomycin and rifampin, but plasmid-mediated resistance to gentamicin is prevalent. For PVE caused by methicillin-resistant organisms that are susceptible to rifampin and gentamicin, the most effective treatment combines vancomycin with rifampin and gentamicin [*see Table 5*].[49,80,90] If the infecting strain is resistant to gentamicin, either a fluoroquinolone or another aminoglycoside to which the strain is susceptible should be substituted for gentamicin. If the staphylococci are susceptible to methicillin, a penicillinase-resistant penicillin should be used in lieu of vancomycin [*see Table 5*]. Antimicrobial therapy for PVE caused by *S. aureus* should utilize the same regimens as those used for coagulase-negative staphylococcal PVE [*see Table 5*].

Rigorous susceptibility testing is required, but the coagulase-negative staphylococci that cause native valve endocarditis are usually sensitive to β-lactam antibiotics unless the infection has been acquired nosocomially. Native valve endocarditis caused by coagulase-negative staphylococci can be treated with the same regimens that are employed to treat native valve endocarditis caused by *S. aureus*.[49,80]

HACEK Organisms

Some HACEK organisms that cause endocarditis have been found to produce β-lactamase, which makes those organisms resistant to ampicillin. Accordingly, endocarditis caused by HACEK organisms should be treated with a third-generation cephalosporin for 4 weeks [*see Table 5*].[49,80]

Corynebacteria (Diphtheroids)

The combination of penicillin and gentamicin acts synergistically in vitro to kill *Corynebacterium* species (diphtheroids) that are susceptible to gentamicin (MIC ≤ 4 μg/ml). Synergism is not achieved against strains that are resistant to gentamicin. For PVE caused by gentamicin-sensitive strains, therapy with intravenous penicillin G (20 million units daily) plus gentamicin (1 mg/kg I.M. or I.V. every 8 hours) is recommended. Vancomycin, which is bactericidal against diphtheroids, is recommended as the initial treatment of diphtheroid PVE, with subsequent treatment determined on the basis of antibiotic-susceptibility test results. Vancomycin is also the drug of choice if the infecting strain is resistant to gentamicin or if the patient is allergic to penicillin.

Gram-Negative Bacilli

Endocarditis caused by gram-negative bacilli should be treated with a potent β-lactam antibiotic, such as a third-generation cephalosporin, with or without an aminoglycoside, depending on the susceptibility of, and the published experience with, the

specific gram-negative bacillus involved. Effective treatment of endocarditis caused by *P. aeruginosa* requires the synergistic combination of ticarcillin or piperacillin (either drug at a dosage of 3 g I.V. every 4 hours) and tobramycin (2.7 mg/kg I.M. or I.V. every 8 hours; 8 mg/kg daily) for at least 6 weeks. Because pseudomonal endocarditis is difficult to cure with antibiotics alone and is often complicated by perivalvular abscesses, aggressive early surgical intervention is recommended.[91]

Fungi

Yeasts Endocarditis caused by yeasts is more difficult to cure than most bacterial infections. The best results are achieved by a combination of antifungal chemotherapy and valve debridement or replacement.[21,92] Amphotericin B is the drug of choice for initial treatment for most patients. Synergism in vitro between flucytosine and amphotericin B has been reported with some strains of *Candida*, and their combined use for this life-threatening infection seems warranted. Serum levels of flucytosine should be monitored to reduce toxicity. The drug, which is cleared by the kidney, should not be given to patients with renal failure, because excessive serum concentrations can induce severe bone marrow depression. Long-term (possibly lifelong) suppressive therapy with oral fluconazole may be effective in patients with *Candida* endocarditis if they have no intracardiac complications. Similarly, long-term suppressive therapy has been advocated for patients with *Candida* endocarditis who have responded to combined antifungal and surgical treatment. Some yeast infections can be suppressed (and a few even cured) by treatment with fluconazole or itraconazole. With optimal therapy, the overall survival rate for patients with yeast endocarditis is about 50%.

Dimorphic and mycelial fungi Endocarditis caused by dimorphic and mycelial fungi is usually fatal despite therapy, with overall survival of less than 10%.[21] Aggressive treatment with amphotericin B and valvular surgery will save a few patients. If the fungus is susceptible to flucytosine, combination treatment with this agent plus amphotericin B may be tried.

Culture-Negative Endocarditis

When blood cultures are negative or culture results are not available, the selection of empirical therapy requires careful consideration of any clinical clues that might suggest the identity of the causative organism. Appropriate special cultures and serologic studies should be pursued in an attempt to identify microorganisms that are difficult to culture.[70,71]

In the absence of information suggesting a probable cause of native valve SBE, the same treatment as that used for enterococcal endocarditis (i.e., ampicillin plus gentamicin) is advised. If the course is acute and *S. aureus* is an etiologic consideration, the addition of vancomycin to this regimen is recommended. Because blood cultures rarely remain negative in patients with ABE, such empirical therapy can usually be revised later on the basis of the culture results. The spectrum of microorganisms causing PVE differs from that implicated in native valve endocarditis, and culture-negative PVE should be treated with a combination of vancomycin, gentamicin, and a third-generation cephalosporin. The third-generation cephalosporin is directed against the HACEK organisms, which are important causes of late PVE and can be difficult to isolate from blood cultures. Before embarking on therapy for culture-negative endocarditis, the physician should carefully consider the possibility that the patient does not have endocarditis, but rather has one of the conditions that mimics endocarditis.

Monitoring Clinical Response

A decrease in fever is usually evident within 1 to 7 days after the start of antibiotic therapy.[93] Fever that persists during the second week of appropriate antimicrobial therapy suggests the possibility of uncontrolled infection, which could be from antibiotic resistance, intracardiac abscess, metastatic foci of infection, or nosocomial complications of therapy. Thorough evaluation is indicated.[78,93] Petechiae and embolic phenomena may occur for several weeks after initiation of recommended treatment; such findings do not necessarily indicate that therapy is ineffective, particularly if other signs indicate that the patient is getting better.

If the patient does not become afebrile within a few days after antibiotic therapy is initiated, repeat blood cultures should be obtained. If the patient is afebrile and otherwise responding well, routine follow-up blood cultures during treatment are usually negative and are not recommended. Similarly, follow-up cultures after the end of antibiotic treatment are unnecessary in most patients, unless nonstandard antimicrobial regimens have been used or there is reason to suspect relapse, such as recurrent fever. Although the ESR slowly returns to normal in the months after effective treatment, this test is not useful for monitoring therapy or as a test of cure. C-reactive protein levels fall within 1 to 2 weeks after starting effective therapy; some clinicians use this test to provide reassurance that treatment is working.[94] The serum bactericidal test, in which dilutions of the patient's serum are tested for bactericidal activity against the organism causing the endocarditis, is no longer recommended for assessment of therapy in patients receiving standard antimicrobial regimens.[80]

SURGICAL INTERVENTION

Operative intervention to debride infected perivalvular tissue or to replace or reconstruct a dysfunctioning valve is important in the management of complicated infective endocarditis that involves either a native or a prosthetic valve.[49,51,80,91,95,96] Overall, surgery is indicated in 25% to 40% of patients with infective endocarditis. Several observations have prompted earlier and more frequent surgical intervention in active endocarditis: (1) the mortality of patients undergoing valve surgery early during active endocarditis is not greater than that for patients treated medically or operated upon later (after microbial cure), and in many series is lower[91,97]; (2) the risk that recurrent endocarditis from the same infecting organism will develop on a prosthesis newly implanted for treatment of endocarditis is very low[97]; and (3) some intracardiac complications of endocarditis can be corrected only with surgery. For patients with PVE, early aggressive surgical treatment is an essential element of therapy in about 45%.

The currently accepted indications for surgical treatment of active native or prosthetic valve endocarditis have been developed through a retrospective analysis of many cases. Moderate to severe heart failure from valvular dysfunction is the most widely accepted indication for valve replacement and accounts for the majority of decisions to operate. Cardiac surgery for these indications has been convincingly shown to improve survival. Bulky vegetations that obstruct the valve orifice may also produce heart failure and are a compelling indication for surgery. Surgical intervention is also indicated when there is clinical evidence of perivalvular invasion and abscess formation

and when infection remains uncontrolled for more than 1 to 3 weeks despite maximal antimicrobial therapy. Fungal endocarditis, especially in patients with intracardiac complications, is commonly treated surgically.[92] Similarly, PVE that occurs within a year after valve implantation and is caused by coagulase-negative staphylococci frequently results in perivalvular invasion. Regardless of the causative agent, if there is evidence that PVE is complicated by perivalvular extension of the infection, particularly if the valve is unstable, the PVE should be managed surgically. Because left-sided native valve endocarditis and, in particular, PVE caused by *S. aureus* are frequently invasive, early surgery should be considered for patients with these infections who do not show prompt and sustained improvement during antibiotic therapy.[51,91,96-99] Endocarditis caused by *P. aeruginosa* or other gram-negative bacilli that has not responded after 7 to 10 days of maximal antibiotic therapy should be treated surgically. Although patients with native valve endocarditis who experience relapses after appropriate antibiotic therapy often can be cured by repeat courses of antibiotics, surgical treatment improves the outcome in patients with PVE who experience relapse.

Arterial embolization, particularly to the brain, is another factor to be considered when deciding whether to operate. Emboli do not constitute an absolute indication for surgery, because the likelihood of recurrent embolization in any individual patient is highly unpredictable. Furthermore, the incidence of embolization falls rapidly after 1 week of effective antibiotic therapy. Choosing whether to operate therefore requires a consideration of relative risks and benefits in the light of multiple clinical, echocardiographic, and microbiologic observations, including embolization. For example, occurrence of an embolus early in the course, with a residual highly mobile vegetation of more than 10 mm in diameter, would be a relative indication for surgical intervention even if the patient did not have heart failure [*see* Clinical Presentations, Embolic Phenomena, *above*].[34,100-102] Surgery may be beneficial if the vegetations are unusually large and mobile.[102] In surgically treated patients with endocarditis, operative mortality is proportional to the degree of preoperative hemodynamic impairment. If surgery is indicated, it should generally be performed promptly. Delaying surgery to administer additional antibiotic therapy in the presence of uncontrolled infection or deteriorating hemodynamic status may result in a less favorable outcome.[97] On the other hand, delaying surgery (when hemodynamic status permits) may be desirable in patients who have experienced a neurologic complication.[43,103] In patients who have sustained a cerebral infarction, the exacerbation rate for cerebral complications is reduced to 10% if surgery is delayed for at least 15 days and is reduced to below 3% if the delay is 28 days. If cerebral hemorrhage has occurred, cardiac surgery should ideally be delayed for at least 14 days—preferably, 28 days or more.[43] Exclusion of a complicating cerebral mycotic aneurysm should be considered in this setting [*see* Clinical Presentations, Neurologic Manifestations, *above*].

Although implantation of a prosthetic valve in a patient who is actively abusing intravenous drugs may be lifesaving, endocarditis is likely to recur if the drug abuse continues, and it is more difficult to cure PVE than it is to cure native valve endocarditis. To avoid these problems, tricuspid or pulmonary valvulectomy without valve replacement has been employed in the treatment of right-sided endocarditis. However, in IDUs with isolated right-sided *S. aureus* endocarditis, surgery should be postponed because most of these cases can be cured medically.[19]

Table 6 Cardiac Abnormalities and Endocarditis Risk[115]

Higher risk	Prosthetic heart valves
	Previous infective endocarditis
	Cyanotic congenital heart disease
	Surgically constructed systemic-pulmonary shunts and conduits
Lower risk	Mitral valve prolapse with regurgitation
	Mitral regurgitation, stenosis, stenosis with regurgitation
	Aortic regurgitation, stenosis, stenosis with regurgitation
	Tricuspid valve disease
	Pulmonary stenosis
	Ventricular septal defect
	Patent ductus arteriosus
	Coarctation of the aorta
	Asymmetrical septal hypertrophy
	Calcific aortic sclerosis
	Degenerative valvular disease in elderly patients
	Surgically repaired intracardiac lesions with residual hemodynamic abnormality
Negligible risk	Mitral valve prolapse without regurgitation
	Minor valvular regurgitation by echocardiography without major structural abnormality
	Isolated atrial septal defect
	Arteriosclerotic plaques and coronary artery disease
	Cardiac pacemakers
	Surgically repaired intracardiac lesions with minimal or no hemodynamic abnormality

ANTITHROMBOTIC THERAPY

Anticoagulant therapy in a patient with endocarditis carries the potential risk of causing or worsening hemorrhage in the brain or other sites. However, the benefits of anticoagulation probably outweigh the risks if a strong indication exists, such as atrial fibrillation, cardiomyopathy, mural thrombus, or deep vein thrombosis. Anticoagulant therapy may be carefully administered to patients with endocarditis when it is so indicated.[104] In patients with prosthetic valves who require long-term warfarin therapy, such therapy should be continued during treatment of endocarditis unless there are specific contraindications. The prothrombin time should be carefully maintained in the lower therapeutic range, at an international normalized ratio (INR) of 2.0 to 3.0. Anticoagulation should be reversed immediately in the event of CNS complications, especially intracranial hemorrhage. Anticoagulation should be interrupted for 1 to 2 weeks after an acute embolic stroke. Heparin therapy should be avoided during active endocarditis if at all possible.

There is no evidence in humans that antithrombotic therapy can prevent arterial embolization of vegetations or otherwise improve the response of endocarditis to standard treatment regimens.

Prophylaxis

Antibiotics can prevent endocarditis in animal models if given before experimentally induced bacteremias, and may do so in humans if given before medical, dental, and surgical procedures that are known to cause bacteremia. However, this has not been proved. Current guidelines support the administra-

Table 7 Procedures and Endocarditis Risk[115]

Higher risk	Dental procedures that involve the gingival crevice Surgical procedures (including biopsies) inside the oral cavity Genitourinary tract procedures when bacterial infection is present Surgical procedures (including incision and drainage) when bacterial infection is present at the site
Lower risk	Injection of local anesthetic Genitourinary tract procedures, in absence of active bacterial infection Surgery involving gastrointestinal or respiratory mucosa, in absence of active bacterial infection Skin biopsies and dermatologic procedures using standard antisepsis Bronchoscopy, with or without biopsy Gastrointestinal endoscopy, with or without biopsy Cardiac catheterization Transesophageal echocardiography Esophageal dilatation and sclerotherapy of esophageal varices Endotracheal tube insertion Tympanostomy tube insertion *In the absence of active bacterial infection*: urethral catheterization, laparoscopy, sterilization procedures, vaginal delivery, vaginal hysterectomy, cesarean section, therapeutic abortion, dilatation and curettage, insertion or removal of intrauterine devices

tion of prophylactic antibiotic to patients who are at higher risk for endocarditis than the general population when they undergo procedures likely to lead to bacteremia with organisms that commonly cause endocarditis.[105-107] The appropriate selection of patients for this intervention remains controversial. Such selections should be made on the basis of an assessment of the combined risk of the underlying cardiac condition, the likelihood that the procedure will cause endocarditis (which is always low), and the potential morbidity and mortality[86] [*see Tables 6, 7 and 8*]. Patients who are at the highest risk for endocarditis are those with a history of previous infective endocarditis, prosthetic valves, cyanotic congenital heart disease, or surgically constructed shunts or conduits. Prophylaxis is merited in all such cases.

Prophylaxis is recommended for patients at high risk for endocarditis who undergo procedures that involve the oral cavity, the respiratory tract, or the esophagus and are likely to cause streptococcal bacteremia. Prophylaxis is also indicated for patients at high risk for endocarditis who undergo procedures that involve the genitourinary tract, the GI tract distal to the stomach, or the biliary tract and are likely to cause bacteremia with enterococci [*see Table 7*]. For patients who are at unusually high risk for endocarditis, some physicians elect to use prophylaxis even for procedures carrying low likelihood of bacteremia (for which prophylaxis would generally not be recommended).

Antibiotic regimens recommended for prophylaxis in high-risk patients have been designed to cover streptococci or enterococci and are assigned in accordance with the anticipated bacteremic organism [*see Tables 8 and 9*]. When urinary tract manipulation is anticipated, an evaluation for urinary tract infection should be undertaken, and infection, if confirmed, should be treated before the manipulation. If procedures must be performed on infected tissues, prophylaxis is directed against the likely pathogen causing the infection. Such prophylaxis would include use of an antistaphylococcal penicillin or a first-generation cephalosporin when the infection involves skin and adjacent soft tissue, joints, or bone.

If a patient who is at risk for endocarditis is receiving or has recently received one of the agents recommended for prophylaxis, it is prudent to choose for prophylaxis an antibiotic from a different class, rather than increase the dosage of the current regimen. The dosage of penicillin used to prevent recurrences of acute rheumatic fever is insufficient to prevent bacterial endocarditis. Moreover, patients taking oral penicillins or cephalosporins may harbor relatively penicillin-resistant strains of bacteria in their oral cavity. Prophylaxis with clarithromycin or clindamycin is suggested when relatively penicillin-resistant streptococci are anticipated.

Most cases of endocarditis are not related to defined events or procedures. It is likely that these infections arise from bacteremias associated with minor unrecognized infections or ordinary activities such as vigorous chewing in patients with gingival disease or use of oral irrigating jets. In patients with vulnerable cardiac lesions, maintaining optimal dental health is essential if the risk of endocarditis is to be minimized. Dental evaluation and treatment, including extractions and necessary restorative work, should be performed under appropriate antibiotic prophylaxis several weeks before the insertion of prosthetic heart valves.

Prognosis

Overall survival for patients with infective endocarditis is about 75% to 80%.[2,44,108] This rate is little better than that in the 1950s and needs to be improved.[108] The outlook depends largely on the promptness of diagnosis and initiation of therapy, the nature of the infecting organism, the presence of comorbid condi-

Table 8 Recommendations for Endocarditis Prophylaxis[115]

*Procedure**	*Preexisting Cardiac Risk Factors†*		
	Higher risk	*Lower risk*	*Negligible risk*
Higher risk	Prophylaxis recommended	Prophylaxis not recommended but optional; clinical judgment needed	Prophylaxis not recommended
Lower risk	Prophylaxis not recommended but optional; clinical judgment needed	Prophylaxis not recommended	Prophylaxis not recommended

*See Table 7.
† See Table 6.

Table 9 Recommended Regimens for Prophylaxis of Endocarditis[116-119]

Category	*Indication*	*Drug and Dosage*
Standard regimen	For general use when prevention of endocarditis is recommended [*see Table 8*]	Amoxicillin, 2 g orally 1 hr before procedure*
Special regimens	Oral regimens for penicillin-allergic patients (oral and respiratory tract procedures only)	Clindamycin, 600 mg orally 1 hr before procedure, then 300 mg 6 hr later* *or* Clarithromycin, 0.5 g orally 1 hr before procedure
	Parenteral regimen for high-risk patients; also for GI or GU tract procedures	Ampicillin, 2 g I.M. or I.V. *plus* Gentamicin, 1.5 mg/kg I.M. or I.V. 30 min before procedure*
	Parenteral regimen for penicillin-allergic patients	Vancomycin, 1 g I.V. slowly over 1 hr, starting 1 hr before procedure, then *add* Gentamicin, 1.5 mg/kg I.M. or I.V. if GI or GU tract involved*
	Cardiac surgery, including implantation of prosthetic valves	Cefazolin, 2 g I.V. at induction of anesthesia, repeated 8 and 16 hr later*† *or* Vancomycin, 1 g I.V. slowly over 1 hr starting at induction, then 0.5 g I.V. 12 hr later*†‡

Note: No regimen has been proved effective for the prevention of endocarditis, and prevention failures may occur with any regimen. These guidelines are not intended as the standard of care and are not intended to cover all clinical situations; practitioners should use their own judgment on safety and cost-benefit issues in each individual case. One or two additional doses may be given if the period of risk for bacteremia is prolonged.

*Pediatric dosages: ampicillin, 50 mg/kg; cefazolin, 30 mg/kg; clindamycin, for children > 60 lb, use the same as for adults, and for children < 60 lb, use half the adult dose; gentamicin, 2 mg/kg; amoxicillin, for children > 60 lb, use the same as for adults, and for children < 60 lb, use half the adult dose; vancomycin, 20 mg/kg.

†Gentamicin, 1.5 mg/kg I.V. may be given with each dose if postoperative gram-negative infections have occurred with significant frequency.

‡This regimen is recommended for units where coagulase-negative staphylococcal prosthetic valve infections have occurred with significant frequency.

tions, and the development of cardiac and neurologic complications.[44,50] Survival is above 90% when the infecting organism is a viridans-type streptococcus or *S. bovis* but is only about 50% in non-IDUs infected with *S. aureus*. Although outcomes are worse in cases of early PVE, survival rates for patients with late PVE appear to be similar to those for patients with native valve infection.[5,6,89] With the incorporation of earlier surgical intervention for patients with endocarditis, cardiac failure is less strongly associated with increased mortality than it once was. However, major CNS complications and uncontrolled infection, especially abscess formation, continue to be associated with increased mortality.[41] The short-term outlook is relatively good for IDUs with right-sided endocarditis.[19] Endocarditis caused by gram-negative bacilli other than HACEK organisms or caused by yeasts is difficult to cure despite optimal antimicrobial therapy and aggressive surgical intervention. Endocarditis caused by mycelial fungi is usually fatal.[21]

David T. Durack, M.B., D.Phil., F.A.C.P., is an employee of Becton, Dickinson and Company and has received grants for educational activities from the Merieux Foundation.

Adolf W. Karchmer, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hoen B, Alla F, Beguinot I, et al: Changing profile of infective endocarditis: results of a one-year survey in France in 1999. JAMA 288:75, 2002

2. Benn M, Hagelskjaer LH, Tvede M: Infective endocarditis, 1984 through 1993: a clinical and microbiological survey. J Intern Med 242:15, 1997

3. Strom BL, Abrutyn E, Berlin JA, et al: Dental and cardiac risk factors for infective endocarditis: a population-based case-control study. Ann Intern Med 129:761, 1998

4. Cabell CH, Jollis JG, Peterson GE, et al: Changing patient characteristics and the effect on mortality in endocarditis. Arch Intern Med 162:90, 2002

5. Watanakunakorn C, Burkert T: Infective endocarditis at a large community teaching hospital, 1980–1990: a review of 210 episodes. Medicine (Baltimore) 72:90, 1993

6. Sandre RM, Shafran SD: Infective endocarditis: review of 135 cases over 9 years. Clin Infect Dis 22:276, 1996

7. Ruoff KL, Miller SI, Garner CV, et al: Bacteremia with *Streptococcus bovis* and *Streptococcus salivarius:* clinical correlates of more accurate identification of isolates. J Clin Microbiol 27:305, 1989

8. Shapiro DS, Kenney SC, Johnson M, et al: Brief report: *Chlamydia psittaci* endocarditis diagnosed by blood culture. N Engl J Med 326:1192, 1992

9. Hoen B, Selton-Suty C, Lacassin F, et al: Infective endocarditis in patients with negative blood cultures: analysis of 88 cases from a one-year nationwide survey in France. Clin Infect Dis 20:501, 1995

10. Raoult D, Fournier PE, Drancourt M, et al: Diagnosis of 22 new cases of *Bartonella* endocarditis. Ann Intern Med 125:646, 1996

11. Drancourt M, Mainardi JL, Brouqui P, et al: *Bartonella (Rochalimaea) quintana* endocarditis in three homeless men. N Engl J Med 332:419, 1996

12. Spach DH, Koehler JE: *Bartonella*-associated infections. Infect Dis Clin North Am 12:137, 1998

13. Stein A, Raoult D: Q fever endocarditis. Eur Heart J 16(suppl B):19, 1995

14. Abramson MA, Sexton DJ: Nosocomial methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* primary bacteremia: at what costs? Infect Control Hosp Epidemiol 20:408, 1999

15. Gouello JP, Asfar P, Brenet O, et al: Nosocomial endocarditis in the intensive care unit: an analysis of 22 cases. Crit Care Med 28:377, 2000

16. Watanakunakorn C: Increasing importance of intravascular device–associated *Staphylococcus aureus* endocarditis. Clin Infect Dis 28:115, 1999

17. Fang G, Keys TF, Gentry LO, et al: Prosthetic valve endocarditis resulting from nosocomial bacteremia: a prospective, multicenter study. Ann Intern Med 119:560, 1993

18. Fernandez-Guerrero ML, Verdejo C, Azofra J, et al: Hospital-acquired infectious endocarditis not associated with cardiac surgery: an emerging problem. Clin Infect Dis 20:16, 1995

19. Miro JM, del Rio A, Mestres CA: Infective endocarditis and cardiac surgery in intravenous drug abusers and HIV-1 infected patients. Cardiol Clin 21:167, 2003

20. Ellis ME, Al-Abdely H, Sandridge A, et al: Fungal endocarditis: evidence in the world literature, 1965–1995. Clin Infect Dis 32:50, 2001

21. Pierrotti LC, Baddour LM: Fungal endocarditis, 1995–2000. Chest 122:302, 2002

22. Durack DT, Beeson PB: Pathogenesis of infective endocarditis. Infective Endocarditis. Rahimtoola SH, Ed. Grune & Stratton, New York, 1978, p 1

23. Veltrop MH, Langermans JA, Thompson J, et al: Interleukin-10 regulates the tissue factor activity of monocytes in an in vitro model of bacterial endocarditis. Infect Immun 69:3197, 2001

24. Bancsi MJ, Thompson J, Bertina RM: Stimulation of monocyte tissue factor expression in an in vitro model of bacterial endocarditis. Infect Immun 62:5669, 1994

25. Dhawan VK, Bayer AS, Yeaman MR: In vitro resistance to thrombin-induced

platelet microbicidal protein is associated with enhanced progression and hematogenous dissemination in experimental *Staphylococcus aureus* infective endocarditis. Infect Immun 66:3476, 1998

26. Dankert J, Krijgsveld J, van Der WJ, et al: Platelet microbicidal activity is an important defense factor against viridans streptococcal endocarditis. J Infect Dis 184:597, 2001

27. Moreillon P, Que YA, Bayer AS: Pathogenesis of streptococcal and staphylococcal endocarditis. Infect Dis Clin North Am 16:297, 2002

28. Moreillon P, Entenza JM, Francioli P, et al: Role of *Staphylococcus aureus* coagulase and clumping factor in pathogenesis of experimental endocarditis. Infect Immun 63:4738, 1995

29. Que YA, Francois P, Haefliger JA, et al: Reassessing the role of *Staphylococcus aureus* clumping factor and fibronectin-binding protein by expression in *Lactococcus lactis*. Infect Immun 69:6296, 2001

30. Gagliardi JP, Nettles RE, McCarty DE, et al: Native valve infective endocarditis in elderly and younger adult patients: comparison of clinical features and outcomes with use of the Duke criteria and the Duke Endocarditis Database. Clin Infect Dis 26:1165, 1998

31. Cunha BA, Gill MV, Lazar JM: Acute infective endocarditis: diagnostic and therapeutic approach. Infect Dis Clin North Am 10:811, 1996

32. Larbalestier RI, Kinchla NM, Aranki SF, et al: Acute bacterial endocarditis: optimizing surgical results. Circulation 86(5 suppl):II68, 1992

33. Meine TJ, Nettles RE, Anderson DJ, et al: Cardiac conduction abnormalities in endocarditis defined by the Duke criteria. Am Heart J 142:280, 2001

34. Steckelberg JM, Murphy JG, Ballard D, et al: Emboli in infective endocarditis: the prognostic value of echocardiography. Ann Intern Med 114:635, 2001

35. Millaire A, Leroy O, Gaday V, et al: Incidence and prognosis of embolic events and metastatic infections in infective endocarditis. Eur Heart J 18:677, 1997

36. Lancellotti P, Galiuto L, Albert A, et al: Relative value of clinical and transesophageal echocardiographic variables for risk stratification in patients with infective endocarditis. Clin Cardiol 21:572, 1998

37. Conlon PJ, Jefferies F, Krigman HR, et al: Predictors of prognosis and risk of acute renal failure in bacterial endocarditis. Clin Nephrol 49:96, 1998

38. Mansur AJ, Grinberg M, Leao PP, et al: Extracranial mycotic aneurysms in infective endocarditis. Clin Cardiol 9:65, 1986

39. Brust JCM, Dickinson PCT, Hughes JEO, et al: The diagnosis and treatment of cerebral mycotic aneurysms. Ann Neurol 27:238, 1990

40. Bamford J, Hodges J, Warlow C: Late rupture of a mycotic aneurysm after "cure" of bacterial endocarditis. J Neurol 233:51, 1986

41. Mansur AJ, Grinberg M, da Luz PL, et al: The complications of infective endocarditis: a reappraisal in the 1980s. Arch Intern Med 152:2428, 1992

42. Bullock R, Van Dellen JR, Van Den Heever CM: Intracranial mycotic aneurysms: a review of 9 cases. S Afr Med J 60:970, 1981

43. Eishi K, Kawazoe K, Kuriyama Y, et al: Surgical management of infective endocarditis associated with cerebral complications: multi-center retrospective study in Japan. J Thorac Cardiovasc Surg 110:1745, 1995

44. Hasbun R, Vikram HR, Barakat LA, et al: Complicated left-sided native valve endocarditis in adults: risk classification for mortality. JAMA 289:1933, 2003

45. Camarata PJ, Latchaw RE, Rufenacht DA, et al: Intracranial aneurysms. Invest Radiol 28:373, 1993

46. Huston J III, Nichols DA, Luetmer PH, et al: Blinded prospective evaluation of sensitivity of MR angiography to known intracranial aneurysms: importance of aneurysm size. AJNR Am J Neuroradiol 15:1607, 1994

47. Moskowitz MA, Rosenbaum AE, Tyler HR: Angiographically monitored resolution of cerebral mycotic aneurysms. Neurology 24:1103, 1974

48. Sande MA, Lee BL, Mills J, et al: Endocarditis in intravenous drug users. Infective Endocarditis. Kaye D, Ed. Raven Press, New York, 1992, p 345

49. Karchmer AW: Infective endocarditis. Heart Disease: A Textbook of Cardiovascular Medicine. Braunwald E, Ed. WB Saunders Co, Philadelphia, 1996, p 1077

50. Vlessis AA, Khaki A, Grunkemeier GL, et al: Risk, diagnosis and management of prosthetic valve endocarditis: a review. J Heart Valve Dis 6:443, 1997

51. Piper C, Korfer R, Horstkotte D: Prosthetic valve endocarditis. Heart 85:590, 2001

52. Blumberg EA, Karalis DA, Chandrasekaran K, et al: Endocarditis-associated paravalvular abscesses: do clinical parameters predict the presence of abscess? Chest 107:898, 1995

53. Daniel WG, Mugge A, Martin RP, et al: Improvement in the diagnosis of abscesses associated with endocarditis by transesophageal echocardiography. N Engl J Med 324:795, 1991

54. Sett SS, Hudson MP, Jamieson WR, et al: Prosthetic valve endocarditis: experience with porcine bioprosthesis. J Thorac Cardiovasc Surg 105:428, 1993

55. Bayer AS, Bolger AF, Taubert KA, et al: Diagnosis and management of infective endocarditis and its complications. Circulation 98:2936, 1998

56. Durack DT, Bright DK, Lukes AS, et al: New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings. Am J Med 96:200, 1994

57. Hoen B, Selton-Suty C, Danchin N, et al: Evaluation of the Duke criteria versus the Beth Israel criteria for the diagnosis of infective endocarditis. Clin Infect Dis 21:905, 1995

58. Nettles RE, McCarty DE, Corey GR, et al: An evaluation of the Duke criteria in 25 pathologically confirmed cases of prosthetic valve endocarditis. Clin Infect Dis 25:1401, 1997

59. Dodds GA, Sexton DJ, Durack DT, et al: Negative predictive value of the Duke criteria for infective endocarditis. Am J Cardiol 77:403, 1996

60. Hoen B, Beguinot I, Rabaud C, et al: The Duke criteria for diagnosing infective endocarditis are specific: analysis of 100 patients with acute fever or fever of unknown origin. Clin Infect Dis 23:298, 1996

61. Sekeres MA, Abrutyn E, Berlin JA, et al: An assessment of the usefulness of the Duke criteria for diagnosing active infective endocarditis. Clin Infect Dis 24:1185, 1997

62. Sachdev M, Peterson GE, Jollis JG: Imaging techniques for diagnosis of infective endocarditis. Cardiol Clin 21:185, 2003

63. Shively BK, Gurule FT, Roldan CA, et al: Diagnostic value of transesophageal compared with transthoracic echocardiography in infective endocarditis. J Am Coll Cardiol 18:391, 1991

64. Shapiro SM, Young E, De Guzman S, et al: Transesophageal echocardiography in diagnosis of infective endocarditis. Chest 105:377, 1994

65. Roe MT, Abramson MA, Li J, et al: Clinical information determines the impact of transesophageal echocardiography on the diagnosis of infective endocarditis by the Duke criteria. Am Heart J 139:945, 2000

66. Karalis DG, Bansal RC, Hauck AJ, et al: Transesophageal echocardiographic recognition of subaortic complications in aortic valve endocarditis: clinical and surgical implications. Circulation 86:353, 1992

67. Sochowski RA, Chan KL: Implication of negative results on a monoplane transesophageal echocardiographic study in patients with suspected infective endocarditis. J Am Coll Cardiol 21:216, 1993

68. Daniel WG, Mugge A, Grote J, et al: Comparison of transthoracic and transesophageal echocardiography for detection of abnormalities of prosthetic and bioprosthetic valves in the mitral and aortic positions. Am J Cardiol 71:210, 1993

69. Vered Z, Mossinson D, Peleg E, et al: Echocardiographic assessment of prosthetic valve endocarditis. Eur Heart J 16(suppl B):63, 1995

70. Towns ML, Reller LB: Diagnostic methods: current best practices and guidelines for isolation of bacteria and fungi in infective endocarditis. Cardiol Clin 21:197, 2003

71. Tunkel AR, Kaye D: Endocarditis with negative blood cultures. N Engl J Med 326:1215, 1992

72. Fournier PE, Casalta JP, Habib G, et al: Modification of the diagnostic criteria proposed by the Duke Endocarditis Service to permit improved diagnosis of Q fever endocarditis. Am J Med 100:629, 1996

73. Houpikian P, Raoult D: Diagnostic methods: current best practices and guidelines for identification of difficult-to-culture pathogens in infective endocarditis. Cardiol Clin 21:207, 2003

74. Hamed KA, Dormitzer PR, Su CK, et al: *Haemophilus parainfluenzae* endocarditis: application of a molecular approach for identification of pathogenic bacterial species. Clin Infect Dis 19:677, 1994

75. Goldenberger D, Kunzli A, Vogt P, et al: Molecular diagnosis of bacterial endocarditis by broad-range PCR amplification and direct sequencing. J Clin Microbiol 35:2733, 1997

76. Lisby G, Gutschik E, Durack DT: Molecular methods for diagnosis of infective endocarditis. Infect Dis Clin North Am 16:393, 2002

77. Perivalvular abscesses associated with endocarditis: clinical features and prognostic factors of overall survival in a series of 233 cases. Perivalvular Abscesses French Multicentre Study. Eur Heart J 20:232, 1999

78. Blumberg EA, Robbins N, Adimora A, et al: Persistent fever in association with infective endocarditis. Clin Infect Dis 15:983, 1992

79. Doern GV, Ferraro MJ, Brueggmann AB, et al: Emergence of high rates of antimicrobial resistance among viridans group streptococci in the United States. Antimicrob Agents Chemother 40:891, 1996

80. Wilson WR, Karchmer AW, Dajani AS, et al: Antibiotic treatment of adults with infective endocarditis due to viridans streptococci, enterococci, staphylococci and HACEK microorganisms. JAMA 274:1706, 1995

81. Hoen B: Special issues in the management of infective endocarditis caused by Gram-positive cocci. Infect Dis Clin North Am 16:437, 2002

82. Bouvet A: Human endocarditis due to nutritionally variant streptococci: *Streptococcus adjacens* and *Streptococcus defectivus*. Eur Heart J 16(suppl B):24, 1995

83. Martinez E, Miro JM, Almirante B, et al: Effect of penicillin resistance of *Streptococcus pneumoniae* on the presentation, prognosis, and treatment of pneumococcal endocarditis in adults. Clin Infect Dis 35:130, 2002

84. Infective endocarditis caused by beta-hemolytic streptococci. The Infectious Diseases Society of America's Emerging Infections Network. Clin Infect Dis 26:66, 1998

85. Eliopoulos GM: Aminoglycoside resistant enterococcal endocarditis. Infect Dis Clin North Am 7:117, 1993

86. Rice LB, Calderwood SB, Eliopoulos GM, et al: Enterococcal endocarditis: a comparison of prosthetic and native valve disease. Rev Infect Dis 13:1, 1991

87. Levine DP, Fromm BS, Reddy BR: Slow response to vancomycin or vancomycin plus rifampin in methicillin-resistant *Staphylococcus aureus* endocarditis. Ann Intern Med 115:674, 1991

88. Torres-Tortosa M, de Cueto M, Vergara A, et al: Prospective evaluation of a two-week course of intravenous antibiotics in intravenous drug addicts with infective endocarditis. Grupo de Estudio de Enfermedades Infecciosas de la Provincia de Cadiz. Eur J Clin Microbiol Infect Dis 13:559, 1994

89. Calderwood SB, Swinski LA, Waternaux CM, et al: Risk factors for the development of prosthetic valve endocarditis. Circulation 72:31, 1985

90. Karchmer AW, Gibbons DW: Infections of prosthetic heart valves and vascular grafts. Infections Associated with Indwelling Medical Devices. Bisno AL, Waldvogel FA, Eds. American Society for Microbiology, Washington DC, 1994, p 213

91. Olaison L, Pettersson G: Current best practices and guidelines: indications for surgical intervention in infective endocarditis. Cardiol Clin 21:235, 2003

92. Melgar GR, Nasser RM, Gordon SM, et al: Fungal prosthetic valve endocarditis in 16

patients: an 11-year experience in a tertiary care hospital. Medicine (Baltimore) 76:94, 1997

93. Lederman MM, Sprague L, Wallis RS, et al: Duration of fever during treatment of infective endocarditis. Medicine (Baltimore) 71:52, 1992

94. Heiro M: Extensive analysis of the utility of serum C-reactive protein, erythrocyte sedimentation rate and white blood cell count in assessing the outcome of infective endocarditis. (personal communication, 2004)

95. Jault F, Gandjbakhch I, Chastre JC, et al: Prosthetic valve endocarditis with ring abscesses: surgical management and long-term results. J Thorac Cardiovasc Surg 105:1106, 1993

96. Vlessis AA, Hovaguimian H, Jaggers J, et al: Infective endocarditis: ten year review of medical and surgical therapy. Ann Thorac Surg 61:1217, 1996

97. D'Agostino RS, Miller DC, Stinson EB, et al: Valve replacement in patients with native valve endocarditis: what really determines operative outcome? Ann Thorac Surg 40:429, 1985

98. Gordon SM, Serkey JM, Longworth DL, et al: Early onset prosthetic valve endocarditis: the Cleveland Clinic experience 1992–1997. Ann Thorac Surg 69:1388, 2000

99. John MD, Hibberd PL, Karchmer AW, et al: *Staphylococcus aureus* prosthetic valve endocarditis: optimal management and risk factors for death. Clin Infect Dis 26:1302, 1998

100. Mugge A, Daniel WG, Frank G, et al: Echocardiography in infective endocarditis: reassessment of prognostic implications of vegetation size determined by the transthoracic and transesophageal approach. J Am Coll Cardiol 14:631, 1989

101. Rohmann S, Erbel R, Gorge G, et al: Clinical relevance of vegetation localization by transoesophageal echocardiography in infective endocarditis. Eur Heart J 12:446, 1992

102. Sanfilippo AJ, Picard MH, Newell JB, et al: Echocardiographic assessment of patients with infectious endocarditis: prediction of risk for complications. J Am Coll Cardiol 18:1191, 1991

103. Piper C, Wiemer M, Schulte HD, et al: Stroke is not a contraindication for urgent valve replacement in acute infective endocarditis. J Heart Valve Dis 10:703, 2001

104. Wilson WR, Geraci JE, Danielson GK, et al: Anticoagulant therapy and central nervous system complications in patients with prosthetic valve endocarditis. Circulation 57:1004, 1978

105. Durack DT: Prevention of infective endocarditis. N Engl J Med 332:38, 1995

106. Dajani AS, Taubert KA, Wilson WR, et al: Prevention of bacterial endocarditis: recommendations by the American Heart Association. Circulation 96:358, 1997

107. Morris CD, Reller MD, Menashe VD: Thirty-year incidence of infective endocarditis after surgery for congenital heart defect. JAMA 279:599, 1998

108. Durack DT: Evaluating and optimizing outcomes of surgery for endocarditis. JAMA 290:3250, 2003

109. Kazanjian PH: Infective endocarditis: review of 60 cases treated in community hospitals. Infect Dis Clin Pract 2:41, 1993

110. Tornos P, Sanz E, Permanyer-Miralda G, et al: Late prosthetic valve endocarditis: immediate and long-term prognosis. Chest 1:37, 1992

111. Grover FL, Cohen DJ, Oprian C, et al: Determinants of the occurrence of and survival from prosthetic valve endocarditis: experience of the Veterans Affairs Cooperative Study on Valvular Heart Disease. J Thorac Cardiovasc Surg 108:207, 1994

112. Li JS, Sexton DJ, Mick N, et al: Proposed modifications to the Duke criteria for the diagnosis of infective endocarditis. Clin Infect Dis 30:633, 2000

113. Morris AJ, Drinkovic D, Pottumarthy S, et al: Gram stain, culture, and histopathological examination findings for heart valves removed because of infective endocarditis. Clin Infect Dis 36:697, 2003

114. Verhagen DW, van der Feltz M, Plokker HW, et al: Optimisation of the antibiotic guidelines in the Netherlands. VII. SWAB guidelines for antimicrobial therapy in adult patients with infectious endocarditis. Neth J Med 61:421, 2003

115. Durack DT: Prophylaxis of infective endocarditis. Mandell, Douglas and Bennett's Principles and Practice of Infectious Diseases, 6th ed. Mandell GL, Bennett JE, Dolin R, Eds. Elsevier, Philadelphia, 2004, in press

116. Dajani AS, Taubert KA, Wilson W, et al: Prevention of bacterial endocarditis: recommendations by the American Heart Association. Circulation 96:358, 1997

117. ACOG Practice Bulletin. Prophylactic antibiotics in labor and delivery. Number 47, October 2003. American College of Obstetricians and Gynecologists. Obstet Gynecol 102:875, 2003

118. Antibiotic prophylaxis for gastrointestinal endoscopy. American Society for Gastrointestinal Endoscopy. Gastrointest Endosc 42:630, 1995

119. Simmons NA: Recommendations for endocarditis prophylaxis. J Antimicrob Chemother 31:437, 1993

Acknowledgment

Figures 2 and 3 Carol Donner.

XIX BACTERIAL INFECTIONS OF THE UPPER RESPIRATORY TRACT

Harvey B. Simon, m.d.

Overview of Upper Respiratory Tract Infections

Upper respiratory tract infections range from common, benign processes, such as nasopharyngitis, to uncommon, potentially lethal processes, such as epiglottitis. Pathogens range from pyogenic bacteria [*see 7:I Infections Due to Gram-Positive Cocci and 7:II Infections Due to Mycobacteria*] to *Chlamydia* [*see 7:XIII Diseases Due to Chlamydia*], mycoplasmas [*see 7:XII Infections Due to Mycoplasmas*], respiratory viruses [*see 7:XXV Respiratory Viral Infections*], and, on occasion, fungi [*see 7:XXXVII Mycotic Infections*]. Despite this heterogeneity, some general anatomic and physiologic principles apply.

pathophysiology

The upper respiratory tract is composed of two distinct types of epithelial surface. A stratified squamous epithelium lines the oropharynx and nasopharynx [*see Figure 1*]. These regions are normally teeming with a variety of microbial flora, and many potential pathogens can temporarily colonize these epithelial surfaces without causing true infection [*see Table 1*]. Bacteriologic cultures are easily obtained from these areas. As a result, the microbiology laboratory faces the challenge of isolating potential pathogens from the normal mouth flora, and the clinician faces the challenge of distinguishing between the carrier state and active infection. A respiratory epithelium, composed of ciliated columnar cells, goblet cells, and mucous and serous glands, lines the paranasal sinuses, the middle ear, and the airway below the epiglottis. These regions, in contrast to the oropharynx and nasopharynx, are normally sterile. Because these areas are inaccessible for routine culture, it is often necessary to diagnose and treat infections on the basis of clinical findings and statistical probabilities [*see Table 1*] rather than on the basis of bacteriologic data derived from an individual patient.

Many host defense mechanisms protect the upper airway from infection. Mechanical defenses tend to prevent penetration by organisms from the nasopharynx and oral cavity into more vulnerable areas. These defenses are the cough, gag, and sneeze reflexes; the viscous mucous secretions, which entrap particulate material; and ciliary action, which propels entrapped particles outward. In addition, local immunologic defenses attempt to deal with organisms that have breached the mechanical barriers; such defenses include abundant lymphoid tissue, secretory IgA antibodies in respiratory secretions, and a rich vasculature that can rapidly deliver phagocytic leukocytes.

etiology

Upper respiratory tract infections typically occur after insults to host defenses—such as allergic rhinitis, chemical irritation, trauma, and, especially, viral respiratory tract infections—that impair ciliary function and produce increased volumes of thin secretions and mucosal edema that block the narrow channels draining the sinuses and middle ear. Barotrauma, which results from a rapid change in atmospheric pressure, can intensify upper respiratory tract problems, particularly in patients with rhinitis, otitis, or sinusitis.

Many upper respiratory tract infections are caused by viruses, and many of the bacterial infections in this region will resolve without antibiotic therapy.[1] Unfortunately, antibiotics are overprescribed for upper respiratory tract infections; in the United States, antibiotics are prescribed for 50% to 75% of ambulatory patients seen for viral upper respiratory tract infections.[2]

complications

The complications of upper airway infections stem from several factors. The proximity of upper airway structures to the central nervous system and the presence of abundant vascular channels connecting the two account for some of the most severe complications. In addition, the numerous fascial planes of the head and neck provide potential spaces where infection can become sequestered. Furthermore, because the upper airway is narrow, infection in the neck may compromise the patency of the airway itself. Finally, immunosuppression renders persons vulnerable to unusual pathogens and unusual complications.

Sinusitis

Although infections of the paranasal sinuses are common, they are greatly overdiagnosed by patient and physician. Acute sinusitis is characterized by nasal congestion, purulent nasal discharge, fetid breath, facial pain that typically increases when the patient stoops forward, and, often, fever and other systemic symptoms. In most patients, acute sinusitis responds well to medical therapy; in a few persons, however, chronic sinusitis may result. Chronic sinusitis is characterized by purulent discharge, usually without fever, that persists for weeks or months.

etiology

Viral, allergic, or vasomotor rhinitis is frequently an antecedent to sinusitis. Nose blowing propels nasal fluid into the sinuses, potentially introducing viscid fluid and bacteria that can lead to sinusitis.[3] Nasal polyps, deviation of the nasal septum, or hypertrophied adenoids may predispose to purulent sinusitis by obstructing sinus drainage. Cigarette smoke and overuse of topical decongestants impair ciliary action and alter the mucous blanket, predisposing to sinusitis. Contributing factors may include rapid changes in altitude, trauma, intranasal foreign bodies or tumors, cocaine abuse, and such systemic processes as cystic fibrosis and Kartagener syndrome (situs inversus, bronchiectasis, and sinusitis). Factors that predispose to chronic sinusitis include inadequately treated bacterial infections, long-term impairment of mucociliary clearance, obstruction of sinus drainage, and allergies and asthma.

Nosocomial sinusitis occurs as a complication of nasotracheal intubation or, less often, the use of nasoenteric feeding tubes.[4] Unexplained fever may be the presenting symptom, and computed tomography scanning is often required for diagnosis if the patient cannot localize the pain.

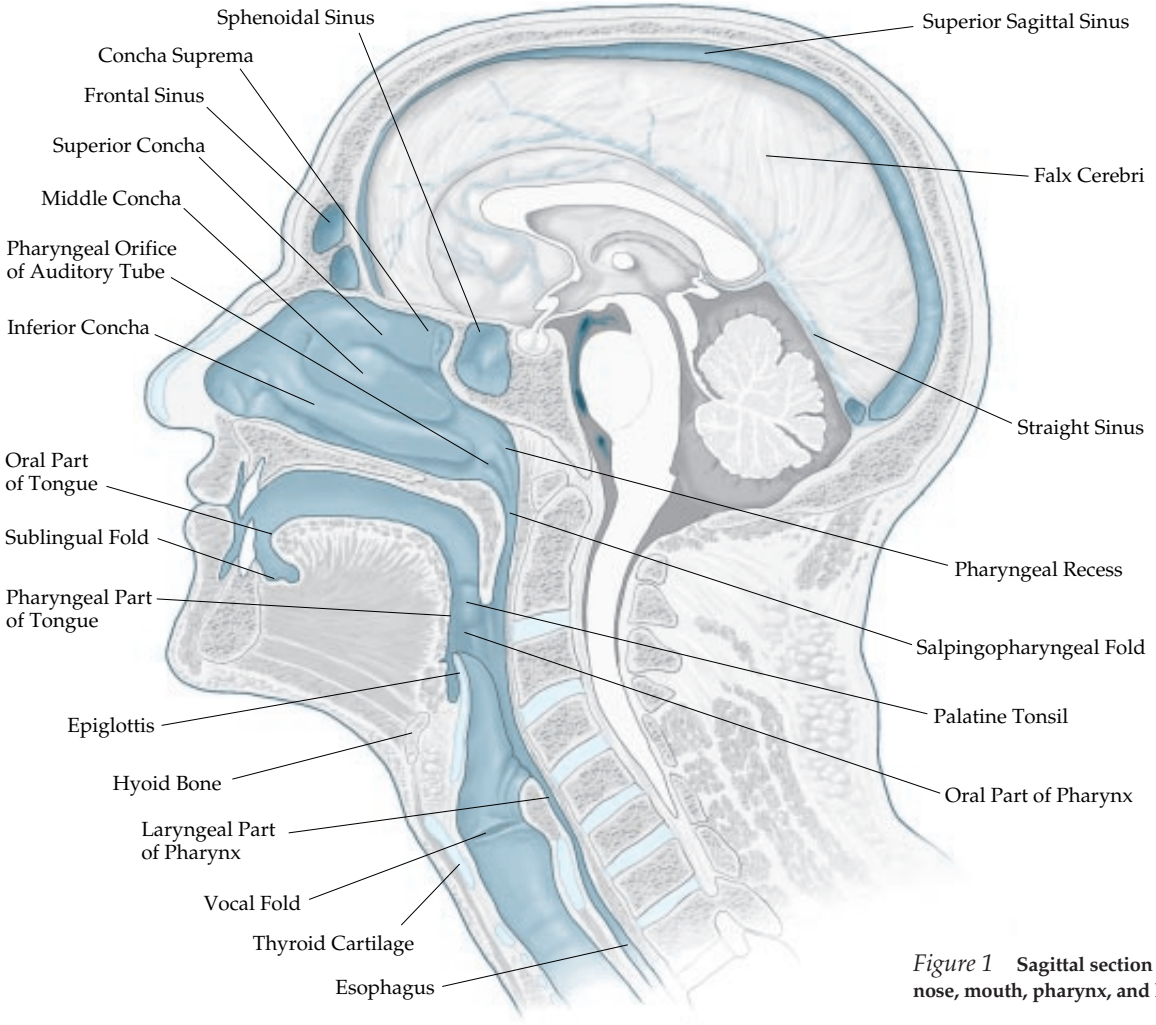


Figure 1 **Sagittal section through the nose, mouth, pharynx, and larynx.**

Immunosuppression increases the frequency and severity of sinusitis. Sinusitis is common in patients with AIDS and is often severe and refractory to treatment.

CLINICAL VARIANTS

The sinuses may be involved singly or, more often, in combination [*see Figure 2*]. Frontal sinusitis and maxillary sinusitis are most common in adults; ethmoiditis is most common in children.

Frontal sinusitis produces pain and tenderness over the lower forehead and purulent drainage from the middle meatus of the nasal turbinates. Maxillary sinusitis produces pain and tenderness over the cheeks. Pain is often referred to the teeth, and the hard palate may be edematous in severe cases. Purulent drainage is present in the middle meatus. Patients with ethmoidal sinusitis complain of retro-orbital pain and may have pain, tenderness, and even erythema over the upper lateral aspect of the nose. The anterior ethmoidal cells drain through the middle meatus, whereas the posterior cells drain through the superior meatus. Isolated sphenoid sinusitis, which occurs in only 3% of patients with sinusitis, is often misdiagnosed. This disorder has traditionally been considered to cause pain at the occiput or vertex; in fact, frontotemporal, retro-orbital, or facial pain is more common. Photophobia and tearing may be present. Purulent drainage is present in the superior meatus.

DIAGNOSIS

Clinical Features and Imaging Studies

Acute sinusitis The diagnosis of acute sinusitis can usually be established on clinical grounds. Useful criteria are a poor response to decongestants; maxillary, facial, or tooth pain; and a duration of more than 7 days.[5] Purulent nasal discharge is not specific for sinusitis and may occur in viral nasopharyngitis.[6] The diagnosis of sinusitis can be confirmed by the finding of opacity when the frontal or maxillary sinuses are transilluminated or by radiographic findings such as mucosal thickening, sinus opacification, or air-fluid levels.[7]

Chronic sinusitis The symptoms of chronic sinusitis are nasal congestion and discharge. Pain and headache are usually mild or absent, and fever is uncommon. Bone erosion may be present in chronic sinusitis. CT scans provide much more information than the traditional sinus x-ray series. Radiologic studies are essential in the diagnosis of severe, refractory, nosocomial, or complicated cases of sinusitis but are not necessary in routine cases.

Bacteriology

Because of technical difficulties in obtaining valid cultures, the bacteriology of sinusitis has been incompletely defined. Nasal

cultures are unreliable because they correlate poorly with actual sinus fluid. Sinus punctures reveal that the most common pathogens in acute sinusitis are *Haemophilus influenzae*, pneumococci, streptococci, and *Moraxella catarrhalis*.[5,7] *Staphylococcus aureus* has been isolated in nasal cultures from some patients with acute sinusitis, but the organisms are probably contaminants rather than the true causative agents. Nosocomial sinusitis is typically polymicrobial; *S. aureus* and gram-negative bacilli are often present.[8]

Table 1 Major Microbial Flora of the Upper Respiratory Tract

Organism	*Oral Cavity*	*Nasopharynx, Tonsils*	*Epiglottis, Larynx*	*Paranasal Sinuses, Middle Ear*
Gram-positive cocci				
Coagulase-negative staphylococci	NF	NF		UP
S. aureus		CC	UP	UP
Pneumococci		CC	UP	CP
Group A streptococci		CC, CP	UP	CP
Groups C and G streptococci		CC, UP		
Other streptococci	NF	NF		CP
Gram-positive bacilli				
Diphtheroids	NF	NF		
Corynebacterium diphtheriae		UC, UP	UP	
Arcanobacterium haemolyticum				
Gram-positive anaerobes				
Anaerobic streptococci	NF	NF		UP*
Anaerobic diphtheroids	NF	NF		
Actinomyces species	NF, UP			
Gram-negative cocci				
Neisseria meningitidis		CC, UP		
N. gonorrhoeae		UC, UP		
Other *Neisseria* species	NF	NF		UP
Moraxella catarrhalis	NF	NF		CP
Gram-negative coccobacilli				
Haemophilus influenzae		CC, UP	CP	CP
Other *Haemophilus* species	NF	NF		
Gram-negative bacilli				
Enterobacteriaceae	UC	UC	UP	
Pseudomonas aeruginosa	UC	UC		
Gram-negative anaerobes				
Prevotella species	NF, UP			UP*
Fusobacterium species	NF, UP			UP*
Veillonella species	NF			
Mycobacterium				
M. tuberculosis	UP	UP	UP	UP
Spirochetes				
Borrelia species	NF, UP	NF, UP		
Treponema pallidum	UP	UP		
Fungi				
Candida albicans	NF	NF, UP		
Aspergillus species				UP
Mucor species				UP
Chlamydia				
C. pneumoniae (TWAR)		UP		
Mycoplasma				
M. pneumoniae		UC, UP	UP	UC, UP
Viruses				
Epstein-Barr virus		CP, CC		
Herpes simplex virus	CC, CP	UP		
Influenza virus		CP	CP	
Parainfluenza virus		CP	CP	
Adenovirus		CP, UC	CP	
Coxsackievirus	CP	CP	CP	
Rhinovirus		CP	CP	
Respiratory syncytial virus		CP	CP	

*Anaerobic bacteria predominate in the normal flora of the upper respiratory tract. They are uncommon pathogens in acute infections but are common pathogens in chronic sinusitis and otitis, as well as in abscesses, soft tissue infections, and periodontal infections; aerobic pathogens participate synergistically in most of these processes.
CC—Common colonizer: organism present in many individuals, often transiently, without causing disease; CP—Common pathogen: organism frequently causes disease in this location; NF—Normal flora: organism present in most healthy individuals; UC—Uncommon colonizer: organism present without causing disease but relatively uncommon; UP—Uncommon pathogen: organism infrequently causes disease in this location.

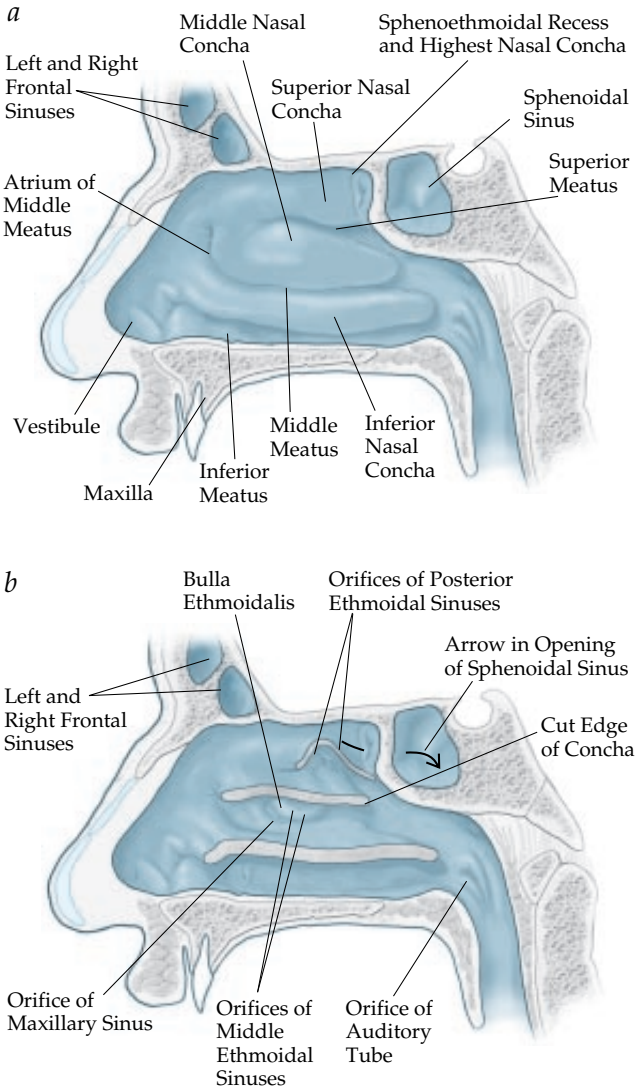


Figure 2 **(*a*) The lateral wall of the right half of the nasal cavity: internal aspect. (*b*) The three nasal conchae have been partially removed.**

Chronic sinusitis is usually a polymicrobial infection. Anaerobic organisms have been isolated in as many as 51% of patients with chronic sinusitis; anaerobic streptococci and *Prevotella* species predominate.[9] Anaerobes are also important in chronic sinusitis because they predominate in brain abscesses associated with sinus infections. *S. aureus* has been recovered from cultures in as many as 33% of patients with chronic sinusitis; enteric gram-negative bacilli may also be present.

Viruses seldom cause clinical sinusitis. Such fungi as *Aspergillus, Mucor,* and *Rhizopus* species produce invasive sinusitis in patients with leukemia or poorly controlled diabetes[10,11] and, occasionally, in normal hosts.[10] In immunocompromised patients, sinusitis may also be caused by *Candida, Alternaria, Pseudomonas, Nocardia, Legionella,* atypical mycobacteria, and certain parasites.[12]

DIFFERENTIAL DIAGNOSIS

Noninfectious processes must be considered in the differential diagnosis of sinusitis. Allergic or vasomotor rhinitis is by far the most common noninfectious cause of sinus symptoms. However, polyps, tumors, cysts, foreign bodies, and vasculitis (such as Wegener granulomatosis) occasionally produce symptoms of sinusitis.

TREATMENT

Medical therapy Acute sinusitis is treated with analgesics and topical heat for patient comfort. Decongestants are of paramount importance. Pseudoephedrine can be administered orally or by nasal spray. The danger of rebound after short-term use of nasal spray has probably been exaggerated. Patients should spray each nostril once and then wait a minute to allow the anterior nasal mucosa to shrink; a repeat spray will reach the upper and posterior mucosa, including the nasal turbinates and sinus ostia. This procedure can be repeated every 4 hours for several days when needed. Antihistamines are not indicated, because they thicken secretions and impair drainage.[5] Intranasal steroids are not necessary in most patients but may accelerate symptomatic improvements in some individuals[13]; however, they may be harmful. Nasal inhalation of warm, humidified vapor may relieve symptoms. Sleeping supine or with the head of the bed elevated 45° promotes sinus drainage; patients with unilateral sinusitis should be encouraged to sleep with the unaffected side on the pillow. Other ancillary measures that may be helpful are nasal saline irrigations and aerobic exercise, which promotes mucous flow. Mucoregulating agents are often recommended but have not been adequately studied.

Many patients with uncomplicated acute sinusitis respond well to decongestants and steam inhalations and do not need antibiotics.[5,14,15] Antibiotics should be used in moderately to seriously ill patients, in patients whose symptoms fail to respond to decongestants, and in patients who have complications. Older agents such as amoxicillin, doxycycline, and trimethoprim-sulfamethoxazole have produced good results.[5] Second-line agents that are often prescribed include amoxicillin–clavulanic acid, cefuroxime, clarithromycin, azithromycin, and levofloxacin; they are no more effective than the less expensive first-line drugs.[16,17] Because both *H. influenzae* and *M. catarrhalis* can produce β-lactamase, the use of antibiotics that resist these enzymes has attracted interest. However, antibiotics that are resistant to β-lactamase have not proved to be superior to those that are susceptible, possibly because comparative studies have been small and because many patients recover without antibiotic therapy. The optimal duration of antimicrobial therapy is no more certain than are the indications for therapy or the choice of agents[18]; 7- to 14-day courses of antibiotics are traditional, but one trial found 3 days of trimethoprim-sulfamethoxazole to be as effective as 10 days of therapy.[19] Clinical guidelines for the diagnosis and treatment of sinusitis are available.[20]

Surgical therapy Surgical intervention should be avoided in acute sinusitis unless the patient fails to respond to medical therapy and unless complications are present. Surgical drainage may be necessary in cases of chronic sinusitis or nosocomial sinusitis[8] and is mandatory in invasive fungal sinusitis.[10,11]

Functional endoscopic sinus surgery has provided an important advance in the surgical management of sinusitis. Patients with recurrent or chronic sinusitis who fail to respond to vigorous medical management may be candidates for this procedure. The patient should undergo a detailed preoperative evaluation, which usually includes a sinus CT scan. Surgery is performed using either local anesthesia with sedation or general anesthesia. A fiberoptic nasal endoscope is used to visualize the sinus ostia;

when obstruction is present, it can be relieved. This technique restores patency and allows normal ciliary action to transport mucus to the natural ostium. Complications are uncommon when the procedure is performed by a skilled surgeon. Nasal endoscopy can also be used as an alternative to sinus puncture when it is necessary to obtain specimens for culture.[21]

COMPLICATIONS

The potentially life-threatening complications of sinusitis have become uncommon in the antibiotic era. However, frontal sinusitis can lead to osteomyelitis of the frontal bones, especially in children. Patients have headache, fever, and a characteristic doughy edema, termed Pott puffy tumor, over the affected bone. The organisms involved are those responsible for the underlying sinusitis, although *S. aureus* is more common in osteomyelitis than in sinusitis. Treatment should include decongestants for the sinusitis, high-dose intravenous antibiotics for the osteomyelitis, and surgical debridement when bone destruction is substantial. It is important to rule out coexisting brain abscess of the frontal lobe, which may be clinically occult.

Because the orbit is bordered on three sides by the paranasal sinuses, orbital infection can result from sinusitis. This disorder is most frequently a complication of ethmoidal sinusitis; it is caused by direct extension of infection through the lamina papyracea. Orbital cellulitis usually begins with edema of the eyelids and rapidly progresses to ptosis, proptosis, chemosis, and diminished extraocular movements. Patients are usually febrile and acutely ill. Pressure on the optic nerve may lead to vision loss, which can be permanent; retrograde spread can lead to intracranial infection. Therapy with high-dose intravenous antibiotics is mandatory. Unless a specific pathogen can be isolated from the blood or from purulent drainage, antibiotics should cover both staphylococci and gram-negative rods. When an orbital abscess is present, surgical drainage may be necessary; CT and ultrasound studies aid this evaluation.

Retrograde spread along venous channels from the sinuses, the orbit, or the nose can produce septic cavernous sinus thrombophlebitis. Patients have high fever and appear toxic. Lid edema, proptosis, and chemosis are present. In cavernous sinus thrombophlebitis (but not in uncomplicated orbital cellulitis), third, fourth, and sixth cranial nerve palsies are prominent; the pupil may be fixed and dilated, and ophthalmoscopic examination may reveal venous engorgement and papilledema. Although at first the process is usually unilateral, spread across the anterior and posterior intercavernous sinuses makes it bilateral. Patients may exhibit alterations of consciousness, and lumbar puncture may reveal inflammatory changes in the cerebrospinal fluid. CT with a contrast agent or venography may help confirm the diagnosis. High-dose intravenous antibiotics, including antistaphylococcal drugs, are mandatory. Anticoagulants are not of established benefit and may even be harmful.

Sinusitis can lead to intracranial suppuration either by direct spread through bone or by spread through venous channels. Various syndromes, such as epidural abscess, subdural empyema, meningitis, and brain abscess, can result. Clinical findings, which vary greatly, range from the subtle personality changes associated with frontal lobe abscesses to headache, symptoms of elevated intracranial pressure, alterations of consciousness, visual symptoms, focal neurologic deficits, seizures, and, ultimately, coma and death. These are true medical emergencies requiring high-dose antibiotics and, except for meningitis, neurosurgical intervention [*see 7:XXXVI Bacterial Infections of the Central Nervous System*].

Otitis

ACUTE OTITIS MEDIA

Acute otitis media is one of the most common childhood infections; in the United States alone, the annual costs of the infection are estimated to be $3 billion to $5 billion.[22] However, the incidence of otitis media declines with age, and it is uncommon in adults. Purulent otitis media results when bacteria ascend from the nasopharynx to the normally sterile middle ear. Abnormal eustachian tube function or obstruction caused by viral or allergic nasopharyngitis is considered important in the pathogenesis[23]; heredity may also play a role.

Etiology

The most common causes of purulent otitis media are the pneumococcus, nontypable strains of *H. influenzae*, and *M. cartarrhalis*[24]; the previously important group A streptococci are now uncommon.[25] Other organisms that can cause acute otitis media are coagulase-negative staphylococci and anaerobic bacteria. Gram-negative bacilli and *S. aureus* can cause acute otitis media in neonates. Viruses and mycoplasmas are uncommon as the primary pathogens in acute otitis media, but in 41% of children with acute bacterial otitis, respiratory tract viruses are also present; respiratory syncytial virus is the most common causal organism, followed by parainfluenza and influenza viruses.[26] Purulent nosocomial otitis is uncommon and occurs in only 4% of patients who have undergone endotracheal intubation; gram-negative bacilli are the responsible agents in this setting.

Diagnosis

Pain and hearing loss are the classic presenting complaints of acute otitis media; fever is present in some cases**,** especially when pneumococci are responsible.[25] The cornerstone of the clinical diagnosis is a bulging tympanic membrane, with impaired mobility and obscuration of the bony landmarks. Tympanic membrane perforation and otorrhea may occur. Concurrent conjunctivitis suggests the presence of *H. influenzae*.[25]

Needle aspiration of the middle ear can be used to confirm the diagnosis of purulent otitis media and to identify the causative organism; however, it is seldom necessary, because the bacteriology of otitis media is well defined and the clinical course is easy to monitor. Routine nose or throat cultures are not diagnostically reliable.

Differential Diagnosis

The major disorder considered in the differential diagnosis of acute otitis media is serous otitis media. Fever and pain are absent in serous otitis media; although fluid is present in the middle ear, the tympanic membrane is usually retracted and the bony landmarks are preserved.

Treatment

Acute otitis media may be treated with analgesics, decongestants, and antibiotics. Except in the Netherlands, where only about 30% of patients with otitis media take antibiotics, nearly all affected patients in developed countries are treated with antibiotics. In the United States, 20% of all prescriptions for oral antibiotics are issued for the treatment of otitis media; the average child consumes 3 months' worth of antibiotics in the first 2 years of life.[27] As a result, *H. influenzae*, *M. catarrhalis*, and *Streptococcus pneumoniae*—the organisms most frequently

responsible for acute otitis media—are manifesting dramatically increased resistance to antibiotics.

The utility of antibiotics is being reappraised.[22,24,28,29] Their benefits appear to be modest; a meta-analysis concluded that to prevent one child from experiencing pain by 2 to 7 days after infection, 17 children must be treated with antibiotics.[30] Further studies are required to determine which patients are most likely to benefit from antibiotics,[31] which drugs are best, and how long therapy should be continued. A new approach that merits study is a delayed-therapy strategy, in which an antibiotic is prescribed when otitis media is diagnosed, but the parents of the child are encouraged to fill the prescription only if the child's condition has not improved after 72 hours.[32] Until more data are available, clinicians in the United States are likely to continue their traditional use of such drugs as ampicillin, amoxicillin, amoxicillin–clavulanic acid, trimethoprim-sulfamethoxazole, and the newer oral cephalosporins and macrolides. As in the case of sinusitis, drugs active against β-lactamase–producing bacteria have not proved to be superior to amoxicillin in the treatment of acute otitis media, perhaps because antibiotics are only slightly more effective than placebos[33]; 81% of patients with acute otitis media recover without any antibiotic therapy.[34] Although antibiotics are generally administered for 10 days, 5 days of treatment may be equally effective in uncomplicated cases.[35,36] A single intramuscular dose of ceftriaxone is also as efficacious as 10 days of oral therapy.[37] Myringotomy does not hasten recovery but is indicated for patients with intractable pain, progressive deafness, or early mastoiditis or for those who respond poorly to medical therapy.

Complications

The prognosis for acute otitis media is excellent. Chronic serous otitis media, hearing loss, and recurrent purulent otitis media are the most common complications. Recurrences are usually caused by reinfection rather than relapse or persistent infection; antibiotics may be used for chemoprophylaxis of recurrent otitis in children.[24] Tympanostomy tubes may be helpful in children with numerous recurrences.

Acute mastoiditis, which is now rare, develops in patients with inadequately treated otitis media. Presenting symptoms consist of fever, otalgia, and postauricular edema and tenderness. The tympanic membrane is abnormal in all cases; perforations and otorrhea are present in about half of cases. Mastoid x-rays are usually abnormal. The species of bacteria responsible for mastoiditis are the same as those that cause acute otitis media. Patients who fail to respond to antibiotic therapy within 48 hours and those with CNS complications should be considered for early mastoidectomy.

Other rare suppurative complications of acute otitis media include purulent labyrinthitis, meningitis, lateral sinus thrombosis, and brain abscess. Bacteremia can be documented in 3% of young febrile children with otitis media, but it does not appear to alter the prognosis or to warrant parenteral antibiotic therapy.

CHRONIC OTITIS MEDIA

Chronic otitis media, which can occur at any age, results from neglected or recurrent acute otitis media. Pain and fever are usually absent but can occur during sporadic flare-ups.[38] Diminished hearing and foul otorrhea are major symptoms.

Physical examination discloses perforation of the tympanic membrane. Central perforations of the pars tensa are associated with benign disease, but peripheral perforations may occur with invasive cholesteatomas. X-rays reveal sclerosis of the mastoid air cells and bone destruction. Various organisms can be cultured from middle-ear drainage, including *P. aeruginosa*, staphylococci, streptococci, and enteric gram-negative bacilli. Anaerobes are present in 16% of cases, almost always in mixed infections with aerobes. Systemic antibiotics are generally of little benefit, but antibiotic eardrops may be helpful; quinolones appear to be more effective and safer than aminoglycosides.[39] Surgery is required in advanced cases. Without therapy, chronic otitis media can cause the same intracranial suppurative complications seen in acute otitis media.

OTITIS EXTERNA

Otitis externa is a common, generally benign inflammatory disorder that is usually precipitated by excessive moisture in or trauma to the external auditory canal. Patients have pruritus or pain, or both, that may be severe. Crusting, inflammation, and discharge in the canal are typical findings. Pain resulting from movement of the external ear helps distinguish otitis externa from otitis media. A broad range of organisms, including gram-positive cocci, gram-negative bacilli, and fungi, can cause otitis externa. Excellent results are produced by topical therapy with eardrops containing polymyxin B, neomycin, and corticosteroids or eardrops containing acetic acid and corticosteroids.[40] Patients occasionally acquire true cellulitis of the external ear; such patients require systemic antibiotics and even debridement of infected cartilage.

Malignant otitis externa, an infection with *P. aeruginosa* that progressively invades the cartilage, soft tissue, and skull, is a rare condition that occurs in diabetic patients. Neurologic complications can be lethal. Prolonged, maximal parenteral therapy with combinations of antipseudomonal agents, such as tobramycin and piperacillin, is generally recommended. Monotherapy with intravenous ceftazidime and prolonged therapy with oral ciprofloxacin have been successful. Other antibiotics that may prove useful alone or in combination are aztreonam, cefepime, imipenem, and meropenem. Aggressive surgical debridement has been a mainstay of treatment but may be required less often in patients who are treated early and aggressively with antibiotics. CT scans are superior to magnetic resonance imaging for early diagnosis, but either technique can be used to monitor patients for bone destruction and neurologic complications; should these sequelae occur, debridement is required. Mortality is high but has decreased since 1990. *Aspergillus* is a rare cause of malignant otitis externa.

Pharyngitis

Although pharyngitis is one of the most common problems in clinical practice, a surprising number of practical questions remain unanswered.[41] Group A streptococci are the most therapeutically important cause of pharyngitis, although in terms of frequency, they cause as few as 5% of the cases of pharyngitis and as many as 38% of the cases of tonsillitis.

STREPTOCOCCAL PHARYNGITIS

Diagnosis

Four clinical criteria have been proposed as suggestive of group A streptococcal pharyngitis: tonsillar exudates, tender anterior adenopathy, absence of cough, and history of fever. Unfortunately, however, there is considerable overlap between streptococcal and viral pharyngitis.

Because streptococcal and viral pharyngitis cannot be distinguished reliably on clinical grounds,[42] laboratory studies are necessary to make a diagnosis of streptococcal pharyngitis. Throat cultures remain the standard method for identifying group A streptococci in the pharynx. In addition, rapid diagnostic tests suitable for office use are available. These procedures entail the extraction of streptococcal antigens from throat swabs and the rapid identification of the antigens through immunologic tests such as latex agglutination or enzyme-linked immunosorbent assay (ELISA). The sensitivity of these tests ranges from 77% to 95%; specificity ranges from 86% to 100%. Thus, these tests are acceptable for clinical practice, and the Infectious Disease Society of America no longer recommends throat cultures to confirm the results of rapid antigen tests.[43]

The diagnosis of streptococcal pharyngitis is suspected on clinical and epidemiologic grounds and confirmed by a rapid antigen test or throat culture. Although the laboratory recognition of group A streptococci is straightforward [*see 7:I Infections Due to Gram-Positive Cocci*], it is difficult to determine whether the organisms are responsible for a patient's symptoms. The carrier rate of group A streptococci in healthy persons approaches 20%, and even among patients with symptomatic pharyngitis, only 50% of those with positive throat cultures demonstrate a serologic response to streptococci. Except for a serologic determination, there is no proven method for distinguishing between the carrier state and active infection. Similarly, controversy exists about the need for obtaining cultures from asymptomatic family contacts of patients with streptococcal pharyngitis, for treating asymptomatic persons with small numbers of group A streptococci in their throat cultures, and for performing follow-up cultures after therapy. On balance, all of these procedures are probably unnecessary unless the patient is highly susceptible to rheumatic carditis because of previous acute rheumatic fever.[44,45]

Treatment

Fortunately, no uncertainty exists about the therapy for streptococcal pharyngitis in symptomatic patients. The disease is ordinarily self-limited, but early antibiotic therapy can alleviate symptoms,[46] as can simple analgesic antipyretics. Antibiotics prevent rheumatic fever and local suppurative complications and limit droplet spread. Because rheumatic fever is now rare in the United States and can be prevented even when therapy is delayed as long as 9 days after the onset of symptoms, treatment need not be started until rapid-antigen assay or throat culture results are available.[47] The goal of therapy is the eradication of group A streptococci from the nasopharynx. In more than 80% of patients, this result can be achieved with oral penicillin V, 250 mg four times a day for 10 days. Twice-a-day dosing is also effective in children and may improve compliance.[48] When patient compliance is in doubt, a single injection of 600,000 to 1.2 million units of benzathine penicillin G will be at least as effective. In patients who are allergic to penicillin, oral erythromycin is an excellent alternative, but it must be continued for 10 days. Tonsillectomy was once widely performed on children to prevent recurrent pharyngitis, but it has largely been abandoned.

Although only a minority of sore throats are caused by group A streptococci, 73% of adults who consult physicians for pharyngitis receive antibiotics; 68% of these prescriptions are for nonrecommended agents.[49] Five-day regimens of newer macrolides or cephalosporins are being studied in Europe, but these drugs are expensive, and the regimens are not currently recommended in the United States.[45]

OTHER FORMS OF BACTERIAL PHARYNGITIS

Other bacterial causes of pharyngitis are uncommon. Groups C and G streptococci can cause pharyngitis, but complications are rare. Gonococci transmitted by oral sex can produce sore throat, erythema, exudates, and lymphadenopathy; asymptomatic colonization of the pharynx may occur. Meningococci sometimes cause symptomatic pharyngitis but are more commonly found in throat cultures of asymptomatic persons. *H. influenzae* pharyngitis can be extremely painful but is rare in adults; epiglottitis is a life-threatening complication of this type of pharyngitis. In patients who have not been immunized, *Corynebacterium diphtheriae* can cause painful membranous pharyngitis, often characterized by dysphagia and prominent edema of the neck. *Arcanobacterium haemolyticum*, formerly known as *Corynebacterium haemolyticum*, can cause pharyngitis and scarlatiniform rash, particularly in teenagers and young adults; administration of penicillin or erythromycin produces rapid relief of symptoms. *Yersinia enterocolitica* is an uncommon cause of pharyngitis; the pharyngitis may be present without enteritis, particularly in adults.

Many other bacterial species can be cultured from the pharynges of both symptomatic and asymptomatic patients [*see Table 1*], but they almost never cause pharyngitis. In particular, although pneumococci and staphylococci commonly reside in the nasopharynx and can cause severe disease in other parts of the respiratory tract, they do not cause pharyngitis. *Mycoplasma pneumoniae* causes pharyngitis in children and adults, but this diagnosis is seldom made in the absence of pneumonitis. *Chlamydophila pneumoniae* (formerly *Chlamydia pneumoniae)* can also cause pharyngitis, which may be protracted.[50]

Treponema pallidum, a spirochete that is not part of the normal flora, can cause pharyngitis in patients with primary or secondary syphilis. Among the other causes of pharyngitis, *Mycobacterium tuberculosis* is very rare. *Candida albicans*, which is part of the normal mouth flora, can produce painful oropharyngeal moniliasis, or thrush, if antibiotics or debilitating illnesses (particularly HIV infection) upset microbial interactions and host defenses. This disease is characterized by a white, cheesy exudate that can be scraped off to demonstrate yeast forms by smear and culture. Nystatin oral suspension is therapeutically effective, but frequent, large doses may be required. Clotrimazole troches are also effective. Patients with HIV infection may require systemic therapy with fluconazole. Other forms of fungal pharyngitis are rare.

Deep Tissue Infections

Infections originating in the pharynx may extend by contiguous spread to the deep tissues of the pharynx and neck. In the neck, numerous fascial planes create a variety of potential spaces where infection can become loculated to form a phlegmon or a full-fledged abscess. Although these processes are now uncommon, prompt recognition is mandatory because antibiotics and surgical drainage are required to control infection and to prevent obstruction of the airway, invasion of vital neurologic and vascular structures, and spread of infection to the mediastinum and bloodstream.

PERITONSILLAR ABSCESS

Peritonsillar abscess, also called quinsy throat, is a complication of streptococcal tonsillitis most often seen in adolescents and young adults. Group A streptococci are the primary cause

of the condition, although most peritonsillar abscesses also harbor mixed oral bacteria, with a predominance of anaerobes. Patients have fever and sore throat, often with pain referred to the ear. Dysphagia prevents the patient from swallowing saliva, commonly causing drooling; edema and pain produce a characteristic muffled, so-called hot-potato voice. The affected tonsil is visibly displaced forward, downward, and toward the midline; the soft palate may be edematous. Trismus occurs in some patients. CT scans are helpful in diagnosis and management. Treatment consists of parenteral penicillin and surgery. The traditional approach consists of immediate incision and drainage, followed by tonsillectomy 4 to 6 weeks later; early tonsillectomy has also produced excellent results. Treatment with needle aspiration and oral antibiotics has also been successful; 80% to 92% of patients with peritonsillar abscess can be cured with this approach, thereby obviating hospitalization and surgery.

RETROPHARYNGEAL AND PARAPHARYNGEAL INFECTIONS

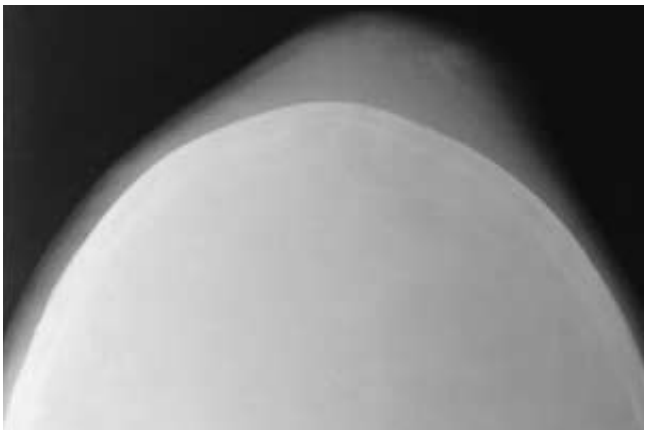
Retropharyngeal infections are most common in childhood because the lymph nodes in this region atrophy during adult life. Patients have fever and systemic toxicity, as well as neck pain, dysphagia, muffled voice, and respiratory stridor. Physical findings include erythema and bulging of the posterior wall of the pharynx. CT scans and lateral-view x-rays of the neck are extremely useful, invariably demonstrating soft tissue swelling and forward displacement of the larynx [*see Figure 3*]. Although the bacterial agent of retropharyngeal infections is not determined in many patients, streptococci and other mouth flora predominate. Penicillin is the traditional antibiotic of choice, but agents that provide a broader spectrum of antibacterial coverage may be justified until culture data are available. Surgical drainage is vital to prevent asphyxiation and extension of infection to the mediastinum. Infections of the retropharyngeal space must be distinguished from infections of the prevertebral space.

The prevertebral space lies posterior to the retropharyngeal space and is separated from it by the prevertebral fascia. Infection of the prevertebral space often originates from osteomyelitis of the cervical vertebrae. This infection is frequently caused by staphylococci and may lead to spinal cord damage unless treated through an approach that combines immobilization, antistaphylococcal antibiotics, and external drainage.

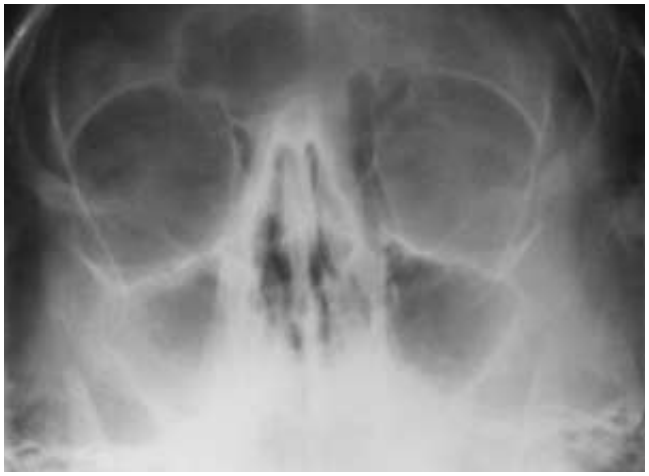
The parapharyngeal space is demarcated by the parotid gland and the internal pterygoid muscle laterally and by the superior constrictor muscle medially. The internal jugular vein, carotid artery, and cranial nerves IX, X, and XII pass through the parapharyngeal space. Infection can reach the parapharyngeal space from the pharynx or from parotid or dental foci. Patients have severe trismus, externally visible inflammation behind the angle of the jaw, and inflammation in the lateral wall of the pharynx, with medial displacement of the tonsil. Treatment consists of intravenous penicillin and drainage; CT-guided needle aspiration may be effective in some patients, but surgical drainage from behind the angle of the jaw is required in others.

Infections of the parapharyngeal space occasionally spread to the jugular vein and cause the syndrome of postanginal sepsis (Lemierre syndrome), which is characterized by the presence of septic phlebitis, septic pulmonary emboli, and anaerobic bacteremia. Pharyngitis and dental infections may also lead to postanginal sepsis. Facial swelling is an early diagnostic clue to this syndrome. CT is important for detecting abscesses, and ultrasonography is helpful for identifying jugular vein throm-

a



b



c

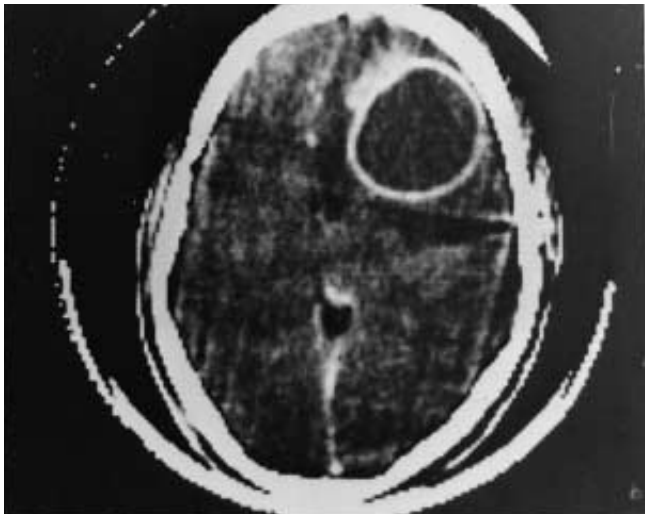


Figure 3 **(*a*) An oblique-view x-ray of the skull of a 16-year-old boy with fever, headache, and doughy edema of the forehead shows soft tissue swelling and a small area of underlying cortical irregularity indicative of osteomyelitis. (*b*) Sinus film reveals severe pansinusitis. (*c*) Because of subtle personality changes, a CT scan was taken; it revealed a large frontal lobe abscess. At surgery, 45 ml of pus was evacuated. Anaerobic streptococci were isolated from the subperiosteal collection and the brain abscess. Prolonged penicillin therapy and surgical management led to recovery.**

bosis. Anaerobic gram-negative bacilli, such as *Fusobacterium* species, are typically responsible for the condition. Management requires intravenous administration of penicillin; metronidazole and clindamycin are useful alternatives. Heparin, venous ligation, and surgical thrombectomy have also been advocated, but data do not allow firm guidelines to be established for these interventions.

LUDWIG ANGINA

Ludwig angina is a cellulitis of the submandibular, sublingual, and submental regions. In 86% of patients, the infection originates from a dental focus. Clinical features include fever, marked toxicity, and a rapidly progressive brawny edema in the floor of the mouth and the anterior neck. Elevation of the tongue impedes swallowing, and airway obstruction may be lethal. Streptococci and mouth flora are the most common etiologic agents, but *H. influenzae*, staphylococci, and gram-negative bacilli have also been implicated; broad antibiotic coverage may be necessary initially. Because endotracheal intubation can provoke laryngeal spasm, tracheostomy may be necessary to preserve the airway. Surgical decompression may be required.

Acute Epiglottitis (Supraglottitis)

No infection of the upper respiratory tract is more rapidly progressive or more lethal than acute epiglottitis, also known as supraglottitis. Acute epiglottitis occurs most commonly in children between 2 and 8 years of age and is more frequent in boys. The incidence of epiglottitis in childhood is declining rapidly in populations that have received *H. influenzae* type b vaccinations. Cases in adults appear to be increasing, however, perhaps because of improved diagnosis.

ETIOLOGY

The major cause of acute epiglottitis in children and adults is *H. influenzae* type b. Other pathogens, including pneumococci, streptococci, staphylococci, and *Klebsiella pneumoniae*, can produce an identical syndrome. Although viral epiglottitis is rare, this syndrome can occur secondary to infection with herpes simplex virus type 1.[51] In the immunocompromised host, additional organisms may be responsible, including *Candida* and *Kingella kingae*; fortunately, such cases are rare.

DIAGNOSIS

Epiglottitis begins with startling rapidity; severe sore throat and fever progress rapidly to dysphagia, with retention of secretions and drooling. Systemic toxicity is marked; if the illness is not treated, dyspnea and progressive respiratory obstruction occur in a matter of hours.

Although uncommon, acute epiglottitis in adults often extends to the supraglottic structures, and edema may be more prominent than acute inflammation. Whereas children with acute epiglottitis usually present with respiratory distress, adults may first complain of pharyngeal symptoms (severe pain and odynophagia) before developing respiratory symptoms.[52] The course of the disease in adults is slower than it is in children, often extending over several days because of the larger diameter of the adult airway.

Acute epiglottitis is a medical emergency. The key to the diagnosis is a swollen, edematous, cherry-red epiglottis. Simple inspection of the pharynx is usually unrewarding. Furthermore, any instrumentation, even a tongue blade, can provoke spasm and total airway obstruction, although adults are at lower risk for this complication. Therefore, unless acute respiratory distress is present, a lateral-view x-ray of the neck should be taken immediately. If the film does not demonstrate epiglottal edema, indirect laryngoscopy can be undertaken; if edema is present, however, the diagnosis is confirmed, and instrumentation is unnecessary. In all cases, patients must be monitored continuously for signs of respiratory obstruction.

DIFFERENTIAL DIAGNOSIS

In the differential diagnosis in children, it is important to distinguish between epiglottitis and croup caused by viral laryngitis with laryngeal and subglottal edema and airway obstruction. Croup usually affects younger patients and progresses more slowly than epiglottitis, and patients with croup present with hoarseness and a characteristic cough. In addition, in croup, the epiglottis is normal or only minimally inflamed on physical or radiographic examination. Cold, humidified oxygen is the mainstay of treatment of croup, and close observation is vital because emergency intubation or tracheostomy may be necessary. In children, bacterial tracheitis or membranous croup may present as severe laryngotracheitis, requiring antibiotic therapy and respiratory support. In patients of all ages, angioneurotic edema of the larynx, foreign bodies lodged in the upper airway, or impingement on the airway by various mass lesions may initially be confused with epiglottitis, but these conditions should be readily recognizable on physical examination and radiography.

TREATMENT

Although tracheostomy is the traditional means of securing a patent airway in patients with epiglottitis, nasotracheal intubation is safe and effective. As a result, direct laryngoscopy by an experienced observer who is prepared to proceed with intubation (or, if necessary, tracheostomy) can be relied on for diagnosis when taking x-rays would produce excessive delay.

Because there is no margin for error, initial therapy with high-dose intravenous antibiotics effective against *H. influenzae* (including penicillinase-producing strains), other gram-negative bacilli, and staphylococci is warranted. Many regimens are available, including cefuroxime or cefamandole, ampicillin-sulbactam, meropenem, imipenem-cilastatin, and a third-generation cephalosporin combined with a penicillinase-resistant penicillin. The choice of drugs depends on the results of throat and blood cultures and on ampicillin-sensitivity testing of *H. influenzae* isolates. A mist tent may be helpful. Steroids are sometimes advocated to reduce the edema, but their effectiveness has not been tested in controlled clinical trials. Above all, preservation of the airway is critical; nasotracheal intubation or tracheostomy is required in about half of all cases. If the airway can be protected while antibiotics take effect, most patients can be saved.

Miscellaneous Upper Respiratory Tract Infections

Many other infectious processes can invade the upper respiratory tract and adjacent regions of the head and neck. Common viral infections of the oropharynx are primary herpes simplex gingivostomatitis and adenovirus and coxsackievirus infections. Spirochetes play a role in ulcerative gingivitis and periodontitis. *Actinomyces israelii* can produce chronic burrowing infections of the gums and jaw, occasionally with secondary involvement of the respiratory tract. Although rare in the United States, rhinoscleroma, believed to be caused by *Klebsiella rhinoscleromatis*, can

occur in immigrants from developing countries or in HIV-infected persons. *S. aureus* may produce acute bacterial sialadenitis or suppurative parotitis; less commonly, gram-negative bacilli or even anaerobes produce parotitis in hospitalized patients.

Many bacterial species have been implicated in a variety of dental infections. Mixed infections with normal mouth flora can occur in debilitated patients. Fusobacteria and spirochetes can cause gingivitis (trench mouth) or necrotic tonsillar ulcers (Vincent angina); patients have foul breath, pain, and dirty-gray membranous inflammation that bleeds easily. Oral irrigations and penicillin are effective therapy. A similar combination of spirochetes and other bacteria can produce a rare but extremely serious invasive gangrene of the mouth, cancrum oris, which occurs only in malnourished infants or in patients with advanced malignant disease or immunosuppression. Anaerobes may play a role in chronic or recurrent tonsillitis.[38] Dental infection is the leading cause of necrotizing cervical fasciitis.

Many noninfectious processes in these regions can produce symptoms that mimic bacterial infection. Aphthous stomatitis may be the most common example. Drug reactions and primary dermatologic diseases cause oral and pharyngeal ulcerations. Vasculitis, carcinomas, lymphomas, and sarcoidosis of the upper airway can also mimic bacterial infection. Even subacute thyroiditis, which may produce fever and referred pain, may be mistaken for pharyngitis or otitis.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Snow E, Mottur-Pilson C, Gonzalez R, et al: Principles of appropriate antibiotic use for treatment of nonspecific upper respiratory tract infections in adults. Ann Intern Med 134:487, 2001

2. Fakih MG, Hilu RC, Savoy-Moore RT, et al: Do resident physicians use antibiotics appropriately in treating upper respiratory infections? A survey of 11 programs. Clin Infect Dis 37:853, 2003

3. Gwaltney JM Jr, Hendley JO, Phillips CD, et al: Nose blowing propels nasal fluid into the paranasal sinuses. Clin Infect Dis 30:387, 2000

4. George DL, Falk PS, Meduri GU, et al: Nosocomial sinusitis in patients in the medical intensive care unit: a prospective epidemiological study. Clin Infect Dis 27:463, 1998

5. Snow V, Mottur-Pilson C, Hickner JM, et el: Principles of appropriate antibiotic use for acute sinusitis in adults. Ann Intern Med 134:495, 2001

6. Gonzalez R: Purulent secretions and antibiotics for URI. J Gen Intern Med 14:151, 1999

7. Gwaltney JM Jr: Acute community-acquired sinusitis. Clin Infect Dis 23:1209, 1996

8. Talmor M, Li P, Barie PS: Acute paranasal sinusitis in critically ill patients: guidelines for prevention, diagnosis, and treatment. Clin Infect Dis 25:1441, 1997

9. Finegold SM, Flynn MJ, Rose FV, et al: Bacteriologic findings associated with chronic bacterial maxillary sinusitis in adults. Clin Infect Dis 35:428, 2002

10. Iwen PC, Rupp ME, Hinrichs SH: Invasive mold sinusitis: 17 cases in immunocompromised patients and review of the literature. Clin Infect Dis 24:1178, 1997

11. DeShazo RD, Chapin K, Swain RE: Fungal sinusitis. N Engl J Med 337:254, 1997

12. Dunand VA, Hammer SM, Rossi R, et al: Parasitic sinusitis and otitis in patients infected with human immunodeficiency virus: report of five cases and review. Clin Infect Dis 25:267, 1997

13. Dolor RJ, Witsell DL, Hellkamp AS, et al: Comparison of cefuroxime with or without intranasal fluticasone for the treatment of rhinosinusitis. The CAFFS trial: a randomized controlled trial. JAMA 286:3097, 2001

14. van Buchem FL, Knottnerus JA, Schrijnemaekers VJJ, et al: Primary-care–based randomised placebo-controlled trial of antibiotic treatment in acute maxillary sinusitis. Lancet 349:683, 1997

15. Bucher HC, Tschudi P, Young J, et al: Effect of amoxicillin-clavulanate in clinically diagnosed acute rhinosinusitis: a placebo-controlled, double-blind, randomized trial in general practice. Arch Intern Med 163:1793, 2003

16. de Ferranti SD, Ioannidis JPA, Lau J, et al: Are amoxycillin and folate inhibitors as effective as other antibiotics for acute sinusitis? A meta-analysis. BMJ 317:632, 1998

17. Piccirillo JF, Mager DE, Frisse ME, et al: Impact of first-line vs second-line antibiotics for the treatment of acute uncomplicated sinusitis. JAMA 286:1849, 2001

18. Gwaltney JM Jr, Wiesinger BA, Patrie JT: Acute community-acquired bacterial sinusitis: the value of antimicrobial treatment and the natural history. Clin Infect Dis 38:227, 2004

19. Williams JW Jr, Holleman DR Jr, Samsa GP, et al: Randomized controlled trial of 3 vs 10 days of trimethoprim/sulfamethoxazole for acute maxillary sinusitis. JAMA 273:1015, 1995

20. The evidence base for diagnosis and treatment of uncomplicated acute bacterial sinusitis. ACP Online. American College of Physicians, Philadelphia www.acponline.org/governors/sci-policy/sinusitis.htm

21. Talbot GH, Kennedy DW, Scheld WM, et al: Rigid nasal endoscopy versus sinus puncture and aspiration for microbiological documentation of acute bacterial maxillary sinusitis. Clin Infect Dis 33:1668, 2001

22. Rovers MM, Schilder AGM, Ziethuis GE, et al: Otitis media. Lancet 363:465, 2004

23. Hendley JW: Otitis media. N Engl J Med 347:1169, 2002

24. O'Neill P: Acute otitis media. BMJ 319:833, 1999

25. Palmu AAI, Herva E, Savolainen H, et al: Association of clinical signs and symptoms with bacterial findings in acute otitis media. Clin Infect Dis 38:234, 2004

26. Heikkinen T, Thint M, Chonmaitree T: Prevalence of various respiratory viruses in the middle ear during acute otitis media. N Engl J Med 340:260, 1999

27. Hirschmann JV: Methods for decreasing antibiotic use in otitis media. Lancet 352:672, 1998

28. Froom J, Culpepper L, Jacobs M, et al: Antimicrobials for acute otitis media? A review from the International Primary Care Network. BMJ 315:98, 1997

29. Culpepper L, Froom J: Routine antimicrobial treatment of acute otitis media: Is it necessary? JAMA 278:1643, 1997

30. Del Mar C, Glasziou P, Hayem M: Are antibiotics indicated as initial treatment for children with acute otitis media? A meta-analysis. BMJ 314:1526, 1997

31. Little P, Gould C, Moore M, et al: Predictors of poor outcome and benefits from antibiotics in children with acute otitis media: pragmatic randomized trial. BMJ 325:22, 2002

32. Little P, Gould C, Williamson I, et al: Pragmatic randomized controlled trial of two prescribing strategies for childhood acute otitis media. BMJ 322:336, 2001

33. Damoiseaux RAMJ, van Balen FAM, Hoes AW, et al: Primary care based randomized, double blind trial of amoxicillin versus placebo for acute otitis media in children aged under 2 years. BMJ 320:350, 2000

34. Rosenfeld RM, Vertrees JE, Carr J, et al: Clinical efficacy of antimicrobial drugs for acute otitis media: meta-analysis of 5400 children from thirty-three randomized trials. J Pediatr 124:355, 1994

35. Kozyrskyl AL, Hildes-Ripstein GE, Longstaffe EA, et al: Treatment of acute otitis media with a shortened course of antibiotics. JAMA 279:1736, 1998

36. Pichichero ME: Changing the treatment paradigm for acute otitis media in children. JAMA 279:1748, 1998

37. Barnett ED, Teele DW, Klein JO, et al: Comparison of ceftriaxone and trimethoprim-sulfamethoxazole for acute otitis media. Pediatrics 99:23, 1997

38. Acuin J: Chronic suppurative otitis media. BMJ 325:1159, 2002

39. Ghosh S, Panarese A, Parker AJ, et al: Quinolone ear drops for chronic otitis media: they are safer and more effective than aminoglycosides. BMJ 321:126, 2000

40. van Balen FA, Smit WM, Zuithoff NP, et al: Clinical efficacy of three common treatments in acute otitis externa in primary care: randomized controlled trial. BMJ 327:1201, 2003

41. Ebell MH, Smith MA, Barry HC, et al: The rational examination: does this patient have strep throat? JAMA 284:2912, 2000

42. Bisno AL: Acute pharyngitis. N Engl J Med 344:205, 2001

43. Bisno AL: Diagnosing strep throat in the adult patient: do clinical criteria really suffice? Ann Intern Med 139:150, 2003

44. Snow V, Mottur-Pilson C, Cooper RJ, et al: Principles of appropriate antibiotic use for acute pharyngitis in adults. Ann Intern Med 134:506, 2001

45. Bisno AL, Gerber MA, Gwaltney JM, et al: Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. Clin Infect Dis 35:113, 2002

46. Zwart S, Sachs APE, Ruijs GHM, et al: Penicillin for acute sore throat: randomized double blind trial of seven days versus three days treatment or placebo in adults. BMJ 320:150, 2000

47. Neuner JM, Hamel MB, Phillips RS, et al: Diagnosis and management of adults with pharyngitis: a cost-effective analysis. Ann Intern Med 139:113, 2003

48. Lan AJ, Colford JM Jr: The impact of dosing frequency on the efficacy of 10-day penicillin or amoxicillin therapy for streptococcal tonsillopharyngitis: a meta-analysis. Pediatrics 105:E19, 2000

49. Linder JA, Stafford RS: Antibiotic treatment of adults with sore throat by community primary care physicians: a national survey, 1989–1999. JAMA 286:1181, 2001

50. Falck G, Engstrand I, Gad A, et al: Demonstration of *Chlamydia pneumoniae* in patients with chronic pharyngitis. Scand J Infect Dis 29:585, 1997

51. Bengualid V, Keesari S, Kandiah V, et al: Supraglottitis due to herpes simplex virus type 1 in an adult. Clin Infect Dis 22:382, 1996

52. Kucera CM, Silverstein MD, Jacobson RM, et al: Epiglottitis in adults and children in Olmstead County, Minnesota, 1976 through 1990. Mayo Clin Proc 71:1155, 1996

Acknowledgment

Figures 1 and 2 Tom Moore.

XX PNEUMONIA AND OTHER PULMONARY INFECTIONS

HARVEY B. SIMON, M.D., F.A.C.P.

Pulmonary infections span a wide spectrum, ranging from self-limited to life-threatening and from acute to chronic. This chapter details the pathophysiology, epidemiology, general features, and treatment of pulmonary infections, particularly bacterial pneumonia [*see Table 1*].

Pneumonia

Although overall hospitalization rates are declining, hospitalizations for acute lower respiratory tract infections have increased steadily since 1980.[1] Taken together, pneumonia and influenza rank as the sixth leading cause of death in the United States and lead all other infectious diseases in this respect. Antibiotics have greatly modified the natural history of pneumonia and have sharply reduced the overall case-fatality rate. At the same time, the widespread use of antimicrobial agents has led to the emergence of drug-resistant strains, thereby altering and expanding the range of pathogens responsible for pneumonia, especially in hospitalized patients. The growing population of patients with chronic obstructive pulmonary disease (COPD) and other debilitating illnesses and the use of respiratory therapy and immunosuppressive drugs have contributed to the increasing incidence of nosocomial and opportunistic pneumonias, which have a very high mortality.

PATHOPHYSIOLOGY

Host Defense Mechanisms

The lung is normally sterile. In healthy people, an intricate series of defense mechanisms maintain that sterility in the face of heavy bacterial colonization of the upper respiratory tract, inhalation of thousands of bacteria in droplet nuclei each day, and nearly universal aspiration of upper airway secretions during normal sleep each night.[2,3] Defects in host defense mechanisms account for most cases of pneumonia [*see Table 2*].

Transmission of Organisms to Lungs

Inhalation of aerosolized droplets accounts for the transmission of respiratory viruses, such as influenza virus, which cause highly contagious infections that often occur in epidemics.[4] Other nonviral agents produce pneumonia through similar means of spread. Such agents include mycoplasmas, which are transmitted from person to person and cause primary atypical pneumonia; *Coxiella burnetii*, which is transmitted from livestock and causes Q fever; and *Chlamydophila* (formerly *Chlamydia*), which is transmitted from birds (*C. psittaci*) or humans (*C. pneumoniae*). *Mycobacterium tuberculosis* is spread from person to person by aerosolized droplets. The organisms that are responsible for causing the systemic mycoses are probably inhaled from sources in nature. Of bacteria that cause pneumonia, *Legionella pneumophila* is the species most likely to be spread by inhalation of aerosolized organisms that originate in contaminated fresh water.

Pneumococci are spread from person to person by aerosolized droplets, but pneumococcal pneumonia is not highly contagious and is caused in many cases by aspiration of nasopharyngeal organisms, the second major mechanism of infection.[4] Aspiration of nasopharyngeal organisms occurs in nearly all persons during sleep and is probably responsible for most bacterial pneumonias, including staphylococcal and gram-negative bacillary pneumonias; it certainly accounts for the necrotizing pneumonitis that results from the aspiration of mixed mouth flora.

Hematogenous seeding, the third and least common mechanism for pneumonia, accounts for occasional cases of staphylococcal pneumonia that complicate tricuspid valve endocarditis or septic thrombophlebitis. This mechanism is also responsible for various gram-negative bacillary pneumonias in patients with bacteremia.

Tissue Responses

Once organisms succeed in bypassing host defense mechanisms to arrive at the alveoli, a variety of tissue responses may ensue, depending on the nature of the pathogen and on the integrity of the host inflammatory response. Although the inflammatory response is essential for the control of infection, it can produce tissue damage, impair ciliary action, and impede phagocytosis.[3]

The inflammatory response to *Streptococcus pneumoniae* or *Haemophilus influenzae* often produces lobar consolidation, but these infections rarely result in tissue necrosis. In contrast, staphylococci and many gram-negative bacilli often produce necrosis, which can lead to cavitation and even frank abscess formation; a peribronchial distribution is characteristic, but lobar consolidation may occur. Viruses generally produce interstitial inflammation rather than air-space exudates. The infection is usually bilateral and causes diffuse alveolar damage and interstitial edema. Similar tissue responses may be initiated by *My-*

Table 1 Major Causes of Pulmonary Infection

Gram-positive cocci: *Streptococcus pneumoniae*, *S. pyogenes*, other streptococci, staphylococci
Gram-positive bacilli: *Bacillus anthracis*
Gram-negative cocci: *Neisseria meningitidis*, *Moraxella catarrhalis*
Gram-negative coccobacilli: *Haemophilus influenzae*
Gram-negative bacilli: *Klebsiella pneumoniae, Pseudomonas* species, *Escherichia coli, Proteus, Serratia* species, *Acinetobacter, Yersinia pestis, Francisella tularensis*, *Enterobacter* species, *Prevotella*, *Legionella* species
Mixed flora (aspiration pneumonia)
Mycobacteria: *Mycobacterium tuberculosis, M. avium* complex
Fungi: *Histoplasma, Coccidioides, Blastomyces*, *Cryptococcus, Candida, Aspergillus,* Mucoraceae
Parasites: *Pneumocystis carinii*, *Toxoplasma gondii*
Mycoplasmas: *Mycoplasma pneumoniae*
Chlamydophila (chlamydiae): *C. pneumoniae, C. psittaci, C. trachomatis*
Rickettsia-like organisms: *Coxiella burnetii*
Viruses: influenza virus, parainfluenza virus, adenovirus, respiratory syncytial virus, rhinovirus, measles virus, varicella-zoster virus, cytomegalovirus

Table 2 Host Defense Mechanisms against Pulmonary Infection

Mechanism	*Modifying Factors*
Normal flora of upper respiratory tract	Antibiotic therapy, respiratory therapy, hospitalization
Aerodynamic properties of upper respiratory tract	Tracheal intubation
Protective reflexes (e.g., cough, sneeze, gag, bronchoconstrictor)	Sedatives and hypnotics, alcohol and drug abuse, neurologic disorders, age and debility, tracheal intubation (gastrointestinal disorders)
Mucous carpet that entraps particles	Viral infections, smoking, chemical irritants, dehydration
Ciliary action, mucociliary transport	Smoking, viral infections, advanced age, aspirin (?)
Antibacterial substances in respiratory secretions (e.g., lysozyme, lactoferrin, α_1-antitrypsin, IgA)	None
Fibronectin in respiratory secretions (competitively inhibits adherence of gram-negative bacilli)	Decreased fibronectin levels in seriously ill patients
Free drainage of tracheobronchial tree	Foreign bodies, obstructing tumors, bronchostenosis
Rich blood supply	Vascular obstruction (e.g., emboli)
Alveolar macrophages	Viral infections, smoking (both increase macrophage numbers but impair function)
Lymphatic tissue	Cytotoxic therapy
Humoral immunity (B cells, IgG and IgA antibodies, complement, polymorphonuclear neutrophils)	Immunosuppressive disorders and therapy
Cellular immunity (T cells, lymphokines, mononuclear phagocytes)	Immunosuppressive disorders and therapy

coplasma, *Chlamydophila*, and *Legionella* species; by gram-negative bacteremia (shock lung); and by other causes of the acute respiratory distress syndrome. Mycobacteria and fungi typically evoke a slow granulomatous response.

EPIDEMIOLOGY AND ETIOLOGY

Community-Acquired Pneumonia

Like other respiratory tract illnesses, pneumonia is most common in the winter because of the seasonal increase in viral infections and the close contact of persons confined indoors. Community-acquired pneumonias are a major problem in the United States, with an estimated four million cases occurring annually.[5] About one million cases require hospitalization,[5] and at least 60,000 result in death.[6] The mortality of community-acquired pneumonia ranges from less than 1% in patients who are not ill enough to require hospitalization to 13.7% for hospitalized patients, 19.6% for bacteremic patients, and 36.5% for patients admitted to intensive care units.[7] Clinical and laboratory data can be used to determine which patients are at greatest risk for death and thus require hospitalization and aggressive therapy. Comorbidity is the strongest risk factor, with neoplastic disease, neurologic disease, and alcoholism being particularly worrisome.[8] Advanced age is another strong predictor of risk,[6] in part because older patients often underreport symptoms.[7-10] Physical findings of high fever, tachypnea, confusion, hypoxia, and hypotension also portend an adverse result.[11] The presence of extensive radiographic abnormalities, especially bilateral pleural effusions, is associated with higher risk, as are laboratory abnormalities such as hypoxia, azotemia, acidosis, hyponatremia, and hypophosphatemia. Postobstructive, aspiration, gram-negative, and staphylococcal pneumonias produce a high mortality. Patients lacking these adverse prognostic indicators have a low risk of death and can usually be treated successfully as outpatients. Patients hospitalized for pneumonia have a greater than fivefold likelihood of requiring subsequent hospitalization for pneumonia than patients with other serious illnesses. They also have substantially higher long-term mortality than age-matched control subjects.[12,13]

In the preantibiotic era, pneumonia was nearly synonymous with *S. pneumoniae* infection. Pneumococci still account for 30% to 60% of all community-acquired pneumonias for which an etiology can be determined.[14] Pneumococci are particularly likely to be responsible for community-acquired pneumonias severe enough to require hospitalization and for pneumonia in persons older than 60 years.[15]

The second most common bacterial cause of community-acquired pneumonia is *H. influenzae,* which accounts for about 10% of cases[14]; patients with COPD are particularly vulnerable. Although infection with *Moraxella catarrhalis* is much less common than infection with *H. influenzae,* it is being recognized increasingly as a cause of community-acquired pneumonia. Like *H. influenzae, M. catarrhalis* has a predilection for patients with cardiopulmonary disease. In rare cases, *M. catarrhalis* causes fulminant pneumonia, bacteremia, or both.

Staphylococci and gram-negative bacilli are much less common but more serious causes of community-acquired respiratory infections. Significant predisposing conditions are required for these organisms to produce pneumonia. In the community setting, staphylococcal pneumonia usually follows influenza. Gram-negative pneumonias in the community setting are most common in patients who have recently been hospitalized and treated with antibiotics, in smokers and others with chronic lung disease, and in immunosuppressed patients.[2,16] Exposure to aerosols of contaminated water is an additional risk factor for infection with *Pseudomonas aeruginosa,*[17] and alcoholism predisposes to *Klebsiella* pneumonia. Meningococcal pneumonia is rare.[18] A variety of other bacteria, including *L. pneumophila,* can cause pneumonia in the community setting.[15] Aspiration of mixed mouth flora is responsible for infection in some patients.

In about half the patients with community-acquired pneumonia, the etiologic agent cannot be identified.[19] Especially in younger patients, many of these infections result from so-called atypical agents, which lack the cell wall structure that characterizes ordinary bacteria. In a study of patients with a mean age of 41 years, for example, *M. pneumoniae* accounted for 22.8% of community-acquired pneumonias, and *C. pneumoniae* for 10.7%; in addition, influenza A accounted for 2.7% .[19] *C. pneumoniae* is also increasingly being recognized as a cause of community-acquired pneumonia in adults with COPD [*see 14:III Chronic Ob-*

structive Diseases of the Lung]. Respiratory tract viruses, including respiratory syncytial virus, adenoviruses, and influenza or parainfluenza viruses,[20] can also cause community-acquired pneumonias in persons of all ages.[1,15]

Hospital-Acquired Pneumonia

Pneumonia is the second most common nosocomial infection in the United States[21]; about 200,000 cases occur annually, accounting for 17.8% of all hospital-acquired infections and 40,000 to 70,000 deaths. Risk factors for nosocomial infections include aspiration, COPD or other chronic severe illnesses, thoracic and upper abdominal surgery, and treatment in an ICU. Patients who require mechanical ventilation are particularly at risk; pneumonia develops in 9% to 24% of patients who require intubation for more than 48 hours.[22] Ventilated patients who acquire nosocomial pneumonia have a much higher mortality (54%) than comparably ill ventilated patients who do not acquire pneumonia (27%).

The bacterial etiologies of hospital-acquired pneumonias are very different from those of community-acquired pneumonias. Many nosocomial pneumonias are polymicrobial; gram-negative bacilli are isolated in 47% of patients, anaerobes in 35%, and *Staphylococcus aureus* in 26%. In contrast, pneumococci account for no more than 10% of hospital-acquired pneumonias. Other organisms that are associated with community-acquired infections can occasionally cause pneumonia in the hospital setting; such organisms include *Legionella* species, *M. pneumoniae,* and *C. pneumoniae*.[23]

Pseudomonas, Klebsiella, and *Escherichia coli* are the most commonly implicated causes of gram-negative pneumonias. Nosocomial gram-negative pneumonias often occur in patients with serious underlying diseases; as a result, mortality is as high as 30% to 50%. Previous antibiotic administration, respiratory therapy, chronic illness, and confinement to bed predispose to oropharyngeal colonization with gram-negative bacilli, which occurs in up to 45% of ICU patients and precedes pneumonia in most cases. Upper intestinal colonization has also been implicated, but its importance is uncertain. Although most nursing home residents are elderly and chronically ill, only 14% of them become colonized with gram-negative bacilli; oropharyngeal colonization in this population is transient and does not appear to predispose to gram-negative pneumonia. Hence, pneumonia in these patients can be approached as a community-acquired infection rather than a hospital-acquired infection.[24,25]

Pneumonia in Immunosuppressed Patients

Immunosuppressed patients, particularly those with AIDS,[26,27] are vulnerable to a broad range of pulmonary pathogens. In addition to being susceptible to the many organisms that produce community- and hospital-acquired pneumonias, these patients are susceptible to many opportunistic microbes that are unlikely to cause pneumonia in immunologically competent hosts.[28] Such organisms include bacteria (e.g., *Pseudomonas, Nocardia,* and *Legionella* species), mycobacteria (e.g., *M. avium* complex), viruses (e.g., cytomegalovirus and herpesvirus), fungi (e.g., *Candida, Aspergillus*, and *Mucor* species), and protozoa (e.g., *Pneumocystis carinii* and *Toxoplasma gondii*). As a result, immunosuppressed patients require an aggressive approach to diagnosis and therapy [*see 7:X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas, 7:XI Infections Due to Brucella, Francisella, Yersinia pestis, and Bartonella, and 7:XXXVIII Mycotic Infections in the Compromised Host*].

DIAGNOSIS

Clinical Features

Classic signs and symptoms of pneumonia include cough, sputum production, chest pain, fever, chills, hypoxia, and dyspnea. Although the results of physical examination in patients with typical pneumonias are often nonspecific, the examination may reveal rales, rhonchi, or bronchial breath sounds, as well as percussion dullness over the involved segments of the lung. Pleural effusions may accompany pneumonia. The chest x-ray shows infiltrates.

Nonbacterial and bacterial pneumonias have differing clinical presentations. Although both types of pneumonia can affect persons of all ages, nonbacterial pneumonias are most common in older children and young adults. Patients with viral, mycoplasmal, or chlamydial pneumonias will often complain of a severe hacking cough, but substantial sputum production is unusual. Sputum production is also minimal in Legionnaires disease, but these patients are usually sicker than those with nonbacterial pneumonias [*see 7:X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas*].

Patients with bacterial pneumonias are more likely to have copious sputum production—as well as an abrupt onset of illness, high temperatures, chills, and development of significant pleural effusions—than are patients with nonbacterial pneumonias.

On physical examination, a patient with bacterial pneumonia generally looks sicker than a patient with nonbacterial pneumonia, and chest examination of patients with bacterial pneumonia usually reveals signs of consolidation or at least localized rales and rhonchi. In contrast, the chest examination of patients with nonbacterial pneumonias typically shows only fine rales, and often, the physical findings are less extensive than the radiologic abnormalities.

Laboratory Studies

Patients with bacterial pneumonias are more likely to have polymorphonuclear leukocytosis. If the chest x-ray reveals lobar or segmental consolidation, abscess formation, or significant pleural effusion, bacterial pneumonia is more likely. A patchy infiltrate can occur in either process, but a true interstitial infiltrate suggests a nonbacterial etiology. Computed tomography is extremely helpful in patients with complex infections.

The sputum examination is central to the etiologic diagnosis of pneumonia.[29] The sputum of patients with bacterial pneumonia is typically thick and either green or brownish and is sometimes blood tinged. A good sputum specimen for microscopic examination and culture is crucial.[30] If the patient cannot expectorate spontaneously, pulmonary physiotherapy, intermittent positive pressure ventilation with humidified air, or nasotracheal suction may be used to obtain the specimen. The Gram stain of sputum from patients with bacterial pneumonia usually reveals abundant polymorphonuclear leukocytes and will often disclose the primary pathogens. Patients with nonbacterial pneumonias or Legionnaires disease generally produce only scant quantities of thin sputum. In influenzal pneumonia, the sputum may be bloody. The Gram stain of sputum from patients with pneumonia from atypical agents reveals an absence of bacteria and a scant cellular response; in patients with mycoplasmal pneumonia, mononuclear cells may predominate.[31] It can be difficult to determine the etiology in patients with nosocomial pneumonias; prolonged hospitalization, antibiotic ad-

ministration, and ventilatory therapy predispose to colonization with organisms that contaminate sputum specimens but can also cause pneumonia. Bronchoalveolar lavage is effective in identifying the responsible pathogen.[32]

Kits for the detection of nucleic acids from *Legionella, Mycoplasma,* and mycobacterial species in sputum are currently available, and similar tests may soon be available for the detection of other pathogens.[33] Urinary antigen assays may assist in the diagnosis of *Legionella*[34] and pneumococcal[35] pneumonias. Blood cultures are of limited value in the management of patients with community-acquired pneumonias.[36]

Invasive Studies

In immunocompromised patients, numerous opportunistic agents can cause pneumonia, and aggressive techniques may be required to obtain a satisfactory specimen. Although invasive procedures rarely are necessary in immunocompetent patients, they may be required in patients who present with unusual features, are critically ill, or fail to respond to conventional therapy. Procedures such as transtracheal aspiration, bronchoscopy (sometimes including transbronchial biopsies), bronchial brushing, or percutaneous lung taps[37] may be necessary; bronchoalveolar lavage is a particularly useful technique and is generally well tolerated. If these less invasive techniques fail to produce a diagnosis, open lung biopsy should be considered.

Diagnosis of Ventilator-Associated Pneumonia

The diagnosis of ventilator-associated pneumonias may be based on clinical criteria, Gram stains and cultures of tracheal aspirates, and pulmonary infection scores or invasive techniques.[38,39] Even with bronchoscopy and lavage, the diagnosis can be difficult; the detection of a marker called soluble triggering receptor expressed on myeloid cells (sTREM-1) in lavage fluid may be an indication of ventilator-associated pneumonia.[40]

DIFFERENTIAL DIAGNOSIS

Noninfectious diseases can be mistaken for infections of the respiratory tract. Asthmatic bronchitis and hypersensitivity pneumonitis are common examples. COPD, including emphysema and bronchiectasis, may be misleading if previous x-rays are not available. Atelectasis, pulmonary infarction, pulmonary edema, and lung tumors may also be confused with pneumonia. Hypersensitivity reactions and toxins—in the form of aerosols, systemic drugs, or chemicals—produce clinical illnesses and pulmonary infiltrates simulating those of infectious pneumonia. Radiation pneumonitis, sarcoidosis, vasculitis, uremic pneumonitis, pulmonary hemorrhage, eosinophilic pneumonia, organizing pneumonia, and lipoid pneumonitis are included in the differential diagnosis [*see 14:V Chronic Diffuse Infiltrative Lung Disease*].

One of the primary considerations in patients with a lower respiratory tract infection is to distinguish between acute bronchitis and pneumonia. The distinction is anatomic rather than etiologic because the same basic range of organisms can cause the two syndromes. As a result, these conditions often overlap clinically, but bronchitis requires less intensive therapy [*see* Acute Bronchitis, *below*].

TREATMENT

Certain general principles are useful in the care of all patients with pneumonia. Adequate hydration is important to help clear secretions; hydration can be achieved by systemic administration of fluids and local airway humidification. Expectorants such as guaifenesin may be helpful in loosening the sputum. Although clinical trials have demonstrated that chest physiotherapy does not hasten the resolution of pneumonia, this traditional therapeutic modality may provide symptomatic benefit to patients with copious airway secretions. In general, the cough reflex should not be suppressed in patients with bacterial infections, because coughing is an important mechanism for clearing secretions. If severe paroxysms of coughing produce respiratory fatigue or harsh pain, however, temporary relief may be obtained with small doses of codeine. Chest pain should be treated with analgesics that do not suppress cough. If hypoxia is present, oxygen should be administered. Persons who have COPD and retain carbon dioxide must be monitored very closely because oxygen therapy can lead to respiratory depression.

Specific antimicrobial therapy depends on the etiologic agent. Whereas culture and sensitivity testing require at least 24 to 48 hours to provide definitive information, the clinical setting,

Table 3 Initial Antibiotic Therapy for Community-Acquired Pneumonia in Outpatients*

Drug	*Typical Dose*†	*Cost/Mo ($)*‡	*Comments*
Fluoroquinolones			Excellent first-line drugs
Levofloxacin	500 mg p.o., q. 24 hr for 10 days, or 750 mg p.o., q. 24 hr for 5 days	279	
Gatifloxacin	400 mg p.o., q. 24 hr	252	
Moxifloxacin	400 mg p.o., q. 24 hr	265	
Gemifloxacin	320 mg p.o., q. 24 hr	NA	
Macrolides			
Erythromycin	250–500 mg p.o., q. 6 hr	34	Cost-effective alternative, but GI intolerance is common
Clarithromycin	250–500 mg p.o., q. 12 hr	235	Better GI tolerability and activity against *Haemophilus* and *Moraxella*; good first-line drug
Azithromycin	500 mg p.o. day 1, then 250 mg p.o. days 2–5	NA	Better GI tolerance and activity against *Haemophilus* and *Moraxella*
Doxycycline	100 mg p.o., q. 12 hr	92	Cost-effective alternative

*See text for details.
†For details, see *7:XIV Chemotherapy of Infection* (especially Tables 3 and 4).
‡Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.
GI—gastrointestinal

Table 4 Initial Antibiotic Therapy for Community-Acquired Pneumonia in Patients Who Require Hospitalization*

Regimen	*Typical Dose*†	*Cost/Mo ($)*‡	*Comments*
Cephalosporins Cefotaxime or ceftriaxone + a macrolide† or a fluoroquinolone†	Cefotaxime, 1–2 g I.V. q. 4 hr; ceftriaxone, 1–2 g I.V. q. 12–24 hr	—	First-line treatment of choice for severely ill patients
Fluoroquinolones			
Levofloxacin	500 mg p.o. or I.V. q. 24 hr	279	First-line treatment, either alone or with a third-generation cephalosporin
Moxifloxacin	400 mg p.o. or I.V. q. 24 hr	265	
Gatifloxacin	400 mg p.o. or I.V. q. 24 hr	252	
Vancomycin + a macrolide§ or a fluoroquinolone†	Vancomycin, 1 g I.V. q. 12 hr	—	Alternative for severely ill patients who are allergic to β-lactams
Linezolid + a macrolide‡ or a fluoroquinolone†	Linezolid, 600 mg p.o. or I.V. q. 12 hr	3,308	For severely ill patients who cannot tolerate β-lactams or vancomycin

*See text for details.
†For details, see *7:XIV Chemotherapy of Infection* (especially Tables 3 and 4).
§For macrolide doses, see Table 3; azithromycin may be administered intravenously.
‡Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.

chest x-ray, and sputum Gram stain usually enable the physician to make a reasonable presumptive diagnosis and to initiate therapy at once. Treatment can then be modified as necessary on the basis of culture results. Because antibiotics penetrate sputum by passive diffusion, it is important to maintain adequate blood levels of these drugs. The administration of antibiotics by aerosol is not indicated for most cases of pneumonia but may help patients with cystic fibrosis and endobronchial *Pseudomonas* infections.[41]

It is best to choose an antibiotic regimen directed specifically at organisms seen on Gram stain and, after 24 to 48 hours, identified from sputum or blood cultures. Even if these data are lacking, however, a reasonable choice of initial antimicrobial therapy can be made on the basis of the epidemiologic setting and clinical features.

Community-Acquired Pneumonia

Several factors are responsible for rapid changes in the empirical treatment of community-acquired pneumonias: the emergence of drug-resistant pneumococci; the increasing population of elderly or chronically ill patients who are vulnerable to infections caused by *H. influenzae* and *M. catarrhalis;* the increased importance of atypical pathogens such as *M. pneumoniae, C. pneumoniae,* and *L. pneumophila;* and the availability of new fluoroquinolones with enhanced activity against gram-positive cocci (including penicillin-nonsensitive pneumococci) and anaerobes (including mouth flora). For patients who do not require hospitalization, several options are available [*see Table 3*].[42] Erythromycin is cost-effective, but the so-called advanced macrolides clarithromycin and azithromycin may be preferable because of their better gastrointestinal tolerability and their activity against *Haemophilus* and *Moraxella* species.[43-45] Doxycycline is an effective and inexpensive alternative.[46] However, because of the increasing prevalence of drug-resistant pneumococci, use of one of the so-called respiratory fluoroquinolones (i.e., levofloxacin, gatifloxacin, moxifloxacin, or gemifloxacin) is recommended.[47] These agents have excellent activity against the major causes of community-acquired pneumonia, and prospective trials have been very favorable.[47,48] They are a particularly good choice for patients who have recently received antibiotic therapy.[49] The use of high-dose levofloxacin (750 mg a day) may enable short-course (5-day) therapy, potentially improving compliance and reducing costs.[50] The respiratory fluoroquinolones may also emerge as drugs of choice for patients with community-acquired pneumonia who require hospitalization; levofloxacin, moxifloxacin, and gatifloxacin are available in preparations for intravenous administration [*see Table 4*]. Levofloxacin-resistant pneumococci are still uncommon, but their emergence is a concern.[51,52] In addition, dual therapy may be preferable for patients with severe pneumococcal pneumonia.[53,54] For that reason, patients with moderate to severe community-acquired pneumonia may benefit from cefotaxime or ceftriaxone in combination with a macrolide or a fluoroquinolone.[42] Because vancomycin is active against virtually all pneumococci, it can be substituted for the third-generation cephalosporin in patients allergic to β-lactams; linezolid is another alternative.[55] When aspiration is suspected, penicillin, clindamycin, or metronidazole is useful; amoxicillin-clavulanate, imipenem, meropenem, and respiratory fluoroquinolones are also active against oral anaerobes [*see Table 5*].

In all cases, antibiotic therapy should be tailored to the results of culture and sensitivity, the clinical response, and the occurrence of side effects. Many patients who require intravenous antibiotics initially can be switched to oral therapy within 3 days,[56] facilitating early hospital discharge.[57] In most patients with uncomplicated pneumococcal pneumonia, antibiotics can be discontinued after 3 afebrile days; most patients with other bacterial pneumonias are treated for 7 to 14 days, and most with atypical pneumonias are treated for 10 to 21 days.

Immunizations may help prevent community-acquired pneumonia. Each fall, influenza vaccine should be offered to patients 50 years of age and older and to other vulnerable persons [*see 7:XXV Respiratory Viral Infections*]. Pneumococcal polysaccharide vaccine should be offered to persons 65 years of age and older and to others who are at increased risk[58] [*see 7:I Infections Due to Gram-Positive Cocci*]. Patients who have recovered from one bout of pneumonia may benefit from both vaccines.

Table 5 Antibiotic Choices for Aspiration Pneumonia*

Drug	*Typical Dose*†	*Cost/mo ($)*‡	*Comments*
Penicillin	500 mg p.o., q. 6 hr, to 1–2 million units I.V. q. 4 hr, depending on severity of infection	22	Traditional drug of choice
Clindamycin	150–300 mg p.o., q. 6 hr, to 600 mg I.V. q. 8 hr, depending on severity of infection	318	May be superior to penicillin
Metronidazole	500 mg p.o., q. 8 hr, to 500 mg I.V. q. 6 hr, depending on severity of infection	28	Excellent alternative
Ampicillin-sulbactam	1–2 g ampicillin + 0.5–1 g sulbactam I.V. q. 6 hr		Alternative useful in hospitalized patients
Imipenem	0.5–1 g I.V. q. 6–8 hr		Alternative useful in hospitalized patients
Meropenem	1 g I.V. q. 8 hr		Alternative useful in hospitalized patients
Fluoroquinolones			
Gatifloxacin	400 mg p.o. or I.V. q. 24 hr	252	Excellent for community-acquired pneumonias but less active against oral anaerobes than penicillin, clindamycin, and metronidazole
Moxifloxacin	400 mg p.o. or I.V. q. 24 hr	265	
Levofloxacin	500 mg p.o. or I.V. q. 24 hr	279	

*See text for details.
†For details, see 7:*XIV Chemotherapy of Infection* (especially Tables 3 and 4).
‡Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.

Hospital-Acquired Pneumonia

Because gram-negative bacilli and *S. aureus* cause many nosocomial pneumonias, patients with hospital-acquired pneumonias require broad antimicrobial coverage until the results of Gram stains, cultures, and sensitivity tests permit focused therapy. Options for the initial treatment of hospital-acquired pneumonia include ticarcillin–clavulanate or piperacillin-tazobactam; meropenem or imipenem-cilastatin; a third-generation cephalosporin plus nafcillin or vancomycin; a first-generation cephalosporin plus an aminoglycoside; or vancomycin plus an aminoglycoside. The prevalence of resistant bacteria in a particular hospital or patient care unit should help guide the initial therapy; for example, if methicillin-resistant staphylococci are common, vancomycin is a desirable component of the initial therapy, and when multidrug-resistant *Klebsiella* organisms are common, meropenem or imipenem-cilastatin should be considered. Linezolid is an effective alternative to vancomycin for the treatment of nosocomial pneumonia caused by methicillin-resistant gram-positive cocci. Patients with ventilator-associated nosocomial pneumonia may respond as well to 8 days of antibiotics as to 15 days of therapy.[59] Because of the high mortality associated with pneumonias in the ICU, strategies to prevent them have been studied,[60] and comprehensive guidelines are available.[61]

Infections Caused by *Legionella* Species

LEGIONNAIRES DISEASE

Epidemiology and Etiology

Since it was first identified in 1976, Legionnaires disease has become recognized as a common cause of both community-acquired and hospital-acquired pneumonias. Worldwide, it accounts for between 2% and 15% of all community-acquired pneumonias severe enough to require hospitalization.[62] It is estimated that 10,000 to 25,000 cases occur in the United States each year, but only 1,200 to 1,500 are reported annually.[63] Legionnaires disease is caused by *L. pneumophila*, a fastidious, filamentous, flagellated, aerobic gram-negative bacillus. The organism can be grown on charcoal–yeast extract agar; optimal growth occurs at 35° C in 5% carbon dioxide, but growth is slow, and a period of 3 to 6 days is required for colonies to form.

At least nine serogroups of *L. pneumophila* exist; most clinical isolates belong to serogroup 1. By special staining techniques, large numbers of the organism can be identified in tissue sections of alveoli, both within macrophages and extracellularly. Virulence factors of *L. pneumophila* and various extracellular enzymes that the organism secretes have been identified. *L. pneumophila* is able to survive intracellularly in host leukocytes. Antibody is not protective, but cell-mediated immunity does promote recovery and prevent reinfection.

In nature, *L. pneumophila* survives principally in water and, to a lesser extent, in soil. Human disease is acquired primarily by inhalation of aerosols contaminated with organisms; person-to-person transmission has not been documented. Contaminated water systems have been responsible for both community-acquired and hospital-acquired outbreaks [*see 7:X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas*].

The attack rate for Legionnaires disease appears to be higher in elderly persons and persons with underlying conditions such as COPD, neoplastic disease, organ transplants, and renal failure. Although *L. pneumophila* is a relatively uncommon pathogen in persons infected with HIV, it can cause severe disease in them.

Diagnosis

Clinical features Legionnaires disease is characterized by a 1-day prodrome of myalgias, malaise, and slight headache after an incubation period of 2 to 10 days. Acute onset of high fever, shaking chills, nonproductive cough, tachypnea, and, often, pleuritic pain ensues. The cough may subsequently become slightly productive, but the sputum is not purulent. Obtundation or toxic encephalopathy is common, but frank meningitis is not a feature. Abdominal pain, vomiting, and, especially, diarrhea may be present. Signs of consolidation on lung examination are present infrequently, but rales are commonly heard. Chest radiographs show patchy or interstitial infiltrates, which often progress to areas of nodular consolidation in a single lobe or multiple lobes; minimal effusions are present in up to one

third of cases. Abscess formation is uncommon but has been observed. Pulmonary fibrosis may occur in some survivors.

Although pneumonia is present in nearly all patients with Legionnaires disease, extrathoracic symptoms can be the presenting or predominant features. Central nervous system, GI, and renal manifestations are especially common. *L. pneumophila* has been isolated from blood cultures, and the organism can be found in many organs both in immunosuppressed patients and in previously normal patients who are afflicted with severe disease. Extrapulmonary manifestations include ocular and pericardial involvement, perirectal abscess, wound infection, peritonitis, cellulitis, rhabdomyolysis and acute renal failure, neutropenia, hemolytic anemia, and thrombotic thrombocytopenic purpura. Implanted devices such as heart valves and hemodialysis fistulas can become colonized.

A nonpneumonic form of legionellosis called Pontiac fever has a short incubation period and a low mortality. It has been responsible for at least four outbreaks of illness, including several related to whirlpools and hot tubs.

Laboratory studies The peripheral white blood cell count is mildly elevated to between 8,000 and 16,000/mm^3. Cold agglutinins are negative. Other laboratory findings may include an elevated erythrocyte sedimentation rate, hypoxia, abnormal liver function test results, and elevated creatine phosphokinase levels.[64] Proteinuria and microscopic hematuria have been observed, and acute renal failure may complicate the course on occasion.

Gram stains of the sputum or tracheal secretions will not reveal *L. pneumophila,* but the organism can be isolated from sputum and other specimens by using charcoal–yeast extract agar. The diagnosis can be established rapidly in about 20% of cases by demonstrating the organism with direct immunofluorescent staining of sputum specimens; bronchoalveolar lavage may be helpful in immunosuppressed patients. A kit that uses radiolabeled complementary DNA is commercially available, but clinical experience is still limited. A very promising polymerase chain reaction assay has been developed.[65]

Another method of rapid diagnosis involves detection of *L. pneumophila* antigen in the urine; this radioimmunoassay test is highly specific and has a sensitivity of about 80% to 90%.[65] However, the test is available only for *L. pneumophila* serogroup 1, which is the most common cause of Legionnaires disease. Most often, however, the diagnosis is established by an indirect fluorescent antibody technique involving staining of the causative bacterium. With this technique, a fourfold or greater rise in titer during the illness or a stable titer of 1:256 or greater is considered diagnostic.

The clinical picture and radiologic findings in Legionnaires disease are not specific. The diagnosis should be considered in patients with segmental, lobar, or interstitial pneumonia in which the etiologic agent is not evident on Gram stains of sputum or tracheal secretions. A mild case may resemble *Mycoplasma* pneumonia or other types of atypical pneumonia.

Treatment

On in vitro susceptibility testing, *L. pneumophila* has been shown to be susceptible to a variety of antimicrobial agents, including erythromycin, clarithromycin, azithromycin, tetracycline, rifampin, and the fluoroquinolones. Current evidence indicates that azithromycin[66] or levofloxacin is the treatment of choice.[67] Occasionally, patients may experience a relapse if antibiotics are discontinued prematurely; recovery occurs during a second, more prolonged course of treatment. A combination of rifampin and either azithromycin or levofloxacin may be considered in patients who fail to respond to monotherapy and in immunologically impaired patients with overwhelming disease. Improvements in diagnosis and therapy have produced a dramatic decline in the case-fatality rate of *L. pneumophila* infection, from 34% in 1980 to 12% in 1998.[68]

INFECTIONS CAUSED BY OTHER *LEGIONELLA* SPECIES

Since 1943, unusual, fastidious *Rickettsia*-like organisms have been identified as causes of isolated cases of pneumonia. Long considered medical curiosities, these organisms were named after their discoverers or the patients from whom they were isolated. Such obscure nomenclature (e.g., TATLOCK, OLDA, HEBA, and WIGA) reflected the absence of knowledge concerning these agents. Subsequent studies of these organisms led to their reclassification as *Legionella* species. Of the more than 30 *Legionella* species that have been identified, at least 19 have been recognized as causes of pneumonia, particularly in immunosuppressed hosts.

The clinical picture is not distinctive. Fever is the most common feature, cough is variable in severity, and sputum production is absent or scant. Pleurisy or dyspnea may develop, and chest x-rays show a patchy or nodular, progressive bronchopneumonia. *L. micdadei* is the most important of these organisms. Previously known as TATLOCK and HEBA, this organism was rediscovered as the Pittsburgh pneumonia agent that caused acute suppurative pneumonia in 13 immunosuppressed patients from two centers. *L. micdadei* can also cause pneumonia in immunologically intact hosts and extrathoracic infections in immunologically impaired hosts. It has been isolated from hospital water supplies and can cause nosocomial infections in immunosuppressed patients.[69]

In the first 13 patients with Pittsburgh pneumonia, the diagnosis was established through a lung biopsy or autopsy that revealed acute alveolar inflammation and short, weakly acid-fast, gram-negative bacilli. More recent cases have been diagnosed by culturing the agent on charcoal–yeast extract agar or by serologic means. Erythromycin has been used with success. Unlike *L. pneumophila, L. micdadei* is susceptible to penicillins and cephalosporins in vitro; however, there are no data on the use of these drugs in clinical *L. micdadei* infection.

Many other *Legionella* species have been identified as causes of pneumonia. These organisms grow on charcoal–yeast extract agar. Some cases in which *Legionella* species have been implicated share common features: patients have been exposed to infected water and have COPD or are immunosuppressed.[70] Therapy with erythromycin, alone or with rifampin, has been suggested.

CHLAMYDOPHILA (CHLAMYDIA) PNEUMONIAE

C. pneumoniae, initially known as the Taiwan acute respiratory disease (TWAR) agent, is an important pathogen, accounting for up to 10% of all pneumonias in the United States.[71] Serologic surveys suggest that a majority of adults have been infected at some point, indicating that most cases are mild or subclinical.

Unlike psittacosis, which is a true zoonosis that spreads only from animals to humans,[72] *C. pneumoniae* spreads from person to person via respiratory droplets. The incubation period of *C. pneumoniae* is long. Unlike *C. trachomatis,* which causes neonatal pneumonia, *C. pneumoniae* infects mainly older children and young adults; however, older patients (especially those with

COPD) can also be affected. Clinically, *C. pneumoniae* pneumonia in young patients resembles *M. pneumoniae* pneumonia[73]; after a prodrome of pharyngitis that lasts no more than 2 weeks, nonproductive cough and fever occur. Pulmonary infiltrates are mild. The infection is usually self-limited. In older patients with underlying COPD, *C. pneumoniae* can cause bronchitis that may be mild or very persistent. Bronchospasm may be prominent.[74] This organism can sometimes cause severe pneumonia in patients with COPD, and fatalities have been reported in debilitated patients. A macrolide, a tetracycline, or a fluoroquinolone is recommended for therapy.

Recurrent Pneumonia

Patients with recurrent pneumonia present a diagnostic and therapeutic challenge. Anatomic abnormalities should be carefully sought in all such patients and should be suspected particularly when the infections recur in one bronchopulmonary segment. Examples of anatomic abnormalities include cysts, blebs, abscess cavities, bronchiectasis, and bronchial obstruction by tumors, foreign bodies, or bronchiostenosis. If shifting locations are involved, systemic abnormalities may be present. Examples include defects of leukocyte function, immunoglobulin deficiencies, α_1-antitrypsin globulin deficiency, and cystic fibrosis. Recurrent aspiration may be responsible, in which case a carefully performed neurologic examination and a barium swallow may reveal the proper diagnosis. It is important to remember that noninfectious processes may produce recurrent pulmonary infiltrates that may be accompanied by cough and fever [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. Examples of such processes include organizing pneumonia, eosinophilic pneumonia, hypersensitivity pneumonitis, vasculitis, pulmonary hemosiderosis, and pulmonary emboli.

Chronic Pneumonia

Often, bacterial pneumonias exhibit delayed resolution, in which radiographic abnormalities persist for weeks or even months after clinical recovery. Infrequently, pyogenic infection pursues a slow but progressively destructive course despite antibiotic therapy; *K. pneumoniae* and other gram-negative bacilli have been implicated in some patients with such chronic bacterial pneumonias. Mycobacteria and *Actinomyces* are common causes of chronic pulmonary infection; fungi are particularly common in patients treated with corticosteroids. Persistent infiltration caused by neoplasia, sarcoidosis, pulmonary hemorrhage, vasculitis, fibrosing alveolitis, alveolar proteinosis, lipoid pneumonia, toxins, and other processes may mimic chronic pulmonary infection. Fiberoptic bronchoscopy is very useful for diagnosis, but lung biopsy may be needed.

Aspiration Pneumonia

EPIDEMIOLOGY AND ETIOLOGY

Although aspiration is probably the mechanism responsible for most bacterial pneumonias, the term aspiration pneumonia is best reserved for infection caused by mixed mouth flora. Seventy percent of patients with depressed consciousness have demonstrable pharyngeal aspiration, often involving much larger volumes of material than that aspirated by healthy persons. Because of the larger volumes of aspirated material, underlying diseases impairing host defenses, and alterations in oropharyngeal flora, patients with altered consciousness are most prone to aspiration pneumonia. In clinical practice, aspiration pneumonia most often results from alcoholism, drug abuse, administration of sedatives or anesthesia, head trauma, and seizures or other neurologic disorders. In addition, aspiration of GI contents is more frequent in patients with abnormalities of deglutition or esophageal motility resulting from placement of nasogastric tubes, esophageal carcinoma, bowel obstruction, or repeated vomiting from any cause. Poor oral hygiene and periodontal disease predispose to aspiration pneumonia because of the increased bacterial flora in these patients.

The clinical results of pulmonary aspiration depend in large part on the nature and volume of material aspirated. Aspiration of gastric contents is a common problem that may produce Mendelson syndrome, a fulminating illness, if a large volume of acidic gastric juice is aspirated.[75] Aspiration of particulate material can produce acute airway obstruction and death by asphyxiation; aspiration of smaller particles may produce atelectasis of a pulmonary segment or even of an entire lung, resulting in dyspnea, wheezing, and cyanosis. Characteristic pulmonary injuries and distinctive clinical syndromes are produced by aspiration of smoke, freshwater or saltwater, and fats or oils (lipoid pneumonitis).

DIAGNOSIS

Clinical Features

Patients with mixed aspiration pneumonia may present with an acute febrile illness, or the illness may follow a more indolent course, extending over many days or even weeks. Fever, cough, and sputum production are the dominant symptoms; the sputum may be copious, foul smelling, or both. Physical examination typically discloses rales and signs of pulmonary consolidation. An evaluation of dental hygiene and of the gag reflex is helpful; disordered pharyngeal sensation is a better predictor of vulnerability than is an absent gag reflex.

Laboratory Studies

Radiographically, infiltrates are most common in dependent areas of the lung, especially the apical segments of the lower lobes and the posterior segments of the upper lobes. Tissue necrosis can occur. Without treatment, aspiration pneumonia may produce multiple small cavities, which reflect a necrotizing pneumonitis. Lung abscesses or empyemas may ensue.

The sputum of patients with classic aspiration pneumonia contains abundant polymorphonuclear leukocytes and mixed mouth flora. If specimens are obtained by transtracheal aspiration or other procedures that avoid contamination of sputum by organisms from the oral cavity, aerobic and anaerobic bacteriologic techniques can reveal the specific causative bacteria. Because anaerobes are the dominant flora of the upper respiratory tract (outnumbering aerobic or facultative bacteria by 10 to 1), it is not surprising that anaerobes are the dominant organisms in aspiration pneumonia. Of particular importance are *Prevotella melaninogenica* and other *Prevotella* (formerly, oral strains of *Bacteroides*) species (slender, pleomorphic, pale gram-negative rods), *Fusobacterium nucleatum* (slender gram-negative rods with pointed ends), and anaerobic or microaerophilic streptococci and *Peptostreptococcus* (small gram-positive cocci in chains or clumps). As expected, multiple organisms are recovered from most patients.

TREATMENT

With the exception of *B. fragilis*, which can be identified along with other anaerobic species in 17% of patients with classic aspiration pneumonia, all the anaerobes found are penicillin sensitive. Penicillin is effective when *B. fragilis* is present in combination with penicillin-sensitive organisms, suggesting that aspiration pneumonias are synergistic infections that can be treated successfully by elimination of most but not necessarily all of the organisms involved. Penicillin dosages of 2.4 to 6.0 million units daily are generally effective. Parenteral therapy is advisable initially, but a 10- to 14-day course of treatment can be concluded with orally administered antibiotics if the patient responds well. Clindamycin represents an excellent alternative agent and may even be superior to penicillin for treatment of necrotizing aspiration pneumonias and lung abscesses.

Hospitalization or antibiotic therapy alters the usual oropharyngeal bacterial flora, so that staphylococci, facultative gram-negative bacilli, or both may be identified in patients. As a result, aspiration pneumonia in hospitalized patients often involves pathogens that are uncommon in community-acquired pneumonias. Gram stains and cultures of sputum are especially important for identifying gram-negative bacilli and staphylococci in the hospital setting. Broad antimicrobial coverage is required until specific pathogens have been identified by culture and sensitivity testing. Although tube feedings are often recommended to prevent aspiration pneumonia, there is no evidence that they are effective.[75]

Other Pulmonary Infections

ACUTE BRONCHITIS

Cough is the chief complaint responsible for an estimated 30 million physician office visits in the United States annually. For about 12 million of these patients, the clinical diagnosis is acute bronchitis.[76]

Acute bronchitis is commonly defined as an acute respiratory tract infection in which cough, with or without sputum production, is a prominent feature. In most cases, an etiologic diagnosis is not established. When sputum is absent or scant, the illness is often attributed to a respiratory tract virus; when purulent sputum is present, the bacteria that cause community-acquired pneumonias are considered likely causes.

Most otherwise healthy persons recover from acute bronchitis in 1 to 3 weeks, but the cough can linger for more than a month in up to 20% of patients.[77] Although 70% to 90% of patients are treated with antibiotics, published trials demonstrate little clinical benefit, even if purulent sputum is present.[77-79] Guidelines of the Infectious Disease Society of America, the American College of Physicians–American Society of Internal Medicine, and the American Academy of Family Physicians state, "Routine antibiotic treatment of uncomplicated acute bronchitis is not recommended, regardless of duration of cough."[80] Suspected cases of pertussis constitute an exception.[81] Clinicians who are confronted with demands for antibiotics may find a delayed prescription strategy useful.[82,83] This strategy involves providing patients with a prescription, but instructing them to fill it only if their symptoms fail to improve within the next few days. Patients with high fever, chills, respiratory distress, underlying pulmonary or immunosuppressive disorders, or physical signs of pulmonary parenchymal infection should be evaluated for pneumonia and treated according to the guidelines for community-acquired pneumonia (see above).

CHRONIC BRONCHITIS

Patients with chronic bronchitis characteristically produce sputum on most days for at least 3 months each year for more than 2 years. The sputum is frequently colonized by *H. influenzae* (nontypable), *S. pneumoniae*, or *M. catarrhalis*, singly or in combination. Although it is not certain whether the bacteria themselves produce additional airway damage, heavy bacterial loads correlate with increased inflammation.[84] Patients who acquire a new strain of bacteria are at increased risk for symptomatic exacerbations of their chronic bronchitis.[85] The role of long-term prophylactic antibiotic therapy in chronic bronchitis is controversial. Long-term antibiotic therapy may provide symptomatic relief in certain patients who experience multiple exacerbations of bronchitis during the winter, but it is not useful in improving or preserving pulmonary function. However, short-term antibiotic therapy is effective in treating acute exacerbations of chronic bronchitis.[86]

BRONCHIECTASIS

True saccular, or cystic, bronchiectasis involves both dilatation of the bronchi and destruction of the bronchial walls. Bronchiectasis results most often from neglected or recurrent infection, especially in childhood; therefore, bronchiectasis has become much less common since the introduction of antibiotics. Aggressive medical therapy has greatly improved the prognosis.

Symptoms include cough that may be dry or productive of copious foul sputum, recurrent lower respiratory tract infection, and hemoptysis. In rare instances, bronchiectasis can present as pleuritic chest pain. In advanced cases, fibrosis can lead to cor pulmonale and respiratory failure. The chest x-ray may show increased lung markings, honeycombing, atelectasis, or pleural changes, but high-resolution or helical chest CT is required for definitive diagnosis[87] [*see 14:III Chronic Obstructive Diseases of the Lung*].

Lung Abscess

EPIDEMIOLOGY AND ETIOLOGY

In the antibiotic era, lung abscesses have become less common and less serious. The most common variety has been termed the primary, simple, nonspecific, or putrid abscess. Primary lung abscess accounts for about 60% of all lung abscesses and originates from a necrotizing suppurative bronchopneumonia caused by the aspiration of mixed oropharyngeal bacteria. Thus, both the predisposing factors and the causative organisms are similar to those identified in aspiration pneumonia. Patients with primary lung abscesses typically have alterations of consciousness because of underlying problems such as alcoholism and neurologic disorders; periodontal disease is often present. The organisms causing the abscess are much more reliably identified by transtracheal aspirates than by sputum cultures, which are invariably contaminated with anaerobes and other mouth flora. Percutaneous lung aspiration and bronchoalveolar lavage may also be useful for bacteriologic diagnosis. Mixed anaerobic bacteria are seen in most cases; *F. nucleatum, P. melaninogenica, Peptostreptococcus,* and anaerobic or microaerophilic streptococci predominate. *B. fragilis* is recovered with other organisms in 15% of cases.

Many other conditions can lead to lung abscess. Necrotizing

bacterial pneumonias caused by *S. aureus, K. pneumoniae,* or other gram-negative bacilli can lead to abscess formation. In other patients, abscess develops as a result of bronchial obstruction caused by tumors, foreign bodies, or bronchial stenosis. Septic pulmonary embolization is a cause of abscess formation. Pulmonary tuberculosis, fungal infection, or actinomycosis often leads to cavity formation. In the immunosuppressed host, *Nocardia* and other opportunistic organisms may also produce cavitation. Lung abscesses in patients with AIDS are caused by a wide array of organisms and respond poorly to therapy.

Noninfectious processes can produce cavitary lung lesions. Primary and metastatic tumors, bullae, cysts, intralobar pulmonary sequestration, pulmonary infarcts, vasculitis (including Wegener granulomatosis), and rheumatoid lung disease must be considered in the differential diagnosis of lung abscess.

DIAGNOSIS

Clinical Features

The clinical presentation of the patient with a lung abscess depends on the type of abscess. Patients with abscesses resulting from necrotizing staphylococcal or gram-negative bacillary pneumonias are usually acutely ill and exhibit clinical features of the underlying pneumonia. Although patients with primary lung abscess may also present acutely with aspiration pneumonia, they more often experience insidiously progressive symptoms for weeks or even months before diagnosis. Cough is present in almost all patients; when the abscess drains into the bronchial tree, production of copious foul-smelling sputum is characteristic. Hemoptysis is present in approximately one third of cases and may occasionally reach life-threatening proportions. Chest pain consisting of either a dull ache or a true pleurisy is common. Most patients have fever, but frank rigors are unusual. Often, patients with a chronic course of lung abscess lasting many weeks have anorexia, weight loss, and debility.

Physical Examination and Imaging

Physical examination of a patient with a lung abscess may disclose pulmonary rales, signs of consolidation, or, rarely, clubbing of the nails. These findings are not diagnostic, however, and chest x-rays or CT scans are required to establish the presence of an abscess. Although any lung segment may be involved, abscesses are most common in the posterior segments of the upper lobes and the apical segments of the lower lobes, because these areas are dependent when a person is recumbent. Abscesses may be single or, less often, multiple. The finding of air-fluid levels signifies rupture into the bronchial tree.

Laboratory Studies

As is the case with other pulmonary infections, examination of the sputum is crucial to the diagnosis of lung abscess. In patients with primary lung abscesses, the sputum is often putrid and contains numerous polymorphonuclear leukocytes and an abundant mixed microbial flora. Sputum cultures reveal only normal mouth flora. Meaningful anaerobic bacteriology depends on obtaining, either by transtracheal aspiration or by bronchoscopy, specimens that have not traversed the oropharynx. Percutaneous needle aspiration can also be very helpful, both diagnostically and therapeutically. In a typical case of aspirational putrid lung abscess, these invasive procedures may not be necessary. They are important, however, if the diagnosis is uncertain. The indications for bronchoscopy in patients with lung abscess are debatable. Although some physicians advocate bronchoscopy in all such patients, other physicians reserve the procedure for patients in whom there is a suspicion of bronchial obstruction by a foreign body or tumor, for patients who fail to respond to medical therapy, and for patients from whom specimens are required to rule out tuberculosis, fungal infection, or carcinoma. Bronchoscopy may be helpful therapeutically by promoting bronchial drainage from cavities that incompletely communicate with the bronchial tree.

TREATMENT

If specific pathogens such as *S. aureus* or *Klebsiella* are present in reliable specimens, therapy should be directed at the causative pathogen. In primary lung abscesses caused by mixed oral flora, penicillin has been the drug of choice. Prospective studies comparing penicillin therapy with clindamycin therapy in 66 patients with lung abscess found clindamycin to be the superior agent; however, the long clinical experience with penicillin warrants retaining it as the drug of choice, with clindamycin an excellent alternative for patients who are allergic to penicillin or who respond poorly to that drug. Despite its excellent bactericidal activity against anaerobic bacteria, metronidazole appears less effective in treating lung abscess. Most centers initiate treatment with I.V. penicillin in a dosage range of 6 to 12 million units a day or with I.V. clindamycin in a dosage of 600 mg every 8 hours. After a clear-cut clinical response is observed, oral penicillin V in a dosage of 750 mg four times a day or oral clindamycin in a dosage of 300 mg every 6 hours can be substituted. Parenteral therapy is often required for 2 to 4 weeks before the occurrence of defervescence, diminished sputum production, and reduction in cavity size. The duration of therapy depends on the clinical course, but prolonged treatment for 4 to 8 weeks is usually required.

In addition to administration of antibiotics, adequate drainage is essential and can usually be achieved with intensive pulmonary physiotherapy and postural drainage. Bronchoscopy can be very useful in promoting drainage and for excluding the diagnosis of cancer. Although surgery was once the mainstay of treatment for lung abscess, antibiotics are now almost always able to control infection, and surgery is needed only when complications occur. Massive hemoptysis is an indication for lung resection. Uncontrolled sepsis may occasionally necessitate lobectomy. CT-guided percutaneous tube drainage may be very helpful in patients who are too ill to tolerate thoracotomy and may be the treatment of choice for lung abscesses that are refractory to medical management. Empyema, another complication of lung abscess, requires external drainage by thoracentesis, chest tube, or rib resection. The persistence of a thin-walled cavity after otherwise successful medical treatment, however, is not an indication for surgery. Recurrent or persistent infection, recurrent hemoptysis, or the suspicion of tumor may mandate operative intervention. Shaggy, thick-walled cavities may be suggestive of tumor. Complications of lung abscess that have become uncommon because of antibiotic therapy include bronchogenic spread of infection to other pulmonary segments, bronchiectasis, and bacteremia with metastatic infection such as brain abscess.

Whereas patients with gram-negative or staphylococcal bacillary pneumonias or serious underlying diseases have a substantial mortality, the prognosis for patients with primary lung abscess is quite good. It is important to prevent recurrent pulmonary infection by treating dental disease and by avoiding factors that predispose to pulmonary aspiration.

Empyema

ETIOLOGY

Bacteria can reach the pleural space by many routes. Most often, empyema results from the direct spread of bronchopulmonary infections, including pneumonias, lung abscesses, and bronchiectasis.[88] Less often, empyema develops as a complication of thoracotomy or, rarely, thoracentesis. Open chest trauma provides another means for the direct introduction of microorganisms. Intra-abdominal infections, especially subphrenic abscesses, can penetrate the diaphragm to cause empyemas. Uncommonly, esophageal rupture can cause spread of infection from the mediastinum to the pleural space. Finally, hematogenous seeding is an infrequent mechanism of empyema formation.

S. aureus, various species of *Streptococcus,* and gram-negative bacilli are the most common causes of empyema; among the gram-negative bacilli, *K. pneumoniae* has been linked with diabetes.[89] Many infections are mixed. Anaerobes have been recognized in 25% to 76% of empyemas and may occur in pure culture or in combination with aerobic or facultative organisms. *Fusobacterium, Prevotella,* and anaerobic gram-positive cocci are the anaerobes most often seen. *M. tuberculosis* has become a relatively rare cause of pleural space infections, and fungi are implicated uncommonly.[90] Transdiaphragmatic rupture of a liver abscess occasionally produces amebic empyema.

DIAGNOSIS

In most patients, the clinical presentation of empyemas includes fever, dyspnea, chest pain, and cough. Hemoptysis is less common than these other symptoms. If diagnosis and treatment are delayed, weight loss and debility may be prominent. The physical findings in patients with empyemas are no different from those in other patients with pleural effusions. In addition, chest wall tenderness may be present, and there may be signs of an underlying pneumonia or intra-abdominal infection. Tachypnea and respiratory distress may occur, and septic shock may complicate advanced cases. Polymorphonuclear leukocytosis is common; other laboratory findings may include anemia and hypoxia. Chest x-rays reveal pleural effusions that are free flowing in early disease but frequently loculated in late cases. Ultrasonography may be necessary to distinguish fluid from pleural fibrosis. Unless surgery or thoracentesis has been performed, air-fluid levels in the pleural space suggest a bronchopleural fistula.

DIFFERENTIAL DIAGNOSIS

In the differential diagnosis of empyema, it is important to consider the many causes of noninfected pleural effusions.[91] Most important is the distinction between sterile parapneumonic effusions and true empyemas. Thoracentesis is mandatory for the diagnosis of empyema. Several thoracenteses may be needed if the fluid is loculated; CT or ultrasound guidance is very helpful in these circumstances. Gross purulence is diagnostic for empyema, but the absence of frank pus does not rule out infection. Like other inflammatory effusions, empyema fluids have the characteristics of exudates: protein levels greater than 3 g/dl and lactic dehydrogenase values in excess of 550 units. Pleural fluid acidosis is characteristic of empyemas, but alkalosis can occur if the infection is caused by a urea-splitting organism such as *Proteus*. Pleural fluid glucose levels are depressed in empyemas, and although white cell counts are variable, counts above 5,000/mm^3 are common, with polymorphonuclear leukocytes predominating. Gram stains of the pleural fluid will often reveal the causative organisms. Both aerobic and anaerobic cultures are mandatory; a foul odor suggests anaerobic infection. Stains and cultures for mycobacteria and fungi are important in selected cases.

TREATMENT

Treatment of empyemas involves both antibiotics and drainage. Antibiotics should be selected on the basis of the causative pathogens. High-dose parenteral therapy is required, and prolonged courses of 3 weeks or more are often needed. Adequate drainage is of paramount importance. In acute empyemas, the pleural cavity is lined by acute fibrinous inflammation, and percutaneous drainage of free-flowing fluid may be possible by repeated thoracentesis or tube thoracostomy. Closed chest tube drainage is the traditional method for draining empyemas, but image-guided catheter drainage is also effective, particularly when the fluid is loculated. Resolution of fever generally signifies satisfactory drainage. If complete drainage cannot be achieved with chest tubes, video-assisted thoracoscopic surgery (VATS) can often disrupt intrapleural adhesions and achieve excellent drainage of loculated effusions[92]; although VATS requires endotracheal intubation and general anesthesia, it is less invasive than the next alternative, rib resection with thoracotomy for decortication. Enzymatic debridement with streptokinase may enable some patients to avoid surgery.

Septic Pulmonary Embolism

Although once uncommon, septic pulmonary embolism is now encountered because of I.V. drug abuse, which accounts for more than 75% of cases; tricuspid valve endocarditis and direct injection of infected material cause most of these cases. *S. aureus* and gram-negative bacilli are the predominant etiologic agents in I.V. drug abusers. Septic pulmonary embolism may also develop in patients with septic phlebitis of peripheral veins (especially phlebitis related to I.V. lines or pelvic infections), abscesses, or other bacteremic infections.

Unlike bland emboli, septic pulmonary emboli produce pulmonary infarction in most instances. Small emboli produce flame-shaped or patchy infiltrates that may shift in location; these manifestations generally resolve with antibiotic therapy. Larger emboli often cavitate and may lead to lung abscess, empyema, or bronchopleural fistula formation. In addition to antibiotics, surgical drainage may be required for such complications. Operative intervention may be needed to control the source of emboli in some patients. Heparin can be useful in patients with septic phlebitis.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Glezen WP, Greenberg SB, Atmar RL, et al: Impact of respiratory virus infections on persons with chronic underlying conditions. JAMA 283:499, 2000

2. Knowles MR, Boucher RC: Mucus clearance as a primary innate defense mechanism for mammalian airways. J Clin Invest 109:571, 2002

3. Strieter RM, Belperio JA, Keane MP: Cytokines in innate host defense in the lung. J Clin Invest 109:699, 2002

4. Musher DM: How contagious are common respiratory tract infections? N Engl J Med 348:1256, 2003

5. Halm EA, Teirstein AS: Management of community-acquired pneumonia. N Engl J Med 347:2039, 2002

6. Loeb M: Pneumonia in older persons. Clin Infect Dis 37:1335, 2003

7. Brown PD, Lerner SA: Community-acquired pneumonia. Lancet 352:1295, 1998

8. Mortensen EM, Coley CM, Singer DE, et al: Causes of death for patients with community-acquired pneumonia: results from the Pneumonia Patient Outcomes Research Team cohort study. Arch Intern Med 162:1059, 2002

9. Marrie TJ: Community-acquired pneumonia in the elderly. Clin Infect Dis 31:1066, 2000

10. Seppä Y, Bloigu A, Honkanen PO, et al: Severity assessment of lower respiratory tract infection in elderly patients in primary care. Arch Intern Med 161:2709, 2001

11. Leroy O, Devos P, Guery B, et al: Simplified prediction rule for prognosis of patients with severe community-acquired pneumonia in ICUs. Chest 116:157, 1999

12. Mortensen EM, Kapoor WN, Chang CC, et al: Assessment of mortality after long-term follow-up of patients with community-acquired pneumonia. Clin Infect Dis 37:1617, 2003

13. Kaplan V, Clermont G, Griffin MF, et al: Pneumonia: still the old man's friend? Arch Intern Med 163:317, 2003

14. Bouchud PY, Moser F, Erard P, et al: Community-acquired pneumonia: a prospective outpatient study. Medicine (Baltimore) 80:75, 2001

15. File TM Jr: Community-acquired pneumonia. Lancet 362:1991, 2003

16. Arancibia F, Bauer TT, Ewig S, et al: Community-acquired pneumonia due to gram-negative bacteria and *Pseudomonas aeruginosa:* incidence, risk, and prognosis. Arch Intern Med 162:1849, 2002

17. Hatchette TF, Gupta R, Marrie TJ: *Pseudomonas aeruginosa* community-acquired pneumonia in previously healthy adults: case report and review of the literature. Clin Infect Dis 31:1349, 2000

18. Winstead JM, McKinsey DS, Tasker S, et al: Meningococcal pneumonia: characterization and review of cases seen over the past 25 years. Clin Infect Dis 30:87, 2000

19. Marrie TJ, Peeling RW, Fine MJ, et al: Ambulatory patients with community-acquired pneumonia: the frequency of atypical agents and clinical course. Am J Med 101:508, 1996

20. Hall CB: Respiratory syncytial virus and parainfluenza virus. N Engl J Med 344:1917, 2001

21. Guidelines for prevention of nosocomial pneumonia. MMWR Morb Mortal Wkly Rep 46(RR-1):1, 1997

22. Morehead RS, Pinto SJ: Ventilator-associated pneumonia. Arch Intern Med 160:1927, 2000

23. Cunha BA: Nosocomial pneumonia: diagnostic and therapeutic considerations. Med Clin North Am 85:79, 2001

24. Mylotte JM: Nursing home–acquired pneumonia. Clin Infect Dis 35:1205, 2002

25. Vergis EN, Brennen C, Wagener M, et al: Pneumonia in long-term care: a prospective case-control study of risk factors and impact on survival. Arch Intern Med 161:2378, 2001

26. Cordero E, Pachón J, Rivero A, et al: Community-acquired bacterial pneumonia in human immunodeficiency virus–infected patients: validation of severity criteria. Am J Respir Crit Care Med 162:2063, 2000

27. Lin JC, Nichol KL: Excess mortality due to pneumonia or influenza during influenza seasons among persons with acquired immunodeficiency syndrome. Arch Intern Med 161:441, 2001

28. Park DR, Sherbin VL, Goodman MS, et al: The etiology of community-acquired pneumonia at an urban public hospital: influence of human immunodeficiency virus infection and initial severity of illness. J Infect Dis 184:268, 2001

29. Roson B, Carratala J, Verdaguer R, et al: Prospective study of the usefulness of sputum Gram stain in the initial approach to community-acquired pneumonia requiring hospitalization. Clin Infect Dis 31:869, 2000

30. Carroll KC: Laboratory diagnosis of lower respiratory tract infections: controversy and conundrums. J Clin Microbiol 40:3115, 2002

31. Reimer LG, Carroll KC: Role of the microbiology laboratory in the diagnosis of lower respiratory tract infections. Clin Infect Dis 26:742, 1998

32. Veber B, Souweine B, Gachot B, et al: Comparison of direct examination of three types of bronchoscopy specimens used to diagnose nosocomial pneumonia. Crit Care Med 28:962, 2000

33. Murdoch DR: Nucleic acid amplification tests for the diagnosis of pneumonia. Clin Infect Dis 36:1162, 2003

34. Fernandez-Sabe N, Roson B, Carratala J, et al: Clinical diagnosis of *Legionella* pneumonia revisited: evaluation of the community-based pneumonia incidence study group scoring system. Clin Infect Dis 37:483, 2003

35. Roson B, Fernandez-Sabe N, Carratala J, et al: Contribution of a urinary antigen assay (Binax NOW) to the early diagnosis of pneumococcal pneumonia. Clin Infect Dis 38:222, 2004

36. Campbell SG, Marrie TJ, Anstey R, et al: The contribution of blood cultures to the clinical management of adult patients admitted to the hospital with community-acquired pneumonia: a prospective observational study. Chest 123:1142, 2003

37. Vuori-Holopainen E, Salo E, Saxén H, et al: Etiological diagnosis of childhood pneumonia by use of transthoracic needle aspiration and modern microbiological methods. Clin Infect Dis 34:583, 2002

38. Fagon JY, Chastre J, Wolff M, et al: Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia: a randomized trial. Ann Intern Med 132:621, 2000

39. Torres A, Ewig S: Diagnosing ventilator-associated pneumonia. N Engl J Med 350:433, 2004

40. Gibot S, Cravoisy A, Levy B, et al: Soluble-triggering receptor expressed on myeloid cells and the diagnosis of pneumonia. N Engl J Med 350:451, 2004

41. Prince AS: Biofilms, antimicrobial resistance, and airway infection. N Engl J Med 347:1110, 2002

42. Mandell LA, Bartlett JG, Dowell SF, et al: Update of practice guidelines for the management of community-acquired pneumonia in immunocompetent adults. Clin Infect Dis 37:1405, 2003

43. Feldman RB, Rhew DC, Wong JY, et al: Azithromycin monotherapy for patients hospitalized with community-acquired pneumonia: a $3^1/_2$-year experience from a Veterans Affairs hospital. Arch Intern Med 163:1718, 2003

44. Sanchez F, Mensa J, Martinez JA, et al: Is azithromycin the first-choice macrolide for treatment of community-acquired pneumonia? Clin Infect Dis 36:1239, 2003

45. Nuermberger E, Bishai WR: The clinical significance of macrolide-resistant *Streptococcus pneumoniae:* it's all relative. Clin Infect Dis 38:99, 2004

46. Ailani RK, Agastya G, Ailani R, et al: Doxycycline is a cost-effective therapy for hospitalized patients with community-acquired pneumonia. Arch Intern Med 159:266, 1999

47. Akpunonu B, Michaelis J, Uy N, et al: Multicenter, postmarketing assessment of levofloxacin in the treatment of adults with community-acquired pneumonia. Clin Infect Dis 38(suppl 1):S5, 2004

48. Kahn JB, Bahal N, Wiesinger BA, et al: Cumulative clinical trial experience with levofloxacin for patients with community-acquired pneumonia-associated pneumococcal bacteremia. Clin Infect Dis 38(suppl 1):S34, 2004

49. Ruhe JJ, Hasbun R: *Streptococcus pneumoniae* bacteremia: duration of previous antibiotic use and association with penicillin resistance. Clin Infect Dis 36:1132, 2003

50. Dunbar LM, Wunderink RG, Habib MP, et al: High-dose, short-course levofloxacin for community-acquired pneumonia: a new treatment paradigm. Clin Infect Dis 37:752, 2003

51. Davidson R, Cavalcanti R, Brunton JL, et al: Resistance to levofloxacin and failure of treatment of pneumococcal pneumonia. N Engl J Med 346:747, 2002

52. Goldstein EJC, Garabedian-Ruffalo SM: Widespread use of fluoroquinolones versus emerging resistance in pneumococci. Clin Infect Dis 35:1505, 2002

53. Waterer GW, Somes GW, Wunderink RG: Monotherapy may be suboptimal for severe bacteremic pneumococcal pneumonia. Arch Intern Med 161:1837, 2001

54. Martinez JA, Horcajada JP, Almela M, et al: Addition of a macrolide to a β-lactam–based empirical antibiotic regimen is associated with lower in-hospital mortality for patients with bacteremic pneumococcal pneumonia. Clin Infect Dis 36:389, 2003

55. Chenoweth CE, Saint S, Martinez F, et al: Antimicrobial resistance in *Streptococcus pneumoniae:* implications for patients with community-acquired pneumonia. Mayo Clin Proc 75:1161, 2000

56. Castro-Guardiola A, Viejo-Rodríguez AL, Soler-Simon S, et al: Efficacy and safety of oral and early-switch therapy for community-acquired pneumonia: a randomized controlled trial. Am J Med 111:367, 2001

57. Battleman DS, Callahan M, Thaler HT: Rapid antibiotic delivery and appropriate antibiotic selection reduce length of hospital stay of patients with community-acquired pneumonia: link between quality of care and resource utilization. Arch Intern Med 162:682, 2002

58. Jackson LA, Neuzil KM, Yu O, et al: Effectiveness of pneumococcal polysaccharide vaccine in older adults. N Engl J Med 348:1747, 2003

59. Chastre J, Wolff M, Fagon J-Y, et al: Comparison of 8 vs 15 days of antibiotic therapy for ventilator-associated pneumonia in adults: a randomized trial. JAMA 290:2588, 2003

60. Collard HR, Saint S, Matthay MA: Prevention of ventilator-associated pneumonia: an evidence-based systemic review. Ann Intern Med 138:494, 2003

61. Kollef MH: The prevention of ventilator-associated pneumonia. N Engl J Med 340:627, 1999

62. Stout JE, Yu VL: Legionellosis. N Engl J Med 337:682, 1997

63. Outbreak of Legionnaires' disease among automotive plant workers, Ohio, 2001. MMWR Morb Mortal Wkly Rep 50:357, 2001

64. Sopena N, Sabria-Leal M, Pedro-Botet ML, et al: Comparative study of the clinical presentation of *Legionella* pneumonia and other community-acquired pneumonias. Chest 113:1195, 1998

65. Waterer GW, Baselski VS, Wunderink RG: *Legionella* and community-acquired pneumonia: a review of current diagnostic tests from a clinician's viewpoint. Am J Med 110:41, 2001

66. Plouffe JF, Breiman RF, Fields BS, et al: Azithromycin in the treatment of *Legionella* pneumonia requiring hospitalization. Clin Infect Dis 37:1475, 2003

67. Edelstein RH: Antimicrobial chemotherapy for Legionnaire's disease: time for a change. Ann Intern Med 129:328, 1998

68. Benin AL, Benson RF, Besser RE: Trends in Legionnaires disease, 1980–1998: declining mortality and new patterns of diagnosis. Clin Infect Dis 35:1039, 2002

69. Knirsch CA, Jakob K, Schoonmaker D, et al: An outbreak of *Legionella micdadei* pneumonia in transplant patients: evaluation, molecular epidemiology, and control. Am J Med 108:290, 2000

70. Harris A, Lally M, Albrecht M: *Legionella bozemanii* pneumonia in three patients with AIDS. Clin Infect Dis 27:97, 1998

71. Kauppinen M, Saikku P: Pneumonia due to *Chlamydia pneumoniae:* prevalence, clinical features, diagnosis, and treatment. Clin Infect Dis 21:S244, 1995

72. Compendium of measures to control *Chlamydia psittaci* infection among humans (psittacosis) and pet birds (avian chlamydiosis), 2000. MMWR Recomm Rep 49(RR-8):3, 2000

73. Principi N, Esposito S, Blasi F, et al: Role of *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* in children with community-acquired lower respiratory tract infections. Clin Infect Dis 32:1281, 2001

74. Miyashita N, Kubota Y, Nakajima M, et al: *Chlamydia pneumoniae* and exacerbations of asthma in adults. Ann Allergy Asthma Immunol 80:405, 1998

75. Marik PE: Aspiration pneumonitis and aspiration pneumonia. N Engl J Med 344:665, 2001

76. Gonzalez R, Sande ME: Uncomplicated acute bronchitis. Ann Intern Med 133:981, 2000

77. Hirshmann JV: Antibiotics for common respiratory tract infections in adults. Arch Intern Med 162:256, 2002

78. Snow V, Mottur-Pilson C, Gonzalez R, et al: Principles of appropriate antibiotic use for treatment of acute bronchitis in adults. Ann Intern Med 134:518, 2001

79. Evans AT, Husain S, Durairaj L, et al: Azithromycin for acute bronchitis: a randomized, double-blind, controlled trial. Lancet 359:1648, 2002

80. Gonzalez R, Bartlett JG, Besser RE, et al: Principles of appropriate antibiotic use for treatment of uncomplicated acute bronchitis: background. Ann Intern Med 134:521, 2001

81. Senzilet LD, Halperin SA, Spika JS, et al: Pertussis is a frequent cause of prolonged cough illness in adults and adolescents. Clin Infect Dis 32:1691, 2001

82. Arroll B, Kenealy T, Goodyear-Smith F, et al: Delayed prescriptions. BMJ 327:1361, 2003

83. Steinman MA, Landefeld CS, Gonzales R: Predictors of broad-spectrum antibiotic prescribing for acute respiratory tract infections in adult primary care. JAMA 289:719, 2003

84. Hill AT, Campbell EJ, Hill SL, et al: Association between airway bacterial load and markers of airway inflammation in patients with stable chronic bronchitis. Am J Med 109:288, 2000

85. Sethi S, Evans N, Grant BB, et al: New strains of bacteria and exacerbations of chronic obstructive pulmonary disease. N Engl J Med 34:465, 2002

86. Snow V, Lascher S, Mottur-Pilson C, et al: Evidence base for management of acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med 134:595, 2001

87. Barker AF: Bronchiectasis. N Engl J Med 346:1383, 2002

88. Bryant RE, Salmon CJ: Pleural empyema. Clin Infect Dis 22:747, 1996

89. Chen KY, Hsueh PR, Liaw YS, et al: A 10-year experience with bacteriology of acute thoracic empyema: emphasis on *Klebsiella pneumoniae* in patients with diabetes mellitus. Chest 117:1685, 2000

90. Ko SC, Chen KY, Hsueh PR, et al: Fungal empyema thoracis: an emerging clinical entity. Chest 117:1672, 2000

91. Light RW: Pleural effusion. N Engl J Med 346:1971, 2002

92. Suzuki T, Kitami A, Suzuki S, et al: Video-assisted thoracoscopic sterilization for exacerbation of chronic empyema thoracis. Chest 119:277, 2001

XXI PERITONITIS AND INTRA-ABDOMINAL ABSCESSES

W. Conrad Liles, m.d., ph.d.
E. Patchen Dellinger, m.d.

Peritonitis

Peritonitis is a diffuse or localized inflammatory process affecting the peritoneal lining.[1,2] Peritonitis has acute and chronic forms and may have a variety of causes, including pyogenic bacteria (e.g., *Escherichia coli*), tuberculosis, fungi, parasites (such as those from ruptured hepatic amebic abscesses or hydatid cysts), carcinomatosis, chemical irritation (such as that from bile), drug-hypersensitivity reaction, foreign-body reaction (such as starch peritonitis), and certain systemic illnesses (such as familial Mediterranean fever and systemic lupus erythematosus). Only acute peritonitis caused by bacteria or fungi, including primary and secondary peritonitis, is discussed here. Primary peritonitis has no underlying intra-abdominal disorder. Secondary peritonitis has an intra-abdominal focus that initiates the infection. Tuberculous peritonitis is considered elsewhere [*see 7:II Infections Due to Mycobacteria*].

SPONTANEOUS BACTERIAL PERITONITIS

Epidemiology and Etiology

Spontaneous bacterial peritonitis (SBP) is being recognized with increasing frequency in patients with either advanced chronic liver disease and concomitant ascites or fulminant hepatic failure and ascites.[3,4] The underlying chronic liver disease is usually alcoholic cirrhosis in older patients, whereas postnecrotic cirrhosis predominates in children and young adults. SBP is both common and serious, occurring in 10% to 30% of patients with cirrhosis and ascites, with an associated mortality between 30% and 50%.[2,5-7] The risk of SBP is increased in cirrhotic individuals with low ascitic protein levels (1 g/dl or lower).[8,9] *E. coli* is the most common cause of SBP and is isolated in about half of patients. Pneumococcal and streptococcal species are each responsible for 15% to 20% of cases, *Klebsiella* species for about 10%, and anaerobic or microaerophilic organisms for about 5%. *Staphylococcus aureus* is an infrequent cause of SBP but a major cause of peritonitis in cirrhotic patients with LeVeen peritoneovenous shunts. A variety of other organisms, including *Listeria monocytogenes*, *Campylobacter coli*, and *Aeromonas* species, have been responsible for isolated cases of SBP. In most instances in which the causative organism is aerobic, a single organism is involved, and concurrent bacteremia is frequent.

Although primary peritonitis occurs most often in children, it can develop in adults; nearly all adult patients are women.[10] Although many patients have had nephrosis, most have not had preexisting ascites. The source of infection is usually occult but may involve the female genital tract. The infecting organisms are almost always pneumococci or group A β-hemolytic streptococci; gram-negative bacilli are only rarely implicated. For reasons that are unclear, the incidence of primary peritonitis has decreased strikingly in the past several decades.

Occasionally, SBP develops in patients with systemic lupus erythematosus and lupus nephritis without detectable ascites, most of whom have been receiving corticosteroid therapy. The most common etiologic agents in this form of SBP are gram-positive cocci such as *Streptococcus pneumoniae* and group B streptococci. The most likely route of infection is bacteremic seeding of ascitic fluid, which may be precipitated by portal hypertension; intrahepatic shunting; intestinal bacterial overgrowth; and impaired host defense mechanisms, including diminished bactericidal activity in ascitic fluid.[3,6] Less often, SBP results from transmural migration of enteric bacteria (possibly associated with diarrhea, a common symptom in cirrhosis). Severe liver disease, hepatocellular carcinoma, gastrointestinal bleeding, and a focus of infection in the urinary tract or elsewhere in the body increase the risk of SBP.[9,11,12] Prior paracentesis may be contributory in a few instances. Penetrating lesions of the biliary tract, peptic ulcer disease, and overt bowel inflammation (such as appendicitis or diverticulitis) do not appear to be sources of infection.

Diagnosis

Clinical features The clinical presentation of SBP is often subtle.[3,6,7] Although ascites is always present, the volume of fluid may occasionally be small enough to necessitate ultrasonography for confirmation. Fever is the most common symptom but is absent in more than 30% of cases. Abdominal pain and hepatic encephalopathy are present in most patients. However, only half of the patients with SBP have abdominal tenderness, and as many as one third may be free of signs and symptoms of infection. Hence, SBP should be suspected in any cirrhotic patient who presents with unexplained clinical deterioration or hypotension.

Laboratory tests The key to diagnosis is examination of the ascitic fluid for bacteria and white blood cells. The polymorphonuclear leukocyte (PMN) count of the ascitic fluid is the best indicator of SBP. Although counts of more than 500 cells/mm^3 are considered specific for SBP, counts of 250 cells/mm^3 or more suggest a diagnosis of SBP and are considered specific enough to mandate treatment of SBP in cirrhotic patients in whom no other evidence of infection is present. An ascitic fluid PMN count below 250 cells/mm^3 excludes the diagnosis of SBP.[6,7]

Because bacterial counts are often very low, Gram stain of ascitic fluid in SBP is typically negative. However, a Gram stain is always useful because visualization of a single bacterial type would be consistent with SBP, whereas the presence of multiple bacterial forms would suggest secondary peritonitis. Because of the low concentration of bacteria, cultures are best performed by inoculation of 10 to 20 ml of ascitic fluid into a blood culture or BACTEC bottle at the bedside.

Three variants of SBP have been recognized on the basis of ascitic fluid PMN counts and cultures. In typical SBP, the PMN count is 250 cells/mm^3 or higher and cultures are positive. When the PMN count is 500 cells/mm^3 or higher but cultures are negative, the syndrome is called culture-negative neutrophilic ascites (CNNA). When the PMN count is below 250 cells/mm^3 but cultures are positive, the syndrome is termed bacterascites (BA). The clinical features and prognosis of SBP and CNNA are indistinguishable, and the two variants should

be managed identically. In contrast, BA can be self-limited; if patients are asymptomatic, they can be managed with careful observation and repeat paracentesis after 48 hours. Antibiotic therapy can be initiated if clinical symptoms develop or if the PMN count of the ascitic fluid rises.

It is important to distinguish SBP from secondary peritonitis resulting from intra-abdominal disease, such as a perforated viscus. An ascitic-fluid white blood cell count of 10,000/mm^3 or higher suggests secondary peritonitis, as does the presence of multiple bacterial species, anaerobes, or fungi. Patients should, of course, always be evaluated clinically and radiologically to exclude an underlying intra-abdominal process that might give rise to secondary peritonitis. Peripheral blood leukocytosis and positive blood cultures are common in both SBP and secondary peritonitis.

Differential Diagnosis

Bacterial peritonitis may be closely mimicked by acute pancreatitis, particularly in a patient with cirrhosis [*see also* Pancreatic Infections, *below*]. Abdominal pain, fever, rebound tenderness, hypotension, and peripheral leukocytosis are common in both bacterial peritonitis and acute pancreatitis. In a patient with pancreatitis, a diagnostic abdominal aspiration may even reveal cloudy fluid, but the turbidity is caused by floating fat globules derived from fat necrosis. Very high serum amylase levels are present in acute pancreatitis, but elevated levels also occur in peritonitis after intestinal perforation or obstruction and in the presence of renal failure.

In a patient with cirrhosis and ascites, a number of conditions may be mistaken for peritonitis, including acute peptic ulcer, cholecystitis, mesenteric artery occlusion, and other intra-abdominal processes. Paracentesis is helpful in arriving at a diagnosis in these circumstances.

Acute bacterial peritonitis may be distinguished from tuberculous peritonitis by several features. Tuberculous peritonitis is marked by a more indolent course, the absence of a peripheral leukocytosis, radiologic evidence of pulmonary tuberculosis, and a mononuclear response in the peritoneal fluid. In the patient with tuberculous peritonitis who does not have cirrhosis and ascites, the abdomen may have the characteristic so-called doughy consistency.

Peritonitis may be superficially suggested by the abdominal pain of acute porphyria, by lead colic, by diabetic acidosis, and by tabetic crisis, but the other features of these illnesses serve to distinguish them from peritonitis. The signs and symptoms of familial Mediterranean fever—high temperature, abdominal pain, abdominal guarding, and peripheral leukocytosis—may suggest bacterial peritonitis. The periodicity of familial Mediterranean fever and its occurrence predominantly in persons of Sephardic, Armenian, and Arab ancestry are helpful in differentiating it from bacterial peritonitis.

SBP may be difficult to diagnose in a patient with systemic lupus erythematosus who experiences acute abdominal pain and fever. These symptoms may stem from a variety of independent surgical problems (e.g., perforated ulcer, intestinal obstruction, and mesenteric occlusion) that must be distinguished from abdominal problems directly related to lupus, such as vasculitis, pancreatitis secondary to vasculitis or corticosteroid therapy, and SBP. Examination of peritoneal fluid obtained by paracentesis, by culdocentesis, or during laparotomy may be the only way to determine the presence of bacterial peritonitis.

Treatment

Until culture results are available, broad coverage should be directed against enteric organisms. Nephrotoxic drugs, including aminoglycosides, should be avoided whenever possible.[13] Cefotaxime (2 g I.V. every 8 hours) has emerged as the favored agent for the empirical treatment of SBP; alternative useful agents include ceftriaxone, ceftazidime, cefonicid, ceftizoxime, ampicillin-sulbactam, meropenem, and imipenem-cilastatin, as well as fluoroquinolones (i.e., ciprofloxacin, levofloxacin, gatifloxacin, and moxifloxacin).[6,7] Traditionally, intravenous antibiotics have been administered for 10 to 14 days, but 5 days of therapy appears to be as effective, provided that the patient is doing well clinically and that the ascitic fluid is sterile, with a PMN count that is below 250 cells/mm^3 before discontinuance of antibiotics.[6,7,14]

Renal failure is a frequent complication of SBP; peripheral vasodilation and renal vasoconstriction are probably responsible.[15] Infusions of albumin (1.5 g/kg at the time of diagnosis and 1 g/kg on day 3) can substantially reduce the incidence of renal failure and mortality in patients with SBP.[16]

Prophylaxis Because patients with cirrhosis are at high risk for primary SBP and because recurrences develop in 43% of these patients within 6 months and in 69% within 1 year after an initial episode of SBP, both primary and secondary prophylaxis regimens are now recommended for certain subgroups of patients. In nonbleeding cirrhotic patients with persistent ascites after an initial episode of SBP, continuous secondary prophylaxis with oral norfloxacin (400 mg daily) is currently recommended. Alternative oral antimicrobial agents for prophylaxis include ciprofloxacin, levofloxacin, trimethoprim-sulfamethoxazole, and amoxicillin-clavulanate. In cirrhotic patients with upper gastrointestinal hemorrhage, primary prophylaxis with oral norfloxacin (400 mg every 12 hours) or alternative systemic therapy (ciprofloxacin, levofloxacin, ampicillin-sulbactam) for a minimum of 7 days is advised. Primary prophylaxis with norfloxacin or another fluoroquinolone should also be considered in cirrhotic patients with low ascitic protein levels (i.e., less than 1.5 g/L).[6,7,17-19] Appropriate prophylaxis not only reduces the incidence of SBP but also improves overall survival.[20]

SECONDARY PERITONITIS

Etiology

Secondary peritonitis occurs as a complication of intra-abdominal disease. It may result from appendicitis, diverticulitis, penetrating abdominal wounds, blunt trauma to the abdomen, perforation of the gastrointestinal tract (e.g., by a peptic ulcer or bowel neoplasm), or rupture of an intra-abdominal abscess. Secondary peritonitis can be divided into spontaneous cases caused by an underlying disease such as appendicitis or diverticulitis and cases that result from a ruptured viscus incurred from an operation, a procedure, or an episode of trauma.[21] Most of these infections are polymicrobial. The pathogens include both anaerobic species (principally *Bacteroides fragilis,* peptococci, and peptostreptococci) and aerobic species (*E. coli*, *Proteus* species, *Klebsiella* species, and various streptococci and enterococci).[22] Bacteremia—which occurs in only 20% to 30% of cases—is most commonly caused by *E. coli*, *Bacteroides* species, or both.[23-25] The prognosis of secondary peritonitis depends on the underlying cause and the patient's physiologic response to the infection. Mortality is lowest in patients with appendicitis or

perforated duodenal ulcer (10%) and highest in those with other intra-abdominal processes (50%) or postoperative peritonitis (60%). Both mortality and the likelihood of complications, including the necessity for a second operation, increase as the patient's physiologic response to the disease is more marked. This is most easily assessed using the Acute Physiology and Chronic Health Evaluation II (APACHE II) score.[23,24]

Tertiary peritonitis is a relatively new term that refers to the persistence of intra-abdominal infection after the initial surgical and medical treatment of secondary peritonitis.[26] Not all authors discussing tertiary peritonitis agree on the definition of this syndrome. In a typical case of tertiary peritonitis, operative exploration in a patient with signs and symptoms of peritonitis after prior treatment for peritonitis will reveal inflammation and bacterial growth despite the absence of a focus for continuing infection, such as a perforated viscus, gangrenous tissue, or abscess. The organisms recovered tend to be considered nonpathogens and tend not to be typical of enteric flora. Tertiary peritonitis can be considered evidence for a type of host defense failure.

Diagnosis

The clinical features of secondary peritonitis most commonly include peritoneal signs such as involuntary guarding, referred percussion tenderness, and abdominal tenderness. There may be abdominal distention. Testing for rebound tenderness is painful to the patient and unhelpful in diagnosis. The abdomen is often not silent. Fever and leukocytosis are usually present. Free air may or may not be visible on plain abdominal films. An ultrasound or CT scan showing peritoneal free fluid or gas in association with a compatible clinical picture confirms the diagnosis. If the diagnosis is clinically obvious, radiographic studies are not required.

In the absence of ascites, the peritoneum has a remarkable capacity to wall off and localize infection to a portion of the abdomen. Thus, in situations such as rupture of the appendix, a localized peritonitis develops and results in a periappendiceal or pelvic abscess [*see* Intra-abdominal Abscesses, *below*].

Treatment

Peritonitis secondary to bowel perforation, ruptured gangrenous appendix, or penetrating trauma requires prompt surgical intervention in addition to antimicrobial therapy. Surgical therapy, known as source control, should be directed at correction of the underlying disease, debridement of the surrounding tissues, and prevention of recurrent microbial soilage. In general, this requires draining, resecting, excluding, or patching the involved viscus during laparotomy or laparoscopy. Intraoperative saline lavage and radical peritoneal debridement have not proved useful; postoperative peritoneal lavage and planned repeat laparotomy have been suggested but are not of established benefit.

Approximately 20% to 30% of patients who require an operation for treatment of peritonitis or intra-abdominal abscess will require a second operative procedure to resolve the infection and establish adequate source control.[27] Surgeons' assessments of the adequacy of the source control achieved with the original operative procedure have been found to be strongly predictive of patients' subsequent need for reoperation and mortality.[28]

The choice of antimicrobial agent depends on the organisms involved in the peritonitis. Initial selection, however, is always made before culture results are available, and it must take into consideration the organisms that are predominant in the colon: *B. fragilis*, enteric gram-negative bacilli, streptococci, and enterococci. Animal studies and clinical experience have shown the importance of using antibiotics that are effective against both aerobic and anaerobic bacteria to treat patients with polymicrobial peritonitis, but clinical trials have failed to establish the superiority of any particular regimen. Newer antibiotics that provide broad-spectrum coverage of many aerobic and anaerobic species allow single antibiotic therapy in many cases. Useful single agents include ampicillin-sulbactam, ticarcillin–clavulanic acid, cefoxitin, cefotetan, piperacillin-tazobactam, ertapenem, meropenem, and imipenem-cilastatin. Effective multidrug regimens include (1) an aminoglycoside combined with either clindamycin or metronidazole, (2) aztreonam combined with clindamycin, and (3) a quinolone or a third-generation cephalosporin combined with either clindamycin or metronidazole[29,30] [*see 7:XIV Chemotherapy of Infection*]. Although multiple drugs provide broader antimicrobial coverage, they do not appear to be more effective than single-drug regimens. In all cases, the final choice of antibiotics should be determined by the results of culture and sensitivity testing and the clinical course.

Most antibiotics attain concentrations in ascitic fluid that are at least half of the simultaneous serum levels and that exceed the minimum inhibitory concentration for the infecting organism. For this reason, systemic therapy alone is generally adequate in the management of bacterial peritonitis in patients with ascites; intraperitoneal instillation of antibiotics does not appear necessary. The necessary duration of antibiotic administration has never been systematically studied. For most patients, antimicrobial agents can be stopped as soon as clinical signs of infection begin to resolve, intestinal function resumes, and temperature and white blood cell count begin to return to normal. It is unusual for patients to require more than 7 days of treatment, and many patients can be managed with less.[29,30]

PERITONITIS IN DIALYSIS PATIENTS

Epidemiology

Infection continues to be a significant problem for peritoneal dialysis patients. Peritonitis develops in as many as 60% of patients undergoing continuous ambulatory peritoneal dialysis during the first year of treatment, and the infection recurs in 20% to 30% of these patients; elderly patients are the most vulnerable.[31]

Diagnosis

Clinical features The development of fever, abdominal pain or tenderness, and leukocytosis and the isolation of a bacterial or mycotic agent from the effluent fluid in a patient on peritoneal dialysis indicate peritonitis. Isolation of bacteria from the dialysate in a patient without these findings often signals contamination rather than infection. Turbidity of the dialysate from neutrophils occurs in 2% to 3% of dialyses. Although turbidity itself does not necessarily indicate peritonitis, it should be considered an indication of infection until the results of the culture are available. Absence of bacteria on a Gram stain of the dialysate sediment does not necessarily confirm the absence of infection, however, because of the extensive dilutions required. Therefore, one cannot rely on a negative Gram stain to discriminate between infection and sterile inflammation.[31]

Bacteriology The principal organisms in peritonitis that complicate peritoneal dialysis are coagulase-negative staphylo-

cocci, *S. aureus, Pseudomonas aeruginosa, E. coli* and other enteric organisms, and *Candida* species.[31-33] Microorganisms may enter the peritoneal cavity exogenously (i.e., after colonization of the abdominal wound area or by contamination of the dialysate) or endogenously (i.e., by bacteremia or by transmural migration of bowel flora, perhaps enhanced by catheter-induced trauma). Most episodes are monomicrobial, but polymicrobial peritonitis can occur.[34,35] Failure to respond to antibiotic therapy within 96 hours often signals infection with gram-negative bacilli (typically, *P. aeruginosa*); the prognosis in these patients is worse than in those who respond rapidly. Removal of the dialysis catheter may be necessary to control infection.

Treatment

Peritonitis caused by *Candida* species occurs most often as a complication of peritoneal dialysis, gastrointestinal surgery, or perforation of an abdominal viscus. Candidal peritonitis that complicates peritoneal dialysis is treated with intravenous amphotericin B, intraperitoneal amphotericin B, or both at a final dialysate concentration of 2 to 4 μg/ml. Fluconazole may also be useful for treatment of peritonitis caused by *C. albicans,* as may caspofungin.[31,36]

Whereas the addition of peritoneal lavage, with or without antibiotics, does not appear to improve on the results achieved by intravenous antibiotics and conventional surgical therapy, intraperitoneal administration of antibiotics may be useful in patients who require peritoneal dialysis. For example, different antibiotics can be added directly to the dialysate in specific concentrations, such as ampicillin in a concentration of 50 mg/L or gentamicin in a concentration of 5 to 10 mg/L. Because bacteremia may occur, antibiotics should also be administered intravenously in these patients in a dosage appropriate to the patient's level of renal function. When peritonitis develops as a complication of peritoneal dialysis or peritoneovenous shunting, it is often necessary to remove or replace the catheter during administration of antibiotics to control the peritonitis.[31,36] In patients with a history of peritonitis caused by *S. aureus,* prophylaxis with either topical mupirocin ointment in the nares or oral rifampin may reduce the incidence of subsequent episodes of staphylococcal peritonitis and peritoneal catheter loss.[31,36,37]

Intra-abdominal Abscesses

Intra-abdominal abscesses may present as complications of an abdominal operation, intra-abdominal conditions (e.g., diverticulitis, appendicitis, biliary tract disease, pancreatitis, perforated viscus, peritonitis), or penetrating abdominal trauma; as fever of obscure origin; or as dysfunction of neighboring organs (e.g., so-called lower lobe pneumonia related to a subphrenic abscess or small bowel obstruction). Bacteremic spread of infection from a distant focus to an intra-abdominal site is a less common cause of intra-abdominal abscesses.

Intra-abdominal abscesses are conveniently classified according to the anatomic location in which they occur: intraperitoneal, retroperitoneal, or visceral. Intraperitoneal abscesses are areas of localized peritonitis in which infection has progressed but has been walled off by omentum, peritoneum, and adjacent viscera. Retroperitoneal infections include pancreatitis-associated infections, perinephric abscesses, and paravertebral abscesses. Visceral abscesses develop within abdominal viscera—predominantly, the liver and, less often, the spleen—and other organs. In general, the location of the abscess does not affect the diagnosis or treatment beyond influencing the choice of percutaneous or surgical drainage.

GENERAL APPROACH TO INTRA-ABDOMINAL ABSCESSES

Diagnosis

Although the location of the abscess determines its particular features, many intra-abdominal infections share common elements. Fever, for example, is almost invariable; it often recurs in a spiking pattern and may be accompanied by rigors. Hypotension and even septic shock may develop. Abdominal pain is a major clue to the presence of an intra-abdominal abscess: when present, it can predominate; when it is absent, the diagnosis can be very difficult. Geriatric patients, in particular, may present atypically, without abdominal pain or fever. On laboratory studies, leukocytosis and elevation of liver enzyme and serum amylase levels are common. Bacteremia, which may be polymicrobial, occurs in up to one fourth of cases.

Intra-abdominal abscesses characteristically contain multiple bacterial species. Anaerobic bacteria can be isolated from 60% to 70% of such abscesses; the bacteria most commonly isolated include *B. fragilis,* peptostreptococci and peptococci, *Clostridium* species, and facultative species such as *E. coli, Enterobacter, Klebsiella,* and enterococci. The specific organisms isolated do not generally provide clues to the nature of the underlying process. However, the presence of *Citrobacter* species strongly suggests a biliary or upper gastrointestinal source; *S. aureus,* otherwise uncommon in intra-abdominal abscesses, suggests bacteremic seeding or vertebral osteomyelitis, which can lead to retroperitoneal or perinephric abscesses.

Plain radiography (kidneys, ureters, and bladder [KUB]; upright and lateral decubitus views) may afford important clues to the diagnosis of intra-abdominal abscesses. For example, air-fluid levels may indicate an intra-abdominal collection, free air may point to perforation of a viscus as the underlying problem, displaced loops of bowel may signify an abscess, and a so-called soap-bubble appearance or loss of the normal psoas shadow may suggest a retroperitoneal collection.

Ultrasonography and computed tomographic scanning, however, are much more sensitive and specific than plain radiography and are now the standard radiologic techniques for evaluating intra-abdominal abscesses.[38] Both are excellent for diagnosis, and both can be used to guide percutaneous abscess drainage [*see* Treatment and Prevention, *below*].[39] CT scanning is the more accurate study; its specificity and sensitivity rates can exceed 90%. Compared with ultrasonography, CT scanning has the additional advantages of allowing simultaneous administration of contrast, of not requiring skin contact (hence, surgical dressings do not interfere with the study), and of producing accurate results even in the presence of ileus and abdominal gas collections. Ultrasonography, however, is less expensive, is often more readily available, can sometimes be done with portable equipment at the bedside, and does not involve exposure to radiation. Ultrasonography is most accurate for detecting abscesses in the left or right upper quadrant of the abdomen and in the true pelvis; it is also sensitive and specific for identifying ascites. In patients with acute abdominal disease, however, ultrasonography is often limited by bowel gas, which obscures any deeper findings.

Magnetic resonance imaging plays a negligible role in the evaluation of intra-abdominal infections. Nuclear medicine studies are also less helpful than CT and ultrasonography. Although early results appeared promising, gallium-67 scanning

and indium-111 scanning both have proved less helpful than CT and ultrasonographic techniques. Cholescintigraphy scans using technetium-99m–labeled hepatoiminodiacetic acid (HIDA, or lidofenin) are useful in evaluating the gallbladder and for demonstrating a bile leak after cholecystectomy or other biliary procedure [*see 4:VI Gallstones and Biliary Tract Disease*]. Arteriography and barium contrast studies are seldom used to diagnose intra-abdominal abscesses. If fistulous tracts are present, however, sinograms may occasionally be helpful.

Treatment and Prevention

The choice of antibiotics depends on the organisms isolated from cultures of blood or abscess material. Until this information is available, the choice of drugs should be guided by the same principles as those that apply to the treatment of peritonitis. Although the use of antibiotics is essential, especially because of the risk of bacteremia, such therapy alone will not eradicate intra-abdominal abscesses and is therefore secondary to prompt, effective abscess drainage.

Until the mid-1970s, surgical drainage was mandatory for the treatment of intra-abdominal abscesses. However, treatment changed dramatically within just a few years of the introduction of percutaneous abscess drainage under ultrasonographic or CT guidance. Ultrasonography may be used to guide drainage of large or superficial collections, but CT is preferable for smaller or deeper abscesses.[40] Many studies have demonstrated that percutaneous abscess drainage is safe and effective for a broad range of intra-abdominal collections; success rates range from 47% to 92%, with most studies reporting better than 80% success, similar to the success rate for surgical drainage.[41] Failure of treatment is more common in immunosuppressed patients and in those with poorly defined phlegmons, multilocular abscesses, thick hematomas or organized infections, or abscesses with associated fistulous tracts.

Radiographic features alone cannot indicate which abscesses will respond to percutaneous drainage. Hence, it seems reasonable to institute percutaneous drainage in all patients who have a safe access route, provided that skilled personnel are available and that the patient does not otherwise require surgical intervention. Surgical drainage may then be used in patients with recurrences, failures, or complications. A surgeon should be involved in the decision regarding the method of drainage, because the surgeon will be called if the initial approach is not successful. Even in the case of abscesses that usually require surgical intervention, such as periappendiceal and diverticular abscesses and peripancreatic infections (see below), percutaneous drainage can provide temporary control of sepsis, allowing the operative procedure to be delayed until conditions are optimal and sometimes allowing a single definitive procedure instead of staged procedures.

A number of effective preventive strategies are available to reduce the likelihood of both wound infections (incisional surgical site infections) and intra-abdominal abscesses (organ/space surgical site infections) after abdominal operations.[42] These include the appropriate use of prophylactic antibiotics,[43] maintenance of normothermia in the operating room,[44,45] provision of high levels of inspired oxygen,[45,46] adequate fluid resuscitation during the operation,[47] and maintenance of euglycemia in the perioperative period.[48-50]

INTRAPERITONEAL ABSCESSES

Intraperitoneal abscesses may form in either of two ways: (1) from diffuse peritonitis in which loculations of pus develop in anatomically dependent areas such as the pelvis, paracolic gutters, and subphrenic areas or (2) by spread of infection from a localized inflammatory process to contiguous peritoneum. About one third of intra-abdominal abscesses are intraperitoneal, and almost one half of intraperitoneal abscesses occur in the right lower quadrant.

Subphrenic Abscesses

About 60% of subphrenic abscesses develop after operations involving the duodenum and stomach, biliary tract, or appendix; 20% to 40% develop after rupture of a hollow viscus (such as perforated peptic ulcer or acute appendicitis), in which the infection is subsequently sealed off. A variable percentage of subphrenic abscesses develop after penetrating or blunt (closed) abdominal trauma, and less than 5% develop without predisposing circumstances. Diagnosis of subphrenic abscesses is sometimes delayed because of their location in the intrathoracic portion of the peritoneal cavity, which is not amenable to examination.

Clinical features The manifestations of a subphrenic abscess range from a severe acute illness to an insidious chronic process characterized by intermittent fever, weight loss, anemia, and nonspecific symptoms. The chronic syndrome is most often observed in patients who have previously received antibiotics; in the past, such an abscess could smolder subclinically for prolonged periods before diagnosis. This is currently uncommon. In any patient with fever of undetermined origin who has had an abdominal operation—even if the operation was performed many months earlier—a chronic intra-abdominal abscess must be suspected and a CT scan should be done.

Spiking fever, abdominal pain and tenderness (most often at the lower costal margin), and weight loss are common manifestations. Features of an intrathoracic process, such as shoulder pain, chest pain, cough, dyspnea, rales, and pleural effusion, are more commonly observed than features of an intra-abdominal condition. Leukocytosis is common. Rarely, patients will have a prolonged, obscure febrile illness complicated by the sudden development of an empyema when the subphrenic abscess ruptures through the diaphragm. Although pleural fluid is present in about 80% of patients with a subphrenic abscess, it is usually a sympathetic transudate. A pleural effusion that develops after an abdominal operation is more commonly caused by inflammation below the diaphragm than inflammation above it.

Diagnosis CT scanning and ultrasonography are the best radiologic techniques for establishing the diagnosis. The plain x-ray findings in patients with subphrenic abscess include pleural effusion, limitation of diaphragmatic movement, elevation of a hemidiaphragm, and lower-lobe pneumonia or atelectasis.

RETROPERITONEAL ABSCESSES

Pyogenic infections of the retroperitoneum present like other intra-abdominal infections. Indeed, many retroperitoneal abscesses arise from disorders of the abdominal viscera; more than two thirds of patients with retroperitoneal abscesses also have underlying debilitating conditions, including malignancies, corticosteroid use, alcoholism, and diabetes. More than 80% of these infections are polymicrobial, involving aerobic and anaerobic enteric organisms.[51] CT scanning is the key to diagnosis of retroperitoneal abscesses. The same is true for primary psoas abscesses, which are often caused by *S. aureus*,[52] and for perinephric abscesses, which usually originate in the urinary

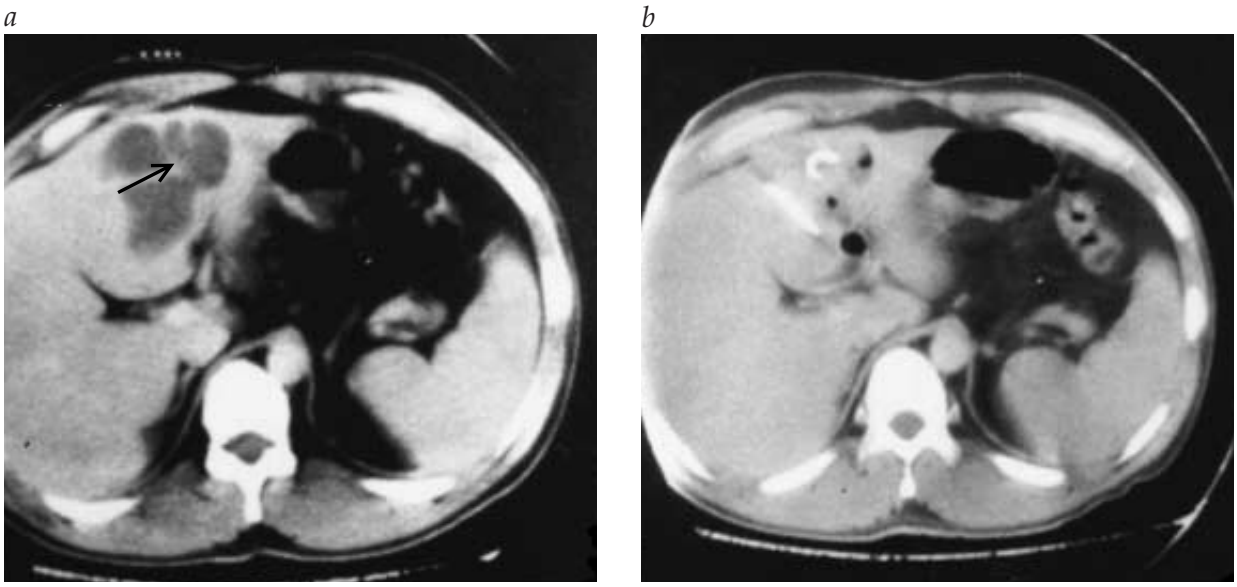


Figure 1 **CT scan (*a*) shows multilobular liver abscess (arrow). Four days after percutaneous abscess drainage, CT scan (*b*) shows resolution of the abscess cavity.**

tract.[53] As with other abscesses, successful management requires prompt percutaneous or surgical drainage and the administration of appropriate antibiotics.[53,54]

Pancreatic Infections

Most pancreatic infections occur as a complication of pancreatitis, which can result from alcoholism (38%), gallstones (11%), surgical trauma (16%), or other factors (35%). Pancreatic infections have been divided into infected pancreatic and peripancreatic necrosis and pancreatic abscesses.[55] Infected pancreatic necrosis tends to occur during the first 3 weeks after the onset of acute necrotizing pancreatitis and is poorly localized within the retroperitoneum. Source control is difficult, and morbidity and mortality can be high. Infected necrosis often requires open operative debridement, although there are promising reports of combined percutaneous drainage and laparoscopically assisted debridement. Pancreatic abscess refers to a more localized infectious process. It can occur either in or adjacent to the pancreas and tends to occur more than 3 weeks after the acute onset of disease. An infected pseudocyst is also a pancreatic abscess. Pancreatic abscesses can often be managed percutaneously but are still likely to require additional source control.[56]

Pancreatic infections are often polymicrobial, typically containing three or four species of bacteria. Most are enteric organisms, including *E. coli,* enterococci, *Klebsiella* species, and anaerobes such as *Bacteroides, Peptococcus, Fusobacterium,* and *Clostridium* species. Nonenteric organisms, including staphylococci, *P. aeruginosa,* and, less often, *Candida* species, may be involved. Bacteremia occurs in about 26% of cases. Recent reports demonstrate a shift in the usual microbial flora of infected pancreatitis, with an increase in gram-positive cocci and fungi (e.g., *Candida* species), probably secondary to the increasingly common use of prophylactic antibiotics for long periods in patients with necrotizing pancreatitis.[57,58]

Clinical features The initial presentation of noninfected acute necrotizing pancreatitis involves fever, leukocytosis, and abdominal pain and tenderness. The clinical features do not allow the differentiation between infected and uninfected patients. Most infections occur after at least 1 week of disease.

Diagnosis The most accurate method for determining whether an area of pancreatic or peripancreatic necrosis is infected is to perform CT scan–guided fine-needle aspiration for Gram stain and culture. This step is indicated if a patient's clinical condition deteriorates after initial stabilization or improvement.[59]

Treatment Source control is mandatory. It can often be accomplished by open surgical debridement, sometimes aided by percutaneous drainage, laparoscopic techniques, or both. The outcome is improved when intervention occurs later in the course of the disease. Antibiotic therapy is the same as for other intra-abdominal infections.

VISCERAL ABSCESSES

Liver Abscesses

Epidemiology and etiology Pyogenic liver abscesses occur in several settings, including biliary tract infection, direct extension from a contiguous site of infection, portal bacteremia from intra-abdominal septic foci, and nonpenetrating trauma.[60-62] Liver abscesses may occur as a result of systemic bacteremia or as complications of abdominal surgery or penetrating abdominal trauma. They may also occur as complications of hepatocellular carcinoma,[63] chronic granulomatous disease,[64,65] or percutaneous transhepatic biliary drainage procedures in patients with cancer and obstructive jaundice. Pyogenic abscesses may be single or multiple.

Like other intra-abdominal abscesses, pyogenic liver abscesses principally involve enteric bacteria; two thirds of these abscesses have polymicrobial origins, and at least one third involve anaerobes. *S. aureus* may be the causative organism in patients with bacteremia and in children. *Klebsiella* species are often responsible for gas-forming liver abscesses, which typically occur in patients with diabetes.[66] Blood cultures are positive

in about half of patients with a pyogenic liver abscess, and metastatic infections may occur.

Clinical features Fever is the most common symptom and is present in nearly 90% of patients. Chills and weight loss occur in about half of cases. Because abdominal pain, abdominal tenderness, or hepatomegaly is present in only half of cases, many of these patients present with fever of undetermined origin. Leukocytosis is present in most cases. Jaundice is infrequent, but the serum alkaline phosphatase level is elevated in almost all patients. Rupture of a liver abscess, although uncommon, is often accompanied by diffuse abdominal pain and septic shock.[67]

Diagnosis CT scanning is the most accurate diagnostic technique [*see Figure 1*], yielding positive results in up to 95% of confirmed cases; ultrasonography is also helpful, yielding positive results in up to 80% of confirmed cases. The initial clue to the diagnosis may come from plain x-rays, which may show an elevated right hemidiaphragm, a right pleural effusion, or an air-fluid level.

The major differential diagnosis is amebic liver abscess [*see 7:XXXIV Protozoan Infections*]. Amebic abscesses are more likely to be solitary and confined to the right lobe of the liver; a history of travel or diarrhea may suggest the diagnosis. Stool ova and parasite examination revealing *Entamoeba histolytica* is highly suggestive, but results are often negative in patients with hepatic amebiasis. However, most patients with amebic liver abscesses have positive amebic serologies. It is important to note that *E. histolytica* has been reclassified into two morphologically similar but genetically distinct species: *E. histolytica,* the pathogenic protozoan that causes amebic dysentery and hepatic abscess, and *E. dispar*, a nonpathogenic commensal protozoan of humans. Specific enzyme-linked immunosorbent assay–based serologic testing to distinguish *E. dispar* colonization from *E. histolytica* infection is recommended before treating amebiasis.[68,69] Less often, hepatic cysts or neoplasms may be confused with liver abscesses.

Treatment Whereas surgery was formerly the mainstay of therapy, percutaneous drainage should now be the initial drainage procedure in most patients with pyogenic liver abscesses. Antibiotics with broad coverage of enteric organisms and staphylococci should be administered intravenously until specific pathogens have been isolated from the abscess or the bloodstream.[70] Mortality depends largely on the underlying disease and is highest in patients with cancer.[71] Surgical therapy is required for ruptured abscesses, but the mortality is high, approaching 44%.[67]

Splenic Abscesses

Splenic abscesses are uncommon.[72] Unlike other intra-abdominal abscesses, they are often bacteremic in origin, especially in patients with endocarditis. In other patients, hemoglobinopathy, vasculitis with splenic infarction, trauma, and immunosuppression may be predisposing factors. Fever and chills and left upper quadrant pain are common. If the upper pole of the spleen is affected, diaphragmatic and pleural and pulmonary symptoms may predominate, but peritoneal symptoms are more common if the lower pole is the site of infection. Responsible organisms include *S. aureus,* streptococci, *Salmonella* species, and enteric bacteria; fungi are important causes in immunocompromised patients. CT scans and ultrasonography are the most useful radiographic studies. Appropriate antimicrobial therapy is essential. Although splenectomy has often been required for effective management in the past, evidence now indicates that percutaneous drainage or even antibiotics alone may suffice in selected cases.[73] More experience is needed before the optimal management of these uncommon infections can be determined.[74]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Johnson CC, Baldessarre J, Levison ME: Peritonitis: update on pathophysiology, clinical manifestations, and management. Clin Infect Dis 24:1035, 1997

2. Hall JC, Heel KA, Papadimitriou JM, et al: The pathology of peritonitis. Gastroenterology 114:185, 1998

3. Such J, Runyon BA: Spontaneous bacterial peritonitis. Clin Infect Dis 27:669, 1998

4. Dhiman RK, Makharia GK, Jain S, et al: Ascites and spontaneous bacterial peritonitis in fulminant hepatic failure. Am J Gastroenterol 95:233, 2000

5. Thuluvath PJ, Morss S, Thompson R: Spontaneous bacterial peritonitis: in-hospital mortality, predictors of survival, and health care costs from 1988 to 1998. Am J Gastroenterol 96:1232, 2001

6. Garcia-Tsao G: Currrent management of the complications of cirrhosis and portal hypertension: variceal hemorrhage, ascites, and spontaneous bacterial peritonitis. Gastroenterology 120:726, 2001

7. Rimola A, Garcia-Tsao G, Navasa M, et al: Diagnosis, treatment, and prophylaxis of spontaneous bacterial peritonitis: a consensus document. International Ascites Club. J Hepatol 32:142, 2000

8. Guarner C, Sola R, Soriano G, et al: Risk of a first community-acquired spontaneous bacterial peritonitis in cirrhotics with low ascitic fluid protein levels. Gastroenterology 117:414, 1999

9. Yoneyama K, Miyagishi K, Kiuschi Y, et al: Risk factors for infections in cirrhotic patients with and without hepatocellular carcinoma. J Gastroenterol 37:1028, 2002

10. Hemsley C, Eykyn SJ: Pneumococcal peritonitis in previously healthy adults: case report and review. Clin Infect Dis 27:376, 1998

11. Hillebrand DJ, Runyon BA: Spontaneous bacterial peritonitis: keys to management. Hosp Pract (Off Ed) 35:87, 2000

12. Llovet JM, Moitinho E, Sala M, et al: Prevalence and prognostic value of hepatocellular carcinoma in cirrhotic patients presenting with spontaneous bacterial peritonitis. J Hepatol 33:423, 2000

13. Hampel H, Bynum GD, Zamora E, et al: Risk factors for the development of renal dysfunction in hospitalized patients with cirrhosis. Am J Gastroenterol 96:2206, 2001

14. Franca A, Giordano HM, Seva-Pereira T, et al: Five days of ceftriaxone to treat spontaneous bacterial peritonitis in cirrhotic patients. J Gastroenterol 37:119, 2002

15. Runyon BA: Albumin infusion for spontaneous bacterial peritonitis. Lancet 354:1838, 1999

16. Sort P, Navasa M, Arroyo V, et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 341:403, 1999

17. Grange JD, Roulot D, Pelletier G, et al: Norfloxacin primary prophylaxis of bacterial infections in cirrhotic patients with ascites: a double-blind randomized trial. J Hepatol 29:430, 1998

18. Gines P, Navasa M: Antibiotic prophylaxis for spontaneous bacterial peritonitis: how and whom? J Hepatol 29:490, 1998

19. Bernard B, Grange JD, Khac EN, et al: Antibiotic prophylaxis for the prevention of bacterial infections in cirrhotic patients with gastrointestinal bleeding: a meta-analysis. Hepatology 29:1655, 1999

20. Soares-Weiser K, Brezis M, Tur-Kaspa R, et al: Antibiotic prophylaxis for cirrhotic patients with gastrointestinal bleeding. Cochrane Database Syst Rev 2:CD002907, 2002

21. Seiler CA, Brugger L, Forssmann U, et al: Conservative surgical treatment of diffuse peritonitis. Surgery 127:178, 2000

22. Wilson SE, Hopkins JA: Clinical correlates of anaerobic bacteriology in peritonitis. Clin Infect Dis 20(suppl 2):S251, 1995

23. Barie PS, Vogel SB, Dellinger EP, et al: A randomized, double-blind clinical trial comparing cefepime plus metronidazole with imipenem-cilastatin in the treatment of complicated intra-abdominal infections. Cefepime Intra-abdominal Infection Study Group. Arch Surg 132:1294, 1997

24. Dellinger EP, Wertz MJ, Meakins JL, et al: Surgical infection stratification system for intra-abdominal infection: multicenter trial. Arch Surg 120:21, 1985

25. Lennard ES, Minshew BH, Dellinger EP, et al: Stratified outcome comparison of clindamycin-gentamicin vs chloramphenicol-gentamicin for treatment of intra-abdominal sepsis. Arch Surg 120:889, 1985

26. Nathens AB, Rotstein OD, Marshall JC: Tertiary peritonitis: clinical features of a complex nosocomial infection. World J Surg 22:158, 1998

27. Dellinger EP: Surgery for intra-abdominal sepsis. Reoperative Gastrointestinal

Surgery. White TT, Harrison RC, Debas HT, et al, Eds. Appleton Century Crofts, New York, 1989, p 63

28. Koperna T, Schulz F: Relaparotomy in peritonitis: prognosis and treatment of patients with persisting intraabdominal infection. World J Surg 24:32, 2000

29. Mazuski JE, Sawyer RG, Nathens AB, et al: The Surgical Infection Society guidelines on antimicrobial therapy for intra-abdominal infections: an executive summary. Surg Infect (Larchmt) 3:161, 2002

30. Mazuski JE, Sawyer RG, Nathens AB, et al: The Surgical Infection Society guidelines on antimicrobial therapy for intra-abdominal infections: evidence for the recommendations. Surg Infect (Larchmt) 3:175, 2002

31. Vas S, Oreopoulos DG: Infections in patients undergoing peritoneal dialysis. Infect Dis Clin North Am 15:743, 2001

32. Zelenitsky S, Barns L, Findlay I, et al: Analysis of microbiological trends in peritoneal dialysis-related peritonitis from 1991 to 1998. Am J Kidney Dis 36:1009, 2000

33. Bibashi E, Memmos D, Kokolina E, et al: Fungal peritonitis complicating peritoneal dialysis during an 11-year period: report of 46 cases. Clin Infect Dis 36:927, 2003

34. Kim GC, Korbet SM: Polymicrobial peritonitis in continuous ambulatory peritoneal dialysis patients. Am J Kidney Dis 36:1000, 2000

35. Szeto CC, Chow KM, Wong TY, et al: Conservative management of polymicrobial peritonitis complicating peritoneal dialysis: a series of 140 consecutive cases. Am J Med 113:728, 2002

36. Keane WF, Bailie GR, Boeschoten E, et al: Adult peritoneal dialysis-related peritonitis treatment recommendations: 2000 update. Perit Dial Int 20:396, 2000

37. Ritzau J, Hoffman RM, Tzamaloukas AH: Effect of preventing *Staphylococcus aureus* carriage on rates of peritoneal catheter-related staphylococcal infections: literature synthesis. Perit Dial Int 21:471, 2001

38. Gupta H, Dupuy DE: Advances in imaging of the acute abdomen. Surg Clin North Am 77:1245, 1997

39. Sheafor DH, Paulson EK, Simmons CM, et al: Abdominal percutaneous interventional procedures: comparison of CT and US guidance. Radiology 207:705, 1998

40. Montgomery RS, Wilson SE: Intraabdominal abscesses: image-guided diagnosis and therapy. Clin Infect Dis 23:28, 1996

41. Levison MA: Percutaneous versus open operative drainage of intra-abdominal abscesses. Infect Dis Clin North Am 6:525, 1992

42. Horan TC, Gaynes RP, Martone WJ, et al: CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. Am J Infect Control 20:271, 1992

43. Dellinger EP, Gross PA, Barrett TL, et al: Quality standard for antimicrobial prophylaxis in surgical procedures. Infectious Diseases Society of America. Clin Infect Dis 18:422, 1994

44. Kurz A, Sessler DI, Lenhardt R: Perioperative normothermia to reduce the incidence of surgical-wound infection and shorten hospitalization: study of wound infection and temperature group. N Engl J Med 334:1209, 1996

45. Sessler DI, Akca O: Nonpharmacological prevention of surgical wound infections. Clin Infect Dis 35:1397, 2002

46. Greif R, Akca O, Horn EP, et al: Supplemental perioperative oxygen to reduce the incidence of surgical-wound infection. Outcomes Research Group. N Engl J Med 342:161, 2000

47. Arkilic CF, Taguchi A, Sharma N, et al: Supplemental perioperative fluid administration increases tissue oxygen pressure. Surgery 133:49, 2003

48. Pomposelli JJ, Baxter JK 3rd, Babineau TJ, et al: Early postoperative glucose control predicts nosocomial infection rate in diabetic patients. J Parenter Enteral Nutr 22:77, 1998

49. van den Berghe G, Wouters P, Weekers F, et al: Intensive insulin therapy in the surgical intensive care unit. N Engl J Med 345:1359, 2001

50. Latham R, Lancaster AD, Covington JF, et al: The association of diabetes and glucose control with surgical-site infections among cardiothoracic surgery patients. Infect Control Hosp Epidemiol 22:607, 2001

51. Brook I, Frazier EH: Aerobic and anaerobic microbiology of retroperitoneal abscesses. Clin Infect Dis 26:938, 1998

52. Santaella RO, Fishman EK, Lipsett PA: Primary vs secondary iliopsoas abscess. Arch Surg 130:1309, 1995

53. Dembry LM, Andriole VT: Renal and perirenal abscesses. Infect Dis Clin North Am 11:663, 1997

54. Meng MV, Mario LA, McAninch JW: Current treatment and outcomes of perinephric abscesses. J Urol 168:1337, 2002

55. Bradley EL: A clinically based classification system for acute pancreatitis. Summary of the International Symposium on Acute Pancreatitis, Atlanta, Ga, September 11 through 13, 1992. Arch Surg 128:586, 1993

56. Dellinger P: Pancreatic infection. Source Control: A Guide to the Management of Surgical Infections. Schein M, Marshall JC, Eds. Springer, Berlin, Germany, 2002, p 154

57. Fernandez-del Castillo C, Rattner DW, Makary MA, et al: Debridement and closed packing for the treatment of necrotizing pancreatitis. Ann Surg 228:676, 1998

58. Howard TJ, Temple MB: Prophylactic antibiotics alter the bacteriology of infected necrosis in severe acute pancreatitis. J Am Coll Surg 195:759, 2002

59. Buchler MW, Gloor B, Muller CA, et al: Acute necrotizing pancreatitis: treatment strategy according to the status of infection. Ann Surg 232:619, 2000

60. Johannsen EC, Sifri CD, Madoff LC: Pyogenic liver abscesses. Infect Dis Clin North Am 14:547, 2000

61. Mohsen AH, Green ST, Read RC, et al: Liver abscess in adults: ten years experience in a UK centre. QJM 95:797, 2002

62. Ralls PW: Inflammatory disease of the liver. Clin Liver Dis 6:203, 2002

63. Rintoul R, O'Riordan MG, Laurenson IF, et al: Changing management of pyogenic liver abscess. Br J Surg 83:1215, 1996

64. Winkelstein JA, Marino MC, Johnston RB Jr, et al: Chronic granulomatous disease: report on a national registry of 368 patients. Medicine (Baltimore) 79:155, 2000

65. Lublin M, Bartlett DL, Danforth DN, et al: Hepatic abscess in patients with chronic granulomatous disease. Ann Surg 235:383, 2002

66. Chou F-F, Sheen-Chen SM, Chen YS: The comparison of clinical course and results of treatment between gas-forming and non-gas-forming pyogenic liver abscess. Arch Surg 130:401, 1995

67. Chou F-F, Sheen-Chen SM, Lee TY: Rupture of pyogenic liver abscess. Am J Gastroenterol 90:767, 1995

68. Hughes MA, Peri WA Jr: Amebic liver abscess. Infect Dis Clin North Am 14:565, 2000

69. Stanley SL Jr: Amoebiasis. Lancet 361:1025, 2003

70. Alvarez Perez JA, Gonzalez JJ, Baldonedo RF, et al: Clinical course, treatment, and multivariate analysis of risk factors for pyogenic liver abscess. Am J Surg 181:177, 2001

71. Yeh TS, Jan YY, Jen LB, et al: Hepatocellular carcinoma presenting as pyogenic liver abscess: characteristics, diagnosis, and management. Clin Infect Dis 26:1224, 1998

72. Ooi LL, Leong SS: Splenic abscesses from 1987 to 1995. Am J Surg 174:87, 1997

73. Thanos L, Dailiana T, Papaioannou G, et al: Percutaneous CT-guided drainage of splenic abscess. AJR Am J Roentgenol 179:629, 2002

74. Ng KK, Lee TY, Wan CF, et al: Splenic abscess: diagnosis and management. Hepatogastroenterology 49:567, 2002

XXII VAGINITIS AND SEXUALLY TRANSMITTED DISEASES

MATTHEW R. GOLDEN, M.D., M.P.H.

Sexually transmitted diseases (STDs) are among the most common causes of infectious illness in the world. The United States leads industrialized nations in the occurrence of STDs, with an estimated 12 million new cases annually, three million of them in teenagers.[1] In many developing nations, STDs (excluding HIV infection) are the second greatest cause of disability-adjusted years of life lost,[2] and highly prevalent bacterial and viral STDs may facilitate HIV transmission.[3]

Of the more than 30 sexually transmitted pathogens that are currently recognized, eight have been identified since 1980, and it seems likely that the full spectrum of STD remains undefined.[4] Antimicrobial resistance has made treatment of some well-established infections (e.g., gonorrhea) more difficult. Finally, decreasing age of menarche, declining median age of populations in developing countries, delayed marriage, increased global travel and trade, urbanization, migration, war and associated social upheaval, and the dissolution of socially restrictive political systems in the former Soviet Union and, to a lesser extent, China all ensure that STDs will remain a major and probably increasing health problem in coming decades.[5]

For the clinician, the increasing recognition of viral STDs and the emergence of screening for chlamydial infections as a population-based STD control strategy has heightened the importance of familiarity with the management of these common infections. This chapter presents general concepts in the epidemiology and approach to patients with STD and reviews important STD syndromes, including urethritis, vulvovaginitis, mucopurulent cervicitis, pelvic inflammatory disease, and genital ulcer disease. Finally, the approach to STD in men who have sex with men and to sexually transmitted enteric infections will be presented. Specific pathogens, including *Chlamydia*, *Neisseria gonorrhoeae,* herpes viruses, and *Treponema pallidum,* are discussed in other chapters. The Centers for Disease Control and Prevention (CDC) issues guidelines for STD/HIV testing and counseling, as well as STD treatment (http://www.cdc.gov/hiv/dhap.htm). Clinicians are advised to refer to these guidelines for updated recommendations.[6-8]

Epidemiology and Transmission Dynamics

The transmission of an STD through a population can be conceptualized mathematically with the formula $R_0 = \beta cD$, in which R_0 is the average number of secondary cases generated by each primary infection in a population (i.e., the reproductive number); β is the average probability of transmission with each sexual partnership, c is the average number of sexual partnerships formed per unit of time; and D is the mean duration of infection.[9] For diseases in which each case generates an average of one additional case, R_0 equals 1 and the prevalence remains stable; values less than 1 and greater than 1 are associated with a declining or rising prevalence, respectively.

Although each of the terms in this equation is complex, the simplification that the equation offers can explain a great deal about the distribution of different STDs in a population and provides a framework for conceptualizing STD epidemiology. For example, gonorrhea is thought to be efficiently transmitted (β = 0.5), but it has a relatively short duration of infection, especially in settings in which medical care and therapy are readily available.[9,10] Consequently, for the reproductive number to remain 1 or greater, c must be relatively high. Thus, infection tends to concentrate in a population of highly sexually active persons, sometimes referred to as a core group.[11] In part because young people tend to be more sexually active than older people, the incidence of gonorrhea, like that of chlamydial infection, is highest among teenagers and persons in their early 20s. (This is less true of men who have sex with men, in whom gonorrhea incidence is less concentrated in the young.) In contrast, herpes simplex virus type 2 (HSV-2) has a very long duration of infection, and R_0 may exceed 1 even in populations with very low rates of partnership change. As a result, the prevalence of genital herpes rises with age, peak incidence likely occurs at a somewhat older age than with *Chlamydia* infection or gonorrhea, and the infection is widely disseminated throughout the population.[12]

This simple model of STD transmission dynamics focuses on average behavior in a population and the host-parasite relationship as determinants of STD epidemiology. However, it neglects the critical role played by variance in sexual behavior and patterns of sexual mixing (i.e., sexual networks) in defining transmission dynamics. The prevalence of STD in a population is in part a function of the extent to which persons who are more sexually active mix primarily with one another (assortative mixing) versus mixing more randomly with others, including persons who are less sexually active.[13] The frequency of concurrent partnerships in a population also exerts a profound influence on STD prevalence; such partnerships allow infections to spread in two directions, connecting groups of people and facilitating rapid transmission of infection.[14,15]

In eliciting a sexual history, clinicians have traditionally focused on the patient's behavior, asking about the number of sexual partners the patient has had and about the use of condoms. In many cases, however, self-reported behavior is not associated with risk of STD; sexual network factors may be more important in defining risk. For example, virtually all studies of selective screening for chlamydial infection have found that self-reported behavior is an insensitive predictor of infection,[16-18] whereas demographic factors such as age, race, socioeconomic status, source of clinical care, and geography are strongly associated with a variety of STDs.[19-21] These factors, which reflect the organization of human society and dictate sexual mixing patterns, play a critical role in defining an individual's risk of infection.[22] STDs exist within a social context; therefore, clinicians should base their assessment of risk on their practice setting and the patient's social milieu. Persons whose behavior would suggest a low risk of STD can in fact be at elevated risk simply by virtue of their sexual network, a population that is often socially determined rather than individually chosen. This knowledge should temper any tendency to entertain stigmatizing stereotypes related to sexual behavior and STD.

STD Prevention

SEXUAL HISTORY AND COUNSELING

What is the best way to elicit a sexual history? Although relatively little research has been done on this question, some general principles can be articulated. In eliciting a sexual history, the clinician must balance the need to collect specific information with the desire to engage the patient in a conversation about sexual risk. Whenever possible, questions should be open ended, allowing the patient to define factors that may have placed him or her at risk for STD (e.g., "What are you doing now, or what have you done in the past, that you think may have put you at risk for a sexually transmitted disease?"). Subsequent questions may be more specific, but the questions should be clear, direct, and phrased nonjudgmentally (e.g., "Do you have sex with men, women, or both men and women?"). Typically, a sexual history should include questions about sexual orientation, the number of sexual partners, the use of condoms, any history of STD, and the sexual repertoire (oral, insertive or receptive anal, and vaginal sex). Persons with HIV or those at high risk for HIV should be asked if they know the HIV status of their sexual partners. Persons with HIV should be asked whether they have informed their partners of their own HIV status.

Clinicians should seek to integrate elicitation of the sexual history with STD prevention counseling. The CDC recommends a client-centered approach to counseling. This approach involves an effort to help patients assess the circumstances and behaviors that place them at risk for STD and then help them commit to a single, defined plan for reducing their risk. Risk-reduction plans should be specific rather than general. For example, a specific goal might be to carry condoms when going out on a date or to ask a specific partner about his or her HIV status, rather than the general goals of using condoms all the time or having safe sex all the time.[23] Client-centered counseling has been shown (in a randomized trial of heterosexual STD clinic patients) to reduce the risk of STDs.[6,24]

Table 1 Recommended STD Screening[8]

Population	*Screening Measures*
All men and women	Retest all patients diagnosed with gonorrhea or *Chlamydia* infection 10–18 wk after initial treatment
Women Sexually active, age ≤ 24 yr Age > 24 yr with a new sexual partner or multiple sexual partners	Annual chlamydial screening
Pregnant women	Test for *Chlamydia trachomatis*, *Neisseria gonorrhoea*, hepatitis B virus infection, and syphilis; offer HIV testing and counseling Test for bacterial vaginosis in women at high risk (e.g., those with history of preterm delivery) Perform Pap smear if none was performed in the past year
Men who have sex with men (MSM)*	Perform the following at least annually: Serologic testing for HIV and syphilis Rectal culture for gonorrhea and chlamydial infection Pharyngeal culture for gonorrhea Consider serologic testing for HSV-2, particularly in HIV-negative MSM Serologic testing for hepatitis A and B antibodies†

*Screening guidelines apply to both HIV-positive and HIV-negative MSM. HIV and hepatitis testing should be performed only for susceptible patients. Current data are insufficient for a recommendation of technologies other than culture to test for rectal gonorrhea or chlamydial infection or pharyngeal gonorrhea. Because herpes simplex virus type 2 (HSV-2) increases the risk of HIV acquisition, HSV-2 screening should be considered to aid in HIV risk assessment and counseling. More frequent screening should be considered for those at highest risk of STD.
†Vaccinate for hepatitis A and B if negative.

STD REPORTING AND SEXUAL PARTNER MANAGEMENT

By law, gonorrhea, syphilis, chancroid, lymphogranuloma venereum (LGV), donovanosis (granuloma inguinale), and, in most parts of the United States, chlamydial infections must be reported to local health departments. In general, health departments in the United States routinely attempt to ensure the treatment of sexual partners of persons with syphilis; they only sometimes attempt to contact persons reported to have HIV to offer them assistance in notifying their sexual and needle-sharing partners; and they seldom make any routine effort to notify the partners of persons with gonorrhea or chlamydial infection. Some health departments will provide such services if specifically asked to do so by a clinician or a person diagnosed with an STD. Although clinicians should make their patients aware that they may be contacted by public health authorities regarding partner notification, in most instances, it is the responsibility of the diagnosing clinician and the patient to ensure that sexual partners are evaluated and treated. Several recent studies have suggested that giving patients medication to give to their sexual partners is feasible and may reduce chlamydial reinfection rates.[25-27] However, at present, there are no guidelines that define the circumstances in which this approach to partner management should be employed.

STD SCREENING

Because STDs are often asymptomatic, screening is a critical component of prevention. Recommendations for screening vary according to population [*see Table 1*].[8] Data in support of STD screening are strongest for chlamydial infection; a randomized trial has demonstrated that chlamydial screening reduces the rate of PID.[28]

Urethritis in Men

EPIDEMIOLOGY

Urethritis is one of the most common STD syndromes in men, resulting in an estimated 200,000 initial physician visits in the United States in 2000.[28] The syndrome is typically divided into urethritis resulting from infection with *N. gonorrhoeae* and nongonococcal urethritis (NGU). Rates of gonococcal urethritis in most developed nations have declined dramatically over the past 20 years, although rates in the United States and Europe now appear to be rising again, particularly among men who have sex with men.[29,30]

ETIOLOGY AND MICROBIOLOGY

Since the mid-1970s, *Chlamydia trachomatis* has been recognized as the most common cause of NGU; *C. trachomatis* has typically been isolated in 30% to 40% of cases of NGU, although the prevalence of *C. trachomatis* in men with NGU may now be declining,[31] and chlamydial infection is less common in older

men with NGU than in younger ones. In areas of the United States where the prevalence of *C. trachomatis* has declined in recent years, most patients with symptomatic urethritis have no evidence of either gonorrhea or chlamydial infection.[32] Other established causes of NGU include *Trichomonas vaginalis*, HSV-2, and, in men who engage in insertive anal intercourse, enteric pathogens. Approximately one third of men with primary genital herpes have dysuria and a urethral discharge. *T. vaginalis* is a more common cause of NGU in older men. However, HSV and *Trichomonas* combined are probably responsible for fewer than 10% of all cases of NGU. *Ureaplasma urealyticum* and *Mycoplasma genitalium* have been associated with NGU in case-control studies, but at present no tests for these organisms are commercially available.[33-35]

DIAGNOSIS

Clinical Manifestations

Clinical manifestations of urethritis include urethral discharge, dysuria, and itching at the distal urethra. Inguinal adenopathy is unusual. Likewise, fever, chills, perineal pain, scrotal mass, genital pain, and other urinary symptoms (e.g., hematuria, frequency, hesitancy, nocturia, or urgency) are unusual and should prompt consideration of alternative diagnoses, such as urinary tract infection (UTI), epididymitis, orchitis, or prostatitis. Although gonorrhea is generally associated with a more abrupt onset of symptoms and a more copious and purulent discharge than NGU, these distinctions are not reliable.

Asymptomatic and subclinical gonoccocal and chlamydial urethral infections probably play an important role in sustaining endemic levels of these STDs, but their incidence is uncertain. A prospective study of gonococcal urethritis found that only 2% of infections remained asymptomatic in the 14 days after acquisition.[36] However, cross-sectional studies have demonstrated that asymptomatic infection is common among the sexual partners of infected women[37] and, at least in some parts of the United States that have a high prevalence of gonorrhea, in the general population of young adults.[38] Prospective data on the frequency of asymptomatic chlamydial urethritis are not available, but as with gonorrhea, cross-sectional studies have demonstrated that asymptomatic or subclinical chlamydial urethritis is common.[38,39]

Physical Examination

Objective evidence of urethral inflammation should be sought in men presenting with dysuria or urethral discharge. Physical examination should include a genital examination, preferably conducted several hours after the patient last urinated; the examination should include a search for purulent or mucopurulent discharge. If no discharge is observed, the examiner should strip the urethra from the base of the penis to the urethral meatus to elicit a discharge.

Laboratory Tests

A urethral Gram stain should be performed on all men with symptoms of urethritis, even those with no discharge evident on physical examination. Urethral specimens for Gram stain are obtained by inserting a thin calcium-alginate–tipped swab 3 to 4 cm into the urethra, then rolling the swab over a glass slide. A diagnosis of urethritis is established by the presence of five or more polymorphonuclear neutrophils (PMNs) per 1,000× oil-immersion field. Alternatively, the diagnosis can be made through use of a centrifuged 10 to 15 ml first-void urine specimen; the diagnosis is established by the finding of 10 or more PMNs per 400× field on at least one of five randomly selected fields. A positive urine leukocyte esterase test is also sufficient for establishing the diagnosis. Among experienced microscopists, the finding of gram-negative intracellular diplococci (GND) is 90% to 95% sensitive and over 95% specific in detecting urethral gonorrhea in symptomatic men,[40,41] and the presence of GND establishes the diagnosis of gonorrhea [*see Figure 1*]. The Gram stain should be considered equivocal if only extracellular organisms are seen. Regardless of Gram stain findings, specific microbiologic testing should be performed for *N. gonorrhoeae* and *C. trachomatis*.

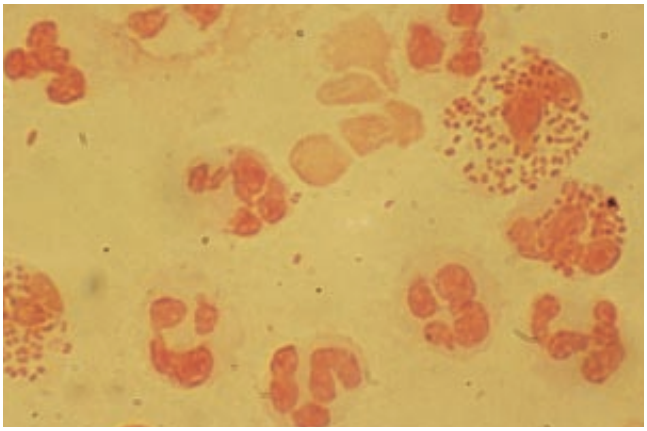


Figure 1 **This figure shows a Gram stain of a urethral discharge from a man with gonorrhea. The gram-negative intracellular diplococci are *Neisseria gonorrhoeae* organisms.**

Because it provides data on antimicrobial susceptibility, culture remains the preferred microbiologic test for gonorrhea. Nonamplified DNA probes offer the advantage of simplified specimen handling and the ability to perform both gonococcal and chlamydial testing on a single specimen. These tests have a specificity of approximately 99%, and their sensitivity is comparable to that of culture for gonorrhea.[42]

Nucleic acid amplification tests (NAATs) of urethral specimens or urine (e.g., ligase chain reaction [LCR], polymerase chain reaction [PCR], transcription-mediated amplification [TMA], and strand displacement amplification [SDA]) have sensitivities comparable or superior to that of culture; although typically more costly, these assays offer the advantage of testing without urethral swabs. In low-prevalence populations, positive NAAT results may require confirmatory testing. The sensitivities of tests for *C. trachomatis* vary widely. Although test performance varies, depending on organism burden and anatomic site of collection, culture has a sensitivity of between 50% and 90%; enzyme immunoassays and nonamplified genetic probes have a sensitivity of 40% to 75%; and NAATs have a sensitivity of more than 90%.[43,44] For that reason, NAATs are preferred, when available.

TREATMENT

Initial Management

Patients with evidence of gonococcal infection on urethral Gram stain should be treated for gonorrhea. Recommended regimens include single doses of the following agents: (1) cefixime, 400 mg orally; (2) ceftriaxone, 125 mg intramuscularly; (3) ciprofloxacin, 500 mg orally; (4) ofloxacin, 400 mg orally; and (5)

levofloxacin, 250 mg orally. Quinolone-resistant *N. gonorrhoeae* has recently emerged as a problem in Asia, the Pacific Islands, and, most recently, California. Consequently, quinolones are no longer recommended for the empirical treatment of gonorrhea in persons in these areas or in their contacts. Because of the high chlamydial coinfection rate, all patients with gonorrhea should also be treated for *Chlamydia,* unless that diagnosis has been microbiologically excluded. Treatment for presumptive chlamydial infection in men with NGU is with azithromycin in a single 1 g oral dose or doxycycline, 100 mg orally twice a day for 7 days.

Treatment of Recurrent or Persistent Urethritis

Although recognition of the pathogenic role of *C. trachomatis* has reduced a major cause of persistent or recurrent urethral symptoms after treatment for gonococcal urethritis, such symptoms continue to affect a minority of patients. Management should include questions regarding adherence to medical therapy and partner treatment, a urethral Gram stain to document evidence of urethral inflammation, and repeat testing for gonorrhea and chlamydial infection. Consideration should be given to possible trichomonal or herpes infection. Erythromycin, 500 mg four times a day for 7 days, with or without a single 2 g dose of metronidazole, is the recommended empirical treatment in patients who are believed to have adhered to their initial regimen and who have not been reexposed to gonorrhea or chlamydial infection.[8]

Lower Genital Tract Infections in Women

Women with STDs involving the lower genital tract may present with dysuria, urethritis or vulvovaginitis, and abnormal or altered vaginal discharge. The initial evaluation of women with these complaints seeks to differentiate urethritis, cystitis, vulvovaginitis, and cervicitis and to identify women with upper genitourinary tract infections (e.g., pyelonephritis or salpingitis). Subsequent microbiologic testing and treatment are guided by this evaluation.

Syndromes Causing Dysuria and Urethritis

STDs that can cause dysuria in women include vulvitis resulting from candidal infection and genital herpes and urethritis caused by *C. trachomatis* or *N. gonorrhoeae*. Dysuria and sterile pyuria (the presence of leukocytes and the absence of more than 10^2 organisms/ml of conventional urinary pathogens in a midstream urine specimen) in a woman are consistent with a diagnosis of urethral infection.[45] Other factors suggesting urethral infection include the absence of other symptoms and signs typical of UTI; risk factors or risk markers for chlamydial infection (young age, new or multiple sexual partners, failure to consistently use condoms, African-American race); symptoms lasting 7 days or longer; purulent vaginal discharge; pelvic pain or tenderness; and evidence of mucopurulent cervicitis.

Women presenting with a syndrome of dysuria and sterile pyuria should be tested for gonorrhea and chlamydial infection. Because HSV-2 can cause urethritis in women, particularly in women with primary HSV infection, the possibility of genital herpes should also be considered. HSV or candidal vulvitis typically causes external, as opposed to internal, dysuria, which occurs when urine comes in contact with the introitus or labia. Women with these infections typically have vulvar irritation or lesions; vaginal discharge; or a history of either HSV or candidal vaginitis.

In the differential diagnosis of dysuria in women, particular attention should be given to bacterial UTI, which is the most common cause of dysuria. Symptoms, signs, and laboratory findings that support the diagnosis of bacterial cystitis include urinary frequency or urgency, a history of UTI, duration of symptoms of less than 4 days, gross or microscopic hematuria, the patient's belief that she has a UTI, suprapubic tenderness, a positive urine nitrite test, and evidence of typical urinary tract pathogens on Gram stain or urine culture.[46-48] Fever or flank pain in a woman with dysuria and other findings consistent with UTI suggests pyelonephritis. Vaginal discharge or irritation is not typical of UTI.

Diagnostic testing for gonococcal and chlamydial urethritis in women should be based on specific tests [*see* Laboratory Tests, *above*]. In women with no evidence of pelvic inflammatory disease (PID), treatment is identical to that for men [*see* Treatment, *above*].

Syndromes Causing Vulvovaginitis and Vaginal Discharge

Abnormal vaginal discharge is one of the most common reasons for women to seek medical attention. Since the 1960s, the number of women receiving care for vulvovaginal infections increased approximately threefold. In 2000, an estimated three million initial physician office visits in the United States were prompted by vulvovaginal infections.[28]

The most common causes of an abnormal vaginal discharge are bacterial vaginosis (BV), vulvovaginal candidiasis (VVC), and trichomonal vaginitis (TV) [*see Table 2*]. Both BV and trichomoniasis have been associated with preterm labor.[49,50] However, to date, treatment of these infections has not definitively been shown to decrease preterm delivery.[51-54] BV has been identified as a risk factor for PID, and both BV and trichomoniasis may increase the risk of HIV acquisition and transmission.[55-59] Consequently, these diagnoses have assumed new importance in HIV prevention.

Evaluation of women with vaginal complaints should include a pelvic examination and a directed laboratory evaluation. Although these infections tend to have different clinical features, a study of patients triaged and selectively treated after a telephone assessment found poor agreement between the diagnosis made by nurses and other providers and the diagnoses obtained after examination and testing.[60] Similarly, a study of women purchasing over-the-counter antifungal therapies found that only one third had vaginal candidiasis, and 53% had a diagnosis other than vulvovaginal candidiasis.[61] These findings emphasize the need for a complete evaluation in women complaining of vaginal discharge or discomfort. Less frequent causes of vaginitis include atrophic vaginitis with secondary bacterial infection, vaginitis associated with foreign bodies or toxins, *Staphylococcus aureus* vaginitis associated with toxic-shock syndrome, group A *Streptococcus*–associated vaginitis, desquamative vaginitis (clindamycin responsive), erosive lichen planus, allergic vaginitis, vaginitis associated with autoimmune disease, and idiopathic vaginitis.[62]

Bacterial Vaginosis

BV is the most common cause of vaginal discharge in women of reproductive age. Prevalence studies have found BV in 10% to 40% of women tested, with higher rates of infection in women tested in STD clinics and in African Americans. Douching and use of intrauterine devices (IUDs) have also been associated with BV.[63]

Table 2 Clinical Features and Management of Vulvovaginitis

Feature	*Normal Vaginal Examination*	*Vulvovaginal Candidiasis*	*Trichomonal Vaginitis*	*Bacterial Vaginosis*
Etiology	Uninfected; lactobacilli predominate	*Candida albicans* most common; candidiasis caused by species other than C. *albicans* may be increasing	*Trichomonas vaginalis*	Loss of normal vaginal lactobacilli; associated with *Gardnerella vaginalis*; increased anaerobic bacteria and mycoplasmas
Symptoms	None	Abnormal vaginal discharge, external dysuria, vulvar itching, pain and/or irritation	Yellow vaginal discharge, external dysuria, vulvar itching	Increased, abnormal, or malodorous vaginal discharge
Discharge Amount Color Consistency	 Variable Clear or white Nonhomogeneous, patchy (floccular)	 Scant White Clumped; adherent plaques	 Profuse Yellow Homogeneous or frothy	 Moderate White or gray Adherent, homogeneous discharge that uniformly coats vagina
Inflammatory findings	None	Vulvar erythema, edema, or fissure; erythema of vaginal epithelium; introitus	Erythema of vaginal and vulvar epithelium; colpitis macularis	None
pH of vaginal fluid*	Usually ≤ 4.5	Usually ≤ 4.5	Usually > 4.5	Usually > 4.5
Amine (fishy) odor with 10% KOH	None	None	May be present	May be present
Microscopy	Normal epithelial cells; lactobacilli predominate	Leukocytes, epithelial cells; mycelia or pseudomycelia† (50%–85% of cases)	Leukocytes; trichomonads seen in 50%–70% of culture-positive cases	Clue cells (81%–94% of cases); few leukocytes; lactobacilli outnumbered by mixed flora
Recommended treatment	—	Intravaginal imidazole (butoconazole, clotrimazole, miconazole, terconazole, tioconazole) for 3–7 days; fluconazole, 150 mg p.o. (single dose)	Metronidazole, 2 g p.o. (single dose); metronidazole, 500 mg p.o., b.i.d., for 7 days	Metronidazole, 500 mg p.o., b.i.d., for 7 days; metronidazole gel, 0.75%, 5 g intravaginally each night for 5 nights; clindamycin cream 2%, 5 g intravaginally each night for 7 days
Sexual partner treatment	—	None if asymptomatic; topical treatment if candidal dermatitis of the penis or balanitis is detected	Metronidazole, 2 g orally (single dose)	None

*pH determination is not useful if blood is present.

†To detect fungal elements, vaginal fluid is digested with 10% KOH before microscopic examination; to examine for other features, fluid is mixed (1:1) with normal saline. Culture may be necessary if microscopy results are negative and the suspicion of *Candida* is high.

Pathophysiology and transmission The etiology of BV is unknown. The syndrome constitutes a disturbance in normal vaginal bacterial flora characterized by a reduction in the concentration of hydrogen peroxide–producing lactobacilli[64] and increased growth of mixed bacterial flora that include *Gardnerella vaginalis,* anaerobes, and *Mycoplasma hominis*. There is evidence that BV can be transmitted sexually. This evidence includes the following: studies have demonstrated that the inoculation of women with vaginal fluid from another woman with BV can induce the syndrome[65,66]; there is a high prevalence of BV in patients being treated at STD clinics; there are high rates of concordant BV among lesbian sexual partners[67]; longitudinal studies have associated BV with having higher numbers of sexual partners and with having new sexual partners[68-70]; and most studies have found that BV is absent in virgins.[71] Evidence against sexual transmission includes the lack of benefit from treating sexual partners[72-74] and inconsistent associations with levels of sexual activity.

Diagnosis Physical examination of women with BV typically reveals a homogeneous, white, uniformly adherent vaginal discharge.[63] The Amsel criteria for diagnosis of BV include the following: (1) presence of a homogeneous, thin vaginal discharge; (2) vaginal pH greater than 4.5; (3) clue cells (bacteria attached to vaginal epithelial cells on wet mount); and (4) presence of an amine (fishy) odor when vaginal fluid is mixed with 10% potassium hydroxide (KOH).[71,75-77] The presence of three of the four criteria establishes the diagnosis [*see Table 3*].

Treatment BV is treated with metronidazole. A meta-analysis found higher cure rates with a dosage of 1 g a day for 7 days than with a single 2 g dose (82% versus 73%).[63] Intravaginal metronidazole and intravaginal clindamycin offer efficacy comparable to 7-day courses of metronidazole, with fewer side effects, but are not effective in the treatment of trichomoniasis and are typically more costly. Recurrence of BV is common, occurring in 50% to 70% of cases. Multiple randomized trials have failed to demonstrate any benefit from treating male partners.[72-74]

Trichomoniasis

T. vaginalis is a sexually transmitted protozoan. In the United States, the number of women seeking care for trichomonal vaginitis declined by over 50% from 1966 to the mid-1980s; in 2000, physicians in the United States saw an estimated 200,000 patients with TV.[28] A cross-sectional study of 13,816 pregnant women in the United States found TV in 13%; the vast majority of those infections were subclinical or asymptomatic. Risk fac-

Table 3 Amsel Criteria for the Diagnosis of Bacterial Vaginosis[71,75,77]

Criterion	*Sensitivity (%)*	*Specificity (%)*
Homogeneous, thin vaginal discharge	52–65	71–97
Vaginal pH > 4.5	92–97	53–62
Clue cells on vaginal wet mount	81–94	94–98
Amine odor when vaginal fluid is mixed with 10% potassium hydroxide (KOH)	43–84	98—99

Note: the presence of three of these four criteria establishes the diagnosis of bacterial vaginosis.

tors for TV included African-American ethnicity, cigarette smoking, unmarried status, and lower educational level.[77] Untreated infections in women are thought to persist for a median of 3 to 5 years.[78]

Diagnosis Clinical manifestations of trichomonal infection include yellow vaginal discharge and vulvar itching. Neither is highly sensitive or specific. On physical examination, signs associated with *Trichomonas* infection include frothy or purulent vaginal discharge, which is sometimes profuse; vulvar or vaginal erythema; and cervical mucopus.[79,80] All of these signs have far greater specificity than sensitivity. The finding of colpitis macularis—punctate cervical hemorrhages and ulcers, sometimes referred to as strawberry cervix—has a specificity of 99% for TV but is seen in fewer than 5% of patients on unaided physical examination; colpitis macularis is much more readily visible on colposcopy.[79] In expert hands, a finding of motile *Trichomonas* on wet-mount examination has a sensitivity of 50% to 70%, although in clinical practice, wet-mount examination is usually considerably less sensitive. Culture on Diamond medium is the traditional diagnostic gold standard, but this technique is not available in most practice settings. Recently, InPouch, a relatively simple and inexpensive culture method, became available. The sensitivity of InPouch is comparable to that of Diamond medium and superior to that of wet mount.[81] PCR has been successfully used in research settings, but no NAAT is commercially available at present. Antigen detection tests are also under investigation.

Treatment A single 2 g dose of metronidazole is the treatment of choice for TV. Reported cure rates are 82% to 88%.[80] Sexual partners should be treated concurrently, and couples should be advised to abstain from sex for 1 week after treatment. Topical metronidazole is not effective.[82] Resistance to metronidazole occurs infrequently, and most cases respond to prolonged courses of metronidazole therapy. Some authors have reported successful treatment of metronidazole-resistant cases using either tinidazole or paromomycin cream.[83]

Vulvovaginal Candidiasis

Because VVC is not a reportable infection, only limited epidemiologic data are available. In the United States, a study of female university students found that over half experienced at least one episode of VVC by 25 years of age,[84] and 6.5% of women who participated in a national random-digit–dialing survey reported that a health care provider had told them they had candidal vaginitis at least once in the preceding 2 months.[85] Higher rates of VVC have been observed in African Americans and in users of oral contraceptives, vaginal sponges, or IUDs.[86] Although VVC is not clearly identified as an STD, it has been associated with the onset of sexual activity in young women and with cunnilingus.[84,86] Other predisposing factors include recent use of antibiotics; diabetes mellitus; pregnancy; and immunodeficiency, including that from HIV infection.

Diagnosis Vulvovaginal pruritus is generally the most common symptom of VVC.[87] Other findings sometimes associated with VVC include a cottage-cheese–like discharge; external dysuria; external genital burning or pain; perineal edema or erythema; and vulvar erythema, edema, and fissures.[87,88] However, several studies have reported the absence of any signs or symptoms significantly associated with VVC.[4,89] As a result, the diagnosis requires microscopic and, at times, microbiologic assessment. A 10% KOH preparation of fluid taken from the vagina has a sensitivity of 50% to 85% in the diagnosis of VVC[90,91]; if this test is negative but the clinical picture is consistent with VVC and there is no alternative diagnosis, culture for yeast should be performed.

Treatment Topical azoles (e.g., butoconazole, clotrimazole, miconazole, econazole, tioconazole, and terconazole) are 80% to 90% effective in treating VCC [*see Table 2*]. Most of these agents are available over the counter. No clear advantage favors one azole over another. Oral azoles (fluconazole or itraconazole) are comparably or slightly more effective and may be more convenient, but these agents also pose a small risk of systemic reactions. Because there are no compelling data favoring any one agent or route of administration, patient preference should guide the choice of treatment. Immunosuppressed patients and those with candidal infections caused by a species other than *Candida albicans* may require more prolonged therapy (e.g., 14 days).

Long-term therapy is indicated for patients with recurrent VVC, which is defined as four or more episodes of VCC in a year. Approximately 5% of women with VVC experience recurrences. Treatment may require 14 days of induction therapy followed by once-weekly maintenance therapy. Patients with *C. glabrata* VVC who do not respond to prolonged courses of azole therapy may benefit from topical boric acid (600 mg once a day for 2 weeks) or topical flucytosine.[92]

Mucopurulent Cervicitis

Mucopurulent cervicitis (MPC) is an inflammatory process affecting the columnar epithelium and subepithelium of the endocervix and adjacent exocervix. As with NGU in men, MPC is common and has most frequently been associated with *N. gonorrhoeae* or *C. trachomatis* and, less frequently, with HSV or *T. vaginalis*. Unlike NGU, MPC typically produces no symptoms; or it may produce nonspecific symptoms, such as a yellow vaginal discharge, that often do not prompt women to seek treatment. In recent years, as the prevalence of gonorrhea and chlamydial infections have decreased in some settings, MPC with no defined microbiologic etiology has come to constitute the majority of cases.[4] MPC is important because of its association with known infections and because patients with MPC have an elevated risk of PID and adverse pregnancy outcome.

Diagnosis Different diagnostic criteria have been used for MPC. According to current CDC guidelines, the diagnosis of

MPC is made on the basis of a finding of a visible purulent or mucopurulent exudate on cervical examination or on endocervical swab. The finding of cervical mucopus is 28% to 52% sensitive and 82% to 94% specific for the presence of either *C. trachomatis* or *N. gonorrhoeae.*[88,93,94] Some investigators use additional criteria for MPC, including a finding of from 20 to 30 PMNs per high-power field (hpf) on cervical Gram stain or easily induced cervical bleeding.[4,94] These factors have been associated with the likelihood of *C. trachomatis* or *N. gonorrhoeae* infection, but they have not consistently been included as diagnostic criteria of MPC; with regard to the use of cervical Gram stain, these findings have not consistently been useful in defining a population in need of empirical therapy.

Treatment The decision to treat MPC is based largely on the local prevalence of *C. trachomatis* or *N. gonorrhoeae* and on the patient's risk. In areas where both gonorrhea and chlamydial infection are common, empirical therapy should be directed at both pathogens. In areas where gonorrhea rates are low, treating for *Chlamydia* infection alone is reasonable. Recent evidence suggests that in areas where the prevalence of both infections is low, older patients (i.e., those older than 30 years) suspected of having MPC need not be treated until microbiologic test results are available, provided follow-up care is ensured.[94]

Pelvic Inflammatory Disease

PID is an inflammatory process involving a variable combination of endometritis, salpingitis, tubo-ovarian abscess, and pelvic peritonitis. PID can be blood-borne (e.g., tuberculosis) or result from extension of an intra-abdominal process. At present, however, PID most often develops when bacteria ascend from the vagina or cervix into the endometrium, fallopian tubes, and pelvic peritoneum. Although the number of women seeking care for PID has declined by over 25% since the 1980s, 8% of participants in the 1995 National Survey of Family Growth, a national representative sample of United States women, reported a history of PID.[95] Identified risk factors for PID include a previous history of PID, higher numbers of lifetime sex partners, douching, and a history of bacterial STD. In the past, IUD use was identified as a risk factor, but its importance beyond the first 30 days after insertion is now controversial; a recent case-control study found no association between the use of currently available copper IUDs and the occurrence of PID.[96] Gynecologic procedures that disrupt the protective cervical barrier (e.g., pregnancy termination, IUD insertion, dilatation and curettage, and hysterosalpingography) elevate the risk of PID and may lead to PID in the absence of classic sexually transmitted pathogens.

MICROBIOLOGY

Studies of PID conducted in the United States and Europe in the 1980s typically implicated *C. trachomatis, N. gonorrhoeae,* or both as a cause of PID in approximately half of cases.[97] Frequently, these bacteria were part of a polymicrobial infection involving diverse normal vaginal flora, including anaerobic bacteria, facultative anaerobes, and genital mycoplasmas. *M. genitalium* has been associated with endometritis and PID.[98] *Actinomyces israelii* is a cause of PID in women with IUDs.

DIAGNOSIS

The diagnosis of PID is difficult. To date, studies have been unable to identify any single clinical finding or constellation of findings that allow accurate identification of women with PID.[99,100] Moreover, PID studies have typically enrolled only women with overt disease and, consequently, have not provided an accurate picture of the full spectrum of the clinical entity. Indeed, most cases of PID probably go undiagnosed. Approximately two thirds of women with postinfectious fallopian tube occlusion report no history of PID, although many have sought care for abdominal pain.[101] When the diagnosis is made clinically, it may not be supported by surgical findings. Only 60% to 70% of women with clinically diagnosed PID typically have laparoscopic evidence of PID.[102]

In clinically detected cases, the cardinal symptom of PID is pelvic or abdominal pain. The pain is typically dull or aching. Onset can be acute or subacute and frequently occurs at the beginning of menses. Typically, patients present after having symptoms for less than 2 weeks. In a Swedish study of 623 patients with PID, all had pelvic or abdominal pain, cervical motion tenderness, and increased inflammatory cells in vaginal or cervical secretions. Other symptoms and laboratory findings included an erythrocyte sedimentation rate (ESR) of 15 mm/hr or higher (75%), leukocytosis greater than 10,000/ml (60%), abnormal vaginal discharge (55%), fever higher than 38° C (100.4° F) (41%), abnormal vaginal bleeding (36%), dysuria (19%), vomiting (10%), and anorectal symptoms (anorectal pain, tenesmus, or rectal bleeding or discharge) (7%).[103] A large study showed that only temperatures higher than 38° C (100.4° F) had a specificity of more than 90% for the diagnosis of PID, although its sensitivity was only 11%.[103]

The differential diagnosis of PID includes other causes of abdominal or pelvic pain. Depending on the clinical circumstances, the physician may need to consider such disorders as appendicitis, endometriosis, bleeding corpus luteum, pelvic adhesions, gastroenteritis, and ectopic pregnancy.

Although laparoscopy has been the traditional gold standard for diagnosing PID, many women with abnormal fimbrial biopsies have normal results on laparoscopy. Moreover, some women have histologic evidence of endometritis without salpingitis,[104] which suggests that laparoscopy may be insensitive for the detection of milder cases or of PID that is restricted to the uterus.

Transvaginal ultrasound (TVUS) should be performed when symptoms are severe, when the physical examination reveals a pelvic mass, or when the diagnosis of PID is uncertain. Studies assessing the performance of different imaging modalities in the diagnosis of PID have been small, with no single study enrolling more than 50 patients with the diagnosis.[102] Small studies of TVUS have reported sensitivities of 81% to 93%, but specificities have been highly variable, ranging from 5% to 100%, with the test performing best in patients with more severe infection.[102] A case-control study of power Doppler TVUS reported a sensitivity of 100% and a specificity of 80%,[105] suggesting it may offer advantages over conventional TVUS. In women with tubo-ovarian abscess, repeat TVUS is often indicated to assess response to therapy. Small studies of CT and pelvic MRI have also reported high sensitivity and specificity. Laparoscopy should be performed if appendicitis, ectopic pregnancy, or ruptured abscess is suspected; laparoscopy should also be considered in women who do not respond to antibiotics.

TREATMENT

Because the diagnosis of PID can be challenging, the sequelae of PID can be severe, and treatment is safe and inexpensive, all

Table 4 Treatment Regimens for Pelvic Inflammatory Disease

Route	*Regimen*
Parenteral	Cefotetan, 2 g I.V. q. 12 hr *or* Cefoxitin, 2 g I.V. q. 6 hr *plus* Doxycycline, 100 mg p.o. or I.V. q. 12 hr
	Clindamycin, 900 mg I.V. q. 8 hr *plus* Gentamicin, 2 mg/kg I.V. or I.M. once, then 1.5 mg/kg I.V. q. 8 hr (single daily dose may be used)
	Ofloxacin, 400 mg I.V. q. 12 hr *or* Levofloxacin, 500 mg I.V. q.d. *plus* Doxycycline, 100 mg p.o. or I.V. q. 12 hr *with or without* Metronidazole, 500 mg I.V. q. 8 hr *or* Ampicillin-sulbactam, 3 g I.V. q. 6 hr
Oral	Ofloxacin, 400 mg p.o., b.i.d. *or* Levofloxacin, 500 mg I.V. q.d. for 14 days *with or without* Metronidazole, 500 mg p.o., b.i.d., for 14 days
	Ceftriaxone, 250 mg I.M. once *or* Cefoxitin, 2 g I.M. once with probenecid, 1 g p.o. *plus* Doxycycline, 100 mg b.i.d. for 14 days *with or without* Metronidazole, 500 mg p.o., b.i.d., for 14 days

*Parenteral therapy can be discontinued after the patient improves clinically, but doxycycline should be continued for 14 days.

patients suspected of having PID should undergo treatment for PID. The CDC recommends initiating treatment of PID in all sexually active young women with adenexal tenderness or cervical motion tenderness.[8] These criteria are likely to be sensitive, but they are also quite nonspecific.[100]

Treatment for PID is directed against *C. trachomatis*, *N. gonorrhoeae*, gram-negative facultative anaerobes, vaginal anaerobes, and streptococci. Numerous regimens have been found acceptable [*see Table 4*]. A recent randomized trial in women with mild to moderate PID found no advantage of inpatient therapy with intravenous cefoxitin and doxycycline over outpatient therapy with a single intramuscular dose of cefoxitin and probenecid followed by oral doxycycline.[106] Indications for hospitalization include the following: (1) inability to exclude a possible surgical emergency (e.g., appendicitis), (2) pregnancy, (3) failure to respond to oral antibiotics, (4) inability to tolerate or adhere to outpatient oral therapy, (5) tubo-ovarian abscess, and (6) inability to reliably ensure follow-up. Patients should show significant improvement within 3 days after starting therapy. Those receiving oral therapy should be reevaluated within 72 hours. Treatment should include efforts to ensure that sexual partners also receive therapy. In addition, patients with *Chlamydia* or *N. gonorrhoeae* infections should be rescreened for those infections 10 to 18 weeks after treatment.

COMPLICATIONS

Although the vast majority of women with PID in developed nations recover fully, long-term sequelae are common; these sequelae include tubal infertility, ectopic pregnancy, and chronic pelvic pain. In the largest study of PID sequelae performed to date, Swedish investigators performed laparoscopy on 1,730 women with suspected PID and then followed them for a mean of 6.9 years. After a single episode of PID, 8% of patients suffered tubal infertility, compared with 1% of control subjects in whom there was no laparoscopic evidence of PID. Of PID patients who subsequently became pregnant, 10% had an ectopic pregnancy, compared with 1% of women without PID. Similarly, pelvic pain lasting longer than 6 months occurred in 17% of women with PID but in only 2% of control subjects.[107] Recurrent episodes of PID multiplied the risk of sequelae [*see Figure 2*], as did more severe PID and longer duration of symptoms before treatment.

Genital Ulcer Disease

Genital ulcers are a frequent presentation of STDs. Epidemiologic studies, as well as studies measuring HIV shedding, suggest that genital ulcer disease (GUD) increases the risk of both HIV acquisition and HIV transmission.[3,108] As a result, the prevention and treatment of GUD is a high public health priority.

ETIOLOGY

Herpes, syphilis, and chancroid are the major causes of genital ulcer disease. Less common causes of GUD include lymphogranuloma venereum (infection with L-serotypes of *C. trachomatis*), donovanosis (infection with *Calymmatobacterium granulomatis*), superinfection of ectoparasitic infections, trauma, neoplasm, Behçet syndrome, Reiter syndrome, and fixed drug eruptions (e.g., from doxycycline or sulfonamides).

Herpes is the most common cause of GUD in developed nations. In the United States in 2000, over two million people sought care for genital herpes. In contrast, a total of 5,979 cases of primary and secondary syphilis and 82 cases of chancroid were reported to the CDC.[28] A 1996 study of 516 STD clinic patients

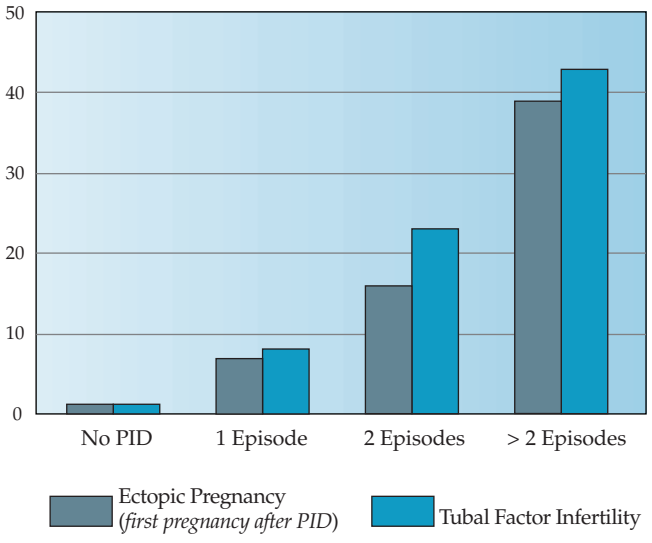


Figure 2 **Proportion of women experiencing an ectopic pregnancy or tubal infertility by number of episodes of pelvic inflammatory disease (PID) among 1,282 patients with PID and 448 control subjects.[97,107]**

Table 5 Clinical Features and Laboratory Diagnosis of Genital Ulcers

Disease	*Etiology*	*Incubation Period*	*Number of Lesions*	*Primary Lesion Type*	*Ulcer Diameter*	*Ulcer Characteristics*	*Pain or Tenderness*	*Lymphadenopathy*	*Laboratory Diagnosis*
Syphilis	*Treponema pallidum*	9–90 days	Usually one	Papule	5–15 mm	Superficial or deep; sharply demarcated; indurated; nonvascular, purulent base	Uncommon	Firm, nontender, bilateral	Darkfield microscopy, RPR/VDRL and FTA, MHA-TP
Herpes	HSV-1 or -2	2–7 days	Multiple, may coalesce	Vesicle	1–2 mm	Superficial; erythematous edges, no induration	Frequently tender	Firm, tender, small; often bilateral with first episode	DFA, culture, serology
Chancroid	*Haemophilus ducreyi*	1–14 days	Multiple, may coalesce	Pustule	Variable	Deep; irregular, undermined edges; purulent base bleeds easily	Usually tender	Tender, may be fluctuant, loculated; usually unilateral	Culture of ulcer base,* NAAT (e.g., PCR, LCR, TMA, SDA)
LGV	L-serotypes of *Chlamydia trachomatis*	3 days to 6 wk	Usually one	Papule, pustule, or vesicle	2–10 cm	Very rarely seen, because of rapid healing; can be superficial, deep, elevated, round, or oval	Variable	Tender, may suppurate or form sinus tracts; loculated, usually unilateral; more common in men than women	Culture, PCR, microimmunofluorescent antibody
Donovanosis	*Calymmatobacterium granulomatis*	1–4 wk (up to 6 mo)	Variable	Papule	Variable	Extensive, indolent ulcer with granulation tissue; elevated, rolled irregular edges on raised ulcer; beefy-red vascular base bleeds easily†	Uncommon	None; pseudobuboes	Giemsa or Wright stain of tissue smear

*Culture of material from bubo seldom positive.

†Less common variants can be hypertrophic, necrotic, or sclerotic.

DFA—direct fluorescent antibody HSV—herpes simplex virus LGV—lymphogranuloma venereum LCR—ligase chain reaction MHA-TP—microhemagglutination assay–*T. pallidum* NAAT—nucleic acid amplification test PCR—polymerase chain reaction RPR—rapid plasma reagin SDA—strand displacement amplification TMA—transcription-mediated amplification VDRL—Venereal Disease Research Laboratory

with genital ulcers found that 62% had HSV, 10% had syphilis, 3% had both syphilis and herpes, 3% had chancroid, and 22% had no identified pathogen.[109]

Traditionally, chancroid and syphilis have been the most common cause of genital ulcers in most developing nations. However, recent studies undertaken in sub-Saharan Africa have documented the increasing importance of herpes as a cause of GUD, particularly in areas where HIV is highly prevalent.[110]

DIAGNOSIS

Clinical Manifestations

When examining patients with genital ulcers, clinicians should note the number and depth of lesions; the presence of vesicles, induration, necrotic material on the ulcer bed, or an undermined ulcer border (i.e., the ulcer invades beneath the superficial edges); the presence or absence of pain; and any associated adenopathy [*see Table 5*]. Although physical findings can be helpful, different GUD etiologies cannot be reliably distinguished by physical examination alone.[111]

Laboratory Tests

Because physical findings are unreliable, clinical assessment should be supported by laboratory evaluation. The laboratory evaluation of GUD typically concentrates on herpes and syphilis. Chancroid, donovanosis, or LGV should be considered if the patient lives in or has traveled to an area where one of those infections is common or if the physical findings are highly suggestive of one of those infections.

When possible, laboratory evaluation should include dark-field microscopy, serologic testing for syphilis (e.g., rapid plasma reagin [RPR] or venereal disease research laboratory (VDRL); and fluorescent treponemal antibody [FTA] or microhemagglutination assay for antibody to *T. pallidum*), and culture for herpes. If available, RPR should be performed. Dark-field microscopy is 70% to 95% sensitive in detecting treponemes, but sensitivity is highly dependent on the expertise of the technician. Culture should seek to distinguish between HSV-1 and HSV-2, because the former typically produces a less severe infection with fewer recurrences. This is particularly important in light of recent data that suggest that HSV-1 is an increasingly common cause of genital herpes.[112] If initial evaluation does not establish a cause of genital ulcers and the clinician's suspicion for chancroid, LGV, or donovanosis remains low, further diagnostic efforts should focus on ruling out genital herpes. Several type-specific serologic tests that target the HSV glycoprotein G-2 (gG-2) are now available.[113] Patients who have no serologic evidence of HSV-2 may have primary infections. Only limited data are available on how soon seroconversion can be detected by commercially available type-specific tests, but the median time from exposure to seroconversion appears to be 2 to 3 weeks. Patients with a clinical syndrome consistent with genital herpes who test negative for HSV-2 should be retested after 6 to 12 weeks if an intervening recurrence of genital ulcers does not establish the diagnosis of HSV infection and the clinical suspicion for genital herpes is high. Older HSV serologic tests are neither sensitive nor specific and should not be used. Clinicians should be aware that type-specific serologic tests have not been studied extensively for HSV-2 screening. Given the

imperfect specificity of these tests, it is likely that widespread testing in populations in which the prevalence of HSV-2 is low will result in large numbers of false positive test results. Because of poor specificity, a positive HSV-2 serologic test result in a patient without signs or symptoms of genital herpes or definite exposure to HSV-2 should be interpreted with caution.

TREATMENT

Treatment of patients with genital ulcers is usually empirical [*see Table 6*]. If patients have physical findings suggestive of syphilis, are residents of or recent travelers to areas where syphilis remains common, or are members of groups at high risk for syphilis (e.g., men who have sex with men, as well as commercial sex workers or their clients), treatment should include benzathine penicillin G, 2.4 million units intramuscularly, and a regimen for genital herpes [*see Table 7*]. If the suspicion for syphilis is low and follow-up can be ensured, initial empirical treatment can focus on genital herpes alone. The treatment and follow-up of patients with genital herpes and syphilis are discussed in other subsections.

Patients with genital herpes should be counseled about the recurrent nature of the infection and advised that subclinical viral shedding is common. The median recurrence rate in the first year after HSV-2 acquisition is 0.33 recurrences monthly.[114] During the first 6 months after HSV-2 acquisition, virus can be isolated by culture on 6% of days and by PCR on 20% to 35% of days.[115] It is not known to what extent HSV can be transmitted by patients whose cultures are negative and whose PCR results are positive. The American Social Health Association Web site (http://www.ashastd.org) is an excellent source of information on STD in general and genital herpes in particular, and it has information on support services for persons with genital herpes.

STDs in Men Who Have Sex with Men and Anorectal STDs in Women

Although surveillance data on STDs in men who have sex with men (MSM) are limited, cases of gonorrhea and syphilis in MSM in selected cities in the United States declined by more than 10-fold in the decade following the first recognition of AIDS.[116] More recently, numerous cities in the United States and Europe have reported rising rates of STDs in MSM.[30,117] Limited data suggest that HIV transmission may also be increasing.[118] Because STDs can enhance HIV transmission, the control of STDs in MSM is a public health priority. Moreover, an STD can be a sentinel event, alerting the clinician to a patient's risk of acquiring HIV infection or transmitting it to others.

GENERAL CONSIDERATIONS IN MSM

Several aspects of the care of MSM merit consideration. First, it is imperative that clinicians adopt a nonjudgmental, direct approach when discussing sexual behavior. In addition to the questions typically included in a sexual history, clinicians should ask patients about the HIV status of their sexual partners and about their anal sexual exposure. The latter can be determined by asking, "Are you a top, a bottom, or both a top and a bottom?" The term top refers to a man who practices insertive anal sex; a bottom practices receptive anal sex.

Second, the spectrum of STD is wider in MSM than in heterosexuals. Several pathogens that are rarely sexually transmitted among heterosexuals are relatively common causes of STD in MSM. These include hepatitis A virus, *Shigella* species, *Salmonella* species, *Campylobacter* species, *Giardia lamblia*, and *Entamoeba histolytica*. *Strongyloides stercoralis* and *Enterobius vermicularis* are occasionally transmitted sexually in MSM.

Third, the anus is a more common sexual organ for MSM than it is for heterosexuals. Consequently, STD should figure prominently in the differential diagnosis of MSM who present with anorectal symptoms, and rectal screening should be part of standard STD screening in MSM.[24] Finally, although there are no guidelines for regular STD screening of heterosexual men, the CDC currently recommends annual STD screening for MSM [*see Table 1*].[8]

PROCTITIS, PROCTOCOLITIS, AND ENTERITIS

Although anorectal STD occurs in both men and heterosexual women, anal STD syndromes are more common in MSM. The symptoms of anorectal infection vary, depending on the level

Table 6 Treatment of Genital Ulcers

Disease	*Regimen*
Syphilis	Benzathine penicillin G, 2.4 million U I.M.*
Chancroid	Azithromycin, 1 g p.o. once *or* Ceftriaxone, 250 mg I.M. once *or* Ciprofloxacin, 500 mg p.o., b.i.d., for 3 days *or* Erythromycin, 500 mg p.o., t.i.d., for 7 days
Lymphogranuloma venereum	Doxycycline, 100 mg p.o., b.i.d., for 21 days
Donovanosis	Doxycycline, 100 mg p.o. for at least 3 wk or until lesion is healed *or* Trimethoprim-sulfamethoxazole, double strength (800 mg/160 mg), one tablet p.o., b.i.d., for at least 3 wk or until lesion is healed
Herpes	See Table 7

Note: treatment should always include evaluation and treatment of sexual partners.
*For primary or secondary syphilis.

Table 7 Treatment of Genital Herpes in Immunocompetent Patients

Primary herpes	Acyclovir, 400 mg p.o., t.i.d., for 7–10 days *or* Valacyclovir, 1 g p.o., b.i.d., for 7–10 days *or* Famciclovir, 250 mg p.o., t.i.d., for 7–10 days
Recurrent herpes	Acyclovir, 800 mg p.o., b.i.d., for 5 days *or* Acyclovir, 800 mg p.o., t.i.d., for 2 days *or* Valacyclovir, 1g p.o., q.d., for 5 days *or* Valacyclovir, 500 mg p.o., b.i.d., for 3–5 days *or* Famciclovir, 125 mg p.o., t.i.d., for 5 days

and extent of anatomic involvement and on the microbiologic etiology.[119] Proctitis is limited to the rectum. It results from direct inoculation of pathogens through anal sex and presents as some combination of rectal pain, constipation, hematochezia, tenesmus, and mucopurulent rectal discharge. Sexually transmitted proctitis is caused by gonorrhea, chlamydial infection (non-LGV), syphilis, or HSV. In proctocolitis, the inflammatory process extends to the colon. As a result, in addition to the symptoms of proctitis, patients may complain of diarrhea, abdominal pain, and bloating or nausea. Most cases of proctocolitis result from oral-genital or oral-anal sex (anilingus, or so-called rimming) and are caused by *Shigella*, *Salmonella*, or *Campylobacter* species; *E. histolytica*; or LGV serovars of *C. trachomatis*. *G. lamblia* infection involves the small bowel alone (enteritis) and typically presents as diarrhea, abdominal pain, and bloating or nausea in the absence of rectal symptoms. The differential diagnosis in patients presenting with symptoms of colitis or proctocolitis should include *Clostridium difficile* infection and inflammatory bowel disease. In persons with HIV infection and $CD4^+$ T cell counts less than $50/mm^3$, cytomegalovirus infection is also a possibility. Depending on their level of immunosuppression, HIV-infected patients presenting with enteritis should also be evaluated for *Mycobacterium avium* complex, cryptosporidium, *Isospora belli, Cyclospora cayetanensis,* and *Microsporidia* organisms.

DIAGNOSIS

History and Physical Examination

Evaluation of a patient with anorectal symptoms should include questions about anal, oral-genital, and oral-anal sex and condom use. In the evaluation, an attempt should be made to differentiate symptoms of proctitis, proctocolitis, and enteritis. Physical examination should include a careful anal examination, digital rectal examination, and anoscopy directed toward finding ulcers consistent with HSV or syphilis, condylomata lata, or rectal discharge or bleeding.

Laboratory Tests

If a rectal exudate is present, a Gram stain should be performed to look for gonorrhea. The reported sensitivity of Gram stain is highly variable (30% to 79%).[120] In one study, rectal Gram stain specimens obtained by anoscopy were more sensitive in detecting gonorrhea than were those obtained blindly (53% versus 79%).[120] In men without rectal symptoms, however, anoscopically obtained rectal cultures do not appear to be more sensitive in detecting gonorrhea than those obtained by blindly inserting a swab 2 to 3 cm into the rectum.[121]

Laboratory evaluation in patients with suspected proctitis or proctocolitis should include cultures for gonorrhea and chlamydial infection, a serologic test for syphilis, and a rapid syphilis test. Rectal ulcers or lesions should be cultured for HSV; when possible, a specimen should be obtained for dark-field evaluation. If symptoms suggest proctocolitis, stool specimens should be obtained for enteric pathogens and *E. histolytica*. Patients with recent antibiotic exposures should also be tested for *C. difficile*. Stool *Giardia* antigen testing should be performed if enteritis is suspected.

TREATMENT

In general, treatment should be directed by laboratory findings. If patients with proctitis have severe symptoms or if follow-up cannot be ensured, empirical therapy should be directed against gonorrhea and chlamydial infection. The CDC recommends treatment with ceftriaxone, 250 mg intramuscularly, and doxycycline, 100 mg a day orally for 7 days. Alternative therapies for gonorrhea (cefixime, 400 mg orally or ciprofloxacin, 500 mg orally once) and chlamydial infection (azithromycin, 1 g orally once) are probably effective but have not been studied. Clinicians should have a low threshold for adding empirical therapy for herpes to this regimen.

The author has received research support from Roche Diagnostics and Gen-Probe.

References

1. Eng TR, Butler WT: The Hidden Epidemic: Confronting Sexually Transmitted Diseases. Committee on Prevention and Control of Sexually Transmitted Diseases, Institute of Medicine (U.S.), National Academy Press, Washington, DC, 1997, p 432

2. Gerbase AC, Rowley JT, Heymann DH, et al: Global prevalence and incidence estimates of selected curable STDs. Sex Transm Infect 74(suppl 1):S12, 1998

3. Fleming DT, Wasserheit JN: From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect 75:3, 1999

4. Holmes KK: Introduction. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p xi

5. Holmes KK: Human ecology and behavior and sexually transmitted bacterial infections. Proc Natl Acad Sci U S A 91:2448, 1994

6. Revised guidelines for HIV counseling, testing and referral. MMWR Recomm Rep 50(RR-19):1, 2001

7. Revised recommendations for HIV screening of pregnant women. MMWR Recomm Rep 50(RR-19):63, 2001

8. Sexually transmitted diseases treatment guidelines 2002. MMWR Recomm Rep 51(RR-6):1, 2002

9. Anderson RM: Transmission dynamics of sexually transmitted infections. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 25

10. Brunham RC, Plummer FA: A general model of sexually transmitted disease epidemiology and its implications for control. Med Clin North Am 74:1339, 1990

11. Thomas JC, Tucker MJ: The development and use of the concept of a sexually transmitted disease core. J Infect Dis 174(suppl 2):S134, 1996

12. Fleming DT, McQuillan GM, Johnson RE, et al: Herpes simplex virus type 2 in the United States, 1976 to 1994. N Engl J Med 337:1105, 1997

13. Garnett GP, Anderson RM: Sexually transmitted diseases and sexual behavior: insights from mathematical models. J Infect Dis 174(suppl 2):S150, 1996

14. Morris M: Concurrent partnerships and syphilis persistence: new thoughts on an old puzzle. Sex Transm Dis 28:504, 2001

15. Kretzschmar M: Sexual network structure and sexually transmitted disease prevention: a modeling perspective. Sex Transm Dis 27:627, 2000

16. Handsfield HH, Jasman LL, Roberts PL, et al: Criteria for selective screening for *Chlamydia trachomatis* infection in women attending family planning clinics. JAMA 255:1730, 1986

17. Marrazzo JM, Fine D, Celum CL, et al: Selective screening for chlamydial infection in women: a comparison of three sets of criteria. Fam Plann Perspect 29:158, 1997

18. van Valkengoed IG, Morre SA, van den Brule AJ, et al: Low diagnostic accuracy of selective screening criteria for asymptomatic *Chlamydia trachomatis* infections in the general population. Sex Transm Infect 76:375, 2000

19. Marrazzo JM, White CL, Krekeler B, et al: Community-based urine screening for *Chlamydia trachomatis* with a ligase chain reaction assay. Ann Intern Med 127:796, 1997

20. Laumann EO, Youm Y: Racial/ethnic group differences in the prevalence of sexually transmitted diseases in the United States: a network explanation. Sex Transm Dis 26:250, 1999

21. Sucato G, Celum C, Dithmer D, et al: Demographic rather than behavioral risk factors predict herpes simplex virus type 2 infection in sexually active adolescents. Pediatr Infect Dis J 20:422, 2001

22. Laumann E, Gagnon J, Michaels S: The Social Organization of Sexuality: Sexual Practices in the United States. University of Chicago Press, Chicago, 1994

23. Sexually transmitted disease and HIV screening guidelines for men who have sex with men. Sex Transm Dis 28:457, 2001

24. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: a randomized controlled trial. Project RESPECT Study Group. JAMA 280:1161, 1998

25. Ramstedt K, Forssman L, Johannisson G: Contact tracing in the control of genital *Chlamydia trachomatis*. Int J STD AIDS 2:116, 1991

26. Kissinger P, Brown R, Reed K, et al: Effectiveness of patient delivered partner medication for preventing recurrent *Chlamydia trachomatis*. Sex Transm Infect 74:331, 1998

27. Golden MR, Whittington WL, Handsfield HH, et al: Partner management for gonococcal and chlamydial infection: expansion of public health services to the private sector and expedited sex partner treatment through a partnership with commercial pharmacies. Sex Transm Dis 28:658, 2001

28. Scholes D, Stergachis A, Heidrich FE, et al: Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. N Engl J Med 334:1362, 1996

29. Gonorrhea among men who have sex with men: selected sexually transmitted disease clinics, 1993–1996. MMWR Morb Mortal Wkly Rep 46:889, 1997

30. Nicoll A, Hamers FF: Are trends in HIV, gonorrhoea, and syphilis worsening in western Europe? BMJ 324:1324, 2002

31. Stamm WE, Hicks CB, Martin DH, et al: Azithromycin for empirical treatment of the nongonococcal urethritis syndrome in men: a randomized double-blind study. JAMA 274:545, 1995

32. Marrazzo JM, Whittington WL, Celum CL, et al: Urine-based screening for *Chlamydia trachomatis* in men attending sexually transmitted disease clinics. Sex Transm Dis 28:219, 2001

33. Bowie WR, Wang SP, Alexander ER, et al: Etiology of nongonococcal urethritis: evidence for *Chlamydia trachomatis* and *Ureaplasma urealyticum*. J Clin Invest 59:735, 1977

34. Tully JG, Taylor-Robinson D, Cole RM, et al: A newly discovered mycoplasma in the human urogenital tract. Lancet 1:1288, 1981

35. Totten PA, Schwartz MA, Sjostrom KE, et al: Association of *Mycoplasma genitalium* with nongonococcal urethritis in heterosexual men. J Infect Dis 183:269, 2001

36. Harrison WO, Hooper RR, Wiesner PJ, et al: A trial of minocycline given after exposure to prevent gonorrhea. N Engl J Med 300:1074, 1979

37. Handsfield HH, Lipman TO, Harnisch JP, et al: Asymptomatic gonorrhea in men: diagnosis, natural course, prevalence and significance. N Engl J Med 290:117, 1974

38. Turner CF, Rogers SM, Miller HG, et al: Untreated gonococcal and chlamydial infection in a probability sample of adults. JAMA 287:726, 2002

39. Klausner JD, McFarland W, Bolan G, et al: Knock-knock: a population-based survey of risk behavior, health care access, and *Chlamydia trachomatis* infection among low-income women in the San Francisco Bay area. J Infect Dis 183:1087, 2001

40. Jacobs NF, Kraus SJ: Gonococcal and nongonococcal urethritis in men: clinical and laboratory differentiation. Ann Intern Med 82:7, 1975

41. Rothenberg RB, Simon R, Chipperfield E, et al: Efficacy of selected diagnostic tests for sexually transmitted diseases. JAMA 235:49, 1976

42. Koumans EH, Johnson RE, Knapp JS, et al: Laboratory testing for *Neisseria gonorrhoeae* by recently introduced nonculture tests: a performance review with clinical and public health considerations. Clin Infect Dis 27:1171, 1998

43. Marrazzo JM, Stamm WE: New approaches to the diagnosis, treatment, and prevention of chlamydial infection. Curr Clin Top Infect Dis 18:37, 1998

44. Black CM, Marrazzo J, Johnson RE, et al: Head-to-head multicenter comparison of DNA probe and nucleic acid amplification tests for *Chlamydia trachomatis* infection in women performed with an improved reference standard. J Clin Microbiol 40:3757, 2002

45. Stamm WE, Wagner KF, Amsel R, et al: Causes of the acute urethral syndrome in women. N Engl J Med 303:409, 1980

46. Hooton TM, Stamm WE: Diagnosis and treatment of uncomplicated urinary tract infection. Infect Dis Clin North Am 11:551, 1997

47. Bent S, Nallamothu BK, Simel DL, et al: Does this woman have an acute uncomplicated urinary tract infection? JAMA 287:2701, 2002

48. Gupta K, Hooton TM, Roberts PL, et al: Patient-initiated treatment of uncomplicated recurrent urinary tract infections in young women. Ann Intern Med 135:9, 2001

49. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. N Engl J Med 333:1737, 1995

50. *Trichomonas vaginalis* associated with low birth weight and preterm delivery. The Vaginal Infections and Prematurity Study Group. Sex Transm Dis 24:353, 1997

51. Metronidazole to prevent preterm delivery in pregnant women with asymptomatic bacterial vaginosis. National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. N Engl J Med 342:534, 2000

52. Klebanoff MA, Carey JC, Hauth JC, et al: Failure of metronidazole to prevent preterm delivery among pregnant women with asymptomatic *Trichomonas vaginalis* infection. N Engl J Med 345:487, 2001

53. Koumans EH, Kendrick JS: Preventing adverse sequelae of bacterial vaginosis: a public health program and research agenda. Sex Transm Dis 28:292, 2001

54. Gulmezoglu AM: Interventions for trichomoniasis in pregnancy. Cochrane Database Syst Rev (3):CD000220, 2002

55. Martin HL, Richardson BA, Nyange PM, et al: Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. J Infect Dis 180:1863, 1999

56. Taha TE, Hoover DR, Dallabetta GA, et al: Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. AIDS 12:1699, 1998

57. Cu-Uvin S, Hogan JW, Caliendo AM, et al: Association between bacterial vaginosis and expression of human immunodeficiency virus type 1 RNA in the female genital tract. Clin Infect Dis 33:894, 2001

58. Sorvillo F, Smith L, Kerndt P, et al: *Trichomonas vaginalis,* HIV, and African-Americans. Emerg Infect Dis 7:927, 2001

59. Wang CC, McClelland RS, Reilly M, et al: The effect of treatment of vaginal infections on shedding of human immunodeficiency virus type 1. J Infect Dis 183:1017, 2001

60. Allen-Davis JT, Beck A, Parker R, et al: Assessment of vulvovaginal complaints: accuracy of telephone triage and in-office diagnosis. Obstet Gynecol 99:18, 2002

61. Ferris DG, Nyirjesy P, Sobel JD, et al: Over-the-counter antifungal drug misuse associated with patient-diagnosed vulvovaginal candidiasis. Obstet Gynecol 99:419, 2002

62. Sobel JD: Vaginitis. N Engl J Med 337:1896, 1997

63. Hillier S, Holmes K: Bacterial vaginosis. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 563

64. Eschenbach DA, Davick PR, Williams BL, et al: Prevalence of hydrogen peroxide-producing *Lactobacillus* species in normal women and women with bacterial vaginosis. J Clin Microbiol 27:251, 1989

65. Gardner H, Dukes C: *Haemophilis vaginalis* vaginitis: a newly defined specific infection previously classified as "nonspcific" vaginitis. Am J Obstet Gynecol 69:962, 1955

66. Criswell BS, Ladwig CL, Gardner HL, et al: *Haemophilus vaginalis*: vaginitis by inoculation from culture. Obstet Gynecol 33:195, 1969

67. Marrazzo JM, Koutsky LA, Eschenbach DA, et al: Characterization of vaginal flora and bacterial vaginosis in women who have sex with women. J Infect Dis 185:1307, 2002

68. Avonts D, Sercu M, Heyerick P, et al: Incidence of uncomplicated genital infections in women using oral contraception or an intrauterine device: a prospective study. Sex Transm Dis 17:23, 1990

69. Barbone F, Austin H, Louv WC, et al: A follow-up study of methods of contraception, sexual activity, and rates of trichomoniasis, candidiasis, and bacterial vaginosis. Am J Obstet Gynecol 163:510, 1990

70. Hawes SE, Hillier SL, Benedetti J, et al: Hydrogen peroxide–producing lactobacilli and acquisition of vaginal infections. J Infect Dis 174:1058, 1996

71. Amsel R, Totten PA, Spiegel CA, et al: Nonspecific vaginitis: diagnostic criteria and microbial and epidemiologic associations. Am J Med 74:14, 1983

72. Vejtorp M, Bollerup AC, Vejtorp L, et al: Bacterial vaginosis: a double-blind randomized trial of the effect of treatment of the sexual partner. Br J Obstet Gynaecol 95:920, 1988

73. Vutyavanich T, Pongsuthirak P, Vannareumol P, et al: A randomized double-blind trial of tinidazole treatment of the sexual partners of females with bacterial vaginosis. Obstet Gynecol 82:550, 1993

74. Colli E, Landoni M, Parazzini F: Treatment of male partners and recurrence of bacterial vaginosis: a randomised trial. Genitourin Med 73:267, 1997

75. Eschenbach DA, Hillier S, Critchlow C, et al: Diagnosis and clinical manifestations of bacterial vaginosis. Am J Obstet Gynecol 158:819, 1988

76. Thomason JL, Gelbart SM, Anderson RJ, et al: Statistical evaluation of diagnostic criteria for bacterial vaginosis. Am J Obstet Gynecol 162:155, 1990

77. Demographic and behavioral predictors of *Trichomonas vaginalis* infection among pregnant women. The Vaginal Infections and Prematurity Study Group. Obstet Gynecol 78:1087, 1991

78. Bowden FJ, Garnett GP: *Trichomonas vaginalis* epidemiology: parameterising and analysing a model of treatment interventions. Sex Transm Infect 76:248, 2000

79. Wolner-Hanssen P, Krieger JN, Stevens CE, et al: Clinical manifestations of vaginal trichomoniasis. JAMA 261:571, 1989

80. Krieger J, Alderete J: *Trichomonas vaginalis* and trichomoniasis. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 587

81. Levi MH, Torres J, Pina C, et al: Comparison of the InPouch TV culture system and Diamond's modified medium for detection of *Trichomonas vaginalis*. J Clin Microbiol 35:3308, 1997

82. Tidwell BH, Lushbaugh WB, Laughlin MD, et al: A double-blind placebo-controlled trial of single-dose intravaginal versus single-dose oral metronidazole in the treatment of trichomonal vaginitis. J Infect Dis 170:242, 1994

83. Sobel JD, Nyirjesy P, Brown W: Tinidazole therapy for metronidazole-resistant vaginal trichomoniasis. Clin Infect Dis 33:1341, 2001

84. Geiger AM, Foxman B, Gillespie BW: The epidemiology of vulvovaginal candidiasis among university students. Am J Public Health 85:1146, 1995

85. Foxman B, Barlow R, D'Arcy H, et al: *Candida* vaginitis: self-reported incidence and associated costs. Sex Transm Dis 27:230, 2000

86. Foxman B: The epidemiology of vulvovaginal candidiasis: risk factors. Am J Public Health 80:329, 1990

87. Abbott J: Clinical and microscopic diagnosis of vaginal yeast infection: a prospective analysis. Ann Emerg Med 25:587, 1995

88. Ryan CA, Courtois BN, Hawes SE, et al: Risk assessment, symptoms, and signs as predictors of vulvovaginal and cervical infections in an urban US STD clinic: implications for use of STD algorithms. Sex Transm Infect 74(suppl 1):S59, 1998

89. Schaaf VM, Perez-Stable EJ, Borchardt K: The limited value of symptoms and signs in the diagnosis of vaginal infections. Arch Intern Med 150:1929, 1990

90. Sobel J: Vulvovaginal candidiasis. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 629

91. Bertholf ME, Stafford MJ: Colonization of *Candida albicans* in vagina, rectum, and mouth. J Fam Pract 16:919, 1983

92. Sobel JD, Chaim W: Treatment of *Torulopsis glabrata* vaginitis: retrospective review of boric acid therapy. Clin Infect Dis 24:649, 1997

93. Myziuk L, Romanowski B, Brown M: Endocervical Gram stain smears and their usefulness in the diagnosis of *Chlamydia trachomatis*. Sex Transm Infect 77:103, 2001

94. Marrazzo J, Handsfield H, Whittington W: Predicting chlamydial and gonococcal cervical infection: implications for management of cervicitis. Obstet Gynecol 100:579, 2002

95. Miller HG, Cain VS, Rogers SM, et al: Correlates of sexually transmitted bacterial infections among U.S. women in 1995. Fam Plann Perspect 31:4, 1999

96. Hubacher D, Lara-Ricalde R, Taylor DJ, et al: Use of copper intrauterine devices and the risk of tubal infertility among nulligravid women. N Engl J Med 345:561, 2001

97. Westrom L, Eschenbach D: Pelvic inflammatory disease. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 783

98. Cohen CR, Manhart LE, Bukusi EA, et al: Association between *Mycoplasma genitalium* and acute endometritis. Lancet 359:765, 2002

99. Kahn JG, Walker CK, Washington AE, et al: Diagnosing pelvic inflammatory disease: a comprehensive analysis and considerations for developing a new model. JAMA 266:2594, 1991

100. Peipert JF, Ness RB, Blume J, et al: Clinical predictors of endometritis in women with symptoms and signs of pelvic inflammatory disease. Am J Obstet Gynecol 184:856, 2001

101. Wolner-Hanssen P: Silent pelvic inflammatory disease: is it overstated? Obstet Gynecol 86:321, 1995

102. Munday PE: Pelvic inflammatory disease: an evidence-based approach to diagnosis. J Infect 40:31, 2000

103. Jacobson L, Westrom L: Objectivized diagnosis of acute pelvic inflammatory disease: diagnostic and prognostic value of routine laparoscopy. Am J Obstet Gynecol 105:1088, 1969

104. Eckert LO, Hawes SE, Wolner-Hanssen PK, et al: Endometritis: the clinical-pathologic syndrome. Am J Obstet Gynecol 186:690, 2002

105. Molander P, Sjoberg J, Paavonen J, et al: Transvaginal power Doppler findings in laparoscopically proven acute pelvic inflammatory disease. Ultrasound Obstet Gynecol 17:233, 2001

106. Ness RB, Soper DE, Holley RL, et al: Effectiveness of inpatient and outpatient treatment strategies for women with pelvic inflammatory disease: results from the Pelvic Inflammatory Disease Evaluation and Clinical Health (PEACH) Randomized Trial. Am J Obstet Gynecol 186:929, 2002

107. Westrom L, Joesoef R, Reynolds G, et al: Pelvic inflammatory disease and fertility: a cohort study of 1,844 women with laparoscopically verified disease and 657 control women with normal laparoscopic results. Sex Transm Dis 19:185, 1992

108. Wald A, Link K: Risk of human immunodeficiency virus infection in herpes simplex virus type 2–seropositive persons: a meta-analysis. J Infect Dis 185:45, 2002

109. Etiology of genital ulcers and prevalence of human immunodeficiency virus coinfection in 10 US cities. The Genital Ulcer Disease Surveillance Group. J Infect Dis 178:1795, 1998

110. Korenromp EL, Bakker R, De Vlas SJ, et al: Can behavior change explain increases in the proportion of genital ulcers attributable to herpes in sub-Saharan Africa? A simulation modeling study. Sex Transm Dis 29:228, 2002

111. DiCarlo RP, Martin DH: The clinical diagnosis of genital ulcer disease in men. Clin Infect Dis 25:292, 1997

112. Lafferty WE, Downey L, Celum C, et al: Herpes simplex virus type 1 as a cause of genital herpes: impact on surveillance and prevention. J Infect Dis 181:1454, 2000

113. Ashley RL: Performance and use of HSV type-specific serology test kits. Herpes 9:38, 2002

114. Benedetti J, Corey L, Ashley R: Recurrence rates in genital herpes after symptomatic first-episode infection. Ann Intern Med 121:847, 1994

115. Corey L, Wald A: Genital herpes. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 285

116. Doll L, Ostrow D: Homosexual and bisexual behavior. Sexually Transmitted Diseases, 3rd ed. Holmes KK, Mårdh P-A, Sparling PF, et al, Eds. McGraw-Hill, New York, 1999, p 151

117. Mayer KH, Klausner JD, Handsfield HH: Intersecting epidemics and educable moments: sexually transmitted disease risk assessment and screening in men who have sex with men. Sex Transm Dis 28:464, 2001

118. Dukers NH, Spaargaren J, Geskus RB, et al: HIV incidence on the increase among homosexual men attending an Amsterdam sexually transmitted disease clinic: using a novel approach for detecting recent infections. AIDS 16:F19, 2002

119. Quinn TC, Stamm WE, Goodell SE, et al: The polymicrobial origin of intestinal infections in homosexual men. N Engl J Med 309:576, 1983

120. William DC, Schapiro CM, Felman YM: Pharyngeal carriage of *Neisseria meningitidis* and anogenital gonorrhea: evidence for their relationship. Sex Transm Dis 7:175, 1980

121. Deheragoda P: Diagnosis of rectal gonorrhoea by blind anorectal swabs compared with direct vision swabs taken via a proctoscope. Br J Vener Dis 53:311, 1977

XXIII URINARY TRACT INFECTIONS

Kalpana Gupta, m.d., m.p.h
Walter E. Stamm, m.d., f.a.c.p.

Definitions

Urinary tract infection (UTI) is the most common of all bacterial infections; it affects persons throughout their life span. The term UTI encompasses a variety of clinical entities, ranging from asymptomatic bacteriuria to cystitis, prostatitis, and pyelonephritis. UTIs may be further characterized as uncomplicated (occurring without an anatomic or other predisposing reason) or complicated (associated with structural or functional abnormalities of the urinary tract and kidney) and as community acquired or nosocomial (generally, catheter associated).

Epidemiology and Risk Factors

UTI occurs far more commonly in females than in males, except at the extremes of age [*see Table 1*]. During the neonatal period, the incidence of UTI is slightly higher in males than in females because of the greater frequency of congenital anomalies of the urinary tract in male infants. After 50 years of age, the incidence of UTI is almost as high in men as in women, presumably because of obstruction from prostatic hypertrophy. In persons between 1 and about 50 years of age, UTI is predominantly a disease of females.

WOMEN

As many as 50% to 80% of women in the general population acquire at least one UTI during their lifetime; most of these infections are uncomplicated cystitis.[1] In a prospective cohort study of sexually active healthy women, the incidence of acute cystitis was 50 to 70 episodes per 100 person-years.[2] Recent use of a diaphragm with spermicide, frequency of sexual intercourse, and a history of UTI were identified as independent risk factors for cystitis in this study.[2] Cystitis has also been temporally related to recent sexual intercourse.

The incidence of acute uncomplicated pyelonephritis is difficult to ascertain because this infection is less common than cystitis and because most episodes are treated in the outpatient setting. In a population-based case-control study of young healthy women, the annual incidence of pyelonephritis was approximately 28 per 10,000 women. Factors independently associated with pyelonephritis in this study included frequency of sexual intercourse, having a new sexual partner, UTI in the past 12 months, maternal history of UTI, diabetes, and incontinence.[3]

Table 1 Incidence of Urinary Tract Infection According to Age and Sex

Age Group	*Incidence (%)*	*Approximate Sex Ratio (Male:Female)*
Neonatal	1.0	1.5:1.0
Preschool age	1.5–3.0	1:10
School age	1.2	1:30
Reproductive age	3–5	1:50
Geriatric	10–30	1:1.5

Thus, many of the factors predisposing women to cystitis also increase the risk of pyelonephritis.

YOUNG MEN

UTI is rare in young men and has traditionally been attributed to the presence of urologic abnormalities. However, it is apparent that uncomplicated UTI can occur in men who acquire uropathogens through direct sexual contact, in the form of unprotected anal intercourse with a man or a woman, or unprotected vaginal intercourse with a woman whose vagina is colonized with uropathogens.[4] Lack of circumcision is also associated with an increased risk of UTI, because of an increased incidence of *Escherichia coli* colonization of the glans and prepuce and the subsequent migration of *E. coli* to the urinary tract.[5,6]

RECURRENT UTI

About 20% to 30% of women who have had one episode of UTI will have recurrent episodes.[4] Recurrence may result from relapse or reinfection. Relapse in either sex is caused by the reappearance of an organism from a sequestered focus, usually within the kidney or prostate, shortly after completion of therapy. Sequestration of infecting organisms in the bladder epithelium has been demonstrated in animals, but the importance of this phenomenon in humans is not yet clear.[7] In reinfection, the course of therapy has successfully eradicated the infection and there is no sequestered focus, but organisms are reintroduced from the fecal reservoir. The majority of recurrences are thought to be reinfections.[4] Studies of the natural history of recurrent UTI in women have found that the rate of recurrence ranges from 0.3 to 7.6 infections per patient per year, with an average rate of 2.6 infections per year.[8] Clustering of episodes occurs; it is not uncommon for multiple recurrences to follow an initial infection. The likelihood of a recurrence decreases with increasing time since the last infection. A case-control study of women with recurrent UTI identified frequency of sexual intercourse, use of spermicide, having a new sexual partner, a history of first UTI occurring before 15 years of age, and a maternal history of UTI as independent risk factors for recurrent UTI.[9]

PREGNANCY

The incidence of asymptomatic bacteriuria in pregnant women is approximately 4% to 10%, which is similar to the rate reported in sexually active nonpregnant women of childbearing age.[10-12] If not treated, 20% to 40% of pregnant women with bacteriuria in the first trimester will acquire acute pyelonephritis later in pregnancy. A recent meta-analysis estimated that treatment of asymptomatic bacteriuria would lead to approximately a 75% reduction in the incidence of pyelonephritis.[12] Premature births and perinatal mortality are increased in pregnancies complicated by UTI.[11,12] There is little evidence that the infections that develop during pregnancy have long-term effects.

DIABETES

The rates of asymptomatic bacteriuria and UTI in diabetic women are twofold to threefold higher than those in nondiabetic women; these differences have not been observed in men.[13] In hospitalized diabetic patients, particularly those with multiple-

organ complications, the incidence of infection and true pyelonephritis also appears to be increased, partly because of poor bladder function and urinary catheterization. Other clinical conditions causing obstruction in urinary flow or incomplete voiding also predispose diabetic patients to infection. In addition, impaired cytokine secretion may contribute to asymptomatic bacteriuria in diabetic women.[14]

Etiology

The spectrum of organisms causing UTI varies by clinical syndrome. In acute uncomplicated cystitis, the etiologic agents are highly predictable: *E. coli* accounts for 75% to 90% of isolates; *Staphylococcus saprophyticus* accounts for 5% to 15% of isolates (particularly in younger women); and *Klebsiella* species, *Proteus* species, enterococci, and other organisms account for 5% to 10% of isolates. The spectrum of agents that cause uncomplicated pyelonephritis is less well studied than, but is similar to, that which causes acute cystitis. In complicated UTIs, *E. coli* remains the predominant organism, but other aerobic gram-negative rods, such as *Klebsiella* species, *Proteus* species, *Citrobacter* species, *Acinetobacter* species, *Morganella* species, and *Pseudomonas aeruginosa* are also frequently isolated. Gram-positive bacteria, such as enterococci, *S. aureus,* and *S. epidermidis,* as well as yeast, are also important pathogens in complicated UTI.

Pathogenesis

Bacteria can establish infection in the urinary tract by traveling from the urethra to the bladder and then up the ureter to the kidney. However, introduction of bacteria into the bladder does not inevitably lead to sustained infection. For example, bacteria often enter the bladder after sexual intercourse, but normal micturition and innate host defense mechanisms in the bladder eliminate these organisms. The bladder mucosal surface has antibacterial properties that eliminate some organisms, presumably through mucus trapping and a polymorphonuclear leukocyte response. In addition, urine that has a low pH, high or very low osmolarity, high urea concentration, or high organic acid content inhibits bacterial growth. Abnormal micturition, a significant residual urine volume, or both will promote true infection. There are also acquired and intrinsic host factors, as well as bacterial virulence factors, which increase the likelihood of development of UTI (see below).

Bacteria can also gain access to the urinary tract through the bloodstream. However, hematogenous spread accounts for fewer than 2% of documented UTIs and usually results from bacteremia caused by relatively virulent organisms, such as *Salmonella* and *S. aureus.* Hematogenous infections may produce focal abscesses or areas of pyelonephritis within a kidney and result in positive urine cultures.

VAGINAL ECOLOGY AND UTI

In women, colonization of the vaginal introitus with organisms from the fecal flora, usually *E. coli,* is the critical initial step in the pathogenesis of UTI. Sexual intercourse and the use of a diaphragm with spermicide or of spermicide alone are strongly associated with an increased risk of *E. coli* vaginal colonization and bacteriuria, probably because of alterations in the normal vaginal microflora.[15]

In postmenopausal women, there is an increased incidence of gram-negative vaginal colonization and bacteriuria. These trends correlate with the changes in the vaginal environment that occur with menopause: disappearance of the previously predominant lactobacilli from the vaginal microflora and a rise in pH. A case-control study of community-dwelling postmenopausal women found significantly lower rates of vaginal colonization with lactobacilli in those women who were not taking hormone replacement therapy than in those who were using systemic or topical estrogen.[16] In a randomized, placebo-controlled trial in postmenopausal women with a history of recurrent UTI, topical estrogen therapy resulted in restoration of the premenopausal vaginal flora and a decrease in both the prevalence of vaginal *E. coli* colonization and the incidence of UTI.[17]

GENETIC FACTORS

There is increasing evidence that genetically determined factors may influence susceptibility to recurrent UTI. Women with recurrent UTI demonstrate a propensity for persistent vaginal colonization with *E. coli,* even during asymptomatic periods. Vaginal and periurethral mucosal cells from women with recurrent UTI bind threefold more uropathogenic bacteria than do mucosal cells from women without recurrent infection. These observations suggest that epithelial cells from susceptible women may possess specific types or greater numbers of receptors to which *E. coli* can bind, thereby facilitating colonization. This increased susceptibility is determined in part by Lewis blood group type and whether the woman secretes blood group antigens into bodily fluids. Vaginal epithelial cells from nonsecretors of blood group antigens bind significantly greater numbers of bacteria, and nonsecretors are particularly at risk for recurrent UTI.[18,19] Mutations in host-response genes (e.g., those coding for Toll receptors and the interleukin-8 receptor) have been linked to severity of UTIs in animals, although they have not yet been linked in humans.[20]

BACTERIAL VIRULENCE

Certain strains of *E. coli* possess chromosomally encoded virulence determinants that confer the ability to infect the anatomically normal urinary tract and produce acute inflammatory disease (e.g., cystitis and pyelonephritis). Characteristics that have been associated with uropathogenicity are the presence of certain O and K surface antigens (the O antigen is the outer polysaccharide portion of the bacterial envelope, and the K antigen is the antiphagocytic capsular antigen), the presence of the siderophore aerobactin, resistance to the bactericidal activity of serum, the ability to produce toxins such as hemolysin and cytotoxic necrotizing factor, and certain intracellular metabolic capabilities.[21]

Also important is the presence of adhesins on the surface of uropathogenic bacteria that mediate binding to specific receptors on the surface of uroepithelial cells. The best-studied adhesion structure is the P fimbriae, which are hairlike protein structures found on the surface of certain pathogenic strains of *E. coli.* P fimbriae interact with a specific receptor on epithelial cells. This epithelial cell receptor contains the carbohydrate moiety α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside, which is found in the P blood group antigens. The prevalence of P-fimbriated *E. coli* correlates with severity of disease the strain is likely to cause: there is a low prevalence (10% to 20%) in *E. coli* strains from the fecal flora of asymptomatic persons; there is a higher prevalence (50% to 60%) in strains that cause cystitis; and the highest prevalence (70% to 100%) is found in strains that cause pyelonephritis.[21,22] P fimbriae also appear to be important

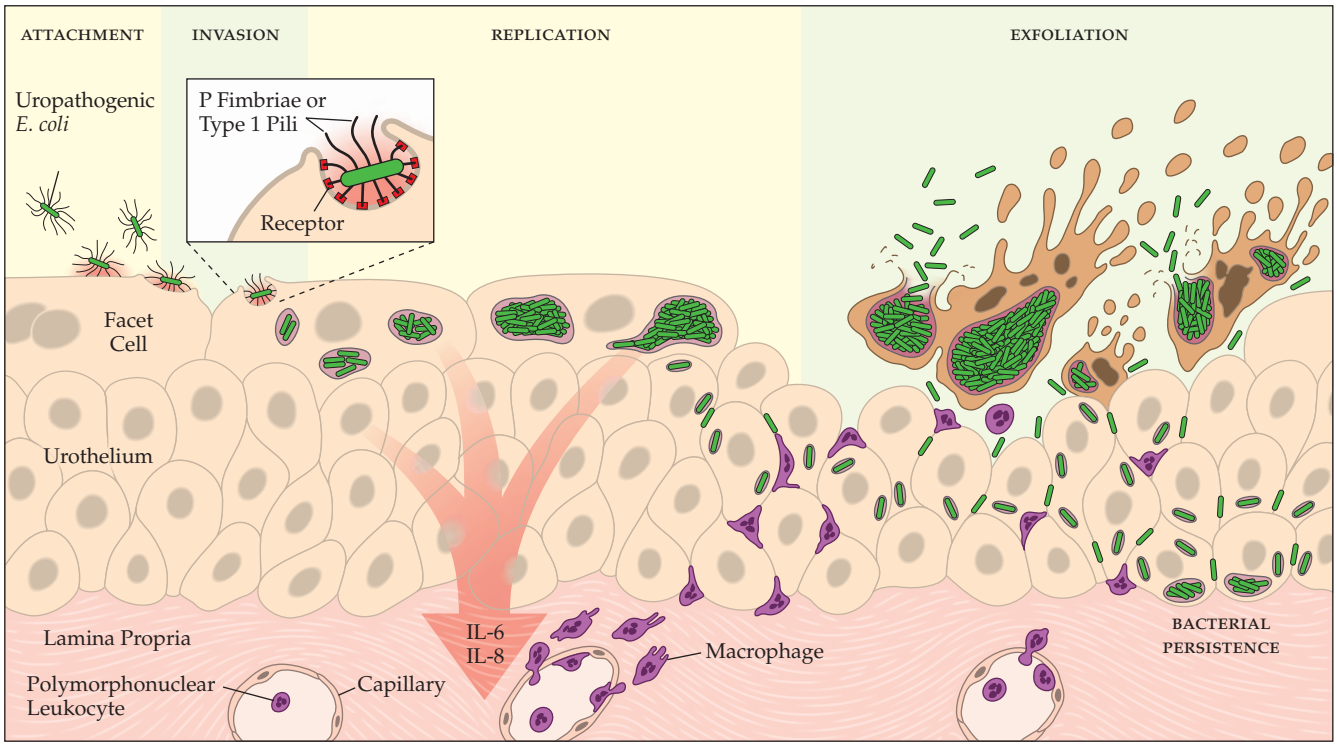


Figure 1 **The pathophysiology of infection by uropathogenic *Escherichia coli* in bladder epithelial cells.[20] Uropathogenic *E. coli* organisms attach to receptors on superficial bladder cells with P fimbriae or type 1 pili. Once contact is established, the bacteria are internalized into the cells, where they can replicate to high levels. However, attachment or invasion can result in the activation of apoptotic pathways within the cells, leading to the eventual exfoliation and clearance of infected host cells. Interactions between *E. coli* and the cells can also result in the induction of inflammatory cytokines, leading to the influx of polymorphonuclear leukocytes into the bladder epithelium. *E. coli* can escape from dying cells, thereby avoiding clearance by exfoliation, and infect surrounding and underlying epithelial cells. Within the bladder epithelium, *E. coli* can escape immune surveillance and persist at subclinical levels. (IL-6 — interleukin-6; IL-8 — interleukin-8)**

in the pathogenesis of bloodstream invasion from the kidney. From 75% to 100% of *E. coli* strains isolated from the blood of otherwise healthy patients with pyelonephritis express P fimbriae. In contrast, *E. coli* strains that cause urosepsis in persons with compromising medical conditions are much less likely to have P fimbriae.[22]

Another adhesion structure is the type 1 pilus (fimbria), which all *E. coli* strains possess. Type 1 pili are also thought to play a key role in initiating *E. coli* bladder infection; they mediate binding to uroplakins, which are mannosylated glycoproteins on the surface of bladder uroepithelial cells.[21] In addition to P fimbriae and type 1 pili, other adhesins may play a role in mediating attachment of uropathogenic strains of *E. coli* to the uroepithelium in selected circumstances.

The binding of uropathogenic *E. coli* to receptors on uroepithelial cells initiates a complex series of intracellular signaling events that alters epithelial cell function [*see Figure 1*].[8] Chemokines and cytokines are synthesized and secreted, inflammatory cells are attracted into the bladder epithelium, and epithelial cells undergo apoptosis and exfoliation, carrying attached *E. coli* away in the urine.

ANATOMIC AND FUNCTIONAL ABNORMALITIES

Persons who have major anatomic and functional abnormalities of the urinary tract, including vesicoureteral reflux, ureteral obstruction, or a foreign body (e.g., a stone, a catheter, or a tumor), are markedly predisposed to UTI, particularly infections involving the kidney. In such persons, UTI can develop as a result of infection with bacteria, including *E. coli* strains, that are not normally uropathogenic.[21] Not surprisingly, infection with nonuropathogenic strains of bacteria is also common in hospitalized patients, presumably because they have anatomic and functional abnormalities. Bacterial invasion of the prostate and incomplete bladder emptying caused by bladder outlet obstruction are the primary predisposing factors in men with UTI. Instrumentation and incomplete bladder emptying are important predisposing factors for UTI in patients with spinal cord injury or diabetes. Inhibition of ureteral peristalsis leading to vesicoureteral reflux is important in the pathogenesis of pyelonephritis in pregnant women.

Vesicoureteral reflux plays a key role in the pathogenesis of renal infection and, more important, in the evolution of chronic renal damage. It is commonly associated with UTI in children but also occurs because of anatomic abnormalities in children without UTI. Reflux provides a direct route for infection to reach the pelvicalyceal system of the kidney; severe reflux may occur intrarenally. Renal scars caused by chronic pyelonephritis are often associated with such reflux. Most reflux-associated renal damage occurs in infancy, because with growth there is a tendency for self-correction of at least mild degrees of reflux. The long-term effects of reflux on kidney structure and function are a major justification for an aggressive radiologic approach in infants and young children with UTI [*see* Management, *below*].[23]

Diagnosis

CLINICAL PRESENTATIONS AND LABORATORY FINDINGS

The clinical presentation of UTI is quite variable, ranging from asymptomatic bacteriuria to typical symptomatic cystitis to acute pyelonephritis. In addition, clinical symptoms do not always correlate with the site of infection (bladder versus kidney) or with the degree of bacteriuria. Approximately 30% of patients with lower urinary tract symptoms also have silent infection of the kidney.[4] Despite considerable effort, researchers have been unable to develop a noninvasive technique for differentiating renal infections from bladder infections. The best noninvasive test to delineate the anatomic site of infection appears to be the response to short-course antibiotic therapy [*see* Management, *below*].

Cystitis

The typical symptoms of cystitis are dysuria, urinary frequency, and urgency. Nocturia and suprapubic or back discomfort are also often present. In addition, the urine may be cloudy, malodorous, or bloody.

A meta-analysis evaluating the probability of UTI on the basis of history and physical findings concluded that the probability of UTI was at least 50% in a woman presenting with one or more symptoms of UTI. If vaginal discharge and complicating factors are absent and risk factors for UTI are present, then the probability of UTI is close to 90% and no laboratory evaluation is needed. Similarly, a combination of dysuria and frequency—the most common symptoms—in the absence of vaginal discharge increases the probability of UTI to 96%. Further laboratory evaluation with dipstick or urine culture can be omitted in such patients, and empirical therapy can be considered. If the history is not clear, then a urine dipstick should be performed. A positive nitrite or leukocyte esterase result makes the probability of UTI about 80%, and empirical treatment can be considered without further testing. In this setting, a negative dipstick result does not rule out UTI, and a urine culture and close clinical follow-up are recommended. Interestingly, the only physical examination finding that increased the probability of UTI was costovertebral angle tenderness (suggesting pyelonephritis), and the authors recommended that the physical examination could be omitted in patients with typical symptoms of acute uncomplicated cystitis.[24]

On dipstick testing, the presence of leukocyte esterase, nitrite, or both has about 75% to 90% sensitivity and 70% to 82% specificity.[4] The pH is typically elevated, and blood may be present. Depending on the clinical circumstances, it may be appropriate to follow dipstick testing with urine culture and antimicrobial-sensitivity testing [*see* Management, *below*].

Urine microscopy reveals pyuria in nearly all cases of cystitis and hematuria in about 30% of cases. Bacteriuria is demonstrable on Gram stain of unspun urine in over 90% of cases with 10^5 bacteria/ml or higher. Approximately two thirds of women presenting with clinical cystitis will have urinary colony counts of 10^5 bacteria/ml or higher. Studies in women with symptoms of cystitis have found that a colony count threshold of greater than 10^2 bacteria/ml has greater sensitivity (95%) and specificity (85%) than a threshold of 10^5 bacteria/ml for the diagnosis of acute cystitis in women[4] [*see* Bacterial Count on Urine Culture, *below*].

Pyelonephritis

Patients with pyelonephritis can present with clinical manifestations that range from mild to relatively severe—from low-grade fever with lower back or costovertebral angle pain to fever, shaking chills, nausea, vomiting, and loin pain. Symptoms are generally acute in onset and may or may not be associated with symptoms of cystitis.[4,25] The primary finding on physical examination is costovertebral angle tenderness on deep palpation. Tachycardia may accompany fever. Pyuria and bacteriuria are usually demonstrable on urine microscopy and Gram stain. Bacteremia may complicate the course of pyelonephritis; but in patients with pyelonephritis, bacteremia is seldom associated with the more serious sequelae of gram-negative infection (i.e., triggering of the complement, clotting, and kinin systems), which may lead to septic shock, disseminated intravascular coagulation, or both. When shock or disseminated intravascular coagulation occurs, the possibility of obstruction must be considered. In a particularly serious form of obstructive uropathy associated with acute papillary necrosis, the sloughed papillae may obstruct the ureter. This should be suspected in diabetic patients who have severe pyelonephritis and persistent bacteremia despite antibiotic therapy. Papillary necrosis may also be evident in some cases of pyelonephritis complicated by obstruction, sickle cell disease, analgesic nephropathy, or combinations of these conditions. Emphysematous pyelonephritis, which is a particularly severe form of pyelonephritis associated with production of gas in renal and perinephric tissues, occurs almost entirely in diabetic patients.

Renal and Perirenal Abscesses

Two unusual forms of UTI are macroscopic renal and perirenal abscesses. In the past, most of these abscesses were secondary to hematogenous infection with *S. aureus.* Currently, most of them are secondary to ascending UTI with the usual Enterobacteriaceae organisms. Such infections are often complicated by renal calculi and obstruction of urinary flow from either the kidney or the ureter. Less commonly, preexisting renal cysts become infected and develop into abscesses. In rare instances, there is contiguous spread from a neighboring site of suppuration, such as the colon or overlying rib. The usual presentation of such infections is insidious, with chronic fever, weight loss, night sweats, and anorexia, and is often associated with flank or back pain. When the abscess is under pressure, usually because of obstruction, a more acute presentation with associated bacteremia may occur. Symptoms specific to the urinary tract, such as dysuria, hematuria, and urinary retention, are sometimes noted but are often absent. On physical examination, costovertebral angle tenderness or even a palpable mass may be found; in 30% to 50% of patients, however, this finding is absent.

Routine laboratory tests are of variable value in patients with renal or perirenal abscesses: leukocytosis may be present, anemia is not unusual, and signs of inflammation (e.g., pyuria or proteinuria) may be evident on urinalysis. In more than half of patients with renal or perirenal abscesses, the organism in the abscess may be isolated on urine culture. Definitive diagnosis, however, depends on radiographic detection of a mass lesion. Gallium and ultrasound scans may be helpful, but a computed tomographic or magnetic resonance imaging scan is considered the diagnostic test of choice.[26] If prompt drainage and antibiotic therapy are not provided, renal or perirenal abscesses may extend to the peritoneal cavity, chest, or skin.

Prostatitis

A common complication of UTI in men is prostatitis. Bacterial prostatitis is usually caused by the same gram-negative bacilli

that cause UTI in females; 80% or more of such infections are caused by *E. coli.* The pathogenesis of this condition is poorly understood. Antibacterial substances in prostatic secretions probably protect against such infections.

A National Institutes of Health expert consensus panel has recommended classifying prostatitis into three syndromes: acute bacterial prostatitis, chronic bacterial prostatitis, and chronic pelvic pain syndrome (CPPS). Acute bacterial prostatitis is a febrile illness characterized by chills; dysuria; urinary frequency and urgency; and perineal, back, or pelvic pain. Bladder outlet obstruction may occur. On physical examination, the prostate is found to be enlarged, tender, and indurated. Pyuria is present, and urine cultures generally grow *E. coli* or another typical uropathogen.

Chronic bacterial prostatitis is a clinically more occult disease and may be manifested only as recurrent bacteriuria or variable low-grade fever with back or pelvic discomfort. Urinary symptoms usually relate to the reintroduction of infection into the bladder, with both pyuria and bacteriuria being present; a chronic prostatic focus is the most common cause of recurrent UTI in men.

CPPS describes the large group of men who present with minimal signs on physical examination but have a variety of irritative or obstructive voiding symptoms; perineal, pelvic, or back pain; and sexual dysfunction. These men can be divided into those with and those without inflammation (defined as > 10 white blood cells per high-power field in expressed prostatic secretions). The etiology and appropriate management in these patients, regardless of inflammatory status, is unknown.[27,28]

INTERPRETATION OF URINE CULTURES

The correlation between clinical symptoms and the presence of infection is imprecise. In addition, it is frequently difficult to obtain a urine specimen that is uncontaminated by the normal microbial flora of the distal urethra, vagina, or skin. When evaluating urine cultures, therefore, the physician should consider both the species and the number of bacteria found.[4]

Species Found on Urine Culture

More than 95% of UTIs result from a single bacterial species. Thus, in most instances, a culture that grows out mixed bacterial species is contaminated and needs to be repeated. Polymicrobial infection may occur in certain settings, however, including long-term catheterization, incomplete bladder emptying because of neurologic dysfunction, and the presence of a fistula between the urinary tract and the GI or genital tract. Organisms that are commonly found in the vaginal introitus and distal urethra, such as *S. epidermidis,* diphtheroids, and lactobacilli, seldom cause UTI except in complicated settings. Another marker of contamination is the presence of squamous epithelial cells in a urine specimen. On the other hand, pyuria is highly associated with inflammation of the urinary tract, which is most commonly caused by infection.

Bacterial Count on Urine Culture

The number of organisms per milliliter of a clean-voided urine specimen is a useful indicator.[4] The growth of 10^5 or more colonies of a single bacterial species per milliliter of urine in two consecutive urine cultures is the diagnostic criterion for asymptomatic bacteriuria in women. In women with acute dysuria, the presence of at least 10^2 organisms/ml of a single coliform species appears to be a more accurate criterion for infection than a threshold of 10^5 organisms/ml. Women with dysuria and pyuria who have less than 10^2 bacteria/ml may have urethritis caused by *Chlamydia trachomatis, Neisseria gonorrhoeae,* or herpes simplex virus or vulvovaginitis caused by *Trichomonas vaginalis* or *Candida* species. Clinical and laboratory criteria can help differentiate these conditions [*see Table 2*]. In men, the minimal level indicating infection appears to be 10^3 organisms/ml. In general, probably 70% of persons with true bacterial UTI will have more than 10^5 organisms/ml; 30% will have lower concentrations of bacteria.

The presence of organisms on a Gram stain of unspun urine is highly suggestive of significant bacteriuria; however, a negative Gram stain does not rule out infection. The Gram stain has excellent sensitivity for demonstrating bacteriuria when the urine colony count exceeds 10^5 organisms/ml but has poor sensitivity at lower colony counts.

Management

ANTIMICROBIAL THERAPY

In general, antimicrobial therapy is warranted for any symptomatic infection of the urinary tract. The choice of antimicrobial agent, dose, and duration of therapy depends on the site of infection and the presence or absence of complicating conditions. Therefore, each category of UTI merits a different approach on the basis of the particular clinical syndrome that is present.

Table 2 Common Causes of Acute Dysuria in Women[68]

Condition	*Pathogen*	*Laboratory Findings*			*Symptoms, Onset, and Factors*
		Pyuria	*Hematuria*	*Urine Culture (cfu/ml)*	
Cystitis	*Escherichia coli* (most common), *Staphylococcus saprophyticus, Proteus* species, *Klebsiella* species	Usual	Often	10^2 to ≥ 10^5	Abrupt onset, multiple symptoms (dysuria, increased frequency, and urgency), suprapubic or low back pain, suprapubic tenderness on examination
Urethritis	*Chlamydia trachomatis, Neisseria gonorrhoeae,* herpes simplex virus	Usual	Rare	< 10^2	Gradual onset, mild symptoms, vaginal discharge or bleeding (caused by concomitant cervicitis), lower abdominal pain, new sexual partner, cervicitis or vulvovaginal herpetic lesions on examination
Vaginitis	*Candida* species, *Trichomonas vaginalis*	Rare	Rare	< 10^2	Vaginal discharge or odor, pruritus, dyspareunia, external dysuria, no increased frequency or urgency, vulvovaginitis on examination

Treatment of Women with Acute Uncomplicated Cystitis

Antimicrobial therapy for healthy reproductive-age women with uncomplicated UTI should have the dual objective of eradicating the infection and eliminating uropathogenic clones of bacteria from the vaginal and GI reservoirs to prevent early recurrences. In general, the species and antimicrobial susceptibilities of the bacteria that cause acute uncomplicated cystitis are highly predictable. Thus, in otherwise healthy women presenting with typical symptoms of acute cystitis (dysuria and frequency without signs or symptoms of vaginitis), it is safe and cost-effective to omit the urine culture and use empirical short-course therapy. Three-day therapy seems to be optimal; single-dose therapy results in higher relapse rates (probably because of failure to eradicate the uropathogen from the vaginal reservoir), and 7-day therapy offers no additional benefit but costs more and causes more side effects. Therapy for 7 days, however, should be considered in women with a history of recent UTI, symptoms of more than 7 days' duration, or diabetes.[29]

Several effective therapeutic regimens for acute uncomplicated cystitis in women are available [*see Table 3*]. Traditionally, trimethoprim-sulfamethoxazole (TMP-SMX) has been recommended as first-line treatment.[30,31] However, resistance of uropathogens to TMP-SMX is increasing and now approaches 15% to 20% in some communities.[32,33] Because in vitro TMP-SMX resistance has been correlated with clinical and bacteriologic failure of TMP-SMX therapy, factors other than the expected prevalence of TMP-SMX resistance should also be considered. These include a history of recent use of TMP-SMX or another antimicrobial and recent travel to an area with high TMP-SMX–resistance rates.[33]

In such settings, an alternative first-line agent should be considered for empirical therapy.[29,32,33] Many strains of *E. coli* that are resistant to TMP-SMX are also resistant to amoxicillin and cephalexin; thus, these drugs should be used only in patients infected with susceptible strains. Other β-lactams, such as amoxicillin-clavulanate or cefpodoxime, can be used but are expensive and may be associated with higher relapse rates, probably because they fail to eradicate the uropathogen from the vaginal reservoir. Nitrofurantoin remains highly active against *E. coli* and most non–*E. coli* isolates (except for *Proteus* species, which are intrinsically resistant to the drug); however, 7 days of therapy may be required to achieve reasonable cure rates with this drug. Most fluoroquinolones are effective for short-course therapy of cystitis. In a randomized trial, 3-day regimens of ciprofloxacin (100 mg twice daily), double-strength TMP-SMX (160/800 mg twice daily), or ofloxacin (200 mg twice daily) were equally effective in the treatment of acute uncomplicated cystitis in women.[34] In separate studies, 3-day regimens of TMP-SMX and ofloxacin were found to be equally cost-effective.[35,36] Higher costs were associated with 3-day regimens of ampicillin, cefadroxil, and nitrofurantoin because of lower efficacy and higher rates of side effects.[35,36] However, increasing fluoroquinolone use has been correlated with increased rates of fluoroquinolone resistance, which suggests that patients with uncomplicated cystitis should be treated with an agent other than a fluoroquinolone, if possible.[37]

If a patient is still symptomatic after therapy, both urinalysis and urine culture are necessary [*see Figure 2*]. If the urinalysis and culture are negative, a 2-day course of the urinary tract analgesic phenazopyridine, 200 mg three times daily after meals, can be prescribed.[38] A pelvic exam for evaluation of alternative diagnoses such as chlamydial, gonococcal, or herpetic infection should be considered, and close clinical follow-up is recommended.[24] If testing shows pyuria but not bacteriuria, pelvic examination for alternative diagnoses should be performed. If the patient has both pyuria and bacteriuria, the antimicrobial susceptibility of the infecting strain should be assessed for resistance and an alternative agent should be given. Finally, a patient who is symptomatic after a short-course regimen and has persistent infection with a uropathogen that is sensitive to the antibiotic used should be regarded as having covert renal infection. In this circumstance, a 14-day course of a fluoroquinolone or TMP-SMX is indicated.[4]

Table 3 Treatment Regimens for Acute Uncomplicated Cystitis[4*]

Host Considerations	*Empirical Treatment Regimens*
Otherwise healthy woman	*Three-Day Regimens* TMP-SMX , 160/800 mg q. 12 hr Trimethoprim, 100 mg q. 12 hr Fluoroquinolones: Ciprofloxacin, 100–250 mg q. 12 hr Ciprofloxacin XR, 500 mg q.d. Gatifloxacin, 200 mg q.d. Levofloxacin, 250 mg q.d. *Five- to Seven-Day Regimens* Nitrofurantoin monohydrate/macrocrystals, 100 mg q. 12 hr Nitrofurantoin macrocrystals, 50–100 mg q.i.d. Amoxicillin, 250 mg q. 8 hr or 500 mg q. 12 hr Cephalexin, 250 mg q. 6 hr, or other cephalosporin
Male sex, diabetes, symptoms for 7 days, recent antimicrobial use, age > 65 yr	*Consider Seven-Day Regimen* TMP-SMX , 160/800 mg q. 12 hr Fluoroquinolones, as per 3-day regimen Cephalexin, 250 mg q. 6 hr, or other cephalosporin
Pregnancy	*Consider Seven-Day Regimen* Amoxicillin, 250 mg q. 8 hr or 500 mg q. 12 hr Nitrofurantoin monohydrate/macrocrystals, 100 mg q. 12 hr Nitrofurantoin macrocrystals, 50–100 mg q.i.d. Cephalexin, 250 mg q. 6 hr, or other cephalosporin TMP-SMX , 160/800 mg q. 12 hr‡

* Characteristic pathogens are *Escherichia coli* (85% to 90%) and *Staphylococcus saprophyticus* (5% to 15%); other organisms, which account for < 5% of cases, are *Proteus mirabilis*, *Klebsiella pneumoniae*, and *Enterococcus* species.

† Treatments listed are those to be prescribed before the etiologic agent is known (Gram stain can be helpful); regimens can be modified once the agent has been identified. The recommendations are limited to drugs currently approved by the Food and Drug Administration, although not all the regimens listed are approved for these indications. Optimal empirical regimens may differ among settings because of differences in the antimicrobial susceptibility profiles of uropathogens. Fluoroquinolones should not be used in pregnancy.

‡ Although TMP-SMX is classified as pregnancy category C, it is widely used; however, avoid use of this drug in the first and third trimesters of pregnancy.

TMP-SMX—trimethoprim-sulfamethoxazole

Treatment of postmenopausal women The antimicrobial approach for postmenopausal women with symptomatic cystitis is similar to that for younger women—namely, short-course therapy with TMP-SMX or a fluoroquinolone initially; longer courses of therapy should be reserved for patients who do not

respond to short-course therapy. In a randomized clinical trial of the treatment of acute cystitis in postmenopausal women, 3-day therapy with ofloxacin was more cost-effective than 7-day therapy with cephalexin.[39] In another randomized trial, 3 days of ciprofloxacin for treatment of cystitis in women 65 years of age and older was equivalent to 7 days of ciprofloxacin and associated with fewer adverse effects.[40] The major difference in the management of older women with UTI is the recognition that topical estrogen replacement, in the form of vaginal estriol cream, may decrease the incidence of recurrent UTI in postmenopausal women.[17] Some studies of postmenopausal women have not found a protective effect with topical estrogen, particularly when it is delivered by pessary. The effects of systemic estrogen replacement on risk of UTI have not been studied in a randomized, controlled trial, but in general, a protective effect has not been demonstrated with this intervention.[41,42]

Treatment of recurrent UTI Recurrence of uncomplicated cystitis in reproductive-age women is common, and some form of preventive strategy is indicated if three or more symptomatic episodes occur in 1 year. A variety of antimicrobial strategies are available, but before embarking on one of them, the patient should try such simple interventions as voiding immediately after sexual intercourse and using a contraceptive method other than a diaphragm and spermicide. There is little evidence to support the former approach, but it is still often recommended, given the difficulty in adequately studying this behavior and the low morbidity associated with implementing it. Ingestion of cranberry juice has been shown to be effective in decreasing bacteriuria with pyuria, but not bacteriuria alone or symptomatic UTI, in an elderly population.[43] Cranberry juice was demonstrated to reduce the rate of recurrent UTI in younger women when combined with lingonberry.[44] A more recent randomized trial comparing placebo, cranberry juice, and cranberry tablets also demonstrated a reduction in the UTI rate in young women from 32% in the placebo group to 20% with juice and 18% with tablets.[45] Thus, although the efficacy of cranberry juice for prevention of UTI needs further evaluation, there is mounting evidence that it may be effective in young otherwise healthy women.[46] This effect appears to be independent of any urinary acidification; rather, it is postulated that cranberry juice contains substances that inhibit the attachment of bacterial adhesins to the uroepithelium.[47]

If simple nondrug measures are ineffective, continuous or postcoital—if the infections are temporally related to intercourse—low-dose antimicrobial prophylaxis with one of several regimens should be considered. A single dose of TMP-SMX (one half of a single-strength tablet, which amounts to 40 mg of trimethoprim and 200 mg of sulfamethoxazole), a fluoroquinolone (one tablet), or nitrofurantoin (50 mg; or 100 mg of nitrofurantoin macrocrystals) can safely and effectively decrease the rate of recurrent infections.[18] Typically, a prophylactic regimen is initially prescribed for 6 months and then discontinued. If the infections recur, the prophylactic program can be instituted for a longer period. Antimicrobial prophylaxis has been effectively used for as long as 5 years in preventing recurrence of infection.[8]

An alternative approach to antimicrobial prophylaxis for women with less frequent recurrences (fewer than four a year) is to supply the patient with TMP-SMX or a fluoroquinolone and allow her to self-medicate with short-course therapy at the first symptoms of infection. The patient is directed to keep track of the number of such episodes and to contact the physician if more than four episodes occur over a 12-month period or if symptoms persist on such therapy. This approach has been shown to be safe and effective in two separate studies of women with recurrent UTI.[48,49]

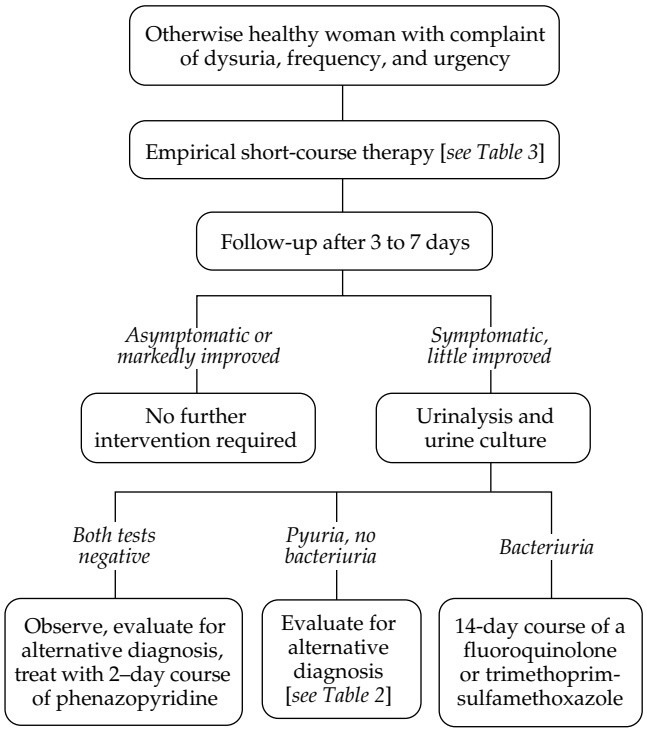


Figure 2 **Clinical approach to acute uncomplicated cystitis in a woman.**

Treatment of relapsing infection The approach to the minority of patients with relapsing infection, as evidenced by finding the same bacterial strain in a UTI that occurs within 2 weeks after completion of antimicrobial therapy, is very different from the management of reinfection. Two factors may contribute to the pathogenesis of relapsing infection in women: deep-tissue infection of the kidney that is suppressed but not eradicated by a 14-day course of antibiotics, and structural abnormality of the urinary tract, particularly calculi. Patients with true relapsing UTIs should undergo radiologic or urologic evaluation and should be considered for longer-term therapy.

Acute Uncomplicated Pyelonephritis

Patients with clear-cut symptomatic pyelonephritis have deep-tissue infection, have (or are at risk for) bacteremia, and merit antimicrobial therapy. Two requirements guide the initial choice of antimicrobial regimens for pyelonephritis: the probability that the infecting organism is sensitive to the regimen should be at least 99%, and therapeutic blood levels should be quickly achievable. Depending on the severity of illness and the presence of comorbid conditions, pyelonephritis can be initially managed with oral outpatient therapy or with parenteral inpatient therapy. Patients with mild disease (low-grade fever and no signs of sepsis) who are otherwise healthy and do not have significant nausea or vomiting can be managed as outpatients with an oral fluoroquinolone or TMP-SMX [*see Figure 3 and Table 4*].[4,29,50] A randomized clinical trial demonstrated that 7 days of therapy with oral ciprofloxacin (with or without an initial intra-

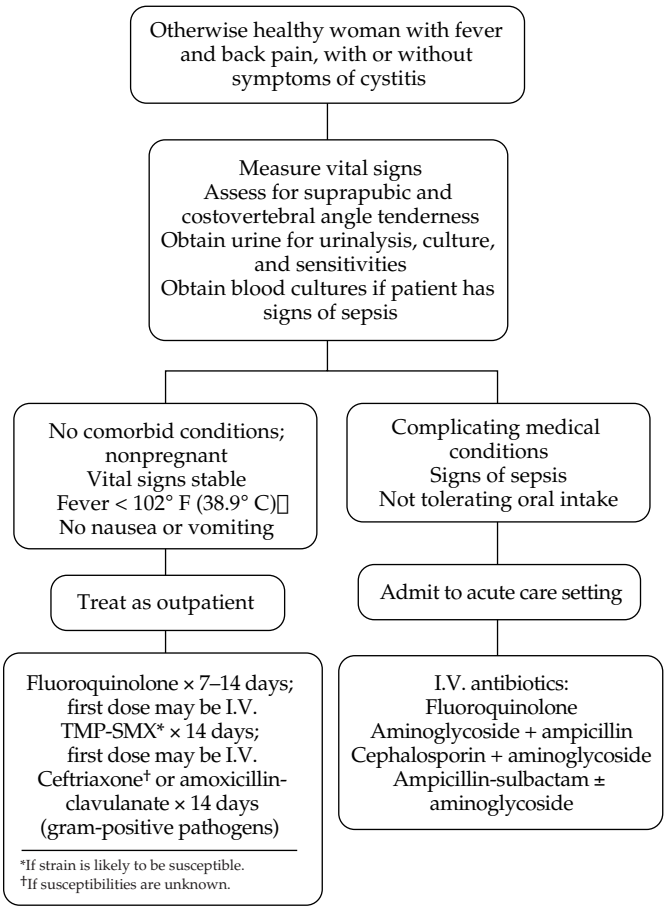


Figure 3 **Clinical approach to acute uncomplicated pyelonephritis in a woman.**

venous dose of the drug) was highly effective for the initial management of pyelonephritis in the outpatient setting.[50] Oral TMP-SMX is also very effective but should not be used unless the prevalence of TMP-SMX resistance in the area is very low or the strain is known to be susceptible.[29] An initial I.V. dose of ceftriaxone should be considered when oral TMP-SMX is being used, because in one study, patients receiving this combination regimen had high success rates even when the infecting strain was resistant to TMP-SMX.[51] Amoxicillin-clavulanate should be considered if the Gram stain suggests enterococci. Regardless of which outpatient oral regimen is chosen, an initial I.V. dose of a fluoroquinolone or a third-generation cephalosporin in an observed setting may be of benefit.[33]

Patients who are severely ill, cannot tolerate oral medication, or have complicating medical conditions should be hospitalized for parenteral therapy. Various regimens can be used, such as ampicillin plus gentamicin; fluoroquinolones; the third-generation cephalosporins (e.g., ceftriaxone); or, for susceptible strains, TMP-SMX. Combinations of a β-lactam and a β-lactamase inhibitor (e.g., ampicillin-sulbactam, ticarcillin-clavulanate, and piperacillin-tazobactam) or imipenem-cilastatin can be used in patients with more complicated histories, previous episodes of pyelonephritis, or recent urinary tract manipulations. After the patient has been afebrile for 24 hours (usually within 72 hours of initiating therapy), there is no benefit in maintaining parenteral therapy. Currently, prescription of oral TMP-SMX or a fluoroquinolone to complete a 14-day course of outpatient therapy appears to be the most effective means of eradicating both tissue infection and residual clones of uropathogens present in the GI tract that could cause early recurrence if allowed to remain.[29]

UTI during Pregnancy

There are three major differences between the approach to UTI in pregnant women and that in nonpregnant women. First, in pregnant women, asymptomatic bacteriuria is actively sought and is as aggressively treated and followed as symptomatic infection; this is clearly not the case in nonpregnant women, for whom screening for asymptomatic bacteriuria is not recommended. Second, although short-course therapy is also the cornerstone of treatment during pregnancy for patients with uncomplicated cystitis (as well as those with asymptomatic bacteriuria), the drugs that can be safely used are far more limited for pregnant women. Third, follow-up of patients with bacteriuria during pregnancy is more intense, with a more rapid deployment of prophylactic strategies in pregnant women with recurrent bacteriuria.[11]

Nitrofurantoin, ampicillin, and the cephalosporins have been considered relatively safe in early pregnancy. Sulfonamides should be avoided in the first trimester because of possible teratogenic effects and avoided near term because of a possible role in the development of kernicterus. Trimethoprim is usually avoided because of evidence of fetal toxicity at high doses in animals, although it has been used successfully in humans during pregnancy without evidence of toxicity or teratogenicity. Fluoroquinolones are avoided because of possible adverse effects on fetal cartilage development. Nitrofurantoin, ampicillin, and the cephalosporins have been used most extensively in pregnancy and are the regimens of choice for the treatment of asymptomatic or minimally symptomatic UTI [*see Table 3*]. For pregnant women with overt pyelonephritis, admission to the hospital for parenteral therapy should be the standard of care; β-lactams with or without aminoglycosides are the cornerstone of therapy.[11,51]

Prevention of UTI, including pyelonephritis, can be accomplished during pregnancy with nitrofurantoin or cephalexin taken prophylactically after coitus or at bedtime without relation to coitus. Such prophylaxis should be considered for patients who have had acute pyelonephritis during pregnancy, patients with bacteriuria during pregnancy who have had a recurrence after a course of treatment, and patients who had recurrent UTI before pregnancy that required prophylaxis.[11,52]

Treatment of Men with UTI

It is uncommon for men to have UTI that is analogous to acute uncomplicated cystitis in women. Even when seemingly uncomplicated UTI does occur, men should never be treated with short-course therapy, because of a high rate of early relapse. Instead, 7- to 14-day regimens of a fluoroquinolone should be prescribed. TMP-SMX is the best alternative drug, assuming susceptibility of the strain.[53]

In men older than 50 years with UTI, bacterial invasion of the prostate and possibly the kidneys should be considered, even in the absence of overt signs of infection at these sites. Acute bacterial prostatitis should be treated with a fluoroquinolone for 2 weeks or with TMP-SMX for at least 4 weeks. Recurrence is common and usually connotes a sustained focus in the prostate that has not been eradicated. Several factors make eradication of prostatic foci difficult. Many antimicrobial agents do not diffuse well across the prostatic epithelium into the prostatic fluid,

where the infection lies. The prostate may harbor calculi, which can block drainage of portions of the prostate gland or act as foreign bodies in which persistent infection can reside. An enlarged (and inflamed) prostate gland can cause bladder outlet obstruction, resulting in pools of stagnant urine in the bladder that are difficult to sterilize.[28]

In view of these factors, intensive therapy for at least 4 to 6 weeks is recommended for chronic bacterial prostatitis. The drugs of choice for this purpose, assuming the organisms are susceptible, are the fluoroquinolones. The best alternative agent is TMP-SMX. Prolonged treatment with any of these drugs has a greater than 60% probability of eradicating infection. Most therapeutic failures result from either anatomic factors or infection by *Enterococcus faecalis* or *P. aeruginosa;* these two organisms are particularly likely to cause relapse after treatment with the antimicrobial agents currently recommended. Relapses should be treated for 12 weeks. If this therapy fails, long-term antimicrobial suppression or repeated treatment courses for each relapse are often needed.[28]

Complicated UTI

Complicated UTI occurs in a heterogeneous group of patients with a wide variety of structural and functional abnormalities of the urinary tract and kidneys. The range of antimicrobial species and their susceptibility to antimicrobial agents are likewise heterogeneous. As a consequence, therapy for patients with complicated UTI requires individualization, although the following guidelines appear to be useful[54]:

- As a rule, only symptomatic UTI requires therapy.
- If symptoms of UTI are present, a urine culture and susceptibilities should always be obtained.
- If the antimicrobial susceptibilities of the infecting organism are not known and symptomatic infection requires immediate therapy, consideration of previous microbiology or recent antimicrobial exposures can help guide initial empirical therapy.
- In patients with mild disease [*see* Pyelonephritis, *above*]), an oral regimen with a fluoroquinolone, or possibly TMP-SMX, is appropriate and can be given in the outpatient or inpatient setting depending on other patient factors.
- In patients who cannot take oral regimens or who have more severe disease, intravenous therapy with fluoroquinolones or broad-spectrum agents such as ampicillin plus gentamicin, imipenem-cilastatin, or piperacillin-tazobactam should be considered until the susceptibilities of the invading organism are identified [*see Table 4*].
- Whenever possible, every effort should be made to correct the underlying complicating factor in conjunction with the antimicrobial therapy.

Asymptomatic Bacteriuria

Bacteriuria detected in the absence of symptoms referable to the urinary tract does not warrant antimicrobial therapy except in specific settings. These include during pregnancy, before surgery or instrumentation of the urinary tract, and after renal transplantation. Treatment of asymptomatic bacteriuria in patients who are immunosuppressed because of transplantation other than renal (i.e., other solid organ or bone marrow) or because of neutropenia has not been well studied and is not currently recommended as standard practice.

In women with diabetes and asymptomatic bacteriuria, a large randomized trial of antimicrobial treatment versus no antimicrobials found no difference in the time to first symptomatic UTI between the groups. The authors conclude that there is no benefit to screening for asymptomatic bacteriuria or treating it in women with diabetes.[55]

RADIOLOGIC EVALUATION TO RULE OUT FUNCTIONAL ABNORMALITIES

A final consideration in the management of UTI is the role of

Table 4 Treatment Regimens for Acute Pyelonephritis[4]

Patient Status	*Regimen*	*Comment*
Outpatient	Fluoroquinolones × 7–14 days*	Fluoroquinolones can be given orally throughout the entire course, if patient can tolerate oral intake
	Ciprofloxacin, 250–500 mg q. 12 hr	
	Ciprofloxacin XR, 1,000 mg q.d.	
	Gatifloxacin, 400 mg q.d.	
	Levofloxacin, 250–500 mg q.d.	
	TMP-SMX, 160/800 mg q. 12 hr × 14 days	If strain is likely to be susceptible
	Amoxicillin, 500 mg q. 8 hr or 875 mg q. 12 hr	
	Amoxicillin-clavulanate, 875/125 mg q. 12 hr × 14 days	If *Enterococcus* is suspected
Inpatient	I.V. fluoroquinolones × 14 days	Patients can be switched to oral therapy after they have been afebrile for 24 hr
	Ciprofloxacin, 200–400 mg q. 12 hr	Particularly useful if enterococcal infection is suspected
	Gatifloxacin, 400 mg q.d.	
	Levofloxacin, 250–500 mg q.d.	
	Ceftriaxone,* 1–2 g I.V. q. 24 hr, or other third- or fourth-generation cephalosporin	
	Ampicillin, 1 g I.V q. 6 hr *plus* gentamicin, 1 mg/kg I.V. q. 8 hr or 3–5 mg/kg I.V. q. 24 hr	
	Ampicillin-sulbactam	
	Ticarcillin-clavulanate	
	Piperacillin-tazobactam	Useful in complicated cases
	Imipenem-cilastatin	

*First dose of a fluoroquinolone or ceftriaxone may be given intravenously.
TMP-SMX—trimethoprim-sulfamethoxazole

radiologic evaluation. Intravenous pyelography, ultrasonography, or CT should be carried out expeditiously to rule out obstruction in any patient with acute pyelonephritis who does not respond to an effective antimicrobial agent and in any patient in whom a persistent bacteremia has been demonstrated. Contrast-enhanced helical CT is the radiologic study of choice for imaging and evaluation of renal infections. The study should be performed without contrast when renal calculi are suspected.[26]

For ambulatory patients who have UTI, guidelines are less clear. Radiographic evaluation, usually with intravenous pyelography and voiding cystourethrography, is carried out for delineation of surgically correctable lesions that might predispose patients to recurrent infection or progressive renal disease. Because congenital anatomic anomalies are particularly prevalent in young children who have a first or second UTI, such studies are obligatory for patients in this age group. In addition, careful prostatic examination and assessment of postvoiding residual urine volume should be considered in males with UTI at any stage of life because such infection is highly unlikely in this population unless anatomic anomalies or specific risk factors are present[55] [*see* Young Men, *above*]. In women with uncomplicated UTI, the incidence of correctable anatomic lesions is so low that radiologic and urologic evaluation should be restricted to patients who have rapid recurrence of infection or recurrent pyelonephritis despite adequate therapy.[26]

Catheter-Associated UTI

UTIs are the most common nosocomial infections, representing 40% of all such infections. Most nosocomial UTIs are related to bladder catheterization. Catheter-associated UTIs are associated with increased mortality and costs.[56,57] Multiple risk factors for catheter-associated UTIs have been identified, including the duration of catheterization, lack of systemic antibiotic therapy, female sex, age older than 50 years, and azotemia.[57] Risk factors for bacteremia related to catheter-associated UTI are not well established.

Several guidelines can be followed to minimize the occurrence of catheter-related infection [*see Table 5*]. Most important, the catheter should be inserted with strict aseptic technique by trained persons. In addition, a closed system should be used at all times. Even with optimal care, however, catheter use for 1 month or longer will eventually result in bladder infection. Apparently, little additional protection is provided by antibiotic rinses for the bladder, antibiotic ointments applied to the urethral meatus, and instillation of disinfectants such as hydrogen peroxide into the urinary collection bag. Although systemic antibiotics are of no value when the closed system is to be in place for an extended period, antibiotics may be protective when the catheter is in place for only a few days.[56] However, this advantage should be balanced against the risk of selecting for resistant flora if catheterization must be prolonged unexpectedly. In patients who require urethral catheterization for 2 to 10 days, use of a silver alloy–coated urinary catheter may offer some protection against infection.[58-60] In one randomized study of hospitalized patients, the use of a silver alloy–coated urinary catheter resulted in a lower rate of UTIs and a significant cost savings.[59] A nitrofurazone-impregnated catheter also has been shown to decrease the incidence of UTIs caused by gram-negative organisms.[60]

Treatment of catheter-associated UTI depends on the clinical circumstances. Symptomatic patients (e.g., those with fever, chills, dyspnea, and hypotension) require immediate antibiotic therapy [*see* Complicated UTI, *above*]. In addition, it may be useful to remove and replace the urinary catheter if it has been in place for a week or longer. This eliminates difficult-to-eradicate organisms in the biofilm on the catheter. In an asymptomatic patient, therapy should be postponed until the catheter can be removed. Patients with persistent asymptomatic bacteriuria and those with lower urinary tract symptoms who have had the catheter removed respond well to short-course therapy.[61]

Patients with long-term indwelling catheters seldom become symptomatic unless the catheter is obstructed or is eroding through the bladder mucosa. In patients who do become symptomatic, appropriate antibiotics should be administered, and the catheter should be changed. Therapy for asymptomatic catheterized patients leads to the selection of increasingly antibiotic-resistant bacteria.[56,57,60] Thus, although long-term bladder catheterization carries a significant risk of chronic pyelonephritis (10% or more if the catheter is in place for more than 90 days), there is no way to avoid this event other than by catheter removal.[62]

The appearance of *Candida* in the urine is an increasingly common complication of indwelling catheterization, particularly for patients in the intensive care unit, on broad-spectrum antimicrobials, or with underlying diabetes mellitus.[63] *C. albicans* is still the most common isolate, although *C. glabrata* and other non-*albicans* species are also frequently isolated. The clinical presentation can vary from an asymptomatic laboratory finding to sepsis. In asymptomatic patients, removal of the urethral catheter results in resolution of the candiduria in as many as one third of cases. For patients with symptomatic candiduria (fever with or without cystitis symptoms), oral fluconazole, 200 mg/day for 7 to 14 days, has been shown to be highly effective. This regimen was even effective for non-*albicans* species of *Candida* that had reduced susceptibility to fluconazole, possibly because of the high concentrations achieved in the urine. For more severely ill patients, the possibility of pyelonephritis and candidemia should be evaluated, and systemic antifungal therapy with fluconazole, 6 mg/kg/day, or amphotericin, 0.6 mg/kg or more a day, should be instituted.

Table 5 Guidelines for Bladder Catheter Care

Use catheters only when absolutely necessary; remove as soon as possible.
Insert catheters aseptically and maintain by trained personnel only.
A sterile closed drainage system is mandatory. The catheter and drainage tube must never be disconnected except when irrigation is necessary to relieve obstruction. Strict aseptic technique is employed under these circumstances.
Obtain urine for culture by aspirating the catheter with a 21-gauge needle after the catheter is prepared with povidone-iodine.
Maintain downhill, nonobstructed flow, with the collection bag always below the level of the bladder and emptied at frequent intervals.
Replace indwelling catheters when obstruction or concretions are demonstrated.
Strict hand-washing precautions should be observed by staff caring for these patients.
Administer prophylactic antibiotics during catheter insertion and removal to patients with cardiac disease (particularly prosthetic valves) that predisposes to bacterial endocarditis.

Prognosis

Although earlier epidemiologic studies suggested that bacteriuria, whether symptomatic or asymptomatic, was associated with increased mortality, more recent findings suggest that bacteriuria does not itself cause an increase in mortality.[64] Rather, bacteriuria appears to be a marker for poor health, particularly in elderly patients. Not surprisingly, then, antimicrobial therapy for asymptomatic bacteriuria in the elderly is of no demonstrable clinical benefit.[65,66]

The relationships between recurrent UTI, chronic pyelonephritis, and renal insufficiency have been widely studied. In the absence of anatomic abnormalities, recurrent infection in children and adults does not lead to chronic pyelonephritis or to renal failure. Also, infection does not play a primary role in chronic interstitial nephritis; the primary etiologic factors in this condition are analgesic abuse, obstruction, reflux, and toxin exposure. In the presence of underlying renal abnormalities, however, infection as a secondary factor can accelerate renal parenchymal damage. In children, the common combination of reflux, congenital anomalies, and infection appears to lead to significant renal parenchymal damage; moreover, kidneys in children appear to be more susceptible to damage than kidneys in adults.[23]

A retrospective questionnaire study has found that men and women newly diagnosed with renal cell carcinoma were more likely to have a history of cystitis or pyelonephritis sometime in their lifetime and to be smokers than persons without renal cell carcinoma.[67] The risk was greater for males than females. Although the study shows a possible association between UTI and renal cell cancer, data on causality, temporal relatedness, dose effect, and putative mechanisms are lacking.

Kalpana Gupta, M.D., M.P.H., has received grant or research support from Bayer Corp. and Procter & Gamble Pharmaceuticals; is a consultant for Bayer Corp.; and has received speaking honoraria from Bayer Corp., Procter & Gamble Pharmaceuticals, and Ortho-McNeil Pharmaceutical.

Walter E. Stamm, M.D., has received grant or research support and has received speaking honoraria from Procter & Gamble Pharmaceuticals, MedImmune, Ortho-McNeil Pharmaceutical, and Bayer Corp.

References

1. Foxman B: Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Am J Med 113:5S, 2002

2. Hooton TM, Scholes D, Hughes J, et al: A prospective study of risk factors for symptomatic urinary tract infection in young women. N Engl J Med 335:468, 1996

3. Scholes D, Hooton TM, Roberts PL, et al: Factors associated with acute pyelonephritis in healthy adult women. Ann Intern Med (in press)

4. Hooton TM, Stamm WE: Diagnosis and treatment of uncomplicated urinary tract infection. Infect Dis Clin North Am 11:551, 1997

5. To T, Agha M, Dick PT, et al: Cohort study on circumcision of newborn boys and subsequent risk of urinary-tract infection. Lancet 352:1813, 1998

6. Spach DH, Stapleton AE, Stamm WE: Lack of circumcision increases the risk of urinary tract infection in young men. JAMA 267:679, 1992

7. Anderson GG, Palermo JJ, Schilling JD: Intracellular biofilm-like pods in urinary tract infections. Science 301:105, 2003

8. Stamm WE, McKevitt M, Roberts PL, et al: Natural history of recurrent urinary tract infections in women. Rev Infect Dis 13:77, 1991

9. Scholes D, Hooton TM, Roberts PL, et al: Risk factors for recurrent urinary tract infection in young women. J Infect Dis 182:1177, 2000

10. Hooton TM, Scholes D, Stapleton AE, et al: A prospective study of asymptomatic bacteriuria in sexually active young women. N Engl J Med 343:992, 2000

11. Patterson TF, Andriole VT: Detection, significance, and therapy of bacteriuria in pregnancy: update in the managed health care era. Infect Dis Clin North Am 11:593, 1997

12. Smaill F: Antibiotics for asymptomatic bacteriuria in pregnancy. Cochrane Database Syst Rev (2):CD000490, 2001

13. Stapleton A: Urinary tract infections in patients with diabetes. Am J Med 113:80S, 2002

14. Geerlings SE, Brouwer EC, Van Kessel KC, et al: Cytokine secretion is impaired in women with diabetes mellitus. Eur J Clin Invest 30:995, 2000

15. Hooton TM, Roberts PL, Stamm WE: Effects of recent sexual activity and use of a diaphragm on the vaginal microflora. Clin Infect Dis 19:274, 1994

16. Pabich WL, Fihn SD, Stamm WE, et al: Prevalence and determinants of vaginal flora alterations in postmenopausal women. J Infect Dis 188:1054, 2003

17. Raz R, Stamm WE: A controlled trial of intravaginal estriol in postmenopausal women with recurrent urinary tract infections. N Engl J Med 329:753, 1993

18. Stapleton A, Stamm WE: Prevention of urinary tract infection. Infect Dis Clin North Am 11:719, 1997

19. Stapleton A, Hooton TM, Fennell C, et al: Effect of secretor status on vaginal and rectal colonization with fimbriated *Escherichia coli* in women with and without recurrent urinary tract infection. J Infect Dis 171:717, 1995

20. Mulvey MA, Schilling JD, Martinez JJ, et al: Bad bugs and beleaguered bladders: interplay between uropathogenic *Escherichia coli* and innate host defenses. Proc Natl Acad Sci USA 97:8829, 2000

21. Johnson JR: Microbial virulence determinants and the pathogenesis of urinary tract infection. Infect Dis Clin North Am 17:261, 2003

22. Otto G, Sandberg T, Marklund BI, et al: Virulence factors and pap genotype in *Escherichia coli* isolates from women with acute pyelonephritis, with or without bacteremia. Clin Infect Dis 17:448, 1993

23. Riccabona M, Fotter R: Urinary tract infection in infants and children: an update with special regard to the changing role of reflux. Eur Radiol Jan 30 (Epub ahead of print), 2004

24. Bent S, Nallamothu BK, Simel DL, et al: Does this woman have an acute uncomplicated urinary tract infection? JAMA 287:2701, 2002

25. Pinson AG, Philbrick JT, Lindbeck GH: Fever in the clinical diagnosis of acute pyelonephritis. Am J Emerg Med 15:148, 1997

26. Browne RF, Zwirewich C, Torreggiani WC: Imaging of urinary tract infection in the adult. Eur Radiol Jan 29 (Epub ahead of print), 2004

27. Schaeffer AJ, Datta NS, Fowler JE, et al: Overview summary statement: diagnosis and management of chronic prostatitis/chronic pelvic pain syndrome. Urology 60(suppl 6A):1, 2002

28. Nickel JC: Recommendations for the evaluation of patients with prostatitis. World J Urol 21:75, 2003

29. Warren JW, Abrutyn E, Hebel JR, et al: Guidelines for antimicrobial treatment of uncomplicated acute bacterial cystitis and acute pyelonephritis in women. Infectious Diseases Society of America (IDSA). Clin Infect Dis 29:745, 1999

30. Saint S, Scholes D, Fihn SD, et al: The effectiveness of a clinical practice guideline for the management of presumed uncomplicated urinary tract infection in women. Am J Med 106:636, 1999

31. Huang ES, Stafford RS: National patterns in the treatment of urinary tract infections in women by ambulatory care physicians. Arch Intern Med 162:41, 2002

32. Manges AR, Johnson JR, Foxman B, et al: Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. N Engl J Med 345:1007, 2001

33. Gupta K, Hooton TM, Stamm WE: Increasing antimicrobial resistance and the management of uncomplicated community-acquired urinary tract infections. Ann Intern Med 135:41, 2001

34. McCarty JM, Richard G, Huck W, et al: A randomized trial of short-course ciprofloxacin, ofloxacin, or trimethoprim-sulfamethoxazole for the treatment of acute urinary tract infection in women. Am J Med 106:292, 1999

35. Hooton TM, Johnson C, Winter C, et al: Single-dose and three-day regimens of ofloxacin versus trimethoprim-sulfamethoxazole for acute cystitis in women. Antimicrob Agents Chemother 35:1479, 1991

36. Hooton TM, Winter C, Tiu F, et al: Randomized comparative trial and cost analysis of 3-day antimicrobial regimens for treatment of acute cystitis in women. JAMA 273:41, 1995

37. Killgore KM, March KL, Guglielmo BJ: Risk factors for community-acquired ciprofloxacin-resistant *E. coli* urinary tract infection. Ann Pharmacother 38:1148, 2004

38. Gaines KK: Phenazopyridine hydrochloride: the use and abuse of an old standby for UTI. Urol Nurs 24:207, 2004

39. Raz R, Rozenfeld S: 3-day course of ofloxacin versus cephalexin in the treatment of urinary tract infections in postmenopausal women. Antimicrob Agents Chemother 40:2200, 1996

40. Vogel T, Verreault R, Gourdeau M, et al: Optimal duration of antibiotic therapy for uncomplicated urinary tract infection in older women. CMAJ 170:469, 2004

41. Jackson SL, Boyko EJ, Scholes D, et al: What predicts urinary tract infection after menopause? A prospective study. Am J Med (in press)

42. Raz R, Colodner R, Rohana Y, et al: Effectiveness of estriol-containing vaginal pessaries and nitrofurantoin macrocrystal therapy in the prevention of recurrent urinary tract infection in postmenopausal women. Clin Infect Dis 36:1362, 2003

43. Avorn J, Monane M, Gurwitz JH, et al: Reduction of bacteriuria and pyuria after ingestion of cranberry juice. JAMA 271:751, 1994

44. Kontiokari T, Sundqvist K, Nuutinen M, et al: Randomised trial of cranberry-lingonberry juice and *Lactobacillus* GG drink for the prevention of urinary tract infections in women. BMJ 322:1571, 2001

45. Stothers L: A randomized trial to evaluate effectiveness and cost effectiveness of naturopathic cranberry products as prophylaxis against urinary tract infection in women. Can J Urol 9:1558, 2002

46. Jepson RG, Mihaljevic L, Craig J: Cranberries for preventing urinary tract infections. Cochrane Database Syst Rev (1):CD0013213, 2004

47. Howell AB, Vorsa N, Der Marderosian A, et al: Inhibition of the adherence of *Es-*

cherichia coli to uroepithelial-cell surfaces by proanthocyanidin extracts from cranberries. N Engl J Med 339:1085, 1998

48. Gupta K, Hooton TM, Roberts PL, et al: Patient-initiated treatment of uncomplicated recurrent urinary tract infections in young women. Ann Intern Med 135:9, 2001

49. Schaeffer AJ, Stuppy BA: Efficacy and safety of self-start therapy in women with recurrent urinary tract infections. J Urol 161:207, 1999

50. Talan DA, Stamm WE, Hooton TM, et al: Comparison of ciprofloxacin (7 days) and trimethoprim-sulfamethoxazole (14 days) for acute uncomplicated pyelonephritis in women: a randomized trial. JAMA 283:1583, 2000

51. Wing DA, Hendershott CM, Debuque L, et al: A randomized trial of three antibiotic regimens for the treatment of pyelonephritis in pregnancy. Obstet Gynecol 92:249, 1998

52. Sandberg T, Brorson JE: Efficacy of long-term antimicrobial prophylaxis after acute pyelonephritis in pregnancy. Scand J Infect Dis 23:221, 1991

53. Lipsky BA: Prostatitis and urinary tract infection in men: what's new; what's true? Am J Med 106:327, 1999

54. Rubenstein JN, Schaeffer AJ: Managing complicated urinary tract infections: the urologic view. Infect Dis Clin North Am 17:333, 2003

55. Harding GKM, Zhanel GG, Nicolle LE, et al: Antimicrobial treatment in diabetic women with asymptomatic bacteriuria. N Engl J Med 347:1576, 2002

56. Warren JW: Catheter-associated urinary tract infections. Internat J Antimicrob Agents 17:299, 2001

57. Saint S, Lipsky BA: Preventing catheter related bacteriuria: Should we? Can we? How? Arch Intern Med 159:800, 1999

58. Saint S, Veenstra DL, Sullivan SD, et al: The potential clinical and economic benefits of silver alloy urinary catheters in preventing urinary tract infection. Arch Intern Med 160:2670, 2000

59. Karchmer TB, Giannetta ET, Muto CA, et al: A randomized crossover study of silver-coated urinary catheters in hospitalized patients. Arch Intern Med 160:3294, 2000

60. Maki DG, Tambyah PA: Engineering out the risk for infection with urinary catheters. Emerg Infect Dis 7:342, 2001

61. Harding GM, Nicolle L, Ronald AR, et al: How long should catheter-acquired urinary tract infection in women be treated? A randomized controlled study. Ann Intern Med 114:713, 1991

62. Warren JW, Muncie HL Jr, Hebel JR, et al: Long-term urethral catheterization increases risk of chronic pyelonephritis and renal inflammation. J Am Geriatr Soc 42:1286, 1994

63. Lundstrom T, Sobel J: Nosocomial candiduria: a review. Clin Infect Dis 32:1602, 2001

64. Bengtsson C, Bengtsson U, Bjorkelund C, et al: Bacteriuria in a population sample of women: 24-year follow-up study: results from the prospective population-based study of women in Gothenburg, Sweden. Scand J Urol Nephrol 32:284, 1998

65. Abrutyn E, Mossey J, Berlin JA, et al: Does asymptomatic bacteriuria predict mortality and does antimicrobial treatment reduce mortality in elderly ambulatory women? Ann Intern Med 120:827, 1994

66. Abrutyn E, Berlin J, Mossey J, et al: Does treatment of asymptomatic bacteriuria in older ambulatory women reduce subsequent symptoms of urinary tract infection? J Am Geriatr Soc 44:293, 1996

67. Parker AS, Cerhan JR, Lynch CF, et al: History of urinary tract infection and risk of renal cell carcinoma. Am J Epidemiol 159:42, 2004

68. Foxman B, Gillespie B, Koopman J, et al: Risk factors for second urinary tract infection among college women. Am J Epidemiol 151:1194, 2000

Acknowledgment

Figure 1 Seward Hung.

XXIV HYPERTHERMIA, FEVER, AND FEVER OF UNDETERMINED ORIGIN

Harvey B. Simon, m.d., f.a.c.p.

Thermoregulation and Abnormal Elevation of Body Temperature

REGULATION OF BODY TEMPERATURE

Like so many biologic functions, human body temperature normally displays circadian rhythmicity, often rising from values of 36.1° C (97.0° F) or lower in the predawn hours to 37.4° C (99.3° F) or higher in the afternoon. This diurnal flux has two important practical consequences. First, fever associated with disease states is superimposed on the normal cycle and tends to peak in the evening; hence, a patient cannot be considered afebrile until his or her temperature has been monitored for at least 24 hours. Second, temperatures exceeding what is mistakenly regarded as the normal value of 37.0° C (98.6° F) are often recorded in perfectly healthy individuals.[1] Unfortunately, many patients with temperature elevations that are entirely physiologic have been subjected to potentially hazardous tests and treatments because their elevated temperatures were incorrectly regarded as pathologic.

Within the limits of the circadian rhythm, however, body temperature is closely regulated by homeostatic mechanisms that strike a balance between heat production and heat dissipation.[2] Heat is a by-product of all metabolic processes. At rest, metabolic activity in the liver and heart produces much of the body's heat; metabolic activity in skeletal muscle accounts for the greatly enhanced thermal load of exercise. Heat is dissipated at the body's surfaces; the skin accounts for about 90% of heat loss, with the lungs contributing most of the remaining 10%. In the basal state, about 70% of the body's thermal load is dissipated by conduction; 30% is removed by the evaporation of insensible perspiration. Radiation and convection are less important mechanisms of heat removal. When the ambient temperature rises or metabolic heat production increases, evaporation accounts for the major share of heat dissipation.

The preoptic nucleus of the anterior hypothalamus functions as the thermal control center and acts to maintain the body temperature at a set value—the so-called hypothalamic thermal set point.[3] In response to elevations in core body temperature (i.e., the temperature of the blood perfusing the internal organs), the hypothalamus stimulates the autonomic nervous system to produce cutaneous vasodilatation and sweating, both of which dissipate heat. In response to either a falling core temperature or a falling skin temperature, the hypothalamus conserves heat by causing cutaneous vasoconstriction. When cold stress is severe, the hypothalamus acts to increase heat production by stimulating muscular activity in the form of shivering; shivering is mediated by the action of somatic nerves, but it is an automatic and involuntary process.

MEASUREMENT OF BODY TEMPERATURE

Body temperature can be measured in several ways. In most clinical circumstances, rectal temperature is an accurate reflection of central (core) temperature. Sublingual (oral) temperature measurements are also useful; they are typically 0.4° C (0.7° F) lower than rectal temperatures[4] and are subject to more technical variation, especially when patients are unable to fully cooperate. Tympanic membrane temperatures are typically 0.4° C (0.7° F) below sublingual readings and have even greater variability.[4,5] Axillary temperature measurements are unreliable[6] but are often used in acutely ill patients.

HYPERTHERMIA AND FEVER

Abnormal elevation of body temperature, or pyrexia, can occur in one of two ways: hyperthermia[7] or fever [*see Figure 1*]. In hyperthermia, thermal control mechanisms fail, so that heat production exceeds heat dissipation. In contrast, in fever, the hypothalamic thermal set point rises, and intact thermal control mechanisms are brought into play to bring body temperature up to the new set point. The distinction between fever and

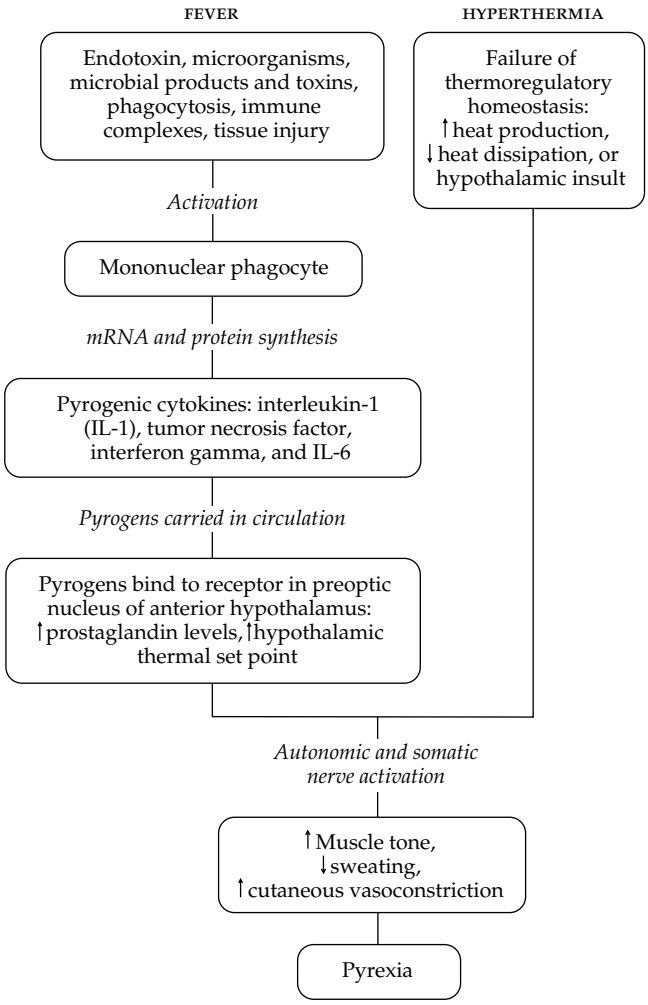


Figure 1 **Body temperature can rise to abnormal levels in one of two ways: hyperthermia or fever. In hyperthermia, thermal control mechanisms fail. In fever, by contrast, the hypothalamic thermal set point rises.**

hyperthermia is more than academic: hyperthermia is best treated with physical cooling methods that promote heat dissipation, whereas fever is best treated with drugs that lower the thermal set point, such as aspirin, other cyclooxygenase inhibitors,[8] or acetaminophen.

Hyperthermia

Hyperthermia is caused by the failure of thermoregulation; the pyrexia associated with it can be mild or severe. The most important causes of severe hyperthermia are heatstroke, neuroleptic syndrome, and malignant hyperthermia of anesthesia.

ETIOLOGY

Numerous clinical disorders can disrupt thermoregulatory homeostasis by causing increased heat production, decreased heat dissipation, or hypothalamic insult, thereby inducing hyperthermia [*see Table 1*]. Mild forms of hyperthermia are common. In dehydration, for example, cutaneous vasoconstriction and cessation of sweating occur in response to the decrease in intravascular volume as a means of conserving further loss of fluid and of minimizing the consequences of fluid loss. As a result, heat dissipation is impaired and body temperature may rise. The hyperthermia of dehydration is usually mild and is readily corrected by fluid replacement. In some cases, however, thermoregulatory disorders cause extreme pyrexia, in which body temperatures rise to 41.1° C (106.0° F) or higher[9]; examples of these thermoregulatory disorders are heatstroke, malignant hyperthermia of anesthesia, thyroid storm, and hypothalamic insult caused by infection, tumor, or drugs [*see 8:I Management of Poisoning and Drug Overdose*].

Numerous exogenous agents, including cocaine[10] and amphetamine, may cause severe hyperthermia. Because salicylates uncouple oxidative phosphorylation in skeletal muscle, excessive heat production is probably responsible for the hyperthermia seen in some children with severe aspirin toxicity.

The hyperthermia that may accompany thyroid storm and pheochromocytoma crisis also results from increased heat production, triggered in these cases by the calorigenic properties of thyroid hormones[11] and catecholamines.[12] Anticholinergic drugs may impair heat dissipation by reducing sweat production.

Hypothalamic disorders are uncommon causes of hyperthermia. Cerebrovascular accidents, however, may be responsible for sustained hyperthermia; this so-called central fever is easily confused with true fever caused by infection or drug hypersensitivity in patients who are hospitalized for major cerebrovascular accidents.

Because infectious and inflammatory diseases can also cause extreme fever, the magnitude of elevation of a patient's temperature cannot be used to distinguish between hyperthermia and fever.[9]

In most cases, hyperthermia is an acute rather than a recurrent problem, and the underlying disorder can be readily diagnosed by careful clinical examination. Exceptions occur in patients with endocrinologic[12] or hypothalamic disorders who present as diagnostic puzzles in the category of fever of undetermined origin (FUO) [*see* Fever of Undetermined Origin, *below*].

HEATSTROKE

Heatstroke is a life-threatening emergency characterized by a body temperature of 40° C (104° F) or higher and neurologic dysfunction that may include delirium, seizures, or coma.[13] Heatstroke occurs in warm, humid conditions: the high ambient temperature impairs heat loss by conduction, and the high ambient humidity limits heat dissipation by the evaporation of sweat. Sustained daily elevations in ambient temperature increase risk.[14] In one form of the syndrome, exertional heatstroke, exercise imposes the additional thermal burden of increased heat production.[15] In classic heatstroke, which is more common than the exertional form, increased heat production does not occur, but heat dissipation is often impaired as a result of a failure of sweating (anhidrosis), as well as of more subtle thermoregulatory problems imposed by underlying diseases and the medications used to treat them.

Table 1 Causes of Hyperthermia

Cause	*Example*
Excessive heat production	Delirium tremens Drug abuse (amphetamines) Exertional hyperthermia Generalized tetanus Heatstroke (exertional)* Lethal catatonia Malignant hyperthermia of anesthesia Neuroleptic malignant syndrome* Pheochromocytoma Salicylate intoxication Serotonin syndrome Status epilepticus Thyrotoxicosis
Diminished heat dissipation	Anticholinergic drugs Autonomic dysfunction Dehydration Heatstroke (classic)* Neuroleptic malignant syndrome* Occlusive dressings Drug abuse (cocaine)
Hypothalamic dysfunction	Cerebrovascular accidents Encephalitis Idiopathic hypothalamic dysfunction Neuroleptic malignant syndrome* Sarcoidosis and granulomatous infections Trauma Tumors

*Pathogenesis of these disorders is mixed.

Epidemiology

Heatstroke is far from rare. Between 1979 and 2002, more than 4,780 deaths in the United States were attributed to excessive heat exposure[16]; it seems likely that many cases of heatstroke go unrecognized or unreported and that hyperthermia also contributes to the morbidity and mortality attributed to underlying diseases.[17]

Classic heatstroke is the most common hyperthermic emergency.[7] Occurring principally during summer heat waves, classic heatstroke is most likely to affect the elderly and patients with serious underlying diseases. The urban poor are particularly vulnerable to classic heatstroke. Women are affected more often than men, and infants are also at risk.

Whereas classic heatstroke occurs in outbreaks, exertional heatstroke occurs sporadically and typically affects young, healthy individuals who engage in strenuous exercise. Exertional

heatstroke is much more common in men than in women. In the United States, military recruits, industrial workers, and athletes such as runners and football players are at greatest risk; in Saudi Arabia, both forms of heatstroke are particularly common during the annual hajj, the pilgrimage to Mecca.

Prevention

Hot, humid conditions are predisposing factors for both classic and exertional heatstroke. In addition to ambient weather conditions and muscular exertion, predisposing factors for exertional heatstroke include inappropriately heavy clothing, exposure to direct sunlight, dehydration,[18] lack of cardiovascular conditioning, lack of acclimatization to heat, and the use of performance-enhancing supplements such as Ma-huang (ephedrine) and creatine.[19] Because all of these factors can be anticipated and corrected, exertional heatstroke is preventable.

Classic heatstroke can be prevented[16] by publicizing community heat-wave alerts; providing cool environments for people at risk; encouraging the use of fans if air conditioners are not available; instructing those at risk to limit physical activity and sun exposure indoors and outdoors, to avoid cooking, to wear lightweight clothing, and to maintain adequate hydration; and promoting judicious use of diuretics, tranquilizers, antidepressants, and anticholinergic medications.

Pathophysiology

The crucial pathophysiologic events causing classic heatstroke are excessive ambient heat and humidity and impaired heat dissipation caused by anhidrosis. The typical victim of classic heatstroke is an elderly or debilitated person who is confined to a poorly ventilated room without benefit of an air conditioner or fan during a summer heat wave. Dehydration, which limits heat dissipation, is an important predisposing condition. Other factors that increase risk include neurologic and cardiovascular diseases, obesity, the use of diuretics, and the use of neuroleptics and other medications with anticholinergic properties; alcohol consumption also contributes to some cases.

The second form of heatstroke is triggered by exercising in a warm, humid environment. The major pathophysiologic mechanism responsible for exertional heatstroke is increased heat production from exercising skeletal muscles. Working at maximal intensity, muscle can increase its energy consumption to levels 20 times its basal rate; because the body's efficiency is only about 25%, much of this energy is converted into heat, which is transferred from muscle to blood, raising the core temperature. Patients who suffer exertional heatstroke do not appear to have intrinsic abnormalities of their thermoregulatory mechanisms, but they may have a high proportion of type II skeletal muscle fibers and lower work efficiency.[20]

Diagnosis

Early diagnosis is critical to prevent life-threatening complications. People who experience weakness or undue fatigue, lack of concentration or confusion, light-headedness or dizziness, headaches, nausea, or muscle cramps in the heat should cease exercise, get to a cool environment, and drink cool fluids to avert more serious heat injury. They should also be sure that medical help is available if symptoms progress.

Clinical features Despite differences in pathophysiology and epidemiology, the major manifestations of the classic and exertional heatstroke syndromes are similar. An abrupt rise in body temperature is universal; most patients with heatstroke have core temperatures in excess of 40.5° C (105° F). Disordered mentation is no less common and may span a spectrum ranging from confusion and lethargy to delirium, stupor, coma, and seizures. Virtually all patients with heatstroke also exhibit cardiovascular abnormalities. Tachycardia is present in most patients, and hypotension is present in many; shock, arrhythmias, myocardial ischemia, and pulmonary edema are most likely to occur early in elderly or debilitated patients with classic heatstroke but are also frequent preterminal events in patients with exertional heatstroke. Most heatstroke victims present with hyperventilation and respiratory alkalosis; pulmonary abnormalities may also include pulmonary edema, acute respiratory distress syndrome, and aspiration pneumonia. Other common clinical manifestations of heatstroke include oliguria, vomiting, and diarrhea; hematuria and gastrointestinal bleeding may reflect disseminated intravascular coagulation (DIC). Cutaneous manifestations of heatstroke may include hemorrhagic lesions; the skin is hot and dry in many patients with classic heatstroke but is often moist and clammy in patients with exertional heatstroke.

Laboratory findings Laboratory abnormalities in patients with heatstroke reflect widespread damage to a number of organ systems. Leukocytosis is common, and hemoconcentration may raise the hematocrit in patients who are volume depleted. Thrombocytopenia is frequently present and may be severe; prolonged prothrombin times, depressed fibrinogen levels, and elevated levels of fibrin split products indicate DIC, which is more common in exertional heatstroke than in classic heatstroke. DIC in heatstroke has been attributed to direct thermal injury to vascular endothelium; clinically significant bleeding may result from the coagulopathy, which can persist for 36 hours after the onset of heatstroke.

Renal abnormalities are common in heatstroke. The urine may sometimes be dilute despite hypotension and oliguria; hematuria, myoglobinuria, proteinuria, and casts are frequently observed. The blood urea nitrogen (BUN) and creatinine levels are usually elevated; acute renal failure occurs in about one third of patients with exertional heatstroke but is less common in patients with classic heatstroke.

Elevated bilirubin levels in patients with heatstroke may reflect hepatic injury, hemolysis, or both. Elevated transaminase levels are the rule, reflecting damage to both liver and muscle cells. Elevated creatine phosphokinase (CPK) and aldolase levels indicate muscle injury; severe rhabdomyolysis with myoglobinuria and renal failure is much more common in exertional than in classic heatstroke.

Heatstroke causes numerous metabolic abnormalities. Respiratory alkalosis is common early in the syndrome, particularly in classic heatstroke; lactic acidosis may supervene, particularly in exertional heatstroke. Potassium levels are variable, sometimes being low early and high later in the course of heatstroke. Hypophosphatemia is the rule; hypokalemia may occur, particularly in exertional heatstroke. Hypoglycemia, perhaps reflecting depletion of glycogen stores, has been reported in many patients with exertional heatstroke. Electrocardiographic abnormalities include conduction disturbances and ST-T segment changes. Elevated levels of pyrogenic cytokines have been reported in heatstroke patients,[21] but it is unclear whether the cytokines participate in the pathogenesis of heatstroke or simply reflect a subsequent host response to tissue injury.[13]

Differential Diagnosis

The differential diagnosis of elevated body temperature with numerous clinical and laboratory abnormalities is broad and includes sepsis, the systemic inflammatory response syndrome, and the other causes of hyperthermia. In practical terms, the major challenge is to distinguish classic heatstroke from sepsis and to distinguish exertional heatstroke from less severe exercise-induced abnormalities, such as exertional hyperthermia and heat exhaustion.[15]

Treatment

Heatstroke necessitates prompt and aggressive therapy. Lowering body temperature is the crucial element in management; because the pathogenesis of heatstroke involves thermoregulatory failure rather than an elevated hypothalamic set point, physical cooling is essential, but antipyretic medications are ineffective.

Field management includes removing the patient's clothing, fanning the patient, and bathing the patient's skin with cool water; ice packs should be applied if they are available. First-aid measures to maintain the patient's airway and prevent injury are important. The patient should be protected from sunlight and should be moved to a cool environment and evacuated to an emergency ward as soon as possible. Hypothermic mattresses or, in urgent circumstances, ice-water immersion may be employed. These treatments are very uncomfortable and should be used only when the hyperthermia is truly deleterious. Cool intravenous fluids should be administered as soon as parenteral therapy becomes available, but oral hydration is inadvisable because of the risk of aspiration.

Physical cooling must continue after the patient arrives in the emergency ward; many techniques are available, but choosing among them is difficult. Immersion in ice water has been a standard approach. Another traditional technique is the application of ice packs to the body surface. Both techniques have the disadvantage of producing cutaneous vasoconstriction, which could impede the transfer of heat from the body's surface; ice-water baths may also pose logistic difficulties in heatstroke patients who are combative and in those with seizures, vomiting, or diarrhea. Despite these drawbacks, ice-water baths are effective; there were no fatalities in a series of 252 marine recruits whose therapy for exertional heatstroke included ice-water immersion.[22] A newer technique for cooling relies on evaporation rather than conduction to dissipate heat: the patient is sprayed with water at a temperature of 15° C; then warm air at a temperature of 30° to 35° C is blown on the patient. A special body-cooling unit that facilitates evaporative cooling has been used with good results in Saudi Arabia.[23] Although alcohol is more volatile than water, it should not be used for evaporative cooling because of the risk of alcohol intoxication, especially in children. Many other cooling techniques, ranging from ice-water enemas and gastric lavage to the use of the downdraft from helicopter blades, have been used to treat heatstroke but are of secondary importance. In general, hospitals in the United States should choose a cooling technique on the basis of their experience and facilities.

Body temperature should be monitored continuously during cooling; in most cases, body temperature will decline significantly in 10 to 40 minutes, at which time cooling should be reduced to avoid hypothermic overshoot with associated shivering and rigors.

The administration of room-temperature intravenous fluids is an important aspect of therapy, helping to lower core temperature and to correct dehydration; meticulous cardiovascular monitoring is essential to prevent fluid overload while ensuring adequate volume replacement. In addition to needing cardiovascular support, patients with heatstroke require careful respiratory monitoring and support, along with fluid and electrolyte monitoring and therapy. Complications such as rhabdomyolysis, renal failure, aspiration pneumonia, DIC, and seizures necessitate additional therapeutic interventions. Despite its efficacy in malignant hyperthermia of anesthesia (see below), dantrolene sodium is ineffective in heatstroke.[21]

Even with aggressive therapy, heatstroke has an appreciable mortality. Most patients who recover have normal thermoregulatory mechanisms and heat tolerance, but all should be instructed on ways to reduce the risk of recurrences.

NEUROLEPTIC MALIGNANT SYNDROME

Etiology

First described in 1960, the neuroleptic malignant syndrome (NMS) is uncommon, occurring in fewer than 1% of patients receiving neuroleptic agents [*see 8:I Management of Poisoning and Drug Overdose and 13:VIII Anxiety Disorders*]. NMS usually occurs within the first 30 days of therapy; haloperidol is most often implicated, but other butyrophenones (e.g., droperidol), various phenothiazines, thioxanthenes, atypical antipsychotics (e.g., resperidone and clozapine), tricyclic antidepressants, and monoamine oxidase (MAO) inhibitors may also be responsible.[24-26] Nonpsychiatric drugs such as metoclopramide have occasionally been linked to NMS.[25] NMS may also be precipitated by the withdrawal of amantadine, levodopa, or other dopaminergic drugs used to treat Parkinson disease.[27] In some cases, dehydration and agitation are contributing factors.

Pathophysiology

NMS is primarily a disorder of excessive heat production. It is probably precipitated by the blockade of dopaminergic receptors in the nigrostriatal tracts, leading to uncontrolled contractions of skeletal muscle, which, in turn, produce excessive body heat. There is no genetic predisposition for NMS, nor does it occur as a result of a drug overdose. Physical exhaustion, dehydration, and underlying organic brain syndromes have been cited as predisposing factors. Some patients who have recovered from NMS have subsequently received neuroleptic agents without suffering recurrences of NMS, but recurrent NMS does develop in 30% of patients who are rechallenged with neuroleptics.

Diagnosis

The clinical features of NMS evolve progressively, usually over a period of 1 to 3 days. Hyperthermia is universal. Other symptoms include bradykinesia, severe muscular rigidity (sometimes described as lead-pipe rigidity); altered sensorium; autonomic dysfunction that produces tachycardia, labile blood pressure, and diaphoresis; and extrapyramidal abnormalities. Laboratory abnormalities include hypernatremia and other electrolyte abnormalities, acidosis, hemoconcentration and leukocytosis, rhabdomyolysis, and abnormal findings on renal and hepatic function tests.

Differential Diagnosis

The differential diagnosis of NMS includes infections, lethal catatonia,[28] an uncommon psychiatric disorder in which agita-

tion progresses to muscular rigidity and hyperthermia, and the serotonin syndrome (see below).

Treatment

The offending neuroleptic agent must be withdrawn in all cases of NMS. Physical cooling as treatment of NMS has not been evaluated in clinical trials but should prove useful in the acute phase of the disorder. Meticulous fluid and electrolyte therapy and careful cardiovascular support are mandatory. Dantrolene sodium, a muscle relaxant, and bromocriptine, a dopamine agonist, have each been reported to reduce the duration of hyperthermia, but these agents have not been evaluated in controlled trials.[29] Aggressive therapy has allowed the mortality associated with NMS to decline to about 10%; patients who recover face a 30% risk of recurrence.[25]

SEROTONIN SYNDROME

The serotonin syndrome[30] has become more prevalent with the increasing use of selective serotonin reuptake inhibitors (SSRIs) and other serotoninergic agents. Because of its broad range of clinical manifestations and the large number of prescription, over-the-counter, and herbal products that have been implicated, the serotonin syndrome is often unrecognized. The actual prevalence is not known, but it occurs in 14% to 16% of patients who take overdoses of SSRIs. The syndrome results from excessive agonism of serotoninergic receptors in the central and peripheral nervous systems.

Diagnosis

The syndrome begins abruptly within 5 weeks of using serotoninergic agents, either singly or in combination. In mild cases, patients exhibit hyperkinesia, intermittent tremors, hyperreflexia and clonus, hyperactive bowel sounds, diarrhea, mydriasis, and tachycardia. Patients with mild cases of the syndrome may exhibit shivering and diaphoresis, but body temperature is normal. In moderate to severe cases, however, hyperthermia is the rule and may be severe. Muscular rigidity, agitation, and ocular clonus or inducible clonus are characteristic features of the severe serotonin syndrome.

Differential Diagnosis

The differential diagnosis includes infection, anticholinergic poisoning, which can be distinguished on the basis of its normal reflexes and absent bowel sounds, and malignant hyperthermia of anesthesia (see below). NMS is the major mimic of the serotonin syndrome, but the symptoms of NMS evolve more slowly, and patients typically display bradykinesia or akinesia and so-called lead-pipe muscular rigidity rather than hyperkinesia, tremors, hyperreflexia, and clonus.

Treatment

Management of the serotonin syndrome requires removal of the responsible drug or drugs and metabolic and hemodynamic support. In mild cases, agitation and tremors can often be controlled with benzodiazepines; however, severely ill patients require sedation, neuromuscular paralysis, and ventilator support. Antipyretics, beta blockers, bromocriptine, and dantrolene are not effective. Cyproheptadine, olanzapine, and chlorpromazine have been useful in isolated cases but have not been studied fully.[30]

MALIGNANT HYPERTHERMIA OF ANESTHESIA

Like NMS, malignant hyperthermia of anesthesia (MHA) is a high-mortality disorder of involuntary skeletal muscle hyperactivity and excessive heat production triggered by a pharmacologic agent. Unlike NMS and the serotonin syndrome, however, MHA is a genetically determined disorder precipitated by anesthetic agents such as halogenated inhalation agents and depolarizing muscle relaxants. Mutations in the gene encoding the skeletal muscle ryanodine receptor type I (*RYRI*) are often responsible for the disorder.[30] Persons who are susceptible to MHA may be identified by a family history of the disorder and by performance of a muscle biopsy and a caffeine-halothane contracture test; mutation analysis and other noninvasive tests to identify susceptible individuals are being developed.[31]

Diagnosis

MHA usually begins shortly after the anesthetic is administered, but it may be delayed for hours. Muscular rigidity and severe hyperthermia are typical; other findings include tachycardia, hypotension, arrhythmias, hyperpnea, hypoxia, hypercapnia, hyperkalemia, lactic acidosis, rhabdomyolysis, and DIC.

Treatment

MHA is an anesthetic emergency. Early detection of hypercapnia during general anesthesia and prompt treatment using intravenously administered dantrolene sodium have markedly improved the prognosis of patients with MHA. Discontinuance of anesthesia is mandatory; physical cooling is therapeutically important, as are cardiopulmonary support and correction of the metabolic abnormalities.

Fever

Fever has been recognized as a cardinal feature of disease since antiquity, but only recently has the pathophysiology of fever come to be understood. Beginning with the work of Dr. Paul Beeson in 1948, it became clear that the ultimate cause of fever is not a bacterial product (a so-called exogenous pyrogen) but a product of host inflammatory cells (i.e., an endogenous pyrogen). For years, the endogenous pyrogen was thought to be a product of polymorphonuclear leukocytes and was referred to as leukocytic pyrogen. Exciting studies, however, have demonstrated that mononuclear phagocytes are the principal source of endogenous pyrogen and that a variety of mononuclear cell products—cytokines—can mediate the febrile response.[32,33] Cytokines are also important as mediators of the acute-phase response to infection and inflammation.

PATHOPHYSIOLOGY

The cytokines interleukin-1 (IL-1), IL-6, interferon gamma, and tumor necrosis factor (TNF) function as pyrogens by acting directly on the hypothalamus to elevate the thermal set point. A variety of stimuli, such as microorganisms, exposure to endotoxin and other bacterial toxins or microbial products, phagocytosis, antigen-antibody immune complexes, and various forms of tissue injury, can initiate IL-1 production by mononuclear phagocytes and many other cells. In the pathogenesis of fever, IL-1 acts as a hormone in that it is carried by the circulation from the local inflammatory site of production to the CNS, where it acts directly on the hypothalamic thermal control center. Its mechanism of action appears to involve induction of phospholipases, which in turn cause the release of arachidonic acids from membrane phospholipids. As a result, prostaglandin levels rise, particularly levels of prostaglandin E. Elevated levels of prostaglandins ap-

pear to be important in raising the hypothalamic thermal set point; this mechanism accounts for why prostaglandin inhibitors such as aspirin are effective antipyretic agents.[8]

Although this classic model of pathogenesis has been validated by numerous studies of fever, further research suggests that several additional pathways may be involved. For example, pyrogenic cytokines may be produced locally in the CNS as well as systemically,[34] and cytokines may signal the thermal control center by neural as well as hormonal mechanisms.[35]

Once the hypothalamic thermal set point has been elevated, thermoregulatory mechanisms are brought into play to raise the body temperature to the level of the new set point. Autonomic efferents lead to heat conversion through cutaneous vasoconstriction and cessation of sweating. Somatic nerves are responsible for increasing heat production via increased skeletal muscle tone or shivering. The myalgias that accompany many febrile states may in part be caused by this increased muscle tone. Rigor, which is a dramatic precursor of some fever spikes, is nothing more than an exaggerated form of shivering that rapidly elevates body temperature in response to an increased hypothalamic thermal set point.

CONSEQUENCES OF ELEVATED BODY TEMPERATURE

Possible Benefits of Fever

Although fever is a common response to infection in many species, there is no direct evidence that it is beneficial to host defense mechanisms in humans. However, insights into the immunostimulatory properties of IL-1 and TNF have led to speculation that fever itself may promote recovery from infection.[36] IL-1 and TNF appear to act across species, order, and class barriers and probably evolved 300 million years ago; this evolutionary stability further suggests that fever plays a role in host defense mechanisms.

In humans, fever appears to decrease serum iron levels. Many microbes need iron for growth, and it has been suggested that fever-induced hypoferremia is a helpful host defense mechanism. Fever is also marked by a metabolic shift away from glucose, an excellent substrate for bacteria, to fat and protein as energy sources. In addition, some microbes, including gonococci and *Treponema pallidum,* are quite heat sensitive and can be killed in experimental animals by artificially induced fevers. However, natural infection never produces body temperatures that are high enough to have this effect.

Extreme pyrexia may be surprisingly well tolerated. It seems likely that the widespread tissue damage, multiple laboratory abnormalities, and high mortality observed in disorders such as heatstroke, MHA, thyroid storm, the serotonin syndrome, and NMS are caused by the underlying disorder rather than the elevated temperature itself. Indeed, in the preantibiotic era, fever therapy was well tolerated, with little evidence of tissue damage, despite the fact that temperatures as high as 41.7° C (107.1° F) were induced.

Therapeutic hyperthermia is being used to treat noninfectious diseases. Regional hyperthermia is a traditional therapy for many musculoskeletal disorders. Heating pads, whirlpools, and ultrasonography have all been used to increase the temperature of injured tissues. Despite widespread endorsements by patients and practitioners, however, heat treatments for musculoskeletal injuries have not been subjected to controlled clinical trials. Hyperthermia is also being investigated for a possible role in treating malignancies. Adjunctive regional hyperthermia often improves the rate of response when compared with radiotherapy or chemotherapy alone; superficial tumors respond most favorably, but better responses may be obtained with internal malignancies as techniques for deep heating are improved.

Complications of Fever

Pyrexia can have deleterious consequences, but complications depend more on the underlying cause of the temperature elevation and the patient's overall condition than on the level of the temperature. Elevated body temperatures are most harmful to the very young and the very old. Because pyrexia increases oxygen consumption, it imposes circulatory demands that may precipitate ischemia, arrhythmias, or congestive heart failure in patients with cardiovascular disease. Fever during pregnancy does not appear to cause fetal death,[37] but during the first trimester, fever may increase the risk of congenital heart disease.[38] Febrile seizures are a risk in children between 6 months and 6 years of age. The patient's age and the degree of fever are the major risk factors; genetics may also influence vulnerability.[39] Although febrile convulsions are generally benign, they are alarming, and it is always necessary to exclude underlying neurologic illnesses, including meningitis.[40] Both intravenous diazepam and intranasal midazolam are effective in controlling febrile convulsions.[41] In the absence of unprovoked seizures, long-term anticonvulsant therapy is usually unnecessary in children with febrile seizures and may even be deleterious; diazepam, administered orally only when fever occurs, safely and effectively reduces the risk of recurrent febrile seizures.[42] The long-term prognosis is excellent.[43]

Febrile seizures do not occur in adults, but fever often results in decreased concentration and sleepiness; high temperatures commonly produce an altered sensorium, including stupor and delirium. Patients with strokes are at risk for additional brain injury from fever; the result is a marked increase in morbidity and mortality,[44] especially when the pyrexia occurs soon after the stroke.[45] Because of this, even modest elevations in temperature should be suppressed in patients with acute strokes.[46] In addition, moderate induced hypothermia may improve the outcome of strokes, even in patients who are afebrile on presentation.[47] Hypothermia may also improve outcome after traumatic brain injury[48] in neonates with hypoxic-ischemic encephalopathy,[49] and in patients receiving cardiopulmonary resuscitation.[50]

DIAGNOSIS

The diagnostic process involves determining whether the patient is febrile, which requires (1) accurate measurements of body temperature and comparisons of body temperature with the diurnal range of normal body temperature and (2) a determination of whether the elevation in temperature is the result of inflammation or infection (fever) rather than thermoregulative failure (hyperthermia).

TREATMENT

The approach to the febrile patient involves three elements: diagnosis and management of the underlying disorder; cardiac, respiratory, and metabolic support; and, when indicated, lowering of body temperature. Infection is the leading cause of elevated body temperature; all febrile patients should be systematically evaluated for infection, and antibiotics should be administered whenever appropriate.

In patients with extreme pyrexia, additional laboratory studies are important both to screen for hyperthermia caused by thermoregulatory defects and to assess tissue damage. Elec-

trolytes, coagulation parameters, muscle enzymes, and arterial blood gases should all be measured, and renal and liver function should be assessed. Studies of thyroid and adrenal function may be indicated. Cardiac function, blood pressure, urine output, and neurologic status should be monitored closely. Adequate hydration is mandatory, and circulatory or respiratory support may be necessary.

Proper management of the elevated temperature itself depends on the clinical circumstances. Although clinicians usually choose to suppress fever, it is unclear whether elevated body temperatures promote or impede recovery from infection; similarly, the risks and benefits of antipyretic therapy have not yet been defined.[8,51,52] Because many patients tolerate high body temperatures very well, antipyretic therapy may actually produce more discomfort than the pyrexia itself. In other patients, myalgias, flushing, fatigue, loss of concentration, shivering, or chills can be very uncomfortable, and antipyretic therapy should be used for symptomatic relief. Significant pyrexia should always be treated in patients with strokes, in young children, in elderly or debilitated patients, and in persons with cardiopulmonary disease, because these patient groups are most likely to suffer adverse consequences. Hyperthermia should be treated in all patients with MHA, heatstroke, NMS, the serotonin syndrome, or thyroid storm. Finally, it may be prudent to treat any patient whose temperature exceeds 40° C (104° F), even a healthy young adult.

The choice of cooling technique depends on the pathogenesis of the temperature elevation [*see Figure 1*]. In patients with fever caused by infection or other inflammatory states, an elevated hypothalamic thermal set point is responsible for the pyrexia. Aspirin or acetaminophen should be used to lower the set point; the drugs seem equally effective as antipyretics, but acetaminophen is preferred in pediatric patients because aspirin may precipitate Reye syndrome in children with influenza or varicella. A broad range of nonsteroidal anti-inflammatory drugs (NSAIDs) and other cyclooxygenase-2 (COX-2) inhibitors can also lower the set point.[8,53] Acetaminophen may act by inhibiting COX-3, a COX-1 variant.[54] If physical cooling methods are used before antipyretic drugs are fully effective, homeostatic mechanisms will continue to operate in an attempt to raise body temperature, resulting in intense vasoconstriction and shivering,[55] which produce adverse cardiovascular and metabolic effects, as well as intense patient discomfort. To avoid these problems, febrile patients who require physical cooling to rapidly lower body temperature should be heavily sedated and ventilated to prevent shivering in conjunction with physical cooling.[56] In contrast, physical cooling is the treatment of choice for patients with hyperthermia. In all patients, careful attention is required to prevent hypothermic overshoot on the one hand and recurrent fever on the other.[57] Although pyrexia may be the most spectacular symptom, meticulous attention to the underlying disorder is of primary importance in all cases.

Fever of Undetermined Origin

FUO presents one of the most challenging and perplexing problems in clinical medicine. Such fevers may persist for weeks or months in the absence of characteristic clinical findings or clues. Ultimately, most such obscure fevers prove to be caused by common diseases presenting in an atypical fashion rather than by rare and exotic illnesses.

Petersdorf and Beeson, in their classic monograph,[58] specified three criteria to define FUO:

1. Duration: at least 3 weeks. This requirement eliminates from consideration most short-lived fevers of indeterminate or viral origin and most postoperative fevers.
2. Magnitude: a temperature higher than 38.3° C (100.9° F) on at least several occasions. This criterion eliminates from consideration persons with unusually prominent circadian fluctuations in body temperature with daily readings on the order of 38.0° C (100.4° F). If such persons are not clinically ill, their elevated temperature should not cause too much concern.
3. Obscure nature: perplexing enough to defy diagnosis after 1 week of study. In the past, inpatient study was required to evaluate FUO, but with changing admission practices, intelligent and thorough study encompassing at least three outpatient visits or 3 days in the hospital is now considered sufficient.[59]

ETIOLOGIC CLASSIFICATION

Although geographic factors are relevant, the leading causes of FUO are reasonably uniform throughout the United States. The relative frequency of the etiologic categories responsible for FUO have been relatively stable over the past 5 decades [*see Table 2*]. In a community hospital study,[60] noninvasive testing provided the diagnosis in 42% of patients, with serologies and lysis centrifugation blood cultures being the most useful noninvasive techniques. Computed tomography–guided percutaneous biopsies were the most useful invasive procedures. Only 9% of patients remained undiagnosed. With an increasing population of

Table 2 Causes of Fever of Undetermined Origin Over 5 Decades

Cause	*Study*				
	New Haven, 1961[58] *(% of 100 Patients)*	*Boston, 1973*[104] *(% of 128 Patients)*	*Seattle, 1982*[105] *(% of 105 Patients)*	*Rhode Island, 1992*[60] *(% of 89 Patients)*	*Leuven, Belgium, 2003*[106] *(% of 290 Patients*)*
Infections	36	35	30	33	20
Neoplasms	19	23	31	24	20
Collagen vascular diseases	13	16	9	16	19
Other specific causes	25	18	18	18	18
Undiagnosed	7	8	12	9	34

*Percentages have been modified to conform to the diagnostic criteria used in the American series.

Table 3 Causes of Fever of Undetermined Origin

Systemic Infections
- Tuberculosis (miliary)
- Infective endocarditis (primarily bacterial endocarditis but also endocarditis with a fungal, Q fever, or chlamydial etiology)
- Bacteremia from an inapparent primary focus
- Chronic meningococcemia
- Brucellosis
- Listeriosis, vibriosis, leptospirosis, relapsing fever (caused by *Borrelia recurrentis*), rat-bite fever (caused by either *Streptobacillus moniliformis* or *Spirillum minus*)
- Miscellaneous
- Psittacosis, toxoplasmosis (disseminated acquired form), Q fever, disseminated deep mycotic infections (e.g., histoplasmosis, blastomycosis, cryptococcosis), cytomegalovirus infection, Whipple disease

Localized Infections and Abscesses
- Hepatic infections
- Liver abscess
- Cholangitis
- Intraperitoneal infections
- Upper abdomen: empyema of gallbladder, pericholecystic and subhepatic abscesses, right or left subphrenic abscesses, lesser sac abscess
- Lower abdomen: periappendiceal and peridiverticular abscesses
- Other intra-abdominal abscesses
- Tubo-ovarian abscess, pelvic inflammatory disease, pelvic abscess
- Retroperitoneal abscess, pancreatic abscess
- Urinary tract infections
- Perinephric abscess
- Renal carbuncle
- Pyelonephritis with ureteral obstruction and pyonephrosis
- Prostatic abscess

Neoplasms
- Hematopoietic malignancies: Hodgkin disease and other lymphomas, leukemias, myeloid metaplasia, malignant histiocytosis
- Other malignancies: renal cell cancer, colon cancer, hepatoma
- Benign neoplasms: left atrial myxoma, pheochromocytoma

Collagen Vascular Diseases
- Temporal arteritis, adult-onset juvenile rheumatoid arthritis, Wegener granulomatosis, polyarteritis, systemic lupus erythematosus, relapsing polychondritis

Granulomatous Diseases
- Sarcoidosis, idiopathic granulomatous hepatitis

Metabolic and Hereditary Disorders
- Familial Mediterranean fever, Fabry disease

Endocrine Disorders
- Adrenal insufficiency, thyrotoxicosis, hyperparathyroidism

Thermoregulatory Disorders
- Hypothalamic dysfunction (e.g., caused by strokes or tumors), encephalitis

Drug Fever
- Antimicrobial agents (β-lactam antibiotics, sulfonamides, nitrofurantoin, isoniazid), antihypertensives (hydralazine), anticonvulsants (phenytoin), allopurinol, and many others

Miscellaneous Conditions
- Pulmonary emboli, alcoholic hepatitis and cirrhosis, inflammatory bowel disease, thrombophlebitis, factitious fever

immunosuppressed and chronically ill patients, the relative importance of the disorders that cause FUO may change. As a result, it may be prudent to consider so-called classic FUO separately from FUO in patients with HIV infection, neutropenia, or nosocomial fevers.[59,61-63] Despite changes in host factors and diagnostic techniques, the basic categorization of the causes of FUO remains valid.

Infections

In the initial evaluation of FUO, the possibility of infection should be carefully considered.

Systemic infections The two major systemic infections to consider in the evaluation of FUO are tuberculosis (usually disseminated but sometimes confined predominantly to the liver and spleen) and infective endocarditis [*see Table 3*]. Most FUO cases caused by miliary tuberculosis arise in elderly patients in whom dissemination has followed activation of quiescent foci. Often, in cases caused by miliary tuberculosis, the intermediate-strength (5 tuberculin units) purified protein derivative skin test is negative, and miliary pulmonary lesions are not present on the chest x-ray. Anemia, leukopenia, or, rarely, a leukemoid reaction caused by bone marrow involvement may be evident; bone marrow biopsy is a very helpful diagnostic test in patients in whom miliary tuberculosis is suspected. An isolated elevation of the serum alkaline phosphatase level may indicate miliary involvement of the liver by tuberculosis, other infection, or neoplasm. The histologic findings on liver biopsy often suggest the diagnosis, and a portion of the specimen should always be cultured for the presence of tubercle bacilli.

Infective endocarditis, usually subacute, is also an important diagnostic consideration. Most patients with subacute bacterial endocarditis have a heart murmur. In about 5% of cases, however, particularly in the elderly, the murmur may be absent or may be considered functional. Blood cultures would be expected to provide the diagnosis in a patient with subacute bacterial endocarditis, particularly because only 5% of patients with endocarditis have negative blood cultures. The leading cause of negative blood cultures in patients with endocarditis is the prior administration of antibiotics. It is therefore very important that a number of blood cultures be obtained, including some as long as 5 to 10 days after antibiotics have been withdrawn. Other causes of culture-negative endocarditis that should be considered in patients with FUO include infection with fastidious bacteria, chlamydial infection, and Q fever. Careful scrutiny for the peripheral stigmas of endocarditis is essential in the evaluation of any patient with FUO. Echocardiography may reveal valvular vegetations in patients with endocarditis; transesophageal studies are more sensitive but more invasive than transthoracic echocardiography. Left atrial myxomas mimic culture-negative endocarditis but may be detected with echocardiography.

Other systemic infections, including bacteremias that occur in the absence of any obvious primary site of involvement, only rarely cause FUO [*see Table 3*]. Viral infections are usually self-limited and do not produce fevers that last longer than 3 weeks. Important exceptions to this generalization are Epstein-Barr virus (EBV)[64] and cytomegalovirus (CMV) infections, which may occasionally present as FUO (often with some mononucleosis-like features) in otherwise healthy individuals. More frequently, CMV infection develops in patients who have received multiple blood transfusions or who have undergone organ transplantation; CMV is the cause of 50% of all febrile episodes in renal transplant recipients.

Localized infections The more common types of localized infection that present as FUO include hepatic abscess, subphrenic abscess, and subhepatic and pericholecystic abscess [*see Table 3*]. Liver abscesses are often occult; the physician should look for a history that includes symptoms of biliary tract disease, recent blunt abdominal trauma, or travel, which might suggest

the diagnosis of amebiasis. Hepatomegaly may be absent initially. The serum alkaline phosphatase level is usually elevated even when the abscess is solitary. Serologic tests for amebiasis are positive in patients with amebic liver abscess. Elevation of the diaphragm, particularly when accompanied by overlying pulmonary atelectasis or a pleural effusion, should raise suspicion of a subphrenic abscess. Ultrasonography and CT are valuable in identifying such collections; gallium scans are less useful for this purpose.

Localized infection in the urinary tract is an important consideration in a patient with FUO; perinephric abscess and renal carbuncle are best diagnosed by ultrasonography or CT. Many other localized infections occasionally present as FUO; occult dental infections are one such example and illustrate the need for thoroughness in the evaluation of patients with obscure fevers.

Neoplasms

Lymphoma, particularly Hodgkin disease, is the most common neoplastic cause of obscure fever. Lymphoma may be difficult to diagnose when the principal site of involvement is the retroperitoneal nodes, but abdominal CT scans greatly facilitate this diagnosis; a skin biopsy may help identify intravascular lymphoma as the cause of an FUO.[65] Although so-called Pel-Ebstein recurrent fevers suggest Hodgkin disease, they are observed in only a minority of patients with this disorder. The development of fever in a patient who has myeloma or chronic lymphocytic leukemia is usually caused by superimposed infection and not by the neoplastic process; in some patients, however, the febrile course appears to be caused by the malignancy itself.[66] Occasionally, a patient with the preleukemia syndrome will present with fever and atypical blood and bone marrow changes, suggesting myeloid metaplasia or a leukemoid response. Only after some months can the hematologic picture be established as leukemia.

Solid tumors can also be associated with fever; hypernephroma is the leading example. As many as 10% of patients with colorectal carcinoma present with fever; either extension of the tumor through the bowel wall, producing a paracolonic abscess, or necrosis and abscess formation in a polypoid intraluminal lesion may be the underlying mechanism. Metastatic cancer may be responsible for continuing fever; hepatic involvement is not necessary for fever to occur. Occasionally, a neuroblastoma involving bone or soft tissues or a pheochromocytoma may have a febrile course. Fevers caused by malignant disease often respond to therapy with NSAIDs; fevers caused by infections may be less likely to respond completely to these agents,[67] but this distinction is not sufficient as a diagnostic test.[68] Hospitalization for FUO often reflects a poor prognosis in patients with malignancies.[69]

Collagen Vascular Disease

A variety of connective tissue disorders and vasculitides may produce prolonged fevers before the development of articular or other characteristic manifestations. In the elderly, polymyalgia rheumatica and the closely related disorder giant cell arteritis (temporal arteritis) are the most common connective tissue disorders presenting as FUO. Malaise, weight loss, muscle weakness, mild arthralgias without overt arthritis, and a markedly elevated erythrocyte sedimentation rate (often > 100 mm/hr) are usual features. Jaw claudication, visual symptoms, and a tender or thickened temporal artery suggest the diagnosis of giant cell arteritis. As many as 15% of cases of giant cell arteritis present as FUO, and in some patients, the vasculitis itself remains occult. Similarly, virtually all patients with adult-onset Still disease are febrile,[70] and systemic symptoms such as fever and weakness may antedate by weeks or months the evolution of the more characteristic clinical manifestations of adult juvenile rheumatoid arthritis. In other patients, involvement of the paranasal sinuses and mastoid or the rapid excavation of a pulmonary lesion suggests Wegener granulomatosis. Many other connective tissue diseases, ranging from classic vasculitides such as systemic lupus erythematosus to uncommon disorders such as relapsing polychondritis, can also present as FUO.

Less Common Etiologic Categories

Granulomatous disease Granulomatous diseases of noninfectious origin may be responsible for FUO. Prolonged fever is uncommon in sarcoidosis, but when it does occur, prominent hilar adenopathy, ocular involvement, erythema nodosum, and hepatic granulomas are usually also present. Biopsy of involved lymph nodes, muscle, or liver usually shows noncaseating granulomas. In addition to sarcoidosis, there are about 40 diseases that may be associated with hepatic granulomas. Treatable infectious granulomatous diseases (e.g., tuberculosis, brucellosis, histoplasmosis, and cat-scratch disease) must be ruled out by cultures, skin tests, serologic tests, and special stains of tissue biopsy specimens. In rare instances, despite extensive investigation and therapeutic trials with antituberculous drugs, an etiologic diagnosis cannot be made in patients with noncaseating hepatic granulomas who have a febrile illness of many months' duration.[71] Beneficial results have been achieved in such cases by giving corticosteroids after excluding the other specific granulomatous diseases; methotrexate also appears to be helpful.[72] Corticosteroids have been beneficial for other patients with idiopathic granulomatosis and FUO.[73] Starch peritonitis represents a febrile granulomatous response to starch introduced on surgical gloves. The nature of the process may not be appreciated for weeks; initially, findings of a doughy abdominal mass and fever are thought to be the result of a postoperative abscess.

Inflammatory bowel disease Bowel symptoms are prominent in almost all patients with idiopathic ulcerative colitis, granulomatous colitis, or regional enteritis, and the diagnosis is obvious in such febrile patients. Occasionally, however, bowel symptoms may not be marked or may be of such long duration that they become accepted as the norm. In this setting, FUO may be the presenting complaint in a patient with inflammatory bowel disease.

Alcoholic hepatitis and cirrhosis Fever is occasionally observed in cases of cirrhosis.[74] Attention should first be directed to possible complicating infections—such as spontaneous bacterial peritonitis, enterogenous bacteremias, or tuberculosis—or to an unrelated process. Active hepatocellular necrosis may occur in the course of alcoholic hepatitis and may account for low-grade fever.

Pulmonary emboli In rare instances, a patient may have multiple small pulmonary emboli, but no significant changes in arterial blood gases or on the chest film will be apparent; the patient will present primarily with a problem of unexplained fever. The fever may exceed 39.0° C (102.2° F), but high-grade fevers caused by pulmonary emboli seldom persist longer than 1 week. Thrombophlebitis itself may be a source of protracted fever, even in the absence of pulmonary emboli.

Drug fever Drug fever frequently occurs in the absence of other manifestations of hypersensitivity, such as rash and eosinophilia. Antimicrobial agents (e.g., β-lactams, sulfonamides, nitrofurantoin, and isoniazid), antihypertensives (e.g., hydralazine and methyldopa), anticonvulsants (e.g., phenytoin), and allopurinol are among the most common offenders, but many other drugs have been implicated. In most instances, the diagnosis of drug fever is considered within the first several weeks of onset of FUO, and any recently administered drugs are discontinued. Several drugs, however, such as phenytoin, methyldopa, and isoniazid, may not produce drug fever until weeks or months after their initial use. Drugs such as these may be overlooked just because they have been administered for some time without producing side effects. Intramuscular injections of analgesics can produce FUO, which may or may not be accompanied by the presence of a sterile abscess or other gross evidence of tissue injury.

Factitious fever In rare instances, a patient may simulate illness by deliberately producing false elevations in temperature.[75] Factitious fever is one of the most challenging etiologic categories of FUO. The patients are usually female and are often paramedical personnel. The underlying problem may be malingering or a more complicated emotional disorder. Discordance between the marked temperature elevations and the pulse rate, distortion of the usual diurnal temperature curve, and absence of diaphoresis when the fever abates suggest the diagnosis.

Miscellaneous causes The hereditary periodic fevers can present as FUOs.[76] The diagnosis of familial Mediterranean fever is suggested by ethnic background, episodic occurrence of fever in association with abdominal pain or other signs of polyserositis, and well-being between attacks [*see 15:VIII Systemic Vasculitis Syndromes*]. The diagnosis can be difficult when recurrent fever is the only symptom[77]; molecular techniques can facilitate the diagnosis of various hereditary periodic fever syndromes.[78]

Whipple disease is a multisystem infection caused by the gram-positive actinomycete *Tropheryma whippelii*.[79] Patients may present with a prolonged febrile illness in association with weight loss, arthralgias, and weakness.[80] The use of special tissue culture and immunodiagnostic tests in diagnosis are being studied[80]; a polymerase chain reaction test for the causative organism is highly sensitive and specific,[81] but false positive reactions have been reported.[82] Inflammatory pseudotumor of intra-abdominal lymph nodes may present as FUO[83]; the clinical features of this disorder may resemble those of Whipple disease, but the pathologic findings are distinctive, and surgical excision of the involved nodes may induce prolonged remissions. Kikuchi-Fujimoto disease is another rare cause of fever and lymphadenopathy.[84]

CNS lesions are decidedly uncommon causes of FUO, except in very obtunded patients with extensive brain damage. Endocrinologic abnormalities, such as subacute thyroiditis, or metabolic disorders, such as hypertriglyceridemia, hypercholesterolemia, or glycosphingolipid storage disease (Fabry disease), may occasionally present as FUO. Many other disorders as diverse as pernicious anemia[85] and xanthogranulomatous pyelonephritis[86] have been identified as rare causes of FUO.

FUO in the Elderly

Advanced age appears to blunt the febrile response to many illnesses ("the older, the colder").[87] Because of this, an elderly patient with a persistent rectal temperature as low as 37.5° C (99.5° F) may qualify for an FUO evaluation.[88] The causes of FUO in the elderly are similar to those in younger patients; however, in elderly patients, tuberculosis is a more common infectious cause and temporal arteritis the most common vasculitic cause.[88,89] The presence of multiple comorbidities may complicate the diagnosis of FUO in the elderly and may temper the use of invasive diagnostic studies.

FUO in Patients with AIDS

Fever is extremely common in patients infected with HIV; typically, the diagnostic challenge is not in determining the source of fever but in deciding which of several potentially pyrogenic processes is most important. However, patients infected with HIV can also present with prolonged, diagnostically obscure fevers. Most often, such patients have advanced AIDS, with $CD4^+$ T cell counts of less than $100/mm^3$.[90] Infections account for more than 75% of FUO cases in such patients[91]; in one series, the more common diagnoses were disseminated *Mycobacterium avium* complex (31%), *Pneumocystis carinii* pneumonia (13%), CMV infection (11%), disseminated histoplasmosis (7%), and lymphoma (7%).[92] Other infectious etiologies are toxoplasmosis, cryptococcosis, salmonellosis, and varicella-zoster virus infections. In Europe, leishmaniasis is also an important cause of FUO in patients with AIDS.[93] Among noninfectious etiologies, lymphoma and drug fever are most prominent. HIV itself is an uncommon cause of FUO.

Blood cultures are the most useful diagnostic tests. Although more invasive than blood cultures, biopsies of bone marrow,[93] liver,[94] or lymph nodes may provide more rapid diagnosis by allowing direct visualization of organisms. Other useful tests include CT of the chest and abdomen and a serum cryptococcal antigen determination.

DIAGNOSIS

History and Physical Examination

With rare exceptions, neither the temperature nor the pattern of fever permits discrimination between the causes of FUO. Detailed review of the history is essential, particularly regarding travel,[95] animal exposure, occupational risks, and other epidemiologic factors; previous trauma or surgery; or features relevant to each of the diagnoses outlined. When physical findings are being evaluated, particular attention should be paid to structures that may be inapparent sources of obscure fever, such as the cardiovascular system, the abdominal viscera, and the genitourinary tract. Lymph nodes should be examined not only in the usual distribution but also in the epitrochlear areas and along the medial aspect of the upper arm. A search of the skin, nail beds, and mucous membranes for petechiae and vasculitic lesions may provide important clues to the diagnosis of endocarditis or of collagen vascular diseases. Funduscopic examination may reveal choroidal tubercles or signs of vasculitis or endocarditis. Rectal examination is particularly important in elderly or obtunded patients, in whom a perirectal or prostatic abscess may be overlooked.

Common Laboratory Studies

Blood cultures Blood cultures should include aerobic (5% to 10% CO_2 tension) and anaerobic cultures that have been incubated for at least 2 weeks. Newer blood culture techniques, including the use of lysis centrifugation and the BACTEC radiometric mycobacterial culture system, may be helpful, particularly in HIV-positive patients.

Serologic tests Serologic tests often employed in cases of FUO include *Brucella* and *Salmonella* agglutinations (usually not very helpful), antistreptolysin O (ASO) titer if acute rheumatic fever is suspected, the Venereal Disease Research Laboratories (VDRL) test for syphilis, serologic tests for less common infections (e.g., psittacosis, toxoplasmosis, and CMV), the test for rheumatoid factor, and antinuclear antibody tests. A serum specimen obtained during an acute-phase host response can be frozen for subsequent comparison with a late-phase or convalescent-phase specimen to look for an increase in titers to specific pathogens. HIV testing should be performed to evaluate host factors that may predispose to FUO.

Sedimentation rate In cases of FUO, a sedimentation rate greater than 100 mm/hr suggests vasculitis but does not differentiate between this disorder and neoplasms, tuberculosis, or pyogenic infections.

Serum enzymes and chemistries The results of liver function tests may indicate primary involvement (e.g., hepatitis or a liver abscess) or secondary infiltration (e.g., miliary tuberculosis) of the liver.

Skin tests Skin testing may aid in diagnosis if other positive results are obtained or if anergy, a characteristic finding in sarcoidosis, Whipple disease, and Hodgkin disease, is demonstrated.

Spinal fluid examination Examination of the spinal fluid is usually unrewarding unless the patient has CNS signs or symptoms, such as a headache or a stiff neck.

Radiologic Studies

In patients with FUO, various radiologic studies in addition to chest films may assist in making the diagnosis:

1. Ultrasonography and CT scans are valuable for detecting intra-abdominal and pelvic abscesses and retroperitoneal adenopathy. The use of CT scans has led to a decrease in the number of biopsies of normal tissues performed in patients with FUO.[96]
2. Radionuclide scans may also be helpful.[97] Bone scans are considerably more sensitive than bone x-rays in the detection of osseous metastases or foci of osteomyelitis. Gallium scans are occasionally useful in detecting occult abscesses, but gallium scanning has had mixed results in patients with FUO.[98] A limited number of studies have examined the usefulness of scans utilizing indium-111–labeled leukocytes in the diagnosis of FUO; the specificity of this test, although not high, appears adequate as a second-line investigative tool.[99] Scans employing indium-111–labeled human immunoglobulin G[100] and technetium-99m–labeled ciprofloxacin are also being investigated.[101] Positron emission tomography may also prove helpful in selected cases.[63,102]
3. Intravenous pyelograms may indicate intrarenal or perirenal abscesses or renal tumors. Some parenchymal lesions can be demonstrated only with the use of a renal angiogram.
4. Upper and lower GI tract x-rays may indicate regional enteritis, ulcerative colitis, or large-bowel neoplasms.
5. Bone x-rays generally are not helpful in the absence of skeletal symptoms.
6. Other radiographic examinations, such as cholangiograms, angiograms, and lymphangiograms, may occasionally be useful but should be performed only when clinical clues suggest specific diagnoses.

Biopsies

All biopsy specimens should be cultured for bacteria, mycobacteria, and fungi and examined histologically. Biopsies that may help determine the diagnosis in patients with FUO include the following:

1. Percutaneous liver biopsy. Neoplastic or granulomatous diseases can generally be detected more easily by percutaneous liver biopsy than by bone marrow biopsy, particularly if hepatomegaly, abnormalities of hepatic function, or both are present.
2. Bone marrow biopsy. This technique may also be used to detect granulomatous diseases or metastatic tumor; it may be rewarding in patients in whom hematologic abnormalities are evident.
3. Lymph node biopsy. Enlarged, matted, or unusually situated nodes are the most favorable for biopsy. Because of the frequent occurrence of chronic inflammation in inguinal lymph nodes, biopsy of these nodes is less satisfactory.
4. Skin and muscle biopsy. Biopsy of skin and muscle may prove useful in diagnosing collagen vascular disease (e.g., polyarteritis or dermatomyositis); biopsy may also be useful in diagnosing sarcoidosis.
5. Temporal artery biopsy. Biopsy of the temporal artery may be the only way to establish the diagnosis of giant cell arteritis in an elderly patient who has FUO, an elevated erythrocyte sedimentation rate, and some thickening of the temporal artery.

Exploratory Laparotomy

In the past, laparotomy was advocated and employed successfully in the diagnosis of FUO. However, noninvasive radiologic techniques, especially when combined with percutaneous needle biopsies (see above), have supplanted laparotomy for the diagnosis of FUO. Laparotomy should be reserved for patients in whom the clinical and laboratory findings point to an intra-abdominal or retroperitoneal source for the fever, particularly when the fever has followed a prolonged and debilitating course.

TREATMENT

Fever is a symptom, not an illness, and treatment of patients with FUO is directed at the underlying illness. Chemotherapeutic trials have generally proved more misleading than helpful when applied to the patient with prolonged FUO. Coincidental temporary defervescence can suggest a specific therapeutic response, thus delaying measures that may provide the correct diagnosis. Occasionally, a therapeutic trial may be reasonable when directed at a specific diagnosis. Thus, a 1- to 2-week trial of a penicillin or vancomycin and an aminoglycoside may be employed when endocarditis is a realistic possibility. Aspirin may be tried in patients who may have adult-type juvenile rheumatoid arthritis. Patients with disseminated tuberculosis presenting as FUO often show a clinical response within 2 weeks after appropriate chemotherapy.

UNDIAGNOSED FUO

In 10% to 15% of patients with FUO, a detailed workup fails to reveal the diagnosis.[103] In about half these cases, the fever resolves spontaneously. Reevaluation of the patient some weeks or

even months later may provide the diagnosis. The prognosis of patients with undiagnosed FUO is surprisingly good[103]; few require empirical corticosteroid therapy, and many can be managed symptomatically with NSAIDs.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Mackowiak PA, Bartlett JG, Borden EC, et al: Concepts of fever: recent advances and lingering dogma. Clin Infect Dis 225:119, 1997

2. Webb P: The physiology of heat regulation. Am J Physiol 268:R838, 1995

3. Boulant JA: Role of the preoptic-anterior hypothalamus in thermoregulation and fever. Clin Infect Dis 31:S157, 2000

4. Rabinowitz RP, Cookson ST, Wasserman SS, et al: Effects of anatomic site, oral stimulation, and body position on estimates of body temperature. Arch Intern Med 156:777, 1996

5. Craig JV, Lancaster GA, Taylor S, et al: Infrared ear thermometry compared with rectal thermometry in children: a systematic review. Lancet 360:603, 2002

6. Craig JV, Lancaster GA, Williamson PR, et al: Temperature measured at the axilla compared with rectum in children and young people: systematic review. BMJ 320:1174, 2000

7. Simon HB: Hyperthermia. N Engl J Med 329:483, 1993

8. Aronoff DM, Neilson EC: Antipyretics: mechanisms of action and clinical use in fever suppression. Am J Med 111:304, 2001

9. Simon HB: Extreme pyrexia. JAMA 236:2419, 1976

10. Crandall CG, Vongpatanasin W, Victor RC: Mechanism of cocaine-induced hyperthermia in humans. Ann Intern Med 136:785, 2002

11. Silva ES: The thermogenic effect of thyroid hormone and its clinical implications. Ann Intern Med 139:205, 2003

12. Simon HB, Daniels GH: Hormonal hyperthermia: endocrinologic causes of fever. Am J Med 66:257, 1979

13. Bouchama A, Knochel JP: Heat stroke. N Engl J Med 346:1978, 2002

14. Wallace RF, Kriebel D, Punnett L, et al: The effects of continuous hot weather training on risk of exertional heat illness. Med Sci Sports Exerc 37:84, 2005

15. Epstein Y, Moran DS, Shapiro Y, et al: Exertional heat stroke: a case series. Med Sci Sports Exerc 31:224, 1999

16. Heat-related mortality—Arizona, 1993–2002, and United States, 1979–2002. MMWR 54:628, 2005

17. Curriero FC, Heiner KS, Samet JM, et al: Temperature and mortality in 11 cities of the eastern United States. Am J Epidemiol 155:80, 2002

18. Sawka MN, Montain SJ: Fluid and electrolyte supplementation for exercise heat stress. Am J Clin Nutr 72:564S, 2000

19. Bailes JE, Cantu RC, Day AL: The neurosurgeon in sport: awareness of the risks of heatstroke and dietary supplements. Neurosurgery 51:283, 2002

20. Epstein Y: Predominance of type II fibres in exertional heat stroke. Lancet 350:83, 1997

21. Duthie DJ: Heat-related illness. Lancet 352:1329, 1998

22. Costrini A: Emergency treatment of exertional heatstroke and comparison of whole body cooling techniques. Med Sci Sports Exerc 22:15, 1990

23. Weiner JS, Khogali M: A physiological body-cooling unit for treatment of heat stroke. Lancet 1:507, 1980

24. Caroff SN, Mann SC: Neuroleptic malignant syndrome. Med Clin North Am 77:185, 1993

25. Guy E, Bandi V: Recognizing drug-induced hyperthemia syndrome: how to distinguish between the possible toxicities. J Crit Illness 17:467, 2002

26. Haddow AM, Harris D, Wilson M, et al: Clomipramine induced neuroleptic malignant syndrome and pyrexia of unknown origin. BMJ 329:1333, 2004

27. Ueda M, Hamamoto M, Nagayama H, et al: Susceptibility to neuroleptic malignant syndrome in Parkinson's disease. Neurology 52:777, 1999

28. Boeve BF, Rummans TA, Philbrick KL, et al: Electrocardiographic and echocardiographic changes associated with malignant catatonia. Mayo Clin Proc 69:645, 1994

29. Rosenberg MR, Green M: Neuroleptic malignant syndrome: review of response to therapy. Arch Intern Med 149:1927, 1989

30. Boyer EW, Shannon M: The serotonin syndrome. N Engl J Med 352:1112, 2005

31. Litman RS, Rosenberg H: Malignant hyperthermia: update on susceptibility testing. JAMA 293:2918, 2005

32. Dinarello CA: Cytokines as endogenous pyrogens. J Infect Dis 179(suppl 2):S294, 1999

33. Mackowiak PA: Concepts of fever. Arch Intern Med 158:1870, 1998

34. Netea MG, Kullberg BJ, Van der Meer JW: Circulating cytokines as mediators of fever. Clin Infect Dis 31:S178, 2000

35. Blattels CM, Sehic E, Li S: Pyrogen sensing and signaling: old views and new concepts. Clin Infect Dis 31(suppl 5):S168, 2000

36. Mackowiak PA: Fever: blessing or curse? A unifying hypothesis. Ann Intern Med 120:1037, 1994

37. Andersen AM, Vastrup P, Wohlfahrt J, et al: Fever in pregnancy with risk of fetal death: a cohort study. Lancet 360:1552, 2002

38. Botto LD, Lynberg MC, Erickson JD: Congenital heart defects, maternal febrile illness, and multivitamin use: a population-based study. Epidemiology 12:482, 2001

39. Vestergaad M, Basso O, Henriksen TB, et al: Risk factors for febrile convulsions. Epidemiology 13:282, 2002

40. Offringa M, Moyer VA: Evidence based management of seizures associated with fever. BMJ 323:1111, 2001

41. Lahat E, Goldman M, Barr J, et al: Comparison of intranasal midazolam with intravenous diazepam for treating febrile seizures in children: prospective randomized study. BMJ 321:83, 2000

42. Rosman NP, Colton T, Labazzo J, et al: A controlled trial of diazepam administered during febrile illnesses to prevent recurrence of febrile seizures. N Engl J Med 329:79, 1993

43. Chang YC, Guo NW, Wang ST, et al: Working memory of school-aged children with a history of febrile convulsions: a population study. Neurology 57:37, 2001

44. Hajat C, Hajat S, Sharma P: Effects of poststroke pyrexia on stroke outcome: a meta-analysis of studies in patients. Stroke 31:410, 2000

45. Kammersgaard LP, Jorgensen HS, Rungby JA, et al: Admission body temperature predicts long-term mortality after acute stroke. Stroke 33:1759, 2002

46. De Keyser J: Antipyretics in acute ischaemic stroke. Lancet 352:6, 1998

47. Hammer MD, Krieger DW: Acute ischemic stroke: is there a role for hypothermia? Cleve Clin J Med 69:770, 2002

48. Shann F. Hypothermia for traumatic brain injury: how soon, how cold, and how long? Lancet 362:1950, 2003

49. Shankaran S, Laptook AR, Ehrenkranz RA, et al: Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. N Engl J Med 353:1574, 2005

50. Boddicker KA, Zang Y, Zimmerman B, et al: Hypothermia improves defibrillation success and resuscitation outcomes from ventricular fibrillation. Circulation 111:3195, 2005

51. Hasday JD, Garrison A: Antipyretic therapy in patients with sepsis. Clin Infect Dis 31:S234, 2000

52. Plaisance KI, Mackowiak PA: Antipyretic therapy: physiologic rationale, diagnostic implications, and clinical consequences. Arch Intern Med 160:449, 2000

53. Simmons DL, Wagner D, Westover K: Nonsteroidal anti-inflammatory drugs, acetaminophen, cyclooxygenase 2, and fever. Clin Infect Dis 31(suppl 5):S211, 2000

54. Chandrasekharan NV, Dai H, Roos KLT, et al: COX-3, a cyclooxygenase-1 variant inhibited by acetaminophen and other analgesic/antipyretic drugs: cloning, structure, and expression. Proc Natl Acad Sci USA 99:13926, 2002

55. Lenhardt R, Negishi C, Sessler DI, et al: The effects of physical treatment on induced fever in humans. Am J Med 106:550, 1999

56. Axelrod P: External cooling in the management of fever. Clin Infect Dis 31(suppl 5):S224, 2000

57. O'Donnell J, Axelrod P, Fisher C, et al: Use and effectiveness of hypothermia blankets for febrile patients in the intensive care unit. Clin Infect Dis 24:1208, 1997

58. Petersdorf RG, Beeson PB: Fever of unexplained origin: report on 100 cases. Medicine (Baltimore) 40:1, 1961

59. Bryan CS: Fever of unknown origin. The evolving definition. Arch Intern Med 163:1003, 2003

60. Knockaert DC, Vanneste LJ, Vanneste SB, et al: Fever of unknown origin in the 1980s: an update of the diagnostic spectrum. Arch Intern Med 152:51, 1992

61. Trivalle C, Chassagne P, Bouaniche M, et al: Nosocomial febrile illness in the elderly. Arch Intern Med 158:1560, 1998

62. Marik PE: Fever in the ICU. Chest 117:855, 2000

63. Knockaert DC, Vanderschueren S, Blockmans D: Fever of unknown origin in adults: 40 years on. J Intern Med 253:263, 2003

64. Borer A, Gilad J, Haikin H, et al: Clinical features and costs of care for hospitalized adults with primary Epstein-Barr virus infection. Am J Med 107:144, 1999

65. Gill S, Melosky B, Haley L, et al: Use of random skin biopsy to diagnose intravascular lymphoma presenting as fever of unknown origin. Am J Med 114:56, 2003

66. Mueller PS, Terrell CL, Gertz MA: Fever of unknown origin caused by multiple myeloma: a report of nine cases. Arch Intern Med 162:1305, 2002

67. Mackowiak PA: Diagnostic implications and clinical consequences of antipyretic therapy. Clin Infect Dis 31(suppl 5):S230, 2000

68. Vanderschueren S, Knockaert DC, Peetermans WE, et al. Lack of value of the naproxen test in the differential diagnosis of prolonged febrile illnesses. Am J Med 115:572, 2003

69. Sorensen HT, Mellernkjaer L, Skriver MV, et al: Fever of unknown origin and cancer: a population-based study. Lancet Oncol 6:851, 2005

70. Pouchot J, Sampalis JS, Beaudet F, et al: Adult Still's disease: manifestations, disease course, and outcome in 62 patients. Medicine (Baltimore) 70:118, 1991

71. Simon HB, Wolff SM: Granulomatous hepatitis and prolonged fever of unknown origin: a study of 13 patients. Medicine (Baltimore) 52:1, 1973

72. Knox TA, Kaplan MM, Gelfand JA, et al: Methotrexate treatment of idiopathic granulomatous hepatitis. Ann Intern Med 122:592, 1995

73. Telenti A, Hermans PE: Idiopathic granulomatosis manifesting as fever of unknown origin. Mayo Clin Proc 64:44, 1989

74. Singh N, Yu VL, Wagener MM, et al: Cirrhotic fever in the 1990s: a prospective study with clinical implications. Clin Infect Dis 24:1135, 1997

75. Sarwari AR, Mackowiak PA: Factitious fever: a modern update. Curr Clin Top Infect Dis 17:88, 1997

76. Drenth JP, Van der Meer JW: Hereditary periodic fever. N Engl J Med 345:1748, 2001

77. Ben-Chetrit E, Levy M: Familial Mediterranean fever. Lancet 351:659, 1998

78. Drenth JP, van der Meer JW: Periodic fevers enter the era of molecular diagnosis (editorial). BMJ 320:1091, 2000

79. Marth T, Raoult D: Whipple's disease. Lancet 361:239, 2003

80. Raoult D, La Scola B, Lecocq P, et al: Culture and immunological detection of *Tropheryma whippelii* from the duodenum of a patient with Whipple disease. JAMA 285:1039, 2001

81. Ramzan NN, Loftus E Jr, Burgart LJ, et al: Diagnosis and monitoring of Whipple disease by polymerase chain reaction. Ann Intern Med 126:520, 1997

82. Ehrbar HU, Bauerfeind P, Dutley F, et al: PCR-positive tests for *Tropheryma whippelii* in patients with Whipple's disease. Lancet 353:2214, 1999

83. Fisher RG, Wright PF, Johnson JE: Inflammatory pseudotumor presenting as fever of unknown origin. Clin Infect Dis 21:1492, 1995

84. Parappil A, Rifaath AA, Doi SAR, et al: Pathan E. Pyrexia of unknown origin: Kikuchi-Fujimoto Disease. Clin Infect Dis 39:138, 2004

85. Mouallem M, Matetzky S, Raichlin E, et al: A man with a prosthetic valve, anaemia, fever, and splenomegaly. Lancet 348:1216, 1996

86. Hamilton I, Nivatpumin P: A 70-year-old man with fatigue. Am J Med 118:20, 2005

87. Roghmann MC, Warner J, Mackowiak PA: The relationship between age and fever magnitude. Am J Med Sci 322:68, 2001

88. Tal S, Guller V, Gurevich A, et al: Fever of unknown origin in the elderly. J Intern Med 252:295, 2002

89. Mourad O, Palda V, Detsky AS: A comprehensive evidence-based approach to fever of unknown origin. Arch Intern Med 163:545, 2003

90. Hirschmann JV: Fever of unknown origin in adults. Clin Infect Dis 24:291, 1997

91. Miralles P, Moreno S, Perez-Tascon M, et al: Fever of uncertain origin in patients infected with the human immunodeficiency virus. Clin Infect Dis 20:872, 1995

92. Armstrong WS, Katz JT, Kazanjian PH: Human immunodeficiency virus–associated fever of unknown origin: a study of 70 patients in the United States and review. Clin Infect Dis 28:341, 1991

93. Benito N, Nunez A, de Gorgolas M, et al: Bone marrow biopsy in the diagnosis of fever of unknown origin in patients with acquired immunodeficiency syndrome. Arch Intern Med 157:1577, 1997

94. Prego V, Glatt AE, Roy V, et al: Comparative yield of blood culture for fungi and mycobacteria, liver biopsy, and bone marrow biopsy in the diagnosis of fever of undetermined origin in human immunodeficiency virus–infected patients. Arch Intern Med 150: 333, 1990

95. O'Brien D, Tobin S, Brown GV, et al: Fever in returned travelers: review of hospital admissions for a 3-year period. Clin Infect Dis 33:603, 2001

96. Rowland MD, Del Bene VE: Use of body computed tomography to evaluate fever of unknown origin. J Infect Dis 156:408, 1987

97. Corstens FH, van der Meer JW: Nuclear medicine's role in infection and inflammation. Lancet 354:765, 1999

98. Knockaert DC, Mortelmans LA, De Roo MC, et al: Clinical value of gallium-67 scintigraphy in evaluation of fever of unknown origin. Clin Infect Dis 18:601, 1994

99. Kjaer A, Lebech AM, Eigtved A, et al: Fever of unknown origin: prospective comparison of diagnostic value of ^{18}F-FDG PET and ^{111}In-granulocyte scintigraphy. Eur J Nucl Med Mol Imaging 31:622, 2004

100. de Kleijn EM, Oyen WJ, Claessens RA, et al: Utility of scintigraphic methods in patients with fever of unknown origin. Arch Intern Med 155:1989, 1995

101. Britton KE, Wareham DW, Das SS, et al: Imaging bacterial infection with ^{99m}Tc-ciprofloxacin. J Clin Pathol 55:817, 2002

102. Blockmans D, Knockaert D, Maes A, et al: Clinical value of [^{18}F]fluoro-deoxyglucose positron emission tomography for patients with fever of unknown origin. Clin Infect Dis 32:191, 2001

103. Knockaert DC, Dujardin KS, Bobbaers HJ: Long-term follow-up of patients with undiagnosed fever of unknown origin. Arch Intern Med 156:618, 1996

104. Jacoby GA, Swartz MN: Fever of undetermined origin. N Engl J Med 289:1407, 1973

105. Larson EB, Featherstone HJ, Petersdorf RG: Fever of undetermined origin: diagnosis and follow-up of 105 cases, 1970–1980. Medicine (Baltimore) 61:269, 1982

106. Vanderschueren S, Knockaert D, Adriaenssens T, et al: From prolonged febrile illness to fever of unknown origin. Arch Intern Med 163:1003, 2003

XXV RESPIRATORY VIRAL INFECTIONS

Frederick G. Hayden, m.d.
Michael G. Ison, m.d., m.sc.

The respiratory tract can be infected by a diverse group of viruses that produce syndromes ranging in severity from mild colds to fulminant pneumonias. Respiratory viral infections are a leading cause of morbidity, hospitalization, and mortality throughout the world; influenza and pneumonia were the most prevalent infectious causes of death during the 20th century in the United States.[1] Respiratory viral infections are also the single most common cause of acute illness and physician visits in the United States and, ironically, a leading cause of prescriptions for antibiotics.[2]

Epidemiology

The high prevalence of respiratory viral infections results from the large number of infectious agents and serotypes and their efficiency of transmission from person to person; from incomplete immunity, with frequent reinfections by some respiratory viruses; and, in the instance of influenza viruses, from frequent changes in viral antigenicity. The frequency of viral respiratory illness is highest in children up to 4 years of age, gradually declines in teenagers, rises again in parents exposed to children, and is generally lowest in the elderly.[3] In families, the average adult contracts two to four acute respiratory illnesses annually, about one quarter of which lead to physician contact. This age-related pattern reflects the gradual acquisition of immunity to some agents and the important role that children play in the transmission of infection. The major reservoir for most respiratory viruses is schoolchildren, who acquire infection in the classroom and introduce it into their homes.

Respiratory viral infections have strong seasonal patterns [*see Table 1*], but sporadic cases or nosocomial outbreaks can occur year-round. The presence or absence of a lipid-containing envelope affects viral survival in the environment. In temperate areas, the enveloped viruses, such as influenza virus, respiratory syncytial virus (RSV), and coronavirus, are characteristically prevalent during midwinter periods, whereas nonenveloped ones, such as rhinoviruses, are found most often in spring through fall. Although respiratory viruses spread from person to person, the relative importance of different routes depends on the virus. Large-droplet spread over short distances (< 1 m) appears common for many viruses. Viruses such as influenza virus and adenovirus spread efficiently in small-particle aerosols. Others, including rhinovirus and RSV, can spread by hand contact with contaminated skin and fomites, with subsequent inoculation onto the nasal mucosa or conjunctiva. A number of respiratory viruses, mainly RSV, influenza virus, and parainfluenza viruses (PIV), cause outbreaks of infection in closed populations, including hospitals, day care centers, and nursing homes. These outbreaks are associated with a high mortality in frail elderly or immunocompromised hosts.

Pathophysiology

Viruses that infect the respiratory tract can be divided into primary respiratory pathogens, whose transmission and replication are generally restricted to the human respiratory tract, and pathogens that affect the respiratory tract as part of a systemic or, sometimes, locally reactivated infection. The latter group includes three types of viruses: (1) those that cause pneumonia occasionally in healthy persons but more often in immunocompromised or pregnant persons, such as measles virus and varicella-zoster virus (VZV) [*see 7:XXVI Herpesvirus Infections*]; (2) those that primarily affect immunocompromised hosts, such as cytomegalovirus (CMV) [*see 7:XXVI Herpesvirus Infections*], herpes simplex virus (HSV) [*see 7:XXVI Herpesvirus Infections*], and, less frequently, human herpesvirus type 6; and (3) several viruses that cause uncommon but severe zoonotic infections, such as severe acute respiratory syndrome coronavirus (SARS-CoV), avian influenza virus, hantaviruses, and zoonotic paramyxoviruses (e.g., Hendra and Nipah viruses) [*see 7:XXXI Viral Zoonoses*]. The primary human respiratory viruses (adenovirus, coronavirus, influenza virus, PIV, human metapneumovirus [hMPV], RSV, and rhinovirus) share global distribution, mucosal sites of infection, and person-to-person transmission, but they differ in many important features, including viral composition,

Table 1 Epidemiologic Features of Principal Human Respiratory Viruses

Virus	*Incubation Period (days)*	*Principal Mode of Spread**	*Peak Seasonality*	*Nosocomial Pathogen*
Adenovirus	4–7	Aerosol, direct	Summer	Yes
Coronavirus (non-SARS)	2–5	Droplet	Winter	Minor
Human metapneumovirus	Probably 2–7	Likely hands, droplet	Early fall to spring	Yes
Influenza A and B viruses	1–4	Aerosol, droplet	Winter	Major
Parainfluenza virus	3–6	Droplet	Types 1 and 2, fall; type 3, spring and summer	Yes
Respiratory syncytial virus	2–8	Hands, droplet	Late fall to spring	Major
Rhinovirus	1–5	Hands, droplet	Early fall, late spring	Yes

*Multiple routes of transmission (i.e., direct, fomites, hand contamination–self inoculation, large droplets, and small-particle aerosols) are possible for most respiratory viruses.
SARS—severe acute respiratory syndrome

seasonality, pathogenesis of disease, and availability of vaccines and specific antiviral agents.

Pathogenesis

Although the extent of viral replication correlates well with severity of illness for most respiratory viruses, the pathogenesis of infection differs for the various groups. Rhinovirus and coronavirus infections are largely limited to the upper respiratory tract, whereas influenza virus, RSV, PIV, SARS-CoV, and adenovirus commonly also infect lower airways . With all respiratory viruses, progression to severe lower respiratory tract disease is more common in patients with impaired immune systems. Respiratory viruses may also produce intra-alveolar inflammatory exudate with hemorrhage, hyaline membrane formation, and differing amounts of epithelial damage. Tracheobronchitis and damage to the respiratory epithelium are typical of influenza virus. RSV infection in young infants is associated with epithelial damage and desquamation, bronchial edema with inflammatory cells, and plugging of small airways. In contrast, damage to the nasal mucosa is modest during rhinovirus or coronavirus colds. Disease manifestations are associated with respiratory tract damage caused by the virus and with host responses to infection, including specific immunologic responses, release of host inflammatory mediators, and neurogenic reflexes. For example, increased levels of histamine, leukotriene C_4, eosinophilic cationic protein, and virus-specific IgE occur in the respiratory secretions of children with RSV bronchiolitis. Influenza is associated with increases in the levels of interleukin-6 (IL-6), interferon gamma (IFN-γ), and tumor necrosis factor–α (TNF-α) in the nasal mucosa and the blood; correlations exist between viral replication, symptom expression, and IL-6 production.[4,5] Severe avian influenza infections in humans have been associated with elevated serum levels of TNF-α, soluble IL-2 receptor, IL-6, and IFN-γ.[6] In persons with colds from rhinoviruses, nasal secretions have higher concentrations of bradykinin, IL-1, IL-6, and IL-8. In addition, respiratory viral infections may alter bacterial colonization patterns, increase bacterial adherence to respiratory epithelium, reduce mucociliary clearance, and alter bacterial phagocytosis by host cells. The impairment of host defenses may foster bacterial infection of areas that are normally sterile, such as the paranasal sinuses, middle ear, and lower respiratory tract.

For many respiratory viruses, the presence of neutralizing antibody in serum and respiratory secretions correlates with protection against infection. Generally, immunity is longer lasting and reinfection less common with virus groups that have many serotypes (e.g., rhinovirus and adenovirus) than in those with only a few serotypes. Although reinfection is frequent with PIV and RSV, severity generally decreases with each episode. In addition to humoral responses, cell-mediated immunity appears to be important for recovery from infection by certain respiratory viruses. In patients who have undergone bone marrow or solid-organ transplantation and in other highly immunocompromised hosts (e.g., patients with acute leukemia or AIDS), many respiratory viruses can cause prolonged infection and severe pulmonary disease, graft dysfunction, and increased mortality.[7]

Diagnosis

CLINICAL MANIFESTATIONS

In general, respiratory viruses cause acute, spontaneously resolving illnesses, although involvement of the lower airways and certain complications (e.g., otitis media, sinusitis, and exacerbations of asthma or chronic obstructive pulmonary disease [COPD]) are common. Clinical diagnosis is difficult because there are a number of respiratory viruses that can cause a variety of overlapping clinical syndromes. Consequently, clinical diagnosis is accurate only under certain circumstances, such as during an influenza epidemic or outbreaks of RSV bronchiolitis. The limited correlation between virus and syndrome emphasizes the importance of rapid virologic techniques for making a specific etiologic diagnosis, particularly in hospitalized or severely affected patients [*see Table 2*]. Improvement in the accuracy of clinical diagnosis will require data from community-based surveillance programs, such as those that track which respiratory viral pathogens are circulating within a community.

LABORATORY TESTS

Most respiratory viruses can be isolated from nose and throat swabs, nasal washes or aspirates, sputum, and other lower respiratory samples. Tests done on nasopharyngeal samples taken early in illness provide the most sensitive results, although early samples in SARS may be falsely negative. However, some respiratory viruses are slow growing [*see Table 2*], and most are pres-

Table 2 Laboratory Methods for Diagnosis of Respiratory Viral Infections

Virus	*Specimen Source*	*Time to Isolation in Cell Culture (days)*	*Rapid Detection Methods and Molecular Techniques*	*Common Serologic Assays*
Adenovirus	Respiratory tract, throat, conjunctiva, urine, stool, blood	2–10	EIA, IF, RT-PCR, IC	CF, IF, NT
Coronavirus (non-SARS)	Nasopharynx	Not routine	Not routine, RT-PCR, IF	Not routine (ELISA)
Influenza A and B viruses	Nasopharynx, throat, lower respiratory tract	2–5	EIA, IF, neuraminidase assay, RT-PCR	CF, HAI, ELISA
Parainfluenza virus	Nasopharynx, lower respiratory tract	3–14	IF, RT-PCR	CF, HAI, ELISA
Respiratory syncytial virus	Nasopharynx, lower respiratory tract	3–14	EIA, IF, RT-PCR	CF, ELISA, NT
Rhinovirus	Nasopharynx	2–7	Not routine, RT-PCR	Not routine (NT)

CF—complement fixation EIA—enzyme immunoassay ELISA—enzyme-linked immunosorbent assay HAI—hemagglutination inhibition IC—immunochromatography IF—immunofluorescence NT—neutralization RT-PCR—reverse transcriptase–polymerase chain reaction SARS—severe acute respiratory syndrome

ent in lower titers in adults than in children, so samples from adults may require longer periods for isolation or have lower yields on rapid diagnostic tests. Because temperature fluctuations and freezing may cause loss of infectivity, samples should be transported at refrigerator temperatures (4° C) or on wet ice. Cell culture inoculation (shell vials) followed by antigen detection has a sensitivity of 80% or greater at 1 to 2 days for adenovirus, influenza virus, PIV, and RSV.[8]

Immunofluorescence testing of exfoliated respiratory cells and enzyme immunoassay for viral antigens are reasonably sensitive, specific, and rapid techniques for detecting some respiratory viruses, particularly RSV and influenza virus.[9] Nucleic acid amplification techniques appear to be the most sensitive direct detection methods for most viruses, including SARS-CoV. However, false negative results are possible (during the first days of infection, SARS-CoV may be undetectable even by polymerase chain reaction), and positive results must be confirmed to ensure accuracy. For simultaneous detection of respiratory viruses (influenza virus types A and B; RSV; and PIV types 1, 2, and 3), multiplex reverse transcriptase PCR (RT-PCR) has good sensitivity,[10] and a commercial assay is available. Real-time quantitative RT-PCR tests offer a sensitive means of detection of many respiratory viruses and, potentially, of monitoring responses to treatment.

Serologic diagnosis requires paired serum specimens in most instances and is intrinsically slower than direct methods. A variety of serologic techniques are used to measure antibodies, including neutralization, hemagglutination inhibition, complement fixation, and enzyme-linked immunosorbent assay (ELISA). Measurement of complement-fixation antibodies is generally less sensitive than the other methods and does not provide a serotype-specific diagnosis. Immunocompromised hosts often fail to develop diagnostic increases in antibody titers. The combined use of cultures, antigen detection, nucleic acid detection, and serology provides the most comprehensive approach for identification of important respiratory viral pathogens in hospitalized adults.

Clinical Syndromes

COMMON COLDS

Diagnosis

The common cold is generally a mild illness of the upper respiratory tract, primarily affecting the nasopharynx and paranasal sinuses. Rhinoviruses, which cause about 50% of colds, and coronaviruses, which cause 10% to 20%, are the most important pathogens. RSV, PIV, influenza virus, adenovirus, and some enteroviruses also cause colds.

Adults with rhinovirus colds typically have prominent upper respiratory tract symptoms (i.e., sneezing, nasal discharge, nasal obstruction, sore or scratchy throat, and cough) and often have headache but little fever or few systemic complaints other than malaise. Physical findings are nonspecific and include nasal discharge and mucosal erythema. Although generally absent in adults, lower respiratory tract manifestations (e.g., cough, sputum, and wheezing) do occur in about 60% of cases in elderly persons, and lower respiratory tract illness and fever are common in young children infected with rhinovirus.

The differential diagnosis of the common cold includes allergic rhinitis, bacterial nasopharyngitis in infants, and, in rare instances, nasal diphtheria or bacterial nasal infection in adults. Because nearly 50% of colds last longer than 1 week and 25% last up to 2 weeks, it is often unclear whether a complicating viral or bacterial sinusitis has occurred. More than 80% of patients with uncomplicated colds have reversible sinus abnormalities demonstrable on computed tomography,[11] and sinusitis is an inherent part of rhinovirus colds. Fever, purulent nasal drainage, cough productive of purulent sputum, increasing symptoms or malaise after 1 week, localized facial pain, or maxillary toothache suggests a bacterial sinusitis.[12] Otitis media complicates approximately 2% of colds in adults and 5% of colds in children.[13]

Treatment

Treatment of the common cold is directed toward specific symptoms. Saline nose drops may be helpful for relief from obstructing secretions, particularly in young children, and warm saline gargles reduce sore throat. Topical decongestants such as phenylephrine and the longer-acting oxymetazoline provide prompt relief of nasal obstruction but may be associated with rebound congestion and possibly throat irritation. Oral pseudoephedrine also provides partial relief of nasal obstruction but may cause anxiousness and insomnia in some patients. The oral decongestant phenylpropanolamine was removed from the market because of increased risk of hemorrhagic stroke. It is uncertain whether topical or oral vasoconstrictors alter the risk of bacterial complications. First-generation oral antihistamines, such as chlorpheniramine and clemastine, reduce sneezing and, to a lesser extent, rhinorrhea, but they are often associated with excess sedation. Nonsedating antihistamines such as loratadine and astemizole are not effective for colds in nonallergic persons. Intranasal ipratropium, a quaternary anticholinergic, reduces cold-associated rhinorrhea by about 30% and may also reduce sneezing, but excess nasal drying may result in blood-tinged mucus.[14] Cough suppressants with codeine or dextromethorphan may be used if the cough is severe or disturbs sleep, but they should be used cautiously in patients with underlying chronic obstructive disease. Specific antiviral therapy is not currently available. Two intranasal antiviral agents from different classes, the 3-C protease inhibitor ruprintrivir and the capsid inhibitor pleconaril, are being investigated. Although effective in treating rhinovirus colds, oral pleconaril was not approved by the Food and Drug Administration because of concerns about safety and drug interactions, most notably with oral contraceptives.[15] A number of other remedies (e.g., *Echinacea* preparations, zinc lozenges or nasal gels, and hot-air inhalation) are of unproven or doubtful value.[16-18]

PHARYNGITIS AND LARYNGITIS

Pharyngitis usually is associated with the same viruses that cause common colds and occurs concurrently with the common cold, but it is also often a prominent complaint of persons with adenovirus and influenza virus infections. Enteroviruses are also important causes of fever and pharyngitis during the summer and early fall.

Diagnosis

Pharyngitis may be associated with characteristic clinical findings, such as palate vesicles and ulcers caused by coxsackievirus-induced herpangina, acute ulcerative stomatitis and pharyngitis caused by HSV, or exudative pharyngitis related to Epstein-Barr virus mononucleosis. Pharyngitis occurs with primary HIV infection and may be associated with mucosal ero-

sions and lymphadenopathy. Pharyngitis caused by group A streptococci and other bacteria (anaerobes, *Corynebacterium diphtheriae, Chlamydophila pneumoniae,* and *Mycoplasma pneumoniae*) are generally not associated with acute rhinorrhea [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*].

Treatment

Treatment for most cases of viral pharyngitis is symptomatic. However, HSV stomatitis, particularly in immunocompromised patients, should be treated with specific antiviral therapy (e.g., acyclovir, famciclovir, or valacyclovir).

Hoarseness caused by laryngitis frequently occurs during acute viral infections and is more common during infections caused by viruses that invade the lower airways. Treatment is primarily voice rest.

ACUTE BRONCHITIS

Diagnosis

Acute bronchitis manifests as a severe or prolonged cough that is usually nonproductive or productive of scant mucoid sputum. Bronchitis can follow infection by any of the respiratory viruses and is a hallmark of influenza. Viral infections commonly cause exacerbation of preexisting airway disease, including chronic bronchitis, emphysema, asthma, and cystic fibrosis. Rhinoviruses are the most frequently implicated viral pathogens.[19] HSV can cause bronchospasm and sometimes ulcerative tracheobronchitis in immunocompromised persons and, less often, in immunocompetent persons.

Cough production in acute bronchitis results from direct viral damage to the respiratory mucosa, release of inflammatory mediators, stimulation of airway irritant receptors, and increased production and decreased clearance of respiratory secretions. Infection may also enhance airway hyperreactivity, characterized by increased sensitivity to cold air and pollutants such as smoke.

The differential diagnosis of acute viral bronchitis includes bacterial, mycoplasmal, chlamydial, and pertussis infections, as well as noninfectious entities such as asthma.

Treatment

Treatment of acute bronchitis in otherwise healthy persons is directed at suppression of the cough with codeine or dextromethorphan. However, these agents should be used cautiously in patients with underlying chronic obstructive disease. Antibiotics are generally not indicated.[20]

INFLUENZA SYNDROME

Diagnosis

The influenza syndrome is characterized by the rapid onset of constitutional and respiratory tract symptoms. Constitutional symptoms, which include fever, chills, prostration, muscle aches, and headache, tend to dominate during the first several days of illness. Respiratory complaints are sore throat, coryza, hoarseness, and, especially, cough. Persistent nonproductive cough, easy fatigability, and asthenia are common in the second week of illness.

Epidemic influenza A and B viruses are the principal causes of the influenza syndrome, but other respiratory viruses, including RSV, PIV, adenovirus, and, sometimes, rhinovirus can cause influenzalike illness.

Treatment

Treatment for the influenza syndrome is usually symptomatic and includes bed rest, oral hydration, antipyretics, and antitussives. Specific antiviral therapy for influenza virus infection is available [*see* Influenza Virus, *below*].[21-23] Fever should be treated in patients in whom it would contribute to disease, such as children with previous febrile convulsions or patients with preexisting cardiac disease. Aspirin should be avoided in pediatric patients because of its association with Reye syndrome.

CROUP

Diagnosis

Croup, or laryngotracheobronchitis, is a characteristic pediatric syndrome characterized by brassy or barking cough, inspiratory stridor, dyspnea, and hoarseness. Symptoms are often preceded by several days of upper respiratory tract illness and typically become worse at night. Croup is seen primarily in children younger than 5 years and is associated most closely with PIV infections and less often with RSV, hMPV, influenza viruses, adenoviruses, and rhinoviruses. Measles virus is an important cause of severe croup. Acute spasmodic croup is characterized by recurrent attacks precipitated by viral infections, allergies, and, possibly, other factors. The differential diagnosis of croup includes foreign-body aspiration or trauma, bacterial epiglottitis or tracheitis, localized abscess, and angioedema.

Inflammation of the larynx and trachea with subglottic narrowing causes localized obstruction; this obstruction is enhanced by inspiration, because of the extrathoracic location of the obstruction and the compliance of the airway walls in children. In addition, inflammatory changes occur throughout the lower respiratory tract. Symptoms usually abate in 3 to 4 days.

Treatment

Most cases of croup can be managed at home, with moist air. Correction of hypoxemia is essential for hospitalized patients. Aerosolized racemic epinephrine provides transient benefit for those with persistent stridor. Systemic glucocorticoids appear to be effective in moderating the severity of illness and shortening hospitalization; inhaled steroids are beneficial in less severe cases.

BRONCHIOLITIS

Diagnosis

Bronchiolitis is an acute inflammatory disorder of the small airways that is characterized by airway obstruction with wheezing, hyperinflation, and atelectasis. It is associated most closely with RSV infection, which is detected in about 75% of infants hospitalized with bronchiolitis, and less often with rhinoviruses, hMPV, PIV, influenza viruses, and adenoviruses. The peak incidence is in infants from 2 to 6 months of age, and over 80% of cases occur in the first year of life. Premature infants and young children with underlying pulmonary conditions are at high risk for RSV bronchiolitis or pneumonia. After several days of coryza and possibly fever, patients experience acute respiratory distress with wheezing, cough, and, often, inspiratory rales; apnea may occur in babies. Characteristic radiologic features are atelectasis and signs of hyperinflation. In mild cases, bronchiolitis symptoms resolve within several days, but prolonged or severe disease may occur in premature infants or those with underlying cardiopulmonary or immunodeficiency disorders. The

major differential diagnosis is asthma, which is uncommon in children younger than 2 years.

The bronchiolitis syndrome results from viral destruction of the respiratory epithelium and associated peribronchiolar mononuclear inflammation, which is perhaps in part immunologically mediated. In infants, because of their disproportionately narrow airways, bronchiolar obstruction causes distal collapse or air trapping and abnormal gas exchange.

Treatment

Correction of hypoxemia is the most important aspect of managing RSV bronchiolitis. Bronchodilators, glucocorticoids, and routine antibiotics are not of proven value. Aerosolized ribavirin therapy is available for selected infants hospitalized with RSV bronchiolitis or pneumonia, but it is an expensive intervention of uncertain value.[24]

REACTIVE AIRWAY DISEASE EXACERBATION

Infection with respiratory viruses, most commonly rhinoviruses,[19] can precipitate exacerbations of chronic respiratory illness. Respiratory viral infections, half of which are rhinovirus infections, can worsen asthma in patients of all ages but are associated with up to 85% of asthma exacerbations in schoolchildren.[25] Young children hospitalized with wheezing illness are in most cases infected with rhinoviruses, RSV, hMPV, or PIV. Exacerbations of COPD often result from respiratory viral infections. Approximately 25% of hospitalizations for COPD exacerbations have been associated with a documented respiratory tract viral infection; this percentage increases to 45% of COPD patients admitted between December and March.[26]

PNEUMONIA

Viral infections have been associated with at least 10% to 15% of community-acquired pneumonias in hospitalized adults and with as many as half of cases in hospitalized infants.[27] The relative importance of different viruses depends heavily on the geographic location of the outbreak, the season, and patient age and immune status. Influenza A and, less often, influenza B viruses are the most common causes of community-acquired viral pneumonia in adults and the elderly. In children younger than 3 years, RSV accounts for about 50% of all hospitalizations for pneumonia[28]; PIV type 3 is also prominent. PIV, hMPV, adenovirus, rhinovirus, and non-SARS respiratory coronavirus infections have been implicated as causes of pneumonia in children and adults.

Respiratory viruses reach the lung by contiguous spread from the upper respiratory tract or inhalation of small-particle aerosols. Depending on the agent, nonrespiratory viruses infect the lung parenchyma by hematogenous dissemination (e.g., CMV, VZV, HSV, and measles) or, less often, by contiguous spread from the tracheobronchial tree (e.g., HSV).

SARS

SARS is a unique form of viral pneumonia. In contrast to most other viral pneumonias, upper respiratory symptoms are usually absent in SARS, although cough and dyspnea occur in most patients. Typically, patients present with a nonspecific illness manifesting fever, myalgia, malaise, and chills or rigors; watery diarrhea may occur as well[29] [*see* SARS Coronavirus, *below*].

Diagnosis

The clinical and radiographic features of most viral pneumonias are not distinctive. Unless a typical exanthem of measles or varicella is present, identification of a specific pathogen requires a careful epidemiologic history and appropriate virologic studies. The clinical and radiologic appearance of viral pneumonia is often indistinguishable from that of bacterial infection, and mixed infections are common. Differentiation between pure viral pneumonias, mixed viral-bacterial pneumonias, and viral infections of the tracheobronchial tree with complicating bacterial pneumonia is usually not possible on clinical grounds.

When taking the history, the clinician should ask about risk factors for exposure, including illness in contacts, institutional residence during an outbreak of illness, travel history, and contact with animals associated with zoonotic transmission. Knowledge of local patterns of viral circulation (e.g., epidemic influenza) and of the patient's immune status are also important in diagnosis. SARS should be suspected in any patient with radiographically confirmed pneumonia who has one or more epidemiologic risk factors for the syndrome[30] [*see* SARS Coronavirus, *below*].

Treatment

Management of viral pneumonia is primarily supportive; ventilatory support has improved the prognosis for patients with extensive viral pneumonia. Specific antiviral therapy for respiratory viruses is limited, and there are no controlled data regarding the use of these agents in pneumonia. Antiviral agents are available to treat influenza[31-33] [*see* Influenza Virus, *below*]. Inhaled ribavirin is available for RSV and other viral infections but is of uncertain value; it has been used intravenously for adenovirus, hantavirus, measles, PIV, and influenza virus infections. Intravenous acyclovir is effective in HSV or VZV pneumonia.

Pneumonia in Immunocompromised Hosts

Immunocompromised hosts, particularly patients undergoing organ transplantation, experiencing chemotherapy-induced neutropenia and immunosuppression, or with HIV-related immunosuppression, are at risk for reactivation of latent viruses (e.g., HSV, CMV, or adenovirus) or acquisition of community respiratory viruses. Bronchoalveolar lavage yields viruses in up to half of immunocompromised patients with acute pneumonia.[34] CMV is the most frequently recovered, accounting for more than 80% of isolates; fewer than 10% of isolates are HSV or various respiratory viruses.[35] In immunosuppressed patients, influenza viruses, PIV, RSV, hMPV, rhinovirus, and adenoviruses can cause severe pneumonia[36] that is often nosocomially acquired and usually associated with a preceding upper respiratory tract illness.[7] It seems prudent to delay chemotherapy or transplantation in the presence of an upper respiratory tract illness.

Intravenous acyclovir is recommended for patients with HSV or VZV pneumonia. Ganciclovir with immunoglobulin appears to be effective in reducing the mortality associated with CMV pneumonia in bone marrow transplant (BMT) recipients and has been used as monotherapy in other patient groups.[37] Aerosolized ribavirin and RSV-specific antibodies (RSV immunoglobulin [RespiGam] or palivizumab), if given early, may be effective for preventing progression of RSV pneumonia or treating established disease in such patients [*see* Respiratory Syncytial Virus, *below*]. Aerosolized and intravenous ribavirin have been used for adenovirus[38] or severe measles. Intravenous cidofovir has been effective in the management of severe adenoviral infection in immunocompromised patients but may cause nephrotoxicity.[39]

Infections Caused by Specific Agents

ADENOVIRUS

Adenoviruses cause a variety of respiratory tract syndromes, ranging from pharyngoconjunctival fever (often contracted while swimming in contaminated water) to severe pneumonia in infants, immunosuppressed patients, and, uncommonly, healthy adults. Epidemics of acute respiratory disease and pneumonia caused by adenovirus have been recognized for decades in United States military recruits and are expected to continue, given the cessation of immunization programs in the military. However, a new manufacturer for adenovirus vaccine has been identified, and vaccination of military personnel is expected to resume.[40]

Classification and Pathogenesis

Adenoviruses are medium-sized (65 to 80 nm), nonenveloped, double-stranded DNA viruses, of which 49 antigenic types (divided into six subgroups) are associated with human infections. The protein coat of the virus is composed of hexagonal and pentagonal subunits (capsomers) with long fibers at each vertex, which are the sites of host cell attachment. Type-specific antigens, which give rise to neutralizing antibody, are present on the hexons and fibers of the capsid. The hexons also contain a group-specific antigen with cross-reactivity among most adenoviruses. Only about one half of serotypes cause disease. Immunity correlates with the presence of type-specific neutralizing antibody.

Epidemiology and Transmission

Adenovirus infections may be endemic or epidemic. Lower-numbered serotypes routinely infect infants and young children, although types 3, 4, and 7 are also acquired later in life and are typically associated with epidemic respiratory disease. In addition to transmission by direct contact with respiratory secretions or infectious aerosols, fecal-oral transmission occurs. Infection may be acquired by pharyngeal or conjunctival inoculation from contaminated water. Adenovirus infection occurs throughout the year but is often recognized during the summer in association with outbreaks of pharyngitis or bronchitis. Adenovirus probably accounts for about 5% of total acute respiratory infections in civilian adults. Nosocomial transmission, including clusters of pneumonia in long-term care facilities and large outbreaks of epidemic keratoconjunctivitis, is well documented.[41,42] Persistent infections occur in the tonsils and gastrointestinal (GI) tract, and prolonged viral shedding is common in immunocompromised hosts and in the GI tract of children.

Diagnosis

Clinical features The incubation period for adenovirus infection of the respiratory tract is usually 4 to 7 days. Adenovirus respiratory infection typically causes a moderate to severe pharyngitis, which is sometimes exudative, and tracheobronchitis. Fever and systemic symptoms are often prominent, and rhinitis, cervical adenitis, and follicular conjunctivitis are common. Pharyngoconjunctival fever, usually associated with serotypes 3, 4, and 7, tends not to be accompanied by lower respiratory tract symptoms. In infants and young children, infections also cause a pertussislike syndrome, croup, and bronchiolitis.

Adenovirus pneumonia is similar to other viral pneumonias; it usually begins with pharyngitis, rhinitis, fever, and cervical adenopathy—conjunctivitis may or may not be present—and then spreads to involve the lower airways and lungs. Bilateral interstitial or alveolar infiltrates are common, although focal consolidation or effusion also occurs.

Laboratory findings The presence of cells with large basophilic, intranuclear inclusions (smudge cells) in lower respiratory tract biopsy specimens may provide early histopathologic diagnosis. A variety of serotypes have been recovered from stool, urine, and lower respiratory tract specimens of immunocompromised patients.[43] Rapid methods for virus detection include enzyme immunoassay (EIA), immunofluorescence, antigen immunochromatography,[44] and RT-PCR. Quantitative PCR to measure blood adenoviral DNA levels is emerging as a useful marker for predicting the risk of progression of disease and for monitoring responses to antiviral therapy in immunocompromised patients.

Complications

Acute complications, which occur mainly in children and in immunocompromised hosts, are leukopenia, lymphocytopenia, thrombocytopenia, rhabdomyolysis, disseminated intravascular coagulation, renal failure, and bacterial infection. Survivors may acquire restrictive lung disease or bronchiectasis; children in particular are at risk for bronchiolitis obliterans. Mortality in BMT recipients with adenovirus pneumonia is as high as 60%.[45]

Adenoviruses are also associated with extrarespiratory infections, including epidemic keratoconjunctivitis; hepatitis and genitourinary infections in immunosuppressed hosts; myocarditis; arthritis; meningoencephalitis; and, especially in children, hemorrhagic cystitis, mesenteric adenitis, intussusception, and gastroenteritis. Adenovirus commonly complicates hematopoietic stem cell transplantation, especially in children, and is often associated with invasive disease, which is associated with reduced survival.[46]

Treatment

Antiviral therapy of proven value is not available, but intravenous ribavirin, ganciclovir, cidofovir, and immunoglobulin have been used to treat severe infections. In invasive adenoviral infections, including pneumonia, mortality appears lower in patients who have received cidofovir than in those who have received intravenous ribavirin, but further studies are needed.[34,47-49] Treatment is otherwise supportive.

Prevention

Live attenuated vaccines for adenovirus types 4 and 7 were used in the military but are not licensed for civilian use.[40,42] Other control measures are swimming-pool chlorination, careful hand washing, and, in the instance of nosocomial keratoconjunctivitis, use of gloves and proper sterilization of equipment.

CORONAVIRUSES

Classification

Coronaviruses are moderate-sized (100 to 150 nm), enveloped, single-stranded RNA viruses named for their distinctive club-shaped surface projections, which resemble a crown. Three antigenic types of human respiratory coronaviruses (229E, OC43, and NL63) have been identified, separately from SARS-CoV.

Epidemiology and Transmission

Coronaviruses are the second most frequently documented

cause of common colds. Coronaviruses account for 4% to 15% of cases of acute respiratory disease annually and as many as 35% during peak periods. Most coronavirus infections occur in winter and early spring, but infections have been detected throughout the year. Immunity to infection is incomplete, and reinfection is common. SARS-CoV causes a clinical syndrome unique among the coronaviruses and is discussed separately [*see* SARS Coronavirus, *below*].

Diagnosis

Clinical features The incubation period of coronavirus colds ranges from 2 to 5 days, longer than that of rhinovirus colds. Infections produce a typical coryzal illness that is indistinguishable from colds caused by other viruses. Lower respiratory tract manifestations include pneumonia in children, military recruits, and possibly immunocompromised hosts; exacerbation of asthma in children; and exacerbation of chronic airway disease in adults. Reported associations with GI disease, including necrotizing enterocolitis in infants, are controversial.

Laboratory findings Virus isolation is insensitive; infections are usually identified by serology or by detection of coronavirus RNA. Culture of virus in human embryonic trachea cells is the most sensitive cell culture system for recovering human respiratory tract coronaviruses but is impractical.

Treatment

Effective antivirals or vaccines are unavailable, although intranasal IFN-α is protective against experimental coronavirus colds. Treatment is symptomatic.

SARS CORONAVIRUS

Epidemiology

SARS-CoV is a novel coronavirus identified as the cause of a clinical syndrome, SARS, that was first recognized as an unusual atypical pneumonia in Guangdong Province, China, in November 2002. Over the ensuing 6 months, the virus spread to multiple areas of the world and resulted in over 8,000 illnesses and 774 deaths. China, Hong Kong, Singapore, Taiwan, and Toronto, Canada, were most severely affected. The virus is of animal origin, but the primary reservoir remains uncertain. It possibly originated from Himalayan palm civets found in live-animal markets, but it is infectious for a wide variety of species, including civets, raccoon dogs, ferrets, cats, mice, and monkeys.

Molecular epidemiology has shown that SARS viruses from outbreaks in Hong Kong, Vietnam, Singapore, Toronto, and Taiwan are clonally related to the index case from Hong Kong, whereas viruses from Guangdong Province are genetically more diverse. It is possible that different strains of SARS-CoV have different characteristics, including efficiency of transmission and virulence. Future SARS outbreaks may differ from the outbreak of 2002–2003. Unlike diseases from other respiratory coronaviruses, SARS does not have any recognized seasonality.

Transmission

SARS-CoV spreads efficiently between humans. The primary mode of transmission appears to be through close personal contact: infectious droplets from the cough or sneeze of an infected person travel through the air a short distance and land on respiratory or, possibly, conjunctival mucous membranes or are deposited there indirectly. Other possible modes of spread include fomites, with hand contamination and subsequent self-inoculation; small-particle aerosols; and perhaps fecal-oral transmission. Transmission may be amplified by certain aerosol-generating procedures, such as endotracheal intubation, bronchoscopy, and treatment with aerosolized medications. Recently, modeling was used to substantiate airborne spread of SARS as the cause of the large outbreak at the Amoy Gardens housing complex in Hong Kong.[50] Person-to-person spread has been especially common in health care and hospital settings, but transmissions have occurred in homes and hotels, in the workplace, on aircraft, and in taxis. Laboratory transmission of SARS-CoV and presumed acquisition from animals have also occurred.[51] No transmission of SARS has been documented before the onset of symptoms or more than 10 days after resolution of fever.

Diagnosis

SARS is diagnosed on the basis of epidemiologic evidence, clinical suspicion, and diagnostic testing. The current case definition of SARS, as well as the most current recommendations on diagnosis, can be found at the Centers for Disease Control and Prevention Web site (http://www.cdc.gov/ncidod/sars).

Epidemiologic evidence SARS should be included in the differential diagnosis of pneumonia in patients of all ages who have one or more of the following risk factors:

- A history of recent travel to mainland China, Hong Kong, or Taiwan or close contact (within 10 days before symptom onset) with ill persons with a history of recent travel to such areas.
- Employment in an occupation that involves particular risk for SARS-CoV exposure, including health care with direct patient contact or work in a laboratory that contains live SARS-CoV.
- Association with a cluster of cases of atypical pneumonia without an alternative diagnosis during periods of known SARS circulation.

Clinical features The incubation period of SARS typically ranges from 2 to 7 days but may be as long as 10 days or even, in rare cases, 14 days. Patients present with fever, myalgia, malaise, and chills and may have a nonproductive cough. Other upper respiratory symptoms are usually absent. Fever may be absent, particularly in the elderly, in the initial phase of the illness. Diarrhea may occur.

Laboratory findings Lymphocytopenia is common in SARS, as are elevations in D-dimer levels and activated partial thromboplastin time. Chest radiographs show evidence of pneumonia in 60% to 100% of patients. Distinctive radiographic findings include a predominant peripheral lung involvement; initial unilateral (typically, in the right lung) focal air-space opacity; and absence of cavitation, lymphadenopathy, and pleural effusion.[52] High-resolution CT scans are abnormal in up to 67% of patients with normal chest radiographs.

Current diagnostic tests for SARS include RNA detection by RT-PCR, virus culture, and serology. Virus is detectable in respiratory secretions (from the nasopharynx, throat, nose, and lower respiratory tract), as well as in feces, serum, and plasma. Initial RT-PCR techniques were associated with poor sensitivity, particularly in early infection, but improved extraction methods and real-time RT-PCR assays have been developed that allow for

more sensitive testing during the first days of infection.[53] Any positive test result must be confirmed by testing another clinical sample, reextracting and testing the original specimen, or retesting with an assay that targets different parts of the viral genome. Measurement of specific antibodies at 21 days or later after illness onset using whole-virus immunofluorescence or ELISA remains the gold standard for retrospective confirmation of SARS-CoV infection, because a negative RNA assay does not rule out infection.

Treatment

Initial treatment in patients with suspected SARS is with empirical antimicrobials directed against the typical pathogens in community-acquired pneumonia. There have been no prospective studies of antiviral agents in the treatment of SARS. Ribavirin, lopinavir-ritonavir, IFN-α, and traditional Chinese medicines were used during the 2002–2003 outbreak. In vitro studies, animal testing, and limited human experience suggest that IFN-α has activity against SARS-CoV; IFN-β also has efficacy in vitro.[54] Ribavirin is not active in vitro at clinically achievable concentrations and was associated with significant toxicity and no obvious antiviral effects in SARS. Neutralizing antibodies, including recently described humanized monoclonal antibodies, are active in animal models[55] and are promising for prophylaxis and perhaps early treatment.

Lung biopsies in patients with SARS suggest a possible immune component; the samples reveal diffuse alveolar damage with desquamation of pneumocytes, inflammatory infiltrates, edema, and hyaline membrane formation. Consequently, high-dose methylprednisolone has been used to modify the disease course and has produced short-term reductions in fever and infiltrates. However, steroid therapy offers uncertain overall benefit and poses such risks as enhanced viral replication, superinfection, and late aseptic necrosis. Thymic peptides, intravenous immune globulin (IVIg), and plasma from patients in the convalescent phase of illness are likewise of unclear benefit. Neutralizing human monoclonal antibodies against S protein are active in experimental animal infection and may prove useful in prophylaxis or possibly early treatment.[56]

Prevention of exposure is the key to limiting future outbreaks. All patients with suspected SARS should be isolated for at least 10 days; in patients with documented SARS, isolation should be continued for at least 10 days after defervescence. Quarantine of contacts has been used in some outbreaks. Current recommendations regarding appropriate infection control practices can be found at the CDC Web site (http://www.cdc.gov/ncidod/sars). In health care settings, precautions against transmission through airborne respiratory droplets and contact should be strictly enforced, with enhanced airborne precautions used during aerosol-generating procedures.

Prognosis

One third of patients with SARS defervesce and improve clinically. The remainder experience progressive illness, with diarrhea and increasing tachypnea, shortness of breath, and abnormal findings on pulmonary examination and radiography; 20% to 30% of patients require admission to an intensive care unit, usually for mechanical ventilation, and approximately 10% succumb. Death typically results from respiratory failure, multiple organ failure, sepsis, or intercurrent medical illness. More severe disease and higher mortality occur in older adults (≥ 60 years) and those with underlying medical conditions.

HUMAN METAPNEUMOVIRUS

hMPV is a respiratory pathogen that causes infections ranging from colds to severe bronchiolitis and pneumonia. The first metapneumovirus associated with human infection, hMPV was discovered in 2001 when viruses that had been isolated from children presenting with RSV-like illnesses failed to be identified by standard techniques. After random-primer RT-PCR was used, the virus was identified as a new metapneumovirus related to turkey tracheitis virus. Serologic studies demonstrated that by 5 years of age, all children were seropositive and that sera originally collected in 1958 were also positive for hMPV antibodies.[57] Two genetic clusters of hMPV that correspond to two different serotypes of hMPV have been recognized on every continent.[58]

Epidemiology

Infections by hMPV occur worldwide and year-round, with a winter predominance. The season peaks between December and February in the northern hemisphere, with a longer season in temperate climates. hMPV can be isolated from 5.5% to 6.5% of children hospitalized with respiratory symptoms. The peak incidence is between 4 and 6 months of age, with most children being younger than 5 years. Coinfection with hMPV and RSV has been documented frequently, and several cases of concomitant hMPV and SARS-CoV infection have been documented.

Asymptomatic or mild illness appears to be much more common with hMPV. hMPV has been detected in 20% of ambulatory young children presenting with respiratory illness in whom no other cause was found. Most cases occurred between December and April.[59] In infirm elderly patients, the incidence is 6.5%. The incidence of hMPV in general-community persons presenting with influenzalike illness is only 2%, whereas hMPV can be isolated from 3% to 6.6% of the general community with acute respiratory illnesses.

Diagnosis

Clinical features hMPV causes clinical syndromes indistinguishable from RSV, including bronchiolitis, croup, asthma exacerbation, and pneumonia. Studies attempting to differentiate hMPV from other respiratory viruses have found that hMPV infection appears to occur at a slightly older age and causes slightly milder symptoms than RSV. Infections in adults manifest as colds, influenzalike illness, bronchitis, exacerbations of underlying airway disease, or, uncommonly, pneumonia. hMPV infections in immunocompromised patients can be severe and have been associated with at least three deaths.

Laboratory tests RT-PCR has become the standard method of detecting hMPV in respiratory samples. Methods that use primers targeted at the polymerase (L) and fusion (F) protein genes have shown adequate sensitivities. The presence of antibodies is not diagnostic by itself, although seroconversion or a greater than fourfold rise in titer is indicative of recent infection. The virus can be grown in cell culture, but the cytopathic effect may not be seen for up to 14 days.

Treatment

In vitro data suggest that ribavirin has activity against hMPV similar to that against RSV. Likewise, neutralizing antibody titers against hMPV in IVIg are similar to titers against RSV. Clinical studies are required to determine the possible clinical efficacy of ribavirin and IVIg.[60]

Classification and Pathogenesis

Influenza viruses belong to the orthomyxovirus family and consist of types A, B, and C. These medium-sized (80 to 120 nm), enveloped, single-stranded RNA viruses contain eight gene segments (seven for influenza C). The segmented nature of the genome allows reassortment of RNA segments between two influenza viruses during dual infection and facilitates antigenic variation. Surface glycoprotein spikes possess either hemagglutinin or neuraminidase activity. Hemagglutinin mediates cell attachment and fusion of virus and cell membranes. By cleaving terminal sialic acid residues and destroying the receptors recognized by hemagglutinin, neuraminidase promotes release of virus from infected cells and spread within the respiratory tract.

Influenza A viruses are further classified into subtypes on the basis of their surface proteins (15 hemagglutinins and nine neuraminidases are recognized in animal influenza viruses). Three A hemagglutinin subtypes (H1, H2, and H3) and two neuraminidases (N1, N2) have caused extensive human infections. Influenza C viruses have seven gene segments and lack a neuraminidase. Influenza viruses are named by the type; location of isolation; isolate number; year of recovery; and, for influenza A type viruses, the subtype (e.g., A/Texas/36/91[H1N1]).

The surface glycoproteins induce host humoral and cellular immune responses and are responsible for the changing antigenicity of influenza viruses. Two major types of antigenic change can occur: drift and shift. Antigenic drift refers to relatively minor changes in hemagglutinin and, less often, neuraminidase antigenicity that occur frequently (usually every few years) and sequentially in the setting of selective immunologic pressure in the population. Drift results from point mutations of the corresponding RNA segment. Antigenic shift occurs only in influenza A viruses and results from acquisition of a new gene segment for hemagglutinin with or without one for neuraminidase. This may occur through genetic reassortment during dual infections with human and animal influenza type A viruses; by the reintroduction of a virus that has not circulated recently in the human population; or by direct transmission to humans of an animal influenza virus that is capable of efficient human-to-human transmission.

Aquatic fowl are the main reservoir for influenza A viruses in nature, although some subtypes also infect other species, including swine, horses, and marine mammals. Swine are susceptible to infection with viruses from both birds and humans and may serve as a so-called mixing vessel, providing an opportunity for the generation of new pathogenic viruses.

Epidemiology and Transmission

Influenza viruses A and B cause annual outbreaks of illness affecting approximately 5% to 10% of adults, with higher rates in children. In the United States, influenza causes an average of over 36,000 deaths[61] and 130,000 to 170,000 hospitalizations during each epidemic.[21] The appearance of a new strain for which most of the population lacks immunity can result in worldwide outbreaks, or pandemics. Pandemic strains are associated with global spread over months and with high attack rates. Three such pandemics occurred during the 20th century; the most severe was the Spanish influenza pandemic in 1918 and 1919, which caused 20 to 40 million deaths worldwide and over 500,000 deaths in the United States.

Epidemic influenza occurs annually in temperate areas, typically between the months of December and March in the Northern Hemisphere, and follows the reintroduction of virus each year. Influenza activity usually occurs in May through August in the Southern Hemisphere and can be year-round in the tropics. Regional outbreaks caused by a particular strain are usually short (6 to 8 weeks), although successive waves of infection by different influenza viruses can occur. Influenza activity in the community is marked by increased medical contacts for febrile respiratory illness, increased absenteeism from school and the workplace, subsequent increased hospitalizations for pneumonia and other cardiopulmonary disorders, and increased mortality. Pneumonia hospitalizations increase by two to five times in high-risk patients. Persons 65 years of age and older constitute nearly 50% of excess hospitalizations and over 85% of deaths from influenza.[62]

Influenza viruses are transmitted principally via large and small aerosolized particles. Direct transmission of influenza A virus to humans from animals has been documented.[63] Although animal-to-human transmission has typically been from swine, direct transmission from birds caused an epizootic of avian influenza A H5N1 subtype virus in Hong Kong and Southeast Asia, with several dozen human infections, many of which were fatal.[64,65] In the cluster of H5N1 avian virus in Hong Kong,[66] the source was domestic poultry, although human-to-human transmission may have occurred in several cases. Additionally, avian H9N2 influenza virus infection has been documented in several children, and a large outbreak of H7N7 avian influenza A virus in the Netherlands was associated with at least one death.[67] H7 subtype viruses have also caused an outbreak, predominantly of conjunctivitis, in British Columbia, Canada,[67] and infected an immunocompromised man in New York. Influenza H5N1 or other avian viruses may pose a reemerging pandemic threat in the future.[68]

Diagnosis

Clinical features The incubation period for influenza virus is short, averaging 2 days (range, 1 to 4 days). Classic influenza is distinguished by abrupt onset of prominent systemic symptoms, including fever, chills, headache, myalgia, malaise, and anorexia. Fever usually lasts for an average of 3 days in adults. Sore throat, dry cough, photophobia, and pain on eye movement occur frequently early in the illness. Mild conjunctivitis, clear nasal discharge, pharyngeal injection, and small, tender cervical lymph nodes are also common. As systemic illness abates, respiratory symptoms become more apparent. The most troubling respiratory symptom is protracted cough, which results from viral tracheobronchitis. Airway hyperactivity and abnormalities in pulmonary function may last from weeks to several months in previously healthy persons. Exacerbations of asthma and other types of preexisting airway disease are often severe. Infections may be subclinical or cause milder illness, including colds.

Primary influenza virus pneumonia is a heterogeneous condition, ranging from mild disease with patchy infiltrates to rapidly fatal infection. Severe pneumonia generally accounts for 2% of influenza-associated pneumonia, but during pandemics, it can account for up to 20%; influenza A viruses cause more than 90% of cases.[69] As many as 40% of patients with influenza pneumonia have no prior underlying disease. Pneumonia usually begins with typical influenza, followed within 1 to 3 days by rapidly progressive dyspnea, cyanosis, diffuse rales, and wheezing. Pleuritic chest pain and blood-tinged sputum or frank he-

moptysis occurs. Patients with influenza virus pneumonia have a high mortality.

Laboratory tests Multiple rapid assays are commercially available in the United States to detect influenza A and B antigens or neuraminidase activity; some of these assays differentiate between influenza A and B, and several can be performed by clinical personnel at the point of care [*see Table 2*]. The specificity of these assays is good to excellent, but the sensitivity varies between approximately 60% and 90%, depending on the sample type, the age of the patient, and the duration of the illness.[68,70] Diagnosis can also be made by viral culture or by RT-PCR. When viral pneumonia is present, Gram stain of sputum shows few to many polymorphonuclear leukocytes, but only rarely does it show bacteria. The chest radiograph shows bilateral infiltrates that may be in the form of diffuse interstitial infiltrates, perihilar pulmonary edema, or dense opacifications. On blood counts, leukocytosis with a left shift is variably present.

A definitive diagnosis of influenza can have a significant impact on medical management. In a pediatric population, detection of influenza A antigen resulted in a decrease in antibiotic use, a decrease in duration of antibiotic use in hospitalized patients, and an increase in antiviral use.[71]

Complications

Secondary bacterial pneumonia should be suspected when fever, increasing cough, and sputum production develop after several days of improvement. *Streptococcus pneumoniae* is the most common bacterial pathogen, but *Staphylococcus aureus,* including community-acquired methicillin-resistant strains, causes up to 25% of cases and is associated with high mortality. *S. aureus* and certain other bacteria produce proteolytic enzymes that activate influenza hemagglutinin and enhance viral replication. *Haemophilus influenzae* and *Streptococcus pyogenes* are also recognized as causes of bacterial complications. Bacterial infections of the sinuses and ears are frequent. Toxic-shock syndrome and invasive meningococcal disease have also been known to complicate influenza.

Uncommon complications are myositis with rhabdomyolysis, renal failure, disseminated intravascular coagulopathy, myocarditis, pericarditis, myelitis, Guillain-Barré syndrome, and Reye syndrome. Neurologic complications, including encephalopathy or encephalitis, are unusual and occur mainly in children.[63,72]

Treatment

A variety of antiviral agents are available for treatment of influenza [*see Table 3*]. The M2 inhibitors amantadine and rimantadine are active against influenza A only, although recent human isolates of avian H5N1 viruses are resistant.[31,68] Oral amantadine and rimantadine reduce the duration of fever and symptoms of uncomplicated influenza A virus infection by 1 to 2 days and provide more rapid overall functional recovery. Effectiveness in preventing complications or treating severe illness in hospitalized patients is uncertain. Resistant virus may arise during treatment and be transmissible on close contact.[73]

The neuraminidase inhibitors, zanamivir and oseltamivir, are active against influenza A and B viruses.[32] Zanamivir is administered by inhalation and may, in rare cases, cause bronchospasm, which can be severe. Oseltamivir is administered orally and is associated with self-limited GI upset in about 10% to 15% of treated patients; this can be reduced by taking the medication with food. Treatment of acute uncomplicated influenza in adults with inhaled zanamivir or oral oseltamivir provides symptomatic relief, reduces time to functional recovery, and decreases the likelihood of lower respiratory tract complications leading to antibiotic use.[22,23] In children 1 to 12 years of age, oseltamivir provides symptomatic relief and also reduces the likelihood of otitis media.[21] Oseltamivir treatment has been associated with reduced risk of hospitalization in both previously healthy and high-risk or elderly adults.[74] Emergence of resistance appears to be very uncommon.

Influenza pneumonia Treatment of influenza virus pneumonia is primarily supportive. Improvements in assisted ventilation techniques have raised the survival rate above 50%, although pulmonary fibrosis develops in some patients. M2 inhibitors, NA inhibitors, and aerosolized or I.V. ribavirin, which is active against influenza A and B viruses, have been used with uncertain benefit. Combination therapies (e.g., an M2 inhibitor plus a neuraminidase inhibitor for influenza A) show enhanced activity in animal models. Although unstudied, the use of neuraminidase inhibitors, either alone or in combination with other agents, seems reasonable in the treatment of influenza virus pneumonia.

Prevention

Chemoprophylaxis Antiviral chemoprophylaxis with amantadine or rimantadine is about 70% to 90% effective in preventing illness caused by influenza A virus. Chemoprophylaxis can provide several weeks' protection to patients immunized after influenza A activity has begun, can be given throughout the season to those who cannot receive influenza vaccine (e.g., those with egg allergy) or who are unlikely to respond to the vaccine, or can be used for protection of high-risk persons when the epidemic strain diverges significantly from the vaccine antigens. Side effects of amantadine are GI upset and minor, reversible central nervous system symptoms (e.g., insomnia, dizziness, and difficulty with concentration). Rimantadine appears to be equally effective as amantadine and is associated with a lower risk of CNS side effects.

The neuraminidase inhibitors are also effective for prophylaxis of influenza A and B infections. Both inhaled zanamivir[75,76] and oral oseltamivir[77,78] prevent influenza when used for seasonal prophylaxis or after exposure (e.g., for family or nursing home contacts), but only oseltamivir has been approved by the Food and Drug Administration for this indication. Oseltamivir has been found to be safe for treatment of children as young as 1 year[79]; inhaled zanamivir has been found to be safe in children as young as 5 years. Of note, both the M2 and the neuraminidase inhibitors may reduce the efficacy of live attenuated influenza vaccine, if the vaccine and the antiviral agent are given together.

Immunization Influenza vaccines are available in two forms: (1) an intramuscular preparation containing formalin-inactivated virus and purified surface antigen and (2) an intranasal spray containing live attenuated viruses.[21] Both are made from egg-grown virus. Vaccine composition is reviewed annually and adjusted to reflect changes in antigenicity and anticipated circulation of viral strains; current vaccines contain one influenza B and two influenza A (H1 and H3 subtypes) antigens.[21] The efficacy of these vaccines is approximately 70% to 90% in young adults, especially when the vaccine antigen and the circulating strain are closely matched. Immunization in

Table 3 Agents Used to Prevent and Treat Influenza[111,112]

Drug	Adult Dosage		Efficacy for Documented Influenza		Dosage Adjustment	
	*Prophylaxis**	*Treatment†*	*Prophylaxis*	*Treatment*	*State*	*Dosage*
Amantadine	100 mg b.i.d.	100 mg b.i.d.	63% efficacy[95]	Shorter duration of fever and symptoms	C_{Cr} 30–50 C_{Cr} 15–30 C_{Cr} < 15 Hemodialysis Elderly	100 mg q.d 100 mg q.o.d. 100 mg q. wk. 100 mg q. wk. ≤ 100 mg q.d.
Rimantadine	100 mg b.i.d. or 200 mg q.d.	100 mg b.i.d. or 200 mg q.d.	72%–83% efficacy[95]	Shorter duration of fever and symptoms	Severe hepatic dysfunction Elderly C_{Cr} < 10	100 mg q.d. 100 mg q.d. 100 mg q.d.
Zanamivir	2 puffs q.d.‡	2 puffs b.i.d.	67%–84% efficacy[96]	1 day faster recovery from illness (up to 2.5 day improvement in high-risk patients) Reduction in severity of illness, number of nights of disturbed sleep, and use of relief medications Reduction in complications that require antibiotic treatment More rapid return to normal function	—	No dosage adjustments
Oseltamivir	75 mg q.d.	75 mg b.i.d.	87% efficacy[96]	24 to 36 hr faster recovery from illness Reduction in severity of illness, number of nights of disturbed sleep, and use of relief medications Reduction in complications that require antibiotic treatment, hospitalization	C_{Cr} < 30	Treatment, 75 mg q.d. Prophylaxis, 75 mg q.o.d.

*Duration of prophylaxis depends on the clinical circumstances: 7–10 days after close contact for postexposure prophylaxis; 2 wk after immunization of an adult; 4–6 wk after a community outbreak; and at least 1 wk (preferably 2 wk) after the last case in outbreak control.
†First dose should be given within 48 hr of the onset of illness to be effective. Therapy should be continued for 5 days.
‡The Food and Drug Administration has not approved zanamivir for this indication.
C_{Cr}—creatinine clearance

healthy working adults is associated with fewer upper respiratory illnesses and fewer visits to physicians' offices.[80,81] The intranasal vaccine is currently approved only for healthy persons 5 to 49 years of age; it appears to be less effective in the elderly but possibly superior to inactivated vaccine in young children. Immunization with inactivated vaccine reduces influenza-related hospitalizations, including acute cardiovascular events and COPD exacerbations,[82] and reduces mortality in ambulatory elderly patients by 40% to 60%.[83] Wide-scale immunization of schoolchildren may reduce influenza-related mortality in older adults.[84] Although inactivated vaccine is less effective in infirm elderly patients, it provides partial protection against pneumonia and death.

The highest priority for vaccination should be given to persons at increased risk for influenza-related complications and their contacts [*see Table 4*]. Administration of the inactivated vaccine is associated with soreness at the injection site in as many as one third of recipients, but fever or systemic reactions are uncommon in adults. Influenza vaccination does not adversely affect $CD4^+$ T cell counts or accelerate progression to AIDS or death in HIV-infected patients.[85]

The intranasal vaccine causes transient coryza and sore throat. Vaccine viruses are recoverable from nasal samples in low titers for up to 1 week in adults and are genetically stable. Although viral transmission is rare between children and has not been documented in adults to date, vaccine recipients should avoid close contact with highly immunocompromised hosts (e.g., stem cell transplant recipients) for 1 week after immunization.[21]

PARAINFLUENZA VIRUSES

PIV infections are the most commonly recognized cause of croup, accounting for up to 75% of cases with a documented viral etiology, and they are a leading cause of lower respiratory tract disease resulting in hospitalization of infants. Infection with PIV induces partial immunity that moderates disease severity; reinfection in older children and adults produces milder respiratory illness and colds.

Classification and Pathogenesis

Human PIVs are medium-sized (120 to 150 nm), enveloped RNA viruses belonging to the Paramyxoviridae family and are classified into four antigenically stable types (1, 2, 3, and 4); type 4 is further classified into subtypes A and B. PIV replicates in epithelial cells of the upper and lower respiratory tract, and antibodies to the two major envelope surface glycoproteins—namely, the hemagglutinin-neuraminidase protein and the fusion protein—confer protection against infection. These proteins are necessary for attachment of virus to host cell receptors and

Table 4 Indications for Influenza Vaccination*[21]

PERSONS AT INCREASED RISK FOR COMPLICATIONS
All adults 65 yr of age and older
Residents of nursing homes and other long-term care facilities that house persons of any age who have chronic medical conditions
Adults and children with chronic cardiopulmonary disorders, including asthma
Adults and children who have required regular medical follow-up or hospitalization during the preceding year because of chronic metabolic diseases (including diabetes), renal dysfunction, hemoglobinopathies, or immunosuppression (including immunosuppression caused by medication or by HIV)
Children and teenagers (6 mo to 18 yr of age) who are receiving long-term aspirin therapy and therefore may be at risk for Reye syndrome
Women who will be pregnant during the influenza season
Children 6–23 mo of age
PERSONS 50–64 YR OF AGE†
PERSONS WHO CAN TRANSMIT INFLUENZA TO THOSE AT HIGH RISK
Physicians, nurses, and other personnel in both hospital and outpatient care settings, including medical emergency response workers (e.g., paramedics and emergency medical technicians)
Employees of nursing homes and long-term care facilities who have contact with patients or residents
Employees of assisted-living and other residences for persons in groups at high risk
Home care providers for persons at high risk
Household contacts of high-risk persons

*Both the inactivated influenza vaccine and live, attenuated influenza vaccine (LAIV) can be used to reduce the risk of influenza. Healthy persons 5–49 yr of age who are not contacts of severely immunosuppressed persons can receive either LAIV or inactivated influenza vaccine; all other persons should receive inactivated influenza vaccine. In addition to persons for whom annual influenza vaccination is recommended, influenza vaccine should be given to any person who wishes to reduce the likelihood of becoming ill with influenza, if vaccine availability permits.
†Vaccination is recommended for persons 50–64 yr of age because of the increased prevalence of persons with high-risk conditions in this group.

membrane-fusing activity. PIV type 1 and type 2, which cause croup and laryngitis, probably replicate principally in the major airways, whereas PIV-3 causes lower respiratory tract infections.

Epidemiology and Transmission

PIV infections occur initially during childhood; PIV-3 infections often occur during infancy and are an important cause of illness in infants younger than 6 months, as well as in immunocompromised persons. The incidence of croup and respiratory disease caused by PIV-1 or PIV-2 infections is highest between the ages of 6 months and 5 years. PIV-1 typically causes epidemics lasting several months, peaking in October or November of alternating years, whereas PIV-2 activity is more sporadic. PIV-3 causes infections throughout the year, with outbreaks during the spring and summer.

PIVs are transmitted from person to person, either by direct contact with respiratory secretions or by large aerosol droplets. Transmission occurs readily in families, and reinfections in older children and adults are common. Outbreaks occur in nurseries, day care centers, and hospitals, with attack rates of 40% or higher in susceptible patients.

Diagnosis

Clinical features The incubation period of PIVs is approximately 3 to 6 days. Virus replication is generally limited to the respiratory tract, although viremia and CNS infections[86] have been rarely documented. Initial infections cause fever, rhinitis, pharyngitis, laryngitis, croup, and bronchitis in children.

Laboratory tests Immunofluorescence testing of respiratory secretions and RT-PCR have been used for rapid PIV detection. Culture is generally slow, requiring days to weeks. Serologic assays, including complement fixation, ELISA, and hemagglutination assays, are used for retrospective diagnosis.

Treatment

Specific antiviral therapy of proven value has not been established. Early treatment with aerosol ribavirin may benefit some immunocompromised patients with serious infections, and intravenous ribavirin has been used for treatment of PIV pneumonia in BMT patients.[87]

Prevention

An effective PIV vaccine is not yet available, but live attenuated vaccines derived from either human or bovine PIV have been developed and appear to be safe and immunogenic in children. Infection control measures, with an emphasis on exclusion of sick caregivers and isolation protocols, may reduce the risk of nosocomial acquisition in compromised hosts.

Complications

In adults and older children, reinfections are often asymptomatic but may cause serious lower respiratory tract disease, including exacerbations of chronic airway disease and pneumonia. During periods of circulation, about 2% to 3% of cases of community-acquired pneumonia in adults are linked to PIV infection.[88] Severe pneumonia, sometimes in the absence of upper respiratory tract illness, occurs in immunosuppressed adults and children, particularly after BMT.[5]

RESPIRATORY SYNCYTIAL VIRUS

Classification and Pathogenesis

RSV is a paramyxovirus with surface glycoproteins that have neither hemagglutinin nor neuraminidase activity. The F protein is responsible for fusion of the viral envelope with the host cell membranes. Antibody against this protein neutralizes RSV infectivity and blocks syncytial cell formation. The G protein is responsible for attachment. Two major subgroups (A and B) are distinguished primarily by antigenic differences in the G glycoprotein, and multiple strains are recognized within each subgroup.

Epidemiology and Transmission

RSV is the major cause of lower respiratory tract disease in infants and young children, accounting for 45% to 90% of bronchiolitis cases and up to 40% of pneumonia cases. Approximately 40% of infections in the first year of life cause lower respiratory tract illness and 1% to 2% result in hospitalization; about 10% of hospitalized infants require ventilatory support.

As many as 100,000 hospitalizations and 510 deaths of infants and children are attributed to RSV each year in the United States.[89] RSV is also increasingly recognized as a cause of lower respiratory tract disease in older adults and immunocompro-

mised persons.[90,91] Mortality is usually less than 1% in previously healthy infants but is much higher in persons with primary immunodeficiency, persons undergoing cancer chemotherapy, and persons with preexisting pulmonary and heart disease.[92] The number of RSV-related deaths in the elderly is estimated to be 10-fold higher than that in infants.[93] In elderly persons, it is estimated that RSV is associated with 2% to 9% of hospitalizations or deaths caused by pneumonia[90] and about 10% of hospitalizations for acute cardiopulmonary conditions.[89] Each year, RSV is responsible for an estimated 11,321 excess deaths in the United States.[61]

RSV causes prolonged outbreaks of infection from late fall to spring annually in temperate climates.[94] Most outbreaks peak between January and March in the Northern Hemisphere, but sporadic cases can occur year-round. Epidemics are associated with increased pediatric hospitalizations and deaths from lower respiratory tract illness. Essentially all children are infected within several years after birth. Immunity is incomplete, and reinfections in children and adults are common. Higher titers of circulating and mucosal antibody occur with successive infections and appear to be associated with milder illness.

RSV appears to be spread by large-particle aerosols during close personal contact and by hand contamination with self-inoculation of the eye or nose. RSV is a major nosocomial pathogen, and high attack rates occur during outbreaks in hospitals, transplantation units, day care centers, and geriatric homes.[95] Illness rates are commonly 20% to 50% among hospital staff and patients during epidemics. In households, secondary infection occurs in about one third of adults.

Diagnosis

Clinical features The incubation period of RSV averages 4 to 5 days but ranges up to 1 week. In infants and young children, almost all primary infections are symptomatic, and 40% or more are associated with bronchiolitis or pneumonia. Febrile upper respiratory tract illness and otitis media are common.

Most recurrent infections in adults are associated with upper respiratory tract illness. Adults typically experience coryza, pharyngitis, and cough, often with low-grade fever. Bronchitis, influenzalike illness, pneumonia, and exacerbations of asthma and chronic bronchitis also occur. In the elderly, the clinical manifestations and outcomes of RSV infections are similar to those of influenza.[89,94] Elderly adults hospitalized with RSV commonly have dyspnea, wheezing, and sputum production. Viral bronchopneumonia or secondary bacterial pneumonia appears to complicate a high proportion of RSV infections in such patients.

RSV causes severe lower respiratory tract disease in immunosuppressed patients, particularly BMT recipients and acute leukemia patients receiving chemotherapy.[6] Most cases are nosocomial; mortality exceeds 60% when pneumonia develops. Rhinitis with sinusitis or otitis often precedes the development of pneumonia by several days and may provide a clue to diagnosis.

Laboratory tests Testing of bronchoalveolar lavage samples with RSV antigen or RNA detection provides rapid diagnosis in pneumonia cases. Commercially available ELISAs are sensitive to RSV in children, but nasopharyngeal samples are an insensitive means of detecting RSV antigen in adults, who have low titers of virus in the upper respiratory tract. RT-PCR is a reasonably sensitive technique to detect RSV infection, whereas culture of virus is slow and less sensitive. Serology is more sensitive than cell culture for diagnosis in adults but is slow because of the need for paired sera.[89]

Treatment

The value of aerosolized ribavirin in the management of RSV disease in hospitalized children remains controversial, and studies of its value in mechanically ventilated infants have yielded conflicting results.[23] Aerosolized ribavirin should be considered for infants hospitalized with RSV infections who are at high risk for severe or complicated illness (e.g., those with congenital heart disease, bronchopulmonary dysplasia, cystic fibrosis, and immunodeficiency), for those who are already severely ill, and for those who are younger than 6 weeks. Treatment with polyclonal or monoclonal RSV antibodies does not provide clear therapeutic benefit for RSV-infected infants and young children. Aerosol ribavirin may reduce progression from upper to lower respiratory tract disease in highly immunocompromised hosts. Early treatment with aerosolized ribavirin plus IVIg, RSV immunoglobulin, or palivizumab appears to benefit immunosuppressed patients with RSV pneumonia.[96]

Prevention

Passive immunoprophylaxis by monthly administration of high-titer anti-RSV immunoglobulin or anti-F monoclonal antibody (palivizumab) reduces the risk of lower respiratory tract RSV disease and hospitalization in high-risk infants and children. Palivizumab is approved for prophylaxis in premature infants and children younger than 2 years with bronchopulmonary dysplasia and is appropriate for infants and young children with hemodynamically significant congenital heart disease.[97] A phase 1 study has found that palivizumab appears to be safe and well tolerated in recipients of hematopoietic stem cell transplants.[98] No vaccine against RSV is available yet, but studies of intranasal live-attenuated vaccine in children and injected subunit vaccine in elderly persons are ongoing.

Interruption of nosocomial transmission may be facilitated by thorough hand washing,[99] decontamination of fomites, early identification and rapid isolation of infected infants and other potential RSV cases,[100] and cohorting of staff with infected infants.[101] Regular use of gowns, gloves, disposable eye-nose goggles, and possibly masks by hospital staff caring for infected children may further reduce the risk of nosocomial RSV spread.

RHINOVIRUS

Classification and Pathogenesis

Rhinoviruses are small (30 nm), nonenveloped viruses that belong to the Picornaviridae family. They have a single-stranded RNA genome and are composed of four structural proteins. Three of the proteins induce neutralizing antibodies and form the basis for serotyping. The large number of serotypes (> 110) makes the development of an effective vaccine unlikely. Nearly 90% of rhinovirus serotypes use intercellular adhesion molecule–1 as a common cellular receptor. Although replication of rhinovirus was thought to be limited to the upper respiratory tract, increasing data suggest that rhinovirus can replicate and cause disease in the lower airways.[102]

Epidemiology and Transmission

Rhinoviruses cause approximately one infection per person per year in adults, and rates are even higher in children. Rhinovirus causes about 50% of colds in adults each year and up to

Internet Resources for Respiratory Viral Infections

Adenovirus
- National Respiratory and Enteric Virus Surveillance System
  http://www.cdc.gov/ncidod/dvrd/revb/respiratory/eadfeat.htm

SARS
- Centers for Disease Control and Prevention
  www.cdc.gov/ncidod/sars
- World Health Organization
  http://www.who.int/csr/sars/en/

Influenza Virus
- World Health Organization
  http://www.who.int/csr/disease/influenza/en/
- Centers for Disease Control and Prevention
  http://www.cdc.gov/ncidod/diseases/flu/fluvirus.htm
- National Foundation for Infectious Diseases
  http://www.nfid.org/library/influenza
- American Lung Association
  http://www.lungusa.org/site/pp.asp?c=dvLUK9O0E&b=35426

Parainfluenza Viruses
- National Respiratory and Enteric Virus Surveillance System
  http://www.cdc.gov/ncidod/dvrd/revb/respiratory/hpivfeat.htm

Respiratory Syncytial Virus
- National Respiratory and Enteric Virus Surveillance System
  http://www.cdc.gov/ncidod/dvrd/revb/respiratory/rsvfeat.htm
- American Lung Association
  http://www.lungusa.org/site/pp.asp?c=dvLUK9O0E&b=35695

Hantavirus
- Centers for Disease Control and Prevention
  http://www.cdc.gov/ncidod/diseases/hanta/hps/index.htm
- Emerging Viruses Research Center Hantavirus Reference Laboratory
  http://hsc.unm.edu/pathology/HjelleLab/
- American Lung Association
  http://www.lungusa.org/site/pp.asp?c=dvLUK9O0E&b=35428

90% during the fall months.[103] Immunity to rhinovirus is serotype specific and relatively enduring after infection, although reinfection can occur. Rhinoviruses cause infections year-round but are most prevalent in early fall and late spring. The relative importance of fomite spread with hand contamination and self-inoculation into the eyes or nose and aerosols in rhinovirus transmission is uncertain under natural conditions.

Diagnosis

Clinical features The incubation period of rhinoviruses averages about 2 days, with a range of 1 to 5 days. Rhinovirus colds vary in severity from mild episodes characterized by 1 to 2 days of coryza or scratchy throat to illnesses with profuse rhinorrhea, pharyngitis, and tracheobronchitis. Symptoms usually peak on the second and third days of illness, and the median duration is 1 week.

Rhinovirus infections are associated with more frequent lower respiratory tract illness in the elderly and with bronchospasm and protracted cough in persons with preexisting airway disease.[104,105] Rhinovirus is the major infectious cause of asthma exacerbations in children and in adults. In addition, rhinovirus infections are associated with acute sinusitis,[106] otitis media,[107] exacerbations of chronic bronchitis, and lower respiratory tract disease in infants and young children. In BMT recipients, rhinovirus infection has been associated with pneumonia and poor prognosis.[1]

Laboratory tests No rapid diagnostic test for rhinovirus infection is routinely available; detection is usually by virus isolation and, more recently, by RT-PCR.[108]

Treatment

No antiviral therapy of proven clinical value is currently available, but several approaches are under investigation. Oral administration of the capsid-binding antirhinoviral compound pleconaril reduces the severity and duration of uncomplicated rhinovirus colds in adults but is associated with risk of drug interactions.[109] Nonselective regimens such as high-dose vitamin C, *Echinacea* preparations, inhalation of heated air, and zinc lozenges are of doubtful value.

Prevention

Barrier control measures may be beneficial, although tissues impregnated with virucides have proved disappointing in this regard. Hand washing and avoidance of finger-to-nose and finger-to-eye contact should reduce the risk of infection. Because of the risk of nosocomial transmission of rhinovirus infections, adherence to infection control precautions in cases of upper respiratory tract illness is important.

HANTAVIRUS

Hantavirus causes hantavirus cardiopulmonary syndrome (HCPS).[110] Most HCPS cases have occurred in young to middle-aged adults with no underlying disease. The largest number of cases have occurred in New Mexico, Arizona, and California, but cases have also been identified throughout western North America and sporadically in the eastern United States. HCPS is typically a zoonosis; the principal animal reservoir is the deer mouse. Human infections occur by inhalation of aerosols of infectious excreta. Detailed discussion of HCPS is provided elsewhere [*see 7:XXXI Viral Zoonoses*].

INTERNET RESOURCES

A listing of Internet resources with information about respiratory viral infections is provided [*see Sidebar* Internet Resources for Respiratory Viral Infections].

Frederick G. Hayden, M.D., has received grant or research support from Abbott Laboratories, Biota Holdings, Loders Croklaan, MedImmune, Perlan Therapeutics, Pfizer/Pharmacia, and ViroPharma; he is a consultant for Biota Holdings, ImmunoRx, Perlan Therapeutics, Roche, and ViroPharma; and he participates in the speakers' bureau for MedImmune and Roche.

Michael G. Ison, M.D. M.Sc., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Amstrong GL, Conn LA, Pinner RW: Trends in infectious disease mortality in the United States during the 20th century. JAMA 281:61, 1999

2. Mainous AG, Hueston WJ: The cost of antibiotics in treating upper respiratory tract infections in a medical population. Arch Fam Med 7:45, 1998

3. Monto AS: Studies of the community and family: acute respiratory illness and infection. Epidemiol Rev 16:351, 1994

4. Kaiser L, Fritz RS, Straus SE, et al: Symptom pathogenesis during acute influenza: interleukin-6 and other cytokine responses. J Med Virol 64:262, 2001

5. Hayden FG, Fritz R, Lobo MC, et al: Local and systemic cytokine responses during experimental human influenza A virus infection: relation to symptom formation and host defense. J Clin Invest 101:643, 1998

6. To KF, Chan PK, Chan KF, et al: Pathology of fatal human infection associated with avian influenza A H5N1 virus. J Med Virol 63:242, 2001

7. Sable CA, Hayden FG: Orthomyxoviral and paramyxoviral infections in transplant patients. Infect Dis Clin North Am 9:987, 1995

8. Olsen MA, Shuck KM, Sambol AR, et al: Isolation of seven respiratory viruses in shell vials: a practical and highly sensitive method. J Clin Microbiol 31:422, 1993

9. Hindiyeh M, Goulding C, Morgan H, et al: Evaluation of BioStar Flu OIA assay for rapid detection of influenza A and B viruses in respiratory specimens. J Clin Virol 17:119, 2000

10. Grondahl B, Puppe W, Hoppe A, et al: Rapid identification of nine microorganisms causing acute respiratory tract infections by single-tube multiplex reverse transcription-PCR: feasibility study. J Clin Microbiol 37:1, 1999

11. Gwaltney JM Jr, Phillips CD, Miller RD, et al: Computed tomographic study of the common cold. N Engl J Med 330:25, 1994

12. Anon JB, Jacobs MR, Poole MD, et al: Antimicrobial treatment guidelines for acute bacterial rhinosinusitis. Sinus and Allergy Health Partnership. Otolaryngol Head Neck Surg 130(1 suppl):1, 2004

13. Arola M, Ruuskanen O, Ziegler T, et al: Clinical role of respiratory virus infection in acute otitis media. Pediatrics 86:848, 1990

14. Hayden FG, Diamond L, Wood PB, et al: Effectiveness and safety of intranasal ipratropium bromide in common colds: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 125:89, 1996

15. Romero JR: Pleconaril: a novel antipicornaviral drug. Expert Opin Investig Drugs 10:369, 2001

16. Melchart D, Linde K, Fischer P, et al: Echinacea for preventing and treating the common cold. Cochrane Database Syst Rev (2):CD000530, 2000

17. Marshall I: Zinc for the common cold. Cochrane Database Syst Rev (2):CD001364, 2000

18. Vickers AJ, Smith C: Homeopathic Oscillococcinum for preventing and treating influenza and influenza-like syndromes. Cochrane Database Syst Rev (2):CD001957, 2000

19. Hayden FG: Rhinovirus and the lower respiratory tract. Rev Med Virol 14:17, 2004

20. Gonzales R, Bartlett JG, Besser RE, et al: Principles of appropriate antibiotic use for treatment of uncomplicated acute bronchitis: background. Ann Intern Med 134:521, 2001

21. Harper SA, Fukuda K, Uyeki TM, et al: Prevention and control of influenza: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 53 (RR-6):1, 2004

22. Kaiser L, Keene ON, Hammond JMJ, et al: Impact of zanamivir on antibiotic use for respiratory events following acute influenza in adolescents and adults. Arch Intern Med 160:3234, 2000

23. Treanor JJ, Hayden FG, Vrooman PS, et al: Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group. JAMA 283:1016, 2000

24. Reassessment of the indications for ribavirin therapy in respiratory syncytial virus infections. American Academy of Pediatrics Committee on Infectious Diseases. Pediatrics 97:137, 1996

25. Folkerts G, Busse WW, Nijkamp FP, et al: Virus-induced airway hyperresponsiveness and asthma. Am J Respir Crit Care Med 157:1708, 1998

26. Glezen WP, Greenberg SB, Atmar RL, et al: Impact of respiratory virus infections on persons with chronic underlying conditions. JAMA 283:499, 2000

27. Greenberg SB: Viral pneumonia. Infect Dis Clin North Am 5:603, 1991

28. Loda FA, Clyde WA Jr, Glezen WP, et al: Studies on the role of viruses, bacteria, and *M. pneumoniae* as causes of lower respiratory tract infections in children. J Pediatr 72:161, 1968

29. Peiris JSM, Phil D, Yue KY, et al: The severe acute respiratory syndrome. N Engl J Med 349:2431, 2003

30. Clinical guidance on the identification and evaluation of possible SARS-CoV disease among persons presenting with community-acquired illness. Centers for Disease Control and Prevention, Atlanta, 2004
http://www.cdc.gov/ncidod/sars/clinicalguidance.htm

31. Hayden FG, Aoki FY: Amantadine, rimantadine, and related agents. Antimicrobial Chemotherapy. Yu VL, Merigan TC, White NJ, et al, Eds. Williams & Wilkins, Baltimore, 1999, p 1344

32. Gubareva LV, Kaiser L, Hayden FG: Influenza virus neuraminidase inhibitors. Lancet 355:827, 2000

33. Hayden FG, Osterhaus AD, Treanor JJ, et al: Efficacy and safety of the neuraminidase inhibitor zanamivir in the treatment of influenza virus infections. N Engl J Med 337:874, 1997

34. Connolly MG Jr, Baugham RP, Dohn MN, et al: Recovery of viruses other than cytomegalovirus from bronchoalveolar lavage fluid. Chest 105:1775, 1994

35. Baughman RP: The lung in the immunocompromised patient: infectious complications part 1. Respiration 66:95, 1999

36. Rabella N, Rodriguez P, Labeaga R, et al: Conventional respiratory viruses recovered from immunocompromised patients: clinical considerations. Clin Infect Dis 28:1043, 1999

37. Duggan JM, Farrehi J, Duderstadt S, et al: Treatment with ganciclovir of adenovirus pneumonia in a cardiac transplant patient. Am J Med 103:439, 1997

38. Shetty AK, Gans HA, So S, et al: Intravenous ribavirin therapy for adenovirus pneumonia. Pediatr Pulmonol 29:69, 2000

39. Safrin S, Cherrington J, Jaffe HS: Clinical uses of cidofovir. Rev Med Virol 7:145, 1997

40. Ludwig SL, Brundage JF, Kelley PW, et al: Prevalence of antibodies to adenovirus serotypes 4 and 7 among unimmunized US Army trainees: results of a retrospective nationwide seroprevalence survey. J Infect Dis 178:1776, 1998

41. Jernigan JA, Lowry BS, Hayden FG, et al: Adenovirus type 8 epidemic keratoconjunctivitis in an eye clinic: risk factors and control. J Infect Dis 167:1307, 1993

42. Uemura T, Kawashitam T, Ostuka Y, et al: A recent outbreak of adenovirus type 7 infection in a chronic inpatient facility for the severely handicapped. Infect Control Hosp Epidemiol 21:559, 2000

43. Hierholzer JC: Adenoviruses in the immunocompromised host. Clin Microbiol Rev 5:262, 1992

44. Tsutsumi H, Ouchi K, Ohsaki M, et al: Immunochromatography test for rapid diagnosis of adenovirus respiratory tract infections: comparison with virus isolation in tissue culture. J Clin Microbiol 37:2007, 1999

45. Couch RB, Englund JA, Whimbey E: Respiratory viral infections in immunocompetent and immunocompromised persons. Am J Med 102:2, 1997

46. Howard DS, Phillips II GL, Reece DE, et al: Adenovirus infections in hematopoietic stem cell transplant recipients. Clin Infect Dis 29:1494, 1999

47. Dagan R, Schwartz RH, Insel RA, et al: Severe diffuse adenovirus 7a pneumonia in a child with combined immunodeficiency: possible therapeutic effect of human immune serum globulin containing specific neutralizing antibody. Pediatr Infect Dis 3:246, 1984

48. DeVincenzo J: Pre-emptive ribavirin for 'RSV' in BMT. Bone Marrow Transplant 26:113, 2000

49. Goetz MB, Mathisen GE: Clinical course and treatment of adults with severe measles pneumonitis. Clin Infect Dis 21:443, 1995

50. Yu IT, Li Y, Wong TW, et al: Evidence of airborne transmission of the severe acute respiratory syndrome virus. N Engl J Med 350:1731, 2004

51. Lim PL, Kurup A, Gopalakrishna G, et al: Laboratory-acquired severe acute respiratory syndrome. N Engl J Med 350:1740, 2004

52. Wong KT, Antonio GE, Hui DS, et al: Severe acute respiratory syndrome: radiographic appearances and pattern of progression in 138 patients. Radiology 228:401, 2003

53. Poon LL, Chan KH, Wong OK, et al: Early diagnosis of SARS coronavirus infection by real time RT-PCR. J Clin Virol 28:233, 2003

54. Tan EL, Ooi EE, Lin CY, et al: Inhibition of SARS coronavirus infection in vitro with clinically approved antiviral drugs. Emerg Infect Dis 10:581, 2004

55. ter Meulen J, Bakker AB, van den Brink EN, et al: Human monoclonal antibody as prophylaxis for SARS coronavirus infection in ferrets. Lancet 363:2139, 2004

56. Traggiai E, Becker S, Subbarao K, et al: An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus. Nat Med 10:871, 2004

57. van den Hoogen BG, de Jong JC, Groen J, et al: A newly discovered human pneumovirus isolated from young children with respiratory tract disease. Nature Medicine 7:719, 2001

58. van den Hoogen BG, Osterhaus ADME, Fouchier RAM: Clinical impact and diagnosis of human metapneumovirus infection. Pediatr Infect Dis J 23(1 suppl):S25, 2004

59. Williams JV, Harris PA, Tollefson SJ, et al: Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. N Engl J Med 350:443, 2004

60. Wyde PR, Chetty SN, Jewell AM, et al: Comparison of the inhibition of human metapneumovirus and respiratory syncytial virus in ribavirin and immune serum globulin in vitro. Antiviral Res 60:51, 2003

61. Thompson WW, Shay DK, Weintraub E, et al: Mortality associated with influenza and respiratory syncytial virus. JAMA 289:179, 2003

62. Lui KJ, Kendal AP: Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985. Am J Public Health 77:712, 1987

63. Kaiser L, Hayden FG: Hospitalizing influenza in adults. Curr Clin Top Infect Dis 19:112, 1999

64. Claas EC, Osterhaus AD, van Beek R, et al: Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. Lancet 351:472, 1998

65. Trampuz A, Prabhu RM, Smith TF, et al: Avian influenza: a new pandemic threat. Mayo Clin Proc 79:523, 2004

66. Subbarao K, Klimov I, Katz J, et al: Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science 279:393, 1998

67. Koopmans M, Wilbrink B, Conyn M, et al: Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. Lancet 363:587, 2004

68. Nicholson KG, Wood JM, Zambon M: Influenza. Lancet 362:1733, 2003

69. Oliveira EC, Marik PE, Colice G: Influenza pneumonia: a descriptive study. Chest 119:1717, 2001

70. Leonardi GP, Leib H, Birkhead GS, et al: Comparison of rapid detection methods for influenza A virus and their value in health-care management of institutionalized geriatric patients. J Clin Microbiol 32:70, 1994

71. Noyola DE, Demmler GJ: Effect of rapid diagnosis on management of influenza A infections. Pediatr Infect Dis J 19:303, 2000

72. Fujimoto S, Kobayashi M, Uemura O, et al: PCR on cerebrospinal fluid to show influenza-associated acute encephalopathy or encephalitis. Lancet 352:873, 1998

73. Hayden FG: Amantadine and rimantadine: clinical aspects. Antiviral Drug Resistance. Richman DD, Ed. John Wiley & Sons, 1996, p 59

74. Kaiser L, Wat C, Mills T, et al: Impact of oseltamivir treatment on influenza-related lower respiratory tract complications and hospitalizations. Arch Intern Med 163:1667, 2003

75. Monto AS, Robinson DP, Herlocher ML, et al: Zanamivir in the prevention of in-

fluenza among healthy adults: a randomized controlled trial. JAMA 282:31, 1999

76. Hayden FG, Gubareva LV, Monto AS, et al: Inhaled zanamivir for the prevention of influenza in families. N Engl J Med 343:1282, 2000

77. Hayden FG, Atmar RL, Schilling M, et al: Use of the selective oral neuraminidase inhibitor oseltamivir to prevent influenza. N Engl J Med 341:1336, 1999

78. Munoz FM, Galasso GJ, Gwaltney JM Jr, et al: Current research on influenza and other respiratory viruses: II international symposium. Antiviral Res 46:91, 2000

79. Whitley RJ, Hayden FG, Reisinger KS, et al: Oral oseltamivir treatment of influenza in children. Pediatr Infect Dis J 20:127, 2001

80. Nichol KL, Lind A, Margolis KL, et al: The effectiveness of vaccination against influenza in healthy, working adults. N Engl J Med 333:889, 1995

81. Nichol KL, Mendelman PM, Mallon KP, et al: Effectiveness of live, attenuated intranasal influenza virus vaccine in healthy, working adults: a randomized controlled trial. JAMA 282:137, 1999

82. Nichol KL: The efficacy, effectiveness and cost-effectiveness of inactivated influenza virus vaccines. Vaccine 21:1769, 2003

83. Gross PA, Hermogenes AW, Sacks HS, et al: The efficacy of influenza vaccine in elderly persons: a meta-analysis and review of the literature. Ann Intern Med 123:518, 1995

84. Reichert TA, Sugaya N, Fedson DS, et al: The Japanese experience with vaccinating schoolchildren against influenza. N Engl J Med 344:889, 2001

85. Sullivan PS, Hanson DL, Dworkin MS, et al: Effect of influenza vaccination on disease progression among HIV-infected persons. AIDS 14:2781, 2000

86. Arisoy ES, Demmler GJ, Thakar S, et al: Meningitis due to parainfluenza virus type 3: report of two cases and review. Clin Infect Dis 17:995, 1993

87. Ison MG, Hayden FG: Viral infections in immunocompromised patients: what's new with respiratory viruses? Curr Opin Infect Dis 15:355, 2002

88. Marx A, Gary HE Jr, Marston BJ, et al: Parainfluenza virus infection among adults hospitalized for lower respiratory tract infection. Clin Infect Dis 29:134, 1999

89. Shay DK, Holman RC, Roosevelt GE, et al: Bronchiolitis-associated mortality and estimates of respiratory syncytial virus–associated deaths among US children 1979–1997. J Infect Dis 183:16, 2001

90. Falsey AR, Cunningham CK, Barker WH, et al: Respiratory syncytial virus and influenza A infections in the hospitalized elderly. J Infect Dis 172:389, 1995

91. Han LL, Alexander JP, Anderson LJ: Respiratory syncytial virus pneumonia among elderly: an assessment of disease burden. J Infect Dis 179:25, 1999

92. Holberg CJ, Wright AL, Martinez FD, et al: Risk factors for respiratory syncytial virus–associated lower respiratory illnesses in the first year of life. Am J Epidemiol 133:1135, 1991

93. Navas L, Wang E, de Carvalho V, et al: Improved outcome of respiratory syncytial virus infection in a high-risk hospitalized population of Canadian children. Pediatric Investigators Collaborative Network on Infections in Canada. J Pediatr 121:348, 1992

94. Respiratory syncytial virus activity—United States, 2000–01 season. MMWR Morb Mortal Wkly Rep 51:26, 2002

95. Falsey AR, McCann RM, Hall WJ, et al: Acute respiratory tract infection in daycare centers for older persons. J Am Geriatr Soc 43:30, 1995

96. Whimbey E, Champlin RE, Englund JA, et al: Combination therapy with aerosolized ribavirin and intravenous immunoglobulin for respiratory syncytial virus disease in adult bone marrow transplant recipients. Bone Marrow Transplant 16:393, 1995

97. Revised indications for the use of palivizumab and respiratory syncytial virus immune globulin intravenous for the prevention of respiratory syncytial virus infections. American Academy of Pediatrics Committee on Infectious Diseases and Committee on Fetus and Newborn. Pediatrics 112:1442, 2003

98. Boeckh M, Berrey MM, Bowden RA, et al: Phase 1 evaluation of the respiratory syncytial virus–specific monoclonal antibody palivizumab in recipients of hematopoietic stem cell transplants. J Infect Dis 184:350, 2001

99. Contreras PA, Sami IR, Darnell ME, et al: Inactivation of respiratory syncytial virus by generic hand dishwashing detergents and antibacterial hand soaps. Infect Control Hosp Epidemiol 20:57, 1999

100. Karanfil LV, Conlon M, Lykens K, et al: Reducing the rate of nosocomially transmitted respiratory syncytial virus. Am J Infect Control 27:91, 1999

101. Madge P, Paton JY, McColl JH, et al: Prospective controlled study of four infection-control procedures to prevent nosocomial infection with respiratory syncytial virus. Lancet 340:1079, 1992

102. Ison MG, Hayden FG, Kaiser L, et al: Rhinovirus infections in hematopoietic stem cell transplant recipients with pneumonia. Clin Infect Dis 36:1139, 2003

103. Arruda E, Pitkaranta A, Witek TJ, et al: Frequency and natural history of rhinovirus infections in adults during autumn. J Clin Microbiol 35:2864, 1997

104. Wald TG, Shult P, Krause P, et al: A rhinovirus outbreak among residents of a long-term care facility. Ann Intern Med 123:588, 1995

105. Nicholson KG, Kent J, Hammersley V, et al: Acute viral infections of upper respiratory tract in elderly people living in the community: comparative, prospective, population based study of disease burden. BMJ 315:1060, 1997

106. Pitkaranta A, Arruda E, Malmberg H, et al: Detection of rhinovirus in sinus brushings of patients with acute community-acquired sinusitis by reverse transcription–PCR. J Clin Microbiol 35:1791, 1997

107. Pitkaranta A, Virolainen A, Jero J: Detection of rhinovirus, respiratory syncytial virus, and coronavirus infections in acute otitis media by reverse transcriptase polymerase chain reaction. Pediatrics 102:291, 1998

108. Andeweg AC, Bestebroer TM, Huybreghs M, et al: Improved detection of rhinoviruses in clinical samples by using a newly developed nested reverse transcription–PCR assay. J Clin Microbiol 37:524, 1999

109. Hayden FG, Herrington DT, Coats TL, et al: Efficacy and safety of oral pleconaril for treatment of colds due to picornaviruses in adults: results of 2 double-blind, randomized, placebo-controlled trials. Clin Infect Dis 36:1523, 2003

110. Update: Hantavirus pulmonary syndrome—United States, 1999. MMWR Morb Mortal Wkly Rep 48:521, 1999

111. Jefferson TO, Demicheli V, Deeks JJ, et al: Amantadine and rimantadine for preventing and treating influenza A in adults. Cochrane Database Syst Rev (2):CD001169, 2000

112. Jefferson T, Demicheli V, Deeks J, et al: Neuraminidase inhibitors for preventing and treating influenza in healthy adults. Cochrane Database Syst Rev (2):CD001265, 2000

XXVI HERPESVIRUS INFECTIONS

MARTIN S. HIRSCH, M.D.

Human and Animal Herpesviruses

The herpes group of viruses is composed of at least eight human viruses and numerous animal viruses [*see 7:XXXI Viral Zoonoses*]. The human herpesviruses include herpes simplex virus types 1 and 2 (HSV-1 and HSV-2), varicella-zoster virus (VZV), cytomegalovirus (CMV), Epstein-Barr virus (EBV), and human herpesvirus types 6 (HHV-6), 7 (HHV-7), and 8 (HHV-8, also known as Kaposi sarcoma–associated herpesvirus [KSHV]).

All human herpesviruses are of similar size and morphology and are characterized by a core, 30 to 60 nm in diameter, that contains a double-stranded DNA genome; a nucleocapsid, 95 to 100 nm, that exhibits icosahedral symmetry; and a lipoprotein envelope with glycoprotein projections that has a diameter of 120 to 250 nm. Herpesviruses replicate primarily in cell nuclei, and a protein envelope is added as the virus passes through the nuclear membrane.

Human herpesviruses share the properties of latency and reactivation. Members of the group can cause productive lytic infections, in which infectious virus is produced and cells are killed, or nonproductive lytic infections, in which viral DNA persists but complete replication does not occur and cells survive. After acute lytic infections, herpesviruses often persist in a latent form for years; periodic reactivations are followed by recurrent lytic infections. Sites of latency vary: HSV and VZV persist in neural ganglion cells, EBV persists in B cells, and CMV probably remains latent in many cell types. The sites of latency for HHV-6 and HHV-7 have not been identified, although both herpesviruses have been detected in salivary glands.

All human herpesviruses have a worldwide distribution [*see Sidebar* Herpesvirus Information on the Internet]. Considerable efforts are being directed toward the development of vaccines and antiviral agents that will be active against herpesviruses.

Herpes Simplex Virus

HSV-1 and HSV-2 can be distinguished by a variety of properties, including clinical and epidemiologic patterns, antigenicity, DNA base composition, biologic characteristics, and sensitivity to various physical and chemical stresses.[1] Advances in molecular biology technology have proved that HSV-1 and HSV-2 share certain antigens (e.g., glycoprotein B) but differ with respect to other antigens (e.g., glycoprotein G). Restriction enzyme analysis of HSV DNA and other molecular techniques are used to identify individual isolates.

EPIDEMIOLOGY AND ETIOLOGY

Humans are the only known natural hosts for HSV, although animals can readily be infected experimentally. HSV-1 primary infection occurs mainly in childhood, whereas HSV-2 infection occurs predominantly in sexually active adolescents and young adults. The prevalence of HSV-2–specific antibodies in the United States increases from less than 6% in those younger than 19 years to more than 25% in those older than 30 years.[2] In older age groups, changes in prevalence are negligible. Independent predictors of HSV-2 seropositivity include female gender, black race, increasing age, less education, more lifetime sex partners, prior occurrence of syphilis or gonorrhea, and lack of HSV-1 antibody.[3] In the United States, changes in sexual mores resulted in an age-adjusted seroprevalence of HSV-2 infection in the 1990s that was 30% higher than the seroprevalence in the 1970s; HSV-2 is now detectable in one of five persons older than 12 years.[2] Approximately one third of new HSV-2 infections are symptomatic.[4] Prevalence of antibody to HSV-1 in different populations ranges from 44% in persons 12 to 19 years of age to above 80% in those older than 60 years.[2] Between 10% and 25% of adults are dually infected with HSV-1 and HSV-2.[2]

Direct contact with infected secretions is the principal mode of transmission. HSV-1 is usually transmitted by an oral route and HSV-2 by a genital route. Although virus titers are higher and

Herpesvirus Information on the Internet

General

http://www.nesfile.com/xlp.htm
Herpes Viruses Weekly

http://www.cdc.gov/std/treatment/TOC2002TG.htm
Sexually Transmitted Disease Treatment Guidelines

http://www.cdc.gov/mmwr/preview/mmwrhtml/00021095.htm
Recommended infection-control practices for dentistry

http://www.cdc.gov/mmwr/preview/mmwrhtml/00001053.htm
Perspectives in disease prevention and health promotion: condoms for prevention of sexually transmitted diseases

HSV

http://www.cdc.gov/std/Herpes/STDFact-Herpes.htm
Patient information on genital herpes

http://www.cdc.gov/mmwr/preview/mmwrhtml/00047449.htm
Case definitions for infectious conditions under public health surveillance

VZV

http://www.cdc.gov/nip/diseases/varicella/vac-chart.htm
Varicella patient information in brief

http://www.cdc.gov/nip/diseases/varicella
Frequently asked questions about varicella and varicella vaccine

http://www.cdc.gov/od/oc/media/fact/chickenp.htm
Facts about chickenpox (varicella) from the CDC Office of Communications, Division of Media Relations

http://www.cdc.gov/epo/mmwr/preview/mmwrhtml/00022690.htm
ACIP recommendation: varicella-zoster immune globulin for the prevention of chickenpox

http://www.cdc.gov/mmwr/preview/mmwrhtml/00042990.htm
Recommendations and Reports—June 27, 1996/Vol. 45/No. RR-1-Preview (The first statement by the Advisory Committee on Immunization Practices [ACIP] on the use of live, attenuated varicella virus vaccine)

CMV

http://www.cdc.gov/ncidod/diseases/cmv.htm
Patient information on cytomegalovirus

http://www.cdc.gov/hiv/pubs/brochure/oi_cmv.htm
Division of HIV/AIDS Prevention Brochures: You Can Prevent CMV

Herpesvirus Simiae

http://www.cdc.gov/epo/mmwr/preview/mmwrhtml/00015936.htm
Guidelines for prevention of herpesvirus simiae (B virus) infection

transmission is more likely when lesions are present, asymptomatic excretion of the virus is common. HSV-2 shedding from the genital tract can occur in seropositive persons who have no history of genital HSV infection.[5] Thus, transmission occurs frequently, even in the absence of lesions. HSV-2 is transmitted more efficiently from males to females than from females to males. Autoinoculation to other skin sites also occurs, more often with HSV-2 than with HSV-1. Spread of infection through contact with oral secretions may be an occupational hazard for respiratory care and dental care providers; thus, gloves should be worn when fingers are placed in patients' mouths. Fomites, including toilet seats and towels, are not important modes of transmission. Recurrences are frequent with both HSV-1 and HSV-2 infections, usually as a result of endogenous reactivation. In the United States, lip or perioral recurrences develop in 20% to 40% of the population. Precipitating factors are sunlight, wind, local trauma, fever, menstruation, and emotional stress. Ocular herpes is present in about 5% of all patients seen at ophthalmology clinics; 25% to 50% of ocular HSV infections recur within 2 years. Of all the primary cases of genital herpes in the United States each year, 60% to 80% will recur. Although most genital recurrences represent reactivation, exogenous reinfection can also occur. Clinically significant recurrences tend to decrease over time.[6]

PATHOGENESIS

After the initial replication of the virus in epithelial cells, cytolysis and local inflammatory reactions develop, resulting in the characteristic lesion—a superficial vesicle on an inflammatory base. Multinucleated cells and Cowdry type A inclusion bodies are present. Subsequent lymphatic spread to regional nodes and viremic spread to other organs may occur, depending on the immune competence of the host. Viremia can be demonstrated in malnourished children, in certain adults with depressed T cell–mediated immunity, and, occasionally, in immunocompetent persons.

After initial infection, HSV travels along sensory nerve pathways to ganglion cells, the site of latent infection. The viral DNA persists, only to become reactivated by certain stresses. After reactivation, the virus reverses its course and spreads peripherally by sensory nerve pathways. Once HSV reaches cutaneous sites, cell-to-cell spread occurs until host immune mechanisms limit further dissemination. Various mechanisms are engaged in host responses to HSV, including T cell and natural killer cell cytotoxicity, macrophage activation, production of antibody, and production of interferon.

CLINICAL SYNDROMES

Oral-Labial Herpes

In patients younger than 5 years, primary HSV-1 infection is most often asymptomatic; when symptomatic, it presents as gingivostomatitis or pharyngitis. After an incubation period of 2 to 12 days, fever and sore throat develop. Small vesicles are observed on the oral mucosa and pharynx. Mouth pain may be severe, breath is fetid, and cervical adenopathy is present. In adolescents and young adults, posterior pharyngitis and tonsillitis may be the primary problem. The differential diagnosis includes streptococcal pharyngitis, aphthous stomatitis, Stevens-Johnson syndrome, herpangina, and infectious mononucleosis. Symptoms and signs often persist for 10 to 21 days, and autoinoculation to other sites, such as fingers and eyes, is common. Oral shedding may persist for several weeks (mean, 7 to 10 days).

Recurrent herpes labialis is a shorter and milder affliction, often heralded by local pain or tingling for a few hours. HSV-1 oral-labial lesions recur more often than HSV-2 oral-labial lesions.[3] Vesicles appear most often on the vermilion border and are painful. The lesion is usually small (< 1 cm^2), and the progression from vesicle to ulcer to crust is rapid (< 96 hours). Healing is complete within 8 to 10 days. Systemic complaints are uncommon in recurrent herpes labialis.

Ocular Herpes

Most ocular herpetic infections are caused by HSV-1. Primary infections may present as unilateral follicular conjunctivitis, blepharitis, or corneal epithelial opacities. Healing is usually complete within 2 to 3 weeks. Recurrences may take the form of keratitis (more than 90% of cases are unilateral), blepharitis, or keratoconjunctivitis. Branching dendritic ulcers, usually detected by fluorescein staining, are virtually diagnostic and are often associated with diminished visual acuity. Deep stromal involvement may result in scarring, corneal thinning, and abnormal vascularization, with resultant blindness or rupture of the globe.

Genital Herpes

HSV-2 is the causative agent in 70% to 95% of primary genital herpesvirus infections. After an incubation period of 2 to 7 days, fever, malaise, and inguinal adenopathy develop; these symptoms are associated with the appearance of vesicular lesions. In men, lesions are often on the glans penis or penile shaft [*see Figure 1*]; in women, lesions may involve the vulva, perineum, buttocks, cervix, or vagina. The differential diagnosis includes syphilis, chancroid, Behçet syndrome, erythema multiforme, and candidiasis. In women, lesions rapidly ulcerate and become covered with exudate, with resultant vaginal discharge. Urethral involvement sometimes results in dysuria or urinary retention. Signs and symptoms of primary genital herpes often persist for several weeks before complete healing. Previous infection with HSV-1 may reduce the severity and duration of a first episode of genital herpes caused by HSV-2.

Extragenital lesions develop during the course of primary infection in 10% to 18% of patients. Aseptic meningitis is not uncommon during primary genital herpes, particularly in women, and in rare instances, herpetic sacral radiculomyelitis occurs.

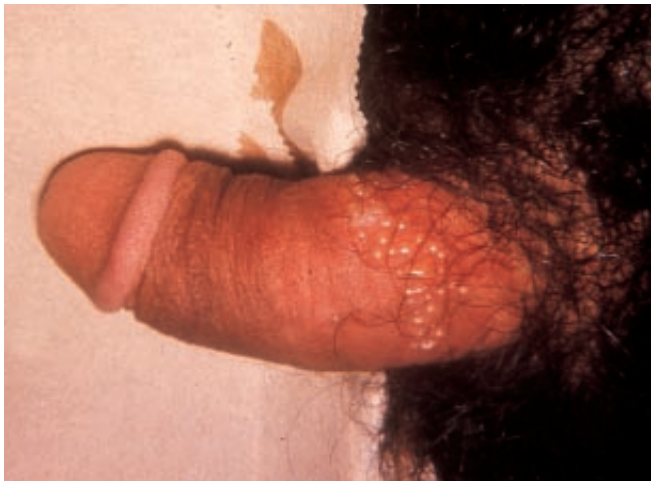


Figure 1 **Herpes simplex virus genital lesions in men often present as grouped vesicles on the penile shaft.**

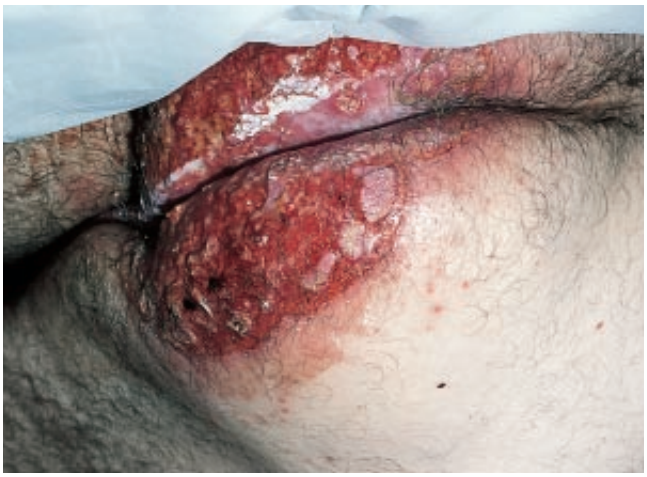


Figure 2 **Perianal herpes simplex infections in patients with compromised immunity may be severe and prolonged.**

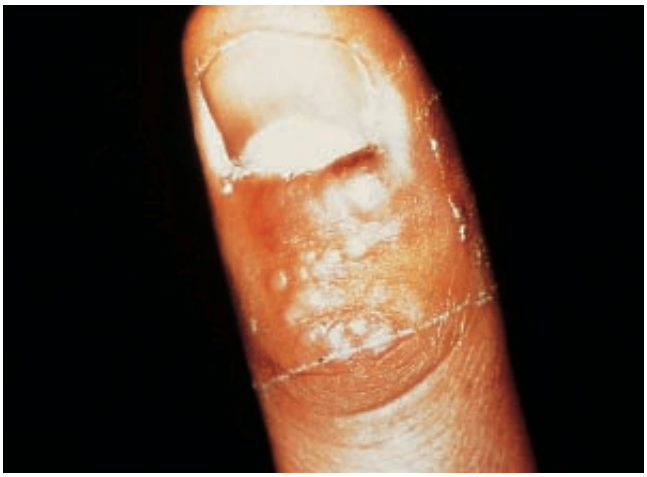


Figure 3 **Herpetic whitlows, which occur on fingers, are often misdiagnosed as staphylococcal infections.**

Urinary retention may occasionally complicate primary genital herpes, particularly in women.

Recurrent episodes of genital herpes are usually shorter and milder than primary episodes but still affect women more severely than men. Genital HSV-2 infections recur more often than genital HSV-1 infections, and on rare occasions, the two infections can be found simultaneously in lesions.[7,8] Moreover, patients dually infected with HSV-1 and HSV-2 have fewer recurrences of genital herpes than those infected with HSV-2 alone.[2] With either virus type, after a variable prodrome of tenderness, itching, or tingling, lesions develop on the penis, labia minora, labia majora, perineum, mons pubis, or buttocks. Healing occurs in 6 to 10 days and is usually uncomplicated; frequent asymptomatic shedding occurs, however, particularly in women.[5,9] HSV-2 may also cause benign, recurrent lymphocytic meningitis that may be associated with recurrent genital lesions.[10]

Perianal and Anal Herpes

Perianal and anal HSV-2 infection is an important problem in men who have sex with men. Pain, itching, tenesmus, discharge, fever, chills, sacral paresthesias, headache, and difficulty in urinating may all occur. Vesicles and ulcerations may lead to an erythematous cryptitis with inguinal adenopathy. Herpes proctitis is often prolonged and severe in patients with AIDS [*see Figure 2*].

Other Herpes Syndromes

Herpetic whitlow Primary finger infections, or whitlows, usually involve one digit and are characterized by intense itching or pain followed by the formation of deep vesicles that may coalesce [*see Figure 3*]. Among the general public, whitlows are most often caused by HSV-2, whereas among medical and dental personnel, HSV-1 is the principal culprit. The lesions gradually resolve in 2 to 3 weeks, unless they are mistakenly incised, in which case healing may be delayed by secondary bacterial infection. Recurrent whitlows commonly appear and are sometimes associated with severe local neuralgia.

Neurologic complication Encephalitis, a severe form of HSV infection, is discussed elsewhere [*see 11:XVI Acute Viral Central Nervous System Diseases*]. HSV-1 has also been implicated as an etiologic agent in Bell palsy[11] and in rare cases of recurrent self-limited meningitis (Mollaret meningitis).[12]

Infection in the immunocompromised host Disorders of T cell–mediated immunity are associated with more severe HSV infections. In clinical settings such as organ transplantation, lymphoreticular neoplasm, or AIDS, HSV infection is often slow to heal and may disseminate cutaneously or to visceral organs. Certain skin conditions, such as eczema and burns, are associated with cutaneous but not visceral dissemination. In rare instances, HSV infection during pregnancy or in the elderly is complicated by visceral dissemination, particularly to the liver. Intubation or catheterization of debilitated patients may facilitate the spread of infection; for instance, herpes esophagitis often complicates long-term use of nasogastric tubes.

Neonatal infection Between one in 2,500 and one in 10,000 births are complicated by HSV infection, usually HSV-2. HSV neonatal infection can be localized or disseminated and results from transmission of the virus to the infant at the time of delivery, either by ascending infection after premature membrane rupture or by passage of the infant through an infected genital tract. The risk of transmission is increased in premature births, after prolonged membrane rupture, and with the use of fetal scalp monitor electrodes. About 40% to 50% of infants born to mothers with primary infections are at risk for the development of severe disease, whereas fewer than 8% of those born to women with recurrent herpes are at risk for severe disease.[1,13] Maternal infection is often asymptomatic at the time of delivery, and asymptomatic shedding may occur in women with no known history of genital herpes. In women with a history of genital HSV infection, genital HSV can be detected at delivery in approximately 2% by use of tissue culture and in 14% by use of polymerase chain reaction (PCR) techniques.[14]

Infection becomes apparent several days to weeks after delivery. Newborns often present with vesicles or conjunctivitis, or a syndrome resembling neonatal sepsis may be evident. Neurologic signs such as seizures, cranial nerve palsies, and lethargy often predominate and are accompanied by cerebrospinal fluid pleocytosis. Disseminated infection may involve the liver, lungs, or adrenal glands. If untreated, disseminated or central nervous system infection is fatal in more than 70% of patients, whereas localized disease is generally self-limited. Treatment has greatly

reduced the mortality from severe infection [*see 11:XVI Acute Viral Central Nervous System Diseases*].

DIAGNOSIS

A variety of tissue culture systems support the replication of HSV, and virus isolation from specimens collected early in the course of infection is the diagnostic method of choice. The characteristic cytopathic effect of the virus is often detectable within a period of 24 to 48 hours. Typing of isolates can be accomplished most readily by immunofluorescence with monoclonal antibodies directed against type-specific antigens. Scrapings or tissue specimens can sometimes be tested directly for herpesvirus antigens by immunofluorescence or immunohistochemistry. Alternatively, scrapings may be prepared by Giemsa or Wright stain and examined for the presence of multinucleated giant cells, which indicates infection with HSV or VZV. Serologic techniques that accurately differentiate HSV-1 from HSV-2 infections are now commercially available. Such tests can be used to confirm a diagnosis of primary HSV infection, but they are seldom helpful in diagnosing recurrences. Serologic techniques can also establish a diagnosis in patients with atypical complaints, identify asymptomatic carriers, and identify persons at risk.[15] PCR detection of HSV DNA in CSF has become the standard means of diagnosing HSV encephalitis. For patients with HSV encephalitis, PCR results are often positive within 24 hours of the onset of symptoms, and test results may remain positive during the first week of illness.[16]

PREVENTION

No HSV vaccine has been approved for general use, although preliminary studies of glycoprotein-D-adjuvant vaccines suggest efficacy in certain populations (i.e., women seronegative for both HSV-1 and HSV-2) but not others.[17] Prophylactic measures that prevent contact with the virus may help in avoiding primary infection. For instance, the use of condoms may prevent sexual transmission when either sexual partner has a history of genital HSV infection. Application of sunscreens to susceptible skin areas before exposure to ultraviolet light can prevent reactivation of HSV. Medical and dental personnel who treat HSV-positive patients should wear gloves to prevent contact with infected areas.

The best strategy for the prevention of neonatal herpes may be close physical examination at the time of labor.[18] Cesarean section is indicated to prevent perinatal infection when lesions consistent with the diagnosis of genital herpes are noted during labor. If primary genital herpes is detected during the third trimester of pregnancy, acyclovir in conventional doses should be considered for the mother during the peripartum period and for the newborn post partum [*see* Treatment, *below*]. Acyclovir in late pregnancy may also be useful in women with recurrent genital HSV infections; placebo-controlled trials have shown that 400 mg three times daily reduced lesions and HSV excretion at delivery.[19,20]

TREATMENT

Two decades of carefully controlled trials and clinical experience have made acyclovir the treatment standard for HSV infections, although the prodrug valacyclovir (which is converted to acyclovir) and famciclovir (which is converted to penciclovir) now provide alternative options [*see Table 1*].[1] Either oral acyclovir (200 mg five times daily) or intravenous acyclovir (5 mg/kg three times daily) for 7 to 10 days is recommended for treatment of primary genital herpes or mucocutaneous herpes (HSV-1 or HSV-2). Alternatively, valacyclovir (500 to 1,000 mg twice daily) or famciclovir (250 mg three times daily) may be used. Short-term (1 to 3 days), high-dose acyclovir or valacyclovir regimens are promising therapies for both genital and orolabial recurrent HSV infections.[21,22]

Suppression of severe and frequently recurring genital herpes can also be accomplished by the administration of oral acyclovir (200 to 400 mg twice daily), oral famciclovir (250 mg twice daily), or oral valacyclovir (500 to 1,000 mg daily) [*see Table 2*]. Once-daily valacyclovir (500 mg) not only suppresses recurrences of genital herpes but also reduces genital excretion of HSV-2 and transmission to partners.[23] Intravenous acyclovir, in a dosage of 10 to 15 mg/kg three times daily for 14 to 21 days, is the treatment of choice for herpes encephalitis. Neonatal HSV infections are treated with intravenous acyclovir in a dosage of 20 mg/kg every 8 hours for 14 to 21 days.

In immunocompromised persons and, in rare instances, in immunocompetent persons, acyclovir-resistant HSV-2 may lead to chronic progressive infections.[24] Intravenous foscarnet (40 mg/kg two to three times daily for 2 to 3 weeks) is useful against acyclovir-resistant HSV-2 infection. Topical preparations of foscarnet, cidofovir, and trifluridine are also under study for acyclovir-resistant HSV.[25-27]

Varicella-Zoster Virus

VZV is the causative agent of varicella (chickenpox) and herpes zoster (shingles). Because it is a strongly cell-associated member of the human herpesvirus group, it is difficult to detect or isolate in cell-free specimens. Its host range is largely limited to human cells. As with other herpesviruses, latency and reactivation are characteristic features of VZV.

EPIDEMIOLOGY

Varicella

Varicella is a highly communicable disease of childhood. More than 90% of patients are children younger than 9 years; most adults have antibody to the virus. Infection is spread by respiratory droplets or by direct contact with active lesions. Fomites are not an important mode of transmission. Chickenpox occurs most frequently in the winter and the spring and spreads rapidly to susceptible household members. The incubation period ranges from 11 to 20 days.

Herpes Zoster

Herpes zoster results from the reactivation of VZV infection. Varicella in one patient cannot produce herpes zoster in another; however, persons who are exposed to patients who have herpes zoster can contract varicella. There is no seasonal pattern for herpes zoster. The likelihood of reactivation is related to both age and immune status.[28] The incidence of disease in patients older than 60 years is two to three times higher than that in younger persons. The incidence in immunosuppressed patients, particularly those with depressed T cell function (e.g., patients who have lymphomas, leukemias, organ transplants, or AIDS), is as much as 100 times higher than that in immunocompetent hosts.

Herpes zoster occurs more frequently in patients with neoplasms. The risk of developing herpes zoster is 13% to 15% for patients with Hodgkin disease, 7% to 9% for patients with other lymphomas, and 1% to 3% for patients with solid tumors. Approximately 15% to 30% of patients with Hodgkin disease and

Table 1 Drug Treatment for Primary Genital or Mucocutaneous Herpes Infection

Drug	*Dose*	*Efficacy Rating*	*Cost/Mo*†	*Comment*
Acyclovir*	200 mg p.o. five times a day for 7–10 days or 5 mg/kg I.V. t.i.d. for 7–10 days	First choice	Oral, $34.98	May be useful in high-dose short-term regimens
Valacyclovir*	500–1,000 mg p.o., b.i.d.	Alternative	$437.94	May be used in high-dose (1–2 g), short-term (1–2 day) regimens
Famciclovir*	250 mg p.o., t.i.d.	Alternative	$286.99	Used when its dosing of three times a day is preferred to acyclovir's five times a day
Foscarnet	40 mg/kg I.V. b.i.d. to t.i.d. for 2–3 wk	Alternative	NA	For chronic acyclovir-resistant HSV-2 infection

*Short-term, high-dose regimens show promise.
†Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.
HSV-2—herpes simplex virus type 2 NA—not available

herpes zoster exhibit significant dissemination; however, mortality in such patients is low. Localized and disseminated herpes zoster are more common in patients with advanced malignant disease who are receiving intensive chemotherapy or irradiation. The incidence of herpes zoster is also increased after organ or bone marrow transplantation. The incidence of herpes zoster is approximately 8.6% in solid-organ transplant recipients, and it may be more common in women and in patients who are following certain immunosuppressive and prophylactic antiviral regimens.[29] In patients with advanced HIV-1 infection, chronic, nonhealing VZV ulcers may persist for several months, and herpes zoster may be precipitated by the introduction of highly active antiretroviral therapy.

CLINICAL SYNDROMES

Varicella

After local replication of VZV at the site of virus entry, the virus is carried by blood leukocytes to focal skin areas and visceral organs. Replication at skin sites results in vesicle formation accompanied by degeneration of epithelial cells, accumulation of edema fluid, and infiltration of inflammatory cells; multinucleated giant cells containing intranuclear inclusions are present.

Clinical features Subclinical infections are rare (< 4% of cases). Illness usually presents as a generalized vesicular eruption that is followed by a variable prodrome of headache, fever, and malaise. Vesicles often appear initially on the face or trunk [*see Figure 4*]. During the next 4 to 7 days, successive crops of lesions appear. Most childhood cases are uncomplicated and resolve in 7 to 10 days without producing scarring.

Complications In adolescents and adults, pneumonia may develop that is characterized by cough, tachypnea, and diffuse reticulonodular infiltrates, which may calcify with time. Pulmonary oxygen diffusion difficulties may persist for a period of weeks to months. Varicella pneumonia may be particularly severe in pregnant women and, among them, is more common in smokers and those with more than 100 skin lesions.[30] Although varicella encephalopathies are uncommon, they can occur at any age and can take several forms; the presenting feature may be hyperexcitability, which can progress to coma or cerebellar ataxia. Other rare complications of varicella infection are Reye syndrome (fulminating encephalopathy and fatty liver), thrombocytopenia, arthritis, ocular involvement, carditis, gastrointestinal bleeding, nephritis, orchitis, and bacterial superinfection. A pregnant woman who contracts varicella during the first 20 weeks of pregnancy has a small (< 2%) risk of giving birth to an infant with congenital varicella syndrome, which includes limb hypoplasia, skin scarring, and ocular defects.[31]

Disseminated disease is more likely to develop after varicella infection in children with leukemia, lymphoma, or Hodgkin disease. Visceral dissemination occurs in 20% to 35% of children with cancer who are receiving chemotherapy; mortality in such patients is 7% to 30%. Clinical reinfection, with resultant atypical generalized varicella, can occur in the severely immunocompromised host; such reinfection tends to be mild and self-limited.

Herpes Zoster

Herpes zoster occurs in patients previously infected with varicella. After an initial episode of varicella infection, the virus persists in neurons of sensory ganglia in a relatively quiescent, ex-

Table 2 Drug Treatment for Suppression of Severe and Frequently Recurring Genital Herpes Infection

Drug	*Dose*	*Efficacy Rating*	*Cost/Mo**	*Comment*
Acyclovir	200–400 mg p.o., b.i.d.	First choice	$27.98	
Valacyclovir	500–1,000 mg/day p.o.	Alternative	$218.97	Has been shown to decrease viral excretion and transmission
Famciclovir	250 mg p.o., b.i.d.	Alternative	$199.98	

*Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.

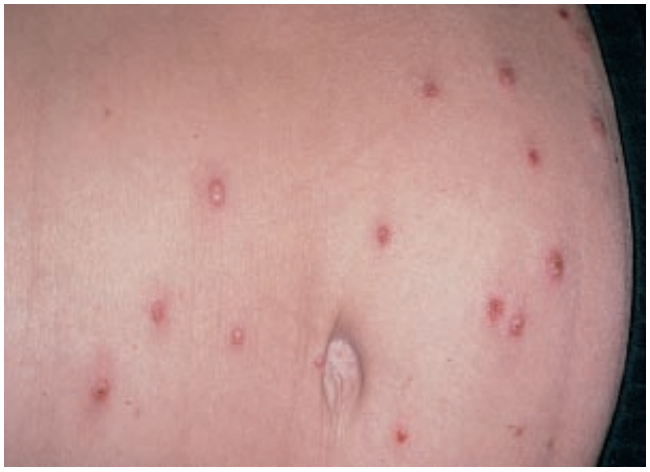


Figure 4 **The lesions characteristic of varicella-zoster virus infection often appear initially on the face and trunk and may present as maculopapules, vesicles, and scabs simultaneously.**

trachromosomal, cytoplasmic state.[32,33] Both subclinical and clinical reactivations occur. Clinical VZV reactivation is enhanced by many factors, including increasing age, immunosuppression, and local skin injury.

Clinical features Attacks of herpes zoster are often preceded by pain, which may persist for several days before lesions appear. The vesicular eruption is unilateral and most often appears on the thorax [*see Figure 5*]. Individual attacks are usually limited to one to three dermatomes, although a few isolated skin lesions at sites distant from this area are not uncommon. Vesicles surrounded by an erythematous base may continue to develop for several days; the lesions then dry and crust. Superinfections are common, and scarring can occur.

Complications The most common complication of herpes zoster is postherpetic neuralgia, which may be severe, particularly in the elderly. Postherpetic neuralgia may be refractory to treatment and persist for months to years [*see 11:XVI Acute Viral Central Nervous System Diseases*]. Other neurologic syndromes associated with herpes zoster are segmental myelitis, Guillain-Barré syndrome, acute retinal necrosis, and Ramsay Hunt syndrome.[33] Ramsay Hunt syndrome is an infection of the geniculate ganglion of the seventh cranial nerve, producing facial paralysis; vesicles on the eardrum and side of the tongue can also occur.

VZV encephalitis, another infrequent complication of herpes zoster, may present in two forms.[33] Large vessel encephalitis (granulomatous arteritis) occurs predominantly in immunocompetent persons; it is characterized by acute focal deficits that occur weeks after the appearance of zoster lesions in the contralateral trigeminal pattern of distribution. Angiography demonstrates segmental narrowing, primarily in the internal carotid and middle or anterior cerebral arteries. Microscopic examination shows arterial inflammation, multinucleated giant cells, and virus particles. Small vessel encephalitis occurs almost exclusively in immunocompromised persons and is characterized by headache, fever, mental changes, seizures, and focal deficits. Imaging studies show ischemic or hemorrhagic infarcts of gray and white matter, and CSF examination demonstrates mononuclear pleocytosis and both VZV DNA and VZV antibody.

DIAGNOSIS

Varicella and herpes zoster can usually be diagnosed solely on clinical grounds. Varicella is sometimes confused with disseminated herpes simplex, impetigo, insect bites, or scabies; on rare occasions, HSV infection presents with zosteriform eruptions. Diagnosis of VZV infection can be confirmed by virus isolation, direct immunofluorescence of lesions, PCR detection of viral DNA, or demonstration of a fourfold rise in antibodies to viral antigens. Herpes zoster provokes a more rapid and greater rise in antibodies than varicella.

PREVENTION

A live attenuated VZV vaccine is available in the United States and appears both safe and effective. It is recommended for persons older than 1 year who are in good health and have no history of clinical VZV infection. A single subcutaneous dose is recommended for children 1 to 12 years of age, whereas two doses separated by 4 to 8 weeks are recommended for susceptible adolescents and adults. Varicella has declined markedly in surveillance areas secondary to vaccine use.[34,35] However, breakthrough infections in vaccinated children suggest that booster doses of vaccine may be necessary to maintain optimal immunity.[36]

Nosocomial transmission of VZV has been reported. Thus, hospitalized patients with varicella or herpes zoster should be isolated to prevent spread of virus to other susceptible persons. Susceptible immunocompromised patients who are exposed to infected persons should receive prophylaxis with varicella-zoster immune globulin (one dose up to 4 days after exposure) to prevent or modify clinical illness. Small studies suggest that high-dose acyclovir (40 to 80 mg/kg for 7 days, beginning 7 to 9 days after exposure) or VZV vaccine (one dose on the day of exposure or up to 3 days after exposure) may also be useful in postexposure prophylaxis.[32]

TREATMENT

Considerable success in the treatment of VZV infections in immunocompromised hosts has been achieved with parenteral administration of acyclovir (for adults, 10 mg/kg every 8 hours for 7 days). High-dose oral acyclovir (800 mg five times

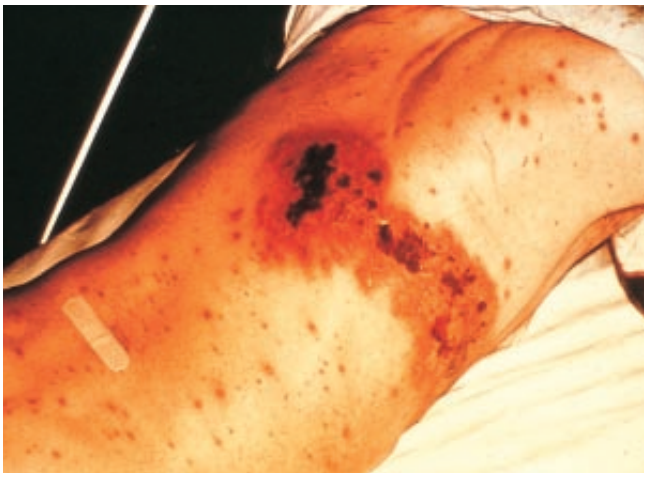


Figure 5 **Herpes zoster is most often thoracic and dermatomal in distribution. Cutaneous dissemination, as seen in this patient, as well as visceral dissemination, may occur when immunity is severely compromised.**

Table 3 Drug Treatment for Varicella-Zoster Virus Infection

Drug	*Dose*	*Efficacy Rating*	*Cost/Mo**	*Comment*
Acyclovir	10 mg/kg I.V. q. 8 hr for 7 days	First choice	NA	For immunocompromised patients
	800 mg p.o. five times/day for 7 days	Alternative	$89.98	May shorten course and reduce severity of herpes zoster and varicella in otherwise healthy patients
Valacyclovir	1 g p.o., t.i.d.	Alternative	$542.99	May shorten course and reduce severity of herpes zoster and varicella in otherwise healthy patients
Famciclovir	500 mg p.o., t.i.d.	Alternative	$599.97	May shorten course and reduce severity of herpes zoster and varicella in otherwise healthy patients
Foscarnet	40 mg/kg I.V. q. 8 hr	Alternative	NA	For acyclovir-resistant VZV infection in immunocompromised patients

*Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.
NA—not available

daily for 7 days), when begun early, may also shorten the course and reduce the severity of herpes zoster in otherwise healthy hosts. Oral valacyclovir (1 g three times daily) or famciclovir (500 mg three times daily) may also be used [*see Table 3*]. The use of corticosteroids in combination with acyclovir for the treatment of acute herpes zoster remains controversial.[37,38] Steroid use may be justified in persons older than 50 years who have no relative contraindications, such as diabetes, hypertension, or glaucoma.[28]

Acyclovir resistance has been described in VZV isolates from patients with HIV infection who received long-term acyclovir therapy. Foscarnet (40 mg/kg every 8 hours) is the drug of choice for acyclovir-resistant VZV infections in immunocompromised hosts [*see Table 3*].[39] Acyclovir, famciclovir, and valacyclovir are also effective in reducing the severity and shortening the course of chickenpox in patients who are immunologically competent, although children are generally not treated for uncomplicated varicella.

Postherpetic neuralgia, the most common complication of herpes zoster, is difficult to treat, particularly in elderly patients; suggested treatments include topical anesthetics, oral analgesics, tricyclic antidepressants, gabapentin, and intrathecal methylprednisolone acetate.[28]

Cytomegalovirus

Although cytomegalic inclusion disease of infants was recognized as a clinical entity at the turn of the 20th century, the etiologic agent, CMV, was not isolated until the mid-1950s. CMV has since become recognized as a nearly ubiquitous virus that plays an important role in many diseases, including a form of mononucleosis and disseminated disease in the immunocompromised host.

EPIDEMIOLOGY AND ETIOLOGY

CMV has a worldwide distribution. Approximately 1% of newborns in developed countries are infected with CMV, and the incidence is much higher in developing areas. In immunocompromised patients (e.g., organ or bone marrow transplant recipients and patients with AIDS), CMV infection is a major cause of morbidity and mortality.

Transmission occurs by close contact with infected persons, most commonly by the sharing of secretions or leukocyte-containing blood products. CMV is shed in semen, cervical secretions, saliva, maternal milk, and urine. In late adolescence and adulthood, genital transmission is common. Primary infection in pregnant women may result in transplacental spread to the fetus, sometimes with devastating consequences. Naturally acquired immunity substantially reduces the likelihood of congenital infection, although infection with a different CMV strain can occur and lead to symptomatic infection of the newborn.[40,41] Perinatal infection is also common because of spread from maternal milk or other secretions. Communal living and poor personal hygiene facilitate transmission. Spread of CMV in day care centers is common; the infected child often serves as a source of transmission to other family members. Blood transfusions and organ transplantation may also transmit CMV from an asymptomatic donor to a recipient. The proportion of CMV antibody–negative liver transplant recipients receiving organs from CMV antibody–positive donors appears to be increasing.[42]

PATHOGENESIS

Once infected, one probably carries the virus for life. Generally, such infections remain dormant, but reactivation occurs during periods of immune compromise. Reactivation is particularly common in patients with compromised T cell function (e.g., transplant recipients or patients with lymphoid neoplasms or AIDS). Reinfections with new strains of CMV can also occur in immunosuppressed patients.

In the fetus, primary infection may result in disseminated cytomegalic inclusion disease with numerous congenital abnormalities, in localized disease affecting the auditory system or the CNS, or in subclinical involvement. Perinatal infection is usually asymptomatic, although protracted interstitial pneumonitis, hepatitis, and failure to thrive may result.

Symptomatic disease is more likely to be produced by primary CMV infection than by viral reactivation. The sites of latent infection are poorly defined, although it appears likely that CMV persists in numerous cell types in many organs. If host immune responses become compromised, virus reactivation and increased replication can occur and result in various syndromes. CMV replication can further suppress immune responses, leading to profound lymphocyte hyporesponsiveness and severe opportunistic infections with protozoa (e.g., *Pneumocystis carinii* and *Toxoplasma gondii*), fungi, or bacteria. The mechanisms by which CMV replication inhibits host responses are complex and involve virus interactions with both host lymphocytes and monocyte-macrophages.

CMV Mononucleosis

Clinical features CMV mononucleosis occurs in patients of any age but is most common in sexually active young adults. A vigorous host T cell response may contribute to the syndrome, which is characterized by fever, malaise, fatigue, and myalgia. Headache and splenomegaly are also often present. Mild liver enzyme abnormalities are common, and atypical lymphocytes are present in the peripheral blood. Heterophile antibodies are not formed in response to CMV infection; however, mild immunologic abnormalities, including the presence of rheumatoid factor and antinuclear antibodies, are common.

Complications Occasionally, hemolytic anemia, thrombocytopenia, neurologic complications (e.g., Guillain-Barré syndrome), or other organ involvement (liver, lung) occurs and is severe.[43] Acute infections generally resolve in 2 to 4 weeks, but postviral asthenia and viral excretion often persist for months.

CMV Infection in the Immunocompromised Host

CMV appears to be the most frequent and important viral pathogen in patients who have had organ transplants. Most commonly, such patients with CMV syndromes present with fever and leukopenia, which may progress to pneumonitis or, in rare instances, to disseminated disease. The period of highest risk is 1 to 4 months after transplantation and appears to relate to the degree of host immunosuppression. CMV may also increase the risk of graft loss after organ transplantation.[44] CMV pneumonia occurs in nearly 20% of marrow transplant recipients; mortality in such patients, if they remain untreated, is about 90%.

CMV is recognized as an important pathogen in patients with AIDS. The virus often contributes to the immunosuppression observed in such patients and may cause disseminated disease affecting the eyes, GI tract, or CNS. At least 50% of patients with AIDS have CMV viremia, and 90% or more have evidence of CMV infection at autopsy.[45] Ulcers of the esophagus, stomach, small intestine, or colon may be present and may lead to bleeding or perforation.[46] It is critical to recognize CMV polyradiculopathy or encephalitis in this setting because this disorder is potentially reversible with therapy.[47,48] Similarly, CMV retinitis is an important treatable cause of blindness in patients with AIDS.[49]

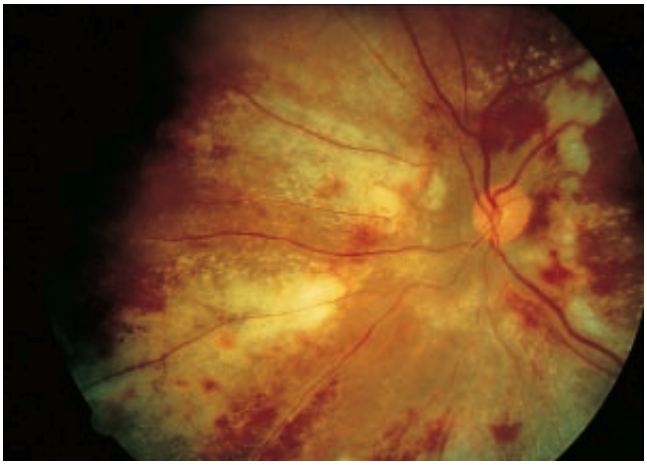


Figure 6 **Cytomegalovirus retinitis may result in hemorrhage exudates and, eventually, blindness.**

Early lesions consist of small white areas of retinal necrosis that progress in a centrifugal manner and are subsequently accompanied by hemorrhage, vessel sheathing, and retinal edema [*see Figure 6*]. CMV and HIV have been shown to potentiate each other's replication in vitro, and it is possible that such interactions occur in dually infected organs (e.g., the retina and brain of an AIDS patient who has CMV infection).

Two forms of CMV encephalitis are clinically and pathologically distinguishable in patients with AIDS: multifocal micronodular encephalitis, which resembles HIV encephalitis and presents with progressive dementia, and ventriculoencephalitis, which is characterized by acute cranial nerve deficits, ventriculomegaly, nystagmus, lethargy, and disorientation.[48,50]

DIAGNOSIS

CMV infection cannot be reliably diagnosed on clinical grounds alone. Virus can be isolated from a variety of secretions or infected tissues, but definitive identification can take several weeks. Use of shell vial centrifugation and immunocytochemical analysis for CMV early antigens can facilitate rapid diagnosis and help monitor outcome, as can the application of immunofluorescence or PCR-based assays to plasma or peripheral blood leukocytes.[51-53] Demonstration of viremia is a better indicator of acute infection than the detection of virus in urine or saliva. The use of PCR to detect CMV DNA in CSF is helpful in the diagnosis of CMV encephalitis and polyradiculopathy.[48,50]

PREVENTION

Several specific and nonspecific prophylactic measures to prevent CMV syndromes are under investigation. Matching of seronegative transplant or transfusion recipients with seronegative organ or blood donors may reduce the likelihood of transmission. Use of blood that has been frozen and thawed, washed with saline, or filtered through cotton wool can prevent transfusion-associated CMV infection. CMV immune globulin may be useful in certain seronegative bone marrow transplant recipients. Experimental CMV vaccines are under investigation.

TREATMENT

Management of CMV Mononucleosis

CMV mononucleosis is generally self-limited and requires only supportive care. Specific antiviral therapy is not indicated.

Management of CMV in the Immunocompromised Host

Intravenous ganciclovir, an analogue of deoxyguanosine, and oral valganciclovir, a prodrug for ganciclovir, are effective treatments for CMV infections (e.g., retinitis) in immunocompromised hosts [*see Table 4*]. Both ganciclovir administered intravenously (5 mg/kg twice daily for 21 days followed by 5 mg/kg daily maintenance) and oral valganciclovir (900 mg twice daily for 21 days followed by 900 mg daily maintenance) result in stabilization or improvement in 70% to 90% of patients with CMV retinitis.[54] If the immunocompromised state persists, prolonged maintenance therapy may be required to maintain improvement of retinitis.[55] Maintenance therapy can be safely discontinued in patients with AIDS whose CD4 cell counts increase substantially with antiretroviral therapy, but relapses occur if CD4 cell counts fall below 50/mm^3.[56]

Foscarnet (90 mg/kg twice daily for 2 to 3 weeks, followed by 90 to 120 mg/kg daily) is effective in treatment of CMV retinitis and GI disease, and cidofovir (5 mg/kg weekly for 2 weeks, fol-

Table 4 Drug Treatment for Cytomegalovirus Infection

Drug	*Dose*	*Efficacy Rating*	*Comment*
Ganciclovir	5 mg/kg I.V. b.i.d. for 21 days, then 5 mg/kg I.V. q.d. for maintenance	First choice	For CMV retinitis and other advanced CMV infections in immunocompromised patients
Ganciclovir + ocular implants	1.5 g p.o., t.i.d.	—	For long-term therapy of CMV retinitis in patients with HIV infection
Valganciclovir	900 mg p.o., b.i.d., for 21 days, then 900 mg p.o., q.d., for maintenance	First choice	For CMV retinitis and other advanced CMV infections in immunocompromised patients
Foscarnet	90 mg/kg I.V. b.i.d. for 2-3 wk, then 90–120 mg/kg I.V. q.d.	Alternative	For CMV retinitis or GI disease; combine with ganciclovir for severe CNS infections in patients with AIDS
Cidofovir	5 mg/kg I.V. once a week for 2 wk, then 5 mg/kg I.V. every 2 wk	Alternative	For CMV retinitis

CMV—cytomegalovirus CNS—central nervous system GI—gastrointestinal NA—not available

lowed by 5 mg/kg every 2 weeks) may be useful in CMV retinitis. Ganciclovir ocular implants in combination with oral ganciclovir (1.5 g three times daily) is effective for long-term treatment of CMV retinitis in persistently immunocompromised patients with HIV infection.[57] The combination of ganciclovir and foscarnet may be necessary for severe CNS infections in patients with AIDS.[58] Resistance to ganciclovir and other anti-CMV drugs may complicate prolonged therapy.[59,60]

Prophylactic or suppressive ganciclovir may be useful in other high-risk patients, such as CMV-seropositive bone marrow or solid-organ transplant recipients. Several preventive or preemptive regimens of ganciclovir or valganciclovir are under study in both susceptible transplant recipients and patients with AIDS.[61-63]

Epstein-Barr Virus

EBV has a very limited cell tropism; its host range is generally restricted to human B cells and epithelial cells of the nasopharynx and uterine cervix. It is the cause of heterophile antibody–positive infectious mononucleosis and has been implicated in a variety of other disorders, including Burkitt lymphoma, nasopharyngeal carcinoma, oral hairy leukoplakia, and a variety of B cell and possibly T cell lymphoproliferative disorders.[64-66]

EPIDEMIOLOGY AND ETIOLOGY

In developing countries, most of the population is exposed to EBV at an early age. In the United States, approximately 50% of the population seroconvert by 5 years of age; there is also a high rate of seroconversion in adolescence and young adulthood. Clinical infectious mononucleosis develops primarily in persons who contract primary EBV infection when they are between 10 and 20 years of age; in younger persons, primary EBV infection is usually clinically inapparent. Thus, most cases of infectious mononucleosis occur in members of higher socioeconomic groups in developed countries, where viral transmission is often delayed until the patients are older than 10 years. Approximately 12% of susceptible college-age young adults seroconvert each year; about 50% of them acquire infectious mononucleosis. As a result, EBV infection has important public health implications for young adult populations, including students and military personnel.

EBV is shed from the oropharynx in approximately 15% of the adult seropositive population at any given time, and all EBV-seropositive individuals may shed virus in their saliva at some time.[67] Shedding occurs much more frequently in patients with infectious mononucleosis and in immunocompromised persons, such as renal transplant recipients and patients with AIDS. Transmission requires close contact, usually oral to oral, with a person shedding EBV. On rare occasions, virus can be transmitted by other routes, including blood transfusion. EBV infection of the uterine cervix has been documented, genital ulcerations have been observed in patients with infectious mononucleosis, and sexual activity is a highly significant risk factor for EBV seropositivity,[68] suggesting the possibility of genital transmission of EBV to sexual partners and newborns.

PATHOGENESIS

Infection is initiated by the binding of EBV envelope proteins gp350/120 to the cell surface molecule CD21. After initial replication in nasopharyngeal cells or B cells, B cells apparently carry the virus to other parts of the body. In the nasopharynx, infectious virus is actively replicated. In contrast, in B cells, the viral genome is present, but mature infectious virus particles are not produced. In B cells in which EBV remains latent, its DNA exists as a circular episome; in cells in which infectious virus is produced, EBV DNA exists in a linear form. Infected B cells are referred to as immortalized (i.e., capable of continuous proliferation) and may produce a variety of antibodies as a consequence of polyclonal B cell activation. One of the antibodies they produce, the heterophile antibody, is a useful diagnostic marker of infection. Once infection is initiated, the host mounts a T cell response against new antigens on infected cells and a B cell antibody response against several EBV-associated antigens, including viral capsid antigen, early antigen, and Epstein-Barr nuclear antigen.[64] Antibody patterns vary according to the stage of infection and the syndrome expressed (see below).

CLINICAL SYNDROMES

Age and immunocompetence greatly influence the expression of EBV infection. Young children usually have asymptomatic or trivial infections and often do not produce heterophile antibodies. In the elderly, EBV infection may present as a persistent febrile syndrome in which the patient tests negative for heterophile antibodies. The most common clinical manifestation of primary EBV infection is infectious mononucleosis.

Infectious Mononucleosis

Clinical features Patients usually present with fever, pharyngitis, and lymphadenopathy. Hepatosplenomegaly, a palatal enanthema, periorbital edema, and jaundice are less common features. A maculopapular diffuse rash occurs in 10% of pa-

tients, particularly in patients who have been given ampicillin. Hematologic abnormalities include a peripheral blood lymphocytosis; usually more than 10% of the leukocytes in the peripheral blood consist of atypical lymphocytes. The total leukocyte count may be normal, low, or high; a relative and absolute neutropenia is observed in 60% to 90% of patients. Thrombocytopenia is also common, and hepatocellular enzymes are abnormal in about 90% of patients. Cryoproteins and antigranulocyte antibodies are frequently present but are usually of little clinical significance. Most cases of infectious mononucleosis resolve in 1 to 3 weeks, although malaise and fatigue occasionally persist for several weeks to months.

Complications Complications of infectious mononucleosis can affect most organ systems. Tonsillar enlargement can be extreme, causing respiratory embarrassment. Splenic rupture is rare but must be considered when abdominal pain develops, particularly during the second and third weeks of illness. Patients with splenic rupture usually have left upper quadrant pain and tenderness, often accompanied by signs of peritoneal irritation and laboratory evidence of a falling hematocrit. Splenectomy may be required when a diagnosis of splenic rupture is confirmed, although splenic preservation is sometimes possible with intensive supportive care.

Neurologic complications occur in fewer than 1% of patients with infectious mononucleosis; these complications include encephalitis, aseptic meningitis, transverse myelitis, Guillain-Barré syndrome, optic neuritis, and peripheral neuropathies. EBV encephalitis often presents as a cerebellitis, but it may mimic the temporal lobe presentation of herpes encephalitis [*see 11:XVI Acute Viral Central Nervous System Diseases*]. Despite the severity of the neurologic complications, most patients with such complications recover completely.

Autoimmune hemolytic anemia mediated by cold agglutinins directed against antigens may occur, usually during the second or third week of illness. Aplastic anemia, thrombocytopenia with bleeding, and severe granulocytopenia with superinfection are rare but potentially fatal complications. EBV has been detected in bone marrow specimens by in situ hybridization. Cardiac and pulmonary complications are rare. A rare chronic syndrome of fever, persistent hepatitis, extensive lymphadenopathy, hepatosplenomegaly, pancytopenia, uveitis, and interstitial pneumonia has been associated with persistent EBV infection.[69]

EBV Infection in the Immunocompromised Host

Immunocompromised hosts are susceptible to overwhelming EBV lymphoproliferative syndromes. A familial X-linked disorder, Duncan disease, has been associated with fatal infectious mononucleosis, as well as agammaglobulinemia and lymphoma. All of these disorders are apparently related to EBV infection in genetically predisposed persons. The gene involved encodes for a 128 amino acid protein that may be involved in T cell signal transduction.[70] Occasionally, in transplant recipients receiving cyclosporine and in patients with AIDS, overwhelming EBV infections develop that culminate in B cell lymphoproliferative syndromes.

Oral hairy leukoplakia is a nonmalignant hyperplastic lesion of epithelial cells that occurs in immunocompromised patients, particularly those with HIV infection, and presents as raised, white, corrugated lesions on the lateral aspect of the tongue. It is associated with active EBV replication and expression of lytic viral proteins.[71]

EBV Infections and Cancer

The nature of the association of EBV with Burkitt lymphoma and nasopharyngeal carcinoma remains unresolved. Nearly all African patients with Burkitt lymphoma and East Asian patients with nasopharyngeal carcinoma have elevated EBV antibody titers, and EBV DNA is found in the tumors themselves. A single clonal form of EBV has been associated with nasopharyngeal carcinoma and with preinvasive lesions, although it is not clear whether or how the virus is involved in the etiology of these neoplasms. EBV has also been associated with some cases of primary CNS B cell lymphoma, T cell lymphoma, smooth muscle tumors, lymphomatoid granulomatosis, and Hodgkin disease.[64-66]

DIAGNOSIS

In most cases of infectious mononucleosis, the diagnosis is confirmed by detection of heterophile antibodies and abnormalities in the blood, characterized by the presence of atypical lymphocytes that frequently account for more than 10% of the leukocytes in the peripheral blood. Heterophile antibodies, often detected with monospot or slide tests, are not directed against viral antigens but rather against erythrocyte antigens of a variety of species, including sheep, horses, and goats. These acute-phase reactants, present in about 90% of patients at some point during the illness, must be distinguished from naturally occurring Forssman antibodies and the antibodies present in patients with serum sickness; differentiation can be accomplished by proper absorption tests. Heterophile antibodies may be present at the onset of clinical illness or may appear later in the course of disease; they usually disappear 3 to 4 months after the onset of illness but may persist for longer periods. False positive reactions are rare.

Heterophile antibody-negative infectious mononucleosis may be caused by CMV, EBV, or other viruses. For heterophile antibody–negative cases of infectious mononucleosis and for other possibly EBV-associated syndromes, measurement of EBV-specific antibodies may be useful. Antibodies to viral capsid antigen appear early in the course of infection and are present in 85% of patients at the time of their initial physician visit. The presence of IgM antibodies to viral capsid antigen suggests acute infection, whereas IgG antibodies to viral capsid antigen indicate past infection. Antibodies to early antigen appear transiently during active infections and are also found in patients with the African variant of Burkitt lymphoma and in patients with nasopharyngeal carcinoma. Antibodies to Epstein-Barr nuclear antigen appear late in the course of illness and persist after recovery. Thus, the pattern of specific EBV antibody responses may indicate the stage of infection.

Cultures are usually not helpful in making a diagnosis. Although EBV can be isolated from the nasopharynx, culture techniques are cumbersome and time consuming. EBV DNA detection by PCR techniques may be useful for monitoring progression in certain diseases, such as encephalitis or lymphoproliferative syndromes.[72-74]

PREVENTION

Experimental EBV vaccines are under study, although appropriate indications for their use are unclear.[64]

TREATMENT

Management of Infectious Mononucleosis

Treatment is supportive. The administration of corticosteroids may be helpful in certain situations, such as impending airway obstruction, severe thrombocytopenia, or hemolytic anemia.

Short-course regimens (1 to 2 weeks) are adequate. Specific antiviral agents such as acyclovir or ganciclovir are not recommended for most EBV-related disorders.[75] Although intravenous acyclovir can reduce viral shedding in patients with infectious mononucleosis, clinical benefit appears to be minimal.

Management of Infection in the Immunocompromised Host

Infusion of donor leukocytes or EBV-specific cytotoxic T cells has also been useful in some EBV-related lymphoproliferative diseases after bone marrow transplantation.[64,76] No antiviral agents have shown benefit in EBV-associated diseases.

Human Herpesvirus Type 6

In 1986, a newly recognized herpesvirus, HHV-6, was isolated from the peripheral blood leukocytes of six persons with various lymphoproliferative disorders.[77] It was initially designated human B lymphotropic virus but was renamed when subsequent studies demonstrated that the virus replicated in cells of T cell origin.[78] Two genetically distinct variants of HHV-6—namely, HHV-6A and HHV-6B—are recognized. Disease is most commonly associated with HHV-6B infection; however, HHV-6A may have greater neurotropism.[79] An immunomodulatory protein, CD46, is a cellular receptor for HHV-6.[80]

HHV-6 infection typically occurs during infancy as the level of maternal antibody wanes (peak between 6 to 9 months), although intrauterine infection can also occur.[81] As many as 80% of adults are seropositive for HHV-6. Infection can present as exanthema subitum (roseola infantum), a common illness characterized by fever followed by a rash or by a febrile illness without rash but often with febrile seizures.[82] HHV-6 has been associated with encephalitis, and it has been suggested, though not proved, that HHV-6 is a causal agent in multiple sclerosis.[83] HHV-6 may be transmitted by saliva and possibly by genital secretions, blood, or transplanted organs, and HHV-6 infection may cause a mononucleosis-like syndrome.[84] After primary infection, the viral genome may persist in peripheral blood mononuclear cells and in salivary glands. Reactivation accompanies immunosuppression, and 30% to 45% of bone marrow transplant recipients develop HHV-6 viremia in the first few weeks after grafting.[85] In immunocompromised adults, HHV-6 may be associated with a rash resembling graft versus host disease and may cause pneumonitis, encephalitis, graft rejection, or disseminated disease, although its etiologic relationship to any of these remains unproven.[85-88] There is no antiviral agent with proven clinical efficacy against HHV-6, although ganciclovir and foscarnet have been suggested as possible therapies.[89]

Human Herpesvirus Type 7

In 1990, another lymphotropic human herpesvirus, tentatively called HHV-7, was isolated from human T cells obtained from a healthy 26-year-old man.[90] Its role in human disease has not been established, although it appears frequently in the saliva of healthy adults.[91] Infection generally occurs by age 5, and approximately 90% of adults are seropositive.[92] The HHV-7 genome can also be detected in peripheral blood leukocytes and cervical secretions. HHV-7 may account for some cases of roseola or febrile seizures[93,94] and may be associated with graft dysfunction after solid-organ transplantation.[88] Although it has been suggested that both HHV-6 and HHV-7 are etiologic agents in chronic fatigue syndrome and pityriasis rosea, supportive evidence is lacking.[95,96]

Human Herpesvirus Type 8

Several reports in late 1994 and 1995 provided convincing evidence of unique herpesvirus-like sequences in Kaposi sarcoma (KS) and body cavity–based lymphoma tissue from patients with AIDS.[97-100] Using representational difference analysis, researchers found that more than 90% of the KS tissue studied contained these sequences; appropriate control tissue tested negative, except for 15% of non-KS tissue from AIDS patients.[97] The same herpesvirus-like DNA sequences were subsequently reported in KS tissue from patients who did not have AIDS[98-100] and in lymph node biopsy specimens from patients with multicentric Castleman disease.[101]

HHV-8 has been propagated in cell cultures and closely resembles a recently identified gamma herpesvirus of rhesus monkeys. Several serologic assays have been developed.[102-104] One, which detects antibody to a latency-associated nuclear antigen, finds seropositivity in 1% to 2% of HIV-negative blood donors; in 30% of HIV-positive hemophiliacs, transfusion recipients, and women; and in more than 80% of patients with HIV infection and KS.

The natural history and clinical spectrum of diseases caused by HHV-8 is not yet well characterized. Primary infection in immunocompetent children may be associated with fever and a maculopapular rash, and a finding of HHV-8 DNA in saliva suggests salivary transmission.[105] One individual developed fever, hepatosplenomegaly, angiolymphoid hyperplasia, and transient KS during seroconversion.[106] Family studies suggest transmission to children via nonsexual routes and between spouses via sexual routes.[107] Transmission has also been demonstrated in the intrapartum and postpartum periods, and intrauterine infection may also occur.[108] Among men who have sex with men, risk factors for transmission include increased number of sex partners, amyl nitrite use, and occurrence of lymphadenopathy within the previous 6 months.[109] Risk factors for transmission in North American women include HIV infection, increasing age, history of syphilis, injection drug use, and black race.[110,111]

Molecular and serologic evidence strongly suggests a causal role for HHV-8 in KS and associations with body cavity–based lymphomas and multicentric Castleman disease. Although it has been suggested that HHV-8 is associated with both multiple myeloma and sarcoidosis, confirmatory evidence is lacking. Studies of HHV-8 susceptibility to antiviral drugs suggest relative resistance to acyclovir and penciclovir but susceptibility to ganciclovir, foscarnet, and cidofovir.[112,113] However, one pilot trial of cidofovir showed no effect on either HHV-8 viral load or on KS lesions.[114] In contrast, effective antiretroviral infection in individuals dually infected with HIV-1 and HHV-8 may reduce the HHV-8 load and may be associated with KS regression.[115]

The author has received grants for clinical research from Takeda Pharmaceuticals North America, Inc., and has served as a consultant to GlaxoSmithKline, Bristol-Myers Squibb Co., and Schering-Plough Corp.

References

1. Whitley RJ, Roizman B: Herpes simplex virus infections. Lancet 357:1513, 2001

2. Xu F, Schillinger JA, Sternberg MR, et al: Seroprevalence and coinfection with herpes simplex virus type 1 and type 2 in the United States, 1988-1994. J Infect Dis 185:1019, 2002

3. Gottlieb SL, Douglas JM Jr, Schmid DS, et al: Seroprevalence and correlates of herpes simplex virus type 2 infection in five sexually transmitted-disease clinics. J Infect Dis 186:1381, 2002

4. Langenberg AG, Corey L, Ashley RL, et al: A prospective study of new infections with herpes simplex virus type 1 and type 2. N Engl J Med 341:1432, 1999

5. Wald A, Zeh J, Selke S, et al: Reactivation of genital herpes simplex virus type 2 infection in asymptomatic seropositive persons. N Engl J Med 342:844, 1999

6. Benedetti JK, Zeh J, Corey L: Clinical reactivation of genital herpes simplex virus infection decreases in frequency over time. Ann Intern Med 131:14, 2000

7. Sucato G, Wald A, Wakabayashi E, et al: Evidence of latency and reactivation of both herpes simplex virus (HSV)-1 and HSV-2 in the genital region. J Infect Dis 177:1069, 1998

8. Engelberg R, Carrell D, Krantz E, et al: Natural history of genital herpes simplex virus type 1 infection. Sex Transm Dis 30:174, 2003

9. Wald A, Corey L, Cone R, et al: Frequent genital herpes simplex virus 2 shedding in immunocompetent women—effect of acyclovir treatment. J Clin Invest 99:1092, 1997

10. Tedder DG, Ashley R, Tyler KL, et al: Herpes simplex virus infection as a cause of benign recurrent lymphocytic meningitis. Ann Intern Med 121:334, 1994

11. Murakami S, Mizobuchi M, Nakashiro Y, et al: Bell palsy and herpes simplex virus: identification of viral DNA in endoneurial fluid and muscle. Ann Intern Med 124:27, 1996

12. Yamamoto LJ, Tedder DG, Ashley R, et al: Herpes simplex virus type 1 DNA in cerebrospinal fluid of a patient with Mollaret's meningitis. N Engl J Med 325:1082, 1991

13. Prober CG, Corey L, Brown ZA, et al: The management of pregnancies complicated by genital infections with herpes simplex virus. Clin Infect Dis 15:1031, 1992

14. Boggess KA, Watts DH, Hobson AC, et al: Herpes simplex virus type 2 detection by culture and polymerase chain reaction and relationship to genital symptoms and cervical antibody status during the third trimester of pregnancy. Am J Obstet Gynecol 176:443, 1997

15. Wald A, Ashley-Morrow R: Serological testing for herpes simplex virus (HSV)-1 and HSV-2 infection. Clin Infect Dis 35(suppl 2): S173, 2002

16. Wildemann B, Erhart K, Storch-Hagenlocher B, et al: Quantification of herpes simplex virus type 1 in DNA of cerebrospinal fluid of patients with herpes simplex encephalitis. Neurology 48:1341, 1997

17. Stanberry LR, Spruance SL, Cunningham AL, et al: Glycoprotein-D-adjuvant vaccine to prevent genital herpes. N Engl J Med 347:1652, 2002

18. Management of herpes in pregnancy. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin 8, 1999, Washington, DC

19. Watts DH, Brown ZA, Money D, et al: A double-blind, randomized, placebo-controlled trial of acyclovir in late pregnancy for the reduction of herpes simplex virus shedding and cesarean delivery. Am J Obstet Gynecol 188:836, 2003

20. Scott LL, Hollier LM, McIntire D, et al: Acyclovir suppression to prevent recurrent genital herpes at delivery. Infect Dis Obstet Gynecol 10:71, 2002

21. Spruance SL, Jones TM, Blatter MM, et al: High-dose, short-duration, early valacyclovir therapy for episodic treatment of cold sores: results of two randomized, placebo-controlled, multicenter studies. Antimicrob Agents Chemother 47:1072, 2003

22. Wald A, Carrell D, Remington M, et al: Two-day regimen of acyclovir for treatment of recurrent genital herpes simplex virus type 2 infection. Clin Infect Dis 34:944, 2002

23. Corey L, Wald A, Patel R, et al: Once-daily valacyclovir to reduce the risk of transmission of genital herpes. N Engl J Med 350:11, 2004

24. Bacon TH, Levin MJ, Leary JJ, et al: Herpes simplex virus resistance to acyclovir and penciclovir after two decades of antiviral therapy. Clin Microbiol Rev 16:114, 2003

25. Javaly K, Wohlfeiler M, Kalayjian R, et al: Treatment of mucocutaneous herpes simplex virus infections unresponsive to acyclovir with topical foscarnet cream in AIDS patients: a phase I/II study. J Acquir Immune Defic Syndr 21:301, 1999

26. Lalezari J, Schacker T, Feinberg J, et al: A randomized, double-blind, placebo-controlled trial of cidofovir gel for the treatment of acyclovir-unresponsive mucocutaneous herpes simplex virus infections in patients with AIDS. J Infect Dis 176:892, 1997

27. Kessler HA, Hurwitz S, Farthing C, et al: Pilot study of topical trifluridine for the treatment of acyclovir-resistant mucocutaneous herpes simplex disease in patients with AIDS (ACTG 172). AIDS Clinical Trials Group. J Acquir Immune Defic Syndr Hum Retrovirol 12:147, 1996

28. Gnann JW Jr, Whitley RJ: Herpes zoster. N Engl J Med 347:340, 2002

29. Gourishankar S, McDermid JC, Jhangri GS, et al: Herpes zoster infection following solid organ transplantation: incidence, risk factors and outcomes in the current immunosuppressive era. Am J Transplant 4:108, 2004

30. Harger JH, Ernest JM, Thurnau GR, et al: Risk factors and outcome of varicella-zoster virus pneumonia in pregnant women. J Infect Dis 185:422, 2002

31. Enders G, Miller E, Cradock-Watson J, et al: Consequences of varicella and herpes zoster in pregnancy: prospective study of 1739 cases. Lancet 334:1548, 1994

32. Cohen JI, Brunnell PA, Straus SE, et al: Recent advances in varicella-zoster virus infection. Ann Intern Med 130:922, 1999

33. Gilden DH, Kleinschmidt-DeMasters BK, LaGuardia JJ, et al: Neurologic complications of the reactivation of varicella-zoster virus. N Engl J Med 342:636, 2000

34. Seward JF, Watson BM, Peterson CL, et al: Varicella disease after introduction of varicella vaccine in the United States, 1995-2000. JAMA 287:606, 2002

35. Vazquez M, LaRussa PS, Gershon AA, et al: The effectiveness of the varicella vaccine in clinical practice. N Engl J Med 344:955, 2001

36. Galil K, Lee B, Strine T, et al: Outbreak of varicella at a daycare center despite vaccination. N Engl J Med 347:1909, 2002

37. Wood MJ, Johnson RW, McKendrick MW, et al: A randomized trial of acyclovir for 7 days or 21 days with and without prednisolone for treatment of acute herpes zoster. N Engl J Med 330:896, 1994

38. Whitley RJ, Weiss H, Gnann JW Jr, et al: Acyclovir with and without prednisone for the treatment of herpes zoster—a randomized placebo-controlled trial. Ann Intern Med 125:376, 1996

39. Breton G, Fillet AM, Katlama C, et al: Acyclovir-resistant herpes zoster in human immunodeficiency virus–infected patients: results of foscarnet therapy. Clin Infect Dis 27:1525, 1998

40. Fowler KB, Stagno S, Pass RF: Maternal immunity and prevention of congenital cytomegalovirus infection. JAMA 289:1008, 2003

41. Boppana SR, Rivera LB, Fowler KB, et al: Intrauterine transmission of cytomegalovirus to infants of women with preconceptional immunity. N Engl J Med 344:1366, 2001

42. Singh N, Wannstedt C, Keyes L, et al: Impact of evolving trends in recipient and donor characteristics on cytomegalovirus infection in liver transplant recipients. Transplantation 77:106, 2004

43. Eddleston M, Peacock S, Juniper M, et al: Severe cytomegalovirus infection in immunocompetent patients. Clin Infect Dis 24:52, 1997

44. de Otoro J, Gavalda J, Murio E, et al: Cytomegalovirus disease as a risk factor for graft loss and death after orthotopic liver transplantation. Clin Infect Dis 26:865, 1998

45. Jacobson MA, Mills J: Serious cytomegalovirus disease in the acquired immunodeficiency syndrome (AIDS): clinical findings, diagnosis, and treatment. Ann Intern Med 108:585, 1988

46. Goodgame RW: Gastrointestinal cytomegalovirus disease. Ann Intern Med 119:924, 1993

47. Kim YS, Hollander H: Polyradiculopathy due to cytomegalovirus: report of two cases in which improvement occurred after prolonged therapy and review of the literature. Clin Infect Dis 17:32, 1993

48. Arribas JR, Storch GA, Clifford DB, et al: Cytomegalovirus encephalitis. Ann Intern Med 125:577, 1996

49. Jacobson M: Treatment of cytomegalovirus retinitis in patients with the acquired immunodeficiency syndrome. N Engl J Med 337:105, 1997

50. McCutchan JA: Cytomegalovirus infections of the nervous system in patients with AIDS. Clin Infect Dis 20:747, 1995

51. Humar A, Wood S, Lipton J, et al: Effect of cytomegalovirus infection on 1-year mortality rates among recipients of allogeneic bone marrow transplants. Clin Infect Dis 26:606, 1998

52. Nokta MA, Holland F, DeGruttola V, et al: Cytomegalovirus (CMV) polymerase chain reaction profiles in individuals with advanced human immunodeficiency virus infection: relationship to CMV disease. J Infect Dis 185:1717, 2002

53. Bowen EF, Emery VC, Wilson P, et al: Cytomegalovirus polymerase chain reaction viraemia in patients receiving ganciclovir maintenance therapy for retinitis. AIDS 12:605, 1998

54. Martin DF, Sierra-Madero J, Walmsley S, et al: A controlled trial of valganciclovir as induction therapy for cytomegalovirus retinitis. N Engl J Med 346:1119, 2002

55. Lalezari J, Lindley J, Walmsley S, et al: A safety study of oral valganciclovir maintenance treatment of cytomegalovirus retinitis. J Acquir Immune Defic Syndr 30:392, 2002

56. Torriani FJ, Freeman WR, Macdonald JC, et al: CMV retinitis recurs after stopping treatment in virological and immunological failures of potent antiretroviral therapy. AIDS 14:173, 2000

57. Martin DF, Kuppermann BD, Wolitz RA, et al: Oral ganciclovir for patients with cytomegalovirus retinitis treated with a ganciclovir implant. Roche Ganciclovir Study Group. N Engl J Med 340:1063, 1999

58. Anduze-Faris BM, Fillet AM, Gozlan J, et al: Induction and maintenance therapy of cytomegalovirus central nervous system infection in HIV-infected patients. AIDS 14:517, 2000

59. Lurain NS, Bhorade SM, Pursell KJ, et al: Analysis and characterization of antiviral drug-resistant cytomegalovirus isolates from solid organ transplant recipients. J Infect Dis 186:760, 2002

60. Chou S, Lurain NS, Thompson KD, et al: Viral DNA polymerase mutations associated with drug resistance in human cytomegalovirus. J Infect Dis 188:32, 2003

61. Zaia JA: Prevention and management of CMV-related problems after hematopoietic stem cell transplantation. Bone Marrow Transplant 29:633, 2002

62. Paya CV: Prevention of cytomegalovirus disease in recipients of solid-organ transplants. Clin Infect Dis 32:596:2001

63. Singh N: Preemptive therapy versus universal prophylaxis with ganciclovir for cytomegalovirus in solid organ transplant recipients. Clin Infect Dis 32:742, 2001

64. Cohen JI: Epstein-Barr virus infection. N Engl J Med 343:481, 2000

65. Sakajiri S, Mori K, Isobe Y, et al: Epstein-Barr virus-associated T-cell acute lymphoblastic leukemia. Br J Haematol 117:127, 2002

66. Timms JM, Bell A, Flavell JR, et al: Target cells of Epstein-Barr virus (EBV)-positive post-transplant lymphoproliferative disease: similarities to EBV-positive Hodgkin's lymphoma. Lancet 361:217, 2003

67. Ling PD, Lednicky JA, Keitel WA, et al: The dynamics of herpesvirus and polyomavirus reactivation and shedding in healthy adults: a 14-month longitudinal study. J Infect Dis 187:1571, 2003

68. Crawford DH, Swerdlow AJ, Higgins C, et al: Sexual history and Epstein-Barr virus infection. J Infect Dis 186:731, 2002

69. Kimura H, Morishima T, Kanegane H, et al: Prognostic factors for chronic active Epstein-Barr virus infection. J Infect Dis 187:527, 2003

70. Sayos S, Wu C, Morra M, et al: The X-induced lymphoproliferative–disease gene product SAP regulates signals induced through the co-receptor SLAM. Nature 395:462, 1998

71. Triantos D, Porter SR, Scully C, et al: Oral hairy leukoplakia: clinicopathologic features, pathogenesis, diagnosis, and clinical significance. Clin Infect Dis 25:1392, 1997

72. Weinberg A, Li S, Palmer M, et al: Quantitative CSF PCR in Epstein-Barr virus infections of the central nervous system. Ann Neurol 52:543, 2002

73. Fan H, Gulley ML: Epstein-Barr viral load measurement as a marker of EBV-related disease. Mol Diagn 6:279, 2001

74. Stevens SJ, Verschuuren EA, Verkuujlen SA, et al: Role of Epstein-Barr virus DNA load monitoring in prevention: early detection of post-transplant lymphoproliferative disease. Leuk Lymphoma 43:831, 2002

75. Torre D, Tambini R: Acyclovir for treatment of infectious mononucleosis: a meta-analysis. Scand J Infect Dis 31:543, 1999

76. Haque T, Wilkie GM, Taylor C, et al: Treatment of Epstein-Barr-virus-positive post-transplantation lymphoproliferative disease with partly HLA-matched allogeneic cytotoxic T cells. Lancet 360:436, 2002

77. Salahuddin SZ, Ablashi DV, Markham PD, et al: Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. Science 234:596, 1986

78. Lopez C, Pellett P, Stewart J, et al: Characteristics of human herpesvirus–6. J Infect Dis 157:1271, 1988

79. Hall CB, Caserta MT, Schnabel KC, et al: Persistence of human herpesvirus 6 according to site and variant: possible greater neurotropism of variant A. Clin Infect Dis 26:132, 1998

80. Santoro F, Kennedy PE, Locatelli G, et al: CD46 is a cellular receptor for human herpesvirus 6. Cell 99:817, 1999

81. Ashshi AM, Cooper RJ, Klapper PE, et al: Detection of human herpesvirus 6 DNA in fetal hydrops. Lancet 335:1519, 2000

82. Caserta MT, Mock DJ, Dewhurst S: Human herpesvirus 6. Clin Infect Dis 33:829, 2001

83. Tyler KL: Human herpesvirus 6 and multiple sclerosis: the continuing conundrum. J Infect Dis 187:1360, 2003

84. Akashi K, Eizuru Y, Sumiyoshi Y, et al: Severe infectious mononucleosis–like syndrome and primary human herpesvirus 6 infection in an adult. N Engl J Med 329:168, 1993

85. Yoshikawa T, Asano Y, Ihira M, et al: Human herpesvirus 6 viremia in bone marrow transplant recipients: clinical features and risk factors. J Infect Dis 185:847, 2002

86. Zerr DM, Gooley TA, Yeung L, et al: Human herpesvirus 6 reactivation and encephalitis in allogeneic bone marrow transplant recipients. Clin Infect Dis 33:763, 2001

87. Amo K, Tanaka-Taya K, Inagi R, et al: Human herpesvirus 6B infection of the large intestine of patients with diarrhea. Clin Infect Dis 36:120, 2003

88. Emery VC: Human herpesviruses 6 and 7 in solid organ transplant recipients. Clin Infect Dis 32:1357, 2001

89. Zerr DM, Gupta D, Huang ML, et al: Effect of antivirals on human herpesvirus 6 replication in hematopoietic stem cell transplant recipients. Clin Infect Dis 34:309, 2002

90. Frenkel N, Schirmer EC, Wyatt LS, et al: Isolation of a new herpesvirus from human CD4+ T cells. Proc Natl Acad Sci USA 87:748, 1990

91. Wyatt LS, Frenkel N: Human herpesvirus 7 is a constitutive inhabitant of adult human saliva. J Virol 66:3206, 1992

92. Black JB, Pellett PE: Human herpesvirus 7. Rev Med Virol 9:245, 1999

93. Levy JA: Three new human herpes viruses (HHV 6, 7, and 8). Lancet 349:558, 1997

94. Hall CB: Human herpesviruses at sixes, sevens, and more. Ann Intern Med 127:481, 1997

95. Reeves WC, Stamey FR, Black JB, et al: Human herpesviruses 6 and 7 in chronic fatigue syndrome: a case-control study. Clin Infect Dis 31:48, 2000

96. Watanabe T, Kawamura T, Jacob SE, et al: Pityriasis rosea is associated with systemic active infection with human herpesvirus-7 and human herpesvirus-6. J Invest Dermatol 119:793, 2002

97. Chang Y, Cesarman E, Pessin MS, et al: Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266:1865, 1994

98. Moore PS, Chang Y: Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. N Engl J Med 332:1181, 1995

99. Cesarman E, Chang Y, Moore PS, et al: Kaposi's sarcoma–associated herpesvirus-like DNA sequences in AIDS-related body-cavity–based lymphomas. N Engl J Med 332:1186, 1995

100. Huang YQ, Li JJ, Kaplan MH, et al: Human herpesvirus-like nucleic acid in various forms of Kaposi's sarcoma. Lancet 345:759, 1995

101. Soulier J, Grollet L, Oksenhendler E, et al: Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease. Blood 86:1276, 1995

102. Kedes DH, Ganem D, Ameli N, et al: The prevalence of serum antibody to human herpesvirus 8 (Kaposi sarcoma–associated herpesvirus) among HIV-seropositive and high-risk HIV-seronegative women. JAMA 277:478, 1997

103. Lennette E, Blackboum DJ, Levy JA: Antibodies to human herpesvirus type 8 in the general population and in Kaposi's sarcoma patients. Lancet 348:858, 1996

104. Gao SJ, Kingsley L, Hoover DR, et al: Seroconversion to antibodies against Kaposi's sarcoma–associated herpesvirus-related latent nuclear antigens before the development of Kaposi's sarcoma. N Engl J Med 335:233, 1996

105. Andreoni M, Sarmati L, Nicastri E, et al: Primary human herpesvirus 8 infection in immunocompetent children. JAMA 287:1295, 2002

106. Oksenhendler E, Cazals-Hatem D, Schulz TF, et al: Transient angiolymphoid hyperplasia and Kaposi's sarcoma after primary infection with human herpesvirus 8 in a patient with human immunodeficiency virus infection. N Engl J Med 338:1585, 1998

107. Mbulaiteye SM, Pfeiffer RM, Whitby D, et al: Human herpesvirus 8 infection within families in rural Tanzania. J Infect Dis 187:1780, 2003

108. Brayfield BP, Phiri S, Kankasa C, et al: Postnatal human herpesvirus 8 and human immunodeficiency virus type 1 infection in mothers and infants from Zambia. J Infect Dis 187:559, 2003

109. Casper C, Wald A, Pauk J, et al: Correlates of prevalent and incident Kaposi's sarcoma-associated herpesvirus infection in men who have sex with men. J Infect Dis 185:990, 2002

110. Greenblatt RM, Jacobson LP, Levine AM, et al: Human herpesvirus 8 infection and Kaposi's sarcoma among human immunodeficiency virus–infected and –uninfected women. J Infect Dis 183:1130, 2001

111. Cannon MJ, Dollard SC, Smith DK, et al: Blood-borne and sexual transmission of human herpesvirus 8 in women with or at risk for human immunodeficiency virus infection. N Engl J Med 344:637, 2001

112. Kedes DH, Ganem D: Sensitivity of Kaposi's sarcoma–associated herpesvirus replication to antiviral drugs: implications for potential therapy. J Clin Invest 99:2082, 1997

113. Neyts J, De Clercq E: Antiviral drug susceptibility of human herpesvirus 8. Antimicrob Agents Chemother 41:2754, 1997

114. Little RF, Merced-Galindez F, Staskus K, et al: A pilot study of cidofovir in patients with Kaposi sarcoma. J Infect Dis 187:149, 2003

115. Gill J, Bourboulia D, Wilkinson J, et al: Prospective study of the effects of antiretroviral therapy on Kaposi sarcoma-associated herpesvirus infection in patients with and without Kaposi sarcoma. J Acquir Immune Defic Syndr 31:384, 2002

XXVIII ENTERIC VIRAL INFECTIONS

Nino Khetsuriani, m.d., ph.d.
Umesh D. Parashar, m.d., m.p.h.

Enterovirus Infections

Human enteroviruses belong to the genus Enterovirus, family Picornaviridae. They are divided into five species (human enterovirus A, B, C, and D, and poliovirus) comprising 68 currently recognized serotypes: polioviruses (types 1 to 3); coxsackieviruses A (types 1 to 14, 16, 17, 19 to 22, and 24) and B (types 1 to 6); echoviruses (types 1 to 7, 9, 11 to 21, 24 to 27, and 29 to 33); and numbered enteroviruses (types 68 to 71 and 73 to 78).[1] Enteroviruses are small (approximately 30 nm), nonenveloped, single-stranded RNA viruses with an icosahedral capsid composed of 60 subunits consisting of four structural proteins (VP1 to VP4). Enterovirus RNA is approximately 7.5 kb long and codes structural proteins, RNA polymerase, other polypeptides necessary for viral replication, and two untranslated regions at the 5´ and 3´ ends of the RNA molecule. New enteroviruses continue to be identified as molecular techniques of enterovirus typing become increasingly available. In addition, some serotypes have been found to be identical to other enteroviruses (e.g., echovirus 8 is the same as echovirus 1) or have been reclassified as members of other genera (e.g., echoviruses 22 and 23 are now human parechoviruses 1 and 2, respectively).

epidemiology

Enteroviruses are among the most common viruses worldwide. In temperate climates, the incidence of enterovirus infections peaks during the summer and fall months (June through October in the United States), whereas in the tropics, transmission occurs year-round.

Several enterovirus serotypes commonly cocirculate in the community, and predominant serotypes tend to change over time. The serotypes most commonly reported in the United States included echoviruses 30, 11, 9, 6, and 7 (from 1997 to 1999[2]), echoviruses 13 and 18 (from 2000 to 2001[3]), and echoviruses 9 and 30 (in 2003).[4] The appearance of new predominant enterovirus serotypes is often accompanied by large-scale outbreaks. For example, large outbreaks of aseptic meningitis were reported in 2001, when previously rare echoviruses 13 and 18 became the predominant serotypes in the United States.[3,5]

Illnesses associated with enteroviruses affect predominantly children, but persons of any age may be susceptible. In addition to young age, predisposing factors for enterovirus illness include male sex, lower socioeconomic status, residence in urban areas, poor sanitation, large household size, and crowded living conditions.[6]

The enteroviruses are universally found in the environment (e.g., in sewage; surface water, such as rivers, lakes, and other reservoirs; and seawater), because they are excreted in large amounts by infected persons. The lack of a lipid envelope and the presence of a dense protein capsid allow enteroviruses to survive in the environment for long periods, especially at low temperatures. Enteroviruses are inactivated by extreme heat, ultraviolet light, drying, and chlorine. A high concentration of organic matter prevents inactivation of enteroviruses by chlorine.[7]

Enteroviruses are most often transmitted from person to person by the fecal-oral or the respiratory route, but transmission by fomites also occurs. Young children are the most important transmitters of enteroviruses and are often the index case in family or school outbreaks.

pathogenesis

Enteroviruses gain entry into the host cell after binding to specific cellular receptors. Several different receptors with tropism to different serotypes have been identified. Initial virus replication takes place in submucosal lymphatic tissues of the pharynx and the gut (Peyer patches). The virus then spreads to the regional lymph nodes, enters the bloodstream ("minor" viremia), and reaches the reticuloendothelial system (i.e., deep lymph nodes, bone marrow, spleen, and liver). If viral replication is not contained by host defense mechanisms at this stage, symptomatic infection occurs, resulting in "major" viremia and dissemination of the virus to target organs. Certain enteroviruses (e.g., polioviruses) may spread along neural pathways as well. Tissue tropism of the serotype determines target organs for further replication. The histopathologic lesions in target organs consist of inflammation and necrosis of various degrees.

The incubation period for enteroviruses usually is 3 to 7 days (range, 1 to 35 days). Shorter incubation periods are observed in children and for certain specific clinical presentations of enterovirus infections (e.g., upper respiratory illnesses and acute hemorrhagic conjunctivitis).

Excretion of enteroviruses by an infected individual starts at the end of the incubation period, and patients are most infectious shortly before and after onset of illness. Depending on the stage of the infection and the clinical syndrome, enteroviruses are found in oropharyngeal secretions, stool, cerebrospinal fluid, blood, and vesicular fluids. In immunocompetent hosts, viral shedding in oropharyngeal secretions and circulation in the blood or CSF usually lasts up to 1 week, but fecal excretion may persist for up to 2 months after infection, even after clinical symptoms resolve. Long-term persistence of enteroviruses (in some cases, for several years) may occur in individuals with impaired humoral immunity.[8,9]

immunity

Immunity to enteroviruses is long lasting and serotype specific and is primarily mediated by humoral mechanisms. Primary infection results in an IgM response and is followed by IgG and IgA production. Secondary infections induce an anamnestic response, resulting in high antibody titers. Circulating IgG and IgM antibodies have a neutralizing effect on enteroviruses during the extracellular phase of replication, and secretory IgA antibodies mediate local mucosal immunity. The importance of antibodies in providing protective immunity against enteroviruses is emphasized by the occurrence of severe chronic enterovirus infections in patients with defective humoral immunity.

clinical syndromes

The vast majority of enterovirus infections are clinically inapparent or result only in a minor febrile illness. A smaller proportion of infected individuals develop more serious diseases, and

Table 1 Diseases Associated with Enteroviruses

Disease	Associated Enteroviruses*					Clinical Features
	Polio	*Coxs (Group A)*	*Coxs (Group B)*	*Echo*	*Numbered EV*	
Febrile illness	+	+	+	+	+	Mild, transient, febrile illness with or without rash or respiratory symptoms; resolves within a few days
Neurologic illnesses						
Aseptic meningitis	+	+	+ (Coxs B5)	+ (Echo 30, 9, 11, 6, 4, 7, 13, 18)	+ (EV 71)	Acute onset of fever, severe headache with meningeal irritation; meningeal symptoms in young infants may not be obvious
						CSF findings: mild to moderate pleocytosis (< 1,000 cells/mm^3) that is predominantly lymphocytic; rarely, normal cell counts; an initial predominance of polymorphonuclear cells in CSF with a subsequent shift to lymphocytosis may be observed, particularly in young children; CSF glucose level, normal; protein level, normal or elevated; CSF bacterial cultures, negative; enteroviruses often are detected in CSF by culture or PCR; often accompanied by other symptoms of enteroviral infection (respiratory symptoms, rashes, or myalgia)
						Usually resolves with no neurologic deficits in about 1 wk; illness and convalescence in adults may last longer
Encephalitis	+	+	+ (Coxs B5)	+ (Echo 30, 11, 6, 9, 24)	+ (EV 71)	Clinical presentation of meningoencephalitis; symptoms of aseptic meningitis plus confusion, lethargy, seizures; rarely, coma, cerebellar ataxia, choreiform movements, and paresthesias; diffuse encephalitis more common than focal; CSF profile similar to that of aseptic meningitis
						Outcome usually favorable; residual neurologic deficits uncommon; severe illness and deaths mostly in neonates and with EV 71 infection
Paralytic illness	+	+ (Coxs A7)	+	+	+ (EV 70, 71)	Usually preceded or accompanied by febrile illness
						Poliomyelitis: acute, usually asymmetrical, rapidly progressive flaccid paralysis; decreased or absent deep tendon reflexes; normal sensory function; respiration and swallowing may be compromised; the extent of paralysis depends on the location of CNS damage (spinal, bulbar, or spinobulbar forms)
						Muscle weakness associated with nonpolio enteroviruses usually milder than polio and transient; other manifestations of enteroviral infection common
						Residual permanent paralysis occurs in paralytic poliomyelitis but is rare after infection with nonpolio enteroviruses
						Death: in fewer than 10% of patients with poliomyelitis, mostly in patients with bulbar involvement; fatal bulbar involvement with nonpolio enteroviruses uncommon
Skin/mucosal syndromes						
Nonspecific rashes		+	+	+ (Echo 9, 11, 16)	+	Variety of exanthems and enanthems (e.g., maculopapular, roseolous, vesicular), rarely as a sole manifestation of enteroviral illness; resolve within a few days
Herpangina		+ (Coxs A10, 16, 2, 4, 5, 8)				Acute onset; fever, headache, malaise, sore throat with 10 to 12 vesicular enanthems, usually ulcerating, on soft palate, anterior tonsillar pillars, or posterior pharynx
						Self-limiting illness; complete recovery in a few days
Hand, foot, and mouth disease		+ (Coxs A16)			+ (EV 71)	Benign illness with acute onset; fever and mild pharyngitis 1 to 2 days before rash onset; lesions initially maculopapular but evolve into whitish-gray tender, flat, often oval vesicles; lesions on oral mucosa, dorsal surface of hands and feet, sometimes buttocks; peripheral distribution of rash on the limbs
						Recovery usually complete in a few days; CNS complications with fatalities observed in outbreaks of hand, foot, and mouth disease caused by EV 71

on relatively rare occasions, conditions with substantial sequelae or death may occur.

Of host factors influencing the clinical outcome of enterovirus infection, age and the degree of immunocompetence are the most important. Agent factors, such as the serotype or intratypic strain of the virus and tissue tropism, are also important. Depending on preferential target organs, enteroviruses can affect several organ systems and tissues, including the respiratory system, central nervous system, heart, skin, muscles, and gastrointestinal tract. Associations between distinctive clinical syndromes and serotypes have been established [*see Table 1*], but there is considerable overlap.

Minor Febrile Illness

Enteroviruses are a common cause of febrile illness, particularly during the summer and fall months. Fever may be the sole symptom, or infection may be accompanied by rashes or respiratory symptoms (the "summer cold"); in infants, symptoms include irritability, lethargy, anorexia, vomiting, and diarrhea. Because of the need to rule out potentially serious bacterial infec-

Table 1 (continued)

Disease	Associated Enteroviruses*					Clinical Features
	Polio	*Coxs (Group A)*	*Coxs (Group B)*	*Echo*	*Numbered EV*	
Muscle diseases						
Myopericarditis		+ (Coxs A4, A16)	+ (All serotypes)	+ (Echo 9, 22)		Both myocardium and pericardium are affected, but signs of either myocarditis or pericarditis predominate Dyspnea, chest pain, fever, malaise; ECG abnormalities (ST segment elevation, nonspecific ST segment and T wave abnormalities, ventricular tachyarrhythmias and heart block); cardiomegaly; serum levels of myocardial enzymes frequently elevated Severity varies; permanent myocardial injury possible in approximately 30% of cases; may later lead to chronic dilated cardiomyopathy Deaths in acute period rare (< 5%), mostly in patients with severe myocardial involvement; mortality in neonates approximately 50%
Pleurodynia			+			Synonyms: epidemic myalgia, Bornholm disease Attacks of spasmodic paroxysmal pain of variable intensity in the chest or upper abdomen with fever, which subsides as the pain recedes; pain usually more intense in adults; affected muscles painful by palpation; may be swollen; chest auscultation and x-ray normal Usually resolves within 1 wk; multiple relapses possible
Systemic infections						
Neonatal systemic infection			+ (All serotypes)	+ (Echo 11)		Overwhelming sepsislike illness Initial symptoms (fever, poor feeding, irritability, lethargy, hypotonus, and vomiting) are followed by multiple organ involvement; dominant syndromes include encephalomyocarditis (severe myocarditis, often accompanied by heart failure, and meningoencephalitis) and hemorrhage-hepatitis (overwhelming hepatitis with hepatic failure and DIC) High mortality (50% to 83%); most deaths within 1 wk of onset; the risk of fatal outcome highest with onset in infants younger than 2 wk
Chronic EV infection in immunocompromised host				+		Occurs in patients with B cell defects; chronic, persistent infection of CNS (meningoencephalitis), skeletal muscles (dermatomyositis-like syndrome), and other organs Progressive course; outcome usually fatal
Acute hemorrhagic conjunctivitis		+ (Coxs A24 variant)			+ (EV 70)	Acute eye pain, photophobia, swelling of eyelids, subconjunctival hemorrhages; transient corneal involvement possible Complete recovery in less than 10 days; permanent poliolike paralysis with onset after a few weeks of the EV 70 illness has been described
Respiratory illnesses		+ (Coxs A21, A24)	+ (Coxs B2)	+ (Echo 11, 9, 4, 18)	+ (EV 71)	Acute febrile illness; both upper (coryza, pharyngitis) and lower (bronchitis, tracheobronchitis, bronchiolitis, pneumonia) respiratory manifestations occur Course usually benign with complete recovery, except in neonatal systemic illness; fatal pulmonary edema observed in Southeast Asian outbreaks of EV 71
Diarrhea		+	+	+ (Echo 11, 14, 18)		Diarrhea usually febrile, accompanying other symptoms of enterovirus infection; recovery complete

*The most commonly associated serotypes are indicated in parentheses. More common serotypes are listed first.
CNS—central nervous system Coxs—coxsackievirus CSF—cerebrospinal fluid DIC—disseminated intravascular coagulation Echo—echovirus EV—enterovirus PCR—polymerase chain reaction Polio—poliovirus

tions, a high proportion of infants with enteroviral febrile illness require evaluation for bacterial sepsis or meningitis.[10]

Aseptic Meningitis

Aseptic meningitis is the most common CNS illness associated with enteroviruses; enteroviruses account for more than 90% of cases of aseptic meningitis for which a causative organism is identified.[11] Patients experience the acute onset of fever, headache, nuchal rigidity, photophobia, nausea, and vomiting. The fever may be biphasic, subsiding for several days and then recurring; development of meningeal symptoms occurs after recurrence of fever. A variable degree of lethargy, irritability, and drowsiness may be present, but considerable changes in mental status are uncharacteristic. Meningeal signs may not be obvious in young infants. Other manifestations of enterovirus infection, such as respiratory symptoms, exanthems, enanthems, or myalgia, are often present.

Examination of the CSF reveals a mild to moderate pleocytosis, with a predominance of lymphocytes (usually < 1,000 cells/mm^3). In some patients—particularly young children—

polymorphonuclear cells may predominate initially, with the subsequent development of lymphocytosis. In rare cases, CSF cell counts are normal. CSF glucose levels are normal; protein levels may be normal or elevated.

Symptoms usually resolve in about 1 week, although CSF pleocytosis may persist for some time thereafter. Severe illness and fatal outcome are uncommon except in the immediate neonatal period. Recovery is usually complete, with no significant sequelae.

Encephalitis

Enteroviral encephalitis results from progression of enteroviral meningitis; therefore, most patients with enteroviral encephalitis have symptoms of meningoencephalitis. Meningoencephalitis usually starts as aseptic meningitis, but it is accompanied by changes in mental status or cerebral involvement, characterized by somnolence, lethargy, generalized or focal seizures, and psychiatric symptoms. Coma, cerebellar ataxia, choreiform movements, and paresthesias are rare. Diffuse symptoms are more common than focal symptoms. The outcome is favorable in the majority of cases, but severe residual neurologic damage and death have been reported. In neonates, the disease is part of a systemic illness. In such patients, the disease follows a much more severe course; the prognosis is guarded, and there is a higher risk of neurologic sequelae and death.[12] In Southeast Asia, fulminant, rapidly fatal bulbar encephalitis has occurred in young children during outbreaks of enterovirus 71 infection.[13]

Poliomyelitis and Poliolike Paralytic Illness

Muscle paralysis is primarily associated with poliovirus infection but has also been observed, sometimes in outbreaks, with several nonpolio enteroviruses, such as coxsackievirus A7, enteroviruses 70 and 71, and several echoviruses. Through successful immunization programs, control of poliomyelitis has been achieved in numerous countries worldwide, including the entire Western Hemisphere, but wild polioviruses continue to circulate in some countries of sub-Saharan Africa and Southeast Asia. It is likely that the large-scale Polio Eradication Initiative, led by the World Health Organization since 1988, will result in global polio eradication[14]; until that occurs, there remains a risk of importation of poliomyelitis into the United States.

Paralysis in poliomyelitis results from the replication of polioviruses in the anterior horn of the spinal cord or in the brain stem, causing the destruction of motor neurons. Depending on the location of the damage, spinal, bulbar, or spinobulbar forms of poliomyelitis may occur. Clinical hallmarks of paralytic poliomyelitis include the acute onset of flaccid, usually asymmetrical, rapidly progressive paralysis that is more severe proximally than distally and that is characterized by decreased or absent deep tendon reflexes without sensory loss. Respiration and swallowing may be compromised if respiratory muscles, cranial nerves, or respiratory centers are affected. Death occurs in fewer than 10% of cases; the majority of deaths involve cases of bulbar poliomyelitis. Recovery of muscle function takes place during the first few months after disease onset, but some degree of permanent paralysis usually remains.

Paralysis associated with nonpolio enterovirus infections results more from inflammatory changes than motor neuron destruction; therefore, it is usually milder and transient, rarely resulting in residual paresis. Fatal bulbar involvement is uncommon.

Myopericarditis

Carditis caused by enteroviruses (predominantly group B coxsackieviruses) occurs mostly in newborns, adolescents, and young adults and usually affects both pericardium and myocardium. Electrocardiographic abnormalities of varying degree are present; these range from ST segment elevation or nonspecific ST segment and T wave abnormalities to ventricular tachyarrhythmias and heart block. Cardiomegaly caused by pericardial effusion or acute cardiac dilatation is common. Serum levels of myocardial enzymes are frequently elevated. Death in the acute period is rare (< 5%) and mostly occurs in patients with severe myocardial involvement. Most patients recover without complications, but permanent myocardial injury occurs in about 30% of cases. Residual damage includes persistent ECG abnormalities, cardiomegaly, and chronic congestive heart failure and may later lead to dilated cardiomyopathy.[15,16]

Pleurodynia

Pleurodynia is characterized by attacks of spasmodic paroxysmal pain in the chest or upper abdomen accompanied by fever, which subsides as the pain recedes. The intensity of pain varies and is usually greater in adults. Affected muscles are painful on palpation and may be swollen. Normal auscultatory examination and chest x-ray are helpful in the differential diagnosis. The disease resolves within 1 week, but some patients experience multiple relapses.

Exanthems and Enanthems

Skin and mucosal lesions seldom appear as a sole presentation of enterovirus infection. Maculopapular rashes are most commonly associated with echoviruses; they appear simultaneously with fever, and spread from the face to the chest, neck, and extremities. Roseoliform rashes appear on the face and upper chest after defervescence. Of the vesicular rashes, the most distinctive clinical syndrome is hand, foot, and mouth disease. This disease is most often caused by coxsackievirus A16, but it can also be caused by enterovirus 71. Most cases of hand, foot, and mouth disease occur in children, who present with fever, vesicular lesions in the mouth, and macular rash on the limbs and sometimes buttocks; these lesions may later become vesicular. The disease generally resolves in a few days, but CNS complications with fatalities occurred in Southeast Asia during large outbreaks of hand, foot, and mouth disease caused by enterovirus 71.[13,17] The most common example of enterovirus enanthems is herpangina, a common illness of young children that is characterized by fever, sore throat, and vesicular enanthems on the soft palate; symptoms resolve within a few days.

Neonatal Systemic Infection

Enterovirus infections in infants usually are benign, self-limited illnesses similar to those observed in older persons. In some neonates, especially during the first 2 weeks of life, enteroviruses may cause an overwhelming sepsislike illness.[18,19] In most cases, infection is transmitted from the sick mother (either transplacentally or during delivery), but nosocomial infections, including nursery outbreaks, also occur. The initial symptoms associated with neonatal enteroviral sepsis are nonspecific and include fever, poor feeding, irritability, lethargy, hypotonia, and vomiting. There subsequently occurs involvement of multiple organs, notably the heart, CNS, liver, lungs, pancreas, and adrenal glands. The dominant features of the encephalomyocarditis syndrome, which is associated with group B coxsackieviruses, are

severe myocarditis, often accompanied by heart failure, and meningoencephalitis. The leading clinical features of the hemorrhage-hepatitis syndrome, predominantly associated with echovirus 11, are overwhelming hepatitis with hepatic failure and disseminated intravascular coagulation. Neonatal enteroviral sepsis is often fatal (reported case fatality ranges from 50% to 83%), and most deaths occur within 1 week of onset.[12,18,19]

Chronic Enterovirus Infection in Immunocompromised Hosts

Neutralizing antibodies play a critical role in the immune response to enteroviruses. Thus, enteroviruses often persist in patients with inherited or acquired defects of humoral immunity; this results in chronic infections of the CNS, skeletal muscles, and the GI system. Most commonly, the condition occurs in children with X-linked agammaglobulinemia, but it may also develop in children with severe combined immunodeficiency syndrome and, rarely, in bone marrow transplant recipients.[8,9] Chronic progressive meningoencephalitis is the predominant clinical feature. More than half of patients experience a dermatomyositis-like syndrome. Chronic hepatitis is commonly present. The course of illness is progressive; although periodic improvements may occur, the overall prognosis is poor, and the disorder is usually fatal.[9]

Acute Hemorrhagic Conjunctivitis

Acute hemorrhagic conjunctivitis occurs in explosive outbreaks, mostly in tropical areas, and is caused by coxsackievirus A24 variant and enterovirus 70. The disease is highly contagious; unlike other enteroviruses, the primary mode of transmission is direct introduction of the virus into the eye by fingers or fomites. The symptoms include eye pain, photophobia, swelling of eyelids, and characteristic subconjunctival hemorrhages of various intensity. The symptoms initially affect one eye and then spread to the other. The disease usually resolves in less than 10 days. Corneal involvement may occur but is transient and leaves no permanent scars. A permanent poliolike paralysis has been reported in persons who had recently had acute hemorrhagic conjunctivitis caused by enterovirus 70.

Other Illnesses

The respiratory syndromes in enterovirus infections involve both upper and lower respiratory tracts and range from the so-called "summer cold" to pneumonia. These infections are usually mild, except when associated with systemic illness in neonates. Fatal pulmonary edema, apparently of neurogenic origin, was reported to have occurred in association with the 1998 outbreak of enterovirus 71 infection in Taiwan.[17]

GI symptoms are often associated with other manifestations of enterovirus infections. In young children, outbreaks of febrile diarrhea associated with some echoviruses have been described, but in general, enteroviruses are not a major cause of gastroenteritis.

Enteroviruses are also known to cause hepatitis and pancreatitis, usually as part of generalized infection. An association of enterovirus infections, particularly group B coxsackievirus infection, with type 1 (insulin-dependent) diabetes mellitus has been suggested, but no conclusive evidence is available.[20,21]

DIAGNOSIS

Isolation of the virus with subsequent serotyping has traditionally been the gold standard for the laboratory diagnosis of enterovirus infections. Most enteroviruses can be grown in susceptible cell lines and identified by observing a characteristic cytopathic effect, but for almost all group A coxsackieviruses, inoculation of suckling mice is needed. The use of several cell lines and multiple specimen types from the patient increases the diagnostic yield. However, viral culture is relatively insensitive, laborious, and time consuming; these factors limit the utility of viral culture for patient care.[22]

The most widely accepted method for the identification of individual enterovirus serotypes is neutralization reaction, using intersecting pools of internationally standardized antisera. Immunofluorescence assay with monoclonal antibodies is also available for several common enteroviruses.

Compared with viral culture for enterovirus detection, reverse transcriptase polymerase chain reaction (RT-PCR) assays have been shown to be more sensitive, as specific, and much more rapid. These assays are being used increasingly in clinical practice with demonstrated clinical utility, particularly for the diagnosis of enterovirus meningitis.[18,22,23] The primers most often used in enterovirus RT-PCR have broad specificity for enteroviruses in general; this precludes serotype identification, but specific primers for individual serotypes are becoming available.[5,24] Molecular typing by sequence analysis, based on the close correlation of the serotype with the nucleotide sequence of the *VP1* gene, is a new modality for enterovirus identification that has led to identification of several previously unknown serotypes.[1,25]

Detection of the virus in normally sterile sites (e.g., CSF, blood, pericardial fluid, tissue specimens) is considered diagnostic. Because enteroviruses are ubiquitous and because asymptomatic infections commonly occur, positive results from testing of nonsterile sites (e.g., stool sample, throat swab) should be interpreted with caution.

Serology has a limited role in the diagnosis of enterovirus infections; it is helpful only if paired sera are available for testing. Demonstration of a greater than fourfold increase in titers of antibodies against the implicated serotype is indicative of recent infection. Of various serologic methods, virus neutralization reaction is preferred.

TREATMENT

Most patients with enterovirus infections recover uneventfully after a few days of supportive care, but patients with potentially serious illnesses require intensive support. No antiviral therapy for enterovirus infections is currently available.

Immunoglobulin preparations are used both for treatment and prophylaxis of enterovirus infections in patients with humoral immunodeficiencies.[9] The occurrence of chronic enterovirus meningoencephalitis has declined notably since the introduction of regular prophylactic immunoglobulin treatment. The benefit for patients with established chronic enterovirus meningoencephalitis is less clear. Most commonly, immunoglobulin preparations are administered intravenously. Results of intramuscular and intrathecal administration are mixed.[9] Intravenous immunoglobulin is sometimes used for neonatal enterovirus infections, but the data on the effectiveness of this therapy are limited.[18,26]

PREVENTION AND CONTROL

Two types of highly effective vaccines—inactivated poliovaccine (IPV) and live, attenuated oral poliovaccine (OPV)—are available for prevention and control of poliovirus infection. Since 2000, IPV has been the recommended vaccine for routine use in the United States.[27]

In the absence of vaccines, prevention and control of nonpolio enteroviral illnesses are primarily accomplished through adherence to good personal hygiene (e.g., thorough hand washing, especially after diaper changes). To prevent neonatal infections, routine infection-control measures in neonatal nurseries must be strictly enforced, and pregnant women near term should be advised to avoid contact with patients with known or suspected enterovirus infections. Control of enterovirus transmission is complicated by the fact that the majority of infections are asymptomatic and by the relatively long duration of viral shedding. For similar reasons, the effectiveness of isolation of symptomatic persons is questionable.

Viral Gastroenteritis

Viral gastroenteritis occurs in two distinct epidemiologic forms: sporadic disease that commonly affects children, and epidemic disease that afflicts both children and adults. Sporadic childhood viral gastroenteritis is primarily caused by rotaviruses (group A), human caliciviruses (including both noroviruses [previously called Norwalk-like viruses] and sapoviruses [previously called Sapporo-like viruses]), enteric adenoviruses, and astroviruses. Epidemic viral gastroenteritis is caused most often by norovirus.

ROTAVIRUS INFECTIONS

In 1973, Bishop and colleagues visualized by electron microscopy a 70-nm triple-layered virus in the duodenal epithelium of children with diarrhea.[28] This virus, designated as rotavirus because of its morphologic appearance (in Latin, *rota* means wheel), belongs to the family Reoviridae; its genome consists of 11 segments of double-stranded RNA. The segmented genome of rotavirus allows reassortment during coinfection, a property that has been utilized in the development of rotavirus vaccines and is probably important in virus evolution. There are seven major groups of rotavirus (groups A to G); human illness is caused by rotaviruses of groups A, B, and C. Two outer capsid proteins, the glycoprotein (G protein) and the protease-cleaved protein (P protein), determine the serotype specificity and form the basis of the binary classification (G and P types) of rotaviruses. Both G protein and P protein induce neutralizing antibodies.

Epidemiology

Rotaviruses are ubiquitous, infecting 95% of children worldwide by the time they are 3 to 5 years of age.[29] Neonatal infections, although common, are often asymptomatic. The incidence of clinical illness peaks in children 4 to 23 months of age. Globally, rotaviruses are the most common agents of severe childhood gastroenteritis; they are estimated to cause about one third of all gastroenteritis hospitalizations and 500,000 deaths a year.[30] Rotavirus infections in adults are usually subclinical but occasionally cause illness in parents of children with rotavirus diarrhea and in immunocompromised individuals, the elderly, and travelers.

In temperate climates, rotavirus disease predominantly occurs during the fall and winter months. A study from Japan showed that, unlike in children, rotavirus diarrhea in adults did not show significant winter seasonality.[30] In the United States, rotavirus activity peaks in the Southwest in autumn (October through December) and migrates across the continent, peaking in the Northeast during spring (March through May).[31] In tropical settings, rotavirus disease occurs year-round.

During episodes of diarrhea, rotaviruses are shed in large amounts in stool; fecal shedding detectable by antigen enzyme immunoassays usually subsides within a week but may persist for more than 30 days in immunocompromised persons. Viral shedding may be detected for longer periods using more sensitive assays, such as PCR. Transmission of rotavirus occurs predominantly through the fecal-oral route. Transmission through respiratory secretions, person-to-person contact, or contaminated environmental surfaces has been postulated.

In humans, most rotavirus diseases, including endemic childhood diarrhea, are caused by group A rotaviruses. Group B rotaviruses have caused large epidemics of severe gastroenteritis in China since 1982 and have also been identified in India.[32,33] Group C rotaviruses have been associated with epidemic gastroenteritis worldwide, and some studies indicate a possible association with extrahepatic biliary atresia in infants.[34]

Rotavirus strains that infect animals differ from those that infect humans. Although some strains of human rotavirus possess a high degree of genetic homology with animal strains, animal-to-human transmission is uncommon.

Pathogenesis

Rotaviruses infect the mature enterocytes in the middle or upper villous epithelium of the small intestine. Ultimately, the epithelium becomes necrotic and sloughs off. The loss of absorptive villous epithelium coupled with proliferation of secretory crypt cells reverses the inherent absorptive state of the epithelium, resulting in secretory diarrhea. Levels of brush-border enzymes characteristic of differentiated cells (e.g., sucrase and lactase) are reduced, leading to the accumulation of unmetabolized disaccharides in the gut lumen, with consequent osmotic diarrhea. The toxinlike effect of a nonstructural rotavirus protein, NSP4, increases intracellular calcium through the opening of a cation channel, causing an efflux of chloride and, thus, sodium and water, contributing to the secretory diarrhea.[35] Furthermore, rotavirus appears to evoke the secretion of intestinal fluids through activation of the enteric nervous system.[36]

Recent data indicate that a transient antigenemia that peaks 1 to 3 days after onset of symptoms is common in children with rotavirus infection. It remains to be proved whether the presence of antigenemia is correlated with increased severity of disease or with extraintestinal manifestations that are sometimes seen in children with rotavirus infection.[37,38]

Immunity

Protection against rotavirus disease is correlated with the presence of virus-specific secretory IgA antibodies in the feces and serum.[39,40] Because virus-specific IgA is short-lived at the intestinal surface, complete protection against natural rotavirus disease is only temporary. Memory B and T cells in the lamina propria are believed to be important in reducing the severity of disease resulting from reinfection.[41]

Diagnosis

Clinical manifestations The clinical spectrum of rotavirus infection ranges from subclinical illness to severe gastroenteritis associated with life-threatening dehydration. The onset of illness is abrupt; vomiting frequently precedes the development of diarrhea. Up to one third of patients may have a temperature greater than 39° C (102.2° F). GI symptoms generally resolve in 3 to 6 days.

Respiratory symptoms have been observed in children with rotavirus infection, as have symptoms of CNS involvement, but

these associations have not been well studied.[42] Rotavirus infection has been observed in patients with a variety of other clinical syndromes, including sudden infant death syndrome, Reye syndrome, necrotizing enterocolitis, intussusception, Kawasaki syndrome, disseminated intravascular coagulation, and Crohn disease. A causal relationship has not been confirmed between any of these syndromes and rotavirus infection.

In immunodeficient children, rotavirus can cause a protracted diarrhea with prolonged viral excretion and, in rare instances, can disseminate systemically and cause hepatic infection.[43]

Laboratory tests Illness caused by rotavirus is difficult to distinguish clinically from that caused by other enteric viruses. Because a large number of viruses are shed in feces, the diagnosis usually can be confirmed by using one of a wide variety of commercial immunoassays or by techniques for detecting viral RNA, such as gel electrophoresis, probe hybridization, or RT-PCR.

Treatment and Prevention

Treatment relies primarily on replacement of fluids and electrolytes; antibiotics and antimotility agents should be avoided.[44] Although oral rehydration therapy is successful in most children, intravenous fluid replacement may be required for children who have severe dehydration or are unable to tolerate oral therapy. A variety of other therapeutic agents have been evaluated, including probiotics,[45] bismuth salicylate,[46] and enkephalinase inhibitors,[47] but their therapeutic roles are not clearly defined.

In 1998, only 25 years after the identification of rotavirus in humans, a live, attenuated rhesus-human reassortant rotavirus vaccine (Rotashield, Wyeth Laboratories, Marietta, Pennsylvania) with 80% efficacy against severe rotavirus disease was licensed in the United States and was recommended for routine immunization of infants. This vaccine was withdrawn, however, in 1999, after it was confirmed that its use was associated with intussusception.[48] Efforts are ongoing to develop other rotavirus vaccines. Two large clinical trials that each involved more than 60,000 infants have recently been completed. The trials examined two leading vaccine candidates—one developed by Merck (a multivalent bovine-human reassortant rotavirus vaccine)[49] and the other by GlaxoSmithKline (an attenuated single human strain rotavirus vaccine).[50] These trials demonstrated the safety of both vaccines with respect to intussusception and other potential adverse events,[51,52] and they indicated that the vaccines have an efficacy of more than 90% against severe rotavirus disease.[49,50] Data on the Merck vaccine, which is administered orally to infants in three doses at 2, 4, and 6 months of age, were submitted for licensure to the Food and Drug Administration in April 2005; a decision is expected in 2006.

HUMAN CALICIVIRUS INFECTIONS

In 1972, using immune electron microscopy, Kapikian and colleagues identified the 27-nm Norwalk agent in stool filtrates of a volunteer challenged with fecal specimens from patients affected by an outbreak of gastroenteritis.[53] Several related but genetically and antigenically diverse single-stranded RNA viruses of positive polarity measuring 27 to 35 nm were subsequently identified; these organisms are currently classified as noroviruses (previously called Norwalk-like viruses), in the family Caliciviridae. The Caliciviridae family also includes the sapoviruses (previously called Sapporo-like viruses), which cause gastroenteritis in children and adults, and Lagovirus and Vesivirus, neither of which is pathogenic for humans.

Epidemiology

Studies indicate that infections with the Norwalk and related caliciviruses are more common than previously believed; a majority of children and nearly all adults demonstrate antibodies to these viruses. Acquisition of antibody occurs at an earlier age in developing countries, as would be expected, given the presumed fecal-oral mode of transmission of these viruses. Although their role in sporadic illness in children and adults is still being defined, noroviruses are clearly recognized as the most common cause of epidemics of gastroenteritis worldwide.[54] Epidemics occur throughout the year and are often linked to consumption of fecally contaminated water and foods; often, food and water become contaminated by an infectious food handler. Dispersion of virus through direct contact, vomitus, or airborne droplets has been postulated in situations in which an alternative mode of transmission cannot be established.

Pathogenesis

Because no animal model exists for calicivirus gastroenteritis, the data on pathogenesis of disease arise exclusively from studies of human volunteers. After challenge of volunteers with either Norwalk virus or Hawaii virus, which is a related calicivirus, biopsy demonstrates lesions in the proximal small intestine; these lesions are characterized by villus shortening, crypt hyperplasia, and infiltration of the lamina propria by polymorphonuclear cells and lymphocytes.[55] The lesions persist for at least 4 days after the person's symptoms disappear and are associated with malabsorption of carbohydrates and fats and a decreased level of brush-border enzymes. No changes are observed in the stomach or colon, but gastric motor function is delayed, which probably contributes to the characteristic nausea and vomiting.[56]

Immunity

Approximately 50% of persons challenged with Norwalk virus become ill and acquire short-term homologous immunity (i.e., immunity against the same strain) that is correlated with serum antibody levels; immunity does not appear to persist for longer than 2 years.[57] Some reports indicate, paradoxically, that persons with higher levels of preexisting antibody to Norwalk virus are more susceptible to illness.[58]

Diagnosis

Clinical manifestations Gastroenteritis caused by Norwalk virus and related enteric caliciviruses occurs with sudden onset after a viral incubation period of 12 to 48 hours. The illness generally lasts 12 to 60 hours and is characterized by nausea, vomiting, abdominal cramps, and diarrhea. In children, vomiting is more prevalent than diarrhea, whereas in adults, diarrhea is more prevalent. Constitutional symptoms, such as headache, fever, chills, and myalgias, are frequently reported. Death is rare and usually occurs from severe dehydration in vulnerable persons, such as the elderly with debilitating health conditions.

Laboratory tests Cloning and sequencing of the genome of Norwalk virus and a few other human caliciviruses has allowed the development of sensitive detection methods based on PCR amplification and Southern blot analysis; strains can be further characterized by sequencing of nucleotide products. Expression

of capsid proteins in a recombinant baculovirus vector produces viruslike particles that have been used to develop immunoassays. Diagnostic assays, however, are not widely available.

Treatment and Prevention

Treatment is generally not required, because the disease is self-limited.[44] Rehydration may be required for patients with volume depletion.

Epidemic prevention relies on situation-specific measures, such as control of contamination of food and water, prevention of the handling of food by persons who are ill, and reduction of person-to-person spread through good personal hygiene. Calicivirus vaccines are being developed and may be particularly useful for groups for whom the impact of short-term morbidity from gastroenteritis is great (e.g., military troops) or those who are at high risk for severe disease (e.g., the elderly in nursing homes).

GASTROENTERITIS CAUSED BY OTHER VIRAL AGENTS

Enteric adenoviruses have double-stranded DNA and measure 70 to 80 nm; they belong to the family Adenoviridae. Serotypes 31, 40, and 41 cause approximately 10% of diarrheal illness in young children. Unlike their counterparts that cause respiratory illness, enteric adenoviruses are difficult to cultivate in cell lines. Their detection requires the use of immunoassays to the adenovirus hexon antigen and subsequent serotyping using monoclonal antibodies.

Astroviruses have positive-sense, single-stranded RNA and measure 28 to 30 nm; they belong to the family Astroviridae. Although at least seven different serotypes have been identified, strains of serotype 1 are most common. Preliminary epidemiologic studies indicate that astroviruses may be important causes of mild to moderate diarrhea in children, causing half as many illnesses as rotavirus.[59] The availability of simple immunoassays to detect virus in fecal specimens and of molecular methods to confirm and further characterize strains will allow for a more comprehensive assessment of the etiologic role of these agents in endemic childhood diarrhea.

Other viruses, including toroviruses, picobirnaviruses, coronaviruses, pestiviruses, and parvoviruses, have been identified in the feces of patients with diarrhea, but their etiologic role has not been well studied. It appears from preliminary studies that some of these agents, such as picobirnaviruses, may be important causes of diarrhea in HIV-infected persons.[60,61]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Stanway G, Brown F, Christian P, et al: Picornaviridae: virus taxonomy: classification and nomenclature of viruses. 8th report of the International Committee on the Taxonomy of Viruses. Fauquet CM, Mayo MA, Maniloff J, et al, Eds. Elsevier Academic Press, Amsterdam, 2005, p 757

2. Enterovirus surveillance—United States, 1997–1999. MMWR Morb Mortal Wkly Rep 49:913, 2000

3. Enterovirus Surveillance—United States, 2000–2001. MMWR Morb Mortal Wkly Rep 51:1047, 2002

4. Outbreaks of Aseptic Meningitis Associated with Echoviruses 9 and 30 and Preliminary Surveillance Reports on Enterovirus Activity—United States, 2003. MMWR Morb Mortal Wkly Rep 52:761, 2003

5. Mullins JA, Khetsuriani N, Nix WA, et al: Emergence of echovirus 13 as a prominent enterovirus. Clin Infect Dis 38:70, 2004

6. Morens MM, Pallansch MA: Epidemiology. Human Enterovirus Infections. Rotbart HA, Ed. ASM Press, Washington, DC, 1995, p 3

7. Feachem R, Garelick H, Slade J: Enteroviruses in the environment. Trop Dis Bull 78:185, 1981

8. Galama JM, de Leeuw N, Wittebol S, et al: Prolonged enteroviral infection in a patient who developed pericarditis and heart failure after bone marrow transplantation. Clin Infect Dis 22:1004, 1996

9. McKinney RE Jr, Katz SL, Wilfert CM: Chronic enteroviral meningoencephalitis in agammaglobulinemic patients. Rev Infect Dis 9:334, 1987

10. Rittichier KR, Bryan PA, Bassett KE, et al: Diagnosis and outcomes of enterovirus infections in young infants. Pediatr Infect Dis J 24:546, 2005

11. Rotbart HA: Viral meningitis. Semin Neurol 20:277, 2000

12. Modlin JF: Update on enterovirus infections in infants and children. Adv Pediatr Infect Dis 12:155, 1996

13. Chan LG, Parashar UD, Lye MS, et al: Deaths of children during an outbreak of hand, foot, and mouth disease in Sarawak, Malaysia: clinical and pathological characteristics of the disease. For the Outbreak Study Group. Clin Infect Dis 31:678, 2000

14. Progress Toward Interruption of Wild Poliovirus Transmission—Worldwide, January 2004–March 2005. MMWR Morb Mortal Wkly Rep 54:408, 2005

15. Smith WG: Coxsackie B myopericarditis in adults. Am Heart J 89:34, 1970

16. Helin M, Savola J, Lapinleimu K: Cardiac manifestations during a coxsackie B5 epidemic. BMJ 3:97, 1968

17. Chang LY, Lin TY, Hsu KH, et al: Clinical features and risk factors of pulmonary oedema after enterovirus-71–related hand, foot, and mouth disease. Lancet 354:1862, 1999

18. Lin TY, Kao HT, Hsieh SH, et al: Neonatal enterovirus infections: emphasis on risk factors of severe and fatal infections. Pediatr Infect Dis J 22:889, 2003

19. Abzug MJ: Presentation, diagnosis, and management of enterovirus infections in neonates. Paediatric Drugs 6:1, 2004

20. Hyoty H, Taylor KW: The role of viruses in human diabetes. Diabetologia 45:1353, 2002

21. Haverkos HW, Battula N, Drotman DP, et al: Enteroviruses and type 1 diabetes mellitus. Biomedicine & Pharmacotherapy 57:379, 2003

22. Tanel RE, Kao SY, Niemiec TM, et al: Prospective comparison of culture vs genome detection for diagnosis of enteroviral meningitis in childhood. Arch Pediatr Adolesc Med 150:919, 1996

23. Ramers C, Hartin M, Ho S, et al: Impact of a diagnostic cerebrospinal fluid enterovirus polymerase chain reaction test on patient management. JAMA 283:2680, 2000

24. Kilpatrick DR, Quay J, Pallansch MA, et al: Type-specific detection of echovirus 30 isolates using degenerate reverse transcriptase PCR primers. J Clin Microbiol 39:1299, 2001

25. Oberste MS, Maher K, Kilpatrick DR, et al: Typing of human enteroviruses by partial sequencing of VP1. J Clin Microbiol 37:1288, 1999

26. Abzug MJ, Keyserling HL, Lee ML, et al: Neonatal enterovirus infection: virology, serology, and effects of intravenous immune globulin. Clin Infect Dis 20:1201, 1995

27. Poliomyelitis prevention in the United States: updated recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 49[RR-5]:1, 2000

28. Bishop RF, Davidson GP, Holmes IH, et al: Virus particles in epithelial cells of duodenal mucosa from children with viral gastroenteritis. Lancet 2:1281, 1973

29. Parashar UD, Bresee JS, Gentsch JR, et al: Rotavirus. Emerg Infect Dis 4:561, 1998

30. Nakajima H, Nakagomi T, Kamisawa T, et al: Winter seasonality and rotavirus in adults. Lancet 357:1950, 2001

31. Torok TJ, Kilgore PE, Clarke MJ, et al: Visualizing geographic and temporal trends in rotavirus activity in the United States, 1991 to 1996. National Respiratory and Enteric Virus Surveillance System Collaborating Laboratories. Pediatr Infect Dis J 16:941, 1997

32. Fang ZY, Ye Q, Ho MS, et al: Investigation of an outbreak of adult diarrhea rotavirus in China. J Infect Dis 160:948, 1989

33. Kang G, Kelkar SD, Chitambar SD, et al: Epidemiological profile of rotaviral infection in India: challenges for the 21st century. J Infect Dis 192(suppl 1):S120, 2005

34. Riepenhoff-Talty M, Gouvea V, Evans MJ, et al: Detection of group C rotavirus in infants with extrahepatic biliary atresia. J Infect Dis 174:8, 1996

35. Ball JM, Tian P, Zeng CQ, et al: Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein. Science 272:101, 1996

36. Lundgren O, Peregrin AT, Persson K, et al: Role of the enteric nervous system in the fluid and electrolyte secretion of rotavirus diarrhea. Science 287:491, 2000

37. Blutt SE, Kirkwood CD, Parreno V, et al: Rotavirus antigenaemia and viraemia: a common event? Lancet 362:1445, 2003

38. Fischer TK, Ashley D, Kerin T, et al: Rotavirus antigenemia in patients with acute gastroenteritis. J Infect Dis 192:913, 2005

39. Coulson BS, Grimwood K, Hudson IL, et al: Role of coproantibody in clinical protection of children during reinfection with rotavirus. J Clin Microbiol 30:1678, 1992

40. Matson DO, O'Ryan ML, Herrera I, et al: Fecal antibody responses to symptomatic and asymptomatic rotavirus infections. J Infect Dis 167:577, 1993

41. Offit PA: Host factors associated with protection against rotavirus disease: the skies are clearing. J Infect Dis 174(suppl 1):S59, 1996

42. Lynch M, Lee B, Azimi P, et al: Rotavirus and central nervous system symptoms: cause or contaminant? Case reports and review. Clin Infect Dis 33:932, 2001

43. Gilger MA, Matson DO, Conner ME, et al: Extraintestinal rotavirus infections in children with immunodeficiency. J Pediatr 120:912, 1992

44. Farthing MJ: Treatment of gastrointestinal viruses. Novartis Found Symp 238:289, 2001

45. Szajewska H, Mrukowicz JZ: Probiotics in the treatment and prevention of acute infectious diarrhea in infants and children: a systematic review of published randomized, double-blind, placebo-controlled trials. J Pediatr Gastroenterol Nutr 33(suppl 2):S17, 2001

46. Figueroa-Quintanilla D, Salazar-Lindo E, Sack RB, et al: A controlled trial of bismuth subsalicylate in infants with acute watery diarrheal disease. N Engl J Med 328:1653, 1993

47. Salazar-Lindo E, Santisteban-Ponce J, Chea-Woo E, et al: Racecadotril in the treatment of acute watery diarrhea in children. N Engl J Med 343:463, 2000

48. Murphy TV, Gargiullo PM, Massoudi MS, et al: Intussusception among infants given an oral rotavirus vaccine. N Engl J Med 344:564, 2001

49. Salinas B, Schael IP, Linhares AC, et al: Evaluation of safety, immunogenicity and efficacy of an attenuated rotavirus vaccine, RIX 4414: a randomized, placebo-controlled trial in Latin American infants. Pediatr Infect Dis J 24:807, 2005

50. Velasquez FR, Abate H, Clemens SA: The human monovalent G1P[8] rotavirus vaccine, Rotarix, is highly efficacious and provides cross protection against G1 and non-G1 serotypes. Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

51. Vesikari T, Matson D, Van Damme P, et al: Incidence of intussusception with the pentavalent (human-bovine) reassortant rotavirus vaccine (PRV) is similar to placebo. Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

52. Vesikari T, Matson D, Dennehy P, et al: Protection against rotavirus gastroenteritis of multiple serotypes by a pentavalent (human-bovine) reassortant rotavirus vaccine (PRV). Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

53. Kapikian AZ, Wyatt RG, Dolin R, et al: Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. J Virol 10:1075, 1972

54. Parashar UD, Monroe SS: "Norwalk-like viruses" as a cause of foodborne disease outbreaks. Rev Med Virol 11:243, 2001

55. Schreiber DS, Blacklow NR, Trier JS: The mucosal lesion of the proximal small intestine in acute infectious nonbacterial gastroenteritis. N Engl J Med 288:1318, 1973

56. Meeroff JC, Schreiber DS, Trier JS, et al: Abnormal gastric motor function in viral gastroenteritis. Ann Intern Med 92:370, 1980

57. Graham DY, Jiang X, Tanaka T, et al: Norwalk virus infection of volunteers: new insights based on improved assays. J Infect Dis 170:34, 1994

58. Johnson PC, Mathewson JJ, DuPont HL, et al: Multiple-challenge study of host susceptibility to Norwalk gatroenteritis in US adults. J Infect Dis 161:18, 1990

59. Clark B, McKendrick J: A review of viral gastroenteritis. Curr Opin Infect Dis 17:461, 2004

60. Grohmann GS, Glass RI, Pereira HG, et al: Enteric viruses and diarrhea in HIV-infected patients. Enteric Opportunistic Infections Working Group. N Engl J Med 329:14, 1993

61. Giordano MO, Martinez LC, Rinaldi D, et al: Diarrhea and enteric emerging viruses in HIV-infected patients. AIDS Res Hum Retroviruses 15:1427, 1999

XXIX MEASLES, MUMPS, RUBELLA, PARVOVIRUS, AND POXVIRUS

MARTIN S. HIRSCH, M.D.

Measles

Measles (rubeola) is a highly infectious disease caused by a paramyxovirus of worldwide distribution. The measles virus was once one of the most common and important global pathogens, but infections caused by this agent are now becoming rare in developed countries, where vaccine use is widespread. In areas where vaccines are not widely used, measles remains a major health problem. It is estimated that worldwide, over 800,000 deaths occur each year from measles.[1]

EPIDEMIOLOGY

Humans are the principal reservoirs of measles virus, which is spread by respiratory droplet aerosols produced by sneezing and coughing. The disease may be contagious from several days before the onset of rash to up to 5 days after lesions appear. The attack rate for exposed susceptible contacts may exceed 90%; asymptomatic infections are rare.

Vaccines have dramatically reduced the incidence of indigenous measles in the United States and have eliminated the disease altogether in some developed countries. However, large outbreaks still occur in unvaccinated or suboptimally vaccinated populations, and an outbreak has even been reported in a highly vaccinated population.[2-4] In the United States, women born after the licensure of measles vaccine transfer less measles antibody to their infants than do older women, which may leave infants of vaccinated mothers with heightened susceptibility to measles.[5]

In undernourished children in developing countries, measles can cause a devastating illness, and case-fatality rates may reach or exceed 25%.[6] Before widespread vaccine use, epidemics occurred every 2 to 3 years in developed countries, and the illness developed in 95% of urban dwellers before they reached 15 years of age. Outbreaks occurred primarily in late winter and early spring. In 1989, the World Health Assembly instituted major efforts to reduce measles morbidity and mortality through implementation of control strategies. By 1996, estimated incidence and death rates were reduced by 78% and 88%, respectively.[1,7,8] In the Americas, between 1990 and 1996, the number of new measles cases decreased from 246,607 to 2,109 (a decrease of 99%), and vaccine coverage increased from 77% to 85%. In 2000, only 100 confirmed measles cases were reported in the United States, the lowest number on record.[1] Molecular epidemiologic studies suggest that most cases in the United States now result from importation of virus.[1,9,10]

PATHOGENESIS

The portals of entry for measles virus include cells of the respiratory tract and possibly the conjunctivae. After undergoing local replication and spreading to regional lymph nodes, measles virus is disseminated to distant sites, particularly skin and mucous membranes, by a leukocyte-associated viremia. Lesions on mucous membranes (Koplik spots) appear as bluish-white specks on an erythematous base. Histologically, Koplik spots are composed of epithelioid giant cells with cytoplasmic and nuclear inclusions that contain microtubular aggregates; inflammatory cells and intercellular edema are present. Lesions of the skin also demonstrate inclusions and microtubular aggregates, suggesting that active viral replication occurs in skin as well as in mucous membranes. Infectious virus can be isolated from leukocytes, urine, conjunctivae, and respiratory secretions. Antibody appears in the serum as viremia ceases. Leukopenia may accompany the illness, together with lymphocyte hyporesponsiveness.

DIAGNOSIS

Clinical Features

Approximately 9 to 11 days after a person is initially exposed to the virus, malaise, fever, conjunctivitis, photophobia, periorbital edema, coryza, and cough develop. Cough may be severe, although generally nonproductive, and temperature may reach 40.6° C (105.1° F). Within 2 to 3 days, Koplik spots may appear on the buccal mucosa and occasionally on the conjunctivae. The skin rash, which erupts 2 to 3 days later, usually appears at the hairline and spreads downward during the next 3 days as systemic symptoms subside [*see Figure 1*]. Lesion density is greatest above the shoulders, where macular lesions may coalesce. The rash lasts 4 to 6 days and then fades from the head downward. Desquamation may be present but is usually not severe. Complete recovery without scarring generally occurs within 7 to 10 days from the onset of the rash.

A severe atypical presentation of measles may appear in persons who were immunized between 1963 and 1967 with killed measles virus vaccine and who were later exposed to the wild virus. After a prodrome of fever, headache, abdominal pain, and myalgias, a rash develops on the hands and feet and advances toward the head. The eruption may be vesicular, urticarial, maculopapular, or hemorrhagic and is most prominent along body creases. Pneumonia, pleural effusion, and hilar lymphadenopathy are common in atypical measles. All persons vaccinated after 1967 received the live attenuated measles vaccine, which is rarely, if ever, associated with the atypical measles syndrome.

Laboratory Tests

In an epidemic setting, observation of the characteristic rash, fever, coryza, and conjunctivitis is sufficient to establish the diagnosis. However, as measles declines in prevalence, clinical acumen in diagnosis also diminishes. Multinucleated giant cells can often be detected in stained smears of nasal secretions, and measles antigen can be demonstrated in such cells by immunofluorescence. The virus can be isolated from nasal secretions or urine by cultivation of such materials on primate cell monolayers. A rise in hemagglutination inhibition antibodies during a period of 2 to 3 weeks confirms the diagnosis. Confirmation by measles-specific IgM enzyme immunoassay is also available.

DIFFERENTIAL DIAGNOSIS

Sporadic cases of measles must be differentiated from other viral exanthems, secondary syphilis, scarlet fever, and drug reactions. Atypical measles may resemble Rocky Mountain spotted fever, meningococcemia, or varicella.

COMPLICATIONS

Measles is usually benign and uncomplicated. Complications occur more commonly in adults, malnourished children, and immunocompromised patients. Measles is associated with severe pulmonary and neurologic complications in up to 80% of immunocompromised children and adults with cancer or HIV infection, with case-fatality rates of 40% to 70%; rash is absent in 30%.[11,12] Measles virus may further suppress host immune responses, leading to reactivation of latent tuberculosis or superimposition of new bacterial pneumonia, particularly in malnourished children.

Measles virus infection often involves the central nervous system, but clinically apparent encephalomyelitis is rare (one in every 1,000 to 2,000 measles patients). CNS involvement may precede the rash but usually begins 4 to 7 days after the eruption appears. The onset of encephalomyelitis is often precipitous and is characterized by a rise in fever, sudden mental deterioration, and seizures; motor defects and cerebellar ataxia are common. An autoimmune demyelinating process may be involved in the pathogenesis of measles encephalomyelitis, which is fatal in about 10% of patients.

Subacute measles encephalitis may complicate measles in immunocompromised hosts 1 to 7 months after exposure; the patient has seizures and altered mental status, but results of cerebrospinal fluid analysis are normal.[13] Diagnosis may require brain biopsy for histology, immunocytochemical analysis, or polymerase chain reaction (PCR) detection of viral RNA.[12] In rare instances, a subacute sclerosing panencephalitis occurs as a complication of measles in children infected before 2 years of age, but the panencephalitis develops after a latent period of several years. Other measles complications include thrombocytopenia with associated purpura or bleeding, myocarditis, pericarditis, hepatitis, and severe keratitis progressing to blindness.

TREATMENT AND PREVENTION

Therapy for measles is mainly supportive. A meta-analysis has suggested that in areas where vitamin A deficiency may be present, an oral dose of 200,000 IU of vitamin A for 2 days can reduce morbidity and mortality in children.[14] Ribavirin, either intravenous or aerosolized, has been proposed for certain complications of measles (i.e., encephalitis or pneumonitis),[11,12] but controlled trials are lacking.

Postexposure prophylaxis can be provided for high-risk patients (e.g., pregnant women and immunosuppressed children) by administering immunoglobulin intramuscularly at a dosage of 0.25 to 0.50 ml/kg body weight within 6 days after exposure.[15]

In vaccination guidelines in the United States, two immunizations are suggested for children before school enrollment to reduce the chance of primary vaccine failure. Vaccination is recommended for nonpregnant women of childbearing years and may be considered for institutionalized adults or individuals working in day care centers [*see CE:V Adult Preventive Health Care*]. Severely immunocompromised individuals should avoid measles vaccination because of the risk of vaccine-associated diseases, such as inclusion body encephalitis or pneumonitis.[16,17]

The effort to eradicate measles from the United States relies on identification and immunization of susceptible children, adolescents, and adults; on strict enforcement of more comprehensive school immunization requirements; and on strengthening of surveillance and outbreak-control measures.

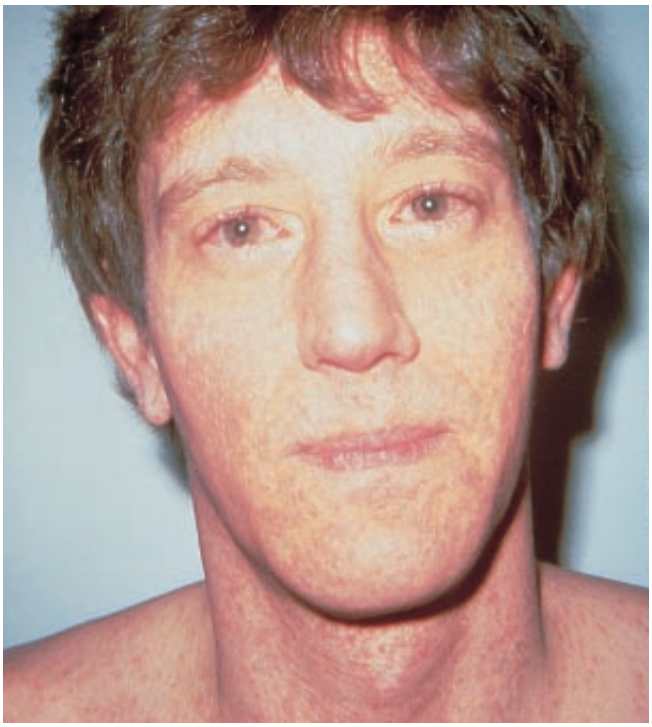


Figure 1 **A case of measles in a young man is characterized by maculopapular rash and conjunctivitis.**

Mumps

Mumps virus is a pleomorphic, enveloped, single-stranded RNA paramyxovirus capable of causing parotitis, epididymoorchitis, and CNS disease. Because of widespread vaccination efforts, its role as a major cause of childhood morbidity has greatly diminished.

EPIDEMIOLOGY

Mumps virus has worldwide distribution, and infection is seen more commonly in winter and spring. Humans are the only natural host. Infection is uncommon in persons younger than 1 year, and congenital infection is rare but can lead to fetal loss in the first trimester. Approximately 11% of cases of mumps are observed in children between 1 and 4 years of age, 52% in children between 5 and 14 years of age, and 11% in children older than 15 years.

A live mumps virus vaccine was approved for use in the United States in 1967; its use was facilitated by subsequent incorporation with measles and rubella (MMR) vaccine. Since its widespread use, mumps cases have fallen by over 99.5%, from 185,691 cases in 1968 to 906 cases in 1995.[15] Healthy People 2010, a comprehensive set of health objectives created under the auspices of the United States Department of Health and Human Services, has targeted mumps for elimination in the United States by 2010.[18]

PATHOGENESIS

Mumps virus spreads from person to person by direct contact with nasopharyngeal secretions. Virus is shed in saliva from as long as 6 days before to 5 days after clinical onset. Incubation ranges from 2 to 4 weeks (mean, 18 days). Initial replication occurs in the pharynx, followed by viremic dissemination. Both humoral and cell-mediated immune responses are induced and correlate with cessation of viremia and salivary virus excretion.

DIAGNOSIS

Clinical Features

Two thirds of cases are symptomatic, with initial symptoms of malaise and fever predominating. Painful swelling of the parotid gland is the characteristic feature of infection. It may be unilateral, and other salivary glands may be involved. However, parotitis occurs in only 30% to 40% of all cases of mumps virus infection.

An unvaccinated child who presents with tender parotitis generally has mumps; further diagnostic testing is not required. In older age groups, other entities (sarcoidosis, tumors, alcohol abuse, drug side effects, and other viral or bacterial infections) should be considered. In persons without parotitis who have orchitis, aseptic meningitis, encephalitis, or other obscure syndromes (e.g., myocarditis or pancreatitis), mumps should be considered [*see* Complications, *below*].

Laboratory Tests

Definitive diagnosis of mumps can be made by virus isolation from the oropharynx, CSF, or urine or by virus serology. Rapid detection by PCR techniques is now possible in some laboratories.[19]

COMPLICATIONS

Epididymo-orchitis occurs in up to 38% of postpubertal males with mumps and is usually unilateral. Subsequent sterility is uncommon, although testicular atrophy may occur.

CSF pleocytosis occurs in at least 50% of patients with mumps,[20] although symptomatic meningitis is less common. A lymphocytic pleocytosis is seen, with an elevated protein level and a normal to low (10%) glucose level. Symptomatic encephalitis occurs in one of 6,000 cases and presents as decreased consciousness and focal neurologic deficits. Most patients with encephalitis recover completely, although 0.5% to 2.3% of patients with mumps encephalitis may die. Other neurologic complications may include hydrocephalus; deafness; and rare cases of demyelinating disorders, transverse myelitis, Guillain-Barré syndrome, and cerebellar ataxia.[21] Deafness may be sudden, unilateral, and permanent.[15]

Pancreatitis, mastitis, and oophoritis have been observed in patients with mumps. Myocarditis occurs and in rare cases can be fatal.[22] Endocardial fibroelastosis, an infrequent sequela of myocarditis, has been associated with mumps infection.[23] Mumps can cause polyarticular or monoarticular arthritis; it generally affects adult males and is self-limited (usually less than 8 weeks).[24]

TREATMENT AND PREVENTION

Treatment of patients with mumps is largely supportive, although anti-inflammatory agents may be useful in cases of severe orchitis or arthritis. Administration of immune globulin does not prevent mumps and is not recommended. Prevention can be achieved in well over 90% of persons by the use of live attenuated mumps vaccine, administered twice as part of the recommended MMR vaccine regimen[15] [*see CE:V Adult Preventive Health Care*].

Rubella

Rubella, or German measles, is usually a benign febrile exanthem, but when it occurs in pregnant women, it can produce major congenital malformations. The infective agent is a single-stranded RNA virus of the Togaviridae family.

EPIDEMIOLOGY

Humans are the only known natural hosts for the rubella virus, which appears to be spread by respiratory droplets. The virus is moderately contagious but less so than measles. Before the introduction of a rubella vaccine, in 1969, epidemics occurred in the United States at 6- to 9-year intervals, predominantly in children younger than 15 years. Since the widespread use of rubella vaccine, the incidence of rubella has decreased by 99%.[25] Outbreaks continue to occur, primarily in young adults in hospitals, colleges, prisons, and prenatal clinics, but no major epidemics have occurred in the United States since 1964. Susceptibility rates in young adults vary, with the greatest proportion of cases now seen among persons of Hispanic origin.[25,26] Rubella outbreaks have recently been identified in processing plants for meat and poultry where many foreign-born workers are employed.[26] Less than 10 confirmed cases of congenital rubella syndrome are reported in the United States annually, almost all of them in infants of foreign-born mothers.[26]

PATHOGENESIS

Initially, the rubella virus replicates in the nasopharynx and regional lymph nodes. After the virus invades the bloodstream, it may spread to the skin and distal organs or, transplacentally, to the developing fetus. The virus may be present in throat washings or blood for several days before the appearance of the rash and up to 2 weeks after its onset. In rare cases of rubella arthropathy, the virus may persist in peripheral leukocytes or in synovial cells for months to years.

A pregnant woman infected with rubella is at risk for transmitting the virus to her fetus. Damage to the fetus is most likely to occur if the mother is infected within the first 2 months of gestation (fetal abnormalities are observed in 40% to 85% of such cases).[27,28] Infection within the third month is associated with fetal defects in 20% to 40% of cases, whereas infection during the fourth month is associated with fetal defects in 10% of cases. Mechanisms of fetal damage are not clear but may include viral cytolysis, chromosomal breaks, reduced cell multiplication, and alteration of fetal blood supply. As a result, fetal growth may be retarded, and defects may develop in multiple organ systems. Despite the production of fetal antibodies, the rubella virus can persist in the fetus and newborn and can be excreted for months to years after birth.

DIAGNOSIS

Clinical Features

After an incubation period of 12 to 23 days, a mild prodrome of malaise, headache, fever, and mild conjunctivitis may develop. Postauricular, suboccipital, and posterior cervical lymphadenopathy often precede the rash, which begins on the face and forehead. Within 1 to 5 days, the discrete maculopapular lesions spread over the trunk and extremities and may coalesce. The rash usually disappears within 3 days.

Laboratory Tests

The presence of rubella can be confirmed by virus isolation, by PCR detection, or by demonstration of seroconversion in response to rubella antigens. Virus isolation is often difficult because rubella virus does not cause cytopathic effects on the cell lines that are generally employed in diagnostic laboratories. Antibodies are often present shortly after rash appears and increase in titer during the next 2 to 3 weeks. Measurements of specific IgM antibodies to rubella virus are particularly useful in newborns: raised IgM levels denote recent infection and are specific

for fetal infection because IgM antibodies do not cross the placenta. Elevated IgM antibodies may return to nondiagnostic levels by 3 to 6 months, and persistence of IgG antibodies beyond this period may also help diagnose neonatal infection.[25]

DIFFERENTIAL DIAGNOSIS

Rubella may be confused with other viral exanthems (e.g., those caused by enteroviruses, parvoviruses, or adenoviruses), scarlet fever, or drug eruptions.

COMPLICATIONS

The most common complications of rubella are arthropathies of the fingers, wrists, and knees; they occur predominantly in young women. Such arthropathies consist of arthralgia or frank arthritis, and recurring joint symptoms may persist for a year or more. Encephalitis and thrombocytopenia are rare complications of acute rubella. Encephalitis occurs in one in 6,000 cases,[15] and thrombocytopenia, which is sometimes associated with purpura or hemorrhage, occurs most often in children.

Congenital rubella syndrome, also a rare complication of acute rubella, may manifest itself as defects in one or many organ systems. Hearing impairment is the most common single defect (60%).[25] Heart malformations, particularly patent ductus arteriosus and peripheral pulmonic stenosis, are also common (45%), as are cataracts (25%), microcephaly (27%), and mental retardation (13%). Malformation of bone metaphyses may also be present, together with hepatosplenomegaly, thrombocytopenia, interstitial pneumonitis, myocarditis, and thrombocytopenic purpura. Congenitally infected infants are often of low birth weight (23%), and they excrete rubella virus for prolonged periods. Late complications may result from imbalances of cellular and humoral immunity, from immune complex deposition, or from prolonged viral replication. Diabetes mellitus and other endocrine abnormalities may be late complications of congenital rubella, as is subacute sclerosing panencephalitis.[29]

TREATMENT AND PREVENTION

Beyond the fetal period, rubella is mild and self-limited. Current treatment of congenital rubella syndrome is only supportive. The Centers for Disease Control and Prevention (CDC) has developed detailed recommendations for dealing with rubella outbreaks focused on patient isolation, identification and vaccination of susceptible persons who have no contraindications to rubella vaccine, and counseling of susceptible pregnant women.[26] These recommendations are available on the Internet at http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5012a1.htm.

Pregnant women infected with rubella virus who are asymptomatic may still transmit rubella to their fetuses. Thus, testing of immune status is advisable for women of childbearing age and for hospital employees who have no history of rubella vaccination. About 10% to 15% of such persons are seronegative and should be vaccinated when not pregnant. Women should avoid becoming pregnant for 28 days after vaccination against rubella.[30]

Measles, Mumps, and Rubella Vaccine

Because they fear possible complications from MMR vaccination, some parents question the safety of the vaccine, and in some cases, they are refusing its use in their children. These parents may cite research suggesting that MMR vaccination may be a risk factor for inflammatory bowel disease[31] and autism.[32] Subsequent research, however, has not supported this hypothesis.[33,34] In addition, although receipt of the MMR vaccine has been associated with an increased risk of febrile seizures 8 to 14 days after vaccination, those children were not found to be at higher risk for subsequent seizures or neurodevelopmental disabilities.[35]

Parvovirus

Although only recently recognized as a human pathogen, parvovirus B19 is now appreciated as a cause of several syndromes in both children and adults. Parvovirus B19, a small (20 to 26 nm), single-stranded DNA virus, causes erythema infectiosum (fifth disease) in normal persons, aplastic crises in persons with underlying hemolytic disorders, chronic anemia in immunocompromised hosts, and fetal loss in pregnant women.[36-40]

EPIDEMIOLOGY

Parvovirus B19 infection occurs most commonly in school-age children in outbreaks during late winter and spring. Only 2% to 15% of pre–school-age children have antibodies, but seroprevalence increases to 35% to 60% by 11 to 19 years of age and to greater than 75% in persons older than 50 years.[41] Respiratory transmission is likely and is facilitated by close contact. Hospital outbreaks have also been described and are often traced to patients with aplastic crises who carry large amounts of virus in blood and respiratory secretions.[42,43] Maternal infection can lead to fetal anemia, hydrops fetalis, heart failure, and death, resulting in spontaneous abortion, most commonly 4 to 6 weeks after infection. When women are infected during the first 20 weeks of pregnancy, the risk of parvovirus-related fetal death is approximately 9% to 10%.[44,45] Routine antenatal screening is not recommended.[38,46]

PATHOGENESIS

Replication of parvovirus B19 has been demonstrated in human erythroid progenitor cells, and the receptor appears to be the P blood group antigen globoside, a neutral glycosphingolipid, which occurs in erythrocytes, erythroblasts, megakaryocytes, endothelial cells, placenta, and fetal liver and heart cells.[47] Expression of this glycosphingolipid in tissues helps to determine parvovirus B19 tropism.[48] Persons who lack erythrocyte P antigen (p phenotype) are naturally resistant to infection,[37,49] and the distribution of parvovirus in infected individuals is linked to the presence of the P antigen. Although little is known about the pathogenesis of parvovirus, antiviral antibodies—particularly those directed against the capsid protein VP1—appear to be responsible for viral clearance. The presence of certain HLA class I and class II alleles may be associated with more symptomatic parvovirus infections.[50]

DIAGNOSIS

Clinical Features

The rash caused by parvovirus B19, erythema infectiosum, usually appears without prodromal symptoms after an incubation period of 4 to 14 days. The exanthem progresses through three stages. Initially, a fiery-red rash develops on both cheeks (giving them the appearance of having been slapped), accompanied by relative pallor around the mouth. From 1 to 4 days later, an erythematous maculopapular eruption appears on the proximal extremities and spreads to the trunk in a lacelike, reticular pattern. The third stage, during which the eruption waxes and wanes, may persist for several weeks and may be precipitated by skin trauma, exposure to sunlight, or extremes of tempera-

ture. Arthralgia and arthritis are seen in up to 80% of infected adults; arthralgia is particularly common in women, may occur without rash, and may linger for weeks. Joint involvement is often symmetrical in the hands, wrists, knees, and ankles. Hemolytic anemias and encephalopathies are rare complications.

Laboratory Tests

Parvovirus-specific IgM antibodies usually appear within 3 days after symptoms develop; these antibodies persist for several weeks and then rapidly decline. IgG antibodies, however, persist for years. Viral DNA can also be detected in blood, tissues, and secretions, although culture techniques for virus isolation are unsatisfactory.

COMPLICATIONS

Transient aplastic crises associated with parvovirus B19 occur in patients who have sickle cell anemia, hereditary spherocytosis, thalassemia, and various other hemolytic anemias.[51] These aplastic crises are abrupt in onset and associated with giant pronormoblasts in the bone marrow. They generally last 1 to 2 weeks and go into remission spontaneously. In immunocompromised hosts (e.g., patients with HIV infection), acute infection may lead to viral persistence and chronic bone marrow suppression.[40] A significant proportion of patients with AIDS who develop severe anemia while receiving zidovudine (AZT) have persistent parvovirus infection.[51] Pneumonia, hepatitis, and myocarditis have also been associated with parvovirus infections in immunocompromised as well as immunocompetent adults and children.[52-57] Although parvovirus B19 has been implicated in a variety of rheumatic diseases, there is no definitive evidence for a causal role.

TREATMENT

Pooled human immune globulin contains anti–parvovirus B19 antibodies and has been used to treat persistent infections as well as acute exposures.[40] Prevention of nosocomial infections is of great concern: pregnant health care workers should not care for patients with aplastic crises. Droplet isolation is recommended for such patients, including the use of gowns, gloves, and masks during close contact. Because certain blood products (e.g., clotting factors) contain parvovirus B19 DNA, screening of products, donors, and recipients has been suggested.[58]

Poxvirus Infections

Poxviruses are the largest (200 to 320 nm) and most complex human viruses. They replicate in cell cytoplasm and may produce eosinophilic cytoplasmic inclusion bodies. They preferentially infect skin epithelial cells and may cause a variety of human diseases. Smallpox (variola), once among the most devastating and feared worldwide pestilences, has been virtually eliminated. Other human poxvirus diseases include vaccinia, molluscum contagiosum, orf (contagious pustular dermatitis), and paravaccinia (milker's nodules).

SMALLPOX

No naturally acquired cases of smallpox (variola) have been observed since 1977, as a result of a global eradication effort that was initiated by the World Health Organization in the 1960s.[59] Several biologic features of the smallpox virus favored its eradication: only one serotype existed, a stable and effective vaccine had been developed, and there were no nonhuman reservoirs and no human carriers of the smallpox virus. For a time, smallpox was considered a disease of purely historical interest. However, laboratory stocks of the virus were never totally destroyed, and the possibility of dissemination of the stores held in the former Soviet Union has raised concerns that smallpox virus might be used as an agent of biological warfare [*see 8:V Bioterrorism*].[60]

Table 1 Contraindications to Nonemergency Smallpox Vaccination

Conditions in the Patient or a Household Member
Eczema or atopic dermatitis (even if it is currently inactive or mild or was experienced in childhood)
Skin conditions such as burns, chickenpox, shingles, impetigo, herpes, severe acne, or psoriasis
Weakened immune system (e.g., from corticosteroid treatment, cancer chemotherapy, posttransplantation immunosuppression, HIV infection, or severe autoimmune disorders)
Pregnancy or intent to become pregnant within 1 mo after vaccination
Conditions in the Patient
Allergy to the vaccine or any of its ingredients
Age less than 12 mo*
Moderate or severe short-term illness
Current breast-feeding
Use of steroid eyedrops

Note: Persons who have been directly exposed to the smallpox virus should be vaccinated, regardless of their health status. More information on smallpox vaccination is available at www.cdc.gov/smallpox.
*The Advisory Committee on Immunization Practices (ACIP) of the Centers for Disease Control and Prevention advises against nonemergency use of smallpox vaccine in children younger than 18 yr, and the vaccine manufacturer does not recommend nonemergency use of the vaccine in geriatric patients.

Pathogenesis

Variola is spread by the respiratory route. Mucosal seeding is followed by spread to lymph nodes, a brief viremia, and spread to the reticuloendothelial system over the next 4 to 14 days. Thereafter a second viremia leads to infection of skin and mucous membranes. Neutralizing antibodies appear during the first week of infection and persist for years.[61] Cytotoxic T lymphocyte reactivity also occurs early and persists.[62]

Diagnosis

Clinical features A prodrome characterized by fever, headache, and backache lasts for 2 to 3 days, after which an oropharyngeal enanthem appears. Smallpox can take several clinical forms [*see 8:V Bioterrorism*]. In its most common form (ordinary smallpox), the disease causes a rash that goes through several stages: papules to vesicles to pustules to crusts over 1 to 2 weeks. Lesions begin on the face and extremities but spread all over the body. Differential diagnosis involves a variety of viral diseases, most notably varicella (chickenpox). The principal clinical differences between smallpox and chickenpox are that, in smallpox, lesions are all in the same stage of development and tend to cluster on the face and extremities (including the palms of the hands and soles of the feet), whereas, in chickenpox, lesions tend to be in various stages of development and to cluster on the torso [*see 8:V Bioterrorism*]. Smallpox must also be differentiated from drug-induced rashes, most notably Stevens-Johnson syndrome.

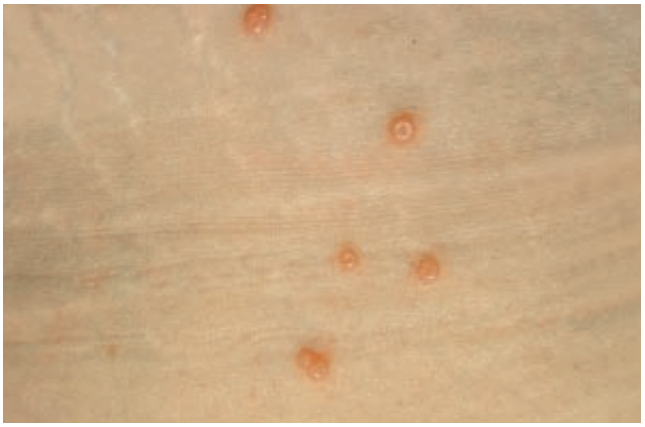


Figure 2 **The pearly white papular lesions of molluscum contagiosum shown here are 2 to 5 cm in diameter, with a central umbilication.**

Complications Visceral involvement can lead to encephalitis (< 1%), arthritis (2%), hypotension, hemorrhage, pneumonia, and death. Case-fatality rates of up to 30% have been reported, often associated with secondary bacteremias.[61]

Laboratory tests If facilities are available, direct examination of specimens by electron microscopy can readily distinguish variola from other viruses. Antigens can be detected by immunohistochemical techniques, DNA can be studied by PCR, and variola virus can be isolated in cell cultures.

Treatment

Strict respiratory and contact isolation is essential, preferably in a room with negative air pressure. Hydration is critical because of fluid losses. Vaccination is recommended in patients with early disease (see below).[61] Cidofovir has activity against related viruses (vaccinia, cowpox, monkeypox) in animal models and may be tried in severe cases, though it has considerable nephrotoxicity and no proven benefit.

Prevention

Smallpox prevention consists of infection control measures and use of vaccine containing live vaccinia virus.[61-63] The exact origin of vaccinia virus is not clear (the virus has no known natural hosts), but vaccinia has long been used as the source of smallpox vaccines. Although the duration of benefit of vaccinia vaccine has never been measured in controlled trials, epidemiologic studies suggest that increased protection against smallpox may persist for more than 10 years after vaccination.[64]

In healthy patients, injection of vaccinia virus usually induces a localized papular eruption at the injection site. However, patients with compromised immune function or with skin conditions such as eczema may experience more severe disease after vaccination. Progressive generalized vaccinia, vaccinia gangrenosa, and eczema vaccinatum may complicate such disorders [*see 8:V Bioterrorism*]. Vaccinia immune globulin and antiviral agents (e.g., cidofovir) have been suggested as possible therapies for vaccinia complications, but their effectiveness has not been established.

Since 1971, routine smallpox vaccination has not been recommended in the United States. Despite concern regarding the use of smallpox as a bioterrorism weapon, vaccination of the general public is still not recommended. Instead, targeted vaccination may be a more effective intervention against bioterrorist smallpox.[65] In October 2002, the CDC's Advisory Committee on Immunization Practices (ACIP) recommended voluntary vaccination of people designated to respond to or care for individuals suspected or confirmed of being infected with smallpox.[66] Because of the risk of complications, nonemergency vaccination is contraindicated in some cases [*see Table 1*]. Recently vaccinated persons may transmit vaccinia through contact with susceptible persons, so infection control measures are important. Current information and recommendations on smallpox and smallpox vaccination are available from the CDC at http://www.bt.cdc.gov/agent/smallpox/reference/resource-kit.asp.

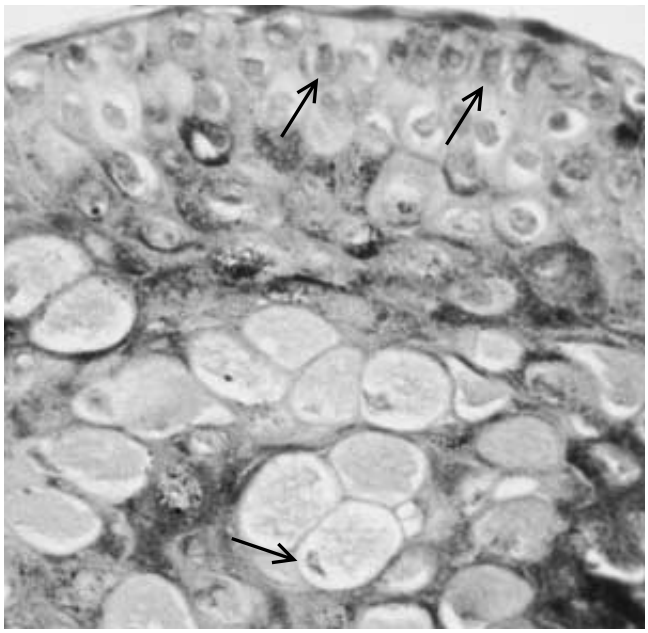


Figure 3 **Micrograph of a molluscum contagiosum lesion demonstrates basophilic molluscum bodies (arrow) in epidermal cell cytoplasm.**

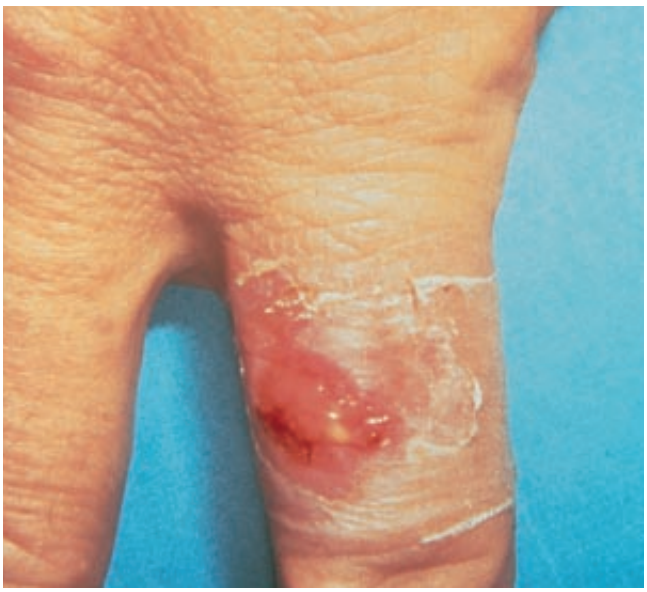


Figure 4 **Papillomatous lesions of orf (pustular dermatitis) can be observed on the finger of a sheep handler.**

MOLLUSCUM CONTAGIOSUM

Molluscum contagiosum is characterized by multiple painless, pearly white nodules 2 to 5 mm in diameter with a central umbilication. They can appear anywhere on the body except the palms and soles [*see Figure 2*]. The nodules, which are most commonly found in anogenital regions, rupture easily and may be spread by sexual routes, by autoinoculation, or by close familial contact under conditions of poor hygiene. Cases occur predominantly in children, sexually active adults, sports participants who have skin-to-skin contact, and persons with impaired cellular immunity.[67] The infection has worldwide distribution, and incidence rates of clinically apparent infection range from 0.1% to 4.5%. Incubation periods vary from several days to several weeks, and lesions may clear rapidly or persist for up to 18 months. Molluscum contagiosum is common in patients with AIDS, in whom the lesions may be large, atypical, and severe[68] [*see 2:I Cutaneous Manifestations of Systemic Diseases*]. Lesions near the eye may be complicated by chronic conjunctivitis or superficial keratitis.

The molluscum contagiosum virus (MCV) has been visualized by electron microscopy but has not been cultivated in vitro. Microscopic observation of large cytoplasmic inclusions, called molluscum bodies [*see Figure 3*], in appropriately stained, expressed lesion contents or histologic sections confirms a clinical diagnosis. Restriction endonuclease cleavage patterns of DNA from purified virus obtained from skin lesions indicate that there are two distinct MCV genotypes.[69]

The lesions resolve spontaneously without scarring. A small number of lesions can be removed by gentle curettage, laser, or caustic chemicals if desired. In immunocompetent patients, successful topical treatment with imiquimod 5% cream has recently been reported.[70]

PARAVACCINIA, ORF, AND MONKEYPOX

Human paravaccinia, orf, and monkeypox infections result from direct contact with natural animal reservoirs of these agents; humans are only incidental hosts.

Paravaccinia is an infection that produces lesions on the teats and oral mucosa of calves and milk cows. When humans are infected by direct contact, so-called milker's nodules develop on the fingers or hands and occasionally are associated with lymphadenitis. Lesions develop over a period of 1 to 2 weeks and resolve in 3 to 8 weeks.

The orf virus causes papillomatous lesions (pustular dermatitis) on the mucous membranes and corneas of sheep and goats. Lesions in humans [*see Figure 4*] are caused by direct contact with infected animals and resemble those caused by paravaccinia, although paravaccinia and orf viruses are distinct. Most cases are benign, but immunocompromised patients have been successfully treated with cidofovir.[71]

Monkeypox is caused by an orthopoxvirus related to the smallpox virus. Vaccination against smallpox also protects against monkeypox. Human monkeypox infections occur sporadically in small villages in African tropical rain forests; monkeypox infections also occur in captive monkeys in European and North American laboratories. Smallpoxlike diseases caused by monkeypox virus have been noted in central Africa in humans who live in close proximity to monkeys.[72] Although the primary reservoir for the monkeypox virus remains uncertain, humans appear to be incidental hosts for the virus. Person-to-person transmission has been described. In animal models, cidofovir has been effective for treating monkeypox infections.[71]

The author has served as a consultant for Bristol-Myers Squibb Company, GlaxoSmithKline, Millennium Pharmaceuticals, Inc., Schering-Plough Corporation, and Takeda Pharmaceuticals North America, Inc.

References

1. Measles—United States, 2000. MMWR Morb Mortal Wkly Rep 51:120, 2002

2. Coughlan S, Connell J, Cohen B, et al: Suboptimal measles-mumps-rubella vaccination coverage facilitates an imported measles outbreak in Ireland. Clin Infect Dis 35:84, 2002

3. Measles outbreak—Netherlands, April 1999–January 2000. MMWR Morb Mortal Wkly Rep 49:299, 2000

4. Transmission of measles among a highly vaccinated school population—Anchorage, Alaska, 1998. MMWR Morb Mortal Wkly Rep 47:1109, 1999

5. Papania M, Baughman AL, Lee S, et al: Increased susceptibility to measles in infants in the United States. Pediatrics 104:e59, 1999

6. Dollimore N, Cutts F, Binka FN: Measles incidence, case fatality, and delayed mortality in children with or without vitamin A supplementation in rural Ghana. Am J Epidemiol 146:646, 1997

7. Progress toward global measles control and elimination, 1990–1996. MMWR Morb Mortal Wkly Rep 46:893, 1997

8. de Quadros CA, Olivé JM, Hersh BS, et al: Measles elimination in the Americas: evolving strategies. JAMA 275:224, 1996

9. Rota JS, Rota PA, Redd SB, et al: Genetic analysis of measles viruses isolated in the United States, 1995–1996. J Infect Dis 177:204, 1998

10. Bellini WJ, Rota PA: Genetic diversity of wild-type measles viruses: implications for global measles elimination programs. Emerging Infect Dis 4:29, 1998

11. Kaplan LJ, Daum RS, Smaron M, et al: Severe measles in immunocompromised patients. JAMA 267:1237, 1992

12. Mustafa MM, Weitman SD, Winick NJ, et al: Subacute measles encephalitis in the young immunocompromised host: report of two cases diagnosed by polymerase chain reaction and treated with ribavirin and review of the literature. Clin Infect Dis 16:654, 1993

13. Budka H, Urbanits S, Liberski PP, et al: Subacute measles virus encephalitis: a new and fatal opportunistic infection in a patient with AIDS. Neurology 46:586, 1996

14. D'Souza RM, D'Souza R: Vitamin A for treating measles in children. Cochrane Database Syst Rev (1):CD001479, 2002

15. Measles, mumps and rubella–vaccine use and strategies for elimination of measles, rubella and congenital rubella syndrome and control of mumps. MMWR Recomm Rep 47(RR8):1, 1998

16. Bitnun A, Shannon P, Durward A, et al: Measles inclusion–body encephalitis caused by the vaccine strain of measles virus. Clin Infect Dis 29:855, 1999

17. Angel JB, Walpita P, Lerch RA, et al: Vaccine-associated measles pneumonitis in an adult with AIDS. Ann Intern Med 129:104, 1998

18. Healthy People 2010. U.S. Department of Health and Human Services. International Medical Publishing, Inc., McLean, Virginia, 2001

http://www.healthypeople.gov/document/html/objectives/14-01.htm

19. Poggio GP, Rodriguez C, Cisterna D, et al: Nested PCR for rapid detection of mumps virus in cerebrospinal fluid from patients with neurological diseases. J Clin Microbiol 38:274, 200**0**

20. Hammer SM, Connolly KJ: Viral aseptic meningitis in the United States: clinical features, viral etiologies, and differential diagnosis. Curr Clin Top Infect Dis 12:1, 1992

21. Nussinovitch M, Volovitz B, Varsano I: Complications of mumps requiring hospitalization in children. Eur J Pediatr 154:732, 1995

22. Baandrup U, Mortensen SA: Fatal mumps myocarditis. Acta Med Scand 216:331, 1984

23. Ni J, Bowles NE, Kim YH, et al: Viral infection of the myocardium in endocardial fibroelastosis: molecular evidence for the role of mumps virus as an etiologic agent. Circulation 95:133, 1997

24. Gordon SC, Lauter CB: Mumps arthritis: a review of the literature. Rev Infect Dis 6:338, 1984

25. Reef SE, Plotkin S, Cordero JF, et al: Preparing for elimination of congenital rubella syndrome (CRS): summary of a workshop on CRS elimination in the United States. Clin Infect Dis 31:85, 2000

26. Control and prevention of rubella: evaluation and management of suspected outbreaks, rubella in pregnant women, and surveillance for congenital rubella syndrome. MMWR Recomm Rep 50(RR12): 1, 2001

27. Miller E, Craddock-Watson JE, Pollock TM: Consequences of confirmed maternal rubella at successive stages of pregnancy. Lancet 2:781, 1982

28. Munro ND, Sheppard S, Smithells RW, et al: Temporal relations between maternal rubella and congenital defects. Lancet 2:201, 1987

29. Townsend JJ, Baringer JR, Wolinsky JS, et al: Progressive rubella panencephalitis: late onset after congenital rubella. N Engl J Med 292:990, 1975

30. Revised ACIP recommendation for avoiding pregnancy after receiving a rubella-containing vaccine. MMWR Morb Mortal Wkly Rep 50(49):1117, 2001

31. Thompson NP, Montgomery SM, Pounder RE, et al: Is measles vaccination a risk factor for inflammatory bowel disease? Lancet 345:1071, 1995

32. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Inflammatory Bowel Disease Study Group. Lancet 351:637, 1998

33. Taylor B, Miller E, Lingam R, et al: Measles, mumps, and rubella vaccination and bowel problems or developmental regression in children with autism: population study. BMJ 324:393, 2002

34. Madsen KM, Hviid A, Vestergaard M, et al: A population-based study of measles, mumps, and rubella vaccination and autism. N Engl J Med 347:1477, 2002

35. Barlow WE, Davis RL, Glasser JW, et al: The risk of seizures after receipt of whole-cell pertussis or measles, mumps, and rubella vaccine. N Engl J Med 345:656, 2001

36. Katta R: Parvovirus B19: a review. Dermatol Clin 20:333, 2002

37. Carper E, Kurtzman GJ: Human parvovirus B19 infection. Curr Opin Hematol 3:111, 1996

38. Parvovirus, erythema infectiosum, and pregnancy. American Academy of Pediatrics Committee on Infectious Diseases. Pediatrics 85:131, 1990

39. Torok TJ: Human parvovirus B19 infection in pregnancy. Pediatr Infect Dis J 9:772, 1990

40. Frickhofen N, Abkowitz JL, Safford M, et al: Persistent B19 parvovirus infection in patients infected with human immunodeficiency virus type 1 (HIV-1): a treatable cause of anemia in AIDS. Ann Intern Med 113:926, 1990

41. Lin KH, You SL, Chen CJ, et al: Seroepidemiology of human parvovirus B19 in Taiwan. J Med Virol 57:169, 1999

42. Bell LM, Naides SJ, Stoffman P, et al: Human parvovirus B19 infection among hospital staff members after contact with infected patients. N Engl J Med 321:485, 1989

43. Pillay D, Patou G, Hurt S, et al: Parvovirus B19 outbreak in a children's ward. Lancet 339:107, 1992

44. Miller E, Fairley C, Cohen BJ, et al: Immediate and long-term outcome of human parvovirus B19 infection during pregnancy. Br J Obstet Gynaecol 105:174, 1998

45. Markenson GR, Yancey ML: Parvovirus B19 infections in pregnancy. Semin Perinatol 22:309, 1998

46. Guidozzi F, Ballot D, Rothberg AD: Human B19 parvovirus infection in an obstetric population: a prospective study determining fetal outcome. J Reprod Med 39:36, 1994

47. Brown KE, Anderson SM, Young NS: Erythrocyte P antigen: cellular receptor for B19 parvovirus. Science 262:114, 1993

48. Coding LLW, Koerner TAW, Naides SJ: Multiple glycosphingolipids determine the tissue tropism of parvovirus B19. J Infect Dis 172:1198, 1995

49. Brown KE, Hibbs JR, Gallinella G, et al: Resistance to parvovirus B19 infection due to lack of virus receptor (erythrocyte P antigen). N Engl J Med 330:1192, 1994

50. Kerr JR, Mattey DL, Thomson W, et al: Association of symptomatic acute human parvovirus B19 infection with human leukocyte antigen class I and II alleles. J Infect Dis 186:447, 2002

51. Naides SJ, Howard EJ, Swack NS, et al: Parvovirus B19 infection in human immunodeficiency virus type 1–infected persons failing or intolerant to zidovudine therapy. J Infect Dis 168:101, 1993

52. Janner D, Bork J, Baum M, et al: Severe pneumonia after heart transplantation as a result of human parvovirus B19. J Heart Lung Transplant 13:336, 1994

53. Hillingso JG, Jensen IP, Tom-Peterson L: Parvovirus B19 and acute hepatitis in adults. Lancet 351:955, 1998

54. Sokal EM, Melchioir M, Cornu C, et al: Acute parvovirus B19 infection associated with fulminant hepatitis of favourable prognosis in young children. Lancet 352:1739, 1998

55. Enders G, Dotsch J, Bauer J, et al: Life-threatening parvovirus B19-associated myocarditis and cardiac transplantation as possible therapy: two case reports. Clin Infect Dis 26:355, 1998

56. Nigro G, Bastianon V, Colloridi V, et al: Human parvovirus B19 infection in infancy associated with acute and chronic lymphocytic myocarditis and high cytokine levels: reports of 3 cases and review. Clin Infect Dis 31:65, 2000

57. Morris CN, Smilack JD: Parvovirus B19 infection associated with respiratory distress. Clin Infect Dis 27:900, 1998

58. Lefrère J-J, Mariotti M, Thauvin M: B19 parvovirus DNA in solvent/detergent-treated antihaemophilia concentrates. Lancet 343:211, 1994

59. Barquet N, Domingo P: Smallpox: the triumph over the most terrible of the ministers of death. Ann Intern Med 127:635, 1997

60. Institute of Medicine: Assessment of future scientific needs for live variola virus. Report of a committee on the assessment of future scientific needs for live variola virus. National Academy Press, Washington, DC, 1999

61. Breman JG, Henderson DA: Diagnosis and management of smallpox. N Engl J Med 346:1300, 2002

62. Frey SE, Newman FK, Cruz J, et al: Dose-related effects of smallpox vaccine. N Engl J Med 346:1275, 2002

63. Frey SE, Couch RB, Tacket CO, et al: Clinical responses to undiluted and diluted smallpox vaccine. N Engl J Med 346:1265, 2002

64. Vaccinia (smallpox) vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP), 2001. MMWR Morb Mortal Wkly Rep 50(RR-10):1, 2001

65. Halloran ME, Longini IM Jr, Nizam A, et al: Containing bioterrorist smallpox. Science 298:1428, 2002

66. Recommendations for using smallpox vaccine in a pre-event vaccination program: supplemental recommendations of the Advisory Committee on Immunization Practices (ACIP) and the Healthcare Infection Control Practices Advisory Committee (HICPAC). MMWR Dispatch 52:1, 2003

67. Smith KJ, Skelton H: Molluscum contagiosum: recent advances in pathogenic mechanisms, and new therapies. Am J Clin Dermatol 3:535, 2002

68. Schwartz MJ, Myskowski PL: Molluscum contagiosum in patients with human immunodeficiency virus infection: a review of twenty-seven patients. J Am Acad Dermatol 27:583, 1992

69. Scholz J, Rosen-Wolff A, Bugert J, et al: Molecular epidemiology of molluscum contagiosum. J Infect Dis 158:898, 1988

70. Skinner RB Jr: Treatment of molluscum contagiosum with imiquimod 5% cream. J Am Acad Dermatol 47(4 suppl):S221, 2002

71. DeClercq E: Cidofovir in the therapy and short-term prophylaxis of poxvirus infections.Trends Pharmacol Sci 23:456, 2002

72. Hutin YJF, Williams RJ, Malfait P, et al: Outbreak of human monkeypox, Democratic Republic of Congo, 1996-1997. Emerg Infect Dis 7:434, 2001

Acknowledgments

Figures 1 and 4 Photographs courtesy of Dr. Howard Baden, Department of Dermatology, Massachusetts General Hospital, and Harvard Medical School, Boston.

Figure 2 Photograph courtesy of the American Academy of Dermatology.

Figure 3 Micrograph courtesy of Dr. George W. Hambrick, Department of Dermatology, New York Hospital, New York.

XXX SEPSIS

Steven M. Opal, m.d.
Christian E. Huber, m.d.

Sepsis, along with the multiorgan failure that often accompanies the systemic inflammatory response syndrome (SIRS), is a leading cause of mortality in the intensive care unit.[1,2] As many as 700,000 patients develop sepsis annually in the United States.[3] Nearly half of these patients manifest severe sepsis and septic shock. The mortality for septic shock remains approximately 35% to 45%, despite a concerted effort to improve the treatment options and outcome.[1-3]

Septic shock has become a major focus of critical care research. Although modest improvements in the prognosis have been made over the past 2 decades and promising new therapies have appeared in the past few years, innovations in the management of septic shock are still required.[4] This chapter reviews some of the remarkable advances achieved in the understanding of the molecular pathophysiology of sepsis, the diagnostic and therapeutic strategies emerging from this research, and the current management of septic shock.

Definitions of Sepsis

Sepsis, septic shock, SIRS, and multiple organ dysfunction syndrome (MODS) were defined at the American College of Chest Physicians/Society for Critical Care Medicine (ACCP/SCCM) Consensus Conference on Definition [*see Table 1*].[5] The definitions take into account the finding that sepsis may result from a multitude of infectious agents and microbial mediators and may or may not be associated with actual bloodstream infection. Despite the clinical logic, intrinsic simplicity, and widespread acceptance of these consensus definitions, their clinical applicability has been justifiably criticized.[6,7] The SIRS definition is so broad and nonspecific that it lacks discriminatory power; many patients admitted to general medical services and most ICU patients have conditions that meet the definition of SIRS. The current definition of sepsis fits virtually every person within the first 24 hours of an episode of influenza because the SIRS criteria are met and an infectious agent (an orthomyxovirus) causes the syndrome; however, clinicians do not generally regard the flu as sepsis. Nevertheless, a consensus conference held in 2001 to address some of these concerns concluded that, apart from a need to expand the list of signs and symptoms of sepsis to reflect clinical bedside experience, no evidence existed to support a change in the definitions.[8]

One of the tenets on which these definitions are based is that the inflammatory response itself, not the infectious organism, drives the septic process. This hypothesis may be largely correct, but the nature of the microbial pathogen responsible for sepsis clearly contributes to the ultimate fate of the patient. Microbial pathogens differ in their susceptibility to host defenses, their potential for developing antimicrobial resistance, and their ability to generate toxins—all of which affect their pathogenicity.[9] Failure to account for these intrinsic differences in microbial virulence limits the utility of current sepsis definitions.

Although the definition of septic shock includes hypotension that is unresponsive to fluid challenge, there is disagreement regarding the level of fluid resuscitation necessary to distinguish between sepsis and hypovolemia. In addition, the amount of vasopressor agent necessary to allow one to confidently conclude that the patient has true septic shock continues to be debated. Controversy also surrounds the difficulty in distinguishing preexisting morbidities and organ dysfunction from morbidity and

Table 1 The Terminology of Sepsis

Term	*Definition*	*Comments*
ARDS* (acute respiratory distress syndrome)[106]	Acute onset Bilateral infiltrates on chest radiograph Hypoxemia (P_aO_2/F_IO_2 ratio ≤ 200 mm Hg No evidence of left atrial hypertension (pulmonary arterial wedge pressure ≤ 18 mm Hg)	Severity can be scored by use of clinical parameters such as presence of hypoxemia, radiologic evidence of lung consolidation and compliance, and PEEP data from mechanical ventilation (Murray lung injury score)[107]
Bacteremia	Detection of viable bacteria in the bloodstream	Transient bacteremia without clinical symptoms is common; may or may not be found in sepsis
SIRS (systemic inflammatory response syndrome)	Temperature > 38.5° C (101.3° F) or < 36° C (96.8° F) Tachypnea (> 20 breaths/minute) Tachycardia (> 90 beats/minute) White blood cell count > 12,000 cells/mm^3 or > 10% immature forms or < 4,000 cells/mm^3	Two or more criteria needed; may be caused by infectious and noninfectious etiologies; clinical features may be caused by release of inflammatory mediators into circulation
Sepsis	SIRS caused by an invasive infection	May be caused by viral, bacterial, fungal, or parasitic pathogens; bloodstream infection need not be present
MODS (multiple organ dysfunction syndrome)	Major organ dysfunction from sepsis	A primary determinant of outcome in sepsis
Severe sepsis	Sepsis accompanied by major organ dysfunction (CNS, renal, pulmonary, hepatobiliary, hematologic, or metabolic)	Sometimes referred to as sepsis syndrome
Septic shock	Severe sepsis with hypotension not responsive to fluid challenge	Systolic blood pressure < 90 mm Hg despite adequate fluid resuscitation

*European-American Consensus Committee on ARDS definition.
F_IO_2—fraction of inspired oxygen P_aO_2—arterial oxygen tension PEEP—positive end-expiratory pressure

organ dysfunction induced by the septic process itself. Many patients with sepsis have major underlying organ dysfunction from a variety of disease entities. The degree to which sepsis contributes to further disordered organ function may be difficult to determine with accuracy. The same can be said for the degree to which sepsis contributes to the mortality in patients who suffer from other serious underlying diseases. All these factors limit the discriminatory value of the consensus definitions of sepsis and jeopardize valid comparisons between different study populations, compromising the ability to pool data and generalize the findings.

Further refinements in sepsis terminology may be possible when rapid diagnostic techniques become available to assess the immune status of septic patients. Functional genomics and proteomics (the study of human gene sequences and protein sequences, respectively) may assist in characterizing septic patients in the future. In the meantime, the current consensus definitions will be used despite their limitations.

Epidemiology

Between 1979 and 2000, the incidence of sepsis increased by 8.7% annually, from 82.7 to 240.4 per 100,000 population.[3,10] This trend will probably continue in the foreseeable future because sepsis has largely become a disease of medical progress. Successful management of a variety of severe medical and surgical diseases has produced a large patient population with critical illness and impaired host defenses; these patients have a greatly increased risk of developing sepsis. Innovations in organ transplantation, implanted prosthetic devices, and long-term vascular access devices continue to expand this patient population. The gradual aging of the population in many developed countries and the increasing prevalence of antibiotic-resistant microbial pathogens also contribute to the rising incidence of septic shock.

Pathogenesis: Microbial Factors

CAUSATIVE MICROORGANISMS

The microbiology of sepsis has undergone a remarkable transition in the past 25 years. The predominant microbial pathogens responsible for sepsis in the 1960s and 1970s were gram-negative bacilli and *Pseudomonas aeruginosa,* but there has been a progressive increase in the incidence of sepsis caused by gram-positive bacteria[4,11,12] and opportunistic fungi. The rapid evolution of antibiotic-resistance genes in gram-positive bacterial pathogens and the frequent occurrence of vascular catheter–related bacterial sepsis may account for the increasing prevalence of gram-positive pathogens as a cause of sepsis.

THE ROLE OF BACTERIAL ENDOTOXIN

Bacterial endotoxin, which is composed of lipopolysaccharide (LPS), is an intrinsic component of the outer membrane of gram-negative bacteria and is essential for the viability of enteric bacteria.[13] The unique potency of endotoxin is illustrated by the recent isolation of an endotoxin-deficient strain of *Neisseria meningitidis* that is at least 100-fold less potent an inducer of cytokine production than wild-type bacteria.[14] Endotoxin may enter the human circulatory system in its free form (released from dead organisms or shed from the membrane of viable organisms) or bound to the cell wall of intact bacteria. Whether free or bound, endotoxin appears to function essentially as an alarm molecule that alerts the host to the presence of invading gram-negative bacteria,[14,15] and its presence provokes a rigorous systemic inflammatory response. The host response to the endotoxin, rather than the endotoxin itself, accounts for the endotoxin's potentially lethal properties. As a species, humans are especially susceptible to the profound immunostimulant properties of endotoxin; even minute doses may be lethal.

Toll-like Receptors

On human immune cells, the Toll-like receptor (TLR) family is the transmembrane receptor for endotoxin and many other microbial mediators, such as peptidoglycan, lipopeptides, and lipoteichoic acid.[16] Ten TLRs have been identified to date. TLR4 is the principal endotoxin receptor,[17] whereas TLR2 and perhaps other TLRs recognize and signal the presence of a variety of microbial mediators, including endotoxin.[18,19] Comparative genomics in mice and fruit flies allowed for the identification of these critical receptors for microbial mediators on human cells.[20] Another TLR, TLR9, has been identified as the cellular receptor for unmethylated CpG motifs found in bacterial DNA but not in eukaryotic DNA. Upon recognition of bacterial DNA, TLR9 mediates an intense inflammatory response.[21,22]

TLRs belong to a family of pattern-recognition molecules that alert the innate immune response system to the presence of a microbial invader. Other pattern-recognition molecules include alternative complement components, mannose-binding lectin,[23] and CD14.[24,25] The innate immune system is by nature a rather nonspecific antimicrobial defense system. It lacks the precision of the acquired immune system (B cells and T cells), but its immediate action—phagocytosis and clearance of pathogens—in the initial stages of infection makes the innate immune response a critical survival mechanism. Activation of the innate immune system and its cellular components (neutrophils, monocytes, macrophages, and natural killer [NK] cells) are primarily responsible for the pathogenesis of septic shock.[26]

LPS Signaling

LPS is a phosphorylated, polar macromolecule that contains hydrophobic elements in the fatty acids of its lipid A core structure and hydrophilic elements in its repeating polysaccharide surface components. LPS forms microaggregates in biologic fluids and then rapidly interacts with a variety of serum or membrane-bound lipophilic proteins. Three receptors for LPS have been recognized in human cells: (1) soluble or membrane-bound CD14 molecules, (2) CD11/CD18 molecules (β2 integrins), and (3) scavenger receptors for lipid molecules. Soluble and membrane-bound CD14 greatly potentiate the host response to small quantities of LPS and other microbial mediators.[27]

In human blood and body fluids, LPS signaling is mediated by interactions with a hepatically derived, acute-phase plasma protein known as LPS-binding protein (LBP).[27] LBP functions primarily as a shuttle molecule that binds to polymeric LPS aggregates and transfers LPS monomers to CD14, which is a glycosyl phosphatidylinositol-linked protein found on the cell surfaces of such immune effector cells as the monocyte-macrophage and the neutrophil. After docking to membrane-bound CD14, LPS is delivered to an adjacent cell surface LPS receptor TLR4, along with an extracellular accessory protein known as MD2. This complex then triggers a signal to the intracellular space, subsequently activating LPS-responsive genes. CD14 also binds to bacterial peptidoglycan and lipopeptides and delivers these microbial ligands to TLR2 for intracellular signaling.

Through a well-characterized sequence of tyrosine and threonine/serine kinases, intracellular signaling leads to phosphorylation of inhibitory κB (IκB). This releases nuclear factor κB (NFκB) from the cytoplasm, allowing it to translocate into the nucleus. Clotting elements and acute phase proteins, cytokines, and nitric oxide synthase genes have NFκB binding sites in their regulatory elements. The outpouring of inflammatory cytokines and other inflammatory mediators after LPS exposure contributes to SIRS and is central to the pathogenesis of septic shock induced by gram-negative bacteria.[28,29]

Bactericidal/Permeability-Increasing Protein

Another important endotoxin-binding protein found in human plasma is bactericidal/permeability-increasing protein (BPI). This protein is 456 amino acids long, is produced by human neutrophils, and is found in greatest quantities in the azurophilic (primary) granules.[30] Its amino acid sequence is 45% homologous with LBP but has a distinctly antagonistic function with respect to LPS handling: unlike LBP, which facilitates LPS delivery to CD14 and thereby activates cells, BPI inhibits LPS delivery to CD14. BPI competes with LBP for LPS binding in biologic fluids.[30] The relative concentrations of these two endotoxin-binding proteins primarily determine the net effect of LPS release.

In human plasma, the concentration of LBP is two to three orders of magnitude higher than that of BPI. The opposite appears to be the case in abscess cavities, where BPI is present in much greater quantities than LBP.[19] This favors LPS-activating activity in the plasma and LPS-inhibitory activity in abscess cavities. Thus, BPI functions as an endogenous anti-endotoxin molecule; it may become a component of treatment for endotoxin-induced injury.[27,30]

Endotoxin Tolerance

The phenomenon of endotoxin tolerance (or reprogramming) has been well characterized in experimental models of sepsis and probably also occurs in human sepsis.[31] Endotoxin tolerance is the desensitization to endotoxin-induced lethality after a priming (small) dose of endotoxin before an otherwise lethal challenge dose of endotoxin. This phenomenon appears to be primarily mediated at the transcription level, with down-regulation of inflammatory cytokine genes. The precise molecular explanation for endotoxin tolerance is not fully characterized. The desensitizing dose of endotoxin may induce endogenous corticosteroids or anti-inflammatory cytokines such as interleukin-10 (IL-10), decrease cell surface expression of TLRs, alter nuclear translocation of signal transduction molecules, or decrease the stability of messenger RNA (mRNA) for cytokine genes.

The significance of endotoxin tolerance in humans is unclear, but some degree of endotoxin tolerance has been observed in patients treated with monophosphoryl lipid A.[31] Unquestionably, endotoxin is an important mediator in the pathogenesis of septic shock. Thus, efforts to limit LPS synthesis, prevent activation of host immune response elements, and enhance the clearance of LPS are important therapeutic strategies for the future treatment of septic shock.

BACTERIAL SUPERANTIGENS

Another important microbial mediator in the pathogenesis of septic shock is bacterial superantigen. Although they comprise a diverse group of protein-based exotoxins from streptococci, staphylococci, and other pathogens, superantigens share an unusual immunologic property: the capacity to activate large numbers of $CD4^+$ T cells in a short period by bypassing the usual mechanism of antigen processing and presentation.[32]

Conventional bacterial antigens are internalized by antigen-presenting cells (APCs) and undergo limited proteolysis and processing within the endosomal component of the macrophage. Appropriate peptide sequences of the microbial antigens (epitopes) are inserted into the central groove of major histocompatibility (MHC) class II molecules and are then expressed on the cell surface of APCs. Specific $CD4^+$ T cells that recognize the unique epitope are then activated. Clonal expansion of this small subset of T cells results in a physiologic immune response to the newly introduced antigen. Superantigens, in contrast, do not require intracellular processing by APCs. Superantigens bind directly to class II antigens adjacent to the epitope-specific peptide groove on APCs. Superantigens also bind to a limited number of Vβ regions of the T cell receptor on $CD4^+$ T cells. This binding brings $CD4^+$ T cells and macrophages into close proximity, which activates both the monocyte-macrophage and T cell populations.

Whereas a conventional peptide antigen stimulates only about one in 105 circulating lymphocytes that can recognize its unique structural epitope, a superantigen (e.g., toxic shock syndrome toxin–1 from *Staphylococcus aureus,* which binds to the $V\beta_2$ region of T cells) stimulates 10% to 20% of circulating human lymphocytes. Thus, a single bacterial superantigen can activate as much as 10% of the entire lymphocyte population.[33] This results in excessive activation of both lymphocytes and macrophages, which, in turn, leads to the uncontrolled synthesis and release of inflammatory cytokines.

Superantigen-induced immune activation may terminate in septic shock if the process is left unchecked. Polymicrobial infections with pathogens that release both bacterial superantigens and endotoxin may be particularly injurious to the host; the toxicity of bacterial endotoxin may be greatly enhanced by superantigens that prime the immune system to react to endotoxin in an overly sensitized manner [*see Figure 1*].

OTHER MICROBIAL MEDIATORS

Peptidoglycan from the cell wall of bacteria, capsular antigens, lipoteichoic acid, lipopeptides, microbial DNA, microbial toxins, and procoagulant substances produced by microbial pathogens may all contribute to the pathogenesis of sepsis. It has been observed that peptidoglycan from gram-positive bacteria interacts with CD14 molecules and activates inflammatory cells via TLR2 in a manner comparable to that observed with bacterial endotoxin.[18,19,33,34] CD14-dependent activation of mononuclear cells may occur from both gram-positive and gram-negative bacteria, although the level of activation is quantitatively less with gram-positive components.[35,36] TLR9-dependent, CD14-independent recognition of the unmethylated CpG motifs found in bacterial DNA also results in a vigorous inflammatory response.[22]

Moreover, gram-positive bacterial and fungal pathogens may induce systemic hypotension, resulting in redistribution of blood flow and in splanchnic vasoconstriction. The ischemia and subsequent reperfusion of the gastrointestinal tract may disrupt the intestinal mucosal barrier to bacterial products. Translocation of intact microbial pathogens as well as bacterial endotoxin from the GI tract to the circulation may occur during periods of severe stress and during periods of hypoperfusion of the GI mucosa.[37] Bacterial endotoxin and perhaps other gut-derived microbial mediators may play a pathogenic role in the ongoing inflammatory process after systemic hypotension produced by infectious and noninfectious insults. This finding has initiated interest in at-

tempts to strengthen the GI mucosal barrier through immunonutrition, epithelial growth factors, and selective decontamination of the GI tract in critical illness. These treatments remain potentially viable options and are areas of active research in the management of sepsis.

Pathogenesis: Host-Derived Mediators

CYTOKINE NETWORKS

Inflammatory cytokines play a pivotal role in the pathogenesis of sepsis. In animal studies, the administration of human tumor necrosis factor–α (TNF-α), an endogenous monocyte-macrophage–derived protein, was shown to have lethal consequences; in human volunteers, dramatic hemodynamic, metabolic, and hematologic changes were observed after administration of TNF-α.[38,39] The injurious effects of systemic levels of IL-1β have also been demonstrated.[40]

The major inflammatory cytokines, TNF-α and IL-1β, induce their hemodynamic and metabolic effects in concert with an expanding group of host-derived inflammatory mediators that work in a coordinated fashion to produce the systemic inflammatory response [*see Figure 1 and Table 2*].[40,41] The cytokine system functions as a network of communication signals between neutrophils, monocytes, macrophages, and endothelial cells. Autocrine and paracrine activation results in synergistic potentiation of the inflammatory response once it is activated by a systemic microbial challenge (e.g., endotoxemia). Much of the inflammatory response is localized and compartmentalized in the primary region of initial inflammation (e.g., lung tissue or the GI tract). If left unchecked, the inflammatory response spills over into the systemic circulation, resulting in a generalized reaction and culminating in diffuse endothelial injury, coagulation activation, and septic shock. The endocrinelike effect of the systemic release of cytokines and chemokines drives the inflammatory process and causes coagulopathies throughout the body.[1,2]

The multitude of inflammatory cytokines and chemokines found in excess quantities in the bloodstream in patients with septic shock is impressive and is matched by an equally daunting group of anti-inflammatory mediators [*see Table 2*]. The inflammatory mediators tend to predominate in the early phases of sepsis (the first 12 to 24 hours), whereas the endogenous anti-inflammatory components often prevail in the later phases of sepsis.[41] It has been observed that mice deficient in T cells and B cells respond to endotoxin challenge in the same way as normal mice.[42] Thus, monocyte-macrophage–generated cytokines are sufficient to drive the early septic process. However, lymphocyte-derived cytokines and interferons become important in the regulation of later phases of sepsis and may ultimately determine the outcome in septic shock.

$CD4^+$ T Cells

Important functional differences exist within $CD4^+$ T cells. Activated, yet uncommitted, $CD4^+$ T cells (T_{H0} cells) have two major pathways of functional differentiation. T_{H0} cells exposed to IL-12 in the presence of IL-2 are driven toward a T_{H1}-type functional development. These cells produce large quantities of interferon gamma (IFN-γ), TNF-α, and IL-2 and promote an inflammatory, cell-mediated immune response. In contrast, T_{H0} cells exposed to IL-4 will preferentially develop into a T_{H2}-type phenotype; T_{H2} cells secrete IL-4, IL-10, and IL-13. These cytokines promote humoral immune responses and attenuate macrophage and neutrophil activity.

Of interest is that T_{H1}-type cytokines suppress the expression of T_{H2}-type cytokines; IFN-γ inhibits the synthesis of IL-10. Conversely, IL-10 from T_{H2} cells is a potent inhibitor of TNF-α and IFN-γ synthesis by T_{H1} cells. The nature of the initial lymphocyte response is critical because the system tends to polarize over time into either a T_{H2}-type or T_{H1}-type response.[43] Evidence now

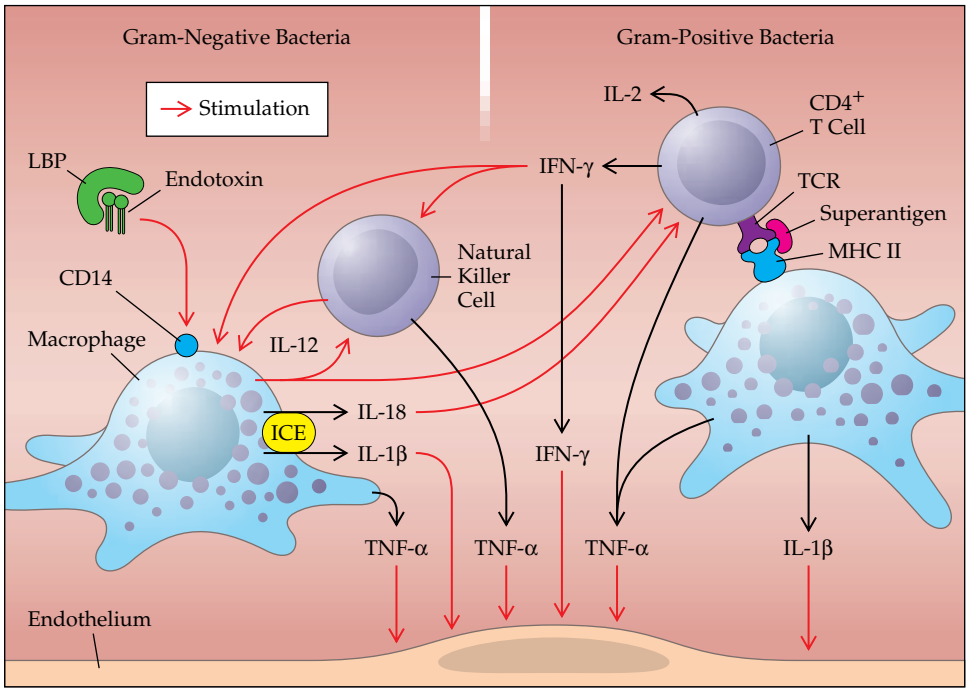


Figure 1 **Interactions between bacterial endotoxin and bacterial superantigens are shown. (ICE—interleukin-1β–converting enzyme; IFN-γ—interferon-gamma; IL—interleukin; LBP—LPS-binding protein; TNF-α—tumor necrosis factor-α)**

Table 2 Host-Derived Inflammatory Mediators in Septic Shock

Proinflammatory Mediators	*Anti-inflammatory Mediators*
Tumor necrosis factor–α	Interleukin-1 receptor antagonist
Interleukin-1β	Soluble tumor necrosis factor receptor
Interferon gamma	Soluble interleukin-1 receptor
Lymphotoxin-α	Transforming growth factor–β
Interleukin-2	Interleukin-4
Interleukin-8	Interleukin-6
Interleukin-12	Interleukin-10
Interleukin-18	Interleukin-11
Complement	Interleukin-13
Leukotriene B_4	Prostaglandin E_2
Platelet-activating factor	Granulocyte colony-stimulating factor
Bradykinin	Antioxidants
Nitric oxide	Interferon alfa
Granulocyte-macrophage colony-stimulating factor	Interferon beta
Chemokines	
Macrophage inhibitory factor	

suggests that similar forms of functional differentiation exist for CD8 cells as well ($CD8^+$ type 1 and type 2 cells).[44]

This process of functional differentiation is clinically relevant because sepsis is often accompanied by a late T_{H2}-type response after an initial septic insult. The stress hormone response in septic shock—expression of adrenocorticotropic hormone, corticosteroids, and catecholamines—promotes a T_{H2} response after systemic injury. This may lead to a phase of relative immune refractoriness (immune paralysis) in which the patient may be at increased risk for secondary bacterial or fungal infection.[41] This pathophysiologic state is associated with endotoxin tolerance; anti-inflammatory cytokine synthesis; and deactivation of monocytes, macrophages, and neutrophils. Methods to detect this immunosuppressed state and restore immune competence are under investigation. Patients with depressed expression of MHC class II antigens (e.g., HLA-DR) on the cell surface of macrophages may be in a functionally immunosuppressed state and may benefit from IFN-γ treatment.[45]

THE COAGULATION SYSTEM

Activation of the coagulation cascade and generation of a consumptive coagulopathy and diffuse microthrombi are well-recognized complications of severe sepsis. Studies of endotoxin challenge and TNF challenge in normal human volunteers indicate that the extrinsic pathway (tissue factor pathway) is the predominant mechanism by which the coagulation system is activated in human sepsis.[46,47] The contact factors in the intrinsic pathway are also activated, which secondarily initiates vasodilation through the generation of bradykinin.[47] Activation of intravascular coagulation results in microthrombi and may contribute to the multiorgan failure that occurs in septic patients. Depletion of coagulation factors and activation of plasmin, antithrombin III, and protein C may subsequently lead to a hemorrhagic diathesis. Depletion of these endogenous anticoagulants may secondarily lead to a procoagulant state and portend a poor prognosis.[48]

Current interest in the administration of tissue factor pathway inhibitor, activated protein C, and antithrombin III[49] for treatment of sepsis attest to the potential therapeutic value of regulation of the coagulation system in sepsis.[50] A phase 3 clinical trial in which 1,690 patients were treated with recombinant human activated protein C (drotrecogin alfa activated) was stopped when an interim analysis revealed a survival benefit for patients receiving activated protein C; mortality was 24.7% in treated patients versus 30.8% in placebo recipients ($P < 0.005$).[51] In contrast, a 2,300-patient multicenter trial with antithrombin III did not show any benefit.[49] Results of a phase 3 trial with tissue factor pathway inhibitor are pending at the time of this writing.

NEUTROPHIL–ENDOTHELIAL CELL INTERACTIONS

The recruitment of neutrophils to an area of localized infection is an essential component of the host inflammatory response. Localization and eradication of invading microbial pathogens at the site of initial infection is the principal objective of the immune response to microbial pathogens. This physiologic process may become deleterious if diffuse neutrophil–endothelial cell interactions occur throughout the circulation in response to systemic inflammation.

Complex mechanisms govern the migration of neutrophils from the intravascular space into the interstitium, where invasive microorganisms may reside [*see Figure 2*]. Activated neutrophils degranulate, exposing endothelial surfaces and surrounding structures to reactive oxygen intermediates, nitric oxide, and a variety of proteases. This process contributes not only to microbial clearance but also to diffuse endothelial injury in the setting of generalized systemic inflammatory responses. Regulation of neutrophil activity may represent a new area for therapeutic intervention in the management of sepsis.[52]

NITRIC OXIDE

Nitric oxide is a highly reactive free radical that plays an essential role in the pathophysiology of septic shock. It has a very short half-life (1 to 3 seconds), which tends to limit its activity to local tissues, where it is first generated by one of three isoforms of nitric oxide synthase. Regulation of the nitric oxide synthases is complex. Full expression of inducible nitric oxide synthase requires TNF-α, IL-1, LPS, and probably other regulatory elements.

Nitric oxide is the major endothelial-derived relaxing factor that initiates the vasodilatation and systemic hypotension observed in septic shock. Nitric oxide activates guanylate cyclase, which increases cyclic guanosine monophosphate levels inside vascular smooth muscle cells. This results in systemic vasodilatation and decreased vascular resistance. Within minutes of administration of an inhibitor of nitric oxide synthesis, blood pressure in hypotensive patients in septic shock moves toward normal levels.[53]

The other major physiologic effects of nitric oxide in septic shock are increased intracellular killing of microbial pathogens and regulation of platelet and neutrophil adherence. Nitric oxide is a highly diffusible gas that does not require specific receptors to enter eukaryotic or prokaryotic cells. In the presence of superoxide anion, nitric oxide leads to the formation of peroxynitrite. The peroxynitrite subsequently decays into highly cytotoxic molecules such as hydroxyl radicals and nitrosyl chloride, which, in turn, initiate lipid peroxidation and cause irreversible cellular damage. Nitric oxide inhibits a variety of key enzymes in the tricarboxylic acid pathway, the glycolytic pathway, DNA repair systems, electron transport pathways, and energy-exchange pathways. Because of its potent reactivity, nitric oxide alters the function of many metalloenzymes, carrier proteins, and structural elements.

Like many other components of the host inflammatory response, nitric oxide may have both advantageous and disadvantageous properties in sepsis. Nitric oxide regulates microcirculation to vital organs and contributes to intracellular killing of mi-

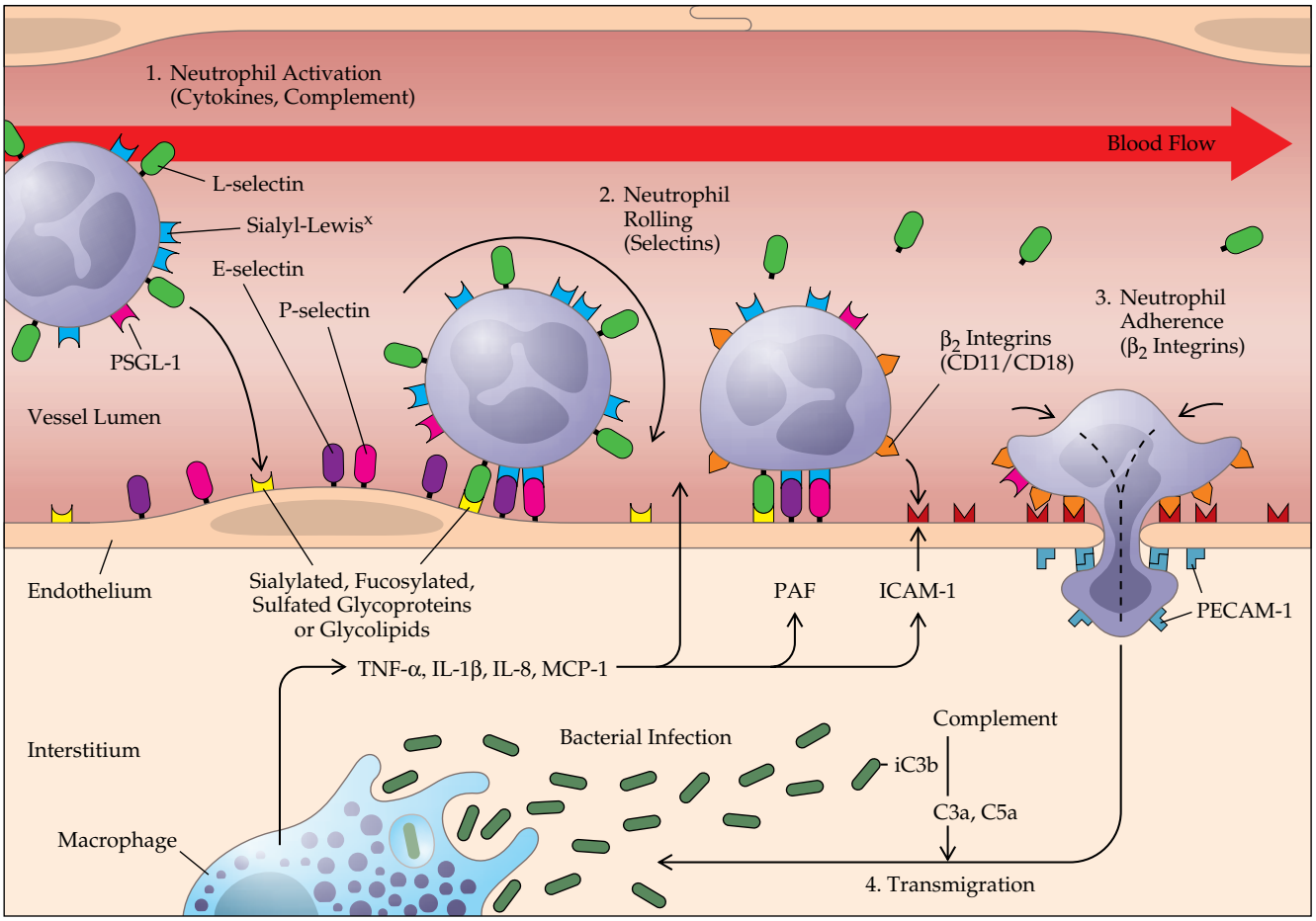


Figure 2 **Neutrophil–endothelial cell interactions in sepsis. (C—complement; ICAM-1—intercellular adhesion molecule–1; IL-1β—interleukin-1β; MCP-1—monocyte chemoattractant protein–1; PAF—platelet-activating factor; PECAM-1—platelet endothelial cell adhesion molecule–1; PSGL-1—P-selectin glycoprotein ligand–1; TNF-α—tumor necrosis factor–α)**

crobial pathogens. However, excessive and prolonged release of nitric oxide results in generalized vasodilatation and the systemic hypotension of septic shock. For those reasons, nitric oxide has become a target for therapeutic strategies in the management of sepsis.[54] Nonselective inhibitors of nitric oxide synthase, for example, have been shown to improve the hemodynamics of septic patients.[53] Unfortunately, this finding was not confirmed in a phase 3 trial.[55]

OTHER HOST-DERIVED MEDIATORS

At least two additional host-derived mediators contribute to the pathogenesis of septic shock. Macrophage migration inhibitory factor (MIF) is a late mediator that activates immune cells, upregulates TLR4 expression, and contributes to lethal septic shock.[56] This corticosteroid-regulated mediator promotes inflammation and has become a target for therapeutic agents in sepsis. High-mobility group–1 (HMG-1) protein also appears to contribute to late-onset inflammatory activities in septic shock. Inhibitors of HMG-1 may prove to have a therapeutic role in sepsis as well.[57]

Pathogenesis: Organ Dysfunction

The diffuse endothelial injury accompanying septic shock results in organ dysfunction distant from the original site of the septic insult. The signal that results in diffuse endovascular injury is thought to be relayed by plasma factors (e.g., inflammatory cytokines, complement, kinins, and other host-derived inflammatory mediators) or by a cellular element found in one or more of the immune effector cells.

Inadequate blood supply to vital tissues produces MODS. The failure of the microcirculation to support tissue maintenance may be the result of hypoperfusion of capillary beds, redistribution of blood flow within vascular beds, functional arteriovenous shunting, obstruction of blood flow from microthrombi, platelet or white blood cell aggregates, or abnormal deformability of red blood cells. Direct endothelial injury from nitric oxide, reactive oxygen intermediates, inflammatory cytokines, and inducers of apoptosis may directly damage endothelial surfaces. Endothelial swelling from the movement of intravascular fluid into the extravascular and intracellular spaces may mechanically obstruct the lumens of the capillary beds as well.

Although the origin of multiorgan failure in sepsis is principally related to microvascular effects, myocardial performance and pulmonary function also diminish over the course of septic shock and may contribute significantly to the development of MODS. Myocardial contractility decreases in response to a variety of myocardial depressant factors found in the plasma of septic patients. TNF-α is a prominent cause of myocardial dysfunction; IL-1, nitric oxide, and other host-derived inflammatory me-

Table 3 Standard Laboratory Values in Sepsis

Laboratory Study	*Typical Findings*	*Comments*
White blood cell count	Leukocytosis or leukopenia	Stress response, increased margination of neutrophils in sepsis; toxic granulation may be seen; occasionally, bacteria may be found in the peripheral blood smear
Platelet count	Thrombocytopenia	Look for evidence of fragmentation hemolysis in the peripheral blood smear; thrombocytopenia may or may not be accompanied by disseminated intravascular coagulation
Glucose	Hyperglycemia or hypoglycemia	Acute stress response, inhibition of gluconeogenesis
Clotting measurements	Elevated prothrombin time, activated partial thromboplastin time, low fibrinogen levels, and evidence of fibrinolysis	Coagulopathy often seen with systemic endotoxin release
Liver enzymes	Elevated alkaline phosphatase, bilirubin, and transaminases; low albumin	—
Blood cultures	Bacteremia or fungemia	The presence of positive blood culture does not make the diagnosis, and its absence does not exclude the diagnosis
Plasma lactate	Mild elevations (> 2.2 mmol/L)	Hypermetabolism, anaerobic metabolism, inhibition of pyruvate dehydrogenase
C-reactive protein	Elevated	Acute-phase reactant, sensitive but not specific for sepsis
Arterial blood gases	Respiratory alkalosis (early); metabolic acidosis (late)	Measurements of O_2 content and mixed venous O_2 saturation useful in management

diators may be contributing factors.[58] Acute lung injury occurs in septic shock as a result of damage to the pulmonary vascular circulation and the alveolocapillary membranes. A supply-dependent dysoxia (cytopathic hypoxia) may contribute to tissue injury in multiorgan failure in sepsis.[59]

Diagnostic Approach to Septic Shock

GENERAL FEATURES

In his classic treatise on human nature (*The Prince*, circa 1505), Machiavelli states, "Hectic fever [i.e., sepsis by current consensus definitions] at its inception is difficult to recognize but easy to treat; left untended, it becomes easy to recognize but difficult to treat." This statement is as true today as it was 500 years ago. Fully developed septic shock is a readily apparent clinical syndrome that is seldom confused with other pathologic states. However, the early phases of septic shock may be quite subtle even in carefully monitored patients. Early signs and symptoms may include confusion, apprehension, or decreased sensorium. Although fever is characteristic, hypothermia may occur and connotes a poor prognosis. An unexplained decrease in urinary output, sudden onset of cholestatic jaundice, unexplained metabolic alkalosis, excess bleeding at venipuncture sites, or even sudden unexplained hypotension may be the presenting finding in septic shock. It is essential that clinicians recognize these early signs and symptoms because successful management of septic shock depends on early recognition and appropriate intervention.

A variety of clinical, laboratory, and hemodynamic abnormalities are recognized in septic shock [*see Tables 3 and 4*]. Unfortunately, no single clinical or laboratory test is sufficiently specific and sensitive to reliably confirm the diagnosis of septic shock.

Table 4 Hemodynamic Findings in Sepsis

Parameter	*Typical Findings*	*Comments*
Heart rate	> 100 beats/min	Major compensatory mechanism for low systemic vascular resistance
Mean arterial blood pressure	< 65 mm Hg	Hallmark of septic shock
Cardiac index (cardiac output/m^2 [surface area])	> 4 L/min/m^2	Cardiac index elevated in early septic shock; may be depressed in late septic shock
Pulmonary arterial wedge pressure	4–10 mm Hg	Must be sure that hypovolemia is not the cause of hypotension; perform fluid resuscitation until pulmonary arterial wedge pressure returns to normal
Systemic vascular resistance (SVR)	< 800 dyne/sec/cm^{-5}	SVR often low in early septic shock; may become elevated in later phases of septic shock
Oxygen delivery (DO_2) Cardiac index (CI) × arterial O_2 content (A)	< 550 ml/min/m^2	Try to provide sufficient DO_2 to maintain adequate mixed venous O_2 saturation
Mixed venous O_2 saturation	< 70%	Low mixed venous O_2 indicates inadequate O_2 delivery to tissues in sepsis
Oxygen consumption (VO_2) (CI) × (A-VO_2) × 10	> 180 L/min/m^2	Typically increased in early septic shock

Table 5 Multiple Organ Dysfunction Syndrome in Severe Sepsis

Organ System	*Clinical-Metabolic Abnormalities*	*Histopathologic Findings*
CNS	Encephalopathy, decreased sensorium	Cerebral edema, microthrombi
Cardiac	Decreased myocardial performance	Altered calcium influx, interstitial edema
Lung	Acute respiratory distress syndrome	Exudation of fluid into the alveolar spaces, neutrophil plugging, hyaline membrane formation
Kidney	Acute tubular necrosis	Hypoperfusion, focal ischemia, microthrombi
Adrenal	Relative adrenal insufficiency, adrenal hemorrhage	Focal or diffuse hemorrhage, ischemic necrosis
Hepatobiliary system	Cholestatic jaundice, decreased hepatic synthesis of albumin, presence of clotting factors	Zonal necrosis, acalculous cholecystitis
Gut	Translocation of bacterial endotoxin and microorganisms, increased permeability	Diffuse interstitial edema, breaks in the epithelial membrane integrity, mucosal necrosis

Patients may have positive or negative blood cultures, leukocytosis or neutropenia, hyperglycemia or hypoglycemia, and respiratory alkalosis or metabolic acidosis. It is the constellation of signs and symptoms that leads to a diagnosis of septic shock.

The most common hemodynamic findings in early septic shock are a high cardiac output and a low systemic vascular resistance state, with initial maintenance of the systolic blood pressure as the heart attempts to compensate for the loss of systemic vascular tone. Myocardial performance is diminished even in the early phases of septic shock.[60] Without adequate intervention, circulating blood volume is continually lost into the interstitial space and intracellular locations. The heart can no longer compensate sufficiently, and systolic hypotension results. Deterioration of myocardial performance, accompanied by diffuse vasoconstriction, marks the late refractory state of septic shock.

MULTIPLE ORGAN DYSFUNCTION SYNDROME

One of the hallmarks of septic shock is the development of MODS. A constellation of clinical and metabolic abnormalities characterizes this syndrome [*see Table 5*]. The development of organ failure at the onset of sepsis or during its course is a poor prognostic factor and is a primary determinant of outcome.[61]

Acute Respiratory Distress Syndrome

The acute respiratory distress syndrome (ARDS) remains a major cause of morbidity and mortality in septic shock.[62] Increased capillary permeability in these patients results in pulmonary edema, which manifests clinically as dyspnea and cough; a standard anteroposterior chest x-ray will typically show bilateral, symmetrical alveolar opacities in all four quadrants [*see 14:X Pulmonary Edema*].

EXPERIMENTAL DIAGNOSTIC TECHNIQUES

The current assays used to diagnose septic shock are cumbersome and slow. Research to develop better assays is ongoing.

Plasma Endotoxin Levels

The measurement of plasma endotoxin levels may prove useful in helping to predict the development of shock.[63] Unfortunately, endotoxin levels are not uniformly elevated in patients with septic shock and may be spuriously elevated in patients with gram-positive infections or other hypotensive disorders. Also, the host responsiveness to endotoxin is variable and does not correlate with circulating endotoxin levels.

Bacterial Superantigen Levels

Circulating levels of bacterial superantigens have been successfully measured in selected patients with toxic shock syndrome.[64] Such measurements may prove useful in specific clinical situations.

IL-6 Levels

IL-6 is a cytokine that has myriad biologic activities, some of which are inflammatory and others of which are anti-inflammatory. IL-6 has been considered an indicator of cytokine activation because it is reliably present after activation of TNF-α and IL-1β. Patients with elevated IL-6 levels may respond favorably to anti-cytokine therapies.[65] In several studies, elevations of IL-6, as well as failure of IL-6 levels to fall after initiation of treatment for sepsis, was associated with a poor outcome. [65-68] Unfortunately, IL-6 levels are not specific for sepsis and may be elevated in a variety of inflammatory and infectious states. This lack of specificity limits the reliability of IL-6 measurement as a diagnostic method for septic shock.

Procalcitonin Levels

Procalcitonin, the propeptide of calcitonin, is normally produced by C cells in the thyroid.[69] In septic patients, procalcitonin is generated by numerous extrathyroidal tissues; its precise origin in this situation is unclear.[70,71] Procalcitonin has attributes that make it a potential marker for sepsis. It has a long half-life (approximately 24 hours), and measured levels will increase from undetectable to over 100 ng/ml during the course of septic shock. Procalcitonin levels do not become elevated as rapidly as IL-6 or IL-8 levels; elevated levels of procalcitonin are seen 4 to 6 hours after a systemic challenge with endotoxin or other septic stimuli.[72]

Of interest is that procalcitonin levels are elevated in severe sepsis but not in localized infections,[71] severe viral infections, or inflammatory conditions of noninfectious origin. In organ transplant recipients, procalcitonin levels may allow differentiation between the fever associated with rejection and that associated with sepsis.[71] Although the precise physiologic role of procalcitonin in sepsis has yet to be defined, procalcitonin elevation appears to be the most sensitive and reasonably specific indicator of severe sepsis currently available.[70,72]

Other Potential Markers

C-reactive protein and plasma lactate have been used as potential markers for sepsis, but their lack of specificity and sensi-

tivity limits their diagnostic value. IL-8 may prove useful, with or without IL-6, as an indicator of sepsis.[73]

Phospholipase A_2 is essential for the generation of platelet-activating factor and arachidonic acid derivatives, including thromboxane, prostacyclin, prostaglandins, and leukotrienes. Phospholipase A_2 or its precursor, type I prophospholipase A_2 propeptide, may prove to be a marker for sepsis, but the diagnostic value of measurements of this enzyme needs to be confirmed in clinical studies.[74]

Management of Septic Shock

There are four goals in the management of septic shock: (1) early recognition and resuscitation; (2) reestablishment of tissue perfusion and arterial blood pressure; (3) provision of optimal supportive care; and (4) timely initiation of treatment to eradicate the causative septic focus. After 2 decades in which the means for achieving these goals remained largely the same, the treatment approach has now changed, with the use of drotrecogin alfa activated,[51] low-dose corticosteroids,[75] and other supportive strategies for the management of septic shock.[76,77] In addition, modest improvements in outcome have accrued as the result of improved nutrition and supportive care and the skillful use of vasopressor agents.

The key determinant in survival is early recognition of sepsis and initiation of treatment while the process is readily reversible. This requires constant vigilance by the clinician caring for patients with a variety of medical and surgical illnesses.

FLUID RESUSCITATION

Fluid resuscitation is a mandatory first step in the treatment of septic shock. The diffuse vascular leak that occurs in septic shock necessitates provision of adequate circulating blood volume to maintain tissue perfusion.

Debate continues regarding the appropriateness of colloid versus crystalloid fluids. The lack of clear evidence of benefit of colloid agents (e.g., albumin, dextran, and plasma expanders) and their high cost have generally resulted in the use of saline solutions for volume expansion. A 1998 meta-analysis of studies comparing colloid versus crystalloid in sepsis found a slight worsening of outcome with colloid solutions.[78] Further review of these data indicate that differences in outcome are equivocal at best, and debate continues on the relative merits of colloids in sepsis.

The optimal amount of fluid for resuscitation of patients in septic shock remains a source of controversy. A delicate balance is required between maintenance of tissue perfusion and prevention of fluid overload, with its attendant risk of lung injury. Decreased myocardial performance in sepsis may necessitate a higher filling pressure for adequate cardiac output; however, exudation of fluids into the alveolar space in lung tissue and into the interstitium in other vital organs continues to be a major problem. Maintenance of a pulmonary arterial occlusion pressure of approximately 12 mm Hg is considered a reasonable starting point in patients who have hemodynamic monitors in place.[79] Rapid resuscitation with the goal of establishing normal mean central venous pressure and oxygen delivery to tissues has been shown to be of clinical value.[76] Diligent care and maintenance of central lines will reduce the frequency of catheter-related sepsis.[80]

VASOPRESSOR THERAPY

When patients fail to improve hemodynamically with fluids alone, vasopressor agents are often employed to reestablish systemic arterial blood pressure. Clinicians may choose among several vasopressor agents; however, the use of any vasopressor agent in septic shock carries with it certain risks and should be reserved for patients with significant hemodynamic instability that is unresponsive to fluid therapy.

Dopamine

Dopamine has been the vasopressor agent of choice for the past 2 decades because of its presumed favorable effects on renal perfusion (through promotion of renal vasodilation) and its modest inotropic effects. However, the validity of the privileged status of this choice has been questioned.[81] Dopamine effects are complicated by the fact that this catecholamine has its own receptors (D_1 and D_2 dopamine receptors) as well as variable affinities for alpha- and beta-adrenergic receptors. The effects of dopamine depend on the receptor density in specific vascular beds, the blood volume, and the dose used. Higher doses of dopamine increase the systemic vascular resistance by the drug's effects on alpha-adrenergic receptors in the peripheral circulation. Dopamine may have adverse effects on splanchnic blood flow,[82] and it has never been shown to be clearly beneficial to septic patients in an adequately controlled clinical trial.

Norepinephrine

Norepinephrine is a potent vasoconstrictor that is being used more frequently to treat the hemodynamic effects of septic shock. Earlier concerns regarding adverse consequences of norepinephrine on renal blood flow may have been overstated; studies suggest that norepinephrine may actually increase urine output and creatinine clearance in septic patients.[83] Norepinephrine may rapidly restore perfusion pressure within the glomerulus and result in improved glomerular filtration in patients with adequate fluid resuscitation.

Vasopressin

Vasopressin, which has its own vascular receptors distinct from adrenergic receptors, has gained favor as a vasopressor in sepsis. The clinical utility of vasopressin and related molecules (e.g., terlipressin) will ultimately be determined through large clinical comparative trials with other vasopressor agents.

Dobutamine

Dobutamine, a beta agonist, may improve cardiac output and oxygen delivery in some patients in septic shock who have low cardiac output. However, dobutamine may result in peripheral vasodilatation, which may be harmful in septic patients. Moreover, dobutamine increases myocardial oxygen consumption by its positive inotropic effects, which also may be detrimental. In one randomized trial in a heterogeneous group of critically ill patients, the use of dobutamine to boost the cardiac index and systemic oxygen delivery failed to improve outcome.[84]

VASODILATOR THERAPY

Another approach to improving the delivery of oxygen to the tissues of patients with septic shock is the use of vasodilators to open up poorly perfused capillary beds. Spronk and colleagues[85] recently presented a study of nitroglycerin therapy after intravascular volume resuscitation. Using an optical device to measure microcirculatory flow (orthogonal polarization spectral imaging), improved microvascular flow rates were achieved in septic patients who received adequate fluid repletion. This is an

Table 6 Suggested Empirical Antibiotic Choices in Severe Sepsis

Source of Infection	*Antimicrobial Choice*
Community-acquired pneumonia	Third-generation cephalosporin with a macrolide (alternative: fluoroquinolones)
Hospital-acquired pneumonia	Third- or fourth-generation cephalosporins, extended-spectrum penicillins with or without an aminoglycoside (alternatives: fluoroquinolones, carbapenems, β-lactam–β-lactamase inhibitor)
Urinary tract infections	Extended-spectrum β-lactam agent with or without an aminoglycoside (add ampicillin or vancomycin when enterococci are present)
Intra-abdominal infections	Third- or fourth-generation cephalosporins with or without metronidazole or clindamycin or extended-spectrum penicillins or β-lactam–β-lactamase inhibitor with or without an aminoglycoside (alternatives: carbapenem, trovafloxacin)
Biliary tract infections	Extended-spectrum penicillin with or without an aminoglycoside
Neutropenic patients	Extended-spectrum β-lactam agent with an aminoglycoside (add vancomycin when there is evidence of gram-positive infection)

appealing strategy to improve tissue oxygenation in septic shock, and it deserves further clinical trials.

The most reliable indicator to assess the adequacy of tissue perfusion in septic shock is not known. Numerous methods of measurement of tissue oxygenation (e.g., gastric tonometry, hepatic venous oxygen measurements, direct tissue oxygen measurements, and microcirculatory probes) have been developed to better measure and understand the critical requirements for oxygen delivery to tissues of septic patients. However, the practical value of these measurements in the clinical management of sepsis remains unclear. Moreover, there remains evidence that even with adequate oxygen delivery, dysoxia may develop in patients with sepsis because intracellular oxygen utilization is impaired by dysfunction of respiratory enzymes.

PREVENTING AND TREATING ARDS

Efforts to prevent lung injury include innovations in respiratory support, avoidance of trauma to the alveolocapillary units from excessive tidal and fluid volume, avoidance of oxidant-induced lung injury, salvage of functional alveolocapillary units through position change (prone position), and judicious fluid management.[86] Measurable improvements in the outcome of patients with ARDS have occurred, but considerable room for improvement remains in regard to the sparing of pulmonary function in patients with septic shock.

A major finding in regard to ARDS management was recognition of the hazard of providing excessive tidal volume through overly high ventilator settings. The resulting overdistention of airways can promote the progression of lung injury and the release of inflammatory mediators into the systemic circulation.[87] The ARDS clinical trials network study demonstrated conclusively that low stretch tidal volume settings (6 ml/kg) are clearly superior to the previous conventional high tidal volume setting (12 ml/kg).[88] Consequently, low tidal volume ventilation is now the standard of practice in the management of most forms of acute lung injury.

BLOOD TRANSFUSIONS

The role of blood transfusions in improving the oxygen-carrying capacity of blood has been the subject of considerable debate.[89] Humans have been shown to be remarkably resistant to adverse effects from isovolumetric reduction in hemoglobin values.[90] Banked, stored RBCs are less deformable, are less efficient at releasing oxygen from their 2,3-biphosphoglycerate–depleted hemoglobin stores, and may have immunosuppressive effects.[91] Promotion of endogenous erythrocyte production with erythropoietin may prove to be superior to blood transfusions[92]; further clinical trials with this treatment approach are warranted.

The hemoglobin level at which transfusion is indicated in septic shock has not been defined, but it appears to be considerably lower than the traditional threshold of less than 10 g/dl. Recent studies in ICU patients indicate that a conservative threshold, set as low as 7 g/dl, may in fact be preferable.

MANAGEMENT OF MULTIORGAN FAILURE

Expert management of acute renal failure [*see 10:VI Acute Renal Failure*], ARDS [*see 14:X Pulmonary Edema*], hepatic decompensation, coagulopathy, acid-base disturbances, and disordered hemodynamics is of paramount importance in the management of sepsis. Multiorgan failure is potentially reversible if rapid interventions correct the hemodynamic and inflammatory abnormalities.

NUTRITIONAL SUPPORT

Nutritional support in septic shock has changed considerably over the past 2 decades. Reliance on total parenteral nutrition has given way to early and extensive use of enteral hyperalimentation. Enteral feeding of septic patients has been shown to benefit enterocyte function, help maintain the intestinal permeability barrier, and help prevent gut-derived endotoxin and cytokine generation.[93] Nutritional supplementation with glutamine, arginine, and omega-3 fatty acids has experimental support and is increasingly being used in septic patients[94] [*see 4:XIII Enteral and Parenteral Nutrition*]. The incremental value of such enriched enteral formulations over standard enteral alimentation has not yet been confirmed in large clinical trials.

MANAGEMENT OF FEVER

Fever is a common concomitant of severe sepsis and appears to be an advantageous response.[95] In experimental animals with *Klebsiella pneumoniae* peritonitis, infection resolved and subjects recovered more rapidly when they were allowed to develop fever, compared with control animals in which normothermia was maintained externally.[96] Heat-shock proteins function as intracellular chaperones to stabilize and prevent denaturation of host proteins. Heat-shock protein induction may actually decrease the mortality associated with experimental endotoxin challenge.[97] Efforts to lower body temperature with cooling blankets are largely ineffective and may not benefit the patient in septic shock. This strategy should generally be avoided unless true hyperthermia is present.[98]

ANTIMICROBIAL THERAPY

The most appropriate antimicrobial therapy in sepsis depends on the source of infection, susceptibility patterns of microbial pathogens within a given institution, prior antimicrobial exposure, presence or absence of pregnancy, hepatic and renal function, and history of drug allergy. In septic shock, combinations of bactericidal antimicrobial agents are generally given on an empirical basis [*see Table 6*]. Antibiotic combinations decrease the

Table 7 Current Adjuvant and Experimental Therapies in the Treatment of Septic Shock

Experimental Agent	*Treatment Target*	*Comment*
Bactericidal/permeability-increasing protein	Endotoxin	Endotoxin-neutralizing human protein
Anti-CD14 antibody	Endotoxin/peptidoglycan	Endotoxin and peptidoglycan blocker
E5564	Endotoxin	Lipid A antagonist
Polymyxin B–binding columns	Endotoxin	Endotoxin-binding antibiotic
Platelet-activating factor (PAF)–acetylhydrolase	PAF	Rapid metabolism of PAF
FAB 2′ anti–tumor necrosis factor (TNF) monoclonal antibody	TNF	Neutralizes TNF in the circulation
Low-dose corticosteroids*	Adrenal function	Treat adrenal hypofunction of sepsis
Activated protein C†	Coagulation system	Inhibits disseminated intravascular coagulation, microthrombi
Hemoperfusion systems	Cytokines and endotoxin	Remove inflammatory mediators and endotoxin during hemoperfusion
Nitroglycerin infusion	Disordered microcirculation	Opens up poorly perfused capillary beds along with I.V. fluids
Arginine, glutamine, nucleic acids, micronutrients	Immunonutrition	Improve immune function and provide antioxidants
Caspase inhibitors	Cellular apoptosis	Block excess apoptosis of immune cells and endothelial cells

*One recent clinical trial demonstrated benefit; another study is ongoing.

†Approved for use in severe sepsis in patients at high risk of mortality (such as an APACHE II [Acute Physiology and Chronic Health Evaluation-II] score > 24 or multiorgan failure); its use in early, less severe sepsis is currently under investigation.

risk that a multidrug-resistant microbial pathogen will be missed and increase the probability that all of the important pathogenic microorganisms will be inhibited by at least one of the antimicrobial agents. Concern continues regarding antibiotic-induced endotoxin release in sepsis, but the clinical relevance of this effect has not been demonstrated and should not affect antibiotic choices for septic shock patients.[99]

MANAGEMENT OF BLOOD GLUCOSE LEVELS

Tight regulation of blood glucose levels is an important supportive management technique in sepsis. Van den Berghe and colleagues[77] reported improved survival, shorter ICU stays, and less bacteremia in patients with strict glycemic control (target was continuous euglycemia) versus conventional care in a cardiovascular ICU setting.

New Therapies for Sepsis

Over the past 15 years, more than 30 double-blind, placebo-controlled, multicenter, phase 2 or phase 3 trials have been conducted to study the efficacy of new experimental agents in the treatment of septic shock.[1,26,100-102] After a long list of disappointments, two recent studies have now shown convincing positive results. The study of recombinant human activated protein C (drotrecogin alfa activated; see above) represents the first successful phase 3 international trial in severe sepsis.[51] This clinical trial resulted in Food and Drug Administration approval of drotrecogin alfa activated (Xigris) for the treatment of adult patients with severe sepsis who have an especially high risk of dying from sepsis. Carefully selected patients benefit from this treatment regardless of the type of infecting microorganism that caused their sepsis. The drug is given as a continuous infusion at 24 µg/kg/hr for 4 days. Because protein C is an endogenous anticoagulant, the major side effect of treatment is bleeding.[50]

The second successful trial was a multicenter clinical study of low-dose corticosteroids. Annane and coworkers reported a significant improvement in survival in patients with vasopressor-dependent septic shock through the use of hydrocortisone (50 mg every 6 hours for 7 days) and fludrocortisone (50 µg/day for 7 days).[75] This treatment strategy is based on the frequent occurrence of relative adrenal insufficiency in patients with septic shock. In fact, low-dose corticosteroid therapy was effective only in those patients who showed evidence of inadequate adrenal response when given a short corticotropin test.[100] A large clinical trial is under way in Europe and Israel to confirm and extend these exciting results; meanwhile, a German study in 40 patients with septic shock has provided supporting evidence.[103] A follow-up study with a larger number of patients would be worthwhile (the Annane study involved 299 patients), but to date, low-dose corticosteroids appear to be a cost-effective, readily available, and relatively safe treatment option for patients with refractory septic shock.

These improvements in treatment strategies for septic patients reflect the newly heightened understanding of molecular events that underlie sepsis pathophysiology. It is anticipated that forthcoming innovations in therapy [*see Table 7*] will lead to further improvements in outcome in severe sepsis/septic shock. The genomics era has already provided insights into variations in the risk of developing sepsis[104] and differential responses to therapeutic agents.[105] Much work remains to be done to provide optimal care of the ever-growing septic patient population.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Bone R: Toward a theory regarding the pathogenesis of the systemic inflammatory response syndrome: what we do and do not know about cytokine regulation. Crit Care Med 24:163, 1996

2. From the bench to the bedside: the future of sepsis research. Executive summary of an American College of Chest Physicians; National Institute of Allergy and Infectious Diseases; and National Heart, Lung, and Blood Institute Workshop. Chest 111:744, 1997

3. Angus DC, Linde-Zwirble WT, Lidicker J, et al: Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med 29:1303, 2001

4. Friedman G, Silva E, Vincent JL: Has the mortality of septic shock changed with time? Crit Care Med 26:2078, 1998

5. Bone RC, Balk RA, Cerra FB, et al: Definitions for sepsis in organ failure and guidelines for the use of innovative therapies in sepsis. Chest 101:1644, 1992

6. Zeni F, Freeman B, Natanson C: Anti-inflammatory therapies to treat sepsis and septic shock: a reassessment. Crit Care Med 25:1097, 1997

7. Bone RC, Fisher CJ Jr, Clemmer TP, et al: Sepsis syndrome: a valid clinical entity. Crit Care Med 17:339, 1989

8. Levy MM, Fink MP, Marshall JC, et al: 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med 31:1250, 2003

9. Cross AS, Opal SM, Sadoff JC, et al: Choice of bacteria in animal models of sepsis. Infect Immun 61:2741, 1993

10. Martin GS, Mannino DM, Eaton S, et al: The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 348:1546, 2003

11. Sands KE, Bates DW, Lanken PN, et al: Epidemiology of sepsis syndrome in eight academic medical centers. JAMA 278:234, 1997

12. Bone RC: Gram-positive organisms in sepsis. Arch Intern Med 154:26, 1994

13. Rietschel ET, Brade H, Holst O, et al: Bacterial endotoxin: chemical composition, biological recognition, host response, and immunological detoxification. Curr Top Microbiol Immunol 216:39, 1996

14. Pridmore AC, Wyllie DH, Abdillahi F, et al: A lipopolysaccharide-deficient mutant of *Neisseria meningitidis* elicits attenuated cytokine release by human macrophages and signals via toll-like receptor (TLR)2 but not via TLR4/MD2. J Infect Dis 183:89, 2001

15. Hoffmann JA, Kafatos FC, Janeway CA, et al: Phylogenetic perspectives in innate immunity. Science 284:1313, 1999

16. Medzhitov R, Preston-Hurlburt P, Janeway CA Jr: A human homologue of the *Drosophila* Toll protein signals activations of adaptive immunity. Nature 388:394, 1997

17. Poltorak A, He X, Smirnova I, et al: Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in the *rlt4* gene. Science 282:2085, 1998

18. Takeuchi O, Hoshino K, Kawai T, et al: Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. Immunity 11:443, 1999

19. Lien E, Sellati TJ, Yoshimura A, et al: Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. J Biol Chem 274:33419, 1999

20. Brightbill HD, Libraty DH, Krutzik SR, et al: Host defense mechanisms triggered by microbial lipoproteins through toll-like receptors. Science 285:732, 1999

21. Hacker H, Vabulas RM, Takeuchi O, et al: Immune cell activation by bacterial CpG-DNA through myeloid differentiation marker 88 and tumor necrosis factor receptor–associated factor (TRAF) 6. J Exp Med 192:595, 2000

22. Hemmi H, Osamu T, Kawai T, et al: A toll-like receptor recognizes bacterial DNA. Nature 408:740, 2000

23. Hibberd ML, Sumiya M, Summerfield JA, et al: Association of variants of the gene for mannose-binding lectin with susceptibility to meningococcal disease. Lancet 353:1049, 1999

24. Frevert CW, Matute-Bello G, Skerrett SJ, et al: Effect of CD14 blockade in rabbits with *Escherichia coli* pneumonia and sepsis. J Immunol 164:5439, 2000

25. Schimke J, Mathison J, Morgiewicz, et al: Anti-CD14 mAb treatment provides therapeutic benefit after in vivo exposure to endotoxin. Proc Natl Acad Sci U S A 59:13875, 1998

26. Vincent JL, Sun Q, Dubois MJ: Clinical trials of immunomodulatory therapies in severe sepsis and septic shock. Clin Infect Dis 34:1084, 2002

27. Fenton MJ, Golenbock DT: LPS-binding proteins and receptors. J Leuk Biol 64:25, 1998

28. Ulevitch RJ, Tobias PS: Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. Annu Rev Immunol 13:437, 1995

29. Schumann RR, Leong SR, Flaggs GW: Structure and function of lipopolysaccharide-binding protein. Science 49:1431, 1992

30. Opal SM, Palardy JE, Marra MN, et al: Relative concentrations of endotoxin-binding proteins in body fluids during infection. Lancet 344:429, 1994

31. Astiz ME, Rackow EC, Still JG, et al: Pretreatment of normal humans with monophosphoryl lipid A induces tolerance to endotoxin: a prospective, double-blind, randomized, controlled trial. Crit Care Med 23:9, 1995

32. Florquine S, Goldman M: Immunoregulatory mechanisms of T-cell-dependent shock induced by bacterial superantigen in mice. Infect Immun 64:3443, 1996

33. Choi Y, Lafferty JA, Clements JRL: Selective expansion of T cells expressing Vb2 in toxic shock syndrome. J Exp Med 172:981, 1990

34. Kusunoki T, Hailman E, Juan TS-C, et al: Molecules from *Staphylococcus aureus* that bind CD14 and stimulate innate immune responses. J Exp Med 182:1673, 1995

35. Cleveland MG, Gorham JD, Murphy TL, et al: Lipoteichoic acid preparations from gram-positive bacteria induce interleukin-12 through a CD14-dependent pathway. Infect Immun 64:1906, 1996

36. Opal SM, Cohen J: Clinical gram-positive sepsis: does it fundamentally differ from gram-negative bacterial sepsis? Crit Care Med 27:1608, 1999

37. Casey LC, Balk RA, Bone RC: Plasma cytokine and endotoxin levels correlate with survival in patients with sepsis syndrome. Ann Intern Med 119:771, 1993

38. Dinarello CA: The proinflammatory cytokines interleukin-1 and tumor necrosis factor and treatment of the septic shock syndrome. J Infect Dis 63:1177, 1991

39. Dinarello CA, Gelfand JA, Wolff SM: Anticytokine strategies in the treatment of the systemic inflammatory response syndrome. JAMA 269:1825, 1993

40. Fantuzzi G, Dinarello CA: The inflammatory response in interleukin-1 beta-deficient mice: comparison with other cytokine-related knockout mice. J Leukoc Biol 59:489, 1996

41. Bone RC: Sir Isaac Newton, sepsis, SIRS and CARS. Crit Care Med 24:1125, 1996

42. Falk LA, McNally R, Perera PY, et al: LPS-inducible responses in severe combined immunodeficiency (SCID) mice. J Endotoxin Res 2:273, 1995

43. Szabo SJ, Dinghe AS, Gubler U, et al: Regulation of the interleukin (IL)-12Rb2 subunit expression in developing T helper 1 (Th1) and Th2 cells. J Exp Med 185:817, 1997

44. Faist E, Kim C: Therapeutic immunomodulatory approaches in the treatment of systemic inflammatory response syndrome and the prevention of sepsis. New Horiz 6:97, 1998

45. Kox WJ, Bone RC, Krausch D, et al: Interferon gamma-1b in the treatment of compensatory anti-inflammatory response syndrome: new approach: proof of principle. Arch Intern Med 157:389, 1997

46. van Deventer SJH, Buller HR, ten Cate JW, et al: Experimental endotoxemia in humans: analysis of cytokine release and coagulation, fibrinolytic and complement pathways. Blood 76:2520, 1990

47. Bauer KA, ten Cate H, Barzegar S, et al: Tumor necrosis factor infusions have a procoagulant effect on the hemostatic mechanism of humans. Blood 74:165, 1989

48. Levi M, ten Cate H, van der Poll T, et al: Pathogenesis of disseminated intravascular coagulation in sepsis. JAMA 270:975, 1993

49. Warren BL, Alain E, Singer P, et al: High-dose antithrombin III in severe sepsis: a randomized controlled trial. JAMA 286:1869, 2001

50. Smith OP, White B, Vaughn D, et al: Use of protein-C concentrate, heparin, and haemodiafiltration in meningococcus-induced purpura fulminans. Lancet 350:1590, 1997

51. Bernard GR, Vincent JL, Laterre PF, et al: Efficacy and safety of recombinant human activated protein C for severe sepsis. Recombinant human protein C Worldwide Evaluation in Severe Sepsis (PROWESS) study group. N Engl J Med 344:699, 2001

52. Luster AD: Chemokines: chemotactic cytokines that mediate inflammation. N Engl J Med 338:436, 1998

53. Petros A, Bennett D, Vallance P: Effect of nitric oxide synthase inhibitors on hypotension in patients with septic shock. Lancet 338:1557, 1991

54. Shenep JL, Toumanen E: Prospective: targeting nitric oxide in the adjuvant therapy of sepsis in meningitis. J Infect Dis 177:766, 1998

55. Grover R, Lopez A, Lorente J, et al: Multicenter, randomized, placebo-controlled, double-blind study of the nitric oxide synthase inhibitor 546C88: effect on survival in patients with septic shock. Crit Care Med 27:433, 1999

56. Calandra T, Spiegel LA, Metz CN, et al: Macrophage migration inhibitory factor is a critical mediator of the activation of immune cells by exotoxins of gram-positive bacteria. Proc Natl Acad Sci U S A 95:11383, 1998

57. Andersson U, Wang H, Palmblad K, et al: High mobility group 1 protein (HMG-1) stimulates pro-inflammatory cytokine synthesis in human monocytes. J Exp Med 192:565, 2000

58. Finkel MS, Oddis CV, Jacob TD, et al: Negative inotropic effect of cytokines on the heart mediated by nitric oxide. Science 257:387, 1992

59. Shoemaker WC, Montgomery ES, Kaplan E, et al: Physiologic patterns in surviving and non-surviving shock patients. Arch Surg 106:630, 1973

60. Parker MN, Suffredini AF, Natanson C, et al: Responses of left ventricular function in survivors and non-survivors of septic shock. Crit Care 4:19, 1989

61. Pittet D, Thievent B, Wenzel RP, et al: Importance of pre-existing co-morbidities for prognosis of septicemia in critically ill patients. Intensive Care Med 19:265, 1993

62. Ware LB, Matthay MA: The acute respiratory distress syndrome. N Engl J Med 342:1335, 2000

63. Opal SM, Scannon P, Vincent JL, et al: The clinical significance of quantitative measurement of plasma endotoxin and LPS binding protein in severe sepsis. J Infect Dis 180:1584, 1999

64. Sriskandan S, Moyes D, Cohen J: Detection of circulating bacterial superantigen and lymphotoxin alpha in patients with streptococcal toxic-shock syndrome. Lancet 348:1315, 1986

65. Damas P, Ledoux D, Nys M, et al: Cytokine serum levels during severe sepsis in humans: IL-6 as a marker of severity. Ann Surg 215:356, 1992

66. Reinhart K, Wiegand-Lohnert C, Grimminger F, et al: Assessment of the safety and efficacy of the monoclonal anti-tumor necrosis factor antibody-fragment, MAK 195F, in patients with sepsis and septic shock: a multicenter, randomized, placebo-controlled, dose-ranging study. Crit Care Med 24:733, 1996

67. Opal SM, Fisher CJ, Pribble JP, et al: The confirmatory interleukin-1 receptor antagonist trial in sepsis: a phase III randomized, double-blind, placebo-controlled, multicenter trial. Crit Care Med

25:1115, 1997

68. Fisher CJ, Dhainaut JF, Opal SM, et al: Recombinant human interleukin-1 receptor antagonist in the treatment of patients with sepsis syndrome: results from a randomized, double-blind, placebo-controlled trial. JAMA 271:1836, 1994

69. Gendrel D, Assicot M, Raymond J, et al: Procalcitonin as a marker for the early di-

agnosis of neonatal infection. J Pediatr 128:570, 1996

70. deWerra I, Jacchard C, Corradin SB, et al: Cytokines, nitrite/nitrate, soluble tumor necrosis factor receptors, and procalcitonin concentration: comparisons in patients with septic shock, cardiogenic shock, and bacterial pneumonia. Crit Care Med 25:607, 1997

71. Staehler M, Hammer C, Meiser B, et al: Procalcitonin: a new marker for differential diagnosis of acute rejection and bacterial infection in heart transplantation. Transplant Proc 29:584, 1997

72. Karzai W, Oberhoffer M, Meier-Helmann A, et al: Procalcitonin: a new indicator of systemic response to severe infections. Infection 6:629, 1997

73. Hack CA, Hart M, van Schijndel RJ, et al: Interleukin-8 in sepsis: relation to shock and inflammatory mediators. Infect Immun 60:2835, 1992

74. Rae D, Porter J, Beechey-Newman N, et al: Type 1-prophospholipase A2 propeptide in acute lung injury. Lancet 344:1472, 1994

75. Annane D, Sebille V, Charpentier C, et al: Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. JAMA 288:862, 2002

76. Rivers E, Nguyen B, Havstad S, et al: Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 345:1368, 2001

77. Van den Berghe G, Wouters P, Weekers F, et al: Intensive insulin therapy in critically ill patients. N Engl J Med 345:1359, 2001

78. Schierhout G, Roberts I: Fluid resuscitation with colloid or crystalloid solutions in critically ill patients: a systemic review of randomized trials. BMJ 316:961, 1998

79. Rackow EC, Kaufman BS, Falk JL, et al: Hemodynamic response to fluid resuscitation in patients with septic shock: evidence for early depression of cardiac performance. Circ Shock 22:11, 1987

80. Mermel LA: Prevention of intravascular catheter-related infections. Ann Intern Med 132:391, 2000

81. Martin C, Papazian L, Perrin G, et al: Norepinephrine or dopamine for the treatment of hyperdynamic septic shock? Chest 103:1826, 1993

82. Germann R, Haiskackl M, Hasibeder W, et al: Dopamine and mucosal oxygenation in porcine jejunum. J Appl Physiol 77:2845, 1994

83. Marin C, Eon B, Saux P, et al: Renal effects of norepinephrine used to treat septic shock patients. Crit Care Med 18:282, 1990

84. Hayes MA, Timmins AC, Yau EHS, et al: Elevation of systemic oxygen delivery in the treatment of critically ill patients. N Engl J Med 330:1717, 1994

85. Spronk PE, Ince C, Gardien MJ, et al: Nitroglycerin in septic shock after intravascular volume resuscitation. Lancet 360:1395, 2002

86. Milberg JA, Davis DR, Steinberg KP, et al: Improved survival of patients with acute respiratory distress syndrome (ARDS): 1983–1993. JAMA 273:306, 1995

87. Stewart TE, Meade MO, Cook DJ, et al: Evaluation of a ventilation strategy to prevent barotrauma in patients at high risk for acute respiratory distress syndrome. N Engl J Med 338:355, 1998

88. Ventilation with lower tidal volumes as compared with traditional tidal volume for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. N Engl J Med 342:1301, 2000

89. Hebert PC, Wells G, Marshall J, et al: Transfusion requirements in critical care. JAMA 273:1439, 1995

90. Weiskopf RB, Viely MK, Feiner J, et al: Human cardiovascular and metabolic response to acute, severe isovolemic anemia. JAMA 279:217, 1998

91. Hebert PC, Wells G, Blajchman MA, et al: A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. N Engl J Med 340:409, 1999

92. Corwin HL, Gettinger A, Rodriguez RM, et al: Efficacy of recombinant human erythropoietin in the critically ill patient: a randomized, double-blind, placebo-controlled trial. Crit Care Med 27:2346, 1999

93. Kudsk KA, Croce MA, Fabian TC, et al: Enteral vs. parenteral feeding: effects on septic morbidity after blunt and penetrating abdominal trauma. Ann Surg 215:503, 1992

94. Bower RH, Cerra FB, Bershadshy B, et al: Early enteral administration of a formula (Impact) supplemented with arginine, nucleotides, and fish oil in intensive care unit patients: results of a multicenter prospective, randomized, clinical trial. Crit Care Med 23:436, 1995

95. Kluger MJ, Kozak W, Conn C, et al: The adaptive value of fever. Infect Dis Clin North Am 10:1, 1996

96. Jiang Q, Cross AS, Singh IS, et al: Febrile core temperature is essential for optimal host defense in bacterial peritonitis. Infect Immun 68:1265, 2000

97. Ribeiro SP, Chu E, Slutsky AS: Heat stress after injection of LPS in rats decreases mortality. Am J Respir Crit Care Med 153:A252, 1996

98. O'Donnell J, Axelrod R, Fisher C, et al: Use and effectiveness of hypothermia blankets for febrile patients in the intensive care unit. Clin Infect Dis 24:1208, 1997

99. Prins JM, van Deventer SJH, Kuijper EJ, et al: Clinical relevance of antibiotic-induced endotoxin release. Antimicrob Agents Chemother 38:1211, 1994

100. Opal SM, Cross AS: Clinical trials for sepsis: past failures and future hopes. Infect Dis Clin North Am 13:285, 1999

101. Abraham E, Matthay MA, Dinarello CA, et al: Consensus Conference Definitions for sepsis, acute lung injury, and ARDS—time for reevaluation. Crit Care Med 28:232, 2000

102. Levin M, Quint PA, Goldstein B, et al: Recombinant bactericidal/permeability-increasing protein (rBPI21) as adjunctive treatment for children with severe meningococcal disease: a randomized trial. Lancet 356:961, 2000

103. Keh D, Boehnke T, Weber-Cartens S, et al: Immunologic and hemodynamic effects of "low-dose" hydrocortisone in septic shock: a double-blind, randomized, placebo-controlled, crossover study. Am J Respir Crit Care Med 167:512, 2003

104. Gibot S, Cariou A, Drouet L, et al: Association between a genomic polymorphism with the CD14 locus and septic shock susceptibility and mortality rate. Crit Care Med 30:969, 2002

105. Lorenz E, Mira JP, Frees KL, et al: Relevance of mutations in the TLR4 receptor in patients with gram-negative septic shock. Arch Intern Med 162:1028, 2002

106. Bernard GR, Artigas A, Brigham KL, et al: The American-European Consensus on ARDS: definitions, mechanisms, relevant outcomes, and clinical trial coordination. Am J Respir Crit Care Med 149:181, 1994

107. Murray JF, Matthay MA, Luce JM, et al: An expanded definition of the adult respiratory distress syndrome. Am Rev Respir Dis 138:720, 1988

Acknowledgment

Figures 1 and 2 Seward Hung.

XXXI VIRAL ZOONOSES

Lyle R. Petersen, m.d., m.p.h.
Duane J. Gubler, sc.d.

Zoonoses are human diseases caused by pathogens that normally infect animals. About 534 zoonotic viruses from eight taxonomic families have been identified, 120 of which are known to cause human illness [*see Tables 1 and 2*]. The natural hosts of zoonotic viruses are usually unaffected by the viruses. Infection in humans may cause no obvious illness, a nonspecific viral syndrome, or more severe illness that generally falls into one of three categories: hemorrhagic fever, encephalitis, or rash arthralgia [*see Table 1*].

Transmission of Zoonotic Viruses

Zoonotic viruses replicate in the reservoir animal host and are usually transmitted to humans by direct contact or the bite of a hematophagous (blood-sucking) arthropod. Transmission by direct contact normally involves a bite by the infected reservoir animal or handling of the animal's tissues or materials contaminated by the animal's body fluids. Most viral zoonoses require a blood-sucking arthropod for transmission to humans. Mosquitoes are the most important arthropod vectors, followed by ticks, sandflies, and midges. Arthropod vector–borne viruses are called arboviruses and are maintained in complex life cycles involving a nonhuman primary vertebrate host and a primary arthropod vector [*see Figure 1*]. The arthropod vector usually becomes infected when it ingests virus while feeding on the blood of a viremic animal. Virus replicates in the arthropod tissues, ultimately infecting the salivary glands. The arthropod then transmits the virus to a new host when it injects infective salivary fluid while taking a blood meal. This extrinsic incubation period (i.e., the time between ingestion and transmission of the virus) is usually 8 to 12 days, depending on environmental factors, the virus, and the vector species.

Arthropod-borne viruses generally remain undetected until humans encroach on the natural enzootic focus or until the virus escapes the primary cycle via a secondary vector or vertebrate host [*see Figure 1*]. Although humans may become ill, they are generally considered dead-end hosts because they do not develop sufficient viremia to infect feeding vectors and thus do not contribute to the transmission cycle. Notable exceptions include dengue, yellow fever, chikungunya, and Ross River virus infection [*see Table 1*].

Hemorrhagic Fevers

Hemorrhagic fevers are diseases, generally viral, that often cause extensive bleeding in humans. Specific laboratory diagnosis of hemorrhagic fevers usually requires special serologic or virologic tests, such as enzyme-linked immunosorbent assays (ELISAs) to detect virus-specific immunoglobulin M (IgM) or immunoglobulin G (IgG) antibody, or other tests such as hemagglutination-inhibition, complement fixation, and neutralization tests on paired serum samples taken during the acute and convalescent phases of illness. Some viruses produce viremia [*see Table 1*] and can be isolated from, or detected in, the acute-phase serum or cerebrospinal fluid by polymerase chain reaction (PCR) or immunohistochemistry (IHC) testing of autopsy tissues. Clinicians who suspect a hemorrhagic fever in one of their patients can have samples sent through their state health department to the Centers for Disease Control and Prevention (CDC) for testing.

VIRUSES OF THE FAMILY FLAVIVIRIDAE

Dengue Fever

The dengue virus complex (family Flaviviridae, genus *Flavivirus*) consists of four antigenically related serotypes (DEN-1, DEN-2, DEN-3, and DEN-4). Although there is extensive cross-reactivity between dengue virus serotypes in serologic tests, there is no lasting cross-protective immunity in humans; cross-protection lasts for only a few months. Thus, individuals can have as many as four dengue infections in their lifetime, one from each serotype.[1]

Epidemiology All four dengue virus serotypes have a worldwide distribution in the tropics and are maintained in tropical rain forests of Asia and Africa in a mosquito-monkey-mosquito cycle and in most tropical urban centers in a mosquito-human-mosquito transmission cycle.[1] The forest cycle is not considered important in terms of public health. A map of countries reporting dengue can be found at www.cdc.gov/ncidod/dvbid/dengue/index.htm. In many urban centers, multiple virus serotypes cocirculate (a phenomenon known as hyperendemicity). An estimated 50 to 100 million infections occur annually. The principal mosquito vector is *Aedes aegypti,* an African species that spread around the world during the 17th, 18th, and 19th centuries via the slave trade and shipping industry. *Ae. aegypti* became well adapted to living in intimate association with humans and is a highly efficient epidemic vector in urban settings. Secondary vectors include other *Aedes (Stegomyia)* species such as *Ae. albopictus, Ae. polynesiensis*, and *Ae. scutellaris*. These secondary vector species can transmit dengue viruses during outbreaks, but they are more important as maintenance vectors.

Diagnosis In the United States, dengue fever should be suspected in a traveler who falls abruptly ill within 2 weeks of returning from the tropics. Infection with dengue viruses can be inapparent or can cause a spectrum of clinical illness ranging from a mild, nonspecific viral syndrome to classic dengue fever to severe and fatal hemorrhagic disease. The classic form usually affects adults and older children; in young children, the illness is usually mild but may be severe. After an infective mosquito bite, there is an incubation period of 3 to 14 days (average, 4 to 7 days), followed by the sudden onset of fever (which is often biphasic—with 2 to 5 days of fever, followed by a 1- to 2-day afebrile period, and then 1 to 2 days of fever), severe headache, chills, retro-orbital pain, and generalized, severe pain in the muscles and joints. A maculopapular rash generally appears on the trunk between the third and fifth days of illness and spreads to the face and extremities. Nausea, vomiting, lymphadenopathy, anorexia, constipation, and altered taste sensation are common. Occasionally, petechiae are seen on the dorsum of the feet, legs, hands, axillae, and palate late in the illness. The illness generally lasts 5 to 7 days, after which recovery is complete, although convalescence may be prolonged. Leukopenia with a relative lym-

Table 1 Important Viral Zoonoses That Cause Human Disease

Family/Virus	*Vector*	*Vertebrate Host*	*Ecology*	*Disease in Humans*	*Geographic Distribution*	*Epidemics*
Togaviridae						
Chikungunya*	Mosquitoes	Humans, primates	U,† S, R	SFI	Africa, Asia	Yes
Ross River*	Mosquitoes	Humans, marsupials	R,† S, U	SFI	Australia, South Pacific	Yes
Mayaro*	Mosquitoes	Birds	R, S, U	SFI	South America	Yes
Ónyong-nyong*	Mosquitoes	?	R	SFI	Africa	Yes
Sindbis	Mosquitoes	Birds	R	SFI	Asia, Africa, Australia, Europe, Americas	Yes
Eastern equine encephalitis	Mosquitoes	Birds	R	SFI, ME	Americas	Yes
Western equine encephalitis	Mosquitoes	Birds, rabbits	R	SFI, ME	Americas	Yes
Venezuelan equine encephalitis*	Mosquitoes	Rodents	R	SFI, ME	Americas	Yes
Barmah Forest*	Mosquitoes	?		SFI	Americas	Yes
Flaviviridae						
Dengue I–IV*	Mosquitoes	Humans, primates	U,† S, R	SFI, HF	Worldwide in tropics	Yes
Yellow fever*	Mosquitoes	Humans, primates	R,† S, U	SFI, HF	Africa, South America	Yes
Kyasanur Forest disease*	Ticks	Primates, rodents, camels	R	SFI, HF, ME	India, Saudi Arabia	No
Omsk hemorrhagic fever	Ticks	Rodents	R	SFI, HF	Asia	No
Japanese encephalitis	Mosquitoes	Birds	R,† S	SFI, ME	Asia	Yes
Murray Valley encephalitis	Mosquitoes	Birds	R	SFI, ME	Australia	Yes
Rocio	Mosquitoes	Birds	R	SFI, ME	South America	Yes
St. Louis encephalitis	Mosquitoes	Birds	R,† S, U	SFI, ME	Americas	Yes
West Nile encephalitis	Mosquitoes	Birds	R,† S, U	SFI, ME	Asia, Africa, North America, Europe	Yes
Tick-borne encephalitis	Ticks	Rodents	R	SFI, ME	Europe, Asia	No
Bunyaviridae						
Sandfly fever*	Sandflies	?	R	SFI	Europe, Africa, Asia	Yes
Rift Valley fever*	Mosquitoes	?	R	SFI, HF, ME	Africa	Yes
La Crosse encephalitis	Mosquitoes	Rodents	R,† S	SFI, ME	North America	No
California encephalitis	Mosquitoes	Rodents	R	SFI, ME	North America, Europe, Asia	Yes
Crimean-Congo hemorrhagic fever*	Ticks	Rodents	R	SFI, HF	Europe, Asia, Africa	Yes
Oropouche*	Midges	?	R,† S, U	SFI	Central and South America	Yes
Hemorrhagic fever with renal syndrome		Rodents	R, S	SFI, HF	Asia, Europe	Yes
Hantavirus pulmonary syndrome		Rodents	R	SFI, HF	United States, Central and South America	Yes
Arenaviridae						
Lassa fever		Rodents	R	SFI, HF	Africa	Yes
Venezuelan hemorrhagic fever		Rodents	R	SFI, HF	Venezuela	Yes
Bolivian hemorrhagic fever		Rodents	R	SFI, HF	Bolivia	Yes
Argentine hemorrhagic fever		Rodents	R	SFI, HF	Argentina	Yes
Filoviridae						
Ebola	?	?	R	SFI, HF	Africa	Yes
Marburg	?	?	R	SFI, HF	Africa	Yes
Rhabdoviridae						
Rabies		Bats, dogs, raccoons	R, S, U	SFI, ME	Global	No
Paramyxoviridae						
Nipah		Pigs, ?bats	R	SFI, ME	Malaysia	Yes
Reoviridae						
Colorado tick fever	Ticks	Rodents, small mammals	R	SFI, ME	Western United States, Canada	No

*Arboviruses that produce significant human viremia.
†The most important ecology.
HF—hemorrhagic fever ME—meningoencephalitis R—rural S—suburban SFI—systemic febrile illness U—urban

phocytosis and thrombocytopenia may occur. Liver enzyme levels may be elevated, and hemorrhagic manifestations may occur. Neurologic manifestations such as encephalopathy and seizures may occur during the disease's febrile stage.[2,3]

The diagnosis of dengue infections should be based on clinical signs and symptoms and on epidemiologic information such as travel history. Laboratory testing is useful only for confirmation of the clinical diagnosis. The IgM-capture ELISA is the serologic test of choice for dengue infection.

Treatment and prevention Treatment of dengue fever is supportive; there are no antiviral agents that are effective for the disease. Prevention consists of environmental control (see below).

Table 2 Principal Hantaviruses That Cause Human Disease

Virus	*Disease*	*Geographic Distribution*	*Primary Host in Nature*
Dobrava	HFRS (severe)	Balkans	*Apodemus flavicollis* (yellow-necked field mouse)
Hantaan	HFRS (severe)	Asia	*Apodemus agrarius* (striped field mouse)
Puumala	HFRS (mild)	Northern, western, central Europe; Balkans; Russia	*Clethrionomys glareolus* (red bank vole)
Seoul	HFRS (moderate)	Principally Southeast Asia, probably worldwide	*Rattus norvegicus*, *R. rattus* (common rat, Norway rat)
Andes	HPS (renal)	Argentina	*Oligoryzomys longicaudatus* (long-tailed pygmy rice rat)
Bayou	HPS (renal)	United States	*Oryzomys palustris* (rice rat)
Black Creek Canal	HPS (renal)	United States	*Sigmodon hispidus* (cotton rat)
Juquitiba	HPS	Brazil	Host unknown
Laguna Negra	HPS	Paraguay, Bolivia	*Calomys laucha* (vesper mouse)
Lechiguanas	HPS (renal)	Argentina	*Oligoryzomys* (long-tailed mouse)
Monongahela	HPS	United States	*Peromyscus leucopus* (white-footed mouse)
New York	HPS	United States	*Peromyscus leucopus* (white-footed mouse)
Oran	HPS (renal)	Argentina	*Oligoryzomys longicaudatus* (long-tailed pygmy rice rat)
Sin Nombre	HPS	North America	*Microtus pennsylvanicus* (meadow vole)

HFRS—hemorrhagic fever with renal syndrome HPS—hantavirus pulmonary syndrome

Dengue Hemorrhagic Fever

Epidemiology Dengue hemorrhagic fever (DHF) is a severe form of dengue infection that is most commonly observed in children younger than 15 years in Southeast Asia and in all age groups in the Americas and the Pacific region.[4,5]

Pathogenesis The pathogenesis of DHF is still not well understood. Classic DHF with a vascular leak syndrome may have a unique immunopathologic basis that is associated with enhancement of viral infection of mononuclear phagocytes in patients with dengue antibodies from a previous infection with a different serotype (heterologous antibody).[6] Infection of mononuclear phagocytes stimulates the release of vasoactive mediators, leading to a cascade of events that result in increased vascular permeability.

Although the risk of DHF is higher in patients experiencing a second dengue infection, DHF also occurs in patients who have primary infections; thus, heterologous dengue antibody (previous infection) is not a prerequisite for DHF. Furthermore, some strains of dengue viruses cannot be enhanced in vitro. Both field evidence and laboratory evidence support a more prominent role of viral factors in the pathogenesis of DHF and suggest that virus strain and serotype are also important risk factors for severe disease.[1,3,6-8] Hemorrhage may occur without vascular leakage, suggesting another pathogenetic mechanism.[3,9]

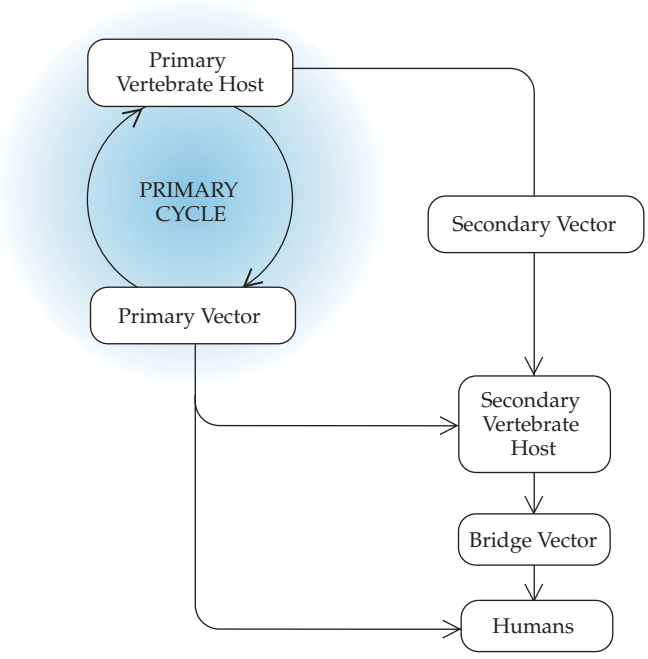


Figure 1 **Generalized arbovirus maintenance cycle.**

Diagnosis DHF is characterized by sudden onset of fever, usually lasting 2 to 7 days, and nonspecific signs and symptoms.[1] The critical stage of DHF occurs between 24 hours before and 24 hours after the patient's temperature falls to or below normal. During this time, hemorrhagic manifestations usually occur, and signs of circulatory failure may appear. The patient may become restless or lethargic, experience acute abdominal pain, and have cold extremities and oliguria, usually on or after the third day of illness. Clinical laboratory tests at this time will show thrombocytopenia (platelet count < 100,000/mm^3), a low serum total protein level, a low albumin level, and a rise in hematocrit secondary to plasma leakage from the vascular compartment. Another indication of vascular leakage is pleural effusion. Loss of intravascular volume may result in hypovolemia, shock, and death if not corrected. The most common hemorrhagic manifestations are skin hemorrhages, but epistaxis, bleeding gums, gastrointestinal hemorrhage, and hematuria may occur.

Treatment and prevention Early diagnosis and prompt management with fluid replacement therapy can substantially reduce case-fatality rates.[10] Initial management and treatment decisions should not be delayed pending results of serologic tests. Clinical laboratory tests should be used to monitor vascular leakage.[10]

There is no licensed vaccine for dengue/DHF, although significant progress is being made toward the development of live attenuated and recombinant candidate vaccines using infectious clone technology.[11,12] Currently, disease prevention depends exclusively on mosquito control and personal protective measures such as the use of mosquito repellents.

Yellow Fever

Epidemiology Yellow fever virus (family Flaviviridae, genus *Flavivirus*) is believed to have originated in Africa. The disease is now present in tropical America and Africa but does

not occur in Asia. Like dengue, yellow fever virus has two transmission cycles: jungle and urban. The jungle or forest transmission cycle involves canopy-dwelling mosquitoes and monkeys. The urban cycle involves humans as the vertebrate host and *Ae. aegypti* as the principal vector. In the past 30 years, *Ae. aegypti* has reinvaded Central and South America, putting the American tropics at the highest risk for urban epidemics of yellow fever in over 60 years. Epidemics in Africa often occur in moist savanna regions; forest or peridomestic *Aedes* mosquitoes and humans are the viremic hosts. In dry areas and urban centers where water storage practices promote the breeding of domestic *Ae. aegypti,* this mosquito is responsible for epidemic transmission. Several hundred thousand people are infected yearly, and outbreaks are frequent. Cases among unvaccinated travelers are rare; however, since 1996, six travelers have died in the United States and Europe of yellow fever acquired in South America and Africa.

Diagnosis Yellow fever varies from an inapparent infection to a deadly fulminating hemorrhagic disease. Three clinical stages are commonly recognized: infection, remission, and intoxication.[13,14] After an incubation period of 3 to 10 days, the period of infection begins with sudden onset of fever, rigors, headache, and backache. In severe cases, the patient is intensely ill and restless, with flushed face, swollen lips, bright-red tongue, congested conjunctivae, and bleeding. There may be bradycardia relative to fever (Faget sign). This stage is followed after 2 to 3 days by a brief period of remission. Remission is often not obvious. The period of intoxication occurs on the third to sixth day after illness onset in about 15% of patients. This period consists of moderate or severe disease with jaundice. Fever returns with relative bradycardia, along with nausea, vomiting, a hemorrhagic diathesis, hypotension, albuminuria, oliguria, and anuria.

Most patients with severe disease will have leukopenia, thrombocytopenia, elevated levels of serum creatinine and liver enzymes, and coagulation defects. The jaundice, which gives the disease its name, is generally apparent only in convalescing patients. In fatal cases, death usually occurs within 7 to 10 days; case-fatality rates vary widely, but they can exceed 50% in clinically ill patients.[15] At autopsy, the organs most affected are the liver, spleen, kidneys, and heart. Typically, the liver shows midzonal hyaline necrosis and Councilman inclusion bodies.

The initial diagnosis of yellow fever should be based on clinical signs and symptoms and on epidemiologic information such as vaccination history and recent travel to an endemic area [*see CE:VII Health Advice for International Travelers*]. Laboratory diagnosis is made by detection of virus, viral antigen, or viral nucleic acid or by serologic examination. Virus is often isolated from blood during the first 4 days of illness. Serologic tests may be negative during the first week of illness. Diagnostic testing for yellow fever may be obtained by request to the CDC through state and local health departments.

Treatment and prevention Treatment is supportive. Yellow fever is an international reportable disease, and immunization is required for travelers to many countries of sub-Saharan Africa and tropical America. The live, attenuated 17D vaccine, delivered as a single 0.5 ml subcutaneous dose, is highly effective. Immunity is probably lifelong, but for travel certification, revaccination is recommended every 10 years. Information about indications for yellow fever vaccine and requirements for international travel are available at www.cdc.gov/travel/reference.htm.[16,17] For patients who require immunization, the locations of designated yellow fever vaccination centers can be obtained through local health departments. The CDC also provides an extensive list of yellow fever vaccination clinics on its Web site.

Although the 17D vaccine is one of the safest vaccines, rare cases of severe and fatal infection from vaccination have been reported. The elderly may experience a higher incidence of serious adverse events.[18,19] Persons with documented egg allergy should not be immunized or should be skin-tested with the vaccine. The vaccine must not be given to children younger than 6 months, in whom there is a risk of postvaccinial encephalitis, and it is best to delay vaccination until 9 months of age. On theoretical grounds, persons with immunosuppression, including those with clinical AIDS, should not be immunized. Immunization during pregnancy is generally contraindicated.

Other Flaviviruses

Other hemorrhagic diseases caused by flaviviruses include Kyasanur Forest disease in India and Saudi Arabia and Omsk hemorrhagic fever in Russia [*see Table 1*]. Both are tick-borne diseases and are rare in comparison with disease from mosquito-borne flaviviruses.

VIRUSES OF THE FAMILY BUNYAVIRIDAE

Crimean-Congo Hemorrhagic Fever

Epidemiology and etiology Crimean-Congo hemorrhagic fever (CCHF) virus (family Bunyaviridae, genus *Nairovirus*) is transmitted by ticks, primarily of the genus *Hyalomma*, over a broad geographic range that includes sub-Saharan Africa, eastern Europe and Russia, the Middle East, and western China.[14,20] Humans become infected after a tick bite; after parenteral exposure to, or contact with, blood from acutely ill patients[20]; or, on occasion, after the slaughter of sick domestic animals. Disease may be seasonal in humans, reflecting the natural abundance of ticks.

Table 3 Arenaviruses That Cause Significant Human Illness

Virus	*Disease*	*Geographic Distribution*	*Primary Host in Nature*
Lymphocytic choriomeningitis virus	Lymphocytic choriomeningitis	Americas, Europe, parts of Asia	*Mus musculus* (house mouse)
Lassa virus	Lassa fever	West Africa	*Mastomys* species (rodent)
Junin virus	Argentine hemorrhagic fever	Argentina	*Calomys musculinus* (rodent)
Machupo virus	Bolivian hemorrhagic fever	Bolivia	*Calomys callosus* (rodent)
Guanarito virus	Venezuelan hemorrhagic fever	Venezuela	*Zygodontomys brevicauda* (rodent) *Sigmodon alstoni* (?) (rodent)
Sabia virus	Hemorrhagic fever	Brazil	Unknown

Diagnosis In the United States, CCHF should be suspected in a traveler who falls abruptly ill less than 2 weeks after returning from an endemic area. The incubation period for CCHF is 3 to 7 days for nosocomial infections and up to 12 days for those infected by a tick bite, followed by abrupt onset of fever, headache, myalgia, weakness, nausea, and vomiting. This initial phase may last 2 to 3 days, followed by remission of several hours' duration in one third to two thirds of patients. The second phase of illness is associated with hemorrhagic manifestations, which may last from 3 days to as long as 10 days and includes, most commonly, petechiae over the chest and abdomen, epistaxis, ecchymoses, bleeding from puncture sites, melena, and hematuria. Those surviving the hemorrhagic phase enter a convalescent phase characterized by normalization of fever, cessation of hemorrhages, occasionally transient hair loss, and prolonged fatigue and dizziness.[20] Mortality ranges from 9% to 40%.

Treatment and prevention Treatment is supportive and focuses on the hemorrhagic manifestations. The antiviral drug ribavirin has been administered to a limited number of patients with apparent success.[21] Persons in endemic areas should take protective measures to prevent exposure to infected ticks, and barrier methods should be used in hospitals where patients with suspected CCHF are being treated.

Hantavirus Infections

Hantaviruses (family Bunyaviridae, genus *Hantavirus*) are rodent-borne viruses causing human diseases known as hemorrhagic fever with renal syndrome (HFRS) in Europe and Asia and as hantavirus pulmonary syndrome (HPS) in the Americas.[22] The prototype hantavirus is Hantaan virus, which causes epidemic hemorrhagic fever in China and Korea, but many different hantaviruses are now recognized. Hantaviruses are maintained in nature by chronic infection of rodent hosts, and each is principally associated with a specific rodent host [*see Table 2*]. Antibodies against hantaviruses are also present in domestic and wild animals such as cats, dogs, pigs, cattle, and deer. German investigators are examining whether hantaviruses can be transmitted from rats to cattle, because the incidental infection of species other than rodents has the potential to influence the pathogenicity and virulence of the virus.[23]

Humans are infected after aerosol exposure to infectious excreta or occasionally by bites. Apparent human-to-human transmission was noted in one outbreak with the Andes virus.[24]

The incidence of hantavirus infections in humans often fluctuates with rodent densities and human activities that increase contact with contaminated materials. Commonly infected groups are military personnel on field maneuvers, shepherds, woodcutters, campers, and others involved in outdoor activities.

Hemorrhagic fever with renal syndrome Along with Hantaan virus, other hantaviruses causing HFRS are Seoul virus, Puumala virus, and Dobrava/Belgrade virus. Classic HFRS caused by Hantaan or Dobrava virus infection has a variable incubation period and is characterized by five phases: (1) a febrile phase of 3 to 7 days, with fever, malaise, headache, abdominal pain, nausea, vomiting, facial flushing, petechiae, and conjunctival hemorrhage; (2) a hypotensive phase of a few hours to 3 days, when hypotension, shock, blurred vision, hemorrhagic signs, and a drop in blood pressure occur; (3) an oliguric phase of 3 to 7 days, during which oliguria or anuria predominates and hemorrhagic manifestations may worsen; (4) a diuretic phase of days to weeks, when polyuria predominates; and (5) a prolonged convalescent phase of weeks to months.[22] Mortality in classic HFRS results from shock, multiorgan hypoperfusion, or uremia and ranges from 1% to 10%. Infection with Puumala virus or Seoul virus produces milder disease, with lower mortality. Serology is the main diagnostic tool, but it should be used only for confirmatory retrospective diagnosis.[22]

Early treatment with ribavirin may reduce hemorrhage, renal failure, and mortality in HFRS.[25] Commercially available inactivated vaccines for Hantaan virus or bivalent Hantaan/Seoul virus vaccines are produced in Korea and China; other vaccines are in development.[26] Control of HFRS relies on reduction of human-rodent contact through good sanitation and waste management, rodent control, and making buildings rodent proof.

Hantavirus pulmonary syndrome Since its discovery in 1993 during an outbreak in the southwestern United States, several hundred cases of HPS have been identified in the Americas.[27] The Sin Nombre virus, which is carried by deer mice, caused the initial outbreak; subsequently, a number of related hantaviruses hosted by other sigmodontine rodents have been identified as a cause of human disease [*see Table 2*].

Diagnosis HPS should be suspected on the basis of the clinical picture, clinical laboratory results, and radiologic findings; confirmation of diagnosis is by serologic testing. Diagnostic information about HPS, including instructions on submitting suspected HPS specimens for serologic testing, can be obtained from the CDC at www.cdc.gov/ncidod/diseases/hanta/hps/index.htm.

The incubation period of HPS probably ranges from 9 to 33 days. Clinical disease can be divided into four phases: febrile, cardiopulmonary, diuretic, and convalescent.[28] The febrile phase, typically lasting 3 to 5 days, is characterized by fever, myalgia, and malaise. Headache, dizziness, anorexia, nausea, vomiting, and diarrhea may occur. The cardiopulmonary phase is marked by pulmonary edema and shock. Once pulmonary edema develops, the rapid onset of circulatory compromise and hypoxia often leads to death. During the diuretic phase, pulmonary edema clears, along with resolution of fever and shock. The convalescent phase may last several months; complete recovery is the rule. Renal insufficiency has been reported with the Sin Nombre virus and is a more constant feature of many of the newly recognized American hantaviruses causing HPS, suggesting that HPS and HFRS are not as clinically distinct as previously thought. Thrombocytopenia is almost universally present. Mortality is approximately 40% but varies with the infecting virus.

Treatment of patients with HPS remains supportive. Early intensive care management is important, with prompt correction of electrolyte, pulmonary, and hemodynamic abnormalities. Vaccines are in development.[26] Despite its in vitro activity against Sin Nombre virus, ribavirin failed to produce a dramatic reduction in case fatality in an open-label trial.[29]

Rift Valley Fever

Epidemiology and etiology Rift Valley fever (RVF) is caused by the RVF virus (family Bunyaviridae, genus *Phlebovirus*). First described during a fatal epizootic in sheep that occurred in the Rift Valley of Kenya in 1931, RVF is a mosquito-borne disease affecting domestic ungulates, especially goats and sheep.[14] Large epizootics occur during periods of heavy rainfall. Humans become infected through the bite of an infected mosquito, by infec-

Table 4 Viral Zoonoses Endemic in the United States

Disease	*Region*	*Vector/Host*	*Clinical Manifestations*
Colorado tick fever	Mountainous areas of western states	Wood ticks	Flulike illness, rash, leukopenia
Eastern equine encephalitis	Focal locations along eastern seaboard, Gulf Coast, and some midwestern states	Mosquitoes	Mild flulike illness to encephalitis
Hantavirus pulmonary syndrome	Most prevalent in southwestern states	Deer mice and other rodents (aerosol exposure to infected excreta, or bites)	Febrile, cardiopulmonary, diuretic, and convalescent phases
La Crosse virus	Widespread; most prevalent in rural upper Midwest	Mosquitoes	Mild flulike illness to encephalitis, often with focal seizures
Powassan virus encephalitis	Northeastern states	*Ixodes* ticks	Encephalitis, with localizing neurologic signs and convulsions
Rabies	All except Hawaii	Wild carnivores, bats	Encephalomyelitis
St. Louis encephalitis	All lower 48 states; epidemics in Midwest and Southeast	Mosquitoes	Febrile headache to encephalitis
West Nile virus infection	Eastern states	Mosquitoes	Febrile illness to encephalitis
Western equine encephalitis	Western states	Mosquitoes	Mild flulike illness to encephalitis

tious aerosols when sick animals are slaughtered, or by occupational exposure (e.g., veterinarians attending to infected animals). Large human outbreaks have occurred in sub-Saharan Africa, Egypt, and more recently on the Saudi Arabian peninsula.[14] RVF virus is a potential agent of bioterrorism.

Diagnosis The incubation period in humans is 2 to 6 days. This is followed by abrupt onset of fever, headache, chills, and malaise. Uncomplicated illness usually resolves within 2 to 3 days. Retinitis, hemorrhagic fever, and encephalitis occur in rare instances.

Treatment and prevention Treatment is supportive, although the antiviral drug ribavirin has been effective in the treatment of related viruses of the family Bunyaviridae and deserves further investigation. Inactivated and live attenuated vaccines have been produced for use in domestic animals and appear to be efficacious. No vaccine is commercially available for use in humans, although an experimental inactivated vaccine has been used to protect select populations at risk for laboratory or occupational exposure.[14]

VIRUSES OF THE FAMILY ARENAVIRIDAE

Significant human illnesses caused by Arenaviridae viruses include lymphocytic choriomeningitis; Lassa fever; and Argentine, Bolivian, and Venezuelan hemorrhagic fever. Arenaviruses are transmitted directly to humans after close rodent-human contact, such as touching objects or eating food contaminated with rodent excreta [*see Table 3*]. Human-to-human transmission of Lassa fever has occurred. Several of the arenaviruses are considered potential agents of bioterrorism.

Diagnosis

Arenaviral infections typically begin with fever, malaise, headache, and GI symptoms. More severe cases may involve the heart, lungs, liver, and kidneys. Fulminant and often fatal hemorrhagic shock occurs in a minority of cases (e.g., about 20% of hospitalized Lassa fever patients). Lassa fever is especially severe in pregnant women, often causing death in both mother and fetus. Hearing loss (of varying degrees) is a common sequela of Lassa fever, even in mild cases.

Several cases of Lassa fever have been imported into Europe, Japan, and the United States in recent years. Consequently, the clinician should consider the diagnosis in patients who experience a febrile illness within 3 weeks after travel to endemic countries in West Africa (i.e., Nigeria, Guinea, Liberia, and Sierra Leone). South American arenaviruses should be considered in the differential diagnosis of hemorrhagic fever in patients with a history of travel to that region. Because its symptoms are varied and nonspecific, arenavirus infection can be confirmed by serology, most commonly with ELISA, which can be requested from the CDC through state and local health departments.

Treatment and Prevention

The antiviral drug ribavirin reduces mortality from Lassa fever, especially if started within the first 6 days after the onset of fever. Ribavirin may be of value in the treatment of other arenaviral infections as well, although efficacy cannot be fully validated because of a lack of clinical experience.[30] A live, attenuated vaccine of proven efficacy has been developed for Junin virus and is available in Argentina[31]; vaccine candidates for Lassa fever are in development.

VIRUSES OF THE FAMILY FILOVIRIDAE

Marburg and Ebola viruses are among the most severe and mysterious viral pathogens to emerge in the 20th century.[32] Filoviruses share common morphology as long, pleomorphic rods but are antigenically and genetically distinct. Current understanding of these agents is largely restricted to investigations of human outbreaks and limited experimental studies conducted at the few laboratories worldwide that are able to safely handle filoviruses.

There is no vaccine or effective chemotherapy for any known filovirus. Marburg and Ebola viruses are considered potential agents of bioterrorism.

Marburg Virus

African green monkeys, *Cercopithecus aethiops,* imported from Uganda for use in research and vaccine production, were the source of the initial 31-person outbreak that led to the discovery of the Marburg virus in 1967. Human fatality was 23%.[32] In 2005, the World Health Organization confirmed an outbreak of Marburg virus infection in 124 persons in Angola.[33] Information for travelers to Angola is available from the CDC at http://www.dcd.gov/travel/other/marburg_vhf_angola_2005.htm. The ecology of Marburg virus remains virtually unknown. Marburg virus disease presents as an acute febrile illness; it can progress within 6 to 8 days to severe hemorrhagic manifestations. Clinical manifestations include fever, chills, headache, myalgia, maculopapular rash, nausea, vomiting, chest pain, and abdominal pain. Signs and symptoms can become increasingly more severe. Clinicians should consider the diagnosis of Marburg virus for febrile patients who, within 10 days before onset of fever, have traveled in northern Angola, had direct contact with blood or other body fluids of a person suspected of having hemorrhagic fever, or have worked in a laboratory that handles hemorrhagic fever viruses. No vaccine or curative treatment is available; treatment is supportive.

Ebola Virus

Epidemiology At least 1,000 persons have died from Ebola virus infection since its discovery in Sudan in 1976. Of the four known genetic subtypes of Ebola virus—Zaire, Côte d'Ivoire, Sudan, and Ebola-Reston—the first three have been associated with human disease in West and central Africa.[32] Ebola-Reston was discovered in macaques imported from the Philippines for medical research.[34] Occupational exposure to Ebola-Reston from nonhuman primates is infrequent and results in asymptomatic infection.[35]

The natural history of Ebola virus remains a mystery. Outbreaks of Ebola hemorrhagic fever are associated most often with the introduction of the virus into the community by one infected person, followed by dissemination by person-to-person transmission, often within medical facilities.[32,36]

Diagnosis Ebola virus infection has an incubation period of approximately 6 days, which is commonly followed by two clinical phases.[37] Early symptoms include fever, asthenia, diarrhea, nausea and vomiting, anorexia, abdominal pain, headaches, arthralgia, and back pain. Bilateral conjunctivitis, nonpruritic rash, and sore throat with odynophagia, when present, suggest Ebola virus infection. The second phase, characterized by hemorrhagic manifestations, neuropsychiatric abnormalities, and oligoanuria, portends a worse outcome. In recent outbreaks, bleeding occurred in a minority of patients.[32,36] Mortality during outbreaks typically exceeds 50%.

The clinical diagnosis is challenging, because the presentation is nonspecific. To exclude other infections, patients with clinical manifestations consistent with Ebola virus infection should have a blood smear examination for malaria, a blood culture, and a stool culture if they have bloody diarrhea. ELISA, PCR, and virus isolation can be used to confirm Ebola virus infection within a few days of the onset of symptoms.[38] Detailed diagnostic information is available from the CDC at http://www.cdc.gov/ncidod/dvrd/spb/mnpages/dispages/ebola.htm.

Treatment and prevention Treatment is supportive. Outbreak control rests on initiation of case finding, case isolation, and other infection-control practices, including barrier nursing procedures.[32,36]

Encephalitis

Viral encephalitis is caused by a number of arboviruses belonging to the families *Flaviviridae, Togaviridae, Bunyaviridae,* and *Reoviridae,* as well as other zoonotic viruses. Specific laboratory diagnosis of encephalitis from zoonotic viruses requires special serologic tests, such as hemagglutination-inhibition, complement fixation, and neutralization tests. ELISAs are used to detect virus-specific IgM or IgG antibody in serum or CSF samples. As with the flaviviruses, serologic tests must be cautiously interpreted, because there is considerable serologic cross-reactivity among many of these viruses of the same genus. Viruses can also be isolated and detected by PCR or immunohistochemistry in autopsy tissue. Assistance with serologic diagnosis can be obtained from the CDC at www.cdc.gov/ncidod/dvbid/index.htm and through state and local health departments.

VIRUSES OF THE FAMILY BUNYAVIRIDAE

La Crosse Encephalitis

Epidemiology La Crosse (LAC) virus (family Bunyaviridae, genus *Bunyavirus*) is the most pathogenic member of the California encephalitis serogroup, which includes the California encephalitis, trivittatus, snowshoe hare, and Jamestown Canyon viruses. LAC is maintained in a cycle involving *Ae. triseriatus* mosquitoes and a number of mammalian hosts, including the eastern chipmunk, tree squirrels, and foxes.

Human infections occur in the central and eastern United States, mostly as sporadic cases in school-age children from July through September [*see Table 4*].[39]

Diagnosis Most infections are asymptomatic. After an incubation period of 3 to 7 days, headache, fever, and vomiting develop. Seizures are a presenting finding in about half of cases, and focal neurologic abnormalities in about one fifth.[39] The combination of fever, focal signs, and focal seizures may mimic herpes simplex encephalitis. Mortality is 1%. About 10% of children have residual neurologic sequelae, including focal neurologic deficits and decreased intelligence. Diagnosis can be confirmed serologically by the CDC on the request of state and local health departments.

Treatment and prevention Treatment is supportive; management of cerebral edema and seizures is important [*see 11:XVI Acute Viral Central Nervous System Diseases*]. Ribavirin has been used, but efficacy is unproved.[39] Prevention rests on avoidance of mosquito bites.

VIRUSES OF THE FAMILY FLAVIVIRIDAE

Japanese Encephalitis

Epidemiology Japanese encephalitis (JE) is the most important global cause of arboviral encephalitis; 30,000 to 45,000 cases are reported annually. JE virus (family Flaviviridae, genus *Flavivirus*) is widespread throughout Asia. In recent years, the disease has been detected in Australia and other areas in the Pacific region. Epidemics occur in late summer in temperate regions, but the virus is enzootic and occurs throughout the year in many tropical areas of Asia. JE virus is maintained in a natural enzoot-

ic cycle involving *Culex* mosquitoes and water birds. The virus is transmitted to humans by *Culex* mosquitoes, primarily *C. tritaeniorhynchus* and related species, which breed in rice fields. Pigs are the primary amplifying hosts in the peridomestic environment.

Diagnosis Only about one in 250 infections results in symptomatic illness. The incubation period of JE is 5 to 14 days. Symptomatic illness is primarily seen in children. Mild clinical illness, such as aseptic meningitis and simple febrile illness with headache, usually goes undetected. In severe cases, the onset of symptoms is usually sudden, with fever, headache, and vomiting. The illness resolves in 5 to 7 days if there is no central nervous system involvement. Patients with CNS involvement commonly are lethargic, with expressionless faces, and have sensory and motor disturbances affecting their speech, eyes, and limbs. They may have confusion and delirium progressing to coma; in children, convulsions are sometimes a presenting sign. Weakness and paralysis may affect any part of the body. Neck rigidity and a positive Kernig sign are found, and reflexes are abnormal. Signs of extrapyramidal involvement are characteristic. Initial leukocytosis is followed by leukopenia. Mortality is 5% to 30%, with higher case-fatality rates in young children. Approximately one third of patients who recover have neurologic sequelae. The diagnosis can be confirmed serologically by the CDC on request of state and local health departments.

Prevention A formalin-inactivated mouse brain vaccine prepared with the Nakayama strain of JE virus is used internationally. Vaccination is recommended to residents in JE-endemic areas and to certain travelers to those areas. The risk to travelers is generally low, but vaccination is recommended for visitors to endemic or epidemic areas during the transmission season, especially when potential exposure will be prolonged and when there is a high likelihood of exposure to vectors.[40] Treatment is supportive. Interferon alfa was not effective against JE in a double-blind, placebo-controlled trial.[41] Further information on JE is available at www.cdc.gov/ncidod/dvbid/jencephalitis/index.htm. Mosquito control and improved animal husbandry and rice-growing practices are needed to decrease transmission risk in endemic areas.

Murray Valley Encephalitis

Epidemiology Murray Valley encephalitis (MVE) virus (family Flaviviridae, genus *Flavivirus*) was first isolated in 1951. MVE occurs only in Australia and New Guinea.[42] Like other flaviviruses, MVE virus is believed to be maintained in a natural cycle involving water birds and *Culex* mosquitoes. Viremia has not been documented in humans, who are likely dead-end hosts.

Diagnosis Only one in 1,000 to 2,000 infections results in clinical illness. Clinical illness resembles JE. Illness is characterized by the sudden onset of fever, headache, nausea and vomiting, anorexia, and myalgias, followed by drowsiness, malaise, irritability, mental confusion, and meningismus. In severe cases, there may be hyperactive reflexes, spastic paresis, convulsions, coma, and death. Of patients with neurologic disease, approximately one third die and one quarter have residual neurologic deficits.

Prevention There is no vaccine for MVE virus. Prevention relies on mosquito control and avoidance of mosquito bites.

Saint Louis Encephalitis

Epidemiology St. Louis encephalitis (SLE) virus (family Flaviviridae, genus *Flavivirus)* is prevalent throughout the Western Hemisphere from Canada to Argentina. In North America, the infection is maintained between wild birds and *Culex* mosquitoes. Although clinical illness has been sporadically reported throughout much of this region, most infections occur in North America during sporadic epidemics in the Midwest and Southeast.

Diagnosis The ratio of infection to clinical illness is high, ranging from 800 to 1 in children younger than 10 years to 85 to 1 in persons older than 60 years. Illness ranges from fever with headache to aseptic meningitis to encephalitis. Advanced age is the strongest risk factor for both symptomatic disease and severe encephalopathy. SLE should be considered in the differential diagnosis of adult viral encephalitis cases during the summer months in the United States. After an incubation period of 4 to 21 days, the illness begins with fever, headache, chills, nausea, and dysuria. Within 1 to 4 days, CNS signs appear, with meningismus, tremor, abnormal reflexes, ataxia, cranial nerve palsies, convulsions (especially in children), stupor, and coma. About 25% of very young infants have residual mental deficits, personality changes, muscle weakness, and paralysis. Overall, the case-fatality rate is about 6%, but the disease is generally milder in children (case-fatality rate of those younger than 5 years is 1%). The diagnosis can be confirmed serologically by the CDC on request of state and local health departments.

Treatment and prevention Treatment is supportive; no specific therapy is available. A pilot study indicated that early initiation of therapy with interferon alfa-2b may reduce the severity and duration of complications such as quadriplegia and quadriparesis, but a randomized trial is required to better assess the efficacy of this therapy.[43] No vaccine is available. Prevention is aimed at mosquito-bite avoidance and mosquito abatement.

Tick-Borne Encephalitis

Epidemiology and etiology Tick-borne encephalitis (TBE) is caused by two closely related viruses of the family Flaviviridae, genus *Flavivirus.*[44] The eastern subtype of the TBE virus is transmitted by *Ixodes persulcatus* and causes Russian spring-summer encephalitis, which occurs from Eastern Europe to China. The western subtype is transmitted by *I. ricinus* and causes Central European encephalitis, which occurs from Scandinavia in the north to Greece and Serbia and Montenegro in the south. Of the two subtypes, Russian spring-summer encephalitis is the more severe infection, having a mortality of 5% to 20%, compared with less than 2% for the western subtype. Both viruses are maintained in natural cycles involving a variety of mammals and ticks. Human exposure occurs through work or recreational activities in the spring and summer months in temperate zones and in fall and winter in the Mediterranean, when the ticks are most active. Infection may also occur through the ingestion of raw milk or cheese from cows, sheep, or goats.

Diagnosis The incubation period of TBE is usually 7 to 14 days. The western subtype typically produces a biphasic illness. Infection usually presents as a mild, influenzalike illness lasting 2 to 7 days, followed by an afebrile or relatively asymptomatic period lasting 2 to 10 days. Approximately one third of patients then develop higher fevers with aseptic meningitis or menin-

goencephalitis. The eastern subtype usually progresses without an asymptomatic phase. Permanent paresis develops in 2% to 10% of patients with the western subtype and in 10% to 25% of patients with the eastern subtype. Rarely, cases occur in United States citizens who visit enzootic areas. Infections can be confirmed serologically by the CDC on request of state and local health departments.

Treatment and prevention Treatment of TBE is supportive. Inactivated vaccines against both eastern and western subtypes of TBE viruses are available in Europe, but they are not commercially available in the United States[44]; candidate vaccines are in development in the United States.[45] Prevention strategies include avoiding tick bites and pasteurizing milk.

West Nile Virus

Epidemiology and etiology West Nile virus (family Flaviviridae, genus *Flavivirus*), first isolated in 1937 in the West Nile district of Uganda, has historically had a wide geographic distribution in Africa, Asia, the Middle East, and Europe.[46] The virus was first recognized in the Western Hemisphere during an epizootic among birds and horses and a human encephalitis outbreak in the New York City area in 1999.[47] From the 1950s to the 1970s, human outbreaks were reported infrequently, mostly in the Middle East. However, since the mid-1990s, outbreaks of neurologic disease in humans and horses, with an increase in death rates, may have marked the evolution of a new West Nile virus variant.[46,48] By 2002, the virus's known geographic distribution extended to southeastern Canada, the Grand Cayman Islands, and throughout the eastern United States.[48,49] Up-to-date maps of the virus's spread in the United States are available at www.cdc.gov/ncidod/dvbid/westnile/surv&control.htm. The virus has a natural maintenance cycle involving wild birds and *Culex* mosquitoes and is thought to spread via migrating birds.[46,48] In temperate climates, the incidence of infection peaks during late summer and early fall; however, year-round transmission is possible in more tropical areas. Human infections result primarily from infectious mosquito bites; however, in 2002, five new modes of transmission were recognized: blood product transfusion, organ transplantation, transplacental transmission, breast-feeding, and occupational exposure in laboratory workers.[48]

Diagnosis The incubation period of West Nile virus ranges from 3 to 14 days. Serologic surveys during recent outbreaks suggest that approximately 20% of persons who are infected develop a systemic febrile illness.[50] Common symptoms are fever; headache; myalgia; GI complaints; and an erythematous macular, papular, or morbilliform skin rash.[46,47] Lymphadenopathy may be present, but it has been reported less frequently in recent outbreaks. Overall, fewer than 1% of infected persons develop encephalitis.[50] Older persons are at increased risk for meningitis or encephalitis; once these complications develop, such individuals have a higher case-fatality rate and a higher incidence of residual neurologic deficits.[47] Of note, muscle weakness and flaccid paralysis, when present, may provide a clue to a West Nile virus etiology. Overall, case-fatality rates in severe cases are about 10%. The diagnosis is most efficiently confirmed serologically; testing can be obtained through state and local health departments.

Treatment and prevention Treatment of West Nile virus infection is supportive. Prevention relies on mosquito control and protection from mosquito bites. There is no human vaccine for West Nile virus infection. An equine vaccine is available.

Powassan Virus

Powassan virus (family Flaviviridae, genus *Flavivirus*) is related to the Eastern Hemisphere's tick-borne encephalitis viruses. It was thought to be a rare cause of encephalitis in eastern Canada and the northern United States; however, West Nile virus surveillance has increased the recognition of this pathogen.[51] The case-fatality rate is 5% to 10%, with a high incidence of residual neurologic dysfunction in survivors. Serologic surveys indicate an antibody prevalence of 1% to 4%. The virus is transmitted between *Ixodes* ticks and rodents; humans become infected via tick bites. The clinical features are those of viral encephalitis, with localizing neurologic signs and convulsions.[44] There is no specific treatment or vaccine.

VIRUSES OF THE FAMILY TOGAVIRIDAE

Eastern Equine Encephalitis

Epidemiology Eastern equine encephalitis (EEE) virus (family Togaviridae, genus *Alphavirus*) is widely distributed throughout North, Central, and South America and the Caribbean; however, little is known about the epidemiology of EEE outside North America. In the United States, human infections are usually sporadic, and small outbreaks occur each summer, mostly in the upper Midwest and along the Atlantic and Gulf coasts. In North America, wild birds and *Culiseta melanura* (a mosquito found in swamp areas that support cedar, red maple, and loblolly bay trees) maintain the virus. In central Alabama, high rates of EEE virus infection were found in *Uranotaenia sapphirina*, a mosquito that commonly feeds on amphibians and reptiles, which suggests that species other than birds may serve as a reservoir for EEE in hardwood swamps in the southeastern United States and elsewhere.[52]

Diagnosis The incubation period of EEE exceeds 1 week; onset is abrupt, with high fever. About 2% of infected adults and 6% of infected children develop encephalitis. EEE causes the most severe of the arboviral encephalitides, with a mortality of 50% to 75%. Symptoms and signs include dizziness, decreasing level of consciousness, tremors, seizures, and focal neurologic signs. Death can occur within 3 to 5 days after onset. Sequelae, which are common in nonfatal encephalitis, include convulsions, paralysis, and mental retardation. Illness from EEE in South America is less severe. Infection can be confirmed serologically by the CDC on request of state and local health departments.

Treatment and prevention No specific treatment is available. Prevention focuses on avoidance of mosquito bites and mosquito control in suburban areas. Inactivated vaccines have been used successfully in horses, and an inactivated vaccine has been used in laboratory workers or others at high risk of exposure, but it is not commercially available.

EEE is a potential agent of bioterrorism through the aerosol route.

Venezuelan Equine Encephalitis

In the Venezuelan equine encephalitis (VEE) virus (family Togaviridae, genus *Alphavirus*) complex, six subtypes (I to VI) have been identified. Five antigenic variants exist in subtype I (IAB, IC, ID, IE, and IF). These subtypes and variants are classified as

epizootic or enzootic on the basis of their apparent virulence and epidemiology. Epizootic variants of subtype I (IAB and IC) cause equine epizootics and are associated with more severe human disease. Enzootic strains (ID to IF, II [Everglades], III [Macambo, Tonate, Paramana], IV [Pixuna], V [Cabassou], VI [unnamed]) do not cause epizootics in horses, but they may produce sporadic disease in man. Epizootic strains are transmitted by many different types of mosquitoes; enzootic strains are transmitted by culicine mosquitoes. VEE has a widespread geographic distribution, from Florida to South America, where it is an important veterinary and public health problem. Focal outbreaks occur periodically, but occasionally, large regional epizootics occur, with thousands of equine and human infections.

VEE is infectious via aerosols, making it an occupational risk to certain laboratory workers and a potential agent of bioterrorism.

Diagnosis After an incubation period of 1 to 6 days, there is a brief febrile illness of sudden onset, characterized by malaise, nausea or vomiting, headache, and myalgia.[53] Fewer than 0.5% of adults and fewer than 4% of children develop encephalitis. Long-term sequelae and fatalities are uncommon.

Treatment and prevention Treatment is supportive. Effective prevention of both human and equine disease can be accomplished by immunizing horses and other equine animals, which serve as the primary amplification host for the epizootic VEE viruses and without which there would be little human disease. During epidemics, mosquito vectors can be controlled by insecticides. Live attenuated and inactivated vaccines have been used for laboratory workers; however, human vaccines are not commercially available.

Western Equine Encephalitis

Western equine encephalitis (WEE) virus (family Togaviridae, genus *Alphavirus*) is a complex of closely related viruses found in North and South America. Flooding, which increases breeding of culicine mosquitoes, may precipitate summer outbreaks of WEE. Large outbreaks in humans and horses occurred in the western United States in the 1950s and 1960s; however, a declining horse population, equine vaccination, and improved vector control have reduced disease incidence. The younger the patient, the greater the likelihood of symptomatic infection: the ratio of asymptomatic to symptomatic infection is less than 1:1,000 in adults but increases to 1:1 in infants.

Diagnosis After an incubation period of about 7 days, headache, vomiting, stiff neck, and backache are typical in WEE. In children, restlessness and irritability are seen, and convulsions are common. Neurologic sequelae are relatively common in infants, but they are rare in older children and adults. The case-fatality rate is 3% to 7%. The diagnosis can be confirmed serologically by the CDC on request of state and local health departments.

Treatment and prevention No specific treatment is available for WEE. Prevention focuses on mosquito control and personal measures to avoid mosquito bites. Inactivated vaccine is available for horses. Although inactivated vaccine has been used for laboratory staff and others at high risk for exposure, it is not commercially available for use in humans. WEE is a potential agent of bioterrorism through the aerosol route.

Rabies

Epidemiology The rabies virus (family Rhabdoviridae, genus *Lyssavirus*) occurs worldwide. Dogs remain the major source of human rabies worldwide. In the United States, however, vaccination has sharply limited canine rabies, and consequently, wildlife rabies has increased in importance. About 90% of all reported cases of animal rabies in the United States now occur in wildlife, particularly wild carnivores (e.g., raccoons, skunks, foxes, coyotes, and bobcats) and bats.[54,55]

The major wildlife reservoir for rabies in the United States is the raccoon, which accounts for 37% of all reported cases of rabies in animals; skunks (29%), bats (17%), and foxes (6%) represent the other major reservoirs for animal rabies.[54] Raccoon rabies is the most prevalent in the eastern states; skunk rabies is predominant in the central and western states.[54] Rodents (e.g., squirrels, hamsters, guinea pigs, gerbils, chipmunks, rats, and mice) and lagomorphs (rabbits and hares) are rarely infected and have not been identified as sources of human rabies in the United States.[55] In the United States since 1990, indigenous rabies virus variants associated with insectivorous bats and foreign canine rabies virus variants have accounted for 30 of the 32 human cases. Although 74% of the 32 cases since 1990 have been attributed to bat-associated variants of the virus, a history of a bite was established in only three cases.[56]

Etiology and pathogenesis In most cases of rabies, an infected animal inoculates saliva containing rabies virus into the patient, and the virus may replicate in muscle cells near the bite. After replication, the virus spreads via retrograde axoplasmic flow in unmyelinated motor or sensory nerves to the CNS. It then replicates in the brain before moving via the nerves into other tissues, including the salivary glands, from which it can be shed.

Clinical manifestations The infectivity of rabies virus varies with the site and mode of transmission. A bite on the face presents a 60% chance of disease; a bite on the hand or arm reduces the chance of disease to between 15% and 40%, and a bite on the leg presents only a 3% to 10% chance of disease. The risk of disease from a bite is almost 50 times greater than the risk from scratches by a rabid animal. Although a less common mode of transmission, the virus also can be inhaled, which accounts for rabies in laboratory workers exposed to viral aerosols and in a few explorers of bat-infested caves.

The incubation period of rabies ranges from 12 days to many years and probably averages 30 to 90 days or less. The clinical course is quite variable. The initial clinical presentation is a prodromal phase typically lasting a day or two. It is marked by pain and paresthesias in the area of the bite, GI and upper respiratory symptoms, irritability, apprehension, and a sense of impending death. Hydrophobia and aerophobia occur in some patients and, like the history of a bite, call attention to this disease. Thereafter, the patient enters an excitation state, marked by hyperventilation, hyperactivity, disorientation, and even seizures. During the next few days, the patient becomes lethargic and begins to show paralysis, particularly in areas innervated by the cranial nerves and in the somatic muscles, bladder, and bowels. Gradual involvement of cardiac muscles and paralysis of respiratory muscles lead to death. Rabies virus infection in humans is uniformly fatal once symptoms occur.

Diagnosis Rabies should be considered if classic signs of hydrophobia, aerophobia, and excited behavior are present or in any case of encephalitis or myelitis of unknown etiology, even in the absence of an exposure history. The CSF shows nonspecific elevation in levels of leukocytes and protein, as occurs in other viral encephalitides. Fluorescent antibody staining, virus isolation, and reverse transcriptase/PCR constitute the most accurate means of diagnosing rabies infection.[57] Circulating antibodies may be detected in unvaccinated persons as early as the sixth day of illness and usually appear within the first 2 weeks after infection. Rabies virus can be isolated from the second day through the second week of illness from a throat swab or saliva sample, as well as from tears, urine sediment, and CSF.

Treatment and prevention Three rabies vaccines are currently available in the United States: human diploid cell rabies vaccine (HDCV), rabies vaccine absorbed (RVA), and purified chick embryo cell vaccine (PCEC).[58] Each is licensed for preexposure or postexposure vaccination. Clinical trials with RVA and PCEC have demonstrated immunogenicity equivalent to that of HDCV.[58] Corticosteroids, other immunosuppressing agents and conditions, and antimalarials may interfere with the development of active immunity after vaccination. Preexposure prophylaxis is indicated for persons in high-risk groups such as certain laboratory workers, persons whose occupation puts them in frequent contact with animal species at risk for having rabies, and certain international travelers.[58]

Disease onset can be prevented by prompt postexposure prophylaxis. Postexposure prophylaxis begins with immediate and thorough washing of all bite wounds with soap and water and irrigation with a virucidal agent such as a povidone-iodine solution. Previously unimmunized persons should receive human rabies immunoglobulin and rabies vaccine. Guidelines for the indications and dosing schedules for postexposure prophylaxis are available from the CDC at www.cdc.gov/ncidod/dvrd/rabies/.[55] Local or state public health officials should be consulted if questions arise about the need for rabies prophylaxis.

At present, therapy for clinical rabies is supportive. In unvaccinated patients, the disease is invariably fatal. A patient with rabies or suspected rabies should be kept in isolation, and standard infection control precautions should be followed, although laboratory-confirmed person-to-person transmission has not been documented.

VIRUSES OF THE FAMILY PARAMYXOVIRIDAE

Nipah virus

Nipah virus (family Paramyxoviridae) is closely related to the Hendra virus, which has rarely caused fever, pneumonia, and encephalitis in persons exposed to ill horses in Australia. Nipah virus was newly discovered in Malaysia during an outbreak involving 265 persons, with 105 deaths in 1998 and 1999; in 2004, outbreaks involving 25 persons occurred in Bangladesh.[59] Most affected persons had contact with live pigs and were pig farmers. The reservoir for Nipah virus is believed to be fruit bats, and it is thought that humans are infected by contact with an infected bat or by contact with an intermediate animal host such as pigs.

Diagnosis The incubation period is unknown, but more than 90% of the hospitalized Malaysian patients had had contact with pigs in the previous 2 weeks. Among hospitalized patients, fever was almost invariably present, with headache, dizziness, and vomiting common presenting symptoms. Predominant neurologic symptoms were decreased level of consciousness, segmental myoclonus, meningismus, and seizures. Pneumonia was noted in 25% of the patients in Singapore, but it was not prominent in the Malaysian patients. The diagnosis can be confirmed serologically by viral culture or by detection of viral nucleic acid.

Treatment Treatment is supportive. An open-label trial suggested that ribavirin reduced mortality.[60] Prevention is avoidance of pig farms.

VIRUSES OF THE FAMILY REOVIRIDAE

Colorado Tick Fever

The Colorado tick fever virus (family Reoviridae, genus *Coltivirus*) is transmitted to humans in the western United States and Canada mainly by the wood tick, *Dermacentor andersoni*. Human incidence corresponds to the wood tick's geographic distribution in mountainous areas at elevations of 4,000 to 10,000 ft. Transmission occurs from March to September, but it peaks from April to June.

Diagnosis The mean incubation period for Colorado tick fever virus is 3 to 4 days. In 90% of cases, the patient reports a tick bite or tick exposure. Fever, chills, myalgias, and prostration are common presenting symptoms. A petechial or maculopapular rash occurs in 15% of patients.[61] Although acute symptoms last about a week, fever may recur several days later. Fatigue is often prolonged. Meningitis or encephalitis develops in 5% to 10% of children; fatal cases with hemorrhage and shock have been rarely reported. Leukopenia is very common. The virus infects marrow erythrocytic precursors, which accounts for the ability to recover the virus from peripheral blood up to 6 weeks after illness onset. Transmission via blood transfusion has been reported. The diagnosis can also be confirmed serologically by the CDC on request through state and local health departments.

Treatment Treatment is supportive. Prevention rests on avoiding tick bites in endemic areas.

Rash Arthralgia

VIRUSES OF THE FAMILY TOGAVIRIDAE

Several alphaviruses belonging to the family Togaviridae can cause a viral syndrome associated with rash and arthralgias or arthritis. Diagnosis requires serologic testing of paired acute-phase and convalescent-phase serum. Virus can also be isolated from or detected in acute-phase serum samples.

There are no vaccines available for general use against alphaviruses. Prevention depends on mosquito control and decreasing exposure to mosquitoes.

Barmah Forest Virus

Barmah Forest virus (family Togaviridae, genus *Alphavirus*) causes sporadic disease and epidemics in Australia.[42] Clinically, Barmah Forest virus causes a Ross River virus–like illness, but the rash tends to be more florid, and true arthritis is less common. Outbreaks have coincided with Ross River virus outbreaks, and Barmah Forest virus has been identified in the same mosquito species as Ross River virus.[42,62]

Chikungunya

Chikungunya (CHIK) virus (family Togaviridae, genus *Alphavirus*) is found in Africa and Asia and is transmitted by *Aedes* mosquitoes. After a 20-year hiatus, outbreaks of CHIK virus infection occurred in Indonesia from September 2001 to March 2003, suggesting a reemergence of the virus.[63] In urban settings, the virus is transmitted from human to human via *A. aegypti* mosquitoes. Explosive urban epidemics occur during the rainy season. The native name for the disease means "doubled up," because of the excruciating joint pains.

Diagnosis In patients with CHIK, after an incubation period of 2 to 4 days, there is a sudden onset of fever and crippling joint pains, accompanied by chills, flushed face, headache, myalgias, backache, and photophobia. Arthralgias are polyarticular, are migratory, and mostly involve the small joints. Papular or maculopapular skin rashes, typically on the trunk and limbs, usually occur during the second to fifth day of illness. The clinical picture resembles that of dengue fever, with which chikungunya is often confused.[64] Most infections are probably asymptomatic. Arthralgias may last several months. In Asia, but not Africa, mild hemorrhagic manifestations have been reported; CHIK virus is not a cause of severe hemorrhagic disease. The diagnosis can be confirmed serologically by the CDC on request through state and local health departments. Confirmation by culture or detection of viral nucleic acid is possible early in illness.

Treatment There is no specific treatment for CHIK. Anti-inflammatory drugs may relieve arthralgia. Chloroquine phosphate has been used for refractory arthralgias. Prevention depends on mosquito control and decreasing mosquito exposure.

Mayaro Virus

Mayaro (MAY) virus (family Togaviridae, genus *Alphavirus*) is closely related to CHIK virus and causes a similar illness. The virus has been isolated from mosquitoes (mostly *Haemagogus*) in various countries in the Caribbean and South America. Little is known about the natural history of the disease. MAY virus causes febrile illness with headache, backache, myalgias, epigastric pain, chills, nausea, photophobia, arthralgias, and maculopapular rash. Polyarthritis occurs and may persist for several weeks. Arthralgia may recur, as indicated in a report of a patient with a previous case of MAY virus infection.[65] Cases of MAY disease have been imported into the United States.[66] MAY virus infection should be considered in the differential diagnosis in patients with a recent travel history to South America. Diagnosis can be confirmed serologically by the CDC on request through state and local health departments.

O'nyong-nyong

O'nyong-nyong (ONN) virus (family Togaviridae, genus *Alphavirus*) was first isolated during an epidemic in Uganda in 1959 and spread to an estimated two million people in neighboring countries by 1962. Another ONN epidemic began in south-central Uganda in 1996.[67] ONN virus is transmitted to humans by *Anopheles* and other mosquitoes. Clinically, ONN fever is similar to CHIK, although fever is less pronounced and lymphadenopathy is more common in ONN.

Ross River Fever

Ross River virus (family Togaviridae, genus *Alphavirus*) has caused so-called epidemic polyarthritis in Australia, southwestern Pacific Islands, and Fiji.[42] Several species of *Aedes* and *Culex* mosquitoes are important vectors.[42] The natural maintenance cycle of Ross River virus is not fully known. Although the magnitude, regularity, seasonality, and locality of outbreaks are wide ranging, rainfall seems to be the single most important risk factor.[68] Humans have significant viremia, and the virus may follow a human-mosquito-human transmission cycle.[69]

After an incubation period of 2 to 21 days, the illness begins suddenly with myalgia and marked arthralgias in the small joints of the hands and feet. True arthritis occurs in over 40% of patients. A maculopapular rash occurs in 50% of patients within 2 days after onset. Myalgia, headache, anorexia, nausea, and tenosynovitis are common, but temperature is only slightly elevated. Arthralgia frequently persists for several weeks or longer—sometimes for longer than a year.

VIRUSES OF THE FAMILY FLAVIVIRIDAE

Dengue

In addition to their role in hemorrhagic fever (see above), dengue viruses are a common cause of rash and arthralgia.[70]

OTHER VIRUSES CAUSING RASH AND ARTHRALGIA

There are a number of other viral zoonoses that cause similar nonspecific febrile illness in humans, but they occur infrequently or are rare. These include Igbo-Ora (family Togaviridae, genus *Alphavirus*) in Africa; Sindbis and Sindbis-like viruses (family Togaviridae, genus *Alphavirus*) in Africa, Asia, Australia, and Europe; Group C arboviruses (family Bunyaviridae, genus *Bunyavirus*) in South America; Oropouche (family Bunyaviridae, genus *Bunyavirus*) in South and Central America; Sandfly fever (family Bunyaviridae, genus *Phlebovirus*) in the Mediterranean, Middle East, West Asia, and South America; Zika virus (family Flaviviridae, genus *Flavivirus*) in Africa and Asia; and vesicular stomatitis virus (family Rhabdoviridae, genus *Vesiculovirus*) in the Americas.

Miscellaneous Viral Zoonoses

MONKEY B VIRUS DISEASE

Monkey B virus (*Herpesvirus simiae* or B virus) causes persistent latent infections in at least 70% of captive adult macaques. Monkey B virus disease in humans usually results from macaque bites or scratches; most documented infections have occurred in laboratory personnel working with apparently healthy rhesus, cynomolgus, or African green monkeys or their tissues, including kidney cell cultures.[71] Human infection from a mucocutaneous exposure to the eye and one human-to-human transmission have been reported. In humans, monkey B virus causes acute ascending myelitis and fulminant meningoencephalitis, which leads to death within days.

Infection can be diagnosed in humans by demonstrating a rise in antibody titer and by isolating the virus from the CNS. Cardiopulmonary support is the most important aspect of management. Human infection carries a high mortality; however, patients treated early with intravenous acyclovir or ganciclovir have survived. Significant neurologic residua are common in survivors. Monkey handlers should wear protective clothing and a face mask.[72] Bites, scratches, or mucosal surfaces exposed to macaque biologic materials should be cleansed thoroughly.[69] Postexposure management should include referral to a medical consultant knowledgeable about monkey B virus.

RUMINANT AND PRIMATE POXVIRUS DISEASES

Several ruminant and primate poxviruses rarely cause human illness. Cowpox (vaccinia virus) is an orthopoxvirus related to variola. In humans, it produces vesicular lesions on the hands. Generalized infections are rare. Monkeypox, the only other orthopoxvirus of significance to humans, is enzootic in monkeys and squirrels in western and central Africa; infection in humans is sporadic and produces a vesicular rash similar to variola. Secondary infections occur. The case-fatality rate is 1.5%.[73] Tanapox virus is a yatapoxvirus that causes vesicular lesions in monkeys along the Tana River in Kenya and Zaire. It produces a monkeypoxlike illness in humans. The parapoxviruses produce disease in humans through direct contact with infected animals. These include bovine papular stomatitis virus, milkers' nodule (pseudocowpox) in cattle, and orf in sheep and goats. Infection results in vesicles that progress to pustules and scabbing at the site of contact with the original infected species or contaminated objects.

NEWCASTLE DISEASE VIRUS INFECTION

Newcastle disease is an often fatal systemic infection of poultry that is caused by a paramyxovirus. The virus is occasionally transmitted to humans from infected birds or in the laboratory, presumably by direct inoculation. In humans, the illness appears as acute, sometimes hemorrhagic conjunctivitis without corneal involvement. It can be accompanied by lymphangitis, headache, malaise, and chills but is usually self-limited. Patients recover within 2 weeks.

VESICULAR STOMATITIS VIRUS INFECTION

Vesicular stomatitis virus is a rhabdovirus whose structure resembles that of rabies virus. The agent is responsible for oral ulcers in cattle. Occasionally, laboratory workers become infected and experience fever, vesicular enanthemas, headache, and myalgias.

FOOT-AND-MOUTH DISEASE

Foot-and-mouth disease is a highly infectious viral infection of cloven-hoofed animals. The causative agent of the disease is aphthovirus, a member of the family Picornaviridae that is indistinguishable morphologically from rhinovirus. Persons contacting infected animals occasionally have fever, vesicular lesions on the hands, and an increase in neutralizing and complement-fixing titers. The infection is mild and transient, but relapses occur.

Lyle R. Petersen, M.D., M.P.H., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Duane J. Gubler, Sc.D., has received grants for educational activities from Novartis Pharmaceutical Corp., Sanofi Pasteur SA, and Hawaii Biotech, Inc.

References

1. Gubler DJ: Dengue and dengue hemorrhagic fever. Clin Microbiol Rev 11:480, 1998

2. Pancharoen C, Thisyakorn U: Neurological manifestations in dengue patients. Southeast Asian J Trop Med Public Health 32:341, 2001

3. Sumarmo, Wulu H, Jahja E, et al: Clinical observations on virologically confirmed fatal dengue infections in Jakarta, Indonesia. Bull World Health Organ 61:693, 1983

4. Rigau-Perez JG, Vorndam AV, Clark GG: The dengue and dengue-hemorrhagic fever epidemic in Puerto Rico, 1994–1995. Am J Trop Med Hyg 64:67, 2001

5. Wichmann O, Hongsiriwon S, Bowonwatanuwong C, et al: Risk factors and clinical features associated with severe dengue infection in adults and children during the 2001 epidemic in Chonburi, Thailand. Trop Med Int Health 9:1022, 2004

6. Kurane I, Takasaki T: Dengue fever and dengue hemorrhagic fever: challenges of controlling an enemy still at large. Rev Med Virol 11:301, 2001

7. Rico-Hesse R: Microevolution and virulence of dengue viruses. Adv Virus Res 59:315, 2003

8. Anantapreecha S, Chanama S, A-nuegoonpipat A, et al: Serological and virological features of dengue fever and dengue haemorrhagic fever in Thailand from 1999 to 2002. Epidemiol Infect 133:503, 2005

9. Krishnamurti C, Kalayanarooj S, Cutting MA, et al: Mechanisms of hemorrhage in dengue without circulatory collapse. Am J Trop Med Hyg 65:840, 2001

10. Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control, 2nd ed. WHO Technical Report Series, World Health Organization, Geneva, Switzerland, 1997 (accessed 12/05)
http://www.who.int/csr/resources/publications/dengue/Denguepublication/en

11. Chang GJ, Kuno G, Purdy DE, et al: Recent advancement in flavivirus vaccine development. Expert Rev Vaccines 3:199, 2004

12. Blaney JE Jr, Matro JM, Murphy BR, et al: Recombinant, live-attenuated tetravalent dengue virus vaccine formulations induce a balanced, broad, and protective neutralizing antibody response against each of the four serotypes in rhesus monkeys. J Virol 79:5516, 2005

13. Monath TP: Yellow fever. The Arboviruses: Epidemiology and Ecology, Vol. V. CRC Press, Boca Raton, Florida, 1988

14. Lacy MD, Smego RA: Viral hemorrhagic fevers. Adv Pediatr Infect Dis 12:21, 1996

15. Nasidi A, Monath TP, DeCock K, et al: Urban yellow fever epidemic in western Nigeria, 1987. Trans R Soc Trop Med Hyg 83:401, 1989

16. Centers for Disease Control and Prevention. Travelers Health. Atlanta: US Department of Health and Human Services, Public Health Service, 2005
http://www.cdc.gov/travel/vaccinat.htm

17. Cetron MS, Marfin AA, Julian KG, et al: Yellow fever vaccine. Recommendations of the Advisory Committee for Immunization Practices (ACIP), 2002. MMWR Recomm Rep 51:1, 2002

18. Martin M, Tsai TF, Cropp B, et al: Fever and multisystem organ failure associated with 17D-204 yellow fever vaccination: a report of four cases. Lancet 358:98, 2001

19. Khromava Ay, Eidex RB, Weld LH, et al: Yellow fever vaccine: an updated assessment of advanced age as a risk factor for serious adverse events. Vaccine 9:3256, 2005

20. Whitehouse CA: Crimean-Congo hemorrhagic fever. Antiviral Res 64:145, 2004

21. Mardani M, Jahromi MK, Naieni KH, et al: The efficacy of oral ribavirin in the treatment of Crimean-Congo hemorrhagic fever in Iran. Clin Infect Dis 36:1613, 2003

22. McCaughey C, Hart CA: Hantaviruses. J Med Microbiol 49:587, 2000

23. Zeier M, Handermann M, Bahr U, et al: New ecological aspects of hantavirus infection: a change of a paradigm and a challenge of prevention: a review. Virus Genes 30:157, 2005

24. Wells RM, Estani SS, Yadon ZE, et al: An unusual hantavirus outbreak in southern Argentina: person-to-person transmission? Emerg Infect Dis 3:171, 1997

25. Huggins JW, Hsiang CM, Cosgriff TM, et al: Prospective, double-blind, concurrent, placebo-controlled clinical trial of intravenous ribavirin therapy of hemorrhagic fever with renal syndrome. J Infect Dis 164:1119, 1991

26. Hjelle B: Vaccines against hantaviruses. Expert Rev Vaccines 1:373, 2002

27. Padula PJ, Colavecchia SB, Martinez VP, et al: Genetic diversity, distribution, and serological features of hantavirus infections in five countries in South America. J Clin Microbiol 38:3029, 2000

28. Enria DA, Briggiler, Pini N, et al: Clinical manifestations of New World hantaviruses. Curr Top Microbiol Immunol 256:117, 2001

29. Mertz GJ, Miedzinski L, Goade D, et al: Placebo-controlled, double-blind trial of intravenous ribavirin for the treatment of hantavirus cardiopulmonary syndrome in North America. Clin Infect Dis 39:1307, 2004

30. Bossi P, Tegnell A, Baka A, et al: Bichat guidelines for the clinical management of haemorrhagic fever viruses and bioterrorism-related fever viruses. Euro Surveill 15:E11, 2004

31. Maiztegui JI, McKee KT Jr, Barrera Oro JG, et al: Protective efficacy of a live attenuated vaccine against Argentine hemorrhagic fever. J Infect Dis 177:277, 1998

32. Peters CJ: Marburg and Ebola: arming ourselves against the deadly filoviruses. N Engl J Med 352:2571, 2005

33. Outbreak of Marburg virus hemorrhagic fever: Angola, October 1, 2004-March 29, 2005. MMWR Morb Mortal Wkly Rep 54:308, 2005

34. Jahrling PB, Geisbert TW, Dalgard DW, et al: Preliminary report: isolation of Ebola virus from monkeys imported to USA. Lancet 335:502, 1990

35. Miranda ME, Ksiazek TG, Retuya TJ, et al: Epidemiology of Ebola (subtype Reston) virus in the Philippines, 1996. J Infect Dis 179(suppl 1):S115, 1999

36. Outbreak of Ebola hemorrhagic fever—Uganda, August 2000–January 2001. MMWR Morb Mortal Wkly Rep 50:73, 2001

37. Bwaka MA, Bonnet MJ, Calain P, et al: Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: clinical observations in 103 patients. J Infect Dis 179(suppl 1):vii, 1999

38. Towner JS, Rollin PE, Bausch DG, et al: Rapid diagnosis of Ebola hemorrhagic fever by reverse transcription-PCR in an outbreak setting and assessment of patient viral load as predictor of outcome. J Virol 78:4330, 2004

39. McJunkin JE, de los Reyes EC, Irazuzta JE, et al: La Crosse encephalitis in children. N Engl J Med 344:801, 2001

40. Inactivated Japanese encephalitis virus vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 42(RR-1):1, 1993
http://www.phppo.cdc.gov/cdcRecommends/

41. Solomon T, Dung NM, Wills B, et al: Interferon alfa-2a in Japanese encephalitis: a

randomized double-blind, placebo controlled trial. Lancet 361:821, 2001

42. Brown A, Bolisetty S, Whelan P, et al: Reappearance of human cases due to Murray Valley encephalitis virus and Kunjin virus in central Australia after an absence of 26 years. Commun Dis Intell 26:39, 2002

43. Rahal JJ, Anderson J, Rosenberg C, et al: Effect of interferon-alpha2b therapy on St. Louis viral meningoencephalitis: clinical and laboratory results of a pilot study. J Infect Dis 190:1084, 2004

44. Gritsun TS, Lashkevich VA, Gould EA: Tick-borne encephalitis. Antiviral Res 57:129, 2003

45. Rumyantsey AA, Chancok RM, Murphy BR, et al: Comparison of live and inactivated tick-borne encephalitis virus vaccines for safety, immunogenicity, and efficacy in rhesus monkeys. Vaccine 24:133, 2006

46. Petersen LR, Roehrig JT: West Nile virus: a reemerging global pathogen. Emerg Infect Dis 7:611, 2001

47. Lanciotti RS, Roehrig JT, Deubel V, et al: Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286:2333, 1999

48. Petersen LR, Marfin AA, Gubler DJ: West Nile virus. JAMA 290:524, 2003

49. Hayes EB, Gubler DJ: West Nile virus: epidemiology and clinical features of an emerging epidemic in the United States. Annu Rev Med, Sept 1, 2005 (epub ahead of print)

50. Mostashari F, Bunning ML, Kitsutani PT, et al: Epidemic West Nile encephalitis, New York, 1999: results of household-based seroepidemiologic survey. Lancet 358:261, 2001

51. Outbreak of Powassan encephalitis–Maine and Vermont, 1999–2000. MMWR Morb Mortal Wkly Rep 50:761, 2001

52. Cupp EW, Klingler K, Hassan HK, et al: Transmission of eastern equine encephalomyelitis virus in central Alabama. Am J Trop Med Hyg 68:495, 2003

53. Watts DM, Lavera V, Callahan J, et al: Venezuelan equine encephalitis and Oropouche virus infections among Peruvian army troops in the Amazon region of Peru. Am J Trop Med Hyg 56:661, 1997

54. Krebs JW, Mandel EJ, Swerdlow DL, et al: Rabies surveillance in the United States during 2003. J Am Vet Med Assoc 225:1837, 2004

55. Human rabies prevention—United States, 1999. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 48 (RR-1):1, 1999

56. Human death associated with bat rabies—California, 2003. MMWR Morb Mortal Wkly Rep 53:33, 2004

57. Noah DL, Drenzek CL, Smith JS, et al: Epidemiology of human rabies in the United States, 1980 to 1996. Ann Intern Med 128:922, 1998

58. Rupprecht CE, Gibbons RV: Prophylaxis against rabies. N Engl J Med 351:25, 2004

59. Bellini WJ, Harcourt BH, Bowden N, et al: Nipah virus: an emergent paramyxovirus causing severe encephalitis in humans. J Neurovirol 11:481, 2005

60. Chong HT, Kamarulzaman A, Tan CT, et al: Treatment of acute Nipah encephalitis with ribavirin. Ann Neurol 49:810, 2001

61. Klasco R: Colorado tick fever. Med Clin North Am 86:435, 2002

62. Passmore J, O'Grady KA, Moran R, et al: An outbreak of Barmah Forest virus disease in Victoria. Commun Dis Intell 26:600, 2002

63. Laras K, Sukri NC, Larasati RP, et al: Tracking the re-emergence of epidemic chikungunya virus in Indonesia. Trans R Soc Trop Med Hyg 99:128, 2005

64. Carey DE: Chikungunya and dengue: a case of mistaken identity? J Hist Med Allied Sci 26:243, 1971

65. Taylor SF, Patel PR, Herold TJ: Recurrent arthralgias in a patient with previous Mayaro fever infection. South Med J 98:484, 2005

66. Tesh RB, Watts, DM, Russell KL, et al: Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America. Clin Infect Dis 28:67, 1998

67. Kiwanuka N, Sanders EJ, Rwaguma EB, et al: O'nyong-nyong fever in South-Central Uganda, 1996–1997: clinical features and validation of a clinical case definition for surveillance purposes. Clin Infect Dis 29:1243, 1999

68. Kelly-Hope LA, Purdie DM, Kay BH: Ross River virus disease in Australia, 1886–1998, with analysis of risk factors associated with outbreaks. J Med Entomol 41:133, 2004

69. Gubler DJ: Transmission of Ross River virus by *Aedes polynesiensis* and *Aedes aegypti*. Am J Trop Med Hyg 30:1303, 1981

70. Chadwick D, Arch B, Wilder-Smith A, et al: Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. J Clin Infect July 28, 2005 [epub ahead of print]

71. Huff JL, Barry PA: B-virus (Cercopithecine herpesvirus 1) infection in humans and macaques: potential for zoonotic disease. Emerg Infect Dis 9:246, 2003

72. Cohen JI, Davenport DS, Stewart JA, et al: Recommendations for prevention of and therapy for exposure to B virus (cercopithecine herpesvirus 1). Clin Infect Dis 35:1191, 2002

73. Human monkeypox–Kasai Oriental, Democratic Republic of Congo, February 1996–October 1997. MMWR Morb Mortal Wkly Rep 46:1168, 1997

XXXII HUMAN RETROVIRAL INFECTIONS

Robert W. Coombs, m.d., ph.d.

Retroviruses have been recognized as important animal pathogens for more than 70 years. However, retroviruses were identified as etiologic agents in human disease only within the past 2 decades.[1,2] Three retroviruses have been recognized to cause infections in humans and to produce well-defined clinical disease: human T cell lymphotropic virus type I (HTLV-I), HIV type 1 (HIV-1), and HIV type 2 (HIV-2). One retrovirus (human T cell lymphotropic virus type II [HTLV-II]) has been associated with certain rare hematologic malignancies, but evidence confirming an etiologic role of HTLV-II in these disorders is lacking. In addition, human foamy virus (HFV), a member of the genus *Spumavirus*, is likely acquired as a zoonotic infection after severe primate bites, but it has not been associated with human disease.[3]

Definition and Classification

Retroviruses are named for the action of a unique viral-encoded enzyme called reverse transcriptase. The normal flow of genetic information is from DNA to RNA to protein. In a reversal of that process, reverse transcriptase makes a double-stranded DNA copy (complementary DNA, or cDNA) of the single-stranded viral RNA genome—hence, the prefix retro. Retroviruses are characterized morphologically by their appearance on electron microscopy and by their genomic, antigenic, and pathophysiologic characteristics.[4] The family Retroviridae comprises a large group of ubiquitous viruses that infect all classes of vertebrates. On the basis of similarities in amino acid sequences in the reverse transcriptase proteins of retroviruses, seven retrovirus genera have been classified. Humans can be or have been infected with several retroviruses from five of the seven retrovirus genera [*see Table 1*]. The alpharetroviruses, betaretroviruses, and gammaretroviruses are considered simple retroviruses; the deltaretroviruses, epsilonretroviruses, lentiviruses, and spumaviruses are considered complex. The simple retroviruses encode only structural and envelope proteins, whereas the complex retroviruses encode these proteins and also small regulatory proteins with a range of functions.[4] With regard to clinical infections in humans, the most important retroviruses (HTLV and HIV) and the least important retroviruses (human retrovirus type 5 [HRV-5][5] and HFV) are acquired by exogenous infection (i.e., by sexual transmission or by transmission from mother to infant through breast-feeding or by parenteral exposure).

A defining feature of retroviruses is the integration of the reverse-transcribed viral cDNA into the genome of the host cell. The integrated viral cDNA is termed the provirus; it serves as the template for viral replication. A persistent infection may result from this process of integration, or the infected cell may be transformed and a malignancy induced. Although many retroviruses can cause cytopathology in the host cell, most replicate without killing the infected cell. For the endogenous human retroviruses (human endogenous retrovirus type K [HERV-K], HERV-W, and several others) [*see Table 1*], multiple copies of endogenous proviral DNA sequences are integrated into chromosomal DNA and are transmitted in the germline. Interestingly, endogenous proviral sequences represent 0.1% or more of human DNA sequences and were acquired sometime in our evolutionary past.[4]

Retroviral Biology

structure and replication

All human retroviruses are similar in structure, genomic organization, and mode of replication.[4] The virus particles are approximately 100 nm in diameter and have a lipid envelope composed of components from the host cell–derived plasma membrane.[6] The envelope renders the virus particles very susceptible to inactivation by environmental drying and by the action of detergents and various chemical disinfectants. The lipid envelope surrounds a dense nucleocapsid core that contains two copies of the unique single-stranded RNA genome. The RNA genomes range from 8 kb in length for HTLV to approximately 10 kb in length for HIV. In the virion, the RNA genomes are associated with the virus-encoded reverse transcriptase, integrase, vpr protein in HIV-1, vpx protein in HIV-2, and a primer transfer RNA (tRNA) that is derived from the host cell.[7] The HIV virion must also incorporate a cellular protein, cyclophilin A, which binds to viral capsid matrix protein p17; this protein is necessary for successful viral disassembly.[8] The tightly packaged HIV nucleocapsid protein protects the enclosed RNA genome from degradation by ribonuclease (RNase) in plasma and other body fluids. As a durable marker of HIV-1 replication, the measurement of viral particle–associated RNA plays a central role in monitoring HIV disease progression and response to therapy.[9,10]

The retroviral RNA genome and the cDNA provirus contain both noncoding and coding sequences **[***see Figure 1***]**. The noncoding sequences, which are important recognition sites for DNA or RNA synthesis, integration, and polyadenylation, are located at the 5′ and 3′ terminal ends of the genome. All retroviruses are terminally redundant and contain identical se-

Table 1 Genera of the Family Retroviridae*

Genus	*Examples*
Alpharetrovirus	Rous sarcoma virus
Betaretrovirus	Human retrovirus type 5
Gammaretrovirus	Human endogenous retrovirus type W* Murine leukemia virus Feline leukemia virus
Deltaretrovirus	Human T cell lymphotropic virus types I and II Bovine leukemia virus
Epsilonretrovirus	Snake retrovirus Walleye dermal sarcoma virus
Lentivirus	Human immunodeficiency virus types 1 and 2 Simian immunodeficiency virus Equine infectious anemia virus Feline immunodeficiency virus Caprine arthritis encephalitis virus Maedi/visna virus
Spumavirus	Human foamy virus Feline foamy virus Bovine foamy virus Simian foamy virus

*Several other human endogenous retroviruses have been identified but are not listed.

quences, called long terminal repeats (LTRs). The coding sequences include the *gag* gene, which encodes group-specific structural antigens; the *pol* gene, which encodes RNA-dependent DNA polymerase or reverse transcriptase, integrase, and protease; and the *env* gene, which encodes envelope structural proteins. The *gag* gene encodes a precursor polypeptide that is cleaved by viral-encoded protease to form several internal structural proteins—namely, matrix protein (MA), capsid protein (CA), and nucleic acid binding protein (nucleocapsid, or NC).[4]

In addition, the 3′ end of the *gag* gene reading frame overlaps with the *pol* gene reading frame to encode for the virus-specific protease. The *pol* gene encodes for a precursor polypeptide that is cleaved by the virus-specific protease enzyme to form three enzymes: protease, reverse transcriptase, and integrase. The *env* gene encodes for a 160 kd precursor protein, which is cleaved posttranscriptionally into two noncovalently associated envelope glycoproteins during transport through the endoplasmic reticulum and the Golgi complex by host proteases, termed furins. The first env protein is gp120, which is a highly charged glycoprotein that is external to the viral envelope; gp120 binds to cell-specific viral receptors (e.g., $CD4^+$ T cell receptor in the case of HIV) and ancillary coreceptors (e.g., chemokine receptors for HIV-1). There are five domains of the *env* gene—V1 through V5. Some of these are highly variable (of these domains, the V3 loop is the most highly variable). These variable domains are responsible for defining multiple variants or quasispecies that contribute to viral evasion from neutralizing antibodies produced by the host in response to viral infection. The second env protein, gp41, is a hydrophobic transmembrane glycoprotein that anchors the oligomeric surface subunit glycoprotein to the viral envelope membrane. The virion surface is studded with approximately 72 knobs; each is composed of three heterodimers of the gp120env / gp41env complex.[7,11] Fusion of the retroviral envelope with the target cell plasma membrane is facilitated by the transmembrane glycoprotein through conformational change mediated by a helix coiled-coil mechanism. This mechanism is present in enveloped viruses other than retroviruses (e.g., influenzavirus)[12] [*see Figure 2*].

The complexity of the human retroviruses is best exemplified by the array of viral proteins that are responsible for regulating viral replication and the host cell response to infection.[4,7] HTLV-I has a region between the *env* gene and the 3′ LTR that encodes for two regulatory proteins, tax and rex, which are produced from messages that are spliced differently from distinct overlapping reading frames. The tax protein induces the expression of cell transcription factors that alter host cell gene expression, and the rex protein regulates the expression of viral messenger RNA (mRNA). The HIV viruses have a larger genome than HTLV. In addition, HIV has a larger translated region (regulatory genes) between the *pol* and *env* genes, which encode portions of several regulatory proteins that depend on the overlapping open reading frame (ORF) into which the mRNA is spliced.

These retroviral regulatory proteins express the following functions[13,14] [*see Table 2*]. The tat protein augments the expression of virus from the LTR region. The rev protein regulates RNA splicing and RNA transport, or both, in HIV and may function in a manner similar to that of the rex protein in HTLV. The nef protein downregulates CD4 protein, which is the cellular receptor for HIV; it alters host T cell activation pathways by decreasing major histocompatibility complex (MHC) class I antigen expression; and it enhances viral infectivity. The vif protein is necessary for the proper assembly of the HIV nucleoprotein core; without the vif protein, viral cDNA is not efficiently produced. The vpr protein (in HIV-1) and the vpx protein (in HIV-2) facilitate transport of the viral cDNA into the nucleus

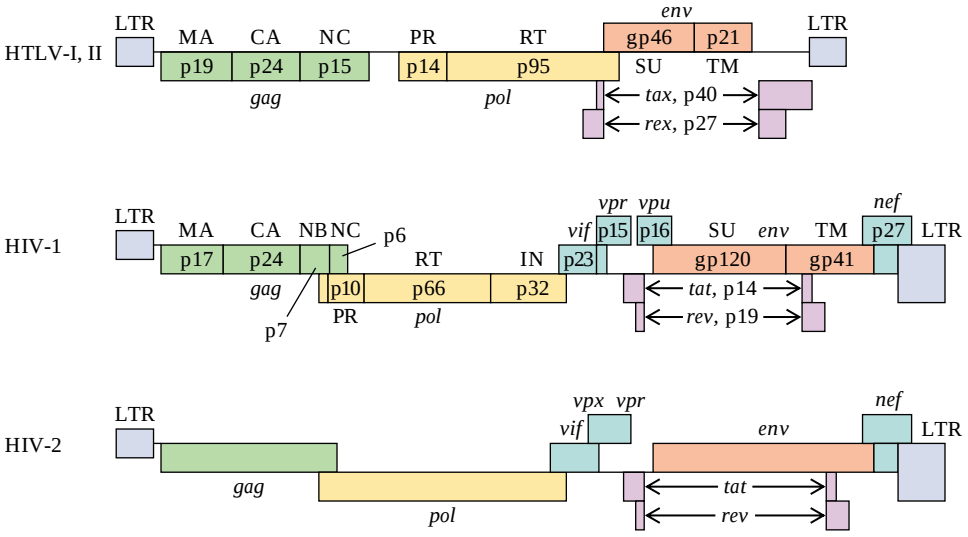


Figure 1 **Depicted is the genomic organization of the three most important human retroviruses. The linear double-stranded proviral DNA forms of HTLV and HIV-1 show similar patterns of genomic organization. The structural genes *gag, pol,* and *env* give rise to several proteins: matrix (MA), capsid (CA), nucleic–acid binding (NB), nucleocapsid core proteins (NC), protease (PR), reverse transcriptase (RT), surface subunit glycoprotein (SU), and a smaller transmembrane protein (TM). In addition, HIV *pol* encodes an integrase (IN). There are additional regulatory gene products translated. HTLV-I and HTLV-II have *tax* and *rex* genes with exons on either side of the *env* gene. HIV-1 and HIV-2 have six accessory gene products: tat, rev, vif, nef, vpr, and either vpu (in HIV-1) or vpx (in HIV-2). Further information about these gene products is provided[131] [*see Table 2*].**

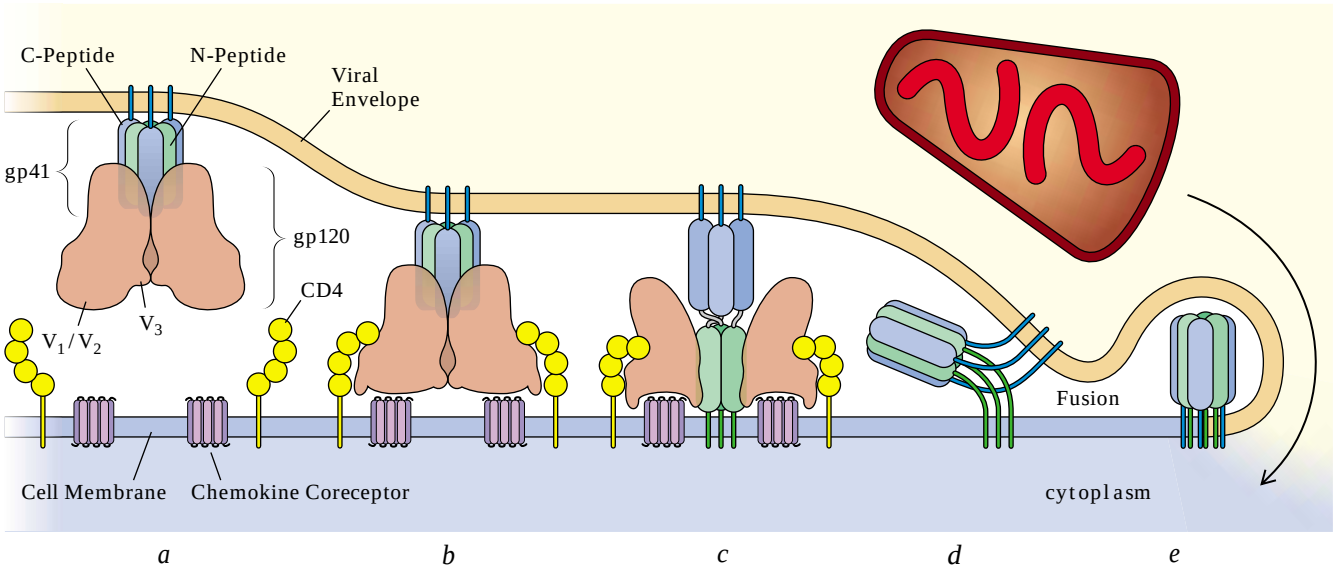


Figure 2 **Illustrated is the binding of HIV-1 with a CD4+ T cell. In unbound virions, gp41 exists in a stable, nonfusogenic conformation in which the fusion peptides are buried within the envelope trimer complex (*a*). When gp120 binds to the CD4 receptor, a conformational change exposes the chemokine receptor attachment site on gp120 (either CXC chemokine receptor–4 [CXCR4] or CC chemokine receptor–5 [CCR5]) (*b*). This in turn triggers a transition of gp41 to the prehairpin intermediate (*c*), with exposure of the fusion peptide attached to the trimeric coiled-coil N-peptide region. The fusin peptide inserts into the target membrane (*c*). In this form, the C-peptide has not associated with the N-peptide because of continued association with gp120; at this stage, the intermediate gp41 polypeptide is vulnerable to C-peptide inhibition (e.g., T-20). When the C-peptide region binds to the N-peptide region coiled-coil, the complex adopts a helical conformation of the fusion-active hairpin, which brings the two membranes into apposition (*d*). The precise mechanism of membrane fusion is not clear, but after fusion is complete, the fusion peptide and transmembrane segment of gp41 lie in the same membrane (*e*). A similar mechanism presumably applies to the fusion of a cell infected with HIV-1 that expresses viral envelope on the plasma membrane surface with an uninfected $CD4^+$ T cell, which leads to syncytium formation among infected and uninfected cells in vitro.[12,132]**

and cause an arrest in the G_2 stage of the cell cycle, which is necessary for optimal viral infection. As such, the vpr protein appears to facilitate the infection of nondividing cells by HIV, a property not shared by HTLV. The vpu protein promotes the degradation of CD4 protein in the endoplasmic reticulum and stimulates the release of virions from infected cells.

pathogenetic tryptic

The biologic properties of the human retroviruses, summarized above, orchestrate a viral pathogenesis tryptic defined by latency, transformation, and cytopathicity; this process occurs through two replication phases [*see Figure 3*]. In the first replication phase, the viral-encoded proteins that are packed within the virion nucleocapsid enter the cell; this eventually results in the formation of the integrated cDNA provirus. In the second replication phase, the enzymatic machinery of the host cell is utilized to replicate the viral RNA genome and to transcribe and translate viral proteins from the provirus. With the exception of retroviral latency, the pathogenesis tryptic leads to cell dysfunction (and sometimes cell death), with eventual immunosuppression and clinical disease. Both HTLV-I and HTLV-II immortalize primary human peripheral blood T cells in vitro. Immortalization is defined as interleuken-2 (IL-2)–dependent, long-term growth in culture. Subsequently, cells become fully transformed—that is, they are subject to continuous cell proliferation in the absence of IL-2. Cellular transformation is an early step in the pathogenic process of HTLV and is distinct from oncogenesis or malignancy.[4] As such, HTLV transformation of T cells results in a pool of proliferating cells that are not oncogenic themselves but that provide a population from which a malignant clone may subsequently arise.[4] This accounts for the long latent period between infection and tumor development in HTLV disease. In contrast, HIV does not transform the infected cell but rather produces a cytopathic effect that results in cell dysfunction and, eventually, cell death.

The first phase of viral replication starts when the retrovirus binds to a specific cell receptor, which for HIV is the CD4 molecule on the cell surface and an associated chemokine coreceptor.[15] The HTLV receptors have not been characterized; however, cellular targets for HTLV-I include several types of T cells, the primary one being the $CD4^+$ T cell. HTLV-II preferentially infects $CD8^+$ T cells.[16] The retroviral nucleocapsid complex enters the cell cytoplasm, and the viral genomic RNA is reverse-transcribed to a slightly longer, double-stranded cDNA molecule that actively enters the nucleus as a nucleoprotein-preintegration complex and integrates permanently at a single random site in the cell chromosome. The integration of linear viral cDNA occurs during division of the cell; as such, nondividing cells are generally resistant to retroviral infection. However, the one exception to this is HIV, which is able to infect resting cells (the G_1 phase of the cell cycle), albeit less efficiently than activated cells in the S phase of the cell cycle. The inhibition of reverse transcription by combinations of nucleotide, nucleoside, or non-nucleoside reverse transcriptase inhibitors (NNRTIs) is a cornerstone of current HIV-1 therapeutics.[10]

The second replication phase begins with the production of viral genomic RNA and mRNA and protein synthesis; the production of mRNA and protein synthesis occur almost exclusive-

Table 2 Genes and Proteins of HIV[13]

Gene	*Gene Product*	*Size (kd)*	*Function*
gag	Capsid matrix protein	p17	Interacts with gp41; binds cyclophilin A in the virion
	Nucleocapsid core protein	p24	Forms nucleocapsid along with p6
	Nucleocapsid core protein	p6	Forms nucleocapsid along with p24; binds to vpr
	Nucleic acid binding protein	p7	Binds to viral RNA
		p1, p2	Cleavage products of gag precursor protein
pol	Protease	p10	Proteolytic cleavage of gag, pol, and env precursor polypeptides
	Reverse transcriptase	p66/51	Polymerase and ribonuclease H activity (p66 only)
	Integrase	p32	Integration of viral DNA (vDNA) into host cell chromosome
env	Envelope protein	gp120	Trimeric external envelope protein; glycosylated protein with five variable domains (V1 through V5) and four invariant regions (C1 through C4); binds to CD4 receptor and chemokine coreceptor
	Transmembrane protein	gp41	Trimeric protein that attaches gp120 to virion envelope; facilitates fusion of the viral envelope with the cell plasma membrane; each molecule is composed of two α-helical regions that form a six-helix bundle; coiled-coil state is basis for "spring-loaded" interaction that leads to virus cell fusion
vif	Virion infectivity protein	p23	Efficient cell free transmission; required for proper assembly of nucleoprotein core
vpr	Viral protein R	p15	Enhances viral replication in primary cells; G2/M phase arrest; nuclear localization in cell; transports vDNA in preintegration complex to the nucleus; virion-associated protein
tat	*Trans*-activator of transcription	p14	Major viral *trans*-activator
rev	Regulator of expression of virion protein	p19	Enhances expression of unspliced and singly spliced RNA molecules; regulates transport of messenger RNA; shuttles back and forth between nucleus and cytoplasm
*vpu**	Viral protein U	p15-16	Enhances release of virion from cells; downregulates CD4 on cell
nef	Negative regulatory factor	p27	Inhibits or enhances viral replication, depending on strain and cell type; down-regulates CD4 and MHC class I receptor
vpx†	Virion protein X	p25	Packaged into the virion

*HIV-1 only.
†HIV-2 only.

ly through the enzymatic machinery of the host cell under the influence of viral gene regulatory products. Processing of virion proteins begins in the endoplasmic reticulum and Golgi complex. Virion assembly begins at the plasma membrane, and the nascent virions are released from the cell surface through budding. The budding viral envelope, which has a phospholipid composition different from that of the plasma membrane of the cell, may incorporate some cell membrane surface proteins (e.g., β_2-microglobulin and MHC class I and II proteins), along with virus-specific glycoproteins.[6,7] Extracellular HIV undergoes further maturation by continued proteolytic cleavage of the nucleocapsid polypeptide, which results in the mature infectious virion with a distinctive cone-shaped nucleocapsid core. The inhibition of viral protease through virus-specific protease inhibitor therapy, in combination with reverse transcriptase inhibitor therapy, is an important aim of current antiretroviral therapeutics. However, inhibition of reverse transcriptase with combinations of nucleoside reverse transcriptase inhibitors (NRTIs) and NNRTIs is equally efficacious in many cases.[10]

Pathogenesis of Retroviral Infection

The clinical consequences of retrovirus infection reflect a precarious balance between the type of invading retrovirus, its tropism for specific cells and tissues, and the patient's natural and acquired resistance. The pathophysiologic model of retrovirus infection can be approached by understanding viral mechanisms of infection, latency and persistence, cell injury, and immune system evasion and modulation. Some aspects of this model are shared by other RNA and DNA viruses; other aspects are unique to the human retroviruses.

viral entry, target cell tropism, and virus receptor interactions

Human retrovirus transmission requires parenteral or intimate mucous membrane exposure to the virus. Transmission by fomites is unlikely because the viral envelope of retroviruses is easily damaged in the environment by exposure to high temperatures, detergents, and chemical disinfectants and through drying. The concentration of virus in body fluids to which a person is exposed is a very important determinant of transmission.[17] For HIV, the blood level of virus may or may not reflect the virus level in genital fluids and breast milk.[18-20]

Receptors

The cellular receptor for HTLV has not been identified. Only T cells are productively targeted by HTLV, but infection of B cells and other cell types has been detected. Much more is known about the manner by which HIV enters the host cell. Susceptible cells must be available for the process of adsorption, penetration, and uncoating of the virus to occur. The receptor for HIV is the CD4 molecule, which is present on T cells and defines the subset of $CD4^+$ helper-inducer T cells. The HIV-1 gp120env protein binds specifically to the CD4 receptor; in ad-

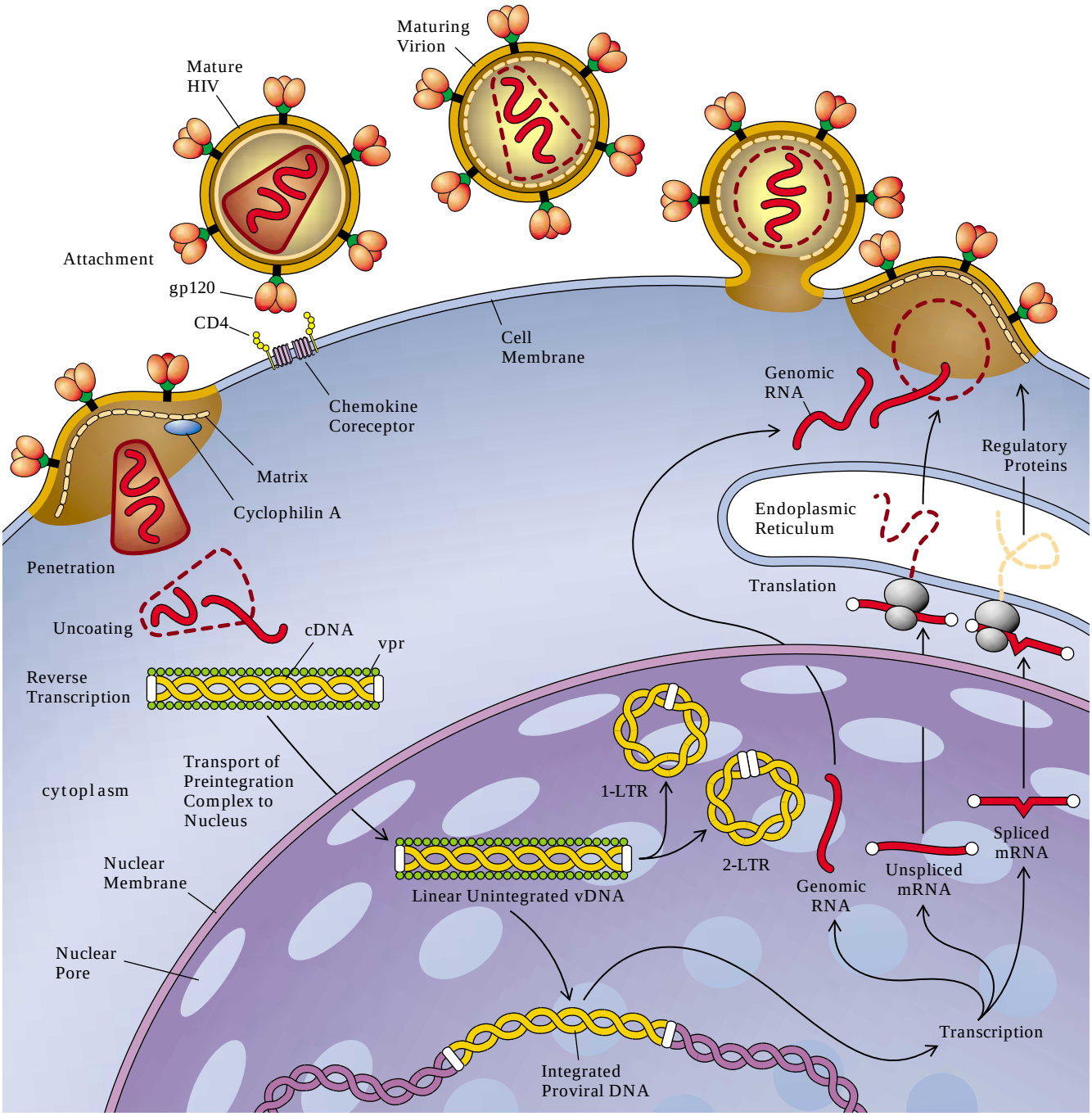


Figure 3 **Illustrated is the replication cycle of HIV-1. The virus must first attach to the CD4 receptor and chemokine coreceptor on the cell surface [*see Figure 2*]. After fusion of the viral envelope with the plasma membrane of the target cell, the nucleocapsid undergoes uncoating, which is facilitated by the presence of cyclophilin A. The viral RNA genome is reverse-transcribed to double-stranded viral DNA (vDNA), which enters the nucleus as a preintegration complex that contains vpr protein and integrase. Only linear vDNA is capable of integrating randomly into the host chromosome; other forms of partially transcribed linear vDNA fragments and 1-long-terminal repeat (LTR) and 2-LTR circularized, episomal vDNA are not capable of integration. The integrated linear vDNA (now termed the provirus) serves as the template for viral transcription. Transcription of the proviral DNA template yields genomic viral RNA, and alternative messenger RNA (mRNA) splicing creates spliced viral mRNA species that encode the viral accessory proteins and the unspliced viral mRNA species that encode the viral structural polyproteins [*see Table 2*]. All of the transcripts are exported to the cytoplasm, where translation processing and assembly begins to occur in the endoplasmic reticulum and Golgi complex. The viral polypeptides, protease, viral RNA, and other constituents of the viral core condense at areas of the plasma membrane that have already accumulated viral envelope proteins (gp120/gp41). Budding of the virion ensues, and the immature virion nucleocapsid core undergoes further proteolytic maturation in the extracellular milieu.[133]**

dition, other cell surface proteins, such as the chemokine receptors, serve as essential coreceptors for different strains of HIV.[15]

Coreceptors

The HIV-1 coreceptors belong to the superfamily of protein-coupled receptors that bind guanosine triphosphate (GTP). These receptors have a characteristic structure of seven transmembrane segments and are coupled with the G protein.[21] The two principal HIV-1 coreceptors, CC chemokine receptor–5 (CCR5) and CXC chemokine receptor–4 (CXCR4), and their chemokine ligands are involved in diverse biologic processes, such as immunomodulation, inflammation, hematopoiesis, and organogenesis.[22] Three cytokines released by $CD8^+$ T cells—RANTES (regulated on activation, normal T cell expressed and secreted), macrophage inflammatory protein–1α (MIP-1α), and MIP-1β—bind to the CCR5 receptor and may suppress HIV macrophage tropic virus. Similarly, stromal cell–derived factor–1α (SDF-1α), which is the natural ligand for CXCR4, and AMD-3100, which is a small molecule inhibitor of SDF-1α, can inhibit R4 virus.[23,24] Clinical development of AMD-3100 was abandoned in May 2001 because of possible cardiac toxicity and limited efficacy. It has been proposed that aberrant signaling through binding of gp120env with the chemokine receptors may cause apoptosis and contribute to the decrease in the $CD4^+$ T cell count in patients with AIDS.[25]

Infection with strains of virus that recognize the CXCR4 coreceptor generally results in cell fusion or syncytium formation in vitro; such strains are termed syncytium-inducing (SI), T-tropic, or R4 strains. Strains of virus that recognize the CCR5 coreceptor generally result in infection without syncytium formation in vitro; such strains are termed non–syncytium-inducing (NSI), monocytotropic, or R5 strains. Dual-tropic strains are able to use both chemokine coreceptors. Other chemokine receptors have also been shown to facilitate HIV infection of susceptible cells in vitro.[26] Thus, chemokine coreceptor use may be a determinant of viral virulence and disease progression.[27] For example, a 32-base-pair deletion in the *CCR5* gene, which is found as a homozygous mutation in less than 1% of the Western European white population, prevents surface expression of the truncated CCR5 receptor protein. This deletion is associated with slower progression of HIV-1 infection and a decrease in HIV-1 infection of peripheral mononuclear blood cells in vitro.[28] Even heterozygosity for this gene defect, which is observed in 20% of Western European whites, may provide some degree of protection against disease progression.[28-30]

HIV-1 gp120env also binds to a dendritic cell–specific C-type lectin, DC-SIGN (dendritic cell–specific intracellular adhesion molecule–grabbing nonintegrin), which is highly expressed on dendritic cells in mucosal tissues.[31] DC-SIGN does not function as a receptor per se for viral entry into dendritic cells; rather, it promotes infection of T cells that express CD4 and chemokine receptors.[32] As such, DC-SIGN efficiently captures HIV-1 in the periphery, stabilizes the virus, and facilitates its transport to secondary lymphoid organs rich in $CD4^+$ T cell targets.[31]

HIV infection is associated with a loss of CD4 antigen on the infected T cell and ultimately leads to the death of the infected T cell. This in turn contributes to a decrease in the ratio of $CD4^+$ T cells to $CD8^+$ T cells—a hallmark of HIV-induced immunosuppression. However, the $CD4^+$ T cell is not the only type of cell subject to infection by HIV. Other cells, such as the monocyte-macrophage, express CD4 molecules and probably disseminate HIV to target organs. In addition, monocyte-macrophages serve as a reservoir for HIV and other lentiviruses. Other cells can be infected with HIV in vitro, including Langerhans cells of the skin, follicular dendritic cells, and several others, some of which do not express the CD4 antigen. For example, astrocytes, oligodendrocytes, M cells of the enteric mucosa, and epithelial cells of the intestine and vagina all express galactosylceramide or a related glycolipid receptor that may act as an alternative virus receptor for gp120env.[33,34]

latency and persistence

For any virus infection to persist, viral gene expression must be restricted to some extent. Three patterns of restricted viral expression are known; all three patterns are important for retroviral infections.[35] The first pattern, latent infection, is characterized by intermittent episodes of acute or subclinical disease; in latent infection, no virus is detectable between episodes. An example of latent infection is herpes simplex virus infection, which is characterized by episodic subclinical shedding. Latent infection may be particularly important for patients infected with HTLV or HIV-1 whose plasma levels of viral RNA have been suppressed below detectable levels through use of antiretroviral therapy. Latent infection must be differentiated from clinical latency, which occurs in the presence of ongoing viral replication.

The second pattern of restricted viral expression is chronic infection. In chronic infection, the virus is usually demonstrable but disease is absent. An example of chronic infection is cytomegalovirus infection. The chronic pattern of infection is seen in approximately one half of the patients infected with HIV-1 who are treated with antiretroviral drugs. In these patients, the plasma levels of HIV-1 RNA are persistently low as a result of potent antiretroviral therapy (HIV-1 RNA counts are greater than 50 RNA copies / ml of plasma but are less than 5,000 copies / ml). However, unlike untreated patients, these patients experience a sustained increase in $CD4^+$ T cells and improved clinical status.

The third pattern of restricted viral expression is persistent infection. This pattern is characterized by a long incubation period with slowly increasing amounts of virus, eventually leading to symptomatic disease. Examples are subacute sclerosing panencephalitis associated with measles virus infection and hepatitis C virus (HCV) infection. This pattern is often seen in HTLV infection, which is characterized by a low probability of developing clinical disease over the infected person's lifetime (the probability is from 5% to 10%) and a period lasting decades in which the patient is asymptomatic. This pattern is also seen in patients with HIV-1 infection; in these patients, the median time to the development of AIDS is approximately 10 years.

Infections with the lentiviruses, particularly HIV, have features characteristic of all three patterns but are best described as persistent and cytopathic. HTLV infections have the characteristic features of latent and persistent infections, but in addition, the infected cell may be transformed.

Both the HIV and HTLV cDNA become integrated with the genome of the infected host cell at a single random site to establish the provirus as a permanent chromosomal resident.[4] In addition, most of the HIV cDNA remains in an extrachromosomal state, either as linear cDNA fragments or as circularized forms (neither of which are replicatively competent). The degree of viral transcription and translation depends on the stage of differentiation of the infected $CD4^+$ T cell. That stage, in turn, is probably related to exogenous antigen stimulation and the action of cytokines (e.g., IL-2) and viral regulatory proteins (e.g., the vpr

protein), which promote cell differentiation and activation of viral promotors (e.g., nuclear factor–κB [NF-κB] inducible transcription factors). Integration of human retroviruses differs from that of certain animal retroviruses in that no transforming genes (oncogenes) are associated with the viral genome, and the proviral genome is not regularly inserted next to a host-transforming gene. Moreover, clinically important human retroviruses are exogenous and are not transmitted in germ cells, as are some endogenous vertebrate retroviruses and the clinically unimportant HERVs [*see Table 1*].

cytopathicity

The cytopathic destruction of lymphoid cells is a characteristic of lentivirus infection in general. In a minority of patients with HIV, an initial period of intense viral replication is followed by an acute and rapidly progressive disease; in most patients with HIV, however, this initial period of viral replication is followed by a period of persistent infection, with clinical disease developing only after a prolonged period. In some patients in whom disease progression is very slow, chronic infection can develop, and the development of such chronic infection is a viable long-term clinical objective of antiretroviral therapy.

The clinical manifestations of HIV are governed by cell and tissue tropism; the clinical signs and symptoms of infection arise directly or indirectly from viral replication within these cells and tissues. The sine qua non of HIV infection is $CD4^+$ T cell depletion, immunosuppression, and the development of opportunistic infections and malignancies. This depletion of $CD4^+$ T cells occurs not only by direct viral replication and cell lysis but also through other mechanisms. For example, HIV env-induced cell fusion may result in syncytia formation; cell surface expression of the *Fas* gene and programmed cell death (apoptosis); aberrant signaling through chemokine receptors; arrest of proliferation at the G_2 phase of the cell cycle, with cytopathic ballooning of the cells; and antibody-dependent or cell-mediated depletion induced by the binding of gp120env. In addition, the functional responses of T cells, such as signaling, are impaired both in vitro and in vivo because of the binding of gp120env with the CD4 molecule.[36]

A number of other cellular and lymphoid abnormalities are observed in persons infected with HIV, including abnormal B cell immunity and humoral immunity, monocyte-macrophage dysfunction, and immune defects in natural killer cells, among many observed immunologic defects.[36] Moreover, the follicular dendritic cells (FDCs) may play a key role in maintaining a large reservoir of protected infectious virus particles on the dendritic cell surface; although these cells do not appear to be infected, the presence of virus particles on the cell surface results in chronic immune activation and cytokine secretion, which leads to further virus replication and dissemination. Eventually, the FDCs are destroyed, which in turn leads to the destruction of the lymphoid architecture and the evolution of clinical disease.[36] In many instances, lymphoid architecture may be partially restored after prolonged antiretroviral therapy.

genomic diversity

HIV undergoes active and continuous replication in the infected person, with infected cells dying after approximately 2 days.[37,38] The level of plasma HIV particles remains relatively constant for each infected person because of the balance between rapid production and clearance of virions. To maintain this steady-state level of HIV RNA, there are an estimated one billion infectious events every day in the untreated person with HIV infection. As such, the large number of replication cycles—estimated to be 300 cycles or more a year—provide ample opportunity for the virus to develop genetic diversity as a primary defence against immune and antiretroviral therapy suppression of replication. For example, within the *env* gene, localized regions of extraordinary hypervariability are interspersed with well-conserved (i.e., nonvariable) regions. One such hypervariable region is the V3 loop, which is a target for neutralizing antibodies. The force driving this variability is probably caused by selective pressure from the host's immune system.

This variability within HIV *env* gene contrasts with the relative genetic stability of HTLV. The mutation rate in the HIV *env* gene is one million times greater than that of DNA viruses and 10 to 100 times greater than that of other retroviruses; this results in a production of so-called quasispecies and is comparable only to the variability noted for another RNA virus, HCV. This extraordinary *env* gene hypervariability is associated with amino acid changes of as much as 35% or more among divergent HIV-1 isolates. Variation arises during reverse transcription of viral RNA to cDNA, a step that is highly prone to error, that lacks proofreading (i.e., correction of transcriptional copying errors) because the RNA template is destroyed by RNase H, and that leads to substitutions and misreading, which may result in the modification of several biologic properties, including tissue tropism, virulence, replication rate, and susceptibility to antiretroviral agents. In addition, recombination between different subtypes has been shown to be an important mechanism for generating diversity. Ultimately, continued genetic diversity of HIV may allow the virus to evade containment by the immune system.[39]

Origins of Human Retroviruses

The extent to which different human and primate retroviruses are related can be determined by comparing nucleic acid sequences through a process called phylogenetic analysis. Through such analysis, the human retroviruses have been shown to be phylogenetically related to the retroviruses of Old World primates. The closest viral relative of HTLV-I is the simian T cell lymphotropic virus type I (STLV-I), which was identified as an agent of naturally occurring infection in many species of Old World monkeys and great apes. Antecedent infection of the human population probably occurred from three geographically distinct interspecies transmission events; once established, the virus followed the migratory patterns of people around the world.[40] Moreover, recent studies have convincingly shown that there have been three separate interspecies transfers of the lentivirus simian immunodeficiency virus (SIV_{cpz}) from chimpanzees (*Pan troglodytes*) to humans; these transmission events occurred within the past century and resulted in the establishment of HIV-1 groups M, N, and O.[41,42] Similarly, HIV-2 has its origin with multiple interspecies transfers of SIV_{smm} from the primate sooty mangaby (*Cerocebus atys*) to man and to primates in captivity (e.g., SIV_{mac} in the rhesus monkey *Macaca mulatta*).

Human T Cell Lymphotropic Viruses

Retroviruses cause many animal malignancies and characteristically immortalize cells in vitro. In humans, the prototype retroviruses associated with malignant transformation are

HTLV-I (and possibly HTLV-II). HTLV-I causes adult T cell leukemia (ATL); direct infection of the central nervous system results in a spastic or ataxic myelopathy termed tropical spastic paraparesis, also referred to as HTLV-I–associated myelopathy (HAM). HTLV-II was first isolated from a patient with a T cell variant of hairy-cell leukemia; it may be associated with certain neurologic, hematologic, and dermatologic diseases, but unequivocal evidence that HTLV-II is the etiologic agent is lacking.[43,44]

HTLV-I and HTLV-II have similar genomic organizations and share approximately 60% nucleotide homology[45] [*see Figure 2*]. The *gag* gene encodes for the structural proteins p19, p24, and p15. The *pol* gene encodes for the protease and the reverse transcriptase. The *env* gene encodes for the external envelope and transmembrane glycoproteins gp46 and gp21. The spliced regulatory proteins tax and rex and the other three open reading frames constitute the 9 kb genome. The gag and env proteins are most immunogenic; the antibodies to these proteins are commonly detected by enzyme immunoassay (EIA) and Western blot (WB) assay.

epidemiology

HTLV-1 has a worldwide distribution; prevalence rates range from 5% to 27%, with the higher rates occurring in certain populations in which HTLV-1 is highly endemic.[46] HTLV-II is endemic in several Native American populations in the Americas and in Pygmy tribes in central Africa; in these populations, prevalence rates range from 7% to 9%.[44] In the United States, the seroprevalence rates of HTLV-I and HTLV-II range from 7% to 49% in injection drug users and prostitutes.[46] Early serologic and epidemiologic studies are confusing, because these studies were unable to effectively differentiate between HTLV-I and HTLV-II; however, this problem has been corrected through the use of more specific recombinant peptide-based WB and nucleic acid amplification methods.[47]

classification of htlv subtypes

The HTLV genome is highly conserved, but greater nucleotide convergence in the LTRs has made possible the development of restriction fragment length polymorphism (RFLP) testing. Through such testing, it is possible to classify both HTLV-I and HTLV-II into genotypic subtypes, and valuable information on viral transmission has become available.[47] There are five major molecular and geographic subtypes of HTLV-I: cosmopolitan (worldwide), Japanese, West African, Central African, and Melanesian.[48] There are three HTLV-II subtypes that reflect population clustering rather than geographic clustering: IIa, IIb, and IIc. Subtype IIa is found in injection drug users worldwide; subtype IIb is found primarily in Native Americans; and subtype IIc is found in Brazilian tribes.[49]

overview of laboratory diagnosis of htlv

Screening of blood donors for HTLV-I and HTLV-II is now routinely used in the United States, Canada, several Caribbean countries, Europe, and Japan.[47] The primary screening assay is the EIA; WB assay is used for confirmatory testing. However, EIA cannot distinguish between HTLV-I and HTLV-II infections because of the significant protein homology between the two viruses. Importantly, antibodies to HTLV do not cross-react with HIV proteins. After repeatedly positive results on EIA, the diagnosis of HTLV infection is confirmed if antibodies to two gene products (gag and env proteins) are detected on WB assay. For example, a specimen demonstrating antibody reactivity to p24gag and to pg46env or gp61/68env, or both, is considered to be positive for HTLV-I or HTLV-II. For specimens that test positive on EIA and that react with any of the WB bands but that do not meet this criterion, results are considered indeterminate. Specimens that test positive on EIA but that display no immunoreactivity to any of the HTLV WB bands are considered to be negative for antibodies to HTLV-I and HTLV-II, and the result of the EIA is considered to be false positive. To better differentiate between HTLV-I and HTLV-II, the WB assay has been modified to contain type-specific recombinant proteins from the external glycoprotein of HTVL-I (rgp46envI) and HTLV-II (rgp46envII) as well as a truncated recombinant peptide, rp21e, from the transmembrane glycoprotein gp21env. In one study, persons with indeterminate WB assay profiles (i.e., persons who did not demonstrate antibody reactivity to p24, p19, or rp21e) who did not have risk factors for HTLV infection were shown by polymerase chain reaction amplification not to be infected with either HTLV-I or HTLV-II.[50] Such indeterminate WB assay results appear to represent antibodies to other viral and cellular antigens that cross-react with HTLV proteins. Nevertheless, blood donors with indeterminate results are deferred and their blood is excluded for transfusion purposes.

PCR has become the reference method for determining infectious status; for validating serologic assays; for distinguishing between HTLV-I and HTLV-II; for studying in vivo viral load and tissue distribution; and for further evaluating patients with risk factors for HTLV infection whose serologic status is either indeterminate or negative.[47] PCR is also an important method of testing infants for HTLV infection, because their serologic status may be unclear, owing to the presence of passively transferred maternal antibodies. In addition, PCR is important for the detection of infection in the window period between exposure, infection, and seroconversion.[47]

HTLV-I

HTLV-I is the etiologic agent of adult T cell leukemia/lymphoma (ATL) and a chronic progressive inflammatory neurologic degenerative disorder known as tropical spastic paraparesis or, more commonly, HTLV-I–associated myelopathy (HAM). However, this virus has been implicated in several other disorders, including an inflammatory arthropathy, uveitis, polymyositis, infectious dermatitis in children, pulmonary disorders, and Sjögren syndrome. There are provocative but unproven associations of HTLV-I with mycosis fungoides and Sézary syndrome. This discussion focuses on ATL and HAM because of the established etiologic linkage to HTLV-I.

epidemiology

HTLV-I is endemic in southwestern Japan and Okinawa, where more than one million persons are infected. HTLV-I infection is also prevalent in Taiwan, the Caribbean basin (including areas of South America and the southeastern United States, central Africa, Israel, and the Arctic, where seroprevalence rates range from 5% to 27% in adults.[46] In the United States, HTLV seroprevalence rates range from 0.025% in blood donors to 7% to 49% in injection drug users and prostitutes.[46] The incidence rates of seroconversion associated with HTLV are estimated to be 1.09 per 100,000 persons per year.[51] HTLV-I infection is transmitted from mother to child; perinatally; sexually; through breast milk and blood transfusions; and from contaminated

needles.[52] In the United States, where blood donors are screened for HTLV, the risk of transmitting HTLV through a unit of blood is estimated to be 1 in 641,000 (95% confidence interval, 256,000 to 2,000,000).[51]

The distribution of both ATL and HAM overlaps the distribution of HTLV-I, with more than 95% of affected persons having serologic evidence of HTLV-I infection. Although early serologic studies suggested that HTLV-I infections were common in injection drug users, confirmatory serologic assays that reliably distinguish between HTLV-I and HTLV-II have subsequently shown that the majority of HTLV infections in injection drug users are HTLV-II infections.[53] The development of ATL in persons infected by blood products is rare; however, 20% of patients with HAM acquire HTLV-I infection from contaminated blood. In ATL, there is a latent period between infection and the emergence of disease of 20 to 30 years or more; in contrast, HAM has a shorter median latency of approximately 3 years, but the latency period can be as long as 20 to 30 years.

Cross-sectional studies have suggested that coinfection with HTLV-I and HIV-1 results in an acceleration of clinical immunosuppression and shortened survival. However, prospective studies of dual infection are not available, and other factors associated with dual infection could be responsible for the clinical findings. Of interest, dual infection with HTLV-I and HIV-1 may affect the interpretation of the results of immunophenotyping because coinfection alters the relationship between $CD4^+$ T cell count and HIV-1 disease stage. As such, patients infected with HIV-1 who have a higher than expected $CD4^+$ T cell count for a given level of clinical immunosuppression should be tested for HTLV-I infection.[16]

pathogenesis

The pathophysiologic basis of HTLV-I disease is less well known than that of HIV-1. Unlike HIV-1, the cellular receptor for HTLV-I has not been identified. $CD4^+$ T cells are productively infected by HTLV-I, but some other cell types, including B cells and $CD8^+$ T cells, are infected occasionally. Like HIV-1, the reverse-transcribed viral cDNA integrates randomly into the host cell to establish the provirus, but in contrast to HIV-1, HTLV-I establishes a latent infection with infrequent expression of viral gene products. As such, HTLV-I has a very low level of disease penetrance; the transformation of an infected cell is a rare event, and the cumulative lifetime risk of developing ATL is 1% to 5% in persons infected with HTLV-I. The latent period from infection to clinical disease is estimated to be 30 to 50 years[54]; most persons with ATL appear to have acquired the infection in childhood.[55]

In contrast to many of the animal retroviruses, and in particular avian retroviruses, HTLV-I does not contain a transforming oncogene. Moreover, the integrated provirus does not locate at a unique site in the genome that would result in the expression of a cellular gene (proto-oncogene), as does, for example, the classical avian leukosis virus, which causes B cell leukemia in birds by inducing the expression of the proto-oncogene *myc*. The only viral gene product expressed in tumor cells transformed by HTLV-I in vivo is tax, a 40 kd protein with exons on either side of the *env* gene. In contrast, the infection of $CD4^+$ T cells in vitro results in transformed cells that produce complete, infectious HTLV-I virions. Tax protein induces the expression of a wide range of host cell proteins, including transcription factors and cytokines (IL-2, tumor necrosis factor [TNF], and others). However, it is not known how this induction of host cell factors leads to neoplastic transformation, and no consistent chromosomal abnormalities have been reported in ATL. The discordant observations between the in vitro and in vivo milieus are a recurrent theme in understanding the complex biology of human retroviral disease.

The pathogenesis of HAM is virus induced. Interestingly, patients with HAM generally have high proviral levels in the blood and a stronger immune response to HTLV-I, with higher antibody titers in the cerebrospinal fluid than in the serum. It has been proposed that the pathology of CNS disease may result from an autoimmune destruction of neuronal tissue by viral-specific $CD8^+$ T cells. HTLV-I has also been associated with uveitis, arthropathy, and infective dermatitis; each of these conditions probably has an autoimmune / inflammatory pathogenesis.[16]

clinical features and diagnosis

HTLV-I–Induced Adult T Cell Leukemia/Lymphoma

Four clinical presentations of ATL have been described: acute, lymphomatous, chronic, and smoldering.[54] All of these malignancies are characterized by the monoclonal expansion of $CD4^+$ T cells that contain HTLV-I provirus and rearrangements of clonal T cell receptor genes. HTLV-I virions are not recovered directly from the infected T cells, but viral tax protein is expressed in vivo. The molecular pathogenesis of this retroviral-induced neoplasm has not been fully elucidated.

Acute ATL Acute ATL accounts for approximately 60% to 80% of cases of ATL. Patients experience a short clinical prodrome, with an average of 2 weeks between the onset of symptoms and diagnosis. ATL has an aggressive natural history; median survival is 6 months. The clinical picture is characterized by rapidly progressive skin lesions (40% of cases ranging from maculopapular rashes to tumorous lesions), pulmonary infiltrates, diarrhea, hypercalcemia (seen in 50% of cases), elevated lactate dehydrogenase (LDH) and alkaline phosphatase levels, and lymphocytosis with pleomorphic mononuclear cells (termed flower cells) that contain lobulated, cloven-hoof–shaped nuclei. The skin lesions may be difficult to distinguish from those in mycosis fungoides and Sézary syndrome. The pulmonary lesions may be the result of a leukemic infiltrate or an opportunistic infection, such as *Pneumocystis carinii* infection or fungal infections. Diarrhea is almost always associated with an opportunistic infection. Hepatosplenomegaly may be present. Lytic bone lesions are common; the lesions are patchy and are composed of osteolytic cells without osteoblastic activity. Leptomeningeal involvement associated with weakness, altered mental status, paresthesias, and headache is found in approximately 10% of patients. The cerebrospinal fluid protein level is usually normal. The diagnosis is confirmed by a finding of malignant cells in the CSF.

Lymphomatous ATL Lymphomatous ATL accounts for approximately 20% of cases of ATL. Lymphomatous ATL is similar to acute ATL but differs in that lymphadenopathy is present and there is an absence of circulating, morphologically abnormal lymphocytes. The diagnosis may be suspected on the basis of the patient's geographic birthplace and the presence of skin lesions and hypercalcemia. The diagnosis is confirmed by HTLV-I serologic testing or the detection of the integrated provirus on PCR testing.

Chronic ATL Chronic ATL accounts for approximately 15% of cases of ATL. This form of ATL is characterized by normal serum levels of calcium, elevated levels of LDH, and the absence of bony involvement or involvement of the CNS or gastrointestinal tract. The median survival for these patients is 2 years. Chronic ATL may progress to acute ATL.

Smoldering ATL Smoldering ATL accounts for fewer than 5% of cases of ATL. Although fewer than 5% of the circulating lymphocytes display the morphologic abnormalities characteristic of ATL or an integrated HTLV-I provirus, there is no hypercalcemia, adenopathy, or hepatosplenomegaly. The CNS, bones, and GI tract are not involved, but typical skin lesions and pulmonary infiltrates may be present. The median survival for patients with the smoldering form of ATL is 5 years or longer.

HTLV-1–Associated Myelopathy/Tropical Spastic Paraparesis

HAM is a slowly progressive thoracic myelopathy that leaves one third of patients bedridden within 10 years of diagnosis.[56] HAM affects women disproportionately for reasons that are not understood.

The onset of symptoms in HAM is insidious; symptoms include weakness or stiffness in one or both legs, back pain, and urinary incontinence. Peripheral neuropathy is usually mild. Physical findings include spastic paraparesis or paraplegia with hyperreflexia, ankle clonus, and extensor plantar responses (Babinski reflex). Cognitive function is usually spared, and cranial nerve involvement is distinctly unusual. Demyelinating lesions are seen in the white matter and paraventricular regions of the brain and spinal cord; the clinical presentation may thus resemble multiple sclerosis or the myelopathy of HIV infection. The pathologic changes in HAM include symmetrical degeneration of the lateral columns, corticospinal tracts, and posterior columns. An inflammatory infiltrate is present on the spinal meninges, and spinal cord parenchyma shows evidence of myelin destruction, similar to that in multiple sclerosis. HTLV-I provirus is usually not found in the cells of the CNS parenchyma, but it may be detected in a few CSF-associated lymphocytes. Although a high proviral serum level of HTLV-I is an important risk factor in the development of HAM, the presence of provirus in the blood is not sufficient to cause disease, and the immunopathogenesis of HAM is probably associated with host genetic factors.

prevention and treatment

The principles of prevention of HIV infection also apply to the prevention of HTLV infection—avoidance of breast-feeding, if possible; screening of blood donors for HTLV antibody; safe sexual practice; and the avoidance of sharing or reusing of needles.

There is no proven, effective therapy for either ATL or HAM.[55,57] Unlike with HIV disease, disease progression of ATL and HAM is not associated with HTLV-I replication; as a result, antiretroviral drugs have not been effective in the treatment of these diseases. The neoplasm is responsive to combination chemotherapy that is used against other forms of lymphoma, but patients usually experience relapse within 1 year of initial remission. The combination of interferon alfa and zidovudine (AZT) is effective therapy and is associated with less toxicity than standard lymphoma regimens; however, the effectiveness of this regimen is most likely the result of the cytotoxic effects of AZT rather than its antiretroviral activity. Anecdotal cases involving the successful use of allogeneic bone marrow transplantation have been reported. Patients with chronic ATL or smoldering ATL should be treated for opportunistic infections and observed for progression to acute ATL, at which time chemotherapy should be initiated. In ATL, hypercalcemia is generally controlled with glucocorticoid therapy and cytotoxic therapy of the neoplasm. In HAM-associated inflammation, glucocorticoids reduce inflammation and may improve uveitis, but antiretroviral therapies are not effective. Danazol, an androgenic steroid, has been reported to benefit women with HAM-associated urinary and fecal incontinence.

HTLV-II

HTLV-II is endemic in certain Native American tribes of North and South America. The modes of transmission are the same as for HTLV-I, although HTLV-II is less readily transmitted sexually than HTLV-I . Because of the similarity in biologic properties between the two viruses and because of the initial inability to distinguish the two viruses serologically, the two viruses were grouped together in early studies. The seroprevalence of HTLV-II in injection drug users in several metropolitan drug treatment programs in the United States was approximately 20%.[58] Approximately 3% of these persons were infected with both HTLV-II and HIV-1. Women are significantly more likely to be infected with HTLV-II than men, a feature of the disease that may reflect the more efficient sexual transmission of HTLV-II from men to women than from women to men.

Although HTLV-II was isolated from a patient with a T cell variant of hairy-cell leukemia,[1] the virus has not been unequivocally associated with a particular disease. Nevertheless, evidence is slowly accumulating that HTLV-I may play a role in certain degenerative neurologic, hematologic, and dermatologic diseases**.** No effective therapy exists for HTLV-II infection.

HIV

The pathogenesis, epidemiology, clinical features, and treatment of HIV-1 infection are described elsewhere [*see 7:XXXIII HIV and AIDS*]. This subsection discusses the principles of laboratory diagnosis and the basis for the laboratory monitoring of HIV-1 infection. It is important that therapy be individualized on the basis of a thorough understanding of the host-virus relationship in the individual patient (see above). In this regard, an understanding of the use of the laboratory for diagnosing and monitoring HIV-1 disease is critical.[59]

laboratory diagnosis

Serologic Diagnosis

The HIV-1 testing algorithm recommended by the United States Public Health Service consists of initial screening with an EIA licensed by the Food and Drug Administration, followed by testing of repeatedly reactive specimens with an FDA-licensed supplemental test (i.e., WB assay or immunofluorescent assay).[60] Although EIAs are highly sensitive and specific, the positive predictive value of EIA is highly dependent on the seroprevalence of HIV-1 antibody in the population from which the individual is being tested. As such, use of both EIA and a supplementary test increases the likelihood of detecting HIV-1 infection. For maximum diagnostic accuracy, the laboratory HIV test results should always be interpreted by the clinician in

conjunction with the clinical and epidemiologic history of the person being tested. At this time, FDA-licensed HIV nucleic acid detection kits are not approved for diagnosis or for supplemental testing but only for disease monitoring (see below).

Enzyme immunoassays The current generation of EIAs have shortened the estimated antibody-negative window period of primary infection to approximately 1 month or less.[61] The specificities of the current commercial EIAs are over 99.5%.[62] False positive reactions arise from nonspecific cross-reacting antibodies in the serum of persons with underlying immunologic disease and of those who are gravid, who have received multiple transfusions, or who have recently received an immunization.[63] Several EIAs that may be used to screen for both HIV-1 and HIV-2 antibody are commercially available.

Immunoblot WB assay detects the serum antibodies directed against specific HIV proteins of varying molecular weights after separation of the viral proteins by gel electrophoresis and blotting onto nitrocellulose paper. WB assay detects antibodies to the following specific HIV-1 proteins: core (p17gag, p24gag, and the precursor p55gag); polymerase (p31pol, p51pol, and p66pol); and envelope (gp41env and gp120env / 160env). The reported analytic specificity of the WB assay is 97.8%.[60] A WB assay is interpreted as being negative when no antibody-antigen band is present; it is interpreted as being positive when antibodies are present to core proteins (p24gag), envelope proteins (gp41env or gp120env / 160env), and, in some cases, polymerase proteins (p31pol). Although several organizations have proposed criteria for interpreting WB assay reactivity, the Centers for Disease Control and Prevention (CDC) endorses interpretative criteria that require the presence of antibodies to two of the three following proteins: p24gag, gp41env, and gp120env / 160env.[60]

Regardless of the HIV-1 antibody seroprevalence, a reactive EIA and confirmatory WB assay together have a positive predictive value of greater than 99.99%.[64] In the blood-donor population, approximately 10 of 10,000 persons (0.1%) who are without risk for HIV-1 infection will have repeatedly reactive results on the HIV-1 EIA. Eight of those 10 low-risk persons with repeatedly reactive results on the HIV-1 EIA will have negative results on the HIV-1 WB assay, and two will have indeterminate results. False positive results for HIV-1 antibody (i.e., both EIA and WB assay are reactive for a person who is not infected with HIV-1) are extremely rare (less than one in 100,000 persons screened).[64] Therefore, an indeterminate WB assay result is more common than a false positive WB assay result.[65]

With the increased use of HIV-1 antibody screening in low-risk populations, including health care workers, it is essential that HIV-1 test results be interpreted accurately. Between 4% and 20% of serum samples that are repeatedly reactive on HIV-1 EIA are interpreted as indeterminate on WB assay.[66] Indeterminate WB assays (IWB assays) in persons infected with HIV-1 may result from early formation of antibody against viral core antigens during primary infection. IWB assays may also occur in conjunction with early detection of HIV-1 antibody by the more sensitive third-generation EIAs. IWB assays rarely occur as a result of the loss of core-specific antibody late in infection because of severe immunosuppression.[67,68] The presence of cross-reacting antibody to HIV-2 has been implicated as a cause of positive results in HIV-1 testing. False positive results on both EIA and WB assay are extremely uncommon, occurring with a frequency of fewer than one in 135,000 tests.[64]

The following recommendations have been proposed for the clinical management of patients with an IWB assay result.[63] Low-risk individuals who have a nonreactive result on repeat testing with EIAs do not need further follow-up. High-risk individuals should be followed serologically for at least 6 months; this is especially important for those with a p24gag band on WB assay. Such a result may reflect the presence of HIV-1 antibody in conjunction with early seroconversion. The early, selective use of supplemental tests for HIV-1 proviral cDNA or plasma RNA may help determine the infection status of high-risk individuals before full seroconversion occurs.[63,66,69] Seroconversion is usually identified within 3 months of primary infection through the use of appropriate supplemental testing.[61] Importantly, negative supplemental test results may help alleviate the anxiety associated with an indeterminate HIV-1 serologic result.[66]

Rapid serologic testing Rapid, reliable, and less expensive alternatives to the use of EIA with a confirmatory WB assay have been sought for use in acute care settings, emergency departments, clinics for sexually transmitted disease, medical field settings, and developing countries. Consideration should be given to the introduction of rapid screening for HIV-1 antibody in certain clinical settings, such as clinics for sexually transmitted disease; such measures would greatly enhance testing programs by preventing delays in the counseling of seronegative patients and by providing preliminary results to seropositive patients.[70] These preliminary results may encourage patients to return for confirmatory testing and to adopt risk-reducing behaviors sooner than they would if following the currently accepted testing algorithms.[71] In addition, the use of rapid screening for HIV-1 after occupational exposure to blood will reduce the duration of antiretroviral prophylaxis therapy for the exposed health care worker, thus minimizing drug-related toxicity and cost and alleviating anxiety sooner should the test result prove negative.[72,73] Rapid serologic testing will likely become the standard of diagnostic care; physicians and other health care professionals should become familiar with the indications for rapid testing.

Detection of HIV-1 subtypes HIV-1 groups and subtypes differ with regard to their diagnostic properties on both serologic and nucleic acid assays. As such, reagents in diagnostic kits have been modified to ensure optimal sensitivity and specificity for antibodies to viruses of groups M and O. Current clinical HIV-1 antibody testing does not distinguish between the various HIV-1 subtypes. However, from an epidemiologic perspective, distinguishing these subtypes by nucleic acid sequencing is very important for understanding the global spread of HIV-1.

As mentioned previously, HIV-1 strains are divided into three groups: M, O, and, most recently, N.[74] To date, in the United States, infection with group O is uncommon, and no group N infections have been reported.[75] Neutral variation in HIV-1 strains of group M has led to a further subdivision of these viruses into nine geographically distributed subtypes, or clades, designated A through K. Subtype B is the most common subtype in the United States and Europe; subtypes A and C are prominent in Africa and Asia. In addition, there are four major circulating recombinant forms: AE is prevalent in Southeast Asia; AG, in west and central Africa; AGI, in Cyprus and

Greece; and AB, in Russia. However, the circulating recombinant forms are expanding in number as more full-genome sequences of HIV-1 become available (e.g., FD, from the Democratic Republic of Congo; BC, from China; and several additional complex recombinants that combine three or more subtypes). Although there are differences in the biologic properties of these subtypes (e.g., there is a low frequency of syncytium-inducing phenotype in subtype C), the overall data are too limited to support any strong influence of subtype on either disease progression or transmission.

Because most of the primer pairs for HIV-1 cDNA PCR amplification have been optimized for group B viruses (see below), it is not surprising that HIV-1 cDNA PCR may also fail to detect HIV-1 group O and some subtypes of groups M and N. To accommodate this deficiency, primer pair modifications have been incorporated into the recent Roche Amplicor HIV-1 DNA and Roche Monitor HIV-1 RNA assays (version 1.5).[76] Because of the large number of *pol*-specific synthetic oligonucleotide target probes used by the branched-chain DNA (bDNA) assay, detection of group O and different group B subtypes has not been a problem for the bDNA assay.[77]

Detection of HIV-2 antibodies HIV-2 infection is less geographically dispersed than HIV-1 infection; the epidemic of HIV-2 is primarily focused in West Africa, and is common in the epidemic in India. In the United States, only a relatively few cases of HIV-2 have been reported. In blood donors in the United States, HIV-2 is extremely rare; only three cases of HIV-2 were detected in screening 74 million blood donations through June 1995.[78] Of the 62 persons reported with HIV-2 infection in the United States, 44 (71%) were born in, had traveled to, or had a sex partner from western Africa. Nevertheless, diagnosis of HIV-2 infection will continue to be an emerging problem in the United States and many other countries, and antibody screening for both viruses is warranted.

HIV-1 and HIV-2 genomes have about 60% homology in conserved genes, such as *gag* and *pol*, and 35% to 45% homology in the *env* gene.[79] The core proteins of HIV-1 and HIV-2 display frequent cross-reactivity, whereas the envelope proteins are more type specific. Despite this cross-reactivity, it has been estimated that the first- and second-generation EIAs used in the United States for screening of blood donors for HIV-1 detect 55% to 91% of HIV-2 infections.[80] The WB assay that is used to detect HIV-1 antibodies may yield positive, negative, or indeterminate results when used to assess sera known to be positive for HIV-2 (when the WB assay is used to confirm a positive result on EIA for HIV-2, p26gag and gp36env correspond to their HIV-1 counterparts p24gag and gp41env, respectively).

Laboratory Detection of HIV-1

The methods of direct detection of HIV-1 include the use of culture, antigen detection, and detection of nucleic acid. Each method has a specific role to play: use of culture is important in research; use of antigen detection is important to public health; and detection of nucleic acid is important for individual patient care. Near the beginning of the HIV-1 pandemic, viral culture and antigen detection were used to evaluate the effectiveness of antiretroviral therapy. However, both methods were replaced by more sensitive and precise methods involving the quantification of viral RNA in plasma through nucleic acid amplification or signal amplification. Nevertheless, both culture and antigen detection continue to have useful roles.

Viral culture Use of mixed-lymphocyte coculture for the detection of HIV-1 is a specialized procedure that has extremely high specificity but lower sensitivity in patients with high $CD4^+$ T cell counts, as compared to methods used to detect viral nucleic acid (see below).[81,82] Use of HIV-1 culture is restricted primarily to research laboratories because of its lower sensitivity than currently available nucleic acid detection methods, cost, time constraints, and highly specialized, technical nature. However, interest is being shown in the use of HIV-1 coculture for assessing viral containment after potent antiretroviral therapy[83]; HIV-1 coculture may also be useful for obtaining primary clinical HIV-1 isolates to study viral syncytium-inducing phenotype and drug-susceptibility phenotype[84] and for assessing viral fitness.[85]

Measurement of HIV-1 p24 antigen With the advent of nucleic acid amplification methods for the monitoring of HIV-1, the measurement of HIV-1 p24 antigen has a much more limited role than it once did. Currently, the primary use of p24 antigen detection is to identify patients who are in the antibody-negative window period of acute HIV-1 infection. Although antigen detection is a less expensive alternative to viral RNA detection in this setting, both the determination of viral RNA levels and the culturing of peripheral blood mononuclear cells are significantly more sensitive than detection of p24 antigenemia,[86] even with the added sensitivity of p24 antigen acid dissociation.[87] However, a tyramide signal-amplification–boosted EIA for quantification of p24 antigen has been reported to have equivalent sensitivity to viral RNA reverse transcriptase polymerase chain reaction (RT-PCR) amplification at 200 to 400 copies / ml.[87,88] Currently, the greatest use of p24 antigen testing is for screening the United States blood supply. This screening program was introduced in 1996 and has been augmented by a plasma HIV-1 RNA screening assay[89,90] to further lower the residual risk of HIV-1 infection, which is estimated to be less than one in 500,000 units (95% confidence interval, 200,000 to 2,800,000).[51]

Detection of viral nucleic acid Commercially available methods for the detection of viral nucleic acid (proviral cDNA or viral RNA) are specific and sensitive and are an effective method for identifying those persons who are infected but who have not undergone seroconversion.[86,91] These methods are also useful in identifying infected infants[92] and in resolving indeterminate HIV-1 antibody serology.[66] In addition, the quantification of plasma viral RNA has assumed a critically important clinical role in assessing disease prognosis and response to antiretroviral therapy.[93-97]

Qualitative HIV-1 DNA PCR amplification is the most commonly used assay for the diagnosis of HIV-1 infection in neonates and infants.[98,99] The Roche Amplicor HIV-1 test (Roche Diagnostic Systems, Inc.) is the only FDA-licensed commercial kit available for clinical use.

The major advantages of HIV-1 DNA PCR over culture are its increased sensitivity and more rapid reporting time (the reporting time is 1 day, as compared to 2 to 4 weeks). There is always a risk of false positive reactivity caused by contamination of the specimen with amplicons (so-called carryover product contamination) or by specimen handling errors[69,99,100-102]; carryover contamination is decreased somewhat by the use of the uracyl *N*–glycosylase enzyme in the commercial assay. False negative results can also occur because of inhibition of the PCR reaction by either hemoglobin or heparin or because there are

fewer target cells in the assay than expected. To control for the latter and improve the precision of the assay, testing for HIV-1 DNA should include concurrent amplification of a cell-associated host gene, such as *HLA-DQ* or the β-globin gene. Diagnostic laboratories should participate in a quality-assurance program to ensure that problems with sensitivity and specificity are quickly identified.[103]

RT-PCR amplification is generally more sensitive than p24 antigen EIA in detecting plasma HIV-1 RNA.[86-88,91] As such, there is much interest in using plasma HIV-1 RNA in diagnostic testing.[69,100,104] As mentioned (see above), HIV-1 RNA has augmented HIV-1 p24 antigen screening; it will likely replace HIV-1 p24 antigen screening of donor blood in the United States after its use has been validated. However, the HIV-1 RNA assays that are commercially available are not currently licensed for use in screening. To avoid incorrectly diagnosing a patient as having HIV-1, the HIV-1 RNA assay should be used diagnostically only as a supplemental test for detecting antibody-negative acute infection. In this particular diagnostic setting, a positive result on HIV-1 RNA assay (particularly one that indicates a viral RNA level of fewer than 1,000 copies / ml) should be confirmed by either another nucleic acid technology (preferably, HIV-1 DNA PCR) or HIV-1 p24 antigen EIA or HIV-1 culture, if available.[100,102,105,106] Alternatively, one can again test for the development of HIV-1–specific antibody, which should be present shortly after viral RNA reaches detectable levels.[107] The presence of HIV-1 RNA in the absence of other laboratory findings requires a correlation with the medical and epidemiologic history and a confirmatory repeat blood test for HIV-1.[100-102,106]

HIV-1 RNA Quantitative Assays

The diversity of viral subtypes of HIV-1 necessitates the use of molecular assays capable of measuring viral RNA concentrations independently of viral sequence or subtype.[108] First-generation viral RNA quantification assays were designed for optimal performance with subtype B, which predominates in Europe and North America. However, the entry of nonsubtype B virus into these areas and the discovery of intersubtype recombinant viruses complicate nucleic acid–based testing.[109] Currently, there are four commercial assays available for quantifying HIV-1 subtype B RNA in ethylenediaminetetraacetic acid (EDTA)–anticoagulated plasma. The current versions of these assays are capable of quantifying a broader range of HIV-1 subtypes than earlier tests.[108] Because these assays are not equivalent, patients should be followed clinically with repeated viral RNA testing using the same manufacturer's assay. Some of the benefits and limitations of viral RNA testing can be better appreciated if the clinician is somewhat familiar with the viral RNA assays that are currently available for clinical use.

The nucleic acid sequence–based amplification (NASBA) (NucliSens QT HIV-1 RNA assay, Organon Teknika, Advanced BioScience Laboratories) assay involves first obtaining nucleic acid through isolation by lysis and binding of the viral RNA to silica dioxide microparticles. The next step involves isothermal amplification (so-called target amplification) through use of the reverse transcriptase RNase H and T7 RNA polymerase. Three internal calibrators are added to the specimen; these are adsorbed along with the specimen before lysis. Amplification covers approximately 1,200 base-pairs in the *gag* and *pol* genes. Detection is by means of chemiluminescence. The sensitivity of the assay is approximately 100 RNA copies / ml; the quantitation limit is 500 copies / ml; and the linear dynamic range is up to 10,000,000 RNA copies / ml.

The bDNA assay (Versant HIV-1 RNA 3.0 assay, Bayer Corp.) is a nonisotopic sandwich nucleic acid hybridization assay that uses a series of target probes of the *pol* gene to hybridize the viral RNA target onto a series of capture probes on the surface of the microwell plate. The bDNA amplifier molecules are hybridized onto the captured viral RNA target. Multiple alkaline phosphatase label probes are attached to the bDNA amplifier molecules to amplify the signal after adding a specific substrate (so-called signal amplification); detection is by chemiluminescence. An external standard curve is used to calculate the HIV-1 RNA copy number. The detection and quantitation levels are approximately 50 to 100 RNA copies / ml, with a linear dynamic range up to 750,000 RNA copies / ml.[110]

The RT-PCR amplification assay (Roche Amplicor HIV-1 Monitor Test, Roche Diagnostic Systems) is the only FDA-licensed assay for assessing HIV-1 disease prognosis. The assay uses the recombinant *Thermus thermophilus* (rTth) enzyme, which serves to catalyze both reverse transcription and DNA amplification of a 142 base *gag* gene sequence in a single reaction tube. An internal quantitative standard is used to adjust for recovery and to calculate the final HIV-1 RNA copy number. The assay has an analytic sensitivity of 200 RNA copies / ml and a quantitation limit of 400 RNA copies / ml, with a dynamic range up to 750,000 RNA copies / ml. A centrifugation step (Amplicor HIV-1 Monitor Ultrasensitive specimen preparation protocol, Roche Molecular Systems) can be used to concentrate virus and thus increase the sensitivity of the assay to 50 copies / ml, and the quantitation limit is 200 copies / ml.[111]

Detection of Antiretroviral Susceptibility Genotype and Phenotype

Incomplete inhibition of HIV-1 replication in vivo may arise because of patient noncompliance with therapy, inadequate potency of the chosen regimen, poor drug absorption, drug-drug interactions, inadequate activation of the antiretroviral drug, or infection with drug-resistant virus variants[112] [*see 7:XXXIII HIV and AIDS*]. This incomplete inhibition may result in the emergence of drug-resistant HIV-1 variants and thus is an important cause of therapy failure.[113] An assessment of drug resistance may be helpful in selecting subsequent antiretroviral therapy, but this has not been rigorously proved. Nevertheless, assays for drug resistance testing in HIV-1 infection are now available, and clinical studies suggest that viral drug resistance correlates with poor virologic response to new therapy.[114] It is recommended that interpretation be made by specialists who are expert in this area, given the complexity of results and assay limitations.[115]

Antiretroviral drug susceptibility is determined either phenotypically by assessing for the susceptibility of the virus (or viral recombinant) ex vivo or genotypically by assessing for mutations that confers resistance. There are two phenotypic approaches: the first approach tests the sensitivity of the proviral population present in the peripheral blood mononuclear cells of the patient; the second approach generates recombinant fragments of the virus that contain *pol* genes obtained from the viral RNA or from the proviral DNA.[116,117] There are several genotypic approaches: DNA sequencing of the entire viral population or clones; selective PCR assay; determination of point mutations; differential probe hybridization; EIA modification of the oligoligase detection reaction assay; and a commercially available HIV-1 reverse transcription line probe assay.[118] Genotypic chang-

es may not always correlate with changes in drug susceptibility of the clinical isolate.[119] Before these techniques can be applied effectively in the clinic, much still needs to be learned about the correlation between genotype and phenotype and the effect of these phenotypic and genotypic approaches on clinical outcome in patients who receive antiretroviral therapy.[120] Nevertheless, despite the recommendation by some experts for the wider use of susceptibility testing in clinical practice,[115,121,122] most decisions to start or change therapy are based on the viral RNA level, $CD4^+$ T cell count, and previous antiretroviral drug history. In addition, the patient should be carefully educated about the importance of compliance with and adherence to the prescribed therapy regimen.[112,123]

Assessment of disease progression is discussed elsewhere [*see 7:XXXIII HIV and AIDS*].

Monitoring of Infection

In general, nadir responses in plasma viral RNA occur in most patients 4 to 12 weeks after initiation of potent antiretroviral therapy. With some potent antiretroviral regimens, plasma viral RNA levels will decrease to less than 500 RNA copies / ml by 16 weeks in 60% to 80% of patients and to less than 50 viral RNA copies / ml by 24 weeks in approximately 70%. Previous antiretroviral therapy is a factor affecting the decrease in viral RNA levels.[94,124] For patients who show a stable response, continued monitoring at intervals of 3 to 4 months seems warranted; more frequent monitoring is warranted if the plasma viral RNA level approaches a value that would cause a change in therapy. It is advisable to plot the plasma viral RNA values on a $\log_{10}$ scale; plasma viral RNA levels should be reassessed only when values exceed the upper 95% confidence interval for the expected variation in plasma viral RNA level (that is, 0.5 $\log_{10}$ RNA copies / ml above the mean nadir response value) and only if such a change in viral RNA levels would necessitate a change in therapy.[125]

By current consensus guidelines, the goal of therapy is to reduce the plasma viral RNA level to an undetectable amount, which is now fewer than 50 RNA copies / ml.[10] Other targets of therapy are a durable reduction in the plasma viral RNA level by at least threefold (0.5 $\log_{10}$) or more from pretherapy levels[125,126] and a reduction in plasma viral RNA level to below 1,000 copies / ml.[9] Although there are only limited data from controlled clinical trials that carefully delineate the decrease in risk of disease progression associated with decreases in the plasma viral RNA levels to less than 1,000 copies / ml,[9] a preliminary analysis of 1,083 patients from the AIDS Clinical Trials Group 320 Study Team showed that only seven of 126 clinical events (5.6%) were associated with a preceding plasma HIV-1 level lower than 500 copies / ml; for the remaining 119 clinical events, the plasma viral RNA level was higher than 500 copies / ml.[94] Nevertheless, many investigators believe that the target for therapy should be at least one $\log_{10}$ value lower than this (i.e., < 50 RNA copies / ml).[10] Occasional blips in detectable levels of viral RNA do not appear to adversely affect outcome. The "as low as you can go" concept in HIV treatment is an appealing virologic concept that has been repeated so often as to have been accepted as the standard. Whether this concept pertains to a chronic viral infection such as HIV-1 requires careful validation by controlled clinical trials. This is particularly important with regard to the treatment of asymptomatic patients with HIV-1 who have a $CD4^+$ T cell count greater than 500 cells / µl; in such patients, the possibility of long-term toxicity associated with the use of potent drug combinations that contain protease inhibitors needs to be weighed against the relatively small clinical benefit derived from reducing plasma viral RNA levels to the lowest level possible.

In addition to reducing plasma viral RNA levels, therapy should also lead to a stabilization of the $CD4^+$ T cell count at or above the expected level of biologic variation; the absolute number of $CD4^+$ T cells should be increased by more than 30%, or the proportion of lymphocytes that are $CD4^+$ T cells should increase by more than 3%. Interpretation of any given reduction in the level of plasma HIV-1 RNA depends on the treatment response of the $CD4^+$ T cell count. To determine clinical response to therapy, changes in both the viral RNA level and the $CD4^+$ T cell count must be considered because of the interaction between them.[127] In an analysis of several clinical studies primarily of nucleoside therapy, patients who did not experience a reduction in the plasma HIV-1 RNA level but who did experience a reduction in the $CD4^+$ T cell count had a 30% greater risk of clinical progression over 2 years than patients who had an increase in $CD4^+$ T cells above pretherapy levels.[9]

An important goal of antiretroviral therapy is to provide an improved sense of patient well-being and to minimize adverse effects. Therefore, in establishing as a target the lowest plasma viral RNA level, consideration must be given to previous antiretroviral drug exposure; in addition, therapy compliance, tolerability of the regimen, and long-term toxicity of the therapy regimen must be weighed.

Use of plasma viral RNA to define therapeutic failure A precise definition of therapeutic failure cannot be based on viral RNA level alone. Such a definition should embrace as a minimum the clinical status of the patient, the tolerability of the antiretroviral regimen, the $CD4^+$ T cell count, and the plasma viral RNA level. The failure of the plasma viral RNA level to decrease by at least 30-fold (1.5 $\log_{10}$) or more from baseline after 4 to 8 weeks of therapy is generally considered a suboptimal virologic response.[9] In addition, many clinicians consider the inability to lower plasma viral RNA to undetectable levels (< 50 RNA copies / ml) after 12 to 24 weeks of therapy to be evidence of therapeutic failure.[10] Lowering viral RNA to undetectable levels is considered a benchmark of success because of the possibility of the development of drug resistance, which can arise when viral replication occurs in the presence of selective pressure by antiretroviral drug regimens.[128] However, many patients fail to achieve undetectable viral RNA levels or experience a rebound in viral RNA after starting antiretroviral therapy.[94,124] Although the pretherapy plasma viral RNA level is itself an independent risk factor for disease progression, the relative benefit of a decline in plasma viral RNA is the same regardless of the level of viral RNA before therapy. For example, a 1 $\log_{10}$ decline in the viral RNA level from 10^5 to 10^4 RNA copies / ml represents the same relative 72% reduction in risk as a change from 10^4 to 10^3 RNA copies / ml.[9] However, these two pretherapy plasma HIV RNA levels confer different relative hazards of disease progression, after controlling for baseline $CD4^+$ T cell count and therapy assignment.[9]

Viral resistance It has been somewhat arbitrarily established that any sustainable 0.5 $\log_{10}$ (threefold) increase in plasma viral RNA level above the therapy-induced plasma viral RNA nadir that is not attributable to intercurrent infection, vaccination, incomplete adherence to the antiretroviral therapy reg-

imen, decreased absorption of antiretroviral drugs, altered drug metabolism, drug-drug interactions, or testing methodology likely represents viral failure caused by the emergence of drug-resistant HIV variants.[10,115] Although genotypic and phenotypic changes associated with drug resistance in vitro are not always synonymous with clinical drug failure in practice, retrospective and prospective clinical data on the predictive value of these tests have supported their adjunctive use in the selection of the antiretroviral regimen to employ next after virologic failure.[114,115] The correct interpretation of viral susceptibility is complicated by the complexity of the mutational patterns, the potential for undetected subpopulations of mutant virus, and residual effects of previous antiretroviral therapy. Thus, the practicing clinician should be cautious about using viral genotypic or phenotypic susceptibility analysis until definitive, prospective clinical trial data are available. Until such data are available, the clinical decision as to when to initiate or change therapy for a patient with HIV should be individualized and should depend primarily on the viral RNA level and $CD4^+$ T cell count, not antiretroviral resistance test results alone.[112]

Discordant viral RNA and $CD4^+$ T cell responses The plasma viral RNA level may increase while the $CD4^+$ T cell count remains stable or continues to rise in response to therapy.[129] This discordance occurs in approximately 14% or more of patients who receive antiretroviral therapy.[9] For many patients, the discordant response is associated with a decrease in disease progression; this raises into question the wisdom of changing antiretroviral therapy solely on the basis of plasma viral RNA response.[130] In the absence of clinical trial data, if the plasma viral RNA level increases in a sustained manner during therapy to either less than 0.5 $\log_{10}$ above the plasma viral RNA therapy nadir or 5,000 (3.7 $\log_{10}$) copies / ml, whichever is less, then it might be prudent to observe carefully for a further deterioration in $CD4^+$ T cell count or clinical progression before considering a change of therapy. A sustained increase in the plasma viral RNA level that exceeds these criteria warrants a reassessment of the antiretroviral regimen for adherence and possible therapy failure because of antiretroviral drug resistance. A more aggressive approach would be to consider a switch in therapy warranted if plasma viral RNA remains at detectable levels.[10] Neither approach, however, has received adequate validation in a controlled clinical trial.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Kalyanaraman VS, Sarngadharan MG, Robert-Guroff M, et al: A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. Science 218:571, 1982

2. Poiesz BJ, Ruscetti FW, Reitz MS, et al: Isolation of a new type C retrovirus (HTLV) in primary uncultured cells of a patient with Sézary T-cell leukaemia. Nature 294:268, 1981

3. Linial M, Weiss RA: Other human and primate retroviruses. Fields Virology, Vol. 2. Knipe DM, Howley MD, Eds. Lippincott Williams & Wilkins, Philadelphia, 2001, p 2123

4. Goff SP: Retroviridae: the retroviruses and their replication. Fields Virology, Vol. 2. Knipe DM, Howley MD, Eds. Lippincott Williams & Wilkins, Philadelphia, 2001, p 1871

5. Weiss RA, Griffiths D, Takeuchi Y, et al: Retroviruses: ancient and modern. Arch Virol 15(suppl):171, 1999

6. Arthur LO, Bess JW Jr, Sowder RC 2nd, et al: Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines. Science 258:1935, 1992

7. Luciw PA: Human immunodeficiency viruses and their replication. Fields Virology, Vol. 2. Fields BN, Knipe DM, Howley PM, Eds. Lippincott-Raven, Philadelphia, 1996, p 1881

8. Sherry B, Zybarth G, Alfano M, et al: Role of cyclophilin A in the uptake of HIV-1 by macrophages and T lymphocytes. Proc Natl Acad Sci USA 95:1758, 1998

9. Marschner IC, Collier AC, Coombs RW, et al: Use of changes in plasma levels of human immunodeficiency virus type 1 RNA to assess the clinical benefit of antiretroviral therapy. J Infect Dis 177:40, 1998

10. Carpenter CC, Cooper DA, Fischl MA, et al: Antiretroviral therapy in adults: updated recommendations of the International AIDS Society–USA Panel. JAMA 283:381, 2000

11. Wyatt R, Kwong PD, Desjardins E, et al: The antigenic structure of the HIV gp120 envelope glycoprotein. Nature 393:705, 1998

12. Chan DC, Kim PS: HIV entry and its inhibition. Cell 93:681, 1998

13. Streicher HZ, Reitz MSJ, Gallo RC: Human immunodeficinecy viruses. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 5th ed, Vol. 2. Mandel GL, Bennett JE, Donlin R, Eds. Churchill Livingstone, New York, 2000, p 1874

14. Emerman M, Malim MH: HIV-1 regulatory / accessory genes: keys to unraveling viral and host cell biology. Science 280:1880, 1998

15. Feng Y, Broder CC, Kennedy PE, et al: HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 272:872, 1996

16. Cleghorn FR, Blattner WA: Human T-cell lymphotropic virus and HTLV infection. Sexually Transmitted Diseases. Holmes KK, Sparling PF, Mardh PA, et al, Eds. McGraw-Hill, New York, 1999, p 259

17. Quinn TC, Wawer MJ, Sewankambo N, et al: Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 342:921, 2000

18. Coombs RW, Speck CE, Hughes JP, et al: Association between culturable human immunodeficiency virus type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: evidence for compartmentalization of HIV-1 between semen and blood. J Infect Dis 177:320, 1998

19. Shepard RN, Schock J, Robertson K, et al: Quantitation of human immunodeficiency virus type 1 RNA in different biological compartments. J Clin Microbiol 38:1414, 2000

20. Pillay K, Coutsoudis A, York D, et al: Cell-free virus in breast milk of HIV-1-seropositive women. J Acquir Immune Defic Syndr 24:330, 2000

21. Berger EA, Murphy PM, Farber JM: Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17:657, 1999

22. Premack BA, Schall TJ: Chemokine receptors: gateways to inflammation and infection. Nat Med 2:1174, 1996

23. Bleul CC, Farzan M, Choe H, et al: The lymphocyte chemoattractant SDF-1 is a ligand for LESTR / fusin and blocks HIV-1 entry. Nature 382:829, 1996

24. Donzella GA, Schols D, Lin SW, et al: AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor. Nat Med 4:72, 1998

25. Herbein G, Mahlknecht U, Batliwalla F, et al: Apoptosis of $CD8^+$ T cells is mediated by macrophages through interaction of HIV gp120 with chemokine receptor CXCR4. Nature 395:189, 1998

26. Pleskoff O, Treboute C, Brelot A, et al: Identification of a chemokine receptor encoded by human cytomegalovirus as a cofactor for HIV-1 entry. Science 276:1874, 1997

27. Connor RI, Sheridan KE, Ceradini D, et al: Change in coreceptor use correlates with disease progression in HIV-1–infected individuals. J Exp Med 185:621, 1997

28. Huang Y, Paxton WA, Wolinsky SM, et al: The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 2:1240, 1996

29. Michael NL, Louie LG, Rohrbaugh AL, et al: The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and disease progression. Nat Med 3:1160, 1997

30. Michael NL, Chang G, Louie LG, et al: The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. Nat Med 3:338, 1997

31. Geijtenbeek TB, Kwon DS, Torensma R, et al: DC-SIGN, a dendritic cell-specific HIV-1–binding protein that enhances trans-infection of T cells. Cell 100:587, 2000

32. Geijtenbeek TB, Torensma R, van Vliet SJ, et al: Identification of DC-SIGN, a novel dendritic cell–specific ICAM-3 receptor that supports primary immune responses. Cell 100:575, 2000

33. Albright AV, Strizki J, Harouse JM, et al: HIV-1 infection of cultured human adult oligodendrocytes. Virology 217:211, 1996

34. Delezay O, Koch N, Yahi N, et al: Co-expression of CXCR4 / fusin and galactosylceramide in the human intestinal epithelial cell line HT-29. AIDS 11:1311, 1997

35. Ahmed R, Morrison LA, Knipe DM: Persistence of viruses. Fields Virology, Vol. 2. Fields BN, Knipe DM, Howley PM, Eds. Lippincott-Raven, Philadelphia, 1996, p 219

36. Koenig S, Fauci AS: Immunology of human immunodeficiency virus. Sexually Transmitted Diseases. Holmes KK, Sparling PF, Mardh PA, et al, Eds. McGraw-Hill, New York, 1999, p 231

37. Ho DD, Neumann AU, Perelson AS, et al: Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373:123, 1995

38. Wei X, Ghosh SK, Taylor ME, et al: Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117, 1995

39. Mortara L, Letourneur F, Gras-Masse H, et al: Selection of virus variants and emergence of virus escape mutants after immunization with an epitope vaccine. J Virol 72:1403, 1998

40. Liu HF, Goubau P, Van Brussel M, et al: The three human T-lymphotropic virus type I subtypes arose from three geographically distinct simian reservoirs. J Gen Virol 77:359, 1996

41. Gao F, Bailes E, Robertson DL, et al: Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes.* Nature 397:436, 1999

42. Korber B, Muldoon M, Theiler J, et al: Timing the ancestor of the HIV-1 pandemic strains. Science 288:1789, 2000

43. Zehender G, De Maddalena C, Osio M, et al: High prevalence of human T cell lymphotropic virus type II infection in patients affected by human immunodeficiency virus type 1–associated predominantly sensory polyneuropathy. J Infect Dis 172:1595, 1995

44. Hall WW, Ishak R, Zhu SW, et al: Human T lymphotropic virus type II (HTLV-II): epidemiology, molecular properties, and clinical features of infection. J Acquir Immune Defic Syndr Hum Retrovirol 13:S204, 1996

45. Franchini G: Molecular mechanisms of human T-cell leukemia / lymphotropic virus type I infection. Blood 86:3619, 1995

46. Kaplan JE, Khabbaz RF: The epidemiology of human T-lymphotropic virus type I and II. Rev Med Virol 3:137, 1993

47. Dezzutti CS, Lal RB: Human T-cell lymphotropic virus types I and II. Manual of Clinical Microbiology. Murray PR, Baron KJ, Pfaller MA, et al, Eds. American Society for Microbiology, Washington, DC, 1999, p 871

48. Vidal AU, Gessain A, Yoshida M, et al: Phylogenetic classification of human T cell leukaemia / lymphoma virus type I genotypes in five major molecular and geographical subtypes. J Gen Virol 75:3655, 1994

49. Eiraku N, Novoa P, da Costa Ferreira M, et al: Identification and characterization of a new and distinct molecular subtype of human T-cell lymphotropic virus type 2. J Virol 70:1481, 1996

50. Lal RB, Rudolph DL, Coligan JE, et al: Failure to detect evidence of human T-lymphotropic virus (HTLV) type I and type II in blood donors with isolated gag antibodies to HTLV-I / II. Blood 80:544, 1992

51. Schreiber GB, Busch MP, Kleinman SH, et al: The risk of transfusion-transmitted viral infections. The Retrovirus Epidemiology Donor Study. N Engl J Med 334:1685, 1996

52. Green PL, Chen SY: Human T-cell leukemia virus types 1 and 2. Fields Virology, Vol. 2. Knipe DM, Howley PM, Eds. Lippincott Williams & Wilkins, Philadelphia, 2001, p 1941

53. Schreiber GB, Murphy EL, Horton JA, et al: Risk factors for human T-cell lymphotropic virus types I and II (HTLV-I and -II) in blood donors. NHLBI Retrovirus Epidemiology Donor Study. J Acquir Immune Defic Syndr Hum Retrovirol 14:263, 1997

54. Shimoyama M: Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma: a report from the Lymphoma Study Group (1984–87). Br J Haematol 79:428, 1991

55. Ifthikharuddin JJ, Rosenblatt JD: Human T-cell lymphotropic virus types I and II. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 5th ed, Vol. 2. Mandel GL, Bennett JE, Donlin R, Eds. Churchill Livingstone, New York, 2000, p 1862

56. Roman GC, Roman LN: Tropical spastic paraparesis: a clinical study of 50 patients from Tumaco (Colombia) and review of the worldwide features of the syndrome. J Neurol Sci 87:121, 1988

57. Blattner WA: Human lymphotropic viruses: HTLV-I and HTLV-II. Clinical Virology. Richman DD, Whitley RJ, Hayden FG, Eds. Churchill Livingstone, New York, 1998, p 683

58. Robert-Guroff M, Weiss SH, Giron JA, et al: Prevalence of antibodies to HTLV-I, -II, and -III in intravenous drug abusers from an AIDS endemic region. JAMA 255:3133, 1986

59. Coombs RW: Clinical laboratory diagnosis of HIV-1 and use of viral RNA to monitor infection. Holmes KK, Sparling PF, Mardh PA, et al, Eds. McGraw-Hill, New York, 1999, p 973

60. Interpretive criteria used to report western blot results for HIV-1–antibody testing—United States. MMWR Morb Mortal Wkly Rep 40:692, 1991

61. Zaaijer HL, van Exel-Oehlers P, Kraaijeveld T, et al: Early detection of antibodies to HIV-1 by third-generation assays. Lancet 340:770, 1992

62. Ascher DP, Roberts C: Determination of the etiology of seroreversals in HIV testing by antibody fingerprinting. J Acquir Immune Defic Syndr 6:241, 1993

63. Celum CL, Coombs RW, Jones M, et al: Risk factors for repeatedly reactive HIV-1 EIA and indeterminate western blots: a population-based case-control study. Arch Intern Med 154:1129, 1994

64. Burke DS, Brundage JF, Redfield RR, et al: Measurement of the false positive rate in a screening program for human immunodeficiency virus infections. N Engl J Med 319:961, 1988

65. Celum CL, Coombs RW: Indeterminate HIV-1 western blots: implications and considerations for widespread HIV testing. J Gen Intern Med 7:640, 1992

66. Celum CL, Coombs RW, Lafferty W, et al: Indeterminate human immunodeficiency virus type 1 Western blots: seroconversion risk, specificity of supplemental tests, and an algorithm for evaluation. J Infect Dis 164:656, 1991

67. Montagnier L, Brenner C, Chamaret S, et al: Human immunodeficiency virus infection and AIDS in a person with negative serology. J Infect Dis 175:955, 1997

68. Zaaijer HL, Bloemer MH, Lelie PN: Temporary seronegativity in a human immunodeficiency virus type 1–infected man. J Med Virol 51:80, 1997

69. Brown AE, Jackson B, Fuller SA, et al: Viral RNA in the resolution of human immunodeficiency virus type 1 diagnostic serology. Transfusion 37:926, 1997

70. Kassler WJ, Dillon BA, Haley C, et al: On-site, rapid HIV testing with same-day results and counseling. AIDS 11:1045, 1997

71. Irwin K, Olivo N, Schable CA, et al: Performance characteristics of a rapid HIV antibody assay in a hospital with a high prevalence of HIV infection. CDC-Bronx-Lebanon HIV Serosurvey Team. Ann Intern Med 125:471, 1996

72. Beltrami EM, Williams IT, Shapiro CN, et al: Risk and management of blood-borne infections in health care workers. Clin Microbiol Rev 13:385, 2000

73. Public Health Service guidelines for the management of health-care worker exposures to HIV and recommendations for postexposure prophylaxis. MMWR Morb Mortal Wkly Rep 47:1, 1998

74. Simon F, Mauclere P, Roques P, et al: Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med 4:1032, 1998

75. Sullivan PS, Do AN, Ellenberger D, et al: Human immunodeficiency virus (HIV) subtype surveillance of African-born persons at risk for group O and group N HIV infections in the United States. J Infect Dis 181:463, 2000

76. Michael NL, Herman SA, Kwok S, et al: Development of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 and performance of an improved AMPLICOR HIV-1 MONITOR test with isolates of diverse subtypes. J Clin Microbiol 37:2557, 1999

77. Coste J, Montes B, Reynes J, et al: Comparative evaluation of three assays for the quantitation of human immunodeficiency virus type 1 RNA in plasma. J Med Virol 50:293, 1996

78. Onorato IM, O'Brien TR, Schable CA, et al: Sentinel surveillance for HIV-2 infection in high-risk US populations. Am J Public Health 83:515, 1993

79. Guyader M, Emerman M, Sonigo P, et al: Genome organization and transactivation of the human immunodeficiency virus type 2. Nature 326:662, 1987

80. Busch MP, Petersen L, Schable C, et al: Monitoring blood donors for HIV-2 infection by testing anti-HIV-1 reactive sera. Transfusion 30:184, 1990

81. Coombs RW, Collier AC, Allain JP, et al: Plasma viremia in human immunodeficiency virus infection. N Engl J Med 321:1626, 1989

82. Ho DD, Moudgil T, Alam M: Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. N Engl J Med 321:1621, 1989

83. Finzi D, Hermankova M, Pierson T, et al: Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278:1295, 1997

84. D'Aquila RT, Johnson VA, Welles SL, et al: Zidovudine resistance and HIV-1 disease progression during antiretroviral therapy. AIDS Clinical Trials Group Protocol 116B / 117 Team and the Virology Committee Resistance Working Group. Ann Intern Med 122:401, 1995

85. Picchio GR, Valdez H, Sabbe R, et al: Altered viral fitness of HIV-1 following failure of protease inhibitor-based therapy. J Acquir Immune Defic Syndr 25:289, 2000

86. Schacker T, Collier AC, Hughes J, et al: Clinical and epidemiologic features of primary HIV infection. Ann Intern Med 125:257, 1996

87. Schupbach J, Tomasik Z, Nadal D, et al: Use of HIV-1 p24 as a sensitive, precise and inexpensive marker for infection, disease progression and treatment failure. Int J Antimicrob Agents 16:441, 2000

88. Ledergerber B, Flepp M, Boni J, et al: Human immunodeficiency virus type 1 p24 concentration measured by boosted ELISA of heat-denatured plasma correlates with decline in CD4 cells, progression to AIDS, and survival: comparison with viral RNA measurement. J Infect Dis 181:1280, 2000

89. Yerly S, Pedrocchi M, Perrin L: The use of polymerase chain reaction in plasma pools for the concomitant detection of hepatitis C virus and HIV type 1 RNA. Transfusion 38:908, 1998

90. Sun R, Jayakar H, Yeh S, et al: Experience with PCR screening. Dev Biol Stand 102:81, 2000

91. Phair JP, Margolick JB, Jacobson LP, et al: Detection of infection with human immunodeficiency virus type 1 before seroconversion: correlation with clinical symptoms and outcome. J Infect Dis 175:959, 1997

92. Shearer WT, Quinn TC, LaRussa P, et al: Viral load and disease progression in infants infected with human immunodeficiency virus type 1. Women and Infants Transmission Study Group. N Engl J Med 336:1337, 1997

93. Coombs RW, Welles SL, Hooper C, et al: Association of plasma human immunodeficiency virus type 1 RNA level with risk of clinical progression in patients with advanced infection. AIDS Clinical Trials Group (ACTG) 116B / 117 Study Team. J Infect Dis 174:704, 1996

94. Hammer SM, Squires KE, Hughes MD, et al: A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. N Engl J Med 337:725, 1997

95. Katzenstein DA, Hammer SM, Hughes MD, et al: The relation of virologic and immunologic markers to clinical outcomes after nucleoside therapy in HIV-infected adults with 200 to 500 CD4 cells per cubic millimeter. AIDS Clinical Trials Group Study 175 Virology Study Team. N Engl J Med 335:1091, 1996

96. O'Brien WA, Hartigan PM, Martin D, et al: Changes in plasma HIV-1 RNA and $CD4^+$ lymphocyte counts and the risk of progression to AIDS. Veterans Affairs Cooperative Study Group on AIDS. N Engl J Med 334:426, 1996

97. Welles SL, Jackson JB, Yen-Lieberman B, et al: Prognostic value of plasma human immunodeficiency virus type 1 (HIV-1) RNA levels in patients with advanced HIV-1 disease and with little or no prior zidovudine therapy. J Infect Dis 174:696, 1996

98. Bremer JW, Lew JF, Cooper E, et al: Diagnosis of infection with human immunodeficiency virus type 1 by a DNA polymerase chain reaction assay among infants enrolled in the Women and Infants' Transmission Study. J Pediatr 129:198, 1996

99. Owens DK, Holodniy M, McDonald TW, et al: A meta-analytic evaluation of the polymerase chain reaction for the diagnosis of HIV infection in infants. JAMA 275:1342, 1996

100. Owens DK, Holodniy M, Garber AM, et al: Polymerase chain reaction for the diagnosis of HIV infection in adults: a meta-analysis with recommendations for clinical practice and study design. Ann Intern Med 124:803, 1996

101. Frenkel LM, Mullins JI, Learn GH, et al: Genetic evaluation of suspected cases of transient HIV-1 infection of infants. Science 280:1073, 1998

102. Schwartz DH, Laeyendecker OB, Arango-Jaramillo S, et al: Extensive evaluation of a seronegative participant in an HIV-1 vaccine trial as a result of false-positive PCR. Lancet 350:256, 1997

103. Brambilla D, Leung S, Lew J, et al: Absolute copy number and relative change in determinations of human immunodeficiency virus type 1 RNA in plasma: effect of an external standard on kit comparisons. J Clin Microbiol 36:311, 1998

104. Steketee RW, Abrams EJ, Thea DM, et al: Early detection of perinatal human immunodeficiency virus (HIV) type 1 infection using HIV RNA amplification and detection. New York City Perinatal HIV Transmission Collaborative Study. J Infect Dis 175:707, 1997

105. Weber B, Fall EH, Berger A, et al: Reduction of diagnostic window by new fourth-generation human immunodeficiency virus screening assays. J Clin Microbiol 36:2235, 1998

106. Rich JD, Merriman NA, Mylonakis E, et al: Misdiagnosis of HIV infection by HIV-1 plasma viral load testing: a case series. Ann Intern Med 130:37, 1999

107. Schacker TW, Hughes JP, Shea T, et al: Biological and virologic characteristics of primary HIV infection. Ann Intern Med 128:613, 1998

108. Jagodzinski LL, Wiggins DL, McManis JL, et al: Use of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 to assess the performance of viral RNA quantitation tests. J Clin Microbiol 38:1247, 2000

109. Robertson DL, Sharp PM, McCutchan FE, et al: Recombination in HIV-1. Nature 374:124, 1995

110. Collins ML, Irvine B, Tyner D, et al: A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules / ml. Nucleic Acids Res 25:2979, 1997

111. Mulder J, Resnick R, Saget B, et al: A rapid and simple method for extracting human immunodeficiency virus type 1 RNA from plasma: enhanced sensitivity. J Clin Microbiol 35:1278, 1997

112. Kartsonis NA, D'Aquila RT: Clinical monitoring of HIV-1 infection in the ERA of antiretroviral resistance testing. Infect Dis Clin North Am 14:879, 2000

113. Brown AJ, Richman DD: HIV-1: gambling on the evolution of drug resistance? Nat Med 3:268, 1997

114. Durant J, Clevenbergh P, Halfon P, et al: Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. Lancet 353:2195, 1999

115. Hirsch MS, Brun-Vezinet F, D'Aquila RT, et al: Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an International AIDS Society—USA Panel. JAMA 283:2417, 2000

116. Nijhuis M, Schuurman R, Boucher CA: Homologous recombination for rapid phenotyping of HIV. Curr Opin Infect Dis 10:474, 1997

117. Hertogs K, de Bethune MP, Miller V, et al: A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. Antimicrob Agents Chemother 42:269, 1998

118. Schuurman R: State of the art of genotypic HIV-1 drug resistance. Curr Opin Infect Dis 10:480, 1997

119. Larder BA, Kemp SD, Harrigan PR: Potential mechanism for sustained antiretroviral efficacy of AZT-3TC combination therapy. Science 269:696, 1995

120. Brun-Vezinet F, Boucher C, Loveday C, et al: HIV-1 viral load, phenotype, and resistance in a subset of drug-naive participants from the Delta trial. The National Virology Groups; Delta Virology Working Group and Coordinating Committee. Lancet 350:983, 1997

121. Hirsch MS, Conway B, D'Aquila RT, et al: Antiretroviral drug resistance testing in adults with HIV infection: implications for clinical management. International AIDS Society—USA Panel. JAMA 279:1984, 1998

122. Rodriguez-Rosado R, Briones C, Soriano V: Introduction of HIV drug-resistance testing in clinical practice. AIDS 13:1007, 1999

123. Andrews L, Friedland G: Progress in HIV therapeutics and the challenges of adherence to antiretroviral therapy. Infect Dis Clin North Am 14:901, 2000

124. Gulick RM, Mellors JW, Havlir D, et al: Treatment with indinavir, zidovudine, and lamivudine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. N Engl J Med 337:734, 1997

125. Brambilla D, Reichelderfer PS, Bremer JW, et al: The contribution of assay variation and biological variation to the total variability of plasma HIV-1 RNA measurements. The Women Infant Transmission Study Clinics; Virology Quality Assurance Program. AIDS 13:2269, 1999

126. Paxton WB, Coombs RW, McElrath MJ, et al: Longitudinal analysis of quantitative virologic measures in human immunodeficiency virus-infected subjects with ≥ 400 CD4 lymphocytes: implications for applying measurements to individual patients. National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group. J Infect Dis 175:247, 1997

127. Grabar S, Le Moing V, Goujard C, et al: Clinical outcome of patients with HIV-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy. Ann Intern Med 133:401, 2000

128. Martinez-Picado J, DePasquale MP, Kartsonis N, et al: Antiretroviral resistance during successful therapy of HIV type 1 infection. Proc Natl Acad Sci U S A 97:10948, 2000

129. Reichelderfer PS, Coombs RW: Cartesian coordinate analysis of viral burden and $CD4^+$ T-cell count in human immunodeficiency virus type-1 infection. Antiviral Res 38:181, 1998

130. Ledergerber B, Egger M, Opravil M, et al: Clinical progression and virological failure on highly active antiretroviral therapy in HIV-1 patients: a prospective cohort study. Swiss HIV Cohort Study. Lancet 353:863, 1999

131. Fauci AS, Longo DL: The human retroviruses. Harrison's Principles of Internal Medicine, 15th ed., Vol. 1. Braunwald E, Fauci A, Kasper D, et al. McGraw-Hill, New York, 2001, p 1131

132. Moore JP, Binley J: HIV: envelope's letters boxed into shape. Nature 393:630, 1998

133. Furtado MR, Callaway DS, Phair JP, et al: Persistence of HIV-1 transcription in peripheral blood mononuclear cells in patients receiving potent antiretroviral therapy. N Engl J Med 340:1614, 1999

Acknowledgment

Figures 1 through 3 Seward Hung.

XXXIII HIV AND AIDS

Robert D. Harrington, m.d.
David H. Spach, m.d.

The HIV/AIDS epidemic has changed dramatically over the years. The gap between AIDS cases in developed countries and those in developing countries has steadily widened as developed countries have continued to reap the tremendous benefits of combination antiretroviral therapy while developing countries have struggled with minimal resources. In response, the World Health Organization (WHO), partnering with the Joint United Nations Programme on HIV/AIDS (UNAIDS), has set an ambitious goal, detailed in their 3 by 5 Initiative, of providing effective antiretroviral therapy to 3 million HIV-infected persons in the developing world by the year 2005.[1] This effort will require an enormous commitment of funds, drugs, and human resources to create the infrastructure needed to diagnose and evaluate patients, distribute medication, and monitor treatment. In developed countries, aggressive antiretroviral therapy has already led to a remarkable decline in the rate of AIDS-related deaths and the incidence of major opportunistic infections,[2,3] but this success has ushered in a new and complicated era of HIV clinical care. This chapter provides an update of key information related to the epidemiology and pathogenesis of HIV infection and to the clinical care of HIV-infected patients. HIV serologic tests, HIV quantitative assays (viral load assays), and HIV resistance tests are described in detail elsewhere [*see 7:XXXII Human Retroviral Infections*].

Origin of HIV

HIV evolved from simian immunodeficiency virus (SIV) in chimpanzees and monkeys. Surprisingly, in its natural host, SIV does not cause disease, despite replicating to high levels in infected animals. This lack of pathogenicity may be from a lack of T cell activation in these hosts. Alternatively, these species may have developed tolerance to the infection through natural selection.[4,5]

Data suggest that humans acquired HIV type 1 (HIV-1) from *Pan troglodytes troglodytes* chimpanzees infected with SIV, and that HIV-1 was introduced into the human population (as SIV) from these chimpanzees on at least three independent occasions.[6] These conclusions arose primarily from studies that used DNA analysis of available SIV and HIV-1 strains and are based on three important findings: (1) all HIV-1 strains known to infect humans are closely related to the SIV lineage found in *P. t. troglodytes* chimpanzees; (2) the HIV-1 group is a mosaic of SIV and HIV-1–related sequences, suggesting an ancestral recombination event in a chimpanzee host; and (3) the natural geographical range of *P. t. troglodytes* corresponds with areas endemic for HIV-1 groups M, N, and O. The exact timing and the mechanism for the transmission of HIV-1 remain unknown; given the fact that some peoples in Africa hunt these chimpanzees for food, transmission may have occurred during the capture, preparation, or ingestion of these chimpanzees.

HIV-2 originated from the sooty mangabey monkey (*Cercocebus atys*). Phylogenetically, strains of SIV in sooty mangabeys are closely related (and in some cases identical) to HIV-2.[7] Moreover, the sooty mangabey's natural habitat corresponds with the epicenter of the HIV-2 epidemic. It has been proposed that at least six independent transmission events of SIV from sooty mangabeys to humans have occurred; presumably, these events occurred through the hunting of this animal or through contact with this animal as a pet.

To date, the oldest confirmed isolate of HIV is from a blood sample collected from a patient in the eastern region of Central Africa in 1959.[8] Investigators have analyzed HIV-1 phylogenetic relationships through use of maximum-likelihood phylogenetic algorithms and have concluded that HIV-1 originated in humans in Africa in 1931 or earlier.[9] Through use of these methods, it has been suggested that HIV entered the United States in approximately 1967.

One investigator has hypothesized that HIV originated in Africa as a result of contaminated polio vaccines that were widely used in Africa in the late 1950s; this theory proposes that chimpanzee kidney cells were used in preparing the vaccines. However, analysis by three independent laboratories of samples of these vaccines stored for more than 40 years failed to show even a trace of HIV or SIV; in addition, these laboratories performed mitochondrial DNA analysis and found no evidence of chimpanzee tissue in the vaccines.[10] Furthermore, the phylogenetic studies mentioned above demonstrate that HIV-1 was in the human population long before the polio vaccine program was introduced and that, by the late 1950s, there was significant diversification of HIV-1. These findings make it highly improbable that HIV was introduced through the polio vaccine.

Classification of HIV and AIDS

In 1982, the Centers for Disease Control and Prevention (CDC) established a surveillance definition of AIDS; persons were classified as having AIDS if they had defects in cell-mediated immunity (without a known cause) and if they had one or more designated opportunistic infections or malignancies. In 1985, the case definition incorporated the newly licensed HIV antibody test. In 1993, the CDC revised the AIDS case definition to include HIV-infected adults and adolescents who had an absolute $CD4^+$ T cell count of less than 200 cells/μl or in whom $CD4^+$ T cells made up less than 14% of lymphocytes.[11] Because $CD4^+$ T cell counts in children are not comparable to those in adults, the 1993 revised AIDS definition does not apply to pediatric patients. In 1993, the CDC also generated a comprehensive classification system that consists of a matrix of $CD4^+$ T cell status (levels 1, 2, and 3) and clinical manifestations (categories A, B, and C). This system became effective on January 1, 1993, and is still in use [*see Table 1*].

Epidemiology

Global epidemiology

The HIV epidemic continues to enlarge worldwide. UNAIDS and the WHO report that as of December 2003, approximately 38 million persons were living with HIV worldwide, exceeding the WHO estimates of 1991 by more than 50%. Children under 15 years of age account for 2.5 to almost 3 million of these. In 2003, a staggering 5 million new infections occurred, and 3 million people died of AIDS [*see Figure 1*].[12] Infected persons' igno-

Table 1 Centers for Disease Control and Prevention Classification System for HIV Infection[11]

CD4⁺ T Cell Categories	*Clinical Categories*		
	A (asymptomatic, first-degree HIV or progressive generalized lymphadenopathy)	*B* (symptomatic, but does not satisfy conditions in categories A or C)*	*C† (AIDS-indicator conditions)*
Category 1: ≥ 500 cells/mm³	A1	B1	C1‡
Category 2: 200–499 cells/mm³	A2	B2	C2‡
Category 3: < 200 cells/mm³	A3‡	B3‡	C3‡

*B conditions include bacterial endocarditis, meningitis, or sepsis; candidiasis (oral); candidiasis (persistent vulvovaginal); cervical dysplasia (or carcinoma in situ); constitutional illness (persistent unexplained fever, diarrhea or weight loss, or disabling weakness); herpes zoster (multidermatomal); listeriosis; myelopathy; nocardiosis; oral hairy leukoplakia; pelvic inflammatory disease; peripheral neuropathy; and thrombocytopenic purpura (idiopathic).

†C conditions include candidiasis (bronchial, esophageal, pulmonary, or tracheal); cervical cancer (invasive); coccidioidomycosis (disseminated or extrapulmonary); cryptococcosis (extrapulmonary); cryptosporidiosis (intestinal, for longer than 1 mo); cytomegalovirus (other than liver, spleen, or nodes); encephalopathy (HIV); herpes simplex virus (ulcers present for longer than 1 mo or esophagitis, bronchitis, or pneumonitis); histoplasmosis (disseminated or extrapulmonary); isosporiasis (intestinal, for longer than 1 mo); Kaposi sarcoma; lymphoma (Burkitt, immunoblastic, or primary in brain); *Mycobacterium avium* complex (disseminated or extrapulmonary), *M. kansasii* (disseminated or extrapulmonary), *M. tuberculosis*, or *Mycobacterium* of other or unidentified species (disseminated or extrapulmonary); *Pneumocystis carinii* pneumonia; recurrent pneumonia (two or more episodes in 1-yr period); progressive multifocal leukoencephalopathy; salmonellosis (recurrent septicemia); toxoplasmosis (brain); wasting syndrome (> 10% weight loss plus either chronic weakness or documented fever for at least 30 days in the absence of a concurrent illness that could explain this finding).

‡AIDS case that satisfies the surveillance definition.

rance regarding their HIV status (which may result from lack of access to testing or refusal to be tested); unsafe sexual practices; and discrimination against and stigmatization of persons who are HIV infected, who are homosexual, or who are injection drug users all promote the spread of HIV.

The global distribution of HIV infection varies widely. Sub-Saharan Africa remains the hardest-hit region, with an estimated 26.6 million persons infected, including 3.2 million who were newly infected in 2003. Injection drug use has sparked the epidemic in Vietnam, Asia, the Middle East, Eastern Europe, and Latin America. Although a few countries, such as Thailand and Uganda, have experienced success in controlling the spread of HIV, the global epidemic continues to expand; any regional stabilization of prevalence simply reflects a rising death toll.

Like the global distribution, local and regional trends in the HIV epidemic also show wide variation. In sub-Saharan Africa, the prevalence of HIV varies greatly between regions within countries (urban areas are more affected that rural areas) and between countries (with rates as low as 1% in West African countries such as Senegal to almost 40% in South African countries such as Botswana and Swaziland). Uganda stands out as the single African nation that experienced a substantial contraction in the epidemic; in the capital city, Kampala, HIV prevalence declined from 30% a decade ago to 8% in 2002. African women are more likely than men to be infected with HIV, because of earlier sexual debut and the practice of younger women having sex with older men.

The HIV epidemic continues to expand in Eastern Europe and Central Asia. The Russian Federation, Ukraine, and the Baltic States are the most affected nations in this region. In 2003 alone, an estimated 230,000 people acquired HIV in Eastern Europe and Central Asia, bringing the total number living with HIV in these areas to 1.5 million. Injection drug use is a particularly common mode for HIV infection in these regions, especially in young males. Some of the Baltic States report that up to 1% of the adult population use injection drugs, and a staggering 12% of teenage males in Moscow report having injected drugs at some time. Low rates of condom use have facilitated the spread of the epidemic from injection drug users, who are predominantly male, to women. Sexual transmission between men who have sex with men (MSM) is undoubtedly occurring as well, but few HIV-infected men admit this as a risk factor, given the stigma associated with homosexuality in these cultures.

Injection drug use also stimulated the epidemic in Cambodia and Thailand and is now the predominant reason for the spread of HIV into regions of China, Vietnam, Indonesia, and other Asian and Pacific countries. Persons using unsafe sexual practices, especially commercial sex workers (CSWs), carry the epidemic forward and spread it to men who often eventually transmit HIV to their spouses. In 2003, HIV infected more than one

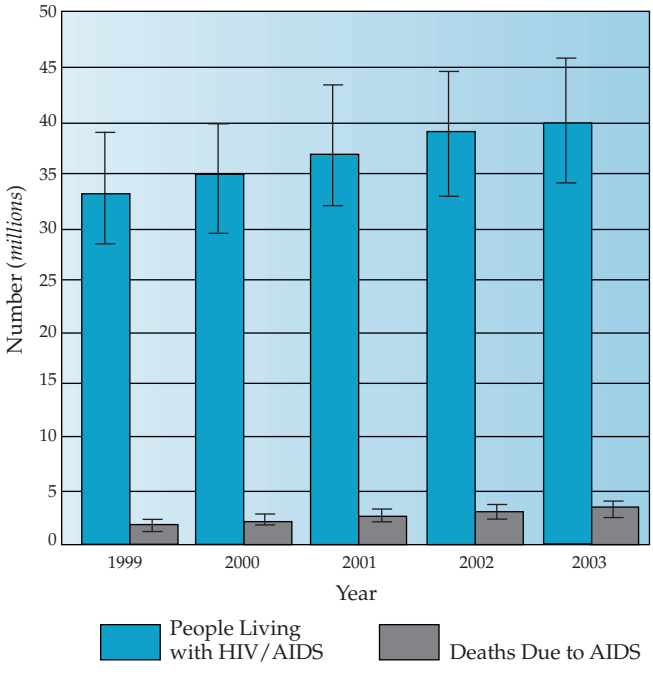


Figure 1 **Estimated number of persons living with HIV/AIDS and global AIDS deaths, 1999–2003.[12]**

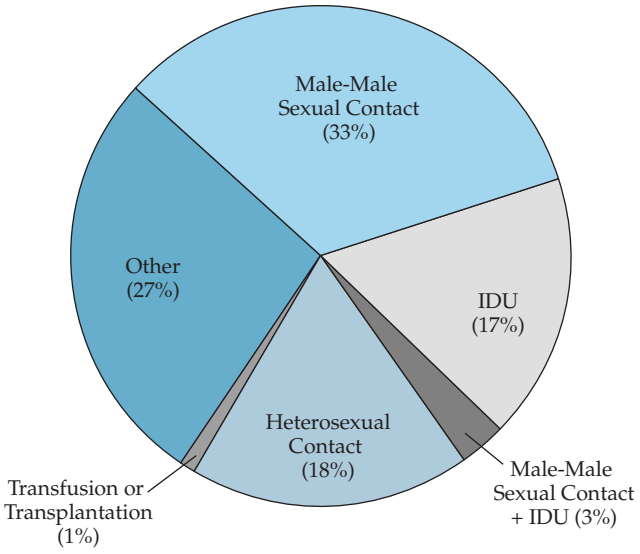


Figure 2 **The percentages of cumulative cases of AIDS of various causes in the United States through December 2002 are depicted in this illustration. These data represent both adult and adolescent AIDS patients.[2] (IDU—injection drug use)**

million persons in Asia, bringing the total number of people living with HIV in this region to 7.4 million. Although countries such as Cambodia and Thailand have posted some success in controlling the spread of HIV, primarily by vigorously promoting universal condom use by CSWs, other countries are just now experiencing the first wave of HIV infections. In these countries, the epidemic is poised to erupt, given their large, vulnerable populations, the lack of safe sex practices, and the high prevalence of contaminated injection drug equipment.

More than two million HIV-infected persons live in Latin America and the Caribbean; the prevalence in some of these countries is among the highest in the world. In Haiti, 5% to 6% of the population is HIV infected; in the Bahamas, Belize, the Dominican Republic, Guyana, and Trinidad and Tobago, the prevalence of HIV in pregnant women is 2% or greater. Different routes of infection predominate in distinct regions: in most of South America, the epidemic is driven by injection drug use; in Haiti, unsafe heterosexual sex is responsible for most infections, whereas both heterosexual and homosexual sex are important modes of transmission in Central America. In Peru and Columbia, most infections are occurring in MSM.

The Middle East and North Africa, except for southern Sudan, have been affected less than other regions of the world. Current data, however, suggest that the number of HIV-infected persons is rising because of localized outbreaks in injection drug users and emigration of HIV-infected persons from adjacent countries where the prevalence of HIV is much higher.

In developed nations other than the United States, the prevalence of HIV is rising because of declining death rates and new infections acquired within these countries or because of immigration of HIV-infected persons from other regions. Sex between men is the major mode of transmission in Germany, Greece, the Netherlands, and Australia, whereas heterosexual transmission is the rule in other Western European countries, whether the infection is imported or acquired locally. Injection drug use is responsible for 10% or less of cases in Australia and Western Europe; the exception is Portugal, where injection drug use caused roughly half of new infections in 2002.

NORTH AMERICAN EPIDEMIOLOGY

Contemporary UNAIDS and WHO estimates of the total number of HIV-infected persons (with or without AIDS) living in North America range from 750,000 to 1,200,000.[12] As of the end of 2002, the cumulative number of AIDS cases in all of the United States reported to the CDC totaled 886,575 [*see Figure 2*].[2] Of the adult patients, 81% were men and 18% were women. The majority of patients were adults or adolescents; only 1% of the patients were children younger than 13 years. Among men, 46% were white, 35% were African American, and 18% were Hispanic. Among women, 21% were white, 59% were African American, and 19% were Hispanic.

Most states have now implemented HIV (not just AIDS) surveillance programs, and many of these programs have matured to the point that the data have stabilized and trends can be monitored. The latest HIV/AIDS surveillance report summarizes cases of HIV and AIDS in the United States through 2002.[13] This report includes catalogued HIV surveillance data from 30 areas (29 states and the U.S. Virgin Islands) where confidential name-based HIV reporting has occurred since at least 1998. Within these 30 areas, several trends in the diagnoses of HIV/AIDS (which includes new diagnoses of HIV infection only, new diagnoses of HIV infection that progressed to AIDS, and concurrent diagnoses of HIV and AIDS) have occurred:

- The diagnoses of HIV/AIDS increased 3.2% from 2001 to 2002, leading to a rate of 19.1 cases per 100,000 population in 2002.
- From 1999 to 2002, increased infection rates were observed in whites, Hispanics, and Asians/Pacific Islanders, while rates in blacks remained stable. Nevertheless, most new infections continued to occur in blacks (54% in 2002).
- From 1999 to 2002, diagnoses of HIV/AIDS increased 7% in men (totaling 71% of cases in 2002), remained stable in women, and decreased in children (92 cases in 2002).
- Rates of infection increased in MSM and heterosexuals but decreased in injection drug users. In 2002, 44% of new cases of HIV/AIDS occurred in MSM and 35% resulted from heterosexual transmission.

The introduction and widespread use of aggressive antiretroviral therapy have produced a remarkable decline in the number of AIDS-related deaths.[2,3] The declining death rate has led to an increase in the prevalence of HIV-infected persons in the United States and other developed countries [*see Figure 1*]. At the end of 2002, there were an estimated 384,906 persons living with AIDS, an increase of almost 100,000 from 1998. Of these patients, approximately 141,000 were white, 162,000 black, 76,000 Hispanic, 298,000 male, 83,000 female, 282,000 between 35 and 54 years of age, and 3,900 children (younger than 13 years). The number of children with AIDS markedly declined during the 1990s, most importantly because of the rapid and widespread implementation of the use of zidovudine to prevent perinatal transmission of HIV. With the current approaches, which use more aggressive combination therapy, rates of perinatal transmission of HIV have declined even further and are currently in the range of 1%.

Approximately 40,000 new HIV infections occur annually in the United States. In recent years, nearly 50% of these new cases have involved African Americans, a group that comprises only 12% of the United States population. Approximately 75% of these infections are acquired sexually; an increasing proportion (about one third) occur through heterosexual sex. Alarming

rates of infection exist in certain demographic groups, such as African-American MSM and African-American women who partner with bisexual men. Supporting evidence for a resurgence in unsafe sexual practices includes recent outbreaks of syphilis and gonorrhea in MSM populations in San Francisco, Seattle, and elswhere.[14] Proposed explanations for this resurgence include belief in the efficacy of medication to treat HIV infection and prevent AIDS, limited exposure of younger persons to AIDS-related deaths of their peers, and the use of drugs (e.g., methamphetamine) and alcohol, which leads to impaired judgment and poor decision making about sexual behavior.

Transmission of HIV

Person-to-person transmission of HIV can occur through several routes, including sexual contact, sharing of contaminated needles used for injecting drugs, receiving a transfusion of a contaminated blood product, transmission from mother to child (during gestation, at the time of delivery, or post partum via breast-feeding), and occupational exposure. The relative risks of acquiring HIV through these routes is difficult to precisely quantitate, but available data suggest that the highest risk is through receiving contaminated blood products, followed by mother-to-child exposure (assuming the mother is not receiving treatment with antiretroviral therapy) and then by the sharing of needles when injecting drugs. Lower risks are associated with needle-stick injuries and sexual transmission.[15] The risk of sexual transmission is highly dependent on the type of encounter, with receptive anal intercourse posing the highest risk. Globally, sexual contact is by far the most common route of HIV transmission, accounting for at least 75% of HIV infections worldwide.[15]

SEXUAL CONTACT

A study from Uganda of 415 couples in which one member of the couple was free of HIV infection (HIV discordant) at the start of the study demonstrated a correlation of increased transmission with increasing HIV RNA levels.[16] Several modeling studies of MSM suggested that transmission during primary infection, which is associated with extremely high HIV RNA levels, may account for up to 50% of infections.[17-19] Longitudinal studies have clearly shown that the consistent use of condoms is highly effective in preventing transmission of HIV via heterosexual intercourse.[20] Several studies have suggested that oral-genital sex, although less risky than genital sex, poses a significant risk of HIV transmission, and a number of cases of HIV transmission have now been reported to have occurred via oral-genital sex.[21,22] Human studies have shown that the tissues in the nasopharyngeal tonsils and adenoids have abundant cells of dendritic origin; these cells can serve as an initial target for HIV and can then assist in transmitting HIV to nearby $CD4^+$ T cells.[23]

A number of factors appear to increase the risk of acquiring HIV via sexual contact; these include disruption and inflammation of the mucosal barrier of the genital tract from genital ulcer disease, urethritis, and cervicitis.[24-26] Evidence that infection with herpes simplex virus (HSV) increases the risk of HIV acquisition and transmission comes from several sources. A meta-analysis suggested that HSV enhanced the risk of HIV acquisition by 2.1-fold to 3.9-fold, depending on the type of study evaluated, although the confounding effect of sexual activity that links both infections could not be eliminated.[27] A subsequent study from India supported the meta-analysis by demonstrating that incident HSV infection increased the risk of HIV acquisition by 3.8-fold.[28] Furthermore, a reanalysis of a discordant-couples study from Uganda showed that HSV-2 seropositive persons had a fivefold increased risk of acquiring HIV per sex act, compared with HSV-2 seronegative persons.[29,30] All of these studies are consistent with reports demonstrating that HSV can increase HIV replication and that HIV is easily cultivated from active HSV lesions.[31,32] Nevertheless, to firmly establish the role of HSV in enhancing HIV transmission, prospective longitudinal studies of sexual partners discordant for both HIV and HSV would need to be performed.

INJECTION DRUG USE

Although injection drug use accounts for fewer cases of HIV transmission than sexual contact, it has played a critical role in the epidemic in certain regions of the world. When injection drug use is coupled with unsafe sexual practices, HIV can rapidly spread through a population. Although needle-exchange programs are not supported by federal funds in the United States, more than 100 exchange programs have been implemented. Multiple studies have clearly shown that needle-exchange programs decrease the incidence of HIV transmission by an estimated 33% without causing an increase in drug use.[33]

USE OF CONTAMINATED BLOOD PRODUCTS

In the United States, the American Red Cross began testing donated blood for antibodies to HIV-1 in 1985; in 1992, this program was expanded to include testing for HIV-2. In 1992 and 1993, the risk of transfusion-acquired HIV in the United States was less than one case of HIV for every 450,000 to 600,000 donations of screened blood, with almost all of the risk related to donors who were infected with HIV shortly before donating their blood and who had not yet undergone seroconversion.[34] With current enzyme immunoassay serologic tests, this so-called window period typically lasts until about day 25 after HIV infection. In 1996, HIV p24 antigen testing was mandated; this test further decreased the risk of HIV transmission. More recent attempts have used nucleic acid testing as an even more sensitive assay to detect donors during the window period. In an interesting case report, a 13-year-old boy received a blood transfusion 2 days before it was discovered that the donor had acute HIV syndrome. The boy received a three-drug antiretroviral postexposure prophylaxis regimen for 9 months and repeatedly tested negative for HIV RNA antibodies, even when tested 6 months after antiretroviral postexposure prophylaxis was discontinued.[35] In China, unsafe blood donation techniques (i.e., reuse of contaminated phlebotomy equipment) led to the infection of some Chinese communities in the 1990s.[12]

MOTHER-TO-CHILD TRANSMISSION

Mother-to-child transmission of HIV accounts for a substantial proportion of HIV cases in developing countries. Such transmission can occur during gestation, at the time of delivery, or post partum via breast-feeding. In the absence of antiretroviral therapy, mother-to-child transmission rates generally range from 20% to 30%,[36] with higher rates occurring in developing countries and in women who breast-feed. Multiple studies have clearly shown the presence of HIV RNA in the mother to be a strong predictor of transmission.[37,38] In addition, in HIV-infected mothers who were not receiving antiretroviral therapy, cesarean section decreased the risk of transmission by about 50%.[39] Although HIV transmission from mother to child can occur early in pregnancy or post partum,[40-42] available data suggest that most

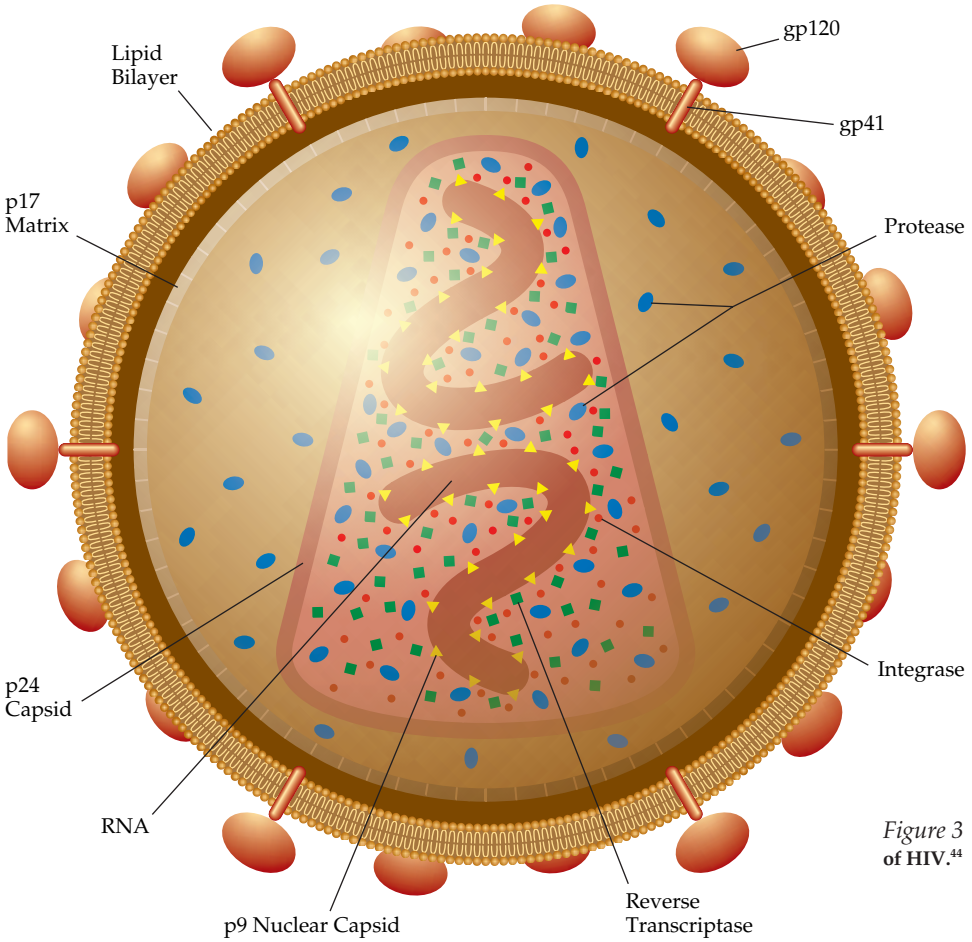


Figure 3 **The structure of HIV.**[44]

perinatal HIV transmission occurs near or at the time of delivery.[41] The widespread use of antiretroviral therapy to prevent mother-to-child HIV transmission has led to a dramatic decline in perinatal HIV transmission in developed countries.[36,41] Current recommendations for the use of antiretroviral therapy to prevent mother-to-child HIV transmission are discussed elsewhere [*see* Antiretroviral Therapy to Prevent Perinatal Transmission, *below*].

OCCUPATIONAL EXPOSURE

As of July 1, 2000, the CDC had reported 56 documented cases of occupational HIV transmission and an additional 138 cases of possible occupational HIV transmission.[2] Of the 56 documented cases, 48 (86%) involved a percutaneous exposure, five (9%) involved a mucocutaneous exposure, two (4%) involved both percutaneous and mucocutaneous exposures, and one (2%) occurred by an unknown route of exposure. In addition, 49 of the 56 cases (88%) involved exposure to blood from an HIV-infected individual. On the basis of available data, the average risk of HIV transmission to a health care worker who experiences a percutaneous exposure to HIV-infected blood but who does not receive postexposure prophylaxis is approximately 0.3%; the risk associated with a mucous membrane exposure to HIV-infected blood is approximately 0.09%.[43] There are no documented cases of HIV transmission in health care workers that resulted from exposure to intact skin. Risk factors associated with increased risk of transmission include deep needle-stick injuries, use of hollow-bore needles, use of a device visibly contaminated with blood, exposure events involving the transference of a large volume of blood, a needle-stick injury in which the needle had been placed directly in an artery or vein of an HIV-infected patient, and the source patient having advanced AIDS.[43] The relative risk of transmission in terms of the source patient's HIV RNA level has not been established for occupational HIV transmission.

Pathogenesis and Progression of Disease

HIV STRUCTURE AND LIFE CYCLE

HIV is a member of the lentivirus family of retroviruses. On electron microscopy, HIV appears as spherical particles that are approximately 110 nm in diameter, with knoblike projections on the surface of the virus and a cone-shaped viral core[44] [*see Figure 3*]. HIV particles contain two copies of an RNA genome, each of which is approximately 10,000 base pairs in length and encodes nine genes.

The genetic structure of HIV-1 is similar to that of all retroviruses. The genome is organized into three major regions (*gag*, *pol*, and *env*) that are flanked by the HIV-1 promoter or long terminal repeat. The *gag* region contains the structural genes for HIV (i.e., matrix, capsid, nucleocapsid, and two small peptides), the *pol* region contains the genes for the viral enzymes needed to carry out the life cycle (i.e., reverse transcriptase, integrase, and protease), and the *env* region encodes the genes for the viral envelope proteins (i.e., gp160, which is cleaved to gp120 and gp41).[45]

HIV-1 has six regulatory genes that are vital for its life cycle and pathogenicity: *tat* (the transactivating gene) upregulates

transcription of the genome; *rev* coordinates the expression of the regulatory and nonregulatory genes by orchestrating the transport of spliced and unspliced RNA transcripts out of the nucleus; *nef* helps the virus evade the host immune response by downregulating expression of CD4 and major histocompatibility complex (MHC) class I molecules on the cell surface and also contributes to viral virulence; *vpu* reduces host cell CD4 expression and is involved in cellular release of virions; *vpr* is important for infection of nondividing cells by facilitating nuclear localization of the viral preintegration complex and also regulates cell cycle arrest; and *vif* is important for virion assembly, infectivity, inactivation of the host cell antiviral factor APOBEC3G,[46] and gp120 membrane insertion.[47] The reverse transcriptase of HIV is very error-prone and introduces mutations at a rate of approximately 1 in 10^4, or about one mutation in every virus produced. In addition, during normal replication, the reverse transcriptase enzyme jumps from one strand of nucleic acid to another to complete the synthesis of daughter strands. This strand-jumping enables recombination between different viral strains infecting the same cell. Mutation and recombination permit the virus to respond rapidly to environmental changes such as those related to receptor availability, host immune responses, and antiretroviral drugs.

Two different HIV species have been identified: HIV-1 and HIV-2. Those isolates of HIV-1 that have been globally identified can be classified into the three major phylogenetic groups: M (main), N (neither M nor O), and O (outlier).[4,48] The M group has predominantly been responsible for the global HIV epidemic. This group can be further subdivided into 10 distinct subtypes or clades, termed subtypes A to J. Patients can be infected with more than one clade, and recombination of viruses from different clades can occur.

The life cycle of HIV is similar to that of other retroviruses. Understanding the life cycle is important for understanding both cellular pathogenesis and the targets of current and future anti-HIV therapies. The first step in the life cycle involves attachment of the virus to the host target cell. The first point of interaction consists of the binding of the HIV envelope surface protein (gp120) with the CD4 receptor on the host cell [*see Figure 4*]. For infection to proceed, cellular coreceptors must also bind gp120, causing its release and the subsequent exposure of the other HIV envelope protein, gp41. The gp41 protein mediates fusion of the virus and cell membranes by stabbing the cell membrane and undergoing a conformational change that brings the two membranes in contact, facilitating fusion and viral entry.

Although investigators have discovered numerous coreceptors that can potentially bind HIV, the coreceptors CCR5 and CXCR4 appear to play the most important role.[49] $CD4^+$ T cells express both CCR5 and CXCR4 receptors, whereas monocytes predominantly express CCR5 [*see Figure 5*]. Available data suggest that initial HIV infection predominantly involves R5 strains of HIV. There is significant genetic variation in the CCR5 coreceptor, and some individuals who are homozygous for a CCR5 deletion mutation (a 32–base-pair deletion) can be repeatedly exposed to HIV without becoming infected.[50] Individuals who are heterozygous for this Δ32 mutation are not protected against HIV infection but may initially manifest a slower rate of disease progression [*see Figure 6*].

After the fusion of the viral and cellular membranes, the viral capsid enters the cell [*see Figure 7*] and the HIV reverse transcriptase enzyme converts the single-stranded HIV RNA into a double-stranded DNA called proviral DNA. The provirus is then integrated into the host cell chromosome by the viral enzyme integrase. Subsequently, cellular enzymes transcribe the provirus into spliced and nonspliced messenger RNA (mRNA) molecules that encode the regulatory genes (*tat* and *rev*) and the structural genes and that serve as full-length genomic transcripts. This process proceeds in an organized fashion with the regulatory genes *tat* and *rev* (on spliced transcripts) expressed first, followed by the transport of full-length transcripts into the cytoplasm, which are then translated into the structural proteins or which serve as genomic RNA for progeny virus. The late stages of viral replication involve both the assembly of the viral particles, with each viral core incorporating two copies of the viral RNA genome, and the budding and release of the virus from the cell surface. The HIV protease enzyme plays an important role in this late process by cleaving the gag polyprotein into smaller functional components, which allows for the formation of mature, infectious viral particles.

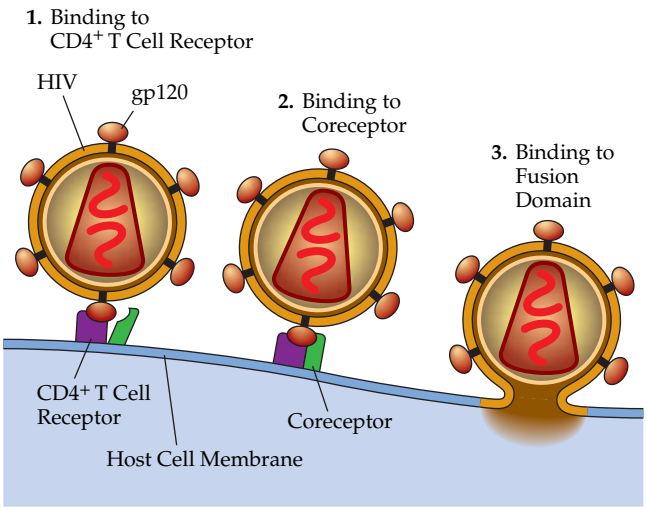


Figure 4 **This illustration shows the major steps involved in HIV binding to and entering the host cell. After binding to the fusion domain, the membranes merge.**

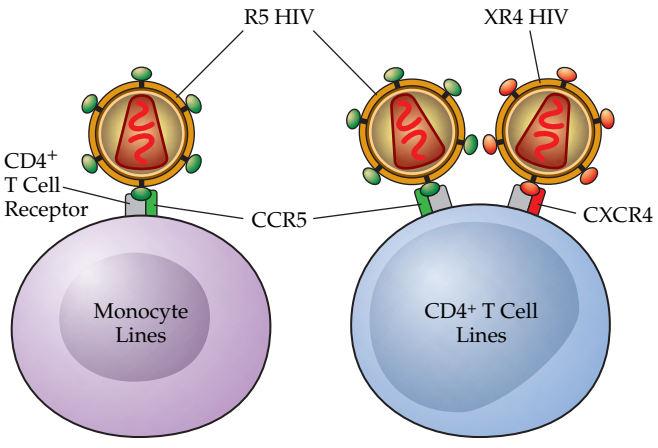


Figure 5 **R5 strains preferentially bind to CC chemokine receptor–5 (CCR5) coreceptors that are present on both monocytes and $CD4^+$ T cells. R4 strains preferentially bind to CXC chemokine receptor–4 (CXCR4) coreceptors (predominantly on $CD4^+$ T cells).**

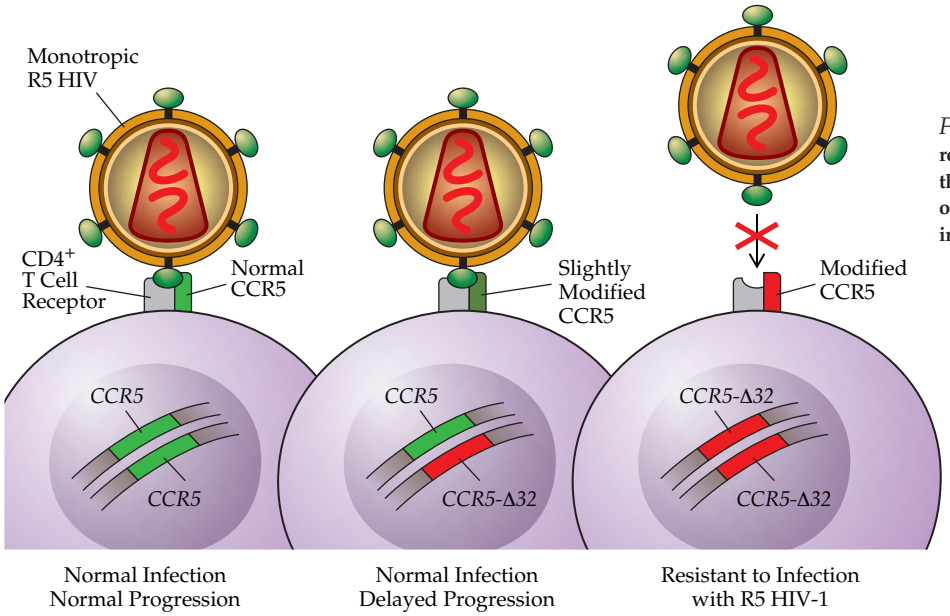


Figure 6 **This diagram shows the relationship of genetic diversity of the CCR5 coreceptor to progression of and susceptibility to HIV infection.**

EARLY EVENTS AND DISEASE PROGRESSION

Worldwide, most HIV infections occur by sexual transmission across mucosal surfaces. The R5 viruses are preferentially transferred across epithelial cell membranes,[51] where they may encounter dendritic cells, $CD4^+$ T cells, or macrophages.[52] HIV may productively infect any of these cell types or be tethered to dendritic cells by means of DC-SIGN (dendritic cell–specific ICAM [intercellular adhesion molecule]-grabbing nonintegrin) or other C-type lectin receptors (CLRs).[53,54] HIV infection of $CD4^+$ T cells is enhanced when the virus is captured and presented to the T cells by dendritic cells expressing DC-SIGN or other CLRs.[55-57] Locally infected T cells or dendritic cells coated with HIV then traffic to regional lymph nodes, where the virus propagates rapidly in the abundant $CD4^+$ T cell pool before disseminating.[58] At this point, patients may experience symptoms of an acute retroviral syndrome before mounting a cytotoxic T cell (CTL) response that partially controls viral replication.[59] However robust this CTL response may be, it never eradicates the infection, in part because the virus preferentially infects and kills the $CD4^+$ T cells that are recruited to the site of HIV replication.[60] This selective depletion of HIV-directed $CD4^+$ T cells cripples the effectiveness of HIV-directed $CD8^+$ T cells, which rely on $CD4^+$ T cell help for proper functioning. Furthermore, the high mutation rate of the virus allows it to escape the control of most potent immune responses. As a result, the infection persists, and continued rounds of replication lead to the gradual depletion of all $CD4^+$ T cells. At the same time, a subset of activated, HIV-infected $CD4^+$ T cells returns to a quiescent state, remains latently infected, and persists with a half-life of up to 44 months.[61] This latent reservoir of HIV can reactivate even after years of suppressive antiviral therapy.[62-64]

The plasma HIV RNA level is a strong independent predictor of the progression to AIDS in untreated HIV-infected persons [*see Figure 8*].[65] In essence, the higher the HIV RNA level, the more rapidly the disease will progress. For example, in patients with a $CD4^+$ T cell count greater than 500 cells/μl, about 5% of

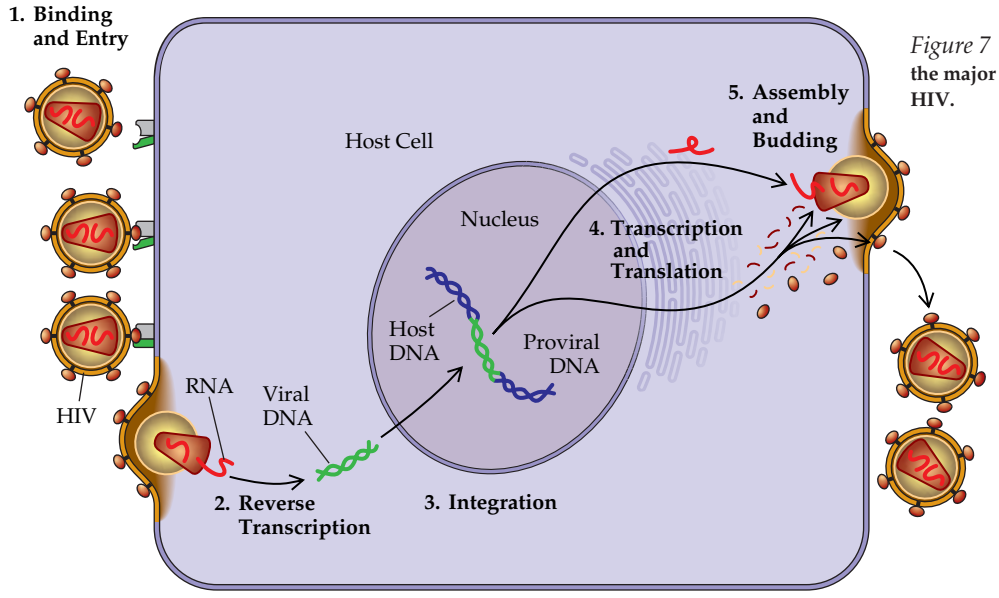


Figure 7 **This illustration depicts the major steps in the life cycle of HIV.**

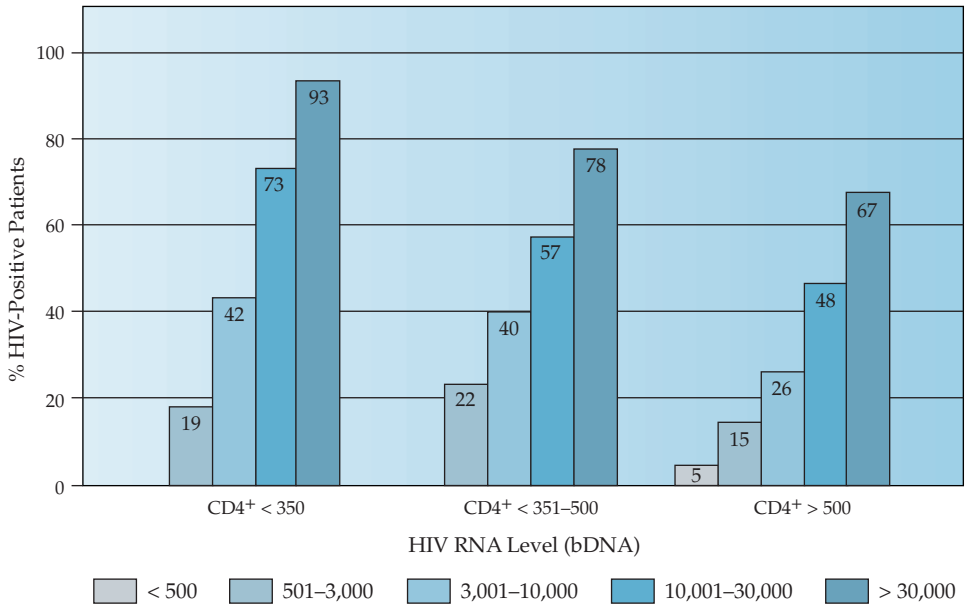


Figure 8 **This graph shows the risk of progression to AIDS at 6 years and the relationship of risk of progression to HIV RNA values and CD4$^+$ T cell counts.[65]**

those who have a baseline HIV RNA level of less than 500 copies/ml will develop AIDS within 6 years, compared with 67% of those who have a baseline HIV RNA level of greater than 30,000 copies/ml. In contrast, the CD4$^+$ T cell count provides an accurate way to assess the current immunologic status.

Within 6 months of primary HIV infection, the plasma level of HIV appears to reach a fairly constant level—a level referred to as the set point. The set point is determined by a number of host and viral factors and varies from person to person. The HIV-directed CTL response appears to be the most important factor in determining the HIV set point. Indeed, the development of the HIV CTL response coincides with the initial decline in HIV RNA levels.[66] Patients who mount a weak CTL response typically have a high set point and rapid progression of HIV disease; in contrast, patients with a strong CTL response have a low set point and slower disease progression [*see Figure 9*].[66] The strength and effectiveness of the T cell response are influenced by the human leukocyte antigen (HLA) genotype of the individual, with certain genotypes reportedly associated with different rates of disease progression; HLA-B27 and HLA-B57, in particular, have been strongly associated with long-term nonprogression.[67-69] The HIV CD8$^+$ CTL response requires a highly coordinated action of multiple cells, including CD4$^+$ helper T cells.[66]

Shortly after acute HIV infection, HIV begins to preferentially destroy HIV-directed CD4$^+$ helper T cells; this process impairs the critical interaction between host CD4$^+$ T cells and CD8$^+$ T cells and thus weakens the host CTL response.

Genetic variations in the host's HIV coreceptors also may play a significant role in the initial progression of HIV disease, most notably with regard to the CCR5-Δ32 heterozygous mutation [*see* HIV Structure and Life Cycle, *above*] and the CCR2-64I mutation.[50] Individuals with the CCR5-Δ32 or CCR2-64I mutation express coreceptors (or coreceptor levels) that are suboptimal for HIV binding, and thus, the level of HIV infection is depressed. Genetic differences between strains of HIV can also affect disease progression. HIV strains can be classified as either R5 or R4, depending on which cellular coreceptor the virus utilizes (CCR5 or CXCR4, respectively). This, in turn, influences whether the virus causes syncytia when grown in vitro.[70] Syncytia-forming HIV strains, which are predominantly R4 (CXCR4-using) viruses, destroy CD4$^+$ lymphocytes more effectively in vivo; their appearance correlates with a rate of CD4$^+$ T cell decline that may be almost three times more rapid than that seen in persons with the nonsyncytium strains.[70] Several reports have also described a small number of persons with very delayed disease progression who are infected with HIV strains containing particular sequence variations or deletions, such as absence of the regulatory gene *nef*. These genetic variations appear to render the virus less pathogenic.[71,72] Finally, individuals with persistent coinfection with the flavivirus GB virus C (GBV-C) may exhibit slower rates of HIV disease progression.[73] Proposed mechanisms for the interference of the HIV life cycle by GBV-C include a shift to a Th1 cytokine profile, downregulation of CCR5, an improved innate immune response to HIV, and decreased viral replication.[74]

Acute Infection

DIAGNOSIS

Clinical Features

After acquiring HIV, infected persons may develop a nonspecific influenzalike or mononucleosis-like illness. The exact proportion of persons with acute HIV infection whose illness reflects primary HIV infection is difficult to ascertain, but estimates range from 40% to 90%.[75] The most common route for HIV inoculation is via the genital mucosa; after genital inoculation, HIV initially infects Langerhans cells, fuses with local CD4$^+$ T cells, and then infiltrates deeper tissues. Within 48 to 72 hours after inoculation, HIV can spread to adjacent inguinal lymph nodes; systemic dissemination and viremia typically develop after about 7 days after initial infection (range, 4 to 11 days).[76] Once viremia develops, HIV extensively seeds lymphoid organs and the central nervous system. During this acute phase of infection, levels of HIV in the genital tract are apparently very high, and the person is likely to be highly infectious.

If a person with primary HIV infection develops a significant clinical illness, it typically begins from 7 to 14 days after the acquisition of the virus. In one study that followed persons considered to be at high risk for HIV, of 23 persons who became infected with HIV, 87% developed an acute HIV illness; of those who developed acute illness, 95% sought medical care, but only 25% were diagnosed as having acute HIV.[22] The correct diagnosis is often missed because of the nonspecific nature of the clinical illness—acute HIV infection may resemble influenza, mononucleosis, secondary syphilis, and more common viral illnesses. The most common signs and symptoms are fever (seen in 80% to 90% of patients), fatigue (70% to 90%), rash (40% to 80%), headache (32% to 70%), lymphadenopathy (40% to 70%), pharyngitis (50% to 70%), and myalgias or arthralgias (50% to 70%).[75] Although the clinical illness caused by acute HIV infection is often nonspecific, clinical findings of a morbilliform rash, acute meningoencephalitis, or mucocutaneous ulcerations may be particularly useful in making the diagnosis. Acute HIV illness typically persists for less than 14 days, but some patients have had illnesses that have extended for longer than 10 weeks.[22] Disease progression is generally more rapid in patients with severe and prolonged symptoms.

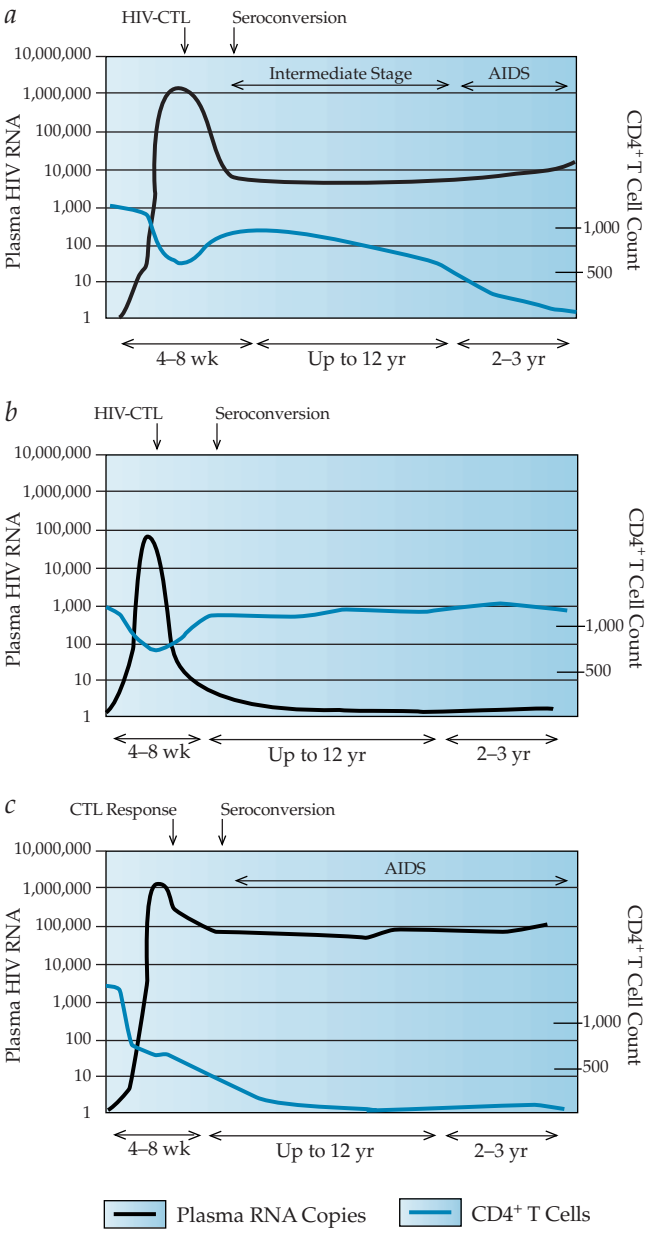


Figure 9 **Differing patterns of HIV-1 RNA and $CD4^+$ T cell levels correlate with different patterns of HIV disease progression. During acute HIV infection, plasma RNA levels rise sharply until the cytotoxic T cell (CTL) response to HIV develops; RNA levels then fall and stabilize at the viral set point. (*a*) Typical progression is marked by the death of many $CD4^+$ T cells, a fair HIV-CTL response, and fair HIV control. (*b*) Long-term nonprogression is marked by some $CD4^+$ T cell death, a good HIV-CTL response, and good HIV control. (*c*) Rapid progression is characterized by considerable $CD4^+$ T cell death, a poor HIV-CTL response, and poor HIV control.**

Laboratory Studies

Routine laboratory studies are generally nonspecific but may show varying degrees of lymphopenia or thrombocytopenia. Results of the standard recombinant enzyme-linked immunoassay (ELISA) are generally negative at the time a patient presents with acute HIV infection; these tests do not typically show a positive result until 22 to 27 days after HIV infection.[77] Although the plasma p24 antigen test is the only currently licensed test for diagnosing acute HIV infection, the HIV RNA assay yields positive results 3 to 5 days earlier than the p24 antigen test. The combination of a strongly positive HIV RNA test result and a negative HIV antibody test result confirms the laboratory diagnosis of acute HIV infection. Almost all persons with acute HIV infection have HIV RNA levels that exceed 50,000 copies/ml, and most have levels that exceed 300,000 copies/ml.[75] Persons presenting weeks or months after their acute HIV illness are likely to have significantly lower levels. A low-level positive result on the HIV RNA assay (i.e., a result of less than 10,000 copies/ml) in a person suspected of being recently infected should be viewed with great caution, as it may represent a false positive result. In this situation, if the p24 antigen test or a repeat HIV RNA test demonstrating a substantial rise cannot be obtained, the HIV serologic test should be repeated 4 to 6 weeks after the clinical illness has resolved.[75,78,79]

Implications of Early Diagnosis

Making a diagnosis of acute HIV is important for several reasons. First, persons with acute HIV are highly infectious and are more likely to transmit HIV. Second, after this initial illness, a very long interval may occur before the infected person develops HIV-related or AIDS-related symptoms, and as a result, the infection may go undiagnosed for years. Third, and most compelling, treatment of persons with acute HIV infection may present a unique opportunity to preserve HIV-directed $CD4^+$ T cells that are preferentially destroyed during acute infection. The preservation of these HIV-directed $CD4^+$ T cells may enable the development of robust $CD8^+$ T cell–mediated CTL responses that effect tight immune control of HIV. This could theoretically alter the natural course of the infection and delay disease progression even if antiretroviral therapy is withdrawn.

TREATMENT

At present, there are limited data on the treatment of patients with acute HIV infection. In the only randomized, placebo-controlled trial that has been performed, patients recently infected with HIV received either zidovudine or placebo. After 6 months, those who received zidovudine experienced a mean

increase in their CD4+ T cell count of 173 cells/μl; those receiving placebo had a mean increase of 6 cells/μl.[80]

Several observational studies have demonstrated that treatment of early or acute HIV infection helps preserve HIV-directed CD4+ and CD8+ T cells.[81-84] However, there are only limited data to suggest that these favorable immunologic profiles translate into long-term clinical benefits. In one study, opportunistic and nonopportunistic infections and progression to AIDS occurred less frequently in 20 patients with early HIV infection who received treatment than in 47 largely untreated historical control patients.[85] A large, randomized, placebo-controlled study has not been performed, however.

The 2004 Department of Health and Human Services (DHHS) guidelines note that certain authorities endorse treatment of acute HIV infection on the basis of the theoretical rationale and limited but supportive clinical trial data. In addition, these guidelines recommend considering therapy for patients who underwent seroconversion within the previous 6 months.[86] However, several questions remain unanswered: (1) How early after infection does treatment have to be initiated? (2) What is the optimal therapeutic regimen? and (3) When and how should HIV therapy be stopped?

Structured Treatment Interruption

One promising strategy for the treatment of primary infection has been the use of structured treatment interruptions. This is a process of scheduled starting and stopping of antiretroviral medications that is intended to enhance CTL responses and thus provide better immunologic control of HIV. The interest in structured treatment interruptions arose from studies that showed the importance of CTL responses in controlling HIV[87] and the observation that CTL responses became very weak in persons who achieved long-term control of HIV with antiretroviral therapy.[88] The goal of structured treatment interruptions is to intermittently expose the immune system to HIV-1 antigen in an attempt to upregulate HIV-specific CTL activity, thus leading to better immunologic control of HIV [*see Figure 10*].

Clinical interest in structured treatment interruption was sparked by a case report of a single patient, known as the Berlin patient, who received early aggressive antiretroviral therapy and then experienced several treatment interruptions; the Berlin patient achieved excellent control of HIV when not receiving antiretroviral therapy and showed excellent HIV-specific CTL responses.[89] A subsequent report showed initially promising results with structured treatment interruption in a carefully selected group of eight patients identified and treated during acute HIV infection. All eight received antiretroviral therapy for longer than 1 year, and all had HIV RNA levels that were consistently lower than 50 copies/ml.[90] With structured treatment interruptions, five of the eight patients were eventually able to maintain HIV RNA levels below 500 copies/ml without treatment. However, expansion of the group to 14 patients with longer follow-up yielded less encouraging results: by intention-to-treat analysis, control of HIV was maintained in only three of the 14 patients without antiretroviral therapy after 2 years.[91] The concept of structured treatment interruptions needs to be evaluated in larger, prospective, randomized studies before it can be recommended.

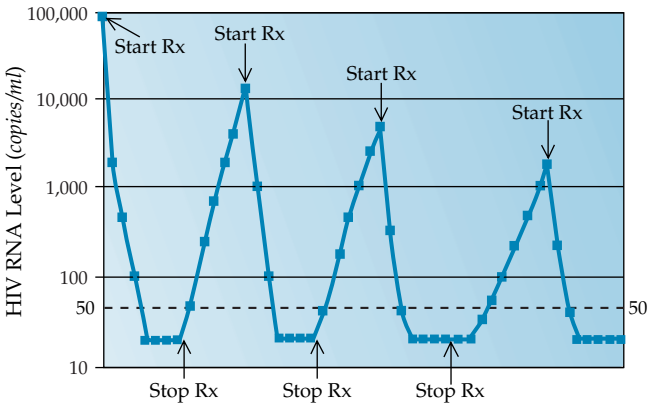


Figure 10 **Conceptual diagram of structured treatment interruption. Medications are started and stopped on several cycles. HIV RNA levels are measured in copies per milliliter. (Rx—therapy)**

Chronic (Established) Infection

INITIAL EVALUATION

The initial evaluation of a patient with HIV who is establishing routine care should include a thorough health history, physical examination, disease screens, vaccinations, and extensive laboratory studies. This initial evaluation process usually requires multiple visits. In the first visit, the health history is obtained, a complete physical examination is performed, and appropriate laboratory studies are ordered. During follow-up visits, which are conducted several weeks after the first, pertinent laboratory data are reviewed; the patient's disease stage and prognosis are discussed; necessary vaccinations are administered; the need for prophylaxis against opportunistic infection is reviewed; and, if appropriate, antiretroviral therapy is discussed.

Medical History

Particular aspects of the HIV-related medical history that should be emphasized are the risk factors for HIV infection, the approximate date of acquisition of HIV, results of the first HIV test that indicated the patient had HIV, a history of an acute seroconversion illness, prior CD4+ T cell and HIV RNA studies, prior HIV-related illnesses, prior and current antiretroviral therapy, and history of other infectious diseases, such as syphilis or hepatitis. Obtaining information regarding the patient's past and current sexual activity and drug use provides insight into both the original acquisition of HIV and current behaviors that may place contacts at risk for HIV infection. In addition to obtaining a complete medical history, it is important to assess the patient's social and psychiatric situation, because social and psychiatric issues can have a critical impact on the patient's adherence to antiretroviral therapy.

Laboratory Studies

The recommended laboratory studies that should be performed at the initial office visit are listed [*see Table 2*].[92]

OVERVIEW OF MANAGEMENT

Prevention Strategies

Recommendations regarding the incorporation of HIV prevention strategies into the care of HIV-infected patients have been developed by the CDC, the Health Resources and Services Administration, the National Institutes of Health, and the HIV

Table 2 Laboratory Studies for the Initial Routine Evaluation of an HIV-Infected Adult

Test	*Comment*
HIV antibody test	Need to firmly establish diagnosis; this test can be omitted if HIV RNA test is convincing
$CD4^+$ T cell count	Critical for staging disease and determining need for antiretroviral therapy
HIV RNA level	Critical for predicting progression of disease and determining need for antiretroviral therapy
Complete blood count	Hematologic problems are more common in HIV-infected persons; baseline value is useful because the number of medications used has effects on marrow
Serum chemistries	Useful to have routine serum chemistries as a baseline
Liver function tests	Useful to have as a baseline, considering potential hepatotoxicity of some antiretroviral medications; useful in patients with active hepatitis B virus (HBV) or hepatitis C virus (HCV) infection
Cholesterol panel	Useful as baseline; significant alterations in lipids can occur with antiretroviral therapy
Serologic testing for syphilis	Important with known higher rates of syphilis among HIV-infected persons
Toxoplasma IgG antibody	Useful in determining whether patient is at risk for developing toxoplasmosis; if negative, patient can receive counseling on preventing acquisition of *T. gondii*
Hepatitis A, B, or C antibody; hepatitis B antigen	Useful for identifying patients coinfected with HIV and either HBV or HCV; useful for identifying patients who are seronegative for hepatitis A virus or HBV and who should therefore receive vaccination; if HCV antibody is positive, testing should be performed to determine whether patient has a positive HCV viral load
Cytomegalovirus serology	IgG levels useful for determining past exposure and risk of reactivation
Urinalysis	Useful as a baseline screen
Purified protein derivative	Considered positive if ≥ 5 mm induration; controls are no longer recommended
Chest x-ray	Useful as a baseline; some specialists obtain a chest x-ray only in patients with a history of pulmonary disease

Medical Association of the Infectious Diseases Society of America.[93] These recommendations are made with the understanding that prevention strategies directed toward persons with known HIV infection can reduce the transmission of HIV and may be particularly effective because they are targeted to a specific group of persons rather than the entire at-risk population. The recommendations include the following: (1) screening patients for risky behaviors, (2) discussing behaviors that are associated with HIV transmission, (3) reinforcing practices that reduce HIV transmission, (4) referring patients with substance abuse problems to appropriate programs, (5) facilitating partner notification and HIV counseling and testing, and (6) regular testing for and treatment of sexually transmitted diseases. These recommendations are particularly important when caring for acutely infected persons, given the high viral loads and high infectivity associated with primary infection.

Vaccination

Asymptomatic HIV-infected patients who have a $CD4^+$ T cell count greater than 500 cells/μl generally produce good antibody responses to vaccines. Responses are frequently poor in patients with lower $CD4^+$ T cell counts, particularly those with AIDS. Persons with more advanced immune suppression may respond better to vaccines administered after immune reconstitution through antiretroviral therapy. Preliminary studies raised concerns that vaccines given to HIV-infected persons could cause significant increases in HIV RNA levels. Further study has shown that although some patients do have a mild to moderate increase in HIV RNA levels after immunization, these increases generally are transient and do not have permanent adverse consequences. Nevertheless, concerns remain regarding the use of live-virus vaccine in HIV-infected patients, particularly those with advanced immune suppression. Vaccines for HIV-infected adults are listed [*see Table 3*].

Routine Follow-up

How often a patient should return for routine follow-up visits depends on the immune status, the clinical status, concomitant medical problems, and whether or not the patient is receiving antiretroviral therapy. In general, patients should have visits at least every 3 months for monitoring of $CD4^+$ T cell counts, HIV RNA levels, and clinical status. More frequent visits are often needed in patients with significant ongoing HIV-related problems or ongoing concomitant medical problems and in those who have recently begun antiretroviral therapy. Visits every 6 months would be appropriate for patients whose disease will probably not progress in the interim, who have a $CD4^+$ T cell count greater than 500 cells/μl, and whose HIV RNA level is less than 10,000 copies/ml.

Antiretroviral Therapy

Before starting antiretroviral therapy, the physician must give careful consideration to the patient's social and medical conditions. In general, the patient should be both interested in and able to take antiretroviral medications consistently. Active social problems (e.g., ongoing drug use, homelessness, or unstable emotional states) can be major barriers to successful antiretroviral therapy. In addition, if a patient has an active severe medical problem, every attempt should be made to defer antiretroviral therapy until that problem is stabilized. Medical indications for antiretroviral therapy are based on the patient's clinical status, the $CD4^+$ T cell count, and the HIV RNA level [*see* Principles for Antiretroviral Therapy, *below*]. There is a wide array of possible antiretroviral drug combinations; in choosing a specific regimen, the clinician should take into account the dosing schedule, possible adverse effects, and underlying patient preferences.

Agents Currently, four different classes of medications, comprising a total of 20 drugs, target HIV: (1) nucleoside reverse transcriptase inhibitors (NRTIs)—abacavir, didanosine, emtricitabine, lamivudine, stavudine, tenofovir disoproxil fumarate (tenofovir-DF), zalcitabine, and zidovudine; (2) nonnucleoside reverse transcriptase inhibitors (NNRTIs)—delavirdine, efavirenz, and nevirapine; (3) protease inhibitors—atazanavir, fosamprenavir, indinavir, lopinavir plus ritonavir, nelfinavir, ritonavir,

Table 3 Vaccinations for HIV-Infected Adults

Vaccine	*Comment*
Haemophilus influenzae type B vaccine	Not generally recommended for adults
Hepatitis A virus vaccine (Havrix, VAQTA)	Recommended for men who have sex with men, injection-drug users, and persons with chronic liver disease; dosage is 1.0 ml given intramuscularly and repeated 6–12 mo later
Hepatitis B virus (HBV) vaccine (Recombivax HB, Engerix-B)	Recommended; regimen for vaccination is 10–20 µg of recombinant HBV vaccine given in the deltoid muscle at 0, 1, and 6 mo; responses are often suboptimal, particularly in those with more advanced immune suppression
Influenza virus vaccine	Recommended; the dosage is 0.5 ml I.M. and should be administered during the period from late October through late December; the goal is to prevent influenza and secondary bacterial pneumonias
Mumps, measles, and rubella vaccine	Recommended, with exceptions; the dosage is 0.5 ml S.C.; patients with severe immunodeficiency should not receive this live vaccine; this vaccine is generally recommended for persons born after 1956 who do not have a documented history of measles or measles vaccination and persons who received the killed vaccine (1963 to 1967) and do not have a documented history of measles or live measles vaccination
Pneumococcal vaccine (polysaccharide 23-valent vaccine)	Recommended if $CD4^+$ T cell count is > 200 cells/µl; the dosage is 0.5 ml I.M. repeated every 5 years; consider repeating vaccine if $CD4^+$ T cell count was less than 200 cells/µl but has increased to greater than 200 cells/µl; there are insufficient data in HIV-infected persons to recommend conjugated 7-valent vaccine
Polio vaccine	Recommended; inactivated vaccine only; live oral polio vaccine is contraindicated in HIV-infected patients; patients who have not been immunized should receive inactivated polio vaccine (trivalent killed poliovirus), 0.5 ml S.C. at 0, 1, and 6 mo
Tetanus-diphtheria toxoid	Recommended; patients previously immunized should receive a booster every 10 years; those never immunized should receive primary vaccination, 0.5 ml I.M. at 0, 1, and 6 mo
Varicella vaccine	Insufficient data to recommend for adults

and saquinavir; and (4) fusion inhibitors—enfuvirtide [*see Figure 11 and Table 4*].[91] The NRTIs—also known as nucleoside analogues (or nucleotide analogues, in the case of tenofovir-DF)—structurally resemble the human nucleosides that HIV uses to make viral DNA. The HIV reverse transcriptase enzyme can mistakenly incorporate the synthetic nucleoside analogue into the elongating strand of viral DNA during the reverse transcriptase process; once incorporated into viral DNA, the nucleoside analogues act as chain terminators because they lack the 3′ hydroxyl group required for chain elongation. The NNRTIs do not act as chain terminators; rather, they directly inhibit the proper functioning of the reverse transcriptase enzyme. The HIV protease inhibitors selectively bind to HIV protease and prevent this enzyme from performing its normal function of cleaving viral polyprotein precursors into individual functional proteins. Successful inhibition of HIV protease causes the formation of deformed HIV particles that generally do not replicate. The fusion inhibitor enfuvirtide represents the newest class of antiretroviral agents. It works by binding to the gp41 envelope protein of HIV to prevent it from mediating fusion of the viral and cell membranes.

Dosing schedules Practical considerations have led to the increasing use of antiretroviral regimens that have infrequent dosing schedules, minimal meal restrictions, and the least possible adverse effects. Most of the NRTIs can be taken twice daily without regard to food consumption; the exceptions are zalcitabine, which is taken three times daily, and didanosine, which is taken a half hour before a meal or 1 hour after a meal. Although the new enteric-coated didanosine capsule is taken once daily, it should not be taken with meals. Alcohol significantly increases the plasma levels of abacavir. The NNRTIs can be taken without regard to food consumption, with the caveat that efavirenz should not be taken after a meal high in fats, because this will increase the plasma level of efavirenz by approximately 50%. Dosing schedules and food restrictions are more complicated with the protease inhibitors. Atazanavir, nelfinavir, saquinavir, and ritonavir should be taken with food; fosamprenavir can be taken with or without meals. Indinavir should be taken 1 hour before or 2 hours after a meal, or it can be taken with a low-fat meal.

Combination therapy Increasingly, pharmacologic boosting of protease inhibitor levels is being used to simplify the dosing schedule, decrease the pill burden, and improve the efficacy of several of the protease inhibitors. Most often, this technique has involved the use of low-dose ritonavir in combination with another protease inhibitor, also at a reduced dosage. Ritonavir significantly increases the trough level of the other protease inhibitor, usually without causing a major increase in the peak levels [*see Figure 12 and Table 5*]. Long-term data from studies on the combination of ritonavir and saquinavir have shown this combination to be highly effective.[94] The fixed combination of ritonavir and lopinavir (Kaletra) is highly effective and is preferred by many providers. Ritonavir is now also commonly used with atazanavir, fosamprenavir, or indinavir to boost the blood level of coadministered drug, simplify the dosing schedule, and reduce the pill burden. Two other fixed combinations are now available: emtricitabine and tenofovir-DF (Truvada) and abacavir and lamivudine (Epzicom).

Principles of Antiretroviral Therapy

An expert panel convened by the DHSS and the Henry J. Kaiser Family Foundation has issued antiretroviral therapy recommendations for medical providers in the United States. These guidelines, which are based on data from clinical trials, basic virologic principles, and clinical experience, were last updated March 23, 2004, and will continue to be modified as needed.[91] A second set of published guidelines was issued by the International AIDS Society[95]; these guidelines had more international input than the DHHS guidelines. As described in the DHHS guidelines, the goal of antiretroviral therapy is to obtain maximal and durable suppression of HIV, restore and preserve im-

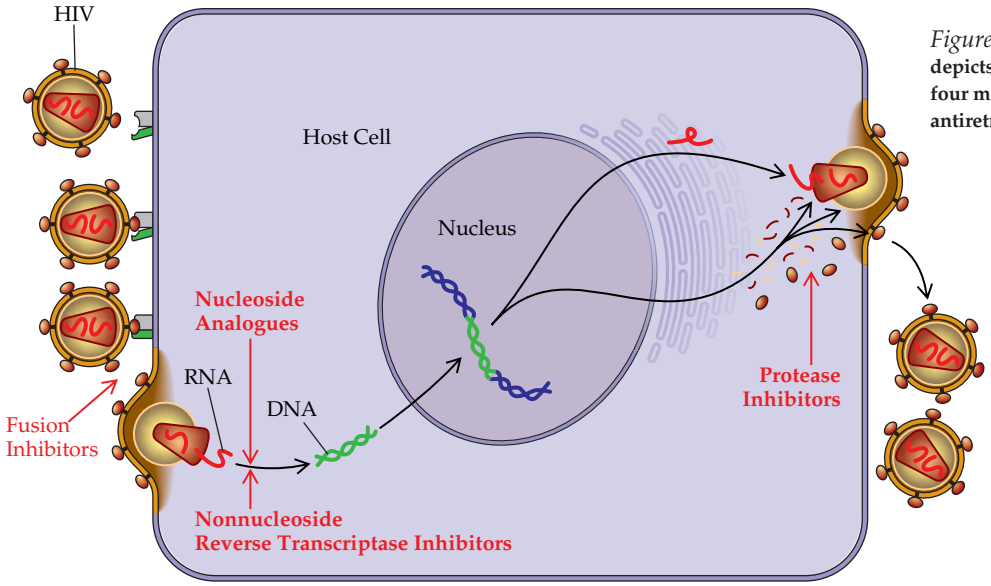


Figure 11 **This illustration depicts the site of action of four major classes of antiretroviral medications.**

mune function, improve quality of life, and reduce HIV-related mortality.

INITIATION OF THERAPY

The DHHS guidelines recommend that antiretroviral therapy be offered to chronically infected patients who have severe HIV-related symptoms or AIDS or who are asymptomatic but meet certain immunologic or virologic thresholds. Asymptomatic patients with $CD4^+$ T cell counts below 200/μl should be offered therapy because of the significant risk of opportunistic infection and death without treatment. Asymptomatic patients with $CD4^+$ T cell counts between 200 and 350/μl should generally be offered treatment unless their plasma HIV RNA level is low, in which case their short-term risk of disease progression and death is also low. Asymptomatic patients with $CD4^+$ T cells above 350/μl can generally be followed without treatment unless their HIV RNA level is high (> 55,000 copies/ml by reverse transcriptase polymerase chain reaction [RT-PCR] or branched DNA [bDNA] assay), in which case some providers recommend treatment because of the risk of more rapid disease progression. Patients who are diagnosed during acute infection or within 6 months of infection may have the unique opportunity to modify their disease history with treatment (see above). Theoretically, early treatment may lower the viral set point, preserve HIV-directed $CD4^+$ T cells, relieve symptoms of acute infection, and prevent transmission. These benefits, however, must be weighed against the toxicity of therapy, the possible promotion of drug-resistant HIV strains, and the lack of long-term clinical data demonstrating a clear benefit of early therapy.

REGIMEN SELECTION

Antiretroviral therapy should consist of a potent regimen that includes a combination of three or more medications [*see Table 6*]. The combination regimens typically consist of (1) two NRTIs plus a protease inhibitor, (2) two NRTIs plus an NNRTI, or (3) two NRTIs plus ritonavir plus a second protease inhibitor. In the last case, ritonavir's potent inhibition of the cytochrome P-450 elimination pathway is used to boost the blood level of the second protease inhibitor. Preferred regimens for antiretroviral-naive patients have been identified in the DHHS guidelines [*see Table 6*].

Enfuvirtide is most effective in patients with $CD4^+$ T cell

Table 4 Antiretroviral Medications

Generic Name	*Trade Name*	*Dosage*
Nucleoside RTIs		
Abacavir (ABC)	Ziagen	300 mg p.o., b.i.d.
Didanosine (ddI)	Videx	400 mg p.o., q.d. (250 mg q.d. if < 60 kg)
Emtricitabine (FTC)	Emtriva	200 mg p.o., q.d.
Lamivudine (3TC)	Epivir	150 mg p.o., b.i.d.
Stavudine (d4T)	Zerit	40 mg p.o., b.i.d. (30 mg p.o., b.i.d., if < 60 kg)
Zalcitabine (ddC)	Hivid	0.75 mg p.o., t.i.d.
Zidovudine (AZT)	Retrovir	300 mg p.o., b.i.d.; or 200 mg t.i.d.
Zidovudine + lamivudine*	Combivir	One tablet p.o., b.i.d.
Zidovudine + lamivudine + abacavir*	Trizivir	One tablet p.o., b.i.d.
Nucleotide RTI		
Tenofovir	Viread	300 mg p.o., q.d.
Nonnucleoside RTIs		
Delavirdine	Rescriptor	400 mg p.o., t.i.d.
Efavirenz	Sustiva	600 mg p.o., q.d. (dose can be split and given 200 mg in A.M. and 400 mg in P.M.)
Nevirapine	Viramune	200 mg p.o., q.d. × 14 days, then 200 mg p.o., b.i.d.
Protease inhibitors		
Atazanavir	Reyataz	400 mg p.o., q.d.
Fosamprenavir	Lexiva	1,400 mg q.d. or 700 mg b.i.d.
Indinavir	Crixivan	800 mg p.o., q. 8 hr
Lopinavir + ritonavir	Kaletra	3 p.o., b.i.d. (lopinavir, 400 mg b.i.d. + ritonavir, 100 mg b.i.d.)
Nelfinavir	Viracept	1,250 mg p.o., b.i.d., or 750 mg p.o., t.i.d.
Ritonavir	Norvir	600 mg p.o., b.i.d. (dose escalation†)
Saquinavir	Fortovase	1,200 mg p.o., t.i.d.

*Standard dosages are used in fixed combinations.
†Ritonavir dose escalation: 300 mg p.o., b.i.d. × 2 days; 400 mg p.o., b.i.d. × 3 days; 500 mg p.o., b.i.d. × 8 days; 600 mg p.o., b.i.d.
RTI — reverse transcriptase inhibitor

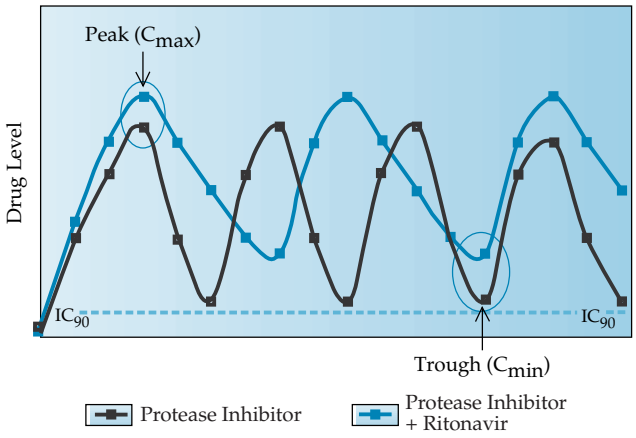


Figure 12 **In this illustration, the drug levels of a protease inhibitor are compared with boosted levels of ritonavir. (IC_{90}—concentration of drug required to inhibit 90% of viral replication)**

counts above 100 cells/μl; in these cases, enfuvirtide is combined with a standard regimen of drugs that includes one or two agents to which the patient's virus is sensitive. However, patients who do not meet these criteria may still experience a virologic response and some immunologic benefit from enfuvirtide. These benefits must be weighed against the cost, difficulty of administration (twice-daily subcutaneous injections), and adverse effects (e.g., injection-site reactions, hypersensitivity, and possible pneumonia).[96]

DRUG SELECTION

Adverse Effects of Drugs

Potential adverse effects can greatly influence the choice of antiretroviral medications and a patient's ability and willingness to stay on long-term antiretroviral therapy [*see Table 7*]. Possible acute life-threatening reactions include didanosine-induced pancreatitis, abacavir-related hypersensitivity syndrome, lactic acidosis secondary to the use of any of the NRTIs but more closely associated with stavudine, Stevens-Johnson syndrome secondary to the use of any of the NNRTIs, and nevirapine-associated liver failure.[97] Several older reports have described fatalities in patients taking hydroxyurea in combination with didanosine and with didanosine plus stavudine, so the use of hydroxyurea as adjuvant treatment with NRTIs is no longer recommended. Unfortunately, the pancreatitis associated with didanosine can occur without warning at any time during therapy.

A number of metabolic abnormalities have been reported in HIV-infected persons, some of which are linked to treatment with antiretroviral agents. Reported metabolic derangements include lactic acidosis, insulin resistance, hyperlipidemia, body-fat redistribution (lipodystrophy), osteonecrosis, and osteopenia.[98]

Hypersensitivity syndrome The abacavir hypersensitivity syndrome occurs in approximately 2% to 3% of patients receiving that drug and is strongly associated with a particular HLA haplotype (HLA-B*5701, HLA-DR7, and HLA-DQ3).[99] It typically develops within the first 3 to 4 weeks after abacavir is started. Symptoms of abacavir hypersensitivity syndrome most often consist of rash, fever, nausea, oral lesions, and cough. Patients often describe an accentuation of these symptoms shortly after taking each dose of abacavir. Patients who begin to develop the hypersensitivity syndrome and then stop taking abacavir can have an acute, potentially fatal reaction on resuming abacavir therapy. Treatment of a patient with mild symptoms that are suggestive of abacavir hypersensitivity should be conducted in consultation with an expert on this matter.

Lactic acidemia Lactic acidemia is defined as an elevated venous lactate level (> 2 mmol/L) and a normal arterial pH. Lactic acidosis is present when the arterial pH is below 7.3. Mild, usually asymptomatic lactic acidemia occurs in 8% to 21% of patients on antiretroviral therapy,[100,101] whereas symptomatic acidemia occurs in less than 2.5%. The leading proposed mechanism is NRTI inhibition of the mitochondrial DNA gamma polymerase. Risk factors for the development of lactic acidemia include longer duration of NRTI exposure (especially with stavudine[102]), female gender, and pregnancy. Symptoms associated with increased serum lactate levels are generally vague and may include nausea, malaise, and anorexia and weight loss, which may go on for months. Symptoms of more severe disease in-

Table 5 Ritonavir Boosted Protease Inhibitor Regimens

Ritonavir Dose	*Other Protease Inhibitor*	*Comment*
100 mg b.i.d.	Lopinavir, 400 mg b.i.d.	FDA-approved fixed combination with excellent potency
400 mg b.i.d.	Saquinavir, 400 mg b.i.d.	Combination with best long-term data; highly effective combination, but intolerance to ritonavir common at this dosage Can use with hard-gel (Invirase) or soft-gel (Fortovase) formulations
100 mg b.i.d.	Saquinavir, 1,000 mg b.i.d.	Can use with hard-gel (Invirase) or soft-gel (Fortovase) formulations, but better pharmacokinetic profile with hard-gel formulation
100 mg b.id. *or* 200 mg b.i.d.	Indinavir, 800 mg b.i.d.	Moderate amount of data suggest regimen very potent; can be taken with meals; increased indinavir toxicity at this dosage
400 mg b.i.d.	Indinavir, 400 mg b.i.d.	Moderate amount of data suggest regimen very potent; ideal pharmacokinetics, but intolerance to ritonavir common at this dosage
200 mg q.d.	Fosamprenavir, 1,400 mg q.d.	Once-daily dosing of fosamprenavir not recommended for protease inhibitor–experienced patients
100 mg b.i.d.	Fosamprenavir, 700 mg b.i.d.	—

Table 6 DHHS Recommendations for Antiretroviral Regimens for the Treatment of Established HIV in Adults and Adolescents[87*]

Regimen Type	*Drugs*	*No. of Pills per Day*
NNRTI based	**Efavirenz + lamivudine + (zidovudine or tenofovir DF or stavudine*)[†]**	**3–5**
	Efavirenz + emtricitabine + (zidovudine or tenofovir DF or stavudine*)[†]	3–4
	Favirenz + (lamivudine or emtricitabine) + (didanosine or abacavir)	3–5
	Nevirapine[‡] + (lamivudine or emtricitabine) + (zidovudine or stavudine* or didanosine or abacavir)	4–5
PI based	**Lopinavir/ritonavir (Kaletra) + lamivudine + (zidovudine or stavudine*)**	**8–10**
	Atazanavir + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)[†]	4–5
	Fosamprenavir + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)	6–8
	Fosamprenavir/ritonavir[§] + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)	6–8
	Indinavir/ritonavir[§] + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)	8–11
	Lopinavir/ritonavir (Kaletra) + emitricitabine + (zidovudine or stavudine* or abacavir)	8–9
	Lopinavir/ritonavir (Kaletra) + lamivudine + abacavir	8–9
	Nelfinavir[‖] + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)	12–14
	Saquinavir (SGC or HGC)/ritonavir[§] + (lamivudine or emtricitabine) + (zidovudine or stavudine* or abacavir)	14–16
Triple NRTI[¶]	Abacavir + lamivudine + (zidovudine or stavudine*)	2–6

Note: Preferred regimens, drugs, and doses are in bold type: regimens are designated as preferred for use in treatment-naive patients when clinical trial data suggest optimal and durable efficacy with acceptable tolerability and ease of use. Alternative regimens are those for which clinical trial data show efficacy, but the preferred agent has better antiviral activity or demonstrated durable effect, tolerability, or ease of use. In some cases, a regimen listed as an alternative regimen in the table may actually be the preferred regimen because of individual patient characteristics.

*Higher incidences of lipoatrophy, hyperlipidemia, and mitochondrial toxicities are reported with stavudine than with other NRTIs.

†Regimen not to be used in women with childbearing potential (i.e., women who want to conceive or who are not using effective contraception).

‡A high incidence of symptomatic hepatic events has been observed in women with pre-nevirapine CD4+ T cell counts > 250 cells/μl (11%) and in men with counts > 400 cells/μl (6.3%). Use with caution in these patients, with close clinical and laboratory monitoring, especially during the first 18 wk of therapy.

§Low-dose (100–200 mg) ritonavir.

‖Nelfinavir is available in a 250 mg or 625 mg tablet.

¶A triple-NRTI regimen is to be used only when an NNRTI-based or a PI-based regimen cannot or should not be used as first-line therapy.

HGC—hard gel capsule NRTI—nucleoside reverse transcriptase inhibitor NNRTI—nonnucleoside reverse transcriptase inhibitor PI—protease inhibitor SGC—soft gel capsule

clude abdominal pain, dyspnea, and encephalopathy. Although severe lactic acidosis is unusual, patients with serum lactate levels greater than 10 mmol/L have a mortality of approximately 80%.[97] Treatment of symptomatic lactic acidemia should include discontinuance of NRTIs and supportive care. Most patients can be successfully rechallenged with other NRTIs after the acidemia has resolved.[103]

Insulin resistance Protease inhibitor–based regimens are associated with significant increases in insulin resistance (i.e., increased insulin concentrations and increased C-peptide levels)[104-106]; in one historical cohort study of 221 patients, the reported incidence of new-onset hyperglycemia was 5%.[107] Other retrospective studies have reported hyperglycemia in 3% to 17% of patients. Postulated mechanisms include a direct effect of HIV therapy on glucose uptake, drug-induced body-fat redistribution (lipodystrophy) (see below) that is associated with increased insulin resistance, and, in the case of indinavir,[108] inhibition of peroxisome proliferator-activated receptor gamma. Reported treatment strategies include substitution of the protease inhibitor with an NNRTI or abacavir.[109-111] Otherwise, treatment should be as recommended for non–HIV-infected patients with diabetes and should include diet and exercise, followed by insulin-sensitizing agents and insulin, if necessary.

Hyperlipidemia Before the widespread use of antiretroviral therapy, reduced levels of high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, and total cholesterol and elevated levels of triglycerides were reported in HIV-infected patients.[112-114] The use of protease inhibitors has been associated with a return of LDL cholesterol levels to normal or with high LDL cholesterol levels in conjunction with marked hypertriglyceridemia.[115-118] Ritonavir has been associated with the greatest increases in triglyceride levels,[117] whereas other protease inhibitors, such as atazanavir and nelfinavir, are less closely linked to this effect. The evidence of an association between NNRTIs and hyperlipidemia is mixed. Improvements in hyperlipidemia are more marked in patients changing from a protease inhibitor to nevirapine than to efavirenz.[110,119,120] Except for one study showing higher lipid levels with stavudine than with tenofovir-DF, the NRTIs have not been associated with hyperlipidemia.[121]

The important clinical question is whether these abnormal lipid values translate into an increased risk of cardiovascular and cerebrovascular disease. Two very large studies reached different conclusions regarding this issue. A study that retrospectively abstracted coded data from records of 36,000 HIV-infected Veterans Affairs patients followed between 1995 and 2001 found a reduction in the rate of hospital admission for cardiovascular or cerebrovascular disease and a reduction in death from all causes that was associated with antiretroviral therapy.[122] In contrast, a study of 23,000 HIV-infected patients enrolled in the multinational Data Collection on Adverse Events of Anti-HIV Drugs (DAD) study demonstrated a 26% increase in the risk of myocardial infarction for each year of antiretroviral treatment.[123] Neither study had the long follow-up periods usually needed to assess cardiovascular disease outcomes, and neither study included an HIV-negative control group. Nevertheless, it is likely that the hyperlipidemia seen in antiretroviral-treated HIV-infected patients is one of several factors that does increase the risk of development of cardiovascular and cerebrovascular disease. However, this potentially adverse effect is far outweighed by the benefits of antiretroviral therapy.

Table 7 Adverse Effects Associated with Antiretroviral Medications

Generic Name	*Trade Name*	*Toxicity*
Nucleoside RTIs		All nucleoside RTIs may be associated with mitochondrial toxicity, lactic acidemia, and lipodystrophy (atrophy)
Abacavir (ABC)	Ziagen	Hypersensitivity reaction in 3% of patients; characterized by fever, rash, nausea, vomiting, or malaise; may be fatal; lactic acidosis (rare)
Didanosine (ddI)	Videx	Peripheral neuropathy, pancreatitis, nausea, lactic acidosis, lipoatrophy
Emtricitabine (FTC)	Emtriva	Few side effects; lactic acidosis (rare)
Lamivudine (3TC)	Epivir	Few side effects; lactic acidosis (rare)
Stavudine (d4T)	Zerit	Peripheral neuropathy; lactic acidosis (rare)
Zalcitabine (ddC)	Hivid	Peripheral neuropathy, stomatitis, lactic acidosis (rare)
Zidovudine (AZT)	Retrovir	Anemia, neutropenia, gastrointestinal problems, headache, lactic acidosis (rare), lipoatrophy
Nucleotide RTIs		
Tenofovir	Viread	Renal dysfunction (unusual)
Nonnucleoside RTIs		
Delavirdine	Rescriptor	Rash (4%); increased transaminase levels; headache
Efavirenz	Sustiva	Rash seen in 2% of patients; central nervous system side effects (confusion, abnormal dreams, agitation, etc.)
Nevirapine	Viramune	Rash seen in 7% of patients; increased transaminase levels; hepatitis (can be fatal)
Protease inhibitors		Most protease inhibitors are associated with lipid abnormalities, lipodystrophy (central fat accumulation) and gastrointestinal side effects
Fosamprenavir	Lexiva	Gastrointestinal intolerance; lipid abnormalities; perioral paresthesias; fat redistribution
Atazanavir	Reyataz	Gastrointestinal intolerance, elevated bilirubin levels
Indinavir	Crixivan	Gastrointestinal intolerance; nephrolithiasis; increase of indirect bilirubin; headache; lipid abnormalities; fat redistribution
Lopinavir + ritonavir	Kaletra	Gastrointestinal intolerance; lipid abnormalities, including increased triglyceride levels; fat redistribution
Nelfinavir	Viracept	Diarrhea; lipid abnormalities; major drug–drug interactions; fat redistribution
Ritonavir	Norvir	Gastrointestinal intolerance; perioral and extremity paresthesias; taste perversions; asthenia; hepatitis; major drug–drug interactions; lipid abnormalities, including increased triglyceride levels; fat redistribution
Saquinavir	Fortovase, Invirase	Gastrointestinal intolerance; headache; lipid abnormalities; fat redistribution

RTI — reverse transcriptase inhibitor

Current recommendations from an International AIDS Society–USA panel for treatment of hyperlipidemia include the following: (1) a possible change to protease inhibitor–sparing antiretroviral regimens, (2) adherence to diet and exercise guidelines, and (3) use of lipid-lowering agents.[98] Specific recommendations for the use of lipid-lowering agents follow those of the United States National Cholesterol Education Program guidelines.[124] Of the statins, pravastatin has the least severe drug interactions with protease inhibitors and is often used as a first-line agent. Although atorvastatin has more interactions with protease inhibitors than does pravastatin, few adverse events have been reported with atorvastatin. In addition, atorvastatin has greater potency than pravastatin, and it has a favorable impact on triglyceride levels.[125] Patients with marked hypertriglyceridemia often require treatment with a fibric acid analogue (e.g., fenofibrate or gemfibrozil).

Lipodystrophy The lipodystrophy syndrome, also known as the fat-redistribution syndrome, has become a problematic long-term complication of antiretroviral therapy. The lipodystrophy syndrome typically presents as central fat accumulation (i.e., accumulation of fat in the abdomen, breasts, or posterior neck) or peripheral fat wasting (i.e., wasting in the face, buttocks, and limbs), or both[126] [*see Figure 13*]. Lipodystrophy is reported in 25% to 75% of HIV-infected patients.[112,113,127-129]

Several mechanisms have been proposed for the lipodystrophy syndrome, but none are proven. Protease inhibitors, because of their role in inducing insulin resistance, have been blamed for the central lipoaccumulation.[112,130] Fat accumulation has also been reported in HIV-infected patients who were not exposed to protease inhibitors, however.[127,131] Lipoatrophy is more closely tied to exposure to NRTIs,[126,132,133] particularly stavudine. In this setting, NRTI-induced mitochondrial poisoning (as occurs in NRTI-induced lactic acidemia) may be responsible for adipocyte death and loss of subcutaneous fat. Other possible risk factors for the syndrome include older age, low body weight before starting antiretroviral therapy, duration of HIV infection, duration and effectiveness of antiretroviral therapy, and white race.[97,107,134-136]

A number of drugs and interventions have been tried as treatment for lipodystrophy, with only modest success. Switching from stavudine to other NRTIs has reversed limb-fat atrophy in some patients,[137] but substitution of protease inhibitors to reverse fat accumulation has not been successful. Exercise may have only a limited benefit in reducing fat accumulation,[138] but it does have other obvious health advantages. Growth hormone has been shown to benefit a small number of patients with fat accumulation,[139] but this approach is not considered practical, given the extreme cost, adverse effects, and the need for continued use to sustain benefit. Promising results in the treatment of lipodystrophy have been reported from preliminary studies of the insulin-sensitizing agent metformin. In a randomized, double-blind, placebo-controlled trial, use of metformin at a dosage of 500 mg twice daily for 3 months was associated with decreased

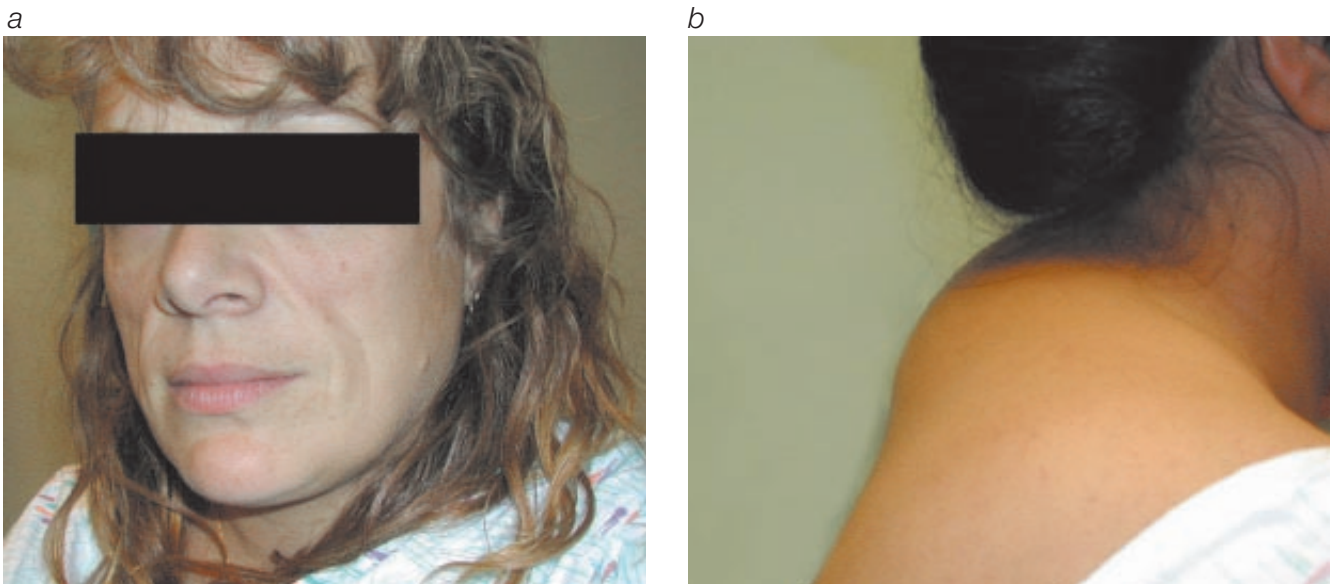


Figure 13 **Typical presentations of the lipodystrophy syndrome include (*a*) wasting of facial fat and (*b*) accumulation of fat in the posterior neck.**

insulin levels, decreased weight, decreased visceral abdominal fat, and decreased subcutaneous abdominal fat.[140] In this study, patients taking metformin did not experience an increase in serum lactate levels or hepatic transaminase levels. Similarly, a study of the insulin sensitizer rosiglitazone showed that this agent can improve insulin sensitivity, reduce free fatty acid levels, and increase subcutaneous leg fat in patients with lipodystrophy.[141] No change was seen in subcutaneous abdominal or visceral fat. The study was small (28 patients) and of short duration (3 months) but was randomized, double blind, and placebo controlled and will hopefully be confirmed by larger studies with longer follow-up. These reports provide further indirect evidence that insulin resistance plays a central role in the pathogenesis of HIV-associated lipodystrophy. Further study is needed to better predict which patients will develop fat-redistribution syndrome and to identify preventive measures.

Bone abnormalities Avascular necrosis (AVN), osteopenia, and osteoporosis are becoming prominent complications of HIV infection and may be linked to antiretroviral therapy. AVN is still an unusual clinical diagnosis, although one study using magnetic resonance imaging scanning documented occult AVN of the femoral head in 5% of HIV-infected patients.[142] Traditional risk factors for AVN include trauma, systemic lupus erythematosus, corticosteroid use, hemoglobinopathies, hypercoagulable states, hyperlipidemia, and alcohol abuse, whereas case-control studies involving HIV-infected patients have demonstrated links only to steroid use and hyperlipidemia.[142-144] HIV-associated AVN occurs more frequently in the hips and knees and is often bilateral, distinguishing it from AVN in HIV-negative persons, which is usually not multifocal. Treatment options for AVN are limited to pain control and, in severe cases, surgery.

Reports of decreased bone mineral density in HIV-infected patients predate the use of potent antiretroviral therapy.[145] However, studies have demonstrated an association between protease inhibitors, osteopenia, and lipodystrophy.[146,147] The pathogenesis of HIV-associated bone abnormalities is unknown. Regular screening for osteopenia in HIV-infected patients is not currently recommended. If patients are discovered to have decreased bone mineral density, treatment should include calcium and vitamin D supplementation. Frank osteoporosis may merit the addition of bisphosphonates, calcitonin, or estrogen receptor modulators.[91]

MONITORING OF RNA LEVELS AND CD4$^+$ T CELL COUNTS

The HIV RNA level should be measured before antiretroviral therapy is started, 2 to 8 weeks after starting therapy, and then at 3 to 4 months after starting treatment to document control of HIV replication. Subsequently, the HIV RNA level and CD4$^+$ T cell count should be monitored approximately every 3 months. In most patients, the HIV RNA level should decrease by at least 1 $\log_{10}$ by week 8 and then to less than 50 copies/ml by weeks 16 to 20. Suppressing HIV levels to less than 50 copies/ml diminishes the number of new HIV virions and thus significantly decreases the chance that resistant strains will emerge.

The CD4$^+$ T cell counts of patients on antiretroviral therapy should be monitored every 3 to 4 months to assess the impact of antiretroviral therapy on immune function. Despite earlier hopes that treatment with an aggressive antiretroviral regimen for 2 to 3 years could eradicate HIV, subsequent studies have shown that HIV eradication is not a realistic goal of antiretroviral therapy. Although potent antiretroviral therapy does have tremendous impact on the total burden of HIV on the body, several cellular and anatomic reservoirs of virus contribute to the long-term persistence of HIV.[148] Resting CD4$^+$ memory T cells that contain latently integrated HIV DNA are the most important cellular reservoirs; the CNS and the male urogenital tract are important anatomic reservoirs. Novel strategies to eliminate the latent HIV reservoirs need to be developed before eradication of HIV becomes a realistic goal.

ADHERENCE TO THERAPY

The degree of adherence to antiretroviral therapy clearly predicts whether or not a patient will sustain long-term viral suppression.[149] Other factors that may affect long-term viral suppression are the inherent potency of the regimen, whether the

patient has previously received antiretroviral therapy, the patient's immune status, absorption of the antiretroviral medications, and drug-drug interactions that may decrease the concentration of the antiretroviral medications. In patients who achieve an HIV RNA level of less than 50 copies/ml, a change in therapy should be considered if there is a significant increase in the HIV RNA level, if there is a persistent decline in the $CD4^+$ T cell count, or if there is evidence of clinical deterioration. The choice of a salvage regimen should take into consideration the patient's prior antiretroviral therapy, resistance testing, the likelihood of adherence with complex salvage regimens, and the stage of the patient's HIV disease. The choice of a salvage regimen should be made in consultation with an expert in antiretroviral therapy.

DRUG RESISTANCE AND RESISTANCE TESTING

Antiretroviral therapy can lead to the emergence of resistant strains of HIV, which can render antiretroviral therapy less effective. Indeed, multiple studies have shown that in cases in which HIV suppression fails, there is usually evidence of HIV drug resistance.

Types of Resistance Assays

Two types of resistance assays are now commercially available to test for HIV resistance: genotype and phenotype assays. The genotype assays identify specific codon mutations associated with resistance, either by amplifying and then sequencing the reverse transcriptase and protease genes of plasma virus or by using standardized probes to detect mutations known to be associated with resistance. The genotype assays can be performed in 1 to 2 weeks, but interpretation of the results requires expertise and review of current and prior antiretroviral therapy.

Phenotype assays measure growth, in the presence of specific antiretroviral medications, of chimeric HIV strains that contain the patient's virus reverse transcriptase and protease genes; the results are reported as the concentration of drug required to inhibit 50% or 90% of viral replication (IC_{50} and IC_{90}, respectively). Chimeric HIV strains are constructed by splicing the amplified reverse transcriptase and protease genes of the patient's virus into a standard HIV backbone that contains an indicator gene, then transfecting these chimeric constructs into cells expressing viral envelope proteins to generate viral stocks. As with genotype results, the phenotype results require expertise to interpret and must be interpreted in the context of current and prior antiretroviral therapy. The strains of HIV identified in the patient are compared with reference strains and reported as the "fold" resistance in IC_{50}. In general, phenotype assays require longer periods than genotype assays and are more expensive.

Both testing techniques have inherent problems: the tests may sample only a minority of the HIV strains; the sampled strains from the blood may not adequately represent strains from other body regions (compartments); resistance to medications taken in the past may not be evident; and for resistance to be determined, plasma HIV RNA levels typically must be in the range of at least 500 to 1,000 copies/ml. Nevertheless, several studies have shown that patients who undergo resistance testing in the setting of virologic failure have better virologic outcomes than those who do not undergo resistance testing.[150,151]

To date, no randomized trials have compared genotypic tests with phenotypic tests. Thus, there is no basis for recommending one over the other. Because the tests will not detect any nonreplicating archived viruses that may contain resistance mutations, they cannot predict with certainty which drugs will work. However, test results indicating the presence of resistance to particular drugs do predict failure if those drugs are used.

Indications for Resistance Testing

The DHHS guidelines recommend resistance testing for patients who experience virologic failure with an antiretroviral regimen, for patients in whom HIV remains at detectable levels after starting therapy, and for patients with acute HIV infection who are considering treatment.[91] Virologic failure is constituted by any one of the following three events: (1) a decrease in HIV RNA of less than 0.75 $\log_{10}$ by week 4 of antiretroviral therapy or of less than 1.0 $\log_{10}$ by week 8; (2) a failure to suppress the level of HIV RNA to less than 50 copies/ml within 6 months of starting or changing therapy; or (3) repeated detection of a high level of HIV RNA (> 5,000 copies/ml) in plasma after initial suppression to less than 50 copies/ml.[91] In addition, some experts consider utilizing resistance testing for those patients with low-level breakthrough viremia (50 to 5,000 copies/ml). The level in such patients must be greater than 1,000 copies/ml to obtain a reliable resistance assay, however. Resistance testing is recommended for patients with acute infection who are considering therapy, because of the known transmission of drug-resistant HIV variants and the demonstration that a significant proportion of persons newly infected with HIV harbor drug-resistant strains.[152]

Drug-resistance testing in chronically infected persons who are not on therapy is controversial. Resistant viruses can persist in the circulation for months or years in the absence of selective drug pressure, so resistance testing may be of value.[153-155] Theoretically, transmitted drug-resistant viruses with mutations that significantly decrease fitness should disappear from the circulating pool (but will be archived) in the absence of drug pressure. However, mutations with limited effects on replication will persist and can be detected by resistance tests. Resistance testing in these individuals may be of benefit, but this has not been formally evaluated in clinical trials. It is reasonable to offer resistance testing to antiretroviral-naive patients who have been infected for less than 2 years and who are contemplating treatment. Resistance testing should be performed for pregnant women starting therapy.[92]

STRUCTURED TREATMENT INTERRUPTIONS

Structured treatment interruptions have been used in an attempt to autoimmunize patients with rebounding HIV viremia in the hope of eliciting better HIV-directed immune responses that will help control HIV replication over the long term. Although this approach held some initial promise in the treatment of acutely infected patients, there is little evidence that it is a viable strategy in chronically infected patients.[156] The desired immunologic responses have not been consistently observed, and patients have not had sufficient CTL responses to permit simplifying or stopping antiretroviral therapy. In addition, stopping and starting antiretroviral therapy in chronically infected patients could have several negative effects. First, drug resistance could develop when the virus rebounds or when the patient resumes therapy. The risk of resistance developing during viral rebound would be of particular concern in persons on regimens that include drugs with long elimination half-lives, such as nevirapine or efavirenz. Second, one study has shown that patients who have long-term viral suppression and quiescent lymph nodes will develop hyperplastic and activated lymph nodes within 1 to 2 months after stopping antiretroviral treatment.[157]

Third, several reports have now described persons with well-controlled HIV who developed an acute retroviral syndrome after stopping antiretroviral therapy.[158,159]

Structured treatment interruption has been suggested for use in patients with uncontrolled HIV replication despite therapy, as a means for allowing repopulation with drug-sensitive virus that would then be killed when treatment was restarted.[160] This tactic, however, is associated with marked reductions in $CD4^+$ T cell counts during the phase of regrowth of wild-type virus, and the archived drug-resistant strains reemerge when treatment is restarted.[161] Consequently, structured treatment interruptions are currently not recommended in patients chronically infected with HIV.

Prophylactic Therapy

ANTIRETROVIRAL THERAPY AFTER OCCUPATIONAL EXPOSURE

Despite widespread efforts to implement universal precautions, health care workers still occasionally suffer accidental exposures to HIV. Current recommendations for postexposure prophylaxis are based on previous retrospective human data, data from animal studies, and extrapolation of experience with antiretroviral therapy in HIV-infected patients. One retrospective case-control study found that use of zidovudine monotherapy for postexposure prophylaxis in health care workers reduced the risk of HIV transmission by approximately 81%.[162] In addition, animal studies have suggested that postexposure prophylaxis decreases the risk of transmission.[43]

Recommendations

A health care worker who has incurred a recognized risk of occupational HIV transmission should first adequately clean the site of exposure and then start postexposure antiretroviral therapy as soon as possible. Combination therapy is recommended for postexposure prophylaxis, because extrapolations from treatment data in HIV-infected patients indicate that two- or three-drug regimens are likely to be more effective than zidovudine alone for preventing transmission of HIV; their superiority is not proven, however. The basic prophylactic regimen contains two drugs; an expanded three-drug regimen is recommended for HIV exposures that pose an increased risk of transmission, such as those involving larger volumes of blood or those in which the source patient has advanced AIDS.

Three basic regimens are recommended by the United States Public Health Service for use in adults: (1) zidovudine (300 mg b.i.d. or 200 mg t.i.d.) plus lamivudine (150 mg b.i.d.); (2) stavudine (40 mg b.i.d.) plus lamivudine (150 mg b.i.d.); and (3) stavudine (40 mg b.i.d.) plus didanosine (400 mg q.d.).[43] For persons who weigh less than 60 kg, the dosage of stavudine should be reduced to 30 mg twice a day, and the dosage of didanosine should be reduced to 300 mg daily. The expanded three-drug regimen consists of a basic regimen plus either nelfinavir (750 mg t.i.d.) or indinavir (800 mg every 8 hours). Newer protease inhibitors, such as lopinavir-ritonavir, atazanavir, or fosamprenavir, are attractive alternatives for use as a third drug. In addition, tenofovir-DF is an excellent option for patients who cannot tolerate zidovudine, especially because it is well tolerated and requires only once-daily administration. Life-threatening hepatotoxicity has been reported in two health care workers who were taking nevirapine for postexposure prophylaxis[163]; accordingly, nevirapine should not be used for this purpose.

The United States Public Health Service guidelines note several special circumstances that warrant consultation with local experts or the National Clinicians' Post-Exposure Prophylaxis Hotline ([PEPline] 1-888-448-4911). These circumstances include cases in which there is a delayed report of exposure; the source person is unknown; the exposed person is pregnant; the source virus is resistant to antiretroviral agents; or the health care worker experiences a toxic reaction to the prophylactic regimen.[43]

Regardless of whether the basic regimen or the expanded regimen is used, it is recommended that the prophylaxis be taken for 28 days. Unfortunately, health care workers frequently have adverse effects with postexposure antiretroviral prophylaxis, even to the extent that many cannot complete their 28-day course.[164]

Health care workers receiving postexposure prophylaxis should undergo drug-toxicity monitoring at baseline and again 2 weeks after therapy is started. In addition, health care workers with occupational exposure to HIV should undergo serologic testing for HIV on first exposure and at 6 weeks, 3 months, and 6 months after exposure. Because three health care providers experienced seroconversion between months 6 and 12 after exposure, some experts recommend also performing follow-up HIV serologic testing at 12 months. In general, health care workers should use safe-sex practices and avoid breast-feeding for at least the first 12 weeks after exposure to HIV.

ANTIRETROVIRAL THERAPY TO PREVENT PERINATAL TRANSMISSION

The Pediatric AIDS Clinical Trials Group Protocol 076 study generated the initial enthusiasm for using antiretroviral therapy to prevent perinatal HIV transmission.[36] In this trial, HIV-infected pregnant women received either placebo or a regimen that consisted of oral zidovudine administered after week 14 of gestation and intravenous zidovudine administered during labor and delivery; in addition, oral zidovudine was given to the newborn until the infant reached 6 weeks of age. The HIV transmission rate in those who received zidovudine was 8.3%, compared with 25.5% in placebo recipients. Longer-term complete results of this study showed transmission rates of 7.6% in the zidovudine group and 22.6% in the placebo group.[165] As perinatal prophylaxis regimens became incorporated into clinical practice, HIV perinatal transmission rates in the United States dramatically declined.[41] Available data have shown no significant differences with regard to congenital abnormalities, growth rates, or neurologic development, nor have the data uncovered other evidence of long-term toxicity in children exposed to zidovudine, either in utero or post partum, compared with children in the general population.[41]

Subsequent international studies have also shown the benefit of antiretroviral therapy in preventing mother-to-child transmission of HIV. In one study in Thailand in which mothers received zidovudine for 4 weeks before and during labor, perinatal transmission was 9% in the zidovudine group and 19% in the placebo group.[166] A study from Uganda found nevirapine to be more effective than zidovudine in the prevention of perinatal HIV transmission.[167] In one arm of the study, a single 200 mg dose of nevirapine was given to the mother at the onset of labor and a single 2 mg/kg dose was given to the infant 48 to 72 hours after birth. In the second arm of the study, a short course of oral zidovudine was given to the mother during labor, and oral zidovudine was given to the infant for 1 week after birth. The HIV infection rate for the newborns at 6 weeks after birth was 12% in

Table 8 Prophylactic Measures Strongly Recommended as Standard of Care for the Prevention of First Episodes of Opportunistic Disease in Adults and Adolescents Infected with HIV[188]

Pathogen	*Indication*	*First Choice**	*Alternatives**
Pneumocystis jiroveci†	$CD4^+$ T cell count < 200/μl *or* Oropharyngeal candidiasis	TMP-SMX, 1 DS q.d. TMP-SMX, 1 SS q.d.	Dapsone, 50 mg b.i.d. or 100 mg q.d. Dapsone, 50 mg q.d. + pyrimethamine, 50 mg q. wk + leucovorin, 25 mg q. wk Dapsone, 200 mg + pyrimethamine, 75 mg + leucovorin, 25 mg q. wk Aerosolized pentamidine, 300 mg q. mo via Respirgard II nebulizer Atovaquone, 1,500 mg q.d. TMP-SMX, 1 DS t.i.w.
Mycobacterium tuberculosis			
Isoniazid-sensitive‡	TST reaction ≥ 5 mm or previous positive TST result without treatment or contact with case of active tuberculosis	Isoniazid, 300 mg + pyridoxine, 50 mg q.d. × 9 mo Isoniazid, 900 mg + pyridoxine, 100 mg b.i.w. × 9 mo	Rifampin, 600 mg q.d. × 4 mo Rifabutin, 300 mg q.d. Pyrazinamide, 15–20 mg/kg q.d. × 2 mo + rifampin, 600 mg q.d. × 2 mo *or* rifabutin, 300 mg q.d. × 2 mo
Isoniazid-resistant	Same; high probability of exposure to isoniazid-resistant tuberculosis	Rifampin, 600 mg × 4 mo Rifabutin, 300 mg q.d. × 4 mo	Pyrazinamide, 15–20 mg/kg q.d. × 2 mo + rifampin, 600 mg q.d. × 2 mo *or* rifabutin, 300 mg q.d. × 2 mo
Multidrug (isoniazid and rifampin)-resistant	Same; high probability of exposure to multidrug-resistant tuberculosis	Choice of drugs requires consultation with public health authorities	None
Toxoplasma gondii§	IgG antibody to *Toxoplasma* and $CD4^+$ T cell count < 100/μl	TMP-SMX, 1 DS q.d.	TMP-SMX, 1 SS q.d. Dapsone, 50 mg q.d. + pyrimethamine, 50 mg q. wk + leucovorin, 25 mg q. wk Dapsone, 200 mg q.d. + pyrimethamine, 75 mg q. wk + leucovorin, 25 mg q. wk Atovaquone, 1,500 mg q.d. ± pyrimethamine, 25 mg q.d. + leucovorin, 10 mg q.d.
Mycobacterium avium complex‖	$CD4^+$ T cell count < 50/μl	Azithromycin, 1,200 mg q. wk Clarithromycin, 500 mg b.i.d.	Rifabutin, 300 mg q.d. Azithromycin, 1,200 mg q. wk + rifabutin, 300 mg q.d.
VZV	Significant exposure to chickenpox or shingles for patients who have no history of either condition or, if available, have negative antibody to VZV	Varicella-zoster immune globulin (VZIG), 5 vials (1.25 ml each) I.M., administered ≤ 96 hr after exposure, ideally within 48 hr	

*All regimens are oral unless indicated otherwise.

†Prophylaxis should also be considered for persons whose $CD4^+$ cells constitute less than 14% of total lymphocytes, for persons with a history of an AIDS-defining illness, and possibly for those with $CD4^+$ T cell counts greater than 200 cells/μl but less than 250 cells/μl. TMP-SMX also reduces the frequency of toxoplasmosis and some bacterial infections. Patients receiving dapsone should be tested for glucose-6-phosphate dehydrogenase deficiency. A dosage of 50 mg q.d. is probably less effective than that of 100 mg q.d. Patients who are being given sulfadiazine-pyrimethamine for toxoplasmosis are protected against *Pneumocystis jiroveci* pneumonia and do not need additional prophylaxis against that illness.

‡It is recommended that patients receiving isoniazid, 900 mg b.i.w., remain under observation during therapy. Isoniazid regimens should include pyridoxine to prevent peripheral neuropathy. Rifampin should not be administered concurrently with protease inhibitors or nonnucleoside reverse transcriptase inhibitors. Rifabutin should not be given with hard-gel saquinavir or delavirdine; caution is also advised when the drug is coadministered with soft-gel saquinavir. Rifabutin may be administered at a reduced dosage (150 mg q.d.) when combined with indinavir, nelfinavir, or amprenavir; at a reduced dosage (150 mg q.o.d. or 150 mg three times weekly) when combined with ritonavir; or at an increased dosage (450 mg q.d.) when combined with efavirenz. Information is lacking regarding coadministration of rifabutin with nevirapine. Exposure to multidrug-resistant tuberculosis might require prophylaxis with two drugs; public health authorities should be consulted. Possible regimens include pyrazinamide plus either ethambutol or a fluoroquinolone.

§Protection against toxoplasmosis is provided by TMP-SMX, dapsone plus pyrimethamine, and possibly atovaquone.

‖As noted in the first footnote above, rifabutin may be administered at a reduced dosage when used in combination with the protease inhibitors indinavir (rifabutin dosage, 150 mg q.d.), nelfinavir (150 mg q.d.), amprenavir (150 mg q.d.), and ritonavir (150 mg q.o.d. or 150 mg three times weekly). It may be administered at an increased dosage with the nonnucleoside reverse transcriptase inhibitor efavirenz (rifabutin dosage, 450 mg q.d.). Information is lacking regarding coadministration of rifabutin with nevirapine.

DS — double-strength tablet SS — single-strength tablet TMP-SMX — trimethoprim-sulfamethoxazole TST — tuberculin skin test VZV — varicella-zoster virus

the nevirapine group and 21% in the zidovudine group.[167] The enthusiasm for these excellent results with short-course nevirapine was tempered by the discovery of high rates of resistance associated with this therapy: in tests conducted 6 weeks after the conclusion of therapy, resistance to nevirapine was found to have developed in 20% of the women.[168] This high rate of resistance after receiving only one dose of nevirapine presumably relates to the very long elimination half-life of this drug. Nevirapine is also associated with severe hypersensitivity reactions involving the skin and liver; these reactions may be more common in pregnant women. These adverse effects have not been reported in women taking only one or two doses of nevirapine to prevent HIV transmission, but they have occurred in women taking nevirapine as part of a potent antiretroviral regimen. The risk of severe hypersensitivity and hepatic failure is increased in persons with $CD4^+$ T cell counts greater than 250 cells/μl and those

with preexisting liver disease.[169,170] Nevirapine should be used with extreme caution in pregnancy; it requires close laboratory and clinical monitoring, and it should be discontinued if the patient shows any clinical or laboratory evidence of hepatitis.

Recommendations

In its recommendations for reducing perinatal HIV transmission, the Public Health Service Task Force stated that pregnancy should influence the decisions regarding timing and choice of antiretroviral therapy but that antiretroviral therapy should not be deferred on the grounds that the patient is pregnant.[41] Several of the unique and important factors to consider with the use of antiretroviral therapy in pregnancy include the physiologic changes associated with pregnancy, which may require altering medication dosages; the potential short-term or long-term adverse effects on the fetus and newborn; and the effectiveness of the regimen in reducing the risk of perinatal transmission. In addition, the efficacy of the regimen for treatment of the mother's HIV infection needs to be taken into consideration. Unfortunately, data regarding combination antiretroviral therapy in pregnancy are insufficient to clarify the safety and efficacy of these regimens. Accordingly, medical providers who offer antiretroviral therapy to HIV-1–infected women during pregnancy need to extensively discuss issues related to the potential short-term and long-term risks and benefits of therapy during pregnancy.

If antiretroviral therapy is given during pregnancy, the three-part zidovudine regimen should be used. In general, clinicians are likely to encounter one of four scenarios: (1) the HIV-infected pregnant woman has not received prior antiretroviral therapy; (2) the HIV-infected woman is receiving antiretroviral therapy at the time the pregnancy is discovered; (3) the HIV-infected woman is in labor and has not been receiving antiretroviral therapy during her pregnancy; or (4) the infant is born of an HIV-infected mother who did not receive antiretroviral therapy during pregnancy or labor. Specific considerations and recommendations for each of these scenarios are discussed in detail in the Public Health Service Task Force recommendations.[41] In addition, these recommendations address issues regarding the mode of delivery, with particular consideration to the performance of cesarean section in women who have HIV RNA levels greater than 1,000 copies/ml near the time of delivery. Women who have HIV RNA levels less than 1,000 copies/ml have a very low rate of HIV transmission, and thus, their child would be unlikely to receive substantial benefit from cesarean section.

VACCINES AGAINST HIV

The development of an effective HIV vaccine has eluded investigators for years. Recently, two large placebo-controlled phase III trials of bivalent recombinant gp120 vaccines (B/B and B/E) demonstrated no benefit from either vaccine.[171,172] Although neutralizing antibodies are produced during natural HIV infection, viral escape develops rapidly because of the high mutation rate of HIV,[173,174] so the failure of these trials is not surprising. Most recent vaccine work has focused on the development of vaccines that elicit potent $CD4^+$ and $CD8^+$ T cell–mediated responses. Theoretically, these vaccines would not prevent infection but might be able to eradicate infection by eliminating HIV-infected cells, or at least they might be able to tightly control the virus and delay or halt disease progression. A variety of vaccines (e.g., killed virus, DNA, viral and bacterial vector) and vaccine strategies (e.g., vaccination, prime-boost with combinations of vaccines) have been tried in animal models and humans, with mixed success.[175-184] Major hurdles are to identify antigens or epitopes that elicit protective responses across a wide variety of HIV types and from which the virus cannot escape. The ideal vaccine would be one that stimulates production of broadly valent neutralizing antibodies that may prevent infection with primary isolates and that elicits potent T cell responses that eliminate or control any infection that is not prevented.[185] Unfortunately, such a vaccine does not currently exist.

Prevention and Management of Opportunistic Infections

With the widespread use of antiretroviral therapy in developed countries, the incidence of serious opportunistic infections has dramatically declined.[3,186] In the United States, the incidence of *Pneumocystis jiroveci* (formerly *P. carinii)* pneumonia, *Mycobacterium avium* complex disease, or cytomegalovirus (CMV) retinitis in HIV-infected patients declined from 21.9 per 100 person-years in 1994 to 3.7 per 100 person-years in mid-1997.[3] In addition, a separate study examined the incidence of the most common opportunistic infections in the United States from 1992 to 1998 and found profound declines during that period; the most pronounced reductions occurred from 1996 to 1998, reflecting the widespread use of aggressive antiretroviral therapy.[186] The incidence of Kaposi sarcoma has also markedly declined, with an estimated 50% reduction occurring in persons receiving potent antiretroviral therapy.[187] Despite these advances, major opportunistic infections continue to cause significant morbidity and mortality in HIV-infected persons. Indeed, prevention of major opportunistic infections remains a critical aspect of HIV care.

PRIMARY PREVENTION OF OPPORTUNISTIC INFECTIONS

The U.S. Public Health Service, in conjunction with the Infectious Diseases Society of America, has issued guidelines for preventing opportunistic infections [*see Table 8*].[188] As the standard of care, these guidelines strongly recommend measures to prevent the first episode of opportunistic disease caused by *M. tuberculosis, P. jiroveci, Toxoplasma gondii,* and *M. avium* complex. Primary prophylaxis against a particular pathogen is recommended when disease from the organism is likely to occur and prophylaxis for that disease is simple, effective, and relatively nontoxic; these four pathogens meet those criteria. The guidelines do not recommend the routine use of prophylaxis to prevent a first episode of *Cryptococcus neoformans, Histoplasma capsulatum,* or CMV infection.

The latest version of the guidelines also addresses discontinuance of primary prophylaxis in patients who have a sustained immunologic response to antiretroviral therapy. The risk of developing *Pneumocystis* pneumonia is extremely small in persons who discontinue primary prophylaxis once they have maintained $CD4^+$ T cell counts of greater than 200 cells/μl for at least 3 months.[189] Similarly, patients can safely discontinue primary prophylaxis for *M. avium* complex when their $CD4^+$ T cell counts have stayed above 100 cells/μl for 3 months or more, and they can discontinue primary prophylaxis for *T. gondii* when their $CD4^+$ T cell counts have stayed above 200 cells/μl for that length of time.[190-192]

MANAGEMENT OF OPPORTUNISTIC INFECTIONS

Tuberculosis

Although tuberculosis can occur in HIV-infected patients regardless of their $CD4^+$ T cell count, the clinical presentation usually varies with the $CD4^+$ T cell count. The clinical picture of

Table 9 Treatment of Acute Opportunistic Infections

Pathogen	*First Choice**	*Alternatives**	*Comment*
Cryptococcal meningitis	Amphotericin B, 0.7–1.0 mg/kg/day I.V. ± flucytosine, 100 mg/kg/day p.o. × 14 days, followed by fluconazole, 400 mg/day for a minimum of 10 wk	Amphotericin B, 0.7–1 mg/kg/day I.V. *or* Fluconazole, 800 mg/day I.V. or p.o. ± flucytosine, 100 mg/kg/day p.o.	Amphotericin is preferred as initial therapy; fluconazole is reasonable as initial therapy only in persons with normal mental status; management of increased intracranial pressure is essential; lumbar punctures should be performed to keep CSF pressure < 200 mm Hg
Cytomegalovirus retinitis	Ganciclovir, 5 mg/kg I.V., b.i.d. × 14–21 days, then 5–6 mg/kg/day I.V. *or* valganciclovir, 900 mg b.i.d. with food × 14–21 days, then 900 mg q.d.	Foscarnet, 60 mg/kg q. 8 hr I.V. × 14–21 days, then 90–120 mg/kg/day Intraocular ganciclovir pellet + ganciclovir, 1,000 mg p.o. t.i.d. Cidofovir, 5 mg/kg I.V. q. wk × 2, then 5 mg/kg every 14 days	Ganciclovir is much easier to administer than foscarnet; intraocular ganciclovir pellets are very effective, but there is risk involved in administering them and they do not provide protection against disease outside the eye
Mycobacterium avium complex	Clarithromycin, 500 mg b.i.d. + ethambutol, 15 mg/kg/day ± rifabutin, 300 mg/day	Azithromycin, 600 mg q.d., can be substituted for clarithromycin Ciprofloxacin, 500 mg b.i.d., can be substituted for rifabutin	Benefit of adding third drug is unclear; clarithromycin and rifabutin can have significant interactions with antiretroviral medications
Mycobacterium tuberculosis	Isoniazid, 300 mg q.d. + rifampin, 600 mg q.d. + pyrazinamide, 25 mg/kg (maximum, 2 g) q.d. + ethambutol, 15–25 mg/kg (maximum, 2.5 g) q.d. × 9 mo	Rifabutin, 150–450 mg q.d., should be substituted for rifampin and the dose adjusted, depending on drug interactions	If the isolate is not drug resistant, then pyrazinamide and ethambutol should be discontinued after 2 mo; if the isolate is drug resistant, the regimen should be chosen in consultation with a tuberculosis expert
Pneumocystis jiroveci	Trimethoprim-sulfamethoxazole, 15–20 mg/kg/day, p.o. or I.V. in three or four divided doses × 21 days	Trimethoprim, 15–20 mg/kg/day in three or four divided doses + dapsone, 100 mg q.d. × 21 days Pentamidine, 3–4 mg/kg I.V. q.d. × 21 days Atovaquone, 750 mg q.d. × 21 days Clindamycin, 450 mg q.i.d. + primaquine, 15 mg base q.d. Trimetrexate, 45 mg/m^2 I.V. q.d. + folinic acid, 20 mg/m^2 q.i.d.	Patients with P_{O_2} < 70 mm Hg or A-a gradient > 35 mm Hg should receive corticosteroids (prednisone, 40 mg b.i.d. × 5 days, then 40 mg q.d. × 5 days, then 20 mg q.d. × 11 days); I.V. pentamidine has been associated with major toxicity
Toxoplasma gondii	Pyrimethamine, 100–200 mg loading dose, then 50–100 mg q.d. + sulfadiazine, 1.5–2.0 g q.i.d. + leucovorin, 10–25 mg q.d.	Clindamycin, 600–900 mg q.i.d., can be substituted for sulfadiazine Other agents that could be substituted for sulfadiazine include azithromycin, 1,200 mg q.d.; clarithromycin, 1.0 g b.i.d.; or atovaquone, 750 mg b.i.d.	After approximately 6–8 wk, patients can change to suppressive regimens that use lower dosages of medications

*All regimens are oral unless noted otherwise.
CSF — cerebrospinal fluid P_{O_2} — oxygen tension

HIV-infected patients with active tuberculosis whose CD4$^+$ T cell count is greater than 350 cells/μl is similar to that of persons not infected with HIV—the illness is characterized by gradual onset of weight loss, cough, fever, upper lobe infiltrates, and cavitary lesions. In contrast, patients with advanced AIDS who develop active tuberculosis often have an atypical presentation that may include diffuse pulmonary lesions, lymph node involvement, blood cultures that are positive for *M. tuberculosis*, and an absence of cavitary lesions. A diagnosis is made by isolating *M. tuberculosis* from any body site; most often, the diagnosis is made on sputum staining or from bronchoscopy. It is critical that drug susceptibility testing be performed on the *M. tuberculosis* isolate.

Initial therapy for *M. tuberculosis* consists of a minimum of three drugs (isoniazid, rifampin, and pyrazinamide); most experts recommend adding ethambutol until susceptibility results become available [*see Table 9*]. Recommendations for the treatment of tuberculosis in HIV-infected patients were published in 2003.[193] These recommendations review the drug interactions between antiretroviral agents and the rifamycins (rifampin and rifabutin), the evidence for and against particular antituberculous regimens in HIV-infected patients, and the importance of directly observed therapy for the treatment of patients who have both HIV infection and tuberculosis.

Pneumocystis *Pneumonia*

Patients with *P. jiroveci* pneumonia generally have a CD4$^+$ T cell count of less than 200 cells/μl; the clinical presentation is characterized by a nonproductive cough of gradual onset, dyspnea, fever, and fatigue. Chest radiographs most often show a diffuse, reticulonodular pattern, predominantly in the perihilar region, but up to 30% of patients have a normal chest radiograph early in the course of their disease. A definitive diagnosis should be established by induced sputum stain or bronchoscopy. For patients with suspected *P. jiroveci* pneumonia, therapy should not be delayed while the patient undergoes a diagnostic workup. First-line therapy consists of trimethoprim-sulfamethoxazole. Patients whose arterial oxygen tension (P_aO_2)

Table 10 USPHS/IDSA Guidelines for the Prevention of Recurrence of Opportunistic Infections in Adults and Adolescents Infected with HIV after Chemotherapy for Acute Disease[188]

Pathogen or Condition	*Indication*	*First Choice*	*Alternatives*
Cryptococcus neoformans	Documented disease	Fluconazole, 200 mg p.o., q.d.	Amphotericin B, 0.6 mg/kg I.V., q.w. to t.i.w. Itraconazole, 200 mg p.o., q.d.
Cytomegalovirus retinitis	Prior end-organ disease	Volganciclovir, 900 mg p.o., q.d. Ganciclovir, 5–6 mg/kg/day I.V., 5–7 days/wk Foscarnet, 90–120 mg/kg I.V. q.d. For retinitis, sustained-release intraocular ganciclovir pellet, q. 6–9 mo + ganciclovir, 1,000 mg t.i.d.	Cidofovir, 5 mg/kg I.V. every 14 days Fomiversen, 1 vial (330 μg) injected into the vitreous, then repeated every 2–4 wk
Pneumocystis jiroveci	Prior *Pneumocystis jiroveci* pneumonia	TMP-SMX, 1 DS p.o., q.d. TMP-SMX, 1 SS p.o., q.d.	Dapsone, 50 mg b.i.d. or 100 mg q.d. Dapsone, 50 mg q.d. + pyrimethamine, 50 mg q.w. + leucovorin, 25 mg q.w. Dapsone, 200 mg + pyrimethamine, 75 mg + leucovorin 25 mg q.w. Aerosolized pentamidine, 300 mg q. mo via Respirgard II nebulizer Atovaquone, 1,500 mg q.d. TMP-SMX, 1 DS t.i.w.
Toxoplasma gondii	Prior *Toxoplasma* encephalitis	Pyrimethamine, 25–75 mg q.d. + sulfadiazine, 0.5–1.0 g q.i.d. + leucovorin, 10–25 mg q.d.	Pyrimethamine, 25–75 mg q.d. + clindamycin 300–450 mg q.i.d. + leucovorin, 10–25 mg q.d. Atovaquone, 750 mg b.i.d. to q.i.d. ± pyrimethamine, 25 mg q.d. + leucovorin, 10 mg q.d.

DS — double-strength tablet IDSA — Infectious Diseases Society of America SS — single-strength tablet TMP-SMX — trimethoprim-sulfamethoxazole
USPHS — U.S. Public Health Service

measurement is less than 70 mm Hg or whose alveolar-arterial difference in oxygen (A-aDO_2) is greater than 35 mm Hg should also receive corticosteroids as part of their initial therapy.

CNS Toxoplasmosis

CNS toxoplasmosis typically develops only in patients who have a $CD4^+$ T cell count of less than 100 cells/μl; it is characterized by fever, headache, altered thought processes, and, in some instances, seizures. The diagnosis is strengthened by serologic evidence of *Toxoplasma* and findings of multiple enhancing lesions on either computed tomography or MRI. At 14 days, nearly all patients with *Toxoplasma* encephalitis show some radiologic improvement in response to therapy. Brain biopsy is rarely performed at the onset of therapy but may be considered in patients who do not show a clinical and radiographic response to therapy.

Cryptococcal Meningitis

Cryptococcal meningitis most often occurs in patients with a $CD4^+$ T cell count of less than 100 cells/μl; the clinical presentation may be similar to that of *Toxoplasma* encephalitis. Fewer than 25% of patients with cryptococcal meningitis have overt symptoms of meningeal irritation, such as photophobia and neck stiffness. More typically, patients have fever, personality changes, and diminished memory. More than 98% of patients with cryptococcal meningitis have a positive result on serum cryptococcal antigen testing; the definitive diagnosis of cryptococcal meningitis is made by isolating *Cryptococcus neoformans* from the cerebrospinal fluid. Current guidelines recommend amphotericin B plus flucytosine as initial therapy for the first 14 days, followed by high-dose fluconazole for a minimum of 10 weeks, followed by lower-dose fluconazole.[194] A pretreatment opening CSF pressure of 250 mm H_2O or greater is generally as-

Table 11 Criteria for Discontinuance of Secondary Prophylaxis for Opportunistic Infections

Disease	*Disease-Specific Treatment Response*	*$CD4^+$ T Cell Counts*
Pneumocystis jiroveci pneumonia	Completion of initial therapy and absence of signs or symptoms of active infection	> 200 cells/μl for at least 6 mo
Mycoplasma avium complex	Absence of symptoms suggestive of ongoing infection after 12 mo of treatment	≥ 100 cells/μl for at least 6 mo
Cytomegalovirus retinitis*	Quiescent disease with treatment	≥ 100–150 cells/μl for at least 6 mo
Toxoplasma gondii encephalitis	Completion of initial therapy and absence of signs or symptoms of active infection	> 200 cells/μl for at least 6 mo
Cryptococcal meningitis	Completion of initial therapy and absence of signs or symptoms of active infection	> 100–200 cells/μl for at least 6 mo

*In patients with extensive or sight-threatening lesions from cytomegalovirus (CMV) infection, the decision to discontinue anti-CMV treatment should be made with an ophthalmologist, because of the risk of vision loss should reactivation of CMV occur.

Internet Resources for Information on HIV and AIDS

The AIDS Education and Training Centers (AETC) National Resource Center (NRC)
http://www.aids-ed.org

AIDS info
http://www.aidsinfo.nih.gov

Infectious Diseases Society of America Primary Care Guidelines
http://www.journals.uchicago.edu/CID/journal/issues/v39n5/34135/34135.html

International AIDS Society–USA
http://www.iasusa.org/pub/index.html

Johns Hopkins AIDS Service
http://www.hopkins-aids.edu

UCSF: HIV InSite
http://hivinsite.ucsf.edu

The Immunodeficiency Clinic, Toronto General Hospital
http://www.tthhivclinic.com

WebMD HIV/AIDS Health Center
http://www.my.webmd.com/medical_information/condition_centers/hiv_aids/default.htm

The 12th Conference on Retroviruses and Opportunistic Infections
http://www.retroconference.org

The Body
http://www.thebody.com

sociated with a higher titer of cerebrospinal cryptococcal antigen, increased abnormal focal neurologic findings, and decreased survival.[195] In addition, outcomes are worse in patients whose CSF pressure increases by at least 10 mm H_2O after therapy is started. Accordingly, CSF opening pressure should be measured in all patients with suspected cryptococcal meningitis; those patients whose CSF opening pressure is 250 mm H_2O or greater should undergo large-volume CSF drainage by lumbar puncture. Such drainage should be repeated as often as necessary.

Disseminated Mycobacterium avium *Complex Infection*

Disseminated *M. avium* complex infection is a frequent complication in patients who have a $CD4^+$ T cell count of less than 50 cells/μl. This disease is characterized by vague constitutional symptoms, such as fever, night sweats, fatigue, weight loss, abdominal pain, and diarrhea. The preferred method of diagnosis is blood culture, which has a sensitivity of at least 90%; detection by culture, however, often takes longer than 7 days. Empirical therapy should generally be discouraged. Therapy for disseminated *M. avium* complex infection consists of clarithromycin (or azithromycin) plus ethambutol; many experts recommend adding either a fluoroquinolone or rifabutin as a third medication, particularly for the first 2 to 3 months of initial therapy.

Cytomegalovirus Retinitis

CMV infection is another common opportunistic infection in patients who have a $CD4^+$ T cell count of less than 50 cells/μl. CMV most often causes a retinitis that generates visual symptoms such as floaters, flashing lights, or visual-field deficits. The diagnosis is established by characteristic findings on ophthalmologic examination; in general, the patient should be referred to an ophthalmologist for diagnosis. Prompt diagnosis and prompt initiation of therapy are essential to minimize permanent visual loss. Initial therapy should consist of an induction course of either intravenous ganciclovir or intravenous foscarnet, followed by long-term maintenance therapy. Most clinicians prefer ganciclovir as initial therapy because it has fewer adverse effects than foscarnet. Newer therapies include long-term suppression with oral ganciclovir, oral valganciclovir (a ganciclovir prodrug), and implantation of ganciclovir pellets in the affected eye. CMV infection can also occur in the upper or lower gastrointestinal tract, as well as in the CNS.

Immune Reconstitution Syndromes

Immune reconstitution syndromes are the clinical manifestation of inflammatory responses directed against known or occult pathogens that develop after patients are started on antiretroviral therapy. These reactions coincide with improvements in immune function as measured by rising $CD4^+$ T cell counts and have been reported to occur with opportunistic and nonopportunistic infections, as well as malignancies.[196] These inflammatory syndromes often involve mycobacterial infections (*M. avium* complex and *M. tuberculosis*) and typically develop within weeks after initiating antiretroviral treatment. The clinical presentations are often atypical and present diagnostic challenges; other possible causes include adverse drug reactions and progressive disease from resistant pathogens or cancers. Reported findings with *M. avium* complex infection include fever, adenitis, organomegaly, and granulomatous masses. Immune reconstitution syndromes with *M. tuberculosis* may present as fever, new lung infiltrates, pleural and pericardial effusions, tuberculomas, and skin findings.[197] Other reported syndromes include the following: (1) vitreitis with CMV infection, (2) neurologic deficits and enhancing MRI lesions with progressive multifocal leukoencephalopathy, (3) inflammatory meningitis with *Cryptococcus,* (4) hepatitis flares with hepatitis B and C infections, (5) adenitis with histoplasmosis, and (6) inflammatory reactions to Kaposi sarcoma. In most cases, treatment of the involved pathogen and antiretroviral therapy can be continued, and the inflammatory response can be managed with anti-inflammatory medications, including steroids. In severe and life-threatening cases, antiretroviral therapy should be discontinued.

SECONDARY PREVENTION OF OPPORTUNISTIC INFECTIONS

The U.S. Public Health Service guidelines include recommendations for preventing recurrent disease after patients have received therapy for acute disease [*see Table 10*]. As with primary prophylaxis, discontinuance of secondary prophylaxis for a number of opportunistic infections is safe for patients who have completed a set length of therapy for the infection and who have had a robust immunologic response to antiretroviral therapy.

HIV/AIDS Telephone Consultation Services

WARMLINE

National HIV Telephone Consultation Service

Available from 7:30 AM to 5:00 PM PST; detailed information on the service can be obtained at the following URL: http://www.ucsf.edu/hivcntr

Telephone: 1-800-933-3413

PEPline

National Clinicians' Post-Exposure Prophylaxis Hotline

Available 24 hrs; detailed information on the service can be obtained at the following URL: http://www.ucsf.edu/hivcntr/PEPline/index.html

Telephone: 1-888-448-4911

Specific criteria have been determined for *P. jiroveci* pneumonia,[198] *M. avium* complex infection,[199,200] CMV retinitis,[201-203] *T. gondii* encephalitis,[204,205] and cryptococcal meningitis[206] [*see Table 11*]. Discontinuance of secondary prophylaxis (or maintenance therapy) for other opportunistic infections may be possible, but there is too little evidence to permit recommendations at this time.

Additional Resources

Because the field of HIV and AIDS continues to change at a very rapid pace, clinicians need easy access to updated material. The resources available in the current electronic era are outstanding, but they are often overwhelming. Several excellent Web sites that provide current information on treatment guidelines, conference summaries, drug-safety warnings, and summaries of articles are listed [*see Sidebar* Internet Resources for Information on HIV and AIDS]. In addition, national consultation services provide expert clinical advice and state-of-the-art information [*see Sidebar* HIV/AIDS Telephone Consultation Services]. These Internet and telephone resources can provide critical information to help practicing physicians optimally manage HIV infection.

Robert D. Harrington, M.D., has received honoraria from Glaxo Smithkline, Bristol-Meyers Squibb, Roche Pharmaceuticals, Gilead Sciences, and Merck & Co, Inc.

David H. Spach, M.D., has received honoraria from Agouron Pharmaceuticals, Inc., Bristol-Myers Squibb, Gilead Sciences, Glaxo SmithKline, Inc., Merck & Co., Inc., and Roche Pharmaceuticals.

References

1. The 3 by 5 Initiative. World Health Organization, 2004 http://www.who.int/3by5/en/

2. National Center for HIV, STD and TB Prevention: HIV/AIDS Surveillance Report. Vol. 14. Centers for Disease Control and Prevention, Atlanta, 2002

3. Palella FJ, Delaney KM, Moorman AC, et al: Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. N Engl J Med 338:853, 1998

4. Stebbing J, Gazzard B, Douek DC: Where does HIV live? N Engl J Med 350:1872, 2004

5. de Groot NG, Otting N, Doxiadis GG, et al: Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci USA 99:11748, 2002

6. Gao F, Bailes E, Robertson DL, et al: Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes.* Nature 397:436, 1999

7. Hirsch VM, Olmsted RA, Murphey-Corb M, et al: An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339:389, 1989

8. Zhu T, Korber BT, Nahmias AJ, et al: An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391:594, 1998

9. Korber B, Muldoon M, Theiler J, et al: Timing the ancestor of the HIV-1 pandemic strains. Science 288:1789, 2000

10. Dickson D: Tests fail to support claims for origin of AIDS in polio vaccine. Nature 407:117, 2000

11. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm Rep 41(RR-17):1, 1992

12. 2004 report on the global AIDS epidemic. Joint United Nations Programme on HIV/AIDS (UNAIDS). Geneva, 2004 http://www.unaids.org/bangkok2004/report.html

13. Cases of HIV infection and AIDS in the United States, 2002 (HIV/AIDS Surveillance Report, Vol.14). Centers for Disease Control and Prevention, Atlanta, 2003 http://www.cdc.gov/hiv/stats/hasr1402.htm

14. Primary and secondary syphilis—United States, 2002. Centers for Disease Control and Prevention. MMWR Morb Mortal Wkly Rep 52:1117, 2003

15. Royce RA, Sena A, Cates W, et al: Sexual transmission of HIV. N Engl J Med 336:1072, 1997

16. Quinn TC, Wawer MJ, Sewankambo N, et al: Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 342:921, 2000

17. Pilcher CD, Eron JJ Jr, Vemazza PL, et al: Sexual transmission during the incubation period of primary HIV infection (letter). JAMA 286:1713, 2001

18. Yerly S, Vora S, Rizzardi P, et al: Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. Swiss HIV Cohort Study. AIDS 15:2287, 2001

19. Koopman JS, Jacquez JA, Welch GW, et al: The role of early HIV infection in the spread of HIV through populations. J Acquir Immune Defic Syndr Hum Retrovirol 14:249, 1997

20. de Vincenzi I: A longitudinal study of human immunodeficiency virus transmission by heterosexual partners. European Study Group on Heterosexual Transmission of HIV. N Engl J Med 331:341, 1994

21. Lifson AR, O'Malley PM, Hessol NA, et al: HIV seroconversion in two homosexual men after receptive oral intercourse with ejaculation: implications for counseling concerning safe sexual practices. Am J Public Health 80:1509, 1990

22. Schacker T, Collier AC, Hughes J, et al: Clinical and epidemiologic features of primary HIV infection. Ann Intern Med 125:257, 1996

23. Frankel SS, Wenig BM, Burke AP, et al: Replication of HIV-1 in dendritic cell–derived syncytia at the mucosal surface of the adenoid. Science 272:115, 1996

24. Mehendale SM, Rodrigues JJ, Brookmeyer RS, et al: Incidence and predictors of human immunodeficiency virus type 1 seroconversion in patients attending sexually transmitted disease clinics in India. J Infect Dis 172:1486, 1995

25. Gray RH, Wawer MJ, Sewankambo NK, et al: Relative risks and population attributable fraction of incident HIV associated with symptoms of sexually transmitted diseases and treatable symptomatic sexually transmitted diseases in Rakai District, Uganda. Rakai Project Team. AIDS 13:2113, 1999

26. Laga M: Interactions between STDs and HIV infection. STD Bull 13:3, 1992

27. Wald A, Link K: Risk of human immunodeficiency virus infection in herpes simplex virus type 2–seropositive persons: a meta-analysis. J Infect Dis 185:45, 2002

28. Reynolds SJ, Risbud AR, Shepher ME, et al: Recent herpes simplex virus type 2 infection and the risk of human immunodeficiency virus type 1 acquisition in India. J Infect Dis 187:1513, 2003

29. Corey L, Wald A, Celum CL, et al: The effects of herpes simplex virus–2 on HIV-1 acquisition and transmission: a review of two overlapping epidemics. J Acquir Immune Defic Syndr 35:435, 2004

30. Serwaddda D, Gray RH, Sewankambo NK, et al: Human immunodeficiency virus acquisition associated with genital ulcer disease and herpes simplex virus type 2 infection: a nested case-control study in Rakai, Uganda. J Infect Dis 188:1492, 2003

31. Schacker T, Zeh J, Hu H, et al: Changes in plasma human immunodeficiency virus type 1 RNA associated with herpes simplex virus reactivation and suppression. J Infect Dis 186:1718, 2002

32. Schacker T, Ryncarz AJ, Goddard J, et al: Frequent recovery of HIV-1 from genital herpes simplex virus lesions in HIV-1–infected men. JAMA 280:61, 1998

33. Vlahov D, Junge B: The role of needle exchange programs in HIV prevention. Public Health Rep 113(suppl 1):75, 1998

34. Lackritz EM, Satten GA, Aberle-Grasse J, et al: Estimated risk of transmission of the human immunodeficiency virus by screened blood in the United States. N Engl J Med 333:1721, 1995

35. Katzenstein TL, Dickmeiss E, Aladdin H, et al: Failure to develop HIV infection after receipt of HIV-contaminated blood and postexposure prophylaxis. Ann Intern Med 133:31, 2000

36. Connor EM, Sperling RS, Gelber R, et al: Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N Engl J Med 331:1173, 1994

37. Garcia PM, Kalish LA, Pitt J, et al: Maternal levels of plasma human immunodeficiency virus type 1 RNA and the risk of perinatal transmission. Women and Infants Transmission Study Group. N Engl J Med 341:394, 1999

38. Mofenson LM, Lambert JS, Stiehm ER, et al: Risk factors for perinatal transmission of human immunodeficiency virus type 1 in women treated with zidovudine. Pediatric AIDS Clinical Trials Group Study 185 Team. N Engl J Med 341:385, 1999

39. The mode of delivery and the risk of vertical transmission of human immunodeficiency virus type 1: a meta-analysis of 15 prospective cohort studies. The International Perinatal HIV Group. N Engl J Med 340:977, 1999

40. John GC, Nduati RW, Mbori-Ngacha DA, et al: Correlates of mother-to-child human immunodeficiency virus type 1 (HIV-1) transmission: association with maternal plasma HIV-1 RNA load, genital HIV-1 DNA shedding, and breast infections. J Infect Dis 183:206, 2001

41. U.S. Public Health Service: Public Health Service Task Force Recommendations for Use of Antiretroviral Drugs in Pregnant HIV-1–Infected Women for Maternal Health and Interventions to Reduce Perinatal HIV-1 Transmission in the United States. Public Health Service and the Perinatal HIV Guidelines Working Group. U.S. Department of Health and Human Services, Hyattsville, Maryland, May 2001 http://www.hivatis.org/trtgdlns.html#Perinatal

42. Embree JE, Njenga S, Datta P, et al: Risk factors for postnatal mother-child transmission of HIV-1. AIDS 14:2535, 2000

43. Updated U.S. Public Health Service Guidelines for the Management of Health-Care Personnel (HCP) Who Have Occupational Exposure to HBV, HCV, and HIV and Recommendations for Postexposure Prophylaxis. Centers for Disease Control and Prevention. MMWR Recomm Rep 50(RR-11):1, 2001

44. Greene WC: AIDS and the immune system. Sci Am 269:98, 1993

45. Streicher HZ, Reitz MS, Gallo RC: Human immunodeficiency viruses. Principles and Practice of Infectious Diseases. Mandell GL, Bennett JE, Dolin R, Eds. Churchill Livingstone, Philadelphia, 2000, p 1874

46. Mangeat B, Turelli P, Caron G, et al: Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 424: 99, 2003

47. Emerman M, Malim MH: HIV-1 regulatory/accessory genes: keys to unraveling viral and host cell biology. Science 280:1880, 1998

48. Sharp P, Robertson D, Gao F, et al: Origins and diversity of human immunodeficiency viruses. AIDS 265:1193, 1994

49. Kinter A, Arthos J, Cicala C, et al: Chemokines, cytokines and HIV: a complex network of interactions that influence HIV pathogenesis. Immunol Rev 177:88, 2000

50. O'Brien SJ, Moore JP: The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. Immunol Rev 177:99, 2000

51. Meng G, Wei X, Wu X, et al: Primary intestinal epithelial cells selectively transfer R5 HIV-1 to CCR5+ cells. Nat Med 8:150, 2002

52. Pope M, Haase AT: Transmission, acute HIV-1 infection and the quest for strategies to prevent infection. Nat Med 9:847, 2003

53. Geijtenbeek TB, Kwon DS, Torensma R, et al: DC-SIGN, a dendritic cell-specific HIV-1–binding protein that enhances trans-infection of T cells. Cell 100:587, 2000

54. Turville SG, Cameron PU, Handley A, et al: Diversity of receptors binding HIV on dendritic cell subsets. Nat Immunol 3:975, 2002

55. Frank I, Pope M: The enigma of dendritic cell–immunodeficiency virus interplay. Curr Mol Med 2:229, 2002

56. Lee B, Leslie G, Soilleux E, et al: cis expression of DC-SIGN allows for more efficient entry of human and simian immunodeficiency viruses via CD4 and a coreceptor. J Virol 75:12028, 2001

57. McDonald D, Wu L, Bohks SM, et al: Recruitment of HIV and its receptors to dendritic cell–T cell junctions. Science 300:1295, 2003

58. Schacker T, Little S, Connick E, et al: Rapid accumulation of human immunodeficiency virus (HIV) in lymphatic tissue reservoirs during acute and early HIV infection: implications for timing of antiretroviral therapy. J Infect Dis 181:354, 2000

59. Koup RA, Safrit JT, Cao Y, et al: Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. J Virol 68:4650, 1994

60. Douek DC, Brenchley JM, Betts MR, et al: HIV preferentially infects HIV-specific $CD4^+$ T cells. Nature 417:95, 2002

61. Siliciano JD, Kajdas J, Finzi D, et al: Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting $CD4^+$ T cells. Nat Med 9:727, 2003

62. Chun TW, Stuyver L, Mizell SB, et al: Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci USA 94:13193, 1997

63. Finzi D, Hermankova M, Pierson T, et al: Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278:1295, 1997

64. Wong JK, Hezareh M, Gunthard HF, et al: Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278:1291, 1997

65. Mellors JW, Rinaldo CR, Gupta P, et al: Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 272:1167, 1996

66. Walker BD, Goulder PJ: AIDS: escape from the immune system. Nature 407:313, 2000

67. Hogan CM, Hammer SM: Host determinants in HIV infection and disease. Part 1: cellular and humoral immune responses. Ann Intern Med 134:761, 2001

68. Hogan CM, Hammer SM: Host determinants in HIV infection and disease. Part 2: genetic factors and implications for antiretroviral therapeutics. Ann Intern Med 134:978, 2001

69. Altfeld M, Addo MM, Rosenberg ES, et al: Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS 17:2581, 2003

70. Koot M, Keet IP, Vos AH, et al: Prognostic value of HIV-1 syncytium-inducing phenotype for rate of $CD4^+$ cell depletion and progression to AIDS. Ann Intern Med 118:681, 1993

71. Alexander L, Weiskopf E, Greenough TC, et al: Unusual polymorphisms in human immunodeficiency virus type 1 associated with nonprogressive infection. J Virol 74:4361, 2000

72. Learmont JC, Geczy AF, Mills J, et al: Immunologic and virologic status after 14 to 18 years of infection with an attenuated strain of HIV-1: a report from the Sydney Blood Bank Cohort. N Engl J Med 340:1715, 1999

73. Williams CF, Klinzman D, Yamashita TE, et al: Persistent GB virus C infection and survival in HIV-infected men. N Engl J Med 350:981, 2004

74. Pomerantz RJ, Nunnari G: HIV and GB virus C: can two viruses be better than one? N Engl J Med 350:963, 2004

75. Kahn JO, Walker BD: Acute human immunodeficiency virus type 1 infection. N Engl J Med 339:33, 1998

76. Niu MT, Jermano JA, Reichelderfer P, et al: Summary of the National Institutes of Health workshop on primary human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 9:913, 1993

77. Busch MP, Lee LL, Satten GA, et al: Time course of detection of viral and serologic markers preceding human immunodeficiency virus type 1 seroconversion: implications for screening of blood and tissue donors. Transfusion 35:91, 1995

78. Daar ES, Little S, Pitt J, et al: Diagnosis of primary HIV-1 infection. Los Angeles County Primary HIV Infection Recruitment Network. Ann Intern Med 134:25, 2001

79. Hecht FM, Busch MP, Rawal B, et al: Use of laboratory tests and clinical symptoms for identification of primary HIV infection. AIDS 16:1119, 2002

80. Kinloch-De Loes S, Hirschel BJ, Hoen B, et al: A controlled trial of zidovudine in primary human immunodeficiency virus infection. N Engl J Med 333:408, 1995

81. Altfeld M, Rosenberg ES, Shankarappa R, et al: Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. J Exp Med 193:169, 2001

82. Oxenius A, Price DA, Easterbrook PJ, et al: Early highly active antiretroviral therapy for acute HIV-1 infection preserves immune function of $CD8^+$ and $CD4^+$ T lymphocytes. Proc Natl Acad Sci USA 97:3382, 2000

83. Rosenberg ES, Altfeld M, Poon SH, et al: Immune control of HIV-1 after early treatment of acute infection. Nature 407:523, 2000

84. Malhotra U, Berrey MM, Huang Y, et al: Effect of combination antiretroviral therapy on T-cell immunity in acute human immunodeficiency virus type 1 infection. J Infect Dis 181:121, 2000

85. Berrey MM, Schacker T, Collier AC, et al: Treatment of primary human immunodeficiency virus type 1 infection with potent antiretroviral therapy reduces frequency of rapid progression to AIDS. J Infect Dis 183:1466, 2001

86. Guidelines for the use of antiretroviral agents in HIV-1–infected adults and adolescents. Panel on Clinical Practices for Treatment of HIV Infection. Department of Health and Human Services, Washington DC, March 23, 2004
http://aidsinfo.nih.gov/guidelines/adult/AA_032304.html

87. Yang OO, Kalams SA, Trocha A, et al: Suppression of human immunodeficiency virus type 1 replication by $CD8^+$ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms. J Virol 71:3120, 1997

88. Kalams SA, Goulder PJ, Shea AK, et al: Levels of human immunodeficiency virus type 1–specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy. J Virol 73:6721, 1999

89. Lisziewicz J, Rosenberg E, Lieberman J, et al: Control of HIV despite the discontinuation of antiretroviral therapy. N Engl J Med 340:1683, 1999

90. Rosenberg ES, Altfeld M, Poon SH, et al: Immune control of HIV-1 after early treatment of acute infection. Nature 407:523, 2000

91. Kaufman D, Lichterfeld, Altfeld M., et al: Limited durability of immune control following acute HIV infection. Oral abstract 24. 11th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2004
http://www.retroconference.org/2004/cd/Abstract/24.htm

92. Aberg JA, Gallant JE, Anderson J, et al: Primary care guidelines for the management of persons infected with human immunodeficiency virus: recommendations of the HIV Medicine Association of the Infectious Diseases Society of America. Clin Infect Dis 39:609, 2004

93. Incorporating HIV prevention into the medical care of persons living with HIV: recommendations of CDC, the Health Resources and Services Administration, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Recomm Rep 52(RR-12):1, 2003

94. Paredes R, Puig T, Arno A, et al: High-dose saquinavir plus ritonavir: long-term efficacy in HIV-positive protease inhibitor–experienced patients and predictors of virologic response. J Acquir Immune Defic Syndr 22:132, 1999

95. Carpenter CC, Cooper DA, Fischl MA, et al: Antiretroviral therapy in adults: updated recommendations of the International AIDS Society—USA Panel. JAMA 283:381, 2000

96. Clotet B, Raffi F, Cooper D, et al: Clinical management of treatment-experienced, HIV-infected patients with the fusion inhibitor enfuvirtide: consensus recommendations. AIDS 18:1137, 2004

97. Carr A, Cooper DA: Adverse effects of antiretroviral therapy. Lancet 356:1423, 2000

98. Schambelan M, Benson CA, Carr A, et al: Management of metabolic complications associated with antiretroviral therapy for HIV-1 infection: recommendations of an International AIDS Society–USA panel. International AIDS Society–USA panel. J Acquir Immune Defic Syndr 31:257, 2002

99. Mallal S, Nolan D, Witt C, et al: Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet 359:727, 2002

100. Carr A, Yarchoan R: AIDS 2000. Clinical treatment: overview. AIDS 14(suppl 3):S187 2000

101. Boubaker K, Flepp M, Soudre P, et al: Hyperlactatemia and antiretroviral therapy: the Swiss HIV Cohort Study. Clin Infect Dis 33:1931, 2001

102. Mokrzycki MH, Harris C, May H, et al: Lactic acidosis associated with stavudine administration: a report of five cases. Clin Infect Dis 30:198, 2000

103. Lonergan JT, Havlir D, Barber E, et al: Incidence and outcome of hyperlactatemia associated with clinical manifestations in HIV-infected adults receiving NRTI-containing regimens. Oral abstract 624. 8th Conference on Retroviruses and Opportunistic Infections (CROI), Chicago, 2001
http://www.retroconference.org//2001/abstracts/abstracts/abstracts/624.htm

104. Walli R, Herfort O, Michl GM, et al: Treatment with protease inhibitors associated with peripheral insulin resistance and impaired oral glucose tolerance in HIV-1–infected patients. AIDS 12:F167, 1998

105. Hadigan C, Meigs JB, Corcoran C, et al: Metabolic abnormalities and cardiovascular disease risk factors in adults with human immunodeficiency virus infection and lipodystrophy. Clin Infect Dis 32:130, 2001

106. Mynarcik DC, McNurlan MA, Steigbigel RT, et al: Association of severe insulin resistance with both loss of limb fat and elevated serum tumor necrosis factor receptor levels in HIV lipodystrophy. J Acquir Immune Defic Syndr 25:312, 2000

107. Tsiodras S, Mantzoros C, Hammer S, et al: Effects of protease inhibitors on hyperglycemia, hyperlipidemia, and lipodystrophy: a 5-year cohort study. Arch Intern Med 160:2050, 2000

108. Caron M, Auclair M, Vigouroux C, et al: The HIV protease inhibitor indinavir impairs sterol regulatory element-binding protein-1 intranuclear localization, inhibits preadipocyte differentiation, and induces insulin resistance. Diabetes 50:1378, 2001

109. Matinez E, Conget I, Lozano L, et al: Reversion of metabolic abnormalities after switching from HIV-1 protease inhibitors to nevirapine. AIDS 13:805, 1999

110. Martinez E, Garcia-Viejo MA, Blanco JL, et al: Impact of switching from human immunodeficiency virus type 1 protease inhibitors to efavirenz in successfully treated adults with lipodystrophy. Clin Infect Dis 31:1266, 2000

111. Walli RK, Michl GM, Bogner JR, et al: Improvement of HAART-associated insulin

resistance and dyslipidemia after replacement of protease inhibitors with abacavir. Eur J Med Res 6:413, 2001

112. Grunfeld C, Pang M, Doerrler W, et al: Lipids, lipoproteins, triglyceride clearance, and cytokines in human immunodeficiency virus infection and the acquired immunodeficiency syndrome. J Clin Endocrinol Metab 74:1045, 1992

113. Constans J, Pellegrin JL, Peuchant E, et al: Plasma lipids in HIV-infected patients: a prospective study in 95 patients. Eur J Clin Invest 24:416, 1994

114. Feingold KR, Krauss RM, Pang M, et al: The hypertriglyceridemia of acquired immunodeficiency syndrome is associated with an increased prevalence of low density lipoprotein subclass pattern B. J Clin Endocrinol Metab 76:1423, 1993

115. Carr A, Samaras K, Burton S, et al: A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors. AIDS 12:F51, 1998

116. Carr A, Samaras K, Thorisdottir A, et al: Diagnosis, prediction, and natural course of HIV-1 protease-inhibitor–associated lipodystrophy, hyperlipidaemia, and diabetes mellitus: a cohort study. Lancet 353:2093, 1999

117. Periard D, Telenti A, Sudre P, et al: Atherogenic dyslipidemia in HIV-infected individuals treated with protease inhibitors. The Swiss HIV Cohort Study. Circulation 100:700, 1999

118. Mulligan K, Grunfeld C, Tai VW, et al: Hyperlipidemia and insulin resistance are induced by protease inhibitors independent of changes in body composition in patients with HIV infection. J Acquir Immune Defic Syndr 23:35, 2000

119. Raffi F, Bonnet B, Ferre V, et al: Substitution of a nonnucleoside reverse transcriptase inhibitor for a protease inhibitor in the treatment of patients with undetectable plasma human immunodeficiency virus type 1 RNA. Clin Infect Dis 31:1274, 2000

120. Ruiz L, Negredo E, Domingo P, et al: Antiretroviral treatment simplification with nevirapine in protease inhibitor–experienced patients with HIV-associated lipodystrophy: 1-year prospective follow-up of a multicenter, randomized, controlled study. J Acquir Immune Defic Syndr 27:229, 2001

121. Gallant JE, Staszewski S, Pozniak AL, et al: Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial. JAMA 292:191, 2004

122. Bozzette SA, Ake CF, Tam HK, et al: Cardiovascular and cerebrovascular events in patients treated for human immunodeficiency virus infection. N Engl J Med 348:702, 2003

123. Friis-Møller N, Sabin CA, Weber R, et al: Combination antiretroviral therapy and the risk of myocardial infarction. N Engl J Med 349:1993, 2003

124. Dube MP, Sprecher D, Henry WK, et al: Preliminary guidelines for the evaluation and management of dyslipidemia in adults infected with human immunodeficiency virus and receiving antiretroviral therapy: recommendations of the Adult AIDS Clinical Trial Group Cardiovascular Disease Focus Group. Clin Infect Dis 31:1216, 2000

125. Henry K, Melroe H, Huebesch J, et al: Atorvastatin and gemfibrozil for protease-inhibitor–related lipid abnormalities. Lancet 352:1031, 1998

126. Carr A, Miller J, Law M, et al: A syndrome of lipoatrophy, lactic acidaemia and liver dysfunction associated with HIV nucleoside analogue therapy: contribution to protease inhibitor–related lipodystrophy syndrome. AIDS 14:F25, 2000

127. Kotler DP, Rosenbaum K, Wang J, et al: Studies of body composition and fat distribution in HIV-infected and control subjects. J Acquir Immune Defic Syndr Hum Retrovirol 20:228, 1999

128. Lipodystrophy, metabolic disorders, and human immunodeficiency virus infection: Aquitaine Cohort, France, 1999. Groupe d'Epidemiologie Clinique du Syndrome d'Immunodeficience Acquise en Aquitaine. Clin Infect Dis 31:1482, 2000

129. Lichtenstein KA, Ward DJ, Moorman AC, et al: Clinical assessment of HIV-associated lipodystrophy in an ambulatory population. AIDS 15:1389, 2001

130. Miller KD, Jones E, Yanovski JA, et al: Visceral abdominal-fat accumulation associated with use of indinavir. Lancet 351:871, 1998

131. Lo JC, Mulligan K, Tai VW, et al: "Buffalo hump" in men with HIV-1 infection. Lancet 351:867, 1998

132. Saint-Marc T, Partisani M, Poizot-Martin I, et al: Fat distribution evaluated by computed tomography and metabolic abnormalities in patients undergoing antiretroviral therapy: preliminary results of the LIPOCO study. AIDS 14:37, 2000

133. Mulligan K, Tai VW, Algren H, et al: Altered fat distribution in HIV-positive men on nucleoside analog reverse transcriptase inhibitor therapy. J Acquir Immune Defic Syndr 26:443, 2001

134. Saves M, Raffi F, Capeau J, et al: Factors Related to the Presence of Fat Redistribution in HIV-Infected Patients Treated With Protease Inhibitor (PI)-Containing Regimens, APROCO Cohort, 1999. Abstract 14. 7th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2000
http://www.retroconference.org/2000/abstracts/14.htm

135. Wanke C, Jacobsen D, Knox T, et al: Prevalence of fat deposition and fat atrophy in a cohort of HIV-infected men and women. Abstract 24. 7th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2000
http://www.retroconference.org/2000/abstracts/24.htm

136. Arpadi SM, Cuff PA, Horlick M, et al: Lipodystrophy in HIV-infected children is associated with high viral load and low CD4+-lymphocyte count and CD4+-lymphocyte percentage at baseline and use of protease inhibitors and stavudine. J Acquir Immune Defic Syndr 27:30, 2001

137. Carr A, Workman C, Smith DE, et al: Abacavir substitution for nucleoside analogs in patients with HIV lipoatrophy: a randomized trial. JAMA 288:207, 2002

138. Roubenoff R, Weiss L, McDermott A, et al: A pilot study of exercise training to reduce trunk fat in adults with HIV-associated fat redistribution. AIDS 13:1373, 1999

139. Engelson ES, Glesby MJ, Mendez D, et al: Effect of recombinant human growth hormone in the treatment of visceral fat accumulation in HIV infection. J Acquir Immune Defic Syndr 30:379, 2002

140. Hadigan C, Corcoran C, Basgoz N, et al: Metformin in the treatment of HIV lipodystrophy syndrome: a randomized controlled trial. JAMA 284:472, 2000

141. Hadigan C, Yawetz S, Thomas A, et al: Metabolic effects of rosiglitazone in HIV lipodystrophy: a randomized, controlled trial. Ann Intern Med 140:786, 2004

142. Miller KD, Masur H, Jones EC, et al: High prevalence of osteonecrosis of the femoral head in HIV-infected adults. Ann Intern Med 137:17, 2002

143. Scribner AN, Troia-Cancio PV, Cox BA, et al: Osteonecrosis in HIV: a case-control study. J Acquir Immune Defic Syndr 25:19, 2000

144. Glesby MJ, Hoover DR, Vaamonde CM: Osteonecrosis in patients infected with human immunodeficiency virus: a case-control study. J Infect Dis 184:519, 2001

145. Paton NI, Macallan DC, Griffin GE, et al: Bone mineral density in patients with human immunodeficiency virus infection. Calcif Tissue Int 61:30, 1997

146. Tebas P, Powderly WG, Claxton S, et al: Accelerated bone mineral loss in HIV-infected patients receiving potent antiretroviral therapy. AIDS 14:F63, 2000

147. Hoy J, Hudson J, Law M, et al: Osteopenia in a randomized, multicenter study of protease inhibitor (PI) substitution in patients with the lipodystrophy syndrome and well-controlled HIV viremia. Abstract 208. 7th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2000
http://www.retroconference.org/2000/abstracts/208.htm

148. Pierson T, McArthur J, Siliciano RF: Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. Annu Rev Immunol 18:665, 2000

149. Paterson DL, Swindells S, Mohr J, et al: Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. Ann Intern Med 133:21, 2000

150. Baxter JD, Mayers DL, Wentworth DN, et al: A randomized study of antiretroviral management based on plasma genotypic antiretroviral resistance testing in patients failing therapy. CPCRA 046 Study Team for the Terry Beirn Community Programs for Clinical Research on AIDS. AIDS 14:F83, 2000

151. Durant J, Clevenbergh P, Halfon P, et al: Drug-resistance genotyping in HIV-1 therapy: the VIRADAPT randomised controlled trial. Lancet 353:2195, 1999

152. Boden D, Hurley A, Zhang L, et al: HIV-1 drug resistance in newly infected individuals. JAMA 282:1135, 1999

153. Little SJ, Koelsch KK, Ignacio CC, et al: Persistence of transmitted drug-resistant virus among subjects with primary HIV infection deferring antiretroviral therapy. Abstract 36LB. 11th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2004
http://www.retroconference.org/2004/cd/Abstract/36LB.htm

154. Palmer S, Boltz V, Maldarelli F, et al: Emergence and long-term persistence of NNRTI-resistant variants in patients starting and stopping NNRTI-containing regimens. Abstract 37. 11th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2004
http://www.retroconference.org/2004/cd/Abstract/37.htm

155. Mellors J, Palmer S, Nissley D, et al: Low-frequency NNRTI-resistant variants contribute to failure of efavirenz-containing regimens. Abstract 39. 11th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2004
http://www.retroconference.org/2004/cd/Abstract/39.htm

156. Oxenius A, Price DA, Gunthard HF, et al: Stimulation of HIV-specific cellular immunity by structured treatment interruption fails to enhance viral control in chronic HIV infection. Proc Natl Acad Sci USA 99:13747, 2002

157. Orenstein JM, Bhat N, Yoder C, et al: Rapid activation of lymph nodes and mononuclear cell HIV expression upon interrupting highly active antiretroviral therapy in patients after prolonged viral suppression. AIDS 14:1709, 2000

158. Colven R, Harrington RD, Spach DH, et al: Retroviral rebound syndrome after cessation of suppressive antiretroviral therapy in three patients with chronic HIV infection. Ann Intern Med 133:430, 2000

159. Kilby JM, Goepfert PA, Miller AP, et al: Recurrence of the acute HIV syndrome after interruption of antiretroviral therapy in a patient with chronic HIV infection: a case report. Ann Intern Med 133:435, 2000

160. Miller V: Structured treatment interruptions in antiretroviral management of HIV-1. Curr Opin Infect Dis 14:29, 2001

161. Deeks SG, Wrin T, Liegler T, et al: Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. N Engl J Med 344:472, 2001

162. Cardo DM, Culver DH, Ciesielski CA, et al: A case-control study of HIV seroconversion in health care workers after percutaneous exposure. Centers for Disease Control and Prevention Needlestick Surveillance Group. N Engl J Med 337:1485, 1997

163. Serious adverse events attributed to nevirapine regimens for postexposure prophylaxis after HIV-exposures—worldwide, 1997–2000. MMWR Morb Mortal Wkly Rep 49:1153, 2001

164. Parkin JM, Murphy M, Anderson J, et al: Tolerability and side-effects of post-exposure prophylaxis for HIV infection. Lancet 355:722, 2000

165. Sperling RS, Shapiro DE, Coombs RW, et al: Maternal viral load, zidovudine treatment, and the risk of transmission of human immunodeficiency virus type 1 from mother to infant. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N Engl J Med 335:1621, 1996

166. Shaffer N, Chuachoowong R, Mock PA, et al: Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: a randomised controlled trial. Bangkok Collaborative Perinatal HIV Transmission Study Group. Lancet 353:773, 1999

167. Guay LA, Musoke P, Fleming T, et al: Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. Lancet 354:795, 1999

168. Jackson JB, Becker-Pergola G, Guay LA, et al: Identification of the K103N resistance

mutation in Ugandan women receiving nevirapine to prevent HIV-1 vertical transmission. AIDS 14:F111, 2000

169. Stern JO, Robinson PA, Love J, et al: A comprehensive hepatic safety analysis of nevirapine in different populations of HIV infected patients. J Acquir Immune Defic Syndr 34(suppl 1):S21, 2003

170. Lyons F, Hopkins S, McGeary A, et al: Nevirapine tolerability in HIV infected women in pregnancy: a word of caution. Abstract LB27. 2nd IAS Conference on HIV Pathogenesis and Treatment, Paris, 2003
http://www.ias.se/abstract/show.asp?abstract_id=11081

171. VaxGen Announces Initial Results of its Phase III AIDS Vaccine Trial. VaxGen, February 24, 2003
http://www.vaxgen.com/pressroom/index.html

172. Pitisutithum P: Efficacy of AIDSVax B/E vaccines in injecting drug users. Abstract 107. 11th Conference on Retroviruses and Opportunistic Infections (CROI), San Francisco, 2004
http://www.retroconference.org/2004/cd/Abstract/107.htm

173. Wei X, Decker JM, Wang S, et al: Antibody neutralization and escape by HIV-1. Nature 422:307, 2003

174. Richman DD, Wrin T, Little SJ, et al: Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci USA 100:4144, 2003

175. Schneider J, Gilbert SC, Blanchard TJ, et al: Enhanced immunogenicity for $CD8^+$ T cell induction and complete protective efficacy of malaria DNA vaccination by boosting with modified vaccinia virus Ankara. Nat Med 4:397, 1998

176. Hanke T, Blanchard TJ, Schneider J, et al: Enhancement of MHC class I–restricted peptide-specific T cell induction by a DNA prime/MVA boost vaccination regime. Vaccine 16:439, 1998

177. McConkey SJ, Reece WH, Moorthy VS, et al: Enhanced T-cell immunogenicity of plasmid DNA vaccines boosted by recombinant modified vaccinia virus Ankara in humans. Nat Med 9:729, 2003

178. Kent SJ, Zhao A, Best SJ, et al: Enhanced T-cell immunogenicity and protective efficacy of a human immunodeficiency virus type 1 vaccine regimen consisting of consecutive priming with DNA and boosting with recombinant fowlpox virus. J Virol 72:10180, 1998

179. Caley IJ, Betts MR, Irlbeck DM, et al: Humoral, mucosal, and cellular immunity in response to a human immunodeficiency virus type 1 immunogen expressed by a Venezuelan equine encephalitis virus vaccine vector. J Virol 71:3031, 1997

180. Xin KQ, Urabe M, Yang J, et al: A novel recombinant adeno-associated virus vaccine induces a long-term humoral immune response to human immunodeficiency virus. Hum Gene Ther 12:1047, 2001

181. Fouts TR, Tuskan RG, Chada S, et al: Construction and immunogenicity of *Salmonella typhimurium* vaccine vectors that express HIV-1 gp120. Vaccine 13:1697, 1995

182. Boyer JD, Cohen AD, Vogt S, et al: Vaccination of seronegative volunteers with a human immunodeficiency virus type 1 env/rev DNA vaccine induces antigen-specific proliferation and lymphocyte production of beta-chemokines. J Infect Dis 181:476, 2000

183. Cao H, Kaleebu P, Hom D, et al: Immunogenicity of a recombinant human immunodeficiency virus (HIV)-canarypox vaccine in HIV-seronegative Ugandan volunteers: results of the HIV Network for Prevention Trials 007 Vaccine Study. J Infect Dis 187:887, 2003

184. Wee EG, Patel S, McMichael AJ, et al: A DNA/MVA-based candidate human immunodeficiency virus vaccine for Kenya induces multi-specific T cell responses in rhesus macaques. J Gen Virol 83:75, 2002

185. McMichael AJ, Hanke T: HIV vaccines 1983–2003. Nat Med 9:874, 2003

186. Kaplan JE, Hanson D, Dworkin MS, et al: Epidemiology of human immunodeficiency virus–associated opportunistic infections in the United States in the era of highly active antiretroviral therapy. Clin Infect Dis 30(suppl 1):S5, 2000

187. Jones JL, Hanson DL, Dworkin MS, et al: Incidence and trends in Kaposi's sarcoma in the era of effective antiretroviral therapy. J Acquir Immune Defic Syndr 24:270, 2000

188. 2001 USPHS/IDSA Guidelines for the Prevention of Opportunistic Infections in Persons Infected with HIV—November 28, 2001. AIDSinfo, U.S. Department of Health and Human Services, Washington, DC 2001
http://www.aidsinfo.nih.gov/guidelines/op_infections/OI_112801.html

189. Furrer H, Egger M, Opravil M, et al: Discontinuation of primary prophylaxis against *Pneumocystis carinii* pneumonia in HIV-1–infected adults treated with combination antiretroviral therapy. Swiss HIV Cohort Study. N Engl J Med 340:1301, 1999

190. Currier JS, Williams PL, Koletar SL, et al: Discontinuation of *Mycobacterium avium* complex prophylaxis in patients with antiretroviral therapy–induced increases in $CD4^+$ cell count: a randomized, double-blind, placebo-controlled trial. AIDS Clinical Trials Group 362 Study Team. Ann Intern Med 133:493, 2000

191. Furrer H, Opravil M, Bernasconi E, et al: Stopping primary prophylaxis in HIV-1–infected patients at high risk of *Toxoplasma* encephalitis. Swiss HIV Cohort Study. Lancet 355:2217, 2000

192. Mussini C, Pezzotti P, Govoni A, et al: Discontinuation of primary prophylaxis for *Pneumocystis carinii* pneumonia and toxoplasmic encephalitis in human immunodeficiency virus type I–infected patients: the changes in opportunistic prophylaxis study. J Infect Dis 181:1635, 2000

193. American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: Treatment of tuberculosis. Am J Respir Crit Care Med 167:603, 2003

194. Saag MS, Graybill RJ, Larsen RA, et al: Practice guidelines for the management of cryptococcal disease. Infectious Diseases Society of America. Clin Infect Dis 30:710, 2000

195. Graybill JR, Sobel J, Saag M, et al: Diagnosis and management of increased intracranial pressure in patients with AIDS and cryptococcal meningitis. The NIAID Mycoses Study Group and AIDS Cooperative Treatment Groups. Clin Infect Dis 30:47, 2000

196. DeSimone JA, Pomerantz RJ, Babinchak TJ: Inflammatory reactions in HIV-1–infected persons after initiation of highly active antiretroviral therapy. Ann Intern Med 133:447, 2000

197. Narita M, Ashkin D, Hollender ES, et al: Paradoxical worsening of tuberculosis following antiretroviral therapy in patients with AIDS. Am J Respir Crit Care Med 158:157, 1998

198. Ledergerber B, Mocroft A, Reiss P, et al: Discontinuation of secondary prophylaxis against *Pneumocystis carinii* pneumonia in patients with HIV infection who have a response to antiretroviral therapy. Eight European Study Groups. N Engl J Med 344:168, 2001

199. Aberg JA, Yajko DM, Jacobson MA: Eradication of AIDS-related disseminated *Mycobacterium avium* complex infection after 12 months of antimycobacterial therapy combined with highly active antiretroviral therapy. J Infect Dis 178:1446, 1998

200. Shafran SD, Mashinter LD, Phillips P, et al: Successful discontinuation of therapy for disseminated *Mycobacterium avium* complex infection after effective antiretroviral therapy. Ann Intern Med 137:734, 2002

201. Macdonald JC, Torriani FJ, Morse LS, et al: Lack of reactivation of cytomegalovirus (CMV) retinitis after stopping CMV maintenance therapy in AIDS patients with sustained elevations in CD4 T cells in response to highly active antiretroviral therapy. J Infect Dis 177:1182, 1998

202. Whitcup SM, Fortin E, Lindblad AS, et al: Discontinuation of anticytomegalovirus therapy in patients with HIV infection and cytomegalovirus retinitis. JAMA 282:1633, 1999

203. Jouan M, Saves M, Tubiana R, et al: Discontinuation of maintenance therapy for cytomegalovirus retinitis in HIV-infected patients receiving highly active antiretroviral therapy. AIDS 15:23, 2001

204. Soriano V, Dona C, Rodriguez-Rosado R, et al: Discontinuation of secondary prophylaxis for opportunistic infections in HIV-infected patients receiving highly active antiretroviral therapy. AIDS 14:383, 2000

205. Miro JM, Lopez JC, Podzamczer D, et al: Discontinuation of primary or secondary *Toxoplasma gondii* prophylaxis is safe in HIV-1 infected patients after immunological recovery with HAART. Final results of the GESIDA 04/98-B study. 40th ICAAC, Toronto, 2000, abstract L-16

206. Mussini C, Pezzotti P, Miro JM, et al: Discontinuation of maintenance therapy for cryptococcal meningitis in patients with AIDS treated with highly active antiretroviral therapy: an international observational study. Clin Infect Dis 38:565, 2004

Acknowledgments

Figures 2, 8, 9, 10, and 12 Marcia Kammerer.

Figures 3, 4, 5, 6, 7, and 11 Seward Hung.

XXXIV PROTOZOAN INFECTIONS

Wesley C. Van Voorhis, m.d., ph.d.
Peter F. Weller, m.d.

Definitions

Parasites are generally subdivided into three categories: protozoans, or single-celled nucleated organisms; helminths, which are metazoan wormlike organisms; and arthropods such as ticks and insects. This chapter focuses on protozoan pathogens [*see Figure 1*]; helminths are discussed elsewhere [*see 7:XXXV Helminthic Infections*].

Parasitism can lead to four different host-parasite states: (1) symbiosis, which is the association of two organisms that cannot exist independently; (2) mutualism, an association in which both organisms benefit; (3) commensalism, in which the parasite benefits and the host is unaffected; and (4) disease, in which the parasite benefits and the host is harmed.

Parasites can be ectoparasites that live on the outside of the host, where they cause an infestation, or endoparasites that live within the host, where they cause an infection. Hosts are classified by the forms of parasites found in them. There are five categories of hosts: definitive, reservoir, incidental, intermediate, and carrier. A definitive host harbors the adult or sexual form of the parasite. Definitive hosts can be reservoir hosts (animals that harbor the same parasite species as humans) or incidental hosts (unnecessary for the maintenance of the parasite in nature). Intermediate hosts harbor a developing larval or asexual form of the parasite. Carrier hosts harbor larval or asexual forms of the parasite without development.

Vectors are objects or organisms responsible for transmitting parasites between hosts. A vector may be biologic, in which the parasite multiplies or develops, or mechanical, which transmits the unchanged parasite from host to host.

In general, protozoa are classified by their organelles of locomotion [*see Figure 1*]. Sporozoa, including *Plasmodium, Toxoplasma*, and *Cryptosporidium*, have no evident organelles of locomotion. Microsporidia also lack evident organelles of locomotion. Ciliates, such as balantidia, move with cilia; flagellates, such as the trypanosomes *Giardia, Leishmania,* and *Trichomonas*, move with flagella; and *Sarcodina*, or amebae, move by means of pseudopodia extension of the cytoplasm.

Only the sporozoa have a clearly identified sexual phase. Other protozoa appear to divide simply by binary mitosis.

Malaria

Four *Plasmodium* species cause human malaria: *P. falciparum, P. vivax*, *P. ovale,* and *P. malariae* [*see Table 1*]. *P. vivax* and *P. falciparum* are the most prevalent worldwide, and falciparum malaria is the one responsible for most deaths. Malaria is transmitted by anopheline mosquitoes. Although malaria transmission in the United States is uncommon, about 300 to 500 million malaria cases, with one to three million deaths, are estimated to occur annually worldwide.[1,2] Malaria is resurgent because of multiple factors, including mosquito vector resistance to insecticides, inability of governmental programs to control mosquito vector populations, increased opportunities for breeding of vector mosquitoes, and increasing resistance of *Plasmodium* to chemotherapeutic agents.

etiology and epidemiology

Malaria is endemic in many countries [*see CE:VII Health Advice for International Travelers*]. Most malaria in the United States is imported. In 2001, 1,383 cases of malaria—11 of them fatal—were reported to the Centers for Disease Control and Prevention (CDC). Of these cases, 891 occurred in United States civilians, 316 in foreign civilians, and 18 in United States military personnel.[1] *P. falciparum* accounted for 50% of cases; *P. vivax,* for 28% . Along with infections acquired abroad, less common means of transmission are by blood transfusion, by sharing of contaminated needles, by organ transplantation, and congenitally in infants who are born of infected mothers.[1,3] In addition, because species of *Anopheles* mosquitoes that can transmit malaria are endemic in the United States, isolated cases of autochthonous malaria have developed in persons in the United States who have been bitten by local mosquitoes that fed on persons with imported malaria.[4] A final means of introduction of malaria into the United States is so-called airport malaria, which arises when an infected mosquito enters the country on an aircraft from a malarious area and transmits the infection in the area around the airport.[5]

pathogenesis

Malaria is normally acquired through the bite of an infected female *Anopheles* mosquito [*see Figure 2*]. Sporozoites inoculated into the bloodstream during the blood meal travel to the liver and infect hepatocytes; this is the preerythrocytic stage of infection. Notably, only two species of *Plasmodium*—*P. vivax* and *P. ovale*—cause persistent infections within the liver; the dormant intrahepatic sporozoites are termed hypnozoites. Within the he-

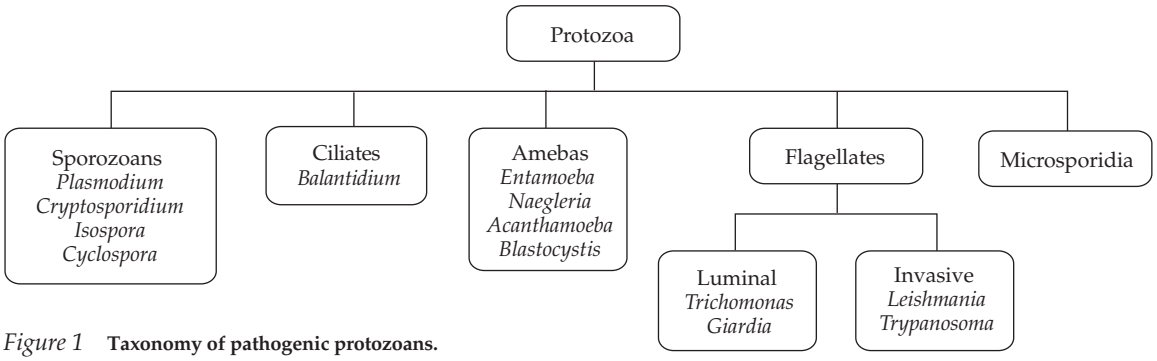


Figure 1 **Taxonomy of pathogenic protozoans.**

Table 1 Differentiating Features of Malaria Species

Species	*Global Prevalence*	*Mortality*	*Clinical Manifestations*	*Length of Erythrocytic Cycle*	*Persistence in Liver*	*Chloroquine Resistance*
P. falciparum	Common	Common (especially if untreated in nonimmune hosts)	Fever, anemia Microvascular blockade with local anoxia Brain (cerebral malaria) Lungs (pulmonary edema) Kidneys (acute renal failure) Intestines (diarrhea) Liver (tenderness, anoxia) Blood (thrombocytopenia)	36–48 hr	No	Common
P. vivax	Common	Uncommon	Fever, anemia Splenic rupture (uncommon)	48 hr	Yes	Uncommon
P. ovale	Uncommon	Rare	Fever, anemia	48 hr	Yes	No
P. malariae	Uncommon	Rare	Mild, chronic parasitemia, often with no symptoms Immune complexes (nephrotic syndrome)	72 hr	No	Rare

patocytes, the sporozoites transform into merozoites. In the erythrocytic stage of malarial infection, merozoites released from infected hepatocytes and later from infected erythrocytes interact with specific erythrocyte membrane proteins and invade the red blood cells. (The Duffy blood group antigen is the requisite erythrocyte receptor for *P. vivax*. Absence of the Duffy blood group—a genetic trait found in many West Africans and their descendants—confers resistance to that species of *Plasmodium*.) Within the infected cells, malarial parasites undergo schizogony to form new merozoites, which are then released to reinfect other erythrocytes. A few parasites differentiate into sexual stages (gametocytes) capable of infecting mosquitoes [*see Figure 3*]. Male and female gametocytes transform and reproduce in the midgut of the mosquito, leading to the production of new sporozoites that localize in the mosquito's salivary glands.

Because malarial parasites can multiply in an infected human, intense infections can develop from minimal inocula of sporozoites. Travelers have developed malaria after being bitten by mosquitoes during brief layovers at airports in malarious areas and during flights on airplanes that have stopped in malarious regions. Consequently, malaria prophylaxis is warranted for even the most limited exposures to mosquitoes in malarious regions.

Infected persons remain asymptomatic during the time between the infecting mosquito bite and the erythrocytic stage of infection, a period that may range from about 1 to 4 weeks for deadly *P. falciparum* infection. Because malaria chemoprophylaxis does not actually prevent malaria but, rather, treats erythrocytic-stage infection, chemoprophylactic medication must be continued for the full 4 weeks after a return from a malarious area. Failure to do so permits the development of *P. falciparum* infection. An exception to this is with so-called causal prophylactic medication, such as atovaquone-proguanil, which also kills liver-stage parasites. This form of prophylaxis can be discontinued a week after leaving a malarious area. Prophylaxis generally does not prevent late emergence of *P. vivax* or *P. ovale* from incubating liver hypnozoites.

Immune responses develop to malarial infections. Development of antibodies to the most indolent and chronic malarial species, *P. malariae*, may lead over years to the onset of immune complex–mediated nephrotic syndrome. Repeated infections with other *Plasmodium* species can elicit a partially protective immunity that will limit the severity of infection but not prevent it. Of note, residents of malarious areas may lose their relative immunity after staying several years in a nonmalarious region; such persons may be fully susceptible to malarial infection upon their return.

DIAGNOSIS

Malaria should be considered as a cause of any febrile illness in immigrants from malarious areas and in persons who have traveled or worked in malarious areas. Febrile illness in recipients of transfused blood or transplanted organs and in neonates of potentially infected mothers should be considered as possible acquired malaria. Finally, even in United States residents, the diagnosis of malaria may need to be considered in febrile patients with compatible illnesses that may have developed from the uncommon autochthonous transmission of malaria.

Infections with the four human malarial species of *Plasmodium* produce distinct clinical syndromes, based in part on the interactions of each species with erythrocytes. The four species of malaria can be distinguished by their characteristics on blood smears [*see Figure 3*]. *P. vivax* and *P. ovale* infect young red cells only, which helps limit the intensity of infection. It also explains why enlarged red cells are characteristic of *P. vivax* and *P. ovale* infections in speciation of malaria on blood films. In contrast, *P. falciparum* infects erythrocytes of all ages. This capacity of falciparum malaria, the greater numbers of merozoites produced by this species, and especially the great propensity of falciparum-infected erythrocytes to adhere to the microvascular endothelium help make falciparum malaria distinctly more severe than other forms of malaria. Erythrocytes infected with *P. falciparum* develop unique surface knobs that mediate binding and adherence to endothelial cells in capillaries and venules. Sequestration of infected erythrocytes in these small vessels results in local anoxia and can lead to severe complications, including cerebral malaria and pulmonary edema [*see Table 1*].

Clinical Features

Symptoms of malaria develop about 1 to 4 weeks after infection and typically include fever and chills. Virtually all patients with acute malaria have episodes of fever. At the outset, fever may occur daily; over time, the paroxysms may take on the typ-

ical pattern of fevers every other day (*P. vivax*, *P. ovale*, *P. falciparum*) or every third day (*P. malariae*) [*see Table 1*]. The paroxysms of fever, which may reach as high as 41.5° C [106.7° F], and chills (with or without rigors) may be irregular, however—especially in falciparum malaria. Other possible symptoms are headache, increased sweating, back pain, myalgias, diarrhea, nausea, vomiting, and cough. The constellations of symptoms are nonspecific and may suggest diagnoses other than malaria.[6] With time, anemia and splenomegaly develop.

Because of the distinct capacity of falciparum-infected erythrocytes to cause microvascular blockade, potentially fatal organ involvement can develop rapidly in falciparum malaria. Cerebral involvement may lead to delirium, focal disorders (e.g., seizures), and coma.[7] Pregnant women are at special risk for death and fetal loss from falciparum infections. Splanchnic involvement may cause protracted nausea, vomiting, diarrhea, melena, and abdominal pain; this syndrome can be readily mistaken for traveler's diarrhea. Lung involvement may cause pulmonary edema and acute respiratory distress syndrome. There may be severe hypoglycemia. A rare syndrome known as blackwater fever reflects hemoglobinuria and acute renal failure from massive intravascular hemolysis.

P. malariae organisms can persist in the blood as an indolent, even asymptomatic, infection for years or even decades.[8]

Laboratory Findings

The white blood cell count is usually in the normal range in malaria patients. Anemia develops but may not be prominent on presentation, especially if the patient is dehydrated. Thrombocytopenia may develop; disseminated intravascular coagula-

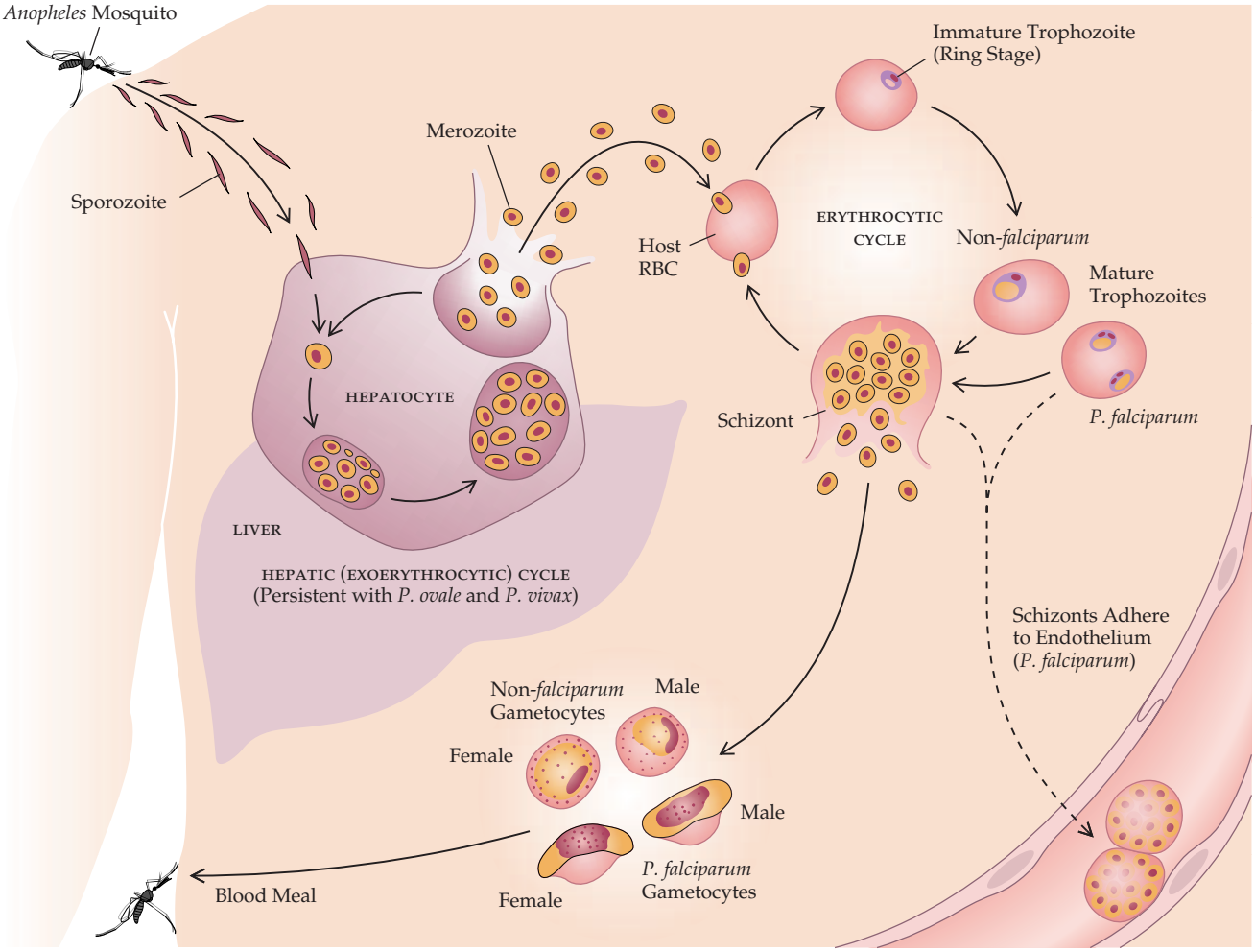


Figure 2 **The life cycle of malaria.[110] All four human malaria species—*Plasmodium falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*—may be transmitted by *Anopheles* mosquito bites or, rarely, introduced by blood (e.g., congenitally, through transfusion, and by sharing of needles). Once in the bloodstream, *Plasmodium* sporozoites travel to the liver and infect hepatocytes. This is the preerythrocytic stage of infection. Only two species of *Plasmodium*, *P. vivax* and *P. ovale*, cause persistent infections within the liver, which can lead to later recrudescence of malaria. Within the hepatocytes, the sporozoites transform into merozoites. In the erythrocytic stage of malarial infection, merozoites released from infected hepatocytes and later from infected erythrocytes interact with specific erythrocyte membrane proteins and invade the red blood cells. Within the infected cells, malarial parasites undergo schizogony to form new merozoites, which are then released to reinfect other erythrocytes. A few parasites differentiate into sexual stages (gametocytes) capable of infecting mosquitoes. Symptoms develop with infection of erythrocytes (erythrocytic cycle), about 1 or more weeks after a mosquito bite; malaria therapy is directed toward the erythrocytic stage. *P. falciparum* causes more severe disease because it is able to invade all stages of red blood cells (RBCs); thus, the parasitemia is higher. In addition, late trophozoites and schizonts of *P. falciparum* adhere to endothelia, causing end-organ damage such as cerebral malaria or placental malaria in pregnancy.**

tion sometimes occurs in falciparum malaria. Liver enzyme levels may be elevated.

The specific diagnosis and speciation of malaria depends on the recognition of parasites in properly stained smears of peripheral blood [*see Figure 3*]. Thick smears are more sensitive than thin smears, but the layering of cells necessitates greater expertise in examining morphology. Smears should be taken repeatedly for several days because of the cyclic nature of the parasitemia. This is especially important in suspected *P. falciparum* infections, in which infected cells may be sequestered in the microvasculature and in which late trophozoites and schizonts are generally not seen. The morphologic features of the parasites (and the infected host erythrocytes) are useful in species identification and in distinguishing *Plasmodium* species from the morphologically similar *Babesia microti*, which causes babesiosis. It is very important to identify and treat *P. falciparum* infection rapidly because of its potential for swift progression if left untreated or if treated improperly. In the absence of a skilled microscopist, the clinician is advised to presumptively treat for falciparum malaria when malarial forms are identified on a blood smear.

The quantitative buffy coat (QBC) technique is sensitive in identifying plasmodial parasitemia, but it does not identify the species. Dipstick antigen-capture assays are available and have the potential to help clinicians detect falciparum infections, particularly where no specialized laboratories are available.[9] DNA probes, polymerase chain reaction (PCR) tests, and serologic tests exist, but these are not usually immediately available for diagnosis of acute malaria at presentation. Thus, examination of blood smears is still the standard in assessing, first, whether a patient has malaria and, second, whether the malaria is caused by *P. falciparum*.

TREATMENT

Treatment of patients with acute malaria requires consideration of the *Plasmodium* species involved and, in cases of falciparum malaria, the likelihood of resistance to antimalarial med-

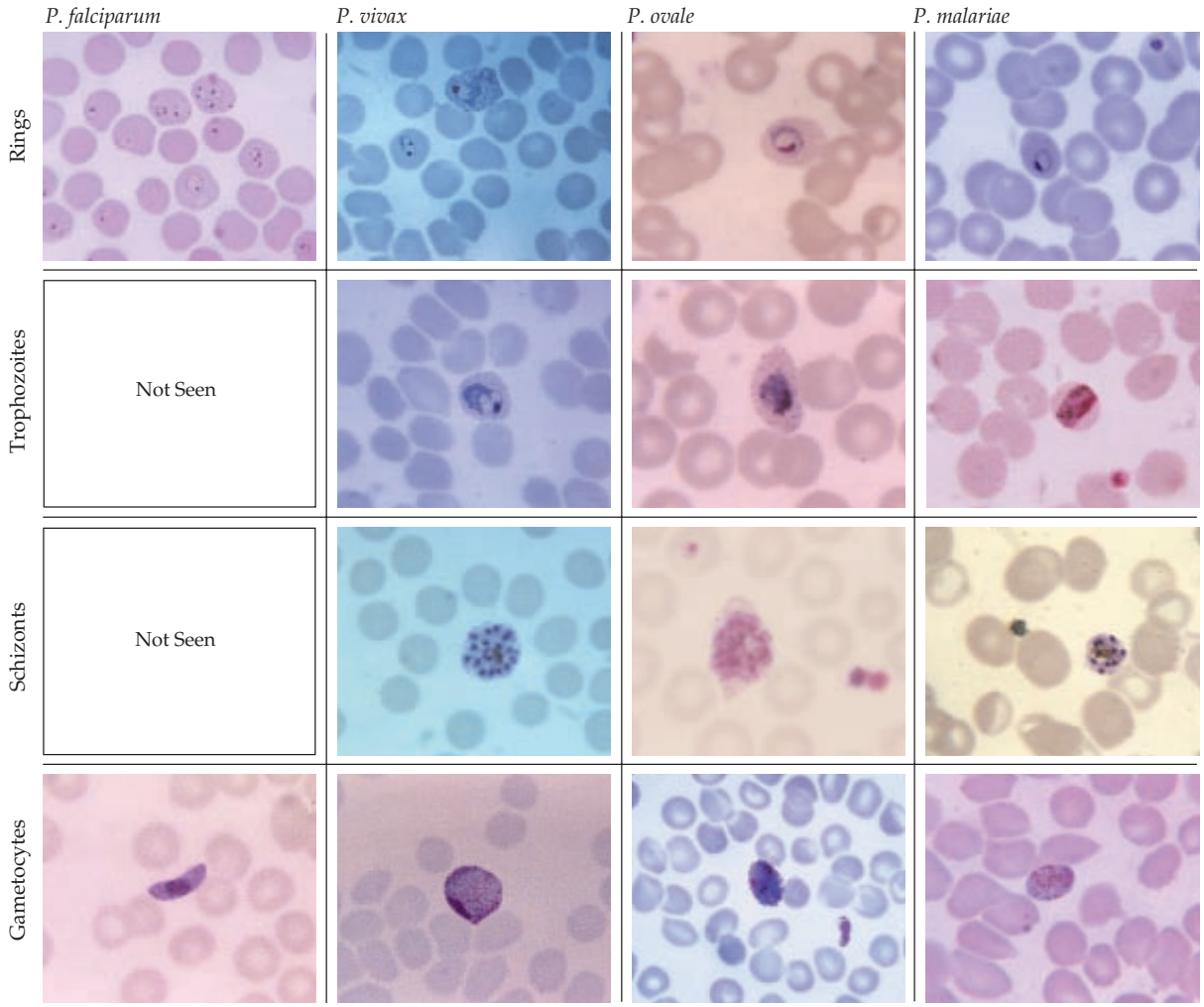


Figure 3 **Identification of species of malaria based on forms seen on blood smears. During the ring stage, multiple parasites are more likely to be seen within the red blood cell in *Plasmodium falciparum* infection. The late trophozoites and schizonts of *P. falciparum* are not usually seen on smears because of their propensity to adhere to the endothelia of the peripheral organs. *P. falciparum* infects mature red blood cells, whereas *P. vivax* and *P. ovale* commonly infect only reticulocytes, which are larger and slightly bluish when stained. Schizonts of *P. vivax* often have more than 14 nuclei, which distinguishes them from other malarial species. The trophozoites of *P. malariae* often have a bandlike appearance across the red blood cell. The spots in late trophozoites are hemozoin deposits; these cells can sometimes be mistaken for granulocytes. Banana-shaped gametocytes are pathognomonic for *P. falciparum* infection, but they are seen only after 10 to 14 days of erythrocytic infection.**

Table 2 Antimalarial Drugs[10]

Category	*Drug (Trade Name)*	*Dosage*	*Cost of Treatment ($)**	*Principal Side Effects*	*Main Limitations*
Quinine derivatives	Chloroquine phosphate	1,000 mg, then 500 mg at 6, 24, and 48 hr	29	Headache, nausea, vomiting, blurred vision, dizziness, fatigue, and confusion	Resistance
	Quinine	650 mg q. 8 hr × 3–7 days	17	Tinnitus, temporary hearing loss, headache, dysphoria, nausea, vomiting, and mild visual disturbances	Compliance, safety, and resistance
	Amodiaquine	—	—	Neutropenia, hepatoxicity	Safety and resistance
	Mefloquine (Lariam)	750 mg, then 500 mg at 12 hr	89	Nausea, dizziness, vivid dreams, and neuropsychiatric disturbances	Safety, resistance, and cost
Artemisinins†	Artemether†	—	—	Infrequent and mild decreases in reticulocyte and neutrophil counts, elevated liver transaminases, abdominal pain, diarrhea, and drug fever	Compliance, safety, and cost
	Artesunate†	4 mg/kg/day × 3 days	—	Same as for artemether (see above)	Compliance, safety, and cost
Antibiotics‡	Tetracycline Doxycycline Clindamycin	500 mg q.i.d × 7 days 100 mg b.i.d. × 7 days 900 mg t.i.d. × 5 days	8 9 40	GI disturbance, photosensitivity, diarrhea, and *Candida* vaginitis	Slowly effective if used alone
Combinations	Sulfadoxine-pyrimethamine (Fansidar)	1,500/75 mg p.o., single dose	4	Rash, fever, and marrow suppression	Resistance
	Atovaquone-proguanil (Malarone)	500/200 mg b.i.d. × 3 days	38	Rash, nausea, and diarrhea	Resistance potential and cost
	Artemether-lumefantrine (Coartem)†	1–4 tablets (by patient weight) at 0, 8, 24, 36, 48, and 60 hr	—	Possible mild CNS disturbances	Compliance, resistance potential, and cost

*Costs are derived from online pharmaceutical sources and are intended to indicate relative costs of available therapies.
†Not available in the United States; artemisinins should be given with another antimalarial.
‡Antibiotics are usually used in combination with quinine or other rapidly killing antimalarials.

ications. Chloroquine is the mainstay of treatment for all malarial species except those strains of *P. falciparum* that are resistant to the drug [*see Table 2*]. Notably, chloroquine-resistant *P. falciparum* (CRPF) malaria is now widespread in all countries where *P. falciparum* is endemic, except Haiti, the Dominican Republic, areas of Central America west of the Panama Canal, and parts of the Middle East [*see CE:VII Health Advice for International Travelers*]. For *P. falciparum* infections acquired in countries with chloroquine-resistant strains, alternative therapies must be utilized [*see Table 2*]; the typical regimen in such cases remains quinine or quinidine combined with a tetracycline. Mefloquine-resistant strains of *P. falciparum* have been identified at the Thailand-Myanmar border. Isolated strains of *P. vivax* resistant to chloroquine have been reported in Africa, Central and South America, Oceania, and Asia. Isolates of *P. malariae* resistant to chloroquine have now been reported, as well.

Acute attacks caused by all species except CRPF are treated with oral chloroquine phosphate at an initial dose of 1 g (600 mg base), followed by 500 mg (300 mg base) at 6 hours and again at 24 and 48 hours.[10] For suspected CRPF malaria, several alternative regimens are available.[10] Quinine sulfate (650 mg every 8 hours for 3 to 7 days) is combined with doxycycline (100 mg b.i.d. for 7 days), sulfadoxine-pyrimethamine (Fansidar) (three 500/25 mg tablets as a single dose on the last day of quinine treatment), or clindamycin (900 mg t.i.d. for 5 days). Alternative therapies for CRPF malaria include mefloquine (750 mg followed by a dose of 500 mg 12 hours later) and atovaquone in combination with proguanil (500 mg and 200 mg b.i.d. within 45 minutes of meals for 3 days).[10]

For patients who are too ill to take oral therapies, intravenous quinidine gluconate and quinine dihydrochloride are the drugs of choice for any species of malaria.[10] Quinidine is effective and well tolerated when appropriate precautions, including hemodynamic and electrocardiographic monitoring, are used. A loading dose of 10 mg/kg (maximum, 600 mg) of quinidine in normal saline is infused slowly over 1 to 2 hours, followed by continuous infusion of 0.02 mg/kg/min. Parenteral therapy should be continued until oral therapy can be tolerated; in most cases, oral therapy can be substituted within 48 to 72 hours. For fulminant falciparum malaria, exchange transfusion can be an adjunct to chemotherapy, but it has not been convincingly shown to decrease mortality.[11] Patients with moderate to severe falciparum malaria should be hospitalized and monitored for hypoglycemia and other complications, as well as response to therapy.

With the exception of atovaquone-proguanil, each of the chemotherapeutic agents above is active only in the erythrocytic stage of infection. Because *P. vivax* and *P. ovale* have persisting hepatic stages that may cause relapses of malaria after chloroquine use, an agent that is active against the exoerythrocytic cycle should be administered after a course of chloroquine. Primaquine is the sole agent used to eradicate hepatic involvement; the dosage is 26.3 mg (15 mg base) orally every day for 2 weeks.

P. vivax from Oceania and Tanzania has been reported to be primaquine tolerant, and 1.5 to 2 times the normal dose is given for 2 to 3 weeks if primaquine-tolerant strains are suspected.[10] Because primaquine may induce hemolysis in persons with glucose-6-phosphate dehydrogenase (G6PD) deficiency, patients should be screened for the disorder before treatment. If only mild G6PD deficiency is found, primaquine may be given at a dosage of 79 mg (45 mg base) orally once a week for 8 weeks.

Advice for Travelers

Most United States residents lack immunity to malaria and are at risk for the morbidity and mortality associated with the disease, especially the falciparum form, if they visit countries where malaria transmission occurs. Of the 891 United States civilians who acquired malaria abroad in 2001, only 20.2% had followed a CDC-recommended chemoprophylactic drug regimen.[1] It is imperative that travelers receive appropriate advice on reducing their risk of acquiring malaria, including the use of appropriate chemoprophylactic regimens [*see CE:VII Health Advice for International Travelers*]; up-to-date information can be found on the CDC Web site (http://www.cdc.gov/travel). Current alternatives, depending on the area of travel and the traveler, include chloroquine, mefloquine, doxycycline, atovaquone-proguanil, and primaquine.

Babesiosis

Babesia organisms are intraerythrocytic protozoan parasites that produce a malarialike illness. Most cases of babesiosis have been reported in the northeastern United States, but cases have also occurred in the upper Midwest, the West Coast, and other regions of the United States, as well as in Europe and elsewhere.[12]

ETIOLOGY AND EPIDEMIOLOGY

Several species of *Babesia* have been recognized as causes of human disease. The first to be recognized was *B. divergens*, a parasite of cattle that has caused several fatal infections in splenectomized persons. Most of these cases have occurred in Europe, although a related species, designated MO1, has caused a fatal case in a splenectomized man in Missouri.[13]

B. microti is the principal cause of babesiosis in the eastern and central United States. This parasite of white-footed mice is transmitted by deer ticks and is prevalent on the islands off Massachusetts, New York, and Rhode Island and in focal areas in Connecticut, Wisconsin, and Minnesota. Risk of *Babesia* infection is not increased by splenectomy, but the disease is more serious in persons who have undergone splenectomy, have HIV infection, or have other immunocompromising conditions.[12,14] Many infected persons have subclinical infections, as evidenced by serologic surveys in endemic areas. Because nymphal ticks are the most efficient at transmitting infection, most cases develop between May and August, when nymphal ticks are most abundant.[12]

Another form of babesiosis develops from infection with an organism found in states along the Pacific coast and designated as WA1.[15] The vector and the reservoir of this emerging *Babesia* species are not yet known. Infections have been recognized in asplenic persons and less commonly in normal hosts; serologic surveys in rural and semirural California indicate that subclinical infection with WA1 may have developed in up to 20% of the population.[12]

Babesiosis may also be acquired perinatally and through blood transfusions.[12,15]

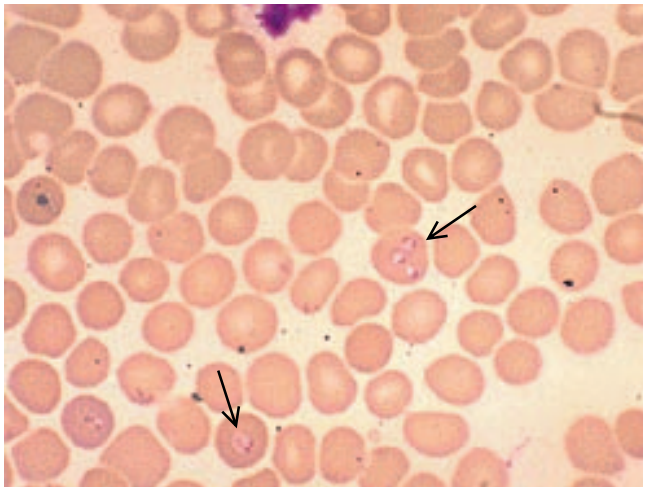


Figure 4 **Blood smear showing *Babesia* parasites in erythrocytes (arrows).**

DIAGNOSIS

Clinical Features

Many persons infected with *B. microti* or WA1 remain asymptomatic. In those who have symptoms, illness develops gradually in the weeks after tick bites or blood transfusions. Malaise, anorexia, and fatigue are followed by the onset of myalgias, fevers, and sweats. Emotional lability, depression, nausea, vomiting, and headache are common. Symptoms tend to abate over several weeks, although fatigue and malaise may persist for months. Splenomegaly is occasionally present. Infections may become more fulminant and persistent in asplenic patients, elderly patients, and patients with HIV infection or other immunocompromising conditions. Intravascular hemolysis, hemoglobinuria, and renal failure may develop. On blood smears, parasitemia is usually less than 1%, but parasites may be present in as many as 85% of red blood cells; this level of parasitemia can be fatal, especially in asplenic persons.

Laboratory Findings

Laboratory findings may include hemolytic anemia, normal to low leukocyte counts, and abnormal liver function test results. The diagnosis is based on the finding of ring forms and pleomorphic intraerythrocytic organisms on Giemsa-stained blood smears [*see Figure 4*]. Occasionally, extracellular merozoites are seen. Unlike malarial parasites, *Babesia* organisms do not produce pigment in red blood cells. In some patients, a parasitemia level of 5% or higher and the presence of several ring forms in a single red cell might suggest *P. falciparum* infection; the absence of gametocytes and intracellular pigmentation will help distinguish babesiosis from malaria. Serologic testing is of value for diagnosing chronic infections with *B. microti* and WA1 when parasites are not detectable on blood smears.

TREATMENT

The treatment of choice for babesiosis is a combination of clindamycin (1.2 g b.i.d. I.V. or 600 mg t.i.d., p.o., for 7 to 10 days) and quinine (650 mg t.i.d., p.o., for 7 days).[10] An alternative is atovaquone (750 mg p.o., b.i.d., for 7 to 10 days) plus azithromycin (600 mg p.o. daily for 7 to 10 days).[10] In a minority of patients who improve symptomatically, parasitemia may nevertheless persist.[16] Occasionally, pulmonary edema may de-

velop after initiation of therapy for babesiosis, as also may occur with malaria. Exchange transfusion can be an adjunct to chemotherapy in cases of fulminant babesiosis.

The tick vector of babesiosis may simultaneously transmit *Borrelia burgdorferi*, the agent of Lyme disease [*see 7:VII Leptospirosis, Relapsing Fever, Rat-Bite Fever, and Lyme Disease*], and *Ehrlichia* (*Anaplasma*) species, the agents of ehrlichiosis [*see 7:XVII Infections Due to Rickettsia, Ehrlichia, and Coxiella*]. Those coinfections may make clinical illness more severe and warrant special therapy.[17]

Toxoplasmosis

Toxoplasma gondii, the cause of toxoplasmosis, is an intracellular protozoan parasite of worldwide distribution. *T. gondii* can infect nearly all animals and birds, making it one of the most widely distributed parasites.[18] Cats are the definitive hosts of *T. gondii*, because only felines harbor the sexual forms; however, humans can develop toxoplasmosis in ways other than exposure to cats.

Humans acquire *T. gondii* infections by the oral and transplacental routes and, less commonly, from blood transfusion and organ transplantation. Infection via the oral route is caused by the ingestion of *T. gondii* tissue cysts in undercooked food or the ingestion of *T. gondii* oocysts, which are found in cat feces [*see Figure 5*]. After ingestion, bradyzoites (from tissue cysts) or sporozoites (from oocysts) invade surrounding cells and develop into tachyzoites, a rapidly dividing form that may disseminate and invade any nucleated cell type. With time and the development of immunity, the parasite will form tissue cysts containing many bradyzoites. Tissue cysts may remain viable for decades without causing disease. Loss of immunity, however, allows reactivation from the latent tissue cysts and the generation of many invasive tachyzoites.

Although toxoplasmosis is extremely common, most acquired and congenital infections are subclinical and are revealed only by the presence of antibodies. The prevalence of infection varies greatly in different population groups and geographic regions. In the United States, serologic evidence of *Toxoplasma* infection varies regionally in prevalence from 3% to 30%.

CLINICAL SYNDROMES

Toxoplasmosis can be divided into five major clinical syndromes: primary toxoplasmosis, toxoplasmosis in immunosup-

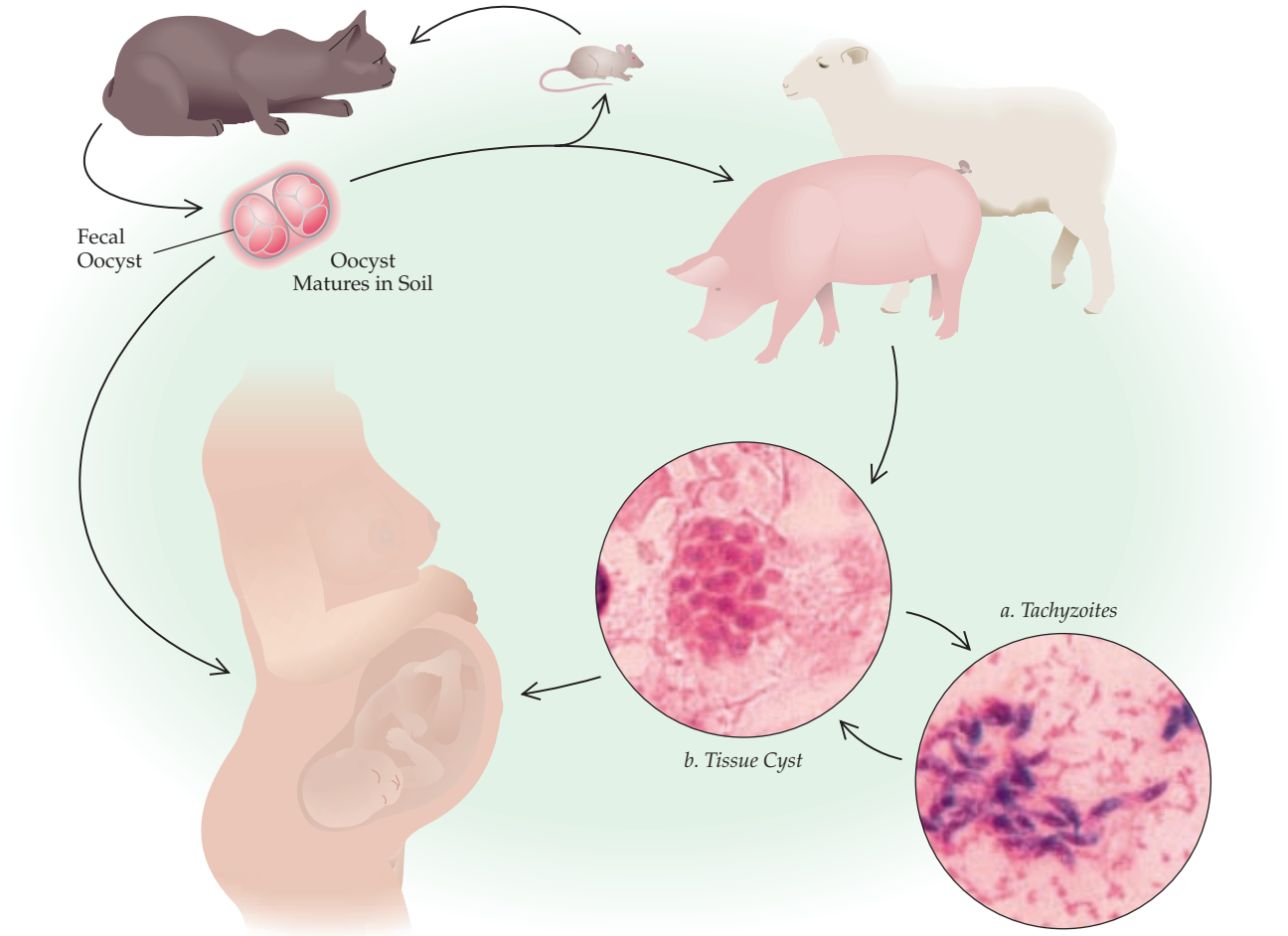


Figure 5 **Life cycle of *Toxoplasma gondii*. Felines, which acquire infection by carnivorism, are the definitive hosts; they are the only species supporting the sexual cycle that leads to oocyst excretion in feces. The oocysts must mature in the environment for at least several days to become infectious to intermediate hosts such as humans. Humans and other carnivores become infected by ingestion of mature oocysts or by ingestion of tissue cysts in undercooked meat. Within the host, both oocysts and tissue cysts convert to rapidly proliferating tachyzoites (*a*); tachyzoites can invade and grow in almost any cell. They form tissue cysts (*b*), which persist in a latent form and can reactivate to proliferate and cause clinical disease if the host becomes immunosuppressed. Reactivation can occur even decades after infection.**

pression, toxoplasmosis in AIDS, congenital toxoplasmosis, and ocular toxoplasmosis.

Primary Toxoplasmosis

An immunocompetent person usually experiences subclinical illness after acquisition of a primary toxoplasmosis infection. Painless lymphadenopathy is the most common symptom. Lymphadenopathy may be localized or generalized and may persist for many months. Isolated cervical lymphadenopathy is the most frequent finding.[19] Enlarged nodes may be the only manifestation; less often, fever, malaise, myalgias, and sore throat are also present. Fatigue and weakness may be pronounced. In addition to lymphadenopathy, physical findings may include pharyngitis, maculopapular rash,[20] and, in a minority of patients, hepatosplenomegaly. Atypical lymphocytosis may be present. The major differential diagnosis includes infectious mononucleosis, cytomegalovirus infection, and lymphoma; less often, sarcoidosis, cat-scratch disease, and other infectious processes may resemble toxoplasmosis. Serologic testing is the key to diagnosis.[21]

Although symptoms of acquired lymphadenopathic toxoplasmosis can persist for weeks or months, the process is almost always self-limited and does not require specific therapy. In severe or protracted cases, pyrimethamine and sulfonamides may be helpful. Chemotherapy should also be used for the rare complications that can occur in the normal host, including chorioretinitis, pericarditis, myocarditis, pneumonitis, myositis, and meningoencephalitis.

Toxoplasmosis in Immunosuppression

In an immunosuppressed person with defective cell-mediated immunity, *T. gondii* can cause devastating neurologic or disseminated disease. Usually, this process develops from reactivation of latent infection and not from a primary infection. Seronegative recipients of organs (especially hearts) transplanted from seropositive donors are at particular risk,[22,23] as are patients with Hodgkin disease, hairy-cell leukemia, and other malignant disorders.[24]

Neurologic abnormalities predominate in at least 50% of these patients. The clinical picture is highly variable and may take the form of a diffuse encephalitis, a meningoencephalitis, or a cerebral mass lesion.[25] The cerebrospinal fluid often shows a mild lymphocytic pleocytosis and an elevation of protein levels; glucose levels, however, remain normal. Pneumonitis may occur in immunosuppressed patients with toxoplasmosis.[26] Fever and dyspnea are present, but cough is absent and no sputum is produced. Chest radiographs typically reveal diffuse bilateral pulmonary infiltrates. In a few cases, the diagnosis has been established by the presence of organisms in the fluid obtained from bronchoalveolar lavage. Other manifestations of disseminated toxoplasmosis in immunosuppressed patients include myocarditis, pericarditis, peritonitis, and lymphadenitis. Many of these immunosuppressed patients have simultaneous infections with other opportunistic pathogens, especially herpesvirus and cytomegalovirus.

Appropriate serologic titers will support the diagnosis of toxoplasmosis; however, antibody levels may be low in these patients, and therefore, brain, lung, or lymph node biopsy may be required for diagnosis. An aggressive diagnostic approach is warranted because pyrimethamine and sulfonamides can be effective in these infections, which are usually fatal if left untreated.

Toxoplasmosis in AIDS

Toxoplasmosis is a particularly severe problem in patients with AIDS; clinically apparent toxoplasmosis develops in 3% to 40% of these patients.[27] Although any of the manifestations of disseminated toxoplasmosis may occur in patients with AIDS,[28] central nervous system abnormalities predominate. Most cases of toxoplasmosis in patients with AIDS result from reactivation of latent *Toxoplasma* cysts acquired before infection with HIV; reactivation is particularly likely when the $CD4^+$ T cell count falls below 100 cells/mm^3.[27] All HIV-infected persons should be tested for antibodies to *T. gondii.* Those who are *Toxoplasma* negative should be cautioned to minimize exposure by cooking meat to an internal temperature of 150° F, washing their hands after contact with raw meat or soil, washing fruits and vegetables before eating them, and avoiding contact with cat feces or contaminated litter (or wearing gloves if these materials must be handled).[29]

In AIDS patients, toxoplasmosis most often presents as necrotizing encephalitis. Symptoms include focal abnormalities (e.g., hemiparesis, sensory loss, visual abnormalities, tremor, cranial nerve palsies, and focal seizures) and generalized neurologic abnormalities (e.g., headache, personality changes, confusion, stupor or coma, and seizures). Although a CSF lymphocytic pleocytosis is consistent with the diagnosis of cerebral toxoplasmosis, computed tomographic or magnetic resonance imaging scans are the crucial diagnostic tests in most cases. Typical findings include single or multiple rounded mass lesions; when a contrast agent is administered, more than 90% of these lesions display ring or nodular enhancement. MRI scans may reveal lesions that were not visualized by CT scanning.[30] Positron emission tomography (PET) may prove useful for distinguishing lesions of toxoplasmosis (which are hypometabolic) from lymphomas (which are hypermetabolic).[31]

Serum antibody tests are useful for screening AIDS patients for cerebral toxoplasmosis; such tests are usually positive because most cases of cerebral toxoplasmosis in patients with AIDS arise from reactivation of latent infection. Although negative antibody tests suggest a diagnosis other than toxoplasmosis, a few seronegative cases have been reported. However, serum antibody tests cannot be relied on in the diagnosis of primary toxoplasmosis in patients with AIDS; antibody titers do not reach the high levels typical of immunocompetent patients with toxoplasmosis, nor are IgM antibodies always present in patients with AIDS.[27] Antibodies against *Toxoplasma* are present in the CSF in nearly two thirds of AIDS patients with cerebral toxoplasmosis, and their detection may assist in the diagnosis [*see* Diagnosis, *below*].[32]

In addition to toxoplasmosis, CNS lesions caused by fungi, mycobacteria, lymphomas, Kaposi sarcoma, metastatic tumors, multifocal leukoencephalopathy, or HIV itself may develop in patients with AIDS. Despite this broad differential diagnosis, empirical treatment of toxoplasmosis is often preferable to early brain biopsy if the clinical and radiologic picture is compatible with the diagnosis and if serum anti-*Toxoplasma* antibodies are present.[33] Because toxoplasmosis is the most common and most treatable cause of cerebral lesions in patients with AIDS, empirical therapy using a combination of sulfonamides and pyrimethamine or the combination of clindamycin and pyrimethamine can be initiated; if the diagnosis is correct, clinical and radiologic improvement is often observed within 1 to 2 weeks. If patients respond poorly to treatment and are seronegative or belong to population groups with a high risk of tuberculosis (e.g., Haitians, Africans, or intravenous drug abusers), biopsy should be strongly considered.[33]

Congenital Toxoplasmosis

Congenital toxoplasmosis arises almost exclusively when the mother develops a primary infection during gestation. Congenital infection almost never develops from latent toxoplasmosis acquired before pregnancy, and the few recognized cases have developed mainly in pregnant women with reactivation of toxoplasmosis secondary to immunosuppressive therapy or HIV infection.[34] The risk of fetal infection depends on when maternal infection occurs, rising from 10% during the first trimester to 60% during the third trimester.[35] The consequences of fetal infection also depend on when infection occurs: fetal infections early in gestation are most likely to result in severe damage.

The clinical spectrum of congenital toxoplasmosis varies widely. About 85% of infected babies appear normal at birth; without treatment, however, about 85% of these infants will experience chorioretinitis, hearing loss, or developmental delay.[35,36] The clinical spectrum of symptomatic congenital toxoplasmosis includes fetal death, neurologic damage (cerebral calcification, seizures, retardation, hydrocephalus, or microcephaly), chorioretinitis, fever, hepatosplenomegaly, and rash. The differential diagnosis includes congenital infection with cytomegalovirus, herpes simplex, rubella, or syphilis. Congenital toxoplasmosis can be diagnosed by serologic methods, by a PCR test performed on amniotic fluid,[37] or by identification of the organism in the placenta or fetal tissues. Infants with congenital toxoplasmosis may benefit from prolonged treatment with pyrimethamine and sulfonamides.[38]

The prevention of congenital toxoplasmosis is of major importance. Pregnant women should minimize contact with cats, especially strays or cats who eat raw meat; they should wash their hands after contact with cats and should have another person empty the litter box daily. In addition, pregnant women should wash all fruits and vegetables before eating them and should not eat undercooked meat. Serologic screening of pregnant women is advisable.

Ocular Toxoplasmosis

Toxoplasmosis may account for about 30% of cases of retinochoroiditis.[39] Many cases result from the reactivation of congenital infection; hence, ocular toxoplasmosis is most common in older children and young adults. Retinochoroiditis may also develop as a manifestation of primary infection.[40] Impairment of vision is the most common symptom, but pain and photophobia may accompany intense inflammation. Typical lesions appear as yellow-white fluffy exudates clustered in the posterior pole. Although a positive *Toxoplasma* serology is needed for the diagnosis of ocular toxoplasmosis, most patients have relatively low titers because the initial infection was acquired years earlier. In most cases, therefore, ocular toxoplasmosis is a clinical diagnosis that depends on the morphology of the lesions. Other conditions considered in the differential diagnosis include tuberculosis, sarcoidosis, syphilis, histoplasmosis, and candidiasis.

DIAGNOSIS

Serologic testing is important in evaluating patients with potential toxoplasmosis. Many tests are available; the most widely used are the indirect fluorescent antibody test and the Sabin-Feldman dye test. Many other tests are utilized by reference laboratories. Tests that measure IgG antibodies show positive results 1 to 3 weeks after infection and continue to show positive results for many years after infection. Diagnoses of recently acquired toxoplasmosis, especially pertinent during pregnancy, and of congenital toxoplasmosis in the newborn require assays of IgM anti-*Toxoplasma* antibodies. Some commercial IgM assays have been noted to generate false positive results and to detect IgM antibodies persisting for over 1 year, potentially confounding the reliability of these assays to detect only recent infections.[41] Positive IgM test results may necessitate confirmation by alternative testing assays, such as avidity testing; IgM, IgG, and IgG Western blot testing to detect new bands bound by infant but not maternal antibodies; and PCR.[42]

Because tissue cysts may be present in tissues for years, definitive evidence of active toxoplasmosis within biopsied tissues requires the detection of tachyzoites. These forms are not readily seen with conventional pathologic stains. Application of immunofluorescent or peroxidase–antiperoxidase antibody staining is needed. Tachyzoites are not usually detectable in biopsied lymph nodes, although characteristic histopathologic features can support the diagnosis.[43] PCR methods are highly sensitive for detecting the organism in amniotic fluid in congenital infections,[37] but neither PCR nor direct culturing of the organisms from blood or tissues has been widely used in diagnosing other forms of toxoplasmosis.

TREATMENT

Primary and AIDS-Associated Toxoplasmosis

Although the majority of patients with primary lymphadenopathic toxoplasmosis do not require treatment, chemotherapy should be considered for patients with unusually prolonged or severe illness. Patients with active chorioretinitis, CNS involvement, or disseminated toxoplasmosis should also be treated, as should immunosuppressed patients with toxoplasmosis, including patients with AIDS.

Combined administration of pyrimethamine and sulfonamides (sulfadiazine or trisulfapyrimidines) is the treatment of choice.[10] Other sulfonamides are less active against *T. gondii* and therefore should be avoided. In adults, a loading dose of 200 mg of pyrimethamine is given on the first day of treatment, followed by the usual dosage of 50 to 75 mg/day. Sulfadiazine or trisulfapyrimidines are usually given to adults in a loading dose of 4 g, followed by 1 to 2 g four times daily. The most common toxic effect of pyrimethamine is marrow suppression; leukocyte, red cell, and platelet counts should be monitored twice weekly, and 10 to 15 mg of folinic acid should be administered daily. Combined pyrimethamine and sulfonamide therapy is usually given for 3 to 6 weeks.

For AIDS patients who are intolerant of sulfonamides, clindamycin at a dosage of 600 mg orally or intravenously four times a day in combination with pyrimethamine has been effective in treating CNS toxoplasmosis.[44] Atovaquone plus pyrimethamine is also effective for sulfa-intolerant patients.[45]

Patients with AIDS who have been treated for toxoplasmosis and those who have not experienced active toxoplasmosis infection but have positive *Toxoplasma* serologies and $CD4^+$ T cell counts below $100/mm^3$ are at risk for reactivation of toxoplasmosis and require prolonged suppressive therapy. A variety of regimens are available for toxoplasmosis prevention in AIDS patients. A regimen consisting of sulfadiazine (500 to 1,000 mg four times a day) and pyrimethamine (25 to 50 mg/day) plus folinic acid (5 mg/day) is most effective.[46] Dapsone-pyrimethamine is also effective for patients intolerant of sulfonamides.[47] Some regimens used for the prevention of *Pneumocystis* pneumonia also provide primary prophylaxis for toxoplasmosis; ex-

amples include trimethoprim-sulfamethoxazole, dapsone-pyrimethamine, and, for sulfa-intolerant patients, atovaquone.[47] If the $CD4^+$ T cell count rises above 200/mm^3 for 3 months, secondary prophylaxis for toxoplasmosis can be stopped.[47]

Congenital Toxoplasmosis

Newly acquired toxoplasmosis documented during pregnancy presents difficult choices. Pyrimethamine is teratogenic and should be avoided in the first trimester. Spiramycin—available in the United States only directly from the manufacturer, Rhône-Poulenc Rorer—given at 3 to 4 g/day, can diminish the risk of transplacental infection.[10] If in utero infection is documented, therapy with pyrimethamine and sulfadiazine should be initiated. Therapeutic abortion might be considered if infection is acquired early in pregnancy; ultrasonography can be used to detect hydrocephalus, intracranial calcifications, and other signs of fetal damage. Children who are born with serologic or clinical evidence of congenital toxoplasmosis should be treated for a year with pyrimethamine and sulfadiazine.[36]

Ocular Toxoplasmosis

Pyrimethamine and sulfonamides are the mainstays of therapy for ocular toxoplasmosis, but results are unpredictable and relapses commonly occur. When there is a threat of visual loss, the patient should receive corticosteroids in combination with antimicrobials.

Intestinal Protozoan Infections

The human intestinal tract may serve as a host for several protozoan parasites. The intestinal protozoa *Entamoeba coli*, *E. hartmanni*, *Endolimax nana*, and *Iodamoeba bütschlii* are nonpathogenic and do not require therapy. Pathogenic intestinal protozoan parasites include five groups: (1) the flagellates (*Giardia lamblia* and *Dientamoeba fragilis*); (2) the amebae, or *Sarcodina* (*Entamoeba histolytica* and, possibly, *Blastocystis hominis*); (3) coccidia (*Cryptosporidium, Isospora, Cyclospora*); (4) a ciliate (*Balantidium coli*); and (5) microsporidia.

GIARDIASIS

G. lamblia inhabits the proximal small intestine; manifestations of *G. lamblia* infections can vary from no symptoms to profound malabsorption. *G. lamblia* and *Cryptosporidium* are the most common pathogenic intestinal protozoan parasites in the United States.[48]

Etiology and Epidemiology

G. lamblia exists in two morphologic forms: the trophozoite and the cyst [*see Figure 6*]. The trophozoite is a pear-shaped, multiflagellated organism that measures 9 to 15 µm long, 5 to 15 µm wide, and 2 to 4 µm thick. It is bilaterally symmetrical and contains two prominent nuclei. On the ventral surface is an adhesive disk with which the trophozoite attaches to the mucosal surface of the duodenum and jejunum. During passage in the bowel, the trophozoite usually encysts. The cysts measure 8 to 12 µm long by 7 to 10 µm wide, have a well-defined outer wall, and contain four nuclei when mature [*see Figure 6*].

Trophozoites may be seen in duodenojejunal fluid and in loose stools but generally are not found in formed stools [*see* Diagnosis, *below*]. Trophozoites are not resistant to external environmental stresses. In contrast, cysts are hardier and may survive in water for at least several months; decreased water temperature enhances their survival. Although infection can occur if trophozoites are ingested in quantities of food that are sufficient to buffer their transit through the stomach, the cyst stage is principally responsible for human infections. In studies, ingestion of as few as 10 cysts has resulted in human infection.

Infection derives from fecally excreted organisms and is spread by direct fecal-oral passage or by foodborne or waterborne transmission. Direct person-to-person transmission accounts for a heightened prevalence of giardiasis in several settings. In institutions where there is fecal incontinence and poor hygiene, giardiasis may be hyperendemic. Particularly in child day care centers, giardiasis can be a cause of intestinal disease.[49] The risk of acquisition and transmission is greatest for young children not yet toilet trained, who may be a source of additional secondary cases within their families.[49] Person-to-person transmission is also responsible for the prevalence of giardiasis in promiscuous male homosexuals. Sexual practices, including anilingus, can allow direct transfer of infectious cysts.

Waterborne transmission is a major source of giardiasis. Because filtration of water through soil removes *Giardia* cysts, deep well water is usually safe. In contrast, surface water, such as mountain streams and reservoirs, can harbor *Giardia* cysts, which are hardy in water and resistant to routine levels of chlorination. Coliform counts are not a reliable measure of giardial contamination. In addition, water-dwelling mammals, such as beavers, can become infected and then serve as continuing sources of water contamination, though whether giardiasis is a zoonotic disease is currently controversial.[50] It may be difficult to recognize a water supply as the common source of giardiasis, because the resultant infection is often asymptomatic.

G. lamblia is widely distributed throughout the world. Travelers to many countries, including those in developed areas, may acquire giardiasis.

Pathogenesis

After *G. lamblia* cysts have been ingested and have passed through the stomach, they liberate trophozoites that proliferate by binary fission. Trophozoites, which localize in the duodenum and the proximal jejunum, may attach to the microvillous border of intestinal epithelial cells by means of their ventral adhesive suckers. They may also reside in the unstirred layer above the epithelium, move around in the luminal contents, or, uncommonly, invade the space between the epithelial cells. Functional changes in the absorptive capabilities of the small bowel may develop. The activities of epithelial brush-border enzymes are diminished, leading to deficiencies in disaccharidases, including lactase deficiency. Jejunal biopsies of patients with giardiasis usually reveal no pathologic findings.[51] The pathogenetic mechanisms of these functional alterations in the small bowel remain uncertain. Both types of alterations return to normal with specific antigiardial therapy, albeit slowly in some patients. Hypochlorhydria predisposes persons to infection. Giardiasis often occurs with greater severity in patients with cystic fibrosis and in those with immunoglobulin deficiencies, possibly because of deficiencies in secretory IgA.[52]

Diagnosis

Clinical features The clinical manifestations of giardiasis may be quite varied. A significant number of persons with giardiasis are asymptomatic. Indeed, in one well-studied outbreak of waterborne giardiasis, two thirds of those infected were asymptomatic. In contrast, others with the infection experience typical acute giar-

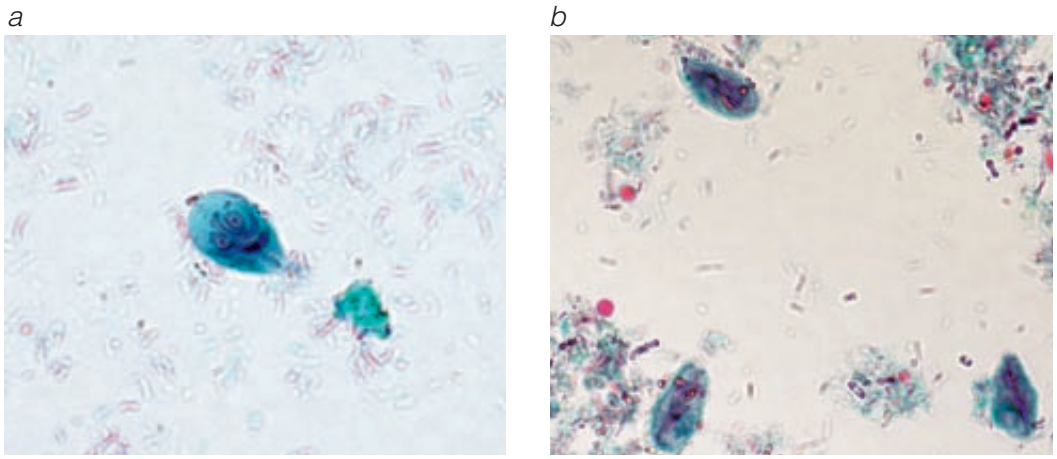


Figure 6 ***Giardia lamblia*** **trophozoites (*a*) and cysts (*b*).**

diasis. After a 1- to 3-week incubation period, the acute onset of the illness is marked by watery diarrhea; abdominal cramping (and other, frequently epigastric, discomfort); nausea (less commonly, vomiting); and systemic symptoms. Fever has been reported but is extremely unusual. Increased intestinal gas production leads to malodorous flatulence and sulfurous eructation. Impaired fat absorption and steatorrhea are common in symptomatic giardiasis. Stools are usually greasy and malodorous and float in water. Blood and mucus in stool are uncommon. These symptoms may be prominent for more than a week, diminishing in intensity over the ensuing weeks. Clinical findings that have helped identify cases of giardiasis in epidemiologic studies include a duration of illness lasting 7 or more days, with at least two of the following six symptoms: diarrhea, flatulence, foul-smelling stools, nausea, abdominal cramps, and excessive fatigue.[53]

A chronic phase of giardiasis may follow the acute phase or may become manifest without an antecedent acute illness. This chronic phase is characterized by loose, but usually not diarrheic, stools that are soft and greasy. Increased abdominal gassiness with cramping, borborygmi, flatulence, and burping occurs. Fever is uncommon, but malaise, fatigue, and depression may ensue. Lactose intolerance can develop with the infection and augment intestinal symptoms after ingestion of milk products. The course can be remitting; asymptomatic periods may alternate with exacerbations of symptoms. For a small number of patients, particularly children, the persistence of infection is associated with moderate to marked malabsorption and weight loss.[54]

Uncommonly, *G. lamblia* spreads from the duodenum to the biliary and pancreatic ducts. Cases of cholecystitis, cholangitis, and granulomatous hepatitis have been reported. Impaired exocrine pancreatic function, manifested by diminished secretion of trypsin and lipase, has been noted.

Laboratory findings Leukocytosis and eosinophilia do not occur in giardiasis. Fecal fat excretion is increased (see above), and the results of other laboratory tests of malabsorption may also be abnormal. An upper GI series usually shows no significant radiologic changes.

Definitive diagnosis of giardiasis requires the morphologic identification of the cyst or trophozoite forms of the parasite [*see Figure 6*]. Fecal examinations are usually positive in acute giardiasis, and evaluation of three sequential daily fecal samples can detect more than 90% of infections. Cysts, which may be present in either loose or formed stools, are hardy, so the search for them does not necessitate prompt analysis of stools. In contrast, examinations to detect trophozoites should be performed on fresh stools or stools preserved in polyvinyl alcohol or merthiolate-iodine-formaldehyde (MIF). Substances that interfere with fecal microscopic evaluations, such as barium, antacids, and mineral oil, should be avoided before stool examinations. Immunologic assays detect giardial antigens in stool with greater sensitivity than a single stool examination.[55]

In chronic giardiasis, the frequency of detection of giardial forms on stool examination diminishes. If three or more stool examinations are unrevealing, the upper intestinal contents should be sampled by duodenojejunal aspiration. Alternatively, a biopsy of the small bowel can be performed. Recognition of trophozoites in biopsied material may require diligent searching of processed tissue; direct examination of mucosal imprint smears from a biopsy can increase detection.

Differential Diagnosis

Other infectious agents that cause gastroenteritis must be considered early in the course of acute giardiasis. Because most of these agents produce illness of short duration, the persistence of symptoms after a week and the prominence of symptoms of malabsorption (flatulence, lactose intolerance, burping) suggest a giardial etiology. However, chronic giardiasis is difficult to distinguish from other chronic small bowel infections, such as *Cyclospora* or *Cryptosporidium*. Chronic giardiasis may resemble other diseases associated with malabsorption.[56]

Treatment

Metronidazole, although not approved by the Food and Drug Administration for giardiasis, is the principal agent used to treat this infection[10,57,58] because quinacrine, the first effective drug for giardiasis, is no longer distributed in the United States. The usual dosage of metronidazole is 250 mg orally three times a day for 5 days, though this may lead to recurrences in up to 40%, or 500 to 750 mg orally three times a day for 10 days, which is 60% to 95% effective.[58] Administration of 2 g of metronidazole once daily for 3 consecutive days is associated with the highest cure rates, yielding 93% to 100% efficacy.[58] Refractory cases can be cured by a combination of quinacrine (available through compounding pharmacies) and metronidazole.[57] Side effects of metronidazole include nausea, headache, and a metallic taste in the mouth; less commonly, dark urine, paresthesias, and dizziness occur. Metronidazole may have a disulfiram-like effect, so

alcohol consumption should be avoided when metronidazole is used. Although not available in the United States, tinidazole (given as a single 2 g oral dose) is highly effective.[58] Another alternative is nitazoxanide (in adults, 500 mg b.i.d.; in children 4 to 11 years of age, 250 mg b.i.d; and in children 1 to 3 years of age, 150 mg b.i.d. for 3 days). A randomized trial showed that for the treatment of symptomatic giardiasis in children, nitazoxanide has an efficacy comparable to that of metronidazole.[59]

For children, furazolidone, available as a suspension, is effective and tolerated.[10] Treatment of giardiasis in pregnancy can be difficult. Metronidazole is often avoided, although studies have not documented teratogenic risks of metronidazole during pregnancy.[60] If symptoms of giardiasis are minimal, therapy can be withheld until delivery. If symptoms are bothersome, one approach is to administer a nonabsorbable aminoglycoside, paromomycin, 25 to 35 mg/kg/day orally in three divided doses for 7 days.[10] This regimen may provide at least symptomatic relief. If giardiasis in pregnancy is associated with dehydration, malabsorption, or severe symptoms, therapy with metronidazole is warranted. In any patient, resolution of malabsorptive symptoms may require months for regeneration of functioning intestinal mucosa after effective antiparasitic therapy, and lactose intolerance may remain indefinitely.

Attention to hygiene is necessary to prevent person-to-person transmission of giardiasis. The risks and benefits of treating asymptomatic infected children in day care centers have not been fully defined; however, treatment of asymptomatic persons who pass cysts is indicated to prevent the spread of infection. Boiling water or heating it to at least 70º C (158º F) for 10 minutes renders water noninfectious. For hikers and campers, iodine-based water treatments are more effective than chlorine-based treatments; iodine disinfection must be carried out for at least 8 hours to be 99.9% effective. High-quality water-filtration units are effective for *Giardia* cyst removal.

DIENTAMOEBA FRAGILIS INFECTION

The large-intestine parasite *D. fragilis* has only a trophozoite stage and no cyst stage [*see Figure 7*]. Although previously grouped with the amebae, histologic and antigenic examination and ribosomal RNA homology have demonstrated that it is closely related to the trichomonad flagellates.[61,62] Because the trophozoite is not resistant to gastric acid, it is not clear how humans acquire infection. The eggs of the pinworm, *Enterobius vermicularis,* might transmit *D. fragilis* trophozoites because the two infections frequently coincide. *D. fragilis* can cause an illness characterized by abdominal pain, anorexia, and loose stools. As with *Isospora* infections, but not other protozoan infections, eosinophilia may accompany infection with *D. fragilis*.[63] The trophozoite stage is not hardy and is difficult to detect; to ensure detection, stool samples must be preserved in polyvinyl alcohol fixative, sodium acetate–acetic acid–formalin fixative, or Schaudinn fluid and must be examined after permanent staining. *D. fragilis* infection can be treated with tetracycline (500 mg q.i.d. for 10 days), paromomycin (25 to 30 mg/kg/day in three doses for 7 days), or iodoquinol (650 mg t.i.d. for 20 days).[10]

AMEBIASIS

Infection with the ameba *E. histolytica* is responsible for human amebiasis. These infections may be limited primarily to the colon. Infections range in severity from asymptomatic to markedly dysenteric and may involve extraintestinal sites, of which the liver is the most common.

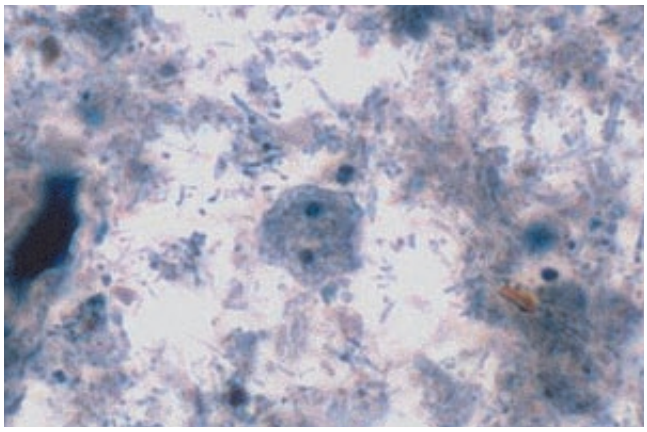


Figure 7 ***Dientamoeba fragilis* trophozoite. Note double nuclei.**

Etiology and Epidemiology

E. histolytica is distinguishable morphologically from the other nonpathogenic intestinal amebae, including *E. coli* and *E. hartmanni* [*see Figure 8*]. *E. histolytica* exists in two forms: as a trophozoite and a cyst. The trophozoite, which usually measures 10 to 20 µm in diameter but may be larger in dysenteric stools, is motile and possesses a single nucleus and a granular cytoplasm [*see Figure 9*]. Trophozoites are passed in loose stool, but this form is not hardy and does not survive outside the body. In contrast, the cyst stage, which arises within the colon from the trophozoite, can survive environmental stresses as well as passage through the acid of the stomach. Cysts measure 10 to 20 µm in diameter and contain one to four nuclei, which have small, centric karyosomes and a pattern of fine peripheral chromatin.

It has long been recognized that not all strains of *E. histolytica* are pathogenic. Nonpathogenic strains can be isolated from asymptomatic patients and have been prevalent in promiscuous male homosexuals in the United States and England. Pathogenic and nonpathogenic strains cannot be distinguished by microscopy, except that pathogenic trophozoites often phagocytose erythrocytes. On the basis of biochemical, immunologic, and genetic data, *E. histolytica* has been redescribed. A new spe-

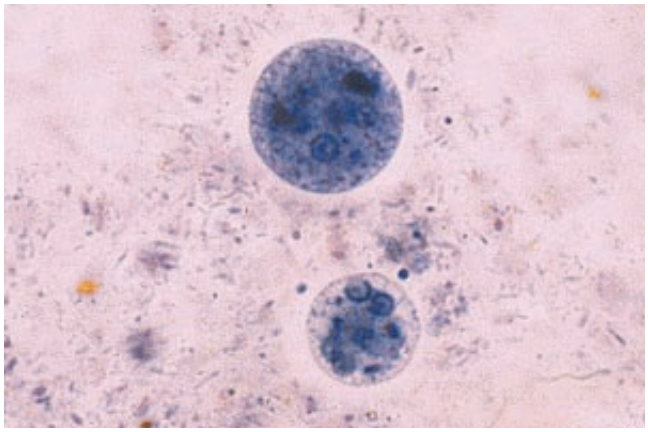


Figure 8 ***Entamoeba histolytica*, which causes amebiasis, is distinguishable morphologically from nonpathogenic intestinal amebae such as *E. coli;* on this photomicrograph, the cyst of *Entamoeba histolytica* is visibly larger than the cyst of *E. coli.***

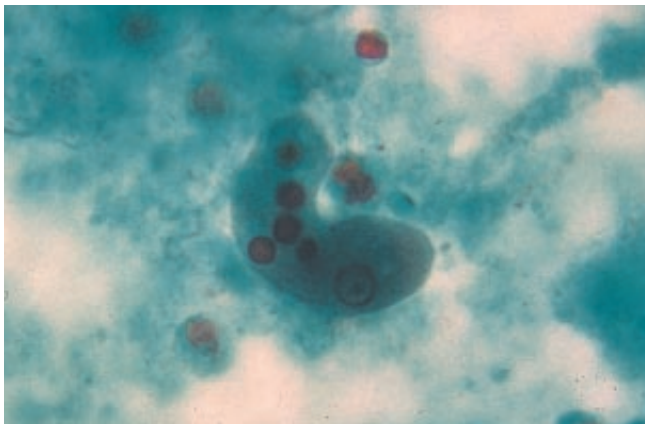


Figure 9 **Trophozoite form of *Entamoeba histolytica*.**

cies, *E. dispar*, now represents the nonpathogenic isolates and is apparently never invasive in humans, whereas *E. histolytica* includes only the potentially pathogenic strains.[64] Because not all potentially pathogenic *E. histolytica* strains invariably produce disease, other processes undoubtedly influence amebic virulence.

An immunoassay to distinguish *E. histolytica* from *E. dispar* has been developed, based on the *E. histolytica*–specific *N*-acetyl-D-galactosamine lectin.[65] There is currently no other method available for a commercial diagnostic laboratory to readily distinguish one organism from the other [*see* Diagnosis, *below*], though these organisms can be distinguished by differences in isoenzymes, restriction fragment patterns, repetitive DNA, and riboprinting.[66]

Cysts passed in human feces are primarily responsible for human infections. Acquisition of infection represents fecal-oral contamination and may occur by waterborne or foodborne transmission, as well as by person-to-person transmission. The latter accounts for the heightened prevalence of *E. histolytica* infection among promiscuous male homosexuals and in institutions where there is fecal incontinence and poor hygiene. Amebiasis is more common in lower socioeconomic groups than in the general population because of poor sanitation and overcrowding. In the United States, cases are seen in persons who have returned from international travel or have immigrated from areas where amebiasis is endemic.

Pathogenesis

Ingested cysts are carried into the intestine, where they excyst to liberate trophozoites that proliferate by binary fission within the colon [*see Figure 9*]. Trophozoites are cytolytic and may invade the bowel wall to produce local necrosis. The resultant ulcers are flask shaped, with a narrow neck through the mucosa and a broader submucosal base. Unless colonic involvement is extensive, intervening areas of bowel are normal. The areas most frequently affected are the cecum and the ascending colon, followed in frequency by the rectosigmoid, the appendix, the descending and transverse colon, and the terminal ileum [*see Figure 10*].

The severity of colonic involvement in patients with amebiasis is quite varied and may range from mild or negligible disease to diffuse and extensive tissue invasion and necrosis. Most persons who harbor *E. histolytica* experience no significant colonic invasion. The determinants of severity are not well understood but may include the inoculum size of *E. histolytica,* the coexistent colonic microbial flora, and the nutritional and physiologic state of the host. Corticosteroid use and pregnancy both diminish host resistance. Complications of colonic involvement include hemorrhage and peritonitis, the latter developing more commonly from transmural leakage across involved colonic tissue than from frank perforation. With chronicity, a granulomatous tissue response can develop at a site of infection (most commonly in the cecum) and can produce a mass lesion termed an ameboma. Colonic strictures may also develop.

Amebiasis may spread hematogenously from the bowel to involve any organ in the body. The liver is most commonly affected, followed in frequency by the lungs, which are principally affected as a result of transdiaphragmatic spread from the liver [*see Figure 10*]. Trophozoites carried via the portal venous system produce necrosis in the liver and cause abscess formation.

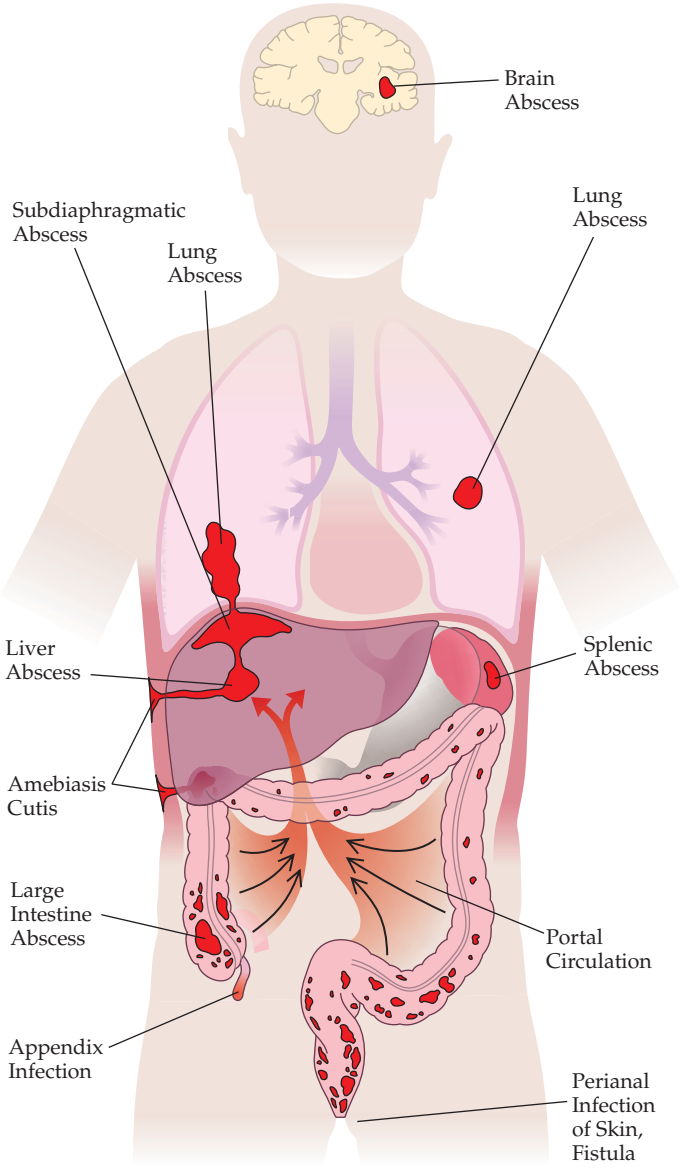


Figure 10 **Amebiasis from *Entamoeba histolytica* may be limited to the colon—most often, the proximal or terminal portions—or may spread by direct extension or hematogenously to any organ in the body. The liver is most commonly affected, followed by the lungs, but brain or spleen abscess may also occur, as may cutaneous involvement.**

About 90% of abscesses are in the right lobe, especially in the superior and anterior aspects. Although multiple abscesses can occur, a solitary abscess ranging in size from a few centimeters to 20 cm in diameter is more common. Amebic liver abscesses are seven to nine times more common in males than in females, although in children, the sex distribution is equal. Uncommonly, an abscess may develop concomitantly with amebic colitis. Depending on the reported series, 50% to 70% of patients with amebic-hepatic abscesses have no history of amebic colitis. Rupture of a right lobe abscess may result in extension into the chest, producing an amebic empyema, pneumonia, or a bronchopleural fistula. Extension into the peritoneal cavity is less common. Rupture from a left lobe abscess can extend into the pericardium, frequently with fatal consequences.

The role of immunity in protection of disease from *E. histolytica* has been controversial, but studies in children in endemic regions suggest that development of intestinal IgA antibodies to an *E. histolytica N*-acetyl-D-galactosamine lectin is highly associated with the development of acquired immunity.[67] This lectin appears essential for cell binding and pathogenesis by *E. histolytica,* and presumably, blocking its function in vivo leads to protection.

Diagnosis

Clinical features Most persons with colonic amebiasis experience no symptoms. Fecal cyst excretion in these asymptomatic individuals is detected only by serendipity. In persons who do experience symptoms, the illness may range from mild diarrhea to fulminant dysentery. The former presentation is common, and most patients are able to continue daily activities as they experience mild diarrhea that may alternate with constipation. The course of illness can be remitting, with subsequent symptomatic relapses. Abdominal discomfort, tenesmus, dull sacral pain, and flatulence are common, but systemic symptoms and fever are less prominent than what is common with bacterial colitis. Blood and mucus are frequently noted in stools. Abdominal findings reflect the severity of the colonic involvement. Tenderness may be absent or mild over the involved areas. In some cases, however, prominent abdominal tenderness, together with high fever and systemic toxicity, is a consequence of extensive colonic disease.[68] Amebomas may be palpable as tender abdominal masses.

In patients with amebic liver abscess, presenting symptoms usually include malaise, fatigue, anorexia, abdominal pain, fever, and weight loss; the duration is usually 1 week to several weeks or, less commonly, many months. The abdominal pain is usually dull and aching and localized over the right upper quadrant or right chest. At times, there is referred pain to the right shoulder. A presentation with pleuritic chest pain, cough, and dyspnea may mistakenly suggest an intrapulmonary infection. An abscess in the left lobe may produce midepigastric or left upper quadrant pain. On examination, hepatomegaly and hepatic punch tenderness or focal tenderness over the abscess are frequent. Frank jaundice is rare. On chest examination, dullness at the right lung base as a result of diaphragmatic elevation and pleural effusion formation can be appreciable. Rales may be heard.

Cutaneous amebiasis, resulting in ulcerative or fungating lesions, most commonly develops on the perineum and genitalia. Lesions arise from invasion by trophozoites derived from fecal contamination or sexual transmission. The lesions can be distinguished from neoplastic, tuberculous, and syphilitic processes by the presence of trophozoites in exudate or biopsied tissue.

Laboratory findings In symptomatic amebic colitis, a mild to moderate leukocytosis may develop. Mild anemia and mild elevations of liver enzyme levels (which are not usually indicative of incipient abscess formation) may occur. Feces usually will contain frank or occult blood.

Leukocytosis, anemia, and an elevated erythrocyte sedimentation rate are common with an amebic liver abscess. In more than two thirds of patients, alkaline phosphatase levels are one to four times higher than normal. Elevated aminotransferase levels occur in fewer than 50% of patients. Bilirubin levels may be elevated, but usually to no more than 2.5 mg/dl. Chest x-ray often shows elevation of the right diaphragm and may reveal a pleural effusion. An isotopic liver scan will reveal a space-occupying lesion; an ultrasound scan will show a round to ovoid hypoechoic lesion and can detect transdiaphragmatic pleural involvement. CT or MRI also can document the abscess and any local extension.

For intestinal amebiasis, definitive diagnosis requires morphologic or antigenic identification of the cysts or trophozoites of *E. histolytica* [*see Figures 8 and 9*]. Unformed stool should be examined immediately for motile trophozoites. Formed stools may be examined directly or after application of concentration techniques for cysts. Stools may also be preserved in polyvinyl alcohol or MIF for subsequent examination. Fecal leukocytes and nonpathogenic amebae are easily confused with *E. histolytica*, and thus, definitive speciation should be performed on stained samples by experienced personnel. Antimicrobial agents, cathartics, antacids, and barium interfere with microscopic detection. Testing of a single stool has only 33% to 50% sensitivity, so multiple stools should be examined when the diagnostic suspicion is high. Because the excretion rate of cysts varies daily, three or more stool samples from different days should be examined. About half of patients with symptomatic amebic colitis have rectosigmoid involvement; thus, aspirates, scrapings, or biopsies of mucosal lesions obtained at sigmoidoscopy can be examined for trophozoites. A fecal antigen detection test for *E. histolytica* has been introduced that is based on a monoclonal antibody to a specific lectin on this organism.[65]

Serologic tests for amebiasis are infrequently positive in asymptomatic cyst passers, but rates of seropositivity rise with increasing extent and duration of amebic colonic involvement. Such tests can be an adjunct to the diagnosis of acute amebic disease and can be especially helpful in the etiologic assessment of chronic colitis. Serologic tests, performed by a variety of methods, may remain positive for months to years after infection; this fact should be considered when interpreting a positive test result.

Serologic tests are positive in 90% to 95% of patients with extraintestinal amebiasis; titers increase with the duration of the disease. A positive titer, together with compatible clinical findings and tests demonstrating a cystic hepatic lesion, allows for the diagnosis of amebic hepatic abscess. Results of stool examinations, either negative or positive, are not etiologically pertinent in such cases. Trophozoites are rarely demonstrable in aspirated abscess fluid, which usually has an anchovy-sauce or chocolate-brown appearance.

Differential Diagnosis

In cases of mild intestinal amebiasis, the diagnosis of irritable bowel syndrome, diverticulitis, or regional enteritis may be suggested by the duration and symptoms of the infection. For more severe intestinal amebiasis with an acute presentation, infections

with *Shigella*, *Salmonella*, *Escherichia coli* O157:H7, and *Campylobacter* can be distinguished by positive stool cultures and by the presence of large numbers of fecal leukocytes, which are not found in amebic colitis. Ulcerative colitis and Crohn disease are to be considered in the differential diagnosis for chronic amebic colitis [*see 4:IV Inflammatory Bowel Diseases*]. Lesions on sigmoidoscopy and barium enema findings may be identical in amebic colitis and ulcerative colitis. It is critical to distinguish between these two conditions because corticosteroid therapy, which may be indicated for ulcerative colitis, would aggravate amebic colitis; therefore, amebic serologic tests and examinations of feces and mucosal lesions for amebae are of cardinal importance. A therapeutic trial of metronidazole cannot be relied on to identify the etiology, because metronidazole may have a salutary effect on inflammatory bowel disease as well as on amebic colitis.

An ameboma may simulate an adenocarcinoma or another granulomatous process. A positive amebic serology and resolution of the mass with metronidazole therapy support the diagnosis of an ameboma. If resolution is not complete, a biopsy is indicated to exclude a coincidental lesion of nonamebic cause.

The differential diagnosis of amebic hepatic abscess is guided by two observations: (1) an isotopic liver scan that shows a space-occupying lesion and (2) an ultrasound, CT, or MRI scan that shows a cystic lesion. Although hepatic cysts and echinococcal cysts are usually not associated with fever and the other symptoms of an amebic lesion, they should be considered in the differential diagnosis. Pyogenic abscesses should be considered as well. A positive amebic serology supports an amebic etiology. In contrast to echinococcal cysts, amebic abscesses calcify very rarely. Imaging studies of echinococcal cysts often demonstrate complicated cysts with internal daughter cysts, which are not seen in amebic abscesses. An amebic abscess and a pyogenic abscess may respond alike to metronidazole therapy. Amebic abscesses are uncommonly infected secondarily. If uncertainty persists about the bacterial or amebic etiology of an abscess, diagnostic aspiration of abscess fluid for bacterial cultures and Gram stain may be necessary.

Treatment

For symptomatic intestinal disease or extraintestinal disease, oral metronidazole (750 mg t.i.d. for 10 days) is usually highly effective and is the preferred therapy.[10] Alternatively, oral tinidazole (800 mg t.i.d. for 3 to 5 days) can be given. Information about the side effects of metronidazole and precautions about its use have been presented in the discussion of giardiasis (see above). Occasional failures in the treatment of hepatic abscesses with metronidazole have been noted. Because relapses of intestinal disease may infrequently occur in the absence of reinfection, follow-up is indicated for a number of months.

Chemotherapy may be unnecessary for many asymptomatic cyst passers because they often harbor nonpathogenic *E. dispar*. If speciation of *Entamoeba* is not available, asymptomatic cyst passers should be treated if they handle food or if they are receiving corticosteroids; they should also be treated in the setting of an amebiasis outbreak.[64] In asymptomatic patients, concentrations of metronidazole in the colonic lumen may be inadequate to eradicate amebae, and one of two luminal amebicides should be used.[10] Paromomycin (25 to 35 mg/kg/day for 7 days) can be used. An alternative is iodoquinol (650 mg t.i.d. for 20 days); it is important not to exceed this dosage because of the potential for causing optic neuritis. Iodoquinol is contraindicated for patients with optic neuropathy or thyroid disease. The administration of a course of iodoquinol is recommended for patients treated for symptomatic intestinal or extraintestinal amebiasis.[10] Another alternative is nitazoxanide, which in one study showed 96% efficacy for elimination of cyst passage[69]; controlled studies have not yet shown efficacy for *E. histolytica* infections.

Therapeutic aspiration is usually not necessary for amebic hepatic abscess, although diagnostic aspiration may be useful in certain cases [*see* Differential Diagnosis, *above*]. Drainage, which can be achieved by percutaneous aspiration, is indicated for those lesions that fail to respond to initial medical therapy or that are in imminent danger of rupturing.

Prevention of amebiasis relies on personal hygiene to prevent person-to-person transmission. In areas where amebiasis is endemic, the provision and use of adequate toilet facilities can decrease the spread of disease. Avoidance of vegetables that grow close to the ground (e.g., lettuce) is advisable because of potential contamination by human feces. Water can be rendered safe by boiling or by use of iodine-based water-treatment tablets.

BLASTOCYSTIS HOMINIS INFECTION

Although *B. hominis* was previously considered to be a nonpathogenic yeast, most investigators now consider this organism to be a protozoan distantly related to ameba or perhaps in its own phylogenetic group.[70] Whether *B. hominis* is capable of causing intestinal illness is controversial.[70] Support for the belief that this organism may be pathogenic stems from the finding of large numbers of *B. hominis* organisms in the feces of patients with diarrhea for which no other cause has been identified. Other investigators, however, have failed to confirm this finding.[70] They have found no concordance between numbers of fecal organisms and extent of the diarrhea and have seen no resolution of symptoms after therapy for *Blastocystis* infection. Until the issue of the pathogenicity of *B. hominis* is settled, patients with a diarrheal illness who are excreting this organism in their feces should be studied for other parasitic, bacterial, or viral infections or other reasons for illness, such as irritable bowel syndrome or inflammatory bowel disease. Stool samples typically reveal the vacuolated form of *B. hominis,* though several other forms can be seen as well [*see Figure 11*]. If diarrheal symptoms are sufficient to warrant therapy and other causes for diarrhea have been ruled out, iodoquinol, 650 mg three times a day for 20 days, or metronidazole, 750 mg three times a day for 10 days, can be used.[10]

COCCIDIOSIS

Coccidia, which are found in the intestines of many domestic and wild animals, are unicellular parasites that reproduce by asexual and sexual cycles in gut epithelium. Three coccidial organisms—*Isospora belli*, *Cryptosporidium*, and *Cyclospora cayetanensis*—have documented pathogenicity as enteric parasites for humans; humans serve as the definitive host and pass infectious cysts in stool [*see Figure 12*]. Contact with infected cattle may result in infection with *Cryptosporidium*, and contaminated water or food sources transmit *I. belli*, *Cryptosporidium,* and *C. cayetanensis*.[71] All three coccidial parasites have been identified as opportunistic pathogens in patients infected with HIV.[72]

Isosporiasis

Diagnosis Infections with *I. belli* usually begin abruptly. Fever and malaise appear first, followed by abdominal pain, diarrhea, and weight loss. In most cases, the illness is self-limited, although chronic infections may last several weeks to several months. Severe diarrhea, steatorrhea, and hepatic involvement

may ensue; in rare instances, severe disease results in death. In patients with AIDS, infection with *I. belli* causes a clinical picture of chronic watery diarrhea and weight loss that is indistinguishable from that produced by *Cryptosporidium*.

Histologic examination of mucosal lesions in patients who have isosporiasis shows shortened villi, crypt hypertrophy, and infiltration with eosinophils, neutrophils, lymphocytes, and plasma cells. Blood eosinophilia, which is not seen with other protozoan infections except *D. fragilis*, may develop with *I. belli* infections. The presence of oocysts in the stool establishes the diagnosis. Although routine stool examinations may fail to detect oocysts, they can be demonstrated with acid-fast staining. If oocysts in the feces are few, incubation of stool at room temperature for 24 to 48 hours can encourage oocyst maturation and the zinc sulfate concentration technique can be used before examining the stool. Parasite forms may also be detected in biopsied intestinal tissue and in intestinal contents.

Treatment *I. belli* infections are treated with double-strength trimethoprim-sulfamethoxazole (160 mg of trimethoprim and 800 mg of sulfamethoxazole) given orally four times a day for 10 days and then twice a day for 3 weeks.[10] Because recurrences of infection are likely in patients who have AIDS, the initial 10-day course of trimethoprim-sulfamethoxazole should be followed by long-term maintenance therapy with either one double-strength tablet of trimethoprim-sulfamethoxazole three times a week or the combination of 25 mg of pyrimethamine and 500 mg of sulfadoxine once a week in those patients. In patients who are intolerant of sulfa drugs, pyrimethamine (75 mg/day) or ciprofloxacin (500 mg b.i.d for 7 days) therapy has been successful; a maintenance dosage of 50 to 75 mg/day of pyrimethamine or 500 mg orally three times a week of ciprofloxacin has prevented relapses of *I. belli* infections.[72,73]

Cryptosporidiosis

Cryptosporidium inhabits the brush border of the small intestine mucosa and can cause enterocolitis in both normal and immunocompromised hosts. Infection may be acquired by ingestion of fewer than 100 oocysts.[74] A common means of transmission is by water, including municipal drinking water supplies[75] and recreational water (pools and water slides).[48,76,77] Cryptosporidiosis has caused outbreaks of diarrheal disease in day care centers and may be acquired by international travelers. Cryptosporidial infection and cryptosporidial disease have their highest incidences in children of the developing world.[76] Because fecal oocysts are infectious, infection can spread nosocomially, via food handlers, and within households.[76,78]

Diagnosis In normal hosts, illness begins after a mean incubation period of about a week. It consists of watery, nonbloody diarrhea that is accompanied at times by such clinical manifestations as abdominal pain, nausea, fever, anorexia, and weight loss. The diarrhea is generally noninflammatory in nature, though children in the developing world and small numbers of infected adults have white blood cells and lactoferrin in the stool. Symptoms may persist for 1 to 2 weeks and are usually self-limited. By contrast, in immunocompromised patients, cryptosporidiosis can be persistent and severe. In HIV-infected patients with $CD4^+$ T cell levels greater than 180/mm^3, cryptosporidiosis can be self-limited. With more profound immunocompromise, however, the secretory diarrhea, which is chronic and profuse, is usually unremitting. In these persons, *Cryptosporidium* organisms may also cause hepatobiliary disease, including cholecystitis, cholangitis, and papillary stenosis.

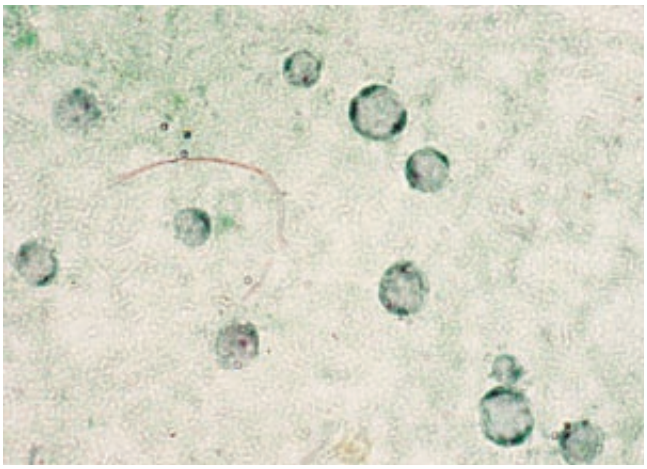


Figure 11 ***Blastocystis hominis*** **on a trichrome stain.**

Cryptosporidium oocysts, which are 4 μm in size, can be detected in fecal smears examined microscopically, either as an iodine wet mount or after staining with a monoclonal antibody or the modified Kinyoun acid-fast reagent [*see Figure 12*]. In addition, direct fluorescent stains and immunoassays of fecal samples can enhance diagnostic yields of *Cryptosporidium*.

Treatment In immunocompetent patients, cryptosporidiosis is a self-limited illness and usually requires only supportive therapy. Chemotherapy would be valuable in immunocompromised patients, but an effective regimen for cryptosporidiosis has not been established.[26] If the patient is receiving immunosuppressive drugs, cessation of these agents may lead to resolution of the diarrhea. Similarly, improvement of $CD4^+$ T cell levels in HIV-infected patients by highly active antiretroviral therapy has led to the cessation of life-threatening cryptosporidial diarrhea.[79] For some HIV-infected patients, paromomycin may be at least partially beneficial in treating cryptosporidiosis, although small controlled trials were unable to show a difference between paromomycin and placebo.[26] Perhaps a better alternative, nitazoxanide (in adults, 500 mg b.i.d.; in children 4 to 11 years of age, 250 mg b.i.d; in children 1 to 3 years of age, 150 mg b.i.d.), is approved by the FDA for the treatment of cryptosporidiosis and giardiasis, albeit in children only. Three randomized trials have shown efficacy in cryptosporidiosis.[80]

Cyclosporiasis

Cyclospora appears to be widely distributed geographically; illness attributable to this protozoan has been described in the United States, Latin America, Africa, Europe, and Asia.[81] Epidemiologic studies indicate that contaminated water[71] and contaminated produce such as raspberries, blackberries, mesclun, and basil[82] have been sources of infection. The oocysts that are passed in stool require days to weeks outside the host in the right environmental conditions to sporulate and become infectious. This implies that person-to-person transmission is unlikely. The median incubation period after ingestion of the infectious sporocyst is about 1 week but is possibly as short as 1 to 2 days.[82,83]

Diagnosis Many patients infected with *Cyclospora* experience prodromal flulike symptoms, diarrhea, and symptoms

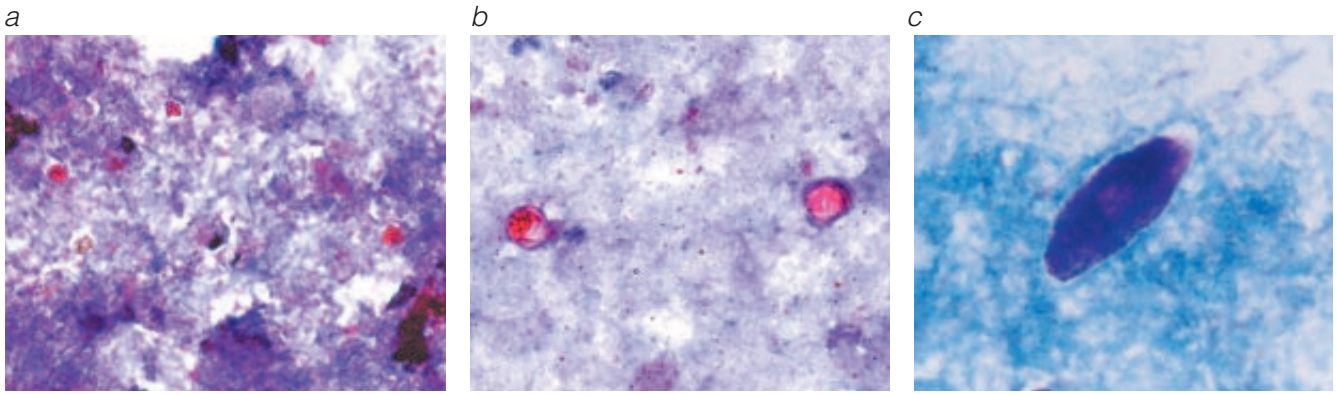


Figure 12 **Cysts of the three coccidial organisms—(*a*) *Cryptosporidium*, (*b*) *Cyclospora*, and (*c*) *Isospora*—with documented pathogenicity as enteral parasites. All are stained with modified Kinyoun (acid-fast) stain. Note that cysts of *Cryptosporidium* and *Cyclospora* can be distinguished only by their size (*Cryptosporidium* is 4 μm in diameter, and *Cyclospora* is 8 μm), whereas *Isospora* cysts are 25 × 15 μm and ovoid in shape.**

common to other small bowel pathogens, including nausea, vomiting, flatulence, and burping.[82] The illness may be confined to a single self-limited episode or may wax and wane, but prolonged diarrhea, anorexia, and upper GI symptoms often occur. Prolonged fatigue and weight loss also frequently occur.

Small-bowel biopsies can detect the parasite in epithelial cells and can detect jejunal inflammation, which is associated with increased numbers of intraepithelial lymphocytes and increased degrees of villous atrophy and crypt hyperplasia. Fecal leukocytes and blood are absent, suggesting that disease involves a noninvasive mechanism. The diagnosis can be made by detection of oocysts in the stool, which, like *Cryptosporidium* oocysts, are apparent in acid-fast stains. Although *Cyclospora* oocysts, which measure 8 to 10 μm in diameter, are larger than *Cryptosporidium* oocysts, caution is needed so that diagnostic testing does not confuse the two distinct protozoan organisms [*see Figure 12*]. Fluorescence microscopy is a rapid and sensitive means to detect oocysts, which are autofluorescent.[84] *Cyclospora* oocysts have been found in patients infected with HIV, and infection in these patients can range from asymptomatic to severe.[72,82]

Treatment Double-strength trimethoprim-sulfamethoxazole (160 mg/800 mg) tablets twice daily for 7 days has proved to be effective therapy for cyclosporiasis.[10] HIV-infected patients, however, may require a higher dosage (four times a day for 7 days) and may need long-term maintenance treatment (three times a week).[10,82] Ciprofloxacin (500 mg b.i.d. for 7 days, then 500 mg three times a week) has been shown to be less effective than trimethoprim-sulfamethoxazole in HIV-infected persons but can be used in patients who are sulfa intolerant.[73] Like cryptosporidiosis, *Cyclospora* infection must be considered in any patient with prolonged diarrhea, anorexia, and upper GI symptoms.

BALANTIDIUM COLI INFECTION

Balantidium coli is the only ciliate that causes human disease. *B. coli* is a rare cause of diarrhea and inflammatory colitis. Humans are incidental hosts. Large mammals, such as pigs, and contaminated food or water are the main sources of human infection. The ciliated trophozoite is found in diarrheal stools and is usually 60 to 70 μm and ovoid in shape; cysts may also be found [*see Figure 13*]. Clinical symptoms usually consist of chronic intermittent diarrhea and weight loss, but acute dysentery occurs in about 5% of cases.[61] In these latter cases, superficial and, rarely, deep colonic ulcerations have been observed. Treatment with tetracycline (500 mg q.i.d for 10 days) or, possibly, metronidazole (750 mg t.i.d. for 5 days) or iodoquinol (650 mg t.i.d. for 20 days) is effective.[10]

MICROSPORIDIOSIS

Microsporidia, which are obligate intracellular, spore-forming organisms, belong to a distinct phylum of protozoans that includes many genera capable of infecting diverse vertebrate and invertebrate hosts. They are most closely related to fungi, although their exact phylogenetic relationship to other eukaryotes remains unclear.[85] Human microsporidial infections have been recognized in recent years, and to date, seven genera of microsporidia—*Enterocytozoon*, *Encephalitozoon* (including *Septata*), *Vittaforma*, *Trachipleistophora*, *Pleistophora*, *Nosema*, and *Brachiola*—have been identified as causes of human disease, especially in persons infected with HIV.[86-88] These microsporidia are differentiated by their size, nuclear morphology, and mode of division, as well as by the intracellular site of proliferation (microsporidia multiply either freely in the cytoplasm or within membrane-bound vacuoles). Despite the ubiquity of microsporidia in other host species, how humans become infected is not known. Because almost all microsporidial infections in humans have been identified in immunocompromised hosts, it is

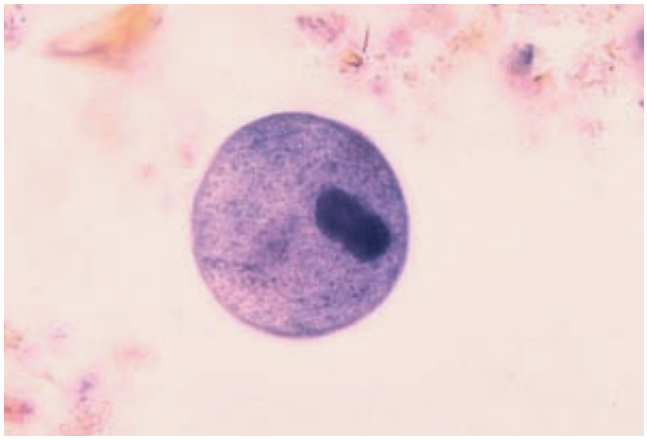


Figure 13 ***Balantidium coli* cyst.**

not clear how frequently immunocompetent hosts are infected or whether infections that develop in immunocompetent hosts are symptomatic or self-limited.

The spectrum of disease attributable to microsporidia is broad, with the disease state apparently depending on the infecting species and the immune status of the host. The most commonly recognized microsporidial infections are *Enterocytozoon bieneusi* infections in HIV-infected patients, usually those with $CD4^+$ T cell levels below $100/mm^3$. In HIV-infected patients, intestinal *E. bieneusi* and *Encephalitozoon intestinalis* (formerly *Septata intestinalis)* infections are causes of chronic diarrhea,[86,88,89] and *E. bieneusi* and *E. intestinalis* may infect the biliary tract, causing cholangitis. *Encephalitozoon* species have caused keratoconjunctivitis marked by a coarse punctate epithelial keratopathy in HIV-infected patients, whereas *Vittaforma* and *Nosema* species have caused stromal keratitis in a few HIV-seronegative patients. *Encephalitozoon* microsporidia have also been associated with peritonitis and hepatitis, as well as nasal and sinus infections and infections in other diverse sites in HIV-infected patients. *Trachipleistophora*, *Pleistophora,* and *Brachiola*[90] species have caused myositis.

The small size of microsporidia, which are gram-positive organisms with mature spores that measure 0.5 to 2 µm by 1 to 4 µm, hinders detection of the parasite. Many microsporidial infections have required electron microscopic tissue evaluation for diagnosis [*see Figure 14*]. Intracellular spores can also be recognized by light microscopy if tissues are stained with hematoxylin-eosin, Giemsa, Gram stain, or modified acid-fast stains. Intestinal microsporidia appear to be spottily distributed and may not be detected on examination of biopsied tissues. Chromotrope-based and fluorochrome (Uvitex 2B and Calcofluor) staining methods facilitate detection of microsporidial spores in smears of either feces or duodenal aspirates.[86]

For intestinal infections with *E. intestinalis*, oral albendazole (400 mg b.i.d. for 21 days) has been beneficial or curative; *E. bieneusi* infections are less responsive, although patients may experience symptomatic improvement without eradication of the infection.[88,89] Oral albendazole has also been used for systemic infections with *E. hellum, E. cuniculi, E. intestinalis, Pleistophora, Trachipleistophora,* and *Brachiola vesicularum.*[10] Oral fumagillin (60 mg daily for 14 days) has been effective in treating *E. bieneusi* infections.[10] For keratoconjunctivitis caused by *Encephalitozoon hellem*, topical therapy with fumagillin suspension combined with oral albendazole (400 mg b.i.d.) has been beneficial.[10]

Infections from Free-Living Amebae

Although amebae such as *E. histolytica* cannot survive and replicate outside of animal hosts, most amebae are free-living in soil or water. Amebae of the genera *Naegleria, Acanthamoeba, Balamuthia,* and *Sappinia* can cause acute meningitis, acute meningoencephalitis, or chronic granulomatous meningoencephalitis [*see Figures 15 and 16*].[91-93] *Naegleria* species, notably *N. fowleri*, cause acute meningoencephalitis in immunocompetent hosts; infection has been reported after trauma in warm freshwater in the southeastern United States. Progression of disease is typically rapid and inexorable, although effective therapy has been reported with amphotericin B, miconazole, and rifampin.[94]

In addition to chronic granulomatous meningoencephalitis, *Acanthamoeba* species have been recognized as a cause of keratitis[95] and, in a small number of HIV-infected patients, of disseminated disease with cutaneous manifestations.[96] *Acanthamoeba* organisms have been isolated from water, airborne dust, hot tubs, and saline solutions used to clean contact lenses. Factors associated with the development of amebic keratitis include the lack of effective disinfection of contact lenses, a history of minor corneal trauma, and exposure to soil or standing water. The lesions, which are usually chronic and severely painful, consist of variable anterior uveitis, epithelial erosion, scleritis, and an infiltrative stromal keratitis that is often ring shaped. Lesions are refractory to the usual antimicrobial medications and must be distinguished from keratitis caused by herpes simplex virus [*see 7:XXVI Herpesvirus Infections*]. The diagnosis can be made by microscopic examination of Giemsa- or trichrome-stained corneal scrapings and by the use of indirect immunofluorescent antibody staining of corneal scrapings. The amebae can be cultured by inoculating corneal tissue into nonnutrient agar seeded with *Escherichia coli.* Acanthamebic keratitis has been treated with topical regimens using chlorhexidine (bis-biguanide) and propamidine or the polymeric equivalent polyhexamethylene biguanide (PHMB). PHMB was originally combined with propamidine but is now combined with hexamidine.[95] This topical therapy is usually effective for patients presenting within 8 weeks, and cure can be achieved in 4 weeks of therapy.[95]

Trichomoniasis

Trichomonas vaginalis is a flagellated protozoan that causes an estimated three million vaginal infections a year.[97] It is a venereal disease, with the highest incidence in women who have multiple sexual partners; thus, persons with *Trichomonas* infection should be screened for other sexually transmitted pathogens, such as *Chlamydia*, *Neisseria gonorrhoeae*, and HIV. *Trichomonas* infection can be passed to neonates, and 2% to 17% of infected women transmit it to their female offspring during birth. It does not have a cyst form but, rather, only the trophozoite form, so that human-to-human transmission is the norm, although *T. vaginalis* can exist outside of the host for several hours if it is in a moist environment. Trichomonads appear to damage genital epithelium by direct contact, and this results in microulcerations

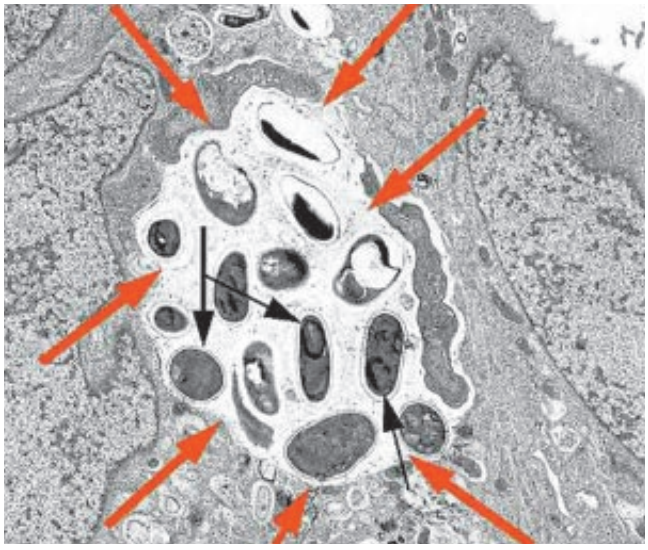


Figure 14 **A transmission electron micrograph shows developing forms of *Encephalitozoon intestinalis* inside a parasitophorous vacuole (red arrows) with mature spores (black arrows).**

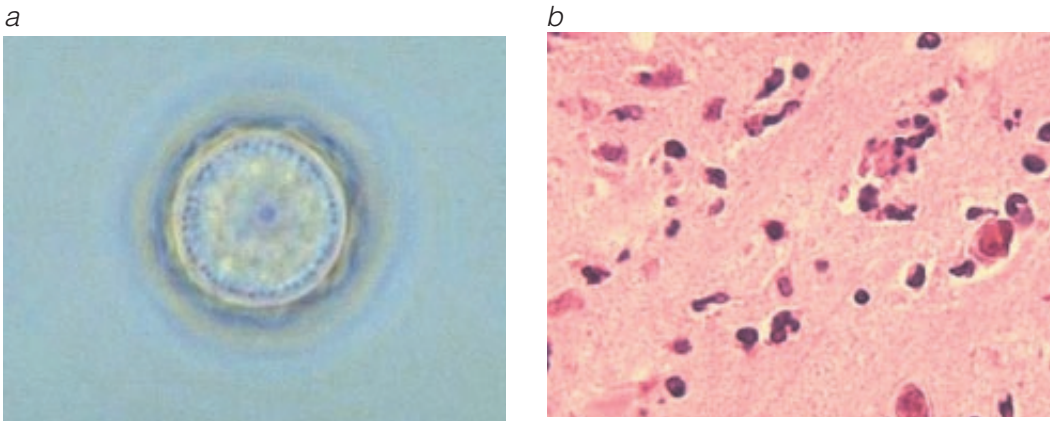


Figure 15 **(*a*) *Acanthamoeba polyphaga* cyst. (*b*) A histopathologic slide shows *A. polyphaga* infection in a mouse brain. Similar histopathologic features are seen in *Acanthamoeba* meningoencephalitis, which generally occurs in immunocompromised persons.**

and inflammation. Management of trichomonal infection is discussed in detail elsewhere [*see 7:XXII Vaginitis and Sexually Transmitted Diseases*].

Leishmaniasis

Leishmania organisms are protozoan hemoflagellates that are obligate intracellular parasites in humans. *Leishmania* species produce a wide spectrum of disease, ranging from generalized visceral involvement to diffuse or circumscribed cutaneous or mucocutaneous lesions. Four species complexes of *Leishmania* may infect humans: *L. donovani*, *L. tropica*, *L. mexicana*, and *L.* (subgenus *Viannia*) *braziliensis*. The resulting patterns of illness arise from the tissue tropism of the leishmanial species and the host's immune response, principally the cell-mediated component of immunity.

VISCERAL LEISHMANIASIS

Etiology and Epidemiology

The *L. donovani* species complex includes several species (e.g., *L. infantum* and *L. chagasi*). These species cause visceral leishmaniasis, or kala-azar, which is endemic in areas of India, China, Central and South America, East and West Africa, and the countries surrounding the Mediterranean. *L. tropica* can also cause a viscerotrophic disease involving bone marrow cells.[98] Sandflies of the genus *Phlebotomus* are the insect vectors that spread *L. donovani*; the species vary in the different areas. In India, no extrahuman reservoirs are known, but in other regions, infection may involve several mammalian species, including dogs, foxes, and wild rodents.

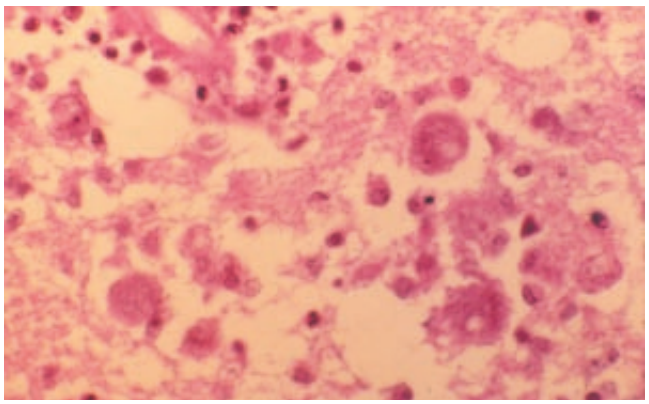


Figure 16 ***Naeglaria* meningoencephalitis in a human brain, on hematoxylin-eosin stain.**

Pathogenesis

The flagellated promastigotes of *L. donovani* are introduced by an insect bite. After entering macrophages of the reticuloendothelial system, these forms change into amastigotes, which multiply in phagocytic cells. Released amastigotes disseminate hematogenously and invade reticuloendothelial cells in the spleen, liver, lymph nodes, bone marrow, and skin. Prospective studies have demonstrated that the ratio of inapparent infection to disease ranges from greater than 6.5:1 in children younger than 5 years, the most susceptible group, to greater than 18:1 in older children and adults.[99] Cell-mediated immunity controls *Leishmania* infection, and compromise of cell-mediated immunity, such as from young age or malnutrition, contributes to susceptibility.

Diagnosis

Clinical features Symptoms of visceral leishmaniasis usually have a gradual onset several months after infection and include weakness, dizziness, weight loss, diarrhea, and constipation. Fever, which almost always develops, may spike twice daily and is sometimes accompanied by chills and sweating. As the disease progresses, the liver and spleen enlarge, the latter often expanding into the iliac fossa. When bone marrow macrophages are parasitized, anemia and leukopenia ensue. The thrombocytopenic patient may bleed from the gingivae, nose, or GI tract, and ecchymoses and petechiae may appear on the skin. Death can result from secondary bacterial infections, severe anemia, or uncontrolled bleeding. Latent infection can become manifest and progressive during immunosuppression, and visceral leishmaniasis can develop as an opportunistic infection in HIV-infected patients. Two thirds of patients with visceral leishmaniasis have typical infections, but leishmanial parasites may localize in unusual sites, including the larynx and throughout the GI tract, and hepatosplenomegaly may be absent.

Laboratory findings Anemia, leukopenia, thrombocytope-

nia, hyperglobulinemia, and hypoalbuminemia suggest visceral leishmaniasis when they are observed in a patient with fever, hepatosplenomegaly, and a history of exposure in endemic areas. The differential diagnosis is wide, including hepatosplenic schistosomiasis, malarial hypersplenism, myeloproliferative diseases, typhoidal *Salmonella* infections, miliary tuberculosis, brucellosis, histoplasmosis, subacute bacterial endocarditis, and infectious mononucleosis. Definitive diagnosis of visceral leishmaniasis requires demonstration of the organism in host tissues cultured on a Novy-MacNeal-Nicolle (NNN) or other medium or detection of Leishman-Donovan bodies (amastigotes) in stained tissue samples. Alternatively, PCR can be performed using genus- or species-specific oligonucleotides. In most cases, the diagnosis can be established by examining bone marrow aspirates. Splenic aspirates have the highest yields but may be risky. Liver biopsy or aspiration of enlarged lymph nodes can also provide diagnostic material.

Treatment

Sodium stibogluconate (pentavalent antimony, 20 mg/kg daily I.M. or I.V. for 20 to 28 days) is the therapy of choice for treating most cases of leishmaniasis and is available as Pentostam through the CDC Drug Service (404-639-3670, days; 404-639-2888, nights and weekends) [*see Sidebar,* Protozoan Infection Information on the Internet].[10] Miltefosine has been shown to be effective in India for treatment of visceral leishmaniasis.[100] The dose of miltefosine is 2.5 mg/kg/day, preferably in two divided doses or a single dose, orally for 28 days. When initial treatment of kala-azar with sodium stibogluconate fails, amphotericin B or pentamidine may be used.[10] Kala-azar that is resistant to sodium stibogluconate may respond to liposomal amphotericin B.[101]

CUTANEOUS AND MUCOCUTANEOUS LEISHMANIASIS

Etiology and Epidemiology

Old World cutaneous leishmaniasis is caused by three species of *Leishmania* that belong to the *L. tropica* complex: *L. tropica* is present in the Middle East and the Mediterranean littoral; *L. major* is found in the Middle East, Arabia, the former Soviet Union, India, and sub-Saharan Africa; and *L. aethiopica* is found principally in Ethiopia and Kenya. *Phlebotomus* sandflies are the principal vectors, although direct contact with an ill person may, in rare cases, result in infection. Infections that are caused by *Leishmania* can be acquired by travelers, as well as by military and other personnel residing in endemic areas. Military personnel in the Middle East have acquired cutaneous leishmaniasis with *L. major* and viscerotropic infections with *L. tropica*.[98]

New World cutaneous leishmaniasis arises from infection with parasites belonging to the *L. mexicana* group or the *L. braziliensis (Viannia* subgenus) group. The *Viannia* subgenus is distinguished from the *Leishmania* subgenus by the differences in development in the sandfly gut. The patterns of illness vary with the nature of the infecting leishmanial organisms, which are found in different regions of North, Central, and South America [*see Table 3*]. In areas of Central and South America, infection with organisms of the *L. mexicana* group produces cutaneous leishmaniasis. A few autochthonous cases have been found in Texas. Infections with strains of *L. viannia*, which are endemic in various areas of South America, cause cutaneous leishmaniasis and, in a small percentage of those infected, result in the later development of mucocutaneous leishmaniasis. Such mucocutaneous disease (espundia) involves the nasal or oropharyngeal mucosa, or both, and may prove fatal. All of these New World leishmanial parasites are transmitted principally by sandfly vectors, although direct human contact may also bring about infection. Various mammals are naturally infected reservoirs of the organisms.

Pathogenesis

Both Old World and New World forms of leishmaniasis are initiated when the bite of an infected sandfly injects promastigotes into the human host. The organisms enter tissue macrophages and capillary endothelial cells, become amastigotes, and multiply. A granulomatous inflammatory response develops at the bite site. With local ischemia, the lesion ulcerates [*see Figure 17*]; a bacterial infection of the necrotic area may extend the ulceration. Resolution of clinical infection is associated with $CD4^+$ T helper type 1 cells that secrete interferon gamma in response to *Leishmania*. Progression of disease appears to be associated with an immune response dominated by interleukin-10, a cytokine that suppresses other cytokine responses.

Diagnosis

Clinical features In Old World cutaneous leishmaniasis, after an incubation period of weeks to months, a papule develops at the inoculation site. This area may resolve spontaneously. More frequently, it ulcerates and a shallow circular lesion appears that is several centimeters in diameter and has a raised margin. Bacterial superinfection may lead to regional lym-

Protozoan Infection Information on the Internet

Centers for Disease Control and Prevention
- CDC Division of Parasitic Diseases
  - **http://www.cdc.gov/ncidod/dpd**
  - Information on parasitic diseases, including DPDx, an online interactive diagnostic service (**http://www.dpd.cdc.gov/dpdx**)
- CDC Drug Service
  - **http://www.cdc.gov/ncidod/srp/drugs/drug-service.html**
  - Information on special immunobiologic agents and drugs distributed through the CDC Drug Service, Scientific Resources Program, and the Division of Quarantine of the National Center for Infectious Diseases
- Emerging Infectious Diseases
  - **http://www.cdc.gov/ncidod/eid**
  - The online edition of the peer-reviewed journal published by the CDC's National Center for Infectious Diseases (**http://www.cdc.gov/ncidod**)

World Health Organization
- Division of Control of Tropical Diseases
  - **http://www.who.int/ctd**
  - Scientific publications and other information from the WHO's lead program for the control of tropical diseases
- Special Programme for Research and Training in Tropical Diseases
  - **http://www.who.int/tdr**
  - Disease information, image library, publications, research guidelines, grant applications, and other information from a scientific collaboration of the United Nations Development Program, UNICEF, the World Bank, and the WHO

Karolinska Institutet
- University Library: Diseases and Disorders—links pertaining to parasitic diseases
  - **http://www.mic.ki.se/Diseases/C03.html**

Table 3 New World Cutaneous Leishmaniasis[97]

Parasite	*Disease in Humans*	*Geographic Areas*
Leishmania viannia group		
L. viannia braziliensis	Usually a single or few lesions but frequently very large, persistent, and disfiguring; later nasopharyngeal involvement (espundia) in a small percentage	Brazil, Peru, Ecuador, Bolivia, Venezuela, Paraguay, Colombia, northern Argentina, Belize, Guatemala
L. viannia colombiensis	Usually a single lesion	Colombia, Panama
L. viannia guyanensis	Pian bois (bush yaws); multiple skin lesions, frequently with metastatic spread along lymphatics	Guyana, Suriname, Belize, northern Brazil
L. viannia panamensis	Usually a single lesion but occasional spread via lymphatics; nasopharyngeal involvement rare	Panama, Costa Rica, Colombia
L. viannia peruviana	Uta; a single or few self-healing skin lesions; no nasopharyngeal involvement	Western slopes of the Peruvian Andes, the Argentinian highlands
L. mexicana group		
L. mexicana mexicana	Frequent involvement of ear pinna (chiclero's ear, bay sore); a single or few skin lesions; no nasopharyngeal lesions; diffuse cutaneous leishmaniasis rare	Mexico, Central America, Texas
L. mexicana amazonensis	Rare human infection; a single or few skin lesions; diffuse cutaneous leishmaniasis	Amazon basin and neighboring areas, Brazil
L. mexicana pifanoi	Simple and diffuse cutaneous leishmaniasis	Venezuela
L. mexicana venezuelensis	Simple cutaneous leishmaniasis	Venezuelan Andes
L. mexicana species	Simple and diffuse cutaneous leishmaniasis	Dominican Republic

phadenopathy. The lesions are often solitary, but multiple bites can produce several concurrent lesions. Healing of the lesions is slow, sometimes requiring more than a year.

L. mexicana infections produce a single lesion or a few lesions on exposed surfaces of the body such as the face and ear.[102] The ulcer usually heals spontaneously over 6 months. An ulcer involving the ear, however, may cause extensive destruction of the pinna. *L. viannia* infection is associated with lesions on the skin or mucous membranes, which may be multiple and may become very large, especially when bacterial superinfection develops. Cutaneous lesions caused by *L. braziliensis* are much less likely to heal spontaneously than those caused by *L. mexicana*.[102] *L. viannia braziliensis* species can invade regional lymph nodes and cause progressive ulcerations along the lymphatics or extend locally and involve mucous membranes. Often, the infection metastasizes to the nasal or oral mucosa after an intervening period of months to years. Metastatic lesions can erode the nasal septum or the hard palate or soft palate. Some patients die of malnutrition or bacterial infection.

Diffuse cutaneous leishmaniasis occurs in parts of Ethiopia, Venezuela, Brazil, and the Dominican Republic. The initial nodule does not ulcerate; instead, multiple nodules evolve on the body. Leishmanial organisms abound in the lesions. Patients with this form of leishmaniasis have a deficiency of cell-mediated immunity, which is similar to the defective immunity that occurs in patients with lepromatous leprosy.

Laboratory findings Definitive diagnosis is made by demonstrating amastigotes on stained smears [*see Figure 18*] of a biopsy or of scrapings from the border of an ulcer. Alternatively, the diagnosis can be made by culturing amastigotes on NNN medium inoculated with lesion material. Use of PCR targeting parasite kinetoplast DNA has allowed detection of organisms that might be missed on histologic section or culturing. Moreover, this technique serves as a rapid method for speciating *L. mexicana* and *L. viannia*. In contrast to *L. tropica* and *L. mexicana* amastigotes, which can be readily cultured and are abundant in lesions, *L. viannia* amastigotes are difficult to culture and are sparse in lesions, especially those of mucocutaneous leishmaniasis. Organisms of these species cannot be distinguished morphologically, and specialized immunologic, enzymatic, or nucleic acid studies may be needed for definitive speciation. Except in diffuse cutaneous leishmaniasis, the leishmanin skin test is usually positive.

Treatment

Although the pentavalent antimonial compounds, sodium stibogluconate (20 mg/kg/day I.V. or I.M. of pentavalent antimonial for 20 days) and meglumine antimoniate, are the general treatments of choice for both Old World and New World cutaneous leishmaniasis,[10] optimal therapeutic regimens with these or alternative agents have not been rigorously defined. Individual species and geographic strains of *Leishmania* respond differently to treatment and have different capacities to cause mucosal

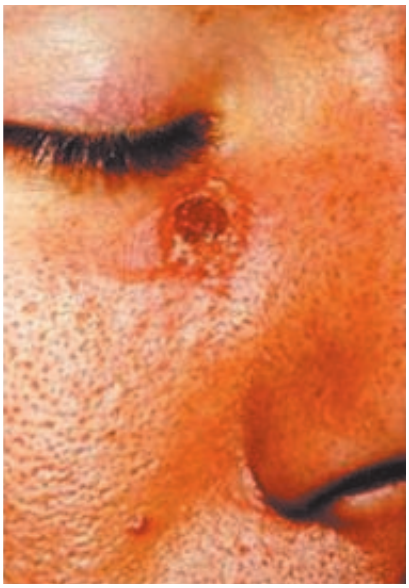


Figure 17 **An ulcerative lesion of *L. (Viannia) braziliensis* acquired in the jungles of Belize.**

disease or to heal spontaneously. Furthermore, it is difficult to distinguish infective species solely on clinical grounds. Therefore, progress in developing rational plans for appropriate treatment has been impeded. Advice on treatment of leishmaniasis is available from the CDC's Division of Parasitic Diseases [*see Sidebar,* Protozoan Infection Information on the Internet]. *Leishmania* organisms have ergosterol in their membrane and are sensitive to amphotericin-containing preparations and azole 14-demethylase inhibitors. Amphotericin B and lipid preparations of amphotericin have been shown to be effective for mucosal leishmaniasis resistant to antimonial compounds. For Old World cutaneous disease with *L. major*, treatment with fluconazole (200 mg daily for 6 weeks) has been shown to be effective.[103] Pentamidine and topical paromomycin have been used as alternatives for treatment of cutaneous disease.[10]

Trypanosomiasis

AMERICAN TRYPANOSOMIASIS (CHAGAS DISEASE)

Infection with the protozoan hemoflagellate *Trypanosoma cruzi* produces American trypanosomiasis, or Chagas disease, which has acute and chronic forms.[104]

Etiology and Epidemiology

Two forms of *T. cruzi* infect mammals: trypanosomes, which are carried in the blood, and amastigotes, which are found in infected cells. The parasite is transmitted by several genera of reduviid or triatomid bugs, commonly called assassin bugs because they prey on other insects or called kissing bugs because of their predilection for biting the face [*see Figure 19*]. Nonhuman reservoirs include cats, dogs, rats and other rodents, raccoons, opossums, and armadillos. The local pattern of transmission of *T. cruzi* infections depends on the species of reduviid bug, the sylvatic and domestic mammalian reservoirs, and housing conditions. Infected animals living around dwellings, usually in rural areas, are likely to transmit the organism to reduviid bugs. These infected bugs inhabit niches in walls and ceilings of poorly constructed houses. At night, they come out and feed on the blood of sleeping humans by biting exposed skin areas such as the face. During their meal, the insects excrete feces containing infective-stage metacyclic trypanosomes, which enter the host through the bite wound, cutaneous abrasions, or mucous membranes of the conjunctiva or lips.

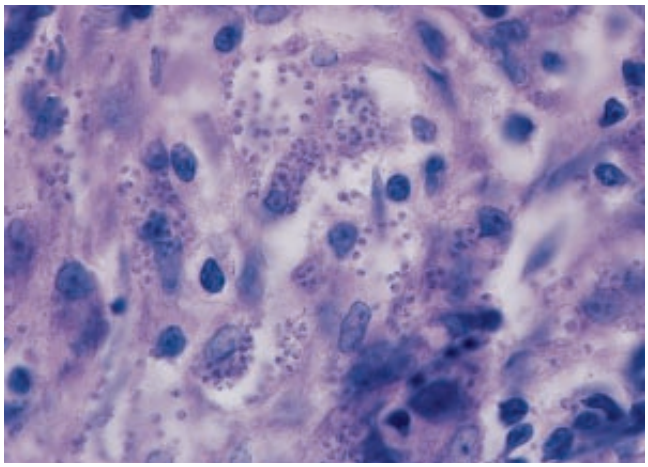


Figure 18 **Amastigotes are demonstrated on Giemsa-stained biopsy tissue from an ulcer in a patient with Old World cutaneous leishmaniasis.**

Although reduviid vectors are present in the United States and infected mammals have been found in several states, the opportunity for transmission of infections seems limited. Only five autochthonous cases of Chagas disease have been reported in the United States.[105] In rural areas of Central and South America, human infections are common where conditions favor access of infected bugs to persons.[104] In addition, infections may be transmitted by blood transfusions, across the placenta, to laboratory workers, and, in rare instances, by ingestion of foodstuffs contaminated with the excreta of infected reduviid bugs.

Pathogenesis

Both acute and chronic forms of Chagas disease are recognized. The acute form, which develops soon after infection, principally affects children in endemic areas.[104] Within several days of infection, an indurated erythematous lesion termed a chagoma appears at the inoculation site. When the inoculation site is the conjunctiva, unilateral periorbital edema develops, called the Romaña sign [*see Figure 20*]. After about 2 weeks, trypanosomes appear in the blood and invade cells, generally those of mesenchymal origin, where they multiply as intracellular amastigote forms. They are even able to proliferate in macrophages, unless the macrophages have been activated by interferon gamma. The resultant intracellular pseudocysts rupture, releasing both trypanosomal and amastigote forms. Both of these forms are infectious to mammalian cells. A combination of humoral and cell-mediated immunity controls high-level parasitemia, but despite this relative immunity, the host remains parasitemic at low levels for life. The reasons *T. cruzi* is able to establish lifelong infection are incompletely understood, but contributors include the following: evasion of complement-mediated cytolysis, intracellular growth in phagocytes, and the display of thousands of antigenically distinct surface proteins that appear to disrupt an effective cell-mediated immune response. Only 10% to 30% of infected persons develop chronic forms of Chagas disease. The reasons some persons develop disease and others don't is poorly understood and may involve the initial parasite burden, continuous inflammation in critical areas, induction of autoimmunity by the chronic infection, or some combination thereof.

Diagnosis

Clinical features During the acute phase, the patient may experience intermittent or continuous fever, malaise, an evanescent rubelliform or petechial rash, hepatosplenomegaly, lymphadenopathy, nonpitting edema of the face or extremities, tender subcutaneous nodules termed hematogenous chagomas, and, in infants, diarrhea. In severe cases, fatal myocarditis or meningoencephalitis can develop. The acute phase is usually self-limited: patients eventually become asymptomatic, and parasites can no longer be detected in the bloodstream except by PCR or by feeding reduviid bugs on the patient's blood and examining them for infection later.

The chronic form of Chagas disease may become manifest either after an acute infection or, more commonly, after a clinically inapparent infection. It usually arises in the second or third decade of life and progresses over subsequent decades. Chronic complications of Chagas disease result from the destruction of autonomic ganglia and from myositis; the pathogenesis of these

Figure 19 ***Triatoma infestans,* commonly known as assassin bugs or kissing bugs, are vectors for Chagas disease.**

lesions is not understood. The organ most frequently involved is the heart,[104] which develops biventricular hypertrophy and a mononuclear cell infiltrative myocarditis. Conduction disorders often include right bundle branch block, partial or complete atrioventricular block, and premature ventricular contractions. Sudden death has occurred in patients with Chagas disease, and fatalities have also resulted from complications of heart failure.

The GI tract is the second most frequently involved organ system.[104] The disease causes denervation leading to impaired motility and dilatation, which results in megaesophagus and megacolon. Neurologic disease is the third most frequently observed complication of chronic Chagas disease and manifests primarily as peripheral neuropathies.[106]

Congenital infections are usually responsible for premature births. Such premature infants may have hepatosplenomegaly, abdominal distention, cardiomegaly, megaesophagus, and meningoencephalitis.

In HIV-infected patients, reactivation of Chagas disease can produce cerebral masses and, in patients with acute infections, necrotizing encephalitis. These CNS infections cannot be distinguished radiographically from toxoplasmosis, and biopsy must be performed in patients with risk factors for both infections.

Laboratory findings In acute Chagas disease, the total leukocyte count often exceeds 18,000 cells/mm^3 (70% to 90% lymphocytes), and parasites are often demonstrable in the blood or in specimens from bone marrow, lymph nodes, CSF, pericardial fluid, or other involved areas. On unstained blood smears, motile trypanosomes may be seen; on Giemsa-stained smears, the organisms appear as C-shaped forms [*see Figure 21*]. If smears do not reveal the organisms, trypanosomes may be found in stained sediment obtained by centrifuging several milliliters of blood after lysing the erythrocytes. Organisms may be cultured from blood on NNN medium or in blood broth. Alternatively, blood may be injected into a laboratory rodent, whose blood is then monitored for evidence of parasitemia. Xenodiagnosis, which is not readily available, is one of the most sensitive diagnostic techniques to detect parasites. In this procedure, laboratory-reared reduviid insects are allowed to feed on a patient. If the blood ingested by the insects contains trypanosomes, the insects will become infected, and such infection can be detected by subsequent examinations of the insects' feces for excreted parasites. PCR to detect circulating parasites is becoming available in clinical labs, is as sensitive as xenodiagnosis, and is easier to implement than xenodiagnosis. Blood containing trypanosomes is infectious and should be handled with care.

In chronic Chagas disease, a chest x-ray may reveal biventricular cardiomegaly and congestive heart failure. Electrocardiographic abnormalities are commonly seen, particularly right bundle branch block. Barium swallow or enema exams may demonstrate megaesophagus or megacolon disease. Methods other than PCR or xenodiagnosis are rarely capable of detecting organisms in the blood. Similarly, it is difficult to demonstrate parasites in affected tissues; commonly, only mononuclear inflammatory cells or fibrosis is seen in pathologic specimens. Indirect fluorescent antibody and enzyme immunoassay *T. cruzi* serology tests are performed by the CDC. Both of these tests rely on crude antigens derived from cultured insect forms of *T. cruzi,* and individuals infected with *Leishmania* may have cross-reactive antibodies that give a false positive result on these tests. Although these serology tests may be positive, their results may only reflect past infection and fail to establish a link between clinical findings and active Chagas disease.

Figure 20 **Periorbital edema of the right eye (Romaña sign) is evident in a child from Panama with acute Chagas disease.**

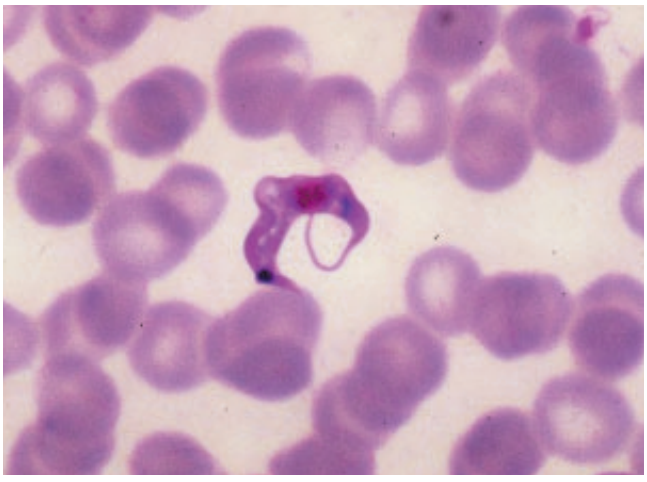


Figure 21 **A trypanosomal form of *Trypanosoma cruzi* is visible on this Giemsa-stained blood smear.**

Treatment

Optimal therapy for *T. cruzi* infections remains to be established. Nifurtimox (available from the CDC Drug Service) eliminates parasitemia [*see Sidebar,* Protozoan Infection Information on the Internet]. It should be administered to patients with acute disease and to patients with chronic disease and demonstrated parasitemia.[10] Side effects are frequent and include hemolytic anemia in patients with G6PD deficiency, peripheral neuritis, and psychosis. Anecdotal evidence suggests that interferon gamma combined with nifurtimox may shorten the duration of acute disease. Some authors believe that benznidazole is the drug of choice for treating *T. cruzi* infection, but it is not available in the United States and has a variety of toxicities, including granulocytopenia, rash, and peripheral neuropathy.[104] Itraconazole and other azoles have activity in blocking the ergosterol synthesis of *T. cruzi,* and itraconazole has some efficacy against chronic disease.[107] Though some small studies have shown benefit of benznidazole or nifurtimox treatment of early chronic Chagas cardiac disease, most experts agree that the cardiovascular and GI complications of chronic Chagas disease should be managed medically. Surgical treatment may be required for megacolon, and balloon dilatation of the lower esophageal sphincter may be needed for megaesophagus. If cardiac transplantation is contemplated, preparations to provide antitrypanosomal therapy should be made because reactivation of latent parasitemia can occur as a complication of the immunosuppressive drugs.

AFRICAN TRYPANOSOMIASIS

Etiology and Epidemiology

African trypanosomiasis, or sleeping sickness, is an acute or chronic parasitic disease caused by protozoan hemoflagellates of two *Trypanosoma brucei* subspecies. The disease is prevalent in a broad periequatorial belt across Africa. Two forms occur in humans: West African and East African, or Rhodesian, sleeping sickness. West African trypanosomiasis is present in the tropical forests of West and Central Africa and is caused by *T. brucei gambiense*, a parasite not carried in any major animal reservoir. In contrast, *T. brucei rhodesiense*, which produces East African sleeping sickness, is prevalent in the savanna and woodlands of tropical East Africa and exists in wild animal reservoirs. Visitors to game parks are at risk for acquiring East African trypanosomiasis.[108] Both types of African trypanosomiasis are transmitted by species of tsetse flies (genus *Glossina*), with the riverine *G. palpalis* group transmitting *T. brucei gambiense* and the savanna *G. morsitans* group transmitting *T. brucei rhodesiense*. *T. brucei* species are able to exist as chronic infections in the bloodstream and, later, in the CNS, partly because of their ability to undergo sequential antigenic variation of the major variant surface glycoprotein (VSG) that covers the trypanosome. As antibodies develop to a given VSG, most of the trypanosomes are eliminated from the circulation, but variants expressing an antigenically distinct VSG grow out and continue the infection. The combination of thousands of genes and pseudogenes for VSGs, plus an ability to create new VSG genes by recombination, allows the trypanosome to stay ahead of the immune response.

Diagnosis

Clinical features Within a few days to a couple of weeks after inoculation of organisms by a tsetse fly, a trypanosomal chancre may develop at the site of the insect bite, which is usually on exposed skin.[109] The chancre initially appears as a papule and, within 2 weeks, evolves into an inflamed, painful nodule that subsequently resolves spontaneously. Trypanosomal chancres commonly occur in non-African patients but usually do not develop in African patients.

During the next phase of African trypanosomiasis, the hemolymphatic phase, trypanosomes invade the bloodstream and lymph nodes. In Africans, the development of symptoms in this phase occurs slowly, over several months, and presenting symptoms include fever, lymphadenopathy, headache, and debility. In non-Africans, however, the onset is abrupt and early, often concomitant with the development of the chancre. In non-Africans, episodes of high fever that last 1 to 7 days and recur after afebrile periods are prominent. Associated symptoms include chills, headache, malaise, and anorexia. Soft, nontender lymphadenopathy develops more prominently with West African trypanosomiasis and may include enlargement of posterior cervical nodes (Winterbottom sign). A characteristic rash, which can be observed on light-skinned individuals, occurs about 6 to 8 weeks after infection and may appear as evanescent, circinate, erythematous patches, usually located on the trunk.

The next phase in the evolution of African trypanosomiasis is CNS invasion leading to diffuse meningoencephalitis or meningomyelitis. In West African trypanosomiasis, which is a slowly evolving illness, the symptoms of sleeping sickness may not develop until years after infection. Increasing lassitude and indifference are complicated by progressive neurologic compromise leading to coma and death by inanition or intercurrent infection. In contrast, the pace of East African trypanosomiasis is much more rapid, and CNS involvement may develop earlier. Even before CNS involvement, such manifestations as somnolence, personality changes, and an inability to concentrate may appear. Pancarditis often complicates acute East African trypanosomiasis and may cause death before the onset of CNS disease. Because East African trypanosomiasis may be acquired by visitors to game preserves and is an acute febrile illness, it may be mistaken for malaria and must be considered in the differential diagnosis of a febrile patient returning from an endemic area.[77]

Laboratory findings The total leukocyte count is usually normal, but the differential may show mononucleosis of 50% to 70%. Serum IgM levels rise 1 to 2 weeks after parasites appear in

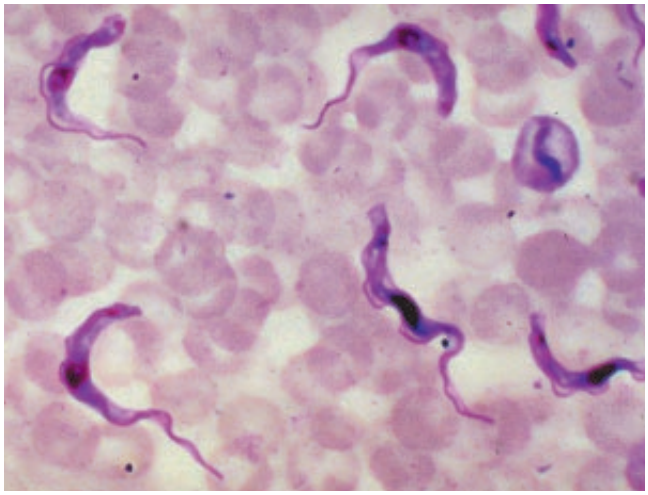


Figure 22 ***Trypanosoma brucei* parasites, the cause of African sleeping sickness, are evident on this Giemsa-stained blood smear.**

the blood and may shoot up to more than seven times the normal level. The definitive diagnosis is made by detecting trypanosomal organisms in blood, bone marrow, fluid from enlarged lymph nodes, or centrifuged CSF [*see Figure 22*]. Trypanosomes may be seen moving rapidly on wet mounts of blood or other aspirates. Organisms stained with Wright, Giemsa, or Leishman stain may be found in a chancre 48 hours before they appear in the blood. If examination of peripheral blood is unrevealing, trypanosomes may be found in fluid aspirated from involved lymph nodes or among buffy coat cells. Even in cases in which neurologic symptoms are absent, CNS disease must be excluded by performing a lumbar puncture. Even if trypanosomes are not seen in centrifuged CSF, CNS involvement may still exist, as indicated by elevations of the cell count or of protein or IgM concentrations. Serologic tests for African trypanosomiasis exist, but treatment is generally reserved for parasitologically confirmed cases because the therapy is so toxic.

Treatment

Eflornithine is effective therapy for *T. brucei gambiense* infections,[10,109] even those refractory to the arsenical agent melarsoprol. Eflornithine is the treatment of choice for both the early hemolymphatic and the later CNS stages of West African trypanosomiasis. Because eflornithine can cause anemia, leukopenia, and thrombocytopenia, blood cell counts should be monitored twice a week during therapy. Eflornithine has not been effective with *T. brucei rhodesiense* infections. The early stage of East African trypanosomiasis is treated with suramin, and the CNS stage is treated with melarsoprol.[10] Both of these drugs are available from the CDC Drug Service [*see Sidebar*, Protozoan Infection Information on the Internet] but have a variety of toxicities; notably, melarsoprol treatment is itself fatal in 4% to 6% of patients.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Filler S, Causer LM, Newman RD, et al: Malaria surveillance—United States, 2001. MMWR Surveill Summ 52:1, 2003

2. Breman JG: The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. Am J Trop Med Hyg 64:S1, 2001

3. Hulbert TV: Congenital malaria in the United States: report of a case and review. Clin Infect Dis 14:922, 1992

4. Local transmission of *Plasmodium vivax* malaria—Palm Beach County, Florida, 2003. MMWR Morb Mortal Wkly Rep 52:908, 2003

5. Jafari S, Durand R, Lusina D, et al: Molecular characterisation of airport malaria: four cases in France during summer 1999. Parasite 9:187, 2002

6. Svenson JE, MacLean JD, Gyorkos TW, et al: Imported malaria: clinical presentation and examination of symptomatic travelers. Arch Intern Med 155:861, 1995

7. Taylor TE, Fu WJ, Carr RA, et al: Differentiating the pathologies of cerebral malaria by postmortem parasite counts. Nat Med 10:143, 2004

8. Vinetz JM, Li J, McCutchan TF, et al: *Plasmodium malariae* infection in an asymptomatic 74-year-old Greek woman with splenomegaly. N Engl J Med 338:367, 1998

9. Labbe AC, Pillai DR, Hongvangthong B, et al: The performance and utility of rapid diagnostic assays for *Plasmodium falciparum* malaria in a field setting in the Lao People's Democratic Republic. Ann Trop Med Parasitol 95:671, 2001

10. Drugs for parasitic infections. Med Lett Drugs Ther 1127:32, 2002

11. Riddle MS, Jackson JL, Sanders JW, et al: Exchange transfusion as an adjunct therapy in severe *Plasmodium falciparum* malaria: a meta-analysis. Clin Infect Dis 34:1192, 2002

12. Telford SR III, Maguire JH: Babesiosis. Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 767

13. Herwaldt BL, Persing DH, Précigout EA, et al: A fatal case of babesiosis in Missouri: identification of another piroplasm that infects humans. Ann Intern Med 124:643, 1996

14. Falagas ME, Klempner MS: Babesiosis in patients with AIDS: a chronic infection presenting as fever of unknown origin. Clin Infect Dis 22:809, 1996

15. Herwaldt BL, Kjemtrup AM, Conrad PA, et al: Transfusion-transmitted babesiosis in Washington State: first reported case caused by a WA1-type parasite. J Infect Dis 175:1259, 1997

16. Krause PJ, Spielman A, Telford SR III, et al: Persistent parasitemia after acute babesiosis. N Engl J Med 339:160, 1998

17. Krause PJ, Telford SR III, Spielman A, et al: Concurrent Lyme disease and babesiosis: evidence for increased severity and duration of the disease. JAMA 275:1657, 1996

18. Dubey JP, Beattie CP: Toxoplasmosis of Animals and Man. CRC Press, Boca Raton, Florida, 1988

19. McCabe RE, Brooks RG, Dorfman RF, et al: Clinical spectrum in 107 cases of toxoplasmic lymphadenopathy. Rev Infect Dis 9:754, 1987

20. Mawhorter SD, Effron D, Blinkhorn R, et al: Cutaneous manifestations of toxoplasmosis. Clin Infect Dis 14:1084, 1992

21. Montoya JG, Remington JS: Studies on the serodiagnosis of toxoplasmic lymphadenitis. Clin Infect Dis 20:781, 1995

22. Gallino A, Maggiorini M, Kiowski W, et al: Toxoplasmosis in heart transplant recipients. Eur J Clin Microbiol Infect Dis 15:389, 1996

23. Renoult E, Georges E, Biava MF, et al: Toxoplasmosis in kidney transplant recipients: report of seven cases and review. Clin Infect Dis 24:625, 1997

24. Israelski DM, Remington JS: Toxoplasmosis in patients with cancer. Clin Infect Dis 17(suppl 2):S423, 1993

25. Hunter CA, Remington JS: Immunopathogenesis of toxoplasmic encephalitis. J Infect Dis 170:1057, 1994

26. Campagna AC: Pulmonary toxoplasmosis. Semin Respir Infect 12:98, 1997

27. Luft BJ, Remington JS: Toxoplasmic encephalitis in AIDS. Clin Infect Dis 15:221, 1992

28. Rabaud C, May T, Amiel C, et al: Extracerebral toxoplasmosis in patients infected with HIV: a French national study. Medicine (Baltimore) 73:306, 1994

29. Richards FO Jr, Kovacs JA, Luft BJ: Preventing toxoplasmic encephalitis in persons infected with human immunodeficiency virus. Clin Infect Dis 21(suppl 1):S49, 1995

30. Knobel H, Guelar A, Graus F, et al: Toxoplasmic encephalitis with normal CT scan and pathologic MRI. Am J Med 99:220, 1995

31. Pierce MA, Johnson MD, Maciunas RJ, et al: Evaluating contrast-enhancing brain lesions in patients with AIDS by using positron emission tomography. Ann Intern Med 123:594, 1995

32. Potasman I, Resnick L, Luft BJ, et al: Intrathecal production of antibodies against *Toxoplasma gondii* in patients with toxoplasmic encephalitis and the acquired immunodeficiency syndrome (AIDS). Ann Intern Med 108:49, 1988

33. Mathews C, Barba D, Fullerton SC: Early biopsy versus empiric treatment with delayed biopsy of non-responders in suspected HIV-associated cerebral toxoplasmosis: a decision analysis. AIDS 9:1243, 1995

34. Minkoff H, Remington JS, Holman S, et al: Vertical transmission of toxoplasmosis by human immunodeficiency virus-infected women. Am J Obstet Gynecol 176:555, 1997

35. Wong S, Remington JS: Toxoplasmosis in pregnancy. Clin Infect Dis 18:853, 1994

36. Guerina NG, Hsu H, Meissner C, et al: Neonatal serologic screening and early treatment for congenital *Toxoplasma gondii* infection. N Engl J Med 330:1858, 1994

37. Romand S, Chosson M, Franck J, et al: Usefulness of quantitative polymerase chain reaction in amniotic fluid as early prognostic marker of fetal infection with *Toxoplasma gondii*. Am J Obstet Gynecol 190:797, 2004

38. McAuley J, Boyer KM, Patel D, et al: Early longitudinal evaluations of treated infants and children and untreated historical patients with congenital toxoplasmosis: the Chicago Collaborative Treatment Trial. Clin Infect Dis 18:38, 1994

39. Holland GN: Ocular toxoplasmosis: a global reassessment: part I: epidemiology and course of disease. Am J Ophthalmol 136:973, 2003

40. Bowie WR, King AS, Werker DH, et al: Outbreak of toxoplasmosis associated with municipal drinking water. Lancet 350:173, 1997

41. Wilson M, Remington JS, Clavet C, et al: Evaluation of six commercial kits for detection of human immunoglobulin M antibodies to *Toxoplasma gondii*. The FDA Toxoplasmosis Ad Hoc Working Group. J Clin Microbiol 35:3112, 1997

42. Remington JS, Thulliez P, Montoya JG: Recent developments for diagnosis of toxoplasmosis. J Clin Microbiol 42:941, 2004

43. Frenkel JK: Toxoplasmosis. Pathology of Infectious Diseases. Connor DH, Chandler FW, Schwartz DA, et al, Eds. Appleton & Lange, Stamford, Connecticut, 1997, p 1261

44. Luft BJ, Harner R, Korzun AH, et al: Toxoplasmic encephalitis in patients with the acquired immunodeficiency syndrome. N Engl J Med 329:995, 1993

45. Torres RA, Weinberg W, Stansell J, et al: Atovaquone for salvage treatment and suppression of toxoplasmic encephalitis in patients with AIDS. Atovaquone/Toxoplasmic Encephalitis Study Group Clin Infect Dis 24:422, 1997

46. Podzamczer D, Miro JM, Bolao F, et al: Twice-weekly maintenance therapy with sulfadiazine-pyrimethamine to prevent recurrent toxoplasmic encephalitis in patients with AIDS. Ann Intern Med 123:175, 1995

47. Masur H, Kaplan JE, Holmes KK: Guidelines for preventing opportunistic infections among HIV-infected persons—2002. Recommendations of the U.S. Public Health Service and the Infectious Diseases Society of America. Ann Intern Med 137:435, 2002

48. Lee SH, Levy DA, Craun GF, et al: Surveillance for waterborne-disease outbreaks—United States, 1999–2000. MMWR Surveill Summ 51:1, 2002

49. Overturf GD: Endemic giardiasis in the United States—role of the daycare center. Clin Infect Dis 18:764, 1994

50. Monis PT, Thompson RC: *Cryptosporidium* and *Giardia*–zoonoses: fact or fiction? Infect Genet Evol 3:233, 2003

51. Oberhuber G, Kastner N, Stolte M: Giardiasis: a histologic analysis of 567 cases. Scand J Gastroenterol 32:48, 1997

52. Lai Ping So A, Mayer L: Gastrointestinal manifestations of primary immunodeficiency disorders. Semin Gastrointest Dis 8:22, 1997

53. Hopkins RS, Juranek DD: Acute giardiasis: an improved clinical case definition for epidemiologic studies. Am J Epidemiol 133:402, 1991

54. Lengerich EJ, Addiss DG, Juranek DD: Severe giardiasis in the United States. Clin Infect Dis 18:760, 1994

55. Zimmerman SK, Needham CA: Comparison of conventional stool concentration and preserved-smear methods with Merifluor *Cryptosporidium/Giardia* Direct Immunofluorescence Assay and ProSpecT Giardia EZ Microplate Assay for detection of *Giardia lamblia*. J Clin Microbiol 33:1942, 1995

56. Thielman NM, Guerrant RL: Persistent diarrhea in the returned traveler. Infect Dis Clin North Am 12:489, 1998

57. Nash TE, Ohl CA, Thomas E, et al: Treatment of patients with refractory giardiasis. Clin Infect Dis. 33:22, 2001

58. Gardner TB, Hill DR: Treatment of giardiasis. Clin Microbiol Rev 14:114, 2003

59. Ortiz JJ, Ayoub A, Gargala G, et al: Randomized clinical study of nitazoxanide compared to metronidazole in the treatment of symptomatic giardiasis in children from northern Peru. Aliment Pharmacol Ther 15:1409, 2001

60. Burtin P, Taddio A, Ariburnu O, et al: Safety of metronidazole in pregnancy: a meta-analysis. Am J Obstet Gynecol 172:525, 1995

61. Hill DR, Nash TE: Intestinal flagellate and ciliate infections. Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 703

62. Silberman JD, Clark CG, Sogin ML: *Dientamoeba fragilis* shares a recent common evolutionary history with the trichomonads. Mol Biochem Parasitol 76:311, 1996

63. Cuffari C, Oligny L, Seidman EG: *Dientamoeba fragilis* masquerading as allergic colitis. J Pediatr Gastroenterol Nutr 26:16, 1998

64. González-Ruiz A, Wright SG: Disparate amoebae. Lancet 351:1672, 1998

65. Haque R, Ali IK, Akther S, et al: Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J Clin Microbiol 36:449, 1998

66. Petri WA, Singh U, Ravdin JI: Enteric amebiasis.Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 685

67. Haque R, Duggal P, Ali IM, et al: Innate and acquired resistance to amebiasis in Bangladeshi children. J Infect Dis 186:547, 2002

68. Chun D, Chandrasoma P, Kiyabu M: Fulminant amebic colitis: a morphologic study of four cases. Dis Colon Rectum 37:535, 1994

69. Diaz E, Mondragon J, Ramirez E, et al: Epidemiology and control of intestinal parasites with nitazoxanide in children in Mexico. Am J Trop Med Hyg 68:384, 2003

70. Tan KSW, Singh M, Yap EH: Recent advances in *Blastocystis hominis* research: hot spots in terra incognita. Int J Parasitol 32:789, 2002

71. Marshall MM, Naumovitz D, Ortega Y, et al: Waterborne protozoan pathogens. Clin Microbiol Rev 10:67, 1997

72. Ackers JP: Gut coccidia—*Isospora*, *Cryptosporidium*, *Cyclospora* and *Sarcocystis*. Semin Gastrointest Dis 8:33, 1997

73. Verdier RI, Fitzgerald DW, Johnson WD Jr, et al: Trimethoprim-sulfamethoxazole compared with ciprofloxacin for treatment and prophylaxis of *Isospora belli* and *Cyclospora cayetanensis* infection in HIV-infected patients: a randomized, controlled trial. Ann Intern Med 132:885, 2000

74. DuPont HL, Chappell CL, Sterling CR, et al: The infectivity of *Cryptosporidium parvum* in healthy volunteers. N Engl J Med 332:855, 1995

75. MacKenzie WR, Hoxie NJ, Proctor ME, et al: A massive outbreak in Milwaukee of *Cryptosporidium* infection transmitted through the public water supply. N Engl J Med 331:161, 1994

76. Leav BA, Mackay M, Ward HD: *Cryptosporidium* species: new insights and old challenges. Clin Infect Dis 36:903, 2003

77. McAnulty JM, Fleming DW, Gonzalez AH: A community-wide outbreak of cryptosporidiosis associated with swimming at a wave pool. JAMA 272:1597, 1994

78. Squier C, Yu VL, Stout JE: Waterborne nosocomial infections. Curr Infect Dis Rep 2:490, 2000

79. Chen XM, Keithly JS, Paya CV, et al: Cryptosporidiosis. N Engl J Med 346:1723, 2002

80. White AC, Robinson P, Okhuysen PC: Human studies provide insight into the pathogenesis, immunology, and treatment of cryptosporidiosis. Clin Infect Dis 37:989, 2003

81. Eberhard ML, Arrowood MJ: *Cyclospora* spp. Curr Opin Infect Dis 15:519, 2002

82. Herwaldt BL: *Cyclospora cayetanensis*: a review, focusing on the outbreaks of cyclosporiasis in the 1990s. Clin Infect Dis 31:1040, 2000

83. Fleming CA, Caron D, Gunn JE, et al: A foodborne outbreak of *Cyclospora cayetanensis* at a wedding: clinical features and risk factors for illness. Arch Intern Med 158:1121, 1998

84. Eberhard ML, Pieniazek NJ, Arrowood MJ: Laboratory diagnosis of *Cyclospora* infections. Arch Pathol Lab Med 121:792, 1997

85. Keeling PJ, Fast NM: *Microsporidia*: biology and evolution of highly reduced intracellular parasites. Annu Rev Microbiol 56:93, 2002

86. Franzen C, Muller A: Microsporidiosis: human diseases and diagnosis. Microbes Infect 3:389, 2001

87. Bryan RT, Weber R, Schwartz DA: Microsporidiosis. Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 840

88. Molina JM, Chang C, Goguel J, et al: Albendazole for treatment and prophylaxis of microsporidiosis due to *Encephalitozoon intestinalis* in patients with AIDS: a randomized double-blind controlled trial. J Infect Dis 177:1373, 1998

89. Dieterich DT, Lew EA, Kotler DP, et al: Treatment with albendazole for intestinal disease due to *Enterocytozoon bieneusi* in patients with AIDS. J Infect Dis 169:178, 1994

90. Coyle CM, Weiss LM, Rhodes LV, et al: Fatal myositis due to the microsporidian *Brachiola algerae*, a mosquito pathogen. N Engl J Med 351:42, 2004

91. Martinez AJ, Visvesvara GS: Free-living, amphizoic and opportunistic amebas. Brain Pathol 7:583, 1997

92. Martinez AJ, Visvesvara GS: Free living amebae: *Naegleria, Acanthamoeba,* and *Balamuthia* infections. Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 814

93. Gelman BB, Rauf SJ, Nader R, et al: Amoebic encephalitis due to *Sappinia diploidea*. JAMA 285:2450, 2001

94. Seidel JS, Harmatz P, Visvesvara GS, et al: Successful treatment of primary amebic meningoencephalitis. N Engl J Med 306:346, 1982

95. Seal DV: *Acanthamoeba* keratitis update—incidence, molecular epidemiology and new drugs for treatment. Eye 17:893, 2003

96. Chandrasekar PH, Nandi PS, Fairfax MR, et al: Cutaneous infections due to *Acanthamoeba* in patients with acquired immunodeficiency syndrome. Arch Intern Med 157:569, 1997

97. Rein MF: Trichomonas vaginalis. Principles and Practice of Infectious Diseases, 5th ed. Mandell GL, Douglas RG Jr, Bennett JE, Eds. Churchill Livingstone, New York, 2000, p 2894

98. Magill AJ, Grogl M, Gasser RA Jr, et al: Visceral infection caused by *Leishmania tropica* in veterans of Operation Desert Storm. N Engl J Med 328:1383, 1993

99. Pearson RD, Jeronimo SMB, Sousa AQS: Leishmaniasis. Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 797

100. Sunder S, Jha TK, Thakur CP, et al: Oral miltefosine for Indian visceral leishmaniasis. N Engl J Med 347:1739, 2002

101. Berman JD: Human leishmaniasis: clinical, diagnostic, and chemotherapeutic developments in the last 10 years. Clin Infect Dis 24:684, 1997

102. Herwaldt BL, Arana BA, Navin TR: The natural history of cutaneous leishmaniasis in Guatemala. J Infect Dis 165:518, 1992

103. Alrajhi AA, Ibrahim EA, De Vol EB, et al: Fluconazole for the treatment of cutaneous leishmaniasis caused by *Leishmania major*. N Engl J Med 346:891, 2002

104. Kirchhoff LV: American trypanosomiasis (Chagas' disease) Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 785

105. Herwaldt BL, Grijalva MJ, Newsome AL, et al: Use of polymerase chain reaction to diagnose the fifth reported US case of autochthonous transmission of *Trypanosoma cruzi*, in Tennessee, 1998. J Infect Dis 181:395, 2000

106. Genovese O, Ballario C, Storino R, et al: Clinical manifestations of peripheral nervous system involvement in human chronic Chagas disease. Arq Neuropsiquiatr 54:190,1996

107. Apt W, Arribada A, Zulantay I, et al: Itraconazole or allopurinol in the treatment of chronic American trypanosomiasis: the regression and prevention of electrocardiographic abnormalities during 9 years of follow-up. Ann Trop Med Parasitol 97:23, 2003

108. Moore AC, Ryan ET, Waldron MA: Case records of the Massachusetts General Hospital. Weekly clinicopathological exercises. Case 20-2002. A 37-year-old man with fever, hepatosplenomegaly, and a cutaneous foot lesion after a trip to Africa. N Engl J Med 346:2069, 2002

109. Pepin J, Donelson JE: African trypanosomiasis (sleeping sickness). Tropical Infectious Diseases: Principles, Pathogens, and Practice. Guerrant RL, Walker DH, Weller PF, Eds. WB Saunders Co, Philadelphia, 1999, p 774

110. Miller LH, Baruch DI, Marsh K, et al: The pathogenic basis of malaria. Nature 415:673, 2002

Acknowledgments

Figures 2, 5, and 10 Seward Hung.

Figures 3, 13, 14, and 21 Centers for Disease Control and Prevention.

Figures 4, 7, 8, 15, and 20 CDC/Dr. Mae Melvin.

Figure 5a CDC/Dr. L.L. Moore, Jr.

Figure 5b CDC/Dr. Edwin P. Ewing, Jr.

Figures 6, 11, and 12 Brett Norquist and Ajit Limaye, University of Washington (Laboratory Medicine).

Figure 9 CDC/Dr. N. J. Wheeler, Jr.

Figure 15 CDC/Dr. George Healy.

Figure 16 Sam Eng, University of Washington Laboratory Medicine.

Figure 17 Wesley C. Van Voorhis, M.D., Ph.D.

Figure 19 CDC/World Health Organization.

Figure 22 CDC/Myron G. Schultz.

XXXV HELMINTHIC INFECTIONS

Wesley C. Van Voorhis, m.d., ph.d., f.a.c.p.
Peter F. Weller, m.d., f.a.c.p

General Considerations

Parasites are generally subdivided into three categories: protozoans, helminths, and arthropods such as ticks and insects. This chapter focuses on helminthic pathogens; protozoans are discussed elsewhere, as are general characteristics of parasite-host relationships [*see 7:XXXIV Protozoan Infections*].

Helminthic parasites are divided into three major groups: nematodes (roundworms), cestodes (tapeworms), and trematodes (flukes) [*see Figure 1*]. Helminthic parasites differ from protozoan parasites in several respects. First, protozoan parasites are unicellular organisms, whereas helminthic parasites are metazoan (multicellular) worms that possess differentiated organ systems. Second, most helminthic parasites do not replicate within the human host; rather, they develop to a certain stage within the human body and mature further outside the human body. During their extrahuman life cycle, helminths exist either as free-living organisms or as parasites within another host species and mature into new developmental stages capable of reinfecting humans. Thus, with only a few exceptions (i.e., those parasites capable of internal reinfection, notably, *Strongyloides stercoralis* and *Capillaria philippinensis*), augmentation of the number of adult helminthic parasites that reside within the human host requires exogenous reinfection.

A third attribute of helminthic parasites but not of protozoan parasites is their tendency to elicit eosinophilia within the tissues and blood of infected humans. The magnitude of eosinophilia tends to correlate with the extent of tissue invasion by larvae or adult helminths. For example, in several helminthic infections, such as paragonimiasis, acute schistosomiasis, and *Ascaris* and hookworm infections, the elicited eosinophilia is greatest during the early phases of infection, when migrations of infecting larvae and the progression of subsequent developmental stages through the tissues are greatest. In established infections, local-tissue eosinophil infiltration is often present around helminths within tissues, but blood eosinophilia may be intermittent, mild, or absent. There may be no eosinophilia in established infections that are well contained within tissues (e.g., intact echinococcal cysts) or confined within the lumen of the intestinal tract (e.g., *Ascaris* and tapeworms). For some established infections, increases in blood eosinophilia may be episodic. Intermittent leakage of fluids from echinococcal cysts can elicit transient increases in blood eosinophil levels, as well as symptoms attributable to allergic or anaphylactic reactions (e.g., urticaria and bronchospasm).[1] With tissue-dwelling helminths, increased eosinophilia may be associated with the migration of adult parasites, as in loiasis and gnathostomiasis.

Certain established infections—including trichinellosis, anisakiasis, gnathostomiasis, visceral larva migrans, echinococcosis, and the several forms of filariasis—are capable of inducing eosinophilia but cannot be diagnosed on the basis of stool examination. In addition, some intestinal helminths may not be readily detectable on routine stool examination. The intestinal nematode that is most likely to cause persistent eosinophilia and that may not be detected on initial stool examination is *S. stercoralis*.[2]

Intestinal Nematode Infections

The major intestinal nematodes are roundworms, hookworms, whipworms, pinworms, and *S. stercoralis* [*see Table 1*]. Infection with these parasites occurs in many hundreds of millions of persons worldwide, especially in tropical areas. Children are particularly affected.

roundworm, hookworm, and whipworm

Roundworms, hookworms, and whipworms are geohelminths, requiring a soil phase for the fecally expelled eggs to develop into their infective stages. Thus, infections from these parasites usually occur in rural areas with poor sanitation. In these areas, highly prevalent infections with these intestinal nematodes are major contributors to malnutrition in children.[3]

Epidemiology, Pathogenesis, and Clinical Features

Hookworm The World Health Organization (WHO) estimates that 700 million persons are infected with *Ancylostoma duodenale* or *Necator americanus.* Hookworm infection is more likely where the following conditions coexist: sanitary practices that permit human fecal contamination of the soil; soil that is damp enough for larval survival; and human contact with contaminated soil. Persons at risk include children, gardeners, plumbers or electricians in contact with soil, and infantry personnel.

Hookworm eggs excreted in feces hatch in the soil, releasing larvae that develop into infective larvae [*see Figure 2*]. Percutaneous larval penetration is the principal mode of human infection, but infections with *A. duodenale* may also be acquired by oral ingestion. Larval penetration of the skin often produces a pruritic, maculopapular eruption at each site of entry. In persons previously infected, serpiginous tracts of intracutaneous larval migration, as in cutaneous larva migrans, can occur. From the skin, hookworm larvae travel via the bloodstream to the lungs. The hookworm larvae enter the alveoli, ascend the tracheobronchial tree to the pharynx, and are swallowed. The development of *A. duodenale* larvae can be arrested for many months before the larvae proceed to the lungs for subsequent maturation. Although transpulmonary larval passage may elicit a transient eosinophilic pneumonitis, this phenomenon is much less common with hookworm infections than with roundworm infections.[4] Larvae and young adult worms in the intestinal tract may cause gastrointestinal symptoms, including nausea, diarrhea, vomiting, abdominal pain (often with postprandial accentuation), and flatulence.

A major health impact of hookworm infection is iron loss resulting from the 0.1 to 0.4 ml of blood ingested daily by each adult worm [*see Figure 3*]. In malnourished hosts, such blood loss can lead to severe iron-deficiency anemia. The number of parasites necessary to cause anemia varies with host iron and protein stores, but hookworm burdens of 40 to 160 worms are associated with hemoglobin levels below 11 g/dl. Severe anemia from the interaction of malnutrition, malaria, and hookworm infection is a major contributor to childhood mortality in areas of the world where these conditions coexist.

Roundworm *Ascaris lumbricoides* is the most prevalent intestinal helminthic parasite, infecting well over one billion people worldwide. Each adult female can produce up to 200,000 eggs a day. Passed in feces, these eggs are remarkably resistant

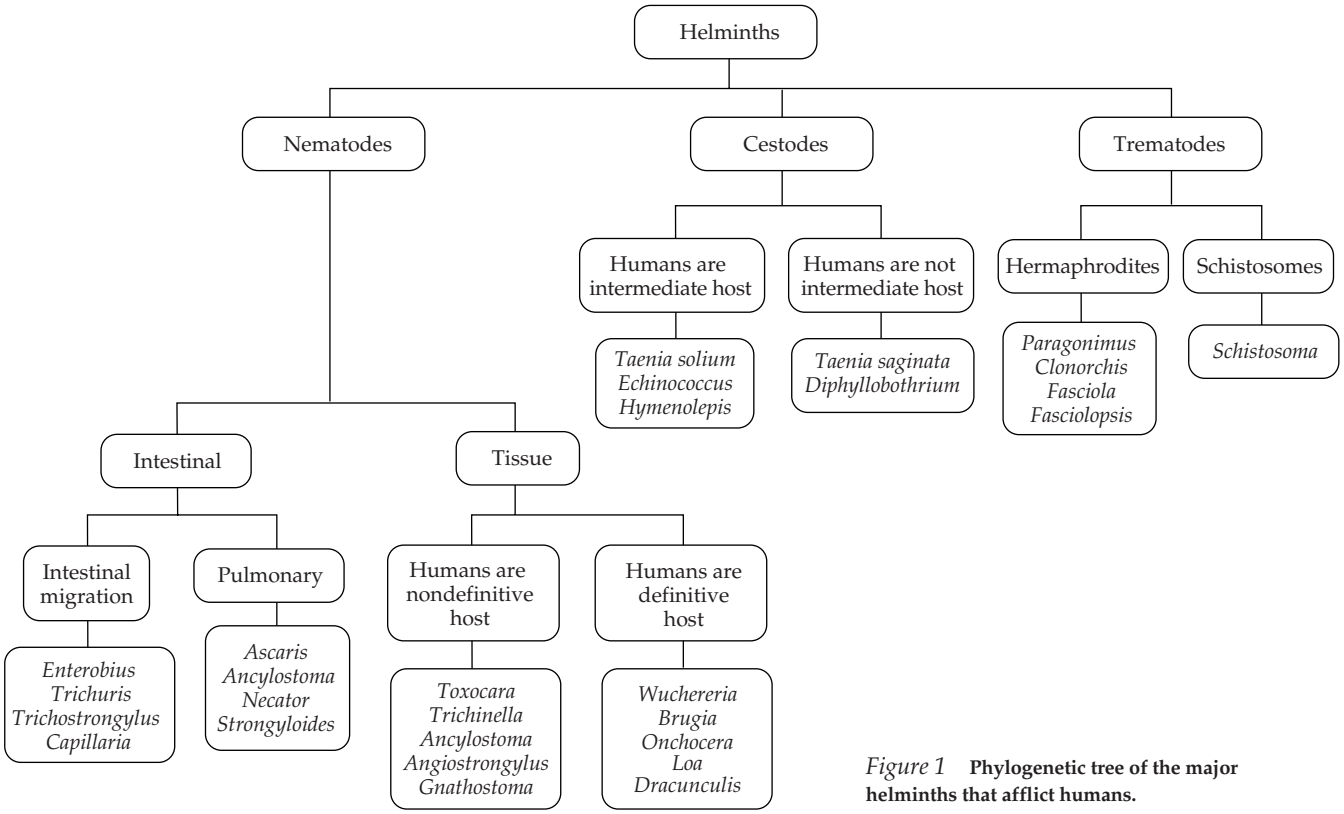


Figure 1 **Phylogenetic tree of the major helminths that afflict humans.**

to environmental stresses and are capable of remaining viable for up to 6 years. With exposure to warm, humid soil, fertilized eggs become embryonated and infectious. Most transmission occurs by ingestion of fertilized eggs on dirty hands, in fecally contaminated agricultural products or other foodstuffs, or through geophagia (the life cycle of *Ascaris* is illustrated in the Centers for Disease Control and Prevention Public Health Image Library [CDC PHIL], at http://phil.cdc.gov/Phil; photograph 5231). Infection is more common in areas with poor sanitation. In regions with large concentrations of *Ascaris* eggs in the soil, eggs may be disseminated in the air, where they are inhaled and swallowed later with respiratory secretions.

From swallowed *Ascaris* eggs, larvae hatch in the intestine within 1 to 2 days and molt into second-stage larvae, which are carried hematogenously to the liver and lungs. Roughly 1 to 2 weeks after infection, larvae penetrate alveoli from the capillary bed and molt into third-stage larvae, which ascend the tracheobronchial tree, are swallowed, and return to the intestine. These larvae mature into adult male and female worms, measuring 10 to 30 cm in length [*see Figure 4*]. The fertilized adult females begin producing eggs 2 to 3 months after the initial infection.

The clinical manifestations of ascariasis occur during early larval migration or with established intestinal infections. During the phase of transpulmonary migration, eosinophilic, often migratory, pulmonary infiltrates (Löffler syndrome) can develop, especially in previously sensitized hosts.[4] The large adult worms in the intestine usually elicit no symptoms, although they may contribute to malnutrition.

Infrequently, adult *Ascaris* worms cause serious complications. In heavy infections, especially in children, the mass of entangled adult worms may cause partial or complete intestinal obstruction.[5,6] A second type of complication can develop even when only a single adult worm migrates from its normal location within the intestinal lumen. Migration into the appendix can precipitate acute appendicitis. Acute or recurrent migrations of one or more adult worms into the biliary tract can cause obstruction leading to acalculous cholecystitis, pyogenic cholangitis, pancreatitis, or liver abscesses.[7] If the bowel wall is thinned or ulcerated by typhoid enteritis or other diseases, adult *Ascaris* worms may penetrate and perforate the intestine.

Heavy worm burden is more frequently seen in children younger than 10 years. Worm burden tends to decrease with increasing age, in correlation with increased secretion of cytokines by type 2 helper T cells (Th2) in response to *Ascaris* antigens.[8] Thus, Th2 cytokine responses may protect against heavy infections.

Whipworm Infection with *Trichuris trichiura* is widespread in tropical regions throughout the world and also occurs in rural areas of the southern United States. Whipworm infection results from ingestion of eggs that have embryonated in the soil for 15 to 30 days (the life cycle of *Trichuris* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil], photograph 5231; a whipworm is shown in photograph 414).

Moderate whipworm infection provokes gastrointestinal complaints; heavy infections (> 200 worms) in children cause the *Trichuris* dysentery syndrome, characterized by chronic diarrhea, anemia, growth retardation, and, occasionally, rectal prolapse.[9]

Laboratory Findings and Imaging Studies

Intestinal infections with roundworm, hookworm, or whipworm are usually easily diagnosed by finding the eggs of the responsible parasite on stool examination. During the pulmonary phase of *Ascaris* or hookworm infections, the diagnosis is made by finding larvae in respiratory secretions. Results of stool examinations are usually negative during the pulmonary phase be-

Table 1 Intestinal Nematodes

	Roundworm (Ascaris lumbricoides)	*Pinworm* (Enterobius vermicularis)	*Hookworm* (Necator americanus *or* Ancylostoma duodenale)	*Whipworm* (Trichuris trichiura)	Strongyloides stercoralis
Infective stage	Egg	Egg	Larva	Egg	Filariform larva
Route of infection	Oral	Oral	Percutaneous Oral (*A. duodenale)*	Oral	Percutaneous or internally with ongoing auto-infection
Extraintestinal penetration					
Larval stage	After egg ingestion	None	During initial infection	None	During initial infection and during ongoing autoinfection
Adult stage	With aberrant migration	Rarely via female genital tract	None	None	None
Adult length	10–30 cm	~ 1 cm	0.6–1.2 cm	3–5 cm	~ 0.2 cm
Principal symptoms	Usually none; occasionally, GI or biliary tract obstruction by adult worms	Perianal pruritus or none	Usually none; iron deficiency anemia in heavy infection	Usually none; GI symptoms or iron deficiency anemia in heavy infection	In uncomplicated infection, none or GI symptoms and skin lesions; in disseminated disease, symptoms related to involvement of GI tract, lungs, or other organs, with possible bacteremia
Diagnostic stage	Eggs in stool	Eggs from perianal skin on cellulose acetate tape	Eggs in fresh stool; larvae in old stool	Eggs in stool	Rhabditiform larvae in stool; filariform larvae in duodenal aspirate, in pulmonary secretions, in stool, or in other fluids or tissues; serology helpful in low-grade infections
Indications for therapy	Presence of adult worms	Symptoms	Heavy worm burden; iron deficiency anemia	Symptoms with heavy infection	Presence of parasite
Therapy (adult dose)	Albendazole (400 mg p.o. once) *or* Mebendazole (100 mg p.o., b.i.d., for 3 days or 500 mg p.o. once) *or* Ivermectin (150–200 μg/kg p.o. once) *or* Nitazoxanide (500 mg p.o., b.i.d., for 3 days)	Pyrantel pamoate (11 mg/kg p.o. in a single dose; maximum: 1 g/day) *or* Mebendazole (100 mg p.o. in a single dose) *or* Albendazole (400 mg p.o. once) For each agent, repeat course after 2 wk	Albendazole (400 mg p.o. once) *or* Mebendazole (100 mg p.o., b.i.d., for 3 days or 500 mg p.o. once) *or* Pyrantel pamoate (11 mg/kg p.o. in a single dose; maximum: 1 g/day)	Mebendazole (100 mg p.o., b.i.d., for 3 days or 500 mg p.o. once) *or* Albendazole (400 mg p.o. for 3 days) *or* Ivermectin (200 μg/kg p.o. daily for 3 days) *or* Nitazoxanide (500 mg p.o. daily for 3 days)	Ivermectin (200 μg/kg p.o., q.d., for 1–2 days) *or* Albendazole (400 mg p.o., b.i.d., for 2 days) *or* Thiabendazole (25 mg/kg p.o., b.i.d.; maximum dose, 3 g/day for 2 days or, in cases of disseminated disease, for 5 days) Immunocompromised persons may require prolonged therapy

cause this phase occurs early in infection, weeks or months before adult worms have matured sufficiently to liberate eggs into the feces. Positive stool samples during the respiratory phase indicate earlier long-standing infection. Thus, in eosinophilic pneumonitis, negative stool examination results do not exclude a parasitic etiology.

Adult *Ascaris* worms can be readily detected in upper GI series; the large worms are outlined by contrast material, and in late follow-up films, the parasite's alimentary tract may be defined by a thin line of ingested contrast medium. Ultrasonography can detect adult worms in the small intestine, facilitating diagnosis of *Ascaris* as the cause of abdominal symptoms.[10] Worms in the biliary tract can be detected by ultrasonography[11] or endoscopic cholangiopancreatography.[12] At times, patients may note the passage of the large, smooth adult *Ascaris* worms in the stool or may cough up an adult worm.

Adult whipworms, which are 3 to 5 cm in length, may be visualized by anoscopy or colonoscopy. Adult hookworms, which are 0.6 to 1.2 cm in length, may be visualized by endoscopy of the proximal small intestine.

Treatment

Therapy for roundworm, hookworm, and whipworm infections utilizes the well-tolerated broad-spectrum anthelmintic agents mebendazole, albendazole, and, in some cases, pyrantel pamoate, ivermectin, and nitazoxanide [*see Table 1*]. Several caveats apply, however. Pyrantel is not used for whipworm or roundworm; ivermectin and nitazoxanide are not used for hookworm. Because mebendazole and albendazole may be teratogenic, they are contraindicated during pregnancy. Albendazole, ivermectin, and nitazoxanide are not yet approved by the Food and Drug Administration for these indications.

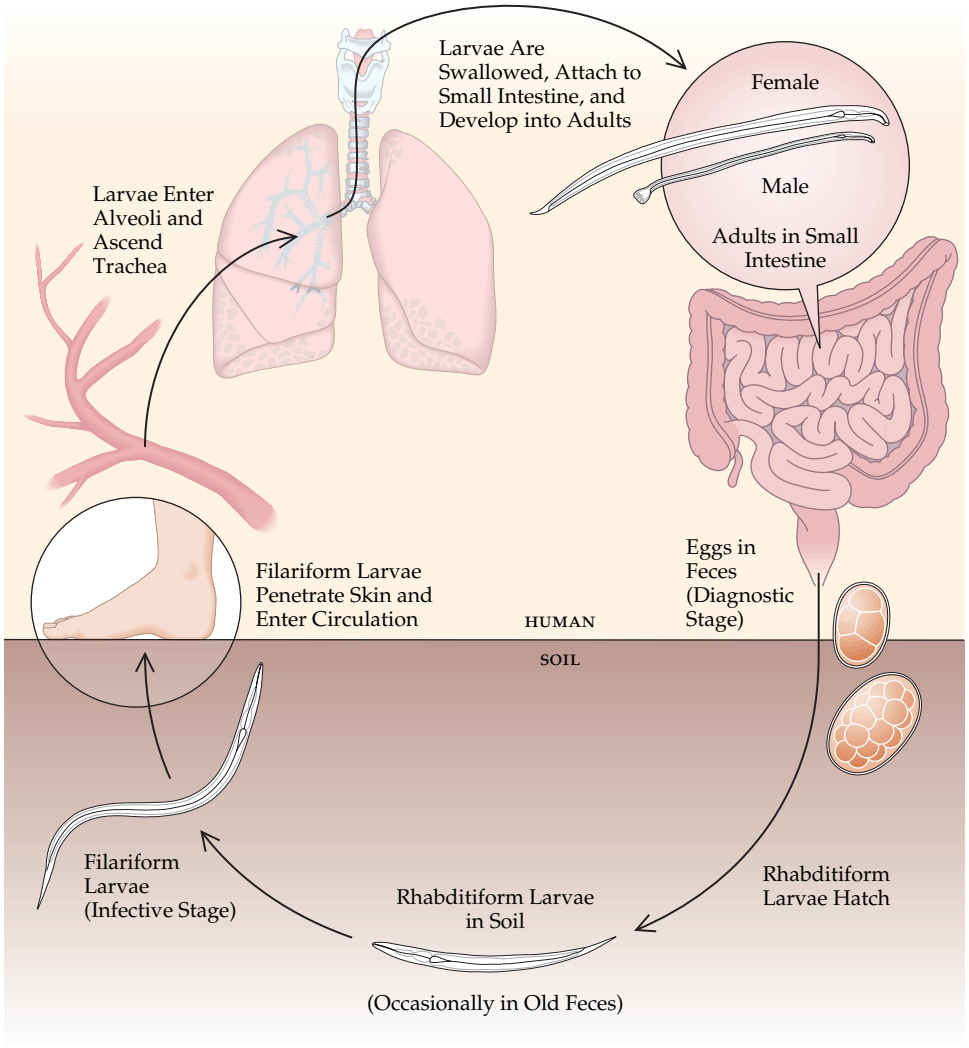


Figure 2 **Hookworm infection typically begins when filariform larvae in soil penetrate the skin and enter the circulation. The larvae travel via the bloodstream to the lungs, where they cross into the alveolar air space and ascend the tracheobronchial tree. They are then swallowed and attach to the small intestine. There they develop into adults, which feed on blood. Adult females lay eggs that are excreted in feces and hatch in the soil, releasing rhabditiform larvae that develop into infective filariform larvae. The hookworm *Ancylostoma duodenale* may also be acquired by oral ingestion of larvae.**

Therapy for *Ascaris* is mandatory to prevent unusual complications resulting from aberrant migration of the large adult worms. Adult *Ascaris* worms causing biliary and pancreatic obstruction may be removed at endoscopy.[12]

Anemia caused by hookworm responds to iron supplementation. Several animal models have shown that vaccination with recombinant hookworm antigens shows great promise in reducing the worm burden in humans and alleviating anemia.[13]

PINWORMS

Pinworms (*Enterobius vermicularis*) often affect groups of children, such as in schools or families, because the eggs are infectious when passed and can be transmitted from person to person and by fomites in the environment (the life cycle of *Enterobius* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 5228). The cardinal symptom of pinworm infection is perianal pruritus, but some infected patients remain asymptomatic. The pruritus often worsens at night, when the worms tend to migrate. Occasionally, migration of adult worms into or through the female genital tract results in vaginitis or peritoneal inflammation associated with granuloma formation.[14] Pinworm larvae may also cause eosinophilic enterocolitis or appendicitis.[15,16]

Diagnosis

Because pinworm eggs are deposited by female worms on the perianal skin, stool examinations usually are not revealing. Pinworm infections are diagnosed by applying cellulose acetate tape to the perianal skin in the morning and microscopically examining the tape on a slide to detect the eggs (a photograph of pinworm eggs is available in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 4818). Adult pinworms, which are about 1 cm in length, may be visualized by anoscopy or colonoscopy.

Treatment

Pinworm infection is treated with mebendazole, pyrantel pamoate, or albendazole (not yet FDA approved for this indication) [*see Table 1*]. Because mebendazole and albendazole may be teratogenic, they are contraindicated during pregnancy. A second treatment should be given 2 weeks after the initial course. Even with repeat treatment, however, reinfection is common. Before treatment, hygienic measures such as frequent baths, the clipping of fingernails, and cleaning of the egg-contaminated household environment should be instituted to reduce the opportunity for reinfection. Because household members are often infected and can provide a source for reinfection, simultaneous treatment of all household members is indicated.

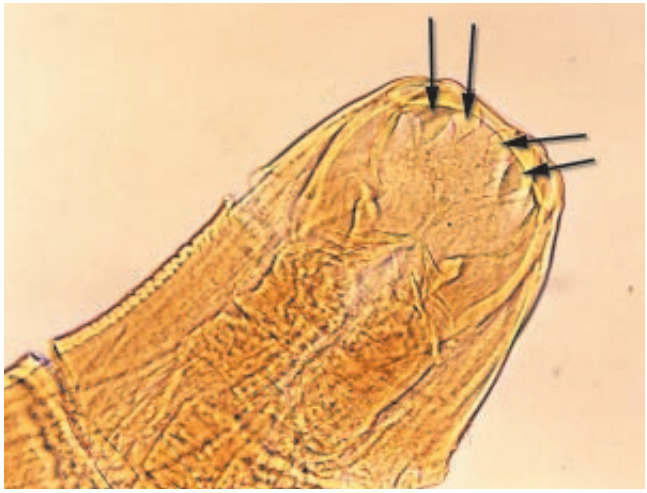


Figure 3 **The mouthparts (arrows) of an *Ancylostoma duodenale* hookworm are visible on an unstained micrograph.**

STRONGYLOIDIASIS

Although *S. stercoralis* is the least prevalent of the intestinal nematodes in the United States, it is widely distributed within the tropics and subtropics. In the United States, *Strongyloides* infection is more prevalent in persons residing in the southern states[17]; in persons living in institutions with poor sanitation[18]; and in immigrants, military veterans, and other persons who have traveled or resided in locales where *Strongyloides* is endemic. *Strongyloides* is one of the few helminths capable of internal reinfection; it can multiply in the human host without reinfection by soil-dwelling larvae. As a result of ongoing internal reinfection, strongyloidiasis may persist for decades, as has been documented in some World War II and Vietnam War veterans. If host immunity is suppressed, the internal reinfection cycle may become unbridled, leading to hyperinfection that can result in an overwhelming and frequently fatal illness.[18]

Pathogenesis and Clinical Features

Eggs of *S. stercoralis* hatch in the intestine into rhabditiform larvae, which develop further by one of three routes [*see Figure 5*]. Two of the routes occur after stool passage: the larvae may develop in the soil into infective filariform larvae either immediately or after an intervening stage as free-living adults. The filariform larvae penetrate the skin and travel via the bloodstream to the lungs, where they enter the alveoli, ascend the trachea, and are swallowed; adult worms subsequently develop in the small intestine.

The third route is the autoinfection cycle, in which the rhabditiform larvae mature directly within the intestine into filariform larvae, which penetrate the colon or the perianal skin to enter the circulation and complete the cycle of maturation into adult worms. This last route can increase the parasitic burden and permit the infection to become chronic.

Invasion of the small intestine wall by adult *S. stercoralis* worms may produce abdominal pain that is often localized to the midepigastrium; the pain is similar to that of peptic ulcer but is aggravated by food consumption. Diarrhea, nausea, and vomiting often occur; less commonly, urticaria, asthma, and weight loss may occur. Patients with heavy infection may experience malabsorption, gastrointestinal bleeding, and a protein-losing enteropathy. In cases of chronic infection, symptoms may be absent or may be mild and intermittent; symptoms include diarrhea, abdominal pain, and recurring episodes of urticaria, especially on the buttocks and the wrists. Less commonly, larva currens, which is a pathognomonic serpiginous, pruritic, elevated eruption, evolves along the tract of larval migration in the skin of the perianal, gluteal, or other body areas.

Disseminated strongyloidiasis may occur, especially in patients who have a malignant disorder or who are immunocompromised as a result of malnutrition or the administration of corticosteroids or other immunosuppressive medications.[18] Disseminated strongyloidiasis occurs but is uncommon in patients with AIDS, although strongyloidiasis is encountered in patients who are infected with human T cell lymphotropic virus type I.[18] In immunocompromised patients, the autoinfection cycle produces large numbers of filariform larvae that may disseminate from the colon and invade any organ system. Colonic leakage or carriage of intestinal bacteria by larvae into the bloodstream often produces concomitant bacterial infection and pneumonia. This progression of events frequently results in a fatal hyperinfection syndrome in the immunocompromised host.

Laboratory Findings

Eosinophilia is elicited in strongyloidiasis but may be minimal or only episodic in chronic infection and absent in the hyperinfection syndrome, especially if the patient has taken corticosteroids. Midepigastric discomfort or unexplained eosinophilia in a patient about to receive immunosuppressive therapy should prompt a diagnostic search for the larvae of *Strongyloides*. Serial stool samples should be examined with the use of concentration techniques. Stool examinations are repeatedly negative in about 25% of patients with strongyloidiasis. An enzyme-linked immunosorbent assay (ELISA) is an especially useful diagnostic test. When hyperinfection is present, *Strongyloides* larvae may be found in other fluids, such as surgical drainage fluid and sputum, and the stool may contain filariform and rhabditiform larvae.

Treatment

Strongyloidiasis must be treated because of the potential for subsequent fatal hyperinfection. The anthelmintic agents used are ivermectin, thiabendazole, and albendazole [*see Table 1*]. Ivermectin, approved in the United States for uncomplicated infection, is safe and effective. Thiabendazole can be used for uncomplicated infection or for disseminated disease.[19] However, thiabendazole often causes unpleasant side effects, and treatment is

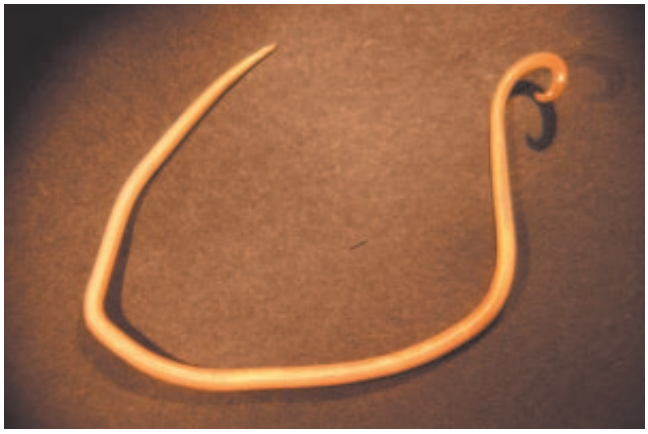


Figure 4 **Adult *Ascaris lumbricoides* roundworms are 10 to 30 cm in length.**

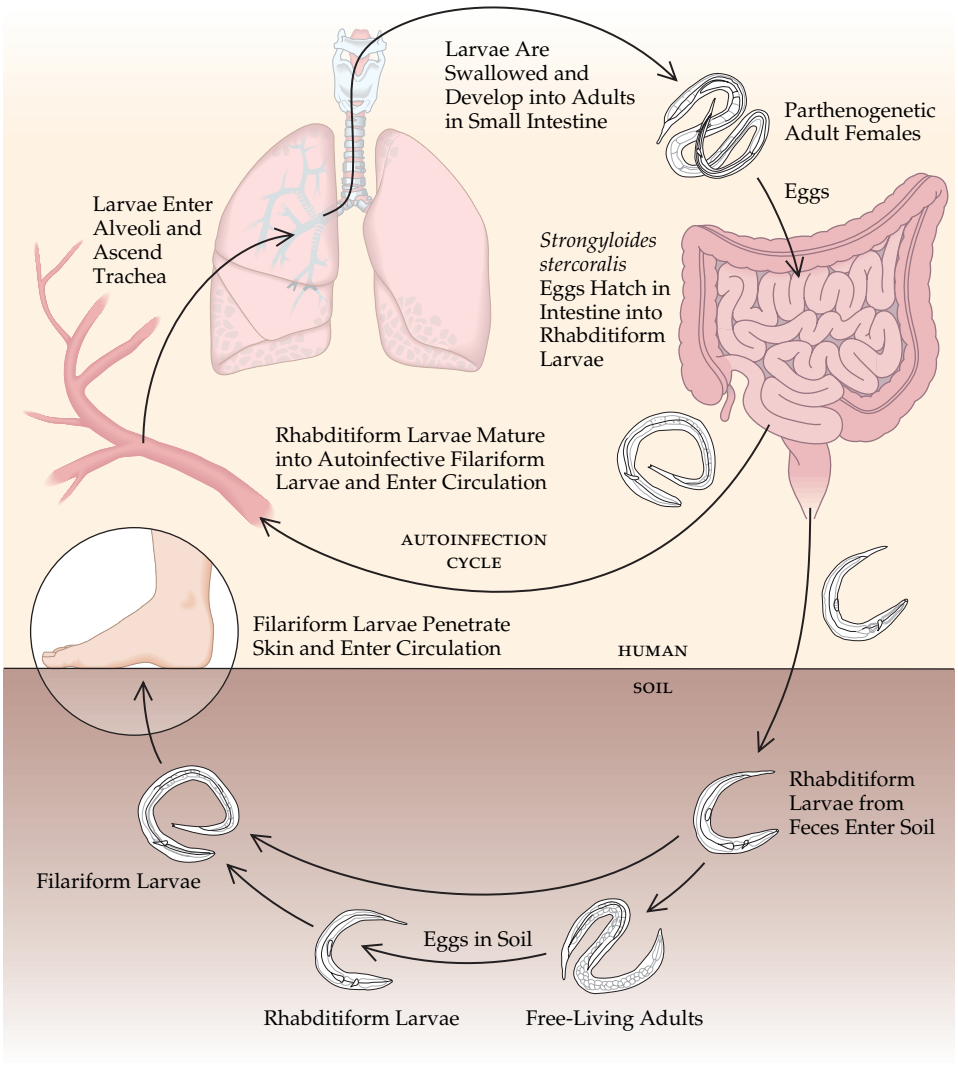


Figure 5 **Filariform larvae of the intestinal nematode *Strongyloides stercoralis* penetrate the skin or mucous membranes and enter the circulation. Larvae travel to the lungs, where they cross into the alveolar air space and ascend the tracheobronchial tree. They are then swallowed and develop into adults, which may mate or undergo parthenogenetic fertilization, penetrate the mucosa of the proximal small bowel, and lay eggs. The eggs hatch within the small intestine and develop into rhabditiform larvae, which then follow one of two paths: a soil phase or an autoinfection cycle. In the soil phase, larvae are passed in the feces and contaminate the soil, where they either transform directly into infectious filariform larvae or develop into free-living adults that generate new infectious larvae. In the autoinfection cycle, rhabditiform larvae penetrate the colonic wall or perianal skin and enter the circulation.**

not uniformly successful. Many experts believe that ivermectin will completely supplant thiabendazole. Although not approved for the treatment of strongyloidiasis in the United States, albendazole is an effective agent but is less effective than ivermectin in clearing larvae from stools.[18] Serial stool examinations should be obtained in the weeks and months after therapy to confirm eradication of the parasite. Immunocompromised hosts may require prolonged anthelmintic therapy.

TRICHOSTRONGYLIASIS

Although *Trichostrongylus* are usually intestinal pathogens of herbivores, several species may infect humans. Infection is most common in Asia and is acquired by consuming food that harbors infectious larvae. Adult worms attach to the proximal small bowel and in most cases do not produce symptoms, but mild GI distress and eosinophilia have been reported. The eggs of *Trichostrongylus* are somewhat larger and have less rounded ends than hookworm eggs. Trichostrongyliasis is treated with pyrantel pamoate (a single oral 11 mg/kg dose, to a maximum of 1 g), mebendazole (100 mg p.o., b.i.d., for 3 days), or albendazole (400 mg p.o., once).[19]

CAPILLARIASIS

Most human infection with *Capillaria* is caused by *C. philippinensis,* which is found primarily in the northern Philippines, Thailand, and Egypt. Humans acquire the infection by eating raw freshwater fish, but birds that eat fish are the major host. The small (3 to 4 mm long) adult worms dwell in the small intestine. Females produce unembryonated eggs that mature in the environment. However, a small percentage of these eggs embryonate and hatch in the intestine, leading to an internal autoinfection similar to that seen in strongyloidiasis. Symptomatic intestinal capillariasis is marked by diarrhea and abdominal pain. When infection is severe, edema, muscle wasting, fat and sugar malabsorption, and a protein-losing enteropathy may ensue.[14] Because of the autoinfection cycle, untreated severe infection may result in death after several months.

Hepatic and pulmonary capillariasis are rare diseases that are found worldwide. Hepatic capillariasis, from *C. hepatica,* manifests as acute or subacute hepatitis; pulmonary capillariasis, from *C. aerophila,* manifests as fever, cough, asthma, and pneumonia. Capillariasis is treated with mebendazole (200 mg orally twice daily for 20 days) or albendazole (400 mg orally daily for 10 days).[19]

CANINE HOOKWORMS

Two canine hookworms, *Ancylostoma ceylanicum* and *A. caninum,* can cause human intestinal infections; a third canine hook-

worm, *A. braziliense,* can cause cutaneous larva migrans [*see* Larva Migrans, *below*]. *A. ceylanicum* is a widely distributed canine and feline parasite of the tropics and subtropics. Like human hookworm parasites, the sexually mature, egg-laying adult worms of this species can cause blood loss and consequent iron deficiency.

Most cases of *A. caninum* infection in humans have been recognized in Australia and Egypt.[20,21] Reported clinical features have been variable, ranging from acute intestinal obstruction to subclinical infections. Eosinophilic enteritis is common, and some patients have experienced acute or recurrent abdominal pain, rectal bleeding, or diarrhea. Worms—often a single worm—have been recovered at surgery or colonoscopy in many patients. The immature worms do not produce eggs and hence are not detectable by stool examinations. An ELISA serologic test has helped with diagnosis. As in human hookworm infections [*see Table 1*], mebendazole or albendazole has been effective.

Tissue Nematode Infections

TRICHINELLOSIS

Trichinella are nematodes whose larvae typically become encysted in striated muscle. Transmission of *Trichinella* occurs via ingestion of infected meat, usually pork or the meat of certain carnivores. Mild to moderate trichinellosis is usually asymptomatic or mildly symptomatic; heavy infection can cause myalgias, periorbital edema, eosinophilia, and, in rare cases, death.

Etiology and Epidemiology

Five species of *Trichinella* are now recognized to infect humans. *T. spiralis,* found in many carnivorous and omnivorous animals, and *T. pseudospiralis,* found in mammals and birds, are distributed throughout the world. *T. nativa* is present in arctic regions and infects bears and arctic mammals; *T. nelsoni* is present in equatorial Africa, where it is common in felid predators and scavenger animals (e.g., hyenas and bush pigs); and *T. bitovi* is present in temperate areas of Europe and western Asia, where it is found in carnivores but not domestic swine.[22,23]

Trichinella is transmitted by the consumption of meat containing encysted larvae [*see Figure 6*]. In its sylvatic life cycle, the parasite is ingested by wild animals that eat trichinous carcasses; in its domestic life cycle, the parasite is ingested by pigs that eat contaminated meat in garbage. Worldwide, most human trichinellosis is caused by ingestion of infected pork products; in the United States, however, ingestion of wild carnivores has surpassed pork in transmitting the infection.[24] Although cattle are

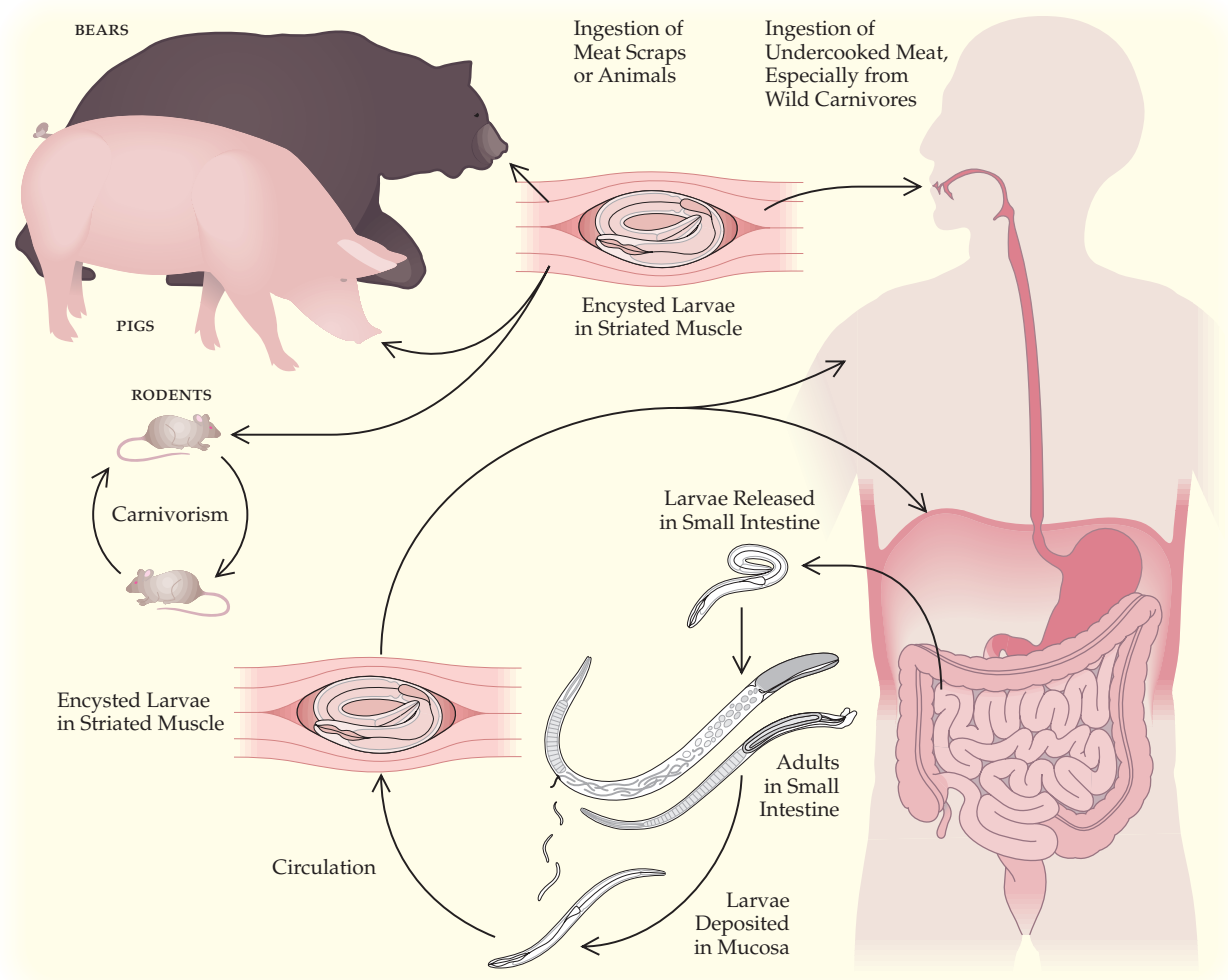


Figure 6 ***Trichinella* infection is acquired by the consumption of meat containing encysted larvae. Digestion liberates the larvae, which penetrate into the small intestine, mature into adult worms, and mate. Adult females produce larvae, which enter the circulation and spread to striated muscle, where they become encysted.**

not natural hosts of the parasite, beef may be implicated in outbreaks because of adulteration with pork, either intentional or incidental, through the common use of a meat grinder. Human trichinellosis can also be acquired from horse meat or the meat of wild animals such as bears, walrus, boars, and cougars.[24,25]

Although meat in the United States is not inspected for *Trichinella* larvae, laws proscribing the feeding of uncooked garbage to pigs have reduced the transmission of the disease.[24] Most cases in the United States are attributable to noncommercial pork, bear, and cougar meat. About 15 cases of trichinellosis, occasionally leading to death, are reported annually in the United States, but more cases, especially mild ones, probably remain undiagnosed.[24]

Larvae in pork may be rendered noninfective by heating to a temperature of 77° C. Freezing at –15° C for 3 weeks, as in a home freezer, will generally kill larvae in meat; however, arctic species of the parasite, which may be present in walrus or bear meat, are more resistant to freezing and thus may remain viable.

Pathogenesis and Clinical Features

The severity of clinical illness in trichinellosis usually correlates with the number of ingested larvae. In most infections, there are only one to 10 larvae per gram of muscle, and prominent symptoms are absent; however, in some infections, there are more than 50 larvae per gram of muscle. The symptoms of these heavier infections may reflect the two phases of the parasite's development (i.e., the intestinal phase and the muscle phase).

Intestinal phase Encysted larvae are liberated from infected meat by the action of gastric acid and peptic enzymes. The larvae burrow into the villi of the small intestine and mature into adult worms. Within 4 to 5 days, the adult worms produce larvae, which enter the circulation. During this enteric phase, which usually lasts from 1 to 7 days, patients may be asymptomatic or may experience nausea, vomiting, constipation, or abdominal aches. Prolonged diarrhea that lasts for weeks, as noted in patients with arctic trichinellosis, may be the result of secondary infections in previously infected and sensitized patients.[26]

Muscle phase After the intestinal phase, larvae are carried by the bloodstream to various organs. In the early muscle phase, periorbital and facial edema, subconjunctival and retinal hemorrhages, and subungual splinter hemorrhages commonly develop.

After about 3 weeks of infection, larval encystment in muscles begins. Patients experience myalgias, muscle edema and weakness, fever, and eosinophilia. Headache, cough, dyspnea, dysphagia, and macular or urticarial skin lesions develop less commonly. Although larvae only become encysted in striated muscle, inflammatory lesions may develop in the heart, lungs, and central nervous system. Deaths usually result from myocarditis. Symptoms generally abate slowly after the third week of infection. Infections with *T. pseudospiralis* may be associated with years of myositis.[27]

Laboratory Findings

Eosinophilia develops in more than 85% of patients with symptomatic trichinellosis.[24] Elevations of muscle enzymes (i.e., serum aspartate aminotransferase and creatine kinase) often accompany symptomatic muscle involvement. Serologic tests are positive in most patients with acute infection, but diagnostic antibody titers may not be detectable until the third week of infection.

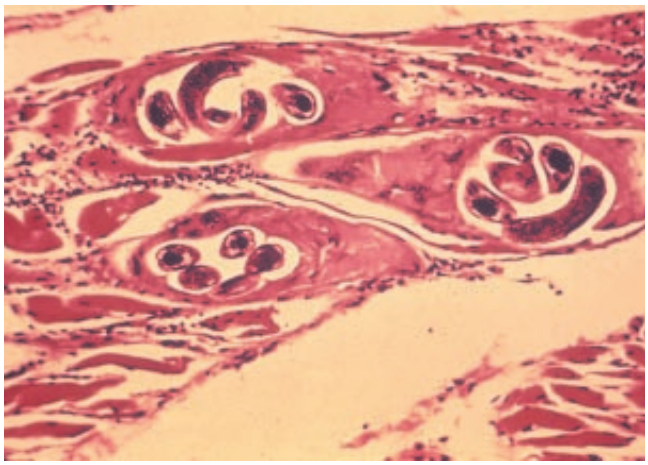


Figure 7 **A micrograph shows developing *Trichinella* cysts in human muscle tissue.**

Definitive diagnosis requires demonstration of larvae in muscle [*see Figure 7*]. The likelihood of obtaining a positive muscle biopsy specimen increases if the specimen is taken from a clinically involved muscle; if it is taken near a tendinous insertion, where the density of larvae is greatest; and if it is obtained after the third week of infection, when larvae are more numerous, larger, and more resistant to digestion. A portion of the muscle biopsy specimen should be submitted for histopathologic examination. In addition, another portion should be compressed between glass slides and examined microscopically for larvae. Because a larger volume of muscle can be examined by this technique, it is more sensitive than histopathologic examination for detecting larvae. Differentiation of *Trichinella* species can be based on the absence of larval encystment with *T. pseudospiralis* and on DNA-based molecular biologic techniques.[23]

Differential Diagnosis

The triad of periorbital edema, myalgias, and eosinophilia strongly suggests the diagnosis of trichinellosis. Most patients with trichinellosis have a history of consuming pork products or the meat of wild mammals. The development of symptomatic infection or eosinophilia in other persons who have eaten the meat incriminated in a case of trichinellosis provides epidemiologic confirmation of the diagnosis. Persons with light infections may have minimal symptoms, or the symptoms may suggest diagnoses other than trichinellosis, such as diarrheal or influenzal syndromes or polymyositis. In rare instances, CNS involvement may suggest aseptic meningitis or encephalitis.[28]

Treatment

Specific therapy for trichinellosis is unsatisfactory. Most infections, however, are not life threatening and are self-limited; bed rest, analgesics, and antipyretics suffice to alleviate myalgias and fever. If serious cardiac, neurologic, or pulmonary involvement occurs, corticosteroid therapy with prednisone (40 to 60 mg daily) is indicated. Prednisone is often given with oral mebendazole (200 to 400 mg t.i.d. for 3 days, then 400 to 500 mg t.i.d. for 10 days) or oral albendazole (400 mg b.i.d. for 8 to 14 days).[19] Limited experience suggests that *T. nativa* and *T. pseudospiralis* infection may respond to albendazole.[27,29]

ANISAKIASIS

Anisakiasis is a zoonotic infection that develops in persons who consume fish infected with anisakine or related parasites.[30] *Anisakis* species are present in the muscle of mackerel, herring, and salmon. In a study in the Pacific Northwest of the United States, all salmon caught in the wild had *A. simplex* larvae, and 87% of the fish had larvae in their edible musculature.[31] Infection caused by *Pseudoterranova* (previously termed *Phocanema*) species is referred to as codworm anisakiasis. *Pseudoterranova* parasites are found most frequently in cod, but they have also been detected in the muscle of pollack, halibut, flatfish, greenling, Pacific rockfish (Pacific red snapper), and squid. Infective third-stage larvae in fish are killed by heating to 60° C, by commercial blast freezing, or by freezing at –20° C for 3 days. Salting, cold smoking, and marinating do not normally kill either *Anisakis* or *Pseudoterranova* larvae.

The number of cases of anisakiasis recognized in the United States, Japan, Europe, and other countries has increased in recent years, probably as a result of three developments.[30] First, raw fish dishes, including sushi, sashimi, and lomi lomi salmon, have become increasingly popular. Second, the population of marine mammals, including sea lions and seals, has increased as a result of legal protection; these marine mammals are the definitive hosts for anisakine nematodes that in a later stage of their life cycle infect fish. Third, there has been an increase in awareness of anisakiasis and, especially in Japan, greater use of endoscopy as a diagnostic modality.

In addition to anisakine nematodes, other parasites, including *Diphyllobothrium latum* [*see* Cestode Infections, Fish Tapeworm, *below*], the intestinal fluke *Nanophyetus salmincola,* and larval *Eustrongylides* nematodes, can be acquired by eating raw fish.[30]

Pathogenesis and Clinical Features

The clinical manifestations of infection after the ingestion of fish containing live anisakine larvae depend in part on the genus of the infecting parasite and on the anatomic localization of the larvae. Typically, in the United States, infection with *Pseudoterranova* species becomes apparent when a patient coughs up a live worm, often within 48 hours after eating raw or undercooked fish. Some patients have experienced mild epigastric pain and a tingling throat and have sensed the worm in the oropharynx or proximal esophagus. In the United States, the infection has only infrequently been associated with serious illness.[32] In Japan, by contrast, codworm anisakiasis has been associated with gastric, but usually not intestinal, invasion. Within several hours after infected fish have been consumed, the larvae can invade the stomach mucosa and produce severe epigastric pain that recurs every 5 to 10 minutes. Patients may experience nausea, vomiting, urticaria, and a tingling throat. Endoscopy at this time will reveal the invading larvae at a site that is swollen and hemorrhagic; often, the site is the posterior wall of the stomach.

Infections with *Anisakis* species can cause either gastric or intestinal involvement. Invasion of the gastric mucosa produces the same clinical syndrome that occurs in gastric codworm anisakiasis but without the tingling throat; the pathologic findings on endoscopy are also the same.[33] Larval invasion of the intestinal mucosa can produce abdominal pain, fever, nausea, vomiting, and diarrhea. Larvae can elicit an eosinophil-rich granulomatous reaction in the submucosa and infrequently penetrate the peritoneum. The clinical presentation of intestinal anisakiasis often mimics that of acute appendicitis, but it can also be confused with gastrointestinal ulcers or neoplasms, acute Crohn disease, or ischemic ileitis. In addition to causing invasive disease, *Anisakis* parasites may elicit allergic reactions.[30] Patients may develop IgE-mediated reactions to *Anisakis* proteins and experience urticaria, angioedema, and anaphylactic reactions after ingesting even adequately cooked infected fish.[30]

Laboratory Findings and Imaging Studies

In gastric anisakiasis, eosinophilia is often present, and occult blood is frequently detectable in gastric fluid or stool. In intestinal anisakiasis, eosinophilia may be absent, and a leukocytosis may develop. In either form of anisakiasis, a history of recent raw fish ingestion before the onset of symptoms can strongly suggest the diagnosis. The definitive diagnosis of anisakiasis involving the stomach or proximal intestine can be made with endoscopy, which will reveal the invading larva.[32,33] The codworm larva resembles a thick piece of yellowish string, whereas the *Anisakis* larva is fine and whitish. *Eustrongylides* larvae, which have infrequently been reported to cause intestinal perforation in humans, have a distinctive bright-red coloration.[34]

Radiographic studies can help identify gastric anisakiasis.[33] In intestinal anisakiasis, such studies may show thickening of the involved bowel area and may detect the larva.[35]

Treatment

Endoscopic or surgical removal of the larva is the only proven treatment. Although surgery for presumed appendicitis is often performed in patients with intestinal anisakiasis, surgical resection may not be needed if the correct diagnosis is considered in time. With conservative care, the illness may be self-limited, and patients may recover without surgical treatment.[35] The efficacy of anthelmintic medications against human anisakiasis has not been properly evaluated. Animal studies suggest that ivermectin and albendazole have activity against *A. simplex,* and there are anecdotal reports of successful treatment of human anisakiasis with albendazole; because anisakiasis is often self-limited, however, the true utility of albendazole therapy is unknown.[19,36]

LARVA MIGRANS

Humans may occasionally act as the nondefinitive host for nematode parasites that normally infect animals. In such cases, the parasites usually do not mature completely, but the introduced larvae may persist and induce an inflammatory reaction in the skin, viscera, or eyes.

Cutaneous Larva Migrans

The syndrome of cutaneous larva migrans (creeping eruption) develops when the larvae of various parasites, including the dog or cat hookworm *A. braziliense,* penetrate human skin. Pruritic, serpiginous cutaneous lesions form along the migratory tracts of the larvae [*see Figure 8*].[37,38] These lesions can be treated with oral albendazole (400 mg daily for 3 days), oral ivermectin (200 μg/kg daily for 1 to 2 days), or topical thiabendazole.[19,37]

Visceral and Ocular Larva Migrans

The syndrome of visceral larva migrans develops when nematode larvae of animal parasites migrate in human tissues. Visceral larva migrans with fatal eosinophilic meningoencephalitis has been caused by the raccoon ascarid *Baylisascaris procyonis,*[39] but visceral larva migrans is most commonly caused by dog or cat ascarids.

Toxocara canis, the canine roundworm, is common in North American dogs. Although fewer than 20% of adult dogs are in-

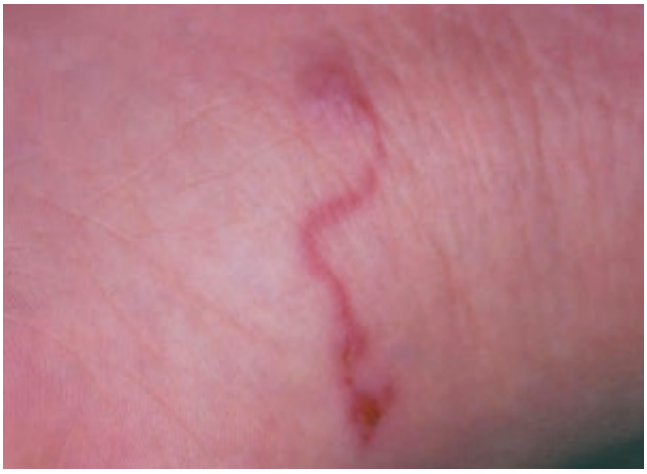


Figure 8 **Cutaneous larva migrans in a person who had recently been on a beach vacation in Haiti is evident as a serpiginous tract that moves over time.**

fected, transplacental migration of larvae from infected females results in an incidence of infection that may exceed 80% in their puppies at 2 to 6 months of age.[40] Both the puppies and the lactating mother shed large numbers of *T. canis* eggs in their stools within a month after parturition. The prevalence of *T. cati* in cats is generally lower than that of *T. canis* in dogs; kittens do not become infected before birth.

After *Toxocara* eggs are shed in the feces, about 3 weeks must elapse before the eggs develop sufficiently to become infectious; thereafter, they may remain viable for several months. Humans acquire infection principally by ingesting eggs in soil; public playgrounds have been implicated [*see Figure 9*]. Direct contact with infected animals is not a major source of infection because of the time required for eggs to become infectious. A history of eating soil is strongly associated with the risk of acquiring infection. Thus, visceral larva migrans is most likely to develop in children with geophagia.

Pathogenesis and clinical features After infectious *Toxocara* eggs are ingested, the larvae hatch and penetrate the gastrointestinal mucosa. They are carried to the liver in the portal circulation; from there, they move into the systemic circulation. When the larvae enter vessels too small to allow their passage, they exit and migrate into the surrounding tissues. The manifestations of visceral larva migrans are a consequence of both the damage done by the migrating *Toxocara* larvae and the induced eosinophilic granulomatous inflammatory reaction.

Many infections are mild and subclinical. Clinically apparent infections may present in one of two distinct patterns: visceral and ocular.[41] In the visceral form of infection, patients are usually 1 to 5 years of age and generally have a history of geophagia. Malaise, irritability, weight loss, wheezing, cough, fever, hepatomegaly, and pruritic cutaneous eruptions are common. Neurologic involvement may produce seizures and behavioral disorders. In rare instances, death results from severe neurologic or myocardial involvement.

In the ocular form of larva migrans, patients are older: the mean age is 7.5 years, and the range includes young adults. A history of geophagia is unusual; a history of an antecedent symptomatic visceral form of the disease is rare. Common symptoms of ocular larva migrans are strabismus and failing vision. The characteristic ocular lesion is a whitish elevated granuloma measuring one to two disk diameters that is located in the posterior pole of the retina; however, the disease may also present as endophthalmitis or uveitis. The ocular lesions may be mistaken for a retinoblastoma, which can lead to unnecessary surgical enucleation of the eye.

Laboratory findings and imaging studies Usual findings in the visceral form of larva migrans are leukocytosis, prominent blood eosinophilia (generally in excess of 1,000/μl), and hypergammaglobulinemia (increased levels of IgG, IgM, and IgE). On chest x-ray, transient pulmonary infiltrates are found in about half of the patients with pulmonary symptoms. All these findings are uncommon, however, in the ocular form of larva migrans.

An ELISA serologic test has proved to be a sensitive, specific diagnostic test for both the visceral and the ocular forms; it can detect subclinical cases that may present only as an unexplained eosinophilia.[40,41]

Treatment The visceral form of larva migrans is often self-limited. However, anthelmintic and anti-inflammatory therapy have been shown to be beneficial. Albendazole (400 mg p.o., b.i.d., for 5 days), though not licensed by the FDA for this indication, is the drug of choice for toxocariasis and has been shown to be superior to thiabendazole.[41] Corticosteroids are used to treat patients with severe respiratory, myocardial, or CNS involvement or, in the ocular form, to suppress active inflammation.

ANGIOSTRONGYLIASIS

Infection of humans with the parasite *Angiostrongylus cantonensis* may produce eosinophilic meningitis.[39] The rat is the definitive host of *A. cantonensis,* but the larval stages develop in mollusks, which serve as intermediate hosts. Humans acquire the infection either by ingesting infected mollusks (i.e., land snails or slugs) or by consuming inadequately cooked carrier hosts (i.e., freshwater shrimp, aquatic and amphibious crabs, and certain marine fish) that have previously ingested infected mollusks. In the United States, *A. cantonensis* is endemic to the Hawaiian Islands, and the parasite has been found in rats in New Orleans. *A. cantonensis* is found throughout the Pacific Basin and in Southeast Asia, Australia, and the Caribbean. Eosinophilic meningitis from this parasite occurs in all these areas. An outbreak causing eosinophilic meningitis in 12 travelers on a cruise ship in Jamaica has been described; ingestion of Caesar salad was highly associated with acquisition of disease.[42]

Pathogenesis and Clinical Features

After *A. cantonensis* larvae are ingested, they migrate to the CNS. There the parasites die, but their passage into the CNS induces an eosinophilic inflammatory reaction. The most common symptom of *A. cantonensis* infection is a severe frontal or bitemporal headache, which is usually the reason patients seek medical attention. More than half of patients experience stiff neck, vomiting, or paresthesias. Extraocular muscle palsies, seizures, and significant fever are uncommon. Although some deaths have been recorded, symptoms generally resolve in 1 to 2 weeks; however, paresthesias may persist for months.[39]

Laboratory Findings

Peripheral blood eosinophilia usually accompanies eosinophilic meningitis caused by *A. cantonensis;* however, the eosinophilia may be mild, and eosinophil levels may not be elevated ini-

tially. In the cerebrospinal fluid, the white blood cell counts ordinarily range from 150 to 1,500 μl, with eosinophils accounting for more than 20% of the leukocytes. CSF protein levels are usually elevated, and glucose levels are at the low end of the normal range. On rare occasions, larvae of *A. cantonensis* are found in the CSF. No serologic tests are available for identifying the organism, so diagnosis is based on a compatible epidemiologic history and clinical course together with eosinophilic pleocytosis in the CSF. Eosinophils in the CSF, which are also found in other diseases,[39] may be misidentified if they are not specifically stained.

Treatment

There is no specific treatment for eosinophilic meningitis caused by *A. cantonensis;* anthelmintic agents have no effect. Corticosteroids have been shown to provide symptomatic relief of headaches caused by eosinophilic meningitis, as has removal of cerebrospinal fluid in patients with increased intracranial pressure.[19,43] For most patients, the illness is self-limited and the prognosis is good. In one report, treatment with the combination of mebendazole and corticosteroids appeared to shorten the duration of illness.[19,39,44]

MAMMOMONOGAMOSIS (SYNGAMOSIS)

Nematodes of the genus *Mammomonogamus* (*Syngamus*) inhabit the trachea of cattle and other herbivores. Although the life cycle of the parasite is unknown, human infection is believed to result from the ingestion of foodstuffs that contain some intermediate host of the parasite. Of the several dozen cases recognized in humans, most have been acquired in the Caribbean area. A chronic, nonproductive cough is characteristic in such patients.[45] No blood eosinophilia is elicited, and no parenchymal lesions are seen on chest x-ray. The diagnosis is made when an

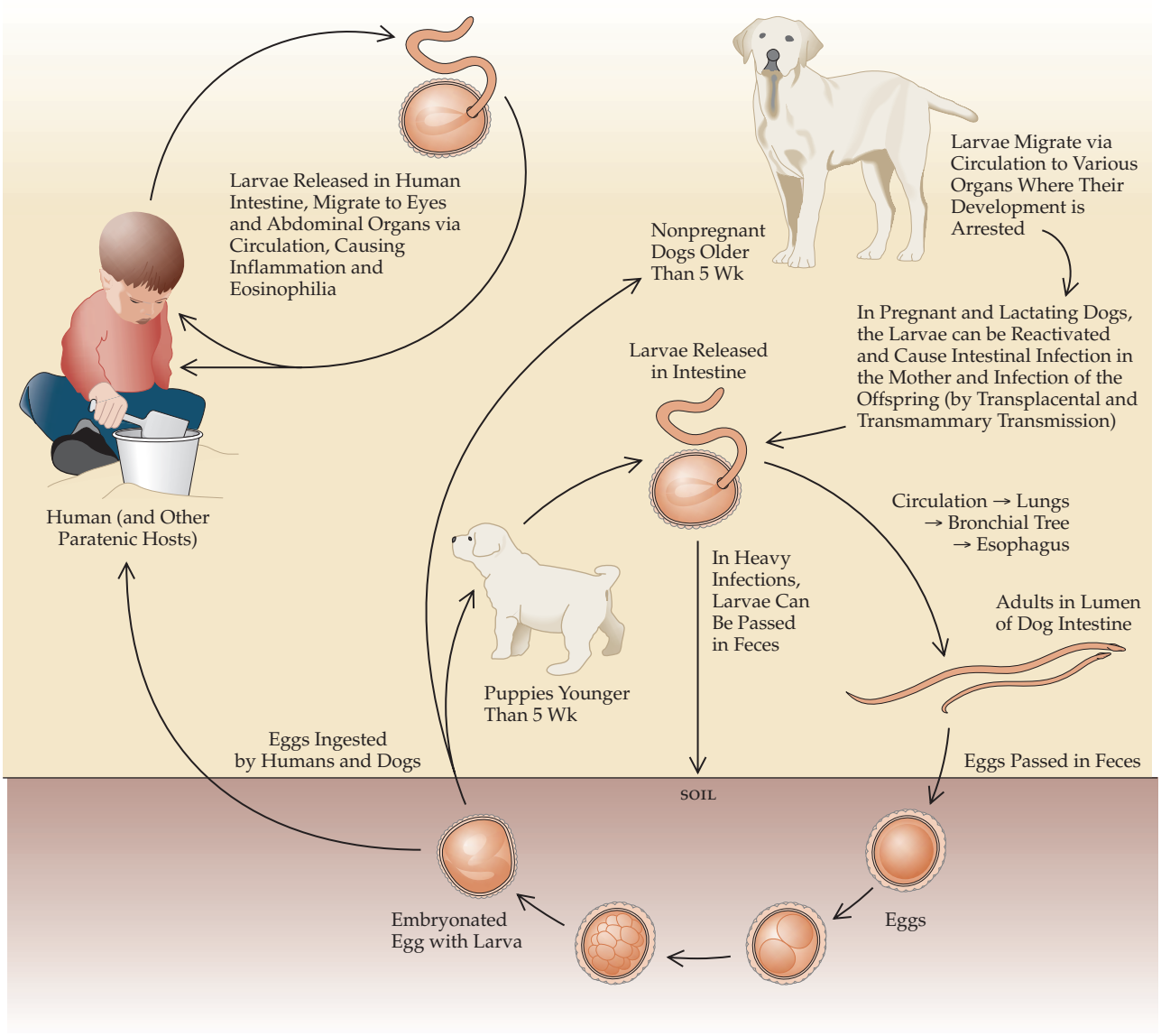


Figure 9 **Life cycle of *Toxocara canis*, one of the principal agents of visceral larva migrans. Humans are nondefinitive hosts for *Toxocara;* after infectious eggs are ingested, the larvae hatch and penetrate the gastrointestinal mucosa. They are carried to the liver in the portal circulation; from there, they move into the systemic circulation. When the larvae enter vessels too small to allow their passage, they exit and migrate into the surrounding tissues, where they may cause an inflammatory reaction.**

Table 2 Filarial Parasites of Humans

	Distribution	*Vector*	*Pathology*	*Localization of Adult Worms*	*Localization of Microfilariae*
Wuchereria bancrofti	Tropics worldwide	Mosquito	Lymphatic, pulmonary	Lymphatic tissues and vessels	Blood
Brugia malayi	Southeast Asia	Mosquito	Lymphatic, pulmonary	Lymphatic tissues and vessels	Blood
Brugia timori	Indonesia	Mosquito	Lymphatic	Lymphatic tissues and vessels	Blood
Onchocerca volvulus	Africa, Central and South America	Blackfly	Cutaneous, ocular	Nodules in subcutaneous tissues	Skin, subcutaneous tissues
Loa loa	Central West Africa	Deerfly	Subcutaneous	Subcutaneous and subconjunctival tissues	Blood
Mansonella perstans	Africa, South America	Midge	Uncertain, possibly allergic	Body cavities	Blood
Mansonella ozzardi	Central and South America	Midge, blackfly	Uncertain	Body cavities	Blood, skin
Mansonella streptocerca	Africa	Midge	Cutaneous	Connective tissues	Skin

adult worm is expectorated or found at bronchoscopy or when eggs are found in sputum or feces. Therapy consists of the removal of the worms during bronchoscopy, but mebendazole and thiabendazole have been used in case reports.

GNATHOSTOMIASIS

Nematodes of the genus *Gnathostoma* are widely distributed and usually infect the intestinal tract of carnivorous animals. Most human infections have been reported from Asia and have been caused by *G. spinigerum,* although human gnathostomiasis has also been described in Central and South America.[46,47] Infectious larvae are found in the tissues of freshwater fish, eels, frogs, and snakes and in animals fed on infected fish (e.g., poultry and pigs). Humans may be infected if they ingest any infected tissues, or possibly by accidental ingestion of the copepod (a tiny freshwater crustacean) that serves as an intermediate host. The larvae fail to develop to maturity within humans, but acute symptoms of nausea, vomiting, and urticaria may develop 1 to 2 days after infection. Larval penetration of the liver produces inflammation and right upper quadrant pain. Subsequently, larvae may migrate within the abdominal or thoracic cavities. Eosinophilia is initially marked but diminishes after about a month, when the larvae reach subcutaneous tissues. Ensuing larval migrations produce foci of serpiginous cutaneous or migratory subcutaneous inflammation and swelling. On occasion, larvae enter the CNS and cause eosinophilic meningitis.[39] Treatment options are oral albendazole (400 mg b.i.d. for 21 days), ivermectin (200 μg/kg/day for 2 days), surgical removal, or both.[19]

DRACUNCULIASIS

Dracunculus medinensis, the guinea worm, infects persons in 12 countries of Central and West Africa. A global eradication effort has reduced the annual incidence from 3.5 million cases in 1986 to less than 33,000 cases in 2003.[48] Infection results from drinking water containing microcrustacean *Cyclops* species that harbor larvae of the parasite. Gastric acid digests the *Cyclops,* releasing the larvae. The larvae then penetrate the duodenum, where they develop into adult worms within the subcutaneous tissues over the course of 60 to 90 days. Male and female worms then mate; after mating, adult male worms die. Adult females, however, move into the lower extremities, grow to 70 cm in length or more, and then emerge through the skin after about a year. This emergence, which may be preceded by a few days of urticaria, fever, and dyspnea, occurs with the formation of a blister that becomes pruritic and painful before rupturing. Larvae are released from the ruptured blister site if the lesion is immersed in water, as occurs during wading to fetch drinking water. The life cycle is completed when larvae are ingested by *Cyclops* and mature within them.

Clinical manifestations of dracunculiasis are usually related to the emergence of the worm and include premonitory symptoms and blister formation. Secondary bacterial infection often develops at the site of the blister and may extend retrograde along the length of the worm.

The conventional method of winding several centimeters of slowly emerging worm around a stick each day until the worm has fully emerged is an effective treatment. Removal is eased by administration of metronidazole (250 mg p.o. t.i.d. for 10 days), but surgical extraction is more rapid and effective.[49] In endemic areas, preventive efforts focus on breaking the *Dracunculus* life cycle by improving the safety of drinking water.

FILARIASIS

Several insect-transmitted filarial nematodes cause chronic infections in humans [*see Table 2*]. The major filarial nematodes are responsible for lymphatic filariasis, onchocerciasis, and loiasis. Other filarial nematodes include *Mansonella ozzardi, M. perstans* (formerly called *Tetrapetalonema perstans*), and *M. streptocerca* (formerly called *T. streptocerca*); these nematode species cause fewer cases of human infection than the major filarial nematodes and are of less certain pathogenicity, but they are thought to cause bronchospastic tropical pulmonary eosinophilic syndrome.

Lymphatic Filariasis

The most common filarial infections are caused by lymphatic tissue–dwelling filarial parasites, three species of which infect humans. It is estimated that 120 million persons in 76 countries are infected with lymphatic filarial parasites and that 44 million have overt clinical disease. However, the Global Alliance to Eliminate Lymphatic Filariasis (http://www.filariasis.org) has

brought the eradication of lymphatic filariasis within reach over the next decade by bringing together a strong group of collaborators whose efforts include the provision of powerful antiparasitic drugs.

Wuchereria bancrofti, the most common lymphatic filarial parasite, is broadly distributed in tropical regions. *Brugia malayi* is found in areas of Southeast Asia, and *B. timori* is found in parts of Indonesia. Mosquitoes serve as an intermediate host for all three species of filarial nematodes and introduce infectious larvae into humans by means of their bite (the life cycle of *B. malayi* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 3379). Different mosquito species are responsible for transmission in different areas; the major vectors include *Culex, Anopheles, Aedes,* and *Mansonia* species. The larvae develop into adult worms, which reside within lymphatic vessels and tissues. Offspring of the adult worms, called microfilariae, are covered by sheaths and may circulate in the bloodstream. In most endemic regions, microfilariae of *W. bancrofti* circulate in the bloodstream in greatest numbers during the night; in the South Pacific, however, no pronounced diurnal variation occurs. The time of day when the greatest numbers of microfilariae circulate correlates with the time of day when the principal mosquito vectors feed. Thousands of mosquito bites are probably required to produce an infection that results in patent microfilaremia. The incubation period until patent microfilaremia for *W. bancrofti* develops ranges from 3 months to the more usual 8 to 12 months.

Pathogenesis and clinical features Most clinical manifestations of lymphatic filariasis are attributable to inflammatory reactions caused by the adult worm; the immunopathogenesis, however, is poorly understood.[50] Many patients with microfilaremia experience no symptoms, although lymphadenopathy of the axillary, epitrochlear, and inguinal-femoral areas is common.

The more prominent clinical manifestations of lymphatic filariasis may be divided into two categories: inflammatory and obstructive. Inflammatory manifestations, which often become recurrent, include lymphadenitis, lymphangitis, funiculitis, orchitis, and epididymitis.[50] A typical presentation is the triad of lymphadenitis, lymphangitis, and fever. In contrast to bacterial ascending lymphangitis, the lymphangitis of filariasis characteristically progresses centripetally down involved extremities. Staphylococcal and streptococcal superinfections are commonly the cause. Fever is variable and may be associated with frank rigors; the patient's temperature may reach 40° C (104° F). Systemic symptoms, such as nausea and vomiting, develop in many of these recurrent episodes. Urticaria and areas of cutaneous erythema commonly develop. The episodes may continue for as long as 7 to 10 days before spontaneously resolving. In filariasis caused by *Brugia* species, suppurative abscesses may form in areas of involved lymphatic channels.

The obstructive phase of filariasis usually develops after decades of exposure to the parasite and reflects lymphatic compromise, the mechanism of which is poorly understood. Inflammatory features may coexist with the obstructive phase. Features of the obstructive phase include hydrocele formation, chyluria, elephantiasis of the extremities (shown in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 373), and, less commonly, elephantiasis of the breast.[50]

Tropical pulmonary eosinophilia develops in some individuals infected with lymphatic tissue–dwelling filarial parasites.[51] This illness is characterized by paroxysmal asthmalike attacks that are often nocturnal; by blood and pulmonary eosinophilia; and by elevated titers of antifilarial IgG, IgM, and especially IgE. Tropical pulmonary eosinophilia should be distinguished from the transient eosinophilic pneumonitis termed Löffler syndrome, which results from a transpulmonary migration of larval forms of *Ascaris* and, less commonly, hookworm and *Strongyloides* [*see* Intestinal Nematode Infections, *above*].

Laboratory findings Definitive diagnosis of lymphatic filariasis requires the demonstration of microfilariae or microfilarial antigens in the blood.[52] Stained blood smears may reveal microfilariae, but often, larger volumes of blood need to be examined either by centrifugal concentration after erythrocyte lysis or by filtration through 3 μm polycarbonate (Nuclepore) filters. In endemic areas where *W. bancrofti* microfilariae display nocturnal periodicity, detection of microfilariae requires examination of blood obtained after midnight or 60 to 90 minutes after daytime administration of a 2 mg/kg provocative dose of diethylcarbamazine (care should be taken if the patient is from a region where onchocerciasis is endemic, because diethylcarbamazine may precipitate a severe reaction as a result of lysis of *Onchocerca* microfilariae). Because microfilarial antigen of *W. bancrofti* can be found in the blood throughout the day, antigen measurement is more convenient than direct microfilaria determination. Microfilariae are rarely found in patients with elephantiasis or with tropical pulmonary eosinophilia and often cannot be found in patients with inflammatory or obstructive filarial manifestations. In males, ultrasonography can detect adult filariae within scrotal lymphatics.[50] Serologic tests may be helpful in making a diagnosis. Mild eosinophilia and elevations of serum IgE are common in lymphatic filariasis.

Treatment Lymphatic filariasis can be treated with diethylcarbamazine (2 mg/kg orally three times daily for 2 to 3 weeks). Treatment with ivermectin or a single dose of diethylcarbamazine (6 mg/kg)[50] is also effective at clearing microfilaremia, and the combination of these two anthelmintics has led to rapid and long-term clearance of microfilariae. There is insufficient reliable research to ascertain whether albendazole, alone or in combination with diethylcarbamazine, is an effective treatment for lymphatic filariasis,[53] but there is growing evidence that such treatment helps to suppress microfilaremia in patients from endemic areas. These short-term and combination therapies have been used more often in treatment of populations in filariasis control projects; for individual treatment, most experts favor the use of diethylcarbamazine for 2 to 3 weeks. Diethylcarbamazine must be used cautiously in patients exposed to *Loa loa* or *Onchocerca volvulus,* because patients infected with these filariae tend to have severe reactions to this drug.

Onchocerciasis

O. volvulus, the causative agent of onchocerciasis (river blindness), is found in equatorial Africa and in elevated regions of Mexico and Guatemala, with smaller foci in Yemen, Brazil, Ecuador, and Venezuela. It is estimated that 17 million people are infected. *Simulium* species (blackflies) transmit the infection. Onchocerciasis control programs, which are based on community therapy with ivermectin, have led to a substantial reduction in the incidence of disease. Repeated therapy is necessary, however, and control programs will have to continue for decades to eliminate onchocerciasis altogether.

Onchocerciasis typically occurs within several kilometers of rapidly flowing rivers and streams where blackflies breed. The

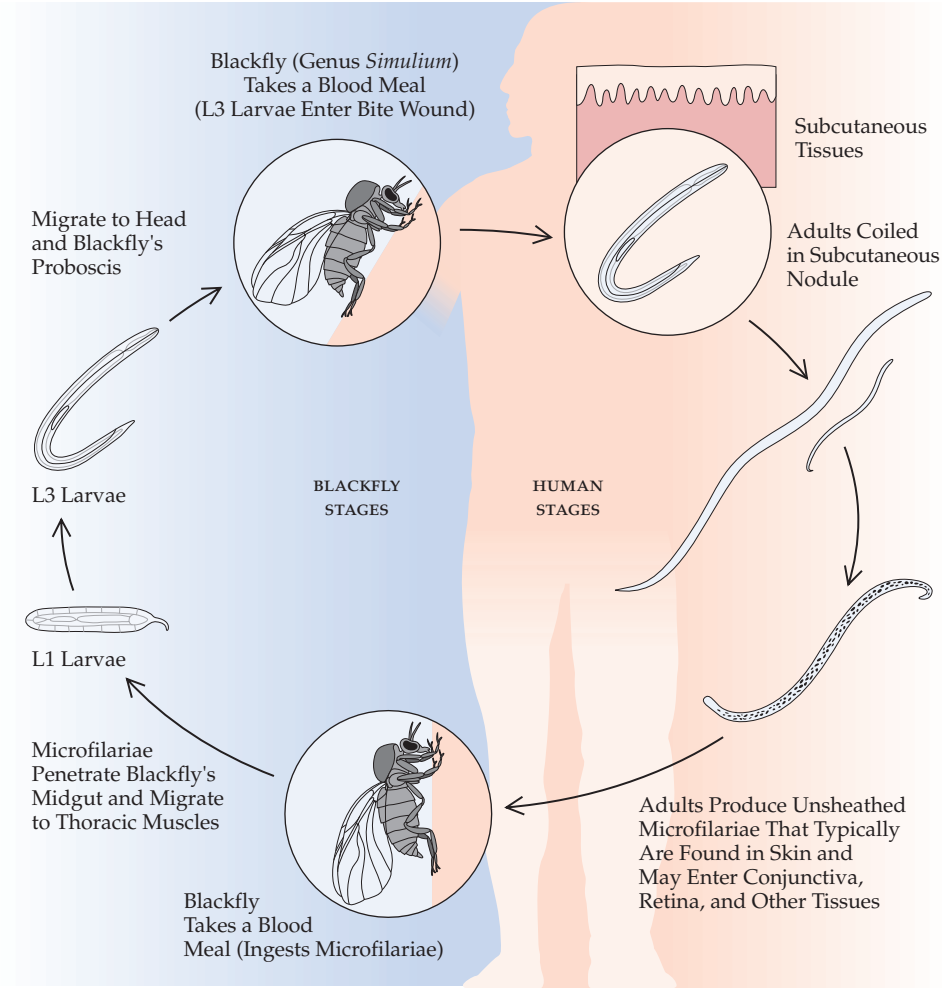


Figure 10 **Life cycle of *Onchocerca volvulus*, which causes river blindness and severe skin disease.**

life cycle of *O. volvulus* is similar to that of species causing lymphatic filariasis [*see Figure 10*]. Adult worms, however, reside in subcutaneous tissues, often enclosed in fibrous nodules. Microfilariae, which in this species lack an enveloping sheath, are released from female adults and localize in skin and subcutaneous tissues. Symbiotic endobacteria (of the genus *Wolbachia*) in filaria may cause some of the inflammatory damage of onchocerciasis.

Clinical features The skin is frequently involved in onchocerciasis, and pruritus is the most common clinical manifestation. With time, such complications as wrinkling, loss of elastic tissue, hypopigmentation or hyperpigmentation, papulovesicular lesions, and localized areas of eczematoid dermatitis may develop. Firm, nontender nodules containing adult worms surrounded by fibrous tissue are often palpable in subcutaneous tissues. In Central America, nodules commonly occur on the head; in Africa, nodules are more common over bony prominences of the body. Regional lymphadenopathy also develops.

Ocular involvement is characteristic of onchocerciasis and may result in blindness. Conjunctivitis with photophobia is common. Punctate keratitis, which is caused by the accumulation of inflammatory cells around dying microfilariae, may develop within the cornea and usually resolves without consequence. However, sclerosing keratitis and chorioretinal lesions may ensue and are the major causes of onchocercal blindness. Anterior uveitis, iridocyclitis, and, less frequently, optic nerve lesions may develop as well.

Laboratory findings The principal method of diagnosis involves finding microfilariae in the skin. A small piece of superficial skin obtained by excision or punch biopsy is weighed and then incubated for several hours in saline or tissue culture media. Microfilariae that exit the skin sample are then counted in the fluid. Care should be taken not to contaminate the skin with blood that might harbor microfilariae of other species. Skin snip sites can include scapular and gluteal areas and, in Africa, leg areas. A count of more than 100 microfilariae/mg of skin indicates a heavy infection.

Another diagnostic method involves administering a 50 mg provocative dose of diethylcarbamazine; the subsequent onset of symptoms, which may include pruritus, rash, fever, and conjunctivitis, constitutes the Mazzotti reaction, which strongly suggests a diagnosis of onchocerciasis. Caution must be exercised in the use of this test because heavily infected patients may experience serious adverse reactions.

Eosinophilia is often prominent during onchocerciasis. Serology can be helpful when parasite demonstration is difficult.

Treatment Therapy for onchocerciasis improved dramatically with the introduction of ivermectin. Single-dose ivermectin therapy is free of most of the immediate cutaneous, ocular, and systemic reactions that complicated therapy with diethylcarbamazine, the agent previously used to treat onchocerciasis. Ivermectin, given orally in a single dose of 150 μg/kg, leads to symptomatic improvement and clearance of microfilariae from the

skin. This dose is repeated every 3 to 12 months for 3 to 4 years.[19] Albendazole combined with ivermectin only transiently halts production of microfilaria, but doxycycline (200 mg daily for 4 to 6 weeks) interrupts microfilaria production for 18 to 24 months.[54] Doxycycline may act by depleting adult filaria of *Wolbachia* endosymbiont bacteria, which correlates with an interruption of embryogenesis.

Loiasis

Loiasis, which is caused by the filarial nematode *Loa loa,* occurs in the rainforest areas of central and western Africa and is transmitted by *Chrysops* species of horseflies and deerflies. Adult worms reside in subcutaneous tissues (the life cycle of *Loa loa* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 3399). The microfilariae, which are sheathed, circulate in the bloodstream with a diurnal periodicity in which peak levels are reached at about noon.

Many residents of endemic areas who have loiasis have asymptomatic microfilaremia. The prominent clinical presentations are related to migrations of adult worms. In the subcutaneous tissues, migrations may produce recurrent lesions termed Calabar swellings, which are erythematous areas of swelling and edema up to 10 cm in diameter that resolve after 1 to 3 days. Infection may also present dramatically when the worm migrates subconjunctivally across the eye.

In contrast, patients who acquire loiasis after brief stays in central and West Africa exhibit different clinical and immunologic responses.[55,56] They are usually free of detectable microfilaremia but may experience more severe and pruritic episodes of angioedema and have brisk immunologic responses, including elevated antifilarial antibody titers, an elevated serum IgE level, and prominent eosinophilia—often greater than 3,000 eosinophils/μl.

Parasitologic diagnosis of loiasis is made by demonstrating adult worms in subconjunctival or subcutaneous tissues or demonstrating microfilariae in the blood by means of a Nuclepore filter concentration technique. In contrast to *Wucheria* microfilaremia, *Loa loa* microfilaremia is found throughout the day. Microfilaremia may not be detectable, however. Eosinophilia may be marked, with elevations of 50% or greater. Filarial antibody titers are usually elevated. In the absence of detectable microfilaremia, the diagnosis is suggested by clinical features, a history of exposure, and eosinophilia.

Therapy consists of diethylcarbamazine (6 mg/kg/day) orally for 3 weeks and may require adjunctive antihistamines or corticosteroids.[19] Patients with microfilaremia should be treated initially with escalating doses of diethylcarbamazine for 3 days (one 50 mg dose on the first day, three 50 mg doses on the second day, and three 100 mg doses on the third day), followed by the full 3-week course. Albendazole therapy is useful to reduce microfilaremia in patients who cannot tolerate diethylcarbamazine. Some patients require repeated courses of therapy.[57]

Side effects are usually mild and include frequent occurrences of pruritus and the development of subcutaneous nodules soon after initiation of therapy. Patients with severe microfilaremia (> 50,000 microfilariae/ml) have developed fatal encephalopathy after diethylcarbamazine or ivermectin therapy. Diethylcarbamazine, given in a dosage of 300 mg orally once a week, is an effective chemoprophylactic agent against loiasis for persons who are planning long-term visits to areas of Africa where this infection is endemic.[58]

Zoonotic Filarial Infections

Although no human filarial parasites are indigenous to the continental United States, a variety of filarial parasites infect animals. In rare cases, these organisms may be transmitted via insect vectors to humans. Within the human host, they may develop into adult worms, which localize in the same organs in humans as in the definitive animal hosts. In this way, dog heartworm (*Dirofilaria immitis*), endemic in dogs along the Atlantic and Gulf coasts, in the Mississippi Valley, and in California, localizes to the human pulmonary arteries. A granulomatous response develops around the worm, producing a pulmonary nodule. Some patients experience chest discomfort, malaise, low-grade fever, cough, and, occasionally, hemoptysis. Typically, however, the pulmonary lesions are detected as a coin lesion on a chest x-ray.[59] Prominent blood eosinophilia is absent, and serologic tests for filariasis are negative. In the absence of reliable diagnostic tests for human pulmonary dirofilariasis, excisional biopsy serves both diagnostic and therapeutic purposes.

Other zoonotic filarial parasites include *D. repens,* which causes subcutaneous abscesses, and *B. beaveri,* which produces focal lymphadenopathy. No chemotherapy is required for these infections or for infections caused by *D. immitis*. Generalized lymphedema has developed in one immunodeficient child with a zoonotic *Brugia* infection.[60]

Trematode Infections

SCHISTOSOMIASIS

Schistosomiasis, a chronic trematode (fluke) infection of humans, constitutes a major worldwide health problem: 200 million persons are infected, 120 million are symptomatic, and 10 million have severe disease.[61] Three major species—*Schistosoma mansoni, S. japonicum,* and *S. haematobium*—infect humans. *S. mansoni* is found in Africa, the Arabian Peninsula, South America, and parts of the Caribbean; *S. japonicum* is found in Japan, China, and the Philippines; and *S. haematobium* is found in Africa and the Middle East. Two minor species, *S. mekongi* and *S. intercalatum,* are found in mainland Indochina and central West Africa, respectively. Transmission of schistosomiasis cannot occur in the United States because of the absence of the specific freshwater snail that is a requisite intermediary host [*see Figure 11*]. However, the disease may be encountered in immigrants or travelers from endemic areas.[62,63]

Diagnosis

The diagnosis of schistosomiasis is suggested by a history of possible exposure, even exposure that occurred many years ago, along with compatible gastrointestinal or urinary tract symptoms, hepatosplenomegaly, eosinophilia, or a combination of these findings.

Clinical features Three stages of disease may occur in schistosomiasis. The first stage, schistosomal dermatitis, may develop acutely within a day of cercarial penetration of the skin. Because this entity develops early after exposure, it usually will have subsided in patients before they are seen by physicians in the continental United States. Swimmer's itch, a similar reaction caused by exposure to animal schistosomes in freshwater and saltwater, is seen in the United States [*see Figure 12*].[64] The schistosomes penetrate human skin and then die, causing no further infection.

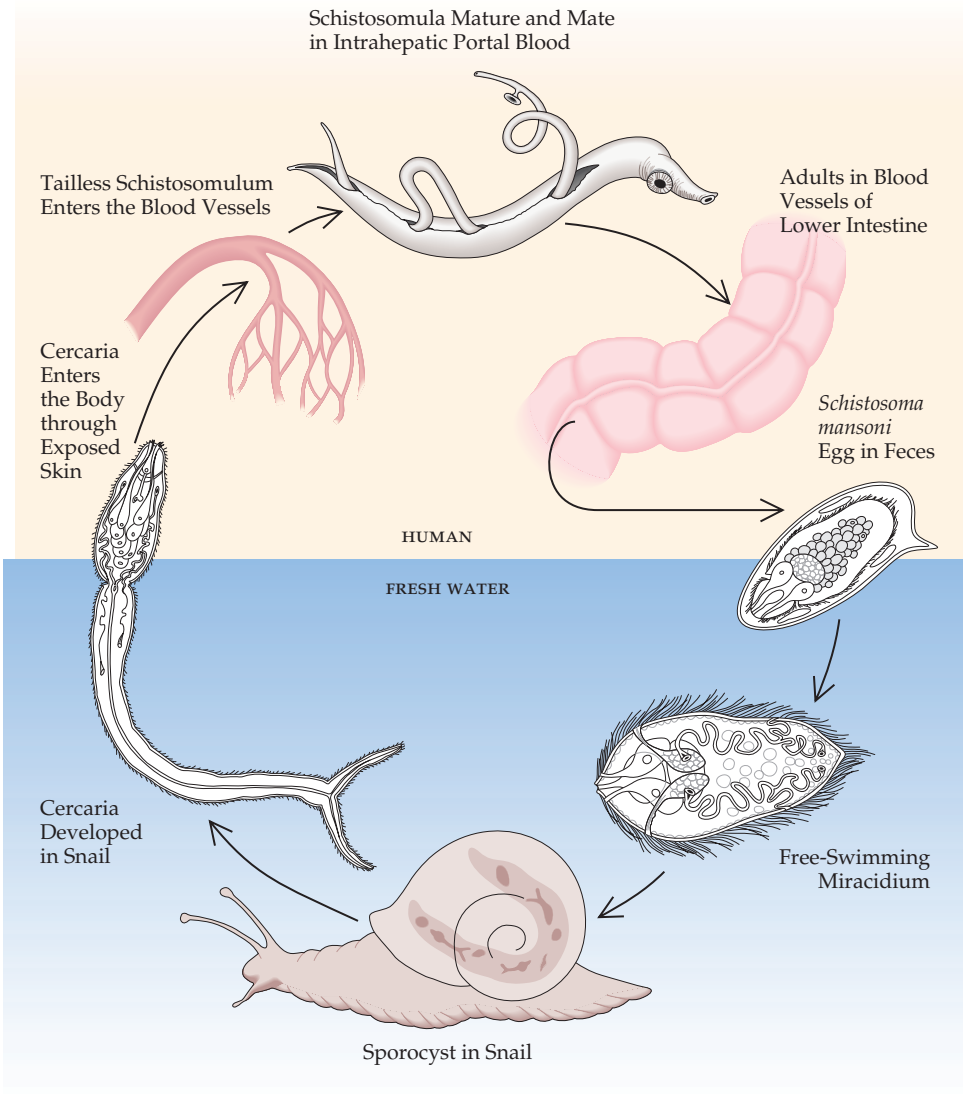


Figure 11 **Freshwater snails and the human body provide the living environment of the schistosomes *Schistosoma mansoni, S. japonicum, S. haematobium, S. mekongi*, and *S. intercalatum*. Eggs of each species hatch into schistosomal miracidia, which enter snails. Free-swimming cercariae released by the snails rapidly penetrate human skin and become tailless schistosomula, which enter the blood vessels. The schistosomula mature in the intrahepatic portal blood, form male-female pairs, and begin to mate. The pairs, still copulating, migrate to various sites: *S. mansoni, S. intercalatum*, and *S. mekongi* ultimately lodge in the blood vessels of the lower intestine; *S. japonicum* lodges in blood vessels throughout the intestine; and *S. haematobium* settles in blood vessels around the bladder. Adult worms subsequently begin laying eggs in the wall of the urinary bladder or intestine. From the intestinal sites, the eggs may be transported in the portal circulation to the liver, or they may be excreted in the feces. *S. hematobium*, by contrast, is excreted in the urine and does not tend to lead to liver disease but rather causes local bladder damage.**

The second stage of disease, acute schistosomiasis, or Katayama fever, develops 4 to 8 weeks after heavy (presumably, primary) infection. This stage is thought to be caused by a severe allergic response at the onset of egg-laying by the schistosomes. Patients have fever, cough (up to 10% also have pulmonary nodules), hepatosplenomegaly, malaise, myalgias, urticaria, and eosinophilia.[65-67] Deaths have ensued. Katayama fever is more severe in infection with *S. japonicum* than with other species, because of the high quantities of eggs produced by *S. japonicum.*

Chronic schistosomiasis, the third stage, is caused by the heavy deposition of eggs in the intestine or bladder and in the liver. In *S. haematobium* infection, the principal symptoms are hematuria, dysuria, and frequent urination. Hydronephrosis and pyelonephritis may develop as a result of fibrosis and infection. In *S. mansoni, S. mekongi,* or *S. japonicum* infection, manifestations may include fever, malaise, abdominal pain, diarrhea, blood in stools, and hepatosplenomegaly. Presinusoidal hepatic trapping of *S. mansoni, S. mekongi,* or *S. japonicum* eggs and the consequent granulomatous reaction induce portal hypertension and collateral esophageal varices. Eggs may then be shunted from the liver to the lung, with the possible sequela of pulmonary hypertension. Death may occur as a result of variceal bleeding. Hepatic encephalopathy rarely develops because the hepatic parenchyma is spared. Coinfection with *S. mansoni* and hepatitis B or C virus is associated with accelerated clinical deterioration.[61] Less common sequelae of chronic schistosomiasis include intestinal polyps, bladder carcinoma, and persistent *Salmonella* infections. An uncommon sequela of both acute and chronic schistosomiasis is focal neurologic dysfunction from aberrant localization of eggs in CNS tissue. Embolic deposition of *S. japonicum* eggs may produce cerebral granulomas, whereas *S. haematobium* and *S. mansoni* eggs may cause transverse myelitis involving the midthoracic or lumbar spinal cord.[68,69]

Laboratory findings and imaging studies Computed tomography, magnetic resonance imaging, or ultrasonography may detect hepatic periportal fibrosis and calcification, colonic wall calcifications, and changes in the bladder and ureter resulting from schistosomiasis. Serologic tests can help confirm the diagnosis; an indirect immunofluorescent test for gut-associated schistosome antigens is especially sensitive for the detection of acute schistosomiasis. Stool examination should include a search for eggs of all *Schistosoma* species [*see Figure 13*]. Urine specimens for detection of *S. haematobium* should be obtained between 10 A.M. and 2 P.M. If stool and urine specimens are negative, microscopic examination of biopsy specimens of rectal or bladder

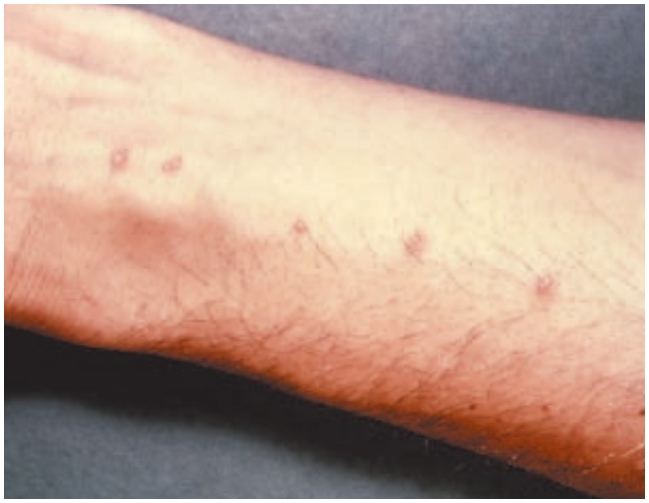


Figure 12 **Swimmer's itch is a cutaneous reaction caused by exposure to animal schistosomes in freshwater and saltwater. The schistosomes penetrate the skin and then die, causing no further infection.**

mucosa may demonstrate eggs of all species. Detection of circulating antigens from adult worms and eggs are promising techniques that may supersede traditional egg demonstration.

Treatment

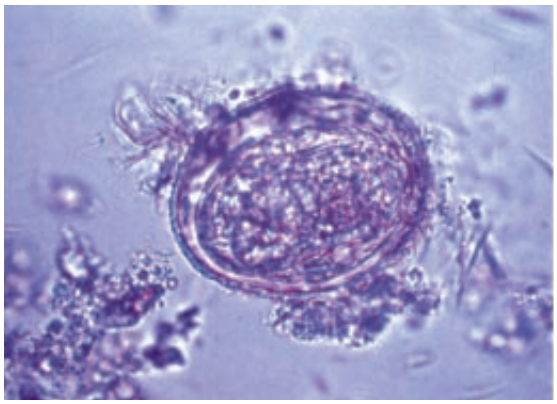
Praziquantel is used to treat infection caused by any of the five *Schistosoma* species; it reliably cures 60% to 90% of infected persons and reduces egg burden substantially in most others. For *S. haematobium* and *S. mansoni,* two oral doses of 20 mg/kg are given in 1 day. For *S. japonicum* and *S. mekongi,* the dosage is 20 mg/kg given orally three times in 1 day. The efficacy, the paucity of side effects, and the convenience of single-day therapy make praziquantel the drug of choice for all forms of schistosomiasis.[19] Resistance to praziquantel has been reported in *S. hematobium* and *S. mansoni* infections in Egypt and Kenya but has not yet become a widespread problem.[61] In addition to praziquantel, corticosteroids are beneficial for patients with spinal cord schistosomiasis[59] and acute schistosomiasis.

PARAGONIMIASIS

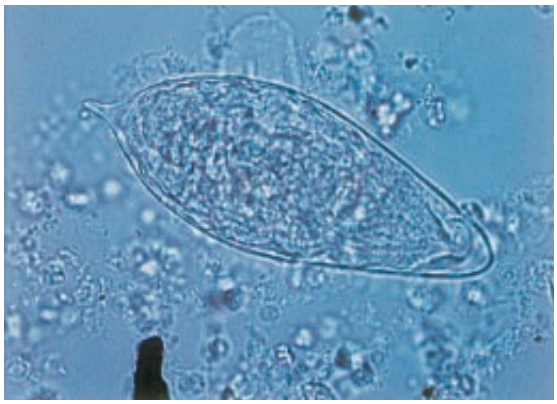
Humans acquire paragonimiasis after consumption of raw, salted, or wine-soaked crustacea (freshwater crabs or crayfish) infested with the metacercarial stage of lung flukes belonging to the genus *Paragonimus* (the life cycle of *Paragonimus* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 3415). It is estimated that 20 million people are infected with *Paragonimus* species.[70] *P. westermani* is endemic in parts of China, Korea, Japan, the Philippines, and Taiwan. Other *Paragonimus* species infect humans in western Africa and Central and South America. Although paragonimiasis is rare in the United States, it has developed in persons in Missouri from indigenous *Paragonimus* species.[71]

Ingested metacercariae undergo excystation in the duodenum and migrate through the wall of the gut into the peritoneal cavi-

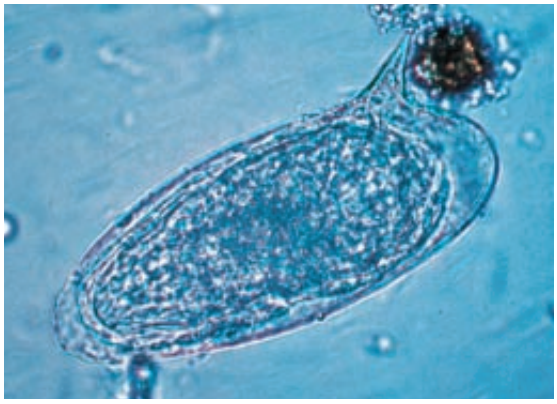
a



b



c



d

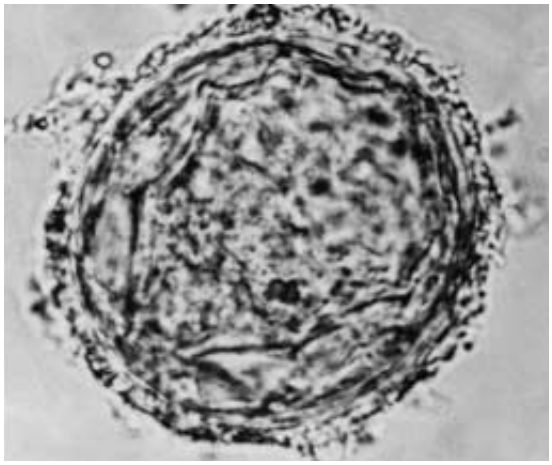


Figure 13 **Eggs of *S. japonicum* (*a*), *S. haematobium* (*b*), or *S. mansoni* (*c*) detected by microscopic examination of the stool, urine, or rectal mucosa biopsy specimens confirm the diagnosis of schistosomiasis. The magnification is about 400 times. (*d*) This *S. mekongi* egg was found in the stool of a Laotian refugee in the United States. The magnification is about 1,000 times.**

ty. Most pass through the diaphragm and penetrate the parenchyma of the lung. Neutrophilic and eosinophilic reactions, followed by a mononuclear leukocytic inflammatory reaction, develop around the fluke. As the lung parenchyma necrotizes, a fibrous capsule begins to surround the fluke. By about 5 to 6 weeks after ingestion, the flukes have matured and start laying eggs, which causes the capsule to enlarge and rupture, often into a bronchiole. The most common presentation of paragonimiasis is the production of brown-tinged sputum or hemoptysis, which derives from the admixture of eggs, inflammatory cells, and blood in the sputum. Sputum is often gelatinous and purulent as well as bloody. Patients usually appear well but may have a chronic cough, pleuritic pain, or night sweats.

Young flukes may migrate to nonpulmonary sites. Localization in the CNS produces signs and symptoms from a cerebral or spinal inflammatory mass lesion, which may calcify.[72] Less commonly, the parasites lodge in cutaneous or peritoneal sites.

During the early stages of infection, when the larvae migrate, blood eosinophilia is prominent. The chest x-ray may show transient, often basilar, infiltrations, as in Löffler syndrome. Later in the course of the disease, blood eosinophilia commonly disappears, and the chest x-ray may show areas of cavitation; ill-defined, so-called cotton-wool and streaky densities; and bubble-like cavities.[73] Pleural reaction, with or without pleural effusion or pneumothorax, can occur.

The diagnosis of paragonimiasis should be considered for patients from endemic areas with compatible clinical presentations.[74] Examination of the sputum for ova may confirm the diagnosis by revealing the operculated eggs of *Paragonimus*. Swallowed eggs appear in the stool; therefore, examining the stool for eggs may be helpful, especially in children. Fine-needle aspiration of pulmonary lesions also yields diagnostic eggs.[75] Paragonimiasis may resemble pulmonary tuberculosis both clinically and radiographically, yet *Paragonimus* eggs are usually not seen in acid-fast stains. Serologic tests are available.[70]

Praziquantel is an effective treatment of paragonimiasis, although its use remains investigational. The dosage is 25 mg/kg orally three times a day for 2 days.[19] Bithionol is somewhat effective, but its use is limited because of side effects such as diarrhea. Triclabendazole has been used and has been reported to cure cases in which praziquantel and bithionol had failed.[70]

CLONORCHIASIS

Clonorchis sinensis, the Chinese liver fluke, infects approximately seven million persons in the Far East, including South China, Hong Kong, Taiwan, Japan, Korea, and Vietnam.[76] Humans become infected with *C. sinensis,* a parasite of freshwater fish, by eating raw or undercooked fish containing encysted metacercariae (the life cycle of *C. sinensis* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 3385). Larvae, liberated by trypsin in the duodenum, migrate to the common bile duct and then to the distal biliary tree. There they mature into adult worms, which may persist for more than 20 years. Less commonly, adult worms are found in the gallbladder and pancreatic ducts.

In the acute phase of the infection, the epithelium of the biliary tree undergoes early desquamation, which is followed by hyperplasia and increased mucin production by the epithelial cells. The hyperplasia may progress to adenomatous changes. With chronic infection, fibrosis develops around dilated bile ducts. Clinically, the syndrome of acute clonorchiasis, manifested by fever, chills, and tender hepatomegaly, may develop 1 week after ingestion of infected fish. Later in the course of the infection, however, most patients with light infection and many with heavy infection are asymptomatic.[77] Occasionally, adult worms block pancreatic ducts, causing pancreatitis. By occluding the biliary tract, worms may contribute to acute suppurative cholangitis. Clonorchiasis predisposes to intrahepatic bile duct stone formation and to recurrent pyogenic cholangitis; infection with *C. sinensis* has been associated with cholangiocarcinoma.[78]

Leukocytosis, prominent eosinophilia, and elevation of alkaline phosphatase levels occur in symptomatic acute clonorchiasis. Sonography often detects diffuse dilatation of small intrahepatic bile ducts without dilatation of large intrahepatic or extrahepatic ducts. Adult flukes may be visualized in the gallbladder by ultrasonography and in the bile ducts by cholangiography.[79] In asymptomatic chronic clonorchiasis, eosinophilia is not present, and liver function tests and liver scans are normal. Egg laying begins within 2 to 3 weeks after infection; eggs may be found either in the stool or in duodenal aspirate. Diagnostic serologic tests are neither sensitive nor specific. The diagnosis should be considered in patients who have a history of travel or residence in the Far East, have eaten undercooked fish, and have a compatible clinical syndrome.

Praziquantel, which is considered investigational for clonorchiasis, constitutes effective and well-tolerated therapy for this disease. Praziquantel is administered at a dosage of 25 mg/kg orally three times a day for 1 day.[19] An alternative investigative therapy is albendazole, 10 mg/kg/day for 7 days.[19] Initially, complications of *C. sinensis* infection, including calculi, cholangitis, and pancreatitis, are managed medically, although surgical drainage may be required.

OPISTHORCHIASIS

Opisthorchiasis represents infection by either *Opisthorchis felineus* or *O. viverrini. O. felineus* infects 1.5 million persons in Kazakhstan, the Ukraine, central Europe, western Siberia, and parts of Asia; *O. viverrini* infects nine million persons in Thailand, Laos, and Cambodia.

Cats and wild carnivores are the definitive hosts of these species. Humans acquire infection by eating raw or undercooked fish that contains metacercariae of the parasite. Metacercariae undergo excystation in the duodenum and migrate into the bile ducts, where they mature.

Clinical features of opisthorchiasis are similar to those of clonorchiasis; complications include the development of cholangiocarcinomas.[77,78,80] The diagnosis is made by finding eggs in feces or in duodenal aspirate.

Therapy, which is considered investigational by the FDA, consists of praziquantel (25 mg/kg orally three times in 1 day).[19] Complications involving biliary tract sepsis require the administration of antibacterial agents.

FASCIOLIASIS

Fasciola hepatica, the liver fluke of sheep and cattle, is a major veterinary problem and can give rise to human fascioliasis.[78,81,82] Human infection with *F. hepatica* shows a wide geographic distribution that includes Europe, China, Africa, and Latin America; the disease is considered a public health problem in the Andean countries of South America, Iran, and western Europe.[82] Despite the prevalence of the parasite in sheep and cattle of the southern and western United States, autochthonous human cases are rare. Humans generally become infected by ingesting parasitic cysts attached to aquatic plants, most notably wild watercress.

Ingested metacercariae burrow through the intestinal wall into the peritoneal cavity and then penetrate the hepatic capsule and parenchyma; they then enter the bile ducts, where they mature into adults after 3 to 4 months. Acute fascioliasis runs its course during the months of penetration and maturation. Symptoms may be minimal or include fever, upper abdominal pain, hepatomegaly, and malaise. Pruritus, urticaria, jaundice, nonproductive coughing, and anemia occur less often.[78,83]

After the flukes mature in the biliary passages, hyperplasia and dilatation of the biliary ducts, as well as periductal fibrosis, develop. Clinical manifestations in the chronic stage of infection are variable and may include the same signs and symptoms experienced in the acute phase. Obstruction of the biliary tract, cholecystitis, and biliary cirrhosis are uncommon. In rare instances, flukes migrate to other tissues, including the lungs, muscles, and CNS.

Ingestion of raw sheep or goat liver containing young flukes produces halzoun, a disease recognized in the Near East. Lodging of the flukes in the pharynx produces a pharyngeal inflammatory mass lesion with attendant dysphagia and dyspnea.

Acute fascioliasis usually produces leukocytosis, marked eosinophilia, and cholestatic-type abnormalities on liver function testing. Eggs are not found until approximately 3 months after infection, when they may be detected in the stool or, with a higher yield, in biliary or duodenal fluid samples. Thus, the triad of fever, marked eosinophilia, and hepatomegaly suggests acute fascioliasis. A history of ingestion of potentially infected watercress supports the diagnosis. Serologic tests are available but may reflect cross-reactions with other helminthic parasites. Nodular hepatic lesions with diminished density may be visualized by CT or MRI.[84,85] In chronic fascioliasis, the extent of the abnormalities in liver function tests and cholangiograms correlates with the magnitude of biliary tract obstruction and hepatocellular damage.

Triclabendazole, given in a single oral dose of 10 mg/kg, is the drug of choice for fascioliasis, but in the United States it is available only from the manufacturer (Novartis). Bithionol (30 to 50 mg/kg on alternate days for 10 to 15 doses) and nitazoxamide (500 mg p.o., b.i.d. for 3 days) are alternatives.[19] Praziquantel is not effective for fascioliasis.

INTESTINAL FLUKES

Fasciolopsiasis

Fasciolopsiasis results from infection with the intestinal fluke *Fasciolopsis buski,* which is found in many parts of Asia.[86] This fluke principally parasitizes the intestine of the pig. Human infection is acquired by ingestion of water plants such as water chestnuts, which bear metacercariae of the parasite. The larvae undergo excystation in the duodenum and develop into large adult flukes, up to 7 cm long, that attach to the mucosa of the proximal small intestine. Inflammation and ulceration may occur at these intestinal sites. Light infection is asymptomatic; heavy infection is associated with abdominal pain, ulceration, hemorrhage, intestinal obstruction, malabsorption, and facial and generalized edema.[87] Blood eosinophilia is common. Diagnosis is made by finding adult flukes or, more commonly, by finding in feces the eggs of *F. buski,* which are difficult to distinguish from the eggs of *Fasciola hepatica*. Fasciolopsiasis is treated, on an investigational basis, with praziquantel, 25 mg/kg orally three times in 1 day,[10] although some sources advise that a single 15 mg/kg dose is effective.[86]

Other Intestinal Flukes

Infection with two small intestinal flukes, *Metagonimus yokogawai* (found in the Far East and Indonesia) and *Heterophyes heterophyes* (found in Tunisia, Egypt, and the Far East), occurs when humans ingest raw or undercooked fish that contains metacercariae of the parasites. The adult flukes are 2 to 3 cm long and attach to the mucosa of the small intestine. Heavy infection may cause abdominal pain and diarrhea.[88] Diagnosis is made by finding eggs, which resemble the eggs of *Clonorchis* species, in feces. Therapy consists of the investigational drug praziquantel (25 mg/kg orally three times in 1 day).[19]

Human infection with the intestinal fluke *Metorchis conjunctus,* acquired near Montreal, Canada, by consumption of the white sucker fish, has been described. Illness consisted of upper abdominal pain, low-grade fever, eosinophilia, and elevated liver enzyme levels. Diagnosis was made by finding eggs in the stool and by serology. Praziquantel (25 mg/kg three times in 1 day) is beneficial.[19,89]

The intestinal fluke *Nanophyetus salmonicola* has been recognized in humans who ate raw or kippered salmon. In most patients, symptoms and infection resolved without therapy, although treatment can be provided with praziquantel (20 mg/kg given three times in 1 day).[19,90]

Cestode Infections

Humans can harbor the adult form of fish, pork, and beef tapeworms (cestodes). These often do not cause disease but are noticed when segments of the worm are passed. Most often, these segments can be distinguished from other worms by their flat, tapelike appearance and their segmented proglottids. Examination of these proglottids or the head of the tapeworm can differentiate these three species. In general, larval forms of tapeworms, such as cysticercus (pork) and *Echinococcus* (dog), cause more severe disease by mass effect and inflammation.

FISH TAPEWORM

Humans acquire fish tapeworm infection by ingestion of inadequately cooked fish containing the infective plerocercoid stage of parasitic *Diphyllobothrium* species, including *D. latum* (the life cycle of *D. latum* is illustrated in the CDC PHIL [http://phil.cdc.gov/Phil]; photograph 5257).[91-93] The growing popularity of raw fish dishes, such as sushi, sashimi, seviche, and Dutch green herring, has increased the risk of acquiring diphyllobothriasis. Freshwater fish, including pike and yellow perch caught in the United States, may harbor *Diphyllobothrium* species, as may anadromous salmon. *Diphyllobothrium* species other than *D. latum* are found in Pacific salmon and Alaskan blackfish.[94] Adult tapeworms may grow up to 15 meters in length and live for 20 years or longer in the small intestine. Human infections are often asymptomatic, although some patients experience anorexia, nausea, or weight loss. Because *D. latum* competes with the host for vitamin B_{12}, megaloblastic anemia and neuropathy from vitamin B_{12} deficiency may develop. Diagnosis is made by finding the operculated eggs in the stool or by recovering proglottids in the stool after a saline purge. Therapy is with praziquantel (5 to 10 mg/kg, given once), which is investigational for this use.[19] An alternative is niclosamide, in a single oral dose of 2 g; however, this drug may be difficult to obtain in the United States and may be available only from compounding pharmacies.

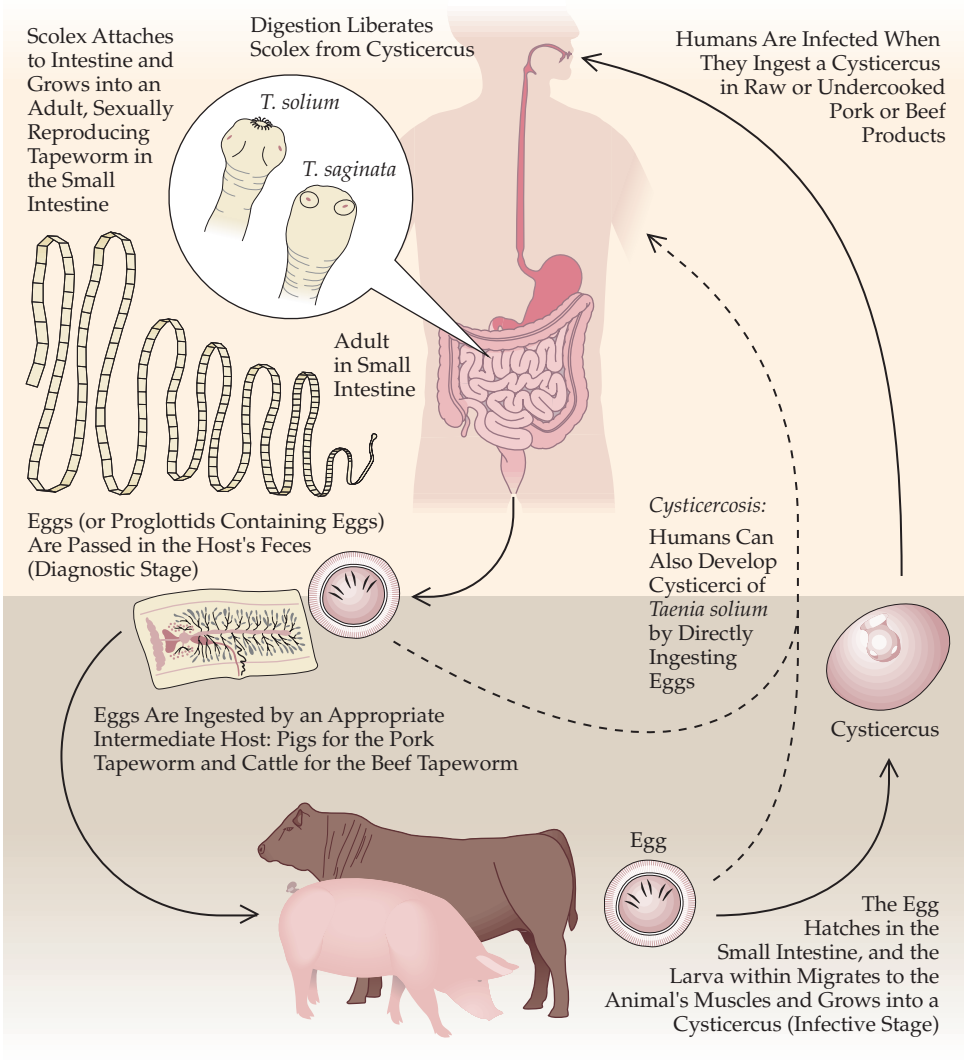


Figure 14 **Life cycle of *Taenia solium*, the pork tapeworm, which causes cysticercosis, and of *T. saginata*, the beef tapeworm.**

PORK TAPEWORM

Taenia solium, or pork tapeworm, causes two distinct types of disease, depending on the stage of the parasite that is ingested [*see Figure 14*].[95] If cysticerci in inadequately cooked pork are ingested, the adult tapeworm develops in the intestine, causing symptoms such as abdominal pain, weight loss, and weakness. Patients may describe passing proglottid segments of the worm. Eggs of *T. solium* are also passed in the feces. Eggs can be detected with greater frequency by applying clear cellulose acetate tape to the perianal skin and examining the tape, as described for pinworm [*see* Intestinal Nematode Infections, *above*]. *T. solium* eggs are indistinguishable from eggs of *T. saginata* (beef tapeworm); diagnostic differentiation between the two species requires recovery of mature proglottids or the head (scolex) from the stool [*see Figure 15*]. Therapy for adults infected with intestinal pork tapeworm consists of praziquantel (5 to 10 mg/kg, given once).[19] Some authorities suggest using niclosamide (2 g in a single oral dose) because it is not absorbed and will not cause neurologic symptoms in patients with asymptomatic neurocysticercosis cysts; however, this drug may be difficult to obtain in the United States and may be available only from compounding pharmacies.[95]

Cysticercosis, the second disease entity, is caused by the ingestion of *T. solium* eggs. In an uninfected person, such ingestion may occur by consumption of food contaminated by egg-containing feces from an infected person. In persons with an intestinal worm, ingestion may occur by autoinfection involving hand-to-mouth fecal carriage or by regurgitation of egg-laden proglottids into the duodenum or stomach. Most infections are encountered in developing countries, where intestinal *T. solium* infections occur frequently. Experience in the United States has

Figure 15 **Hooklets on the scolex (head) of *Taenia solium* give it an "armed" appearance.**

demonstrated, however, that cysticercosis may develop in those who have never traveled abroad, through transmission of infectious eggs from family members, domestic workers, or others infected with *T. solium*.[96-98]

The ingested eggs hatch in the stomach and upper intestine, and the resultant oncospheres circulate in the blood to various tissues. Cysticerci develop most often in subcutaneous tissue, skeletal muscle, and the brain, as well as in other organs, including the eyes, heart, liver, and lungs. Developing cysticerci elicit little host reaction, but as the cysticerci begin to degenerate, usually after several years, inflammation develops. Ultimately, the cysts, which range from 0.5 to about 2.0 cm in diameter, undergo necrosis and may become calcified.

Diagnosis

Clinical features Clinical signs and symptoms depend on the organ compromised by the cysticerci, the specific localization of a cysticercus within the organ, the state of inflammation surrounding the cysticerci, and the viability of the cestode. The most serious forms of cysticercosis are those with ocular, cardiac, and neurologic involvement. The manifestations of neurocysticercosis, which can be quite varied, depending on the site of lesions and the evolution of inflammatory reactions, include seizure disorders and focal deficits, hydrocephalus, arachnoiditis, and intracranial hypertension.[95] In endemic regions, cysticercosis is the most common cause of late-onset seizure disorders.[99]

Laboratory findings and imaging studies The use of CT and MRI has facilitated the recognition of CNS cysticercosis,[94,100,101] has helped delineate the various forms of neurocysticercosis,[83] and has provided a means for assessing the adequacy of therapy.[100,102] CNS imaging studies may reveal cysticerci in the brain parenchyma, within the ventricles, at the surface or the base of the brain, or within the subarachnoid space.

CT may miss early lesions, which have the same x-ray density as the brain, but hypodensity and contrast-enhancing ring lesions are seen with the development of inflammation around the cyst. As sclerosis occurs, a cystic lesion develops, and the cyst wall may calcify. Ultimately, degenerated cysts are replaced by small (1 to 4 mm in diameter) calcified lesions within the brain [*see Figure 16*].

The diagnosis of cysticercosis can be made with certainty only by biopsy of a cyst. Calcified cysts in subcutaneous tissue and muscle, which have a puffed-rice appearance on radiographs, should be sought. Although the signs and symptoms of neurocysticercosis are not specific, this diagnosis is supported by the finding of characteristic multiple cystic or calcified lesions on CT scans in a patient from an endemic area.

A serologic test, which should be used to test serum and CSF for antibody to *T. solium,* is available through the CDC. ELISA testing for antibody, however, may be negative in about 20% of patients with cysticercosis[103] and may be falsely positive in those with echinococcosis. An enzyme-linked immunoelectrotransfer blot assay for antibody is highly sensitive in patients who have several enhancing intracranial lesions; it is less sensitive in those who have only one lesion or calcified lesions.[101,104] Stool examination for *Taenia* eggs may detect concurrent infection with the tapeworm but is not directly pertinent to the diagnosis of cysticercosis.

Treatment

Therapy for cysticercosis may be medical or surgical. Patients with only calcified soft tissue or CNS lesions do not require medical therapy. Surgical excision was once the only approach for viable cysts, but praziquantel and albendazole have proved to be effective against neurocysticercosis.[100,102,105] Despite the improvements noted after medical therapy, the absence of controlled trials specifically comparing medically treated patients with untreated patients has left room for uncertainty concerning the efficacy of medical therapy for neurocysticercosis.[102,106] However, a randomized trial has demonstrated a trend toward fewer seizures in patients treated with albendazole and steroids, as compared with those treated with placebo.[107]

Albendazole is given in a dosage of 400 mg orally twice daily for 10 to 28 days; praziquantel is given in a dosage of 50 to 100 mg/kg/day in three divided doses for 30 days.[19] Albendazole appears to be slightly more effective than praziquantel at killing cysticercosis cysts.[102] Because treatment may cause inflammatory reactions to develop around cysticerci, ocular cysticercosis and spinal cysticercosis are not usually treated medically; an ophthalmologic examination is indicated before drug therapy to rule out intraocular cysticercosis, which could lead to devastating inflammation. For patients with neurocysticercosis, corticosteroids (e.g., dexamethasone, 4 to 16 mg/day, or prednisone, 60 to 100 mg/day) are usually given 1 to 2 days before and during treatment with albendazole or praziquantel to minimize inflammatory reactions. Patients who are taking anticonvulsant medications for neurocysticercosis should continue to use them during this treatment, but many such patients can stop taking antiseizure medications after cysticercosis therapy.[107,108] CT scanning should be repeated 3 to 6 months after therapy, to determine whether any of the cysts are still viable; therapy should be repeated if viable cysts remain.

BEEF TAPEWORM

T. saginata, or beef tapeworm, causes an intestinal infection in persons who have eaten undercooked beef containing cysticerci.[93] The infection is usually asymptomatic, although abdominal pain, weight loss, increased appetite, or passage of lengths of proglottids may be noticed because proglottids of *T. saginata* tend to be more motile than those of the other tapeworms [*see Figure 17*]. As noted, the detection of eggs in the feces or on the perianal skin is diagnostic of tapeworm infection, but differentiation between *Taenia* species requires identification of the proglottids. Therapy consists of praziquantel or niclosamide, as given for *T. solium* intestinal infection.[19] Infection with *T. saginata* does not lead to cysticercosis in humans.

DWARF TAPEWORM

Hymenolepis nana, a tapeworm measuring 3 to 4 cm long and 1 mm wide, has a broad geographic distribution.[91] In the United States, it is most commonly encountered in persons living in the southern states, in institutionalized patients, and in children. Unlike the other tapeworms, the larval and adult stages of *H. nana* develop in the same host. Infection is spread by the fecal-oral route and occurs when eggs, which are immediately infectious, are ingested. Thus, *H. nana* is one of the few helminths (along with *Enterobius, Capillaria,* and *Strongyloides*) that can propagate by autoinfection; children, in particular, can develop huge worm burdens and become symptomatic.

After an egg of *H. nana* has been ingested, an oncosphere hatches from the egg and penetrates the intestinal villi, where it develops into a cerocyst; the cerocyst reenters the lumen of the small intestine and develops into an adult worm. Eggs are liberated from the distal segments of the adult worm, which lives for

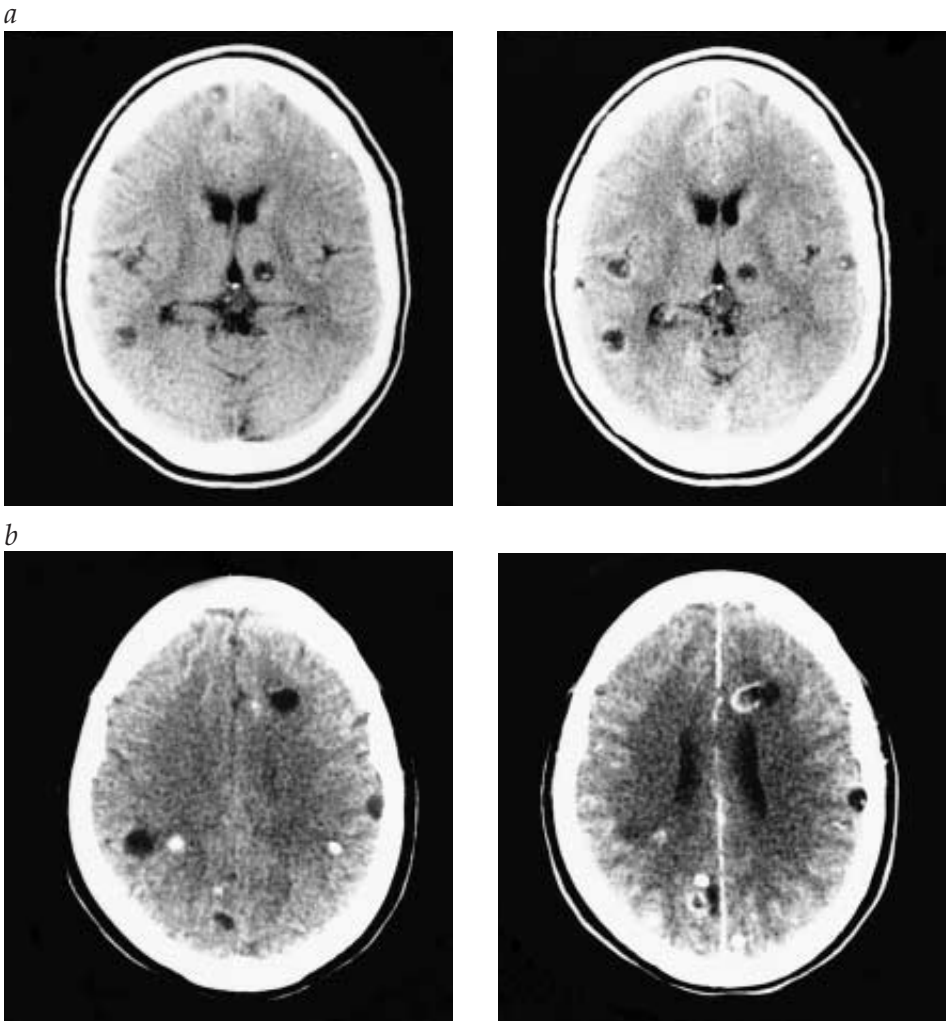


Figure 16 **Shown are CT scans of two patients (*a, b*) with seizures resulting from neurocysticercosis. The CT scans on the left are without contrast; those on the right are with contrast. Multiple cystic and calcified lesions are evident in both patients (the contrast agent enhances the cystic lesions). The scans from the first patient (*a*) show multiple cystic lesions, with a denser central spot that represents the scolex, a pathognomonic finding for neurocysticercosis.**

about 1 year. The eggs may cause internal reinfection. Light infection is usually asymptomatic; diarrhea and abdominal pain may accompany heavy infection. Diagnosis is made by finding eggs in feces. Therapy consists of praziquantel (25 mg/kg given once); this use of praziquantel is currently considered investigational. An alternative investigational therapy is nitazoxanide (500 mg p.o. daily for 3 days).

OTHER TAPEWORMS

Human infection with *H. diminuta,* a tapeworm of mice and rats, occasionally occurs when humans ingest insects that harbor developing cerocysts of the tapeworm; one such insect is the flea. Infection is more common in children and is associated with few or no symptoms. Diagnosis and treatment are the same as described for the dwarf tapeworm *H. nana* (see above).

Children are also more commonly infected with the dog tapeworm *Dipylidium caninum.* Infection is acquired by consuming infected fleas or lice. Adult worms develop in the small intestine and measure 15 to 70 cm in length. Mild intestinal symptoms may or may not be present. Diagnosis is made by finding

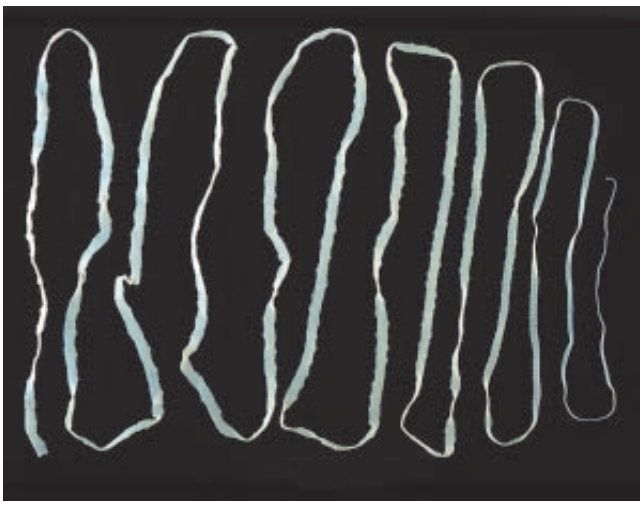


Figure 17 **Persons infected with the beef tapeworm, *Taenia saginata,* may pass lengths of proglottid in the stool. An entire adult beef tapeworm is shown.**

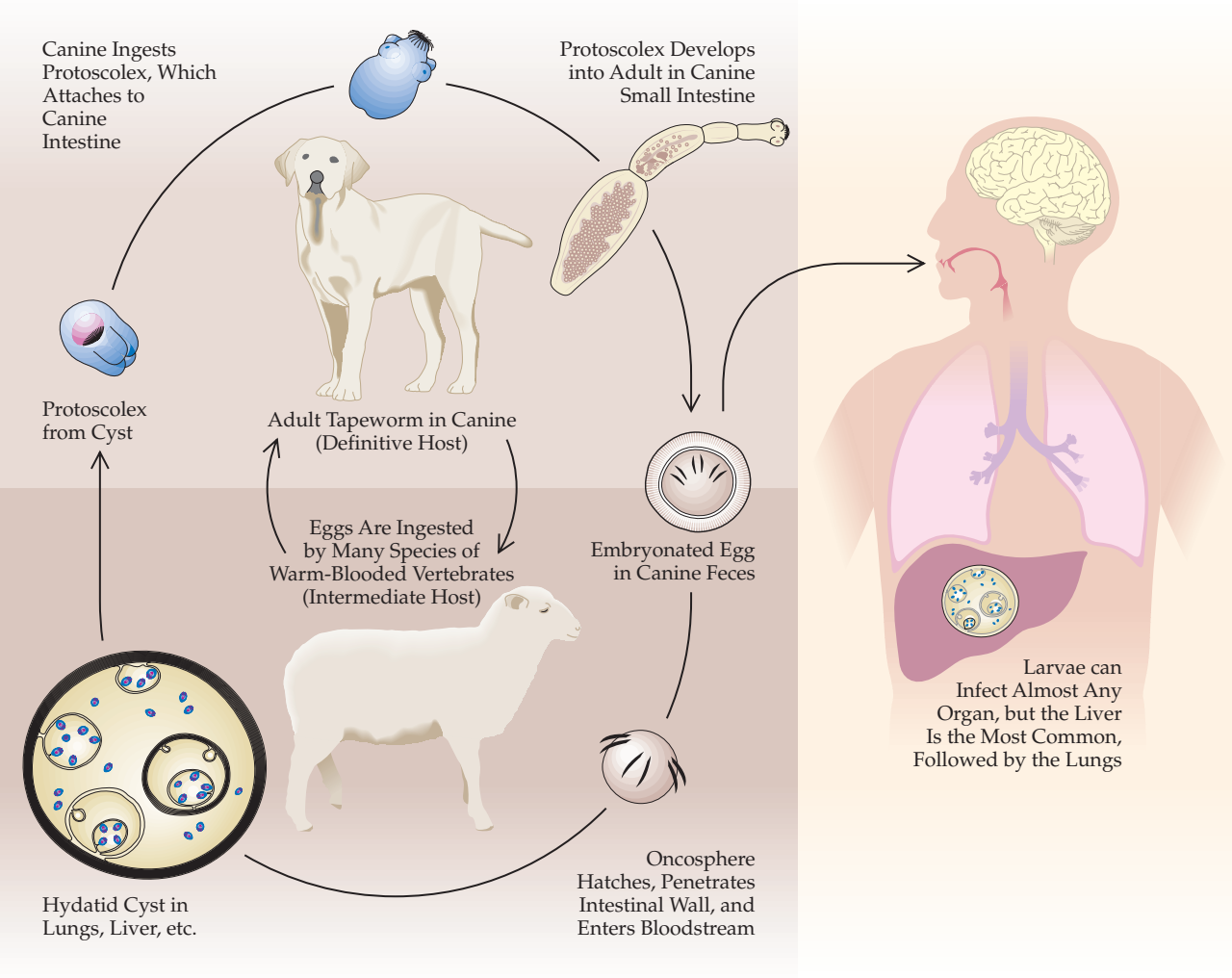


Figure 18 **Life cycle of *Echinococcus granulosus*. Adult *E. granulosus* cestodes live in the intestine of dogs and wolves. Infectious eggs are passed in the feces of these animals and may be ingested by intermediate hosts such as sheep, cattle, or humans. After the eggs are ingested, oncospheres are carried in the bloodstream to the liver, lungs, and other organs, where they form hydatid cysts. The cycle is maintained when dogs or wolves ingest the carcasses of intermediate hosts.**

proglottids or eggs in feces. Therapy for adults consists of praziquantel (5 to 10 mg/kg given once).[19]

Human infection with dog tapeworms of the genus *Multiceps* results in a syndrome termed coenurosis, which is similar to cysticercosis.[109] At present, surgical excision forms the basis of diagnosis and treatment.

Sparganosis represents infection by larval tapeworms of the genus *Spirometra,* which are closely related to tapeworms of the genus *Diphyllobothrium*. Most such infections occur in the Far East. Human infection results from drinking water containing microcrustacean *Cyclops* species that harbor procercoid larvae (spargana) of the parasite. Human infection may also be acquired by ingesting raw flesh of amphibians or snakes that contains larvae of the parasite or from applying such flesh to the skin as a poultice. After ingestion, the procercoid larvae migrate into subcutaneous tissues, where they usually present as a painless subcutaneous nodule that enlarges during the course of many months. Larvae may also migrate to the CNS.[110,111] Blood eosinophilia is commonly elicited. Surgical excision of larvae-containing nodules remains the basis of diagnosis and treatment.

HYDATID DISEASE

Echinococcosis in humans may result from infection with *Echinococcus granulosus,* which causes cystic hydatid disease; *E. multilocularis,* which causes alveolar hydatid disease; or *E. vogeli* or *E. oligarthus,* which causes polycystic hydatid disease and is found in areas of Central and South America.[93,112,113] Adult *E. granulosus* cestodes live in the intestine of dogs and wolves. Infectious eggs are passed in the feces of these animals and may be ingested by intermediate hosts such as sheep, cattle, or humans [*see Figure 18*]. The cycle is maintained when dogs or wolves ingest the carcasses of intermediate hosts. *E. granulosus* infection is most prevalent in sheep- and cattle-raising countries. It is also found in a number of western states, Alaska, and Canada, where autochthonous cases of human infection have been recorded.[114]

After the eggs are ingested, oncospheres are carried in the bloodstream to the liver, lungs, and other organs. Unilocular cysts, which may contain daughter cysts, develop most commonly in the liver; the second most common site is the lungs. In children, pulmonary involvement may be more common than hepatic involvement. Unilocular hydatid cysts enlarge concen-

trically, increasing in diameter by about 1 to 5 cm a year, depending on the density of the organs in which they are located. A cyst may attain a large size before the initial symptoms develop; these symptoms are usually attributable to a space-occupying mass lesion. The onset of symptoms has been reported to occur from before 1 year of age to 75 years of age, but in large series, most persons become symptomatic between 4 and 15 years of age. Pathologic fractures or neurologic symptoms can occur with osseous or CNS localization, respectively. When cysts leak, patients may experience bronchospasm, urticaria, or anaphylaxis; blood eosinophilia, which is otherwise usually not prominent, may increase. Communicating rupture of pulmonary or hepatic cysts can lead to the release of cyst contents into the bronchial or biliary systems. Because cysts contain multiple infective protoscolices, rupture of cysts can lead to dissemination of infection and the generation of new cysts from each released protoscolex.

E. multilocularis lives in the intestine of foxes and dogs. Its intermediate hosts are mice and other small mammals. *E. multilocularis* is found only in the Northern Hemisphere, including central western Europe, Russia, the central Asian republics, China, northern Japan, Canada, Alaska, and the north central United States.[113] Because the cysts of *E. multilocularis* lack a containing capsule, they progressively invade involved tissues and produce honeycombed alveolar hydatid cysts. The liver is most commonly affected. Severe damage caused by extensive alveolar hydatid cysts can result in jaundice and portal hypertension. The mortality in untreated alveolar disease is 90%.[115]

The diagnosis of hydatid disease can be strongly suggested by the results of radiographic studies.[113,116] Plain films detect pulmonary cysts[117] but often do not visualize cysts in other organs unless they are calcified—a process that occurs mostly in hepatic cysts. On CT and MRI, echinococcal cysts appear as well-defined, thick- or thin-walled cysts that may have calcified rims. [116-118] In older lesions, where scolices and daughter cysts form hydatid sand that settles in the dependent portion of the cyst, a layer of fluid can be visualized. Dependent movement of calcified hydatid sand, on repositioning and ultrasound monitoring, is strongly suggestive, if not pathognomonic, of a hydatid cyst. A pathognomonic CT finding in intact cysts is the presence of daughter cysts that are either free within the cyst or adherent to the inner germinal layer. Separation and collapse of cyst wall layers and the introduction of air into the space between the layers can be detected on plain films (as the meniscus, double arch, or water lily signs) and by CT scan.

Serologic tests can be helpful in making the diagnosis of echinococcosis but are not uniformly sensitive or specific.[92,119] A sensitive assay, such as ELISA or indirect hemagglutination, is performed first. Because of the possibility of false positive results produced by cross-reacting helminthic infections, specificity is confirmed with a less sensitive but more specific assay, such as an antigen-specific immunoblot or a gel diffusion assay for the *Echinococcus*-specific arc 5 immunoprecipitin band. Even with these assays, 5% to 25% of patients with neurocysticercosis have false positive results. Conversely, negative tests do not exclude the diagnosis, because approximately 50% of patients with isolated pulmonary cysts and 10% to 15% of those with hepatic cysts lack detectable antibodies against *Echinococcus*. Although leakage of cyst fluid poses risks of anaphylaxis or dissemination of infection, percutaneous aspiration of a cyst in a seronegative patient, with guidance provided by CT or ultrasonography, can yield diagnostic protoscolices or hydatid membranes.[120]

Therapeutic approaches to cystic and alveolar hydatid disease are quite complex, and consultation for expert advice is highly recommended.[113,115] For cystic hydatid disease, treatment should be reserved for symptomatic lesions or those affecting vital anatomic structures, because 75% of asymptomatic persons with cysts remain symptom-free for more than a decade.[113] When therapy is needed, the options are surgery, PAIR (puncture, aspiration, injection, and reaspiration; see below), or chemotherapy. Surgery offers the potential for total parasite removal and complete cure; nevertheless, cysts recur in 2% to 25% of patients treated with surgery.[113] Preoperative and postoperative albendazole therapy is highly advisable. The past practice of injecting protoscolicidal solutions (7% to 90% ethanol, 0.5% percent cetrimide, or 15% to 20% hypertonic saline) before resection is no longer recommended, because these agents are of uncertain efficacy and potentially dangerous to the patient because of chemical cholangitis and other complications.

The PAIR procedure, which consists of percutaneous aspiration, injection of protoscolicidal agents (e.g., 95% ethanol or 0.5% cetrimide), and reaspiration, has demonstrated efficacy as an alternative to surgery. A randomized, controlled trial showed PAIR to be as effective as surgery for hepatic cystic echinococcosis; moreover, patients treated with PAIR had lower postprocedure morbidity and a shorter hospital stay.[121-123] Like surgery, PAIR should be performed with albendazole therapy before and after the procedure to minimize dissemination of any leaked fluids containing infectious protoscolices.[124] PAIR is best used for liver cysts of 5 cm or greater that are anechoic, multiseptate, or multiple.[113] The PAIR procedure should not be performed if hepatic cysts communicate with the biliary tract. Complications are more common when PAIR is used to treat pulmonary lesions. PAIR may provoke acute allergic reactions.[113]

If neither surgery nor PAIR therapy is an option, medical therapy with albendazole is administered (mebendazole is an alternative that is considered less effective). Albendazole is given as a 400 mg dose twice a day for 28 days, often with additional 28-day courses given in subsequent months, with each course separated by 14-day treatment-free periods.[125] Most commonly, albendazole therapy is continued for 6 months. Patients need to be monitored for complications during long-term albendazole therapy, because hepatotoxicity, nausea, and neutropenia are common. Neither mebendazole nor albendazole is completely effective: 20% to 40% of patients experience no improvement. Albendazole therapy results in a cure in about one third of patients and leads to regression in cyst size and improvement in symptoms in another third. Both agents are contraindicated during pregnancy. Serial titers of echinococcal antibodies can be useful in monitoring therapeutic success.

Because alveolar hydatid disease is an aggressive and often fatal illness, the preferred therapy is radical resection of cysts, followed by albendazole therapy (up to 20 mg/kg/day) for at least 2 years.[113] Long-term chemotherapy can benefit patients who are not candidates for operation, as well as those patients who have undergone nonradical resections or liver transplantation.[113,126,127]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Mooraki A, Rahbar MH, Bastani B: Spontaneous systemic anaphylaxis as an unusual presentation of hydatid cyst: report of two cases. Am J Trop Med Hyg 55:302, 1996

2. Weller PF: Eosinophilia in travelers. Med Clin North Am 76:1413, 1992

3. Stephenson LS: Helminth parasites, a major factor in malnutrition. World Health Forum 15:169, 1994

4. Weller PF: Parasitic pneumonias. Respiratory Infections: Diagnosis and Management, 3rd ed. Pennington JE, Ed. Raven Press, New York, 1994, p 695

5. de Silva NR, Guyatt HL, Bundy DA: Morbidity and mortality due to *Ascaris*-induced intestinal obstruction. Trans R Soc Trop Med Hyg 91:31, 1997

6. Villamizar E, Mendez M, Bonilla E, et al: *Ascaris lumbricoides* infestation as a cause of intestinal obstruction in children: experience with 87 cases. J Pediatr Surg 31:201, 1996

7. Khuroo MS: Ascariasis. Gastroenterol Clin North Am 25:553, 1996

8. Turner JD, Faulkner H, Kamgno J, et al: Th2 cytokines are associated with reduced worm burdens in a human intestinal helminth infection. J Infect Dis 188:1768, 2003

9. Stephenson LS, Holland CV, Cooper ES: The public health significance of *Trichuris trichuria*. Parasitology 121:S73, 2000

10. Mahmood T, Mansoor N, Quraishy S, et al: Ultrasonographic appearance of *Ascaris lumbricoides* in the small bowel. J Ultrasound Med 20:269, 2001

11. Khuroo MS, Zargar SA, Yattoo GN, et al: Sonographic findings in gallbladder ascariasis. J Clin Ultrasound 20:587, 1992

12. Reddy DN, Sriram PV, Rao GV: Endoscopic diagnosis and management of tropical parasitic infections. Gastrointest Endosc Clin North Am 13:765, 2003

13. Loukas A, Bethony JM, Williamson AL, et al: Vaccination of dogs with a recombinant cysteine protease from the intestine of canine hookworms diminishes the fecundity and growth of worms. J Inf Dis 189:1952, 2004

14. Grencis RK, Cooper ES: *Enterobius, Trichuris*, *Capillaria*, and hookworm, including *Ancylostoma caninum*. Gastroenterol Clin North Am 25:579, 1996

15. Liu LX, Chi J, Upton MP, et al: Eosinophilic colitis associated with larvae of the pinworm *Enterobius vermicularis*. Lancet 346:410, 1995

16. Arca MJ, Gates RL, Groner JI, et al: Clinical manifestations of appendiceal pinworms in children: an institutional experience and a review of the literature. Pedatr Surg Int 20:372, 2004

17. Kitchen LW, Tu KK, Kerns FT: *Strongyloides*-infected patients at Charlestown Area Medical Center, West Virginia, 1997–1998. Clin Infect Dis 31:E5, 2000

18. Keiser PB, Nutman TB: *Strongyloides stercoralis* in the immunocompromised population. Clin Microbiol Rev 17:208, 2004

19. Drugs for parasitic infections. Med Lett Drugs Ther , in press

20. Bahgat MA, El Gindy AE, Mahmoud LA, et al: Evaluation of the role of *Ancylostoma caninum* in humans as a cause of acute and recurrent abdominal pain. J Egypt Soc Parasitol 29:873, 1999

21. Croese J, Loukas A, Opdebeeck J, et al: Human enteric infection with canine hookworms. Ann Intern Med 120:369, 1994

22. Murrell KD, Bruschi F: Clinical trichinellosis. Prog Clin Parasitol 4:117, 1994

23. Capo V, Despommier DD: Clinical aspects of infection with *Trichinella* species. Clin Microbiol Rev 9:47, 1996

24. Roy SL, Lopez AS, Schantz PM: Trichinellosis surveillance—United States, 1997–2001. MMWR Surveill Summ 52:1, 2003

25. Dworkin MS, Gamble HR, Zarlenga DS, et al: Outbreak of trichinellosis associated with eating cougar jerky. J Infect Dis 174:663, 1996

26. MacLean JD, Poirier L, Gyorkos TW, et al: Epidemiologic and serologic definition of primary and secondary trichinosis in the Arctic. J Infect Dis 165:908, 1992

27. Andrews JR, Ainsworth R, Abernethy D: *Trichinella pseudospiralis* in humans: description of a case and its treatment. Trans R Soc Trop Med Hyg 88:200, 1994

28. Mawhorter SD, Kazura JW: Trichinosis of the central nervous system. Semin Neurol 13:148, 1993

29. Schellenberg RS, Tan BJ, Irvine JD, et al: An outbreak of trichinellosis due to consumption of bear meat infected with *Trichinella nativa*, in 2 northern Saskatchewan communities. J Infect Dis 188:835, 2003

30. Audicana MT, Ansotegui IJ, de Corres LF, et al: *Anisakis simplex*: dangerous—dead or alive? Trends Parasitol 18:20, 2002

31. Deardoff TL, Kent ML: Prevalence of larval *Anisakis simplex* in pen-reared and wild-caught salmon (*Salmonidae*) from Puget Sound, Washington. J Wildl Dis 25:416, 1989

32. Kakizoe S, Kakizoe H, Kakizoe K, et al: Endoscopic findings and clinical manifestations of gastric anisakiasis. Am J Gastroenterol 90:761, 1995

33. Sugimachi K, Inokuchi K, Ooiwa T, et al: Acute gastric anisakiasis: analysis of 178 cases. JAMA 53:1012, 1985

34. Wittner M, Turner JW, Jacquette G, et al: Eustrongylidiasis: a parasitic infection acquired by eating sushi. N Engl J Med 320:1124, 1989

35. Shirahama M, Koga T, Ishibashi H, et al: Intestinal anisakiasis: US in diagnosis. Radiology 185:789, 1992

36. Dziekonska-Rynko J, Rokicki J, Jablonski Z: Effects of ivermectin and albendazole against *Anisakis* simplex in vitro and in guinea pigs. J Parasitol 88:395, 2002

37. Blackwell V, Vega-Lopez F: Cutaneous larva migrans: clinical features and management of 44 cases presenting in the returning traveler. Br J Dermatol 145:434, 2001

38. Jelinek T, Maiwald H, Nothdurft HD, et al: Cutaneous larva migrans in travelers: synopsis of histories, symptoms, and treatment of 98 patients. Clin Infect Dis 19:1062, 1994

39. Weller PF, Liu LX: Eosinophilic meningitis. Semin Neurol 13:161, 1993

40. Glickman LT, Magnaval JF: Zoonotic roundworm infections. Infect Dis Clin North Am 7:717, 1993

41. Despommier D: Toxocariasis: clinical aspects, epidemiology, medical ecology, and molecular aspects. Clin Microbiol Rev 16:265, 2003

42. Slom TJ, Cortese MM, Gerber JI, et al: An outbreak of eosinophilic meningitis caused by *Angiostrongylus cantonensis* in travelers returning from the Caribbean. N Engl J Med 346:668, 2002

43. Chomongkol V, Sawanyawisuth K, Thavornpitak Y: Corticosteroid treatment of eosinophilic meningitis. Clin Infect Dis 31:660, 2000

44. Kliks MM, Palumbo NE: Eosinophilic meningitis beyond the Pacific Basin: the global dispersal of a peridomestic zoonosis caused by *Angiostrongylus cantonensis*, the nematode lungworm of rats. Soc Sci Med 34:199, 1992

45. Nosanchuk JS, Wade SE, Landolf M: Case report of and description of parasite in *Mammomonogamus laryngeus* (human syngamosis) infection. J Clin Microbiol 33:998, 1995

46. Rusnak JM, Lucey DR: Clinical gnathostomiasis: case report and review of the English-language literature. Clin Infect Dis 16:33, 1993

47. Moore DA, McCroddan J, Dekumyoy P, et al: Gnathostomiasis: an emerging imported disease. Emerg Infect Dis 9:647, 2003

48. Progress toward global eradication of dracunculiasis, January–June 2003. MMWR Morb Mortal Wkly Rep 52:881, 2003

49. Rohde J, Sharma BL, Patton H, et al: Surgical extraction of guinea worm: disability reduction and contribution to disease control. Am J Trop Med Hyg 48:71, 1993

50. Melrose WD: Lymphatic filariasis: new insights into an old disease. Int J Parasitol 32:947, 2002

51. Ong RK, Doyle RL: Tropical pulmonary eosinophilia. Chest 113:1673, 1998

52. Simonsen P, Dunyo S: Comparative evaluation of three new tools for diagnosis of Bancroftian filariasis based on detection of specific antigens. Trans R Soc Trop Med Hyg 93:278, 1999

53. Addiss D, Critchley J, Ejere H, et al: Albendazole for lymphatic filariasis. Cochrane Database Syst Rev (1):CD003753, 2004

54. Hoerauf A, Buttner DW, Adjei O, et al: Onchocerciasis: science, medicine, and the future. BMJ 326:207, 2003

55. Klion AD, Massougbodji A, Sadeler BC, et al: Loiasis in endemic and nonendemic populations: immunologically mediated differences in clinical presentation. J Infect Dis 163:1318, 1991

56. Churchill DR, Morris C, Fakoya A, et al: Clinical and laboratory features of patients with loiasis (*Loa loa* filariasis) in the U.K. J Infect 33:103, 1996

57. Klion AD, Ottesen EA, Nutman TB: Effectiveness of diethylcarbamazine in treating loiasis acquired by expatriate visitors to endemic regions: long-term follow-up. J Infect Dis 169:604, 1994

58. Nutman TB, Miller KD, Mulligan M, et al: Diethylcarbamazine prophylaxis for human loiasis: results of a double-blind study. N Engl J Med 319:752, 1988

59. Shah MK: Human pulmonary dirofilariasis: review of the literature. South Med J 92:276, 1999

60. Orihel TC, Beaver PC: Zoonotic *Brugia* infections in North and South America. Am J Trop Med Hyg 40:638, 1989

61. Ross AG, Bartley PB, Sleigh AC, et al: Schistosomiasis. N Engl J Med 346:1212, 2002

62. Cetron MS, Chitsulo L, Sullivan JJ, et al: Schistosomiasis in Lake Malawi. Lancet 348:1274, 1996

63. Grobusch MP, Muhlberger N, Jelinek T, et al: Imported schistosomiasis in Europe: sentinel surveillance data from TropNetEurop. J Travel Med 10:164, 2003

64. Gonzalez E: Schistosomiasis, cercarial dermatitis, and marine dermatitis. Dermatol Clin 7:291, 1989

65. Rocha MO, Pedroso ER, Greco DB, et al: Pathogenetic factors of acute schistosomiasis *mansoni*: correlation of worm burden, IgE, blood eosinophilia and intensity of clinical manifestations. Trop Med Int Health 1:213, 1996

66. Schwartz E, Rozenman J, Perelman M: Pulmonary manifestations of early schistosome infection among nonimmune travelers. Am J Med 109:718, 2000

67. Visser LG, Polderman AM, Stuiver PC: Outbreak of schistosomiasis among travelers returning from Mali, West Africa. Clin Infect Dis 20:280, 1995

68. Haribhai HC, Bhigjee AI, Bill PLA, et al: Spinal cord schistosomiasis: a clinical, laboratory and radiological study, with a note on therapeutic aspects. Brain 114:709, 1991

69. Ferrari TC, Moreira PR, Cunha AS: Spinal cord schistosomiasis: a prospective study of 63 cases emphasizing clinical and therapeutic aspects. J Clin Neurosci 11:246, 2004

70. Blair D, Xu ZB, Agatsuma T: Paragonimiasis and the genus *Paragonimus*. Adv Parasitol 42:113, 1999

71. Pachucki CT, Levandowski RA, Brown VA, et al: American paragonimiasis treated with praziquantel. N Engl J Med 311:582, 1984

72. Cha SH, Chang KH, Cho SY, et al: Cerebral paragonimiasis in early active stage: CT and MR features. AJR Am J Roentgenol 162:141, 1994

73. Im JG, Whang HY, Kim WS, et al: Pleuropulmonary paragonimiasis: radiologic findings in 71 patients. AJR Am J Roentgenol 159:39, 1992

74. Yee B, Hsu JI, Favour CB, et al: Pulmonary paragonimiasis in Southeast Asians living in the central San Joaquin Valley. West J Med 156:423, 1992

75. Rangdaeng S, Alpert LC, Khiyami A, et al: Pulmonary paragonimiasis: report of a case with diagnosis by fine needle aspiration cytology. Acta Cytol 36:31, 1992

76. King S, Scholz T: Trematodes of the family Opisthorchiidae: a minireview. Korean J Parasitol 39:209, 2001

77. Harinasuta T, Pungpak S, Keystone JS: Trematode infections: opisthorchiasis, clonorchiasis, fascioliasis, and paragonimiasis. Infect Dis Clin North Am 7:699, 1993

78. Watanapa P, Watanapa WB: Liver fluke–associated cholangiocarcinoma. Br J Surg 89:962, 2002

79. Lim JH: Radiologic findings of clonorchiasis. AJR Am J Roentgenol 155:1001, 1990

80. Pungpak S, Viravan C, Radomyos B, et al: *Opisthorchis viverrini* infection in Thailand: studies on the morbidity of the infection and resolution following praziquantel treatment. Am J Trop Med Hyg 56:311, 1997

81. Arjona R, Riancho JA, Aguado JM, et al: Fascioliasis in developed countries: a review of classic and aberrant forms of the disease. Medicine (Baltimore) 74:13, 1995

82. Mas-Coma MS, Esteban JG, Bargues MD: Epidemiology of human fascioliasis: a review and proposed new classification. Bull World Health Organ 77:340, 1999

83. Price TA, Tuazon CU, Simon GL: Fascioliasis: case reports and review. Clin Infect Dis 17:426, 1993

84. Han JK, Choi BI, Cho JM, et al: Radiological findings of human fascioliasis. Abdom Imaging 18:261, 1993

85. Han JK, Han D, Choi BI, et al: MR findings in human fascioliasis. Trop Med Int Health 1:367, 1996

86. Graczyk TK, Gilman RH, Fried B: Fasciolopsiasis: is it a controllable food-borne disease? Parasitol Res 87:80, 2001

87. Plaut AG, Kampanart-Sanyakorn C, Manning GS: A clinical study of *Fasciolopsis buski* infection in Thailand. Trans R Soc Trop Med Hyg 63:470, 1969

88. Adams KO, Jungkind DL, Berquist EJ, et al: Intestinal fluke infection as a result of eating sushi. Am J Clin Pathol 86:688, 1986

89. MacLean JD, Arthur JR, Ward BJ, et al: Common-source outbreak of acute infection due to the North American liver fluke *Metorchis conjunctus*. Lancet 347:154, 1996

90. Eastburn RL, Fritsche TR, Terhune CA Jr: Human intestinal infection with *Nanophyetus salmincola* from salmonid fishes. Am J Trop Med Hyg 36:586, 1987

91. Schantz PM: Tapeworms (cestodiasis). Gastroenterol Clin North Am 25:637, 1996

92. Butt AA, Aldridge KE, Sanders CV: Infections related to the ingestion of seafood. Part II: parasitic infections and food safety. Lancet Infect Dis 4:294, 2004

93. Raether W, Hanel H: Epidemiology, clinical manifestations and diagnosis of zoonotic cestode infections: an update. Parasitol Res 91:412, 2003

94. Curtis MA, Bylund G: Diphyllobothriasis: fish tapeworm disease in the circumpolar north. Arctic Med Res 50:18, 1991

95. Garcia HH, Gonzalez AE, Evans CA, et al: *Taenia solium* cysticercosis. Lancet 362:547, 2003

96. Schantz PM, Moore AC, Munoz JL, et al: Neurocysticercosis in an Orthodox Jewish community in New York City. N Engl J Med 327:692, 1992

97. Kruskal BA, Moths L, Teele DW: Neurocysticercosis in a child with no history of travel outside the continental United States. Clin Infect Dis 16:290, 1993

98. Locally acquired neurocysticercosis—North Carolina, Massachusetts, and South Carolina, 1989–1991. MMWR Morb Mortal Wkly Rep 41:1, 1992

99. Medina MT, Rosas E, Rubio-Donnadieu F, et al: Neurocysticercosis as the main cause of late-onset epilepsy in Mexico. Arch Intern Med 150:325, 1990

100. Shandera WX, White AC Jr, Chen JC, et al: Neurocysticercosis in Houston, Texas: a report of 112 cases. Medicine (Baltimore) 73:37, 1994

101. Garcia HH, Herrera G, Gilman RH, et al: Discrepancies between cerebral computed tomography and Western blot in the diagnosis of neurocysticercosis. Am J Trop Med Hyg 50:152, 1994

102. Nash TE: Human case management and treatment of cysticercosis. Acta Tropica 87:61, 2003

103. Ramos-Kuri M, Montoya RM, Padilla A, et al: Immunodiagnosis of neurocysticercosis: disappointing performance of serology (enzyme-linked immunosorbent assay) in an unbiased sample of neurological patients. Arch Neurol 49:633, 1992

104. Wilson M, Bryan RT, Fried JA, et al: Clinical evaluation of the cysticercosis enzyme-linked immunoelectrotransfer blot in patients with neurocysticercosis. J Infect Dis 164:1007, 1991

105. Takayanagui OM, Jardim E: Therapy for neurocysticercosis: comparison between albendazole and praziquantel. Arch Neurol 49:290, 1992

106. Carpio A, Santillan F, Leon P, et al: Is the course of neurocysticercosis modified by treatment with antihelminthic agents? Arch Intern Med 155:1982, 1995

107. A trial of antiparasitic treatment to reduce the rate of seizures due to cerebral cysticercosis. Cysticercosis Working Group in Peru. N Engl J Med 350:249, 2004

108. Vazquez V, Sotelo J: The course of seizures after treatment for cerebral cysticercosis. N Engl J Med 327:696, 1992

109. Pau A, Perria C, Turtas S, et al: Long-term follow-up of the surgical treatment of intracranial coenurosis. Br J Neurosurg 4:39, 1990

110. Tsai MD, Chang CN, Ho YS, et al: Cerebral sparganosis diagnosed and treated with stereotactic techniques: report of two cases. J Neurosurg 78:129, 1993

111. Moon WK, Chang KH, Cho SY, et al: Cerebral sparganosis: MR imaging versus CT features. Radiology 188:751, 1993

112. Torgerson PR, Heath DD: Transmission dynamics and control options for *Echinococcus granulosus*. Parasitology 127(suppl):S143, 2003

113. McManus DP, Zhang W, Li J, et al: Echinococcosis. Lancet 362:1295, 2003

114. Rausch RL: Cystic echinococcosis in the Arctic and sub-Arctic. Parasitology 127(suppl):S73, 2003

115. Pawlowski I, Eckert J, Vuitton D, et al: Echinococcosis in humans: clinical aspects, diagnosis, and treatment. WHOI/OIE Manual on Echinococcosis in Humans and Animals: Public Health Problem of Global Concern. Eckert J, Gemmell M, Meslin F-X, et al, Eds. World Health Organisation for Animal Health, Paris, 2001, p 20

116. von Sinner WN: New diagnostic signs in hydatid disease: radiography, ultrasound, CT and MRI correlated to pathology. Eur J Radiol 12:150, 1991

117. Jerray M, Benzarti M, Garrouche A, et al: Hydatid disease of the lungs: study of 386 cases. Am Rev Respir Dis 146:185, 1992

118. Taourel P, Marty-Ane B, Charasset S, et al: Hydatid cyst of the liver: comparison of CT and MRI. J Comput Assist Tomogr 17:80, 1993

119. Biava MF, Dao A, Fortier B: Laboratory diagnosis of cystic hydatid disease. World J Surg 25:10, 2001

120. Filice C, Di Perri G, Strosselli M, et al: Parasitologic findings in percutaneous drainage of human hydatid liver cysts. J Infect Dis 161:1290, 1990

121. Filice C, Brunetti E: Use of PAIR in human cystic echinococcosis. Acta Trop 64:95, 1997

122. Ammann RW, Eckert J: Cestodes: *Echinococcus*. Gastroenterol Clin North Am 25:655, 1996

123. Khuroo MS, Wani NA, Javid G, et al: Percutaneous drainage compared with surgery for hepatic hydatid cysts. N Engl J Med 337:881, 1997

124. Khuroo MS, Dar MY, Yattoo GN, et al: Percutaneous drainage versus albendazole therapy in hepatic hydatidosis: a prospective, randomized study. Gastroenterology 104:1452, 1993

125. Horton RJ: Albendazole in treatment of human cystic echinococcosis: 12 years of experience. Acta Trop 64:79, 1997

126. Ammann RW, Ilitsch N, Marincek B, et al: Effect of chemotherapy on the larval mass and the long-term course of alveolar echinococcosis. Hepatology 19:735, 1994

127. Ishizu H, Uchino J, Sato N, et al: Effect of albendazole on recurrent and residual alveolar echinococcosis of the liver after surgery. Hepatology 25:528, 1997

Acknowledgments

Figures 3, 7,12, and 15 Centers for Disease Control and Prevention Public Health Image Library.

Figure 4 University of Washington Microbiology Laboratory.

Figures 2, 5, 6, 9, 10, 11, 14, and 18 Seward Hung.

Figure 8 Courtesy of Drs. Herbert Cushing and Ken Fife, Indiana University.

Figures 13a through 13c Courtesy of Dr. Steve Pan, Department of Tropical Public Health, Harvard School of Public Health.

Figure 13d Courtesy of Dr. Allen W. Cheever, National Institutes of Health.

Figure 16 Courtesy of Dr. Theodore E. Nash, National Institutes of Health.

XXXVI BACTERIAL INFECTIONS OF THE CENTRAL NERVOUS SYSTEM

Jan V. Hirschmann, m.d.

Several barriers protect the central nervous system from bacterial infection. Considerable force is necessary to breach the skull and allow ingress of organisms into the brain or its coverings. Tight junctions between cells in the cerebral vasculature form the blood-brain barrier, limiting access by blood-borne pathogens. The vertebrae and the dura mater enveloping the spinal cord present strong defenses against incursions by microbes from contiguous areas.

Nevertheless, bacteria occasionally overwhelm or bypass these barriers, resulting in CNS infections. They usually arrive by one of three routes: (1) penetration as a result of trauma or surgery, (2) migration from an adjacent site of infection or colonization, or (3) hematogenous spread from another, often distant, location. The major bacterial infections that result are meningitis, brain abscess, epidural abscess, subdural empyema, and septic thrombophlebitis of the cerebral veins.

Acute Bacterial Meningitis

EPIDEMIOLOGY AND ETIOLOGY

Because vaccination has markedly decreased childhood *Haemophilus influenzae* type b infections, acute bacterial meningitis, previously most common in children, has become predominantly a disease of adults.[1] Bacterial meningitis may occur in otherwise healthy persons, but it usually afflicts those with significant underlying disorders, such as hypogammaglobulinemia, sickle cell anemia, alcoholism, cirrhosis, and concurrent infections of the ears, paranasal sinuses, lungs, or cardiac valves.[2] Patients with HIV disease have a substantially increased risk of bacterial meningitis, usually when their $CD4^+$ T cell count is less than $200/mm^3$. The most common organism is *Streptococcus pneumoniae;* bacteremia is nearly always present.[3] Patients with a splenectomy or poor splenic function may develop overwhelming sepsis, often including meningitis, from encapsulated bacteria, primarily *S. pneumoniae*. Deficiencies of some complement components increase susceptibility to infections with *Neisseria meningitidis,* and head trauma that disrupts the dura mater can lead to meningitis, usually from *S. pneumoniae.* Meningitis that occurs as a complication of neutropenia or cranial surgery is most commonly caused by gram-negative bacilli [*see Table 1*]. A viral upper respiratory tract infection may predispose to meningitis by allowing bacteria that are colonizing the respiratory tract, especially *N. meningitidis,* to enter the bloodstream and invade the meninges.

S. pneumoniae causes approximately 40% to 60% of adult cases of bacterial meningitis in the United States; these cases often occur secondary to bacteremic pneumococcal pneumonia. Other major organisms in community-acquired cases are *N. meningitidis,* primarily in young adults; *Listeria monocytogenes,* especially in immunocompromised or elderly hosts; *H. influenzae;* and group B streptococci (*S. agalactiae*). Usually, the organisms reach the meninges through bacteremia from a mucosal site or a distant infection. Occasionally, the bacteria enter the subarachnoid space from adjacent infectious foci or through direct inoculation via trauma, surgery, or invasive medical procedures, such as lumbar puncture.

DIAGNOSIS

Clinical Features

The duration of symptoms before patients seek medical attention varies from less than 24 hours to more than 1 week. A prodrome resembling a viral upper respiratory tract infection may occur; it is characterized by sore throat, rhinorrhea, and nasal congestion, sometimes accompanied by myalgias. The most common features of bacterial meningitis are headache, fever, nuchal rigidity, and neurologic findings. Less than half of patients have the classic triad of fever, neck stiffness, and a change in mental status; however, almost all have at least two of the four manifestations of headache, fever, neck stiffness, and altered mental status.[4] Fever is present in approximately 95% of patients and typically lasts 4 to 8 days after appropriate therapy has begun.[2,5,6] Stiff neck is apparent in about 90% of patients. Two physical findings that, like nuchal rigidity, arise from meningeal irritation and the neuromuscular response to it are the Kernig sign (when the hip is flexed at 90°, attempted extension of the knee meets resistance at 135°) and the Brudzinski sign (passive flexion of the neck causes knee and hip flexion).[7] Mental changes ranging from lethargy to confusion, stupor, and coma occur in about 80% of patients, somewhat more frequently in the elderly. Seizures, either focal or generalized, occur in about 10% to 30% of patients, usually within 24 hours of admission. Many such seizures arise from alcohol withdrawal. Focal neurologic findings other than seizures develop in about 30% of patients. They include cranial nerve palsies, aphasia, and hemiparesis as a consequence of infection-induced vasculitis or pressure of purulent exudate on neurologic structures. Papilledema occurs only occasionally; its absence does not exclude significantly increased intracranial pressure.

Laboratory Findings

Most patients with acute bacterial meningitis have leukocytosis, and some have hyponatremia, presumably because of inappropriate secretion of antidiuretic hormone. Blood cultures, which are positive in approximately 50% to 60% of cases, are always indicated. The most important diagnostic procedure is a lumbar puncture to obtain cerebrospinal fluid to determine the white cell count and differential, to measure protein and glucose levels, and for culture and Gram stain.

Table 1 Causes of Bacterial Meningitis in Adults[1,2]

Organism	*Approximate Frequency (%)*
Streptococcus pneumoniae	40–60
Neisseria meningitidis	15–25
Listeria monocytogenes	10–15
Haemophilus influenzae	5–10
Other	5–20
Culture negative	10–15

Clinicians commonly obtain a CT scan before doing a lumbar puncture, on the basis of two beliefs: that brain herniation is a frequent risk in meningitis and that CT scans can accurately predict its development. Neither belief is correct. Herniation occurs in approximately 1% of patients with bacterial meningitis,[2] sometimes without a preceding lumbar puncture.[8] CT scans are often normal in those who later experience brain herniation, and most patients who develop this complication have focal neurologic findings before the lumbar puncture that suggest that herniation has already begun: dilated, fixed pupils; Cheyne-Stokes respiration; decerebrate posturing; hemiplegia; and coma.[8] In addition to focal neurologic deficits and abnormal level of consciousness, findings that justify performing CT before lumbar puncture include the following: immune compromise; a history of CNS disease (e.g., mass lesion, stroke, or focal infection); papilledema (although the presence of venous pulsations suggests that the patient does not have increased intracranial pressure); and seizures within 1 week before presentation (some experts will not perform a lumbar puncture in patients with prolonged seizures and will delay lumbar puncture for 30 minutes in patients who have experienced short, convulsive seizures).[9] In the absence of such findings in patients with suspected bacterial meningitis, clinicians should not delay lumbar puncture to obtain a CT scan [*see Figure 1*].

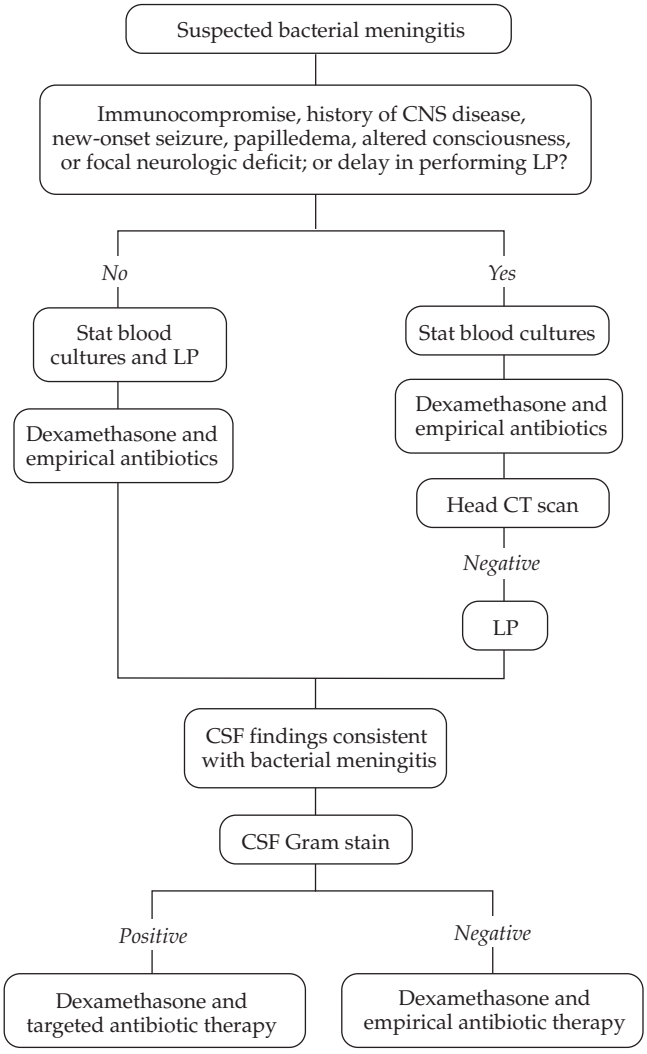


Figure 1 **Management of adult patients with suspected bacterial meningitis.[9] (CSF—cerebrospinal fluid; LP—lumbar puncture)**

In approximately 90% of patients, the opening pressure is above 180 mm H_2O; in 20%, it is above 400 mm H_2O; and in about 5%, it is above 500 mm H_2O. The CSF protein level exceeds 40 mg/dl in approximately 85%, and the glucose level is below 40 mg/dl in about 60%.[2] When bacterial meningitis develops in patients with diabetes and hyperglycemia, the CSF glucose level may be normal, but the ratio of CSF glucose to blood glucose in these patients is usually less than 0.31.[10] The CSF white cell count is greater than 100/mm^3 in about 90% of patients and exceeds 1,000/mm^3 in 15% to 20%. Neutrophils nearly always predominate, constituting at least 80% of cells in 80% to 90% of patients. Occasionally, lymphocytes constitute the majority, especially when the white cell count is very low; lymphocytes constitute the majority in about 25% of patients with meningitis caused by *L. monocytogenes*.[11] Several conditions can cause neutrophilia with low CSF glucose levels and can mimic acute bacterial meningitis [*see Table 2*].

Gram stain is positive in approximately 60% to 90% of patients, indicating a concentration of bacteria exceeding 10^5 organisms/ml. Misinterpretation may occur, especially through mistaking pneumococci for *Listeria*. In listerial meningitis, the Gram stain is positive in only 30% of cases.[12]

The CSF culture is positive in approximately 80% of patients. Previous oral antibiotic therapy has little effect on the WBC or differential cell counts, protein level, or glucose level, but it decreases the number of positive Gram stains by about 20% and the number of positive cultures by about 30%.[13] The effect of parenteral therapy is presumably much greater, especially if given more than a few hours before the lumbar puncture. In patients with negative Gram stains and cultures, latex agglutination tests of the CSF are highly specific for *H. influenzae* type b, meningococci, and pneumococci, but their sensitivity ranges from 50% to 80%.

Imaging Studies

Imaging techniques are ordinarily unnecessary in bacterial meningitis unless specific complications are suspected. The subarachnoid space may be distended on CT or MRI studies; however, this finding is often difficult to interpret in adults, in whom a degree of brain atrophy is common—especially in the elderly. Several days after infection, meningeal enhancement may occur after contrast injection because of meningeal inflammation and vascular congestion. Enhancement of areas in the cerebral cortex usually indicates brain infarction from occlusion of inflamed vessels. Brain edema may occur because of disturbed cerebrovascular autoregulation, dural sinus thrombophlebitis, leakage of fluid from damaged vessels, or cytotoxic edema from injured brain cells. Communicating or noncommunicating hydrocephalus occasionally develops because inflammation obstructs the normal flow of CSF, either at the arachnoid granulations where it is absorbed or at the sites where it leaves the ventricles.

TREATMENT

Intravenous Antibiotics

The most important therapeutic maneuver is prompt initiation of intravenous antibiotics after lumbar puncture, because delay in beginning treatment increases the mortality [*see Tables 3 and 4*].[14] In the rare situation in which lumbar puncture is justifiably postponed—for example, when signs of brain herniation

Table 2 Causes of Neutrophilia with Low Glucose Levels in the CSF

	Infectious Cause	*Noninfectious Cause*
Common	Bacterial meningitis	—
Uncommon	Viral meningitis* (in very early phase only) Certain parameningeal infections Subdural empyema* (usually produces lymphocytic–normal glucose profile) ?Early brain abscess at bacterial cerebritis stage* Cerebral abscess with leakage or rupture into ventricle Embolic cerebral infarction* (bacterial endocarditis) Amebic meningoencephalitis Tuberculous meningitis (only very early in the disease and only in a small percentage of cases) Acute hemorrhagic leukoencephalitis	Chemical meningitis Exogenous (e.g., contrast media, detergents used in cleaning needles) Endogenous (release of material into CSF from tumors: dermoids, craniopharyngiomas) Unusual diseases* Initial phase of Mollaret meningitis Behçet syndrome Hypersensitivity meningitis Drug-induced* (sulfonamides, tolmetin, ibuprofen, isoniazid)

*Glucose level in CSF usually normal.

are present (see above)—blood cultures should be obtained and antibiotics begun, even though they may reduce the yield of the subsequent CSF Gram stain and culture. The antimicrobial agent chosen should be bactericidal against the suspected pathogens and should achieve good CSF levels. In a Danish study, ceftriaxone plus penicillin was an appropriate choice for empirical treatment in 97% of the adult patients with bacterial meningitis in the study population.[15]

The findings on CSF Gram stain may help guide the choice of antibiotic; when the stain is negative, however, ceftriaxone (2 g I.V. every 12 hours, or 4 g I.V. every 24 hours) is a good choice in adults.[15] Ceftriaxone provides coverage for meningococci, *H. influenzae,* group B streptococci, and *S. pneumoniae,* including many penicillin-resistant strains. In a Danish study, ceftriaxone plus penicillin was an appropriate choice for empirical treatment in 97% of the adult patients with bacterial meningitis in the study population.[16]

With suspected high-level penicillin resistance, use of vancomycin (2 g every 12 hours) is prudent until culture and susceptibility test results are available. If Gram stain of the CSF is negative or if a distinction between pneumococci and *Listeria* is uncertain, the patient should receive, in addition, ampicillin at a dosage of 2 g every 4 hours to treat possible *L. monocytogenes* infection.

Although controlled studies have not delineated the optimal length of antibiotic therapy, the duration is usually 7 days for *H. influenzae* and 10 to 14 days for most other organisms. For meningococcal disease, treatment for 4 to 5 days is of proven efficacy, and in one study, 3 days of treatment with intravenous benzylpenicillin was effective.[17]

Repeat lumbar puncture during or after therapy is usually unnecessary unless treatment seems to be failing.

Systemic Corticosteroids

In adults with acute pneumococcal meningitis, adjuvant treatment with dexamethasone has been shown to lower the risk of an unfavorable outcome and to lower mortality, without increasing the likelihood of gastrointestinal bleeding.[18,19] Consequently, guidelines from the Infectious Diseases Society of America recommend that adults with suspected or proven pneumococcal meningitis receive adjunctive dexamethasone.[9] The dosage is 0.15 mg/kg every 6 hours for 2 to 4 days, with the first dose given 10 to 20 minutes before or at least concomitant with the first dose of antibiotics. Dexamethasone should be continued only if

Table 3 Empirical Antibiotic Therapy for Purulent Meningitis in Adults[9]

Patient Factor	*Common Bacterial Pathogens*	*Antibiotic*
Age 2–50 yr	*Neisseria meningitidis, Streptococcus pneumoniae*	Vancomycin plus a third-generation cephalosporin*; consider adding rifampin if dexamethasone is given
Age > 50 yr	*S. pneumoniae*, *N. meningitidis*, *Listeria monocytogenes*, aerobic gram-negative bacilli	Vancomycin plus ampicillin plus a third-generation cephalosporin*; consider adding rifampin if dexamethasone is given
Basilar skull fracture	*S. pneumoniae*, *Haemophilus influenzae*, group A β-hemolytic streptococci	Vancomycin plus a third-generation cephalosporin*
Penetrating head trauma	*Staphylococcus aureus*, coagulase-negative staphylococci, especially *S. epidermidis*; aerobic gram-negative bacilli, including *Pseudomonas aeruginosa*	Vancomycin plus cefepime, vancomycin plus ceftazidime, or vancomycin plus meropenem
Postneurosurgery	Aerobic gram-negative bacilli, including *P. aeruginosa*; *S. aureus*; coagulase-negative staphylococci, especially *S. epidermidis*	Vancomycin plus cefepime, vancomycin plus ceftazidime, or vancomycin plus meropenem
CSF shunt	Coagulase-negative staphylococci, especially *S. epidermidis*; *S. aureus*; aerobic gram-negative bacilli, including *P. aeruginosa*; *Propionibacterium acnes*	Vancomycin plus cefepime, vancomycin plus ceftazidime, or vancomycin plus meropenem

*Ceftriaxone or cefotaxime.
CSF—cerebrospinal fluid

Table 4 Antibiotic Therapy for Bacterial Meningitis in Adults[9]

Organism	*Antibiotic*	*Dosage*	*Alternative Antibiotic*
Streptococcus pneumoniae			
Penicillin MIC < 0.1 μg/ml	Penicillin G *or* Ampicillin	4 million U I.V. q. 4 hr 2 g I.V. q. 4 hr	Third-generation cephalosporin* or chloramphenicol, 1 g I.V. q. 6 hr
Penicillin MIC 0.1–1.0 μg/ml	Ceftriaxone	2 g I.V. q. 12 hr	Cefepime, 2 g I.V. q. 8 hr, or meropenem, 2 g I.V. q. 8 hr
Penicillin MIC > 1.0 μg/ml	Vancomycin *plus* third-generation cephalosporin*‡	30–45 mg/kg I.V.	Fluoroquinolone†
Cefotaxime or ceftriaxone MIC > 1 μg/ml	Vancomycin *plus* third-generation cephalosporin*‡	30–45 mg/kg I.V.	Fluoroquinolone†
Neisseria meningitidis			
Penicillin MIC < 0.1 μg/ml	Penicillin G *or* Ampicillin	4 million U I.V. q. 4 hr 2 g I.V. q. 4 hr	Third-generation cephalosporin* or chloramphenicol, 1 g I.V. q. 6 hr
Penicillin MIC 0.1–1.0 μg/ml	Third-generation cephalosporin*		Chloramphenicol, 1 g I.V. q. 6 hr, fluoroquinolone,† meropenem, 2 g I.V. q. 8 hr
Listeria monocytogenes	Ampicillin§ *or* Penicillin G§	2 g I.V. q. 4 hr 4 million U I.V. q. 4 hr	Trimethoprim (TMP)-sulfamethoxazole (SMX): 10–20 mg/kg TMP, 50–100 mg/kg SMX q.d. in two to four divided doses; or meropenem, 2 g I.V. q. 8 hr
Haemophilus influenzae			
β-Lactamase negative	Ampicillin	2 g I.V. q. 4 hr	Third-generation cephalosporin*; cefepime, 2 g I.V. q. 8 hr; chloramphenicol, 1 g I.V. q. 6 hr or fluoroquinolone†
β-Lactamase positive	Third-generation cephalosporin*		Cefepime, 2 g I.V. q. 8 hr; chloramphenicol, 1 g I.V. q. 6 hr; or fluoroquinolone†
Group B *Streptococcus* (*S. agalactiae*)	Ampicillin§ *or* Penicillin G§	2 g I.V. q. 4 hr 4 million U I.V. q. 4 hr	Third-generation cephalosporin*
Escherichia coli, *Klebsiella*, *Proteus*, and other enteric gram-negative bacilli	Third-generation cephalosporin*		Aztreonam, 2 g q. 6–8 hr; fluoroquinolone†; meropenem, 2 g I.V. q. 8 hr; TMP-SMX: 10–20 mg/kg TMP, 50–100 mg/kg SMX, q.d. in two to four divided doses; or ampicillin, 2 g I.V. q. 4 hr
Pseudomonas aeruginosa	Cefepime§ *or* Ceftazidime§	2 g I.V. q. 8 hr 2 g I.V. q. 8 hr	Aztreonam, 2 g q. 6–8 hr§; ciprofloxacin, 400 mg q. 8–12 hr§; or meropenem, 2 g I.V. q. 8 hr§
Staphylococcus aureus			
Methicillin sensitive	Nafcillin *or* Oxacillin	1.5–2 g I.V. q. 4 hr 1.5–2 g I.V. q. 4 hr	Vancomycin, 30–45 mg/kg I.V., or meropenem, 2 g I.V. q. 8 hr
Methicillin resistant	Vancomycin‖	30–45 mg/kg I.V.	TMP-SMX: 10–20 mg/kg TMP, 50–100 mg/kg SMX, q.d. in two to four divided doses; or linezolid, 600 mg I.V. q. 12 hr
Staphylococcus epidermidis	Vancomycin‖	30–45 mg/kg I.V.	Linezolid, 600 mg I.V. q. 12 hr
Enterococcus species			
Ampicillin susceptible	Ampicillin *plus* Gentamicin	2 g I.V. q. 4 hr 2 mg/kg first dose I.V., then 1.7 mg/kg I.V. q. 8 hr with normal renal function (or 4–8 mg q.d. intrathecally) 30–45 mg/kg I.V.	—
Ampicillin resistant	Vancomycin *plus* Gentamicin	2 mg/kg first dose I.V., then 1.7 mg/kg I.V. q. 8 hr with normal renal function (or 4–8 mg q.d. intrathecally)	—
Ampicillin and vancomycin resistant	Linezolid	600 mg I.V. q. 12 hr	

*Ceftriaxone, 2 g I.V. q. 12 hr, or cefotaxime, 2 g q. 4–6 hr.
†Gatifloxacin, 400 mg I.V. q.d., or moxifloxacin, 400 mg I.V. q.d.
‡Consider adding rifampin, 600 mg q.d., if the ceftriaxone MIC is > 2 μg/ml.
§Consider adding an aminoglycoside.
‖Consider adding rifampin, 600 mg q.d.
MIC—minimum inhibitory concentration

the CSF Gram stain reveals gram-positive diplococci or if blood or CSF cultures are positive for *S. pneumoniae.* Dexamethasone treatment is unlikely to improve the outcome in adult patients who have already received antimicrobial therapy and hence should not be used in this circumstance.

COMPLICATIONS

Many complications may occur in meningitis patients, including secondary nosocomial infections, systemic venous thromboembolism, and adverse effects from the medications administered. The most common systemic complications occur from the initial infection, however, and include septic shock, disseminated intravascular coagulation, and the acute respiratory distress syndrome.[20] The most frequent neurologic complications other than seizures and altered mentation are cerebral edema, hydrocephalus, and cerebrovascular complications.[21]

In bacterial meningitis, the internal carotid, basilar, and vertebral arteries and their branches lie within a purulent subarachnoid exudate, which can provoke vascular inflammation and constriction. Smaller arteries may demonstrate vessel wall irregularities, occlusion, narrowing, or widening. Thrombosis of the superior sagittal sinus or the cortical veins can occur. These abnormalities may cause discrete areas of brain damage—which is evidenced by focal neurologic findings—or increased intracranial pressure that is produced by leakage of cerebral vessels (vasogenic edema), swelling of damaged neurons (cytotoxic edema), or increased blood volume that occurs when sinus venous thrombosis impedes drainage of blood from the brain.

PROGNOSIS

The mortality associated with meningitis is approximately 10% in meningococcal infections and is typically 20% to 30% in meningitis caused by other organisms. Factors associated with poorer prognoses include infection with *S. pneumoniae,* advanced age (> 60 years), onset of seizures during the first 24 hours after infection, hypotension, and coma or obtundation on hospital admission.[4,22] Most survivors recover completely. Approximately 30% of patients with pneumococcal meningitis have moderate to severe sequelae, including dementia, seizures, hearing loss, and gait disturbances; about 20% have mild problems, such as dizziness, slightly impaired memory, headaches, and phonophobia.[23]

Bacterial Meningitis in Special Circumstances

POSTTRAUMATIC BACTERIAL MENINGITIS

Epidemiology and Etiology

Bacterial meningitis develops in about 1% of persons who receive medical attention for blunt cranial injuries, most commonly from motor vehicle accidents.[24] At risk are those in whom the force of impact fractures the base of the skull, tearing the dura mater, the underlying arachnoid, and adjacent soft tissues. The trauma, however, need not be severe: coma or retrograde amnesia is absent or brief in 30% to 40% of cases.

The most common cause of posttraumatic meningitis is *S. pneumoniae,* which accounts for about 65% of cases. In the remainder of cases, the causes are primarily other streptococci, *H. influenzae,* meningococci, and *Staphylococcus aureus*. Enteric gram-negative bacilli are rare causes, except in patients who have previously received antimicrobial therapy.

Pathogenesis

The dural rent resulting from blunt cranial trauma creates a fistula between the subarachnoid space and the nasal cavity, paranasal sinuses, or ear. The most common site for these abnormal communications is the cribriform plate; the disruptions are multiple, however, in about 40% of cases. When these rents are large, they are evident as CSF rhinorrhea; smaller leaks may go unrecognized, and drainage may be delayed or intermittent when tissue occludes the dural laceration. When trauma weakens but does not tear the dura, CSF rhinorrhea may begin abruptly after sudden increases in cerebrospinal pressure from coughing, sneezing, or straining. When the dural fistula occurs in the middle cranial fossa because of a fractured temporal bone, CSF may exit through the ear canal (if the tympanic membrane ruptures) or drain into the pharynx through the eustachian tube and go unrecognized. Evidence of basilar skull fractures may appear shortly after the injury as periorbital bruising, anosmia, hemotympanum, bloody ear drainage, or Battle sign (ecchymoses behind the ear).

Diagnosis

Clinical features The usual clinical features of bacterial meningitis, including fever, headache, and stiff neck, are present in posttraumatic meningitis. Mortality is comparatively low (10%), however, perhaps because the fistula allows drainage of CSF (relieving intracranial hypertension) or because many patients with cranial trauma are young and otherwise healthy. The interval between blunt cranial injury and meningitis is usually less than 1 month, but infection may occur years later; when evaluating patients with unexplained meningitis, clinicians should therefore ask about any significant head trauma, even if such trauma occurred in the remote past.

Recognition of CSF rhinorrhea or otorrhea allows a diagnosis of a dural fistula. The discharge often increases with sudden movement, lowering of the head, performance of the Valsalva maneuver, or jugular vein compression. Unlike normal nasal secretions, the CSF fluid is not sticky, does not stiffen a handkerchief on drying, may taste salty, and has a glucose level greater than 30 mg/dl in the absence of infection. The most sensitive and specific way to identify CSF is by detection of a β_2-transferrin found only in that body fluid.[25]

Imaging studies Several techniques help localize the leak of CSF. CT scans may demonstrate the fistula site by disclosing sinus fractures, displaced bony fragments, or intracranial air (pneumocephalus). Radionuclide-labeled albumin and various dyes injected in the subarachnoid space may leak into the middle and superior meatus of the nose and become detectable on cotton pledgets placed in these locations. Magnetic resonance cisternography, a noninvasive and rapid procedure, seems highly accurate in detecting fistulas.[25] This technology relies on the fact that CSF gives a high signal on T_2-weighted images: demonstration of a continuous high signal through the cribriform plate or the paranasal sinuses that is similar to the signal given by the CSF in the basal cisterns indicates a dural fistula.

Treatment

Because many dural fistulas resolve spontaneously, repair is usually unnecessary unless problems persist for 2 to 3 weeks after the injury. Patients with facial fractures or meningitis are exceptions. Those with fracture sites accessible through the nose typically undergo transnasal closure with fibrin glue, fascia lata,

or local flaps. Otherwise, intracranial repair is required. Recurrences may develop if the dural fistula remains uncorrected. Vancomycin plus a third-generation cephalosporin (e.g., ceftriaxone or cefotaxime) is a good antibiotic choice for bacterial meningitis after closed head trauma when the CSF Gram stain is negative. A randomized trial of patients with acute traumatic pneumocephalus after mild head injury found that ceftriaxone was not effective in preventing meningitis in these patients.[26]

RECURRENT BACTERIAL MENINGITIS NOT CAUSED BY TRAUMA

Although a dural fistula created by blunt cranial trauma is the most common cause of recurrent bacterial meningitis, seemingly spontaneous leaks of CSF occasionally occur; most of these probably arise from a congenital or acquired weakness in the dura. Defects of the skull or of the floor of the middle cranial fossa are other potential sources of recurrent meningitis, as are bony complications from chronic ear infections. Congenital malformations can lead to dermal sinus tracts, often associated with intradural dermoid tumors, that connect the skin to the CSF. They commonly occur at the lumbodorsal spine but may be present anywhere along the midline. Abnormalities at the cutaneous exit site can include small tufts of hair, a nevus, or a dimple from which purulent drainage may emanate. Other conditions associated with recurrent bacterial meningitis are hypogammaglobulinemia, asplenia, and deficiencies of certain components of complement, which especially predispose to meningococcal meningitis [*see Table 5*].

Table 5 Causes of Recurrent Meningitis

*Bacterial Meningitis**
Predisposing anatomic defects
Congenital: meningomyeloceles, dermal sinuses
Traumatic: skull fractures through cribriform plate or petrous ridge
Postoperative: craniotomy, transsphenoidal hypophysectomy
Tumors: direct invasion through the dura
Empty sella syndrome
Parameningeal infections*
Mastoiditis, sinusitis, osteomyelitis of skull
Immunologic defects*
Immunoglobulin deficiencies: congenital or acquired (e.g., myeloma)
Asplenic state: splenectomy or functional asplenia (e.g., sickle cell anemia)
Complement deficiencies: C6, C7, or C8 deficiency
Endogenous Chemical Meningitis
Tumors†
Craniopharyngioma, epidermoid cyst
Unusual Nonbacterial Recurrent Meningitides
Mollaret meningitis†
Behçet syndrome†
Drug hypersensitivity†
Sulfonamides, ibuprofen
Systemic lupus erythematosus†
Uveoencephalitides‡ (Vogt-Koyanagi syndrome, Harada syndrome)

*Neutrophilic CSF pleocytosis.
†Either neutrophilic or lymphocytic CSF pleocytosis.
‡Lymphocytic CSF pleocytosis.

CEREBROSPINAL FLUID SHUNT INFECTIONS

CSF shunts, used to relieve hydrocephalus, usually consist of a ventricular catheter; a pressure-regulating valve and reservoir, commonly placed just outside the skull; and distal tubing that drains the CSF, most often into the peritoneum but sometimes into the right atrium, pleura, or other sites.

Etiology

Mechanisms of infection of CSF shunts include implantation of microbes during shunt insertion, entry through a break in the overlying skin, hematogenous dissemination from a distant site, and retrograde spread of organisms from the distal end (in the peritoneum or another body cavity).[27] The most common origin is contamination of the device by skin flora at the time of surgery. Most infections become apparent within a few weeks of the shunt's insertion. Entry of microbes into the shunt from overlying skin is probably the next most frequent cause of infection. Retrograde and hematogenous infections seem to be rare except in the case of ventriculoatrial shunts, in which the distal end lies within the bloodstream.

The bacteriology of CSF shunt infections reflects the probable cutaneous source of infection in most cases. The most common isolates are staphylococci, with coagulase-negative species accounting for approximately 50% to 60% of cases and *S. aureus* for approximately 20%. Streptococci, anaerobes, diphtheroids, and polymicrobial infections, primarily from these and other cutaneous species, constitute most of the remaining cases. About 10% of isolates are gram-negative bacilli, and about 5% are the same organisms that commonly cause community-acquired meningitis in patients without shunts: pneumococci, meningococci, and *H. influenzae*.

Diagnosis

Clinical features The manifestations of CSF shunt infections differ from the usual findings of bacterial meningitis and depend on the site involved. Fever is usually present. When ventriculitis occurs, nuchal rigidity rarely develops, because infected CSF typically does not communicate from the ventricle to the meninges. Mental changes, including headache, nausea, lethargy, and confusion, are the most common abnormalities; these may result from infection in the ventricle or from increased intracranial pressure caused by obstruction of the shunt anywhere along its route. Evidence of infection over the shunt may sometimes be detected by the presence of wound dehiscence, cellulitis, or purulent drainage. When the distal end lies within the atrium, bacteremia may occur, sometimes complicated by endocarditis. In chronic bacteremia, usually caused by coagulase-negative staphylococci, renal disease (so-called shunt nephritis) produced by deposition of immune complexes in the kidney may develop. When the distal end lies in the peritoneum, infection may cause peritonitis, manifested by abdominal pain and tenderness, anorexia, and fever. Often, the infection is more insidious, with enlarging cysts developing around the end of the catheter as the inflamed peritoneum fails to absorb the CSF.

Laboratory findings The peripheral white cell count is often normal. Blood cultures are characteristically positive in patients with infected ventriculoatrial shunts but are usually negative when the distal end lies elsewhere. The culture of CSF obtained through needle aspiration of the shunt reservoir or the culture of any fluid in contact with the shunt is the best approach to delineating the cause. The CSF will usually show pleocytosis, which is often mild, and an elevated protein level; the glucose level is typ-

ically normal or slightly decreased. Gram stain may reveal bacteria, but it is often negative.

Treatment

Treatment of shunt infections usually entails systemic antibiotics, shunt removal, and temporary placement of an external ventricular drain to control intracranial pressure (and to provide a conduit for intraventricular antibiotic administration). The systemic antimicrobials used for the treatment of shunt infections are nafcillin for susceptible strains of *S. aureus* and vancomycin for methicillin-resistant *S. aureus* and coagulase-negative staphylococci, often supplemented by oral rifampin (600 mg daily). Ceftriaxone is used for susceptible gram-negative bacilli; ceftazidime is appropriate for ceftriaxone-resistant organisms, including *Pseudomonas aeruginosa.* Empirical antibiotic treatment of CSF shunt infections is with vancomycin plus cefepime, ceftazidime, or meropenem. Intraventricular antibiotics are commonly employed in patients with ventriculitis (generally, vancomycin at a dosage of 10 to 20 mg daily for gram-positive organisms and gentamicin at a dosage of 4 to 8 mg daily for gram-negative bacilli). Once the infection has resolved (usually after several days), a new shunt can be inserted.

NOSOCOMIAL MENINGITIS AND MENINGITIS AFTER NEUROSURGERY

Epidemiology and Etiology

Nosocomial meningitis is quite rare except in patients who have recently undergone neurosurgery. A study of 51 hospitalized patients who underwent lumbar puncture to exclude bacterial meningitis revealed no cases, despite the presence of mental status changes in 78%, fever in 47%, and meningeal signs or headache in 22%.[28] By contrast, 92% of patients with community-acquired bacterial meningitis had either meningeal signs or headache. The conclusion is that lumbar puncture can be safely withheld in most hospitalized patients who have fever and changes in mental status; the exceptions to this rule are patients who have headache or nuchal rigidity and patients who have recently undergone neurosurgery.

Even after neurosurgery, bacterial meningitis is very uncommon, especially after spinal surgery or craniotomy; it is somewhat less rare, but still unusual, after transsphenoidal procedures. During neurosurgery, bacteria most commonly enter the subarachnoid space from contiguous sites of colonization on the skin and mucous membranes. Less frequent sources or causes of infection include contaminated instruments or irrigation solutions, airborne organisms, breaches of sterile technique, and organisms from adjacent sites of suppuration. Postoperative sources are wound infections at the operative site, indwelling catheters that allow organisms to enter the ventricle or subarachnoid space, and CSF leaks. A very uncommon cause is hematogenous spread from a distant site of infection, such as the lungs or urinary tract.

Approximately one half of cases of bacterial meningitis after craniotomy or spinal surgery occur in the first postoperative week, approximately 25% in the second week, and the remainder after the second week. After transsphenoidal surgery, about 85% of cases occur in the first week. The onset of meningitis may be inconspicuous, however, because headache, nuchal rigidity, and impaired consciousness are common after cranial surgery.

Diagnosis

Clinical features Nearly all patients with nosocomial or postneurosurgical meningitis have fever, often accompanied by worsening meningismus, deteriorating mentation, and erythema or purulence at the incision site. CSF rhinorrhea may appear in patients with dural defects.

Laboratory findings The peripheral white cell count is usually elevated; the elevation is accompanied by neutrophilia and increased bands. Blood cultures are positive in about 30% of cases. The most important diagnostic test is sampling of the CSF, usually after obtaining a CT scan to exclude brain abscesses or other fluid collections. In most cases of postoperative meningitis, the CSF protein level is 100 to 500 mg/dl and the white cell count exceeds 1,000/mm^3, with neutrophilic predominance. The CSF glucose level is decreased in 85% of cases; Gram stain discloses the responsible bacteria in 50% to 80%.

The infecting organisms are primarily gram-negative bacilli, including *P. aeruginosa* and species of *Klebsiella* and *Enterobacter*. In patients with ventricular drains, coagulase-negative staphylococci predominate. The bacteria isolated in meningitis after transsphenoidal surgery are more diverse, reflecting various nasal mucosal organisms—especially as altered by hospitalization and previous antimicrobial therapy. Causes include gram-negative bacilli, streptococci, staphylococci, diphtheroids, anaerobes, *Haemophilus* species, and, sometimes, polymicrobial isolates.

Differential Diagnosis

A confounding diagnostic problem in neurosurgical patients with a negative Gram stain is that an aseptic meningitis may develop several days after surgery, probably in response to blood in the subarachnoid space. Fever, nuchal rigidity, nausea, vomiting, and altered mentation may all occur, and the findings on lumbar puncture resemble a bacterial infection, including marked pleocytosis, neutrophilic predominance, increased protein, and, sometimes, decreased glucose. However, in one study, no patients with postneurosurgical chemical meningitis had a CSF white cell count above 7,500/mm^3 or a CSF glucose level below 10 mg/dl.[29]

Treatment

Antimicrobial therapy should be directed by the results of CSF Gram stain and culture. For gram-negative bacilli, ceftazidime (2 g I.V. every 8 hours) is the drug of choice unless the organism is known to be susceptible to ceftriaxone. Often, ceftazidime is combined with intrathecal or intraventricular injection of gentamicin (4 to 8 mg daily). For gram-positive organisms, parenteral vancomycin, supplemented by oral or parenteral rifampin and intrathecal or intraventricular vancomycin (10 to 20 mg daily), is a good regimen. When the Gram stain is negative, a combination of ceftazidime and vancomycin is reasonable until culture results become available.

When the distinction between bacterial and aseptic meningitis is unclear, a reasonable approach is to initiate appropriate empirical antimicrobial therapy. If the cultures remain negative, it is appropriate to withdraw the antibiotics and administer systemic corticosteroids, which seem useful in this form of chemical meningitis.[30]

Brain Abscess

EPIDEMIOLOGY AND ETIOLOGY

Brain abscesses are about two to three times more frequent in males than in females. About 50% arise from contiguous infec-

tions, such as sinusitis and otitis media, with the organisms arriving in the brain from direct extension or retrograde spread through the venous system. Pathogens from distant areas of suppuration may reach the brain via a hematogenous route through the cerebral arteries. Occasionally, abscesses develop after penetrating trauma or surgery, with organisms directly inoculated into the cerebral tissue. In approximately 20% to 30% of brain abscesses, the source of origin is inapparent.[31]

Chronic otitis media is a common predisposing factor, with the abscesses most frequently forming in the temporal lobe or cerebellum. Those occurring secondary to frontal sinusitis usually develop in the frontal lobe, whereas abscesses originating from the sphenoidal sinus are typically in the temporal lobe. Hematogenous infections from distant sites are commonly multiple and most likely to cause suppuration in the territories supplied by the distal middle cerebral artery. Frequently, these arise when blood-borne organisms bypass the ordinary filtering mechanisms of the lung in such conditions as left-sided endocarditis, pulmonary infections, and right-to-left cardiac or pulmonary shunts. In patients with shunts, concurrent hypoxemia and secondary erythrocytosis may cause brain damage that predisposes them to subsequent infection. Immunocompromised patients, especially patients who have undergone bone marrow transplantation or solid organ transplantation, have an increased risk of developing brain abscesses, usually through hematogenous spread.

Brain abscesses begin as localized—but poorly demarcated—areas of encephalitis, often called cerebritis. At this stage, tissue has not liquefied, and needle aspiration or surgery is unhelpful. Over several days, necrosis occurs, with pus forming in the center of the infection. With time, a capsule of fibrous tissue develops, delimiting the suppuration, but pus may rupture through it into the ventricle or adjacent cerebral tissue, forming satellite abscesses.

The most common organisms isolated from brain abscesses are aerobic, microaerophilic, or anaerobic streptococci; other anaerobic bacteria; staphylococci; and gram-negative bacilli (especially when the abscess arises from chronic ear infection). Common associations are *Klebsiella pneumoniae* and viridans streptococci in brain abscess from hematogenous spread; *S. aureus*, *K. pneumoniae*, and viridans streptococci in postneurosurgical brain abscess; viridans streptococci from paranasal sinusitis infections; and *Proteus* species in polymicrobial brain abscesses of otogenic origin.[32] *K. pneumoniae* is an especially common pathogen in patients with diabetes.[33]

DIAGNOSIS

Clinical Features

Most patients with brain abscess have symptoms for several days to weeks before they seek medical attention. Fever is present in only about 40% to 50% of cases. The most common symptom is headache, which may be diffuse or localized to the abscess area. Other frequent findings are altered mental states, such as confusion, aberrant behavior, and somnolence; generalized or focal seizures; and specific neurologic abnormalities, depending on the abscess site. With temporal lobe involvement, contralateral sensory or motor signs, aphasia, and an upper homonymous quadrantanopia may occur. In cerebellar abscesses, ataxia in the ipsilateral extremities, nystagmus, signs of increased intracranial pressure, and ataxic gait are characteristic. Frontal lobe involvement commonly results in hemiparesis, aphasia, and impaired mentation, whereas the main finding in occipital lobe abscesses is homonymous hemianopsia. Typical features of parietal lobe abscesses include hemianesthesia, homonymous hemianopsia, neglect of one half of the body, alexia, and impaired spatial perception. Papilledema is infrequent, irrespective of the abscess site.

Laboratory Findings

Leukocytosis occurs in approximately 50% of patients, and an elevated erythrocyte sedimentation rate occurs in about 60%. Lumbar punctures are not recommended unless meningitis is strongly suspected, because of the low risk of brain herniation and the low likelihood of positive cultures. Examination of the CSF usually shows a pleocytosis with mixed neutrophils and lymphocytes, elevated protein levels, normal glucose levels, and a negative Gram stain. Blood cultures are positive in about 10% to 20% of patients.

Imaging Studies

During the cerebritis phase, CT scans show low-density abnormalities with mass effect, sometimes with patchy enhancement. On MRI, the characteristic findings are areas of low density on T_1-weighted images and, on proton-density or T_2-weighted images, high-intensity areas surrounded by areas of patchy enhancement with gadolinium. In a mature abscess, the CT scan shows a low-density lesion with uniform ring enhancement and a surrounding area of low density, representing cerebral edema [*see Figure 2*]. On T_1-weighted images, the encapsulated abscess appears as a round, low-intensity lesion with mass effect and a surrounding area of low density, signifying edema. On proton-density and T_2-weighted images, the abscess has a high-intensity signal in the center and in the surrounding parenchyma as a consequence of the adjacent cerebral swelling. Ring enhancement occurs with gadolinium.

TREATMENT AND PROGNOSIS

Most patients should undergo prompt stereotactic-guided needle aspiration of pus[34] or craniotomy with excision of the abscess. If possible, antimicrobials should be withheld until a specimen is obtained, to avoid reducing the yield on cultures. Gram stain and culture determine antibiotic choice; appropriate empirical therapy that can be used until the microbiologic results become available include agents effective against anaerobes, streptococci, staphylococci, and gram-negative bacilli. A good regimen is a combination of ceftriaxone and metronidazole.[35] Patients with significant cerebral edema should receive dexamethasone. Some patients may respond to antibiotic therapy alone—especially those with cerebritis only, those with numerous abscesses that are inaccessible to drainage, or those with small abscesses who have stable neurologic function.

Mortality may range from 5% to 25%. Factors associated with poor prognosis include a low Glasgow Coma Scale score, immunodeficiency, or the presence of underlying disease.[36] Many patients have residual neurologic problems, most commonly focal seizures.

Spinal Epidural Abscess

PATHOGENESIS AND PATHOPHYSIOLOGY

The epidural space, an area posterolateral to the spinal cord containing fat and blood vessels, can potentially extend anteriorly to where the dura adheres to the posterior vertebral bodies.

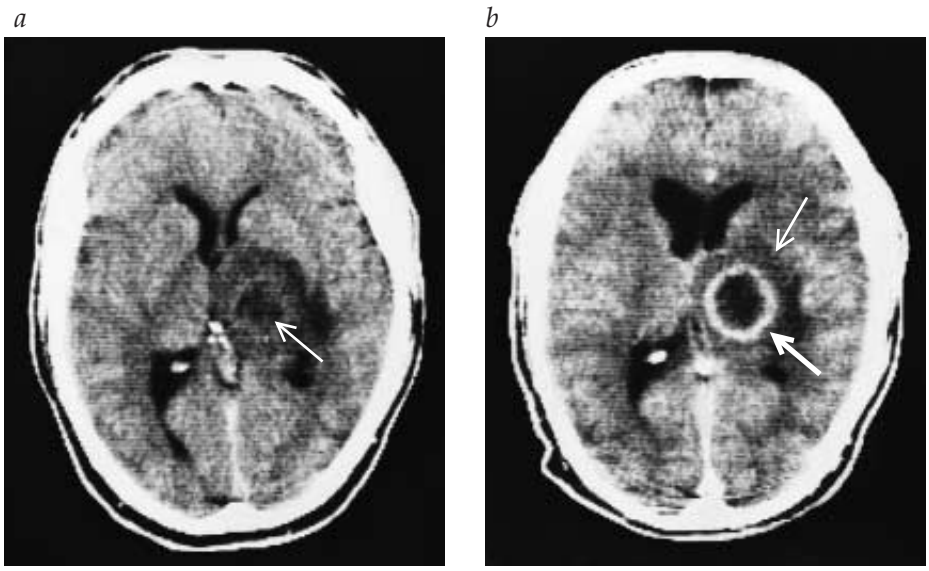


Figure 2 **(*a*) CT scan of a 39-year-old patient reveals a large thalamic abscess. Plain scan demonstrates abscess and a surrounding zone of what may be either increased vascularity or an actual capsule formation (arrow). (*b*) Intravenous administration of contrast media enhances definition of the surrounding zone (thick arrow). A wide area of periabscess edema (thin arrow) is also evident. The mass effect can be seen clearly in the asymmetry of the ventricles.**

Most epidural abscesses occur in midthoracic and lumbar locations, where the space is largest. Cervical abscesses are less common, in part because the epidural space is very small in that area. Because no anatomic barriers prevent superior and inferior extension, abscesses typically involve several (an average of four) adjacent vertebral segments.[37]

Organisms may arrive in the epidural space from contiguous infection, especially vertebral osteomyelitis, but also from retropharyngeal, psoas, perivertebral, or perinephric abscess. All these infections are predominantly anterior to the spinal cord. Microbes also enter the epidural space as a result of penetrating trauma, surgery, and epidural catheterization for analgesic administration. Hematogenous spread may occur from distant foci of infection, including the skin and soft tissue, urinary tract, heart valves, and respiratory tract. Patients with frequent or prolonged bacteremias, such as intravenous drug users and those undergoing hemodialysis, are at increased risk.[38] Diabetes mellitus, alcoholism, and underlying disease in the affected spinal area are other predisposing factors. Many patients report recent trauma near the affected location, suggesting hemorrhage at the site with subsequent bacteremic superinfection. Often, the source of infection is inapparent. Males are more frequently affected than females, for unclear reasons.

Pus and granulation tissue in the epidural space can damage the spinal cord, in part by compression, which may impede arterial flow and cause ischemia. Inflammation and increased pressure can also provoke thrombophlebitis or venous thrombosis, which can lead to cord edema and infarction.

ETIOLOGY

S. aureus is the most common isolate from spinal epidural abscess; most of the other isolates are streptococci and gram-negative bacilli. Usually, only a single species of organism is responsible, but polymicrobial infections occur in approximately 5% to 10% of cases. In the United States, *Mycobacterium tuberculosis* is an occasional cause of epidural abscess. In such cases, the infection usually arises from an adjacent vertebral osteomyelitis. Patients may feel back pain for quite some time (weeks to months) before neurologic symptoms appear. Most of these patients have no evidence of pulmonary tuberculosis at the time of epidural infection.

DIAGNOSIS

Clinical Features

The most common symptom of spinal epidural abscess is back pain over the infected area. It is typically persistent and unrelieved by altered position or bed rest. Localized tenderness is often present; fever is common but is not invariable. With time, patients may develop radicular pain, primarily along the route of the affected level of the spinal cord. The next stage is usually marked by muscle weakness, typically paraparesis with thoracic involvement or quadriparesis with cervical infection. At this stage, findings of upper motor neuron disease are usually predominant; such findings include spasticity, hyperreflexia, and extensor plantar responses. Bilateral sensory loss and impaired bladder and bowel function may occur. Without treatment, paralysis develops, sometimes abruptly, and is usually irreversible.

Laboratory Findings

Leukocytosis is common in patients with epidural abscess, but white cell counts may be normal, especially in patients with protracted symptoms. The erythrocyte sedimentation rate is usually elevated. A lumbar puncture is not ordinarily recommended, except when needed for myelography; however, when a lumbar puncture is performed, the specimen typically demonstrates pleocytosis of up to a few hundred cells, commonly with a mixture of lymphocytes and neutrophils. The glucose level is normal; the protein level is usually elevated, sometimes by more than 1 g/L when the spinal canal is completely blocked. Gram stain and culture of CSF are usually normal. Blood cultures, which are commonly positive in epidural abscess, are worth obtaining in all suspected cases.

Imaging Studies

When the abscess arises from vertebral osteomyelitis, plain films of the spine may demonstrate erosion of the cortical margins of the vertebrae, narrowed or obliterated intervertebral disk spaces, and bone destruction. Myelograms show evidence of blockage of CSF circulation at the level of the epidural abscess but are rarely necessary when CT or MRI scans are available. CT scans detect vertebral osteomyelitis well; characteristic findings

include narrowed intervertebral disk spaces, eroded vertebral end plates, and bony destruction of the vertebral bodies. The epidural abscess is present as a low-density focal epidural mass, which may be better delineated with CT myelography. The best imaging modality, however, is MRI, which discloses a low-intensity or isointense image on T_1-weighted scans (when compared with the spinal cord image) and a high signal on proton-density or T_2-weighted scans. Epidural abscesses typically appear enhanced with gadolinium. Demonstration of the extent of the mass is simple on MRI with both longitudinal and transverse sections, and vertebral osteomyelitis is easily visualized as high-intensity areas on T_2-weighted images.

TREATMENT

Neurosurgery

The most important element of therapy is urgent neurosurgery for removal of pus and granulation tissue—which, rather than purulent material, is sometimes the major or sole finding. Decompressive laminectomies are performed for posterior epidural abscesses, whereas more complicated anterior approaches with bone graft fusion to establish spinal stability may be necessary for ventral collections. Occasionally, patients with absent or minimal neurologic findings respond to antimicrobial therapy alone,[39] but even these patients may benefit from surgery when imaging procedures show significant evidence of spinal cord compression. In such patients, the organism should be identified by blood culture or abscess aspiration, and clinicians should realize that neurologic decline may occur abruptly.[39]

Antimicrobial Therapy

Antimicrobial therapy alone may also be appropriate for patients who are at high surgical risk because of concurrent medical problems, for those who have very extensive spinal involvement, and for patients with paralysis that has lasted more than 48 hours, after which significant neurologic recovery is highly unlikely.

If the infecting organism is unknown, empirical antibiotic therapy should include an agent effective against gram-negative bacilli (e.g., gentamicin or a third-generation cephalosporin) plus an agent active against *S. aureus* (e.g., a penicillinase-resistant penicillin such as nafcillin).

When methicillin-resistant *S. aureus* is a possibility, vancomycin should be used. Results of microbiologic evaluation of surgical specimens or blood cultures can then dictate the appropriate choice. In general, treatment is continued for 3 to 4 weeks in the absence of vertebral osteomyelitis and continued for 6 to 8 weeks when it is present. Patients with tuberculous epidural abscesses should receive antituberculous chemotherapy for at least 6 months.

PROGNOSIS

The prognosis in spinal epidural abscess depends on several factors, the most important of which is the extent of neurologic deficiency at the time of surgery. Poor outcomes are common when paresis or, especially, paralysis is present. Other unfavorable prognostic features are advanced age (> 60 years), substantial thecal sac compression on neuroimaging, and longer duration of symptoms.[40] About 40% of patients recover completely, 25% have residual weakness, 20% have permanent paralysis, and 15% die.

Subdural Empyema

PATHOGENESIS AND PATHOPHYSIOLOGY

A subdural empyema is a collection of pus between the dura mater and the underlying arachnoid. Because this space contains no septations, except where the arachnoid granules lie within the dura, purulent material easily spreads over the surface of the brain. Most cases occur as a complication of sinusitis in older children and young adults, especially males, who account for about 75% of cases.[41] The frontal sinus is nearly always involved, usually along with one or more of the other paranasal sinuses. Another cause is ear infection, either otitis media or mastoiditis; in such cases, the empyema may lie below the tentorium.

In these infections, pathogens probably reach the subdural space via blood draining from the infected sites through the venous sinuses in the dura. Subsequent septic thrombophlebitis of these vessels spreads organisms into the subdural space. Direct extension from contiguous infectious foci through erosion of the bones of the sinus or the ear may also occur. Other possible sources of the organisms causing subdural empyema are bacteremic infection of a preexisting subdural hematoma and direct inoculation of microbes into the area from penetrating head trauma or intracranial surgery.

Once the subdural space is infected, pus may spread widely over the convexities of the ipsilateral cerebrum; it may also travel along the falx separating the two hemispheres. The clinical features arise from pressure exerted on the underlying brain, the systemic effects of infection, and septic thrombophlebitis that spreads to the cortical and subcortical veins, causing brain infarction or infection.

ETIOLOGY

In most cases, only one species of organism is isolated from the pus, but polymicrobial infections can occur. The usual pathogens are aerobic or anaerobic streptococci and other anaerobes, especially with infections arising from sinusitis. Staphylococci and gram-negative bacilli are common causes in postoperative infections. Approximately 20% of cultures are negative, possibly because of inadequate microbiologic techniques (especially in isolating anaerobic bacteria) or previous antimicrobial therapy.

DIAGNOSIS

Clinical Features

Many patients have had a preceding sinusitis or a nonspecific upper respiratory tract illness; in some patients, however, symptoms of subdural empyema arise suddenly. The most common features are fever, stiff neck, and headache, which can be diffuse or localized to the area of empyema or underlying sinusitis. Neurologic abnormalities are usually present, including hemiparesis, seizures, aphasia, hemianesthesia, hemianopsia, and altered mentation, such as confusion, drowsiness, disorientation, and even coma. Papilledema may occur, indicating increased intracranial pressure, which can also cause palsies of the third and sixth cranial nerves. Sinus tenderness and periorbital edema may develop, reflecting the underlying sinusitis. With infratentorial empyema, which usually originates from neglected otic infections, the clinical features include stiff neck, altered consciousness, signs of increased intracranial pressure, otorrhea, and fever.[42] Patients typically deteriorate rapidly.

The presence of fever, headache, nuchal rigidity, and focal neurologic signs, especially in an adolescent boy or young man

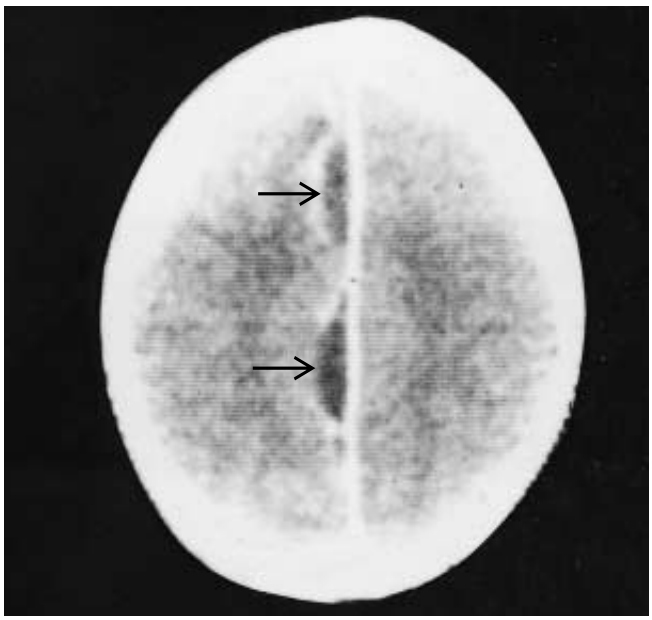


Figure 3 **CT scan of a 17-year-old patient with a subdural empyema shows two loculations (arrows) that are restricted to the interhemispheric area. Contrast-enhanced margins surround the two loculations.**

with sinusitis, should strongly suggest the presence of a subdural empyema and the necessity of immediate neuroimaging with CT or, preferably, MRI.

Laboratory Findings

A peripheral leukocytosis is common. Lumbar puncture is inadvisable because of the risk of brain herniation from increased intracranial pressure, but it is often performed because the clinical features suggest acute bacterial meningitis. Fortunately, serious complications rarely occur. The opening pressure is elevated, and pleocytosis of several hundred white cells or fewer is present. There is commonly a mixture of neutrophils and lymphocytes; either may predominate, but frequently the two cell types are equal in number. The protein level is typically increased, but the glucose level is almost always normal, and both Gram stain and culture are characteristically negative.

Imaging Studies

CT scans may show the empyema as isodense to low-density collections over one hemisphere and in the interhemispheric space, with a contrast-enhanced rim [*see Figure 3*]. MRI is a more sensitive and accurate imaging technique, however, revealing the empyema as isointense masses on T_1-weighted images but showing high signal on proton density or T_2-weighted images. The rim may be enhanced with gadolinium.[41] These imaging techniques may also demonstrate concurrent complications, which are relatively common and include brain abscess, cranial epidural abscess, cortical thrombophlebitis, and venous infarction.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis for subdural empyema includes bacterial meningitis, brain abscess, and encephalitis. Clinical distinction among these infections may be difficult; however, focal findings are unusual in meningitis, and a stiff neck is uncommon with brain abscesses and encephalitis.

TREATMENT

The most important therapy is removal of the subdural pus through use of multiple bur holes or craniotomy. It is unclear which method is better; the choice may depend on the location of the infection and the patient's neurologic status. With either procedure, but especially with bur holes, multiple operations are often necessary because of inadequate drainage of pus. Antibiotic therapy, usually continued for 3 to 4 weeks after surgery, can be dictated by the results of Gram stain and culture of pus. The combination of ceftriaxone and metronidazole is a reasonable empirical regimen to use until the results of the Gram stain and culture are received.

PROGNOSIS

The prognosis depends greatly on the patient's state of consciousness at the time of surgery, with worsening outcomes accompanying decreasing levels of awareness. Overall mortality is approximately 15%. Significant neurologic sequelae, such as speech abnormalities and hemiparesis, occur in approximately 15% to 20% of survivors, and seizures occur in about 30%. These problems may emerge long after the surgery.

Septic Thrombophlebitis of the Major Cerebral Veins

The dural sinuses drain blood from the brain into the jugular veins. These sinuses lack valves, allowing flow in either direction, depending on the prevailing pressure gradient. Septic thrombophlebitis of these vessels, which may result from intracranial suppuration or spread of infection from extracranial veins, causes increased intracranial pressure, focal cerebral edema, or brain infarction. Usually, the initiating site of infection is clinically obvious in the middle ear, mastoid, paranasal sinuses, or facial skin. The cavernous, lateral, and superior sagittal sinuses are the most commonly involved vessels.[43]

CAVERNOUS SINUS THROMBOPHLEBITIS

Cavernous sinus thrombophlebitis most frequently develops from infections in the paranasal sinuses or the facial skin, especially on the medial third near the eyes and nose. Most cases are caused by *S. aureus*.[44] Other causative pathogens are streptococci, gram-negative bacilli, and anaerobes.

The earliest symptom is usually headache, which often precedes fever and other findings by several days. The pain typically involves areas innervated by the ophthalmic and maxillary divisions of the trigeminal nerve, which traverse the cavernous sinus. Hyperesthesia or decreased sensation may be demonstrable in the dermatomes served by these nerves. Other focal findings may arise from obstructed ophthalmic veins: unilateral chemosis, proptosis, and edema of the ipsilateral eyelids, nose, and forehead. With time, the findings become bilateral as the phlebitis extends to the opposite cavernous sinus. Compression of cranial nerves III, IV, and VI, which course through the cavernous sinus, leads to varying degrees of ophthalmoplegia. The pupils may dilate and fail to react to light. Retinal veins engorge, visual acuity may lessen, and hemorrhages and papilledema can occur. Lethargy and coma may supervene, reflecting increased intracranial pressure and neuronal damage. Occasionally, bacteremia with metastatic foci of infection develops.

LATERAL SINUS THROMBOPHLEBITIS

Lateral sinus thrombophlebitis is almost always a complication of mastoiditis. Anaerobes, staphylococci, and gram-negative

bacilli, especially *Proteus* organisms, are the most common isolates.

The clinical course is usually subacute, with symptoms lasting for several weeks. Earache and drainage are often the first symptoms. Persistent, severe unilateral headache, followed by nausea and vomiting, indicates the development of lateral sinus thrombophlebitis. Fever, chills, and manifestations of increased intracranial pressure, including confusion and papilledema, are commonly present. Postauricular tenderness, venous engorgement, and edema represent involvement of the mastoid emissary vein. Extension into the jugular vein may occur.

Otoscopic examination commonly reveals a perforated or inflamed tympanic membrane. Focal neurologic findings are usually absent except for unilateral sixth nerve palsy, reflecting compression of the inferior petrosal sinus.

SUPERIOR SAGITTAL SINUS THROMBOPHLEBITIS

Superior sagittal sinus thrombophlebitis complicates bacterial meningitis, paranasal sinusitis, contiguous osteomyelitis, and dural infection. The most common organisms are *S. pneumoniae*, other streptococci, and gram-negative bacilli. Thrombosis of only the anterior portion of the sinus is often asymptomatic, but posterior involvement causes increased intracranial pressure. The predominant feature is acute headache, often with nausea and vomiting, followed in a few days by confusion and eventually coma. Focal or generalized seizures are common, and most patients are febrile. Extension of the thrombophlebitis to cortical veins may cause infarction of the underlying cortex, producing such neurologic manifestations as focal seizures and hemiparesis.

DIAGNOSIS

With septic thrombophlebitis in any of the cerebral veins, examination of the CSF may yield normal results or may demonstrate findings consistent with a parameningeal focus of infection: increased protein level, normal glucose level, and a pleocytosis of mixed neutrophils and lymphocytes that rarely exceeds a few hundred cells. Increased opening pressure is also common in all types.

Plain films of the skull are rarely helpful except in cases of septic lateral sinus thrombosis, in which they nearly always reveal mastoiditis. Certain findings on CT scan suggest cerebral venous thrombosis, but MRI provides a much better study. Acutely thrombosed cerebral veins lack a flow void. They are isointense on T_1-weighted images and hypointense on T_2-weighted images. In subacute thrombosis, the methemoglobin in the veins produces areas of high intensity on both images.

TREATMENT

Management of cerebral vein thrombophlebitis includes appropriate antibiotic therapy. For each of these disorders, when the infecting organisms are unknown, a good empirical antibiotic regimen is a combination of ceftriaxone and metronidazole. With cavernous sinus thrombosis in which *S. aureus* is prominent, vancomycin may be added if methicillin-resistant strains are likely.

Occasionally, treatment of cerebral vein thrombophlebitis includes surgery for the areas of purulence from which the infection arose. When paranasal sinusitis is the underlying cause, drainage of the infected sinuses may be necessary to obtain a good clinical response. In lateral sinus thrombophlebitis, radical mastoidectomy and exploration over the lateral sinus are commonly performed to remove purulent material.

Other medical treatment may include mannitol and dexamethasone in superior sagittal thrombophlebitis to help relieve the severe cerebral edema that is often present. The role of anticoagulation in cerebral vein thrombophlebitis remains unsettled; it seems valuable in cavernous sinus thrombosis[45] but appears potentially hazardous in superior sagittal and lateral sinus thrombophlebitis because of the frequent concurrent hemorrhagic venous cortical infarcts that it may exacerbate.

PROGNOSIS

Septic thrombophlebitis of the superior sagittal and cavernous sinuses imposes a high mortality, even with appropriate therapy; however, most patients with involvement of the lateral sinus have a good prognosis.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Schuchat A, Robinson K, Wenger JD, et al: Bacterial meningitis in the United States in 1995. N Engl J Med 337:970, 1997

2. Durand ML, Calderwood SB, Weber DJ, et al: Acute bacterial meningitis in adults: a review of 493 episodes. N Engl J Med 328:21, 1993

3. Almirante B, Saballs M, Ribera E, et al: Favorable prognosis of purulent meningitis in patients infected with human immunodeficiency virus. Clin Infect Dis 27:176, 1998

4. van de Beek D, de Gans J, Spanjaard L, et al: Clinical features and prognostic factors in adults with bacterial meningitis. N Engl J Med 351:1849, 2004 (erratum in N Engl J Med 352:950, 2005)

5. Sigurdardottir B, Bjornsson OM, Jonsdottir KE, et al: Acute bacterial meningitis in adults: a 20-year overview. Arch Intern Med 157:425, 1997

6. Gorse GJ, Thrupp LD, Nudleman KL, et al: Bacterial meningitis in the elderly. Arch Intern Med 144:1603, 1984

7. Verghese A, Gallemore G: Kernig's and Brudzinski's signs revisited. Rev Infect Dis 9:1187, 1987

8. Rennick G, Shann F, de Campo J: Cerebral herniation during bacterial meningitis in children. BMJ 306:953, 1993

9. Tunkel AR, Hartman BJ, Kaplan SL, et al: Practice guidelines for bacterial meningitis. Clin Infect Dis 39:1267, 2004

10. Powers WJ: Cerebrospinal fluid to serum glucose ratios in diabetes mellitus and bacterial meningitis. Am J Med 71:217, 1981

11. Powers WJ: Cerebrospinal fluid lymphocytosis in acute bacterial meningitis. Am J Med 79:216, 1985

12. Mylonakis E, Hohmann EL, Calderwood SB: Central nervous system infection with *Listeria monocytogenes*. Medicine (Baltimore) 77:313, 1998

13. Talan DA, Hoffman JR, Yoshikawa TT, et al: Role of empiric parenteral antibiotics prior to lumbar puncture in suspected bacterial meningitis: state of the art. Rev Infect Dis 10: 365, 1988

14. Proulx N, Frechette D, Toye B, et al: Delays in the administration of antibiotics are associated with mortality from adult acute bacterial meningitis. QJM 98:291, 2005

15. Cabellos C, Viladrich PF, Verdaguer R, et al: A single daily dose of ceftriaxone for bacterial meningitis in adults: experience with 84 patients and review of the literature. Clin Infect Dis 20:1164, 1995

16. Meyer CN, Samuelsson IS, Galle M, et al: Adult bacterial meningitis: aetiology, penicillin susceptibility, risk factors, prognostic factors and guidelines for empirical antibiotic treatment. Clin Microbiol Infect 10:709, 2004

17. Briggs S, Ellis-Pegler R, Roberts S, et al: Short course intravenous benzylpenicillin treatment of adults with meningococcal disease. Intern Med J 34:383, 2004

18. de Gans J, van de Beek D, European Dexamethasone in Adulthood Bacterial Meningitis Study investigators: Dexamethasone in adults with bacterial meningitis. N Engl J Med 347:1549, 2002

19. van de Beek D, de Gans J, McIntyre P, et al: Steroids in adults with acute bacterial meningitis: a systematic review. Lancet Infect Dis 4:139, 2004

20. Pfister HW, Feiden W, Einhaupl KM: Spectrum of complications during bacterial meningitis in adults: results of a prospective clinical study. Arch Neurol 50:575, 1993

21. Pfister HW, Borasio GD, Dirnagl U, et al: Cerebrovascular complications of bacterial meningitis in adults. Neurology 42:1497, 1992

22. Aronin SI, Peduzzi P, Quagliarello VJ: Community-acquired bacterial meningitis: risk stratification for adverse clinical outcome and effect of antibiotic timing. Ann Intern Med 129:862, 1998

23. Kastenbauer S, Pfister HW: Pneumococcal meningitis in adults: spectrum of complications and prognostic factors in a series of 87 cases. Brain 126:1015, 2003

24. Hirschmann JV: Bacterial meningitis following closed cranial trauma. Bacterial Meningitis. Sande MA, Smith AL, Root RK, Eds. Churchill Livingstone, New York, 1985, p 95

25. Walsh M, Curran AJ: Cerebrospinal fluid rhinorrhoea. Br J Neurosurg 11:189, 1997

26. Eftekhar B, Ghodsi M, Nejat F, et al: Prophylactic administration of ceftriaxone for the prevention of meningitis after traumatic pneumocephalus: results of a clinical trial. J Neurosurg 101:757, 2004

27. Morris A, Low DE: Nosocomial bacterial meningitis, including central nervous system shunt infections. Infect Dis Clin North Am 13:735, 1999

28. Metersky ML, Williams A, Rafanan AL: Retrospective analysis: are fever and altered mental status indications for lumbar puncture in a hospitalized patient who has not undergone neurosurgery? Clin Infect Dis 25:285, 1997

29. Forgacs P, Geyer CA, Friedberg SR: Characterization of chemical meningitis after neurological surgery. Clin Infect Dis 32:179, 2001

30. Kaufman BA, Tunkel AR, Pryor JC, et al: Meningitis in the neurosurgical patient. Infect Dis Clin North Am 4:677, 1990

31. Mathisen GE, Johnson JP: Brain abscess. Clin Infect Dis 25:763, 1997

32. Lu CH, Chang WN, Lin YC, et al: Bacterial brain abscess: microbiological features, epidemiological trends and therapeutic outcomes. QJM 95:501, 2002

33. Kao PT, Tseng HK, Liu CP, et al: Brain abscess: clinical analysis of 53 cases. J Microbiol Immunol Infect 36:129, 2003

34. Boviatsis EJ, Kouyialis AT, Stranjalis G, et al: CT-guided stereotactic aspiration of brain abscesses. Neurosurg Rev 26:206, 2003

35. Jansson AK, Enblad P, Sjolin J: Efficacy and safety of cefotaxime in combination with metronidazole for empirical treatment of brain abscess in clinical practice: a retrospective study of 66 consecutive cases. Eur J Clin Microbiol Infect Dis 23:7, 2004

36. Xiao F, Tseng MY, Teng LJ, et al: Brain abscess: clinical experience and analysis of prognostic factors. Surg Neurol 63:442, 2005

37. Darouiche RO, Hamill RJ, Greenberg SB, et al: Bacterial spinal epidural abscess: review of 43 cases and literature survey. Medicine (Baltimore) 71:369, 1992

38. Obrador GT, Levenson DJ: Spinal epidural abscess in hemodialysis patients: report of three cases and review of the literature. Am J Kidney Dis 27:75, 1996

39. Curry WT Jr, Hoh BL, Amin-Hanjani S, et al: Spinal epidural abscess: clinical presentation, management, and outcome. Surg Neurol 63:364, 2005

40. Khanna RK, Malik GM, Rock JP, et al: Spinal epidural abscess: evaluation of factors influencing outcome. Neurosurgery 39:958, 1996

41. Greenlee JE: Subdural empyema. Curr Treat Options Neurol 5:13, 2003

42. Nathoo N, Nadvi SS, van Dellen JR: Infratentorial empyema: analysis of 22 cases. Neurosurgery 41:1263, 1997

43. Southwick FS, Richardson EP Jr, Swarz MN: Septic thrombosis of the dural venous sinuses. Medicine (Baltimore) 65:82, 1986

44. Ebright JR, Pace MT, Niazi AF: Septic thrombosis of the cavernous sinuses. Arch Intern Med 161:2671, 2001

45. Bhatia K, Jones NS: Septic cavernous sinus thrombosis secondary to sinusitis: are anticoagulants indicated? A review of the literature. J Laryngol Otol 116:667, 2002

XXXVII MYCOTIC INFECTIONS

Carol A. Kauffman, m.d.

Overview of Endemic Mycosis

The endemic mycoses (histoplasmosis, blastomycosis, coccidioidomycosis, and sporotrichosis) are caused by fungi that share several important characteristics but differ greatly in other respects. The fungi are all dimorphic, existing as molds in the environment and as either yeasts or spherules in tissues. Each organism occupies a different ecological niche; infection is directly related to exposure to the mycelial or mold phase of the organism in the environment. These fungi are true pathogens in that they are capable of causing infection in otherwise healthy individuals. The severity of infection is determined both by the extent of the exposure to the organism and by the immune status of the patient. Three of the four organisms (*Histoplasma capsulatum*, *Blastomyces dermatitidis*, and *Coccidioides immitis*) are inhaled, have the propensity to disseminate hematogenously, and are able to reactivate years later; the fourth organism, *Sporothrix schenckii*, usually causes primary infection after cutaneous inoculation and does not reactivate. Although many of the disease manifestations overlap, each of the four has its own distinctive characteristics.

The endemic mycoses mimic many common infections involving lungs, skin, and other organs, and frequently, the diagnosis of fungal infection is not entertained; this is especially true for patients who present outside of the areas endemic for these specific mycoses. Because of an increase in travel and leisure activities over the past several decades, patients often present with illness outside endemic areas. When faced with an elderly retiree who spends winters in Arizona and who has developed fever, headache, and visual complaints, a physician who practices in Ohio and who may have never seen coccidioidomycosis must remember that this fungal infection causes chronic meningitis. Similarly, an AIDS patient with fever, hepatosplenomegaly, and pancytopenia who lives in Seattle but who spent his childhood in Arkansas and has not returned there for the past 20 years may well be experiencing reactivation of disseminated histoplasmosis.

The most useful diagnostic tests differ for each of the endemic mycoses, but as a general rule, histopathologic demonstration of the fungi in biopsy specimens is the most expeditious way to make a diagnosis, especially in a severely ill patient; growth of the organism in vitro is definitive. For most patients with an endemic mycosis, treatment will be with an azole antifungal agent. For those who are severely ill, initial therapy with amphotericin B followed by consolidation therapy with an azole is standard.

A review of antifungal agents and guidelines for their use in the treatment of mycotic infections is found after the descriptions of specific manifestations of each endemic mycosis [*see* Antifungal Therapy, *below*].

Histoplasmosis

epidemiology

H. capsulatum is a dimorphic fungus that every year infects hundreds of thousands of individuals in the United States. The organism is endemic in the Mississippi and Ohio River valleys and Central America. Soil containing high concentrations of bird or bat guano supports the profuse growth of the mycelial phase of *H. capsulatum*. Exposure typically occurs as a result of activities that generate aerosols containing the organism. Although most cases are sporadic and the exact source of exposure is unknown, many point-source outbreaks have been well described in association with disruption of soil; the cleaning of attics, bridges, or barns; tearing down old structures laden with guano; and spelunking.[1-3] Massive outbreaks of infection with *H. capsulatum* occurred when urban demolition exposed hundreds of thousands of people to the organism.[1] Infection is very common in the endemic area, most persons having been infected before adulthood.

pathogenesis

During the mycelial phase of the organism's life cycle, the microconidia are inhaled into the alveoli, causing a localized pulmonary infection. Neutrophils and macrophages phagocytize the organism, which converts to the yeast phase; the organism is then able to survive and travel within the macrophage. Spread to the hilar and mediastinal lymph nodes ensues, and hematogenous dissemination subsequently occurs throughout the reticuloendothelial system before specific immunity has developed. After several weeks, cell-mediated immunity that is specific for *H. capsulatum* activates macrophages, which then kill the organism.[4] Of all the human mycoses, histoplasmosis exemplifies best the pivotal importance of the cell-mediated immune system. The corollary of this observation is that most patients with severe infection are those with cellular immune deficiencies.

The extent of disease is determined by the number of conidia that are inhaled and the immune response of the host. A healthy individual may develop severe life-threatening pulmonary infection if a large number of conidia are inhaled. This might occur during demolition or renovation of an old building or as a result of spelunking in a heavily infested cave. Conversely, a small inoculum can cause severe pulmonary infection or progress to acute symptomatic disseminated histoplasmosis in a patient with advanced HIV infection whose cell-mediated immune system is unable to contain the organism.

Most persons who have been infected experience asymptomatic dissemination; only rarely will this lead to symptomatic acute or chronic disseminated histoplasmosis. However, by virtue of this dissemination, latent infection can persist for a lifetime. Reactivation of quiescent infection can occur years later if immunosuppression occurs.[5] Reinfection has also been documented, though rarely, in persons previously known to have had histoplasmosis. Such reinfection almost always occurs in the setting of exposure to a heavy inoculum of *H. capsulatum* conidia.

clinical presentations

Pulmonary Histoplasmosis

Infection is asymptomatic in the vast majority of persons who have been infected with *H. capsulatum*. Most patients who have symptomatic pulmonary infection will have a self-limited illness characterized by fever, chills, and cough that is usually nonproductive; such infection is often associated with anterior chest discomfort, myalgias, arthralgias, and fatigue. A patchy lobar or multilobar nodular infiltrate is noted on chest radiography.[6] The diagnosis is usually not made in individual cases until the patient fails to respond to several courses of antibiotics given for atypical pneumonia.

The diagnosis of acute pulmonary histoplasmosis is more easily made when the patient has been involved in an outbreak.[2] A careful history of the patient's activities as well as the patient's colleagues' activities—especially if the history reveals participation in outdoor activities, activities around a demolition site, or spelunking several weeks before the onset of symptoms—may point to histoplasmosis. High spiking fever, prostration, dyspnea, and cough are prominent in severe cases in which the number of conidia inhaled is large. Diffuse nodular infiltrates are noted on the chest radiograph, and development of the acute respiratory distress syndrome (ARDS) can ensue.

In patients with cell-mediated immune deficiencies, including those who have advanced HIV infection, those with a hematologic malignancy, those who have received a transplant, and those who are receiving immunosuppressive medications, pulmonary infection is more severe than in otherwise healthy patients. Although pulmonary infection may be the only manifestation of histoplasmosis, in most immunosuppressed patients, pulmonary involvement is merely one component of widespread dissemination.[7] Prostration, fever, chills, marked dyspnea, and hypoxemia are prominent, and chest radiographs show diffuse infiltrates.

Chronic cavitary pulmonary histoplasmosis is a progressive, fatal form of histoplasmosis that usually develops in older patients who have chronic obstructive pulmonary disease (COPD). Typical symptoms include fever, fatigue, anorexia, weight loss, cough that is productive of purulent sputum, and hemoptysis.[8,9] On chest radiography, infiltrates can be either unilateral or bilateral and are almost always located in the upper lobes. Cavities may be multiple and are frequently quite large; extensive fibrosis occurs in the lower lobes.[6]

Disseminated Histoplasmosis

Hematogenous dissemination typically occurs in immunosuppressed patients and young children early in the course of infection with *H. capsulatum*. Patients who have HIV infection and whose $CD4^+$ T cell counts are less than 150 cells / mm^3 are at great risk for acquiring histoplasmosis, which almost always presents as disseminated infection. Symptoms and signs of acute disseminated histoplasmosis include chills, fever, malaise, anorexia, weight loss, dyspnea, hepatosplenomegaly, and skin and mucous membrane lesions. Pancytopenia, diffuse pulmonary infiltrates seen on chest radiographs, and blood cultures that yield *H. capsulatum* are common.[7,10] Adrenal insufficiency may also be present.

A chronic progressive disseminated form of histoplasmosis that occurs mostly in middle-aged to elderly men who have no known immunosuppressive illness is characterized by fever, night sweats, weight loss, and fatigue.[11] Patients appear chronically ill and frequently have hepatosplenomegaly, mucocutaneous ulcerations, and signs of adrenal insufficiency. Typical findings include an increase in the erythrocyte sedimentation rate, an elevation in the alkaline phosphatase level, pancytopenia, and diffuse pulmonary infiltrates seen on chest radiograph. If not diagnosed and treated appropriately, this form of histoplasmosis is fatal.

diagnosis

Culture

The definitive diagnostic test for histoplasmosis is growth of *H. capsulatum* in culture, which, unfortunately, may take as long as 6 weeks.[5] Sputum, bronchoalveolar lavage fluid, or tissue biopsy material should be sent to the laboratory for culture. For those patients who have evidence of dissemination, cultures of blood are useful. The best yield is with the lysis-centrifugation system (Isolator aerobic blood culture system, Wampole Laboratories). The laboratory should be informed that pulmonary histoplasmosis is a diagnostic consideration, so that a special medium that decreases the growth of commensal fungi, such as *Candida*, can be used for the culture of pulmonary samples. As soon as growth of a mold has been detected, a DNA probe specific for *H. capsulatum* confirms the identity of the organism.

Serologic Studies

In contrast to several other endemic mycoses, serology plays an important adjunctive role in the diagnosis of histoplasmosis.[12] Both complement fixation (CF) and immunodiffusion (ID) tests are available. ID is more specific than CF (> 95% specificity versus 85% to 90%); both are only modestly sensitive (75% to 85% sensitivity). False negative serologic test results occur in immunosuppressed patients who cannot mount an antibody response. However, most patients with pulmonary histoplasmosis are not immunosuppressed, and thus, in this group, antibody tests are helpful.

Patients with chronic cavitary pulmonary histoplasmosis almost always have an elevated CF antibody titer (> 1:32), and either an M precipitin band or both H and M precipitin bands are seen on ID in these patients. In a patient with acute pneumonia, a documented fourfold increase in CF titer or the appearance of an M precipitin band on ID establishes the diagnosis of histoplasmosis. Serologic tests are less definitive in patients with mediastinal lymphadenopathy and should always be confirmed by tissue biopsy. False positive test results, especially false positive results on CF, occur in patients with lymphoma, tuberculosis, sarcoidosis, and other fungal infections, all of which may present as mediastinal masses. In most patients with the chronic disseminated form of histoplasmosis, both precipitin and CF antibodies are present.

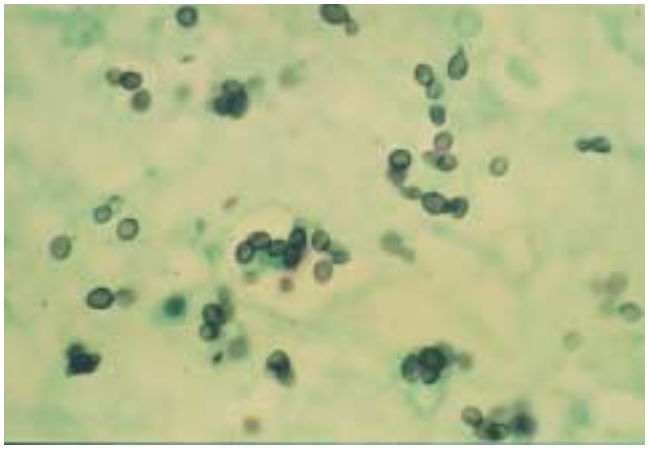
Antigen Tests

Enzyme immunoassays for the detection of *H. capsulatum* polysaccharide antigen in urine and serum are extremely helpful in AIDS patients who have disseminated infection and a large fungal burden.[13,14] They are less useful for patients with pulmonary histoplasmosis; antigen is detected in only about 20% of patients with pulmonary histoplasmosis. Cross-reactivity occurs in a small number of patients with blastomycosis, paracoccidioidomycosis, and penicilliosis.

Tissue Biopsy

For an acutely ill person, evidence of *H. capsulatum* on tissue biopsy allows the physician to make the diagnosis of histoplasmosis in a timely fashion.[5] The organisms appear as distinctive, fairly uniformly shaped, oval budding yeasts 2 to 4 μm in diameter. Tissue biopsy is most useful for the diagnosis of disseminated disease; bone marrow, liver, and mucocutaneous lesions reveal many organisms [*see Figure 1*]. For most patients with acute pulmonary histoplasmosis, biopsy is not indicated unless the patient is severely ill. For patients with chronic pulmonary histoplasmosis or granulomatous mediastinitis, biopsy of lung or lymph nodes may reveal the organism. Routine hematoxylin-eosin staining will not show the tiny *H. capsulatum* yeasts; biopsy material must be stained with methenamine-silver stain, periodic acid–Schiff stain, or both. In contrast to larger yeasts, such as *B. dermatitidis*, it is extremely unusual to find *H. capsulatum* on cytologic examination of sputum or bronchoalveolar lavage fluid.

a



b

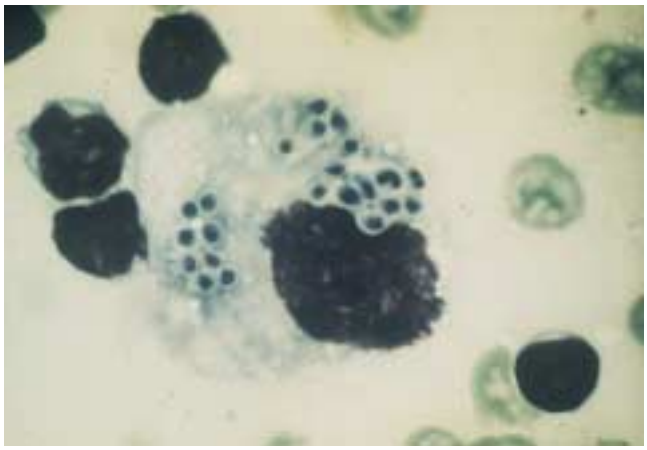


Figure 1 **Shown are biopsy samples from an elderly man with chronic progressive disseminated histoplasmosis. (*a*) Tongue ulcer stained with methenamine-silver stain shows several budding yeast forms measuring 2 to 4 μm. (*b*) A smear from a lung biopsy sample, which is stained with Giemsa stain, shows small intracellular yeasts inside a macrophage.**

Skin Tests

Skin tests with histoplasmin antigen are not useful for the diagnosis of histoplasmosis in individual patients. In the endemic area, as many as 85% of adults have positive results on skin testing because of previous exposure to the organism; cross-reactions occur in those with other fungal infections, especially blastomycosis; and negative test results occur frequently in patients who have severe infection.

differential diagnosis

Acute pulmonary histoplasmosis is most difficult to differentiate from acute pulmonary blastomycosis. The clinical and radiographic findings of these two diseases are similar, the regions in which these diseases are endemic overlap, and the CF antibody tests for the two organisms show cross-reactivity. For either disease, culture of the causative organism from sputum is diagnostically definitive. However, results of sputum cultures are often negative for patients with acute pneumonia caused by either fungus. Treatment, fortunately, is the same for the two diseases. Atypical pneumonias caused by *Mycoplasma*, *Legionella*, and *Chlamydia* are included in the differential diagnosis of acute pulmonary histoplasmosis. Hilar and mediastinal lymphadenopathy, which is very common with histoplasmosis, is uncommon in patients with atypical pneumonia caused by these other organisms.

Chronic pulmonary histoplasmosis mimics tuberculosis in regard to symptoms, signs, and radiographic findings. Other chronic fungal pneumonias, especially blastomycosis and sporotrichosis, and nontuberculous mycobacterial infections also must be differentiated from this form of histoplasmosis. Sputum culture and serology are most useful in differentiating these infections.

Patients with acute disseminated histoplasmosis can present with a sepsis syndrome associated with ARDS and disseminated intravascular coagulation that is indistinguishable from sepsis of any bacterial or viral etiology. Helpful findings supporting the diagnosis of histoplasmosis are pancytopenia, diffuse nodular pulmonary infiltrates, and hepatosplenomegaly; the hematology laboratory may make the diagnosis when they find small yeasts inside white cells on the peripheral smear. In AIDS patients, the main infections that must be excluded are those caused by cytomegalovirus, *Mycobacterium avium* complex, and *M. tuberculosis*. Biopsy of involved tissues should be performed as soon as possible to help differentiate between these conditions. Culture of blood by means of lysis-centrifugation (Isolator system) for fungus and *Mycobacteria* and use of urinary antigen testing for *H. capsulatum* are very helpful diagnostic tests in this circumstance.

Patients with chronic disseminated histoplasmosis usually present with fever of unknown origin. The disease that mimics histoplasmosis most closely is miliary tuberculosis; lymphoma, brucellosis, and sarcoidosis also must be excluded. Serology is helpful, but histopathologic evidence of yeasts in tissue granulomas and confirmatory culture of the organism from bone marrow, mucous membrane lesions, or liver are definitive. It should be emphasized that the diagnosis of sarcoidosis requires firm evidence that the patient does not have histoplasmosis or tuberculosis; corticosteroid use for the treatment of suspected sarcoidosis in patients who in fact have histoplasmosis frequently leads to defervescence for a short period, followed by further worsening of the infection.[15]

treatment

Guidelines for the treatment of histoplasmosis have recently been published by the Mycoses Study Group and the Infectious Diseases Society of America.[16] Efficacy has been defined primarily by open-label trials in patients with and without AIDS and through anecdotal experience[17-22]; there are no studies that directly compare azoles with amphotericin B. The only blinded, randomized treatment trial studied liposomal amphotericin B versus amphotericin B deoxycholate in a very restricted patient population: AIDS patients with severe disseminated histoplasmosis.[23] In spite of the lack of controlled treatment trials, it is clear that for the treatment of most patients with histoplasmosis, itraconazole is the drug of choice.[18-20] Fluconazole should be considered a second-line agent; primary response rates are lower for fluconazole in patients with and without AIDS, and relapse rates for AIDS patients receiving fluconazole are higher than those noted in previous studies with itraconazole.[21,22] If a patient does not tolerate itraconazole, fluconazole can be used, but the dosage should be 400 to 800 mg daily. Side effects, drug absorption, and drug-drug interactions are important considerations in antifungal therapy,

particularly in long-term regimens such as those frequently required for AIDS patients [*see* Antifungal Therapy, *below*].

Pulmonary Histoplasmosis

Treatment recommendations are based on the type of histoplasmosis and the immune status of the host.[16] Treatment is generally not recommended for patients with acute pulmonary histoplasmosis; in fact, the diagnosis is often not made until after the patient's symptoms have resolved. However, if the patient remains symptomatic after 4 weeks, therapy with itraconazole, 200 mg daily for 6 to 12 weeks, is recommended. Children with acute pulmonary histoplasmosis should be treated with itraconazole, 5 to 7 mg / kg / day.

Although pneumonia resolves without treatment in most patients with acute outbreak-related histoplasmosis, those with severe infection should receive treatment. Immunosuppressed patients with acute pulmonary histoplasmosis should always be treated. Initial therapy should be administration of amphotericin B, 0.7 mg / kg / day. After a favorable response is noted, which usually occurs quite soon after initiation of therapy, therapy can be changed to oral itraconazole. Treatment should continue until the infiltrate has resolved.

Antifungal therapy is required for all patients with chronic pulmonary histoplasmosis.[8,16] Itraconazole, 200 mg once or twice daily for 12 to 24 months, is the treatment of choice. Even with appropriate antifungal therapy, outcome is poor.

Disseminated Histoplasmosis

All patients with symptomatic disseminated histoplasmosis should receive antifungal therapy.[11,16] Patients with acute disseminated disease who have only mild to moderate symptoms and most patients with chronic progressive disseminated histoplasmosis should be treated with itraconazole, 200 mg twice daily; this recommendation applies both to patients with AIDS and to those who do not have AIDS. For patients who do not have AIDS, a total of 12 months of therapy is usually adequate, but the length of therapy will be determined by the patient's clinical course; this is especially true for patients with chronic progressive disease. Patients who have AIDS should initially receive itraconazole twice daily for 12 weeks; after that period, they should receive a maintenance course of long-term suppressive therapy with itraconazole, 200 mg daily.[20]

Young infants, who frequently have overwhelming infection, and immunosuppressed patients with moderately severe to severe symptoms should be treated initially with amphotericin B, 0.7 to 1.0 mg / kg / day. For most adults, therapy can be changed to itraconazole after they become afebrile and are able to take oral medications. Infants should receive a full course of amphotericin B because experience with itraconazole is limited in this population group. When therapy is with amphotericin B deoxycholate only, the total dose should be 35 mg / kg.[16]

In a blinded, randomized clinical trial, liposomal amphotericin B, 3 mg / kg / day, was found to be superior to amphotericin B deoxycholate, 0.7 mg / kg / day, for initial treatment of AIDS patients who had severe disseminated histoplasmosis. In this study, the 51 patients receiving liposomal amphotericin B became afebrile sooner than the 22 receiving amphotericin B deoxycholate (P = 0.01); in addition, mortality was decreased and toxicity was lessened in the liposomal amphotericin B arm.[23] However, it is not clear whether these results should affect clinical decision making regarding the treatment of disseminated histoplasmosis in patients who do not have AIDS. The main drawback to the use of the liposomal formulation of amphotericin B is that it is extremely costly in comparison with the standard formulation.

complications

Granulomatous mediastinitis is an uncommon complication of pulmonary histoplasmosis in which lymphadenopathy persists for months.[5] The enlarged nodes can erode into the esophagus; can lead to the formation of tracheoesophageal fistulas; and can cause traction diverticula. Chest radiography and CT scanning reveal enlarged hilar and mediastinal lymph nodes; necrosis and calcium deposition in the nodes are commonly noted on CT scan. Mediastinoscopy and biopsy reveal caseous material, which may contain a few yeastlike organisms typical of *H. capsulatum*. In most patients, the disease follows a self-limited, although protracted, course, and patients do not develop fibrosing mediastinitis. Patients may benefit from itraconazole, 200 mg twice daily, but there are no clinical trials proving efficacy.[16]

Fibrosing mediastinitis is a rare complication of pulmonary histoplasmosis in which an excessive amount of fibrosis develops in the mediastinal lymph nodes in response to infection.[24,25] The exuberant fibrous tissue entraps the great vessels, causing heart failure, pulmonary emboli, and superior vena cava syndrome. Odynophagia, dyspnea, cough, wheezing, and hemoptysis are commonly noted. There is no effective therapy for this progressive disease.

Pericarditis is a self-limited illness that occurs as an inflammatory response to acute pulmonary histoplasmosis.[5] Pleural effusions and mediastinal lymphadenopathy are commonly noted with pericarditis. The pericardial fluid is exudative. Drainage is required if tamponade occurs, but this is rare. Pericarditis should be treated with nonsteroidal anti-inflammatory drugs; corticosteroids should be used for severe cases. Antifungal agents should not be used.[16] The long-term outcome is excellent; only rarely has constrictive pericarditis occurred.

prognosis

Most patients with histoplasmosis respond quite well to antifungal agents. Even patients with advanced AIDS respond quickly and completely to therapy and continue to do well as long as they receive suppressive therapy with an azole. Patients with chronic progressive disseminated histoplasmosis usually experience a return to their baseline function, but it often takes months for this to occur. The patients who do poorly are those whose underlying illness precludes a return to normal function. For example, patients with chronic cavitary pulmonary histoplasmosis frequently have progressive respiratory insufficiency because of their severe underlying emphysema and the fibrosis caused by the fungal infection.

Blastomycosis

epidemiology

Although the dimorphic fungus *B. dermatitidis* is present in many diverse geographic areas worldwide, most cases of blastomycosis are reported from the south central and north central United States.[26] The endemic area for blastomycosis overlaps that for histoplasmosis but extends further north into Wisconsin, Minnesota, and the southern portions of the Canadian provinces of Manitoba, Saskatchewan, and Alberta. Most cases are reported from Arkansas, Mississippi, and Kentucky.

The ecology of *B. dermatitidis* is not as well understood as that of the other endemic mycoses. It is assumed that the natural niche is soil, but the yield of soil cultures is quite low. Most cases occur sporadically, but several well-described outbreaks have helped define the presumed natural habitat.[27,28] The largest outbreak to date involved 95 students in northern Wisconsin who camped along a beaver pond and explored a beaver lodge.[27] Decomposed wood on the pond bank and samples from the lodge yielded *B. dermatitidis.*

Most, but certainly not all, patients who develop blastomycosis are men who have an outdoor occupation or hobby. Cases have been described in hunters and their dogs; in these cases, it is presumed that the hunters and their dogs were exposed to the same environmental source. For most sporadic cases, the source of exposure to the fungus is unknown.

pathogenesis

Most patients with blastomycosis have no underlying illnesses, but the disease is more severe in immunosuppressed patients.[29] Blastomycosis is acquired by inhaling the conidia of *B. dermatitidis* into the alveoli. The organisms change to the yeast form in the lungs and then reproduce by budding. The immune response to *B. dermatitidis* appears to include both neutrophils and cell-mediated immune mechanisms involving T cells and macrophages.

Even though a major clinical manifestation of blastomycosis is cutaneous lesions, blastomycosis is only very rarely spread through inoculation. Examples of such cases include accidental inoculation in laboratory workers, conjugal inoculation from a genital lesion, and inoculation from the bite of an infected dog. The organism travels to the skin and other common target organs, including osteoarticular structures and the genitourinary tract, by the hematogenous route. Most persons infected with *B. dermatitidis* are asymptomatic or have mild pulmonary symptoms. Hematogenous dissemination occurs without clinical manifestations, and only later do patients present with skin lesions in the absence of other obvious organ involvement.

clinical presentations

Pulmonary Blastomycosis

Blastomycosis begins as a pulmonary infection, but many patients never develop symptoms, and others have an acute illness with fever, cough, and a pulmonary infiltrate that is diagnosed as an atypical pneumonia. Overwhelming pulmonary disease with ARDS can occur; this appears to be more common in older adults.[30] Whether severe infection occurs because the patient has been exposed to a large number of conidia or has a defect in his or her immune response has not been clarified.

A more common presentation is a subacute to chronic pulmonary infection that is clinically and radiographically similar to tuberculosis. Fever, night sweats, weight loss, and fatigue are common symptoms. Cough, sputum production, hemoptysis, and dyspnea are noted. The lesions may be cavitary, nodular, fibrotic, or masslike in appearance.[31,32] Hilar and mediastinal lymphadenopathy and pleural effusions are uncommonly seen.

Disseminated Blastomycosis

Many patients with blastomycosis have no pulmonary symptoms.[26,31] They present with cutaneous lesions that are well circumscribed, nonpainful papules, nodules, or plaques that often become verrucous and develop punctate drainage areas in the center. The lesions are common on the face and extremities but can appear anywhere. There may be only a single lesion, or there may be multiple lesions. An uncommon manifestation seen mostly in immunocompromised patients is the appearance of hundreds of acute pustular lesions associated with widespread visceral involvement and an acute downhill course.

Additional manifestations of disseminated blastomycosis include prostatitis, septic arthritis, osteomyelitis, laryngeal and oropharyngeal nodules, and, less commonly, central nervous system involvement. Genitourinary involvement may be asymptomatic, or it may be associated with signs of prostatism. Although osteoarticular infection can be associated with a contiguous skin lesion, bone involvement often occurs at sites distant from the cutaneous lesions. It is helpful to obtain a bone scan for all patients with disseminated blastomycosis because of the propensity of the organism to seed into bone and because osteomyelitis requires prolonged antifungal therapy. All patients with manifestations of dissemination, even patients with only one skin lesion, need to be treated with systemic antifungal therapy to prevent progression of disease.[33]

diagnosis

Culture

The most definitive method for diagnosing blastomycosis is growth of the organism from an aspirate, tissue biopsy, sputum, or body fluid.[34] For patients with disseminated blastomycosis, urine obtained before and after prostatic massage should be sent for fungal culture. *B. dermatitidis* generally takes several weeks to grow at room temperature. Once growth has occurred, highly specific and sensitive DNA probes are used to rapidly identify the mold as *B. dermatitidis.*

Biopsy and Cytologic Studies

Biopsy specimens of skin lesions characteristically show pseudoepitheliomatous hyperplasia; when distinctive, large, thick-walled yeasts with a single broad-based bud are seen, a firm diagnosis of blastomycosis can be made before the results of cultures are known. In patients who have pulmonary lesions, the distinctive yeast forms should be sought on potassium hydroxide (KOH) preparations or Papanicolaou stains performed for cytologic examination of sputum or bronchoalveolar lavage fluid and on lung biopsy specimens [*see Figure 2*].

Serologic Studies

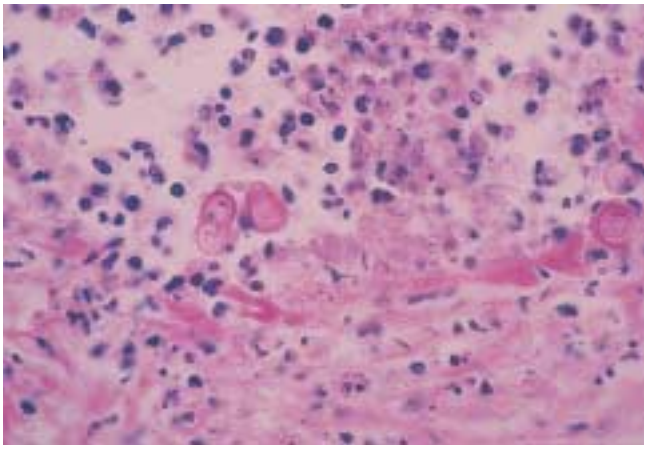
Standard CF and ID serologic assays for blastomycosis are neither sensitive nor specific. Several different enzyme immunoassays have been developed that enhance both specificity and sensitivity,[34] but they are not currently available.

differential diagnosis

The skin lesions of blastomycosis clinically mimic those caused by mycobacteria, especially nontuberculous mycobacteria; other fungal infections, especially coccidioidomycosis or paracoccidiodomycosis; mycosis fungoides; the lesions associated with bromide use; and atypical pyoderma granulosum. The histopathologic characteristics of pseudoepitheliomatous hyperplasia occur in response to other chronic infections and to bromoderma. Culture of biopsy material is of crucial importance in differentiating between these conditions.

Acute pulmonary blastomycosis is almost always misdiagnosed as a bacterial or atypical pneumonia and is then treated

a



b

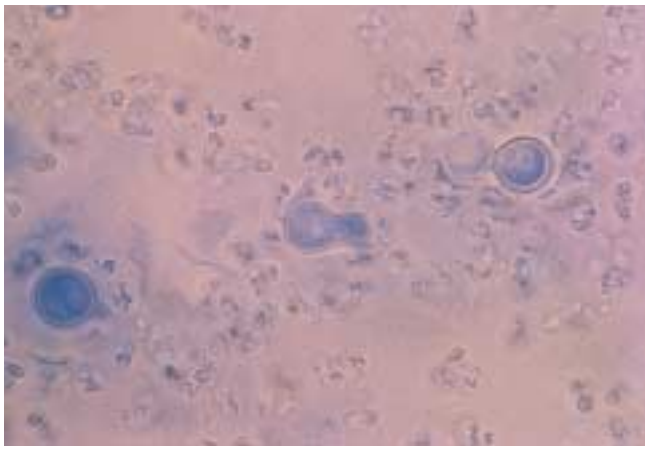


Figure 2 **Thick-walled, broad-based budding yeasts typical for *B. dermatitidis* are seen on biopsy material; they allow a definitive diagnosis of blastomycosis. (*a*) A lung biopsy specimen is stained with periodic acid–Schiff stain. The patient presented with a mass lesion in the right lower lobe and was initially thought to have lung carcinoma. (*b*) Purulent material was aspirated from one of numerous skin lesions in a patient with disseminated blastomycosis. Lactophenol cotton blue stain was used to prepare the slide.**

with antibiotics. Development of skin lesions is a strong clue suggesting blastomycosis. The masslike pulmonary lesion that is frequently noted with blastomycosis is usually thought to be lung cancer until biopsy results are reported. Likewise, lesions in the upper airways, which are usually firm and nodular, are frequently assumed to be squamous cell carcinoma until biopsy is performed. The clinical and radiographic signs and symptoms of patients with chronic pulmonary blastomycosis are indistinguishable from those associated with tuberculosis, histoplasmosis and other fungal infections, and sometimes sarcoidosis. Histopathologic studies and culture studies from biopsy material are most important in differentiating between these conditions.

treatment

Guidelines for the treatment of blastomycosis have been published by the Mycoses Study Group under the auspices of the Infectious Diseases Society of America.[33] The current recommendations are based on several multicenter, nonrandomized, open-label treatment trials and a few retrospective and prospective reports from individual institutions.[17,18,35,36] There have been no trials comparing amphotericin B with azoles for the treatment of blastomycosis.

In general, patients with pulmonary or disseminated forms of blastomycosis can be treated in a similar manner. Those patients who have mild to moderate illness should be treated with an azole; patients who have severe blastomycosis, either pulmonary or disseminated, and all patients with CNS involvement should receive amphotericin B as initial therapy.[33]

Itraconazole is the drug of choice for the treatment of mild to moderate blastomycosis.[18,33] Although lesions often begin to resolve within the first month of therapy, treatment should be continued for 6 to 12 months to achieve a mycologic cure and to prevent relapse. The usual dosage is 200 mg once or twice daily for 6 to 12 months. Children with blastomycosis should be treated with itraconazole, 5 to 7 mg/kg/day. Osteoarticular involvement requires therapy for at least a year. Fluconazole is not as effective as itraconazole for blastomycosis.[35,36] However, if a patient is unable to take itraconazole because of drug interactions or problems with absorption, fluconazole can be used; the dosage should be high—a total daily dosage of 400 to 800 mg given in once-daily or twice-daily doses for 6 to 12 months.

Amphotericin B should be reserved for patients who have severe blastomycosis, are immunosuppressed, or have CNS involvement.[33] The daily dosage is 0.7 to 1.0 mg/kg until a total of 1 to 2 g has been given. An alternative and preferred approach is to first administer amphotericin B until the patient has improved and then to switch to itraconazole, 200 mg twice daily, for a total of 6 to 12 months of therapy.

Side effects, drug absorption, and drug-drug interactions are important considerations in the use of antifungal therapy [*see* Antifungal Therapy, *below*].

prognosis

In patients with pulmonary and cutaneous manifestations of blastomycosis who undergo treatment with itraconazole, the clinical and mycologic response rates are 90% to 95%.[18] If relapse occurs, a second treatment with itraconazole is generally successful. Most reported deaths from blastomycosis occur early in the course of disease in patients who present with overwhelming pneumonia and ARDS. The outcome for patients with osteoarticular infection is good, but antifungal therapy should be given for at least a year to effect a cure.

Coccidioidomycosis

epidemiology

The dimorphic fungus *Coccidioides immitis*, the cause of coccidioidomycosis, occupies the most specific ecological niche of the endemic mycoses commonly seen in the United States.[37,38] The organism grows in the semiarid desert regions known as the Lower Sonoran Life Zone, which encompasses portions of California, Arizona, New Mexico, Nevada, and western Texas. The organism is also found in discrete areas of Central and South America.

The mycelial form of the organism produces arthroconidia, which can be widely dispersed and are highly infectious. Although persons who work outdoors are at highest risk for infection, the wide dispersal of arthroconidia ensures that most of the people who live in the endemic areas become infected before they reach adulthood. In recent years, as the allure of the sun belt has increased, there has been an increase in coccidioidomycosis in older adults. When patients with coccidioidomycosis were sub-

jected to a case-control analysis, the greatest risk factor for the development of infection was a recent move into an area endemic for *C. immitis*.[39]

Periodic outbreaks of coccidioidomycosis in the endemic area have been reported; these outbreaks are probably related to environmental cycles of rain and drought in the desert.[40] Severe infection can follow massive exposure, as might occur during participation in activities such as anthropologic excavations, in which many arthroconidia are dispersed. In addition, catastrophic massive dust storms and earthquakes have been linked to the spread of *C. immitis* to areas far beyond those normally considered endemic for the disease.[40,41]

pathogenesis

Coccidioidomycosis is acquired after the arthroconidia are inhaled into the alveoli. In the lung, the arthroconidia undergo morphologic changes that result in the formation of spherules, which are large (20 to 80 μm), thick-walled structures that become filled with huge numbers of endospores. When filled, the spherule ruptures, releasing the endospores, each of which forms a new spherule and thus propagates the infection. *C. immitis* does not form yeasts in tissues, as do other endemic mycoses.

The primary host defense against *C. immitis* is cell-mediated immunity. Neutrophils are present in most lesions but are ineffective at eradicating spherules. In hosts with deficient T cell immunity, such as patients with AIDS, the organism has the propensity to disseminate widely by the hematogenous route. Not yet elucidated is the reason for the well-known observation that dark-skinned races are at high risk for dissemination and severe infection. The highest risk appears to be among African Americans; higher but less dramatic risks for dissemination are also noted in persons from the Philippines, Native Americans, and Hispanics. Women who develop primary coccidioidomycosis while pregnant, especially in the third trimester, are also at increased risk for disseminated infection.

clinical presentations

Pulmonary Coccidioidomycosis

Most persons with acute infection with coccidioidomycosis have no symptoms or have symptoms that are seen in many types of pneumonias, such as fever, nonproductive cough, chest pain, dyspnea, fatigue, myalgias, and arthralgias. Chest radiographs show a patchy pneumonitis. Erythema nodosum occurs during acute pulmonary coccidioidomycosis in as many as 25% of women and a smaller percentage of men and is a clue to the diagnosis. For most otherwise healthy patients, acute pulmonary coccidioidomycosis is a self-limited infection.

In rare cases, when exposure to the organism is extensive or the host is immunosuppressed, acute overwhelming pneumonia, characterized by high fevers, hypoxemia, and diffuse reticulonodular infiltrates, is seen. This is most commonly noted in patients with advanced HIV infection; in many cases, the pulmonary symptoms are merely one manifestation of widely disseminated infection.

It is estimated that fewer than 5% of patients will have residual manifestations of the acute pulmonary infection.[38] Other than nodules or coccidioidomas, the most common manifestations are cavitary lesions that persist for months to years. These are generally asymptomatic, solitary, thin-walled, peripherally located lesions. Although almost half of these cavities will resolve, for some patients they can last for years and cause hemoptysis, with or without the presence of a mycetoma; they rarely rupture into the pleura. Chronic progressive coccidioidal pneumonia is rare but is more likely to occur in older patients who have COPD and diabetes mellitus; this form of coccidioidomycosis mimics tuberculosis or chronic cavitary histoplasmosis clinically and radiographically.[42]

Disseminated Coccidioidomycosis

As with other endemic mycoses, hematogenous dissemination of *C. immitis* is probably common and mostly asymptomatic. However, approximately one in every 200 people will develop symptoms of extrapulmonary coccidioidomycosis.[37] The manifestations of extrapulmonary infection with *C. immitis* are protean; they may be focal and related to only one organ system that has been seeded, or they may be systemic. The sites involved most often are skin, bones, subcutaneous tissues, and the meninges. In a recent study, risk factors for disseminated disease included African-American race, low socioeconomic status, and pregnancy.[42] Other studies have noted an increased risk of dissemination in patients with hematologic malignancies, those who have received an organ transplant, and, not surprisingly, those with AIDS.[37,38]

The skin lesions of coccidioidomycosis are typically papular or pustular initially and then become plaquelike or verrucous; facial lesions are common. Ulceration can occur, but drainage is usually minimal. Subcutaneous abscesses are commonly noted in disseminated coccidioidomycosis. Sinus tracts form, and they may drain intermittently for years. There often is underlying osteomyelitis contiguous with these subcutaneous abscesses. Bony lesions may be single or multiple and can appear in any area of the body; vertebral involvement is particularly common. Although skin and soft tissue abscesses respond fairly well to antifungal therapy, osteomyelitis and coccidioidal arthritis are characterized by an indolent, relapsing course in spite of antifungal therapy.

C. immitis can infect any organ system, but the most dreaded complication of dissemination is meningitis. Chronic progressive meningeal infection is typically concentrated in the basilar area. Meningitis may be only one of several manifestations of severe disseminated coccidioidomycosis, or it may be the sole manifestation of coccidioidomycosis. The symptoms include headache, cranial nerve palsies, and signs of increased intracranial pressure. Vasculitis occurs throughout the brain in a minority of patients with coccidioidal meningitis, and spinal cord involvement at any level of the spinal cord is not uncommon. If not treated, this form of coccidioidomycosis is fatal in virtually all patients within 2 years of diagnosis. Even with treatment, which must be lifelong, the outcomes are poor.

diagnosis

Culture

The definitive diagnostic test is growth of the organism. The mold phase of *C. immitis* grows within several days to a week on most laboratory media. DNA probes can identify the organism very specifically once growth has occurred. Also very helpful is histopathologic identification of the very distinctive large spherules of *C. immitis* in tissue biopsy, on cytologic preparations from body fluids, and even on KOH smears from sputum or from purulent material of an abscess [*see Figure 3*].

Serologic Studies

Serology is quite helpful in the diagnosis of various forms of coccidioidomycosis, as discussed in several excellent reviews.[38,43]

For serologic tests to be as useful as possible, it is important that they be sent to a reference laboratory that is experienced in performing fungal serologic analyses. Early appearance of IgM antibodies, detected through an ID test, is helpful for the diagnosis of acute coccidioidomycosis, but this test is not always sensitive enough to detect early infection. An enzyme immunoassay that is not standardized appears to be more sensitive, but false positive results are common. For patients with chronic infections and to follow the course of infection, an IgG CF assay is generally used. A decrease in the CF antibody titer or reversion to a negative titer is associated with a good clinical response, and a stable high (> 1:16) titer or a rising titer is a poor prognostic sign. A positive CF antibody titer obtained on cerebrospinal fluid is diagnostic of coccidioidal meningitis; cultures of CSF very infrequently yield *C. immitis.*

Skin Tests

Skin testing with coccidioidal antigens is helpful for the diagnosis of acute infection in a person who has just entered an endemic area, but it is of little benefit for residents of endemic areas. A negative skin test result in a patient with known coccidioidomycosis implies absence of cell-mediated immunity and portends a poor prognosis.

differential diagnosis

Coccidioidomycosis mimics many other infections. Acute pulmonary coccidioidomycosis is often diagnosed as an atypical pneumonia caused by a virus, *Mycoplasma,* or *Chlamydia.* Chronic cavitary pneumonia is radiographically similar to tuberculosis and other endemic mycoses. The diffuse pulmonary infiltrates seen in immunosuppressed patients are no different from those noted in patients with *Pneumocystis carinii* infection, histoplasmosis, tuberculosis, nocardiosis, and many other infections. The importance of tissue biopsy for histopathology and culture in these settings cannot be overemphasized.

The cutaneous lesions can appear similar to those seen in patients with blastomycosis, or they can mimic squamous or basal cell carcinomas. The bone involvement associated with coccidioidomycosis, which is frequently accompanied by subcutaneous abscesses, is typical of that seen in patients with blastomycosis or tuberculosis. The differential diagnosis of chronic coccidioidal meningitis includes tuberculosis, other fungal infections (cryptococcosis, histoplasmosis, and, rarely, sporotrichosis or blastomycosis), and sarcoidosis and other disorders of noninfectious etiology.

treatment

The Mycoses Study Group and the Infectious Diseases Society of America have recently published guidelines for the treatment of coccidioidomycosis.[44] Except for one recently completed randomized, blinded comparative trial,[45] the appropriate treatment for various forms of coccidioidomycosis has been defined by open-label, nonrandomized, multicenter trials and anecdotal experience from experienced clinicians who practice in areas endemic for coccidioidomycosis.[37,38,45-49] In contrast to the other endemic mycoses, either fluconazole or itraconazole can be used to treat coccidioidomycosis. A head-to-head, blinded, randomized comparison of fluconazole and itraconazole (each given in dosages of 400 mg daily for 1 year) showed no overall statistically significant difference in efficacy between the drugs.[45] However, for skeletal coccidioidomycosis, itraconazole was significantly superior to fluconazole (the response rates were 70% versus 30% after 1 year of therapy).[45]

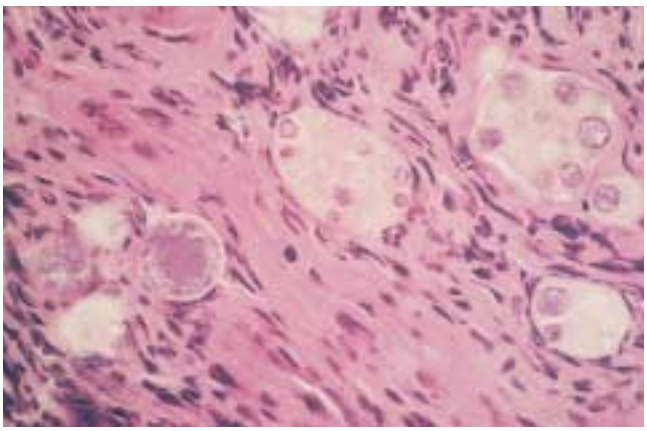


Figure 3 **Shown is a transbronchial lung biopsy specimen, stained with hematoxylin-eosin stain, from a patient with advanced HIV infection and diffuse reticulonodular pulmonary infiltrates. Large spherules (measuring 60 to 80 μm) of *C. immitis* in various stages of maturation are evident.**

Side effects, drug absorption, and drug-drug interactions are important considerations in the use of antifungal therapy [*see* Antifungal Therapy, *below*].

Pulmonary Coccidioidomycosis

Most patients with acute pulmonary coccidioidomycosis do not require therapy with an antifungal agent, and in fact, most are not seen by a physician or the diagnosis is not made until after improvement has occurred. For patients with symptoms lasting 3 to 4 weeks who show no sign of improvement, therapy with either itraconazole, 200 mg twice daily, or fluconazole, 400 mg daily, for 3 to 6 months is recommended.[44] Any patient with underlying immunosuppression, especially immunosuppression associated with HIV infection, solid-organ or stem cell transplantation, or corticosteroid therapy, should be treated because the risk of dissemination is high. Pregnant women are also at risk for dissemination and should be treated. However, azoles cannot be used in pregnant women; rather, amphotericin B, which is safe in pregnancy, is required. Consideration should be given for treatment of African-American and Filipino patients because of the high risk of dissemination in these populations.

When diffuse pulmonary infiltrates are present, patients should receive amphotericin B, 0.7 mg / kg / day initially, followed by azole therapy, for a total of at least 1 year of therapy. Patients who have chronic cavitary coccidioidomycosis should receive therapy with itraconazole, 200 mg twice daily, or fluconazole, 400 mg daily for 1 to 2 years, depending on the patient's response. However, patients with asymptomatic cavities are often observed, and treatment with an azole is begun only if the cavities increase in size or complications such as rupture into the pleural cavity have occurred or seem imminent. Surgical removal of cavities is preferred if the patient is a good operative candidate.

Disseminated Coccidioidomycosis

Disseminated coccidioidomycosis should always be treated. Patients who are seriously ill require therapy with amphotericin B, 0.7 mg / kg / day; patients with mild to moderate symptoms can be treated with an azole.[44] Treatment should be given for at least 1 year; for some patients, especially those with HIV infection and African Americans, lifelong azole therapy may be required to prevent relapse.

The most difficult form of coccidioidomycosis to treat is meningitis. The preferred treatment is fluconazole; the dosage probably should be 800 mg daily, although initial studies of this therapy employed doses of only 400 mg daily.[48,49] Itraconazole has been reported to be as effective as fluconazole, but experience with itraconazole is limited and most physicians prefer fluconazole. Approximately 20% to 25% of patients will not respond to fluconazole. Amphotericin B administered by both the intravenous and the intrathecal routes is used for those in whom fluconazole therapy fails. Intrathecal amphotericin B is not well tolerated. Repeated administration of amphotericin B into the lumbar area leads to arachnoiditis, and this is an inefficient method of delivering drug to the basilar meninges. Ideally, the drug should be administered into the cistern, but most physicians cannot perform this procedure. An alternative method is delivery through an intraventricular reservoir, but this method also does not deliver drug to the basilar meninges as efficiently as intracisternal injection. Therapy must be individualized and must be given for life.[49]

prognosis

Of all the endemic mycoses, infection with *C. immitis* is the least responsive to currently available antifungal agents. Success rates with azoles are highest in patients with soft tissue infections (approximately 70% of such patients respond after 12 months of therapy); success rates are lower in patients with chronic pulmonary infection (response rates of 50% to 60% are seen in these patients).[45-47] Relapses are more common with coccidioidomycosis than with other endemic mycoses. For patients with chronic relapsing disease, who are often of African-American descent, lifelong azole therapy may be required. For patients with meningitis, cures are rare, and suppressive therapy must be continued indefinitely.[49]

Sporotrichosis

epidemiology

The dimorphic fungus *Sporothrix schenckii* is found worldwide. The environmental niches for the organism include sphagnum moss, decaying vegetation, hay, and soil. Infection is seen most often in persons whose vocation or avocation brings them into contact with the environment. Landscaping, rose gardening, Christmas tree farming, topiary production, baling hay, and motor vehicle accidents have all been associated with sporotrichosis.[50-52] Less commonly, pulmonary sporotrichosis results from inhalation of *S. schenckii* conidia from soil. Cases of sporotrichosis usually occur sporadically, but outbreaks have been described. The largest outbreak in the United States involved 84 patients in 25 states and was traced back to conifer seedlings that had been packed in sphagnum moss from Wisconsin.[50]

S. schenckii can also be acquired through exposure to animals that are either infected or are able to passively transfer the organism from soil through scratching or biting. A variety of animals have been reported to transmit sporotrichosis, but cats with ulcerated skin lesions appear to be the most infectious. Clusters of sporotrichosis involving families and veterinarians caring for infected cats have been described.[53]

pathogenesis

Sporotrichosis typically develops in an otherwise healthy person who becomes exposed to the fungus while outdoors. The typical clinical picture is that of localized cutaneous or lymphocutaneous disease. After inoculation of *S. schenckii* conidia, the organism converts to the yeast form, which reproduces by budding. Strains that grow poorly at temperatures higher than 35° C tend to be found in fixed cutaneous lesions; these strains do not have the ability to spread along lymphatics, as do most strains of *S. schenckii*.[54]

In patients with underlying illnesses, including alcoholism, diabetes mellitus, COPD, and HIV infection, *S. schenckii* can disseminate to involve osteoarticular structures, lungs, meninges, and other organs. Neutrophils and macrophages are able to ingest and kill the yeast phase of *S. schenckii* in the presence of nonimmune human serum. A role for cell-mediated immunity as a host defense against *S. schenckii* is suggested by the observation that sporotrichosis is more severe in those with HIV infection.[55]

clinical presentations

Lymphocutaneous Sporotrichosis

Days to weeks after cutaneous inoculation of the fungus, a papule develops at the site of inoculation. The primary lesion can become nodular, but most often, it ulcerates. The drainage is not grossly purulent and has no odor, and the lesion is not terribly painful. Similar lesions subsequently occur along the lymphatic channels proximal to the original lesion. A fixed cutaneous lesion that is verrucous or ulcerative and that is not associated with lymphatic extension can also occur.[56]

Visceral Sporotrichosis

Pulmonary sporotrichosis occurs most often in middle-aged men who have COPD and abuse alcohol.[57] In contrast to most forms of sporotrichosis, systemic symptoms, including fever, night sweats, weight loss, and fatigue, are common. Dyspnea, cough, purulent sputum, and hemoptysis are common respiratory symptoms. A chest radiograph shows unilateral or bilateral upper lobe cavities with variable amounts of fibrosis and nodular lesions.

Osteoarticular sporotrichosis is found most often in middle-aged men and occurs more frequently in patients with alcoholism.[58] Although some patients experience osteoarticular involvement after local inoculation, this form of infection most often develops through hematogenous spread. Infection may involve one joint or multiple joints. The joints most commonly affected are the knee, elbow, wrist, and ankle. Isolated cases of bursitis and tenosynovitis, sometimes presenting as nerve entrapment syndromes, have been reported.

S. schenckii has rarely been reported to cause localized infection of the meninges, pericardium, eye, perirectal tissues, larynx, breast, epididymis, spleen, liver, bone marrow, or lymph nodes.[54] Disseminated sporotrichosis is very uncommon, with cases occurring primarily in patients with advanced HIV infection.[55] Most patients have widespread ulcerative cutaneous lesions and may or may not have visceral dissemination.

diagnosis

Culture

Growth of *S. schenckii* from material aspirated from a lesion, a tissue biopsy specimen, sputum, or body fluid is the most sensitive method for establishing a diagnosis of sporotrichosis. Growth of the mold phase of *S. schenckii* is usually evident a few days after inoculation onto Sabouraud agar. Synovial tissue provides a better yield than synovial fluid.

Biopsy

Histopathologic examination of biopsy material shows a mixed granulomatous and pyogenic process. The organisms often are not visualized, even with special stains for fungi, because they are rarely present in large numbers. When visualized, *S. schenckii* yeasts are 3 to 5 μm in diameter; they are oval to cigar-shaped and may show multiple buds.

Serologic Studies

Serology is less useful in the diagnosis of sporotrichosis than in histoplasmosis and coccidioidomycosis. A tube agglutination test is available at the Centers for Disease Control and Prevention; however, the sensitivity and specificity of this assay has not been established.

differential diagnosis

The differential diagnosis of lymphocutaneous sporotrichosis includes atypical mycobacterial infections, especially *M. marinum* infections; *Nocardia* infections, particularly *N. brasiliensis* infections; *Leishmania brasiliensis* infections; and tularemia.[59] Both clinically and radiographically, pulmonary sporotrichosis mimics tuberculosis; nontuberculous mycobacterial infections; other fungal infections, especially histoplasmosis; and sarcoidosis. Osteoarticular sporotrichosis is often thought to have a bacterial etiology and is often treated with antibiotics, to no avail. Growth of the organism in culture is the most important method of differentiating between these conditions.

treatment

Guidelines for the management of the various forms of sporotrichosis have recently been published by the Mycoses Study Group and the Infectious Diseases Society of America.[60] No comparative randomized, blinded treatment trials have been performed comparing various antifungal agents for the treatment of sporotrichosis. The guidelines are based entirely on open treatment trials and anecdotal experience.[58,61-64] Because most *S. schenckii* infections are subacute to chronic and are localized, oral antifungal agents are preferred; amphotericin B is reserved for the uncommon visceral infections.

Lymphocutaneous Sporotrichosis

Itraconazole is the drug of choice for lymphocutaneous sporotrichosis.[60] The usual dose is 200 mg/day. Treatment should continue for several weeks after all lesions have disappeared, usually for a total of 3 to 6 months.[58,62,63] Fluconazole is less active against *S. schenckii* but may be effective when a daily dose of 400 mg is used.[64] Saturated solution of potassium iodide (SSKI) is also effective and is much less costly than therapy with an azole. The initial dose is 5 to 10 drops three times daily in water or juice; the dose is increased each week to a maximum of 40 to 50 drops three times daily. SSKI is difficult for many patients to administer and tolerate; side effects include a metallic taste, salivary gland swelling, rash, and fever.[61]

Terbinafine appears to be effective for sporotrichosis, but few patients have been treated with this agent to date.[65] Local hyperthermia, induced by a variety of different warming devices or baths, also has been used with minimal side effects for localized cutaneous sporotrichosis.[66]

Visceral Sporotrichosis

For a seriously ill patient with pulmonary sporotrichosis, amphotericin B, 0.7 to 1.0 mg/kg/day, should be used as initial therapy.[60] After the patient has shown improvement, therapy can be changed to itraconazole, 200 mg twice daily. Azole therapy should be continued for at least 1 to 2 years. For patients who are not acutely ill, therapy can be initiated with itraconazole, 200 mg twice daily.[58] Surgical resection has proved to be useful for patients who have focal lesions and whose pulmonary function is adequate to withstand a lobectomy.[61]

Almost all patients who have osteoarticular sporotrichosis can be treated with itraconazole, 200 mg twice daily.[60] Therapy should continue for 1 to 2 years.[58] Other azoles are less effective, and SSKI is ineffective. Intravenous amphotericin B is rarely required.

Amphotericin B, 0.7 to 1.0 mg/kg/day, is the drug of choice for disseminated sporotrichosis, including meningeal infection.[60] Itraconazole, 200 mg twice daily, can be used in those patients with disseminated skin lesions and other nonmeningeal manifestations as soon as the patient's condition has stabilized. AIDS patients with disseminated sporotrichosis should receive lifelong maintenance therapy with itraconazole, 200 mg daily.[55]

prognosis

The success rate for treatment of lymphocutaneous sporotrichosis is 90% to 100%, but it is much lower for all other forms of the disease. This is undoubtedly related partly to delays in diagnosis and partly to the underlying diseases of the hosts who have noncutaneous forms of sporotrichosis. Joint function after cure of sporotrichal arthritis is often poor, and pulmonary function after treatment of pulmonary sporotrichosis is often marginal. The outcome of disseminated sporotrichosis in patients with HIV infection has been poor,[55] but almost all reports are from studies conducted before the era of highly active antiretroviral therapy. Presumably, combination antifungal therapy and highly active antiretroviral therapy would improve the prognosis.

Antifungal Therapy

For treatment of the endemic mycoses, the choice of azoles or amphotericin B is dependent on the severity of the infection and the underlying conditions and the immune status of the host. In general, most mild to moderately ill patients can be very effectively treated with an azole; most severely ill patients should receive amphotericin B as initial therapy.[67] Induction amphotericin B therapy can almost always be followed by consolidation therapy with an azole. The length of therapy is measured in months to years and is obviously managed primarily in the outpatient settting. For some patients (e.g., those with AIDS), lifelong maintenance therapy is required to prevent relapse. Thus, issues of compliance, absorption, and drug-drug interactions assume great importance in current treatment regimens for the endemic mycoses.

use of azoles to treat endemic mycoses

There are three azole antifungal agents currently available: ketoconazole, itraconazole, and fluconazole. Ketoconazole, which is available only in tablet form, is used infrequently now because of its greater toxicity, poor absorption, and modest spectrum of antifungal activity when compared with the other azoles.[68] However, it is still used for some patients who require long-term therapy with an azole because it is much less expensive than itraconazole and fluconazole. For most patients, itraconazole has supplanted ketoconazole because it is both more effective and better tolerated. Fluconazole is less active than itraconazole against most of the endemic mycoses,[21,22,35,36,64] and except for coc-

cidioidomycosis,[45,48] it is a second-line agent. The newer azoles, voriconazole, posaconazole, and ravuconazole, have not been studied for the treatment of the endemic mycoses, with the exception of a very small trial that used posaconazole for the treatment of coccidioidomycosis. These agents are not indicated for the treatment of the endemic mycoses at this time.

Side Effects

Idiosyncratic, non–dose-related hepatitis, which can sometimes be severe, occurs rarely with all azoles.[68,69] Before therapy is started, liver enzyme and bilirubin levels should be measured; they should be measured again after several weeks of therapy and then every 1 to 2 months in patients receiving long-term therapy. Mild elevations in the levels of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) (i.e., a twofold to threefold increase over normal levels) do not require discontinuance of the drug, but such increases do require very careful follow-up. If the ALT and AST levels continue to rise, the drug should be discontinued. If therapy with an azole is required, another azole can be cautiously given after the ALT and AST levels return to normal, but liver enzyme levels should be assessed weekly, and the drug must be stopped if the ALT and AST levels rise threefold or more while the patient is receiving the second azole.

It is uncommon for hypertension, edema, and hypokalemia to occur with itraconazole therapy; such side effects usually occur in older adults and require discontinuance of the drug.[70] Whether these effects are related to the recently described myocardial dysfunction noted with itraconazole has not yet been elucidated.[71]

Rash and nausea can occur with all azoles. Long-term therapy with fluconazole has been associated with alopecia, which is reversible when the drug is discontinued.[72] All azoles are contraindicated during pregnancy; teratogenicity in animals has been noted, and several cases of fetal malformations have been reported in women taking fluconazole during pregnancy.[73]

Absorption Issues

Both ketoconazole and itraconazole require gastric acid for absorption, and the capsule formulation of itraconazole also requires food for absorption. Therefore, histamine receptor antagonists (H_2 receptor blockers), antacids, and proton pump inhibitors cannot be administered to patients receiving these agents.[68] Older adults, who are often achlorhydric, may not absorb these agents well. For the treatment of endemic mycoses, itraconazole is commonly used at a daily dose of 400 mg; regardless of whether solution or capsules are prescribed, the drug should be administered in twice-daily doses of 200 mg, rather than a once-daily dose of 400 mg, to achieve appropriate serum drug levels. At doses higher than 200 mg, decreased absorption of the drug occurs.

The oral-suspension formulation of itraconazole was developed specifically to overcome the poor absorption characteristics of the capsules. The suspension requires neither acid nor food for absorption. When taken on an empty stomach, absorption is approximately 30% better than when the capsule formulation is given with both food and acid.[74] Although compliance with long-term use of the oral suspension may be difficult, the suspension is preferred to ensure adequate serum levels. In patients with serious infection and those who do not respond to therapy, serum itraconazole levels should be obtained at a reference mycology laboratory to ensure that absorption is adequate.

Use of the intravenous formulation of itraconazole, solubilized in cyclodextrins, avoids absorption problems.[75] The intravenous formulation has been approved for use for periods of only 2 weeks because of concerns about the nephrotoxicity of the cyclodextrin vehicle, which is used to solubilize the drug. The intravenous formulation cannot be given if the rate of creatinine clearance is less than 30 ml / min. The dosage of the intravenous formulation is 200 mg twice daily for 2 days, followed by 200 mg daily for the next 12 days; intravenous therapy is usually followed by treatment with the oral supension.

Fluconazole, which is nearly 100% bioavailable, has none of the absorption problems that have been so troublesome with itraconazole. It distributes into most body compartments, including the eye and the CSF; is excreted as active drug in the urine; and can be given in a once-daily dose.

Drug-Drug Interactions

The azoles interact with many other drugs primarily through their interactions with cytochrome P-450 3A4 but also to a lesser extent through their interactions with several of the other cytochrome enzymes and with P-glycoprotein.[69] These interactions

Table 1 Effects of Concomitantly Administered Drugs on Serum Levels of Azoles

Drug Affecting Azole Serum Level	*Effects That Other Drugs Have on Azole Serum Levels*		
	Ketoconazole	*Itraconazole*	*Fluconazole*
Antituberculous drugs			
Rifampin	Decreased	Decreased	Decreased
Rifabutin	None known	Decreased	No effect
Isoniazid	Decreased	None known	None known
Anticonvulsants			
Phenytoin	Decreased	Decreased	None known
Carbamazepine	None known	Decreased	None known
Gastric acid lowering agents*			
Antacids	Decreased	Decreased	No effect
H_2 receptor blockers	Decreased	Decreased	No effect
Proton pump inhibitors	Decreased	Decreased	No effect
Sucralfate	Decreased	None known	No effect

* Gastric acid–lowering agents should not be used with ketoconazole or itraconazole capsules; if necessary, they can be used with itraconazole suspension.

Table 2 Effects of Azoles on Serum Levels of Other Drugs

Drug Affected by Azole	*Effects That Azoles Have on Other Drugs' Serum Levels*		
	Ketoconazole	*Itraconazole*	*Fluconazole*
Immunosuppressants			
Cyclosporine	Increased*	Increased*	Increased*
Tacrolimus	Increased*	Increased*	Increased*
Anticonvulsants			
Phenytoin	Increased*	Increased*	Increased*
Carbamazepine	Increased*	Increased*	Increased*
Antihistamines			
Loratadine	Increased	None known	None known
Terfenadine	Increased†	Increased†	No effect
Astemizole	Increased†	Increased†	None known
Sedatives			
Triazolam	Increased*	Increased*	Increased*
Midazolam	Increased*	Increased*	Increased*
Cholesterol-lowering agents			
Lovastatin	None known	Increased†	None known
Simvastatin	None known	Increased†	None known
Antiretroviral agents			
Indinavir	Increased	None known	None known
Saquinavir	Increased	None known	None known
Ritonavir	Increased	Increased	Increased
Nelfinavir	Increased	None known	None known
Antituberculous drugs			
Rifampin	Decreased	None known	None known
Isoniazid	Decreased	None known	None known
Rifabutin	None known	Increased	Increased
Other drugs			
Cisapride	Increased†	Increased†	Increased†
Warfarin	Increased*	Increased*	Increased*
Digoxin	None known	Increased*	None known
Sulfonylureas	Increased*	Increased*	Increased*
Quinidine	Increased†	Increased†	None known

*Significant interaction; monitoring of drug serum levels, clinical status, or both is required.
†Life-threatening interaction; these agents should not be used together.

can be serious and even life-threatening. It is absolutely essential that the patient's concomitant medications be reviewed and the package insert read before use of any azole for the treatment of any fungal infection. Major interactions for the azoles are detailed [*see Tables 1 and 2*].

Itraconazole and ketoconazole have the most drug-drug interactions, but fluconazole shares several significant interactions. Increased serum levels of warfarin, phenytoin, and oral hypoglycemic agents occur when azoles are given with these commonly used drugs. Itraconazole increases serum levels of digoxin in some but not all patients. Thus, serum levels of these agents must be closely monitored in those receiving an azole.

Itraconazole and ketoconazole should never be given to patients receiving cholesterol-lowering agents, such as simvastatin and lovastatin, because of the potential for life-threatening rhabdomyolysis, nor should these agents be given to those receiving midazolam or triazolam because of the possibility of markedly increased sedation. Coadministration of an azole with cisapride, astemizole, or terfenadine is contraindicated because azoles potentiate the prolongation of the QT interval that is induced by these drugs.

Several commonly used drugs have the effect of inducing the metabolism of the azoles and thus decreasing serum azole concentrations and diminishing effectiveness. This has been reported with rifampin, phenytoin, and carbamazepine.

use of amphotericin b to treat endemic mycoses

Amphotericin B is fungicidal against a broad range of fungi. The major drawbacks are the need for intravenous administration and the inherent toxicity of the drug.[76] Some, but not all, of the infusion-related reactions to amphotericin B can be minimized with the use of preinfusion medications[77]; however, nephrotoxicity will almost always develop during the course of therapy with this agent.[78]

In an attempt to decrease the toxicity of amphotericin B, the drug has been incorporated into liposomes and other lipid delivery vehicles. Three formulations are currently available: liposomal amphotericin B (AmBisome, L-AmB); amphotericin B lipid complex (Abelcet, ABLC); and amphotericin B colloidal dispersion (Amphotec or Amphocil, ABCD). Each differs from the others and from amphotericin B deoxycholate with regard to composition, pharmacologic parameters, toxicity, recommended

dosages, and cost.[79,80] All of these preparations are clearly less nephrotoxic than amphotericin B deoxycholate, and two of the three cause fewer infusion-related reactions than amphotericin B deoxycholate. However, they are all much more costly than standard amphotericin B. Because of cost considerations, many hospitals restrict the use of these preparations to patients who have preexisting renal disease or have developed some degree of renal failure related to amphotericin B deoxycholate administration (as evidenced, for example, by a serum creatinine level of more than 2 or 3 mg / dl).

Other than two studies that show liposomal amphotericin B to be superior to amphotericin B deoxycholate for the treatment of severe histoplasmosis in AIDS patients and for the empirical treatment of febrile neutropenic patients,[23,81] there is little evidence to suggest that lipid formulations of amphotericin B are more efficacious than the standard formulation. Thus, for most patients, a lipid formulation of amphotericin B will not be used as a first-line antifungal agent but will be reserved for patients who cannot tolerate amphotericin B deoxycholate.

Nephrotoxicity Nephrotoxicity is seen in most patients receiving amphotericin B. Diminished creatinine clearance is the major manifestation, but profound hypokalemia, hypomagnesemia, or both are also common. Patients with underlying renal disease show a more rapid rise in creatinine levels than patients with normal renal function. Use of other nephrotoxic drugs, such as aminoglycosides, cisplatinum, cyclosporine, or tacrolimus, should be avoided when amphotericin B is being used.

Sodium loading is now routinely employed in an attempt to decrease nephrotoxicity[82] [*see Table 3*]. Potassium and magnesium losses can be large and can contribute to other organ dysfunction; for this reason, electrolytes should be monitored carefully, and electrolytes should be replaced as soon as the serum levels show even a slight decrease. In many patients, intravenous repletion is ultimately required to keep pace with the renal loss. Early oral repletion will obviate the need for intravenous repletion in many patients.

Infusion-related reactions Infusion-related reactions are the most common adverse events experienced by patients receiving amphotericin B. Chills or rigors, fever, nausea, headache, and myalgias occur in the majority of patients treated with amphotericin B. Most of these side effects can be diminished by prescribing a variety of medications before infusion of amphotericin B[77] [*see Table 3*].

Patients with life-threatening infections that require immediate treatment with large doses of amphotericin B are more likely to have infusion-related reactions. Pretreatment medications should be given before the first dose of amphotericin B in these patients. For others who have subacute to chronic infections, the dosage of amphotericin B can be increased gradually over a few days, decreasing the risk of immediate reactions.

Other side effects Most patients receiving amphotericin B develop a normocytic, normochromic anemia, but transfusion is rarely required.[83] Other side effects, including leukopenia, thrombocytopenia, hepatotoxicity, neuropathy, and acute pulmonary edema, are rare.

Table 3 Administration of Amphotericin B for Treatment of the Endemic Mycoses

General Guidelines

- Do not use in-line filters; no need to cover drug while infusing
- Use central I.V. catheter if at all possible. With peripheral I.V. catheter, 1,000 units of heparin can be added to the drug in an attempt to decrease phlebitis
- Dilute drug to a concentration of 0.1 mg / ml
- For the initial dose, infuse very slowly for the first 20 minutes and assess for the rare occurrence of anaphylaxis or arrhythmias
- Infuse drug over 2 to 4 hours; for patients with infusion-related reactions, slower infusion times are often better tolerated; for others, 2-hr infusion times are appropriate
- For subacute or chronic infections, start with 0.2 mg / kg and increase daily by the same increment until the desired daily dose is reached; use pretreatment medications only if required
- For life-threatening infections, give the desired daily dose within the first 24 hr; preinfusion medications should be given before the first dose because reactions are frequent

Measures to Decrease Infusion-Related Reactions (chills, fever, nausea and vomiting, headache, myalgia)

- Medications before beginning infusion of amphotericin B: acetaminophen (650 mg p.o.), diphenhydramine (50 mg p.o.), prochlorperazine (25 mg p.o. for nausea and vomiting), and meperidine (50–75 mg I.V. or I.M. for rigors); administer hydrocortisone (25–50 mg I.V.) if reactions do not improve with the above drugs

Measures to Reduce Nephrotoxicity

- Assess volume status before treatment; stop diuretics, if possible; do not restrict sodium
- Before each dose, infuse 500 ml 0.9% saline over 1 hr

Required Laboratory Studies to Assess Toxicity

- Baseline creatinine, blood urea nitrogen, potassium, magnesium levels; blood count, liver function tests
- Monitor creatinine, blood urea nitrogen, potassium, magnesium levels every other day
- Monitor blood count weekly; repeat liver function tests if clinically indicated

The author has received grant or research support from Pfizer, Inc., and participates in the speakers' bureaus for Fujisawa Healthcare, Inc.; Ortho Biotech Products, L.P.; and Pfizer, Inc.

References

1. Wheat J: Histoplasmosis: experience during outbreaks in Indianapolis and review of the literature. Medicine (Baltimore) 76:339, 1997

2. Cano M, Hajjeh RA: The epidemiology of histoplasmosis: a review. Semin Respir Infect 16:109, 2001

3. Jones TF, Swinger GL, Craig AS, et al: Acute pulmonary histoplasmosis in bridge workers: a persistent problem. Am J Med 106:480, 1999

4. Newman SL: Cell-mediated immunity to *Histoplasma capsulatum*. Semin Respir Infect 16:102, 2001

5. Wheat LJ, Kauffman CA: Histoplasmosis. Infect Dis Clin North Am (in press)

6. Gurney JW, Conces DJ Jr: Pulmonary histoplasmosis. Radiology 199:297, 1996

7. Wheat LJ, Connolly-Stringfield PA, Baker RL, et al: Disseminated histoplasmosis in the acquired immune deficiency syndrome: clinical findings, diagnosis and treatment, and review of the literature. Medicine (Baltimore) 69:361, 1990

8. Goodwin RA Jr, Owens FT, Snell JD, et al: Chronic pulmonary histoplasmosis. Medicine (Baltimore) 55:413, 1976

9. Wheat LJ, Wass J, Norton J, et al: Cavitary histoplasmosis occurring during two large urban outbreaks: analysis of clinical, epidemiologic, roentgenographic, and laboratory features. Medicine (Baltimore) 63:201, 1984

10. Johnson PC, Khardori N, Najjar AF, et al: Progressive disseminated histoplasmosis in patients with acquired immunodeficiency syndrome. Am J Med 85:152, 1988

11. Goodwin RA Jr, Shapiro JL, Thurman GH, et al: Disseminated histoplasmosis: clinical and pathologic correlations. Medicine (Baltimore) 59:1, 1980

12. Wheat LJ: Laboratory diagnosis of histoplasmosis: update 2000. Semin Respir Infect 16:131, 2001

13. Wheat LJ, Kohler RB, Tewari RP: Diagnosis of disseminated histoplasmosis by detection of *Histoplasma capsulatum* antigen in serum and urine specimens. N Engl J Med 314:83, 1986

14. Durkin MM, Connolly PA, Wheat LJ: Comparison of radioimmunoassay and enzyme-linked immunoassay methods for detection of *Histoplasma capsulatum* var. *capsulatum* antigen. J Clin Microbiol 35:2252, 1997

15. Gulati M, Saint S, Tierney LM Jr: Impatient inpatient care. N Engl J Med 342:37, 2000

16. Wheat J, Sarosi G, McKinsey D, et al: Practice guidelines for the management of patients with histoplasmosis. Clin Infect Dis 30:688, 2000
http: / / www.journals.uchicago.edu / IDSA / guidelines

17. Dismukes WE, Cloud G, Bowles C, et al: Treatment of blastomycosis and histoplasmosis with ketoconazole: results of a prospective randomized clinical trial. Ann Intern Med 103:861, 1985

18. Dismukes WE, Bradsher RW Jr, Cloud GC, et al: Itraconazole therapy for blastomycosis and histoplasmosis. Am J Med 93:489, 1992

19. Wheat J, Hafner R, Korzun AH, et al: Itraconazole treatment of disseminated histoplasmosis in patients with the acquired immunodeficiency syndrome. Am J Med 98:336, 1995

20. Wheat J, Hafner R, Wulfsohn M, et al: Prevention of relapse of histoplasmosis with itraconazole in patients with the acquired immunodeficiency syndrome. Ann Intern Med 118:610, 1993

21. McKinsey DS, Kauffman CA, Pappas PG, et al: Fluconazole therapy for histoplasmosis. Clin Infect Dis 23:996, 1996

22. Wheat J, MaWhinney S, Hafner R, et al: Treatment of histoplasmosis with fluconazole in patients with acquired immunodeficiency syndrome. Am J Med 103:223, 1997

23. Johnson PC, Wheat LJ, Cloud GA, et al: Safety and efficacy of liposomal amphotericin B compared with conventional amphotericin B for induction therapy of histoplasmosis in patients with AIDS. Ann Intern Med (in press)

24. Davis AM, Pierson RN, Loyd JE: Mediastinal fibrosis. Semin Respir Infect 16:119, 2001

25. Sherrick AD, Brown LR, Harms GF, et al: The radiographic findings of fibrosing mediastinitis. Chest 106:484, 1994

26. Pappas PG: Blastomycosis. Seminars in Respiratory and Critical Care Medicine 18:217, 1997

27. Klein BS, Vergeront JM, Weeks RJ, et al: Isolation of *Blastomyces dermatitidis* in soil associated with a large outbreak of blastomycosis in Wisconsin. N Engl J Med 314:529, 1986

28. Baumgardner DJ, Buggy BP, Mattson BJ, et al: Epidemiology of blastomycosis in a region of high endemicity in north central Wisconsin. Clin Infect Dis 15:629, 1992

29. Pappas PG, Threlkeld MG, Bedsole GD, et al: Blastomycosis in immunocompromised patients. Medicine (Baltimore) 72:311, 1993

30. Meyer KC, McManus EJ, Maki DG: Overwhelming pulmonary blastomycosis associated with the adult respiratory distress syndrome. N Engl J Med 329:1231, 1993

31. Bradsher RW: Histoplasmosis and blastomycosis. Clin Infect Dis 22(suppl):S102, 1996

32. Sheflin JR, Campbell JA, Thompson GP: Pulmonary blastomycosis: findings on chest radiographs in 63 patients. AJR Am J Roentgenol 154:1177, 1990

33. Chapman SW, Bradsher RW Jr, Campbell GD Jr, et al: Practice guidelines for the management of patients with blastomycosis. Clin Infect Dis 30:679, 2000
http: / / www.journals.uchicago.edu / IDSA / guidelines

34. Areno JP IV, Campbell GD Jr, George RB: Diagnosis of blastomycosis. Semin Respir Infect 12:252, 1997

35. Pappas PG, Bradsher RW, Chapman SW, et al: Treatment of blastomycosis with fluconazole: a pilot study. Clin Infect Dis 20:267, 1995

36. Pappas PG, Bradsher RW, Kauffman CA, et al: Treatment of blastomycosis with higher doses of fluconazole. Clin Infect Dis 25:200, 1997

37. Stevens DA: Coccidioidomycosis. N Engl J Med 332:1077, 1995

38. Galgiani JN: Coccidioidomycosis. West J Med 159:153, 1993

39. Leake JAD, Mosley DG, England B, et al: Risk factors for acute symptomatic coccidioidomycosis among elderly persons in Arizona, 1996–1997. J Infect Dis 181:1435, 2000

40. Kirkland TN, Fierer J: Coccidioidomycosis: a reemerging infectious disease. Emerg Infect Dis 2:192, 1996

41. Schneider E, Hajjeh RA, Spiegel RA, et al: A coccidioidomycosis outbreak following the Northridge, Calif, earthquake. JAMA 277:904, 1997

42. Rosenstein NE, Emery KW, Werner SB, et al: Risk factors for severe pulmonary and disseminated coccidioidomycosis: Kern County, California, 1995–1996. Clin Infect Dis 32:708, 2001

43. Pappagianis D, Zimmer BL: Serology of coccidioidomycosis. Clin Microbiol Rev 3:247, 1990

44. Galgiani JN, Ampel NM, Catanzaro A, et al: Practice guidelines for the treatment of coccidioidomycosis. Clin Infect Dis 30:658, 2000
http: / / www.journals.uchicago.edu / IDSA / guidelines

45. Galgiani JN, Catanzaro A, Cloud GA, et al: Comparison of oral fluconazole and itraconazole for progressive, nonmeningeal coccidioidomycosis: a randomized, double-blind trial. Ann Intern Med 133:676, 2000

46. Graybill JR, Stevens DA, Galgiani JN, et al: Itraconazole treatment of coccidioidomycosis. Am J Med 89:282, 1990

47. Catanzaro A, Galgiani JN, Levine BE, et al: Fluconazole in the treatment of chronic pulmonary and nonmeningeal disseminated coccidioidomycosis. Am J Med 98:249, 1995

48. Galgiani JN, Catanzaro A, Cloud GA, et al: Fluconazole therapy for coccidioidal meningitis. Ann Intern Med 119:28, 1993

49. Dewsnup DH, Galgiani JN, Graybill JR, et al: Is it ever safe to stop azole therapy for *Coccidioides immitis* meningitis? Ann Intern Med 124:305, 1996

50. Coles FB, Schuchat A, Hibbs JR, et al: A multistate outbreak of sporotrichosis associated with sphagnum moss. Am J Epidemiol 136:475, 1992

51. Hajjeh, R, McDonnell S, Reef S, et al: Outbreak of sporotrichosis among tree nursery workers. J Infect Dis 176:499, 1997

52. Dooley DP, Bostic PS, Beckius ML: Spook house sporotrichosis: a point-source outbreak of sporotrichosis associated with hay bale props in a halloween haunted house. Arch Intern Med 157:1885, 1997

53. Reed KD, Moore FM, Geiger GE, et al: Zoonotic transmission of sporotrichosis: case report and review. Clin Infect Dis 16:384, 1993

54. Kauffman CA: Sporotrichosis. Clin Infect Dis 29:231, 1999

55. Bolao F, Podzamczer D, Ventin M, et al: Efficacy of acute phase and maintenance therapy with itraconazole in an AIDS patient with sporotrichosis. Eur J Clin Microbiol Infect Dis 13:609, 1994

56. Pappas PG, Tellez I, Deep AE, et al: Sporotrichosis in Peru: description of an area of hyperendemicity. Clin Infect Dis 30:65, 2000

57. Pluss JL, Opal SM: Pulmonary sporotrichosis: review of treatment and outcome. Medicine (Baltimore) 65:143, 1986

58. Sharkey-Mathis PK, Kauffman CA, Graybill JR, et al: Treatment of sporotrichosis with itraconazole. Am J Med 95:279, 1993

59. Kostman JR, DiNubile MJ: Nodular lymphangitis: a distinct but often unrecognized syndrome. Ann Intern Med 118:883, 1993

60. Kauffman CA, Hajjeh R, Chapman SW: Practice guidelines for the management of patients with sporotrichosis. Clin Infect Dis 30:684, 2000
http: / / www.journals.uchicago.edu / IDSA / guidelines

61. Kauffman CA: Old and new therapies for sporotrichosis. Clin Infect Dis 21:981, 1995

62. Restrepo A, Robledo J, Gómez I, et al: Itraconazole therapy in lymphangitic and cutaneous sporotrichosis. Arch Dermatol 122:413, 1986

63. Conti Díaz IA, Civila E, Gezuele E, et al: Treatment of human cutaneous sporotrichosis with itraconazole. Mycoses 35:153, 1992

64. Kauffman CA, Pappas PG, McKinsey DS, et al: Treatment of lymphocutaneous and visceral sporotrichosis with fluconazole. Clin Infect Dis 22:46, 1996

65. Pappas PG, Bustamante B, Nolasco D, et al: Treatment of lymphocutaneous sporotrichosis with terbinafine: results of a randomized double-blind trial (abstr 648). 39th Annual Meeting of the Infectious Diseases Society of America, San Francisco, October, 2001

66. Hiruma M, Kawada A, Noguchi H, et al: Hyperthermic treatment of sporotrichosis: experimental use of infrared and far infrared rays. Mycoses 35:293, 1992

67. Dismukes WE: Introduction to antifungal drugs. Clin Infect Dis 30:653, 2000

68. Como JA, Dismukes WE: Oral azole drugs as systemic antifungal therapy. N Engl J Med 330:263, 1994

69. Kauffman CA, Carver PL: Use of azoles for systemic antifungal therapy. Adv Pharmacol 39:143, 1997

70. Kauffman CA: Fungal infections in older adults. Clin Infect Dis 33:550, 2001

71. Ahmad SR, Singer SJ, Leissa BG: Congestive heart failure associated with itraconazole. Lancet 357:1766, 2001

72. Pappas PG, Kauffman CA, Perfect J, et al: Alopecia associated with fluconazole therapy. Ann Intern Med 123:354, 1995

73. Pursley TJ, Blomquist IK, Abraham J, et al: Fluconazole-induced congenital anomalies in three infants. Clin Infect Dis 22:336, 1996

74. Barone JA, Moskovitz BL, Guarnieri J, et al: Enhanced bioavailability of itraconazole in hydroxypropyl–β cyclodextrin solution versus capsules in healthy volunteers. Antimicrob Agents Chemother 42:1862, 1998

75. Stevens DA: Itraconazole in cyclodextrin solution. Pharmacotherapy 19:603, 1999

76. Gallis HA, Drew RH, Pickard WW: Amphotericin B: 30 years of clinical experience. Rev Infect Dis 12:308, 1990

77. Goodwin SD, Cleary JD, Walawander CA, et al: Pretreatment regimens for adverse events related to infusion of amphotericin B. Clin Infect Dis 20:755, 1995

78. Kauffman CA: Amphotericin B. Seminars in Respiratory and Critical Care Medicine 18:281, 1997

79. Hiemenz JW, Walsh TJ: Lipid formulations of amphotericin B: recent progress and future directions. Clin Infect Dis 22(suppl 2):S133, 1996

80. Wong-Beringer A, Jacobs RA, Guglielmo BJ: Lipid formulations of amphotericin B: clinical efficacy and toxicities. Clin Infect Dis 27:603, 1998

81. Walsh TJ, Finberg RW, Arndt C, et al: Liposomal amphotericin B for empirical therapy in patients with persistent fever and neutropenia. N Engl J Med 340:764, 1999

82. Branch RA: Prevention of amphotericin B–induced renal impairment: a review of the use of sodium supplementation. Arch Intern Med 148:2389, 1988

83. Lin AC, Goldwasser E, Bernard EM, et al: Amphotericin B blunts erythropoietin response to anemia. J Infect Dis 161:348, 1990

XXXVIII MYCOTIC INFECTIONS IN THE COMPROMISED HOST

Jo-Anne van Burik, m.d., f.a.c.p.

Opportunistic fungal infections have become increasingly important over the past several decades, paradoxically because advances in medical practice have improved the survival of debilitated and immunosuppressed patients. Opportunistic fungal pathogens may originate as human commensals (e.g., *Candida albicans*) or have an environmental reservoir (e.g., *Aspergillus fumigatus*). They may have yeast (budding, unicellular) or mold (branching, tubular hyphal cell) phenotypes. Timely diagnosis of opportunistic fungal infection depends on an understanding of host characteristics, environmental risk factors, clinical presentation, and diagnostic testing. In treatment, correction of predisposing conditions can be as important as the use of antifungal medication. Localized infections may require surgery. Immunocompromised patients require either prophylaxis or early empirical treatment during high-risk periods.

Candidiasis

EPIDEMIOLOGY

C. albicans is a yeast that is virtually ubiquitous in humans. Acquired at or soon after birth, it becomes part of the normal flora of the gastrointestinal and vaginal tracts. In healthy persons, overgrowth of *Candida* at those commensal sites is a common cause of minor skin and mucous membrane infections—notably, oral candidiasis, diaper rash in infants, and vaginal candidiasis in women. In persons with immune system aberrations or iatrogenic factors that predispose to infection, candidal infections can become invasive; the organism may enter the bloodstream and disseminate to normally sterile tissues.

C. albicans has the best adherence mechanisms of the various *Candida* species and is therefore the predominant yeast colonizing the mouth, gut, and vagina. Treatment with systemic antifungal drugs may result in the replacement of *C. albicans* by other candidal species and their appearance as opportunistic pathogens. This phenomenon has grown with the wider use of such agents. In particular, *C. glabrata* and *C. krusei* bloodstream infections have become more common since the introduction of fluconazole in the early 1990s.[1,2] Other *Candida* species regarded as frequent human pathogens include *C. parapsilosis*, *C. tropicalis*, *C. lusitaniae*, *C. dubliniensis*, and *C. guilliermondii*.[3] *C. lusitaniae* is notable for resistance to amphotericin. *C. parapsilosis* is notable for nosocomial transmission from an exogenous source, such as hyperalimentation fluids. In a hospital, a series of infections with the same non-*albicans Candida* species in several patients over a short period of time suggests a common nosocomial point source.

ETIOLOGY

Candidal infection often reflects a combination of environmental and host defense factors; this is especially true of localized mucocutaneous infection from overgrowth of commensal *C. albicans.* Wearing diapers or dentures provides an environment conducive to candidal growth, as does antibiotic treatment that reduces the commensal bacteria that normally help keep *Candida* in check. Host factors include the normal life phases of infancy and pregnancy (especially third trimester) and pathologic conditions such as vitamin deficiencies and malnourishment, diabetes mellitus and other endocrine abnormalities, congenital or acquired defects in cell-mediated immunity, and malignancy. Iatrogenic contributors include radiotherapy, systemic or inhaled steroids, and other immunosuppressive medications.[4]

PATHOGENESIS

Candida can reach normally sterile tissues through iatrogenic means: contamination of plastic catheters may lead to localized infection, such as peritonitis from peritoneal dialysis catheters and cystitis or upper urinary tract infection from urinary (Foley) catheters.[5] Alternatively, *Candida* may spread from colonized surfaces to contiguous tissues. Depending on the original site, such invasion can lead to esophagitis, sinusitis, mastoiditis, and, in rare cases, primary pneumonia. A breakdown of GI mucosa through damage from cancer chemotherapy, radiation, trauma, or concurrent viral ulcers may allow a commensal *Candida* strain to gain access to the bloodstream. Tissue abscesses from hematogenous dissemination can form in the lung, brain, liver, spleen, kidney, bone, joints, heart valves, skin, and eye. In short, *Candida* infections can affect almost every organ system in the body.

DIAGNOSIS

Clinical Manifestations

Oropharyngeal candidiasis Patients with oropharyngeal candidiasis may complain of sore throat or raw tongue; eating may be so uncomfortable that patients lose weight. Pseudomembranous candidiasis (thrush) presents as loosely adherent white patches and plaques on the buccal mucosa, palate, oropharynx, or tongue. Erythematous candidiasis presents as redness without pseudomembranes. An erythematous, smooth plaque on the dorsal surface of the tongue characterizes median rhomboid glossitis. Angular cheilitis presents as erythema, maceration, and fissuring at the corners of the mouth.

Gastrointestinal candidiasis Esophageal candidiasis causes painful swallowing, a feeling of obstruction on swallowing, and substernal or retrosternal chest pain. Thrush may also be evident.[6] Candidiasis elsewhere in the GI tract presents as pain in the abdominal quadrant where the affected organs are located.

Genital candidiasis Candidal vaginitis may present as intense pruritus of the vulva and a cervical discharge that can vary from scant to thick and white. There may be edema of the vulva and erythema of the vagina and labia that extends onto the perineum.

Balanitis can begin as itching or burning, with vesicles or white patches on the penis. Plaques may extend onto the thighs, buttocks, and scrotum.

Diaper rash *Candida* diaper rash causes itching and maceration in the perianal area that may spread to the entire skin region of diaper contact. Satellite lesions may be present. Girls may have a vaginal discharge.

Hematogenous candidiasis Patients with hematogenous *Candida* infection and major organ involvement may have fever, with or without additional manifestations in the organ system involved (central nervous system, lungs, heart, urinary tract, bones, joints, liver, spleen, gallbladder, or eyes). Even in the absence of ocular symptoms, patients with positive blood cultures for *Candida* should have a dilated eye examination to look for the white cotton ball–like chorioretinitis lesions of hematogenous *Candida* endophthalmitis, which can result in permanent blindness unless treated appropriately.

Candiduria A laboratory report of *Candida* in the urine may represent contamination of the sample by vaginal flora, colonization of a urinary catheter or the lower urinary tract, or infection. Asymptomatic colonization of the lower urinary tract is the most common source of candiduria; treatment is of questionable value in such patients, because the candiduria may resolve without treatment or return despite initially successful treatment.[7,8] Systemic antifungal treatment is typically reserved for symptomatic patients and for those who are at increased risk for *Candida* infection; this population includes recent renal transplant recipients, patients with neutropenia, patients with diabetes, and those undergoing genitourinary tract manipulation.[9]

Laboratory Tests

Candidal lesions can be so distinctive that diagnosis does not require testing. When the appearance is uncertain, pseudomembranes should be scraped and sent for staining and culture. Fungal forms (4 to 6 μm oval yeast cells, pseudohyphae, and hyphae) can be visualized using 40× to 100× magnification after staining with potassium hydroxide—with or without calcofluor—or with Gram stain. Culture plates specific for yeast include bromocresol green and Sabouraud dextrose agar, although candidal organisms are easily cultured on blood agar. In immunosuppressed patients, endoscopy specimens from esophageal and other GI sites should be submitted for viral culture in addition to fungal culture, because such patients can have concurrent herpesvirus infections.

Diagnosis of hematogenous and major organ candidiasis requires culture of blood, radiographic imaging of the organ system involved, and culture of any localized fluid collections. Blood culture is used for diagnosis of fungemia, catheter-related infection, vascular infection, endocarditis, and liver and spleen infection. Even repeated cultures—both of blood and of biopsy tissue—are often negative in visceral (hepatosplenic) candidiasis, however.[10]

The current generation of automated blood culture systems has improved the ability to detect *Candida*, and *Candida* species are now the fourth most common nosocomial bloodstream isolate. When culture growth is noted, a germ tube test can be performed to distinguish *C. albicans* from other species of *Candida.* The organism is inoculated into serum and incubated for 2 to 3 hours at 37° C. If an elongated hypha (a germination tube) is seen extending from the yeast, the species is *C. albicans.*

Antifungal susceptibility testing should be requested for the first bloodstream isolate from a new infection episode. Susceptibility testing is available for azoles (e.g., fluconazole, voriconazole) but not for echinocandins.

DIFFERENTIAL DIAGNOSIS

In patients with mucocutaneous manifestations that do not fit the classic picture of *C. albicans* infection, non-*albicans* candidal species, bacteria, mold, and viruses need to be considered. Bacterial infection is especially likely in patients with findings such as abscesses or fever, whereas viral infection may be present in those with ulcerations of the mucosal surface. Infections in severely debilitated patients may have multiple copathogens. Candidiasis must also be distinguished from noninfectious syndromes such as aphthous ulcers.

TREATMENT

Active Infection

Management of candidal infections should include attempts to correct any factors that can predispose to candidal overgrowth. In some cases, this will involve local measures: with diaper rash, for example, the affected areas should be kept dry. Other cases require general measures, such as tighter control of glucose levels in diabetic patients or minimizing the number and dosage of antibacterial or immunosuppressive medications.

Antifungal drug therapy for candidiasis may be topical or systemic[11] [*see Table 1*]. For mucocutaneous infections in particular, it is important to try topical agents before giving systemic ones. The use of topical agents can prevent the replacement of *C. albicans* with other candidal strains and the development of drug resistance, and it avoids the side effects of systemic treatment.

When systemic treatment is required, fluconazole is often the drug of choice for *C. albicans* infections. *C. glabrata* infections should be considered fluconazole resistant and treated with a different agent until sensitivities are known. Importantly, *C. krusei* is innately resistant to fluconazole, so infections caused by this organism should be treated with a different agent.

In the first years of the 21st century, several new antifungal agents have gained important roles in the treatment of candidiasis and other fungal infections. Voriconazole, a new triazole with wide-spectrum antifungal activity and high bioavailability, has proved effective as empirical antifungal therapy in patients with neutropenia and persistent fever.[12] Caspofungin is the first in a new class of antifungal agents called echinocandins, which inhibit synthesis of an integral component of the fungal cell wall, β-(1,3)-D-glucan. The Food and Drug Administration has approved caspofungin for the first-line treatment of candidemia, esophageal candidiasis, and other *Candida* infections (e.g., intra-abdominal abscesses, peritonitis, and pleural space infections).[13,14] A comparison study found caspofungin to be at least as effective as amphotericin B for the treatment of invasive candidiasis and, more specifically, candidemia.[15] Caspofungin is also approved by the FDA for the empirical treatment of presumed fungal infection in febrile neutropenic patients, and it appears to be as effective as, as well as generally better tolerated than, liposomal amphotericin B when used for this purpose.[16] Micafungin, the second echinocandin to become clinically available, is approved by the FDA for intravenous treatment of esophageal candidiasis and for *Candida* prophylaxis in hematopoietic stem cell transplant recipients.[17,18]

Adjunctive measures are important in controlling severe candidal infections and in boosting the immune function of severely debilitated patients. For patients with catheter-related infections, strong consideration should be given to catheter removal from any affected site (intravascular, genitourinary, and peritoneal) if possible. Neutropenic patients may benefit from colony-stimulating factors or granulocyte transfusions. In patients with a central venous catheter, infections accompanied by tenderness or erythema along the catheter tunnel tract may require surgical de-

Table 1 Treatment of Infections Caused by *Candida* Species

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Denture stomatitis	Fluconazole	100–400 mg p.o. daily for 1–14 days	First-choice agent	Remove dentures at night
Gingivostomatitis	Nystatin	Swish 4–6 ml of 100,000 units/ml q.i.d., or suck 200,000-unit lozenges q.i.d., daily for 2 wk	First-choice topical agent	—
	Clotrimazole	Suck 10 mg troches five times daily for 2 wk	Alternative topical agent	Less bitter than nystatin
	Gentian violet	Apply 1% solution to oropharynx once, repeat weekly as needed	Alternative topical agent	Stains tissue
	Fluconazole	100–400 mg p.o. daily for 1–14 days	First-choice systemic agent	May lead to replacement of *C. albicans* with other *Candida* strains
	Itraconazole	100 mg p.o., b.i.d., or 200 mg daily of oral solution for 14 days	Alternative systemic agent	Capsule absorption improved by acidic stomach contents
	Amphotericin	0.5 mg/kg I.V. daily for 2 wk beyond the last clinical symptoms, then once or twice weekly as maintenance	Alternative systemic agent	—
Sinusitis/ mastoiditis	Fluconazole	400–800 mg p.o. or I.V. daily for 3 wk	First-choice agent	—
	Amphotericin	0.5 mg/kg I.V. daily for 3 wk	Alternative agent	—
	Itraconazole	200 mg p.o. or I.V. twice daily or 400 mg daily of oral solution for 3 wk	Alternative agent	Capsule absorption improved by acidic stomach contents
Pneumonia	Amphotericin	0.5 mg/kg I.V. daily for 2 wk ± 50 mg nebulized daily	First-choice agent	—
	Fluconazole	400–800 mg p.o. or I.V. daily for 2 wk	Alternative agent	—
	Itraconazole	200 mg p.o. or I.V. b.i.d., or 400 mg daily of oral solution for 2 wk	Alternative agent	Capsule absorption improved by acidic stomach contents
	Flucytosine	25 mg/kg p.o., q.i.d., for duration of amphotericin therapy	Adjunct to amphotericin	Follow levels and adjust dose for renal insufficiency
Meningitis				Remove foreign bodies
	Amphotericin	0.7–1.0 mg/kg I.V. daily for 2–6 wk ± 1 mg intrathecal or intraventricular daily	First-choice agent	—
	Flucytosine	37.5 mg/kg p.o., q.i.d., for duration of amphotericin therapy	Adjunct to amphotericin	Follow levels and adjust dose for renal insufficiency
	Fluconazole	400–800 mg p.o. or I.V. daily for duration of risk factors that led to infection	First-choice agent	May be used for maintenance after amphotericin
Brain macroabscess		[*See* Meningitis, *above*] ; continue therapy until lesion has resolved		Debulk or drain abscess
Esophageal infection		Continue *all* therapy for 2 wk after clinical resolution		—
	Fluconazole	100–400 mg p.o. daily	First-choice oral systemic agent	—
	Micafungin	150 mg daily I.V	First-choice I.V. systemic agent	—
	Itraconazole	100–200 mg p.o., b.i.d., or 200–400 mg daily of oral solution	Alternative agent	Capsule absorption improved by acidic stomach contents
	Caspofungin	50 mg daily by slow I.V. infusion	Alternative agent	Reduce dose for hepatic insufficiency
	Voriconazole	200 mg q. 12 hr p.o.	Alternative agent	Taken ≥ 1 hr before or after meal
	Amphotericin	0.5 mg/kg I.V. daily	Alternative agent	—
Liver/spleen infection				Correct neutropenia
	Amphotericin	0.5–1.0 mg/kg I.V. daily until fevers have resolved and lesions appear smaller on CT scan	First-choice agent	—
	Fluconazole	6–12 mg/kg p.o. or I.V. daily for the duration of lesions on serial CT scans	First-choice agent	May be used for maintenance after amphotericin

Table 1 (*continued*)

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Spleen infection		[*See* Liver/Spleen Infection, *above*]		May require splenectomy
Peritoneum infection				Remove peritoneal catheter
	Amphotericin	0.5–1.0 mg/kg I.V. daily until fevers have resolved and peritoneal fluid appears less infected, for 3-4 wk after last positive culture	First-choice agent	—
	Fluconazole	6–12 mg/kg p.o./I.V. daily until fevers have resolved and peritoneal fluid appears less infected, for 3-4 wk after last positive culture	First-choice agent	May be used for maintenance after amphotericin
	Caspofungin	50 mg daily by slow I.V. infusion, for 3-4 wk after last positive culture	Alternative agent	Reduce dose for hepatic insufficiency
Cystitis, uncomplicated				Asymptomatic candiduria may represent colonization or specimen contamination rather than infection, and may not require treatment; remove urinary catheter
	Fluconazole	200 mg p.o./I.V. on first day, then 100 mg p.o. daily for 4 more days	First-choice agent	If poor response, rule out fungus ball in renal pelvis
	Amphotericin	Bladder irrigation with 5 mg in 100 ml water at 42 ml/hr for 1–2 days, or 0.3 mg/kg I.V. once	Alternative agent	If poor response, rule out fungus ball in renal pelvis
Pyelonephritis	Amphotericin	0.5–1.0 mg/kg I.V. daily for 2 wk, or longer if fevers have not resolved	First-choice agent	Resolve concomitant nephrolithiasis
	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 3–4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr for 2–4 wk after cultures become negative	First-choice agent	Oral form taken ≥ 1 hr before or after meal
Renal/perinephric abscess	Amphotericin	0.5–1.0 mg/kg I.V. daily until several weeks after scan shows resolution of abscess	First-choice agent	Drain abscess
	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 3–4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr until several weeks after scan shows resolution of abscess	First-choice agent	Oral form taken ≥ 1 hr before or after meal
Vulva/cervix infection				Stop antibacterial agents
	Clotrimazole	One applicator per vagina each evening for 3–7 days	First-choice topical agent	—
	Gentian violet	Apply 0.25%–1% solution to vagina two to three times daily for 3 days	Alternative topical agent	Stains tissue
	Fluconazole	100–400 mg p.o. daily for 1–14 days	First-choice systemic agent	—
	Itraconazole	100 mg p.o., b.i.d, or 200 mg daily of oral solution for 14 days	Alternative systemic agent	Capsule absorption improved by acidic stomach contents
Balanitis/ balanoposthitis				Reduce predisposing factors
	Clotrimazole	1% cream applied twice daily for 1 wk	First-choice topical agent	—
	Fluconazole	Single oral 150 mg dose	First-choice systemic agent	—
Diaper/perianal rash				Keep affected area dry
	Nystatin	Cream or powder applied after every diaper change for 1 wk	First-choice topical agent	—
	Miconazole	Ointment applied after every diaper change for 1 wk	First-choice topical agent	—
Osteomyelitis/ septic arthritis				May require surgery
	Amphotericin	0.5–1.0 mg/kg I.V. daily for 4–12 wk	First-choice agent	Follow ESR
	Flucytosine	25 mg/kg p.o., q.i.d., for duration of amphotericin therapy	Adjunct to amphotericin	Follow levels and adjust dose for renal insufficiency
	Fluconazole	6–12 mg/kg p.o. daily for 1-6 mo	Maintenance	Follow ESR

Table 1 (*continued*)

Infection	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Endocarditis/ pericarditis				May require surgery
	Amphotericin	0.5–1.0 mg/kg I.V. daily, often continued for 6–10 wk after corrective surgery	First-choice agent	Follow ESR
	Flucytosine	25–37.5 mg/kg p.o., q.i.d., for duration of amphotericin therapy	Adjunct to amphotericin	Follow levels and adjust dose for renal insufficiency
	Fluconazole	6–12 mg/kg p.o. daily for 1–6 mo	Maintenance	Follow ESR
Chronic mucocutaneous candidiasis	Ketoconazole	400 mg p.o. daily for 3–9 mo	—	Take with food
Folliculitis	Fluconazole	100–400 mg p.o. daily for 1–6 wk	First-choice agent	—
	Econazole	1% cream rubbed into affected area twice daily	Adjunctive agent	—
	Ketoconazole	400 mg p.o. daily for 1–6 wk	Alternative agent	Take with food
	Itraconazole	200 mg p.o. daily for 1–6 wk	Alternative agent	—
Paronychia/ onychomycosis				Avoid moisture
	Fluconazole	6–12 mg/kg p.o. daily for 2–6 mo	First-choice agent	—
	Itraconazole	200 mg p.o. daily for 2–6 mo	Alternative agent	Capsule absorption improved by acidic stomach contents
Keratitis				May require adjunctive penetrating keratoplasty
	Natamycin	1 drop 5% suspension every 1–2 hr for first 2 days, then decrease gradually over 3 wk	First-choice agent	—
	Fluconazole	0.2% topical, 1 drop to affected eye q. 5 min for four doses for 4–6 wk	Alternative agent	—
Endophthalmitis	Amphotericin	Intravitreal 0.005–0.010 mg in 0.1 ml once, with repeat dosing based on ophthalmologic opinion	Ophthalmologist will make decision to use	Advanced vitritis may require systemic amphotericin and/or vitrectomy
	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 3–4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr; duration determined by ophthalmologist		
	Caspofungin	50 mg daily by slow I.V. infusion; duration determined by ophthalmologist		
Systemic infection				Halt fungemia
	Amphotericin	0.5–1.0 mg/kg I.V. daily for 7 days after the last positive blood culture, then switch to fluconazole, 6–12 mg/kg p.o. daily for 7 additional days; fluconazole may be continued longer if neutropenia has not resolved; if the *Candida* species recovered is resistant to fluconazole, continue amphotericin for a total of 14 days after the last positive blood culture	First choice for patients who are neutropenic, deteriorating, or otherwise unstable	Use lipid formulation, not generic, for patients with nephrotoxicity or infusion toxicity; remove catheters and replace at new sites; check for metastatic lesions (e.g., endophthalmitis)
	Fluconazole	6–12 mg/kg I.V. or p.o. daily for 14 days after the last positive culture	First choice for stable patients who are not neutropenic	Remove catheters and replace at new sites; check for metastatic lesions (e.g., endophthalmitis)
	Caspofungin	50 mg daily by slow I.V. infusion for 14–28 days after the last positive culture	Alternative agent	Reduce dose for hepatic insufficiency
	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 3–4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr for 14–28 days after the last positive culture	Alternative agent	Oral form taken ≥ 1 hr before or after meal
Prevention of candidiasis in hematopoietic stem cell transplant recipients	Micafungin	50 mg I.V. daily until 5 days after engraftment	—	—

Note: Major pathogens include *C. albicans*, *C. tropicalis*, *C. parapsilosis*, and *C. pseudotropicalis*. *C. lusitaniae* and *C. guilliermondii* are usually resistant to amphotericin. *C. glabrata* and *C. krusei* are usually resistant to fluconazole.
ESR—erythrocyte sedimentation rate

Table 2 Treatment of Infections Caused by *Cryptococcus* Species

Infection Site	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
CNS	Amphotericin	0.7–1.0 mg/kg I.V. daily for AIDS patients, 0.5–0.8 mg/kg I.V. daily for non-AIDS patients; continue for 2–6 wk	First-choice agent	Continue until patient is stable and afebrile
	Flucytosine	37.5 mg/kg p.o., q.i.d., for duration of amphotericin therapy	Adjunct to amphotericin	Adjust dose for renal insufficiency
	Fluconazole	6 mg/kg p.o. daily for 8–10 wk, then 3 mg/kg for duration of suppressed cell-mediated immunity	Maintenance/prophylaxis	—
Pulmonary	Fluconazole	6 mg/kg I.V./p.o. daily for 2–6 mo	First-choice agent	—
	Amphotericin	0.5–1.0 mg/kg I.V. daily for 2–6 wk, then change to fluconazole	Alternative agent	—

bridement of the tract. Macroabscesses may require surgical drainage or debulking. Patients with recurrent positive cultures despite receiving both systemic medication and adjunctive measures should be reexamined clinically and radiologically to look for an occult focus of *Candida* organisms, such as an infected thrombus or abscess.

Prophylaxis

Secondary pharmacologic prophylaxis for mucocutaneous infection may be started in patients who have experienced several episodes of thrush (in advanced AIDS) or vaginitis (in pregnancy). Profoundly neutropenic patients with sufficient mucosal breakdown to provide portals of entry for *Candida* into the bloodstream require primary pharmacologic prophylaxis until the mucosal barriers have recovered their protective function. Examples of such mucosal barrier breakdown include, but are not limited to, mucositis during bone marrow transplant procedures, GI graft versus host disease (GVHD), and viral gastroenteritis. Prophylactic medications can be topical or systemic, depending on the site involved and the patient's tolerance of individual agents.

The use of antifungal chemotherapy can shift GI yeast flora from *C. albicans* to fluconazole-resistant species. For that reason, candidal infections that occur during fluconazole prophylaxis require culture and sensitivity testing to determine definitive treatment.

PROGNOSIS

Fortunately, most candidal infections are readily manageable, provided that further diagnostic investigation (e.g., susceptibility testing and radiography) is pursued for poorly responsive infections. Morbidity and mortality remain highest for patients with hematogenous and major-organ candidiasis.

Cryptococcosis

EPIDEMIOLOGY

Cryptococcus neoformans is a yeast that is widely distributed in nature. Environmental sources of *C. neoformans* include aged pigeon droppings, pigeon nesting areas, dust, and eucalyptus trees.[19] However, most persons exposed to environmental sources of *C. neoformans* do not experience symptomatic disease. Suppressed cell-mediated immunity is the most important risk factor for symptomatic infection; currently, AIDS patients with $CD4^+$ T cell counts below $100/mm^3$ account for 80% to 90% of cases of clinical cryptococcosis. In the United States, the annual incidence of cryptococcosis in AIDS patients decreased substantially during the 1990s, with more than two thirds of cases occurring in patients who did not receive highly active antiretroviral therapy.[20] Person-to-person transmission has not been documented other than through transplanted organs.

ETIOLOGY

Cryptococcus infection begins with the inhalation of aerosolized organisms and localized proliferation with pulmonary invasion. In immunocompetent individuals, pulmonary infection may be asymptomatic and resolve spontaneously. Immunocompromised persons may have an acute, symptomatic pulmonary infection, which may disseminate by hematogenous spread—most often to the CNS, but also to the skin, soft tissue, genitourinary tract, bone, or joints. Cryptococcal CNS infection may become symptomatic while pulmonary infection clears because cerebrospinal fluid lacks several soluble anticryptococcal factors that are present in serum, such as complement components. Organ transplant recipients who receive calcineurin inhibitors such as tacrolimus as their primary immunosuppressive agent may be protected from cryptococcal infections, because calcineurin is thought to be a potential virulence factor for the yeast. Moreover, only a minority of cryptococcal infections in organ transplant recipients involve the CNS.[21] Rarely, *Cryptococcus* can be directly inoculated through intact skin as a route of infection.

PATHOGENESIS

C. neoformans is a round or oval yeastlike structure, 4 to 6 μm in diameter, that grows well at body temperature. A large protective polysaccharide capsule surrounds each cell. The highly negative surface charge may contribute to resistance to leukocyte phagocytosis.

DIAGNOSIS

Clinical Manifestations

CNS infection Cryptococcosis of the CNS may present as mild and nonspecific complaints, such as persistent headache, nausea, dizziness, ataxia, impaired memory and judgment, irritability, somnolence, clumsiness, confusion, and obtundation. Patients may or may not have fever, and most have minimal or no nuchal rigidity. Papilledema is noted in up to one third of cases and cranial nerve palsies in about one fifth. If the cranial nerves become involved, patients may experience decreased vi-

sual acuity, diplopia, facial numbness or weakness, or catastrophic vision loss.[22] As the disease progresses, seizures may occur.

Respiratory infection Pulmonary cryptococcosis may present as cough, dyspnea, blood-streaked sputum, and a dull ache in the chest.

Cutaneous infection Skin lesions may be single or multiple and commonly begin as painless lesions of the face or scalp. Skin lesions may take the form of erythematous or umbilicated papules, pustules, acneiform lesions, indurated plaques, palpable purpura, soft subcutaneous masses, sinus tracts, cellulitis, vesicles, or large ulcers with undermined edges.

Prostatic infection The prostate can provide a sequestered focus of active cryptococcosis after therapy for systemic cryptococcal infection. These foci may occur even in patients who did not have prostatic involvement initially, and the foci can be a source of recurrent systemic infection. Prostatic cryptococcosis may present as a peripheral prostatic nodule, but patients often have no symptoms. Diagnosis requires culture of urine obtained after prostatic massage.

Laboratory Tests

Routine blood studies remain normal in cryptococcosis. Lumbar puncture is indicated in immunosuppressed patients with CNS abnormalities; characteristic findings include elevated opening pressure, depressed glucose level, increased protein concentration, and lymphocytic pleocytosis. The latex agglutination test detects antigen in CSF or serum in more than 90% of patients with cryptococcal meningitis, whereas India ink smear of CSF detects cryptococci in 25% to 60%.

In cryptococcal CNS infection, CT or MRI scans of the head may be normal or reveal hydrocephalus, cerebral edema or atrophy, or a focal space-occupying mass lesion. Gelatinous cryptococcal pseudocysts may appear as nonenhancing lesions.

In pulmonary cryptococcosis, chest x-rays most often show one or more circumscribed masses or nodules, often in the upper lobes, without hilar involvement. Less common radiographic patterns include segmental pneumonia, single thick-walled cavities, pleural effusion, and miliary disease.

Antigen titers in either serum or CSF can be used to follow the course of disease, but a lack of standardization among manufacturers of cryptococcal antigen tests means that reliable results can be obtained only if the same kit is used for serial measurements. For definitive diagnosis of cryptococcosis, positive antigen test results must be confirmed by culture. Sputum cultures are often negative, however, and may be falsely positive. Bone lesions appear on radiographic studies as round, lytic lesions.

DIFFERENTIAL DIAGNOSIS

CNS cryptococcosis may resemble coccidioidomycosis, histoplasmosis, tuberculosis, brucellosis, syphilis, viral meningoencephalitis, meningeal metastases, sarcoidosis, and chronic benign lymphocytic meningitis. Cryptococcomas may resemble pyogenic, nocardial, or mold-associated abscesses; tuberculosis; toxoplasmosis; hemorrhage; or lymphoma or other neoplasms. Pulmonary cryptococcosis may be indistinguishable from tuberculosis, histoplasmosis, pneumocystosis, and neoplasm. Cutaneous cryptococcosis may resemble comedones, acne, lipoma, syphilis, tuberculosis, sarcoidosis, molluscum contagiosum, and basal cell carcinoma. Bone lesions resemble those of other mycoses and tuberculosis.

TREATMENT

In patients with cryptococcal CNS infection, therapy begins with intravenous amphotericin B plus oral flucytosine [*see Table 2*].[23] Once the patient is clinically stable and afebrile, which usually takes 2 to 6 weeks to achieve, those agents are replaced with fluconazole. After 8 to 10 weeks of maintenance therapy, fluconazole is continued at a reduced dose for the duration of suppressed cell-mediated immunity, to prevent relapse.[24]

A condition that resembles immune reconstitution syndrome has been reported in organ transplant recipients with *C. neoformans* infection. Onset of this condition was observed a median of 5.5 weeks after the initiation of antifungal therapy and was marked by worsening of clinical manifestations despite negative cultures for *C. neoformans.* In some patients, this condition may be misconstrued as a failure of therapy. Immunomodulatory agents may have a role as adjunctive therapy.[25]

Pulmonary disease from *C. neoformans* can be treated with amphotericin or fluconazole. Secondary prophylaxis for pulmonary disease is often not indicated in patients whose immunosuppression is not severe.

COMPLICATIONS

Elevated CSF pressure in CNS cryptococcosis is associated with blindness and death. An absolute pressure under 250 to 300 mm H_2O can be maintained by removing CSF to decrease pressure by 50%.[26] Hydrocephalus can lead to permanent loss of cognitive function even in patients whose infection is considered cured. Relief of hydrocephalus with a shunt is vital to ensure an optimal outcome.[27]

PROGNOSIS

In CNS cryptococcosis, a good prognosis is associated with normal mental status, a CSF leukocyte level of more than 20 cells/mm^3, and a CSF cryptococcal antigen titer of less than 1:1,024. Even in patients who respond to therapy initially, up to 40% may have residual neurologic defects and up to 25% of the 40% may experience relapse.

For HIV-negative patients who had cryptococcosis between 1990 and 1996, overall mortality was 30% and attributable mortality 12%. Significant predictors of mortality included age greater than 60 years, hematologic malignancy, and organ failure.[28] For solid-organ transplant recipients, renal failure at the time of hospital admission has been identified as an independent predictor of death.[21]

Pneumocystosis

EPIDEMIOLOGY

Common and apparently harmless in the lungs of healthy persons, *Pneumocystis jiroveci* (formerly known as *P. carinii*) can cause pneumonia in those with prolonged lymphopenia. AIDS patients remain the single largest group at risk, although prophylaxis has markedly reduced the incidence of the disease in that population.[29] The risk of *Pneumocystis* pneumonia (PCP) is also higher in the setting of primary immune deficiencies, severe malnutrition (which accounted for epidemics in Central Europe during World War II), organ transplantation, and long-term corticosteroid treatment with monthly doses above 20 mg of pred-

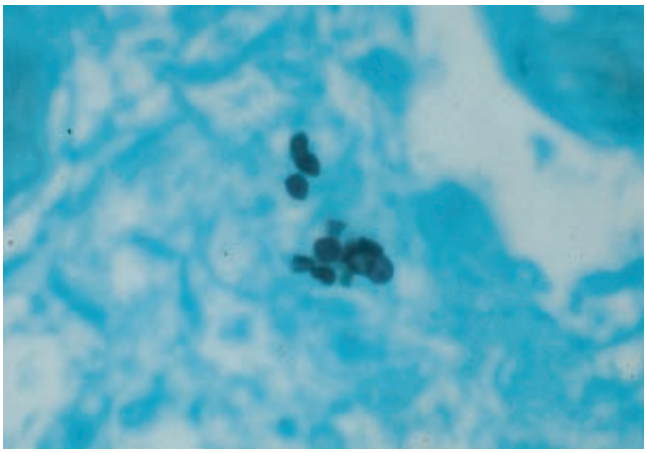


Figure 1 **Cysts of *Pneumocystis*, which stain black and measure approximately 4 to 5 μm, are visible on a methenamine-silver stain of a cytospin preparation from a bronchoscopic alveolar lavage fluid sample.**

nisone or the equivalent. For solid-organ transplant (except lung transplant) recipients, the incidence of *P. jiroveci* pneumonia is highest during the first year after transplantation.[30]

ETIOLOGY

Once *P. jiroveci* is inhaled, it attaches to lung epithelial cells. Pulmonary infection focuses on the interstitium and the alveoli.

Uncommonly, *P. jiroveci* infection spreads beyond the lungs. Patients receiving aerosolized pentamidine as prophylaxis are at risk for upper lung lobe and extrapulmonary disease because this is primarily a topical treatment; the aerosol does not penetrate into the less well ventilated portions of the lungs, and the drug is not absorbed into the circulation. Sites of extrapulmonary involvement include lymph nodes, abdominal organs, bone marrow, eyes, and thyroid.[31]

PATHOGENESIS

Although *Pneumocystis* was formerly classified as a protozoan, nucleic acid comparison studies have identified this organism as a fungus. Host immune defects in humoral and cellular immunity allow unchecked replication of the organism. In the absence of $CD4^+$ T cells, alveolar macrophages are unable to contain infection.

DIAGNOSIS

Clinical Manifestations

Most patients with PCP have a fever.[32] Tachypnea and tachycardia are common in acutely ill patients. Clinical manifestations may develop insidiously and may not interfere with the person's daily routine initially. Consequently, some patients present with several weeks of pulmonary symptoms, including dyspnea, nonproductive cough, hypoxemia, chest pain, and hemoptysis. Nevertheless, lung auscultation may reveal only scant abnormal findings.[33]

Laboratory Tests

Chest x-rays show bilateral infiltrates in most patients. Patients who have received aerosolized pentamidine as prophylaxis but experienced breakthrough infection are more likely to have disease confined to the apices.[31]

Methenamine-silver staining of induced sputum or bronchoscopic alveolar lavage specimens can usually confirm the diagnosis [*see Figure 1*]. At 4 μm in diameter, the cysts of *Pneumocystis* are similar in size to the 5 μm cysts of *Histoplasma.* Pulse oximetry and arterial blood gases are among the methods for testing the level of oxygenation that are used to evaluate disease severity and monitor progression.

DIFFERENTIAL DIAGNOSIS

Any immunocompromised patient with respiratory symptoms, fever, and an abnormal chest radiograph should be considered to have PCP. Nevertheless, many other infectious agents and noninfectious diseases can mimic PCP.

TREATMENT

Active Infection

The treatment of choice for *P. jiroveci* infection is high-dose intravenous trimethoprim-sulfamethoxazole (TMP-SMX). Severely hypoxemic patients should receive adjunctive corticosteroid therapy.[34] Approximately 25% of patients receiving high-dose TMP-SMX have a poor therapeutic response or hypersensitivity reactions. Alternative treatment regimens include pentamidine, trimethoprim-dapsone, clindamycin-primaquine, trimetrexate-leucovorin, and atovaquone [*see Table 3*].

Prophylaxis

Primary prophylaxis is indicated for HIV and organ transplant patients whose $CD4^+$ T cell count is below $200/mm^3$, and secondary prophylaxis is indicated for those with a history of PCP. In HIV-infected patients who respond to highly active antiretroviral therapy, primary and secondary *Pneumocystis* prophylaxis can be safely discontinued once the $CD4^+$ T cell count has remained above $200/mm^3$ for more than 3 months.[35]

TMP-SMX is the first choice for pharmacologic prophylaxis. Along with helping prevent PCP, TMP-SMX has the added benefit of providing protection against toxoplasmosis, salmonellosis, *Haemophilus* infection, and staphylococcal infection.[36] The dosage for TMP-SMX prophylaxis is specific to different immunocompromised patient populations.

There is a significant incidence of adverse reactions to TMP-SMX (rash, nausea, and, in rare cases, myelosuppression or cholestasis). However, because of its superior efficacy and low cost, desensitization should be attempted before switching to a second-choice agent in patients who experience side effects from TMP-SMX [*see Table 4*]. Other options for prophylaxis include dapsone, aerosolized pentamidine, atovaquone, and trimetrexate[37] [*see Table 3*]. Breakthrough infection with *P. jiroveci* may occur despite prophylaxis, especially in patients who have unrecognized poor compliance with a regimen of TMP-SMX or high-dose dapsone (100 mg daily).

PROGNOSIS

Prognosis is related to the degree of hypoxemia at presentation, degree of infiltrates on chest radiographs, elevated lactate dehydrogenase level, the presence of copathogens, prior lung damage, severity of the underlying disease, and general markers of debility. Spontaneous pneumothorax is associated with a poor prognosis.

Aspergillosis

Aspergillus is an environmental saprobe whose spores (coni-

dia) readily become airborne. Conidia of *A. fumigatus,* the principal *Aspergillus* species involved in human infection, are the ideal size for deposition into lungs and sinuses, with a diameter of 2 to 5 μm. *Aspergillus* thrives in wet areas such as crawl spaces, basements, and water-treatment facilities. Renovations may result in the release of spore-bearing dust. Potted plants and marijuana have been suggested as sources of spores.

After it is inhaled, *Aspergillus* can cause a localized infection that may result in allergic or invasive disease. Allergic bronchopulmonary aspergillosis (ABPA) and invasive aspergillosis are the two major clinical variants.

ALLERGIC BRONCHOPULMONARY ASPERGILLOSIS

Epidemiology

ABPA affects patients with a hypersensitive immune status, such as those with atopy, asthma, extrinsic alveolitis, and cystic fibrosis.[38] Genetic determinants of immune response are thought to help explain why ABPA develops in only some patients with those conditions. ABPA is equally distributed between males and females with asthma and probably also in those with cystic fibrosis. Diagnosis is usually made in the late teenage years or in the 20s.

Etiology

In susceptible individuals, exposure to *Aspergillus* antigens causes the formation of IgE antibodies directed at the antigen. Reexposure will then result in mast cell degranulation and eosinophilic infiltration, with wheezing and fleeting pulmonary infiltrates.

Pathogenesis

Aspergillus does not invade tissue in ABPA; rather, it colonizes pulmonary secretions. The severity of ABPA probably varies with the intensity of exposure to spores, the species of *Aspergillus* involved, and the immune status of the host. Initially, fungal spores are trapped within thick bronchial mucus. As the fungus proliferates, the bronchi fill with mucus and become dilated; fungal hyphae can be identified in mucoid impactions. ABPA has been divided into five stages [*see Table 5*]. Patients do not necessarily progress through these stages in linear fashion.[39]

Diagnosis

Clinical manifestations Patients with ABPA may have cough, wheezing, expectoration of sputum containing brown plugs, fever, weight loss, dyspnea, malaise, pleuritic chest pain, sweats, and hemoptysis.

Table 3 Treatment and Prophylaxis of *Pneumocystis jiroveci* Pneumonia

	Drug	*Dosage*	*Relative Efficacy*	*Comments*
Treatment				All treatments last 21 days
	Trimethoprim sulfamethoxazole (TMP-SMX)	Two double-strength tablets p.o., or 5 mg/kg of the trimethoprim component I.V. q. 8 hr	First-choice agent	Rash or fever in 19% of patients
	Corticosteroid (oral prednisone or I.V. prednisolone)	40 mg b.i.d. for 5 days, then 40 mg daily for 5 days, then 20 mg daily for 11 days; administer 30 min before TMP-SMX	Adjunctive agent	If arterial oxygen pressure < 70 mm Hg or arterial-alveolar gradient > 35 mm Hg
	Pentamidine	4 mg/kg I.V. daily	Alternative agent for severe disease	—
	Trimethoprim-dapsone	Trimethoprim, 5 mg/kg p.o., t.i.d., and dapsone 100 mg p.o. daily	Alternative regimen for mild to moderate disease	Rash or fever in 10% of patients
	Clindamycin-primaquine	Clindamycin 300–450 mg p.o., q.i.d., or 600 mg I.V. q. 8 hr, and primaquine base 15 mg p.o.daily	Alternative regimen	Give clindamycin I.V. for more severe infections; rash or fever in 21% of patients
	Atovaquone	750 mg suspension p.o., b.i.d.	Alternative agent for mild to moderate disease	Take with food; less effective than TMP-SMX but fewer side effects
	Trimetrexate-leucovorin	Trimetrexate, 45 mg/m^2 I.V. daily; leucovorin, 20 mg/m^2 q. 6 hr for 3 days after trimetrexate	Alternative regimen	—
Prophylaxis		Discontinue when CD4$^+$ cell count > 200/mm^3 for 3 mo		
	TMP-SMX	One double-strength or single-strength tablet, minimum three times per week but usually taken daily	First-choice agent	Provides cross-protection for toxoplasmosis and some bacteria
	Dapsone	100 mg p.o. daily	Alternative agent	Well tolerated if glucose-6-phosphatase levels are normal
	Pentamidine	300 mg in 6 ml sterile water by aerosol every 4 wk	Alternative agent	Breakthrough disease occurs in upper lung lobes
	Atovaquone	1,500 mg p.o. daily	Alternative agent	Take with food
	Dapsone-pyrimethamine–leucovorin	Once weekly: dapsone, 200 mg p.o.; pyrimethamine, 75 mg p.o.; leucovorin 25 mg p.o.	Alternative regimen	Provides cross-protection for toxoplasmosis

Table 4 Desensitization of Adult Patients with Sulfa Allergy*

TMP-SMX Preparation	*Day*	*Dose (ml)*	*SMX Equivalent (mg)*
Pediatric oral suspension (40 mg/200 mg per 5 ml), 1 ml diluted with 9 ml normal saline, to equal 4 mg SMX/ml	1	0.25	1
	2	0.5	2
	3	1	4
	4	2	8
	5	4	16
Pediatric oral suspension, full strength	6	0.5	20
	7	1	40
	8	2	80
	9	4	160
Single-strength tablet	10–30	One tablet	400

*If the patient has had no reaction after 30 days of continuous therapy, the full dose can be given.
TMP-SMX—trimethoprim-sulfamethoxazole

Lung auscultation in patients with active, recurrent, or corticosteroid-dependent disease (stages I, III, or IV) gives variable results, ranging from normal to localized rales in areas of consolidation to wheezes. In fibrotic disease (stage V), patients may have digital clubbing or cyanosis.

Laboratory tests Primary criteria for a diagnosis of ABPA include episodic bronchial obstruction (asthma), immediate skin reactivity to *A. fumigatus* antigen, precipitating antibodies against *A. fumigatus* antigen, increased total serum IgE concentration (> 1,000 ng/ml), history of transient or fixed pulmonary infiltrates, peripheral blood eosinophilia, and central bronchiectasis. Secondary criteria include *A. fumigatus* in sputum, history of expectoration of brown plugs or flecks, and late skin reactivity to *Aspergillus* antigen. Eosinophilia has diurnal variation and may decrease with steroid use or during remission (stage II). Pulmonary function tests are useful in defining the stage of disease and underlying asthmatic activity.

Differential Diagnosis

Allergic bronchopulmonary syndromes can be triggered by molds other than *A. fumigatus.* Allergic fungal sinusitis is histologically identical to ABPA; if it is caused by *A. fumigatus*, the serologic findings will be positive in the absence of asthma and lower respiratory findings.

Treatment

Active infection ABPA is treated with systemic corticosteroids supplemented with oral itraconazole [*see Table 6*]. Itraconazole reduces the burden of colonizing fungal organisms, thereby decreasing chronic antigen stimulation, permitting a lower corticosteroid dose, and decreasing future recurrences of ABPA.[40-42] Measures of clinical response can be followed serially and include decreasing serum IgE concentration, improvement in exercise tolerance and pulmonary function, and resolution of pulmonary infiltrates.

Corticosteroid-dependent (stage IV) patients will have disabling wheezing and dyspnea after oral corticosteroids are tapered or stopped. Patients with fibrotic disease (stage V) require supplemental oxygen and moderate doses of alternate-day prednisone.

Prophylaxis Serum IgE levels should be monitored monthly for 2 years and then every 2 months in patients who have had ABPA. Pulmonary function tests should be performed annually. An increase in serum IgE above remission levels should prompt an evaluation for exacerbation.

Table 5 Clinical Stages of Allergic Bronchopulmonary Aspergillosis

Stage	*Clinical Findings*	*IgE Level*	*Eosinophilia*	Aspergillus fumigatus *Precipitating Antibodies*	*Chest Films*
I – Active	Asthma	Extremely elevated	Extreme	Present	Infiltrates
II – Remission	Normal	Below active levels but not normal	None	± Slightly elevated	Clear
III – Exacerbation	Asthma	Twice normal	Extreme	Not followed	Infiltrates
IV – Corticosteroid dependent	Tapering steroid therapy worsens asthma	Usually elevated but may be normal	—	Not followed	Central bronchiectasis
V – Fibrotic	Dyspnea, cyanosis, rales, clubbing, cor pulmonale	—	—	Not followed	Fibrosis

Note: IgE levels and eosinophilia are less often followed in patients with stage IV or V disease.

Table 6 Treatment of Disease Caused by *Aspergillus* Species

Disease	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Allergic bronchopulmonary aspergillosis	Corticosteroid (prednisone)	0.5 mg/kg daily for 2 wk, then 0.5 mg/kg every other day for 3 mo, then taper to discontinuance over another 3 mo	First-choice agent	Follow serum IgE monthly
	Itraconazole	200 mg p.o., b.i.d., for 4 mo	Adjunctive agent	Capsule absorption improved by acidic stomach contents
Invasive aspergillosis	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr; treat for minimum of 12 wk	First-choice agent	Oral form taken ≥ 1 hr before or after meal
	Caspofungin	70 mg I.V. loading dose, then 50 mg I.V. daily; treat for minimum of 12 wk	Alternative agent	Reduce dose for hepatic insufficiency
	Amphotericin	1.0–1.5 mg/kg I.V. daily for 2–12 wk, then switch to maintenance therapy	Alternative agent	Use lipid formulation, not generic, for patients with nephrotoxicity or infusion toxicity
	Itraconazole	200–300 mg p.o., b.i.d., for 2–12 mo or duration of clinical disease	Maintenance or alternative agent	Solution has better bioavailability than capsule; document blood levels

INVASIVE ASPERGILLOSIS

Epidemiology

Invasive aspergillosis typically occurs during periods of immune suppression, including neutropenia, treatment of GVHD, and treatment of threatened rejection in transplant recipients. Other factors that may predispose immunosuppressed patients to invasive infection include diminished cough reflex, impairment of mucociliary function, and ischemic damage in the perioperative period or at surgical sites. Except in outbreaks, the incidence of invasive aspergillosis should not be higher than 10% to 15%, even in high-risk patients.

Etiology

In bone marrow transplant recipients, invasive aspergillosis has two distinct periods of onset. The first peak, termed early disease, occurs at a median of 16 days after transplantation, when recipients of both autologous and allogeneic bone marrow have absolute neutropenia and lymphopenia, allowing any form of silent colonization to develop rapidly into invasive and often fatal disease. The second peak, late disease, occurs at a median of 96 days. Late disease develops indolently in recipients of allogeneic bone marrow who have recovered from absolute neutropenia but remain immunosuppressed[43] [*see Figure 2*]. Risk factors for early aspergillosis include allogeneic and especially un-

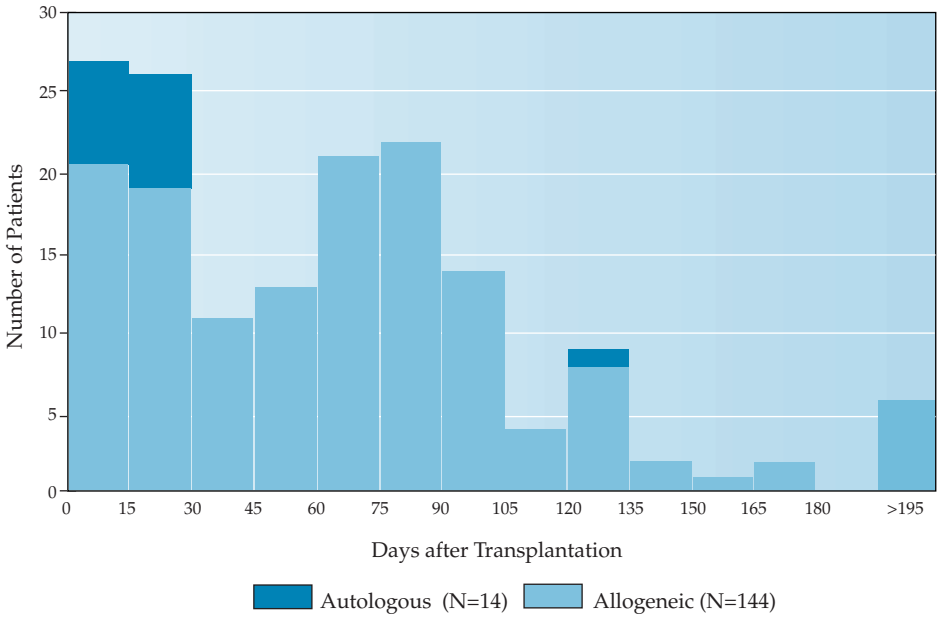


Figure 2 ***Aspergillus* infection during the first 6 months after bone marrow transplantation has a bimodal distribution. The first peak (early disease) occurs at a median of 16 days. Early disease is associated with profound neutropenia and develops rapidly. The second peak (late disease) occurs at a median of 96 days. Late disease is associated with corticosteroid use and develops indolently.[43]**

related donors, male gender, summer season, and transplantation outside of laminar airflow rooms. Risk factors for late aspergillosis are construction in the vicinity of the hospital, acute GVHD, and corticosteroid therapy. The risk for both early and late disease is higher with older age and increasing severity of the underlying oncologic condition. The 1-year survival estimate for bone marrow transplant patients with invasive aspergillosis is less than 10%.

Pathogenesis

Along with *A. fumigatus* and *A. flavus,* emerging species involved in invasive aspergillosis include *A. terreus*, *A. nidulans*, *A. niger*, *A. glaucus*, *A. versicolor*, and *A. ustus*. Sinusitis or pneumonia accounts for 90% of invasive disease and develops after inhalation of *Aspergillus* spores in patients with prolonged neutropenia or other intense forms of immunosuppression.[44] *Aspergillus* spores can also enter the body after direct inoculation through intact skin or ingestion. Intestinal *Aspergillus* infection has led to fatal GI bleeding in severely immunocompromised patients. For that reason, health-food supplements made from natural substances that are not quality controlled for mold spores are contraindicated in immunocompromised patients. Dissemination of invasive disease can occur either as contiguous extension (such as erosion of paranasal sinus infection into the brain) or by hematogenous spread after the organism erodes into blood vessels.

Diagnosis

Clinical manifestations Symptoms and signs of invasive aspergillosis reflect the sites of infection.[45] Patients with pulmonary disease may have fever, dyspnea, tachypnea, cough, brown sputum, pleuritic chest pain, hemoptysis, wheezing, and rhonchi. Sinus infection may present as headache and sinus pressure or tenderness. Cutaneous disease may present as skin lesions (with or without eschar) at the site of primary inoculation or, in the case of secondary dissemination, in multiple noncontiguous sites. Invasive GI disease may present as abdominal discomfort or distention, peritonitis, small bowel obstruction, or melena; it can involve the liver, spleen, appendix, or peritoneum or cause bowel infarction. Liver transplant recipients are at particular risk for *Aspergillus* infection of the abdominal suture line or peritoneum.

Laboratory tests CT scans of the lungs and sinuses should be done in all patients suspected of having an invasive mold infection. Air crescents or so-called ground-glass halos around lung lesions on CT are considered pathognomonic for invasive aspergillosis. Patients with indeterminate pulmonary abnormalities (e.g., new infiltrates or nodules) should undergo bronchoscopy unless the lesions are peripheral. Transbronchial biopsy is performed with bronchoalveolar lavage if the platelet count and bleeding time indicate a low risk of bleeding. A patient with a solitary unilateral pulmonary nodule should undergo CT-guided fine-needle aspiration or in toto resection if the nodule is discrete and large enough.[46] CT or MRI scan of the head should be performed before surgical resection to confirm a lack of hematogenous dissemination.

Lung transplant recipients are at risk for localized infection of the bronchial anastomosis, which can result in circumferential anastomotic dehiscence with herniation of perianastomotic fat.[47] In the early stages of tracheobronchitis, onset of pulmonary symptoms may precede the appearance of radiographic abnormalities. Bronchoscopy will reveal lesions of the mainstem and segmental bronchi. Although most of the lesions are ulcerative, other lesions such as plaquelike lesions, pseudomembranes, or nodules occasionally develop. Disease can vary from localized tracheobronchitis discovered incidentally to extensive bronchial obstruction contributing to respiratory failure.

Opacified sinuses and sinuses containing air-fluid levels should be cultured by an otolaryngologist. However, sinuses showing mucosal thickening and viscous air-fluid levels are inflamed and provide low yields upon culture. If cultured by swab rather than by aspiration or biopsy, they have a predictably low yield of pathogens. If a CT scan demonstrates bony erosion of sinus walls, bone biopsy samples should be taken and examined histopathologically.

Although the diagnostic tests used on a given specimen depend on the body site from which the specimen was taken, specimens are usually submitted for culture, cytology, and histopathologic evaluation. Recovery of the organism in culture and demonstration of hyphae invading tissue make the diagnosis definitive. Growth occurs after 2 to 14 days of incubation. Hyphal cells or cell fragments should be visible with hematoxylin-eosin or methenamine-silver staining. Typical *Aspergillus* hyphal cells have frequent septations and join at 45° angles (dichotomous branching).

An enzyme-linked immunosorbent assay for the detection of *Aspergillus* galactomannan antigen in serum is commercially available, and has become an important tool for the early diagnosis of invasive aspergillosis.[48] In one study, the test had a sensitivity of 75% and a specificity of 100%.[49] The assay is increasingly being used on specimens of body fluids other than serum, including urine, bronchoalveolar lavage fluid, and CSF.[50]

Differential Diagnosis

Although invasive aspergillosis is the most common cause of pulmonary nodules in immunocompromised patients, in at least 10% of such cases the pathogen (or copathogen) will be a non-*Aspergillus* fungus or a bacterium (e.g., *Nocardia*, *Legionella*, or atypical *Mycobacterium*). Other conditions that must be excluded are septic emboli, pulmonary edema, diffuse alveolar hemorrhage, drug reactions, acute respiratory distress syndrome, idiopathic pneumonia syndrome, and bronchiolitis obliterans.

Treatment

Active infection The recommended agent for treatment of invasive aspergillosis is voriconazole.[51,52]

An alternative regimen is amphotericin B, either conventional or lipid formulations, in dosages equivalent to 1 mg/kg/day of standard amphotericin B. Caspofungin is approved by the FDA for patients who fail to tolerate or to improve with standard therapy (salvage treatment) [*see Table 6*].[53,54] Adjunctive measures include colony-stimulating factors, surgical resection of unilateral pulmonary nodules (in patients with no evidence of CNS dissemination), and granulocyte transfusions.

Prophylaxis Environmental and pharmaceutical prophylaxis options are used in patients who are profoundly neutropenic for 3 weeks or longer, because in such patients, invasive aspergillosis occurs with 5% to 10% frequency and mortality is high. Environmental prophylaxis includes high-efficiency particulate air (HEPA) filtration for allogeneic hematopoietic stem cell transplant recipients, rather than laminar airflow; in addition, severely immunocompromised patients should use high-efficiency respiratory-protection devices (e.g., N95 respirators)

when they leave their rooms if dust-generating activities are ongoing in the building.[55]

Pharmacologic antimold prophylaxis is not standardized. In a 2004 clinical trial comparing micafungin with fluconazole for fungal prophylaxis during the preengraftment period of hematopoietic stem cell transplantation, there was a trend toward fewer *Aspergillus* infections in micafungin-treated patients during the study period (one infection versus seven infections, $P = 0.08$).[18] Some centers begin pharmacologic prophylaxis at or before the onset of neutropenia in select patient populations.

Secondary prophylaxis should be started at the onset of neutropenia in patients with a history of aspergillosis who are entering major periods of immunosuppression, such as hematopoietic stem cell transplant recipients.[56] Voriconazole and caspofungin have been replacing amphotericin B for secondary prophylaxis in these patients; the lower toxicity of these agents permits wider use of drug prophylaxis, but the agents are being used in off-label fashion. Adjunctive measures such as nonmyeloablative conditioning procedures and granulocyte transfusions may be used.[57] In general, however, granulocyte transfusions are not effective for secondary prevention of fungal infection during profound neutropenia.

Prognosis

In the 1990s, the overall case-fatality rate for aspergillosis was 58%. Mortality was highest (> 85%) in bone marrow transplant recipients and patients with CNS or disseminated aspergillosis.[43,44] Death often results from the erosion of blood vessels by fungal invasion, with massive hemoptysis or other fatal hemorrhage.

Zygomycosis

EPIDEMIOLOGY

The agents of zygomycosis (a broad term that includes the better-known mucormycosis) cause disease that closely resembles invasive aspergillosis. However, this group of organisms is taxonomically different, belonging to the phylum Zygomycota. The principal pathogens include *Rhizopus*, *Mucor*, and *Rhizomucor;* occasional pathogens include *Absidia*, *Cunninghamella*, *Mortierella*, *Saksenaea*, *Cokeromyces*, and *Apophysomyces.*

Different populations are at risk for zygomycosis at different sites. Cutaneous infection most often develops in patients who are severely ill from trauma; only 50% have systemic immunocompromising disease. Conditions that increase risk for pulmonary, rhinocerebral, and disseminated zygomycosis include poorly controlled diabetes (particularly ketoacidosis) and immunosuppression (e.g., neutropenia, sustained immunosuppressive therapy, or long-term prednisone use). Iron overload, such as that from repeated dialysis or blood transfusions, increases susceptibility to zygomycosis because iron is an essential element for the growth and metabolism of Zygomycetes. Infection may occur in patients receiving the iron chelator deferoxamine to reduce iron overload, because Zygomycetes can readily use iron even when it is bound to deferoxamine.

Persons at risk for zygomycosis of the GI tract are those with some form of severe protein-calorie malnutrition, such as uremia, kwashiorkor, or chronic diarrhea. In the United States, a *Mucor* liver abscess developed in a transplant recipient who consumed health-food supplements made from natural substances that were not quality controlled for mold spores.[58]

ETIOLOGY

Cutaneous infection usually occurs by traumatic inoculation; it is preceded by skin trauma in greater than 80% of cases. Sources of trauma have included surgery, burns, motor vehicle accidents, knives, and even insect bites. A well-publicized series of infections occurred between 1978 and 1980 from elasticized surgical dressings that were found to be contaminated with *Rhizopus*.[59]

Pulmonary, rhinocerebral, and disseminated zygomycosis are acquired via airborne spores deposited in the nasal turbinates or pulmonary alveoli. GI tract disease follows ingestion of fungal spores in moldy food.

PATHOGENESIS

Onset of cutaneous infection usually occurs a few days after the initiating trauma. The pace of cutaneous illness varies: necrosis may spread slowly over weeks, resembling an arterial insufficiency ulcer, or progress rapidly over days, resembling synergistic gangrene. Secondary cutaneous dissemination is unusual, except perhaps in burn victims.

Infection that starts as colonization in the nasal turbinates or paranasal sinuses may extend contiguously. When infection extends directly along veins that drain the orbit and facial tissues, cavernous sinus thrombosis results. With posterior extension, frontal lobe necrosis with brain abscess ensues. Zygomycetes can also enter the CNS by hematogenous spread.

DIAGNOSIS

Clinical Manifestations

Cutaneous disease can occur anywhere on the body, but it is most often found on the extremities.[60] Cutaneous lesions are usually characterized by central dermal necrosis surrounded by a margin of red-to-purple edematous cellulitis, sometimes with bullae indicating epidermal necrosis. Mold may be visible on the edge of a wound. If dermal necrosis is present, pain is common. Fever may be present or absent.

Pulmonary disease typically displays the usual pneumonia symptoms of fever, dyspnea, pleuritic chest pain, hoarseness, and gross hemoptysis. Rhinocerebral disease may begin with fever, leukocytosis, ketoacidosis, facial pain, sinus drainage, or headache. A black eschar on the nasal or palatal mucosa should immediately raise suspicion for rhinocerebral disease. With orbital extension, clinical findings may include orbital swelling or cellulitis, extraocular muscle paresis, proptosis, and chemosis.

Rhinocerebral disease may result in cranial nerve abnormalities, most commonly a seventh-nerve palsy. In patients with cavernous sinus thrombosis, venous engorgement from outflow obstruction is noted in the retina, conjunctiva, and eyelid. Paresis of cranial nerves III, IV, and VI may result in diplopia. Less often, proptosis, meningeal irritation, epistaxis, and involvement of cranial nerve V may occur.

Laboratory Tests

Most patients with pulmonary or disseminated disease will have abnormal chest imaging results, but there is no specific radiographic appearance or lobar predilection. In neutropenic patients with early disease, chest films may show little infiltrate, perhaps because these patients may lack inflammatory cells to cause infiltrate. If pulmonary disease progresses to tissue necrosis and hemorrhage, however, the radiographic changes will become obvious.

A CT scan of the head is indicated in patients with clinical signs of rhinocerebral disease. Scans may show bone destruc-

tion, a brain infarct, or a space-occupying brain lesion and can be used to guide surgical intervention.

Definitive diagnosis is made by recovery of Zygomycetes in culture. Growth occurs after 2 to 5 days of incubation. Wound surfaces can be swabbed for fungal culture, but specimens should also be taken from rapidly spreading infections, nonwound cutaneous lesions, and lesions such as a black eschar on the hard palate or nasal mucosa.

Specimens should be immediately examined with potassium hydroxide, calcofluor white stains, or both. Histologic studies should be done by using hematoxylin-eosin and methenamine-silver staining. In cutaneous disease, hyphae are seen in the middle of an acute neutrophilic infiltrate. Zygomycetes hyphal cells should have infrequent septations, be of variable diameter, join at 90° angles, and twist in ribbonlike fashion upon one another. In the uncommon instance where all those histologic findings are obvious, a tentative pathologic diagnosis of Zygomycetes infection can be made. However, the usual situation is that the tangle of fungal hyphae cannot be distinguished from *Aspergillus*.

Histology may be the only basis for diagnosis, however, given that Zygomycetes grow less frequently in culture than other filamentous molds. One reason is that Zygomycota mold cells are large, because they have few or no septations (i.e., they are pauciseptate or aseptate). Consequently, when the microbiology technologist uses a razor blade to mince tissue being prepared for fungal culture, one slice through the cell wall can lead to leakage of cytoplasm and death, leaving no intact cells for growth. Homogenization of tissue submitted for culture also leads to a high rate of false negative results.

DIFFERENTIAL DIAGNOSIS

Cutaneous zygomycosis can look like ecthyma gangrenosum, which is more often caused by *Aspergillus* and *Pseudomonas.* Pulmonary zygomycosis should be differentiated from aspergillosis, pneumonia caused by other less common bacteria and molds, and pulmonary embolism. Rhinocerebral zygomycosis can resemble orbital tumors, cavernous sinus thrombosis from *Staphylococcus* infection, or aspergillosis.

TREATMENT

Active Infection

Zygomycosis is currently treated with maximal-dosage intravenous amphotericin, 1.0 to 1.5 mg/kg daily for 2 to 12 weeks or for the duration of clinical disease. In addition to intravenous infusions, intracavitary or topical amphotericin may be required for CNS disease; topical amphotericin can be applied to cutaneous disease. Posaconazole, an orally administered triazole, is effective against zygomycosis[61] but is not yet approved by the FDA.

Treatment should include adjunctive measures such as debridement of any adherent mycelial masses at accessible sites of infection. Debridement should be repeated as often as every other day until cultures of the debrided tissues are negative. Some patients may be candidates for granulocyte transfusions or hyperbaric oxygen.

Prophylaxis

Pharmacologic prophylaxis of zygomycosis infections is unusual because the frequency of infection with this group of organisms is low even in high-risk patient groups. Additionally, the medication for prophylaxis is intravenous amphotericin, which is toxic and is expensive when the lipid formulations are used; however, posaconazole may have a role in the future. Topical or systemic preemptive therapy may be appropriate for individual patients, such as burn victims in whom surveillance cultures of wounds demonstrate fungal hyphae.

PROGNOSIS

Hosts with zygomycoses may not be as immunocompromised as hosts with some of the more common opportunistic mold infections; hence, zygomycoses may be associated with better survival.

Fusariosis

EPIDEMIOLOGY

Fusarium is a soil-borne mold that can cause human disease if a portal of entry is available, such as skin, nails, the airways, or the eyes. Species responsible for the majority of human disease are *F. solani*, *F. moniliforme*, and *F. oxysporum*. The number of organisms and the depth of penetration may determine the course of infection. Eye infection can occur after contamination of soft-contact-lens paraphernalia through inadequate washing of soil-borne spores from the hands or contamination of the lens during use; under windy conditions, aerosolized spores can be blown into the eye.

The single most important host risk factor for invasive fusariosis in immunocompromised patients is neutropenia.[62,63] Good clinical outcomes have clearly been linked to recovery from neutropenia, and fatal relapse of disease has been linked to recurrent neutropenia.[64] Other immune system defects linked to infection include heatstroke, systemic corticosteroid use, and topical steroid use at the site of colonization.[65]

Some *Fusarium* species produce secondary metabolites called mycotoxins. Under favorable environmental conditions, those species can grow on grains and contaminate them with mycotoxins. Consumption of mycotoxin-laden foodstuffs can result in both infection and poisoning—especially in malnourished persons, who have enhanced susceptibility to the toxin. An important historical example is an outbreak of toxic alimentary aleukia that occurred in the Soviet Union at the end of World War II. Starving people harvested grains that had been left in the fields over the winter and had become heavily contaminated with *Fusarium* mycotoxin. Those who ate more than 2 kg of food made from the grains suffered an illness that progressed from mucosal inflammation to pancytopenia to mucosal necrosis with fatal bacterial infection. Hundreds of thousands of people died of the illness.

ETIOLOGY

Opportunistic *Fusarium* infection begins by colonization of broken skin, the upper airways, sinuses, or conjunctivae. Reported causes of skin-barrier breakdown include catheters (central venous and peritoneal dialysis), burns, trauma, surgical procedures (coronary bypass), and denuding of skin from infection (as from varicella). When the host has an immune system defect, tissue inflammatory reactions at the site of colonization are reduced, enabling the organism to grow and spread.

PATHOGENESIS

The mold form of *Fusarium* consists of septate hyphae with branches at acute angles. Growth of hyphae in skin lesions may result in subcutaneous granulomas. Once infection has pene-

Table 7 Treatment of Infections Caused by *Fusarium* Species

Infection Site	*Drug*	*Dosage*	*Relative Efficacy*	*Comments*
Invasive fusariosis				Surgery may be needed
	Amphotericin	1.0–1.5 mg/kg I.V. daily for 2–12 wk	First-choice agent	Use lipid formulation, not generic, for patients with nephrotoxicity or infusion toxicity
	Voriconazole	6 mg/kg q. 12 hr for first 24 hr, then 4 mg/kg q. 12 hr; when oral medication tolerated, 200 mg p.o., q. 12 hr; continue treatment for several months, or for minimum 2 wk after recovery from neutropenia, with resumption if patient becomes neutropenic again after an infection	Alternative agent	Oral form taken ≥ 1 hr before or after meal
Keratitis/corneal ulcer	Natamycin 5% suspension	1 drop q. 1–2 hr for first 2 days, then decrease gradually over 3–6 wk	First-choice agent	May require surgery
Skin lesions (as part of invasive disease)				Treat primarily as invasive disease; continue topical therapy until lesions are dry scabs
	Nystatin	Cream twice daily to affected areas	Adjunct to I.V. amphotericin	—
	Amphotericin	Topical compound twice daily to affected areas	Adjunct to I.V. amphotericin	Compound locally

trated into the lumen of blood vessels, hyphae can disseminate through the body to produce embolic infection whose hallmark is necrotic skin lesions and, less often, endophthalmitis or brain abscess.[66]

DIAGNOSIS

Clinical Manifestations

Manifestations of focal *Fusarium* infection vary with the organ involved. Keratitis presents as a corneal ulcer; if therapy is delayed, infection may progress to endophthalmitis. Primary cutaneous infections present in many forms, including ulcers, necrosis, pustules, vesicles, painful nodules, mycetomas, and panniculitis. Onychomycosis presents as a milky-white area in the nail, which may spread until the entire nail becomes opaque; the free border of the nail may become thickened.

With disseminated infection, secondary cutaneous lesions may be the first characteristic finding. These may take the form of multiple erythematous subcutaneous nodules, painful lesions with progressive central infarction, or target lesions with ecthyma gangrenosum–like centers surrounded by a rim of erythema.[67]

Laboratory Tests

Fusarium is the only opportunistic mold that can be easily recovered from the bloodstream; blood cultures will be positive in 40% to 60% of patients with disseminated infection. Many clinical laboratories identify *Fusarium* to only the genus level; speciation would not alter the prognosis or treatment. Patients with fever and neutropenia should have cultures of blood, as well as biopsy and culture from any body sites suspicious for infection. Because *Fusarium* often colonizes nonsterile tissues, the mycologic data should be combined with histopathologic data to distinguish colonization from tissue infection. Biopsies of suspicious skin lesions should be taken from the center of the lesion and should extend to subcutaneous fat, because mold organisms easily invade blood vessels of the dermis and subcutis, resulting in an overlying ischemic cone. Part of the biopsy specimen, or a second biopsy specimen, should be sent for culture.

DIFFERENTIAL DIAGNOSIS

In addition to *Fusarium,* a variety of other organisms can be responsible for new skin lesions in an immunocompromised host; these include *Aspergillus*, other molds, *Pseudomonas,* and varicella.

TREATMENT

Fusarium species are relatively resistant to treatment with antifungal agents, and sensitivity testing for *Fusarium* is not standardized. Fluconazole, itraconazole, and 5-fluorocytosine have no activity against *Fusarium* species, and amphotericin, ketoconazole, miconazole, and terbinafine have limited activity.

Active Infection

Invasive disease is usually treated with high-dose systemic amphotericin [*see Table 7*]. Topical amphotericin or nystatin cream is sometimes added to the systemic amphotericin regimen in attempts to increase local delivery of drug. Voriconazole is approved by the FDA for salvage therapy for *Fusarium* infections. Corneal ulcers caused by *Fusarium* infection are treated with topical amphotericin or natamycin.

Because outcome is clearly linked to recovery from neutropenia, adjunctive measures are essential.[64] Colony-stimulating-factor injections and neutrophil transfusions have been used to increase the neutrophil count. Further episodes of electively induced myelosuppression should be avoided because there is a high incidence of recurrence and multiple-organ failure in patients with subsequent neutropenic episodes.

Surgical treatment has an important role in the management of localized infection; examples include enucleation of an affected eye, wide excision, and nail removal. In an extreme situation, amputation of a limb can be used to cure fusarial soft tissue infections.

Prophylaxis

Prevention of primary *Fusarium* infection involves the same environmental measures that would be used to prevent other mold infections in patients at risk. In general, patients with chronic neutropenia or those about to receive myelosuppressive

therapy should wash their hands regularly and avoid contact with soil or inhalation of dust. Myelosuppressive therapy should be given in an inpatient environment with air filtration to reduce the number of aerosolized spores.

Secondary prophylaxis is required when a patient with known recent *Fusarium* infection has an episode of neutropenia. A systemic antifungal agent with activity against *Fusarium* is used at treatment doses. Efforts to shorten the duration of neutropenia with colony-stimulating factors are sometimes employed. The use of neutrophil transfusions as secondary prophylaxis for patients without fever or skin lesions is difficult to justify and is rarely done.

Infection by Dematiaceous Fungi

EPIDEMIOLOGY

Dematiaceous fungi live in soil and rotting wood. Current taxonomy has more than 100 species of dematiaceous fungi spread out among more than a dozen different families, falling under several classes and orders. All contain melanin pigment in the walls of hyphae or spores, resulting in a color that is dark brown, greenish gray, or black. They are an important group of opportunists, because some species have a tendency to cause CNS infection.[68] Species that appear to be particularly neurotropic include *Cladophialophora*, *Cladosporium*, *Bipolaris*, *Exserohilum*, *Dactylaria*, *Exophiala*, and *Fonsecaea*.

Most reports of infection by dematiaceous fungi emphasize cerebral disease in an immunocompromised patient.[69] However, there is a growing literature describing these infections in specific at-risk patient populations (such as solid-organ-transplant recipients) and in immunocompetent patients, who may experience nasal polyps and an obstructive fungal ball destroying the sinuses.[70,71]

ETIOLOGY

Injury to the skin can implant dematiaceous fungus in subcutaneous tissues; inhalation of spores can cause a minimal infection in the lung or sinuses. Sinus infection can erode through bone into the brain, or unrecognized pulmonary infection can metastasize to the brain.

PATHOGENESIS

Primary cutaneous disease usually exhibits one of three histopathologic patterns: solid granulomas, stellate abscesses, or cavitary lesions. Cystic lesions probably start as solid granulomas, which undergo focal necrosis to form stellate abscesses. The abscesses can then coalesce into a unilocular fluctuant abscess.

DIAGNOSIS

Clinical Manifestations

Dematiaceous fungi cause primarily skin and soft tissue infections.[70] Analogous syndromes seen in tropical and subtropical areas of the world include mycetoma (tumorous growth of the skin with tissue granules) and chromoblastomycosis (verrucous skin infection with sclerotic bodies). Immunocompetent persons develop an asymptomatic initial lesion: a discrete, well-encapsulated, erythematous to flesh-colored plaque or nodule. The nodule center tends to become necrotic, forming a localized, encapsulated abscess. In immunocompromised persons, subcutaneous lesions may not encapsulate and may spread contiguously or systemically.

Systemic infections may present as unexplained fever during neutropenia that persists despite administration of broad-spectrum antibiotics. Metastatic disease may involve the lungs, joints, esophagus, heart, peritoneum, or bone, but the favored site is the CNS.

Laboratory Tests

Histopathology and culture of a tissue specimen are generally necessary to identify an infection as dematiaceous. Fungi are visible in cells lining an abscess, in the lumen of an abscess, or free in the granulation tissue of a cyst wall. The fungi appear as yeastlike cells that can be solitary or arranged in short chains, pseudohyphal-like fungal elements, or septate hyphae of varying lengths. The hyphae may differ from *Aspergillus*, *Fusarium*, or the Zygomycetes by having thicker walls, irregular diameter, variable branching, or constrictions at the septations. Because the degree of visible pigment can vary between patients infected with the same species, a melanin stain such as Masson-Fontana can be used to reveal cell wall pigment in tissue sections.

In patients with CNS involvement, dematiaceous infection often presents as rim-enhancing brain abscesses on CT scan.

DIFFERENTIAL DIAGNOSIS

On histopathologic examination, the septate fungal hyphae of dematiaceous fungi bear close resemblance to *Aspergillus.* The pathologist will usually label such specimens as *Aspergillus*-like, because less than 0.1% of them will culture out as a dematiaceous fungus. The Masson-Fontana melanin stain can discriminate dematiaceous fungus from *Aspergillus* but is rarely used in histopathology laboratories, so the distinction is usually made by culture.

TREATMENT

Complete surgical excision appears critical in the management of cutaneous lesions. If the entire lesion is removed in an immunocompetent host, antifungal therapy may not be needed. Disseminated disease requires systemic antifungal medication, sometimes combined with surgery. Amphotericin, 0.8 to 1.2 mg/kg I.V. daily for 2 to 12 weeks, is typically used, but the choice of medications should be evaluated on an individual basis. The new triazoles and echinocandins may come to replace amphotericin for this purpose.[72]

The duration of therapy is determined by the immune competence of the host, the site and extent of involvement, disease responsiveness, and the susceptibility of the organism. Neutrophil transfusions may be added for patients with systemic infection and low levels of circulating neutrophils, but no controlled studies of this approach have been performed.

The author has received grant or research support from Schering-Plough Research Corporation, Inc., Astellas Pharma US, Inc., and Enzon Pharmaceuticals, Inc.; participates in the speakers' bureaus for Schering-Plough Research Corporation, Inc., Pfizer, Inc., and Astellas Pharma US, Inc.; and is a consultant for Schering-Plough Research Corporation, Inc., and Astellas Pharma US, Inc.

References

1. Marr KA: Invasive *Candida* infections: the changing epidemiology. Oncology (Williston Park) 18(14 suppl 13):9, 2004

2. Snydman DR: Shifting patterns in the epidemiology of nosocomial *Candida* infections. Chest 123(5 suppl):500s, 2003

3. Abi-Said D, Anaissie E, Uzun O, et al: The epidemiology of hematogenous candidiasis caused by different *Candida* species. Clin Infect Dis 24:1122, 1997

4. Wright WL, Wenzel RP: Nosocomial *Candida*: epidemiology, transmission, and prevention. Infect Dis Clin North Am 11:411, 1997

5. Lundstrom T, Sobel J: Nosocomial candiduria: a review. Clin Infect Dis 32:1602, 2001

6. Darouiche RO: Oropharyngeal and esophageal candidiasis in immunocompromised patients: treatment issues. Clin Infect Dis 26:259, 1998

7. Kauffman CA: Candiduria: diagnostic and treatment conundrums. Curr Treat Options Infect Dis 4:513, 2002

8. Sobel JD, Kauffman CA, McKinsey D, et al: Candiduria: a randomized, double-blind study of treatment with fluconazole and placebo. The National Institute of Allergy and Infectious Diseases (NIAID) Mycoses Study Group. Clin Infect Dis 30:19, 2000

9. Edwards JE Jr, Bodey GP, Bosden RA, et al: International Conference for the Development of a Consensus on the Management and Prevention of Severe Candidal Infections. Clin Infect Dis 25:43, 1997

10. Kontoyiannis DP, Luna MA, Samuels BI, et al: Hepatosplenic candidiasis: a manifestation of chronic disseminated candidiasis. Infect Dis Clin North Am 14:721, 2000

11. Practice guidelines for the treatment of candidiasis. Infectious Diseases Society of America. Clin Infect Dis 30:662, 2000

12. Walsh TJ, Pappas P, Winston DJ, et al: Voriconazole compared with liposomal amphotericin B for empirical antifungal therapy in patients with neutropenia and persistent fever. N Engl J Med 346:225, 2002

13. Kartsonis NA, Saah A, Lipka CJ, et al: Second-line therapy with caspofungin for mucosal or invasive candidiasis: results from the caspofungin compassionate-use study. J Antimicrob Chemother 53:878, 2004

14. Arathoon EG, Gotuzzo E, Noriega LM, et al: Randomized, double-blind, multicenter study of caspofungin versus amphotericin B for treatment of oropharyngeal and esophageal candidiases. Antimicrob Agents Chemother 46:451, 2002

15. Mora-Duarte J, Betts R, Rotstein C, et al: Comparison of caspofungin and amphotericin B for invasive candidiasis. Caspofungin Invasive Candidiasis Study Group. N Engl J Med 347:2020, 2002

16. Walsh TJ, Teppler H, Donowitz GR, et al: Caspofungin versus liposomal amphotericin B for empirical antifungal therapy in patients with persistent fever and neutropenia. N Engl J Med 351:1391, 2004

17. Micafungin (Mycamine) for fungal infections. Med Lett Drugs Ther 47:51**,** 2005

18. van Burik JA, Ratanatharathorn V, Stepan DE, et al: Micafungin versus fluconazole for prophylaxis against invasive fungal infections during neutropenia in patients undergoing hematopoietic stem cell transplantation. National Institute of Allergy and Infectious Diseases Mycoses Study Group. Clin Infect Dis 39:1407, 2004

19. Perfect JR, Casadevall A: Cryptococcosis. Infect Dis Clin North Am 16:837, 2002

20. Mirza SA, Phelan M, Rimland D, et al: The changing epidemiology of cryptococcosis: an update from population-based active surveillance in 2 large metropolitan areas, 1992–2000. Clin Infect Dis 36:789, 2003

21. Husain S, Wagener MM, Singh N: *Cryptococcus neoformans* infection in organ transplant recipients: variables influencing clinical characteristics and outcome. Emerg Infect Dis 7:375, 2001

22. Rex JH, Larsen RA, Dismukes WE, et al: Catastrophic visual loss due to *Cryptococcus neoformans* meningitis. Medicine (Baltimore) 72:207, 1993

23. Practice guidelines for the management of cryptococcal disease. Mycoses Study Group Subproject, Infectious Diseases Society of America. Clin Infect Dis 30:710, 2000

24. Singh N, Barnish MJ, Berman S, et al: Low-dose fluconazole as primary prophylaxis for cryptococcal infection in AIDS patients with CD4 cell counts of ≤ 100/mm^3: demonstration of efficacy in a prospective, multicenter trial. Clin Infect Dis 23:1282, 1996

25. Singh N, Lortholary O, Alexander BD, et al: An immune reconstitution syndrome–like illness associated with *Cryptococcus neoformans* infection in organ transplant recipients. Clin Infect Dis 40:1756, 2005

26. Diagnosis and management of increased intracranial pressure in patients with AIDS and cryptococcal meningitis. NIAID Mycoses Study Group and AIDS Cooperative Treatment Groups. Clin Infect Dis 30:47, 2000

27. Park MK, Hospenthal DR, Bennett JE: Treatment of hydrocephalus secondary to cryptococcal meningitis by use of shunting. Clin Infect Dis 28:629, 1999

28. Pappas PG, Perfect JR, Cloud GA, et al: Cryptococcosis in human immunodeficiency virus–negative patients in the era of effective azole therapy. Clin Infect Dis 33:690, 2001

29. Morris A, Lundgren JD, Masur H, et al: Current epidemiology of *Pneumocystis* pneumonia. Emerg Infect Dis 10:1713**,** 2004

30. Gordon SM, LaRosa SP, Kalmadi S, et al: Should prophylaxis for *Pneumocystis carinii* pneumonia in solid organ transplant recipients ever be discontinued? Clin Infect Dis 28:240, 1999

31. Ng VL, Yajko DM, Hadley WK: Extrapulmonary pneumocystosis. Clin Microbiol Rev 10:401, 1997

32. Predictors for failure of *Pneumocystis carinii* pneumonia prophylaxis. Multicenter AIDS Cohort Study. JAMA 273:1197, 1995

33. Thomas CF Jr, Limper AH: *Pneumocystis* pneumonia. N Engl J Med 350:2487, 2004

34. Consensus statement on the use of corticosteroids as adjunctive therapy for *Pneumocystis* pneumonia in the acquired immunodeficiency syndrome. National Institutes of Health–University of California Expert Panel for Corticosteroids as Adjunctive Therapy for Pneumocystis Pneumonia. N Engl J Med 323:1500, 1990

35. A randomized trial of the discontinuation of primary and secondary prophylaxis against *Pneumocystis carinii* pneumonia after highly active antiretroviral therapy in patients with HIV infection. Grupo de Estudio del SIDA. N Engl J Med 344:159, 2001

36. Dworkin MS, Williamson J, Jones JL, et al: Prophylaxis with trimethoprim-sulfamethoxazole for human immunodeficiency virus–infected patients: impact on risk for infectious diseases. Clin Infect Dis 33:393, 2001

37. Hughes WT: Use of dapsone in the prevention and treatment of *Pneumocystis carinii* pneumonia: a review. Clin Infect Dis 27:191, 1998

38. Cockrill BA, Hales CA: Allergic bronchopulmonary aspergillosis. Annu Rev Med 50:303, 1999

39. Patterson R, Greenberger PA, Radin RC, et al: Allergic bronchopulmonary aspergillosis: staging as an aid to management. Ann Intern Med 96:286, 1982

40. Nepomuceno IB, Esrig S, Moss RB: Allergic bronchopulmonary aspergillosis in cystic fibrosis: role of atopy and response to itraconazole. Chest 115:364, 1999

41. Stevens DA, Schwartz HJ, Lee JY, et al: A randomized trial of itraconazole in allergic bronchopulmonary aspergillosis. N Engl J Med 342:756, 2000

42. Wark PA, Gibson PG, Wilson AJ: Azoles for allergic bronchopulmonary aspergillosis associated with asthma. Cochrane Database Syst Rev (3):CD001108, 2004

43. Wald A, Leisenring W, van Burik J-A, et al: Epidemiology of *Aspergillus* infections in a large cohort of patients undergoing bone marrow transplantation. J Infect Dis 175:1459, 1997

44. Lin S-J, Schranz J, Teutsch SM: Aspergillosis case-fatality rate: systematic review of the literature. Clin Infect Dis 32:358, 2001

45. Patterson TF, Kirkpatrick WR, White M, et al: Invasive aspergillosis: disease spectrum, treatment practices, and outcomes. I3 *Aspergillus* Study Group. Medicine (Baltimore) 79:250, 2000

46. Yeghen T, Kibbler CC, Prentice HG, et al: Management of invasive pulmonary aspergillosis in hematology patients: a review of 87 consecutive cases at a single institution. Clin Infect Dis 31:859, 2000

47. Sole A, Morant P, Salavert M, et al: *Aspergillus* infections in lung transplant recipients: risk factors and outcome. Valencia Lung Transplant Group. Clin Microbiol Infect 11:359, 2005

48. Mennink-Kersten MA, Donnelly JP, Verweij PE: Detection of circulating galactomannan for the diagnosis and management of invasive aspergillosis. Lancet Infect Dis 4:349, 2004

49. Rovira M, Jimenez M, De La Bellacasa JP, et al: Detection of *Aspergillus* galactomannan by enzyme immunoabsorbent assay in recipients of allogeneic hematopoietic stem cell transplantation: a prospective study. Transplantation 77:1260, 2004

50. Klont RR, Mennink-Kersten MA, Verweij PE: Utility of *Aspergillus* antigen detection in specimens other than serum specimens. Clin Infect Dis 39:1467, 2004

51. Mouas H, Lutsar I, Dupont B, et al: Voriconazole for invasive bone aspergillosis: a worldwide experience of 20 cases. Voriconazole/Bone Invasive Aspergillosis Study Group. Clin Infect Dis 40:1141, 2005

52. Herbrecht R, Denning DW, Patterson TF, et al: Voriconazole versus amphotericin B for primary therapy of invasive aspergillosis. N Engl J Med 347:408, 2002

53. Benson CA, Kaplan JE, Masur H, et al: Treating opportunistic infections among HIV-exposed and infected adults and adolescents: recommendations from CDC, the National Institutes of Health, and the Infectious Diseases Society of America. MMWR Recomm Rep 53(RR-15):1, 2004

54. Maertens J, Raad I, Petrikkos G, et al: Efficacy and safety of caspofungin for treatment of invasive aspergillosis in patients refractory to or intolerant of conventional antifungal therapy. Caspofungin Salvage Aspergillosis Study Group. Clin Infect Dis 39:1563, 2004

55. Tablan OC, Anderson LJ, Besser R, et al: Guidelines for preventing health-care–associated pneumonia, 2003: recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee. MMWR Recomm Rep 53(RR-3):1, 2004

56. Offner F, Cordonnier C, Ljungman P, et al: Impact of previous aspergillosis on the outcome of bone marrow transplantation. Clin Infect Dis 26:1098, 1998

57. Hermann S, Klein SA, Jacobi V, et al: Older patients with high-risk fungal infections can be successfully allografted using non-myeloablative conditioning in combination with intensified supportive care regimens. Br J Haematol 113:446, 2001

58. Oliver MR, Van Voorhis WC, Boeckh M, et al: Hepatic mucormycosis in a bone marrow transplant recipient who ingested naturopathic medicine. Clin Infect Dis 22:521, 1996

59. Gartenberg G, Bottone EJ, Keusch GT, et al: Hospital-acquired mucormycosis (*Rhizopus rhizopodiformis*) of skin and subcutaneous tissue: epidemiology, mycology and treatment. N Engl J Med 299:1115, 1978

60. Prabhu RM, Patel R: Mucormycosis and entomophthoramycosis: a review of the clinical manifestations, diagnosis and treatment. Clin Microbiol Infect 10(suppl 1):31, 2004

61. Keating GM: Posaconazole. Drugs 65:1553, 2005

62. Gamis AS, Gudnason T, Giebink GS, et al: Disseminated infection with *Fusarium* in recipients of bone marrow transplants. Rev Infect Dis 13:1077, 1991

63. Boutati EI, Anaissie EJ: *Fusarium*, a significant emerging pathogen in patients with hematologic malignancy: ten years' experience at a cancer center and implications for management. Blood 90:999, 1997

64. Invasive *Fusarium* infections: a retrospective survey of 31 cases. French 'Groupe d'Etudes des Mycoses Opportunistes' GEMO. J Med Vet Mycol 35:107, 1997

65. Sampathkumar P, Paya CV: *Fusarium* infection after solid-organ transplantation. Clin Infect Dis 32:1237, 2001

66. Louie T, el Baba F, Shulman M, et al: Endogenous endophthalmitis due to *Fusarium*: case report and review. Clin Infect Dis 18:585, 1994

67. Prins C, Chavaz P, Tamm K, et al: Ecthyma gangrenosum–like lesions: a sign of disseminated *Fusarium* infection in the neutropenic patient. Clin Exp Dermatol 20:428, 1995

68. Fleming RV, Walsh TJ, Anaissie EJ: Emerging and less common fungal pathogens. Infect Dis Clin North Am 16:915, 2002

69. Osiyemi OO, Dowdy LM, Mallon SM, et al: Cerebral phaeohyphomycosis due to a novel species: report of a case and review of the literature. Transplantation 71:1343, 2001

70. Singh N, Chang FY, Gayowski T, et al: Infections due to dematiaceous fungi in organ transplant recipients: case report and review. Clin Infect Dis 24:369, 1997

71. Zieske LA, Kopke RD, Hamill R: Dematiaceous fungal sinusitis. Otolaryngol Head Neck Surg 105:567, 1991

72. Al-Abdely HM: Management of rare fungal infections. Curr Opin Infect Dis 17:527, 2004

XXXIX INFECTIONS DUE TO MYCOBACTERIUM LEPRAE AND NONTUBERCULOUS MYCOBACTERIA

Michael K. Leonard, Jr., m.d.
Henry M. Blumberg, m.d., f.a.c.p.

Mycobacterium leprae infection (i.e., leprosy) is a disease that has been recognized—and has often been misunderstood—since ancient times. Nontuberculous mycobacteria (NTM) (i.e., mycobacteria other than *M. tuberculosis* complex [*M. tuberculosis, M. bovis, M. africanum,* and *M. microti*] and *M. leprae*) have been recognized to cause human disease since at least the 1950s. The emergence of the AIDS pandemic and the development of newer culture methodologies and molecular diagnostic tools have brought about increased interest in the epidemiology, diagnosis, and treatment of human infections from NTM.

More than 100 species of NTM have been identified; approximately 50 of these may be pathogenic for humans, causing a broad spectrum of disease.[1] Most NTM organisms are readily recovered from the environment, including environmental and drinking water, soil, and aerosols.[1,2] NTM organisms are important environmental pathogens that can cause a broad spectrum of diseases. NTM disease usually develops in immunocompromised hosts, such as development of *M. avium* complex [MAC] infection and other NTM infections in persons with HIV infection; but occasionally, NTM disease does occur in immunocompetent persons, such as those with underlying lung disease. Other types of NTM infection are skin and soft tissue infections from rapidly growing mycobacteria in postoperative patients and *M. marinum* infection after aquatic exposure. NTM infections also occasionally develop in otherwise healthy patients. This chapter covers both *M. leprae* and selected NTM organisms, including MAC; *M. kansasii*; *M. marinum;* and rapidly growing mycobacteria such as *M. chelonae*, *M. fortuitum,* and *M. abscessus*. Discussion of other NTM pathogens is beyond the scope of this chapter, but these other organisms are discussed in several reviews.[1,2-5] *M. ulcerans*, the causative agent of Buruli ulcer, which is common only in children in rural tropical areas, primarily in West Africa (i.e., Ghana, Côte d'Ivoire, and Benein), is also discussed elsewhere.[6,7]

Leprosy (*M. leprae* Infection)

A classic scourge described in ancient medical texts and the Bible, leprosy (Hansen disease) is a mycobacterial disease caused by *M. leprae* that affects the skin and peripheral nerves and can lead to severe disfigurement. Worldwide, about one to two million people are affected, with 763,917 new cases reported in 2002.[8] In the United States, however, fewer than 100 new cases are diagnosed each year, almost all of them in immigrants from endemic areas.[9] Secondary transmission from imported cases has not been recognized. Prompt recognition is imperative to limit morbidity from irreversible nerve damage. The disease is curable with multidrug therapy.

M. leprae has two unique properties: it is thermolabile, growing best at 27° to 30° C, and it divides extremely slowly—the generation time is 12 to 14 days (in the mouse footpad model; the organism cannot be cultured in artificial media). Because of this slow growth, the incubation period in humans is long: a minimum of 2 to 3 years, with the average incubation time thought to be 5 to 7 years and the maximum time as long as 40 years. Because of the long and variable incubation period of the disease, its epidemiology and pathophysiology remain incompletely understood. Humans are the principal reservoir for *M. leprae*, but leprosy has also been found in armadillos and primates.[9] The main modes of disease transmission are by aerosolized droplets from lepromatous persons and, much less commonly, by direct skin contact.[9] Transmission is thought to occur by the respiratory route because large numbers of bacilli are often found in the nasal discharge of untreated patients with multibacillary disease [*see* Diagnosis, Laboratory Studies, *below*]. Although these are believed to be the main modes of transmission, many patients have no identifiable contacts.[10]

Host genetics are thought to play an important role in both development of disease in exposed persons and the pattern of disease that develops. The vast majority of the world's population is not susceptible to leprosy; however, familial clustering of leprosy has been demonstrated, and twin studies have revealed high concordance rates.[10] Susceptibility appears to be governed at least partly by the *nramp1* gene, which controls susceptibility to mycobacteria in mice. A human leukocyte antigen (HLA) association appears to play a role in the clinical spectrum of disease, with the HLA-DR3 genotype overrepresented in tuberculoid leprosy and the HLA genotype DQ1 or MT1 more often seen in lepromatous disease.[10]

Diagnosis

Clinical Manifestations and Classification

Because *M. leprae* grows at cooler temperatures, leprosy primarily affects skin and nerves in the peripheral tissues. The clinical features are a result of interaction of the organism with the host's immune system. Infection of sensory neurons results in the loss of the sensation of pain; this leads to the mutilation that is characteristic of advanced leprosy. Infection of motor neurons causes paralysis. The manifestations of leprosy depend on the infected person's immune response to the causative agent, *M. leprae*.[10] Clinically, leprosy may resemble many dermatologic and neurologic conditions; thus, a high index of suspicion is necessary for accurate diagnosis.

Early leprosy and indeterminate leprosy, which are characterized by hypopigmented, ill-defined skin lesions, often heal on their own and may even be ignored by some patients. Nevertheless, all such patients should receive chemotherapy. If healing does not occur, the disease progresses along a clinical spectrum [*see Figure 1*]. Leprosy typically presents as anesthetic skin lesions associated with thickened peripheral nerves.[10] The appearance of the skin lesions varies according to the spectrum of disease.

Advanced infection can be divided into two polar forms: tuberculoid leprosy and lepromatous leprosy. The stages of the spectrum of disease, in order of decreasing cell-mediated im-

mune response to *M. leprae,* are tuberculoid leprosy, which is characterized by few skin lesions and low bacterial loads; borderline tuberculoid leprosy; borderline leprosy; borderline lepromatous leprosy; and lepromatous leprosy, which is characterized by diffuse skin lesions and high bacterial loads. The cell-mediated immunity of patients with tuberculoid leprosy is intact. Patients who have lepromatous leprosy are anergic and show severely impaired cell-mediated immunity to *M. leprae* and thus have a high burden of disease.

Tuberculoid leprosy Patients at the tuberculoid pole of the disease spectrum typically present with a limited number (i.e., less than six) of large, asymmetrical, well-defined skin lesions. The skin is rough, anhidrotic, and usually anesthetic; central healing of lesions may occur. Neurologic involvement is limited to a few peripheral nerves. Enlargement of a single nerve is common, and marked nerve damage can occur early in the course of tuberculoid disease, often resulting in wristdrop, clawing of the hand, and footdrop.[10] Tuberculoid leprosy often involves the greater auricular, radial cutaneous, ulnar, common peroneal, and posterior tibial nerves. Early treatment is key to minimizing nerve damage. Skin lesions of patients with tuberculoid leprosy contain predominantly helper T cells, well-formed granulomas, and few organisms.

Lepromatous leprosy Patients at the lepromatous pole of the spectrum of disease present with skin lesions that are widely and symmetrically disseminated, often demonstrating only slight hypopigmentation or erythema. The major defect in lepromatous leprosy appears to be a specific inability of T cells to respond to *M. leprae*; as a result, these patients are unable to generate the chemokines and lymphokines that normally activate macrophages and thereby enhance killing of the organisms. Because of the poor cell-mediated immune response to *M. leprae,* the bacterial burden becomes quite large in patients with lepromatous leprosy. Typically, skin lesions are numerous. At times, the entire skin surface can be involved, yielding a diffuse, waxy appearance. Loss of eyebrows and hypertrophy of earlobes produce the characteristic leonine facies. Neurologic involvement is widespread. A symmetrical peripheral neuropathy progresses proximally; sensory loss precedes paralysis. Other cool tissues are often involved, including the anterior chamber of the eye, the upper airway, and the testes. Sustained bacillemia with multisystem infection can occur.

On biopsy, skin lesions of patients with lepromatous leprosy are practically devoid of granulomas and helper T cells; instead, they contain suppressor T cells and numerous bacilli. A biopsy specimen showing many organisms and foam cells but no granulomas confirms the diagnosis, as does the finding of numerous acid-fast bacilli (AFB) in smears of skin-slit preparations. If left untreated, lepromatous leprosy is relentlessly progressive.

Borderline leprosy Many patients with leprosy have disease that falls between the tuberculoid and lepromatous forms. These patients are classified as having borderline tuberculoid or borderline lepromatous disease [*see Figure 1*].

Laboratory Studies

According to the World Health Organization (WHO) classification of leprosy, patients have paucibacillary disease when no bacilli are demonstrated on skin smears, and have multibacillary disease when bacilli are seen on skin smears.[11] The diagnosis of leprosy is usually suspected on clinical grounds. However, demonstration of acid-fast bacilli in slit-skin smears provides laboratory confirmation of the diagnosis in cases of multibacillary disease. Samples for skin smears are obtained by using a scalpel blade to scrape the openings of small slits made in pinched skin. The tissue fluid obtained is smeared on a slide and stained for AFB by the Fite method.[10,11]

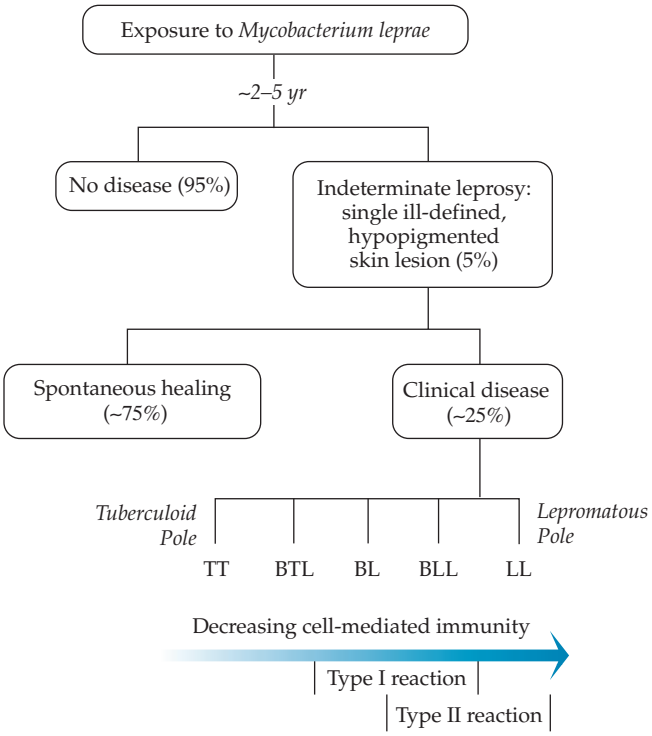


Figure 1 **The natural history and clinical spectrum of leprosy. (BL—borderline leprosy; BLL—borderline lepromatous leprosy; BTL—borderline tuberculoid leprosy; LL—lepromatous leprosy; TL—tuberculoid leprosy)**

Skin biopsy is another method of diagnosing leprosy, especially in patients with paucibacillary disease, and it can be used to classify the disease according to the clinical spectrum. Bacilli are frequently not seen in specimens from patients with disease in the tuberculoid end of the spectrum. In such cases, a biopsy that shows well-formed granulomas but few or no organisms can help establish the diagnosis of tuberculoid leprosy; the finding of inflamed nerves usually confirms the diagnosis.[10]

Serologic testing has little clinical utility in diagnosing leprosy.

TREATMENT

Patients with leprosy should be treated by physicians who are experienced in the management of this disease. Antimicrobial agents that are used include rifampin, dapsone, clofazimine, ofloxacin, and minocycline. The WHO has recommended treatment regimens for leprosy [*see Table 1*].

Rifampin is the most effective agent against *M. leprae* and can render lepromatous patients noninfectious within days. Patients with paucibacillary disease that is indeterminate, tuberculoid, or borderline tuberculoid receive monthly rifampin (under supervision) and daily dapsone (self-administered) therapy for 6 months. In the past, dapsone was used alone, but this is no longer recommended, because monotherapy led to the emergence of drug resistance. Patients with a single-lesion paucibacil-

Table 1 World Health Organization Recommendations for Treatment of Leprosy[12]

Disease Classification (Leprosy Type)	*Drugs*	*Dosages*	*Duration*
Paucibacillary (indeterminate, tuberculoid, or borderline tuberculoid)	Rifampin Dapsone	600 mg once monthly, supervised 100 mg daily, self-administered	6 mo
Single-lesion, paucibacillary	Rifampin Ofloxacin Minocycline	600 mg 400 mg 100 mg	Single dose
Multibacillary (borderline, borderline lepromatous, or lepromatous)	Rifampin Dapsone Clofazimine	600 mg once monthly, supervised 100 mg daily, self-administered 300 mg once monthly, supervised *or* 50 mg daily, self-administered	12 mo

lary disease can be treated with triple-drug therapy consisting of a single dose of rifampin, ofloxacin, and minocycline. Multibacillary disease requires a minimum of 12 months of therapy with a three-drug regimen that includes daily dapsone and monthly rifampin therapy plus clofazimine therapy either monthly or daily; monthly medication is given under supervision, whereas daily medication is self-administered.[12] Long-term follow-up is often suggested because late relapses may occur.

Leprosy Reactions

After initiation of therapy for *M. leprae,* patients may experience episodic immunologically mediated acute inflammatory responses called reactions, which are important mechanisms for causing nerve damage [*see Table 2*]**.** These immune reconstitution reactions can be characterized by swelling and edema in preexisting skin lesions or by peripheral neuropathy and neuritis, which can cause pain, tenderness, and loss of function. If not recognized and treated aggressively, these reactions can lead to irreversible nerve damage and permanent limb deformity.

There are two types of leprosy reactions. Type I reactions, or reversal reactions, are characterized by cellular hypersensitivity. Type II reactions, or erythema nodosum leprosum (ENL), are characterized by a systemic inflammatory response to immune complex distribution. Reversal reactions typically occur after initiation of leprosy treatment but can occur spontaneously before or after multidrug therapy. Treatment of type I reactions includes the use of anti-inflammatory agents; thalidomide is indicated for treatment of type II reactions.[13] A decline in ENL has been observed since the introduction of multidrug therapy, perhaps in part because of the anti-inflammatory effects of clofazimine.

CHALLENGES

The enormous progress in chemotherapy, which has helped to reduce the number of leprosy cases by nearly 90% over the past 10 years, makes the worldwide elimination of leprosy a realistic goal. However, there are challenges to eradicating this disease: many patients present late in the course; multidrug therapy is still not used in some endemic areas, so relapses may occur; and unfortunately, the disease still carries a stigma, which can interfere with effective disease management.

Nontuberculous Mycobacteria Infections

NTM infections are associated with several major clinical syndromes [*see Table 3*]. Selected NTM infections are associated with pulmonary disease, lymphadenitis, skin and soft tissue disease, skeletal infection, and infections related to catheters and other foreign bodies. In addition, disseminated disease (e.g., caused by

Table 2 Characteristics and Treatment of Leprosy Reactions[13]

Type	*Clinical Features*	*Risk Factors*	*Pathogenesis*	*Management*
Type I: reversal reactions	Occurs in BT and BB, but most common in BL; skin lesions are new, or there is increased inflammation in preexisting lesions; acute inflammation of nerve trunks may or may not occur	Recent pregnancy, facial plaques, extensive skin involvement, and preexisting neuritis	Increase in cell-mediated immunity to bacilli in dermis and Schwann cells, leading to inflammation of skin and nerve trunks	Prolonged anti-inflammatory therapy, analgesia, and physical support of active neuritis; high-dose steroids (1 mg/kg) for 4–6 mo; azathioprine and cyclosporine are also effective; surgical decompression of swollen nerves if medical therapy is unsuccessful
Type II: erythema nodosum leprosum	Occurs in BL and LL; skin lesions are new, small, tender, erythematous subcutaneous nodules; fever, arthralgia, neuritis, vasculitis, adenopathy, iridocyclitis, orchitis, and dactylitis are present	Pregnancy, age < 40 yr	Systemic inflammatory response to deposition of extravascular immune complexes formed from *Mycobacterium leprae* antigen	Thalidomide is the drug of choice; systemic symptoms and pain are alleviated in 24–48 hr, and nodules involute in 3 days; starting dose, 200 mg, may be decreased with control of symptoms; pentoxifylline and colchicine may be effective in mild cases; neuritis requires systemic steroids

BB—borderline leprosy BL—borderline lepromatous leprosy BT—borderline tuberculoid leprosy LL—lepromatous leprosy

Table 3 Major Clinical Syndromes Associated with Nontuberculous Mycobacterial Infections[1]

Clinical Disease	*Selected Etiologic Species*			
	Common	*Less Common*	*Unusual*	*More Unusual*
Chronic bronchopulmonary disease	*M. avium* complex *M. kansasii* *M. abscessus*	*M. malmoense* *M. xenopi*	*M. simiae* *M. szulgai* *M. fortuitum* *M. celatum* *M. gordonae*	*M. asiaticum* *M. shimodii* *M. smegmatis* *M. haemophilum*
Lymphadenitis	*M. avium* complex *M. scrofulaceum* *M. malmoense*	—	*M. fortuitum* *M. chelonae* *M. kansasii* *M. abscessus* *M. haemophilum*	*M. lentiflavum* *M. interjectum* *M. heidelbergense* *M. bohemicum*
Skin and soft tissue infection	*M. ulcerans* *M. marinum*	*M. abscessus* *M. chelonae* *M. fortuitum*	*M. kansasii* *M. haemophilum* *M. malmoense*	*M. smegmatis*
Otitis media	*M. abscessus*	*M. chelonae*	—	—
Tenosynovitis	*M. marinum* *M. avium* complex	—	*M. fortuitum* *M. abscessus* *M. chelonae* *M. kansasii* *M. haemophilum* *M. scrofulaceum*	*M. smegmatis* *M. nonchromogenicum* *M. malmoense* *M. xenopi* *M. szulgai*
Osteomyelitis	*M. fortuitum*	*M. abscessus*	*M. xenopi* *M. marinum*	*M. kansasii*
Catheter-related infections	*M. fortuitum* *M. abscessus* *M. chelonae*	—	*M. mucogenium* *M. neoaurum* *M. aurum*	*M. avium* complex *M. smegmatis*
Prosthetic valve infections	*M. fortuitum*	*M. chelonae*	*M. gordonae*	—
Surgical site infections	*M. fortuitum*	*M. chelonae*	*M. abscessus*	*M. simiae*

MAC) is common in persons with advanced HIV infection.

Traditionally, NTM infections have been categorized on the basis of characteristic colony morphology, growth rate, and pigmentation (i.e., the Runyon system of classification). The availability of rapid molecular diagnostics has reduced the usefulness and importance of this system. However, growth rates and colony pigmentation continue to provide a practical way for grouping mycobacteria specimens in the clinical microbiology laboratory. Rapidly growing mycobacteria include nonpigmented and pigmented species that produce mature growth on agar plates within 7 days. So-called slowly growing mycobacteria include species that require more than 7 days to reach mature growth on solid media; this group includes MAC and *M. kansasii*, which are pigmented and require 7 to 10 days to mature on solid media. *M. marinum* grows optimally at 28° to 30° C (which in large part explains its proclivity to cause skin and soft tissue infections), whereas *M. gordonae* (a common laboratory contaminant that rarely causes human disease) grows best at 35° to 37° C.

Although growth characteristics on solid media may be used to help categorize NTM infections, molecular diagnostic tests provide a definitive diagnosis for many NTM infections. Current guidelines recommend that for culture of NTM, as with culture for *M. tuberculosis*, both solid and liquid media be used to optimize recovery of mycobacterial species.[14,15] Mycobacteria grow more quickly in broth or liquid media than on solid media, which can decrease time to detection of a positive culture; in addition, liquid media may be more sensitive than solid media for some mycobacteria. Commercially available genetic probes for identification of selected NTM are available, including an acridinium ester–labeled nonradioactive DNA probe based on the detection of ribosomal RNA (rRNA). These probes are approved by the Food and Drug Administration for identification of *M. tuberculosis* and some common species of NTM, including *M. kansasii*, MAC, and *M. gordonae*.[14] No commercial probes are available for identification of other NTM species, but polymerase chain reaction (PCR) has proved useful for identification of rapidly growing mycobacteria. High-performance liquid chromatography, which is generally available at large reference laboratories (e.g., state public health laboratories, the Centers for Disease Control and Prevention) is useful in identifying *Mycobacterium* species.

MYCOBACTERIUM AVIUM COMPLEX

MAC comprises two closely related species, *M. avium* and *M. intracellulare*. MAC is encountered most often as an opportunistic pathogen in immunosuppressed patients, especially those with advanced HIV infection or AIDS. However, MAC infection can occur in immunocompetent hosts; it is especially likely to cause pulmonary infections in persons with underlying chronic

lung disease. The environmental reservoirs for MAC include both water and soil. The portal of entry is the respiratory tract or the gastrointestinal tract. When isolated from immunocompetent patients, MAC may be a simple contaminant or a true pathogen.[16] Because MAC is an organism of relatively low virulence, especially compared with *M. tuberculosis*, its presence does not always represent disease; colonization by MAC must be excluded through consideration of the clinical, radiographic, and pathologic features of the disease.

MAC Infection in the Immunocompetent Host

Chronic pulmonary disease is the clinical condition that is most often associated with MAC infection in the immunocompetent host; lymphadenitis and, rarely, disseminated disease are also seen. MAC pulmonary disease was initially described predominantly in men with underlying lung disease; often, such patients had a history of tobacco or alcohol consumption, and some had a prior history of tuberculosis.[17] The radiographic presentation is similar to that of tuberculosis, with upper lobe fibronodular or cavitary disease.[18] Patients with these radiographic abnormalities may present with pulmonary and systemic symptoms. CT findings may include multiple nodular infiltrates and bronchiectasis.[17]

Since the late 1980s, pulmonary MAC has been described in other immunocompetent patients, such as nonsmoking elderly women with fibronodular bronchiectasis and those with limited radiographic changes.[19,20] In addition, carriage of a cystic fibrosis or an abnormal α_1-antiproteinase gene appears to predispose to the development of MAC lung disease.[18] Older women without previous lung disease may present with fewer systemic symptoms and less pronounced radiographic findings. Pulmonary MAC infection in women, typically elderly women, with a radiographic pattern of nodules initially in the middle or lingual lobes has been termed the Lady Windermere syndrome and has been attributed to habitual voluntary suppression of cough.[21] Pulmonary MAC infection may develop in patients with cystic fibrosis. A hypersensitivity pneumonitis has been reported in persons exposed to MAC aerosolized from contaminated hot tubs or pools.[22] The radiographic picture is similar to that of other hypersensitivity pneumonitides.

Criteria for the diagnosis of NTM infection include repeated isolation of a potentially pathogenic species, the absence of other pathogens, and a compatible clinical, radiologic, or pathologic picture. The American Thoracic Society has developed criteria for the diagnosis of NTM infection, including pulmonary infection by MAC [*see Table 4*].[23]

Lymphadenitis from MAC typically affects previously healthy children who are 1 to 5 years of age.[24] Submandibular lymph nodes are most often involved. The patient usually does not have evidence of pulmonary or systemic illness. The lymph node may become tender and erythematous, and it may eventually drain. Diagnosis is made by pathologic examination of the excised node, which in most cases reveals caseating granulomas and positive AFB smears. Without culture data or nucleic acid testing, it is impossible to distinguish NTM lymphadenitis from tuberculous lymphadenitis. Excision of the infected node is the treatment of choice, because incision and drainage may lead to the formation of a draining fistulous tract.[25] Treatment with antimicrobials is rarely indicated except for serious disease or recurrences.

MAC Infection in AIDS

Disseminated MAC infection is an opportunistic disease that typically occurs late in the course of AIDS—in patients with $CD4^+$ T cell counts lower than $50/mm^3$—when other opportunistic infections and neoplasia have already occurred. The mode of acquisition of MAC in patients with AIDS is not clear. The genetic diversity of MAC isolates from patients with AIDS implies that the organisms may be acquired from multiple environmental sources.

MAC produces widely disseminated infection in patients with AIDS.[26] Systemic symptoms predominate, including fever, chills, night sweats, and profound weight loss. Fever and night sweats are the most common manifestations, occurring in more than 75% of patients.[27] Gastrointestinal symptoms, notably diarrhea, are present in at least a third of patients. Pulmonary symptoms are often present, but their significance is difficult to interpret because other opportunistic pathogens (e.g., cytomegalovirus and *Pneumocystis*) are often present, as well. Anemia and an elevated alkaline phosphatase level are frequently observed laboratory abnormalities.

Disseminated MAC infection can be readily documented by recovery of the organism from AFB and mycobacterial blood

Table 4 American Thoracic Society Criteria for Diagnosis of Nontuberculous Mycobacterial Lung Disease[21*]

Radiographic Criteria	*Laboratory Criteria*
Infiltrate, reticulonodular infiltrate, or cavitary disease on chest x-ray *or* Multifocal bronchiectasis or multiple small nodules on high-resolution chest CT	Two positive respiratory (sputum and/or bronchial wash) cultures within 12 mo, if one or both specimens are AFB smear positive *or* Three positive sputum/bronchial wash cultures within 12 mo, if none of the specimens are AFB smear positive *or* In patients unable to produce sputum, one positive bronchial wash culture with a 2+, 3+, or 4+ AFB smear and/or growth on solid media *or* A transbronchial biopsy yielding NTM *or* Biopsy showing mycobacterial histopathologic features (granulomatous infiltration and/or AFB) and one or more sputa/bronchial washings are positive for NTM

*For symptomatic patients who are HIV seropositive or HIV seronegative.
AFB—acid-fast bacilli CT—computed tomography NTM—nontuberculous mycobacteria

cultures. Blood is the preferred source of diagnostic specimens, but organisms may also be seen (and recovered by AFB culture) in biopsy specimens of the lung, liver, spleen, bone marrow, and lymph nodes. Despite the presence of many mycobacteria in macrophages, well-formed granulomas are typically absent; this so-called lepromatous histology reflects the profound impairment of cell-mediated immunity and explains the inability to contain the MAC infection. Intestinal or pulmonary infection may precede MAC bacteremia, but smears and cultures of these sites are not sensitive enough to be used as screening tests for early infection. Positive intestinal and pulmonary smear and culture results may represent colonization; however, positive cultures of stool and sputum are strongly predictive of subsequent dissemination in patients with low $CD4^+$ T cell counts.[28]

Treatment

Immunocompetent patients Treatment of pulmonary MAC is lengthy (longer than 1 year), but success rates have increased with the availability of newer macrolide drugs. Multidrug therapy is indicated to prevent the development of resistance. The cornerstone of therapy is the administration of a macrolide; clarithromycin is the preferred agent, with azithromycin an alternative. Ethambutol and rifabutin are used in conjunction with a macrolide. In severe cases, an aminoglycoside (e.g,. amikacin) may be needed. Treatment should be continued for at least 12 months after sputum cultures have converted to negative. Surgical resection may be indicated in patients with severe, refractory disease or in those who develop severe complications (e.g., life-threatening hemoptysis).

HIV-infected patients The treatment and survival of AIDS patients with MAC infection have improved greatly, especially since the advent of highly active antiretroviral therapy (HAART). More effective agents for MAC prophylaxis and treatment are now available, and the use of HAART can reduce the burden of disseminated MAC.[27] Patients with $CD4^+$ T cell counts of $50/mm^3$ or lower should receive MAC prophylaxis with azithromycin (1,200 mg weekly) or clarithromycin (500 mg twice daily).[29] The drugs are equally effective, but azithromycin is often preferred for MAC prophylaxis because of its greater convenience. Rifabutin is an alternative prophylactic agent for patients who cannot tolerate a macrolide.[30] An AFB blood culture should be obtained before the initiation of MAC prophylaxis to rule out disseminated disease.

Macrolides are the cornerstone of treatment of MAC disease. In HIV-infected patients with disseminated MAC, treatment should consist of clarithromycin or azithromycin plus ethambutol [*see Table 5*]. The addition of rifampin or rifabutin may be indicated in patients with advanced AIDS who are not receiving antiretroviral treatment and who have a high mycobacterial load.[31] Clarithromycin is the preferred macrolide for initial therapy. Clarithromycin may be slightly more active than azithromycin, but it is associated with more gastrointestinal side effects.[23] The debate continues as to which rifamycin is best. Rifabutin has lower minimum inhibitory concentrations than rifampin, but rifampin achieves higher serum levels.[32] Rifampin is also associated with many more drug interactions than rifabutin and cannot be given with protease inhibitors; this is an important consideration in patients receiving treatment for HIV infection. In some patients who do not respond to therapy, the addition of a third and even a fourth drug may be indicated. Amikacin or a fluoroquinolone such as levofloxacin is used for this purpose.[31,33]

HIV-infected patients who start antiretroviral therapy can discontinue primary prophylaxis for MAC infection when their $CD4^+$ T cell counts exceed $100/mm^3$ for more than 6 months.[34] Maintenance therapy (i.e., secondary prophylaxis) for MAC infection may be discontinued after 12 months if the patient remains asymptomatic and the $CD4^+$ T cell count has been above $100/mm^3$ for at least 3 months.[29,31]

Immune reconstitution syndromes can be an adverse event in the treatment of patients with HIV infection who have disseminated MAC. These syndromes sometimes occur in HIV-infected patients with unrecognized MAC who are started on antiretroviral agents. Pathogen-specific immune responses to MAC can occur in the first few months after the initiation of highly active antiretroviral therapy.[35] Immune reconstitution reactions to MAC usually manifest themselves as fever and lymphadenopathy (peripheral, intrathoracic, or intra-abdominal). Bacteremia does not occur. In severe cases, treatment with corticosteroids or nonsteroidal anti-inflammatory drugs may be necessary to ameliorate the symptoms associated with the immune response.

MYCOBACTERIUM KANSASII

M. kansasii is found in the United States, Europe, and Japan; but in the United States, most cases are clustered in the Midwest—hence the name *M. kansasii*. Cases are also clustered in the South. Water is the only known reservoir; this organism has not been found in soil. *M. kansasii* is the second most common organism responsible for NTM pulmonary disease. Risk factors include cigarette smoking, pneumoconiosis, and HIV infection. *M. kansasii* is also the second most common NTM infection causing disseminated disease in patients with AIDS.[36] Infection probably occurs through the respiratory route. Clinical symptoms of *M. kansasii* infection can be identical to those of pulmonary tuberculosis (i.e., fever, cough, weight loss, and hemoptysis). The radiographic presentation usually involves fibronodular or fibrocavitary upper lobe disease resembling pulmonary tuberculosis, except that the cavities may be thin walled and look like bullae.[37] In patients with HIV infection who have $CD4^+$ T cell counts above $200/mm^3$, *M. kansasii* pulmonary disease can be cavitary. Patients with lower $CD4^+$ T cell counts are more likely to have noncavitary pulmonary disease, sometimes in association with disseminated disease.

Treatment

Treatment for *M. kansasii* includes isoniazid, rifampin, and ethambutol given for a total of 18 months, with continuation of therapy for at least 12 months after the patient is culture negative [*see Table 5*].[23] *M. kansasii* is resistant to pyrazinamide; hence, this drug is not used. Treatment failures may result from rifampin resistance. Fluoroquinolones and aminoglycosides are alternative agents in selected cases. HIV-infected patients may benefit from the addition of clarithromycin as a fourth agent.[1]

RAPIDLY GROWING MYCOBACTERIA

Rapidly growing mycobacteria constitute a group of nonpigmented and pigmented NTM species that produce growth on agar plates within 7 days. Of the rapidly growing mycobacteria, three groups account for most disease in humans: the *M. fortuitum* group, the *M. chelonae-abscessus* group, and the *M. smegmatis* group [*see Table 6*].[3,38] The use of molecular tools, including 16S ribosomal gene (rDNA) sequencing, has led to a recategorization of rapidly growing mycobacteria and recognition of new taxa. This is particularly the case with the *M. chelonae-abscessus* group,

Table 5 Treatment Regimens for Selected Nontuberculous Mycobacterial Infections[1]

Species	*Clinical Syndrome*	*Treatment*	*Duration*
M. avium complex	Pulmonary disease (e.g., HIV seronegative patients)	Clarithromycin, 500 mg b.i.d. (preferred) *or* Azithromycin, 600 mg q.d.* *plus* Ethambutol, 15–25 mg/kg/day *plus* Rifabutin, 300 mg q.d. *or* Rifampin, 600 mg q.d. *with or without* Streptomycin, 1 g q. 3–5 days/wk for 6–12 wk Amikacin, 7.5–15 mg/kg I.V. q.d.† (may add initially for cavitary disease)	Until 12 mo after sputum cultures become negative (usually at least 18 mo of therapy)
	Disseminated disease (e.g., HIV-infected patients)	Clarithromycin, 500 mg b.i.d. (preferred) *or* Azithromycin, 600 mg q.d.* *plus* Ethambutol, 15 mg/kg/day *with or without* Rifabutin, 300 mg q.d.‡ *or* Rifampin, 600 mg q.d.§	For HIV-infected patients, duration depends on $CD4^+$ T cell count
M. kansasii	Infiltrate/invasive pulmonary disease Disseminated disease	Rifampin, 600 mg q.d. *plus* Ethambutol, 15 mg/kg/day *plus* Isoniazid, 300 mg q.d. For severe disease, consider streptomycin, 0.5–1 g I.M. three times/wk or clarithromycin, 500 mg b.i.d., p.o., for first 2–4 mo	18 mo total therapy, with 12 mo of therapy after patient becomes culture negative
	Disease in HIV-infected patients	As for pulmonary disease, but replace rifampin with rifabutin, 150 µg q.d. for patients on a protease inhibitor; clarithromycin, 500 mg b.i.d., can be substituted for rifampin/rifabutin in patients unable to take either drug	
M. abscessus	Pulmonary, disseminated, or extensive cutaneous disease	Amikacin, 15 mg/kg/day I.V. in one or two doses, with peaks at 20 µg/ml if given in two divided doses *plus* Clarithromycin, 500 mg b.i.d. *plus* Cefoxitin, 200 mg/kg, usually 3 g q. 6 hr I.V. *or* Imipenem, 500–750 mg t.i.d. I.V.	Unknown; surgical resection is the only curative therapy for pulmonary disease, but intermittent treatment with clarithromycin or azithromycin alone for weeks to months may be an alternative in selected patients
	Cutaneous localized disease	Clarithromycin, 500 mg b.i.d.	4–6 mo
M. fortuitum	Pulmonary or cutaneous disease	Clarithromycin, 500 mg b.i.d. *plus* Doxycycline, 100 mg b.i.d. *or* Trimethoprim-sulfamethoxazole, 800/160 mg b.i.d. *or* Levofloxacin, 500–750 mg q.d. *or* Gatifloxacin, 400 mg q.d.	For pulmonary disease, 6–12 mo; for cutaneous disease, 4–6 mo
M. marinum	Cutaneous disease	Ethambutol, 15 mg/kg/day *plus* Rifampin, 600 mg q.d. *or* Clarithromycin, 500 mg b.i.d. *or* Minocycline, 100 mg b.i.d. *or* Doxycycline, 100 mg b.i.d.	Minimum, 3 mo (see text)
M. ulcerans	Cutaneous disease	Rifampin, 600 mg q.d. *or* Amikacin, 15 mg/kg b.i.d. I.M. *or* Streptomycin, 15 mg/kg/day I.M. *with or without* Surgical resection, if possible	4–6 wk

*Maximum dosage used in clinical study; adjust on individual basis depending on drug toxicity; dosages of 250 mg/day, 500 mg/day, or 500 mg three times/wk have been discussed.
†A possible alternative is to give amikacin two or three times a week in doses of 15 to 20 mg/kg.
‡Rifabutin dose may need to be adjusted in HIV-infected patients on antiretroviral therapy.
§Rifampin cannot be given with certain antiretroviral agents, including protease.

Table 6 Principal Species of Rapidly Growing Mycobacteria That Are Human Pathogens

Group	*Species*
M. fortuitum	*M. fortuitum* (formerly *M. fortuitum* biovar *fortuitum*) *M. peregrinum* (formerly *M. fortuitum* biovar *peregrinum*) *M. mucogenicum* (formerly *M. chelonae*–like organism) *M. senegalense* *M. septicum* *M. magertiense* *M. porcinum* *M. houstonense* *M. bonickei* *M. neworleansense*
M. chelonae-abscessus	*M. chelonae* (formerly *M. chelonae* subsp. *chelonae*) *M. immunogenum*
M. smegmatis	*M. smegmatis* sensu stricto *M. goodie* *M. wolinskyi*

which consists of *M. chelonae* (formerly *M. chelonae chelonae*) and *M. abscessus* (formerly *M. chelonae abscessus*), along with *M. immunogenum*. Rapidly growing mycobacteria are associated with a variety of clinical syndromes, but they most commonly cause skin and soft tissue infections, both in the community and in hospitalized postoperative patients; other infections include skeletal infections (e.g., bone, joint, tendon), pulmonary disease (*M. abscessus*), and, on occasion, catheter-related infections [*see Table 3*].

Skin, Soft Tissue, and Skeletal Infections

The skin and soft tissues are the most common sites of infection from rapidly growing mycobacteria; indeed, of all NTM species, rapidly growing mycobacteria are the most common cause of skin and soft tissue infections. The *M. fortuitum* group and the *M. chelonae-abscessus* group cause localized infections, usually after local trauma or surgery.[1,38] Unlike infections with the *M. chelonae-abscessus* group, however, most *M. fortuitum* infections occur in patients who have no underlying chronic disease or immunosuppression. Furunculosis from *M. fortuitum* and *M. fortuitum*–group organisms have been reported in association with whirlpool footbaths in customers of nail salons.[39,40] Shaving the legs with a razor before a pedicure was a risk factor for infection. In these cases, molecular typing techniques were used to match rapidly growing mycobacteria recovered from the footbaths with isolates recovered from patients.

Like other NTM species, rapidly growing mycobacteria can cause infections of tendon sheaths, bursae, bones, and joints after direct inoculation through accidental trauma, surgery, puncture, or injection.[1] Surgical debridement is often necessary for the diagnosis and treatment of such infections, along with chemotherapy (see below).

Health Care–Related Infections

There have been a number of reports of nosocomial or health care–associated infections caused by rapidly growing mycobacteria.[41-43] These include infections from colonization of long-term venous access devices or peritoneal dialysis catheters, postinjection abscesses, and surgical wound infections (e.g., after cardiac bypass surgery, as well as augmentation mammoplasty, facelifts, and other plastic surgery; postoperative keratitis has been reported after ophthalmic surgery). Clusters of infections and outbreaks of true infection and pseudoinfections have occurred from contaminated fluids, irrigation with or exposure to tap water, injectable medicines, and topical skin solutions and markers. Most have involved *M. fortuitum* or *M. abscessus,* but *M. chelonae* has also been reported as a pathogen.[41]

Pulmonary Infections

M. abscessus is the third most common cause of pulmonary infection from NTM, after MAC and *M. kansasii*. *M. abscessus* causes 80% of the pulmonary disease from rapidly growing mycobacteria; *M. fortuitum* accounts for about 15%.[44] *M. chelonae* and other rapidly growing mycobacteria cause pulmonary disease less commonly. Adolescents with cystic fibrosis are at increased risk for NTM-related pulmonary disease; MAC is the most common pathogen in these cases, but other mycobacteria, especially *M. abscessus*, have been reported with increasing frequency.

Signs and symptoms of pulmonary disease caused by rapidly growing mycobacteria are variable and nonspecific. Most patients have a cough that becomes productive as the disease progresses. Fatigue and weight loss may also occur as the disease progresses. In many patients, symptoms have been present for months to years and have been attributed to chronic bronchitis or bronchiectasis.

The chest radiograph usually demonstrates patchy, reticular, nodular opacities—usually in the upper lobes, although any portion of the lung can be involved. Cavitation occurs in less than 20% of cases.[44] CT scan may show findings similar to those noted in patients with pulmonary MAC infection.

The diagnostic approach to lung disease that is suspected to be caused by rapidly growing mycobacteria is the same as that followed with infections from MAC and other NTM infections [*see Table 4*]. All patients with suspected mycobacterial infection should have three sputum specimens examined microscopically for AFB and cultured for mycobacteria.

Disseminated Cutaneous Infections

Disseminated cutaneous disease from rapidly growing mycobacteria is unusual. When it does occur, the pathogen is usually from the *M. abscessus-chelonae* group. More than 90% of patients with disseminated cutaneous disease from rapidly growing mycobacteria are immunocompromised. They generally have risk factors such as chronic renal failure, renal transplantation, or, especially, long-term use of corticosteroid therapy; few have HIV infection, however.[38]

Treatment

Selection of antimicrobial agents for infections from rapidly growing mycobacteria depends on the site of infection and the presumed pathogen. Empirical therapy is usually required initially, because final species identification and susceptibility testing are often performed at reference laboratories, and it may take weeks to obtain the results. Susceptibility testing should be performed on all clinically significant initial isolates, as well as on isolates recovered from patients who experience therapeutic failure or relapse.[21] Standardization guidelines for susceptibility testing of all *Mycobacterium* species, including rapidly growing mycobacteria and other NTM species, have been recommended by the Clinical and Laboratory Standards Institute (formerly the

National Committee for Clinical Laboratory Standards).[15] To meet these standards, rapidly growing mycobacteria should be tested against clarithromycin, amikacin, cefoxitin, imipenem, tobramycin, doxycycline, ciprofloxacin, and trimethoprim-sulfamethoxazole; other drugs that may be included are linezolid and the newer fluoroquinolones (e.g., levofloxacin, gatifloxacin, and moxifloxacin).

Treatment regimens for infections from rapidly growing mycobacteria have been developed [*see Table 5*]. Recommendations are based on in vitro and case series reports, because data from controlled trials are lacking. For pulmonary disease from *M. abscessus*, antimicrobial therapy alone is usually unsuccessful.[45] Surgical resection of the involved lung is recommended, preferably after an initial period of antimicrobial therapy.[1] Unfortunately, some patients with *M. abscessus* lung disease have bilateral disease and thus are not surgical candidates.

Treatment of *M. fortuitum* pulmonary disease with antimicrobial therapy alone has been much more successful. A 6- to 12-month regimen is recommended, starting with clarithromycin plus a second drug (e.g., doxycyline, trimethoprim-sulfamethoxazole, or a newer fluoroquinolone such as levofloxacin, gatifloxacin, or moxifloxacin) [*see Table 5*]. Therapy should be adjusted on the basis of susceptibility-testing results.

Generally, the length of treatment with any of the current antimicrobials for most skin and soft tissue infections from rapidly growing mycobacteria has been 4 months for mild disease and 6 months for serious disease.[38] For more serious disease, treatment may include the use of injectable agents; however, these are usually limited to the first 2 to 6 weeks of therapy, to minimize cost and drug toxicity. Surgical excision or debridement of the wound site combined with appropriate antimicrobial therapy is recommended because of better outcomes (e.g., healing without relapse).[38] Treatment of cutaneous infection from rapidly growing mycobacteria generally includes the use of a newer macrolide such as clarithromycin [*see Table 5*]. Treatment of skeletal infections such as osteomyelitis from rapidly growing mycobacteria has been accomplished by surgical wound debridement plus a minimum of 6 months of drug therapy with agents selected on the basis of the in vitro susceptibilities of the isolate. Treatment of disseminated cutaneous disease involves drainage of abscesses plus appropriate antimicrobial therapy (usually including clarithromycin) for at least 6 months. Because of the risk of the development of clarithromycin resistance with monotherapy for disseminated disease (estimated to be about 10% to 20% in this setting), therapy for the first 3 to 6 weeks should include other drugs, selected on the basis of in vitro susceptibilities whenever possible.[38]

MYCOBACTERIUM MARINUM

M. marinum is an NTM organism (group I of the Runyon classification) that causes skin infections in humans. *M. marinum* causes disease in many fish species from cold water, warm water, freshwater, or saltwater; and human infection follows contact with diseased fish or contaminated water.[46] *M. marinum* is distributed widely in aquatic environments, especially in relatively still or stagnant water, such as in fish tanks and in naturally occurring bodies of water; swimming pools are rarely reported as a source at present, likely because of chlorination. First described as so-called swimming-pool granulomas, *M. marinum* skin infections have also been termed fish-tank granulomas, because infection is most often acquired from aquarium maintenance.[46] Additional investigations are needed to further develop preventive strategies; some investigators have suggested that gloves be worn when cleaning fish tanks.

Infection is acquired by direct inoculation of the bacterium through broken skin in an aquatic environment; typically, this is seen after lacerations or abrasions are exposed to freshwater or saltwater or in injuries from fish spines. Human skin and soft tissue infections are manifested by cutaneous ulcers, nodules, or nodular lymphangitis; such infections can result in significant morbidity.[47] Although infection is most commonly limited to skin, it can spread to deeper structures, resulting in tenosynovitis, arthritis, and osteomyelitis. Disseminated infections are rare. Delays in diagnosis have been noted and can lead to adverse outcomes. In a number of cases, delayed or missed diagnoses led to the use of corticosteroids, which may have facilitated the spread of infection.[46] Clues in the clinical history, such as exposure to fish, natural bodies of water, or swimming pools, can expedite diagnosis and therapy in patients presenting with cutaneous infections.[47] The median incubation period after exposure to *M. marinum* is about 3 weeks, but the incubation period was longer than 30 days for more than a third of the patients in one series, and the incubation period can be up to months after exposure.[47]

A definitive diagnosis of *M. marinum* infection is made by biopsy and AFB culture results. The optimal therapy for *M. marinum* infection is unknown; there are no controlled trials on which to base treatment decisions, and data on *M. marinum* susceptibility are scarce.[46] Data from case series suggest that most skin infections can be adequately treated with antibiotics alone, with surgical intervention considered for those who do not respond to initial antimicrobial therapy. Adjunctive surgical debridement has been suggested for patients with deeper infections.[46] The length of treatment is also not standardized and has ranged from 3.5 months to 8 months in different series.[48] Treatment includes clarithromycin or minocycline (as monotherapy or in combination with other agents) or rifampin plus ethambutol [*see Table 5*]. Some authors recommend using at least two drugs.[48] A minimum of 3 months of therapy is recommended, with longer courses of therapy advocated for patients with more invasive infection (beyond skin and soft tissue) and those requiring surgical debridement. Some authors have suggested that treatment of *M. marinum* should include two drugs for 1 to 2 months after resolution of lesions, typically 3 to 4 months in total.[48] Although routine susceptibility testing has not been recommended for *M. marinum*,[1] it may be considered in special situations, such as for patients who do not respond clinically after several months of therapy and for those who continue to have positive cultures.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Wagner D, Young LS: Nontuberculous mycobacterial infections: a clinical review. Infection 32:257, 2004

2. Falkinham JO 3rd: Nontuberculous mycobacteria in the environment. Clin Chest Med 23:529, 2002

3. Brown-Elliott BA, Griffith DE, Wallace RJ Jr: Newly described or emerging human species of nontuberculous mycobacteria. Infect Dis Clin North Am 16:187, 2002

4. van der Werf TS, van der Graaf WT, Tappero JW, et al: *Mycobacterium ulcerans* infection. Lancet 354:1013, 1999

5. Primm TP, Lucero CA, Falkinham JO 3rd: Health impacts of environmental mycobacteria. Clin Microbiol Rev 17:98, 2004

6. Asiedu K, Scherpbier R, Raviglione M: Buruli ulcer: *Mycobacterium ulcerans* infection. Geneva: World Health Organization, 2000
http://www.who.int/gtbburuli/publications/PDF/Buruli_ulcer_monograph.PDF

7. Wansbrough-Jones M, Phillips R: Buruli ulcer. BMJ 330:1402, 2005

8. Hansen's Disease (Leprosy). Centers for Disease Control and Prevention, Atlanta, December 2003
http://www.cdc.gov/ncidod/dbmd/diseaseinfo/hansens_t.htm

9. Ooi WW, Moschella SL: Update on leprosy in immigrants in the United States: status in the year 2000. Clin Infect Dis 32:930, 2001

10. Boggild AK, Keystone JS, Kain KC: Leprosy: a primer for Canadian physicians. CMAJ 170:71, 2004

11. World Health Organization: Guidelines for skin smears. Int J Lepr Other Mycobact Dis 55:421, 1987

12. WHO Expert Committee on Leprosy: 7th report (WHO Technical Report Series, No. 874). World Health Organization, Geneva, 1998
http://www.who.int/lep/research/expert1.htm

13. Britton WJ, Lockwood DN: Leprosy. Lancet 363:1209, 2004

14. Brown-Elliott BA, Griffith DE, Wallace RJ Jr: Diagnosis of nontuberculous mycobacterial infections. Clin Lab Med 22:911, 2002

15. National Committee for Clinical Laboratory Standards: Susceptibility testing of mycobacteria, *Nocardia*, and other aerobic actinomycetes: approved standard M24-A. National Committee for Clinical Laboratory Standards, Wayne, Pennsylvania, 2003

16. Oliver A, Maiz L, Canton R, et al: Nontuberculous mycobacteria in patients with cystic fibrosis. Clin Infect Dis 32:1298, 2001

17. Field SK, Fisher D, Cowie RL: *Mycobacterium* complex pulmonary disease in patients without HIV infection. Chest 126:566, 2004

18. Chalermskulrat W, Gilbey JG, Donohue JF: Nontuberculous *Mycobacterium* in women, young and old. Clin Chest Med 23:675, 2002

19. Prince DS, Peterson DD, Steiner RM, et al: Infection with *Mycobacterium avium* complex in patients without predisposing conditions. N Engl J Med 321:863, 1989

20. Chitty SA, Ali J: *Mycobacterium avium* complex pulmonary disease in immunocompetent patients. South Med J 98:646, 2005

21. Reich JM, Johnson RE: *Mycobacterium avium* complex pulmonary disease presenting as an isolated lingular middle-lobe pattern. Chest 101:1605, 1992

22. Rickman OB, Ryu JH, Fidler ME, et al: Hypersensitivity pneumonitis associated with *Mycobacterium avium* complex and hot tub use. Mayo Clin Proc 77:1233, 2002

23. Diagnosis and treatment of disease caused by nontuberculous mycobacteria. American Thoracic Society. Am J Respir Crit Care Med 156:S1, 1997

24. Schaad UB, Votteler TP, McCracken GH Jr, et al: Management of atypical mycobacterial lymphadenitis in childhood: a review based on 380 cases. J Pediatr 95:356, 1979

25. Taha AM, Davidson PT, Bailey WC: Surgical treatment of atypical mycobacterial lymphadenitis in children. Pediatr Infect Dis 4:664, 1985

26. Horsburgh CR Jr: The pathophysiology of disseminated *Mycobacterium avium* complex disease in AIDS. J Infect Dis 179(suppl 3):S461, 1999

27. Horsburgh CR Jr, Gettings J, Alexander LN, et al: Disseminated *Mycobacterium avium* complex disease among patients infected with human immunodeficiency virus, 1985–2000. Clin Infect Dis 33:1938, 2001

28. Chin DP, Reingold AL, Horsburgh CR Jr, et al: Predicting *Mycobacterium avium* complex bacteremia in patients infected with HIV: a prospectively validated model. Clin Infect Dis 19:668, 1994

29. Kaplan JE, Masur H, Holmes KK, et al: Guidelines for preventing opportunistic infections among HIV-infected persons—2002. Recommendations of the U.S. Public Health Service and the Infectious Diseases Society of America. MMWR Recomm Rep 51(RR-8):1, 2002

30. Benson CA, Williams PL, Cohn DL, et al: Clarithromycin or rifabutin alone or in combination for primary prophylaxis of *Mycobacterium avium* complex disease in patients with AIDS: a randomized, double-blind, placebo-controlled trial. J Infect Dis 181:1289, 2000

31. Benson CA, Kaplan, Masur H, et al: Treating opportunistic infections among HIV-infected adults and adolescents. MMWR Recomm Rep 53(RR-15):1, 2004

32. Iseman MD: Medical management of pulmonary disease caused by *Mycobacterium avium*. Clin Chest Med 23:633, 2002

33. Koletar SL: Treatment of *Mycobacterium avium* in human immunodeficiency virus–infected individuals. Am J Med 102:16, 1997

34. El-Sadr WM, Burman WJ, Bjorling L, et al: Discontinuation of prophylaxis against *Mycobacterium avium* complex disease in HIV-infected patients who have a response to antiretroviral therapy. N Engl J Med 342:1085, 2000

35. Lawn SD, Bekker L, Miller RF: Immune reconstitution associated with mycobacterial infections in HIV-infected individuals receiving antiretrovirals. Lancet Infect Dis 5:361, 2005

36. Horsburgh CR: The epidemiology of disseminated nontuberculous mycobacterial infection in the acquired immunodeficiency syndrome. Am Rev Respir Dis 139:4, 1989

37. Ahn CH, McLarty JW, Ahn SS, et al: Diagnostic criteria for pulmonary disease caused by *Mycobacterium kansasii* and *Mycobacterium intracellulare*. Am Rev Respir Dis 125:388, 1982

38. Brown-Elliot BA, Wallace RJ Jr: Clinical and taxonomic status of pathogenic nonpigmented or late-pigmenting rapidly growing mycobacteria. Clin Microbiol Rev 15:716, 2002

39. Gira AK, Reisenauer AH, Hammock L, et al: Furunculosis due to *Mycobacterium mageritense* associated with footbaths at a nail salon. J Clin Microbiol 42:1813, 2004

40. Winthrop KL, Abrams M, Yakrus M, et al: An outbreak of mycobacterial furunculosis associated with footbaths at a nail salon. N Engl J Med 346:1366, 2002

41. *Mycobacterium chelonae* infections associated with face lifts—New Jersey, 2002–2003. MMWR Morb Mortal Wkly Rep 53:192, 2004 (Erratum in MMWR Morb Mortal Wkly Rep 53:246, 2004)

42. Phillips MS, von Reyn CF: Nosocomial infections due to nontuberculous mycobacteria. Clin Infect Dis 33:1363, 2001

43. Raad II, Vartivarian S, Khan A, et al: Catheter-related infections caused by the *Mycobacterium fortuitum* complex: 15 cases and review. Rev Infect Dis 13:1120, 1991

44. Daley CL, Griffith DE: Pulmonary disease caused by rapidly growing mycobacteria. Clin Chest Med 23:623, 2002

45. Griffith DE, Girard WM, Wallace RJ Jr: Clinical features of pulmonary disease caused by rapidly growing mycobacteria: an analysis of 154 patients. Am Rev Respir Dis 147:1271, 1993

46. Aubry A, Chosidow O, Caumes E, et al: Sixty-three cases of *Mycobacterium marinum* infection: clinical features, treatment, and antibiotic susceptibility of causative isolates. Arch Intern Med 162:1746, 2002

47. Jernigan JA, Farr BM: Incubation period and sources of exposure for cutaneous *Mycobacterium marinum* infection: case report and review of the literature. Clin Infect Dis 31:439, 2000

48. Lewis FM, Marsh BJ, von Reyn CF: Fish tank exposure and cutaneous infections due to *Mycobacterium marinum*: tuberculin skin testing, treatment, and prevention. Clin Infect Dis 37:390, 2003

I MANAGEMENT OF POISONING AND DRUG OVERDOSE

Kent R. Olson, m.d.
Manish M. Patel, m.d.

Drug overdose and poisoning are leading causes of emergency department visits and hospital admissions in the United States, accounting for more than 500,000 emergency department visits[1] and 11,000 deaths[2] each year. Exposure to poison can occur in several ways. The patient may have ingested it accidentally or for the purpose of committing suicide, may be a victim of accidental intoxication from acute or long-term exposure in the workplace, may be suffering from unexpected complications or overdose after intentional drug abuse, or may be a victim of an assault or terrorist attack. Poisons can include drugs, chemicals, biotoxins in plants or foods, and toxic gases. In all cases of poisoning, the clinician has several priorities: (1) immediately stabilize the patient and manage life-threatening complications; (2) perform a careful diagnostic evaluation, which includes obtaining a directed history, performing a physical examination, and ordering appropriate laboratory tests; (3) prevent further absorption of the drug or poison by decontaminating the skin or gastrointestinal tract; and (4) consider administering antidotes and performing other measures that enhance the elimination of the drug from the body. For expert assistance with identification of poisons, diagnosis and treatment, and referral to a medical toxicologist, the clinician should consider consulting with a regional poison-control center.

Initial Stabilization

In many cases of poisoning, the patient is awake and has stable vital signs, which allows the clinician to proceed in a stepwise fashion to obtain a history and to perform a physical examination. In other cases, however, the patient is unconscious, is experiencing convulsions, or has unstable blood pressure or cardiac rhythm, thus requiring immediate stabilization [*see Table 1*].

The first priority is airway patency. The airway's reflex protective mechanisms may be impaired because of drug-induced central nervous system depression (e.g., from opioids or sedative-hypnotic agents), excessive bronchial and oral secretions (e.g., from organophosphate insecticides), or swelling or burns (e.g., from corrosive agents or irritant gases). The airway should be cleared by the use of suction and by repositioning the patient; if the patient has an impaired gag reflex or other evidence of airway compromise, a cuffed endotracheal tube should be inserted. The adequacy of ventilation and oxygenation should be determined by clinical assessment, pulse oximetry, measurement of arterial blood gases, or a combination of these techniques. Supplemental oxygen should be administered, and if necessary, ventilation should be assisted with a bag/valve/mask device or a ventilator.[3] Even if the patient is not unconscious or hemodynamically compromised on arrival in the emergency department, continued absorption of the ingested drug or poison may lead to more serious intoxication during the next several hours. Therefore, it is prudent to keep the patient under close observation, with continuous or frequent monitoring of alertness, vital signs, the electrocardiogram, and pulse oximetry.

Management of Common Complications

COMA

Poisoning or drug overdose depresses the sensorium, the symptoms of which may range from stupor or obtundation to unresponsive coma. Deeply unconscious patients may appear to be dead because they may have nonreactive pupils, absent reflexes, and flat electroencephalographic tracings; however, such patients may have a complete recovery without neurologic sequelae as long as they receive adequate supportive care, including airway protection, oxygenation, and assisted ventilation.[4]

All patients with a depressed sensorium should be evaluated for hypoglycemia because many drugs and poisons can directly reduce or contribute to the reduction of blood glucose levels. A finger-stick blood glucose test and bedside assessment should be performed immediately; if such testing and assessment are impractical, an intravenous bolus of 25 g of 50% dextrose in water should be administered empirically before the laboratory report arrives.[5] For alcoholic or malnourished patients, who may have vitamin deficiencies, 50 to 100 mg of vitamin B_1 (thiamine) should be administered I.V. or I.M. to prevent the development of Wernicke syndrome.[5] If signs of recent opioid use (e.g., suspicious-looking pill bottles or I.V. drug paraphernalia) are in evidence or if the patient has clinical manifestations of excessive opioid effect (e.g., miosis or respiratory depression), the administration of naloxone may have both therapeutic and diagnostic value. Naloxone is a specific opioid antagonist with no intrinsic opioid-agonist effects.[6,7] Initially, a dose of 0.2 to 0.4 mg I.V. should be administered; if there is no response, repeated doses of up to 4 to 5 mg should be given. Doses as high as 15 to 20 mg may be administered if overdose with a resistant opioid (e.g., propoxyphene, codeine, or some fentanyl derivatives) is suspected.[6,7] Patients with opioid intoxication usually become fully awake within 2 to 3 minutes after administration of naloxone.

Table 1 The ABCDs of Initial Stabilization of the Poisoned Patient

Airway
Position the patient to open the airway; suction any secretions or vomitus; evaluate airway-protective reflexes; consider endotracheal intubation

Breathing
Determine adequacy of ventilation; assist ventilation, if necessary; administer supplemental oxygen

Circulation
Evaluate perfusion, blood pressure, and cardiac rhythm; determine QRS complex; attach continuous cardiac monitor

Dextrose
Quickly determine blood glucose by finger-stick test; give dextrose if patient is suspected of having hypoglycemia

Decontamination
Perform surface and gastric decontamination to limit absorption of poisons

Table 2 Mechanisms of Drug-Induced Hypotension

Mechanism	*Selected Causes*
Hypovolemia	
Vomiting and diarrhea	Iron; arsenic; food poisoning; organophosphates and carbamates; mushroom poisoning; thallium
Sweating	Organophosphates and carbamates
Venodilatations	Barbiturates; other sedative-hypnotic agents
Depressed cardiac contractility	Tricyclic antidepressants; beta blockers; calcium antagonists; class IA and class IC antiarrhythmic agents; sedative-hypnotic agents
Reduced peripheral vascular resistance	Theophylline; beta$_2$-adrenergic stimulants; phenothiazines; tricyclic antidepressants; hydralazine

Failure to respond to naloxone suggests that (1) the diagnosis is incorrect [*see* Differential Diagnosis, *below*]; (2) other, nonopioid drugs may have been ingested; (3) a hypoxic insult may have occurred before the victim was found and resuscitated; or (4) an inadequate dose of naloxone was given.

Flumazenil, a short-acting, specific benzodiazepine antagonist with no intrinsic agonist effects, can rapidly reverse coma caused by diazepam and other benzodiazepines.[5,7] However, it has not found a place in the routine management of unconscious patients with drug overdose, because it has the potential to cause seizures in patients who have been consuming large quantities of benzodiazepines on a long-term basis or who have ingested an acute overdose of benzodiazepines and a tricyclic antidepressant or other potentially convulsant drug [*see* Sedative-Hypnotic Agents, *below*].[5,7,8]

HYPOTENSION AND CARDIAC DYSRHYTHMIAS

The hypotension that commonly complicates drug intoxication has many possible causes [*see Table 2*].[9,10] Hypotension may result from volume depletion caused by severe drug-induced vomiting or diarrhea. In addition, relative hypovolemia may be caused by the venodilating effects of many drugs. Certain drugs or poisons can have direct negative inotropic or chronotropic effects on the heart, reducing cardiac output. Others can cause a severe reduction in peripheral vascular resistance. Some drugs or poisons can cause shock by a combination of these mechanisms.

Treatment of drug-induced shock includes rapid assessment of the likely cause, which is suggested by the history of exposure and the clinical findings. Hypotension with tachycardia suggests that the cause is volume depletion or reduced peripheral vascular resistance, whereas hypotension with bradycardia suggests that the cause is a disturbance of cardiac rhythm or that shock is a result of the generalized cardiodepressant effects of the drug. Regardless of the etiology, most patients benefit from an I.V. bolus of fluid (e.g., 0.5 to 1 L of normal saline) and empirical pressor therapy with dopamine or norepinephrine.[11] However, if hypoperfusion persists, it may be necessary to insert a pulmonary arterial catheter to obtain more specific information about volume and hemodynamic status.

A variety of cardiac dysrhythmias may occur as a result of drug intoxication or poisoning [*see Table 3*]. In addition to the direct pharmacologic actions of the drug or poison, impaired ventilation and oxygenation may trigger disturbances of cardiac rhythm.[11]

Treatment of a cardiac dysrhythmia depends on its etiology. Because conventional advanced cardiac life support (ACLS) protocols were not designed with poisoning in mind, use of these guidelines may have inappropriate or dangerous effects.[12] For example, a patient with tricyclic antidepressant intoxication (see below) may have wide-complex tachycardia resulting from severe depression of sodium-dependent channels in the myocardial cell membrane. However, use of the ACLS protocols for wide-complex tachycardia or possible ventricular tachycardia may lead the treating physician to administer procainamide, a class IA antiarrhythmic agent with cardiodepressant effects that are additive to those of the tricyclic antidepressants.[11] A patient with multiple premature ventricular contractions or who experiences episodes of ventricular tachycardia after intoxication with chloral hydrate or inhalation of a chlorinated solvent would respond more readily to a beta blocker than to lidocaine, the drug recommended by the ACLS protocols.[13] Finally, cardiac dysrhythmias from digitalis intoxication are most appropriately treated with digoxin-specific antibodies (see below).

HYPERTENSION

Although hypertension is not commonly recognized as a serious pharmacologic effect of drug intoxication, it may have life-threatening consequences and requires aggressive treatment. Hypertension may result from generalized CNS and sympathetic stimulation (e.g., by amphetamines or cocaine) or from the peripheral actions of drugs such as phenylpropanolamine, a potent alpha-adrenergic agonist.[14] (Although the Food and Drug Administration removed phenylpropanolamine from the market in the United States in November 2000, phenylpropanolamine is still available in other countries.) In addition, hypertension may result from the pharmacologic interaction of two agents, such as in the use of a stimulant or the ingestion of an inappropriate food by a person taking monoamine oxidase (MAO) inhibitors.[15] Severe hypertension can lead to intracranial hemorrhage, aortic dissection, or other catastrophic complications.[16,17]

Hypertension may be accompanied by tachycardia, as commonly occurs in cases of intoxication with generalized stimu-

Table 3 Causes of Cardiac Disturbances

Type of Disturbance	*Selected Causes*
Sinus tachycardia	Anticholinergic agents (e.g., diphenhydramine, atropine, tricyclic antidepressants); theophylline and caffeine; cocaine and amphetamines; volume depletion
Bradycardia or atrioventricular block	Beta blockers; calcium antagonists; tricyclic antidepressants; class IA and class IC antiarrhythmic agents; organophosphate and carbamate insecticides; digitalis glycosides; phenylpropanolamine (hypertension with reflex bradycardia)
Widening of the QRS complex	Tricyclic antidepressants; class IA and class IC antiarrhythmic agents; diphenhydramine; thioridazine; propoxyphene; hyperkalemia; hypothermia
Ventricular tachycardia or ventricular fibrillation	Tricyclic antidepressants; cocaine and amphetamines; theophylline; digitalis glycosides; fluoride or hydrofluoric acid burns (hypocalcemia); trichloroethane and numerous other chlorinated, fluorinated, and aromatic solvents; chloral hydrate; agents that cause prolongation of the QT interval (e.g., quinidine, sotalol)

lants such as cocaine and amphetamine derivatives. Hypertension may also be accompanied by bradycardia or even atrioventricular (AV) block, which may occur after phenylpropanolamine overdose because of the reflex baroreceptor response.

Treatment is directed at the cause of the hypertension. In patients who have taken cocaine, amphetamines, or other generalized stimulants, mild or moderate increases in blood pressure may be reduced simply by providing a quiet environment and administering a sedative agent such as diazepam. In persons who have taken an overdose of phenylpropanolamine or other alpha-adrenergic stimulant, administration of a specific alpha-adrenergic antagonist, such as phentolamine (2 to 5 mg I.V.), is extremely effective and usually leads to normalization of the slow heart rate or reversal of the AV block.[14] In general, beta blockers should not be used as single agents in the treatment of drug-induced hypertension, because their use may lead to unopposed alpha-adrenergic activity with paradoxically worsened hypertension.[18]

SEIZURES

Seizures may result from a number of factors, including a variety of drugs and poisons. The drugs that most commonly induce seizures are tricyclic antidepressants, bupropion, cocaine and related stimulants, antihistamines, and isoniazid [*see Table 4*].[19,20] Prolonged or repeated convulsions can lead to serious complications, including hyperthermia, rhabdomyolysis, brain damage, and death. In addition, seizure activity causes metabolic acidosis, which may worsen cardiotoxicity in patients who have taken an overdose of a tricyclic antidepressant.[11,19] Seizures can also result from hypoxia, hypoglycemia, head trauma, stroke, or serious CNS infections [*see* Differential Diagnosis, *below*].

Treatment of seizures includes taking immediate steps to protect the airway and provide oxygen while administering anticonvulsant drugs. The blood glucose level should be determined and dextrose administered if needed [*see* Coma, *above*]. Initial anticonvulsant therapy consists of diazepam (5 to 10 mg I.V.), lorazepam (1 to 2 mg I.V.), or midazolam (3 to 5 mg I.V. or, if I.V. access is not immediately available, 5 to 10 mg I.M.). Repeated doses are given if the initial therapy is ineffective. Because it is often ineffective, phenytoin is not a first-line anticonvulsant agent for drug- or toxin-induced seizures.[19] If convulsions persist, phenobarbital should be administered at a dosage of 15 to 20 mg/kg (1 to 1.5 g) I.V. over 20 to 30 minutes.[21] If seizure activity continues, the physician should consult with a neurologist and consider administering pentobarbital, another short-acting barbiturate, or propofol.[21] In addition, inducing neuromuscular paralysis (e.g., with pancuronium) should be considered to control the muscle hyperactivity, which may be necessary for controlling hyperthermia, rhabdomyolysis, or metabolic acidosis. If neuromuscular paralysis is induced, however, the physician should be aware that seizure activity in the brain may persist but may not be apparent.[21] If isoniazid poisoning is suspected, 5 g of pyridoxine (vitamin B_6) should be administered intravenously; or if more than 5 g of isoniazid was ingested, pyridoxine should be administered in an amount (in grams) equal to that of the isoniazid overdose.

HYPERTHERMIA

Hyperthermia is an underrecognized complication of poisoning and drug overdose that is associated with high morbidity and mortality.[22] It may result from the pharmacologic effects of the agent or as a consequence of prolonged muscle hyperactivity or seizures [*see Table 5*]. Severe hyperthermia (rectal temperature > 104° F [40° C]) that goes untreated may lead to brain damage, coagulopathy, rhabdomyolysis, hypotension, and, ultimately, death.[22]

Because it is immediately life threatening, hyperthermia warrants immediate and aggressive treatment.[22] Therapy is directed at the underlying cause, which is usually excessive muscle activity or rigidity. For mild or moderate cases, the physician should use appropriate pharmacologic agents (e.g., sedatives for cases of stimulant-induced psychosis and hyperactivity and anticonvulsants for cases of seizure), remove the patient's clothing, and maximize evaporative cooling by spraying the exposed skin with tepid water and fanning the patient. For severe cases, the most rapidly effective treatment is neuromuscular paralysis accompanied by maximal evaporative cooling.[22] In some cases, a specific antidote or therapeutic agent may be available [*see Table 5*].

Table 4 Drug-Induced Seizures

Common Causes	*Comments*
Tricyclic antidepressants	Seizure activity and resulting metabolic acidosis often aggravate cardiotoxicity; protracted seizures with absent sweating may lead to hyperthermia; phenytoin worsens cardiotoxicity in animal models; treat with benzodiazepines or phenobarbital
Cocaine and amphetamines	Seizures are usually brief and self-limited and are often preceded by tremors, agitation, hallucinations, or tachycardia; bupropion most commonly implicated in seizures, sometimes even with therapeutic use
Theophylline	Seizures are often prolonged, recurrent, and refractory to anticonvulsant therapy; phenytoin is ineffective in animal models; administer high-dose phenobarbital (at least 15–20 mg/kg I.V.); for patients with serum theophylline levels > 100 mg/L or status epilepticus, consider hemoperfusion or hemodialysis
Diphenhydramine	Seizures are usually brief and self-limited; in patients with massive intoxication (e.g., > 4–5 g), tricyclic-like cardiotoxicity may also occur
Isoniazid	Seizures are often accompanied by severe lactic acidosis; the specific antidote for seizures and coma is vitamin B_6 (pyridoxine), 5–10 g I.V., or, if the amount of ingested isoniazid is known, the equivalent gram-for-gram amount of vitamin B_6

HYPOTHERMIA

Hypothermia may accompany drug overdose and is usually caused by environmental exposure combined with inadequacy of the patient's response mechanisms. These inadequate mechanisms may include impaired judgment (in patients who have taken opioids, sedative-hypnotic agents, or phenothiazines or who have underlying mental disorders), a reduced shivering response (in those who have taken phenothiazines or sedative-hypnotic agents), and peripheral vasodilatation (in those who have taken phenothiazines or vasodilators).[23] Severe hypothermia (core temperature < 82° F [28° C]) may cause the patient to appear to be dead and may be associated with barely perceptible blood pressure, heart rate, or neurologic reflexes. Hypotension, bradycardia, and ventricular arrhythmias may fail to respond to pharmacologic treatment until the patient is warmed.[23,24] Because no controlled trials comparing rewarming methods exist, man-

Table 5 Drug-Induced Hyperthermia

Mechanisms	*Selected Causes and Comments*
Increased metabolic activity	Causes include salicylates, dinitrophenol, and cocaine and amphetamines
Reduced sweating	Causes include anticholinergic agents (e.g., tricyclic antidepressants, antihistamines, many plants, and some mushrooms)
Increased muscle activity or exertion	Causes include cocaine and amphetamines, phencyclidine, and exertional heatstroke
Neuroleptic malignant syndrome	Causes include haloperidol, related antipsychotic agents, and lithium; patients have lead-pipe rigidity, acidosis, and an elevated creatine kinase level that are caused by CNS dopamine blockade; specific treatment is bromocriptine (2.5–10.0 mg by nasogastric tube two to six times daily)[157]; treat severe hyperthermia with neuromuscular paralysis
Malignant hyperthermia	An inherited disorder of muscle cell function, commonly triggered by certain anesthetic agents (e.g., succinylcholine or halothane); causes severe muscle rigidity and acidosis not responsive to neuromuscular paralysis; treatment is dantrolene (2–5 mg/kg I.V.)[158]
Serotonin syndrome	Associated with the use of serotonin-enhancing agents (e.g., meperidine, dextromethorphan, fluoxetine, paroxetine, sertraline, L-tryptophan, or trazodone), especially in patients taking monoamine oxidase inhibitors; causes muscle rigidity, acidosis, and hyperthermia; treatment is neuromuscular paralysis; for mild cases, consider cyproheptadine (4 mg p.o. every hour for three or four doses)[159] or methysergide (2 mg p.o. every 6 hr for three or four doses)[160]

agement protocols vary institutionally and are often controversial.[23] Treatment of hypothermia is generally administered gradually because more aggressive management may precipitate cardiac dysrhythmias. Passive external rewarming is an acceptable treatment if the patient's condition is stable. Administration of a warmed mist inhalation or warmed I.V. fluids may be helpful, as may gastric or peritoneal lavage with warmed fluids, although the heat transfer involved in these measures is variable. For profound hypothermia accompanied by evidence of severe hypoperfusion (e.g., cardiac arrest or ventricular fibrillation), more aggressive measures, such as partial cardiopulmonary or femorofemoral bypass, may be required.[23,24] Of note is that patients with severe hypothermia can withstand cardiorespiratory arrest longer than normothermic patients—hence the old adage, "No one is dead until warm and dead."

RHABDOMYOLYSIS

Rhabdomyolysis, a common complication of severe poisoning or drug overdose, may result from direct myotoxic effects of the agent, from prolonged or recurrent muscle hyperactivity or rigidity, or from prolonged immobility with mechanical compression of muscle groups.[25] Severe rhabdomyolysis (usually associated with markedly elevated serum creatine kinase levels) may cause massive myoglobinuria that results in acute tubular necrosis and renal failure. Myoglobinuria is usually recognized by the pink or reddish hue of spun serum or by a positive dipstick test for hemoglobin in the urine, with few or no red blood cells seen on microscopic examination. Severe rhabdomyolysis may also cause hyperkalemia, which results from loss of potassium from dead or injured cells.

Treatment of rhabdomyolysis includes measures to prevent further muscle breakdown (e.g., control of muscle hyperactivity and treatment of hyperthermia) and to prevent deposition of toxic myoglobin in the renal tubules. Unequivocally, the mainstay of treatment in rhabdomyolysis is aggressive volume expansion with normal saline early in the disease to maintain urine output of 200 ml/hr in those who can tolerate the fluid load.[25] Nonrandomized trials have also shown alkalinization of urine to be beneficial, but the role of mannitol and furosemide in rhabdomyolysis is less clear.[25]

Clinical Evaluation

Although the history recounted by patients who have intentionally taken a drug overdose may be unreliable, it should not be overlooked as a valuable source of information. If the patient is unwilling or unable to specify which drugs were taken and when they were ingested or to provide a pertinent medical history, family and friends may be able to do so. Family members should be asked about other medications available in the household and about exposure in the workplace and through hobbies. In addition, paramedics should be asked for any pill bottles or drug paraphernalia that they may have obtained at the scene.

A directed toxicologic physical examination may yield important clues about the drugs or poisons that have been taken. Pertinent variables include the patient's vital signs, pupil size, lung sounds, peristaltic activity, skin moisture and color, and muscle activity; the presence or absence of unusual odors; and the presence or absence of track marks associated with I.V. drug

Table 6 Autonomic Syndromes Induced by Drugs or Poisons

Autonomic Syndrome	*Selected Causes*	*Empirical Interventions*
Sympathomimetic (agitation; dilated pupils; elevated BP and HR; sweaty skin; hyperthermia)	Cocaine; amphetamines; pseudoephedrine	Induce sedation; initiate aggressive cooling; treat hypertension with phentolamine; treat tachycardia with beta blockers
Sympatholytic (lethargy or coma; small pupils; normal or low BP and HR; low temperature)	Barbiturates; opioids; clonidine; benzodiazepines	Give naloxone for suspected opioid overdose; consider flumazenil for benzodiazepine overdose
Cholinergic (pinpoint pupils; variable HR; sweaty skin; abdominal cramps and diarrhea)	Organophosphate and carbamate insecticides; chemical warfare nerve agents	Give atropine and pralidoxime; obtain measurements of serum and RBC cholinesterase activity
Anticholinergic (agitation; delirium; dilated pupils; tachycardia; decreased peristalsis; dry, flushed skin)	Atropine and related drugs; antihistamines; phenothiazines; tricyclic antidepressants	Obtain immediate ECG tracing to evaluate for poisoning with tricyclic antidepressants; consider physostigmine only if tricyclics are not involved

Table 7 Use of the Clinical Laboratory in the Initial Diagnosis of Poisoning

Test	*Finding*	*Selected Causes*
Arterial blood gases	Hypoventilation (elevated Pco_2)	CNS depressants (e.g., opioids, sedative-hypnotic agents, phenothiazines, and ethanol)
	Hyperventilation	Salicylates; carbon monoxide; other asphyxiants
Electrolytes	Anion-gap metabolic acidosis	Salicylates; methanol; ethylene glycol; carbon monoxide; cyanide; iron; isoniazid; theophylline
	Hyperkalemia	Digitalis glycosides; fluoride; potassium
	Hypokalemia	Theophylline; caffeine; beta-adrenergic agents (e.g., albuterol); soluble barium salts
Glucose	Hypoglycemia	Oral hypoglycemic agents; insulin; ethanol
Osmolality and osmolar gap	Elevated osmolar gap*	Ethanol; methanol; ethylene glycol; isopropyl alcohol; acetone
ECG	Wide QRS complex	Tricyclic antidepressants; quinidine and other class IA and class IC antiarrhythmic agents
	Prolongation of the QT interval	Quinidine and related antiarrhythmic agents
	Atrioventricular block	Calcium antagonists; digitalis glycosides
Plain abdominal x-ray	Radiopaque pills or objects	Iron; lead; potassium; calcium; chloral hydrate; some foreign bodies
Serum acetaminophen	Elevated level (> 140 mg/L 4 hr after ingestion)	Acetaminophen (may be the only clue to a recent ingestion)

*Osmolar gap = measured osmolality – calculated osmolality. Measured osmolality is performed in the laboratory using a freezing-point-depression device (do not use the vaporization method). Calculated osmolality = 2(Na) + [BUN/2.8] + [glucose/18]. The normal osmolar gap is 0 ± 5 mOsm/L.
BUN—blood urea nitrogen Pco_2—carbon dioxide tension

abuse. Signs of one of the so-called autonomic syndromes [*see Table 6*] may suggest diagnostic possibilities and potential empirical interventions.[10]

The clinical laboratory may provide useful information that obviates an expensive and time-consuming toxicology screen. Recommended laboratory tests in the patient with an overdose of unknown cause include a complete blood count; measurements of glucose, electrolytes, blood urea nitrogen, creatinine, aspartate aminotransferase (AST), and serum osmolality (both measured and calculated); ECG; and plain abdominal x-ray (KUB [kidneys, ureters, and bladder] view) [*see Table 7*]. A quantitative serum acetaminophen level should be obtained immediately because acetaminophen overdose may be difficult to diagnose in the absence of a complete and reliable history, does not produce suggestive clinical or laboratory findings, and requires prompt administration of an antidote in patients with a serious acute ingestion if hepatic injury is to be prevented.[10,26,27]

Obtaining a thorough history, performing a careful physical examination, and using the clinical laboratory in a logical manner can often enable the physician to make a tentative diagnosis and to order specific quantitative measurements of certain drugs (e.g., salicylates, valproic acid, or digoxin) when the results of such tests may alter therapy. It is rarely useful, especially in the emergency management of a poisoning victim, to order a comprehensive toxicology screen. Generally, this test is performed at an outside reference laboratory at considerable expense, and patients often awaken and confirm the tentative diagnosis before results are available (usually 1 to 2 days after testing). In addition, many common dangerous drugs and poisons (e.g., isoniazid, digitalis glycosides, calcium antagonists, beta blockers, metals, and pesticides) are not included in the screening procedure; thus, a negative toxicology screen does not rule out the possibility of poisoning.[28] So-called drugs-of-abuse screens for opioids, amphetamines, and cocaine are commonly performed by hospital laboratories and are useful in identifying intoxication by these substances, but they should not be mistaken for a comprehensive toxicologic screening test.

Differential Diagnosis

Whenever a patient with suspected poisoning or drug overdose is evaluated, the possibility that other illnesses are mimicking or complicating the presentation should always be considered. These illnesses include head trauma (e.g., in the ethanol-intoxicated patient, who often falls); cerebrovascular accident; meningitis; metabolic abnormalities, such as hypoglycemia, hyponatremia, and hypoxemia; underlying liver disease; and the postictal state. In any patient with altered mental status, computed tomography of the head and lumbar puncture should be considered.

Management Issues

DECONTAMINATION AFTER ACUTE INGESTION

Nowhere in the field of toxicology is there more controversy than in the debate about gastrointestinal decontamination.[29-31] Techniques for gut decontamination include emesis, gastric lavage, administration of activated charcoal, and whole bowel irrigation [*see Table 8*].

Ipecac-induced emesis has been almost completely abandoned in the clinical setting.[32] In 2003, the American Academy of Pediatrics advised against routine home use of ipecac and recommended disposing of all ipecac found in the home.[33] One reason it has fallen out of favor is that treated patients run the risks of sudden, unexpected deterioration from the effects of the overdose and subsequent pulmonary aspiration; more important, however, is the lack of evidence of the efficacy of ipecac-induced emesis, especially when emesis is induced more than 1 hour after the ingestion.[29,30]

Gastric lavage is still an accepted method for gut decontamination in hospitalized patients who are obtunded or comatose, but several prospective, randomized, controlled trials failed to show that lavage in conjunction with the administration of activated charcoal provides better clinical results than administration of activated charcoal alone. In one study, patients given a

Table 8 Methods of Gastrointestinal Decontamination

Method and Technique	*Useful Situations*	*Comments*
Emesis: give syrup of ipecac, 30 ml p.o. in adults (15 ml in children), along with one to two glasses of water; may repeat after 30 min if no emesis occurs; alternatively, give 1–2 tbsp of liquid handwashing or dishwashing soap	Possible benefit in rare circumstances after a potentially lethal ingestion when medical care is more than 60 min away, but only under the guidance of a poison control center	No longer used in emergency departments; American Academy of Pediatrics recommends against routine home use of ipecac[32]; contraindicated in ingestions of corrosive agents and most hydrocarbons, when the patient is lethargic, or when the ingested substance is likely to cause abrupt onset of coma or seizures
Gastric lavage: insert large-bore nasogastric or orogastric tube, empty stomach contents, and lavage with 100–200 ml aliquots of water or saline until clear	Useful in obtunded or comatose patients, in recent ingestions (< 1 hr), or in ingestion of anticholinergic agents or salicylates (delayed gut emptying)	Obtunded patient should have prior endotracheal intubation to protect airway; best position is left lateral decubitus to reduce movement of poison into small intestine
Activated charcoal: give 50–60 g of charcoal slurry p.o. or by gastric tube; goal is approximately 10:1 ratio of charcoal to ingested poison; usually given with one dose of a cathartic agent	Often useful because it adsorbs most drugs and poisons; may be equally effective when given alone as when given after emesis or lavage	Not effective for ingestions of iron, lithium, potassium, sodium, or alcohols; may need to repeat two or three times or more for large ingestions; repeated dosing may also enhance elimination of some drugs
Whole bowel irrigation: give Colyte or GoLYTELY, 1–2 L/hr p.o. or by gastric tube, until rectal effluent is clear or x-ray is negative for radiopaque materials	Useful in ingestions of iron, lithium, sustained-release or enteric-coated pills, and drug packets or other foreign bodies	Generally well tolerated; no significant fluid or electrolyte gain or loss occurs; most useful in awake, ambulatory patients; may reduce effectiveness of activated charcoal

regimen of activated charcoal and patients given a combination regimen of gastric lavage and charcoal showed no significant differences in all outcome parameters, including clinical deterioration, length of hospital stay, complications, and mortality.[34] Studies of volunteers have shown that only about 30% of ingested material is returned with gastric lavage.[29,31] However, many authors agree that it may still be useful if the ingested material has caused slowing of peristalsis (e.g., in the case of anticholinergic agents or opioids) or pyloric spasm (e.g., in the case of salicylates) or if a potentially life-threatening amount of poison (e.g., 5 g of a tricyclic antidepressant) was ingested.[29] Some investigators have suggested that gastric lavage is associated with an increased rate of complications, although adverse events are rare in clinical practice.[31,35]

Activated charcoal—a fine powder produced from the distillation of various organic materials—has a large surface area that is capable of adsorbing many drugs and poisons.[31] Studies of volunteers and clinical trials have suggested that administration of activated charcoal without gastric lavage may be as effective as, or superior to, its administration after gut emptying.[29,31,34] Although it seems logical that gastric lavage in combination with the use of activated charcoal would be more effective than the use of activated charcoal alone, this hypothesis has not been proved. Most clinicians now employ oral activated charcoal without prior gut emptying in the awake patient who has taken a moderate overdose of a drug or poison; some clinicians still recommend lavage after a massive ingestion of a highly toxic drug.

There is no consensus about the use of cathartic agents with activated charcoal, although it seems logical to hasten passage of the charcoal-drug material from the intestinal tract. If a cathartic agent is used, it should be limited to a single dose, and the potential adverse effects should be taken into account.[36] Adverse effects may more likely occur in the very young or old (who may not be able to tolerate fluid shifts associated with osmotic cathartics such as sorbitol) or in patients with renal insufficiency (who may not be able to tolerate large doses of magnesium or sodium).

Whole bowel irrigation is a technique that was introduced for gut cleansing before surgical or endoscopic procedures and that has recently been adopted for gut decontamination after certain ingestions.[31,37] It involves the use of a large volume of an osmotically balanced electrolyte solution, such as Colyte or GoLYTELY, that contains nonabsorbable polyethylene glycol and that cleans the gut by mechanical action without net gain or loss of fluids or electrolytes. Whole bowel irrigation is well tolerated by most awake patients. Although no controlled clinical trials to date have demonstrated improved outcome, it is recommended for those who have ingested large doses of poisons that are not well adsorbed to charcoal (e.g., iron or lithium), for those who have ingested sustained-release or enteric-coated products, and for those who have ingested drug packets or other potentially toxic foreign bodies.[31,38]

ENHANCED ELIMINATION

Measures to enhance the elimination of drugs and poisons are less popular than they were 20 years ago, primarily because it has since been recognized that the available techniques do not have a significant effect on total drug elimination of many of the most commonly ingested products and that they have little effect on the clinical course of intoxication.[9] In addition, hemodialysis is an invasive procedure that requires systemic anticoagulation and that is associated with potential morbidity. For a drug or poison to be considered for removal by hemodialysis, it should have a relatively small volume of distribution, have a slow intrinsic rate of removal (clearance), and cause life-threatening intoxication that is poorly responsive to supportive measures.[9] Only a few drugs and poisons meet these criteria [*see Table 9*]. Continuous renal replacement therapy has been utilized for enhanced removal of a few poisons (e.g., lithium), but data on its efficacy are limited.[39]

Repeated oral doses of activated charcoal can reduce the elimination half-life of some drugs and poisons by interrupting enterohepatic or enteroenteric recirculation.[29,40] This technique was introduced in the late 1970s, after studies reported its efficacy in volunteers, and it was considered a benign, noninvasive treatment. However, reports of fluid depletion and shock caused by

excessive coadministration of sorbitol, as well as the paucity of evidence of clinical benefit, have reduced the initial optimism about this treatment.[29,40]

Specific Drugs and Poisons

ACETAMINOPHEN

Acetaminophen is a widely used analgesic and antipyretic drug that is found in a number of over-the-counter and prescription products. When it is taken in combination with another drug that has acute toxic effects (e.g., an opioid), the more obvious and more rapidly apparent manifestations of the second drug may cause the clinician to overlook the subtle and nonspecific symptoms of acetaminophen poisoning. As a result, the opportunity to administer the highly effective prophylactic antidote acetylcysteine may be missed.

Acetaminophen is metabolized by various processes in the liver and, to a lesser extent, in the kidneys. One of the minor pathways of acetaminophen metabolism in the liver involves the cytochrome P-450 system (CYP 2E1), which generates a highly reactive intermediate metabolite. Normally, this toxic intermediate metabolite is readily scavenged by the intracellular antioxidant glutathione. In overdose, however, exhaustion of glutathione stores by production of the toxic intermediate metabolite allows the metabolite to react with cellular macromolecules, leading to cell injury and death. A similar process occurs in kidney cells.

The minimum acutely toxic single dose of acetaminophen is approximately 150 to 200 mg/kg, or about 7 to 10 g in adults.[41] Alcoholics are at risk for toxicity at lower doses, particularly when the drug is taken for several days, presumably because they have increased cytochrome P-450 metabolic activity and reduced glutathione stores.[42] Enhanced susceptibility to toxic effects has also been reported in persons who are fasting and in patients receiving long-term anticonvulsant therapy[43] or taking isoniazid.[42] Severe toxicity may result in fulminant hepatic and renal failure.[41]

Table 9 Methods of and Indications for Enhanced Drug Removal

Drug or Poison	*Preferred Elimination Method and Indications*
Carbamazepine	Hemoperfusion is indicated for severe poisoning with status epilepticus or cardiotoxicity; repeated doses of charcoal are of possible benefit for mild to moderate poisoning and for gut decontamination
Ethanol, isopropyl alcohol	Hemodialysis is rarely indicated because supportive care is generally successful; consider hemodialysis for deep coma with refractory hypotension
Lithium	Hemodialysis is indicated for severe neurologic manifestations (deep coma or seizures); I.V. saline is fairly effective for mild to moderate intoxication
Methanol, ethylene glycol	Hemodialysis is indicated for severe acidosis or for estimated or measured drug levels > 20–50 mg/dl
Phenobarbital	Hemoperfusion is indicated for refractory shock and drug levels > 200 mg/L; repeated doses of charcoal are of questionable clinical benefit
Salicylates	Hemodialysis is indicated for severe acidosis and drug levels > 100 mg/dl; consider hemodialysis at lower salicylate levels (> 60 mg/dl) in elderly patients with chronic, accidental intoxication
Theophylline	Hemoperfusion or hemodialysis is indicated for drug levels > 100 mg/L or status epilepticus; repeated doses of charcoal are indicated for less severe cases
Valproic acid	Hemodialysis or hemoperfusion is indicated for severe cases (coma, acidosis, and drug levels > 1,000 mg/L); repeated doses of charcoal are of theoretical benefit

Diagnosis

Early after acute ingestion of acetaminophen, the patient may have few or no symptoms.[44] Vomiting is not uncommon in those who have taken large doses. Other than what can be found in the patient's history, the only reliable early diagnostic clue is provided by a quantitative measurement of the serum acetaminophen level, which can be provided immediately by most hospital laboratories. Clinical evidence of liver and kidney damage is usually delayed for 24 hours or more after ingestion. An acute massive overdose of acetaminophen (i.e., levels greater than 500 to 600 mg/L) can cause transient metabolic acidosis.[45]

The earliest evidence of toxicity in most patients is elevated levels of hepatic aminotransferases (i.e., AST and alanine aminotransferase [ALT]), followed by a rising prothrombin time (PT) and bilirubin levels. Hypoglycemia, metabolic acidosis, and encephalopathy are signs of a poor prognosis.[41]

Treatment

Oral activated charcoal should be administered. Ipecac-induced emesis is not recommended, because it often leads to protracted vomiting, which makes administration of the oral antidote difficult. A serum acetaminophen level should be obtained approximately 4 hours after ingestion, and the result should be plotted on the Rumack-Matthew nomogram [*see Figure 1*]. Ingestion of massive quantities of acetaminophen or a modified-release preparation or the coingestion of a drug that slows gastric emptying may result in delayed peak serum acetaminophen levels; in such cases, repeated measurements of serum concentrations should be obtained. If the acetaminophen level is above the "probable toxicity" line (many clinicians use the "possible toxicity" line instead), treatment should be initiated with acetylcysteine. If the patient has additional risk factors for hepatotoxicity (e.g., long-term alcohol abuse, long-term use of anticonvulsants or isoniazid, or an unreliable history of time of ingestion), it is prudent to treat for toxicity even with levels below the lower possible toxicity line. Acetylcysteine, an antioxidant that substitutes for glutathione as a scavenger, is highly effective in preventing liver damage from acetaminophen toxicity, especially if therapy is initiated within 8 to 10 hours after the ingestion of acetaminophen.[44] It is less effective when initiated 12 to 16 hours after acetaminophen ingestion; but it should be given in such cases anyway because it still has beneficial effects, presumably owing to its antioxidant and anti-inflammatory properties and because it increases survival in patients with hepatic failure.[44] The dose of oral acetylcysteine is 140 mg/kg initially (diluted in soda or juice) followed by 70 mg/kg every 4 hours. The treatment protocol approved by the FDA for the oral administration of acetylcysteine stipulates that 17 doses (approximately 3 days of therapy) be administered; however, shorter courses have been shown to be equally effective in patients who were treated within 8 to 10 hours after ingestion of acetaminophen.[44,46] At our institution, we usually administer oral acetylcysteine until 36 hours after the in-

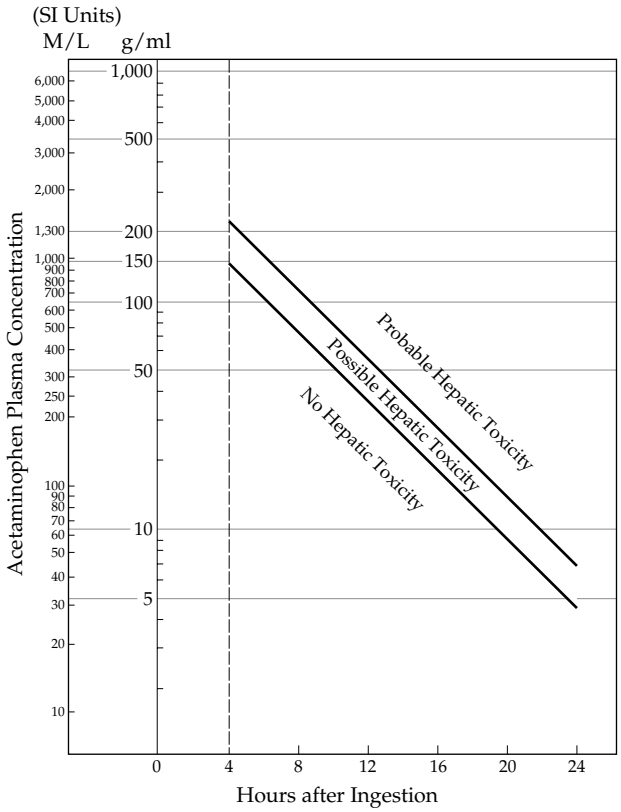


Figure 1 **The Rumack-Matthew nomogram for acetaminophen poisoning.**

gestion and then stop its administration if the liver enzymes (e.g., AST and ALT) reach normal levels. A retrospective study showed that the 36-hour regimen has a safety and efficacy profile similar to that of the traditional 72-hour protocol.[47] A longer course may be given to high-risk patients (e.g., patients who arrive in the emergency department late in the course of overdose or who have evidence of liver injury).[44]

Aggressive intervention is recommended to ensure that the loading dose is given within the first 8 hours of overdose. Occasionally, however, patients cannot tolerate oral acetylcysteine because the drug has a disagreeable odor and they are already vomiting. In such cases, it is advisable to administer the drug by the I.V. route. In 2004, the FDA approved a 20-hour acetylcysteine protocol for the treatment of acetaminophen overdose.[48] The initial loading dose is 150 mg/kg in 200 ml of 5% dextrose in water (D5W) over 15 minutes. This is followed by 50 mg/kg in 500 ml D5W over 4 hours, then 100 mg/kg in 1 L D5W over the next 16 hours. Intravenous administration can cause an anaphylactoid reaction (i.e., skin flushing and hypotension), and we usually slow the rate of the initial loading dose, administering it over 45 to 60 minutes.[49]

ANTICHOLINERGIC AGENTS AND ANTIHISTAMINES

Intoxication with anticholinergic agents can involve a variety of over-the-counter and prescription products, including antihistamines, antispasmodic agents, antipsychotic drugs, and antidepressants. In addition, several plants and mushrooms (e.g., *Datura stramonium* [angel's trumpet],[50] *Atropa belladonna,* and *Amanita phalloides*) contain potent anticholinergic alkaloids [*see Amanita phalloides* Mushrooms, *below*]. Anticholinergic agents competitively inhibit the action of acetylcholine at muscarinic receptors. Antihistamines are commonly found in a variety of over-the-counter and prescription medications for the treatment of cough and cold symptoms, itching, dizziness, nausea, and insomnia. The most commonly used nonprescription antihistamine is diphenhydramine.

Diagnosis

Clinical manifestations of intoxication with anticholinergic agents include delirium, flushed skin, dilated pupils, tachycardia, ileus, urinary retention, jerky muscle movements, and, occasionally, hyperthermia. Coma and respiratory arrest may occur. Tricyclic antidepressants (see below) and phenothiazines may also cause seizures and quinidinelike cardiac conduction abnormalities. Therefore, an ECG should be obtained and the QRS complex and cardiac rhythm monitored in any patient who displays anticholinergic manifestations of intoxication.

Antihistamine intoxication is similar to anticholinergic poisoning and may also be associated with seizures[19] and tricyclic-like cardiac conduction abnormalities.[51] The older nonsedating antihistamines terfenadine and astemizole were associated with prolongation of the QT interval and the occurrence of atypical (torsade de pointes) ventricular tachycardia both after overdose and after coadministration of macrolide antibiotics or other drugs that interfere with their elimination.[52] Because safer agents are available, both of these drugs were removed from the United States market by the manufacturers in 1999.[53]

Treatment

Activated charcoal and a cathartic agent should be administered to patients with anticholinergic or antihistamine intoxication. Gastric lavage should be considered in cases of a large ingestion; this measure may be appropriate even if some time has passed since ingestion, because ileus may delay gastric emptying. Coma and respiratory depression should be treated with the usual supportive measures. The physician should consider administering physostigmine, 0.5 to 2.0 mg in a slow I.V. infusion, in patients with pure anticholinergic intoxication (i.e., intoxication with agents other than tricyclic antidepressants or antihistamines) and severe delirium.[54] Drowsiness, confusion, and sinus tachycardia usually resolve without aggressive intervention. Prolongation of the QT interval and atypical ventricular tachycardia can be treated with magnesium, 1 to 2 g I.V., or overdrive pacing.

ANTICOAGULANTS

The anticoagulants include warfarin and the so-called superwarfarin rodenticides. Accidental intoxication with warfarin may result from long-term therapeutic overmedication or from the addition of a drug that interacts with it (e.g., allopurinol, cimetidine, nonsteroidal anti-inflammatory drugs, quinidine, salicylates, or sulfonamides). Acute ingestion of a single dose of warfarin rarely causes significant anticoagulation. However, a single dose of brodifacoum or one of the other superwarfarins can cause severe and prolonged anticoagulation that lasts for weeks to months.[55]

Diagnosis

All anticoagulants inhibit the hepatic production of clotting factors II, VII, IX, and X and prolong the PT. Circulating factors are not affected; the peak effect of anticoagulants on the PT is not seen until 36 to 48 hours after administration, when circulating

factors are degraded. Severe anticoagulation can result in hemorrhage, which may be fatal.[55]

Treatment

Acute superwarfarin overdose should be treated with oral activated charcoal and a cathartic agent. A baseline PT should be obtained on presentation and 24 and 48 hours later. If prolongation of the PT occurs, the physician should administer oral vitamin K_1 (phytonadione), 25 to 50 mg/day, and monitor the PT; in rare instances, as much as 150 to 200 mg/day may be necessary to correct the PT. It may also be necessary to continue treatment for several weeks or even months.[56] Patients should not be treated prophylactically with vitamin K_1 after an acute ingestion, because such treatment would mask the rise in PT for about 3 to 5 days or more, preventing early diagnosis. As a result, the patient would require prolonged follow-up even in the case of a subtoxic ingestion.

Vitamin K_1 may be given subcutaneously or, cautiously, by the I.V. route to patients with severe prolongation of the PT. However, because vitamin K_1 does not restore clotting factors immediately, patients who have active bleeding may require fresh frozen plasma or whole blood. Because coagulopathy after a superwarfarin overdose may last for weeks to months, high-dose oral vitamin K_1 therapy (5 mg/kg over 24 hours) may be necessary for outpatient therapy.[57]

Beta Blockers

Beta blockers are used for the treatment of hypertension, angina pectoris, migraine, and cardiac arrhythmias. Propranolol is the prototypical beta blocker but is also the most toxic [*see Table 10*].[58] All of these agents act competitively at beta-adrenergic receptors; at therapeutic doses, some have a degree of selectivity for $beta_1$- or $beta_2$-adrenergic receptors that is not apparent at high doses. Propranolol and a few of the other agents also have depressant effects on the myocardial cell membrane that are similar to those of quinidine and the tricyclic antidepressants.[59]

Beta blockade typically causes hypotension and bradycardia. Severe overdose may cause cardiogenic shock and asystole. Bronchospasm and hypoglycemia may also occur. In addition, propranolol overdose may cause widening of the QRS complex and CNS intoxication, including seizures and coma.[59] Most patients with beta-blocker poisoning manifest symptoms within 6 hours after an acute ingestion.[60]

Treatment

Treatment of overdose with a beta blocker includes aggressive gut decontamination. In cases of a large or recent ingestion, gastric lavage and the administration of activated charcoal and a cathartic agent should be initiated.

Hypotension and bradycardia are unlikely to respond to beta-adrenergic–mediated agents such as dopamine and isoproterenol; instead, the patient should receive high dosages of glucagon (5 to 10 mg I.V. followed by 5 to 10 mg/hr). Glucagon is a potent inotropic agent that does not require beta-adrenergic receptors to activate cells.[59,61] When glucagon fails, an epinephrine drip may be more beneficial in increasing heart rate and contractility than isoproterenol or dopamine. If pharmacologic therapy is unsuccessful, transvenous or external pacing should be used to maintain heart rate.[59,61] Use of hemodialysis in atenolol poisoning has been reported.[59]

Table 10 Toxicity of Common Beta Blockers[161]

Drug	*Usual Daily Dose (mg)*	*Cardioselective*	*Myocardial Cell Membrane Depression*
Acebutolol	400–800	+	+
Atenolol	50–100	+	–
Labetalol	200–800	–	+
Metoprolol	100–450	+	Variable
Nadolol	40–240	–	–
Propranolol	40–360	–	++

CALCIUM ANTAGONISTS

Calcium channel blockers are used for the treatment of angina pectoris, hypertension, hypertrophic cardiomyopathy, migraine, and supraventricular tachycardia. These agents have a relatively low toxic-to-therapeutic ratio, and life-threatening toxicity can occur after accidental or intentional overdose. Calcium antagonists block the influx of calcium through calcium channels and act mainly on vascular smooth muscle, resulting in vasodilatation, reduced cardiac contractility, and slowed AV nodal conduction and sinus node activity. The most commonly used calcium antagonists in the United States are nifedipine, verapamil, diltiazem, and amlodipine. Although each of these agents has a different spectrum of activity, this selectivity is usually lost in overdose.[59]

Diagnosis

Manifestations of intoxication with a calcium antagonist include hypotension and bradycardia. Bradycardia may result from AV block or sinus arrest with a junctional escape rhythm. The QRS complex is usually normal. Severe poisoning may cause profound shock followed by asystole. Overdose with sustained-release products, which are very popular, may be associated with delayed onset of toxicity.[62]

Treatment

Treatment of overdose of an orally administered calcium antagonist includes aggressive gut decontamination. Gastric lavage and administration of activated charcoal are recommended. For patients who have ingested a large dose of a sustained-release preparation, the physician should consider whole bowel irrigation[62] in combination with administration of repeated doses of activated charcoal; in such cases, the patient should be observed closely for possible delayed-onset effects.

Hypotension should be initially treated with boluses of fluid, vasopressors, and I.V. calcium chloride (10 ml of a 10% solution) or calcium gluconate (20 ml of a 10% solution).[12] Doses of calcium should be repeated as needed; in some case reports, as much as 10 g of calcium has been given.[63] Calcium administration may improve cardiac contractility but has less effect on AV nodal conduction or peripheral vasodilatation. Infusion of glucagon (5 to 10 mg I.V.) or epinephrine has been recommended for patients with unresponsive hypotension; in one reported case, cardiopulmonary bypass was also shown to be effective. In a verapamil-toxic canine model, the survival rate was higher with high-dose insulin therapy (i.e., insulin-dextrose infusion) than with high doses of epinephrine, calcium chloride, or glucagon. A small, uncontrolled case series of patients with calcium channel blocker poisoning showed improvement with high-dose insulin therapy,

but a prospective, controlled trial is still pending.[64] Hemodialysis is not effective.[59]

CARBON MONOXIDE

Carbon monoxide is a colorless, odorless, nonirritating gas that is produced by the combustion of organic material. It is responsible for more than 5,000 deaths in the United States each year, most occurring from suicidal inhalation. Sources of carbon monoxide include motor vehicle exhaust, improperly vented gas or wood stoves and ovens, and smoke generated by fire. Children riding under closed canopies in the backs of pickup trucks have been poisoned from the exhaust, and campers have been poisoned by using propane stoves or charcoal grills inside their tents.[65] The blizzards that hit the eastern United States in the winter of 1996 produced reports of carbon monoxide poisoning associated with snow-obstructed vehicle exhaust systems.[66] In 2005, use of portable generators in hurricane-damaged areas of Florida led to increased cases of carbon monoxide poisoning.[67]

Tissue hypoxia, which occurs as a consequence of the high affinity of carbon monoxide for hemoglobin, is the major pathophysiologic disturbance in carbon monoxide poisoning: at a carbon monoxide concentration of only 0.1%, as many as 50% of hemoglobin binding sites may be occupied by carbon monoxide. In addition to reducing the oxygen-carrying capacity of the blood, carbon monoxide interferes with the release of oxygen to the tissues. Carbon monoxide may also inhibit intracellular oxygen utilization by binding to myoglobin and cytochromes.[68]

Diagnosis

Carbon monoxide poisoning produces the symptoms and signs commonly associated with hypoxia, such as headache, confusion, tachycardia, tachypnea, syncope, hypotension, seizures, and coma. Clinical manifestations depend on the duration and intensity of exposure: an acute, sizable exposure may produce rapid unconsciousness, seizures, and death, whereas prolonged, low-level exposure may cause vague and nonspecific symptoms such as headache, dizziness, nausea, and weakness. Mild cases may be mistakenly diagnosed as influenza or migraine headache. So-called classic features of carbon monoxide poisoning, such as cherry-red skin coloring and bullous skin lesions, are not always present. Survivors of severe carbon monoxide poisoning may be left with permanent neurologic sequelae. These sequelae can include gross deficits, such as a permanent vegetative state or parkinsonism, or more subtle deficits, such as memory loss, depression, and irritability. In some cases, delayed neurologic deterioration may occur after 1 to 2 weeks.[68,69]

Laboratory findings may include metabolic acidosis and cardiac ischemia on ECG. The oxygen tension is usually normal because carbon monoxide binds to hemoglobin but does not disturb levels of dissolved oxygen; therefore, the calculated oxygen saturation is falsely normal. Furthermore, indirect measurement of oxygen saturation by pulse oximetry is inaccurate because of the similar absorption characteristics of oxyhemoglobin and carboxyhemoglobin.[70] Thus, correct diagnosis depends on direct spectrophotometric measurement of oxyhemoglobin and carboxyhemoglobin in a blood sample or direct measurement of exhaled carbon monoxide. Carboxyhemoglobin levels greater than 20% to 30% are usually associated with moderate symptoms of intoxication, and levels greater than 50% to 60% are associated with a serious or fatal outcome. There is considerable variability, however, and levels do not always correlate with symptoms.[68]

Treatment

The victim of carbon monoxide poisoning should immediately be removed from the site of exposure and given supplemental oxygen in the highest available concentration. Oxygen competes with carbon monoxide for hemoglobin binding sites, and administration of 100% oxygen can reduce the half-life of carboxyhemoglobin to approximately 40 to 60 minutes, thereby restoring normal oxygen saturation within about 2 to 3 hours. It should be noted that it is difficult to deliver 100% oxygen unless the patient is endotracheally intubated. Hyperbaric oxygen (HBO) administered in a sealed chamber can deliver oxygen at a pressure of 2.5 to 3.0 atm and has been reported to speed recovery and reduce neurologic sequelae.[68,71] Proponents of HBO therapy assert that this treatment can reduce cerebral edema and quell lipid peroxidation and other postinjury mechanisms of cellular destruction.[68,72] However, hyperbaric chambers are not readily available, and until recently, the few clinical studies to have compared HBO therapy with 100% oxygen at ambient pressure produced conflicting or inconclusive results or were otherwise unsatisfactory.[73]

In Australia in 1999, a randomized, double-blind, placebo-controlled trial (using sham HBO treatments) compared HBO with normobaric oxygen in a large number of patients with significant carbon monoxide poisoning; the authors found that HBO provided no greater benefit than normobaric oxygen.[74] A more recent study of similar design from the United States found a small but statistically significant reduction in cognitive sequelae 6 weeks after treatment.[71] Proponents of HBO generally advise its use for patients who have a history of unconsciousness, a detectable neuropsychiatric abnormality on bedside testing, or a carboxyhemoglobin level greater than 25%.[68] Because of concerns about the higher affinity of carbon monoxide for fetal hemoglobin, the recommended threshold for treatment of young infants and pregnant women is usually lower.[75] However, there are no controlled studies evaluating HBO therapy in pregnancy. It also remains unclear whether HBO may be useful in patients presenting many hours after exposure or with milder degrees of poisoning.

In patients with carbon monoxide poisoning associated with smoke inhalation, consideration should be given to the potential role of other toxic gases produced during combustion, such as cyanide, phosgene, nitrogen oxides, and hydrogen chloride, as well as the possibility that inhaled soot or steam has caused direct thermal injury to the airway and respiratory tract.

COCAINE, AMPHETAMINES, AND OTHER STIMULANTS

The 2004 National Survey on Drug Use and Health reports that 5.7 million Americans used cocaine in the past year.[76] This figure is down from 5.9 million in 2003. In 2002, there were an es-

Table 11 Common Stimulant Drugs

Drug	*Street Names*
Cocaine	Coke, crack (free-base cocaine)
Methamphetamine	Speed, crystal, ice
3,4-Methylenedioxymethamphetamine (MDMA)	Ecstasy
Methylphenidate	Ritalin*
Methcathinone†	Cat

*Ritalin is the trade name, not the street name.
†An illegally synthesized ephedrine derivative.

Table 12 Corrosive Agents

Corrosive or Caustic Agent	*Comments*
Mineral acids (e.g., hydrochloric, sulfuric, nitric, and phosphoric acids)	Produce rapidly painful coagulation necrosis of skin and eyes; inhalation of mists or vapors can cause irritation, bronchospasm, and chemical pneumonitis
Hydrofluoric acid	Highly electronegative fluoride ion causes deep tissue injury, which may have a delayed onset; systemic absorption from the skin or after ingestion may cause fatal hypocalcemia or hyperkalemia[162,163]
Caustic alkalis (e.g., sodium, potassium, calcium, and ammonium hydroxides)	Injury is often progressive and deep because of tissue saponification and resulting liquefaction necrosis
Phenol (carbolic acid)	Liquid and vapor are rapidly absorbed across the skin, causing severe systemic toxicity (shock, convulsions, and coma)[164]; isopropyl alcohol may speed its removal from skin[165]
Paraquat	Ingestion causes severe corrosive injury; systemic absorption leads to progressive and ultimately fatal pulmonary fibrosis[166]

timated 199,198 cocaine-related emergency department visits.[1]

Cocaine and the amphetamines [*see Table 11*] stimulate the CNS and the sympathetic nervous system and may act directly on peripheral adrenergic receptors.[76,77] Although cocaine also has local anesthetic properties and may cause sodium channel blockade in high doses, the clinical manifestations and treatment of cocaine overdose are essentially the same as those of amphetamine overdose. These drugs can be taken orally or can be snorted, smoked, or injected. So-called crack cocaine is a crudely prepared nonpolar derivative of the hydrochloride salt that is more easily volatilized and is thus the preferred form for smoking. The combined use of ethanol and cocaine may create the highly potent metabolite cocaethylene, which has a longer half-life than does cocaine and may contribute to the development of delayed toxic effects.[77,78]

Another common drug of abuse, particularly among teenagers and young adults, is methylenedioxymethamphetamine (MDMA), or ecstasy. National surveys suggest a marked increase in the prevalence of MDMA use in the United States. In 1993, there were 168,000 new users of MDMA; in 2001, the number of new users had soared to 1.8 million.[79] The 2001 National Survey on Drug Use and Health reports that 3.2% of teenagers (aged 12 to 17) and 13.1% of young adults (aged 18 to 25) have used MDMA.[79] Additionally, MDMA-related emergency department visits increased significantly from 253 visits in 1999 to 5,542 visits in 2001 and then declined to 4,026 in 2002.[80] Although MDMA is an amphetamine derivative with psychoactive properties similar to those of the hallucinogen mescaline, MDMA toxicity appears to be related to its stimulant properties. The subjective effects of MDMA include euphoria, sexual arousal, enhanced sensory perception, increased endurance, and greater sociability.[81] Adverse reactions from MDMA abuse reported in the literature include hyperthermia, hyponatremia, seizures, hepatitis, cerebrovascular accidents, and cardiac arrhythmias.[81] As MDMA use rises, health care providers are likely to see more patients with adverse reactions from this drug.[82]

Very limited national data regarding abuse of prescription stimulants, particularly methylphenidate (Ritalin), indicate that rates of abuse in children and teenagers are declining. Past-year rates of abuse have been tracked only since 2001; the data indicate an overall decrease from 2001 to 2004 among eighth graders (2.9% to 2.5%) and 10th graders (4.8% to 3.4%). Among 12th graders, past-year rates of abuse fluctuated between 5.1% and 4.0%. Data show that the past-year rate of abuse of methylphenidate among young adults was 2.9% in both 2002 and 2003.[83] Methylphenidate toxicity is most commonly the result of therapeutic error in children treated with the drug.[84] Abuse of methylphenidate has been reported; a national survey indicated the prevalence among college students in the United States varies by region, ranging from zero to 25%.[85]

Diagnosis

Clinical manifestations of mild stimulation include euphoria, alertness, and anorexia. More severe intoxication causes agitation, psychosis, tachycardia, hypertension, and diaphoresis. The pupils are usually dilated. Severe poisoning may result in convulsions, hypertensive crisis (e.g., intracerebral hemorrhage or aortic dissection), and hyperthermia.[17,18] Consequences of severe hyperthermia include shock, brain damage, coagulopathy, and hepatic and renal failure.[22]

The differential diagnosis includes acute functional psychosis, acute exertional heatstroke, and intoxication with other drugs. Phencyclidine, a ketaminelike dissociative anesthetic, may produce stimulant effects, but victims of overdose often have a waxing-and-waning encephalopathy with periods of flaccid stupor or coma.[17] Anticholinergic agents (see above) may also cause dilated pupils, tachycardia, and agitation, but these toxins usually cause the skin to be dry and flushed; stimulants generally cause the skin to be pale, clammy, and diaphoretic.

Treatment

Mild or moderate intoxication with a stimulant can often be successfully managed by administering a sedative agent, such as diazepam or lorazepam, and by providing the patient with a quiet room. If hypertension is severe and does not improve after sedation, phentolamine (2 to 5 mg I.V. at 5- to 10-minute intervals) or nitroprusside (0.5 to 10 μg/kg/min) should be administered. For patients with tachycardia or ventricular arrhythmias, a short-acting beta blocker such as esmolol (50 to 100 μg/kg/min) is recommended, although it should be cautioned that beta blockers may worsen hypertension because of unopposed alpha-adrenergic effects of the stimulant drug.[18] Wide-complex dysrhythmias in cases of cocaine overdose should be treated with sodium bicarbonate.[76] Severe hyperthermia should be treated aggressively to prevent brain damage and multiorgan complications [*see* Hyperthermia, *above*].

Because acute myocardial infarction may occur even in young persons with normal coronary arteries, all patients with chest pain should be evaluated carefully for evidence of ischemia.[18] Other causes of chest pain in these patients may include mechanical trauma to the chest wall, pneumomediastinum from hard coughing or the Valsalva maneuver, or pectoral muscle ischemia.[18]

CORROSIVE AGENTS

A number of agents with caustic or corrosive properties [*see Table 12*] are used for a variety of purposes in industry, as cleaning agents in the home, and in hobbies. Exposure to these agents may occur accidentally or as a result of suicidal ingestion. In some cases, the corrosive effect of these agents is a direct result of the high concentration of hydrogen (H^+) or hydroxyl (OH^-) ions

and can be predicted from the very low or very high pH of the product. In other cases, toxicity may result from the product's oxidizing or alkylating or from other cytotoxic effects. Systemic toxicity can occur as a result of absorption across burned skin or after ingestion (e.g., in the case of hydrofluoric acid, phenol, or paraquat) [*see Table 12*].[86]

Diagnosis

Manifestations of toxicity usually occur immediately after exposure to the corrosive or caustic agent and include burning pain and erythema at the site of exposure. Immediate effects occur most commonly with acids. Injury caused by alkali burns can evolve over several hours and takes the form of a penetrating liquefaction necrosis. Burns may also be delayed in cases of exposure to hydrofluoric acid (hydrogen fluoride in aqueous solution); the toxicity of this agent is mediated through its fluoride component, which combines with calcium and magnesium ions. With hydrofluoric acid burns, pain and swelling may not be apparent until several hours after exposure, especially after exposure to relatively dilute solutions.

Treatment

Treatment of toxicity from corrosive or caustic agents must be initiated rapidly to reduce injury. Exposed areas should be flushed with copious amounts of plain water and any contaminated clothing removed (health providers must be careful not to become exposed while assisting victims). For patients whose eyes have been exposed to the agent, the physician should use an eyewash fountain or should splash water into the face, then pour water directly over the eyes from a pitcher or glass. Patients who have ingested a corrosive agent should drink one to two glasses of water. Although use of gastric lavage is controversial because of concerns about possible mechanical damage to the esophagus, our gastrointestinal consultants recommend gastric intubation with a small flexible tube as soon as possible after corrosive-liquid ingestion, to remove as much of the injurious material as possible. Neutralizing agents should not be administered in an attempt to normalize the pH; they may modify the pH too far in the opposite direction, and the heat of neutralization may cause thermal injury. There are a few exceptions to this rule; for example, after exposure to hydrofluoric acid, soaking the skin in a solution or gel that contains calcium (e.g., 2.5% calcium gluconate gel) or magnesium or in benzalkonium chloride may bind the toxic fluoride ion before it can be absorbed[87]; calcium is sometimes injected subcutaneously or by the intra-arterial route for deeper burns. For management of exposure to hydrofluoric acid, the physician should consult a regional poison-control center, a medical toxicologist, or a plastic or hand surgeon.

CYANIDE

Cyanide (the CN^- anion or a salt that contains this ion) is a highly toxic chemical that is used in a variety of industries, including electroplating, chemical synthesis, and laboratory analysis.[88] Cyanide is also released in the I.V. administration of nitroprusside. Acetonitrile, which is found in some glue removers for artificial fingernails, is metabolized to cyanide and has caused death in children.[88] Natural sources of cyanide (cyanogenic glycosides) include cassava, apricot pits, and several other plants and seeds. Hydrogen cyanide gas is generated from the combustion of many natural and synthetic materials that contain nitrogen and is a common component of the smoke generated by fire [*see* Smoke Inhalation, *below*].[89]

Cyanide is a highly reactive chemical that binds to intracellular cytochrome, blocking the utilization of oxygen. The resulting cellular asphyxia leads to headache, confusion, dyspnea, syncope, collapse, and death.[88,90] Although these effects occur rapidly after inhalation of hydrogen cyanide gas, symptoms of intoxication may be delayed for minutes after the ingestion of cyanide salts or even for hours after the ingestion of cyanogenic glycosides or acetonitrile.[88]

Diagnosis

A diagnosis of cyanide poisoning is based on a history of possible exposure (e.g., in a laboratory worker who attempts to commit suicide; in a person who has ingested laetrile, a cyanogenic glycoside; in a victim of smoke inhalation; or in a patient who has received a rapid high-dose infusion of nitroprusside) and the presence of characteristic symptoms. Any victim of smoke inhalation who has altered mental status should be suspected of having been poisoned with cyanide as well as with carbon monoxide. After cyanide ingestion, the victim may detect a smell of bitter almonds, but only about 50% of the general population has the ability to perceive this odor. Severe lactic acidosis is usually present. Because cyanide blocks the cellular utilization of oxygen, the oxygen content of venous blood may be elevated; a venous oxygen saturation of greater than 90% suggests the diagnosis.

Treatment

Once cyanide poisoning is suspected, immediate measures must be taken to prevent further exposure and to provide an antidote. For an ingestion, oral activated charcoal should be immediately administered; although the adsorption of cyanide to charcoal is relatively low, a standard dose of charcoal (e.g., 50 to 60 g) is sufficient to adsorb several hundred milligrams of cyanide salts. If charcoal is not available and there will be a delay before the patient reaches the hospital, emesis should be induced with ipecac. If ipecac is not available, emesis should be induced by mechanical gagging.

The antidotes for cyanide poisoning consist of nitrites, which oxidize hemoglobin to methemoglobin; in turn, methemoglobin binds free cyanide ions. If I.V. access is not immediately available, a pearl of amyl nitrite should be broken and the victim should inhale the contents. As soon as possible, sodium nitrite, 300 mg I.V., should be administered. The other antidote is sodium thiosulfate (12.5 g I.V.), which enhances the conversion of cyanide to the less toxic thiocyanate by the endogenous enzyme rhodanese. Although nitrites produce serious side effects (e.g., methemoglobinemia reduces the oxygen-carrying capacity, and vasodilatation may cause hypotension), sodium thiosulfate is relatively benign and can be used empirically as a single agent when the diagnosis is uncertain. Other potential antidotes include cobalt ethylenediaminetetraacetic acid (cobalt EDTA) and vitamin B_{12a} (hydroxocobalamin), but these agents have not been approved for use in the United States, and hydroxocobalamin, although used in the United States for the treatment of pernicious anemia, is not available in a concentrated high-strength form needed for antidotal treatment of cyanide poisoning.[88,91]

DIGITALIS GLYCOSIDES

Digitalis glycosides are found in a variety of plants, including foxglove, oleander, and rhododendron,[92] and have been used for centuries to treat heart failure. Digoxin is the most commonly prescribed digitalis glycoside. Digitalis poisoning may occur af-

Table 13 Dosing of Digoxin-Specific Antibodies

Type of Intoxication	*Dose Needed to Provide Complete Binding of Digoxin*
Acute ingestion*	Administer one vial (40 mg) for each 0.5 mg of digoxin expected to be absorbed (because bioavailability is 80%, multiply ingested dose by 0.8 to estimate absorbed dose)
Chronic intoxication†	Use the following formula to calculate the number of vials needed: $\frac{\text{Serum digoxin level (ng/ml)} \times \text{body weight (kg)}}{100}$

*Dose of digoxin-specific antibodies is based on the estimated amount of digoxin ingested.
†Dose of digoxin-specific antibodies is based on the steady-state serum digoxin level.

ter accidental or suicidal acute overdose, as a result of long-term accumulation (usually because of renal insufficiency or overmedication), or as a drug interaction. There have been many reports of elevated digoxin levels resulting from the interaction of digoxin with commonly used drugs, such as quinidine, amiodarone, and macrolide antibiotics.[93] Digitalis glycosides inhibit the sodium pump (Na^+,K^+-ATPase), which returns potassium to cells and increases the intracellular calcium concentration.[94]

Diagnosis

After an acute overdose, serum potassium levels are often elevated and AV nodal conduction is impaired, leading to varying degrees of AV block. Additionally, gastrointestinal symptoms of nausea, vomiting, and anorexia are often described after acute digitalis poisoning. With chronic poisoning, in contrast, ventricular dysrhythmias (e.g., ventricular ectopic beats or bidirectional ventricular tachycardia) predominate, and the potassium level is often normal or low, perhaps in part because of long-term coadministration of diuretic agents. The digitalis level is usually markedly elevated; however, if the sample is drawn within a few hours of overdose or within a few hours after receiving the last therapeutic dose, the result may be misleading because the drug would not have been fully distributed to tissues.[95]

Treatment

Management of acute digitalis poisoning includes gut decontamination with the oral administration of activated charcoal and, if the ingestion was large and occurred shortly before presentation, gastric lavage. Activated charcoal administered in multiple doses is effective in reducing deaths and life-threatening cardiac arrhythmias after yellow oleander poisoning.[96] This treatment has important implications for areas of the world where antidotal therapy with digoxin-specific antibodies is not available. Initially, sinus bradycardia or uncomplicated AV block should be treated with atropine (0.5 to 2 mg I.V.). A temporary pacemaker may be needed in patients with persistent symptomatic bradycardia; however, such patients should also receive digoxin-specific antibodies.

Digoxin-specific antibodies (e.g., Digibind, DigiFab) are indicated for patients with manifestations of severe intoxication (i.e., marked hyperkalemia and symptomatic dysrhythmias). These antibodies are derived from sheep and then cleaved so as to leave only the Fab fragment, which is small enough to be filtered and eliminated by the kidney after binding to digoxin. Extensive clinical experience with dixogin-specific antibodies has shown that they are safe and highly effective, with peak activity occurring within 20 to 30 minutes after administration.[7] The dose of digoxin-specific antibodies depends on the type of intoxication [*see Table 13*]. After acute ingestion, the serum level of drug does not predict the body burden because of ongoing tissue distribution[95]; therefore, the dose of digoxin-specific antibodies is calculated by estimating the amount of drug ingested. In patients with chronic poisoning in whom a steady-state digoxin level can be obtained, the body burden can be estimated on the basis of the serum level and the average apparent volume of distribution. When the ingested dose is not known or a steady-state level cannot be obtained, patients should be treated empirically: initially, one to five vials should be administered, depending on the severity of toxicity. It may also be appropriate to start with small doses and to titrate them to clinical effect in patients who have preexisting disease that requires residual digitalis effect (e.g., those with congestive heart failure or atrial fibrillation).

ETHANOL, METHANOL, AND ETHYLENE GLYCOL

Ethanol (grain alcohol) is probably the most widely used drug in the United States, and complications related to acute intoxication, as well as related medical illness and trauma, are commonly encountered. Ethanol-related illnesses account for nearly 20% of the national expenditure for hospital care, and ethanol is involved in about 50% of all fatal motor vehicle accidents.[97] Ethanol is frequently ingested with other drugs, both in suicide attempts and in recreational drug abuse. Ethylene glycol (antifreeze) and methanol (wood alcohol) are other alcohols that cause profound and often fatal poisoning when mistakenly ingested as substitutes for ethanol.

Ethanol

Diagnosis Acute ethanol intoxication produces an easily recognized state of inebriation that includes disinhibition, slurred speech, ataxia, stupor, and coma.[17] Loss of protective reflexes in the airway may permit pulmonary aspiration of gastric contents, possibly causing respiratory arrest in those who are in a deep coma. In most states, a blood ethanol level above 80 to 100 mg/dl is considered sufficient evidence to charge a driver of a car with the crime of driving while intoxicated. A level above 300 mg/dl is generally considered sufficient to cause deep coma and respiratory arrest; however, because tolerance to ethanol develops, persons with a long history of ethanol abuse who have ethanol blood levels above 300 mg/dl are often awake and even able to ambulate.[17] Acute ethanol ingestion can also cause hypoglycemia because of the inhibitory effect of ethanol on gluconeogenesis.

Treatment Treatment of ethanol intoxication usually consists of supportive care. The blood ethanol level decreases at an average (but variable) rate of about 20 mg/dl/hr,[17] and most patients are awake and ambulatory within 6 to 12 hours or less. The physician should protect the airway and, if necessary, intubate the trachea and assist ventilation. The patient should be evaluated for hypoglycemia, and glucose-containing fluids should be given as necessary; vitamin B_1, 100 mg I.V. or I.M., should be administered to malnourished patients or patients with chronic alcoholism. Hypotension, although uncommon, may result from vasodilatation and dehydration and usually responds to an I.V. bolus of fluid. Although such patients often come to medical attention because of falls, even those without a history of trauma should be examined for occult injuries (especially to the head,

neck, and abdomen) because inebriated patients often have such injuries. In addition, serious infections, vitamin deficiencies (especially of vitamin B_1 and folic acid), and metabolic abnormalities also occur frequently in patients with chronic alcoholism[18]; if any of these are present, they should be treated.

Methanol and Ethylene Glycol

Diagnosis Methanol or ethylene glycol poisoning produces an initial clinical picture that is similar to that of ethanol intoxication. However, these alcohols are gradually metabolized to highly toxic organic acids that can have disastrous effects [*see Table 14*]. After a delay of up to several hours, the patient develops severe metabolic acidosis and evidence of end-organ injury from the accumulation of the toxic acid metabolites. A diagnosis of methanol or ethylene glycol poisoning is based on the patient's history of exposure and the presence of severe metabolic acidosis. The osmolar gap is usually elevated, especially early after ingestion when the parent compounds are present, but toxic products can be present with a seemingly normal osmolar gap.[98] The serum lactate level is relatively low despite a large anion gap.[99]

Treatment If methanol or ethylene glycol poisoning is suspected, immediate measures should be instituted to reduce absorption, prevent metabolism, and remove the toxic acid metabolites.[99] If the ingestion occurred shortly before presentation (i.e., < 1 hour), gastric aspiration should be performed to remove as much of the ingested liquid as possible; activated charcoal does not efficiently adsorb the alcohols. Metabolism of the alcohols can be prevented by giving ethanol or fomepizole (4-methylpyrazole), which competitively inhibits the enzyme alcohol dehydrogenase. If ethanol is used, a loading dose of approximately 750 mg/kg orally or I.V. usually produces an ethanol level of about 100 mg/dl[99]; an infusion of 100 to 150 mg/kg/hr is given to maintain this level. An ethanol drip is difficult to manage, and the ethanol may contribute to obtundation. Fomepizole is easier to administer, has few side effects, and, if initiated early after ethylene glycol ingestion, may eliminate the need for dialysis (this is not the case for methanol). Although costly, fomepizole therapy may be less expensive than the combined costs of hemodialysis, intensive care, and serial blood work during an ethanol drip.[99] Administration of folic acid (50 mg I.V. every 4 hours), vitamin B_1 (100 mg I.M. or I.V. every 6 hours), and pyridoxine (50 mg I.V. every 6 hours) is also recommended to enhance the metabolism of the toxic organic acids. In addition, sodium bicarbonate should be given as needed to restore normal serum pH and enhance renal elimination of the toxic acid metabolites.

If the measured or estimated serum level of the toxic alcohol is greater than 50 mg/dl or if severe metabolic acidosis is present, hemodialysis is indicated to remove the parent compounds and their metabolites. During hemodialysis, the ethanol infusion is usually increased twofold, and fomepizole is administered every 4 hours to replace the respective drugs that are lost during the procedure.[99]

γ-HYDROXYBUTYRIC ACID

γ-Hydroxybutyric acid (GHB) is a naturally occurring four-carbon compound that was first synthesized in 1960. Since then, the drug has been used for various clinical purposes, including induction of general anesthesia, treatment of alcohol withdrawal and narcolepsy, and even as a protective agent during tissue ischemia.[100] In the United States, it has been available only under an FDA investigational new drug exemption for the treatment of narcolepsy. However, in the late 1980s, GHB gained popularity among some bodybuilders who believed it could enhance muscle mass through stimulation of growth hormone release. It is now promoted popularly as a sleep aid, a diet agent, and a euphorigenic drug. Its increasing use has been accompanied by a number of reports of severe and fatal effects. Its illegal recreational abuse has become a part of the underground drug culture (e.g., at rave parties and dance clubs). It has also been used to facilitate rape and assault because it produces rapid loss of consciousness. Innovative ways to continue GHB use despite FDA and Drug Enforcement Administration restrictions have included the sale of precursors of the drug such as γ-butyrolactone (GBL) and 1,4-butanediol, marketed as dietary supplements at health food stores and on the Internet under several trade names (e.g., Renewtrient and Revivarant). These precursors are metabolized to GHB in the body, and toxic effects are similar or identical to those of GHB.[101] After numerous reports of adverse reactions to these agents, including one death, the FDA asked manufacturers on January 21, 1999, to recall their GBL-containing products and warned consumers to avoid taking these products.[101]

Diagnosis

Clinically, patients poisoned by GHB or its analogues usually present with profound CNS and respiratory depression, with possible loss of laryngeal reflexes and apnea. Symptoms usually last less than 4 to 6 hours, and patients often have sudden awakening and agitation, particularly in response to painful stimuli (e.g., intubation).[102] Concurrent sinus bradycardia, myoclonic movements, and vomiting are common. Delirium and tonic-clonic seizures have been reported. There is an additive effect of GHB when it is taken in conjunction with sedative agents or al-

Table 14 Poisoning with Ethylene Glycol, Isopropyl Alcohol, or Methanol

Alcohol	*Metabolic Products*	*Treatment*
Ethylene glycol	Oxalic, hippuric, and glycolic acids cause severe anion-gap metabolic acidosis; calcium oxalate crystals precipitate in tissues and kidneys[167]	Fomepizole or ethanol infusion; perform hemodialysis if there is severe acidosis, if serum level > 20–50 mg/dl, or if osmolar gap > 10 mOsm/L
Isopropyl alcohol	Acetone causes characteristic odor; toxicity includes CNS depression, but there are no toxic acid by-products[168]	Isopropyl alcohol is a potent CNS depressant and gastric irritant, but its toxicity is usually managed supportively
Methanol	Formic acid causes severe anion-gap metabolic acidosis and visual disturbances that can lead to blindness and death[169]	Fomepizole or ethanol infusion; perform hemodialysis if there is severe acidosis, if serum level > 20–50 mg/dl, or if osmolar gap > 10 mOsm/L

cohol. GHB is absorbed within 10 to 15 minutes, and because of its short half-life of 27 minutes, plasma blood levels are undetectable within 4 to 6 hours of therapeutic ingestion.[100] Evidence suggests that GHB dependence may lead to severe withdrawal after sudden discontinuance.[100] Symptoms are similar to those of alcohol withdrawal but may last 7 to 14 days; these patients often require very large doses of benzodiazepines and barbiturates to control agitation.[103]

Treatment

There is no specific antidote for GHB. Therapy consists of airway protection, with rapid-sequence intubation if needed [*see* Initial Stabilization, *above*]. Because of the short half-life of GHB, patients without complications from GHB (e.g., prolonged hypoxia, aspiration, or untoward effects of mechanical ventilation) are often extubated and discharged from the emergency department within 3 to 7 hours.[102] Symptomatic bradycardia can be successfully treated with atropine.[102] Decontamination measures, such as gastric lavage and activated charcoal, are of little benefit because of GHB's rapid absorption, although it should be considered for large overdoses or if a coingestion is suspected. GHB withdrawal can be treated in the same manner as alcohol withdrawal, although physicians should recognize the potential need for higher doses of benzodiazepines and a longer treatment period.[104]

IRON

Iron poisoning is typically seen in children who accidentally ingest their parents' iron supplements, but intentional overdose occasionally occurs in adults.[105] Iron in large quantities is corrosive to the gastrointestinal tract, causes nausea and vomiting, and sometimes causes bloody emesis and diarrhea. Intestinal perforation occasionally occurs. Shock may result from volume loss and fluid shifts, as well as from iron-induced peripheral vasodilatation. In addition, free iron is cytotoxic, and coma, metabolic acidosis, and liver failure may develop from excessive, acute systemic absorption.[105]

Diagnosis

The diagnosis of acute iron poisoning may be based on a history of exposure or may be suspected in a patient with severe gastroenteritis and hypotension, especially if such a patient also has metabolic acidosis, hyperglycemia, and leukocytosis.[105] A plain x-ray of the abdomen (KUB view) may reveal radiopaque iron tablets. Serum iron levels in patients with severe poisoning are usually higher than 600 to 1,000 μg/dl, although lower levels may be seen if the sample is drawn late in the course of intoxication. In the past, it was common to estimate the quantity of free iron by subtracting the total iron-binding capacity (TIBC) from the serum iron level. However, it has since been shown that the TIBC is falsely elevated during iron poisoning, and this value is no longer considered useful for the purpose.[106]

Treatment

Treatment of acute iron overdose includes gut decontamination, I.V. administration of fluids, and, possibly, chelation with deferoxamine. Patients who are in shock should receive vigorous I.V. fluid replacement. Because activated charcoal does not bind iron, it should not be given unless an overdose of other drugs is also suspected. Gastric lavage may be useful in patients who have taken liquid iron preparations or chewable products; however, if intact tablets are seen on x-ray, it is unlikely that they can be removed through even the largest-bore gastric hose. Attempts to render the iron insoluble by gastric lavage with bicarbonate- or phosphate-containing solutions have proved ineffective or dangerous. Currently, the recommended method of gut decontamination in patients with large ingestions is whole bowel irrigation,[26,31] which is achieved by administering polyethylene glycol-electrolyte solution (e.g., GoLYTELY or Colyte), 1 to 2 L/hr by nasogastric tube for several hours, until the rectal effluent is clear and the x-ray shows no radiopacities.

Therapy with deferoxamine, a specific chelator of iron, is indicated in patients who have evidence of severe poisoning, but such therapy should not replace thorough gut decontamination and aggressive volume replacement.[105] The I.V. route is preferred, and an initial dosage of 10 to 15 mg/kg/hr should be given. Dosages as high as 40 to 50 mg/kg/hr may be given in particularly severe cases of poisoning. The iron-deferoxamine complex imparts an orange or vin rosé color to the urine that is sometimes used as evidence of the continued presence of chelatable (free) iron. Inasmuch as serum iron levels are readily available in most hospitals, the so-called vin rosé test is seldom used as an indication to continue therapy. Many clinicians stop administering deferoxamine as soon as the serum iron level is lower than 350 μg/dl, because prolonged infusions have been associated with acute respiratory distress syndrome (ARDS).[7]

ISONIAZID

Isoniazid is widely used in the treatment of tuberculosis. Long-term use of isoniazid has been associated with hepatitis and peripheral neuropathy [*see 4:VIII Chronic Hepatitis and 7:XIV Antimicrobial Therapy*]. Acute overdose of isoniazid is a well-known cause of seizures and metabolic acidosis.[107] Isoniazid causes acute toxicity by competing with pyridoxal 5´-phosphate (the active form of vitamin B_6), resulting in lowered γ-aminobutyric acid (GABA) levels in the brain. It also inhibits the hepatic metabolism of lactate to pyruvate. As little as 1.5 g of isoniazid may cause toxicity, with severe toxicity likely to occur after administration of 5 to 10 g.

Diagnosis

Acute overdose of isoniazid causes confusion, seizures, and coma; the onset is abrupt, often occurring within 30 to 60 minutes of ingestion. Lactic acidosis is often severe, and its severity is disproportional to the duration or intensity of seizure activity. Diagnosis is based on a history of isoniazid ingestion and should be suspected in any person who experiences the acute onset of seizures and who may be taking the drug (e.g., persons who have tuberculosis or AIDS and recent immigrants who test positive on the purified protein derivative [PPD] skin test). Results of testing for serum isoniazid levels are not generally available immediately, and routine toxicology screens do not ordinarily test for the drug.

Treatment

Activated charcoal should be administered to any person who is suspected of having isoniazid intoxication. Emesis should not be induced because of the risk of the abrupt onset of seizures and coma. Gastric lavage is appropriate in cases of large, recent ingestion. Seizures should be treated initially with diazepam, 5 to 10 mg I.V., or with lorazepam, 1 to 2 mg I.V. Vitamin B_6 is a specific antidote and should be given to all patients who have taken more than 3 to 5 g of isoniazid. In cases in which the amount of isoniazid ingested is unknown, the dose is 5 to 10 g I.V.; if the

amount is known, an equivalent gram-for-gram amount of vitamin B_6 should be given.[107] Administration of vitamin B_6 effectively stops resistant seizures and improves metabolic acidosis. It has also reportedly reversed isoniazid-induced coma.[108]

LEAD

Lead poisoning primarily occurs in the occupational setting, with exposure occurring over a period of months or years. However, lead is a ubiquitous metal found in the paint of older houses, car batteries and radiators, some pottery glazes and solders, and some folk medicines[109]; thus, it may be encountered by hobbyists, home-repair buffs, and those who use ceramic cookware.

Diagnosis

The clinical manifestations of lead poisoning are sufficiently variable and nonspecific that lead poisoning should be suspected in any patient who has multisystem illness, especially if the illness involves the neurologic, hematopoietic, and gastrointestinal systems.[109] Lead poisoning rarely results from a single ingestion, although such occurrences have been reported.[38] More commonly, exposure occurs repeatedly and gradually. Patients typically have cramplike abdominal pain or nausea and may have chronic systemic symptoms such as irritability, malaise, and weight loss. Other manifestations of lead poisoning include peripheral motor neuropathy (wristdrop) and anemia, which is often microcytic and accompanied by basophilic stippling. Lead encephalopathy, manifested by coma and seizures, is rare.

Chronic lead poisoning has been misdiagnosed as porphyria, in part because they both involve alteration of heme metabolism.[110] Diagnosis of lead poisoning is usually based on the lead level in whole blood. Symptoms generally occur in patients with lead levels above 25 to 40 μg/dl, but lower levels have been associated with impaired neurobehavioral development in children.[111] Lead levels above 80 μg/dl are often associated with severe overt toxicity. The free erythrocyte protoporphyrin (FEP) concentration, which is elevated (> 35 μg/dl) in persons with chronic intoxication, has been used to screen large populations for lead poisoning but is not sufficiently sensitive for the identification of low blood lead levels (< 30 μg/dl) in children.

Treatment

For patients with an acute ingestion of lead (e.g., a fishing weight, bullet, or curtain weight), a plain x-ray of the abdomen should be obtained. If the object is in the stomach, there is a risk that the action of stomach acid may create enough absorbable lead to cause systemic toxicity; therefore, the object should be removed by the use of cathartic agents, whole bowel irrigation, or endoscopy. Objects that clearly lie beyond the pylorus are likely to pass uneventfully into the stool, but confirmation of this supposition should be obtained by close follow-up with repeated x-rays and measurement of blood lead levels.[38]

Several chelating agents are available for the treatment of patients with acute or chronic intoxication who are symptomatic and have elevated blood lead levels.[111] The oldest chelating agent, dimercaprol, is reserved for patients with lead encephalopathy (but even this use is controversial). For less severe intoxication, the physician should administer I.V. calcium EDTA or oral succimer (meso-2,3-dimercaptosuccinic acid [DMSA]). Triple-chelation therapy with dimercaprol, EDTA, and oral succimer has been used in conjunction with whole bowel irrigation following an extremely high lead level in a 3-year-old child with encephalopathy.[111] A recent trial suggests that succimer does not provide any benefit in children with chronically elevated blood lead levels between 20 and 44 μg/dl.[112] However, the findings of this study, the indications for treatment, and the recommended agents and doses are controversial; the physician should consult with a specialist in occupational medicine or toxicology or contact a regional poison-control center for specific advice about the doses and side effects of these drugs.

Health care providers should be aware that the Occupational Safety and Health Administration (OSHA) has provided specific guidelines for monitoring and managing workers who have been exposed to lead [*see CE:VI Occupational Medicine*]; these guidelines stipulate that such workers be removed from exposure if a single blood lead level exceeds 60 μg/dl or if the average of a series of three successive periodic screening levels exceeds 50 μg/dl.[113] For further information, a regional OSHA office or an occupational medicine specialist should be consulted. (A directory of regional offices is available at the OSHA Web site, at http://www.osha.gov.) Finally, because household members of persons who have been occupationally exposed to lead may be contaminated by the poisoned individual, household members should also be evaluated for lead poisoning even if they are apparently asymptomatic, and measures should be taken to reduce or prevent further exposure.

LITHIUM

Lithium is a simple cation that is widely used for the treatment of manic-depressive illness and other psychiatric disorders. It is also used to elevate the white blood cell count in patients with severe leukopenia. Lithium is excreted renally, and severe intoxication usually results from drug accumulation caused by renal impairment or excessive overmedication. An acute single overdose, however, is less likely to result in severe poisoning.

Diagnosis

The usual therapeutic level of lithium is 0.6 to 1.2 mEq/L. Chronic intoxication can occur with levels only slightly above 1.2 mEq/L, but patients with acute overdose may remain asymptomatic despite having much higher levels early after ingestion of the drug.[114] Manifestations of lithium intoxication include confusion, lethargy, tremors, and muscle twitching. The ECG may show flattening of T waves, the presence of U waves, and prolongation of the QT interval. In severe cases, coma and convulsions may occur.[114] Symptoms may take several days to weeks to resolve, and some patients are left with permanent neurologic impairment.[115] Other toxic effects of lithium intoxication are nephrogenic diabetes insipidus and neuroleptic malignant syndrome [*see Table 5*]. These effects can also occur at therapeutic levels of the drug.

Treatment

Treatment of acute lithium overdose consists mainly of gut decontamination and fluid therapy. Because lithium is poorly adsorbed to activated charcoal, administration of this agent is not necessary unless the physician suspects that another drug has also been ingested. Gastric lavage may reduce the gastric burden of lithium. Whole bowel irrigation should be considered, especially if the patient has ingested a sustained-release form of the drug.[114] Limited experimental and anecdotal evidence suggests that administration of sodium polystyrene sulfonate reduces absorption and enhances elimination of lithium, although its role in acute lithium overdose remains to be established.[116]

Fluid therapy is an essential part of treatment of lithium intoxication. Volume should be restored with 1 to 2 L of normal saline; the I.V. administration of fluids should be continued at a rate sufficient to produce urine at a rate of about 100 ml/hr. The indications for hemodialysis in the setting of lithium toxicity are controversial. A recent review article recommends the following guidelines for hemodialysis: a lithium level greater than 6 mEq/L in any patient; a lithium level greater than 4 mEq/L in any patient on long-term lithium therapy (in contrast to an acute overdose); or a lithium level of 2.5 to 4.0 mEq/L in any patient with severe neurologic symptoms, renal insufficiency, hemodynamic instability, or neurologic instability.[114] However, a poison-control center–based study did not report any significant difference in patients with lithium toxicity in whom hemodialysis was recommended by the poison-control center but not performed and in those for whom hemodialysis was performed.[117] These authors recommended reserving hemodialysis for severe cases of lithium toxicity. Blood should be drawn at least 8 to 12 hours after the last dose of lithium is given to prevent misinterpretation, which can occur as a result of the serum level being falsely elevated before the drug is distributed in tissues. Serial lithium measurements should be obtained until the level clearly drops, to exclude ongoing absorption or rebound after hemodialysis. Consultation with a regional poison-control center, medical toxicologist, and nephrologist should be obtained early to help manage a lithium-toxic patient.

METHEMOGLOBINEMIA-INDUCING AGENTS

Methemoglobin is an oxidized form of hemoglobin that is incapable of carrying and delivering oxygen normally. A number of oxidant drugs and chemicals can convert hemoglobin to its oxidized form, causing methemoglobinemia.[118] These agents include local anesthetics (e.g., benzocaine and lidocaine), antimicrobial agents (e.g., chloroquine, dapsone, primaquine, and sulfonamides), analgesics (e.g., phenazopyridine and phenacetin), nitrites and nitrates (e.g., amyl nitrite, butyl nitrite, isobutyl nitrite, and sodium nitrite), and several miscellaneous drugs and chemicals (e.g., aminophenol, aniline dyes, bromates, chlorates, metoclopramide, nitrobenzene, nitrogen oxides, and nitroglycerin). Benzocaine-containing sprays used for topical anesthesia before certain procedures (e.g., endoscopy, intubation, and nasogastric lavage) are a common cause of methemoglobinemia.[119] Persons with glucose-6-phosphate dehydrogenase (G6PD) deficiency and congenital methemoglobin reductase deficiency are more likely than persons without these conditions to accumulate methemoglobin after exposure to an oxidant.

Diagnosis

Methemoglobinemia causes cellular asphyxia. Symptoms of mild to moderate methemoglobinemia include headache, nausea, dizziness, and dyspnea. Methemoglobin levels as low as 15% can cause the patient to appear cyanotic despite having a normal oxygen tension. The blood usually has a dark or chocolate-brown appearance. Although pulse oximetry is abnormal, the reported drop in oxygen saturation does not correlate with the actual reduction in oxyhemoglobin saturation, and specific testing for methemoglobinemia should be performed.[120]

Treatment

Mild methemoglobinemia (methemoglobin levels < 15% to 20%) usually resolves spontaneously and requires no treatment. Patients who have more severe intoxication should be given the antidote methylene blue (1 to 2 mg/kg I.V. [0.1 to 0.2 ml/kg of a 1% solution] over several minutes).[118] The dosage may be repeated once. Although symptoms and signs usually resolve quickly, methemoglobinemia may recur with the administration of long-acting oxidants such as dapsone [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*].[118]

OPIOIDS

The opioids and opiates include several synthetic and naturally occurring compounds that are widely used for their analgesic properties. Common opium derivatives include morphine, heroin, hydrocodone, and codeine. Synthetic opioids include fentanyl, methadone, and butorphanol. Preparations of hydrocodone or codeine for oral use commonly contain aspirin or acetaminophen, which may themselves be responsible for serious toxicity in an overdose. Opioids stimulate several receptors in the CNS, resulting in sedation and reduced sympathetic outflow.[6,7] Excessive opioid effect may cause coma and blunting of the respiratory response to hypercapnia. Buprenorphine is a mixed opioid agonist-antagonist that has been introduced as an alternative to methadone in outpatient drug-treatment programs. Its use in a patient addicted to opioids may precipitate acute withdrawal symptoms.[121] The opioids meperidine and dextromethorphan may cause serious rigidity and hyperthermia in persons who are taking MAO inhibitors or other serotoninergic drugs (e.g., selective serotonin reuptake inhibitors [SSRIs]).[122]

Diagnosis

Patients may have opioid intoxication as a result of unintentional overdose or attempted suicide. Signs of intoxication include lethargy or coma, pinpoint pupils, and respiratory depression. Acute noncardiogenic pulmonary edema may occur.[6] Seizures are not typical but may occur with acute propoxyphene overdose; repeated therapeutic doses of meperidine can also cause seizures, especially in persons with renal failure because of the accumulation of the metabolite normeperidine.

Diagnosis of opioid intoxication is usually not difficult in a person who is in a coma and has pinpoint pupils and apnea.[5,6] Paramedics may discover I.V. drug paraphernalia or empty prescription bottles at the scene. Exposure to other drugs, however, may complicate the clinical picture.

Treatment

The physician should immediately establish that the airway is not obstructed and that ventilation is adequate. Supplemental oxygen should then be administered as necessary. After these initial measures, the specific opioid antagonist naloxone should be given (0.2 to 2 mg I.V. or S.C.). A recent trial has shown similar results with subcutaneous and intravenous naloxone.[123] Persons who are suspected of long-term narcotic abuse should be started with smaller doses of naloxone to minimize the severity of an acute withdrawal reaction. Patients usually become fully awake within a few minutes after administration. If the initial dose is not effective, additional doses (up to 15 to 20 mg if opioid intoxication is strongly suspected) should be given until a satisfactory response is achieved. The plasma half-life of naloxone is about 60 minutes, which is shorter than that of most of the opioids whose actions it reverses; therefore, patients who respond to the antidote should be observed for at least 3 hours after the last dose for the recurrence of sedation. Traditionally thought to be an innocuous drug, naloxone has been associated with an approximately 1.6% complication rate. Com-

plications include asystole, seizures, pulmonary edema, and severe agitation.[6]

Oral ingestion of an opioid should be treated with activated charcoal. Gastric lavage should be considered in cases of large or recent overdose. There is no role for hemodialysis or other enhanced removal procedures in the treatment of opioid overdose.

ORGANOPHOSPHATES AND RELATED AGENTS

Organophosphates and carbamates are widely used as pesticides,[124] and several of the nerve agents (e.g., VX, soman, sarin) developed for chemical warfare[125] are potent organophosphates. All of these poisons inhibit the enzyme acetylcholinesterase, preventing the breakdown of acetylcholine at cholinergic synapses. Whereas the organophosphates may cause permanent damage to the enzyme, carbamates have a transient and reversible effect. Many of these agents are well absorbed through intact skin. Persons may be exposed accidentally while working with or transporting the chemicals or as a result of accidental or suicidal ingestion.

Diagnosis

Excessive activity of acetylcholine may occur at nicotinic, muscarinic, and CNS cholinergic receptors. The most common presenting symptoms of poisoning are abdominal cramps and vomiting accompanied by sweating and hypersalivation [*see Table 15*]. The patient usually has small or pinpoint pupils. Because of the mixed effects of poisoning on sympathetic ganglia and parasympathetic synapses, the heart rate may be either slow or fast. Life-threatening manifestations of acetylcholinesterase inhibition include muscle weakness with respiratory arrest, as well as severe bronchospasm. Significant volume loss may result from excessive sweating, salivation, vomiting, and diarrhea.[124]

Treatment

Contaminated clothing should be removed immediately and all exposed areas washed thoroughly with soap and water. Rescue personnel should take precautions to avoid secondary contamination from direct contact with the victim's skin, clothing, or vomitus. Xylene or other solvent vapors emanating from the victim are not life threatening to medical personnel but may cause dizziness, nausea, and headache. In patients who have ingested an organophosphate or a carbamate, gastric lavage should be performed with the use of a closed-container unit, and activated charcoal should be administered.

Specific therapy includes administration of atropine and pralidoxime (2-PAM). Atropine is not a physiologic antidote but can reverse excessive muscarinic stimulation, thereby alleviating abdominal cramps, bronchospasm, and hypersalivation. It does not reverse muscle weakness. All patients with organophosphate poisoning should also be given 2-PAM because it can chemically restore the enzyme acetylcholinesterase; in persons who go untreated, the organophosphate's binding to acetylcholinesterase may become permanent (the so-called aging effect). Because carbamates have a transient effect, 2-PAM therapy is not needed in patients who have been poisoned with these agents. However, because the exact product causing cholinergic excess is often not known initially or because the cholinergic excess may be the the result of a mix of organophosphate and carbamate, 2-PAM may be initiated empirically. Additionally, several case reports suggest that 2-PAM may be useful in carbamate poisoning.[126]

Table 15 Manifestations of Excessive Activity of Acetylcholine

Site of Activity	*Clinical Manifestations*
Postganglionic muscarinic receptors	Bradycardia; miosis; salivation; lacrimation; bronchorrhea; increased peristalsis; sweating
Autonomic ganglia	Tachycardia; hypertension
Skeletal muscle nicotinic receptors	Muscle fasciculations followed by weakness; neuromuscular paralysis
CNS cholinergic receptors	Agitation; seizures

The dosage of 2-PAM is 1 to 2 g I.V. initially, followed by a continuous infusion of 200 to 500 mg/hr, depending on the patient's response. The infusion should be continued until the patient can be weaned from the drug without experiencing recurrence of weakness or muscarinic manifestations. This process may take several days in persons who have been exposed to highly lipid-soluble agents such as fenthion or dichlorvos.[7] A so-called intermediate syndrome has been described in which some patients experience recurrent muscle weakness several days after initially successful treatment[127]; this syndrome may be caused by neurotoxic components of the agent, continued toxicity from a lipid-soluble product, or inadequate 2-PAM therapy.

SALICYLATES

Aspirin (acetylsalicylic acid) and other salicylates are widely used for their antipyretic, anti-inflammatory, and analgesic effects and can be found alone or in combination in a number of prescription and over-the-counter products (e.g., oil of wintergreen, Pepto-Bismol). Salicylates interfere with the metabolism of glucose and fatty acids; they also uncouple oxidative phosphorylation, leading to inefficient production of adenosine triphosphate, accumulation of lactic acid, and production of heat. Poisoning may result from an acute single ingestion (usually in a dose > 200 mg/kg) or from long-term overmedication.[128] Long-term poisoning occurs most commonly in elderly persons who regularly take large doses of aspirin (e.g., for osteoarthritis) and who gradually begin to take larger doses or in whom renal insufficiency develops. In such cases, the diagnosis of salicylism is often overlooked, and patients may be assumed to have sepsis, gastroenteritis, or pneumonia on admission to the hospital.[128]

Diagnosis

The most common initial manifestation of salicylate poisoning is hyperventilation, which occurs largely as a result of central stimulation of the respiratory drive and partly in response to metabolic acidosis. Measurement of arterial blood gases usually reveals respiratory alkalosis with predominant alkalemia and underlying metabolic acidosis. Other findings include tinnitus, confusion, and lethargy. Patients with severe intoxication may experience coma, seizures, hyperthermia, noncardiogenic pulmonary edema, and circulatory collapse. The serum salicylate level in such cases usually exceeds 100 mg/dl (1,000 mg/L), although patients with chronic intoxication may experience severe effects with much lower serum levels.[129]

Treatment

For patients with an acute ingestion, activated charcoal

should be administered and gastric lavage considered if the ingestion was large (e.g., > 10 to 15 g). Because salicylates cause pylorospasm and delay gastric emptying, lavage may be successful even after a delay of several hours. For a patient who has taken a massive ingestion, extra dosages of activated charcoal (50 to 60 g every 4 to 6 hours for the first 1 to 2 days) may be needed to achieve the desired 10-to-1 ratio of charcoal to drug. Massive ingestions, as well as those involving enteric-coated aspirin, may lead to prolonged or delayed absorption and the potential for catastrophic worsening after 1 to 2 days.[128] In such cases, close observation of the patient should be maintained, and measurement of the serum salicylate level should frequently be performed until the level clearly drops into the therapeutic range (10 to 20 mg/dl).

Enhanced elimination procedures can effectively reduce elevated salicylate levels. Alkalinization of the urine traps the ionized form of salicylate in the kidney tubules, increasing renal elimination.[120] To initiate alkalinization, the physician should add 100 mEq of sodium bicarbonate to 1 L of 5% dextrose in quarter-normal (0.225%) saline, then infuse the solution at 200 ml/hr while monitoring the pH of the urine (the goal is to achieve a pH of 7 to 8). It may be difficult to perform alkalinization in patients with volume and potassium deficits without first replacing these losses. Hemodialysis rapidly lowers serum salicylate levels and can restore fluid and electrolyte balances. Hemodialysis is recommended for patients who are unable to tolerate fluid challenges (e.g., as in cerebral edema or pulmonary edema) and those who have worsening renal insufficiency, severe metabolic acidosis, or a serum salicylate level greater than 100 mg/dl (1,000 mg/L).

SEDATIVE-HYPNOTIC AGENTS

The sedative-hypnotic agents include the barbiturates (e.g., phenobarbital, pentobarbital, butalbital, and amobarbital) and the benzodiazepines (e.g., alprazolam, diazepam, lorazepam, and triazolam), as well as several other drugs, such as meprobamate, glutethimide, ethchlorvynol, chloral hydrate, zolpidem, zaleplon, and buspirone. These drugs cause generalized depression of CNS activity and are commonly used to alleviate anxiety or to induce sleep. The mechanisms of action and pharmacokinetics are different for each drug group.[13,130,131]

Diagnosis

Overdose of a sedative-hypnotic drug causes lethargy, ataxia, and slurred speech. In patients with severe poisoning, coma and respiratory arrest may occur, especially when sedative-hypnotic drugs are combined with other depressants, such as ethanol. The blood pressure and pulse rate are usually decreased, the temperature may be low because of exposure and venodilatation, and the pupils are usually small (although they may be dilated in patients with glutethimide overdose). Patients who are in a deep coma may appear to be dead because they may have absent reflexes, fixed pupils, and even flat EEG tracings.[132] In patients with chloral hydrate overdose, ventricular ectopy and ventricular tachycardia may develop; these effects are caused by generation of the metabolite trichloroethanol, which, like other chlorinated hydrocarbons, can sensitize the myocardium to the effects of epinephrine.[133] In cases of phenobarbital overdose, blood levels of the drug can be obtained in most hospital laboratories, but in cases of overdose of most of the other sedative-hypnotic agents, blood levels are neither clinically useful nor readily available.

Treatment

An unobstructed airway should be maintained and supplemental oxygen should be administered. The trachea should then be intubated and assisted ventilation initiated, if necessary. Uncomplicated hypothermia should be treated with gradual passive external rewarming. I.V. crystalloids should be administered to patients with low blood pressure; if necessary, dopamine and other pressor agents should be given. For patients with ventricular arrhythmias caused by chloral hydrate overdose, propranolol (1 to 5 mg I.V.) or esmolol (25 to 100 μg/kg/min) should be given.[134] Activated charcoal should be administered. For cases of massive ingestion, gastric lavage should be considered.

Flumazenil is a specific benzodiazepine antagonist that has been proved effective in reversing the coma caused by benzodiazepine overdose. It has a rapid onset of action after I.V. administration (0.5 to 3.0 mg); because its effects last for only about 2 to 3 hours, resedation may occur. Flumazenil is contraindicated in patients with a known or suspected overdose of a tricyclic antidepressant and in patients who have been given a benzodiazepine for the control of status epilepticus, because flumazenil may induce seizures in these patients. It should also not be used in patients who have increased intracranial pressure and who are receiving benzodiazepines for sedation. The use of flumazenil in persons who have been taking large quantities of benzodiazepines for long periods may provoke an acute withdrawal syndrome.[5,7,8]

Enhanced removal procedures are rarely needed in patients with sedative-hypnotic overdose because most will recover with airway management, assisted ventilation, and other supportive measures. When supportive measures fail, hemodialysis can effectively reduce blood concentrations of phenobarbital.[134]

THEOPHYLLINE

Although no longer a first-line drug, theophylline is still occasionally used for the treatment of asthma and other bronchospastic disorders, congestive heart failure, and neonatal apnea. It is available in regular and sustained-release formulations for oral use. Aminophylline, the ethylenediamine salt of theophylline, is used for I.V. infusions. Theophylline intoxication may occur after an acute single overdose or as a result of long-term overmedication.[135] Chronic intoxication may also be caused by reduced theophylline metabolism resulting from the addition of an interfering drug (e.g., cimetidine or erythromycin) or from an intercurrent illness (e.g., congestive heart failure or liver failure). The normal elimination half-life, 4 to 6 hours, may be prolonged to more than 20 hours in theophylline overdose.

Diagnosis

Acute theophylline overdose causes vomiting, tremors, and tachycardia. Laboratory findings include hypokalemia, hypophosphatemia, and hyperglycemia. These metabolic effects, as well as tachycardia and vasodilatation, are thought to be mediated through excessive $beta_2$-adrenergic stimulation. If serum theophylline levels exceed 100 mg/L, seizures, hypotension, and ventricular arrhythmias are likely to develop.[135] The seizures are often refractory to anticonvulsant therapy. Serum drug levels may not peak for 16 to 24 hours after theophylline ingestion, especially if the drug was in a sustained-release formulation.

Chronic intoxication may develop gradually, with toxicity possibly occurring at serum drug levels that are much lower than those associated with acute overdose: seizures have been

reported to occur at levels as low as 14 to 35 mg/L.[135] Unlike the findings in acute overdose, hypokalemia and hypotension are not common.

Treatment

In cases of acute ingestion of theophylline, activated charcoal should be given. Gastric lavage should be considered for large ingestions (i.e., more than 15 to 20 tablets). However, it is unlikely that lavage will remove intact sustained-release tablets, and severe or fatal intoxication may ensue despite aggressive attempts at decontamination.[136] Although some toxicologists have suggested administering repeated doses of activated charcoal in combination with whole bowel irrigation for massive ingestions of sustained-release medications, this approach remains controversial.[136]

Hypotension should be treated with esmolol (25 to 100 μg/kg/min) rather than a beta-adrenergic agonist because the hypotension is probably caused by beta$_2$-adrenergic–mediated vasodilatation.[137] Seizures should be treated with phenobarbital (15 to 20 mg/kg I.V.) rather than with phenytoin, which is ineffective.[136] For patients with recurrent seizures and for those with serum theophylline levels of around 100 mg/L or greater, excess theophylline should be removed as quickly as possible by hemodialysis or hemoperfusion.[138] Administration of multiple repeated doses of activated charcoal [*see* Enhanced Elimination, *above*] can effectively shorten the elimination half-life of theophylline, but such administration is often not practical in the critically ill patient.

TRICYCLIC ANTIDEPRESSANTS AND RELATED COMPOUNDS

Tricyclic antidepressants, also known as cyclic antidepressants, were once a leading cause of seizures and death from acute drug overdose.[134] Although most of the newer SSRI antidepressants are much less toxic [*see Table 16*], tricyclic antidepressants are still commonly used for the treatment of depression, enuresis, and other disorders.

The toxicity of the tricyclic antidepressants is caused by various pharmacologic properties of this class of agents, including anticholinergic activity, inhibition of norepinephrine reuptake, alpha-adrenergic blockade, and, most important, depression of the fast sodium channel in cardiac cells (the so-called quinidine-like or membrane-depressant effect). This last property is responsible for prolongation of conduction and depressed cardiac contractility.[139] Ingestion of approximately 1 g of a tricyclic antidepressant is likely to produce severe toxicity.

Diagnosis

Initially, persons with tricyclic antidepressant overdose have anticholinergic signs, including tachycardia; dilated pupils; reduced peristalsis; muscle twitching; and dry, flushed skin. Lethargy and slurred speech are common. The abrupt onset of seizures, coma, and hypotension signals severe toxicity, which may occur within 30 to 60 minutes of ingestion or may be delayed because of slowed gut absorption. In patients with severe intoxication, the ECG shows a QRS complex that is usually wider than 0.12 second[139,140]; however, this finding may initially be absent if the drug has not been absorbed or in cases of overdose with amoxapine or another noncardiotoxic drug. In some patients, right-axis deviation of the terminal 40 msec of the QRS complex may represent early evidence of a conduction disturbance.[140] Death may result from profound depression of cardiac conduction and contractility; respiratory arrest; or complications of pulmonary aspiration, aspiration pneumonia, or hyperthermia (caused by muscle twitching and seizures coupled with the absence of sweating).

Table 16 Common Tricyclic and Other Antidepressants

Tricyclic antidepressants and related agents (may induce cardiotoxicity, including widening of the QRS complex)
- Amitriptyline
- Desipramine
- Doxepin
- Imipramine
- Maprotiline
- Nortriptyline

Newer-generation antidepressants (cardiotoxicity is unlikely but seizures may occur)
- Amoxapine
- Bupropion
- Fluoxetine
- Paroxetine
- Sertraline
- Trazodone
- Venlafaxine

Treatment

The physician should administer activated charcoal. Gastric lavage should be considered for patients with massive ingestions (e.g., > 4 to 5 g), especially if less than 1 hour has elapsed since the overdose. All patients should be monitored closely for at least 6 hours; any person with altered mental status, evidence of anticholinergic toxicity, or cardiac conduction abnormalities should be admitted to the hospital and monitored closely. The physician should maintain an unobstructed airway, intubate the trachea, and assist ventilation if needed.

Seizures should be treated with benzodiazepines and phenobarbital (see above). Physostigmine should not be administered, because it may cause seizures and can worsen cardiac conduction disturbances. Initially, hypotension should be treated with I.V. boluses of normal saline. If there is evidence of depression of the sodium channel (i.e., a wide QRS complex), sodium bicarbonate should be administered at a dosage of 50 to 100 mEq I.V.[11,139] Repeated doses may be given as needed, although the serum pH should be monitored for excessive alkalemia. If hypotension does not respond to administration of fluids and sodium bicarbonate, dopamine or norepinephrine should be given. Norepinephrine may be more effective than dopamine in some patients, possibly because of tricyclic antidepressant–induced depletion of norepinephrine, but in one study, no difference between these agents was found.[139] Partial cardiopulmonary bypass has been suggested for patients with refractory hypotension and agonal cardiac rhythm, although there is little likelihood of survival.[12] There is no known role for hemodialysis in this setting.

Food Poisoning

A variety of toxins may produce illness after consumption of fish, shellfish, or mushrooms. Illness caused by bacterial or viral contamination of food, including botulism, is discussed elsewhere [*see 7:V Anaerobic Infections*].

SEAFOOD

The mechanism of toxicity varies with each toxin [*see Table 17*]. In general, the seafood-associated toxins are heat stable; therefore, cooking does not render the food safe to eat. In some cases (e.g., ciguatera and paralytic shellfish poisoning [PSP]), the poisons are highly potent neurotoxins elaborated by dinoflagellates, which are then consumed by fish or concentrated by filter-feeding clams and mussels. Scombroid poisoning results from bacterial overgrowth in inadequately refrigerated fish (although the fish may look and smell fresh); scombrotoxin is a mixture of histamine and histaminelike compounds produced by the breakdown of histidine in the fish flesh. Tetrodotoxin is produced by microorganisms associated with the puffer fish (as well as the California newt and some species of South American frogs) and concentrated in various internal organs. Although the fish is deadly and ranks as the leading cause of fatal food poisoning in Japan, it is also considered a delicacy; extreme care is required in preparation of this fish by specially trained chefs to separate the edible muscle from the toxin-containing organs. Poisoning from saxitoxin (the culprit in PSP) has recently been reported in persons who ate puffer fish caught in waters near Titusville, Florida.[141]

Diagnosis

Signs and symptoms of seafood poisoning vary with the toxin [*see Table 17*]. Diagnosis is based on the clinical presentation and history of ingested seafood. In some cases, laboratory confirmation can be carried out with the assistance of the regional or state health department.

Treatment

In general, treatment is supportive. For neurotoxic poisonings such as PSP and tetrodotoxin, prompt medical attention may be required to prevent death from sudden respiratory arrest. Scombroid poisoning is often treated with H_1 and H_2 histamine blockers (e.g., diphenhydramine and cimetidine). For ciguatera poisoning, previous anecdotal reports have suggested benefit from mannitol, but a recent randomized, controlled blinded trial showed that mannitol did not relieve symptoms of ciguatera poisoning and resulted in more side effects than normal saline.[142] Ciguatera poisoning can produce chronic symptoms, which may resemble multiple sclerosis or chronic fatigue syndrome.[143] Improvement in chronic symptoms has been reported in patients treated with amitriptyline or fluoxetine[144,145]; polyneuropathy has responded to gabapentin.[146] Recurrence of symptoms, which may be worse than the initial attack, can be triggered by ingestion of fish or alcohol.

AMANITA PHALLOIDES MUSHROOMS

The *A. phalloides* mushroom ("death cap") has been known and feared for at least two millennia and continues to cause serious illness and death, although in recent years, mortality has declined because of the availability of orthotopic liver transplantation for patients with fulminant liver failure. This mushroom, as well as several others that contain the cellular toxin amanitin (also known as amatoxin), are found throughout Europe and the United States. Most victims are amateur or novice mushroom hunters who mistake this mushroom for another, edible species. The toxin is heat stable and is not destroyed by cooking. Once absorbed, it binds to RNA polymerase and inhibits cellular protein synthesis. Hepatocytes and rapidly dividing cells are most sensitive.

Diagnosis

Severe abdominal cramps, vomiting, and diarrhea begin about 8 to 12 hours or longer after a meal. Diarrhea can be so severe that it results in severe volume depletion and cardiovascular collapse. After apparent recovery from the gastrointestinal syndrome, patients can develop rapidly progressive hepatic failure.

Treatment

Treatment of suspected amatoxin poisoning includes aggressive fluid replacement and administration of activated charcoal by mouth to bind any unabsorbed toxin in the gut and to prevent enterohepatic reabsorption, which can be significant.[147] Patients who develop severe liver injury with encephalopathy are candidates for emergency liver transplantation. Various antidotes have been described over the years, including high-dose intravenous penicillin G, corticosteroids, thioctic acid, and silibinin (an extract of the milk thistle plant), but none have proved to be effective in controlled studies, and neither thioctic acid nor silibinin is available as a pharmaceutical in the United States.[147] (Milk thistle extract can be found in some stores selling dietary and nutritional supplements, however.)

MONOSODIUM GLUTAMATE

Monosodium glutamate (MSG) is a food additive used to enhance flavor and add body to prepared foods. It is also found as a component of hydrolyzed vegetable protein. Consumption of MSG can invoke, in susceptible persons, a syndrome originally

Table 17 Seafood Poisonings[170]

Type	*Onset*	*Common Sources*	*Syndrome*	*Treatment*
Ciguatera	1–6 hr	Barracuda, red snapper, grouper	Gastrointestinal upset, paresthesias, sensation of hot and cold reversal, itching, weakness, myalgias, orthostatic hypotension	Supportive; ?mannitol
Paralytic shellfish poisoning	30 min	Bivalve mollusks (mussels, clams), associated with algae bloom (red tide)	Gastrointestinal upset, paresthesias, ataxia, weakness, respiratory muscle paralysis, respiratory arrest	Supportive
Scombroid	Minutes to hours	Tuna, mahi-mahi, bonito, mackerel	Gastrointestinal upset, flushed skin, urticaria, wheezing	Antihistamines
Tetrodotoxin	30 min	Puffer fish (fugu), sunfish, porcupine fish	Vomiting, paresthesias, perioral tingling, muscle weakness, respiratory paralysis, respiratory arrest	Supportive

coined the Chinese-restaurant syndrome and now known as the MSG symptom complex. The syndrome, which begins about 15 to 30 minutes after ingestion, includes a burning sensation or pressure in the face, behind the eyes, and in the chest, neck, shoulders, forearms, and abdomen. Headache, syncope, and, rarely, cardiac arrhythmias have been described. Not everyone who ingests MSG experiences the reaction. The etiology of the syndrome is not clearly understood. Symptoms usually last no more than 2 to 3 hours, and there is no specific treatment.[148,149]

HERBAL REMEDIES AND DIETARY SUPPLEMENTS

In 2002, about 62% of adults in the United States reported using at least one form of alternative medicine within the previous year[150] [*see CE:XII Complementary and Alternative Medicine*]. Herbal products are not subject to FDA approval, because they do not undergo the scientific testing required of conventional therapies. They cannot be promoted specifically for treatment, prevention, or cure of a disease. However, the Dietary Supplement Health and Education Act of 1994 allows these products to be sold and labeled with statements describing their professed effects. With the increasing use and availability of herbal medications, poison-control centers and health care providers are commonly encountering patients with adverse effects from impure products, drug interactions, and intentional ingestions. Ginkgo biloba has been suggested to have antiplatelet effects, and cases of spontaneous hyphema and bilateral subdural hematomas have been reported.[151] The additional risk of warfarin must be considered in patients taking Ginkgo biloba. Ephedra (ma huang) is a common ingredient in herbal weight-loss products (herbal fen-phen), stimulants (herbal ecstasy), decongestants, and bronchodilators. The active moiety in ephedra is ephedrine and related alkaloids. Serious adverse reactions, including hypertension, seizures, arrhythmias, heart attack, stroke, and death, have been reported.[152] In 2004, the FDA declared dietary supplements containing ephedra to be unsafe and banned ephedra-containing supplements.[153] However, a federal judge reversed the ban in early 2005; the future of this substance remains uncertain. St. John's wort (*Hypericum perforatum*), touted as a natural antidepressant, has been shown to inhibit serotonin, dopamine, and norepinephrine reuptake and thus presents the possibility of interaction with MAO inhibitors and other serotoninergic drugs.[152]

Adverse events associated with most herbal products are largely undescribed, and there are few specific antidotes. Emergency and supportive measures should therefore be instituted as necessary [*see* Management of Common Complications, *above*]. To enhance research and knowledge in this area, all such events should be reported to poison-control centers and to the FDA's MedWatch Program (800-FDA-1088; http://www.fda.gov/medwatch).

Smoke Inhalation

Smoke inhalation injury is the most common cause of mortality among fire victims, accounting for up to 75% of deaths.[90] Fires produce heat and smoke, although the latter is the chief culprit in inhalation injuries.[154] Smoke comprises a varying mixture of particles and gaseous chemicals that are pyrolysis products of substances that become toxic only when burned.[155] Smoke components can be broken down into simple asphyxiants, chemical asphyxiants, and irritants. Simple asphyxiants (e.g., methane and carbon dioxide) displace oxygen, thus decreasing fraction of inspired oxygen (F_IO_2) and resulting in hypoxemia. Chemical asphyxiants (e.g., carbon monoxide, cyanide, and hydrogen sulfide) cause systemic toxicity and cellular hypoxia by interrupting transport or utilization of oxygen [*see* Specific Drugs and Poisons, *above*].

Irritant gases have a direct cytotoxic effect on the oropharynx and the respiratory tract. Toxicity depends on the physical and chemical properties of the gas, which are often divided into two major groups on the basis of their water solubility. Highly water-soluble gases (e.g., ammonia, acrolein, hydrogen chloride, and sulfur dioxide) are readily absorbed in the mucous membranes along the upper respiratory tract, causing local irritation of the eyes, nose, and throat. Compounds with intermediate solubility (e.g., chlorine and isocyanates) cause upper and lower respiratory tract injury. Substances that are less water soluble (e.g., phosgene and nitrogen dioxide) do not dissolve readily in the mucous membranes of the upper respiratory tract and can reach the distal airway, producing delayed-onset pulmonary toxicity.[90,155]

Diagnosis

Clinical symptoms vary with the location of tissue injury, which in turn depends on the solubility and the concentration of exposure. Manifestations of toxicity may include conjunctival irritation, rhinitis, oropharyngeal erythema and burns, coryza, hoarseness, stridor, wheezing, coughing, and noncardiogenic pulmonary edema. Onset of pulmonary edema may be delayed from 12 to 24 hours or longer when the patient has been exposed to low-solubility gases such as phosgene and nitrogen dioxide.[90]

Treatment

Management at the scene of the exposure should include evacuation of all persons from further exposure to the smoke. Rescuers should take precautions to avoid personal exposure and should use a self-contained breathing apparatus. Although the clinician rarely has access to information regarding the constituents of the smoke, initial treatment of all victims should focus on the airway [*see* Initial Stabilization, *above*]. All patients should receive supplemental oxygen in the highest concentration while arterial blood gas and carboxyhemoglobin levels are pending [*see* Carbon Monoxide, *above*]. For patients who do not require immediate airway protection (e.g., those who are without respiratory distress, coma, or stridor), a careful plan should be sought for identifying those at high risk for potential deterioration. Many authors recommend fiberoptic bronchoscopy to help identify supraglottic and subglottic airway injury.[90] An important caveat is that lack of upper airway injury (e.g., oropharyngeal burns or singed nasal hairs) neither precludes nor predicts future airway demise. Patients should be risk-stratified on the basis of history (e.g., closed-space fire, particular materials in the fire, loss of consciousness, or history of reactive airway disease) before final disposition. Patients with any sign of airway injury or clinically significant smoke inhalation should be observed overnight. A normal initial chest radiograph is not a reliable indicator of pulmonary injury.[156] If exposure to a low-solubility toxin is likely (e.g., phosgene or nitrogen dioxide), manifestation of pulmonary injury may be delayed for 12 to 24 hours. Bronchodilators should be used for bronchospasm, but unlike treatment of patients with asthma and chronic obstructive pulmonary disease, use of steroids has not been shown to be beneficial in patients with smoke inhalation.[90] Patients with suspected cyanide poisoning should receive sodium thiosulfate [*see* Cyanide, *above*].

Kent R. Olson, M.D., has served as a paid expert consultant to McNeil Con-

sumer Products, the makers of Tylenol-brand acetaminophen.

Manish M. Patel, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Emergency department trends, final estimates 1995–2002: data from the Drug Abuse Warning Network (DAWN). US Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration (DHS Publication No [SMA] 03-3780), Rockville, Maryland, 2003
http://dawninfo.samhsa.gov/old_dawn

2. Mortality data: Data from the Drug Abuse Warning Network (DAWN). US Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration (DHS Publication No [SMA] 02-3633), Rockville, Maryland, 2002
http://dawninfo.samhsa.gov/old_dawn

3. Kharasch M, Graff J: Emergency management of the airway. Crit Care Clin 11:53, 1995

4. Powner DJ: Drug-associated isoelectric EEGs: a hazard in brain-death certification. JAMA 236:1123, 1976

5. Hoffman RS, Goldfrank LR: The poisoned patient with altered consciousness: controversies in the use of a 'coma cocktail.' JAMA 274:562, 1995

6. Sporer KA: Acute heroin overdose. Ann Intern Med 130:584, 1999

7. Bowden CA, Krenzelok EP: Clinical applications of commonly used contemporary antidotes: a US perspective. Drug Saf 16:9, 1997

8. Seger DL: Flumazenil: treatment or toxin. J Toxicol Clin Toxicol 42:209, 2004

9. Vernon DD, Gleich MC: Poisoning and drug overdose. Crit Care Clin 13:647, 1997

10. Krenzelok EP, Leikin JB: Approach to the poisoned patient. Dis Mon 42:509, 1996

11. Kolecki PF, Curry SC: Poisoning by sodium channel blocking agents. Crit Care Clin 13:829, 1997

12. Albertson TE, Dawson A, de Latorre F, et al: TOX-ACLS: toxicologic-oriented advanced cardiac life support. Ann Emerg Med 37:S78, 2001

13. Sing K, Erickson T, Amitai Y, et al: Chloral hydrate toxicity from oral and intravenous administration. J Toxicol Clin Toxicol 34:101, 1996

14. Leo PJ, Hollander JE, Shih RD, et al: Phenylpropanolamine and associated myocardial injury. Ann Emerg Med 28:359, 1996

15. Volz HP, Gleiter CH: Monoamine oxidase inhibitors: a perspective on their use in the elderly. Drugs Aging 13:341, 1998

16. Hsue PY, Salinas CL, Bolger AF, et al: Acute aortic dissection related to crack cocaine. Circulation 105:1592, 2002

17. Brust JC: Acute neurologic complications of drug and alcohol abuse. Neurol Clin 16:503, 1998

18. Hoffman RS, Hollander JE: Evaluation of patients with chest pain after cocaine use. Crit Care Clin 13:809, 1997

19. Olson KR, Kearney TE, Dyer JE, et al: Seizures associated with poisoning and drug overdose. Am J Emerg Med 12:392, 1994

20. Shepherd G, Velez LI, Keyes DC: Intentional bupropion overdoses. J Emerg Med 27:147, 2004

21. Bleck TP: Management approaches to prolonged seizures and status epilepticus. Epilepsia 40(suppl 1):S59, 1999

22. Halloran LL, Bernard DW: Management of drug-induced hyperthermia. Curr Opin Pediatr 16:211, 2004

23. Kempainen RR, Brunette DD: The evaluation and management of accidental hypothermia. Respir Care 49:192, 2004

24. Antretter H, Dapunt OE, Bonatti J: Management of profound hypothermia. Br J Hosp Med 54:215, 1995

25. Huerta-Alardin AL, Varon J, Marik PE: Bench-to-bedside review: rhabdomyolysis: an overview for clinicians. Crit Care 9:158, 2005

26. Salgia AD, Kosnik SD: When acetaminophen use becomes toxic: treating acute accidental and intentional overdose. Postgrad Med 105:81, 1999

27. Kirk M, Pace S: Pearls, pitfalls, and updates in toxicology. Emerg Med Clin North Am 15:427, 1997

28. Belson MG, Simon HK: Utility of comprehensive toxicologic screens in children. Am J Emerg Med 17:221, 1999

29. Manoguerra AS: Gastrointestinal decontamination after poisoning. Where is the science? Crit Care Clin 13:709, 1997

30. Bond GR: The role of activated charcoal and gastric emptying in gastrointestinal decontamination: a state-of-the-art review. Ann Emerg Med 39:273, 2002

31. Krenzelok E, Vale A: Position statements: gut contamination. American Academy of Clinical Toxicology: European Association of Poisons Centres and Clinical Toxicologists. J Toxicol Clin Toxicol 35:695, 1997

32. Position paper: ipecac syrup. J Toxicol Clin Toxicol 42:133, 2004

33. Poison treatment in the home. American Academy of Pediatrics Committee on Injury, Violence, and Poison Prevention. Pediatrics 112:1182, 2003

34. Pond SM, Lewis-Driver DJ, Williams GM, et al: Gastric emptying in acute overdose: a prospective randomised controlled trial. Med J Aust 163:345, 1995

35. Vale JA, Kulig K: Position paper: gastric lavage. J Toxicol Clin Toxicol 42:933, 2004

36. Position paper: cathartics. J Toxicol Clin Toxicol 42:243, 2004

37. Position paper: whole bowel irrigation. J Toxicol Clin Toxicol 42:843, 2004

38. Mowad E, Haddad I, Gemmel DJ: Management of lead poisoning from ingested fishing sinkers. Arch Pediatr Adolesc Med 152:485, 1998

39. van Bommel EF, Kalmeijer MD, Ponssen HH: Treatment of life-threatening lithium toxicity with high-volume continuous venovenous hemofiltration. Am J Nephrol 20:408, 2000

40. Bradberry SM, Vale JA: Multiple-dose activated charcoal: a review of relevant clinical studies. J Toxicol Clin Toxicol 33:407, 1995

41. Makin AJ, Wendon J, Williams R: A 7-year experience of severe acetaminophen-induced hepatotoxicity (1987–1993). Gastroenterology 109:1907, 1995

42. Whitcomb DC, Block GD: Association of acetaminophen hepatotoxicity with fasting and ethanol use. JAMA 272:1845, 1994

43. Bray GP, Harrison PM, O'Grady JG, et al: Long-term anticonvulsant therapy worsens outcome in paracetamol-induced fulminant hepatic failure. Hum Exp Toxicol 11:265, 1992

44. Jones AL: Mechanism of action and value of *N*-acetylcysteine in the treatment of early and late acetaminophen poisoning: a critical review. J Toxicol Clin Toxicol 36:277, 1998

45. Roth B, Woo O, Blanc P: Early metabolic acidosis and coma after acetaminophen ingestion. Ann Emerg Med 33:452, 1999

46. Brok J, Buckley N, Gluud C: Interventions for paracetamol (acetaminophen) overdoses (Cochrane Review). Cochrane Database Syst Rev (3):CD003328, 2002

47. Woo OF, Mueller PD, Olson KR, et al: Shorter duration of oral *N*-acetylcysteine therapy for acute acetaminophen overdose. Ann Emerg Med 35:363, 2000

48. Safety labeling changes approved by FDA Center for Drug Evaluation and Research (CDER)—April 2004. Acetadote (acetylcysteine) injection, accessed 2005
http://www.fda.gov/medwatch/SAFETY/2004/apr_PI/Acetadote_PI.pdf

49. Bailey B, McGuigan MA: Management of anaphylactoid reactions to intravenous *N*-acetylcysteine. Ann Emerg Med 31:710, 1998

50. Greene GS, Patterson SG, Warner E: Ingestion of angel's trumpet: an increasingly common source of toxicity. South Med J 89:365, 1996

51. Doig JC: Drug-induced cardiac arrhythmias: incidence, prevention and management. Drug Saf 17:265, 1997

52. de Abajo FJ, Rodriguez LA: Risk of ventricular arrhythmias associated with nonsedating antihistamine drugs. Br J Clin Pharmacol 47:307, 1999

53. Gottlieb S: Antihistamine drug withdrawn by manufacturer. BMJ 319:7, 1999

54. Burns MJ, Linden CH, Graudins A, et al: A comparison of physostigmine and benzodiazepines for the treatment of anticholinergic poisoning. Ann Emerg Med 35:374, 2000

55. Chua JD, Friedenberg W: Superwarfarin poisoning. Arch Intern Med 158:1929, 1998

56. Sheen SR, Spiller HA, Grossman D: Symptomatic brodifacoum ingestion requiring high-dose phytonadione therapy. Vet Hum Toxicol 36:216, 1994

57. Bruno GR, Howland MA, McMeeking A, et al: Long-acting anticoagulant overdose: brodifacoum kinetics and optimal vitamin K dosing. Ann Emerg Med 36:262, 2000

58. Love JN, Litovitz TL, Howell JM, et al: Characterization of fatal beta blocker ingestion: a review of the American Association of Poison Control Centers data from 1985 to 1995. J Toxicol Clin Toxicol 35:353, 1997

59. DeWitt CR, Waksman JC: Pharmacology, pathophysiology, and management of calcium channel blocker and beta-blocker toxicity. Toxicol Rev 23:223, 2004

60. Love JN, Handler JA: Toxic psychosis: an unusual presentation of propranolol intoxication. Am J Emerg Med 13:536, 1995

61. Miller MB: Arrhythmias associated with drug toxicity. Emerg Med Clin North Am 16:405, 1998

62. Ashraf M, Chaudhary K, Nelson J, et al: Massive overdose of sustained-release verapamil: a case report and review of literature. Am J Med Sci 310:258, 1995

63. Buckley N, Dawson AH, Howarth D, et al: Slow-release verapamil poisoning: use of polyethylene glycol whole-bowel lavage and high-dose calcium. Med J Aust 158:202, 1993

64. Yuan TH, Kerns WP, Tomazszewski CA, et al: Insulin-glucose as adjunctive therapy for severe calcium channel antagonist poisoning. J Toxicol Clin Toxicol 37:463, 1999

65. Carbon monoxide poisoning deaths associated with camping—Georgia, March 1999. MMWR Morb Mortal Wkly Rep 48:705, 1999

66. Carbon monoxide poisonings associated with snow-obstructed vehicle exhaust systems in Philadelphia and New York City, January 1996. MMWR Morb Mortal Wkly Rep 45:1, 1996

67. Carbon monoxide poisoning from hurricane-associated use of portable generators—Florida, 2004. MMWR Morb Mortal Wkly Rep 54:697, 2005

68. Hardy KR, Thom SR: Pathophysiology and treatment of carbon monoxide poisoning. J Toxicol Clin Toxicol 32:613, 1994

69. Weaver LK: Carbon monoxide poisoning. Crit Care Clin 15:297, 1999

70. Hampson NB: Pulse oximetry in severe carbon monoxide poisoning. Chest 114:1036, 1998

71. Weaver LK, Hopkins RO, Chan KJ, et al: Hyperbaric oxygen for acute carbon monoxide poisoning. N Engl J Med 347:1057, 2002

72. Van Meter KW, Weiss L, Harch PG, et al: Should the pressure be off or on in the use of oxygen in the treatment of carbon monoxide-poisoned patients? (editorial). Ann Emerg Med 24:283, 1994

73. Buckley NA, Isbister GK, Stokes B, et al: Hyperbaric oxygen for carbon monoxide poisoning: a systematic review and critical analysis of the evidence. Toxicol Rev 24:75, 2005

74. Scheinkestel CD, Bailey M, Myles PS, et al: Hyperbaric or normobaric oxygen for acute carbon monoxide poisoning: a randomised controlled clinical trial. Med J Aust 170:203, 1999

75. Elkharrat D, Raphael JC, Korach JM, et al: Acute carbon monoxide intoxication and hyperbaric oxygen in pregnancy. Intensive Care Med 17:289, 1991

76. Results from the 2004 National Survey on Drug Use and Health: national findings. Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, 2005
http://oas.samhsa.gov/nsduh/2k4nsduh/2k4Results/apph.htm

77. Benowitz NL: Clinical pharmacology and toxicology of cocaine. Pharmacol Toxicol 72:3, 1993

78. Henning RJ, Wilson LD, Glauser JM: Cocaine plus ethanol is more cardiotoxic than cocaine or ethanol alone. Crit Care Med 22:1896, 1994

79. Ecstasy, LSD, PCP, and other hallucinogens: data from 2001 National Household Survey on Drug Abuse. Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.
http://oas.samhsa.gov/ecstasy.htm#Trends

80. Club Drugs, 2002 Update. Drug Abuse Warning Network [DAWN]. The DAWN Report, 2002
http://oas.samhsa.gov/2k4/clubdrugs/clubdrugs.cfm

81. Kalant H: The pharmacology and toxicology of "ecstasy" (MDMA) and related drugs. CMAJ 165:917, 2001

82. Patel MM, Wright DW, Ratcliff JJ, et al: Shedding new light on the "safe" club drug: methylenedioxymethamphetamine (ecstasy)-related fatalities. Acad Emerg Med 11:208, 2004

83. Pharmaceuticals: national drug threat assessment 2005. National Drug Intelligence Center, February 2005
http://www.usdoj.gov/ndic/pubs11/12620/pharma.htm

84. White SR, Yadao CM: Characterization of methylphenidate exposures reported to a regional poison control center. Arch Pediatr Adolesc Med 154:1199, 2000

85. McCabe SE, Knight JR, Teter CJ, et al: Non-medical use of prescription stimulants among US college students: prevalence and correlates from a national survey. Addiction 100:96, 2005

86. Kao WF, Dart RC, Kuffner E, et al: Ingestion of low-concentration hydrofluoric acid: an insidious and potentially fatal poisoning. Ann Emerg Med 34:35, 1999

87. Matsuno K: The treatment of hydrofluoric acid burns. Occup Med (Lond) 46:313, 1996

88. Beasley DM, Glass WI: Cyanide poisoning: pathophysiology and treatment recommendations. Occup Med (Lond) 48:427, 1998

89. Shusterman D, Alexeeff G, Hargis C, et al: Predictors of carbon monoxide and hydrogen cyanide exposure in smoke inhalation patients. J Toxicol Clin Toxicol 34:61, 1996

90. Bizovi KE, Leikin JD: Smoke inhalation among firefighters. Occup Med 10:721, 1995

91. Sauer SW, Keim ME: Hydroxocobalamin: improved public health readiness for cyanide disasters. Ann Emerg Med 37:635, 2001

92. Dasgupta A, Emerson L: Neutralization of cardiac toxins oleandrin, oleandrigenin, bufalin, and cinobufotalin by digibind: monitoring the effect by measuring free digitoxin concentrations. Life Sci 63:781, 1998

93. Marik PE, Fromm L: A case series of hospitalized patients with elevated digoxin levels. Am J Med 105:110, 1998

94. Derlet RW, Horowitz BZ: Cardiotoxic drugs. Emerg Med Clin North Am 13:771, 1995

95. Williamson KM, Thrasher KA, Fulton KB, et al: Digoxin toxicity: an evaluation in current clinical practice. Arch Intern Med 158:2444, 1999

96. de Silva HA, Fonseka MM, Pathmeswaran A, et al: Multiple-dose activated charcoal for treatment of yellow oleander poisoning: a single-blind, randomised, placebo-controlled trial. Lancet 361:1935, 2003

97. Freedland ES, McMicken DB, D'Onofrio G: Alcohol and trauma. Emerg Med Clin North Am 11:225, 1993

98. Glaser DS: Utility of the serum osmol gap in the diagnosis of methanol or ethylene glycol ingestion. Ann Emerg Med 27:343, 1996

99. Jacobsen D, McMartin KE: Antidotes for methanol and ethylene glycol poisoning. J Toxicol Clin Toxicol 35:127, 1997

100. Galloway GP, Frederick SL, Staggers FE Jr, et al: Gamma-hydroxybutyrate: an emerging drug of abuse that causes physical dependence. Addiction 92:89, 1997

101. Adverse events associated with ingestion of gamma-butyrolactone—Minnesota, New Mexico, and Texas, 1998–1999. MMWR Morb Mortal Wkly Rep 48:137, 1999

102. Chin RL, Sporer KA, Cullison B, et al: Clinical course of gamma-hydroxybutyrate overdose. Ann Emerg Med 31:716, 1998

103. Dyer JE, Roth B, Hyma BA: Gamma-hydroxybutyrate withdrawal syndrome. Ann Emerg Med 37:147, 2001

104. Addolorato G, Caputo F, Capristo E, et al: A case of gamma-hydroxybutyric acid withdrawal syndrome during alcohol addiction treatment: utility of diazepam administration. Clin Neuropharmacol 22:60, 1999

105. McGuigan MA: Acute iron poisoning. Pediatr Ann 25:33, 1996

106. Siff JE, Meldon SW, Tomassoni AJ: Usefulness of the total iron binding capacity in the evaluation and treatment of acute iron overdose. Ann Emerg Med 33:73, 1999

107. Romero JA, Kuczler FJ Jr: Isoniazid overdose: recognition and management. Am Fam Physician 57:749, 1998

108. Temmerman W, Dhondt A, Vandewoude K: Acute isoniazid intoxication: seizures, acidosis and coma. Acta Clin Belg 54:211, 1999

109. Graeme KA, Pollack CV Jr: Heavy metal toxicity, part II: lead and metal fume fever. J Emerg Med 16:171, 1998

110. Markowitz SB, Nunez CM, Klitzman S, et al: Lead poisoning due to hai ge fen: the porphyrin content of individual erythrocytes. JAMA 271:932, 1994

111. Gordon RA, Roberts G, Amin Z, et al: Aggressive approach in the treatment of acute lead encephalopathy with an extraordinarily high concentration of lead. Arch Pediatr Adolesc Med 152:1100, 1998

112. Rogan WJ, Dietrich KN, Ware JH, et al: The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. N Engl J Med 344:1421, 2001

113. Staudinger KC, Roth VS: Occupational lead poisoning. Am Fam Physician 57:719, 1998

114. Timmer RT, Sands JM: Lithium intoxication. J Am Soc Nephrol 10:666, 1999

115. Kores B, Lader MH: Irreversible lithium neurotoxicity: an overview. Clin Neuropharmacol 20:283, 1997

116. Gehrke JC, Watling SM, Gehrke CW, et al: In-vivo binding of lithium using the cation exchange resin sodium polystyrene sulfonate. Am J Emerg Med 14:37, 1996

117. Bailey B, McGuigan M: Comparison of patients hemodialyzed for lithium poisoning and those for whom dialysis was recommended by PCC but not done: what lesson can we learn? Clin Nephrol 54:388, 2000

118. Coleman MD, Coleman NA: Drug-induced methaemoglobinaemia: treatment issues. Drug Saf 14:394, 1996

119. Moore TJ, Walsh CS, Cohen MR: Reported adverse event cases of methemoglobinemia associated with benzocaine products. Arch Intern Med 164:1192, 2004

120. Sinex JE: Pulse oximetry: principles and limitations. Am J Emerg Med 17:59, 1999

121. Gowing L, Ali R, White J: Buprenorphine for the management of opioid withdrawal. Cochrane Database Syst Rev 18:CD002025, 2004

122. Mills KC: Serotonin syndrome: a clinical update. Crit Care Clin 13:763, 1997

123. Wanger K, Brough L, Macmillan I, et al: Intravenous vs subcutaneous naloxone for out-of-hospital management of presumed opioid overdose. Acad Emerg Med 5:293, 1998

124. Bardin PG, van Eeden SF, Moolman JA, et al: Organophosphate and carbamate poisoning. Arch Intern Med 154:1433, 1994

125. Holstege CP, Kirk M, Sidell FR: Chemical warfare: nerve agent poisoning. Crit Care Clin 13:923, 1997

126. Lifshitz M, Rotenberg M, Sofer S, et al: Carbamate poisoning and oxime treatment in children: a clinical and laboratory study. Pediatrics 93:652, 1994

127. De Bleecker J, Van den Neucker K, Colardyn F: Intermediate syndrome in organophosphorus poisoning: a prospective study. Crit Care Med 21:1706, 1993

128. Yip L, Dart RC, Gabow PA: Concepts and controversies in salicylate toxicity. Emerg Med Clin North Am 12:351, 1994

129. Chui PT: Anesthesia in a patient with undiagnosed salicylate poisoning presenting as intraabdominal sepsis. J Clin Anesth 11:251, 1999

130. Fraser AD: Use and abuse of the benzodiazepines. Ther Drug Monit 20:481, 1998

131. Coupey SM: Barbiturates. Pediatr Rev 18:260, 1997

132. Hojer J, Baehrendtz S, Gustafsson L: Benzodiazepine poisoning: experience of 702 admissions to an intensive care unit during a 14-year period. J Intern Med 226:117, 1989

133. Zahedi A, Grant MH, Wong DT: Successful treatment of chloral hydrate cardiac toxicity with propranolol. Am J Emerg Med 17:490, 1999

134. Herrington AM, Clifton GD: Toxicology and management of acute drug ingestions in adults. Pharmacotherapy 15:182, 1995

135. Shannon M: Hypokalemia, hyperglycemia and plasma catecholamine activity after severe theophylline intoxication. J Toxicol Clin Toxicol 32:41, 1994

136. Shannon M: Life-threatening events after theophylline overdose: a 10-year prospective analysis. Arch Intern Med 159:989, 1999

137. Kempf J, Rusterholtz T, Ber C, et al: Haemodynamic study as guideline for the use of beta blockers in acute theophylline poisoning. Intensive Care Med 22:585, 1996

138. Shannon MW: Comparative efficacy of hemodialysis and hemoperfusion in severe theophylline intoxication. Acad Emerg Med 4:674, 1997

139. Shanon M, Liebelt E: Targeted management strategies for cardiovascular toxicity from tricyclic antidepressant overdose: the pivotal role for alkalinization and sodium loading. Pediatr Emerg Care 14:293, 1998

140. Harrigan RA, Brady WJ: ECG abnormalities in tricyclic antidepressant ingestion. Am J Emerg Med 17:387, 1999

141. Update: Neurologic illness associated with eating Florida pufferfish, 2002. MMWR Morb Mortal Wkly Rep 51:414, 2002

142. Schnorf H, Taurarii M, Cundy T: Ciguatera fish poisoning: a double-blind randomized trial of mannitol therapy. Neurology 58:873, 2002

143. Ting JY, Brown AF: Ciguatera poisoning: a global issue with common management problems. Eur J Emerg Med 8:295, 2001

144. Davis RT, Villar LA: Symptomatic improvement with amitriptyline in ciguatera fish poisoning. N Engl J Med 315:65, 1986

145. Berlin RM, King SL, Blythe DG: Symptomatic improvement of chronic fatigue with fluoxetine in ciguatera fish poisoning. Med J Aust 157:567, 1992

146. Perez CM, Vasquez PA, Perret CF: Treatment of ciguatera poisoning with gabapentin. N Engl J Med 344:692, 2001

147. Yamada EG, Mohle-Boetani J, Olson KR, et al: Mushroom poisoning due to amatoxin. Northern California, Winter 1996–1997. West J Med 169:380, 1998

148. Yang WH, Drouin MA, Herbert M, et al: The monosodium glutamate symptom

complex: assessment in a double-blind, placebo-controlled, randomized study. J Allergy Clin Immunol 99:757, 1997

149. Tarasoff L, Kelly MF: Monosodium L-glutamate: a double blind study and review. Food Chem Toxicol 31:1019, 1993

150. Barnes PM, Powell-Griner E, McFann K, et al: Complementary and alternative medicine use among adults: United States, 2002. Advance Data from Vital and Health Statistics, no 343. National Center for Health Statistics, Hyattsville, Maryland, 2004 http://www.cdc.gov/nchs/data/ad/ad343.pdf

151. Matthews MK Jr: Association of Ginkgo biloba with intracerebral hemorrhage (letter). Neurology 50:1933, 1998

152. Haller CA, Benowitz NL: Adverse cardiovascular and central nervous system events associated with dietary supplements containing ephedra alkaloids. N Engl J Med 343:1833, 2000

153. Final rule declaring dietary supplements containing ephedrine alkaloids adulterated because they present an unreasonable risk: final rule. Fed Regist 69:6787, 2004

154. Hill IR: Reactions to particles in smoke. Toxicology 115:119, 1996

155. Weiss SM, Lakshminarayan S: Acute inhalation injury. Clin Chest Med 15:103, 1994

156. Wittram C, Kenny JB: The admission chest radiograph after acute inhalation injury and burns. Br J Radiol 67:751, 1994

157. Harpe C, Stoudemire A: Aetiology and treatment of neuroleptic malignant syndrome. Medical Toxicology 2:166, 1987

158. Sessler DI: Malignant hyperthermia. J Pediatr 109:9, 1986

159. Goldberg RJ, Huk M: Serotonin syndrome from trazodone and buspirone (letter). Psychosomatics 33:235, 1992

160. Sternbach H: The serotonin syndrome. Am J Psychiatry 148:705, 1991

161. Benowitz NL: Beta-adrenergic blockers. Poisoning & Drug Overdose, 4th ed. Olson KR, Ed. Appleton & Lange, Norwalk, Connecticut, 2004

162. Stremski ES, Grande GA, Ling LJ: Survival following hydrofluoric acid ingestion. Ann Emerg Med 21:1396, 1992

163. Bertolini JC: Hydrofluoric acid: a review of toxicity. J Emerg Med 10:163, 1992

164. Spiller HA, Quadrani-Kushner DA, Cleveland P: A five year evaluation of acute exposures to phenol disinfectant (26%). J Toxicol Clin Toxicol 31:307, 1993

165. Hunter DM, Timerding BL, Leonard RB, et al: Effects of isopropyl alcohol, ethanol, and polyethylene glycol/industrial methylated spirits in the treatment of acute phenol burns. Ann Emerg Med 21:1303, 1992

166. Suzuki K, Takasu N, Arita S, et al: Evaluation of severity indexes of patients with paraquat poisoning. Hum Exp Toxicol 10:21, 1991

167. Karlson-Stiber C, Persson H: Ethylene glycol poisoning: experiences from an epidemic in Sweden. J Toxicol Clin Toxicol 30:565, 1992

168. Gaudet MP, Fraser GL: Isopropanol ingestion: case report with pharmacokinetic analysis. Am J Emerg Med 7:297, 1989

169. Becker CE: Methanol poisoning. J Emerg Med 1:51, 1983

170. Kim S: Food poisoning: fish and shellfish. Poisoning & Drug Overdose, 4th ed. Olson KR, Ed. Appleton & Lange, Stamford, Connecticut, 2004

Acknowledgment

Figure 1 Tom Moore. Data from "Acetaminophen poisoning and toxicity," by B. H. Rumack and H. Matthew, in *Pediatrics* 55:871, 1975.

II BITES AND STINGS

Lawrence M. Lewis, m.d.
William H. Dribben, m.d.
Mark D. Levine, m.d.

Mammalian Bites

EPIDEMIOLOGY

There were an estimated four and a half million cases of dog bites in the United States in 1994, the most recent year for which published statistics are available.[1] Over a lifetime, over half of all Americans are bitten by a dog or cat.[2] Although the majority of victims with bite wounds treat themselves, nearly 370,000 dog-bite victims required medical attention in an emergency department in 2001.[3] This is an increase of 10% to 15% over that reported for 1992 to 1994.[4] The highest incidence of dog-bite injuries is in boys 5 to 9 years of age (60.7 per 10,000 person-years).[4] Children are also more likely to be bitten on the face, head, and neck than are adults.[4] Infection, the most common complication of bite wounds, arises from microbes either on the victim's skin or in the mouth of the human or animal inflicting the bite wound. The relative risk of infection is determined by a number of factors, including the species of animal inflicting the wound, the location of the bite, the size and depth of the wound, host factors, and the type of wound care given.[5-7] Most infections that result from mammalian bites are polymicrobial, with mixed aerobic and anaerobic species.[8] Nonetheless, certain species tend to cause infections with characteristic bacteriologic pathogens. In addition, transmission of the rabies virus may occur through bites from mammals of certain species.

DOMESTIC ANIMALS

Dog Bites

Dogs are responsible for more bite wounds in patients who seek medical attention than all other animals combined.[9] Most bites are inflicted by animals known to the victim, with strays accounting for less than 10% of reported bite-wound injuries.[9] Wounds occur most frequently on the extremities, except in young children, in whom bite wounds to the face, head, and neck are most common.[4,10] Specifically, bite wounds to the hand occur in anywhere from one fifth to one half of reported cases; such wounds are associated with an increased risk of infection, particularly tenosynovitis, closed-space compartment infection, and septic arthritis.[11] Although deaths from dog attacks are rare, this tragic scenario occurs about 12 to 15 times a year in the United States (238 dog-bite–related deaths from 1979 through 1998). Pit-bull–type dogs and Rottweilers are responsible for over half of the reported deaths.[12]

The oral bacterial flora of dogs includes *Pasteurella multocida*, *Staphylococcus aureus, Capnocytophaga canimorsus, S. epidermidis, Streptococcus* species, and a number of anaerobes. Mixed aerobic and anaerobic infection is present in about half of infected dog bites.[8] The aerobic organisms most commonly isolated from infected dog bites include *Pasteurella*, *Staphylococcus*, and *Streptococcus* species.[8,13] The anaerobic organisms most commonly isolated from dog bites include *Fusobacterium, Bacteroides, Porphyromonas*, and *Prevotella* species.[8,13] Anaerobic organisms are significantly more common in cultures from abscesses than from other types of infection.[6]

Dog bites are unlikely to become infected, with infection rates usually reported to be on the order of 5% to 10%.[14,15] However, the risk of infection is higher in older persons and in persons with diabetes, vascular disease, chronic alcoholism, or immunosuppression. Infection risk is also higher in puncture wounds, in wounds on the hand or foot or over a joint, and in wounds associated with crush injuries.[11]

Dog bites infected with *Capnocytophaga canimorsus* may cause an overwhelming sepsislike picture associated with high fever, leukocytosis, disseminated intravascular coagulation (DIC), and multiorgan failure. This complication of dog-bite injuries is seen most commonly in immunocompromised patients (e.g., those with asplenia, alcoholism, or hematologic malignancy) and carries a 25% mortality.[11]

Globally, dogs are the major reservoirs for rabies.[16] In developed countries, however, vaccination programs have reduced the prevalence of canine rabies; for example, only 117 rabid dogs were reported in the United States in 2003.[17] Rabies is discussed in detail elsewhere [*see 7:XXXI Viral Zoonoses*].

Cat Bites

Cat bites are the second most common mammalian bites in the United States, accounting for 5% to 15% of all reported bites. About two thirds of all cat bites occur on the upper extremity. Cat bites are more often puncture wounds than tearing lacerations and often appear innocuous initially. However, the infection rate for cat bites is reported to be 15% to 30%, or almost triple that for dog bites.[18,19] *P. multocida* is the major pathogen associated with cat bites, being found in about three fourths of infected cat-bite wounds.[8] *P. multocida* infection progresses rapidly, with pain, swelling, and erythema usually occurring within 24 hours.[9] The types of anaerobic organisms found in cat bites are similar to those found in dog bites (see above), although anaerobic isolates are somewhat more common in cat bites. Penetration into deep tissues with resultant osteomyelitis or septic arthritis is more common with cat bites than dog bites.[13] Cat-scratch disease (CSD) is an infection arising from a rickettsia-like organism, *Bartonella henselae*; CSD is discussed in detail elsewhere [*see 7:XI Infections Due to Brucella, Francisella, Yersinia pestis, and Bartonella*].

Cats continue to outnumber dogs by more than 2 to 1 as the most common domestic rabid animal.[15,20] Overall, however, domestic animals account for only a small minority of animal rabies cases in the United States; about 90% occur in wildlife [*see 7:XXXI Viral Zoonoses*].

Ferret Bites

Ferrets have become increasingly popular as pets; a 1996 survey suggested that there are almost 800,000 ferrets in the United States.[21] Ferret attacks are uncommon but can result in severe injury, especially to infants and small children.[22,23] These attacks (in contrast to dog or cat bites) are usually unprovoked.[22,23]

The bacteriologic flora in ferrets has not been well studied. One study showed that facultative anaerobic gram-positive cocci were the predominant organisms, followed by *Pasteurella* and *Corynebacterium* species; few strict anaerobes were detected.[24]

Ferrets are clearly capable of contracting and carrying the ra-

bies virus.[25] However, it is not known how long infected ferrets can shed virus before showing clinical signs of disease, thus making quarantine recommendations problematic. Although there is an approved rabies vaccine for use in ferrets, its efficacy in preventing rabies is currently not known.[26] For ferret bites involving animals suspected of being rabid, current recommendations are to give the patient rabies postexposure prophylaxis immediately. For cases involving ferrets that are not suspected of being rabid, the recommendation is to withhold vaccination and to observe the animal for 10 days; the patient should be vaccinated only if the animal shows signs of rabies during that period.[27]

HUMAN BITES

Human bites are the third most common mammalian bite in the United States, accounting for approximately 5% to more than 20% of bite wounds seen in urban emergency departments.[11,18,28] Most human-bite wounds occur on the extremities, with an unusually high percentage being over the metacarpal-phalangeal joint secondary to a clenched fist contacting a tooth. Traditionally, human-bite wounds have had a reputation for frequent and severe complications. Current data, however, suggest an infection rate from human-bite wounds on the order of 10% to 50%, depending on the wound type and location.[11,16,28] Occlusional/simple bite wounds to areas other than the hand probably are no more at risk for infection than any other type of bite wound and minimally more than for nonbite lacerations.[11,16,17,29,30] However, human-bite wounds to the hand are associated with infection rates of almost 50%.[18] A clenched-fist injury is considered the most serious of all human-bite wounds.[1] These injuries may appear innocent at first but progress to serious infections that may include the joint, tendons, or various compartments of the hand. These injuries require meticulous wound care, appropriate antibiotic therapy, and consultation with a hand surgeon. Bacterial pathogens associated with human-bite wounds include a number of anaerobes similar to those recovered from dog and cat bites, but with a much higher percentage of β-lactamase producers.[19,31,32] The predominant aerobes are *Staphylococcus* and *Streptococcus* species. About 10% to 30% of human-bite wounds have been shown to contain *Eikenella corrodens,* a facultative anaerobe.[18,28,33,34] *E. corrodens* is present in 25% of clenched-fist injuries and often causes serious, chronic infections.[34] Besides bacterial infection, human bites can transmit the hepatitis B virus, HIV, herpes simplex virus, tuberculosis, and even syphilis.[11,13,28,35] Prophylactic therapy against hepatitis B or HIV should be considered for patients bitten by persons considered at high risk for these diseases.

NONDOMESTIC ANIMALS

Rats

Rat bites are uncommon, representing less than 2% of the bite wounds seen in one urban emergency department.[36] Although the list of potential pathogens that could be transmitted from rat bites is daunting, infections from rat bites, including rabies, are, in fact, very infrequent.[37] Rat-bite fever is a disease caused by *Streptobacillus moniliformis,* a gram-negative rod. It is associated with fever, chills, headache, myalgia, and rash and usually begins abruptly about 3 to 10 days after inoculation [*see 7:VII Lyme Disease and Other Spirochetal Zoonoses*].

Bats

Although bat bites typically produce only trivial trauma, bat bites were responsible for almost 75% of all human rabies cases reported in the United States since 1990 and for 90% of all cases acquired in the United States from 1981 through 1998.[37,38] Studies suggest that cleaning a bite wound with soap and a virucidal agent is effective in lowering the risk of rabies transmission.[39] Rabies prophylaxis is recommended for any bat exposures unless immediate brain testing of the animal can be performed.

Table 1 Bite Wounds Requiring Prophylactic Antibiotics

Wound characteristics	Puncture wounds Full-thickness wounds Hand or foot wounds Wounds requiring surgical repair Treatment delay (> 24 hr) Human bites* Cat bites*
Patient characteristics	Age > 50 yr Immunosuppression (e.g., asplenia, alcoholism, corticosteroid use) Diabetes mellitus Peripheral vascular disease

*There is debate, but many authors recommend prophylactic antibiotic treatment for virtually all human and cat bites because of the high rate of infection.

Other Mammals

Skunks, raccoons, and foxes are also important animal reservoirs of rabies in the United States.[38] Bites from raccoons, skunks, and foxes should be regarded as likely to be rabid until proved otherwise, and the use of rabies immune globulin and rabies vaccine is warranted, particularly if rabies is endemic to the area or the animal's behavior is deemed abnormal.[39]

Bite wounds from wild animals are rare; most occur from exposure at zoos or from owning or harboring exotic animals.[40] The incidence of serious injury from wild-animal attacks among the three million visitors to Yellowstone National Park is reported to be lower than the chance of being struck by lightning.[40]

Bite wounds from nonhuman primates are rare. In addition to transmitting bacterial infection, these bites may transmit *Herpesvirus simiae* or monkey B virus [*see 7:XXXI Viral Zoonoses*]. If left untreated, monkey B virus infection often causes encephalitis, resulting in death or permanent neurologic impairment.[41] Monkey B virus is enzootic in North African and Asian monkeys, including the macaque and rhesus. Thorough scrubbing of bites or scratches with soap or detergent and irrigation for 15 minutes have been shown to reduce the viral inoculum.[42]

TREATMENT

The goals of bite-wound care are to recognize and treat serious injury (e.g., nerve or tendon laceration), avoid infection (both local and systemic), and achieve a good cosmetic result. The treatment of mammalian-bite wounds begins with a history and physical examination. The history should include when and where the bite occurred, the events leading to the bite, what type of animal was responsible for the bite, and any background information on the animal. Any treatment rendered before arrival at the facility, as well as the patient's tetanus immunization history, should be documented.

A careful physical examination to assess for arterial or major venous injury and nerve, joint, bone, or tendon involvement should be performed and documented. The wound should be explored for foreign bodies, including teeth or tooth fragments.

Table 2 Common and Important Pathogens and Antibiotic Selection for Various Mammalian Bite Wounds

Animal	*Pathogen*	*Antibiotics*
Dog	*Staphylococcus, Streptococcus, Capnocytophaga canimorsus, Pasteurella multocida,* anaerobes	Amoxicillin–clavulanic acid, third-generation fluoroquinolones, doxycycline*
Cat	*P. multocida, Staphylococcus, Streptococcus,* anaerobes	Amoxicillin–clavulanic acid, third-generation fluoroquinolones, doxycycline*
Human	*Staphylococcus, Streptococcus,* anaerobes, *Eikenella corrodens*	Amoxicillin–clavulanic acid, third-generation fluoroquinolones
Rodents	*Staphylococcus, Streptococcus, S. moniliformis*	Penicillin G (for *S. moniliformis*), amoxicillin–clavulanic acid, doxycycline*
Nonhuman primates	*Staphylococcus, Streptococcus,* monkey B virus	Amoxicillin–clavulanic acid; third-generation fluoroquinolones; acyclovir, valacyclovir, or famciclovir (for monkey B virus)

*Consider doxycycline in penicillin-allergic patients but do not use in pregnant patients or young children.

After the examination and provision of adequate anesthesia, the wound should be meticulously cleaned and irrigated. Wound soaking or scrubbing is to be avoided. Irrigation with 200 to 250 ml of normal saline or dilute povidone-iodine solution using moderate pressure (20 psi, the pressure generated using a syringe and a 19-gauge needle) has been shown to decrease wound infection fivefold.[16,43] Careful debridement of nonviable or grossly contaminated tissue may be necessary. If the risk of rabies is high, a benzalkonium chloride scrub should be used, because povidone-iodine irrigation has not been shown to be effective in reducing viral load.[16] Benzalkonium chloride should be rinsed out to avoid tissue irritation.

Whether lacerations associated with bite wounds should be treated with primary closure is an area of debate. A number of studies suggest that it is safe to suture bite-wound lacerations that are greater than 1 to 2 cm in length and less than 12 hours old—especially those about the head and face, where there is good circulation and greater concern for a good cosmetic result.[44,45] Bite wounds to the hand and to the lower extremities often result in complications and probably have increased rates of infection with primary closure.[44,46]

The use of prophylactic antibiotics for any bite wound is debatable,[47] but there is general consensus that certain wounds in all patients and most wounds in certain patients deserve prophylactic antibiotics [*see Table 1*]. The antibiotic of choice for prophylaxis of most mammalian-bite wounds is amoxicillin–clavulanic acid. For the penicillin-allergic patient, a third-generation fluoroquinolone (e.g., moxifloxacin)[48] or a cephalosporin (e.g., cefotaxime)[49] serves as a good alternative [*see Table 2*]. The timing of prophylactic antibiotics is important. Prophylactic antibiotics should be given as soon after the bite injury as possible. A systematic review suggests that prophylactic antibiotics may reduce the incidence of infection in all hand-bite wounds and human bites, regardless of location.[49]

Table 3 Recommendations for Tetanus Prophylaxis after Animal Bites

Primary Tetanus Immunization Series Received?	*Wound*	*Time since Last Tetanus Toxoid Dose*	*Recommended Prophylaxis*
Yes	Clean, minor	> 10 yr	Td*
	All other	> 5 yr	Td*
Uncertain	Clean, minor	—	Td*
	All other	—	Td*, TIG†

*Adult (> 7 yr) Td dose is 0.5 ml I.M.; DTaP (diphtheria, tetanus, acellular pertussis) vaccine may be used in patients younger than 7 yr.
†Dose of TIG is 250 units I.M.; give in opposite arm from Td.
Td—diphtheria and tetanus toxoids adsorbed TIG—tetanus immune globulin

Tetanus prophylaxis should be given to those patients who have not been immunized in the previous 10 years [*see Table 3*]. The standard adult dose is 0.5 ml of tetanus and diphtheria toxoids adsorbed, given intramuscularly. If tetanus immune globulin is required, it is usually given in a single dose of 250 units intramuscularly, but not in the same arm as the tetanus toxoid.

Postexposure rabies prophylaxis consists first and foremost of appropriate wound care. The use of rabies immune globulin and vaccine administration depend on local epidemiology, the animal involved (species and behavior), and the type of exposure[16] [*see 7:XXXI Viral Zoonoses*]. Postexposure rabies prophylaxis for domestic animal bites is warranted in any of the following circumstances: (1) the animal is observed to be abnormal, (2) the animal is not available for observation and the rate of endemic rabies in domestic animals for the region is not exceedingly low, or (3) the animal exhibited abnormal behavior, including an unprovoked attack [*see Table 4*].[26]

Snakebites

EPIDEMIOLOGY

Over 3,000 species of snakes exist worldwide. Snakes are found everywhere on Earth except for the Arctic and Antarctic, New Zealand, Madagascar, and a few small islands. Snakes live in almost all land environments and in both saltwater and freshwater.

Approximately 10% of snakes are venomous. Of the 14 families of snakes, only five include venomous species: the Colubridae, Hydrophidae (sea snakes), Elapidae (cobras, kraits, mambas, and coral snakes), Viperidae (Russell viper, puff adder, Gaboon viper, saw-scaled viper, and European viper), and Crotalidae (rattlesnake, water moccasin, copperhead, bushmaster, and fer-de-lance). Snakes are carnivores; venomous snakes use their venom to immobilize prey for digestive purposes.

The number of snakebites in the United States is estimated to be approximately 8,000 a year. Many bites occur when patients are hiking, walking, or handling a snake. Frequently, the patient is intoxicated at the time of the bite.[50] Most snakebites do not result in envenomation, but nine to 15 deaths occur annually from

Table 4 Recommendations for Rabies Postexposure Prophylaxis[27]

Animal Type	*Evaluation and Disposition of Animal*	*Postexposure Prophylaxis Recommendations*
Dogs, cats, ferrets	Healthy and available for 10 days of observation	Do not begin prophylaxis unless animal develops clinical signs of rabies*
	Rabid or suspected rabid	Immediately vaccinate
	Unknown (e.g., escaped)	Consult public health officials
Skunks, raccoons, foxes, most other carnivores; bats	Regarded as rabid unless animal proved negative by laboratory tests†	Consider immediate vaccination
Livestock, small rodents, lagomorphs (rabbits and hares), large rodents (woodchucks and beavers), other mammals	Consider individually	Consult public health officials; bites of squirrels, hamsters, guinea pigs, gerbils, chipmunks, rats, mice, other small rodents, rabbits, and hares almost never require antirabies postexposure prophylaxis

*During the 10-day observation period, begin postexposure prophylaxis at first sign of rabies in a dog, cat, or ferret that has bitten someone. If the animal exhibits clinical signs of rabies, it should be euthanized immediately and tested.
†The animal should be euthanized and tested as soon as possible. Holding for observation is not recommended. Discontinue vaccine if immunofluorescence test results of the animal are negative.

bites that do result in envenomation.[51] In the United States, rattlesnake bites most frequently result in significant envenomation and death.

Snakes are most active in the spring, when they begin to mate and are no longer hibernating[52]; however, the incidence of snakebite is highest in the summer months. Snakes remain active throughout the day and night, but because they are poikilothermic, they must contain their activity within a narrow temperature range of approximately 25° to 35° C.

CORAL SNAKES

The range of the eastern coral snake extends from North Carolina south and west to Texas. The western coral snake is found mainly in Arizona and New Mexico. Coral snakes are nocturnal and shy away from human contact.

Coral snakes are identified by their color, pattern, and permanently erect fangs. The nose of the coral snake is black, and the body has black, red, and yellow bands. The black bands do not separate the red and yellow bands, as they do on the nonvenomous but similarly banded kingsnake. This pattern is commonly remembered through the rhyme "red on black, venom lack; red on yellow, kills a fellow" [*see Figure 1*]. Coral snakes release their venom slowly, so they attach themselves to their prey and envenomate through a chewing motion. Instead of the puncture wounds typical of most snakebites, the chewing leaves what appear to be scratches on the skin.[52,53]

The bite of the eastern coral snake can be fatal. There are no confirmed fatalities from western coral snake envenomations.

PIT VIPERS

In the United States, pit vipers (Crotalidae) are found in all states except Maine, Alaska, and Hawaii. South America has nine subspecies of rattlesnakes; Mexico and Central America have four subspecies of rattlesnakes. These snakes can be found in a variety of habitats and at elevations up to 14,000 ft. The eastern and western diamondback rattlesnakes (*Crotalus adamanteus* and *C. strox*) [*see Figure 2*] are the largest and most dangerous in the United States and are found in the southwestern states and in Nevada, California, and Oklahoma.[54,55] The timber rattlesnake (*C. horridus*) is the second most dangerous rattlesnake common to the eastern United States, but it is rarely found in Delaware, Maine, Michigan, or Washington, D.C. Pigmy rattlesnakes (*Sistrurus catenatus* and *S. miliarius*) are found in areas ranging from New York to Michigan and from Texas to Arizona; they have the least toxic venom of all of the rattlesnakes.[51] Overall, rattlesnakes are responsible for 65% of envenomations in the United States. Their venom is 2.5 to 5 times more toxic than other North American species of venomous snakes.[50] Cottonmouths (*Agkistrodon piscivorus*), also known as water moccasins, live in the southern and southeastern states along streams and in low-lying trees. Copperheads (*A. contortix*) [*see Figure 3*] are found in mountains, rock piles, and sawdust piles. Their range extends from Massachusetts southwest to Texas. Cottonmouths and copperheads have only moderately toxic venom; their bite is painful but rarely fatal. In a study of 400 copperhead bites, 32 of which were treated with antivenin, 88% of bites responded to the antivenin, as evidenced by a cessation of local tissue injury progression.[56]

Pit vipers are identified by a small depression (pit) between the eyes and the nostrils bilaterally. They have a triangular-shaped head, an elliptical pupil, and fangs that fold back when the mouth is closed and unfold via a hingelike mechanism when the mouth is opened. The pit is a heat-sensitive organ that enables the snake to locate live, warm-blooded prey. Snakes can detect movement at a distance of about 40 ft. They can strike at a distance of approximately half their body length. Rattlesnakes use their rattle when threatened or endangered, not necessarily just when they are about to attack. The pit viper is aggressive and will stand its ground when provoked or cornered.[54,55] The venom is stored in glands that are located on each side of the head above the maxillae and behind the eyes. The glands are similar in function to the human submaxillary glands. The snake may discharge anywhere from 25% to 75% of its venom when biting a human. The fangs are either hollow or grooved. Even young snakes are venomous, and the venom of young snakes may be 12 times as strong as the venom of adult snakes.

ENVENOMATION

Toxicology

Snake venom has both neurotoxic and hematotoxic properties. The venom is a complex mixture of hydrolases, polypeptides, glycoproteins, and low-molecular-weight compounds. Snake venom, especially that of the Elapidae and Hydrophidae families, contains polypeptides that produce neuromuscular blockade at the presynaptic or postsynaptic terminals, or both, causing a flaccid paralysis. Composition of the venom varies greatly between species and between individual snakes. Viper venom is mainly cytotoxic, Elapidae (cobra, coral) venom is usu-

Figure 1 **The nose of the coral snake is black, and the body has black, red, and yellow bands. The black bands do not separate the red and yellow bands, as they do on the nonvenomous but similarly banded kingsnake. Snake shown is an eastern coral snake, *Micrurus fulvius*.**

Figure 2 **Diamondback rattlesnakes are the largest and most dangerous rattlesnakes in the United States. Shown is an eastern diamondback rattlesnake, *Crotalus adamanteus*.**

Figure 3 **The copperhead (*Agkistrodon contortix*) has a geographic range that extends from Massachusetts southwest to Texas. Bites from these snakes are painful but rarely fatal.**

ally neurotoxic, and Hydrophidae (sea snake) venom is mainly myotoxic.[57]

Snake venom has profound effects on coagulation pathways, causing a hypercoagulable state. Over the first few hours after a person is bitten, thrombocytopenia occurs, with a platelet count of less than 10,000/mm^3, a decrease in fibrinogen, and an increase in fibrin degradation products. The venom proteins may induce distention of the vascular basement membrane and capillary matrix.[58] Prothrombin time and partial thromboplastin time increase with severe envenomation. Usually, these increases occur because consumption of coagulation factors results in clinical anticoagulation. In dog experiments, activation of fibrinolysis may be preceded by thrombus formation, with clotting of critical vessels in the coronary vasculature, which can lead to cardiac arrest and death. This may also explain pulmonary emboli, as the thrombus formation may occur in the legs and cause deep vein thrombosis.[59] Drops in hematocrit may also occur, along with so-called burring of erythrocytes.[54] Approximately 53% of patients experience coagulopathy 2 to 14 days after envenomation. In one study, 76% of patients with pit viper envenomations developed coagulopathy during their hospital course.[60] The coagulopathy may last up to 26 days.[59]

Clinical Features

From 30% to 50% of snakebites do not result in envenomation. The snake can control the amount of venom injected and may inject up to 90% of its venom to immobilize its prey. Other factors involved in the injection of venom include the health of the snake; its satiety; the condition of the fangs; the toxicity of the venom; whether the snake is injured; and the size, age, and health of the victim.

Minor pit viper envenomation causes local pain and swelling (edema with a diameter of approximately 1 to 5 in.), without systemic symptoms or signs. Moderate envenomation is characterized by greater edema (diameter of 6 to 12 in.), weakness, sweating, nausea, fainting, dizziness, ecchymoses, and tender adenopathy.[51] As the envenomation becomes more severe, the symptoms increase to include tachycardia, tachypnea, hypothermia, hypotension, ecchymoses, paresthesias, fasciculations, gingival bleeding, hematemesis, hematuria, melena, oliguria, epistaxis, intracerebral hemorrhage, or coma.[58] Fasciculations are a characteristic manifestation of bites from the eastern diamondback rattlesnake.[61] The skin around and over the snakebite will develop a tense, discolored bulla with serous or hemorrhagic fluid. Death usually results from hemorrhage, increased vascular permeability, and thromboembolic events secondary to disruption of the coagulation pathways.

Coral snake envenomation is painful and has the appearance of scratch marks with no surrounding edema. Systemic symptoms are delayed by about 1 to 6 hours. They begin with paresthesias around the wound margins, followed by weakness, apprehension, giddiness, nausea, vomiting, and a sense of euphoria. Excess salivation is nearly always present.[61] Bulbar and cranial nerve paralysis and ptosis may develop. Ptosis is very common and is often the first sign of coral snake envenomation. Diplopia, papillary dilatation, salivation, dysphagia, and respiratory failure may occur. The paralysis may last up to 14 hours, and full strength may not return for 6 to 8 weeks. In fatal cases, the usual cause of death is respiratory failure.

Viperid venom may increase vascular permeability, leading to bleeding into the gastrointestinal or genitourinary tract. In addition to the obvious signs and symptoms, renal failure may oc-

cur secondary to hemorrhage, coagulopathy, or secondary shock. Intracranial hemorrhage, especially into the anterior pituitary gland (leading to Sheehan syndrome) has been seen in envenomation by the Russell viper. Nephrotic syndrome, glomerulonephritis, hemolytic-uremic syndrome, and DIC have an incidence of 1.4% to 28%, especially in envenomations by the Russell viper, puff adder, and sea snake.[59,62]

TREATMENT

First Aid

Treatment in the field should focus on preventing systemic absorption of the toxin. This may be done with compressive dressings and immobilization of the bitten extremity. Stabilization may be accomplished via an inflatable splint. Nothing should be given by mouth.

If signs of envenomation begin to occur, a constriction band to impede lymphatic flow should be placed on the extremity, proximal to the bite.[63] The Commonwealth Serum Laboratory technique (Australia) uses an elastic band or air splint for wrapping the extremity. The Monash method uses a thick pad and tight bandage over the wound site to impede flow of the venom. Both of these methods have proven efficacy only with Elapidae bites. Transport to a hospital should take place immediately, because the absorption of neurotoxic venoms may result in respiratory compromise or arrest. In patients with a facial envenomation, edema may cause airway obstruction, so emergency response personnel may have to establish immediate airway control.[64]

The site should be wiped off and cleaned. However, the old practice of incising the bite site and applying suction to remove the snake venom should not be used. This practice, which dates from the 1920s, was tested in animal models and found not to increase survival. In fact, incision and suction at the wound site poses more hazard than benefit. The incision may aggravate bleeding, damage nerves and tendons, introduce infection, and delay healing. Cryotherapy (e.g., placing ice on the bite site), which was once thought to lower venom enzyme activity and absorption into the systemic circulation, has also been shown to provide no significant benefit; rather, it causes tissue loss, cold injury, and possible permanent disability.

Extraction therapy has also fallen out of favor. In this procedure, a suction device is placed over the fang wounds, and suction is applied to remove the venom from the bite site and the surrounding tissue without an incision. Prehospital personnel who find a suction device already in place when they arrive at the scene should remove the device, provided there is no fluid accumulating in the cup.[64] The use of field first-aid methods such as incision and suction, tourniquets, and cryotherapy has been associated with a threefold increase in the likelihood of the need for surgical intervention.[50]

Although popular belief has it that snakebites kill within minutes, in fact, the toxicity from snake venom usually does not even begin to affect the body for several hours. In one review, 64% of deaths from snakebite occurred between 6 and 48 hours after the patient was bitten.[65]

Emergency Department Management

History When a snakebite victim arrives at the hospital, the history of the bite should be obtained. This should include (1) a description of the snake, (2) the time elapsed since the bite, (3) the circumstances surrounding the bite, (4) the number of bites, (5) the location of the bite, (6) the type of first aid administered, and (7) any symptoms that have occurred since the bite. The patient's past medical history and allergy history should be reviewed briefly. In particular, the clinician should ask whether the patient has ever experienced allergic symptoms around horses or on exposure to horse serum and whether the patient has asthma, hay fever, or urticaria.

Physical examination Special attention should be paid to the area around the snakebite. The wound should be examined for fang marks, edema, petechiae, ecchymoses, and bullae. Thorough neurologic and cardiovascular examinations are indicated. If the patient was bitten on an extremity, circumferential measurements of the extremity should be taken at the site of injury and 5 in. proximal to the site. Distal pulses and neurologic status should be assessed and recorded, because edema from snakebites may result in elevated compartment pressures, leading to compartment syndrome.[50,64] The patient should be monitored in an intensive care setting.

Laboratory tests All patients should have baseline laboratory studies performed, including a complete blood count, urinalysis, electrocardiogram, prothrombin time, partial thromboplastin time, fibrinogen levels, fibrin split products, serum electrolytes, blood urea nitrogen, and serum creatinine. Blood should be typed and screened. In severe envenomations, arterial blood gas determinations also are indicated. In patients with extremity edema, arterial Doppler evaluation and, in some cases, compartmental pressure determinations may be necessary.

Antivenin Therapy

Antivenins are available for bites of North American pit vipers and eastern coral snakes. Water moccasin and copperhead bites are typically managed without the use of antivenin. The choice whether to use antivenin is based on many factors, including clinical signs and symptoms of envenomation and the physiologic status of the victim. Antivenin is indicated only for severe envenomations.[66]

Antivenin can be obtained through hospital pharmacies, veterinarians, local zoos, and poison control centers. Antivenin is most therapeutic when given within 4 hours after the bite. It is of limited value when given after 12 hours.[52]

Classification of envenomation Envenomations are classified according to a five-level system. The amount of antivenin given correlates with the grade of envenomation.

In grade 0 envenomations, the patient may have fang marks or superficial abrasions of the skin at the bite site but has minimal local edema or pain and no associated systemic manifestations.

Grade 1 envenomations involve some pain or throbbing at the bite site, with 1 to 5 in. of edema and erythema surrounding it. There are no systemic manifestations.

Grade 2 envenomations produce more severe pain over a larger area. The edema spreads toward the trunk, and petechiae and ecchymosis are present in the edematous area. There may be systemic involvement consisting of nausea, vomiting, and temperature elevation.

Grade 3 envenomation is considered severe. Edema spreads up the extremity and may move to the trunk. There may be generalized ecchymosis and petechiae. The patient may have a rapid pulse, hypotension, and hypothermia and may go into shock.

Grade 4 envenomation is very severe and usually results from the bite of a large snake or from a very large venom load. Edema,

petechiae, ecchymosis, and necrosis rapidly overtake the extremity and a large portion of the trunk. Muscle fasciculations, sweating, nausea, vomiting, cramping, pallor, weak pulse, incontinence, convulsions, and coma may all occur.

Antivenins Multiple types of antivenin are on the market. The first marketed antivenin (Antivenin [Crotalidae] polyvalent [ACP]) was a horse-serum–based, whole antibody preparation. The dosage for that preparation was three to five ampules of antivenin diluted in 500 ml of intravenous fluid. Up to 54% of patients treated in studies were allergic to the ACP antivenin. Rash, hypotension, wheezing, and phlebitis occurred in 20% of patients.[67] Nevertheless, clinicians would frequently forgo skin testing for allergy to ACP because it delayed administration of the antivenin.

Although the ACP antivenin is still produced and is used in some areas, a polyvalent crotalid (ovine) Fab antivenin (CroFab) has been introduced. This antivenin minimizes the risk of immediate hypersensitivity and prevents delayed serum sickness. It is based on sheep serum and is four to five times more potent than ACP.[60] CroFab is made by immunizing sheep with crotaline snake venom and digesting the immune serum with papain to produce antibody fragments (Fab and Fc); the antigenic Fc segment is removed during purification.[66] In a study of 1,000 treated patients, none showed evidence of true anaphylaxis.[68] Each vial of CroFab contains 750 mg of Fab and is reconstituted in 10 ml of normal saline; four vials are diluted in 250 ml of normal saline. Studies have shown improvement at the 4-hour mark in all patients given this regimen, although some patients subsequently worsened.[68] The half-life of Fab antivenin is less than 12 hours, compared with 61 to 194 hours for ACP,[69,70] so repeat dosing of Fab may be needed to maintain therapeutic serum levels.[71] In studies, only 16% of patients experienced serum sickness after administration of Fab antivenin, and the severity of the serum sickness was classified as only mild to moderate; in one study, the only reaction in 64 of 65 patients receiving antivenin was simple urticaria.[72] The Fab antivenin is given in interval doses, with the first dose given to achieve initial control (defined as cessation of all symptoms—local, systemic, and coagulopathy) and subsequent doses given 6, 12, and 18 hours after the first dose. It is presumed that in some cases, coagulopathy may recur after initial neutralization of the venom. Recurrence may result from a depot of unneutralized venom at the bite site that is released into the circulation after the venom-antivenin complexes are cleared. A combination of edema, circulatory injury, and a lesser amount of subcutaneous tissue at the site of the bite may inhibit the antivenin from reaching the venom depot.[60,73] Alternatively, uncleared complexes may dissociate, leaving free venom to recirculate.[60,68]

Adjunctive Therapy

A number of adjunctive therapies have been proposed for snakebite envenomations. Excision of tissue around the snakebite to remove the depot of venom was proposed at one time, but this approach is no longer used. The strategy of excising only necrotic-appearing tissue has likewise proved inadvisable, because histologic examination of the excised tissue revealed live muscle fibers interwoven with the macroscopically necrotic tissue. Aggressive debridement and antibiotic therapy may be indicated in the event of complications from infection. This may be seen with necrotizing fasciitis from either *Vibrio vulnificans* or *Aeromonas hydrophila*.[74]

Extremity edema from a snakebite may mimic compartment syndrome, but true compartment syndrome is rare in such cases. Most often, the subcutaneous tissue rather than the deep compartmental space is involved. When a deep envenomation occurs and a true compartment syndrome does develop, first-line treatment is antivenin administration, which diminishes the compartmental pressure and swelling. Compartmental pressures greater than 30 mm Hg may indicate a need for fasciotomy, but fasciotomy and debridement should be avoided if possible because this procedure is associated with worse functional results. Fasciotomy is recommended only if a patient's fingertip was bitten and has swelled, with loss of neurovascular or functional activity. Such patients are candidates for so-called digit dermotomy. The incision should be made on the lateral or medial aspect of the finger, through the skin only, and should extend from the web to the middle of the distal phalanx.[61,75]

General Management

A regional poison control center or the local zoo should be used as a resource when dealing with a venomous snakebite. This is especially true if the snake is not believed to be native to the area, as might occur with hobbyists who keep exotic snakes as pets. For cases in which an expert is not available, the Department of Surgery at the University of California, San Diego, School of Medicine has established a Web site that lists protocols (including antivenin availability) for management of snakebites from venomous species around the world. This information is available online at http://www.surgery.ucsd.edu/ENT/DAVIDSON/snake/index.htm.

Other therapeutic measures are keyed to specific symptoms. Isotonic fluid replacement should be given if the patient is hypotensive. Abnormalities of the clotting mechanism should be corrected with blood product replacement as necessary, but this should be done only after antivenin therapy has been started. In fact, common treatments for standard coagulopathies are ineffective or dangerous for snakebite-induced coagulopathies. Instead, the effects of the venom should be treated (with antivenin) before usual coagulopathy treatment is initiated.[59] Corticosteroids are contraindicated during the acute stages of envenomation, but they may be used if the patient experiences serum sickness from antivenin use. Studies in Costa Rica have shown that in viper bites, the release of inflammatory cytokines leads to clinical and pathologic alterations similar to those found in trauma patients.[76] Further research on this reaction and on the potential use of steroids is indicated. Patients should be placed on oxygen and should be given mechanical support if necessary for signs of trismus, laryngeal spasm, or excessive salivation. Tetanus therapy should be given if indicated [*see Table 3*]. Antibiotics are recommended only if signs of infection are present. In one study, there were no wound infections in patients with nonenvenomated snakebite.[77]

The wounds should be examined daily. Superficial necrosis and hemorrhagic blebs should be debrided at days 3 through 10. Debridement may need to be done in stages.

SPECIAL CONSIDERATIONS

Snakebite in Pregnancy

In pregnant women who have been bitten by snakes, what is best for the mother will usually be best for the fetus. Fetal outcome may depend on the gestational age of the fetus, with younger age associated with a negative outcome. The miscarriage rate after snake envenomation may be as high as 43%.[78] Miscarriage may result from shock, uterine contractions, pyrex-

ia, or placental or uterine bleeding. Venom may cross the placental barrier and cause some systemic poisoning of the fetus, even if the mother remains symptom free.[78]

In pregnant snakebite victims, airway compromise and shock states should be corrected to ensure perfusion of the placenta and uterus and thereby prevent fetal hypoxia. Circulatory support with vasopressors should be avoided because they reduce uterine blood flow and are detrimental to the fetus. Pregnant women are already in a hypercoagulable state and therefore are even more susceptible to DIC.[78] Abruptio placentae from hypercoagulability has occurred after snakebites.

Antivenin is the therapy of choice in pregnant patients, but there is no reliable information regarding the risks of administration of antivenin during pregnancy. Serum sickness and anaphylactic reactions remain the highest risks associated with antivenin administration.[79]

Snakebite in Children

Snakebites in children are often on the lower extremities or—if the child was handling the snake—on the hands. Signs and symptoms are typically similar to those in adults.[80] Because of their smaller blood volume, however, children may experience a more severe envenomation syndrome.[81] In a study of 67 children with severe envenomation, 72% had systemic involvement and 50% developed coagulopathy; 61% received antivenin, and 36% of those treated experienced adverse reactions.[82] In another case study of 12 children with rattlesnake envenomation who were treated with with CroFab, no evidence of acute or delayed hypersensitivity was noted.[83] Recurrence of local swelling was seen in one patient despite repeated treatment with antivenin. In children, dosing of antivenin should be based on venom load and severity of signs and symptoms rather than on patient age or weight.[82] Although envenomation can lead to multiorgan failure in children, most of the symptoms are localized and limited to erythema, edema, and blisters at the snakebite site.[83] Although most patients are admitted to an intensive care unit, only hemodynamically unstable patients must be admitted to such a setting. Mild and moderate symptoms can be treated in the emergency department or the floor setting.[83]

Disposition

All victims of suspected snakebites should be observed for a minimum of 4 to 6 hours. If there is no sign of envenomation after 6 hours and it is believed that the snake was either nonvenomous or a pit viper, the patient can be discharged. The patient should be given instructions to return to the emergency department immediately if any symptoms of envenomation occur. A patient who has minimal edema and pain should be observed for at least 12 hours. If the swelling has begun to diminish and the pain has resolved, the patient may be discharged with the same discharge instructions.

Pit viper envenomation may result in significant hypofibrinogenemia and thrombocytopenia lasting up to 2 weeks, which may lead to complications from surgery or trauma.[84] If coagulation abnormalities have resolved, however, no further workup of the coagulation system is needed. The best predictor of late hypofibrinogenemia is early hypofibrinogenemia.[60]

Any patient who has been bitten by a Mojave rattlesnake, coral snake, or other exotic snake should be admitted to the intensive care unit, with cardiorespiratory and dialysis equipment readily available. Antivenin should be administered to such patients.[52]

SNAKEBITE PREVENTION

Snakes should not be handled except by a professional herpetologist. Even a dead snake can envenomate its handler. Persons spending time outdoors in areas known to be heavily populated with snakes should wear long pants and closed shoes. Because of the varying size of snake fangs, loose pants are preferred to tight-fitting trousers. Heavy leather boots are recommended rather than sandals, open-toed shoes, or sneakers. Persons who are going to handle snakes, even dead ones, should also wear protective gloves.

Spider Bites

The class Arachnida contains the largest number of known venomous species, including 20,000 venomous spiders. In the United States, only 50 species can envenomate humans, partly because most spiders' fangs are not long enough to penetrate human skin. All spiders are carnivorous; they capture their prey either by hunting or trapping it. Trappers spin webs and wait for prey to become ensnared. They have limited vision and sense prey on their web with their jaw, which can detect movement. Hunters have better eyesight than trappers, and most hunters eat their prey on the spot. For a true diagnosis of a spider bite with possible envenomation, there should be evidence of a bite with pain or discomfort (except with bites from *Loxosceles* species, which are considered painless), collection of the spider at the time of the bite, and identification of the spider by an expert arachnologist. Other signs should be the appearance of fang marks, redness at the site of injury, immediate or delayed itching, and the presence of spines and swelling (which may not be seen in all cases).[85]

BLACK WIDOW SPIDERS

The black widow spider, *Latrodectus mactans* [*see Figure 4*], is found throughout the United States and southern Canada, with other closely related species found mainly in the western United States. There have been no recorded findings of the spider in Alaska. The female is twice the size of the male and is the only sex that is able to envenomate. The male does have venom, but because of its smaller size and less powerful fangs, it is unable to bite and envenomate a human. The female is glossy black, with a bright-red marking on the abdomen that may appear to be two spots or have an hourglass shape. Occasionally, it has red stripes. Immature females are red, brown, and cream colored. The spider's body is about $^1/_2$ in. long; with the legs, it measures approximately $1^1/_2$ in. in length. It is usually found under rocks and in woodpiles, outhouses, and stables, and it is not aggressive unless guarding its eggs. The black widow is a trapper. The webs are close to the ground to have access to crawling insects and are usually in secluded, dim areas.

Envenomation

Toxicology Black widow venom contains various proteinaceous compounds. The venom paralyzes the prey and begins the digestive process by liquefying the victim's tissues. The toxin depletes acetylcholine from the presynaptic nerve terminals, thereby destabilizing nerve cell membranes and opening ionic channels. The toxin of the black widow is thought to cause a massive release of acetylcholine and then block its reuptake, which leads to both sympathetic and parasympathetic stimulation.[86] There is a patchy paralysis of skeletal muscles with various changes in the autonomic nervous system. This paralytic syndrome has been likened to polio.[87]

Figure 4 **The mature female black widow spider is glossy black, with a bright-red marking on the abdomen that may appear to be two spots or have an hourglass shape.**

Clinical features On being bitten by a black widow spider, a person may feel a pinprick sensation, with minimal local swelling and erythema. Two small fang marks may be visible. The middle of the bite site may be white, with surrounding erythema and a reddish-blue border. Within an hour, the patient may feel a dull ache or crampy pain in the area of envenomation. This feeling may spread throughout the body shortly thereafter. The pain will increase over the first hour in approximately one half of the cases.[85] The pain spreads to the chest from upper extremity bites and to the abdomen from lower extremity bites. This pain may mimic pancreatitis, appendicitis, or a peptic ulcer. The abdomen may have boardlike rigidity but will not necessarily be painful to palpation. In addition, there may be other myopathic signs such as facial trismus, muscle fibrillation, tonic contractions, or so-called facies latrodectismica, a constellation of symptoms that includes blepharoconjunctivitis, flushing, and contortions.[88] Other systemic symptoms may include dizziness, nausea, vomiting, headache, itching, conjunctivitis, diaphoresis, piloerection, priapism, anxiety, and dyspnea. There is wide variability in the percentage of each symptom experienced by people bitten by the spider.[85] The symptoms may last for 2 to 3 days, abating slightly after a few hours. Patients with preexisting hypertension, cerebrovascular disease, or cardiovascular disease are at risk for a worsening of those conditions. Rare complications of the bites also include compartment syndrome, rhabdomyolysis, and obstruction of the venous outflow in the affected extremity.[89]

Treatment

First aid Patients bitten by a black widow spider should have cool compresses applied to the bite and be transported to a hospital. Laboratory experiments have shown that a Sawyer extraction device may be helpful in removing venom if applied to the skin within 3 minutes after the bite.[88] Basic or advanced life support should be given on the way to the hospital. If possible, the spider should be brought along with the patient, because many nonvenomous species may resemble the black widow.

Emergency department management A complete history and physical examination should be done, with attention paid to the circumstances surrounding the bite, a description of the spider, and allergies to other bites or to horse serum. The site should be inspected and cleansed. Tetanus immunization should be updated, if necessary. Laboratory studies should be obtained, including a complete blood count, serum electrolytes, clotting studies, urinalysis, electrocardiogram, blood urea nitrogen, and serum creatinine.

Patients with envenomations should be treated symptomatically. Nitroprusside may be used for hypertensive episodes related to the envenomation. Abdominal cramps may be alleviated with calcium gluconate (10 ml of a 10% solution given intravenously over 20 minutes). Serum calcium levels should be followed. Diazepam may be given to alleviate muscle spasms.[90] Dantrolene has also been shown to provide muscle relaxation in these patients.[91]

Antivenin therapy *Latrodectus* antivenin exists and is indicated for patients who have hypertensive heart disease, underlying respiratory disease, or severe envenomation. Patients between 16 and 65 years of age may receive the antivenin. The antivenin is derived from horse serum and may cause allergic reactions. The intravenous route is preferred. The contents of the vial are diluted in 50 ml of normal saline and given over 15 minutes. The typical dose is one to two vials. Widow antivenin is available in Australia, South Africa, and the United States, which are the primary locations of the widow spiders. Intramuscular administration in Australian studies has shown few adverse reactions. However, American black widow antivenin has a higher incidence of early allergic reactions.[92]

Special Considerations

Spider bites in children Children who are bitten by a black widow spider may have severe symptoms that may lead to death. Possibly because of the smaller volume in which the venom is circulated, a dose that would be tolerable in an adult may cause fatal cardiovascular or respiratory decompensation in a child. A retrospective Australian study of red-back spider bites showed systemic symptoms of diaphoresis, irritability, and hypertension in 85% of pediatric patients admitted to the hospital.[93]

Spider bites in pregnancy A pregnant woman who is bitten by a black widow spider will have signs and symptoms that are otherwise typical, but she may not have a rigid abdomen because of the stretching and laxity of the gravid abdominal wall. However, the cramping may be severe enough to induce miscarriage. The toxin does not seem to have a direct effect on the fetus, possibly because it is not able to cross the placental barrier, nor is it able to cross the blood-brain barrier. Antivenin is indicated for the symptomatic pregnant patient.[78]

Disposition

Patients with signs or symptoms of black widow spider envenomation should be admitted to the hospital. The patient should be observed for a minimum of 2 hours and, if totally asymptomatic, may be discharged with instructions to return if any symptoms develop.

BROWN RECLUSE SPIDERS

The brown recluse spider, *Loxosceles reclusa* [*see Figure 5*], is approximately 1 in. long and ranges in color from tan to dark brown. The female has a violin-shaped dark-brown spot on its abdomen. Males and young spiders may also have a darkened violin pattern on the thorax.[94] The brown recluse has three sets of eyes arranged in pairs called dyads.

The brown recluse is frequently found in clothing, bedsheets, and blankets in a closet; in woodpiles in a shed; or under rocks. The brown recluse hunts for its food but can live for 6 months without food or water.[95] It forages at night but is not aggressive unless threatened, hence the "recluse" name.[95]

The brown recluse is found in the south central United States, especially in Missouri, Kansas, Arkansas, Louisiana, east Texas, and Oklahoma. It is occasionally found in other states, but these cases likely represent spiders uprooted and transported from the endemic areas.[94] The diagnosis of true *Loxosceles* spider bites in the United States is considered to be far less than the number of patients who are treated for recluse spider envenomations.[95] Strict inclusion criteria for recluse envenomation include sighting of the biting spider and identification of that specific spider. In studies from areas endemic for *Loxosceles,* patients present with almost no true *Loxosceles* bites or envenomations despite exposure to dozens to hundreds of spiders; these results suggest that nonendemic areas should have minimal to no true *Loxosceles* bites.[85,96]

Loxosceles spiders are also found outside of North America, most notably in Brazil and elsewhere in South America. *L. laeta* is the most toxic of these species.[97]

Envenomation

Toxicology Brown recluse venom is protein based, and it has antigenic and locally destructive properties. Esterases, hyaluronidases, and proteases have been isolated from recluse spider venom.[95] The components of the venom are cytotoxic to endothelial cells and red blood cells. Sphingomyelinase-D acts directly on red blood cells to cause lysis.[98] Unlike black widow venom, brown recluse venom has no known neurotoxic effects.

Clinical features Brown recluse spider venom produces both localized and systemic symptoms. These are referred to as cutaneous and viscerocutaneous symptoms.[85] Cutaneous loxoscelism presents initially as pain in the area of the bite, or pain may be absent at first and then develop over the next 3 to 4 hours. Soon thereafter, a ring of pallor from vasoconstriction appears around the bite, with a surrounding area of erythema. A bleb develops in the center and, after a few days, becomes necrotic. The bleb may spread with gravity-dependent flow. This necrotic area spreads over the next few days, involving both superficial and deep tissues. An eschar usually forms days later, and the wound may not heal for months after the eschar separates.[99,100] Subcutaneous fat may liquefy below the eschar, leaving a depressed scar. This occurs in areas with more subcutaneous tissue, such as the thigh. Healing of the eschar takes from 5 days to 17 weeks.[101] Viscerocutaneous loxoscelism may present as systemic effects that include fever, chills, rash, nausea, vomiting, shock, renal failure, hemorrhage, DIC, or pulmonary edema. There have been case reports of transverse myelitis and paralysis from the brown recluse spider bite; these are believed to result from microthrombosis at the anterior vertebral artery.[102] Also, loss of cutaneous sensation may be caused by damage to or destruction of a nerve or its branch by the venom itself or ischemia from the edema.[98]

The differential diagnosis of the skin changes caused by brown recluse spider envenomation includes Stevens-Johnson syndrome, toxic epidermal necrolysis, erythema nodosum, purpura fulminans, diabetic ulcer, allergic dermatitis, Lyme disease, and pyoderma gangrenosum.[88] Patients who have had numerous brown recluse spider bites have demonstrated an antibody response and may have a decreased response to subsequent bites.[95]

Figure 5 **The brown recluse spider (7 to 12 mm body length) is a timid arachnid that may be encountered in basements, closets, and woodpiles.**

In the few cases of fatal bites, death usually results from DIC and renal failure. In the United States, however, brown recluse spider envenomations frequently result in little more than an inflammatory reaction.[94]

Spider bites in children Brown recluse spider bites in children may result in severe systemic reactions, among which are hemolytic anemia, hypotension, and anemia.[103] It is thought that sphingomyelinase-D in the venom leads to hemolysis, platelet aggregation, thrombosis, and vasoconstriction in children. The severity of any reaction is a function of the amount of venom injected and the location of the bite. Swelling of the neck and subsequent airway obstruction has been reported in a child.[104] Death is a rare but known complication.

Treatment

Prehospital care should include basic or advanced life support, cool compresses, and immobilization of the extremity. On the patient's arrival at the emergency department, a complete history and physical examination should be done, with attention paid to the circumstances surrounding the bite, the description of the spider, and allergies to other bites or to horse serum. Early diagnosis is most easily accomplished if the patient brings the spider to the emergency department. The site should be inspected and cleansed. Tetanus toxoid should be administered, if appropriate.

Brown recluse antivenin is available and has been shown in some studies to be successful in limiting necrosis, but it must be given within the first 24 hours of the bite. During this time, however, it is difficult to assess the future severity of the cutaneous lesion.[105] A 1983 study found some evidence that systemic steroids should be used with brown recluse bites if the patient is seen within 24 hours.[106] Other studies have shown that the intralesional injection of steroids may help control inflammation or thrombosis but does not alter eschar size or outcome[107,108] Dapsone has also been shown to be helpful in treating the local effects of the venom. However, it should be used only in adults who have been screened for glucose-6-phosphate dehydrogenase (G6PD) deficiency. Dapsone is most effective in the first few hours after a

bite, but its side effects usually outweigh its benefits for prophylactic use.[105] The complications of dapsone use include hemolysis, agranulocytosis, aplastic anemia, methemoglobinemia, rashes, toxic epidermal necrolysis, and fatal reactions.[106] Phentolamine has not been shown to have any appreciable effect on the necrotic activity.[95] In animal models, hyperbaric oxygen therapy has not been conclusively proved to be effective, but studies have shown some positive outcomes.[109] In one study, electric-shock treatment of the envenomation site in a human showed a potentially positive outcome to the lesion; however, no benefit was shown in animal models, and delayed healing was actually found to occur in some cases. Therefore, electric-shock therapy has not been supported by the literature.[94,110] Excision has not been shown to improve outcome and, in fact, may be detrimental. Antibiotics may be prescribed if signs of infection are present. Analgesia should also be offered to the patient.[98] Patients who have signs or symptoms of envenomation should be admitted to the hospital. If acute renal failure develops, dialysis may be necessary. Tissue necrosis at the wound site may necessitate surgical intervention, but only if the lesion is large, has persisted for 6 to 8 weeks, and has stopped progressing in size.

Disposition

A patient who has been bitten by a brown recluse spider should be observed in the emergency department for a minimum of 6 hours. If no local or systemic symptoms develop, the patient may be discharged with instructions to return if any symptoms appear.

TARANTULAS

At least 30 species of tarantula (*Theraphosidae*) live in the deserts of the western United States. Tarantulas are hunters that eat nocturnal insects. Because of the location of its fangs, a tarantula must raise itself on its hind legs to inflict a bite. In addition, when it is handled, a tarantula releases the hairs of its abdomen, which cause a local urticarial reaction in humans.

Bites from the tarantula are relatively innocuous and result in a low-grade histamine reaction. However, they should be cleansed, and tetanus immunization should be updated if necessary.

Bites from tarantulas from the Panama Canal Zone may cause paresthesias and local discomfort. The South American tarantula has a more toxic bite, for which antivenin is available.

Treatment of all tarantula bites should be supportive, with the administration of antihistamines and oral analgesics. The tarantula hairs, which may be barbed, can be removed with adhesive tape.

Scorpion Stings

EPIDEMIOLOGY

Scorpion stings are common in tropical and subtropical regions of the world. For example, Tunisia reports almost 40,000 stings a year, which result in approximately 1,000 hospitalizations and 100 deaths.[111] Deaths from severe envenomation are the result of cardiogenic shock and pulmonary edema.[112]

Scorpion envenomation is not rare in the United States, with 14,569 consultations for scorpion stings reported in 2001 by the American Association of Poison Control Centers.[113] Most scorpion stings occur in the southwestern states. The scorpion responsible for severe envenomations in the United States is *Centruroides exilicauda*.[114]

Table 5 Grading System for Severity of *Centruroides exilicauda* Envenomation[117]

Grade	*Features*
I	Pain or paresthesia at the site of envenomation
II	Local findings plus pain or paresthesia remote from the sting site
III	Cranial nerve dysfunction* or somatic-skeletal neuromuscular dysfunction†
IV	Cranial nerve dysfunction* and somatic-skeletal neuromuscular dysfunction†

*Blurred vision, wandering eye movements, hypersalivation, trouble swallowing, tongue fasciculation, problems with upper airway, slurred speech.
†Restlessness, severe involuntary shaking, and jerking of extremities (may be mistaken for seizures).

ENVENOMATION

Toxicology

The toxin of *C. exilicauda* is a heat-stable neurotoxin that increases permeability of neuronal sodium channels, causing depolarization of the nerve and myocyte.[115] Severe envenomation results in stimulation of both cholinergic and adrenergic neurons by its action on presynaptic cell membranes.[116] Increased permeability of neuronal sodium channels in the autonomic nervous system results in tachycardia, agitation, hypertension, hypersalivation, dysphagia, and gastrointestinal symptoms.[117,118]

Clinical Features

Scorpion envenomation can produce effects ranging from local to life threatening. Envenomations are categorized by severity from grade I (pain or paresthesia at the sting site) to grade IV (combined cranial nerve and somatic-skeletal neuromuscular dysfunction) [*see Table 5*].[116]

Treatment

Most adults stung by *C. exilicauda* experience only local pain and paresthesia and can be managed as outpatients.[119] Young children, however, often present with severe involuntary motor activity, agitation, and respiratory symptoms requiring intensive supportive care.[119,120]

Initial hypertension and tachycardia may occur in close to half the patients with milder envenomation.[121] These patients normally respond well to treatment with an antihypertensive agent such as prazosin[121] or captopril.[122]

Patients with severe envenomation may have hypotension, left ventricular failure, and pulmonary edema.[122] Supportive care and afterload reduction with vasodilators appear to reduce mortality.[121,122]

The use of specific antivenin in scorpion stings is controversial.[123-127] Antivenin specific to *C. exilicauda* is not available except in Arizona,[115] where it is commonly used for severe grade III and almost all grade IV envenomations.[128] It is rarely associated with anaphylaxis but commonly results in mild serum sickness.[115,128] A majority of opinions recommend the treatment of scorpion stings with antivenin, but this recommendation depends on the dose of the antivenin and its route of administration. Intravenous administration is recommended because of the slow absorption of the intramuscular route compared with the rapid distribution and absorption of scorpion venom. In studies using the intravenous administration of antivenin, overall mortality

was reduced to less than 0.05%.[129] However, some of the symptoms, such as pulmonary edema, labile vital signs, and dysrhythmias, were still seen in envenomated patients.

The use of sedative-hypnotics has been advocated for the severe agitation and motor restlessness associated with *C. exilicauda* envenomation. Careful assessment (to ensure adequate airway, breathing, and circulation) and monitoring of patients treated with sedative-hypnotics is essential because respiratory depression and even respiratory arrest have been reported with this therapy.[115,120]

Insect Bites and Stings

Insect bites are medically important primarily because insects can act as vectors for pathogenic microorganisms by directly inoculating their human hosts while feeding on blood or tissue fluids. In addition, stinging insects have venom, and the exoskeleton, hair, and secretions may act as irritants or allergens. Allergic reactions to Hymenoptera venom are discussed elsewhere [*see 6:XV Allergic Reactions to Hymenoptera*].

Although insect bites and stings are an essentially universal human experience, the exact incidence of serious morbidity and mortality is difficult to determine because of different reporting practices, regional differences of endemic species, and the wide spectrum of clinical effects, particularly severity. In 2003, the American Association of Poison Centers reported 94,247 bites and envenomations (representing 3.5% of all toxic exposures), with six deaths (0.25% fatal exposure cases).[130] These statistics include bites and stings not only from insects but also from marine species and mammals.

FIRE ANTS

Two species of fire ants were imported into Alabama in the early part of the 1900s. Since that time, they have spread throughout the southeastern United States and Texas.[131] Fire ants both bite and sting, anchoring themselves with their mandibles to leverage the thrust of their stinger. Their venom is primarily composed of an insoluble alkaloid that has local hemolytic and necrotic effects.[132] Fire ants have a propensity to swarm, resulting in multiple bites. Initially, stings cause an erythematous papule or wheal that develops into a pruritic sterile pustule over the course of 6 to 24 hours and may persist for weeks [*see 6:XV Allergic Reactions to Hymenoptera*]. Because of the alkaloid nature of the venom, reactions typically remain local; systemic or anaphylactic reactions are rare. Occasionally, secondary infection occurs.[133]

Treatment consists of local wound care, cold packs, antihistamines, and topical steroids. Extensive involvement may necessitate oral steroids. Anaphylactic reactions are managed in the same manner as those from other causes. Secondary infection requires treatment with antibiotics. Desensitization therapy should be considered for patients who experience life-threatening reactions.[134]

KISSING BUGS AND BEDBUGS

The kissing bug (*Triatoma* species), also known as the assassin, cone-nosed, or reduviid bug, is found mostly in the southern and western regions of the United States. It feeds on the blood of vertebrates, including humans, mostly at night. Kissing bugs possess a long proboscis from which they suck blood from the victims without pain. Bites commonly occur on the face because that area is usually exposed during the night.[132] In Central and South America, kissing bugs are the vector for *Trypanosoma cruzi,* the causative agent of Chagas disease.[135]

Bedbugs (*Cimex* species) are also nocturnal bloodsucking insects that have adapted to human environments. They are found throughout the United States and live in baseboards, furniture, clothing, and bedding. Bedbugs are not known to be vectors for human disease, but allergic reactions to their bites can present as multiple clustered, erythematous, pruritic papules that can last over a week.[136] Systemic effects are rarely encountered. Treatment consists of symptomatic care with antihistamines and topical steroids.

CATERPILLARS AND MOTHS

The order *Lepidoptera* includes venomous caterpillars and moths. Some of these caterpillars have hollow spines among their body hairs; injection of venom through these spines can cause symptoms ranging from local dermatitis to generalized systemic reactions.[137] The puss caterpillar, or woolly slug, is found in the southeastern United Sates and Texas and accounts for most of the envenomations from this insect family for which patients seek medical care. Stings produce small, erythematous, painful papules at the site of contact. Fever and muscle cramps may occur, but serious systemic effects are rare.[137]

Gypsy moths infest much of the eastern United States but also are present elsewhere in the country. Skin contact with gypsy moth caterpillars can cause dermatitis from delayed hypersensitivity. Treatment is symptomatic, with topical steroids and oral antihistamines. Analgesics occasionally are needed for pain control.[135]

BLISTER BEETLES

The blister beetle is found throughout the United States.[132] Blister beetles do not have a toxic bite or sting, but they secrete cantharidin, which acts as a vesicating agent, causing skin irritation and blisters several hours after contact.[132] The blisters can range from a few millimeters to several centimeters in diameter. Pulverized blister beetles were used as an aphrodisiac known as Spanish fly, which causes urethral irritation when ingested orally. Local blisters from blister beetle contact are treated as a chemical burn, with diligent wound care and prevention of secondary infection.[135]

TICKS

Ticks are found throughout the world and are members of the class Arachnida. They painlessly attach to their host (mostly mammals) to feed on blood. The primary medical importance of ticks is as vectors for infection. Infectious diseases carried by ticks include Lyme disease, Rocky Mountain spotted fever, babesiosis, ehrlichiosis, tularemia, Colorado tick fever, and relapsing fever.[138]

Some *Ixodid* species of ticks produce a neurotoxin in their salivary gland that can induce a syndrome known as tick paralysis.[139] The toxin is usually transmitted by an engorged, gravid female and causes an ascending flaccid paralysis approximately 2 to 7 days after the tick begins feeding. Children are the ones who are most often affected. Respiratory failure may occur, and ventilatory support is required in some cases. The diagnosis is confirmed by finding an embedded tick on the victim. Removal of the tick is essential to recovery. Symptoms improve several hours to days after the tick is removed.[140]

Ticks generally attach themselves to their hosts after approximately 1 to 2 hours, so persons in tick-infested areas should perform frequent checks of their clothing and body. If a tick does attach, it should be removed promptly to minimize the risk of dis-

ease transmission. The tick should be grasped as close to the skin as possible, using blunt forceps, tweezers, or protected fingers. Steady pressure should be applied while pulling out the tick. After removal, standard wound care should be employed. If the mouth parts are only partially removed, foreign-body reactions, secondary infections, and granuloma formation can occur.

Marine Envenomations

Almost 75% of the earth's surface is covered with water, and approximately 80% of our planet's organisms live in this environment. Only a few of these marine creatures pose a threat to humans, but the dangers have been recognized since ancient times. Significant human morbidity and mortality, ranging from minor dermatitis to life-threatening infections, envenomations, and trauma, may result from exposure to marine life.

Over recent decades, human exposure to the aquatic environment has greatly increased, thanks to scientific and technological advancement and exploration, increased sport diving and recreational activities, increased harvesting of marine resources for food, private and commercial saltwater aquariums, and more travel to and greater accessibility of exotic locations. These factors have increased the risk of exposure to marine organisms. Consequently, it is imperative for clinicians, not only in coastal areas but also inland, to be familiar with the hazards.

Marine organisms that are harmful to humans range from one-celled diatoms and dinoflagellates that cause poisoning by being bioamplified up the food chain (e.g., ciguatera and amnestic shellfish poisoning) to invertebrates with lethal toxins (e.g., jellyfish poisoning) to large vertebrates, such as sharks, that can inflict massive trauma. This discussion reviews some of the more common and clinically relevant envenomations (and their associated injuries and infections) that humans may incur in freshwater and saltwater around the United States.

TOXIC INVERTEBRATES

Coelenterates

The phylum Cnidaria (which includes the former phylum Coelenterata) is divided into three classes: (1) *Hydrozoa* (which includes Portuguese man-of-war, feather hydroids, and fire coral); (2) *Schyphozoa* (true jellyfish, sea nettles, and box jellyfish); and (3) *Anthozoa* (sea anemones, stony corals, and soft corals).[141]

Coelenterates are characterized by venomous stinging organoids called nematocysts. The nematocyst is a fluid-filled capsular structure that encloses a tightly coiled, hollow, sharply pointed tubule that bursts forth into the victim when it is discharged after contact. The venom is a complex mixture of proteins, carbohydrates, and other nonproteinaceous substances.[142]

Clinical effects The clinical features of coelenterate envenomation are fairly constant but have a range of severity from mild dermatitis to rapid cardiovascular collapse resulting in death. Factors that determine severity include the following: species (the Australian box jellyfish is the most deadly of all stinging marine life[143]); season; number of nematocysts triggered; size and age of the victim; location and surface area of the sting; and the sensitivity of the victim to the venom.[141,144] Knowledge of the species of coelenterates indigenous to the geographic location is important in predicting the potential severity of the envenomation. Most hydroids and hydroid corals (which inhabit both temperate and tropical waters off the Atlantic and Pacific coasts of the United States) initially produce a stinging sensation, paresthesias, and pruritus, with local edema, blistering, and wheal formation. Occasionally, the injury can progress over several days to local necrosis, ulceration, and secondary infection.

Physalia physalis (Portuguese man-of-war) consists of a violet-blue floating sail with several nematocyst-bearing tentacles that can be up to 30 m in length.[145] They are widely distributed but are prevalent in the tropical and semitropical waters off the southeastern coast of the United States and in the Gulf of Mexico.

The man-of-war's sting produces an intense pain radiating up the involved extremity, with the development of linear, edematous, erythematous, cutaneous eruptions. Systemic involvement can occur, involving multiple organ systems, with nausea, vomiting, headache, myalgias, respiratory distress, hypotension, anaphylaxis, and cardiovascular collapse.[145,146]

Jellyfish from the class Scyphozoa display a wide variety of colors, shapes, and sizes and vary in toxic potential. They are the most common coelenterates that produce clinical injuries and that cause people to seek medical attention. They have a worldwide distribution and can range in size from a few millimeters to greater than 2 m at the bell with tentacles up to 36 m in length.[147] Because the tentacles are so long in some species, it is possible to undergo a significant envenomation without ever seeing the bell. Organisms that have washed ashore also pose a risk, because undischarged nematocysts can fire if an unwary person steps on or picks up a tentacle or part of it. Envenomations from most jellyfish are of mild to moderate severity, with clinical symptoms similar to those of the hydroids and *Physalia* species.[148]

Treatment Treatment of coelenterate stings includes advanced-life-support measures, symptomatic care, pain control, and prevention of further envenomation by nematocyst inactivation. In the case of box jellyfish envenomations, an antivenin is available (Commonwealth Serum Laboratories, Melbourne, Australia) and should be administered as soon as possible.[149] Anaphylaxis and bronchospasm should be treated with epinephrine, oxygen, intravenous fluids, glucocorticoids, antihistamines, bronchodilators, and vasopressors if needed. The area of the sting should immediately be rinsed with saltwater (not freshwater) or 5% acetic acid (vinegar) for nematocyst inhibition. In the case of a Portuguese man-of-war sting, vinegar should be avoided and only saltwater used. Tentacles should then be removed with a gloved hand, hemostats, or a towel to prevent envenomation of the treating individual.[141] Alternatively, isopropyl alcohol (40% to 70%) may be effective. For *Chrysaora* (sea nettle) or *Cyanea* (lion's mane) stings, a baking-powder slurry applied to the affected area is an effective treatment.[144] In patients with seabather's eruption, a vesicular or morbilliform, pruritic dermatitis caused by larval forms of certain coelenterates [*see 2:VIII Parasitic Infestations*], a papain (meat tenderizer) solution is effective.[150] Topical anesthetics, antihistamines, and corticosteroids may be of benefit. Prophylactic antibiotics are generally not indicated, but appropriate tetanus prophylaxis and proper wound care should be provided.

Echinodermata

The phylum Echinodermata consists of poisonous species of starfish, sea urchins, and sea cucumbers. Toxic sea urchins (mostly found in the Pacific and Indian oceans and the Red Sea) have sharp, brittle, venom-filled spines and may also possess pincerlike seizing organs termed pedicellariae.[147] Spines can eas-

ily penetrate wet suits and skin to lodge in the victim.[151] The immediate reaction consists of intense local pain, erythema, edema, and bleeding. Subcutaneous staining from pigments in the spine can also occur. If multiple spines have penetrated the skin, the patient may develop systemic envenomation symptoms, such as nausea, vomiting, paresthesias, muscular paralysis, hypotension, and respiratory distress. Nontoxic sea urchins can also cause significant morbidity by foreign-body reactions and soft tissue infections or septic arthritis if a spine enters a joint.[145]

Most starfish are nontoxic, but the crown-of-thorns sea star (*Ancanthaster planci*) is covered with thorny spines that deliver a secreted venom produced in specialized glandular tissue.[152] Sea cucumbers are free-living bottom dwellers that produce a mild toxin that is concentrated in the tentacular organs. Direct contact can cause dermatitis. The venom is usually diluted in the seawater, and the greatest risk is exposure of the corneas and conjunctiva, resulting in an intense inflammation.[148] Treatment consists of immersion in hot water (110° to 115° F) for 30 to 90 minutes or until the patient has significant relief of pain. Spines should be localized under direct visualization or by appropriate radiographic technique and carefully removed; retained spines may produce a granulomatous reaction requiring surgical excision. Routine antibiotics are not necessary, but diligent wound care and tetanus prophylaxis should be employed.[147]

Mollusca

Two classes in the phylum Mollusca that have potential toxicities include gastropods (coneshells) and cephalopods (octopuses). Coneshells are univalve animals found in shallow Indo-Pacific waters.[147] They inject potent neurotoxins (conotoxins) into their victims via a detachable, dartlike, radular tooth.[153] Local symptoms resemble a bee or wasp sting, followed by ischemia, cyanosis, and numbness in the area surrounding the wound. In severe envenomations, local paralysis can progress to generalized weakness, paralysis, and respiratory failure.[145] Octopus bites are rare, but deaths have been reported from the Australian blue-ringed octopus. This creature possesses a potent neurotoxin with at least one fraction identical to tetrodotoxin (also found in tissues from the pufferfish), which blocks neuronal sodium conduction.[153] Local symptoms are usually minimal, but significant envenomations can lead to generalized paresthesias, followed by paralysis and respiratory failure.[145] Treatment of octopus envenomation is mostly supportive, with close attention to the need for respiratory support.

Porifera

The phylum Porifera consists of over 5,000 species of sponges. Sponges are stationary acellular animals that attach to the ocean floor or coral beds and may be colonized by other animals.[147] Sponges inhabit waters off both the Atlantic and Pacific coasts of the United States, as well as Hawaii. Sponge diver's disease is usually caused by colonization of the sponges by coelenterates and manifests as a dermatitis or a local necrotic skin reaction.[152] Two primary syndromes can also occur from contact with sponges.[145] The first is a contact dermatitis similar to that caused by plants. The second is an irritant dermatitis caused by the penetration of small spicules of silica or calcium carbonate into the skin. Because it is difficult to distinguish between the two different reactions, treatment of both should be initiated. The skin should be dried, and the spicules should be removed with adhesive tape or a facial peel. Acetic acid 5% soaks should be applied for 10 to 30 minutes three or four times a day (isopropyl alcohol, 40% to 70%, may be substituted). Topical steroids may help with secondary inflammation but should not be used initially.[145]

Annelida

The phylum Annelida includes bristleworms, or fireworms, which are covered by cactuslike bristles and spines. On contact, these bristles easily enter the skin and break off, causing an intense inflammation with a burning sensation and erythema. Bristles should be removed with adhesive tape or a facial peel, with subsequent treatment similar to that used for sponge envenomations (see above).[145,147]

TOXIC VERTEBRATES

Stingrays

Stingrays are found in temperate to tropical waters worldwide and represent the most common source of human envenomations from vertebrates.[145] The stingray is armed with one to four venomous spines on a whiplike tail. Envenomations usually occur when an unwary swimmer steps on a buried stingray. The fish reflexively whips its tail upward, thrusting its spine into the victim. The venomous spine has a sharp tip with serrated edges that often cause a jagged laceration in addition to a puncture wound. Occasionally, a spine breaks off and remains in the wound.[154] Most envenomations occur on the ankle or foot. Initial symptoms include an intense, localized pain that may radiate centrally. Local edema and variable bleeding may occur. The pain intensifies, peaking after approximately 30 to 60 minutes.[155] The wound often appears dusky or cyanotic and progresses to an erythematous, hemorrhagic stage, occasionally with deep tissue involvement and frank necrosis.[154] Systemic effects can occur and include muscle cramps, nausea, vomiting, weakness, headache, diaphoresis, dizziness, and, in rare cases, seizures, paralysis, cardiovascular collapse, and death.[145] Initial treatment consists of hot-water immersion for 60 minutes to deactivate the venom, which is heat labile, and to provide pain relief. The wound should then be thoroughly irrigated and debrided, if needed. Hot-water immersion often provides adequate pain relief, but if pain continues, analgesics should be given, as well as tetanus prophylaxis. These wounds are at risk for infection, and antibiotics are often necessary, along with diligent wound care and close follow-up.[155]

Scorpionfish and Lionfish

The Scorpaenidae family consists of over 80 species and includes stonefish, lionfish, and zebrafish. Most Scorpaenidae species are reef-dwelling fish found in tropical waters of the Pacific and Indian oceans and the Red Sea.[156] They are exquisitely camouflaged, and envenomations occur when a victim inadvertently grasps or steps on one of these fish. Lionfish are often kept in home saltwater aquariums because of their beauty and, therefore, may be involved in accidental envenomations. These fish possess venomous spines covered by an integumentary sheath. The venom consists of a complex mixture of inflammatory mediators and heat-labile proteins and is injected after the spine punctures the skin.[154] The spine may also fracture and remain in the puncture site, causing a foreign-body reaction or acting as a nidus for secondary infection.

Clinical manifestations of envenomation vary from mild toxicity with the lionfish to severe life-threatening toxicity with the stonefish.[156] Initial symptoms include an intense pain at the sting site, with central radiation, that can continue to intensify for sev-

eral hours. The wound appears ecchymotic and cyanotic initially; subsequently, it often becomes erythematous and swollen. Localized tissue necrosis and skin sloughing can occur after several days.[155] Systemic symptoms can include nausea, vomiting, abdominal pain, headache, myalgias, weakness, tremor, syncope, hypotension, paralysis, seizures, cardiac arrhythmias, hypotension, cardiopulmonary arrest, and death.[154] Initial treatment is immediate immersion of the affected area in hot water (110° to 115° F) for 60 to 90 minutes in an attempt to denature the venom proteins and provide pain relief.[155] Supportive care for systemic symptoms, diligent local wound care (including evaluation for retained spines), tetanus prophylaxis, and analgesics are also the mainstay of treatment. An antivenin is available but is usually reserved for the more serious stonefish envenomations.[156]

Catfish

There are over 1,000 species of catfish worldwide that inhabit both freshwater and saltwater environments.[156] Freshwater catfish are typically sedentary bottom dwellers of slow-moving and often dirty waters. Saltwater catfish, on the other hand, travel in schools and typically stay on the move. Catfish have a smooth, scaleless skin and derive their name from the perioral barbels (which, contrary to popular belief, are incapable of inflicting any stings or envenomation). However, catfish do possess serrated dorsal and pectoral spines that have venom glands enclosed in an integumentary sheath. The spines can produce a deep puncture wound, and the glands release their venom after being traumatized.[157] Most stings occur when the fish is handled after capture, although occasionally, a wader will be stung on the foot. The initial symptom is intense pain, which radiates up the affected limb and is out of proportion to the mechanical trauma. The pain is followed by an intense inflammatory reaction that can include erythema, swelling, local hemorrhage, and tissue necrosis.[158] Systemic reactions are rare but can include nausea, vomiting, weakness, hypotension, syncope, and respiratory distress.[155]

The main concern with catfish stings, as with all aquatic trauma, is the risk of secondary infection. Catfish spines can be retained in the puncture site. Because spines are often radiopaque, x-rays should be obtained to locate them.[157] If there is evidence of a retained spine, the wound should be surgically explored and the foreign matter removed. Initial treatment includes immersion of the affected site in hot water (100° to 115° F) for 60 to 90 minutes, meticulous wound care and debridement, tetanus prophylaxis, and analgesics.[155]

Weeverfish

The weeverfish is the most venomous fish of the temperate zone and is found in the Mediterranean Sea, the Black Sea, and the eastern Atlantic coastal waters from North Africa to the North Sea.[156] It has four to eight sharp dorsal spines and two opercular spines, with which it can cause one of the most painful envenomations known. Weeverfish inhabit flat sandy or muddy bays and hide on the bottom with only their head exposed. Injuries usually occur to the hands or feet of fishermen or unwary swimmers. Weeverfish are typically sedentary but can be aggressive when provoked and can remain alive for many hours after removal from the water.[155]

Initially, weeverfish envenomations cause an intense pain that spreads to involve the entire limb; the pain is often so severe that the victim acts irrationally or loses consciousness. The wound becomes swollen, ecchymotic, and warm and may take months to heal. Systemic symptoms may occur, and secondary infection is common.[145] The affected area should be immediately placed in hot water, as for other marine invertebrate envenomations (see above), although this may not be as effective for weeverfish envenomations. The wound should be cleaned and gently explored, although retention of spines is uncommon. Tetanus prophylaxis should be administered. Systemic analgesics and possibly infiltration of local anesthetics or regional nerve blocks may be needed for pain relief.[155]

Table 6 Pathogens Associated with Aquatic Sources of Infections

Freshwater Organisms	*Saltwater Organisms*
Aeromonas hydrophila	*Aeromonas hydrophila*
Agrobacterium sanguineum	*Bacteroides fragilis*
Chromobacterium violaceum	*Clostridium perfringens*
Escherichia coli	*Erysipelothrix rhusiopathiae*
Pseudomonas aeruginosa	*Mycobacterium marinum*
Staphylococcus aureus	*Salmonella enteritidis*
Streptococcus species	*Staphylococcus aureus*
Vibrio parahaemolyticus	*Streptococcus* species
	Vibrio species

WATER-BORNE INFECTIONS

Freshwater and saltwater environments provide a medium for a host of microbes not typically encountered in traumatic wounds of nonaquatic origin. The concentration and diversity of these organisms vary, depending on temperature, sunlight, depth, salinity, nutrients, coexisting lifeforms, and pollutants. However, most aquatic microbes are heterotrophic, motile, gram-negative rods that are facultative anaerobes [*see Table 6*].[159,160] Antibiotic coverage should include gram-negative organisms, as well as *S. aureus* and *Streptococcus* species, which are the pathogens in most secondary aquatic wound infections.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Sacks JJ, Kresnow M, Houston B: Dog bites: how big a problem? Inj Prev 2:52, 1996

2. Statistical Abstracts of the United States, 112th ed. United States Department of Commerce, 1992

3. Nonfatal dog bite–related injuries treated in hospital emergency departments—United States, 2001. MMWR Morb Mortal Wkly Rep 52:605, 2003

4. Weiss HB, Friedman DI, Coben JH: Incidence of dog bite injuries treated in emergency departments. JAMA 279:51, 1998

5. Boenning DA, Fleisher GR, Campos JM: Dog bites in children: epidemiology, microbiology, and penicillin prophylactic therapy. Am J Emerg Med 1:17, 1983

6. Rest JG, Goldstein EJ: Management of human and animal bite wounds. Emerg Med Clin North Am 3:117, 1985

7. Goldstein EJ: Bite wounds and infection. Clin Infect Dis 14:633, 1992

8. Talan DA, Citron DM, Abrahamian FM, et al: Bacteriologic analysis of infected dog and cat bites. Emergency Medicine Animal Bite Infection Study Group. N Engl J Med 340:85, 1999

9. Garcia VF: Animal bites and *Pasteurella* infections. Pediatr Rev 18:127, 1997

10. Dire DJ: Emergency management of dog and cat bite wounds. Emerg Med Clin North Am 10:719, 1992

11. Griego RD, Rosen T, Orengo IF, et al: Dog, cat, and human bites: a review. J Am Acad Dermatol 33:1019, 1995

12. Sacks JJ, Sinclair L, Gilchrist J, et al: Breeds of dogs involved in fatal human attacks in

the United States between 1979 and 1998. J Am Vet Med Assoc 217:836, 2000

13. Goldstein EJ, Richwald GA: Human and animal bite wounds. Am Fam Physician 36:101, 1987

14. Dire DJ, Hogan DE, Riggs MW: A prospective evaluation of risk factors for infections from dog-bite wounds. Acad Emerg Med 1:258, 1994

15. Medeiros I, Saconato H: Antibiotic prophylaxis for mammalian bites. Cochrane Database Syst Rev (2):CD001738, 2001

16. Rupprecht CE, Gibbons RV: Prophylaxis against rabies. N Engl J Med 351:2626, 2004

17. Krebs JW, Mandel EJ, Swerdlow DL, et al: Rabies surveillance in the United States during 2003. J Am Vet Med Assoc 225:1837, 2004

18. Callaham M: Human and animal bites. Top Emerg Med 4:1, 1982

19. Glaser C, Lewis P, Wong S: Pet-, animal-, and vector-borne infections. Pediatr Rev 21:219, 2000

20. Compendium of animal rabies prevention and control, 2000. National Association of State Public Health Veterinarians, Inc. MMWR Recomm Rep 49(RR-8):21, 2000

21. Gehrke C: Results of the AVMA survey of US pet-owning households on companion animal ownership. J Am Vet Med Assoc 211:169, 1997

22. Paisley JW, Lauer BA: Severe facial injuries to infants due to unprovoked attacks by pet ferrets. JAMA 259:2005, 1988

23. Applegate JA, Walhout MF: Childhood risks from the ferret. J Emerg Med 16:425, 1998

24. Fischer RG, Edwardsson S, Klinge B, et al: The effect of cyclosporin-A on the oral microflora at gingival sulcus of the ferret. J Clin Periodontol 23:853, 1996

25. Krebs JW, Strine TW, Smith JS, et al: Rabies surveillance in the United States during 1994. J Am Vet Med Assoc 207:1562, 1995

26. Compendium of animal rabies control, 1996. National Association of State Public Health Veterinarians, Inc. J Am Vet Med Assoc 208:214, 1996

27. Human rabies prevention—United States, 1999. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 48(RR-1):1, 1999

28. Callaham M: Controversies in antibiotic choices for bite wounds. Ann Emerg Med 17:1321, 1988

29. Lindsey D, Christopher M, Hollenbach J, et al: Natural course of the human bite wound: incidence of infection and complications in 434 bites and 803 lacerations in the same group of patients. J Trauma 27:45, 1987

30. Broder J, Jerrard D, Olshaker J, et al: Low risk of infection in selected human bites treated without antibiotics. Am J Emerg Med 22:10, 2004

31. Goldstein EJ, Citron DM, Vagvolgyi AE, et al: Susceptibility of bite wound bacteria to seven oral antimicrobial agents, including RU-985, a new erythromycin: considerations in choosing empiric therapy. Antimicrob Agents Chemother 29:556, 1986

32. Brook I: Microbiology of human and animal bite wounds in children. Pediatr Infect Dis J 6:29, 1987

33. Goldstein EJ: Infectious complications and therapy of bite wounds. J Am Podiatr Med Assoc 79:486, 1989

34. Glass KD: Factors related to the resolution of treated hand infections. J Hand Surg [Am] 7:388, 1982

35. Wiley JF, 2nd: Mammalian bites: review of evaluation and management. Clin Pediatr (Phila) 29:283, 1990

36. Ordog GJ, Balasubramanium S, Wasserberger J: Rat bites: fifty cases. Ann Emerg Med 14:126, 1985

37. Smith JS, Fishbein DB, Rupprecht CE, et al: Unexplained rabies in three immigrants in the United States: a virologic investigation. N Engl J Med 324:205, 1991

38. Krebs JW, Smith JS, Rupprecht CE, et al: Mammalian reservoirs and epidemiology of rabies diagnosed in human beings in the United States, 1981–1998. Ann N Y Acad Sci 916:345, 2000

39. Moran GJ, Talan DA, Mower W, et al: Appropriateness of rabies postexposure prophylaxis treatment for animal exposures. Emergency ID Net Study Group. JAMA 284:1001, 2000

40. Freer L: North American wild mammalian injuries. Emerg Med Clin North Am 22:445, 2004

41. Ostrowski SR, Leslie MJ, Parrott T, et al: B-virus from pet macaque monkeys: an emerging threat in the United States? Emerg Infect Dis 4:117, 1998

42. Holmes GP, Chapman LE, Stewart JA, et al: Guidelines for the prevention and treatment of B-virus infections in exposed persons. B-Virus Working Group. Clin Infect Dis 20:421, 1995

43. Callaham ML: Treatment of common dog bites: infection risk factors. JACEP 7:83, 1978

44. Maimaris C, Quinton DN: Dog-bite lacerations: a controlled trial of primary wound closure. Arch Emerg Med 5:156, 1988

45. Chen E, Hornig S, Shepherd SM, et al: Primary closure of mammalian bites. Acad Emerg Med 7:157, 2000

46. Garbutt F, Jenner R: Best evidence topic report: wound closure in animal bites. Emerg Med J 21:589, 2004

47. Cummings P: Antibiotics to prevent infection in patients with dog bite wounds: a meta-analysis of randomized trials. Ann Emerg Med 23:535, 1994

48. Talan DA, Abrahamian FM, Moran GJ, et al: Clinical presentation and bacteriologic analysis of infected human bites in patients presenting to emergency departments. Clin Infect Dis 37:1481, 2003

49. Turner TW: Evidence-based emergency medicine/systematic review abstract. Do mammalian bites require antibiotic prophylaxis? Ann Emerg Med 44:274, 2004

50. Offerman SR, Smith TS, Derlet RW: Does the aggressive use of polyvalent antivenin for rattlesnake bites result in serious acute side effects? West J Med 175:88, 2001

51. Gold BS, Barish RA: Venomous snakebites: current concepts in diagnosis, treatment and management. Emerg Med Clin North Am 10:249, 1992

52. Smith TA, Figge HL: Treatment of snakebite poisoning. Am J Hospital Pharm 48:2190, 1991

53. Norris R: Snake envenomations: coral. eMedicine Jan 19, 2005 http://eugen.leitl.org/eMedicine_Snake_Envenomations_Coral_Article_Robert_Norris_MD.pdf

54. Sanford JP: Snakebites. Cecil Textbook of Medicine. Wyngaarden JB, Smith LH, Bennett JC, Eds. WB Saunders Co, Philadelphia, 1992

55. Bush SP: Snake envenomations: rattle. eMedicine Nov 2, 2004 http://www.emedicine.com/emerg/topic540.htm

56. White J: Snake venoms and coagulopathy. Toxicon 45:951, 2005

57. Nasu K, Ueda T, Miyakawa M: Intrauterine fetal death caused by pit viper poisoning in early pregnancy. Gynecol Obstet Invest 57:114, 2004

58. Bartholdi D, Selic C, Meier J, et al: Viper snakebite causing symptomatic intracerebral haemorrhage. J Neurol 251:889, 2004

59. Lavonas EJ, Gerardo CJ, O'Malley G: Initial experience with Crotalidae polyvalent immune Fab (ovine) antivenom in the treatment of copperhead snakebite. Ann Emerg Med 43:200, 2004

60. Boyer LV, Seifert SA, Clark RF, et al: Recurrent and persistent coagulopathy following pit viper envenomation. Arch Intern Med 159:706, 1999

61. Watt CH: Treatment of poisonous snakebite with emphasis on digit dermotomy. South Med J 78:694, 1985

62. Karthik S, Phadke KD: Snakebite-induced acute renal failure: a case report and review of the literature. Pediatr Nephrol 19:1053, 2004

63. McKinney PE: Out-of-hospital and interhospital management of crotaline snakebite. Ann Emerg Med 72:168, 2001

64. Lewis JV, Portera CA: Rattlesnake bite of the face: case report and review of the literature. Am Surgeon 60:681, 1994

65. Parrish HM: Analysis of 460 fatalities from venomous animals in the United States. Am J Med Sci 245:129, 1963

66. Dart RC, McNally J: Efficacy, safety, and use of snake antivenoms in the United States. Ann Emerg Med 37:181, 2001

67. Jurkovich GJ, Luterman A, McCullar K, et al: Complications of Crotalidae antivenin treatment. J Trauma 28:1032, 1988

68. Dart RC, Seifert SA, Carroll L, et al: Affinity-purified, mixed monospecific crotalid antivenom ovine Fab for the treatment of crotalid venom poisoning. Ann Emerg Med 30:33, 1997

69. Gillissen A, Theakston RD, Barth J, et al: Neurotoxicity, haemostatic disturbances and haemolytic anaemia after a bite by a Tunisian saw-scaled or carpet viper (Echis 'pyramidium'-complex): failure of antivenom treatment. Toxicon 32:937, 1994

70. Hardy DL, Jeter M, Corrigan JJ: Envenomation by the Northern blacktail rattlesnake (*Crotalus molossus molossus*): report of two cases and the in vitro effects of the venom on fibrinolysis and platelet aggregation. Toxicon 20:487, 1982

71. Seifert SA, Boyer LV, Dart RC, et al: Relationship of venom effects to venom antigen and antivenom serum concentrations in a patient with *Crotalus atrox* envenomation treated with a Fab antivenom. Ann Emerg Med 30:49, 1997

72. LoVecchio F, DeBus DM: Snakebite envenomation in children: a 10-year retrospective review. Wilderness Environ Med 12:184, 2001

73. Berlinger FG, Flowers HH: Some observations of the treatment of snakebites in Vietnam. Military Med 138:139, 1973

74. Wu CH, Hu WH, Peng YC, et al: Snakebite complicated with *Vibrio vulnificus* infection. Vet Hum Toxicol 34:283, 2001

75. Hall EL: Role of surgical intervention in the management of crotaline snake envenomation. Ann Emerg Med 37:175, 2001

76. Avila-Aquero ML, Paris MM, Hu S, et al: Systemic cytokine response in children bitten by snakes in Costa Rica. Snakebite Study Group. Pediatr Emerg Care 17:425, 2001

77. Weed HG: Nonvenomous snakebite in Massachusetts: prophylactic antibiotics are unnecessary. Ann Emerg Med 22:220, 1993

78. Pantanowitz L, Guidozzi F: Management of snake and spider bite in pregnancy. Obstet Gynecol Survey 51:615, 1996

79. Parrish HM, Kahn MS: Snakebite during pregnancy. Obstet Gynecol 27:468, 1966

80. Cruz NS, Alvarez RG: Rattlesnake bite complications in 19 children. Pediatr Emerg Care 10:30, 1994

81. Weber RA, White RR: Crotalidae envenomation in children. Ann Plast Surg 31:141, 1993

82. Schmidt JM: Antivenom therapy for snakebites in children: is there evidence? Curr Opin Pediatr 17:234, 2005

83. Offerman SR, Bush SP, Moynihan JA, et al: Crotaline Fab antivenom for the treatment of children with rattlesnake envenomation. Pediatrics 110:968, 2002

84. Mammen EF: Fibrinogen abnormalities. Semin Thromb Hemost 9:1, 1983

85. Isbister GK, White J: Clinical consequences of spider bites: recent advances in our understanding. Toxicon 43:477, 2004

86. Zukowski CW: Black widow spider bite. J Am Board Fam Pract 6:279, 1993

87. Sternlicht H, Fosson A: Partial paralysis following a black widow spider bite. J Kentucky Med Assoc 9:531, 1987

88. Allen C: Arachnid envenomations. Emerg Med Clin North Am 19:269, 1992

89. Isbister GK, White J, Currie BJ: Letter to the editor: spider bites: addressing mythol-

ogy and poor evidence. Am J Trop Med Hyg 72:361, 2005

90. Otten EJ: Venomous animal injuries. Emergency Medicine—Concepts and Clinical Practice. Rosen P, Barkin RM, Braen GR, et al, Eds. Mosby Year Book, St. Louis, 1992

91. Ryan PG: Preliminary report: experience with the use of dantrolene sodium in the treatment of bites by the black widow spider *Latrodectus hesperus*. J Toxicol Clin Toxicol 21:487, 1984

92. Jelinek GA: Widow spider envenomation (latrodectism): a worldwide problem. Wilderness Environ Med 8:226, 1997

93. Trethewy CE, Bolisetty S, Wheaton G: Red-back spider envenomation in children in Central Australia. Emerg Med (Fremantle) 15:170, 2003

94. Swanson DL, Vetter RS: Bites of brown recluse spiders and suspected necrotic arachnidism. N Engl J Med 352:700, 2005

95. Young VL, Pin P: The brown recluse spider bite. Ann Plast Surg 20:447, 1988

96. Vetter RS, Cushing PE, Crawford RL, et al: Diagnoses of brown recluse spider bites (loxoscelism) greatly outnumber actual verifications of the spider in four western American states. Toxicon 42:413, 2003

97. de Oliveira KC, Goncalves de Andrade RM, Piazza RM, et al: Variations in *Loxosceles* spider venom composition and toxicity contribute to the severity of envenomation. Toxicon 45:421, 2005

98. Gross AS, Wilson DC, King LE: Persistent segmental cutaneous anesthesia after a brown recluse spider bite. South Med J 83:1321, 1990

99. Wasserman G: Wound care of spider and snake envenomations. Ann Emerg Med 17:1331, 1988

100. Wong R, Hughes S, Voorhees J: Spider bites. Arch Dermatol 123:99, 1987

101. Sams HH, Hearth SB, Long LL, et al: Nineteen documented cases of *Loxosceles reclusa* envenomation. J Am Acad Dermatol 44:603, 2001

102. Sauer GC: Transverse myelitis and paralysis from a brown recluse spider bite. Missouri Med 72:603, 1975

103. Elbahlawan LM, Stidham GL, Bugnitz MC: Severe systemic reaction to *Loxosceles reclusa* spider bites in a pediatric population. Pediatr Emerg Care 21:177, 2005

104. Goto CS, Abramo TJ, Ginsburg CM: Upper airway obstruction caused by brown recluse spider envenomization of the neck. Am J Emerg Med 14:660, 1996

105. Glenn JB, Lane JE, Clark EK: Arachnid envenomation from the brown recluse spider. Clin Pediatr (Phila) 42:567, 2003

106. King LE, Rees RS: Dapsone treatment of brown recluse bite. JAMA 250:648, 1983

107. Gutowicz M, Fritz RA, Sonoga AL: Brown recluse spider bite: a literature review and case report. J Am Podiatric Med Assoc 79:142, 1979

108. Elston DM, Miller SD, Young RJ, et al: Comparison of colchicine, dapsone, triamcinolone, and diphenhydramine therapy for the treatment of brown recluse spider envenomation: a double-blind, controlled study in a rabbit model. Arch Dermatol 141:595, 2005

109. Wright SW, Wrenn KD, Murray L: Clinical presentation and outcome of brown recluse spider bites. Ann Emerg Med 30:28, 1997

110. Ben Welch E, Gales BJ: Use of stun guns for venomous bites and stings: a review. Wilderness Environ Med 12:111, 2001

111. Abroug F, Nouira S, Haguiga H, et al: High-dose hydrocortisone hemisuccinate in scorpion envenomation. Ann Emerg Med 30:23, 1997

112. Abroug F, Boujdaria R, Belghith M, et al: Cardiac dysfunction and pulmonary edema following scorpion envenomation. Chest 100:1057, 1991

113. Litovitz TL, Klein-Schwartz W, Rodgers GC Jr, et al: 2001 annual report of the American Association of Poison Control Centers Toxic Exposure Surveillance System. Am J Emerg Med 20:391, 2002

114. Bush SP, Gerardo C: Scorpion envenomations. eMedicine July 30, 2003 http://www.emedicine.com/emerg/topic524.htm

115. Gibly R, Williams M, Walter FG, et al: Continuous intravenous midazolam infusion for *Centruroides exilicauda* scorpion envenomation. Ann Emerg Med 34:620, 1999

116. Sofer S: Scorpion envenomation. Intensive Care Med 21:626, 1995

117. Curry SC, Vance MV, Ryan PJ, et al: Envenomation by the scorpion *Centruroides sculpturatus*. J Toxicol Clin Toxicol 21:417, 1983

118. Likes K, Banner W, Chavez M: *Centruroides exilicauda* envenomation in Arizona. West J Med 141:634, 1984

119. Rachesky IJ, Banner W, Dansky J, et al: Treatments for *Centruroides exilicauda* envenomation. Am J Dis Child 138:1136, 1984

120. Berg RA, Tarantino MD: Envenomation by the scorpion *Centruroides exilicauda* (*C. sculpturatus*): severe and unusual manifestations. Pediatrics 87:930, 1991

121. Bawaskar HS, Bawaskar PH: Prazosin for vasodilator treatment of acute pulmonary oedema due to scorpion sting. Ann Tropic Med Parasitol 81:719, 1987

122. Karnad DR: Haemodynamic patterns in patients with scorpion envenomation. Heart 79:485, 1998

123. Abroug F, ElAtrous S, Nouira S, et al: Serotherapy in scorpion envenomation: a randomised controlled trial. Lancet 354:906, 1999

124. Belghith M, Boussarsar M, Haguiga H, et al: Efficacy of serotherapy in scorpion sting: a matched-pair study. J Toxicol Clin Toxicol 37:51, 1999

125. Amaral CF, Rezende NA: Treatment of scorpion envenoming should include both a potent specific antivenom and support of vital functions. Toxicon 38:1005, 2000

126. Bond GR: Antivenin administration for *Centruroides* scorpion sting: risks and benefits. Ann Emerg Med 21:788, 1992

127. de Rezende NA, Chavez-Olortegui C, Amaral CF: Is the severity of *Tityus serrulatus* scorpion envenoming related to plasma venom concentrations? Toxicon 34:820, 1996

128. LoVecchio F, Welch S, Klemens J, et al: Incidence of immediate and delayed hypersensitivity to *Centruroides* antivenom. Ann Emerg Med 34:615, 1999

129. Hamed MI: Treatment of the scorpion envenoming syndrome: 12 years experience with serotherapy. Int J Antimicrob Agents 21:170, 2003

130. Watson WA, Litovitz TL, Klein-Schwartz W, et al: 2003 annual report of the American Association of Poison Control Centers Toxic Exposure Surveillance System. Am J Emerg Med 22:335, 2004

131. Kemp SF, DeShazo RD, Moffitt JE, et al: Expanding habitat of the imported fire ant (*Solenopis invicta*): a public health concern. J Allergy Clin Immunol 105:683, 2000

132. Handbook of Clinical Toxicology and Animal Venoms and Poisons. Meier J, White J, Eds. CRC Press, New York, 1995, p 331

133. deShazo RD, Williams DF, Moak ES: Fire ant attacks on residents in health care facilities: a report of two cases. Ann Intern Med 131:424, 1999

134. Elgart GW: Ant, bee, and wasp stings. Dermatol Clin 8:229, 1990

135. Sherman AM, Bechtel HB, EricksonTB: North American arthropod envenomation and parasitism. Wilderness Medicine, 4th ed. Mosby, St. Louis, 2001, p 863

136. Elston DM, Stockwell S: What's eating you? Bedbugs. Cutis 65:262, 2000

137. Rosen T: Caterpillar dermatitis. Dermatol Clin 8:245, 1990

138. Spach DH, Liles WC, Campbell GL, et al: Tick-borne diseases in the United States. N Engl J Med 329:936, 1993

139. Grattan-Smith PJ, Morris JG, Johnston HM, et al: Clinical and neurophysiological features of tick paralysis. Brain 120:1975, 1997

140. Dworkin MS, Shoemaker PC, Anderson DE: Tick paralysis: 33 human cases in-Washington State. Clin Infect Dis 29:1435, 1999

141. Fenner PJ: Dangers in the ocean: the traveler and marine envenomation. I. Jellyfish. J Travel Med 5:135, 1998

142. Burnett JW, Calton GJ: Venomous pelagic coelenterates: chemistry, toxicology, immunology and treatment of their stings. Toxicon 25:581, 1987

143. Bonnet MS: The toxicology of the *Chironex fleckeri* jellyfish. Br Homeopath J 88:62, 1999

144. Burnett JW: Human injuries following jellyfish stings. Maryland Med J 41:509, 1992

145. Auerbach PS: Envenomation by aquatic invertebrates, envenomation by aquatic vertebrates.Wilderness Medicine, 4th ed. Auerbach PS, Ed. Mosby, St. Louis, 2001, p 1450

146. Kaufman MB: Portuguese man-of-war envenomation. Pediatr Emerg Care 8:27, 1992

147. Handbook of Clinical Toxicology and Animal Venoms and Poisons. Meier J, White J, Eds. CRC Press, New York, 1995, p 27

148. Brown CK, Shepherd SM: Marine trauma, envenomations, and intoxications. Emerg Med Clin North Am 10:385, 1992

149. Fenner PJ, Williamson JA, Burnett JW, et al: First aid treatment of jellyfish stings in Australia: response to a newly differentiated species. Med J Aust 158:498, 1993

150. Kumar S, Hlady WG, Malecki JM: Risk factors for seabather's eruption: a prospective cohort study. Public Health Rep 112:59, 1997

151. Baden HP, Burnell JW: Injuries from sea urchins. South Med J 70:459, 1997

152. Auerbach PS: Marine envenomations. N Engl J Med 325:486, 1991

153. Wu CH, Narahashi T: Mechanism of action of novel marine neurotoxins on ion channels. Annu Rev Pharmacol Toxicol 28:141, 1998

154. Cooper NK: Stone fish and stingrays: some notes on the injuries that they cause to man. J R Army Med Corps 137:136, 1991

155. McGoldrick J, Marx JA: Marine envenomations. J Emerg Med 9:497, 1991

156. Gwee MC, Gopalakrishnakone P, Yuen R., et al: A review of stonefish venoms and toxins. Pharmacol Ther 64:509, 1994

157. Shepherd S, Thomas SH, Stone K: Catfish envenomation. J Wilderness Med 5:67, 1994

158. Bolmkalns AL, Otten EJ: Catfish spine envenomation: a case report and literature review. Wilderness Environ Med 10:242, 1999

159. Auerbach PS, Yajko DM, Nassos PS, et al: Bacteriology of the freshwater environment: implications for clinical therapy. Ann Emerg Med 16:1016, 1987

160. Auerbach PS, Yajko DM, Nassos PS, et al: Bacteriology of the marine environment: implications for clinical therapy. Ann Emerg Med 16:643, 1987

Acknowledgments

Figures 1, 2, and 3 Photography by Bill Love/Blue Chameleon Ventures.

Figure 4 Photography from Animals Animals/Jim Bockowski. Used by permission.

Figure 5 Photography by B. J. Kaston (Trans. no K13623). Courtesy of the Department of Library Services, the American Museum of Natural History, New York.

III CARDIAC RESUSCITATION

TERRY J. MENGERT, M.D.

Out-of-hospital sudden cardiac arrest claims the lives of more than 300,000 persons in the United States each year, making it the leading cause of death.[1-4] In fact, approximately 50% of all cardiac deaths are sudden deaths.[5] In hospitals, a minimum of 370,000 patients also suffer a cardiac arrest, followed by an attempted, but only sometimes successful, resuscitation.[6] Although most victims of sudden death have underlying coronary artery disease (70% to 80%), sudden death is the first manifestation of the disease in half of these persons.[2] Other causes and contributing factors include abnormalities of the myocardium (i.e., chronic heart failure or hypertrophy from any cause), electrophysiologic abnormalities, valvular heart disease, congenital heart disease, and miscellaneous inflammatory and infiltrative disease processes (e.g., myocarditis, sarcoidosis, and hemochromatosis).[7-9]

The pathophysiology that culminates in a sudden cardiac death is complex and poorly understood. It likely represents a mix of electrical abnormalities combined with acute functional triggers, such as myocardial ischemia, central and autonomic nervous system effects, electrolyte abnormalities, and even pharmacologic influences.[1] Classically, most sudden deaths that occur in adults in the community are thought to be secondary to ventricular tachycardia (VT) that quickly degenerates into ventricular fibrillation (VF). In a 10-year study in the Seattle area, the different arrhythmias found in prehospital cardiac arrest patients presumed to have underlying cardiovascular disease were VF (45%), asystole (31%), pulseless electrical activity (PEA; 10%), VT (1%), and other arrhythmias (14%).[3] Studies indicate that the out-of-hospital incidence of VF has decreased in recent years, probably because of the decrease in mortality from coronary artery disease.[10]

The Chain of Survival

The resuscitation of an adult victim of sudden cardiac arrest should follow an orderly sequence, no matter where the patient's collapse occurs. This sequence is called the chain of survival.[11] It comprises four elements, all of which must be instituted as rapidly as possible: activation of the emergency medical service network, cardiopulmonary resuscitation (CPR), early defibrillation, and provision of advanced care.

ACTIVATION OF EMERGENCY MEDICAL SERVICES

A person in cardiac arrest is unresponsive and pulseless, although agonal respirations may last for minutes. Confirm unresponsiveness by speaking loudly and gently shaking the patient. If the patient is truly unresponsive, immediately call for help by activating the emergency medical service in the community (in most locales, this means calling 911); or if the patient is already in the hospital, call a code (e.g., code blue, code 199). If an automated external defibrillator (AED) is available, have it brought to the resuscitation scene. AEDs are both easily used and lifesaving.[12-16]

INITIATION OF CPR

While awaiting the arrival of a defibrillator and advanced help, the rescuer assesses the patient's airway, breathing, and circulation [*see* The Primary Survey, *below*], and initiates CPR [*see Table 1*]. When CPR is started within 4 minutes of collapse, the likelihood of patient survival at least doubles.[17,18]

INITIATION OF DEFIBRILLATION

When the AED or monitor-defibrillator arrives, attach it appropriately to the patient and analyze the patient's rhythm; if the patient is in VF or pulseless VT, a defibrillatory shock should be rapidly applied [*see Tables 2 and 3*]. If required, two additional shocks may be administered sequentially. The importance of rapid access to defibrillation cannot be overemphasized. In a patient who is dying from a shockable rhythm, the chance of survival declines by 7% to 10% for every minute that defibrillation is delayed.[19]

INITIATION OF ADVANCED CARE

If the patient remains pulseless despite the steps described above, resume CPR; establish a definitive airway, confirm its correct placement, and then secure it; establish intravenous access; then administer appropriate medications as determined by the rhythm and the arrest circumstances. If the patient is in VF or pulseless VT, repeated attempts at defibrillation are interspersed with delivery of vasoactive and antiarrhythmic drugs [*see Table 4*].

RESUSCITATION OUTCOME

When every link in the chain of survival is quickly and sequentially available, the patient is provided an optimal opportunity for return of spontaneous circulation.[19-22] In the United States, individual communities report survival rates of 4% to 40% or more in cases of sudden cardiac death.[23-27] Prehospital victims of VF have had survival rates to hospital discharge of

Table 1 Initial Resuscitation Steps in the Unresponsive Patient

- Confirm unresponsiveness
- Activate the emergency medical system
  - In most community locales, call 911
  - In the hospital, activate the appropriate code response
- Call for an automatic external defibrillator (AED)
- Begin basic life support (CPR)
  - Open airway
  - Check breathing; if not breathing, deliver two initial breaths
  - Check for a carotid pulse; if pulseless, do the following:
    - Begin chest compressions at the rate of 100 compressions/min, depressing the sternum 1.5–2 in. per compression in patients older than 8 yr
    - Intersperse ventilations with chest compressions: in nonintubated patients, deliver 15 compressions, pause for two breaths, then repeat; in intubated patients, deliver one breath every 5 sec, with no pause in compressions
- Reassess for return of spontaneous circulation every 1–3 min
- When defibrillator arrives, immediately analyze and treat arrhythmia
  - Attach patient to AED [*see Table 2*] or the monitor-defibrillator [*see Table 3*]
  - Analyze arrhythmia and treat as appropriate [*see Figure 2*]

Table 2 Using an Automatic External Defibrillator in Patients Older than 8 Years

Automatic external defibrillator (AED) arrives (CPR is in progress)
- Place AED beside patient.
- Turn on the AED.
- Attach the electrodes to the AED (they may be preattached).
- Attach the electrode pads to the patient (as diagrammed on the pads).

AED analyzes patient's rhythm
- Stop CPR (and ensure no one is touching the patient).
- Press the Analyze button on the AED (some devices analyze the rhythm automatically as soon as the pads are placed on the patient).

AED instructs rescuers (via an audible voice prompt and/or on-screen instructions)
- Shock is indicated: clear the patient (ensure no one is touching the patient) and push the Shock button.
- After delivering shock, press the Analyze button again; the sequence of analysis followed by shock (if so indicated) may be performed a total of three times.

or

- Shock not indicated: reassess the patient for signs of circulation; if present, assess the adequacy of breathing; if there are no signs of circulation, resume CPR for 1–2 min. After 1 min of CPR, assess the patient again for signs of circulation; if present, assess the adequacy of breathing. If the patient is still pulseless, repeat rhythm analysis, followed (if indicated) by shock steps.

greater than 50% when an AED was expeditiously used.[28] Many other factors also influence patient survival, however; these include whether the patient's collapse was witnessed, the rapidity and effectiveness of bystander CPR, the rhythm associated with the cardiac arrest, and underlying comorbidities.[29,30] With inpatient cardiac arrest, for example, overall survival rates vary from 9% to 32%,[31-37] but in one study, survival to hospital discharge was 30% for patients with primary heart disease, 15% for patients with infectious diseases, and only 8% for patients with other end-stage diseases (e.g., cancer, lung disease, liver failure, or renal failure).[38]

Such statistics underline the importance of using cardiac resuscitation appropriately and with discrimination. Cardiac resuscitation provides rescuers with powerful tools that save the lives of thousands of people every year. These techniques are capable of returning patients who would otherwise die to productive and meaningful lives. However, cardiac resuscitation should not be employed to reverse timely and natural death. Under those circumstances, it has the potential to lengthen the dying process and to increase human suffering. All practitioners are well advised to remember that "death is not the opposite of life, death is the opposite of birth. Both are aspects of life."[39] It is untimely death that requires immediate intervention with cardiac resuscitation.

The Primary and Secondary Surveys of Cardiac Resuscitation

A cardiac resuscitation is a stressful event for everyone involved. Too often, clinic and inpatient cardiac arrests and their management are episodes of chaos in the busy lives of resident and attending physicians. Yet, it has been eloquently stated that a good resuscitation team should function like a fine symphony orchestra.[40] Such skill levels require dedicated individual and team practice and careful code-team organization. Mastery in cardiac resuscitation is in fact a lifelong pursuit that requires training and retraining in advanced cardiac life support (ACLS); regular practice and review; and leadership and team skill development. Its key elements include not only the resuscitation itself but the response to the announcement of a code, postresuscitation stabilization of the patient, notification of the family and primary care provider, and code critique and debriefing. To help practitioners learn and apply some of the most essential techniques used in cardiac resuscitation more easily and effectively, the American Heart Association (AHA) has developed the concepts of primary and secondary surveys of a patient in atraumatic cardiac arrest.[41]

Table 3 Using a Manual Defibrillator[52,89]

Defibrillator arrives (CPR is in progress)
- Place defibrillator beside patient.
- Turn defibrillator on (initial energy level setting is typically 200 J).*
- Set Lead Select switch to Paddles. Alternatively, if patient is already attached to monitor leads, set Lead Select switch to lead I, II, or III; ensure all three leads are correctly attached to the patient and the defibrillator: white to right shoulder, black to left shoulder, red to ribs on left side.
- Apply gel to paddles or place conductor pads on patient's chest. Some devices use disposable electrode patches that are prepasted with a conducting gel. In either case, the appropriate positions of the paddles with applied gel, conductor pads, or disposable paddles are as follows: sternal paddle is placed to the right of the sternum, just below the right clavicle; apex paddle is placed to the left of the left breast, centered in the left midaxillary line at the fifth intercostal space.

Analyze rhythm

Briefly withhold CPR
- If using paddles to assess rhythm, apply paddles as described with firm pressure (25 lb of pressure to each paddle) and visually assess rhythm on monitor (if using leads, briefly withhold CPR and assess rhythm in leads I, II, or III). If rhythm is either pulseless VT or VF, proceed as follows:

Defibrillate, then reassess
- Announce to resuscitation team, "Charging defibrillator, stand clear!" and press Charge button on either paddles or defibrillator (initial energy, 200 J, not synchronized).*
- Warn resuscitation team that a defibrillatory shock is coming: "I am going to shock on three! ONE, I'm clear; TWO, you're clear, THREE, everybody's CLEAR!" Simultaneously with these statements, visually ensure that no resuscitation team member is in contact with patient.
- Press the Discharge buttons on both paddles simultaneously to deliver a defibrillatory shock.
- Reassess rhythm on monitor; if patient is still in VT or VF, recharge defibrillator (now 300 J)* and repeat process of loudly informing team members by giving the warning statements as above, and then apply defibrillatory shock.
- Reassess rhythm on monitor; if patient is still in VT or VF, recharge defibrillator (now 360 J)* and repeat process of loudly informing team members by giving the warning statements as above, and then apply defibrillatory shock.
- Reassess rhythm on monitor; if patient is still in VT or VF, resume CPR and continue with resuscitation sequence [*see Figure 2*].

*Note: if using a biphasic defibrillator, a lower initial defibrillatory energy level (< 200 J) without energy escalation on subsequent shocks is acceptable.

VF—ventricular fibrillation VT—ventricular tachycardia

Table 4 Drugs Useful in Cardiac Arrest[3,90]

Category	*Drug and Doses Supplied*	*Indications in Cardiac Arrest*	*Adult Dosage*	*Comments*
Vasopressors	Epinephrine, 1 mg in 10 ml emergency syringe; 1 mg/ml (1 ml and 30 ml vials)	Pulseless VT or VF unresponsive to initial defibrillatory shocks; PEA; asystole	1 mg I.V. push; may repeat every 3–5 min for as long as patient is pulseless; can also be given via the endotracheal route: 2–2.5 mg diluted with normal saline (NS) to 10 ml total volume	I.V. boluses of epinephrine (1 mg) are appropriate only in pulseless cardiac arrest patients; if continued epinephrine is required postresuscitation, a continuous infusion should be started (1–10 μg/min). High-dose epinephrine (up to 0.2 mg/kg I.V. per dose) does not improve survival to hospital discharge in cardiac arrest patients and is no longer recommended in adults.
	Vasopressin, 20 IU/ml (1 ml vial)	Pulseless VT or VF unresponsive to initial defibrillatory shocks	40 IU I.V. push, single dose only; can also be given via endotracheal tube: same dose, diluted with NS to 10 ml total volume	If no response after 10 min of continued resuscitation, administer epinephrine, as above.
Antiarrhythmics	Amiodarone, 50 mg/ml (3 ml vial)	Pulseless VT or VF unresponsive to initial defibrillatory shocks and epinephrine plus shock(s)	VT/VF: 300 mg diluted in 20–30 ml; NS or D5W rapid I.V. push; a repeat dose of 150 mg may be given if required in 5 min; maximum dose in 24 hr should not exceed 2,200 mg	Side effects may include hypotension and bradycardia in the postresuscitation phase.
	Lidocaine, 50 mg or 100 mg in 5 ml emergency syringes; premixed bag, 1 g/250 ml or 2 g/250 ml	Pulseless VT or VF unresponsive to initial defibrillatory shocks and epinephrine plus shock(s)	Initial dose: 1–1.5 mg/kg I.V.; for refractory VF or unstable VT, may repeat 1–1.5 mg/kg I.V. in 3–5 min; maximum dose, 3 mg/kg May also be given endotracheally: 2–4 mg/kg diluted with normal saline to 10 ml total volume	If lidocaine is effective, initiate continuous I.V. infusion at 2–4 mg/min when patient has return of a perfusing rhythm (but do not use if this rhythm is an idioventricular rhythm or third-degree heart block with an idioventricular escape rhythm). Continuous infusion should begin at 1 mg/min in congestive heart failure or chronic liver disease or in elderly patients.
	Magnesium sulfate, 500 mg/ml (2 ml and 10 ml vials), or 10 ml emergency syringe	Pulseless VT or VF unresponsive to initial defibrillatory shocks and epinephrine plus shock(s) if suspected hypomagnesemic state	Administer 1–2 g diluted in 100 ml D5W I.V. over 1–2 min Total body magnesium deficits should be replaced gradually after initial therapy has stabilized the emergency: administer 0.5–1 g/hr for 3–6 hr, then reassess continued need	Measured magnesium levels correlate only approximately with the actual level of deficiency. Patients with renal insufficiency are at risk for dangerous hypermagnesemia; use appropriate caution. Side effects may include bradycardia, hypotension, generalized weakness, and temporary loss of reflexes.
	Procainamide, 100 mg/ml (10 ml injection); 500 mg/ml (2 ml vial)	Recurrent or intermittent pulseless VT or VF	20–30 mg/min I.V. (up to 50 mg/min if situation is critical); maximum dose is 17 mg/kg over time (but maximum dose is reduced to 12 mg/kg in setting of cardiac or renal dysfunction) Maintenance infusion is 1–4 mg/min	Administer procainamide during a perfusing rhythm. Stop procainamide administration when arrhythmia is adequately suppressed, hypotension occurs, QRS widens to > 50% of original duration, or maximum dose is administered.
Anticholinergic	Atropine, 1 mg in 10 ml emergency syringe	Asystole or PEA (if rate of rhythm is slow)	For asystole or PEA: 1 mg I.V. every 3–5 min up to 3 mg May be given via ET tube: 2–3 mg diluted with normal saline to 10 ml	Minimal adult dose is 0.5 mg. Avoid use in type II second-degree heart block or third-degree heart block.
Miscellaneous	Bicarbonate, 50 mEq in 50 ml emergency syringe	Significant hyperkalemia Significant metabolic acidosis unresponsive to optimal CPR, oxygenation, and ventilation Certain drug overdoses, including tricyclic antidepressants and aspirin	Hyperkalemia therapy: 50 mEq I.V. Metabolic acidosis: 1 mEq/kg slow I.V. push; may repeat half initial dose in 10 min; ideally, ABGs should help guide further therapy Use in overdose: discuss with toxicologist	In non–dialysis-dependent hyperkalemic patients, bicarbonate is most useful if metabolic acidosis is also present; bicarbonate is less effective in dialysis-dependent renal failure patients. The use of bicarbonate in metabolic acidosis management in cardiac arrest patients is controversial. Side effects may include sodium overload, hypokalemia, and metabolic alkalosis.
	Calcium chloride, 100 mg/ml in 10 ml prefilled syringe	Significant hyperkalemia Calcium channel blocker drug overdose Profound hypocalcemia of other causes	In hyperkalemia: 5–10 ml slow I.V. push; may repeat if required In calcium channel blocker overdose: discuss with toxicologist	Do not use if hyperkalemia is suspected to be caused by acute digoxin poisoning. Do not combine in same I.V. with sodium bicarbonate. Calcium chloride is not a routine medication in cardiac arrest.

Note: All medications used during cardiac arrest, when given via a peripheral venous site in an extremity, should be followed by a 20 ml I.V. saline bolus and elevation of the extremity for 10 to 20 sec.

ABG—arterial blood gases D5W—5% dextrose in water ET—endotracheal PEA—pulseless electrical activity VF—ventricular fibrillation VT—ventricular tachycardia

THE PRIMARY SURVEY

The primary survey for the victim of sudden cardiac arrest consists of the appropriate assessment of the patient's airway (A), breathing (B), and circulation (C) and the simultaneous application of expert CPR until defibrillation (D) becomes possible (assuming the patient is in VF or pulseless VT). Thus, the primary survey includes the second and third links in the chain of survival (see above).

In 1958, Kouwenhoven noted that when his research fellow forcefully applied external defibrillating electrodes on a dog's chest in the laboratory, an arterial pressure wave occurred.[42] Further study and refinements led to the technique of closed-chest CPR, the careful description of which was published in 1960.[43] The first report of the use of this technique in patients was in 1961.[44] Since those early days, the fundamentals of closed-chest CPR have remained relatively unchanged. Mouth-to-mouth, mouth-to-mask, or bag-valve-mask ventilation oxygenates the blood. Chest compressions produce forward blood flow. This flow appears to result from a combination of direct compression of the heart and intrathoracic pressure changes.[45,46]

CPR in isolation does not defibrillate the heart. Its main benefit is to extend patient viability until a defibrillator and advanced interventions become available and, one hopes, succeed in restoring spontaneous circulation in the patient. CPR is not nearly as effective as a contracting heart; systolic arterial pressure peaks of 60 to 80 mm Hg may be generated, but diastolic blood pressure remains low, and a cardiac output of only 25% to 30% of normal can be achieved even under optimal conditions.[47] Still, effective CPR is critical to keeping the patient alive. It is worth remembering that the most important rescuers at a cardiac resuscitation are those who are performing expert CPR, because it is only through their efforts that the patient's heart and brain are kept viable until defibrillation and other advanced interventions can restore spontaneous circulation.

After unresponsiveness is confirmed, the emergency medical system is activated and an AED is called for; the primary survey (A, B, C, and D) proceeds as described (see below) until the AED arrives.

Airway Optimization

Open the patient's mouth and optimize the airway in the nontrauma patient by use of the head-tilt and chin-lift maneuver. A jaw-thrust maneuver should be used instead of the head-tilt technique if cervical spine injury is suspected. In patients with suspected spine injury, proper spine alignment must be maintained throughout all phases of the resuscitation. In such circumstances, as equipment becomes available, the patient's spine requires immobilization with a padded backboard, hard cervical collar, appropriate bolstering around the patient's head to prevent movement, and strapping of the patient to the backboard.[48]

Breathing Assessment

To assess breathing, the rescuer places his or her cheek close to the patient's mouth and looks, listens, and feels for patient respirations. If the respirations are agonal or the patient is apneic, the rescuer then delivers two initial breaths. Each breath is delivered over 1.5 to 2.0 seconds. The patient's chest should rise with each delivered breath, and exhalation is allowed for between breaths. Breaths may be delivered using the mouth-to-mouth technique with appropriate barrier precautions (the patient's nose should be pinched if the mouth-to-mouth technique is used) or mouth-to-mask technique. The ideal device, if available, is a bag-valve-mask device attached to high-flow oxygen; this allows the delivery of a substantially higher oxygen concentration to the patient. If the patient cannot be ventilated, the rescuer repositions the airway and attempts the technique again. If the airway is still obstructed, up to five abdominal thrusts are then applied, followed by a finger sweep of the oropharynx to relieve suspected obstruction, and then ventilation attempts are repeated. Definitive intervention for an obstructed airway in the hospital setting may involve laryngoscopic visualization of the cause of obstruction and foreign-body removal. If an adequate airway cannot be established by less invasive means, cricothyrotomy may be required.

CPR Initiation

The health care rescuer next checks for a carotid pulse in the unresponsive patient but should allow no more than 10 seconds to do so. (The AHA no longer recommends pulse checks for rescuers who are not health care providers.[49] Instead, lay rescuers should initiate chest compressions if the patient is not breathing, coughing, or moving after the initial two breaths). If the patient has no carotid pulse, begin chest compressions. The patient should be on a firm surface, and the heel of the rescuer's hand should be in the center of the inferior half of the patient's sternum (but cephalad to the xiphoid process). The rescuer's other hand is placed on top of the lower hand, with the fingers interlocked.

The rescuer's arms are held straight, with the force of each compression coming from the rescuer's trunk. In patients older than 8 years, the sternum is smoothly compressed by 1.5 to 2.0 inches, then released. The duration of the compression-release cycle is divided equally between compression and release. The rate of chest compression is 100 compressions/min in patients older than 8 years. The chest should be allowed to rebound to its precompression dimensions between compressions, but the resuscitator's palm closest to the patient should remain in contact with the sternum.

In nonintubated patients, chest compressions are regularly interrupted for the delivery of ventilations. The sequence is the same, regardless of whether one-rescuer or two-rescuer CPR is being performed: the rescuer delivers 15 compressions, pauses for two breaths (each given over 2 seconds), then resumes compressions. In endotracheally intubated patients, no pause for ventilation is necessary; every 5 seconds, one ventilation is delivered over a period of 1 to 2 seconds, while compressions continue.[18]

The optimal timing and ratio of ventilations to compressions in CPR is an ongoing area of research, which may lead to changes in the current recommendations. In the porcine model, for example, optimal neurologic outcome was achieved with the use of only compressions for the first 4 minutes, followed by a compression-ventilation ratio of 100:2.[50] In the prehospital setting, when rapid advanced care is available within minutes, bystander-initiated mouth-to-mouth ventilation combined with chest compressions offers no advantage over chest compressions alone.[51]

Good technique is critical throughout CPR delivery. The patient should have carotid pulses with chest compressions and should have appropriate breath sounds and chest movement with ventilations. Interestingly, femoral pulsations with CPR do not necessarily indicate effective CPR; these pulsations often are venous rather than arterial. Quantitative end-tidal carbon dioxide levels can be monitored, if practical. Higher levels correlate with more effective CPR and increased survival.[52] The patient should be reassessed for return of spontaneous circulation every 1 to 3 minutes [*see Table 1*].

Defibrillation

When the monitor-defibrillator or AED arrives, it is attached to the patient; the rhythm is analyzed, and if the patient is in VF or pulseless VT, defibrillation is provided [*see Tables 2 and 3*].

Defibrillation is thought to work by simultaneously depolarizing a sufficient mass of cardiac myocytes to make the cardiac tissue ahead of the VT or VF wavefronts refractory to electrical conduction. Subsequently, the sinus node or another appropriate pacemaker region of the heart with inherent automaticity can resume orderly depolarization-repolarization, with return of a perfusing rhythm.[15,53] The sooner defibrillation occurs, the higher the likelihood of resuscitation. When defibrillation is provided immediately after the onset of VF, its success rate is extremely high.[54] In a study of sudden cardiac arrest patients in Nevada gambling casinos, the survival rate to hospital discharge was 74% for patients who received their first defibrillation no later than 3 minutes after a witnessed collapse.[28] In this study, defibrillation was delivered via an AED operated by casino security officers.

Early defibrillation is so critical that if a defibrillator is immediately available, its use traditionally takes precedence over CPR for patients in VF or pulseless VT of recent onset. If CPR is already in progress, it should of course be halted while defibrillation takes place. Newer defibrillators can compensate for thoracic impedance, ensuring that the selected energy level is in fact the energy that is delivered to the myocardial tissue. In addition, defibrillators that deliver biphasic defibrillation waveforms instead of the standard monophasic damped sinusoidal waveforms allow effective defibrillation at lower energy levels (< 200 joules) and without the need for energy-level escalation during subsequent shocks.[15,55-58] In the Optimized Response to Cardiac Arrest (ORCA) study, which involved 115 patients with prehospital VF, the 150-joule biphasic-shock AED was more effective than the traditional high-energy monophasic-shock AED in four respects: it was more successful in producing defibrillation with the first shock (96% versus 59%); it led to a higher rate of ultimate success with defibrillation (100% versus 84%); it had a better rate of return of spontaneous circulation (76% versus 54%); and its use was associated with a higher rate of good cerebral performance in the survivors (87% versus 53%).[59] There were no differences, however, in terms of survival to hospital admission or discharge, and replication of the ORCA findings is lacking at this time. Current AHA guidelines state that lower-energy biphasic waveform defibrillators are safe and have equivalent or higher efficacy for termination of VF, as compared with the standard monophasic waveform defibrillator.[15,49]

Ongoing research suggests that the duration of VF is a consideration in deciding whether to defibrillate immediately and as soon as a defibrillator is available or to perform CPR for a brief period first to "prime the pump" before proceeding to defibrillation. In the porcine model in the setting of prolonged VF (> 10 minutes), CPR before countershock provides several physiologic benefits.[60] Studies have found that patients with VF of longer than 5 minutes' duration had better return of spontaneous circulation, survival to hospital discharge, and 1-year survival if ambulance personnel provided 3 minutes of CPR before performing defibrillation than if ambulance personnel performed defibrillation immediately after arriving at the scene; however, some experts question the validity of these results, on the basis of study design.[61,62]

THE SECONDARY SURVEY

The secondary survey for a victim of persistent cardiac arrest takes place after completion of the primary survey. Like the primary survey, the secondary survey follows an ABCD format, which in this case consists of advanced airway interventions (A); optimized oxygenation and ventilation by confirmation of endotracheal (ET) tube placement and repeated reassessment of the adequacy of delivered breaths (B); intravenous access and appropriate medication delivery to the patient's circulation (C); and definitive therapy (D), based on a differential diagnosis that considers the specific disease processes thought to be responsible for, or contributing to, the cardiac arrest. The secondary survey includes the fourth link in the chain of survival, rapid advanced care (see above).

Placement of an Advanced Airway

Patients who remain in cardiac arrest after completion of the primary survey require placement of an advanced airway. Depending on the setting and the experience of the rescuers, this advanced airway may be a laryngeal mask airway, an esophageal-tracheal Combitube (a tracheal tube bonded side by side with an esophageal obturator), or an ET tube.[36,63,64] The laryngeal mask airway and the Combitube can be placed by personnel with less training than that required for endotracheal intubation, and they do not require additional special equipment or visualization of the vocal cords. Nevertheless, oral endotracheal intubation is generally the preferred advanced airway technique for cardiac resuscitation, especially in the hospital setting, where experienced intubators are generally present; in the prehospital setting, the evidence supporting endotracheal intubation remains inconsistent. Endotracheal intubation isolates the airway, maintains airway patency, helps protect the trachea from the ever-present risk of aspiration, helps permit optimal oxygenation and ventilation of the patient, allows for tracheal suctioning, and even provides a route for delivery of some medications to the systemic circulation (via the pulmonary circulation) if intravenous access is unobtainable or lost.[63]

Optimization of Breathing and Ventilation

When a cardiac arrest patient undergoes endotracheal intubation, correct positioning of the ET tube must be immediately confirmed and regularly reconfirmed during and after the resuscitation [*see Table 5*]. Routine use of an esophageal detector device or end-tidal CO_2 detector is recommended, along with careful patient examination. Caution is necessary with qualitative colorimetric end-tidal CO_2 detectors because both false positive and false negative results have been documented during cardiac arrests.[65] Breath sounds should be present during auscultation over the anterior and lateral chest walls, and the patient's chest should rise and fall with delivered ventilations. No gurgling should be heard when the epigastrium is auscultated. The ET tube should be inserted to the appropriate depth marking: for average-sized adults, this is 21 cm at the corner of the mouth in a woman and 23 cm in a man. The patient's skin color should be reasonable (i.e., not dusky or cyanotic), provided that the patient's pigmentation allows such assessment.

Once correct positioning is confirmed, the ET tube is then appropriately secured to prevent its dislodgment. When feasible, an arterial blood gas (ABG) measurement will help further confirm the adequacy of oxygenation and ventilation as the resuscitation proceeds.

Establishment of Circulation Access

Access to the patient's venous circulation is mandatory; such

Table 5 Confirmation of Oral Endotracheal Tube Placement

Intubation process
Vocal cords are visualized by intubator
Tip of ET tube is seen passing between the cords
Cuff of ET tube also passes cords by 1 cm
Postintubation checks
Esophageal detector device or end-tidal CO_2 detector confirms ET tube placement in trachea
Breath sounds are symmetrical (auscultate over lateral anterior chest and in midaxillary line bilaterally)
No gurgling heard with auscultation over epigastrium
Patient's chest rises and falls appropriately with ventilation
ET tube depth is appropriate: 21 cm at the corner of the mouth in women, 23 cm in men
Secure the ET tube to prevent dislodgment
Reassess the adequacy of oxygenation and ventilation throughout the resuscitation (bedside patient assessment; also obtain ABGs when feasible)
Postresuscitation, obtain a portable chest radiograph

ABG—arterial blood gas ET—endotracheal

access may be achieved by a code-team member or members simultaneously while other resuscitators pursue steps A and B of the secondary survey. Ideally, a large intravenous cannula is placed in a prominent upper-extremity vein or the external jugular vein to optimize delivery of needed medications. If a peripheral line is not achievable, additional access possibilities include central line placement via the internal jugular, subclavian (via the supraclavicular approach), or, less ideally, femoral vein; even intraosseous access is possible (intraosseous access is a common emergency vascular access site in pediatric patients, but it is an unusual route of access in adults). It is useful to remember, as already noted, that some important resuscitation medications can be delivered via the ET tube in cases of failed intravenous access; such medications include naloxone, atropine, vasopressin, epinephrine, and lidocaine (mnemonic: NAVEL).

The commonly used medications in cardiac resuscitation may be grouped into the following general categories: vasopressors (epinephrine or vasopressin), antiarrhythmics (amiodarone, lidocaine, magnesium, and procainamide), anticholinergic agents (atropine, if the arrest arrhythmia is asystole or PEA is slow), and miscellaneous drugs used to treat specific problems contributing to the arrest state, such as sodium bicarbonate (for severe metabolic acidosis, hyperkalemia, and certain drug overdoses), and calcium chloride (for hyperkalemia, calcium channel blocker drug overdose, or severe hypocalcemia) [*see Table 4*].

Persons in cardiac arrest (which can result from pulseless VT, VF, PEA, or asystole) require a vasopressor for as long as they remain pulseless. Typically, this consists of 1 mg of epinephrine intravenously every 3 to 5 minutes. Epinephrine stimulates adrenergic receptors, which leads to vasoconstriction and optimization of CPR-generated blood flow to the heart and brain. Vasopressin (40 units I.V. once only) is a reasonable alternative to epinephrine, at least initially. Vasopressin in the recommended dose is a potent vasoconstrictor. It also has the theoretical advantage over epinephrine of not increasing myocardial oxygen consumption or lactate production in the arrested heart.[66] Despite its potential advantages, however, in a study of 200 inpatient cardiac arrest patients, vasopressin did not provide a better survival rate than epinephrine.[67] Vasopressin was also found to be comparable to epinephrine in out-of-hospital cardiac arrests when the rhythm was VF or PEA but superior to epinephrine for patients in asystole.[68,69]

During resuscitation with ongoing CPR, medication delivery through an intravenous cannula needs to be followed by a 20 ml saline bolus; if the cannula is in a peripheral vein, the extremity containing the cannula should then be elevated for 10 to 15 seconds to augment delivery of the medication to the central circulation. This is especially important because of the low-flow circulatory state with closed-chest CPR.

Differential Diagnosis and Definitive Care

The most challenging part of the secondary survey, as well as cardiac resuscitation management in general, is the problem-solving required when spontaneous circulation does not return despite appropriate initial interventions. This situation poses a critical question to the resuscitators: Why is this patient dying right now? The intellectual challenge of that question, which the resuscitators must try to answer expeditiously and at the bedside, is compounded by the emotional intensity that pervades most cardiac resuscitations.

The solvable problems that may interfere with resuscitation can be grouped into three broad categories: technical [*see Table 6*], physiologic, and anatomic [*see Table 7*]. Technical problems consist of difficulties with the resuscitators' equipment or skills; such difficulties include ineffective CPR, inadequate oxygenation and ventilation, ET tube complications, intravenous access difficulties, and monitor-defibrillator malfunction or misuse. The physiologic and anatomic problems consist of life-threatening but potentially treatable conditions that may have led to the cardiac arrest in the first place. This differentiation between physiology and anatomy is admittedly artificial, given that physiology is always involved in a cardiac arrest, but it has some usefulness as a teaching and problem-solving tool. Physiologic problems classically include hypoxia, acidosis, hyperkalemia, severe hypokalemia, hypothermia, hypoglycemia, and drug overdose. Anatomic problems are hypovolemia/hemorrhage, tension pneumothorax, cardiac tamponade, myocardial infarction, and pulmonary embolism.[41]

Whenever possible, the patient's medical and surgical history and the circumstances and symptoms immediately before the cardiac arrest should be sought from family members, bystanders, or hospital staff as the resuscitation proceeds. This information may contain important clues to the principal arrest problem and how it may be expeditiously treated. For example, a patient who presents to an emergency department with chest pain and then suffers a VF cardiac arrest is probably dying of a massive myocardial infarction, pulmonary embolism, or aortic dissection, with tension pneumothorax or cardiac tamponade also being possibilities.

Specific questions to consider include the following: Does the patient have risk factors for heart disease, pulmonary embolism, or aortic disease? What was the quality of the patient's pain and its radiation before the cardiac arrest? What were the prearrest vital signs and physical examination findings? What did the prearrest ECG show (if available)? Can any of this information be used now, at the bedside, to dictate the needed resuscitation interventions during the D phase of the secondary survey? For example, if the prearrest ECG showed prominent ST segment elevation in leads V1 through V4 consistent with a large anterior myocardial infarction, if the patient's resuscitation is failing de-

Table 6 Technical Problems That May Prevent a Successful Resuscitation

Problem	*Patients at Risk*	*Recommendations*
Ineffective CPR	All cardiac arrest patients	Ensure technically perfect CPR. Confirm carotid pulses with CPR. If arterial line was in place before cardiac arrest, confirm adequate arterial waveform with CPR on arterial line monitor. Monitor end-tidal CO_2 if available (higher levels correlate with better CPR and improved patient survival). Confirm adequate oxygenation with an ABG when feasible.
Inadequate oxygenation and ventilation	All cardiac arrest patients	Ensure optimal airway positioning and control. Have suction immediately available to manage pharyngeal and airway secretions. Ensure use of properly fitting, tightly sealed face mask for bag-valve mask (BVM) ventilation until a definitive airway is established. Apply cricoid pressure to prevent gastric distention during BVM ventilation until a definitive airway is established. Ensure that supplemental oxygen is flowing to BVM at 15 L/min. Deliver an appropriate tidal volume per breath (6–7 ml/kg if oxygen is available) at the rate of 12–15 breaths/min. Confirm bilateral and equal breath sounds with ventilation. Confirm that patient's chest rises with each ventilation. Allow adequate time for exhalation between breaths. Confirm optimal oxygenation and ventilation with an ABG when feasible.
ET tube difficulties	All patients intubated with ET tube	Allow ≤ 20–30 sec/intubation attempt. Intubator should see tip of ET tube and cuff pass between vocal cords at time of intubation. After intubation, immediately confirm correct ET tube placement; regularly reconfirm ET tube placement throughout resuscitation. Confirm adequacy of oxygenation and ventilation with an ABG. After intubation, consider nasogastric tube placement to decompress stomach and optimize diaphragmatic excursions with ventilation.
Intravenous line difficulties	All cardiac arrest patients	Place one or more 18-gauge or larger I.V. cannulas in an antecubital or external jugular vein site. Check for I.V. infiltration regularly throughout the resuscitation. Follow all medications administered through a peripheral I.V. site with a 20 ml saline bolus and elevation of the extremity containing the I.V. for 10–15 sec (if possible). Consider central line placement if the resuscitation is prolonged. Be aware of all I.V. infusions the patient is receiving. Stop all nonessential medications that had been started before the cardiac arrest. During the resuscitation, the only infusions the patient should receive are normal saline, blood products (if clinically indicated), and pertinent medications necessary to assist with return of spontaneous circulation. Pulmonary artery catheters and central lines occasionally act as an arrhythmogenic focus within the right ventricle. If applicable, deflate all relevant balloons on the catheter and withdraw the catheter to a superior vena cava position.
Monitor-defibrillator difficulties	All cardiac arrest patients	Make sure Synchronization Mode button is in the off position when defibrillating patients in pulseless VT or VF. Make sure electricity is not arcing over the patient's chest because of perspiration or smeared conducting gel; dry patient's chest with a towel except for areas directly beneath pads or paddles. Do not administer shock through nitroglycerin paste or patches. If the patient has an internal cardioverter-defibrillator (ICD) or a pacemaker, the patient may still be manually defibrillated, but do not shock directly over the internal device. Under these circumstances, place the pads or paddles at least 1 in. away from the patient's internal device. If the ICD is intermittently firing but not defibrillating the patient and if the ICD is thought to be compromising the resuscitation, turn the device off with a magnet so that manual defibrillation may take place without interference. Maximize the gain or electrocardiography "size" and check the rhythm in several leads (or change the axes of the paddles if reading the rhythm in Paddles mode) to confirm asystole when the initial rhythm appears to be asystole.

ABG—arterial blood gas ET—endotracheal VF—ventricular fibrillation VT—ventricular tachycardia

spite appropriate interventions, and if there appear to be no technical problems hampering the resuscitation, a working diagnosis of massive myocardial infarction can be made; intravenous thrombolytic therapy may then be a reasonable and needed step in such a resuscitation.[70]

Thoughtful consideration of the possible reasons why a resuscitation is failing will regularly push the code-team captain's and resuscitation team's expertise and clinical skills to the limits. Nevertheless, the failure to consider these formidable issues will deprive the patient of an optimal opportunity to survive the cardiac arrest.

Cardiac Resuscitation Based on Rhythm Findings

When a monitor-defibrillator arrives at the scene of a cardiac arrest, the patient's rhythm is immediately analyzed.

Table 7 Potentially Treatable Conditions That May Cause or Contribute to Cardiac Arrest[3]

Condition	*Clinical Setting*	*Diagnostic and Corrective Actions*
Acidosis	Preexisting acidosis, diabetes, diarrhea, drugs, toxins, prolonged resuscitation, renal disease, shock	Obtain stat ABG. Reassess technical quality of CPR, oxygenation, and ventilation. Confirm correct endotracheal tube placement. Hyperventilate patient (P_aCO_2 of 30–35 mm Hg) to partially compensate for metabolic acidosis. If pH < 7.20 despite above interventions, consider I.V. sodium bicarbonate, 1 mEq/kg I.V. slow push.
Cardiac tamponade	Hemorrhagic diathesis, malignancy, pericarditis, postcardiac surgery, postmyocardial infarction, trauma	Initiate large-volume I.V. crystalloid resuscitation. Confirm diagnosis with emergent bedside echocardiogram, if available. Perform pericardiocentesis. Immediate surgical intervention is appropriate if pericardiocentesis is unhelpful but cardiac tamponade is known or highly suspected clinically.
Hypoglycemia	Adrenal insufficiency, alcohol abuse, aspirin overdose, diabetes, drugs, toxins, liver disease, renal disease, sepsis, certain tumors	Consider clinical setting and obtain finger-stick glucose or stat blood glucose measurements (may be obtained on ABG specimen). If glucose < 60 mg/dl, treat: 50 ml = 25 g of D50W I.V. Follow glucose levels closely posttreatment.*
Hypomagnesemia	Alcohol abuse, burns, diabetic ketoacidosis, severe diarrhea, diuretics, drugs (e.g., cisplatin, cyclosporine, pentamidine), malabsorption, poor intake, thyrotoxicosis	Obtain stat serum magnesium level. Treat: 1–2 g magnesium sulfate I.V. over 2 min. Follow magnesium levels over time, because blood levels correlate poorly with total body deficit.
Hypothermia	Alcohol abuse, burns, central nervous system disease, debilitated and elderly patients, drowning, drugs, toxins, endocrine disease, exposure history, homelessness, poverty, extensive skin disease, spinal cord disease, trauma	Obtain core body temperature. If severe hypothermia (< 30° C), limit initial shocks for pulseless VT/VF to three, initiate active internal rewarming and cardiopulmonary support, and hold further resuscitation medications or shocks until core temperature > 30° C.† If moderate hypothermia (30°–34° C), proceed with resuscitation (space medications at intervals greater than usual), passively rewarm, and actively rewarm truncal body areas.
Hypovolemia, hemorrhage, anemia	Major burns, diabetes, gastrointestinal losses, hemorrhage, hemorrhagic diathesis, malignancy, pregnancy, shock, trauma	Initiate large-volume I.V. crystalloid resuscitation. Obtain stat hemoglobin level on ABG specimen. Emergently transfuse packed red blood cells (O negative if type-specific blood not available) if hemorrhage or profound anemia is contributing to arrest. Emergently consult necessary specialty for definitive care. Emergent thoracotomy with open cardiac massage is a consideration if experienced providers are available for the patient with penetrating trauma and cardiac arrest.
Hypoxia	All cardiac arrest patients are at risk	Reassess technical quality of CPR, oxygenation, and ventilation. Confirm correct ET tube placement. Obtain stat ABG to confirm adequate oxygenation and ventilation.
Myocardial infarction	Consider in all cardiac arrest patients, especially those with risk factors for coronary artery disease, a history of ischemic heart disease, or prearrest picture consistent with an acute coronary syndrome	Review prearrest clinical presentation and ECG. Continue resuscitation algorithm; proceed with definitive care as appropriate for the immediate circumstances (e.g., thrombolytic therapy, cardiac catheterization/coronary artery reperfusion, circulatory assist device, emergent cardiopulmonary bypass).

This step constitutes the beginning part of the defibrillation stage, or step D, of the AHA's primary survey). There are four rhythm possibilities [*see Figure 1*]: (1) pulseless VT; (2) VF; (3) organized or semiorganized electrical activity despite the absence of a palpable carotid pulse, which defines PEA; and (4) asystole. The detailed management of these different cardiac resuscitation scenarios is based on the recommendations of the AHA[49] and the International Liaison Committee on Resuscitation.[71] In following these guidelines, the clinician should remember that, with the exception of early CPR and early defibrillation for VF and pulseless VT, many of the recommendations that form the foundation of modern resuscitation are evidence supported or consensus based (rather than evidence based, as would be ideal). Because of the nature of cardiac arrest and the multiple variables involved, it is exceptionally difficult to perform high-quality research in cardiac resuscitation.

PULSELESS VENTRICULAR TACHYCARDIA OR VENTRICULAR FIBRILLATION

The appearance of either VF or pulseless VT on the rhythm monitor in a patient with ongoing CPR is a relatively favorable finding, because there is reasonable hope for a successful outcome with these rhythms. In addition, the interventions and medications sequentially used in the resuscitation are plainly delineated, and the initial course of action is clear. VF and pulseless VT are managed identically.

Initiation of Defibrillation

Defibrillation with 200 joules should be attempted immediately. However, if the time from onset of arrest to CPR to the availability of defibrillation is estimated to be longer than 5 minutes, it may be reasonable to continue CPR for another 3 minutes before initiating defibrillation [*see* Defibrillation, *above*]. If the VF or VT persists after the initial shock, subsequent attempts

Table 7 (*continued*)

Condition	*Clinical Setting*	*Diagnostic and Corrective Actions*
Poisoning	Alcohol abuse, bizarre or puzzling behavioral or metabolic presentation, classic toxic syndrome, occupational or industrial exposures, history of ingestion, polysubstance abuse, psychiatric disease	Consider clinical setting and presentation; provide meticulous supportive care. Emergently consult toxicologist (through regional poison center) for resuscitative and definitive care advice, including appropriate antidote use. Prolonged resuscitation efforts are appropriate. If available, immediate cardiopulmonary bypass should be considered.
Hyperkalemia	Metabolic acidosis, excessive administration, drugs and toxins, vigorous exercise, hemolysis, renal disease, rhabdomyolysis, tumor lysis syndrome, significant tissue injury	Obtain stat serum potassium level on ABG specimen. Treatment: calcium chloride 10% (5–10 ml I.V. slow push [do not use if hyperkalemia is secondary to digitalis poisoning]), followed by glucose and insulin (50 ml of D50W and 10 U regular insulin I.V.); sodium bicarbonate (50 mEq I.V.); albuterol (15–20 mg nebulized or 0.5 mg I.V. infusion).‡
Hypokalemia	Alcohol abuse, diabetes, diuretic use, drugs and toxins, profound gastrointestinal losses, hypomagnesemia, excess mineralocorticoid states, metabolic alkalosis	Obtain stat serum potassium level on ABG specimen. If profound hypokalemia (K^+ < 2–2.5 mEq/L) is contributing to cardiac arrest, initiate urgent I.V. replacement (2 mEq/min I.V. for 10–15 mEq) then reassess.§
Pulmonary embolism	Hospitalized patients, recent surgical procedure, peripartum, known risk factors for venous thromboembolism (VTE), history of VTE, prearrest presentation consistent with acute pulmonary embolism	Review prearrest clinical presentation; initiate appropriate volume resuscitation with I.V. crystalloid and augment with vasopressors as necessary. Attempt emergent confirmation of diagnosis, depending on availability and clinical circumstances; consider emergent cardiopulmonary bypass to maintain patient viability. Continue resuscitation algorithm; proceed with definitive care (thrombolytic therapy, embolectomy via interventional radiology, or surgical thrombectomy) as appropriate for immediate circumstances and availability.
Tension pneumothorax	Post–central line placement, mechanical ventilation, pulmonary disease (including asthma, COPD, necrotizing pneumonia), postthoracentesis, trauma	Consider risks and clinical presentation (prearrest history, breath sounds, neck veins, tracheal deviation). Proceed with emergent needle decompression, followed by chest tube insertion.

*Unrecognized hypoglycemia can cause significant neurologic injury and can be life threatening, but caution with I.V. glucose is appropriate in the setting of cardiac arrest. Available evidence indicates that hyperglycemia may contribute to impaired neurologic recovery in cardiac arrest survivors.

†Active internal or core rewarming includes warm (42°–46° C) humidified oxygen delivered through the endotracheal tube; warm I.V. fluids; peritoneal lavage; esophageal rewarming tubes; bladder lavage; and extracorporeal rewarming if immediately available. Active external rewarming includes warming beds, hot-water bottles, heating pads, and radiant heat sources applied externally to the patient.

‡Glucose is not necessary initially if patient is already hyperglycemic, but glucose levels should be followed closely after administration of I.V. insulin because of the risk of hypoglycemia (especially in patients with renal failure, because of the long duration of action of I.V. insulin in such patients). Sodium bicarbonate is most helpful in patients with concomitant metabolic acidosis; it is less effective in lowering serum potassium in dialysis-dependent renal failure patients. High-dose nebulized albuterol should lower serum potassium by 0.5 to 1.5 mEq/L within 30 to 60 min, but administration during cardiac arrest may be difficult.

§In a non–cardiac arrest situation, usual I.V. potassium replacement guidelines for patients requiring parenteral therapy are generally 10 to 20 mEq/hr with continuous electrocardiographic monitoring. If profound hypokalemia is contributing to cardiac arrest, however, these usual replacement rates are not timely enough, given the critical nature of the situation. Under these circumstances, potassium chloride, 2 mEq/min I.V. for 10 to 15 mEq, is reasonable, but reassessment and careful attention to changing levels, redistribution, and ongoing clinical circumstances are essential to prevent life-threatening hyperkalemia from developing.

ABG—arterial blood gas COPD—chronic obstructive pulmonary disease D50W—50% dextrose in water ET—endotracheal VF—ventricular fibrillation VT—ventricular tachycardia

should be made with 200 to 300 joules and then 360 joules [*see Figure 2*].

A lower, nonescalating equivalent biphasic energy level is acceptable, if the defibrillator offers this option. Manually checking the patient's carotid pulse between shocks is no longer recommended, but the displayed rhythm on the monitor must be carefully assessed after each defibrillation attempt. If there are any doubts concerning the rhythm or if there is suspicion of a dysfunctional lead or paddle cable, then a manual pulse check would be appropriate. If VF or pulseless VT persists, CPR is resumed, the patient endotracheally intubated, correct ET tube placement confirmed, and the tube secured. Simultaneously, intravenous access should be established.

Initiation of Drug Therapy

In patients with ongoing VF, drug therapy begins with the administration of a vasoconstrictor (either epinephrine or vasopressin) [*see Table 4*]. If there is no intravenous access, the drug can be given endotracheally. After each intravenous dose, drug delivery is followed by a 20 ml saline bolus and the extremity containing the I.V. line is elevated. Rescuers continue CPR for 30 to 60 seconds to allow the drug to reach the heart, then attempt defibrillation again with one to three shocks at 360 joules each. As long as the patient remains pulseless, epinephrine is administered every 3 to 5 minutes, with each dose followed by one to three attempts at defibrillation. When vasopressin is the chosen initial drug, only a single dose is given; if the resuscitation continues 10 minutes or longer after vasopressin is administered, epinephrine should be substituted for vasopressin for the remainder of the code. If VF or pulseless VT persists despite the initial administration of a vasoconstrictor and repeated defibrillation attempts, parenteral antiarrhythmic drug therapy is added; amiodarone or lidocaine is an appropriate agent [*see* Choice of Antiarrhythmic Drugs, *below*]. Throughout all of these steps, the code-team leader is also actively looking for and correcting any technical and physiologic or anatomic problems that may be preventing a successful resuscitation [*see Tables 6 and 7*].

Emergency Laboratory Tests

If spontaneous circulation does not return after the first round of antiarrhythmic drug therapy, the resuscitation team must also endeavor to identify and treat the clinically relevant conditions

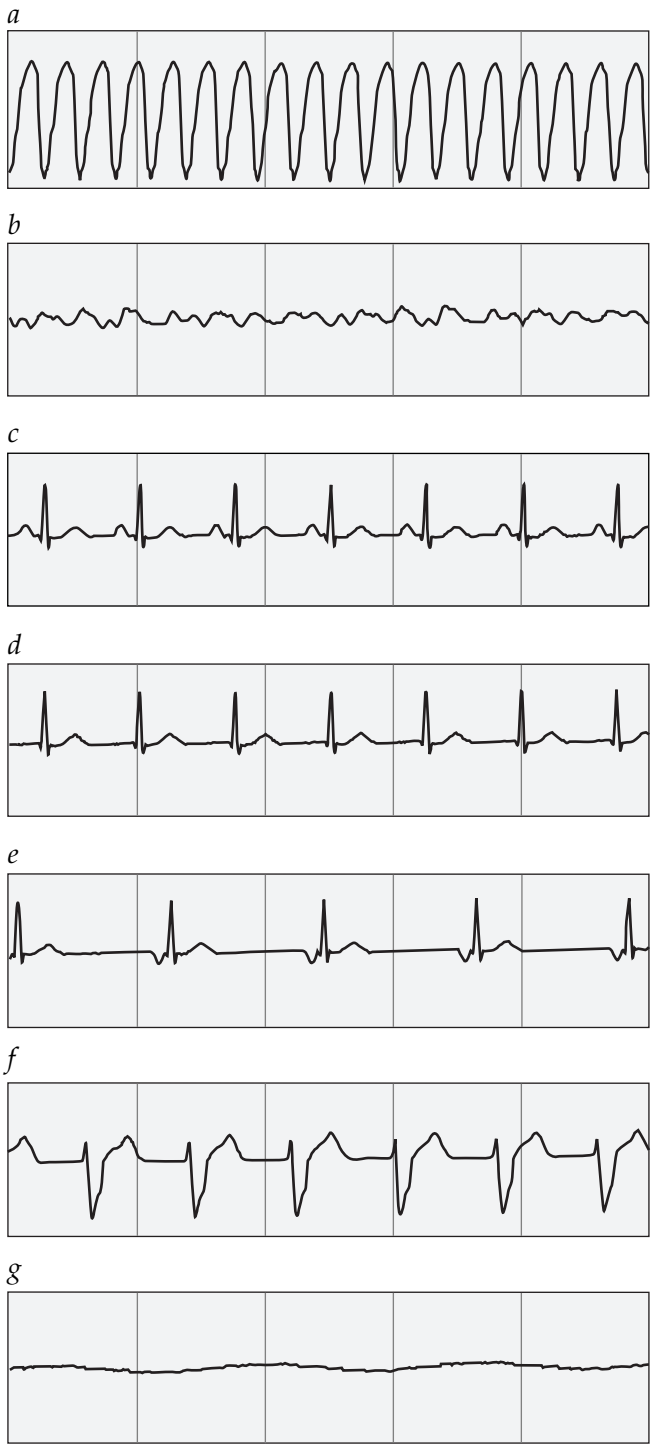


Figure 1 **The sudden cardiac arrest arrhythmias. (*a*) Ventricular tachycardia. (*b*) Ventricular fibrillation. Pulseless electrical activity encompasses any of several forms of organized electrical activity in the pulseless patient; these include (*c*) normal sinus rhythm, (*d*) junctional rhythm, (*e*) bradycardic junctional rhythm, and (*f*) idioventricular rhythm. (*g*) Asystole.**

causing or contributing to the cardiac arrest [*see Table 7*]. In theory, the interventions conducted to this point should have resulted in a perfusing rhythm. The code team must ask why this has not occurred and then attempt to answer this question as the resuscitation continues. Emergency laboratory studies that may prove helpful include a stat ABG measurement and measurements of hemoglobin, potassium, magnesium, and blood glucose levels (most of which can be obtained from the ABG specimen).

Choice of Antiarrhythmic Drugs

Four antiarrhythmic drugs are used in cardiac resuscitation: amiodarone, lidocaine, magnesium (if the patient is thought or proved to be hypomagnesemic), and procainamide (for intermittent or recurrent VT or VF that initially responds to defibrillation).[72] It is not known which one of these drugs or which combination of them will optimize the chances of patient survival to hospital discharge. Despite many years of routine use, there are no controlled studies demonstrating a survival benefit with lidocaine, versus placebo, in the management of VF or pulseless VT. Two studies in patients with shock-refractory prehospital VF showed that survival to hospital admission was better with amiodarone than with placebo (44% versus 34%; $P = 0.03$)[73] or with lidocaine (22.8% versus 12.0%; $P = 0.009$).[74] Neither of these studies demonstrated an improved survival to hospital discharge in the amiodarone groups, but neither study had the statistical power to demonstrate such a difference. Amiodarone is also considerably more expensive than lidocaine.

The optimal role and the exact benefit of antiarrhythmic medications in cardiac resuscitation are yet to be fully elucidated. According to AHA guidelines, either amiodarone or lidocaine is an acceptable initial antiarrhythmic drug for the treatment of patients with VF or pulseless VT that is unresponsive to initial shocks, CPR, airway management, and administration of epinephrine or vasopressin plus shocks. On the basis of available evidence, however, amiodarone may be the antiarrhythmic agent of first choice in the setting of prehospital refractory VF, allowing for optimal survival to hospital arrival.[72-74]

PULSELESS ELECTRICAL ACTIVITY

Community ACLS providers are encountering nonventricular arrhythmias (i.e., PEA and asystole) with increasing frequency. Classically, the prognosis for PEA has been poor, with outpatient survival rates generally reported as 0% to 7%.[75,76] The sequence of resuscitation steps in the management of PEA is as follows: activation of the emergency medical or code response, primary survey (CPR and rhythm evaluation), and secondary survey (intubation and confirmation of correct ET tube placement, optimal oxygenation and ventilation, establishment of I.V. access, epinephrine administration, and, finally, problem solving for technical difficulties and establishment of the cause of the cardiac arrest) [*see Figure 2*]. The two core drugs for PEA management are epinephrine (repeated every 3 to 5 minutes for as long as the patient is pulseless) and atropine (up to 3 mg over time if the PEA rhythm on the monitor is inappropriately slow). Although not currently on the AHA PEA algorithm, vasopressin is probably a reasonable alternative to epinephrine. The best hope for a successful resuscitation is to find and treat the cause of PEA; therein lies the exceptionally challenging aspect of PEA resuscitation management [*see Tables 6 and 7*]. Because coronary artery thrombosis and pulmonary thromboembolism are common causes of cardiac arrest, a trial evaluated the efficacy of tissue plasminogen activator (t-PA) in the setting of PEA of unknown or presumed cardiovascular cause in 233 patients in prehospital and emergency department settings.[77] No benefit was found with thrombolytic therapy for PEA in this study; the proportion of patients with return of spontaneous circulation was 21.4% in the t-PA group and 23.3% in the placebo group.

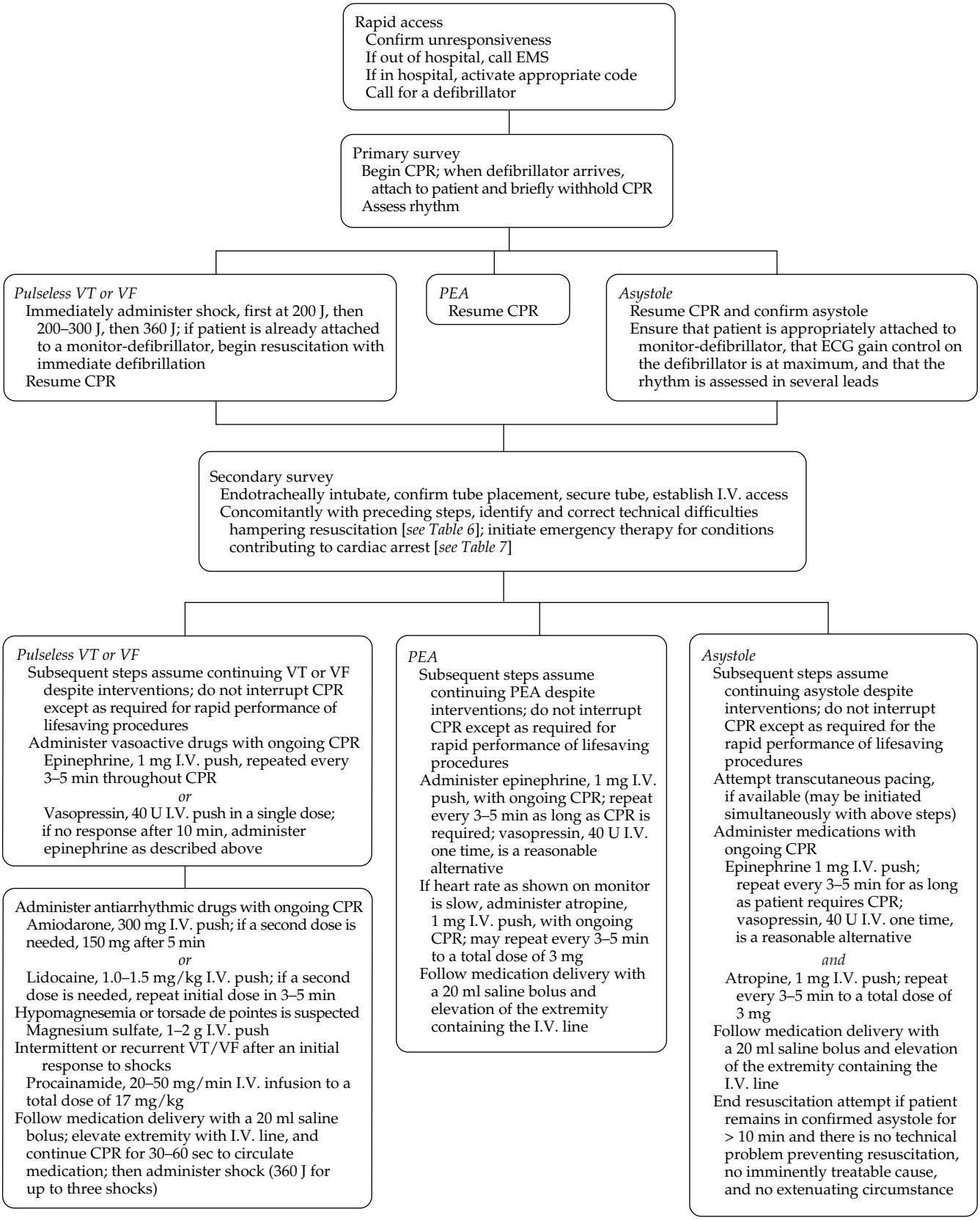


Figure 2 **Treatment algorithm for patients with VT, VF, PEA, or asystole. (PEA—pulseless electrical activity; VF—ventricular fibrillation; VT—ventricular tachycardia)**

ASYSTOLE

The prognosis for asystole is generally regarded as dismal unless the patient is hypothermic or there are other extenuating but treatable circumstances. The sequence of resuscitation steps in the management of asystole is as follows: activation of the emergency medical or code response, primary survey (CPR, rhythm evaluation, and asystole confirmation), and secondary survey (intubation and confirmation of correct ET tube placement, optimal oxygenation and ventilation, I.V. access with epinephrine and atropine administration, immediate transcutaneous pacing, if available, and problem solving for technical difficulties and establishment of the cause of cardiac arrest) [*see Figure 2*]. The two core drugs for asystole management are epinephrine (repeated every 3 to 5 minutes for as long as the patient is pulseless) and atropine (up to 3 mg over time). As with PEA, vasopressin appears to be a reasonable and possibly beneficial substitute for epinephrine in asystole. A single dose of aminophylline (250 mg I.V.) may also be beneficial in atropine-resistant asystole.[78] Potentially treatable causes of asystole include hypoxia, acidosis, hypothermia, hypokalemia, hyperkalemia, and drug overdose. Resuscitation efforts should stop if asystole persists for longer than 10 minutes despite optimal CPR, oxygenation and ventilation, and epinephrine or atropine administration; if extenuating circumstances (e.g., hypothermia, cold-water submersion, or drug overdose) are not present; and if no other readily treatable condition is identified.

Immediate Postresuscitation Care

Even when the resuscitation is successful, the patient's situation remains tenuous and continued meticulous patient care is essential. When the cardiac monitor indicates what should be a perfusing rhythm, the rescuer should immediately confirm that the patient has a palpable pulse. If there is a pulse, the patient's blood pressure is then obtained. Simultaneously, resuscitation team members need to quickly reassess the adequacy of the patient's airway, the ET tube position, oxygenation and ventilation, and the patient's level of consciousness and comfort.

If the patient is hypotensive, appropriate blood pressure management depends on the presence or absence of fluid overload, as judged at the bedside. If the patient is clinically volume overloaded or in frank pulmonary edema and hypotensive, dopamine is started at inotropic doses (5 μg/kg/min I.V.) and titrated to a target systolic blood pressure of 90 to 100 mm Hg. If the patient's clinical status suggests normovolemia or hypovolemia, intravenous crystalloid boluses (in 250 to 500 ml increments) can be administered instead of dopamine to support adequate tissue perfusion. In patients who are regaining consciousness, their level of comfort mandates careful assessment and administration of analgesia and sedation, as appropriate.

If the arrest rhythm was either VT or VF, the parenteral antiarrhythmic drug used immediately before the return of spontaneous circulation is continued as a maintenance infusion (amiodarone, 1 mg/min for 6 hr, then 0.5 mg/min for 18 hr as blood pressure allows; or lidocaine, 2 to 4 mg/min). If an antiarrhythmic drug has not yet been administered, it is usually started at this point to prevent the recurrence of VF or pulseless VT. There are important exceptions to this guideline, however. If the perfusing postarrest arrhythmia is an idioventricular rhythm or third-degree heart block accompanied by an idioventricular escape rhythm, an antiarrhythmic medication should not be started at this time, because the antiarrhythmic agent could eliminate the ventricular perfusing focus and return the patient to a pulseless state.

Initial postresuscitation studies usually include an ECG; portable chest radiography; and measurement of ABGs, a serum electrolyte panel, fingerstick or blood glucose, serum magnesium and cardiac enzyme levels, and hemoglobin and hematocrit. The resuscitated patient requires urgent transfer to the optimal site for continued definitive care. Depending on the circumstances, this may be either the cardiac catheterization laboratory or the intensive care unit.

Ongoing research continues to look at optimal postresuscitation management strategies to improve neurologic outcome and survival to hospital discharge.[79] Hyperthermia and hyperglycemia compromise postresuscitation neurologic outcome, whereas mild to moderate induced hypothermia appears to improve neurologic outcome and decrease mortality.[80-83]

Ending a Resuscitation Attempt

Throughout the resuscitation, the team leader must speak with calmness and authority, and all resuscitations should be orchestrated with clarity and finesse. If possible, the code captain should make clinical decisions without directly performing specific procedures. Cardiac arrests are emotionally charged, but the leader must insist on a composed, orderly, and technically sound resuscitation. It is appropriate to invite suggestions from team members and to ensure that all members are comfortable with the decision to stop the resuscitation, should that time arrive.

The decision whether to stop a cardiac resuscitation is burdensome. Clearly, the circumstances of the event, patient comorbidities, the nature of the lethal arrhythmia, and the resuscitation team's ability to correctly identify and treat potential contributing causes of the arrest are all important considerations. Resuscitation efforts beyond 30 minutes without a return of spontaneous circulation are usually futile unless the cardiac arrest is confounded by intermittent or recurrent VF or pulseless VT, hypothermia, cold-water submersion, drug overdose, or other identified and readily treated contributing conditions.[84,85]

With nontraumatic cardiac arrest in the prehospital setting (assuming proper equipment and medications are available and no extenuating circumstances suggest otherwise), full resuscitation efforts take place at the scene of the arrest in preference to rapid transport to an emergency department. A prehospital resuscitation that has been appropriately conducted but has not resulted in at least temporary return of spontaneous circulation to the patient may be discontinued. It is important that certain criteria are adhered to, however, including the following: high-quality CPR provided, an adequate airway successfully placed, appropriate oxygenation and ventilation delivered, intravenous access established, appropriate medications specific to the arrest scenario administered, and resuscitation attempted for at least 10 minutes; in addition, the patient must not be in persistent VF, and there can be no extenuating circumstances that mandate in-hospital continuation of the resuscitation (e.g., hypothermia, drug overdose). The decision whether to cease resuscitation efforts in the field is bolstered by direct discussion with EMS physicians. It is also essential that social services be available to provide immediate assistance and support to the family and loved ones of the patient who has now died.

Discontinuing in-hospital resuscitations is advisable when three criteria are met: (1) the arrest was unwitnessed, (2) the ini-

tial rhythm was other than VF or VT, and (3) spontaneous circulation does not return after 10 minutes of ongoing resuscitation.[86] In a study of this three-component decision rule, only 1.1% of patients (three out of 269) who met these criteria survived to hospital discharge, and none of the three survivors were capable of independent living.[87] In a study of 445 prospectively recorded resuscitation attempts in hospitalized patients, no patient survived who suffered a cardiac arrest between 12 A.M. and 6 A.M. if the arrest was unwitnessed and if it occurred in an unmonitored bed.[38]

A resuscitation attempt in a persistently asystolic patient should not last longer than 10 minutes, assuming all of the following conditions apply: asystole is confirmed through proper rhythm monitoring and assessment; high-quality CPR is taking place; ET intubation is correctly performed and confirmed; adequate oxygenation and ventilation are provided; intravenous access is secured; appropriate medications (epinephrine or vasopressin and atropine) have been administered; and the patient is not the victim of hypothermia, cold-water submersion, drug overdose, or other readily identified and reversible cause.

After all resuscitation attempts, the code-team captain should debrief the team so that all may learn from the experience. Finally, marked empathy and skill are needed to carefully and compassionately inform family members about the outcome of the resuscitation.[88]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Callans DJ: Management of the patient who has been resuscitated from sudden cardiac death. Circulation 105:2704, 2002

2. Zipes DP, Wellens HJ: Sudden cardiac death. Circulation 98:2334, 1998

3. Eisenberg MS, Mengert TJ: Cardiac resuscitation. N Engl J Med 344:1304, 2001

4. 1999 Heart and Stroke Statistical Update. American Heart Association, Dallas, 1998

5. Huikuri HV, Castellanos A, Myerburg R: Sudden death due to cardiac arrhythmias. N Engl J Med 345:1473, 2001

6. Ballew KA, Philbrick JT: Causes of variation in reported in-hospital CPR survival: a critical review. Resuscitation 30:203, 1995

7. Myerburg RJ, Castellanos A: Cardiac arrest and sudden cardiac death. Heart Disease: A Textbook of Cardiovascular Medicine. Braunwald E, Ed. WB Saunders Co, Philadelphia, 1997, p 742

8. Osborn LA: Etiology of sudden death. Cardiac Arrest: The Science and Practice of Resuscitation Medicine. Paradis NA, Halperin HR, Nowak RM, Eds. Williams & Wilkins, Philadelphia, 1996, p 243

9. Maron BJ: Sudden death in young athletes. N Engl J Med 349:1064, 2003

10. Cobb LA, Fahrenruch CD, Olsufka M, et al: Changing incidence of out-of-hospital ventricular fibrillation, 1980–2000. JAMA 288:3008, 2002

11. Cummins RO, Ornato JP, Thies W, et al: Improving survival from cardiac arrest: the chain of survival concept: a statement for health professionals from the Advanced Cardiac Life Support Subcommittee and the Emergency Cardiac Care Committee, American Heart Association. Circulation 83:1832, 1991

12. Capussi A, Aschieri D, Piepoli MF, et al: Tripling survival from sudden cardiac arrest via early defibrillation without traditional education in cardiopulmonary resuscitation. Circulation 106:1065, 2002

13. Callaham M, Madsen CD: Relationship of timeliness of paramedic advanced life support interventions to outcome in out-of-hospital cardiac arrest treated by first responders with defibrillators. Ann Emerg Med 27:638, 1996

14. Marenco JP, Wang PJ, Link MS, et al: Improving survival from sudden cardiac arrest: the role of the automated external defibrillator. JAMA 285:1193, 2001

15. Peberdy MA: Defibrillation. Cardiol Clin 20:13, 2002

16. Public-access defibrillation and survival after out-of-hospital cardiac arrest. The Public Access Defibrillation Trial Investigators. N Engl J Med 351:637, 2004

17. Cummins RO, Eisenberg MS: Prehospital cardiopulmonary resuscitation: is it effective? JAMA 253:2408, 1985

18. Stapleton ER: Basic life support cardiopulmonary resuscitation. Cardiol Clin 20:12, 2002

19. Valenzuela TD, Roe DJ, Cretin S, et al: Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model. Circulation 96:3308, 1997

20. Eisenberg MS, Bergner L, Hallstrom A: Cardiac resuscitation in the community: the importance of rapid delivery of care and implications for program planning. JAMA 241:1905, 1979

21. Weaver WD, Cobb LA, Hallstrom AP, et al: Considerations for improving survival from out-of-hospital cardiac arrest. Ann Emerg Med 15:1181, 1986

22. Larsen MP, Eisenberg MS, Cummins RO, et al: Predicting survival from out-of-hospital cardiac arrest: a graphic model. Ann Emerg Med 270:1211, 1993

23. Eisenberg MS, Horwood BT, Cummins RO, et al: Cardiac arrest and resuscitation: a tale of 29 cities. Ann Emerg Med 19:179, 1990

24. Lombardi G, Gallagher J, Gennis P: Outcome of out-of-hospital cardiac arrest in New York City: the Pre-Hospital Arrest Survival Evaluation (PHASE) study. JAMA 271:678, 1994

25. Becker LB, Ostrander MP, Barrett J, et al: Outcome of CPR in a large metropolitan area: where are the survivors? Ann Emerg Med 20:355, 1991

26. Killien SY, Geyman JP, Gossom JB, et al: Out-of-hospital cardiac arrest in a rural area: a 16-year experience with lessons learned and national comparisons. Ann Emerg Med 28:294, 1996

27. Bunch TJ, White RD, Gersh BJ, et al: Long-term outcomes of out-of-hospital cardiac arrest after successful early defibrillation. N Engl J Med 348:2626, 2003

28. Valenzuela TD, Roe DJ, Nichol G, et al: Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. N Engl J Med 343:1206, 2000

29. Eisenberg M, Bergner L, Hallstrom A: Sudden Cardiac Death in the Community. Praeger, Philadelphia, 1984

30. Becker L: The epidemiology of sudden death. Cardiac Arrest: The Science and Practice of Resuscitation Medicine. Paradis NA, Halperin HR, Nowak RM, Eds. Williams & Wilkins, Philadelphia, 1996, p 28

31. Jastremski MS: In-hospital cardiac arrest. Ann Emerg Med 22:113, 1993

32. Rosenberg M, Wang C, Hoffman-Wilde S, et al: Results of cardiopulmonary resuscitation failure to predict survival in two community hospitals. Arch Intern Med 153:1370, 1993

33. Ballew KA, Philbrick JT, Caven DE, et al: Predictors of survival following in-hospital cardiopulmonary resuscitation: a moving target. Arch Intern Med 154:2426, 1994

34. De Vos R, Koster RW, De Haan RJ, et al: In-hospital cardiopulmonary resuscitation: prearrest morbidity and outcome. Arch Intern Med 159:845, 1999

35. Goodlin SJ, Zhong Z, Lynn J, et al: Factors associated with use of cardiopulmonary resuscitation in seriously ill hospitalized adults. JAMA 282:2333, 1999

36. Van Walraven C, Forster AJ, Stiell IG: Derivation of a clinical decision rule for the discontinuation of in-hospital cardiac arrest resuscitations. Arch Intern Med 159:129, 1999

37. Zoch TW, Desbiens NA, DeStefano F, et al: Short- and long-term survival after cardiopulmonary resuscitation. Arch Intern Med 160:1969, 2000

38. Dumot JA, Burval DJ, Sprung J, et al: Outcome of adult cardiopulmonary resuscitations at a tertiary referral center including results of "limited" resuscitations. Arch Intern Med 161:1751, 2001

39. Meade M: Men and the Water of Life. Harper, San Francisco, 1993, p 442

40. Burkle FM Jr, Rice MM: Code organization. Am J Emerg Med 5:235, 1987

41. ACLS Provider Manual. American Heart Association, Dallas, 2001

42. Safar P: On the history of modern resuscitation. Anesthesiol Clin North Am 13:751, 1995

43. Kouwenhoven WB, Jude JR, Knickerbocker GG: Closed-chest cardiac massage. JAMA 173:1064, 1960

44. Jude JR, Kouwenhoven WB, Knickerbocker GG: Cardiac arrest: report of application of external cardiac massage on 118 patients. JAMA 178:1063, 1961

45. Halperin HR: Mechanisms of forward flow during external chest compression. Cardiac Arrest: The Science and Practice of Resuscitation Medicine. Paradis NA, Halperin HR, Nowak RM, Eds. Williams & Wilkins, Philadelphia, 1996, p 252

46. Ornato JP, Peberdy MA: Cardiopulmonary resuscitation. Textbook of Cardiovascular Medicine. Topol EJ, Ed. Lippincott-Raven, Philadelphia, 1998, p 1779

47. Paradis NA, Martin GB, Goetting MG, et al: Simultaneous aortic, jugular bulb, and right atrial pressures during cardiopulmonary resuscitation in humans: insights into mechanisms. Circulation 80:361, 1989

48. Daya MR, Mariani RJ: Out-of-hospital splinting. Clinical Procedures in Emergency Medicine, 3rd ed. Roberts JR, Hedges JR, Eds. WB Saunders Co, Philadelphia 1998, p 1297

49. Guidelines 2000 for cardiopulmonary resuscitation and emergency cardiovascular care: international consensus on science. Circulation 102(suppl I):1, 2000

50. Sanders AB, Kern KB, Berg RA, et al: Survival and neurologic outcome after cardiopulmonary resuscitation with four different chest compression–ventilation ratios. Ann Emerg Med 40:553, 2002

51. Hallstrom A, Cobb L, Johnson E, et al: Cardiopulmonary resuscitation by chest compression alone or with mouth-to-mouth ventilation. N Engl J Med 342:1546, 2000

52. Levine RL, Wayne MA, Miller CC: End-tidal carbon dioxide and outcome of out-of-hospital cardiac arrest. N Engl J Med 337:301, 1997

53. Hedges JR, Greenberg MI: Defibrillation. Clinical Procedures in Emergency Medicine, 3rd ed. Roberts JR, Hedges JR, Eds. WB Saunders Co, Philadelphia, 1998, p 1297

54. Hossack KF, Hartwig R: Cardiac arrest associated with supervised cardiac rehabilitation. J Cardiac Rehab 2:402, 1982

55. Bardy GH, Marchlinski FE, Sharma AD, et al: Multicenter comparison of truncated biphasic shocks and standard damped sine wave monophasic shocks for transthoracic ventricular defibrillation. Circulation 94:2507, 1996

56. Gliner BE, White RD: Electrocardiographic evaluation of defibrillation shocks delivered to out-of-hospital sudden cardiac arrest patients. Resuscitation 41:129, 1999

57. Gliner BE, Jorgenson DB, Poole JE, et al: Treatment of out-of-hospital cardiac arrest with a low-energy impedance-compensating biphasic waveform automatic external defibrillator. Biomed Instrum Technol 32:631, 1998

58. Poole JE, White RD, Kanz KG, et al: Low-energy impedance-compensating biphasic waveforms terminate ventricular fibrillation at high rates in victims of out-of-hospital cardiac arrest. J Cardiovasc Electrophysiol 8:1373, 1997

59. Schneider T, Martens PR, Paschen H, et al: Multicenter, randominzed, controlled trial of 150-J biphasic shocks compared with 200- to 360-J monophasic shocks in the resuscitation of out-of-hospital cardiac arrest victims. Circulation 102:1780, 2000

60. Berg RA, Hilwig RW, Kern KB, et al: Precountershock cardiopulmonary resuscitation improves ventricular fibrillation median frequency and myocardial readiness for successful defibrillation from prolonged ventricular fibrillation: a randomized, controlled swine study. Ann Emerg Med 40:563, 2002

61. Wik L, Hansen TB, Fylling F, et al: Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation. JAMA 289:1389, 2003

62. Weisfeldt ML, Becker LB: Resuscitation after cardiac arrest: a 3-phase time sensitive model. JAMA 288:3035, 2002

63. Aehlert B: ACLS: Quick Review Study Guide, 2nd ed. CV Mosby, St Louis, 2001

64. Rumball CJ, MacDonald D: The PTL, Combitube, laryngeal mask, and oral airway: a randomized prehospital comparative study of ventilatory device effectiveness and cost-effectiveness in 470 cases of cardiorespiratory arrest. Prehosp Emerg Care 1:1, 1997

65. Garnett AR, Ornato JP, Gonzales ER, et al: End-tidal carbon dioxide monitoring during cardiopulmonary resuscitation. JAMA 257:512, 1987

66. Paradis NA, Wenzel V, Southall J: Pressor drugs in the treatment of cardiac arrest. Cardiol Clin 20:61, 2002

67. Stiell IG, Hebert PC, Wells GA, et al: Vasopressin versus epinephrine for inhospital cardiac arrest: a randomized controlled trial. Lancet 358:105, 2001

68. Wenzel V, Krismer AC, Arntz HR, et al. A comparison of vasopressin and epinephrine for out-of-hospital cardiopulmonary resuscitation. N Engl J Med 350:105, 2004

69. McIntryre KM: Vasopressin in asystolic cardiac arrest. N Engl J Med 350:179, 2004

70. Tiffany PA, Schultz M, Stueven H: Bolus thrombolytic infusions during CPR for patients with refractory arrest rhythms: outcome of a case series. Ann Emerg Med 31:124, 1998

71. Cummins RO, Chamberlain DA: Advisory statements of the International Liaison Committee on Resuscitation. Circulation 95:2172, 1997

72. Kudenchuk PJ: Advanced cardiac life support antiarrhythmic drugs. Cardiol Clin 20:79, 2002

73. Kudenchuk PJ, Cobb LA, Copass MK, et al: Amiodarone for resuscitation after out-of-hospital cardiac arrest due to ventricular fibrillation. N Engl J Med 341:871, 1999

74. Dorian P, Cass D, Schwartz B, et al: Amiodarone as compared with lidocaine for shock-resistant ventricular fibrillation. N Engl J Med 346:884, 2002

75. Myerburg RJ, Conde CA, Sung RJ, et al: Clinical, electrophysiologic, and hemodynamic profile of patients resuscitated from prehospital cardiac arrest. Am J Med 68:568, 1980

76. Stratton SJ, Niemann JT: Outcome from out-of-hospital cardiac arrest caused by nonventricular arrhythmias: contribution of successful resuscitation to overall survivorship supports the current practice of initiating out-of-hospital ACLS. Ann Emerg Med 32:448, 1998

77. Abu-Laban RB, Christenson JM, Innes GD, et al: Tissue plasminogen activator in cardiac arrest with pulseless electrical activity. N Engl J Med 346:1522, 2002

78. Mader TJ, Smithline HA, Durkin L, et al: A randomized controlled trial of intravenous aminophylline for atropine-resistant out-of-hospital asystolic cardiac arrest. Acad Emerg Med 10:192, 2003

79. Angelos MG, Menegazzi JJ, Callaway CW: Bench to bedside: resuscitation from prolonged ventricular fibrillation. Acad Emerg Med 8:909, 2001

80. Zeiner A, Holzer M, Sterz F, et al: Hyperthermia after cardiac arrest is associated with an unfavorable neurological outcome. Arch Intern Med 161:2007, 2001

81. Moghissi E: Hospital management of diabetes: beyond the sliding scale. Cleve Clin J Med 71:801, 2004

82. Bernard SA, Gray TW, Buist MD, et al: Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N Engl J Med 346:557, 2002

83. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. The Hypothermia After Cardiac Arrest Study Group. N Engl J Med 346:549, 2002

84. Bonnin MJ, Pepe PE, Kimball KT, et al: Distinct criteria for termination of resuscitation in the out-of-hospital setting. JAMA 270:1457, 1993

85. Kellermann AL, Hackman BB, Somes G: Predicting the outcome of unsuccessful prehospital advanced cardiac life support. JAMA 270:1433, 1993

86. Van Walraven C, Forster AJ, Stiell IG: Derivation of a clinical decision rule for the discontinuation of in-hospital cardiac arrest resuscitations. Arch Intern Med 159:129, 1999

87. Van Walraven C, Forster AJ, Parish DC, et al: Validation of a clinical decision aid to discontinue in-hospital cardiac arrest resuscitations. JAMA 285:1602, 2001

88. Iserson K: Grave Words: Notifying Survivors about Sudden, Unexpected Deaths. Galen Press, Tucson, Arizona, 1999

89. Cummins RO, Field JM, Hazinski MF, et al: ACLS Provider Manual. American Heart Association, Dallas, 2001, p 36

90. Part 1: introduction to the international guidelines 2000 for CPR and ECC: a consensus on science. Circulation 102(8 suppl):I1, 2000

IV PREOPERATIVE ASSESSMENT AND CARE OF THE SURGICAL PATIENT

MICHAEL F. LUBIN, M.D., F.A.C.P.

Surgery is changing, and so too is the preoperative evaluation of surgical patients. Newer surgical procedures (e.g., laparoscopic cholecystectomy), along with many common, well-established procedures, are being done in outpatient facilities and on increasingly older and much sicker patients. Increasing numbers of patients are having elective surgical procedures. The evaluation of a patient who is being considered for surgery has changed a great deal in the years since the first papers on cardiac risk assessment were published in the late 1970s.[1] Despite much progress and new data, expert opinion and clinical judgment are still critical in this area of medicine. This chapter reviews risk assessment and provides current recommendations for the evaluation and management of perioperative medical problems [*see Sidebar* Approach to the Preoperative Consultation].

Goldman and coworkers developed the first systematic method to evaluate surgical patients for potential cardiac complications [*see Table 1*].[1] Their cardiac risk index for noncardiac surgery factored in decompensated heart failure, recent myocardial infarction, and arrhythmias, all of which were found to significantly contribute to the risk of cardiac complications and cardiac death. Impairment of pulmonary, renal, or hepatic function added additional risk; other risk factors included age, certain surgical sites (e.g., abdomen, thorax, and aorta), and emergency procedures. A subsequent study confirmed these risk factors and emphasized that severe and unstable angina added significantly to the risk of cardiac complications and cardiac death [*see Table 2*].[2]

Although these methods could identify groups of patients for whom surgery posed higher risk, they were not very helpful in selecting out individual patients who were at increased risk. Except in the highest risk group, the chances of cardiac complications varied from only 1% to 11%, and the risk of death varied from 0 to 2%[2]; these data demonstrate that a large majority of patients can tolerate surgery without problems, even if they have higher risk scores. Since the original indices were published, other methods to assess risk have been developed that employ a variety of information.[3,4] Although the use of any of these indices yields results that are better than the results that come from chance alone, all of them have a high false positive rate that limits their usefulness.[5]

How then should a physician proceed in evaluating a patient for surgery? Initially, two very important questions that have nothing to do with individual risk assessment need to be asked: First, is the surgical procedure being contemplated important to the patient's health and well-being? Second, what are the intrinsic risks of the anesthesia and proposed surgery?

It is quite clear that in a young patient who has a compound fracture, surgical repair is essential to the patient's health, and the risk of the surgery itself is rather low. On the other hand, extensive liposuction in a patient with advanced heart failure who also has renal failure is unlikely to have an acceptable risk-to-benefit ratio no matter how stable the patient appears to be. Many procedures fall between these two extremes, however, and risk assessment in such cases can be much more challenging. For example, bypass surgery for moderately severe angina in an 80-year-old patient with diabetes may or may not be useful, and the risks of the operation are not low.

Therefore, the decision to proceed with surgery requires the input of all the medical personnel involved. The internist or family physician has a critical role in assessing the findings from the history and physical examination; such an assessment will determine the degree of risk in the contemplated operation and the need for further testing to define risk. The surgeon must evaluate the risk-to-benefit ratio for the proposed procedure. The anesthesiologist needs to evaluate the risks of anesthesia and present this assessment to the patient. Finally, of course, the patient will have to understand the risks of the procedure and have confidence that they are outweighed by the potential benefits.

Approach to the Preoperative Consultation*

Assessment

Cardiac risk

- Use Goldman or Detsky score [*see Tables 1 and 2*] but provide explanation if the score does not reflect the true assessed risk (e.g., patient does not have angina but has many risk factors, including being wheelchair bound)

Pulmonary risk

- Assess as low, medium, or high, using clinical judgment

Other risks

- Hypertension (well or poorly controlled)
- Diabetes mellitus
- Heart failure (uncontrolled)
- Bleeding risk
  - Use of aspirin or other NSAIDs
- DVT risk
  - Assess as low, medium, or high [*see 1:XVIII Venous Thromboembolism*]

Plan

Cardiac risks

- Use routine cardiac monitoring or monitor as if for stable coronary disease; consider perioperative beta blockade

Pulmonary risks

- Use routine pulmonary care or preoperative and postoperative incentive spirometry with or without preoperative and postoperative chest physiotherapy

Hypertensive patients

- Continue hypertensive medications; hold diuretics on day of surgery

Diabetic patients

- Monitor blood glucose and treat with sliding-scale regular insulin, q. 4 hr, or give NPH insulin half dose, then follow with sliding-scale regular insulin

Thrombosis risk

- Hold aspirin for 5 days preoperatively; if DVT prophylaxis is necessary, give low-dose heparin, 5,000 U q. 12 hr, or LMWH

*This sidebar presents an example of possible assessments and plans.
DVT—deep vein thrombosis LMWH—low-molecular-weight heparin
NSAIDs—nonsteroidal anti-inflammatory drugs

Table 1 Calculating Operative Morbidity and Mortality[1]

Information Source	*Criterion*	*Points*
History	Age > 70 yr	5
	MI < 6 mo ago	10
	Chronic liver disease	3
	Bedridden	3
Physical examination	S_3 heart sound or JVD	11
	Significant aortic stenosis	3
Electrocardiogram	Rhythm other than sinus or PACs; > 5 PVCs/min	7
Laboratory tests	P_{O_2} < 60 mm Hg or P_{CO_2} > 50 mm Hg	3
	Serum potassium < 3 mmol/L	3
	Serum bicarbonate < 20 mEq/L	3
	Serum creatinine > 3 mg/dl	3
	BUN > 50 mg/dl	3
	Elevated AST	3
Intended operation	Peritoneal, thoracic, or aortic	3
	Emergency	4

Risk Class (Total Points)	*Morbidity*	*Mortality*
1 (0–5)	0.7%	< 0.2%
2 (6–12)	5%	2%
3 (13–25)	11%	2%
4 (> 25)	22%	56%

AST—aspartate aminotransferase BUN—blood urea nitrogen JVD—jugular venous distention MI—myocardial infarction PAC—premature atrial contraction PVC—premature ventricular contraction

Decisions about proceeding to surgery require good clinical judgment. High-risk patients need individual assessment. Despite high risk, those with life-threatening indications, such as an uncontrollable upper gastrointestinal bleed, will usually proceed to surgery. Less important procedures that are not likely to sustain or enhance life will more likely be postponed to permit treatment of alterable conditions, replaced with a lower-risk procedure, or canceled if the risks appear to be too high and the benefits not high enough.

Clinical Evaluation

In determining surgical risk, the history and physical examination are of great importance. The systems most involved in postoperative morbidity and mortality are the cardiovascular and pulmonary systems.[1,2] It is also essential to evaluate renal, hepatic, and endocrine function.

ASSESSING CARDIOVASCULAR RISK

The most critical parts of the cardiac assessment are those that involve heart failure and coronary artery disease (CAD). Assessment of anginal chest pain is particularly important because proper management of CAD can decrease the risks of surgery. On the other hand, extensive cardiac testing in a patient who is unlikely to have significant CAD will delay procedures that may be important and will increase the cost of care.

Heart Failure

Uncontrolled heart failure is the most important risk factor for cardiac death or complications. A history of functional limitation appears to be the most helpful of all the historical points in this assessment. Patients who can perform activities that require four metabolic equivalents (METs) have a good chance of survival for most surgical procedures [*see Table 3*]; such patients require no further testing. Some authors have used stair climbing to assess patients' functional capacity; the ability to walk up approximately two flights of stairs or to walk four level city blocks has proved helpful.[6] The presence of such symptoms as orthopnea and paroxysmal nocturnal dyspnea in a patient with poor functional capacity increases the likelihood of heart failure but does not have predictive value for perioperative outcome.

The physical examination should identify signs of heart failure, such as distended neck veins, an S_3 gallop, rales, and edema. Other signs of cardiovascular disease, such as hypertension, decreased pulses, or bruits, should be recorded. These signs will help identify patients with ventricular dysfunction and peripheral vascular disease that may increase overall risk.

The use of echocardiography as a predictive tool is controversial. Although many experts advocate echocardiography as a good tool for assessing heart failure control, the procedure may provide little prognostic information beyond that available from a careful history and physical examination.

The most important preoperative use of echocardiography is in the differentiation of systolic dysfunction from diastolic dysfunction in patients with new-onset heart failure. A careful history and physical examination can suggest which is more likely, but the echocardiographic differentiation is significantly more reliable. The distinction is important, because data clearly show that systolic dysfunction, in a patient with substantial clinical manifestations (i.e., overt congestive failure), adds significantly to the risk of surgery.[1] On the other hand, there are no data showing that echocardiographic evidence of systolic dysfunction in a patient without symptoms or signs of heart failure has any prognostic implications.[7]

There are also no good data indicating that diastolic dysfunction increases risk significantly. Although diastolic dysfunction probably adds some degree of risk, patients with echocardiographic findings of diastolic dysfunction but who have no signs and symptoms are probably not at substantially higher risk. Thus, there is no reason to order preoperative echocardiography when the diagnosis of diastolic dysfunction has already been established.

Patients who are able to perform tasks that require more than four METs can undergo most surgery without a significant increase in risk[8]; therefore, such patients would not benefit from an echocardiogram. Additionally, patients who have no clear-cut symptoms or signs of heart failure are unlikely to benefit from the test. Patients who do have symptoms and signs (e.g., elevated neck veins, S_3 gallop, orthopnea, or paroxysmal nocturnal dyspnea) are definitely at increased risk; preoperative echocardiography is unnecessary in such cases because it provides no additional information on risk.[7] Treating the heart failure until the symptoms have stabilized and the signs have improved will lower risk substantially.

Coronary Artery Disease

The preoperative evaluation of the patient with established or probable CAD is of great importance. In general, most patients who will be having cardiac surgery or major vascular

surgery are evaluated by cardiologists. Recent myocardial infarction is second only to decompensated heart failure as a risk factor for perioperative complications.[1] Decisions regarding the evaluation of chest pain in patients without a history of CAD can be difficult under any circumstance. Preoperatively, a decision must be made about the likelihood of CAD; again, if the chances of having the disease are low, there will be many false positive tests [*see* General Laboratory Testing, *below*], delays in surgery, and interventions that will not help many patients and may actually harm some of them. If a patient has known CAD or if angina is a serious consideration, however, then the severity and stability of the CAD must be ascertained. The clinician should then proceed with the evaluation process (see below).

The American College of Cardiology and the American Heart Association (ACC/AHA) have developed guidelines for preoperative cardiovascular evaluation for noncardiac surgery (http://www.acc.org/clinical/guidelines/perio/clean/perio_index.htm).[8] These guidelines are widely used. The ACC/AHA guidelines include an eight-step algorithm for patient risk stratification and subsequent determination of appropriate cardiac evaluation[9]; this algorithm is available on the ACC Web site (www.acc.org/clinical/guidelines/perio/update/fig1.htm).

The American College of Physicians has also developed guidelines on the perioperative assessment and management of risk from CAD.[10] Perhaps the most important message in the ACP's clinical guideline is the following: most patients who do not have an independent clinical need for coronary revascularization can proceed to surgery without further cardiac investigation.[10] In other words, if there is no prior reason to perform coronary artery bypass surgery, further cardiac investigation usually does not need to be carried out for the anticipated surgery, unless there is some other overriding consideration. This guideline does not apply to the patient with significant peripheral vascular disease, particularly if the patient is to undergo major vascular surgery.

Unstable angina or severe angina needs evaluation before surgery if the procedure is not immediately lifesaving. For patients who have stable, less than severe angina and for those whose symptoms occur only after significant exertion (e.g., walking at 4 miles an hour or engaging in moderate lifting), it is not necessary to proceed with coronary artery evaluation for most surgical procedures.[8] This is particularly true if the operative procedure is of low risk (see above). It may also be true in those at moderately higher risk when the following factors are considered: there may be an unacceptable increase in risk because of the delay induced by the bypass surgery, if such surgery is done before the anticipated operation (e.g., resection of potentially curable lung cancer or repair of a hip fracture); the combined mortality of bypass surgery and of the planned procedure may exceed the risk of simply proceeding with the operation (e.g., hip arthroplasty or herniorrhaphy).

The preoperative use of noninvasive testing for CAD is controversial. Despite the indications given in the ACC/AHA algorithm, it is important to consider the consequences of doing these tests, along with their rate of false positive and false negative results. Specifically, it is important to determine whether bypass surgery is truly an option before testing for CAD; if it is not an option, there is no need to test, because surgery will not be done no matter what the result. Some patients refuse bypass surgery even when significant disease is found.

Another consideration is the reliability and accuracy of the test or tests being considered. Exercise tests can be helpful, but only for those patients who can exercise enough to have a valid result. Those whose capacity to exercise is limited by weight, age, arthritis, or other physical or mental ailments will not achieve the level of exercise necessary to give valid results.

Thallium testing, particularly in women, can also be quite difficult. The specificity of these tests may be as low as 50% in

Table 2 Calculating the Likelihood of Postoperative Events[2]

Risk Factor	*Points*
CAD	
MI < 6 mo ago	10
MI > 6 mo ago	5
Canadian CVS angina	
Class III	10
Class IV	20
Unstable angina < 6 mo	10
Alveolar pulmonary edema	
< 1 wk	10
Ever	5
Critical aortic stenosis	20
Arrhythmia	
Other than sinus or PACs	5
> 5 PVCs/min	5
Poor medical status	5
Age > 70 yr	5
Emergency surgery	10

Class (Total Points)	*Likelihood Ratio**
1 (0–15)	0.42
2 (16–30)	3.58
3 (> 30)	14.93

*Likelihood ratio for postoperative events, defined as myocardial infarction, pulmonary edema, ventricular tachycardia or fibrillation, and cardiac death.
CAD—coronary artery disease CVS—Cardiovascular Society MI—myocardial infarction PAC—premature atrial contraction PVC—premature ventricular contraction

Table 3 Metabolic Equivalents of Selected Activities*

Activity	*METs*
Baking	2
Golfing with cart	2.5
Playing a musical instrument (various)	1.8–2.3
Mowing lawn (power mower)	3
Bicycling (leisurely)	3.5
Calisthenics (no weights)	4
Golfing without cart	5
Chopping wood	5

*A metabolic equivalent (MET) is an approximation of the energy expenditure involved in a particular physical activity, expressed as a unit of oxygen uptake; 1 MET is defined as 3.5 ml O_2/kg/min. Representative levels are as follows: 1 MET, resting; 2 METs, walking on level ground at 2 mph; 4 METs, walking on level ground at 4 mph.

women[11]; that is, half of the patients without disease may test positive. These patients will be subjected to invasive catheterization despite the absence of clinically significant disease.

Dobutamine echocardiography is now being used to identify patients with reversible ischemia. This test appears to have fairly good sensitivity and specificity when an adequate test can be done, but it is not definitive, and no controlled studies have proved its usefulness.[12,13] The results of dobutamine echocardiography may also be less accurate in patients in whom the study is difficult (e.g., obese patients).

ELECTROCARDIOGRAPHY

There is general agreement on indications for electrocardiography, and there is some additional clinical support for its use.[1,2] ECG testing is recommended for men older than 45 years and women older than 55 years, although these age limits are arbitrary. Many elderly patients (arbitrarily defined as those older than 65 years) will have ECG abnormalities even without a history of any cardiovascular disease.[14] The ECG has been shown to have modest predictive value in anticipating morbidity and mortality; it is also useful as a baseline to guide care in patients who develop postoperative symptoms or complications. In addition, any patient with hypertension, cardiovascular disease, or pulmonary disease should have an ECG. For these latter indications, it is probable that an ECG done within the past 3 to 6 months can be used, provided there has been no intervening clinical change.

ASSESSING PULMONARY RISK

The pulmonary evaluation process is unfortunately much more subjective than the cardiac evaluation. Despite many attempts to find methods to identify patients at increased pulmonary risk, no method has been shown to be definitively useful for an individual patient; thus, clinical judgment is critical. The importance of the surgery is often the deciding factor, even in those who are thought to be at high risk.

Significant lung disease, particularly from smoking, is often identified at the time of surgery. For that reason, taking a smoking history is an important part of the preoperative assessment. In the pulmonary assessment, it is important to note any dyspnea on exertion, wheezing, or coughing with sputum production, particularly if acute. Acute reversible disease, such as asthma or a respiratory tract infection, must be identified so that it can be treated and reversed before the procedure, if possible. Patients who can exercise without significant symptoms are at low risk.[3] Shortness of breath on exercise, in the absence of heart disease, identifies patients at higher risk. In general, the ability to climb a flight of stairs or walk several blocks without stopping indicates a lower risk level. Advancing age is a minor risk factor.[1,2] On physical examination, a respiratory rate above normal; poor chest wall movement; use of accessory muscles on inspiration; and poor breath sounds, wheezes, or rales on lung auscultation suggest significant underlying disease.

Clinical detection of obstructive airway disease has not been carefully studied. One study in 309 patients found only four clinical elements that were significantly associated with the diagnosis of obstructive airway disease: smoking for more than 40 pack-years, a self-reported history of chronic obstructive airway disease, a maximum laryngeal height of 4 cm (corrected) or less, and an age of 45 years or older.[15] In this study, the presence of all four findings indicated obstructive airways disease (likelihood ratio, 220); the absence of all four ruled out obstructive airways disease (likelihood ratio, 0.13). In patients with more severe chest disease, formal testing of respiratory function is indicated.

Pulmonary Function Tests

Preoperative use of pulmonary function tests (PFTs) is controversial.[16,17] PFTs do not readily identify individual patients who are at prohibitive risk for mortality; there is poor correlation between PFT results and mortality, despite some statistical correlation.[18] If the history and physical examination do not suggest significant pulmonary disease, there is no advantage in performing PFTs. If the intended surgery is of major importance, such surgery will have to be done regardless of the PFT results. The most important place for PFTs is in patients whose surgery may be of some benefit but whose risks may be prohibitively high. Many patients with very poor PFT results have been shown to survive surgery despite serious underlying disease.[19]

Exercise tolerance may be a good way to test for risk of death from pulmonary causes. It appears that for patients who can walk up two to four flights of stairs without stopping, the risk associated with major surgery is reasonable.[18]

Patients with chronic obstructive pulmonary disease and asthma should have their condition under the best possible control before surgery is attempted. In those whose symptoms are not controlled, surgery should be delayed until control is maximized; if this is not possible, additional drugs, including steroids, can be given before admission or in hospital to achieve as much control as possible before the procedure. It may be necessary to admit patients to the hospital one day or more before the operation to control their disease if it is particularly resistant to treatment.

Chest X-ray

Most experts believe that any patient older than 60 years should have a baseline chest x-ray. This has been shown to be clinically useful in elderly patients, who often have abnormalities that would not have been expected from history and physical examination.[20] Although the results of the chest x-ray may have little impact on preoperative risk assessment, I have found them to be useful in the care of patients who have postoperative pulmonary complications.

Clearly, any patient with cardiovascular or pulmonary disease needs a chest x-ray. Those with acute pulmonary symptoms and perhaps those with a history of tuberculosis or tuberculosis exposure probably derive some benefit from a chest x-ray. Again, in the absence of a change in symptoms or signs, a recent x-ray is usually adequate; a repeat x-ray is not needed in the immediate preoperative period.

ASSESSING RENAL RISK

Decreased renal function, most commonly defined as a serum creatinine concentration of approximately 2.0 mg/dl or higher, increases the risk of morbidity and mortality.[3] Because most renal disease results from hypertension and diabetes, a history of these diseases should lead to testing of renal function. Polyuria and nocturia may also indicate underlying kidney disease.

ASSESSING FOR LIVER DISEASE

Although significant liver disease is not common, failure to identify underlying liver disease can lead to catastrophic complications, particularly bleeding. The leading cause of significant liver disease is ethanol, so a history of alcohol intake is very

important. Hepatitis B and C are also important risk factors for impaired hepatic function; these diseases may be suggested by a history of bleeding episodes, transfusions, drug abuse, or sexual promiscuity.

ASSESSING FOR ENDOCRINE DISEASE

Endocrine abnormalities affect surgical risks. Diabetes is the most common endocrine abnormality; it is especially important because of its association with cardiovascular disease, which may be silent in diabetic patients.[21] Consequently, a history of hyperglycemic symptoms should be sought.

A history of thyroid disease is important, although most patients with hypothyroidism are at low risk unless they are profoundly affected.[22,23] On the other hand, identifying patients with undiagnosed or undertreated hyperthyroidism is very important, because it is well known that surgery can precipitate thyroid storm.[24] Thus, a history of hyperthyroidism or its symptoms (e.g., tremor, palpitations, weight loss, and anxiety) must be sought. Physical findings of tremor, lid lag, hyperreflexia, and tachycardia are useful clues. If there is any indication that the patient may be hyperthyroid, the thyroid-stimulating hormone level should be measured.

Addison disease is uncommon but can be deadly with the stress of surgery.[25] Profound weakness, orthostatic hypotension, and hyperpigmentation suggest adrenal insufficiency. For patients with adrenal insufficiency and those receiving long-term corticosteroid therapy, supplementary doses of corticosteroids must be given during the perioperative period.

ASSESSING FOR NEUROLOGIC DISEASE

Neurologic disease may significantly increase surgical risks. The most common and important neurologic disease is cerebrovascular disease. Hypertension, atherosclerosis, and diabetes are important antecedent processes. A history of episodic or fixed neurologic deficits, as well as a history of stroke, is important to identify. A history of memory loss or other cognitive deficits must also be sought, because strokes and dementia increase the risk of postoperative delirium.[26]

On examination, neurologic deficits must be noted and an etiology sought. I believe a baseline mini–mental status examination should be done on any patient suspected of having a dementing illness. With elderly patients, I recommend a mini–mental status examination even in those with no clear evidence of dementia. Particularly in elderly patients, any signs of dementia on mental status examination should be noted and clearly documented so that any suspected change after surgery can be compared with the presurgical state.

Medication Review and Adjustment

Review of medications, particularly in the elderly, is critical. In some cases, it may be necessary to discontinue a drug or adjust its dosage before surgery. As a side benefit, the medication review may reveal that the patient is taking unnecessary medications; this may be especially likely in elderly patients, many of whom are on multiple medications. Stopping unnecessary drugs simplifies the care of all patients.

The most commonly prescribed medications that require adjustment of dosages with surgery are those for hypertension and diabetes. Diuretics and oral diabetic medications are usually not administered on the day of surgery.[27,28]

Anticoagulant therapy obviously requires thoughtful perioperative management [*see* Thrombosis, *below*]. There are good reviews of the management of perioperative anticoagulation.[29] Because of their antiplatelet actions, nonsteroidal anti-inflammatory drugs (NSAIDs), including aspirin, should be discontinued 5 days before surgery. NSAIDs also increase the risk of acute renal failure, particularly in patients with renal insufficiency.

Many patients take over-the-counter medications or herbal remedies. It is important to specifically ask about such products when taking a medication history. These products may affect coagulation or interact with medications the patient will receive perioperatively. All such products should be stopped before surgery.

The clinician should review any previous surgery the patient has undergone. This review may reveal problems with anesthetic agents, such as malignant hyperthermia, that might otherwise be missed.

General Laboratory Testing

Laboratory testing before surgery has been studied extensively over the past 2 decades. Unfortunately, there are few data to help clinicians determine which patients truly need specific tests; most published recommendations are based on expert opinion. Many clinicians worry about the medicolegal risks of not performing tests before surgery, but preoperative testing also has drawbacks. Testing subjects the patient to increased risks from more invasive procedures such as catheterization and may even erroneously indicate disease that is not present.

Noninvasive tests are often far from perfect for predicting the presence or absence of disease. Most of these tests have rather good sensitivity but only moderate specificity. The lower the prevalence of a disease in the population being tested for it, the greater the odds of false-positive results, which almost inevitably lead to further costly, invasive, and perhaps risky testing. This limitation of testing underlines the importance of a good history and physical examination for selecting patients for testing. Methods for calculating the likelihood of disease on the basis of a given test result are discussed in detail elsewhere [*see CE:VIII Quantitative Aspects of Clinical Decision Making*.]

For many years, most preoperative patients at most institutions underwent a standard panel of tests: a complete blood count (CBC), including platelets; electrolytes and renal panel; extended serum chemistry studies with liver function tests; prothrombin time and partial thromboplastin time; and chest x-ray and electrocardiogram. Kaplan and colleagues showed that many tests were not needed for the care of these patients and that many of the abnormal test results were only slightly out of the normal range.[30] Almost all of those abnormal results were ignored without consequence or were pursued with no benefit to the patients. Pursuing these abnormal results simply added more tests and caused delays in surgery. A number of subsequent studies confirmed these findings.[31,32]

Most experts agree that patients who are older than 60 years and those undergoing procedures that pose a substantial risk of bleeding should have their hematocrit measured. It is not clear, however, that even this limited testing makes a clinical difference. Obviously, patients with a history of anemia should have a CBC; those with any underlying disease associated with anemia should also be tested. Measurement of renal function and electrolyte levels may be indicated for all patients older than 60 years, but there are no data to support this recommendation. Such tests should be performed on the following patients: those

with hypertension, diabetes, cardiovascular disease, or renal disease; those who are taking diuretics, angiotensin-converting enzyme (ACE) inhibitors, or angiotensin receptor blockers (ARBs); and those who will undergo bowel preparation for surgery. The Cockcroft-Gault equation should to used to estimate renal function from the serum creatinine; this equation estimates creatinine clearance, corrected for age and weight:

$$\text{Creatinine clearance (ml/min)} = \frac{(140 - \text{age}) \times \text{wt (in kg)}}{(\text{serum creatinine} \times 72)}$$

Routine measurement of the platelet count, prothrombin time, and partial thromboplastin time are not necessary.[30] Indications for testing include a history of prolonged bleeding, easy bruising, postoperative bleeding, or a family history of bleeding.[33] Patients on anticoagulants obviously also need testing.

Other areas of testing are more controversial. Liver enzyme, bilirubin, and albumin measurements are not needed unless there are indications of liver dysfunction on history and physical examination. Other serum chemistry tests, such as calcium and phosphate concentrations, are rarely indicated.

Management of Surgical Risk Factors

CORONARY ARTERY DISEASE

It is now clear that the use of perioperative beta blockers can prevent complications after surgery, both short term and long term.[34-36] Patients with known CAD who can tolerate beta blockers should already be taking these drugs. If they are not, a beta blocker should be started. Many experts recommend starting beta blockade before surgery in patients at high risk for CAD; such patients include those older than 70 years and those who have symptomatic heart failure, a history of stroke, renal insufficiency (i.e., a serum creatinine concentration greater than 2.0), or diabetes.[20] A number of agents have proved useful for perioperative treatment, including labetolol, esmolol, and bisoprolol. Because only a few studies have been done, no specific recommendations can be made; both intravenous and oral regimens have been used with success.[36]

BLOOD PRESSURE

Patients with diastolic blood pressures below 110 mm Hg are not at significant added risk and do not require specific blood pressure management.[1] For elective surgery, however, I believe it is useful that blood pressure be reasonably controlled (i.e., that diastolic blood pressure be lower than 100 mm Hg). It is clear that in patients with poorly controlled blood pressure who undergo surgery, blood pressure may swing widely, both in hypertensive and hypotensive directions. High blood pressure and low blood pressure can each cause problems perioperatively. Unless surgery is urgent or emergent, hurried attempts at blood pressure control are not advised. Excessive diuretic use can lead to volume depletion and hypokalemia; heavy doses of other medications can lead to unanticipated hypotension. Maintaining reasonable control for a period of about a week on an outpatient basis appears to be quite safe. Diuretics are usually withheld on the day of surgery to avoid preoperative volume decrease.[27] Other antihypertensive medications should be taken as usual. Patients with very high blood pressure who need surgery may require treatment with intravenous doses of drugs such as nitroprusside, labetalol, or enalaprilat.

RENAL INSUFFICIENCY

Although patients with significant renal disease are known to be at increased risk for perioperative morbidity and mortality, including increased cardiac complications,[1,2] there is little that can be done to decrease the risks for most of these patients. Foreknowledge aids in watching for complications, however. Perhaps the most common problem seen is volume depletion from high doses of diuretics. A blood urea nitrogen:creatinine ratio of approximately 10 is usually optimal; higher levels should increase suspicion of volume depletion. The presence of such signs as tachycardia, orthostatic changes in blood pressure and pulse, and poor skin turgor help to confirm this suspicion.

Other problems that are seen frequently in patients with renal insufficiency include hyponatremia, hyperkalemia, volume overload, and metabolic acidosis. These abnormalities should be treated to the extent possible before the operation. In patients with acute renal insufficiency, acute complications should be anticipated. Delirium from uremia, serositis (particularly pericarditis), and severe acidosis may be present or may develop quickly. Some of these patients may require dialysis before surgical intervention.

The management of patients on replacement therapy can be quite complicated. In general, to be in optimal condition for the operation, most patients undergo dialysis either on the day before or the day of surgery.

HYPERGLYCEMIA AND DIABETES MELLITUS

Hyperglycemia can lead to poor wound healing and perioperative infections, particularly wound infections.[37] Control of blood sugars to below approximately 200 mg/dl appears to be sufficient. There is now evidence that controlling blood sugars, even in patients without previous diabetes, can improve outcomes in patients in intensive care units.[38]

Oral antidiabetic agents should be discontinued on the day of surgery. In patients whose diabetes is controlled by diet and oral agents, normoglycemia can usually be maintained by giving regular insulin every 4 hours on a sliding scale, until diet and oral medications can be resumed.[39] Patients on insulin generally must be given insulin during and after surgery. It is best to perform surgery in the morning in these patients so that the dose of insulin can be more easily determined.

There are a number of methods of giving insulin.[39] There are no good studies to prove that one method gives better clinical results than another. The most expensive and time consuming but the one that provides the best glucose control is an intravenous drip of regular insulin. This method is probably not necessary in most cases, but in difficult, brittle cases, it will give the best glucose control. In most cases, it is reasonable to give one half to two thirds of the usual dose of neutral protamine Hagedorn (NPH) or other long-acting insulin, monitor regularly, and give regular insulin or glucose as needed.[39] Patients with hyperglycemia or acidosis will require more intensive dosing with regular insulin.

DECREASED HEMATOCRIT

The hematocrit level necessary for minimizing surgical risk has been controversial for many years. There are no good data to support any specific limits for all patients, although young, healthy patients clearly can tolerate much lower hematocrits than older patients with multiple underlying diseases, particularly those of the heart, vascular system, and lungs. Most patients without underlying disease can tolerate surgery if their

hematocrit is above 28%; those with underlying heart or lung disease, particularly elderly patients, may need to undergo transfusion to raise the hematocrit above 35%.[40] Transfusion appears to decrease the development of postoperative delirium.

THYROID DISEASE

Patients with hypothyroidism do not have significant problems with surgery and do not require special treatment, provided they are functional.[23] Patients who are clinically hyperthyroid are at risk for thyroid storm perioperatively, so hyperthyroidism should be well controlled before surgery is undertaken. If surgery is necessary, it is best to consult an endocrinologist to manage the hyperthyroidism perioperatively.[41]

LIVER DISEASE

Underlying liver disease does not appear to be a major risk factor for surgery under most circumstances,[1] nor do elevated levels of liver enzymes appear to be associated with increased risk. Active inflammation from alcoholic hepatitis, however, substantially increases the risk of mortality. Patients with fever, elevated white blood cell counts, and jaundice should not have surgery unless their condition is clearly not the result of alcoholic hepatitis and they are facing a surgical emergency.[42] The main risks from underlying liver disease are variceal bleeding in patients with varices and hepatic encephalopathy in those at risk. Ascites also presents difficult management problems, with many complicated issues that are beyond the scope of this chapter [*see 4:IX Cirrhosis of the Liver*].

CORTICOSTEROID THERAPY

Patients who are taking corticosteroids (e.g., for rheumatic disease or asthma) usually need replacement therapy perioperatively.[25] In normal persons, the daily output of cortisone is approximately 30 mg; peak stress levels are approximately 300 mg a day. Unfortunately, no good studies have been done to determine which patients definitely need supplementation and how long the increase in dose should be maintained. Current practice is to give additional medication to patients who are taking the equivalent of 30 mg or more of hydrocortisone a day. For patients at lower dose ranges, supplemental doses (e.g., 50 mg) given twice daily are adequate; for those taking more than 150 mg a day, three doses a day of 50 to 100 mg are usually recommended. It is unusual to give more than 100 mg three times a day. For most patients, the dose can be tapered back to baseline in 2 to 3 days.

ALCOHOL ABUSE

Patients who abuse alcohol are at risk for withdrawal syndromes when admitted for surgery. It is best to get patients to stop drinking at least a week before surgery. If efforts to get a patient to stop do not succeed (or if such efforts were not made, because the patient gave a misleading drinking history), then therapy should be begun for withdrawal as soon as the initial signs and symptoms are noted. Most often, these consist of tremor and tachycardia with agitation. Oral benzodiazepines (e.g., chlordiazepoxide, lorazepam, diazepam), are the usual choice for treating withdrawal; they may be given every 4 to 6 hours until symptoms abate.[43] The dose will depend on the patient's alcohol consumption; the dose can be remarkably high, as compared with the dose in patients who do not abuse alcohol. Requiring the equivalent of 400 mg of chlordiazepoxide is not uncommon.

THROMBOSIS

All surgical procedures increase the risk of deep vein thrombosis (DVT), and most major surgery carries a significant risk of both DVT and pulmonary embolism. Patients older than 40 years who will have an operation lasting longer than 30 minutes will require DVT prophylaxis. Most general and gynecologic surgery patients require low-dose heparin at 5,000 U every 12 hours; some experts believe that 5,000 U every 8 hours should be given for patients having long and complicated surgical procedures. In orthopedic surgery, particularly hip and knee operations, low-molecular-weight heparin is the agent of choice. There is increasing and convincing evidence that continuing DVT prophylaxis, particularly after orthopedic hip and knee surgery, for 4 to 6 weeks postoperatively prevents morbidity and mortality from late thromboses and pulmonary emboli. This is particularly true in light of the shorter hospitalizations of patients after surgical procedures.[44,45] Warfarin can be used as well. For surgical procedures in which any increased risk of bleeding is unacceptable (e.g., neurosurgery or open prostatectomy), intermittent compression stockings are most often used to prevent DVT.

ADVANCED AGE

Surgery in elderly patients has become routine over the past decade or so. Improvements in anesthetic techniques, better understanding of the management of medical diseases, and improved and less invasive surgical techniques have allowed older and much sicker patients to undergo surgery. Many studies have shown that an elderly patient can undergo most procedures that are important to survival or quality of life without being exposed to a large increase in perioperative risk.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Goldman L, Caldera DL, Nussbaum SR et al: Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 297:845, 1977

2. Detsky AS, Abrams HB, McLaughlin JR, et al: Predicting cardiac complications in patients undergoing noncardiac surgery. J Gen Intern Med 1:211, 1986

3. Lee TH, Marcantonio ER, Mangione CM, et al: Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 100:1043, 1999

4. Palda VA, Detsky AS: Perioperative assessment and management of risk for CAD. Ann Intern Med 127:313, 1997

5. Gilbert K, Larocque BJ, Patrick LT: Prospective evaluation of cardiac risk indices for patients undergoing noncardiac surgery. Ann Intern Med 133:356, 2000

6. Reilly DF, McNeely MJ, Doerner D, et al: Self reported exercise tolerance and the risk of serious perioperative complications. Arch Intern Med 159:2185, 1999

7. Halm EA, Browner WS, Tubau JF, et al: Echocardiography for assessing cardiac risk in patients having non-cardiac surgery. Ann Intern Med 125:433, 1996

8. Eagle KA, Berger PB, Calkins H, et al: ACC/AHA guideline update for perioperative cardiovascular evaluation of noncardiac surgery—executive summary: a report of the ACC/AHA task force on practice guidelines (Committee to Update the 1996 Guidelines on Perioperative Cardiovascular Evaluation for Noncardiac Surgery). J Am Coll Cardiol 39:542, 2002

9. Mukherjee D, Eagle KA: Perioperative cardiac assessment for noncardiac surgery: eight steps to the best possible outcome. Circulation 107:2771, 2003

10. Guidelines for assessing and managing the perioperative risk from CAD associated with major non-cardiac surgery. American College of Physicians. Ann Intern Med 127:309, 1997

11. Kwok Y, Kim C, Grady D et al: Meta-analysis of exercise testing to detect CAD in women. Am J Cardiol 83:660, 1999

12. Grayburn PA, Hillis LD: Cardiac events in patients undergoing noncardiac surgery: shifting the paradigm from noninvasive risk stratification to therapy. Ann Intern Med 138:506, 2003

13. Geleijnse ML, Elhendy A, Fioretti PM, et al. Dobutamine stress myocardial perfusion imaging. J Am Coll Cardiol 36:2017, 2000

14. Seymour DG, Pringle R, MacLennan WJ: The role of the routine preoperative elec-

trocardiogram in the elderly surgical patient. Age Ageing 12:97, 1983

15. Straus SE, McAlister FA, Sackett DL, et al: The accuracy of patient history, wheezing, and laryngeal measurements in diagnosing obstructive airway disease. CARE-COAD1 Group. Clinical Assessment of the Reliability of the Examination—Chronic Obstructive Airways Disease. JAMA 283:1853, 2000

16. Lawrence V, Page CP, Harris GD: Preoperative spirometry before abdominal operations. Arch Intern Med 149:280, 1989

17. Kocabas A, Kara A, Ozgur G, et al: Value of preoperative spirometry to predict postoperative pulmonary complications. Respir Med 90:25, 1996

18. Smetana GW: Preoperative pulmonary evaluation. N Engl J Med 340:937, 1999

19. Arozullah A, Daley J, Henderson W, et al: Multifactorial risk index for predicting postoperative respiratory failure in men after noncardiac surgery. Ann Surg 232:242, 2000

20. Seymour DG, Pringle R, Shaw JW: The role of the routine chest x-ray in the elderly general surgical patient. Postgrad Med 58:742, 1982

21. Consensus development conference on the diagnosis of coronary heart disease in people with diabetes. American Diabetes Association. Diabetes Care 12:1551, 1998

22. Ladenson PW, Levin AA, Ridgeway EC, et al: Complications of surgery in hypothyroid patients. Am J Med 77:261, 1984

23. Drucker DJ, Burrows GN: Cardiovascular surgery in the hypothyroid patient. Arch Intern Med 145:1585, 1985

24. Lamphier TA, Wickman W: Postoperative thyroid storm. Postgrad Med 15:493, 1954

25. Salassa RM, Bennett WA, Keating FR, et al: Postoperative adrenal cortical insufficiency: occurrence in patients previously treated with cortisone. JAMA 152:1509, 1953

26. Winawer N: Postoperative delirium. Med Clin North Am 85:1229, 2001

27. Kroenke K, Gooby-Toedt D, Jackson JL: Chronic medications in the perioperative period. South Med J 91:358, 1998

28. Lawrence R, Walter RM: Diabetes mellitus. Medical Management of the Surgical Patient. Lubin MF, Walker HK, Smith RB III, Eds. Lippincott Williams & Wilkins, Philadelphia, 1995, p 317

29. Geerts W, Heit J, Claggett G, et al: Prevention of venous thromboembolism. Chest 119:132s, 2001

30. Kaplan EB, Sheiner LB, Boeckmann AJ, et al: The usefulness of preoperative laboratory screening. JAMA 253:3576, 1985

31. Narr BJ, Hansen TR, Warner MA: Preoperative laboratory screening in healthy Mayo patients. Mayo Clinic Proc 66:15, 1991

32. McKee RF, Scott EM: The value of routine preoperative investigation. Ann R Coll Surg Engl 69:160, 1987

33. Rapaport SI: Preoperative hemostatic evaluation: which tests if any? Blood 61:229, 1983

34. Mangano DT, Layug EC, Wallace A, et al: Effect of atenolol on mortality and cardiovascular morbidity after non-cardiac surgery. N Engl J Med 335:1713, 1996

35. Poldermans D, Boersma E, Bax JJ, et al: The effect of bisoprolol on perioperative mortality and myocardial infarction in high risk patients undergoing vascular surgery. N Engl J Med 341:1789, 1999

36. Auerbach AD, Goldman L: Beta blockers and reduction of cardiac events in noncardiac surgery: scientific review. JAMA 287:1435, 2002

37. Golden SH, Peart-Vigilance C, Kao WH, et al: Perioperative glycemic control and the risk of infectious complications in a cohort of adults with diabetes. Diabetes Care 22:1408, 1999

38. Van den Berghe G, Wouter R, Weekers F, et al: Intensive insulin therapy in critically ill patients. N Engl J Med 345:1359, 2001

39. Hoogwerf BJ: Management of the diabetic patient. Med Clin North Am 85:1213, 2001

40. Carson JL, Duff A, Berlin JA, et al: Perioperative blood transfusion and postoperative mortality. JAMA 279:199, 1998

41. Langley RW, Burch HB: Perioperative management of the thyrotoxic patient. Endocrinol Metab Clin North Am 32:519, 2003

42. Powell-Jackson P, Greenway B, Williams R: Adverse effects of exploratory laparotomy in patients with unsuspected liver disease. Br J Surg 69:449, 1982

43. Chang PH, Steinberg MB: Alcohol withdrawal in postoperative patients. Med Clin North Am 85:1191, 2001

44. Hull RD, Pineo GF, Stein PD, et al: Extended out-of-hospital low-molecular-weight heparin prophylaxis against deep vein thrombosis in patients after elective hip arthroplasty: a systematic review. Ann Intern Med 135:858, 2001

45. Eikelboom JW, Quinlan DJ, Couketis JD: Extended-duration prophylaxis against venous thromboembolism after total hip or knee replacement: a meta-analysis of the randomised trials. Lancet 358:9, 2001

V BIOTERRORISM

Jeffrey Duchin, m.d.

Well before the 2001 anthrax outbreak, public health and government leaders in the United States recognized the need for increased preparedness to detect and respond to acts of biologic terrorism. Concern about the vulnerability of the United States to a biologic attack grew with revelations about the offensive biologic weapons programs of the former Soviet Union and Iraq, as well as uncertainty about the whereabouts of and accountability for biologic agents produced through those programs; the successful chemical attack on the Tokyo subway system by the Aum Shinrikyo cult, coupled with information that the cult was actively experimenting with biologic agents; and information about the potential for domestic bioterrorism.[1-5]

In April 2000, the Centers for Disease Control and Prevention (CDC) published a strategic plan for preparedness and response to biologic and chemical terrorism.[6] This subsection describes the clinician's role in recognizing and responding to biologic terrorism, as presented in the CDC plan; summarizes current information on the diagnosis and management of the most likely agents of bioterrorism; and describes current resources for authoritative information and guidelines related to bioterrorism.

The Clinician's Role in Bioterrorism Preparedness and Response

For clinicians, the response to a bioterrorism attack is in many ways the same as the response to naturally occurring outbreaks of communicable disease.[7,8] Both situations typically require early identification of ill or exposed persons, rapid implementation of preventive therapy, special infection control considerations, and collaboration or communication with the public health system. Examples of naturally occurring communicable diseases that require such a response include meningococcal disease[9]; enteric infection with *Escherichia coli* 0157:H7, *Salmonella,* or *Shigella*[10]; pertussis, rubella, measles, or chickenpox occurring in health care facilities and clinics[11-14]; unusual or newly emerging infections such as West Nile virus and hantavirus pulmonary syndrome[15-17]; and the inevitable reappearance of pandemic influenza.[18] The 2002 outbreak of severe acute respiratory syndrome (SARS) exemplifies the type of unexpected, naturally occurring disease for which the need for preparedness and the impact on clinicians have much in common with that of biologic terrorism.[19,20]

The first indication of an unannounced biologic attack will likely be an increase in the number of persons seeking care from primary care physicians. In the 2001 anthrax outbreak, as well as in outbreaks of *E. coli* 0157:H7 disease and hantavirus pulmonary syndrome in 1993 and West Nile virus in 1999, alert clinicians initiated the public health response by recognizing an unusual clinical syndrome, ordering appropriate laboratory tests, and notifying public health officials.[10,16,17] Similarly, primary care physicians and subspecialists alike must be familiar with both the specific clinical syndromes associated with agents of bioterrorism and the ways to rapidly notify public health authorities. In addition to identifying cases and treating ill patients, clinicians also play a critical role in managing postexposure prophylaxis and its complications, as well as psychological and mental health problems brought on by the event.

During both bioterrorism attacks and naturally occurring outbreaks, clinicians are faced with the challenge of excluding the outbreak disease in persons who are worried about potential exposure or who are ill with signs and symptoms similar to those of the outbreak disease. The clinician must have knowledge of the signs and symptoms, modes of transmission, incubation periods, and communicable periods of these diseases, as well as skill in both clinical evaluation and eliciting an appropriate and thorough history, including relevant occupational, social, and travel information. In both the 2001 anthrax bioterrorism attack and the 2002 SARS outbreak, the epidemiologic setting of cases played an important role in guiding clinical management, diagnostic tests, and treatment.[21-23] The primary care clinician has the best opportunity to obtain relevant information early in the evaluation; this is important because such information may be more difficult to obtain as time goes on, particularly if the patient's condition deteriorates.

Physicians and other health care providers should have a working knowledge of the basic classes of isolation and infection control measures recommended for patients exposed to agents of potential bioterrorism. Again, these measures are also used in the management of common communicable diseases.[14,24-26]

Recognition of Potential Bioterrorism Agents

The CDC has developed a list of bacteria, viruses, and toxins thought to pose the greatest risk for use in a bioterrorist attack [*see Table 1*].[27] Agents were included in the list on the basis of their ability to cause disease that (1) is easily disseminated or transmitted from person to person; (2) has high mortality, with potential for major public health impact; (3) may result in panic and social disruption; and (4) requires special action for public health preparedness. Category A agents are thought to pose the highest immediate risk for use as biologic weapons; and category B agents, the next highest risk. Category C agents are thought to pose a potential, but not immediate, risk for use as biologic weapons.

As in naturally occurring outbreaks, early recognition of a bioterrorist attack is critical for rapid implementation of preventive measures and treatment. Early recognition can be challenging, however, because patients presenting for medical care after exposure to a biologic agent may initially exhibit nonspecific symptoms, and pathogens that ordinarily occur in the community, particularly enteric organisms, may be used in a biologic attack.[28,29] A heightened level of suspicion, plus knowledge of the relevant epidemiologic clues, should help physicians recognize changes in illness patterns, including clusters and increases in observed cases over the number expected [*see Table 2*].[30] Physicians should also be able to recognize diagnostic clues in single cases of a syndrome of concern (e.g., inhalational anthrax, plague and tularemia, botulismlike illness, and possible smallpox).[31] Familiarity with the clinical features of diseases from potential bioterrorist agents and diseases prevalent in the community will allow recognition of potentially significant differences from naturally occurring cases. One of the most important lessons learned from the 2001 anthrax attack was that clinical ill-

Table 1 Critical Biologic Agent Categories for Public Health Preparedness

Category	*Biologic Agent*	*Disease*
A (highest immediate risk)	Variola major	Smallpox
	Bacillus anthracis	Anthrax
	Yersinia pestis	Plague
	Clostridium botulinum (botulinum toxins)	Botulism
	Francisella tularensis	Tularemia
	Filoviruses and arenaviruses (e.g., Ebola virus, Lassa virus)	Viral hemorrhagic fevers
B (next highest risk)	*Coxiella burnetii*	Q fever
	Brucella species	Brucellosis
	Burkholderia mallei	Glanders
	Burkholderia pseudomallei	Melioidosis
	Alphaviruses (VEE, EEE, WEE)	Encephalitis
	Rickettsia prowazekii	Typhus fever
	Toxins (e.g., ricin, staphylococcal enterotoxin B)	Toxic syndromes
	Chlamydia psittaci	Psittacosis
	Food-safety threats (e.g., *Salmonella* species, *E. coli* 0157:H7)	
	Water-safety threats (e.g., *Vibrio cholerae*, *Cryptosporidium parvum*)	
C (potential, but not immediate, risk)	Emerging-threat agents (e.g., Nipah virus, hantavirus)	

EEE—eastern equine encephalitis VEE—Venezuelan equine encephalitis WEE—western equine encephalitis

ness caused by agents prepared as biologic weapons may differ from typical natural infections.

The identification of a bioterrorist attack requires clinicians to be prepared, alert, and open-minded [*see Sidebar* Internet Resources on Bioterrorism]. Many local and state health departments post current information about communicable diseases on their Web sites and distribute informational newsletters with relevant data. The CDC's weekly bulletin, *Morbidity and Mortality Weekly Report (MMWR),* contains current information on medical conditions of public health importance in the United States. Subscriptions to MMWR are available online at http://www.cdc.gov/mmwr/mmwrsubscribe.html.

Communication with Authorities

Once a potential outbreak or significant cluster or event has been detected, prompt consultation with appropriate medical specialists and public health authorities is indicated. Clinicians must have reliable, around-the-clock contact information for emergency resources in the geographic area where they practice; these resources include specialist consultants (e.g., consultants in infectious disease, dermatology, or pulmonary medicine) and infection control professionals or hospital epidemiologists. All clinicians should know how to contact their local or state public health department 24 hours a day to report suspicious or otherwise immediately notifiable cases or for consultation. Many local and state health departments have such contact numbers on their Web sites. Clinicians should have these numbers readily accessible and keep them current.

Clinicians must also ensure that they have a reliable way to promptly receive urgent communications from public health authorities, both for naturally occurring outbreaks of local significance and for a bioterrorist event or outbreak. Increasingly, public health authorities are disseminating health alerts over the Internet, through Web sites and e-mail listserves.

Smallpox

Smallpox is caused by variola virus, an orthopox virus unique to humans. No known animal or insect reservoirs or vectors exist.[32] Related orthopox viruses infecting humans include vaccinia (smallpox vaccine), monkeypox, and cowpox. Smallpox existed in two forms: variola major, which accounted for most morbidity and mortality, and a milder form, variola minor. Variola major is the type of concern in the context of biologic terrorism.

Smallpox was declared eradicated in 1980, 3 years after the last naturally occurring case was reported from Somalia. Stocks of smallpox virus were retained, however, by World Health Organization (WHO) reference laboratories at the Institute of Virus Preparations in Moscow, Russia, and at the CDC in Atlanta, Georgia. In the late 1990s, allegations were published describing the production of large quantities of smallpox virus by the former Soviet Union. These stores, which may have become dis-

Table 2 Epidemiologic Clues of a Biologic Attack

- Presence of a large epidemic
- Unusually severe disease or unusual routes of exposure
- Unusual geographic area, unusual season, or absence of normal vector
- Multiple simultaneous epidemics of different diseases
- Outbreak of zoonotic disease
- Unusual strains of organisms or antimicrobial-resistance patterns
- Higher attack rates in persons with common exposures
- Credible threat, as determined by authorities, of biologic attack
- Direct evidence of biologic attack

Internet Resources on Bioterrorism

General

CDC portal to bioterrorism information for laboratory and health professionals
http://www.bt.cdc.gov/index.asp

Contact information for U.S. state and local health departments
http://www.cdc.gov/other.htm#states

CDC Emergency Response Hotline (24 hours): 770-488-7100

American College of Physicians–American Society of Internal Medicine (ACP-ASIM) bioterrorism resources
http://www.acponline.org/bioterro/?hp

Association for Professionals in Infection Control and Epidemiology, Inc.
http://www.apic.org/bioterror/

Center for Infectious Disease Research & Policy (CIDRAP), University of Minnesota
http://www1.umn.edu/cidrap/content/bt/bioprep

FDA bioterrorism site
http://www.fda.gov/oc/opacom/hottopics/bioterrorism.html

Infectious Diseases Society of America
bioterrorism: **http://www.idsociety.org/BT/ToC.htm**
practice guidelines: **http://www.idsociety.org/PG/toc.htm**

Johns Hopkins University Center for Civilian Biodefense Strategies
http://www.hopkins-biodefense.org

National Academies' Expert-selected Web Resources for "First Responders" on Bioterrorism
http://www.nap.edu/shelves/first

National Institute of Allergy and Infectious Diseases
http://www.niaid.nih.gov/publications/bioterrorism.htm

National Library of Medicine Specialized Information Services
http://www.sis.nlm.nih.gov/Tox/biologicalwarfare.htm

St. Louis University Center for Study of Bioterrorism
http://bioterrorism.slu.edu

Treatment of Biological Warfare Agent Casualties (July 17, 2000), U.S. Army Field Manual on Treatment of Biological Warfare Casualties
http://www.nbc-med.org/SiteContent/MedRef/OnlineRef/FieldManuals/Fm8_284/fm8_284.pdf

USAMRIID's Medical Management of Biological Casualties Handbook (Blue Book)
http://www.usamriid.army.mil/education/bluebook.html
word format: **http://www.nbc-med.org/SiteContent/HomePage/WhatsNew/MedManual/Feb01/handbook.htm**

seminated after the breakup of the Soviet Union, would presumably be the source for a bioterrorist attack involving smallpox.

Smallpox is stable and highly infectious in the aerosol form. The risk for a smallpox attack currently is considered low but not zero.[1,4,33,34]

CLASSIFICATION

On the basis of a study from India, the WHO has classified smallpox into five clinical forms: ordinary, flat-type, hemorrhagic, modified, and sine eruptione.[35] These forms reflect different host reactions to the same strain of virus.

Ordinary Smallpox

Ordinary smallpox is the most common form seen in nonimmune persons; it accounted for 90% of cases in the WHO study and had an average case-fatality rate of 30%. The incubation period is 7 to 17 days (mean, 10 to 12 days). Symptoms of the prodromal phase include the acute onset of high fever, malaise, headache, backache, and prostration. Other prominent symptoms include vomiting and abdominal pain.

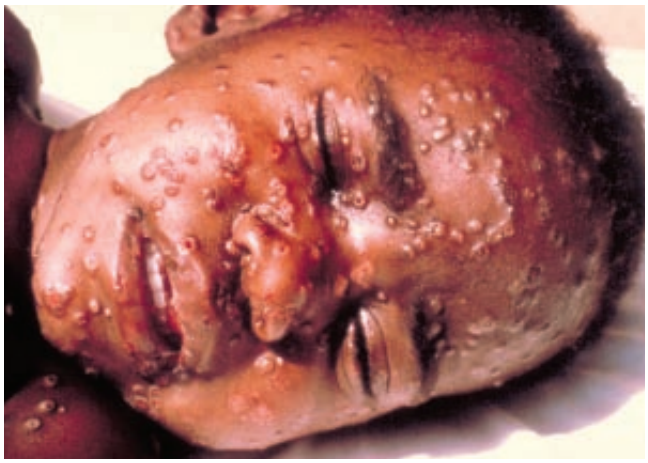


Figure 1 **Lesions of smallpox.**

The characteristic rash occurs 2 to 3 days later, appearing first on the face and forearms. An enanthem involving the oropharyngeal mucosa precedes the rash by a day. The rash progresses slowly, from macules to papules to vesicles and pustules and finally to scabs, with each stage lasting 1 to 2 days. The lesions are firm, discrete vesicles or pustules (4 to 6 mm in diameter) deeply embedded in the dermis; they may become umbilicated or confluent as they evolve [*see Figure 1*]. The patient remains febrile throughout the evolution of the rash, which may become painful as pustules enlarge. A second fever spike 5 to 8 days after onset of the rash may signify a secondary bacterial infection. Pustules remain for 5 to 8 days, after which umbilication and crusting occur. Lesions are in the same stage of development on any given part of the body. They are peripherally distributed, more concentrated on the face and distal extremities than on the trunk, and may involve the palms and soles. Scarring occurs with scab separation from destruction of sebaceous glands.

Experience during the global smallpox eradication program suggests that the onset of communicability coincides with the development of rash, approximately 2 days after the onset of the acute febrile prodrome. However, because the oropharyngeal enanthem and associated release of virus into oral secretions may precede rash onset, it is recommended that for the purposes of postexposure management, anyone who has contact with smallpox patients from the time of onset of fever should be considered potentially exposed [*see* Infection Control, *below*].[36]

Complications of smallpox include fluid and electrolyte disturbances; extensive desquamation that clinically resembles burns; bronchitis and pneumonitis; panophthalmitis and blindness from viral keratitis or secondary infection of the eye; arthritis (developing in up to 2% of children); and encephalitis (less than 1% of cases). Death results from toxemia associated with circulating immune complexes and variola antigens.[37]

Other Forms of Smallpox

Flat-type (or malignant) smallpox occurs in 5% to 10% of cases and is severe, with a 97% case-fatality rate among unvaccinated persons. In this form, lesions are flat and become densely confluent, evolving slowly and coalescing with a soft, velvety texture. Hemorrhagic smallpox was reported in less than 3% of cases, occurring particularly in pregnant women. It is a severe, rapidly progressive, uniformly fatal illness. A dusky erythema develops, followed by hemorrhages into the skin and mucous membranes. Both hemorrhagic and flat-type smallpox have an accelerated and more severe prodromal phase and are thought to be associated with underlying immune dysfunction.

Modified smallpox is a mild form that accounted for 2% of cases in unvaccinated patients and 25% in previously vaccinated patients. This form rarely resulted in death, and these patients had fewer, smaller, more superficial, and more rapidly evolving lesions. Smallpox sine eruptione (without rash) occurs in previously vaccinated persons or children with maternal antibodies to smallpox. It is a mild or asymptomatic illness that has not been documented to be transmissible.[35,37-39]

DIAGNOSIS

A suspected case of smallpox is a public health emergency. Local and state health authorities, the hospital epidemiologist, and other members of a hospital response team for biologic emergencies should be notified immediately (see the CDC Interim Smallpox Response Plan and Guidelines at http://www.bt.cdc.gov/documentsapp/SmallPox/RPG/ContactInfo.asp).

The differential diagnosis of smallpox includes other illnesses that can cause fever and a rash [*see Table 3*]. Severe varicella is the disease most likely to be confused with smallpox. However, familiarity with the clinical features of the two diseases, particularly the rash, should help differentiate them [*see Table 4*]. Additional information that may be useful in differentiating smallpox from chickenpox includes a history of exposure to persons with chickenpox, a personal history of chickenpox, a history of vaccination against varicella or smallpox, and the clinical course of illness.

If shingles or disseminated herpes infection is a consideration, direct fluorescent antibody testing for varicella-zoster virus can rapidly confirm varicella-zoster virus and herpes simplex virus infection in patients not considered at high risk for smallpox. Such testing should not be done in patients who are considered at high risk, to avoid exposing laboratory workers to smallpox virus. Certain laboratories can also perform polymerase chain reaction (PCR) testing for herpes simplex virus and varicella-zoster virus. Consultation with an infectious disease specialist, a dermatology specialist, or both is recommended.

Flat-type and hemorrhagic smallpox may be difficult to recognize because of the absence of the characteristic rash of ordinary smallpox, yet these cases are highly infectious. Hemorrhagic smallpox cases may be mistaken for meningococcemia or acute leukemia. All patients with potential smallpox should be asked about their travel history, history of varicella and vaccinia vaccination, level of immunocompetence, and current medications.

The local or state health department should be contacted to facilitate specimen collection for smallpox testing (http://www.statepublichealth.org). Protocols for specimen collection for smallpox testing have been published by the CDC and are available on the Internet (http://www.cdc.gov).

Diagnostic testing is available at designated biosafety level 4 (BSL-4) laboratories and includes electron microscopy, immunohistohemical tests, and viral culture with PCR and restriction fragment length polymorphism (RFLP) testing. Only personnel who have undergone successful smallpox vaccination recently (within 3 years) and who are wearing appropriate barrier protection (gloves, gown, and shoe covers) should be involved in specimen collection for suspected cases of smallpox. Respiratory protection is not needed for personnel with recent, successful vaccination. Masks and eyewear or face shields should be used if splashing is anticipated. If unvaccinated personnel must collect specimens, only those who are without con-

Table 3 Diagnosis of Smallpox

Incubation Period	*Clinical Presentation*	*Differential Diagnosis*	*Diagnostic Testing*
7–17 days; mean, 10–12 days	Severe, acute febrile prodrome 1–4 days before rash onset, with temperature ≥ 101° F (38.3° C), headache, backache, chills, vomiting, abdominal pain, prostration Enanthem on oropharyngeal mucosa, followed by rash on face, forearms, distal extremities, then trunk; lesions most concentrated on face and distal extremities Lesions evolve slowly from macules to papules to deep-seated, firm, nodular, round, well-circumscribed vesicles or pustules to scabs over 1–2 days per stage; are in same stage of evolution on a given area of the body; may become umbilicated or confluent Hemorrhagic smallpox: bleeding into skin and mucous membranes Flat-type/malignant smallpox: lesions remain soft and flattened, coalesce Modified smallpox: less severe with fewer, more superficial and rapidly evolving lesions	Varicella (chickenpox); disseminated herpes zoster and simplex; drug eruptions; erythema multiforme; enteroviral infections; secondary syphilis; contact dermatitis; impetigo; scabies; molluscum contagiosum Hemorrhagic smallpox: meningococcemia, Rocky Mountan spotted fever, ehrlichiosis, gram-negative bacterial sepsis, severe acute leukemia Malignant smallpox: hemorrhagic chickenpox	Diagnostic testing at BSL-4 laboratory, including skin biopsy, electron microscopic examination of vesicular and pustular fluid, culture, PCR; serology Appropriate infection control precautions

Note: The clinical manifestations of infections acquired during a biologic attack may differ from those of naturally occurring infections. Clinicians should remain alert for compatible syndromes that vary from the descriptions given.
BSL-4—biosafety level 4 PCR—polymerase chain reaction

Table 4 Differentiating Features of Smallpox and Chickenpox

Clinical Feature	*Smallpox*	*Chickenpox*
Prodromal illness	Febrile prodrome lasting 1–4 days; patient appears ill or toxic	No or mild prodrome; patients typically do not appear ill
Appearance of lesions	Firm, round, well-circumscribed, deep-seated lesions; may be umbilicated	Superficial lesions
Stage of lesions on any one part of the body	Lesions are all at the same stage of development on a given area of the body	Lesions occur in crops with various stages of development evident on a given area of the body
Initial lesions	Oral mucosa, face, or forearms	Face, then trunk
Oral lesions	Early; may not be evident	May occur
Severity of illness	Typically severe	Typically not severe
Distribution of rash	Centrifugal: lesions concentrated on the face and extremities, with relative sparing of the trunk	Centripetal: lesions concentrated on the trunk with relative sparing of the face and extremities
Lesions on palms or soles	Lesions on palms and soles in majority of cases	Lesions on palms and soles uncommon
Rate of evolution of rash	Slow evolution of lesions from macules to papules to pustules over days	Rapid evolution from macules to papules to crusted lesions within 24 hours
Presence of pruritus	Lesions may be painful and are not usually pruritic until scabbing occurs	Often pruritic, typically not painful in the absence of secondary infection
Hemorrhagic lesions	Can occur	Can occur

traindications to vaccination should do so, because they would require immediate vaccination if the diagnosis of smallpox were confirmed. Vesicular or pustular fluid, scabs, punch biopsies of skin lesions, blood, and tissue from autopsy specimens should be obtained, packaged, and transported according to CDC protocol (http://www.bt.cdc.gov/labissues/PackagingInfo.pdf; http://www.bt.cdc.gov/documentsapp/SmallPox/RPG/index.asp).[36,39]

The CDC has developed a protocol in poster format for evaluating patients with an acute vesicular or pustular rash illness and for determining the risk of smallpox. The protocol, including color pictures of smallpox lesions, is available on the Internet at the CDC smallpox Web site (http://www.bt.cdc.gov/agent/smallpox/index.asp).

INFECTION CONTROL AND POSTEXPOSURE ISOLATION

In the event of a limited outbreak, patients should be admitted to the hospital and confined to rooms that are under negative atmospheric pressure and equipped with high-efficiency particulate air (HEPA) filtration. Standard, contact, and airborne precautions, including use of gloves, gowns, and masks, should be strictly observed. Unvaccinated personnel caring for patients suspected of having smallpox should wear fit-tested N95 or higher-quality respirators. Once successful vaccination is confirmed, care providers are no longer required to wear an N95 mask.[39] Patients should wear a surgical mask and be wrapped in a gown or sheet to cover the rash when they are not in a negative-airflow room. All laundry and waste should be placed in biohazard bags and autoclaved before being laundered or incinerated. Surfaces that may be contaminated with smallpox virus can be decontaminated with disinfectants that are used for standard hospital infection control, such as hypochlorite and quaternary ammonia.

Persons suspected of being infected with smallpox should be immediately isolated, and all their household members and others who have had face-to-face contact with the infected patient after the onset of fever should be vaccinated and placed under surveillance. Because persons who have had contact with an infected patient would not be contagious until the onset of rash, they should take their temperatures at least once daily, preferably in the evening. Any temperature higher than 101° F (38.3° C) during the 17-day period after the last exposure to the infected patient would suggest the possibility of the development of smallpox. This would be cause for immediate isolation until the diagnosis can be determined clinically, by laboratory examination, or both.

In the event of an outbreak, the following high-risk groups should be given priority for vaccination: (1) persons exposed to the initial release of the virus; (2) contacts of suspected or confirmed smallpox patients; (3) personnel who are directly involved in medical or public health evaluation of suspected or confirmed smallpox patients, as well as the care or transportation of such patients; (4) laboratory workers involved in the collection or processing of possible smallpox specimens; (5) other persons who may be in contact with infectious material, such as hospital laundry, medical waste, and mortuary workers; (6) other groups essential to response activities, such as law enforcement, emergency response, or military personnel; and (7) all persons in a hospital where there is a smallpox patient who is not isolated appropriately. Employees for whom vaccination would be contraindicated (see below) should be furloughed.[36,39]

Smallpox Vaccine

Vaccinia vaccine does not contain smallpox (variola) virus. The only currently available vaccine was prepared from calf lymph with a seed virus derived from the New York City Board of Health (NYCBOH) strain of vaccinia virus (Dryvax vaccine, Wyeth Laboratories, Marietta, Pennsylvania). A supply of licensed Dryvax vaccine was used in 2003 to immunize smallpox health care and public health teams as part of the National Smallpox Vaccination Plan. A reformulated vaccine, produced by using cell-culture techniques, is being developed.

The immune status of those vaccinated more than 27 years ago is not clear. Studies have demonstrated persistence of T cell and humoral responses, but absolute levels of neutralizing antibodies decline substantially during the first 5 to 10 years after vaccination. Interpretation of the data is complicated by the fact that the laboratory correlates of immunologic protection against smallpox infection are not currently established. Epidemiologic studies demonstrate that an increased level of protection against smallpox persists for less than 5 years after primary vaccination, and substantial but waning immunity can persist for longer than 10 years. Antibody levels after revaccination can remain high longer, conferring a greater period of immunity than occurs after primary vaccination alone.[32,35]

Complications of smallpox vaccination Current data on complication rates after primary vaccination are derived from observations made when smallpox vaccine was in routine use in the United States, over 30 years ago, and from the initial reports from the 2002 to 2003 United States military and civilian smallpox vaccination programs, involving 450,293 and 38,257 vaccinees, respectively. In these vaccination programs from 2002 to 2003, licensed undiluted Dryvax vaccine was used.[32,40-42] No cases of eczema vaccinatum or progressive vaccinia or vaccine-attributable deaths occurred. These reports suggest that for selected healthy vaccinees who have been carefully screened for vaccine contraindications and educated about adverse effects, adverse-event rates can be comparable to or less than historical rates. Should smallpox vaccine need to be administered to the general public, higher rates of vaccine complications could occur, given the increased number of persons with vaccine contraindications, including atopic dermatitis and other medical conditions, as well as the number of persons receiving medications that compromise the immune system. Contact transmission of vaccinia from immunized persons to close contacts, particularly those at risk for serious adverse events, is of particular concern.[43] Twenty-one cases of contact transmission were reported from the military program (47 per million vaccinees); no cases of transmission from health care providers to patients occurred.

Cell-mediated immunity is important in controlling smallpox and vaccinia infection. However, the level of immunosuppression that correlates with increased risk of adverse events is not known with certainty. In HIV-infected persons, progressive vaccinia infection after smallpox vaccination is the adverse event of greatest concern (see below); the risk presumably correlates with CD4 cell count.[44] During the 2002 to 2003 United States military smallpox vaccination program, 10 service members with HIV infection and CD4 cell counts ranging from 303 to 752 cells/μl were inadvertently vaccinated before recognition of their HIV infection. All 10 responded successfully to vaccination without adverse events.[40]

Moderate and severe complications of vaccinia vaccination include eczema vaccinatum, generalized vaccinia, progressive vaccinia, and postvaccinial encephalitis. These complications are rare but are at least 10 times more common after primary vaccination than after revaccination; they occur more frequently in infants than in older children and adults.

The most common complication of smallpox vaccination, occurring in 529.2 cases per million doses, is localized vaccinia infection resulting from inadvertent transfer (autoinoculation) of vaccinia from the vaccination site to other parts of the body. In addition, transmission of vaccinia virus can occur when a recently vaccinated person has contact with a susceptible person; in one study, approximately 30% of eczema vaccinatum cases were persons who had had such contact.[32,45] Inadvertent transfer of vaccinia from the vaccination site to other parts of the body can be prevented by careful hand washing after touching the vaccination site and by keeping the site covered.

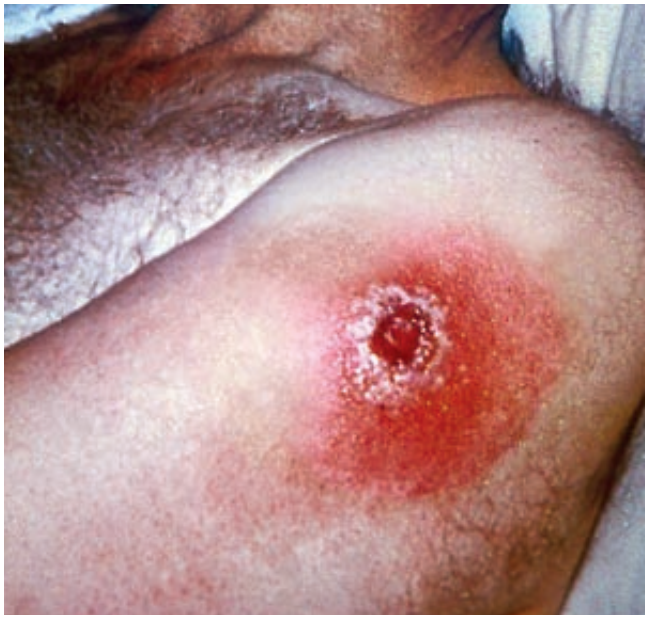


Figure 2 **Progressive vaccinia (vaccinia necrosum) at the site of smallpox vaccination in a 64-year-old man.**

Eczema vaccinatum (38.5/million doses) is a localized or systemic dissemination of vaccinia virus that occurs in persons who have eczema or a history of eczema or other chronic or exfoliative skin conditions (e.g., atopic dermatitis). Illness is usually mild and self-limited but can be severe or fatal. Severe cases have also been observed in persons with active eczema or a history of eczema, after contact with recently vaccinated persons.

Generalized vaccinia (241.5/million doses) is characterized by a vesicular rash of varying extent that can occur in persons without underlying illness. The rash is generally self-limited and requires minor or no therapy except in patients whose condition might be toxic or who have serious underlying immunosuppressive illnesses.

Progressive vaccinia (vaccinia necrosum, 1.5/million doses) is a severe, potentially fatal illness characterized by progressive necrosis in the area of vaccination, often with metastatic lesions [*see Figure 2*]. It has occurred almost exclusively in persons with cellular immunodeficiency.

The most common serious complication is postvaccinial encephalitis (12.3/million doses). It occurs mostly in infants younger than 1 year and, less often, in adolescents and adults receiving a primary vaccination. Rates of this complication were influenced by the strain of virus used in the vaccine and were higher in Europe than in the United States. The principal strain of vaccinia virus used in the United States—the NYCBOH strain—was associated with the lowest incidence of postvaccinial encephalitis. Approximately 15% to 25% of affected vaccinees with this complication die, and 25% have permanent neurologic sequelae.

Fatal complications caused by vaccinia vaccination are rare, with approximately one death per million primary vaccinations and 0.25 deaths per million revaccinations. Death is most often the result of postvaccinial encephalitis or progressive vaccinia.

In 2003, cases of myocarditis, pericarditis, or both (myopericarditis) occurred at higher than expected rates in both the civilian and the military smallpox vaccination programs. Cases of cardiac ischemic events, including myocardial infarction and angina, were also reported in persons who had received vaccinations; however, the rates of ischemic events were not clearly elevated above expected background rates, and the association between smallpox vaccination and cardiac ischemic events is unclear.[46,47]

Among 450,293 United States military service members vaccinated from December 2002 to May 2003, 37 cases of suspected, probable, or confirmed myopericarditis were observed; the rate was 1:12,195 among male primary vaccinees 21 to 33 years of age. No cases were reported among 132,836 previously vaccinated vaccinees.[40] Through May 9, 2003, 21 cases of myopericarditis (90% of revaccinees) were reported among 36,217 civilian vaccinees, representing a rate of 1:1,700 vaccinees (the rate decreases to 1:36,000 vaccinees when vaccinees without elevations in cardiac enzyme levels are excluded).[46]

Symptoms of myopericarditis after smallpox vaccination began 7 to 19 days after vaccination (range, 1 to 42 days). Symptoms included prodromal myalgias, arthralgias, or both, as well as subsequent pleuritic precordial chest pain with variable shortness of breath, dry cough, or both. Electrocardiogram findings varied; such findings included ST segment and T wave abnormalities and dysrhythmias, including paroxysmal atrial fibrillation, atrial ectopy, supraventricular tachycardia, and ventricular ectopy. ECG findings included pleural effusion and wall motion abnormalities; these findings were not seen in all patients. All patients among the United States military vaccinees, as well as one civilian patient, were reported to have had elevated cardiac enzyme levels. All the patients who were reported to have myopericarditis recovered. In cases where alternative etiologies were sought, none were established, and in no case was a virologic diagnosis made. The biologic mechanism for myopericarditis after smallpox vaccination is not established; it may involve the viral cytopathic effect, an immune-mediated reaction, or both.[47] A surveillance case definition for myopericarditis has been developed to monitor smallpox vaccine adverse events (this definition is not to be used for clinical diagnosis). That survillance case definition is available on the Internet.[46]

In response to these reports, new cardiovascular screening and exclusion criteria were published for use of smallpox vaccine in persons not exposed to smallpox.[48] Vaccinees experiencing chest pain, shortness of breath, or other symptoms of cardiac disease within 2 weeks of smallpox vaccination are advised to seek medical attention. Although myopericarditis after smallpox vaccination had been previously reported, a causal association with the vaccine had not been recognized.

Focal folliculitis and generalized folliculitis were reported after smallpox vaccination in 7.4% and 2.7%, respectively, of 148 primary vaccinees enrolled in a clinical trial of a smallpox vaccine that is not currently in use but that was derived from the same strain of vaccinia virus as the NYCBOH strain used in Dryvax.[49] Follicular erythematous papules developed from 9 to 11 days after vaccination; these papules progressed to pustules that lasted 3 to 5 days. Lesions appeared primarily in areas where there were larger numbers of hair follicles or sebaceous glands (e.g., the extremities, face, and back); concurrent lesions were in different stages of development. Vaccinia was not isolated from specimens that were available for culture. This new observation is significant because of the potential for this apparently benign condition to be confused with generalized vaccinia.

Contraindications Groups at special risk for complications include persons with eczema or other acute, chronic, or exfoliative skin conditions; patients with immune system suppression, including leukemia, lymphoma, or generalized malignancy who are receiving therapy with alkylating agents, antimetabolites, radiation, or large doses of corticosteroids; patients with HIV infection and persons with hereditary immune disorders; children younger than 1 year; and women who are pregnant or breast-feeding (because of the risk of contact transmission to the child). On the basis of reports of adverse cardiac events in 2003, persons should currently be excluded from preexposure smallpox vaccination if they have known underlying heart disease or three or more known cardiac risk factors (i.e., hypertension, smoking, diabetes, hypercholesterolemia, and heart disease occurring by 50 years of age in a first-degree relative). Deferring vaccination of persons with active inflammatory disease of the eye requiring steroid treatment is advised until the condition resolves and treatment is complete.[48] In persons with contraindications who require vaccination because of exposure to smallpox virus from a bioterrorist attack, the risk of complications can be reduced by giving vaccinia immune globulin (VIG; see below) simultaneously with the vaccine. However, current stores of VIG are insufficient to allow its prophylactic use. Even if VIG is not available, vaccination may still be warranted, given the far higher risk of an adverse outcome from smallpox than from vaccination.

Current information about smallpox vaccine, including recommendations for vaccine use, contraindications, screening of potential vaccinees, prevention of contact transmission, and management of adverse effects, is available on the CDC Web site (http://www.cdc.gov).

Vaccinia immune globulin Complications of vaccinia vaccination can be prevented or treated with VIG, which is an isotonic sterile solution of the immunoglobulin fraction of plasma from persons vaccinated with vaccinia vaccine. For prophylactic use, in persons with contraindications who require vaccination, VIG is given along with vaccinia vaccine.[32] Very large amounts are required: VIG is administered intramuscularly in a dose of 0.3 ml/kg (e.g., 22.5 ml I.M. for a 75 kg patient) At present, however, supplies of VIG are so limited that its use should be reserved for treatment of patients with the most serious vaccine complications.

For treatment of vaccinia vaccination complications, VIG is administered intramuscularly; 0.6 ml/kg is given in divided doses over a 24- to 36-hour period. A repeat dose may be given 2 to 3 days later if improvement does not occur. VIG is effective for treatment of eczema vaccinatum and certain cases of progressive vaccinia; it might be useful also in the treatment of ocular vaccinia resulting from inadvertent implantation. VIG is contraindicated for the treatment of vaccinial keratitis. VIG is recommended for severe generalized vaccinia if the patient is extremely ill or has a serious underlying disease. VIG provides no benefit in the treatment of postvaccinial encephalitis and has no role in the treatment of smallpox.[32,36]

Anthrax

Anthrax is a zoonotic disease caused by the spore-forming bacterium *Bacillus anthracis,* a large, nonmotile, nonhemolytic, gram-positive rod [*see 7:IV Infections Due to Gram-Positive Bacilli*]. The organism is distributed worldwide in soil. Animals, pri-

marily herbivores, become infected through grazing in contaminated areas. Under natural conditions, humans contract the disease after close contact with infected animals or contaminated animal products such as hides, wool, or meat.[50] Hardy spores resistant to heat and environmental degradation are the usual infective form. The spores develop in response to exposure to ambient air. On exposure to favorable, nutrient-rich environmental conditions such as tissues or blood of an animal or human host, the spores germinate, producing vegetative cells.[51]

CLASSIFICATION AND EPIDEMIOLOGY

Anthrax occurs in three clinical forms in humans: inhalational, cutaneous, and gastrointestinal. In a biologic attack, aerosol exposure to anthrax spores would be most likely.[33] Only 18 cases of inhalational anthrax were reported in the United States in the 20th century, none of them after 1976. Sixteen of these cases were attributable to an industrial source of infection, and two cases were laboratory associated.[52] Before 2001, exposure to powdered anthrax spores in an envelope or package was not thought to be an efficient means of causing inhalational disease. However, exposure to anthrax spores sent through the United States mail in the 2001 anthrax attack resulted in 11 cases of inhalational anthrax and 11 cases of cutaneous disease.[21,53,54] Recent research has demonstrated the unanticipated potential for significant dispersion of respirable aerosol particles of spores through opening of a contaminated envelope.[55] In addition, expected clinical findings based on previous experience with naturally occurring anthrax infections did not entirely correspond to the clinical presentation in persons exposed to anthrax in the context of a biologic attack, although there was considerable overlap between the two.

Cutaneous anthrax accounts for the majority of naturally occurring anthrax cases worldwide. It results from inoculation of spores subcutaneously through a cut or abrasion.[56] Given that cutaneous anthrax cases occurred during the 2001 anthrax outbreak, it is possible that a bioterrorist attack could be detected through recognition of cutaneous anthrax cases.[21] Gastrointestinal and oropharyngeal anthrax occur in rural parts of the world where anthrax is endemic. They result from ingestion of meat contaminated with spores or large numbers of vegetative cells.[57] No cases of gastrointestinal anthrax occurred during the 1979 accidental release of anthrax from a military facility in Sverdlovsk, Russia, in which 77 inhalational cases occurred, or during the 2001 outbreak in the United States. Because of the logistic difficulty of effectively contaminating food and water supplies, it is thought that this form of anthrax would be less likely to occur as a result of a biologic attack.[33]

PATHOPHYSIOLOGY

Anthrax is a toxin-mediated disease. In inhalational anthrax, 1 to 5 μm particle–bearing spores are deposited in the terminal airways or alveoli, phagocytized by alveolar macrophages, and transported to mediastinal and peribronchial lymph nodes. Spores may stay in the mediastinal lymph nodes for extended periods and can germinate for up to 60 days or longer.[58] Cases of inhalational anthrax occurred up to 43 days after exposure in the Sverdlovsk outbreak.[59] Spores germinate into vegetative cells, which escape from the macrophages, multiply in the lymphatics, and ultimately gain access to the bloodstream, where they can reach high concentrations (107 to 108 organisms per milliliter of blood). Hemorrhagic meningitis is a complication of bacteremic spread; it develops in up to one half of cases.

In anthrax, tissue damage is mediated by two toxins: edema toxin and lethal toxin. These two toxins are composed of various combinations of edema factor, lethal factor, and protective antigen. The three components of edema toxin and lethal toxin are produced by vegetative cells. Vegetative cells also produce an antiphagocytic capsule that is necessary for virulence.[60] Lethal toxin is a combination of lethal factor and protective antigen that interferes with cellular protein synthesis; it causes macrophages to release tumor necrosis factor and interleukin-1. In severe cases, it contributes to sudden death from toxemia. Edema toxin is a combination of edema factor and protective antigen that causes increased cellular levels of cyclic adenosine monophosphate (cAMP) and altered water homeostasis, resulting in massive edema. Together, edema toxin and lethal toxin cause edema, hemorrhage, necrosis, and shock. In cutaneous and gastrointestinal anthrax, toxin production results in a similar pathophysiologic process that causes edema and hemorrhagic necrosis in the skin and gastrointestinal mucosa, respectively. Pathologic studies of eight cases of inhalational anthrax related to the bioterrorism attack of 2001 demonstrated hemorrhagic mediastinitis without pneumonia; pulmonary infiltrates seen on chest radiographs corresponded to pulmonary edema and hyaline membrane formation. Large numbers of anthrax bacilli and cell wall and capsular antigens were observed in pleural tissues—findings that related to persistent pleural effusions seen clinically.[61]

INHALATIONAL ANTHRAX

Clinical Presentation and Diagnosis

Recent information on the clinical manifestations of inhalational anthrax from the 2001 anthrax outbreak both confirms many of the features reported in naturally occurring anthrax cases and reveals unanticipated differences.[52,58,62-64] The infectious dose of anthrax is not known with certainty. Animal data suggest that the median lethal dose (LD_{50}, which is the dose sufficient to kill 50% of exposed subjects) is 2,500 to 55,000 inhaled spores. Data from naturally occurring cases and from two cases in the 2001 outbreak suggest that the infectious dose may be very low in some persons, particularly those with underlying pulmonary disease.[58,63,65]

Clinical symptoms develop rapidly after germination of anthrax spores. The incubation period for inhalational disease is most commonly reported as 1 to 6 days but may be prolonged by antibiotic administration or, presumably, a low infectious dose.[66,67] In the 2001 anthrax outbreak, the median incubation period was 4 days (range, 4 to 6 days) for the six cases in which it could be calculated.

Inhalational anthrax has been described as a two-stage disease. The initial stage is a nonspecific, flulike illness lasting from several hours to a few days. In the 2001 bioterrorism-associated anthrax cases, symptoms at presentation included some combination of fever, chills, weakness, cough, dyspnea, chest discomfort, nausea or vomiting, myalgia, and headache. Profound malaise and drenching sweats were prominent symptoms, and most patients reported nausea or vomiting. Tachycardia was the most consistent clinical sign. Fever was documented in the majority of patients, although it was sometimes low grade. Sore throat, nasal symptoms, and abdominal pain were common. Classically, the initial stage is followed 1 to 3 days later, sometimes after brief improvement, by the rapidly progressive second stage, characterized by fever, dyspnea, diaphoresis,

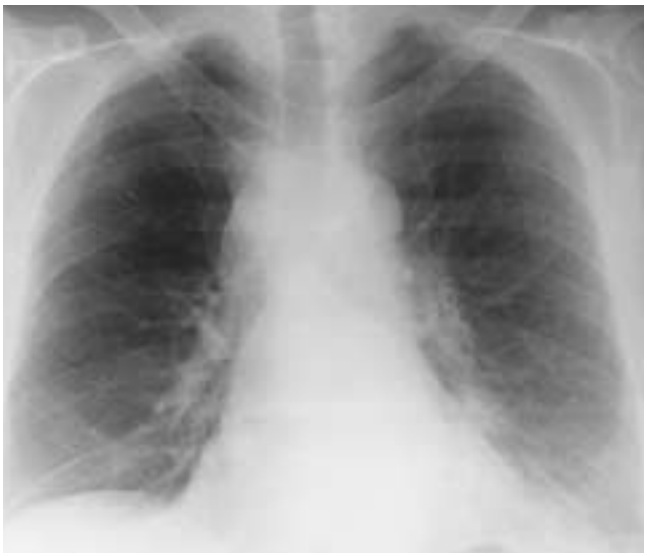


Figure 3 **Chest x-ray of a patient with inhalational anthrax showing mediastinal widening and a small left pleural effusion.**

cyanosis, and shock. In the 2001 cases, no brief improvement between stages was observed.

Laboratory studies are nonspecific or unremarkable during the early stage of disease. The majority of patients experienced elevations in hepatic transaminase levels; neutrophilia without leukocytosis; hyponatremia; hypoalbuminemia; and hypocalcemia.[62,68] Chest x-rays were abnormal on initial presentation in all 11 recent cases, although one patient had subtle abnormalities that were initially interpreted as normal. Only seven patients had the classic finding of mediastinal widening [*see Figure 3*]. Pleural effusions were present in all cases. These effusions were often small on presentation and were progressive, requiring drainage in the majority of patients. In contrast to previous descriptions, seven patients had pulmonary infiltrates consistent with pneumonia at presentation, and one patient was thought to have heart failure with pulmonary congestion. Other abnormalities included paratracheal and hilar fullness. The CT scan was valuable in further characterizing abnormalities in the lungs and mediastinum and was more sensitive than the chest x-ray in revealing mediastinal changes. Blood cultures can be diagnostic, although appropriate antibiotic therapy rapidly reduces the likelihood of isolating the organism. In the 2001 cases, *B. anthracis* was isolated from blood cultures obtained before antibiotic therapy was given, but not from those obtained afterward.

The initial manifestations of inhalational anthrax are nonspecific and are consistent with flulike illnesses caused by a variety of respiratory viruses, as well as with community-acquired bacterial infections. Adults can average one to three episodes of flulike illness a year, and millions of cases occur throughout the United States.[69] Because of the high frequency of flulike illnesses and the low likelihood of inhalational anthrax in a given patient, a combination of epidemiologic, clinical, and (if indicated) laboratory testing should be used to evaluate potential cases of inhalational anthrax [*see Figure 4*]. According to CDC guidelines, consideration of inhalational anthrax hinges on a history of exposure or occupational/environmental risk within 2 to 5 days before illness onset.[19] Whenever possible, exposure and risk determinations should be made in consultation with public health authorities before initiating treatment or preventive therapy.

The clinical presentation of inhalational anthrax may be difficult to distinguish from that of community-acquired pneumonia.[68] According to the CDC, diagnostic testing for anthrax should be done in patients whose signs and symptoms are consistent with anthrax and when one or more of the following conditions are present: a history of a recent anthrax case or outbreak in the community; a credible threat of anthrax exposure, as determined by law enforcement and public health authorities; a cluster of anthraxlike cases characterized by rapid deterioration. Anthrax should also be considered in any patient with compatible symptoms and rapid deterioration. Alternatively, a set of five symptoms that are compatible with inhalational anthrax, in combination with fever and tachycardia, have been suggested as criteria for pursuing additional diagnostic evaluation for patients with possible inhalational anthrax who are without epidemiologic risk factors.[70] All cases of *suspected* anthrax should be reported immediately to local or state public health authorities and the hospital epidemiologist (http://www.statepublic health.org). The clinical laboratory should also be alerted when diagnostic specimens of suspected anthrax are submitted to ensure that appropriate precautions are taken to protect laboratory staff, facilitate proper evaluation of the isolate, and expedite confirmatory testing at the nearest laboratory that belongs to the public health Laboratory Response Network.[6]

There is no rapid screening test to diagnose inhalational anthrax in its early stages. In persons with a compatible clinical illness for whom there is a heightened suspicion of anthrax based on clinical and epidemiologic data, the appropriate initial diagnostic tests are a chest x-ray or chest CT scan, or both, and culture and smear of peripheral blood. On chest x-rays, the posteroanterior and lateral view may be more sensitive than the anteroposterior (portable) view in detecting pulmonary abnormalities. Mediastinal widening or hyperdense mediastinal lymphadenopathy (secondary to hemorrhagic lymph nodes) on a nonenhanced CT scan should raise the suspicion of pulmonary anthrax [*see Figure 5*]. Most persons with flulike illnesses do not have radiologic findings of pneumonia; such findings occur most often in the very young, the elderly, and persons with chronic lung disease.

Pleural fluid and cerebrospinal fluid, as well as biopsy specimens taken from the pleura and lung, are also potentially useful for culture and other testing when disease is present in these sites, whereas sputum culture and Gram stain are unlikely to be useful. In highly suspicious cases, local or state health departments can arrange for additional diagnostic testing, including immunohistochemical staining and PCR at the CDC. Serologic testing is not useful in clinical management but may be used in epidemiologic investigations. Similarly, nasal swabs are of potential value in epidemiologic investigations for determining the route and extent of spread of anthrax in a population, but they have no role in clinical management.

A rapid influenza test can be used when influenza itself is a consideration in a patient with flulike illness, but these kits have limited value because their sensitivity can be relatively low (45% to 90%). However, rapid influenza testing with viral culture can help indicate whether influenza viruses are circulating among certain populations, and this epidemiologic information can be useful in diagnosing flulike illnesses.[69]

Treatment

Early intravenous antibiotic treatment may improve survival in inhalational anthrax.[71] In contrast to the reported case-fatality

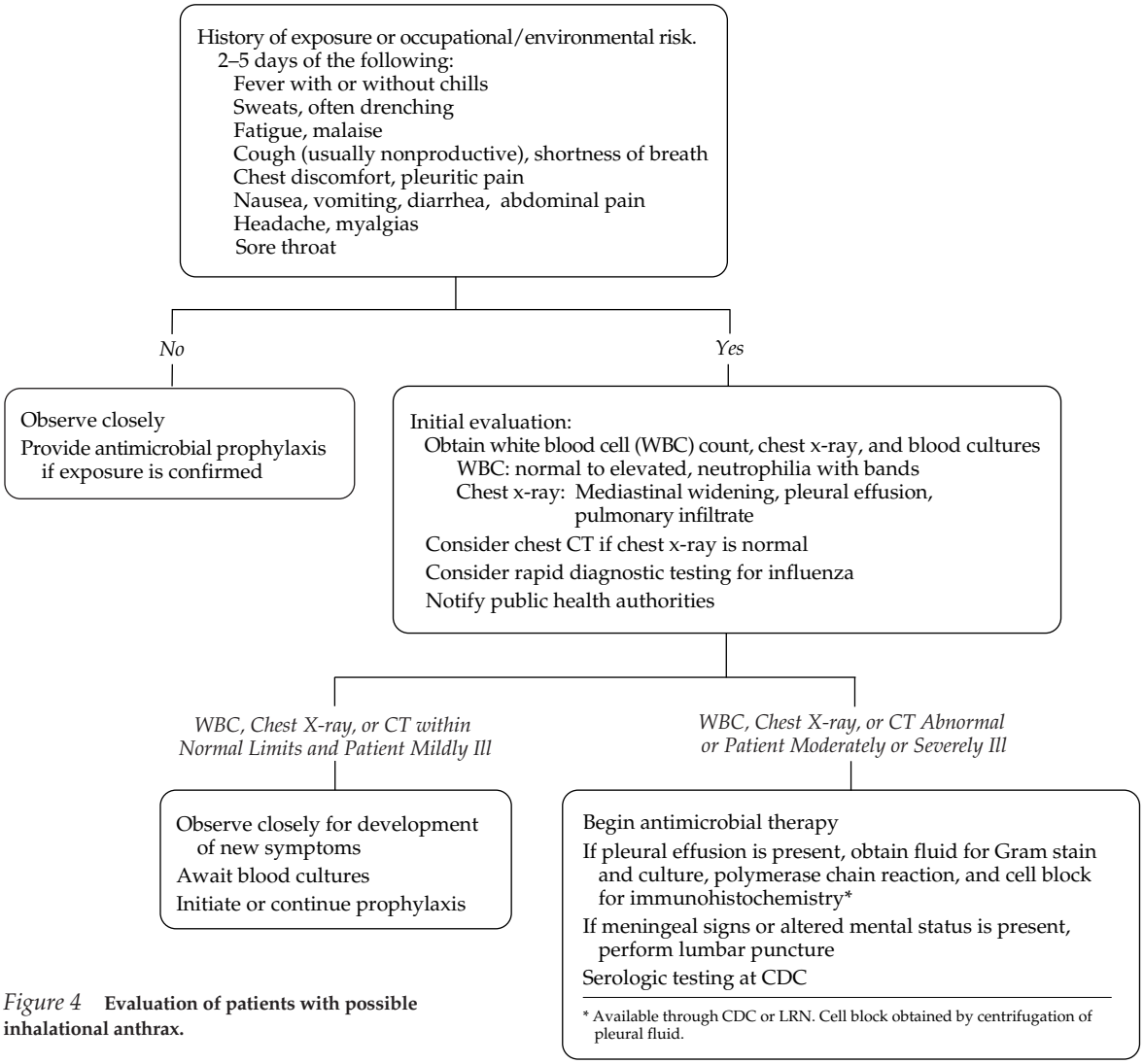


Figure 4 **Evaluation of patients with possible inhalational anthrax.**

rate of 85% for 20th-century inhalational anthrax cases, 6 of 11 patients in the 2001 outbreak survived; all the survivors presented during the initial phase of the illness and received treatment the same day with antibiotics active against *B. anthracis.* Fatal cases occurred in patients who had severe disease by the time they first received antibiotics with activity against *B. anthracis.* Aggressive supportive care—including attention to fluid, electrolyte, and acid-base disturbances and drainage of pleural effusions—also played an important role in treatment.[62]

Current CDC treatment recommendations and related guidelines and information can be obtained at http://www.bt.cdc.gov/HealthProfessionals/index.asp. Before initiating treatment, clinicians should review this site to stay informed of revisions and updates. The Working Group on Civilian Biodefense has published similar recommendations with a detailed accompanying text.[58]

At present, intravenous ciprofloxacin or doxycycline plus one or two additional antimicrobials with in vitro activity against *B. anthracis* are recommended for initial empirical treatment [*see Table 5*]. Antibiotic therapy should be modified according to the results of antimicrobial susceptibility testing to ensure that the most effective and least toxic regimen is used. The duration of antimicrobial therapy should be at least 60 days. Once clinical improvement occurs, it may be possible to complete the course of treatment with one or two agents given orally. Corticosteroid therapy has been suggested as adjunct therapy for inhalational anthrax associated with extensive edema, respiratory compromise, and meningitis.[21,56,58]

Prevention

Ciprofloxacin and doxycycline are recommended first-line agents for prophylaxis in persons exposed to inhalational anthrax. In vivo data suggest that other fluoroquinolone antibiotics would have efficacy equivalent to that of ciprofloxacin.[58] High-dose amoxicillin is an option when ciprofloxacin or doxycycline is contraindicated [*see Table 6*]. Postexposure prophylaxis should continue for at least 60 days.[72] Given the uncertainty about the length of time viable spores can persist in the lungs, patients should be instructed to seek prompt medical evaluation if symptoms compatible with anthrax develop after discontinuance of postexposure prophylaxis. Because of uncertainty about the length of time that anthrax spores can remain viable in the lungs, the United States Department of Health and Human Services made two additional options available for preventive treatment for persons exposed to inhalational anthrax in the 2001 outbreak. These options were to follow a 60-day course of antibiotic treatment with either (1) an additional 40 days of an-

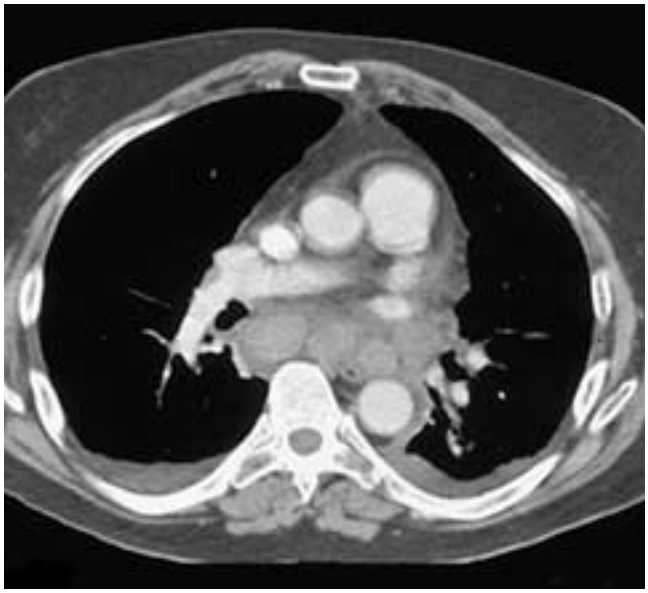


Figure 5 **CT scan of the chest of a patient with inhalational anthrax showing mediastinal lymphadenopathy and small bilateral pleural effusions.**

tibiotic treatment or (2) an additional 40 days of antibiotic treatment plus three doses of anthrax vaccine over a 4-week period.[73]

Anthrax vaccine The only licensed human anthrax vaccine available in the United States is anthrax vaccine adsorbed (AVA). This is an inactivated, cell-free filtrate of a nonencapsulated attenuated strain of *B. anthracis* (BioPort Corporation, Lansing, Michigan).[67] Primary vaccination consists of three subcutaneous injections at 0, 2, and 4 weeks and three booster vaccinations at 6, 12, and 18 months. To maintain immunity, the manufacturer recommends an annual booster injection. The basis for this recommended schedule of vaccination is not well defined.

Vaccination of adults with the licensed vaccine induced an immune response, as measured by indirect hemagglutination, in 83% of vaccinees 2 weeks after the first dose and in 91% of vaccinees who received two or more doses. Approximately 95% of vaccinees undergo seroconversion after three doses, with a fourfold rise in titers of IgG against protective antigen (the principal antigen responsible for inducing immunity). However, the precise correlation between antibody titer (or concentration) and protection against infection is not defined. The vaccine has shown efficacy in experiments involving animal models of inhalational anthrax in preexposure settings and, in combination with antibiotics, in postexposure settings.[58,69]

Anthrax vaccine is considered acceptably safe by the Advisory Committee on Immunization Practices and the Institute of Medicine.[67,74] Supplies of anthrax vaccine are limited and are held by the United States Department of Defense. A combination of antibiotics and anthrax vaccine, if available, is recommended for exposed persons after a biologic attack.[58,75,76] At this time, preexposure use of anthrax vaccine is recommended only for certain laboratory workers and others at occupational risk for repeated exposures to *B. anthracis* spores; it is not recommended for the general public.

CUTANEOUS ANTHRAX

After an incubation period of approximately 7 days (range, 1 to 12 days), the primary lesion of cutaneous anthrax appears as a nondescript, painless, pruritic papule, usually on an exposed area such as the face, head, neck, or upper extremity. The papule enlarges and develops a central vesicle or bullae with surrounding brawny, nonpitting edema. The central vesicle enlarges and ulcerates over 1 to 2 days, becoming hemorrhagic, depressed, and necrotic and leading to a central black eschar [*see Figure 6*]. Satellite vesicles may be present. The eschar dries and falls off over the next 1 to 2 weeks. The findings of a painless lesion and edema out of proportion to the size of the lesion and the fact that pustules are rarely present in cutaneous anthrax are clinically useful. Tender regional lymphadenopathy, fever, chills, and fatigue may occur. Systemic disease has been

Table 5 Treatment of Inhalational Anthrax

Patients	*Medication and Dosage*	*Comments*
Adults, including pregnant women and immunocompromised persons	Ciprofloxacin, 400 mg I.V., q. 12 hr *or* Doxycycline, 100 mg I.V., q. 12 hr *and* One or two additional antimicrobials*	If meningitis is suspected, doxycycline may be less optimal because of poor central nervous system penetration Modify regimen on the basis of susceptibility testing of isolate; can switch to p.o. after patient is clinically stable; continue treatment for at least 60 days Consider corticosteroids for meningitis, severe edema, or respiratory compromise
Children, including those who are immunocompromised	Ciprofloxacin, 10–15 mg/kg I.V., q. 12 hr, not to exceed 1 g/day *or* Doxycycline If > 8 yr and > 45 kg, give adult dosage If ≤ 8 yr or if > 8 yr but ≤ 45 kg, give 2.2 mg/kg q. 12 hr (maximum, 200 mg/day) *and* One or two additional antimicrobials*	If meningitis is suspected, doxycycline may be less optimal because of poor central nervous system penetration Modify regimen on the basis of susceptibility testing of isolate; can switch to p.o. after patient is clinically stable; continue treatment for at least 60 days Consider corticosteroids for meningitis, severe edema, or respiratory compromise

Note: Treatment recommendations may change over time and according to antimicrobial susceptibility test results during a biologic attack and to availability of selected antimicrobial agents. Before initiating treatment, clinicians should consult with an infectious disease specialist and public health authorities and should check for revisions and updates at http://www.bt.cdc.gov/index.asp. This information is adapted from CDC and Working Group on Civilian Biodefense recommendations and may not represent FDA-approved uses.
*Other agents with in vitro activity against anthrax include rifampin, vancomycin, penicillin, ampicillin, chloramphenicol, imipenem, clindamycin, and clarithromycin.

Table 6 Postexposure Prophylaxis for Anthrax in the Setting of a Bioterrorist Attack

Patients	*Medication*	*Comments*
Adults, including immunocompromised persons	Ciprofloxacin, 500 mg p.o., q. 12 hr *or* Doxycycline, 100 mg p.o., q. 12 hr *or, if strain proved susceptible,* Amoxicillin, 500 mg p.o., q. 8 hr	Give prophylaxis for at least 60 days
Pregnant women	Ciprofloxacin, 500 mg p.o., q. 12 hr *or, if strain proved susceptible,* Amoxicillin, 500 mg p.o., q. 8 hr	Give prophylaxis for at least 60 days
Children, including those who are immunocompromised	Ciprofloxacin, 10–15 mg/kg p.o., b.i.d., not to exceed 1g/day *or, if strain proved susceptible,* Amoxicillin If ≥ 20 kg: 500 mg p.o., q. 8 hr If < 20 kg: 80 mg/kg/day p.o., in divided doses q. 8 hr (maximum, 500 mg/dose)	Give prophylaxis for at least 60 days

Note: Prophylaxis recommendations may change over time and according to antimicrobial susceptibility test results during a biologic attack and to availability of selected antimicrobial agents. Before initiating prophylaxis, clinicians should consult with an infectious disease specialist and public health authorities and should check for revisions and updates at http://www.bt.cdc.gov/index.asp. This information is adapted from CDC and Working Group on Civilian Biodefense recommendations and may not represent FDA-approved uses.

reported to have a mortality of 20% if untreated. Cutaneous anthrax of the face or neck may lead to respiratory compromise from massive edema.[56,58,60,77]

The differential diagnosis of cutaneous anthrax includes other causes of eschar and ulceration and the ulceroglandular syndrome.[75] Guidelines for the diagnosis of cutaneous anthrax have been published by the American Academy of Dermatology (http://www.aad.org/BioInfo/anthrax.html).

For patients with the typical appearance and progression of cutaneous anthrax, a Gram stain and culture of the skin lesion should be obtained using a dry swab for unroofed vesicle fluid and a moist swab for the base of the ulcer and edges underneath the eschar. Blood cultures are also recommended. If the patient is taking antimicrobial drugs or if the Gram stain and culture are negative for *B. anthracis* or clinical suspicion remains high, two punch biopsies for culture (with the specimen placed in saline) and immunohistochemical staining should be performed; PCR (with the specimen placed in formalin) should be performed, or both should be considered [*see Figure 7*]. Immunohistochemical staining and PCR testing at the CDC should be arranged through local public health authorities.[21,58]

Management

Antibiotic treatment is curative in cutaneous anthrax and can be initiated pending confirmation of anthrax infection. Ciprofloxacin and doxycycline are first-line agents for the empirical treatment of cutaneous anthrax and may be administered orally. Intravenous therapy with multiple drugs, as for inhalational anthrax (see above), is recommended for patients with signs of systemic involvement, extensive edema, or lesions of the face and neck.[71]

GASTROINTESTINAL AND OROPHARYNGEAL ANTHRAX

Symptoms appear 2 to 5 days after ingestion of contaminated food and include nausea, vomiting, fever, malaise, and abdominal pain. Severe bloody diarrhea with rebound abdominal tenderness develops. Ulcerative lesions occur primarily in the terminal ileum and cecum. Gastric ulcers with hematemesis, hemorrhagic mesenteric lymphadenitis, and marked ascites may occur. Mediastinal widening has also been reported with gastrointestinal anthrax. Morbidity results from blood loss, fluid and electrolyte imbalances, and shock. The case-fatality rate is reportedly greater than 50%; death results from toxemia or intestinal perforation.[33,57,58]

Oropharyngeal anthrax is characterized by sore throat, fever, dysphagia, and marked edema and lymphadenitis. Ulcerative lesions may have an associated pseudomembrane. Specimens for diagnosis of gastrointestinal anthrax may include ascitic fluid for Gram stain and culture, blood cultures, and tissue samples from affected mucosal sites.

Treatment for gastrointestinal anthrax and oropharyngeal anthrax is the same as that for inhalational anthrax (see above).

INFECTION CONTROL

Person-to-person transmission of anthrax is not known to occur. Patients may be hospitalized in a standard hospital room

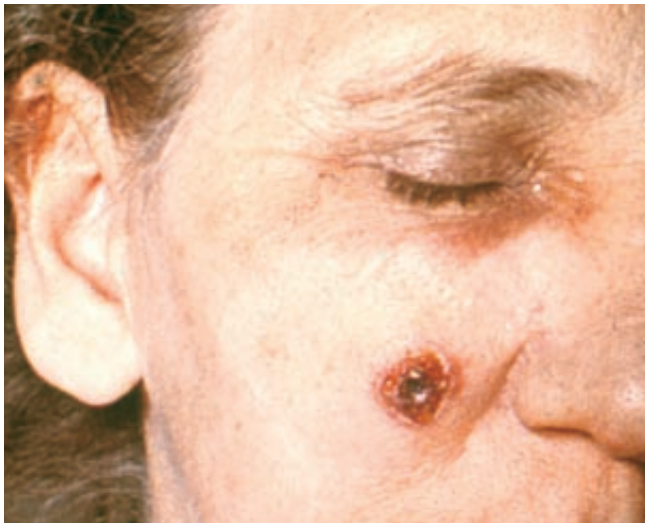
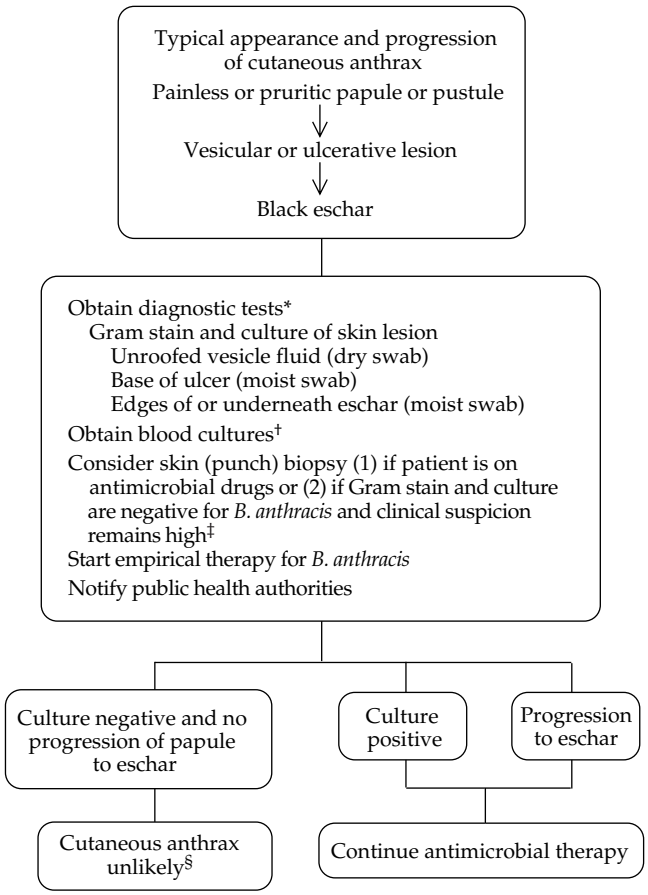


Figure 6 **Cutaneous anthrax lesion, 11 days old.**



* Serologic testing available at CDC may be an additional diagnostic technique for confirmation of cases of cutaneous anthrax.

† If blood cultures are positive for *B. anthracis*, treat with antimicrobials as for inhalational anthrax.

‡ Punch biopsy should be submitted in formalin to CDC. Polymerase chain reaction can also be done on formalin-fixed specimen. Gram stain and culture are frequently negative for *B. anthracis* after initiation of antimicrobials.

§ Continue antimicrobial prophylaxis for inhalational anthrax for 60 days if aerosol exposure to *B. anthracis* is known or suspected.

Figure 7 **Evaluation of patients with possible cutaneous anthrax.**

with standard barrier isolation precautions. No treatment is necessary for contacts of cases.

The microbiology laboratory should be notified upon suspicion of anthrax to ensure that appropriate precautions are taken under BSL-2 conditions when specimens are processed for culture.[78] Sporicidal solutions approved for use in hospitals and commercially available bleach or a 0.5% hypochlorite solution (1:10 dilution of household bleach) are effective for decontamination of contaminated areas. Precautions should be taken during autopsies, and cremation of human remains should be considered to prevent further transmission of disease.[24]

Plague

Plague is caused by the gram-negative coccobacillus *Yersinia pestis,* of the family Enterobacteriaceae. Wild rodents are the animal reservoir for the disease. Under natural conditions, plague is transmitted to humans by the bite of an infectious flea and, less frequently, by direct contact with infectious body fluids or tissues of an infected animal or by inhaling infectious droplets.[79] Plague has a long history of use and development as a biologic weapon, including the catapulting of plague victims' corpses over the walls of a besieged city in the 14th century. The most likely presentation after a biologic attack is primary pneumonic plague.[33] Additional information on plague, including the nonpneumonic forms (bubonic and septicemic plague), microbiology, and pathogenesis, is available elsewhere [*see 7:XI Infections Due to Brucella, Francisella, Yersina Pestis, and Bartonella*].

CLINICAL PRESENTATION

Plague is a severe febrile illness. Pneumonic plague, the most fatal form of the infection, can develop from inhalation of plague bacilli (primary pneumonic plague) or from hematogenous spread secondary to septicemic plague. Approximately 12% of cases of bubonic and primary septicemic plague develop into secondary pneumonic plague. Conversely, septicemic plague can be secondary to primary pneumonic plague.

The incubation period for pneumonic plague is typically 2 to 4 days (range, 1 to 6 days). Presenting symptoms typically include the acute onset of malaise, high fever, chills, headache, chest discomfort, dyspnea, and cough concomitant with or followed rapidly by clinical sepsis. Hemoptysis is a classic sign that should suggest plague in the appropriate clinical context, but sputum may be watery or purulent. Gastrointestinal symptoms may be prominent with pneumonic plague; these include nausea, vomiting, diarrhea, and abdominal pain. A cervical bubo is infrequently present.

The disease is rapidly progressive, with increasing dyspnea, stridor, and cyanosis. Rapidly progressive respiratory failure and sepsis within 2 to 4 days of onset of illness is typical of pneumonic plague. Abnormalities on chest x-ray are variable but frequently show bilateral patchy infiltrates or consolidation. The mortality for pneumonic plague is reported to be 57% and is extremely high when initiation of treatment is delayed beyond 24 hours after symptom onset.[80] Complications of septicemic plague include disseminated intravascular coagulation (DIC), purpuric skin lesions and gangrene of extremities (so-called black death), acute respiratory distress syndrome (ARDS), meningitis, and multiorgan failure with shock.[33,81-83]

DIAGNOSIS

During a confirmed outbreak of pneumonic plague after a biologic attack, a presumptive diagnosis can be made on the basis of symptoms, especially if there is a high index of suspicion. However, other causes of severe pneumonia or rapidly progressive respiratory infection with or without sepsis should be considered. Suspected cases of plague should be immediately reported to the local public health department and the hospital epidemiologist.

There are no widely available, rapid confirmatory tests for *Y. pestis.* Specimens for bacteriologic and serologic testing should be collected before initiating therapy. Sputum, blood, and lymph node aspirate should be submitted for Gram stain and culture. Microscopic examination of clinical specimens or buffy coat may show a gram-negative coccobacillus; Wright, Giemsa, or Wayson stains may show bipolar (safety pin) staining. Sera for acute and convalescent antibody detection should be obtained, but findings are primarily of epidemiologic value. Additional diagnostic testing, including antigen detection, IgM immunoassay, immunostaining, PCR testing, and antimicrobial susceptibility testing, is available through the CDC and through designated public health laboratories (http://www.

Table 7 Antimicrobial Treatment of Pneumonic Plague

Patients	*Drug*	*Comments*
Adults	Streptomycin, 1 g I.M., b.i.d. *or* Gentamicin, 5 mg/kg I.M. or I.V., q.d., or 2 mg/kg loading dose followed by 1.7 mg/kg I.M. or I.V., t.i.d. *Alternative choices:* Doxycycline, 100 mg I.V., b.i.d., or 200 mg I.V., q.d. Ciprofloxacin, 400 mg I.V., b.i.d. Chloramphenicol, 25 mg/kg I.V., q.i.d.	Treat for 10 days; during a community outbreak of pneumonic plague, all persons developing a temperature ≥ 101.3° F (38.5° C) or a new cough should begin parenteral antibiotic treatment; oral treatment may be given when resources for parenteral treatment are limited; pregnant women should not receive streptomycin or chloramphenicol; tetracycline may be substituted for doxycycline
Children	Streptomycin, 15 mg/kg I.M., b.i.d. (maximum, 2 g/day) *or* Gentamicin, 2.5 mg/kg I.M. or I.V., t.i.d. *Alternative choices:* Doxycycline: if ≥ 45 kg, adult dosage; if < 45 kg, 2.2 mg/kg b.i.d. (maximum, 200 mg/day) Ciprofloxacin, 15 mg/kg I.V., b.i.d. Chloramphenicol, 25 mg/kg I.V., q.i.d.; maintain serum concentration between 5 and 20 μg/ml	Use chloramphenicol in plague meningitis

Note: Treatment recommendations may change over time and according to antimicrobial susceptibility test results during a biologic attack and to availability of selected antimicrobial agents. Before initiating treatment, clinicians should consult with an infectious disease specialist and public health authorities and should check for revisions and updates at http://www.bt.cdc.gov/index.asp. This information is adapted from CDC and Working Group on Civilian Biodefense recommendations and may not represent FDA-approved uses.

statepublichealth.org). Specimen submission should be arranged through local public health authorities. The laboratory should be notified whenever plague is suspected, to help prevent exposures to staff and to facilitate appropriate testing.[33,80,83]

Laboratory findings are consistent with the systemic inflammatory response syndrome. The leukocyte count is elevated and the differential shows a neutrophil predominance, including immature forms. Platelets may be normal or low. Coagulation abnormalities include increased fibrin degradation products, hypofibrinogenemia, and prolongation of the prothrombin time (PT) and partial thromboplastin time (PTT). Elevated liver function tests and abnormal renal function tests are seen with systemic disease.

TREATMENT

When plague is suspected, antibiotic treatment should begin before laboratory confirmation of the diagnosis [*see Table 7*]. Whenever possible, specimens should be collected for bacteriologic and serologic testing before the start of therapy. Antibiotic resistance is rare with naturally occurring *Y. pestis* but may be present in strains used as biologic weapons. Treatment should be continued for 10 days or for 3 days after defervescence and improvement in symptoms. The route of administration can be changed from intravenous to oral after the patient is clinically stable. The choice of antibiotic may be modified after microbial sensitivity testing is completed. The CDC bioterrorism Web site or local public health authorities should be consulted for updated treatment recommendations.[33,80,83]

Postexposure Prophylaxis for Pneumonic Plague

All persons potentially exposed to aerosolized *Y. pestis* and all persons in close contact with pneumonic plague patients (close contact is defined as exposure within 2 m [6.5 ft]) should be treated for 7 days after the last exposure [*see Table 8*]. Persons receiving prophylactic antibiotic treatment should seek medical evaluation immediately if fever or illness with cough develops.

There is no currently available vaccine for pneumonic plague. The previously available licensed plague vaccine in the United States was discontinued in 1999. That vaccine was demonstrated to reduce the severity of illness with bubonic plague but not pneumonic plague.[81]

Communicability and Infection Control Considerations

Pneumonic plague is transmitted person to person through respiratory droplets. Aerosol transmission has not been demonstrated. For patients with pneumonic plague, respiratory droplet precautions as well as standard precautions are recommended, including the use of gowns, gloves, eye protection, and surgical masks for the first 48 hours of antimicrobial therapy and until clinical improvement occurs. Hospitalized patients should remain in isolation for the first 48 hours of antimicrobial therapy and until clinical improvement occurs. Hospitalized patients should wear a mask during transport.

Y. pestis is rapidly destroyed by sunlight and drying. Environmental surfaces can be decontaminated with a standard disinfectant. Persons exposed to aerosolized plague bacilli during a biologic attack should shower with warm water and soap. Clothing of persons exposed to an aerosol of *Y. pestis* and linens of plague patients should be washed in hot water.[24,81,82]

Botulism

Botulism is a paralytic illness caused by a potent neurotoxin produced by *Clostridium botulinum,* an anaerobic, spore-forming bacterium. Natural forms of the disease are foodborne botulism, wound botulism, and infant botulism. Foodborne botulism results from ingestion of improperly processed foodstuffs containing preformed toxin produced by *C. botulinum.* Wound botulism results from production of botulinum toxin by *C. botulinum* organisms that contaminate wounds. Infant botulism results from the colonization of the intestinal tract of infants after ingestion of spores. Botulinum toxin has been developed as

a biologic weapon. An aerosol attack is considered the most likely use of botulinum toxin for bioterrorism, although intentional contamination of food supplies is possible.[33,84] Additional information about the pathogenesis and epidemiology of noninhalational forms of botulism is available elsewhere [*see 7:V Anaerobic Infections*].

Botulinum toxin is the most potent lethal toxin known. The estimated toxic dose of type A botulinum toxin is 0.001 μg/kg of body weight. There are seven distinct antigenic types of botulinum neurotoxins—types A through G—produced by different strains of *C. botulinum.* Human botulism is caused primarily by toxin types A, B, and E. Botulinum toxin acts to block neurotransmission by binding irreversibly to the presynaptic nerve terminal at the neuromuscular junction and preventing the release of acetylcholine, resulting in bulbar palsies and skeletal muscle weakness. The toxin is colorless, odorless, and presumably tasteless.[33,85,86]

CLINICAL PRESENTATION

The incubation period for foodborne botulism is 2 hours to 8 days; the typical incubation period is 12 to 72 hours. The incubation period for inhalational botulism is not established. Aerosol exposures of monkeys and accidental aerosol exposure of humans have resulted in clinical illness developing 12 to 80 hours after exposure. Type A toxin is associated with more severe disease and a higher fatality rate than type B or E. The neurologic features of all forms of botulism are similar.[33,85,86] Although initial symptoms in foodborne botulism may include nausea, vomiting, abdominal cramps, and diarrhea, these symptoms are thought to result from other bacterial metabolites in contaminated food and may not occur in inhalational botulism.

The so-called classic triad of botulism summarizes the clinical presentation: an afebrile patient, symmetrical descending flaccid paralysis with prominent bulbar palsies, and a clear sensorium.[85-87] Symptoms of cranial nerve abnormalities nearly always begin in the bulbar musculature; patients typically present with difficulty seeing, speaking, or swallowing. Clinical hallmarks include ptosis, blurred vision, and the so-called four Ds: diplopia, dysarthria, dysphonia, and dysphagia. Cranial nerve abnormalities and bulbar weakness are followed by symmetrical descending weakness and paralysis with progression from the head to the arms, thorax, and legs. The extent of paralysis and rapidity of onset of symptoms are proportional to the dose of toxin absorbed into the circulation. Recovery depends on the regeneration of new motor axon twigs to reinnervate paralyzed muscle fibers; recovery may take weeks to months.

Anticholinergic symptoms are common, including dry mouth, ileus, constipation, nausea and vomiting, urinary retention, and mydriasis. Other symptoms include dizziness and sore throat. Sensory findings are not present, with the exception of circumoral and peripheral paresthesias secondary to hyperventilation resulting from anxiety. Botulinum toxin does not cross the blood-brain barrier. Cranial nerve dysfunction and facial nerve weakness may make communication difficult; these symptoms may be mistaken for lethargy and signs of central nervous system involvement.

DIAGNOSIS

Initiation of treatment with botulinum antitoxin should be based on the clinical diagnosis and should not await laboratory confirmation. A clinician who suspects botulism should immediately contact the local or state health department to facilitate procurement of antitoxin for treatment; arrangements should be made for confirmatory diagnostic testing and initiation of an epidemiologic investigation to identify the source of infection. In cases of potential foodborne botulism, any leftover foodstuffs or containers should be held for testing by the public health laboratory.

Demonstration of botulinum toxin in serum samples by mouse bioassay is diagnostic. Samples of serum (in adults, > 30 ml blood in a tiger-top or red-top tube) obtained before administration of botulinum antitoxin should be submitted for testing. For potential foodborne botulism, samples of stool, gastric aspirate, emesis, and suspect foods should also be submitted.[86] The likelihood of finding toxin in the sera of affected patients decreases with time; it is detectable in only 13% to 28% of patients more than 2 days after ingestion.[88]

The possibility of a bioterrorist attack should be considered in any outbreak of botulism. A bioterrorist attack should especially be considered when a cluster of cases occurs; when an outbreak has a common geographic location but there is no common dietary exposure (suggestive of possible aerosol exposure); when there is an outbreak of an unusual botulinum toxin type; or when multiple simultaneous outbreaks occur. A careful dietary and travel history must be taken to help identify the

Table 8 Postexposure Prophylaxis of Pneumonic Plague

Patients	*Drug*	*Comments*
Adults, including pregnant women	Doxycycline, 100 mg p.o., b.i.d. Ciprofloxacin, 500 mg p.o., b.i.d. *Alternative:* Chloramphenicol, 25 mg/kg p.o., q.i.d.	Asymptomatic household contacts, hospital contacts, or other close contacts should receive postexposure prophylaxis for 7 days; contacts who develop fever or cough while receiving prophylaxis should begin antibiotic treatment for plague.
Children	Doxycycline: if ≥ 45 kg, adult dosage; if < 45 kg, 2.2 mg/kg p.o., b.i.d. (maximum, 200 mg/day) Ciprofloxacin, 20 mg/kg p.o., b.i.d. *Alternative:* Chloramphenicol, 25 mg/kg p.o., q.i.d.	Asymptomatic household contacts, hospital contacts, or other close contacts should receive postexposure prophylaxis for 7 days; contacts who develop fever or cough while receiving prophylaxis should begin antibiotic treatment for plague.

Note: Prophylaxis recommendations may change over time and according to antimicrobial susceptibility test results during a biologic attack and to availability of selected antimicrobial agents. Before initiating prophylaxis, clinicians should consult with an infectious disease specialist and public health authorities and should check for revisions and updates at http://www.bt.cdc.gov/index.asp. This information is adapted from CDC and Working Group on Civilian Biodefense recommendations and may not represent FDA-approved uses.

source. Patients should be asked if they know of others with similar symptoms.

The differential diagnosis of botulism includes stroke and other neuromuscular disorders.[85,86] A CT scan of the head may be used to exclude cerebrovascular accident, although it is relatively insensitive in early ischemic stroke [*see 11:IV Cerebrovascular Disorders*]. Patients with myasthenia gravis will often have characteristic electromyographic findings and serum antibody tests. A test dose of edrophonium (Tensilon) may briefly reverse paralysis in patients with myasthenia gravis but also, reportedly, in some cases of botulism. Guillain-Barré syndrome typically results in ascending paralysis and sensory abnormalities. Cerebrospinal fluid protein is normal in patients with botulism and is normal or elevated in patients with Guillain-Barré syndrome. The rare Miller-Fisher variant of Guillain-Barré syndrome is characterized by descending paralysis and may be confused with botulism. Other conditions that mimic botulism include tick paralysis; poliomyelitis; Eaton-Lambert syndrome; paralytic shellfish poisoning; pufferfish ingestion; and anticholinesterase intoxication with organophosphates, atropine, carbon monoxide, or aminoglycosides.

The electromyogram (EMG) can help distinguish different causes of paralysis. The EMG in botulism demonstrates normal nerve conduction velocity, normal sensory nerve function, and small amplitude motor potentials with facilitation to repetitive stimulation at 50 Hz.[89]

TREATMENT

The mainstay of treatment for botulism is supportive care, including intensive care, mechanical ventilation, and parenteral nutrition. Morbidity and mortality are usually from pulmonary aspiration secondary to loss of the gag reflex and dysphagia leading to inability to control secretions, respiratory failure secondary to inadequate tidal volume from diaphragmatic and accessory respiratory muscle paralysis, and airway obstruction from pharyngeal and upper airway muscle paralysis. Careful and frequent monitoring of the gag and cough reflexes, swallowing, oxygen saturation, vital capacity, and inspiratory force are critical. Airway intubation is indicated for inability to control secretions and impending respiratory failure. Secondary infections are common and should be sought in patients who develop fever.

Trivalent (ABE) equine antitoxin is available from the CDC through state and local health departments and should be administered as soon as possible after clinical diagnosis. Antitoxin can prevent progression of disease caused by subsequent binding of toxin but does not reverse the effects of already bound toxin. For this reason, antitoxin is not useful if the patient is no longer showing progression of disease or is improving from maximum paralysis. The amount of neutralizing antibody present in the standard treatment dose of antitoxin far exceeds maximum serum toxin concentrations in foodborne botulism patients, and repeat doses are usually not required. In a biologic attack, however, patients may be exposed to unusually high concentrations of toxin, so serum toxin levels should be assessed after initiation of treatment in such cases to determine the need for repeat doses. Botulism caused by toxin types other than A, B, or E would not respond to the trivalent antitoxin. Limited quantities of an investigational heptavalent (A-G) antitoxin are held by the United States Army. However, because of the time delay involved in typing the toxin, the utility of this product in a biologic attack is probably minimal.[85,87]

Hypersensitivity reactions, including anaphylaxis, have occurred after administration of botulism antitoxin. For that reason, all patients should undergo a skin test before receiving the antitoxin, and resuscitation equipment should be immediately available. Patients showing a positive hypersensitivity reaction on the skin test can be desensitized over several hours.[90,91]

Before administering antitoxin, physicians should carefully review the package insert for dosage and adverse effects. Standard regimens can be used in children, pregnant women, and immunocompromised persons with botulism. Botulism immune globulin intravenous is an investigational human-derived neutralizing antibody that is available only for treatment of infant botulism from the California Department of Health Services, Berkeley. The CDC bioterrorism Web site or local public health authorities should be consulted for updated treatment recommendations.[33,85,86]

Transmissibility and Infection Control

Botulism is an intoxication, not an infection, and thus is not transmitted from person to person. Botulinum toxin does not penetrate intact skin. Standard infection-control precautions are adequate unless meningitis is suspected, in which case droplet precautions are indicated. Clothes of persons exposed to an aerosol release of botulinum toxin should be removed and washed. Exposed persons should shower with soap and hot water. Exposed environmental surfaces can be decontaminated with 0.1% hypochlorite bleach solution.[86]

Tularemia

Tularemia is a zoonotic infection caused by *Francisella tularensis,* a small, nonmotile, gram-negative, pleomorphic coccobacillus. The disease is typically acquired through contact with blood or tissue fluids of infected animals or through the bite of an infected deerfly, tick, or mosquito.[92] Inhalation of organisms aerosolized from the environment and the drinking of contaminated water can also result in human infection.[93] *F. tularensis* was developed for use as a biologic weapon by the United States (before its offensive biologic weapons program was terminated) and other countries.[33] The epidemiology, pathogenesis, and clinical manifestations of the naturally occurring forms of tularemia are discussed in more detail elsewhere [*see 7:XI Infections Due to Brucella, Francisella, Yersina Pestis, and Bartonella*].

CLINICAL PRESENTATION

Tularemia can take several forms in humans, depending on the route of infection. Ulceroglandular, oculoglandular, glandular, typhoidal, and pharyngeal tularemia are discussed elsewhere [*see 7:XI Infections Due to Brucella, Francisella, Yersina Pestis, and Bartonella*]. Inhalational tularemia is a term used to describe infection resulting from an aerosol release of *F. tularensis*.[94] Most patients with inhalational tularemia develop pleuropulmonary tularemia (tularemia pneumonia), but many patients may present with an undifferentiated febrile illness. The infectious dose is as low as one to 50 organisms, and the incubation period is typically 3 to 5 days (range, 1 to 14 days).[33]

The clinical course of inhalational tularemia is less rapidly progressive than that of pulmonary anthrax or plague. Illness onset is acute, with some combination of fever, chills, sweats, myalgias, headache, coryza, and sore throat. Nausea, vomiting, diarrhea, and abdominal pain are common. Anorexia and weight loss may occur as the illness continues. Cough may be

dry or mildly productive. Hemoptysis is uncommon. Pleuritic chest pain, substernal chest discomfort, and dyspnea may be present. Chest x-rays may be normal or minimally abnormal or show a variety of abnormalities, including peribronchial patchy infiltrates, effusions, and hilar adenopathy.[95]

F. tularensis infection may be mild and nonspecific or rapidly progressive. Any form of tularemia may result in hematogenous spread with secondary pleuropneumonia, sepsis, and, rarely, meningitis. If left untreated, tularemia can progress to respiratory failure; liver, kidney, and splenic involvement; meningitis; sepsis; shock; and death. There is usually complete recovery with early diagnosis and treatment. Mortality is less than 2% if the patient is treated; it can be as high as 60% for untreated severe disease and pneumonia.[94,96,97]

DIAGNOSIS

A clustering of sudden, severe pneumonias in previously healthy patients should raise the possibility of an intentional aerosolized release of tularemia. Clusters of patients with tularemia and cases in which there is no natural explanation for the disease should be reported immediately to the local or state health department (http://www.statepublichealth.org). There are no rapid confirmatory tests for *F. tularensis*. Gram stain of sputum is not diagnostic but may identify other potential etiologies.[97,98] In the context of a known or suspected outbreak, a presumptive diagnosis can be made on the basis of symptoms. A chest x-ray should be obtained for patients with suspected pleuropulmonary tularemia. The x-ray may show infiltrates, effusion, hilar adenopathy, or subtle abnormalities, or it may be normal. Recent experience with inhalational anthrax suggests that chest CT scans of patients with tularemia may show pulmonary abnormalities, including infiltrates, effusions, and adenopathy, before they are evident on x-ray.[62]

Specimens of respiratory secretions and blood for bacteriologic and serologic testing should be collected before initiating therapy. Pharyngeal washings, sputum specimens, fasting gastric aspirates, and blood can be cultured for *F. tularensis*. Growth may be slow, so cultures should be held for 10 days. Cysteine-enriched culture media should be used to improve yield. Direct examination (by direct fluorescent antibody staining or immunohistochemical testing, antigen detection, microagglutination antibody testing, PCR, and other research tests) is available through designated public health laboratories. Acute and convalescent serologies are valuable for epidemiologic purposes.[94,98]

TREATMENT AND POSTEXPOSURE PROPHYLAXIS

When the index of suspicion is high, antibiotic treatment should be started before diagnosis is confirmed. Streptomycin or gentamicin is the preferred agent. All persons potentially exposed to aerosolized *F. tularensis* should be treated with doxycycline or ciprofloxacin. Close contacts of patients with tularemia pneumonia do not need prophylactic antibiotics. No vaccine for tularemia is currently available. The CDC bioterrorism Web site, local public health authorities, or both should be consulted for updated treatment recommendations.[33,94,99]

Transmissibility and Infection Control

Tularemia is not transmitted from person to person, and isolation of patients with tularemia is not necessary. Standard precautions are recommended for all patients with tularemia. Microbiology staff must be alerted when tularemia is suspected, so they can take precautions to prevent laboratory-acquired infection from culture plates and other infectious materials. Contaminated environmental surfaces can be disinfected with a 10% bleach solution followed by cleansing with 70% alcohol.[94]

Hemorrhagic Fever Viruses

Hemorrhagic fever viruses (HFVs) are RNA viruses classified in several taxonomic families. HFVs cause a variety of disease syndromes with similar clinical characteristics, referred to as acute hemorrhagic fever syndromes [*see 7:XXXI Viral Zoonoses*]. The pathophysiologic hallmarks of HFV infection are microvascular damage and increased vascular permeability. HFVs that are of concern as potential biologic weapons include Arenaviridae (Lassa, Junin, Machupo, Guanarito, and Sabia viruses, which are the causative agents of Lassa fever and Argentine, Bolivian, Venezuelan, and Brazilian hemorrhagic fevers, respectively); Filoviridae (Ebola and Marburg viruses); Flaviviridae (yellow fever, Omsk hemorrhagic fever, and Kyasanur Forest disease viruses); and Bunyaviridae (Rift Valley fever [RVF]). Under natural conditions, humans are infected through the bite of an infected arthropod or through contact with infected animal reservoirs. Hemorrhagic fever viruses are highly infectious by aerosol; are associated with high morbidity and, in some cases, high mortality; and are thought to pose a serious risk as biologic weapons.[33] All suspected cases of HFV infection should be reported immediately to the local or state health department and the hospital epidemiologist.

PATHOPHYSIOLOGY

The exact pathogenesis for HFVs varies according to the etiologic agent. The major target organ is the vascular endothelium. Immunologic and inflammatory mediators are thought to play an important role in the pathogenesis of HFVs. All HFVs can produce thrombocytopenia, and some also cause platelet dysfunction. Infection with Ebola and Marburg viruses, Rift Valley fever virus, and yellow fever virus causes destruction of infected cells. DIC is characteristic of infection with Filoviridae. Ebola and Marburg viruses may cause a hemorrhagic diathesis and tissue necrosis through direct damage to vascular endothelial cells and platelets with impairment of the microcirculation, as well as cytopathic effects on parenchymal cells, with release of immunologic and inflammatory mediators. Arenaviridae, on the other hand, appear to mediate hemorrhage via the stimulation of inflammatory mediators by macrophages, thrombocytopenia, and the inhibition of platelet aggregation. DIC is not a major pathophysiologic mechanism in arenavirus infections.[100,101]

CLINICAL PRESENTATION

The incubation period of HFVs ranges from 2 to 21 days. The clinical presentations of these diseases are nonspecific and variable, making diagnosis difficult. It is noteworthy that not all patients will develop hemorrhagic manifestations. Even a significant proportion of patients with Ebola virus infections may not demonstrate clinical signs of hemorrhage.[102]

Initial symptoms of the acute HFV syndrome may include fever, headache, myalgia, rash, nausea, vomiting, diarrhea, abdominal pain, arthralgias, myalgias, and malaise. Illness caused by Ebola, Marburg, Rift Valley fever virus, yellow fever virus, Omsk hemorrhagic fever virus, and Kyasanur Forest disease virus are characterized by an abrupt onset, whereas Lassa fever and the diseases caused by the Machupo, Junin, Guarinito, and Sabia viruses have a more insidious onset. Initial signs may include fever, tachypnea, relative bradycardia, hypotension

(which may progress to circulatory shock), conjunctival injection, pharyngitis, and lymphadenopathy. Encephalitis may occur, with delirium, seizures, cerebellar signs, and coma. Most HFVs cause cutaneous flushing or a macular skin rash, although the rash may be difficult to appreciate in dark-skinned persons and varies according to the causative virus. Hemorrhagic symptoms, when they occur, develop later in the course of illness and include petechiae, purpura, bleeding into mucous membranes and conjunctiva, hematuria, hematemesis, and melena. Hepatic involvement is common, and renal involvement is proportional to cardiovascular compromise.[33,100,102,103]

Laboratory abnormalities include leukopenia (except in some cases of Lassa fever), anemia or hemoconcentration, and elevated liver enzymes; DIC with associated coagulation abnormalities and thrombocytopenia are common. Mortality ranges from less than 1% for Rift Valley fever to 70% to 90% for Ebola and Marburg virus infections.[33,100,102-104]

DIAGNOSIS

The nonspecific and variable clinical presentation of the HFVs presents a considerable diagnostic challenge. Clinical diagnostic criteria based on WHO surveillance standards for acute hemorrhagic fever syndrome include temperature greater than 101° F (38.3° C) of less than 3 weeks' duration; severe illness and no predisposing factors for hemorrhagic manifestations; and at least two of the following hemorrhagic symptoms: hemorrhagic or purple rash, epistaxis, hematemesis, hematuria, hemoptysis, blood in stools, or other hemorrhagic symptom with no established alternative diagnosis. Any suspected case of HFV should result in immediate notification of the hospital epidemiologist, local public health department, and clinical laboratory personnel.[101,105] Laboratory testing is currently available only at the CDC and the United States Army Medical Research Institute for Infectious Diseases. Laboratory techniques for the diagnosis of HFVs include antigen detection, IgM antibody detection, isolation in cell culture, visualization by electron microscopy, immunohistochemical techniques, and reverse transcriptase–polymerase chain reaction. Submission of clinical specimens, including processing and transport, should be arranged through consultation with local public health authorities. The CDC's Packaging Protocols for Biologic Agents/Diseases are available at (http://www.bt.cdc.gov/agent/vhf/index.asp).

TREATMENT

Therapy for HFVs is largely supportive. Treatment of other suspected causes of infection should be administered pending confirmation of HFV infection. Hypotension and shock may require early administration of vasopressors and hemodynamic monitoring with attention to fluid and electrolyte balance, circulatory volume, and blood pressure. HFV patients tend to respond poorly to fluid infusions and rapidly develop pulmonary edema.

Secondary infections may occur and should be diagnosed and treated. Intravenous lines, catheters, and other invasive procedures should be avoided unless they are clearly indicated. The management of bleeding is controversial. Recent recommendations include not treating mild bleeding and use of replacement therapy and heparin for severe bleeding with DIC.[33] Intramuscular injections and medications that interfere with platelet function or coagulation should be avoided.

No treatments of HFVs have been approved by the Food and Drug Administration. Ribavirin is a nucleoside analogue with activity against some Arenaviridae and Bunyaviridae (including the viruses that cause Lassa fever, Argentine hemorrhagic fever, and Crimean-Congo hemorrhagic fever) but not against Filoviridae or Flaviviridae. Ribavirin may be used under an IND protocol for the empirical treatment of HFV patients while awaiting identification of the etiologic agent. Current treatment protocols and dosing recommendations for ribavirin should be obtained through local public health authorities or the CDC's bioterrorism Web site.

Postexposure Prophylaxis

Postexposure prophylaxis is currently recommended only for persons potentially exposed to HFV and for known high-risk contacts or close contacts of HFV patients who develop fever or other clinical criteria of HFV infection with no alternative diagnosis, unless the etiologic agent is known to be a filovirus or a flavivirus.[100]

Infection Control Considerations

Ebola virus, Marburg virus, Lassa fever virus, and the New World arenaviruses are transmissible from person to person through direct contact with blood and body fluids. Airborne transmission of HFVs is unlikely but cannot be completely ruled out. The risk of person-to-person transmission is highest during the latter stages of illness, which are characterized by vomiting, diarrhea, shock, and, often, hemorrhage. The most important step in preventing transmission of HFVs is strict attention to implementation of appropriate barrier infection control measures, including double gloves, impermeable gowns, face shields, eye protection, and leg and shoe coverings.

Airborne precautions are recommended during care of patients with possible HFV infections. Airborne precautions include high-efficiency particulate respirators such as N-95 masks or powered air-purifying respirators (PAPRs) for all persons entering the patient's room. Patients should be placed in a negative-pressure isolation room with 6 to 12 air changes per hour.[101,106]

High-risk contacts of HFV patients include persons having contact with mucous membranes (e.g., through kissing or sexual intercourse) or with secretions, excretions, or blood (through percutaneous injury) of the infected person. Close contacts are persons who have other direct contact with the patient (e.g., shaking hands or hugging), provide medical care to the patient, or process laboratory specimens from a patient with HFV before initiation of infection-control precautions.

Persons potentially exposed to HFVs in a bioterrorist attack and their close and high-risk contacts should be placed under medical surveillance for 21 days from the day of exposure. Temperatures should be recorded twice daily, and any temperature of 101° F (38.3° C) or higher should be reported to the designated clinical or public health authority. Therapy with ribavirin should be initiated promptly unless an alternative diagnosis is established or the etiologic agent is known to be a filovirus or a flavivirus [*see* Treatment, *above*].[100]

HFVs are highly infectious in the laboratory setting through small-particle aerosols generated through procedures such as centrifugation. Laboratory personnel should be alerted when HFV infections are suspected, and appropriate personal-protection precautions and laboratory biosafety procedures should be implemented.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Davis CJ: Nuclear blindness: an overview of the biological weapons programs of the former Soviet Union and Iraq. Emerg Infect Dis 5:509, 1999

2. Stern J: The prospect of domestic bioterrorism. Emerg Infect Dis 5:517, 1999

3. Kortepeter MG, Parker GW: Potential biological weapons threats. Emerg Infect Dis 5:523, 1999)

4. Alibek K: Biohazard. Random House, Inc. New York, 1999

5. Henderson DA: The looming threat of bioterrorism. Science 283:1279, 1999

6. Biological and chemical terrorism: strategic plan for preparedness and response. Recommendations of the CDC Strategic Planning Workgroup. MMWR Morb Mortal Wkly Rep 49:1, 2000

7. Gerberding JL, Hughes JM, Koplan JP: Bioterrorism preparedness and response: clinicians and public health agencies as essential partners. JAMA 287:898, 2002

8. Lane HC, Fauci AS: Bioterrorism on the home front: a new challenge for American medicine. JAMA 286:2595, 2001

9. Meningococcal disease and college students. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 49:13, 2000

10. Bell BP, Goldoft M, Griffin PM, et al: A multistate outbreak of Escherichia coli 0157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers: the Washington experience. JAMA 272:1349, 1994

11. Weber DJ, Rutala WA: Pertussis: a continuing hazard for healthcare facilities. Infect Control Hosp Epidemiol 22:736, 2001

12. Measles, mumps, and rubella: vaccine use and strategies for elimination of measles, rubella, and congenital rubella syndrome and control of mumps. Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 47:1, 1998

13. Prevention of varicella. Updated Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 48:1, 1999

14. Bolyard EA, Tablan OC, Williams WW, et al: Guideline for infection control in healthcare personnel, 1998. Hospital Infection Control Practices Advisory Committee. Infect Control Hosp Epidemiol 19:386, 1998

15. Suspected brucellosis case prompts investigation of possible bioterrorism-related activity–New Hampshire and Massachusetts, 1999. MMWR Morb Mortal Wkly Rep 49:509, 2000

16. Duchin JS, Koster FT, Peters CJ, et al: Hantavirus pulmonary syndrome: a clinical description of 17 patients with a newly recognized disease. The Hantavirus Study Group. N Engl J Med 330:949, 1994

17. Fine A, Layton M: Lessons from the West Nile viral encephalitis outbreak in New York City, 1999: implications for bioterrorism preparedness. Clin Infect Dis 32:277, 2001

18. Pandemic influenza: confronting a re-emergent threat. Proceedings of a meeting. Bethesda, Maryland, 11-13 December 1995. J Infect Dis 176:S1, 1997

19. Rota PA, Oberste MS, Monroe SS, et al: Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300:1394, 2003

20. Wenzel RP, Edmond MB: Managing SARS amidst uncertainty. N Engl J Med 348:1947, 2003

21. Update: investigation of bioterrorism-related anthrax and interim guidelines for clinical evaluation of persons with possible anthrax. MMWR Morb Mortal Wkly Rep 50:941, 2001

22. Updated interim surveillance case definition for severe acute respiratory syndrome (SARS)—United States, April 29, 2003. MMWR Morb Mortal Wkly Rep 52:391, 2003

23. Interim guidance on infection control precautions for patients with suspected severe acute respiratory syndrome (SARS) and close contacts in households. Centers for Disease Control and Prevention, Atlanta, August 18, 2003
http://www.cdc.gov/ncidod/sars/ic-closecontacts.htm.

24. Control of Communicable Diseases Manual, 17th ed. Chin JE, Ed. American Public Health Association, Washington DC, 2000

25. Garner JS: Guideline for isolation precautions in hospitals. Infect Control Hosp Epidemiol 17:53, 1996

26. Immunization of health-care workers. Recommendations of the Advisory Committee on Immunization Practices and the Hospital Infection Control Practices Advisory Committee. MMWR Morb Mortal Wkly Rep 46:1, 1997

27. Rotz LD, Khan AS, Lillibridge SR, et al: Public health assessment of potential bioterrorism agents. Emerg Infect Dis 8:225, 2002

28. Torok TJ, Tauxe RV, Wise RP, et al: A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. JAMA 278:389, 1997

29. Kolavic SA, Kimura A, Simons SL, et al: An outbreak of Shigella dysenteriae type 2 among laboratory workers due to intentional food contamination. JAMA 278:396, 1997

30. Pavlin JA: Epidemiology of bioterrorism. Emerg Infect Dis 5:528, 1999

31. Recognition of illness associated with the intentional release of a biologic agent. MMWR Morb Mortal Wkly Rep 50:893, 2001

32. Vaccinia (smallpox) vaccine. Recommendations of the Advisory Committee on Immunization Practices (ACIP), 2001. MMWR Morb Mortal Wkly Rep 50:1, 2001

33. Franz DR, Jahrling PB, Friedlander AM, et al: Clinical recognition and management of patients exposed to biological warfare agents. JAMA 278:399, 1997

34. Draft Supplemental Recommendation of the ACIP. Use of Smallpox (Vaccinia) Vaccine. Centers for Disease Control and Prevention, Atlanta, June 2002

35. Fenner F, Henderson DA, Arita I, et al: Smallpox and its eradication. World Health Organization, Geneva, 2001
http://www.who.int/emc/diseases/smallpox/Smallpoxeradication.html

36. Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. JAMA 281:2127, 1999

37. Breman, JG, Henderson DA: Diagnosis and management of smallpox. N Engl J Med 346:1300, 2002

38. Henderson DA: Smallpox: clinical and epidemiologic features. Emerging Infect Dis 5:537, 1999

39. Smallpox response plan and guidelines. Centers for Disease Control and Prevention, Atlanta, November 26, 2002
http://www.bt.cdc.gov/agent/smallpox/response-plan/index.asp

40. Grabenstein JD, Winkenwerder W: US military smallpox vaccination program experience. JAMA 289:3278, 2003

41. Update: cardiac and other adverse events following civilian smallpox vaccination—United States, 2003. MMWR Morb Mortal Wkly Rep 52:639, 2003

42. Update: adverse events following civilian smallpox vaccination—United States, 2003. MMWR 52:819, 2003

43. Neff JM, Lane JM, Fulginetti VA, et al: Contact vaccinia: transmission of vaccinia from smallpox vaccination. JAMA 288:1901, 2003

44. Bartlett J: Smallpox vaccination and patients with human immunodeficiency virus infection or acquired immunodeficiency syndrome. Clin Infect Dis 36:468, 2003

45. Lane JM, Ruben FL, Neff JM, et al: Complications of smallpox vaccination, 1968: results of ten statewide surveys. J Infect Dis 122:303, 1970

46. Update: cardiac-related events during the civilian smallpox vaccination program—United States, 2003. MMWR Morb Mortal Wkly Rep 52:492, 2003
http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5221a2.htm

47. Halsell JS, Riddle JR, Atwood JE, et al: Myopericarditis following smallpox vaccination among vaccinia naive US military personnel. JAMA 289:3283, 2003

48. Supplemental recommendations on adverse events following smallpox vaccine in the pre-event vaccination program: recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 52:282, 2003

49. Talbot TR, Bredenberg HK, Smith M, LaFleur BJ, Boyd A, Edwards KM. Focal and generalized folliculitis following smallpox vaccination among vaccinia-naive recipients. JAMA 289:3290, 2003

50. Acha PN, Szyfres B: Zoonoses and Communicable Disease Common to Man and Animals, 3rd ed. Vol I. Pan American Health Organization, Washington, DC, 2001

51. Swartz MN: Recognition and management of anthrax: an update. N Engl J Med 345:1621, 2001

52. Brachman PS: Bioterrorism: an update with a focus on anthrax. Am J Epidemiol 155:981, 2002

53. Update: investigation of bioterrorism-related anthrax and adverse events from antimicrobial prophylaxis. MMWR Morb Mortal Wkly Rep 50:973, 2001

54. Cieslak TJ, Eitzen HM Jr: Bioterrorism: agents of concern. J Public Health Manag Pract 6:19, 2000

55. Kournikakis B, Armour SJ, Boulet CA, et al: Risk assessment of anthrax threat letters. Technical Report 2001-048. Defense Research Establishment, Suffield, Canada, 2001

56. Dixon TC, Meselson M, Guillemin J, et al: Anthrax. N Engl J Med 341:815, 1999

57. Sirisanthana T, Brown AE: Anthrax of the gastrointestinal tract. Emerg Infect Dis 8:649, 2002

58. Inglesby TV, O'Toole T, Henderson DA, et al: Anthrax as a biological weapon, 2002: updated recommendations for management. JAMA 287:2236, 2002

59. Meselson M, Guillemin J, Hugh-Jones M, et al: The Sverdlovsk anthrax outbreak of 1979. Science 266:1202, 1994

60. LaForce FM. Anthrax. Clin Infect Dis 19:1009, 1994

61. Guarner J, Jernigan JA, Sheih W, et al: Pathology and pathogenesis of bioterrorism-related inhalational anthrax. Am J Pathol 163:701, 2003

62. Jernigan JA, Stephens DS, Ashford DA, et al: Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg Infect Dis 7:933, 2001

63. Barakat LA, Quentzel HL, Jernigan JA, et al: Fatal inhalational anthrax in a 94-year-old Connecticut woman. JAMA 287:863, 2002

64. Mina B, Dym JP, Kuepper F, et al: Fatal inhalational anthrax with unknown source of exposure in a 61-year-old woman in New York City. JAMA 287:858, 2002

65. Brachman PS: Inhalation anthrax. Ann NY Acad Sci 353:83, 1980

66. USAMRIID's Medical Management of Biological Casualties Handbook (USAMRIID Blue Book), 4th ed. United States Army Medical Research Institute of Infectious Diseases, Fort Detrick, Maryland, February, 2001
http://www.usamriid.army.mil/education/bluebook.html

67. Use of anthrax vaccine in the United States. Recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 49:1, 2000

68. Kuehnert MJ, Doyle TJ, Hill HA, et al: Clinical features that discriminate inhalational anthrax from other acute respiratory diseases. Clin Infect Dis 36:328, 2003

69. Considerations for distinguishing influenza-like illness from inhalational anthrax. MMWR Morb Mortal Wkly Rep 50:984, 2001

70. Mayer TA, Morrison A, Bersoff-Matcha S, et al: Inhalational anthrax due to bioterrorism: would current Centers for Disease Control and Prevention guidelines have identified the 11 patients with inhalational anthrax from October through November 2001? Clin Infect Dis 36:1275, 2003

71. Update: investigation of bioterrorism-related anthrax and interim guidelines for exposure management and antimicrobial therapy, October 2001. MMWR Morb Mortal Wkly Rep 50:909, 2001

72. Update: investigation of anthrax associated with intentional exposure and interim public health guidelines, October 2001. MMWR Morb Mortal Wkly Rep 50:889, 2001

73. Statement by the Department of Health and Human Services: Regarding additional options for preventive treatment for those exposed to inhalational anthrax. US Department of Health and Human Services, December 18, 2001
http://www.hhs.gov/news/press/2001pres/20011218.html

74. Committee to Assess the Safety and Efficacy of the Anthrax Vaccine. Medical Follow-Up Agency: The Anthrax Vaccine: Is It Safe? Does It Work? Joellenbeck LM, Zwanziger LL, Durch JS, et al, Eds: Institute of Medicine, National Academy Press, Washington, DC, 2002
http://www.iom.edu/iom/iomhome.nsf/Wfiles/Anthrax-8-pager1FINAL/$file/Anthrax-8-pager1FINAL.pdf

75. Additional options for preventive treatment for persons exposed to inhalational anthrax. MMWR Morb Mortal Wkly Rep 50:1142, 2001

76. Use of anthrax vaccine in response to terrorism. Supplemental recommendations of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 51:1024, 2002

77. Cutaneous anthrax management algorithm. AAD Ad Hoc Task Force on Bioterrorism. American Academy of Dermatology, November 21, 2001
http://www.aad.org/BioInfo/Biomessage2.html

78. Biosafety in Microbiological and Biomedical Laboratories (BMBL) 4th Edition. U.S. Department of Health and Human Services Centers for Disease Control and Prevention and National Institutes of Health, May 1999. US Government Printing Office, Washington, DC, 1999
http://www.cdc.gov/od/ohs/biosfty/bmbl4/bmbl4toc.htm

79. Perry RD, Fetherston JD: Yersinia pestis—etiologic agent of plague. Clin Microbiol Rev 10:35, 1997

80. Gage KL, Dennis DT, Orloski KA, et al: Cases of cat-associated human plague in the western U.S., 1977–1998. Clin Infect Dis 30:893, 2000

81. Inglesby TV, Dennis DT, Henderson DA, et al: Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. JAMA 283:2281, 2000

82. Prevention of plague. Recommendations of the Advisory Committee on Immunization Practice. MMWR Morb Mortal Wkly Rep 45:1 1996

83. McGovern TW, Friedlander AM: Plague. Medical Aspects of Chemical and Biological Warfare. Textbook of Military Medicine Series. Part I, Warfare, Weaponry and the Casualty. Sidell FR, Takafuji ET, Franz DR, Eds. TMM Publications, Washington, DC, 1997

84. Shapiro RL, Hatheway C, Becher J, et al: Botulism surveillance and emergency response: a public health strategy for a global challenge. JAMA 278:433, 1997

85. Shapiro RL, Hatheway C, Swerdlow DL: Botulism in the United States: a clinical and epidemiologic review. Ann Intern Med 129:221, 1998

86. Arnon SS, Schechter R, Inglesby TV, et al: Botulinum toxin as a biological weapon: medical and public health management. JAMA 285:1059, 2001

87. Cherington M: Clinical spectrum of botulism. Muscle Nerve 21:701, 1998

88. Woodruff BA, Griffin PM, McCroskey LM, et al: Clinical and laboratory comparison of botulism toxin types A, B and E in the United States, 1975–1988. J Infect Dis 166:1281, 1992

89. Angulo FJ, Getz J, Taylor JP, et al: A large outbreak of botulism: the hazardous baked potato. J Infect Dis 178:172, 1998

90. Eitzen E: Medical management of biological casualties, 3rd ed. U.S. Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland, 1998

91. Black RE, Gunn RA: Hypersensitivity reactions associated with botulinal antitoxin. Am J Med 69:567, 1980

92. Tularemia—United States, 1999–2000. MMWR Morb Mortal Wkly Rep 51:181, 2002

93. Feldman KA, Enscore RE, Lathrop SL, et al: An outbreak of primary pneumonic tularemia on Martha's Vineyard. N Engl J Med 345:1601, 2001

94. Dennis DT, Inglesby TV, Henderson DA, et al: Tularemia as a biological weapon: medical and public health management. JAMA 285:2763, 2001

95. Choi E: Tularemia and Q fever. Med Clinics North Am 86:393, 2002

96. Evans ME, Gregory DW, Schaffner W, et al: Tularemia: a 30-year experience with 88 cases. Medicine (Baltimore) 64:251, 1985

97. Gill V, Cunha BA: Tularemia pneumonia. Semin Respir Infect 12:61, 1997

98. Evans ME, Friedlander AM: Tularemia. Medical Aspects of Chemical and Biological Warfare. Textbook of Military Medicine Series. Part I, Warfare, Weaponry and the Casualty. Sidell FR, Takafuji ET, Franz DR, Eds. TMM Publications, Washington, DC, 1997

99. Limaye AP, Hooper CJ: Treatment of tularemia with fluoroquinolones: two cases and review. Clin Infect Dis 29:922, 1999

100. Borio L, Inglesby T, Peters CJ, et al: Hemorrhagic fever viruses as biological weapons: medical and public health management. JAMA 287:2391, 2002

101. Khan AS, Sanchez A, Pflieger AK: Filoviral hemorrhagic fevers. Br Med Bull 54:675, 1998

102. Bwaka MA, Bonnet MJ, Calain P, et al: Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: clinical observations in 103 patients. J Infect Dis 179:S1, 1999

103. Jahrling PB: Viral hemorrhagic fevers. Textbook of Military Medicine Series. Part I, Warfare, Weaponry and the Casualty. Sidell FR, Takafuji ET, Franz DR, Eds. TMM Publications, Washington, DC, 1997

104. Isaacson M: Viral hemorrhagic fever hazards for travelers in Africa. Clin Infect Dis 33:1707, 2001

105. Acute hemorrhagic fever syndrome. World Health Organization.
http://www.who.int/emcdocuments/surveillance/docs/whocdscsrisr992.html/41Acute%20haemorrhagic%20fever%20syndrome.htm

106. Update: management of patients with suspected viral hemorrhagic fever—United States. MMWR Morb Mortal Wkly Rep 44:475, 1995

Acknowledgments

Figures 1, 3, 4, and 5 Centers for Disease Control and Prevention Public Health Image Library.

Figure 2 Centers for Disease Control and Prevention Public Health Image Library (Dr Duma).

VIII ASSESSMENT OF THE GERIATRIC PATIENT

HELEN K. EDELBERG, M.D.

The conventional medical approach, which focuses on single diseases and organ systems, does not meet the complex medical and psychosocial needs of many elderly patients. Over the past 20 years, geriatric assessment (GA) has evolved as a way to account for the atypical or nonspecific presentation of illness, the impact of medical comorbidity, and the influence of psychosocial stressors on the health of elderly persons.[1,2]

Different models of GA have been devised, utilizing a variety of settings, personnel, and approaches. Most models also include treatment and follow-up of some kind. Hence, the term assessment is something of a misnomer; these would more accurately be described as evaluation and management programs. Despite their diversity, many GA programs have resulted in improved diagnostic accuracy, functional status, affect, cognition, placement outcomes, and survival. They have also had a beneficial effect on health care utilization, reducing the number of hospital admissions (and readmissions) and institutionalization, as well as total health care costs. A recent multicenter randomized, controlled trial of 1,388 frail elderly persons reported significant reductions in functional decline with inpatient GA and improvements in mental health with outpatient GA, with no increase in costs.[2]

Studies suggest that for a GA program to be successful, it should be targeted to those patients most likely to benefit from the intervention.[3] For all intents and purposes, frail patients (i.e., those 80 years of age and older and those suffering from multiple medical and psychosocial problems) can be considered appropriate candidates for GA programs that focus on rehabilitation or chronic disease management and improved care coordination.[4,5] A number of criteria can be used for the selection of these high-risk elderly patients [*see Table 1*]. In contrast, GA programs that aim to identify risk factors for functional decline and develop interventions to prevent or delay impairment are more likely to benefit low-risk elderly persons.[6]

Principles of Geriatric Assessment

General features of GA include the following: (1) an interdisciplinary team approach to patient care; (2) a focus on prevention, including the prevention of decline (maintaining functional status); and (3) a feedback loop to promote adherence to recommendations by other health care providers, patients, and caregivers, as well as to promote patient self-efficacy or confidence in the ability to perform specific activities.[7]

TEAM APPROACH

The traditional components or functional domains of GA are physical, cognitive, psychological, and social [*see Table 2*]. Teams are more effective in assessing these domains and creating an effective care plan than are professionals working alone.[8] The core disciplines of GA are medicine, nursing, and social work. Other professionals—such as a physical, occupational, or speech therapist; a pharmacist; a psychologist; a nutritionist; a dentist; a visiting nurse; a podiatrist; or a member of the clergy—may participate in the GA team or serve as consultants. The composition and structure of the team reflect the treatment goals and resources of the specific setting, whether it is a specialized geriatric ward, consultation service, outpatient clinic, patient's home, or a nursing home.

In general, team leadership rotates, with the key provider of care reporting on the patient's progress. For example, if the major concern is the medical condition of the patient, a physician should lead the team meeting and introduce the team to the patient and family members. To be effective team members, physicians must be knowledgeable about geriatric medicine, familiar with the patient, dedicated to the team process, and have good communication skills. As team members, physicians explain the medical conditions and differential diagnoses that affect care, incorporate the team advice into medical orders, and alert the patient, family members, and caregivers about team decisions.[8] To create, monitor, or revise the care plan, interdisciplinary teams must communicate openly, freely, and regularly. Core team members must collaborate with trust and respect for the contributions of others and coordinate (i.e., delegate, share accountability, and jointly implement) the care plan. Some team members work together at the same site, so communication can be informal and expeditious. Teams should set deadlines for reaching their goals and have regular meetings to discuss team structure, process, and communication. Team effectiveness should be defined by specific goals at the outset and monitored by continuous quality-improvement measures.[8]

When formal teams do not exist, physicians may be able to achieve the benefits of an interdisciplinary approach by assembling informal teams through referrals to other professionals and community-based organizations, such as adult day programs and home care agencies. The Eldercare Locator service (1-800-677-1116), operated by the National Association of Area Agencies on Aging, provides information on and refers callers to local

Table 1 Suggested Criteria for Geriatric Assessment

Impairment of activities of daily living
Incontinence
Confusion or dementia
Chronic and disabling illness that prevents patient self-care
Impaired mobility
Malnutrition or weight loss
Falls
Depression
Socioeconomic or family problems
Sensory impairment (e.g., visual or hearing)
Cerebrovascular accident
Prolonged bed rest
Restraints
Pressure sore
Polypharmacy

Table 2 Domains of Geriatric Assessment

Physical health	Illness Incontinence Balance and gait Falls Nutrition Dental Hearing Vision Sexual functioning Polypharmacy Prevention
Mental health	Cognition Depression Anxiety Psychosis
Functioning	Basic activities of daily living Instrumental activities of daily living Advanced activities of daily living
Social support	Caregiver burden Finances Service needs Values and advance directives Communication and emergency support Elder abuse and neglect
Environmental adequacy	Safety Impact of climate

services for elderly persons. Resources are also available on the Internet [*see Table 3*].

FOCUS ON PREVENTION

Primary and secondary prevention for the elderly have become increasingly important as the population has continued increasing in average age and average life expectancy.[9] Because primary care practice guidelines do not always address the specific needs of elderly persons, physicians may be compelled to rely on clinical judgment to modify preexisting recommendations that were designed for middle-aged adults.[10] Assessment categories unique to elderly patients include functional status (e.g., activities of daily living [ADL] and instrumental ADL) [*see Table 4*], sensory perception, and injury prevention.[11] Interventional areas that are common to other age groups but have special implications for elderly patients include immunizations, diet and exercise, and sexuality [*see Table 5*]. Cognitive ability and mental health issues should also be evaluated within the context of the individual patient's social situation—not by screening all patients but by being alert to the occurrence of any change in mental function. The prevalence of undetected, correctable conditions and comorbid diseases is high in elderly persons. The American College of Physicians has advocated incorporating GA into routine medical care for elderly persons, particularly patients 75 years of age and older.[12]

FEEDBACK LOOP

Communication among team members and with patients and caregivers—whether it is face to face or by telephone, e-mail, or fax—is the key to effective GA programs. The team and patient must develop ways to communicate honestly, to prevent the patient from suppressing an opinion and agreeing to every suggestion. For example, patients can help the team to set goals (e.g., advance directives and end-of-life care). They can also discuss drug treatment, rehabilitation, dietary plans, and other forms of therapy. If the team learns that the patient will not take a particular drug or change certain dietary habits, the care plan can be modified accordingly. Caregivers, including family members, may help by identifying realistic and unrealistic expectations on the basis of the patient's habits and lifestyle.[8]

Settings for Geriatric Assessment

GA originated in the inpatient setting, but it has since taken on significance in the outpatient setting and across the continuum of care in integrated delivery systems. It may play a critical role in hospital discharge planning and during the transition from independent living to assisted living or nursing home. Formal GA programs are available on a limited basis only, because of the nationwide shortage of trained geriatricians, and are more likely to be found in large regional or academic medical centers. Local community hospitals, visiting nurse agencies, adult day programs and senior centers may provide some of the resources necessary to provide comprehensive geriatric care, such as social work support, home nursing care and safety evaluations, and physical and occupational therapy.

INPATIENT MODELS

Community Hospital

The benefits of GA at a community rehabilitation hospital were first reported in 1990, in a 1-year randomized, controlled trial involving 155 functionally impaired elderly patients recovering from acute medical or surgical illnesses and considered at

Table 3 Selected Internet Resources for Geriatric Care

Organization	*Web Site*
Administration on Aging	www.aoa.gov
Aging Network Services	www.agingnets.com
Alzheimer's Association	www.alz.org
American Academy of Home Care Physicians	www.aahcp.org
American Association of Homes and Services for the Aging	www.aahsa.org
American Geriatrics Society	www.americangeriatrics.org
American Hospital Association	www.aha.org
American Medical Directors Association	www.amda.com
American Seniors Housing Association	www.seniorshousing.org
American Society of Consultant Pharmacists	www. ascp.com
Assisted Living Federation of America	www.alfa.org
CareGuide	www.careguide.net
Family Caregiver Alliance	www.caregiver.org
Gerontological Society of America	www.geron.org
Medicare	www.medicare.gov
National Adult Day Services Association	www.ncoa.org/nadsa
National Association for Home Care	ww.nahc.org
National Association of Area Agencies on Aging	www.n4a.org
National Association of Professional Geriatric Care Managers	www.caremanager.org
National Institute on Aging	www.nia.nih.gov
Visiting Nurse Associations of America	www.vnaa.org

Table 4 Activities of Daily Living and Instrumental Activities of Daily Living

Activities of daily living (ADL)	Feeding Dressing Ambulation Toileting Bathing Transfer (from bed to chair) Continence Grooming
Instrumental activities of daily living (IADL)	Using telephone Shopping Food preparation Housekeeping Laundry Travel, use of transportation Managing medications Managing finances

risk for nursing home placement. Compared with usual care, treatment in a GA unit resulted in improved function and decreased risk of nursing home placement. The beneficial effects of inpatient GA on mortality and function appeared greatest for patients at a moderate rather than high risk of nursing home placement.[13] The researchers of this study also found that improved outcomes from GA required an investment in rehabilitation that was not totally offset by decreased institutional charges in the following year.[14] Since then, less costly and labor-intensive approaches to inpatient GA have been developed and implemented, including limited support from a geriatric nurse practitioner and use of current hospital personnel to educate attending nurses and staff physicians.[15] These options include wards or beds assigned to elderly patients and consultation services dedicated to the care of elderly persons.

Geriatric Acute Care Units

In 1990, the University Hospitals of Cleveland established an Acute Care for Elders (ACE) unit that reengineered the process of caring for elderly patients to improve functional outcomes. This program comprised four key elements: (1) a specially prepared environment (e.g., uncluttered hallways, large clocks and calendars, and handrails); (2) patient-centered care emphasizing independence, including specific protocols for prevention of disability and for rehabilitation; (3) discharge planning with the goal of returning the patient to his or her home; and (4) intensive review of medical care to minimize the adverse effects of procedures and medications. A randomized, controlled trial of 651 acutely ill patients 70 years of age or older demon-

Table 5 Evidence-Based Preventive Services Recommended for the General Population 65 Years of Age and Older[9]

Service	*Activity*	*Frequency*	*Condition to Detect or Prevent*
Screening	Blood pressure	At least annually	Hypertension
	Mammography	Every 2–3 yr	Breast cancer
	Fecal occult blood testing	Annually	Colorectal cancer
	and/or		
	Flexible sigmoidoscopy	Every 3–5 yr	—
	or		
	Colonoscopy	Once	—
	Pap smear	Every 3 yr or less*	Cervical cancer
	Height and weight	At least annually	Obesity, malnutrition
	Alcoholism questionnaire	†	Alcoholism
	Serum lipids in persons with angina, prior MI, DM	Annually	Risk factor for CAD
	Vision testing	Annually	Visual impairment
	Hearing testing	Annually	Hearing loss
Counseling	Low-fat, well-balanced diet	Annually	Obesity, CAD
	Adequate calcium intake	Annually	Osteoporosis
	Physical activity	Annually	Immobility, CAD, osteoporosis
	Injury prevention	Annually	Falls, motor vehicle accidents, burns, other injuries
	Smoking cessation	‡	COPD, many cancers, CAD
	Regular dental visits	Annually	Malnutrition, oral cancers, endentulism
Immunization	Influenza vaccination	Each fall	Influenza
	Pneumococcal vaccination	§	Pneumococcal disease
	Tetanus booster	Every 10 yr	Tetanus
Chemoprophylaxis	Discussion, implementation of HRT in women	¶	Osteoporosis
	Aspirin therapy after MI	Daily	Additional MI, TIA, or stroke

*May stop screening at age 65 if patient had regularly normal smears up to that age; if never tested before age 65, may stop after two normal annual smears.

†Perform at initial visit and whenever problem drinking is suspected.

‡Discuss at every visit of patients who smoke.

§Immunize once at age 65 for immunocompetent patients; revaccinate high-risk immunocompromised patients after 7–10 yr.

¶Discuss at menopause and at least one time after age 65.

CAD—coronary artery disease COPD—chronic obstructive pulmonary disease DM—diabetes mellitus HRT—hormone replacement therapy MI—myocardial infarction TIA—transient ischemic attack

Table 6 Checklist for Assessing Environmental Hazards in the Home

House or rooms	Lack of cleanliness Inadequate insulation Inadequate lighting Hot-water temperature not regulated at 110° F Inaccessible entries and exits
Stairs or furniture	Unsafe stairs Steps with sharp or broken edges Pointed, broken, or hanging furniture (e.g., suspended chairs) Wheels on furniture
Bathroom	Absence of handrails or secure railings Absence of skid strips for bathtub Outdated prescription drugs Absence of working toilet or acceptable substitute
Signs of neglect	Old food in refrigerator Unwashed dishes Insufficient supply of food Accumulated clothing or newspapers Other signs of disrepair
Obvious dangers	Loose, frayed, or worn carpets; scatter rugs Loose or waxed flooring Absence of smoke alarm Frayed electrical cords or overloaded outlets Exposed pipes and radiators

strated that at the time of hospital discharge, patients admitted to the ACE unit were better able to perform basic ADL and more likely to return home. However, by 90 days after discharge, the functional status of ACE patients was similar to that of standard care patients.[16] However, the costs to the hospital for ACE unit care were less than those for usual care.[17] A recently published trial of 1,531 community-dwelling patients found that an ACE unit in a community hospital improved the process of care and patient and provider satisfaction without increasing hospital length of stay or costs.[18] Publications such as *The ACE (Acute Care for Elders) Manual: Meeting the Challenge of Providing Quality and Cost-Effective Hospital Care to Older Adults*[19] provide valuable information and practical advice for health care providers interested in establishing an ACE unit at their local hospital.

Geriatric Evaluation and Management Units

In the United States, the Department of Veterans Affairs helped to pioneer inpatient geriatric evaluation and management units (GEMUs), which undertake interdisciplinary diagnosis to improve the health of frail elderly patients. The GEMU team, consisting of a geriatrician, a social worker, and a nurse, follows standard protocols for geriatric evaluation and management, with specific instructions in five areas: (1) complete the history taking and physical examination, including screening for geriatric syndromes such as falls or incontinence; (2) develop a list of problems; (3) assess the patient's functional, cognitive, affective, and nutritional status; (4) evaluate the caregiver's capabilities; and (5) assess the patient's social situation. A care plan is developed on the basis of that evaluation, and the team meets at least twice a week to discuss the plan. Preventive and management services (e.g., physical and occupational therapy, nutrition, clinical pharmacy) are coordinated to address the problems identified, with a general emphasis on maintaining functional status.[2] Although the Department of Veterans Affairs mandated that every VA facility have a GEMU by 1996, there are at present only 110 GEMUs functioning in the United States.[20] In the United Kingdom, the counterpart to the GEMU, the geriatric day hospital, continues to play a major role in the multidisciplinary assessment and rehabilitation of elderly people, despite limited evidence to support its efficacy.[21]

Inpatient Consultation Services

If GA is performed as a consultation, special care must be taken in transmitting the findings and recommendations to the primary care provider.[6] The transition between the GA team and the primary care physician frequently presents the greatest obstacle to achieving optimal care of the elderly patient.[3] The GA literature depicts a range of consulting services, from single-focus services (addressing delirium, falls, and other geriatric syndromes) to more comprehensive programs. For example, the Hospital Elder Life Program, developed by Inouye and colleagues, employs an interdisciplinary team that works closely with primary nurses. The team includes a geriatric nurse specialist and an elder life specialist. The latter is a team member with a bachelor's degree in human services or a health care–related field who has experience in geriatrics and management; this person is responsible for program operations, interventions, and volunteer coordination. The Elder Life Program is unique in its hospital-wide focus, in providing skilled staff and volunteers to implement interventions, and in targeting practical interventions toward evidence-based risk factors. In elderly persons who are cognitively impaired, sleep deprived, immobile, dehydrated, and vision or hearing impaired, the program has successfully implemented a nonpharmacologic sleep protocol, documented a reduction in incidence delirium, and prevented cognitive and functional decline.[22]

OUTPATIENT MODELS

Primary Care

Outpatient GA programs may be used as an adjunct to or a substitute for routine primary care. A consultative or primary care team—usually consisting of a physician, nurse (or nurse practitioner), and social worker—provides GA services in the ambulatory setting. Whenever appropriate, referrals are made to other providers (e.g., podiatry, neurology, physical therapy). These programs utilize self-administered questionnaires to screen for common conditions in elderly persons. Patients can complete these questionnaires in the waiting room or with the assistance of a trained receptionist.[23] For example, Moore and colleagues developed and tested a 10-minute office-staff administered screening measure to evaluate malnutrition or weight loss, visual impairment, hearing loss, cognitive impairment, urinary incontinence, depression, physical limitations, and reduced leg mobility. In a large randomized, controlled trial, the measure was associated with more frequent detection and assessment of hearing loss but did not appear to affect detection and management of other conditions included in the screen or health status at 6 months.[24]

Preventive medical care may be provided by the medical staff and psychosocial issues addressed by the nurse or social worker. Lachs and colleagues proposed a short, simple approach that primary care providers in office practice can use to routinely screen

the functional status of elderly patients. The approach focuses on a limited number of functions that often cause problems for elderly patients but that conventional histories and physical examinations often fail to uncover. The screening instrument, which employs brief questions and easily observed tasks, consists of carefully selected tests of vision, hearing, arm and leg function, urinary incontinence, mental status, instrumental and basic ADL, environmental hazards, and social support systems; and it may be administered by a physician or nurse.[25]

Outpatient GA can benefit caregivers as well as patients. In a randomized trial of 568 high-risk, community-dwelling elderly persons who received outpatient GA or usual care for 6 months, caregivers of participants in the GA group were less than half as likely as the usual-care group (16.7% versus 38.5%)to report increased burden during the 1-year follow-up period.[26] Moreover, participants given outpatient GA were significantly less likely than the usual-care patients to lose functional ability, to experience increased health-related restrictions in their daily activities, to have possible depression, or to use home health care services during the 12 to 18 months after randomization. Mortality, use of most health services, and total Medicare payments did not differ significantly between the two groups. The intervention cost $1,350 a person.[27]

In 1997, Beck and colleagues described another GA model for use in primary care. In this model, elderly patients with high health services utilization and one or more chronic conditions had monthly group visits with their primary care physician and nurse. Visits included health education and prevention measures, along with opportunities for socialization, mutual support, and one-to-one consultations with physicians, when necessary. A 12-month randomized, controlled trial of 321 chronically ill HMO enrollees 65 years of age and older showed that group visits resulted in reduced repeat hospital admissions and emergency care use, reduced cost of care, more effective delivery of certain preventive measures, and increased patient and physician satisfaction.[28]

Veterans Affairs Medical Centers

There is a long-established tradition of GA within the VA system. In a randomized, controlled trial involving 160 frail elderly outpatients that compared ongoing outpatient GA through the VA system with usual care, GA proved significantly more effective in reducing mortality, increasing patient satisfaction, and improving the quality of health and social care. It did not, however, reduce health care utilization or costs.[29] In a recent 2-year randomized, controlled trial, 128 veterans 65 years of age or older were treated with outpatient GA or usual care. GA resulted in greater improvement in perceived health; smaller increases in numbers of clinic visits and impairment in instrumental ADL; improved social activity; and greater improvement in measures of depression, general well-being, life satisfaction, and cognitive status. There were no significant treatment effects in ADL scores, number of hospitalizations, or mortality.[30]

Consultation Services

GA can be a consultative service provided through hospital-based outpatient clinics. In a 12-month multicenter, randomized, controlled trial of outpatient consultative GA versus usual care in 442 elderly patients with a health problem or recent change in health status, outpatient consultative GA led to significantly improved diagnosis of cognitive impairment, depression, and incontinence; to psychological and emotional benefits for patients; and to reduced levels of caregiver stress.[31]

Even a single outpatient GA consultation, coupled with subsequent intervention to improve adherence to consultative recommendations, can be beneficial in elderly patients with identified problems.[32,33] In a 15-month randomized, controlled trial of 363 community-dwelling elderly persons in whom screening identified one or more specific geriatric conditions (falls, urinary incontinence, depressive symptoms, or functional impairment), intervention not only prevented functional and health-related quality-of-life decline but also was cost effective.[32]

In-Home Geriatric Assessment

GA in the home may be provided as part of an ongoing office-based program, as an extension of hospitalization through a post–acute care program,[34-36] or as a freestanding entity.[37] In-home GA includes the assessment of environmental hazards, which can serve an important preventive function [*see Table* 6].

The physician's role in in-home GA is to serve as a member of an interdisciplinary team composed of nurses, therapists (speech, physical, occupational, and respiratory), social workers, personal care attendants, home medical equipment suppliers, and informal caregivers (e.g., family members and friends).[38] In-home GA is used to establish a baseline, monitor the course of an illness, and evaluate the effects of an intervention. A recent meta-analysis of 18 randomized, controlled trials found that preventive home-visitation programs (based on multidimensional GA that included multiple follow-up home visits) helped prevent nursing home admission and functional decline, particularly in low-risk elderly persons.[39]

INTEGRATED DELIVERY SYSTEMS

GA programs have been developed and implemented by health systems to improve the evaluation and management of older adults. Managed care organizations and other integrated delivery systems have employed registered nurses, nurse practitioners, social workers, or public health workers as care or case managers to follow, organize, and coordinate care for elderly persons.[40,41] Chronic disease management programs incorporate the basic principles of GA in a variety of different settings, including outpatient clinics, senior centers, and home care.[42-44]

Programs of All-Inclusive Care for the Elderly (PACE)

More than 6,000 Medicare nursing home certifiable beneficiaries receive care from PACE, a program designated for expansion under the Balanced Budget Act of 1997.[45] PACE replicates the model of comprehensive, community-based geriatric care pioneered by On Lok. This model enrolls frail elderly persons who meet states' criteria for nursing home care and uses interdisciplinary teams to assess the participants and to deliver care in appropriate settings. Analysis of data from PACE's minimum data set reveals that short-term hospital utilization among PACE participants is lower than that for other older and disabled populations.[46]

Reimbursement

There is currently no explicit Medicare reimbursement mechanism to cover hospital or physician GA services in the United States. Elderly persons can receive GA in the context of hospitalization for acute illness, provided that the assessment does not prolong the hospital length of stay. A number of Diagnostic Related Groups (DRGs) of the Medicare system apply to frail elderly patients and can be used to justify the hospitalization of pa-

tients with coexisting conditions who require GA. Documentation of the need for hospitalization is critical.[47] As managed care, PACE receives capitated payment from Medicare and Medicaid on the basis of the rate structure of the Medicare+Choice payment system, adjusted for the comparative frailty of PACE enrollees and other factors deemed to be appropriate by the Secretary of Health and Human Services.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Fried LP, Storer DJ, King DE, et al: Diagnosis of illness presentation in the elderly. J Am Geriatr Soc 39:117, 1991

2. Cohen HJ, Feussner JR, Weinberger M, et al: A controlled trial of inpatient and outpatient geriatric evaluation and management. N Engl J Med 346:905, 2002

3. Stuck AE, Siu AL, Wieland GD, et al: Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet 342:1032, 1993

4. Winograd CH, Gerety MB, Chung M, et al: Screening for frailty: criteria and predictors of outcomes. J Am Geriatr Soc 39:778, 1991

5. Manton KG, Gu X: Changes in the prevalence of chronic disability in the United States black and nonblack population above age 65 from 1982 to 1999. Proc Natl Acad Sci U S A 98:6354, 2001

6. Osterweil D, Brummel-Smith K, Beck JC: Comprehensive Geriatric Assessment. McGraw-Hill Book Co, New York, 2000

7. Holman H, Lorig K: Patients as partners in managing chronic disease: partnership is a prerequisite for effective and efficient health care. BMJ 320:526, 2000

8. Fulmer TT, Hyer K: Geriatric interdisciplinary teams. The Merck Manual of Geriatrics. Beers MH, Berkow R, Eds. Merck & Co, Whitehouse Station, New Jersey, 2000

9. Bloom HG, Edelberg HK: Prevention.Geriatrics Review Syllabus: A Core Curriculum in Geriatric Medicine, Fifth ed. Cobbs EL, Duthie EH, Eds. Kendall/Hunt Publishing Co, Dubuque, Iowa, 2002, p 61

10. Edelberg HK, Wei JY: Primary-care guidelines for community-living older persons. Clin Geriatrics 7:42, 1999

11. U.S. Preventive Services Task Force. Guide to Clinical Preventive Services: Report of the U.S. Preventive Services Task Force. Williams & Wilkins, Baltimore, 1996

12. Comprehensive functional assessment for elderly patients. American College of Physicians Health and Public Policy Committee. Ann Intern Med 109:70, 1988

13. Applegate WB, Miller ST, Graney MJ, et al: A randomized, controlled trial of a geriatric assessment unit in a community rehabilitation hospital. N Engl J Med 322:1572, 1990

14. Applegate WB, Graney MJ, Miller ST, et al: Impact of a geriatric assessment unit on subsequent health care charges. Am J Public Health 81:1302, 1991

15. Cefalu CA, Colbourne G, Duffy M, et al: A university-affiliated community hospital inpatient geriatrics program functioning in an administrative and educational capacity. J Am Geriatr Soc 45:355, 1997

16. Landefeld CS, Palmer RM, Kresevic DM, et al: A randomized trial of care in a hospital medical unit especially designed to improve the functional outcomes of acutely ill older patients. N Engl J Med 332:1338, 1995

17. Covinsky KE, Palmer RM, Kresevic DM, et al: Improving functional outcomes in older patients: lessons from an acute care for elders unit. Jt Comm J Qual Improv 24:63, 1998

18. Counsell SR, Holder CM, Liebenauer LL, et al: Effects of a multicomponent intervention on functional outcomes and process of care in hospitalized older patients: a randomized controlled trial of Acute Care for Elders (ACE) in a community hospital. J Am Geriatr Soc 48:1572, 2000

19. Counsell SR, Holder C, Liebenauer MA, et al: The ACE (Acute Care for Elders) Manual: Meeting the Challenge of Providing Quality and Cost-Effective Hospital Care to Older Adults. Summa Health Systems, Akron, Ohio, 1998

20. Wieland D, Hedrick SC, Rubenstein LZ, et al: Inpatient geriatric evaluation and management units: organization and care patterns in the Department of Veterans Affairs. Gerontologist 34:652, 1994

21. Black DA: The modern geriatric day hospital. Hosp Med 61:539, 2000

22. Inouye SK, Bogardus ST Jr, Baker DI, et al: The Hospital Elder Life Program: a model of care to prevent cognitive and functional decline in older hospitalized patients. J Am Geriatr Soc 48:1697, 2000

23. Wasson JH, Jette AM, Johnson DJ, et al: A replicable and customizable approach to improve ambulatory care and research. J Ambulatory Care Manage 20:17, 1997

24. Moore AA, Siu AL, Partridge JM, et al: A randomized trial of office-based screening for common problems in older persons. Am J Med 102:371, 1997

25. Lachs MS, Feinstein AR, Cooney LM Jr, et al: A simple procedure for general screening for functional disability in elderly patients. Ann Intern Med 112:699, 1990

26. Weuve JL, Boult C, Morishita L: The effects of outpatient geriatric evaluation and management on caregiver burden. Gerontologist 40:429, 2000

27. Boult C, Boult LB, Morishita L, et al: A randomized clinical trial of outpatient geriatric evaluation and management. J Am Geriatr Soc 49:351, 2001

28. Beck A, Scott J, Williams P, et al: A randomized trial of group outpatient visits for chronically ill older HMO members: the Cooperative Health Care Clinic. J Am Geriatr Soc 45:543, 1997

29. Toseland RW, O'Donnell JC, Engelhardt JB, et al: Outpatient geriatric evaluation and management: results of a randomized trial. Med Care 34:624, 1996

30. Burns R, Nichols LO, Martindale-Adams J, et al: Interdisciplinary geriatric primary care evaluation and management: two-year outcomes. J Am Geriatr Soc 48:8, 2000

31. Silverman M, Musa D, Martin DC, et al: Evaluation of outpatient geriatric assessment: a randomized multi-site trial. J Am Geriatr Soc 43:733, 1995

32. Reuben DB, Frank JC, Hirsch SH, et al: A randomized clinical trial of outpatient comprehensive geriatric assessment coupled with an intervention to increase adherence to recommendations. J Am Geriatr Soc 47:269, 1999

33. Keeler EB, Robalino DA, Frank JC, et al: Cost-effectiveness of outpatient geriatric assessment with an intervention to increase adherence. Med Care 37:1199, 1999

34. Naylor MD, Brooten D, Campbell R, et al: Comprehensive discharge planning and home follow-up of hospitalized elders: a randomized clinical trial. JAMA 281:613, 1999

35. Siu AL, Kravitz RL, Keeler E, et al: Postdischarge geriatric assessment of hospitalized frail elderly patients. Arch Intern Med 156:76, 1996

36. Landi F, Gambassi G, Pola R, et al: Impact of integrated home care services on hospital use. J Am Geriatr Soc 47:1430, 1999

37. Rockwood K, Stadnyk K, Carver D, et al: A clinimetric evaluation of specialized geriatric care for rural dwelling, frail older people. J Am Geriatr Soc 48:1080, 2000

38. Eleazer GP: Community-based care. Geriatrics Review Syllabus: A Core Curriculum in Geriatric Medicine, Fifth ed. Cobbs EL, Duthie EH, Eds. Kendall/Hunt Publishing Co, Dubuque, Iowa, 2002, p 61

39. Stuck AE, Egger M, Hammer A, et al: Home visits to prevent nursing home admission and functional decline in elderly people: systematic review and meta-regression analysis. JAMA 287:1022, 2002

40. Marshall BS, Long MJ, Voss J, et al: Case management of the elderly in a health maintenance organization: the implications for program administration under managed care. J Health Manag 44:477, 1999

41. Gagnon AJ, Schein C, McVey L, et al: Randomized controlled trial of nurse case management of frail older people. J Am Geriatr Soc 47:1118, 1999

42. Rich MW, Beckham V, Wittenberg C, et al: A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. N Engl J Med 333:1190, 1995

43. Leveille SG, Wagner EH, Davis C, et al: Preventing disability and managing chronic illness in frail older adults: a randomized trial of a community-based partnership with primary care. J Am Geriatr Soc 46:1191, 1998

44. Coleman EA, Grothaus LC, Sandhu N, et al: Chronic care clinics: a randomized controlled trial of a new model of primary care for frail older adults. J Am Geriatr Soc 47:775, 1999

45. Temkin-Greener H, Meiners MR, Gruenberg L: PACE and the Medicare+Choice risk-adjusted payment model. Inquiry 38:60, 2001

46. Wieland D, Lamb VL, Sutton SR, et al: Hospitalization in the Program of All-Inclusive Care for the Elderly (PACE): rates, concomitants, and predictors. J Am Geriatr Soc 48:1373, 2000

47. Fillit H: Challenges for acute care geriatric inpatient units under the present Medicare prospective payment system. J Am Geriatr Soc 42:553, 1994

IX MANAGEMENT OF COMMON CLINICAL DISORDERS IN GERIATRIC PATIENTS

Robert M. Palmer, m.d.

The process of aging predisposes elderly patients to homeostatic failure, chronic disease, and a loss of independence in the performance of daily activities (functional decline). Consequently, clinical disorders often manifest as geriatric syndromes, which are multifactorial health conditions that occur when the accumulated effect of impairments in multiple systems renders a person vulnerable to situational challenges.[1] Geriatric syndromes increase in frequency and clinical importance in advanced age, growing more prevalent in patients older than 75 years. These syndromes are frequently encountered by physicians in ambulatory, hospital, and long-term care settings.

Geriatric syndromes reflect the presence of common chronic conditions in the elderly population. In the United States, an estimated 90 million people live with at least one chronic condition that threatens their independent self-care or compromises the quality of their lives. Rates of chronic conditions are highest among the elderly, 88% of whom have at least one chronic condition.[2] By identifying, evaluating, and treating these common conditions, physicians can help older patients maximize their physical and mental status.

Delirium

Common causes of cognitive dysfunction in elderly patients are delirium, dementia, and depression. Dementia and depression are discussed elsewhere [*see CE:X Symptom Management in Palliative Medicine, 11:IV Cerebrovascular Disorders, 11:VII Anoxic, Metabolic, and Toxic Encephalopathies, 11:XI Alzheimer Disease and the Dementias, 11:XIV Pain, and 13:II Depression and Bipolar Mood Disorders*].

Delirium, an acute disorder of attention and global cognitive function, is a common and potentially preventable cause of adverse health outcomes. Elderly patients who develop delirium are at risk for future functional decline (a loss of independence in their performance of daily activities), cognitive decline, and increased mortality.[3,4] Delirium is also associated with prolonged hospitalization, increased risk of nursing home placement, persistent functional decline, and debilitating complications (e.g., falls, injury, and immobility).

The criteria for delirium caused by a general medical condition include the following: disturbance of consciousness (i.e., reduced awareness of the environment) with reduced ability to focus, sustain, or shift attention; change in cognition (e.g., memory deficit, disorientation, language disturbance) or perceptual disturbance that is not better accounted for by a preexisting, established, or evolving dementia; increased or reduced psychomotor activity; disorganized sleep-wake cycle; acute onset of disturbance (usually hours to days) with fluctuation over the course of the day; and evidence from the history, physical examination, or laboratory findings that the disturbance is caused by an etiologically related general medical condition.[5,6] The cognitive disturbances contribute substantially to problems in identifying and managing delirium.

epidemiology

Medical Illness

Delirium is most often identified in hospitalized patients. It is present in 10% to 15% of older patients at hospital admission for treatment of medical illnesses, and it develops in 10% to 15% of older medical patients during hospitalization.[3] Independent risk factors for delirium in elderly medical patients during hospitalization include the use of psychoactive medications, severe illness, cognitive impairment (dementia), vision impairment, and a high ratio of blood urea nitrogen to creatinine, implying dehydration.[3,7] Precipitating factors for delirium in hospitalized elderly persons include the use of physical restraints, more than three medications added to the patient's drug regimen, bladder catheterization, and any iatrogenic event (e.g., unintentional injury).[8] Benzodiazepines, opioids (meperidine), and medications that have central anticholinergic effects, including diphenhydramine, are most often associated with incident delirium.[9] These findings establish ways to both identify high-risk patients and prevent incident delirium through appropriate medical, behavioral, and pharmacologic approaches.

Postoperative Delirium

Postoperative delirium occurs in 10% to 15% of older general surgery patients and in 30% to 60% of older patients admitted with hip fractures or who are undergoing knee arthroplasty.[10] The incidence of delirium or other cognitive impairments is not related to the type of anesthesia administered, whether general or epidural. In a large prospective study, postoperative cognitive dysfunction was present in 25.8% of patients 60 years of age and older 1 week after surgery.[11] Although both surgery and anesthesia are factors contributing to the development of both short-term and long-term postoperative cognitive impairment, uncertainty remains about the specific factors that contribute to postoperative delirium. Putative risk factors for postoperative delirium include polypharmacy, the use of preoperative anticholinergic drugs, cognitive impairment, and older age.[10] Other contributing factors are intraoperative hypoxemia, perioperative hypotension, and postoperative complications.

Among low-risk adult patients undergoing elective noncardiac surgery, postoperative delirium occurred in 9% and was independently correlated with an age of 70 years or older, self-reported alcohol use, poor cognitive status, poor functional status, markedly abnormal preoperative serum sodium level, and thoracic surgery.[12] In these patients, the postoperative use of meperidine and benzodiazepines increased the risk of delirium.[12]

risk factors and etiology

Virtually any acute physical stress can precipitate delirium in elderly patients, particularly those with risk factors [*see Table 1*]. In medically ill patients, delirium is most commonly associated with acute infections, such as pneumonia and urosepsis; hypoxemia; hypotension; and use of psychoactive medications.[13] The last include many antiarrhythmic agents, tricyclic antidepressants, neuroleptics, gastrointestinal medications, and antihistamines. These agents used in large doses or in combination at therapeutic doses

Table 1 Prevention and Management of Delirium

	Predisposing or Precipitating Factors	*Interventions*
Prevention	Psychotropic medications Benzodiazepines Alcohol Anticholinergic agents Opiates	Avoid or discontinue antihistamines (e.g., diphenhydramine) and benzodiazepines Observe for possible alcohol withdrawal Titrate doses of opiates to achieve analgesia Promote sleep by applying relaxation techniques and keeping hallways quiet at night
	Preexisting cognitive impairment or dementia Alzheimer disease Vascular cognitive impairment Major depressive disorder	Orient to date, place, caregivers, and upcoming events Use psychotropic medications judiciously to treat specific symptoms Encourage visits by family members or sitters
	Dehydration High BUN-to-creatinine ratio Hypotension Poor skin turgor	Prescribe fluids, and monitor intake and discharge Utilize intravenous hydration if oral intake is inadequate Have nursing staff assist with eating and drinking Monitor basic electrolytes
	Severe illness High fever Infection Hypotension	Diagnose illness quickly and follow with empirical therapy (e.g., intravenous antibiotics for community-acquired pneumonia) Optimize hemodynamic status (ensure adequate oxygenation and cardiac output)
	Sensory impairment Vision Hearing	Utilize visual aids (e.g., corrective lenses, large-print books, diffuse illumination) Utilize hearing aids (e.g., portable amplifying devices, cerumen disimpaction)
	Immobility/physical restraint Bed-rest orders Mechanical or chemical restraints	Ambulate or apply bedside exercises Obtain physical therapy consultation for gait assessment and strengthening exercises Obtain occupational therapy consultation for functional assessment and adaptive devices
	Interventions	*Suggested Guidelines*
Management	Medications Neuroleptics Anxiolytics Opiates	Judiciously use neuroleptics to treat disturbing psychiatric symptoms (e.g., delusions, hallucinations, aggressiveness) Administer haloperidol, 0.5–1.0 mg orally or parenterally, every 6–8 hr (higher doses are justified when there is a risk that patient will disrupt life-sustaining therapies) Administer risperidone, 0.5–1.0 mg orally daily in divided doses, when long-term neuroleptic therapy is warranted Administer lorazepam, 0.5–1.0 mg every 6–8 hr orally or parenterally, to patients with disturbing anxiety (higher doses for benzodiazepine or alcohol withdrawal) Control acute pain with morphine sulfate, administered slowly in low initial doses (e.g., start with 2–4 mg parenterally every 3 hr); monitor pain relief with visual analogue scale Avoid meperidine
	Nursing care	Keep sitter at bedside Move patient closer to nurses' station Implement preventive measures Educate patient and family about delirium
	Environmental measures	Keep hallways or corridors quiet Dim lights at night Provide recreational activities outside patient's room Reduce clutter in rooms and hallways

BUN—blood urea nitrogen

may induce delirium. Antibiotics (e.g., ciprofloxacin), analgesics (e.g., nonsteroidal anti-inflammatory drugs), opiates (e.g., meperidine), and H_2 receptor antagonists (e.g., cimetidine) have been associated with delirium.[14] Other important causes of delirium include alcohol and drug intoxication and withdrawal and neurologic illness (e.g., stroke, tumor, or infection).

PATHOPHYSIOLOGY

The pathophysiology of delirium in elderly patients is complex. Patients with risk factors for delirium are regarded as vulnerable on the basis of the hypothetical assumption of limited homeostatic (brain) reserves. Global functional disturbances are associated with neurotransmitter disturbances (e.g., acetylcholine, dopamine, and serotonin). For example, depressed acetylcholine transmission secondary to anticholinergic drugs with central effects can disturb normal cognition.[14] Patients with Alzheimer disease appear to be sensitive to the central anticholinergic effects of various drugs and show improved cognition after treatment with cholinesterase inhibitors, which potentiate the effects of acetylcholine. In some conditions, such as he-

patic encephalopathy, the presence of delirium correlates with the accumulation of toxic metabolites (e.g., ammonia). Disturbances in the elaboration of cytokines, including interleukin-1 (IL-1) and IL-2, have also been associated with delirium.[15]

DIAGNOSTIC EVALUATION

Clinical Features

The patient with delirium presents with an acute change in mental status and clinical features of disturbed consciousness, impaired cognition, and a fluctuating course. The patient has a reduced ability to focus (i.e., sustain or shift) attention, which is associated with incoherent or tangential speech and disorganized or erratic thought processes. Perceptual disturbances, such as misperceptions, illusions or frank delusions, and hallucinations, are often accompanied by increased psychomotor activity. Most patients with delirium vacillate between hypoalertness and hyperalertness. However, the majority of elderly patients with delirium have a so-called quiet confusion, with fluctuations in behavior and level of cognition throughout the day (e.g., increased confusion in the evening, referred to as sundowning).

Several instruments or scales have been found to have significant likelihood ratios and reasonable specificity for the diagnosis of delirium.[3] Of these tools, the confusion assessment method is often used to make a diagnosis in clinical studies and is practical for bedside application.[3] The diagnosis of delirium by the confusion assessment method requires (1) the presence of inattention that has an acute onset and a fluctuating course and (2) disorganized thinking or altered level of consciousness, or both [*see Table 2*]. The confusion assessment method has been adapted for use in patients who are admitted to intensive care units.[16]

Table 2 Confusion Assessment Method for Diagnosis of Delirium*

Factor	*Assessment*
1. Delirium of acute onset and fluctuating course	Indicated by positive response to the following questions: Is there evidence of acute change in the patient's mental status? Does the abnormal behavior fluctuate during the day (appear to come and go or increase or decrease in intensity)?
2. Inattention	Indicated by positive response to the following questions: Does the patient have difficulty focusing attention? Does the patient have difficulty keeping track of what is being said?
3. Disorganized thinking	Indicated by positive response to the following questions: Is the patient's thinking disorganized or incoherent (rambling, irrelevant conversation, unclear or illogical flow of ideas, unpredictable changes in topic)?
4. Altered level of consciousness	Indicated by an answer other than "alert" to the following question: Overall, how would you rate this patient's level of consciousness? Normal = alert Hyperalert = vigilant Drowsy, easily aroused = lethargic Difficult to arouse = stupor Unarousable = coma

*The diagnosis of delirium requires the presence of features 1 and 2 and either 3 or 4.

Table 3 Differential Features of Delirium and Dementia

Feature	*Delirium*	*Dementia*
Onset	Rapid	Insidious
Duration	Hours to days (transient)	Months to years (persistent)
Attention	Decreased (digit span test ≤ 4 words, distractible, fluctuating)	Usually normal (mild to moderate dementia)
Awareness	Always impaired	Usually normal
Alertness	Fluctuates	Usually normal
Consciousness	Depressed	Normal
Memory	Impaired (varies)	Impaired (recent is better than remote)
Language	Normal or incorrect naming	Aphasia, anomia, paraphasia
Perception	Misperceptions, illusions, hallucinations (common)	Usually normal (or delusions)
Psychomotor activity	Increased, decreased (varies)	Usually normal
Sleep-wake pattern	Disrupted (reversal)	Normal or fragmented

Laboratory Tests

The evaluation of a patient with delirium focuses on a search for the most probable etiologies and the need to treat life-threatening illnesses. Often, the precipitant of delirium is obvious, but careful monitoring is needed to exclude other causes or contributing conditions (e.g., pneumonia accompanied by hypoxemia, electrolyte imbalance, and adverse drug effects). A neuroimaging head scan is rarely useful unless patients have focal neurologic findings. Because infection is so frequently the precipitant of delirium, a complete blood count and a chemistry panel are obtained for all patients when infection cannot be excluded. In select cases, blood and urine cultures are warranted, as are a chest x-ray, an electrocardiogram, arterial blood gas analysis, a lumbar puncture, a computed tomographic scan of the head, and an electroencephalogram. The lumbar puncture is rarely diagnostic in patients who are not immunocompromised or in patients with no signs suggesting a serious diagnosis (e.g., meningitis). Although the EEG is useful in distinguishing delirium from depression or dementia, it only occasionally suggests a diagnosis (e.g., partial-complex seizure disorder) that could not be established on the basis of clinical findings.[3,15]

DIFFERENTIAL DIAGNOSIS

Delirium is often misdiagnosed as dementia, depression, or functional psychosis. The clinical features of delirium and dementia may overlap, but the findings of inattention and altered level of consciousness and the duration of symptoms usually enable a separation of these two diagnoses [*see Table 3*]. Psychosis caused by late-onset schizophrenia or major depression is not characterized by impaired attention or altered level of consciousness and fluctuating mental status. Although many depressed patients perform poorly on formal tests of cognition, they remain alert and attentive, and they do not display the fluctuating course of delirium. The depressed patient often refuses to complete a mental status test or gives many "I don't know" responses. Likewise, many anxious patients have impaired concentration and may appear inattentive, but they do not have the fluctuating course and altered level of consciousness seen in delirium.

Protocols Targeted at Risk Factors

Delirium can be prevented in patients who are at risk for the disorder [*see Table 1*]. In the Elder Life Program, patients at risk for incident delirium were identified shortly after hospital admission and followed daily with the confusion assessment method.[17] An array of protocols targeted at specific risk factors served to optimize cognitive function (reorientation and therapeutic activities), prevent sleep deprivation (relaxation and noise reduction), avoid immobility (ambulation and exercises), improve vision (visual aids and illumination), improve hearing (hearing devices), and treat dehydration (volume repletion). In 852 patients studied, the incidence of delirium was 9.9% in the intervention group, compared with an incidence of 15% in the usual-care group. The total number of days and number of episodes of delirium were also significantly reduced in the intervention group, but no significant effects were seen on the severity of delirium or on recurrence rates. The prevention of postoperative delirium in patients undergoing hip surgery for fracture was demonstrated in a pilot study. Geriatric consultation accompanied by targeted recommendations resulted in an incidence of delirium of 32% in intervention-group patients, compared with an incidence of 50% in usual-care patients.[18]

Increased Socialization

Other nursing and environmental interventions that are often helpful in the management of delirium include placement of a patient in a room near the nurses' station for greater observation and socialization; social visits with family members, a caregiver, or hired sitter; the avoidance of physical restraints, which can aggravate agitation; and the promotion of normal sleep cycles through noise control and dim lighting at night.

Pharmacotherapy

Pharmacologic interventions are often warranted for patients with specific indications, including bothersome hallucinations and delusions, physical aggression and risk of harm to self or others, or severe personal distress or discomfort. Pharmacologic interventions are often utilized in patients in whom disruption of medical therapy must be avoided, such as those who are critically ill. However, no agents have been approved by the Food and Drug Administration specifically for the treatment of delirium. Definitive clinical trials of the drug therapy for elderly patients with delirium are lacking. As a general recommendation, for patients with bothersome symptoms, haloperidol is the usual drug of choice. It is administered in dosages of 0.5 to 1.0 mg given orally or intramuscularly every 6 to 8 hours. Higher doses or more frequent administration (e.g., every 30 minutes) may be needed to control agitation when patients require endotracheal intubation or other lifesaving procedures. Once the agitation is relieved, the doses of haloperidol should be gradually tapered over several days. Adverse effects of neuroleptics include extrapyramidal effects, dystonic reactions, and torsade de pointes, which is associated with high doses of intravenous haloperidol.

For treatment of withdrawal from alcohol or benzodiazepine use or severe anxiety and sleep disturbance, lorazepam, in doses of 0.5 to 1.0 mg given orally or parenterally every 6 to 8 hours, is the usual drug of choice; higher doses are often required for treatment of alcohol withdrawal.

A neuroleptic such as haloperidol or risperidone, 0.5 to 1.0 mg/day given orally, can be used as an adjunct to control psychotic symptoms associated with withdrawal. Haloperidol and lorazepam may also be useful in preparing patients for neuroimaging scans or when invasive or supportive therapies are necessary (e.g., insertion of a central line). Adverse effects of benzodiazepines include paradoxical confusion, amnesia, and falls. Severe pain associated with agitation should be treated parenterally with morphine sulfate with lower than usual adult dosages (e.g., 4 to 6 mg every 3 hours; with lower doses used in very old or frail patients). In general, repeated doses of meperidine should be avoided to prevent the neurotoxic effect of its metabolite, normeperidine. Novel approaches to the prevention or reversal of delirium (e.g., use of cholinesterase inhibitors and newer neuroleptics) await clinical trials.

PROGNOSIS

Treatment of the underlying cause of delirium usually results in rapid improvement, although full resolution of cognitive symptoms can take several days to several months. Functional and cognitive deficits often persist after hospital discharge. In a multicenter cohort study, delirium that occurred in hospitalized elderly patients was associated with an increased risk of nursing home admission and mortality, underscoring the seriousness of the diagnosis of delirium and the need to closely monitor such patients after discharge from the hospital.[6] The persistence of cognitive deficits may be related to underlying conditions, especially dementia.

Urinary Incontinence

Urinary incontinence, the involuntary loss of urine of sufficient severity to be a social or health problem, is a common, costly, and potentially disabling condition that is never a consequence of normal aging; it is always treatable and often curable.[19] Urinary incontinence is a source of social embarrassment for older patients, results in a loss of self-esteem and physical independence, and increases the patient's risk of institutionalization.

EPIDEMIOLOGY

The prevalence of urinary incontinence increases from approximately 10% to 15% in women 65 years of age to more than 25% in men and women 85 years of age and older.[19] The prevalence of incontinence approaches 50% in nursing home residents and frail homebound elderly patients. Despite the publication of guidelines for screening and management of incontinence, this condition is still underrecognized and undertreated by primary care physicians.[20]

RISK FACTORS AND ETIOLOGY

Urinary incontinence in older patients results from an interaction of age-related changes in the genitourinary tract, comorbid conditions, altered cognition, impaired mobility, and environmental barriers to toileting. An overactive bladder associated with changes in the smooth muscle of the bladder, prostatic hyperplasia in men, bladder wall relaxation or prolapse in women, medication side effects, and cognitive impairment are the most common factors predisposing older patients to urinary incontinence. Fecal impaction and urinary tract infections (UTIs) are common causes of acute incontinence, which resolve with appropriate treatment.

PATHOPHYSIOLOGY

Incontinence results from neurologic or anatomic defects that

interfere with normal urinary micturition. The urinary bladder is responsible for the storage and emptying of urine. Lesions that interfere with bladder contraction or emptying will predispose patients to urinary incontinence. Urine storage occurs when the muscular wall (detrusor) relaxes, and micturition occurs when the detrusor contracts. Bladder relaxation and filling are promoted by sympathetic stimulation via the hypogastric plexus (T11-L2), which inhibits parasympathetic tone and augments beta-adrenergic tone. Bladder contraction is initiated through parasympathetic (cholinergic) stimulation via the sacral complex (S2-4) and relaxation of the pelvic floor musculature (external sphincter), which is under control of the somatic (pudendal) nerve. As the urinary bladder fills, contraction of the detrusor muscle is inhibited by closure of the bladder neck under sympathetic tone, somatic innervation of pelvic floor muscles and striated muscles around the urethra, and activation of inhibitory pathways from the brain stem and cerebral (frontal) cortex. Detrusor pressure increases, resulting in involuntary contractions and the subjective sensation of a need to void. During voluntary urination, the increase in detrusor pressure overcomes inhibitory pathways and exceeds urethral resistance.

Lesions that interfere with bladder contraction and emptying (e.g., sensory neuropathy) predispose patients to incontinence. Contractions of the detrusor muscle at low bladder filling volumes (detrusor overactivity or instability) may occur in patients with lesions of the central nervous system (e.g., stroke) or with increased sensory stimulation from the bladder, which may be seen in UTI or prostatic hyperplasia. Incompetence of the internal urethral sphincter (e.g., caused by pelvic relaxation or intrinsic sphincter deficiency) allows urine to leak from the bladder (stress incontinence) whenever intra-abdominal pressure exceeds urethral resistance. Loss of detrusor contractility (acontractile bladder) or bladder outlet obstruction may result in a distended bladder and leakage of urine (overflow incontinence) with or without a subjective sensation of urinary urgency.

DIAGNOSTIC EVALUATION

Incontinence may present as either an acute (transient) or chronic (established) condition. The clinical type and most likely cause of urinary incontinence can be determined by a focused medical history and physical examination, a few laboratory studies, and simple office procedures.[21]

Clinical Features

Acute incontinence Acute incontinence typically has a sudden onset and is associated with an acute illness (e.g., infection or delirium) or iatrogenic event (e.g., polypharmacy or restricted mobility). UTI is the most commonly recognized cause of transient incontinence in ambulatory elderly patients. In hospitalized patients, delirium or acute confusion, excessive infusions of intravenous fluids, and metabolic disorders, such as hyperglycemia with glucosuria, will predispose the patient to incontinence. Acute urinary incontinence is also associated with functional impairments such as an inability to toilet quickly enough or a lack of awareness of the need to urinate (e.g., in patients with dementia or acute confusion). Despite the common distinction between acute and chronic causes of incontinence, the two are often difficult to distinguish and often share a similar pathophysiology.

Established incontinence Four basic types of established urinary incontinence occur in elderly patients: stress, urge, overflow, and functional incontinence [*see Table 4*]. Urinary incontinence is often caused by a combination of two or more of the subtypes of incontinence, in which case it is classified as mixed incontinence.

Medical History

The medical history, the key to diagnosis, includes a description of the onset, duration, and characteristics of incontinence; the most significant symptoms; the frequency, timing, and quantity of episodes; and the precipitants of incontinence, such as the relationship of incontinence to exercise, previous surgery, or onset of acute diseases. In men, symptoms related to prostatism should be elicited (i.e., nocturia, dysuria, hesitancy, or decreased stream). The patient's bowel habits (e.g., constipation or impaction) and fluid intake should be noted. A written record or diary of incontinent episodes recorded for 7 days (or a continence record for nursing home residents) is a reliable measure of the frequency of incontinent episodes; it also provides clues to the etiology and severity of urinary incontinence.[22]

Physical Examination

The physical examination focuses on conditions affecting genitourinary function. The abdominal examination seeks abdominal masses caused by bladder distention or fecal impaction. The rectal examination evaluates patients for fecal impaction, rectal masses, and, in men, abnormalities of the prostate gland. The pelvic examination evaluates women for pelvic organ prolapse, genital atrophy, and urethral abnormalities. Neuromuscular examination identifies patients with impaired mobility or diseases of the central nervous system that predispose to urinary incontinence (e.g., stroke, Parkinson disease, or dementia). Sacral root integrity is evaluated by assessment of anal sphincter tone, anal wink, and perineal sensation.

Laboratory Tests

The laboratory evaluation of an incontinent elderly patient is driven by the clinical type, frequency, and significance of the incontinence. Patients with acute urinary incontinence associated with a UTI should first be treated with antibiotics before further urodynamic evaluation. Although many patients with established incontinence have asymptomatic bacteriuria, there is no evidence that treatment of the infection will lead to resolution of the incontinence.[23] A urinalysis will suggest the presence of infection, possible inflammation, or the need for further diagnostic evaluation to rule out tumor or stones (e.g., if the patient has hematuria or pyuria).

In patients with established incontinence, blood tests should measure renal function, electrolytes, blood glucose, and serum calcium, all of which help to exclude polyuric conditions that may cause incontinence.

The most useful bedside test of lower urinary tract function is measurement of the postvoiding residual (PVR) urine. Accurate measurement of the PVR is most often accomplished by straight catheterization of the urinary bladder after the patient attempts complete voiding. Pelvic ultrasonography and portable bladder scanning are safe and accurate alternative methods of estimating PVR.[24] A PVR of less than 50 ml of urine is considered normal. A PVR of greater than 150 ml is abnormal even in elderly patients and indicates the need for further urologic evaluation or repeat measurement of PVR.

When the cause of incontinence is not apparent after these tests, additional studies often performed by urologic specialists

Table 4 Types, Characteristics, and Treatments of Urinary Incontinence

Types	*Characteristics*	*Treatments*		
		Nondrug	*Drug*	*Precautions*
Stress	Associated with urethral hypermobility or intrinsic sphincter deficiency (ISD); urinary leakage with an increase in intra-abdominal pressure resulting from coughing, sneezing, or physical exertion; continuous leakage with ISD	*Behavioral:* pelvic muscle exercises (repeated contraction of pelvic muscles); scheduled toileting (urinate at defined intervals), bladder training (interrupt urination, avoid straining) *Surgical:* needle bladder neck suspension for urethral hypermobility; vaginal sling or periurethral (collagen) injections for ISD	Estrogens Conjugated, 0.3–0.625 mg orally daily (plus progestin if uterus is intact) Topical, 0.5–1.0 g/application Alpha-adrenergic agonists Pseudoephedrine, 30–60 mg three times daily	Oral estrogens increase the risk of endometrial cancer, elevated blood pressure, and venous thrombosis Alpha-adrenergic agonists are not FDA-approved treatment of stress incontinence; they may increase the risk of stroke, hypertension, headache, and tachycardia
Urge	Urinary urgency and frequency (small to moderate volumes); overactive bladder caused by detrusor instability	Bladder training (scheduled or prompted voiding), pelvic muscle rehabilitation (contraction and relaxation of pelvic muscles), biofeedback (reduction of intensity of uninhibited contractions)	Anticholinergics/bladder relaxants Oxybutynin, 2.5–5.0 mg three times a day (or a single dose at night for nocturia and urgency) Oxybutynin XL, 5–10 mg every day Tolterodine, 1–2 mg twice a day (or a single dose at night for nocturia and urgency) Tolterodine LA, 2–4 mg every day	Common side effects are dry mouth, constipation, and blurred vision; potential risk of increased intraocular pressure, delirium, and tachyarrhythmias
Overflow	Associated with either an acontractile bladder or bladder outlet obstruction (BPH, stricture, or bladder neck constriction); bladder distention with frequent or constant dribbling, with or without urgency	Intermittent or continuous catheter drainage for acontractile bladder Surgical relief of urethral obstruction (prostatectomy) Bladder retraining and scheduled toileting can be attempted after acute urinary retention	Cholinergic agonist Bethanechol, 10–30 mg three times daily (initiate with lower doses) $Alpha_1$-adrenergic antagonists Terazosin, 1–5 mg at night Doxazosin, 1–4 mg at night Tamsulosin, 0.4–0.8 mg at night	Cholinergic agonists should be used judiciously for acontractile bladder; bethanechol may exacerbate bradycardia, hypotension, bronchoconstriction, and peptic ulcer exacerbation $Alpha_1$-adrenergic antagonists should be used only for mild outlet obstruction caused by BPH; possible side effects are postural hypotension, fatigue, and sexual dysfunction
Functional	Associated with inability or unwillingness to perform toileting (physical disability, environmental barriers, or psychiatric illness)	*Behavioral:* toileting schedule and prompted voiding; assisted toileting, bedside commode or urinal; assistive devices to aid ambulation and toileting (canes, walkers, rails, grab bars); discontinue physical restraints; increase exercise *Medical:* absorbent pads and garments; indwelling catheters for severely debilitated patients	Change treatments (reduce diuretic and vasodilator dosages) Treat depression Treat Parkinson disease	Minimize use of drugs that worsen functional impairments (e.g., neuroleptics with extrapyramidal side effects and vasodilators that cause postural hypotension)

Note: Elderly patients frequently have mixed incontinence (urge and stress) and functional impairments that precipitate incontinence. Treatments should not be started until after a basic evaluation is completed to exclude reversible causes of acute incontinence and overflow incontinence. Many frail patients have detrusor hyperactivity with impaired contractility presenting with urinary urgency and small-volume voids.
BPH—benign prostatic hyperplasia LA—long acting XL—extended length

are warranted. Outpatient cystometry is able to determine bladder filling capacity, detrusor compliance and contractility, and PVR. A detailed urodynamic evaluation may be needed to determine the cause and most appropriate treatment of incontinence, particularly before operative treatments of incontinence. For example, stress incontinence is diagnosed and categorized according to abdominal-leak point pressures.[25] Urodynamic studies are essential for making a definitive diagnosis of urinary obstruction or detrusor hyperactivity with impaired contractility and for determining the type of stress incontinence.

MANAGEMENT

Strategies for the management of urinary incontinence include behavioral modification techniques, medications, patient and caregiver education, surgical procedures, catheter placement, and incontinence supplies [*see Table 4*].[21] The acute onset of incontinence should be evaluated and treated promptly. UTI, acute urinary retention, stool impaction, and adverse effects of medications (e.g., diuretics) should be excluded. After the initial diagnostic evaluation, most patients should be treated on the basis of the most likely type of incontinence. This empirical ap-

proach will lead to successful management of a large percentage of incontinent patients.[26]

Stress Incontinence

In female outpatients, behavioral interventions (e.g., anorectal biofeedback, which helps patients identify pelvic floor muscles so as to relax them selectively), bladder training, and pelvic muscle exercises are effective first-line therapies for established stress incontinence.[27-29] Daily repeated pelvic floor exercises will strengthen the voluntary periurethral and perivaginal muscles, augment urethral pressure, and inhibit urinary leakage. Pelvic muscle exercise is assumed to enhance urethral resistance by increasing the strength and endurance of the periurethral and perivaginal muscles and by improving the anatomic support to the bladder neck and proximal urethra.

Pelvic muscle exercises entail tensing the perivaginal muscles and anal sphincter as if to control urination or defecation while applying minimal contraction of abdominal, buttock, or inner thigh muscles. Typically, pelvic floor exercises are conducted for 10-second intervals several times a day for at least 8 weeks—most often, indefinitely. Another technique for exercising pelvic muscles is to interrupt midstream voiding by contracting the urinary sphincter while keeping the other muscles relaxed. The effectiveness of these exercises depends on patient compliance; benefits may not be seen for several weeks. Vaginal cones and electrical stimulation are sometimes used as adjuncts to exercises, with less certain results. For mild cases of stress incontinence, elevation of the urethra with a foam pessary often suffices.

Medications play a modest role in the treatment of stress incontinence. Middle-aged women with stress incontinence and hypoestrogenism may benefit from intravaginal, oral, or transdermal estrogen replacement therapy.[19] However, current concerns regarding the long-term risks of estrogen replacement therapy on cardiovascular disease, cancer, and venous thromboembolism limit the role of these treatments in elderly women.[30] Alpha-adrenergic agonists (e.g., pseudoephedrine, 15 to 30 mg) are thought to reduce the frequency of stress incontinence by increasing internal sphincter tone and bladder outflow resistance. These agents have potential side effects, including exacerbation of hypertension, and should be used with caution in elderly patients.

Surgical options for treatment of urethral hypermobility include bladder neck resuspension (colposuspension) or vaginal slings. The long-term effectiveness of percutaneous needle operations has been disappointing and suggests that conservative approaches should be fully explored before resorting to major operations.[31] Vaginal slings or periurethral injections of collagen are the treatments of choice for patients with intrinsic sphincter deficiency. In one study, collagen-injection therapy for intrinsic sphincter deficiency in women showed excellent 1-year responses. Of 94 patients, 67% achieved continence, 38.3% became dry, and 28.7% became socially continent.[25] Long-term studies find collagen to be a safe and moderately effective therapy; however, many women experience recurrent leakage that may not resolve with reinjection.[32]

In elderly men, stress incontinence is usually caused by intrinsic sphincter deficiency resulting from trauma to the bladder outlet, most commonly secondary to prostatectomy. Stress incontinence caused by intrinsic sphincter deficiency may be treated with either periurethral injections or the placement of an artificial sphincter.[19] Success rates with collagen are less impressive in men than in women. However, very high continence rates have been reported in men in whom artificial sphincters (especially models having a double cuff) have been implanted for treatment of postprostatectomy incontinence.[33]

Urge Incontinence

Elderly women with detrusor instability with urge incontinence often respond to behavioral therapies, such as bladder training (i.e., a combination of patient education, scheduled voiding, urge-suppression techniques, and pelvic muscle exercises)[34]; biofeedback has been shown to work well in younger and cognitively intact women.[27] Bladder training is assumed to improve cortical inhibition over lower urinary tract functioning and has been used primarily in the treatment of urge incontinence. Cognitively impaired patients also benefit from bladder training with scheduled toileting (e.g., beginning every 2 hours and progressing to every 3 to 4 hours); prompted voiding in which the caregiver toilets the patient may also be useful. Bladder-relaxant medications that also have anticholinergic properties are the most effective drug therapies for urge incontinence. Most commonly, either oxybutynin or tolterodine is used for treating detrusor overactivity.[19] Oxybutynin is short acting and often produces significant anticholinergic side effects of constipation, dry mouth, and exacerbation of glaucoma. A long-acting preparation of oxybutynin (Ditropan XL) may enhance compliance and reduce the incidence of anticholinergic side effects.[35] Tolterodine is available in both short-acting and long-acting forms.[36] Direct comparison of the extended-release preparations of oxybutynin and tolterodine in women with an overactive bladder showed similar reductions in incontinence episodes, with only minor differences in side effects.[37] Because of anticholinergic side effects, these agents need to be used cautiously in cognitively impaired patients, especially patients taking cholinesterase inhibitors for treatment of Alzheimer disease.

For the frail elderly, behavioral and environmental interventions are most effective in the treatment of urge incontinence. Strategies that maintain or improve mobility may prevent incontinent episodes.[19] Devices, including urinals, bedside commodes, or other external collecting devices, may help these patients to achieve continence. Other helpful aids are canes, walkers, or wheelchairs for patients with impaired ambulation; elevated toilet chairs; and the avoidance of physical or chemical restraints that impede the patient's toileting ability. For patients with intractable incontinence, absorbent undergarments or adult diapers are frequently used.

Combined Stress and Urge Incontinence

Because mixed forms of urinary incontinence are common in women, patients may benefit from both behavioral therapies and medications. One randomized clinical trial contrasted a biofeedback-based behavioral approach with both drug treatment (oxybutynin chloride, 2.5 to 5.0 mg three times daily) and a placebo.[28] Women 55 years of age or older with urge or mixed urge-stress incontinence were assigned to behavioral treatment consisting of anorectal biofeedback and pelvic muscle exercises performed at home. Episodes of incontinence decreased nearly 81% on average with behavioral treatment, which was significantly better than the decreases that occurred with drug treatment (68.5%) or placebo (39%). A toileting schedule may also be effective in combination with behavioral and pharmacologic approaches. Patients should be encouraged to urinate at regular intervals and before physical exertion. As in treatment of stress incontinence, vaginal cones, electronic devices, pessaries, or ele-

vating devices (e.g., surgical sling procedures) can be useful for individual patients with mixed incontinence, although their long-term effectiveness has not been established.[38]

Overflow Incontinence

Overflow incontinence occurs when there is either bladder outlet obstruction or an underactive (acontractile) bladder. Acute urinary retention is a common cause of overflow incontinence in hospitalized elderly patients. It is often precipitated by medications (e.g., anticholinergic drugs), anesthesia, or urethral manipulation, and it may be treated with intermittent urethral catheterization until the acute precipitating event subsides. The most common cause of bladder outlet obstruction in men is prostatic hyperplasia; less common causes are carcinoma and urethral stricture. Patients with incontinence resulting from bladder outlet obstruction will require either surgical correction or intermittent catheterization. Occasionally, men with prostatic hyperplasia respond to alpha-adrenergic antagonists, which may reduce internal sphincter tone. In women, overflow incontinence occurs as a complication of anti-incontinence operations or severe pelvic organ prolapse. An underactive or acontractile bladder may occur secondary to neurologic diseases such as diabetic neuropathy, spinal cord injury, and idiopathic detrusor underactivity.

Intermittent self-catheterization is ideal in patients who have atonic or neurogenic bladders. Long-term indwelling urinary catheterization is usually reserved for patients who cannot be catheterized intermittently because of discomfort or terminal illness. External (condom) catheters for males are used selectively because they often fail and can lead to local skin infection or UTI.

Functional Incontinence

Patients with neuropsychiatric diseases may experience functional incontinence in the absence of bladder disease. Immobile patients may suffer incontinence when they are unable to toilet because of physical illness, restraints, or environmental barriers. If urinary retention with overflow incontinence is excluded, patients with functional incontinence can be managed through exercise, toileting schedules, and assistive devices that enhance toileting ability.

Fecal Incontinence

Fecal incontinence is the continuous or recurrent uncontrolled passage of fecal material for at least 1 month. Acute and chronic fecal incontinence occur commonly in elderly patients with comorbid conditions and are often a socially embarrassing and incapacitating problem.[39]

EPIDEMIOLOGY

The prevalence of fecal incontinence increases with age, occurring in 3% to 7% of all persons 65 years of age and older,[39] in up to 50% of elderly patients in nursing homes,[40] and in 30% of elderly patients in hospitals.[41] Constipation (i.e., the passage of infrequent, hard, or difficult-to-pass stools) is often associated with incontinence, particularly in patients with fecal impactions.

RISK FACTORS AND ETIOLOGY

Fecal incontinence is most often a result of fecal impaction caused either by constipation or dysfunction of the internal or external anorectal sphincters. Factors that contribute to the risk of fecal incontinence include disordered muscle integrity, decreased rectal sensation or compliance, declining mental function, and loss of physical mobility.[39] Common causes of constipation include a diet low in fiber and fluids, dehydration, immobility, and medications.[41]

PATHOPHYSIOLOGY

The process of normal defecation requires the integrity of skeletal and striatal muscles involved in anal sphincter function, cognitive awareness and ability to get to the toilet, and normal function of the pelvic floor muscles and nerves.[41] A disruption of any of these links predisposes the patient to incontinence. Rectal trauma, pudendal nerve injury, autonomic neuropathies, rectal prolapse, hyperosmolar diets, and fecal impaction are common physiologic factors contributing to incontinence.[41] Fecal impaction occurs most often in the distal or rectosigmoid colon. Mucus and fluids are secreted proximal to the impaction and leak around the mass or are passed after therapeutic disimpaction. Medications, especially opiates and anticholinergic agents, are common causes of constipation, impaction, and incontinence. Acute fecal incontinence may be seen in diarrheal states, and intermittent incontinence is often seen in patients with dementia, delirium, pelvic floor denervation, or excessive laxative use.[42]

Table 5 Fecal Incontinence: Prevention and Management

	Intervention	*Example*
Prevention of Constipation	Physical activity	Aerobic exercises (e.g., walking, water aerobics)
	High-fiber (8–12 g daily) diet* (to induce bulking effects, alter microbial ecology) plus fluids	High-fiber vegetables, bran supplements, psyllium supplement, methylcellulose supplement
	Laxatives (when fiber is ineffective or not tolerated)	Lactulose or sorbitol preferred (osmolar agents); irritant laxatives (for patients intolerant of osmolar agents): bisacodyl, senna; enemas (for patients with colonic dysfunction)
	Avoid constipating medications or use judiciously	Anticholinergic agents; antispasmodics, antiparkinsonian drugs, tricyclic antidepressants, neuroleptics, iron supplements, opiates, calcium channel blockers (verapamil)
Management of Fecal Incontinence	Treatment of fecal impaction†	Enemas: saline, water, sodium phosphate, bisphosphonate; colonic irrigation; high-fiber diet (after disimpaction); oral hyperosmolar solutions
	Treatment of colitis (e.g., from radiation, inflammatory bowel disease)	Medical: soluble fiber supplements, opiates (loperamide)
	Modification of behavior	Biofeedback (motivated, cognitively intact patients), toileting schedule for physically or cognitively impaired patients
	Treatment of anatomic abnormalities	Surgical repair of rectal prolapse or anal sphincter

*Fiber should be avoided until fecal impaction resolves.
†Often, a combination of soluble fiber, hyperosmolar agents, and periodic enemas is needed to prevent impaction in bedridden patients and those with chronic colonic dysfunction (e.g., chronic laxative use, diverticular disease).

DIAGNOSTIC EVALUATION

Medical History and Physical Examination

The evaluation of fecal incontinence includes a careful review of the patient's cognitive status, anorectal and neurologic function, and a rectal examination. In hospitalized or institutionalized patients, the diagnosis of a fecal impaction is suggested by the passage of watery stools laden with mucus. This suspicion is confirmed by rectal examination, which generally reveals firm or hard stool in the ampulla, often associated with a patulous rectum. With high impactions, feces may be palpable during abdominal examination or confirmed by x-ray of the abdomen. Mental status examination identifies the patient with dementia or delirium who has lost self-toileting capacity. The absence of anal sphincter tone or anal wink may suggest denervation of the pudendal nerve (S2-4), resulting from a local or spinal cord lesion. The rectal and abdominal examinations also help identify inflammatory diseases or tumors as causes of incontinence.

Laboratory Tests

Diagnostic studies are needed when either the diagnosis or the appropriate management of the incontinent patient remains uncertain. Useful studies include anoscopy or flexible sigmoidoscopy for confirmation of masses or inflammation. Infrequently, anorectal manometry to measure intraluminal pressure, anal endosonography to identify mass lesions, pudendal nerve conduction measurement to diagnose neuropathy, or defecography to define intrinsic lesions are needed to ascertain the cause of fecal incontinence.[39]

PREVENTION AND MANAGEMENT

The prevention of fecal incontinence begins with an assessment of risk factors for fecal impaction and functional incontinence. The most common approaches to the prevention of fecal incontinence and constipation include changes in diet, increased physical activity, the judicious use of laxatives and enemas, and surgical correction in patients with anatomic abnormalities (e.g., rectocele) [*see Table 5*]. Biofeedback can be effective in highly motivated and cognitively normal patients with sphincteric dysfunction.

Laxatives

Laxatives are prescribed for patients with incontinence resulting from constipation. Stimulant laxatives (e.g., senna and bisacodyl), hyperosmolar laxatives (e.g., sorbitol and lactulose), rectal suppositories (e.g., glycerin and bisacodyl), or enemas (e.g., tap water) are often sufficient to treat constipation. Saline laxatives (e.g., milk of magnesia) can also be used, but hypermagnesemia limits their usefulness in patients with renal insufficiency.[39] Fecal disimpaction can be performed manually with an anesthetic lubricant or by mineral oil enemas. Mixed electrolyte solutions containing polyethylene glycol, a major hyperosmolar laxative,[39] are effective in relieving fecal impactions; the electrolyte solution is given orally, usually in large amounts (e.g., 2 to 3 L/day in divided doses). Sodium phosphate solution should be avoided in most elderly patients with fecal impaction because of its increased absorption and the risk of electrolyte disturbances associated with its use.

High-Fiber Diet

Once the impaction is resolved and colonic function is restored, patients should be placed on a high-fiber diet or given fiber supplements (e.g., psyllium or methylcellulose) along with liquids. If constipation is associated with hard stools, fiber and fluids should be gradually increased over several weeks to ensure a soft stool. The role of diet in the prevention of impaction is unclear; diet is probably not as important as colonic cleansing in preventing impaction.[39]

Suppositories and Enemas

The intermittent use of glycerin or bisacodyl suppositories is warranted if rectosigmoid outlet delay or difficult passage of a soft stool is the primary concern.[37] Patients unable to retain a suppository can be treated with periodic enemas or hyperosmolar solutions.

Falls and Gait Disturbances

Accidental falls, defined as unintentionally coming to rest on the ground, floor, or other lower level, are common and potentially preventable causes of morbidity and mortality in elderly adults. The risk factors for falls and the effectiveness of multifactorial interventions to prevent recurrent falls in carefully targeted patients are well established.[43] Falls often result from an interaction of conditions that involve host (intrinsic) or environmental (extrinsic) predisposing or situational risk factors [*see* Risk Factors, *below*].[44] Falls can be associated with syncopal or presyncopal episodes, resulting from cardiac arrhythmias, postural hypertension, or postprandial hypotension [*see 1:I Approach to the Cardiovascular Patient*].

EPIDEMIOLOGY

Each year, falls occur in about one third of community-dwelling persons older than 65 years and in about half of persons 80 years of age or older. About half of these individuals who fall experience multiple falls. Falls account for serious injuries that include hip fractures and soft tissue trauma and often lead to an older person's loss of functional independence and a fear of falling.[43] Unintentional injury, most often attributed to falls, is the seventh leading cause of death in elderly persons. About 5% of falls by community-dwelling elderly persons result in a fracture. Falling increases the probability of hospitalization, nursing home placement, and death. About half of falls by elderly patients result in soft tissue injuries such as bruises, lacerations, and abrasions. Less common complications are subdural hematomas and cervical fractures. Most hip fractures that occur in elderly people result from falls. More than half of the survivors of hip fracture are discharged to a nursing home, and half of those remain in a nursing home a year later. After hip fracture, fewer than 30% of patients regain their prefracture level of physical functioning.

RISK FACTORS

Intrinsic risk factors include lower extremity weakness, poor grip strength, gait and balance deficits, impaired performance of daily activities, visual impairment, cognitive impairment, and depression.[44] Extrinsic risk factors include use of four or more prescription drugs and environmental impediments such as poor lighting, loose carpets, and absence of bathroom-safety equipment.[44] Risk factors for hip fractures resulting from falls include low bone mineral density, use of long-acting benzodiazepines (e.g., diazepam), vision impairment, reduced mobility and physical independence, and cognitive dysfunction. About a third of fallers develop a fear of falling that is itself predictive of

an increased risk of balance and gait problems, a decline in self-care abilities, and an increased risk of falling.[45]

ETIOLOGY

Accidental falls stem from the combination of environmental hazards and the increased susceptibility to falls related to aging or diseases. Accidents, simple slips, and trips are the most common causes of falls occurring in the community-dwelling elderly population and are usually associated with environmental hazards. In nursing home patients, however, environmental factors are the most common immediate cause of falls, followed by weakness, gait or balance disturbances, drop attacks, dizziness or vertigo, and confusion. Less common causes of falls are postural hypotension, syncope [*see 1:I Approach to the Cardiovascular Patient*], acute illness (e.g., infection), and medications. Gait impairments with falls are associated with lower extremity weakness from deconditioning, stroke, cardiovascular disease (e.g., arrhythmias), and neurologic disease (e.g., Parkinson disease). Dizziness, vertigo, delirium, postural hypotension, visual disorders, alcohol use, and medications (e.g., psychotropic agents) are other causes of falls.[43] Medications associated with falls most notably include those that cause postural hypotension, such as loop diuretics, vasodilators, or adrenergic antagonists, and those with psychotropic properties, such as antidepressants and sedative-hypnotic agents.

PATHOPHYSIOLOGY

The maintenance of normal balance and gait requires the successful integration of sensory (afferent), central nervous (brain and spinal cord), and musculoskeletal systems. A disturbance in sensory input (e.g., peripheral neuropathy), central nervous system functioning (e.g., dementia), or motor function (e.g., arthritis or muscle weakness) will predispose elderly patients to falls. The aging process may also predispose patients to falls by increasing postural sway and reducing adaptive reflexes. Postural-stability limits appear to decrease with aging. However, changes in gait, such as slowing of walking speed, reduced stride length, and prolonged double support, which are often attributed to the aging process, likely represent adaptations to the fear of falling.

DIAGNOSTIC EVALUATION

Patients at risk for falls can be identified through a medical history, physical examination, and a few laboratory studies. A review of risk factors, medications (e.g., vasodilators, adrenergic blockers, and psychotropic agents), and screening instruments (vision, mental status, balance, and gait) help identify patients at risk. Older persons should be asked at least once a year whether they experience falls.[42] For those persons reporting a fall, a review of the circumstances surrounding the fall, including symptoms before and after the event, provides clues to the likely causes. For example, vertigo may precede loss of balance and a fall; a

Table 6 Interventions to Reduce the Risk of Falls

Risk Factors	*Interventions*
Medications Use of ≥ 4 medications Alcohol reduction Use of any benzodiazepine or other sedative-hypnotic agent Vasodilators (arterial and venous)	Review medications, and reduce, taper, or discontinue use of alcohol and psychotropics (benzodiazepines and sedative-hypnotics); if necessary, replace with antidepressants (for insomnia, depression, or anxiety) Use nonpharmacologic therapy for sleep disorders (sleep restriction, no long daytime naps, aerobic exercise early in day, utilization of relaxation techniques such as music and massage) Use alternatives to vasodilators (beta blockers and calcium channel blockers)
Postural hypotension Drop in systolic blood pressure ≥ 20 mm Hg Postural dizziness or light-headedness	Change doses of medications (e.g., loop diuretics, vasodilators) If syncopal or presyncopal, consider workup for autonomic nervous system or hemodynamic causes If venous pooling occurs in legs, try leg pumps, salt repletion, or graded compression stockings
Lower-extremity weakness Generalized decreased strength of quadriceps, knee extensors, or flexors Deconditioning resulting from recent illness, hospitalization, or immobility	Low-intensity resistive exercises (bands, tubes, pulleys, or weights) under therapist supervision Endurance exercises (walking, biking, or water exercise)
Balance/gait impairment Postural instability Inability to transfer safely to bathtub or toilet Antalgic gait Fear of falling	Physical-therapy consultation for gait assessment Prescription of assistive devices (canes, walkers) Training in transfer skills Gait training and balance exercises (e.g., Tai chi) Analgesics for pain relief Environmental alterations (raised toilet seats, commodes, grab bars, and handrails)
Sensory impairments Hearing Vision	Hearing aids (headset microphones, reduced background noise) Visual aids (corrective lenses, improved illumination of room)
Environmental hazards Uneven surfaces Frayed rugs Loose cords Poor lighting Uneven steps or stairs	Occupational-therapy consultation (home visit, safety evaluation) Diffuse illumination of rooms Install handrails and grab bars Resurface slippery floors Install ramps Rearrange furniture Remove frayed rugs and cords

loss of consciousness preceding the fall suggests a diagnosis of syncope.

Observation of the patient's balance and gait is the most useful aspect of the examination. Performance-based measures have been validated and are predictive of gait and balance impairments. The timed get-up-and-go test is particularly useful, is quickly performed, and appears to be predictive of falling. The test requires a patient to stand up, walk 10 ft, turn, walk back, and sit down. Older adults at risk for falls require more than 20 seconds to complete this task. During the test, postural instability, lower extremity weakness, reduced steppage, increased lateral sway, stride variability, and ataxia can be easily identified. Individuals who have difficulty with this task or demonstrate unsteadiness require further assessment.[45] In the outpatient setting, patients who have had two or more falls or who experience balance or gait difficulties should have a detailed assessment of the factors that predispose them to falls [*see Table 6*].[44]

The further diagnostic evaluation is based on the circumstances surrounding the fall and a judgment about the most likely causes. An examination of the patient's vision, gait and balance, lower extremity strength and function, and mental status, as well as a basic neurologic examination, is recommended.[45] An extensive diagnostic workup is not usually warranted in nonsyncopal falls. However, in the emergency evaluation of an elderly faller, the history of head trauma and the finding of focal neurologic deficits suggest the need for a neuroimaging procedure (e.g., CT scan of the head).

PREVENTION AND MANAGEMENT

Multicomponent Interventions

The most consistently effective interventions to prevent recurrent falls have been targeted at both intrinsic and extrinsic risk factors [*see Table 6*].[44] Successful components of interventions used in clinical trials include review and alterations of medications, balance and gait training, muscle-strengthening exercises, improvement of postural hypotension, home-hazard modifications, and specific medical and cardiovascular treatments.[44] In a clinical trial involving community-residing persons 70 years of age or older with risk factors for falling, those who received interventions that included an adjustment of medications, behavioral instructions, and an exercise program (e.g., balance exercises, gait training, and low-intensity resistive exercises) had fewer falls in the subsequent year than did control subjects.[46] Meta-analyses support the effectiveness of multicomponent interventions to prevent falls.[47]

Balance and Strengthening Exercises

Tai Chi exercise Tai Chi exercises to enhance balance and body awareness when combined with balance training may also reduce the rate of falls.[48] A randomized trial of Tai Chi exercise for 15 weeks in 200 persons 70 years of age and older resulted in a 47% decrease in falls after a 4-month follow-up. Fear of falling is reduced as well, which may account for the benefits of exercise rather than improved postural balance.[6,49]

Therapist-conducted exercise In one study, home visits of women 80 years of age and older made by a physiotherapist led to a 41% reduction in self-reported falls in 1 year, a decreased risk of fall with injury, and improved balance.[50,51] The therapist conducted an exercise program designed to increase balance and muscle strength. Benefits persisted for 2 years in compliant patients.

Resistive exercise Clinical trials demonstrate the effectiveness of both low-intensity and progressive high-intensity resistive exercise in improving lower-extremity strength.[44,47] Bands, tubes, pulleys, and weight machines have been used under therapist supervision in various studies.[52]

Prevention of Hip Fractures Caused by Falls

Hip protectors have been found to reduce the risk of hip fracture in nursing home patients.[53] Studies of community-residing older people have produced mixed results.[54,55] In part, this reflects the low level of adherence to the use of hip protectors, as well as differences in the efficacy of available brands.

Recommended Treatment Plan

Clinicians who are evaluating an elderly patient with a history of falls should begin with a discussion of the circumstances surrounding the falls and determine if the falls were syncopal or nonsyncopal. If the falls were nonsyncopal, the next step is to review the patient's medications, perform a test of mobility such as the get-up-and-go test, and evaluate the patient's muscles and joints (especially knees, hips, and ankles) for range of motion, strength, and stability. If the causes of falls remain unclear and the falls occur primarily at home, the clinician should consider a home-safety evaluation by a physical therapist. For patients with lower extremity weakness, a referral of the patient to a physical therapist for gait assessment and muscle strengthening is advisable. The referral should request a gait assessment and low-intensity resistive exercises of the lower extremities, including hip and knee extensors. Clinicians should also treat comorbid conditions that increase the risk of injurious falls, notably osteoporosis, drug intoxication, and being underweight.

Immobility

Prolonged bed rest produces many physiologic changes, including decreases in blood volume and cardiac output, orthostatic hypotension, hypoxemia, muscle atrophy and generalized weakness, and decreased muscle oxidative capacity. In hospitalized elderly patients, immobility increases the risks of functional dependency; nursing home placement after discharge; and medical complications, including deep vein thrombosis, urinary incontinence, pressure sores, joint contractures, cardiac deconditioning and muscle weakness, and falls. Immobile patients in hospitals or nursing homes are at risk for pressure ulcers.

Pressure Ulcers

A pressure ulcer occurs when soft tissue is compressed between a bony prominence and an external surface for a sustained period. Pressure ulcers result in substantial morbidity, increased health care costs, and reduced quality of life for older patients.[56] Pressure ulcers are common in immobilized patients and increase in incidence with aging. Early-stage ulcers should be identified and managed aggressively to prevent their progression to more severe stages. The National Pressure Ulcer Advisory Panel[57] has defined four stages of pressure ulcers:

Stage I: nonblanchable (persistent) erythema of intact skin.
Stage II: partial-thickness skin loss involving the epidermis, the dermis, or both.

Stage III: extension into subcutaneous tissues to the deep fascia with or without undermining.
Stage IV: extension into muscle, bone, or both.

EPIDEMIOLOGY

The prevalence of stages II, III, and IV ulcers in patients in acute care hospitals ranges from 3% to 11%, with an incidence during hospitalization of 1% to 3%. The rate is higher in bedridden patients. Stage I and stage II pressure ulcers are the most common. The prevalence of pressure ulcers in nursing home residents in one study was 11.3% for stage II, III, or IV pressure ulcers.[55] For residents admitted to the nursing home without pressure ulcers, the incidence was 13.2% at 1 year and increased to 21.6% after 2 years of nursing home stay.[58] About 80% of pressure ulcers in nursing home patients develop over the sacrum or coccyx, hips (femoral trochanter), ischia, and heels.[59]

RISK FACTORS AND ETIOLOGY

Most prospective studies of stage II, III, or IV pressure ulcers in hospitalized patients have implicated impaired mobility, incontinence, undernutrition, and impaired consciousness as significant risk factors for pressure ulcers.[59] In hospitalized patients who are bedridden or chairbound, risk factors include hypoalbuminemia, fecal incontinence, and fractures.[59] A prospective study of hospitalized patients with activity limitation (e.g., bedridden or hip fracture) identified the following five independent predictors of incident pressure ulcers in patients 55 years of age or older: nonblanchable erythema (stage I ulcer), lymphopenia, immobility, dry skin, and decreased body weight.[59] In general, impaired mobility is the most important risk factor for pressure ulcers.

PATHOPHYSIOLOGY

Four factors have been implicated in the pathogenesis of pressure ulcers: pressure, shearing forces, friction, and moisture. Sustained pressure over bony prominences (e.g., for more than 2 hours) results in ischemic damage to muscle and subcutaneous tissues. Shearing forces are tangential forces that lower the amount of pressure required to cause damage to the epidermis. Shearing forces occur, for example, when a patient is shifted in bed and dragged across underlying sheets. Repeated exposures to pressure will cause skin necrosis at lower pressures. The loss of subcutaneous tissue also lowers the threshold for skin breakdown caused by pressure. Friction is the force that results in skin abrasion when a patient is dragged across bedsheets, for instance. Moisture secondary to incontinence or perspiration can result in skin maceration and predispose the patient to pressure ulcers.[59] Any disease process leading to immobility and limited activity levels increases the risk of pressure ulcers. The aging skin predisposes patients to pressure ulcers: it is more susceptible to shearing forces, has decreased vascularity, and, in malnourished patients, has decreased subcutaneous fat. Furthermore, patients with cognitive impairment, depression, or spinal cord injury are likely to be immobile or unable to report symptoms of pain and discomfort, thereby delaying detection of an ulcer.

DIAGNOSTIC EVALUATION

Although stage I ulcers are not true ulcerations, their identification and early treatment are critical to prevent progression to stage II, III, or IV. Full-thickness injury is often manifested by eschar, which must be removed before staging can be completed. The most common sites for pressure ulcers in elderly patients include the scapula, iliac crest, sacrum, ischium, trochanter, lateral malleolus, heel, and lateral edge of the foot.[59] Pressure ulcers often present as a skin blister, which evolves to frank ulceration with exudate or plaque eschar over several days. The ulcer represents the tip of the iceberg as the ischemic injury extends in a widening triangular fashion down to subcutaneous tissue. Clinically, the lesion is identified as a stage III ulcer when the injury extends beneath the dermis to involve muscle fascial structures. In stage IV ulcers, undermining and sinus tracts may be present. Sinus tracts in stage IV lesions can be identified by probing the ulcer margin. Pressure ulcers also can become colonized by bacteria, resulting in cellulitis or bacteremia, especially in debilitated or immunocompromised patients.

PREVENTION AND MANAGEMENT

Preventive strategies for pressure ulcers include recognizing risk factors, decreasing the effects of pressure, assessing nutritional status, avoiding excessive bed rest, and preserving the integrity of the skin.[60] Patients at risk for an ulcer can be identified with the use of the Norton or Braden scales. These scales relate the risk of pressure ulcer to impaired sensory perception, increased skin moisture, decreased physical activity, immobility, poor nutrition, and friction and shearing force.[56]

Treatment principles include assessing the severity of the wound; reducing pressure, friction, and shearing forces; optimizing wound care; removing necrotic debris; managing bacterial contamination; and correcting nutritional deficits.[60] For all stages of pressure ulcers, the first step is to reduce pressure. The principles of stage I ulcer management apply to all patients with stage II, III, or IV ulcers [*see Table 7*].[61] For stage II, III, or IV ulcers, necrotic tissue is eliminated by either medical or surgical debridement.[56] Necrotic tissue is a barrier to epithelialization and serves as a nidus for infection.

The prevention and treatment of pressure ulcers in hospitalized and nursing home patients is a measure of quality of care. A multicenter study of hospitals in the United States identified six processes of care for prevention of pressure ulcers in at-risk patients: use of daily skin assessment; use of a pressure-reducing device; documentation of being at risk; repositioning for a minimum of 2 hours; nutritional consultation initiated for patients with nutritional risk factors; and staging of pressure ulcer.[62]

Wound Care

Wound management strategies such as use of wound dressings, debridement techniques, physical therapies, antibiotics, and antiseptics are used in the management of pressure ulcers. Occlusive dressings such as transparent films in hydrocolloid dressings or topical collagen improve healing of stage II pressure ulcers. The dressings remain in place for several days and facilitate epithelial migration.[58] Topical hydrocolloid and collagen treatments were found to have similar healing outcomes in a study of nursing home patients with stage II or III ulcers, in whom healing occurred in most cases within 8 weeks; however, hydrocolloid dressing has the advantage of being less expensive and is changed weekly rather than daily.[63]

A 2-week trial of a topical antibiotic such as silver sulfadiazine can be considered for clean pressure ulcers that are not healing or are continuing to produce exudate after 2 to 4 weeks of optimal management. Povidone-iodine should not be used, because it is toxic to fibroblasts. When necrotic tissue is present, debridement should be performed. Debridement can be accomplished

Table 7 Stages and Usual Treatments for Pressure Ulcers

Stage	*Presentation*	*Usual Treatment*
I	Nonblanchable erythema of intact skin; the heralding lesion of skin ulceration	Reduce pressure over ulcer and bony prominence; bedridden patients with sacral, ischial, or back ulcers are repositioned (e.g., side-to-side at 30° angle) at least every 2 hr. For chair-bound patients, use pressure-reducing device (e.g., foam, gel, or air), reposition hourly; treat dry skin with moisturizers (e.g., creams, lotions, ointments, or lubricants); protect skin from moisture; maintain proper positioning, transferring, and turning techniques; provide nutritional support; apply semipermeable polyurethane film (change weekly)
II	Partial-thickness skin loss involving epidermis, dermis, or both; the ulcer is superficial and presents clinically as an abrasion, blister, or shallow crater	Debride devitalized tissue. Medical debridement: cleanse wound with saline; use wet-to-dry saline dressing, thin-film polymer dressing, or hydrocolloid dressing (wet-to-dry saline dressings when exudate is present, wet-to-moist when ulcer base is free of exudate and eschar; change hydrocolloid dressing every 4–7 days [reduces caregiver time]); schedule exercise for patients able to walk Surgical debridement: use scalpel if eschar or advancing cellulitis is evident; use thick foam mattress or air mattress as a support surface
III	Full-thickness skin loss involving damage or necrosis of subcutaneous tissue, which may extend down to but not through underlying fascia; the ulcer presents clinically as a deep crater with or without undermining of adjacent tissue	Intervention for stage II ulcers Medical: use hydrogel or alginate, moist gauze packs (saline), enzymatic debridement (noninfected ulcers), or topical antibiotic (if exudate or nonhealing persists after ≥ 2 wk of optimal care) for 2 wk (gram-negative, gram-positive, anaerobic coverage) Surgical: debride large eschar and devitalized tissue; use air-fluidized bed or low-air-loss bed (for deep, large, or multiple ulcers) as a support surface
IV	Full-thickness skin loss with extensive destruction, tissue necrosis, or damage to muscle, bone, or supporting structures (e.g., tendon or joint capsule)	Interventions for stage II–III ulcers Medical: systemic antibiotics for bacteremia, sepsis, advancing cellulitis, or osteomyelitis; suspect osteomyelitis in patients with deep ulcers that fail to heal with appropriate therapy and in patients with elevated temperature and white blood cell count or abnormal bone scan Surgical: create myocutaneous flap (after debridement) of large (wide) ulcers; use split-thickness skin grafts

with a sharp blade; by mechanical approaches such as wet-to-dry dressings, hydrotherapy, irrigation, or dextranomers; by enzymatic approaches (collagenase); or by autolytic techniques (synthetic dressing cover).[56,61] Once an ulcer is clean and epithelialization occurs, a moist wound environment should be maintained; wet-to-dry dressings are avoided. Hydrocolloid dressings or gels are continued until healthy skin appears.

Most stage I and stage II pressure ulcers heal within 60 days with the usual optimal therapies. The choice of dressings and debridement techniques has little impact on this good prognosis. In general, a moist saline gauze wound dressing is as effective as the application of hydrogel dressings. The choice of treatment for these ulcers is based on considerations of cost, convenience, and caregiver preferences.

Stage III and IV ulcers are often slow to heal or refractory to usual medical therapies. A promising technique for curing these ulcers is the sequential use of calcium alginate dressings and hydrocolloid dressings. The sequential approach has been shown to promote faster healing of ulcers than treatment with hydrocolloid dressings alone.[64] Another promising therapy is the use of topical treatment of pressure ulcers with nerve growth factor. In a 6-week clinical trial, compared to standard care, daily application of a solution of nerve growth factor on foot ulcers produced acceleration of ulcer healing.[65]

Pressure Relief

Pressure relief can contribute to prevention of pressure ulcers by two principal methods: (1) use of a conforming support surface to distribute the body weight over a large area and (2) use of an intermittent support surface in which inflatable cells are alternately inflated and deflated. Treatment of pressure ulcers also requires reduction of pressure to the damaged skin. Special support surfaces aimed at pressure redistribution (e.g., beds, mattresses, and cushions) are widely used.[66] Air or foam products may be helpful, but some patients require the use of specialized beds, such as an air-fluidized bed or a low-air-loss bed. Air-fluidized beds contain microspheres of ceramic glass, and warm pressurized air is forced up through the beads, causing them to take on the characteristics of a fluid. Patients float on the beads, with pressure reduced under prominences. Low-air-loss beds consist of large fabric cushions that are constantly inflated with air. The cushions of low-air-loss beds are fitted on a regular hospital bed frame, which is more practical for patient transfers. The use of these beds should be considered when patients have large, multiple, or full-thickness (stage III or IV) pressure ulcers; when an individual has fewer than two turning surfaces free of pressure ulcers; or when a patient has experienced recurrent ulceration and an inability to heal on a static pressure-reducing device.[66] In a clinical trial conducted in nursing homes, low-air-loss beds provided significantly better healing rates of stage II or greater pressure ulcers than foam mattresses.[67]

Malnutrition

EPIDEMIOLOGY

In the United States, protein-calorie malnutrition, also known as protein-energy undernutrition, is uncommon among community-residing elderly persons. In community surveys, however, the prevalence of nutrient-specific malnutrition is greater in people older than 65 years.[68] In the National Health and Nutrition Examination Survey III, median energy intake for elderly adults was below the recommended levels, with higher intakes for whites than for African Americans. Minorities tended to have

lower median energy, dietary, and mineral intakes than whites.[65] A survey of patients 65 years of age and older who were admitted to general medicine, orthopedic, general surgery, and neuroscience services revealed that 41% were well nourished, 44% had moderate risk for malnutrition, and 15% were malnourished.[69]

In long-term care, malnutrition has been identified in 50% of residents and is an independent predictor of subsequent mortality. Poor nutritional status and protein-calorie malnutrition have been associated with altered immunity, impaired wound healing, reduced functional status, increased health care utilization, and increased mortality. Weight loss of 5% or more of usual body weight is associated with increased morbidity and mortality.

RISK FACTORS AND ETIOLOGY

Nutritional disorders in older patients may be caused by numerous intrinsic and extrinsic factors. Many acute and chronic diseases predispose older patients to malnutrition (e.g., hyperthyroidism, diabetes, cancer, diseases of malabsorption, cardiovascular disease, renal disease, chronic obstructive pulmonary disease, and conditions associated with increased tumor necrosis factor or cachectin). Dementia is a common cause of undernutrition, as are adverse effects of medications (e.g., digoxin, some antidepressants). Furthermore, dysgeusia can be caused by medications (e.g., antihistamines, angiotensin-converting enzyme inhibitors, levodopa, lithium), zinc deficiency, and changes in sense of smell and taste resulting from physiologic aging. Inflammatory bowel disease, pancreatic insufficiency, and gluten enteropathy can cause malabsorption.[67] Other factors contributing to malnutrition include poor oral dentition, ill-fitting dentures, dry mouth, functional disability, and impaired vision. Less common causes are endocrine disorders (e.g., thyrotoxicosis or uncontrolled diabetes mellitus).

Studies of involuntary weight loss suggest that depression, cognitive impairment (dementia), gastrointestinal disorders (peptic ulcer or motility disorders), and cancer are the most common causes. Often, inflammation caused by infections and catabolic states contributes to the weight loss, hypoalbuminemia, and other serum markers of chronic disease.[70]

Psychosocial factors contribute to the risk of inadequate nutrition in older adults. Many older patients live on fixed incomes, have reduced access to food (social isolation), have poor knowledge of nutrition, or are dependent on others (caretakers or institutions) for food preparation. They may also suffer from depression, bereavement, dementia, or alcohol use. Other contributing factors that affect shopping for food and selecting or preparing food are impaired strength and impaired mobility.

PATHOGENESIS

During middle age, there is an increase in body mass and percentage of body fat in both men and women. At 70 years of age and beyond, declines in both lean body mass (sarcopenia) and body fat occur. Normal aging is associated with many physiologic changes that have a great impact on nutrition, notably in the gastrointestinal, sensory, hormonal, renal, and musculoskeletal systems. The complex interaction between these physiologic changes, chronic illness, and psychosocial and environmental factors determines the overall nutritional health of the elderly adult.[70]

Sarcopenia

The relevant changes associated with usual aging are a decrease in lean body mass (both skeletal and visceral), bone density, and total body water and an increase in total fat. From 25 to 75 years of age, lean body mass declines 19% in males and 12% in females, probably because of both inactivity and decreases in protein synthesis. Sarcopenia, or skeletal muscle wasting, results from a constellation of factors, including biologic changes associated with aging, inactivity, protein-calorie malnutrition, and catabolic diseases (e.g., congestive heart failure, chronic obstructive pulmonary disease, cancer, and hyperthyroidism). In both frail elderly and healthy elderly adults, type II, or fast-twitch, fibers are the only type of muscle fiber to atrophy. This selective atrophy of the type II fibers suggests that sarcopenia is caused by disuse, implying the potential for prevention. Sarcopenia, which is the result of a physiologic process, has to be differentiated from cachexia, a loss of both fat and muscle associated with chronic diseases such as malignancy.[70]

Anorexia of Aging

Anorexia of aging is a physiologic concept that explains the decrease in nutritional intake in old age; it can be regarded as a syndrome of inadequate nutritional consumption.[70] Anorexia of aging is caused by a combination of factors that interfere with the normal regulation of appetite and food consumption. Decreased olfactory discrimination, altered taste perception, and xerostomia diminish the flavor of food, which discourages the consumption of an adequate total caloric intake. Abnormal gastric emptying, decreased adaptive relaxation, and decreased gastric acid secretion probably caused by atrophic gastritis can interfere with absorption of vitamin B_{12} and iron, resulting in micronutrient deficiencies.

DIAGNOSTIC EVALUATION

Clinical Features

Major indicators of poor nutritional status include significant weight loss over time (10% or more of body weight in 6 months or involuntary weight loss), significantly low weight for height (e.g., 20% below desirable weight, or a body mass index of less than 19 kg/m^2), serum albumin of less than 3.5 g/dl, change to dependence in two activities of daily living (e.g., bathing and dressing), sustained inappropriate food intake (e.g., excessive alcohol use or dietary imbalance), reduction in midarm circumference to less than 10th percentile of ideal circumference, decrease in triceps skinfold thickness to less than 10th percentile of ideal thickness, and presence of nutrition-related disorders (e.g., osteoporosis, vitamin B_{12} deficiency, or folate deficiency).

Laboratory Tests

The serum albumin level is generally the most reliable, although nonspecific, indicator of chronic malnutrition. An albumin level below 3.5 g/dl, an unexplained normocytic anemia, and a very low serum cholesterol level (< 160 mg/dl) are compatible with a diagnosis of protein-calorie malnutrition. Reduced levels of serum transferrin and prealbumin and a low total lymphocyte count also suggest protein-calorie malnutrition.

PREVENTION AND MANAGEMENT

For healthy older people, a well-balanced diet is recommended and includes adequate amounts of calories, protein and essential amino acids, essential fatty acids, fiber, and complex carbohydrates and sufficient amounts of minerals and vitamins. The food-guide pyramid may be used as a general guideline for counseling patients about a balanced diet (http://www.nal.

usda.gov/fnic). Patients who are chronically ill, who have recent weight loss resulting from illness or surgery, or who have an unbalanced diet should be advised to take a multivitamin and mineral supplement daily. Calcium supplementation to ensure a daily consumption of 1 g or more of elemental calcium, along with vitamin D supplementation, is advisable. When treating chronically ill patients, education of the patient and family caregivers is an important step in preventing malnutrition. Patients should take advantage of nutritional programs available in the community, such as Meals on Wheels (http://www.projectmeal.org) and Title III nutrition services. A referral to the local office on aging or the Area Agency on Aging (http://www.aoa.dhhs.gov) will enable the patient to access these services.

Patients with social, physical, and psychological risk factors for malnutrition can be readily identified using screening instruments and a targeted physical examination.[70] Subjective global assessment (SGA), a validated measure of nutritional status based on medical history and physical examination findings, accurately classifies patients as severely malnourished, moderately nourished, or well nourished. SGA combines elements of the patient's nutritional history (weight loss in previous 6 months) and physical examination (e.g., muscle wasting) to generate a valid and subjective impression of nutritional status (http://www.eneph.com/feature_archive/nutrition/v25n4p190.html).[70,71] When combined with a review of biochemical indicators of nutritional status, this information should drive the process of further evaluation and treatment of protein-calorie malnutrition.

Balanced Diet and Nutritional Supplements

A well-balanced diet with calorie-dense foods should be prescribed, along with vitamin and mineral supplements. Commercially available canned nutritional supplements containing 250 to 350 calories of a nutritionally balanced formula are recommended as between-meal supplements (e.g., midmorning and midafternoon) for undernourished older patients. In hospitalized or community-dwelling elderly patients, oral protein and energy supplements reduce all-cause mortality.[72] Available interventions include frequent meals and snacks, enhanced flavors of favorite foods, protein-calorie supplements, multivitamins, appetite stimulants, and enteral and parenteral nutrition [*see Table 8*]. In a randomized, controlled trial of 88 nursing-home patients, an oral supplement was well accepted and resulted in increased daily protein and energy intake, body weight, and nutritional status in most malnourished patients and in those at risk for malnutrition.[73]

Acutely ill, delirious, or demented patients are at great risk for oropharyngeal dysphagia and aspiration pneumonia. A swallowing evaluation may help enhance the safety and success of oral feeding.

Despite their effectiveness, nutritional oral or enteral supplements are often underutilized in hospitalized patients. A prospective cohort study found that 20% of elderly patients had an average daily in-hospital nutrient intake of less than 50% of their calculated maintenance energy requirements. These patients often had orders for nothing by mouth, and canned supplements were often ordered but not consumed by the patients.[74]

Nutritional Support

High caloric intake is recommended for ill adults (i.e., 25 to 30 kcal/kg/day in women and 30 to 35 kcal/kg/day for men). Even higher amounts of calories are needed in severely malnourished or catabolic individuals.[72] Nutritional support (i.e., nonoral feeding) is considered if prevention or treatment of protein-calorie malnutrition will improve prognosis or quality of life and if the nutritional requirements cannot be met with oral foods and supplements [*see Table 8*].[75] If the gastointestinal tract is functional, enteral nutrition (via nasogastric or nasoenteric tubes) is preferred over parenteral nutrition. Percutaneous tube placement is indicated when long-term tube feeding is not anticipated for weeks to months or for patient comfort. Total parenteral nutrition (intravenous feeding) is essential for survival in patients who cannot eat for extended periods of time and who are not candidates for enteral support (e.g., because of bowel obstruction). However, gastrostomy tube feedings are not recommended for patients with severe dementia, given the absence of data to show that tube feedings improve clinical outcomes in such patients.[76]

Table 8 Nutritional Support

Route	*Patient Characteristics*	*Nutritional Intervention*
Oral	Alert, normal swallowing, mildly to moderately malnourished	High-calorie, high-protein diet; calorie-dense foods (high fat); nutritional supplements between meals
Intravenous fluids	Acutely ill, decreased oral fluid intake, unable to swallow	Glucose solution with electrolytes for ≤ 48 hr as sole nutritional source
Enteral	Oropharyngeal dysphagia, cognitive dysfunction, critically ill (e.g., intubated), or severely malnourished; ability to resume oral feeding within a few weeks (or cyclic use)	Nasoenteric tube feedings (continuous or cyclic) with lactose-free enteral feeding solutions (normal-calorie or normal-calorie, high-nitrogen solutions)
Peripheral parenteral nutrition	Oral and nasoenteric routes temporarily contraindicated (e.g., acute pancreatitis, bowel obstruction)	Peripheral intravenous infusion with isotonic glucose-electrolyte-lipid solution
Total parenteral nutrition	Severely malnourished; hypermetabolic state (e.g., sepsis syndrome); enteral route contraindicated or inadequate	Central intravenous catheter (e.g., in subclavian vein): infusion of high-calorie, hypertonic, balanced (protein, carbohydrate, and amino acid) solutions
Percutaneous endoscopic gastrostomy	Severely malnourished; oropharyngeal dysphagia contraindicates oral feeding (e.g., stroke or advanced dementia) for prolonged time (e.g., ≥ 2 mo) or indefinitely*	Enteral solutions: isotonic or hypertonic (limited by diarrhea; high- or normal-nitrogen formulas)

*Severely demented patients not included.

Correction of Nutrient Deficiencies

Specific nutrient deficiencies result from dietary imbalance, chronic disease, or medications. The most commonly recognized examples are vitamin B_{12} (cobalamin), calcium, and iron deficiencies. Cobalamin deficiency can occur in the absence of classic hematologic or neurologic findings of pernicious anemia. A low serum cobalamin level accompanied by an elevated serum level of methylmalonic acid supports the diagnosis of B_{12} deficiency. Dietary calcium deficiency is common in elderly women. Dietary supplementation is often needed to maintain a daily consumption of 1.2 to 1.5 g of elemental calcium. Vitamin D supplementation, 800 IU daily, is often recommended for the treatment of elderly patients who lack sun exposure or have evidence of osteoporosis or osteomalacia. Iron deficiency is more common in the elderly because of long-term internal or external causes of blood loss.

Sensory Impairment

Hearing and visual losses are the most important and common sensory impairments in elderly people. However, they often go undetected because physicians frequently do not screen for them. Sensory impairments adversely affect the older patient's physical, cognitive, and social functioning. In community-dwelling elders, mood and social relationships are particularly affected by vision impairment, and performance of daily activities is strongly reduced by hearing impairment.

HEARING IMPAIRMENT

Epidemiology

Hearing loss affects between 25% and 40% of the population 65 years of age or older. The prevalence rises with age, ranging from 40% to 66% in patients older than 75 years and more than 80% in patients older than 85 years.[77] Significant hearing loss has been reported in 70% to 90% of nursing-home residents. Hearing loss is associated with emotional and social dysfunction, including depression and social isolation.[77]

Etiology

Hearing loss is categorized as sensorineural, conductive, or mixed. Sensorineural hearing loss is caused by cochlear or retrocochlear diseases and is characterized by decreased thresholds for both air and bone conduction. Presbycusis is the most common cause of sensorineural hearing loss in elderly patients. Other causes include ototoxicity from medications (e.g., aminoglycosides and chemotherapeutic agents), infections involving the eighth cranial nerve, and injury caused by vascular events or tumors of the eighth cranial nerve. Most hearing loss in older adults is sensorineural and caused by presbycusis; however, cerumen impaction and chronic otitis media may be present in up to 30% of elderly patients with hearing loss.[78]

Presbycusis is a bilateral, symmetrical cause of hearing loss at high frequencies, especially at frequencies above 2,000 Hz, and is associated with impaired speech discrimination and loudness recruitment. Presbycusis begins in middle age and causes hearing loss that is typically gradual and bilateral. The most severe changes occur in the inner ear, which is responsible for sensitivity, sound, understanding of speech, and maintenance of equilibrium. Signs and symptoms of presbycusis include a history of progressive high-frequency hearing loss and difficulty in understanding speech, especially in noisy environments. The cause of presbycusis remains uncertain. Presbycusis is also associated with an auditory processing disorder, which makes understanding speech more difficult than would be predicted on the basis of loss of peripheral hearing sensitivity.

Conductive hearing loss occurs when there is impairment of sound transmission to the inner ear. Lesions of the external or middle ear may cause conductive hearing loss. Typically, bone-conduction thresholds are better than air-conduction thresholds. The most common causes of conduction hearing loss are cerumen impaction and otosclerosis. Less common causes include tumors and degenerative disorders (e.g., Meniere disease), trauma, vasculitis, and hemorrhagic disorders.[79]

Diagnostic Evaluation

Hearing impairment may be obvious during casual conversation with the patient. Patients with hearing impairments should be first examined for cerumen impaction. Six independent factors for hearing loss can be identified by brief self-reports of patients 55 to 74 years of age. These factors are (1) age 70 years or older, (2) male gender, (3) 12 or more grades of education, (4) having seen a doctor for deafness or hearing loss, (5) inability to hear a whisper across a room, and (6) inability to hear a normal voice across the room.[80] Several office measures of hearing impairment are useful to clinicians. The whisper test can be performed by asking patients to repeat a short list of whispered numbers from an examiner positioned 2 ft behind them. Although a reasonably sensitive measure, the whisper test lacks reproducibility. A more quantitative approach is a handheld otoscope with a built-in audiometer. This device has high sensitivity and specificity but is expensive and requires some skill to be used correctly.

Patients with hearing loss should undergo formal audiologic assessment. The evaluation includes a pure-tone audiogram to document the decibel loss across frequency ranges and to determine whether the loss is sensorineural or mixed.

Management

Management of hearing loss may be surgical, medical, or rehabilitative. Patient education is important with any treatment. The treatment of presbycusis entails referral to specialists for counseling regarding hearing strategies, emotional support from family members and health care professionals, and the use of hearing aids or various types of assistive listening devices. Referrals for hearing loss are made to audiologists, otolaryngologists, or both. Audiologists have expertise in hearing testing, use of assistive listening devices (e.g., telephone amplifiers, infrared systems, pocket talkers, and visual/tactile alerts for the doorbell, telephone, and smoke alarm), and the selection and fitting of hearing aids.[77] Surgical approaches are indicated for patients with obstructive lesions of the external auditory canal and remedial causes of conductive hearing loss. Stapedectomy is the most common surgical intervention for otosclerosis. Cochlear implants are of value to patients with profound sensorineural deafness for whom a conventional hearing aid is not feasible.[77,79] A change in medical treatments may be indicated for patients who experience sensorineural hearing loss. Serial evaluations of hearing are helpful for patients who are receiving high doses of potentially ototoxic drugs (e.g., anticancer drugs and high-dose loop diuretics).

Most patients with sensorineural or cochlear disease will benefit from aural rehabilitation. Aural rehabilitation includes treatment modalities such as hearing aids, auditory training, and

training in lip reading. Hearing aids remain the usual treatment of patients with sensorineural hearing loss. A variety of devices are available, including programmable hearing aids and amplification circuits that reduce distortion. Hearing aids may fit entirely within the external auditory canal, attach over the patient's earlobe, or sit behind the ear.[77] Assistive listening devices include television listening systems, alerting devices, telephone amplifiers, large-area amplification systems, and remote microphone systems. Remote microphone systems with headsets can be purchased at radio supply stores. They are inexpensive, practical, and capable of improving communication with even severely hearing-impaired patients.

To improve communication with a hearing-impaired patient, an attempt should be made to get the listener's attention before speaking, to face the listener directly to afford visual cues, to reduce background noise, to use facial expressions and gestures, to speak slowly and clearly, to speak only slightly louder than normal and not shout, to rephrase the message if the listener does not understand rather than repeating it, to alert the listener to changes in the topic, and to not turn and walk away while talking.

VISION IMPAIRMENT

Age-related changes in vision, especially presbyopia or farsightedness, are common causes of increasing vision impairment. Other important changes that affect vision include reduced pupillary dilatation, which contributes to poor night vision; discoloration of the crystalline lens; and changes in the vitreous fluid, which may produce dots in the visual field. With normal aging and refractive error, little change occurs in acuity that cannot be corrected or compensated for with ease. Major causes of vision impairment associated with low vision or blindness (i.e., vision worse than 20/200 in either one or both eyes with correction) include cataracts, glaucoma, macular degeneration, and diabetic retinopathy [*see 9:VI Diabetes Mellitus*].

In a study of older adults from three communities, the prevalence of functional blindness increased from 1% in persons 71 to 74 years of age to 17% in those 90 years of age and older. Functional vision impairment increased from 7% in persons 71 to 74 years of age to 39% in those 90 years of age and older.[81] Racial differences are reflected in patterns of vision loss, with whites being more likely to have age-related macular degeneration and African Americans being more likely to have primary open-angle glaucoma. African Americans have a twofold greater prevalence of blindness and vision impairment than whites; there is no difference in prevalence by gender. In nursing homes, the prevalence of bilateral blindness is 17% and the prevalence of vision impairment (i.e., worse than 20/40 but better than 20/200) is 19%. The prevalence of blindness is 29% in nursing-home residents 90 years of age and older.[82]

Cataract

A cataract is an opacity or clouding of the crystalline lens that may affect visual acuity, contrast sensitivity, and light perception. Senile cataracts are often classified as cortical, subcapsular, or nuclear. Although the prevalence of opacification of the lens increases with aging to nearly 100% of those older than 90 years, functional impairment occurs in only half of people with cataracts. Cortical or cuneiform cataracts present as translucent spokes, flakes, or wedges of opacity around the nucleus. Cortical cataracts progress slowly and may eventually involve the entire cortex of the lens. Subcapsular cataracts are more common in younger patients and are associated with use of corticosteroids. They often appear as irregular granules and crystals with various colors. Nuclear cataracts are the most common in elderly white patients; they appear as a yellow or brown discoloration and are associated with increasing myopia and vision deterioration.

Risk factors and etiology The possible risk factors for cataract include exposure to ultraviolet B radiation, a history of diabetes mellitus, alcohol consumption, cigarette smoking, a vitamin-deficient diet, and corticosteroid use.[83] The prevalence of age-related cataract increases from less than 5% in persons younger than 65 years to 46% in persons 75 to 85 years of age.[83] Cigarette smoking has been shown to be an important independent risk factor for age-related cataract. In a prospective cohort study of nearly 21,000 physicians free of cataract at baseline, the relative risk of age-related cataract in men was 36% lower in those who never smoked. Smoking cessation appears to reduce the risk of cataract primarily by limiting total dose-related damage to the lens.[84]

Diagnostic evaluation Initial symptoms of a cataract include glare and poor contrast sensitivity. Contrast sensitivity is the ability to discern subtle variations in shade. It can be tested through the use of figures that vary in contrast, luminance, and spatial frequency.[85] The glare is caused by excess refraction of light rays penetrating the clouded lenses and is most troublesome in bright sunlight or during night driving. Decreased contrast sensitivity is manifested by difficulty distinguishing field objects in poorly illuminated settings. Near visual acuity is more often reduced in posterior subcapsular cataract, and patients often complain of disabling glare during daytime. Patients with cortical cataract complain of glare with oncoming headlights while driving at night. A nuclear sclerotic cataract affects distance vision more than near vision. Nuclear and cortical cataracts have distinct characteristics on examination. Nuclear cataract may be invisible against the red reflex until the cataract is fairly mature. The cataract then appears as a poorly defined central fog. The cortical cataract is typically composed of radially oriented, sharply defined, spokelike opacities.[83]

Management Symptomatic cataracts are managed surgically. The decision whether to perform cataract surgery is based on the likely degree of visual improvement that will occur and its impact on the quality of life weighed against the risks and cost of surgery. A cataract is considered to be clinically significant if it causes a decrease in visual acuity or function that interferes with the patient's performance of activities of daily living.[85] Older drivers with cataracts are at greater risk of vehicular crashes than are cataract-free drivers. Drivers who undergo cataract surgery, compared to those who elect not to have surgery, have a 50% reduction in the risk of a motor vehicle accident.[86] Most cataract surgery in the United States is performed under regional or local anesthesia in an ambulatory setting. Extracapsular cataract surgery with implantation of a posterior chamber lens implant is the procedure of choice. Approximately 95% of patients without other ocular comorbidity who undergo cataract extraction achieve a visual acuity of 20/40 or better.

Glaucoma

Glaucoma is a chronic progressive optic neuropathy characterized by excavation of the optic nerve head and loss of visual

field in the midperiphery. Two anatomic classifications of glaucoma are based on whether the angle of the anterior chamber is open or narrow; the more common open-angle glaucoma is a chronic disease, whereas the less common angle-closure glaucoma is usually an acute disease.

Risk factors and etiology The major risk factor for open-angle glaucoma is thought to be elevation of the intraocular pressure beyond the statistical norm of 21 mm Hg.[87] The high intraocular pressure originates from an increased resistance to drainage of aqueous humor through the trabecular meshwork.[87]

However, many patients with glaucoma have normal-pressure glaucoma characterized by pressures of less than 21 mm Hg.[87] Putative risk factors for open-angle glaucoma include high intraocular pressure, African-American ancestry, positive family history, myopia, and possibly diabetes and systemic hypertension.[87]

Angle-closure glaucoma results from obstruction of aqueous humor flow from the posterior chamber through the pupil into the anterior chamber through the pores of the Schlemm canal. Production of aqueous humor continues, resulting in very high intraocular pressures. The main characteristic of the angle-closure glaucomas is a relative pupillary block with a forward bulging of the iris face, thereby obstructing the aqueous humor flow at the chamber angle.[86] Acute angle-closure glaucoma can be precipitated by the use of dilating eyedrops and is a medical emergency. In contrast, primary open-angle glaucoma is an insidious disease most often discovered during routine examinations.

Diagnostic evaluation Symptoms of glaucoma include seeing halos around lights and experiencing reduced vision under low levels of illumination. Glaucoma is frequently asymptomatic and detected only during ophthalmologic examination; unless it comes to clinical attention, glaucoma will cause progressive loss of visual field and eventually lead to blindness. The diagnosis of glaucoma is based on an eye examination that includes tonometry, gonioscopy (which examines the angle of the anterior chamber), inspection of the optic disk and nerve fiber layer, and visual-field testing. With injury to the optic nerve, there is the appearance of cupping of the optic disk, with an increase in the cup-to-disk ratio and visual dysfunction in the midperipheral field of vision. As the disease progresses, deterioration of central visual functions, including acuity, becomes evident.

Intraocular pressure is measured by determination of the force required to flatten the central cornea (applanation). Applanation tonometry can be performed with an optical measuring device or an electrical strain gauge.

Management The treatment of glaucoma includes both medical and surgical approaches. Intraocular pressure is reduced either by decreasing the amount of aqueous humor produced by the ciliary body or by increasing its outflow through the trabecular meshwork, through the uveoscleral pathway, or through a surgically created path.[85] Medical treatments are targeted at the physiology of intraocular pressure. The cell membranes of the nonpigmented ciliary epithelial cells contain α-adrenoceptors and β-adrenoceptors, carbonic anhydrase, and sodium and potassium activated adenosine triphosphatases (ATPases). By stimulation or inhibition of these enzymes or receptors, the active transport of aqueous humor across the blood-aqueous barrier can be modulated to reduce intraocular pressure.[87]

Topical eyedrops are the most common medical treatment of open-angle glaucoma. Agents either decrease the production of aqueous or increase its outflow and absorption. Therapeutic options include: topical carbonic anhydrase inhibitors (e.g., brinzolamide); beta-adrenergic blockers (e.g., timolol, betaxolol); alpha$_2$-adrenergic agonists (e.g., brimonidine); and prostaglandin analogues (e.g., latanoprost and bimatoprost).[88,89]

The aim of medical treatment is to obtain a target pressure at which progression of visual-field defects is halted.[85] Combinations of these therapies can be attempted if treatment with one of these agents fails to reduce or maintain target pressures or if there is progression of visual-field defects.

Surgical therapy for glaucoma includes iridectomy to enhance flow of aqueous humor in the treatment of angle-closure glaucoma and, for open-angle glaucoma, argon-laser therapy to improve outflow through the trabecular meshwork. Surgery for glaucoma is usually reserved for patients in whom target pressures cannot be achieved with medical or laser therapy. Good success rates have been described with argon-laser trabeculectomy, but additional surgery is often needed.

Age-Related Macular Degeneration

Age-related macular degeneration (AMD) causes sudden worsening and distortion of central vision, or scotoma, progressing rapidly over weeks or months until scarring is complete and no further vision is lost (legal blindness).[90] AMD impairs central vision that is required for reading, driving, face recognition, and all fine visual tasks. The insidious loss of central vision results in the initial symptoms of reduced visual perception and visual sensitivity to light and gradual progression to legal blindness despite preservation of peripheral vision.

Risk factors and etiology Epidemiologic studies show that cigarette smoking increases the risk for the development of AMD and the consumption of dietary carotenoids and vitamins A, C, and E may prove protective. However, cigarette smoking is the only modifiable risk factor for AMD and maculopathy that has been consistently identified in cohort studies.[91,92]

Diagnostic evaluation AMD is highlighted by the presence of abnormalities in the macular area. These abnormalities include soft drusen, yellow-white deposits of extracellular material containing debris external to the retinal pigment epithelium, and hyperpigmentation or hypopigmentation (or both) of the retinal pigment epithelium. Neurosensory detachment, retinal hemorrhages, and retinal scarring gradually result in decreased visual function of photoreceptors in the central vision.[93] Late maculopathy includes both dry and neovascular AMD. The dry form of AMD is more common and occurs with atrophy of the retinal pigment epithelium. Neovascular, or exudative, AMD is characterized by choroidal neovascularization with vascular leakage into subretinal spaces. Recognition of the less common exudative AMD is important because argon-laser photocoagulation is effective in reducing loss of visual acuity.[93]

Management There is no effective treatment of the dry form of AMD. Clinical trials, however, have found that laser photocoagulation of exudative forms of AMD decreases the rate of severe vision loss and preserves contrast sensitivity.[93] Laser photocoagulation therapy is the only treatment of AMD with proven long-term benefit. However, other therapies have shown promise, including photodynamic therapy (verteporfin), which delays or prevents loss of vision during at least 1 year of

follow-up in patients with predominantly classic neovascular lesions.[93] To identify patients who may benefit from laser therapy, home monitoring for symptoms of scotoma or distorted vision is performed with the use of an Amsler grid.[93]

Treatment of AMD with oral antioxidants may prove effective for subgroups of nonsmoking patients who have either (1) extensive intermediate-size drusen and noncentral geographic atrophy in one or both eyes or (2) advanced AMD or unilateral vision loss caused by AMD.[91] In a clinical trial, supplementation with vitamins C and E, β-carotene, and zinc slowed progression to advanced AMD.[91]

Iatrogenic Illness

Iatrogenic, or physician-induced, illness results from a diagnostic procedure or therapeutic intervention that is not a natural consequence of the patient's disease.[94] Iatrogenic illnesses include complications of drug therapy and of diagnostic or therapeutic procedures, nosocomial infections, fluid and electrolyte disorders, and trauma.[95]

POLYPHARMACY

The most common documented cause of iatrogenic illness is adverse drug reactions, usually associated with polypharmacy.[95] Adverse drug events are more likely to occur in elderly patients because of the age-related changes in drug metabolism, the occurrence of multiple comorbidities, and the use of polypharmacy. The incidence of adverse drug reactions increases with advancing age and number of chronic diseases requiring drug therapy. The concomitant use of several medications increases the risk of drug interactions, unwanted effects, and adverse reactions. Drug-drug interactions resulting in hospitalization have been attributed to hypoglycemia associated with use of the oral hypoglycemic agent glyburide, digoxin toxicity associated with use of the antibiotic clarithromycin, and hyperkalemia in patients treated with angiotensin-converting enzyme inhibitors.[96] Suboptimal drug prescribing, including the inappropriate use or underuse of medications, is common in older outpatients and inpatients and is associated with significant morbidity.[97] Adverse drug events occur in nursing-home patients treated with opioids, antidepressants, antipsychotic agents, and antiseizure medications.[98] A United States consensus panel has identified medications or medication classes that should generally be avoided in persons 65 years age or older because the agents are either ineffective or pose unnecessarily high risk for older persons when safer alternatives are available.[99] The panel's recommendations also included medications to be avoided in older persons known to have specific medical conditions.[99] Many of these agents are psychotropic and anticholinergic drugs (http://www.seniorcarepharmacist.com/inappropriate).

Age-Related Changes in Drug Metabolism

Many medications should be used with special caution in elderly patients because of age-related changes in drug pharmacokinetics (drug disposition) and pharmacodynamics (target-tissue effects). Although drug absorption is not reduced in healthy elderly persons, absorption of medications can be reduced by disease states (e.g., malabsorption) or concomitant administration of drugs that decrease absorption of medications (e.g., antacids). Drug distribution is altered by aging, primarily because of body-composition changes, with a decrease in total body water and lean body mass and a relative increase in body fat. Consequently, water-soluble drugs achieve a higher serum concentration, whereas lipid-soluble drugs have a prolonged elimination half-life. This change is especially important in regard to drugs that are lipid soluble and penetrate the blood-brain barrier (e.g., diazepam). Although serum protein levels are not significantly affected by aging, many elderly patients have reduced levels of serum albumin resulting from acute or chronic disease or malnutrition. Consequently, displacement of a drug by one that binds very highly to albumin enhances the delivery of that agent to the target site, thereby increasing the risk of an adverse reaction. For example, bleeding (excessive anticoagulation) may occur when patients treated with warfarin are given drugs such as sulfas or phenytoin.

A decrease in hepatic blood flow with usual aging will decrease the rate of metabolism of drugs that undergo a high degree of first-pass extraction (e.g., propranolol). Aging and diseases affect phase I hepatic metabolism, the microsomal enzyme mixed-function oxidase system. Active metabolites of drugs that undergo phase I metabolism may prolong the effects of the parent medication (e.g., diazepam). Phase II metabolism, the conjugation of drugs, is not significantly affected by aging. Consequently, the elimination of agents that undergo phase II metabolism is unaffected by normal aging.

Drug elimination is mainly influenced by renal function. The age-associated decrease in renal function, which results in decreased creatinine clearance, necessitates lower maintenance doses of renally excreted drugs in elderly patients.

The effect of aging on target-organ responsiveness to medications is less well established. However, decreased beta-adrenergic receptor sensitivity and increased sensitivity to opiates is well established. Many elderly patients are also more susceptible to the adverse effects of anticholinergic drugs, notably constipation, dry mouth, and delirium.

Diagnostic Evaluation

Because of unpredictable effects of aging on the metabolism of drugs, physicians should suspect an adverse drug reaction whenever any new symptom occurs. Also, blood levels of medications that have narrow therapeutic windows (e.g., aminoglycosides) should be considered.

Prevention and Management

The prevention of iatrogenic illness from inappropriate drug prescribing begins with an understanding of the rational use of medications in elderly patients. Consensus criteria for appropriate drug prescribing offer useful guidelines.[99]

In general, prescribing the fewest medications at the lowest needed dosage is a rational approach to the prevention of iatrogenic illness. Knowing the pharmacology of prescribed drugs and the age-related alterations in drug disposition and tissue sensitivity and using lower than standard doses of most drugs when the therapeutic dose is uncertain should help physicians avoid adverse drug reactions. When prescribing a new medication for an elderly patient, a practical approach is to use the following criteria:

1. Determine whether the drug is lipid or water soluble.
2. Determine whether the drug is highly bound to albumin.
3. Determine whether the drug undergoes cytochrome P-450 metabolism (substrate, inducer, or inhibitor of microsomal enzymes).

Table 9 Long-term Care Options

Site	*Patient Characteristics*	*Comments*
Home (independent living, house, apartment)	Independent performance of ADL; stable chronic diseases; adequate social support network	Acute medical illness may require short-term home care (skilled nursing services); home safety evaluation for environmental hazards
Home with formal support services	Needs assistance with performance of ADL (e.g., bathing, dressing); adequate social support network; acute illness or convalescence from hospitalization; requires skilled nursing care or physical therapy	Skilled services covered for limited period by Medicare; custodial services not covered long-term (e.g., home health aide); risk of elder abuse or neglect because of caregiver strain or loss of informal supports (e.g., illness of spouse)
Community services	Physically or cognitively impaired (limits performance of ADL); limited finances; limited informal supports; caregivers need respite (relief from caregiving)	Options: adult day care; nutritional programs (e.g., Meals-on-Wheels); protective services (suspected elder abuse); transportation services; case management services (professionally coordinated); categorized services (e.g., PACE) for eligible patients
Residential care facilities (assisted living, continuing care retirement communities)	Patients for whom independent living is either no longer desired or feasible; retirees, persons able to perform ADL with minimal or no assistance	Assisted living is ideal for demented patients who are too functional for a nursing home but are unable to safely live at home; retirement communities are vertically integrated, permitting residents to move from independent living (apartment) to assisted living or nursing care as needed
Rehabilitation/hospital	Categorical illness (e.g., stroke or hip fracture); able to tolerate physical therapy (e.g., ≥ 3 hr daily); good informal home supports; likely return home	Interdisciplinary care with focus on returning patient to independent living; patients must be able to participate in rehabilitative services and demonstrate potential to improve ambulation and performance of ADL
Skilled nursing facility	Dependence in ADL or ambulation preventing discharge to home; requires skilled nursing care or physical therapy; too impaired for rehabilitation hospital (noncategorical illness, cannot perform therapy for ≥ 3 hr daily); inadequate social supports at home	The choice of skilled nursing facility, home care, or subacute care unit is often an issue of patient choice, disease severity, availability of services (geographic), and family or caregiver availability; limited Medicare coverage for skilled services forces many patients to eventually enter a long-term care facility or return home
Subacute care unit	Similar to patients in skilled nursing care but typically requiring augmented professional services (e.g., hyperalimentation, respiratory care, peripheral intravenous catheters, or increased nursing staff time)	Subacute care units bridge the gap between acute hospitalization and the patient's return home; market forces have encouraged growth of these units (e.g., hospitals reduce patient length of stay and are reimbursed for cost-based skilled nursing care); short-term stay (< 2 mo) is typical
Long-term care facility	Dependent in performance of ADL; unable to return home (temporarily or permanently) to independent living; ineligible for skilled, subacute, or rehabilitative services; inadequate social supports (e.g., lives alone, no caregivers)	Self-pay and Medicaid are usual sources of payment; many states offer a Medicaid waiver program to provide home services to frail patients who would otherwise require nursing home placement
Palliative (hospice) care	Patients with terminal illness (prognosis ≤ 6 mo) (e.g., metastatic cancer or end-stage heart or renal failure)	Provides comfort measures in home (or inpatient unit where available); Medicare covers palliative (comfort-related) not curative (e.g., elective surgery) services; underutilized by patients (delayed recommendation by physician; home hospice services limited compared with skilled nursing care)

ADL—activities of daily living PACE—Program for All-inclusive Care of Elderly

4. Determine whether the drug is renally excreted (use lower maintenance doses in older patients and in patients with renal insufficiency).

This information can be found by literature searches, a review of newsletters on drug prescribing, or consultation with a pharmacologist or clinical pharmacist.

MEDICAL ERRORS

Medical errors are common in the care of elderly hospitalized patients. These errors often involve incorrect drug dispensing that sometimes causes patient morbidity or mortality. Most medical errors are the result of systems errors in hospitals. Concerns over the reported incidence of medical errors led the Institute of Medicine to release a report advocating dramatic, systemwide changes in hospitals, such as the use of computerized medical information systems and other support systems, to reduce the rate of errors.[100] Computerized medical information systems improve antibiotic selection, limit the emergence of antibiotic-resistant pathogens, and lessen the risk of adverse drug events.[100]

NOSOCOMIAL INFECTION

Nosocomial pathogens are primarily transmitted through contact with hospital or nursing home personnel. Urinary tract, skin (intravascular), lung, and wound infections are common examples of nosocomial infection. Of growing concern for elderly patients are infections with resistant strains of gram-negative bacilli, methicillin-resistant *Staphylococcus aureus*, and vancomycin-resistant enterococci. Resistant urinary tract infections are common after prolonged indwelling urinary catheterization.

Risk Factors and Etiology

Factors promoting nosocomial pneumonia include gastric aspiration, spread of pathogens by poorly cleansed hands of medical and nursing personnel, fecal-oral spread of pathogens, and cross-contamination from other patients. Patients with physical debility, patients who have had a prolonged hospital stay, and

patients who have been exposed to broad-spectrum antibiotics are at risk for nosocomial infections.

Vancomycin-resistant enterococci are likely transmitted from patient to patient by the unwashed hands of health care workers, by contaminated medical equipment, or on environmental surfaces (e.g., bed rails or blood pressure cuffs). Nosocomial pneumonia results from colonization of the upper respiratory and gastrointestinal tracts and occurs most often in critically ill and ventilator-dependent patients.

Prevention and Management

Nosocomial infection can be prevented by washing hands and cleaning medical equipment (e.g., stethoscopes) between patient contacts, wearing gloves during invasive procedures or contact with wounds or mucous membranes, using aseptic techniques when inserting or changing urinary catheters, isolating infected patients (e.g., in nursing homes), elevating the patient's head (to lessen the risk of aspiration), replacing broad-spectrum antibiotics with narrow-spectrum antibiotics on the basis of bacterial-sensitivity reports, and limiting the use of urinary catheters. Prophylactic antimicrobial therapies and routine catheter replacement are not recommended.[95]

Long-term Care

Long-term care refers to the provision of comprehensive health care services, including personal health and social services, delivered over an extended period to people with limited functional capacity.

AVAILABLE CARE SERVICES

The spectrum of long-term care ranges from home and social services provided to patients living in the community to residential and long-term care facilities [*see Table 9*].

In the United States, most elderly persons live in the community. Only 5% at any time are residing in long-term care facilities (i.e., nursing homes). However, the probability of nursing home use increases sharply with age: 25% of persons 85 years of age and older reside in nursing homes, and the lifetime risk of entering a nursing home for people who turned 65 years of age in 1990 is estimated at 40%.[101] About 25% of individuals in the United States will spend at least 1 year in a nursing home, with more women than men having total lifetime nursing home use of 5 years or more.[101] The high cost of institutional care and the growing number of frail Americans have provided an impetus to develop less expensive levels of care (e.g., residential care facilities), ambulatory health centers (e.g., Program for All-inclusive Care of Elderly [PACE]), and community services that will enable the frail elderly patient to remain at home [*see CE:IX Palliative Medicine*].

ADVANCE DIRECTIVES

Health care institutions are required to ascertain whether patients have advance directives and to include copies of them in the medical record. A living will is an advance directive by which a person specifies the circumstances in which life-sustaining treatment is to be provided or discontinued in the event of terminal illness if the person is unable to communicate with health professionals. A durable power of attorney for health care is an advance directive by which a person designates a proxy to represent his or her wishes if he or she loses decision-making capacity. In discussions with physicians, patients should communicate their wishes for end-of-life care, including cardiopulmonary resuscitation, intensive care (e.g., ventilator support), and nutritional support during acute or end-stage illness [*see CE:IX Palliative Care*].[102]

The author participates in the speakers' bureau of Pfizer, Inc.

The drug haloperidol, which is discussed in this chapter, has not been approved by the FDA for use as a treatment of delirium.

References

1. Tinetti ME, Williams CS, Gill TM: Dizziness among older adults: a possible geriatric syndrome. Ann Intern Med 132:337, 2000

2. Hoffman C, Rice D, Sung HY: Persons with chronic conditions: their prevalence and costs. JAMA 276:1473, 1996

3. Inouye SK: Delirium in hospitalized older patients. Clin Geriatr Med 14:745, 1998

4. McCusker J, Cole M, Abrahamowicz M, et al: Delirium predicts 12-month mortality. Arch Intern Med 162:457, 2002

5. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC, 1994

6. Meagher DJ: Delirium: optimising management. BMJ 322:144, 2001

7. Inouye SK, Viscoli CM, Horwitz RI, et al: A predictive model for delirium in hospitalized elderly medical patients based on admission characteristics. Ann Intern Med 119:474, 1993

8. Inouye SK, Charpentier PA: Precipitating factors for delirium in hospitalized elderly persons. JAMA 275:852, 1996

9. Agostini JV, Leo-Summers LS, Inouye SK: Cognitive and other adverse effects of diphenhydramine use in hospitalized older patients. Arch Intern Med 161:2091, 2001

10. Dyer CB, Ashton CM, Teasdale TA: Postoperative delirium: a review of 80 primary data-collection studies. Arch Intern Med 155:461, 1995

11. Moller JT, Cluitmans P, Rasmussen LS, et al: Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International study of post-operative cognitive dysfunction. Lancet 351:857, 1998

12. Marcantonio ER, Goldman L, Mangione CM, et al: A clinical prediction rule for delirium after elective noncardiac surgery. JAMA 271:134, 1994

13. Tune LE: Serum anticholinergic activity levels and delirium in the elderly. Semin Clin Neuropsychiatry 5:149, 2000

14. Brown TM: Drug-induced delirium. Semin Clin Neuropsychiatry 5:113, 2000

15. Trzepacz PT: Delirium: advances in diagnosis, pathophysiology, and treatment. Psychiatr Clin North Am 19:429, 1996

16. Ely EW, Margolin R, Francis J, et al: Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the intensive care unit (CAM-ICU). Crit Care Med 29:1481, 2001

17. Inouye SK, Bogardus ST Jr, Charpentier PA, et al: A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med 340:669, 1999

18. Marcantonio ER, Flacker JM, Wright RJ, et al: Reducing delirium after hip fracture: a randomized trial. J Am Geriatr Soc 49:516, 2001

19. Fantl JA, Newman DK, Colling J, et al: Urinary incontinence in adults: acute and chronic management. Clinical Practice Guideline, No 2. US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research (AHCPR Publication No 96-0682), Rockville, Maryland, March 1996

20. Bland DR, Dugan E, Cohen SJ, et al: The effects of implementation of the Agency for Health Care Policy and Research urinary incontinence guidelines in primary care practices. J Am Geriatr Soc 51:979, 2003

21. Assessment and treatment of urinary incontinence. Scientific Committee of the First International Consultation on Incontinence. Lancet 355:2153, 2000

22. Locher JL, Goode PS, Roth DL, et al: Reliability assessment of the bladder diary for urinary incontinence in older women. J Gerontol A Biol Sci Med Sci 56:M32, 2001

23. Ouslander JG, Schapira M, Schnelle JF, et al: Does eradicating bacteriuria affect the severity of chronic urinary incontinence in nursing home residents? Ann Intern Med 122:749, 1995

24. Marks LS, Dorey FJ, Macairan ML, et al: Three-dimensional ultrasound device for rapid determination of bladder volume. Urology 50:341, 1997

25. Smith DN, Appell RA, Winters JC, et al: Collagen injection therapy for female intrinsic sphincteric deficiency. J Urol 157:1275, 1997

26. Seim A, Sivertsen B, Eriksen BC, et al: Treatment of urinary incontinence in women in general practice: observational study. BMJ 312:1459, 1996

27. Burgio KL, Locher JL, Goode PS, et al: Behavioral vs drug treatment for urge urinary incontinence in older women: a randomized controlled trial. JAMA 280:1995, 1998

28. Burgio KL, Goode PS, Locher JL, et al: Behavioral training with and without biofeedback in the treatment of urge incontinence in older women: a randomized controlled trial. JAMA 288:2293, 2002

29. Holroyd-Leduc JM, Straus SE: Management of urinary incontinence in women: scientific review. JAMA 291:986, 2004

30. Manson JE, Martin KA: Clinical practice. Postmenopausal hormone-replacement therapy. N Engl J Med 345:34, 2001

31. Tebyani N, Patel H, Yamaguchi R, et al: Percutaneous needle bladder neck suspension for the treatment of stress urinary incontinence in women: long-term results. J Urol 163:1510, 2000

32. Winters JC, Chiverton A, Scarpero HM, et al: Collagen injection therapy in elderly women: long-term results and patient satisfaction. Urology 55:856, 2000

33. O'Connor RC, Gerber GS, Avila D, et al: Comparison of outcomes after single or DOUBLE-CUFF artificial urinary sphincter insertion. Urology 62:723, 2003

34. Ouslander JG: Management of overactive bladder. N Engl J Med 350:786, 2004

35. Gleason DM, Susset J, White C, et al: Evaluation of a new once-daily formulation of oxybutynin for the treatment of urinary urge incontinence. Urology 54:420, 1999

36. Millard R, Tuttle J, Moore K, et al: Clinical efficacy and safety of tolterodine compared to placebo in detrusor overactivity. J Urol 161:1551, 1999

37. Diokno AC, Appell RA, Sand PK, et al: Prospective, randomized, double-blind study of the efficacy and tolerability of oxybutynin and tolterodine for overactive bladder: results of the OPERA trial. Mayo Clin Proc 78:681, 2003

38. Thakar R, Stanton S: Regular review: management of urinary incontinence in women. BMJ 321:1326, 2000

39. Stevens TK, Soffer EE, Palmer RM: Fecal incontinence in elderly patients: common, treatable, yet often undiagnosed. Cleve Clin J Med 70:441, 2003

40. Nelson RL: Epidemiology of fecal incontinence. Gastroenterology 126:S3, 2004

41. DeLillo AR, Rose S: Functional bowel disorders in the geriatric patient: constipation, fecal impaction, and fecal incontinence. Am J Gastroenterol 95:901, 2000

42. Camilleri M, Lee JS, Viramontes B: Insights into the pathophysiology and mechanisms of constipation, irritable bowel syndrome, and diverticulosis in older people. J Am Geriatr Soc 48:1142, 2000

43. Tinetti ME: Clinical practice: preventing falls in elderly persons. N Engl J Med 348:42, 2003

44. Guideline for the prevention of falls in older persons. American Geriatrics Society, British Geriatrics Society, American Academy of Orthopaedic Surgeons Panel on Falls Prevention. J Am Geriatr Soc 49:664, 2001

45. Maki BE: Gait changes in older adults: predictors of falls or indicators of fear. J Am Geriatr Soc: 45:313, 1997

46. Tinetti ME, Baker DI, McAvay G, et al: A multifactorial intervention to reduce the risk of falling among elderly people living in the community. N Engl J Med 331:821, 1994

47. Gillespie LD, Gillespie WJ, Robertson MC, et al: Interventions for preventing falls in elderly people. Cochrane Database Syst Rev 3: 2003

48. Wu G: Evaluation of the effectiveness of Tai Chi for improving balance and preventing falls in the older population—a review. J Am Geriatr Soc 50:746, 2002

49. Wolf SL, Barnhart HX, Kutner NG, et al: Reducing frailty and falls in older persons: an investigation of Tai Chi and computerized balance training. Atlanta FICSIT Group. Frailty and Injuries: Cooperative Studies of Intervention Techniques. J Am Geriatr Soc 44:489, 1996

50. Campbell AJ, Robertson MC, Gardner MM, et al: Randomised controlled trial of a general practice programme of home based exercise to prevent falls in elderly women. BMJ 315:1065, 1997

51. Robertson MC, Devlin N, Gardner MM, et al: Effectiveness and economic evaluation of a nurse delivered exercise programme to prevent falls. 1: Randomised control trial. BMJ 322:697, 2001

52. Palmer R: Falls in elderly patients: predictable and preventable. Cleve Clin J Med 68:303, 2001

53. Parker MJ, Gillespie LD, Gillespie WJ: Hip protectors for preventing hip fractures in the elderly. Cochrane Database Syst Rev 3:CD001255, 2003

54. Kannus P, Parkkari J, Niemi S, et al: Prevention of hip fracture in elderly people with use of a hip protector. N Engl J Med 343:1506, 2000

55. van Schoor NM, Smit JH, Twisk JW, et al: Prevention of hip fractures by external hip protectors: a randomized controlled trial. JAMA 289:1957, 2003

56. Bergstrom N, Bennett MA, Carlson CE, et al: Treatment of pressure ulcers. Clinical Practice Guideline No 15. US Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research (AHCPR Publication No 95-0652), Rockville, Maryland, December 1994

57. Pressure ulcers: prevalence, cost and risk assessment: consensus development conference statement. The National Pressure Ulcer Advisory Panel. Decubitus 2:24, 1989

58. Brandeis GH, Morris JN, Nash DJ, et al: The epidemiology and natural history of pressure ulcers in elderly nursing home residents. JAMA 264:2905, 1990

59. Allman RM: Pressure ulcer prevalence, incidence, risk factors, and impact. Clin Geriatr Med 13:421, 1997

60. Thomas DR: Prevention and treatment of pressure ulcers: what works? what doesn't? Cleve Clin J Med 68:704, 2001

61. Orlando PL: Pressure ulcer management in the geriatric patient. Ann Pharmacother 32:1221, 1998

62. Lyder CH, Preston J, Grady JN, et al: Quality of care for hospitalized Medicare patients at risk for pressure ulcers. Arch Intern Med 161:1549, 2001

63. Graumlich JF, Blough LS, McLaughlin RG, et al: Healing of pressure ulcers with collagen or hydrocolloid: a randomized, controlled trial. J Am Geriatr Soc 51:147, 2003

64. Belmin J, Meaume S, Rabus MT, et al: Sequential treatment with calcium alginate dressings and hydrocolloid dressings accelerates pressure ulcer healing in older subjects: a multicenter randomized trial of sequential versus nonsequential treatment with hydrocolloid dressings alone. J Am Geriatr Soc 50:269, 2002

65. Landi F, Aloe L, Russo A, et al: Topical treatment of pressure ulcers with nerve growth factor: a randomized clinical trial. Ann Intern Med 139:635, 2003

66. Cullum N, Deeks J, Sheldon TA, et al: Beds, mattresses and cushions for pressure sore prevention and treatment. Cochrane Database Sys Rev (3):CD001735, 2003

67. Ferrell BA, Osterweil D, Christenson P: A randomized trial of low-air-loss beds for treatment of pressure ulcers. JAMA 269:494, 1993

68. Marwick C: NHANES III health data relevant for aging nation. JAMA 277:100, 1997

69. Azad N, Murphy J, Amos SS, et al: Nutrition survey in an elderly population following admission to a tertiary care hospital. CMAJ 161:511, 1999

70. Meyyazhagan S, Palmer RM: Nutritional requirements with aging: prevention of disease. Clin Geriatr Med 18:557, 2002

71. Covinsky KE, Martin GE, Beyth RJ, et al: The relationship between clinical assessment of nutritional status and adverse outcomes in older hospitalized medical patients. J Am Geriatr Soc 47:532, 1999

72. Milne AC, Potter J, Avenell A: Protein and energy supplementation in elderly people at risk from malnutrition. Cochrane Database Syst Rev (3): CD003288, 2002

73. Lauque S, Arnaud-Battandier F, Mansourian R, et al: Protein-energy oral supplementation in malnourished nursing-home residents: a controlled trial. Age Ageing 29:51, 2000

74. Sullivan DH, Sun S, Walls RC: Protein-energy undernutrition among elderly hospitalized patients: a prospective study. JAMA 281:2013, 1999

75. Haddad RY, Thomas DR: Enteral nutrition and enteral tube feeding: review of the evidence. Clin Geriatr Med 18:867, 2002

76. Finucane TE, Christmas C, Travis K: Tube feeding in patients with advanced dementia: a review of the evidence. JAMA 282:1365, 1999

77. Yueh B, Shapiro N, MacLean CH, et al: Screening and management of adult hearing loss in primary care. JAMA 289:1976, 2003

78. Bogardus ST, Yueh B, Shekelle PG: Screening and management of adult hearing loss in primary care: clinical applications. JAMA 289:1986, 2003

79. Nadol JB Jr: Hearing loss. N Engl J Med 329:1092, 1993

80. Reuben DB, Walsh K, Moore AA, et al: Hearing loss in community-dwelling older persons: national prevalence data and identification using simple questions. J Am Geriatr Soc 46:1008, 1998

81. Salive ME, Guralnik J, Christen W, et al: Functional blindness and visual impairment in older adults from three communities. Ophthalmology 99:1840, 1992

82. Tielsch JM, Javitt JC, Coleman A, et al: The prevalence of blindness and visual impairment among nursing home residents in Baltimore. N Engl J Med 332:1205, 1995

83. Das A: Prevention of visual loss in older adults. Clin Geriatr Med 15:131, 1999

84. Christen WG, Glynn RJ, Ajani UA, et al: Smoking cessation and risk of age-related cataract in men. JAMA 284:713, 2000

85. Solomon R, Donnenfeld ED: Recent advances and future frontiers in treating age-related cataracts. JAMA 290:248, 2003

86. Owsley C, McGwin G, Sloane M, et al: Imapct of cataract surgery on motor vehicle crash involvement by older adults. JAMA 288:841, 2002

87. Hoyng PF, Van Beek LM: Pharmacological therapy for glaucoma: a review. Drugs 59:411, 2000

88. Alward WL: Medical management of glaucoma. N Engl J Med 339:1298, 1998

89. Novack GD, O'Donnell MJ, Molloy DW: New glaucoma medications in the geriatric population: efficacy and review. J Am Geriatr Soc 50:956, 2002

90. Bird AC, Bressler NM, Bressler SB, et al: An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group. Surv Ophthalmol 39:367, 1995

91. Chopar A, Chakravarthy U, Verma D: Age related macular degeneration. BMJ 326:485, 2003

92. McCarty CA, Mukesh BN, Fu CL, et al: Risk factors for age-related maculopathy. Arch Ophthalmol 119:1455, 2001

93. Fine SL, Berger JW, Maguire MG, et al: Age-related macular degeneration. N Engl J Med 342:483, 2000

94. A randomized, placebo-controlled, clinical trial of high-dose supplementation with vitamins C and E, beta carotene, and zinc for age-related macular degeneration and vision loss (AREDS report no. 8). Age-Related Eye Disease Study Research Group. Arch Ophthalmol 119:1417, 2001

95. Riedinger JL, Robbins LJ: Prevention of iatrogenic illness: adverse drug reactions and nosocomial infections in hospitalized older adults. Clin Geriatr Med 14:681, 1998

96. Juurlink DN, Mamdani M, Kopp A: Drug-drug interactions among elderly patients hospitalized for drug toxicity. JAMA 289:1652, 2003

97. Hanlon JT, Schmader KE, Ruby CM, et al: Suboptimal prescribing in older inpatients and outpatients. J Am Geriatr Soc 49:200, 2001

98. Field TS, Gurwitz JH, Avorn J, et al: Risk factors for adverse drug events among nursing home residents. Arch Intern Med 162:1629, 2001

99. Fick DM, Cooper JW, Wade WE, et al: Updating the Beers criteria for inappropriate medication use in older adults: results of a US consensus panel of experts. Arch Intern Med 163:2716, 2003

100. To Err Is Human: Building a Safer Health System. Kohn L, Corrigan J, Donaldson M, Eds. National Academy Press, Washington, DC, 2000

101. Kemper P, Murtaugh CM: Lifetime use of nursing home care. N Engl J Med 324:595, 1991

102. Miller DL, Bolla LR: Patient values: the guide to medical decision making. Clin Geriatr Med 14:813, 1998

Acknowledgment

Table 2 Adapted from "Clarifying the confusion: the confusion assessment method. A new method for detection of delirium," by S. K. Inouye, C. H. van Dyck, C. A. Alessi, et al, in *Annals of Internal Medicine* 113:941, 1990.

X REHABILITATION OF GERIATRIC PATIENTS

Stephanie Studenski, m.d., m.p.h.
Cynthia J. Brown, m.d., p.t.

The Role of the Generalist in Rehabilitation

The generalist plays a key role in the recovery of function in older patients who have had disabling illnesses. In most health care settings other than formal inpatient rehabilitation units, the primary care physician is the coordinator of medical rehabilitation services. Access to and duration of rehabilitation services are undergoing rapid change as Medicare and managed-care policies become more restrictive. Medicare has converted rehabilitation services in the inpatient, postacute, and home settings to prospective payment systems with extensive regulatory and documentation requirements.[1] The primary care physician must be able to assess rehabilitation potential, determine specific patient needs, and match them with the appropriate setting to optimize patient care.

To fulfill this responsibility, the physician must be able to assess the degree of disability, the effect of comorbidity on function, the possibility of recovery, and the indications for specific rehabilitation therapies.[2] Rehabilitation service planning is usually based on a model of diagnosis and treatment that differs from that of traditional medical care. The revised World Health Organization International Classification of Functioning, Disability and Health (ICIDH-2) can be used to assess the causes of disability, plan treatment approaches, and determine the outcomes of care [*see Figure 1*].[3] In this framework, abnormalities in organ system structure or physiologic function are called impairments. They lead to difficulties with individual activities and participation in society. Environmental factors (e.g., stairways and crosswalks) and personal factors (e.g., culture and education) can influence the effect of organ system impairments on activities and participation. Thus, a patient who has recently had a stroke may have impairments of sensory, motor, and language function that limit mobility and self-care, and these limitations can interfere with work and family roles.

The rehabilitation assessment includes evaluation of impairments in body structure and function, as well as the patient's activity limitations and participation goals. The treatment plan includes efforts to reverse or modify impairments, such as decreased strength, and to enable the patient to adapt to activity limitations, such as difficulty in walking. An evaluation of the effectiveness of the service may include measurement of gains in organ function, activities, and participation. The evaluation should also take into account environmental and personal factors. The generalist team can also implement early exercise-rehabilitation services during hospitalization for acute illness. This practice has been shown to contribute to improved function.[4] A form of preventive rehabilitation termed prehab has recently been shown to reduce the onset of new disability in high-risk older adults, but this service is not currently covered by Medicare.[5]

Stroke

Each year, more than 700,000 people in the United States suffer a stroke. Stroke is the leading cause of major long-term functional limitation. Each year, more money is spent in the rehabilitation of stroke patients than is spent on the rehabilitation of any other patient group. Guidelines for poststroke rehabilitation have been developed by an interdisciplinary panel. These guidelines consist of evidence-based recommendations for assessment, referral, and patient management.[6]

Assessment

The primary care physician should assess the type of stroke, the extent of neurologic deficits [*see Table 1*], the presence or absence of comorbid conditions, and the patient's prestroke functional status. Stroke guidelines provide extensive recommendations to help assess these features, measure recovery, and determine the benefits of rehabilitation.[6] A compendium of the recommended assessments has been gathered into a so-called stroke toolbox, which can be found on the Internet (http://www2.kumc.edu/coa/Stroke_Toolbox/stroke-tool.htm).[7]

Selecting the Best Setting for Care

The setting for rehabilitation should be determined as soon as the patient's neurologic function and medical condition are reasonably stable. The appropriate setting is determined by the presence, severity, and complexity of the patient's functional limitations; cognitive status (especially ability to learn); ability to tolerate up to 3 hours of therapy daily; need for close medical monitoring; and the availability of social support [*see Figure 2*]. Under managed care, stroke rehabilitation is shifting from settings that provide acute rehabilitation to settings that provide subacute care, sometimes with a negative effect on outcome. Prospective payment in all settings has led to shorter duration of service, with unclear effects on recovery of function.[1,8,9] Community-based rehabilitation programs may be appropriate at some point in the recovery process for many persons with stroke and can lead to improved rates of personal care independence.[10]

Organized inpatient multidisciplinary stroke care is defined

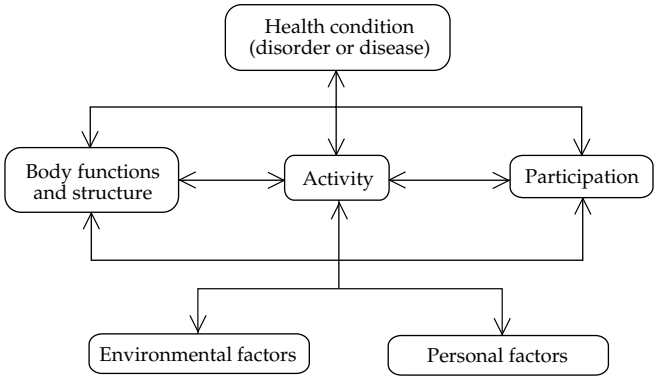


Figure 1 **Current understanding of interactions between the components of ICIDH-2.**[3]

Table 1 Selected Results of the Neurologic Examination of a Patient with Stroke[6]

Type of Deficit	*Tests Used*	*Key Findings*	*Effect of Persistent Deficit on Rehabilitation*
Altered level of consciousness	Repeated observation and testing of responses to external stimuli; Glasgow Coma Scale	Drowsiness, stupor, coma	An altered level of consciousness is a contraindication to rehabilitation
Cognitive deficits in higher functions, memory, ability to learn	Observation; questions to probe mental functions; standardized screening test	Various degrees and types of deficits	A severe deficit is a contraindication to rehabilitation; a moderate deficit may impede rehabilitation and must be incorporated into the rehabilitation management plan
Motor deficits	Tests of strength and tone in muscles of the upper and lower extremities and face	Various degrees and sites of weakness, incoordination, abnormal movements	Motor deficits are the primary indications for rehabilitation; absence of any voluntary movement is a sign of poor prognosis
Disturbances in balance and coordination	Tests of coordination, sitting, standing, walking	Various degrees and types of deficits	Deficits impede but are not a contraindication to rehabilitation
Somatosensory deficits	Specific tests for sensory modalities (e.g., pain, touch); complex sensory tests	Various degrees and types of deficits	Deficits impede but are not a contraindication to rehabilitation
Disorders of vision	Tests of pupillary responses, ocular motility, optic fundus, visual fields, acuity	Visual loss or field defect; conjugate gaze deficits	Severe visual loss or ocular motility disturbances impede rehabilitation
Unilateral neglect	Observation; description of complex picture by patient; sensory testing	One side of body or external environment is ignored (often clears spontaneously)	Neglect impedes but is not a contraindication to rehabilitation
Speech and language deficits	Observation of spontaneous speech and language use, including language comprehension and, if possible, simple reading and writing skills	Aphasia, dysarthria, apraxia of speech	Severe problems in communication impede rehabilitation; treatment becomes an integral part of rehabilitation
Swallowing disorder (dysphagia)	History; test of ability to swallow liquids and solids; cineradiography with barium swallow	Abnormal swallowing mechanism; aspiration	Dysphagia requires careful attention if aspiration and pneumonia are to be prevented
Affective disorder	History; observation; depression screening test	Symptoms of depression	Depression may impede rehabilitation if it is not treated
Pain	Description of pain by patient; observation of restrictions in range of motion; observation of facial expressions or resistance to movement	Location, severity, and precipitating causes of pain	Pain impedes rehabilitation and may require specific treatment or medication

as care from a team of physicians, nurses, and therapists whose work is coordinated through regular weekly meetings and is dedicated to rehabilitation. A systematic review has shown that organized inpatient multidisciplinary stroke care, compared with conventional inpatient medical care or multidisciplinary care in a different setting, reduces mortality, institutionalization, and dependency. For every 100 patients receiving organized inpatient multidisciplinary stroke rehabilitation, an additional five will return home in an independent state.[11] A large multisite observational study showed that adherence to Agency for Health Care Policy and Research (AHCPR) stroke rehabilitation care guidelines results in improved motor recovery and increased patient satisfaction.[12,13] A pooled analysis found that the use of care pathways to enhance adherence to guidelines appears to improve stroke recovery, but evidence is limited to observational studies.[14] Coordinated discharge planning with support for patients' needs after release can shorten length of stay without short-term adverse effects on disability or need for institutionalization. However, the effect of early discharge on long-term outcome is not known.[15]

Age itself does not influence recovery. Problems such as memory loss, diminished ability to follow instructions, urinary or fecal incontinence, and visuospatial deficits such as unilateral neglect (i.e., ignoring one side of the body or failing to be attentive to people and objects on one side of the body) are associated with poorer outcome.[16] The prognosis is also affected by coexisting or associated medical problems, such as delirium and fecal impaction. Simple assessment measures such as the Orpington Prognostic Scale [*see Figure 3*], which contains items for motor function, sensory function, balance, and cognition, can be completed in 5 to 10 minutes and are easily obtained in the first weeks after stroke.[7] The initial Orpington score is a strong predictor of functional status at 3 and 6 months. A score of 2.4 or less, suggesting mild deficits in the first 2 weeks after stroke, is associated with an 80% chance of being independent in personal care and homemaking activities at 6 months, whereas a score of 4.4 or higher, suggesting more severe early deficits, is associated with about a 20% chance of achieving independence in these activities.[17] Motor evoked potentials elicited by transcranial magnetic stimulation may be able to predict potential for motor recovery after stroke, according to preliminary studies.[18]

KEY ASPECTS OF REHABILITATION

Optimal rehabilitation for stroke starts in the acute care setting. Mobilization and efforts at self-care should be instituted as soon as they are medically feasible. Urinary continence and bladder function programs should be initiated; indwelling catheters should not be used unless they are medically necessary, as in cases of urinary retention or cases involving wounds that must be kept clean. An educational program aimed at informing the patient and the patient's family about stroke and stroke recovery should be started.

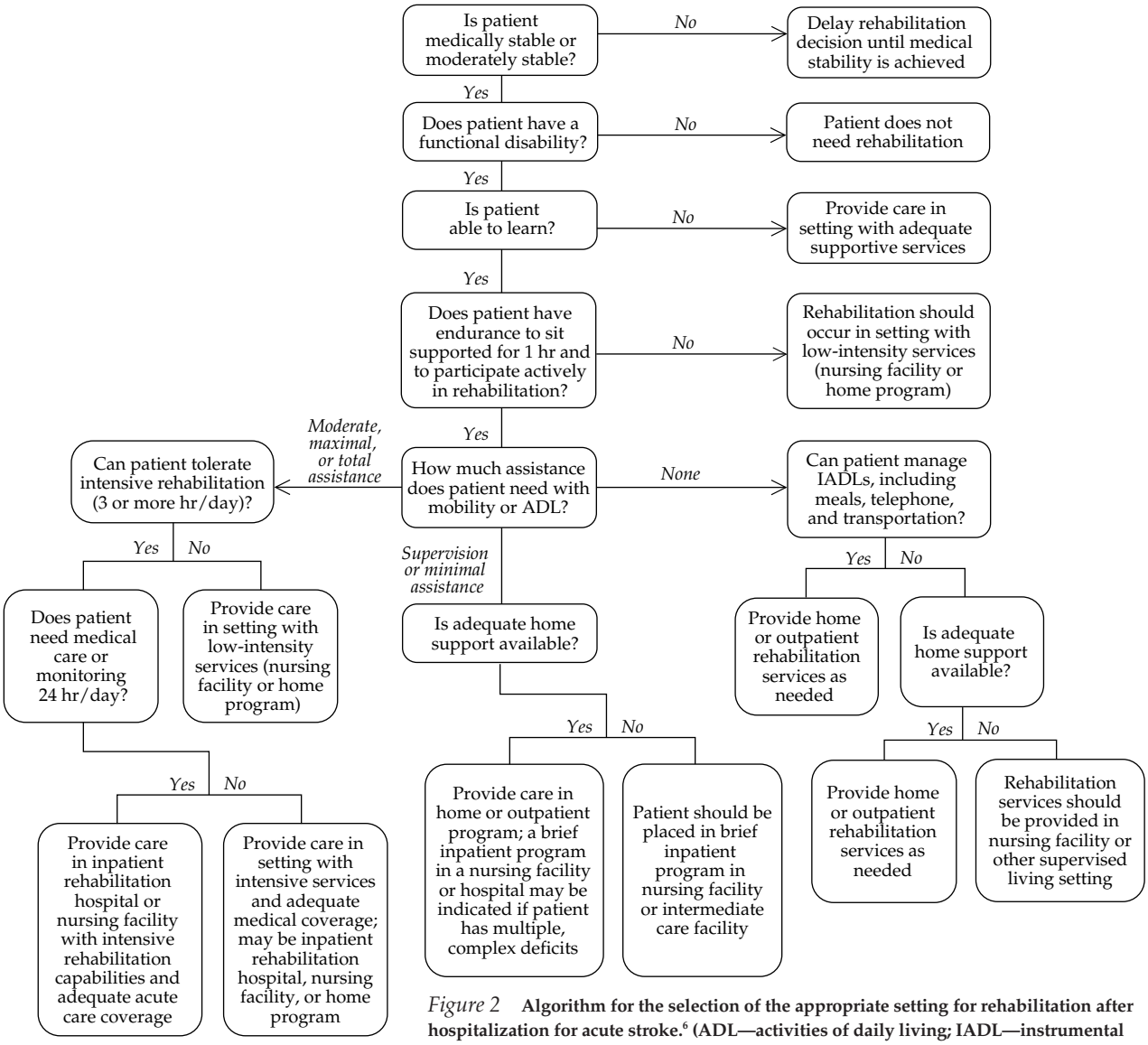


Figure 2 **Algorithm for the selection of the appropriate setting for rehabilitation after hospitalization for acute stroke.[6] (ADL—activities of daily living; IADL—instrumental activities of daily living)**

The next stage in rehabilitation is the defining of functional goals on the basis of the patient's own goals, deficits, and potential. Therapies are devised to ameliorate impairments in strength, range of motion, endurance, balance, sensorimotor coordination, and mobility; to facilitate self-care and household maintenance; to relieve pain; and to educate the patient and caregivers on maintaining functional gains. There are many proposed approaches to physiotherapy after stroke, and no single strategy is superior.[19] However, a clinical trial comparing usual home health care with an intensive home-based program that included endurance training, strengthening, balance, and upper extremity activities found greater gains in mobility, balance, and endurance in the intensive-therapy group.[20] High-intensity strength training improves functional performance after stroke.[21] Complex areas of cognition, such as attention, are affected by stroke[22]; formal cognitive rehabilitation interventions are under development, but results have been inconsistent.[23,24] Adaptive devices can help the patient eat, bathe, use the toilet, dress, walk, transfer between the bed and chairs and commode, and engage in leisure activities. Other aspects of treatment include further education and involvement of the patient and family and attention to the emotional and psychological sequelae of stroke.[25,26] Telephone support for caregivers may improve caregiver skills and satisfaction with the health care system.[27] Good coordination and communication must be maintained during transitions between rehabilitation settings.

Until recently, there was little scientific basis for stroke rehabilitation practice. Previously, it was believed that nerve cells were irreplaceable and that their death led inevitably to losses of neurologic function. Studies in animal models and in humans have now demonstrated, however, that after nerve cells die, their functions can be taken over by other nerve cells. Thus, after destruction of motor neurons controlling the hand, specific rehabilitation approaches can induce new motor neurons to control hand movement. Neuroplasticity in the adult after neurologic injury has been demonstrated by use of functional magnetic resonance imaging and transcranial magnetic stimulation.[28] The potential for neuroplastic change is revolutionizing approaches to neurologic rehabilitation.[29] Systematic studies of treatment approaches that maximize neural reorganization are

INSTRUCTIONS: Circle the appropriate response.

A. Motor deficit in arm
Lying supine, patient flexes shoulder to 90° and is given resistance.

0.0 = MRC Grade 5	(normal power)
0.4 = MRC Grade 4	(diminished power)
0.8 = MRC Grade 3	(movement against gravity)
1.2 = MRC Grade 1–2	(movement with gravity eliminated or trace)
1.6 = MRC Grade 0	(no movement)

B. Proprioception (eyes closed)
Locates affected thumb:
0.0 = Accurately
0.4 = Slight difficulty
0.8 = Finds thumb via arm
1.2 = Unable to find thumb

C. Balance
0.0 = Walks 10 ft without help
0.4 = Maintains standing position (unsupported for 1 min)
0.8 = Maintains sitting position
1.2 = No sitting balance

D. Cognition
Score 1 point for each correct answer:

_____ 1. Age of patient
_____ 2. Time (to the nearest hour)
I am going to give you an address, please remember it and I will ask you later: *42 West Street*
_____ 3. Name of hospital
_____ 4. Year
_____ 5. Date of birth of patient
_____ 6. Month
_____ 7. Year of the Second World War
_____ 8. Name of the President
_____ 9. Count backwards (20–1)
_____ 10. What is the address I asked you to remember: *42 West Street*

0.0 = Mental test score of 10
0.4 = Mental test score of 8–9
0.8 = Mental test score of 5–7
1.2 = Mental test score of 0–4

TOTAL SCORE:

1.6 + _____ + _____ + _____ + _____ = _____
Motor Proprioception Balance Cognition

Figure 3 **The Orpington Prognostic Scale is useful for estimating stroke recovery.[82]**

under way. One successful approach is based on intensive, repetitive practice using rhythm or robot aids for the upper extremity and treadmill training for the lower extremity.[30,31] Another method is to force the use of the damaged extremity by limiting the use of the uninvolved extremity, an approach termed constraint-induced movement therapy.[32]

The potential for neural reorganization is also changing the timing of neurologic rehabilitation. Previously, rehabilitation focused on the immediate poststroke period, because the natural history of recovery suggests that neurologic status plateaus 6 months after injury. However, studies have now documented significant gains in function in stroke patients with long-term impairment who undergo treatments that focus on repetitive and forced practice.[33-35]

Neuropharmacologic interventions that promote neural reorganization are being studied as a means to enhance neurologic recovery. One area of interest is the use of stimulants such as amphetamines, which have been shown to accelerate motor and sensory recovery in animal models. It is postulated that increased central noradrenergic—rather than serotoninergic or dopaminergic—mechanisms are the neurochemical basis for motor and sensory recovery secondary to treatment with amphetamines. Human studies have been small and have yielded inconsistent results.[36,37] A small trial has suggested that levodopa treatment after stroke may enhance motor recovery.[38]

PREVENTING COMPLICATIONS AND FUTURE STROKES

Stroke and other disabling conditions can have physical, functional, and psychological complications [*see Table 2*]. Mobilization of the patient is always preferable to inactivity; when in-

activity is inevitable, however, bed exercise and skin care protocols should be instituted. Prevention of deep vein thrombosis during immobilization is a standard component of care after stroke. Low-dose or low-molecular-weight heparin is the preferred treatment, but warfarin, intermittent pneumatic compression, and elastic stockings are acceptable alternatives [*see 1:XVIII Venous Thromboembolism*].[39]

In patients with a swallowing disorder, long-term enteral feeding and cessation of oral feeding may prevent aspiration; however, this measure is not always effective and has many ethical implications for quality of life. Incontinence may be exacerbated by one or more factors: a mobility problem that prevents the patient from getting to the bathroom, constipation, a urinary tract infection, or drug-induced urinary retention.[40] Stroke can cause an uninhibited bladder, characterized by frequent episodes of urgency and incontinence. Treatment with bladder relaxants may help, but these drugs can worsen confusion.

Seizures can occur at any time after a stroke as a result of scarred brain tissue, which serves as a focus of irritation. Standard anticonvulsants are indicated. Major depression occurs in at least one third of patients after stroke. The guidelines for poststroke rehabilitation emphasize the need to screen for and treat this commonly overlooked complication.[6] Increased socialization, counseling, and medication are appropriate. Prophylactic antidepressant medication has been proposed for the prevention of depression after stroke and is currently under study.[41,42] Cognitive impairment caused by poststroke depression may improve with treatment of the depression.[43] Preferred medications are those that cause less sedation and fewer anticholinergic side effects, such as most selective serotonin reuptake inhibitors or low-level anticholinergic tricyclic antidepressants such as nortriptyline. Methylphenidate, given in the morning and at noon, can reduce apathy and increase motivation early in poststroke rehabilitation. It has a rapid onset of effect [*see Table 3*].

Shoulder pain and dislocation are especially likely in the patient with a flaccid upper extremity, because the joint capsule is normally stabilized by the surrounding muscles. Measures to prevent dislocation include supporting the shoulder in a normal position, using a pull sheet rather than the underarm to reposition the patient in bed, and restricting range of motion to 90° of flexion and abduction. Reflex sympathetic dystrophy (chronic regional pain syndrome) occurs in as many as 25% of patients with hemiplegia. Early signs are exquisite cutaneous sensitivity and diffuse swelling of the hands. Treatment includes the use of compression gloves, anti-inflammatory agents, steroids, analgesics, injections for sympathetic nerve blockade, and, most important, aggressive and consistent range-of-motion exercises.[44] Spasticity is a common long-term complication of stroke. Because stroke is a form of upper motor neuron lesion, gradually increasing sensitivity of the lower motor neuron after stroke can lead to hypertonia, muscle spasm, pain, and functional limitations. Current treatments for spasticity (e.g., surgery or medications such as baclofen) have limited benefit. Botulinum toxin injections have been tested in clinical trials, with mixed results.[45-47]

Stroke recurs in 7% to 10% of survivors every year. Screening

Table 2 Assessment and Management of Complications of Stroke and Other Disabling Conditions[6]

Complication	*Cause*	*Assessment*	*Intervention*
Pressure sores	Immobility, malnutrition, incontinence	Monitor skin over pressure sites	Mobilize patient; provide protective bedding and pads; keep skin dry; maintain good nutrition and local skin care
Deep vein thrombosis	Inactivity, loss of muscle pump action in calf	Look for swelling and pain in leg; however, there are often no findings	Prevent with low-dose or low-molecular-weight heparin
Swallowing disorders	Cranial nerve dysfunction	Observe swallowing; listen for hoarse, wet voice; perform cineradiography	Reposition patient; alter food consistency; suggest chewing adaptations; switch to enteral feeding
Incontinence	Mobility problems, infection, impaction, drug side effects, urinary retention, outlet obstruction	Assess frequency and volume of incontinent episodes; assess postvoiding residual volume and ability to get to toilet; perform urinalysis and rectal examination; monitor drugs	Treat superimposed problems; institute toileting program; give oxybutynin for uninhibited bladder
Depression	Local brain effects, sensory isolation, situational factors	Administer depression screens	Give antidepressants
Shoulder pain	Loss of muscle tone around the shoulder, poor repositioning technique	Monitor symptoms and shoulder stability	Use careful positioning technique; give analgesics; conduct ROM exercises
Contractures	Increased muscle tone, immobility	Look for brisk DTRs, clonus, reduced ROM of joints	Conduct ROM exercises; use appropriate braces, heat/cold modalities, antispasticity drugs (baclofen), motor point blocks, botulinum injections, serial casting
Deconditioning	Immobility, depression, malnutrition	Test for reduced endurance and orthostatic hypotension; look for lack of motivation and poor exercise tolerance	Mobilize patient; conduct graded exercise program with scheduled rest breaks
Secondary osteoporosis	Dietary factors, smoking, immobility, medication	Test functional status; perform the following laboratory tests: vitamin D level, CBC, TFT, PTH level, 24 hr urine; measure bone mass	Modify diet; begin hormone therapy; conduct weight-bearing exercises and mobilize patient
Obesity	Dietary factors, immobility	> 20% of ideal body weight	Institute aerobic program; modify diet

CBC—complete blood count DTR—deep tendon reflex PTH—parathyroid hormone ROM—range of motion TFT—thyroid function tests

Table 3 Pharmacologic Therapy for Poststroke Depression

Drug Class	*Drug Name*	*Initial Dose*	*Maintenance Dose*	*Controlled Trial Evidence in Stroke*	*Side Effects*	*Advantages*	*Comments*
Heterocyclics	Nortriptyline	10–25 mg	25–100 mg	Yes	Sedation, orthostasis	Blood levels measurable; goal, 50–150 ng/ml	—
	Desipramine	10 mg	25–100 mg	No	Orthostasis	Less sedation; blood levels measurable; goal, 125–300 ng/ml	—
	Trazodone	25–50 mg	25–200 mg	Yes	Hypotension; sedation	Useful in sleep disturbance	—
Selective serotonin reuptake inhibitors	Fluoxetine	5–10 mg	5–60 mg	No	Nausea; tremor; insomnia	Long half-life	Morning dosing
	Sertraline	25 mg	—	No	Nausea; tremor; insomnia	Few drug interactions; not sedating	Morning dosing
	Citalopram	10 mg	20–40 mg	Yes	Nausea; tremor	Few drug interactions	Morning dosing
	Paroxetine	5–10 mg	5–40 mg	No	Nausea; tremor	Mild sedation	Evening dosing
Stimulants	Methylphenidate	2.5 mg morning and noon	5–30 mg	Yes	Nervousness; insomnia; anorexia	Rapid onset; useful in early period for treatment of apathy	Recommended for short-term use during rehabilitation only; change to other preparation for long-term use

and intervention to prevent recurrence must be considered in all patients except, perhaps, those with severe brain damage.[35] Carotid endarterectomy may be indicated if stenosis of 70% or more is detected in a vessel that feeds a large area of viable brain tissue. Anticoagulation with warfarin in patients with atrial fibrillation significantly reduces risk of stroke; in one meta-analysis, patients receiving warfarin had a third as many strokes as did control subjects.[48] Warfarin therapy is the standard of care for patients with atrial fibrillation who are not at increased risk of bleeding or falls. For an acceptable risk-to-benefit ratio, the international normalized ratio (INR) should be maintained at 2 to 3. Antiplatelet agents such as aspirin, ticlopidine, clopidogrel, and combinations of aspirin and dipyridamole can be used instead of warfarin. Control of hypertension and hyperlipidemia and smoking cessation reduce the risk of stroke.

Peripheral Vascular Disease and Amputation

REHABILITATION FOR PERIPHERAL VASCULAR DISEASE

Although little is known about the benefits of exercise in patients with critical leg ischemia, multiple prospective studies have demonstrated the effectiveness of exercise training in patients with claudication. Enrollment in a medically supervised training program, as well as treadmill testing before the initiation of training, is recommended because patients often have other comorbidities, such as coronary artery disease or diabetes, that put them at higher risk for an adverse event during exercise. The exercise training entails walking until the patient experiences a moderate degree of discomfort; after the onset of discomfort, the patient takes a short rest, then resumes walking. Three to five such sessions are conducted each week of training. The ultimate goal is for the patient to perform 50 minutes of intermittent walking. Improvement has been noted in pain-free walking time, maximal walking distance, and the ability to carry out routine daily activities.[49] Despite use of exercise, pharmacotherapy, angioplasty, and bypass surgery, 5% to 10% of patients with peripheral vascular disease will ultimately require amputation.

SELECTING CANDIDATES FOR PROSTHESES

About 50,000 lower-extremity amputations are performed each year in the United States. In the past, most amputations were performed because of trauma; today, as many as 90% are performed because of peripheral vascular disease. Lower-extremity ischemia often occurs in the setting of widespread atherosclerosis, which includes cerebrovascular, cardiovascular, and renovascular disease. Patients with diabetes are at increased risk for amputation because of vascular disease, peripheral neuropathy that causes the feet to become insensitive, and an increased vulnerability to infection. Thus, the candidate for amputation is very likely to have potentially disabling comorbid conditions.[50]

Prosthetic rehabilitation is most likely to succeed in patients who are ambulatory before undergoing amputation, can bear weight on the contralateral leg, have stabilized medical conditions, and are able to follow the caregiver's instructions. In patients with prostheses, the presence of a natural knee joint significantly reduces the energy cost of walking; therefore, every effort should be made to preserve the joint by performing a below-the-knee procedure, even though this procedure is associated with a higher risk of poor wound healing than above-the-knee amputation. The prognosis for successful prosthetic ambulation is best for patients who undergo unilateral below-the-knee amputation. Patients who undergo a second amputation or

those in whom other major disabling conditions develop are most likely to regain the ability to walk if they were able to walk after the initial amputation. Older patients who undergo bilater-al above-the-knee amputation rarely learn to walk with prostheses.

KEY ASPECTS OF REHABILITATION

Rehabilitation comprises three phases: preoperative preparation, postoperative wound healing, and prosthetic ambulation. Before undergoing amputation, the patient should be educated about the healing and mobilization process and be prepared for life as an amputee. After surgery, the primary care physician should focus attention on the common problems of the hospitalized older patient, including polypharmacy, immobility, delirium, and depression. Measures that promote proper wound healing should be instituted and the stump prepared for weight bearing and mobilization. Traditionally, weight bearing on the residual limb was discouraged until full healing took place; this sometimes resulted in prolonged inactivity for the patient. It is now recognized that earlier mobilization using a rigid, removable dressing and sometimes a temporary artificial leg can simultaneously protect the fragile healing tissues and prevent the complications caused by prolonged immobility.[50] Such complications include deconditioning and flexion contractures, which may limit the patient's success with a prosthesis.

Successful prosthetic ambulation depends on selection of an appropriate device, progressive mobilization, and management of concurrent problems. The permanent prosthesis for an older adult must be lightweight and easy to put on and take off. A comfortable prosthesis fits snugly and has one or more liners, socks, sleeves, or removable attachments. There are many options available for the components (materials, feet, and suspensions) of prostheses. The primary care physician should form a relationship with a reputable prosthetist who can integrate the technical issues involved in designing a prosthesis with the medical and functional status of the older amputee. The prosthetic limb should be adapted to existing comorbidity; for example, patients with heart failure or advanced renal disease often have wide variations in limb edema that must be accommodated by altering the size of the socket and by use of fillers. Progressive mobilization is accomplished by gradually lengthening the period during which the prosthesis is worn and changing the means of external support from parallel bars to walker to cane. Because increasing activity affects diabetes control, salt and water balance, myocardial oxygen demand, and local weight-bearing tissues, the physician should monitor the patient's glucose levels, weight, orthostatic blood pressure, and anginal symptoms and should adjust therapy as needed. The stump must be examined for signs of skin breakdown, edema, and infection.

PREVENTING COMPLICATIONS AND FUTURE AMPUTATIONS

About 25% of patients who undergo unilateral amputation because of peripheral vascular disease and 50% of patients who undergo the procedure because of diabetes will need to have the other leg amputated within 5 years.[51] Care of the contralateral extremity is essential; every amputee should have a program of regular foot care, should check for foot lesions, and should practice control of peripheral vascular disease. Footwear should protect and support the foot. Shoes should be wide at the toe and very roomy, and they should have cushioned soles or inserts. Many athletic shoes meet these criteria. Measures used to control vascular disease include cessation of smoking, control of diabetes, management of cholesterol levels, and a program of exercise of the lower extremity.[52] Invasive treatment of claudication can either delay or prevent the development of critical ischemia, which may lead to a decreased rate of amputation; however, further study of this approach is needed.[53,54]

Hip Fracture

CONSEQUENCES OF HIP FRACTURE

Each year in the United States, hip fracture occurs in about 250,000 persons, of whom 75% are women. Age-adjusted rates of hip fracture are higher in nursing home residents and patients with dementia.

Hip fracture in elderly patients results in increased mortality and loss of functional independence. One-year mortality is nearly 25%.[55] In those who survive, early decline in function is common; up to 50% require temporary nursing home stays. Although most survivors return to their original functional status within 1 year, about 25% require long-term care.[56]

TYPES OF HIP FRACTURE AND REPAIR

Hip fractures occur at the femoral neck in about one third of cases. Such fractures are likely to disrupt the blood supply to the femoral head, particularly if the fractures are displaced; this in turn may lead to avascular necrosis or nonunion. Nondisplaced fractures of the femoral neck are usually treated by internal fixation with pins or nails, whereas displaced fractures may be treated either by reduction and fixation or by hemiarthroplasty with a prosthetic femoral head. Selected patients with significant underlying bony acetabular disease may benefit from complete hip arthroplasty for displaced fracture of the femoral neck.

Two thirds of hip fractures occur across the trochanter. Intertrochanteric fractures are often associated with significant bleeding into the surrounding soft tissue. Surgical treatment usually consists of open reduction and internal fixation with a variety of mechanisms, such as compression screws or sliding nails.

Preoperative stabilization of concurrent medical problems and early surgical intervention is usually recommended. Complications such as deep vein thrombosis, pneumonia, pressure sores, urinary tract infection, malnutrition, delirium, and deconditioning should be prevented or treated.[57] Delirium has been shown to be an independent predictor of poor outcome up to at least 1 year after injury.[58]

KEY ASPECTS OF REHABILITATION

Rehabilitation should be offered to all patients in the absence of near-terminal conditions or possibly to bedridden patients with end-stage dementia. Such patients may be treated nonsurgically with early mobilization from bed to chair, control of pain, and treatment of complications. Patients with mild to moderate dementia can benefit from rehabilitation after fracture.[59,60]

After hip fracture, the care setting may not influence outcome as much as it does in stroke; in an observational study, outcomes in settings involving intensive rehabilitation were not better than outcomes in subacute programs.[61,62] Well-defined home-based programs can be effective.[63] Coordinated multidisciplinary approaches to inpatient rehabilitation of older patients have been found to have borderline effectiveness, with about a 10% reduction in combined outcomes such as death or institutionalization.[64]

Early intervention with early surgery, minimal narcotic analgesia, intense daily therapy, and multidisciplinary management reduced length of stay without affecting function or survival.[65]

The goal of rehabilitation is to regain prefracture function. Short-term goals include pain control, prevention of medical complications, maintenance of range of motion and muscle strength in other joints, early mobilization, and gradual improvement in the movement of the affected hip. Early mobilization reduces all the complications of immobility, including bedsores, constipation, loss of strength, and risk of thromboembolism. If a prosthetic femoral head is placed by the posterior approach, the risk of dislocation is reduced by the intentional limitation of hip motion to 90° of flexion, with no internal rotation and no adduction. Shortening of the affected leg can occur after fracture, resulting in an abnormal gait, which can be corrected with shoe lifts.

Weight bearing after hip fracture is controversial. Surgical factors and surgeons' preferences affect recommendations. Pain, the hazards of inactivity, and the practical difficulties in limiting weight bearing must also be considered. In one study, both weight-bearing and non–weight-bearing exercises during hospitalization after a hip fracture resulted in similar improvements in strength, balance, and functional performance.[66] Early mobilization with unrestricted weight bearing after hip fracture results in mild spontaneous weight shifting and accelerates hospital discharge without adversely affecting healing or putting the patient at higher risk for additional surgery.[67,68] A synthesis of clinical trials comparing mobilization strategies after hip fracture, including twice-daily physical therapy, treadmill training, neuromuscular stimulation, and early mobilization, found no significant effects on the outcomes assessed.[69]

Pharmacologic approaches to improving rehabilitation outcomes in hip fracture are also under study. A clinical trial of human growth hormone in the immediate postfracture phase demonstrated reduced functional decline and increased rates of return to independent living in the subset of patients older than 75 years[70]; however, another large trial found that 6 months of oral growth hormone therapy did not improve physical function.[71]

PREVENTING FUTURE FRACTURES

Efforts should be made to prevent future fractures. Preventive practices include treatment of osteoporosis [*see 3:VI Diseases of Calcium Metabolism and Metabolic Bone Disease*], evaluation and management of any tendency toward falling,[72] and instruction on how to avoid injury in the event of a fall.[57] Higher levels of physical activity have been shown to be associated with lower rates of hip fracture in epidemiologic and case-control studies. As yet, no prospective, randomized clinical trials of physical activity as a preventive approach to hip fracture have been carried out. Hip protectors have been extensively studied, with varied results. For patients living in the community, hip protectors do not appear to decrease the incidence of hip fractures. However, for those living in an institutional setting where the incidence of fracture is high, hip protectors may offer some benefit. Compliance remains an issue.[73]

Rheumatoid Arthritis

Management of rheumatoid arthritis involves pain relief, preservation of strength and joint function, and prevention of deformities [*see 15:II Rheumatoid Arthritis*]. Patients with arthritis can safely participate in exercise programs and often experience relief of pain and of disability as a result.[74]

Internet Resources Relevant to Geriatric Rehabilitation

National Stroke Association
www.stroke.org

American Stroke Association
www.strokeassociation.org

Center on Aging
Stroke toolbox
www2.kumc.edu/coa/Stroke_Toolbox/stroke-tool.htm

National Institute of Neurological Disorders and Stroke
Stroke information page
www.ninds.nih.gov/health_and_medical/disorders/stroke.htm

National Aphasia Association
www.aphasia.org

National Center for Injury Prevention and Control
Falls and hip fractures among older adults
www.cdc.gov/ncipc/factsheets/falls.htm

Preventing Falls and Fractures
http://home.nyc.gov/html/dfta/pdf/fallsandfractures.pdf

Family Caregivers Online
Home safety for elders
www.familycaregiversonline.com/family_caregiver_module_6.htm

The International Center for Disability Resources on the Internet
Main Stroke Page
www.icdri.org/stroke/stroke_main.htm

Osteoporosis and Related Bone Diseases
National Resource Center
www.osteo.org

American Academy of Orthopaedic Surgeons
Links to patient education booklets, including exercise guides after total hip replacement, total knee replacement, and other procedures.
http://orthoinfo.aaos.org

Missouri Arthritis Rehabilitation Research and Training Center
www.hsc.missouri.edu/arthritis

Resources for Rehabilitation
Publications to help people with disabilities and chronic conditions remain independent
www.rfr.org

National Institute of Arthritis and Musculoskeletal and Skin Diseases
www.nih.gov/niams

Arthroplasty

Each year, over 100,000 Americans undergo total hip replacement (THR) and over 50,000 have a total knee replacement (TKR). Joint replacement should be considered in persons with structural damage to the joint who experience pain and loss of function despite nonsurgical management. Most candidates have osteoarthritis, are older than 60 years, and have multiple other medical conditions.

Surgical interventions for uncontrolled knee pain and mobility limitation include lavage with or without debridement, osteotomy, and knee replacement. Lavage can produce significant temporary relief. Osteotomy is preferred in active persons younger than 60 years. Knee replacement can be total or limited to one compartment. Joint replacement relieves pain and improves function in most patients, irrespective of their age. A retrospective cohort study examined complication rates, length of stay in acute care facilities, pain scores, and functional abilities

for selected patients older than 80 years who underwent either a hip or knee arthroplasty. Joint replacement in this selected population led to greatly reduced pain and increased function. Given that surgical risk appears to be tolerable in the increasing population of persons older than 80 years who are in good health with stable chronic conditions, age alone should not be used as a criterion for eligibility for joint replacement.[75]

KEY ASPECTS OF REHABILITATION

Hip

Good pain control and anticoagulation with warfarin to prevent thrombosis are the major goals of the immediate THR postoperative period. Weight bearing often begins on the second postoperative day. The rehabilitation program emphasizes progressive range of motion, strengthening, and gait training. Hip abductors, which are often weak as a direct result of surgery, are a focus of the strengthening program. Patients are taught to avoid motions that increase risk of dislocation; such motions include deep squatting and crossing the knees. A raised toilet seat is recommended for use during the first months after surgery to prevent excess hip flexion. The appropriate site and the duration of rehabilitative services are determined on the basis of medical and functional status, as well as on the availability of caregivers. Many low-risk patients can be discharged from the acute care hospital within 5 days.[76] Patients at higher risk (i.e., those older than 70 years or those with two or more comorbid conditions) can benefit from inpatient rehabilitation, beginning as early as the third postoperative day, and can experience faster recovery of mobility and reduced total length of stay.[77]

Knee

Initial goals after TKR are pain control, wound drainage, and joint stabilization. The patient is usually allowed to bear weight with a straight leg by the second postoperative day. Improved range of motion is a critical step in recovery and is often aided by the use of a continuous passive motion (CPM) machine. Early postoperative CPM combined with physical therapy (PT) has been shown to be more effective than PT alone in increasing active range of motion and shortening length of stay. CPM also decreases the need for postoperative manipulation.[78] Therapy with a CPM machine involves gradually increasing passive range of motion in the operated knee over a period of several days. Good recovery of knee motion is represented by full extension and at least 90º of flexion. In a clinical trial, home CPM produced satisfactory range of motion at about half the cost of home physical therapy.[79] Strength training is often deferred for several weeks to promote stable healing of tissues. Isometric and resistive exercise with gradually increasing loads generally can be introduced safely by 8 weeks after surgery. Low-risk TKR patients appear to do well with discharge from acute care on the fourth postoperative day, and high-risk patients benefit from early transfer to intensive rehabilitation.[76,77] Long-term outcomes include significant relief of pain and improved function,[80] although many patients do not achieve levels of strength or mobility comparable to those of age-matched control subjects.[81]

Additional Information

Additional information on topics discussed in this chapter can be found at various Internet sites [*see Sidebar,* Internet Resources Relevant to Geriatric Rehabilitation].

Stephanie Studenski, M.D., M.P.H., has received grants for clinical research from Eli Lilly and Company during the past 12 months.

Cynthia J. Brown, M.D., P.T., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Amphetamine, methylphenidate, and levodopa, which are discussed in this chapter, have not been approved by the FDA for use in promoting stroke recovery. Oral human growth hormone, which is discussed in this chapter, is experimental and has not been approved by the FDA.

References

1. McCall N, Korb J, Petersons A, et al: Reforming Medicare payment: early effects of the 1997 Balanced Budget Act on postacute care. Milbank Q 81:277, 2003

2. Hoenig H, Nusbaum N, Brummel-Smith K: Geriatric rehabilitation: state of the art. J Am Geriatr Soc 45:1371, 1997

3. ICIDH-2. International Classification of Functioning, Disability and Health. Classification, Assessment, Surveys and Terminology Team, World Health Organization, Geneva, Switzerland, December 2000

4. Siebens H, Aronow H, Edwards D, et al: A randomized controlled trial of exercise to improve outcomes of acute hospitalization in older adults. J Am Geriatr Soc 48:1545, 2000

5. Gill TM, Baker DI, Gottschalk M, et al: A program to prevent functional decline in physically frail, elderly persons who live at home. N Engl J Med 347:1068, 2002

6. Gresham GE, Duncan PW, Stason WB, et al: Post-stroke rehabilitation (Clinical Practice Guideline, No. 16). US Dept of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research (AHCPR Publication No. 95-0662), Rockville, Maryland, May 1995

7. Duncan PW, Lai SM, van Culin V, et al: Development of a comprehensive assessment toolbox for stroke. Clin Geriatr Med 15:885, 1999

8. Kramer AM, Kowalsky JC, Lin M, et al: Outcome and utilization differences for older persons with stroke in HMO and fee-for-service systems. J Am Geriatr Soc 48:726, 2000

9. Yip JY, Wilber KH, Myrtle RC: The impact of the 1997 Balanced Budget Amendment's prospective payment system on patient case mix and rehabilitation utilization in skilled nursing. Gerontologist 42:653, 2002

10. Therapy-based rehabilitation services for stroke patients at home. Outpatient Service Trialists. Cochrane Database Syst Rev (1):CD002925, 2003

11. Langhorne P, Duncan P: Does the organization of postacute stroke care really matter? Stroke 32:268, 2001

12. Hoenig H, Duncan PW, Horner RD, et al: Structure, process, and outcomes in stroke rehabilitation. Med Care 40:1036, 2002

13. Reker DM, Duncan PW, Horner RD, et al: Postacute stroke guideline compliance is associated with greater patient satisfaction. Arch Phys Med Rehabil 83:750, 2002

14. Kwan J, Sandercock P: In-hospital care pathways for stroke. Cochrane Database Syst Rev (2):CD002924, 2002

15. Services for reducing duration of hospital care for acute stroke patients. Early Supported Discharge Trialists. Cochrane Database Syst Rev (1):CD000443, 2002

16. Galski T, Bruno RL, Zorowitz R, et al: Predicting length of stay, functional outcome, and aftercare in the rehabilitation of stroke patients: the dominant role of higher-order cognition. Stroke 24:1794, 1993

17. Studenski S, Wallace D, Duncan PW, et al: Predicting stroke recovery: three- and six-month rates of patient-centered functional outcomes based on the Orpington Prognostic Scale. J Am Geriatr Soc 49:308, 2001

18. Hendricks HT, Zwarts MJ, Plat EF, et al: Systematic review for the early prediction of motor and functional outcome after stroke by using motor-evoked potentials. Arch Phys Med Rehabil 83:1303, 2002

19. Pollock A, Baer G, Pomeroy V, et al: Physiotherapy treatment approaches for the recovery of postural control and lower limb function following stroke. Cochrane Database Syst Rev (2):CD001920, 2003

20. Duncan P, Studenski S, Richards L, et al: Randomized clinical trial of therapeutic exercise in subacute stroke. Stroke 34:2173, 2003

21. Weiss A, Suzuki T, Bean J, et al: High intensity strength training improves strength and functional performance after stroke. Am J Phys Med Rehabil 79:369, 2000

22. McDowd JM, Filion DL, Pohl PS, et al: Attentional abilities and functional outcomes following stroke. J Gerontol B Psychol Sci Soc Sci 58:P45, 2003

23. Cuesta GM: Cognitive rehabilitation of memory following stroke: theory, practice, and outcome. Adv Neurol 92:415, 2003

24. Bowen A, Lincoln NB, Dewey M: Cognitive rehabilitation for spatial neglect following stroke. Cochrane Database Syst Rev (2):CD003586, 2002

25. Evans RL, Hendricks RD, Haselkorn JK, et al: The family's role in stroke rehabilitation: a review of the literature. Am J Phys Med Rehabil 71:135, 1992

26. Downhill JE Jr, Robinson RG: Longitudinal assessment of depression and cognitive impairment following stroke. J Nerv Ment Dis 182:425, 1994

27. Grant JS, Elliott TR, Weaver M, et al: Telephone intervention with family caregivers of stroke survivors after rehabilitation. Stroke 33:2060, 2002

28. Liepert J, Bauder H, Wolfgang HR, et al: Treatment-induced cortical reorganization

after stroke in humans. Stroke 31:1210, 2000

29. Wittenberg GF, Chen R, Ishii K, et al: Constraint-induced therapy in stroke: magnetic-stimulation motor maps and cerebral activation. Neurorehabil Neural Repair 17:48, 2003

30. Volpe BT, Krebs HI, Hogan N: Robot-aided sensorimotor training in stroke rehabilitation. Adv Neurol 92:429, 2003

31. Hesse S, Werner C, von Frankenberg S, et al: Treadmill training with partial body weight support after stroke. Phys Med Rehabil Clin North Am 14(1 suppl):S111, 2003

32. Dromerick AW, Edwards DF, Hahn M: Does the application of constraint-induced movement therapy during rehabilitation reduce arm impairment after ischemic stroke? Stroke 31:2984, 2000

33. Taub E, Miller NE, Novack TA, et al: Technique to improve chronic motor deficit after stroke. Arch Phys Med Rehabil 74:347, 1993

34. Whitall J, McCombe Waller S, Silver KH, et al: Repetitive bilateral arm training with rhythmic auditory cueing improves motor function in chronic hemiparetic stroke. Stroke 31:2390, 2000

35. van der Lee JH, Wagenaar RC, Lankhorst GJ, et al: Forced use of the upper extremity in chronic stroke patients: results of a single-blind randomized clinical trial. Stroke 30:2369, 1999

36. Goldstein LB: Pharmacotherapy in stroke rehabilitation. Adv Neurol 92:447, 2003

37. Sonde L, Nordstrom M, Nilsson CG, et al: A double-blind placebo-controlled study of the effects of amphetamine and physiotherapy after stroke. Cerebrovasc Dis 12:253, 2001

38. Scheidtmann K, Fries W, Muller F, et al: Effect of levodopa in combination with physiotherapy on functional motor recovery after stroke: a prospective, randomised, double-blind study. Lancet 358:787, 2001

39. Kamran SI, Downey D, Ruff RL: Pneumatic sequential compression reduces the risk of deep vein thrombosis in stroke patients. Neurology 50:1683, 1998

40. Gelber DA, Good DC, Laven LJ, et al: Causes of urinary incontinence after acute hemispheric stroke. Stroke 24:378, 1993

41. Whyte EM, Mulsant BH: Post stroke depression: epidemiology, pathophysiology, and biological treatment. Biol Psychiatry 52:253, 2002

42. Rasmussen A, Lunde M, Poulsen DL, et al: A double-blind, placebo-controlled study of sertraline in the prevention of depression in stroke patients. Psychosomatics 44:216, 2003

43. Kimura M, Robinson RG, Kosier JT: Treatment of cognitive impairment after post-stroke depression: a double-blind treatment trial. Stroke 31:1482, 2000

44. Stanton-Hicks M, Janig W, Hassenbusch S, et al: Reflex sympathetic dystrophy: changing concepts and taxonomy. Pain 63:127, 1995

45. Brashear A, McAfee AL, Kuhn ER, et al: Treatment with botulinum toxin type B for upper-limb spasticity. Arch Phys Med Rehabil 84:103, 2003

46. Intramuscular injection of botulinum toxin for the treatment of wrist and finger spasticity after a stroke. Botox Post-Stroke Spasticity Study Group. N Engl J Med 347:395, 2002

47. Rowland LP: Stroke, spasticity, and botulinum toxin. N Engl J Med 347:382, 2002

48. Matchar DB, McCrory DC, Barnett HJM, et al: Medical treatment for stroke prevention. Ann Intern Med 121:41, 1994

49. Stewart KJ, Hiatt WR, Regenstein JG, et al: Exercise training for claudication. N Engl J Med 347:1941, 2002

50. Cutson T, Bongiorni D: Rehabilitation of the older lower limb amputee: a brief review. J Am Geriatr Soc 44:1388, 1996

51. Coletta EM: Care of the elderly patient with lower extremity amputation. J Am Board Fam Pract 13:23, 2000

52. Gardner AW, Poehlman ET: Exercise rehabilitation programs for the treatment of claudication pain: a meta-analysis. JAMA 274:975, 1995

53. Holtzman J, Caldwell M, Walvatne C, et al: Long-term functional status and quality of life after lower extremity revascularization. J Vasc Surg 29:395, 1999

54. Jamsen TS, Manninen HI, Tulla HE, et al: Infrainguinal revascularization because of claudication: total long-term outcome of endovascular and surgical treatment. J Vasc Surg 37:808, 2003

55. Braithwaite RS, Col NF, Wong JB: Estimating hip fracture morbidity, mortality and costs. J Am Geriatr Soc 51:364, 2003

56. Egol KA, Koval KJ, Zuckerman JD: Functional recovery following hip fracture in the elderly. J Orthop Trauma 11:594, 1997

57. Morrison RS, Chassin MR, Siu AL: The medical consultant's role in caring for patients with hip fracture. Ann Intern Med 128:1010, 1998

58. Marcantonio ER, Flacker JM, Michaels M, et al: Delirium is independently associated with poor functional recovery after hip fracture. J Am Geriatr Soc 48:618, 2000

59. Goldstein FC, Strasser DC, Woodard JL, et al: Functional outcome of cognitively impaired hip fracture patients on a geriatric rehabilitation unit. J Am Geriatr Soc 45:35, 1997

60. Huusko TM, Karppi P, Avikainen V, et al: Randomised, clinically controlled trial of intensive geriatric rehabilitation in patients with hip fracture: subgroup analysis of patients with dementia. BMJ 321:1107, 2000

61. Kramer AM, Steiner JF, Schlenker RE, et al: Outcomes and costs after hip fracture and stroke. JAMA 277:396, 1997

62. Roder F, Schwab M, Aleker T, et al: Proximal femur fracture in older patients: rehabilitation and clinical outcome. Age Ageing 32:74, 2003

63. Tinetti ME, Baker DI, Gottschalk M, et al: Systematic home-based physical and functional therapy for older persons after hip fracture. Arch Phys Med Rehabil 78:1237, 1997

64. Cameron ID, Handoll HH, Finnegan TP, et al: Co-ordinated multidisciplinary approaches for inpatient rehabilitation of older patients with proximal femoral fracture. Cochrane Database Syst Rev (3):CD000106, 2001

65. Swanson CE, Day GA, Yelland CE, et al: The management of elderly patients with femoral fractures: a randomised controlled trial of early intervention versus standard care. Med J Aust 169:515, 1998

66. Sherrington C, Lord SR, Herbert RD: A randomized trial of weight-bearing versus non–weight-bearing exercise for improving physical ability in inpatients after hip fracture. Aust J Physiother 49:15, 2003

67. Koval KJ, Friend KD, Aharonoff GB, et al: Weight bearing after hip fracture: a prospective series of 596 geriatric hip fracture patients. J Orthop Trauma 10:526, 1996

68. Koval KJ, Sala DA, Kummer FJ, et al: Postoperative weight-bearing after a fracture of the femoral neck or an intertrochanteric fracture. J Bone Joint Surg Am 80:352, 1998

69. Handoll HH, Parker MJ, Sherrington C: Mobilisation strategies after hip fracture surgery in adults. Cochrane Database Syst Rev (1):CD001704, 2003

70. Van Der Lely AJ, Lamberts SW, Jauch KW, et al: Use of human GH in elderly patients with accidental hip fracture. Eur J Endocrinol 143:585, 2000

71. Bach M, Rockwood K, Zetterberg C, et al: The effects of MK-0677, an oral growth hormone secretogogue, in patients with hip fracture. J Am Geriatr Soc (in press)

72. Gillespie L, Gillespie W, Robertson M, et al: Interventions for preventing falls in elderly people. Cochrane Database Syst Rev (4):CD000340, 2003

73. Parker MJ, Gillespie LD, Gillespie WJ: Hip protectors for preventing hip fractures in the elderly. Cochrane Database Syst Rev (3):CD001255, 2003

74. Ettinger WH Jr: Physical activity, arthritis, and disability in older people. Clin Geriatr Med 14:633, 1998

75. Brander VA, Malhotra S, Jet J, et al: Outcome of hip and knee arthroplasty in persons aged 80 and over. Clin Orthop 345:67, 1997

76. Weingarten S, Riedinger MS, Sandhu M, et al: Can practice guidelines safely reduce hospital length of stay? Results from a multicenter interventional study. Am J Med 105:33, 1998

77. Munin MC, Rudy TE, Glynn NW, et al: Early inpatient rehabilitation after elective hip and knee arthroplasty. JAMA 279:847, 1998

78. Milne S, Brosseau L, Robinson V, et al: Continuous passive motion following total knee arthroplasty. Cochrane Database Syst Rev (2):CD004260, 2003

79. Worland RL, Arredondo J, Angles F, et al: Home continuous passive motion machine versus professional physical therapy following total knee replacement. J Arthroplasty 13:784, 1998

80. Hawker G, Wright J, Coyte P, et al: Health-related quality of life after knee replacement. J Bone Joint Surg Am 80:163, 1998

81. Walsh M, Woodhouse LJ, Thomas SG, et al: Physical impairments and functional limitations: a comparison of individuals 1 year after total knee arthroplasty with control subjects. Phys Ther 78:248, 1998

82. Lai S, Duncan PW, Keighley J: Prediction of functional outcome after stroke. Stroke 29:1838, 1998

Acknowledgment

Figure 2 Marcia Kammerer.

II DIAGNOSIS AND TREATMENT OF DYSLIPIDEMIA

JOHN D. BRUNZELL, M.D., F.A.C.P.
R. ALAN FAILOR, M.D.

Disorders of lipoprotein metabolism, in conjunction with the prevalence of high-fat diets, obesity, and physical inactivity, have resulted in an epidemic of atherosclerotic disease in the United States and other developed countries. The interaction of common genetic and acquired disorders of lipoproteins with these adverse environmental factors leads to the premature development of atherosclerosis. In the United States, mortality from coronary artery disease (CAD), particularly in persons younger than 60 years, has been declining since 1970; however, atherosclerotic cardiovascular disease remains the most common cause of death among both men and women.

Formerly, hyperlipidemia was defined as elevation of a lipoprotein level in the population. The recognition that a low level of high-density lipoprotein (HDL) and the presence of small, dense low-density lipoprotein (LDL) are clinically important in the pathophysiology of lipid disorders has led to the use of the term dyslipidemia to describe a range of disorders that include both abnormally high and low lipoprotein levels, as well as disorders in the composition of these particles. Dyslipidemias are clinically important, principally because of their contribution to atherogenesis. Pancreatitis and fatty liver disease are less common but clinically significant manifestations of lipid disorders.

Lipoprotein Physiology

LIPOPROTEIN COMPOSITION AND METABOLISM

Lipoproteins are spherical macromolecular complexes of lipid and protein [*see Figure 1*]. Clinically important lipids in the blood include cholesterol (both unesterified and esterified) and triglyceride (molecules consisting of three fatty acids attached to a glycerol backbone). Cholesterol has three primary functions: it plays a role in the structure of cell membranes, in the synthesis of steroid hormones, and in the formation of bile acids. The major functions of triglyceride are energy storage (in fat) and energy use (by muscle). Because fat cannot readily dissolve in plasma, cholesterol and triglyceride are made miscible by incorporation into lipoproteins (e.g., very low density lipoprotein [VLDL], LDL, and HDL). Apolipoproteins are the protein component of lipoproteins; they aid in the lipid transport and delivery process in three ways: they serve as structural elements, as ligands for receptors, and as regulatory cofactors [*see Table 1*].

LIPOPROTEIN STRUCTURE AND CLASSIFICATION

A mature lipoprotein particle is a sphere consisting of a central core of lipids (triglyceride and cholesteryl ester) surrounded by a monolayer surface of phospholipid, unesterified cholesterol, and apolipoproteins [*see Figure 1*]. Operationally, the lipoproteins can be described on the basis of their size and buoyancy characteristics [*see Figure 2*].

Chylomicron Chylomicrons are the largest of the lipoprotein particles. The major structural protein is apolipoprotein B-48 (apo B-48). The bulk (~80%) of the lipid core consists of triglyceride. Synthesized and secreted from the intestine, chylomicrons transport exogenous cholesterol, fatty acids, and fat-soluble vitamins absorbed from digested food [*see* Exogenous Pathway, *below*].

VLDL This triglyceride-rich particle (~80% of the lipid core consists of triglyceride) is synthesized in the liver, delivers triglyceride to the periphery, and is the precursor for intermediate-density lipoproteins (IDLs) and LDL. The major structural protein of this lipoprotein is apo B-100 [*see* Endogenous Pathway, *below*].

IDL The remnant of VLDL is of IDL density. It is formed after triglyceride in VLDL is hydrolyzed by lipoprotein lipase. The core is roughly 50% triglyceride and 50% cholesteryl ester. Approximately half of the body's IDL particles are cleared from the plasma into the liver; the other half are further processed to form LDL [*see* Endogenous Pathway, *below*]. In clinical practice, assessment of LDL levels includes the determination of cholesterol in both IDL and LDL fractions.

LDL This lipoprotein results from the hepatic processing of VLDL remnants. The core is rich in cholesteryl ester and accounts for the majority of cholesterol circulating in the blood. LDL plays a major role in the development of atherosclerosis [*see* LDL Catabolism, *below*].

HDL HDL forms from the unesterified cholesterol and phospholipid removed from peripheral tissues and the surface of triglyceride-rich proteins [*see* Function and Regulation of HDL, *below*]. The major structural protein is apo A-I; the core is predominantly cholesteryl ester. HDL mediates the return of lipoprotein and tissue cholesterol to the liver for excretion in the process referred to as reverse cholesterol transport. Another of its functions is to shuttle apo E and apo C-II to and from chylomicrons and VLDL.

LIPOPROTEIN ASSEMBLY AND CATABOLISM

Exogenous Pathway

After a meal, intestinal cells absorb fatty acids and cholesterol, esterify them into triglyceride and cholesteryl ester, and incorporate them into the core of chylomicrons.[1] Triglyceride greatly predominates over cholesterol ester in the chylomicron core. The chylomicrons are secreted into plasma, where apo C-II on the chylomicron surface activates endothelial-bound lipoprotein lipase (LPL). LPL in turn hydrolyzes the chylomicron's core triglyceride and releases free fatty acids, which are taken up by adipose tissue for storage and by muscle for energy. During lipolysis, the chylomicron decreases in size, and some surface components are transferred to HDL; the remaining particle is the chylomicron remnant particle. This chylomicron remnant next acquires apo E from HDL and is subsequently taken up by the liver after binding to sites that recognize apo E. It is then degraded, thereby delivering dietary cholesterol to the liver.

Endogenous Pathway

The liver secretes triglyceride-rich VLDL into plasma, where they too acquire apo C-II from HDL. As with chylomicrons,

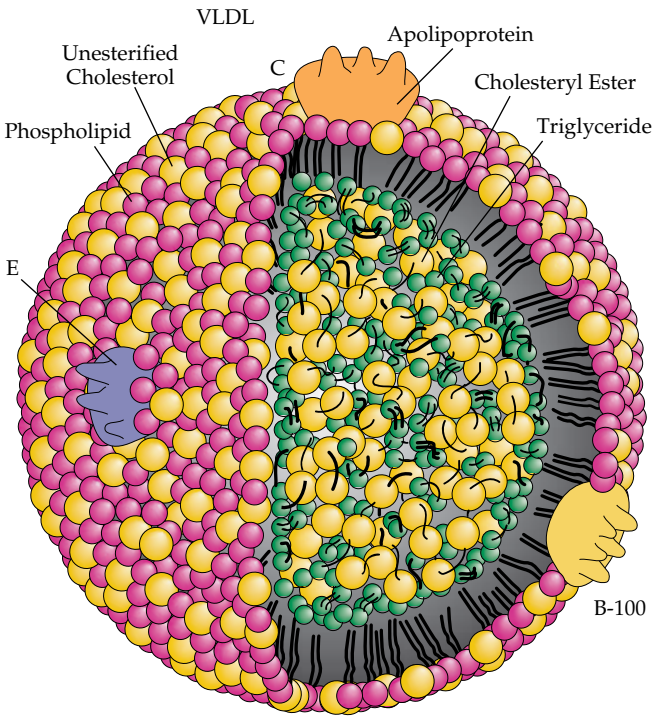


Figure 1 **Lipoproteins transport water-insoluble triglyceride and cholesterol through the bloodstream. All apo B–containing lipoproteins have a structure similar to that shown for very low density lipoproteins (VLDL). The core is composed of triglyceride and cholesteryl ester, whereas the monolayer surface is composed of phospholipid, unesterified cholesterol, and protein in the form of apolipoproteins. VLDL contains apolipoproteins B-100, C-I, C-II, and E. Low-density lipoprotein (LDL), which transports most of the cholesterol found in blood, contains primarily apo B-100.**

VLDL interacts with LPL on the capillary endothelium, and the core triglyceride is hydrolyzed to provide fatty acids to adipose and muscle tissues.[1] About half of the catabolized VLDL remnants (IDL density) are taken up by hepatic receptors that bind to apo E for degradation; the other half—apo B-100 particles, depleted of triglyceride relative to cholesteryl ester— are converted by the liver to cholesteryl ester-rich LDL. As IDL is converted to LDL, apo E becomes detached, leaving only one apolipoprotein, apo B-100. Each particle in this cascade from VLDL to LDL contains one molecule of apo B-100.

In the metabolism of both chylomicrons and VLDL, apo C-II permits the hydrolysis of triglyceride by lipoprotein lipase, and apo E prompts hepatic uptake of remnants. A major difference in the metabolism of these particles is that chylomicrons contain a truncated form of apo B (i.e., apo B-48), whereas VLDL contains the complete form (i.e., apo B-100). Another difference is that chylomicron remnants are degraded after they are absorbed by the liver, whereas many of the VLDL remnants are most likely processed in the hepatic sinusoids to become LDL.

REGULATION OF LIPOPROTEIN METABOLISM

There are four major clinically significant physiologic steps in the lipoprotein cascade from VLDL to LDL—namely, VLDL assembly, hydrolysis by LPL, remnant catabolism, and LDL catabolism [*see Figure 3*].[1,2] Defects at any step in the cascade can lead to hyperlipidemia. These defects can be genetic or acquired (i.e., secondary to disease or the effects of drugs) or the result of an interaction of genetic and acquired factors.

Table 1 Major Apolipoproteins and Their Functions

Apolipoprotein	*Function*
Apo A-I	Structural protein of HDL; activates lecithin-cholesterol acyltransferase
Apo A-II	Structural protein of HDL
Apo B-48	Structural protein of chylomicron
Apo B-100	Structural protein of VLDL, IDL, and LDL; ligand for LDL receptor
Apo C-II	Activator of LPL
Apo C-III	Potential inhibitor of apo C-II and apo E functions
Apo E	Ligand for chylomicron remnant receptor and LDL receptor
Apo(a)	Function unknown; antagonizes plasminogen

HDL—high-density lipoprotein IDL—intermediate-density lipoprotein LDL—low-density lipoprotein LPL—lipoprotein lipase VLDL—very low density lipoprotein

Lipoprotein Assembly

Apo B-100 is synthesized constitutively in the endoplasmic reticulum of the hepatocyte, and much of it is degraded in the endoplasmic reticulum. Triglyceride is added to the surviving apo B that will be secreted as VLDL. It is transported to the Golgi complex, where it acquires additional core lipid, forming the nascent VLDL particle. This particle is secreted into plasma, where it acquires apolipoproteins (e.g., apo C-II and apo E) from HDL.[1]

Abnormalities in VLDL secretion can occur in two genetic forms of hyperlipidemia: familial hypertriglyceridemia (FHTG) and familial combined hyperlipidemia (FCHL). FHTG is characterized by the overproduction of triglyceride contained within a normal number of VLDL particles; this results in each particle's having an excessive amount of triglyceride. In FCHL, an excessive amount of apo B-100 is secreted into VLDL or LDL particles; these particles tend to be smaller than normal.[3]

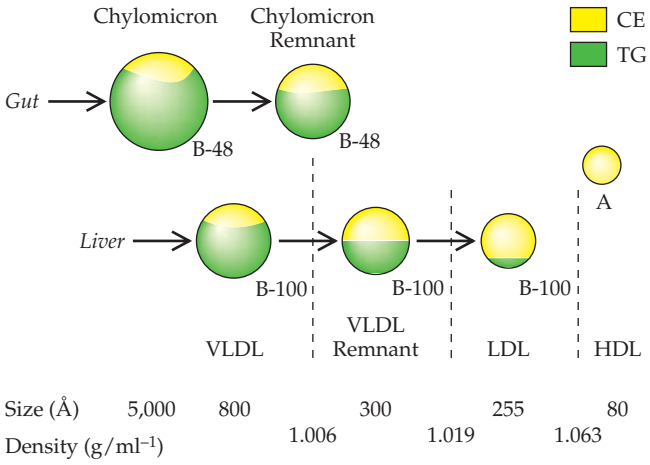


Figure 2 **Size and buoyancy characteristics of lipoproteins. Chylomicrons, which are composed largely of triglyceride, are the largest and most buoyant of the lipoproteins. High-density lipoprotein (HDL) particles are substantially smaller and denser and are composed mostly of cholesteryl ester. (CE—cholesteryl ester; TG—triglyceride)**

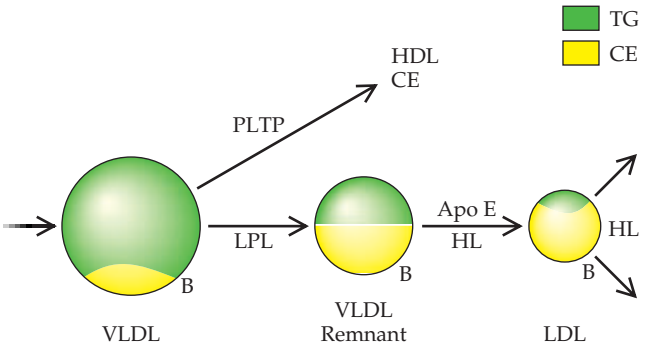


Figure 3 **The apolipoprotein B-100 (apo B-100) cascade. VLDL is secreted from the liver with one apo B on the surface and triglyceride and cholesteryl ester in the core. Core triglyceride is hydrolyzed by lipoprotein lipase and becomes a remnant lipoprotein that is recognized by the liver—in part, by apo E. The remnant lipoprotein is further processed to form LDL, which has a cholesterol-rich core and an intact apo B on its surface. The LDL particle can be removed by peripheral or hepatic LDL receptors. As the VLDL core is hydrolyzed, the unesterified cholesterol and phospholipid are transferred to HDL by phospholipid transfer protein to become the cholesteryl ester of HDL. (CE—cholesteryl ester; HL— hepatic lipase; LPL—lipoprotein lipase; PLTP—phospholipid transfer protein; TG—triglyceride)**

The metabolic syndrome, which is a common condition in the general population, is a component of most cases of FCHL and also contributes to the residual dyslipidemia seen in patients with type 2 diabetes mellitus who have been treated with insulin or insulin secretagogues. The molecular basis of the hepatic triglyceride or apo B oversecretion in these disorders is unknown.

A deficiency in lipoproteins containing apo B is referred to as hypobetalipoproteinemia; an absence of apo B is termed abetalipoproteinemia. Abetalipoproteinemia may occur because of a defect involving both apo B genes that prevents the production of apo B. It also may occur in individuals who are homozygous for mutations in the microsomal triglyceride transport protein, which is critical for apo B transport in the endoplasmic reticulum. Homozygous hypobetalipoproteinemia and abetalipoproteinemia lead to deficiencies in fat-soluble vitamins because each of these conditions results in a shortage of apo B–containing lipoproteins, which are needed to transport fat-soluble vitamins. Hypobetalipoproteinemia, which is characterized by apo B levels of 50% normal, can be caused by a defect in a single apo B gene.[4]

Lipoprotein(a) Lipoprotein(a) [Lp(a)] is a specific class of lipoprotein particles that are synthesized in the liver and that have a lipid composition similar to that of LDL. Lp(a) differs from LDL by the presence of apolipoprotein(a) [apo(a)], a protein whose structure is homologous to plasminogen.[5] The apo(a) protein is bound by a disulfide linkage to apo B-100 to form Lp(a). High levels of Lp(a) are both prothrombotic and atherogenic.[5] Levels of Lp(a) in plasma are almost completely determined by genetic variation in the *Lp(a)* gene.

Lipoprotein Catabolism

Lipoprotein lipase–mediated triglyceride removal LPL is synthesized in adipose tissue and muscle and then transported to the luminal surface of the endothelial lining of the adjacent capillary, where it acts on triglyceride-rich lipoproteins. The fatty acids that are released during the processing of triglyceride-rich particles (i.e., chylomicrons and VLDL) can be used for energy by muscle, or they can be reesterified into triglyceride and stored in adipocytes for later use.[6] Apo C-II, the LPL activator, is carried on the triglyceride-rich lipoproteins chylomicrons and VLDL.

Genetic defects that result in impaired lipoprotein lipase synthesis or function are rare autosomal recessive causes of hyperlipidemia. Usually, these mutations present in neonates or infants as severe hypertriglyceridemia. Heterozygote parents of these children often have mild hypertriglyceridemia. Acquired defects of LPL, such as untreated diabetes or uremia, are more common causes of hyperlipidemia. When an acquired defect of LPL is associated with a disorder characterized by excessive input of VLDL, marked hypertriglyceridemia can ensue. The coexistence of two or more disorders that independently increase the level of triglycerides in plasma (e.g., FHTG or FCHL coexistent with untreated diabetes) can lead to marked hypertriglyceridemia.[6]

Remnant catabolism Both chylomicron and VLDL remnants acquire apo E from HDL before they can bind to hepatic receptors for either uptake and degradation or further processing to LDL. Three alleles of the *APOE* gene (i.e., *APOE*E2, APOE*E3,* and *APOE*E4*) result in six possible combinations. The *APOE*E4* allele product has the greatest affinity for hepatic receptors, followed by the *APOE*E3* allele product; the *APOE2* allele product has markedly reduced receptor affinity.

Individuals who are homozygous for the *APOE*E2* allele (E2/E2) have marked impairment of hepatic remnant lipoprotein uptake, which results in the accumulation of these remnants in the plasma and in very low levels or the absence of LDL. Interestingly, individuals with E2/E2 typically have either normal or low cholesterol levels because of the paucity of LDL particles characteristic of this disorder.[7] If, however, an individual who is homozygous for the *APOE*E2* allele (E2/E2) has a defect—either inherited or acquired—that causes excessive input of VLDL, then excessive accumulation of VLDL remnants and hyperlipidemia occur. This results in remnant removal disease. Because chylomicron and VLDL remnants contain roughly equal amounts of triglyceride and cholesterol, the hyperlipidemia of remnant removal disease is characterized by both hypercholesterolemia and hypertriglyceridemia.[7]

LDL catabolism The final step at which a defect in lipoprotein metabolism can occur is in LDL catabolism. Apo B-100 on the surface of LDL binds to its receptor on the cell surface; LDL is then absorbed into the cell, where it is catabolized [*see Figure 4*]. After hydrolysis of the core lipids, unesterified cholesterol is used by cells for synthesis of membranes, bile acids, and steroid hormones and for various regulatory actions that prevent overaccumulation of cholesterol within the cell. The vast majority of LDL particles in plasma are taken up by the liver by means of the LDL receptor.

Mutations of the LDL receptor (as found in familial hypercholesterolemia [FH]) or, less commonly, mutations in the apo B-100 molecule (as found in familial defective apo B-100) lead to an impairment in the interaction of LDL with its receptor; this can result in elevated LDL levels. LDL levels also can be influenced by dietary factors. For example, dietary cholesterol delivered to the liver by chylomicron remnants can suppress hepatic LDL receptors, leading to impaired LDL removal from plasma. Dietary saturated fats also may reduce LDL receptor activity and may increase LDL production. Hypothyroidism can also be associated with defective LDL receptor–mediated cholesterol removal.[8]

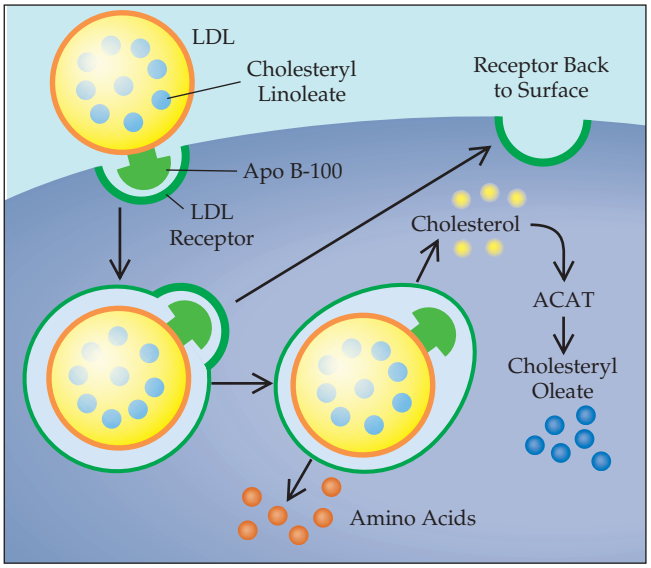


Figure 4 **LDL is absorbed by cells through the LDL receptor. This receptor recognizes apo B-100, the apolipoprotein on the surface of LDL. Once internalized, the lipoprotein is catabolized, releasing cholesterol and amino acids. The free cholesterol is converted to cholesteryl oleate by the enzyme acyl–coenzyme A: cholesterol acyltransferase (ACAT). The LDL receptor is recycled back to the cell surface.**

Function and Regulation of HDL

The major HDL apolipoproteins are apo A-I and apo A-II, which are formed in the liver and small intestine.[9] Apo A-I is secreted with phospholipid in a disklike structure called nascent HDL. Most of the apolipoproteins and phospholipid destined to become nascent HDL are initially secreted on the surface of chylomicrons and VLDL. After LPL hydrolyzes triglyceride in chylomicrons and VLDL, the core lipid content in these lipoprotein particles becomes smaller, and redundancies of unesterified cholesterol and phospholipid occur in the surface layer. These redundant surface components are transferred to HDL by phospholipid transfer protein. Nascent HDL particles also pick up excess unesterified cholesterol and phospholipid from peripheral tissues via the transporter ABCA1. This HDL cholesterol then undergoes esterification by the plasma enzyme lecithin-cholesterol acyltransferase (LCAT). LCAT is activated by apo A-I on the HDL surface to esterify free cholesterol into cholesteryl ester, causing it to move into the core. In this process, the particle becomes the larger, more buoyant HDL_3 particle and progresses to the even larger HDL_2 particle.[9,10] At some point, apo A-II may be added to the HDL_2 particle, which then is directed to deliver cholesteryl ester to the liver by cholesteryl ester transfer protein (CETP). Hepatic lipase activity on the liver surface hydrolyzes the phospholipid and triglyceride in the HDL_2 particle, promoting the decrease in size and density to HDL_3 and then to even smaller HDL particles.[10] Recycling of some of the apo A-I causes the process to repeat itself [*see Figure 5*].

Abnormally high or low levels of HDL cholesterol may be caused, rarely, by genetic defects. Elevations in the HDL cholesterol level may result from genetic hyperalphalipoproteinemia or CETP deficiency. Markedly reduced HDL cholesterol levels may be caused by apo A-I structural mutation; homozygosity for mutations in *ABCA1*,[11] leading to Tangier disease; or homozygosity for mutations in the enzyme LCAT, leading to LCAT deficiency and fish-eye disease. Factors associated with an increase in HDL levels include female sex, aerobic exercise, weight reduction, high-fat diets, and certain drugs (e.g., alcohol, estrogens, fibrates, and nicotinic acid) [*see Table 2*]. Factors associated with a decrease in HDL levels include male sex, central obesity, cigarette smoking, low-fat diets, hypertriglyceridemia, uremia, being heterozygous for Tangier disease, and certain drugs (e.g., androgens, progestins, and some antihypertensive agents) [*see Table 2*]. Low HDL particle number is commonly associated with increased triglyceride levels, as seen in the metabolic syndrome.

Function of Hepatic Lipase

Hepatic lipase is synthesized in the hepatocyte, binds to endothelial surfaces in the liver sinusoids, and acts on lipoproteins.[10] After triglyceride-rich VLDL particles exchange triglyceride for the cholesteryl ester in LDL and HDL, hepatic lipase can hydrolyze the phospholipid and triglyceride in LDL and HDL [*see Figure 6*]. This process leads to the formation of small, dense LDL and converts HDL_2 to HDL_3. This process may be driven by the presence of excessive levels of triglyceride-rich VLDL in the presence of normal hepatic lipase activity or by increases in the level of hepatic lipase. Factors such as male sex and the accumulation of intra-abdominal fat predispose to increased hepatic lipase levels and are associated with an increase in small, dense LDL levels and a decrease in HDL_2 levels. Increased hepatic lipase activity is an important factor in the dyslipidemia of the metabolic syndrome.[10,12] Hepatic lipase also may facilitate hepatic recognition and uptake of chylomicron and VLDL remnant lipoproteins.

Clinical Manifestations of Dyslipidemia

The main clinical consequences of hyperlipidemia are premature atherosclerosis; pancreatitis, which is usually associated with the chylomicronemia syndrome; and nonalcoholic fatty liver disease. Atherosclerosis is most clearly associated with elevated levels of LDL cholesterol and reduced levels of HDL cholesterol. In both pancreatitis and fatty liver disease, the underlying lipid disorder is hypertriglyceridemia.

DYSLIPIDEMIA IN ATHEROSCLEROSIS

There is consensus that elevated plasma LDL levels and reduced HDL levels are associated with an increased risk of atherosclerosis. The role of hypertriglyceridemia as a cardiovascular risk factor is more complex. Hypertriglyceridemia may be a marker for other lipoprotein abnormalities (e.g., increased levels of small, dense LDL particles; low levels of HDL; or remnant accumulation) that are part of the dyslipidemic pattern associated with FCHL, type 2 diabetes mellitus, and the metabolic syndrome. In these settings, hypertriglyceridemia is a predictor of increased premature cardiovascular risk. However, other forms of hypertriglyceridemia may not be associated with premature cardiovascular disease [*see* Familial Hypertriglyceridemia, *below*]. The precise mechanisms whereby increased levels of LDL result in increased atherosclerotic risk are unclear. Very high levels of large, buoyant LDL particles, such as occur in FH and familial defective apo B-100, as well as the presence of more moderate numbers of small, dense LDL particles, are associated with an increased risk of cardiovascular disease. Accumulating evidence suggests that LDL needs to be modified before it becomes atherogenic.[13] Oxidation of LDL may increase its atherogenicity. Oxidized LDL has many biologic properties that may cause it to become atherogenic. The atherogenicity of small, dense LDL particles may result from the ability of LDL to enter the arterial

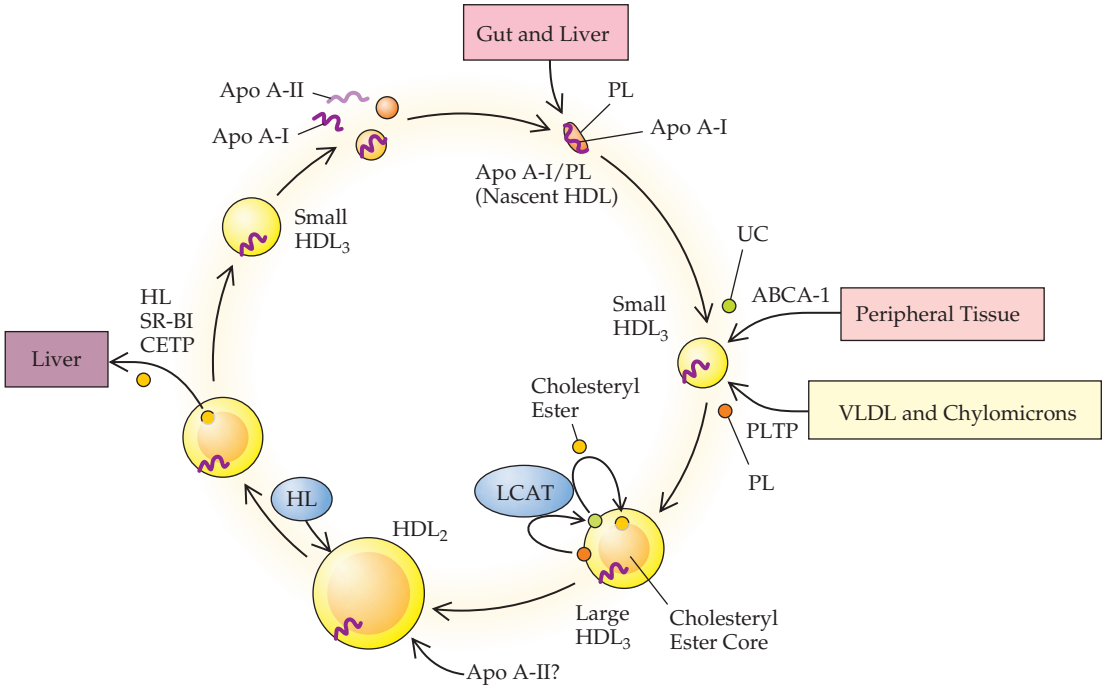


Figure 5 **The circular pathway of HDL formation and degradation.[11] HDL begins as an apo A-1 phospholipid complex. Unesterified cholesterol and phospholipid are added to the nascent HDL via adenosine triphosphate–binding cassette transporter A-1 and phospholipid transfer protein to begin the formation of the smaller HDL_3 particle. LCAT transfers a fatty acid from phospholipid to unesterified cholesterol to cholesteryl ester, which moves to the HDL core. In this process, the HDL particle becomes the larger, more buoyant HDL_3 particle and progresses to the even larger HDL_2 particle. Cholesteryl ester transfer protein contributes to the transfer of cholesteryl ester from HDL_2 to the liver and various lipoproteins; with this loss of cholesteryl ester, the HDL particle shrinks in size. Hepatic lipase hydrolyzes the phospholipid and triglyceride in the HDL_2 particle, promoting the decrease in size and density to HDL_3 and then to even smaller HDL particles, including apo A-1. Recycling of some of the apo A-I causes the process to repeat itself. The role of apo A-II in this process in humans is not clear. (ABCA1—ATP-binding cassette transporter A1; CETP—cholesteryl ester transfer protein; LCAT—lecithin-cholesterol acyltransferase; LPL—lipoprotein lipase; PL—phospholipid; PLTP—phospholipid transfer protein; SR-BI—scavenger receptor BI; UC—unesterified cholesterol)**

intima, where it is retained by matrix molecules and undergoes oxidation more readily than larger, more buoyant LDL particles. The antiatherogenic properties of HDL are probably related to its role in reverse cholesterol transport, and HDL may have anti-inflammatory and antioxidant effects.

DYSLIPIDEMIA IN THE CHYLOMICRONEMIA SYNDROME

Pancreatitis is associated with chylomicronemia, usually with elevated levels of VLDL. The mechanism by which chylomicronemia causes pancreatitis is unclear. Pancreatitis is believed to be caused by the release of free fatty acids and lysolecithin from chylomicrons in excess of their binding capacity in the capillaries of the pancreas by pancreatic lipase.

The chylomicronemia syndrome occasionally occurs when LPL is defective as a result of genetic variation in the enzyme or its cofactor, apo C-II. Much more commonly, chylomicronemia is caused by the coexistence of a genetic form of hypertriglyceridemia combined with an acquired disorder of plasma triglyceride metabolism, the most common being untreated diabetes. Other conditions may be implicated (e.g., hypothyroidism and nephrotic syndrome), as may the use of drugs that raise triglyceride levels.

The chylomicronemia syndrome is associated with abdominal pain, eruptive xanthomas, and transient memory loss. Eruptive xanthomas occur most frequently on the buttocks and the extensor surfaces of the upper limb. A reversible loss of memory, particularly for recent events, and peripheral neuropathy, which sometimes mimics the carpal tunnel syndrome, also may occur. The retinal vessels occasionally demonstrate lipemia retinalis. If the chylomicronemia syndrome is not corrected, it may lead to acute pancreatitis. Acute pancreatitis can be fatal and is often recurrent until low triglyceride levels are maintained. The risk of pancreatitis caused by severe hypertriglyceridemia markedly increases with triglyceride levels over 2,000 mg/dl.

DYSLIPIDEMIA IN NONALCOHOLIC FATTY LIVER DISEASE

Fatty liver disease seems to occur in both genetic and acquired hypertriglyceridemia. It usually is caused by the synthesis of hepatic triglyceride in amounts that are excessive relative to the amount of apo B that is synthesized; this leads to accumulation of triglyceride in the liver, rather than the hepatic secretion of VLDL triglyceride. Fatty liver disease also may occur in heterozygous familial hypobetalipoproteinenemia because of the decreased synthesis of hepatic apo B associated with this disorder. Alcoholic fatty liver disease also occurs with increased hepatic triglyceride synthesis in the face of impaired apo B synthesis.[14] Fatty liver disease has been associated with the metabolic syndrome,[12] which is related to central obesity, insulin resistance, and hypertriglyceridemia.

Any severe form of hypertriglyceridemia with defective VLDL catabolism also can be associated with fatty liver and hepatosplenomegaly. In fact, familial LPL deficiency—a form of hypertriglyceridemia caused entirely by an extrahepatic defect

Table 2 Effects of Selected Drugs on Lipoprotein Levels

Drug	*VLDL*	*LDL*	*HDL*
Alcohol*	+	0	+
Estrogens, estradiol*	+	–	+
Androgens, testosterone	+	+	–
Progestins	–	+	–
Glucocorticoids*	+	0	+
Cyclosporines	+	+	+
Tacrolimus	+	+	+
Thiazide diuretics*	+	+	–
Beta blockers*	+	0	–
Calcium channel blockers	0	0	0
Angiotensin-converting enzyme inhibitors	0	0	0
Sertraline*	Possible+	+	0
Protease inhibitors*	+	0	0
Valproate and related drugs	+	0	–
Isotretinoin*	+	0	–

*Can cause severe hypertriglyceridemia and chylomicronemia syndrome in patients with a familial form of hypertriglyceridemia or type 2 diabetes mellitus.

in triglyceride hydrolysis—is commonly associated with fatty liver disease; in this setting, fatty liver disease regresses rapidly with restriction of dietary fat. In some patients, fatty liver disease progresses to steatohepatitis that is associated with fibrosis and necrosis; the reasons for such a progression are not clear. Perhaps a second insult is needed for these patients to develop nonalcoholic steatohepatitis and then progress to cirrhosis.

Approach to the Patient with Abnormal Lipid Levels

PATIENTS WITH ISOLATED ELEVATION OF LDL CHOLESTEROL LEVELS

A patient's cholesterol level is said to be "above desirable" in an individual with low atherosclerotic risk if the LDL cholesterol level exceeds 130 mg/dl. High LDL levels are those above 190 mg/dl. The patient's triglyceride level is by definition normal,[15] and the HDL cholesterol level is variable but is often normal. The lipid disorders in these patients are usually discovered through routine cholesterol screening. Although some observers question the cost-effectiveness of screening men and women older than 20 years, the high prevalence of elevated LDL cholesterol in the United States warrants population screening, as recommended by the National Cholesterol Education Program (NCEP) and other authorities.

Severely elevated cholesterol levels are an indication of FH. The ability to diagnose FH is valuable because affected individuals will require drug therapy from a relatively young age [*see* Familial Hypercholesterolemia, *below*].

Isolated hypercholesterolemia may be present intermittently in patients with FCHL. A family history that is strongly positive for premature cardiovascular disease, or the presence of any of the other criteria for FCHL, should provide clues to the diagnosis of this disorder [*see* Familial Combined Hyperlipidemia, *below*]. Not all cases of mild isolated hypercholesterolemia are indicative of FH or FCHL; such cases may result from interactions of acquired and environmental factors, particularly dietary factors, with unknown genetic factors that confer susceptibility to hypercholesterolemia.

Most current treatment guidelines are based primarily on LDL cholesterol levels, because reduction of LDL has been shown to reduce cardiovascular disease by as much as 50%.[16] Reduction in the consumption of dietary saturated fat and cholesterol usually leads to a modest reduction in LDL cholesterol levels; such a reduction depends in part on the baseline diet [*see CE:IV Diet and Exercise*]. Lifestyle changes, including diet and weight loss, will suffice in some individuals for reducing LDL cholesterol levels to an acceptable range. However, this approach is unlikely to suffice in patients with familial forms of dyslipidemia, such as FH or FCHL.

In patients with familial forms of the disease or in patients for whom lifestyle measures alone fail to bring LDL cholesterol levels within guideline goals, cholesterol-lowering drugs should be added to the treatment regimen [*see* Drug Therapy in Dyslipidemia, *below*]. Diet therapy can reduce LDL cholesterol levels an additional 5% to 15% beyond reductions achieved with drugs.[17] Diet therapy can therefore lead to a reduction in the dosages of required drugs and should be used in combination with drug therapy. The major class of drugs used to reduce LDL cholesterol is the statins. However, bile acid–binding resins and drugs that block cholesterol absorption are of value in patients who do not respond adequately to statins alone, and they can be used in combination with statins and other drugs.

PATIENTS WITH ISOLATED ELEVATION OF TRIGLYCERIDE LEVELS

An isolated elevation in triglyceride levels may be caused by a primary disorder of lipid metabolism (e.g., FHTG or FCHL); it

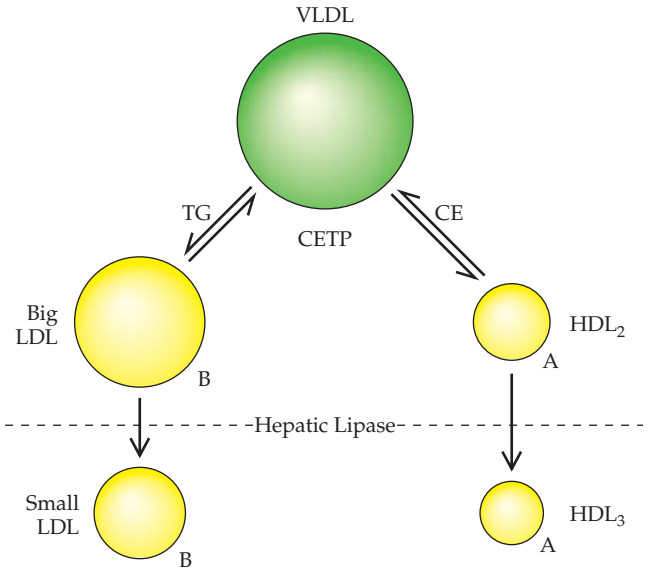


Figure 6 **Dyslipidemia in the metabolic syndrome. Triglyceride-rich VLDL exchanges triglyceride for the cholesteryl ester in LDL and HDL particles. This change in lipoprotein composition is initiated by cholesteryl ester transfer protein. Hepatic lipase hydrolyzes the triglyceride and phospholipid in large LDL and HDL particles, decreasing the size of each particle. (CE—cholesteryl ester; TG—triglyceride; CETP—cholesterol ester transfer protein)**

may arise secondary to the use of therapeutic drugs; or it may be a component of the metabolic syndrome or type 2 diabetes mellitus. Unlike with cholesterol levels, it has been difficult to determine the level of triglyceride at which the risk of CAD increases or decreases. It is valuable to ascertain the cause of the hypertriglyceridemia, because the therapeutic approaches may differ.

For example, it is important to distinguish FHTG, which confers no risk of premature CAD, from FCHL, which is associated with a high incidence of premature atherosclerosis.[17] However, it can be difficult to distinguish these disorders when FCHL is associated with hypertriglyceridemia. A positive personal or family history of premature atherosclerosis suggests FCHL. In addition, patients with FCHL frequently have nonlipid cardiovascular risk factors (i.e., central obesity, hypertension, insulin resistance, impaired glucose tolerance, increased levels of plasminogen activator inhibitor–1, (PAI-1) and increased levels of circulating inflammatory markers). Hypertriglyceridemia present in FCHL indicates the presence of increased numbers of small, dense LDL particles and confers an increased risk of premature cardiovascular disease.[12] Similarly, hypertriglyceridemia associated with type 2 diabetes mellitus and the metabolic syndrome is an important cardiovascular risk factor. Other cardiovascular risk factors are usually present in patients with type 2 diabetes mellitus, the metabolic syndrome, or FCHL. Therefore, the therapeutic strategy must consider factors beyond the lipid disorder.

Patients with FHTG do not appear to be at significantly increased risk for developing premature CAD. However, they are at increased risk for developing the chylomicronemia syndrome when secondary forms of hypertriglyceridemia are present, such as the hypertriglyceridemia caused by the use of triglyceride-raising drugs. The chylomicronemia syndrome occurs in FCHL in combination with other causes of hypertriglyceridemia as well. In patients with pancreatitis caused by hypertriglyceridemia, triglyceride levels are above 2,000 mg/dl and can be much higher. It is recommended that plasma triglyceride levels be maintained below 2,000 mg/dl to prevent recurrent acute pancreatitis. A safe goal would be a level of less than 1,000 mg/dl.

PATIENTS WITH ELEVATIONS IN CHOLESTEROL AND TRIGLYCERIDE LEVELS

Patients with elevations in the levels of both total plasma cholesterol and triglyceride fall into three categories. In the first category, there is an elevation in VLDL and in LDL, as seen in FCHL. In the second category, there is an elevation in VLDL remnants and chylomicron remnants, as in remnant removal disease. The third category consists of patients with very high triglyceride levels in whom the increase in total cholesterol is a result of the cholesterol in VLDL and chylomicrons.

In patients with FCHL, an increase in triglycerides and in LDL cholesterol is often seen. These patients have elevated apo B levels and small, dense LDL particles. Therapy for these individuals often requires several drugs, one aimed at lowering the triglyceride level and one aimed at reducing the amount of small, dense LDL particles [*see* Drug Therapy in Dyslipidemia, *below*].

In patients with remnant removal disease, the levels of plasma cholesterol and triglyceride are often equal. It is important to consider remnant removal disease in these circumstances. Therapy in this case is related to decreasing hepatic lipoprotein secretion with statins, fibrates, or niacin.

In patients with severe hypertriglyceridemia, the increase in total plasma cholesterol is a result of the cholesterol in VLDL and chylomicrons. Fibrates are often the drug of choice. However, it is very important to determine the etiology of the severe hypertriglyceridemia and remove any offending drug [*see Table 2*] or treat any secondary cause for the hypertriglyceridemia.

PATIENTS WITH LOW HDL CHOLESTEROL LEVELS

Many if not most patients with hypertriglyceridemia have a concomitant reduction in HDL cholesterol levels. Therefore, the management of low HDL cholesterol levels should be considered in the context of the management of the underlying disorder (e.g., FCHL or type 2 diabetes mellitus) [*see* Patients with Isolated Elevation of Triglyceride Levels, *above*]. Isolated low HDL cholesterol levels of 20 to 30 mg/dl without concomitant hypertriglyceridemia or other changes in lipid and lipoprotein levels are rare, but such low levels are a risk factor for cardiovascular disease.[9] In the past, these reductions in HDL levels were often not identified; the screening strategies that were employed were based on the assessment of total cholesterol levels, and total cholesterol levels often are not elevated in patients with isolated reductions in HDL. Specific measurement of HDL cholesterol is required to identify these patients. The treatment of the rare patients with isolated low levels of HDL cholesterol remains somewhat controversial. There are no currently available drugs that effectively increase HDL cholesterol levels only.[18] Gemfibrozil, a fibrate that decreases VLDL triglyceride levels, also raises HDL cholesterol levels. Many studies of fibrate therapy for atherosclerosis have been inconclusive. However, the Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial (VA-HIT) demonstrated a reduction in cardiovascular events.[19] Nicotinic acid, which acts at many metabolic sites, also raises HDL cholesterol. It decreased cardiovascular death in the Coronary Drug Project of the 1970s.[20] Few other studies have evaluated the effect of niacin on atherosclerotic events.

PATIENTS WITH ATHEROSCLEROSIS AND NORMAL LIPID LEVELS

On rare occasions, a middle-aged patient with established atherosclerosis is seen to have no detectable lipid or lipoprotein abnormality. In addition to the standard lipid profile, measurement of apo B and Lp(a) will often reveal subtle lipoprotein abnormalities, such as increased numbers of small, dense LDL particles in these patients. Assessment of nonlipoprotein risk factors (e.g., homocysteine and inflammatory markers such as C-reactive protein [CRP]) also may be of value in assessing cardiovascular risk factors. Although the levels of some of these risk factors can be reduced by various strategies (e.g., homocysteine by folate therapy), the use of statins in all categories of high-risk individuals, particularly those who have established vascular disease, has been shown to be of benefit, even if lipid levels are apparently normal.

Genetic Disorders of Lipoprotein Metabolism

Primary disorders of lipoprotein metabolism are those that arise from genetic defects in the metabolic pathways of lipoproteins (i.e., familial disorders caused by increased hepatic secretion of lipoproteins or by catabolic defects). The disorders that cause increased lipoprotein secretion are the metabolic syndrome, familial combined hyperlipidemia, type 2 diabetes mellitus, and FHTG; elevations of Lp(a) can also cause increased lipoprotein secretion. Disorders of LDL catabolism are FH and familial defective apo B-100. Remnant removal disease is a defect in remnant catabolism.

METABOLIC SYNDROME

The metabolic syndrome consists of a central distribution of adiposity or visceral obesity; insulin resistance; elevations in plasma free fatty acid levels; impaired glucose tolerance; hypertension; dyslipidemia; and an abnormal procoagulant state. Many features of this syndrome are known to predispose men and women to premature CAD.[12]

Etiology and Risk Factors

An accumulation of visceral rather than subcutaneous fat has been observed in individuals with the central body fat distribution characteristic of the metabolic syndrome. Men have more visceral fat than premenopausal women, even when matched for body mass index. It has been suggested that these differences in visceral fat and the associated changes in lipoproteins and blood pressure could account, in part, for the difference in risk of premature CAD between men and premenopausal women.[21,22] Increased visceral fat is associated with insulin resistance, hyperinsulinemia, low plasma adiponectin, and elevations in plasma free fatty acid levels.[23] It has been suggested that the accumulation of visceral fat precedes and causes insulin resistance and the resultant hyperinsulinemia, because insulin sensitivity increases and free fatty acid levels fall when visceral fat is decreased after caloric restriction.[24]

The levels of insulin, glucose, triglyceride, HDL cholesterol, blood pressure, PAI-1, and other inflammatory markers are increased above the mean normal in patients with the metabolic syndrome. Although these variables are usually shifted to high levels, some of these variables are in the high-normal range in some affected individuals. HDL levels tend to be lower than mean normal. Genetic and environmental factors appear to affect the distribution of these variables in both normal persons and those with the metabolic syndrome. Because the metabolic syndrome is associated with multiple cardiovascular risk factors, individuals with the metabolic syndrome are at increased risk for CAD. Whether all individuals who meet the NCEP guidelines for the metabolic syndrome[25] [*see* Diagnosis, *below*] are at increased risk for premature CAD is unknown. However, type 2 diabetes mellitus and FCHL are specific disorders of which the metabolic syndrome is a component.[12] These two disorders account for at least 40% to 50% of premature CAD and need to be considered in the context of the metabolic syndrome.

The risk of abdominal fat patterning, dyslipidemia, impaired glucose metabolism, and hypertension—the sentinel symptoms of the metabolic syndrome—increases with age.[26] Central obesity associated with the metabolic syndrome may be evident in young adults after completion of adolescent growth; however, it is more typical for central obesity and insulin resistance to manifest in midlife. Whereas elevations in LDL cholesterol levels may not predict the onset of atherosclerosis in the elderly, central obesity, hypertension, and insulin resistance are risk factors for atherosclerosis, and their prevalence increases with age,[26-30] possibly because of the metabolic syndrome.

Pathophysiology

Although the association of central obesity and insulin resistance with dyslipidemia is well established, the underlying cause remains unclear. One mechanism that would explain the association of central obesity and insulin resistance with dyslipidemia is an increase in portal vein long-chain free fatty acids. Such an increase would inhibit hepatic apo B from undergoing degradation in the endoplasmic reticulum and would increase the likelihood of apo B undergoing hepatic secretion as triglyceride-containing lipoproteins. This would account for the increased levels of triglyceride and the increased number of VLDL and LDL particles seen in patients in insulin-resistant states.[31] Another effect of long-chain free fatty acids is to increase hepatic lipase on the surface of hepatic cells. Hepatic lipase hydrolyzes triglyceride and phospholipid in LDL and HDL, decreasing the size of each particle [*see Figure 6*].[12] However, CETP also contributes to this lipoprotein remodeling process; whether hepatic lipase or CETP has the predominant effect on the size and density of LDL and HDL particles depends on the triglyceride content of VLDL and the secretion rate of VLDL. The differences in LDL particle size and HDL_2 levels between men and premenopausal women can largely be accounted for by differences in visceral fat in men and women.

Diagnosis

The National Cholesterol Education Program Adult Treatment Panel III (NCEP ATPIII) has suggested five clinical variables as diagnostic criteria for the metabolic syndrome: (1) increased waist circumference, (2) increased triglyceride level, (3) decreased HDL cholesterol level, (4) increased blood pressure, and (5) elevated level of fasting plasma glucose [*see Table 3*].[15] A diagnosis of the metabolic syndrome is made when three or more of these clinical variables are present. When these five variables were assessed in a survey of 8,814 adult men and women, approximately 24% of those surveyed met the diagnostic criteria for diagnosis of the metabolic syndrome.[32,33] The World Health Organization (WHO) also has criteria for the metabolic syndrome. An attempt to harmonize the two sets of criteria is in progress.

Visceral obesity and insulin resistance are major contributors to the dyslipidemia associated with the metabolic syndrome. The following lipid abnormalities are associated with the metabolic syndrome: increased levels of triglyceride; increased numbers of small, dense LDL particles; increased apo B levels; and decreased levels of HDL cholesterol. However, in normal, randomly selected populations, isolated visceral obesity and insulin resistance were associated with only a slight increase in triglyceride levels and only a slight decrease in HDL cholesterol levels.[23] In contrast, visceral obesity and insulin resistance can contribute to a more severe dyslipidemia, such as that associated with type 2 diabetes mellitus and FCHL.[12]

The dyslipidemia of the metabolic syndrome can be diagnosed by demonstrating mild to moderate increases in plasma triglyceride and apo B levels, decreased levels of HDL cholesterol, and normal levels of LDL. Although the LDL level is normal in patients with this disorder, the number of LDL particles is generally increased; the predominant form is small, dense LDL

Table 3 Clinical Features of the Metabolic Syndrome[15]

The presence of three or more variables indicates a diagnosis of metabolic syndrome
Abdominal obesity: waist circumference > 35 in (women) or > 40 in (men)
Triglycerides ≥ 150 mg/dl
HDL cholesterol < 50 mg/dl (women) or < 40 mg/dl (men)
Blood pressure ≥ 130/85 mm Hg
Fasting plasma glucose ≥ 110 mg/dl

particles, which are cholesterol poor relative to large, buoyant LDL particles. The presence of small, dense LDL particles can be determined by direct measurement of LDL size or density. The routine measurement of plasma apo B levels in clinical practice is not necessary for the diagnosis of this disorder; however, measurement of plasma apo B levels can indicate the presence of increased numbers of small, dense LDL particles. Similarly, total HDL levels reflect changes in the HDL_2 levels, indicating that HDL subfractions do not need to be measured.[34]

Treatment

Aerobic exercise and a diet low in saturated fat are indicated as therapy for most people with the metabolic syndrome. If the metabolic syndrome is severe or FCHL or type 2 diabetes mellitus is present, more aggressive therapy is indicated [*see CE:IV Diet and Exercise* and *see* Drug Therapy in Dyslipidemia, *below*].

FAMILIAL COMBINED HYPERLIPIDEMIA

FCHL is an autosomal dominant disorder that accounts for up to half of the familial causes of CAD[35]; it was first described in families of survivors of myocardial infarction.[36-38] FCHL is characterized by elevations in triglyceride or cholesterol levels, or both, in affected relatives. In addition to increases in triglyceride and cholesterol levels, patients with FCHL characteristically have elevations in apo B levels and increased numbers of small, dense LDL particles.[39]

Genetic linkage analysis suggests that the inheritance of the lipid phenotype in FCHL involves separate gene effects[40] for the elevation in apo B levels[41] and the increased numbers of small, dense LDL particles that are present in FCHL families. Further evidence for genetic heterogeneity comes from studies that found that in one third of individuals with FCHL, the activity level of LPL in postheparin plasma was reduced by half.[42] Visceral obesity and insulin resistance contribute to the dyslipidemia seen in FCHL but cannot account for the elevation in apo B levels.[43]

In the Familial Atherosclerosis Treatment Study (FATS), intensive lipid-lowering therapy with nicotinic acid or lovastatin in combination with colestipol led to decreased hepatic lipase activity; decreased numbers of small, dense LDL particles; and elevated levels of HDL_2 cholesterol, with subsequent regression of CAD, as evidenced by angiography.[44] Intensive lipid lowering resulted in subsequent regression of atherosclerosis, particularly in individuals with small, dense LDL particles who had FCHL or who had elevated Lp(a) levels at baseline.

An aggressive approach to modify reversible cardiovascular risk factors should be undertaken in individuals affected by this disorder. Diet therapy and therapeutic lifestyle modification that includes physical activity should be undertaken [*see CE:IV Diet and Exercise*], together with lipid-lowering drug therapy [*see* Drug Therapy in Dyslipidemia, *below*][18] and management of other cardiovascular risk factors. Which lipid-lowering drug to use depends to some extent on whether the primary lipid manifestation is hypercholesterolemia, hypertriglyceridemia, or combined elevations of cholesterol and triglyceride. If hypercholesterolemia is the primary manifestation, the approach should be the same as that for the hypercholesterolemic patient [*see* Patients with Isolated Elevation of LDL Cholesterol Levels, *above*]. If hypertriglyceridemia is the major abnormality, the initial approach might be that used for patients with isolated hypertriglyceridemia. However, most patients will have elevations in both triglyceride and LDL levels and will require combination therapy; regimens may combine a statin and niacin, fibrate, or ezetimibe [*see* Drug Therapy in Dyslipidemia, *below*].

TYPE 2 DIABETES MELLITUS

Patients undergoing treatment of type 2 diabetes mellitus characteristically have visceral obesity and insulin resistance. A defect in insulin secretion is present in insulin-resistant individuals who develop hyperglycemia. First-degree relatives of individuals with type 2 diabetes mellitus may be centrally obese and insulin resistant or may experience decreased insulin secretion in response to glucose; first-degree relatives who are both centrally obese and who have a defect in insulin secretion invariably develop type 2 diabetes mellitus. Although the genes contributing to central obesity, insulin resistance, and defective insulin secretion are mostly unknown, type 2 diabetes mellitus is a classic example of an oligogenic disorder. Determining all of the genes involved will require careful phenotypic characterization of subsets of individuals with type 2 diabetes mellitus.

The dyslipidemia of untreated diabetes mellitus and hyperglycemia is discussed later in this chapter under acquired disorders [*see* Endocrine Disorders That Cause Dyslipidemia, *below*]. The dyslipidemia of treated type 2 diabetes mellitus is similar to that of the metabolic syndrome and FCHL; it is characterized by a mild increase in triglyceride levels, decreased HDL_2 cholesterol levels, and increased numbers of small, dense LDL particles. Treatment entails diet therapy, increased physical activity, and lipid-lowering drug therapy [*see* Drug Therapy in Dyslipidemia, *below*].[45]

FAMILIAL HYPERTRIGLYCERIDEMIA

FHTG is a common inherited disorder, thought to be autosomal dominant, that affects about 1% of the population. FHTG is characterized by an increase in triglyceride synthesis resulting in VLDL particles enriched with triglyceride secreted in normal numbers. Affected people have elevated VLDL levels but low levels of LDL and HDL and are generally asymptomatic unless severe hypertriglyceridemia (i.e., chylomicronemia syndrome) develops. FHTG does not appear to be associated with an increase in the risk of premature CAD.[17]

A diagnosis is made by family history and examination of fasting lipoprotein profiles of the patient and relatives. The triglyceride level ranges from 250 to 1,000 mg/dl in approximately one half of first-degree relatives; a strong family history of premature CAD usually is lacking; and elevated LDL levels should not be present.

Patients with FHTG should lose weight if necessary, exercise regularly, and reduce their intake of saturated fatty acids and cholesterol. Alcohol, exogenous estrogens, and other drugs that increase VLDL levels might need to be restricted. Diabetes, if present, should be well controlled. Hypertriglyceridemia in patients with FHTG often responds to these measures. If triglyceride levels exceed 500 mg/dl after 6 months of nonpharmacologic therapy, drug therapy with a fibrate should be considered[19]; at levels above 1,000 mg/dl, drug therapy should be instituted.

Fibrates are the drugs of choice to reduce elevated triglyceride levels in patients with familial hypertriglyceridemia [*see* Drug Therapy in Dyslipidemia, *below*]. In familial combined hyperlipidemia, niacin can be very useful. Niacin has several additional beneficial effects on blood lipids—it increases HDL cholesterol levels; it reduces levels of small, dense LDL particles; and it may reduce Lp(a) levels. Despite having a less dramatic effect on triglycerides than fibrates, statins have been shown to be of value in high-risk patients with moderate hypertriglyceridemia and increased levels of small, dense LDL particles, such as occur in patients with type 2 diabetes mellitus and FCHL.

FAMILIAL HYPERCHOLESTEROLEMIA

FH is an autosomal dominant disorder caused by a mutation in the gene encoding the LDL receptor protein. The extremely rare homozygote with FH has two mutant alleles at the LDL receptor locus, leaving the person with an absolute or nearly absolute inability to clear LDL from the circulation by the LDL receptor.[8] Heterozygotes with FH possess one normal allele, giving them approximately one half of the normal receptor activity. Because the LDL receptor contributes to VLDL remnant clearance from the plasma, a deficiency of LDL receptors may lead to some accumulation of remnant lipoproteins. High concentrations of LDL result in nonreceptor-mediated uptake of LDL by the extracellular matrix, including that of the arterial wall, which leads to the formation of xanthomas and atherosclerosis. The heterozygous form of this disorder has a prevalence of about one in 500 people, making it one of the more common genetic diseases.[8]

Diagnosis

Hypercholesterolemia can be detected at birth in umbilical cord blood. If FH is not detected at birth, various associated conditions may suggest the diagnosis later in life. Tendon xanthomas are a highly specific sign of FH; typically, they begin to appear by 20 years of age and may be present in up to 70% of older patients. Occasionally, xanthomas are seen on the patellar tendon. Because xanthomas are subtle, careful examination of the dorsal hand tendons and Achilles tendon is required for their detection. Xanthelasma (cutaneous xanthomas on the palpebra) and corneal arcus are common in patients with FH after 30 years of age; however, they also occur in normocholesterolemic persons. Early corneal arcus is seen superiorly and inferiorly in the eyes and later becomes totally circumferential.

CAD develops early, with symptoms often manifesting in men in the fourth or fifth decade. Approximately 5% of all cases of premature MI occur in patients with heterozygous FH.[8] Before the development of statin therapy, at least 50% of men with heterozygous FH experienced MI by 60 years of age; in women, symptoms tend to develop about 10 years later. The total cholesterol level in heterozygous patients generally ranges from 350 to 550 mg/dl. The triglyceride level may be mildly elevated, and the HDL cholesterol level is reduced in about 10% of heterozygotes. LDL receptor function can be measured only in special laboratories.

Heterozygous FH should be suspected when severe hypercholesterolemia from elevated LDL is detected. If tendon xanthomas are present, the diagnosis is virtually certain. If tendon xanthomas are absent, secondary causes of hypercholesterolemia (e.g., hypothyroidism) should be sought, but the diagnosis of familial hypercholesterolemia is not excluded. A comprehensive family history should reveal a strong history of premature CAD and hypercholesterolemia without hypertriglyceridemia; the disorder affects approximately one half of first-degree relatives. The presence of hypercholesterolemia and tendon xanthomas in a parent or sibling is virtually diagnostic, as is hypercholesterolemia in a child in the family. Careful screening of family members is mandatory, because 50% of first-degree relatives will be affected and will require aggressive lipid-lowering therapy.[18,46]

Treatment

Management of FH requires both dietary intervention and drug therapy. The goal of therapy is to lower the LDL cholesterol level to less than 130 mg/dl, or even lower if the patient exhibits CAD. In patients with heterozygous FH, effective treatment is possible with combinations of statins, intestinally active drugs, and nicotinic acid. Because LDL cholesterol levels tend to be very high, combination therapy with two drugs is often required, and three drugs may be necessary [*see* Drug Therapy in Dyslipidemia, *below*]. Although diet therapy alone is not sufficient for patients with heterozygous FH, reducing saturated fatty acid and cholesterol intake will lower LDL levels and reduce the amount of medication required. This is particularly important in children and adolescents before initiation of drug therapy. Tendon xanthomas have been shown to regress when LDL levels are maintained in a desirable range. Aggressive reduction of LDL cholesterol in men and women who have heterozygous FH may cause a regression of coronary atherosclerosis.

FAMILIAL DEFECTIVE APOLIPOPROTEIN B-100

A mutation in apo B-100 that inhibits its binding to the LDL receptor is another genetic cause of elevations in the LDL level. The prevalence of this disorder is unknown but is estimated to be 5% to 10% that of FH. LDL receptor structure and function are normal. A full-length apo B-100 molecule is produced with a single amino acid substitution; this results in apo B that binds poorly to LDL receptors, leading to LDL accumulation in the plasma.

Affected individuals are clinically indistinguishable from patients with heterozygous FH: they may present with severe hypercholesterolemia, tendon xanthomas, and premature atherosclerosis. Treatment with statins appears to lower LDL cholesterol levels in patients with this disorder. Specialized tests available only in selected research laboratories are required to distinguish affected people with defective apo B from those with defective LDL receptors.

INCREASED LEVELS OF LIPOPROTEIN(A)

Lp(a) is a specific class of lipoprotein particles synthesized in the liver.[5] An important component of Lp(a) is apo(a), which has a structure homologous with plasminogen, a key protein in the coagulation cascade. Plasma concentrations of Lp(a) vary markedly among individuals, ranging from undetectable to 200 mg/dl. Lp(a) plasma concentration is strongly controlled by genetic factors.

Most epidemiologic studies suggest that Lp(a) is a risk factor for CAD and stroke. If Lp(a) is atherogenic, it may be because of its LDL-like properties: Lp(a) has been shown to undergo endothelial uptake and oxidative modification and to promote foam cell formation. Because Lp(a) has a high degree of homology with plasminogen, it may play a role in thrombosis by interfering with the binding of plasminogen to fibrin. Elevated Lp(a) levels appear to increase the atherogenicity of other cardiovascular risk factors, with earlier onset of cardiovascular events.

Data suggest that reducing LDL cholesterol levels in patients with high levels of Lp(a) may be an effective strategy to slow the progression of atherosclerosis and to prevent coronary events. The Lp(a) level itself can be reduced with high-dose niacin, estrogen, or tamoxifen, as well as with LDL apheresis. Insufficient data exist regarding the efficacy of lowering the Lp(a) level per se to inhibit atherosclerosis or to prevent coronary events.[5]

REMNANT REMOVAL DISEASE

Remnant removal disease, also called type III hyperlipoproteinemia, dysbetalipoproteinemia, and broad-beta disease, is defined as the presence of VLDL particles that migrate in the beta position on electrophoresis (normal VLDL particles migrate in

the pre-beta location). Beta-VLDL particles are chylomicron and VLDL remnants.

Remnant removal disease is caused in part by a mutation in the *APOE* gene[7] [*see* Regulation of Lipoprotein Catabolism, *above*]; this mutation leads to an impairment in the hepatic uptake of apo E–containing lipoproteins and stops the conversion of VLDL and IDL to LDL. Without the presence of additional genetic, hormonal, or environmental factors, remnants do not accumulate to a degree sufficient to cause hyperlipidemia, because they are cleared by hepatic receptors that also bind, with less avidity, to apo B-48 and apo B-100. Remnant removal disease results when an apo E defect (almost always the E2/E2 genotype) occurs in conjunction with a second genetic or acquired defect that causes either overproduction of VLDL (such as occurs with FCHL) or a reduction in LDL receptor activity (such as occurs in heterozygous FH or hypothyroidism). The E2/E2 genotype is found in 1% of the white population and in virtually all persons with remnant removal disease.

Diagnosis

Persons with remnant removal disease have elevations in both cholesterol and triglyceride levels and are likely to develop premature CAD. For reasons that are not understood, these patients are at particularly increased risk for peripheral vascular disease. Hyperlipidemia usually does not develop before adulthood. Palmar xanthomas (xanthoma striata palmaris)—orange-yellow discolorations of the palmar creases—are pathognomonic for genetic remnant removal disease, but they are not always present. Palmar xanthomas may be difficult to see and should be carefully sought using good lighting. Tuboeruptive xanthomas are occasionally found at pressure sites, particularly the elbows, buttocks, and knees.

The diagnosis of remnant removal disease should be suspected in a person with elevated total cholesterol and triglyceride levels, elevated VLDL and IDL cholesterol levels, and reduced LDL and HDL cholesterol levels. Cholesterol and triglyceride levels range from 300 to 1,000 mg/dl and are roughly equal, except during an acute exacerbation, at which time hypertriglyceridemia tends to predominate. Beta-migrating VLDL is present on electrophoresis, although this test is seldom used today. Ultracentrifugation demonstrates that the ratio of VLDL cholesterol to total plasma triglyceride is greater than 0.3. Definitive diagnosis is made by detecting the E2/E2 phenotype by isoelectric focusing of plasma lipoproteins or the genotype by gene analysis.

Treatment

Generally, therapy for remnant removal disease is the same as that for other forms of hypertriglyceridemia. A low-fat diet, weight loss, and exercise can have a major effect on lipid levels. Fibrates, statins, and nicotinic acid have been used successfully in this disorder. However, drugs that increase triglyceride levels, such as bile acid-binding resins, must be avoided.

RARE DISORDERS

Severe hypertriglyceridemia can present in childhood as a result of LPL deficiency or, extremely rarely, as apo C-II deficiency. These patients are at risk for acute pancreatitis with severe hypertriglyceridemia and must be treated with moderate to severe dietary-fat restriction until plasma triglyceride levels are below 1,000 to 2,000 mg/dl.

Homozygous FH is extremely rare and leads to severe hypercholesterolemia, atherosclerosis, and death, often in the first two decades of life. Patients with homozygous FH may benefit from LDL apheresis. At the other extreme, the absence of apo B-containing lipoproteins can result from defects in the synthesis of apo B (e.g., homozygous hypobetalipoproteinemia) or from defects in the transport of apo B into the hepatic endoplasmic reticulum. Individuals with very low apo B levels are not at risk for atherosclerosis.

The absence of HDL can occur in persons with homozygosity for defects in the cholesterol and phospholipid transporter ABCA-1. The heterozygous state is an uncommon cause of isolated low-HDL cholesterolemia[11] (i.e., hypoalphalipoproteinemia).

MISCELLANEOUS COMMON DYSLIPIDEMIAS

Polygenic hypercholesterolemia was once thought to be common. Polygenic hypercholesterolemia is a term used to refer to the occurrence of mild elevations in LDL cholesterol in the apparent absence of a familial form of dyslipidemia or of dyslipidemia of secondary cause. This category of dyslipidemia continues to shrink as LDL variants such as Lp(a) and small, dense LDL particles are discovered.

Mild to moderate hypertriglyceridemia may occur in the presence of modest defects in LPL. Typically, it presents as an increase in VLDL levels in conjunction with a decrease in HDL cholesterol levels. It is seen in the obligate heterozygote parents of children with LPL deficiency. This defect may predispose to premature CAD.

Secondary Disorders of Lipoprotein Metabolism

Secondary dyslipoproteinemias are caused by acquired defects in lipoprotein metabolism that result in hypercholesterolemia, hypertriglyceridemia, or combined hyperlipidemia; the HDL level may or may not be low. Secondary hypertriglyceridemia in conjunction with a common genetic form of hypertriglyceridemia may be severe enough to cause chylomicronemia with pancreatitits. Dyslipoproteinemia may also be caused by selected medications.

ENDOCRINE DISORDERS THAT CAUSE DYSLIPIDEMIA

Untreated Hyperglycemia

Untreated hyperglycemia in patients with diabetes mellitus causes an increase in VLDL synthesis, a reduction in VLDL catabolism with an accompanying reduction in LPL activity, or both. These abnormalities result in hypertriglyceridemia and a reduction in the level of HDL. The LDL level usually is normal. Fasting chylomicronemia occurs when there is a coexisting primary form of hypertriglyceridemia. VLDL and chylomicrons compete to interact with LPL, and both lipoproteins may accumulate. A low HDL level results from impaired lipolysis of triglyceride-rich lipoproteins, which supply lipid components for HDL development. These defects occur in both untreated type 1 and untreated type 2 diabetes mellitus. Lipid levels should approach normal with comprehensive treatment of diabetes; if they fail to do so, additional causes should be sought [*see* Genetic Disorders of Lipoprotein Metabolism, *above*]. In diabetic patients with persistent moderate to severe hypertriglyceridemia, a fibric acid is suitable because it reduces the secretion of VLDL and enhances the activity of LPL. Nicotinic acid may be used, but with care, particularly in patients with type 2 diabetes mellitus, because it may exacerbate hyperglycemia.[47] Statins are effective in reducing coronary events in diabetic patients.[45]

Hypothyroidism

Hypothyroidism may cause a severe elevation of LDL levels because of reduced LDL receptor activity; in addition, it frequently causes hypertriglyceridemia and an associated reduction in the HDL level as a result of reduced LPL activity. Remnants of chylomicrons and VLDL may also accumulate and unmask remnant removal disease. The dyslipoproteinemia that occurs with hypothyroidism is corrected by thyroid hormone replacement.

Dyslipidemia Secondary to Estrogen and Progestin Therapy

Oral contraceptives that contain a combination of estrogen and progestin can have variable effects on lipoproteins, depending on the specific combination used. Estrogen tends to raise VLDL and HDL levels and lower LDL levels. Progestins tend to lower VLDL and HDL levels and raise LDL levels, but the effect varies considerably. Postmenopausal estrogen replacement therapy lowers LDL levels and raises HDL levels; the addition of progesterone to protect the uterus lessens these effects but does not eliminate them.[48] Estrogen may increase triglycerides to severe levels in women who have an underlying primary triglyceride disorder, leading to pancreatitis; therefore, triglyceride levels should be closely monitored in these patients.[6] Oral combination therapy with estrogen and progesterone was associated with a mild increase in CAD[43] in the Women's Health Initiative Study. In this randomized study of 16,608 women, use of oral hormone replacement therapy also was associated with an excess rate of breast cancer. In women who have undergone hysterectomies, estrogen therapy has been shown to increase the risk of stroke [*see* Managing Dysplipidemia in Women, *below*].[49] These studies have led to a decrease in the use of postmenopausal hormone replacement therapy.

RENAL DISORDERS THAT CAUSE DYSLIPIDEMIA

Nephrotic Syndrome

The nephrotic syndrome causes enhanced hepatic secretion of apo B-100–containing lipoproteins (i.e., VLDL) in response to the loss of albumin and other proteins in the urine. Hepatic synthesis of cholesterol is also increased. The LDL level is typically elevated, and it may be severely elevated. The VLDL level elevation may be associated with a reduction in the HDL level as lipolysis becomes impaired.[50] Patients with the nephrotic syndrome are at increased risk for CAD, and the lipid disorder should be treated aggressively. Dietary change, weight loss, and exercise may improve lipoprotein levels, but pharmacologic therapy is necessary to achieve desirable lipoprotein levels. Nicotinic acid should be effective in the treatment of this disorder because it inhibits hepatic secretion of apo B-100–containing lipoproteins; however, it has not been studied extensively for this use. The statins are useful in lowering LDL cholesterol levels in patients with the nephrotic syndrome. Combination drug therapy with statins, nicotinic acid, fibrates, or ezetimibe may be necessary for the reduction of LDL cholesterol and triglyceride levels [*see* Drug Therapy in Dyslipidemia, *below*]. Studies are needed to evaluate the effects of various drug combinations on cardiovascular outcomes.

Chronic Renal Failure

Chronic renal failure produces hypertriglyceridemia as a result of a decrease in LPL and hepatic triglyceride lipase.[50] Triglyceride levels typically range from 150 to 750 mg/dl, and the HDL level is usually low; the risk of CAD is increased. Dietary measures should be initiated while drug treatment is being considered. Gemfibrozil, a drug that enhances LPL activity, has been shown to be effective in lowering triglyceride levels in patients with renal insufficiency.[51] Gemfibrozil is preferred over other fibrates (e.g., fenofibrate and clofibrate) in this setting because gemfibrozil is partly cleared by the liver; as such, it carries a lower risk of drug-induced myopathy than do fibrates that are cleared by the kidneys. Nonetheless, because gemfibrozil is partially excreted renally, the drug should be administered in the lowest effective dose. Nicotinic acid and statins have been less well studied in this condition. Combination therapy with nicotinic acid, statins, or gemfibrozil may be necessary to attain the therapeutic goal.

GASTROINTESTINAL DISORDERS THAT CAUSE DYSLIPIDEMIA

Primary biliary cirrhosis is the most significant gastrointestinal cause of dyslipidemia. In the early stages of primary biliary cirrhosis, when some hepatocellular function remains, mild elevations of VLDL and LDL levels occur because of elevations in the levels of remnant lipoproteins and HDL. Terminal liver disease with cirrhosis results in severe elevation in cholesterol levels because of increased production of lipoprotein X—an abnormal lipoprotein particle containing albumin and other plasma components that is rich in free cholesterol and phospholipid. Treatment of this terminal disorder requires liver transplantation.

OTHER CAUSES OF SECONDARY DYSLIPIDEMIA

Many commonly used drugs have adverse effects on lipoproteins [*see Table 2*]. Discontinuance of the drug often will improve lipid levels. An increase in VLDL, LDL, and HDL cholesterol levels can result from the use of drugs for the prevention of rejection after organ transplantation. Pravastatin is the drug of choice for lowering LDL levels because of its unique catabolic pathways. Immunosuppressive agents such as cyclosporine compete with atorvastatin and simvastatin for the cytochrome P-450 3A4 system. The use of antifungal agents also can interfere with the metabolism of these statins. The predominant dyslipidemia that is seen in patients with AIDS is similar to the dyslipidemia that occurs in patients with the metabolic syndrome; mild hypertriglyceridemia is common, and low HDL cholesterol is seen in some patients.[52] In others, extreme hypertriglyceridemia can result from the use of HIV drugs, and the resultant hypertriglyceridemia may be associated with pancreatitis. The etiology of dyslipidemia in AIDS is complex: excessive free fatty acid mobilization is seen, along with the development of lipodystrophy and insulin resistance. In addition, AIDS patients typically use dyslipidemia-causing drugs. The specific therapy in each patient needs to be individualized.

Prevention and Treatment of Coronary Artery Disease

PRIMARY PREVENTION

The treatment of lipid disorders in individuals who do not have clinical evidence of CAD is considered primary prevention. Primary prevention is based on the assumption that modification of lipid risk factors will alter the natural history of the untreated condition—the so-called lipid hypothesis. An association between cholesterol and CAD has been known since the early 1950s; however, it was not until the publication of the Lipid Research Clinics Coronary Primary Prevention Trial (LRC-CPPT) in 1984 that there were data to support the lipid hypothesis.

The LRC-CPPT enrolled almost 4,000 men with moderate hypercholesterolemia; patients were followed for 7 years. The treat-

ment group was prescribed cholestyramine, which resulted in LDL cholesterol levels being 12.6% lower than those of the control subjects, who were given placebo. The cholestyramine group had a 19% reduction in CAD deaths and nonfatal myocardial infarcts ($P < 0.05$), although no decrease in total mortality was observed.[53] Further analysis demonstrated that the extent of benefit depended upon the achieved reduction in serum cholesterol (reflecting drug compliance). Use of a proportional hazards model indicated that a 25% decrease in total cholesterol or a 35% decrease in LDL cholesterol would be expected to decrease the risk of a CAD event by 50%.[54]

The Helsinki Heart Study used the fibrate gemfibrozil to treat dyslipidemic men without CAD. After 6 years of follow-up, a 34% reduction in CAD events was seen in the treatment group, compared with the group receiving placebo.[55] Again, no decrease in CAD mortality was demonstrated.

In both the Helsinki Heart Study and the LRC-CPPT, the sample size was calculated on the power to detect CAD events, not on fatal outcomes alone. As such, the lack of an effect on mortality was not surprising, but an increase (not statistically significant) in noncoronary death in the treatment groups of both these studies was troublesome[53-55] and confounded the recommendations for primary preventive therapy in hypercholesterolemic patients. These concerns were not completely addressed until 1995, when results of the West of Scotland Coronary Prevention Study (WOSCOPS) were published.[56]

The WOSCOPS trial evaluated the effect of 5 years of treatment with pravastatin on the incidence of nonfatal myocardial infarction (MI) and CAD deaths in 6,595 men. The men were middle-aged (45 to 64 years of age) and moderately hypercholesterolemic (LDL cholesterol level above 155 mg/dl). The treatment group manifested a 20% reduction in total cholesterol, a 26% reduction in LDL cholesterol, a 12% decrease in triglycerides, and a 5% increase in HDL cholesterol, as compared with the control group. On the basis of intention-to-treat principles, these changes were associated with a 31% risk reduction in nonfatal MI or CAD deaths ($P < 0.001$), a 32% risk reduction in all cardiovascular deaths ($P = 0.033$), and a 22% risk reduction in total mortality ($P = 0.051$). In addition, coronary interventions (i.e., angiography, angioplasty, and coronary artery bypass surgery) were reduced 31% to 37% ($P < 0.01$).

The reduction in clinical events began within 6 months of randomization and were independent of other risk factors, such as diabetes, smoking, blood pressure, family history of CAD, and the ratio of total cholesterol to HDL cholesterol.[57] Although there was no risk reduction without a decrease in LDL cholesterol, a decrease in LDL cholesterol of approximately 24% was adequate to see the full benefit of treatment. The treatment effect was proportionately the same regardless of baseline lipid levels and the reduction in LDL cholesterol. As such, LDL reduction alone did not account for all the benefits of treatment with pravastatin.[58] Importantly, there was no increase in noncoronary deaths, which was reported in earlier primary preventive trials.

The Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCAPS) was the first large primary intervention trial to study the effects of cholesterol lowering in individuals with average cholesterol levels.[59] That is, the mean total and LDL cholesterol levels were nearer the average value for the general population (221 mg/dl and 150 mg/dl, respectively). In addition, it was the first large study to include women (997 of a total of 6,605 patients). Lovastatin was the treatment agent in this randomized, placebo-controlled trial. LDL cholesterol reduction was 25%, and follow-up was 5 years. The total absolute benefit was 2%, meaning that 50 patients had to be treated for 5 years to prevent one event. The treatment group had a 28% reduction in cardiovascular hospitalizations, a 23% decrease in angioplasty, and a 32% reduction in coronary bypass surgery. An analysis of the cost-effectiveness of lovastatin treatment demonstrated a 27% (or $524 per patient) reduction in cardiovascular health care costs for the lovastatin group, as compared with the group that received placebo.[60]

Persons with average cholesterol levels were also evaluated in the Anglo-Scandinavian Cardiac Outcomes Trial–Lipid Lowering Arm (ASCOT-LLA).[61] Nearly 20,000 hypertensive patients were randomized to one of two antihypertensive regimens. The lipid-lowering arm of this study randomized 10,305 patients with total cholesterol levels of 251 mg/dl or lower to treatment with atorvastatin or placebo. The study was halted early because a significant benefit was observed in the treatment group. Median follow-up was 3.3 years. The study did not demonstrate statistically significant reductions in cardiovascular or all-cause mortality; however, significant reductions were seen in total coronary events, total cardiovascular events and procedures, and stroke.[61]

The effect of atorvastatin (10 mg/day) on primary prevention of cardiovascular disease in diabetic patients was examined in the Collaborative Atorvastatin Diabetes Study (CARDS).[45] CARDS randomized almost 3,000 diabetic patients with LDL levels of 160 mg/dl or lower, triglyceride levels of 600 mg/dl or lower, and at least one of the following: retinopathy, albuminuria, smoking habit, or hypertension. The trial was stopped 2 years early because predetermined criteria had been met. The atorvastatin group demonstrated a 36% reduction in coronary events, a 31% reduction in coronary revascularization procedures, a 48% decrease in stroke, and a 27% reduction in all-cause deaths, as compared with the placebo group.

These studies support lipid-lowering therapy as primary prevention for patients with high LDL and average LDL values. There are virtually no data on primary prevention in patients with other lipid abnormalities, such as isolated low HDL cholesterol levels or elevated triglyceride levels.

SECONDARY PREVENTION

Lipid-lowering therapy in patients with documented CAD is considered secondary prevention. Lipid levels have a significant influence on CAD death rates in those with and without CAD; however, the impact is significantly greater in patients with established CAD.[62]

Several trials have investigated the effect of aggressive lifestyle intervention in patients with CAD. The Saint Thomas Atherosclerosis Regression Study (STARS) randomized men with CAD and total cholesterol levels above 232 mg/dl to conventional care or a low-fat, low-cholesterol diet. Despite relatively modest changes in lipid levels (the intervention group had an LDL cholesterol average of 162 mg/dl), the progression of CAD decreased and the rate of regression increased in the intervention group. Angina symptoms also improved.[63]

The effects of a Mediterranean diet (increased α-linoleic acid) were compared with those of a prudent Western diet in the Lyon Diet Heart Study.[64] All study participants had had a first MI. Those consuming the Mediterranean diet had lower rates of primary (death and MI) and secondary (unstable angina, stroke, heart failure) end points than those on the prudent Western diet at 27 months. This effect persisted after 4 years of follow-up. The group on the Mediterranean diet had a rate of combined prima-

ry and secondary end points of 2.59 events per 100 patients per year, compared with 9.03 events per 100 patients per year in the group on the prudent diet.[65]

A variety of pharmacologic agents have been used alone and in combination in secondary prevention trials. Some trials have used angiographic end points in assessing progression or regression of CAD, whereas others have used clinical end points. The Familial Atherosclerosis Treatment Study (FATS) examined the effect of several lipid-reducing regimens in men with elevated apo B levels. The two most aggressive regimens (nicotinic acid–colestipol and lovastatin-colestipol) were equally effective; both regimens were associated with delayed progression (21% and 25%, respectively, versus 46% in the placebo-colestipol group) and an increased likelihood of regression of coronary artery stenoses (32% and 39%, respectively, versus 11% in the placebo-colestipol group). Clinical end points (death, MI, worsening angina, and revascularization) were also reduced in the more aggressively treated groups (4.2% and 6.5%, respectively, versus 19% in the placebo-colestipol group).[44] This was the first major study to document the regression of CAD with aggressive lipid-lowering therapy. A subsequent analysis of these patients correlated the change in CAD severity with therapy-induced changes in LDL buoyancy and hepatic lipase activity.[66]

The Scandinavian Simvastatin Survival Study (4S) evaluated 4,444 patients with known CAD and moderate to severe hypercholesterolemia at baseline (total cholesterol concentration ranging from 212 to 309 mg/dl).[67] Patients were randomized to a regimen of diet plus simvastatin or diet plus placebo. At 5.4 years, there was a significant reduction in total mortality (8% on simvastatin versus 12% on placebo), major coronary events (19% versus 28%), CAD deaths (42% reduction), and cerebrovascular events (2.7% versus 4.3%). The reduction in cardiovascular events correlated with total cholesterol and LDL cholesterol levels and with changes from baseline.[68]

The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) trial randomized approximately 9,000 men and women with a history of recent MI or unstable angina to receive either placebo or pravastatin.[69] The study was stopped prematurely at 60 months because of a significant benefit associated with pravastatin therapy. CAD death was reduced in the treatment arm of the study (6.4% versus 8.3%), as were total mortality (11% versus 14%), stroke (20% relative decrease), need for bypass surgery (8.9% versus 11.3%), and MI (7.4% versus 10.1%). The benefit was primarily related to changes in lipid levels and was seen in all predefined subgroups. The greatest reduction in coronary events was seen in those patients thought to be at highest risk, as assessed by concomitant risk factors.[70]

The Cholesterol and Recurrent Events (CARE) trial evaluated 4,159 patients with relatively low lipid levels. The average total cholesterol level was 209 mg/dl, and the average LDL cholesterol level was 139 mg/dl. Treatment with pravastatin over 5 years resulted in significant reductions in coronary death or nonfatal myocardial infarction (10.2% versus 13.2% for placebo), need for revascularization (14.1% versus 18.8%), and frequency of stroke (2.6% versus 3.8%).[71] However, in contrast to the results seen with 4S and LIPID, the absolute or percentage reductions in LDL had little relationship to coronary events.[72] The benefits were seen only in patients with LDL levels above 125 mg/dl.[72]

The Heart Protection Study enrolled over 20,000 persons with a history of cardiovascular disease (coronary, cerebrovascular, or peripheral vascular disease), diabetes, or treated hypertension.[73] As such, it was a mixture of primary and secondary intervention. One third of the individuals had baseline LDL cholesterol levels below 116 mg/dl, and 25% had initial LDL levels ranging from 116 to 135 mg/dl. Participants were randomized to receive simvastatin or placebo. After an average follow-up of 5.5 years, the lipid-lowering group showed a 24% reduction in major cardiovascular events, an 18% reduction in cardiovascular deaths, and a 13% reduction in all-cause mortality, as compared with the placebo group. The percentage reductions in events were similar in all three tertiles of baseline LDL cholesterol levels and in patients with LDL cholesterol levels below 100 mg/dl at baseline. These results differ from those reported in the CARE study, but they are consistent with results from the 4S and LIPID trials. The results of the Heart Protection Study also suggest that there may not be a threshold beyond which increased LDL-lowering therapy ceases to improve outcome, at least in patients at high risk for recurrent coronary events.

Aggressive LDL-lowering therapy appears to be more effective than standard lipid-lowering treatment. The Pravastatin or Atorvastatin Evaluation and Infection Therapy (PROVE-IT) trial compared standard LDL-lowering treatment (pravastatin, 40 mg daily) with intensive LDL-lowering treatment (atorvastatin, 80 mg daily) in more than 4,000 patients recently hospitalized with an acute coronary syndrome.[74] The average follow-up was 24 months. The median LDL cholesterol level achieved with atorvastatin was 62 mg/dl, compared with 96 mg/dl in the group treated with pravastatin. The primary composite end point was death from any cause, MI, unstable angina not requiring hospitalization, coronary revascularization, and stroke. The rate of reaching the primary end point was 22.4% in the atorvastatin group and 26.3% in the pravastatin group. The benefit of aggressive therapy with atorvastatin was apparent as early as 30 days after initiating therapy and was consistent over time.

The Reversal of Atherosclerosis with Aggressive Lipid Lowering (REVERSAL) trial also compared moderate LDL-lowering therapy (pravastatin, 40 mg daily) with more intensive LDL-lowering therapy (atorvastatin, 80 mg daily).[75] The study used coronary intravascular ultrasound, a sensitive means of measuring plaque volume, as a baseline measurement and primary end point. The median percentage change in atheroma volume was –0.4% in the atorvastatin group, compared with +2.7% in the pravastatin group. This finding correlated with mean LDL cholesterol levels of 79 mg/dl in the atorvastatin group and 110 mg/dl in the pravastatin group. These results gave further support to the view that aggressive LDL-lowering therapy is superior to standard LDL-lowering therapy.

Few studies have examined the benefit of raising HDL cholesterol levels in the secondary prevention of CAD. The VA-HIT trial enrolled 2,531 patients with known CAD and with LDL cholesterol levels below 140 mg/dl, HDL choleserol levels of 40 mg/dl or above, and triglyceride levels of 300 mg/dl or below.[19] The patients were randomized to receive gemfibrozil or placebo. The subsequent mean HDL cholesterol level in the gemfibrozil group was 6% higher than that in the placebo group, and the mean triglyceride level in the gemfibrozil group was 31% lower. The mean LDL cholesterol levels were 113 mg/dl in both groups. The combined primary end point of cardiac death and nonfatal myocardial infarction was 17% in the gemfibrozil group and 22% in the placebo group (relative risk reduction, 22%). The beneficial effect of gemfibrozil did not become apparent until 2 years after randomization.

Combination therapy using a statin to lower LDL cholesterol levels and niacin to raise HDL cholesterol levels has been shown to provide increased cardioprotection. In one study, patients

were randomized to one of four groups: simvastatin plus niacin, vitamins, simvastatin-niacin plus antioxidants, or placebos. At entry, the HDL choleseterol level was below 35 mg/dl, and the LDL cholesterol level was below 145 mg/dl. The mean LDL and HDL cholesterol levels were unaltered in the antioxidant and placebo groups but were changed significantly in the simvastatin plus niacin groups (mean LDL cholesterol level reduced by 42% and mean HDL cholesterol level raised by 26%). At 3 years, the reduction of clinical events in the simvastatin and niacin groups was greater than that which is usually reported in studies of statins alone (relative risk, 0.1 to 0.4 compared with placebo), suggesting that a benefit may be associated with the elevation of HDL cholesterol levels. The antioxidants provided no additional benefit and may even have attenuated the benefits of combination therapy.[76]

RISK STRATIFICATION

CAD risk factors seldom occur in isolation, and the risk associated with each varies widely in combination with other risk factors. The variability in risk prompted the NCEP ATPIII to standardize guidelines for risk assessment of CAD. Over time, the guidelines were revised to recommend more aggressive lipid-lowering targets as a means of reducing CAD risk. This evolution in guidelines is the result of consistently emerging data that extend our understanding of dyslipidemia, associated risk factors and their relationship to CAD, and the utility of new therapeutic options.

The ATP guidelines focus primarily on LDL cholesterol levels as the major lipid risk factor. More recently, low HDL levels have become a factor in risk assessment. In ATPIII, the metabolic syndrome was added as a risk factor in an attempt to assess risk for CAD in centrally obese patients who have modest elevations in triglyceride levels, low HDL cholesterol levels, and small, dense LDL particles, as well as type 2 diabetes mellitus or FCHL. In an effort to better identify those at highest risk for CAD events, the NCEP recognizes several CAD equivalents. They include diabetes mellitus, peripheral arterial disease, abdominal aortic aneurysm, symptomatic carotid artery disease, and multiple risk factors that confer a 10-year risk of CAD greater than 20%.[15] The presence of these CAD equivalents requires a level of therapeutic aggressiveness equal to that recommended for patients with established CAD.

The American College of Physicians (ACP) has adopted a somewhat less aggressive recommendation for treatment of individuals with type 1 diabetes mellitus. The ACP reserves the use of statins for patients with type 2 diabetes and other CAD risk factors.[77] The ATPIII guidelines do not differentiate between the risk of CAD in patients with type 1 diabetes and that in patients with type 2 diabetes. An argument can be made that the CAD risk is greater in type 2 diabetes and that treatment guidelines should differentiate between these entities.

The ATPIII guidelines use the Framingham scoring system for estimating the 10-year risk of CAD. Some studies indicate that the Framingham score overestimates risk in Japanese-American and Hispanic men, Native-American women, and some European and Asian populations.[78-80] It also has been suggested that the Framingham score weights age too heavily as a risk factor. The Pravastatin in Elderly Individuals at Risk of Vascular Disease (PROSPER) study, which is the only prospective study to assess statin therapy in men and women older than 70 years, demonstrated that statin therapy was of no benefit in those without preexisting atherosclerosis.[81] Any age bias present in the Framingham scoring system is eliminated when the system is used to predict risk in nonelderly patients.

A multicenter, international study confirmed the validity of risk stratification. In this study of over 15,000 patients with acute MI from 52 countries, over 90% of the population-attributable risk could be accounted for by nine potentially modifiable risk factors.[82] Most of the risk is accounted for by an elevated apo B to apo A1 ratio, smoking, hypertension, and diabetes. These risk factors were more important in younger than in older individuals. As such, principles of cardiovascular disease prevention are similar worldwide and have the potential to have a major impact.

DRUG THERAPY IN DYSLIPIDEMIA

Drugs Used to Lower LDL Cholesterol Levels

Several classes of drugs lead to a reduction in LDL cholesterol levels [*see Table 4*].[16] Before the introduction of statins in the mid-1980s, the major drugs used for this purpose were the bile-acid sequestrants and niacin. The introduction of statins, with their powerful effects on LDL cholesterol, their tolerability, and their relative lack of toxicity, provided a significant advance in the management of patients with hypercholesterolemia. The introduction of intestinally active drugs has provided additional approaches both for monotherapy—especially for individuals who are unable to tolerate statins—and, more particularly, for combination therapy.

Statins Several statins are now available, and new ones continue to be introduced. To date, statins have been highly effective in clinical trials in reducing clinical events, including stroke. Although some of the benefits of statins have been attributed to the so-called pleotropic effects of this class of drugs, the extent of reduction in LDL cholesterol levels nonetheless appears to be the major determinant of risk reduction.

Intestinally active compounds Bile-acid sequestrants were among the earliest drugs to become available for the treatment of hypercholesterolemia, and they were the first class of drugs to demonstrate that the reduction of LDL cholesterol was associated with reduced risk of CAD; however, their use was limited by their very poor tolerability and their modest effect in reducing LDL cholesterol. Moreover, triglyceride levels tend to increase with their use in patients with high baseline plasma triglyceride levels. The introduction of a more tolerable bile-acid sequestrant, colesevelam, resulted in improved compliance with this class of drugs, especially when used in combination therapy in patients with very high LDL cholesterol levels (e.g., for patients with FH).

Unlike bile-acid sequestrants, the intestinally active drug ezetimibe directly inhibits cholesterol absorption. Although clinical data are limited, it appears ezetimibe is able to reduce LDL cholesterol by approximately 20%, whether used as monotherapy or in combination with other lipid-lowering agents.[83] In addition, ezetimibe does not cause an increase in plasma triglyceride levels, as occurs with bile-acid sequestrants. Ezetimibe has not yet been reported in clinical trials with cardiovascular end points.

Drugs Used Primarily to Lower Triglyceride Levels

The preferred drugs for treatment of hypertriglyceridemia are the fibrates and niacin. Niacin is the best drug currently available for raising HDL cholesterol levels. It also produces modest re-

Table 4 Drug Treatment of Lipid Disorders

Drug	*Dosage*	*Cost per Month*	*Comment*
Bile acid–binding resins	Start with one packet (2 g for colestipol tabs) b.i.d., increase over 1–2 wk to desired dose		For elevated LDL, normal triglycerides; take other drugs 1 hr before or 4 hr after; may be used with nicotinic acid, statins, or fibrates
Cholestyramine	Maximum 24 g/day b.i.d. or t.i.d.	$69	
Colestipol	Maximum 30 g/day b.i.d. or t.i.d.	$305	t.i.d. more effective
Colestipol tablets	Maximum 16 g/day	$267	
Colesevelam	Three 625 mg tablets b.i.d. with meals or six tablets/day with a meal; maximum seven tablets/day	$142	Better tolerated than other resins
Ezetimibe	10 mg/day	$72	Can reduce LDL cholesterol by ~20% without increasing plasma triglyceride levels
Fenofibrate	200 mg/day	$73	For elevated triglycerides and patients in whom both LDL and triglycerides are elevated; may be used with bile acid–binding resins or nicotinic acid; decrease dose with severe renal disease
Gemfibrozil	600 mg b.i.d. before meals	$17	
Niacin	Start with 250 mg q.d. with meals; increase to 1.5–2.0 g/day; maximum 6 g/day	$28	For elevated LDL, triglycerides, or both; may be used with bile acid–binding resins or fibrates
Statins			For elevated LDL; possibly useful for patients in whom both LDL and triglycerides are elevated; may be used with bile acid–binding resins
Atorvastatin	Start with 10 mg/day; maximum 80 mg/day	$95	
Fluvastatin	Start with 20 mg b.i.d. or at bedtime; maximum 80 mg/day	$68	
Lovastatin	Start with 20 mg b.i.d. or with dinner; maximum 80 mg/day	$126	
Pravastatin	Start with 10 mg at bedtime; maximum 40 mg/day	$120	May be used with drugs that are cleared by hepatic enzymes CYP450, CYP3A4
Simvastatin	Start with 10 mg at bedtime; maximum 40 mg/day	$124	

ductions in LDL and reduces apo B levels, but because it worsens insulin sensitivity, its use in patients with type 2 diabetes mellitus is limited. Fibrates are the drugs of choice for patients with marked hypertriglyceridemia, for whom the primary goal of therapy is the prevention of pancreatitis and other features of the chylomicronemia syndrome. They also are of use in hypertriglyceridemic states (e.g., the familial forms of hypertriglyceridemia and in some patients with diabetic dyslipidemia), especially when triglyceride levels are more than mildly elevated. Fibrates also have a modestly beneficial effect on HDL cholesterol levels. Both fibrates and niacin are useful in combination therapy, primarily with statins.

Omega-3 fatty acids (e.g., those found in marine oils) have been used for the treatment of hypertriglyceridemia, especially when other modalities of therapy have failed to reduce markedly elevated levels of triglycerides.

Combination Therapy

Combinations of drugs often need to be used when both LDL cholesterol and triglyceride levels are elevated. Combination therapy also is of use when monotherapy, especially with statins, fails in achieving target lipid and lipoprotein levels, especially LDL cholesterol levels. Commonly used combinations include statins and fibrates—although little is known of their additive benefit in reducing clinical events—and statins and niacin. Statins and bile-acid sequestrants also are a useful combination, and the use of the new cholesterol absorption inhibitors with other classes of drugs, particularly statins, is likely to be of value. In some cases, triple therapy (e.g., statins, niacin, and an intestinally active agent) is required.

Special Issues in the Management of Dyslipidemia

SCREENING FOR HYPERCHOLESTEROLEMIA IN CHILDREN

Numerous autopsy studies demonstrate that coronary atherosclerosis begins in childhood and adolescence and that lipoprotein levels are consistently associated with the extent of such atherosclerosis. Children in families with FH and early CAD have higher cholesterol levels, and childhood cholesterol levels are significant predictors of adult levels. However, a significant proportion of children and adolescents who have mildly elevated cholesterol level will not as adults develop cholesterol levels high enough to warrant intervention; screening all children for high cholesterol would risk labeling many young people as diseased. All children older than 2 years would benefit from a diet that is low in saturated fat; this goal should be a part of any population strategy for controlling epidemic atherosclerosis. However, the safety and efficacy of long-term drug therapy have not been established in this age group, and treatment must be approached cautiously.

Considering these and other issues, the recommendations of the NCEP's Expert Panel on Blood Cholesterol Levels in Children and Adolescents seem appropriate.[84] Physicians should advise patients younger than 55 years who have a known CAD or a lipid disorder that their children or grandchildren should undergo regular cholesterol testing, and patients with a genetically well-defined lipid disorder should obtain appropriate genetic counseling. Physicians who care for patients younger than 20 years who have markedly elevated LDL levels should exhaust all lifestyle interventions before considering medications. If such measures are ineffective, resins should be used, and referral to a specialty clinic should be considered.

Treatment of young adults with elevated cholesterol levels is controversial. The strategy of matching the intensity of intervention with the level of risk of atherosclerosis has been proposed, but for young adults, a short-term (e.g., 10-year) risk assessment may be inadequate for estimating the potential benefit of cholesterol lowering. It is incorrect to argue that all treatment can be safely deferred to later life or until the occurrence of an atherosclerotic event. Population-level prevention and lifestyle interventions should still be favored for young adults, but advances in technology that better enable the identification of asymptomatic patients (of any age) who should take steps to reduce risk are greatly needed. Such advances may make it possible to reliably identify or quantify vulnerable plaques; markers of inflammation; or noninvasive measurements of endothelial dysfunction.

MANAGING DYSLIPIDEMIA IN WOMEN

Before menopause, women have a lower incidence of CAD than men of the same age. Although rare, CAD does occur in premenopausal women, usually in association with multiple genetic and environmental risk factors, such as in patients with familial forms of dyslipidemia or in diabetic patients who smoke cigarettes.

After menopause, some women develop the metabolic syndrome, characterized by visceral obesity, insulin resistance, hypertension, and dyslipidemia.[21] There is some evidence that estrogen replacement therapy can reverse these findings. However, the Women's Health Initiative Study demonstrated that combined oral estrogen and progesterone did not protect women from CAD and that it in fact had adverse effects.[85] The estrogen-alone component of the Women's Health Initiative Study indicated that estrogen therapy carried modest risk of CAD (i.e., MI or CAD death); the study was halted prematurely because of increased risk of stroke.[49]

MANAGING DYSLIPIDEMIA IN OLDER PATIENTS

Age is the most significant risk factor for the development of atherosclerosis. CAD is currently a major cause of disability and mortality in older populations; however, the relative risk associated with any single coronary risk factor decreases with age because of the comordid conditions and noncardiovascular mortality that affect an aging population. One implication of the complex relationships between risk factors and comorbid conditions in the pathogenesis of coronary-related events in the elderly is represented by the multiple effects of treatment of single risk factors, such as the decrease in LDL cholesterol levels and inflammation markers yielded by statins. A growing body of evidence from clinical trials indicates that statin therapy is effective in the elderly; lipid-lowering therapy is probably indicated in this population in persons who are at high risk for atherosclerosis or who have preexisting atherosclerosis.[73,81] Primary intervention with drug therapy in persons not at high atherosclerotic risk is controversial. In the PROSPER trial of persons older than 70 years, no benefit was seen with statin therapy in those who did not have preexisting clinical atherosclerosis. Indeed, there was a suggestion of increased gastrointestinal cancer with statin therapy in these elderly patients.[81] Attention to other concomitant diseases and the nutritional state, as well as to capabilities of the elderly, are important considerations in the management of older patients with dyslipidemia.[86]

John D. Brunzell, M.D., has served as consultant for GlaxoSmithKline, Merck, Abbott Laboratories, and Novartis Pharmaceuticals Corp.

R. Alan Failor, M.D., has served as a consultant for GlaxoSmithKline.

References

1. Havel R, Kane J: Introduction: structure and metabolism of plasma lipoproteins. The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2705

2. Brunzell JD, Chait A, Bierman EL: Pathophysiology of lipoprotein transport. Metabolism 27:1109, 1978

3. Brunzell JD, Albers JJ, Chait A, et al: Plasma lipoproteins in familial combined hyperlipidemia and monogenic familial hypertriglyceridemia. J Lipid Res 24:147, 1983

4. Kane JP, Havel RJ: Disorders of the biogenesis and secretion of lipoproteins containing the B apolipoproteins. The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001

5. Utermann G: Lipoprotein (a). The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CS, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2753

6. Brunzell J, Deeb S: Familial lipoprotein lipase deficiency, apo CII deficiency, and hepatic lipase deficiency. The Metabolic and Molecular Bases of Inherited Disease, 8th ed., Vol 2. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2789

7. Mahley R, Rall S: Type III hyperlipoproteinemia (dysbetalipoproteinemia): the role of apolipoprotein E in normal and abnormal lipoprotein metabolism. The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2835

8. Goldstein JL, Hobbs HH, Brown MS: Familial hypercholesterolemia. The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CS, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2863

9. Tall A, Breslow J, Rubin E: Genetic disorders affecting plasma high-density lipoproteins. The Metabolic and Molecular Bases of Inherited Disease, 8th ed. Scriver CS, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill Book Co, New York, 2001, p 2915

10. Deeb SS, Zambon A, Carr MC, et al: Hepatic lipase and dyslipidemia: interactions among genetic variants, obesity, gender, and diet. J Lipid Res 44:1279, 2003

11. Frikke-Schmidt R, Nordestgaard BG, Jensen GB, et al: Genetic variation in ABC transporter A1 contributes to HDL cholesterol in the general population. J Clin Invest 114:1343, 2004

12. Carr MC, Brunzell JD: Abdominal obesity and dyslipidemia in the metabolic syndrome: importance of type 2 diabetes and familial combined hyperlipidemia in coronary artery disease risk. J Clin Endocrinol Metab 89:2601, 2004

13. Griffin BA: Lipoprotein atherogenicity: an overview of current mechanisms. Proc Nutr Soc 58:163, 1999

14. Steinberg D, Pearson TA, Kuller LH: Alcohol and atherosclerosis. Ann Intern Med 114:967, 1991

15. Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. JAMA 285:2486, 2001

16. Brown CD, Higgins M, Donato KA, et al: Body mass index and the prevalence of hypertension and dyslipidemia. Obes Res 8:605, 2000

17. Austin MA, McKnight B, Edwards KL, et al: Cardiovascular disease mortality in familial forms of hypertriglyceridemia: a 20-year prospective study. Circulation 101:2777, 2000

18. Knopp RH: Drug treatment of lipid disorders. N Engl J Med 341:498, 1999

19. Rubins HB, Robins SJ, Collins D, et al: Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. Veterans Affairs High-Density Lipoprotein Cholesterol Intervention Trial Study Group. N Engl J Med 341:410, 1999

20. Coronary Drug Project Research Group: Clofibrate and niacin in coronary heart disease. JAMA 231:360, 1975

21. Carr MC: The emergence of the metabolic syndrome with menopause. J Clin Endocrinol Metab 88:2404, 2003

22. Lemieux S, Despres JP, Moorjani S, et al: Are gender differences in cardiovascular disease risk factors explained by the level of visceral adipose tissue? Diabetologia 37:757, 1994

23. Nieves D, Cnop M, Retzlaff B, et al: The atherogenic lipoprotein profile associated with obesity and insulin resistance is largely attributable to intra-abdominal fat. Diabetes 52:172, 2003

24. Purnell JQ, Kahn SE, Albers JJ, et al: Effect of weight loss with reduction of intra-abdominal fat on lipid metabolism in older men. J Clin Endocrinol Metab 85:977, 2000

25. Grundy SM: Approach to lipoprotein management in 2001 National Cholesterol Guidelines. Am J Cardiol 90:11i, 2002

26. Cefalu WT, Wang ZQ, Werbel S, et al: Contribution of visceral fat mass to the insulin resistance of aging. Metabolism 44:954, 1995

27. Bermudez OI, Tucker KL: Total and central obesity among elderly Hispanics and the association with type 2 diabetes. Obes Res 9:443, 2001

28. Lempiainen P, Mykkanen L, Pyorala K, et al: Insulin resistance syndrome predicts coronary heart disease events in elderly nondiabetic men. Circulation 100:123, 1999

29. Mykkanen L, Kuusisto J, Haffner SM, et al: Hyperinsulinemia predicts multiple atherogenic changes in lipoproteins in elderly subjects. Arterioscler Thromb 14:518, 1994

30. Cefalu WT, Werbel S, Bell-Farrow AD, et al: Insulin resistance and fat patterning with aging: relationship to metabolic risk factors for cardiovascular disease. Metabolism 47:401, 1998

31. Ginsberg HN: Insulin resistance and cardiovascular disease. J Clin Invest 106:453, 2000

32. Ford E, Giles W, Dietz W: Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. JAMA 287:356, 2002

33. Alexander CM, Landsman PB, Teutsch SM, et al: NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. Diabetes 52:1210, 2003

34. Lamarche B, Moorjani S, Cantin B, et al: Associations of HDL2 and HDL3 subfractions with ischemic heart disease in men. Arterioscler Thromb Vasc Biol 17:1098, 1997

35. Williams RR, Hopkins PN, Hunt SC, et al: Population-based frequency of dyslipidemia syndromes in coronary-prone families in Utah. Arch Intern Med 150:582, 1990

36. Goldstein JL, Hazzard WR, Schrott HG, et al: Hyperlipidemia in coronary heart disease. I. Lipid levels in 500 survivors of myocardial infarction. J Clin Invest 52:1533, 1973

37. Goldstein JL, Schrott HG, Hazzard WR, et al: Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. J Clin Invest 52:1544, 1973

38. Hazzard WR, Goldstein JL, Schrott MG, et al: Hyperlipidemia in coronary heart disease. III. Evaluation of lipoprotein phenotypes of 156 genetically defined survivors of myocardial infarction. J Clin Invest 52:1569, 1973

39. Ayyobi AF, McGladdery SH, McNeely MJ, et al: Small, dense LDL and elevated apolipoprotein B are the common characteristics for the three major lipid phenotypes of familial combined hyperlipidemia. Arterioscler Thromb Vasc Biol 23:1289, 2003

40. Jarvik GP, Brunzell JD, Austin MA, et al: Genetic predictors of FCHL in four large pedigrees. Influence of ApoB level major locus predicted genotype and LDL subclass phenotype. Arterioscler Thromb 14:1687, 1994

41. Austin MA, Horowitz H, Wijsman E, et al: Bimodality of plasma apolipoprotein B levels in familial combined hyperlipidemia. Atherosclerosis 92:67, 1992

42. Babirak SP, Brown BG, Brunzell JD, et al: Familial combined hyperlipidemia and abnormal lipoprotein lipase. Arterioscler Thromb 12:1176, 1992

43. Purnell JQ, Kahn SE, Schwartz RS, et al: Relationship of insulin sensitivity and ApoB levels to intra-abdominal fat in subjects with familial combined hyperlipidemia. Arterioscler Thromb Vasc Biol 21:567, 2001

44. Brown G, Albers JJ, Fisher LD, et al: Regression of coronary artery disease as a result of intensive lipid-lowering therapy in men with high levels of apolipoprotein B. N Engl J Med 323:1289, 1990

45. Colhoun HM, Betteridge DJ, Durrington PN, et al: Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicentre randomised placebo-controlled trial. Lancet 364:685, 2004

46. Marks D, Thorogood M, Neil HA, et al: A review on the diagnosis, natural history, and treatment of familial hypercholesterolaemia. Atherosclerosis 168:1, 2003

47. Grundy SM, Vega GL, McGovern ME: Efficacy, safety, and tolerability of once-daily niacin for the treatment of dyslipidemia associated with type 2 diabetes: results of the assessment of diabetes control and evaluation of the efficacy of Niaspan trial. Arch Intern Med 162:1568, 2002

48. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial. JAMA 273:199, 1995

49. Anderson GL, Limacher M, Assaf AR, et al: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 291:1701, 2004

50. Joven J, Villabona C, Vilella E, et al: Abnormalities of lipoprotein metabolism in patients with the nephrotic syndrome. N Engl J Med 323:579, 1990

51. Samuelsson O, Attman PO, Knight-Gibson C, et al: Effect of gemfibrozil on lipoprotein abnormalities in chronic renal insufficiency: a controlled study in human chronic renal disease. Nephron 75:286, 1997

52. Khovidhunkit W, Memon RA, Feingold KR, et al: Infection and inflammation-induced proatherogenic changes of lipoproteins. J Infect Dis 181:S462, 2000

53. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. JAMA 251:351, 1984

54. The Lipid Research Clinics Coronary Primary Prevention Trial results. II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. JAMA 251:365, 1984

55. Frick MH, Elo O, Haapa K, et al: Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia: safety of treatment, changes in risk factors, and incidence of coronary heart disease. N Engl J Med 317:1237, 1987

56. Shepherd J, Cobbe SM, Ford I, et al: Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. N Engl J Med 333:1301, 1995

57. Baseline risk factors and their association with outcome in the West of Scotland Coronary Prevention Study. The West of Scotland Coronary Prevention Study Group. Am J Cardiol 79:756, 1997

58. Influence of pravastatin and plasma lipids on clinical events in the West of Scotland Coronary Prevention Study (WOSCOPS). Circulation 97:1440, 1998

59. Downs JR, Clearfield M, Weis S, et al: Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. JAMA 279:1615, 1998

60. Gotto AM Jr, Boccuzzi SJ, Cook JR, et al: Effect of lovastatin on cardiovascular resource utilization and costs in the Air Force/Texas Coronary Atherosclerosis Prevention Study (AFCAPS/TexCAPS). AFCAPS/TexCAPS Research Group. Am J Cardiol 86:1176, 2000

61. Sever PS, Dahlof B, Poulter NR, et al: Prevention of coronary and stroke events with atorvastatin in hypertensive patients who have average or lower-than-average cholesterol concentrations, in the Anglo-Scandinavian Cardiac Outcomes Trial—Lipid Lowering Arm (ASCOT-LLA): a multicentre randomised controlled trial. Lancet 361:1149, 2003

62. Pekkanen J, Linn S, Heiss G,et al: Ten-year mortality from cardiovascular disease in relation to cholesterol level among men with and without preexisting cardiovascular disease. N Engl J Med 322:1700, 1990

63. Watts GF, Lewis B, Brunt JN, et al: Effects on coronary artery disease of lipid-lowering diet, or diet plus cholestyramine, in the St Thomas' Atherosclerosis Regression Study (STARS). Lancet 339:563, 1992

64. de Lorgeril M, Salen P, Martin JL, et al: Effect of a Mediterranean type of diet on the rate of cardiovascular complications in patients with coronary artery disease: insights into the cardioprotective effect of certain nutriments. J Am Coll Cardiol 28:1103, 1996

65. de Lorgeril M, Salen P, Martin JL, et al: Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. Circulation 99:779, 1999

66. Zambon A, Hokanson JE, Brown BG, et al: Evidence for a new pathophysiological mechanism for coronary artery disease regression: hepatic lipase-mediated changes in LDL density. Circulation 99:1959, 1999

67. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 344:1383, 1994

68. Pedersen TR, Olsson AG, Faergeman O, et al: Lipoprotein changes and reduction in the incidence of major coronary heart disease events in the Scandinavian Simvastatin Survival Study (4S). Circulation 97:1453, 1998

69. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. N Engl J Med 339:1349, 1998

70. Simes RJ, Marschner IC, Hunt D, et al: Relationship between lipid levels and clinical outcomes in the Long-term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial: to what extent is the reduction in coronary events with pravastatin explained by on-study lipid levels? Circulation 05:1162, 2002

71. Sacks FM, Pfeffer MA, Moye LA, et al: The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. N Engl J Med 335:1001, 1996

72. Sacks FM, Moye LA, Davis BR, et al: Relationship between plasma LDL concentrations during treatment with pravastatin and recurrent coronary events in the Cholesterol and Recurrent Events trial. Circulation 97:1446, 1998

73. MCR/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. Heart Protection Study Collaborative Group. Lancet 360:7, 2002

74. Cannon CP, Braunwald E, McCabe CH, et al: Intensive versus moderate lipid lowering with statins after acute coronary syndromes. N Engl J Med 350:1495, 2004

75. Nissen SE, Tuzcu EM, Schoenhagen P, et al: Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. JAMA 291:1071, 2004

76. Brown BG, Zhao XQ, Chait A, et al: Simvastatin and niacin, antioxidant vitamins, or the combination for the prevention of coronary disease. N Engl J Med 345:1583, 2001

77. Snow V, Aronson MD, Hornbake ER, et al: Lipid control in the management of type 2 diabetes mellitus: a clinical practice guideline from the American College of Physicians. Ann Intern Med 140:644, 2004

78. D'Agostino RB Sr, Grundy S, Sullivan LM, et al: Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 286:180, 2001

79. Liu J, Hong Y, D'Agostino RB Sr, et al: Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. JAMA 291:2591, 2004

80. Brindle P, Emberson J, Lampe F, et al: Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. BMJ 327:1267, 2003

81. Shepherd J, Blauw GJ, Murphy MB, et al: Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet 360:1623, 2002

82. Yusuf S, Hawken S, Ounpuu S, et al: Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 364:937, 2004

83. Davidson MH, McGarry T, Bettis R, et al: Ezetimibe coadministered with simvastatin in patients with primary hypercholesterolemia. J Am Coll Cardiol 40:2125, 2002

84. Cholesterol in childhood. American Academy of Pediatrics, Committee on Nutrition. Pediatrics 101:141, 1998

85. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

86. Carlsson CM, Carnes M, McBride PE, et al: Managing dyslipidemia in older adults. J Am Geriatr Soc 47:1458, 1999

Acknowledgments

Figures 1 and 4 Andy Christie.

Figures 2, 3, 5, and 6 Seward Hung.

V THE PORPHYRIAS

Shigeru Sassa, m.d., ph.d.
Attallah Kappas, m.d.

The porphyrias are uncommon disorders caused by deficiencies in the activities of the enzymes of the heme biosynthetic pathway. The enzymatic defects that cause porphyrias may be either inherited or acquired, and there is significant interplay between the gene defect and acquired or environmental factors in the expression of clinical symptoms.

Acute forms of the porphyrias may be life threatening and may be misdiagnosed because of the nonspecific nature of the clinical presentations (e.g., acute abdominal pain, psychiatric disturbances, and polyneuropathies). The course of the acute forms of disease is characterized by long latent periods interrupted by acute attacks, which are associated with substantial morbidity and mortality.

Classification

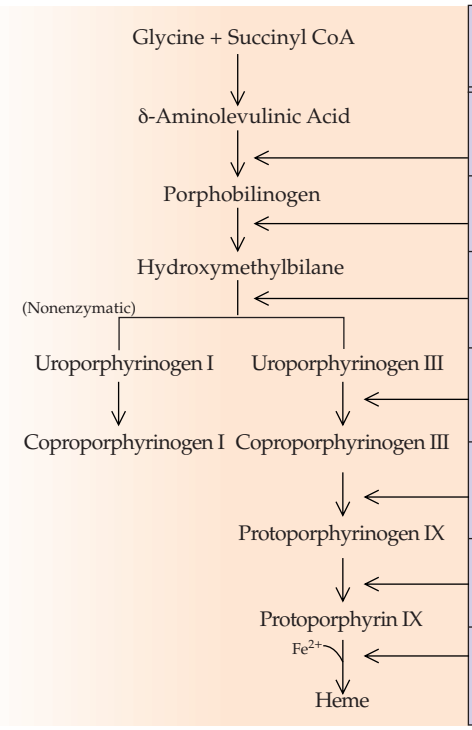
Porphyrias may be classified as neurovisceral or photosensitive, depending on their prominent clinical characteristic, but some porphyrias have both symptoms [*see Figure 1*]. Alternatively, the porphyrias can be classified as hepatic or erythropoietic, depending on the principal site of expression of the specific enzymatic defect involved, but in some porphyrias the expressions overlap.

Pathophysiology

The porphyrias are best understood by the examination of the basic scheme of heme synthesis [*see Figure 2*]. The rate of synthesis is controlled by the mitochondrial enzyme δ-aminolevulinic acid (ALA) synthase. Subsequently, in the cytoplasm, the tetrapyrrole rings remain in the reduced state (porphyrinogens), but the number of carboxyl residues progressively decreases. The last three enzymatic reactions take place in the mitochondrion, resulting in heme, which represses the production of ALA in the liver, the initial and rate-limiting enzyme in this metabolic pathway. The loss of carboxyl groups makes each successive compound less water soluble.

The oxidation of porphyrinogen by the removal of hydrogen atoms results in a series of porphyrins that absorbs light with a wavelength of approximately 400 nm (the Soret band), accounting for the fluorescence characteristic of all porphyrins. Porphyrinogen intermediates oxidize spontaneously, especially in the presence of light. The resulting porphyrins are excreted in the urine, the stool, or both, depending on their relative water solubility.

Each specific abnormality in the pattern of excretion of porphyrins and porphyrin precursors is caused by a reduction in the activity of one of the enzymes from ALA dehydratase (ALAD) to ferrochelatase. The two major organs that are active in heme synthesis are the liver and the erythroid bone marrow, and the consequences of inherited enzymatic defects in the por-



Enzyme	Disease	Major Symptoms	Accumulation Products
δ -Aminolevulinate Dehydratase	ALAD Deficiency Porphyria	Neurovisceral	Urinary ALA
Porphobilinogen Deaminase	Acute Intermittent Porphyria	Neurovisceral	Urinary ALA and PBG
Uroporphyrinogen III Cosynthase	Congenital Erythropoietic Porphyria	Photosensitivity	Erythrocyte and Urinary Uroporphyrin I and Coproporphyrin I
Uroporphyrinogen Decarboxylase	Porphyria Cutanea Tarda; Hepatoerythropoietic Porphyria	Photosensitivity	Urinary 7-Carboxylate Porphyrin; Fecal Isocoproporphyrin
Coproporphyrinogen Oxidase	Hereditary Coproporphyria	Neurovisceral and Photosensitivity	Urinary ALA, PBG, and Coproporphyrin; Fecal Coproporphyrin
Protoporphyrinogen Oxidase	Variegate Porphyria	Neurovisceral and Photosensitivity	Urinary ALA and PBG; Fecal Protoporphyrin
Ferrochelatase	Erythropoietic Protoporphyria	Photosensitivity	Erythrocyte, Plasma, and Fecal Protoporphyrin

Figure 1 **Classification and major symptoms of the porphyrias. δ-Aminolevulinate dehydratase deficiency and porphobilinogen deaminase deficiency are accompanied by acute hepatic porphyria, but not by photocutaneous porphyria, because their enzymatic blocks result in a decrease in porphyrin precursor synthesis. Enzymatic defects beyond uroporphyrinogen III cosynthase are all associated with photocutaneous porphyrias, because they produce excessive amounts of various porphyrins. Both hereditary coproporphyria and variegate porphyria are additionally associated with acute hepatic porphyria.**

phyrias are mainly expressed in these tissues. There is significant tissue-specific regulation for some enzymes in the heme biosynthetic pathway.[1] For example, there are two genes for ALA synthase (ALAS)—namely, the erythroid-specific ALAS gene (*ALAS2*, also called *ALAS-E*) and the housekeeping ALAS gene (*ALAS1*, also called *ALAS-N*). Heme-mediated regulation of ALAS is also tissue specific—namely, *ALAS1* expression in the liver is suppressed by heme, whereas *ALAS2* in the erythroid bone marrow is not. Although deficiency of ALAS does not cause porphyria, *ALAS2* gene defects are associated with X-linked sideroblastic anemia. Genetic deficiency of *ALAS1* has not been described, suggesting that it may not be compatible with life. In addition, there are the erythroid-specific and housekeeping messenger RNAs (mRNAs) for ALAD, porphobilinogen deaminase (PBGD), and uroporphyrinogen cosynthase (UCS), as well as the erythroid-specific and housekeeping enzymes for ALAS and PBGD.

The presence of functional or nonfunctional enzyme proteins can be verified by cross-reaction with specific antibodies to the normal enzymatic protein (called cross-reacting immunologic material [CRIM]). Because individual mutations define differing protein structures, gradations of enzyme activities are encountered and can account for some of the differences in clinical severity.

Clinical Presentation of the Porphyrias

PORPHYRIAS ASSOCIATED WITH NEUROVISCERAL ATTACKS

Porphyrias associated with neurovisceral attacks are also called acute hepatic porphyrias; they include acute intermittent porphyria (AIP), variegate porphyria (VP), hereditary coproporphyria (HCP), and ALAD-deficiency porphyria (ADP). Along with having neurovisceral attacks, patients with HCP and VP may also present with photosensitivity [*see Figure 1*].

Acute Intermittent Porphyria

AIP is an autosomal dominant disorder resulting from a partial deficiency of PBGD.[2] In most patients, the deficient enzyme activity (i.e., approximately 50% of normal) is found in all tissues, including erythrocytes, whereas in 5% of patients, it is found in nonerythroid cells but not in erythrocytes. The cardinal pathobiologic defect of the disease is a neurologic dysfunction that may affect the peripheral, autonomic, and central nervous systems. The majority of persons with this inherited enzyme deficiency remain clinically normal throughout life. Clinical expression of the disease is usually linked to environmental or acquired factors (e.g., nutritional status, drugs, sex steroids, and other chemicals of endogenous or exogenous origin), indicating that there is a significant environmental effect on the clinical expression of the primary gene defect.

Epidemiology AIP is the severest form of the acute hepatic porphyrias and probably the most common of the genetic porphyrias. The highest incidence of AIP occurs in Lapland, Scandinavia, and the United Kingdom, although it has been reported in many population groups. The disorder is expressed clinically almost invariably after puberty and more often in women than in men.

Molecular defects and pathophysiology PBGD catalyzes the condensation of four molecules of PBG to yield hydroxymethylbilane [*see Figure 2*]. In the presence of the next enzyme, uroporphyrinogen III cosynthase, hydroxymethylbilane is converted to uroporphyrinogen III, which involves an intramolecular rearrangement of the D-ring pyrrole.[3] Both the hepatocyte and erythrocyte PBGD are derived from a gene containing 15 exons, but the mRNA is spliced differently in the two tissues, allowing heme synthesis to be regulated differently.[4] In the hepatocyte, transcription starts from exon 1 and encodes an enzyme slightly larger (i.e., 17 additional amino acids at the amino-terminal of the enzyme but otherwise identical) than the enzyme in the erythrocyte. In the erythrocyte, transcription starts at exon 2 rather than at exon 1.

More than 170 different point mutations of the human *PBGD* gene have been described in AIP. Patients with AIP can be classified into three subsets: type I, type II, and type III.

Patients with type I mutations are characterized by PBGD mutations that are negative for CRIM, and they exhibit 50% reduction in enzyme activity and PBGD protein. Mutations found in type I AIP are mostly single-base substitutions or deletions that lead to a single amino acid change or to truncated proteins, which result in the loss of expression of the enzyme protein.

Type II mutations are observed in fewer than 5% of AIP patients and are characterized by decreased PBGD activity in nonerythroid cells (e.g., liver cells) but normal erythroid PBGD activity. The mutations found in type II AIP are single-base substitutions that occur in the exon-intron boundary of exon 1, resulting in a splicing defect that affects the nonspecific form of PBGD but not the erythroid-specific PBGD.[5]

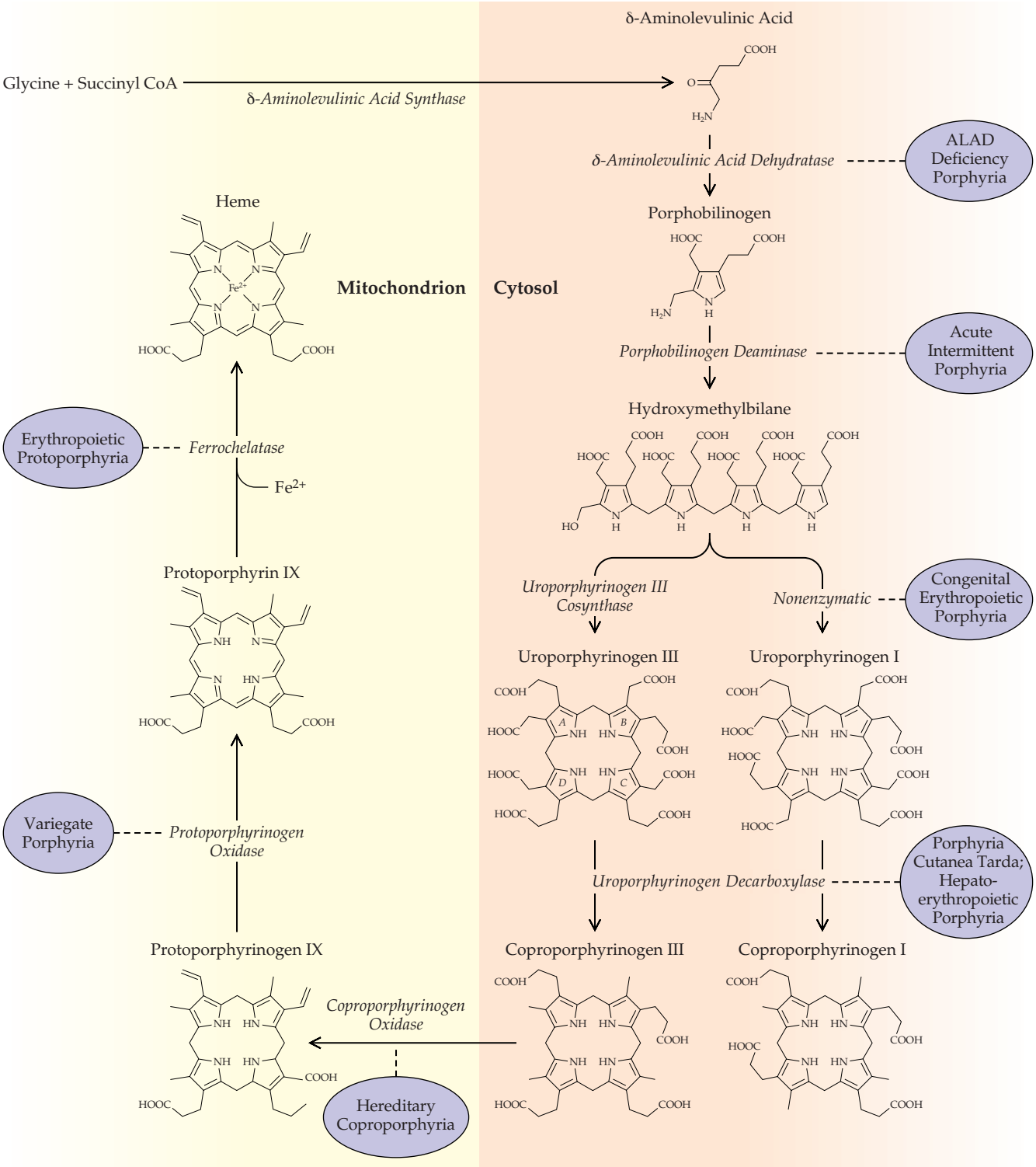
Patients with type III mutations are characterized by CRIM-positive mutations—that is, decreased enzyme activity and the

Figure 2 **The initial rate-limiting step in the synthesis of heme is the synthesis of δ-aminolevulinic acid (ALA) from glycine and succinyl–coenzyme A (succinyl-CoA). This step is catalyzed by the intramitochondrial enzyme ALA synthase. The next step, which takes place in the cytosol, consists of the condensation of two molecules of ALA by the enzyme ALA dehydratase (ALAD) to form the ring pyrrole, porphobilinogen. Porphobilinogen contains two carboxyl residues, acetate (– CH_2COOH) and propionate (– CH_2CH_2COOH) and is water soluble. Porphobilinogen deaminase catalyzes the third step in the biosynthesis of heme, involving the sequential condensation of four molecules of porphobilinogen, yielding the first tetrapyrrole in the biosynthetic pathway, hydroxymethylbilane. In the presence of uroporphyrinogen III cosynthase, hydroxymethylbilane is rapidly converted to uroporphyrinogen III, the vital substrate for subsequent porphyrin and heme formation. In the absence of uroporphyrinogen III cosynthase, hydroxymethylbilane is converted nonenzymatically to the nonphysiologic porphyrin isomer, uroporphyrinogen I. Only uroporphyrinogen III is a precursor of heme. Partial decarboxylation of uroporphyrinogen III yields coproporphyrinogen III, which contains four carboxyl groups. Coproporphyrinogen III reenters the mitochondrion and is oxidized by coproporphyrinogen oxidase to form protoporphyrinogen IX, with a loss of two more propionate residues. The enzyme protoporphyrinogen oxidase catalyzes futher oxidation of the molecule, with a loss of six hydrogen atoms, to form protoporphyrin IX. Finally, the insertion of Fe^{2+} by the mitochondrial enzyme ferrochelatase completes the synthesis of heme. Heme may inhibit ALA-synthase activity directly but also acts to repress synthesis of the enzyme. The porphyrias are caused by inherited defects in specific enzymes in the heme synthetic pathway. A partial blockage at any step in the pathway, when combined with certain endogenous or exogenous factors, results in an overproduction of intermediates or of products normally produced only in small amounts and is associated with overt clinical disease.**

presence of structurally abnormal enzyme protein.[6] Mutations characterizing type III AIP, which mostly occur in exons 10 and 12, are observed in the region that is essential for catalytic activity.

Diagnosis Abdominal pain is the most common symptom of AIP and is often the initial symptom of an acute attack. Other gastroenterologic features may include nausea, vomiting, constipation or diarrhea, abdominal distention, and ileus. Urinary retention, incontinence, and dysuria are frequently observed. Tachycardia, hypertension, and, less commonly, fever, sweating, restlessness, and tremor are also observed. Peripheral neuropathy and muscle weakness are common features of AIP. Muscular weakness can progress to quadriparesis and respiratory paralysis and arrest, which may resemble the Guillain-Barré syndrome.[7]

Acute attacks of AIP may be accompanied by seizures, espe-



cially in patients with hyponatremia caused by vomiting, inappropriate fluid therapy, or the syndrome of inappropriate antidiuretic hormone. No cutaneous manifestations are associated with this enzyme deficiency. Patients with clinically expressed AIP excrete increased amounts of ALA and PBG in the urine during attacks and sometimes between attacks. In severe cases, the urine develops a port-wine color from a high content of porphobilin, an auto-oxidation product of PBG.

In persons with latent disease, an acute attack may be precipitated by endogenous or exogenous environmental factors. Precipitating factors of AIP fall into five categories: (1) drugs that induce ALAS1 in the liver; (2) endocrine factors that facilitate ALAS1 induction; (3) reduced caloric intake that derepresses the synthesis of ALAS1; (4) drugs that induce hepatic cytochrome P-450, thereby driving heme synthesis; and (5) various oxidative stresses (e.g., infections and surgery) that induce heme oxygenase–1, which results in excessive heme catabolism and leads to the derepression of ALAS1. In women, relapses occur, particularly premenstrually. One study showed oral contraceptives precipitated acute attacks in 24% of the persons studied, whereas menopausal hormone replacement therapy only rarely affected the disorder.[8] Nevertheless, as a precaution, it is practical to restrict use of exogenous sex hormones in all women who have AIP.

The diagnosis of clinically expressed AIP requires demonstration of increased urinary excretion of PBG and ALA. A urine sample can be rapidly screened for PBG by the Watson-Schwartz or the Hoesch test, both of which utilize the Ehrlich reagent to detect a chromogen. Twenty-four-hour urine samples should be collected, placed in opaque containers, refrigerated, and delivered to a qualified laboratory for quantitative analysis of ALA, PBG, and porphyrins. Because the other hepatic porphyrias (i.e., VP, HCP, and ALAD deficiency) may produce identical neuropathic syndromes and may be marked by excess excretion of PBG, ALA, or both, definitive diagnosis should be sought by the erythrocyte PBGD assay. The diagnosis of AIP types I and III can be made by demonstrating decreased PBGD activity using an erythrocyte PBGD assay.

Treatment The same treatment is applied to all acute hepatic porphyrias, including AIP. During latent periods, when clinical manifestations are absent, management of AIP comprises adequate nutritional intake, avoidance of drugs and chemicals known to exacerbate porphyria, and prompt treatment of concurrent diseases or infections. Drug treatment should be prescribed only after reviewing a drug list. Some drugs are generally agreed to be safe in acute porphyria; some are considered unsafe [*see Table 1*]. Reactions to a drug, however, may be variable depending on the individual patient.

Severe cases of AIP should be treated with I.V. administration of carbohydrate (dextrose) to provide a minimum of 300 g of carbohydrate a day.[9] Intravenous hematin (4 mg/kg every 12 hours) is probably most effective in reducing ALA and PBG excretion, as well as in curtailing acute attacks.[10]

Variegate Porphyria

VP, also known as South African genetic porphyria, is caused by a heterozygous deficiency in protoporphyrinogen oxidase (PPO) activity and is inherited as an autosomal dominant trait.[2] The disease is called variegate because patients with this disorder can present with neurovisceral symptoms, photosensitivity, or both.

Epidemiology VP is rarer than AIP but has a substantially higher incidence in South Africa (three per 1,000 population) than elsewhere.[11] In 1980, it was estimated that there were 10,000 affected individuals in South Africa, and good evidence suggests that they were all descendants of a single union between two Dutch settlers in 1688.[12,13] With the exception of the incidence of disease in South Africa, VP probably has no racial or geographic predilection. Very rare forms of VP are seen with homozygous deficiency in PPO activity.[14]

Molecular defects and pathophysiology VP is attributed to a 50% reduction of PPO activity. PPO catalyzes the oxidation of protoporphyrinogen to yield protoporphyrin [*see Figure 2*]. In homozygous VP patients, which are rare, PPO mutations are found that are usually not found in heterozygous VP patients, suggesting that the PPO mutations in homozygotes are associated with residual activity.[15]

Diagnosis Patients with VP can have neurovisceral symptoms, photosensitivity, or both. The neurovisceral findings in VP are identical to those observed in the other acute hepatic porphyrias (e.g, abdominal pain, gastroenterologic symptoms, peripheral neuropathy, and muscle weakness). Photosensitivity and the resulting lesions tend to be more chronic in VP than in hereditary coproporphyria. Typically, lesions appear in sun-exposed areas. Blisters, as well as superficial ulcers, are present in various stages of healing and scarring. Mechanical fragility is common, especially at sites of bony protuberances such as knuckles and ankles. Skin lesions and neuropathic lesions may occur separately or together. Because the skin lesions are identical to those seen in hereditary coproporphyria and porphyria cutanea tarda, VP must be differentiated from these two conditions. The presence of neuropathic lesions without cutaneous manifestations in VP necessitates differentiation from AIP. When both cutaneous and neuropathic lesions occur in a person or kindred, hereditary coproporphyria should be considered in the differential diagnosis.

PPO activity in most patients with VP is decreased by 50%. In very rare cases of homozygous VP, however, a virtual absence of PPO activity has been documented.[16] Symptoms in homozygous VP patients are severe photosensitivity, growth and mental retardation, and marked neurologic abnormalities. Onset of the homozygous disease is in childhood. Protoporphyrin is excreted by the liver and partially reabsorbed in the gut. It must be sought in plasma and feces, where increased levels of protoporphyrin are found in 95% of VP cases. Thus, the biochemical hallmark of VP is an elevated concentration of plasma and fecal protoporphyrin IX. Plasma invariably shows a fluorescence emission when excited by longwave ultraviolet light (maximum, ~ 400 nm), which is caused by the presence of a protoporphyrin-peptide conjugate.

Treatment The treatment of neurovisceral symptoms in VP is the same as that in AIP. Cutaneous lesions are limited to sun-exposed areas, but management of VP skin lesions remains inadequate. Protection from sunlight, therefore, is an important management consideration and may include wide-brimmed hats, protective clothing, and sunscreen.[17]

Hereditary Coproporphyria

HCP is a disease caused by a heterozygous deficiency of coproporphyrinogen oxidase (CPO) activity that is inherited as an

Table 1 Safe and Unsafe Drugs for Patients with Acute Intermittent Porphyria, Variegate Porphyria, Hereditary Coproporphyria, or ALA Dehydratase Deficiency Porphyria

Safe Drugs	**Unsafe Drugs**
Acetaminophen	Barbiturates†
Acetazolamide	Captopril
Acyclovir	Chloramphenicol†
Allopurinol	Chlordiazepoxide†
Amiloride	Chlorpropamide†
Ampicillin	Diazepam‡
Aspirin	Diltiazem
Atropine	Diphenhydramine
Bumetanide	Doxycycline
Bupivacaine	Ergot compounds†
Buprenorphine	Erythromycin
Chlorothiazide*	Estrogen
Codeine phosphate	Ethanol†
Corticosteroids*	Furosemide‡
Deferoxamine	Griseofulvin†
Demerol	Hydralazine
Digoxin	Hydrochlorothiazide‡
Fentanyl	Imipramine†
Follicle-stimulating hormone (FSH)	Lidocaine
Gabapentin	Methyldopa†
Gentamicin	Metoclopramide‡
Glipizide	Metronidazole
Haloperidol	Nifedipine
Heme arginate	Oral contraceptives‡
Heparin	Orphenadrine†
Insulin	Oxycodone
Iron	Pentazocine†
Lithium salts	Phenobarbital†
Meperidine	Phenytoin†
Mequitazine	Piroxicam
Metformin	Pivampicillin†
Metoprolol	Progesterone†
Morphine	Pyrazinamide†
Nadolol	Sodium valproate‡
Oxytocin	Terfenadine
Penicillin	Tetracyclines‡
Procaine	Theophylline†
Propofol	Trimethoprim
Propylthiouracil	Verapamil
Quinine	
Ranitidine*	
Salbutamol	
Senna	
Temazepam	
Thyroxine	
Warfarin	

Note: See also the American Porphyria Foundation Web site, at http://www.porphyriafoundation.com. The American Porphyria Foundation charges a nominal fee to access their site.

*Has produced conflicting results (occasionally positive but mainly negative) in experiments on porphyrinogenicity. None of the safe drugs listed has been associated with human porphyric attacks.

†Has been associated with acute attacks of porphyria.

‡Has produced conflicting results (some positive, some negative) in experiments on porphyrinogenicity.

autosomal dominant trait.[2] Clinically, the disease is similar to AIP, although it is often milder; additionally, HCP may be associated with cutaneous photosensitivity. Very rarely, homozygous deficiency of CPO occurs and is associated with a more severe form of HCP.[18]

Molecular defects and pathophysiology CPO is a mitochondrial enzyme that catalyzes the removal of the carboxyl group and two hydrogens from the propionic groups of pyrrole rings A and B of coproporphyrinogen to form vinyl groups at these positions [*see Figure 2*]. Molecular analysis of several families with HCP revealed a variety of mutations in the *CPO* gene. These include missense, nonsense, and splicing mutations, as well as insertions and deletions. CPO activity in HCP is typically reduced by about 50% in heterozygotes and by 90% to 98% in homozygotes, who are rare.[18]

Diagnosis Neurovisceral symptoms predominate in HCP and are essentially indistinguishable from those seen in other acute hepatic porphyrias. Abdominal pain, vomiting, constipation, neuropathies, and psychiatric manifestations are common. Cutaneous photosensitivity is a feature in about 30% of cases. Attacks can be precipitated by the same factors as those that are known to aggravate AIP.

The biochemical hallmark of HCP is hyperexcretion of coproporphyrin (predominantly type III) in urine and feces. Hyperexcretion of ALA, PBG, and coproporphyrin into the urine may accompany exacerbation of the disease; however, in contrast to such findings in AIP, these findings in HCP generally return to normal between attacks.

HCP should be suspected in patients with the signs, symptoms, and clinical course characteristic of the acute hepatic porphyrias but in whom PBGD activity is normal. Urinary excretion of heme precursors in HCP is similar to that in VP, but the predominant or exclusive presence of fecal coproporphyrin is highly suggestive of HCP. Fecal or urinary predominance of harderoporphyrin, with greatly reduced CPO activity, was reported in a case of harderoporphyria,[19] a variant form of HCP.

Treatment Treatment of and prophylaxis for HCP are the same as those for VP and AIP. The use of drugs in patients with HCP should be the same as that in patients with AIP [*see Table 1*].

ALAD-Deficiency Porphyria

ADP is the rarest of the porphyrias; only four cases have been reported as confirmed by molecular diagnosis. The symptomatology is similar to that seen in AIP, but ADP can be differentiated from AIP by the lack of PBG overproduction. Urinary ALA excretion is greatly increased, whereas urinary PBG excretion is within the normal range. Patients with ADP display markedly

decreased activity of ALAD in erythrocytes, as well as in nonerythroid cells (less than 2% of normal), and their parents typically show 50% decreases in enzyme activity.

ADP results from a marked reduction in ALAD activity caused by a homozygous enzyme deficiency [*see Figure 2*].[2] ALAD catalyzes the reaction that converts two molecules of ALA to form a monopyrrole, PBG, by the removal of two molecules of water. The enzyme is a homo-octamer, with a subunit size of 36,274 daltons,[20] and requires an intact sulfhydryl group and one zinc atom per subunit for full activity.[21] The human *ALAD* genomic structure is 16 kb in length, with two promoter regions and two alternative noncoding exons, 1A and 1B, that generate housekeeping and erythroid-keeping transcripts, respectively.[22] The two transcripts encode the same amino acid sequence, because translation begins in exon 2. Lead displaces zinc from the enzyme, resulting in an inactive enzyme and neurologic disturbances, some of which resemble those of ADP.[23] The most potent inhibitor of the enzyme is succinylacetone, a structural analogue of ALA, which is found in urine and blood of patients with hereditary tyrosinemia, who frequently develop symptoms similar to those of ADP.[24,25]

All ALAD point mutations have been studied by bacterial expression. The studies have revealed that eight of nine different mutations are unique and have markedly decreased enzyme activity, indicating the highly heterogeneous nature of the enzyme phenotypes.[26]

The clinical management of ADP is essentially identical to that of AIP.

PORPHYRIAS ASSOCIATED WITH CUTANEOUS PHOTOSENSITIVITY

The porphyrias associated with cutaneous photosensitivity are porphyria cutanea tarda (PCT), hepatoerythropoietic porphyria (HEP), erythropoietic protoporphyria (EPP), and congenital erythropoietic porphyria (CEP) [*see Figure 1*]. CEP, HEP, and EPP are also called erythropoietic porphyrias.

Porphyria Cutanea Tarda

PCT refers to a heterogeneous group of cutaneous porphyric diseases caused by uroporphyrinogen decarboxylase (UROD) deficiency, which may be either inherited or, more commonly, acquired.[2,27]

Epidemiology PCT is probably the most common of all forms of the porphyrias, but its exact incidence is not clear. The disease is recognized worldwide, and there is no racial predilection except for the high incidence of hemochromatosis found in the Bantus in South Africa. PCT was once considered more common in men, possibly because of greater alcohol intake; however, the incidence in women has recently matched that of men, which may be explained by women's use of contraceptive steroids and postmenopausal estrogens and increased alcohol intake .

Molecular defects and pathophysiology UROD is a cytosolic enzyme that catalyzes the removal of the four carboxyl groups of the carboxymethyl side chains from uroporphyrinogen to yield coproporphyrinogen [*see Figure 2*]. In contrast to the unique erythroid-expression mechanism for the first four enzymes of the heme biosynthetic pathway, the *UROD* gene has only a single promoter, and the gene is transcribed as a single mRNA.[28] Both PCT and the much rarer HEP [*see* Hepatoerythropoietic Porphyria, *below*] are characterized by a partial and a nearly complete UROD deficiency, respectively. Inherited PCT is caused by heterozygous UROD deficiency, whereas HEP is caused by homozygous UROD deficiency. Both inherited and noninherited forms of PCT display reductions in hepatic UROD activity, but erythrocyte UROD activity may or may not be decreased, depending on the clinical subtype. There are three types of PCT. Type I PCT, which accounts for 80% to 90% of all cases, is an acquired disease that typically presents in adults as decreased hepatic UROD activity but not decreased erythroid UROD activity. The disease may occur spontaneously, but it more commonly occurs in conjunction with precipitating environmental factors, such as use of alcohol, estrogen, or drugs. Type II PCT is inherited as an autosomal dominant trait and is associated with decreased UROD activity in all tissues. Type III PCT is also inherited, but the defect is confined to the UROD activity in the liver and in erythrocytes, and its protein concentrations are normal.

A variety of UROD mutations causing type II familial PCT have been identified, including missense, nonsense, and splice-site mutations; several small and large deletions; and a small insertion. In contrast to genetic defects in type II PCT, UROD mutations are not found in type I PCT.[29]

The pathogenetic processes of PCT are not fully understood. Porphyrin-mediated activation of the complement system after irradiation is implicated as one of the possible mechanisms and has been demonstrated in patients with PCT; activation of the complement system is presumed to result from the generation of reactive oxygen species, most likely singlet oxygen.[30]

Diagnosis The pathognomonic clinical feature of PCT is the formation of vesicles on sun-exposed areas of the skin, particularly the dorsal aspects of the hands, as well as on the face, forearms, and legs. The vesicles are superseded by crusting, superficial scarring, or milia formation and by residual pigmentation. Facial hypertrichosis may be present and is conspicuous in women. Iron, estrogens, alcohol, viral hepatitis, and chlorinated hydrocarbons can aggravate PCT.[31] Mild iron overload is nearly always present in patients with PCT. Iron plays a particularly important role in the symptomatology of PCT, in that phlebotomy to decrease hepatic iron overload is effective in treating PCT [*see* Treatment, *below*], whereas iron supplementation results in relapse of PCT. A significant number of PCT patients have associated hemochromatosis gene (*HFE*) mutations.[32] Isocoproporphyrin, unique to PCT and HEP, may be detected in serum or in stool and is diagnostic of PCT. Plasma porphyrin levels are increased in PCT, HEP, and other photosensitizing porphyrias. In PCT, serum ferritin levels are also typically elevated, which is not the case in other cutaneous porphyrias.

Although many PCT patients have moderately excessive alcohol intake or hepatitis C infection, few have advanced liver disease at the time of initial presentation. However, liver abnormalities are seen even in patients without heavy alcohol intake or hepatitis C, indicating PCT itself is associated with liver damage. PCT appears to increase the risk for hepatocellular carcinoma in patients with chronic liver disease.[33]

Treatment In type I PCT, the identification and avoidance of precipitating factors represent the first line of treatment. Abstinence from alcohol ingestion should be recommended to the patient.

The cornerstone of therapy for all types of PCT is depletion of

iron, even in patients lacking biochemical evidence of iron overload. Repeated phlebotomy of one unit of blood twice monthly for a total of 5 to 10 L, with treatment guided by the patient's hematocrit and ferritin levels, decreases both uroporphyrin excretion and photosensitivity. Improvement occurs within several months to a year. Patients who are on hemodialysis, who are anemic, or who cannot tolerate phlebotomy should receive erythropoietin. Subcutaneous infusion of desferrioxamine by portable syringe pump (1.0 to 1.5 g in 8 to 10 ml of sterile water for 8 to 10 hours 5 nights a week for 2 to 5 months) is also effective.[34]

If phlebotomy is ineffective or contraindicated, low-dose chloroquine therapy may also be considered. Chloroquine, which forms complexes with uroporphyrin, has produced improvement over 3 to 6 months; however, hepatotoxicity may occur. Resolution of type 1 PCT has followed successful interferon therapy for HCV infection.[35]

Hepatoerythropoietic Porphyria

HEP is a rare form of porphyria resulting from a homozygous deficiency of UROD.[36] Individuals in HEP families who have heterozygous UROD deficiency usually do not have clinical symptoms. HEP is characterized clinically by the childhood onset of severe photosensitivity and skin fragility and is indistinguishable from CEP. Some 20 cases have been reported worldwide.

Molecular defects and pathophysiology As in PCT, HEP is caused by a UROD deficiency [*see Figure 2*]. A variety of UROD mutations have been identified in HEP patients, indicating the molecular heterogeneity of the disease. Most UROD mutations in HEP have not been found in familial PCT and are associated with residual UROD activity.[29]

Diagnosis Clinical findings of HEP are very similar to those of CEP. Pink urine, severe photosensitivity leading to scarring and mutilation of sun-exposed areas of skin, sclerodermoid changes, hypertrichosis, erythrodontia, anemia (often hemolytic), and hepatosplenomegaly are characteristic features of HEP. In contrast to PCT, serum iron concentrations are usually normal, and phlebotomy has no beneficial effects in patients with HEP. Isocoproporphyrin concentrations are equal to or greater than concentrations of coproporphyrin found in urine and feces, and although the reasons are unclear, an elevated erythrocyte zinc protoporphyrin concentration is commonly observed.

Urinary fluorescence under ultraviolet light and quantitation and identification of isocoproporphyrin by thin-layer or high-performance liquid chromatography establish the diagnosis of HEP.

As in other photosensitizing porphyrias, plasma porphyrin levels are elevated in HEP. Fecal porphyrin levels are often elevated. The detection of isocoproporphyrin in feces is diagnostic of HEP and PCT. The diagnosis of HEP should be suspected in patients with severe photosensitivity and should especially be included in the differential diagnosis of CEP.

Treatment The identification and avoidance of precipitating factors represent the first line of treatment for PCT. In contrast to treatment for PCT, avoidance of the sun and the use of topical sunscreens are essentially all that can be recommended to patients with HEP; phlebotomy, which is most useful for treatment of PCT, provides no beneficial response in patients with this disorder.

Erythropoietic Protoporphyria

EPP is due to a partial deficiency of ferrochelatase and is inherited as an autosomal dominant trait.[2] Biochemically, this defect results in massive accumulations of protoporphyrin in erythrocytes, plasma, and feces. Clinically, the disease is characterized by the childhood onset of cutaneous photosensitivity in light-exposed areas, but skin lesions are milder and less disfiguring than those in CEP. EPP is the most common form of the erythropoietic porphyrias. There is no racial or sexual predilection, and onset is typically in childhood.

Molecular defects and pathophysiology Ferrochelatase catalyzes the final reaction in heme biosynthesis—that is, the insertion of iron into protoporphyrin IX. Unlike other steps in the heme biosynthetic pathway, this mitochondrial enzyme utilizes protoporphyrin IX, rather than its reduced form (i.e., protoporphyrinogen IX), as substrate. However, the enzyme specifically requires the reduced form of iron (i.e., ferrous, not ferric, iron) [*see Figure 2*]. The gene for human ferrochelatase has been assigned to chromosome 18q21.3.

Molecular analysis of ferrochelatase mutations causing EPP has revealed a variety of alterations, including missense, nonsense, and splice-site mutations, as well as insertions and deletions. Of these alterations, splice-site mutations are the most frequent. EPP patients have only 10% to 25% of normal ferrochelatase activity, whereas their asymptomatic family members typically have 50% ferrochelatase activity. A normal coding ferrochelatase sequence allele, expressed at a lower than normal level,[37] is present in about 10% of the white population. Inheritance of a ferrochelatase mutation in *cis* and the low expression allele in *trans* appears to account for the markedly low ferrochelatase activity and clinical expression of the disease.[37]

Light-excited porphyrins generate free radicals and singlet oxygen.[38] Such radicals, notably singlet oxygen, can lead to peroxidation of lipids and cross-linking of membrane proteins, which, in erythrocytes, can result in reduced deformability and thus hemolysis. Protoporphyrin IX, but not zinc protoporphyrin IX, can be released from erythrocytes after irradiation.[39] This finding explains why patients with EPP exhibit elevated levels of free protoporphyrin in plasma and manifest photosensitivity, whereas patients with lead intoxication and iron deficiency, who have elevated zinc protoporphyrin levels in erythrocytes, do not exhibit photosensitivity.[40]

Diagnosis Cutaneous photosensitivity of EPP is quite different from that of CEP or PCT. Stinging or painful burning sensations of the skin occur within 1 hour after exposure to the sun and are followed several hours later by erythema and edema. Petechiae or, more rarely, purpura, vesicles, and crusting may develop and may persist for several days after sun exposure. Symptoms are usually worse during spring and summer and occur in light-exposed skin. Excoriations secondary to scratching may be present. Recurrence of the lesions as a result of chronic sun exposure may result in onycholysis, scarring, altered pigmentation, lichenification, and premature aging of the skin. Gallstones, sometimes presenting at an early age, are fairly common in patients with EPP, and hepatic disease, although unusual, may be severe and associated with significant morbidity.

The biochemical hallmark of EPP is excessive concentrations of protoporphyrin in erythrocytes, plasma, bile, and feces—but not in urine, because of the poor solubility of protoporphyrin in water.

Photosensitivity should suggest the diagnosis of EPP, which can be confirmed by the demonstration of increased concentrations of free protoporphyrin in erythrocytes, plasma, and stool and normal levels of urinary porphyrins. The presence of protoporphyrin in both plasma and erythrocytes is a finding specific for EPP. Fluorescent reticulocytes on examination of peripheral blood smear also suggest the diagnosis.

Treatment Avoidance of the sun and use of topical sunscreen agents are helpful in the management of EPP. Oral administration of β-carotene can afford systemic photoprotection, resulting in improved, although highly variable, tolerance to the sun. The recommended serum β-carotene level of 600 to 800 μg/dl is usually achieved with oral dosages of 120 to 180 mg daily, and beneficial effects are typically seen 1 to 3 months after the therapy is begun. The mechanism of this beneficial effect of β-carotene probably involves quenching of activated oxygen radicals.[41]

Congenital Erythropoietic Porphyria

CEP, which is also referred to as Günther disease, is an autosomal recessive disorder caused by a homozygous deficiency of the cytosolic enzyme, uroporphyrinogen cosynthase (UCS). The enzymatic defect results in accumulation and hyperexcretion of predominantly type I porphyrins.[2] Fewer than 200 cases have been reported, and some of these cases may have been PCT or HEP. No clear racial or sexual predilection is apparent.[2]

Molecular defects and pathophysiology UCS catalyzes the formation of uroporphyrinogen III (UROIII) from hydroxymethylbilane (HMB). In the absence of UCS, HMB is converted nonenzymatically to uroporphyrin I, which is then enzymatically converted to coproporphyrin I [*see Figure 2*]. Excess porphyrin causes the staining of bones and teeth (erythrodontia), hemolysis, dark urine, and photosensitivity, all of which are usually identified early in infancy.

Similar to the *ALAD* and *PBGD* genes, the *UCS* gene has alternative promoters that generate housekeeping and erythroid transcripts.[39] A heterogeneity of mutations in the *UCS* gene is found in patients with CEP. A $Cys^{73} \rightarrow$ Arg point mutation appears to occur more frequently than others, because it has been found in eight of 21 unrelated patients with CEP (about 21% of CEP alleles).[42]

Diagnosis The first clue suggesting the diagnosis of CEP at birth is pink to dark-brown staining of diapers, which is caused by large amounts of porphyrins in the urine. Early onset of cutaneous photosensitivity is characteristic and is exacerbated by exposure to sunlight. Subepidermal bullous lesions progress to crusted erosions that heal with scarring and either hyperpigmentation or, less commonly, hypopigmentation. Hypertrichosis and alopecia are frequent and erythrodontia (appearing as red fluorescence under ultraviolet light) is pathognomonic of CEP. Patients may display symptoms and signs of hemolytic anemia with splenomegaly and porphyrin-rich gallstones. Bone marrow shows erythroid hyperplasia, which may result in pathologic fractures or vertebral compression-collapse and shortness of stature. Urinary levels of uroporphyrins and coproporphyrins are always elevated (20- to 60-fold), with predominant elevations of type I isomers.

Pink urine or the onset of severe cutaneous photosensitivity in infancy (and rarely in adults) suggests the diagnosis of CEP. Demonstration of increased levels of urinary, fecal, and erythrocyte porphyrins, together with elevated type I isomers of uroporphyrin and coproporphyrin, establish the diagnosis of CEP.

Treatment The avoidance of sunlight, trauma to the skin, and infections is the most important preventive measure. Topical sunscreens may be of some help, as may oral administration of β-carotene. Transfusions with packed erythrocytes transiently decrease hemolysis and its attendant drive to increased erythropoiesis and also decrease porphyrin excretion by suppressing erythropoiesis in the bone marrow. Bone marrow transplantation is curative.[43]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Sassa S: The hematological aspects of porphyrias. Williams Hematology. Beutler E, Lichtman MA, Coller BS, et al, Eds. McGraw-Hill, New York, 2001, p703

2. Kappas A, Sassa S, Galbraith RA, et al: The porphyrias. The Metabolic and Molecular Basis of Inherited Disease. Scriver CR, Beaudet AL, Sly WS, et al, Eds. McGraw-Hill, New York, 1995, p 2103

3. Grandchamp B, Beaumont C, de Verneuil H, et al: Genetic expression of porphobilinogen deaminase and uroporphyrinogen decarboxylase during the erythroid differentiation of mouse erythroleukemic cells. Porphyrins and Porphyrias. Nordmann Y, Ed. John Libbey & Company Ltd, London, 1986, p 35

4. Romeo PH, Chretien S, Dubart A, et al: Erythroid specific promoter of a housekeeping gene. Prog Clin Biol Res 251:55, 1987

5. Grandchamp B, Picat C, Mignotte V, et al: Tissue-specific splicing mutation in acute intermittent porphyria. Proc Natl Acad Sci USA 86:661, 1989

6. Grandchamp B, Picat C, de Rooij F, et al: A point mutation G → A in exon 12 of the porphobilinogen deaminase gene results in exon skipping and is responsible for acute intermittent porphyria. Nucl Acids Res 17:6637, 1989

7. Elder GH, Hift RT, Meissner PN: The acute porphyrias. Lancet 349:1613, 1997

8. Andersson C, Innala E, Backstrom T: Acute intermittent porphyria in women: clinical expression, use and experience of exogenous sex hormones: a population-based study in northern Sweden. J Intern Med 254:176, 2003

9. Sassa S: Diagnosis and therapy of acute intermittent porphyria. Blood Reviews 10:53, 1996

10. Mustajoki P, Tenhunen R, Pierach C, et al: Heme in the treatment of porphyrias and hematological disorders. Semin Hematol 26:1, 1989

11. Eales L, Day RS, Blekkenhorst G.H: The clinical and biochemical features of variegate porphyria: an analysis of 300 cases studied at Groote Schuur Hospital, Cape Town. Int J Biochem 12:837, 1980

12. Dean G.: The porphyrias: a study of inheritance and environment. Pitman Medical, London, 1971

13. Meissner PN, Dailey TA, Hift RJ, et al: A R59W mutation in human protoporphyrinogen oxidase results in decreased enzyme activity and is prevalent in South Africans with variegate porphyria. Nature 13:95, 1996

14. Roberts AG, Whatley SD, Daniels J, et al: Partial characterization and assignment of the gene for protoporphyrinogen oxidase and variegate porphyrias to human chromosome 1q23. Human Mol Genet 4:2387, 1995

15. Frank J, McGrath J, Lam H, et al: Homozygous variegate porphyria: identification of mutations on both alleles of the protoporphyrinogen oxidase gene in a severely affected proband. J Invest Dermatol 110:452, 1998

16. Kordac V, Martasek P, Zeman J, et al: Increased erythrocyte protoporphyrin in homozygous variegate porphyria. Photo-Dermatology 2:257, 1985

17. Muhlbauer JE, Pathak MA, Tishler PV, et al: Variegate porphyria in New England. JAMA 247:3095, 1982

18. Grandchamp B, Phung N, Nordmann Y: Homozygous case of hereditary coproporphyria. Lancet 2:1348, 1977

19. Nordmann Y, Grandchamp B, de Verneuil H, et al: Harderoporphyria: a variant hereditary coproporphyria. J Clin Invest 72:1139, 1983

20. Wetmur JG, Bishop DF, Ostasiewicz L, et al: Molecular cloning of a cDNA for human δ-aminolevulinate dehydratase. Gene 43:123, 1986

21. Sassa S: δ-Aminolevulinic acid dehydratase assay. Enzyme 28:133, 1982

22. Kaya AH, Plewinska M, Wong DM, et al: Human δ-aminolevulinate dehydratase (ALAD) gene: structure and alternative splicing of the erythroid and housekeeping mRNAs. Genomics 19:242, 1994

23. Granick JL, Sassa S, Kappas A: Some biochemical and clinical aspects of lead intoxication. Advances in Clinical Chemistry. Bodansky O, Latner AL, Eds. Academic Press, New York, 1978, p 287

24. Lindblad B, Lindstedt S, Steen G: On the genetic defects in hereditary tyrosinemia. Proc Natl Acad Sci USA 74:4641, 1977

25. Sassa S, Kappas A: Hereditary tyrosinemia and the heme biosynthetic pathway. Profound inhibition of δ-aminolevulinic acid dehydratase activity by succinylacetone. J Clin Invest 71:625, 1983

26. Maruno M, Furuyama K, Akagi R, et al: Highly heterogenous nature of δ-aminolevulinate dehydratase (ALAD) deficiencies in ALAD porphyria. Blood 97:2972, 2001

27. de Verneuil H, Nordmann Y, Phung N, et al: Familial and sporadic porphyria cutanea: two different diseases. Int J Biochem 9:927, 1978

28. Romeo PH, Raich N, Dubart A, et al: Molecular cloning and tissue-specific expression analysis of human porphobilinogen deaminase and uroporphyrinogen decarboxylase. Porphyrins and Porphyrias. Nordmann Y, Ed. John Libbey & Company Ltd, London, 1986, p 25

29. Elder GH: Porphyria cutanea tarda. Semin Liver Dis 18:67, 1998

30. Lim HW, Poh-Fitzpatrick MB, Gigli I: Activation of the complement system in patients with porphyrias after irradiation in vivo. J Clin Invest 74:1961, 1984

31. Blauvelt A, Harris HR, Hogan DJ, et al: Porphyria cutanea tarda and human immunodeficiency virus infection. Int J Dermatol 31:474, 1992

32. Bulaj ZJ, Phillips JD, Ajioka RS, et al: Hemochromatosis genes and other factors contributing to the pathogenesis of porphyria cutanea tarda. Blood 95:1565, 2000

33. Fracanzani AL, Taioli E, Sampietro M, et al: Liver cancer risk is increased in patients with porphyria cutanea tarda in comparison to matched control patients with chronic liver disease. J Hepatol 35:498, 2001

34. Gilbertini P, Rocchi E, Cassanelli A, et al: Advances in the treatment of porphyria cutanea tarda: effectiveness of slow subcutaneous desferrioxamine infusion. Liver 4:280, 1984

35. Sheikh MY, Wright RA, Burruss JB: Dramatic resolution of skin lesions associated with porphyria cutanea tarda after interferon-alpha therapy in a case of chronic hepatitis C. Dig Dis Sci 43:529, 1998

36. Elder GH, Smith SG, Herrero C, et al: Hepatoerythropoietic porphyria: a new uroporphyrinogen decarboxylase defect or homozygous porphyria cutanea tarda? Lancet 1:916, 1981

37. Gouya L, Puy H, Lamoril J, et al: Inheritance in erythropoietic protoporphyria: a common wild-type ferrochelatase allelic variant with low expression accounts for clinical manifestation. Blood 93:2105, 1999

38. Spikes JD: Porphyrins and related compounds as photodynamic sensitizers. Ann NY Acad Sci 244:496, 1975

39. Aizencang GI, Bishop DF, Forrest D, et al: Uroporphyrinogen III synthase. an alternative promoter controls erythroid-specific expression in the murine gene. J Biol Chem 275:2295, 2000

40. Sandberg S, Talstad I, Hovding G, et al: Light-induced release of protoporphyrin, but not of zinc protoporphyrin, from erythrocytes in a patient with greatly elevated erythropoietic protoporphyria. Blood 62:846, 1983

41. Mathews-Roth MM: Systemic photoprotection. Dermatol Clin 4:335, 1986

42. Solis C, Aizencang GI, Astrin KH, et al: Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. J Clin Invest 107:753, 2001

43. Harada FA, Shwayder TA, Desnick RJ, et al: Treatment of severe congenital erythropoietic porphyria by bone marrow transplantation. J Am Acad Dermatol 45:279, 2001

VI DIABETES MELLITUS

Saul Genuth, m.d.

Definition and Overview

Diabetes mellitus is a metabolic disease characterized by hyperglycemia that results from defects in insulin secretion, insulin action, or both. Important abnormalities in fat and protein metabolism are also present. Nonetheless, the diagnosis still rests upon demonstrating elevated plasma glucose levels. The chronic hyperglycemia of diabetes mellitus is specifically associated with long-term damage, dysfunction, and failure of various organs, especially the retina and lens of the eye, the kidneys, and both somatic and autonomic nervous systems. The heart, arterial system, and microcirculation are also adversely affected.

A variety of pathogenic processes are involved in the development of different forms of diabetes. These processes range from autoimmune destruction of the beta cells of the pancreatic islets with consequent insulin deficiency to mutations in the insulin receptor gene with consequent resistance to insulin action. The basis for the metabolic abnormalities of diabetes mellitus is deficient action of insulin on its major target tissues, including skeletal muscle, cardiac muscle, adipose tissue, and liver. Loss of proper insulin regulation of metabolism results from inadequate secretion of insulin, from diminished tissue responses to insulin at one or more points in the complex pathways of insulin action, or from both processes. Impairment of insulin secretion and defects in insulin action coexist in many patients, and in these patients, it is often unclear which abnormality is the primary cause of the hyperglycemia.

Acute life-threatening consequences of diabetes mellitus are ketoacidosis and nonketotic hyperglycemic hyperosmolar coma. Overtreatment of hyperglycemia can lead to hypoglycemia, which may be severe enough to cause seizures and loss of consciousness. Symptoms of poorly controlled hyperglycemia include polyuria, polydipsia, blurred vision, weight loss, polyphagia, stunting of growth, and vulnerability to infections or susceptibility to a more virulent or chronic course when infected.

Specific long-term complications of diabetes include (l) retinopathy with potential loss of vision, (2) nephropathy leading to end stage renal disease (ESRD), and (3) neuropathy with risk of foot ulcers, amputation, Charcot joints, sexual dysfunction, and potentially disabling dysfunction of the stomach, bowel, and bladder. Numerous mechanisms have been discovered that may mediate the specific tissue damage caused by hyperglycemia. Diabetic patients are also at increased risk for atherosclerotic cardiovascular, peripheral vascular, and cerebrovascular disease. These conditions may be related to hyperglycemia as well as to hypertension and abnormal lipoprotein profiles that are often found in diabetic patients.

Sufficient hyperglycemia to cause pathologic and functional changes in target tissues may be present for some time before clinical symptoms lead to a diagnosis of diabetes in many patients. At an even earlier stage, an incipient abnormality in glucose metabolism can be identified on plasma glucose testing, which indicates that the patient is at considerably increased risk for the full clinical disorder.

Classification

The classification of diabetes mellitus has recently been revised by a task force of the American Diabetes Association that included representation from Europe.[1] Major etiologic classes of the disease, along with more esoteric examples, have been categorized [*see Table 1*]. The vast majority of cases of diabetes mellitus are either type 1 (insulin-dependent) or type 2 (non–insulin-dependent) in an approximate ratio of 1:9.

type 1 and type 2 diabetes mellitus

Type 1 and type 2 diabetes were formerly known as insulin-dependent diabetes mellitus (IDDM) and non–insulin-dependent diabetes mellitus (NIDDM), respectively. This classification was abandoned largely because it was difficult to distinguish patients with IDDM from those patients with NIDDM who eventually required insulin treatment to mitigate hyperglycemia. Physicians, nurses, hospital-record-room personnel, health insurers, and even sometimes researchers were hard put to distinguish between these two forms of diabetes using the old terminology. The new classification, dependent on etiology rather than mode of treatment, puts a greater emphasis on the history and characteristics of the patients to determine the probable etiology and type. Two categories of blood glucose elevation, impaired glucose tolerance (IGT) and impaired fasting glucose (IFG), that lie between normal glucose levels and overt diabetes have also been established [*see Impaired Glucose Tolerance, below*].[2]

gestational diabetes mellitus

Gestational diabetes mellitus (GDM) constitutes a separate category for cases of diabetes first detected during pregnancy.[3] When diabetes is detected early in pregnancy, it is likely to be type 1 or type 2 diabetes mellitus that is presenting symptomatically and was probably precipitated or worsened by the pregnant state. Diabetes is commonly detected in the second and third trimester (i.e., in 4% of pregnant women) and is likely to be specific for the pregnant state, to be transient, and to reverse to normal glucose tolerance or to IGT on follow-up oral glucose tolerance testing 6 weeks after delivery. However, GDM is associated with a high risk of future diabetes, especially in women who have IGT post partum or who remain obese.[3] Permanent diabetes will develop in approximately 50% of patients within 10 years of GDM. The greatest importance of any single episode of GDM lies in the risks it poses to the fetus. These risks include intrauterine mortality, neonatal mortality, respiratory distress syndrome, hypoglycemia, hypocalcemia, jaundice, and macrosomia, which can cause trauma such as shoulder dystocia during passage through the birth canal.

secondary forms of diabetes mellitus

Of the categories of secondary diabetes [*see Table 1*], endocrinopathies and drug- or chemical-induced diabetes are noteworthy because they represent instances of diabetes that are potentially reversible if they are recognized and the physician can cure the endocrinopathy or discontinue the offending drug. The category of genetic defects in beta cell function illustrates how the classification will grow ever more detailed as knowledge increases. For example, the single diabetes mellitus phenotype formerly called maturity-onset diabetes of the young (MODY) can now be

more precisely classified into at least four genetic varieties, each of which arises from mutation of a different gene.

Diabetes caused by chronic pancreatitis, pancreatectomy, or occasionally carcinoma of the pancreas is usually type 1 in character. Because patients with this disease have glucagon as well as insulin deficiency, they are somewhat less likely to go into ketoacidosis[4] but are quite vulnerable to hypoglycemia. Because they are deficient in pancreatic enzymes, their digestion and subsequent absorption of nutrients is somewhat erratic, even though replacement enzymes are ingested with meals. If alcoholism, often the cause of chronic pancreatitis, is irremediable, it also contributes to blood glucose instability, as does the often accompanying irregular lifestyle. Small frequent doses of lispro insulin should be helpful, but safety may require less stringent blood glucose goals in such patients.

Although many individual drugs have been incriminated as a cause of hyperglycemia, the continued use of pharmacologic anti-inflammatory or immunosuppressive doses of synthetic glucocorticoids is an especially important continuing problem. Up to 25% of renal transplant patients develop so-called steroid diabetes.[5] In a case-control study, use of glucocorticoids for up to 45 days was a risk factor for diabetes that required pharmacologic treatment.[6] The odds ratio rose from 1.77 at a prednisone equivalent of 10 mg/day to an odds ratio of 10.3 at 30 mg/day. Obesity and family history of diabetes increased the risk of steroid diabetes. Although insulin resistance in the liver and muscle is a well-recognized effect of glucocorticoids, an action on the beta cells to limit the compensatory response to hyperglycemia[7] adds to the diabetogenic effect at higher steroid doses. Patients treated with glucocorticoids for more than a few days need to be warned to watch for and report clinical symptoms of hyperglycemia promptly. Ketoacidosis is rare, but hyperglycemic hyperosmolar nonketotic coma can occur. Insulin treatment is usually necessary for symptomatic patients and for those with a fasting plasma glucose (FPG) level greater than 200 mg/dl, but sulfonylurea drugs are sometimes effective. There is little systematic information on the efficacy of the other oral agents. In most instances, steroid diabetes is transient, but in a minority of cases, diabetes persists even after withdrawal of the glucocorticoids.

Table 1 Etiologic Classification of Diabetes

- Type 1 diabetes mellitus* (β cell destruction, usually leading to absolute insulin deficiency)
  - Immune mediated
  - Idiopathic
- Type 2 diabetes mellitus* (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly insulin secretory defect with insulin resistance)
- Other specific types of diabetes
  - Genetic defects of β cell function
    - Chromosome 12, HNF-1α (formerly MODY3)
    - Chromosome 7, glucokinase (formerly MODY2)
    - Chromosome 20, HNF-4α (formerly MODY2)
  - Genetic defects in insulin action
    - Type A insulin resistance
  - Disease of the exocrine pancreas
    - Pancreatitis
    - Trauma/pancreatectomy
    - Neoplasia
  - Endocrinopathies
    - Acromegaly
    - Cushing syndrome
    - Glucagonoma
  - Drug- or chemical-induced
    - Nicotinic acid
    - Glucocorticoids
    - Thiazides
  - Infections
    - Congenital rubella
    - Cytomegalovirus
  - Uncommon forms of immune-mediated diabetes
    - Stiff-man syndrome
    - Anti-insulin receptor antibodies
  - Other genetic syndromes associated with diabetes
    - Down syndrome
    - Turner syndrome
    - Friedreich ataxia
    - Myotonic dystrophy
- Gestational diabetes mellitus (GDM)

Note: The list of other specific types of diabetes is not comprehensive. There are many other such syndromes.
*Patients with any form of diabetes may require insulin treatment at some stage of their disease. Such use of insulin does not, of itself, classify the patient.

Screening for Diabetes

Screening for type 1 diabetes mellitus by office glucose testing is currently indicated in high-risk patients. Current American Diabetes Association criteria for office screening of asymptomatic individuals for type 2 diabetes mellitus employ FPG levels.[1] Screening is recommended in all individuals 45 years of age and older at 3-year intervals. Younger individuals should be screened if they are obese (> 120% desirable body weight or a body mass index ≥ 27), have a first-degree relative with diabetes, are members of a high-risk ethnic population (African American, Hispanic American, Native American, Asian American), have delivered a baby weighing more than 9 lb, have previously had GDM, are hypertensive (blood pressure ≥ 140/90 mm Hg), have atherogenic dyslipidemia (high-density lipoprotein [HDL] cholesterol levels ≤ 35 mg/dl or triglyceride levels ≥ 250 mg/dl) or had IFG or IGT on previous testing.[1] Mass indiscriminate public screening is not justified, because there is as yet no proof of population benefit.

Epidemiology

TYPE 1 DIABETES MELLITUS

Available, but not up-to-date, studies suggest the prevalence of type 1 diabetes mellitus in the United States is 1.7 per 1,000 in individuals younger than 19 years and 2.1 per 1,000 in adults.[8] A total prevalence of approximately 500,000 is estimated. Current estimates of annual incidence are 18 per 100,000 population in the 0- to 19-year age range and 9 per 100,000 population in those older than 20 years.[8] Approximately 30,000 cases of type 1 diabetes mellitus are estimated to occur yearly in the United States, and it is more common in whites than in African Americans. Worldwide, the highest annual incidence of type 1 diabetes mellitus is found in Finland (35 cases per 100,000) and the lowest is found in Korea (< 1 per 100,000).

TYPE 2 DIABETES MELLITUS

Analysis of data from the third National Health and Nutrition Examination Survey (NHANES III), conducted from 1988 to 1994,[9] indicates a prevalence of 5.1% for adults at least 20 years of

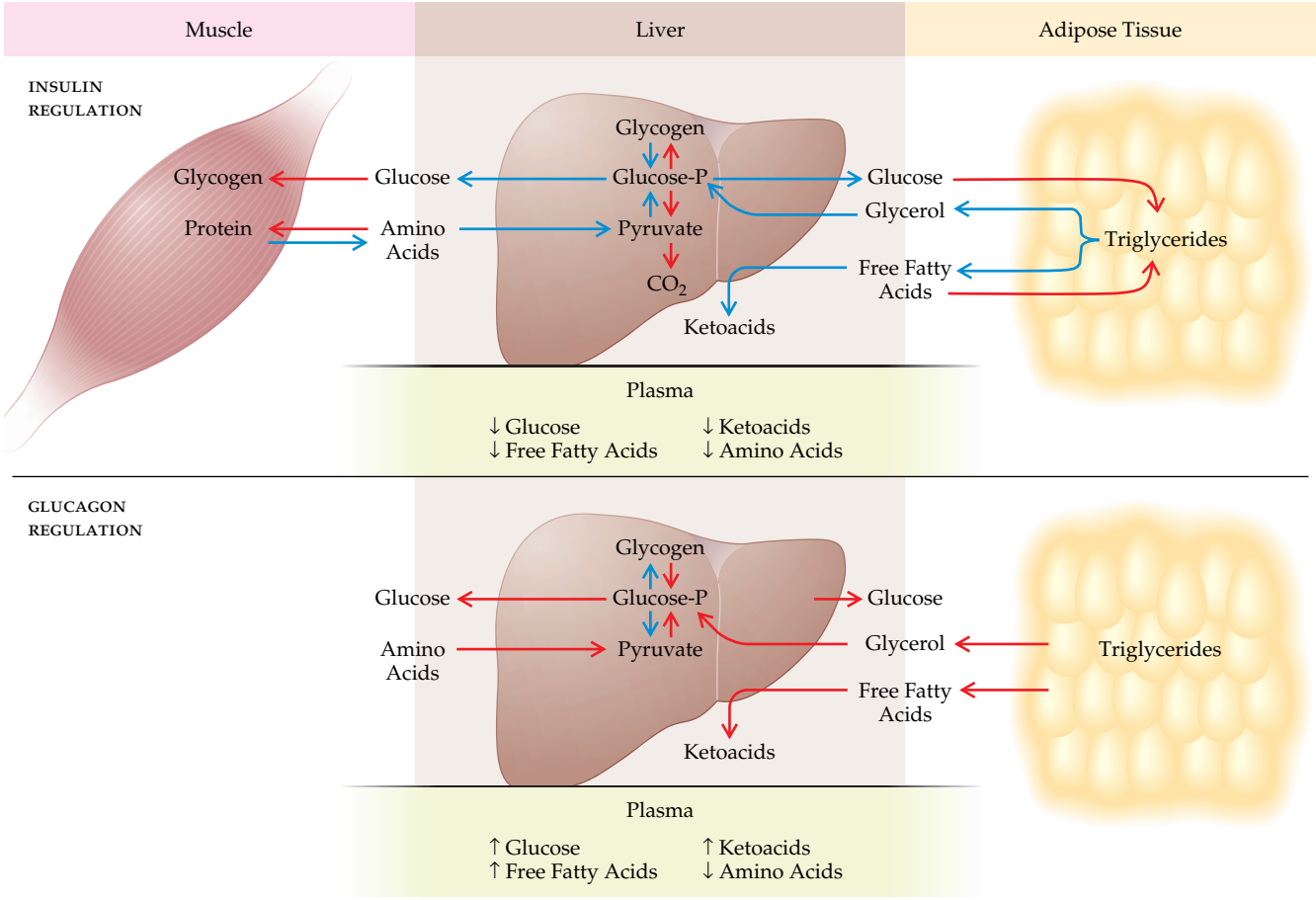


Figure 1 **The opposing actions of insulin and glucagon, particularly within the liver, on substrate flow and plasma levels are seen here. The two hormones have directly opposite effects on key enzymes, such as glycogen synthase and phosphorylase. Thus, stimulatory effects of glucagons on glucose and ketoacid production are magnified when insulin is deficient, as in type 1 diabetes mellitus. Red arrows indicate stimulation. Blue arrows indicate inhibition.**

age in the United States and a prevalence of 2.7% of undiagnosed diabetes (FPG ≥ 126 mg/dl). A prevalence of 12.3% (diagnosed plus undiagnosed) was estimated for individuals 40 to 74 years of age. There are an estimated 10.2 million diagnosed and 5.4 million undiagnosed cases of diabetes in the United States. The estimated number of persons with IGT approximately equals the number with diabetes. Non-Hispanic African-American and Mexican-American women have nearly twice the prevalence of diabetes as non-Hispanic white women. Non-Hispanic African-American men have a slightly higher risk than non-Hispanic white men, but Mexican-American men have about a 50% greater risk than non-Hispanic white men.[9]

Annual incidence of type 2 diabetes mellitus per 100,000 population ranges from 180 in 25 to 44 year olds to a peak of 860 in 65 to 74 year olds. Approximately 625,000 cases of type 2 diabetes mellitus develop yearly in the United States.[10] The prevalence is expected to rise from 15 million in the year 2000 to 21 million in 2025. Worldwide, the prevalence of type 2 diabetes mellitus will likely increase from 150 million to 300 million during that time.[11] The increase[12] reflects aging of the population, strikingly increased obesity,[13] and a sedentary lifestyle. This rise in the number of cases is especially troubling in regard to high-risk ethnic minorities whose access to medical care may be limited.[14,15]

Obesity is a major risk factor for type 2 diabetes mellitus.[16] The current definition of obesity employs the body mass index (BMI) (body weight in kilograms divided by height in meters squared). A person with a BMI of at least 25 but less than 30 is defined as overweight.[17] A BMI of 30 or more is defined as obesity,[16] and a BMI of 40 and above is associated with a 15-fold increased risk of type 2 diabetes mellitus.[15] Abdominal obesity, defined as a waist circumference greater than 100 cm in men and greater than 88 cm in women or a waist-to-hip ratio greater than 0.9, is an especially strong risk factor for type 2 diabetes mellitus. A large preponderance of patients with type 2 diabetes mellitus are obese; even those with normal BMI may have an increased percentage of their body weight accounted for by fat.[18] Longer duration of obesity further increases the risk of diabetes, emphasizing the importance of early efforts to control weight. Many patients with type 2 diabetes mellitus have a strong family history of that disease in first-degree relatives. An extraordinary example is found among the Arizona Pima Indians on the Gila River reservation, where 50% of the adult population has type 2 diabetes mellitus. Other risk factors for the disease include physical inactivity, hypertension, dyslipidemia, gestational diabetes, low birth weight, low income, low level of education, and low socioeconomic status.[1,19]

Hormonal Regulation of Metabolism

Diabetes involves the most fundamental aspects of human metabolism. The following are all affected by the hormonal abnormalities of diabetes: energy production and expenditure; the proportioning of carbohydrate, fat, and protein as energy sourc-

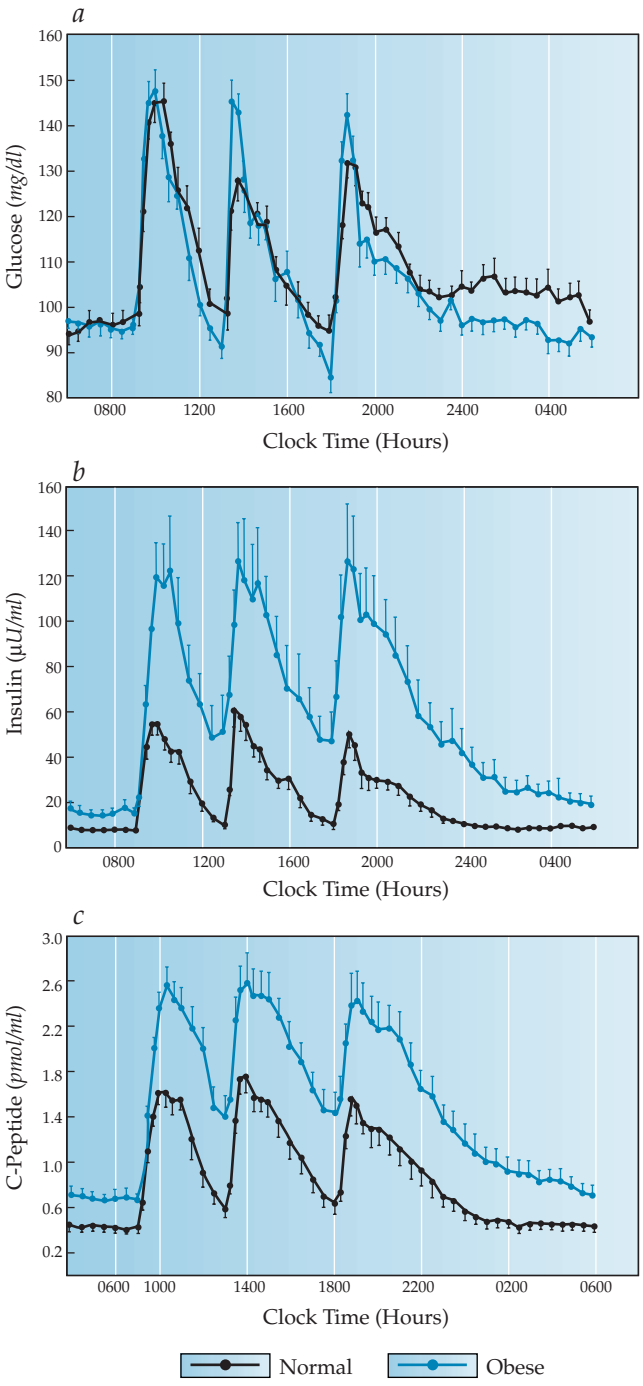


Figure 2 **Plasma glucose (*a*) is normally kept within a narrow range throughout the day, largely because of beta cell function. Plasma insulin (*b*) and plasma C-peptide (*c*) rise sharply from their basal levels with each meal and, after reaching peaks, return promptly to basal levels, which are maintained throughout the night. Note also that plasma insulin and C-peptide levels are elevated in obese individuals who are insulin resistant.**

es; the storage of energy as carbohydrate and fat; and the balance between protein synthesis (anabolism) and degradation (catabolism). To understand the pathogenesis of diabetes, it is useful to start with a brief review of normal metabolism.[20]

A proper balance between insulin and glucagon is one crucial hormonal regulator of basal metabolic homeostasis.[20] Insulin primarily facilitates storage of glucose as glycogen, free fatty acids in triglycerides, and amino acids in protein, and it inhibits glycogenolysis, lipolysis, ketogenesis, proteolysis, and gluconeogenesis [*see Figure 1*]. Glucagon stimulates mobilization of glucose, free fatty acids, and glycerol and stimulates hepatic uptake of amino acids and the conversion of their carbon skeletons to glucose. Glucagon also stimulates ketogenesis from free fatty acids. The normal steady-state levels of insulin and glucagon help maintain the overnight FPG level at 60 to 110 mg/dl, free fatty acid levels at less than 0.7 mmol/L, ketoacids at less than 0.2 mmol/L, and each amino acid at its unique level. After a mixed meal, plasma insulin rises sharply [*see Figure 2*] and, with it, the insulin-glucagon ratio. This condition reverses all the previously described processes. Dietary carbohydrate is stored in muscle and liver glycogen, free fatty acids are reesterified and stored as triglycerides in adipose tissue, and protein metabolism shifts back toward anabolism. When all the nutrients have been assimilated and plasma glucose returns to its basal preprandial level, plasma insulin [*see Figure 2*] and the insulin-glucagon ratio promptly return to basal levels, preventing an overshoot of insulin action that would otherwise cause hypoglycemia. Thus, an immediate rise, an early peak, and a prompt fall in insulin secretion are requisite to normal postprandial metabolism [*see Figure 2*].

Insulin is synthesized in pancreatic islet beta cells from a larger molecule called proinsulin, which is then split to yield insulin and an intramolecular connecting peptide called C-peptide [*see Figure 3*]. The two molecules are stored in the same granules and secreted in an equimolar ratio when the beta cell is stimulated. Thus, plasma C-peptide levels are a faithful marker of beta cell function [*see Figure 3*].

Insulin acts via a plasma insulin receptor that leads to the generation of multiple mediators of insulin's numerous intracellular cytoplasmic and nuclear effects [*see Figure 4*]. Insulin regulates both the activities and syntheses of target enzymes. Sensitivity of target tissues to insulin is the other major determinant of insulin action. Insulin sensitivity is best measured in humans by infusing insulin to establish steady-state plasma insulin levels [*see Figure 5*]. Simultaneously, the baseline plasma glucose is maintained at a constant level by a variable glucose infusion. The amount of glucose required to prevent plasma glucose from decreasing under the effect of insulin is equal to the increased amount of glucose being used per unit time under insulin stimulation (assuming that insulin has completely suppressed hepatic glucose output by the liver). The quantity of glucose used per unit time divided by the plasma insulin level provides an index of whole body sensitivity to insulin in the sphere of glucose metabolism.

A feedback loop exists between insulin responsiveness in target tissues and insulin secretion by beta cells. This relation operates to increase insulin secretion in individuals relatively resistant to insulin action and to decrease insulin release in individuals very sensitive to insulin action. The result is one critical mechanism for maintaining fasting and postprandial plasma glucose levels within narrow normal ranges.

Pathogenesis of Microvascular Complications in Diabetes

A distinctive feature of diabetes—the microvascular complications—were only revealed or commonly appreciated after the introduction of insulin therapy in 1922 allowed patients with type 1 diabetes mellitus to live long enough to experience these complications. It should be borne in mind that the descriptions and pathogenetic sequences presented below reflect a former

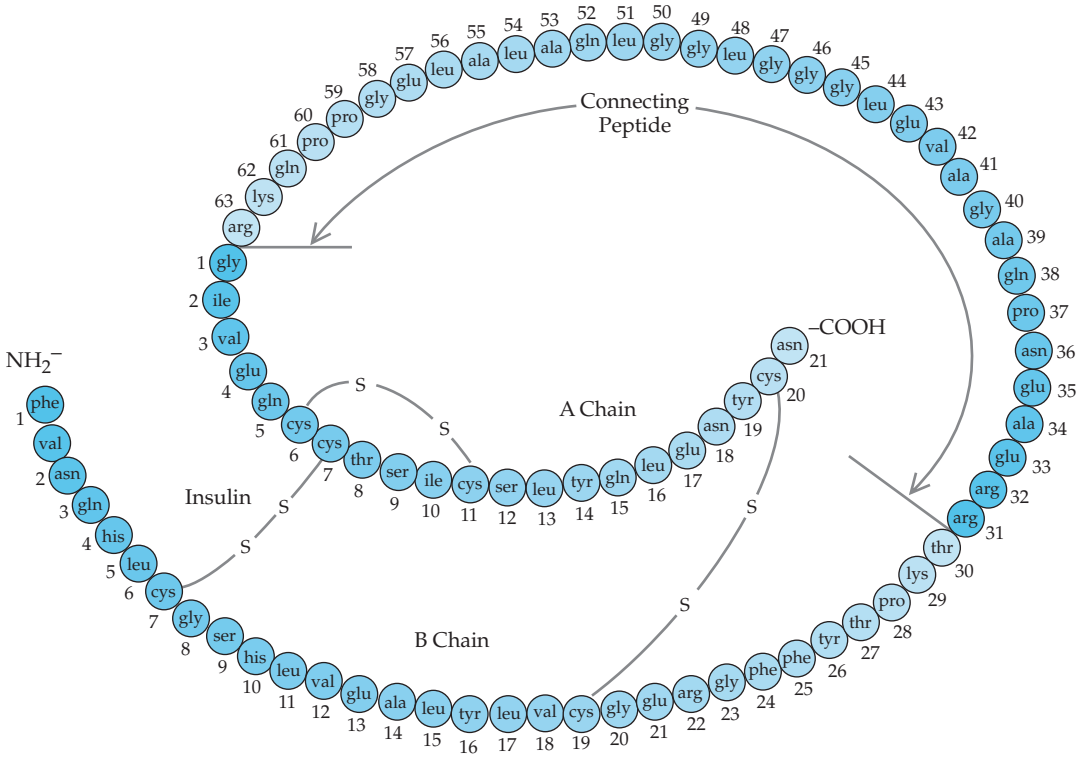


Figure 3 **The structure of human proinsulin, the precursor molecule to insulin. The peptide that connects the amino terminus (NH_2^-) of the A chain to the carboxyl terminus (–COOH) of the B chain is called connecting peptide (C-peptide). Proinsulin is converted to insulin and C-peptide, and these two molecules are packaged together in the secretory granule. On stimulation of the beta cell, C-peptide and insulin are secreted in equimolar proportions. Thus, C-peptide levels reflect beta cell functional capacity.**

commonly practiced degree of metabolic control no longer considered acceptable. Prevention of these complications is a major goal of current therapeutic policy and recommendations for all but transient forms of diabetes [*see* Prevention and Treatment of Microvascular Complications, *below*].

RETINOPATHY

Given a long enough duration, retinopathy occurs in almost all patients with type 1 diabetes mellitus and in most patients with type 2 diabetes mellitus who are on conventional treatment that does not come close to normalizing glycemic levels [*see Table 2*].[21] The most common form of retinopathy is nonproliferative retinopathy (also termed background retinopathy). It begins with loss of capillary pericytes, the supporting cells of the retinal vasculature, a loss leading to capillary dilatations that are seen on direct fundoscopy as microaneurysms [*see Figure 6a*]. Microaneurysms measure 50 to100 µm in diameter and can occur anywhere in the retina. However, they tend to cluster near the macula, the area responsible for central vision and visual acuity. Small dot hemorrhages form when microaneurysms leak blood. Hard lipid exudates form on leakage of serum [*see Figure 6a*]. These lesions are usually benign unless they occur quite close to the macula and in sufficient number to cause clinically significant macular edema. The latter is a feared complication that can decrease central vision and acuity. Capillary closure, which actually begins in the phase of background retinopathy, increases; and in the phase of preproliferative retinopathy, enough capillaries become obstructed to cause ischemia of the retina. Infarctions of the retinal nerve layer appear as soft (cotton wool) exudates. The retina responds to further ischemia with proliferation of new blood vessels from its surface [*see Figure 6b*]. In this phase of proliferative retinopathy, ischemic retina releases vascular endothelial growth factor (VEGF), which stimulates new vessel formation. These new vessels grow forward into the vitreous. They are extremely fragile and can bleed into the vitreous, causing temporary loss of vision until the blood is reabsorbed. If no reabsorption occurs, blindness can result unless successful vitrectomy is carried out. Proliferative vessels that cover more than one fourth of the disk diameter and that occur within 1 disk diameter of the disk [*see Figure 6b*] are especially likely to bleed. Even after reabsorption of the vitreous blood, fibrous scars form that can cause traction on the retina and can lead to retinal detachment, another cause of profound and often permanent loss of vision.

NEPHROPATHY

Diabetic nephropathy [*see Figure 7*] is the complication associated with the highest mortality. Between 35% and 45% of patients with type 1 diabetes mellitus and a somewhat smaller percentage of patients with type 2 diabetes mellitus experience significant nephropathy.[22-24] Histologically, the earliest change is thickening of the capillary basement membrane. Subsequently, mesangial material accumulates diffusely throughout the glomerulus [*see Figure 8*]. Ultimately, there is loss of podocytes and development of peritubular fibrosis. Excretion of low but abnormal levels of albumin in the urine is a marker of the incipient phase of nephropathy.[25] As glomeruli become increasingly filled with mesangial matrix products, albuminuria increases and eventually gross proteinuria appears. Microalbuminuria is defined as excretion of 30 to 300 mg of albumin a day or an albumin-creatinine ratio between 30 and 300 in a random urine spec-

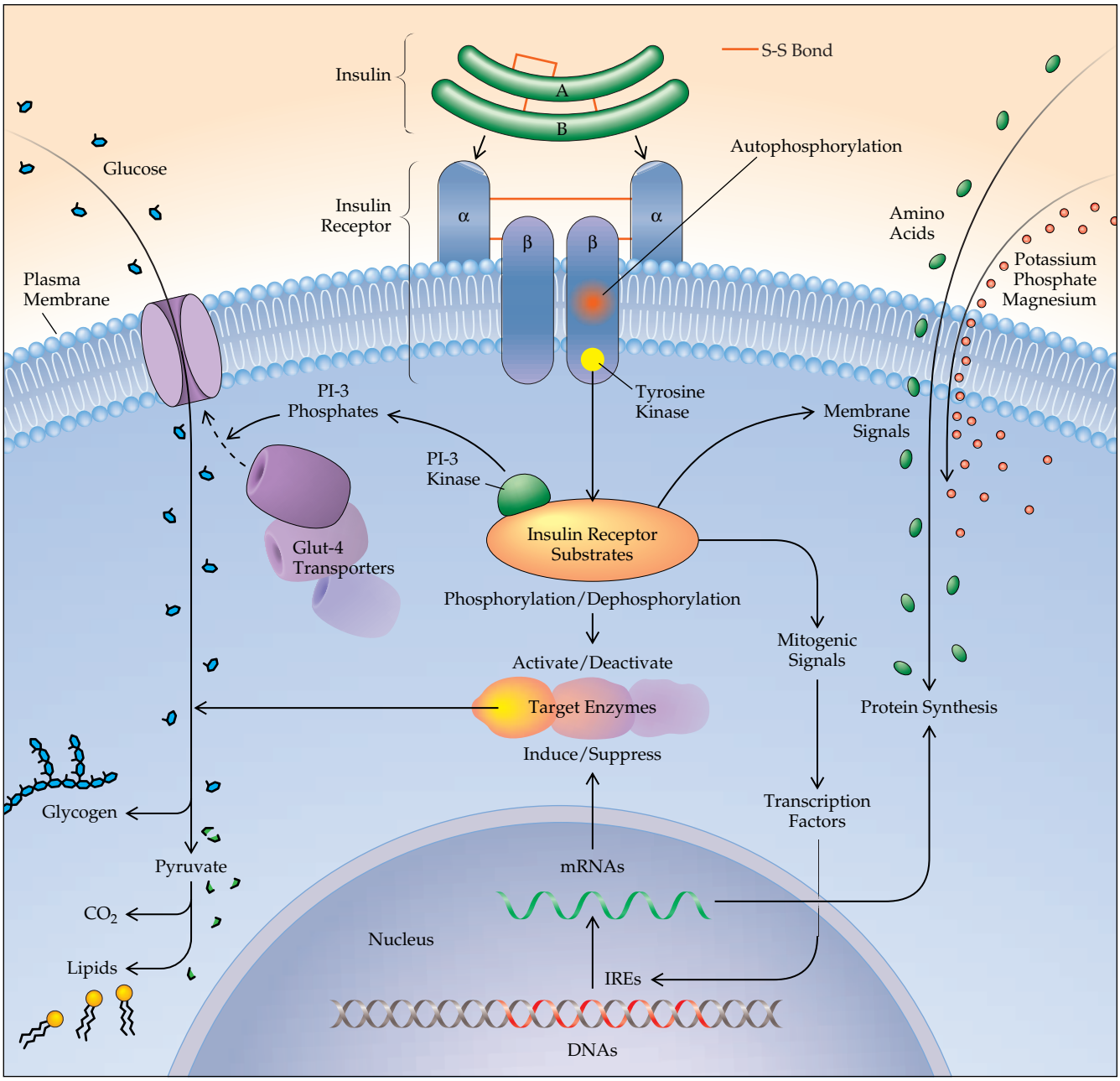


Figure 4 **The cellular actions of insulin begin with binding to its plasma membrane receptor. As a result, certain tyrosine molecules in the intracellular portion of the transmembrane receptor are autophosphorylated, creating tyrosine kinase activity in the receptor. Several intracellular insulin receptor substrates (IRS) are then tyrosine phosphorylated by the receptor. Phosphorylated IRS docks and either activates or inactivates numerous enzymes (e.g., phosphatidylinositol-3-kinase [PI-3 kinase]) and other mediating molecules. Among the chief effects of these insulin-stimulated cascades are translocation of glucose (Glut-4) transporters to the plasma membrane, where they facilitate glucose diffusion into the cell; shifting of intracellular glucose metabolism toward storage as glycogen by activating glycogen synthase; stimulation of cellular uptake of amino acids, phosphate, potassium, and magnesium; stimulation of protein synthesis and inhibition of proteolysis; and regulation of gene expression via insulin regulatory elements (IRE) in target DNA molecules. Numerous intermediates in these various pathways, along with the molecules mentioned above, are products of candidate genes whose mutation could produce the state of insulin resistance characteristic of type 2 diabetes mellitus. Red connectors between insulin chains A and B and among insulin receptor subunits α and β indicate S-S bonds. The A chain also has an intramolecular S-S bond.**

imen. Clinical proteinuria is defined as excretion of more than 0.5 g of total protein a day. This level of excretion can be detected by a positive dipstick urine test for protein. The nephrotic syndrome may also eventually occur.

Early in type 1 diabetes mellitus, kidney size and glomerular filtration rate (GFR) may actually be greater than normal. However, in both types of diabetes, GFR begins to decline, and after clinical proteinuria develops, GFR almost inexorably falls to the level of ESRD [*see Figure 7*]. Unlike the risk of retinopathy, the risk of nephropathy does not continue to rise with increasing duration. The incidence of nephropathy peaks at approximately 15 to 17 years and declines somewhat thereafter.[26] The prevalence of nephropathy remains approximately constant after that time. If the dipstick test has not revealed proteinuria by 25 to 30 years

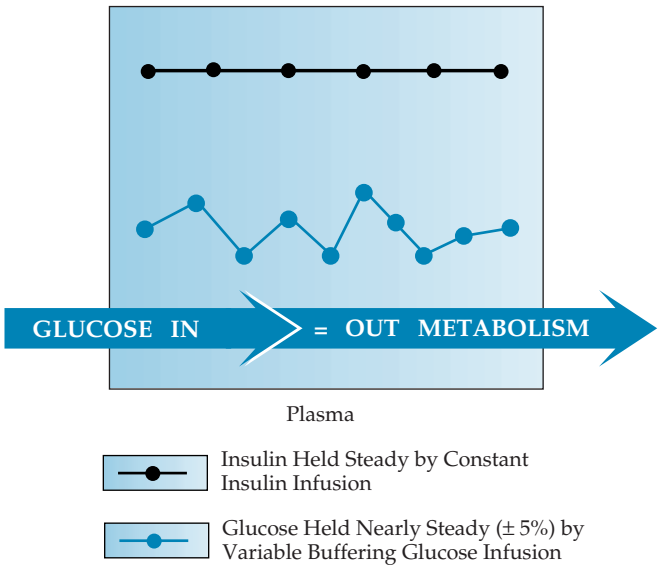


Figure 5 **The diagram represents the gold standard for measuring the sensitivity of glucose metabolism to insulin, utilizing a glucose insulin clamp. When steady state is reached, glucose metabolized/unit time = glucose infused/unit time. Assuming endogenous glucose production is suppressed to zero, insulin sensitivity = (glucose metabolized/unit time) ÷ plasma insulin. For each dose of insulin, the more exogenous glucose required to sustain plasma glucose at its basal levels, the greater the insulin sensitivity. Conversely, individuals who require lesser amounts of glucose than usual to maintain the basal plasma glucose level are insulin resistant. The latter is usually the case in type 2 diabetes mellitus.**

of diabetes duration, the risk of ESRD decreases. Coincident with or shortly after the development of microalbuminuria, hypertension often appears. Hypertension in turn further aggravates diabetic nephropathy and is an important component in the progression to renal failure.

NEUROPATHY

Neuropathy has protean manifestations in diabetes. The most common presentation is peripheral symmetrical sensorimotor neuropathy, which causes numbness or tingling in the toes and feet.[27] At this point, symptoms are only mildly disturbing and require no specific treatment. These symptoms may even abate over time as neuropathy becomes more severe and hypoesthesia or anesthesia takes the place of paresthesias and dysesthesias. Ultimately, insensate feet become very vulnerable to trauma, and neuropathic foot ulcers are frequent causes of hospitalization and even amputation. Testing sensation with a nylon monofilament providing a calibrated 10 g point pressure is an effective way to screen for high risk of foot ulcers. Patients who cannot detect the pressure of the nylon filament have a 30- to 40-fold increased risk of foot ulcer.[28] In some instances, neuropathy is manifested by severe pain that can interfere with sleep and normal daily activities. The distribution of pain can suggest mononeuropathy and radiculopathy. Abrupt onset of cranial neuropathies that most commonly give rise to extraocular muscle weakness and diplopia has been attributed to microinfarcts caused by thrombosis of nutrient blood vessels. Carpal tunnel syndrome and other entrapment syndromes are more frequent in diabetic patients than in nondiabetic patients.

Involvement of the autonomic nervous system is also common and can become debilitating. Manifestations include male impotence and female anorgasmia, difficulty voiding and urinary retention, impaired gastric emptying with early satiety and emesis, diarrhea, orthostatic hypotension, and decreased sweating and vasomotor tone in the lower extremities. The combination of decreased sympathetic tone and loss of vagal control of the heart rate can produce persistent resting sinus tachycardia; sudden death can result.

A form of diabetic neuropathy called amyotrophy occurs most commonly in elderly men with diabetes. It is manifested by severe, unremitting pain and weakness in the thigh muscles. Severe depression, cachexia, and weight loss may mark the 1- to 2-year course of this form of neuropathy. Sometimes confused with painful neuropathy are rare muscle infarcts, usually occurring in the thigh muscles. These infarcts are marked by abrupt onset of severe pain lasting several months. Magnetic resonance imaging of the affected area can demonstrate the presence of necrosis.

Diabetic neuropathy may be another microvascular complication, but the pathogenesis is still not completely understood.[29] Demyelinization of nerves is manifested by decreases in motor and sensory nerve conduction velocities. Axonal degeneration is reflected in decreased amplitudes of action potentials. Histologically, swelling is seen at the axonal nodes. An inflammatory component to diabetic neuropathy has also been suggested.[30]

RELATION OF MICROVASCULAR COMPLICATIONS TO GLYCEMIA

The appearance of microvascular complications in the 1930s generated a 50-year debate about whether diabetic retinopathy, nephropathy, and neuropathy were the direct result of the metabolic abnormalities, most notably hyperglycemia, or whether they were a parallel independent consequence of diabetes that had formerly been usually preempted by death from extreme metabolic disequilibrium (i.e., diabetic coma). This debate ultimately came to encompass type 2 diabetes mellitus as well. The debate was not merely academic, because it was reflected in quite different approaches to treatment. A belief in the metabolic hyperglycemic cause of retinopathy, nephropathy, and neuropathy impelled the physician to work with inadequate means to help the patient achieve as close to normal blood glucose levels as possible. Conversely, a belief in the metabolically independent nature

Table 2 Diabetic Retinopathy

*Stage**	*Pathologic Process*	*Manifestations*
Background	Loss of capillary integrity Leakage, exudation, diapedesis Early capillary closure	Microaneurysms Dot hemorrhages Hard exudates Macular edema
Preproliferative	Capillary closure Microinfarcts Ischemia	Blot hemorrhages Soft exudates Intraretinal microvascular abnormalities Venous beading Macular edema
Proliferative	Forward growth of new large vessels Fibrosis Traction on retina or vitreous	Preretinal hemorrhage Vitreous hemorrhage Retinal detachment Macular edema

*Loss of visual acuity may occur from macular edema at any stage. Blindness may occur from severe macular edema, vitreous hemorrhage, or retinal detachment.

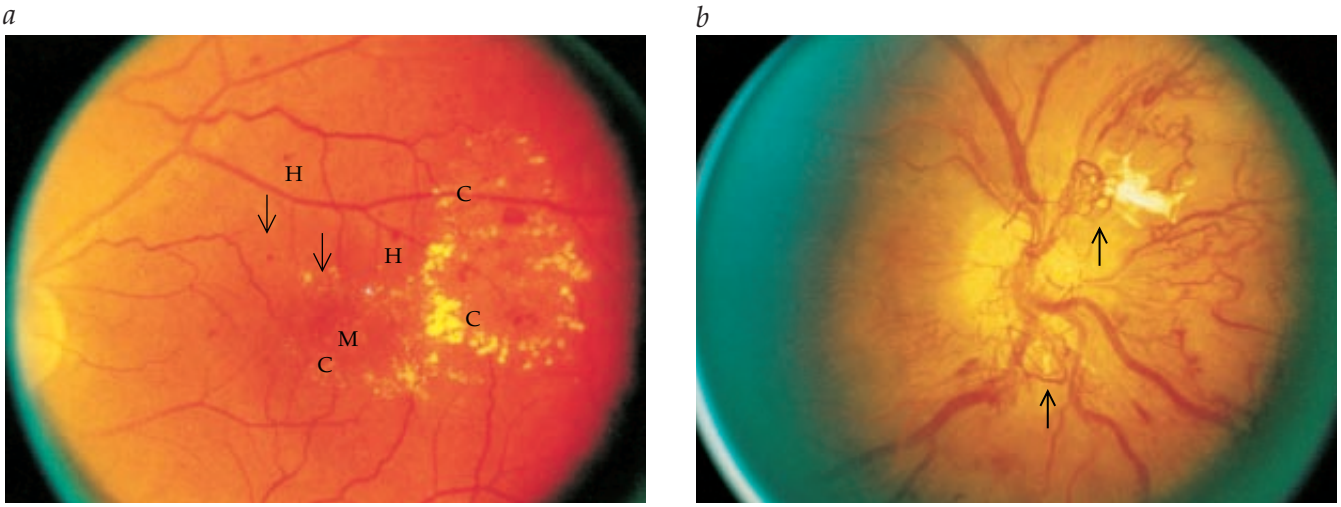


Figure 6 **(*a*) This fundus photograph reveals nonproliferative (or background) retinopathy in a diabetic patient. Microaneurysms (arrows) occur at end capillaries. Punctate (or dot-and-blot) hemorrhages (H) and hard exudates (C) can also be seen. The hard exudates form three distinct circles (termed circinate retinopathy), which indicate leakage of plasma proteins from abnormal vessels located in the centers of the three circles. Lesions in the area of the macula (M) are potentially more dangerous, as they may lead to macular edema requiring laser therapy. (*b*) In proliferative retinopathy, new vessels grow from the retina into the vitreous. This fundus photograph reveals fine, tangled, new vessels originating from several areas of the disk (arrows). The vessels often form arcades and characteristically have thin walls and are fragile. They tend to bleed into the vitreous; the scars that form can cause retinal detachment and loss of vision. Proliferation within one disk diameter of the disk (termed neovascularization of the disk) is particularly dangerous, as these vessels are especially prone to bleed and form traction scars.**

of these complications encouraged a somewhat more laissez-faire approach, which attempted primarily to eliminate the immediate symptoms, such as polyuria, that were produced by plasma glucose levels exceeding the renal threshold (> 180 mg/dl). Furthermore, the risks associated with the more aggressive approach to hyperglycemia reinforced the arguments of the conservative practitioners. A large body of evidence was eventually built up that supported but did not prove the so-called glucose hypothesis.[31] The Diabetes Control and Complications Trial (DCCT) and United Kingdom Prospective Diabetes Study (UKPDS) ended this debate for type 1 and type 2 diabetes mellitus, respectively.

The DCCT[32] was a randomized clinical trial that enrolled 1,441 nonobese patients, aged 13 to 39 years, with type 1 diabetes mellitus. Half of the patients with diabetes of 1 to 5 years' duration participated in a primary prevention trial that excluded all patients with retinopathy or microalbuminuria, and half of the patients with diabetes of 1 to 15 years' duration participated in a secondary intervention trial that included only patients who already had mild to moderate nonproliferative diabetic retinopathy but less than 200 mg/day of urinary albumin excretion. In both of these DCCT trials, patients were randomly assigned either to receive conventional treatment (no more than two insulin injections a day) or to receive intensive treatment (three to four insulin injections a day or use of a continuous subcutaneous insulin infusion [CSII] pump; self-monitoring of blood glucose at least four times a day; premeal target blood glucose levels of 70 to 120 mg/dl; glycated hemoglobin [HbA_{1c}] goal of less than 6.05%; and very frequent contacts between patient and treatment team). An HbA_{1c} difference of 1.8% (8.9% versus 7.1%) was maintained between the two treatment groups for up to 9 years.[33]

Over a mean follow-up of 6.5 years, intensive treatment produced substantial benefits. The risks of de novo development (primary prevention trial) or of progression (secondary intervention trial) of retinopathy were reduced by 27% to 76%; the development of microalbuminuria was reduced by 35%; macroalbu-

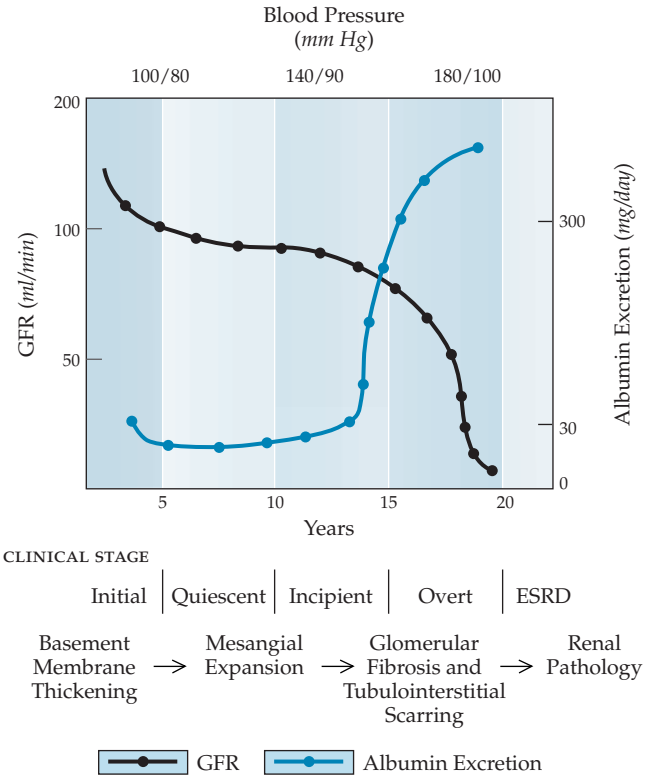
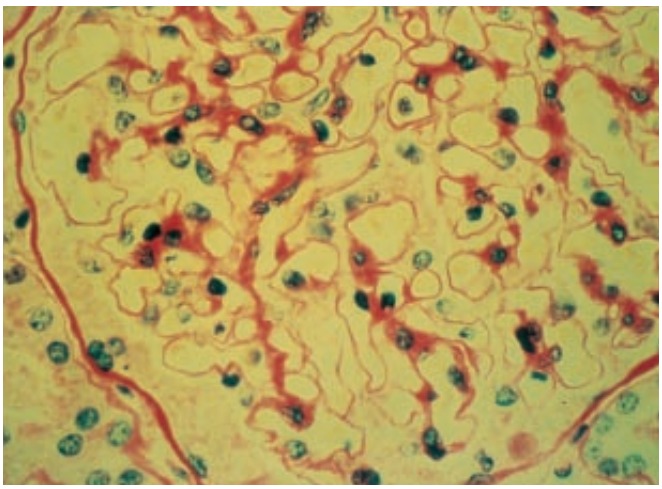


Figure 7 **Relation of the developing histopathologic changes in the kidney to the development of renal functional abnormalities. Note that GFR is actually elevated early, corresponding to early renal hypertrophy. The appearance of microalbuminuria (albumin excretion > 30 mg/day) indicates that the patient is at considerable risk for overt nephropathy and end-stage renal disease (ESRD), but not all such individuals suffer this fate. Blood pressure begins to rise at about the time that microalbuminuria appears, and hypertension further damages the kidney.**

a



b

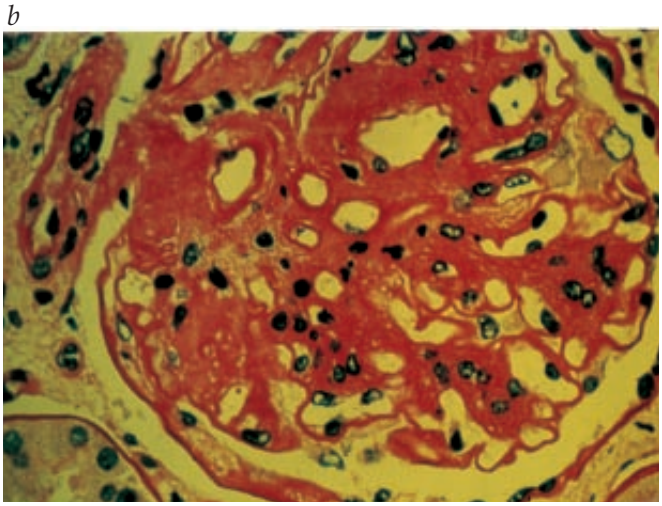


Figure 8 **(*a*) The normal glomerulus with a large filtration surface has a lacy appearance. (*b*) There is diffuse deposition of extramesangial material throughout, as well as thickening of capillary basement membranes in a diabetic glomerulus. The GFR through such a glomerulus is reduced.**

minuria (i.e., proteinuria) was reduced by 56%; and development of clinical neuropathy, confirmed by abnormal nerve conduction velocities or autonomic nervous system function tests, was reduced by 60%.[32] Patients in the primary prevention cohort, with a mean diabetes duration of 2.5 years, had a greater response to intensive treatment than did patients in the secondary prevention cohort, with a mean diabetes duration of 8.5 years.

The main adverse effect of intensive treatment was a threefold increase in the risk of severe hypoglycemic episodes characterized by coma, convulsions, or the required assistance of others to treat and reverse the episode.[32,34] At least one such event per year was experienced by 25% of intensively treated patients, and 50% had experienced more than one such episode by the end of the study[34];14% experienced 10 or more episodes. The overall rate of severe hypoglycemia was 62 events per 100 patient-years for intensive treatment, compared with 19 events per 100 patient-years for conventional treatment. In addition, intensive treatment caused greater weight gain; one third of the patients exceeded 120% of ideal body weight (approximate BMI, 27) by the end of the study.[32] Intensive treatment was also more expensive than conventional treatment.[35] However, the cost was partly offset by projected decreased costs of a lower rate of complications,[36] and the estimated cost per year of quality life gained was $28,661, a figure thought to represent a good value.

The UKPDS[37,38] enrolled 5,102 patients with newly diagnosed type 2 diabetes mellitus, a mean age of 53 years, and a mean BMI of 28. After a 3-month dietary run-in, 1,138 patients were randomly assigned to a continuation of diet treatment only as long as their FPG remained below 270 mg/dl and they had no hyperglycemic symptoms. In the study, 2,729 patients were randomly assigned to intensive treatment, 1,573 to receive one of three sulfonylurea (SU) drugs, and 1,156 to receive insulin. In two thirds of the clinical sites, 342 patients were also randomized to intensive treatment with metformin. The goal of intensive treatment was an FPG of less than 108 mg/dl. Of the conventional-treatment patients, 80% ultimately required drugs to maintain their treatment goals of an FPG of less than 270 mg/dl and freedom from symptoms, although nearly 60% of their total treatment time was spent on diet therapy alone. Likewise, in the intensive-treatment groups, metformin therapy had to be added to the SU therapy, and insulin had to be substituted for or added to oral-drug therapy to maintain the stringent treatment goal.

Despite these drug crossovers, after 10 years of follow-up, patients who received intensive treatment showed a 25% decrease in the risk of serious microvascular complications (vitreous hemorrhage, need for laser treatment, and renal failure), compared with patients given conventional treatment.[37] This important benefit was associated with an HbA_{1c} difference of 0.9% (7.9% for conventional therapy; 7.0% for intensive therapy). Serious hypoglycemia occurred in 3% of insulin-treated patients each year and in 1% to 2% of SU-treated patients. These rates were much lower than that experienced with intensive treatment in patients with type 1 diabetes mellitus in the DCCT.

These two trials provided experimental proof that microvascular and neuropathic complications could be prevented or at least substantially delayed by maintaining blood glucose levels as near to normal as treatment techniques would safely allow. Although these two experimental trials did not prove that hyperglycemia caused microvascular complications, both trials provided additional strong evidence supporting that hypothesis. In the DCCT, the risk of retinopathy was directly related to the preceding mean HbA_{1c} difference in a similar exponential fashion in each of the two treatment groups.[39] The risk of retinopathy was decreased by about 44% for each proportional 10% decrease in HbA_{1c} (e.g., a decrease in HbA_{1c} from 10% to 9.0%). Microalbuminuria and neuropathy showed similar risk relations with glycemia. In the UKPDS, the risk of microvascular complications was also directly related to the mean HbA_{1c} in an exponential fashion.[40] The risk of these complications was decreased by about 37% for every absolute decrease of 1% in HbA_{1c}. These similarities suggest that similar biologic processes are at work. Neither the UKPDS nor the DCCT analyses indicated any glycemic threshold in the diabetic range of HbA_{1c}, below which there was no further risk of microvascular complications.[40,41] This observation sets normoglycemia as the ultimate goal of treating type 1 and type 2 diabetes mellitus. Furthermore, the benefits of previous intensive treatment (or the adverse effects of previous conventional treatment) are still demonstrable 7 years after the DCCT was completed, during which time interval the mean HbA_{1c} concentrations in both groups were nearly identical (approximately 8.0%).[42] Thus, sustained periods of glycemic exposure are associated with pro-

longed consequences. An unacceptable level of hyperglycemia continues to have adverse effects even after some improvement in metabolic control, and a marked reduction in hyperglycemia with intensive treatment continues to have beneficial effects even after some worsening in metabolic control.

Multiple mechanisms by which increased glucose concentrations may cause damage to the retina, kidney, and nerves have been discovered [*see Figure 9*]. (1) Glucose itself can react nonenzymatically with free amino groups in N-terminal amino acids and lysine residues of proteins. $HbA_{1c}$is one such molecule. This reaction sets into motion cross-linking of proteins that ultimately generate harmful advanced glycation end products (AGEs).[43,44] Such products include carboxymethyllysine and pentosidine. Concentrations of long-lived AGEs were higher in tissues of conventionally treated patients in the DCCT than in tissues of intensively treated patients in the DCCT.[45] AGEs correlated with HbA_{1c} and, independent of HbA_{1c}, with the presence of retinopathy, nephropathy, and neuropathy.[45] (2) Three-carbon dicarbonyl products of glucose and lipid metabolism, glyoxal and methylglyoxal, also react readily with amino groups in proteins and produce other AGEs, one of which is argpyrimidine. AGEs react with specific cellular receptors and can stimulate numerous potentially dangerous processes.[43,44] (3) Hyperglycemia can also secondarily produce oxidative stress in tissues, with depletion of glutathione and formation of reactive oxygen species and damaging free radicals.[46] (4) When glucose is insufficiently metabolized by insulin-stimulated routes [*see Figure 1*], it can overflow into the sorbitol (polyol) pathway via the enzymes aldose reductase and sorbitol dehydrogenase.[47] Accumulation of sorbitol and fructose in vulnerable tissues such as nerves produces osmotic damage, loss of myoinositol essential to nerve membrane integrity, and reduction of Na^+, K^+–ATPase activity.[47] (5) Elevated glucose levels increase protein kinase C, an enzyme whose activity influences numerous cellular processes with damaging potential,[48] such as stimulating neovascularization and epithelial cell proliferation, increasing collagen synthesis, increasing vascular permeability, increasing apoptosis (programmed cell death), increasing oxidative stress, and mediating the actions of VEGF and transforming growth factor–β. (6) Elevated glucose levels also increase the production of VEGF, a molecule that stimulates angiogenesis. VEGF is present in high concentrations in human diabetic ocular tissues and in kidneys of animals with experimentally produced diabetes. It is a logical candidate to mediate development of proliferative retinopathy. (7) Hyperglycemia stimulates nitric oxide synthase to produce nitric oxide, a molecule that itself generates damaging free radicals.[46] (8) Excess blood glucose also overflows into the hexosamine pathway, resulting in deleterious products.[49] A single mitochondrial defect that leads to overproduction of reactive oxygen species can result in at least three of the above pathways and has been proposed as the primary culprit.[50] A number of these pathways are also mutually reinforcing, setting up vicious circles that can accelerate tissue damage.

The therapeutic importance of elucidating the mechanistic links between hyperglycemia and microvascular/neuropathic complications lies in our current inability to normalize blood glucose consistently. Therefore, drug therapies that intercept pathogenetic processes downstream from glucose hold promise for preventing these complications, even in the presence of hyperglycemia. An inhibitor of AGE formation, aminoguanidine, has been successful in animal experiments, but human trials

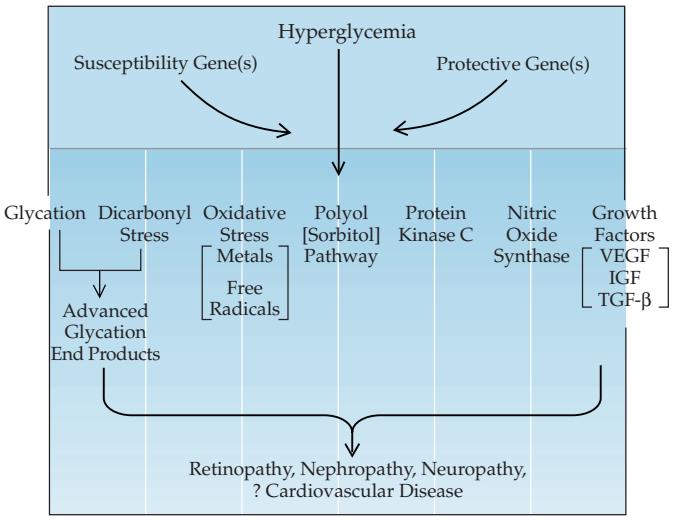


Figure 9 **Multiple pathways have been described that may link high blood glucose levels to the microvascular and neuropathic complications of diabetes (see text). There are good reasons to believe that genetic factors, possibly operating through such pathways, may explain the observation that some individuals with consistently high blood glucose levels do not experience complications, whereas other individuals with near-normal blood glucose levels do experience complications.**

have revealed unacceptable toxicity. Several inhibitors of aldose reductase, catalyzing the first step in the polyol pathway, have been studied in clinical trials, but none have shown sufficient clinical benefit or an acceptable adverse-effect profile to warrant approval in the United States. Nonetheless, such drugs have been effective in animal models. Current clinical trials are testing the effects of antioxidants such as vitamin E and a relatively nontoxic oral inhibitor of protein kinase C. Antagonists to VEGF and other growth factors to be administered by systemic or local injection are also in development.

GENETICS OF MICROVASCULAR COMPLICATIONS

There is considerable evidence from several studies that diabetic nephropathy clusters in families.[51] Thus, either genetic susceptibility or genetic protection is likely to explain the fact that nephropathy develops in only 35% to 40% of patients with diabetes. One likely influence on the development of nephropathy is the family of genes that code for the components of the renin-angiotensin system. Both positive and negative findings have been reported concerning involvement of the gene for angiotensin-converting enzyme (ACE) and the gene for angiotensinogen in the risks for nephropathy and retinopathy. A family study conducted in the DCCT showed no evidence for familial clustering of diabetic retinopathy per se. In view of the nearly 100% prevalence of retinopathy in patients with type 1 diabetes mellitus of many years' duration, it is not likely that a genetic factor is involved in the initiation of retinopathy. The DCCT analyses did, however, show evidence of familial clustering of severe diabetic retinopathy and confirmed familial clustering of nephropathy.[52]

Type 1 Diabetes Mellitus

PATHOGENESIS OF TYPE 1 DIABETES MELLITUS

Type 1 diabetes mellitus is characterized by absolute insulin deficiency, making patients dependent on exogenous insulin re-

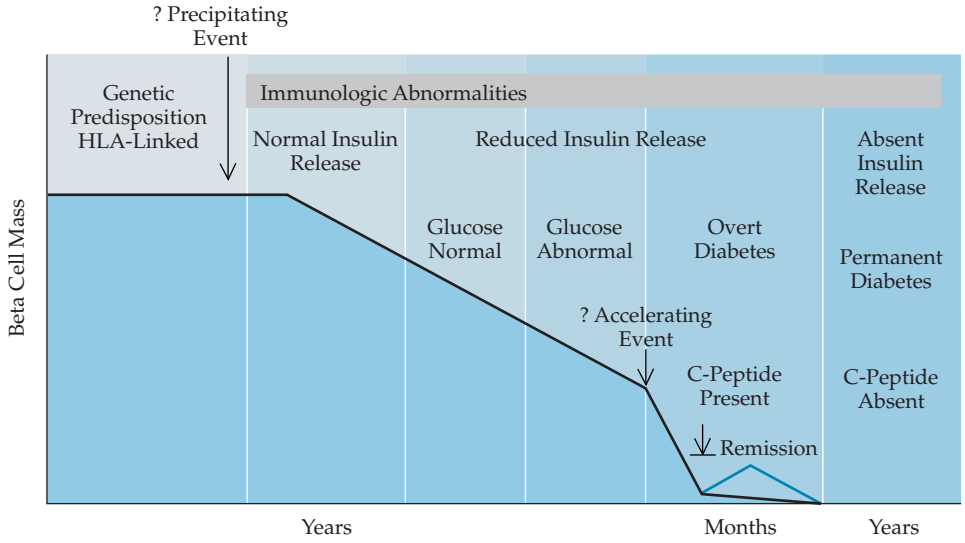


Figure 10 **Current view of the pathogenesis of type 1 autoimmune diabetes mellitus. In some individuals, HLA-linked genes set in motion an autoimmune attack on islet cells, predominantly beta cells. In other individuals, HLA-linked genes protect against the autoimmune destructive response. An initiating event, such as exposure to a virus with an antigenic epitope that resembles a beta cell antigen or to a toxin, may start the process of self-destruction. Disappearance of the beta cells may occur because the viral antigen accelerates the normal rate of apoptosis (programmed cell death). As time passes, insulin production and secretion diminish, despite increasing hyperglycemia. When insulin release falls to trivial amounts or none, diabetic ketoacidosis results. Another external event may trigger this final beta cell catastrophe. A few beta cells may survive, because after this, a brief period of remission marked by reappearance of C-peptide in plasma may ensue if plasma glucose levels are controlled very tightly with exogenous insulin. Eventually, all beta cell function ceases, leading to metabolic instability.**

placement for survival.[53] Insulin deficiency results from destruction or disappearance of the insulin-producing beta cells[54] that constitute 80% of the pancreatic islets of Langerhans. When 90% of the beta cells have been eliminated, clinical diabetes occurs [*see Figure 10*].

Autoimmune Factors

There is strong evidence for a cell-mediated autoimmune process being involved in the destruction of beta cells in the majority of cases of type 1 diabetes mellitus.[55-57] In a number of cases in which death occurred from an accident or from an illness other than diabetes shortly after diagnosis of type 1 diabetes mellitus, a mononuclear lymphocytic infiltrate was found in the islets. In this form of insulinitis, T cell distribution shows an increase in CD8 suppressor-inducer T cells and a decrease in CD4 helper-inducer T cells.[55] A similar immunocellular response has been found in animal models of spontaneous insulin-deficient diabetes.[57] In some instances, experimental manipulations that prevent T cell lymphocytic responses also prevent the development of diabetes. Furthermore, transfer of diabetes from affected animals to nonaffected animals by lymphocytes has also been described. Interleukins and other cytokines have been shown to exhibit toxic effects on the beta cells and to inhibit insulin secretion.

Autoantibodies to a variety of beta cell and islet autoantigens are present in the sera of patients with type 1 diabetes mellitus at the time of diagnosis.[58] The autoantigens include the enzymes glutamic acid decarboxylase (GAD), carboxypeptidase H, a protein tyrosine phosphatase labeled ICA512 or IA-2, and insulin itself.[58-60] Some, but not all, studies have shown that islet autoantibodies are capable of inhibiting insulin secretion in vitro or even causing lysis of beta cells. Other evidence supports the importance of autoimmune phenomena in the pathogenesis of type 1 diabetes mellitus. In cases of transplantation of pancreases from nondiabetic identical twins to patients with type 1 diabetes mellitus who were not given immunosuppressive therapy, the pancreas was rejected by the diabetic host's immune system, which apparently recognized as self, identical antigens in the normal twin's pancreatic islets. If treatment of type 1 diabetes mellitus with the immunosuppressive agent cyclosporine is initiated within 2 to 6 weeks after clinical onset of diabetes, dependency upon insulin can be eliminated or insulin doses markedly reduced, but only as long as immunosuppression is maintained.[61,62] The toxicity associated with cyclosporine and other immunosuppressive agents has precluded use of this form of therapy in clinical practice.

It is now clear that the autoimmune phenomena begin long before clinical onset of the disease. Islet or beta cell autoantibodies can be found in 2% to 4% of first-degree relatives of patients with type 1 diabetes mellitus, which is 10 to 20 times the prevalence of control subjects. Longitudinal studies have shown that type 1 diabetes mellitus is much more likely to develop in clinically unaffected relatives with high autoantibody titers than in relatives without such antibodies, and that the disease will develop in such patients within a few years.[63-65] Longitudinal serial testing of plasma insulin responses to intravenous glucose injection demonstrates progressively declining beta cell function in autoantibody-positive relatives before the clinical onset of diabetes.[66]

Environmental Factors

Because only 30% to 50% of unaffected monozygotic identical twins of patients with type 1 diabetes melllitus will eventually develop the disease, it is likely that an environmental factor may be required to trigger the autoimmune destructive process.[67] A

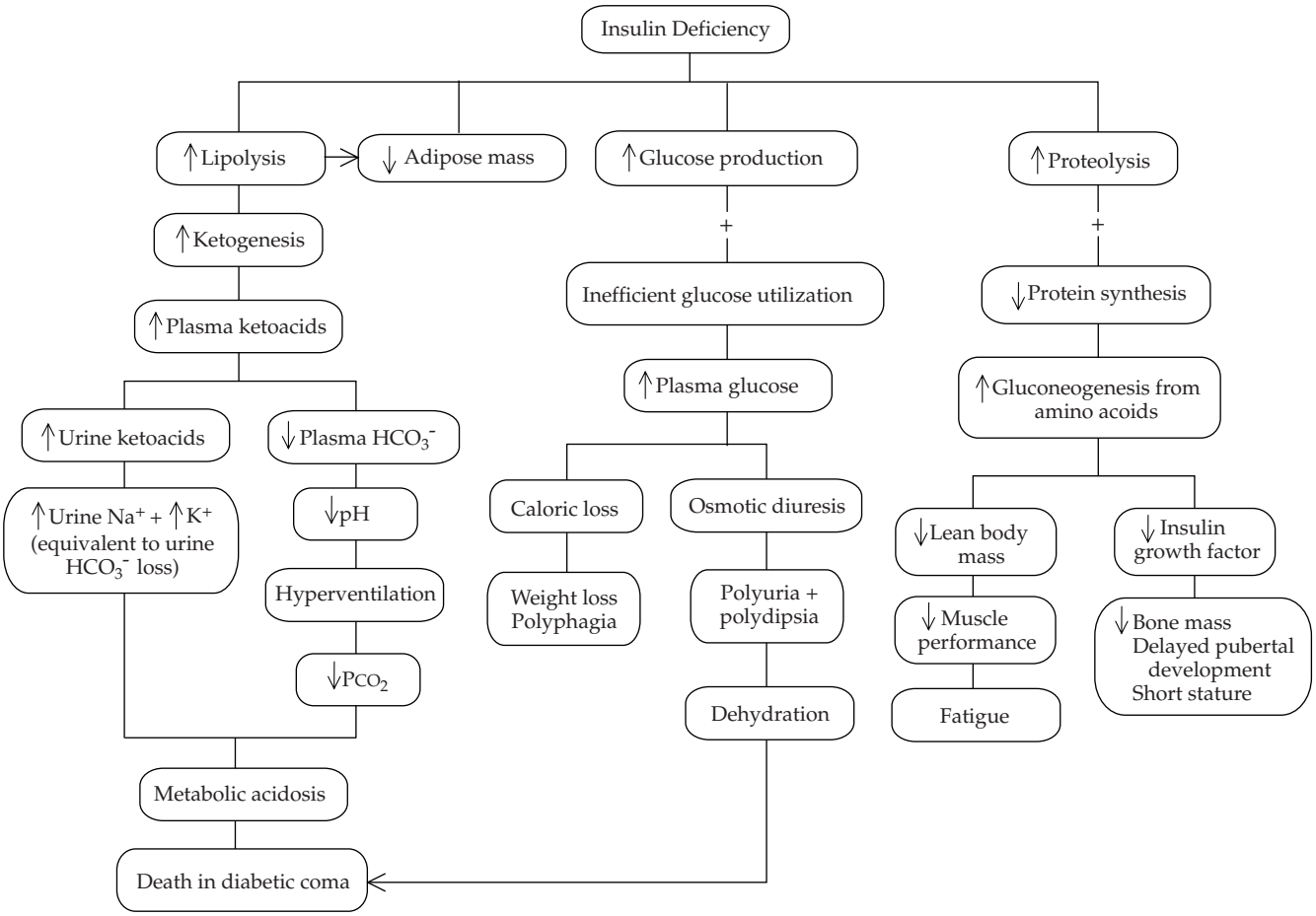


Figure 11 **Shown are the pathways that lead from insulin deficiency to the major clinical manifestations of type 1 diabetes mellitus. Note that a decrease in insulin growth factor also results from insulin deficiency and decreases growth rate.**

number of viral candidates have been proposed.[67] The only certain association is that offspring of women who are infected with rubella during pregnancy are at increased risk for type 1 diabetes mellitus. A small amount of indirect evidence also associates coxsackievirus B with type 1 diabetes mellitus.[68] Toxins in the environment or diet might also initiate the destruction of genetically vulnerable beta cells.

Temporal Sequence of Beta Cell Destruction

At the time of clinical onset of type 1 diabetes mellitus, at least a small number of beta cells are still potentially capable of function.[69,70] After several weeks of exogenous insulin treatment, particularly if exemplary metabolic control has been established,[71] dependency on exogenous insulin decreases or ceases entirely for weeks to months in some patients. This temporary so-called honeymoon remission phase is marked by an increase in serum C-peptide levels, which indicates an increase in endogenous insulin secretion [*see Figure 10*].[70] However, within 5 years after diagnosis of childhood type 1 diabetes mellitus, C-peptide virtually disappears from the serum.[72]

Type 1 diabetes mellitus does not develop in all autoantibody-positive individuals. Moreover, the latency period between initiation of beta cell destruction and appearance of the clinical disorder may be many years,[73] as the disease does not appear in some patients until considerably later in life. The gradual, indolent nature of the disease in these autoantibody-positive individuals is also suggested by the fact that some can be treated with beta cell–stimulating drugs before absolutely requiring insulin.[74] A trial sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases unfortunately found no evidence that type 1 diabetes mellitus can be prevented by inducing immune tolerance to exogenous human insulin given subcutaneously or orally to relatives of patients with high islet autoantibody titers.

Genetic Factors

Although a family history of type 1 diabetes mellitus is more likely to be absent than present in index cases, it is nonetheless true that offspring and siblings of patients with type 1 diabetes mellitus are at increased risk for the disease. There is a genetic basis for susceptibility to type 1 diabetes mellitus but not for inevitable development of the disease.[75] The disease will develop in 5% to 10% of first-degree relatives of patients with type 1 diabetes mellitus and in 20% of persons who have two first-degree relatives (e.g., both parents) with the disease. Association and linkage studies have incriminated a number of genes involved in the risk of type 1 diabetes mellitus. Polymorphism of HLA genes in the MHC locus on chromosome 6 account for 50% of the genetic risk.[75] DR3 and DR4 are susceptibility alleles that appear to operate synergistically. Individuals heterozygous for DR3 and DR4 are at greater risk than either homozygous DR3 or homozygous DR4 individuals. The DR2 allele decreases the risk and dominates the susceptibility effect of DR3 or DR4 when either is accompanied by DR2. The HLA-DQ locus also is associated with increased risk of diabetes.[76] Substitution of alanine, valine, or ser-

ine for the more usual aspartic acid at position 57 of DQ β chain or the presence of arginine at position 52 of DQ α chain increases the risk of type 1 diabetes mellitus. A number of mechanisms have been suggested to explain how HLA class II molecules might predispose to or protect against the disease.[77] Despite the accumulation of considerable knowledge, type 1 diabetes mellitus still cannot be predicted with complete certainty.[78]

Type 1 diabetes mellitus is associated with at least 15 additional loci on nine other chromosomes.[78] Of particular interest is that a variable number of tandem repeats in the promoter region of the insulin gene has been associated with the disease. However, the insulin molecule itself is apparently normal in structure in patients with type 1 diabetes mellitus. With the human genome soon to be fully known and advanced genetic technology becoming cost-effective, it is likely that the genetic components of type 1 diabetes mellitus will be sorted out in a way that will make it possible to identify susceptible individuals who might benefit from preventive therapies.

The clinical and biochemical manifestations of type 1 diabetes mellitus can all be accounted for as consequences of insulin deficiency [*see Figures 1 and 11*].[79] Loss of the stimulating effect of insulin on glucose uptake by muscle and adipose tissue coupled with loss of the suppressive effect of insulin on hepatic glucose output lead to severe hyperglycemia. FPG rises typically to 300 to 400 mg/dl, and postprandial glucose levels rise to 500 to 600 mg/dl in patients before treatment.[79] This increase presents a high filtered load of glucose to the renal tubules, causing a severe osmotic diuresis, manifested by polyuria and compensatory polydipsia. Loss of the lipogenic and antilipolytic effects of insulin on adipose tissue leads to high plasma levels and increased hepatic uptake of free fatty acids. This condition enhances ketogenesis, and ultimately, high plasma ketoacid levels cause metabolic acidosis. Protein breakdown is favored in the absence of the anticatabolic and anabolic actions of insulin. The proteolysis of muscle protein provides amino acids that sustain high rates of gluconeogenesis. Body-weight loss thus includes fat and lean body mass, and it is further aggravated by an increase in basal energy expenditure.[80] The negative nitrogen balance, accompanied by losses of potassium, magnesium, and phosphate in the urine, impairs growth and development in children.

DIAGNOSIS OF TYPE 1 DIABETES MELLITUS

The diagnosis of type 1 diabetes mellitus is still almost always made on the basis of symptom history confirmed by a blood or plasma glucose level greater than 200 mg/dl, with the presence of glucosuria and often ketonuria. The classic symptoms are polyuria, polydipsia, weight loss with normal or even increased food intake, fatigue, and blurred vision, commonly present 4 to 12 weeks before the symptoms are noticed. In the future, however, before clinical onset of type 1 diabetes mellitus, diagnosis may be possible with serologic methods, complemented by beta cell function tests.

MANAGEMENT OF TYPE 1 DIABETES MELLITUS

Of all chronic diseases, diabetes is unique because its therapy involves daily self-management by the patient and a host of lifestyle adaptations. For optimal metabolic control, patients must prick their fingers to test blood glucose at least four times daily, inject insulin at least three times daily, pay regular attention to the timing and content of their meals, and try to follow a scheduled exercise program. The patient is truly at the center of his or her care. Patient self-management requires intensive education with regard to the skills of injection and blood glucose monitoring, urine ketone testing on sick days, meal planning, detection and treatment of hypoglycemia, and management of intercurrent illness. Family members and close associates of the patient need to be included as is appropriate, particularly with regard to recognition and treatment of hypoglycemia. Ideally, the patient should understand the pathophysiology of diabetes and its long-term complications almost as well as health care professionals. Some aspects of care require periodic educational reinforcement, which is often stimulated by some therapeutic mishap, such as a preventable episode of severe hypoglycemia.

The clinical goals of treatment include (1) decreasing plasma glucose levels and urine glucose excretion to eliminate polyuria, polydipsia, polyphagia, caloric loss, and adverse effects such as blurred vision from lens swelling and susceptibility to infection, particularly vaginitis in women, (2) abolishing ketosis, (3) inducing positive nitrogen balance to restore lean body mass and physical capability and to maintain normal growth, development, and life functioning, (4) preventing or greatly minimizing the late complications of diabetes previously discussed. After publication of the DCCT results, The American Diabetes Association revised their standards of care accordingly [*see Table 3*] to include firm biochemical goals[81]: (1) maintaining preprandial capillary whole blood glucose levels at 80 to 120 mg/dl, bedtime blood glucose levels at 100 to 140 mg/dl, and postprandial peak blood glucose levels at less than 180 mg/dl, and (2) maintaining

Table 3 American Diabetes Association Standards* for Glycemic Control in Diabetes Mellitus[288]

Biochemical Index	*Normal*	*Goal*	*Additional Action Suggested*
Capillary whole blood values† (mg/dl)			
Average preprandial glucose level	< 110	80–120	< 80 > 140
Average bedtime glucose level	< 120	100–140	< 100 < 160
HbA_{1c} (%)	< 6	< 7	> 8

*The values shown in this table are by necessity generalized to the entire population of individuals with diabetes. Patients with comorbid diseases, the very young, older adults, and patients with unusual conditions or circumstances may warrant different treatment goals. These values are for nonpregnant adults. Additional action suggested depends on individual patient circumstances. Such actions may include enhanced diabetes self-management education, comanagement with a diabetes team, referral to an endocrinologist, change in pharmacologic therapy, initiation of or increase in self-monitored blood glucose testing, or more frequent contact with the patient. HbA_{1c} is referenced to a nondiabetic range of 4.0% to 6.0% (mean, 5.0%; SD, 0.5%).
†To convert to plasma glucose values, add 10 mg/dl to whole blood values, except for 160 mg/dl, which becomes 180 mg/dl.

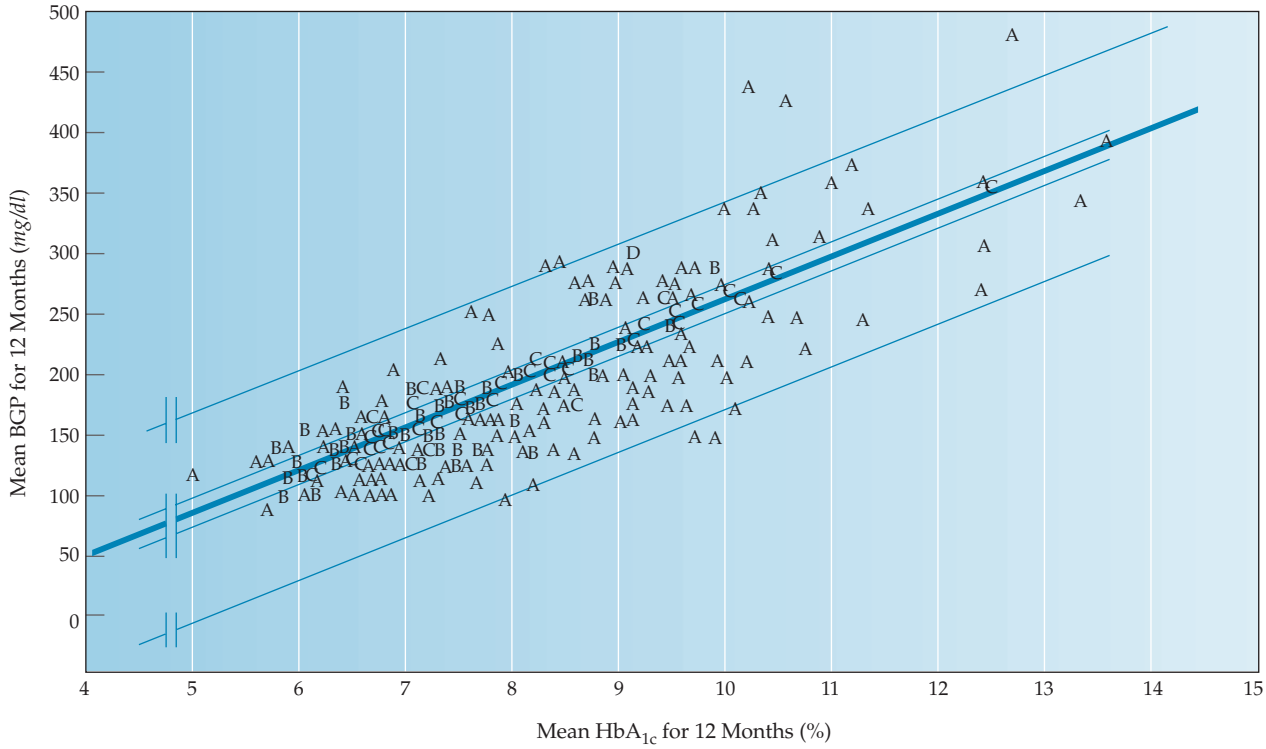


Figure 12 **The 12-month mean value of all seven-sample-a-day blood glucose profile values measured quarterly in the Diabetes Control and Complications Trial central biochemistry laboratory is plotted against the 12-month mean of quarterly HbA_{1c} values in the same patients. (A = 1 point, B = 2 points, C = 3 or more points; r = 0.80, $P < 0.001$)**

an HbA_{1c} of less than 7.0% (relative to a nondiabetic DCCT range of approximately 4.0% to 6.0%). Realistically, current therapeutic tools make it difficult to achieve these stringent goals in many patients with type 1 diabetes mellitus, particularly those with absolutely no endogenous insulin secretion. The exponential relation between the risk of microvascular complications and HbA_{1c} predicts that only normal HbA_{1c} levels would completely prevent the complications. However, maintaining an HbA_{1c} below 7.0% will remove much of the absolute risk from most patients. Efforts to achieve an HbA_{1c} of less than 7.0% should continue as long as hypoglycemia can be minimized.

Monitoring of Glycemic Control

In the past 5 to 10 years, blood glucose meters have undergone continuous development and improvement. They are now smaller, use less blood and more sites for puncture, are less vulnerable to inaccuracy because of patient errors, and have memory programs that allow the patient or caregiver to assess the pattern of blood glucose control over the previous 2 months, largely eliminating the problem of incorrect or fabricated written transcription of results. Devices that can accurately estimate blood glucose without a blood sample have been in development for a number of years but have not reached a state of reliability suitable for clinical practice. An indwelling subcutaneous catheter for blood glucose monitoring that can be used for 3 days and provide frequent readings is now available. Although the recorded profile can provide only a brief window into a lifetime of blood glucose fluctuation, such a profile can guide periodic adjustments of the regimen.

Currently, even with its imperfections, blood glucose testing before each meal or large snack is essential if the patient is to adjust each dose of rapid-acting insulin to the level of blood glucose before the meal and to the amount of carbohydrate about to be ingested. Blood glucose levels also need to be periodically checked after meals to ensure that undue postprandial hyperglycemia is not occurring. Patients should also check blood glucose levels before or after intensive exercise to prevent or abort hypoglycemia. It is very important to check blood glucose levels before driving to prevent motor vehicle accidents brought on by severe hypoglycemia, which can have adverse effects on drivers' judgment and reaction times. Occasional 3:00 A.M. blood glucose readings are useful in monitoring for otherwise unrecognized frequent nocturnal hypoglycemia. Most important, during intercurrent illnesses, especially those accompanied by nausea, vomiting, and limitation of fluid and caloric intake, patients must test blood frequently to guide insulin treatment. In addition, under these circumstances, the risk of ketoacidosis mandates testing of urine or blood for ketoacids. The presence of significant levels of ketoacids is a signal to call the caregiver immediately and establish frequent contact for instructions regarding insulin doses and carbohydrate intake.

A critical supplement to home blood glucose testing is monitoring of HbA_{1c} in the physician's office. It is now well established that this product of nonenzymatic glycation provides an excellent index of average blood glucose levels [*see Figure 12*] for approximately the preceding 2 months.[82,83] In at least one study, patients whose HbA_{1c} was measured periodically had a better health status, lower glycemic levels, and fewer hospitalizations than a randomly selected group of patients whose HbA_{1c} level remained unknown to both the patient and the physician.[84] Quarterly HbA_{1c} measurements are satisfactory except during pregnancy, when monthly levels should be obtained. Because methods and results vary among laboratories, a national glycohemoglobin standardization program is under way, and HbA_{1c} should be measured in laboratories certified to provide DCCT-equivalent results.[85] Use of rapid-turnaround, point-of-service

HbA_{1c} assays improves the efficiency with which diabetes caregivers can modify patients' regimens on office visits and improves treatment results.[86] Assays of other products of nonenzymatic glycation, such as fructosamine and glycated albumin, that reflect shorter periods of chronic glycemia are less useful in routine diabetes management.[85]

Insulin Types and Delivery

Correction of insulin deficiency is the most critical component in managing type 1 diabetes mellitus. Before the availability of insulin, patients with type 1 diabetes mellitus and complete insulin deficiency inevitably followed a predictable downhill course [*see Figure 11*] and died either in diabetic coma or essentially of starvation and inanition. Insulin extracted from beef and pork pancreas and purified to increasingly high levels was the mainstay of therapy until recombinant DNA technology made it possible to produce authentic human insulin in large quantities. Although animal insulins are therapeutically bioequivalent to human insulin, they disappeared from the market as manufacturers switched over to making only human insulin. In rare instances of local allergy to human insulin, lispro insulin (see below) can be substituted. In emergency situations, patients with systemic allergy to human insulin can be desensitized by administering extremely small amounts and gradually increasing the dose over 6 to 24 hours until the patient is tolerant and responsive to human insulin.

The basic principle of insulin replacement[87,88] is to provide a slow, long acting, continuous supply that mimics the nighttime and interprandial basal secretion by normal beta cells. In addition, a rapid and relatively short-acting form of insulin delivered before meals mimics the normal meal-stimulated burst of insulin secretion [*see Figure 2*]. A number of insulin preparations for subcutaneous administration are currently available [*see Table 4*]. It is important to recognize that there is considerable variability in the pharmacokinetic characteristics of these insulins both from individual to individual and within the same individual from day to day. Rates of insulin absorption from the skin vary with the injection site, the depth and angle of injection, ambient temperature, and exercise of an injected limb. Injection into the subcutaneous tissue of the abdomen produces the least variable results. The expected therapeutic action can also be affected by fluctuations in sensitivity to insulin from time to time in patients. Despite the variability of results, certain average patterns can be expected from the multiple daily injection regimens in common use [*see Figure 13*]. CSII by use of an external pump provides smooth basal delivery and somewhat more predictable acute increases in plasma insulin for meals. Only crystalline zinc insulin (regular insulin) and lispro insulin are used in such pumps, which is one reason for their greater consistency of effect.

Synthetic Insulin Analogues

Lispro was the first of what undoubtedly will be many new insulin analogues with structures designed to provide pharmacokinetics that more closely mimic physiologic insulin secretion and needs.[89] One of the features of natural (or synthetic) human insulin is that six molecules associate with a zinc molecule to form hexamers. Insulin hexamers must disassociate to monomers before they can be absorbed from subcutaneous injection sites. This requirement is the main reason that crystalline zinc insulin (regular insulin) has a peak action 2 to 4 hours after injection and must be taken 30 to 60 minutes before eating to have any chance of limiting postprandial hyperglycemia. By simply exchanging lysine and proline at positions 28 and 29 of the B chain of insulin [*see Figure 3*], hexamer formation is prevented and the monomer is rapidly absorbed from an injection site. Lispro insulin action begins within 15 minutes, the peak effect is reached at 1 to 2 hours, and the duration of action is only 4 to 6 hours. Thus, lispro insulin injected just before a meal provides a postprandial plasma insulin profile similar to that of normal human insulin secretion [*see Figure 2*]. The chief benefits of using lispro insulin are to reduce postprandial blood glucose peaks and to somewhat decrease the hypoglycemia that can result from the late tail of regular insulin action.[90,91] However, loss of that late action can lead to recurrent hyperglycemia before the next meal. Hence patients switched from regular insulin to lispro insulin may have no reduction in HbA_{1c} unless their doses of basal insulin (neutral protamine Hagedorn [NPH], Lente, or Ultralente or the basal rate in CSII) are increased.[92] It may even prove useful to combine lispro insulin with regular insulin in a single injection to optimize postprandial control.

Another synthetic rapid-acting analogue, insulin aspart, replaces proline with aspartic acid at position B28 [*see Figure 3*]. This substitution leads to a profile of action and therapeutic benefits that are very similar to those of lispro insulin.[93] A long-acting analogue, glargine, has also been synthesized as a basal insulin with no discernible peak and a longer duration of action than Ultralente insulin.[94,95] Glargine has two additional arginines at the carboxyl terminus of the B chain, B31 and 32, and has a glycine for arginine substitution at position A21 [*see Figure 3*]. Glargine is giv-

Table 4 Insulin Preparations

Insulin Type	*Onset (hr)*	*Duration (hr)*	*Peak (hr)*
Rapid acting (regular, crystalline zinc insulin [CZI])	0.5–1.0	6–8	2–3
Very rapid acting			
Lispro	0.25–0.5	4–6	1–2
Insulin aspart	0.25–0.5	4–6	1–2
Intermediate acting			
Lente, neutral protamine Hagedorn (NPH)	1	10–14	4–8
Long acting			
Ultralente	1	18–24	Minimal at 10–14
Glargine	1.5	30	None

en as a single bedtime injection to provide basal insulin for 24 hours with less nocturnal hypoglycemia.[96] For reasons that should now be clear, intensive treatment regimens are the preferred form of therapy and should be implemented early in as many patients as is safely possible. Different combinations of insulin preparations can be used to approximate (but never reliably reproduce) normal plasma insulin profiles [*see Figure 13*]. Type 1 diabetes mellitus can almost never be satisfactorily controlled on less than two injections a day of intermediate- or long-acting insulin combined with rapid-acting insulin. Only in patients experiencing a honeymoon remission or in patients with late-onset autoimmune type 1 diabetes mellitus in adults can satisfactory metabolic control be established with a single injection of insulin daily. Such success is made possible only by the presence of some normally regulated endogenous insulin secretion.

Insulin Regimens

As a rule of thumb, basal insulin and mealtime insulin pulses each constitute approximately 50% of the average total daily dose (0.6 to 0.7 U/kg) in intensive-therapy regimens. The dose of regular or lispro insulin or insulin aspart before each meal is chosen by the patient on the basis of the blood glucose level, the estimated amount of carbohydrate to be eaten, or both. A typical regimen would call for 1 to 2 extra units of insulin for each 50 mg/dl increment in blood glucose above the dose called for by the preprandial target of 80 to 120 mg/dl, or 1 U/10 to 15 g of extra carbohydrate to be ingested above the usual amount of carbohydrate prescribed by the nutrition plan. Very sophisticated patients can combine both guidelines. Glargine is rapidly achieving dominance as the basal insulin.

Fixed-dose mixtures of insulin are not physiologically very suitable for patients with type 1 diabetes mellitus. However, for patients who can or will implement only such conventional treatment, a typical regimen might be a total daily dose of 0.6 to 0.7 U/kg. Two thirds to three fourths of the dose would be given before breakfast and the remainder before supper; the ratio of intermediate-acting insulin to rapid-acting insulin might be 2:1 to 4:1 before breakfast and 1:1 before supper. Because giving NPH or Lente insulin before supper increases the risk of hypoglycemia between 2:00 and 4:00 A.M., patients on conventional treatment should be urged to switch to a three-injection regimen, taking the evening dose of intermediate-acting insulin at bedtime to avoid nocturnal hypoglycemia and to better control the prebreakfast blood glucose level. Glargine insulin may also be helpful in minimizing nocturnal hypoglycemia.[87,88]

Insulin requirements are increased by greater caloric and especially carbohydrate intake, by weight gain of both lean body mass and fat mass, by the onset of puberty, by infections and other medical or surgical stresses, by pregnancy, by glucocorticoid administration, and sometimes by the physiologic changes that pre-

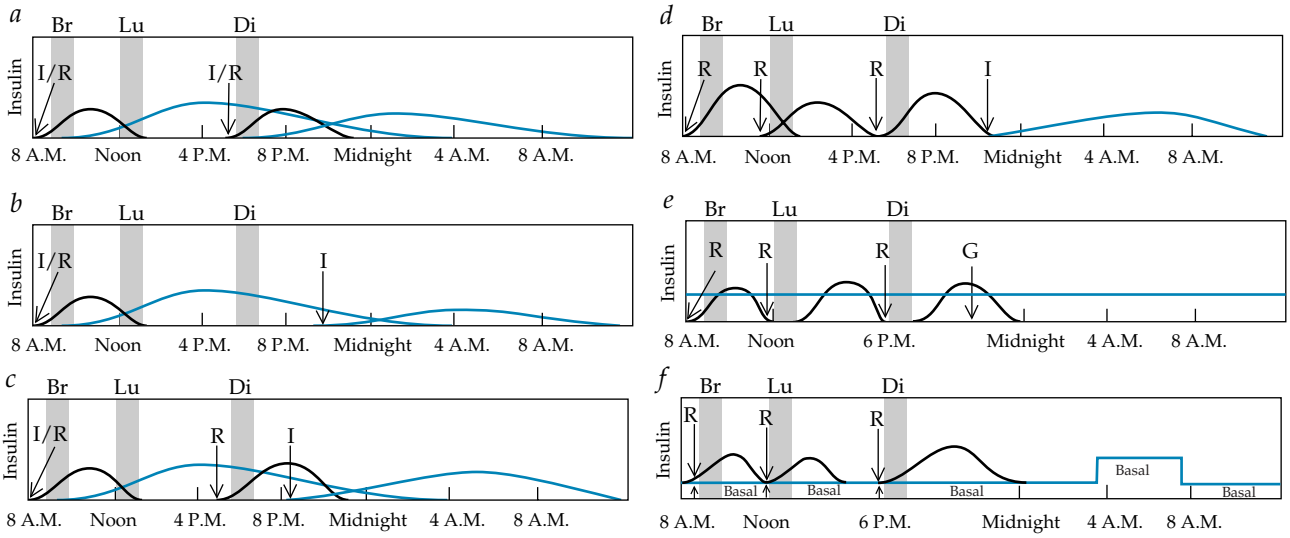


Figure 13 **Different combinations of various insulin preparations can be employed in establishing glycemic control in type 1 diabetes mellitus (and in those patients with type 2 diabetes mellitus who eventually reach an equivalent degree of insulin deficiency). Arrows indicate time of injection. Red curves represent rapid-acting (R) regular or lispro insulin. Blue curves represent intermediate-acting (I) NPH or Lente insulin. Gray curves represent long-acting insulin glargine (G). (*a*) A mixed injection of I and R insulin is administered before breakfast and dinner in this average regimen. In addition to the risk of hypoglycemia before lunch and in the late afternoon, the predinner administration of I insulin predisposes patients to hypoglycemia from 2:00 A.M. to 4:00 A.M. (*b*) This average regimen combines a mixed injection of I and R insulin given before breakfast with an injection of I insulin given before bed. The I insulin administered at bedtime provides safer, more effective overnight glucose control; without predinner insulin, however, glucose levels may rise to unacceptably high levels after dinner. (*c*) In this intensive regimen, the patient receives three injections: a mixed injection before breakfast, R insulin before dinner, and I insulin before bed. (*d*) This intensive regimen combines three preprandial injections of R insulin with one injection of I insulin before bed. Preprandial doses of R insulin are adjusted according to glucose levels and meal size. (*e*) This intensive regimen uses long-acting insulin glargine to replace basal insulin secretion. Preprandial doses of R insulin are adjusted according to blood glucose levels and anticipated meal carbohydrate content. (*f*) This intensive regimen provides only R insulin as regular or lispro insulin. A pump-driven continuous subcutaneous infusion of R insulin replaces basal insulin secretion. Basal rates can vary during different times of day or activities. For example, the basal rate can be lowered or even suspended during periods of intensive aerobic exercise. The nocturnal basal rate can be increased 1.5 to 2.0 times from 3:00 A.M. to 4:00 A.M. until breakfast to accommodate the rising early morning insulin requirement known as the dawn phenomenon. Preprandial bolus doses are individually dialed in and rapidly pumped in, adjusted according to blood glucose levels and anticipated meal carbohydrate content. (Br = breakfast, Lu = lunch, Di = dinner)**

cede the onset of menses. During acute illnesses, patients will require extra doses of rapid-acting insulin when hyperglycemia accelerates and especially if ketosis occurs. Frequent telephone contact with caregivers allows timely professional guidance of the extra insulin doses, nutrient intake to prevent hypoglycemia, and fluid intake to prevent dehydration. Lispro insulin or insulin aspart is especially useful in these circumstances because the effect of an overdose is short lived and hypoglycemia is less likely.

CSII has improved considerably since its introduction in the 1970s.[97] Modern insulin infusion pumps permit programming with multiple basal rates, allowing flexibility during the day as well as automatic adjustment of doses while sleeping at night. Frequently, the basal rate needs to be lower in the first half of the night and then increased to accommodate the so-called dawn phenomenon [*see Figure 13*]. The latter is a slow rise in the plasma glucose level before the patient awakens, demonstrable in normal individuals but exaggerated in individuals with type 1 diabetes mellitus who cannot limit it by increasing endogenous insulin secretion. On the other hand, interruption of insulin delivery from a pump for as little as 8 hours can result in extreme hyperglycemia, diabetic ketoacidosis (DKA), and hyperkalemia. In the DCCT, patients who used an insulin pump had a slightly but significantly higher DKA event rate (1.8 per 100 patient-years) than patients on multiple daily injection regimens (0.8 per 100 patient-years).[98] There was no difference in risk of severe hypoglycemia between patients treated with insulin pumps and patients treated with multiple daily injections, although episodes resulting in coma or seizure were more common in CSII-treated patients.[98] The rate of infection at catheter sites was kept very low by frequent change of catheters and preemptive use of antibiotics at the first visible signs of infection. Pump use has grown exponentially in the past 10 years; 200,000 patients now use pumps.

Avant-garde Therapy

Implantable pumps delivering insulin into the peritoneal cavity and resulting in a more physiologic first pass of insulin through the liver have provided acceptable HbA_{1c} levels with a lower frequency of severe hypoglycemia.[99] Difficulties with obstruction of insulin delivery and infection have occurred, and they are not yet approved for commercial use. Closed-loop insulin-delivery devices that would measure the patient's blood glucose level very frequently and would automatically adjust insulin delivery still await the development of a practical and long-lived indwelling continuous glucose sensor.

Insulin can be absorbed through the mucosa of the nose and also through the lungs. A nasal preparation of insulin has been effective in short-term clinical trials, but the disadvantages of high cost (10 times the subcutaneous insulin dose is needed to achieve the same blood glucose lowering) and failure to develop a vehicle that does not cause allergic nasal symptoms have prevented this preparation from being used in practice.[100] Inhaled insulin, with various delivery devices, is still undergoing clinical trials.[101,102] The pharmacokinetic properties of inhaled insulin resemble those of lispro insulin and insulin aspart, so inhaled insulin is suitable for preprandial use. Various attempts to package insulin for oral administration so as to prevent its degradation in the gastrointestinal tract have also been investigated, as has transdermal insulin.

Pancreas transplantation remains controversial as a routine form of insulin replacement therapy.[103] Over the period of 1994 to 1997, 1-year graft survival rates were 82% when a pancreas was transplanted with a needed kidney transplant and 62% when a pancreas was transplanted alone.[104] Successful pancreas transplants provide nondiabetic HbA_{1c} levels and free the patient from the rigors of diet, blood glucose testing, and insulin injection, and they virtually eliminate episodes of hypoglycemia.[104] Quality of life is usually improved. On the negative side, the patient incurs the risk of operative mortality and morbidity and must remain on immunosuppressive therapy with its attendant risks of infection and malignant disease.[103] Length of stay, readmission rates, morbidity, and the number of acute rejection episodes are higher for pancreas transplants than for kidney transplants. From 1994 to 1996, the 1-year pancreas transplant survival was 81%, compared with a kidney transplant survival of 88%.[103] The large majority of pancreas transplantations are still performed as an option in conjunction with a necessary kidney transplant.

Transplantation of isolated islets can be accomplished without major surgery. Furthermore, the ability to immunomodulate isolated islets in the laboratory (by masking or removing cell surface antigens) may someday allow transplantation with little or no immunosuppression. Alternatively, islets can be placed in semipermeable hollow tubes that allow glucose to enter and insulin to leave but shield the islets from inflammatory reactions to a foreign body. Islet transplantation with function lasting at least 1 year has been achieved in less than 10% of attempts worldwide. A Canadian group has reported on seven successive cases of islet injection into the liver, with persistent function and independence from insulin injections for up to 15 months, using a new immunosuppressive regimen.[105] This technique is undergoing a multicenter trial.

Nutritional Therapy and Exercise

Intensive and conventional insulin treatment will produce unsatisfactory results unless it is appropriate for the nutrient intake. To facilitate the matching of insulin doses to meals and to prevent hypoglycemia, patients with type 1 diabetes mellitus should eat consistent regular meals comprising about 50% carbohydrate calories, less than 30% total fat calories, and less than 300 mg cholesterol a day.[106] Various methods of teaching patients how to assess amounts of foods and their nutrient and caloric content have been utilized. These methods include exchange lists that place foods into six categories; each category has approximately the same quantity of carbohydrate, protein, and calories per serving. These exchange categories are bread, meat, milk, fruit, fat, and vegetable. Another approach is to focus only on the carbohydrate content of foods because carbohydrates cause most of the postprandial hyperglycemia. Because different carbohydrates are digested and absorbed at different rates and therefore have different effects on plasma glucose levels, glycemic indices have been developed for common foods that help adjust for their different effects.[107] It is noteworthy that numerous studies have disproved the myth that sucrose raises blood glucose more than equivalent amounts of other carbohydrates.[108] For optimal instruction and reinforcement of diet therapy, a dietitian should be part of the diabetes care team.

Exercise is another important component of diabetes care because it helps maintain cardiovascular conditioning, insulin sensitivity, and general well-being.[109] However, patients must be instructed how to adjust their meals, their insulin doses and timing, or both to prevent hypoglycemia during, immediately after, or even 6 to 12 hours after exercise as muscle glycogen stores are replenished from plasma glucose. High-impact sports are contraindicated for patients with advanced retinopathy who are at risk for vitreous hemorrhage or for patients with peripheral neuropathy or vascular disease who are at risk for foot trauma, because such sports can be hazardous.

Table 5 Typical Laboratory Findings and Monitoring in Diabetic Ketoacidosis

Test	*Average*	*Range*
Plasma glucose	600 mg/dl (33 mmol/L)	200–2,000 mg/dl (11– 110 mmol/L)
Plasma ketones (positive)	1:16	1:2–1:64
Blood betahydroxybutyrate (mmol/L)	—	3–25
Plasma HCO_3 (mEq/L)	10	4–15
Blood pH	7.15	6.80–7.30
Pco_2 (mm Hg)	20	14–30
Plasma anion gap (Na^+ – [CI + HCO_3]) (mEq/L)	23	16–30

Perform complete blood count, serum urea nitrogen measurement, serum creatinine measurement, urinalysis, appropriate cultures, and chest radiography.

1. Weigh on admission and every 12 hr.
2. Record cumulatively intake and output every 1 to 2 hr (Foley catheter if incontinent).
3. Check blood pressure, pulse, respiration, mental status every 1 to 2 hr and temperature every 8 hr.
4. Check blood (fingerstick) or plasma (laboratory) glucose every 1 to 2 hr.
5. Check serum potassium every 2 to 4 hr; check other electrolytes and serum ketones or betahydroxybutyrate every 4 hr.
6. Check arterial blood pH and gases on admission (in children, venous pH may be substituted; add 0.1 to result). If pH < 7.0 on admission, recheck as required until pH exceeds 7.1.
7. Check serum phosphate, magnesium, and calcium levels on admission. If low, repeat every 4 hr; otherwise, every 8 to 12 hr.
8. Spot-check voidings for ketones and glucose.
9. Perform ECG on admission; repeat if follow-up serum potassium level is abnormal or unavailable.

Note: 1–9 should be carried out until the patient is stable, glucose levels have reached and are maintained at 250 mg/dl, and acidosis is largely reversed (plasma HCO_3 > 15–18, plasma anion gap < 16). An intensive care setting is preferred.

DIABETIC EMERGENCIES IN TYPE 1 DIABETES MELLITUS

Diabetic Ketoacidosis

DKA is the ultimate result of insulin deficiency[110,111] [*see Figure 11*], which is aggravated by stress-induced elevations of glucagon, cortisol, growth hormone, epinephrine, and norepinephrine[110] that add a component of insulin resistance.[112] DKA occurs in 2% to 5% of patients with type 1 diabetes mellitus a year. In the closely followed DCCT patients, overall event rates were 2.0 per 100 patient-years in the intensively treated group and 1.8 per 100 patients-years in the conventionally treated group.[34] Reported mortality varies worldwide from as low as 0% to as high as 10%. Most cases occur in patients already diagnosed with type 1 diabetes mellitus, but DKA still can be the first manifestation of diabetes, especially in children. Self-monitoring of blood glucose and urine ketones and close contact with the diabetes care team should facilitate recognition and abortion of evolving DKA by early and aggressive treatment with extra insulin and fluids at home. Approximately half the cases of DKA are precipitated by infection. Sepsis, myocardial infarction, and other major intercurrent illnesses are more often the cause of death than the metabolic disequilibrium itself. In children, cerebral edema rarely occurs. It usually appears 6 to 12 hours after treatment is initiated when biochemical improvement is manifest; yet it is often fatal.

Presenting features DKA presents with signs and symptoms of dehydration secondary to osmotic diuresis and vomiting and, sometimes, to diarrhea caused by concurrent gastroenteritis; of compensatory hyperventilation to eliminate CO_2; and of various degrees of depressed mentation or decreased consciousness. Seizures are notably not a result of DKA. Complete coma almost certainly indicates a long period of DKA before medical attention. DKA yields a number of characteristic laboratory findings [*see Table 5*]. The anion gap metabolic acidosis is secondary to elevated levels of acetoacetate and betahydroxybutyrate with small contributions from lactate and free fatty acids. Although serum potassium and phosphate levels are usually normal or even high initially, this finding masks a profound total body depletion of these electrolytes, along with magnesium. Deviations from the customary pattern create pitfalls in diagnosis. Ketones, which current tests detect only as acetoacetate or acetone, may be missing from the serum if the redox potential of the patient is very high and the equilibrium of the ketoacids is shifted toward the reduced partner betahydroxybutyrate (as may occur in alcohol intoxication). Serum bicarbonate levels may be normal if there is coexisting respiratory acidosis. Arterial blood pH may be normal if there is coexistent metabolic alkalosis caused by diuretic ingestion or pernicious vomiting. Occasionally, plasma glucose levels are less than 250 mg/dl because of fasting,[113] high alcohol intake, profound inanition, or pregnancy.

Treatment Treatment of DKA[110,114,115] requires careful monitoring of the patient [*see Table 5*]. Volume repletion is as important as insulin therapy.[116] Intravenous 0.9% saline should be started even before the diagnosis is established. After an initial liter in 30 to 60 minutes, fluid therapy should continue aggressively until the circulating volume is replenished, as indicated by an increase in blood pressure to normal and a reduction in compensatory tachycardia. Subsequent total volume repletion is carried out more slowly at 150 to 500 ml/hr with 0.45% saline, switching to 5% glucose-containing solutions once plasma glucose has decreased to 250 mg/dl. Typical fluid deficits range from 50 to 100 mEq/kg. Average sodium deficits are 7 mEq/kg, and most important, potassium deficits may be as high as 7 mEq/kg. The effective depletion of total body bicarbonate through loss of the strong organic acids acetoacetate and betahydroxybutyrate in the urine is revealed later, when a hyperchloremic metabolic acidosis often ensues. Potassium repletion (10 to 40 mEq/hr) should begin promptly after insulin administration and as soon as hyperkalemia and oliguria or anuria have been ruled out [*see Table 5*]. Otherwise, serious hypokalemia will result as insulin stimulates potassium uptake by cells [*see Figure 4*]. If the serum potassium level is less than 40 mEq/L on admission, a very large deficit exists and repletion should be at a faster rate to maintain a level no lower than 3.5 to 4.0 mEq/L. Insulin should be withheld in such circumstances until serum potassium reaches 4.0 mEq/L. Hypokalemia is the most tragic cause of death resulting from therapeutic misjudgment.

Although DKA can be managed satisfactorily with insulin given intramuscularly or subcutaneously, intravenous administration is far more reliable and results in fewer instances of hypokalemia and hypoglycemia. A bolus of 10 U or 0.1 U/kg is followed by the same dose given hourly by intravenous infusion, preferably with a pump and through its own intravenous line. Routine addition of sodium bicarbonate or potassium phosphate has not been found to hasten recovery in ordinary cases of DKA.[115] Possible indications for administration of sodium bicarbonate (50 to 200 mEq)

include arterial pH less than 7.0, ECG changes of hyperkalemia, hypotension that does not respond to rapid infusion of 0.9% saline, and left ventricular failure. If bicarbonate therapy is given, serum potassium and arterial pH should be monitored hourly and extra potassium given to prevent hypokalemia. Rhabdomyolysis, hemolysis, and central nervous system deterioration can be caused by severe hypophosphatemia (<1.5 mg/dl) and call for intravenous administration of 60 mmol (approximately 2 g) of phosphate as the potassium salt over 6 hours. Once the anion gap has decreased to near normal and bicarbonate has risen to 15 to 18 mEq/L, the insulin infusion rate can be decreased to 2 U/hr. In general, it is best to maintain the insulin infusion at 1 to 2 U/hr with accompanying 5% glucose infusion, aimed at keeping the plasma glucose level at around 150 mg/dl until the following morning, when a subcutaneously mixed insulin regimen can be started or resumed along with a diet.

It is preferable to treat patients with DKA in an intensive care unit to ensure close monitoring. Persistent vomiting calls for gastric intubation, and the airway of an obtunded patient should be protected to prevent aspiration. Any suspicion of sepsis mandates treatment with broad-spectrum antibiotics.

Hypoglycemia

Hypoglycemia is a more common emergency than DKA and potentially as dangerous. Clinical hypoglycemia can range from annoying symptoms accompanying a biochemically low blood glucose level (< 50 to 60 mg/dl) to confusion, seizures, or coma. Any episode that requires intervention by another person to reverse is categorized as severe hypoglycemia. Severe hypoglycemia can have disastrous consequences, particularly if the patient is driving any sort of vehicle, working at heights, or operating potentially dangerous machinery.

The most common causes of hypoglycemia are missed meals and snacks,[117] insulin dosage errors, exercise, alcohol, and drugs such as beta-adrenergic blockers. During the DCCT, 55% of hypoglycemic episodes occurred during sleep.[117] Such episodes often go undetected.[118]

Glucagon and epinephrine are the major counterregulatory hormones that are secreted in response to hypoglycemia.[119] Both restore glucose levels by increasing hepatic glucose output, while epinephrine also decreases the sensitivity of muscles to insulin. Furthermore, catecholamine secretion alerts the patient to treat the episode because it produces the sympathoadrenal symptoms noted below. Cortisol and growth hormone are also secreted in response to hypoglycemia[119] and play a role in maintaining glucose levels but not in rapid recovery from hypoglycemia.

Presenting features The most common symptoms of early mild hypoglycemia are adrenergic and include tachycardia, tremulousness, anxiety, and sweating.[120] The last symptom requires sympathetic activation of cholinergic nerves innervating the sweat glands.

Factors affecting severity of hypoglycemic episodes The development of primary or secondary adrenal insufficiency, hypopituitarism, and hypothyroidism may increase the risk of hypoglycemia by increasing sensitivity to insulin, decreasing appetite, or both. Stress, exercise, or use of alcohol or illicit drugs may blunt or prevent recognition of hypoglycemia. Patients who do recognize incipient hypoglycemia but who consciously do not respond expeditiously (for example, they may wait for a meal in a restaurant or continue to drive after symptoms first appear) are also at increased risk for severe hypoglycemia. Moreover, some risk factors for hypoglycemia have multiple effects that can precipitate, prolong, or worsen the severity of hypoglycemia. Alcohol, for instance, impairs judgment and inhibits gluconeogenesis and hepatic glucose output, thereby delaying recovery. When hypoglycemia is inadequately treated, more severe hypoglycemia often ensues.

Finally, because glucagon and epinephrine are the major defense hormones against prolonged hypoglycemia, their absence promotes longer and more severe episodes by two mechanisms: (1) compensatory hepatic glucose output is decreased when not stimulated by glucagon or epinephrine and (2) the familiar adrenergic symptoms may cease in the absence of epinephrine, resulting in failure to recognize the episode.[119] The glucagon response to hypoglycemia often wanes in patients after they have had type 1 diabetes mellitus for a few years. In the absence of glucagon, epinephrine secretion still provides adequate counterregulatory defense; however, epinephrine response can also be lost eventually, sometimes in association with other autonomic neuropathies and sometimes selectively. Many patients lose the ability to counterregulate against hypoglycemia during the first 10 years that they have type 1 diabetes mellitus.

Given the importance of intensive regimens to prevent microvascular complications from hyperglycemia, it is most unfortunate that a lowered glucose threshold for release of glucagon and epinephrine in response to hypoglycemia has been observed, particularly in patients undergoing intensive insulin therapy.[121] The lowered glucose level needed to stimulate counterregulation narrows the safety margin of therapy. For instance, the first symptom of hypoglycemia may occur only at glucose levels as low as 35 mg/dl (as opposed to 55 to 60 mg/dl) and may consist of confusion or loss of judgment, which interferes with self-treatment. Some evidence suggests that unawareness of hypoglycemia is self-generating, because each episode may lower the threshold at which autonomic counterregulation begins in subsequent episodes.[119] The converse of this is that a period free of hypoglycemia, produced by daily therapeutic contact with caregivers, may restore hypoglycemia awareness,[122,123] though it may not restore normal counterregulatory responses.[122] Increased uptake of glucose by the brain in the presence of hypoglycemia[124,125] is a likely explanation for the relative infrequency of clinical hypoglycemic catastrophes.

Treatment Patients recognize most episodes of hypoglycemia quickly and can effectively treat themselves with a promptly absorbed oral carbohydrate. Approximately 15 g of carbohydrate is sufficient to restore blood glucose levels to normal. This amount is provided by approximately 6 oz of orange juice, 4 oz of a cola drink, 3 to 4 tsp of table sugar, five Life Savers, or three glucose tablets (each containing 5 g of glucose). The use of complex carbohydrates and foods with a high fat content, such as chocolate, may delay digestion and absorption of the glucose and are not first choices for treatment of hypoglycemia. If the patient cannot swallow or cooperate, a gel form of glucose and simple carbohydrates can be administered by mouth, applying it between the gums and cheeks, from where it slowly and generally safely trickles down into the stomach. Glucagon (1 mg administered subcutaneously or intramuscularly) will also usually raise blood glucose levels sufficiently within 15 to 30 minutes, when the patient can then take oral carbohydrates. Glucagon comes in emergency kits, and it should always be on hand for patients with a history of severe hypoglycemic episodes. Glucagon may

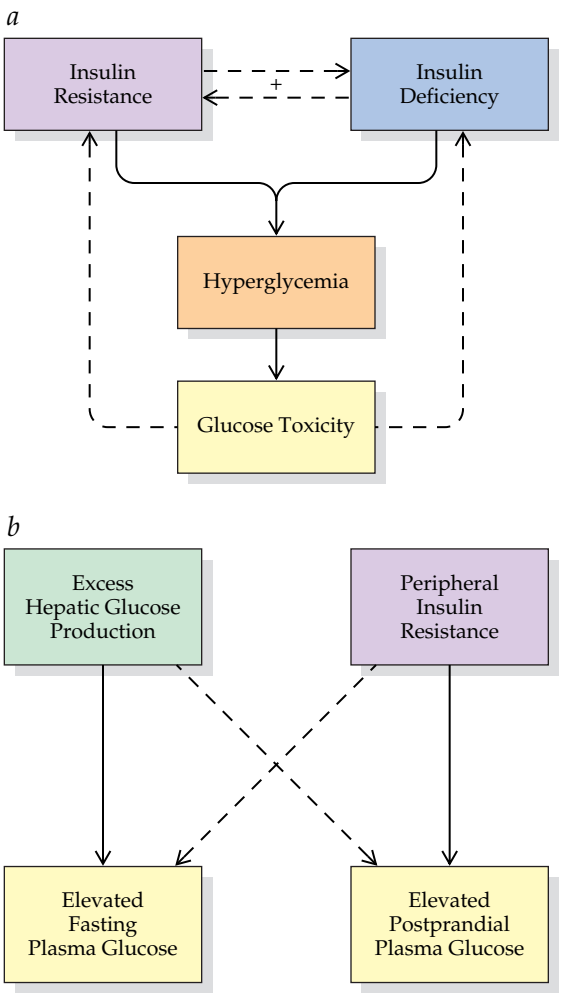


Figure 14 **(*a*) The interrelations of insulin resistance, insulin deficiency, and glucose toxicity that create overall hyperglycemia in type 2 diabetes mellitus are depicted. Insulin resistance and insulin deficiency are mutually reinforcing factors. Glucose toxicity refers to the secondary aggravating effects of hyperglycemia that both increase insulin resistance and reduce beta cell function. The glucose toxicity is diminished or eliminated by any therapy that lowers blood glucose. (*b*) Once fasting glucose levels are abnormal, they are correlated with and largely driven by the excess hepatic glucose production. Abnormal postprandial glucose levels are largely a consequence of peripheral insulin resistance that makes glucose utilization in muscle and adipose tissue inefficient. Insulin deficiency plays an increasingly important role in elevating both fasting and postprandial glucose levels as time goes on.**

cause nausea, vomiting, and headache, especially in children. When all else fails, intravenous glucose must be given by emergency medical service personnel or in an emergency room, whichever is quicker. When the timing of an episode suggests it was caused by intermediate- or long-acting insulin or by prior exercise, blood glucose may fall to hypoglycemic levels again and re-treatment may be necessary. Thus, a patient who has required assistance from others in reversing hypoglycemia should be kept under surveillance for some time thereafter.

Patients with severe hypoglycemia usually respond rapidly to treatment, although patients who are postictal or in a prolonged coma may require days to regain normal mental status and cognitive function. Quite often, there is amnesia for such extended episodes, including a period preceding the onset of hypoglycemia. In rare instances, neurologic deficits can be permanent. In general, however, long-term consequences of hypoglycemia have not been detected in adults.[126,127] In view of the potential consequences of prolonged episodes, hypoglycemia should always be treated immediately.

Prevention Patients should be instructed to treat themselves as though they have hypoglycemia whenever they suspect it, even if they are unable to do a confirmatory blood glucose test at the time. The threshold for symptoms of hypoglycemia varies from person to person and even varies in the same person on different occasions. Therefore, whenever possible, a confirmatory blood glucose test should be done to help the patient discriminate nonspecific symptoms from true hypoglycemia. Patients at increased risk for severe hypoglycemia should monitor their blood glucose levels more frequently.

Type 2 Diabetes Mellitus

PATHOGENESIS OF TYPE 2 DIABETES MELLITUS

Insulin Resistance and Insulin Deficiency

The pathogenesis of type 2 diabetes mellitus[128] is even more complex than that of type 1 diabetes mellitus. Insulin resistance, reported in 92% of one large group of people with type 2 diabetes mellitus,[129] plays a major role in generating hyperglycemia.[128] In addition, some degree of functional insulin deficiency exists [*see Figure 14*].[128] Certain studies suggest that insulin resistance is primary[128,130] and that impaired insulin secretion is only really evident when fasting hyperglycemia supervenes.[128, 131-133] Other investigators find evidence of early abnormal beta cell function in type 2 diabetes mellitus,[134-136] in IGT,[134,137] and in first-degree glucose-tolerant relatives of patients with type 2 diabetes mellitus.[138] Regardless of which comes first, the loss of compensatory beta cell hyperfunction to overcome insulin resistance is a key factor in the progression from genetic susceptibility to established type 2 diabetes mellitus.[139] Furthermore, insulin resistance may cause secondary insulin deficiency, and insulin deficiency tends to lead to insulin resistance; thus, they are mutually reinforcing defects, partly through an effect commonly referred to as glucose toxicity.[140] Some period of hyperglycemia has a secondary noxious effect that aggravates both insulin resistance and insulin deficiency; thus, hyperglycemia begets hyperglycemia. Therefore, any form of treatment of type 2 diabetes mellitus that lowers plasma glucose levels is self-reinforcing and may gain momentum with time by virtue of the added early benefit of eliminating the effects of glucose toxicity. For this reason, aggressive early treatment (e.g., with insulin) can sometimes be replaced with oral drugs or even diet.[141]

The exact locus of insulin resistance in type 2 diabetes mellitus remains unidentified. Indeed, there may be various sites because the disease is considered likely to be a heterogeneous disorder.[142,143] Numerous candidate genes for defective insulin action, including the insulin receptor, glucose transporter, insulin receptor substrate, and insulin target enzymes, such as glycogen synthase, have been largely excluded as common primary causes of insulin resistance[144,145] in type 2 diabetes mellitus.

As in type 1 diabetes mellitus, the loss of effective insulin action directly leads to unrestrained hepatic glucose production and inefficient peripheral glucose utilization [*see Figure 14*]. Exces-

Table 6 Definitions of the Metabolic Syndrome[155,156]

National Cholesterol Education Program Adult Treatment Panel III	*World Health Organization*
At least three of the following:	Diabetes, IGT, or IFG and/or insulin resistance*
Fasting plasma glucose ≥ 110 mg/dl	*plus at least two of the following:*
Abdominal obesity: waist circumference 35 in. in women or > 40 in. in men	Abdominal obesity: waist-to-hip ratio > 0.85 in women or > 0.9 in men and/or body mass index > 30 kg/m^2
Triglycerides > 150 mg/dl; HDL < 50 mg/dl in women or < 40 mg/dl in men	Triglycerides > 150 mg/dl and/or HDL < 40 mg/dl in women or < 35 mg/dl in men
Blood pressure ≥ 130/85 mm Hg	Blood pressure ≥ 140/90 mm Hg
	Microalbuminuria: urinary albumin excretion ≥ 20 µg/min or albumin-to-creatinine ratio ≥ 30 mg/g

*Insulin resistance assessed as fasting insulin ÷ (fasting glucose × 22.5).
HDL—high density lipoprotein IFG—impaired fasting glucose IGT—impaired glucose tolerance

sive hepatic glucose output largely accounts for elevation of FPG levels.[128] Resistance to the antilipolytic action of insulin in adipose tissue leads to elevated plasma free fatty acid (FFA) levels and increased FFA delivery to the liver. There, the oxidation of FFA generates energy (adenosine triphosphate [ATP]) needed to sustain gluconeogenesis; in addition, the latter process is stimulated by FFA metabolites such as acyl coenzyme A (acyl-CoA). In this indirect manner, insulin resistance also contributes to elevated glucose production in the liver.[146] Moreover, the elevation of FFA levels also contributes to insulin resistance in muscle.[147] The presence of some residual insulin secretion in type 2 diabetes mellitus, however, is ordinarily enough to restrain ketogenesis and prevent DKA. Elevated hepatic glucose output largely sustains an elevated FPG, whereas reduced peripheral glucose utilization especially causes elevation of postprandial glucose levels [*see Figure 14*].

The ratio of proinsulin to insulin in plasma is high and remains so even after glucose-lowering therapy,[148,149] suggesting an early abnormality in processing of proinsulin to insulin in the beta cell [*see Figure 3*]. Insulin is normally secreted in cyclic pulses that can be entrained by rapid changes in plasma glucose levels. Disruption of this close concordance between plasma glucose and plasma insulin fluctuations is a subtle lesion that is demonstrable early in patients with type 2 diabetes mellitus and, to a lesser extent, in some patients with only impaired glucose tolerance.[134,136] Finally, the plasma insulin response to abrupt elevation of plasma glucose levels normally shows a first sharp, spikelike phase.[150] Before the plasma insulin level returns to baseline, it slowly rises again to produce a second plateau phase of more prolonged insulin release. The immediate first-phase response to glucose decreases in type 2 diabetes mellitus, as it does in the preclinical phase of type 1 diabetes mellitus, and is completely lost when the FPG level exceeds the normal range.[151]

Other Beta Cell Abnormalities

Another, previously neglected abnormality in type 2 diabetes mellitus is the presence of amyloid in close proximity to the islet beta cells. The amyloid fibrils have been found to contain amylin, a peptide that is cosecreted with insulin.[152] Amylin deficiency parallels insulin deficiency in type 2 diabetes mellitus.[152] Whether the accumulation of amyloid impairs beta cell function or is an epiphenomenon resulting from beta cell hyperfunction with increased amylin secretion in the early phases of the disease remains unclear.

More than 10% of some patient populations presenting with the clinical phenotype of type 2 diabetes mellitus have serum islet cell autoantibodies typical of type 1 diabetes mellitus, such as antibodies to GAD.[153] This combination has been referred to as latent autoimmune diabetes in adults (LADA). These individuals exhibit a rapid decline in beta cell function, as shown by serum C-peptide levels, and they are likely to need insulin replacement therapy, even if their hyperglycemia is initially alleviated by oral beta cell stimulants.[154,155]

Metabolic Syndrome (Insulin-Resistance Syndrome)

The metabolic (insulin-resistance) syndrome is closely associated with, and often a forerunner of, type 2 diabetes mellitus. The metabolic syndrome has been defined both by the National Cholesterol Education Program and the World Health Organization (WHO) [*see Table 6*].[156,157] Only the WHO definition includes insulin resistance per se, assessed by determining the ratio of the fasting plasma insulin level to the glucose level. By either definition, the syndrome represents a collection of risk factors not only for diabetes regulation but also for cardiovascular disease, and it presages both diseases. One obvious link between the components of the metabolic syndrome is obesity, which is a cause of insulin resistance[158] and a contributor to the insulin resistance of type 2 diabetes mellitus.[159] Weight gain presages diabetes,[160] and weight loss in obese individuals prevents progression of IGT to full-blown diabetes.[161] Most patients with type 2 diabetes mellitus have abdominal obesity and many have dyslipidemia, hypertension, and other features of the metabolic syndrome.[129] Abdominal obesity is itself a risk factor for type 2 diabetes mellitus and cardiovascular disease.[162] The interrelationship of the metabolic syndrome, diabetes, and cardiovascular disease is exemplified in the NHANES III study. The overall prevalence of the metabolic syndrome in the United States is 23% in men and women older than 20 years and 44% in those older than 50 years.[163] This huge prevalence of the metabolic syndrome reflects the burgeoning of obesity in the population. In the over-50 age group, 87% of those with diabetes, 71% of those with IFG, and 33% of those with IGT also had the metabolic syndrome.[164] The prevalence of coronary heart disease (CHD) was 8.7% in those with neither diabetes nor the metabolic syndrome and 7.5% in those with only diabetes, but it increased to 13.9% in those with only the metabolic syndrome and to 19.2% in those with both the metabolic syndrome and diabetes. These cross-sectional data suggest that the metabolic syndrome is more potent than diabetes per se as a risk factor for CHD, but hyperglycemia (diabetes) aggravates the risk inherent in the metabolic syndrome. This relation explains much but not all of the vulnerability of patients with type 2 diabetes mellitus to cardiovascular complications resulting from accelerated atherosclerosis. On the other hand, not all patients with the metabolic syndrome and IGT go on to experience full-blown type 2 diabetes mellitus. A large randomized controlled trial is currently testing the ability of lifestyle changes (weight reduction and regular exercise) and the drug metformin to reduce the risk of progressing from IGT to type 2 diabetes mellitus.[165]

Genetic Factors

Type 2 diabetes mellitus has a strong hereditary component. In virtually all monozygotic twinships, the disease develops in both individuals, often within a few years of each other.[166] Offspring and siblings of diabetic patients are at great risk for the disease.

Table 7 American Diabetes Association Plasma Glucose Diagnostic Criteria for Diabetes Mellitus

Diagnosis	*Test Condition*	
	Plasma Glucose (mg/dl)	
	Fasting ≥ 8 hr	*2 hr after 75 g Oral Glucose*
Normal	< 110	< 140
Impaired glucose tolerance (IGT)	< 126	≥ 140–< 200
Impaired fasting glucose (IFG)	≥ 110–< 126	< 200
Diabetes mellitus	≥ 126	—
Diabetes mellitus	< 126	≥ 200
Diabetes mellitus (Classic symptoms + casual plasma glucose, ≥ 200 mg/dl)	—	—
	Plasma Glucose (mg/dl)	
	Fasting	*After 100 g Oral Glucose*
Gestational diabetes mellitus (GDM)	> 105*	1 hr ≥ 190*
		2 hr ≥ 165*
		3 hr ≥ 145*

Note: The Fourth International Workshop-Conference on Gestational Diabetes Mellitus has proposed lower criteria, which would increase the percentage of cases from 4% to 7% in white women. These criteria are fasting, 95; 1 hour, 180; 2 hours, 155; and 3 hours, 140, after 100 g oral glucose.
*Two of these four criteria must be met for diagnosis of GDM.

No HLA markers have been identified for type 2 diabetes mellitus, in contrast to type 1 diabetes mellitus. Most current thinking is that the common forms of type 2 diabetes mellitus represent a complex multigenic disorder. Examination of the mechanism of action of insulin [*see Figure 4*] suggests many logical candidate genes, mutations of which could lead to type 2 diabetes mellitus by causing primary insulin resistance. Thus far, genes for insulin, the insulin receptor, insulin receptor substrate, glucose transporter, protein tyrosine phosphatase (which inactivates the insulin receptor), muscle hexokinase, glycogen synthase, and other insulin target enzymes have all been excluded as the cause of so-called garden-variety type 2 diabetes mellitus.[167] Because of the association with obesity, genes that could cause obesity are also being investigated (e.g., leptin, uncoupling protein, and beta$_3$-adrenergic receptor). The positional cloning approach being used in populations with high diabetes prevalence, such as Pima Indians and Mexican Americans, has yielded hints of loci on certain chromosomes that require confirmation.

There is one form of diabetes, MODY [*see Table 1*], that does have genetic specificity. In this disorder, mutations of several different genes on different chromosomes lead to a common phenotype resembling type 2 diabetes mellitus, but the disorder begins at an early age.[168] One of the genes codes for glucokinase, an enzyme that plays a key role in stimulation of insulin secretion by glucose.[169] Another mutation occurs in a molecule known as insulin production factor–1, a transcription factor responsible for differentiation of precursor cells into beta cells capable of insulin secretion.[168] Two other genes responsible for MODY code for hepatic transcription factor–1 and hepatic transcription factor–4, which, despite their names, operate in beta cells to regulate the glucose responsive pathway of insulin secretion.[168] All of these genetic abnormalities more likely explain type 2 diabetes mellitus caused by beta cell dysfunction than that caused by peripheral insulin resistance. Their functional relation to the diabetic diathesis is still obscure. Even in a phenotypically well defined monogenic form of diabetes such as MODY, the existence of many alleles for hepatic transcription factor–1 indicates the genetic complexity of diabetes. Although the mutations responsible for MODY account for only a minute fraction of all cases of type 2 diabetes mellitus, they encourage the view that genes contributing to most or all cases of type 2 diabetes mellitus will eventually be found.

IMPAIRED GLUCOSE TOLERANCE

The state known as IGT [*see Table 7*] is associated with a future risk of development of diabetes of 1% to 10% a year, with different levels of risk for different ethnic groups. Equally important is the association of IGT with the metabolic syndrome [*see Table 6*], which includes hyperinsulinemia, glucose intolerance, dyslipidemia, hypertension, and impaired fibrinolysis. Presence of this syndrome constitutes a high risk for atherosclerosis, cardiovascular disease, thrombotic events, and mortality. The category of impaired fasting glucose was established by the American Diabetes Association as an intermediate zone between the upper limit of normal and the lower limit for diabetes.[1] IFG is also associated with increased risk of diabetes and cardiovascular disease. IFG and IGT are not identical states. About one third of people with IGT have IFG, one third of those with IFG have IGT, and one third of affected individuals have both.[2] The pathophysiologic bases and clinical significance of the differences between IFG and IGT remain to be determined. Both conditions can be thought of as early stages of type 2 diabetes mellitus and can be referred to as prediabetes.

PREVENTION OF TYPE 2 DIABETES

Five randomized clinical trials have recently demonstrated that the risk of progression from IGT to diabetes can be significantly reduced by lifestyle modifications or pharmacologic interventions. The Diabetes Prevention Program (DPP),[170] the Finnish Diabetes Prevention Study,[171] and the Da Qing IGT and Diabetes Study[172] showed that intensive diet and exercise therapy brought reductions ranging from 42% to 58% in the progression from IGT to diabetes over 3 to 6 years. The weight loss achieved and the amount of exercise performed were modest—5.6 kg and 150 minutes of brisk walking a week in the DPP. The DPP also had a placebo-controlled metformin (850 mg twice daily) treatment arm, which showed a 31% reduction in diabetes. Most of this effect persisted after a 1-week washout from metformin. In the STOP-NIDDM trial, 100 mg of acarbose three times daily, compared with placebo, reduced diabetes development by 25%.[173] Finally, in a group of Hispanic women with previous gestational diabetes, 400 mg of troglitazone daily, compared with placebo, reduced development of diabetes.[174] This benefit was still present after an 8-month drug washout.

The efficacy, safety, and consistency of lifestyle interventions are impressive, but long-term follow-up is needed to determine how long such patients will continue to implement this therapy and how durable the benefits will be from either lifestyle changes or drugs. Equally important is whether car-

diovascular disease events will be reduced eventually. A preliminary report from the first 3 years of the STOP-NIDDM trial suggests an encouraging significant reduction in myocardial infarction and total cardiovascular disease events.[175]

DIAGNOSIS OF TYPE 2 DIABETES MELLITUS

Although patients with type 2 diabetes mellitus may present with symptoms as florid as those of type 1 diabetes mellitus (but usually not exhibiting spontaneous ketonuria), most patients with type 2 disease have relatively mild polyuria and polydipsia, and many cases are diagnosed only by office screening or other health checks.

The preferred test for type 2 diabetes mellitus on the grounds of reproducibility, convenience, and cost is an FPG . Oral glucose tolerance testing (OGTT) is more sensitive than FPG but is not recommended for routine use, because it is less reproducible, more inconvenient, and more costly.[1] Moreover, the recommended treatment for almost all overweight or obese patients who would be candidates for OGTT would be the same regardless of OGTT results: a combined regimen of nutrition therapy, weight loss, and exercise. OGTT may be considered in unusually high risk patients and in those with IFG.

MANAGEMENT OF TYPE 2 DIABETES MELLITUS

The same glycemic goals discussed earlier [*see Table 5*] are appropriate for type 2 diabetes mellitus. However, these goals may sometimes have to be modified if severe cardiovascular disease, concurrent life-shortening malignancy, hypoglycemia unawareness, or inadequate family or social support make intensive treatment of diabetes dangerous or unlikely to benefit the patient in the long run. Self-monitoring of blood glucose when patients with type 2 diabetes mellitus are treated with diet plus exercise or with oral drugs is of less well established utility in patients with type 2 diabetes mellitus than in patients with type 1 diabetes mellitus. However, fasting and postprandial blood glucose levels both correlate with HbA_{1c} levels, and postprandial values can help reveal inadequate attention to diet and insufficient effectiveness of certain oral agents.

Nutritional Therapy and Exercise

An excellent short-term glycemic response to caloric reduction in patients with type 2 diabetes mellitus who are even modestly overweight can be expected.[175, 176] On the basis of the degree of obesity and with the help of a dietitian, the patient should be provided with individualized culturally appropriate instructions to reduce intake by at least 250 to 500 calories a day. Such a decrease generally leads to an overall weight loss of 0.5 to 1.0 lb a week. There should be periodic reinforcement by the dietitian and physician. In the absence of a dietitian, the patient's basal metabolic rate can be estimated at 10 cal/lb (20 cal/kg) of ideal body weight. A caloric prescription less than this amount will perforce decrease energy intake below the total daily energy expenditure. Consensus guidelines recommend that the calories should consist of less than 30% total fat, less than 10% saturated fat, less than 10% polyunsaturated fat, 10% to 15% monounsaturated fat, 10% to 20% protein, and 50% to 55% carbohydrate.[177] Table sugar and other concentrated forms of carbohydrates are allowable in small portions at any one time (e.g., 5 g or 1 tsp of table sugar). Adding high-fiber foods can also lower plasma glucose modestly.[178] Teaching patients to count the contemplated grams of carbohydrate before each meal helps them limit elevation of postprandial plasma glucose (PPG).

In massively obese individuals with BMI greater than 40 who are very symptomatic from hyperglycemia, a very low calorie diet (400 to 800 total calories a day using special high-protein supplements) can be very effective for the initial 2 to 3 months, but this strategy requires close medical monitoring.[179]

Weight losses of 5% to 10% (10 to 20 lb) produce significant decreases in FPG and HbA_{1c} over 1 to 3 months.[176] In the UKPDS, mean HbA_{1c} fell from 9% to 7% during the 3-month dietary run-in period before randomization of the study patients.[180,181] However, many patients are unable to maintain a calorie-restricted diet and even their initial weight loss. Pharmacologic aids for weight loss can be considered in such cases, but their efficacy is limited. These drugs include orlistat,[182] a gastrointestinal lipase inhibitor that causes malabsorption of fat calories, and sibutramine, an inhibitor of dopamine, norepinephrine, and serotonin reuptake. Even after the addition of a weight-loss drug to therapy, appropriate diet therapy is essential. The patient should not be blamed for recidivism, because inability to lower body weight to ideal and keep it there may well be a central nervous system manifestation of or contributor to type 2 diabetes mellitus and out of the patient's consistent control.[183] Surgical therapy for obesity by reduction of gastric volume[184,185] can effectively control type 2 diabetes mellitus and is gaining acceptance in very obese individuals who are unresponsive to other therapy. Of great interest, such treatment may work by altering the concentrations of humoral signals from the gastrointestinal tract that regulate appetite.

Additional benefits accrue from gradually increased aerobic exercise[186] aimed at achieving at least 60% of maximal heart rate (220 minus age), such as walking 45 minutes at a brisk pace (approximately 3 to 5 miles an hour) three to five times a week. Exercise decreases insulin resistance and glycemia, contributes modestly to weight loss, reduces the risk of future cardiovascular disease, improves prognosis should a myocardial infarction occur, and enhances the patient's sense of well-being and physical fitness. Conversely, physical inactivity predicts mortality in men with type 2 diabetes mellitus.[187] In the presence of known coronary artery disease (CAD), the exercise should be prescribed with input from the patient's cardiologist. If type 2 diabetes mellitus has existed for more than 5 to 10 years or if the patient already has peripheral vascular or cerebral vascular disease, autonomic neuropathy, microalbuminuria, dyslipidemia, or a history of smoking, an electrocardiogram is essential and an electrocardiographic exercise tolerance test is prudent before initiating a formal exercise program.

Pharmacologic Monotherapy

The array of pharmacologic agents available for treatment of type 2 diabetes mellitus is increasing steadily. Drugs can be specifically directed at the known pathophysiologic defects in type 2 diabetes mellitus [*see Figure 15*]. Although patient compliance favors initial use of monotherapy, none of the available agents can alone be expected to adequately control hyperglycemia indefinitely.[188] Therefore, diabetologists are beginning to consider using combinations of drugs from the outset of the need for any pharmacotherapy.[189] Clinical trial data have established the efficacy and safety of various drugs [*see Table 8*] . The degrees to which these drugs lower HbA_{1c} are fairly similar; the higher the initial dose of these agents, the greater the decrease in HbA_{1c}.

Sulfonylurea agents SU agents, the oldest oral hypoglycemic drugs, continue to have an important place in treat-

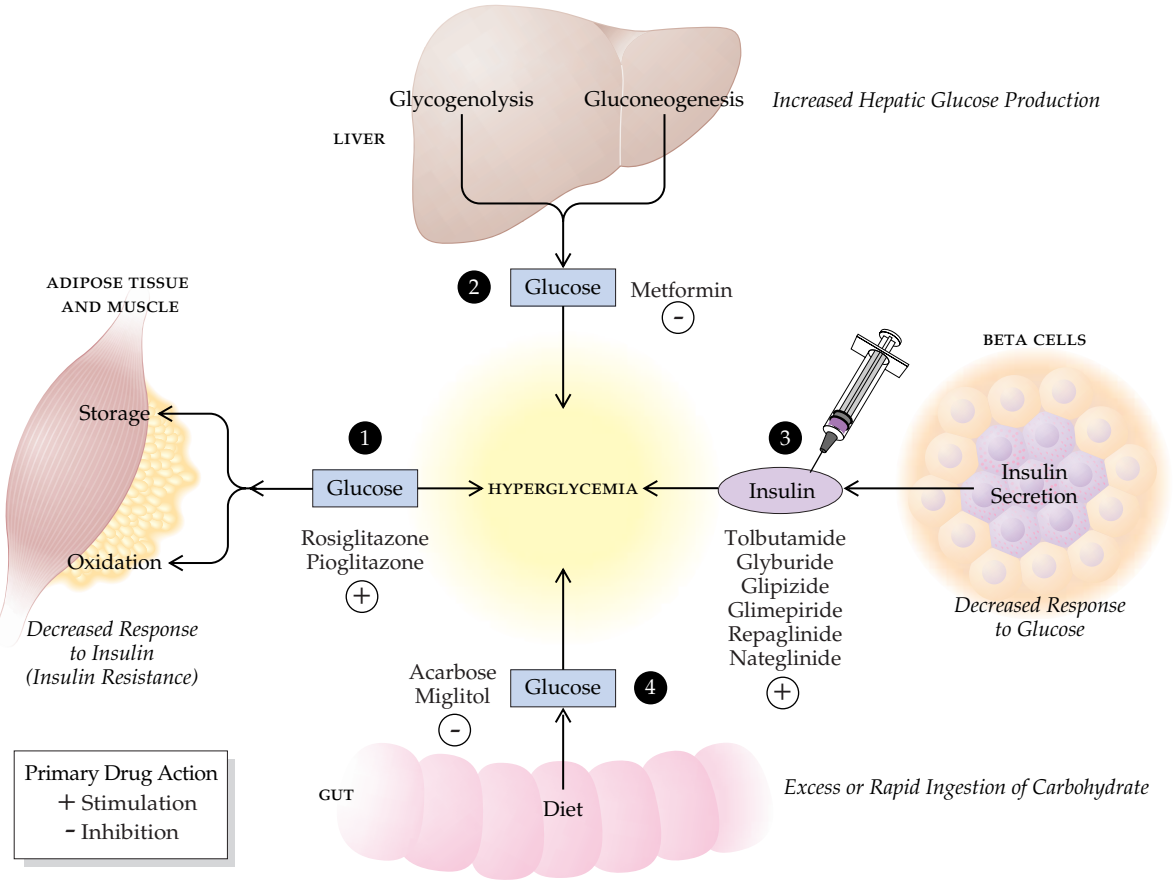


Figure 15 **Multiple drug classes with different predominant therapeutic effects are available for use singly or in numerous combinations. (1) Glitazones (thiazolidinediones) increase the sensitivity to insulin of glucose uptake by muscle and adipose tissue. (2) Metformin, the only approved biguanide drug, inhibits glucose production by the liver. (3) Sulfonylureas, repaglinide, and nateglinide stimulate insulin secretion, and insulin itself can be provided by injection. (4) α-Glucosidase inhibitors slow the digestion and absorption of carbohydrates from the diet. Plus signs indicate stimulation. Minus signs indicate inhibition.**

ment. Their primary mechanism of action is to close ATP-sensitive potassium channels in the beta cell (and other cell) membranes, which leads to an influx of calcium and stimulation of exocytosis of insulin storage granules. They are most effective in normal-weight or modestly obese individuals who have had diabetes for less than 5 years and who can still secrete considerable amounts of insulin. The SU drugs in common use stimulate the beta cells more or less continuously and secondarily decrease insulin resistance. These effects [*see Figure 14*] result in decreases in FPG of 50 to 70 mg/dl and HbA_{1c} of 1.0% to 2.0%.[190] Peak PPG levels fall approximately as much as FPG. For most patients, treatment is initiated with the lowest recommended dose, and the dose is increased every 1 to 2 weeks until target blood glucose levels are attained or a practical maximal dose is reached [*see Table 9*]. Modern SU drugs can be taken as a single daily dose but occasionally are more effective when split into twice-daily doses. In symptomatic patients with FPG greater than 250, the patient may begin with half the maximal recommended dose. Hypoglycemia, in particular, and weight gain are adverse effects of SU drugs. The highest prevalence of hypoglycemia occurs with glyburide and chlorpropamide,[191] drugs with long biologic half-lives. Elderly patients who live alone and lack concerned family, friends, or neighbors are at a special risk for severe, even fatal, hypoglycemia.[192] The shortest-acting SU drug, tolbutamide, may be the safest to use in such cases.

Patients who present to an emergency room in hypoglycemic coma from any SU drug should be given restorative treatment with intravenous boluses of glucose and then admitted to the hospital because SU drugs can have durations of biologic action for up to 7 days. A blood glucose should be maintained at 150 to 200 mg/dl on intravenous glucose, oral carbohydrate, or both until this level can be sustained by administration of only 5 g/hr of one of the therapeutic agents. In the UKPDS, SU drugs did not increase cardiovascular disease events or mortality.[37] This observation relieves much of the concern previously raised by the University Group Diabetes Program trial in 1970, which found that tolbutamide was associated with an excess of cardiovascular and total deaths.[193] However, interactions of even modern SU drugs with cardiac muscle are reported, particularly inhibition of ischemic preconditioning, a cardioprotective mechanism.[194] SU drugs are contraindicated in hepatic insufficiency and are dangerous when combined with alcohol ingestion. Glimepiride, the newest SU drug,[195] has been given safely to patients with renal insufficiency, although these patients are susceptible to hypoglycemia for other reasons. SU drugs are subject to interactions with other drugs that can either exaggerate or interfere with their effects.

Beta cell stimulants Repaglinide and the phenylalanine derivative nateglinide represent a new class of beta cell stimulants [*see Figure 15*] that differ in structure and timing of action

Table 8 Oral Drugs for Type 2 Diabetes Mellitus

Drugs	*Lowest Effective Single Dose (mg)*	*Practical Maximum Daily Dose (mg)*	*Hypoglycemia with Monotherapy*
Sulfonylureas*			
Glyburide	1.25	10	Yes
Micronized glyburide	1.5	6	Yes
Glipizide	5	20	Yes
Glipizide (gastrointestinal therapeutic system)	5	20	Yes
Glimepiride	0.5	8	Yes
Meglitinides†			
Repaglinide	0.5	4‡	Yes
Biguanides			
Metformin	500	2,000	No
Thiazolidinediones§			
Rosiglitazone	2	8	No
Pioglitazone	15	45	No
α-Glucosidase inhibitors			
Acarbose	25	100‡	No
Miglitol	25	100‡	No

*Tolbutamide, chlorpropamide, and acetohexamide are also still available.
†Nateglinide has a similar action to meglitinides but is technically phenylalanine derivative.
‡This maximal dose must be taken each time with meals.
§Troglitazone was the initial drug approved in this class but was later withdrawn because of serious liver toxicity.

from those of SU drugs.[196] Although they may act in part through SU drug mechanisms in the beta cells,[197] they do so rapidly, with a peak effect at about l hour, and transiently, with a duration of about 4 hours. Their major action is the decrease of PPG by 50 to 60 mg/dl, although FPG also declines somewhat as glucose toxicity is relieved. As monotherapy, these new beta cell stimulants are most logically used early in type 2 diabetes mellitus, when FPG is not greatly elevated. They lower HbA_{1c} by about 1.0%. Repaglinide and nateglinide must be taken 15 to 30 minutes before a meal and should never be taken without eating. Their short half-lives and the fact that, unlike SU drugs, they are active only in the presence of glucose are expected to reduce the likelihood of severe prolonged hypoglycemic episodes.[196] Weight gain may occur secondary to improved glycemic control.

Table 9 Combination Oral Drug Therapy for Type 2 Diabetes Mellitus

Combinations reported in the literature
- Sulfonylurea + metformin
- Sulfonylurea + thiazolidinedione
- Metformin + thiazolidinedione
- Metformin + repaglinide
- Repaglinide + thiazolidinedione
- Sulfonylurea + metformin + thiazolidinedione
- Acarbose + any other drug except repaglinide
- Miglitol + sulfonylurea
- Insulin + any other drug

Potentially useful combination
- Repaglinide + metformin + thiazolidinedione
- Nateglinide + metformin + thiazolidinedione

α-Glucosidase inhibitors α-Glucosidase inhibitors[197,198] are a class of drugs represented by acarbose and miglitol, which are poorly absorbed but act within the gut to inhibit the digestion of polysaccharides [*see Figure 15*]. This action results in a slow release of glucose from food and therefore slow absorption from the GI tract. PPG levels decrease by 60 to 70 mg/dl, but FPG decreases by only 15 to 20 mg/dl.[199] HbA_{1c} generally falls 0.5% to 0.8%.[198,199] These drugs are useful only as monotherapy when postprandial hyperglycemia is the main problem. They must be taken at the start of a meal. Flatulence, abdominal cramping, and diarrhea are frequent side effects that result from undigested carbohydrate reaching bacteria in the lower bowel. These side effects often limit patient acceptance of treatment with α-glucosidase inhibitors. Treatment should start with the smallest dose, and doses should be raised very gradually to enhance tolerance. With the exception of rare elevations of alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels, these agents are nontoxic. Although hypoglycemia does not occur with monotherapy, it can do so when α-glucosidase inhibitors are added to SU drugs or insulin. In those instances, patients must be warned to treat hypoglycemia only with pure glucose (e.g., glucose tablets) because the therapeutic benefits of complex carbohydrates and even sucrose will be delayed by slow digestion.

Biguanide agents Metformin,[200] the only drug approved in the biguanide class, acts primarily by decreasing excessive hepatic glucose production [*see Figure 15*],[201,202] most likely through inhibiting gluconeogenesis. Because insulin also inhibits gluconeogenesis [*see Figure 1*][20] and metformin requires the presence of insulin to be effective, metformin may be considered a hepatic insulin sensitizer. During metformin treatment, plasma insulin levels tend to decrease relative to glucose levels as a result of the decrease in insulin resistance. The chief action of this drug is to lower FPG by 50 to 70 mg/dl, with peak PPG levels following

suit.[203] Hypoglycemia almost never occurs with metformin monotherapy. Weight is not gained and may even be lost.[200] Metformin also decreases plasma triglyceride and low-density lipoprotein (LDL) cholesterol levels, and it increases HDL cholesterol levels to some degree. In addition, plasma plasminogen activator inhibitor–1 (PAI-1) activity declines.[204] The weight loss, the improvement in the dyslipidemia typical of type 2 diabetes mellitus, and the reduction in antifibrinolytic activity could explain one of the most interesting UKPDS observations. Compared with conventional diet treatment, metformin monotherapy substantially decreased the incidence of myocardial infarction, diabetes-related death, and all-cause mortality in an obese type 2 subcohort of the trial.[38]

The most common side effects of metformin therapy are diarrhea, which may be severe; abdominal cramps; and gastric upset. To reduce the likelihood of these symptoms, the starting dose should not exceed 500 mg twice a day, and the drug should be used with special caution, if at all, in patients who have inflammatory GI disease. The maximum effective dose is 2,000 mg/day.[205] The most feared, although rare, adverse effect is lactic acidosis.[206] This often fatal complication occurs in 30 per million patients a year, and it usually occurs when the drug is used inappropriately, such as when the serum creatinine level is elevated or the patient is dehydrated. Hemodialysis along with bicarbonate administration can be effective therapy for metformin-induced lactic acidosis.[207] The following are contraindications to use of metformin: serum creatinine level greater than 1.4 mg/dl in women and greater than 1.5 mg/dl in men; intravenous administration of radiographic iodinated contrast media; acute myocardial infarction; congestive heart failure; and any ischemic condition. Nausea, vomiting, tachypnea, and change in mental status call for measurements of serum electrolytes and lactate to rule out lactic acid metabolic acidosis. Although metformin monotherapy can be effective in both normal-weight and obese patients with type 2 diabetes mellitus, obese patients especially benefit from metformin therapy because of the absence of weight gain as glucose levels gradually fall.

Thiazolidinediones The newest class of oral drugs are thiazolidinediones (TZDs) [*see Table 8*],[208] which were exemplified by the no longer available but best studied drug, troglitazone . These agents work predominantly in muscle and adipose tissue to decrease insulin resistance [*see Figure 15*].[202,209] Because insulin resistance is seen in almost all patients with type 2 diabetes mellitus, the advent of the TZDs raised expectations that they might be singularly effective. Like metformin, TZD drugs need the presence of insulin and are especially effective in obese patients. They also decrease hepatic glucose production to some extent.[202] As monotherapy, they decrease FPG by 50 to 70 mg/dl and PPG by slightly more than that.[209] HbA_{1c} decreases by about 1.0% to 1.5%. Plasma insulin levels also decrease as glucose levels fall.[209] TZD drugs act by binding to a metabolically important receptor, the peroxisome proliferator-activated receptor (PPAR), and they thereby regulate the expression of multiple genes.[210] Their clinical effects take 4 to 12 weeks to become evident. In patients with marked elevation of FPG, a midrange dose is appropriate to start with (e.g., 4 mg rosiglitazone or 30 mg pioglitazone). Otherwise, the lowest dose is appropriate, and dose changes should be made at 4- to 12-week intervals.

TZD drugs can cause weight gain, which is partly fat tissue and partly extracellular fluid.[208] The accumulation of extracellular fluid presents as edema, which can be troublesome when concurrent congestive heart failure exists, and a small dilutional fall in hemoglobin and hematocrit. Some findings suggest that the adipose tissue gain is largely subcutaneous rather than visceral.[211] In clinical research trials, troglitazone therapy was accompanied by elevations of ALT and AST to more than three times the upper limit of normal in 2% of treated patients, compared with 0.6% of patients given placebo. However, after troglitazone was approved by the FDA, more than 60 cases of hepatic failure that necessitated liver transplantation or resulted in death, or both, were reported out of a user base of about one million persons. The FDA eventually withdrew approval, first for the prescription of troglitazone as monotherapy and later for all indications. In clinical research trials, neither rosiglitazone nor pioglitazone caused AST and ALT elevations in excess of those caused by placebo, but these drugs have not been used enough to provide a guarantee that they will never cause idiosyncratic liver toxicity similar to that caused by troglitazone. Therefore, the FDA has mandated that these drugs be monitored by ALT measurements every 2 months for the first year of use. Neither rosiglitazone nor pioglitazone should be prescribed if the ALT level is greater than 2.5 times the upper limit of normal, and the drugs should be stopped if such levels are reached.

TZD drugs have exhibited effects in addition to glucose lowering that may be beneficial for treating cardiovascular complications.[212-219] Troglitazone and pioglitazone suppress formation of PAI-1, an action that enhances fibrinolysis.[213,214] TZD drugs tend to decrease serum triglyceride levels and increase serum HDL cholesterol levels, but they also increase serum LDL cholesterol levels.[208] In addition, troglitazone has been reported to shift the LDL spectrum from small, dense atherogenic particles to larger, more buoyant, less atherogenic particles.[219] Endothelial function also likely improves.[216,217] A decrease in carotid artery intimal-medial thickness[217]—a marker of atherosclerosis—as well as decreases in vasospastic angina[215] and in recurrence of intimal hyperplasia after coronary angioplasty[218] have been reported in small series of patients treated with troglitazone for only 6 months.[216,217] Whether TZD drugs will decrease rates of cardiovascular events through such actions remains to be seen.

Insulin About 40% of patients with type 2 diabetes mellitus in the United States are estimated to be taking insulin. A small proportion of these patients may have delayed-onset type 1 diabetes mellitus and may offer serologic evidence of beta cell autoimmunity. However, most of these patients represent the end stage of type 2 diabetes mellitus. A small number of such patients, some of whom are even obese, present initially with clear-cut biologic evidence of insulin deficiency. This evidence includes marked recent loss of weight and muscle mass, debilitating fatigue and weakness, severe polyuria and polydypsia, considerable hypertriglyceridemia, ketonuria, and FPG often exceeding 300 mg/dl. These patients should be started on insulin immediately—as in patients with type 1 diabetes mellitus. After usually rapid clinical and biochemical improvement, insulin-dose requirements may decrease progressively. Patients can sometimes be tapered off insulin and be given a trial of an SU drug. This sequence has been reported in certain groups of African Americans.[220,221]

Much more commonly, the need for insulin treatment has arisen because of eventual failure of oral drug therapy, particularly SU drugs. For normal-weight individuals in this situation, it is best to simply switch them to insulin. Some patients may still be managed on a single dose of intermediate- or long-acting insulin (starting dose of NPH, Lente, or glargine of 0.15 to 0.20 U/kg) in

the morning[222] or at bedtime if the FPG is being specifically targeted.[223] The latter is a particularly attractive way to lessen glycemia without stimulating weight gain.[224] Other patients may need intermediate- or long-acting insulin twice a day, usually in a ratio of breakfast dose to bedtime dose of from 1:1 to 2:1. As endogenous postprandial insulin secretion declines further, regular or lispro insulin or insulin aspart [*see* Type 1 Diabetes Mellitus, *above*] must be added before meals. To approach normal glycemia, the doses of rapid, short-acting insulin are best adjusted according to the premeal blood glucose level, the carbohydrate content of the meal, or both. For all practical purposes, some patients with type 2 diabetes mellitus closely resemble patients with type 1 diabetes mellitus in the insulin regimens they require.

For stable patients incapable of accurately mixing different insulins in one syringe because of visual or cognitive impairment, premixed combinations of NPH with regular or lispro insulin are available in varying proportions; these combinations include 70% NPH/30% regular, 50% NPH/50% regular, and 75% NPH/25% lispro. These mixtures all suffer from the inflexibility of neither the dose of NPH nor the dose of the regular or lispro insulin being able to be altered individually. For example, a patient with satisfactory postbreakfast or prelunch blood glucose levels but elevated predinner blood glucose levels would benefit from an increase in the morning NPH insulin dose but not necessarily from an increase in the morning regular or lispro insulin dose. Premixed insulins are not suitable for bedtime use unless an uncommonly large snack is eaten. Despite the above objections, premixed insulins are convenient for patients and for family members who have therapeutic responsibilities.

Whatever insulin regimens are chosen, obese patients with type 2 diabetes mellitus often need large daily doses, which many practitioners are unaccustomed to prescribing. Doses of 1 U/kg body weight are not unusual, and doses of up to 400 U daily have been required to achieve glycemic targets in morbidly obese individuals.[225,226] Concern has been raised in the past that insulin might have atherogenic effects because epidemiologic studies (mostly in nondiabetic individuals) have shown an association between insulin resistance, fasting or postprandial plasma insulin levels, and future risk of cardiovascular disease.[227] In the UKPDS[37] and in the University Group Diabetes Program,[228] exogenous insulin did not increase the rate of myocardial infarction or of cardiovascular death. It can be argued that the insulin doses used in those trials were not very large or that an adverse effect from an atherogenic property of insulin was offset by a beneficial effect resulting from a decrease in glycemia. In any event, there is not enough evidence of cardiovascular danger from exogenous insulin to justify withholding doses necessary to achieve near-normal glycemia. In two randomized clinical trials[37,229] and in a large retrospective study,[192] the incidence of serious hypoglycemic episodes was about two to three events per 100 patient-years. In elderly patients, however, one out of 20 severe hypoglycemic events can be accompanied by such complications as stroke, transient ischemic attack, myocardial infarction, injury, and death.[190] Weight gain—in rare cases, even to degrees that have resulted in sleep apnea—is a major adverse effect of insulin therapy and is one justification for combining insulin with a drug such as metformin, which can restrict weight gain to some extent.[230]

Combination Therapy

Improved understanding of the pathogenesis of hyperglycemia in type 2 diabetes mellitus and longer experience with oral drug monotherapy have greatly increased interest in and popularity of using combinations of oral drugs. In patients with considerable hyperglycemia, none of the current drugs reliably normalize HbA_{1c} when used alone, probably because they act primarily by correcting single abnormalities [*see Figure 15*]. Thus, to reach aggressive therapeutic targets [*see Table 3*], combinations are needed [*see Table 9*]. Moreover, all forms of monotherapy—including insulin used conventionally—fail after a number of years, with the possible exception of TZD drugs, for which long enough experience is still lacking. This need for combination therapy was best shown by the UKPDS experience.[37,188] Combination therapies attack two or more different causes of hyperglycemia simultaneously—for example, reducing insulin resistance in the liver with metformin while increasing insulin secretion with an SU drug[203] or meglitinide.[231] Moreover, as a practical matter, when monotherapy fails after initial success, substituting another drug from a different drug class has not been effective (except for insulin), as has been shown in trials that unsuccessfully attempted to substitute metformin[203] or a TZD drug[232] for an SU drug. By contrast, addition of either metformin[203] or a TZD[232] to an SU drug did lower HbA_{1c} significantly. The combinations of metformin with repaglinide,[231] metformin with a TZD drug,[202] and repaglinide with a TZD drug[233] have also been more effective than any of these agents given alone. Pharmaceutical companies have responded to these considerations by marketing combination pills containing metformin and either glyburide or rosiglitazone for use as initial drug therapy. The single tablet may improve compliance with a two-drug regimen.

α-Glucosidase inhibitors complement the different actions of each of the other drugs, including insulin,[198] and their combinations. All other oral drugs are effective when added to SU drugs, except possibly repaglinide or nateglinide, for which data are still lacking. Metformin and a TZD drug also work in triple combination with an SU drug or meglitinide. Combinations of oral drugs may at least postpone having to switch the patient to progressively more intensive insulin therapy as last recourse. Furthermore, the progressive rise in plasma glucose levels seen in patients on monotherapy in the UKPDS was attributable to declining beta cell function.[234] Thus, pathophysiologically rational combinations of oral drugs, if used much earlier in patients with type 2 diabetes mellitus, might even preserve beta cell function longer than was previously achieved with the initial monotherapy approach.

When adding drugs, particularly to insulin, it is usually wise to start with the lowest dose of the drug being added to the regimen and to increase the dose as though it were being used as monotherapy. Self–blood glucose testing at times of the day appropriate for the added drug should be used as a safety check and a guide to efficacy. For metformin, TZD drugs, SU drugs, and bedtime insulin, the FBG is especially helpful. For repaglinide, nateglinide, α-glucosidase inhibitors, premeal regular insulin, lispro insulin, or insulin aspart, postprandial blood glucose levels are important guides to therapy. Patients should be given blood glucose guidelines for when to call the physician (e.g., when FBG is consistently less than 100 mg/dl).

Combination therapy can be quite expensive, even when it results in a lower insulin requirement.[235] The primary aim should always be to decrease the HbA_{1c}. Reduction in insulin dose, number of injections, or both should be thought of only as a secondary benefit.

MANAGEMENT OF HYPEROSMOLAR HYPERGLYCEMIC NONKETOTIC COMA

Type 2 diabetes mellitus seldom gives rise to DKA unless the patient experiences a severe medical stress. On the other hand,

hyperosmolar hyperglycemic nonketotic coma (HHNC) is a common and feared acute complication characterized by extreme hyperglycemia (> 600 mg/dl) and serum hyperosmolarity (> 320 mOsm/L) but with little or no ketosis.[110,236,237] The CNS effects of extreme hyperosmolarity range from somnolence or confusion to coma but notably can also include focal or generalized seizures as well as focal neurologic deficits that disappear with treatment. The absence of severe ketonemia is usually attributed to enough residual insulin secretion that lipolysis is not as unrestrained as in type 1 diabetes mellitus with DKA. HHNC is marked by extreme dehydration, in which the deficit of free water is prominent and the circulatory volume is often seriously compromised. Thus, hypotension; extemely dry skin and mucous membranes; and gross elevation of hematocrit, urea nitrogen, creatinine, and albumin are frequent. Secondary lactic acidosis[110] is not uncommon, so that the serum bicarbonate level may be low and the anion gap increased. The increased viscosity of the blood predisposes to thrombotic events in the cerebral and coronary artery circulations. However, stroke and myocardial infarction, along with pancreatitis and sepsis, may also precipitate the syndrome. It has also been caused by drugs such as hydrochlorothiazide, phenytoin, and glucocorticoids. Elderly patients living in nursing homes are particularly vulnerable to HHNC because their thirst mechanisms are less sensitive to a rising serum osmolality and because dementia, increasing obtundity, or institutional conditions may combine to reduce water intake to less than urinary and insensible water losses. At presentation, serum sodium level is usually elevated or surprisingly normal in the face of extreme hyperglycemia (i.e., the expected pseudohyponatremia is absent). Whatever the presenting level of serum sodium is, it will rise, sometimes markedly, when glucose levels decline with insulin treatment.

Fluid replacement is the most important component of therapy. Restoration of circulating volume is an urgent first priority. One to two liters of isotonic 0.9% saline is therefore given rapidly initially, followed by 0.45% saline. Later, when plasma glucose levels have declined to 250 to 300 mg/dl, 5% glucose in water or in 0.2% saline is given. Total fluid deficits of as much as 12 L may have to be replaced. Insulin treatment, as for DKA, is started after at least 1 or 2 L of 0.9% saline has been administered. Potassium must be added to intravenous fluids to prevent hypokalemia caused by insulin action. It may take days of fluid replacement, the tonicity of which must be carefully adjusted to achieve a gradual steady decrease in serum osmolality and sodium levels, before central nervous system function returns to normal or at least to baseline. The mortality in HHNC is still high. Infection, especially of the urinary tract, even if only suspected, should be treated with broad-spectrum antibiotics. Papillary necrosis may be seen. Patients with histories of arterial and venous thrombosis can benefit from low-dose prophylactic heparin administration.

Cardiovascular Complications of Diabetes Mellitus

Diabetes as an independent risk factor for cardiovascular disease[238] in women is now well established and is so great that it equalizes the risk of cardiovascular disease in men.[239] The risk of a first myocardial infarction in patients with diabetes is equal to that in nondiabetic individuals who have already suffered such an event.[240] Furthermore, acute and subsequent mortality is greater with diabetic-related myocardial infarctions than with nondiabetic myocardial infarctions.[241] In type 1 diabetes, cardiovascular disease is often a fatal accompaniment of end-stage renal disease (ESRD),[242] although even in patients without ESRD, cardiovascular complications may occur earlier in life than usual. Cardiovascular complications are the most prominent cause of morbidity and the most frequent cause of mortality in type 2 diabetes mellitus.[243,244] Mortality in individuals with diabetes is higher than that in nondiabetic persons of all age and racial groups and both sexes.[244] The decline in heart disease mortality noted in recent years in the United States was less in diabetic persons than in nondiabetic persons, and mortality even increased in women with diabetes.[245] The same common cardiovascular disease risk factors important in nondiabetic individuals are clustered in individuals with type 2 diabetes mellitus[129] as part of the metabolic syndrome. The pathologic picture of atherosclerosis in diabetic persons is similar to that in nondiabetic individuals, and the same processes lead to ischemic events. Thus, in regard to cardiovascular disease, the difference between diabetes and nondiabetes appears largely quantitative, although diabetes remains an independent risk factor even after adjusting for other known risk factors.[238]

Intensive treatment of type 1 or type 2 diabetes mellitus to achieve near-normal glycemia has not been proved to reduce the incidence of cardiovascular complications. The UKPDS reported that intensive treatment with insulin or SU drugs decreased myocardial infarction by 16%, with a *P* value of less than 0.052.[37] Data from several populations,[246-249] and the UKPDS[250] have shown that HbA_{1c} is a risk factor for cardiovascular events and death. Randomized clinical trials are under way to test the question of whether improved glycemic control or particular blood glucose lowering strategies will decrease cardiovascular outcomes in various stages of type 2 diabetes mellitus. Because we still cannot always eliminate whatever risk is incurred from hyperglycemia or insulin resistance per se, we must work assiduously to minimize or negate the adverse effects of hypertension, dyslipidemia, smoking, obesity, and physical inactivity on the cardiovascular system.

HYPERTENSION

Aggressive treatment of hypertension in diabetes mellitus is mandatory for three important reasons: (1) it decreases the risk of cardiovascular disease and mortality,[251] (2) it reduces or at least delays progression of diabetic nephropathy to ESRD,[252-254] and (3) it may decrease the risk of hemorrhage from proliferative retinopathy. The most recent American Diabetes Association guidelines[81] recommend a target blood pressure of 130/80 mm Hg, equivalent to a mean blood pressure of 97 mm Hg (mean blood pressure is easily calculated as one third systolic plus two thirds diastolic). However, even lower blood pressure targets may eventually prove to be advisable.

ACE inhibitors have enjoyed first-choice status in treatment because even in nonhypertensive diabetic patients, these agents decrease albumin excretion and the rate of decline in glomerular filtration rate.[254,255] Similar benefits are offered by angiotensin receptor blockers (ARBs)[256-258]; the ARB losartan also decreased cardiovascular disease events more than did the beta blocker atenolol.[259] Low-dose diuretics (chlorthalidone or hydrochlorothiazide) and beta blockers are also very effective antihypertensive agents in diabetes; they decrease risks of cardiovascular disease events and mortality as well as renal failure. However, the possible adverse effects of these agents on glycemic control and serum lipids must be monitored. Atenolol and captopril were equally effective in lowering blood pressure and reducing cardiovascular events and death in the UKPDS.[260] In the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial (ALLHAT),[261] which included approximately 12,000 diabetic patients, 12.5 to 25 mg of chlorthalidone was at least as effective

as lisinopril and amlodipine in blood pressure reduction and was associated with equal rates of nonfatal myocardial infarction and CHD. However, the cheaper chlorthalidone was superior to the other two agents in preventing heart failure and was superior to the ACE inhibitor lisinopril in preventing strokes and a combined cardiovascular disease end point. The role of calcium channel blockers is somewhat uncertain, as a study with nisoldipine has shown an adverse effect in diabetic patients.[262] At present, calcium channel blockers are probably best used after ACE inhibitors, diuretics, and possibly beta blockers fail to achieve the target blood pressure.[254,255] Central alpha$_2$ agonists (e.g., clonidine), alpha$_1$ antagonists (e.g., prazosin, terazosin, and doxazosin), and combined alpha and beta antagonists (e.g., labetalol) also can be used, although orthostatic hypotension may limit their utility, particularly in patients with autonomic neuropathy.

DYSLIPIDEMIA

Severe hypertriglyceridemia may complicate DKA in type 1 diabetes mellitus, but it clears rapidly with insulin treatment. Serum triglyceride levels are usually elevated—sometimes strikingly so—in uncontrolled type 2 diabetes mellitus, and they are almost invariably accompanied by decreased HDL levels, an atherogenic combination. LDL levels are normal or slightly elevated; however, the LDL component may include a higher proportion of small, dense, more atherogenic particles. Restriction of saturated fat and calories, elimination of excess weight, exercise, and improved glycemic control reduce triglycerides and increase HDL.[262] When these measures are insufficient, gemfibrozil, fenofibrate, or benzafibrate should be prescribed with the purpose of decreasing triglycerides to less than 150 mg/dl and increasing HDL to greater than 35 mg/dl in men and greater than 45 mg/dl in women. For LDL levels greater than 100 mg/dl in patients with or without established coronary artery disease, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins) are the drugs of choice. Simvastatin, pravastatin, and lovastatin have all been shown to decrease cardiovascular events in diabetic patients. Atorvastatin may have the greatest efficacy in lowering LDL and triglyceride levels. If necessary, tolerated doses of bile acid resins may be added. Niacin would be ideal monotherapy because it powerfully lowers triglyceride and LDL levels and increases HDL levels; however, its side effects often discourage compliance, and niacin can also increase blood glucose levels. An extended-release form of niacin, Niaspan, may be more acceptable and effective. If niacin is used as monotherapy or in combination with statins, the daily dosage should not exceed 3 g. Both niacin and statins require the monitoring of serum ALT levels.

MEASURES TO REDUCE RISK OF CARDIOVASCULAR DISEASE

Smoking

Referral to successful smoking-cessation programs and use of oral or dermal nicotine preparations during withdrawal from tobacco should be employed as needed to rid patients of this serious risk factor for cardiovascular disease. Success appears to be directly related to the amount of counseling and support provided by physicians or other professionals.[263]

Aspirin

In the Early Treatment of Diabetic Retinopathy Study, administration of 650 mg of aspirin a day resulted in a statistically significant 17% reduction in the risk of fatal plus nonfatal myocardial infarctions.[264] All-cause mortality and cardiovascular disease mortality tended to decrease, whereas strokes tended to increase, but none of these differences were statistically significant. Preventive use of aspirin is now recommended by the American Diabetes Association for patients who already have cardiovascular disease or who have other risk factors for cardiovascular disease.[265,266]

ACE Inhibitors

In the large, multicentered, randomized 5-year Heart Outcomes Prevention Evaluation (HOPE) trial,[267] ramipril in a single daily dose of 10 mg decreased major cardiovascular events, including myocardial infarction, stroke, heart failure, revascularization procedures, and death by 20% to 32% when compared with placebo. The 9,300 patients were at high risk at entry, and 38% had diabetes. All the diabetic patients benefited from ramipril therapy. Notably, many of the patients were normotensive at baseline, and the beneficial effect of the ACE inhibitor was not thought to be accounted for by the small decrease in average blood pressure. Patients with previous cardiovascular disease also benefited. The data indicated that treatment of 100 patients with ramipril for 4 years would prevent 15 events in seven patients.

Antioxidants

The HOPE study also compared the effect of 400 IU of vitamin E daily with that of placebo. Subjects received no benefit from vitamin E.[268] Although some observational and experimental studies have shown an association between antioxidants and protection from atherosclerosis, there are no firm data on which to base a recommendation for their routine use in diabetes.

MANAGEMENT OF SYMPTOMATIC CORONARY ARTERY DISEASE

Beta blockers, nitrates, and calcium channel blockers can all be used as in nondiabetic individuals, with the proviso that patients treated with insulin or beta cell stimulants should be cautioned about hypoglycemia. When a revascularization procedure has been deemed necessary, coronary artery bypass surgery has been reported to be superior to angioplasty in 5-year survival and recurrent myocardial infarction rates in patients receiving pharmacologic treatment for type 2 diabetes mellitus.[269] In patients without mandatory indications for immediate surgical intervention, such as significant left main coronary artery stenosis, a clinical research trial is currently attempting to determine whether a prompt revascularization procedure is superior to aggressive medical therapy. One study has shown that normalization of blood glucose levels with intravenous insulin during the first 3 days of an acute myocardial infarction, followed by intensive blood glucose control on an outpatient basis for at least 3 months, significantly decreased mortality for up to 3.5 years.[270] If intensive control of glycemia is used in patients with cardiovascular disease, prevention of hypoglycemia should especially be emphasized because, in rare instances, it may precipitate myocardial infarction or stroke.

Prevention and Treatment of Microvascular Complications

As noted above, intensive treatment of both type 1 and type 2 diabetes mellitus, aiming at normoglycemia, reduces the risks of development or progression of diabetic retinopathy, nephropathy, and neuropathy. The earlier such treatment is begun, the greater the benefit.[39] However, once these complications have reached stages of major clinical impact, their response to intensive glycemic control is unknown or at least unproved, with the possible exception of pancreas transplantation.[104] Fortunately,

there are forms of therapy for advanced complications that may ameliorate or prevent their worst manifestations.[271,272]

RETINOPATHY

Laser treatment of high-risk proliferative retinopathy and of macular edema has been demonstrated to preserve vision.[273] For proliferative retinopathy, panretinal scatter photocoagulation is performed to ablate ischemic retina in the periphery capable of producing VEGF. For macular edema, finely focused laser treatment is performed to close visibly leaking perimacular vessels that are demonstrated by fluorescein angiography. The role of the internist and ophthalmologist is to detect retinopathy requiring laser therapy before irreversible damage and loss of vision occur. Although fundus photography is the most sensitive means of detecting early retinopathy, ophthalmologists and even well-trained endocrinologists and internists can detect retinopathy by direct ophthalmoscopy.[274] An examination with the pupil dilated is preferable, but examination in a completely blackened room can be reasonably effective. In type 1 diabetes mellitus, significant retinopathy (beyond microaneurysms) seldom occurs before 5 years' duration, so that regular yearly ophthalmologic examinations do not need to commence until then. By contrast, 20% to 40% of patients with type 2 diabetes mellitus already have detectable retinopathy at the time of clinical onset and diagnosis.[275] Therefore, yearly ophthalmologic examinations should begin at the time of diagnosis. Pregnancy is a recognized risk factor for progression of retinopathy in type 1 diabetes mellitus,[276] and ophthalmologic examinations should be performed at the beginning of pregnancy and thereafter with a frequency dependent on the findings of the first examination. For patients with vitreous hemorrhage that does not clear or significant vitreous scarring and debris, vitrectomy can be performed. Fibroproliferative scars can be excised, and a detached retina can be reattached. The vitreous is replaced with a salt solution. In selected cases, these procedures can restore vision.

NEPHROPATHY

The best preventive approach for diabetic nephropathy in both type 1 and type 2 diabetes mellitus is to maintain a normal blood pressure.[257,277] In normotensive patients with type 1[278] and type 2[279] diabetes mellitus who develop microalbuminuria (30 to 300 mg/day), clinical trials have shown that ACE inhibitor or ARB treatment decreases the rate of progression from microalbuminuria to proteinuria to early renal insufficiency. Maintaining blood glucose near normal with intensive treatment also significantly reduces the risk of diabetic nephropathy.[32,37] If ESRD does develop, a renal transplant is the preferred replacement therapy; home peritoneal dialysis is superior to chronic hemodialysis because the latter is often complicated by vitreous hemorrhage, amputations, and septic episodes. With all forms of therapy for ESRD, mortality is higher in diabetic patients than nondiabetic patients largely because of cardiovascular complications.[280]

NEUROPATHY

Management of diabetic neuropathy is still largely symptomatic[271,272] and often inadequate. Gabapentin[281] in doses of up to 3 g/day has been added to the list of agents that include bedtime tricyclic antidepressants (e.g., nortriptyline), carbamazepine, and topical capsaicin for relief of pain and dysesthesias. Intensive blood glucose control may benefit patients with diabetic amyotrophy and radiculopathy. Prevention of foot ulcers remains very important; patient self-examination of the feet daily and physician-nurse examination at each office visit unequivocally reduce the risk of foot ulcer and amputation.[282] When a foot ulcer does occur, it should be treated aggressively with broad-spectrum antibiotics effective against staphylococci and anaerobes, vigorous debridement as necessary, radiographic examination for osteomyelitis, and sometimes special weight-bearing casts.[283] The use of locally applied growth factors appears promising to reduce healing time.[283] Aggravating effects of ischemia may be alleviated by revascularization of the leg when it is still possible to abort gangrene. Appropriate specialists should be consulted early for achievement of the best outcomes.

Management of autonomic neuropathy is especially challenging. Gastroparesis can benefit from frequent small feedings and either parenteral or liquid oral preparations of metoclopramide[284] or erythromycin.[285] Intermittent intubation to decompress a dilated full stomach may be required to relieve persistent vomiting or painful bloating. A feeding jejunostomy can be considered for intractable cases. Diarrhea sometimes responds to tetracycline antibiotics; clonidine and occasionally somatostatin are effective. Bladder dysfunction may be improved by oral bethanechol and regular timed voiding, but self-catheterization is necessary in severe cases of atony. Use of indwelling catheters should be minimized because of the danger of bacterial or fungal infection. Orthostatic hypotension is benefited by compression stockings, ample sodium intake, and fluorohydrocortisone. The use of midodrine is limited by the risks of excessive hypotension or urinary retention. Male impotence can be satisfactorily treated by penile injection or urethral insertion of alprostadil; by use of a simple vacuum pump; or, increasingly rarely, by implantation of a penile prosthesis. Sildenafil is effective for diabetic impotence,[286] but it may be dangerous in diabetic men with established or unsuspected coronary disease.

Diabetes Mellitus during Pregnancy

Women in their reproductive years with known diabetes of any type should be instructed to inform their physicians when they have decided to have a child. Conception when diabetes control is inadequate markedly increases the risk of major congenital abnormalities. This risk can be reduced to the nondiabetic background rate when control is excellent.[287,288] Therefore, the patient's HbA_{1c} should be brought as close to normal as possible before conception. One recent recommendation is that the average of preprandial and postprandial home blood glucose test results should be less than 126 mg/dl and HbA_{1c} should be brought to at least less than 7.0%.[288] The patients taking oral hypoglycemic agents should be switched to insulin and excellent control established before conception. Nondiabetic women should be screened for gestational diabetes mellitus (GDM) during weeks 24 to 28 of pregnancy by glucose loading.

Throughout pregnancy, normoglycemia (relative to the normal pregnant state) is required to prevent intrauterine death and perinatal morbidity and mortality. Preprandial blood glucose targets during pregnancy are 60 to 90 mg/dl and postprandial targets are less than 120 to 140 mg/dl.[289,290] Most patients with GDM detected by routine screening can be tried on diet treatment for 1 to 2 weeks. In obese women, either 1,500 kcal or 35 kcal/kg prepregnancy weight has been recommended.[287] In 15% to 20% of cases, persistence of FBG of at least 105 mg/dl or 2-hour postprandial values of at least 120 to 140 mg/dl mandates institution of insulin treatment. The blood glucose control targets can then often be achieved with injections of NPH plus regular insulin before breakfast, regular insulin before supper, and NPH

insulin at bedtime. Lispro insulin has not yet been approved for use during pregnancy. Pregnant women with type 1 diabetes mellitus need to continue intensive treatment as described previously. Pregnant women with type 2 diabetes mellitus often respond to insulin regimens as described for GDM. After delivery, insulin requirements disappear almost instantaneously in patients with GDM and may decrease strikingly from those of the third trimester in women with type 1 diabetes mellitus.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. American Diabetes Association: Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 23(suppl 1):S4, 2000

2. DECODE Study Group: Will new diagnostic criteria for diabetes mellitus change phenotype of patients with diabetes? Reanalysis of European epidemiological data. BMJ 317:371, 1998

3. Kjos SL, Buchanan TA: Gestational diabetes mellitus. Current Concepts 341:1749, 1999

4. Barnes AJ, Bloom SR, Goerge K, et al: Ketoacidosis in pancreatectomized man. N Engl J Med 296:1250, 1977

5. Hirsch IB, Paauw DS: Diabetes management in special situations. Curr Therap Diabetes 26:631, 1997

6. Gurwitz JH, Bohn RL, Glynn RJ, et al: Glucocorticoids and the risk for initiation of hypoglycemic therapy. Arch Intern Med 154:97, 1994

7. Matsumoto K, Yamasaki H, Akazawa S, et al: High-dose but not low-dose dexamethasone impairs glucose tolerance by inducing compensatory failure of pancreatic β-cells in normal men. J Clin Endocrinol Metab 81:2621, 1996

8. LaPorte RE, Matsushima M, Chang Y-F: Prevalence and incidence of insulin-dependent diabetes. Diabetes in America, 2nd ed. NIDDK NIH Publication No. 95-1468, 1995

9. Harris MI, Flegal KM, Cowie CC, et al: Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. Diabetes Care 21:518, 1998

10. Kenny SJ, Aubert RE, Geiss LS: Prevalence and incidence of non-insulin-dependent diabetes. Diabetes in America, 2nd ed. NIDDK NIH Publication No. 95-1468, 1995

11. King H, Aubert RE, Herman WH: Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. Diabetes Care 21:1414, 1998

12. Mokdad AH, Ford ES, Bowman BA, et al: Diabetes trends in the U.S.: 1990–1998. Diabetes Care 23:1278, 2000

13. Mokdad AH, Serdula MK, Dietz WH, et al: The spread of the obesity epidemic in the United States, 1991–1998. JAMA 282:1519, 1999

14. Harris MI: Health care and health status and outcomes for patients with type 2 diabetes. Diabetes Care 23:754, 2000

15. Harris MI, Eastman RC, Cowie CC, et al: Racial and ethnic differences in glycemic control of adults with type 2 diabetes. Diabetes Care 22:403, 1999

16. National Task Force on the Prevention and Treatment of Obesity: Overweight, obesity, and health risk. Arch Intern Med 160:898, 2000

17. Expert Panel on the Identification, Evaluation, and Treatment of Overweight in Adults: Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: executive summary. Am J Clin Nutr 68:899, 1998

18. Ruderman N, Chisholm D, Pi-Sunyer X, et al: The metabolically obese, normal-weight individual revisited. Diabetes 47:699, 1998

19. Rewers M, Hamman RF: Risk factors for non-insulin-dependent diabetes. Diabetes in America, 2nd ed. NIDDK NIH Publication No. 95-1468, 1995

20. Genuth S: Hormones of the pancreatic islets. Physiology, 4th ed. Berne RM, Levy MN, Eds. Mosby, St. Louis, 1998, p 822

21. Davis MD: Diabetic retinopathy. Diabetes Care 15:1844, 1992

22. Goldfarb S: Diabetic nephropathy. Primer on Kidney Diseases. Greenberg A, Cheung AK, Falk RJ, et al, Eds. Academic Press, San Diego, 1994, p 112

23. Breyer JA: Diabetic nephropathy in insulin-dependent patients. Am J Kidney Dis 20:533, 1992

24. Ritz E, Stefanski A: Diabetic nephropathy in type II diabetes. Am J Kidney Dis 27:167, 1996

25. Mogensen CE, Christensen CK: Predicting diabetic nephropathy in insulin-dependent patients. N Engl J Med 311:89, 1984

26. Andersen AR, Christiansen JS, Andersen JK, et al: Diabetic nephropathy in type 1 (insulin-dependent) diabetes: an epidemiological study. Diabetologia 25:496, 1983

27. Eaton S, Tesfaye S: Clinical manifestations and measurement of somatic neuropathy. Diabetes Reviews 7:312, 1999

28. Rith-Najarian SJ, Stolusky T, Gohdes DM: Identifying diabetic patients at high risk for lower-extremity amputation in a primary health care setting: a prospective evaluation of simple screening criteria. Diabetes Care 15:1386, 1992

29. Malik RA: Pathology and pathogenesis of diabetic neuropathy. Diabetes Reviews 7:253, 1999

30. Krendel DA, Costigan DA, Hopkins LC: Successful treatment of neuropathies in patients with diabetes mellitus. Arch Neurol 52:1053, 1995

31. Genuth SM: The case for blood glucose control. Advances in Internal Medicine, Vol 40. Schrier RW, Abboud FM, Baxter JD, et al, Eds. Mosby, St. Louis, 1995, p 573

32. The Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 329:977, 1993

33. The Diabetes Control and Complications Trial Research Group: Implementation of treatment protocols in the Diabetes Control and Complications Trial. Diabetes Care 18:361, 1995

34. The Diabetes Control and Complications Trial Research Group: Adverse events and their association with treatment regimens in the Diabetes Control and Complications Trial. Diabetes Care 18:1415, 1995

35. The Diabetes Control and Complications Trial Research Group: Resource utilization and costs of care in the Diabetes Control and Complications Trial. Diabetes Care 18:1468, 1995

36. The Diabetes Control and Complications Trial Research Group: Lifetime benefits and costs of intensive therapy as practiced in the Diabetes Control and Complications Trial. JAMA 276:1409, 1996

37. UK Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 352:837, 1998

38. UK Prospective Diabetes Study (UKPDS) Group: Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet 352:854, 1998

39. The Diabetes Control and Complications Trial Research Group: The relationship of glycemic exposure (HbA_{1c}) to the risk of development and progression of retinopathy in the Diabetes Control and Complications Trial. Diabetes 44:968, 1995

40. Stratton IM, Adler AI, Neil HA, et al: Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 321:405, 2000

41. The Diabetes Control and Complications Trial Research Group: The absence of a glycemic threshold for the development of long-term complications: the perspective of the Diabetes Control and Complications Trial. Diabetes 45:1289, 1996

42. Effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus. Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. JAMA 287:2563, 2002

43. Vlassara H, Bucala R, Striker L: Pathogenic effects of advanced glycosylation: biochemical, biologic, and clinical implications for diabetes and aging. Lab Invest 70:138, 1994

44. Brownlee M: Negative consequences of glycation. Metabolism 49(suppl 1):9, 2000

45. Monnier VM, Bautista O, Kenny D, et al: Skin collagen glycation, glycoxidation, and crosslinking are lower in subjects with long-term intensive versus conventional therapy of type 1 diabetes: relevance of glycated collagen products versus HbA_{1c} as markers of diabetic complications. Diabetes 48:870, 1999

46. Baynes JW, Thorpe SR: Role of oxidative stress in diabetic complications: a new perspective on an old paradigm. Diabetes 48:1, 1999

47. Carrington AL, Litchfield JE: The aldose reductase pathway and nonenzymatic glycation in the pathogenesis of diabetic neuropathy: a critical review for the end of the 20th century. Diabetes Reviews 7:275, 1999

48. Ishii H, Koya D, King GL: Protein kinase C activation and its role in the development of vascular complications in diabetes mellitus. J Mol Med 76:21, 1998

49. Brownlee M: Biochemistry and molecular cell biology of diabetic complications. Nature 414:813, 2001

50. Nishikawa T, Edelstein D, Du XL, et al: Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. Nature 404:787, 2000

51. Quinn M, Angelico MC, Warram JH, et al: Familial factors determine the development of diabetic nephropathy in patients with IDDM. Diabetologia 39:940, 1996

52. The Diabetes Control and Complications Trial Research Group: Clustering of long-term complications in families with diabetes in the Diabetes Control and Complications Trial. Diabetes 46:1829, 1997

53. Thai A-C, Eisenbarth GS: Natural history of IDDM. Diabetes Reviews 1:1, 1993

54. Foulis AK, Liddle CN, Farquharson MA, et al: The histopathology of the pancreas in type 1 (insulin-dependent) diabetes mellitus: a 25-year review of deaths in patients under 20 years of age in the United Kingdom. Diabetologia 29:267, 1986

55. Eisenbarth GS: Type 1 diabetes mellitus: a chronic autoimmune disease. N Engl J Med 314:1360, 1986

56. Faustman D, Eisenbarth G, Daley J, et al: Abnormal T-lymphocyte subsets in type 1 diabetes. Diabetes 38:1462, 1989

57. Rossini AA, Greiner DL, Friedman HP, et al: Immunopathogenesis of diabetes mellitus. Diabetes Reviews 1:43, 1993

58. Littorin B, Sundkvist G, Hagopian W, et al: Islet cell and glutamic acid decarboxylase antibodies present at diagnosis of diabetes predict the need for insulin treatment. Diabetes Care 22:409, 1999

59. Falorni A, Lernmark A: Humoral autoimmunity. Diabetes Mellitus: A Fundamental and Clinical Text, 1st ed. LeRoith D, Taylor SI, Olefsky JM, Eds. Lippincott-Raven Publishers, Philadelphia, 1996, p 298

60. Palmer JP, Asplin CM, Clemons P, et al: Insulin antibodies in insulin-dependent diabetics before insulin treatment. Science 222:1337, 1983

61. Feutren G, Papoz L, Assan R, et al: Cyclosporin increases the rate and length of remissions in insulin-dependent diabetes of recent onset: results of a multicentre double-blind trial. Lancet 2:119, 1986

62. Bougneres PF, Landais P, Boisson C, et al: Limited duration of remission of insulin dependency in children with recent overt type 1 diabetes treated with low-dose cy-

closporin. Diabetes 39:1264, 1990

63. Riley WJ, Maclaren NK, Krischer J, et al: A prospective study of the development of diabetes in relatives of patients with insulin-dependent diabetes. N Engl J Med 323:1167, 1990

64. Tarn AC, Thomas JM, Dean BM, et al: Predicting insulin-dependent diabetes. Lancet 1:845, 1988

65. Ziegler AG, Herskowitz RD, Jackson RA, et al: Predicting type 1 diabetes. Diabetes Care 13:762, 1990

66. Srikanta S, Ganda OP, Gleason RE, et al: Pre-type 1 diabetes: linear loss of beta cell response to intravenous glucose. Diabetes 33:717, 1984

67. Rayfield EJ, Ishimura K: Environmental factors and insulin-dependent diabetes mellitus. Diabetes Metab Rev 3:925, 1987

68. Jones DB, Armstrong NW: Coxsackie virus and diabetes revisited. Nat Med 1:284, 1995

69. Agner T, Damm P, Binder C: Remission in IDDM: prospective study of basal C-peptide and insulin dose in 268 consecutive patients. Diabetes Care 10:164, 1987

70. Heinze E, Beischer W, Keller L, et al: C-peptide secretion during the remission phase of juvenile diabetes. Diabetes 27:670, 1978

71. Shah SC, Malone JI, Simpson NE: A randomized trial of intensive insulin therapy in newly diagnosed insulin-dependent diabetes mellitus. N Engl J Med 320:550, 1989

72. The Diabetes Control and Complications Trial Research Group: Effects of age, duration and treatment of insulin-dependent diabetes mellitus on residual beta-cell function: observations during eligibility testing for the Diabetes Control and Complications Trial (DCCT). J Clin Endocrinol Metab 65:30, 1987

73. Gorsuch AN, Spencer KM, Lister J, et al: Evidence for a long prediabetic period in type I (insulin-dependent) diabetes mellitus. Lancet 2:1363, 1981

74. Karjalainen J, Salmela P, Ilonen J, et al: A comparison of childhood and adult type I diabetes mellitus. N Engl J Med 320:881, 1989

75. Pugliese A: Unraveling the genetics of insulin-dependent type 1A diabetes: the search must go on. Diabetes Reviews 7:39, 1999

76. Baisch JM, Weeks T, Giles R, et al: Analysis of HLA-DQ genotypes and susceptibility in insulin-dependent diabetes mellitus. N Engl J Med 322:1836, 1990

77. Nepom GT: Immunogenetics and IDDM. Diabetes Reviews 1:93, 1993

78. Palmer JP: Predicting IDDM: use of humoral immune markers. Diabetes Reviews 1:104, 1993

79. Genuth SM: Plasma insulin and glucose profiles in normal, obese, and diabetic persons. Ann Intern Med 79:812, 1973

80. Nair KS, Halliday D, Garrow JS: Increased energy expenditure in poorly controlled type I (insulin-dependent) diabetic patients. Diabetologia 27:13, 1984

81. American Diabetes Association: Standards of medical care for patients with diabetes mellitus. Diabetes Care 24:533, 2001

82. The DCCT Research Group: Diabetes Control and Complications Trial (DCCT): results of feasibility study. Diabetes Care 10:1, 1987

83. Goldstein DE, Little RR, Lorenz RA, et al: Tests of glycemia in diabetes. Diabetes Care 18:896, 1995

84. Larsen ML, Horder M, Mogensen EF: Effect of long-term monitoring of glycosylated hemoglobin levels in insulin-dependent diabetes mellitus. N Engl J Med 323:1021, 1990

85. American Diabetes Association: Tests of glycemia in diabetes. Diabetes Care 23(suppl 1):S80, 2000

86. Cagliero E, Levina EV, Nathan DM: Immediate feedback of HbA_{1c} levels improves glycemic control in type 1 and insulin-treated type 2 diabetic patients. Diabetes Care 22:1785, 1999

87. Skyler JS: Insulin treatment. Therapy for Diabetes Mellitus and Related Disorders, 3rd ed. Lebovitz HE, Ed. American Diabetes Association, Alexandria, Virginia, 1998, p 186

88. Hirsch IB: Type 1 diabetes mellitus and the use of flexible insulin regimens. Am Fam Physician 60:2343, 1999

89. Lee WL, Zinman B: From insulin to insulin analogs: progress in the treatment of type 1 diabetes. Diabetes Reviews 6:73, 1998

90. Garg SK, Carmain JA, Braddy KC, et al: Pre-meal insulin analogue insulin lispro vs Humulin® R insulin treatment in young subjects with type 1 diabetes. Diabet Med 13:47, 1996

91. Anderson JH Jr, Brunelle RL, Koivisto VA, et al: Reduction of postprandial hyperglycemia and frequency of hypoglycemia in IDDM patients on insulin-analog treatment. Diabetes 46:265, 1997

92. Del Sindaco P, Ciofetta M, Lalli C, et al: Use of the short-acting insulin analogue lispro in intensive treatment of type 1 diabetes mellitus: importance of appropriate replacement of basal insulin and time-interval injection-meal. Diabet Med 15:592, 1998

93. Raskin P, Guthrie RA, Leiter L, et al: Use of insulin aspart, a fast-acting insulin analog, as the mealtime insulin in the management of patients with type 1 diabetes. Diabetes Care 23:583, 2000

94. Heinemann L, Linkeschova R, Rave K, et al: Time-action profile of the long-acting insulin analog insulin glargine (HOE901) in comparison with those of NPH insulin and placebo. Diabetes Care 23:644, 2000

95. Rosenstock J, Park G, Zimmerman J: Basal insulin glargine (HOE 901) versus NPH insulin in patients with type 1 diabetes on multiple daily insulin regimens. Diabetes Care 23:1137, 2000

96. Ratner RE, Hirsch IB, Neifing JL, et al: Less hypoglycemia with insulin glargine in intensive insulin therapy for type 1 diabetes. Diabetes Care 23:639, 2000

97. Mecklenburg RS: Insulin-pump therapy. Therapy for Diabetes Mellitus and Related Disorders, 3rd ed. Lebovitz HE, Ed. American Diabetes Association, Alexandria, Virginia, 1998, p 204

98. Diabetes Control and Complications Trial Research Group: Implementation of treatment protocols in the Diabetes Control and Complications Trial. Diabetes Care 18:361, 1995

99. Dunn FL, Nathan DM, Scavini M, et al: Long-term therapy of IDDM with an implantable insulin pump. Diabetes Care 20:59, 1997

100. Salzman R, Manson JE, Griffing GT, et al: Intranasal aerolized insulin: mixed-meal studies and long-term use in type I diabetes. N Engl J Med 312:1078, 1985

101. Jendle JH, Karlberg BE: Effects of intrapulmonary insulin in patients with non-insulin-dependent diabetes. Scand J Clin Lab Invest 56:555, 1996

102. Selam JL: Inhaled insulin for the treatment of diabetes: projects and devices. Expert Opin Pharmacother 4:1373, 2003

103. Robertson RP, Holohan TV, Genuth S: Therapeutic controversy: pancreas transplantation for type I diabetes. J Clin Endocrinol Metab 83:1868, 1998

104. Robertson RP, Davis C, Larsen J, et al: Pancreas and islet transplantation for patients with diabetes. Diabetes Care 23:112, 2000

105. Shapiro J, Lakey JRT, Ryan EA, et al: Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. N Engl J Med 343:230, 2000

106. American Diabetes Association: Nutrition recommendations and principles for people with diabetes mellitus. Diabetes Care 23:543, 2000

107. Jenkins DJ, Wolever TM, Taylor RH, et al: Glycemic index of foods: a physiological basis for carbohydrate exchange. Am J Clin Nutr 34:362, 1981

108. Franz MJ, Horton ES, Bantle JP, et al: Nutrition principles for the management of diabetes and related complications. Diabetes Care 17:490, 1994

109. American Diabetes Association: Diabetes mellitus and exercise. Diabetes Care 23(suppl 1):S50, 2000

110. Genuth SM: Diabetic ketoacidosis and hyperglycemic hyperosmolar coma. Current Therapy in Endocrinology and Metabolism, 6th ed. Bardin CW, Ed. Mosby–Year Book, St. Louis, 1997, p 438

111. Foster DW, McGarry JD: The metabolic derangements and treatment of diabetic ketoacidosis. N Engl J Med 309:159, 1983

112. Barrett EJ, DeFronzo RA, Bevilacqua S, et al: Insulin resistance in diabetic ketoacidosis. Diabetes 31:923, 1982

113. Burge MR, Hardy KJ, Schade DS: Short-term fasting is a mechanism for the development of euglycemic ketoacidosis during periods of insulin deficiency. J Clin Endocrinol Metab 76:1192, 1993

114. DeFronzo RA, Matsuda M, Barrett EJ: Diabetic ketoacidosis: a combined metabolic-nephrologic approach to therapy. Diabetes Reviews 2:209, 1994

115. Kitabchi AE, Wall BM: Diabetic ketoacidosis. Med Clin North Am 79:9, 1995

116. Waldhausl W, Kleinberger G, Korn A, et al: Severe hyperglycemia: effects of rehydration on endocrine derangements and blood glucose concentration. Diabetes 28:577, 1979

117. The DCCT Research Group: Epidemiology of severe hypoglycemia in the Diabetes Control and Complications Trial. Am J Med 90:450, 1991

118. Gale EA, Tattersall RB: Unrecognised nocturnal hypoglycaemia in insulin-treated diabetics. Lancet 1:1049, 1979

119. Cryer PE: Hypoglycemia: the limiting factor in the management of IDDM. Diabetes 43:1378, 1994

120. Heller SR, Macdonald IA, Herbert M, et al: Influence of sympathetic nervous system on hypoglycaemic warning symptoms. Lancet 2:359, 1987

121. Mokan M, Mitrakou A, Veneman T, et al: Hypoglycemia unawareness in IDDM. Diabetes Care 17:1397, 1994

122. Dagogo-Jack S, Rattarasarn C, Cryer PE: Reversal of hypoglycemia unawareness, but not defective glucose counterregulation, in IDDM. Diabetes 43:1426, 1994

123. Fanelli C, Pampanelli S, Epifano L, et al: Long-term recovery from unawareness, deficient counterregulation and lack of cognitive dysfunction during hypoglycaemia, following institution of rational, intensive insulin therapy in IDDM. Diabetologia 37:1265, 1994

124. Boyle PJ, Kempers SF, O'Connor AM, et al: Brain glucose uptake and unawareness of hypoglycemia in patients with insulin-dependent diabetes mellitus. N Engl J Med 333:1726, 1995

125. Wahren J, Ekberg K, Fernqvist-Forbes E, et al: Brain substrate utilisation during acute hypoglycaemia. Diabetologia 42:812, 1999

126. The Diabetes Control and Complications Trial Research Group: Effects of intensive diabetes therapy on neuropsychological function in adults in the Diabetes Control and Complications Trial. Ann Intern Med 124:379, 1996

127. Kramer L, Fasching P, Madl C, et al: Previous episodes of hypoglycemic coma are not associated with permanent cognitive brain dysfunction in IDDM patients on intensive insulin treatment. Diabetes 47:1909, 1998

128. Boden G: Pathogenesis of type 2 diabetes: insulin resistance. Endocrinol Metab Clin North Am 30:801, 2001

129. Haffner SM, D'Agostino R, Mykkänen L, et al: Insulin sensitivity in subjects with type 2 diabetes. Diabetes Care 22:562, 1999

130. Haffner SM, Stern MP, Hazuda HP, et al: Increased insulin concentrations in nondiabetic offspring of diabetic parents. N Engl J Med 319:1297, 1988

131. Lillioja S, Mott DM, Howard BV, et al: Impaired glucose tolerance as a disorder of insulin action: longitudinal and cross-sectional studies in Pima Indians. N Engl J Med 318:1217, 1988

132. Saad MF, Knowler WC, Pettitt DJ, et al: Sequential changes in serum insulin concentration during development of non-insulin-dependent diabetes. Lancet 1:1356, 1989

133. Warram JH, Martin BC, Krolewski AS, et al: Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic patients. Ann Intern Med 113:909, 1990

134. O'Meara NM, Sturis J, Van Cauter E, et al: Lack of control by glucose of ultradian insulin secretory oscillations in impaired glucose tolerance and in non-insulin-dependent diabetes mellitus. J Clin Invest 92:262, 1993

135. Porte D Jr: Beta-cells in type II diabetes mellitus. Diabetes 40:166, 1991

136. Hollingdal M, Juhl CB, Pincus SM, et al: Failure of physiological plasma glucose excursions to entrain high-frequency pulsatile insulin secretion in type 2 diabetes. Diabetes 49:1334, 2000

137. Polonsky KS: Evolution of beta-cell dysfunction in impaired glucose tolerance and diabetes. Exp Clin Endocrinol Diabetes 107(suppl 4):S124, 1999

138. Pimenta W, Korytkowski M, Mitrakou A, et al: Pancreatic beta-cell dysfunction as the primary genetic lesion in NIDDM: evidence from studies in normal glucose-tolerant individuals with a first-degree NIDDM relative. JAMA 273:1855, 1995

139. Cavaghan MK, Ehrmann DA, Polonsky KS: Interactions between insulin resistance and insulin secretion in the development of glucose intolerance. J Clin Invest 106:329, 2000

140. Rossetti L, Giaccari A, DeFronzo RA: Glucose toxicity. Diabetes Care 13:610, 1990

141. Cohen M, Crosbie C, Cusworth L, et al: Insulin—not always a life sentence: withdrawal of insulin therapy in non-insulin-dependent diabetes. Diabetes Res 1:31, 1984

142. Saltiel AR: The molecular and physiological basis of insulin resistance: emerging implications for metabolic and cardiovascular diseases. J Clin Invest 106:163, 2000

143. Stern MP: Strategies and prospects for finding insulin resistance genes. J Clin Invest 106:323, 2000

144. Del Prato S, Bonadonna RC, Bonora E, et al: Characterization of cellular defects of insulin action in type 2 (non-insulin-dependent) diabetes mellitus. J Clin Invest 91:484, 1993

145. Pessin JE, Saltiel AR: Signaling pathways in insulin action: molecular targets of insulin resistance. J Clin Invest 106:165, 2000

146. Reaven GM, Chen YD: Role of abnormal free fatty acid metabolism in the development of non-insulin-dependent diabetes mellitus. Am J Med 85:106, 1988

147. Shulman GI: Cellular mechanisms of insulin resistance. J Clin Invest 106:171, 2000

148. Røder ME, Porter D Jr, Schwartz RS, et al: Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. J Clin Endocrinol Metab 83:604, 1998

149. Rachman J, Levy JC, Barrow BA, et al: Relative hyperproinsulinemia of NIDDM persists despite the reduction of hyperglycemia with insulin or sulfonylurea therapy. Diabetes 46:1557, 1997

150. Genuth SM: Hormones of the pancreatic islets. Physiology, 4th ed. Berne RM, Levy MN, Eds. Mosby, St. Louis, 1998, p 822

151. Brunzell JD, Robertson RP, Lerner RL, et al: Relationships between fasting plasma glucose levels and insulin secretion during intravenous glucose tolerance tests. J Clin Endocrinol Metab 42:222, 1976

152. Kahn SE, Andrikopoulos S, Verchere CB: Islet amyloid: a long-recognized but underappreciated pathological feature of type 2 diabetes. Diabetes 48:241, 1999

153. Groop LC, Bottazzo GF, Doniach D: Islet cell antibodies identify latent type 1 diabetes in patients aged 35-75 years at diagnosis. Diabetes 35:237, 1986

154. Turner R, Stratton I, Horton V, et al: UKPDS 25: autoantibodies to islet-cell cytoplasm and glutamic acid decarboxylase for prediction of insulin requirement in type 2 diabetes. Lancet 351:1288, 1997

155. Torn C, Landin-Olsson M, Ostman J, et al: Glutamic acid decarboxylase antibodies (GADA) is the most important factor for prediction of insulin therapy within 3 years in young adult diabetic patients not classified as type 1 diabetes on clinical grounds. Diabetes Metab Res Rev 16:442, 2000

156. Executive Summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation and treatment of high blood cholesterol in adults (Adult Treatment Panel III). JAMA 285:2486, 2001

157. Alberti KG, Zimmet PZ: Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus: provisional report of a WHO consultation. Diabetic Med 15:539, 1998

158. Kahn BB, Flier JS: Obesity and insulin resistance. J Clin Invest 106:473, 2000

159. Gastaldelli A, Baldi S, Pettiti M, et al: Influence of obesity and type 2 diabetes on gluconeogenesis and glucose output in humans. Diabetes 49:1367, 2000

160. Colditz GA, Willett WC, Rotnitzky A, et al: Weight gain as a risk factor for clinical diabetes mellitus in women. Ann Intern Med 122:481, 1995

161. Long SD, O'Brien K, MacDonald KG Jr, et al: Weight loss in severely obese subjects prevents the progression of impaired glucose tolerance to type II diabetes: a longitudinal interventional study. Diabetes Care 17:372, 1994

162. Montague CT, O'Rahilly S: The perils of portliness: causes and consequences of visceral adiposity. Diabetes 49:883, 2000

163. Ford ES, Giles WH, Dietz WH: Prevalence of the metabolic syndrome among US adults: findings from the Third National Health and Nutrition Examination Survey. JAMA 287:356, 2002

164. Alexander CM, Landsman PB, Teutsch SM, et al: NCEP-defined metabolic syndrome, diabetes, and prevalence of coronary heart disease among NHANES III participants age 50 years and older. Diabetes 52:1210, 2003

165. The Diabetes Prevention Program Research Group: The Diabetes Prevention Program: design and methods for a clinical trial in the prevention of type 2 diabetes. Diabetes Care 22:623, 1999

166. Tattersall RB, Pyke DA: Diabetes in identical twins. Lancet 2:1120, 1972

167. Stern MP: Strategies and prospects for finding insulin resistance genes. J Clin Invest 106:323, 2000

168. Habener JF, Stoffers DA: A newly discovered role of transcription factors involved in pancreas development and the pathogenesis of diabetes mellitus. Proc Assoc Am Physicians 110:12, 1998

169. Froguel P, Zouali H, Vionnet N, et al: Familial hyperglycemia due to mutations in glucokinase: definition of a subtype of diabetes mellitus. N Engl J Med 328:697, 1993

170. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. Diabetes Prevention Program Research Group. N Engl J Med 346:393, 2002

171. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. Finnish Diabetes Prevention Study Group. N Engl J Med 344:1343, 2001

172. Pan X-R, Li G-W, Hu Y-H, et al: Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and diabetes study. Diabetes Care 20:537, 1997

173. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomized trial. STOP-NIDDM Trial Research Group. Lancet 359:2072, 2002

174. Buchanan TA, Xiang AH, Peters RK, et al: Preservation of pancreatic β-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk Hispanic women. Diabetes 51:2796, 2002

175. Acarbose treatment and the risk of cardiovascular disease and hypertension in patients with impaired glucose tolerance: the STOP-NIDDM Trial. STOP-NIDDM Trial Research Group. JAMA 290:486, 2003

176. Bosello O, Armellini F, Zamboni M, et al: The benefits of modest weight loss in type II diabetes. Int J Obes Relat Metab Disord 21(suppl 1):S10, 1997

177. American Diabetes Association: Nutrition recommendations and principles for people with diabetes mellitus. Diabetes Care 24(suppl 1):S44, 2001

178. Chandalia M, Garg A, Lutjohann D, et al: Beneficial effects of high dietary fiber intake in patients with type 2 diabetes mellitus. N Engl J Med 342:1392, 2000

179. The National Task Force on the Prevention and Treatment of Obesity: Very low-calorie diets. JAMA 270:967, 1993

180. UK Prospective Diabetes Study Group: UK Prospective Diabetes Study (UKPDS): VIII: study design, progress and performance. Diabetologia 34:877, 1991

181. UKPDS Group: UK Prospective Diabetes Study 7: response of fasting plasma glucose to diet therapy in newly presenting type II diabetic patients. Metabolism 39:905, 1990

182. Kelley DE, Jneidi M: Orlistat in the treatment of type 2 diabetes mellitus. Expert Opin Pharmacother 3:599, 2002

183. Porte D Jr, Seeley RJ, Woods SC, et al: Obesity, diabetes and the central nervous system. Diabetologia 41:863, 1998

184. MacDonald KG, Long SD, Swanson MS, et al: The gastric bypass operation reduces the progression and mortality of non-insulin-dependent diabetes mellitus. J Gastrointest Surg 1:213, 1997

185. O'Brien PE, Brown WA, Smith A, et al: Prospective study of a laparoscopically placed, adjustable gastric band in the treatment of morbid obesity. Br J Surg 85:113, 1999

186. Zinman B, Ruderman N, Campaigne BN, et al: Physical activity/exercise and diabetes mellitus. Diabetes Care 26(suppl 1):S73, 2003

187. Wei M, Gibbons LW, Kampert JB, et al: Low cardiorespiratory fitness and physical inactivity as predictors of mortality in men with type 2 diabetes. Ann Intern Med 132:605, 2000

188. Turner RC, Cull CA, Frighi V, et al: Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus. JAMA 281:2005, 1999

189. Riddle M: The 2 defects of type 2 diabetes: combining drugs to treat both insulin deficiency and insulin resistance. Am J Med 108(suppl 6a):S1, 2000

190. Ashcroft FM: Mechanisms of the Glycaemic Effects of Sulfonylureas. Horm Metab Res 28:456, 1996

191. Shorr RI, Ray WA, Daugherty JR, et al: Individual sulfonylureas and serious hypoglycemia in older people. J Am Geriatr Soc 44:751, 1996

192. Shorr RI, Ray WA, Daugherty JR, et al: Incidence and risk factors for serious hypoglycemia in older persons using insulin or sulfonylureas. Arch Intern Med 157:1681, 1997

193. University Group Diabetes Program: A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. I: Design, methods and baseline results. II: Mortality results. Diabetes 19:747, 1970

194. Brady PA, Terzic A: The sulfonylurea controversy: more questions from the heart. J Am Coll Cardiol 31:950, 1998

195. Schade DS, Jovanovic L, Schneider J: A placebo-controlled, randomized study of glimepiride in patients with type 2 diabetes mellitus for whom diet therapy is unsuccessful. J Clin Pharmacol 38:636, 1998

196. Lebovitz HE: Insulin secretagogues: old and new. Diabetes Reviews 7:139, 1999

197. Fuhlendorff J, Rorsman P, Kofod H, et al: Stimulation of insulin release by repaglinide and glibenclamide involves both common and distinct processes. Diabetes 47:345, 1998

198. Lebovitz HE: Alpha-glucosidase inhibitors. Current Therapies for Diabetes 26:539, 1997

199. Coniff RF, Shapiro JA, Robbins D, et al: Reduction of glycosylated hemoglobin and postprandial hyperglycemia by acarbose in patients with NIDDM: a placebo-controlled dose-comparison study. Diabetes Care 18:817, 1995

200. Kirpichnikov D, McFarlane SI, Sowers JR: Metformin: an update. Ann Intern Med 137:25, 2002

201. Stumvoll N, Nurjhan N, Perriello G, et al: Metabolic effects of metformin in non-insulin-dependent diabetes mellitus. N Engl J Med 333:550, 1995

202. Inzucchi SE, Maggs DG, Spollett GR, et al: Efficacy and metabolic effects of metformin and troglitazone in type II diabetes mellitus. N Engl J Med 338:867, 1998

203. Multicenter Metformin Study Group: Efficacy of metformin in patients with non-insulin-dependent diabetes mellitus. N Engl J Med 333:541, 1995

204. Lebovitz HE: Effects of oral antihyperglycemic agents in modifying macrovascular risk factors in type 2 diabetes. Diabetes Care 22(suppl 3):S41, 1999

205. Garber AJ, Duncan TG, Goodman AM, et al: Efficacy of metformin in type II diabetes: results of a double-blind, placebo-controlled, dose-response trial. Am J Med 102:491, 1997

206. Lalau JD, Lacroix C, Compagnon P, et al: Role of metformin accumulation in metformin-associated lactic acidosis. Diabetes Care 18:779, 1995

207. Lalau JD, Westeel PF, Debussche X, et al: Bicarbonate haemodialysis: an adequate treatment for lactic acidosis in diabetics treated by metformin. Intensive Care Med 13:383, 1987

208. Henry RR: Thiazolidinediones. Current Therapies for Diabetes 26:553, 1997

209. Maggs DG, Buchanan TA, Burant CF, et al: Metabolic effects of troglitazone monotherapy in type 2 diabetes mellitus: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 128:176, 1998

210. Olefsky JM: Treatment of insulin resistance with peroxisome proliferator-activated receptor γ agonists. J Clin Invest 106:467, 2000

211. Akazawa S, Sun F, Ito M, et al: Efficacy of troglitazone on body fat distribution in type 2 diabetes. Diabetes Care 23:1067, 2000

212. Minamikawa J, Tanaka S, Yamauchi M, et al: Potent inhibitory effect of troglitazone on carotid arterial wall thickness in type 2 diabetes. J Clin Endocrinol Metab 83:1818, 1998

213. Kato K, Satoh H, Endo Y, et al: Thiazolidinediones down-regulate plasminogen activator inhibitor type 1 expression in human vascular endothelial cells: a possible role for PPARgamma in endothelial function. Biochem Biophys Res Commun 258:431, 1999

214. Yamakawa K, Hosoi M, Fukumoto S, et al: Pioglitazone inhibits plasminogen activator inhibitor–1 expression in human vascular smooth muscle cell (abstr). Diabetes 47:A366, 1998

215. Murakami T, Mizuno S, Ohsato K, et al: Effects of troglitazone on frequency of coronary vasospastic-induced angina pectoris in patients with diabetes mellitus. Am J Cardiol 84:92, 1999

216. Avena R, Mitchell ME, Nylen ES, et al: Insulin action enhancement normalizes brachial artery vasoactivity in patients with peripheral vascular disease and occult diabetes. J Vasc Surg 28:1024, 1998

217. Kotchen TA, Zhang HY, Reddy S, et al: Effect of pioglitazone on vascular reactivity in vivo and in vitro. Am J Physiol 270:R660, 1996

218. Takagi T, Yoshida K, Akasaka T, et al: Troglitazone reduces intimal hyperplasia after coronary stent implantation in patients with type 2 diabetes mellitus: a serial intravascular ultrasound study (abstr). J Am Coll Cardiol 33(suppl A):100A, 1999

219. Tack CJJ, Smits P, Demacker PNM, et al: Troglitazone decreases the proportion of small, dense LDL and increases the resistance of LDL to oxidation in obese subjects. Diabetes Care 21:796, 1998

220. Winter WE, Maclaren NK, Riley WJ, et al: Maturity-onset diabetes of youth in black Americans. N Engl J Med 316:285, 1987

221. Banerji MA, Lebovitz HE: Insulin-sensitive and insulin-resistant variants in NIDDM. Diabetes 38:784, 1989

222. Genuth S: Insulin in non–insulin-dependent patients. Adv Endocrinol Metab 3:89, 1992

223. Riddle MC: Evening insulin strategy. Diabetes Care 13:676, 1990

224. Yki-Jarvinen H, Kauppila M, Kujansuu E, et al: Comparison of insulin regimens in patients with non-insulin-dependent diabetes mellitus. N Engl J Med 327:1426, 1992

225. Genuth SM, Martin P: Control of hyperglycemia in adult diabetics by pulsed insulin delivery. Diabetes 26:571, 1977

226. Henry RR, Gumbiner B, Ditzler T, et al: Intensive conventional insulin therapy for type II diabetes. Diabetes Care 16:21, 1993

227. Elliott TG, Viberti G: Relationship between insulin resistance and coronary heart disease in diabetes mellitus and the general population: a critical appraisal. Baillieres Clin Endocrinol Metab 7:1079, 1993

228. Genuth S: Exogenous insulin administration and cardiovascular risk in non–insulin-dependent and insulin-dependent diabetes mellitus. Ann Intern Med 124:104, 1996

229. Abraira C, Colwell JA, Nuttall FQ, et al: Veterans Affairs Cooperative Study on Glycemic Control and Complications in Type II Diabetes (VA CSDM): results of the feasibility study. Diabetes Care 18:1113, 1995

230. Giugliano D, Quatraro A, Consoli G, et al: Metformin for obese, insulin-treated diabetic patients: improvement in glycaemic control and reduction of metabolic risk factors. Eur J Clin Pharmacol 44:107, 1993

231. Moses R, Slobodniuk R, Boyages S, et al: Effect of repaglinide addition to metformin monotherapy on glycemic control in patients with type 2 diabetes. Diabetes Care 22:119, 1999

232. Horton ES, Whitehouse F, Ghazzi MN, et al: Troglitazone in combination with sulfonylurea restores glycemic control in patients with type 2 diabetes. Diabetes Care 21:1462, 1998

233. Raskin P, Jovanovic L, Berger S, et al: Repaglinide/troglitazone combination therapy: improved glycemic control in type 2 diabetes. Diabetes Care 23:979, 2000

234. U.K. Prospective Diabetes Study Group: U.K. Prospective Diabetes Study 16 – overview of 6 years' therapy of type II diabetes: a progressive disease. Diabetes 44:1249, 1995

235. Casner PR: Insulin-glyburide combination therapy for non-insulin-dependent diabetes mellitus: a long-term double blind, placebo-controlled trial. Clin Pharmacol Ther 44:594, 1988

236. Cruz-Caudillo JC, Sabatini S: Diabetic hyperosmolar syndrome. Nephron 69:201, 1995

237. Lorber D: Nonketotic hypertonicity in diabetes mellitus. Med Clin North Am 79:39, 1995

238. Stamler J, Vaccaro O, Neaton JD, et al: Diabetes, other risk factors, and 12-yr cardiovascular mortality for men screened in the Multiple Risk Factor Intervention Trial. Diabetes Care 16:434, 1993

239. Barrett-Connor EL, Cohn BA, Wingard DL, et al: Why is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? The Rancho Bernardo Study. JAMA 265:627, 1991

240. Haffner SM, Lehto S, Rönnemaa T, et al: Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med 339:229, 1998

241. Singer DE, Moulton AW, Nathan DM: Diabetic myocardial infarction: Interaction of diabetes with other preinfarction risk factors. Diabetes 38:350, 1989

242. Jensen T, Borch-Johnsen K, Kofoed-Enevoldsen A, et al: Coronary heart disease in young type 1 (insulin-dependent) diabetic patients with and without diabetic nephropathy: incidence and risk factors. Diabetologia 30:144, 1987

243. Grundy SM, Benjamin IJ, Burke GL, et al: Diabetes and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. Circulation 100:1134, 1999

244. Gu K, Cowie CC, Harris MI: Mortality in adults with and without diabetes in a national cohort of the U.S. population, 1971–1993. Diabetes Care 21:1138, 1998

245. Gu K, Cowie CC, Harris MI: Diabetes and decline in heart disease mortality in US adults. JAMA 281:1291, 1999

246. Moss SE, Klein R, Klein BEK, et al: The association of glycemia and cause-specific mortality in a diabetic population. Arch Intern Med 154:2473, 1994

247. Kuusisto J, Mykkänen L, Pyörälä K, et al: NIDDM and its metabolic control predict coronary heart disease in elderly subjects. Diabetes 43:960, 1994

248. Andersson DKG, Svärdsudd K: Long-term glycemic control relates to mortality in type II diabetes. Diabetes Care 18:1534, 1995

249. Wei M, Gaskill SP, Haffner SM, et al: Effects of diabetes and level of glycemia on all-cause and cardiovascular mortality: the San Antonio Heart Study. Diabetes Care 21:1167, 1998

250. Stratton IM, Adler AI, Neil HAW, et al: Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 321:405, 2000

251. UK Prospective Diabetes Study Group: Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. BMJ 317:703, 1998

252. Nielson FS, Rossing P, Gall MA, et al: Long-term effect of lisinopril and atenolol on kidney function in hypertensive NIDDM subjects with diabetic nephropathy. Diabetes 46:1182, 1997

253. American Diabetes Association: Diabetic nephropathy. Diabetes Care 23:569, 2000

254. Pahor M, Psaty BM, Alderman MH, et al: Therapeutic benefits of ACE inhibitors and other antihypertensive drugs in patients with type 2 diabetes. Diabetes Care 23:888, 2000

255. Barkis GL, Williams M, Dworkin L, et al: Preserving renal function in adults with hypertension with diabetes: a consensus approach. Am J Kidney Dis 36:646, 2000

256. Effects of losartan on renal and cardiovascular outcome in patients with type 2 diabetes and nephropathy. RENAAL Study Investigators. N Engl J Med 345:861, 2001

257. The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. Irbesartan in Patients with Type 2 Diabetes and Microalbuminuria Study Group. N Engl J Med 345:870, 2001

258. Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. Collaborative Study Group. N Engl J Med 345:851, 2001

259. Cardiovascular morbidity and mortality in patients with diabetes in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomized trial against atenolol. LIFE Study Group. Lancet 359:1004, 2002

260. UK Prospective Diabetes Study Group: Efficacy of atenolol and captopril in reducing risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. BMJ 317:713, 1998

261. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Collaborative Research Group. JAMA 288:2981, 2002

262. Estacio RO, Jeffers BW, Hiatt WR, et al: The effect of nisoldipine as compared with enalapril on cardiovascular outcomes in patients with non-insulin-dependent diabetes and hypertension. N Engl J Med 338:645, 1998

263. Haire-Joshu D, Glasgow RE, Tibbs TL: Smoking and diabetes. Diabetes Care 22:1887, 1999

264. ETDRS Investigators: Aspirin effects on mortality and morbidity in patients with diabetes mellitus: Early Treatment Diabetic Retinopathy Study report 14. JAMA 268:1292, 1992

265. American Diabetes Association: Aspirin therapy in diabetes. Diabetes Care 24(suppl 1):S62, 2001

266. Colwell JA: Aspirin therapy in diabetes. Diabetes Care 20:1767, 1997

267. The Heart Outcomes Prevention Evaluation Study Investigators: Effects of an angiotensin-converting–enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. N Engl J Med 342:145, 2000

268. The Heart Outcomes Prevention Evaluation Study Investigators: Vitamin E supplementation and cardiovascular events in high-risk patients. N Engl J Med 342:154, 2000

269. The Bypass Angioplasty Revascularization Investigation (BARI) Investigators: Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. N Engl J Med 335:217, 1996

270. Malmberg K, for the DIGAMI (Diabetes Mellitus, Insulin Glucose Infusion in Acute Myocardial Infarction) Study Group: Prospective randomised study of intensive insulin treatment on long term survival after acute myocardial infarction in patients with diabetes mellitus. BMJ 314:1512, 1997

271. Nathan DM: Long-term complications of diabetes mellitus. N Engl J Med 328:1676, 1993

272. Clark CM Jr, Lee DA: Prevention and treatment of the complications of diabetes mellitus. N Engl J Med 332:1210, 1995

273. Ferris FL III, Davis MD, Aiello LM: Treatment of diabetic retinopathy. Drug Therapy 341:667, 1999

274. Nathan DM, Fogel HA, Godine JE, et al: Role of diabetologist in evaluating diabetic retinopathy. Diabetes Care 14:26, 1991

275. Harris MI, Klein R, Welborn TA, et al: Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis. Diabetes Care 15:815, 1992

276. Effect of pregnancy on microvascular complications in the diabetes control and complications trial. The Diabetes Control and Complications Trial Research Group. Diabetes Care 23:1084, 2000

277. Cooper ME: Pathogenesis, prevention, and treatment of diabetic nephropathy. Lancet 352:213, 1998

278. Lewis EJ, Hunsicker LG, Bain RP, et al: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. N Engl J Med 329:1456, 1993

279. Ravid M, Savin H, Jutrin I, et al: Long-term stabilizing effect of angiotensin-converting enzyme inhibition on plasma creatinine and on proteinuria in normotensive type II diabetic patients. Ann Intern Med 118:577, 1993

280. Foley RN, Culleton BF, Parfrey PS, et al: Cardiac disease in diabetic end-stage renal disease. Diabetologia 40:1307, 1997

281. Backonja M, Beydoun A, Edwards KR, et al: Gabapentin for the symptomatic treatment of painful neuropathy in patients with diabetes mellitus. JAMA 280:1831, 1998

282. Litzelman DK, Slemenda CW, Langefeld CD, et al: Reduction of lower extremity clinical abnormalities in patients with non–insulin-dependent diabetes mellitus. Ann Intern Med 119:36, 1993

283. Mason J, O'Keeffe C, Hutchinson A, et al: A systematic review of foot ulcer in patients with type 2 diabetes mellitus. II: Treatment. Diabetic Medicine 16:889, 1999

284. McCallum RW, Ricci DA, Rakatansky H, et al: A multicenter placebo-controlled clinical trial of oral metoclopramide in diabetic gastroparesis. Diabetes Care 6:463, 1983

285. Janssens J, Peeters TL, Vantrappen G, et al: Improvement of gastric emptying in diabetic gastroparesis by erythromycin: preliminary studies. N Engl J Med 322:1028, 1990

286. Sildenafil Diabetes Study Group: Sildenafil for treatment of erectile dysfunction in men with diabetes: a randomized controlled trial. JAMA 281:421, 1999

287. Kühl C: New approaches for the treatment of pregnant diabetic women. Diabetes Reviews 3:621, 1995

288. Langer O: Is normoglycemia the correct threshold to prevent complications in the pregnant diabetic patient? Diabetes Reviews 4:2, 1996

289. de Veciana M, Major CA, Morgan MA, et al: Postprandial versus preprandial blood glucose monitoring in women with gestational diabetes mellitus requiring insulin therapy. N Engl J Med 333:1237, 1995

290. Kjos SL, Buchanan TA: Gestational diabetes mellitus. N Engl J Med 341:1749, 1999

Acknowledgments

I am grateful to the previous author of this section, Dr. David Nathan, for providing such an excellent template and for generously permitting me to retain certain sections that required only minor updating. I also wish to thank Eileen Campbell and Molly Genuth for skilled assistance in preparing the manuscript.

Figures 1, 4, 14, and 15 Seward Hung.

I RENAL FUNCTION AND DISORDERS OF WATER AND SODIUM BALANCE

Richard H. Sterns, m.d., f.a.c.p.

Overview of Body Fluid Homeostasis

Life takes place in an aqueous solution. Cells, the blood bringing nutrients and oxygen to them, and the interstitial fluid bathing them are all mostly water. Each day, water and salt are lost and replaced. To maintain stability of the internal milieu, body fluids are processed by the kidney, guided by intricate physiologic control systems that regulate fluid volume and composition.

Distribution and Composition of Body Water

Water accounts for approximately half of an adult human's body weight. Because fat contains little water, individuals with more body fat have less body water. On average, total body water constitutes 60% of lean body weight in young men, 50% in young women and older men, and 45% in older women. Two thirds of body water is intracellular, and the remainder is contained in the extracellular fluid compartment, which includes intravascular (plasma) and interstitial fluid. Small amounts of water are also contained in bone, dense connective tissue, digestive secretions, and cerebrospinal fluid.[1]

Extracellular solutes are predominantly sodium salts (primarily a mixture of NaCl and $NaHCO_3$). Thus, extracellular fluid can be thought of as saltwater. Except for protein (present at a higher concentration in plasma [approximately 1 mmol/L] than in interstitial fluid), the compositions of the intravascular and interstitial subdivisions of the extracellular fluid compartment are similar.

The sodium-potassium adenosine triphosphatase (Na^+,K^+-ATPase) pump on cell membranes keeps intracellular sodium at low levels. Potassium, the dominant intracellular cation, is electrically balanced, in large part, by anionic charges on impermeant macromolecules. Stability in the number of intracellular anionic charges makes the total solute content of cells much less variable than that of the extracellular fluid.

Osmolality

Extracellular and intracellular fluids contain different types of solutes, but the concentrations of solutes inside and outside of cells are equal. Concentration differences exist only transiently because they create an extremely strong force for water movement across cell membranes. Osmotic pressure moves water rapidly to the fluid compartment with the higher solute concentration until concentrations once again become equal. The osmotic pressure responsible for water movement across cell membranes depends on the total number of solute particles (osmoles) dissolved in solution, a property known as osmolality.[2] Osmolality is usually expressed as milliosmoles of solute per kilogram of solvent (mOsm/kg), but it can be thought of more simply as the number of millimoles of solute particles per liter of solution. A solute particle's contribution to osmolality is independent of its charge and molecular size. Ionic substances such as sodium chloride that dissociate in solution contribute more than one osmotically active particle. Sodium salts, glucose, and urea, commonly measured as blood urea nitrogen (BUN), are responsible for most of the solute particles normally present in extracellular fluid. Plasma osmolality can be measured directly with an osmometer or can be estimated with reasonable accuracy from the concentrations of the major extracellular solutes, as follows:

$$P_{osm} \cong 2 \times \text{plasma } [Na^+] + \frac{[\text{glucose}]}{18} + \frac{\text{BUN}}{2.8}$$

The multiple of 2 reflects the anions accompanying sodium ions, and 18 and 2.8 are the corrections required to convert glucose and urea nitrogen concentration from mg/dl (the units used by most laboratories in the United States) to mmol/L. Exogenous solutes (e.g., ethanol, methanol, ethylene glycol, glycine, mannitol) are measured by osmometers but are not included in the formula shown above. A discrepancy between the measured and the calculated plasma osmolality values (an osmolar gap) is useful clinically as a way to recognize the presence of an exogenous solute.[2]

Fluid Movement between Body Fluid Compartments

The intravascular and interstitial subdivisions of the extracellular fluid compartment are separated by capillary walls that are freely permeable to small extracellular solutes but relatively impermeable to plasma proteins. Protein-free saltwater continuously moves across the capillary endothelial barrier by filtration, driven by a hydrostatic pressure gradient (generated by contractions of the heart), which forces fluid from the capillary into the interstitium, and an oncotic pressure gradient (the consequence of the osmotic force created by intravascular protein), which draws interstitial fluid into capillaries [*see Figure 1*]. These so-called Starling forces, which regulate the disposition of fluid within the extracellular compartment, determine how much of the extracellular saltwater is contained in intravascular plasma and how much is in interstitial fluid [*see* Disorder of Saltwater Excess: Edematous States, *below*]. Sodium salts, urea, glucose, and other small extracellular solutes freely cross the capillary wall, achieving similar concentrations in the interstitial fluid and plasma. Thus, changes in plasma osmolality do not influence water movement between the intravascular and the interstitial fluid compartments.

Fluid movement between the extracellular and the intracellular fluid compartments is unaffected by Starling forces. Rather, transcellular water movement is driven by osmotic forces, a function of the concentration of solutes in the extracellular fluid. A decrease in extracellular solute concentration (hypotonicity) drives water into cells, causing cell swelling; an increase in the concentration of exclusively extracellular solutes (hypertonicity) draws fluid out of cells, dehydrating them.

Not all solutes contribute to the tonicity of extracellular fluid.[2] Permeant solutes such as urea and ethanol readily cross cell membranes, achieving equal concentrations in the extracellular and intracellular fluid compartments without driving transcellular water movement and without affecting cell volume. Such solutes, which increase plasma osmolality without altering plasma tonicity, are sometimes called ineffective osmoles. Impermeant solutes are excluded from cell water either by active trans-

port (e.g., sodium ions) or because the cell membrane is impermeable to them (e.g., mannitol). Such solutes, which cause both hyperosmolality and hypertonicity, are sometimes called effective osmoles.

Body Fluid Tonicity and the Plasma Sodium Concentration

Normally, sodium salts are the major effective osmoles in extracellular fluid, and potassium salts are the major effective osmoles in cells. Given that effective osmolality is equal in all fluid compartments, body fluid tonicity can be described by the following equation:

$$\text{Tonicity} \cong$$

$$2 \times \text{plasma } [Na^+] \cong \frac{2\,(\text{exchangeable } Na^+ + \text{exchangeable } K^+)}{\text{total body water}}$$

therefore,

$$\text{Plasma } [Na^+] \cong \frac{\text{exchangeable } Na^+ + \text{exchangeable } K^+}{\text{total body water}}$$

(Only the exchangeable fractions of sodium and potassium are included in the equation, because one third of body sodium is bound to bone and is osmotically inactive.)

Thus, the plasma sodium concentration is, in effect, a measure of the concentration of tonicity of all body fluids.[3] In the absence of an osmolar gap, the plasma sodium concentration is a more valid measure of body fluid tonicity than is plasma osmolality (which includes the ineffective osmole urea). With a few exceptions, most notably hyperglycemia, a low plasma sodium concentration indicates hypotonicity and cell swelling, whereas a high plasma sodium concentration indicates hypertonicity and cellular dehydration.

RENAL PROCESSING OF BODY FLUIDS

Glomerular Filtration

Approximately 170 L of extracellular saltwater containing over 25,000 mmol of sodium are filtered by the glomerulus each day. Although glomerular hydrostatic pressure is considerably higher than the pressure of other capillary beds, the Starling forces that control fluid movement between intravascular and interstitial fluid also drive glomerular filtration. The glomerular filtrate contains the same concentrations of sodium and other solutes as interstitial fluid and is nearly protein free.

Tubular Reabsorption

On a conventional diet, all but 2 L of filtered fluid and all but 175 mmol of filtered sodium is reabsorbed by the renal tubules. Regulation of tubular reabsorption of salt and water is the key to renal regulation of body fluid balance.[4]

At the end of the proximal tubule, the remaining filtrate has the same sodium concentration as plasma; as the filtrate passes through downstream tubular segments, it undergoes major changes in composition. In these more distal segments, sodium and water reabsorption are uncoupled; salt can be reabsorbed without water, and water can be reabsorbed without salt. Thus, depending on conditions, the sodium concentration of the final urine can vary from less than 1 mEq/L to nearly 300 mEq/L, and urine osmolality can vary from one sixth (50 mOsm/kg) to four times (1,200 mOsm/kg) that of plasma.

The process of sodium reabsorption is mediated by carriers or channels embedded in the tubular cell's luminal and basolateral (blood side) membranes. In each nephron segment, sodium reabsorption is powered by the Na^+,K^+-ATPase, which is located on the blood side of the tubular cell [*see Figure 1*]. This sodium pump exports sodium from the cell, lowering the intracellular sodium concentration. The lowered concentration of cellular sodium creates an electrochemical gradient driving sodium from the tubular lumen into the cell. Tubular segments at various regions of the nephron utilize different luminal mechanisms for sodium reabsorption [*see Table 1 and Figure 2*]. Luminal exchangers, cotransporters, and ion channels along the nephron are subject to physiologic control, and they can be inhibited pharmacologically by specific diuretic agents.[5] Mutations in these transport proteins are responsible for well-defined clinical disorders.[6]

The luminal membrane of the collecting duct is impermeable to water in the absence of arginine vasopressin, an antidiuretic hormone (ADH). Thus, when plasma ADH levels are low, this segment progressively reduces the osmolality and sodium concentration of the final urine and permits the excretion of large volumes (as much as 20 L daily) of dilute urine. In the presence of ADH, water channels—called aquaporins—are inserted in the luminal membrane of the distal tubule and collecting duct.[7,8] When plasma levels of ADH are high, water is attracted osmotically from the tubular lumen to the hypertonic medullary interstitium, permitting excretion of a small volume (as little as 0.5 L daily) of concentrated urine.

REGULATION OF BODY FLUID VOLUMES

Saltwater (isotonic saline) is confined to the extracellular space. Accumulation of saltwater expands extracellular volume; loss of saltwater causes volume depletion. In either case, changes in saltwater balance do not alter the plasma sodium concentration or cell volume. By contrast, so-called electrolyte-free water, or pure water, is distributed throughout body fluids, affecting both extracellular and intracellular fluid compartments. Because only one third of body water is extracellular, electrolyte-free water has only one third the impact on extracellular

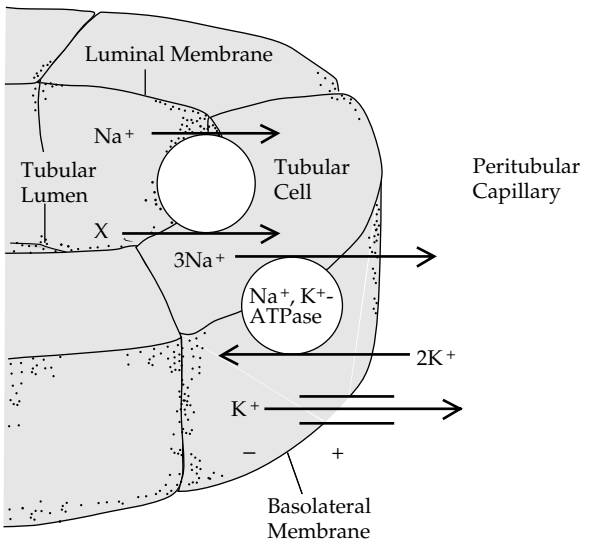


Figure 1 **Sodium reabsorption by the renal tubules is driven by the Na^+, K^+-ATPase pump in the basolateral (blood side) membrane, which maintains a low sodium concentration within the tubular cell. Filtered sodium in the tubular lumen enters the cell down its concentration gradient via a transmembrane carrier (which may also transport another solute, X) or via a channel in the luminal membrane [*see Figure 2*].**

Table 1 Sodium Transport in Different Nephron Segments

Nephron Segment	*Glomerular Filtrate Reabsorbed*	*Mechanism of Luminal Sodium Entry*	*Physiologic Regulation*	*Diuretic Site of Action*
Proximal tubule	60%–70%	Na^+-H^+ exchange; cotransport with glucose and other organic solutes	Angiotensin II; renal nerves; peritubular Starling forces	Carbonic anhydrase inhibitors (e.g., acetazolamide)
Loop of Henle	20%–25%	Na^+-K^+-$2Cl^-$ cotransport	Flow dependent; peritubular Starling forces	Loop diuretics (e.g., furosemide, bumetanide, ethacrynic acid)
Distal tubule	5%	Na^+-Cl^- cotransport	Flow dependent	Thiazide diuretics
Collecting tubule	4%	Na^+ channels	Aldosterone; atrial natriuretic factor	Potassium-sparing diuretics (e.g., amiloride, triamterene, spironolactone)

volume that saltwater has; however, unlike saltwater balance, electrolyte-free water balance has a major impact on the plasma sodium concentration, body fluid tonicity, and cell volume.

Extracellular and intracellular fluid volumes are maintained by separate but interacting control systems [*see Table 2*]; the extracellular system primarily regulates urinary sodium excretion, whereas the intracellular system regulates the intake and excretion of water. Extracellular fluid volume maintains a proper degree of vascular fullness, a variable that is sensed by atrial stretch receptors and arterial baroreceptors. Intracellular volume is regulated by hypothalamic osmoreceptor cells that swell or shrink in response to changes in plasma tonicity.

Control of Extracellular Fluid Volume

In a healthy person, the amount of sodium in the extracellular space can vary considerably, depending on dietary salt intake; however, the extracellular sodium concentration remains almost constant because of physiologic control systems that tightly regulate water intake and excretion. In healthy persons, more salt in the extracellular space means an expanded extracellular fluid volume, and less salt means a smaller extracellular volume; but in either case the extracellular sodium concentration does not change.

Sodium balance and intravascular volume are affected by numerous hormonal and nonhormonal mediators; in addition to aldosterone and angiotensin—the best known mediators of sodium excretion—the sympathetic nervous system, natriuretic peptides, and changes in the renal circulation all play important regulatory roles [*see Table 2*].[9] Because of redundancy and overlap in the control system, failure of a single factor does not cause major, sustained abnormalities in intravascular volume. The relative importance of the various mediators that affect urinary sodium excretion are incompletely understood, and it is likely that some sodium regulatory factors remain undiscovered.

Control of Intracellular Fluid Volume

Water balance and cell volume are controlled by a single hormonal mediator, arginine vasopressin [*see* Tubular Reabsorption, *above*], which is released into the systemic circulation by the neurohypophysis [*see Table 2 and Figure 3*]. The hormone activates V_2 receptors on the basolateral membrane of principal cells in the renal collecting duct, initiating a cyclic adenosine monophosphate–dependent (cAMP-dependent) process that culminates in the insertion of water channels (aquaporins) into the cells' luminal membranes.[7,8] Modulation of the number of aquaporins controls urine osmolality and the rate of water excretion by the kidney. Vasopressin's short half-life in the circulation and continuous shuttling of aquaporins between the collecting duct's cell membrane and cytosol ensure that urinary water excretion responds rapidly to changes in body fluid tonicity.

Vasopressin levels are normally unmeasurable when the plasma sodium concentration falls to 135 mEq/L or lower [*see Figure 3*]. Low levels of the hormone result in the excretion of large volumes of a maximally dilute urine (50 mOsm/kg). Above a sodium level of 135 mEq/L, plasma vasopressin levels are linearly related to the plasma sodium concentration and increase measurably in response to changes in the plasma sodium concentration of as little as 1 mEq/L. Once the plasma sodium concentration reaches approximately 142 to 144 mEq/L, plasma vasopressin levels are high enough to promote the excretion of maximally concentrated urine (1,200 mOsm/kg). A rising plasma sodium concentration also causes hypothalamic cell volume receptors to relay signals to nearby thirst centers. Mediated by thirst and changes in vasopressin secretion, the plasma sodium concentration is normally prevented from rising above 144 mEq/L or falling below 135 mEq/L.

Under day-to-day conditions, water intake, vasopressin secretion, and urinary free-water excretion primarily respond to changes in the plasma sodium concentration created by variations in electrolyte-free water balance. Unlike sodium excretion, which is affected only by changes in intravascular volume, free-water excretion can be affected by two types of stimuli: intravascular volume and tonicity. Under pathologic conditions, osmotic control of vasopressin secretion and thirst can be overridden by hemodynamic stimuli. The hypothalamic neurons that secrete vasopressin receive neural input from baroreceptors in the great vessels and volume receptors in the atria. When these receptors are stimulated by hypotension or by a major reduction in plasma volume, impulses are carried via cranial nerves IX and X.[8] Vasopressin and thirst responses to hypovolemia and hypotension can be regarded as backup systems that serve to maintain arterial blood volume under emergency conditions, sacrificing tonicity to tissue perfusion.

CELL VOLUME REGULATION IN HYPOTONICITY AND HYPERTONICITY

Cell volume is determined by the amount and concentration of intracellular solute. Because intracellular and extracellular solute concentrations must be equal, the relation between cell water and extracellular osmolality can be described by the following equation:

$$\text{Cell water} = \frac{\text{cell solute content}}{\text{extracellular osmolality}}$$

Normally, water intake and excretion are modulated to maintain body fluid tonicity within a narrow physiologic range. However, under pathologic conditions, body cells can be exposed to a hypotonic or hypertonic milieu.[10] The first response to osmotic stress is a compensatory adjustment to intracellular electrolytes: loss of potassium in hypotonicity and accumulation of sodium and potassium in hypertonicity. With time, changes in organic solutes dominate the response.

Most cells maintain relatively high concentrations of small, osmotically active organic molecules known as organic osmolytes. Organic osmolytes are nonperturbing solutes; unlike sodium and potassium, their intracellular concentrations may vary widely without affecting tertiary protein structure. Cells accumulate organic osmolytes under hypertonic conditions and lose them when confronted with hypotonicity.

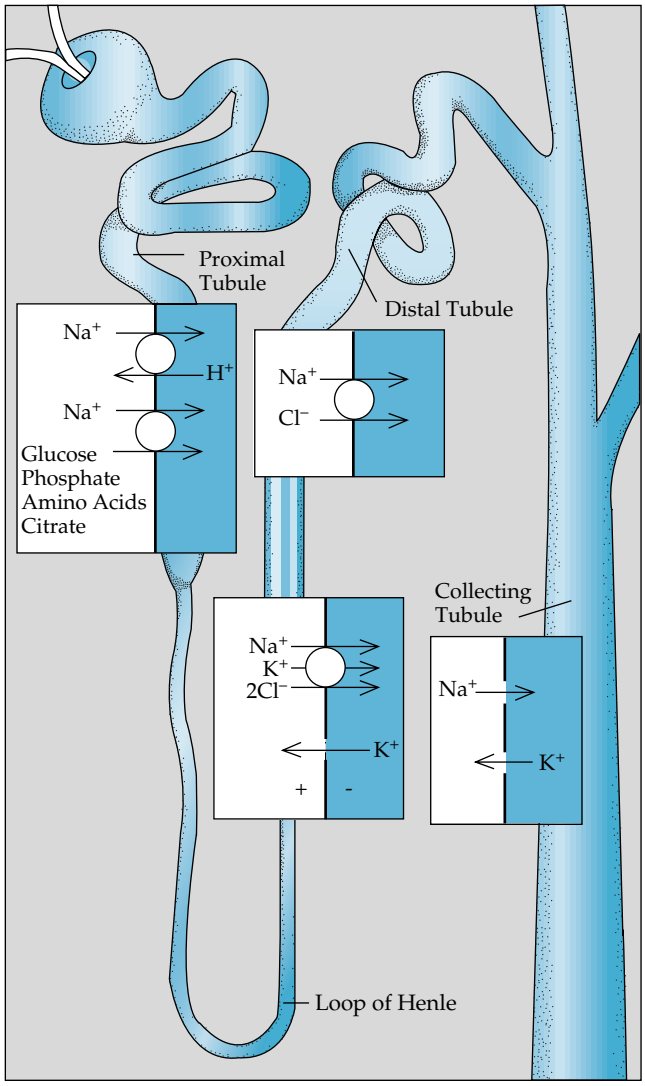


Figure 2 **In each tubular segment, sodium enters the tubular cell passively, down the favorable electrochemical gradient created by the Na^+, K^+-ATPase pump. Luminal sodium enters by a different mechanism in each of the nephron segments. The proximal tubule reabsorbs filtered bicarbonate and other solutes, such as glucose, phosphate, amino acids, and citrate. An Na^+, K^+, $2Cl^-$ transporter mediates Na^+ entry into the ascending limb of the loop of Henle. The distal tubule has an Na^+, Cl^- cotransporter. Na^+ enters the principal cells of the cortical collecting tubules through Na^+ channels in the luminal membrane.**

The need for cell volume regulation is most imperative in the brain, where the rigid calvaria places sharp limits on the degree of tissue expansion or contraction that can be tolerated.[10-13] An increase in brain water content of more than about 5% to 10% is incompatible with life. Variations in the intracellular concentration of organic osmolytes provide the brain with an astonishing ability to adapt to chronic osmotic disturbances. However, because changes in the osmolyte content of brain cells require a few days to develop fully, the brain is imperiled by rapid osmotic changes. Thus, acute hyponatremia or hypernatremia may be fatal at plasma sodium concentrations that are well tolerated chronically.

With sustained osmotic disturbances, adaptations that protect against brain swelling and shrinkage also predispose to injury when the osmotic disturbance is suddenly corrected. In chronic hyponatremia, cellular solutes lost in the adaptive phase must be recovered when the plasma sodium concentration returns to normal—a process that may require several days. Unless solute recovery keeps pace with the rising extracellular osmolality, brain cells will become dehydrated.[11,12] This phenomenon may cause clinical complications [*see* Complications of Therapy: Myelinolysis and Osmotic Demyelination Syndrome, *below*]. Similarly, in chronic hypernatremia, accumulated solutes must be shed during correction of the electrolyte disturbance.[10,13] Cells that have become acclimated to a hypertonic environment lose organic osmolytes slowly because of slow turnover of the efflux mechanism, slow downregulation of hypertonically stimulated uptake pathways, or both. Thus, when chronic hypernatremia is corrected rapidly, brain cells swell to a greater than normal volume.

Disorder of Water Excess: Hyponatremia

Hyponatremia simply means a low plasma sodium concentration. In most cases, hyponatremia is associated with a low plasma osmolality level and body fluids that are too dilute (hypotonic hyponatremia). However, there are exceptions to this rule [*see* Differential Diagnosis for Hyponatremia, *below*].

PATHOGENESIS OF HYPONATREMIA

Hypotonic hyponatremia results from two basic mechanisms, individually or together: (1) massive water intake, exceeding the capacity to excrete electrolyte-free water, or (2) impaired water excretion. Normally, the capacity for water excretion is rather large. In the absence of vasopressin, urine osmolality falls to approximately 50 mOsm/kg. A typical United States diet provides 600 to 900 mOsm of electrolytes and urea that must be excreted each day. At this rate of solute excretion, the volume of maximally dilute urine equals 12 to 18 L. Water intake can occasionally exceed the normal excretory capacity, primarily in psychotic patients who frantically ingest gallons of water over a few hours[14] and in very heavy beer drinkers who ingest large volumes of fluid but take in small amounts of salt and protein.[15] More commonly, hyponatremia occurs in patients with a diminished ability to excrete free water.[8,12]

Impaired Water Excretion

Water excretion is obviously compromised in severe renal failure; oliguric patients become hyponatremic if they are given too much water. However, most cases of hyponatremia occur in patients whose normal kidneys are unable to excrete maximally di-

Table 2 Control of Body Fluid Volumes

	Saltwater Balance	*Electrolyte-Free Water Balance*	
		Day to Day	*Emergency Backup*
Regulated variable	Extracellular volume Vascular fullness	Cell volume	Arterial filling
Clinical indicator	Blood pressure Edema	Plasma sodium concentration	Blood pressure Edema
Sensors	Baroreceptors, atrial volume receptors	Hypothalamic osmoreceptors	Baroreceptors, atrial volume receptors
Mediators	Renin-angiotensin-aldosterone system Sympathetic nervous system Atrial natriuretic peptide Starling forces in peritubular capillaries	Antidiuretic hormone (arginine vasopressin) Thirst	Antidiuretic hormone (arginine vasopressin) Thirst
Affected variable	Urinary sodium excretion	Urine osmolality Water intake	Urine osmolality Water intake

lute urine. A pathologically low plasma sodium concentration occurs when water is taken in at a time when renal diluting mechanisms are not functioning maximally because either (1) diuretics or tubular transport defects are blocking sodium reabsorption in the renal diluting segments or (2) ADH levels are elevated.

Nonosmotic Release of Vasopressin

Vasopressin is a water-retaining hormone that is released when water is needed. Because hypotonic hyponatremia normally inhibits vasopressin secretion, detectable vasopressin in a patient who is hyponatremic indicates that a nonosmotic stimulus for vasopressin release must be present. Vasopressin action increases the urine osmolality, which can be thought of as a bioassay for the hormone.

Hemodynamic stimuli for vasopressin Hypovolemia, heart failure, and cirrhosis are the most common nonosmotic stimuli for ADH secretion.[16-18] The hemodynamic abnormalities that stimulate vasopressin release also promote sodium reabsorption by the renal tubules; thus, these conditions result in both sodium and water retention.

Inappropriate antidiuretic hormone secretion Nonosmotic release of vasopressin without a hemodynamic stimulus to account for it is considered "inappropriate." Patients with the syndrome of inappropriate antidiuretic hormone secretion (SIADH) retain water because of nonosmotic release of vasopressin and abnormal thirst mechanisms but have no abnormality in sodium balance, evidence of volume depletion, or tendency to form edema; in the steady state, sodium excretion matches intake.[8,12,19,20] Because of water retention, SIADH causes mild, subclinical volume expansion. Any additional volume expansion is met by a brisk increase in urinary sodium excretion.

Reset osmostat Reset osmostat is a variant of SIADH, commonly seen in patients with chronic, debilitating illness; it is also a characteristic of normal pregnancy. Patients with this condition are able to dilute their urine normally but at a lower set point than in normal individuals. Such patients are thus mildly hyponatremic, but unlike other patients with SIADH, they are not predisposed to progressive water retention and do not require dietary water restriction or other measures used to treat chronic hyponatremia.[11] Reset osmostat can, however, be seen in malignancies, and like other causes of SIADH, it requires a diagnostic evaluation to determine its cause.

Urinary Electrolyte Losses: Desalination and Hyponatremia

If the urine is concentrated, urinary sodium and potassium losses can contribute to the pathogenesis of hyponatremia. The plasma sodium concentration can be reduced either by loss of sodium or potassium or by water gain. However, to lower the plasma sodium concentration, electrolytes must be lost in urine that has a higher electrolyte concentration than plasma. The combination of high vasopressin levels (which concentrate the

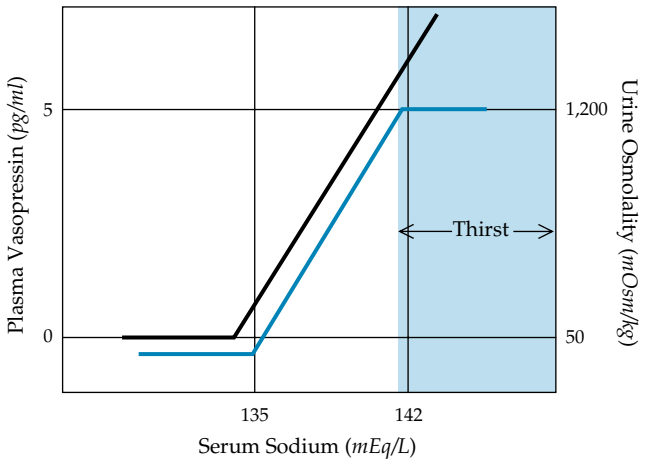


Figure 3 **Graph depicts the normal relation between plasma vasopressin levels and urine osmolality (black line) and the plasma sodium concentration (blue line). Plasma vasopressin levels change within minutes in response to changes in plasma sodium, and urine osmolality changes within minutes in response to changes in vasopressin levels. When hydration reduces the plasma sodium level below 135 mEq/L, plasma vasopressin becomes undetectable and the urine becomes maximally dilute (osmolality, 50 mOsm/kg). Between sodium concentrations of 135 and 142 mEq/L, vasopressin levels are linearly related to the plasma sodium, causing nearly a 100 mOsm/kg increase in urine osmolality for every 0.5 mEq/L increase in sodium concentration. Above a plasma sodium concentration of 142 mEq/L, the urine is maximally concentrated; increased water intake, mediated by thirst, then becomes the major defense against progressive hypernatremia.**

urine) and a high rate of sodium and potassium excretion can yield hypertonic urine capable of generating free water, which in essence desalinates the plasma.[21]

DIFFERENTIAL DIAGNOSIS FOR HYPONATREMIA

Several conditions can lower the plasma sodium concentration without causing hypotonicity and are referred to as nonhypotonic hyponatremia [*see Table 3*]. The diagnostic and therapeutic approach to these conditions differs fundamentally from the approach to hypotonic hyponatremia. Thus, it is important that nonhypotonic hyponatremia be excluded whenever a low plasma sodium concentration is encountered.

Hyperglycemic Hyponatremia

Hyperglycemia lowers the plasma sodium concentration; in the absence of insulin, glucose is an effective osmole that holds water in the extracellular space, diluting extracellular sodium. A variety of correction factors have been offered to quantify this effect.[22] However, a precise correction factor is probably unobtainable because in practice, hyperglycemia develops, in part, from the ingestion of glucose with water and resolves, in part, from the urinary excretion of glucose with water.[23] As a rough estimate, the serum sodium concentration decreases 2 mEq/L for every 100 mg/dl increase in blood glucose.

Exogenous solutes such as mannitol and maltose (a sugar contained in intravenous immunoglobulin preparations) are confined to the extracellular space and have an effect on the plasma sodium concentration similar to that of hyperglycemia. When the clinical setting suggests that these solutes might be responsible for hyponatremia, their presence can be confirmed by measuring the plasma osmolality and comparing it with the calculated value to identify an osmolar gap [*see* Osmolality, *above*].[2,11]

Postprostatectomy Syndrome and Hysteroscopic Hyponatremia

Irrigants containing mannitol, sorbitol, or glycine are used for endoscopic transurethral and intrauterine procedures [*see Table 3*].[2,24] Occasionally, several liters of irrigant may be absorbed systemically, reducing the plasma sodium in a matter of minutes. Immediately after surgery, the serum sodium concentration is much lower than would be anticipated, because the electrolyte-free solution is initially confined to the extracellular space. Glycine, the most commonly used irrigant in the United States, is metabolized to ammonia and eventually to urea and glucose. Hyperammonemia may be responsible for most of the symptoms in patients with postprostatectomy syndrome and hysteroscopic hyponatremia, and glycine itself has direct neuroinhibitory effects and may cause hypotension, bradycardia, and visual disturbances.

Pseudohyponatremia

High plasma concentrations of lipid or protein cause mild nonhypotonic hyponatremia because of an artifact of laboratory measurement [*see Table 3*].[2,11,25] With extremely high concentrations of triglycerides (enough to give serum a milky appearance), hypercholesterolemia with lipoprotein X from obstructive or cholestatic jaundice, or very high serum protein levels (from multiple myeloma or Waldenström macroglobulinemia), plasma water may constitute a smaller fraction of the plasma sample than normal, which can result in an underestimate of the "true" sodium concentration. The plasma osmolality and the sodium concentration in plasma water (as measured in an undiluted sample by a sodium-sensitive electrode) are unaffected. There are no symptoms, and no therapy is required.

ACUTE HYPONATREMIA (WATER INTOXICATION)

The term water intoxication was coined in the early 1920s to describe a neurologic syndrome that develops when large volumes of water are retained within a relatively short period of time (< 48 hours). The syndrome is often referred to as acute hyponatremia.[12,26]

Etiology

Acute hyponatremia develops when water intake is high and electrolyte-free water excretion is impaired. Potentially, hyponatremia can develop rapidly in any patient predisposed to water

Table 3 Causes of Nonhypotonic Hyponatremia

Condition	*Plasma Osmolality*	*Pathogenesis*	*Therapeutic Implications*
Hyperglycemia	High	Extracellular glucose osmotically draws water into the ECF, diluting extracellular sodium	During treatment of hyperglycemia, anticipate 3 mEq/L increase in serum sodium for every 200 mg/dl reduction in blood sugar
Intravenous hypertonic mannitol therapy	High	Water shift from ICF to ECF as with hyperglycemia	Mannitol is rapidly excreted when renal function is normal
Intravenous γ-globulin therapy	High	Maltose present in solution acts like mannitol	Measure plasma osmolality when hyponatremia is suspected
Irrigant absorption (prostatectomy or intrauterine surgery)	Normal or low (when hypoosmolar irrigants are used)	Absorbed solute—mannitol, sorbitol, or glycine (most common)—initially confined to ECF, causing severe hyponatremia but little change in plasma osmolality	Mannitol is rapidly excreted; sorbitol is metabolized, causing late-onset hypotonic hyponatremia; glycine is neurotoxic and causes transient blindness and is metabolized to ammonia, causing encephalopathy; consider hemodialysis
Pseudohyponatremia (severe hyperlipidemia, multiple myeloma, macroglobulinemia)	Normal	Laboratory artifact; plasma water constitutes a smaller fraction of the plasma sample, causing a more serious underestimate of the true sodium concentration	Suspect when serum is lactescent; compare measured plasma osmolality with calculated osmolality or measure plasma sodium with direct-reading sodium electrode

ECF—extracellular fluid ICF—intracellular fluid

Table 4 Causes and Treatment of Acute Hyponatremia

Causes	*Pathogenesis*	*Effect of Treatment*	*Recommendations*
Postoperative stress*	Vasopressin is secreted in response to surgical stress for 2 or more days; free water from hypotonic I.V. fluids is retained and sodium and potassium are excreted in urine at high concentrations	Normal saline ineffective for correction—administered sodium is excreted in concentrated urine, "desalinating" isotonic fluid and causing water retention	Avoid hypotonic fluid (e.g., D5W, 0.45% saline) and excessive volumes of isotonic fluid (lactated Ringer solution or 0.9% saline) after surgery; treat symptomatic hyponatremia with 3% saline and furosemide
Oxytocin	Used in obstetrics to induce labor; direct antidiuretic effect of drug mimics SIADH; free water from I.V. fluids retained	Urine becomes dilute when oxytocin is discontinued	Avoid administration of oxytocin in or with hypotonic fluids; treat hyponatremia by discontinuing drug
Cyclophosphamide	Drug has antidiuretic effect that persists for as long as 12 hours; patients are encouraged to drink large volumes of water to prevent chemically induced cystitis	Normal saline ineffective for correction as in other causes of persistent SIADH	Treat symptomatic hyponatremia with 3% saline and furosemide
Psychotic self-induced water intoxication	Extreme polydipsia (> 1 L/hr) common in patients with severe psychosis; retained water causes hyponatremia by late afternoon or evening, and water diuresis restores normonatremia by morning	Normal ability to dilute urine in most patients so hyponatremia self-corrects when water intake stops; some patients have vasopressin release (often transient) from stress, smoking, or medications (e.g., carbamazepine)	Monitor diurnal weight in institutionalized patients for early detection; avoid antidiuretic medications; treat hyponatremia with water restriction; use hypertonic saline and furosemide for occasional patient with SIADH
Marathon running	Extracellular volume depletion caused by saltwater losses from sweating and possibly stress are nonosmotic stimuli for vasopressin secretion; large volumes of sugar water consumed during race are retained	Isotonic saline restores ability to dilute urine	3% saline without furosemide for seizures; isotonic saline and water restriction for more moderate symptoms
Ecstasy (methylenedioxymethamphetamine [MDMA]) use	Excessive fluid intake and inappropriate antidiuretic hormone secretion, induced by MDMA, is implicated	Isotonic saline ineffective; self-correction typical but may be delayed	Hypertonic saline for severe symptoms

*Excluding irrigant absorption syndromes [*see Table 3*].
D5W—5% dextrose in water SIADH—syndrome of inappropriate antidiuretic hormone

retention who takes in a large volume of water in a short period of time. However, this is likely to occur in a limited number of settings [*see Table 4*], and such instances account for most cases of severe symptomatic hyponatremia and for most of the recorded fatalities.

Postoperative hyponatremia Vasopressin is released immediately after surgical procedures in what appears to be a stress response [*see Table 4*].[21,27] Particularly during the first 24 hours, the concentration of urinary cations (sodium plus potassium) may greatly exceed the plasma sodium concentration. As a result, even isotonic fluids may be "desalinated" and can lower the plasma sodium concentration.[21] Thus, all hypotonic fluids and excessive amounts of isotonic fluids should be avoided after surgery. As noted, endoscopic prostatectomy and intrauterine procedures can cause hyponatremia if the irrigant used in the procedures is absorbed systemically. The management of irrigant absorption syndromes differs from that of other causes of postoperative hyponatremia [*see* Postprostatectomy Syndrome and Hysteroscopic Hyponatremia, *above*].[27]

Oxytocin infusions Oxytocin, which is used in obstetrics to induce labor, has a direct antidiuretic effect. If the drug is administered in 5% dextrose in water (D5W), which was formerly a common practice, symptomatic hyponatremia may emerge after the infusion of less than 3 L of fluid [*see Table 4*].[11] Termination of the infusion permits a water diuresis and correction of hyponatremia; however, the syndrome is best avoided by using isotonic saline as a vehicle for the drug.

Cyclophosphamide infusion Intravenous cyclophophamide impairs water excretion by an unknown mechanism.[28] The antidiuretic effect of the drug begins 4 to 12 hours after injection and persists for as long as 12 hours. Patients receiving cyclophosphamide are particularly susceptible to hyponatremia because they are encouraged to drink large volumes of water to prevent chemically induced cystitis [*see Table 4*].

Psychotic self-induced water intoxication Extreme polydipsia is relatively common in patients with psychiatric illnesses, particularly schizophrenia, and it may lead to symptomatic hyponatremia [*see Table 4*].[14] Daily intake of 10 to 15 L has been documented, and much of the intake may take place over a few hours. Many patients become hyponatremic in the late afternoon and evening; however, water diuresis typically restores normonatremia by the following morning. Occasionally, individuals drink enough water to produce seizures. By monitoring diurnal changes in body weight, water intoxication can be recognized before the onset of severe neurologic symptoms. Transient release of vasopressin (most commonly provoked by nau-

sea and vomiting) may contribute to water retention. There is little evidence that any of the major tranquilizers has a significant antidiuretic effect; however, carbamazepine, an anticonvulsant, enhances sensitivity to vasopressin.

Water intoxication during exercise Hyponatremia is disturbingly common in nonelite marathon runners; it is associated with slow finishing times and with excessive consumption of fluids while running, as evidenced by substantial weight gain.[29,30] Severe symptomatic hyponatremia has mostly been reported after participation in marathons or ultramarathons, but symptomatic hyponatremia may also occur after recreational running and military fitness training.

Water intoxication from the drug ecstasy During the 1990s, 3,4-methylenedioxymethamphetamine (MDMA, or ecstasy) gained widespread popularity as a recreational drug taken at dances.[31] When malignant hyperthermia was recognized as a complication associated with this drug, MDMA users were advised in underground magazines and the lay press to drink plenty of fluids. Subsequently, acute water intoxication emerged as a potentially lethal complication of the drug [*see Table 4*]. Excessive fluid intake and SIADH, induced by MDMA, have been implicated.

Diagnosis

Symptoms of water intoxication include headaches, weakness, nervousness, and vomiting, progressing to disorientation, delirium, tremulousness, and ultimately convulsions and coma.[12,26] The pupils are often dilated, and bilateral Babinski signs may be present. On occasion, patients may present with hemiparesis, mimicking a cerebrovascular accident. The syndrome reflects cerebral edema, which can lead to herniation of the brain and death. Clinical findings may emerge explosively. Complaints of headache and mild confusion may be followed within hours by respiratory arrest and, in some cases, neurogenic pulmonary edema. For reasons that remain obscure, almost all reported fatalities from acute postoperative hyponatremia have been in women (usually of childbearing age) and young children. Fatal cases of acute hyponatremia from other etiologies have been recorded in men and women.

Acute hyponatremia should be suspected in any patient who has unexplained neurologic symptoms, especially in psychiatric patients, marathon runners, users of ecstasy, and patients receiving hypotonic fluids intravenously (e.g., after surgery). Serum electrolyte levels should be obtained immediately. In the proper setting, a tentative diagnosis of water intoxication is advisable when symptoms develop in a patient whose serum sodium concentration is lower than 130 mEq/L (provided that causes of nonhypotonic hyponatremia have been excluded). Although severe neurologic symptoms do not usually appear until the sodium level has fallen below 120 mEq/L, some patients (particularly young women and children) may be unusually susceptible to brain edema when they become acutely hyponatremic; in rare cases, fatalities have been reported at plasma sodium concentrations between 120 and 128 mEq/L.[21,27]

Elderly patients can tolerate acute hyponatremia better than the young, because brain atrophy affords more room for brain cell swelling. The same water load per kilogram of body weight can cause a much more severe degree of acute hyponatremia when water is ingested rapidly, especially if the person has a much smaller muscle mass (the reservoir for a water load that limits brain cell swelling). When the serum sodium concentration is falling rapidly, the arterial sodium concentration (to which the brain responds) may be lower than the venous sodium concentration (which is measured in most clinical electrolyte assays).[32]

Computed tomography demonstrates cerebral edema in severe cases of water intoxication, and it rules out other potential explanations for neurologic findings. However, when symptoms are severe, therapy should not be delayed while imaging studies are being obtained.

Treatment

Free-water intake should be stopped immediately whenever water intoxication is suspected. Hypertonic saline is the treatment of choice for water-intoxicated patients who cannot autocorrect their electrolyte disturbance, including patients with neurogenic pulmonary edema.[26,27] Each 1 ml of 3% saline contains 0.5 mEq of sodium. Because there are approximately 0.5 L of body water for every 1 kg of body weight, 1 ml of 3% saline per 1 kg of body weight can be expected to increase the plasma sodium concentration by 1 mEq/L. For patients with severe neurologic symptoms, an infusion of 3% saline at 1 to 2 ml/kg/hr will increase the plasma sodium concentration by approximately 1 to 2 mEq/L/hr, a rate that is considered appropriate for initial therapy. Hypertonic saline is best infused in 100 ml containers to avoid inadvertently giving an excessive dose. Concurrent administration of a loop diuretic (furosemide, bumetanide, or torsemide) is advisable. The diuretic prevents volume overload and, by blocking sodium reabsorption in the loop of Henle, impedes the formation of concentrated urine.

The goal of therapy in acute hyponatremia is to decrease the severity of cerebral edema and to stop seizures. A 4 to 6 mEq/L increase in plasma sodium concentration is usually sufficient to accomplish these goals. Thus, the plasma sodium concentration should be monitored frequently during therapy, and emergency treatment with hypertonic saline should be stopped after 2 to 3 hours. Once initial therapy with high-dose hypertonic saline has been completed, more conservative measures should be substituted to gradually return the plasma sodium concentration to normal. To avoid complications from excessive correction of hyponatremia, the plasma sodium concentration should not be intentionally increased by more than 12 mEq/L during the first day of therapy or by more than 6 mEq/L/day thereafter.

CHRONIC HYPONATREMIA

The distinction between acute and chronic hyponatremia is somewhat arbitrary. Commonly, hyponatremia is considered chronic when it has evolved over the course of 48 hours or more.[11,12,26] Although the precise duration of an electrolyte disturbance cannot be known when it develops outside the hospital (except for psychotic water drinkers, marathon runners, and users of ecstasy), outpatients can be assumed to have chronic hyponatremia.[33] Prolonged hyponatremia cannot occur unless there is a sustained defect in water excretion. Except for patients with renal failure, virtually all chronically hyponatremic patients have some abnormality in vasopressin secretion.

Etiology

Advanced renal failure A low glomerular filtration rate limits the ability to excrete electrolyte-free water. Many patients with advanced renal failure excrete urine that has the same osmolality as plasma regardless of physiologic conditions (fixed

isosthenuria). In acute oliguric renal failure, the ability to excrete free water is virtually nil; administration of hypotonic fluids must be scrupulously avoided to avoid hyponatremia.

Diuretics Thiazide diuretics are commonly the sole cause or a major contributing factor of hyponatremia requiring hospital admission.[33,34] For unknown reasons, severe hyponatremia caused by thiazides affects elderly women much more often than other groups. By blocking the reabsorption of sodium and chloride in the distal tubule, thiazides and metolazone prevent the generation of maximally dilute urine.[35] Because sodium reabsorption in the ascending limb of the loop of Henle is left unaffected by these agents, they permit excretion of maximally concentrated, hypertonic urine and can lead to simultaneous retention of water and depletion of sodium and potassium. Extraordinarily severe hyponatremia can result from thiazides, with plasma sodium levels as low as 100 mEq/L. Vasopressin levels are usually elevated in patients who present with thiazide-induced hyponatremia, sometimes because of diuretic-induced volume depletion but more often because of the stress of minor intercurrent illnesses. Patients with thiazide-induced hyponatremia do not usually appear clinically volume depleted, presumably because retained water partially sustains extracellular fluid volume. Patients who have become hyponatremic on thiazides should not be given these agents again; recurrent episodes of severe hyponatremia are common.

Hypovolemia Hypovolemic hyponatremia is most often associated with gastrointestinal fluid losses caused by vomiting, diarrhea, or laxative abuse. Surprisingly, particularly in alcoholics, patients who continue to drink while vomiting repeatedly can still absorb enough ingested water to become hyponatremic. Electrolyte losses in the vomitus, combined with urinary sodium and potassium losses that result from metabolic alkalosis, lower the plasma sodium concentration.

Beer potomania Patients who subsist on beer (a practice known as beer potomania) are susceptible to hyponatremia because of their low rates of solute excretion (beer contains little protein or electrolyte). Reduced delivery of the glomerular filtrate to distal diluting sites limits the amount of water that can be excreted. Nonosmotic stimuli to vasopressin secretion caused by nausea or gastrointestinal fluid losses or by treatment with thiazide diuretics are often contributing factors.[15]

Edematous conditions Any disease that can cause edema also predisposes to water retention and hyponatremia. The same hemodynamic factors that promote sodium retention are nonosmotic stimuli for vasopressin release.[8,16-18] Elevated vasopressin levels have been reported in hyponatremic patients with congestive heart failure, cirrhosis, and nephrotic syndrome. In heart failure, hyponatremia is associated with a low cardiac output and a poor prognosis.

SIADH Nonosmotic release of vasopressin that has no hemodynamic explanation is termed inappropriate [*see Table 5*].[12,19] A number of tumors (most commonly small cell carcinoma of the lung) ectopically synthesize and secrete vasopressin.[36] Unexplained, persistent hyponatremia should be considered a marker for an underlying malignancy.

SIADH may also complicate the course of a wide variety of conditions in which there is damage to or inflammation of the central nervous system.[8,12] In patients with subarachnoid hemorrhage, natriuretic peptides released by the brain may directly promote urinary sodium loss, regardless of extracellular volume (cerebral salt wasting).[37,38] Urinary salt losses combined with vasopressin-induced water retention are responsible for hyponatremia. SIADH is a common complication of chest infection. Antidiuretic activity has been demonstrated by bioassay in patients with tuberculous lung tissue, and tuberculosis causes SIADH.[11] In pneumonia, vasopressin levels are increased during the acute phase of the disease and return to baseline within a few days. Isolated glucocorticoid deficiency caused by anterior pituitary dysfunction also causes hyponatremia; patients with hypopituitarism develop SIADH but, unlike patients with Addison disease, have normal levels of mineralocorticoid and do not become hypovolemic or hyperkalemic. Hyponatremia caused by glucocorticoid deficiency promptly resolves when cortisol is replaced. Hypothyroidism also causes SIADH; hyponatremia gradually resolves when thyroid hormone replacement is given.[39]

A number of therapeutic agents can induce SIADH.[12,40] Nonsteroidal anti-inflammatory drugs (NSAIDs) decrease water excretion because they inhibit formation of prostaglandin E_2,

Table 5 Causes of the Syndrome of Inappropriate Antidiuretic Hormone

Tumors	Bronchogenic (small cell) Pancreatic Duodenal Urethral Nasopharyngeal Leukemia Hodgkin disease Thymoma
Neurologic disorders	Psychosis Trauma Neoplasms (primary and metastatic) Vascular (hemorrhage, infarction, and vasculitis) Infection (meningitis, brain abscess, and encephalitis) Miscellaneous (Guillain-Barré syndrome, multiple sclerosis, hydrocephalus, Shy-Drager syndrome)
Pulmonary disorders	Infectious (bacterial, viral, and fungal pneumonia and tuberculosis) Functional (asthma, acute respiratory failure, and mechanical ventilation)
Endocrine diseases	Glucocorticoid deficiency (hypopituitarism) Hypothyroidism
Drugs	Antidiuretic hormones (vasopressin, DDAVP, and oxytocin) Psychotropic agents (tricyclic antidepressants, serotonin reuptake inhibitors, monoamine oxidase inhibitors, and carbamazepine) Ecstasy (MDMA) Antineoplastic agents (cyclophosphamide, vincristine, and vinblastine) Nonsteroidal anti-inflammatory drugs Diabetic agents (chlorpropamide and tolbutamide) Miscellaneous (bromocriptine and nicotine)
Other causes	Postoperative stress Alcohol withdrawal AIDS Nausea

DDAVP—1-desamino-8-D-arginine vasopressin

which modulates vasopressin action.[8] Rare cases of hyponatremia solely attributable to NSAIDs have been reported, but these commonly used agents may exacerbate other causes of hyponatremia.

Hyponatremia in AIDS Hyponatremia is an extremely common finding in AIDS patients.[41] Many AIDS patients have features of SIADH associated with opportunistic infections that cause pneumonia and meningitis. Others have clinical signs of volume depletion without low urine sodium values, a finding that may indicate coexistent renal disease or adrenal insufficiency.[42] Hyponatremia often occurs when antibiotics are administered in hypotonic intravenous solutions.

Diagnosis

Hyponatremia should be approached in a systematic fashion. First, the various disorders that can lower the plasma sodium concentration without causing hypotonicity should be excluded [*see* Differential Diagnosis for Hyponatremia, *above*]. Once it has been established that hypotonic hyponatremia is present, the mechanism for impaired water excretion is identified (hypovolemia versus an edematous condition versus SIADH), and the differential diagnosis that applies to that mechanism is considered. The most challenging goals of the diagnostic process are to determine whether chronic SIADH is present and, if it is, to define the specific disease responsible for the syndrome.

Clinical manifestations Because cerebral edema is usually not severe, the symptoms of chronic hyponatremia are much more subtle, vague, and nonspecific than those of acute water intoxication[12,26]; indeed, patients with chronic hyponatremia are often asymptomatic at sodium levels that may be lethal to a patient with acute water intoxication. As the plasma sodium concentration falls below 115 to 120 mEq/L, patients often experience anorexia, nausea, vomiting, muscle weakness, and muscle cramps. They may be irritable and show personality changes, becoming uncooperative, confused, hostile, or simply slow to respond. At plasma sodium concentrations below 110 mEq/L, gait disturbances, falling, stupor, tremulousness, and, more rarely, seizures may occur.

Chronic hyponatremia itself is rarely, if ever, fatal. However, chronically hyponatremic patients may develop life-threatening acute hyponatremia if they rapidly drink water or are infused with a large volume of hypotonic fluid. Because hyponatremia can be a marker for severe underlying illness, hospitalized patients with hyponatremia often have a high mortality, dying with but not of chronic hyponatremia. There is little evidence that chronic hyponatremia itself leads to permanent sequelae, even when the plasma sodium concentration falls below 105 mEq/L.[33,43] However, patients with prolonged, severe hyponatremia are susceptible to iatrogenic injury if their electrolyte disturbance is corrected too rapidly [*see* Treatment, *below*].

History and physical examination The history in patients with chronic hyponatremia should include information about diet, fluid intake, gastrointestinal fluid losses, and use of diuretics, antidepressants, or other antidiuretic drugs. During the physical examination, physicians should look for clinical signs of volume depletion or an edematous condition. Evidence of volume depletion may not always be definitive, however. For example, vomiting may be a symptom rather than the cause of hyponatremia; extreme hyponatremia may occasionally impair baroreceptor reflexes, causing postural hypotension and a false impression of volume depletion; and retained water may mask underlying volume depletion. When the distinction between hyponatremia caused by hypovolemia and hyponatremia caused by SIADH is not obvious, laboratory clues may helpful.

Laboratory tests Measurement of the urinary sodium concentration, chloride concentration, or both is often the most helpful test.[44] Water retention caused by hypovolemia or by an edematous condition is usually associated with a urinary sodium concentration lower than 20 mEq/L in a spot sample. Hypovolemia caused by upper gastrointestinal fluid losses is an important exception. Loss of gastric fluid causes a metabolic alkalosis that may increase urinary sodium excretion despite volume depletion; the diagnosis can be made by measuring the urine chloride concentration, which is reduced in this condition. In SIADH, urinary sodium matches intake; because the urine is usually concentrated, the urinary sodium concentration exceeds 40 mEq/L unless dietary sodium intake is very low. Measurements of the BUN and serum uric acid complement these measurements. When a hemodynamic abnormality is responsible for hyponatremia, the kidney is underperfused, urea and uric acid clearances are diminished, and the BUN and serum uric acid levels are usually elevated. Conversely, SIADH is a volume-expanded state, and BUN and uric acid levels are usually low. Uric acid is a more reliable indicator of volume status than the BUN, because the latter value is affected by dietary protein intake as well as renal clearance.

Assessment of acid-base and potassium balance may provide helpful clues to the diagnosis. The serum potassium and bicarbonate levels are normal in SIADH. Hypokalemia and metabolic alkalosis suggest diuretic therapy or vomiting, which can be surreptitious. Hyperkalemia and metabolic acidosis suggest the possibility of adrenal insufficiency. Hypokalemia and acidosis can result from diarrhea, and their presence may raise the possibility of surreptitious laxative abuse.

Withdrawal of hyponatremic drugs When a patient is taking a drug that can cause hyponatremia, it is important to exclude another underlying cause of hyponatremia before attributing the electrolyte disturbance to the medication. For example, thiazide diuretics can exacerbate hyponatremia caused by SIADH. The best way to confirm a diagnosis of drug-induced hyponatremia is to eliminate the offending agent and be sure that water excretion returns to normal when the patient is off the drug. Full resolution of hyponatremia and full recovery of diluting function may be delayed for a week or two in patients with thiazide-induced hyponatremia. During repair of sodium and potassium deficits, transient resetting of the osmostat is common and should not necessarily prompt an extensive search for an underlying cause.

Response to therapy On occasion, evidence regarding the cause of hyponatremia can be equivocal. In such cases, the patient's response to isotonic saline (or a generous oral salt intake and the passage of time) is the best clue to the diagnosis. Patients with subclinical edematous conditions will retain the administered sodium, developing clinically obvious edema. Volume-depleted patients initially retain the administered sodium, but as soon as hypovolemia is corrected, the urine becomes dilute, the rate of urinary sodium excretion increases to match intake, and hyponatremia improves as water is excreted in the urine. Uri-

nary sodium excretion promptly increases in patients with SIADH, but the urine remains concentrated and hyponatremia persists. Isotonic saline should be given with extreme caution to patients with very low plasma sodium concentrations; in SIADH, saline can exacerbate hyponatremia, whereas in volume depletion, hyponatremia may correct too rapidly.

Identifying a specific cause for SIADH SIADH is a mechanism for developing hyponatremia, not a diagnosis. In all patients with SIADH, a specific etiology for inappropriate vasopressin secretion should be sought. When hyponatremia develops during hospitalization, the cause is sometimes obvious (e.g., pneumonia, meningitis, or acute respiratory failure) and no further testing is indicated. In a patient with clinical features of SIADH but no obvious cause for it, a more extensive evaluation is indicated. The workup should include a careful search for malignancy and central nervous system pathology and an endocrine evaluation to exclude hypothyroidism and hypocortisolism. Sometimes, no cause for SIADH is found, especially in elderly patients and patients with psychiatric disorders, mental retardation, or alcoholism.[45] Careful follow-up is important, because malignancies may become clinically apparent after several years in so-called idiopathic SIADH.

Treatment

Patients with very low plasma sodium concentrations usually have some neurologic symptoms, and they are at risk of sustaining injuries from falls. However, unlike acute water intoxication, chronic hyponatremia poses little risk of an explosive onset of seizures or a fatal outcome, provided that water is withheld and the plasma sodium concentration is not allowed to fall any further. On the other hand, patients with chronic hyponatremia are at considerable risk for neurologic injury caused by overaggressive correction. Thus, there are four major goals in managing chronic hyponatremia: (1) prevention of a progressive decrease in plasma sodium concentration; (2) amelioration of hyponatremic symptoms by promptly but carefully increasing the plasma sodium concentration (an increase of at least 4 mEq/L/day is desirable in most cases); (3) avoidance of excessive correction; and (4) gradual restoration and maintenance of a normal plasma sodium concentration.

Free-water restriction should be instituted in all patients until the plasma sodium concentration has begun to increase. Intravenous fluids should be at least isotonic, and oral fluid intake should be limited to 500 to 1,000 ml/day, depending on the severity of the electrolyte disturbance. In patients with reversible defects in water excretion, limitations on free-water intake should be lifted once the plasma sodium concentration has begun to increase.

Attempts to calculate the dose of sodium chloride needed to correct hyponatremia are doomed to failure. The increase in plasma sodium concentration depends on the amounts of administered sodium and potassium that the body retains, as well as on the amount of electrolyte-free water that is eliminated in the urine. Indeed, in some cases, the plasma sodium concentration will return to normal solely because of a water diuresis, with no sodium given.

The measures required to increase the plasma sodium concentration, along with the likelihood of inadvertent rapid correction, vary depending on the cause of hyponatremia. For therapeutic purposes, the causes can be divided into reversible and persistent defects in water excretion.

Reversible defects in water excretion Hyponatremia corrects easily when the cause of defective water excretion can be eliminated by volume expansion, by withdrawal of a therapeutic agent, or by treatment of an underlying illness [*see Table 4*]. In patients with reversible defects in water excretion, avoiding excessive correction may become a major challenge.

Hypovolemic hyponatremia responds readily to 0.9% sodium chloride because the sodium concentration of isotonic saline is higher than the cation concentration of the excreted urine. Once the volume deficit is repaired and the hemodynamic stimulus to vasopressin secretion is removed, the urine becomes dilute and a water diuresis may rapidly return the plasma sodium concentration to normal. Similarly, patients with diuretic-induced hyponatremia are extremely susceptible to rapid correction; restoration of the renal diluting mechanism when the diuretic is discontinued and replacement of sodium and potassium deficits contribute to the increase in plasma sodium concentration.

Intravenous saline should be discontinued once clinically apparent hypovolemia has been corrected and the plasma sodium concentration has begun to increase. Saline should be given cautiously, if at all, to hypokalemic patients who require potassium replacement. During repair of a potassium deficit, potassium enters cells, displacing sodium, which then returns to the extracellular fluid; administered potassium is therefore as effective as sodium in raising the plasma sodium concentration. Diuretic-induced hyponatremia does not usually necessitate use of intravenous saline; for most patients, an adequate diet, replacement of potassium deficits, and discontinuance of thiazide diuretics are sufficient. In severely hyponatremic patients, the plasma sodium concentration should be monitored every 6 to 8 hours for the first 2 to 3 days of therapy. If it appears that a water diuresis is going to increase the plasma sodium by more than the desired amount, replacement of fluid losses with oral water or D5W may become necessary.

Persistent defects in water excretion: SIADH Patients with SIADH tend to be resistant to rapid changes in plasma sodium concentration (unless the cause of SIADH is short-lived). Water restriction is the cornerstone of therapy, but if used alone, water restriction often leads to an extremely slow resolution of hyponatremia. Isotonic saline is ineffective and may even be counterproductive. Furosemide and other loop diuretics are often useful therapeutic adjuncts because by blocking sodium reabsorption in the ascending limb of the loop of Henle, they interfere with the renal-concentrating mechanism, partially blocking the effect of vasopressin. Loop diuretics can be combined with oral salt or a slow infusion (approximately 15 ml/hr) of 3% saline. Oral and intravenous urea have been used extensively to treat SIADH in some parts of Europe, but experience with this agent in the United States is very limited. Demeclocycline, a tetracycline that blocks the effect of vasopressin on the collecting duct, is another therapeutic option in chronic SIADH; however, its expense and long duration of action limit its effectiveness. Several orally active vasopressin receptor blockers have been developed and are currently in clinical trials.[46,47]

Persistent defects in water excretion: edematous conditions and renal failure Saline should rarely, if ever, be given to correct hyponatremia in edematous patients or patients with renal failure (except for those with prerenal azotemia). Because it has no effect on water excretion, 1 L of 0.9% saline will increase the plasma sodium concentration by only 1 mEq/L.[12] In addition,

saline exacerbates edema and ascites in patients with cirrhosis and may cause pulmonary edema in patients with heart failure or renal failure.

Although thiazide diuretics are contraindicated, loop diuretics are the mainstay of treatment of hyponatremia for patients with edematous conditions because they increase free-water excretion and improve hyponatremia, particularly when dietary salt intake is increased. There is a natural inclination to discontinue loop diuretics when severely edematous patients develop hyponatremia. The usual problem, however, is oliguria and diuretic resistance rather than overdiuresis; the proper response is to increase the dose of loop diuretics and restrict water intake. The combination of a loop diuretic and an angiotensin-converting enzyme (ACE) inhibitor is particularly effective in patients with heart failure. The beneficial effect of an ACE inhibitor can be explained by reduced thirst and vasopressin secretion attributable to angiotensin II and by a direct effect on the hydro-osmotic effect of vasopressin, mediated by prostaglandins.[11]

Hyponatremia in edematous conditions is mediated by vasopressin. Clinical trials have shown that vasopressin receptor antagonists can be effective in managing patients with hyponatremia and edema.[46,47]

Treatment of hyponatremic seizures A small percentage of chronically hyponatremic patients with very low plasma sodium concentrations present with seizures. Regardless of the suspected duration or cause of the electrolyte disturbance, active seizures may be resistant to anticonvulsants alone and should be treated with hypertonic saline. The therapeutic approach is similar to that used for patients with acute water intoxication, except that even more vigilance is required to prevent an excessive increase in plasma sodium concentration once emergency measures have been discontinued.[12,26]

Table 6 Causes of Hypernatremia

Electrolyte-free water losses	Extrarenal Insensible loss (skin and lungs) Renal Neurogenic (central) diabetes insipidus Nephrogenic diabetes insipidus Congenital X-linked (V_2 vasopressin receptor defect) Recessive (aquaporin defects) Acquired Electrolyte abnormalities: hypokalemia, hypercalcemia Drugs: lithium, demeclocycline, methoxyflurane Pregnancy (vasopressinase) Excess urea excretion
Hypotonic losses	Extrarenal Sweat Upper GI tract Osmotic cathartics Renal Glycosuria, mannitol, glycerol, diuretics
Salt poisoning	Oral Parenteral $NaHCO_3$, 3% or 5% I.V. saline, therapeutic abortion (inadvertent 29% I.V. saline) Hemodialysis with hypertonic dialysate

Complications of Therapy: Myelinolysis and Osmotic Demyelination Syndrome

Excessive correction of chronic hyponatremia may be complicated by neurologic injury.[43,48] Typically, the patient's hyponatremic symptoms improve as the plasma sodium concentration increases, but after a delay of one to several days, new findings emerge. The patient may become confused and may exhibit psychotic or catatonic behavior, pathologic crying, or a movement disorder. Swallowing dysfunction, progressive unresponsiveness, and a spastic quadriparesis may develop. In severe cases, locked-in syndrome occurs—that is, the patient is awake but unable to move or respond. The stereotypical pattern of delayed neurologic deterioration after rapid correction of hyponatremia has been named the osmotic demyelination syndrome, because these clinical features are associated with brain lesions (myelinolysis) characterized by disruption of myelin and sparing of neurons and axons.[48-50] Lesions, which are best identified by magnetic resonance imaging, are typically found in the center of the basal pons (central pontine myelinolysis), but histologically similar lesions may also occur in a symmetrical distribution in extrapontine areas of the brain where there is a close admixture of gray and white matter. The osmotic demyelination syndrome has been reproduced in animal studies[49]; these experiments have shown that the disorder is a complication of rapid correction of hyponatremia rather than the electrolyte disturbance itself. Observational studies in severely hyponatremic patients suggest that this therapeutic complication can be avoided if correction rates are maintained below 10 to 12 mEq/L/day and 18 mEq/L/48 hr. It should be emphasized that these values are limits and not goals. Because large increases in the serum sodium concentration are seldom required to relieve hyponatremic symptoms and because unintentional excessive correction is common, the goal of therapy should be to increase serum sodium concentration by 8 mEq/L/day or less.[12]

Disorder of Water Deficiency: Hypernatremia

PATHOGENESIS

Persistent hypernatremia results from one of two basic mechanisms: water is lost and not adequately replaced or, less commonly, too much salt is taken in without enough water [*see Table 6*].[13,51-53] In either case, electrolyte-free water is needed to return the plasma sodium concentration to normal. Because thirst is the primary defense against hypertonicity, persistent hypernatremia indicates a defect in water intake. A maximally concentrated urine minimizes but does not prevent water losses. Insensible water losses from the skin and lungs are unavoidable, and urea excretion obligates some urinary losses. Maintenance of a normal serum sodium concentration (135 to 142 mEq/L) requires that daily water losses be replaced.

Most hypernatremic patients are too sick, too young, or too old to obtain water themselves or ask for it.[13,51] Sometimes, the thirst sensation itself is impaired, so that the patient has no desire to drink when the plasma sodium concentration increases above the normal range. Inadequate water intake by itself will lead to hypernatremia. When impaired intake is coupled with excessive water losses, severe hypernatremia results.

ETIOLOGY

Electrolyte-free water can be lost as pure water, with no accompanying electrolyte, or it can be lost in hypotonic fluids,

which have lower electrolyte concentrations than plasma. Hypotonic losses can be thought of as mixtures of isotonic fluid and free water. Pure-water and hypotonic fluid losses, the most common causes of hypernatremia, are typically associated with a contracted extracellular fluid volume.[13] However, this is not always the case. When hypernatremia is caused by a rapid intake of salt (acute salt poisoning), the extracellular volume expands because of water drawn from the intracellular space.[51] In critically ill patients, extracellular volume expansion with edema often coexists with hypernatremia[52,54]; the finding reflects free-water losses in patients who become edematous after fluid resuscitation for shock or underlying conditions such as congestive heart failure, renal disease, and hepatic cirrhosis.

Pure-Water Losses

When pure-water losses are responsible for hypernatremia, each body-fluid compartment loses an equal percentage of its volume.[13,51] Plasma constitutes only one twelfth of total body water (one quarter of extracellular fluid volume), and plasma volume is defended by oncotic pressure, which increases with water loss. Thus, plasma volume contracts by less than 83 ml for each 1 L of water lost; clinical signs of hypovolemia are unusual unless the water deficit is extremely large.

Insensible water losses Water is constantly lost by evaporation from the skin and lungs and must be replaced to avoid dehydration. Daily insensible water losses, normally about 0.5 L, can be increased severalfold by high environmental temperature, fever, or hypermetabolic states such as thyrotoxicosis.

Increased urea excretion Although urea is an ineffective osmole that freely crosses most cell membranes, urinary urea excretion can play an important role in water balance. High rates of urea excretion caused by very high protein diets, catabolism, or recovery from renal failure obligate increased rates of water loss. When the urine solute is composed almost exclusively of urea, the urine becomes an electrolyte-free water solution, regardless of its osmolality.

Diabetes insipidus Because sodium excretion is unaffected in diabetes insipidus (see below), the excess fluid lost in the urine is pure water. As long as water is available and the patient is able to drink, hypernatremia does not occur. Without water replacement, however, hypernatremia develops within a few hours.

Hypotonic Losses

Hypernatremia caused by hypotonic fluid loss is associated with extracellular volume depletion.

Sweat Sweat is a hypotonic solution containing water, sodium, potassium, and chloride. Sweat glands respond to aldosterone by lowering the sodium concentration and increasing the potassium concentration of their secretions.

Gastric fluid losses Fluid lost by vomiting or nasogastric suction is hypotonic to plasma. Without adequate water replacement, large gastric fluid losses can cause hypernatremia.

Osmotic cathartics Fecal losses of water contain electrolytes at a concentration comparable to that of plasma, except when osmotic cathartics such as sorbitol or lactulose are given. These cathartic agents osmotically attract electrolyte-free water to the intestinal lumen, leading to hypotonic fluid losses. Oral sorbitol is a nonabsorbable solute, given with sodium polystyrene sulfonate (Kayexalate) to treat hyperkalemia or with charcoal to treat poisoning; the sorbitol osmotically attracts electrolyte-free water into the intestinal lumen, where it is eliminated in the stool. Similarly, lactulose, which is used to treat hepatic encephalopathy, can promote large electrolyte-free water losses, causing a high incidence of hypernatremia unless the lost water is replaced.

Osmotic diuretics and glycosuria Glucose in the extracellular fluid acts as an effective osmole that attracts water to the extracellular fluid, dehydrating cells and lowering the plasma sodium concentration.[23,51] Excretion of glucose in the urine acts as an osmotic diuretic that can provoke the loss of several liters of hypotonic fluid. Electrolyte-free water losses induced by glycosuria raise the plasma sodium concentration, offsetting the hyponatremic effect of the high blood glucose levels. Intravenous hypertonic mannitol has a similar effect on body fluids.

Acute Salt Poisoning

Water losses increase the serum sodium concentration over hours or days. The oral ingestion of large amounts of salt without water—1 tbsp of salt contains nearly 350 mEq of NaCl, enough to increase the plasma sodium concentration by 8 mEq/L—or the intravenous infusion of hypertonic salt solutions can increase the plasma sodium concentration much more rapidly (i.e., cause acute salt poisoning).[13,51]

DIAGNOSIS

Clinical Manifestations

An acute onset of hypernatremia (seen almost exclusively in acute salt poisoning) causes the brain to shrink, leading to vascular injury and intracranial bleeding. Patients present with seizures, coma, hyperventilation, hyperreflexia, hypertonia, and high fever. Acutely hypernatremic patients with plasma sodium levels above 170 mEq/L often die.[13,51]

Given time to adapt, brain cells protect their volume by accumulating organic osmolytes, preventing the hemorrhages caused by acute hypernatremia. Thus, the clinical manifestations of chronic hypernatremia are less dramatic than those seen in acute salt poisoning, ranging from lethargy to coma, depending on the severity of the electrolyte disturbance.[13,51,53]

The clinical signs of pure-water loss and acute salt poisoning are primarily neurologic. Hypotonic fluid losses may be associated with signs and symptoms of extracellular fluid volume depletion in addition to symptoms related to hypernatremia.

Recognition of Water Deficit

The plasma or serum sodium concentration can be used to determine how much water is needed to restore normotonicity; it seriously underestimates the magnitude of the water deficit in diabetic patients with hyperglycemic dehydration [*see* Diabetic Dehydration, *below*]. In patients without severe hyperglycemia, the percentage increase in the serum sodium concentration approximates the percentage decrease in total body water, as stated more precisely in the following equation:

$$\text{Water deficit} = \text{normal body water} \times (1 - \text{serum } [Na^+]/140)$$

The value for body water is based on the patient's usual body weight (often an estimate), age, and sex.

The calculated water deficit is the amount of water that will return the serum sodium concentration to normal. It reveals nothing about the volume status of the extracellular fluid. Extracellular fluid volume deficits (or surfeits) must be estimated from the history and physical examination, not from the serum sodium concentration.

TREATMENT

Correction of severe extracellular volume depletion takes precedence over correction of hypernatremia. When the patient is hypotensive, initial therapy should include a rapid infusion of isotonic saline to quickly achieve hemodynamic stability. In hemodynamically stable patients, pure-water losses should be replaced with pure water, and isotonic saline is not required. Edematous patients with hypernatremia can be given diuretics along with electrolyte-free water to replace urinary electrolyte-free water losses; the net effect is reduction of the extracellular volume surfeit and restoration of normotonicity and cell volume.

Electrolyte-free water can be given intravenously as D5W to patients who are unable to drink. Dextrose solutions cannot be infused more rapidly than approximately 500 ml/hr. Faster infusions provide more glucose than can be metabolized and therefore cause hyperglycemia, glycosuria, and urinary water losses, which are counterproductive to the correction of hypertonicity. Water replacement should not be based on formulas alone; the serum sodium concentration and urine output should be monitored frequently so that the fluid prescription can be adjusted appropriately.

Rate of Correction

In the vast majority of cases, the onset of hypertonicity is slow enough for brain adaptations to minimize cerebral dehydration. Organic osmolytes that accumulate in the adaptation to hypernatremia are slow to leave the cell during rehydration. If hypernatremia is corrected too rapidly, cerebral edema results.[13,51] To be safe, the serum sodium concentration should be reduced by no more than 10 to 12 mEq/L/day. To achieve the desired rate of correction, electrolyte-free water intake should exceed free-water losses by no more than 2 L daily.

Acute salt poisoning causes devastating brain injury that is largely irreversible. In rare cases when acute salt poisoning can be rapidly diagnosed (e.g., in a case of inadvertent intravenous infusion of hypertonic saline during therapeutic abortion), an effort to prevent a neurologic catastrophe can be made with rapid infusions of electrolyte-free water along with a loop diuretic before the results of the serum electrolyte measurements are known.

Diabetic Dehydration

Hypertonicity associated with diabetes mellitus is a complex disorder.[23] The osmotic diuresis induced by glycosuria results in both saltwater and electrolyte-free water losses; the accumulation of glucose in the extracellular fluid adds impermeant solute, which contributes to hypertonicity and neurologic symptoms. Severely dehydrated hyperglycemic patients may not appear hypovolemic at first, because the high glucose concentration in the extracellular fluid osmotically attracts water from cells, masking the loss of saltwater. With correction of hyperglycemia, marked hypovolemia may emerge. Initial treatment should include 1 to 2 L of isotonic saline in anticipation of this complication, even in patients who are initially normotensive. With volume expansion, excess glucose will be excreted in the urine, creating an ongoing requirement for both saline and electrolyte-free water. An infusion of 0.45% saline at a rate that exceeds urine output will serve to replace the electrolyte-free water deficit and remaining saltwater deficits. The serum sodium concentration, blood glucose level, and urine output should be monitored carefully so that fluid replacement can be tailored to the patient's needs.

Rapid correction of hypertonicity should be avoided in severely hyperglycemic patients because of the risk of cerebral edema. This problem is of particular concern in young children with diabetic ketoacidosis. For this reason, in young patients, the serum sodium concentration should be allowed to increase as the blood glucose level falls (i.e., 0.5 mmol/L increase in serum sodium for every 18 mg/dl decrease in plasma glucose), maintaining a near-constant effective osmolality, especially in the first 12 to 16 hours. In children, hypotonic fluids should probably be avoided during the first day of therapy.[55]

Patients with oliguric renal failure do not become dehydrated when they become severely hyperglycemic. Such patients often experience hypertension or congestive heart failure because of fluid shifts from cells to the extracellular fluid. Even after adjusting for the effect of hyperglycemia, the serum sodium concentration is often low. Insulin is the only required treatment; neither isotonic saline nor 0.45% saline is indicated.

Disorder of Water Conservation: Diabetes Insipidus

PATHOGENESIS

Diabetes insipidus (DI) may be neurogenic or nephrogenic. Neurogenic DI is caused by deficient secretion of vasopressin[8,56,57]; nephrogenic DI results from the kidney's unresponsiveness to normally secreted hormone.[58] In both disorders, patients present with polyuria (loosely defined as the passage of excessive volumes of urine—generally more than 3 to 4 L daily) and polydipsia (excessive thirst). Most patients with polyuria do not become hypernatremic, because thirst maintains electrolyte-free water balance. The causes and treatment of neurogenic, or central, DI are discussed elsewhere [*see 3:V Pituitary*].

Defective responsiveness to vasopressin (nephrogenic DI) may be inherited as an X-linked trait, caused by a mutation in the gene encoding for the V_2 vasopressin receptor, or as an autosomal recessive trait caused by a mutation in the gene encoding for the vasopressin-responsive water channel (aquaporin 2). Acquired nephrogenic DI may be caused by lithium or demeclocycline therapy, hypokalemia, or hypercalcemia; or it may complicate a number of renal diseases [*see Table 6*]. Vasopressin-resistant DI may emerge during the late stage of pregnancy as a result of vasopressinase released by the placenta; many affected patients have underlying, subclinical partial neurogenic or nephrogenic DI that has been exacerbated by increased catabolism of circulating vasopressin.

DIAGNOSIS

Clinical Manifestations

Patients with DI complain of polyuria, nocturia (the need to urinate during the night), and polydipsia. The only significant physical findings or laboratory abnormalities are those of the underlying cause.

Laboratory Tests

A diagnosis of DI can be made if the urine osmolality is less

than 250 mOsm/kg despite hypernatremia.[8] When the disorder is suspected in a polyuric patient whose serum sodium concentration is normal or borderline, the urine osmolality level can be monitored while the patient is deprived of water, allowing the serum sodium level to increase to higher than 144 mEq/L. Exogenous vasopressin increases urine osmolality by more than 150 mOsm/kg in patients with neurogenic, but not nephrogenic, DI. It is possible to misdiagnose DI in patients who actually have a primary thirst disorder. Excessive water intake suppresses vasopressin secretion and causes polyuria with dilute urine. Because patients with primary polydipsia secrete vasopressin normally, they do not become hypernatremic during diagnostic water deprivation. Correlation with plasma vasopressin levels is often necessary in borderline cases. Polyuric patients whose urine osmolality equals or exceeds plasma osmolality should be distinguished from patients with DI; polyuria in such cases is usually caused by excessive excretion of salt, urea, or glucose or by an osmotic diuretic (solute diuresis).

Magnetic resonance imaging of the brain can be helpful in the evaluation of patients with suspected DI.[8,57] In 85% to 90% of healthy adults and children, the posterior pituitary emits a hyperintense signal, or so-called bright spot, on T_1-weighted magnetic resonance images, apparently related to the vasopressin release of the gland. This bright spot is also normal in 85% to 90% of patients with primary polydipsia, but it is almost always absent or greatly diminished in patients with pituitary DI.

TREATMENT

When access to water is limited, patients with DI are more susceptible to dehydration than normal persons. Thus, water losses must be carefully replaced during superimposed illnesses. Neurogenic DI is best treated with a synthetic antidiuretic hormone, 1-desamino-8-D-arginine vasopressin (DDAVP), which can be given parenterally or intranasally.[8,57] Chlorpropamide or carbamazepine (both of which enhance vasopressin action) or thiazide diuretics (which limit the ability to maximally dilute the urine) can be used in patients with mild disease. Limiting dietary salt and protein intake is also helpful. Nephrogenic DI can also be treated with dietary measures or with thiazides and indomethacin (which helps concentrate the urine by inhibiting prostaglandin synthesis). Lithium-induced nephrogenic DI may be improved by amiloride, which blocks lithium entry into the collecting duct cell.

Disorder of Saltwater Excess: Edematous States

Edema, a swelling of the soft tissues that can be indented or pitted by the examiner's fingers, is the clinical manifestation of an expanded interstitial fluid volume. To be detected clinically, interstitial volume must increase by at least 2.5 to 3 L, nearly equaling the total amount of fluid in the intravascular space. Thus, generalized edema requires an increase in the total amount of saltwater in the extracellular space, and it implies retention of dietary or infused sodium, with an impaired ability to excrete saltwater.

PATHOGENESIS

Excess fluid collects in the interstitial space in response to Starling forces, which govern the movement of extracellular fluid into and out of the vasculature. Edema occurs when there is increased capillary blood pressure, decreased plasma oncotic pressure, increased capillary permeability to protein, or obstruction to lymph flow.

When fluid overflows from an overfilled vascular space into the interstitium, impaired sodium excretion is clearly implicated. However, even when edema formation results from decreased oncotic pressure or increased capillary permeablity, renal sodium retention is required to replace the saltwater that has been lost from the vasculature.

ETIOLOGY

Primary Renal Sodium Retention: Edema Caused by Renal Disease

Nephrotic syndrome The nephrotic syndrome is characterized by heavy urinary protein losses (in excess of 3 g/day), hypoalbuminemia, and edema.[59] The syndrome, which can be seen in a variety of glomerular diseases, is caused by increased permeability of the glomerular capillary to protein. Traditionally, edema in the nephrotic syndrome has been ascribed to decreased plasma oncotic pressure. This no longer appears to be the sole explanation. Correction of hypoalbuminemia by infusing albumin does not consistently improve the edema, and in steroid-responsive cases, edema may resolve before hypoalbuminemia improves. Thus, in most patients, primary sodium retention by the kidney, independent of an underfilled vasculature, plays a major contributing role.

Nephritic edema Glomerular diseases characterized by proliferation of mesangial cells (e.g., diffuse proliferative glomerulonephritis and membranoproliferative glomerulonephritis) often cause primary sodium retention that is not associated with heavy proteinuria or hypoalbuminemia (the nephritic syndrome). Patients with nephritic edema are typically hypertensive and may present with congestive heart failure because of an overexpanded vascular volume.

Secondary Renal Sodium Retention: Edema Caused by Extrarenal Disease

In congestive heart failure and hepatic cirrhosis, the body responds as if it were volume depleted. Despite an expanded interstitial fluid volume, as well as increased total body sodium content, the kidney avidly retains salt and water. The normal renal response to a high salt intake is lost, and progressive salt retention occurs. In these conditions, volume-regulatory mechanisms are responding to reduced fullness of the arterial portion of the vascular system, which normally contains about 15% of the total blood volume.

Congestive heart failure Advanced stages of the many disorders that affect the pericardium, myocardium, or heart valves can produce congestive heart failure, a disorder characterized by renal sodium retention and interstitial edema in systemic or pulmonary capillary beds.[16,17] Arterial receptors are activated when cardiac output falls (low-output heart failure) or when cardiac output is not high enough to compensate for decreased peripheral resistance (high-output failure).

Cirrhosis Patients with severe liver disease may exhibit profound salt retention, often excreting less than 10 mEq of sodium in the urine each day. Scarring of the hepatic parenchyma increases resistance to blood flow in the postsinusoidal venules, resulting in high sinusoidal pressures and venous hypertension throughout the portal system. Portal hypertension and hypoalbuminemia promote the formation of ascites. In addition, cirrho-

sis results in vasodilatation from endotoxins, vasodilatory prostaglandins, nitric oxide, various gut hormones, and other mediators. The combined effects of blood pooling in the splanchnic circulation (caused by portal hypertension) and systemic vasodilatation lead to underfilling of the arterial circulation and subsequent activation of sodium-retaining factors.[60]

Idiopathic edema Idiopathic edema is a benign disorder of young, menstruating women who have no cardiac, hepatic, or renal disease.[61] Fluid retention often begins premenstrually and then becomes persistent. Depression and neurotic symptoms are commonly present, and affected patients are often weight conscious and markedly concerned about even minor degrees of edema. Some patients episodically fast for days at a time and then accumulate edema on refeeding. In many others, diuretics play an important role in the pathogenesis of idiopathic edema. Long-term diuretic or cathartic use leads to persistent hypovolemia and chronic activation of sodium-retaining mechanisms, which include hypertrophy of the nephron segments distal to the site of action of the diuretic. When the diuretic is stopped, marked sodium retention occurs because the sodium-retaining forces cannot be shut off rapidly. The patient thus becomes convinced of the need for diuretics, and the cycle continues.

DIAGNOSIS

The symptoms and laboratory findings associated with edematous conditions depend on the underlying cause. Dyspnea on exertion and orthopnea provoked by pulmonary interstitial edema are prominent features in patients with left ventricular failure or nephritic edema, but these symptoms are usually absent when edema is caused by right heart failure, nephrotic syndrome, or cirrhosis. Mild peripheral edema develops in the dependent portions of the anatomy and is usually asymptomatic. Although more severe edema, which can extend to the thighs and buttocks, may be uncomfortable, it is usually harmless. A large volume of ascites not only causes discomfort but also may elevate the diaphragm, causing shortness of breath; may promote reflux of gastric fluid, causing bleeding from esophageal varices; or may become infected spontaneously.

The diagnosis of edema should be approached systematically; the physician should look for evidence of heart, renal, or liver disease. A diagnosis can usually be made from the history and physical examination, urinalysis, liver function tests, and chest x-ray. More puzzling cases may require echocardiography or, rarely, right heart catheterization. Plasma levels of B-type natriuretic peptide (BNP) are increased in patients with heart failure. Used in conjunction with other clinical information, rapid-measurement BNP is useful in establishing or excluding the diagnosis of congestive heart failure in patients who present with acute dyspnea.[62]

TREATMENT

Dietary salt restriction is important for patients with edema, but this measure alone is impractical or insufficient when urinary sodium excretion is reduced to very low levels. Thus, most edematous patients whose underlying condition cannot be reversed require treatment with diuretics.[5,63] Salt restriction and diuretics are adjunctive treatments for heart failure. Therapy is also directed at improving cardiac performance by using digoxin and vasodilators to reduce afterload. Recombinant human BNP (nesiritide) has become available for the treatment of acute decompensated heart failure. The agent reduces pulmonary capillary wedge pressure and systemic vascular resistance, improves cardiac performance, and has a diuretic effect, in part because of its effect on sodium reabsorption in the distal nephron.[64]

Ascitic fluid is a separate compartment of the extracellular fluid compartment that is much more difficult to mobilize than peripheral edema. Thus, cirrhotic patients who have ascites but no peripheral edema are susceptible to intravascular volume depletion when they are treated with diuretics; weight loss should thus be limited to 0.5 kg daily. Repeated large-volume paracenteses combined with intravenous albumin is a safe and effective alternative to diuretics that avoids intravascular volume depletion.[65] The subsequent administration of diuretics prevents reaccumulation of ascitic fluid.

Use of Diuretics

Diuretics increase saltwater excretion by impairing tubular reabsorption of the sodium filtered by the glomerulus. The diuretic effect is dose-dependent; the maximum response is determined by the diuretic's site of action within the nephron, the filtered load of sodium, and the amount of sodium reabsorbed by nephron segments unaffected by the diuretic.

Mechanism of action All diuretics except spironolactone are specific inhibitors of luminal transporters and must gain access to the tubular fluid to block sodium reabsorption.[5,63] Because diuretic agents are highly protein bound, they are not readily filtered at the glomerulus; instead, they are actively transported into the urine by the organic acid (in the case of agents such as acetazolamide, thiazides, and loop diuretics) or organic base (in the case of agents such as amiloride and triamterene) via secretory pumps in the proximal tubule. A dose-response curve links the amount of drug reaching the urine to the amount of sodium excretion that is elicited [*see Figure 4*]. Spironolactone binds to the cytosolic receptor for aldosterone, and its diuretic action, unlike that of other diuretics, does not depend on secretion into the tubular lumen.

The most potent agents are those that block sodium transport in the loop of Henle. At high doses, loop diuretics almost totally block sodium reabsorption in this nephron segment, causing about 20% of the filtered load of sodium to be excreted in the urine; at low glomerular filtration rates, the same percentage of filtered sodium is excreted, but the total amount is reduced. Conditions such as volume depletion, heart failure, and cirrhosis, which cause avid sodium reabsorption in the proximal and distal tubules, blunt the maximum response to the diuretic. Because gastrointestinal absorption of diuretics is often delayed in edematous conditions (presumably because of bowel edema), higher oral doses must be used to achieve adequate blood levels. In renal disease and cirrhosis, organic anions such as hippurate and bile acids compete with the diuretic for secretion into the proximal tubule; thus, higher plasma levels may be required to achieve adequate drug levels in the urine. Similarly, severe hypoalbuminemia can diminish drug secretion into the tubular lumen, because albumin binding of most diuretics maximizes the rate of diuretic delivery to the organic anion secretory pump in the proximal tubule. Reduced renal blood flow also limits delivery of drug to the tubular lumen. Some patients with advanced cirrhosis who are resistant to furosemide respond to spironolactone, a generally weak diuretic whose effectiveness does not depend on tubular secretion.

Agents that act in the proximal tubule, loop of Henle, or distal tubule cause potassium wasting and hypokalemia because they

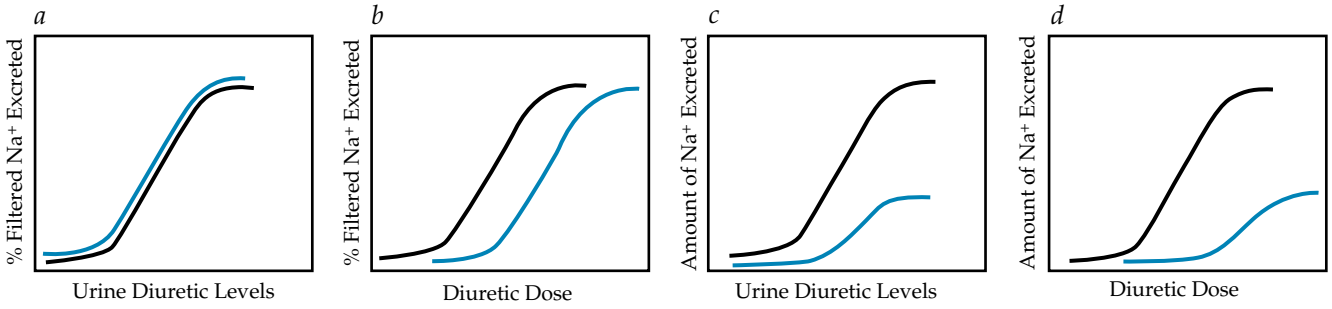


Figure 4 **Dose-response curves for a loop diuretic in patients with normal (black line) and reduced (blue line) renal function. Urinary sodium excretion responds to diuretic levels in the tubular lumen (as reflected by urinary drug levels). The diuretic effect reaches a maximum at approximately 20% of the filtered load of sodium regardless of renal function (*a*). Secretion of the diuretic into the tubular lumen is reduced in renal failure; thus, higher doses are required in azotemic patients to achieve the same urinary drug levels found in patients with normal renal function (*b*). Because the filtered load of sodium is reduced in patients with renal failure, absolute sodium excretion is also reduced, even at high doses (*c* and *d*).**

increase delivery of tubular fluid to the cortical collecting tubule, where potassium secretion is flow dependent. Potassium-sparing diuretics, which act in the cortical collecting tubule, cause hyperkalemia because sodium reabsorption at this site favors potassium secretion. The carbonic anhydrase inhibitor acetazolamide causes metabolic acidosis, as do the potassium-sparing diuretics. Thiazides and loop diuretics cause metabolic alkalosis because of increased distal delivery of sodium to sites where sodium reabsorption stimulates hydrogen ion secretion.

Clinical strategies Diuretic doses should be adjusted to achieve explicit therapeutic goals. Outpatient therapy is usually designed to produce a gradual loss of fluid, with the dose being increased until a desired target weight is reached. The patient should be instructed to keep a daily log that records weight and diuretic dose. Patients are instructed to stop the diuretic if their weight falls too low, resuming at a lower dose when enough saltwater has been retained to restore the target weight.

Inpatient diuretic management should also employ the target-weight concept, but dose adjustments can be made more often and more aggressively, particularly at the start of therapy. It is important to rapidly define the dose that can deliver enough drug to the tubular lumen to reach the steep portion of the dose-response curve. Once an effective dose is defined, larger doses of diuretic provide little benefit. If a greater response is needed, the effective dose should be repeated several times during the day, or alternatively, a continuous infusion can be given to maintain effective urinary drug levels. Continuous infusion of loop diuretics induces a slightly larger natriuretic response than does bolus administration and is associated with a shorter hospital stay in patients with advanced heart failure.[66]

Diuretic resistance Resistance to high doses of loop diuretics may be overcome by administering loop diuretics in combination with a thiazide or metolazone. Acetazolamide may be used along with or in place of a thiazide or metolazone. This strategy blocks sodium reabsorption at several sites along the nephron, avoiding resistance caused by increased sodium reabsorption proximal or distal to the loop of Henle. Careful monitoring is extremely important, because these combinations can be extremely potent, causing large potassium and sodium losses.

Diuretic complications All diuretic agents may cause volume depletion and azotemia, but these complications are most likely to occur with loop diuretics.[4] Hypokalemic alkalosis, hyperglycemia, and hyperuricemia (sometimes with clinical gout) are common dose-dependent complications of both thiazides and loop diuretics. Thiazides decrease calcium excretion and may cause hypercalcemia in patients with underlying conditions that increase gastrointestinal calcium absorption (e.g., sarcoidosis) or bone reabsorption (e.g., hyperparathyroidism). Thiazides are also much more likely to cause hyponatremia than other agents and should be avoided in patients who habitually drink large amounts of fluid. Potassium-sparing agents (e.g., triamterene, amiloride, and spironolactone) may cause hyperkalemia; these agents should generally not be given with potassium supplements, and they should be used with caution in patients with renal insufficiency (particularly diabetic nephropathy) and patients taking ACE inhibitors or angiotensin receptor blockers. Loop diuretics can predispose to hearing loss, particularly when high doses are administered by bolus injection to patients receiving other ototoxic drugs.[35] Hearing loss from ethacrynic acid is more likely to be permanent.

Disorder of Saltwater Deficiency: Volume Depletion

PATHOGENESIS

Volume depletion occurs when saltwater is lost from the extracellular fluid at a rate that exceeds intake. Saltwater can be lost from the gastrointestinal tract, kidney, or skin, or it can result from extravascular sequestration (third-space losses) in the abdominal cavity or in traumatized tissues.

Underfilling of the arterial circulation triggers a cascade of physiologic responses that preserve blood flow to vital organs. Volume receptors and baroreceptors activate the sympathetic nervous system and the renin-angiotensin-aldosterone system. Except when renal salt wasting is the cause, these responses reduce urinary sodium excretion so that nearly all ingested salt is retained. Volume-depleted persons also become thirsty; ingested water is retained because vasopressin, released in response to volume depletion, concentrates the urine, decreasing water excretion. The plasma sodium concentration can be high, normal, or low in volume-depleted persons, depending on electrolyte-free water intake and excretion. Vasoconstriction maintains the systemic blood pressure and also reduces renal blood flow. Initially, efferent arteriolar resistance, mediated by angiotensin II, predominates, sustaining intraglomerular pressure and the

glomerular filtration rate; in more severe hypovolemia, renal blood flow is further reduced and glomerular filtration falls.

ETIOLOGY

Because renal sodium conservation can reduce urinary sodium losses to less than 10 mmol/day, volume depletion is unlikely to occur from decreased intake alone. The small bowel and colon are the most common sources of isotonic fluid loss. Spectacular amounts of isotonic saltwater can be lost in diarrhea. For example, rice-water stool losses in cholera can reach 20 L/day, causing death within a few hours without fluid replacement. Small bowel obstruction causes pooling of several liters of saltwater within the bowel lumen. Fluid may also be sequestrated in the abdominal cavity in patients with pancreatitis or peritonitis. Sequestration of fluid in the soft tissues may also complicate crush injuries with rhabdomyolysis or burns.

Renal salt wasting can cause volume depletion, but only a few disorders can cause enough renal salt loss to be clinically apparent. Diuretics and osmotic diuresis caused by glycosuria are the most frequent causes of renal salt wasting. Transient renal salt wasting may occur in the recovery phases of acute tubular necrosis or obstructive uropathy, and it can also occur in toxic nephropathies. Renal salt wasting also occurs in adrenal insufficiency.

DIAGNOSIS

Clinical Manifestations

Minor degrees of volume depletion (less than 10% of plasma volume, equivalent to the loss of one unit of blood) cause an increase in heart rate and may also be associated with complaints of fatigue, thirst, or muscle cramps. With modest hypovolemia, arteriolar vasoconstriction is sufficient to maintain the blood pressure when the patient is recumbent. However, dizziness and hypotension emerge on standing or during physical exertion. Severe fluid losses cause hypotension in recumbency and, ultimately, signs of tissue ischemia and shock (e.g., cool, clammy extremitites, decreased urine output, lethargy, and confusion). Irreversible tissue injury may occur if this condition is allowed to continue.

Loss of weight within a short period is the most reliable sign of volume depletion. Physical findings include a low jugular venous pulse rate and orthostatic changes in blood pressure and heart rate.[67,68] However, because postural hypotension can occur in up to 30% of normovolemic persons older than 65 years, these changes must be interpreted with caution. Decreased skin turgor and dry mucous membranes are generally unreliable findings in volume-depleted adults; these signs can be absent in severe hypovolemia, and they can be present (particularly in mouth breathers and the elderly) when the patient is actually volume overloaded. The presence of edema makes true volume depletion unlikely.

Laboratory Tests

Laboratory findings are related to the decreased volume of intravascular saltwater and to decreased renal perfusion. The hematocrit increases in proportion to the contraction of plasma volume, and the serum albumin may be increased as well. Urinary sodium is usually less than 20 mEq/L except in metabolic alkalosis (in which the urine chloride is low) or when renal sodium wasting is the cause of the condition.[44,69] Renal blood flow is reduced, but unless the patient is frankly hypotensive, the glomerular filtration rate is maintained by vasoconstriction of the efferent glomerular arteriole. Thus, except in severe volume depletion, the serum creatinine changes very little. Unlike creatinine, urea is reabsorbed from the glomerular filtrate. Thus, in volume depletion (prerenal azotemia), the BUN is increased disproportionately to the increase in creatinine.[69] Azotemia may be blunted in patients with a poor dietary-protein intake and may be exacerbated in patients who are catabolic, bleeding, or receiving steroid therapy.

TREATMENT

Patients with mild volume depletion can be treated by increasing their dietary intake of salt, relying on normal thirst mechanisms to provide the appropriate amount of water. For most patients, the familiar (but misguided) order to drink fluids should be replaced with an order to salt one's food. Even severe volume depletion can be treated with oral solutions containing electrolytes, sugar, and amino acids.[70] Glucose and amino acids promote intestinal absorption of sodium through cotransport mechanisms similar to those found in the proximal tubule of the kidney. Rice-based oral replacement solutions have been a major advance in the treatment of diarrhea in developing countries.

Intravenous fluids are necessary when fluids cannot be taken orally. If the patient is hypotensive, isotonic saline should be given as rapidly as possible until tissue perfusion is adequate. Colloid-containing solutions have no proven advantage over crystalloids.[71] There is no accurate way to estimate the total fluid deficit in hypovolemia other than continued clinical observation of the patient's response to therapy.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Edelman IS, Leibman J: Anatomy of body water and electrolytes. Am J Med 27:256, 1959

2. Oster JR, Singer I: Hyponatremia, hyposmolality, and hypotonicity: tables and fables. Arch Intern Med 159:333, 1999

3. Mange K, Matsuura D, Cizman B, et al: Language guiding therapy: the case of dehydration versus volume depletion. Ann Intern Med 127:848, 1997

4. Greger R: Physiology of renal sodium transport. Am J Med Sci 319:51, 2000

5. Rasool A, Palevsky PM: Treatment of edematous disorders with diuretics. Am J Med Sci 319:25, 2000

6. Scheinman SJ, Guay-Woodford LM, Thakker RV, et al: Genetic disorders of renal electrolyte transport. N Engl J Med 340:1177, 1999

7. Kozono D, Yasui M, King LS, et al: Aquaporin water channels: atomic structure molecular dynamics meet clinical medicine. J Clin Invest 109:1395, 2002

8. Robertson G: Antidiuretic hormone: normal and disordered functions. Endocrinol Metab Clin North Am 30:671, 2001

9. Levin ER, Gardner DG, Samson WK: Natriuretic peptides. N Engl J Med 339:321, 1998

10. McManus ML, Churchwell KB, Strange K: Regulation of cell volume in health and disease. N Engl J Med 333:1260, 1995

11. Sterns RH, Silver SM, Spital A: Hyponatremia. The Kidney: Physiology and Pathophysiology. Seldin DW, Giebisch G, Eds. Lippincott Williams & Wilkins, Philadelphia, 2000

12. Han DS, Cho BS: Therapeutic approach to hyponatremia. Nephron 92(suppl 1):9, 2002

13. Adrogue HJ, Madias NE: Hypernatremia. N Engl J Med 342:1493, 2000

14. Kawai N, Baba A, Suzuki T, et al: Roles of arginine vasopressin and atrial natriuretic peptide in polydipsia-hyponatremia of schizophrenic patients. Psychiatry Res 101:39, 2001

15. Fenves AZ, Thomas S, Knochel JP: Beer potomania: two cases and review of the literature. Clin Nephrol 45:61, 1996

16. Cadnapaphornchai MA, Gurevich AK, Weinberger HD, et al: Pathophysiology of sodium and water retention in heart failure. Cardiology 96:122, 2001

17. Schrier RW, Abraham WT: Hormones and hemodynamics in heart failure. N Engl J Med 341:577, 1999

18. Cardenas A, Gines P: Pathogenesis and treatment of fluid and electrolyte imbalance

in cirrhosis. Semin Nephrol 21:308, 2001

19. Schwartz WB, Bennett W, Curelop S, et al: A syndrome of renal sodium loss and hyponatremia probably resulting from inappropriate secretion of antidiuretic hormone. J Am Soc Nephrol 12:2860, 2001

20. Smith D, Moore K, Tormey D, et al: Downward resetting of the osmotic threshold for thirst in patients with SIADH. Am J Physiol Endocrinol Metab 287:E1019, 2004

21. Steele A, Gowrishankar M, Abrahamson S, et al: Postoperative hyponatremia despite near-isotonic saline infusion: a phenomenon of desalination. Ann Intern Med 126:20, 1997

22. Hillier TA, Abbott RD, Barrett EJ: Hyponatremia: evaluating the correction factor for hyperglycemia. Am J Med 106:399, 1999

23. Davids MR, Edoute Y, Stock S, et al: Severe degree of hyperglycaemia: insights from integrative physiology. QJM 95:113, 2002

24. Istre O, Jellum E, Skajaa K, et al: Changes in amino acids, ammonium, and coagulation factors after transcervical resection of the endometrium with a glycine solution used for uterine irrigation. Am J Obstet Gynecol 172:939, 1995

25. Turchin A, Seifter JL, Seely EW: Clinical problem solving. Mind the gap. N Engl J Med 349: 1365, 2003

26. Verbalis JG: Adaptation to acute and chronic hyponatremia: implications for symptomatology, diagnosis, and therapy. Semin Nephrol 18:3, 1998

27. Gowrishankar M, Lin SH, Mallie JP, et al: Acute hyponatremia in the perioperative period: insights into its pathophysiology and recommendations for management. Clin Nephrol 50:352, 1998

28. Spital A, Ristow S: Cyclophosphamide induced water intoxication in a woman with Sjögren's syndrome. J Rheumatol 24:2473, 1997

29. Almond CS, Shin AY, Fortescue EB, et al: Hyponatremia among runners in the Boston Marathon. N Engl J Med 352:1550, 2005

30. Ayus JC, Varon J, Arieff AI: Hyponatremia, cerebral edema, and noncardiogenic pulmonary edema in marathon runners. Ann Intern Med 132:711, 2000

31. Hartung TK, Schofield E, Short AI, et al: Hyponatraemic states following 3,4-methylenedioxymethamphetamine (MDMA, 'ecstasy') ingestion. QJM 95:431, 2002

32. Shafiee MA, Charest AF, Cheema-Dhadli S, et al: Defining conditions that lead to the retention of water: the importance of the arterial sodium concentration. Kidney Int 67:613, 2005

33. Sterns RH: Severe symptomatic hyponatremia: treatment and outcome: a study of 64 cases. Ann Intern Med 107:656, 1987

34. Sonnenblick M, Friedlander Y, Rosin AJ: Diuretic-induced severe hyponatremia: review and analysis of 129 reported patients. Chest 103:601, 1993

35. Akalin E, Chandrakantan A, Keane J, et al: Normouricemia in the syndrome of inappropriate antidiuretic hormone secretion. Am J Kidney Dis 37:E8, 2001

36. Sorensen JB, Andersen MK, Hansen HH: Syndrome of inappropriate secretion of antidiuretic hormone (SIADH) in malignant disease. J Intern Med 238:97, 1995

37. Diringer MN, Wu KC, Verbalis JG, et al: Hypervolemic therapy prevents volume contraction but not hyponatremia following subarachnoid hemorrhage. Ann Neurol 31:543, 1992

38. Maesaka JK, Gupta S, Fishbane S: Cerebral salt-wasting syndrome: does it exist? Nephron 82:100, 1999

39. Hanna FW, Scanlon MF: Hyponatraemia, hypothyroidism, and role of arginine-vasopressin. Lancet 350:755, 1997

40. Wilkinson TJ, Begg EJ, Sainsbury R: Incidence and risk factors for hyponatremia following treatment with fluoxetine or paroxetine in elderly people. Br J Clin Pharmacol 47:211, 1999

41. Vitting KE, Gardenswartz MH, Zabetakis PM, et al: Frequency of hyponatremia and nonosmolar vasopressin release in the acquired immunodeficiency syndrome. JAMA 263:973, 1990

42. Gebers AL, Gulberg V, Gines P, et al: Therapy of hyponatremia in cirrhosis with a vasopressin receptor antagonist: a randomized double-blind multicenter trial. Gastroenterology 124:933, 2003

43. Sterns RH, Cappuccio JD, Silver SM, et al: Neurologic sequelae after treatment of severe hyponatremia: a multicenter perspective. J Am Soc Nephrol 4:1522, 1994

44. Narins RG, Jones ER, Stom MC, et al: Diagnostic strategies in disorders of fluid, electrolyte and acid-base homeostasis. Am J Med 72:496, 1982

45. Anpalahan M: Chronic idiopathic hyponatremia in older people due to the syndrome of inappropriate antidiuretic hormone secretion (SIADH) possibly related to aging. J Am Geriatr Soc 49:788, 2001

46. Greenberg A: Diuretic complications. Am J Med Sci 319:10, 2000

47. Shimizu K: Combined effects of vasopressin V2 receptor antagonist and loop diuretics in humans. 59:164, 2003

48. Sterns RH, Riggs JE, Schochet SS: Osmotic demyelination syndrome following correction of hyponatremia. N Engl J Med 314:1535, 1986

49. Laureno R, Karp BI: Myelinolysis after correction of hyponatremia. Ann Intern Med 126:57, 1997

50. Brown WD: Osmotic demyelination disorders: central pontine and extrapontine myelinolysis. Curr Opin Neurol 13:691, 2000

51. Feig PU, McCurdy DK: The hypertonic state. N Engl J Med 297:1444, 1977

52. Kahn T: Hypernatremia with edema. Arch Intern Med 159:93, 1999

53. Palevsky PM, Bhagrath R, Greenberg A: Hypernatremia in hospitalized patients. Ann Intern Med 124:197, 1996

54. Sterns RH: Hypernatremia in the intensive care unit: instant quality—just add water. Crit Care Med 27:1041, 1999

55. Carlotti APCP, Bohn D, Halperin ML: Importance of timing of risk factors for cerebral oedema during therapy for diabetic ketoacidosis. Arch Dis Child 88:170, 2003

56. Maghnie M, Cosi G, Genovese E, et al: Central diabetes insipidus in children and young adults. N Engl J Med 343:998, 2000

57. Singer I, Oster JR, Fishman LM: The management of diabetes insipidus in adults. Arch Intern Med 157:1293, 1997

58. Bichet DG: Nephrogenic diabetes insipidus. Am J Med 105:431, 1998

59. De Santo NG, Pollastro RM, Saviano C, et al: Nephrotic edema. Semin Nephrol 21:262, 2001

60. Martin PY, Gines P, Schrier RW: Nitric oxide as a mediator of hemodynamic abnormalities and sodium and water retention in cirrhosis. N Engl J Med 339:533, 1998

61. Kay A, Davis CL: Idiopathic edema. Am J Kidney Dis 34:405, 1999

62. Maisel AS, Krishnaswamy P, Nowak RM, et al: Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. Breathing Not Properly Multinational Study Investigators. N Engl J Med 347:161, 2002

63. Brater DC: Diuretic therapy. N Engl J Med 339:387, 1998

64. Colucci WS, Elkayam U, Horton DP, et al: Intravenous nesiritide, a natriuretic peptide, in the treatment of decompensated congestive heart failure. Nesiritide Study Group. N Engl J Med 343:246, 2000

65. Peltekian KM, Wong F, Liu PP, et al: Cardiovascular, renal, and neurohumoral responses to single large-volume paracentesis in patients with cirrhosis and diuretic-resistant ascites. Am J Gastroenterol 92:394, 1997

66. Howard PA, Dunn MI: Aggressive diuresis for severe heart failure in the elderly. Chest 119:807, 2001

67. Carlson JE: Assessment of orthostatic blood pressure: measurement technique and clinical applications. South Med J 92:167, 1999

68. McGee S, Abernethy WB, Simel DL: The rational clinical examination. Is this patient hypovolemic? JAMA 281:1022, 1999

69. Klahr S, Miller SB: Acute oliguria. N Engl J Med 338:671, 1998

70. Farthing MJ: Oral rehydration: an evolving solution. J Pediatr Gastroenterol Nutr 34:S64, 2002

71. Schierhout G, Roberts I: Fluid resuscitation with colloid or crystalloid solutions in critically ill patients: a systematic review of randomised trials. BMJ 316:961, 1998

Acknowledgment

Figures 3 and 4 Marcia Kammerer.

II DISORDERS OF ACID-BASE AND POTASSIUM BALANCE

Robert M. Black, M.D.

Acid-Base Disorders

The blood pH is normally maintained at 7.38 to 7.42. Any deviation from this range indicates a change in the hydrogen ion concentration ($[H^+]$) because blood pH is the negative logarithm of $[H^+]$, as expressed by the following equation:

$$pH = -\log_{10}[H^+]$$

The $[H^+]$ at a physiologic blood pH of 7.40 is 40 nEq/L [*see Figure 1*]. An increase in the $[H^+]$—a fall in the blood pH—is termed acidemia. A decrease in the $[H^+]$—a rise in the blood pH—is termed alkalemia. The disorders that cause these changes in the blood pH are acidosis and alkalosis, respectively. Because abnormalities of acid-base metabolism are often associated with potassium imbalance, clinical approaches to hypokalemia and hyperkalemia are also discussed in this chapter.

Normal Acid-Base Physiology

The normal adult diet generates an excess 70 to 100 mEq of acid that must be eliminated every day. Failure to do so results in a persistent fall in the blood pH resulting from a rise in the plasma H^+ ion concentration. The balance of acid-base homeostasis is maintained in part by the relation between the arterial carbon dioxide tension (P_aCO_2) and plasma bicarbonate concentration ($[HCO_3^-]$), as noted in the following equation, a nonlogarithmic expression of the Henderson-Hasselbach equation[1]:

$$[H^+] = 24 \times P_aCO_2/[HCO_3^-]$$

A fall in plasma $[HCO_3^-]$ caused by either gastrointestinal or renal bicarbonate losses also increases the $[H^+]$ and lowers blood pH.

Renal Reabsorption of Bicarbonate

The plasma $[HCO_3^-]$ is normally maintained at approximately 25 mEq/L by the daily reabsorption of the filtered bicarbonate load (about 4,500 mEq) by the kidneys. If the filtered bicarbonate were not reabsorbed, the plasma $[HCO_3^-]$ would fall, as would the blood pH. Thus, maintenance of a normal plasma $[HCO_3^-]$ requires reabsorption of essentially all of the bicarbonate filtered across the glomerular capillaries each day.

Most bicarbonate reabsorption (almost 90%) occurs in the proximal convoluted tubule [*see Figure 2*]; in contrast, the distal nephron reclaims very little bicarbonate. The difference is a result of a complex process, which is facilitated by the greater quantity of carbonic anhydrase in the lumen of the proximal tubule.

Renal Excretion of Acid

In addition to reabsorbing essentially all filtered bicarbonate, the kidneys excrete the daily dietary acid load, derived mainly from sulfur-containing amino acids. The hydrogen ions that are excreted in the final urine are secreted mainly in the collecting tubules [*see Figure 3*]. This secretory process is facilitated indirectly by aldosterone and directly by acid-sensing renal tubular cells.[2]

The daily acid load is excreted into the collecting tubules by the H^+-ATPase pumps located in the luminal membrane of the intercalated cells. This secretory process is inhibited by a trivial quantity of free hydrogen ions that lower the urine pH below the critical level of 4.0 to 4.5. This limitation is normally overcome by the presence of urinary buffers that combine with free hydrogen ions, thus permitting continued secretion of acid. There are several urinary buffers, the most important of which is ammonia because it is the only buffer that can increase substantially in the presence of an acid load. Limitation of the capacity to generate adequate urinary ammonia, as occurs in renal insufficiency, usually leads to acidosis.

The major site of ammonia production in the kidney is the proximal tubule [*see Figure 4*]; ammonia moves from the proximal tubule to the collecting tubule, where it is eliminated [*see Figure 5*]. The quantity of ammonia produced is stimulated both by acidemia and by hypokalemia. Conversely, alkalemia and hyperkalemia limit renal tubular ammonia production and acid excretion.

Metabolic Acidosis

Metabolic acidosis results whenever a primary decrease in the plasma $[HCO_3^-]$ occurs. Such a decrease may be caused by several factors: exogenous acid administration, endogenous acid production, impaired renal hydrogen secretion, and bicarbonate losses from the kidney or in gastrointestinal secretions. Calculation of the plasma anion gap is particularly useful in identifying the specific cause of metabolic acidosis and in narrowing the differential diagnosis.

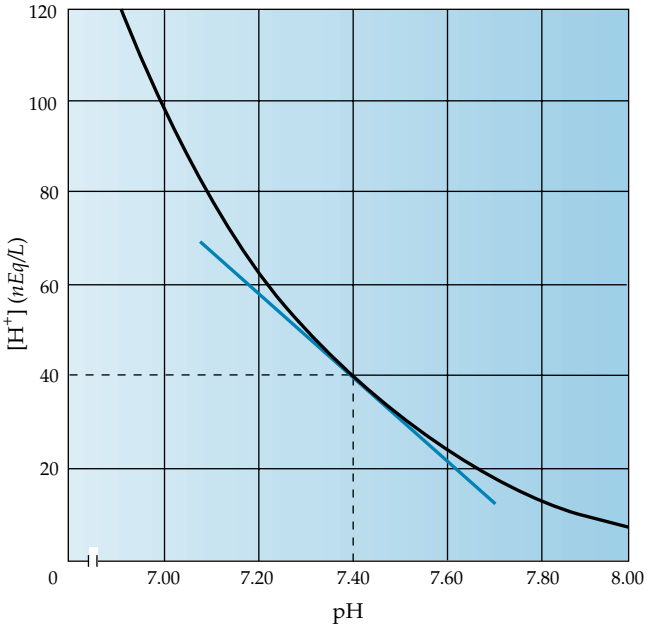


Figure 1 **The relation between the plasma hydrogen ion concentration ($[H^+]$) and the pH of the blood ($pH = -\log_{10}[H^+]$).**

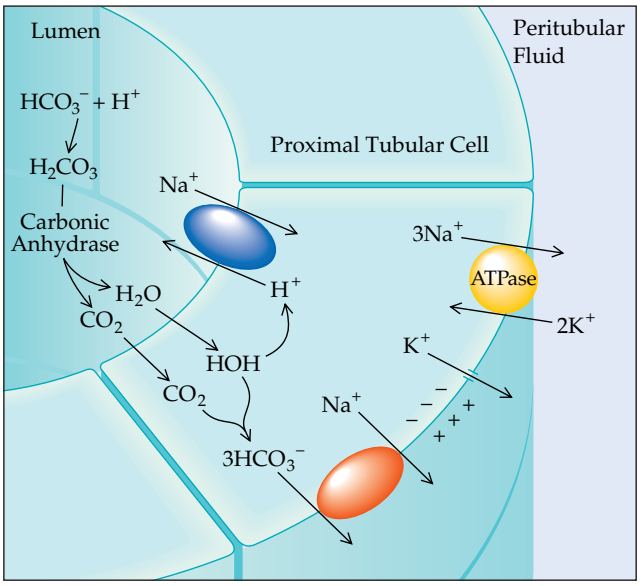


Figure 2 **Proximal tubular bicarbonate reabsorption is activated by the Na^+,K^+-ATPase pump in the peritubular cell membrane. Exchanging peritubular K^+ for intracellular Na^+ keeps the intracellular [Na^+] low, allowing Na^+ to move down its concentration gradient from the tubular lumen through the Na^+-H^+ antiporter to the cell interior. HCO_3^- filtered across the glomerular capillaries combines with secreted H^+ to form H_2CO_3. Rapid dissociation of H_2CO_3 to CO_2 and H_2O in the presence of luminal carbonic anhydrase permits movement into the cell, where redissociation occurs. Ultimately, the reabsorbed H^+ is resecreted in exchange for Na^+, and HCO_3^- moves down an electrical gradient from the cell interior to the peritubular space, where it is reabsorbed into the systemic circulation.**

Anion Gap in Metabolic Acidosis

The anion gap (measured in mEq/L) refers to the difference between the plasma concentrations of the major measured cation (sodium) and the major measured anions (chloride and bicarbonate), as described by the following equation:

$$\text{Anion gap} = [Na^+] - ([HCO_3^-] + [Cl^-])$$

The normal anion gap ranges from 3 to 13 mEq/L, with a mean of about 10 mEq/L. It is composed mainly of plasma proteins (primarily albumin) that carry a negative charge. In the setting of severe hypoalbuminemia, the baseline anion gap may be less than 3 mEq/L, because the anion gap falls by approximately 2.5 mEq/L for every 1 g/dl reduction in the serum albumin concentration.

The most important clinical use of the anion gap is to identify the etiology of metabolic acidosis.[3] The disorders that cause metabolic acidosis fall into two categories: (1) those that cause a fall in the plasma [HCO_3^-] while concomitantly raising the anion gap and (2) those that cause a fall in the plasma [HCO_3^-] without affecting the anion gap. In the latter setting, the plasma [Cl^-] increases, giving rise to a hyperchloremic metabolic acidosis.

Causes of Metabolic Acidosis with a High Anion Gap

Several disorders, as well as the ingestion of toxins, can cause metabolic acidosis with an increased anion gap [*see Table 1*].

Renal failure Advanced renal failure is the most common cause of metabolic acidosis with an increased anion gap in the outpatient setting. The retention of hydrogen ions leads to a fall in the plasma [HCO_3^-]. Because sulfate and phosphate (which are the accompanying anions) are excreted in the urine while chloride is retained, the anion gap remains normal during the early course of acidosis in renal failure. As renal failure progresses (creatinine > 3.0 mg/dl), these ingested anions and metabolic waste products can no longer be excreted normally. At this point, the anion gap increases. There is, however, no linear correlation between the degree of acidemia (or hypobicarbonatemia) and the level of the anion gap. In uncomplicated renal failure, the plasma [HCO_3^-] rarely decreases to less than 12 mEq/L, and the anion gap characteristically remains less than 20 mEq/L.

Lactic acidosis Lactic acidosis is the most common cause of high-anion-gap metabolic acidosis observed in hospitalized patients. Lactic acid production usually increases as a result of hypotension or sepsis, both of which cause true or relative tissue is-

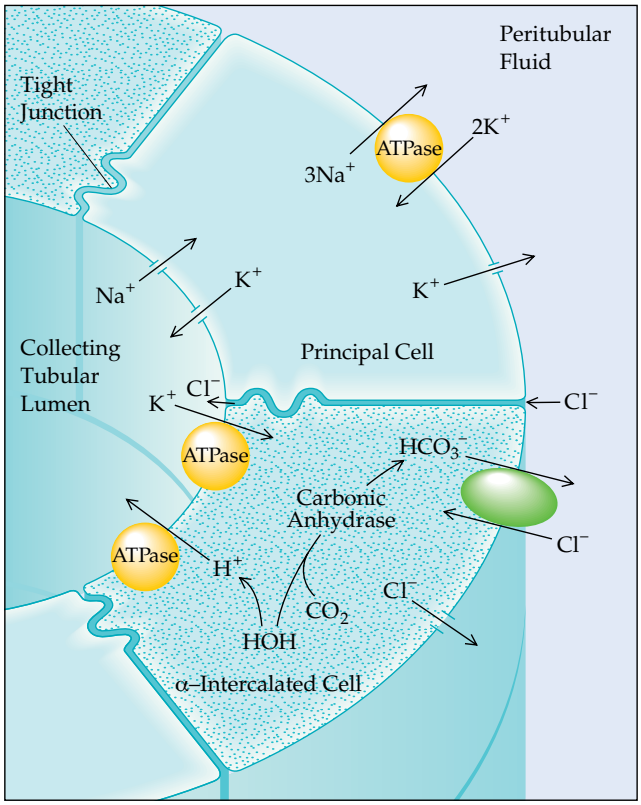


Figure 3 **Secretion of H^+ from the cortical collecting tubule is indirectly linked to Na^+ reabsorption. Intracellular potassium is exchanged for sodium in the principal cells, whereas H^+ is actively transported by an ATPase pump from the α-intercalated cells. Aldosterone stimulates H^+ secretion by entering the principal cell, where it opens Na^+ channels in the luminal membrane and increases Na^+,K^+-ATPase activity. The movement of cationic Na^+ into the principal cells then creates a negative charge within the tubular lumen. K^+ moves from the principal cells and H^+ from the α-intercalated cells down this electrochemical gradient and into the lumen. (When K^+ is depleted, principal cell K^+ secretion is reduced, and K^+ reabsorption via an ATPase pump in the α-intercalated cell is stimulated.) Aldosterone apparently also stimulates the H^+-ATPase directly in the intercalated cell, further enhancing H^+ secretion. HCO_3 is returned to the blood across the peritubular membrane in exchange for Cl^-, thus maintaining electroneutrality.**

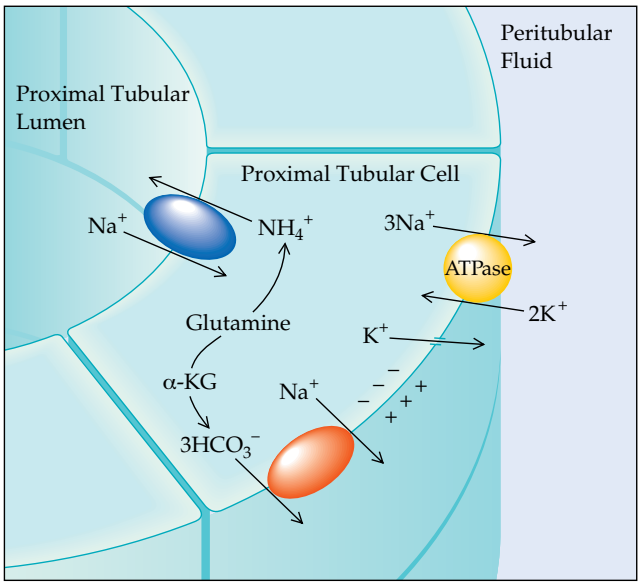


Figure 4 **All of the ammonia used to buffer urinary H^+ in the collecting tubule is synthesized in the proximal convoluted tubule, and glutamine is assumed to be the main source of this ammonia. As glutamine is metabolized, α-ketoglutarate (α-KG) is formed, which ultimately breaks down to bicarbonate that is then secreted into the peritubular fluid by an Na^+-HCO_3^- cotransporter.**

chemia. The oxidative pathways of pyruvate metabolism become markedly impaired in states of mitochondrial dysfunction, such as those induced by tissue hypoxemia. This setting enhances the conversion of pyruvate to lactate. The liver and, to a lesser degree, the kidneys are the major organs that remove lactate from the circulation. In both organs, lactate is converted back to pyruvate and then to carbon dioxide and water via the tricarboxylic acid cycle. When normal tissue perfusion is restored, metabolism of lactate ($CH_3CHOHCOO^-$) in these organs rapidly regenerates the bicarbonate used initially to buffer the acid load, a process that is largely independent of renal acid excretion and that is summarized by the following equation:

$$CH_3CHOHCOO^- + 3O_2 \rightarrow 2CO_2 + 2H_2O + HCO_3^-$$

The normal rate of lactate production can reach 320 mEq/hr (e.g., during exercise), a rate that is usually greater than the rate of lactate production in lactic acidosis. This finding indicates that for lactic acid to accumulate, lactate metabolism must also be impaired. In shock, for example, the marked reduction in hepatic perfusion slows lactate clearance.

Certain medications can also cause lactic acidosis—for example, metformin therapy in diabetics with renal failure and antiretroviral therapy in people with AIDS.[4]

The anion gap is almost always elevated above baseline in lactic acidosis. Because the renal excretory threshold for lactate is 6 to 8 mEq/L and the normal lactate level is less than 1 mEq/L, excess lactate accumulates in the blood, rather than being eliminated in the urine, thus contributing to the increased anion gap.

D-Lactic acidosis is an unusual form of lactic acidosis that is most often seen in patients who have undergone ileal bypass or small bowel resection.[5] In each of these cases, short bowel syndrome can occur, resulting in increased bacterial metabolism of carbohydrate to D-lactic acid because of local overgrowth. In D-lactic acidosis, the anion gap rises initially but may fall over time because renal tubular reabsorption of D-lactate is inefficient in comparison with the reabsorption of L-lactate.

Ketoacidosis Acidosis caused by overproduction of the keto acids acetoacetic acid and β-hydroxybutyric acid occurs when insulin deficiency, fasting, or insulin resistance impairs glucose use. In these settings, ketone bodies are overproduced (a condition termed ketosis) and serve as an alternative source of energy for many cells. The initial ketone formed is acetoacetic acid, which may then be reduced to β-hydroxybutyric acid or nonenzymatically decarboxylated to acetone. Although acetone is chemically neutral, the other ketones are organic acids, and their accumulation leads to metabolic acidosis. The following equation summarizes the reactions:

$$\text{Acetoacetic acid} \longleftrightarrow \beta\text{-hydroxybutyric acid} \longrightarrow \text{acetone}$$

The anion gap characteristically increases in ketoacidosis. Acetoacetate and β-hydroxybutyrate are the major unmeasured anions that accumulate, although a concomitant lactic acidosis may be observed in some patients. The plasma [HCO_3^-] may be markedly depressed in diabetic ketoacidosis; in contrast, the acidemia is generally mild and the plasma bicarbonate rarely less than 18 mEq/L in starvation, or fasting, ketosis.

In patients with ketoacidosis, isotonic fluid replacement leads to ketonuria. At the initiation of fluid replacement, the anion gap is elevated, but it begins to fall as the rate of urinary excretion of acetoacetate and β-hydroxybutyrate exceeds the rate of production. Ultimately, normalization of the anion gap results at a time when the plasma [HCO_3^-] is still low.

Rhabdomyolysis Massive muscle breakdown is an important cause of metabolic acidosis.[6] The retention of metabolic acids and such inorganic anions as phosphate appear to contribute to the increased anion gap. Metabolic acidosis is more likely to develop when concomitant renal failure is present [*see 10:VI Acute Renal Failure*].

Ingested agents and toxins Salicylates and the alcohols ethylene glycol (a component of antifreeze and solvents) and methanol, or wood alcohol (a component of shellac, varnish, deicing solutions, and other commercial preparations), are the most frequent causes of metabolic acidosis among ingested agents and toxins. Ethyl alcohol ingestion is not associated with a high-anion-gap acidosis unless accompanied by either lactic acidosis or ketoacidosis.

The most common acid-base abnormality observed with salicylate intoxication in adults is respiratory alkalosis caused by direct stimulation of the medullary respiratory center [*see* Respiratory Acidosis and Alkalosis, *below*]. In adults, the presence of respiratory alkalosis appears to be a prerequisite for the development of metabolic acidosis, because an isolated metabolic acidosis is rare. Moderate to severe salicylate intoxication, on the other hand, leads to a mixed respiratory alkalosis and a high-anion-gap metabolic acidosis. Methanol and ethylene glycol can produce fatal intoxication, and both can cause a plasma osmolal gap, which refers to the difference between the plasma osmolality (P_{osm}) measured by the laboratory and that calculated by the following formula:

$$\text{Calculated } P_{osm} = (2 \times \text{plasma } [Na^+]) + (\text{glucose concentration}/18) + (\text{blood urea nitrogen}/2.8)$$

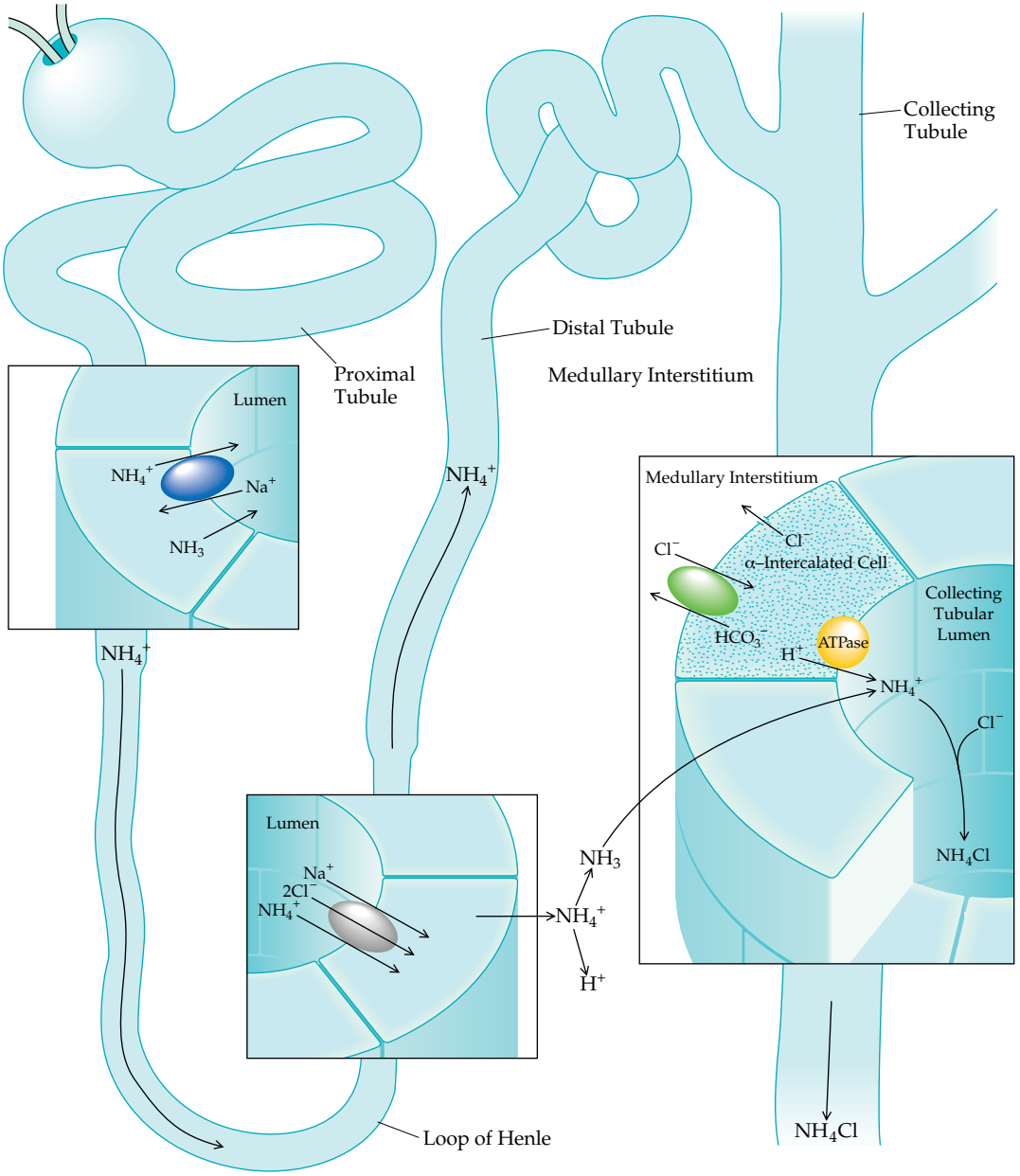


Figure 5 **The ammonia used to buffer urinary hydrogen ions is synthesized in the proximal convoluted tubule. It then diffuses into the proximal tubular lumen or can become acidified within the cell, forming ammonium, which can enter the tubular lumen by substituting for hydrogen ions on the Na^+-H^+ antiporter. Ammonium flows through the thick ascending limb of the loop of Henle, where it is transported from the tubule into the medullary interstitium by replacing potassium on an Na^+-K^+-$2Cl^-$ transporter. In the interstitium, ammonium dissociates to ammonia, which diffuses down its concentration gradient into the lumen of the collecting tubule. Here, ammonia combines with secreted H^+ to form ammonium; NH_4^+ is then excreted as NH_4Cl to maintain electroneutrality. A bicarbonate molecule is regenerated for each H^+ eliminated in the urine.**

A high plasma osmolal gap (with the measured level at least 10 mOsm/kg higher than the calculated value) can be detected only when the P_{osm} is measured by freezing-point depression. In contrast, the osmotic contribution of volatile alcohols is not included when a vapor-pressure osmometer, which assumes that only water is in the vapor phase, is used.[7] Furthermore, the presence of an osmolal gap is not specific for these ingestions, and confirmation of the diagnosis requires plasma assays for the individual drugs [*see* Diagnosis, *below*].

Causes of Metabolic Acidosis with a Normal Anion Gap

A normal anion gap in metabolic acidosis can arise from several causes, including administration of inorganic acids such as hydrochloric acid, loss of bicarbonate from the gastrointestinal tract (e.g., in diarrhea) or the kidneys (e.g., type 2 renal tubular acidosis [RTA]), and impaired renal excretion of hydrogen ions (e.g., renal insufficiency and type 1 RTA) [*see Table 2*]. In the resultant conditions, electroneutrality is maintained by a decrease in plasma bicarbonate and replacement by chloride. Conse-

Table 1 Causes of High-Anion-Gap Metabolic Acidosis

Renal failure*	Ingested agents and toxins
Lactic acidosis	Salicylates
Ketoacidosis*	Ethylene glycol
Rhabdomyolysis	Methanol
	Toluene*

*May also be associated with a normal-anion-gap acidosis.

quently, these disorders are sometimes collectively referred to as the hyperchloremic acidoses.

Acid and chloride administration The infusion of amino acid solutions during parenteral nutrition supplies abundant hydrochloric acid. In certain settings, administration of a sodium chloride solution also may lower the plasma [HCO_3^-] by dilution (a condition termed expansion acidosis), leading to a fall in blood pH.

Bicarbonate or other alkali losses Bicarbonate can be lost from the body via the gastrointestinal tract or the kidneys. Compared with blood, bowel contents are alkaline because pancreatic and biliary secretions add bicarbonate, which is later exchanged for chloride in the ileum and colon to maintain a normal acid-base status. Gastrointestinal losses of bicarbonate (or bicarbonate precursors such as lactate and acetate) are most commonly observed in patients with severe diarrhea. The diagnosis of diarrhea as the cause of a normal-anion-gap acidosis is usually apparent from the patient's history [*see* Diagnosis, *below*]. Hypokalemia resulting from stool potassium losses also supports the diagnosis.

Less commonly, metabolic acidosis resulting from gastrointestinal alkali losses is caused by pancreatic fistulas, biliary drainage, or urinary diversions to the colon or small bowel.[8] It also occurs after pancreatic transplantation in patients who lose bicarbonate via a pancreas-bladder anastomosis.

Renal bicarbonate losses cause the acidemia observed in type 2, or proximal, RTA and in patients who are posthypocapnic. In type 2 RTA, the normal threshold for bicarbonate reabsorption is reduced. Therefore, bicarbonate can no longer be reabsorbed at a rate adequate to maintain the normal plasma level of about 25 mEq/L. During this phase, bicarbonaturia occurs, leading to a urine pH of greater than 5.3. Bicarbonate wasting ceases, however, once the fall in plasma [HCO_3^-] has stabilized; at this time, the urine pH may be less than 5.

Proximal RTA can be observed in a number of circumstances, including the administration of carbonic anhydrase inhibitors and in some patients with multiple myeloma. In multiple myeloma, toxic damage to the proximal tubular cells by the filtered myeloma light chains often leads to a generalized reduction in proximal reabsorption (including glucose and phosphate) in addition to bicarbonate. This group of maladies, caused by proximal tubular dysfunction, is called the Fanconi syndrome.[9,10]

Hypocapnia leads to a fall in the proximal tubular reabsorption of bicarbonate. After 1 to 3 days, the plasma level of bicarbonate decreases. Because this renal adaptive process requires the same time to cease, a sudden increase in the P_aCO_2 does not immediately alter proximal bicarbonate reabsorption. This posthypocapnic metabolic acidosis resolves spontaneously within 24 to 72 hours.

Table 2 Causes of Metabolic Acidosis with a Normal Anion Gap

- Acid and chloride administration
  - Hyperalimentation*
  - NH_4Cl, HCl for treatment of severe metabolic acidosis*
- Bicarbonate (or other alkali) losses
  - Gastrointestinal alkali loss†
    - Diarrhea, pancreatic-intestinal-biliary fistulas, pancreatic transplantation with drainage into the urinary bladder
  - Ureteral diversions†
    - Ureterosigmoidostomy, ileal bladder (if obstructed)
  - Type 2 RTA†
  - Recovery from ketoacidosis†
  - Posthypocapnia†
- Reduced renal and hydrogen excretion
  - Renal insufficiency*
  - Type 1 RTA [*see Table 3*]
    - Hypokalemic forms*
    - Hyperkalemic forms†
  - Type 4 RTA* [*see Table 4*]

* Normal or increased plasma K^+.
† Normal or decreased plasma K^+.
RTA—renal tubular acidosis

Reduced renal hydrogen ion excretion Reduced renal acid excretion is observed in three conditions: renal insufficiency, type 1 (distal) RTA, and type 4 RTA (hypoaldosteronism) [*see Tables 3 and 4*]. The acidosis of renal failure is primarily caused by a reduction in the number of nephrons. In contrast, type 1 RTA is characterized by a reduction in renal acid excretion by each nephron. Because the total quantity of ammonia that can be synthesized is also reduced in renal failure, the urine pH is lower than 5.3 in most patients.

Type 1 RTA, which may be acquired in association with a number of disorders, such as Sjögren syndrome, occurs most commonly when hydrogen ions cannot be pumped out of the α-intercalated cells into the collecting tubular lumen or when they leak back into the cell from the tubule lumen [*see Figure 3*]. The mutations underlying congenital type 1 RTA may be caused by mechanisms similar to those found in the acquired form or by translocation of the Cl^-/HCO_3^- transporter, from the basolateral to the apical (luminal) membrane of the α-intercalated cells.[11] As a result, the urine cannot be maximally acidified, and the urine pH is always higher than 5.3. Furthermore, hypokalemia is characteristically present, in part because of enhanced distal nephron Na^+-K^+ exchange, a process that is necessary to maintain sodium balance because hydrogen ions cannot be secreted in response to sodium reabsorption.

A hyperkalemic form of distal RTA also has been described; it occurs most often in patients with urinary tract obstruction. This disorder is characterized by extensive tubular injury involving aldosterone-sensitive hydrogen secretion as well as aldosterone-independent hydrogen secretion. The urine pH in patients with this disorder is above 5.3.

The most important clinical complication of hypokalemic type 1 RTA is the formation and deposition of calcium phosphate salts that can cause calculi throughout the kidney (nephrocalcinosis). A major factor contributing to crystal formation is hypocitraturia

Table 3 Causes of Type 1 RTA

Hypokalemic Forms	*Hyperkalemic Forms*
Primary	Urinary tract obstruction
Idiopathic	Sickle cell anemia
Genetic	Systemic lupus erythematosus
Familial	Renal transplant rejection
Marfan syndrome	
Ehlers-Danlos syndrome	
Nephrocalcinosis	
Chronic hypercalcemia	
Medullary sponge kidney	
Hypergammaglobulinemic states	
Amyloidosis*	
Cryoglobulinemia	
Cirrhosis	
Drugs and toxins	
Amphotericin B	
Lithium carbonate	
Toluene†	
Autoimmune diseases	
Sjögren syndrome*	
Thyroiditis	
Chronic active hepatitis	
Primary biliary cirrhosis	

*May also cause type 2 RTA [*see Table 5*].
†May also cause metabolic acidosis with an elevated anion gap [*see Table 1*].

caused by an increase in proximal tubular citrate reabsorption. Because calcium citrate is significantly more soluble than calcium phosphate, a decrease in urinary citrate facilitates the precipitation of calcium phosphate crystals in the collecting tubular lumen.

Type 4 RTA may be caused by a number of medications, including nonsteroidal anti-inflammatory drugs (NSAIDs), angiotensin-converting enzyme (ACE) inhibitors, angiotensin II blockers, cyclosporine, and heparin. However, it is most often observed in patients with diabetes mellitus [*see* Plasma Potassium Disorders, Hyperkalemia, *below*]. The most common electrolyte disturbance in type 4 RTA is hyperkalemia (plasma $[K^+] > 5.0$ mEq/L), which is caused by impaired luminal reabsorption of sodium ions and, thus, reduced potassium ion secretion by the principal cells of the collecting tubule [*see Figure 3*]. This defect is caused by a reduction in aldosterone production or action. In contrast to hypokalemia, high plasma potassium levels impair renal ammonia production. Insufficient urinary buffer in the form of ammonia limits hydrogen ion secretion in the collecting tubule. Despite this reduced ability to excrete hydrogen ions, the urine is usually acidified (pH < 5.3). This apparently paradoxical finding occurs because the limited urinary buffer permits the free urine $[H^+]$ to exceed the secretory capacity of the α-intercalated cells, thus limiting further hydrogen ion transport from the cells and into the collecting tubular lumen. In contrast to patients with type 1 RTA, patients with type 4 RTA can excrete some acid, albeit in inadequate quantities; consequently, the urine pH is typically lower than 5.3 in patients with this disorder.

Potassium Imbalance in Metabolic Acidosis

Hyperkalemia may be observed in patients with normal anion gap metabolic acidosis; in respiratory acidosis, the shift of potassium from cells to plasma is smaller. A fall in the plasma $[HCO_3^-]$ leads to an exchange of extracellular hydrogen ions for intracellular potassium ions. This defense mechanism permits intracellular buffering of hydrogen ions; the H^+-K^+ exchange maintains electroneutrality. Therefore, the patient who has diarrhea with a normal-anion-gap metabolic acidosis and a low plasma $[K^+]$ has significantly greater potassium depletion than the patient with the same plasma $[K^+]$ and a normal blood pH.

In contrast to what occurs in patients with a normal-anion-gap metabolic acidosis, potassium shifts caused by acidemia are less pronounced in patients with endogenous organic acidoses (e.g., lactic acidosis and ketoacidosis).[12] When hyperkalemia develops in this setting, it is typically caused by renal failure, cellular catabolism that permits leakage of potassium from cells (lactic acidosis), or insulin deficiency and hyperglycemia, both of which promote potassium exit from cells.

Diagnosis

Clinical manifestations Kussmaul respirations suggest the presence of metabolic acidosis. The increase in tidal volume, rather than respiratory rate, that characterizes these ventilatory changes results from stimulation of the brain stem respiratory system by the low pH.

Secondary hypotension also may be observed in severely acidemic persons. In this setting, the reduced blood pressure results from depressed myocardial contractility and arterial vasodilatation, which are induced by the decreased blood pH. Initially, elevated levels of circulating catecholamines counter the cardiovascular effects of acidemia; however, at a blood pH below 7.15 to 7.20, the effects of acidemia may predominate.[13] Reentrant arrhythmias and a reduction in the threshold for ventricular fibrillation can occur, whereas the defibrillation threshold remains unaltered.

The symptoms of metabolic acidosis caused by renal failure depend on the cause, the rapidity with which the renal failure develops, and concomitant conditions that may be present (e.g.,

Table 4 Causes of Type 4 RTA and Aldosterone Resistance

Disorder	*Cause*
Type 4 RTA	Reduced activity of the renin-angiotensin system
	Hyporeninemic type 4 RTA (diabetes most common)
	Nonsteroidal anti-inflammatory drugs (with the possible exception of sulindac)
	Angiotensin-converting enzyme inhibitors
	Angiotensin II receptor blockers
	Cyclosporine
	AIDS*
	Reduced aldosterone synthesis
	Low cortisol levels
	Primary adrenal insufficiency
	Enzymatic deficiencies (primarily adrenal hyperplasia)
	Normal cortisol levels
	Heparin
	Immediately after removal of adrenal adenoma in primary aldosteronism
	Enzymatic deficiencies
Aldosterone resistance (normal or increased aldosterone levels)	Potassium-sparing diuretics, trimethoprim
	Pseudohypoaldosteronism (hereditary or acquired)
	Hyperkalemic type 1 RTA†

*Adrenalitis causing type 4 RTA may also occur in persons who are seropositive for HIV.
†In this setting, the urine is alkaline (pH > 5.3).

congestive heart failure). There may be no abnormal clinical symptoms, despite the presence of azotemia and acidemia.

Ketoacidosis is often associated with increased thirst and polyuria. It is frequently precipitated by an unrelated insult, such as an infection, that may dominate the clinical presentation.

The distinctive characteristics of rhabdomyolysis are myalgia and myoglobinuria, the latter of which leads to reddish urine that has no significant red cells. In many patients, however, neither finding is present. Despite the lack of findings, serum creatine kinase levels are uniformly elevated.

The symptoms and signs of lactic acidosis are characteristically those of the underlying disturbance causing the disorder—namely, hypotension or sepsis. Most commonly, patients exhibit evidence of hypoperfusion, such as low blood pressure and cool or mottled extremities. Less commonly, a medication (e.g., metformin) may be the cause. Patients with D-lactic acidosis may present with encephalopathy. L-Lactate dehydrogenase is the standard assay for the diagnosis of lactic acidosis; because this test does not measure D-lactic acid, a specific enzymatic assay for D-lactate (i.e., one that uses D-lactic dehydrogenase) is necessary to confirm the diagnosis of D-lactic acidosis.

In addition to acid-base changes, clinical findings that may accompany severe salicylate intoxication include tinnitus, hyperpyrexia, vasodilatation leading to shock, and peripheral or pulmonary edema. Symptoms of acidosis caused by ingestion of methanol or ethylene glycol [*see* Ingested Agents and Toxins, *above*] may develop 12 to 36 hours after ingestion. In addition to acid-base changes, the initial symptoms that occur after ingestion of methanol include weakness, nausea, headache, and decreased vision, which can progress to blindness, coma, and death. Fundoscopic examination may reveal a retinal sheen caused by retinal edema. After ethylene glycol is ingested, the earliest findings are neurologic abnormalities that range from drunkenness to coma. If the patient is not treated, these changes may be followed first by cardiopulmonary symptoms (tachypnea and pulmonary edema) and then by flank pain and renal failure caused by calcium oxalate crystal deposition, which may be seen in the urine sediment [*see Figure 2 in 10:XII Nephrolithiasis*].

Laboratory tests The diagnosis of metabolic acidosis is made relatively easily in the presence of a low blood pH and a low plasma [HCO_3^-]. The anion gap can then be used to identify a specific cause. The finding of concomitant hypokalemia or hyperkalemia may also be useful.

Once the presence or absence of a high serum anion gap is determined, the respiratory defense against acidemia can be evaluated. The respiratory response begins immediately, although it may not be maximal for 12 to 24 hours. The appropriate respiratory compensation can be calculated by using the Winter[14] formula:

$$\text{Expected } P_{CO_2} = 1.5\,[HCO_3^-] + 8 \pm 2$$

The respiratory compensation is inadequate when the measured P_aCO_2 is higher than the expected value and is excessive when the P_aCO_2 is lower than the expected value. When these inadequate or excessive responses occur, a superimposed respiratory acidosis or respiratory alkalosis, respectively, is present [*see* Respiratory Acidosis and Alkalosis, *below*]. For example, if the serum [HCO_3^-] were 16 mEq/L in a patient with a low blood pH (metabolic acidosis), the expected P_{CO_2} would be approximately 32[(1.5 × 16) + 8] mm Hg. A P_{CO_2} below this value indicates a superimposed respiratory alkalosis.

Another laboratory calculation, the urinary anion gap, may be useful in defining the cause of metabolic acidosis when the serum anion gap is normal. For example, although the diagnosis of diarrhea is usually apparent from the patient's history and the presence of hypokalemia, a profile of plasma electrolytes similar to that in persons with diarrhea may be observed in patients with type 1 RTA. These disorders can usually be distinguished by the urine pH; the urine tends to be acidic (pH < 5.3) in patients with diarrhea and tends to be alkaline (pH > 5.3) in patients with type 1 RTA. In some patients with diarrhea, however, the urine may be alkaline, presumably because ammonia production (induced by hypokalemia) increases to such an extent that urinary buffer is produced in excess of hydrogen ion secretion.

Calculation of the urinary anion gap, as shown in the following equation, may be quite useful in providing an estimate of urinary ammonia secretion:

$$\text{Urinary anion gap} = (\text{urinary } [Na^+] + \text{urinary } [K^+]) - \text{urinary } [Cl^-]$$

Whenever secreted hydrogen ions are excreted as ammonium chloride ($NH_4^+Cl^-$), an increase in urinary chloride excretion results. The increase in urinary chloride excretion decreases the urinary anion gap, leading to a negative value in most patients with diarrhea. By comparison, in type 1 RTA, the urinary anion gap is positive. As an example, a patient with a normal anion gap metabolic acidosis (e.g., [HCO_3^-] = 10 mEq/L), a low potassium level (hypokalemia), and an alkaline urine pH (6.0) could have diarrhea or type 1 RTA. If the urine [Na^+] were 50 mEq/L, the urine [K^+] were 28 mEq/L, and the urine [Cl^-] were 55 mEq/L, the urinary anion gap would be +23, supporting a diagnosis of RTA.

Despite the potential value of the urinary anion gap in estimating urinary $NH_4^+Cl^-$ excretion, the presence of certain anions (e.g., β-hydroxybutyrate and acetoacetic acid) can contribute to a positive urinary gap. This may erroneously suggest that metabolic acidosis is caused by RTA rather than diarrhea. The urinary osmolal gap, which largely represents ammonium salts, can be used to confirm that the positive urinary anion gap is the result of type 1 RTA**.** The urinary osmolal gap is the difference between the urinary osmolality (U_{osm}) measured by the laboratory and that calculated by the following formula:

$$\text{Calculated } P_{osm} = (2 \times \text{plasma } [Na^+]) + (\text{glucose concentration}/18) + (\text{blood urea nitrogen}/2.8)$$

The multiple of 2 accounts for the anions accompanying sodium and potassium, whereas the divisors 2.8 and 18 reflect adjustments required to convert the blood urea nitrogen and glucose findings from the routinely used units of mg/dl to mmol/L or mOsm/kg. The normal urinary osmolol gap is 80 to 100 mmol/L, reflecting an NH_4^+ excretion rate of about half this value—namely, 40 to 50 mEq/L—as a result of the accompanying anion chloride. In the presence of metabolic acidosis, the value should be significantly increased; a normal or lower value in an individual with metabolic acidemia strongly supports impaired $NH_4^+Cl^-$ excretion—and, ultimately, type 1 RTA—as the underlying cause for the positive anion gap.

Treatment

Treatment of metabolic acidosis is aimed at correcting both acidemia and the underlying disorder. The likelihood that alkali administration is needed and that it will be effective depends on

the blood pH, the compensatory mechanisms, and the underlying condition.

Until the arterial blood pH falls below 7.15 to 7.20, the adverse effects of acidemia are usually compensated for by elevated plasma catecholamines [*see* Diagnosis, *above*]. To maintain adequate buffer reserves, alkali therapy should be considered to keep [HCO_3^-] higher than 10 to 12 mEq/L. Alkali administration is usually unnecessary, however, if the acidosis is likely to resolve spontaneously (as in lactic acidosis after a grand mal seizure). Therapy must take into consideration the underlying cause of metabolic acidosis (see below).

Chronic renal failure Of interest is the finding that the acidemia tends to be more severe in nondiabetic patients with chronic renal failure than in diabetic patients with similar degrees of renal insufficiency.[15] This difference may be the result of more efficient extrarenal bicarbonate generation in diabetic patients.

As long as the metabolic acidosis is mild, many adults with renal failure are not treated with alkali replacement, partly because of a concern that sodium bicarbonate will exacerbate the volume expansion and hypertension that are commonly present.

Some studies, however, have suggested several reasons for the use of alkali replacement therapy in patients with renal failure.[16-18] Among the reasons are the likelihood that acidemia can enhance the breakdown of skeletal muscle, reduce albumin synthesis, and (by activating the complement system) contribute to tubulointerstitial injury and that bone buffering of hydrogen ions can lead to bone resorption.[19] These findings have led some physicians to advocate the early use of alkali therapy to maintain plasma [HCO_3^-] above 22 mEq/L. Definitive studies are still needed, however, to ascertain the benefit of such treatment.

Physicians should be aware that electrolyte studies in hemodialysis patients who use non–hospital-based dialysis units may be performed in central laboratories several hundred miles away. Compared with results of samples obtained locally, samples tested at the central laboratories appear to show consistent and clinically important decreases in serum [HCO_3^-].[20] This in vitro enhancement of metabolic acidosis should be confirmed before treatment is initiated.

Lactic acidosis Correcting the underlying disorder is the primary therapy for lactic acidosis. Reversal of circulatory failure, hypoxemia, or sepsis reduces the rate of lactate production and enhances its removal.

As in chronic renal failure, the use of alkali therapy in lactic acidosis is controversial.[21,22] The principal rationale for bicarbonate administration is the potential maintenance of normal cardiovascular homeostasis. This possible advantage must be weighed against deleterious side effects, such as volume overload, hypernatremia, and alkalosis (when excessive bicarbonate is administered).

Clinical studies also suggest that sodium bicarbonate therapy may not improve either the blood pH or the survival of patients with lactic acidosis. The administration of sodium bicarbonate has been observed to reduce cardiac performance in patients with cardiac disease, congestive heart failure, or acute myocardial infarction.[23] This lack of efficacy possibly results from an associated increase in net lactic acid production, although hyperosmolality of the administered alkali solution also may be important. The finding that dichloroacetate (which promotes the conversion of pyruvate into the Krebs cycle instead of permitting its conversion to lactate) can lower lactate levels and raise the blood pH in patients with lactic acidosis without seeming to improve survival lends further support to the concept that treating the underlying cause of lactic acidosis is more important than treating the acidemia. In view of these observations, a reasonable approach may be to administer bicarbonate to maintain the arterial blood pH at higher than 7.15 and the plasma bicarbonate at higher than 10 mEq/L. If one of the complications of bicarbonate therapy ensues, the benefit of continuing alkali therapy should be reevaluated.

Therapy for D-lactic acidosis must address the two underlying factors that contribute to the overproduction of D-lactic acid: the presence of intestinal bacteria that produce D-lactic acid and the enhanced delivery of carbohydrates to the colon secondary to short bowel syndrome. Treatment usually consists of antibiotic therapy to decrease the number of D-lactate–producing organisms. Alkali therapy has been used but has not been evaluated in a controlled fashion.

Ketoacidosis The trend toward normal-anion-gap metabolic acidosis during recovery from ketoacidosis has an important effect on the rate at which acidemia can be corrected. A maximal rate of 500 ml of fluid an hour appears to be effective replacement therapy for patients with ketoacidosis. Significantly higher rates of fluid administration can result in a decreased anion gap without an increase in the plasma [HCO_3^-]. Once this occurs, correction of the acidosis requires regeneration of bicarbonate by the kidney, a process that may take at least 2 to 3 days[24]; consequently, more vigorous fluid administration (above 500 ml/hr) may delay recovery from acidosis. The rate of fluid administration should be slowed after intravascular volume compromise—manifested by reduced blood pressure or increased plasma creatinine and urea nitrogen concentrations—has been corrected.

Although exceptions occur, the administration of sodium bicarbonate is not usually necessary in ketoacidosis, because there appears to be no difference in mortality between patients treated with sodium bicarbonate and control subjects. Bicarbonate therapy may be a risk factor for the subsequent development of cerebral edema in this setting.[25]

Salicylate, ethylene glycol, and methanol intoxication Salicylate removal is enhanced by urinary alkalinization (to maintain the blood pH between 7.45 and 7.5), which increases urinary excretion even in the setting of dialysis.[26] The usual treatment of ethylene glycol or methanol intoxication is to administer an agent that will reduce conversion of the nontoxic alcohol to its toxic metabolic by-products and to use dialysis in the presence of tissue damage or acidemia. Although ethyl alcohol has been administered historically, fomepizole (4-methylpyrazole) is the only potent inhibitor of alcohol dehydrogenase that has been studied prospectively and that has been approved by the Food and Drug Administration for this condition.[27]

Hyperchloremic acidoses The therapeutic approach to the patient with diarrhea and metabolic acidosis depends upon the severity of the two disturbances. Alkali administration is not necessary with mild to moderate reductions in the plasma [HCO_3^-] if the diarrhea is controlled (thereby minimizing further bicarbonate loss) and renal function is normal (thereby allowing acid excretion to increase). However, some patients will require treatment with bicarbonate.

The acidemia in type 1 RTA can be corrected with bicarbonate

or with a bicarbonate precursor, such as citrate. The usual requirement is 1 to 3 mEq/kg/day. Correcting the acidemia reduces tubular citrate reabsorption, which leads to an increase in urinary excretion and a decrease in the tendency toward nephrolithiasis and nephrocalcinosis. Usually, a potassium salt, such as potassium citrate, is administered because it corrects the potassium deficit as well.

In adults, treatment of type 2 RTA is aimed at the underlying cause (e.g., multiple myeloma). Because the acidemia is typically mild, alkali therapy may not be necessary. Correction of acidemia is appropriate in children because in this age group, acidemia is more likely to impair growth and contribute to metabolic bone disease.

Type 4 RTA, if mild, may not require treatment. Aldosterone replacement with fludrocortisone (0.1 to 1 mg/day) may increase acid secretion by lowering the serum [K^+]; however, many patients cannot tolerate the side effects (e.g., edema and hypertension) associated with this therapy. In these individuals, correction of serum [K^+] may be achieved by administration of a loop diuretic, dietary potassium restriction, or the elimination of drugs that promote hyperkalemia. When fludrocortisone therapy is required, concomitant administration of a loop diuretic may limit the development of high blood pressure and fluid retention.

METABOLIC ALKALOSIS

Primary metabolic alkalosis is characterized by an elevated plasma [HCO_3^-] and an arterial pH greater than 7.42. When there is concomitant metabolic acidosis, however, the blood pH may be increased, decreased, or normal. Furthermore, hyperbicarbonatemia alone is not diagnostic of primary metabolic alkalosis, because it may also represent the appropriate physiologic response to chronic respiratory acidosis. These conditions can usually be easily distinguished by measuring the arterial blood pH, which is reduced in respiratory acidosis.

Causes of Metabolic Alkalosis

Metabolic alkalosis is a relatively common clinical problem that is most often induced by diuretic therapy or the loss of gastric secretions as a result of vomiting or nasogastric suction.[28] Two abnormalities must be present for metabolic alkalosis to develop and to be sustained. First, there must be an initial increase in the plasma [HCO_3^-] caused by hydrogen loss in gastrointestinal secretions or in the urine, hydrogen movement into cells, alkali administration, or volume contraction around a relatively constant amount of extracelluar bicarbonate (called a contraction alkalosis) [*see Table 5*]. Second, one of three factors (in the absence of advanced renal failure) must be present to sustain the high plasma [HCO_3^-] once the initiating event has been terminated: effective circulating volume depletion, chloride depletion, and hypochloremia or hypokalemia [*see* Additional Factors Contributing to Metabolic Alkalosis, *below*].

Gastrointestinal hydrogen loss Gastric juice contains a high concentration of hydrochloric acid and a lesser amount of potassium chloride. Each 1 mEq of hydrogen lost generates 1 mEq of bicarbonate. In normal persons, gastric hydrogen secretion does not lead to metabolic alkalosis, because it is matched by pancreatic bicarbonate secretion that is stimulated as the acid enters the duodenum. There is no stimulus to bicarbonate secretion, however, when vomiting or nasogastric tube drainage prevents the hydrogen ions from reaching the duodenum. Vomiting can be surreptitious in some cases, such as in patients with eating disorders. In some patients, laxative abuse can also lead to a metabolic alkalosis, most likely resulting from a hypokalemia-induced increase in urinary ammonia production, which in turn leads to increased ammonium chloride excretion in the urine.

Table 5 Causes of Metabolic Alkalosis

Disorder	*Cause*
Hydrogen loss	Gastrointestinal losses Removal of gastric secretions (vomiting or nasogastric suction)* Chloride-losing diarrheal states Laxative abuse Renal losses Loop or thiazide diuretics* Mineralocorticoid excess Posthypercapnic alkalosis Hypercalcemia (including milk-alkali syndrome) High-dose intravenous penicillin derivatives Bartter syndrome
Intracellular hydrogen shift	Hypokalemia
Bicarbonate retention	Administration of alkali (as either a bicarbonate or a bicarbonate precursor with massive blood transfusions or with absorbable antacids)†
Contraction alkalosis	Diuretics* Loss of gastrointestinal secretions having high [Cl^-] and low [HCO_3^-] when compared with plasma (usually with vomiting)*

*These are the most common causes.
†For alkalosis to be maintained, renal bicarbonate excretion also must be impaired by either reduced filtration or enhanced reabsorption by the proximal tubule.

Renal hydrogen loss An inappropriate renal acid loss may occur when there is an increase in hydrogen ion secretion by the distal nephrons. Mineralocorticoids, including aldosterone, act here both by directly stimulating the secretory H^+-ATPase pump and by making the tubular lumen more electronegative through the stimulation of sodium reabsorption [*see Figure 6*]. Distal potassium secretion is also enhanced in this setting and results in concurrent hypokalemia.

Metabolic alkalosis associated with loss of renal hydrogen may be caused by mineralocorticoid excess, administration of loop or thiazide diuretics, posthypercapnic alkalosis, or hypercalcemia (e.g., milk-alkali syndrome) [*see Table 5*]. Metabolic alkalosis also occurs in Bartter and Gitelman syndromes. These two disorders produce electrolyte abnormalities similar to those caused by diuretic therapy because they are associated with defects in the transporters in the loop of Henle and the distal tubule, respectively, that inhibit mechanisms also inhibited by loop and thiazide diuretics [*see* Bartter Syndrome and Gitelman Syndrome, *below*]. Any cause of mineralocorticoid excess, such as primary aldosteronism, can result in metabolic alkalosis. Primary aldosteronism is generally accompanied by hypertension and hypokalemia. In contrast, untreated patients with secondary aldosteronism caused by congestive heart failure or cirrhosis usually do not present with metabolic alkalosis or hypokalemia. In such cases, the effect of aldosterone is counteracted by decreased distal sodium delivery (unless diuretics are administered) and reduced urine volume; these factors limit the quantity of acid and potassium secreted into, as well as excreted in, the final urine.

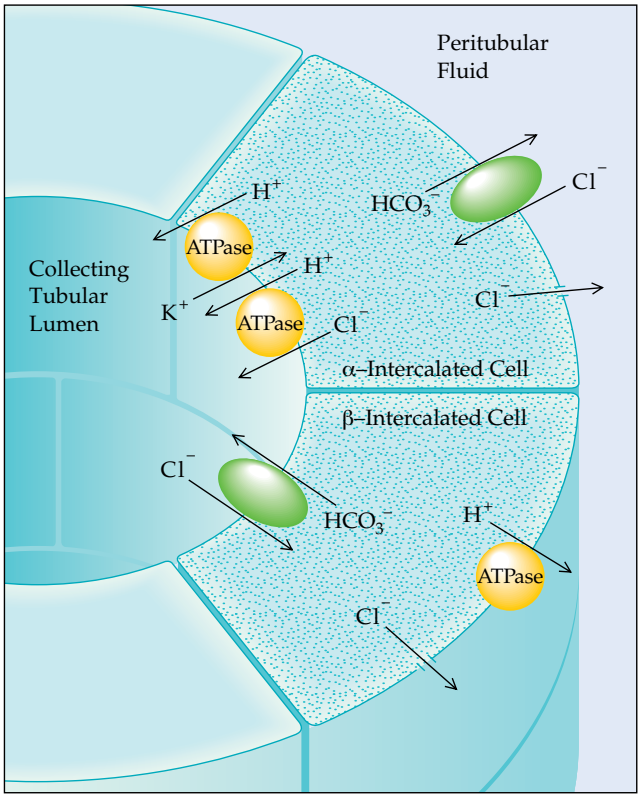


Figure 6 **In metabolic alkalosis, the high ratio of α-intercalated to β-intercalated cells decreases. Unlike α-intercalated cells, which regenerate bicarbonate and add it to the venous blood, β-intercalated cells promote urinary HCO_3^- excretion by exchanging HCO_3^- for Cl^- in the glomerular filtrate. As a result, chloride administration serves an important function in the treatment of most persons with metabolic alkalosis.**

When patients are treated with either thiazide or loop diuretics, adequate distal delivery of sodium chloride and increased secretion of aldosterone occur. The increase in distal hydrogen ion secretion and volume contraction, if the diuresis has been large, contribute to the development of metabolic alkalosis.

Chronic respiratory acidosis brings about an appropriate increase in hydrogen secretion, as the rise in the plasma [HCO_3^-] raises the pH toward normal. Rapid lowering of the P_aCO_2, usually by mechanical ventilation, leads to metabolic alkalosis, as the patient is left with an elevated plasma [HCO_3^-]. This abnormality is called a posthypercapnic metabolic alkalosis.

Hypercalcemia Hypercalcemia increases renal tubular bicarbonate reabsorption. Significant metabolic alkalosis in hypercalcemic patients, however, is more commonly seen in those with milk-alkali syndrome. In milk-alkali syndrome, an increased alkaline load (caused by the ingestion of calcium carbonate) and hypercalcemia-induced renal failure increase bicarbonate production and diminish bicarbonate excretion. Metabolic alkalosis can be observed in any severe hypercalcemic state that is caused by volume contraction and by impaired sodium chloride and potassium chloride absorption resulting from tubular injury.

Intracellular hydrogen shift In addition to being caused by hydrogen loss, metabolic alkalosis can be caused by a shift of hydrogen into the cells. Both vomiting and diuretic therapy directly induce potassium and hydrogen loss. Hypokalemia produces a transcellular shift in which potassium leaves cells to replete extracellular stores; to maintain electroneutrality, hydrogen enters the cells. This shift not only raises the extracellular pH but also decreases the intracellular pH; the latter promotes proximal tubular bicarbonate reabsorption and distal hydrogen ion secretion.

Alkali administration The administration of sodium bicarbonate at a dosage as high as 1,000 mEq/day does not normally induce metabolic alkalosis in normal persons, because the excess bicarbonate is rapidly excreted in the urine. If the ability to excrete bicarbonate is impaired, however, metabolic alkalosis can occur when a very large quantity of bicarbonate or a bicarbonate precursor (e.g., lactate, citrate, or acetate) is administered, as occurs with citrate in large-volume blood transfusions.

Contraction alkalosis Contraction alkalosis develops when loss of relatively large volumes of bicarbonate-free fluid occurs. In this setting, the plasma [HCO_3^-] rises because the extracellular volume contracts around a relatively constant quantity of extracellular bicarbonate.

The most common cause of contraction alkalosis is administration of a loop diuretic to induce rapid fluid removal in a patient with marked edema. Similarly, contraction alkalosis occurs under other conditions in which fluid with a high chloride concentration and a low [HCO_3^-] is lost. Among these causes are receipt of thiazide diuretics; loss of gastric secretions (even in patients with achlorhydria); sweat losses in patients with cystic fibrosis; and diarrhea in some patients with villous adenomas or congenital chloridorrhea.

Additional factors contributing to metabolic alkalosis In the absence of advanced renal failure, one of three factors must be present to sustain high plasma [HCO_3^-]: effective circulating volume depletion, chloride depletion, or hypokalemia.

Both the fall in the glomerular filtration rate (GFR) and the associated sodium avidity seen with hypovolemia limit the excretion of sodium bicarbonate. Most bicarbonate is reabsorbed in the proximal tubule [*see Figure 2*]. An important stimulus for enhanced reabsorption in this nephron segment is increased activity of the Na^+-H^+ antiporter in the tubular cell membranes. Volume contraction promotes Na^+-H^+ exchange in this nephron segment, in part through the release of angiotensin II. The hydrogen ions secreted into the lumen combine with filtered bicarbonate, ultimately leading to a higher rate of transport back into the tubular cells. More bicarbonate is returned to the venous blood at the level of the collecting tubules, partly under the influence of secondary aldosteronism. Volume depletion does not cause metabolic alkalosis without selective chloride losses. For example, blood loss does not increase the [HCO_3^-], because the chloride losses are identical to those in an equal volume of plasma.

Chloride depletion can both promote bicarbonate regeneration and decrease distal bicarbonate secretion. Bicarbonate generation in the α-intercalated cells in the cortical collecting tubule is mediated by hydrogen ion secretion via H^+-ATPase pumps in the luminal membrane [*see Figure 6*]. Passive cosecretion of chloride is required to maintain electroneutrality. Intracellular bicarbonate is returned to the systemic circulation in exchange for chloride.

The β-intercalated cells in the cortical collecting tubule (which increase in number when metabolic alkalosis develops) are able to secrete bicarbonate directly by reversing the location

of the transporters [*see Figure 6*]. Thus, the Cl-HCO_3^- exchangers are located in the luminal membrane, leading to bicarbonate secretion into the tubular lumen; the H^+-ATPase pumps are located in the basolateral membrane. Although the activity of these cells is appropriately enhanced by alkalemia in an attempt to excrete the excess bicarbonate, the associated fall in the tubular fluid [Cl^-] diminishes the favorable inward gradient for chloride, thereby reducing bicarbonate secretion. However, chloride depletion in humans is almost always associated with effective volume depletion.

Hypokalemia directly increases bicarbonate reabsorption by at least two different mechanisms. First, the fall in plasma [K^+] shifts potassium from hydrogen into cells. The ensuing intracellular acidosis stimulates hydrogen secretion and bicarbonate reabsorption in the proximal and collecting tubules. Second, distal hydrogen and potassium secretion is mediated in exchange for luminal sodium [*see Figure 6*]. In states of potassium depletion, the rate of hydrogen secretion in exchange for sodium increases. As a result, hypokalemia and hyperaldosteronism, which stimulate hydrogen ion secretion, can have a potentiating effect on the development and maintenance of metabolic alkalosis.

Diagnosis

Clinical manifestations Some individuals with metabolic alkalosis have severe cramping, paresthesias, or even tetany, but others with similar electrolyte levels do not; the reason for this difference in presentation is unclear. Other clinical findings are the result of the underlying etiology (e.g., hypertension with primary hyperaldosteronism).

The diagnosis of metabolic alkalosis is usually evident from the patient's history of vomiting or receipt of diuretic therapy. In some cases, however, no cause for the metabolic alkalosis is apparent. In such a setting, the most likely diagnosis is surreptitious vomiting caused by an eating disorder, use of diuretics, or one of the causes of mineralocorticoid excess (e.g., primary aldosteronism). The first two factors induce effective volume depletion, whereas primary hyperaldosteronism is usually associated with mild volume expansion as a result of the stimulatory effect of aldosterone on renal sodium reabsorption.

Several findings on physical examination may suggest surreptitious vomiting as the cause, including dental erosion from repeated exposure to acid gastric secretions and the presence of ulcers and calluses on the dorsum of the hand caused by sticking a finger in the back of the throat to induce vomiting.

Laboratory tests Measurement of the [Na^+] in a random urine specimen is used in many conditions to distinguish between volume depletion (urinary [Na^+] usually < 20 mEq/L) and euvolemia (urinary [Na^+] > 40 mEq/L). However, metabolic alkalosis is one of the conditions in which volume depletion may not lead to a low urinary [Na^+]. The capacity to retain sodium in this setting may be antagonized by the need to excrete bicarbonate (as the sodium salt) in an attempt to correct the alkalosis. In such cases, a random urinary [Cl^-] determination is more useful.

Sodium wasting is most likely to occur during the first few days of vomiting, when the plasma [HCO_3^-] and thus the filtered bicarbonate load are increased. Early in the course of vomiting, the ability to enhance bicarbonate reabsorption has not yet occurred. The net effects are a high urinary [Na^+] and urinary [K^+] and a urine pH of greater than 7.0 caused by bicarbonaturia.

As a result, the urinary [Na^+] is not necessarily an accurate reflection of the patient's volume status in metabolic alkalosis. The presence of underlying hypovolemia can be detected more accurately by the finding of a urinary [Cl^-] below 25 mEq/L. The appropriate chloride conservation is caused both by volume depletion and by hypochloremia induced by chloride losses in gastric secretions. The urinary chloride concentration may be inappropriately elevated, however, if a defect in chloride reabsorption is present. Such a defect most commonly occurs in patients who receive diuretic therapy. Thus, in patients with metabolic alkalosis caused by vomiting, the urinary chloride concentration is typically low; it is higher when diuretics are administered. In both circumstances, the urinary [Na^+] may be elevated.

Bartter syndrome and Gitelman syndrome Bartter and Gitelman syndromes are disorders of sodium chloride reabsorption in the loop of Henle and the distal tubule, respectively.[29] Bartter syndrome is a rare disorder that causes hypokalemic metabolic alkalosis. Because the urinary [Na^+] and urinary [Cl^-] are usually higher than 25 mEq/L, surreptitious diuretic use is the major disorder to be considered in the differential diagnosis. Classic Bartter syndrome generally presents in early life and may be associated with delayed growth and mental retardation. The spectrum of findings, including hypercalciuria, is most compatible with a primary defect in sodium chloride reabsorption in the medullary thick ascending limb of the loop of Henle.

Gitelman syndrome is a more benign condition that may be inherited as an autosomal recessive disease; it may not be diagnosed until late childhood or even adulthood. In contrast to patients with Bartter syndrome, who have a defect in urinary concentrating ability (normal function of the loop of Henle is needed to generate a high interstitial osmotic gradient), patients with Gitelman syndrome can demonstrate normal urinary concentrating ability and have hypocalciuria. This finding suggests that the defect resides in the distal tubule, because similar findings are observed in patients treated with thiazide diuretics. Bartter and Gitelman syndromes are diagnosed only after diuretic use—the other, much more common, cause of these findings—has been excluded.

Treatment

In patients with true volume depletion caused by vomiting, nasogastric suction, villous adenomas, or diuretic therapy, metabolic alkalosis can be corrected by the administration of sodium chloride.[30] The administration of potassium chloride to patients with concurrent hypokalemia also contributes to correction of the alkalemia. Measuring the urinary [Cl^-] may be useful in determining when volume depletion has been corrected; once renal perfusion has been restored, this value should be above 40 mEq/L.

Edematous states Therapy is different for edematous patients with metabolic alkalosis usually caused by heart failure, cor pulmonale, or advanced liver disease. In these disorders, sodium chloride is contraindicated because it increases the degree of edema. Administration of the carbonic anhydrase inhibitor acetazolamide (beginning with 250 mg p.o., q.d. or b.i.d., or 125 mg I.V., q.d. or b.i.d.), however, may be particularly effective. This drug preferentially inhibits proximal tubular reabsorption of sodium bicarbonate, thereby correcting both the alkalosis and the fluid overload.

A potential side effect of therapy with a carbonic anhydrase inhibitor is the development or worsening of hypokalemia. Although hypokalemia can be treated with potassium supple-

ments, an alternative approach to therapy is to administer a potassium-sparing diuretic (e.g., spironolactone) instead of a carbonic anhydrase inhibitor. Potassium-sparing diuretics impair sodium reabsorption in the collecting tubules and, as a result, limit further potassium and hydrogen secretion [*see Figure 6*]. In patients with advanced liver disease, the aldosterone antagonist spironolactone or eplerenone may be the most effective diuretic.

On extremely rare occasions, metabolic alkalosis may be so severe that administration of hydrochloric acid is required to correct the problem. The standard 0.1N (decinormal) hydrochloric acid solution contains 100 mEq H^+/L. Because of the corrosive nature of hydrochloric acid solution, administration of hydrochloric acid should be used only when other measures to correct metabolic alkalosis have failed. The hydrochloric acid solution must always be administered into a central vein.

Bartter syndrome and Gitelman syndrome The tubular defect in patients with Bartter syndrome or Gitelman syndrome cannot be corrected. As a result, treatment is directed at minimizing the electrolyte and metabolic abnormalities. The combination of an NSAID, including cyclooxygenase-2 inhibitors (because prostaglandin levels are secondarily increased), and a potassium-sparing diuretic can raise the plasma [K^+] toward normal and largely reverse the metabolic alkalosis. Most patients, however, require continued oral potassium and magnesium supplements because drug therapy is rarely completely effective.

RESPIRATORY ACIDOSIS AND ALKALOSIS

Alveolar ventilation provides the oxygen necessary for oxidative metabolism and eliminates the carbon dioxide produced by these metabolic processes. It is therefore appropriate that the main physiologic stimuli to respiration are a reduced arterial oxygen tension (P_aO_2), termed hypoxemia, and an elevated P_aCO_2. Carbon dioxide stimulates ventilation through its pH effect. In contrast, the initial hypoxemic enhancement of ventilation is mostly mediated by chemoreceptors in the carotid bodies, which are located near the bifurcation of the carotid arteries. A number of disorders can be responsible for acute and chronic respiratory acidosis and alkalosis [*see Tables 6 and 7*].

Table 6 Causes of Respiratory Acidosis

Disorder	*Cause*
Suppression of the medullary respiratory center	Sedative medications Oxygen administration in chronic lung disease Sleep apnea (also caused by extreme obesity) Central nervous system lesions (uncommon) Cardiopulmonary arrest
Reduced respiratory muscle function	Acute Muscle weakness or paralysis (myasthenia gravis, periodic paralysis, intraperitoneal aminoglycosides, Guillain-Barré syndrome, botulism, severe hypokalemia, severe hypophosphatemia) Chronic Muscle weakness: poliomyelitis, amyotrophic lateral sclerosis, myxedema Kyphoscoliosis
Upper airway obstruction	Aspiration of a foreign body or vomitus Obstruction in sleep apnea Laryngospasm
Disorders affecting pulmonary gas exchange	Acute Acute respiratory distress syndrome Acute cardiac pulmonary edema Severe asthma or pneumonia Pneumothorax or hemothorax Chronic Obstructive pulmonary diseases
Inadequate mechanical ventilation	—

Table 7 Causes of Respiratory Alkalosis

Disorder	*Cause*
Hypoxemia	Pulmonary disease: pneumonia, emboli, edema, interstitial fibrosis Congestive heart failure Severe anemia High-altitude exposure
Direct stimulation of the medullary respiratory center	Hyperventilation syndrome Hepatic encephalopathy Sepsis or fever Salicylate intoxication* After rapid correction of metabolic acidosis Pregnancy (i.e., increased progesterone) Neurologic disorders (cerebrovascular accidents, pontine tumors)
Excessive mechanical ventilation	—

*Respiratory alkalosis is the initial disorder observed, although metabolic acidosis develops later if intoxication is severe.

Diagnosis

Clinical manifestations Severe respiratory acidosis can produce a variety of neurologic abnormalities. Initial symptoms include headache, blurred vision, restlessness, and anxiety, which can progress to tremors, asterixis, delirium, and a somnolence termed carbon dioxide narcosis. Some of these signs, including papilledema, appear to be caused by the acidemia-induced increase in cerebral blood flow. Overall, these signs seem to result from the fall in cerebrospinal fluid pH, not from the changes in the arterial pH or P_aCO_2.

The symptoms produced by respiratory alkalosis are related to increased irritability of the central and peripheral nervous systems and include light-headedness, altered consciousness, paresthesias of the extremities and circumoral area, cramps, carpopedal spasm (indistinguishable from that caused by hypocalcemia), and syncope. In critically ill patients, a variety of arrhythmias may also occur. These abnormalities are thought to relate to the ability of alkalosis to reduce cerebral blood flow and to increase membrane excitability; a decrease in ionized calcium or magnesium also contributes to heightened membrane excitability.

Laboratory tests Primary respiratory alkalosis can be diagnosed when the blood pH is greater than 7.42 in the presence of a reduced P_{CO_2}. It can also be identified as a second primary disorder in patients with metabolic acidosis by use of the Winter formula [*see* Metabolic Acidosis, *above*].

Treatment

Usually, respiratory failure causes acute respiratory acidosis. As a result, treatment is normally indicated, and the patient may

require mechanical ventilation [*see 14:VIII Respiratory Failure*]. In contrast, treatment of respiratory alkalosis is generally not necessary; evaluation should be aimed at diagnosing and correcting the underlying disorder.

Plasma Potassium Disorders

Potassium is the major intracellular cation, with much higher concentrations inside cells (123 to 140 mEq/L) than in the extracellular space (3.5 to 5.0 mEq/L). This concentration difference is preserved by the Na^+,K^+-ATPase pump, which actively transports sodium ions out of most cells and potassium ions into most cells. The difference between potassium concentration inside and outside the cell is the major determinant of membrane excitability. Thus, despite the comparatively small quantity of extracellular potassium, slight changes can have dramatic effects on muscle contraction and nerve conduction.

POTASSIUM HOMEOSTASIS

In the United States, the daily dietary potassium intake varies from 40 to 120 mEq (i.e., 1,560 to 4,680 mg). Under normal conditions, about 90% of dietary potassium is excreted in the urine; most of the remainder is eliminated in the stool. Gastrointestinal potassium losses increase in patients with renal failure, but the significance of this adaptation is uncertain.

Only about 50% of an oral or intravenous potassium load appears in the urine during the first 4 hours after administration. Marked and potentially life-threatening hyperkalemia can occur if the remaining potassium is confined to the extracellular fluid; this volume is only about 14 L in a 70 kg male. Transport of most of this potassium into cells before excretion in the urine minimizes the rise in the plasma $[K^+]$.

The most important factors in the transport of dietary potassium into cells are the influences of insulin and $beta_2$-adrenergic receptors. Insulin stimulates the Na^+,K^+-ATPase pump, leading to a more rapid rate of potassium entry. Recent evidence suggests that the rate of disposal of an intravenous potassium load is lower in African Americans than in white Americans.[31] Activation of $beta_2$-adrenergic receptors also promotes potassium movement into cells. Aldosterone is the most important hormone involved in potassium secretion by epithelial surfaces, including epithelial cells of the renal tubule. Aldosterone appears to be less important to the transport of potassium into other cells.

Renal Regulation of Potassium Secretion

Potassium is freely filtered at the glomerulus. The concentration of potassium ions entering the early proximal tubule is approximately 4 mEq/L, which is identical to the plasma $[K^+]$. By the time the glomerular filtrate reaches the distal tubule, 90% of the filtered potassium has been reabsorbed. Thus, renal potassium excretion occurs almost exclusively by secretion in the collecting tubule.

Potassium secretion into the collecting tubule takes place in the principal cells [*see Figure 3*]. Movement of potassium ions from the tubular cells into the lumen is controlled by the rates of (1) dietary potassium intake; (2) sodium reabsorption, which generates a negative electrical gradient that favors potassium movement from the cell interior to the tubule lumen; and (3) distal urine flow, which maintains a high concentration gradient of tubular cell to tubular lumen concentration by washing away secreted potassium.

Aldosterone produced in the adrenal glands enters the principal cells from the antiluminal, or capillary, surface. Once inside, aldosterone binds to receptors that increase the number of open sodium channels on the luminal cell membrane. The number and activity of Na^+,K^+-ATPase pumps in the cell membrane also increase. The ensuing rise in cellular potassium leads to secretion into the collecting tubular lumen down the favorable concentration and electrochemical gradient.

In states of potassium depletion, secretion of potassium ions by the principal cells is reduced (provided the kidney is not the source of potassium wasting) and potassium reabsorption is stimulated. This process occurs in the adjacent intercalated cells [*see Figure 3*].

HYPOKALEMIA

Causes of Hypokalemia

Hypokalemia, which can be defined as a plasma $[K^+]$ lower than 3.5 mEq/L, may be caused by low potassium intake, a shift of potassium into cells, or potassium losses from the body [*see Table 8*]. In most persons, potassium losses occur from the gastrointestinal tract, skin, or kidneys.

Low potassium intake Inadequate potassium intake is uncommon in healthy individuals in developed nations because potassium is abundant in most foods. Furthermore, if intake is diminished, urinary and intestinal losses can be reduced to less than 15 mEq/day.

Altered potassium distribution Even in the presence of normal potassium stores in the body, several conditions can cause transport of potassium into cells and hence reduce the plasma $[K^+]$.

Metabolic and, to a much lesser degree, acute respiratory alkalosis may be associated with hypokalemia. The principal mechanism in this setting is the transfer of hydrogen ions out of cells as part of the buffering response that minimizes the rise in the extracellular pH; electroneutrality is maintained in part by potassium entry into cells. The relation between the degree of hypokalemia and the increase in blood pH varies greatly. Potassium loss from the body also plays an important role in the hypokalemia observed with the metabolic alkalosis caused by vomiting or diuretic use.

A catecholamine surge, which occurs during an acute myocardial infarction or delirium tremens, can result in an acute intracellular shift of potassium caused by stimulation of $beta_2$-adrenergic receptors. This phenomenon, in which epinephrine can convert mild hypokalemia to severe hypokalemia, may contribute to the apparent increase in coronary mortality observed in studies of patients with left ventricular hypertrophy and mild hypertension who have been treated with moderate- to high-dose thiazide diuretics. The plasma $[K^+]$ typically falls during insulin administration in patients with diabetic ketoacidosis, despite a normal or increased level on presentation. This decrease results in part from insulin-stimulated transfer of potassium into cells. A similar problem can occur when intravenous dextrose is given to nondiabetic patients; for example, 5% dextrose in water can temporarily worsen hypokalemia and possibly lead to cardiac arrhythmias.

Rare causes of redistribution-induced hypokalemia include folic acid or vitamin B_{12} administration in the treatment of megaloblastic anemias (in which the rapid production of new cells results in the uptake of potassium from the extracellular fluid) and

Table 8 Causes of Hypokalemia

Disorder	*Cause*
Inadequate potassium intake	Low dietary content
Cellular translocation of potassium	Metabolic and, to a lesser degree, acute respiratory alkalosis $Beta_2$-adrenergic stimulation (myocardial infarction, delirium tremens) Insulin administration (or glucose administration in nondiabetics) Increased cell proliferation during the treatment of megaloblastic anemia Hypokalemic periodic paralysis Hypothermia Barium poisoning Theophylline toxicity (may be caused by catecholamine or insulin-induced shifts)
Losses of potassium from the body	Extrarenal losses Gastrointestinal (i.e., diarrhea) Skin losses Profuse sweating Extensive burns Renal losses Normotensive Diuretics Vomiting or nasogastric suction* Hypomagnesemia Nonreabsorbable anions (high-dose penicillins) Levodopa Tubular disorders: classic type 1 RTA, Bartter and Gitelman syndromes, drugs (cisplatin, aminoglycosides, amphotericin B), and lysozymuria (in leukemia) Hypertensive with high mineralocorticoid activity Low plasma renin activity: primary aldosteronism, licorice ingestion, and Liddle syndrome Normal or high plasma renin activity: renal artery stenosis, malignant hypertension, and Cushing syndrome

*Gastric losses of potassium also contribute but are minor.

poisoning with barium salts (which block cell membrane channels that normally allow potassium to leave the cells). Patients undergoing gastrointestinal radiographic procedures are not at risk for poisoning with barium salts, because barium sulfate, which is used in gastrointestinal studies, is not absorbed into the systemic circulation. Severe theophylline overdoses also may cause hypokalemia resulting from catecholamine-induced potassium shifts or glucose-induced insulin release.

Hypokalemic periodic paralysis is a rare disorder of uncertain cause that is characterized by potentially fatal episodes of muscle weakness or paralysis resulting from the sudden movement of potassium into cells. Acute attacks are often precipitated by rest after exercise, stress, or a carbohydrate meal, events that are often associated with increased release of epinephrine or insulin. Attacks tend to last for 6 to 24 hours but can be longer.

The disorder may be familial, and the abnormal gene in most patients seems to code for the part of the calcium channel in skeletal muscle that is blocked by dihydropyridine calcium channel blockers (e.g., nifedipine). However, some patients with hyperthyroidism, particularly Asian males, acquire the disease[32]; in these patients, excess thyroid hormone may predispose to paralytic episodes by increasing Na^+,K^+-ATPase activity in cells [*see* Treatment, *below*].

Potassium losses from the body The gastrointestinal tract can be an important site of potassium wasting, particularly in patients with vomiting or diarrhea. However, the [K^+] in gastric secretions (5 to 10 mEq/L) is much lower than that in intestinal secretions, where it can reach 75 mEq/L. As a result, the loss of large volumes of gastric secretions would be required to produce substantial potassium depletion. The decreased [K^+] that is seen with vomiting results primarily from increased urinary, rather than gastric, losses (see below). By comparison, oral sodium phosphate administered for preparatoin of the colon for colonoscopy may lead to hypokalemia in elderly individuals.[33]

Two conditions can cause cutaneous potassium losses resulting in hypokalemia: exercise in a hot, humid environment and severe burns. Persons who undergo intense physical exercise may lose more than 10 L of sweat a day. This loss can lead to significant potassium depletion even though the [K^+] in sweat is only about 5 mEq/L. Potassium losses in the urine occur most commonly in patients who are either vomiting or receiving diuretic therapy. Extracellular volume depletion promotes aldosterone release, leading to enhanced Na^+-K^+ exchange in the collecting tubule [*see Figure 3*]. Aldosteronism alone is insufficient to permit increased potassium excretion, however, because distal urine flow must be maintained for secreted potassium to be excreted. Distal urine flow is maintained in the presence of vomiting or diuretic therapy because sodium and water escape reabsorption. Diuretics inhibit reabsorption of water and sodium in either the loop of Henle or the distal tubule. The increased tubular bicarbonate load present during vomiting allows sodium and water to escape reabsorption in the proximal tubule. Direct inhibition of potassium reabsorption also contributes to the hypokalemia observed with the use of loop diuretics.

As many as 40% of hypokalemic persons also have hypomag-

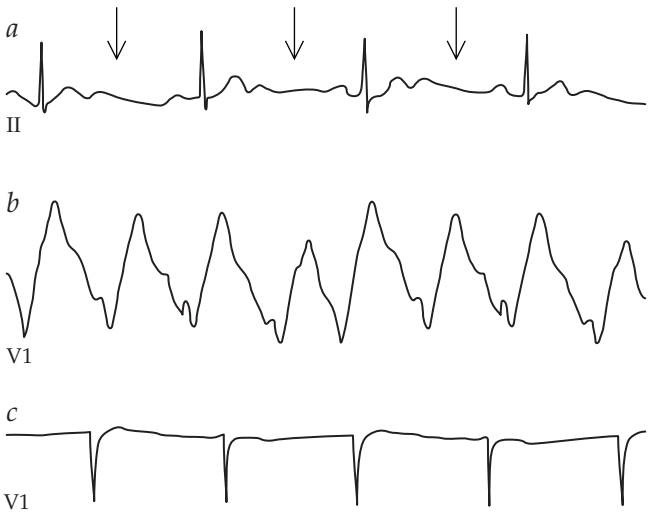


Figure 7 **Both hypokalemia and hyperkalemia can cause changes in the patient's electrocardiogram. (*a*) The ECG from a patient with moderate hypokalemia shows prominent U waves (arrows). (*b*) Marked hyperkalemia results in peaked T waves and widened QRS complexes in this ECG. Ten minutes after the patient receives intravenous calcium, however, all ECG manifestations of hyperkalemia have resolved (*c*).**

nesemia. At times, the cause of potassium wasting, such as diuretic or cisplatin treatment, is also responsible for impaired magnesium reabsorption. However, potassium depletion usually cannot be corrected without restoring magnesium balance. This connection may exist because magnesium can act as a potassium channel blocker. In the loop of Henle, for example, potassium may leak through the cell membrane into the tubular lumen when magnesium levels are low, thereby contributing to ongoing urinary losses.

Several less common causes of tubular dysfunction can also lead to urinary potassium losses. These causes include classic type 1 RTA; drugs (e.g., cisplatin, amphotericin B, and the aminoglycosides); both acute and, less commonly, chronic leukemia; and levodopa administration. Reduced sodium reabsorption, leading to volume depletion, stimulates aldosterone secretion and plays a role in many of these disorders. Primary aldosteronism should be considered, however, whenever unprovoked potassium wasting is detected in a hypertensive person.

Liddle syndrome is a rare cause of hypertension and metabolic alkalosis resulting from a genetic or acquired defect in the epithelial sodium channel, an abnormality that leads to increased sodium reabsorption in the collecting tubule. This disorder is associated with normal or suppressed aldosterone levels. Consequently, it can be treated with potassium-sparing diuretics, such as amiloride, but does not respond to aldosterone antagonists.

Diagnosis

Clinical manifestations In most circumstances, mild hypokalemia (plasma [K^+] 3.0 to 3.5 mEq/L) causes no symptoms. The major disturbances seen with more severe potassium deficiency result from changes in cardiovascular, neuromuscular, and renal function. Cardiac toxicity may be manifested by serious arrhythmias, which occur because hyperpolarization of the myocardial cell membrane leads to a prolonged refractory period and increased susceptibility to reentrant arrhythmias. Other electrocardiographic changes of hypokalemia include T wave depression and prominent U waves [*see Figure 7*].

Hyperpolarization also slows nerve conduction and muscle contraction, which may contribute to muscle weakness, cramps, and paresthesias, although these symptoms are usually not observed until the plasma [K^+] is lower than 2.5 mEq/L. When hypokalemia is severe, it can impair respiratory muscle function, leading to hypoventilation. Because potassium normally permits vasodilatation in response to muscle contraction, severe hypokalemia also predisposes to rhabdomyolysis.

The primary renal manifestations of hypokalemia are polyuria resulting from stimulated thirst and resistance to antidiuretic hormone action. The latter condition is termed nephrogenic diabetes insipidus. Increased thirst is an appropriate response to polyuria, but it is also caused by direct stimulation of the hypothalamic thirst center by hypokalemia. The resistance to antidiuretic hormone is caused by a hypokalemia-induced fall in the number of collecting tubular water channels.[34] Chronic hypokalemia can lead to chronic interstitial nephritis and a decline in glomerular filtration.

When hypokalemia is severe (plasma [K^+] < 2 mEq/L), the ability of the renal tubules to reabsorb sodium and chloride may be impaired. This condition can lead to volume depletion and to the loss of urinary sodium and chloride, even in states of reduced renal perfusion. The mechanism may involve suppressed activity of tubular chloride transporters, many of which also facilitate sodium reabsorption.

Physiologic tests The history and physical examination usually suggest the diagnosis in the hypokalemic patient. Excretion of more than 30 mEq of potassium a day indicates some renal potassium wasting. Patients with extrarenal potassium losses excrete less potassium daily, as do patients who have discontinued diuretic therapy. Once urinary potassium excretion is measured, the following diagnostic possibilities should be considered in the patient with hypokalemia of uncertain origin:

1. In an asymptomatic patient, metabolic acidosis with a low rate of urinary potassium excretion suggests lower gastrointestinal losses because of diarrhea, laxative abuse, or villous adenomas.
2. Metabolic acidosis with renal potassium wasting is most often caused by diabetic ketoacidosis or type 1 RTA.
3. Metabolic alkalosis with potassium wasting and a normal blood pressure is most often caused by surreptitious vomiting (early phase) or diuretic use. In this setting, measurement of the urinary [Cl^-] is often helpful. It is low in patients who have been vomiting, whereas the urinary [Na^+] and urinary potassium excretion may be relatively high. This diagnosis can be determined at the bedside from the urine pH, which should be 7.0 or higher if significant bicarbonaturia is present.
4. Metabolic alkalosis with potassium wasting and hypertension suggests surreptitious diuretic use by a patient with underlying hypertension, renovascular hypertension, malignant hypertension, or primary aldosteronism.
5. In hypokalemic periodic paralysis, a history of exercise or carbohydrate ingestion may be obtained. Urinary potassium is normally conserved during an episode of hypokalemic paralysis.

Treatment

Normalizing the plasma [K^+] is indicated in most cases, although the means of normalization and the quantity and route of potassium administration vary greatly. In the absence of factors that cause transcellular shifts of potassium, a fairly predictable relation exists between the degree of hypokalemia and the extent of total body potassium depletion. For each 1 mEq/L fall in the plasma [K^+], potassium stores fall by 200 to 400 mEq, until the plasma [K^+] drops below 2.0 mEq/L. At that point, the total deficit may exceed 1,000 mEq. With both oral and parenteral replacement, potassium enters the plasma before it is transferred into cells. Consequently, potassium supplementation, particularly when given intravenously, carries the risk of hyperkalemia if the dose is too large or if it is administered too rapidly. Therefore, potassium should be administered orally whenever possible.

Potassium replacement using potassium-rich foods can be effective when metabolic acidosis or renal failure is present. However, dietary potassium is not effective in correcting the potassium deficit that occurs with metabolic alkalosis. In most foods, potassium is bound to poorly reabsorbable anions such as phosphate. Therefore, chloride salts (e.g., potassium chloride, which is used when metabolic alkalosis is caused by diuretics) are usually required to restore potassium losses. An exception to this may be in the preventive treatment of renal calculi with thiazide diuretics, where potassium citrate can maintain a normal serum potassium level.[35]

Potassium can be administered intravenously into a peripheral vein in concentrations as high as 40 mEq/L; higher concentrations can cause phlebitis and thus should be infused only into a

large vein. Except in unusual settings, the rate of administration should probably not exceed 20 to 40 mEq/hr, although dosages as high as 100 mEq/hr have been infused in selected patients with paralysis or life-threatening arrhythmias.[36] Glucose-containing solutions should be avoided because insulin stimulation can drive potassium into cells, thereby exacerbating the hypokalemia.

For patients taking diuretics for high blood pressure, several alternatives are available to correct hypokalemia without use of potassium chloride supplementation. Because patient compliance decreases as the number of medications (including potassium supplements) increases, reducing the diuretic dose (e.g., to 12.5 mg of hydrochlorothiazide) or substituting an alternative agent should be considered first.

Edematous individuals or patients with primary aldosteronism who also experience hypokalemia can be treated with potassium-sparing diuretics (amiloride or spironolactone) until more definitive therapy can be performed (e.g., the surgical removal of an adrenal adenoma in primary aldosteronism).

In patients with hypokalemic periodic paralysis [*see* Altered Potassium Distribution, *above*], administration of potassium chloride can abort acute attacks within minutes; however, long-term potassium chloride supplementation does not usually prevent attacks. Hyperthyroidism should be treated when present, and the administration of a nonselective beta-adrenergic blocker (e.g., propranolol) may prevent episodes in patients with the familial form of the disorder.

HYPERKALEMIA

Hyperkalemia is a common electrolyte disorder. Because a variety of frequently used medications can interfere with normal potassium homeostasis, hyperkalemia is often iatrogenic and therefore preventable. Severe hyperkalemia is associated with altered neuromuscular or cardiac function.

Causes of Hyperkalemia

Hyperkalemia can be caused by excessive potassium intake, increased potassium release from cells, or reduced renal excretion of potassium [*see Table 9*].

Increased potassium intake The increased plasma [K^+] that occurs after an acute oral or intravenous potassium load depends on four factors: (1) the quantity of potassium administered, (2) the ability of some of the excess potassium to enter cells, (3) urinary potassium excretion, and (4) the preceding level of potassium intake—a process called adaptation. The rise in plasma [K^+] is minimized if potassium intake has increased slowly. During a slow increase, renal excretion and, perhaps, cellular entry become more efficient. Increased aldosterone levels and increased Na^+,K^+-ATPase activity in the cortical collecting tubule play important roles in this process. Even in the absence of adaptation, the kidney ultimately excretes the excess potassium, although more slowly. Thus, hyperkalemia resulting from an acute load is transient unless renal potassium excretion is concomitantly reduced.

Increased potassium release from cells Both insulin deficiency and $beta_2$-adrenergic blockade can increase the plasma [K^+]. In patients with end-stage renal disease who are on hemodialysis, for example, propranolol or labetalol, but not the selective $beta_1$-adrenergic blocker atenolol, can raise the plasma [K^+] level by about 1 mEq/L. At higher doses of atenolol, however, this selectivity may be lost.

An elevation in plasma osmolality results in osmotic water movement from the cells into the extracellular fluid. This event is accompanied by the movement of potassium from cells via two mechanisms. First, the increased cellular [K^+] caused by the loss of water creates a favorable gradient for passive exit of potassium through the potassium channels in the cell membrane. Second, friction forces between solvent (water) and solute carry potassium along with water through the water channels (aquaporins) in the cell membrane.

Acute acidemia can also increase the plasma [K^+] by shifting potassium from cells to plasma. This shift is most likely to occur during the infusion of acids; when arginine hydrochloride is infused, for instance, potassium is exchanged for cationic arginine entering the cells. This shift, however, appears to be important in inorganic acidoses only. In comparison, hyperkalemia in lactic acidosis is caused primarily by cell breakdown, and hyperkalemia in ketoacidosis results from insulin deficiency and hyperglycemia.

Massive cell breakdown that occurs during rhabdomyolysis, during abscess formation, with cell necrosis (e.g., bowel infarction), or after chemotherapy for leukemia or lymphoma releases potassium into the plasma. Less common causes of an increased plasma [K^+] include cardiac glycoside intoxication, a result of partial inhibition of the Na^+,K^+-ATPase pump; succinylcholine administration, which depolarizes the cell membrane and thus permits potassium to leave the cells; and the rare disorder hyperkalemic periodic paralysis.

Impaired renal excretion of potassium The kidney, the major site of potassium excretion, has the distinct ability to increase the excretion rate as the plasma [K^+] rises. Thus, sustained hyperkalemia is virtually always associated with some impairment of renal potassium elimination. Distal nephron hydrogen ion secre-

Table 9 Causes of Hyperkalemia

Disorder	*Cause*
Pseudohyperkalemia	Traumatic hemolysis during blood drawing Thrombocytosis Marked leukocytosis
Increased intake of potassium*	Increased dietary potassium Increased intake of potassium-containing medications Increased release of endogenous potassium (as occurs during rhabdomyolysis)
Increased release from cells	Insulin deficiency $Beta_2$-adrenergic blockade Acute metabolic acidosis Cell breakdown Rhabdomyolysis Trauma or infection with abscess formation Tumor lysis syndrome after chemotherapy Marked exercise (transient; resolves with rest) Massive cardiac glycoside intoxication (e.g., digoxin) Succinylcholine Arginine infusion Hyperkalemic periodic paralysis
Impaired renal potassium excretion	Oligoanuric renal failure Marked volume depletion (prerenal diseases) Type 4 RTA Hyperkalemic type 1 RTA [*see Table 3*]

*Sustained hyperkalemia associated with increased intake is always accompanied by impaired renal excretion.

Table 10 Evaluation of the Renal Defect in Hyperkalemia

Disorder	*Urine pH*	*Urinary [Na^+] (mEq/L)*	*Plasma Aldosterone Level*
Renal failure	< 5.3	> 20	Normal or increased
Marked volume depletion	Variable	< 20	Increased
Type 4 RTA	< 5.3	> 20	Low*
Hyperkalemic type 1 RTA	> 5.3	> 20	Normal or increased

*A normal aldosterone level is inappropriate in the presence of hyperkalemia.
Urinary [Na^+]—urinary sodium concentration

tion is frequently impaired as well, leading to a normal-anion-gap metabolic acidosis, because the collecting tubule is the site of both potassium and hydrogen ion secretion. Renal potassium excretion can be impaired by oligoanuric renal failure, reduced effective arterial volume, type 4 RTA, and hyperkalemic type 1 RTA.

Oligoanuric renal failure impairs, but does not halt, potassium excretion. Although renal failure predisposes to potassium retention, excretion of dietary potassium usually continues in chronic renal disease until the GFR is less than 5 to 10 ml/min. Although less efficient in oligoanuric renal failure, renal excretion can still occur because most urinary potassium is derived from tubular secretion, not glomerular filtration.

When the effective arterial volume is reduced, delivery of sodium and water to the collecting tubule is also diminished. Adequate distal delivery of sodium and water is essential to normal potassium excretion. Distal delivery is characteristically reduced, for example, in patients with advanced congestive heart failure or cirrhosis. Both a fall in GFR and a rise in proximal sodium and water reabsorption, mediated in part by angiotensin II, play a role in the reduction of potassium excretion in this setting. Although aldosterone enhances sodium reabsorption and potassium secretion in these conditions, the increase is limited because less sodium is available for reabsorption. Furthermore, secreted potassium may not be eliminated completely if urine flow is markedly decreased.

Aldosterone deficiency or resistance (i.e., type 4 RTA) is responsible for the hyperkalemia in more than 75% to 85% of those cases in which patients have persistent hyperkalemia in the absence of any apparent cause, such as advanced renal failure, an offending drug, or a potassium load.

Type 4 RTA is also associated with diminished hydrogen ion secretion, which is caused in part by impaired ammoniagenesis in the proximal tubule, a possible result of hyperkalemia-induced intracellular alkalosis. The result is a diminished supply of ammonia in the more distal nephron segments. Consequently, an insufficiency of buffer in the urine plays an important role in the limited hydrogen ion secretion and explains why the urine pH is characteristically acidified (< 5.3) in this disorder.

Several conditions cause type 4 RTA [*see Table 4*]. Hyporeninemic type 4 RTA is the most common form of the disorder and occurs most often in patients with mild renal insufficiency. Diabetic nephropathy or chronic interstitial nephritis is often present. Medications such as NSAIDs can also contribute to this disorder.

The pathophysiology of type 4 RTA in hyporeninemic patients is incompletely understood. Decreased angiotensin II production is clearly important, because potassium and angiotensin II are the major physiologic stimuli for aldosterone release. However, at least some patients must also have a concurrent adrenal defect, because observations indicate that the normal adrenal gland can release angiotensin II. It has also been suggested that in some patients, the reduced renin may be secondary. For example, volume expansion, which increases atrial natriuretic peptide, inhibits both renin and aldosterone release.

In other cases of type 4 RTA, the defect in aldosterone synthesis and release seems confined to the adrenal glands. Heparin (or perhaps its preservative, chlorbutol) is a potent inhibitor of aldosterone production, even when given subcutaneously in low doses. Heparin use is not widely associated with hyperkalemia, however, because moderate aldosterone deficiency is not sufficient to cause significant potassium retention in the great majority of patients who still have normal renal function. Hyperkalemia can occur, however, if renal disease is present or if agents that impair the renin-angiotensin-aldosterone system, such as NSAIDs or ACE inhibitors, are concomitantly administered.

The possibility of panadrenal insufficiency (i.e., a deficiency of both aldosterone and cortisol, also termed Addison disease) should be assessed before the diagnosis of selective aldosterone deficiency is made. When panadrenal insufficiency is present, hypoadrenocorticism often impairs water excretion, leading to hyponatremia in addition to the hyperkalemia resulting from aldosterone deficiency.

Although classic type 1 RTA is associated with hypokalemia, some patients with impaired hydrogen ion secretion by the collecting tubule also cannot secrete potassium normally. This condition, hyperkalemic type 1 RTA, is most often seen with urinary tract obstruction but can also be caused by tubulointerstitial damage from such disorders as sickle cell nephropathy, lupus nephritis, or transplant rejection. The hyperkalemia and metabolic acidosis are caused by damage to both the potassium-secreting principal cells and the acid-secreting intercalated cells. Hyperkalemic type 1 RTA can be distinguished from type 4 RTA by the persistently alkaline urine (pH > 5.3) and by a normal or increased plasma aldosterone level [*see Table 10*].

Finally, it is important to remember the effects of medications that can limit potassium secretion directly.[37] Obvious agents are the potassium-sparing diuretics: amiloride, triamterene, and spironolactone. Another, less common agent is trimethoprim-sulfamethoxazole. Trimethoprim in this combination or alone can impair potassium secretion by blocking sodium transport from the tubular lumen into the collecting tubule cell, in a mechanism similar to that of amiloride.

Evidence suggests that aldosterone may be vasculotoxic, cardiotoxic, and nephrotoxic.[38] As a consequence, the combined use of ACE inhibitors or angiotensin receptor blockers (ARBs) with spironolactone has increased. Initial studies of combined ACE and ARB therapy showed very little toxicity from hyperkalemia; however, the risk of hyperkalemia has been found to be associated with combination therapy.[39,40] When multiple medications are administered, each of which can reduce urinary potassium secretion, care must be taken to avoid the morbidity and potential mortality associated with severe hyperkalemia.[41] This is particularly important in patients with impaired renal function.

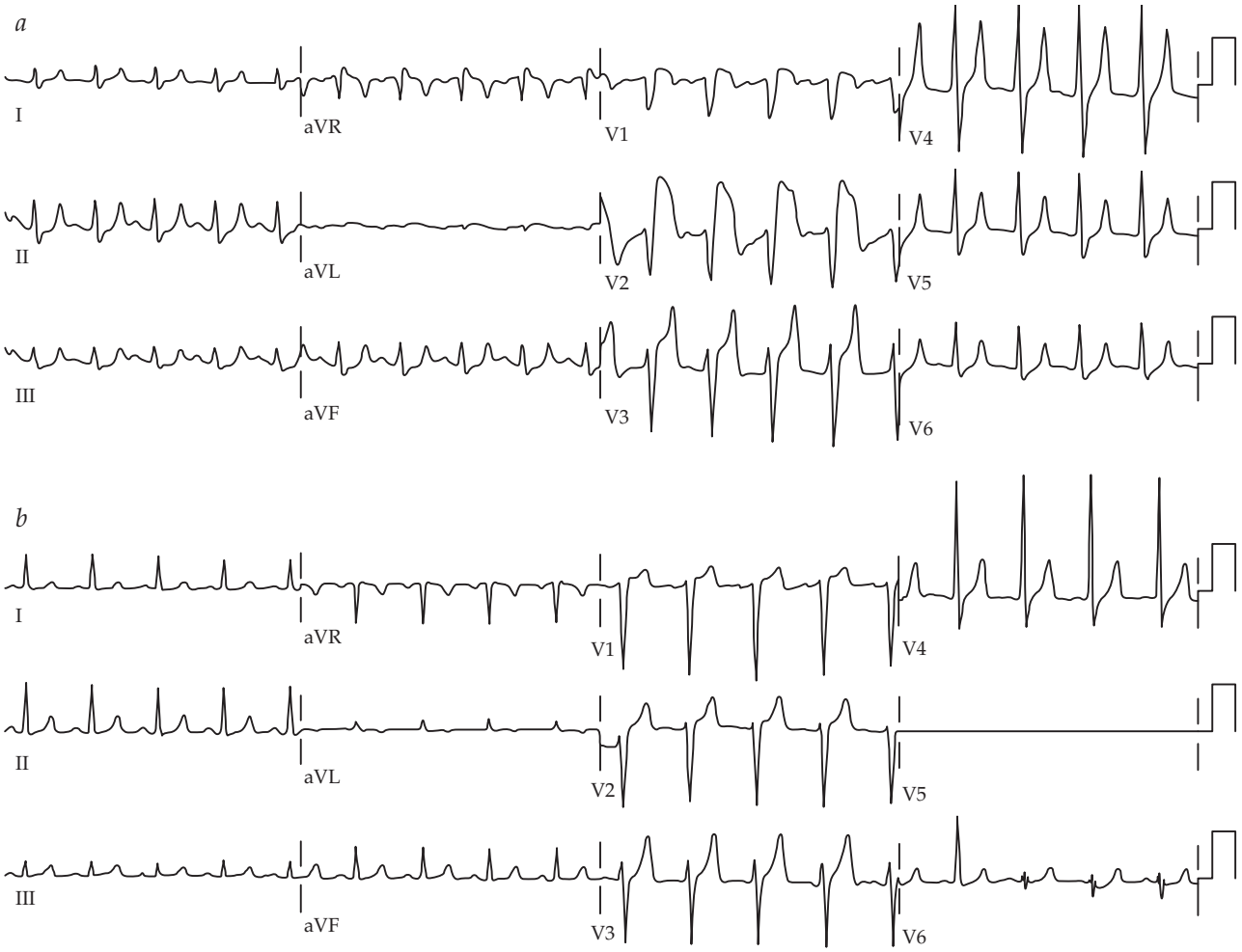


Figure 8 **Pseudoinfarction pattern caused by hyperkalemia. (*a*) ECG from a patient with a serum potassium level of 7.9 mEq/L. The anterior chest leads show ST segment elevation, and there is peaking of the T waves, most notably in leads V3 and V4. (*b*) Resolution of the ECG changes after correction of the serum potassium level.**

Diagnosis

Clinical manifestations The symptoms of hyperkalemia are related to impaired neuromuscular transmission. However, the neuromuscular manifestations are not specific. The earliest findings are paresthesias and weakness, which can progress to paralysis affecting respiratory muscles. These symptoms are similar to those seen with hypokalemia; cranial nerve function, however, characteristically remains unaffected.

The ease of generating an action potential (called membrane excitability) is related both to the magnitude of the resting membrane potential and to the activation state of membrane sodium channels. The opening up of these sodium channels, leading to the passive diffusion of extracellular sodium into the cells, is the primary step in this process.

According to the Nernst equation, the resting membrane potential is related to the ratio of the intracellular [K^+] to the extracellular [K^+]. An elevation in the (extracellular) plasma [K^+] decreases this ratio and therefore partly depolarizes the cell membrane (i.e., it makes the resting membrane potential less electronegative). This change initially increases membrane excitability because less of a depolarizing stimulus is required to generate an action potential. However, persistent depolarization inactivates sodium channels in the cell membrane, thereby producing a net decrease in membrane excitability that may be manifested clinically by impaired cardiac conduction, muscle weakness, or paralysis.

In general, severe symptoms of hyperkalemia do not occur until the plasma [K^+] is higher than 7.5 mEq/L. There is substantial variability among patients, however, because factors such as concomitant hypocalcemia, metabolic acidosis, and the rate at which hyperkalemia develops can increase the toxicity of excess potassium.[42]

Physiologic tests Careful monitoring of the electrocardiogram and muscle strength is necessary to assess the functional consequences of hyperkalemia. At a plasma [K^+] higher than 7.5 to 8.0 mEq/L, severe muscle weakness or marked ECG changes are potentially life threatening and require immediate treatment with almost all of the available modalities [*see* Treatment, *below*].

The earliest ECG abnormality is symmetrical peaking of T waves, which is followed by reduced P wave voltage and widening of the QRS complexes [*see Figure 7*]. If left untreated, severe hyperkalemia may ultimately cause a sinusoidal ECG pattern, with one oscillation representing a wide QRS complex and the complementary oscillation representing an abnormal T wave. ECG changes do not usually occur until the plasma [K^+] exceeds 6.5 mEq/L, and they are more likely to develop when the rise in potassium occurs rapidly. However, there is no absolutely predictive relation between the severity of the elec-

trolyte disturbance and the ECG pattern. In rare cases, the ECG pattern can remain unchanged even in patients with a plasma [K^+] above 9.0 mEq/L. In others, a pseudoinfarction pattern may be observed [*see Figure 8*].[43]

When hyperkalemia is sustained, reduced renal elimination is usually the cause [*see Table 9*]. When type 4 RTA is documented, the plasma cortisol concentration should be measured to determine whether complete adrenal insufficiency is present.

Differential Diagnosis

Pseudohyperkalemia should be considered when there is evidence of hemolysis in the sample or when the platelet or white blood cell count (WBC) is markedly increased. In contrast to true hyperkalemia, which is associated with altered neuromuscular or cardiac function, pseudohyperkalemia does not put patients at risk.

In this condition, the plasma [K^+] measured by the clinical laboratory is increased because potassium is released during clotting that occurs after the blood specimen has been obtained. This condition is most frequently caused by hemolysis resulting from traumatic blood drawing but is also seen with marked thrombocytosis (platelet count > 1,000,000/μl) or leukocytosis (WBC > 100,000/μl). It can also be observed as familial pseudohyperkalemia, in which potassium leaks out of cells at room temperature after clotting, without overt hemolysis.[44]

The diagnosis of pseudohyperkalemia is made by demonstrating that the [K^+] is normal in a nonhemolyzed plasma sample; such a sample is collected by drawing the blood into a heparinized tube, thus preventing clotting. In cases in which no cause for hyperkalemia can be determined (e.g., normal renal function) and the serum [K^+] is over 1 mEq/L higher than the plasma level, the cause of hyperkalemia is an in vitro phenomenon if the patient has a normal serum potassium level after blood is drawn without a tourniquet.[45] Treatment of pseudohyperkalemia is not indicated, because the in vivo (i.e., plasma) [K^+] is normal.

Treatment

In true hyperkalemia, particularly when it is severe (plasma [K^+] > 6.0 mEq/L), discontinuance of all medications that adversely affect potassium balance is mandatory.[46] These medications include nonselective beta blockers, ACE inhibitors, potassium-sparing diuretics, NSAIDs, and trimethoprim. Salt substitutes, which contain potassium chloride, should also be avoided. Persons with mild hyperkalemia (plasma [K^+] < 6.0 mEq/L) can usually be treated conservatively with reduction of daily intake to less than 2 g and, if indicated, with the addition of a loop diuretic.

Active treatment to lower the plasma [K^+] or to antagonize its effects on the cell membrane should be started if the plasma [K^+] has risen acutely to 6.0 mEq/L or higher, particularly if ECG manifestations of hyperkalemia are present; several therapeutic options are available [*see Table 11*].

An infusion of calcium rapidly normalizes the ECG. An increase in the [Ca^+] raises the threshold potential, thereby returning membrane excitability to normal. Calcium should be administered, however, in the presence of only those ECG manifestations of hyperkalemia that may precede ventricular fibrillation. Because calcium can exacerbate or precipitate glycoside-induced cardiac arrhythmias, it should be used only when necessary and with great care in patients receiving digoxin or other digitalis preparations.

Calcium is administered intravenously as the gluconate or chloride salt (10 ml of 10% calcium gluconate over 2 to 3 minutes, with ECG monitoring). Normalization of the ECG may persist for less than 60 minutes; furthermore, this treatment does not alter the plasma [K^+] or total potassium body stores. Consequently, measures that will lower the plasma level and remove potassium from the body should be initiated simultaneously.

Sodium bicarbonate, glucose (with or without insulin), and beta$_2$-adrenergic receptor stimulation lower the plasma [K^+] by promoting the entry of potassium from the extracellular fluid into cells. The administration of sodium bicarbonate reduces the hydrogen ion concentration in the extracellular fluid. Theoretically, if the Na^+-H^+ exchanger were in an active mode, the administration of sodium bicarbonate would favor the movement of H^+ out of cells as Na^+ enters. The subsequent exit of sodium by the Na^+,K^+-ATPase pump would shift potassium into the cell. In the steady state, however, the Na^+-H^+ exchanger appears to be inactive.[47] By comparison, the Na^+-H^+ exchanger is activated by intracellular acidosis.

Several studies have found sodium bicarbonate therapy to be ineffective in the treatment of acute hyperkalemia, despite a rise in the serum [HCO_3^-]. Furthermore, bicarbonate does not appear to potentiate the hypokalemic action of beta$_2$-adrenergic agonists or the combination of glucose and insulin, which are more effective in this setting.[48] Because of these findings, sodium bicarbonate, if used in the treatment of hyperkalemia, should be used in combination with other agents.[47]

Beta$_2$-adrenergic agonists can also transiently lower the serum potassium level. Albuterol can be administered at a dosage of 10 to 20 mg in 4 ml of saline by nasal inhalation over 10 minutes or by a 0.5 mg I.V. infusion. The blood potassium level usually falls by 0.5 to 1.5 mEq/L within 30 minutes after the I.V. infusion, but more than 1 hour is required for a peak effect when albuterol is inhaled.[49]

The hypokalemic effect of insulin is observed within 30 to 60 minutes, and it can be achieved by simply administering 25 g of glucose intravenously, because a glucose infusion rapidly stimulates insulin release. Alternatively, 10 units of regular insulin can be given intravenously with this glucose load to patients with diabetes, although this regimen can cause hypoglycemia in normal persons. Regardless of the method, care must be taken to avoid producing severe hyperglycemia by the glucose infusion, because a high plasma glucose concentration can exacerbate hyperkalemia by causing water, which contains potassium, to shift from the intracellular compartment into the plasma.

Like the administration of calcium, the administration of glucose and insulin, bicarbonate, and beta$_2$-adrenergic agonists is a temporary measure because total body potassium stores are not reduced. Therefore, additional therapy is required to ensure that the plasma [K^+] does not return to pretreatment levels. Furthermore, in some patients, lowering the plasma [K^+] with these

Table 11 Treatment of Hyperkalemia

Agent	*Result*
Calcium	Antagonism of hyperkalemic membrane effects
Glucose infusion (with insulin in diabetics) Sodium bicarbonate Beta$_2$-adrenergic agonists	Potassium movement into cells
Loop or thiazide diuretics Cation exchange resins Hemodialysis or peritoneal dialysis	Potassium removal from the body

agents shortly before dialysis can lead to reduced dialytic potassium removal, possibly resulting in rebound hyperkalemia after the dialysis treatment has been completed.[49]

Increasing urine flow with diuretics is sometimes useful, but renal insufficiency frequently limits the effectiveness of diuretics. However, the Na^+-K^+ cation exchange resin sodium polystyrene sulfonate can be administered in sorbitol (to promote diarrhea) orally or as a retention enema. The use of sodium polystyrene sulfonate can exchange 4 mEq of Na^+/g for 4 mEq of K^+; therefore, in theory, 30 g of sodium polystyrene sulfonate could remove 120 mEq of potassium. However, because the quantity of gastrointestinal potassium available for exchange may be as little as 5 mEq/day, the effectiveness of sodium polystyrene sulfonate has been questioned.[47] In some studies, sodium polystyrene sulfonate was no more effective than the laxative in which it was administered. Some experts no longer recommend use of sodium polystyrene sulfonate in sorbitol; however, the diarrhea this agent induces may in itself be helpful in reducing the serum [K^+]. I recommend use of sodium polystyrene sulfonate if the urine output is low and if dialysis is not readily available.

Dialysis should be considered in severe or refractory cases of hyperkalemia, particularly when there is advanced renal failure. When available, hemodialysis is preferable in the acute setting because it removes potassium much more quickly than peritoneal dialysis does.

If aldosterone deficiency has been documented in cases of chronic hyperkalemia that are inadequately controlled by diet or diuretics, aldosterone replacement with fludrocortisone acetate may be useful. The combination of fludrocortisone and a loop diuretic may limit the development of hypertension or edema. However, because its onset of action may take several days or longer, this therapy is not sufficiently effective to be used in the treatment of acute, life-threatening hyperkalemia.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Kraut JA, Madias NE: Approach to patients with acid-base disorders. Respir Care 46:392, 2001

2. Gluck SL: Acid sensing in renal epithelial cells. J Clin Invest 114:1696, 2004

3. Lolekha PH, Vanavanan S, Lolekha S: Update on value of the anion gap in clinical diagnosis and laboratory evaluation. Clin Chim Acta 307:33, 2001

4. Shaer AJ, Rastegar A: Lactic acidosis in the setting of antiretroviral therapy for the acquired immunodeficiency syndrome: a case report and review of the literature. Am J Nephrol 20:332, 2000

5. Uribarri J, Oh MS, Carroll HJ: D-lactic acidosis: a review of clinical presentation, biochemical features, and pathophysiologic mechanisms. Medicine (Baltimore) 77:73, 1998

6. Vanholder R, Sever MS, Erek E, et al: Rhabdomyolysis. J Am Soc Nephrol 11:1553, 2000

7. Sweeney T, Beuchat C: Limitation of methods of osmometry: measuring the osmolality of biological fluids (editorial). Am J Physiol 264:R469, 1993

8. Cruz DN, Huot SJ: Metabolic complications of urinary diversions: an overview. Am J Med 102:477, 1997

9. Messiaen T, Deret S, Mougenot B, et al: Adult Fanconi syndrome secondary to light chain gammopathy: clinicopathologic heterogeneity and unusual features in 11 patients. Medicine (Baltimore) 79:135, 2000

10. Igarashi T, Sekine T, Inatomi J, et al: Unraveling the molecular pathogenesis of isolated proximal renal tubular acidosis. J Am Soc Nephrol 13:2171, 2002

11. Rungroj N, Devonald AW, Cuthbert F, et al: A novel missense mutation in AE1 causing autosomal dominant distal renal tubular acidosis retains normal transport functions, but is mistargeted in polarized epithelial cells. J Biol Chem 279:13833, 2004

12. Graber M: A model of hyperkalemia produced by metabolic acidosis. Am J Kidney Dis 22:436, 1993

13. Adrogue HJ, Madias N: Management of life-threatening acid-base disorders: first of two parts. N Engl J Med 338:26, 1998

14. Winter SD, Pearson R, Gabow PA, et al: The fall of the serum anion gap. Arch Intern Med 150:311, 1990

15. Caravaca F, Arrobas M, Pizarro JL, et al: Metabolic acidosis in advanced renal failure: differences between diabetic and nondiabetic patients. Am J Kidney Dis 33:892, 1999

16. Gadola L, Noboa O, Marquez, et al: Calcium citrate ameliorates the progression of chronic renal injury. Kidney Int 65:1224, 2004

17. Mehrotra R, Kopple J, Wolfson M: Metabolic acidosis in maintenance dialysis patients: clinical considerations. Kidney Int 64:S13, 2004

18. Eustace J, Astor B, Muntner P, et al: Prevalence of acidosis and inflammation and their association with low serum albumin in chronic kidney disease. Kidney Int 65:1031, 2004

19. Boirie Y, Broyer M, Gagnadoux MF, et al: Alterations of protein metabolism by metabolic acidosis in children with chronic renal failure. Kidney Int 58:236, 2000

20. Kirschbaum B: Spurious metabolic acidosis in hemodialysis patients. Am J Kidney Dis 35:1068, 2000

21. Forsythe SM, Schmidt GA: Sodium bicarbonate for the treatment of lactic acidosis. Chest 117:260, 2000

22. Luft FC: Lactosis acidosis update for critical care clinicians. J Am Soc Nephrol 12(suppl 17):S15, 2001

23. Kraut J, Kurtz I: Use of base in the treatment of severe acidemic states. Am J Kidney Dis 38:703, 2001

24. Okuda Y, Adrogue HJ, Field JB, et al: Counterproductive effects of sodium bicarbonate in ketoacidosis. J Clin Endocrinol Metab 81:314, 1996

25. Glaser N, Barnett P, McCaslin I, et al: Risk factors for cerebral edema in children with diabetic acidosis. N Engl J Med 344:264, 2001

26. Higgins RM, Connolly JO, Hendry BM: Alkalinization and hemodialysis in severe salicylate poisoning: comparison of elimination techniques in the same patient. Clin Nephrol 50:178, 1998

27. Brent J, McMartin K, Phillips S, et al: Fomepizole for the treatment of methanol poisoning. N Engl J Med 344:424, 2001

28. Galla JH: Metabolic alkalosis. J Am Soc Nephrol 11:369, 2000

29. Kurtz I: Molecular pathogenesis of Bartter's and Gitelman's syndromes. Kidney Int 54:1396, 1998

30. Adrogue HJ, Madias NE: Management of life-threatening acid-base disorders: second of two parts. N Engl J Med 338:107, 1998

31. Suh A, DeJesus E, Rosner K, et al: Racial differences in potassium disposal. Kidney Int 66:1523, 2004

32. Manoukian MA, Foote JA, Crapo LM: Clinical and metabolic features of thyrotoxic periodic paralysis in 24 episodes. Arch Intern Med 159:601, 1999

33. Beloosesky Y, Grinblat J, Weiss A, et al: Electrolyte disorders following oral sodium phosphate adminstration for bowel cleansing in elderly patients. Arch Intern Med 163:803, 2003

34. Marples D, Frokiaer J, Dorup J, et al: Hypokalemia-induced downregulation of aquaporin-2 water channel expression in rat kidney medulla and cortex. J Clin Invest 97:1960, 1996

35. Odvina C, Preminger G, Lindberg J, et al: Long-term combined treatment with thiazide and potassium citrate in nephrolithiasis does not lead to hypokalemia or hypochloremic metabolic alkalosis. Kidney Int 63:240, 2003

36. Kruse J, Carlson R: Rapid correction of hypokalemia using concentrated intravenous potassium chloride infusions. Arch Intern Med 150:613, 1990

37. Perazella MA: Drug-induced hyperkalemia: old culprits and new offenders. Am J Med 109:307, 2000

38. Hollenberg N: Aldosterone in the development and progression of renal injury. Kidney Int 66:1, 2004

39. Juurlink D, Muhammad M, Lee D, et al: Rates of hyperkalemia after publication of the Randomized Aldactone Evaluation Study. N Engl J Med 351:543, 2004

40. McMurray J, O'Meara E: Treatment of heart failure with spironolactone: trial and tribulations. N Engl J Med 351:526, 2004

41. Palmer B: Managing hyperkalemia caused by inhibitors of the renin-angiotensin-aldosterone system. N Engl J Med 351:585, 2004

42. Charytan D, Goldfarb DS: Indications for hospitalization of patients with hyperkalemia. Arch Intern Med 160:1605, 2000

43. Pseudoinfarction pattern due to hyperkalemia. N Engl J Med 351:593, 2004

44. Iolascon A, Stewart GW, Ajetunmobi JF, et al: Familial pseudohyperkalemia maps to the same locus as dehydrated hereditary stomatocytosis (hereditary xerocytosis). Blood 93:3120, 1999

45. Wiederkehr MR, Moe OW: Factitious hyperkalemia. Am J Kidney Dis 36:1049, 2000

46. Weiner ID, Wingo CS: Hyperkalemia: a potential silent killer. J Am Soc Nephrol 9:1535, 1998

47. Kamel K, Wei C: Controversial issues in the treatment of hyperkalaemia. Nephrol Dial Transplant 18:2215, 2003

48. Allon M, Shanklin N: Effect of bicarbonate administration on plasma potassium in dialysis patients: interactions with insulin and albuterol. Am J Kidney Dis 28:508, 1996

49. Allon M, Shanklin N: Effect of albuterol treatment on subsequent dialytic potassium removal. Am J Kidney Dis 26:607, 1995

Acknowledgments

Figure 1 Janet Betries.

Figures 2 through 6 Tom Moore.

Figure 7 Electrocardiograms kindly provided by Dr. David Spodick, Professor of Medicine at the University of Massachusetts Medical School, Worcester.

III APPROACH TO THE PATIENT WITH RENAL DISEASE

Biff F. Palmer, m.d.

Diseases of the kidney can present as a variety of clinical syndromes. In some cases, the clinical presentation is directly referable to the kidney, as with the finding of proteinuria or an increased serum creatinine concentration. In other instances, the presentation reflects the impact of impaired renal function on other organ systems, such as edema or shortness of breath resulting from renal salt retention. Still other patients are asymptomatic and are simply found to have an abnormal urinalysis result on routine examination. In all of these situations, a systematic approach will promote an efficient and accurate diagnosis.

First, the clinician should determine the duration of the disease. Second, renal function should be measured, to determine whether the patient has suffered a loss in renal function and, if so, to what degree. Third, the specific syndrome should be identified, on the basis of information obtained through the history and physical examination, routine laboratory testing, and imaging of the kidneys. Assessment of volume status deserves particular attention, because volume abnormalities are common in patients with renal disease and offer an important clue not only to the presence of renal failure but also to its management. Approaching the case in this manner will allow the clinician to establish the correct diagnosis and estimate the duration, course, and severity of the renal disease, as well as to institute appropriate therapy.

Determining Disease Duration

Whether renal disease is acute or chronic is important both in asymptomatic patients with normal renal function and in patients with clinical evidence of renal insufficiency. There are two main reasons to differentiate acute from chronic renal disease. First, acute renal disease is more likely to be self-limited and hence to have a better prognosis. Second, the treatment of renal disease may vary, depending on whether the disease is of recent onset or is long-standing. This is particularly true for treatment of acute renal failure versus treatment for chronic renal failure. There are now specific therapies for certain types of acute renal failure that do not apply to chronic renal failure. Moreover, such therapy may conceivably prevent acute renal failure from becoming chronic.

Several tools can help determine disease duration [*see Table 1*]. Use of old medical records is particularly valuable in dating the onset of an increased serum creatinine concentration, proteinuria, or hematuria. In patients with impaired renal function, measurement of kidney size by renal sonography or plain film of the abdomen is quite useful. Small kidneys (i.e., less than 8 cm in total length, in an adult), are an almost certain sign of chronic renal failure. If the kidneys are normal in size, one cannot be certain whether the patient has acute or chronic renal disease. Radiographic evidence of renal osteodystrophy in the distal clavicles, hands, ribs, skull, spine, and pelvis strongly supports the diagnosis of chronic renal disease. However, patients in early stages of chronic renal failure may not have detectable bone lesions even when the kidney size is small. The most precise way to differentiate acute renal disease from chronic renal disease is to perform a renal biopsy; however, this is not practical or necessary in most cases. Renal biopsy is useful in patients who are suspected of having chronic renal failure because of their history and whose kidneys are of normal size on ultrasound or plain films of the abdomen (kidneys, ureters, and bladder [KUB] view).

Assessment of Renal Function

Once renal disease is discovered, the presence or degree of renal dysfunction should be determined. The glomerular filtration rate (GFR) is generally considered the best measure of renal function. Serial assessment of GFR can allow the clinician to determine the course of the underlying disease by demonstrating either rapid or slow rates of decline in renal function. Accurate determination of renal function also helps the clinician to make adjustments in the dosing of pharmacologic agents so as to prevent the accumulation of drugs and metabolites and, thereby, potential toxicities.

Serum Creatinine

Measurement of the serum creatinine concentration is the most commonly used method for determining the level of renal function. The creatinine concentration can be used to estimate GFR because creatinine varies inversely with the level of renal function. Normal creatinine concentrations range from 0.6 to 1.0 mg/dl in women and 0.8 to 1.3 mg/dl in men.

The major limitation of the serum creatinine level is its insensitivity to mild to moderate reductions in renal function. The relationship between creatinine level and GFR is nonlinear. A change in creatinine from 0.6 mg/dl to 1.2 mg/dl reflects a decline in GFR of approximately 50%. If a previous baseline value for creatinine does not exist for comparison, a creatinine of 1.2 would not draw clinical attention to a potential reduction of GFR. On the other hand, nephrologists are often consulted emergently when a patient's creatinine concentration rises from 5 mg/dl to 8 mg/dl, which is far less critical, because GFR has fallen from approximately 20 ml/min to 15 ml/min, a 25% decline. Nevertheless, it is important to detect changes in GFR at relatively low creatinine values, when renal injury may still be reversible.

Changes in the serum creatinine concentration are also slow to reflect acute changes in renal function; accumulation of creatinine in the blood and achievement of a new steady state occur

Table 1 Tools Used to Differentiate Acute from Chronic Renal Disease

Old medical and hospital records
Determination of kidney size (normal or small)
Radiographic evidence of renal osteodystrophy
Renal biopsy

Note: degree of anemia, if present, and serum phosphate measurement are not useful for this purpose.

gradually. For example, if acute renal failure occurs and the GFR suddenly falls from 100 ml/min to 10 ml/min, the serum creatinine would not rise correspondingly for approximately 7 days.

CREATININE CLEARANCE

A more accurate way to assess the GFR is with a 24-hour urine collection to determine the creatinine clearance. Creatinine is an endogenous marker of filtration that is produced at a relatively constant rate. Creatinine clearance (C_{Cr}) may be calculated by the following equation:

$$C_{Cr}\,(\text{ml/min}) = \frac{\text{urine creatinine (mg/dl)} \times \text{urine volume (ml/min)}}{\text{plasma creatinine (mg/dl)} \times 1{,}440}$$

Normal C_{Cr} is approximately 95 ± 20 ml/min in women and 125 ± 25 ml/min in men. Creatinine is freely filtered and is not reabsorbed; it is thus excreted primarily by filtration. However, 15% to 20% of urinary creatinine is derived from tubular secretion via an organic acid pump in the proximal tubule. Therefore, use of C_{Cr} to assess GFR may be expected to result in an overestimation of the GFR.[1] However, the plasma creatinine (P_{Cr}), as measured by certain picric acid methods (e.g., the Jaffe reaction), is typically inflated because of the presence of noncreatinine chromogens, such as glucose, acetoacetate, pyruvate, uric acid, and proteins in plasma. Thus, the 15% to 20% increase in urine creatinine levels caused by tubular secretion may be counterbalanced by the inflated level of P_{Cr}, resulting in a reasonably close correlation between C_{Cr} and GFR.

The reliability of C_{Cr} assessments, however, is diminished by the variability in tubular secretion of creatinine and by the inability of most patients to accurately collect timed urine samples.[2,3] The accuracy of the urine samples can be estimated on the basis of the normal daily rate of creatinine excretion: 15 to 20 mg/kg lean body weight in women and 20 to 25 mg/kg lean body weight in men. A creatinine excretion rate that is significantly less than the normal daily rate usually indicates an incomplete urine collection. Additionally, tubular secretion of creatinine increases proportionally with declining levels of renal function, resulting in further overestimations of GFR.[4] Prolonged storage of urine can also invalidate urinary creatinine clearance measurements, because high temperatures and low pH promote the conversion of creatine to creatinine in urine.[5]

To avoid the inaccuracies and inconvenience of timed urine collections, the creatinine clearance can be estimated at the bedside by using the patient's age and P_{Cr}, which correlate inversely with GFR, and the patient's ideal body weight (IBW), which correlates directly with GFR:

$$\text{Estimated } C_{Cr} = \frac{(140 - \text{age}) \times (\text{IBW in kg})}{72 \times P_{Cr}}$$

When this equation is used for estimating C_{Cr} in women, the results should be multiplied by 0.85. This equation provides a quick and reasonably accurate estimate of GFR at the bedside and is particularly useful in determining dosage adjustments for pharmacologic agents excreted by the kidney.

Another equation, derived from the Modification of Diet in Renal Disease study, can also be used to more accurately assess renal function[6]:

$$\text{GFR} = 170 \times [\text{Scr}]^{-0.999} \times [\text{age}]^{-0.176} \times [0.762 \text{ if female}] \times [1.180 \text{ if patient is black}] \times [\text{BUN}]^{-0.170} \times [\text{albumin}]^{0.318}$$

This equation estimates GFR rather than creatinine clearance and factors in ethnicity, gender, and serum albumin concentration. The normal GFR is approximately 95 ± 20 ml/min in women and 125 ± 25 ml/min in men.

INULIN CLEARANCE

Inulin is a fructose polymer found in chicory and the Jerusalem artichoke; it is freely filtered by the glomerulus and is neither reabsorbed nor secreted by the tubule. Thus, inulin clearance (C_{In}) is one of the most accurate methods of quantifying renal function. C_{In} can be determined by the following equation:

$$C_{In}\,(\text{ml/min}) = \frac{\text{urine inulin (mg/dl)} \times \text{urine volume (ml/min)}}{\text{plasma inulin (mg/dl)}}$$

Inulin is an exogenous compound and must therefore be administered as a continuous infusion to achieve steady-state concentrations in the blood. The need for continuous infusion and the cost and limited supply of inulin prevent the routine clinical implementation of inulin-clearance testing.

IOHEXOL CLEARANCE

To avoid the aforementioned problems with GFR assessment, investigators use plasma-clearance techniques. Until recently, such measurements depended on radiolabeled markers of GFR, such as technetium-99m–diethylenetriamene pentaacetic acid, chromium-51–ethylenediaminetetraacetic acid, and iodine-125–iothalamate, which are accurate but costly and involve exposure to radioactive material. Plasma clearance of iohexol, a nonionic and nonradiolabeled contrast agent, provides an accurate measurement of GFR.[7,8] Low concentrations of iodine can be measured with x-ray fluorescence techniques in a single blood sample after a bolus injection of iohexol. Because

Table 2 Imaging Studies Used in Evaluating Patients with Renal Disease

Imaging Study	*Use*
Plain radiograph of abdomen (kidneys, ureters, and bladder [KUB])	Determining kidney size and shape; detection of nephrolithiasis (radiopaque) and nephrocalcinosis
Ultrasonography and CT scanning	Determining kidney size and shape, detecting urinary obstruction and radiolucent stones, distinguishing between simple and complex cysts, early evaluation of polycystic kidney disease, evaluation of renal mass
Intravenous pyelogram	Determining kidney size and shape and calyceal anatomy, diagnosis of medullary sponge kidney and papillary necrosis, detection of site and cause of obstruction
Radionuclide studies	Detection of urinary obstruction and urine leak, screening for renal artery stenosis, assessing renal arterial flow
Renal arteriography	Detection of renal artery stenosis, assessing for evidence of vasculitis, distinguishing vascular versus solid masses
Voiding cystourethrogram	Detection of vesicoureteral reflux
Retrograde or antegrade pyelography	Determining site of obstruction, placement of a ureteral stent
Magnetic resonance imaging	Detection of renal mass, detection of renal vein thrombosis

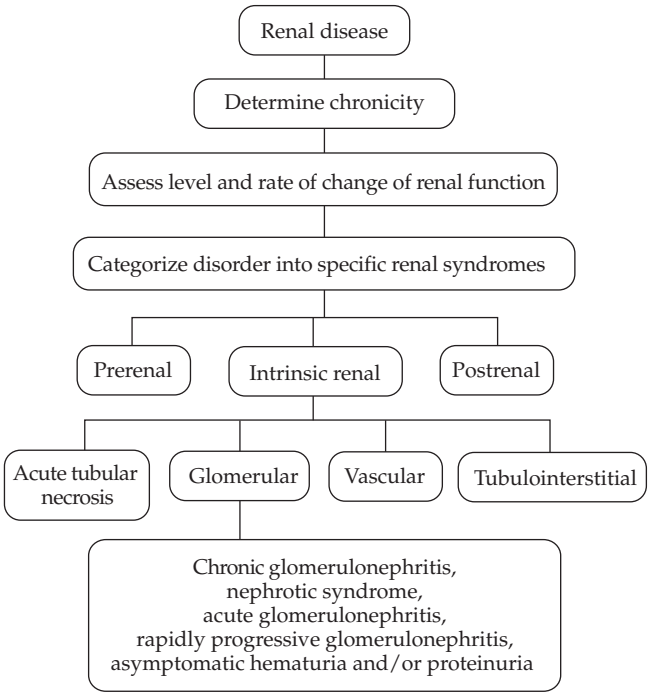


Figure 1 **Approach to the patient with renal disease.**

iohexol is excreted by glomerular filtration, the fall in plasma iodine levels can be converted to a plasma clearance value. Iohexol-derived GFR determinations are accurate over a wide range of values, and iohexol does not appear to be nephrotoxic even at extremely low levels of GFR (< 10 ml/min).[9,10]

Imaging Techniques for the Genitourinary System

Selected use of imaging studies can be an effective tool in the initial evaluation of patients with renal disease [*see Table 2*]. Renal ultrasonography provides information on kidney size, structure, and symmetry; cortical thickness; and the presence of hydronephrosis and nephrolithiasis. At the same time, ultrasound avoids the nephrotoxicity of large-volume radiocontrast procedures, such as the intravenous pyelogram. A plain film of the abdomen (KUB) can be used to assess kidney size and detect calcifications indicative of renal stones or nephrocalcinosis. Magnetic resonance imaging of the genitourinary tract is also a useful, noninvasive imaging modality.[11] In addition to providing a detailed structural image of the kidney, ureters, and bladder, MRI can be used to assess the renal vasculature. Magnetic resonance angiography, in which a gadolinium-based (non–iodine-containing) contrast agent is given intravenously, can be used in the evaluation of renovascular disease. This approach avoids the nephrotoxicity of the iodinated contrast used in conventional angiography. MRI is also an effective tool to detect renal vein thrombosis.

Major Syndromes of Renal Disease

After determining chronicity and current level of renal function, the clinician should attempt to classify the patient's renal disease into one of several syndromes, on the basis of the renal structures most affected [*see Figure 1*]. This classification is based on the information obtained in the history, physical examination, laboratory tests, and selected imaging studies [*see Table 3*]. It is particularly important to identify prerenal and postrenal disorders because they are often readily reversible.

PRERENAL DISEASE

Prerenal disease refers to azotemia that results from a reduction in effective arterial blood volume (EABV). The concept of EABV provides a meaningful way to explain why heart failure (a condition associated with total body volume overload) and severe vomiting (a condition associated with total blood volume depletion) are characterized by a similar renal response. In both of these conditions, the kidney behaves as if renal perfusion is decreased. The kidney is structurally normal and is simply responding to a perceived or actual underfilling in the circu-

Table 3 Urinalysis Findings and Selected Features of Renal Disease Syndromes

Renal Disease	*Urinalysis Findings*	*Other Features*
Prerenal	Specific gravity (SG) 1.020, glucose negative, protein negative, occasional hyaline casts	Signs and symptoms of decreased effective arterial blood volume, heart failure
Postrenal	SG 1.007, protein negative, blood negative, 0–3 erythrocytes/hpf, 0–3 leukocytes/hpf	Decreased urine flow, symptoms of prostatism in men
Acute tubular necrosis	SG 1.010, protein trace, blood negative, 1–3 leukocytes/hpf, many pigmented granular casts, 5–10 renal tubular casts/hpf, occasional renal tubular cells	History of hypotension, sepsis, or administration of nephrotoxins (e.g., radiocontrast, aminoglycosides, cisplatinum)
Nephrotic syndrome	SG 1.015, protein 4+, oval fat bodies, fatty casts	Edema, hypoalbuminemia
Acute tubulointerstitial nephritis	SG 1.012, protein 1+, blood 2+, 20–30 leukocytes/hpf, + stain for eosinophils, few leukocyte casts, 1–15 erythrocytes/hpf	Fever, morbilliform rash, eosinophilia, recent administration of antibiotic
Chronic tubulointerstitial nephritis	SG 1.012, protein 1+, glucose 1+ with normal serum glucose, 10–15 leukocytes/hpf, no bacteria	Type IV renal tubular acidosis, anemia, sodium wasting
Acute glomerulonephritis	SG 1.020, 2+ protein, 3+ blood, 15–20 erythrocytes/hpf, 3–5 erythrocyte casts/hpf, 0–3 leukocytes/hpf	Recent upper respiratory tract infection, nephritic syndrome
Rapidly progressive glomerulonephritis	Same as acute glomerulonephritis	Nephritic syndrome, hemoptysis, rapid loss of renal function

lation. The findings utilized to diagnose prerenal disease are reviewed in detail elsewhere [*see 10:VI Acute Renal Failure*].

POSTRENAL DISEASE

Obstruction of urine flow can occur anywhere along the urinary tract from the renal pelvis to the urethra. Anuria suggests complete urinary obstruction, although anuria can also be a feature of bilateral renal artery thrombosis, acute cortical necrosis, or severe acute tubular necrosis. In the absence of complete obstruction, urine flow may not necessarily be decreased and in fact is often increased. Chronic partial obstruction of the ureters leads to ureteral dilatation, which overcomes the blockage of urine flow. In addition, obstruction impairs the urinary-concentrating ability and thus contributes to a polyuric state. Common causes of urinary obstruction include nephrolithiasis, prostate enlargement, neurogenic bladder in diabetic patients, and an enlarging cervical cancer. The urinalysis result is typically unremarkable in obstructive uropathy. The diagnosis is most often made by demonstrating ureteral dilatation on renal sonography.

INTRINSIC RENAL DISEASE

After excluding prerenal and postrenal disease, the clinician should focus on various causes of intrinsic renal disease. Intrinsic renal disease can be subdivided into acute tubular necrosis, vascular renal disease, tubulointerstitial disease, and glomerular disease [*see Figure 1*].

Acute Tubular Necrosis

Acute tubular necrosis is the most common cause of intrinsic renal disease. This disorder most commonly results from ischemic or nephrotoxic injury to the kidney.[12] Acute ischemic injury is generally the consequence of severe hypotension, as occurs in cardiogenic shock, sepsis, or hemorrhage. Common causes of nephrotoxic injury include aminoglycosides and radiocontrast agents. Characteristic findings on urinalysis include granular and epithelial cell casts [*see Figure 2*]. Free renal tubular cells are also seen. The clinical characteristics of acute tubular necrosis are reviewed in detail elsewhere [*see 10:VI Acute Renal Failure*].

Renal Vascular Disease

Vascular disease of the kidney can involve either the arteries or the veins. Thrombosis of the renal veins is a complication of the nephrotic syndrome. Acute renal vein thrombosis presents as sudden flank pain and hematuria, whereas chronic renal vein thrombosis is often asymptomatic. In some cases, renal vein thrombi may be the source of a pulmonary embolus.

Renal artery diseases may present in one of several ways. Renal artery stenosis can cause secondary hypertension or give rise to ischemic nephropathy and a slow progressive loss of renal function.[13] Administration of an angiotensin-converting enzyme inhibitor or an angiotensin receptor blocker can precipitate acute renal failure in the setting of bilateral renal artery stenosis or an obstructed solitary kidney. A comprehensive review of renal vascular disease is provided elsewhere [*see 10:VII Vascular Diseases of the Kidney*].

Tubulointerstitial Renal Disease

Tubulointerstitial renal disease refers to diseases that principally involve the tubules or interstitium, with relative sparing of the glomeruli. These diseases can be primary to the kidney or can be part of a systemic disorder. Tubulointerstitial disease can be further subdivided into acute or chronic, on the basis of the clinical presentation and laboratory findings.

Acute interstitial renal disease most commonly occurs in patients taking a sulfa or β-lactam antibiotic, but this disease can occur with virtually any drug. Clinically, this condition presents as a systemic allergic disorder with fever and rash, beginning 7 to 10 days after the drug is started. Laboratory examination reveals an increased serum creatinine concentration and peripheral eosinophilia. The urinalysis demonstrates white blood cells (WBCs), WBC casts, and eosinophiluria [*see Figure 2*]. Proteinuria tends to be low grade, with no more than 1 to 2 g of protein in a 24-hour specimen. Histologically, there is a marked inflammatory cell infiltrate in the interstitium, with evidence of tubulitis.

Chronic tubulointerstitial renal disease is much more subtle in presentation, often manifested only by a slowly rising creatinine concentration over months to years. The characteristic histologic finding in chronic tubulointerstitial nephritis is interstitial fibrosis. Laboratory tests are characterized by evidence of tubular dysfunction. Characteristic findings include an inability to concentrate the urine, hyperkalemia out of proportion to the reduction in GFR, early development of a hyperchloremic normal-gap metabolic acidosis, and a tendency toward salt wasting. Sterile pyuria is often present on urinalysis. Anemia is common and tends to be more severe than in glomerular disease, because erythropoietin is primarily synthesized by cells in the interstitium. Tubulointerstitial renal disease is discussed in detail elsewhere [*see 10:VIII Tubulointerstitial Diseases*].

GLOMERULAR DISEASE

Diseases of the glomerulus can present as one of five clinical syndromes: chronic glomerulonephritis, nephrotic syndrome, acute glomerulonephritis, rapidly progressive glomerulonephritis, or asymptomatic hematuria or proteinuria [*see Figure 1*]. The clinical characteristics and the urinalysis results are useful for determining which syndrome is present in a patient with suspected glomerular disease [*see Table 3*].

Diseases that cause active inflammation in the glomerulus are typically associated with a constellation of clinical features referred to as the nephritic syndrome. This syndrome is characterized by signs and symptoms suggestive of primary salt retention by the kidney, resulting in volume overload and circulatory congestion. Manifestations include hypertension, congestive heart failure, and peripheral edema. The urinalysis result also reflects active glomerular inflammation and is described as nephritic in nature. A nephritic urinary sediment is characterized by a variable number of red blood cells (RBCs), WBCs, and RBC casts [*see Figure 2*]. The finding of RBC casts is pathognomonic of active glomerulonephritis. Proteinuria is usually modest, ranging from 2 to 6 g in a 24-hour collection, and the GFR is often severely reduced.

The nephrotic syndrome results from diseases that produce little in the way of active glomerular inflammation but, instead, cause massive leakage of protein across the glomerular basement membrane. The development of hypoalbuminemia and subsequent decline in oncotic pressure cause intravascular fluid to translocate into the extravascular compartment, giving rise to a contracted EABV. In turn, the decline in EABV results in secondary renal salt retention. Unlike the nephritic syndrome, hypertension is less prominent and findings of circulatory congestion are absent. Typically, the GFR is only mildly reduced. The urinalysis result also reflects the lack of inflammation in the glomerulus. Characteristic findings include free-fat droplets,

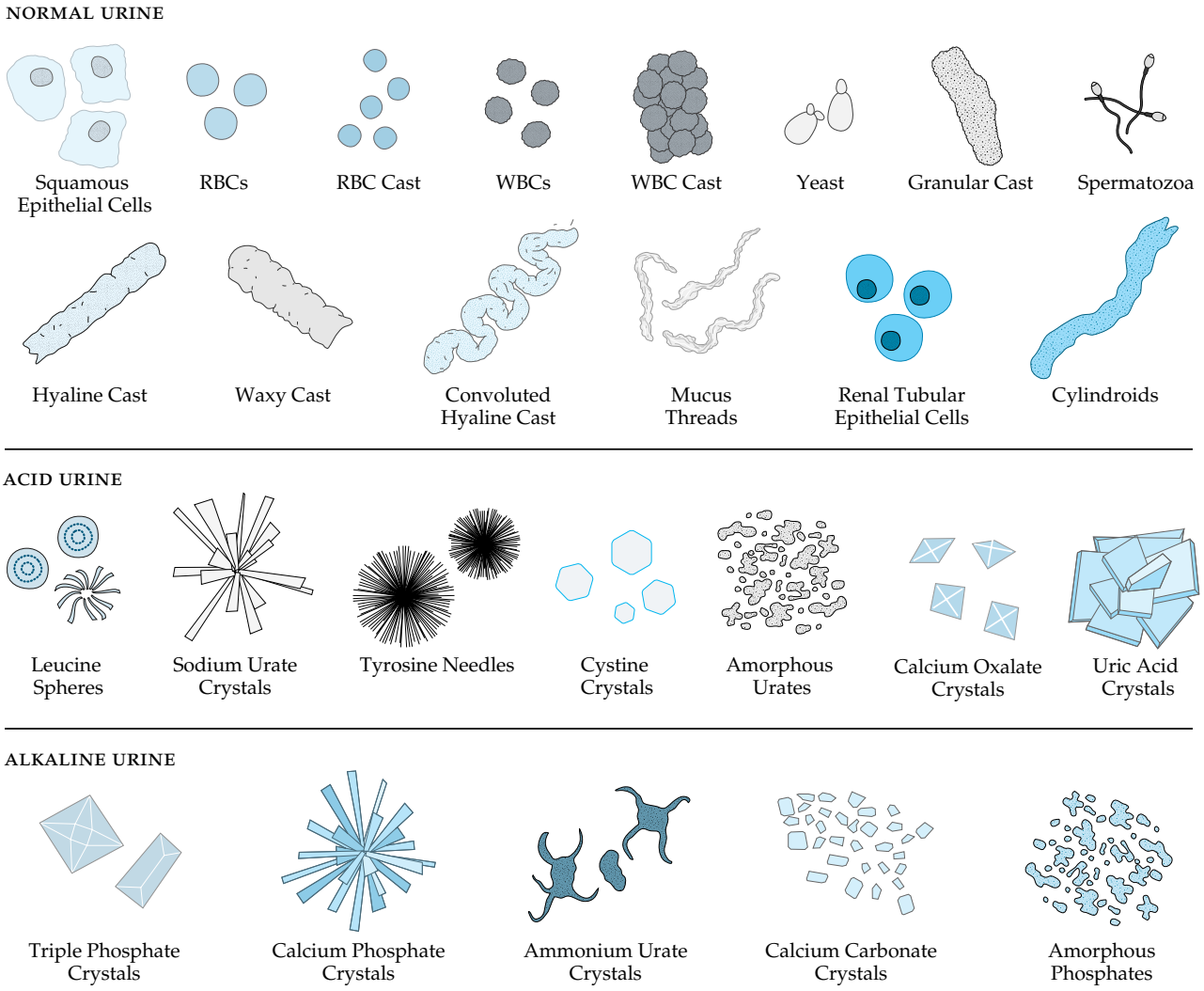


Figure 2 **Components of urinary sediment.**

oval fat bodies, fatty casts, and only minimal amounts of cellularity. These findings, along with large amounts of protein in the urine, constitute a nephrotic urinary sediment.

Chronic Glomerulonephritis

Chronic glomerulonephritis describes a patient with evidence of chronic renal failure who shows characteristics of glomerular disease. The urinalysis demonstrates a few RBCs and WBCs but is mostly nonspecific. A variable amount of proteinuria is present. Kidney size is typically small, reflecting the presence of advanced fibrosis and glomerulosclerosis. Patients with this syndrome presumably have a glomerular disease that has progressed to end stage, leaving the kidneys irreversibly damaged.

Nephrotic Syndrome

The nephrotic syndrome describes a patient who presents with a nephrotic clinical picture and a nephrotic urinary sediment. In addition to edema, hypoalbuminemia, and large amounts of protein in the urine, a number of other findings are typical of the nephrotic syndrome. These include hyperlipidemia, hypercoagulability, and a predisposition to infection with encapsulated gram-positive organisms because of hypogammaglobulinemia. As with other clinical syndromes, the underlying cause of the nephrotic syndrome can be a primary renal disease or secondary kidney involvement as part of a systemic disorder. Because the kidney can respond to injury in only a few ways, systemic disorders often cause the same histologic picture as primary renal disease. For example, membranous glomerulopathy is a primary renal disease, but this same histologic pattern can develop as a result of systemic lupus erythematosus or exposure to gold in the treatment of rheumatoid arthritis.

The approach to the patient with nephrotic syndrome is directed toward distinguishing primary from secondary glomerular diseases. Routine clinical evaluation may be sufficient to make this distinction, but in many cases, additional laboratory tests are required [*see 10:V Glomerular Diseases*].

Acute Glomerulonephritis

Acute glomerulonephritis describes a patient who presents with the abrupt onset of a nephritic clinical syndrome accompanied by a nephritic urinary sediment. There is evidence of circulatory congestion, manifested by hypertension and occasionally

by pulmonary congestion. Examination of the urine demonstrates RBCs, WBCs, and RBC casts. The GFR is relatively stable, in contrast to the rapid fall in GFR seen with the syndrome of rapidly progressive glomerulonephritis. Acute glomerulonephritis is often associated with a preceding infection and can resolve spontaneously. The workup is designed to determine whether the underlying disorder is a primary renal disease or a systemic disorder affecting the kidney [*see 10:V Glomerular Diseases*].

Rapidly Progressive Glomerulonephritis

Rapidly progressive glomerulonephritis should be considered in patients who present with a nephritic clinical picture and who have a nephritic urinary sediment. This condition is distinguished from acute glomerulonephritis by the rapid loss of renal function, which is defined as a rise in the serum creatinine concentration of more than 2 mg/dl over a 3-month period. This syndrome has a much less consistent temporal relationship with infection, and there is little tendency for spontaneous recovery.

This syndrome needs to be recognized early so that renal biopsy can be done and therapy instituted immediately, if indicated. With patients in whom renal biopsy shows a crescentic glomerulonephritis, immunofluorescent studies provide a useful classification of the diseases that most commonly give rise to this clinical syndrome.

Asymptomatic Hematuria or Proteinuria

Asymptomatic hematuria or proteinuria describes a patient who has hematuria or proteinuria, or both, but who is otherwise without clinical symptoms. The approach to these patients is described below.

Asymptomatic Urinary Sediment Abnormalities

PROTEINURIA

Proteinuria is relatively common, with 5% to 10% of urine dipstick tests showing positive results in adults during screening examinations.[14] Because many disease states and mechanisms can underlie proteinuria, it is important to review renal handling of protein under normal conditions to formulate a differential diagnosis on the basis of structural-functional abnormalities.

Transient and Orthostatic Proteinuria

Two types of asymptomatic, isolated proteinuria warrant consideration: transient or functional proteinuria and orthostatic proteinuria. The underlying mechanism for transient proteinuria is not yet understood. Common factors such as fever, heart failure, seizure activity, pancreatitis, and exercise are frequently associated with transient proteinuria.[15] The proteinuria resolves with treatment of the underlying cause. Orthostatic proteinuria occurs in up to 5% of adolescents but is uncommon in older age groups.[16] Typically, such patients demonstrate proteinuria while in the upright position, which then resolves when the recumbent position is assumed. Long-term follow-up studies demonstrate no deterioration in renal function over time. Generally, orthostatic proteinuria resolves spontaneously.[16]

Persistent Proteinuria

Persistent proteinuria generally connotes significant renal pathophysiology. Renal protein excretion is determined primarily by three factors: the characteristics of the protein (its size and charge, as well as its concentration in the blood); the permeability of the glomerular capillary wall; and tubular reabsorption of filtered protein. Smaller and more positively charged proteins have greater rates of filtration into the urinary space than do larger and more negatively charged proteins. Normally, 1 to 3 g of protein (approximately two thirds globulin and one third albumin) are filtered daily. Over 90% of the filtered load is reabsorbed by the tubule, so that 150 mg of protein is maximally excreted each day.

There are three categories of persistent proteinuria: overflow, glomerular, and tubulointerstitial. In overflow proteinuria, proteins that are normally filtered are produced in such high quantities that they overwhelm the reabsorptive capacity of the tubule and are spilled into the urine. The glomerular filtration barrier is intact and functions normally. Paraprotein disorders such as multiple myeloma, in which excess immunoglobulin light chains are produced, result in Bence-Jones proteins in the urine. Lysozymuria, which occurs in patients with myelogenous leukemia, and myoglobinuria caused by rhabdomyolysis are two additional examples of overflow proteinuria. Standard urine dipsticks will react positively only with albumin (and other negatively charged proteins) and thus will be negative in dysproteinemic states when paraproteins (positively charged proteins) appear in the urine. When such disorders are suspected, sulfosalicylic acid or heat acetic precipitation methods should be used to detect nonalbumin species.

In glomerular proteinuria, the glomerular filtration barrier is structurally abnormal (as in diabetic nephropathy, amyloidosis, and membranous lesions) or loses its fixed negative charge (as in minimal change disease). Both the filtration barrier of the glomerular basement membrane and the reabsorptive capacity of the tubule are exceeded. As a result, proteinuria is often heavy, with albumin being the most prevalent protein. When urinary protein excretion exceeds 3.5 g in a 24-hour collection, nephrotic-range proteinuria is said to be present.

Tubulointerstitial proteinuria occurs when tubular dysfunction and injury preclude normal reabsorptive pathways for filtered proteins, primarily low-molecular-weight globulins. Additionally, damaged tubules may secrete excess proteinaceous material into the urinary space (Tamm-Horsfall mucoprotein). Typically, protein excretion amounts to 1 to 2 g/24 hr but can exceed 3 g/24 hr as tubulointerstitial disease progresses, causing secondary glomerular injury. Acute tubulointerstitial nephritis, tubular toxins such as aminoglycosides, and Fanconi syndrome can underlie this type of proteinuria. So-called indeterminate proteinuria (0.3 to 3.0 g/24 hr, primarily albumin) can be seen in diseases such as hypertensive nephrosclerosis, obstructive uropathy, and chronic tubulointerstitial nephritis.

Diagnosis of Proteinuria

Quantifying the amount of protein excreted by a patient is conventionally done with a 24-hour urine collection. An efficient and accurate alternative is measurement of the protein-to-creatinine ratio in a spot urine specimen.[17] The numerical value of this ratio represents the amount of urinary protein excreted in grams per day. For example, if a patient has 150 mg/dl of protein and 50 mg/dl of creatinine in a spot urine specimen, a total protein excretion rate of 3 g/24 hr would be expected. This simple procedure is accurate and independent of the amount of proteinuria, degree of renal insufficiency, or underlying disease state.

The urine dipstick test for protein can give false positive results when various pharmacologic agents and gross hematuria

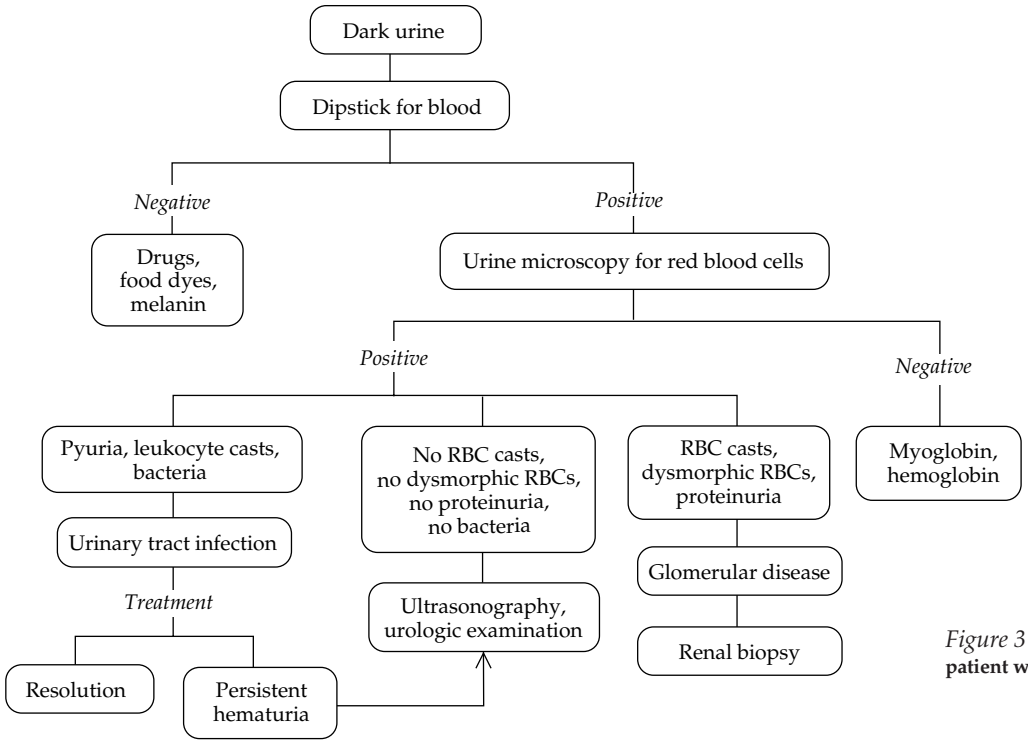


Figure 3 **Approach to the patient with hematuria.**

are present or when urine pH is high (> 7.0). False negative results highlight the fact that dipstick testing primarily detects albumin but misses positively charged proteins such as paraproteins (e.g., Bence-Jones immunoglobulin light chains).

HEMATURIA

Hematuria is defined as an RBC count greater than 1 RBC per high-power field for men and a count greater than 3 RBCs per high-power field for women. Alternatively, urinary tract blood loss that is greater than 8,000 to 15,000 RBCs/ml of urine is considered abnormal. Hematuria can be microscopic or gross, but the degree of hematuria is generally of little diagnostic or prognostic value. As little as 1 ml of blood per liter of urine can cause a visible color change. Not all reddish-brown urine is the result of hematuria. Hemoglobinuria and myoglobinuria can cause similar discoloration, but only the supernatant of a spun urine sample is discolored; RBCs are absent on urine microscopy in these conditions. Drugs such as rifampin, sulfasalazine, and the phenothiazines can also cause red discoloration of urine [*see Figure 3*].

Hematuria is a relatively common clinical finding. Froom and colleagues reported a 40% incidence on routine urinalysis in 1,000 men between 18 and 33 years of age; 13% of postmenopausal women were reported to have hematuria on routine screening.[18,19] Crystalluria caused by hypercalciuria or hyperuricosuria may often account for so-called idiopathic hematuria in asymptomatic persons. Crystal aggregates or microcalculi are thought to injure tubular epithelium, resulting in hematuria. Andres and colleagues found that more than 30% of children with isolated hematuria had hypercalciuria and a positive family history of nephrolithiasis.[20] If left untreated, these children have an increased risk of kidney stones later in life. Similar patterns have been observed in adults; correction of hypercalciuria or hyperuricosuria leads to resolution of hematuria in most cases.

Differential Diagnosis

The differential diagnosis of hematuria is extensive but can be divided into two broad categories: renal, or so-called upper tract, sources; and nonrenal, or so-called lower tract, sources.[21] In most cases, hematuria is nonrenal in origin. Mariani and colleagues determined the etiology for asymptomatic, isolated hematuria in adults in 1,000 consecutive cases.[22] Of 883 cases in which causes were identified, only 13 (1.5%) were glomerular in origin, whereas urethritis, prostate disease, bladder carcinoma, and cystitis accounted for more than 75% of cases.

RBCs originating from the glomerulus have a distinctive dysmorphic appearance when viewed on phase-contrast microscopy, whereas hematuria originating from the lower urinary tract (nonglomerular hematuria) is characterized by uniform, intact RBCs similar to those seen on a peripheral smear. Diagnostically, the use of RBC dysmorphism is somewhat limited by the need for phase-contrast microscopy and a trained urine microscopist.

If glomerulonephritis appears to be an unlikely diagnosis on the basis of clinical findings, serial urinalyses should be done to determine whether hematuria is transient or persistent. Transient hematuria, particularly in young persons, is quite common and rarely indicative of significant pathology.[23] When present in patients older than 50 years, however, transient hematuria warrants a comprehensive evaluation to rule out malignancy. Persistent hematuria in patients of any age mandates a thorough diagnostic workup.

Historical clues can provide key diagnostic information in the evaluation of patients with hematuria. For example, inherited forms of renal disease are likely if a patient has a family history of autosomal dominant polycystic kidney disease, deafness (Alport disease), or sickle cell trait or disease. Flank pain radiating to the groin may indicate urinary tract obstruction caused by stones or blood clots. Dysuria, urinary frequency, and urgency are consistent with an infectious source. A recent upper respiratory tract infection may have triggered IgA nephropathy or postinfectious glomerulonephritis. Hematuria may develop spontaneously in patients receiving anticoagulants. Finally, in-

tense exercise, such as long-distance running, can precipitate hematuria, as can prostate enlargement.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Shemesh O, Golbetz H, Kriss JP, et al: Limitations of creatinine as a filtration marker in glomerulopathic patients. Kidney Int 28:830, 1985

2. Jones CA, McQuillan GM, Kusek JW, et al: Serum creatinine levels in the US population: third National Health and Nutrition Examination Survey. Am J Kidney Dis 35:178, 2000

3. Rahn KH, Heidenreich S, Bruckner D: How to assess glomerular function and damage in humans. J Hypertens 17:309, 1999

4. Levey AS, Berg RL, Gasman JJ, et al: Creatinine filtration, secretion, and excretion during progressive renal disease. Kidney Int 36(suppl 27):S73, 1989

5. Fuller NJ, Elia M: Factors influencing the production of creatinine: implications for the determination and interpretation of urinary creatinine and creatinine in man. Clin Chim Acta 175:199, 1988

6. Levey AS, Bosch JP, Lewis JB, et al: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461, 1999

7. Gaspari F, Perico N, Ruggenenti P, et al: Plasma clearance of nonradioactive iohexol as a measure of glomerular filtration rate. J Am Soc Nephrol 6:257, 1995

8. Brandstrom E, Grzegorczyk A, Jacobsson L, et al: GFR measurement with iohexol and ^{51}Cr-EDTA: a comparison of the two favoured GFR markers in Europe. Nephrol Dial Transplant 13:1176, 1998

9. Lundqvist S, Holmberg G, Jakobsson G, et al: Assessment of possible nephrotoxicity from iohexol in patients with normal and impaired renal function. Acta Radiol 39:362, 1998

10. Swan SK, Halstenson CE, Kasiske BL, et al: Determination of residual renal function with iohexol clearance in hemodialysis patients. Kidney Int 49:232, 1996

11. Francois M, Tostivint I, Mercadal L, et al: MR imaging features of acute bilateral renal cortical necrosis. Am J Kidney Dis 35:745, 2000

12. Esson ML, Schrier RW: Diagnosis and treatment of acute tubular necrosis. Ann Intern Med 137:744, 2002

13. Safian RD, Textor S: Renal artery stenosis. N Engl J Med 344:431, 2001

14. Larson TS: Evaluation of proteinuria. Mayo Clin Proc 70:308, 1994

15. Wingo CS, Clapp WL: Proteinuria: potential causes and approach to evaluation. Am J Med Sci 320:188, 2000

16. Springberg PD, Garrett LE Jr, Thompson AL Jr, et al: Fixed and reproducible orthostatic proteinuria: results of a 20-year follow-up study. Ann Intern Med 97:516, 1982

17. Levey A, Coresh J, Balk E, et al: National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Ann Intern Med 139:137, 2003.

18. Froom P, Ribak J, Benbassat J: Significance of microhaematuria in young adults. Br Med J Clin Res Ed 288:20, 1984

19. Mohr DN, Offord KP, Owen RA, et al: Asymptomatic microhematuria and urologic disease: a population-based study. JAMA 256:224, 1986

20. Andres A, Praga M, Bello I, et al: Hematuria due to hypercalciuria and hyperuricosuria in adult patients. Kidney Int 36:96, 1989

21. Cohen R, Brown R: Microscopic hematuria. N Engl J Med 348:30, 2003

22. Mariani AJ, Mariani MC, Macchioni C, et al: The significance of adult hematuria: 1,000 hematuria evaluations including a risk-benefit and cost-effectiveness analysis. J Urol 141:350, 1989

23. Murakami S, Igarashi T, Hara S, et al: Strategies for asymptomatic microscopic hematuria: a prospective study of 1,034 patients. J Urol 144:99, 1990

IV MANAGEMENT OF CHRONIC KIDNEY DISEASE

Biff F. Palmer, m.d., f.a.c.p.
Michael K. Hise, m.d.

End-stage renal disease (ESRD) is an important source of long-term morbidity and mortality. Efforts to reduce or prevent the development of this disorder are of paramount importance. One of the problems that nephrologists often encounter is that patients are referred late in the course of disease, when treatments are largely ineffective. Thus, it is imperative to identify patients early and to treat them aggressively. Recently published clinical practice guidelines are aimed at accomplishing this goal. This chapter reviews the management of patients with chronic kidney disease (CKD), giving particular emphasis to the following: (1) methods that can be clinically used to accurately stage and monitor the level of renal function; (2) identification of common comorbidites of CKD; (3) the pathophysiology of CKD; (4) therapies designed to slow the progression of CKD; (5) management of common complications of CKD (e.g., anemia, fluid and electrolyte disturbances, and renal osteodystrophy); and (6) timely referral to a nephrologist so as to ensure a smooth transition to renal replacement therapy when indicated. Information on the causes and manifestations of CKD are covered elsewhere [*see 10:X Chronic Renal Failure and Dialysis*].

Overview of Clinical Practice Guidelines for Chronic Kidney Disease

It is estimated that 20 million persons in the United States population have chronic kidney disease. The typical course of patients with CKD is progression to kidney failure and development of cardiovascular disease. Increasing evidence suggests that treatment strategies initiated at earlier stages of CKD are effective in slowing the progression toward kidney failure. In addition, addressing cardiovascular risk factors at earlier stages of CKD should be effective in reducing cardiovascular events and mortality, both before and after the onset of kidney failure.

To date, implementation of measures designed to detect and treat CKD at an early stage have been hampered, in part by lack of agreement on a definition and classification of stages in the progression of chronic kidney disease. In addition, there has not been a uniform application of tests for detection and evaluation of such patients. In February 2002, the Kidney Disease Outcomes Quality Initiative (K/DOQI) clinical practice guidelines were published in an attempt to address the growing burden of CKD in the United States.[1] These guidelines emphasize the need for prevention, early diagnosis, and aggressive treatment of CKD.

definition and staging of chronic kidney disease

CKD is defined as either kidney damage or a glomerular filtration rate (GFR) of less than 60 ml/min/1.73 m^2 for more than 3 months. Kidney damage is defined as pathologic abnormalities or markers of damage, including abnormal findings in blood or urine tests or imaging studies. In patients with CKD, the stage is defined by the level of the GFR. The level of GFR is considered to be the best measure of overall kidney function in health and disease.

Inulin or iothalamate clearance is regarded as the gold standard for measuring GFR. However, these methods are cumbersome and expensive and are not available to many practicing physicians. By contrast, the GFR can be easily estimated by formulas that take into account the patient's age, gender, ethnicity, and serum creatinine concentration [*see Table 1*]. The Modification of Diet in Renal Disease (MDRD) and Cockcroft-Gault equations are two such formulas that provide useful estimates of GFR in adults. Of these, the MDRD equation is more accurate and precise for persons with a GFR of less than 90 ml/min/1.73 m^2. The equation was based on a large database encompassing persons with various diseases and including both European Americans and African Americans. The K/DOQI clinical practice guidelines state that clinicians should not use serum creatinine concentration as the sole means to assess the level of kidney function. In addition, measurement of the 24-hour creatinine clearance level to assess GFR is no more accurate than estimating GFR from the MDRD equation. Certain clinical conditions, however, may necessitate use of creatinine clearance as a way to better assess the GFR [*see Table 2*].

Evaluation of all patients with CKD should include testing for proteinuria. Persistent proteinuria is usually a marker of kidney damage, and in the setting of a normal or increased GFR, it signifies the presence of stage 1 CKD. Proteinuria is also a strong and independent predictor of increased risk for cardiovascular morbidity and mortality, particularly in high-risk groups such as diabetics, hypertensives, the elderly, and those with CKD.

The preferred method for measuring proteinuria is the ratio of protein or albumin to creatinine (mg/g) in an untimed urine specimen. Timed urine collections are inconvenient to obtain and may be associated with errors; therefore, they are usually not necessary for measuring proteinuria. In addition to being utilized for the detection of CKD, measurement of the protein- or albumin-to-creatinine ratio is a useful way to monitor proteinuria in patients with known CKD. Changes in the degree of proteinuria are directly associated with the risk of loss of kidney function.

evaluation and treatment

Once a patient is identified as having CKD, a clinical action plan is implemented on the basis of the patient's stage of disease [*see Table 3*]. Although specific therapies vary according to the underlying cause of kidney disease, many aspects of treatment are common to all types of CKD. Treatment of comorbid conditions, interventions to slow the progression of kidney disease, and measures to reduce cardiovascular disease should begin during stages 1 and 2. Particular attention needs to be given in

Table 1 Equations to Estimate Glomerular Filtration Rate

Abbreviated MDRD (Modification of Diet in Renal Disease Study) equation:

GFR (ml/min/1.73 m^2) = $186 \times (Scr)^{-1.154} \times (age)^{-0.203} \times 0.742$ (if female) $\times$ 1.210 (if African American)

Cockcroft-Gault equation:

Creatinine clearance (ml/min) = [140 – age $\times$ weight (kg) / 72 $\times$ serum creatinine] $\times$ 0.85 (if female)

Table 2 Clinical Conditions That May Require Measurement of GFR Using Clearance Methods

- Extremes of age and body size
- Severe malnutrition or obesity
- Diseases of skeletal muscle
- Paraplegia or quadriplegia
- Vegetarian diet
- Rapidly changing kidney function
- Calculation that potentially toxic drugs may be excreted by the kidney

stages 1 and 2 to aggressively control blood pressure. Evaluation and treatment of anemia, malnutrition, bone disease, neuropathy, and decreased quality of life should be undertaken in patients with stage 3 CKD. Preparations for renal replacement therapy should begin as patients enter stage 4. Kidney failure (stage 5) is defined either as a GFR of less than 15 ml/min/1.73 m^2 or as complications of decreased GFR severe enough to increase the risk of mortality and morbidity unless initiation of kidney replacement is undertaken. Some patients who have symptoms of uremia may require kidney-replacement therapy although their GFR is greater than 15 ml/min/1.73 m^2.

The K/DOQI guidelines recommend that all routine health encounters include an assessment to determine whether an individual is at increased risk for developing kidney disease. Persons deemed to be at increased risk should undergo testing for proteinuria. Detection of proteinuria in patients at increased risk would allow for early implementation of measures designed to reduce proteinuria, slow the progression of kidney disease, and reduce the risk of cardiovascular mortality and morbidity. Numerous clinical factors are identified with increased risk of developing kidney disease [*see Table 4*]. Some of these factors are discussed in more detail below.

Comorbidites of Chronic Kidney Disease

In the United States, millions of patients have an increased serum creatinine concentration.[2,3] In many of these patients, the renal insufficiency is advanced. Diabetes and hypertension are the underlying causes of renal disease in the majority of these patients. The remaining cases can be accounted for by a variety of primary glomerular diseases, cystic renal diseases, obstruction uropathy, and renal diseases related to infections.

Regardless of the underlying cause, it should be appreciated that CKD carries with it the burden of other comorbidities that directly influence patient outcome. In addition, abnormal renal function confers its own inherent risk. In the Hypertension Detection and Follow-up Program, cardiovascular disease was strongly associated with a serum creatinine level of 1.7 mg/dl or more, and this risk was independent of other factors.[4] In the 10,768 participants in this program, the risk of death increased progressively with increasing creatinine concentration. There was an almost fivefold increase in the 8-year mortality between patients in the lowest stratum of creatinine concentration and those in the highest. Similar findings were reported in the Heart Outcomes and Prevention Evaluation.[5] In this study, the incidence of cardiovascular death, myocardial infarction, and stroke increased with each quartile of serum creatinine concentration. This association between renal function and mortality is particularly strong in the elderly.[6]

STROKE

The prospective British Regional Heart Study, comprising 7,690 men 40 to 59 years of age, demonstrated that the risk of stroke was significantly increased when the creatinine concentration was above the 90th percentile—a serum creatinine concentration of 116 mmol/L—whereas all-cause mortality and cardiovascular mortality were significantly increased when the serum creatinine concentration was only above the 97.5 percentile.[7] In elderly persons admitted to an acute care hospital with stroke, the serum creatinine concentration was a highly significant predictor of survival.[8]

CARDIAC DISEASE

The relationship between renal function and myocardial infarction (MI) was studied in 417 patients.[9] These individuals did not have diabetes, and their baseline creatinine values ranged from 0.7 to 1.9 mg/dl. Each 0.1 mg/dl increase in baseline creatinine concentration was associated with a 36% increase in the relative risk of subsequent overall mortality and a 47% increase in the relative risk of subsequent mortality from coronary artery dis-

Table 3 Stages of CKD and Clinical Action Plans

Stage	*Description*	*GFR (ml/min/1.73 m^2)*	*Action*
0	At increased risk	≥ 90 (with CKD risk factors)	Estimate GFR and determine presence or absence of proteinuria; institute steps to reduce CKD risk
1	Kidney damage with normal or ↑ GFR	≥ 90	Diagnosis and treatment, treatment of comorbid conditions, slow progression, CKD risk reduction
2	Kidney damage with mild ↓ GFR	60–89	Estimate progression
3	Moderate ↓ GFR	30–59	Evaluate and treat complications
4	Severe ↓ GFR	15–29	Prepare for kidney replacement therapy
5	Kidney failure	< 15	Kidney replacement (if uremia present)

CKD—chronic kidney disease ↑ GFR—increased glomerular filtration rate ↓ GFR—decreased glomerular filtration rate

Table 4 Risk Factors for CKD Requiring Testing for Proteinuria and Determination of an Estimated GFR

Clinical factors	Diabetes Hypertension Autoimmune diseases Urinary tract infection Nephrolithiasis Urinary tract obstruction Cancer Family history of CKD Reduction in kidney mass Exposure to nephrotoxins Low birth weight
Sociodemographic factors	Elderly Ethnic minority status (African American, Native American, Hispanic) Exposure to chemical or environmental hazard Low income or education

CKD—chronic kidney disease GFR—glomerular filtration rate

ease. A similar trend between mortality and serum creatinine levels was observed in another study of post-MI patients.[10] Nearly 75% of individuals starting dialysis have left ventricular hypertrophy (LVH), and there is an inverse relationship between renal function and the incidence of LVH in patients before undergoing dialysis.[11] There is also a high prevalence of congestive heart failure. The incidence of coronary artery disease is approximately 40% in patients receiving dialysis, and there also appears to be an increased incidence of peripheral vascular disease in patients with abnormal renal function.

Risk Factors for Progression of Kidney Disease

Most patients with established CKD continue to experience a decrease in renal function over time.[3] This section reviews various factors known to contribute to loss of renal function. In addition, therapies that can slow the progression of CKD are discussed.

DIABETES MELLITUS

Animal models have provided a clear picture of the mechanisms underlying diabetic glomerular destruction. What is not known is why only approximately 40% of patients with type 1 (insulin-dependent) diabetes mellitus are at risk. Obviously, genetics plays an important role, but other potentially manageable factors are also important. Hyperglycemia through nonenzymatic glycosylation, production of oxidants, or stimulation of the production of diacylglycerol may activate protein kinase C (PKC). It has been shown that levels of PKC are increased in the diabetic kidney.[12,13] PKC causes an increase in production of transforming growth factor–β (TGF-β), which is a potent stimulator of collagen formation and basement membrane thickening. A variety of other alterations, including permeability changes, altered fibrinolysis, changes in gene expression, and vascular contractility, may also be stimulated by PKC.[14] Interestingly, inhibition of PKC in the kidney and retina reduces diabetes-related abnormalities.[15] Micropuncture studies in diabetic animals showed that glomerular intracapillary pressures were increased.[16] These changes may lead to structural alterations that are important in the pathogenesis of diabetes-related complications.

Type 1 Diabetes Mellitus

The most important study regarding the treatment of type 1 diabetes is the Diabetes Control and Complications Trial (DCCT).[17] This study examined 1,441 patients with type 1 diabetes. Of these patients, 726 had no retinopathy at baseline; 715 had mild retinopathy. After a mean follow-up period of nearly $6\frac{1}{2}$ years, intensive insulin therapy reduced the occurrence of microalbuminuria by 39% and the occurrence of macroalbuminuria (> 300 mg/24 hr) by 54% in the combined groups. Improvement in nephropathy was mirrored by a slowing in the progression of retinopathy and neuropathy. A follow-up to this study indicated that the effects of intensive treatment are long lasting.[18]

Type 2 Diabetes Mellitus

Although type 1 diabetes has received a great deal of attention, the incidence of type 2 (non–insulin-dependent) diabetes continues to rise at a rapid pace, particularly in members of minority groups, including African Americans and Native Americans. In patients with nephropathy and type 2 diabetes, familial clustering has been observed. Polymorphism of the angiotensin-converting enzyme (ACE) gene may influence the rate of renal disease progression, and several genes have been linked to renal disease in specific populations.[19] Although the renal lesions in type 2 diabetes are similar to those in type 1, a greater proportion of patients with type 2 diabetes have nonspecific lesions, including ischemic changes.[20] Control of hyperglycemia in patients with type 2 diabetes is a controversial issue. Some studies have failed to show an association between good glucose control and the maintenance of functional nephron mass. On the other hand, Hsu and colleagues demonstrated that hemoglobin A_{1c} (HbA_{1c}) levels were independent predictors of a decline in renal function.[21]

Treatment of Diabetes Mellitus in Patients with CKD

Treatment of type 1 diabetes mellitus Treatment of type 1 diabetes in patients with CKD requires the use of insulin therapy [*see 9:VI Diabetes Mellitus*]. Management should include a carefully structured diet and exercise. Care must be taken to avoid malnutrition with advancing renal insufficiency because anorexia is part of the uremic syndrome. Continuing podiatric care is an important part of the exercise program. The use of insulin requires special care in patients with renal insufficiency because of the prolonged half-life of insulin. Furthermore, CRF patients with type 1 diabetes frequently have long-standing disease, and there may be few warning signs of hypoglycemia. Gastroparesis with erratic food absorption further increases the risk of hypoglycemia. Home glucose monitoring is an invaluable element in the care of these patients, as is careful follow-up by providers skilled in diabetes management. Symptomatic hypoglycemia is much more common with tight glucose control, and it is difficult to maintain proper glucose levels within the framework of traditional practice. This is clearly evident in the follow-up of the patients in the DCCT trial.[17,18] In the patients assigned to receive tight control of glucose levels, HBA_{1c} increased from 7.2% to 7.9% in the 4-year interval after the conclusion of the original study. Transplantation is gaining increasing favor as a method of treating patients with type 1 diabetes, especially those with advancing renal disease, because of the use of simultaneous kidney and pancreas transplantation.

Treatment of type 2 diabetes mellitus Treatment of patients with type 2 diabetes who have renal insufficiency should

also begin with a carefully designed diet and exercise program. Weight loss plays an important role in improving insulin resistance and should be employed as well. It should be kept in mind that there are a number of drugs that are associated with worsening of insulin resistance [*see Table 2 and 9:VI Diabetes Mellitus*]. One should consider altering the drugs in patients with type 2 diabetes whose glucose levels are difficult to control. Early in the course of disease, when the patient has some pancreatic insulin reserves, hypoglycemic agents can be considered. Their use, however, is complicated in patients with renal disease. Metformin and other biguanides may cause lactic acidosis and should not be used in patients with even mild degrees of renal dysfunction. Troglitazone may be used in this situation because of its hepatic excretion; however, there is a high risk of acute hepatitis with this agent, necessitating monthly monitoring of liver function. One should avoid the older sulfonylureas, such as chlorpropamide, because of their reduced renal excretion. Glipizide is inactivated by the liver and may be a better choice. Acarbose, which reduces carbohydrate absorption, should be avoided. Unfortunately, the majority of patients with type 2 diabetes require insulin therapy. These individuals frequently gain weight with therapy. Other complications include hypoglycemia and hyperinsulinemia. The latter is of concern because of its association with cardiovascular complications. Patients whose glucose levels cannot be controlled within an acceptable range should be referred to specialized clinics. Investigations are needed to understand and reverse the trend toward increasing obesity in the United States, as it is a contributing factor in the development of type 2 diabetes, particularly in susceptible minority populations.

HYPERTENSION

The mechanisms of renal injury in hypertension were presumed to be ischemic in nature, but these mechanisms remained largely unexplained until the work of Brenner and colleagues provided detailed studies of the glomerular alterations in hypertensive rats with reduced renal mass.[22,23] These studies demonstrated that afferent arteriolar vasodilatation accompanied by similar but relatively less efferent arteriolar changes increased transcapillary filtration pressures. Although these changes were associated with improving glomerular filtration rates in the short term, they were eventually seen to be detrimental, leading to progressive glomerular sclerosis. In a direct test of this hypothesis in patients, a collaborative study demonstrated that the ACE inhibitor captopril was effective in reducing the development of nephropathy in patients with type 1 diabetes and in retarding the progression of established renal dysfunction over and above its effects in lowering blood pressure.[24] A similar benefit was recently shown with use of angiotensin receptor blockers in patients with type 2 diabetes and established nephropathy.[25,26]

The United Kingdom Prospective Diabetes Study demonstrated that moderate reductions in blood pressure were effective in reducing renal abnormalities; in this study, there were no apparent differences in effectiveness between the beta blocker atenolol and captopril. However, the target blood pressure values may have been too high to substantiate the beneficial effects of the ACE inhibitor.[27,28] The Reduction of Endpoints in NIDDM with Angiotensin II Antagonist Losartan (RENAAL) study comprised 1,513 participants who were followed for an average of 3.4 years. It demonstrated that an antihypertensive program containing the drug, compared with placebo, significantly decreased the risk of the doubling of serum creatinine, ESRD, and death.[26] A repeat analysis of these data was performed to look specifically at systolic blood pressure and pulse pressure; every 10 mm Hg rise in baseline systolic blood pressure increased the risk for ESRD or death by 6.7%; a similar rise in diastolic blood pressure decreased the risk by 10.9%. Pulse pressure was a significant risk factor that was dramatically improved by losartan.[29]

Hypertension in Diabetic Patients

Although there is a great deal of focus on negative outcomes in patients with CRF, the Heart Outcomes and Prevention Evaluation (HOPE) trial provides some reasons for optimism. This study followed 3,577 individuals with diabetes for a median of 4.5 years. This group included 1,139 patients with microalbuminuria and 333 patients with renal insufficiency (i.e., serum creatinine level less than 2.3 mg/dl). During the follow-up period, serum creatinine values remained stable in those persons without overt proteinuria, as well as in those persons with renal insufficiency, microalbuminuria, or both. It should be noted that the patients were treated aggressively with lipid-lowering agents and antihypertensives. In addition, the study included only a few smokers.[30]

Hypertension in Nondiabetic Patients

Blood pressure control is equally important in nondiabetic patients with CKD. In the Seventh Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC VII), a blood pressure goal of less than 130/85 mm Hg is recommended.[31] It should be anticipated that multiple medications will be required to achieve these levels of control in most CKD patients.

Treatment of Hypertension in Patients with CKD

The MDRD study helped define the target levels of blood pressure in persons with renal disease. The results suggest that in patients with less than 1 g of protein in a 24-hour urine collection, the target blood pressure should be 130/80 mm Hg or less.[32] For those with findings of more than 1 g of protein, a target blood pressure value of 125/75 mm Hg should be sought. A recent consensus group sponsored by the National Kidney Foundation has suggested that the target blood pressure value for adults with hypertension and diabetes should be 130/80 mm Hg.[33] In most cases, these lower blood pressure goals can be obtained without significantly increasing side effects [*see 1:III High Blood Pressure*]. The findings of the African American Study of Kidney Disease and Hypertension (AASK) trial provides evidence that factors other than blood pressure play important roles in the functional decline in CKD. A low blood pressure, (e.g., average of 128/78, compared with 141/85) did not improve protection in the participants. However, an ACE inhibitor proved superior to a dihydropyridine calcium channel blocker in retarding progression of hypertensive kidney disease.[34]

PROTEINURIA

In a study of elderly persons, urinary albumin excretion was found to be associated with cardiovascular disease.[6] In this study, prognostic factors were male sex, serum creatinine concentration, and hypertension. In a study of 11,343 German patients, the presence of microalbuminuria served to identify hypertensive patients with cardiovascular risks.[35] In diabetic patients, microalbuminuria is a well-known risk factor for progression to ESRD. A study from the Royal Infirmary in Edinburgh, United Kingdom, examined the course of 190 patients with type I diabetes who had had the disease for least 30 years'

duration; the study followed the participants for at least 5 years. Microalbuminuria was absent in 66% of participants at baseline; 11% of these individuals died during the follow-up period. Individuals with microalbuminuria at baseline were 22% of the total; of these, 26% died during the follow-up period. Of the 8% of patients who had macroalbuminuria at baseline, 44% died during the follow-up period. These data provide strong evidence for the influence of albuminuria on overall mortality.[36]

Interestingly, microalbuminuria is also a predictor of renal disease in patients who do not have diabetes.[37] The association between proteinuria and a decline in renal function is particularly strong in patients with a finding of more than 1 g of protein in a 24-hour urine collection. The mechanism behind the association between proteinuria and a decline in renal function appears to be in part related to the fact that high levels of urinary protein stimulate interstitial inflammation and fibrosis.[38] For these reasons, drug therapy that minimizes urinary protein loss seems prudent. Because angiotensin II stimulates *TGF-β1* gene expression and type IV collagen production, treatment with agents that block the effects of angiotensin are likely to be effective.

SMOKING

Smoking is a well-known risk factor for the development of cardiovascular disease. The World Health Organization estimates that as many as three million deaths a year worldwide are related to tobacco use. This number of deaths will likely increase substantially in the future. Smoking is an independent risk factor for microalbuminuria in both hypertensive and nonhypertensive patients.[39] The relationship between smoking and progression of renal disease is not widely appreciated; however, the risk of the progression of diabetic nephropathy has been known for over 20 years. A multicenter case-control European study of 582 patients with either IgA nephropathy or polycystic kidney disease demonstrated an increase in the risk of progression to ESRD with increased levels of smoking, as measured in pack-years; this risk was mitigated by the use of ACE inhibitors.[40] Bleyer and colleagues analyzed data from the Cardiovascular Health Study Cohort. In 4,142 nondiabetic patients who were at least 65 years of age, the number of cigarettes smoked per day was found to correlate with an increase in creatinine concentration of at least 0.3 mg/dl. In general, these patients had a relatively low incidence of progressive renal deterioration.[41] The mechanisms of smoking-related renal injury are complex. Possible alterations include changes in intrarenal hemodynamics, altered sympathetic activity, direct tubular damage, changes in intrarenal hormones, and oxidative stress. Physicians should be aware of the specific data regarding the overall risk of smoking to health. The renal risks of smoking should not be underestimated, and smoking should trigger an aggressive approach by the health care provider [*see Clinical Essentials: III Reducing Risk of Injury and Disease*].

PROTEIN INTAKE

Protein intake has long been considered a potential risk factor for the progression of renal disease; protein restriction, however, remains a controversial topic in the treatment of chronic kidney disease. Interest in protein restriction in patients with renal disease has been fueled by animal studies conducted over several decades that demonstrate that diets high in protein lead to histologic abnormalities in the kidney and to proteinuria and that such diets are associated with high death rates. These effects have been shown to be reduced or eliminated by protein or calorie restriction. The fact that the controversy still exists suggests that the effects of protein restriction are likely to be small and difficult to achieve. A meta-analysis of 890 nondiabetic patients reported in 46 studies suggested that protein restriction was beneficial in reducing ESRD.[42] There was a 46% decrease in the number of patients requiring replacement therapy. Another meta-analysis, of 1,413 patients, demonstrated a reduction of ESRD or mortality in both diabetic patients and nondiabetic patients who were assigned to receive a low-protein diet.[43] A third meta-analysis, of 1,919 randomized patients and 2,248 nonrandomized patients, found that a reduction in dietary protein had only a weak effect on the development of ESRD, with diabetic patients tending to benefit the most and with more benefit being associated with longer follow-up.[44]

The most ambitious test of the low-protein-diet hypothesis was the MDRD study, in which 585 patients with GFRs of 25 to 55 ml/min were randomized to receive either 1.3 or 0.58 g protein/kg/day. A second group, with GFRs of 13 to 24 ml/min, were randomized to receive either 0.58 or 0.29 g protein/kg/day.[45] The diets of patients in the latter group were supplemented with an amino acid–keto acid mixture that equalized the protein intake. Nuclear clearances with iodine-125–iothalamate were employed; the follow-up period was slightly longer than 2 years. During the initial 4 months of the study, renal function worsened more rapidly in the low-protein group; after that initial 4-month period, the rate of decline in GFR was lower. The decline in renal function seen at the end of the study did not differ between the two patient groups. Similarly, there was no clear difference in renal function in the group receiving a very low protein diet, but there was a trend toward better outcomes in the group receiving the lowest amount of protein. A more recent review of the data indicates that individuals in the lowest GFR category whose diets contained the lowest amount of protein may have benefited from the intervention. A multicenter trial in Europe assigned 456 adults with renal disease to either a low-protein diet (0.6 g/kg/day) or a normal-protein diet (1 g/kg/day).[46] The end point of the study was a doubling of the baseline creatinine level or a need for dialysis. The difference between the groups in cumulative renal survival was of borderline significance ($P < 0.06$). Compliance was good in the group receiving the higher amount of protein and poor in the low-protein group. Protein restriction is of particular concern in children because of the added complication of growth. During a 3-year study, a low-protein diet did not affect renal outcome in children from 2 to 18 years of age.[47]

Several caveats are important when one considers diet therapy in patients with CKD. First, one must ensure an adequate caloric intake of approximately 35 kcal/kg/day. Second, physicians caring for these patients should understand the techniques used to monitor overall nutrition, including anthropometric measurements and use of serum transferrin and serum albumin levels. Protein intake can be estimated through the use of 24-hour urine urea nitrogen measurements plus an estimate of nonurea nitrogen losses. Third, physicians should understand that uremic symptoms respond to diets low in protein. These diets normally entail reduced intake of sodium, potassium, phosphate, and acid. Thus, renal osteodystrophy, hypertension, and metabolic acidosis are improved. Fourth, low-protein diets can be used without provoking malnutrition; however, patients should be closely monitored. A low-protein diet of 0.6 to 0.7 g/kg/day can be safely used as adjunctive therapy in individuals with CKD to ameliorate symptoms and to reduce complications of uremia. In certain patients, particularly those

with diabetes, such diets may slow the rate of renal disease progression.

HYPERLIPIDEMIA

A substantial number of patients with nephrotic syndrome without impaired renal function will have lipid abnormalities, including patients with diabetes and heavy proteinuria. The etiology appears to be related to an increase in hepatic synthesis of lipids. In an elegant study, Appel and colleagues demonstrated an inverse correlation between the total plasma cholesterol level and serum albumin level.[48] There was also an inverse relationship between cholesterol level and plasma oncotic pressure. On the other hand, there was no correlation between cholesterol level and plasma viscosity. The incidence of hyperlipidemia is also increased in patients with CRF. Kasiske has estimated that approximately 30% of patients with CKD and proteinuria in the nonnephrotic range have total cholesterol values higher than 240 mg/dl. Levels of triglycerides and lipoprotein(a) are estimated to be higher than 200 mg/dl and 30 mg/dl, respectively, in approximately 60% of patients. High-density lipoprotein (HDL) cholesterol levels tend to be low, whereas only approximately 10% will have elevated low-density lipoprotein (LDL) values.[49] One issue that must be considered is the relationship between lipid abnormalities and the deterioration of renal function. In a study by Bleyer and coworkers, there was no association between cholesterol level and a rise of 0.3 mg/dl in creatinine level, with measurements made at least 3 years apart.[41] On the other hand, there are several studies, including studies with diabetic patients, that do demonstrate a relationship between lipid abnormalities and the progression of CRF. Because of the high incidence of cardiovascular disease, all patients with CRF should be screened. Recently, the guidelines for treatment were revised by the National Cholesterol Education Program.[50] Diet should be the first line of therapy, but the addition of lipid-lowering drugs is almost always necessary. Newer HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A) reductase inhibitors lower both LDL and triglyceride levels. Fibric acid analogues are effective in reducing triglyceride levels; however, myositis and rhabdomyolysis limit use of these agents. The dose of the fibric acid analogues should be adjusted appropriately for the degree of renal function.

Management of Complications of Chronic Kidney Disease

SODIUM AND WATER IMBALANCE

As renal mass becomes progressively reduced, the fractional excretion of salt and water increases in the remaining nephrons. The solute diuresis that occurs in the remaining nephrons results in a relatively fixed level of salt and water excretion. Thus, in patients with CKD, renal salt and water excretion is limited to a range that is quite a bit narrower than that of healthy persons.

The optimal salt intake will differ from patient to patient; once a level has been prescribed, salt intake will need to be constantly monitored, because requirements will vary as renal function changes. The goal should be a salt intake that results in the patient's being normotensive and maintaining a constant weight, with only trace edema present. A diet that restricts the amount of salt to 6 to 8 g/day is a useful starting point [*see Table 5*]. If the patient's weight begins to decrease over a period of several days and the patient becomes more azotemic, a higher salt intake is required. In addition, during intercurrent illness, supplemental salt can be given in the form of bouillon cubes if a deficit in extracellular fluid volume develops. By contrast, if the patient's weight increases over time and is accompanied by increasing edema and worsening hypertension, further salt restriction is indicated. Once the estimated GFR falls below 20 ml/min, even salt-restricted diets may exceed the excretory capacity of the kidney, and diuretic therapy will have to be utilized to prevent progressive expansion of the extracellular fluid volume.

Table 5 Typical Diet for Patients with Renal Insufficiency

0.8 protein/kg body weight	800–1,000 mg phosphate
6–8 g sodium	1,000–1,500 mg elemental calcium
70 mEq potassium	1,000–1,500 ml free water in excess of urine output

The ability to maximally concentrate or dilute the urine becomes progressively impaired as renal function declines. As a result, patients with CKD are at risk of developing positive water balance and resultant hyponatremia, as well as negative water balance and hypernatremia. In general, fluid intake should be equal to urine output plus an additional 1,000 to 1,500 ml/day to account for insensible losses. The treatment of hyponatremia depends on the existing extracellular fluid balance. In volume-overloaded patients, further water restriction is indicated. In hypovolemic patients, water restriction with judicious administration of salt and the withdrawing of diuretic therapy is the appropriate treatment.

POTASSIUM IMBALANCE

Potassium balance is generally maintained within normal limits until the GFR falls to less than 10 ml/min. This balance is achieved by an increased potassium excretion rate per remaining nephron and by an increase in extrarenal potassium excretion, primarily effected via the colon. The development of hyperkalemia at higher levels of renal function suggests the presence of tubulointerstitial disease or disturbances in the renin-angiotensin-aldosterone axis. In addition, there are several commonly used medications that can predispose to hyperkalemia in the patient with renal failure [*see Figure 1*].

The initial approach to the treatment of hyperkalemia is the institution of a low-potassium diet (50 to 70 mEq/day). Should hyperkalemia persist, administration of a loop diuretic is a reasonable second step, especially if the patient has demonstrable edema or is hypertensive. Loop diuretics increase distal sodium delivery and thus serve to increase potassium secretion from the distal tubule. If the patient is acidotic, sodium bicarbonate administration is an effective way to lower the serum potassium concentration. This agent also increases distal sodium delivery and, therefore, potassium secretion from the distal tubule. In addition, correction of the underlying acidotic state causes a shift of potassium into cells. Some patients continue to remain hyperkalemic despite this therapy. In these patients, a potassium-binding resin such as sodium polystyrene sulfonate (Kayexalate) may have to be given daily or every other day. This agent should be given with a bowel cathartic such as sorbitol to prevent constipation. Constipation can actually worsen hyperkalemia because potassium secretion by the colon is substantial in patients

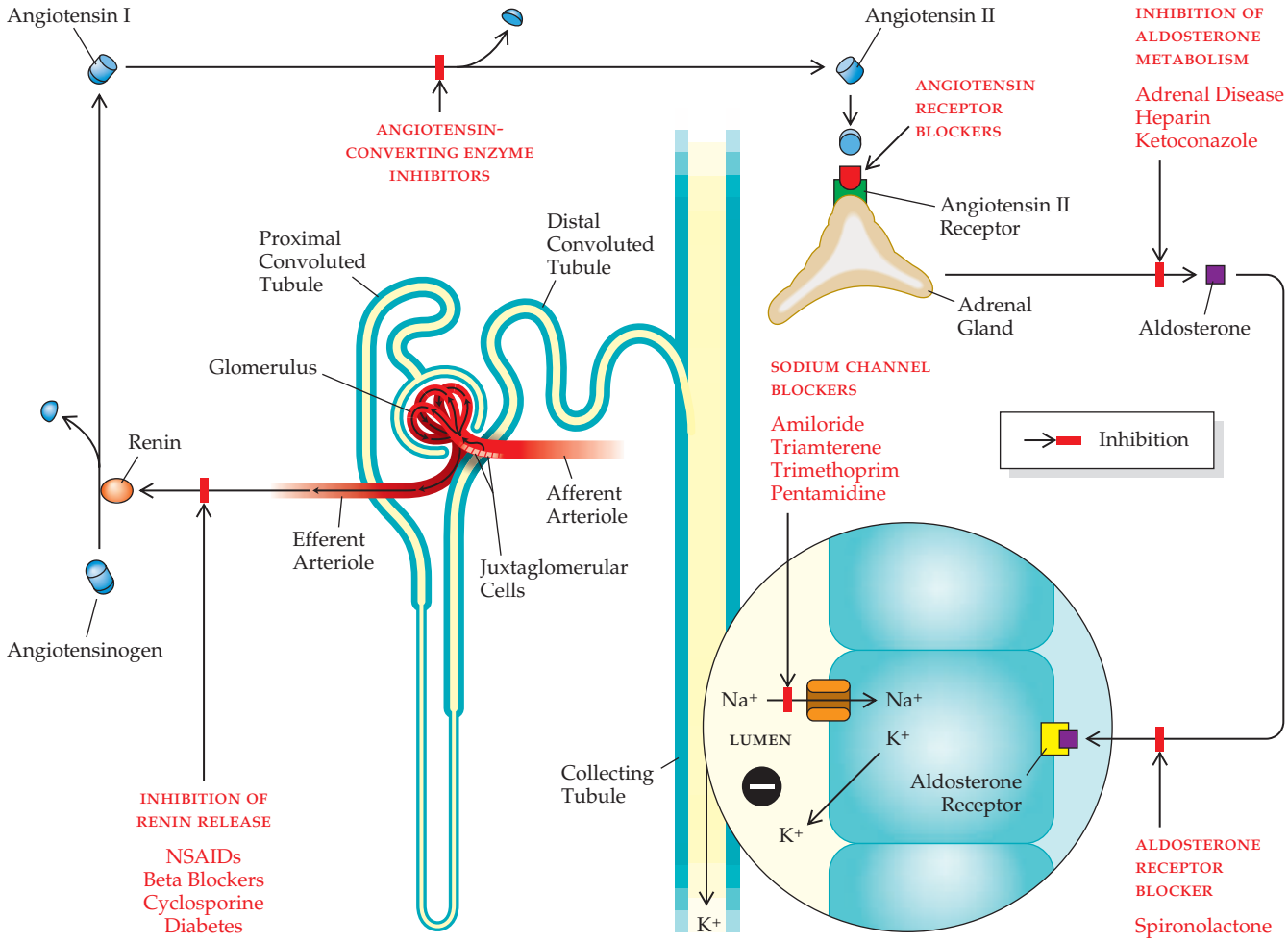


Figure 1 **This diagram depicts the renin-angiotensin-aldosterone cascade. Aldosterone stimulates sodium reabsorption in the collecting duct, which in turn generates a lumen-negative potential. The luminal electronegativity serves as a driving force for potassium excretion. Drugs that interfere with this process are depicted according to mechanism of action. Use of these agents in the setting of chronic renal insufficiency can predispose to the development of hyperkalemia. (AI—angiotensin I; AII—angiotensin II; K—potassium; Na—sodium; NSAIDs—nonsteroidal anti-inflammatory drugs)**

with advanced renal insufficiency. Cathartics that contain magnesium should be avoided because of the risk of inducing hypermagnesemia in the setting of renal insufficiency.

METABOLIC ACIDOSIS

Under normal conditions, the kidney is responsible for regenerating consumed bicarbonate as a result of the buffering of daily net acid production. As renal insufficiency progresses, patients typically become acidotic. Initially, the acidosis is of the non–anion gap type, but as renal insufficiency becomes far advanced, an anion gap acidosis supervenes. If left untreated, acidosis leads to bone resorption; in addition, it may contribute to protein catabolism and can result in malaise and dyspnea.[51]

Measurement and monitoring of the serum bicarbonate should be part of the routine electrolyte analysis in patients with CKD. In patients with stage 3 CKD, the bicarbonate should be measured at least every 12 months; in patients with stage 4 or 5 CKD, it should be measured every 3 months. Every effort should be made to keep the bicarbonate concentration above 22 mEq/L to avoid adverse effects on bone histology and protein catabolism.

Alkali therapy can be administered in the form of sodium bicarbonate tablets. Each 650 mg tablet contributes 8 mEq of bicarbonate. A useful starting dosage is one tablet three times a day. Alternatively, a sodium citrate solution (Bicitra) can be given; this solution contributes 1 mEq of bicarbonate per millimeter of solution. Citrate-containing alkali should not be administered to patients receiving aluminum-containing phosphate binders, because citrate is known to enhance the GI absorption of aluminum. Alkali therapy contains a substantial sodium load, and therefore, the patient needs to be monitored closely for the development of volume overload.

CALCIUM AND PHOSPHORUS IMBALANCE

Disturbances in calcium and phosphate metabolism regularly accompany CKD and contribute to many of the manifestations of uremia. As GFR declines, the serum phosphate level begins to increase, causing a reciprocal decrease in the serum calcium concentration. In response, parathyroid hormone (PTH) is released, resulting in increased phosphate excretion in each of the remaining nephrons; thus, calcium and phosphorus levels return to

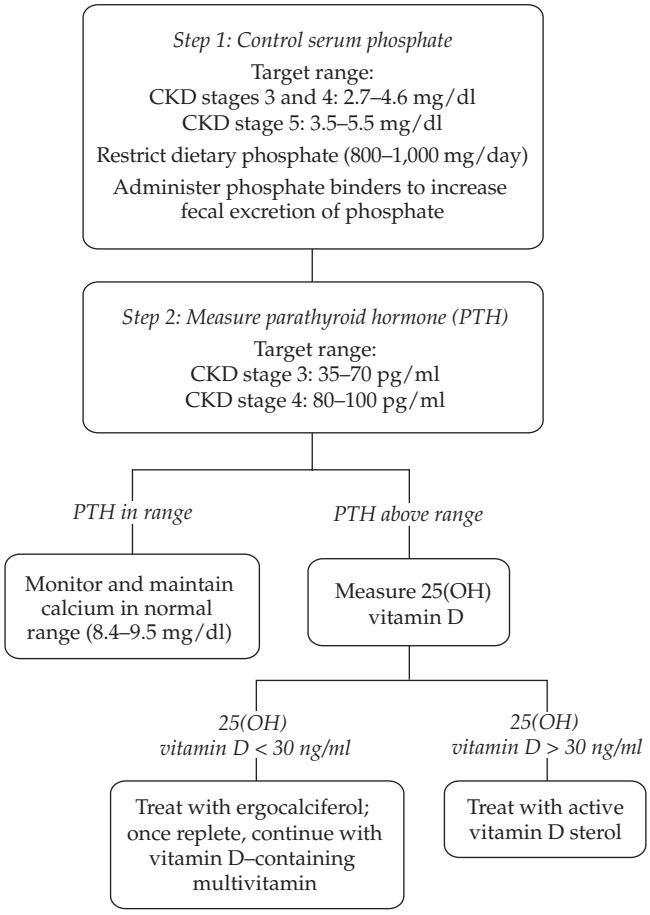


Figure 2 **Steps to manage calcium and phosphorus abnormalities that accompany chronic kidney disease.**

normal. As renal function continues to decline, calcium and phosphorus levels remain within the normal range but at the expense of an ever increasing level of PTH. Patients with stage 3 CKD commonly have elevated PTH levels and may already demonstrate evidence of osteitis fibrosa cystica on bone biopsy. Ultimately, loss of renal mass is so great that hyperphosphaturia per nephron is insufficient to prevent phosphate retention, so that hyperphosphatemia becomes sustained.

As renal mass declines, the circulating level of 1,25-dihydroxyvitamin D [1,25-(OH_2)D] also begins to fall. 25-Hydroxyvitamin D [25-(OH)D] undergoes 1α-hydroxylation in the kidney to form 1,25-(OH_2)D. Lack of 1,25-(OH_2)D contributes to the development of hypocalcemia because this hormone normally serves to increase calcium absorption from the GI tract and enhances the ability of PTH to mobilize calcium from bone. Decreased absorption of calcium from the intestine is further compounded by the low calcium content in the diet of patients with CKD. Low levels of active vitamin D also contribute to the development of secondary hyperparathyroidism, because this hormone normally exerts a direct inhibitory effect on the release of PTH from the parathyroid gland. Finally, during the course of advancing renal insufficiency, the set point at which calcium suppresses PTH release becomes higher, leading to further elevations in the level of PTH.

Clinical practice guidelines regarding the management of calcium and phosphorus disturbances in CKD patients are now available.[52] The primary goal in the management of patients with CKD is to maintain the serum phosphorus level within normal limits [*see Figure 2*]. The serum level of phosphorus should be maintained between 2.7 and 4.6 mg/dl in patients with stage 3 or 4 CKD. In patients with stage 5 disease, the serum level of phosphorus should be maintained between 3.5 and 5.5 mg/dl. To achieve these levels, the patient should initially be placed on a phosphate-restricted diet (800 to 1,000 mg/day); the serum phosphorus level should be monitored monthly.[52] Dietary sources particularly rich in phosphate must be restricted; these include eggs, dairy products (e.g., cream, milk, and cheese), and meat products. Although a few patients may be able to maintain the serum phosphate level within normal limits on a restricted diet alone, most patients with advanced CKD will require treatment with a phosphate binder to increase fecal excretion of phosphate.

Oral phosphate binders are available as either calcium- or non–calcium-containing drugs. In patients with stage 3 or 4 CKD, calcium-containing phosphate binders are usually effective in controlling the serum phosphorus. In patients with stage 5 disease, control of the serum phosphorus level may require a combination of both calcium- and non–calcium-containing binders.

The decision as to which class of binder to use should be based on the starting phosphate level and the calcium-times-phosphorus product. Every effort should be made to keep the calcium-phosphorus product lower than 55. In patients with a serum phosphorus level higher than 7 mg/dl or a calcium-phosphorus product greater than 63, a non–calcium-containing binder is the appropriate choice. Sustained use of calcium-containing binders in patients with a high product will result in the development of metastatic calcification; therefore, these binders should be avoided in such patients.

Sevelamer (RenaGel) is a calcium- and aluminum-free phosphate binder that is increasingly being used in the care of patients with ESRD. It has been shown to control serum phosphorus levels and to reduce PTH levels without inducing hypercalcemia.[53] In addition, this agent lowers serum cholesterol levels. Several trials are comparing the degree of major artery calcifications associated with calcium-containing and calcium-free phosphate binders. Lanthanum carbonate is another non–calcium-containing binder that may soon become available for clinical use.

Because of concerns about long-term toxicity, aluminum-containing binders have largely fallen from favor. Aluminum-containing compounds such as aluminum hydroxide and aluminum carbonate may be used as a short-term therapy (1 month) but should be replaced thereafter by other non–calcium- containing phosphate binders, such as sevelamer. As soon as the serum phosphate level is reduced to less than 7 mg/dl and the calcium-phosphorus product is less than 63, calcium-containing binders can be utilized.

The calcium-containing phosphate binders are available as the calcium salts of carbonate, acetate, and citrate. Recent evidence suggests that calcium acetate is the most potent phosphate binder in this class. To be most effective, all the phosphate binders should be given with meals. The effectiveness can be further enhanced by varying the dose of the binder in proportion to the phosphate content of each meal.

As previously discussed, patients with advancing renal insufficiency tend to develop negative calcium balance because of decreased GI absorption of calcium and decreased calcium content in the diet. To remain in calcium balance, most patients with CKD require 1,000 to 1,500 mg/day of elemental calcium. This level is difficult to achieve with diet alone because many foods that are high in calcium are also high in phosphorus and are

therefore restricted. To overcome this problem, supplemental calcium needs to be administered; supplemental calcium can be given as calcium carbonate or calcium acetate. When administered for this indication, calcium should be given between meals. Once again, to prevent calcium phosphate deposition in the tissues, calcium should not be given until the serum phosphate level is normalized.

Although calcium-containing binders provide an effective means of controlling phosphorus, their use may not be without risk. Calcium excess induced by the prescription of large doses of calcium-containing phosphate binders has been associated with calcifications of the aorta and the carotid and coronary arteries.[54] Use of these drugs has also been implicated in the development of calciphylaxis. Given these concerns, the total dose of elemental calcium provided by calcium-based phosphate binders should not exceed 1,500 mg/day, and the total intake of elemental calcium, including that derived from dietary sources, should not exceed 2,000 mg/day.

Chronic kidney disease is associated with the development of secondary hyperparathyroidism. Monitoring of plasma levels of intact PTH may help prevent the development of secondary hyperparathyroidism. In patients with stage 3 CKD, the target plasma level of intact PTH is 35 to 70 pg/ml; in patients with stage 4 disease, it is 80 to 100 pg/ml. For patients with values above the target range, 25-(OH)D levels should be obtained at first encounter; if the serum level is normal, the test should be repeated annually. If the serum level of 25-(OH)D is less than 30 ng/ml, supplementation with vitamin D_2 (ergocalciferol) should be initiated. Once vitamin D levels are replenished, the patient should be maintained on a multivitamin containing vitamin D.

Serum levels of 25-(OH)D are considered the measure of body stores of vitamin D. In patients with a GFR of 20 to 60 ml/min, levels of 25-(OH)D below 30 ng/ml are common. The prevention and treatment of vitamin D deficiency in patients with stage 3 or 4 CKD is believed to decrease the frequency and severity of secondary hyperparathyroidism.

Serum levels of calcium and phosphorus need to be monitored every 3 months after starting therapy with ergocalciferol. If the total corrected serum calcium level exceeds 10.2 mg/dl, vitamin D therapy should be discontinued. If the serum phosphorus level exceeds 4.6 mg/dl, phosphate binders should be initiated or the dose increased. Persistently increased serum phosphorus levels should prompt the discontinuance of vitamin D.

In patients who have stage 3 or 4 CKD, serum 25-(OH)D levels greater than 30 ng/ml, and plasma levels of PTH above the target range, therapy with an active form of vitamin D (calcitriol, alfacalcidol, or doxercalciferol) is indicated. There were early concerns that administration of the active form of vitamin D would hasten the loss of renal function by causing hypercalcemia, hyperphosphatemia, and hypercalciuria. Reports to date have generally shown no change in renal function in association with vitamin D therapy, provided that prolonged hypercalcemia is avoided. As a result, close monitoring of both the serum calcium level and the phosphate concentration are required, because vitamin D enhances the GI absorption of these electrolytes. Treatment should be initiated only if the total corrected serum calcium level is less than 9.5 mg/dl and the serum phosphorus level is less than 4.6 mg/dl. The serum levels of calcium and phosphorus should be monitored monthly for the first 3 months after initiation of therapy and every 3 months thereafter. The active vitamin D sterol should be held for calcium values that exceed 9.5 mg/dl or serum levels of phosphorus greater than 4.6 mg/dl.

Plasma levels of PTH also need to be monitored during therapy when the active form of vitamin D is being used. The target values for PTH in patients with CKD are higher than normal because of evidence that higher levels are required for normal bone remodeling, presumably as a result of the end-organ resistance to PTH in patients with uremia. Suppression of PTH to normal nonuremic values is not desirable, because such PTH levels are associated with a higher prevalence of a dynamic bone disease. After the initiation of therapy with vitamin D, plasma PTH levels should be measured every 3 months. Vitamin D should be withheld when PTH values fall below the target range.

Once patients reach stage 5 CKD, levels of PTH are almost always elevated. A plasma level of intact PTH of 300 pg/ml should prompt the initiation of active vitamin D therapy, with the goal of reducing PTH levels to a target range of 150 to 300 pg/dl. As with earlier stages of CKD, close monitoring of serum calcium and phosphorus levels is required. In these patients, treatment with ergocalciferol is not indicated, because there is inadequate renal mass to convert 25–(OH)D to the active vitamin D sterol.

ANEMIA

Patients with CKD almost uniformly develop a normocytic, normochromic anemia that tends to worsen in parallel with advancing azotemia. Anemia can develop in CKD patients who have a serum creatinine level as low as 2 mg/dl (occasionally lower), particularly in individuals with reduced muscle mass. The anemia seen in CKD patients is primarily caused by a decrease in the biosynthesis of erythropoietin from the kidney. Recombinant human erythropoietin is now the most definitive treatment of the anemia of CKD. In addition to freeing the patient from repetitive exposures to blood-borne pathogens, iron overload, and sensitization, the use of erythropoietin has been demonstrated to improve cardiovascular and cognitive function and the overall quality of life of patients with CRF. Although transfusions are clearly indicated for the treatment of acute hemorrhage and cardiovascular instability, this form of therapy should no longer be considered routine in the management of anemia of patients undergoing peritoneal dialysis or hemodialysis.

Clinical practice guidelines are now available for the management of anemia in patients with CKD.[1] Before the initiation of erythropoietin therapy, the patient should undergo a workup to exclude causes of anemia other than CKD. The evaluation should at least include hemoglobin levels, red blood cell indices, iron parameters, and testing for occult blood in the stool. Monitoring of iron levels should include testing for serum iron, total iron-binding capacity, percent transferrin saturation, and serum ferritin. If it is concluded that the anemia is the result of CKD, erythropoietin therapy can be initiated. Measurement of erythropoietin levels is usually not indicated.

Erythropoietin is administered subcutaneously in doses of 50 to 150 U/kg up to three times a week. The target range for hemoglobin (hematocrit) is Hgb 11 to 12 g/dl (33% to 36%). The need to administer erythropoietin two to three times a week can be bothersome to the patient with early CKD who is largely asymptomatic and ambulatory. Darbepoetin alfa, an erythropoiesis-stimulating protein, offers the advantage of less frequent administration. Darbepoetin alfa has a much longer half-life than recombinant human erythropoietin. Dosing once or twice weekly can often maintain hemoglobin concentrations in the target range. The optimum starting dose is 0.45 μg/kg once or twice weekly administered subcutaneously.

Failure to respond to erythropoietin therapy is most commonly the result of iron deficiency. A transferrin saturation of less than 25% or a serum ferritin level of less than 100 mg/L indicates inadequate iron stores; such a condition requires iron supplementation, usually given as ferrous sulfate, 325 mg twice or three times a day. Oral iron is best absorbed when it is ingested without food or medications. Intravenous iron is usually reserved for patients who are already receiving hemodialysis or peritoneal dialysis. The occasional patient with early CKD who requires intravenous iron supplementation can be given 500 to 1,000 mg of iron dextran administered intravenously in a single infusion. An initial test dose of 25 mg should precede an infusion of 1,000 mg. During the course of therapy, the transferrin saturation and serum ferritin level should be monitored frequently to ensure that iron deficiency does not develop. Other causes of a suboptimal response include the presence of an underlying inflammatory illness, aluminum intoxication, and the presence of marrow fibrosis due to long-standing hyperparathyroidism.

Patients with advanced CKD typically develop a qualitative defect in platelet function. In patients at risk for bleeding complications, three forms of therapy have been shown to be effective in lowering the prolonged bleeding time associated with uremia. First, one can administer desmopressin intravenously at a dose of 0.3 mg/kg in 50 ml of normal saline solution infused over 30 minutes. Alternatively, one can administer cryoprecipitate (10 bags) infused intravenously over 30 minutes. Finally, conjugated estrogens given at a dosage of 0.6 mg/kg I.V. daily for 5 consecutive days have also been shown to be effective.

The Case for Early Referral to a Nephrologist

Optimal care of the patient with CKD involves a multifaceted treatment approach that includes close monitoring of renal function and aggressive institution of measures designed to slow the progression of loss of renal function. Interventions to reduce the comorbidities that accompany CKD should be initiated early. Metabolic and hematologic complications of uremia should be prevented; if already present, they should be treated judiciously. Patients with advanced CKD should be adequately prepared so that referral for renal replacement therapy is smooth and timely.

Disturbingly, recent reports indicate that pre-ESRD care in the United States is suboptimal for a substantially large number of patients.[55-58] Evidence suggests that less than ideal management of the pre-ESRD patient may be an important factor contributing to the high morbidity and mortality of patients receiving dialysis. For example, many patients are significantly anemic and have not been treated with erythropoietin before initiation of dialysis. Severe anemia contributes to the development of left ventricular hypertrophy, which in turn is an important predictor of subsequent cardiac morbidity and mortality in patients receiving dialysis. Hypoalbuminemia is also a common finding at the time dialysis is initiated. Hypoalbuminemia is a strong predictor of subsequent morbidity and mortality in dialysis patients. Although the cause of hypoalbuminemia is multifactorial in this setting, it is likely that lack of supervision by a qualified dietitian early in the course of the disease is a contributing factor.

There is also evidence that many patients are not adequately prepared for initiation of dialysis. Optimal preparation for initiation of dialysis involves educating the patient and family about the various forms of renal replacement therapy. In those who choose hemodialysis, a vascular-access device needs to be placed several months before initiation so that it can be used for the first treatment. Because of the time commitment required to be trained for peritoneal dialysis, patients referred late are more likely to be started on hemodialysis, effectively limiting patient choice. The lack of a permanent access device at the time of initiation necessitates placement of a temporary catheter and increases the likelihood that the patient will receive an arteriovenous graft rather than an arteriovenous fistula. Finally, evidence is accumulating that early initiation of dialysis is associated with improved patient outcome.

Given the complexities involved in management, patients with CKD should be referred to a specialist for consultation and comanagement if the primary care provider cannot adequately evaluate and treat the patient. A nephrologist should participate in the care of patients who have an estimated GFR of less than 30 ml/min/1.73 m^2. To achieve optimal management of all CKD patients, education targeting patients, generalists, and nephrologists is required.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Clinical practice guidelines for chronic kidney disease. Am J Kidney Dis 39(suppl 1):S1, 2002

2. Trivedi HS, Pang MMH: Discrepancy in the epidemiology of nondiabetic chronic renal insufficiency and end-stage renal disease in black and white Americans: The Third National Health and Nutrition Examination Survey and United States Renal Data System. Am J Nephrol 23:448, 2003

3. Fink JC, Salmanullah M, Hise MK et al: The inevitability of renal function loss in patients with hypercreatinemia. Am J Nephrol 21:386, 2001

4. Shulman NB, Ford CE, Hall WD, et al: Prognostic value of serum creatinine and effect of treatment of hypertension on renal function: results from the hypertension detection and follow-up program. The Hypertension Detection and Follow-up Program Cooperative Group. Hypertension 13(suppl):I80, 1989

5. Mann JF, Gerstein HC, Pogue J, et al: Renal insufficiency as a predictor of cardiovascular outcomes and the impact of ramipril: the HOPE randomized trial. Ann Intern Med 134:629, 2001

6. Damsgaard EM, Froland A, Jorgensen OD, et al: Microalbuminuria as predictor of increased mortality in elderly people. BMJ 300:297, 1990

7. Wannamethee SG, Shaper AG, Perry IJ: Serum creatinine concentration and risk of cardiovascular disease: a possible marker for increased risk of stroke. Stroke 28:557, 1997

8. Friedman PJ: Serum creatinine: an independent predictor of survival after stroke. J Intern Med 229:175, 1991

9. Matts JP, Karnegis JN, Campos CT, et al: Serum creatinine as an independent predictor of coronary heart disease mortality in normotensive survivors of myocardial infarction. POSCH Group. J Fam Pract 36:497, 1993

10. Furberg CD, Byington RP: What do subgroup analyses reveal about differential response to beta-blocker therapy? The Beta-Blocker Heart Attack Trial experience. Circulation 67(suppl 1):I98, 1983

11. Levey AS, Beto JA, Coronado BE, et al: Controlling the epidemic of cardiovascular disease in chronic renal disease: what do we know? What do we need to learn? Where do we go from here? Am J Kidney Dis 32:853, 1998

12. Hise MK, Mehta PS: Characterization and localization of calcium/phospholipid-dependent protein kinase-C during diabetic renal growth. Endocrinology 123:1553, 1988

13. Craven PA, DeRubertis FR: Protein kinase C is activated in glomeruli from streptozotocin diabetic rats. J Clin Invest 83:1667, 1989

14. Koya D, King GL: Protein kinase C activation and the development of diabetic complications. Diabetes 47:859, 1998

15. Ishii H, Jirousek MR, Koya D, et al: Amelioration of vascular dysfunctions in diabetic rats by an oral PKC beta inhibitor. Science 272:728, 1996

16. Zatz R, Dunn BR, Meyer TW, et al: Prevention of diabetic glomerulopathy by pharmacological amelioration of glomerular capillary hypertension. J Clin Invest 77:1925, 1986

17. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. N Engl J Med 329:977, 1993

18. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. N Engl J Med 342:381, 2000

19. Krolewski AS: Genetics of diabetic nephropathy: evidence for major and minor gene effects. Kidney Int 55:1582, 1999

20. Fioretto P, Mauer M, Brocco E, et al: Patterns of renal injury in NIDDM patients with microalbuminuria. Diabetologia 39:1569, 1996

21. Hsu CY, Bates DW, Kuperman GJ, et al: Diabetes, hemoglobin A1c, cholesterol, and the risk of moderate chronic renal insufficiency in an ambulatory population. Am J Kidney Dis 36:272, 2000

22. Anderson S, Meyer TW, Rennke HG, et al: Control of glomerular hypertension limits glomerular injury in rats with reduced renal mass. J Clin Invest 76:612, 1985

23. Anderson S, Rennke HG, Brenner BM: Therapeutic advantage of converting enzyme inhibitors in arresting progressive renal disease associated with systemic hypertension in the rat. J Clin Invest 77:1993, 1986

24. Lewis EJ, Hunsicker LG, Bain RP, et al: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. N Engl J Med 329:1456, 1993

25. Lewis EJ, Hunsicker LG, Clarke WR, et al: Renoprotective effect of the angiotensin-receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. N Engl J Med 345:851, 2000

26. Brenner BM, Cooper ME, DeZeeuw D, et al: Effect of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. N Engl J Med 345:861, 2001

27. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: United Kingdom Prospective Diabetes Study 38. United Kingdom Prospective Diabetes Study Group. BMJ 317:703, 1998

28. Efficacy of atenolol and captopril in reducing risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 39. United Kingdom Prospective Diabetes Study Group. BMJ 317:713, 1998

29. Bakris GL, Weir MR, Shanifar S, et al: Effects of blood pressure level on progression of diabetic nephropathy: results from the RENAAL study. Arch Intern Med 163:1555, 2003

30. Mann JFE, Gerstein HC, Yi, Q-L: Progression of renal insufficiency in type 2 diabetes with and without microalbuminuria: results of the heart outcomes and prevention evaluation (HOPE) randomized study. Am J Kidney Dis 42:936, 2003

31. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (NIH Publication No. 03-5233). National Institutes of Health, Bethesda, Maryland, 2003

32. Peterson JC, Adler S, Burkart JM, et al: Blood pressure control, proteinuria, and the progression of renal disease. The Modification of Diet in Renal Disease (MDRD) Study Group. Ann Intern Med 123:754, 1995

33. Bakris GL, Williams M, Dworkin L, et al: Preserving renal function in adults with hypertension and diabetes: a consensus approach. The National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group. Am J Kidney Dis 36:646, 2000

34. Agodoa LY, Appel L, Bakris GL et al: African American Study of Kidney Disease and Hypertension (AASK) study group: Effects of ramipril vs amlodipine on renal outcomes in hypertensive nephrosclerosis: a randomized control trial. JAMA 285:2719, 2001

35. Agrawal B, Berger A, Wolf K, et al: Microalbuminuria screening by reagent strip predicts cardiovascular risk in hypertension. J Hypertens 14:223, 1996

36. Allen KV, Walker JD: Microalbuminuria and mortality in long-duration type 1 diabetes. Diabetes Care 26:2389, 2003

37. Ruggenenti P, Perna A, Mosconi L, et al: Urinary protein excretion rate is the best independent predictor of ESRF in non-diabetic proteinuric chronic nephropathies. Kidney Int 53:1209, 1998

38. Palmer BF: The renal tubule in the progression of chronic renal failure. J Investig Med 45:346, 1997

39. Orth SR: Smoking: a renal risk factor. Nephron 86:12, 2000

40. Orth SR, Stockmann A, Conradt C, et al: Smoking as a risk factor for end-stage renal failure in men with primary renal disease. Kidney Int 54:926, 1998

41. Bleyer AJ, Shemanski LR, Burke GL, et al: Tobacco, hypertension, and vascular disease: risk factors for renal functional decline in an older population. Kidney Int 57:2072, 2000

42. Fouque D, Laville M, Boissel JP, et al: Controlled low protein diets in chronic renal insufficiency: meta-analysis. BMJ 304:216, 1992

43. Pedrini MT, Levey AS, Lau J, et al: The effect of dietary protein restriction on the progression of diabetic and nondiabetic renal diseases: a meta-analysis. Ann Intern Med 124:627, 1996

44. Kasiske BL, Lakatua JD, Ma JZ, et al: A meta-analysis of the effects of dietary protein restriction on the rate of decline in renal function. Am J Kidney Dis 31:954, 1998

45. Klahr S, Levey AS, Beck GJ, et al: The effects of dietary protein restriction and blood-pressure control on the progression of chronic renal disease. The Modification of Diet in Renal Disease Study Group. N Engl J Med 330:877, 1994

46. Locatelli F, Alberti D, Graziani G, et al: Prospective randomised, multicentre trial of effect of protein restriction on progression of chronic renal insufficiency. Lancet 337:1299, 1991

47. Wingen AM, Fabian-Bach C, Schaefer F, et al: Randomised multicentre study of a low-protein diet on the progression of chronic renal failure in children. European Study Group of Nutritional Treatment of Chronic Renal Failure in Childhood. Lancet 349:1117, 1997

48. Appel GB, Blum CB, Chien S, et al: The hyperlipidemia of the nephrotic syndrome. N Engl J Med 312:1544, 1985

49. Kasiske BL: Hyperlipidemia in patients with chronic renal disease. Am J Kidney Dis 32(suppl 3):S142, 1998

50. Executive Summary of the Third Report of the National Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 285:2486, 2001

51. Alpern RJ, Sakhaee K: The clinical spectrum of chronic metabolic acidosis: homeostatic mechanisms produce significant morbidity. Am J Kidney Dis 29:291, 1997

52. Clinical practice guidelines for bone metabolism and disease in chronic kidney disease. Am J Kidney Dis 42(suppl 3):S1, 2003

53. Slatopolsky EA, Burke SK, Dillon MA: RenaGel, a nonabsorbed calcium- and aluminum-free phosphate binder, lowers serum phosphorus and parathyroid hormone. The RenaGel Study Group. Kidney Int 55:299, 1999

54. Block GA: Control of serum phosphorus: implications for coronary artery calcification and calcific uremic arteriolopathy (calciphylaxis). Curr Opin Nephrol Hypertens 10:741, 2001

55. Arora P, Obrador G, Ruthazer R, et al: Prevalence, predictors, and consequences of late nephrology referral at a tertiary care center. J Am Soc Nephrol 10:1281, 1999

56. Obrador G, Ruthazer R, Arora P, et al: Prevalence of and factors associated with suboptimal care before initiation of dialysis in the United States. J Am Soc Nephrol 10:1793, 1999

57. Nissenson AR, Collins A, Hurley J, et al: Opportunities for improving the care of patients with chronic renal insufficiency: current practice patterns. J Am Soc Nephrol 12:1713, 2001

58. Kausz AT, Khan S, Abichandani R, et al: Management of patients with chronic renal failure in the northeastern United States. J Am Soc Nephrol 12:1501, 2001

Acknowledgment

Figure 1 Seward Hung.

V GLOMERULAR DISEASES

Raimund H. Pichler, m.d.
Stuart J. Shankland, m.d

Clinical Classification of Glomerular Disease

A number of inflammatory and noninflammatory diseases can affect glomerular integrity and function. In general, nephrologists use the term glomerulonephritis to denote glomerular diseases that result from an inflammatory process; these diseases also usually have an immunologic component. Important examples of glomerular diseases that are not caused by an inflammatory process include diabetic nephropathy [*see 9:VI Diabetes Mellitus*], hypertensive nephrosclerosis [*see 1:III Hypertension*], amyloidosis, and hereditary nephropathies (e.g., Alport syndrome). Diabetic nephropathy and hypertensive nephrosclerosis are the most common causes of chronic renal failure and end-stage renal disease in North America.

The clinical approach to glomerular disease often appears more complicated than it perhaps should be. This stems in part from the variety of classification systems used for glomerular disease, which categorize these disorders on the basis of different criteria—clinical syndromes, underlying renal pathology (including the percentage of glomeruli involved), molecular and cellular mechanisms underlying the injury,[1] the localization of immune deposits,[2] levels of serum complement,[3] and specific antibodies involved.[4] Moreover, although glomerular diseases may be named according to descriptive histologic findings, such as proliferative glomerulonephritis, these classic names do not distinguish between primary renal disease (i.e., disease limited to the kidneys) and glomerular involvement that is secondary to a multisystem disease.

To simplify matters, this chapter classifies glomerular disease into two main groups, on the basis of the clinical presentation: nephritic syndrome (caused by glomerulonephritis) and nephrotic syndrome [*see Table 1 and Figure 1*]. Occasionally, a patient presents with features of both syndromes (nephritic-nephrotic syndrome). Two additional points should be noted. First, although patients typically present with nephritic or nephrotic syndrome, they may manifest an underlying systemic disease, such as systemic lupus erythematosus (SLE) or systemic vasculitis. Also, in early stages of glomerular diseases, patients may present with only hematuria or low-grade proteinuria before developing full-blown nephritic or nephrotic syndrome. Early recognition of the underlying disease process—with referral to a nephrologist—renal biopsy if indicated, and initiation of appropriate therapy are essential in preventing progressive renal injury. Second, patients may present with complications of glomerular disease, such as hypertension, electrolyte and acid-base disturbances, and edema. Regardless of the clinical presentation, the physician should correlate the clinical syndrome with the underlying pathophysiologic events.

Epidemiology

Diseases of the glomerulus are the most common causes of end-stage renal disease worldwide. In the United States, the major glomerular diseases causing end-stage renal disease include diabetes mellitus (43%), hypertension (26%), and glomerulonephritis (9%). Worldwide, by far the leading cause of end-stage renal disease is glomerulonephritis, probably because of the prevalence of infectious diseases in developing countries. It is estimated that only 10% to 20% of patients with glomerulonephritis show clinical symptoms; thus, the prevalence of glomerular diseases may be underestimated.

Pathogenesis

Each kidney contains approximately one million glomeruli. The glomerulus is a complex structure comprising a series of capillary loops that derive from the afferent arteriole and drain into the efferent arteriole. There are four glomerular cell types: mesangial cells, glomerular endothelial cells, podocytes (also called visceral epithelial cells), and parietal epithelial cells [*see Figure 1a*]. Each glomerular cell type serves distinct functions, and injury to each cell type results in a different renal histologic pattern and clinical picture [*see Figure 1b*].

The normal glomerular filtration rate (GFR) is about 80 to 120 ml/min. Approximately 30% of the plasma flow is filtered at the level of the glomerulus. The major barrier to filtration of plasma constituents is the glomerular capillary wall, which consists of glomerular endothelial cells, glomerular basement membrane

Table 1 Clinical Presentation and Common Causes of Glomerular Disease

	Nephritic Syndrome	*Nephrotic Syndrome*
Clinical findings	Hypertension Edema Reduced glomerular filtration rate Active urinary sediment (red cell casts) Proteinuria < 3.5 g/24 hr	Hypoalbuminemia Edema Hyperlipidemia Oval fat bodies in urine Proteinuria > 3.5 g/24 hr
Primary renal disease	IgA nephropathy Poststreptococcal GN*† Idiopathic necrotizing GN Anti-GBM disease	Focal segmental glomerulosclerosis Membranous nephropathy Minimal change disease Membranoproliferative GN*‡
Systemic disease	Systemic lupus erythematosus*† Vasculitis (ANCA-related): Wegener granulomatosis Microscopic polyarteritis Hypersensitivity vasculitis Henoch-Schönlein purpura Goodpasture syndrome Cryoglobulinemia	Diabetes mellitus Amyloidosis

*Low complement.
†May present with the nephrotic syndrome.
‡May present with the nephritic syndrome.
ANCA—antineutrophil cytoplasmic antibody GBM—glomerular basement membrane GN—glomerulonephritis

(GBM), and podocytes.[5] The normal rate of urinary protein excretion is less than 150 mg/day, of which less than 20 mg is albumin. The normal filtration barrier limits passage of molecules by size and charge selectivity.[6] The negative charge of the glomerular endothelial cells, GBM, and podocytes limits the passage of anionic substances, such as albumin.[7] Most glomerular diseases cause an alteration in both size and charge selectivity, which leads to proteinuria.

Glomerular diseases can be caused by a variety of mechanisms, including immune mediated, hemodynamic (e.g., hypertension, reduced renal mass), metabolic (e.g., diabetes, metabolic syndrome), and hereditary (e.g., defects of the GBM in Alport syndrome). The most common cause of glomerulonephritis is immune-mediated injury.[8]

Two mechanisms underlie antibody-mediated glomerular disease. One is the deposition of circulating antibodies directed against specific antigens in the glomerulus. The major example is Goodpasture syndrome, in which an antibody is directed against the noncollagenous portion of the α3 chain of type IV collagen in the GBM.[9,10] Antibodies may be directed against an intrinsic glomerular antigen; for example, in idiopathic membranous nephropathy, the antibody is directed against an antigen on the podocyte.[11] Alternatively, glomerular disease may result when antibody-antigen complexes circulating in the bloodstream are deposited in the subendothelial space or mesangium during the normal filtration process. Examples of such immune complex deposition include poststreptococcal glomerulonephritis (PSGN) and membranoproliferative glomerulonephritis (MPGN) from hepatitis C.[12]

Immune complexes generally activate either the alternative or the classical pathway of the complement cascade, thereby initiating injury.[13] The site of the immune complex deposition determines which glomerular cell is injured and, therefore, which clinical manifestations develop.

In contrast to immune-complex mechanisms, certain glomerular diseases develop primarily from cell-mediated immunity.[14] It is postulated that T cells sensitized to exogenous or endogenous antigens in the glomerulus recruit macrophages into the glomerulus, leading to a delayed-type hypersensitivity reaction. A classic example is minimal change disease, which has been postulated to occur when a T cell product injures podocytes and induces a permeability defect.

Diagnosis

Although renal biopsy remains the definitive diagnostic tool in glomerular disease, a thorough clinical evaluation is essential in determining the etiology and nature of these cases.

HISTORY

A carefully obtained history may help define a possible cause of glomerular disease. A family history of renal disease may suggest Alport syndrome, especially if affected individuals have hearing loss or another of the manifestations that are characteristic of this syndrome; there are also familial forms of IgA nephropathy, focal segmental glomerulosclerosis (FSGS), and hemolytic-uremic syndrome.

A number of glomerular diseases may be caused by drugs or toxins. For example, minimal change disease has been associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs) and interferon. Gold, penicillamine, NSAIDs, and mercury have been shown to cause membranous nephropathy. Thrombotic microangiopathy has been associated with the use of cyclosporine, tacrolimus, mitomycin C, and oral contraceptives.

Certain glomerular disorders are associated with malignancies. These include membranous nephropathy, which is associated with lung, breast, and gastrointestinal cancers; minimal change disease with Hodgkin disease; MPGN with non-Hodgkin lymphoma; and amyloid with renal cell carcinoma.

SIGNS AND SYMPTOMS

Patients with glomerular disease may be asymptomatic or may present with manifestations ranging from minimal findings to full-blown nephritic or nephrotic syndrome. Many patients have only mild to moderate proteinuria (150 mg to 3 g/24 hr) or microscopic hematuria (> two red blood cells [RBCs] per high-power field). Others present with painless gross hematuria. The classic example is hematuria immediately following an intercurrent infection, which occurs in IgA nephropathy; onset of hematuria 2 to 3 weeks after infection is typical of postinfectious glomerulonephritis. Patients with rapidly progressive glomerulonephritis (RPGN) present with rapid onset of renal failure over days or weeks with hematuria (from RBC casts) and proteinuria.

Patients with chronic glomerulonephritis generally have hypertension, renal insufficiency, and proteinuria. On renal ultrasonography, the kidneys are somewhat smaller than normal and display increased echogenicity.

RENAL BIOPSY

The ultimate diagnostic tool in many renal diseases is a renal biopsy, which is performed percutaneously with a spring-loaded biopsy needle, using local anesthesia, or surgically as an open biopsy with the patient under general anesthesia. Renal biopsies should be performed only after the patient has been evaluated by a nephrologist, but they also should be performed in a timely fashion, because early initiation of therapy is essential in preserving renal function and preventing further, potentially irreversible, injury to the kidney.

Although renal biopsy is generally a safe procedure, complications include hematoma, gross hematuria, arteriovenous fistula, and infection. In rare cases, renal biopsy leads to complications requiring surgical intervention.

Four situations mandate consideration of renal biopsy: (1) nephrotic syndrome (except in children, who may be assumed to have minimal change syndrome, and in diabetics, who are likely to have diabetic nephropathy); (2) renal disease in the setting of a systemic disorder (e.g., SLE, myeloma, amyloid, or vasculitis); (3) acute renal failure in the setting of glomerular disease; and (4) renal abnormalities in renal transplant recipients (when there is a possibility of rejection or of recurrent or de novo glomerular disease). Some patients with nonnephrotic proteinuria, hematuria, and chronic renal failure may also benefit from a renal biopsy for diagnostic and prognostic purposes.

The renal biopsy establishes a diagnosis, determines whether disease is mediated by antibody or by complement, and assists in determining disease-specific therapy. The biopsy also provides important information on the degree of glomerular and interstitial fibrosis. The latter is particularly important as a prognostic index. The renal pathology report includes a description of the kidney by light microscopy, immunofluorescence, and electron microscopy.

a

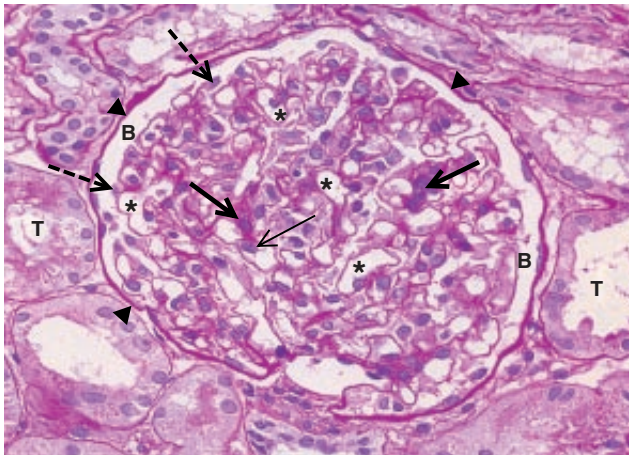
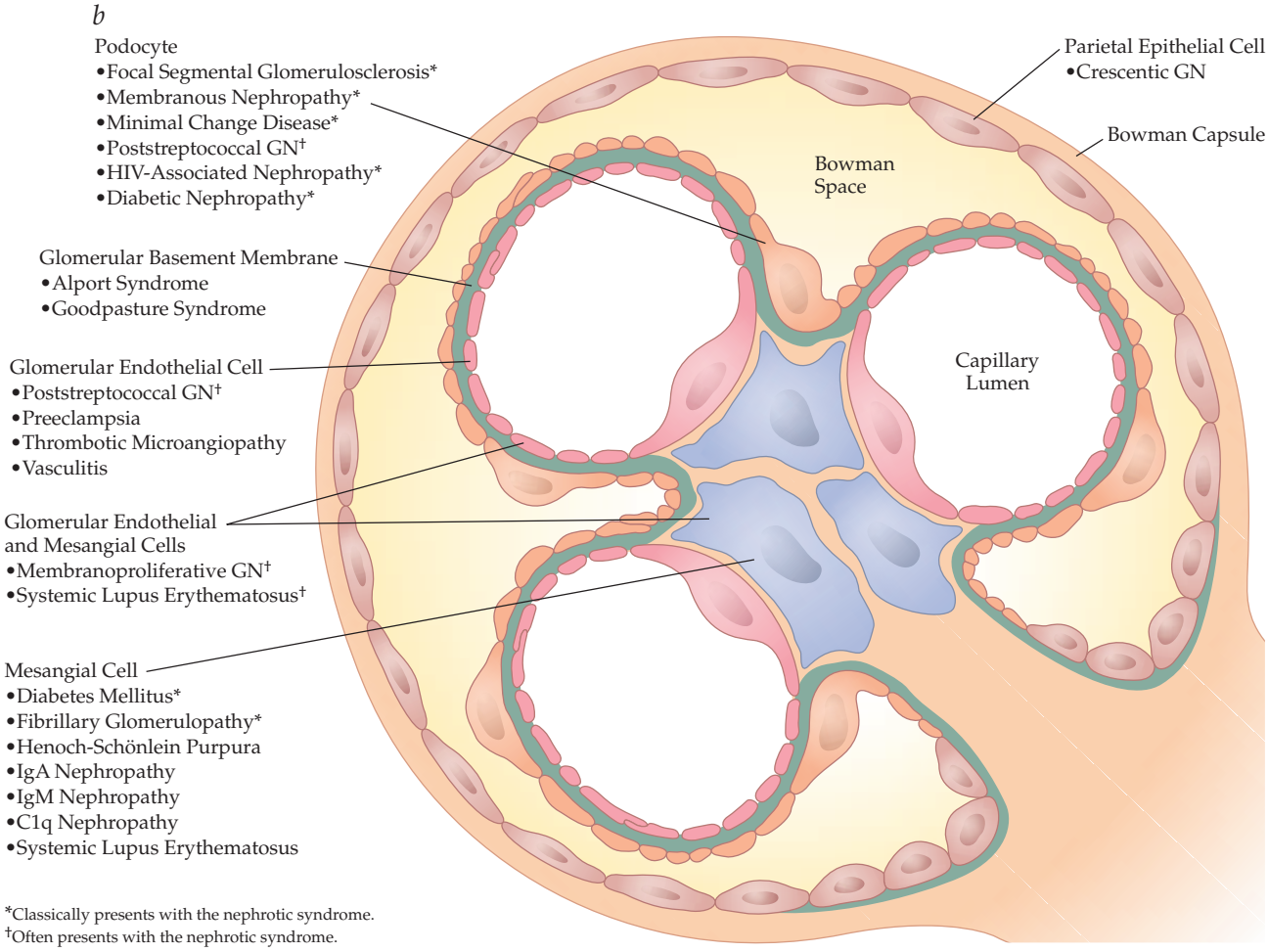


Figure 1 **(*a*) Light micrograph illustrating the different components of a normal glomerulus. Thick arrows illustrate mesangial cells; thin arrow illustrates glomerular endothelial cell; * illustrates capillary lumen; B illustrates Bowman space; dashed arrows illustrate podocytes; arrowheads illustrate Bowman capsule; and T illustrates tubule. (*b*) Schematic representation showing that injury to individual glomerular cell types or glomerular components causes specific forms of glomerular disease or glomerulonephritis.**

b



*Classically presents with the nephrotic syndrome.
†Often presents with the nephrotic syndrome.

Light Microscopy

Light microscopy describes glomerular cellularity—that is, whether the number of glomerular cells is normal or increased (hypercellularity). Often, light microscopy can distinguish which cell type (resident glomerular cells or infiltrating cells such as neutrophils) is increased; whether the GBMs are thickened and whether the capillary loops are patent, collapsed, or filled with material such as hyaline; and the presence or absence of glomerulosclerosis (scarring). Although the glomerulus is the primary site of injury in glomerular disease, the tubules and the interstitium must be carefully inspected because the degree of tubulointerstitial fibrosis is the best predictor of the prognosis in renal disease.

Light microscopy is also used to classify glomerular disease as focal or diffuse. If less than 50% of glomeruli are involved by the disease process, the disease is termed focal; if 50% or more glomeruli are involved, the disease is called diffuse. If a small portion (or segment) of an individual glomerulus is involved,

the disease is described as segmental; if most of an individual glomerulus is involved, it is called global.

The presence of glomerular crescents can also be detected on light microscopy. Crescents are layers of cells (parietal epithelial cells, podocytes, lymphocytes, and macrophages) in the Bowman space, and their presence signifies severe disease.

Immunofluorescence

Immunofluorescent immunostaining determines the presence or absence of any underlying immune processes. Staining is directed against specific antibodies (e.g., IgG, IgA, and anti-GBM) and individual complement components (e.g., C3, C4, and C5b-9). The pattern of the immune components is also diagnostic. A granular pattern is typical of antigen-antibody complexes, such as in membranous nephropathy, whereas a linear pattern occurs in anti-GBM disease. The location of antibody or complement (e.g., in the mesangium in IgA nephropathy) also provides clues to the diagnosis. Immunostaining can determine the presence of matrix proteins (silver stain), amyloid fibrils (Congo red), and viral inclusions.

Electron Microscopy

Electron microscopy provides information about the presence and subcellular location of immune complexes (which are seen as electron-dense deposits), the degree of injury to glomerular cells, and the consistency of the basement membrane. Electron microscopy also detects fibrils and provides information on the ultrastructure of the kidney, such as podocyte effacement and flattening, which cannot be readily detected by light microscopy.

General Management

THERAPEUTIC PRINCIPLES

Four principles govern the treatment of glomerular disease. First, treat any acute complications of renal failure such as pericarditis; hyperkalemia; acidosis; hypertension with or without myocardial, cerebral, or other vascular bed crisis; and volume overload causing pulmonary edema. Second, use disease-specific therapeutic agents to treat any underlying primary and secondary glomerular disease. Third, to reduce the risk of further glomerular damage, use nonspecific strategies to lower the intraglomerular capillary pressure, reduce systemic blood pressure, and decrease elevated serum cholesterol levels. Fourth, anticipate and treat potential complications of nephrotic syndrome, such as hypercoagulability, infection, and skeletal abnormalities.

NONSPECIFIC TREATMENT

To reduce glomerular filtration pressure and proteinuria, give angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARBs), or both. The treatment goal is to reduce proteinuria to less than 1 g/24 hr. A 24-hour urine collection remains the gold standard for assessing proteinuria, but for patient convenience, measurement of the protein-to-creatinine ratio in a spot urine specimen can be used to follow proteinuria. In such cases, both protein and creatinine concentration are measured in mg/dl. Dividing the protein concentration by the creatinine concentration yields an estimate of the amount of protein excreted in grams per 24 hours.

Hypertension requires aggressive treatment. The goal is a blood pressure of less than 130/80 mm Hg. Hyperlipidemia is treated with a statin drug. The goal is a low-density lipoprotein (LDL) cholesterol level of less than 100 mg/dl. In addition to lowering lipid levels, statins have been shown to reduce proteinuria and to help prevent progression of glomerular disease by lipid-independent mechanisms.

Specific Glomerular Diseases

DISEASES PRESENTING AS NEPHRITIC SYNDROME

The clinical features of nephritic syndrome are caused by a severe and typically acute inflammatory injury to the glomeruli, which leads to decreased renal function. The clinical characteristics include hematuria, proteinuria, hypertension, and increased creatinine levels. Hematuria in nephritic syndrome may be macroscopic or microscopic.[15] RBCs and RBC casts in the urine, which represent an active urinary sediment, are pathognomonic for glomerulonephritis. Proteinuria in nephritic syndrome is generally less than 3.5 g/24 hr. In contrast, nephrotic syndrome is distinguished by a marked increase in glomerular permeability and more marked proteinuria (> 3.5 g/24 hr).

The major complications of nephritic syndrome are hypertension, volume overload, hyperkalemia, and metabolic acidosis; patients may present with one or more of these. The decrease in glomerular filtration in nephritic syndrome causes marked salt and water retention, leading to hypertension and edema. Acute hypertensive crisis with congestive heart failure, cerebral failure, and pulmonary edema may occur. An acute decrease in GFR can also cause hyperkalemia and metabolic acidosis.

The more common forms of glomerulonephritis that present as nephritic syndrome can be classified on the basis of their underlying disease mechanisms into three categories: immune complex mediated (e.g., postinfectious glomerulonephritis, IgA nephropathy), antibody induced (e.g., anti-GBM disease), and pauci-immune (i.e., with no immune complexes detected by immunofluorescence; for example., vasculitis associated with antineutrophil cytoplasmic antibodies [ANCAs]). The clinical course of individual glomerular diseases may vary. For example, PSGN is usually acute and self-limiting, IgA nephropathy is intermittent and chronic, and anti-GBM disease and ANCA-associated vasculitis have a rapid onset and are progressive, leading to end-stage renal failure within a short period if not treated.

Glomerular Diseases Associated with Infection

Acute glomerulonephritis can result from bacterial infections (*Streptococcus, Staphylococcus aureus, Salmonella, Escherichia coli, Legionella,* and *Mycoplasma*), viral infections (hepatitis B, hepatitis C, HIV, and cytomegalovirus), parasitic infections (*Schistosoma*, *Plasmodium*, *Trypanosoma*), and fungal infections (*Aspergillus, Histoplasma, Candida*). Renal disease associated with infection is typically immune complex related; the complexes comprise antigens from the infectious agent and host antibodies directed against the antigens.

The classic glomerular disease associated with infection is PSGN, which produces a different glomerular histologic picture than most other postinfectious forms of glomerulonephritis.[16] At present, hepatitis C[17] and HIV[18,19] are the leading causes of infection-associated glomerulonephritis in developed countries, whereas PSGN remains common in other parts of the world. Hepatitis C–associated glomerulonephritis and PSGN typically present as nephritic syndrome, whereas HIV causes nephrotic syndrome.

Hepatitis C virus–associated glomerulonephritis Glomerular disease is a common extrahepatic manifestation of chronic hepatitis C infection. Hepatitis C infection is now the major cause of essential mixed cryoglobulinemia. Patients can present with different histologic and clinical syndromes,[20] with renal disease carrying the worst prognosis. About 40% to 50% of patients with renal disease have systemic indications of cryoglobulinemia, manifesting as the triad of arthralgias, purpura, and weakness; 50% of patients have renal manifestations only, in the form of proteinuria, microhematuria, or renal dysfunction. The proteinuria may be nephrotic (> 3.5 g /24 hr) or nonnephrotic (< 3.5 g/24 hr). The majority of patients have hypocomplementemia (low C3 and C4 levels) and elevated rheumatoid factor levels (indicative of cryoglobulins). Occasionally, patients present with signs of severe vasculitis and a rapid decline in renal function. Overt liver disease is unusual, however.

The most common histologic finding in patients with hepatitis C–associated glomerulonephritis is MPGN type I with mixed cryglobulinemia, but there are also reports of MPGN type I without cryoglobulinemia in these patients [*see* Membranoproliferative Glomerulonephritis, *below*]. MPGN is characterized by deposits of preformed hepatitis C antigen-antibody complexes that are trapped in the glomerular subendothelial space and mesangial cells, which causes glomerular cell proliferation. The mesangium increases in size and interposes into the GBM, causing basement membrane splitting (hence the term membranoproliferative). Less frequently, hepatitis C causes membranous nephropathy, fibrillary glomerulonephritis, FSGS (in African Americans) and thrombotic microangiopathy (in renal transplant recipients).

Treatment of hepatitis C–associated MPGN is directed against the virus and consists of polyethylene glycol (PEG) interferon alpha-2b and ribavirin (if renal function allows) to reduce the viral burden. Severe acute disease is treated with plasmapheresis to remove cryoglobulins, plus immunosuppressive therapy with corticosteroids and cyclophosphamide, followed by initiation of antiviral therapy.

Poststreptococcal glomerulonephritis PSGN is typically a self-limiting disease that develops after a pharyngeal infection or skin infection (e.g., impetigo or pyoderma) with so-called nephritogenic strains of group A streptococci. Occasionally, cases have been associated with group C streptococci. PSGN most often affects children 2 to 10 years of age and is twice as common in males as in females. The disease is usually sporadic but can occur in epidemics.

PSGN is a classic immune complex–mediated form of glomerulonephritis.[21] The glomerular injury is induced when circulating preformed immune complexes (comprising the streptococcal antigen and an antibody directed against it) are deposited in the mesangium and subendothelium. Two streptococcal antigens are most likely involved in the pathogenesis of PSGN: plasmin receptor protein (related to preabsorbing antigen and endostreptosin) and streptococcal zymogen (exotoxin B). Antibodies to these antigens correlate with clinical disease; both antigens have been localized in glomeruli in a large percentage of patients. The formation of immune complexes explains the typical 2- to 3-week delay between streptococcal infection and the onset of PSGN.

On light microscopy, PSGN is characterized by diffuse glomerular involvement [*see Figure 2*]. The histologic hallmark of PSGN is glomerular hypercellularity, which is caused by the proliferation of resident mesangial and endothelial cells and an influx of infiltrating cells such as neutrophils. Severe PSGN, which occurs in less than 5% of patients, is marked by crescent formation in the Bowman space. The clinical presentation of patients with crescents on renal biopsy is typically that of RPGN (see below).

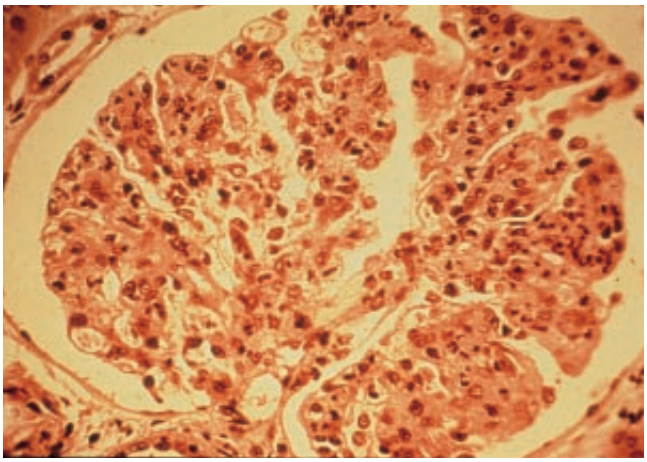


Figure 2 **Light micrograph of poststreptococcal glomerulonephritis showing increased glomerular hypercellularity.**

There are three patterns of immunofluorescent staining in PSGN; these are termed starry sky, garland, and mesangial. The classic starry-sky pattern of immune deposition is seen in about 30% of PSGN cases and occurs early in the disease. It consists of fine granular deposits of IgG and C3, without C1q and C4, on all capillary walls and the mesangium. The garland pattern, which consists of coarse granular deposits of large amounts of IgG and C3 along the capillary loops, with relatively few mesangial deposits, also is found in about 30% of patients. The mesangial pattern of IgG deposition, which is often associated with mesangial cell proliferation, occurs in about 40% of patients.

Electron microscopy in PSGN shows multiple large, dome-shaped, immune complex deposits that are referred to as subepithelial humps because of their location at the base of the podocyte [*see Figure 3*].[22] The differential diagnosis of subepithelial immune deposits includes membranous nephropathy, lupus nephritis, and MPGN type III [*see* Membranoproliferative Glomerulonephritis, *below*]. Biopsies taken early in the course of PSGN may show immune complex deposits in the mesangium and subendothelium.

Although PSGN presents as an acute disease, there is a latent phase between the initiating pharyngitis (mean, 10 days) or skin infection (mean, 21 days) and the onset of the nephritic syndrome.[16] Patients present with pronounced edema, often of the upper body; diminished urine output; dark or smoky urine containing RBCs and RBC casts; and signs and symptoms of hypertension.

The serum creatinine level is often normal; a rise in creatinine reflects underlying acute renal failure. Proteinuria is typically less than 500 mg/day during the acute illness. Throat and skin cultures are usually negative, but serologic measures of recent streptococcal infection, such as antistreptolysin O, are elevated and remain so for 3 to 6 months. Antistreptolysin O titers may be blunted by recent antibiotic usage and are less marked after skin infections. Because PSGN is induced by immune complexes via the alternative pathway, patients demonstrate marked decreases

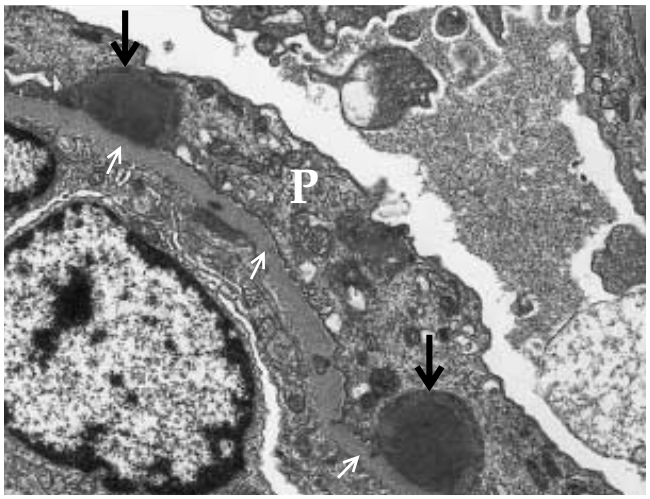


Figure 3 **Transmission electron micrograph (magnification: × 10,500) showing characteristic subepithelial hump-shaped deposits (illustrated by thick arrows) in postinfectious glomerulonephritis lying below the podocytes (P); small arrows indicate the glomerular basement membrane.**

in levels of CH50 and C3, but not of C1q, C2, or C4, which are activated by the classical complement pathway [*see 6:II Innate Immunity*]. Depression of the C3 level that lasts longer than 8 weeks should suggest another cause.

There is no role for antibiotic therapy in PSGN, because the inciting infection has typically occurred weeks earlier and the antigen-antibody complexes have already formed. Treatment is therefore supportive. The majority (> 95%) of patients recover normal renal function; the remainder, however, develop end-stage renal failure. Although macroscopic hematuria and edema usually resolve within 2 weeks, microscopic hematuria may persist for a year; 20% of patients develop nephrotic-range proteinuria, typically during the recovery phase of the disease, that resolves spontaneously. However, proteinuria in the nonnephrotic range can last as long as 2 years. Recurrent PSGN is exceedingly rare.

Other forms of postinfectious glomerulonephritis Bacterial infections (e.g., infective endocarditis,[23] infected shunts,[24] and visceral abscesses[25]), as well as certain viral infections (e.g., hepatitis B and hepatitis C[26]), can cause MPGN, which is associated with alternative-pathway hypocomplementemia characterized by low C3 and normal C4 levels. Antibiotic therapy eliminates infection in endocarditis and allows for resolution of the immune response. A persistently low C3 level suggests that infection is not cleared.

IgA Nephropathy

IgA nephropathy derives its name from the glomerular deposition of IgA, which causes an injurious inflammatory response.[27] IgA nephropathy is the most prevalent pattern of glomerulonephritis found in countries where renal biopsy is widely used as an investigative tool.[28] IgA nephropathy is particularly prevalent in Asia and southern Europe but is less common in the United States and Canada, which suggests a possible genetic predisposition to the disease. IgA nephropathy affects more males than females (3-to-1 ratio) and occurs most commonly in children and young adults. Although most cases of IgA nephropathy are idiopathic, a number of diseases have been associated with secondary IgA nephropathy [*see Table 2*], and hereditary forms have also been described.

Pathogenesis and histology Mesangial cell injury in IgA nephropathy results from the deposition of IgA on mesangial cells [*see Figure 4*].[29] The tendency for flares of macroscopic or microscopic hematuria to occur within 24 hours after a mucosal infection (e.g., upper respiratory tract or bowel) indicates that the glomerular disease may be initiated by an impaired mucosal IgA response to infection. However, studies have suggested that IgA derived from bone marrow may be increased in these patients and that hepatic clearance of IgA may be decreased, which may in part reflect abnormal glycosylation.[30]

The glomerular lesions that follow mesangial cell IgA deposition are variable.[29] The characteristic lesion is mesangial cell proliferation [*see Figure 5*], which appears on light microscopy as mesangial cell hypercellularity [*see Figure 6*]. Glomerular involvement may be focal or diffuse. In severe cases that progress to end-stage renal failure, mesangial scarring is noted. Glomerular crescents form occasionally, and as in most forms of glomerular disease, tubulointerstitial fibrosis is present in those patients with renal failure.

Immunofluorescent staining reveals IgA [*see Figure 4*] and, occasionally, IgG and IgM in the mesangium; C3 is often detected without C1q and C4, suggesting activation of the alternative complement pathway. Mesangial immune deposits can occasionally be seen as electron densities on electron microscopy.

Clinical and laboratory findings IgA nephropathy has six different clinical-presentation patterns[31]: episodic macroscopic hematuria, asymptomatic microscopic hematuria and nonnephrotic-range proteinuria, nephrotic syndrome, acute renal failure, chronic renal failure and hypertension, and concomitant systemic disease.

Episodic macroscopic hematuria is the classic presentation of IgA nephropathy. It occurs in 40% to 50% of patients. The hematuria is often characterized by RBC casts and occurs within 24 hours after a mucosal infection, commonly of the upper respiratory tract. This presentation is also called synpharyngitic hematuria. Episodic macroscopic hematuria differs from the delayed-onset hematuria that appears 2 to 3 weeks after pharyngitis in PSGN. Macroscopic hematuria usually resolves spontaneously

Table 2 Diseases Associated with IgA Deposits

- Dermatologic diseases
  - Dermatitis herpetiformis
  - Psoriasis
  - Reiter syndrome
- Gastrointestinal diseases
  - Celiac disease
  - Crohn disease
  - Ulcerative colitis
- Hepatic diseases
  - Alcoholic liver diseases
  - Hepatic cirrhosis
- Neoplastic diseases
  - Bronchial carcinoma
  - IgA monoclonal gammopathy
  - Mycosis fungoides
  - Sézary syndrome
- Rheumatic diseases
  - Ankylosing spondylitis
  - Rheumatoid arthritis
- Unclassified
  - Diabetes
  - Idiopathic pulmonary hemosiderosis
  - Properdin deficiency
  - Retroperitoneal sclerosis
  - Sarcoidosis

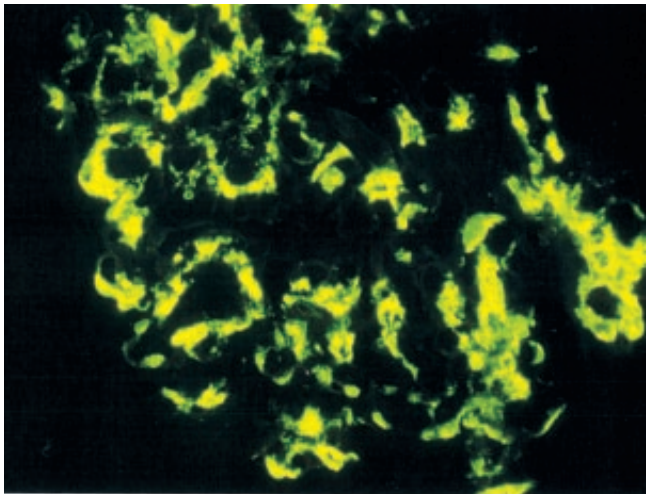


Figure 4 **IgA immunostaining showing IgA deposition in a typical mesangial cell distribution in IgA nephropathy.**

within days, and microscopic hematuria often occurs between mucosal infections.

Asymptomatic microscopic hematuria and nonnephrotic-range proteinuria constitute 30% to 40% of IgA nephropathy; nephrotic syndrome constitute approximately 5%. Acute renal failure, which constitutes less than 5%, can occur in association with crescents (i.e., crescentic IgA nephropathy) or, possibly, as a result of acute tubular necrosis from excessive hematuria. In rare instances, IgA nephropathy presents as part of a systemic disease [*see Table 2*].

Laboratory investigations are usually not helpful in diagnosing IgA nephropathy. Serum IgA levels are increased in about a third of patients, and IgA-fibronectin aggregates are increased in about half. Mesangial deposition of C3 is found in more than half of patients. Serum complement levels are usually normal. Taken together, the diagnosis is clinical, supported by a renal biopsy.

The differential diagnosis of persistent isolated microscopic glomerular hematuria includes hereditary glomerulonephritis (Alport syndrome), postinfectious glomerulonephritis, thin basement membrane disease, SLE, FSGS, and minimal change disease. A major consideration in the differential diagnosis of IgA nephropathy is Henoch-Schönlein purpura (HSP). HSP is a systemic vasculitis that is characterized by deposition of IgA-containing immune complexes in tissues, including mesangial cells, and thus shares many features with IgA nephropathy.[32] HSP is manifested clinically by a classic tetrad of purpuric rash (typically over the extremities), arthralgias (in knees and ankles), abdominal pain, and renal disease. Although HSP is more common in children younger than 5 years, renal involvement is more common in older children and adults. Like IgA nephropathy, HSP often follows an upper respiratory tract infection. A typical urinary finding is mild proteinuria and an active urinary sediment (containing RBCs and RBC casts), which is often transient. Most patients are asymptomatic, with a normal or slightly elevated serum creatinine level, but nephrotic syndrome develops in 20% of cases. The diagnosis of HSP can often be made clinically. Confirmation requires evidence of tissue deposition of IgA on either a skin biopsy (which shows a leukocytoclastic vasculitis) or renal biopsy (which shows an IgA immunostaining in a typical mesangial distribution). HSP is discussed in detail elsewhere [*see 10:VII Vascular Diseases of the Kidney*].

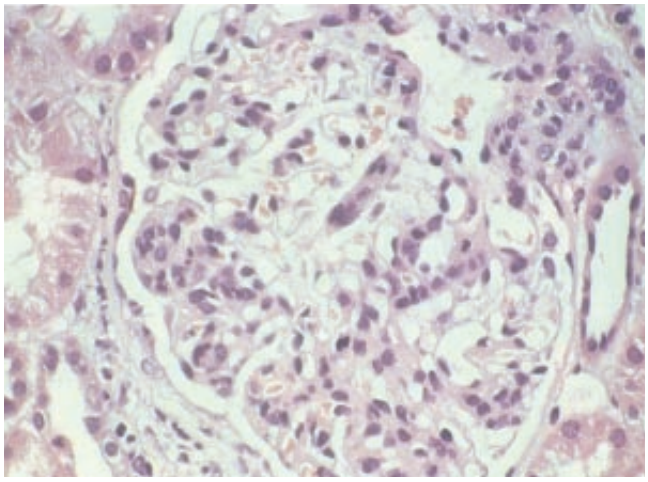


Figure 5 **Light micrograph from a patient with IgA nephropathy showing the characteristic glomerular hypercellularity.**

Treatment Although a variety of treatment strategies have been tried, therapy for IgA nephropathy remains disappointing.[33] Several studies suggest that the use of ACE inhibitors, even in normotensive patients, may slow progression by lowering glomerular filtration pressure and reducing proteinuria.[34-36] Combining ACE inhibitors with ARBs may enhance the antiproteinuric effect and further slow the progression of renal disease.[37] In patients with aggressive or progressive IgA nephropathy, slowing or arresting of disease progression may be accomplished with the combination of cytotoxic agents (oral cyclophosphamide, 1.5 mg/kg daily for 3 months, then converted to azathioprine) and prednisolone.[38] Clinical trials evaluating the efficacy of fish oil have produced conflicting results, and a meta-analysis concluded that a minor benefit was likely.[39]

All patients with IgA nephropathy should receive treatment with an ACE inhibitor, an ARB, or both. Corticosteroid therapy for 18 to 36 months may be associated with less proteinuria and better outcome.[40-43] However, studies showing a benefit from corticosteroid therapy were performed before the current aggressive use of angiotensin antagonists. A subset of proteinuric patients without hematuria (usually with nephritic syndrome and minimal glomerular changes on light microscopy and effacement of the foot processes on electron microscopy) seem to respond very well to steroid treatment. Some experts recommend steroid treatment only for patients with nephrotic syndrome and minimal apparent changes on histology; progressive active disease despite the use of ACE inhibitors, ARBs, or both; or severe disease on biopsy.[44]

Prognosis Most patients with IgA nephropathy have an indolent course, with chronic intermittent hematuria. Risk factors for disease progression include an elevated serum creatinine level and elevated blood pressure on presentation (although these may instead reflect severe disease), persistent proteinuria (> 1 g/day), older age, and interstitial fibrosis on renal biopsy. For reasons that are still unclear, recurrent macroscopic hematuria is often associated with a better prognosis than persistent microscopic hematuria and proteinuria. Serum IgA levels and gender are not prognostic factors. IgA nephropathy is a significant cause of end-stage renal failure, with 20% of patients requiring renal replacement therapy within 20 years of presentation.

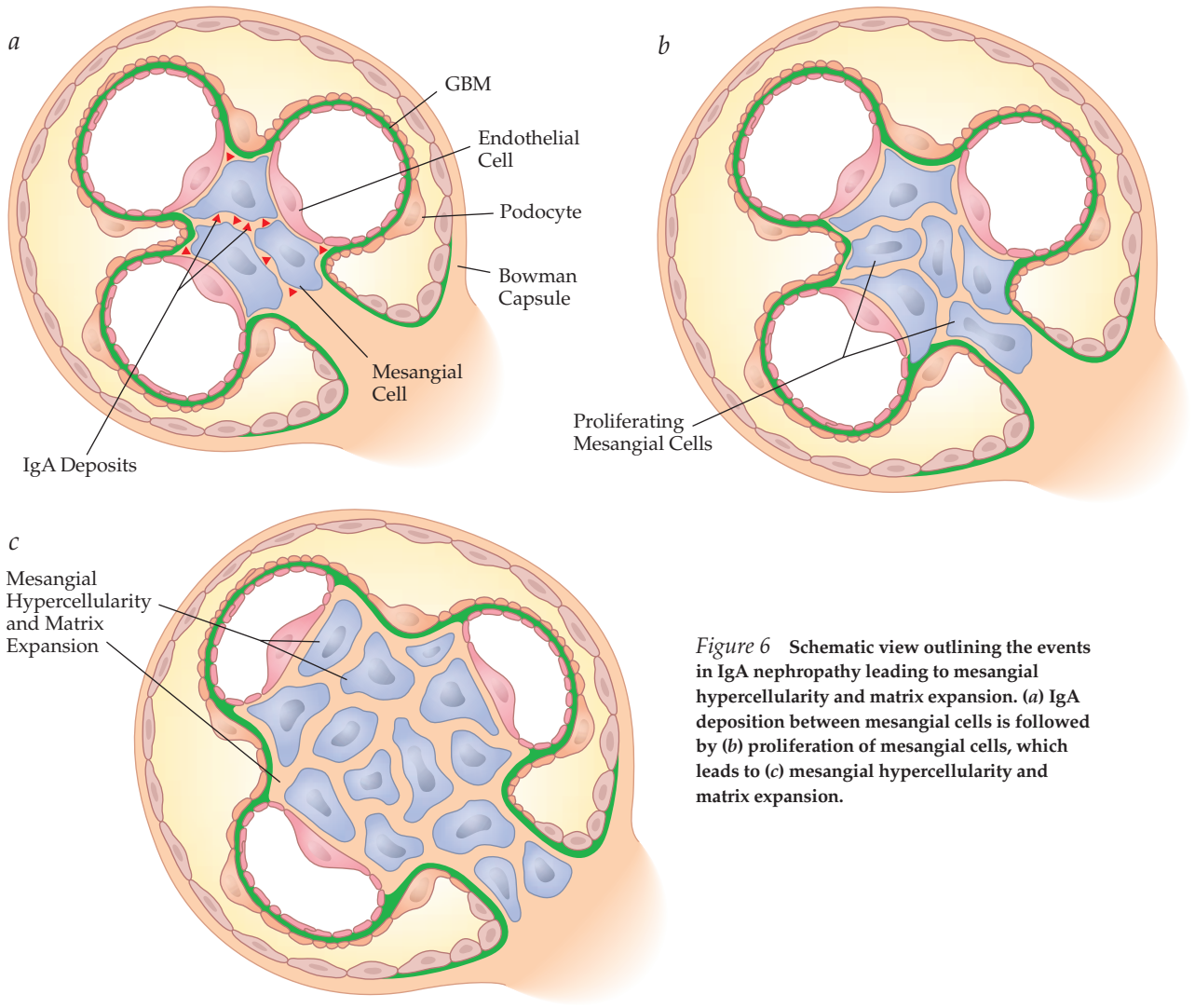


Figure 6 **Schematic view outlining the events in IgA nephropathy leading to mesangial hypercellularity and matrix expansion. (*a*) IgA deposition between mesangial cells is followed by (*b*) proliferation of mesangial cells, which leads to (*c*) mesangial hypercellularity and matrix expansion.**

Anti–Glomerular Basement Membrane Disease

Epidemiology Anti-GBM disease is rare, with an estimated annual incidence of 0.5 to 0.9 cases per million population in whites and an even lower incidence in other races. Most patients are male and in either the second to third or the sixth to seventh decades of life, but the disease has been reported in both men and women and at all ages. Exposure to pulmonary toxins, such as hydrocarbons, increases the incidence of anti-GBM disease.

Pathogenesis and histology Anti-GBM disease is caused by a nephritogenic antibody directed against a specific epitope on the noncollagenous (NC1) domain of the α3 chain of type IV collagen in the GBM.[9,45] Type IV collagen also occurs in the alveolar basement membrane, so these antibodies may also result in pulmonary involvement. Anti-GBM disease is the term used if the disorder is restricted to the kidney (primary renal disease); Goodpasture syndrome refers to the constellation of glomerulonephritis, pulmonary hemorrhage, and anti-GBM antibodies[46] [*see 10:VII Vascular Diseases of the Kidney*].

The characteristic finding on renal biopsy is a diffuse necrotizing proliferative lesion (with increased cellularity because of proliferation) and crescents (layers of cells in the Bowman space comprising parietal epithelial cells, podocytes, macrophages, and leukocytes) in more than 50% of glomeruli [*see Figure 7*]. Although the initial injury occurs within the glomerular tuft, breaks in the GBM allow plasma products to leak from the capillaries into the Bowman space [*see Figure 8*]. The early cellular and later fibrocellular crescents compress the glomerular tuft [*see Fig-*

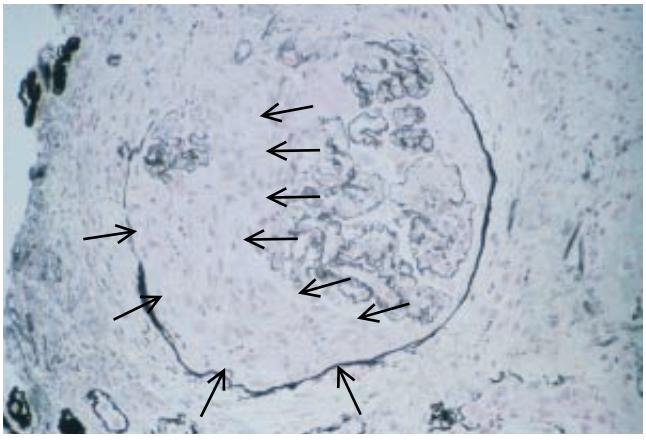


Figure 7 **A large multilayer of cells in the Bowman space (arrows) represents a typical crescent in anti–glomerular basement membrane disease.**

ure 8], so disease severity is related in part to the extent of crescents. Pathognomonic findings on renal biopsy in both anti-GBM disease and Goodpasture syndrome are linear deposition of IgG and C3 along the GBM [*see Figure 9*]. These findings are in contrast with the granular pattern of IgG immunostaining in immune complex glomerulonephritis (i.e., PSGN, IgA nephropathy, MPGN, and lupus nephritis).

Clinical and laboratory findings The typical presentation of anti-GBM disease consists of a rapid decline in renal function in a patient with a so-called active urinary sediment (i.e., containing RBCs and RBC casts and, occasionally, white blood cell [WBC] casts) and mild proteinuria, with or without the complications of pulmonary hemorrhage. There is a direct correlation between the initial plasma creatinine concentration and the percentage of glomeruli with crescents; crescents are present in more than 75% of glomeruli when plasma creatinine concentration is above 5 mg/dl.

Patients with pulmonary-renal syndrome (exemplified by Goodpasture syndrome) present with pulmonary hemorrhage, acute renal failure, and anemia; these patients may have a history of viral infection or exposure to pulmonary toxins such as cigarette smoke or hydrocarbons.[47] Blood loss from pulmonary hemorrhage often leads to iron deficiency anemia.

The diagnosis is confirmed by an increase in serum anti-GBM antibody titers. Serum complement levels are normal, and ANCA assays are positive in 20% to 30% of patients. Despite the presence of ANCAs in some patients, anti-GBM disease is not classified as a renal vasculitis.

Treatment and prognosis Patients with anti-GBM disease require urgent therapy to prevent a progressive decline in renal

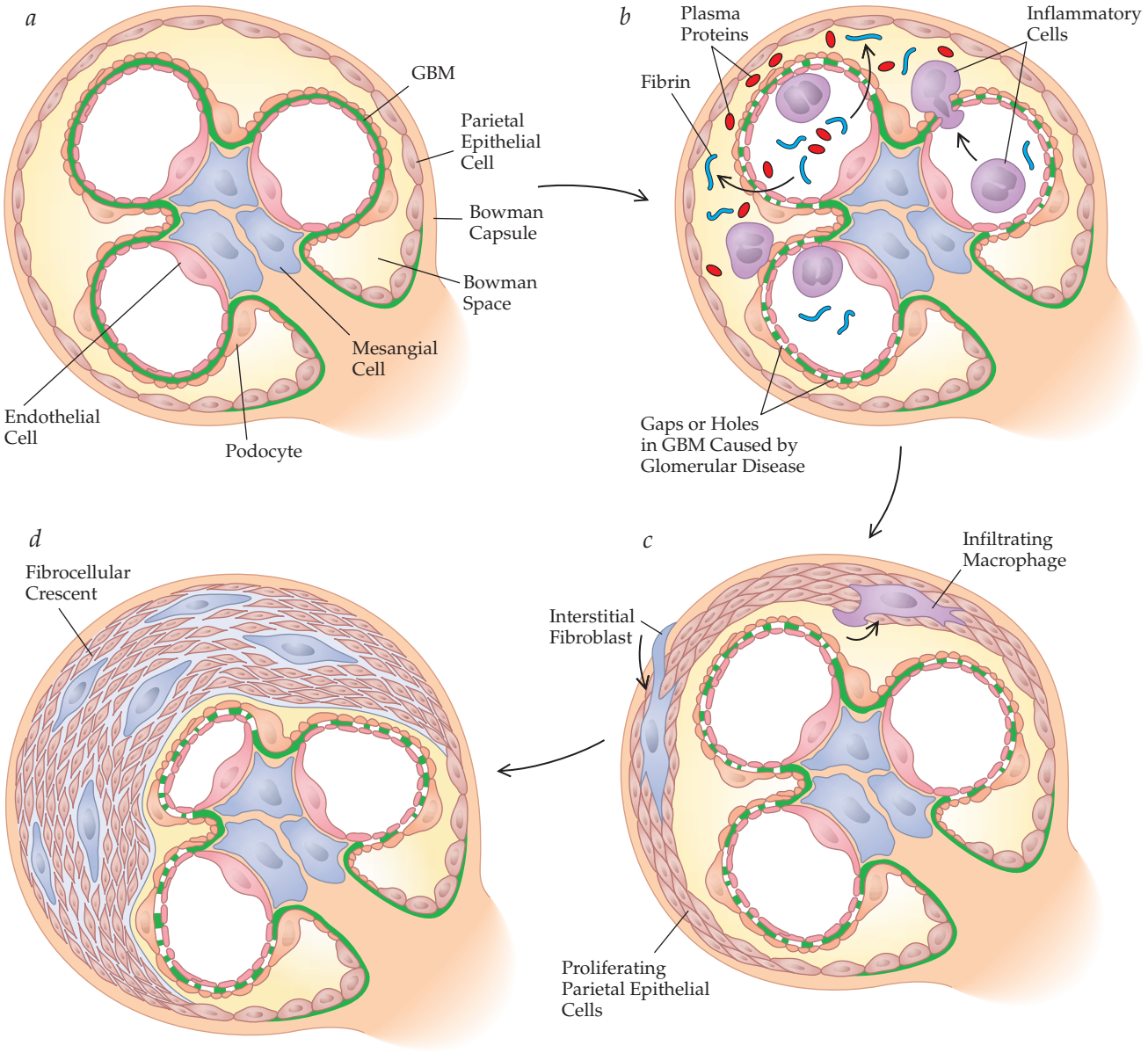


Figure 8 **Schematic view illustrating the events leading to crescent formation. (*a*) Normal glomerulus. (*b*) Glomerular disease leads to gaps or holes in the glomerular basement membrane (GBM), resulting in (*c*) proliferation of parietal epithelial cells and ultimately (*d*) fibrocellular crescent formation.**

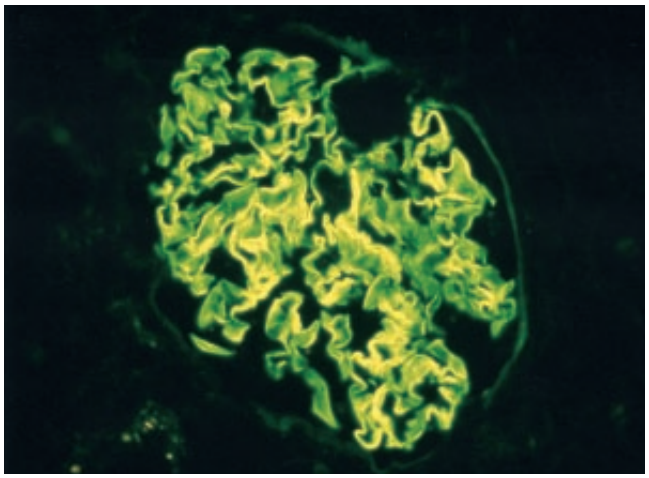


Figure 9 **Linear pattern of IgG immunostaining along the GBM in a patient with anti-GBM disease.**

function. The treatment of choice is triple therapy, consisting of plasmapheresis plus prednisone and cyclophosphamide. The rationale is that plasmapheresis removes circulating anti-GBM antibodies and other mediators of inflammation, whereas prednisone and cyclophosphamide minimize new antibody formation. Pulse steroids (methylprednisone, 30 mg/kg or 1,000 mg I.V.) are given for 3 days, followed by high-dose oral prednisone (1 mg/kg/day) given for several months, with plasma exchanges (4 L/day) given daily or on alternate days until the anti-GBM antibody is undetectable.[48,49] Cyclophosphamide is given orally at a dosage of 2 mg/kg/day. Therapy is generally continued for 9 to 12 months.

More than 70% of treated patients with anti-GBM disease recover renal function. Oliguria and a serum creatinine level greater than 6 mg/dl are poor prognostic findings; patients with this high a creatinine level are generally considered not to be candidates for triple therapy. Renal transplantation should be avoided until the anti-GBM antibody has been undetectable for 12 months.

Renal and Systemic Vasculitis

The kidney is often involved in systemic vasculitis; examples include microscopic polyangiitis, Wegener granulomatosis, Churg-Strauss syndrome, and isolated pauci-immune crescentic glomerulonephritis. These diseases are discussed in detail elsewhere [*see 10:VII Vascular Diseases of the Kidney and 15:VIII Systemic Vasculitis Syndromes*].

Rapidly Progressive Glomerulonephritis

RPGN, an acute nephritic syndrome, is distinguished from other forms of nephritic syndrome by its clinical presentation, glomerular histology, and etiology [*see Figure 10*].[50] RPGN is diagnosed clinically as a doubling in serum creatinine level over 3 months because of an underlying glomerulonephritis.[50] Frequently, renal function deteriorates over as little as days or weeks. Most authorities also require the presence of glomerular crescents on renal biopsy to confirm the diagnosis. Typically, if more than 50% of glomeruli contain crescents, renal function decreases acutely (i.e., over days to weeks). RPGN constitutes a medical emergency.

RPGN can occur as a primary renal disorder or as a consequence of systemic disease. Conventionally, RPGN is classified according to the underlying disease mechanism into one of three categories: 50% of cases are from pauci-immune disease, which is often associated with ANCAs; 30% are from immune complex disease (e.g., postinfectious glomerulonephritis, lupus nephritis, IgA nephropathy, and MPGN); and 20% of cases are from anti-GBM.[51]

DISEASES PRESENTING AS NEPHROTIC SYNDROME

Nephrotic syndrome is characterized by heavy proteinuria (> 3.5 g/24 hr), hypoalbuminemia (< 3 g/dl), peripheral edema, hyperlipidemia (elevated total and LDL cholesterol levels) and lipiduria (oval fat bodies in urinary sediment). Nephrotic syndrome results from a marked increase in glomerular permeability to protein and other macromolecules. Although glomerular injury is often severe in diseases causing nephrotic syndrome, it is unusual for patients to present with a decrease in renal function early in the course of disease. A rise in the serum creatinine level is usually a feature of more advanced disease. The urinary sediment is typically devoid of red and white cells or casts and is termed an inactive urinary sediment. This contrasts with the active urinary sediment in nephritic syndrome. However, some patients who present with a combined nephritic-nephrotic picture will have an active urinary sediment. In nephrotic patients with a marked increase in serum lipid levels, fat and cholesterol casts can be seen in the urine because of markedly increased glomerular permeability.

The hallmark of the nephrotic syndrome is proteinuria greater than 3.5 g/24 hr. It is very unusual for proteinuria of this magnitude to be caused by conditions other than glomerular disease. In the absence of a 24-hour urine collection, a spot urine collection can be used to estimate the degree of proteinuria.[52] A normal urinary protein-to-creatinine ratio (mg/mg) is less than 0.15, which indicates a protein excretion of less than 150 mg/day; a ratio greater than 3.5 is consistent with excretion of more than 3.5 g of protein, which defines nephrotic syndrome. A urine dipstick detects only albumin, not other proteins, and is positive only when protein excretion exceeds 300 to 500 mg/day. Caution must also be used when interpreting dipstick units, however, because urine tonicity or specific gravity affects the measurement. In certain circumstances (e.g., Bence Jones proteinuria in multiple myeloma), the dipstick will be negative for protein.

Lesser degrees of proteinuria (< 3.5 g/24 hr) may result from glomerular or nonglomerular causes. These include heart failure; renal vascular disease; increased GFR from exercise, pregnancy, diabetes, or fever; tubulointerstitial disorders; increased protein delivery from myeloma; and orthostatic proteinuria (a benign condition in which low-grade proteinuria occurs only when the patient is not recumbent).

Although nephrotic syndrome is caused by an underlying glomerular disease, progressive renal failure is caused by tubulointerstitial inflammation and scarring that results from ongoing proteinuria. Thus, the renal biopsy must also be evaluated for the presence of tubulointerstitial fibrosis.

The etiology of nephrotic syndrome is separated into primary and secondary causes. The most common primary glomerular diseases that cause nephrotic syndrome in adults are FSGS, membranous nephropathy, minimal change disease, and MPGN [*see Table 1*]. Although these disease entities are usually primary (and idiopathic), each can also be secondary to an underlying systemic disease. However, the glomerular histology is indistinguishable in primary and secondary forms. Diabetes mellitus and amyloidosis are the leading causes of secondary nephrotic syndrome.

The target of injury in FSGS, membranous nephropathy, minimal change disease, and even diabetes is the podocyte [*see Figure 1b*]. A critical function of the podocyte, which sits on the outer aspect of the GBM, is to prevent proteinuria. Thus, the characteristic clinical consequence of podocyte damage is massive proteinuria. Moreover, because podocytes are located on the outer aspect of the GBM and therefore not in close proximity to the capillary loops, podocyte injury in nephrotic syndrome is typically not associated with red or white cells or cell casts in the urine.

Therapeutic Principles in Nephrotic Syndrome

Four principles guide the management of the nephrotic syndrome. First, treat any complications: reduce fluid overload with diuretics, provide anticoagulation therapy for patients at high risk for venous thrombosis (e.g., those with a serum albumin level < 2.0 g), aggressively treat hypertension (blood pressure goal < 130/80 mm Hg), and decrease hyperlipidemia with statins. Second, lower proteinuria to less than 1 g/24 hr, ideally through combination therapy with ACE inhibitors and ARBs. Third, use disease-specific therapy when possible. Fourth, treat the underlying secondary cause if one is present.

Complications of Nephrotic Syndrome

Major complications of nephrotic syndrome include hypoalbuminemia, edema, hypercholesterolemia, and hypercoagulability. Although the liver is capable of synthesizing large amounts of albumin, proteinuria is associated with increased tubular uptake of protein. In the tubules, protein is catabolized rather than recycled for further use.[53] As a result, serum albumin levels decrease. Hypoalbuminemia decreases the plasma oncotic pressure, resulting in a decrease in effective circulating volume. This is followed by an increased proximal tubular sodium resorption because of increased angiotensin II, aldosterone, and epinephrine. A concomitant increase in antidiuretic hormone levels results in fluid retention, which manifests clinically as edema and hypertension. Hypoalbuminemic nephrotic patients are often resistant to loop diuretics such as furosemide, because loop diuretics are inactivated by binding to albumin in the tubular lumen. Edema also appears to result from a primary defect in sodium excretion that leads to an expanded plasma volume, followed by transudation of fluid.

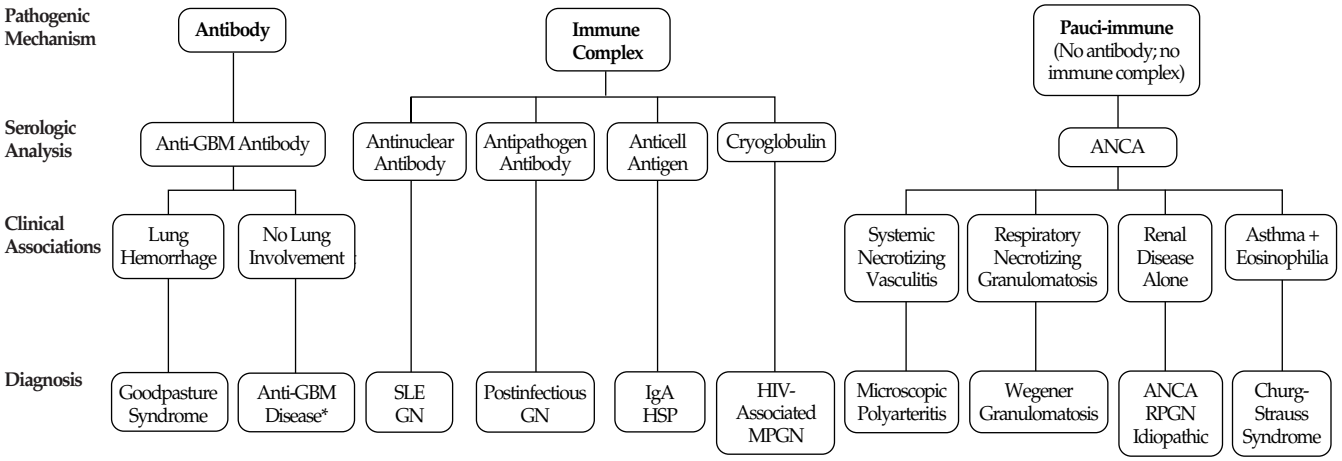
Hepatic cholesterol and lipid synthesis are increased in nephrotic patients, probably in response to decreased oncotic pressure.[54] Moreover, there is decreased catabolism, which explains in part the increase in levels of very low density lipoprotein. Hyperlipidemia is a major risk factor for cardiovascular disease in nephrotic patients. It may also contribute to the progression of glomerular diseases. Treatment of hyperlipidemia should be aggressive and includes cholesterol-lowering agents such as statin drugs [*see 9:II Diagnosis and Treatment of Dyslipidemia*].

Patients with nephrotic syndrome are in a hypercoagulable state: 10% to 40% will experience arterial or venous thromboembolism.[55] Factors that underlie increased coagulation include urinary losses of the anticoagulant antithrombin, increased platelet activation, and alterations in the fibrinolytic system. Patients may present with lower extremity, pulmonary, or renal vein thrombosis. Renal vein thrombosis complicates all forms of the nephrotic syndrome—especially membranous nephropathy—and may present acutely as a sudden decrease in renal function, loin pain, hematuria, or even systemic emboli. However, thrombosis of the renal vein may also be asymptomatic, especially if collateral veins have been established. Diagnosis is made by renal vein Doppler sonography, CT angiography, or MR angiography. Treatment includes anticoagulation therapy with heparin followed by warfarin. Many authorities use prophylactic low-dose warfarin when the plasma albumin concentration is less than 2 g/dl [*see 5:XIV Thrombotic Disorders*].

Other complications of the nephrotic syndrome include an increased susceptibility to infection (caused by urinary losses of IgG), bone disease (caused by loss of vitamin D), and loss of trace elements.

Focal Segmental Glomerulosclerosis

FSGS is the leading cause of nephrotic syndrome in adults, responsible for 35% of all cases and more than 50% of cases in



* May be associated with ANCA (~30% of cases).

ANCA—antineutrophil cytoplasmic antibody GBM—glomerular basement membrane GN—glomerulonephritis HSP—Henoch-Schönlein purpura MPGN—membranoproliferative GN RPGN—rapidly progressive GN SLE—systemic lupus erythematosus

Figure 10 **Rapidly progressive glomerulonephritis (RPGN) is classified by the underlying disease mechanism. Specific serologic tests determine whether RPGN is caused by an antineutrophil cytoplasmic antibody (ANCA)–negative or ANCA-positive glomerular disease. Clinical findings further differentiate each disease entity.**

Table 3 Causes of Focal Segmental Glomerulosclerosis (FSGS)

Primary
Idiopathic
Minimal change disease
Collapsing FSGS
Hereditary
Secondary
Infection (e.g., HIV)
Drugs (e.g., heroin)
Obesity
Reduced nephron number (e.g., reflux nephropathy, unilateral renal agenesis/congenital renal dysplasia, surgical reduction, sickle cell disease, or any advanced renal disease causing reduced nephron number)
Collapsing FSGS

African Americans. The name derives from the histologic findings: less than 50% of glomeruli are involved (focal); and within the involved glomeruli, a segment is scarred (sclerotic).

Pathogenesis and histology FSGS may be primary, hereditary, or secondary [*see Table 3*]. Except in so-called collapsing FSGS (see below), the histologic findings are the same regardless of the cause. Because FSGS typically begins at the corticomedullary junction and progresses to the outer cortex, an inadequate renal biopsy may miss the lesion. Although sclerosis is visible only in segments of a minority of glomeruli, all glomeruli leak protein and have foot process fusion (a primary alteration in the podocyte).

The mechanisms of FSGS depend on the underlying cause. Hereditary FSGS results from mutations in the genes that code for podocyte proteins. The structure and function of several podocyte proteins have been identified and characterized. The slit diaphragm, the major size barrier to molecules on the outer side of the GBM, comprises a complex and active network of proteins called nephrin, podocin, and CD2AP. These proteins signal and communicate with the rich actin cytoskeleton of the podocyte, which is also regulated by the actin-binding protein, α-actinin 4. Abnormalities of any of these proteins can lead to abnormal function and, therefore, to nephrotic syndrome, which is usually characterized by the histologic lesion of FSGS.

In idiopathic and hereditary FSGS, the primary target of injury is the podocyte, and the histologic findings are secondary to this injury. There is evidence that a circulating glomerular permeability factor plays a role in the genesis of the podocyte injury in idiopathic FSGS, but the exact nature of this role needs to be determined.[56] In contrast, in secondary forms of FSGS, podocyte injury is a consequence of the underlying disease. For example, secondary injury to podocytes from increased intraglomerular pressure [*see Figure 11*] leads to progressive podocyte dysfunction and ultimately to cell loss. Obesity is an increasing cause of FSGS, but it is usually not associated with proteinuria in the nephrotic range.

Light microscopy in glomerulosclerosis shows collapse of the glomerular capillaries [*see Figure 12*], which are filled with proteinaceous material. It is hypothesized that injured podocytes detach from the GBM, which results in denudation of the underlying GBM.[57] This promotes the formation of adhesions and synechial attachments between the glomerular tuft and the Bowman capsule, which may be accompanied by mesangial hypercellularity and, in severe cases, by tubulointerstitial fibrosis.

Because FSGS is not an immune-mediated disease, immunofluorescent staining is typically negative for IgG and complement components. However, some IgM and C3 may be found in the sclerotic areas as a consequence of passive trapping.

Electron microscopy shows fused, effaced, and flattened foot processes. One distinguishing feature between idiopathic and secondary FSGS is that foot process fusion is diffuse in the former but localized to the site of injury in the latter.

In addition to the classic histologic findings, there is a histologic variant known as collapsing or cellular FSGS.[58] This entity occurs in HIV infection, but it also has been described with the use of bisphosphonates.[59] In contrast to classic FSGS, collapsing FSGS is associated with podocyte proliferation and a marked collapse of the glomerular tuft. The decline in renal function in patients with collapsing FSGS occurs more rapidly than in patients with classic FSGS.

Clinical and laboratory findings Approximately 90% of children and 70% of adults with FSGS present with clinical and laboratory findings of the nephrotic syndrome. It is unclear why the incidence and severity of FSGS is increased in African Americans. Approximately 50% of adults present with hypertension and hematuria and 30% present with renal insufficiency.[60] As in most diseases involving the podocyte, urinalysis in FSGS typically shows no red, white, or tubular cell casts; however, hematuria is present in some patients. There are no laboratory findings specific for the diagnosis of idiopathic FSGS. Specific tests, such as those for HIV infection, may assist in ruling out secondary causes. Although studies have shown that a permeability factor is increased in many patients, an assay for this factor is not available commercially.[61]

Treatment and prognosis As in all glomerular diseases, treatment of FSGS is directed toward any underlying diseases and toward the complications of the nephrotic syndrome [*see Figure 13*]. Disease-specific therapy should be offered to patients with nephrotic-range proteinuria and to those whose serum creatinine level is below 3 mg/dl.[62] No disease-specific therapy is indicated for patients with nonnephrotic-range proteinuria. It is important to stress that FSGS caused by mutations in podocyte-specific proteins is steroid resistant. However, the tests for diagnosing these mutations are not widely available.

The first line of disease-specific therapy in adults with primary FSGS is prednisone (1 mg/kg/day) for 3 to 4 months. To minimize toxicity, many authors now recommend 2 mg/kg of prednisone every other day. In the 55% of patients who respond, prednisone should then be changed to 0.5 mg/kg/day for 8 weeks (or 1 mg/kg on alternate days) and tapered over the next 6 weeks. The duration of therapy is critical in preventing relapses. In the 45% of patients who are steroid resistant and in steroid-responsive patients who encounter relapses, the next line of therapy includes cyclosporine at a dosage of 7 mg/kg/day in divided doses, aiming for trough levels between 125 and 225 μg/L.[63] Patients who experience frequent relapses and those who are steroid dependent may benefit from cyclophosphamide (2 mg/kg/day) for 3 months and high-dose prednisone (1 mg/kg/day) for 1 month, followed by a steroid taper. Initiating therapy with cytotoxic agents and prednisone has no added benefit over therapy with prednisone alone.

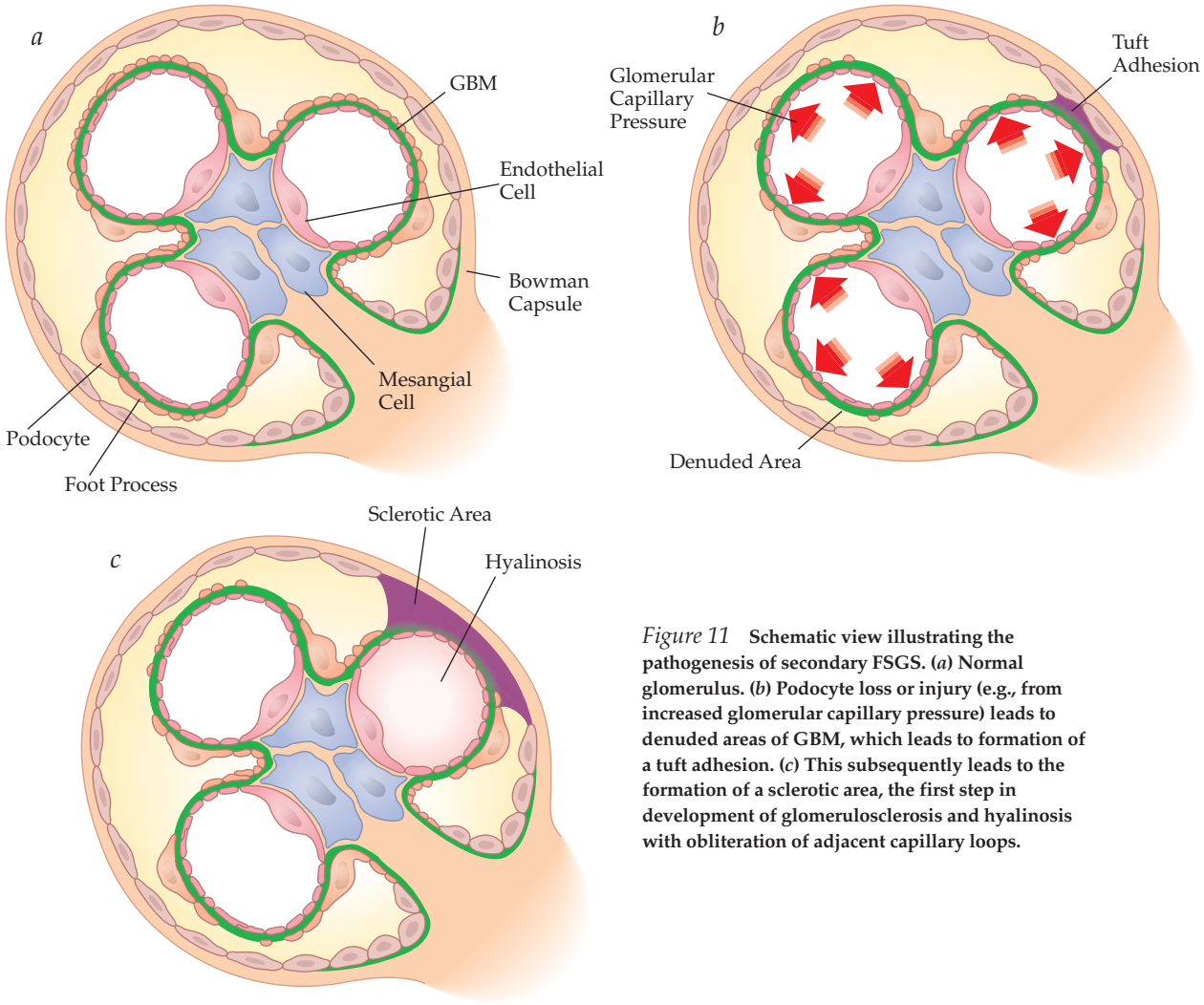


Figure 11 **Schematic view illustrating the pathogenesis of secondary FSGS. (*a*) Normal glomerulus. (*b*) Podocyte loss or injury (e.g., from increased glomerular capillary pressure) leads to denuded areas of GBM, which leads to formation of a tuft adhesion. (*c*) This subsequently leads to the formation of a sclerotic area, the first step in development of glomerulosclerosis and hyalinosis with obliteration of adjacent capillary loops.**

An increased serum creatinine level at presentation and heavy proteinuria are poor prognostic indicators in FSGS. Control of proteinuria is associated with greater renal survival than the persistence of proteinuria in the nephrotic range. Surprisingly, the number of scarred glomeruli is not a prognostic factor. However, the presence of diffuse mesangial hypercellularity does correlate with a poor prognosis. Progression to renal failure in FSGS has no significant correlation with age or gender or the presence of hematuria or hypertension on presentation.

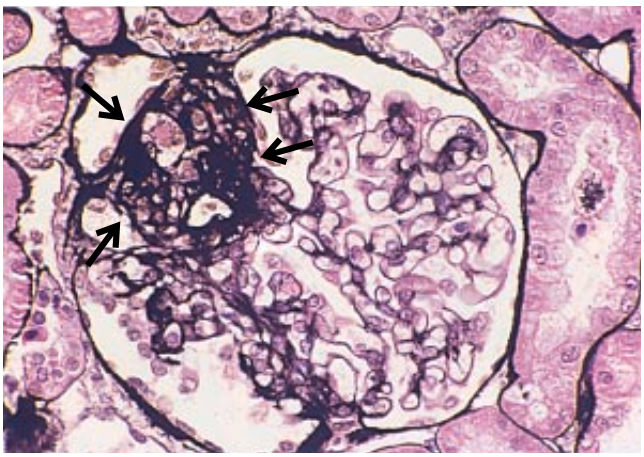


Figure 12 **Light micrograph showing the characteristic features of focal segmental glomerulosclerosis, with collapse of capillaries, hyalinosis, and adhesion (area highlighted by arrows).**

Patients with end-stage renal failure from FSGS are excellent transplant candidates. However, the disease subsequently occurs in the transplanted kidney in 20% to 30% of patients. Plasmapheresis has proved helpful in some patients with recurrent FSGS after transplantation, who likely have a circulating factor causing their disease.[64]

HIV-Associated Nephropathy

HIV infection can cause glomerular lesions.[65] The most common presentation is nephrotic syndrome caused by FSGS (see above). HIV infection typically is associated with the collapsing form of FSGS, also known as HIV-associated nephropathy (HIVAN). HIVAN occurs predominantly in African Americans. Histologically, HIVAN is characterized by proliferation of podocytes and collapse of the glomerular tuft. HIV infection can also cause hemolytic-uremic syndrome and thrombotic thrombocytopenic purpura, IgA nephropathy, mesangial proliferative glomerulonephritis, and MPGN.

Patients with HIVAN usually have advanced HIV infection, with $CD4^+$ T cell counts below 200 cells/μl. In about 30% of pa-

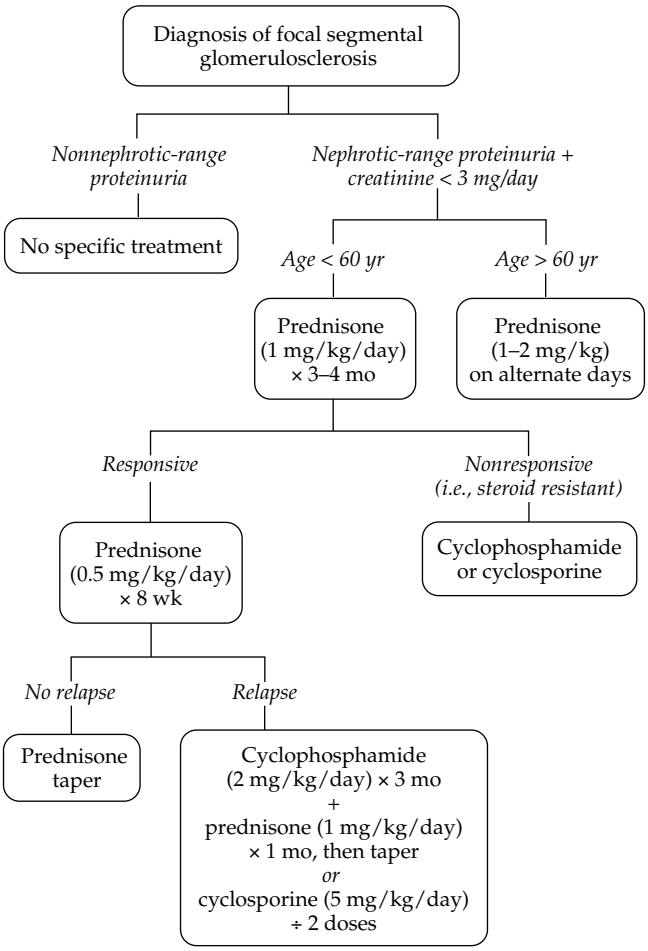


Figure 13 **Treatment algorithm for patients with focal segmental glomerulosclerosis.**

tients, HIVAN may be the presenting symptom that prompts a workup for HIV infection. Steroids, antiretroviral therapy, and ACE inhibitors have been found to be helpful in the treatment of HIVAN.

In the past, HIVAN was characterized by progression to end-stage renal disease within months. With current antiretroviral therapy, however, many patients can avoid dialysis for years. In contrast to the small kidney size that characterizes most forms of end-stage renal disease, HIV-associated nephropathy is associated with an enlarged kidney.

Membranous Nephropathy

Although membranous nephropathy derives its name from the characteristic thickening of the GBM seen on light microscopy, the primary site of injury is the podocyte. In turn, podocyte dysfunction results in increased extracellular matrix protein accumulation, causing the basement membrane thickening that is the hallmark of the disease.

Membranous nephropathy is the most common cause of the nephrotic syndrome in white adults and is the leading cause of nephrotic syndrome in persons older than 60 years. In adults, 80% of cases are idiopathic; the remaining 20% are from secondary causes [*see Table 4*]. The most common secondary causes are SLE; hepatitis B infection[66]; drugs (e.g., gold, penicillamine); and, in the elderly, solid tumors—malignancy underlies 20% of cases in patients older than 50 years.[67] Membranous nephropathy is typically a disease of men (the male-to-female ratio is 2:1 to 3:1) and has a biphasic age distribution of 30 to 40 years and 50 to 60 years.

Pathogenesis and histology Although thickening of the GBM characterizes membranous nephropathy, it occurs late in the course of disease. The pathognomonic finding is diffuse deposition of finely granular immune deposits at the base of the podocytes on the outer surface of the GBM, which are referred to as subepithelial or subpodocyte immune deposits [*see Figure 14*]. The location of the immune deposits and evidence from experimental models of disease suggest that the target of disease is one or more antigens on the podocytes.[68] The inciting antigen (or antigens) remains to be identified in humans. However, the concurrent presence of subendothelial and mesangial deposits is more suggestive of circulating antigens or antigen-antibody complexes. Antibodies such as antineuroendopeptidase have been linked with the onset of membranous nephropathy. Membranous nephropathy is mediated by a humoral immune response, leading to deposition of IgG and C3, but not C4, indicating activation of the alternative complement pathway. Immunostaining for C5b-9 (the membrane attack complex) occurs in a pattern similar to that for IgG [*see Figure 15*]. Consequences of C5b-9 injury to podocytes include production of oxidants, cytokines, proteases, and growth factors that further damage the podocyte and the underlying GBM. The presence of C1q on the biopsy suggests lupus nephritis class V as the secondary cause.

Later in the course of disease, thickening of the GBM is evident on light microscopy. Silver immunostaining, which is used to detect organized matrix, shows the characteristic spike pattern of membrane thickening forming around the subepithelial deposits. Electron microscopy shows flattening, fusion, and effacement of the podocytes, which augments proteinuria. As in most forms of progressive glomerular disease, further deterioration in renal function is caused by progressive tubulointerstitial fibrosis.

Clinical and laboratory findings Patients with membranous nephropathy present with the complications of marked nephrotic-range proteinuria (> 3.5 g/day), including edema and fluid retention. The proteinuria is nonselective (i.e., including not only albumin but also larger proteins) and is usually in the range of 5 to 15 g/day. However, 20% of patients with membranous

Table 4 Conditions Associated with Membranous Nephropathy

Tumors (e.g., bowel, breast, lung)
Infections (e.g., hepatitis B, hepatitis C, syphilis, filariasis, hydatid disease, schistosomiasis, malaria, leprosy)
Drugs and toxins (e.g., gold, penicillamine, nonsteroidal anti-inflammatory drugs, mercury, captopril, formaldehyde, hydrocarbons)
Immune diseases (e.g., systemic lupus erythematosus, diabetes mellitus, rheumatoid arthritis, Hashimoto disease, Graves disease, mixed connective tissue disease, Sjögren syndrome, primary biliary cirrhosis, bullous pemphigoid, small bowel enteropathy syndrome, dermatitis herpetiformis, ankylosing spondylitis, graft versus host disease, Guillain-Barré syndrome)
Miscellaneous conditions (e.g., sarcoidosis, Kimura disease [angiolymphoid hyperplasia with eosinophilia], angiofollicular lymph node hyperplasia)

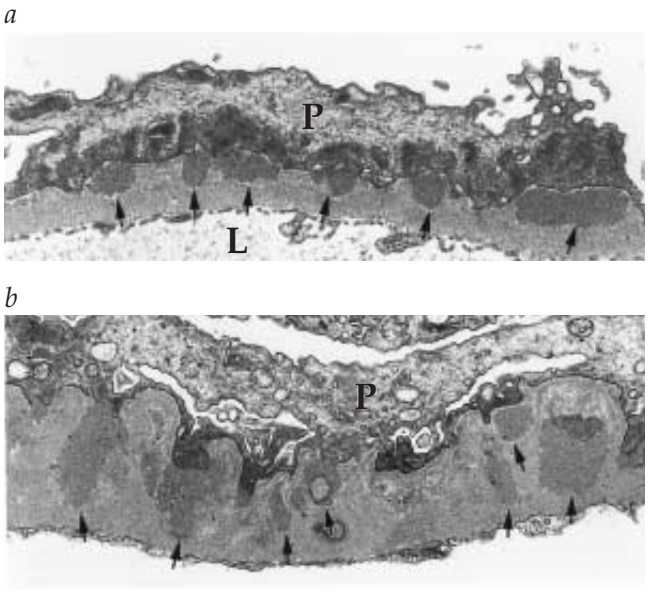


Figure 14 **Transmission electron micrograph showing characteristic subepithelial cell deposits (arrows) in (*a*) early and (*b*) late membranous nephropathy; P illustrates podocytes, and L shows the capillary lumen.**

nephropathy have non–nephrotic-range proteinuria and therefore may be asymptomatic. Clinical clues to secondary causes may be evident at presentation in patients who have systemic symptoms (e.g., a rash in patients with SLE). Microscopic hematuria is noted in 50% of patients, yet the urine is devoid of RBC casts and WBCs. Because the decline in GFR occurs slowly in membranous nephropathy, renal function is normal in 80% of patients at presentation. Hypertension occurs in 30% of cases.

Positive laboratory studies in membranous nephropathy include those providing evidence of nephrotic syndrome (i.e., hypoalbuminemia, hypercholesterolemia, and proteinuria). Although hypercoagulability can complicate any form of the nephrotic syndrome, renal vein thrombosis is most commonly found in membranous nephropathy and may occur in up to 50% of patients. Despite the local activation of complement in the glomerulus, serum total complement and individual complement components are normal in idiopathic membranous nephropathy. In contrast, low complement levels support secondary causes, such as SLE, hepatitis B, and hepatitis C. Investigations for specific secondary causes should be sought, especially in patients older than 50 years.

Treatment and prognosis Management of membranous nephropathy includes the treatment of the underlying disease in secondary forms of membranous nephropathy and prevention or reduction of complications of nephrotic syndrome. The prognosis of idiopathic membranous nephropathy has traditionally been known by the rule of thirds[69]: about one third of patients have a spontaneous, complete remission within 3 to 5 years, with resolution of proteinuria and normalization of the GFR; one third have partial remission, with persistent proteinuria of less than 2 g/day but a normal GFR; and one third progress to end-stage renal disease. Some studies have shown that closer to 50% of patients progress to end stage. In these patients, GFR drops 50% over 3 to 4 years. Poor prognostic markers include male sex, age over 60 years, hypertension, proteinuria that is massive in both quantity and duration (e.g., > 8 g for 6 months, > 6 g for 9 months, or > 4 g for 12 months), and a decrease in GFR.

Renal injury can be reduced by controlling hypertension and lowering proteinuria or by the judicious use of ACE inhibitors, ARBs, or both. In patients who are at high risk for progressive re-

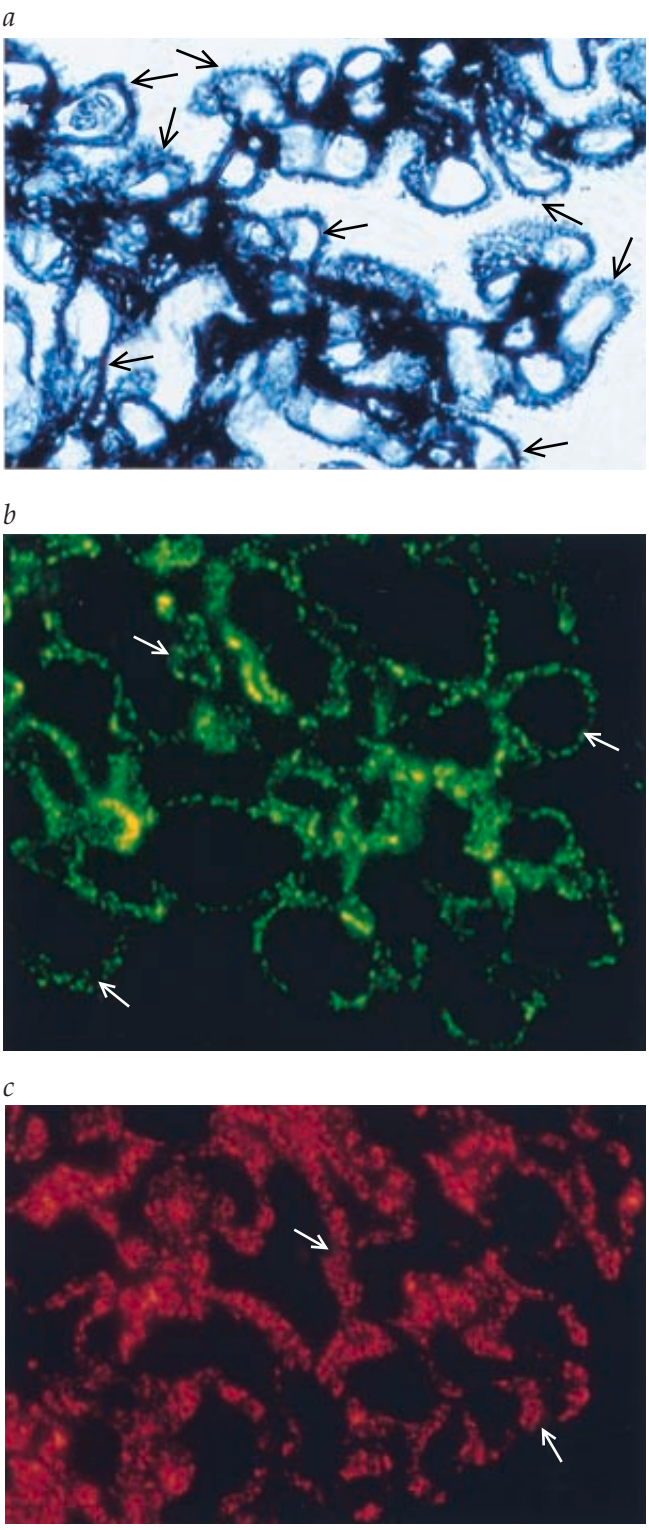


Figure 15 **(*a*) Silver staining showing thickened GBM and spike formation in membranous nephropathy. (*b*) Immunofluorescent staining showing the typical granular pattern of IgG in membranous nephropathy. (*c*) C5b-9 immunostaining in a pattern similar to that of IgG.**

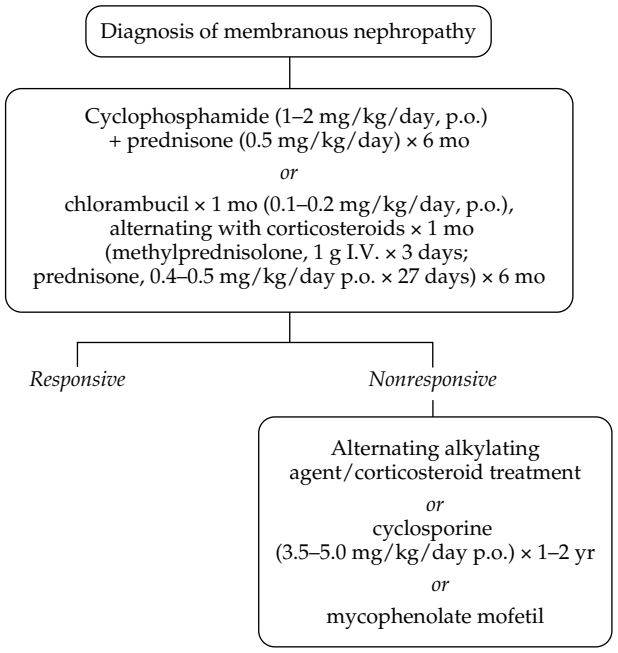


Figure 16 **Treatment algorithm for high-risk patients with membranous nephropathy.**

nal failure, disease-specific treatment should be implemented in addition to these nonspecific measures [*see Figure 16*].[70]

In contrast to idiopathic FSGS and minimal change disease, membranous nephropathy cannot be adequately treated with corticosteroids alone; instead, disease-specific therapy includes both alkylating agents and steroids. Two regimens are used. The first is oral cyclophosphamide (1 to 2 mg/kg/day) with prednisone (0.5 mg/kg/day) for 6 months.[71-73] A second, proposed protocol involves alternating cycles of oral chlorambucil (0.1 to 0.2 mg/kg/day) for 1 month and steroids for 1 month (methylprednisolone, 1g I.V., followed by 0.4 to 0.5 mg/kg/day of oral prednisone for 27 days).[74]

If alkylating agents are contraindicated or if the patient experiences relapses, the second line of therapy is oral cyclosporine (3.5 to 5.0 mg/kg/day) for 1 to 2 years.[75] Alternative and experimental therapies include intravenous immunoglobulin and mycophenolate mofetil and, more recently, rituximab. These therapies should be administered by a nephrologist. Renal transplantation for patients with end-stage renal failure from membranous nephropathy is encouraged, although there is a slight (5%) risk of subsequent occurrence in the donor kidney.

Minimal Change Disease

Minimal change disease derives its name from the lack of clear-cut histologic abnormalities visible on light microscopy of renal biopsy specimens.[76] Minimal change disease is the most common form of primary glomerular disease causing nephrotic syndrome in children and is the third most common cause of nephrotic syndrome in adults.[77] There is no gender predilection in adults, whereas twice as many boys as girls are affected.

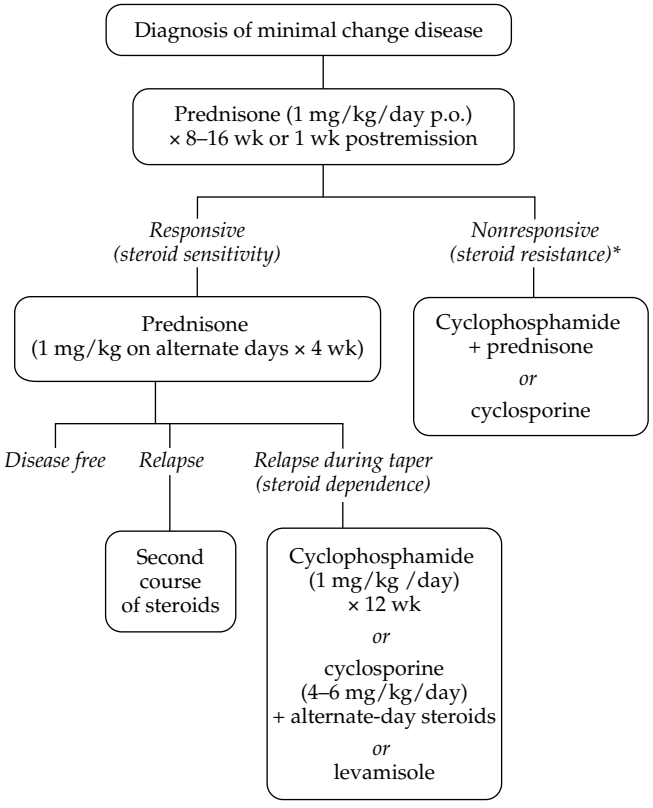
Although most forms of minimal change disease are idiopathic, secondary causes need to be ruled out [*see Table 5*]. A combination of the nephrotic syndrome and renal failure should alert the clinician to NSAID-induced minimal change disease complicated by interstitial nephritis.[78,79]

Table 5 Causes of Minimal Change Disease

Drugs (e.g., nonsteroidal anti-inflammatory drugs, interferon alfa, lithium, gold)
Allergies (e.g., pollen, house dust, insect stings, immunizations)
Malignancies (e.g., Hodgkin lymphoma, leukemia)

Pathogenesis and histology Minimal change disease is probably caused by a defect in cell-mediated immunity, particularly T cells. Podocytes are the target of injury. However, no abnormalities are seen on light microscopy, and immunofluorescent staining is negative for immunoglobulins and complement factors. On electron microscopy, the classic finding is flattening, effacement, and fusion of the foot processes.[80] Podocyte injury leads to a decrease in production of negatively charged proteoglycans in the underlying basement membrane and foot processes, which underlies the development of proteinuria.

Clinical and laboratory findings Because the onset of proteinuria is typically abrupt in minimal change disease, patients present with acute clinical findings of the nephrotic syndrome. The laboratory abnormalities usually reflect the severity of the nephrotic syndrome. Hypertension and a decline in renal function occur in fewer than 20% of affected adults. Proteinuria is typically massive (> 6 g/day), urine sediment is unremarkable, and hematuria is rare.



*Heavy proteinuria for more than 12–16 wk.

Figure 17 **Treatment algorithm for patients with minimal change disease.**

Table 6 Causes of Membranoproliferative Glomerulonephritis Types I–III

Type	*Causes*
I	Idiopathic Collagen vascular diseases (e.g., rheumatoid arthritis, Sjögren syndrome, systemic lupus erythematosus) Complement deficiencies Acquired (e.g., C4 nephritic factor) Hereditary (e.g., C1q, C2, C3, C4) Cryoglobulinemia (types 2 and 3) Infections Viral (e.g., hepatitis C, chronic hepatitis B, HIV) Bacterial (e.g., infective endocarditis, visceral abscess, infected shunt) Other (e.g., *Plasmodium malariae, Schistosoma, Mycoplasma*) Malignancies (e.g., chronic lymphocytic leukemia, lymphoma, renal cell carcinoma) Other (e.g., α_1-antitrypsin deficiency, other chronic liver diseases, hypocomplementemic urticaria vasculitis)
II	Idiopathic Associated with C3 nephritic factor Associated with factor H defect Partial lipodystrophy
III	Idiopathic Associated with terminal complement nephritic factor Same causes listed for type I (see above)

Treatment and prognosis Corticosteroids are the treatment of choice for minimal change disease [*see Figure 17*]. In adults, prednisone (1 mg/kg/day) should be given for 8 to 16 weeks. In those patients who respond (> 95%), prednisone is then switched to alternate days at a dosage of 1 mg/kg for 4 weeks, followed by tapering over the next few months.[80] A first relapse should be treated with another course of steroids. In patients who have relapses more than twice after completion of the steroid course and in patients who have a relapse during the taper (i.e., steroid-dependent patients), cyclophosphamide (2 mg/kg/day) should be given for 8 to 12 weeks, combined with prednisone at an increased dosage. Alternatives to cytotoxic therapy include cyclosporine (4 to 6 mg/kg/day) along with alternate-day prednisone. Limited published experience with mycophenolate mofetil suggests some benefit in patients with refractory or relapsing disease.[81-83] In patients who do not respond to prednisone within 12 to 16 weeks (i.e., steroid-resistant patients), cyclophosphamide or cyclosporine may be used. Levamasole, an immune stimulant, has also been used with some success in steroid-dependent children with nephritic syndrome from minimal change disease.[84] In children who have frequent relapses after treatment with alkylating agents, levamisole may serve as an alternative.[85]

DISEASES PRESENTING AS MIXED NEPHRITIC-NEPHROTIC SYNDROME

Membranoproliferative Glomerulonephritis

MPGN, which is also known as mesangiocapillary glomerulonephritis, is an immune complex–mediated glomerular disease.[86] It can present as nephritic, nephrotic, or nephritic-nephrotic syndrome. The name derives from the classic finding of splitting of the basement membrane and proliferation of mesangial and endothelial cells. MPGN accounts for 5% to 20% of all primary nephrotic syndromes. The incidence of the disease is highest in South America and Africa, and it occurs equally in males and females. MPGN is mostly idiopathic, but the histologic diagnosis should always prompt a search for secondary causes [*see Table 6*].[87]

Pathogenesis and histology MPGN is classified into types I, II, and III, according to the sites of deposition of glomerular immune complexes (IgG and C3). These are preformed immune complexes, such as hepatitis C antigen and anti–hepatitis C antibody. Each type of MPGN has a different cause and involves activation of different components of the complement pathway, giving rise to differences in serum complement levels.

MPGN type I refers to immune complex deposits that are subendothelial and mesangial and result from diseases characterized by chronic immune complex formation that activates the classic complement pathway [*see Figure 18*].[88] The most common cause of MPGN type I is hepatitis C, with or without mixed cryoglobulinemia.[26] Less frequent causes include mixed cryoglobulinemia that does not result from hepatitis C infection, hepatitis B infection, or subacute bacterial endocarditis. The classic double contouring (so-called tram tracks) seen on histology is a result of interposition of mesangial cells, leukocytes, or endothelial cells in the capillary wall, with the synthesis of new basement membrane material. Resident mesangial and endothelial cells and infiltrating cells proliferate in MPGN. Immunofluorescent staining is positive for IgG, IgM, and C3 along the capillary wall; these immune complexes take on a granular pattern, which is distinct from the linear pattern in anti-GBM disease.

MPGN type II occurs primarily in young adults and children. It is caused by activation of the alternative complement pathway and is associated with the presence of dense deposits in the basement membrane itself that are positive for C3 but negative for IgG and for the classic complement components C1q and C4.[89] Accordingly, type II MPGN is also called dense-deposit disease. Its etiology is unknown, but possible causes include the absence of a constitutive inhibitor of the alternative complement pathway (factor H) or the presence of a circulating autoantibody (C3 nephritic factor) that binds to and prevents inactivation of the alternative pathway C3 convertase.[90]

MPGN type III has lesions that are similar to those of type I but have additional subepithelial deposits. Occasionally, intramembranous deposits are also present.[91]

Clinical and laboratory findings MPGN has four clinical presentations: microscopic hematuria and nonnephrotic proteinuria, which is seen in 35% of patients; nephrotic syndrome and a mild decrease in renal function, seen in 35% of patients; chronic glomerulonephritis, usually progressive, in 20% of patients; and acute renal failure with proteinuria and an active urine sediment with RBC casts, in 10%. Hypertension is a very common presenting finding, occurring in 50% to 80% of patients, and the urinary sediment is often active. One quarter of patients with MPGN type II may also manifest partial lipodystrophy that preferentially involves the face and upper body.

Although patients present with evidence of glomerular disease, clues to a secondary cause may also be present, such as features of cryoglobulinemia (e.g., weakness, arthralgia, purpura), viral infection (e.g., hepatitis B, hepatitis C, HIV), bacterial infection (e.g., endocarditis, abscess), and liver disease. Laboratory evidence for MPGN can be categorized as follows: the presence

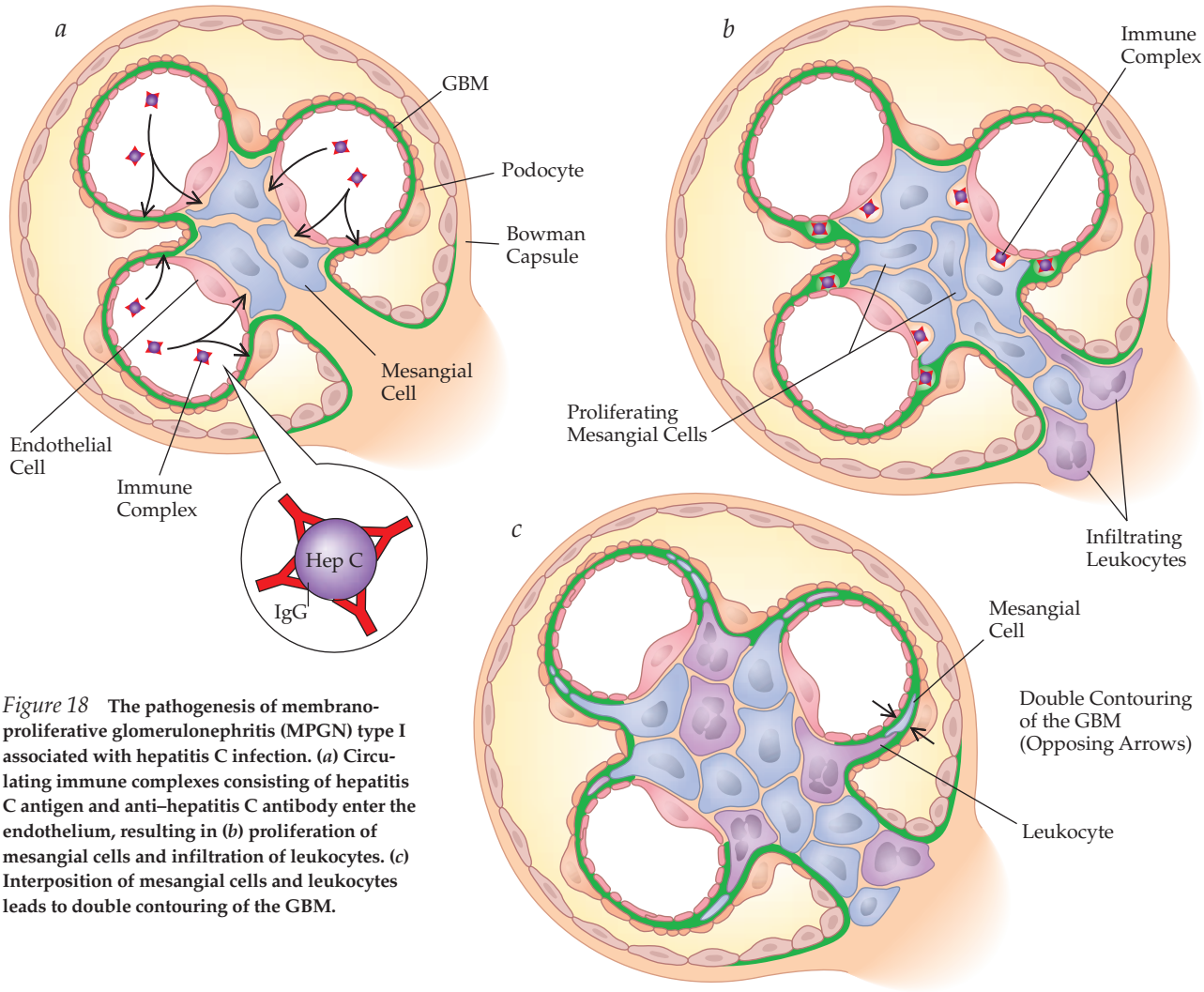


Figure 18 **The pathogenesis of membranoproliferative glomerulonephritis (MPGN) type I associated with hepatitis C infection. (*a*) Circulating immune complexes consisting of hepatitis C antigen and anti–hepatitis C antibody enter the endothelium, resulting in (*b*) proliferation of mesangial cells and infiltration of leukocytes. (*c*) Interposition of mesangial cells and leukocytes leads to double contouring of the GBM.**

of the nephrotic syndrome or renal failure, or both, and the complications thereof; activation of the classic complement cascade, with low levels of C1q, C4, and CH_{50}, which is seen in MPGN types I and III; and activation of the alternative pathway, with low levels of C3 and CH_{50} and normal levels of C1q and C4, which is characteristic of MPGN type II. In addition, the clinician should look for evidence of an underlying cause of MPGN, such as the presence of circulating immune complexes, cryoglobulins, rheumatoid factor, and specific viruses (e.g., hepatitis C virus RNA, hepatitis B markers, HIV).

Treatment and prognosis Hepatitis C–associated MPGN with cryoglobulinemia is usually treated with antiviral therapy (i.e., PEG-interferon alpha-2b and ribavirin, if renal function allows). Remission occurs in 60% of patients, but more than 80% of patients experience relapses. Other infection-associated MPGN usually responds to specific antimicrobial therapy, which reduces the antigen levels and therefore the immune complexes. Persistent signs of renal disease and hypocomplementemia usually indicate inadequate therapy. Steroids have been administered on alternate days in children with MPGN, but the use of steroids remains controversial in adults. On the basis of pediatric studies, long-term low-dose prednisone may be tried in adults with nephrotic-range proteinuria. For MPGN associated with rapidly progressive renal failure, combination therapy with steroids and cytotoxic agents is frequently used.[92]

The prognosis of MPGN depends in part on the presence of any underlying disease, which must be treated with specific therapy. Idiopathic MPGN has a poor prognosis: 40% to 50% of patients progress to renal failure within 10 years. Optimal therapy has not been established. In adults, studies have shown potential benefit with antiplatelet therapy such as aspirin (975 mg/day) and dipyridamole (225 mg/day), anticoagulation therapy such as warfarin, and steroids. Controlled trials of alkylating agents have not shown long-term benefit, and the therapeutic benefit of cyclosporine has not been adequately proved. MPGN recurs in renal transplant patients with type I disease (20% to 50% of patients) and type II disease (80% of patients) but is usually milder than in native kidneys.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hricik DE, Chung-Park M, Sedor JR: Glomerulonephritis. N Engl J Med 339:888, 1998

2. Madaio MP, Harrington JT: Current concepts: the diagnosis of acute glomerulonephritis. N Engl J Med 309:1299, 1983

3. Hebert LA, Cosio FG, Neff JC: Diagnostic significance of hypocomplementemia. Kidney Int 39:811, 1991

4. Falk RJ, Jennette JC: ANCA small-vessel vasculitis. J Am Soc Nephrol 8:314, 1997

5. Adler S: Glomerular epithelial cells. Immunologic Renal Diseases. Neilson EG, Couser WG, Eds. Lippincott-Raven, Philadelphia, 1997, p 655

6. Guasch A, Deen WM, Myers BD: Charge selectivity of the glomerular filtration barrier in healthy and nephrotic humans. J Clin Invest 92:2274, 1993

7. Kanwar YS, Liu ZZ, Kashihara N, et al: Current status of the structural and functional basis of glomerular filtration and proteinuria. Semin Nephrol 11:390, 1991

8. Couser WG: Glomerulonephritis. Lancet 353:1509, 1999

9. Kalluri R, Wilson CB, Weber M, et al: Identification of the alpha 3 chain of type IV collagen as the common autoantigen in antibasement membrane disease and Goodpasture syndrome. J Am Soc Nephrol 6:1178, 1995

10. Kalluri R, Sun MJ, Hudson BG, et al: The Goodpasture autoantigen: structural delineation of two immunologically privileged epitopes on alpha3(IV) chain of type IV collagen. J Biol Chem 271:9062, 1996

11. Couser WG, Abrass CK: Pathogenesis of membranous nephropathy. Annu Rev Med 39:517, 1988

12. Fries JW, Mendrick DL, Rennke HG: Determinants of immune complex-mediated glomerulonephritis. Kidney Int 34:333, 1988

13. Hebert LA: Complement and complement regulatory proteins in renal disease. Immunologic Renal Diseases. Neilson EG, Couser WG, Eds. Lippincott-Raven, Philadelphia, 1997, p 377

14. Main IW, Atkins RC: The role of T-cells in inflammatory kidney disease. Curr Opin Nephrol Hypertens 4:354, 1995

15. Pollock C, Liu PL, Gyory AZ, et al: Dysmorphism of urinary red blood cells: value in diagnosis. Kidney Int 36:1045, 1989

16. Lewy JE, Salinas-Madrigal L, Herdson PB, et al: Clinico-pathologic correlations in acute poststreptococcal glomerulonephritis: a correlation between renal functions, morphologic damage and clinical course of 46 children with acute poststreptococcal glomerulonephritis. Medicine (Baltimore) 50:453, 1971

17. Johnson RJ, Willson R, Yamabe H, et al: Renal manifestations of hepatitis C virus infection. Kidney Int 46:1255, 1994

18. Humphreys MH: Human immunodeficiency virus–associated glomerulosclerosis. Kidney Int 48:311, 1995

19. Rao TK: Clinical features of human immunodeficiency virus associated nephropathy. Kidney Int Suppl 35:S13, 1991

20. D'Amico G: Renal involvement in hepatitis C infection: cryoglobulinemic glomerulonephritis. Kidney Int 54:650, 1998

21. Rodriguez-Iturbe B, Carr RI, Garcia R, et al: Circulating immune complexes and serum immunoglobulins in acute poststreptococcal glomerulonephritis. Clin Nephrol 13:1, 1980

22. Tornroth T: The fate of subepithelial deposits in acute poststreptococcal glomerulonephritis. Lab Invest 35:461, 1976

23. Neugarten J, Baldwin DS: Glomerulonephritis in bacterial endocarditis. Am J Med 77:297, 1984

24. Schoeneman M, Bennett B, Greifer I: Shunt nephritis progressing to chronic renal failure. Am J Kidney Dis 2:375, 1982

25. Kim Y, Michael AF: Chronic bacteremia and nephritis. Annu Rev Med 29:319, 1978

26. Johnson RJ, Gretch DR, Yamabe H, et al: Membranoproliferative glomerulonephritis associated with hepatitis C virus infection. N Engl J Med 328:465, 1993

27. Donadio JV Jr, Grande JP: Immunoglobulin A nephropathy: a clinical perspective. J Am Soc Nephrol 8:1324, 1997

28. D'Amico G: The commonest glomerulonephritis in the world: IgA nephropathy. Q J Med 64:709, 1987

29. Emancipator SN: IgA nephropathy: morphologic expression and pathogenesis. Am J Kidney Dis 23:451, 1994

30. Feehally J, Allen AC: Structural features of IgA molecules which contribute to IgA nephropathy. J Nephrol 12:59, 1999

31. Haas M: Histologic subclassification of IgA nephropathy: a clinicopathologic study of 244 cases. Am J Kidney Dis 29:829, 1997

32. Kauffmann RH, Herrmann WA, Meyer CJ, et al: Circulating IgA-immune complexes in Henoch-Schonlein purpura: a longitudinal study of their relationship to disease activity and vascular deposition of IgA. Am J Med 69:859, 1980

33. Nolin L, Courteau M: Management of IgA nephropathy: evidence-based recommendations. Kidney Int Suppl 70:S56, 1999

34. Rekola S, Bergstrand A, Bucht H: Deterioration rate in hypertensive IgA nephropathy: comparison of a converting enzyme inhibitor and beta-blocking agents. Nephron 59:57, 1991

35. Cattran DC, Greenwood C, Ritchie S: Long-term benefits of angiotensin-converting enzyme inhibitor therapy in patients with severe immunoglobulin a nephropathy: a comparison to patients receiving treatment with other antihypertensive agents and to patients receiving no therapy. Am J Kidney Dis 23:247, 1994

36. Praga M, Gutierrez E, Gonzalez E, et al: Treatment of IgA nephropathy with ACE inhibitors: a randomized and controlled trial. J Am Soc Nephrol 14:1578, 2003

37. Nakao N, Yoshimura A, Morita H, et al: Combination treatment of angiotensin-II receptor blocker and angiotensin-converting-enzyme inhibitor in non-diabetic renal disease (COOPERATE): a randomised controlled trial. Lancet 361:117, 2003

38. Ballardie FW, Roberts IS: Controlled prospective trial of prednisolone and cytotoxics in progressive IgA nephropathy. J Am Soc Nephrol 13:142, 2002

39. Dillon JJ: Fish oil therapy for IgA nephropathy: efficacy and interstudy variability. J Am Soc Nephrol 8:1739, 1997

40. Yoshikawa N, Ito H, Sakai T, et al: A controlled trial of combined therapy for newly diagnosed severe childhood IgA nephropathy. The Japanese Pediatric IgA Nephropathy Treatment Study Group. J Am Soc Nephrol 10:101, 1999

41. Tamura S, Ueki K, Ideura H, et al: Corticosteroid therapy in patients with IgA nephropathy and impaired renal function. Clin Nephrol 55:192, 2001

42. Katafuchi R, Ikeda K, Mizumasa T, et al: Controlled, prospective trial of steroid treatment in IgA nephropathy: a limitation of low-dose prednisolone therapy. Am J Kidney Dis 41:972, 2003

43. Pozzi C, Bolasco PG, Fogazzi GB, et al: Corticosteroids in IgA nephropathy: a randomised controlled trial. Lancet 353:883, 1999

44. Floege J: Evidence-based recommendations for immunosuppression in IgA nephropathy: handle with caution. Nephrol Dial Transplant 18:241, 2003

45. Couser WG, Abrass CK: Pathogenesis of membranous nephropathy. Ann Rev Med 39:517, 1988

46. Salant DJ: Immunopathogenesis of crescentic glomerulonephritis and lung purpura. Kidney Int 32:408, 1987

47. Donaghy M, Rees AJ: Cigarette smoking and lung haemorrhage in glomerulonephritis caused by autoantibodies to glomerular basement membrane. Lancet 2:1390, 1983

48. Madore F, Lazarus JM, Brady HR: Therapeutic plasma exchange in renal diseases. J Am Soc Nephrol 7:367, 1996

49. Levy JB, Pusey CD: Still a role for plasma exchange in rapidly progressive glomerulonephritis? J Nephrol 10:7, 1997

50. Couser WG: Rapidly progressive glomerulonephritis: classification, pathogenetic mechanisms, and therapy. Am J Kidney Dis 11:449, 1988

51. Angangco R, Thiru S, Esnault VL, et al: Does truly 'idiopathic' crescentic glomerulonephritis exist? Nephrol Dial Transplant 9:630, 1994

52. Schwab SJ, Christensen RL, Dougherty K, et al: Quantitation of proteinuria by the use of protein-to-creatinine ratios in single urine samples. Arch Intern Med 147:943, 1987

53. Harris RC, Ismail N: Extrarenal complications of the nephrotic syndrome. Am J Kidney Dis 23:477, 1994

54. Appel G: Lipid abnormalities in renal disease. Kidney Int 39:169, 1991

55. Llach F: Hypercoagulability, renal vein thrombosis, and other thrombotic complications of nephrotic syndrome. Kidney Int 28:429, 1985

56. Savin VJ, McCarthy ET, Sharma M: Permeability factors in focal segmental glomerulosclerosis. Semin Nephrol 23:147, 2003

57. Kriz W: Progressive renal failure: inability of podocytes to replicate and the consequences for development of glomerulosclerosis. Nephrol Dial Transplant 11:1738, 1996

58. D'Agati V: The many masks of focal segmental glomerulosclerosis. Kidney Int 46:1223, 1994

59. Markowitz GS, Appel GB, Fine PL, et al: Collapsing focal segmental glomerulosclerosis following treatment with high-dose pamidronate. J Am Soc Nephrol 12:1164, 2001

60. Beaufils H, Alphonse JC, Guedon J, et al: Focal glomerulosclerosis: natural history and treatment. A report of 70 cases. Nephron 21:75, 1978

61. Savin V, Sharma R, Sharma M, et al: Circulating factor associated with increased glomerular permeability to albumin in recurrent focal segmental glomerulosclerosis. N Engl J Med 334:878, 1996

62. Cameron JS: Focal segmental glomerulosclerosis in adults. Nephrol Dial Transplant 18(suppl 6):vi45, 2003

63. Cattran DC, Appel GB, Hebert LA, et al: A randomized trial of cyclosporine in patients with steroid-resistant focal segmental glomerulosclerosis. North America Nephrotic Syndrome Study Group. Kidney Int 56:2220, 1999

64. Artero ML, Sharma R, Savin VJ, et al: Plasmapheresis reduces proteinuria and serum capacity to injure glomeruli in patients with recurrent focal glomerulosclerosis. Am J Kidney Dis 23:574, 1994

65. Klotman PE: HIV-associated nephropathy. Kidney Int 56:1161, 1999

66. Lai KN, Li PK, Lui SF, et al: Membranous nephropathy related to hepatitis B virus in adults. N Engl J Med 324:1457, 1991

67. Alpers CE, Cotran RS: Neoplasia and glomerular injury. Kidney Int 30:465, 1986

68. Kerjaschki D, Neale TJ: Molecular mechanisms of glomerular injury in rat experimental membranous nephropathy (Heymann nephritis). J Am Soc Nephrol 7:2518, 1996

69. Donadio JV Jr, Torres VE, Velosa JA, et al: Idiopathic membranous nephropathy: the natural history of untreated patients. Kidney Int 33:708, 1988

70. Muirhead N: Management of idiopathic membranous nephropathy: evidence-based recommendations. Kidney Int Suppl 70:S47, 1999

71. West ML, Jindal KK, Bear RA, et al: A controlled trial of cyclophosphamide in patients with membranous glomerulonephritis. Kidney Int 32:579, 1987

72. Bruns FJ, Adler S, Fraley DS, et al: Sustained remission of membranous glomerulonephritis after cyclophosphamide and prednisone. Ann Intern Med 114:725, 1991

73. Jindal K, West M, Bear R, et al: Long-term benefits of therapy with cyclophosphamide and prednisone in patients with membranous glomerulonephritis and impaired renal function. Am J Kidney Dis 19:61, 1992

74. Ponticelli C, Zucchelli P, Passerini P, et al: A randomized trial of methylprednisolone and chlorambucil in idiopathic membranous nephropathy. N Engl J Med 320:8, 1989

75. Ponticelli C: Treatment of the nephrotic syndrome with cyclosporin A. J Autoimmun 5(suppl A):315, 1992

76. Carrie BJ, Salyer WR, Myers BD: Minimal change nephropathy: an electrochemical disorder of the glomerular membrane. Am J Med 70:262, 1981

77. Nolasco F, Cameron JS, Heywood EF, et al: Adult-onset minimal change nephrotic syndrome: a long-term follow-up. Kidney Int 29:1215, 1986

78. Lowenstein J, Schacht RG, Baldwin DS: Renal failure in minimal change nephrotic syndrome. Am J Med 70:227, 1981

79. Bargman JM: Management of minimal lesion glomerulonephritis: evidence-based recommendations. Kidney Int Suppl 70:S3, 1999

80. Bohman SO, Jaremko G, Bohlin AB, et al: Foot process fusion and glomerular filtration rate in minimal change nephrotic syndrome. Kidney Int 25:696, 1984

81. Briggs WA, Choi MJ, Scheel PJ Jr: Successful mycophenolate mofetil treatment of glomerular disease. Am J Kidney Dis 31:213, 1998

82. Choi MJ, Eustace JA, Gimenez LF, et al: Mycophenolate mofetil treatment for primary glomerular diseases. Kidney Int 61:1098, 2002

83. Day CJ, Cockwell P, Lipkin GW, et al: Mycophenolate mofetil in the treatment of resistant idiopathic nephrotic syndrome. Nephrol Dial Transplant 17:2011, 2002

84. Levamisole for corticosteroid-dependent nephrotic syndrome in childhood. British Association for Paediatric Nephrology. Lancet 337:1555, 1991

85. Filler G: Treatment of nephrotic syndrome in children and controlled trials. Nephrol Dial Transplant 18(suppl 6):vi75, 2003

86. Ooi YM, Vallota EH, West CD: Classical complement pathway activation in membranoproliferative glomerulonephritis. Kidney Int 9:46, 1976

87. Rennke HG: Secondary membranoproliferative glomerulonephritis. Kidney Int 47:643, 1995

88. Cameron JS, Turner DR, Heaton J, et al: Idiopathic mesangiocapillary glomerulonephritis: comparison of types I and II in children and adults and long-term prognosis. Am J Med 74:175, 1983

89. Bennett WM, Fassett RG, Walker RG, et al: Mesangiocapillary glomerulonephritis type II (dense-deposit disease): clinical features of progressive disease. Am J Kidney Dis 13:469, 1989

90. Halbwachs L, Leveille M, Lesavre P, et al: Nephritic factor of the classical pathway of complement: immunoglobulin G autoantibody directed against the classical pathway C3 convertase enzyme. J Clin Invest 65:1249, 1980

91. Strife CF, Jackson EC, McAdams AJ: Type III membranoproliferative glomerulonephritis: long-term clinical and morphologic evaluation. Clin Nephrol 21:323, 1984

92. Levin A: Management of membranoproliferative glomerulonephritis: evidence-based recommendations. Kidney Int Suppl 70:S41, 1999

Acknowledgments

Figures 1b, 6, 8, 11, and 18 Seward Hung.

Figures 3, 5, and 9 Courtesy of Dr. Vivette D'Agati, Columbia University.

Figure 4 Courtesy of Dr. Agnes Fogo, Vanderbilt University School of Medicine.

Figure 6 Courtesy of Dr. Peter Furness, University of Leicester.

Figures 7 and 8 Courtesy of Dr. Charles E. Alpers, University of Washington.

Figures 11 and 12 Courtesy of Dr. Dontscho Kerjaski, Vienna, Austria.

VI ACUTE RENAL FAILURE

Mary Jo Shaver, m.d.
Sudhir V. Shah, m.d., f.a.c.p.

Acute renal failure (ARF) can be broadly defined as an abrupt decrease in renal function occurring over hours to days and sufficient to result in retention of nitrogenous waste (i.e., azotemia), as measured by elevations in blood urea nitrogen (BUN) and creatinine levels. The usual criteria for ARF are an increase in serum creatinine of 0.5 mg/dl, a 25% increase in serum creatinine, or a 25% decrease in glomerular filtration rate (GFR). However, it is now recognized that even a modest rise in serum creatinine of 0.3 mg/dl is associated with increased mortality and morbidity, and this may lead to revision of the criteria for ARF.

ARF is further defined pathophysiologically, depending on whether it results from a decrease in renal blood flow (prerenal azotemia), intrinsic renal parenchymal diseases (renal azotemia), or obstruction of urine flow (postrenal azotemia) [*see Table 1*]. The most common intrinsic renal disease that leads to ARF is acute tubular necrosis (ATN). Clinical recognition is based largely on exclusion of prerenal and postrenal causes of sudden azotemia, followed by exclusion of other causes of intrinsic ARF (e.g., glomerulonephritis, acute interstitial nephritis [AIN], and vasculitis). Although the term acute tubular necrosis is not an entirely valid histologic description of this syndrome, the term is ingrained in clinical medicine and is therefore used in this chapter.

Despite major advances in dialysis and intensive care, the mortality in patients with severe ARF who require dialysis has not decreased significantly over the past 50 years. Because of that and because even a minor decline in renal function is associated with poor patient outcome, prevention [*see Tables 2 and 3*] and early diagnosis of ARF are the cornerstones of management. Discovery of biomarkers that can predict the development of renal dysfunction before the serum creatinine level rises may lead to early and successful treatment of ARF.

Epidemiology

The frequency of ARF varies greatly, depending on the clinical setting. The incidence of ARF has been cited as 1% on admission to the hospital, from 2% to 5% during hospitalization, in as many as 20% of patients treated in intensive care units, and in 4% to 15% of patients after cardiovascular surgery.[1-3] In a study of 2,216 medical and surgical patients, ARF developed in 5% of patients; of those cases, 55% were associated with iatrogenic factors—mostly adverse drug effects—and sepsis.[1] In a study of 681 ICU patients with ARF, the average age was 59.5 years, and many patients had complicated medical conditions; 30% had chronic kidney disease, 37% had coronary artery disease, 29% had diabetes mellitus, and 21% had chronic liver disease.[4] Female sex has been found to be an independent risk factor for developing ARF after open heart surgery.[5]

The incidence of severe ARF in adults in Western countries is approximately 140 per million population per year; 50 to 70 per million population per year require dialysis.[6]

Acute Renal Failure in the Hospitalized Patient

etiology

Prerenal azotemia is the single most common cause of ARF, accounting for 30% to 60% of all cases.[7] Prerenal azotemia, depending on its severity and duration, may lead to ATN in a susceptible person. Therefore, early identification and aggressive treatment of prerenal azotemia are important. In the hospital setting, postrenal azotemia causes 1% to 10% of cases of ARF.[7] ARF associated with postrenal azotemia should be a consideration in elderly men, especially if they are receiving medications that could impair bladder function.

ATN is the most common intrinsic renal disease leading to ARF in hospitalized patients. When prerenal and postrenal causes of ARF have been excluded, about 75% of hospitalized patients with ARF have ATN.[7] Multiple insults are usually present, but prerenal azotemia is often the predisposing factor. Iatrogenic causes, including sepsis and nephrotoxins (e.g., antibiotics, contrast dye), contribute to a large portion of these cases. Approxi-

Table 1 Causes of Acute Renal Failure

I. Prerenal azotemia
  A. Absolute decrease in effective blood volume
    Hemorrhage, skin losses (burns, sweating), gastrointestinal losses (diarrhea, vomiting), renal losses (diuretics, glycosuria), fluid pooling (peritonitis, burns)
  B. Relative decrease in blood volume (ineffective arterial volume)
    Congestive heart failure, sepsis, anaphylaxis, liver failure
  C. Arterial occlusion
    Bilateral thromboembolism, thromboembolism of solitary kidney
II. Renal azotemia
  A. Vascular causes
    Vasculitis, malignant hypertension, microscopic polyarteritis
  B. Acute glomerulonephritis
    Postinfectious glomerulonephritis, anti–basement membrane–antibody disease
  C. Acute interstitial nephritis
    Drug-associated acute interstitial nephritis (methicillin nephrotoxicity)
  D. Acute tubular necrosis
    1. Ischemia
      Prerenal azotemia (if severe enough), postsurgical complication
    2. Sepsis syndrome
    3. Nephrotoxicity
      a. Exogenous nephrotoxins
        Antibiotics (aminoglycosides, cephalosporin, amphotericin B); iodinated contrast agents; chemotherapeutic agents (cisplatin); solvents (carbon tetrachloride, ethylene glycol)
      b. Endogenous nephrotoxins
        Intratubular pigments (hemoglobinuria, myoglobinuria), intratubular proteins (myeloma), intratubular crystals (uric acid, oxalate), tumor lysis syndrome
III. Postrenal azotemia (obstruction of collecting system)
  Bladder outlet obstruction, bilateral ureteral obstruction (unusual), ureteral obstruction in a solitary kidney

Table 2 Prevention of Acute Renal Failure

1. Identify patients with risk factors
   - Old age
   - Abnormal renal function or diabetes
   - Volume depletion
   - Recent vascular surgery
   - Recent trauma
2. Avoid nephrotoxic agents
   - Nonsteroidal anti-inflammatory drugs
   - Aminoglycosides
   - Amphotericin B
   - Chemotherapeutic agents (e.g., cisplatin)
   - Angiotensin-converting enzyme inhibitors in volume-depleted patients
3. Apply preventive strategies in specific circumstances [*see Table 3*]
   - Use of contrast media
   - Tumor lysis syndrome
   - Rhabdomyolysis
   - Surgical procedures

mately 40% to 60% of ATN cases occur in the postoperative or trauma setting. Depending on the clinical setting, other diagnoses to be considered are AIN (e.g., secondary to use of methicillin), glomerulonephritis, atheromatous emboli (in association with previous aortic surgery, aortography, or both), ureteral obstruction (in association with pelvic or abdominal pathology or secondary to complications of pelvic or abdominal surgery), or intrarenal obstruction (e.g., acute uric acid nephropathy).

PATHOPHYSIOLOGY

Prerenal Azotemia

Prerenal ARF occurs when glomerular perfusion falls, as a result of either an absolute reduction in the volume of extracellular fluid or a reduction in circulating volume despite a normal total extracellular fluid volume, as may occur in conditions such as heart failure, advanced cirrhosis, and septic states. The kidney has the capacity to autoregulate GFR and blood flow simultaneously during renal hypoperfusion through the independent regulation of afferent and efferent arteriolar tone. While the afferent arteriole dilates in response to renal hypoperfusion, the efferent arteriole constricts, maintaining glomerular intracapillary pressure, which is the driving force for glomerular filtration. Thus, during the early phase of mild to moderate prerenal conditions, renal blood flow and GFR are maintained within normal ranges, and BUN and creatinine levels remain normal. When prerenal conditions become severe and renal adaptive mechanisms cannot compensate, GFR falls and the BUN and creatinine levels begin to increase.

Prerenal ARF is characterized by renal hypoperfusion, a high ratio of BUN to serum creatinine, and a low urine volume. Levels of antidiuretic hormone are elevated, thus increasing the reabsorption of both water and urea. Prerenal failure is a response to severe volume depletion, heart failure, or sepsis. Prerenal failure can also occur with the presentation of a large solute load to the kidney, along with stimulation of volume loss by an osmotic diuresis, as occurs in hyperglycemia.[8]

Acute Tubular Necrosis

ATN is a clinical syndrome of abrupt and sustained decline in GFR that is triggered by an acute ischemic or nephrotoxic event and develops within minutes to days after the insult. Leakage of glomerular ultrafiltrate from the tubular lumen into the renal interstitium across the damaged renal tubular cells, obstruction of flow within the tubule by debris or crystals in the lumen, and a decrease in the glomerular capillary ultrafiltration coefficient have all been proposed as playing a pathophysiologic role in ATN.

A variety of biochemical and cellular changes may be involved in cell injury in ARF. These include mitochondrial dysfunction, adenosine triphosphate (ATP) depletion, phospholipid degradation, elevation in cytosolic free calcium levels, a decrease in Na^+,K^+-ATPase activity, alterations in substrate metabolism, lysosomal changes, and the production of oxygen free radicals. It is not clear yet which changes are causative and which are by-products of advanced cell injury.

Ischemic ARF culminating in ATN is a very complex physiologic process resulting from a mismatch between oxygen and

Table 3 Prevention of Acute Renal Failure in High-Risk Circumstances

Procedures Involving the Use of Contrast Media
- Hydration
  - Normal saline at 1 ml/kg/hr starting 8–12 hr before the procedure and continuing for 8–12 hr afterward
  - Contrast medium
    - Limit the volume of contrast used
    - Use iso-osmolar contrast (iodixanol) for high-risk patients
  - Pharmacologic pretreatment
    - Acetylcysteine, 600 mg p.o., b.i.d., the day before and the day of the procedure

Tumor Lysis Syndrome
- Hydration and forced diuresis
  - Infuse normal saline to maintain a urine output of 3–5 L/day
- Urinary alkalinization
  - Infuse 2 to 3 ampules of sodium bicarbonate (100–150 mEq) in 5% dextrose in water to maintain urine pH > 7
  - Infuse acetazolamide, 1 g/m^2
  - Monitor basic metabolic panel and avoid significant metabolic alkalosis
- Pharmacologic treatment
  - Allopurinol, 300–600 mg/day, starting 3 days before chemotherapy; adjust dose for patients with renal impairment

Rhabdomyolysis
- Hydration
  - Infuse normal saline to replace volume if patient is volume depleted, then give 200–300 ml/hr and follow hemodynamic status; balance fluid input with urine output
- Pharmacologic treatment
  - Add 2 ampules of mannitol (25 g/100 ml) and 2 ampules of sodium bicarbonate (100 mEq/100 ml) to 800 ml of 5% dextrose in water; infuse at a rate of 250 ml/hr while frequently monitoring fluid intake, urine output, and basic metabolic panel; if the urine output is good, continue infusing until the myoglobinuria resolves; if the patient becomes oliguric (urine output < 400 ml/24 hr), stop I.V. fluids and manage as established renal failure

Surgical Procedures
- Recognize patients with high-risk conditions
  - Preexisting renal disease
  - Chronic liver disease
  - Cardiac failure
  - Old age
- Avoid volume depletion
- Avoid hypotension
- Avoid nephrotoxic agents

nutrient delivery and energy demand of the nephrons. The most important pathophysiologic component of ischemic ARF may be a reduction in the local blood flow to the outer medulla.[9] There appears to be an excess of constrictor hormones relative to hormones that dilate the afferent arteriole; subsequently, there is a drop in glomerular perfusion pressure and depression of the GFR. The brunt of the postischemic injury is in the proximal tubular cells, with loss of cell polarity and redistribution of Na^+/K^+-ATPase from its normal location on the basolateral plasma membrane. This leads to an increase in the delivery of sodium to the macula densa that stimulates afferent arteriole vasoconstriction.[10]

Inflammatory cells appear to be involved in the pathogenesis of ATN.[9,11] In addition, whereas previous studies focused on the renal tubular injury to epithelial cells, current research also emphasizes the importance of endothelial cells in ATN. As endothelial cells are injured, the cells swell, cell adhesion molecules are expressed, leukocytes are activated, and renal injury is potentiated. Vasoactive mediators promote vasoconstriction and further compromise in local blood flow and tubular cell metabolism.[12]

Despite the common use of the term ATN, necrosis of the tubules is seen infrequently in either ischemic or nephrotoxic ARF. Cell death may be of two kinds: apoptotic or necrotic. Advances in the understanding of cell death have led to the recognition that the pathways traditionally associated with apoptosis may be critical in determining the form of cell injury associated with necrosis. Apoptotic pathways, in which endonucleases play an important role, appear to be regulated by mediators such as oxidants, caspases, and ceramide. Which pathway the cell follows seems to depend on both the nature and the severity of the insult. It is likely that the pathway followed is to a large extent affected by the expression of the many genes involved in cell cycle regulation and by inflammatory and chemotactic genes. It is also likely that the cascades that lead to either apoptotic or necrotic cell death are activated almost simultaneously and may share some common pathways.

Natural history Persistent renal impairment from ATN may be oliguric or nonoliguric and typically lasts for 1 to 2 weeks, but it can persist for 4 to 6 weeks. This is followed by the diuretic phase, which eventually leads to normalization of kidney function over a few days to a week. Most patients achieve complete recovery of renal function, except for perhaps some defect in the concentrating ability of the kidney. However, there is increased recognition that a small but significant percentage of patients with ATN do not recover renal function and may require long-term dialysis.

DIAGNOSIS

To determine the cause of ARF, the physician must follow a systematic approach, which should start with excluding or correcting both prerenal and postrenal azotemia. Often, the difficulty in arriving at a correct etiologic diagnosis of ARF in a hospitalized patient is not in identifying a possible cause but, rather, in selecting the actual cause from among several possible ones.

Correct diagnosis depends on basic knowledge of the natural history of ARF from different causes, review of the chronologic sequence of events in the deterioration in the patient's renal function, and analysis of the available patient data. Despite the exhaustive list of conditions that can cause acute azotemia in hospitalized patients, a careful history and physical examination and simple laboratory tests often suffice for diagnosis.

Clinical Manifestations

In a hospital setting, ARF is most commonly recognized when patients exhibit oliguria, rising BUN and creatinine levels, or both oliguria and rising BUN and creatinine. Thus, it is unusual for ARF to go unrecognized until the patient develops symptoms suggestive of uremia, such as fatigue, weakness, nausea, vomiting, loss of appetite, metallic taste in the mouth, itching, confusion, fluid retention, and hypertension.

Chart Review, History, and Physical Examination

Evaluation of the patient with ARF should start with a complete medical history and a review of the hospital records. Some of the important data that should be sought from chart review are presented [*see Tables 4 and 5 and Figure 1*].

Reduced body weight, marked orthostatic decrease in blood pressure, an increase in pulse, and lack of jugular venous distention all suggest a reduction in extracellular fluid volume. Patients with prerenal azotemia can appear to be experiencing volume overload in association with extracellular fluid expansion (e.g., cardiac failure, cirrhosis, nephrotic syndrome), but the effective blood volume is decreased and thus renal perfusion is impaired.

Careful abdominal examination may reveal a distended, tender bladder, indicating lower urinary tract obstruction. In men in whom lower tract obstruction is suspected as a cause of acute azotemia, examination of the prostate and a sterile postvoid bladder catheterization with urinalysis should be performed.

Additional findings that may be helpful are fever and rash, which occur in some patients with AIN. A history of a recent aortic catheterization (e.g., cardiac catheterization) and the finding of livedo reticularis are diagnostic clues for cholesterol or atheromatous emboli.

Differentiating prerenal azotemia from ATN may be difficult, partly because of the difficulty in evaluating the volume status in a critically ill patient and also because prerenal azotemia from any cause, if severe enough, may lead to ATN. Evaluation of the urine volume and urine sediment and a number of urinary indices (most useful in patients with oliguria) are particularly helpful in making the correct diagnosis.

Laboratory Tests

Ratio of blood urea nitrogen to creatinine Initial laboratory tests include measurement of BUN and serum creatinine, sodium, chloride, potassium, and bicarbonate levels. These tests are important not only for the diagnosis but for assessment of complications of ARF. In prerenal conditions resulting from enhanced salt and water avidity, there is a disproportionate increase in the ratio of BUN to creatinine (> 20:1). Other causes of BUN elevation include gastrointestinal bleeding, use of systemic steroids, catabolism caused by the underlying medical condition, or a high-protein diet. An elevation in the creatinine level that exceeds the elevation in BUN suggests rhabdomyolysis; such patients will also have an elevated creatine phosphokinase (CPK) or myoglobin level.

Urinary volume ARF is traditionally divided into anuric (urine output of < 100 ml/day), oliguric (urine output of 100 to 400 ml/day), and nonoliguric (urine output of > 400 ml/day). In a hospitalized patient, a normal urine output in the face of a rise in creatinine level most often signifies ATN, because prerenal azotemia should lead to oliguria. Wide variations in daily urinary output suggest obstruction. Complete anuria (no urine output) suggests obstruction or an acute vascular catastrophe, such as renal vein or

Table 4 Diagnostic Evaluation of a Patient with Acute Renal Failure

Types of Evaluation	*Comments*
Chart Review	
Previous renal function	Patients with previous renal insufficiency are more susceptible to ARF
Medications	Aminoglycosides are important causes of ATN in hospitalized patients (nonoliguric and in the first 2 wk of therapy); certain antibiotics, NSAIDs, and a host of medications can cause AIN; contrast agents are an important cause of ARF
Start date	
Drug levels	
Nephrotoxins?	
NSAIDs?	
Aminoglycosides?	
Amphotericin B?	
Contrast agent used?	
Surgery	Cardiac and vascular surgery patients are particularly susceptible to ATN
Type and duration	
Hemodynamics	
Blood loss	
Anesthetic used	Methoxyflurane and enflurane (which is related to methoxyflurane but is less toxic) can cause nonoliguric ATN
Infection	Infection or sepsis can cause ARF even in the absence of hypertension
Fever?	
Positive blood cultures?	
Hemodynamics	Hypotension can cause prerenal azotemia and ischemic ATN
Blood pressure	
Heart rate	
Weight (in and out)	
Physical Examination	
Volume status	
Edema?	Pretibial in the ambulatory patient, sacral in the bedridden patient
Jugular vein distention?	
Crackles?	
S_3 gallop?	
Skin	
Diffuse rash?	Skin rash seen in ARF induced by some drugs (e.g., antibiotics and allopurinol); a purpuric skin rash may be seen with Henoch-Schönlein purpura and cryoglobulinemia
Livedo reticularis?	
Atheroemboli?	
Bladder	Distended bladder
Suprapubic fullness	To assess postvoid residual urinary volume and to relieve bladder obstruction
Bladder catheterization	
Urinalysis and Sediment Analysis	
Elements	
Protein and blood	
Cellular	
Cast	
Urine Indices *See Table 6*	
Additional Tests	
Renal ultrasonography	A finding of kidneys of different size suggests vascular disease in the smaller kidney
Size and symmetry	
Corticomedullary differentiation	
Obstruction	
Indication for renal biopsy	

AIN—acute interstitial necrosis ARF—acute renal failure ATN—acute tubular necrosis NSAIDs—nonsteroidal anti-inflammatory drugs

renal artery thrombosis or large emboli in the renal arteries, although it can be seen in ATN and AIN. To cause total anuria, a vascular event must affect both kidneys or a single functioning kidney.

Urinalysis and urine sediment The following is a list of findings on urinalysis, including urine sediment analysis, and their implications for renal failure:

1. In prerenal failure, a moderate number of hyaline and finely granular casts may be seen, but coarsely granular and cellular casts are infrequent. Finely granular casts can also be seen in the sediment of patients with chronic renal failure.
2. In ATN, a characteristic sediment is found in 70% to 80% of patients, particularly those with oliguric ATN; it consists of dirty-brown, coarse, granular casts, free renal tubular epithelial cells, and epithelial cell casts. With more severe renal injury, there may be a 24- to 48-hour delay before the appearance of dirty-brown casts.
3. A seemingly benign urine sediment containing few formed elements suggests possible obstruction.
4. Proliferative glomerulonephritis is characterized by urine containing 3+ to 4+ protein; 2+ to 3+ blood; and active sediment, defined as sediment containing red blood cells (RBCs) and RBC casts. An accurate history and careful physical examination (which may, for example, suggest systemic lupus erythematosus), determination of complement levels, antinuclear antibody testing, and kidney biopsy (if the kidney is of normal size) generally help clarify the diagnosis.
5. Findings of only a few RBCs in the urine sediment and strongly heme-positive urine or heme-positive urine supernatant (after removal of the RBCs by centrifugation) usually indicate myoglobinuria or hemoglobinuria. Patients with rhabdomyolysis have a marked increase in the level of muscle enzyme, such as CPK. The urine sediment in patients with myoglobinuria may show RBCs, pigmented casts, granular casts, and numerous uric acid crystals.

Table 5 Daily Evaluation and Management of the Hospitalized Patient with Acute Renal Failure

Weight	Weight loss of 0.5 to 1 lb/day in a patient with ATN will help prevent volume overload
Jugular vein distention or crackles	Indicate volume overload with possible congestive heart failure; restrict volume, add diuretics, consider dialysis
CVP, PCWP	Assessment of these parameters may be indicated to differentiate volume overload from the presence of noncardiogenic pulmonary infiltrates; low PCWP suggests noncardiogenic pulmonary edema
Intake/output	In euvolemic patients, assess volume of previous day's urine output (in stable patients, add 400 ml for insensible losses); insensible losses may be higher in catabolic, febrile, or agitated patients
Blood urea nitrogen	Disproportionately high values suggest gastrointestinal bleeding, steroid use, hypercatabolic state, or prerenal ARF
Creatinine	Disproportionately high values suggest muscle breakdown, as caused by rhabdomyolysis
Electrolytes	[*see* Complications, *in text*]

ARF—acute renal failure ATN—acute tubular necrosis CVP—central venous pressure PCWP—pulmonary capillary wedge pressure

					First Day for Elevated Creatinine		
Date							
Weight							
Input							
Output							
Blood Pressure							
Pulse							
Temperature							
BUN							
Creatinine							
Sodium							
Potassium							
Chloride							
CO_2							
WBC							
Hgb							
Hct							
Platelet							
Surgery							
Contrast							
Medications							

Figure 1 **This figure shows a worksheet for the evaluation of acute renal failure in hospitalized patients. The worksheet may be used to chart a chronologic outline of the events that may have contributed to the insult. In using such an approach, particular attention should be paid to the events that occurred a few days before the acute rise in the serum creatinine level. (BUN—blood urea nitrogen; CO_2—carbon dioxide; Hgb—hemoglobin; Hct—hematocrit; WBC—white blood cell count)**

6. The presence of white blood cells (WBCs) in clumps and in casts, in the absence of evidence of bacteria, suggests AIN. This can be associated with microscopic hematuria and RBC casts. The presence of eosinophils in the urine is less helpful in the diagnosis of AIN than was once thought.
7. A urine sediment with abundant uric acid crystals or oxalate crystals suggests uric acid deposition (associated with hyperuricemia after chemotherapy) or intratubular oxalate (e.g., methoxyflurane anesthesia), respectively.

Urinary indices Urinary indices are often used to differentiate prerenal azotemia from ATN. However, interpretation of urinary indices in a patient with chronic kidney disease is often difficult because of underlying sodium wasting in the baseline state that cannot be corrected rapidly with volume depletion. The rationale for the use of these indices is as follows: The ratio of urine to plasma creatinine (U/P_{Cr}) provides an index of the fraction of filtered water excreted. If it is assumed that all of the creatinine filtered at the glomerulus is excreted into the urine and that relatively little is added by secretion, any incremental increase in the concentration of creatinine in urine over that in plasma must be the result of the removal of water. In prerenal azotemia, because of the reduction in the amount of glomerular filtrate entering each nephron and because of an increase in the retention of salt and water, U/P_{Cr} typically is considerably greater than it is in ATN, and urinary sodium concentrations are characteristically lower [*see Table 6*]. In contrast, in ARF associated with ATN, the nephrons excrete a large fraction of their filtered sodium and water, resulting in lower U/P_{Cr} and a higher fractional excretion of sodium (FE_{Na}). Fractional excretion of urea (FE_{Urea}) has also been reported to help differentiate prerenal azotemia from renal azotemia [*see Table 6*]. The use of FE_{Urea} appears to have a particular diagnostic advantage over FE_{Na} in patients who have been receiving diuretics; in one study, the FE_{Na} was low in only 48% of diuretic-treated patients with prerenal azotemia.[13] It must be kept in mind, however, that FE_{Urea} values in chronic kidney disease have not been standardized.

Interpretations of urinary indices must be made in conjunction with other assessments of the patient because there are clinically important exceptions to these generalizations. For example, patients with certain types of ATN, such as radiographic dye–induced renal injury, may present with all the clinical characteristics of ATN but with FE_{Na} rates of less than 1%.

Imaging Studies

If the diagnosis of prerenal azotemia or ATN is reasonably certain and the clinical situation does not require that other causes of acute azotemia be excluded, generally no further diagnostic evaluation is necessary. Further diagnostic evaluation is indicated under the following circumstances: (1) the diagnosis is uncertain, especially if the clinical situation suggests other possibilities (e.g., obstruction or vascular accident); (2) clinical findings (e.g., total anuria) make the diagnosis of prerenal azotemia or ATN less likely; or (3) the oliguria persists longer than 4 weeks.

Renal ultrasonography is the initial diagnostic procedure of choice in renal failure because it is noninvasive and reliable. A finding of normal-size kidneys in a patient with advanced azotemia generally suggests acute rather than chronic renal fail-

Table 6 Urinary Diagnostic Indices

Indices	*Prerenal Azotemia*	*Acute Tubular Necrosis*
Urinary sodium (U_{Na}) (mEq/L)	< 20	> 40
Urine osmolarity (U_{osm}) (mOsm/kg H_2O)	> 500	< 450
Fractional excretion of sodium ($U_{Na}P_{Cr}/P_{Na}U_{Cr}$) (100)	< 1%	> 1%
BUN-to-creatinine ratio	> 20:1	—
Urine creatinine–plasma creatinine [U_{Cr} (mg/dl)/P_{Cr} (mg/dl)]	> 40	< 20

BUN—blood urea nitrogen P_{Cr}—plasma creatinine P_{Na}—plasma sodium

ure; however, several important causes of chronic renal failure, including diabetes mellitus, HIV infection, multiple myeloma, and amyloidosis, may be associated with normal-sized kidneys. The renal ultrasound examination is helpful in confirming or excluding obstruction, identifying polycystic kidney disease, determining whether one or two kidneys are present, and localizing the kidney for renal biopsy.

A high-resolution CT scan is considered the test of choice for suspected urinary tract calculi. Radionuclide methods are available for assessing renal blood flow and excretory (secretory) function. Blood flow studies can be used to easily determine whether the kidneys are perfused and, if so, whether blood flow to the two kidneys is symmetrical; such tests are less accurate in quantitating flow rates. Magnetic resonance angiography is recommended to evaluate renal arterial or venous thrombosis or obstruction. Radionuclide WBC scans have a limited role in diagnosing AIN.

Renal Biopsy

Renal biopsy is rarely required for ARF that occurs in the hospital setting. In contrast, renal biopsy is indicated somewhat more frequently for ARF that occurs outside the hospital.

Community-Acquired Acute Renal Failure

Azotemia that first occurs outside the hospital may be either acute or chronic. Useful points in determining whether the renal failure is acute or chronic are summarized [*see Table 7*]. The majority of patients who present with advanced azotemia have chronic renal failure.

Most cases of ARF occur in the elderly, possibly because of the anatomic and physiologic changes of aging. In a study evaluating 748 patients presenting with ARF, 36% were older than 70 years.[14] Frequent causes of community-acquired ARF are hypovolemia, ingestion of over-the-counter nonsteroidal anti-inflammatory drugs (NSAIDs) and prescription medications, and obstruction. Before a detailed evaluation is carried out, priority should be given to identifying complications of renal failure that may be lethal unless treated promptly. Some of these, such as marked fluid overload and pericardial tamponade, may be detected on clinical examination. Other life-threatening complications, however, such as severe hyperkalemia or extreme metabolic acidosis, require laboratory evaluation. The electrocardiogram is valuable in assessing the effects of hyperkalemia on the heart [*see* Treatment of Complications, Hyperkalemia, *below*]. Even before the underlying disease that is causing the azotemia is known, often a decision whether to initiate dialysis has to be made. Dialysis should be instituted promptly in patients with severe hyperkalemia, acidosis, marked fluid overload, or signs or symptoms of uremia. Although many manifestations of uremia are nonspecific, a pericardial rub or neurologic manifestations (e.g., asterixis or confusion) are indications for prompt dialysis.

After serious electrolyte and acid-base imbalances have been addressed, other clinical information important in the evaluation can be ascertained.

DIAGNOSIS

Clinical Manifestations

A patient presenting with ARF may have very nonspecific complaints, including fatigue, weakness, restlessness, loss of appetite, nausea, vomiting, decreased urine output, swelling, and hiccups. When levels of uremic toxins are markedly elevated, changes in mental status and seizures may occur. A history of having recently started new medications may suggest drug-related renal injury. A history of nausea, vomiting, diarrhea, or other volume losses suggests ARF resulting from a prerenal condition. Recent trauma with muscle injury suggests rhabdomyolysis as the cause of ARF. Fever, skin rash, and joint pains raise the possibility of a rheumatic disease, such as systemic lupus erythematosus, vasculitis, endocarditis, or drug allergy with AIN. A history of shortness of breath or pulmonary hemorrhage suggests a pulmonary renal syndrome, such as Goodpasture syndrome, Wegener granulomatosis, Churg-Strauss disease, or pulmonary edema associated with acute glomerulonephritis. Abrupt anuria suggests acute obstruction, severe glomerulonephritis, or a sudden vascular event. Painless hematuria suggests acute glomerulonephritis, whereas painful hematuria is more consistent with obstruction.

Clinical and Laboratory Evaluation

Use of physical examination, laboratory tests, and imaging studies in community-acquired ARF is largely identical to that in patients with hospital-acquired ARF (see above).

Specific Causes of Acute Renal Failure

EXOGENOUS NEPHROTOXINS

Radiocontrast

Radiocontrast is a common cause of ARF.[15] In patients with normal renal function who have no other risk factors for ARF, the incidence of contrast-induced renal injury is low (< 1%). The incidence of contrast-induced nephropathy is approximately 150,000 cases a year in the United States.[16,17] Contrast-induced nephropathy is defined as ARF occurring shortly after exposure to intravenous contrast—typically within 48 hours—in the absence of other causes of renal failure. The most important risk factor is preexisting renal insufficiency; other risk factors include diabetes,[16,18] volume depletion, advanced age, heart failure, states of reduced renal perfusion, high total dose of contrast, and concomitant exposure to other nephrotoxins. A meta-analysis of 31 controlled, randomized trials, which included more than 5,000 patients, compared high-osmolar agents with low-osmolar agents; in all but nine studies, a favorable outcome was reported for patients receiving low-osmolar agents.[19] The odds ratio for an increase in serum creatinine of greater than 0.5 mg/dl

(44 μmol/L) with low-osmolar agents was 0.61 times that of the odds ratio associated with high-osmolar agents. Low-osmolar contrast media are the first choice for patients at risk for contrast media–induced ARF. Iodixanol, a nonionic, iso-osmolar contrast medium, has been shown to be beneficial in preventing contrast nephropathy in high-risk patients.[20] Extracellular fluid volume expansion with saline plays a key role in the prevention of contrast-induced ARF.[21-23] High-risk patients should be kept well hydrated by administration of normal saline at the rate of 1 ml/kg/hr for 8 to 12 hours before and after the procedure. The addition of diuretics and mannitol to accelerate diuresis has not been shown to provide any benefit over volume expansion alone, and it may be harmful. However, diuretics may be useful for fluid management in patients with volume overload or low cardiac output. For patients at high risk for contrast-induced ARF, consideration should be given to the volume of contrast used, because there is an increased risk of renal injury with higher doses and with doses that are repeated within 48 hours. In high-risk patients, consideration should be given to the use of diagnostic procedures that do not require contrast.

Acetylcysteine is commonly used to minimize toxicity of contrast in high-risk patients. Clinical studies of acetylcysteine for preventing contrast media–induced ARF have yielded conflicting results, however, and further studies are in order before this measure can be routinely recommended.[24,25] The use of fenoldopam, a selective D_1 dopamine receptor agonist that causes both systemic and renal arteriolar vasodilatation, has been advocated for helping to prevent contrast-induced nephropathy. In a randomized, controlled, multicenter study of patients with chronic renal insufficiency, however, fenoldopam was ineffective in preventing further deterioration of renal function after administration of contrast dye.[26]

Hydration with sodium bicarbonate before administration of contrast medium appears to be more effective than hydration with sodium chloride. In a randomized, controlled trial in 199 patients whose serum creatinine level was at least 1.1 mg/dl (1.1 to 3.7 mg/dl), contrast-induced nephropathy occurred in eight patients (13.6%) who received sodium chloride by infusion but in only one patient (1.7%) who received sodium bicarbonate.[27] These encouraging data need to be confirmed by other studies. In the typical hydration protocol, 154 mEq/L of sodium bicarbonate in 5% dextrose in water (D5W) was infused at a rate of 3 ml/kg/hr for 1 hour immediately before radiocontrast administration, followed by 1 ml/kg/hr during contrast administration and for 6 hours after the procedure.

Nephrotoxic Drugs

Aminoglycosides The most important manifestation of aminoglycoside nephrotoxicity is ARF secondary to ATN, which occurs in about 10% to 20% of patients receiving aminoglycosides (e.g., tobramycin, gentamicin, amikacin).[28,29] Maintaining blood levels in the therapeutic range reduces but does not eliminate the risk of nephrotoxicity. Aminoglycoside-associated ARF is usually mild and nonoliguric; it is manifested by an increase in the serum creatinine level after about 1 week of aminoglycoside therapy. Patients with aminoglycoside nephrotoxicity may present with polyuria and hypomagnesemia; these conditions occur as a result of a decrease in the urinary concentrating ability and enhanced urinary loss, respectively.[28] Several clinical factors or conditions can potentiate the effect of aminoglycosides on the kidney and can thus potentiate nephrotoxicity; these include renal ischemia induced by hypotension or volume depletion, the dosing schedule and the serum levels of aminoglycosides, sepsis, administration of other nephrotoxins, and liver disease. Once-daily dosing of aminoglycosides is as effective in controlling infection as more frequent dosing, with less nephrotoxic effect; there is no difference in the incidence of ototoxicity with once-daily dosing versus more frequent dosing.[30] Individualized pharmacokinetic dosing of aminoglycosides has been shown to reduce the incidence of nephrotoxicity while allowing for higher doses of the drug to be administered.[31]

The only treatment for aminoglycoside nephrotoxicity is to discontinue the medication and to support the patient during the period of ARF. The prognosis for recovery of renal function after several days is excellent, although some patients may need dialysis before full recovery is achieved.

Amphotericin B Amphotericin B is a relatively frequent cause of ARF. Hydration with normal saline before the infusion of amphotericin B decreases the incidence of ARF from this medication. Amphotericin B in lipid emulsion may be less nephrotoxic than nonlipid formulations and is recommended for patients at high risk for nephrotoxicity. Permanent renal failure from amphotericin B nephrotoxicity is very uncommon.

Table 7 Features That Help Differentiate between Acute and Chronic Renal Failure

	Acute Renal Failure	*Chronic Renal Failure*
Previous laboratory assay	Normal renal function	Abnormal renal function is documented on old laboratory studies
Medical history	None	Longstanding and poorly controlled diabetes, hypertension, and severe vascular disease are risk factors
Renal ultrasonography	Normal	Small, echogenic kidneys; patients with diabetes, HIV infection, multiple myeloma/amyloidosis, and polycystic kidney disease may have large kidneys
Bone films	Normal	Possible evidence of renal osteodystrophy with osteitis fibrosa, osteomalacia, mixed and adynamic bone lesions, and dialysis-related amyloidosis; subperiosteal erosions of the phalanges and tuft; bone cysts with amyloidosis
Hemoglobin/hematocrit	Anemia is possible, but a normal hemoglobin level in a patient with advanced azotemia is presumptive evidence of acute renal failure	Anemia is common

Nonsteroidal anti-inflammatory drugs NSAIDs are potent inhibitors of prostaglandin synthesis, a property that contributes to their nephrotoxic potential in certain high-risk patients who require prostaglandins for maintaining renal vasodilatation. The most frequent pattern of injury related to the use of NSAIDs is prerenal azotemia; this is particularly the case for patients with volume depletion or a reduced effective circulating volume. Susceptible patients include those with heart failure, cirrhosis, diabetes, chronic renal disease, nephrotic syndrome, and septic shock, as well as those of advanced age or who require use of a diuretic.[32]

A hyperchloremic metabolic acidosis, often associated with hyperkalemia, has also been recognized as an effect of NSAIDs, particularly in patients with preexisting chronic interstitial renal disease. In such persons, hyporeninemic hypoaldosteronism occurs during states of renal prostaglandin inhibition. NSAIDs have been associated with the development of AIN, which is often associated with renal insufficiency and marked proteinuria.[33] This complication appears to be an idiosyncratic reaction and is particularly associated with the use of propionic acid derivatives, such as ibuprofen, naproxen, and fenoprofen.[34] In contrast to AIN associated with other drugs, AIN associated with NSAIDs is marked by a low incidence of hypersensitivity symptoms and eosinophilia. This disorder usually resolves with discontinuance of the offending agent.[35] High-risk patients should be educated about the risk of using NSAIDs, and they should be advised to avoid these medications if possible. ARF caused by cyclooxygenase-2 inhibitors has been reported; thus, the same cautions apply to these agents as to nonspecific NSAIDs.

Cisplatin Renal injury is a well-recognized complication of the use of cisplatin for the management of many carcinomas. Cisplatin-associated nephrotoxicity affects a significant percentage of such patients; 25% to 35% develop a mild and reversible decrease in GFR after their first dose of cisplatin. With subsequent doses, the incidence and severity of renal failure increases, until irreversible renal injury occurs. Hypomagnesemia caused by renal losses of magnesium may be severe and can occur in as many as 50% of patients. Patients should be well hydrated with normal saline (200 to 250 ml/hr) before administration of cisplatin; other known nephrotoxins should be avoided whenever possible.[36] The usual lesion is that of ATN, but with severe damage or recurrent administration of the drug, chronic interstitial disease in the setting of persistent inflammation may ensue.[11]

Angiotensin-converting enzyme inhibitors ARF that is associated with the use of angiotensin-converting enzyme (ACE) inhibitors is thought to be hemodynamic in origin; it is believed to occur as a result of the loss of autoregulation of renal blood flow and has typically been reported when ACE inhibitors are given to patients with bilateral renal artery stenosis. A 30% increase from baseline in serum creatinine levels is acceptable when ACE inhibitors are initiated.[37] If the serum creatinine level continues to rise, the ACE inhibitor should be discontinued, and an investigation for renal vascular disease should be considered. ACE inhibitors are not directly nephrotoxic; therefore, discontinuance of the medication should allow renal function to return to baseline if no other insults have occurred.

Ethylene Glycol Toxicity

Ethylene glycol is a colorless, odorless, sweet liquid found in solvents and antifreeze. Ingestion of ethylene glycol, usually in the form of antifreeze, produces a syndrome of severe metabolic acidosis characterized by a high anion gap and a large osmolar gap. Anion gap and osmolar gap are defined as follows:

$$\text{Anion gap} = [Na^+] - ([Cl^-] + [HCO_3^-])$$

$$\text{Plasma osmolality (calculated)} = 2(Na^+) + (\text{BUN}/2.8) + (\text{glucose}/18) + (\text{ethanol}/4.7)$$

$$\text{Osmolar gap} = \text{measured osmolality} - \text{calculated osmolality}$$

The normal value for the anion gap has been 12 ± 4 mEq/L. However, in a retrospective analysis of 222 patients with normal serum creatinine and albumin levels, the range for the anion gap was found to be much narrower: 6.6 ± 2 mEq/L. This difference was thought to be the result of the use of automated laboratory analysis techniques that use ion-selective electrodes.[38]

The osmolar gap is the difference between the measured and the calculated osmolality. The calculated osmolar gap is derived by use of the serum sodium, glucose, and urea levels. The addition of solutes to plasma can contribute to the osmolar gap; if such solutes are contributing to the osmolar gap, the measured osmolality will be found to be higher than the calculated osmolality. Alcohol intoxication is probably the most common cause of an osmolar gap.

Ethylene glycol is metabolized by alcohol dehydrogenase to glycolic acid, which is believed to be the major contributor to acidosis.[39] The key clinical findings in patients who have ingested ethylene glycol are initial disorientation and agitation, with progression to central nervous system depression, renal failure, metabolic acidosis, respiratory failure, and circulatory insufficiency. Hypocalcemia is a prominent feature that occurs as a result of the deposition of calcium oxalate in multiple tissues; it may be aggravated by a decrease in parathyroid hormone response. Calcium oxalate crystals are typically found in the urine sediment. ARF generally manifests after 48 to 72 hours.

Aggressive intervention should be initiated at the time of diagnosis. Intervention should consist of intravenous infusion of sodium bicarbonate to enhance renal clearance of glycolate through ion trapping; intravenous infusions of ethanol or fomepizole to block the metabolism of ethylene glycol; and hemodialysis for the removal of ethylene glycol and glycolate.[40] Regular monitoring of the osmolar gap (corrected for ethanol level if I.V. ethanol is being used during treatment) and the anion gap will help guide therapy during hemodialysis.[39]

ENDOGENOUS NEPHROTOXINS

Rhabdomyolysis

Since the first description of the causative association between rhabdomyolysis and ARF in victims of crush injuries in World War II, the spectrum of recognized causes of rhabdomyolysis, myoglobinuria, and renal failure has markedly broadened. The most frequent causes are trauma or other injury that leads to muscle compression; ischemia; excess muscle activity, such as occurs during exercise or seizures; metabolic derangements; drugs; and infections.[41] Some important metabolic derangements that can cause rhabdomyolysis include hypokalemia and hypophosphatemia; the risk of rhabdomyolysis associated with these electrolyte imbalances is increased in patients with chronic alcoholism. Cocaine use, neuroleptic malignant syndrome, and the use of statin drugs can also contribute to or cause rhabdomyolysis.

Muscle pain and dark-brown urine that is positive for blood on dipstick testing (orthotoluidine positive) but that does not

contain RBCs are important diagnostic clues. However, the diagnosis must be confirmed by findings of elevated CPK and myoglobin levels. About one third of patients with rhabdomyolysis develop ARF; many of these patients have hyperkalemia, hyperuricemia, hyperphosphatemia, early hypocalcemia, and a reduced BUN-to-creatinine ratio because of excessive creatinine release from muscle. Late hypercalcemia is also a typical feature of the disease.

Early recognition of rhabdomyolysis and initiation of treatment are the keys to minimizing ARF. These patients commonly develop hypovolemic shock, particularly when the injury is associated with trauma and massive muscle crushing.

The most important aspect of management of rhabdomyolysis is rapid volume repletion. Patients who are encountered in the field should receive immediate treatment with intravenous normal saline, 200 to 300 ml/hr. If urine output increases in 4 to 6 hours, the infusion should be continued (at a rate that matches the urine output) until the rhabdomyolysis resolves. However, if the patient continues to be oliguric (i.e., urine output is less than 400 ml/day), the infusion should be discontinued and the patient treated for ARF. Experience from recent disasters has shown that early aggressive hydration and alkalinization can prevent myoglobinuric ARF by protecting the kidney from the nephrotoxicity of myoglobin and urate.[42] This therapy consists of sodium bicarbonate and D5W, in a ratio of three ampules of sodium bicarbonate to 1 L of D5W, infused at a rate of 250 ml/hr. The metabolic alkalosis induced will help protect the patient from hyperkalemia, which can be a lethal complication of rhabdomyolysis. Urinary alkalinization is controversial; some investigators consider it no better than saline diuresis, as well as potentially hazardous because it may contribute to intratubular calcium and phosphate deposition.[43]

A clinical presentation that is similar to that of rhabdomyolysis occurs after the release of heme pigments after intravascular hemolysis.

Hyperuricemic Acute Renal Failure

ARF may occur in patients with malignancies that are associated with a high rate of tumor cell turnover (tumor lysis syndrome) [*see 12:XII Oncologic Emergencies*]. Such turnover may occur either spontaneously or after chemotherapy; high cell turnover is particularly associated with poorly differentiated lymphomas and acute lymphoblastic leukemia. Uric acid production and hyperuricosuria may increase, causing uric acid nephropathy. In addition, during massive cell lysis, phosphate and potassium are released in large amounts, resulting in hyperphosphatemia and hyperkalemia. In some patients, the precipitation of calcium and phosphate in the renal tubules can induce ARF independently of and in addition to uric acid deposition. The peak uric acid level often exceeds 20 mg/dl, and a ratio of urinary uric acid to creatinine concentrations greater than 1 to 1 suggests the diagnosis of acute uric acid nephropathy.

Prevention of ARF from tumor lysis syndrome involves establishing a urine output of greater than 3 to 5 L/24 hr and initiating treatment with allopurinol before institution of cytotoxic therapy. Establishing a high urinary output results in high intratubular pressure and thereby helps prevent intratubular obstruction. Allopurinol, which blocks the synthesis of uric acid by inhibiting xanthine oxidase, should be administered in dosages of 300 to 600 mg/day; therapy should begin 3 days before initiation of chemotherapy. In patients with underlying renal impairment, the allopurinol dose should be adjusted for renal function. In patients with initial hyperuricemia, allopurinol should be started and chemotherapy delayed until the serum uric acid concentration has become normal. Urinary alkalinization increases the solubility of xanthine and enhances its excretion. Urinary alkalinization can be achieved by the infusion of sodium bicarbonate in amounts sufficient to keep the urinary pH above 7 or by the administration of acetazolamide, which inhibits the reabsorption of sodium bicarbonate in the proximal tubule, thereby making the tubular fluid and the urine alkaline. Sodium bicarbonate and acetazolamide can be used in combination.

Rasburicase is a recombinant form of urate oxidase, which is a nonhuman proteolytic enzyme that oxidizes uric acid to allantoin. Rasburicase has been shown to reduce serum uric acid levels with associated diuresis more effectively and much faster than allopurinol, and it should be considered in the treatment of hyperuricemia and ARF.[44]

The development of oliguria with hyperuricemia may be an indication for dialysis. Early dialysis in these patients, started before severe renal impairment or uremia manifests, can help minimize further injury.

Hepatorenal Syndrome

The hepatorenal syndrome (HRS) is defined as kidney failure in patients with severely compromised liver function in the absence of clinical, laboratory, or anatomic evidence of other known causes of renal failure. HRS closely resembles prerenal failure, except that it does not respond to conventional volume replacement. The etiology of HRS is thought to involve significant reductions in renal perfusion associated with splanchnic vasodilatation.[45] In the United States and Europe, the great majority of HRS cases occurs in patients with advanced alcoholic cirrhosis. In persons who have cirrhosis with ascites, the probability of developing HRS is 18% at 1 year and 39% at 5 years.[46]

HRS may begin insidiously over a period of weeks to months or may appear suddenly and cause severe azotemia within days. Common precipitating causes are deterioration of liver function, sepsis, the use of nephrotoxic antibiotics or NSAIDs, overzealous use of diuretics, diarrhea, or GI bleeding. Large-volume paracentesis has the potential to initiate and exacerbate hepatorenal dysfunction. Support with intravenous albumin infusion during large-volume paracentesis may prevent circulatory dysfunction and decreased renal perfusion.[47] HRS can, however, occur without any apparent precipitating cause. The diagnosis is one of exclusion and should be suspected in any patient with advanced acute or chronic liver disease, portal hypertension, and progressive renal insufficiency associated with an increase in serum BUN and creatinine levels.

The initial step in management is to search diligently for and treat correctable causes of azotemia. All nephrotoxic agents should be discontinued. It is important to exclude reversible prerenal azotemia. Because HRS and prerenal azotemia produce similar urinary diagnostic indices, differentiating these two entities often requires a functional approach, such as volume expansion. Once a diagnosis of HRS is established, there is no specific treatment; management is conservative. If the patient is hypotensive, normalization of blood pressure may lead to improvement in renal perfusion and thus to stabilization of renal function. In patients with spontaneous bacterial peritonitis, intravenous albumin infusion (1.5 g/kg at the time of diagnosis followed by 1 g/kg on day 3) plus antibiotic treatment has proved more effective than antibiotic treatment alone in decreasing the incidence of renal impairment and death.[48]

Several agents have been investigated for pharmacologic treatment of hepatorenal syndrome; all are intended to increase renal blood flow. Octreotide is known to cause selective splanchnic vasoconstriction and may be beneficial for prolonged therapy and subcutaneous administration while patients await liver transplantation.[49] Midodrine, octreotide, and albumin therapy followed by a transjugular intrahepatic portosystemic stent shunt has been shown to be beneficial in a few cases of type 1 HRS, which is characterized by rapid and progressive renal impairment (doubling of the serum creatinine level to more than 2.5 mg/dl in less than 2 weeks, commonly with associated oliguria or anuria) and typically is precipitated by spontaneous bacterial peritonitis.[50] The combination of albumin infusion with the vasopressin analogue terlipressin has proved effective in reversing HRS in cirrhotic patients.[51]

The prognosis for patients with hepatorenal syndrome is poor unless their liver function can be improved. This typically requires liver transplantation.

ACUTE RENAL FAILURE RELATED TO PREGNANCY

In industrialized nations, ARF rarely occurs in association with pregnancy; the incidence is approximately one in 20,000 deliveries.[52] The low incidence is directly related to legalization of abortion in many countries.[52]

Infection after an abortion may precipitate ARF by causing hypotension, hemorrhage, sepsis, and disseminated intravascular coagulopathy. Hemolysis may result from toxin production. Although many organisms can be involved, *Clostridium* species are the most common pathogens, and cause the most serious of such infections. Aggressive management with broad-spectrum antibiotics and dialysis is the mainstay of therapy for these patients.

Urinary tract infections are among the most common medical complications of pregnancy and can progress to pyelonephritis. Approximately 25% of pregnant patients with pyelonephritis develop a transient decline in GFR.[53] These patients should be treated initially with intravenous antibiotics, followed by up to 2 weeks of therapy with oral antibiotics.

Third-trimester and peripartum complications of pregnancy that may lead to ARF include preeclampsia, postpartum hemorrhage, amniotic fluid embolism, placental abruption, and retained fetal or placental parts. A renal failure pattern similar to that of ATN is seen in patients with preeclampsia and peripartum hemorrhage. Bilateral cortical necrosis may follow any type of ischemic renal injury; the incidence seems disproportionately higher in pregnant patients than in nonpregnant adult patients. Although abruptio placentae can also cause ATN, renal cortical necrosis is the most common cause of ATN. The HELLP (hemolysis, elevated liver enzymes, low platelets) syndrome has been associated with ARF in up to 7.7% of pregnant patients.

Postpartum ARF, also known as postpartum hemolytic-uremic syndrome (HUS), is characterized by hypertension and microangiopathic hemolytic anemia. It can occur from postpartum day 1 to several months after delivery, but peak onset is from postpartum week 2 to week 5. Glomerular lesions resemble those found in adult HUS and are characterized by fibrin deposition, thickened capillary walls, and subendothelial swelling associated with large granular subendothelial deposits. Genetic factors may be involved in the etiology of postpartum HUS, given that women who have experienced one episode are at increased risk for recurrence with subsequent pregnancies, and occurrence of the disease has been documented in sisters. Postpartum HUS is associated with elevations in the lactate dehydrogenase (LDH) level; the HELLP syndrome, in contrast, is associated with elevations in transaminase levels. This difference is useful in distinguishing between these two syndromes. The mainstay of treatment for postpartum HUS is plasma exchange. Before the advent of plasma exchange therapy, mortality was 90%; current maternal survival rates are 70% to 80%.[54]

Other renal disorders may cause ARF during pregnancy. If active urine sediment occurs in a pregnant patient before the 20th week of gestation, acute glomerulonephritis should be considered. Lupus nephritis may develop in patients with a history of lupus nephritis or extrarenal lupus. Because complement levels are typically elevated during pregnancy, the finding of low complement levels in a pregnant woman (with additional supportive serology) is especially suggestive of lupus nephritis. Renal biopsy is not contraindicated during pregnancy. If a proliferative lesion is suspected, renal biopsy should be performed to facilitate early initiation of aggressive therapy. Any of the other prerenal and nephrotoxic causes of ARF can affect pregnant patients and should be considered. After the 20th week of gestation, ARF in a patient with hypertension is most likely the result of preeclampsia.

ACUTE INTERSTITIAL NEPHRITIS

AIN is often underdiagnosed as a cause of ARF. The diagnosis should not be overlooked, because this disorder requires specific intervention. Antibiotics are a major cause of AIN. This etiology is especially important to consider when a patient who is recovering from sepsis-induced ARF develops ARF while receiving antibiotics.

AIN is sometimes an immunologically induced hypersensitivity reaction to an antigen, usually a drug. AIN is detected in 2% to 3% of all renal biopsies but in up to 25% of renal biopsies that are performed specifically in the setting of drug-induced ARF. The drugs most commonly involved in the induction of AIN are NSAIDs and antibiotics[55] [*see 10:VIII Tubulointerstitial Diseases and Table 2*].

AIN is suggested by laboratory findings that indicate an abrupt deterioration of renal function and by urinalysis and urine sediment findings of blood, protein, and WBCs in clumps and in casts in the absence of evidence of infection. Clinically, the only complaint may be flank pain, caused by distention of the renal capsule; there may also be systemic signs and symptoms of a hypersensitivity reaction, such as rash, joint pain, and fever. The complete blood count may reveal eosinophilia.

Definitive diagnosis is made by renal biopsy. Noninvasive techniques suggestive of AIN, which may include an increased urinary eosinophil count (using Hansel stain) and a positive gallium scan, are not specific.

Management of Acute Renal Failure

Management of ARF entails close monitoring of the patient, supportive therapy, and targeted treatment of specific complications that may arise. The signs and symptoms of ARF reflect loss of the regulatory, excretory, and endocrine functions of the kidney. The loss of the excretory ability of the kidney is reflected in a rise in the plasma concentration of specific substances normally excreted by the kidney. The most widely monitored indices are the concentrations of BUN and creatinine in the serum. In patients without complications, BUN increases by 10 to 20 mg/dl/day, and bicarbonate decreases to a steady-state level of 16 to 18 mEq/L. The serum potassium concentration need not

rise appreciably unless the patient experiences a hypercatabolic state, GI bleeding, or extensive tissue trauma.

PREVENTION

The estimated cost of dialysis and aggressive care of a critically ill patient with ARF is approximately $128,000 per quality-adjusted life-year saved (QALYS). With the accepted upper limit of cost-effective care being $50,000 per QALYS, ARF places an overwhelming burden on the health care system.[56] Thus, prevention is the cornerstone of patient care.

Approximately 25% of all hospital-acquired cases of ARF are related to the use of one or more nephrotoxic agents.[7,57,58] Consequently, the best strategy for prevention of ARF is to avoid drug-related nephrotoxicity, especially in high-risk situations. Patients at risk are the elderly, those with preexisting renal disease, and those with volume depletion. The serum creatinine level is a poor marker of actual renal function, especially in elderly persons. An estimate of renal function is needed, especially in elderly patients, to permit proper management of the drug levels of all potential nephrotoxic agents and to enhance awareness of diminished renal function when the serum creatinine level falls within the normal range. The Modification of Diet in Renal Disease (MDRD) formula gives an acceptable estimate of GFR. Calculators for the MDRD GFR are available on the Internet (http://www.nkdep.nih.gov/healthprofessionals/tools). Measurement of serum cystatin C can help with recognition of the initial stages of renal impairment.[59]

Correcting fluid deficiencies before surgery and providing adequate hydration for patients who are particularly at risk before use of radiocontrast studies are useful measures. Nephrotoxic drugs should be used only when necessary; when such drugs are used, the patient should be monitored carefully. Pretreatment with allopurinol before chemotherapy of massive tumors diminishes uric acid excretion and thus lowers the potential for nephrotoxicity.

EMERGENT INTERVENTION

Hyperkalemia is a life-threatening complication of ARF that often necessitates urgent intervention [*see* Treatment of Complications, Hyperkalemia, *below*]. The electromechanical effects of hyperkalemia on the heart are potentiated by hypocalcemia, acidosis, and hyponatremia. Thus, the ECG, which measures the summation of these effects, is a better guide to therapy than a single determination of serum potassium. It must be emphasized that hyperkalemia is the biochemical abnormality that is most often responsible for death in patients with ATN. In contrast, moderate acidosis is generally well tolerated and usually does not require treatment; treatment may be necessary as adjunct therapy to control hyperkalemia or if the plasma bicarbonate level falls below 15 mEq/L.

SUPPORTIVE THERAPY

Because ATN is inherently a catabolic disorder, patients with ATN generally lose about 0.5 lb a day. Further weight loss may be minimized by providing adequate calories (1,800 to 2,500 kcal or 35 kcal/kg/day) and about 1.0 to 1.4 g protein/kg/day. The use of hyperalimentation with 50% dextrose and essential amino acids has had little effect on reducing mortality and morbidity in patients with ATN, except in patients who also have significant burns.

DIALYSIS

Indications for initiating dialysis are as follows: (1) severe hyperkalemia, acidosis that is not easily controlled by medical treatment, or both; (2) fluid overload that is not responsive to fluid restriction, diuretics, or both; and (3) signs or symptoms of uremia (e.g., a pericardial friction rub, asterixis, mental status changes that are not accounted for by some other disorder, or seizures). In the absence of any of these indications, most nephrologists advocate dialysis when the BUN level reaches 80 to 100 mg/dl, because the goal of modern therapy is to avoid the occurrence of uremic symptoms. Thus, the patient undergoes dialysis as frequently as necessary to keep the BUN level below 80 mg/dl. When this approach is used, most patients do not develop uremic symptoms, the diet and fluid intake can be liberalized, and the overall management of the patient is easier. It is critical to carefully review the indications for, and the doses of, all drugs administered to patients with ARF. Monitoring of blood concentrations of drugs is an important adjunct to effective treatment. Dialysis is discussed in detail elsewhere [*see 10:X Chronic Renal Failure and Dialysis*].

In a patient who receives appropriate therapy with early dialysis, many of the uremic manifestations associated with ARF either do not develop or are minimal. Infection remains the main cause of death despite vigorous dialysis. Thus, meticulous aseptic care of intravenous catheters and wounds and avoidance of the use of indwelling urinary catheters are important in the management of such patients.

DOPAMINE

There are no clinical data to support the use of low-dose dopamine for the protection or improvement of renal function in patients with ATN or ARF. In a multicenter study comparing dopamine with placebo, dopamine was not found to improve survival or eliminate the need for dialysis.[60] In both animal and human studies, use of low-dose dopamine was found to be associated with an increase in renal blood flow and natriuresis in the context of euvolemia and normal renal function.

TREATMENT OF COMPLICATIONS

Volume Overload

Volume overload is one of the first manifestations of ARF caused by salt and water retention secondary to a decrease in GFR. In addition, volume overload may be exacerbated by intravenous fluids given in an aggressive attempt to reverse ARF or to treat oliguria. Consequently, volume status needs to be evaluated daily in patients with ARF. This evaluation includes assessment of weight, blood pressure, and heart rate. Physical examination is also important and should be directed specifically toward the detection of skin turgor, peripheral edema, pulmonary edema, and a third heart sound.

Records of the patient's daily fluid intake and output should be reviewed, but insensible fluid losses, as well as weight loss resulting from a high catabolic state, must be taken into consideration.[61,62] A normal adult loses about 200 to 300 g of body weight a day as a result of catabolism, and the average 70 kg person has insensible losses of around 850 to 1,000 ml/day. With fever, the insensible water loss increases by about 13% for each degree centigrade rise above normal (7% for each degree Fahrenheit). Daily assessment and modification of fluid therapy is essential. Water is continuously generated from endogenous sources by the oxidation of protein (41 ml/100 g), fat (107 ml/100 g), and carbohydrates (55 ml/100 g). Adequate carbohydrates will help reduce protein metabolism, with a small decrease in water generation.

The most useful therapy for volume overload is loop diuretics.[63,64] Furosemide can be given intravenously in a bolus or by continuous infusion [*see Table 8*]. If started in the early stages of ARF, this intervention, along with fluid restriction, can be very beneficial in preventing or minimizing volume overload. The main risk of using high doses of furosemide is deafness, which can be transient or permanent.[65] This risk is increased if the serum albumin level is low and if other ototoxic medications are being used simultaneously with furosemide.[66] There is no good evidence that the use of diuretics alters the course of ATN; however, in patients with fluid overload, diuretics may be useful in increasing urine output and preventing the need for dialysis. If diuretic treatment does not correct volume overload and the patient remains oliguric, early intervention with dialysis should be considered. However, there is evidence that outcomes may be worse in patients who receive diuretics than in patients who do not.[67]

Table 8 Furosemide Dosing

Bolus dosing	Initially, 20 mg I.V.; if no response in 30 min to 1 hr, double the dose to 40 mg If no response, increase the dose slowly until a response is achieved Give repeat doses every 8 to 12 hr to maintain diuresis Single bolus doses should not exceed 240 mg
Continuous infusion	Safer and more effective than bolus dosing in patients with refractory disease Initial I.V. bolus of 0.1 mg/kg (or a bolus dose previously determined to initiate diuresis) Continuous infusion of 0.1 mg/kg/hr; double every 2 hr to a maximum of 0.4 mg/kg/hr *or* After an initial bolus, continuous infusion of 20 mg/hr If diuresis is not sustained, give a second, larger bolus and increase the infusion rate to 40 mg/hr Risk of toxicity increases with continuous infusion rates above 80 mg/hr
For refractory cases	Consider the addition of an intravenous thiazide diuretic Consider fluid removal via hemofiltration (dialysis)

Hyponatremia

Hyponatremia is a common problem in patients who have ARF that is related to a decrease in GFR and impaired tubular function. Usually, such hyponatremia is associated with hypervolemia. Sources of excess water intake include the administration of hypotonic solutions such as D5W or free-water intake through enteral or parenteral feeding.

The clinical manifestations of hyponatremia are primarily neurologic in nature. Symptoms are related to cell swelling and may include headache, behavioral disturbances, lethargy, ataxia, and seizures; symptoms can progress to coma, respiratory depression, and death.

Symptomatic hyponatremia should be treated aggressively, but also with caution, because overly aggressive correction of a low serum sodium level can lead to central pontine myelinolysis if the duration of electrolyte imbalance has been longer than 48 hours [*see 11:VII Anoxic, Metabolic, and Toxic Encephalopathies*]. The initial approach in a patient who has volume overload and hyponatremia is to administer loop diuretics and restrict free water. If symptoms of hyponatremia occur in a patient with ARF that is unresponsive to diuretics or saline replacement, the removal of free water may be warranted as the initial step in renal replacement therapy. Correction of hyponatremia should proceed gradually, with the rise in sodium level targeted at a rate of 1 to 2 mEq/L/hr until symptoms resolve or until the serum sodium level approaches 120 mEq/L. Therefore, continuous monitoring of the patient's condition and measurement of serum electrolytes every 1 to 2 hours is warranted.

Hyperkalemia

Hyperkalemia is common in patients with ARF. Causal factors include a decrease in GFR, a low rate of urinary flow to the collecting duct, distal tubular damage, and concomitant conditions that contributed to or are associated with ARF, such as rhabdomyolysis, acidosis, and the hypercatabolic state. ACE inhibitors and NSAIDs may also play a role.

A small portion of the total body potassium—1.5% to 2.5%, or around 60 mEq—is found in the extracellular compartment. The intracellular compartment has a potassium concentration about 38-fold higher than that of the extracellular compartment. Thus, any process that leads to cellular destruction can contribute to hyperkalemia, especially in patients with ARF.

The cardiac effects of hyperkalemia are primarily associated with the blunting of the magnitude of the action potential in response to a depolarizing stimulus. The sequential ECG changes observed in hyperkalemia are peaked T waves, prolongation of the PR interval, widening of the QRS complex, and a sine wave pattern; these findings call for prompt treatment. Other symptoms consistent with hyperkalemia include paresthesias, muscular weakness, and depressed deep tendon reflexes. These symptoms can progress to flaccid paralysis and acute respiratory failure. Severe hyperkalemia is life-threatening and should be considered a medical emergency.

The treatment of hyperkalemia associated with ECG changes starts with the administration of calcium chloride or calcium gluconate, which will work to immediately antagonize the effect of potassium on the heart. Insulin and glucose, inhaled beta-adrenergic receptor agonists, or intravenous sodium bicarbonate in patients known to have acidosis will help redistribute potassium to the intracellular space. These measures require 30 to 60 minutes to take effect and have a short duration of action. In patients with renal failure, the only certain way of lowering the potassium is to remove it from the body by using sodium polystyrene sulfonate (orally or as an enema), dialysis, or both.

Dietary intake of potassium should be restricted to less than 50 mEq/day; intravenous or oral sources of potassium should be discontinued, and medications that inhibit potassium excretion (e.g., potassium-sparing diuretics, NSAIDs, ACE inhibitors, and angiotensin receptor blockers) should be stopped.

Acidosis

A healthy adult produces 1 mEq/kg/day of hydrogen ions; this production is influenced to a large degree by dietary intake. The kidney plays a major role in acid-base balance through the excretion of nonvolatile acids and by the reabsorption and regeneration of bicarbonate. In patients with ARF, the serum bicarbonate level decreases by 1 to 2 mEq/day and typically stabilizes around 16 to 18 mEq/L. A more severe drop may be precipitated by a hypercatabolic state, infection, inadequate nutrition, and other causes of metabolic acidosis.

Clinical manifestations of acidosis are partly related to the rapidity with which the acidosis develops. When metabolic acidosis develops rapidly, the body's attempt to compensate by blowing off carbon dioxide may result in Kussmaul respiration, which is characterized by deep inspiration and a normal or reduced respiratory rate. Other clinical consequences include cardiac arrhythmias, depressed myocardial contractility, peripheral vasodilatation, abdominal pain, nausea and vomiting, headache and lethargy, and an increased catabolic rate.

Metabolic acidosis is defined as a low arterial pH in conjunction with a low serum bicarbonate level. Mild to moderate metabolic acidosis with a pH greater than 7.2 can be treated with oral sodium bicarbonate. Severe metabolic acidosis with a pH of 7.1 or less that is associated with ARF should be treated more aggressively, with intravenous sodium bicarbonate, renal replacement therapy, or both; continuous renal replacement therapy is the preferred approach. Because intravenous sodium bicarbonate replacement poses many risks, critically ill patients who receive it require close monitoring. If needed, intravenous bicarbonate can be administered as an isotonic solution by adding 3 ampules of sodium bicarbonate (50 mEq/50 ml) to 1 L of D5W; the solution is infused at a variable but slow rate, depending on the patient's condition and volume status. Particular attention should be given to cardiac, respiratory, and hemodynamic status. Serum electrolyte levels, including the serum calcium level, should be monitored and corrected as needed.

Early in the recovery phase of ARF, there may be modest reductions in GFR and in the ability to concentrate and acidify the urine. Thus, patients may continue to need support with intravenous fluids and oral or intravenous bicarbonate.

Calcium and Phosphate Imbalances

Hypocalcemia is a common finding in patients with ARF, but it rarely requires intervention. In patients with ARF, hypocalcemia can result from several mechanisms, the most common being hyperphosphatemia. Resistance to parathyroid hormone, altered vitamin D metabolism, and calcium sequestration in tissues all play a role. Transfusions of blood products that have been stored in citrate and infusions of sodium bicarbonate can contribute to a decrease in the serum calcium level. Calcium supplementation should be undertaken only if it is clinically indicated and should be carried out with caution.

Hypercalcemia is rare in patients with ARF and may be indicative of other underlying medical conditions. The most common underlying disorders are malignancies, particularly multiple myeloma. Hypercalcemia occurs in about 30% of patients during the recovery phase of ARF-associated rhabdomyolysis.

Hyperphosphatemia is common in patients with ARF. The elevation in the phosphate level results primarily from a decrease in renal excretion; tissue destruction and shifts from the intracellular to extracellular space as a result of acidosis and catabolism can be contributing factors. In patients who can ingest food, hyperphosphatemia should be treated with phosphate binders; these agents prevent absorption of dietary phosphorus into the bloodstream. Phosphate binders include calcium salts (calcium carbonate or calcium acetate), aluminum hydroxide, and sevelamer, which is a cationic polymer that binds phosphate through ion exchange. If the calcium-times-phosphorus product is high (i.e., > 70), a non–calcium-based binder such as sevelamer should be administered to minimize metastatic calcification in soft tissues.

Anemia

A normochromic-normocytic anemia develops in 65% to 95% of patients with ARF. These patients have normal serum iron levels and normal or hypercellular bone marrow. An increase in the rate of destruction of RBCs as a result of increased erythrocyte fragility may play a role in the early decrease in the hemoglobin level in patients with ARF. The most important cause of ARF-associated anemia is inadequate production of erythropoietin, but in critically ill patients, a decrease in erythropoietin responsiveness also plays a role.

Blood loss may contribute to anemia in patients with ARF. These patients have an increased tendency to bleed because of platelet dysfunction secondary to azotemia. Conjugated estrogen and 1-desamino-8-D-arginine vasopressin (DDAVP or desmopressin) may help correct this bleeding tendency. DDAVP is given at a dosage of 0.3 μg/kg infused as an intravenous piggyback over 20 to 30 minutes.[68]

Because anemia can produce significant complications in critically ill patients, hemoglobin levels should be maintained above 10 g/L. Erythropoietin is increasingly used in such patients to reverse anemia and to reduce the need for blood transfusions [*see 5:III Anemia: Production Defects*].

Prognosis of Patients with Acute Renal Failure

Mortality can be as high as 50% to 80% in patients who have ARF in association with sepsis, hypotension, and respiratory failure.[57] The prognosis for hospitalized patients with ARF depends largely on the setting (i.e., hospital ward or ICU); the presence of comorbidities; the underlying cause of the renal failure; the severity of the renal failure; and how early in its course the condition is diagnosed and treatment initiated. In a review of 22,589 patients who underwent open heart surgery, mortality was 61.2% in those with ARF requiring dialysis, 14.1% in those with ARF not requiring dialysis, and only 0.68% in those who did not develop ARF.[5] In a study of ICU patients with ARF, 64% required dialysis; in-hospital mortality was 37%, and 50% of patients died or failed to recover renal function.[4]

In hospitalized patients with ARF caused by ATN, about one fourth to one third of deaths occur during the diuretic phase. This is not surprising, because with the availability of dialysis, the most important determinant of the outcome is not the uremia itself but, rather, the underlying disease causing the ATN. Infection is a common cause of death in patients with ATN-related ARF. In patients who survive the acute episode, renal function essentially returns to normal.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Liano F, Pascual J: Epidemiology of acute renal failure: a prospective, multicenter, community-based study. Madrid Acute Renal Failure Study Group. Kidney Int 50:811, 1996

2. Kaufman J, Dhakal M, Patel B, et al: Community-acquired acute renal failure. Am J Kidney Dis 17:191, 1991

3. Thadhani R, Pascual M, Bonventre JV: Acute renal failure. N Engl J Med 334:1448, 1996

4. Mehta RL, Pascual MT, Soroko S, et al: Spectrum of acute renal failure in the intensive care unit: the PICARD experience. Kidney Int 66:1613, 2004

5. Thakar CV, Liangos O, Yared JP, et al: ARF after open-heart surgery: influence of gender and race. Am J Kidney Dis 41:742, 2003

6. Doherty CC: Epidemiology of acute renal failure. Oxford Textbook of Clinical

Nephrology. Davison AM, Cameron JS, Grunfeld JP, et al, Eds. Oxford University Press, Oxford, 1998, p 1521

7. Nolan CR, Anderson RJ: Hospital-acquired acute renal failure. J Am Soc Nephrol 9:710, 1998

8. Blantz RC: Pathophysiology of pre-renal azotemia. Kidney Int 53:512, 1998

9. Bonventre JV: Pathophysiology of ischemic acute renal failure: inflammation, lung-kidney cross-talk, and biomarkers. Contrib Nephrol 144:19, 2004

10. Kwon O, Corrigan G, Myers BD, et al: Sodium reabsorption and distribution of Na+/K+-ATPase during postischemic injury to the renal allograft. Kidney Int 55:963, 1999

11. Li S, Gokden N, Okusa MD, et al: Anti-inflammatory effect of fibrate protects from cisplatin-induced ARF. Am J Physiol (in press)

12. Sheridan AM, Bonventre JV: Cell biology and molecular mechanisms of injury in ischemic acute renal failure. Curr Opin Nephrol Hypertens 9:427, 2000

13. Carvounis CP, Nisar S, Guro-Razuman S: Significance of the fractional excretion of urea in the differential diagnosis of acute renal failure. Kidney Int 62:2223, 2002

14. Liano F, Pascual J: Course and prognosis of acute renal failure in the very old: results of a multi-centre survey in Madrid, Spain (abstr). J Am Soc Nephrol 5:400, 1994

15. Berg KJ: Nephrotoxicity related to contrast media. Scand J Urol Nephrol 34:317, 2000

16. McCullough PA, Sandberg KR: Epidemiology of contrast-induced nephropathy. Rev Cardiovasc Med 4(suppl 5):S3, 2003

17. Gruberg L, Mehran R, Dangas G, et al: Acute renal failure requiring dialysis after percutaneous coronary interventions. Catheter Cardiovasc Interv 52:409, 2001

18. Kandzari DE, Rebeiz AG, Wang A, et al: Contrast nephropathy: an evidence-based approach to prevention. Am J Cardiovasc Drugs 3:395, 2003

19. Barrett BJ, Carlisle EJ: Metaanalysis of the relative nephrotoxicity of high- and low-osmolality iodinated contrast media. Radiology 188:171, 1993

20. Erdogan A, Davidson CJ: Recent clinical trials of iodixanol. Rev Cardiovasc Med 4(suppl 5):S43, 2003

21. Mueller C, Buerkle G, Buettner HJ, et al: Prevention of contrast media–associated nephropathy: randomized comparison of 2 hydration regimens in 1620 patients undergoing coronary angioplasty. Arch Intern Med 162:329, 2002

22. Erley CM, Duda SH, Rehfuss D, et al: Prevention of radiocontrast-media–induced nephropathy in patients with pre-existing renal insufficiency by hydration in combination with the adenosine antagonist theophylline. Nephrol Dial Transplant 14:1146, 1999

23. Morcos SK, Thomsen HS, Webb JA: Contrast-media–induced nephrotoxicity: a consensus report. Contrast Media Safety Committee, European Society of Urogenital Radiology (ESUR). Eur Radiol 9:1602, 1999

24. Misra D, Leibowitz K, Gowda RM, et al: Role of *N*-acetylcysteine in prevention of contrast-induced nephropathy after cardiovascular procedures: a meta-analysis. Clin Cardiol 27:607, 2004

25. Gill NK, Piccione EA, Vido DA, et al: Gender as a risk factor for contrast nephropathy: effects of hydration and N-acetylcysteine. Clin Cardiol 27:554, 2004

26. Stone GW, McCullough PA, Tumlin JA, et al: Fenoldopam mesylate for the prevention of contrast-induced nephropathy: a randomized controlled trial. CONTRAST Investigators. JAMA 290:2284, 2003

27. Merten GJ, Burgess WP, Gray LV, et al: Prevention of contrast-induced nephropathy with sodium bicarbonate: a randomized controlled trial. JAMA 291:2328, 2004

28. Swan SK: Aminoglycoside nephrotoxicity. Semin Nephrol 17:27, 1997

29. Streetman DS, Nafziger AN, Destache CJ, et al: Individualized pharmacokinetic monitoring results in less aminoglycoside-associated nephrotoxicity and fewer associated costs. Pharmacotherapy 21:443, 2001

30. Prins JM, Buller HR, Kuijper EJ, et al: Once versus thrice daily gentamicin in patients with serious infections. Lancet 341:335, 1993

31. Bartal C, Danon A, Schlaeffer F, et al: Pharmacokinetic dosing of aminoglycosides: a controlled trial. Am J Med 114:194, 2003

32. Garella S, Matarese RA: Renal effects of prostaglandins and clinical adverse effects of nonsteroidal anti-inflammatory agents. Medicine (Baltimore) 63:165, 1984

33. Bender WL, Whelton A, Beschorner WE, et al: Interstitial nephritis, proteinuria, and renal failure caused by nonsteroidal anti-inflammatory drugs: immunologic characterization of the inflammatory infiltrate. Am J Med 76:1006, 1984

34. Radford MG Jr, Holley KE, Grande JP, et al: Reversible membranous nephropathy associated with the use of nonsteroidal anti-inflammatory drugs. JAMA 276:466, 1996

35. Henrich WL, Agodoa LE, Barrett BJ, et al: Analgesics and the kidney: summary and recommendations to the Scientific Advisory Board of the National Kidney Foundation from an ad hoc committee of the National Kidney Foundation. Am J Kidney Dis 27:162, 1996

36. Arany I, Safirstein RL: Cisplatin nephrotoxicity. Semin Nephrol 23:460, 2003

37. Bakris GL, Weir MR: Angiotensin-converting enzyme inhibitor–associated elevations in serum creatinine: is this a cause for concern? Arch Intern Med 160:685, 2000

38. Sadjadi SA: A new range for the anion gap. Ann Intern Med 123:807, 1995

39. Moreau CL, Kerns W 2nd, Tomaszewski CA, et al: Glycolate kinetics and hemodialysis clearance in ethylene glycol poisoning. META Study Group. J Toxicol Clin Toxicol 36:659, 1998

40. Poldelski V, Johnson A, Wright S, et al: Ethylene glycol–mediated tubular injury: identification of critical metabolites and injury pathways. Am J Kidney Dis 38:339, 2001

41. Knochel JP: Nontraumatic rhabdomyolysis. Acute Renal Failure: A Companion to Brenner & Rector's The Kidney. Molitoris BA, Finn WF, Eds. WB Saunders, Philadelphia, 2001, p 220

42. Better OS, Stein JH: Early management of shock and prophylaxis of acute renal failure in traumatic rhabdomyolysis. N Engl J Med 322:825, 1990

43. Zager RA: Rhabdomyolysis and myohemoglobinuric acute renal failure. Kidney Int 49:314, 1996

44. Ronco C, Inguaggiato P, Bordoni V, et al: Rasburicase therapy in acute hyperuricemia and renal dysfunction. Contrib Nephrol 147:115, 2005

45. Arroyo V: Pathophysiology, diagnosis and treatment of ascites in cirrhosis. Ann Hepatol 1:72, 2002

46. Gines A, Escorsell A, Gines P, et al: Incidence, predictive factors, and prognosis of the hepatorenal syndrome in cirrhosis with ascites. Gastroenterology 105:229, 1993

47. Garcia-Compean D, Blanc P, Larrey D, et al: Treatment of cirrhotic tense ascites with dextran-40 versus albumin associated with large volume paracentesis: a randomized controlled trial. Ann Hepatol 1:29, 2002

48. Sort P, Navasa M, Arroyo V, et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 341:403, 1999

49. Kaffy F, Borderie C, Chagneau C, et al: Octreotide in the treatment of the hepatorenal syndrome in cirrhotic patients. J Hepatol 30:174, 1999

50. Wong F, Pantea L, Sniderman K: Midodrine, octreotide, albumin, and TIPS in selected patients with cirrhosis and type 1 hepatorenal syndrome. Hepatology 40:55, 2004

51. Uriz J, Gines P, Cardenas A, et al: Terlipressin plus albumin infusion: an effective and safe therapy of hepatorenal syndrome. J Hepatol 33:43, 2000

52. Nzerue CM, Hewan-Lowe K, Nwawka C: Acute renal failure in pregnancy: a review of clinical outcomes at an inner-city hospital from 1986–1996. J Natl Med Assoc 90:486, 1998

53. Gilstrap LC 3rd, Ramin SM: Urinary tract infections during pregnancy. Obstet Gynecol Clin North Am 28:581, 2001

54. Takahashi Y, Imai A, Hayasaki Y, et al: Postpartum microangiopathic hemolytic anemia: cases of successful and dismal outcome assisted with plasma therapy. Eur J Obstet Gynecol Reprod Biol 89:213, 2000

55. Rossert J, Fischer EA: Acute interstitial nephritis. Comprehensive Clinical Nephrology. Johnson RJ, Feehally J, Eds. Mosby, London, 2000, p 62.1

56. Hamel MB, Phillips RS, Davis RB, et al: Outcomes and cost-effectiveness of initiating dialysis and continuing aggressive care in seriously ill hospitalized adults. Ann Intern Med 127:195, 1997

57. Brivet FG, Kleinknecht DJ, Loirat P, et al: Acute renal failure in intensive care units—causes, outcome, and prognostic factors of hospital mortality: a prospective, multicenter study. French Study Group on Acute Renal Failure. Crit Care Med 24:192, 1996

58. Liano F, Pascual J: Epidemiology of acute renal failure: a prospective, multicenter, community-based study. Madrid Acute Renal Failure Study Group. Kidney Int 50:811, 1996

59. Herget-Rosenthal S, Marggraf G, Husing J, et al: Early detection of acute renal failure by serum cystatin C. Kidney Int 66:1115, 2004

60. Chertow GM, Sayegh MH, Allgren RL, et al: Is the administration of dopamine associated with adverse or favorable outcomes in acute renal failure? Auriculin Anaritide Acute Renal Failure Study Group. Am J Med 101:49, 1996

61. Leverve X, Barnoud D: Stress metabolism and nutritional support in acute renal failure. Kidney Int 66:S62, 1998

62. Druml W, Mitch WE: Nutritional management of acute renal failure. Therapy in Nephrology and Hypertension. Brady HR, Wilcox CS, Eds. WB Saunders, Philadelphia, 1999, p 65

63. Unwin RJ, Capasso G, Wilcox CS: Therapeutic use of diuretics. Therapy in Nephrology and Hypertension. Brady HR, Wilcox CS, Eds. WB Saunders, Philadelphia, 1999, p 654

64. Kellum JA: Use of diuretics in the acute care setting. Kidney Int 66:S67, 1998

65. Ellison DH, Wilcox CS: Diuretic resistance. Therapy in Nephrology and Hypertension. Brady HR,Wilcox CS, Eds. WB Saunders, Philadelphia, 1999, p 665

66. Suki WN: Use of diuretics in chronic renal failure. Kidney Int 59:S33, 1997

67. Mehta RL, Pascual MT, Soroko S, et al: Diuretics, mortality, and nonrecovery of renal function in acute renal failure. JAMA 288:2547, 2002

68. Soslau G, Schwartz AB, Putatunda B, et al: Desmopressin-induced improvement in bleeding times in chronic renal failure patients correlates with platelet serotonin uptake and ATP release. Am J Med Sci 300:372, 1990

VII VASCULAR DISEASES OF THE KIDNEY

Abhijit V. Kshirsagar, m.d., m.p.h.
Ronald J. Falk, m.d.

A number of local and systemic disease processes affect the renal vascular tree [*see Figure 1*]. Although the underlying mechanisms may differ, vascular diseases of the kidney all share an important feature: they characteristically cause varying degrees of vessel obstruction, eventually leading to an impairment of renal blood flow. Reduction of renal blood flow often leads to two important physiologic adaptations: elevated blood pressure and reduced glomerular filtration rate (GFR). Hypertension and renal insufficiency are often important clues to the diagnosis of renal vascular disease. Rapid diagnosis is necessary to prevent the development of irreversible kidney disease, including end-stage renal disease (ESRD).

This chapter discusses the major vascular diseases of the kidney, categorized according to the size of the affected vessel. Large vessels include the main renal arteries and their major branches, as well as the renal vein and its main branches. Medium-sized vessels refer mainly to arteries. Small vessels include arterioles, capillaries, and venules.

Large Vessel Diseases

RENAL ARTERY STENOSIS

Renal artery stenosis (RAS) refers to a group of diseases characterized by narrowing of the main renal artery and its major branches. Atherosclerosis and fibromuscular dysplasia (FMD) are the two major causes of RAS; atherosclerosis is responsible for over 90% of cases, with FMD accounting for almost all of the remainder.

Atherosclerotic Renal Artery Stenosis

Atherosclerotic RAS has come to be defined as lumen narrowing of at least 50%, although some researchers have advocated that significant narrowing is 70% or greater. Atherosclerosis of the renal artery is a relatively common finding in both men and women after the fifth decade of life, but the exact prevalence of atherosclerotic RAS is not known, because there are no economical screening tests for the general population. Screening for RAS in highly selected populations consisting of patients with a high burden of atherosclerosis (i.e., those who have diabetes, who have recently had a stroke or myocardial infarction, or who have peripheral vascular disease) has revealed a prevalence of 10% to 30%.[1-4] Although early reports suggested that there was a higher prevalence in whites, current evidence suggests that atherosclerotic RAS is also found in significant numbers of African Americans.[5,6] There are no reported gender differences in prevalence. Risk factors for the development of atherosclerotic RAS include hypercholesterolemia, advanced age, cigarette smoking, hypertension, and diabetes mellitus.

The lesions of atherosclerosis occur primarily in the large and medium-sized arteries. They are characterized by the infiltration of the intimal layer by macrophages and monocytes and the deposition of oxidized low-density lipoprotein. Most cases of atherosclerotic RAS involve the opening of the renal artery from the aorta (ostium), the proximal third of the main renal artery, or the perirenal aorta. Segmental and diffuse intrarenal atherosclerosis may be observed in advanced cases of RAS, especially in patients with diabetes mellitus.

Renal Artery Stenosis from Fibromuscular Dysplasia

FMD refers to a group of uncommon diseases known to cause stenosis of several vascular beds, notably the renal and carotid arteries. They are characterized by noninflammatory and non-

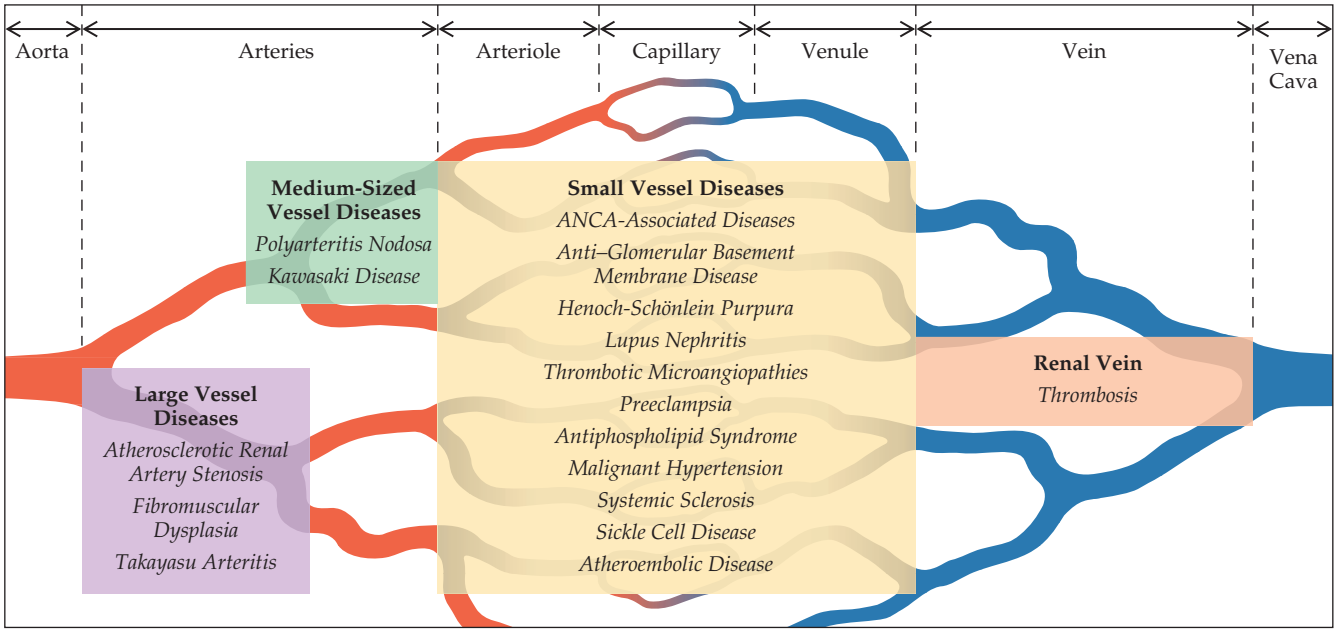


Figure 1 **Overview of vascular disease of the kidney. (ANCA—antineutrophil cytoplasmic autoantibody)**

atherosclerotic dysplastic changes of any of the three layers of the artery—the intima, media, or adventitia. The prevalence of FMD is not clearly known, but the condition is probably rare. FMD accounts for about 10% of all cases of RAS. In turn, RAS is thought to cause less than 5% of all cases of hypertension.

The renal vasculature is most frequently affected by FMD; renovascular FMD accounts for 60% to 75% of all cases of FMD. Bilateral renal involvement occurs in 35% of all cases of renal involvement. FMD of the renal vessels tends to affect women younger than 50 years—typically, younger than 30 years. The cause of FMD is not currently known.

Medial fibroplasia, the most common subtype of FMD, is characterized by a predominance of fibrotic material in the media, with sparing of the intima and adventitia. It affects the distal two thirds of the main renal artery and its branches. The segmental nature of medial fibroplasia, together with localized poststenotic dilatation, result in the classic so-called beads-on-a-string appearance on angiography.

Perimedial fibroplasia is characterized by the appearance of an area of elastic tissue circumscribing the entire diameter of the vessel at the junction of the media and adventitia. Angiographically, perimedial fibroplasia typically appears as focal stenosis, with the development of extensive collateral circulation.

Medial hyperplasia, the last type of medial FMD, is exceedingly rare and is characterized by proliferation of smooth muscle cells. Angiographically, medial hyperplasia is indistinguishable from intimal fibroplasia.

Intimal fibroplasia is characterized by a predominance of fibrotic material in the intima. It is the second most common type of FMD, accounting for about 10% of all cases. It affects the midportion of the main renal artery and its branches. Angiographic evaluation typically reveals long, smooth narrowing or focal, concentric stenosis.

Adventitial hyperplasia is often termed periarterial because of the hyperplastic changes that seem to surround the entire artery. An extremely rare disorder, it is reported to affect the middle to the distal portions of the renal artery.

Pathophysiology

Narrowing of the renal arteries may lead to two important clinical consequences: elevated blood pressure and renal insufficiency. The mechanism of elevated blood pressure is akin to that demonstrated by Goldblatt in the 1930s with the so-called one-clip model, in which ligation of one renal artery led to the production of vasoactive hormones distal to the stenosis and an increase in systemic arterial pressure.[7]

The mechanisms of renal insufficiency are not clear; renal insufficiency is not merely the direct result of decreased renal blood flow. Experimental models offer an imperfect guide, because they use methods that induce acute narrowing rather than the progressive narrowing that occurs with atherosclerosis and FMD. In addition, the clustering of atherosclerosis with other risk factors for renal insufficiency—notably, diabetes mellitus—makes it difficult to ascribe the renal insufficiency to atherosclerotic RAS alone. When renal insufficiency occurs in the setting of atherosclerotic RAS, it is often termed ischemic nephropathy.

Diagnosis

Certain clinical and laboratory findings are suggestive of RAS. These include the abrupt onset or worsening of hypertension, hypokalemia, abdominal bruits, peripheral vascular disease, carotid artery disease, and various renal abnormalities. The renal abnormalities include unexplained azotemia, azotemia induced by the use of angiotensin-converting enzyme (ACE) inhibitors or angiotensin II receptor blockers (ARBs), and asymmetrical kidney size on radiologic testing. In addition, recurrent episodes of flash pulmonary edema may be associated with bilateral RAS.

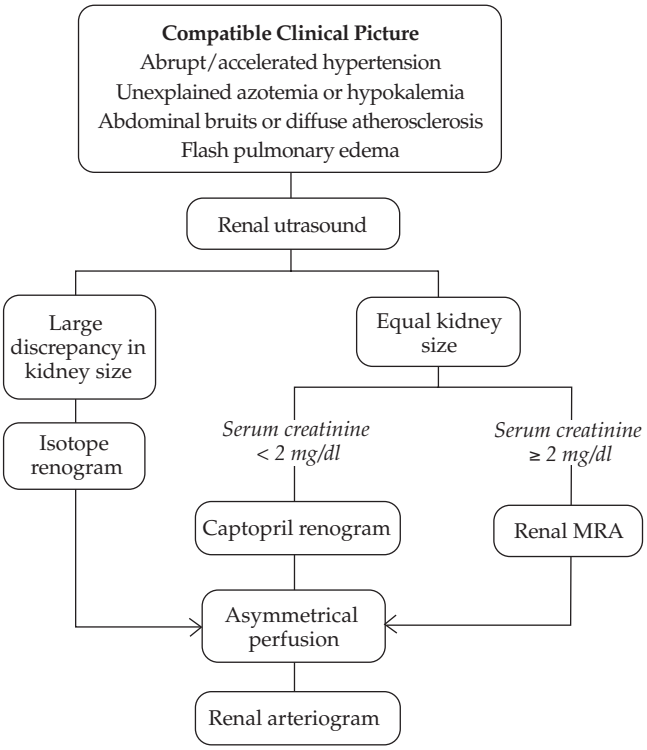


Figure 2 **Workup of suspected renal artery stenosis. (MRA—magnetic resonance angiogram with gadolinium)**

In patients with a compatible clinical picture, evaluation for RAS starts with renal ultrasonography to measure kidney size [*see Figure 2*]. If there is a large discrepancy in kidney sizes (i.e., > 2 cm in the longitudinal axis), significant arterial stenosis is or was likely present. An isotope renogram can then be performed to determine the relative perfusion to each kidney. If asymmetrical perfusion is seen, invasive treatment may be indicated, although this is a matter of debate [*see* Treatment, *below*].

Even if the ultrasound scan shows that the kidneys are equal in size, further diagnostic testing is required. The choice of procedures is determined by the level of renal function: patients with a serum creatinine level below 2 mg/dl should undergo renography; those with a serum creatinine of 2 mg/dl or greater should undergo magnetic resonance angiography (MRA). The captopril renogram, using a variety of radioisotopes, has a sensitivity of 93% and a specificity of 70% for RAS.[8] Gadolinium-enhanced MRA has a sensitivity of 88% and a specificity of 98% for RAS.[9]

The gold standard for the diagnosis of RAS remains a renal arteriogram. Given the known risks of contrast media exposure and manipulation of atherosclerotic plaques, it is important to limit renal arteriography to patients with positive results on noninvasive renal testing.

Treatment

Atherosclerotic RAS Radiographic studies have demonstrated that the lesions of atherosclerotic RAS progress, yet functional studies that measure GFR or surrogate indices have not consistently demonstrated progressive renal insufficiency. It is

therefore not clear whether intervention alters the natural course of the disease process. Thus, the choice between medical management and intervention for atherosclerotic RAS remains controversial. Medical management with various antihypertensive agents remains the most common method of managing hypertension and renal insufficiency. Agents that preferentially reduce efferent arterial tone—ACE inhibitors and ARBs—are contraindicated because of the risk of inducing acute renal failure. Percutaneous angioplasty has been presented as a favorable alternative, but numerous studies have demonstrated its equivalence to conventional pharmacologic therapy.[10] Angioplasty with stenting has been touted as technically superior to angioplasty alone,[11] but a comprehensive review showed that this procedure cures hypertension in fewer than 20% of patients and that it provides no apparent benefit in renal function.[12] Similarly, long-term follow-up studies have failed to demonstrate that surgical revascularization of the stenotic lesion is superior to pharmacologic therapy.[13] Thus, a consensus has been forming that pharmacologic therapy is probably equivalent to more invasive interventions for the vast majority of patients with atherosclerotic RAS.

Some researchers have attempted to determine whether certain variables predict success or failure with intervention in atherosclerotic RAS. Factors that correlate with a lack of improvement include unilateral involvement, stenosis less than 50% of the luminal diameter, baseline kidney size of less than 8 to 9 cm longitudinally, diabetic nephropathy, and severe proteinuria. Recent interest has focused on the resistive index, an estimate of the cumulative resistance of the segmental arteries, and possibly the smaller vessels. Univariate analysis has demonstrated that a high resistive index is a better predictor than level of renal function, kidney size, and proteinuria for a poor outcome with RAS revascularization.[14]

Fibromuscular dysplasia Most forms of FMD progress radiographically, but it is unclear whether there is any associated development of renal insufficiency. In medial FMD, 30% of patients have progressive narrowing without dissection or thrombosis. Both intimal and periarterial FMD are associated with dissection and thrombosis and, rarely, acute renal infarction. Percutaneous intervention has been the standard of care, but large comparative trials are not feasible, given the relative rarity of these conditions. Angioplasty and stenting completely cure hypertension in about 22% of patients.[15] Surgical revascularization has a higher cure rate, about 79% at 1 year,[16] but probably causes more short-term morbidity than angioplasty. Surgical repair of aneurysms (the "beads" seen on arteriography) is required if their diameter is greater than 1.5 cm and if patients have uncontrolled hypertension or are pregnant.

TAKAYASU ARTERITIS

Takayasu arteritis is an inflammatory arteritis involving the aorta and its branches. It primarily affects the ascending aorta and its arterial branches (the carotid, subclavian, and axillary arteries). Involvement of the lower branches of the aorta, including the main renal artery and its branches, occurs as the disease process progresses.

Takayasu arteritis was first described in Japan in the 1800s. Cases occur primarily in women, with a female-to-male ratio as high as 9:1. The incidence varies dramatically worldwide. In Asia, new cases occur up to 100 times more frequently than in the Western world. The usual age at presentation is from 10 to 40 years.

Histologically, Takayasu arteritis may be divided into an acute inflammatory phase and a chronic fibrotic phase. In the acute phase, changes occur in all three layers of the artery: smooth muscle cell proliferation and mucopolysaccharide deposition in the intima; infiltration of media by lymphocytes and giant cells; and vasa vasoritis in the adventitia, characterized by infiltration of T cells, B cells, and dendritic cells. The chronic phase is marked by fibrosis and destruction of the vessel architecture.

The pathogenesis of Takayasu arteritis remains unknown, but it likely involves a combination of host and environmental factors. For example, Takayasu arteritis has been associated with certain human leukocyte antigens (HLA), including HLA B52, DRBI*1502, DRB5*0102, and DQAI*0103. Furthermore, chronic infections, such as tuberculosis, may trigger an immune response that eventually leads to Takayasu arteritis.

Diagnosis

Clinical suspicion of Takayasu arteritis is typically raised when a female presents with uncontrolled hypertension. Diminished or absent pulses are found in 84% to 96% of patients, along with limb claudication and blood pressure discrepancies.[17,18] Vascular bruits, often multiple, are present in 80% to 94% of patients. Aortic regurgitation, a result of dilatation of the ascending aorta, may be found in up to 25% of patients.

Histologically, Takayasu arteritis may be indistinguishable from giant cell arteritis, another systemic vasculitis of large arteries. Takayasu arteritis affects the aorta and its branches, however, and giant cell arteritis primarily affects the external carotid and its branches. Furthermore, renal artery involvement is extremely rare in giant cell arteritis. Other demographic factors also help to distinguish the two disorders. Onset of giant cell arteritis typically occurs after 50 years of age, whereas onset of Takayasu arteritis is typically before age 40. Takayasu arteritis is more common in Asians, whereas giant cell arteritis is more common in persons of European descent. The female-to-male ratio, which is as high as 9:1 for Takayasu arteritis, is 3:2 for giant cell arteritis.

In 1990, the American College of Rheumatology developed specific diagnostic criteria for Takayasu arteritis [*see Table 1*].[19] To make the diagnosis, at least three of the six criteria must be met. The diagnostic strategy for establishing renal involvement is similar to the workup for renal artery stenosis [*see Figure 2*].

Table 1 American College of Rheumatology Classification Criteria for Takayasu Arteritis*

Age of onset ≤ 40 yr
Claudication of extremities†
Decreased brachial pulse
Blood pressure difference > 10 mm Hg‡
Bruit over subclavian arteries or aorta
Arteriogram abnormalities§

*At least three of the six criteria must be met to make the diagnosis.

†Fatigue and discomfort in muscles of one or more extremities, especially the upper extremities.

‡Typically, systolic blood pressure, between arms.

§Narrowing or occlusion of aorta, primary branches, or large arteries in the proximal upper or lower extremity not caused by other conditions (e.g., atherosclerosis, fibromuscular dysplasia).

Treatment

Treatment of Takayasu arteritis consists of antihypertensive therapy, immunosuppressive therapy, and surgical bypass of severely stenotic arterial lesions. Most antihypertensive agents are suitable for use in Takayasu arteritis, although the use of ACE inhibitors or ARBs should be closely monitored because of the potential for abrupt changes in the GFR. Steroids are the first-line immunosuppressive therapy, with approximately 50% of patients achieving remission.[20] Prednisone at 45 to 60 mg/day orally or another steroid at an equivalent dose is administered until symptoms improve. The dose is then tapered; if symptoms recur, the dose is restored to the starting dose. Given the waxing and waning nature of the disease, the optimal duration of steroid therapy is variable. The second-line therapy, methotrexate, is generally used because of its low side-effect profile compared with other cytotoxic agents, such as cyclophosphamide. Approximately 50% of patients respond to oral methotrexate at a dosage of 5 to 20 mg a week, along with corticosteroids.[21] Azathioprine has been used, with a reported effect similar to that of cyclophosphamide.[22] Agents such as mycophenolate mofetil have been studied, but the data are too preliminary to support formal recommendations.

Surgical correction of arterial stenosis is indicated for patients with critical renal artery stenosis and hypertension, extremity claudication, cerebrovascular disease, splanchnic ischemia, aortic regurgitation, or coronary disease. Surgical intervention should take place during a quiescent phase of disease activity, usually after immunosuppressive therapy.

Prognosis

The prognosis for patients with Takayasu arteritis varies with the extent of vessel involvement. Five-year survival for all patients with Takayasu arteritis is between 80% and 90%,[18,23] but it drops to 60% for those patients with multiple arterial involvement.

RENAL VEIN THROMBOSIS

Renal vein thrombosis (RVT) is a complication of the nephrotic syndrome. Loss of glomerular basement membrane selectivity leads to the excretion of numerous serum proteins, including anticoagulant proteins such as antithrombin III, protein C, and protein S. Furthermore, the blood in the renal veins may be relatively hemoconcentrated after ultrafiltration of the plasma.

The exact prevalence of RVT is not known. Estimates range widely; the disorder may develop in anywhere from 5% to 60% of patients with nephrotic syndrome. RVT has been most frequently associated with idiopathic and secondary membranous nephropathy; 30% of these patients may have RVT. Pulmonary embolism may develop in up to 30% of patients with RVT, although alarmingly, the vast majority of these cases are asymptomatic.

The classic clinical presentation of RVT is acute lower back pain and gross hematuria. Patients typically do not have renal insufficiency or hypertension. RVT can be diagnosed by computed tomography, magnetic resonance imaging, and contrast venography. Doppler ultrasound imaging is notoriously operator dependent and therefore should not be used for diagnosis of RVT.

Anticoagulation with warfarin is indicated for patients with RVT. The therapeutic goal is an international normalized ratio (INR) of 2 to 3. The appropriate duration of therapy is likely lifelong. Thrombolytic therapy or surgical thrombectomy is sometimes considered when bilateral RVT occurs in association with acute renal failure. The value of prophylactic anticoagulation in patients with nephrotic syndrome, especially membranous nephropathy, is open to debate. The prognosis for patients with RVT is generally good.

Disease of Medium-Sized Renal Vessels

Diseases affecting medium-sized arteries and veins of the kidney are relatively rare. Two types of systemic vasculitis that infrequently affect the kidney are polyarteritis nodosa and Kawasaki disease.

POLYARTERITIS NODOSA

Polyarteritis nodosa (PAN) was first described by Kussmaul and Maier in 1866 as periarteritis nodosa. It is a systemic arterial vasculitis that can involve the kidney. For many years, conditions now known to be associated with antineutrophil cytoplasmic autoantibodies (ANCA) may have been incorrectly diagnosed or classified as PAN. The Chapel Hill Consensus Conference provided strict criteria for the diagnosis of ANCA-associated diseases and PAN.[24]

Polyarteritis nodosa is now defined by three criteria: (1) necrotizing vasculitis affecting small arteries; (2) sparing of the smaller vessels (i.e., arterioles, capillaries, and venules); and (3) no association with a primary or secondary glomerulopathy. The condition is rare; idiopathic PAN affects two to nine persons per million population annually,[25,26] although annual incidence rates as high as 77 cases per million have been reported in areas endemic for hepatitis B virus infection.[27] No racial, ethnic, or gender predilection for idiopathic PAN has been observed.

The pathogenesis of PAN is unclear. There appears to be an association with hepatitis B viral infection. In the past, most reported cases of idiopathic PAN may actually have been long-term complications of hepatitis B resulting from immune complex deposition. Childhood cases of PAN may be associated with streptococcal infection.

Histologically, PAN appears as a pleomorphic cellular infiltration of the adventitia. Polymorphonuclear cell degranulation leads to leukocytoclasis and, eventually, to transmural necrosis. Segmental necrosis of the artery may occur, which then leads to the formation of aneurysms. In later stages of the inflammatory process, endothelial damage and thrombosis result in complete obstruction of an affected vessel.

Clinically, PAN develops in an insidious manner, with constitutional symptoms appearing over many weeks to even months. Early nonspecific signs and symptoms include fevers, weight loss, malaise, and anorexia. Large-joint arthralgias without true arthritis may occur in the knees, ankles, elbows, and wrists. Findings that are more suggestive of PAN include livedo reticularis (a purplish rash over the lower extremities and abdominal wall), dermal nodules or ulcerations, and digital ischemia—all of which are sequelae of arteritis and disordered blood flow through medium-sized arteries. PAN may also affect a host of organ systems, including the gut, heart, and kidney. Renal involvement typically leads to hypertension. If extensive renal involvement is present, renal insufficiency may also develop.

The diagnosis of PAN is made by demonstration of the characteristic lesion in an artery. Biopsy of a cutaneous nodule is preferred because of its relative accessibility, but such nodules are notoriously uncommon. Arteriography may represent a suitable alternative. Serologic tests are not diagnostic in PAN, but low-titer antibodies to rheumatoid factor (RF) and nuclear antigen

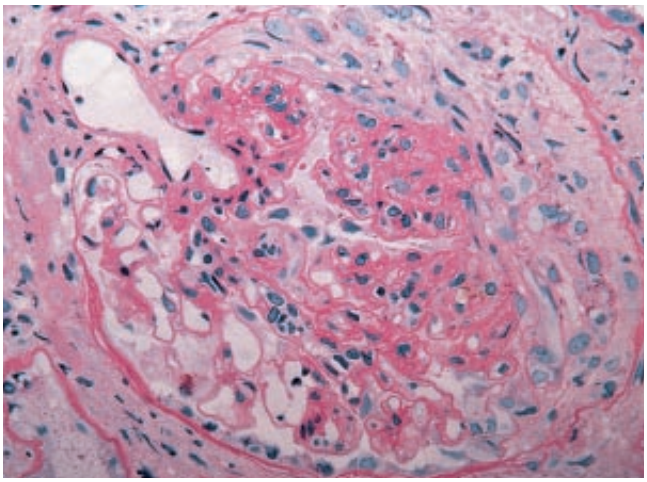


Figure 3 **Light microscopy of a renal biopsy specimen from a patient with crescentic glomerulonephritis shows compression of the glomerular tuft by a large crescent. Although an identical renal lesion can be observed in a variety of renal disorders, the finding of negative immunofluorescent staining for antigen-antibody complexes is highly suggestive of a systemic necrotizing vasculitis, such as Wegener granulomatosis or polyarteritis nodosa. Most other causes of crescentic glomerulonephritis are typified by characteristic patterns of immune staining, such as the linear immunofluorescent pattern seen in Goodpasture syndrome.**

(ANA) may be present. PAN secondary to hepatitis B has been associated with hypocomplementemia. Finally, immunofluorescence antibody staining for cytoplasmic and perinuclear ANCA may be positive, but the more specific test, serum enzyme-linked immunosorbent assay (ELISA) titers for antibodies against both serine protease 3 (PR3) and myeloperoxidase (MPO), is negative.

If left untreated, patients with PAN have a poor prognosis. In such cases, patients are at risk for ischemia of numerous organ systems; the major causes of morbidity and mortality include renal failure, mesenteric ischemia, and cerebrovascular accidents. Corticosteroids and cytotoxic agents have been the mainstays of therapy for idiopathic PAN, although the optimal therapy remains unknown.

Approximately 50% of patients with idiopathic PAN achieve remission with high-dose steroids (e.g., prednisone, 1 mg/kg/day) for 3 to 6 months.[28] Cyclophosphamide, either intravenous (0.6 g/m^2/mo) or oral (2 mg/kg/day), for up to 1 year is used in patients whose disease does not respond to steroids or in patients who are at risk for serious complications. A method for determining the need for cytotoxic therapy is the five-factor score, which comprises the following: (1) gastrointestinal signs, including bleeding, perforation, infarction, or pancreatitis; (2) renal insufficiency; (3) proteinuria; (4) CNS involvement; and (5) cardiac involvement.[29] The five factor score is slightly dated; according to the Chapel Hill nomenclature, proteinuria would no longer be a valid criterion for PAN. ACE inhibitors and ARBs should be used cautiously in patients with PAN, because renal involvement may produce a functional equivalent of classic RAS.

KAWASAKI DISEASE

Kawasaki disease (KD) is another systemic vasculitis affecting medium-sized arteries. Renal involvement in KD has been varied. Vasculitic lesions of the renal vessels are rare, for unclear reasons. Sterile pyuria, interstitial nephritis, and proteinuria may be associated with KD in the kidney. The diagnosis and treatment of KD are discussed in detail elsewhere [*see 15:VIII Systemic Vasculitis Syndromes*].

Small Vessel Disease

CONDITIONS ASSOCIATED WITH GLOMERULOPATHY

Many small vessel diseases of the kidney are associated with glomerulopathy. Although these diseases are often considered to be forms of glomerulonephritis, the sentinel lesion occurs at the level of the glomerular capillary, where in situ and exogenous antigens incite an aggressive inflammatory response. This can lead to a significant breakdown of glomerular architecture, including rupture of the Bowman capsule (crescent formation), as well as leakage of protein and red blood cells into the tubular space. Several of these vascular diseases have been described as rapidly progressive glomerulonephritis (RPGN) because of the loss of renal function within days to weeks [*see Figure 3*].

These small vessel diseases share a similar clinical presentation. Most patients have evidence of systemic illness. They have a history of malaise, anorexia, weight loss, fevers, rash, arthralgias, and discolored urine. With pulmonary-renal syndromes, patients may also have hemoptysis, epistaxis, or chronic sinusitis. Key features of the physical examination for all these diseases include assessment of blood pressure, joints, skin, nasopharynx, lymph nodes, abdominal organs, and eyes (retina). Urinalysis shows hematuria or proteinuria, and microscopic evaluation of the urine reveals red blood cell casts, oval fat bodies, and dysmorphic red blood cells.

Rapidly Progressive Glomerulonephritis

ANCA-associated vasculitis Vasculitis associated with ANCA is the most common small-vessel vasculitis in adults. It includes three main categories: Wegener granulomatosis, microscopic polyangiitis (MPA), and Churg-Strauss disease. All three conditions are associated with a pauci-immune glomerulonephritis, although traditionally this is seen at a lower frequency in Churg-Strauss disease than in the other two ANCA diseases. Wegener granulomatosis manifests primarily in the kidney, upper respiratory tract, and lung, whereas MPA occurs primarily in the kidney. Churg-Strauss disease is found typically in the lungs and is associated with asthma and eosinophilia [*see 15:VIII Systemic Vasculitis Syndromes*].

ANCA-associated vasculitis may affect persons of all ages, but it is most common in middle-aged and older adults (> 50 years). It is equally distributed among men and women, and in the United States, it is most common in whites. Its documented incidence is between one and two cases per 100,000 population.[30,31]

Information on the pathogenesis of ANCA-associated vasculitis is growing. There are likely environmental and infectious stimuli. For example, viral and bacterial respiratory tract infections often precede the first occurrence of disease. Furthermore, therapy to suppress staphylococcal species decreases the frequency of recurrences.[32] The pathologic effect of the autoantibodies may result from activation of neutrophils.

The histologic appearance of ANCA-associated vasculitis is leukocytoclastic infiltration of arterioles, capillaries, and venules throughout the body. Fibrinoid necrosis of the muscularis layer, often extending into the perivascular tissue, may also be present.

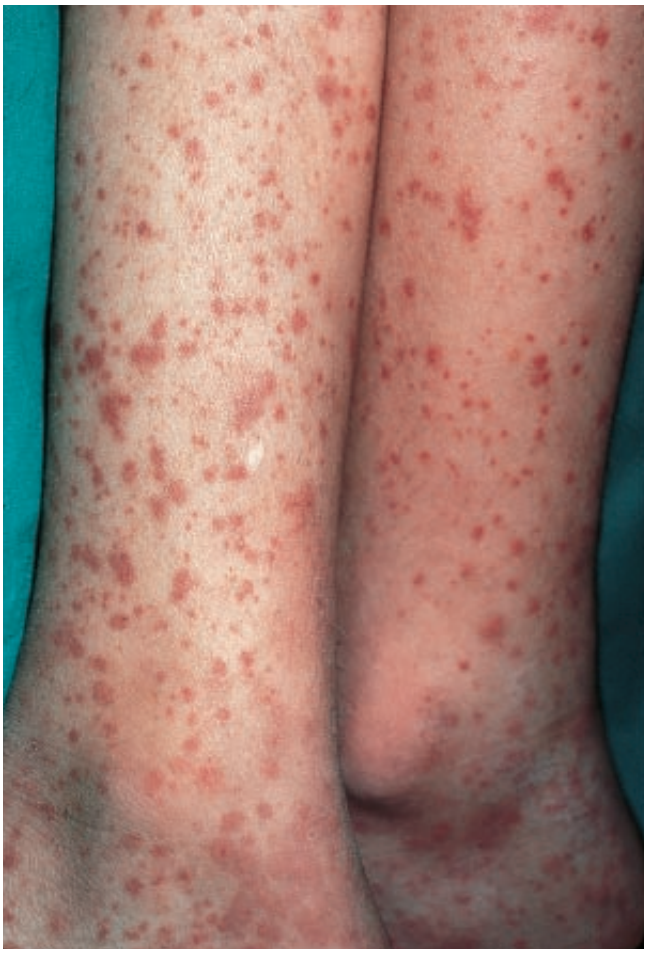


Figure 4 **Purpuric eruptions can be seen in a variety of small vessel vasculitides, Henoch-Schönlein purpura, hypersensitivity vasculitis, and ANCA-associated disease. The lesions are palpable, and immunofluorescent staining of the dermal vessels on skin biopsy may be diagnostic. For example, IgA deposition is highly suggestive of Henoch-Schönlein purpura in this setting.**

The histologic lesion, fibrinoid necrosis, is indistinguishable from that seen in other vasculitides; the type and location of the involved vessel help identify the specific disease process.

The clinical presentation of small-vessel vasculitis often starts with constitutional signs and symptoms such as fever, malaise, myalgias, and arthralgias. Blood vessels throughout the body—including skeletal muscle, peripheral nerves, gastrointestinal tract, kidneys, respiratory tract, and skin—may be affected; thus, signs and symptoms are variable. Leukocytoclastic angiitis of dermal vessels leads to the most common skin finding—purpura [*see Figure 4*]. These purpuric lesions are often raised, and for unclear reasons, they tend to form on the lower extremities, often distal to the knees. Angiitis of vessels supplying peripheral nerves may lead to mononeuritis multiplex, the most common neurologic manifestation.[33,34] Renal involvement typically leads to glomerulopathy, with subsequent hematuria and proteinuria. Respiratory tract involvement leads to epistaxis and hemoptysis. Asthma is specific to Churg-Strauss disease.

The diagnosis of ANCA-associated vasculitis is aided by serum ANCA testing. The former ANCA nomenclature has been replaced by more specific terminology, with cytoplasmic ANCA (c-ANCA) succeeded by PR3-ANCA (serine protease 3) and perinuclear ANCA (p-ANCA) replaced by MPO-ANCA (myeloperoxidase). Approximately 80% of patients with Wegener granulomatosis have high-titer antibodies to PR3-ANCA, whereas 80% of those with microscopic polyangiitis have high-titer antibodies to MPO-ANCA. Up to 10% of patients with a typical clinical presentation for ANCA-associated vasculitis may have negative or low-titer serum antibodies to MPO or PR3.[35]

The prognosis for patients with untreated ANCA-associated vasculitis is poor. Death results from renal failure or pulmonary hemorrhage. Consequently, intervention early in the course of disease is important, preferably before the serum creatinine concentration reaches 4 mg/dl.[36] Corticosteroids and cyclophosphamide are the mainstays of induction therapy [*see Table 2*]. This regimen produces complete remission in about 75% of patients and improvement in up to 90%.[31] Plasma exchange should be initiated promptly with cytotoxic therapy in the setting of concomitant pulmonary hemorrhage. Plasma exchange is performed every second day, with each exchange consisting of 1 to 1.5 plasma volumes, for a total of three to six procedures.

The optimal maintenance therapy has yet to be determined. Both cyclophosphamide (oral and intravenous) and oral mycophenolate mofetil have been used. Attention has recently focused on an old agent, azathioprine; oral azathioprine has proved to be equivalent to oral cyclophosphamide with respect to relapses and severe adverse effects.[37] Relapses of ANCA-associated vasculitis are generally treated as de novo disease, with high-dose corticosteroids and cyclophosphamide.

Anti–glomerular basement membrane disease Nearly 20% of RPGN cases are caused by anti–glomerular basement membrane (anti-GBM) disease. This condition is defined by the presence of circulating IgG antibodies to type IV collagen—specifically, to the NC1 domain of the α3 chain.

Anti-GBM disease has been reported to occur at the extremes of age: the youngest patient was 4 years old, and the oldest was 81 years old. Classically, the disease occurs in two distinct groups[38]: men younger than 30 years and women older than 60 years. Young men with anti-GBM disease tend to have both kidney and lung involvement. This condition is termed Goodpasture syndrome, in honor of Ernest Goodpasture, who is credited with the first description of anti-GBM disease, in a young man

Table 2 Induction Therapy for Rapidly Progressive Glomerulonephritis*

Pulse corticosteroids:
Methylprednisolone, 7 mg/kg/day I.V. (maximum dose, 500 mg) × 3 days

then

Oral corticosteroids:
Prednisone, 1 mg/kg/day p.o. × 3–6 mo

plus

Cyclophosphamide:
Oral: 2–3 mg/kg/day × 6 mo; round dose down to nearest 50 mg; reduce to 2 mg/kg/day in patients > 55 yr

Intravenous: 0.5 g/m^2 monthly for six doses

plus

Exchange transfusion, if indicated

*See text for details on individual diseases.

with acute renal failure and pulmonary hemorrhage; interestingly, the presentation was ascribed to the influenza pandemic of the early 1900s. Older women typically present with disease limited to the kidneys. In both men and women, there appears to be a strong racial predilection for whites.

The agent or agents that incite the development of autoantibodies in anti-GBM disease remain unknown. Isolated reports have suggested that the disease develops after pulmonary infections or exposure to aerosolized hydrocarbons, yet large epidemiologic studies have yet to substantiate these observations. Type IV collagen is found predominantly in the GBM and in the alveolar basement membrane in the lungs. Interestingly, there are persons with a hereditary lack of type IV collagen (Alport syndrome) who may experience anti-GBM disease if they receive a kidney transplant from an unrelated donor. Alveolar type IV collagen is relatively unexposed to the serum, which may help to explain the lower frequency of lung involvement. Injury to the barrier in the lung, by such mechanisms as cigarette smoking, may unmask the type IV collagen.

On light microscopy, the histologic appearance of anti-GBM disease is similar to that of the other forms of RPGN. The lesions demonstrate aggressive changes: diffuse cellular and mesangial expansion, karyorrhexis, and crescent formation (breakdown of Bowman capsule). Anti-GBM disease is distinguished from these other forms by immunohistologic staining that reveals linear IgG along the basement membrane.

High serum antibody titers to GBM help confirm the diagnosis. Over 30% of patients with anti-GBM disease may also demonstrate ANCA antibodies.[39] The presence of ANCA antibodies predicts a better response to immunosuppressive therapy.

The prognosis for patients with untreated anti-GBM disease is dismal; nearly all patients die of either renal failure or pulmonary hemorrhage. With therapy, however, the prognosis is good. One-year survival is about 75% to 90%.[38] As in ANCA disease, early intervention provides a substantial survival benefit. Corticosteroids, cyclophosphamide, and plasma exchange are the cornerstones of therapy [*see Table 2*]. These three modalities act to eliminate the pathogenic antibody from circulation, thereby limiting additional glomerular and alveolar damage. Plasma exchange consists of a 4-liter exchange for 5% human albumin solution, performed daily for 14 days. Because these exchanges may deplete intrinsic clotting factors, fresh frozen plasma should be given within 3 days of any planned invasive procedure (e.g., biopsy) or in the event of pulmonary hemorrhage. Relapses are exceedingly rare, and immunosuppressive therapy can generally be discontinued after the initial course of treatment.

Immune complex glomerulonephritis The immune complex glomerulonephritides are a diverse group of conditions accounting for 30% to 40% of RPGN. They often develop in patients with a preexisting glomerular disease; histopathologic study shows an aggressive variation of the underlying glomerulonephritis. IgA nephropathy, lupus nephritis, membranoproliferative glomerulonephritis, and postinfectious glomerulonephritis all may present as an immune complex RPGN. Diagnostic testing depends on the underlying glomerular disease.

Lupus nephritis typically presents as renal insufficiency with a nephritic sediment in urine. This disorder affects significantly more women than men (ratio of 9 to 1), and generally occurs between 20 and 40 years of age. Histologically, lupus nephritis is characterized according to the extent of mesangial and cellular proliferation in the glomerulus and is categorized as World Health Organization (WHO) classes I through V.[40] Class III (focal proliferative glomerulonephritis) and class IV (diffuse proliferative glomerulonephritis) have particular clinical significance. In addition, patients with lupus and lupus anticoagulant antibodies may have renal disease with features of a thrombotic microangiopathy (TMA) (see below).

As is true of systemic lupus erythematosus (SLE), the cause of lupus nephritis is not yet known. The diagnosis of lupus nephritis is aided by the fact that it usually occurs along with other features of SLE, such as rash or arthritis. Serum complement levels (both C3 and C4) are typically depressed. Other helpful serologic tests include measurement of ANA and of antibodies against double-stranded DNA (dsDNA). ANA testing has high specificity, with titers greater than 1:160 strongly suggesting disease. Furthermore, a negative ANA confirms the absence of disease. A positive dsDNA antibody assay helps confirm the presence of SLE [*see 15:IV Systemic Lupus Erythematosus*]. Lupus nephritis is confirmed with a kidney biopsy.

Without therapeutic intervention, the prognosis is poor for patients with class III or IV lupus nephritis or TMA. Most patients progress to ESRD or die. As with other types of RPGN, therapy consists of corticosteroids, cytotoxic agents, and, occasionally, plasma exchange [*see Table 2*]. Plasma exchange is indicated with concomitant pulmonary hemorrhage or TMA; the regimen is one exchange every second day, with each exchange consisting of 1 to 1.5 plasma volumes, for a total of three to six procedures. Data have also been mounting regarding the efficacy of oral mycophenolate mofetil (500 to 2,000 mg daily for up to 2 years) as either primary therapy[41] or maintenance therapy[42] for class III or IV lupus nephritis.

Henoch-Schönlein Purpura

Henoch-Schönlein purpura (HSP) is a systemic vasculitis characterized by IgA-dominant immune complex deposition in the venules, capillaries, and arterioles. Renal involvement in HSP is indistinguishable from idiopathic IgA nephropathy. The involvement of other organ systems, such as the skin, nerves, or gastrointestinal tract, helps distinguish HPA from IgA nephropathy.

HSP is a disease of childhood, with peak occurrence at 5 years of age. Renal involvement tends to occur in older children and in adults. HSP in adults is generally more severe than that in children. There is a stronger predilection for HSP in Europeans and Asians than in other ethnic groups, and there may be a slightly higher frequency of the disease in men than in women.

The etiologic agent of HSP is not known. The disease often begins after a respiratory infection. Exposure to viruses such as adenovirus may trigger an immune complex response leading to the deposition of IgA subtype A1 in arterioles, venules, and capillaries throughout the body.

In the kidney, IgA deposition is visualized by immunofluorescence staining. It is often accompanied by mesangial matrix expansion and hypercellularity (neither finding is specific to HSP). In the skin, the characteristic lesion is a leukocytoclastic vasculitis.

The clinical presentation of HSP is variable. The classic tetrad of signs and symptoms comprises rash, arthralgias, abdominal pain, and renal disease. The rash is purpuric. The arthralgias are symmetrical and affect small joints. The abdominal pain is diffuse and waxes and wanes. The renal disease, typically with hematuria and proteinuria, is found in over 50% of patients with HSP. Renal insufficiency affects only 10% to 20% of patients. In rare cases, pulmonary disease and peripheral neuropathy may occur.

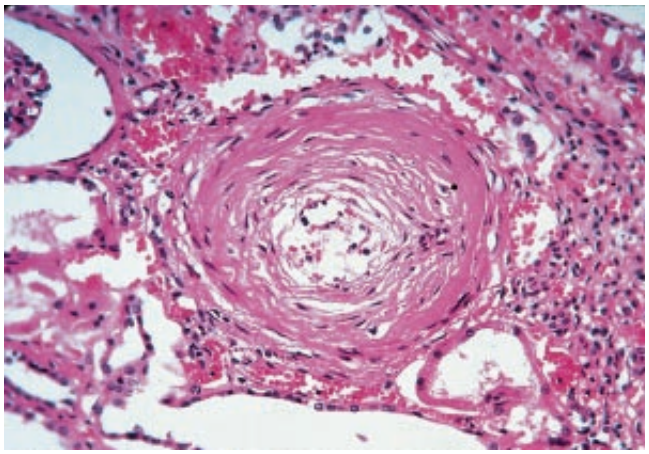


Figure 5 **Light microscopy reveals the characteristic onion-skin lesion seen in scleroderma renal crisis. This finding is typical of disorders that present as a thrombotic microangiopathy, such as malignant hypertension, thrombotic thrombocytopenic purpura, hemolytic-uremic syndrome, and catastrophic antiphospholipid syndrome.**

The prognosis for most patients with HSP is good. Almost 95% of children and 90% of adults have complete recovery. Supportive therapy, therefore, is sufficient for most patients. Progressive renal insufficiency leading to ESRD occurs in about 5% of patients. The optimal therapy in such cases is not known. Immunosuppressive agents (e.g., corticosteroids, cyclophosphamide, and azathioprine) and anticoagulants have been used in HSP, with inconsistent results. Thus, therapy is usually reserved for aggressive forms of HSP marked by renal insufficiency and diffuse gastrointestinal involvement; it includes a combination of corticosteroids, cyclophosphamide, and plasma exchange.

THROMBOTIC MICROANGIOPATHY

TMA comprises a group of occlusive vascular diseases characterized by endothelial disruption, systemic or intrarenal platelet aggregation, thrombocytopenia, and mechanical injury to erythrocytes. The platelet thrombi contain large amounts of von Willebrand factor (vWF) but no fibrin or fibrinogen. Furthermore, intimal thickening of the small arteries and arterioles occurs diffusely and has an appearance that has been described as an onion-skin lesion.

The classic TMAs include thrombotic thrombocytopenic purpura (TTP) and hemolytic-uremic syndrome (HUS). Antiphospholipid syndrome (APS), preeclampsia, systemic sclerosis with renal crisis, malignant hypertension, and certain drugs can produce a clinical picture that is phenotypically similar to classic TMA [*see Figures 5 and 6*]. All of these conditions often present as acute renal failure and can cause acute cortical necrosis, which is generally considered irreversible. Furthermore, hemolytic anemia and thrombocytopenia are present. The renal lesions are remarkably similar in all of these diseases, so the history is often crucial in determining the exact type of TMA.

Thrombotic Thrombocytopenic Purpura

Eli Moschowitz first described TTP in 1924, in a 16-year-old girl with abrupt onset of fever and petechiae followed by paralysis and death. Autopsy revealed diffuse hyaline thrombi affecting the terminal arterioles and capillaries in numerous organs. Subsequently, TTP came to be defined by a pentad of clinical findings: fever, thrombocytopenia, microangiopathic hemolytic anemia, neurologic abnormalities, and renal dysfunction.

TTP can occur in any age group. There are no apparent differences in occurrence of the disease by race or gender.

The critical role of a metalloprotease (ADAMTS-13 [a disintegrin and metalloprotease with thrombospondin type 1 motif, 13 isoform]) in the pathogenesis of TTP has emerged in the past 10 years. ADAMTS-13, which is found on the surface of endothelial cells, normally cleaves large multimers of the von Willebrand antigen as they are secreted by the cell. These large multimers bind more efficiently than the cleaved von Willebrand antigen to platelets (at the Ibα component of platelet glycoprotein Ib/IX/V vWA receptor). Plasma ADAMTS-13 activity must be severely depressed for TTP to develop. The level of ADAMTS-13 plasma activity is modestly depressed in patients with liver disease, disseminated cancer, chronic metabolic and inflammatory conditions, and pregnancy, as well as in newborns.[43] When plasma activity of ADAMTS-13 falls to less than 5% of normal, these large von Willebrand antigens predominate, bind to platelets, and cause aggregation and thrombi in small vessels.

Several mechanisms have been postulated for the abnormal functioning of ADAMTS-13 and noncleavage of von Willebrand antigens. Familial or chronic relapsing TTP may result from a mutation in the gene that codes for ADAMTS-13; this would explain why infusion of fresh frozen plasma, which is known to contain ADAMTS-13, is the only effective therapy in these cases. Autoantibodies against ADAMTS-13 may form with exposure to certain drugs (e.g., ticlopidine) or in idiopathic single-episode TTP. Removal of the autoantibody against ADAMTS-13 would explain the efficacy of plasmapheresis in the treatment of TTP in these cases.

Another newly postulated mechanism for the development of the familial form of TTP is a deficiency in plasma factor H. Factor

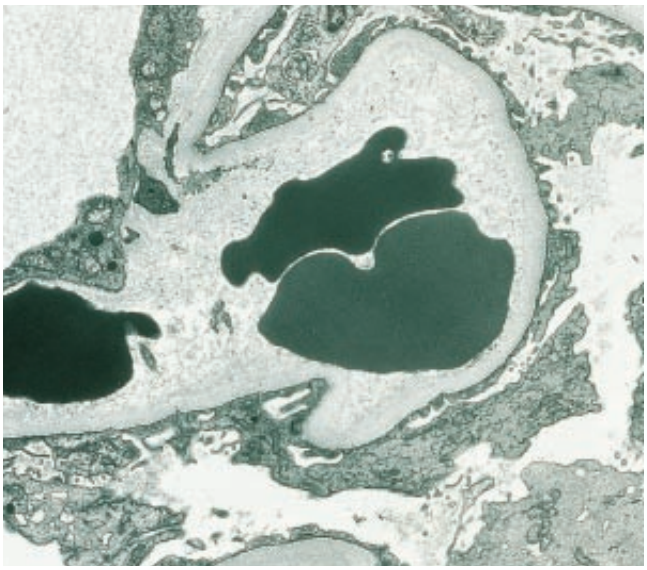


Figure 6 **Electron microscopy shows a glomerular capillary in a patient with thrombotic thrombocytopenic purpura. The capillary lumen is in the center and contains three red cells. The fluffy-appearing material between the red cells and the glomerular basement membrane is characteristic of this disorder and may represent incompletely polymerized fibrin; it has separated the endothelial cells from the basement membrane. In more chronic cases, a vascular lesion similar to that observed in scleroderma [*see Figure* 5] may be seen.**

H is a 150 kd plasma protein that inhibits the activation of the alternative pathway of complement.[44-46] Specifically, factor H exposes C3b to cleavage by factor I. A deficiency in factor H therefore leads to unregulated C3, potentiating immune complex–mediated or autoantibody-mediated glomerular endothelial injury. Exposure of the glomerular subendothelium leads to the adhesion and aggregation of platelets. Thus far, most of the defects in factor H have been determined to be frameshifts, deletions, and point mutations.

A mutation in the membrane cofactor protein (MCP) has been suggested to cause a form of TTP that is phenotypically similar to factor H deficiency.[47] The MCP, a widely expressed transmembrane glycoprotein, serves as a necessary cofactor to factor I. An abnormal MCP would prevent factor I–driven inactivation of C3b, resulting in overactivation of the alternative complement system.

Although TTP is often defined by the classic pentad of findings (see above), a presumptive diagnosis can often be based on a triad of laboratory observations: thrombocytopenia, schistocytes, and elevated levels of serum lactate dehydrogenase (from shredded erythrocytes). The differential diagnosis includes other TMAs. A drug history that reveals exposure to agents such as cyclosporine, tacrolimus, mitomycin C, quinine, ticlopidine, and clopidogrel suggests drug-induced TTP. Quinine is no longer an over-the-counter medication, but it is still found in tonic water and herbal medications containing bark from the cinchona tree. A history of ingesting undercooked meat or unpasteurized dairy products or juices and, possibly, an antecedent diarrheal illness suggest HUS. In a pregnant woman, the triad of laboratory findings raises the possibility of preeclampsia and its variant, HELLP syndrome (hemolysis, elevated liver enzymes, and low platelets). The renal crisis phase of systemic sclerosis (scleroderma) may also be difficult to distinguish from TTP; a history of scleroderma, compatible physical findings, and the presence of antibodies to Scl-70 help confirm scleroderma renal crisis. Finally, a history of poorly controlled hypertension and severe elevations of blood pressure (generally, a diastolic pressure greater than 130 mm Hg) and papilledema suggest malignant hypertension.

Conditions other than TMA may also be difficult to distinguish from TTP. Disseminated intravascular coagulation (DIC), thought to be triggered by sepsis, shock, or obstetric complications, leads to intravascular activation of the coagulation cascade and consumption of coagulation factors. The formation of intravascular fibrin thrombi leads to a microangiopathic hemolytic anemia. It may be possible to presumptively distinguish DIC from TTP by measuring the prothrombin time (PT) and partial thromboplastin time (PTT); both the PT and the PTT should be elevated in DIC.

The prognosis is poor for patients with untreated TTP. Therapy is targeted at the postulated underlying mechanism. Because familial and chronic relapsing TTP are thought to result from the production of functionally deficient ADAMTS-13, these conditions are treated with infusion of platelet-poor fresh frozen plasma or cryoprecipitate-poor plasma, rather than being treated with plasmapheresis. Plasma contains active metalloproteases, including ADAMTS-13, and a plasma transfusion will increase plasma activity for up to 3 weeks. Gene therapy may soon be possible, because the sequence of the gene that codes for ADAMTS-13 has been determined.

Acquired TTP is treated with plasma exchange. Plasma exchange has two components: (1) plasmapheresis to remove the von Willebrand multimers and autoantibodies to ADAMTS-13 and (2) infusion of fresh frozen plasma to help reconstitute plasma volume and proteins, including coagulation factors. Plasma exchange is typically administered every other day, for a total of six sessions. Patients treated with plasma exchange have a 90% chance of surviving an acute episode of TTP.[48] Antiplatelet therapy with aspirin is no longer routinely recommended for TTP, because it can cause severe bleeding in patients with severe thrombocytopenia.[49]

Hemolytic-Uremic Syndrome

HUS is defined as a TMA that occurs after exposure to *Escherichia coli* 0157:H7 or other enteropathogenic serotypes of *E. coli*, *Shigella dysenteriae*, and, occasionally, some other enteropathogenic organism. It is predominantly a disease of children, especially those younger than 5 years. HUS occurs in 9% to 30% of children who experience bloody diarrhea from *E. coli* 0157:H7; the HUS follows the diarrhea approximately 1 week later[50-52] [*see 7:VIII Infections Due to Escherichia coli and Other Enteric Gram-Negative Bacilli*]. Cattle are reservoirs of *E. coli* 0157:H7. The organism is harmless to cattle because it is unable to penetrate their intestinal mucosa. The toxins produced by the bacteria can be transmitted via unpasteurized milk products, as well as by undercooked meat.

In humans, *E. coli* 0157:H7 is able to adhere to the colonic mucosal cells, invade, and replicate. The ensuing inflammatory response leads to the disruption of the normal mucosal architecture, with bloody diarrhea and systemic bacterial invasion. Systemically, Shiga toxins play two key roles in the development of thrombi. First, they impair the secretion of ADAMTS-13 through an unspecified mechanism, resulting in large von Willebrand multimers. Second, they activate platelet adherence via the glycoprotein Ibα component of glycoprotein Ib/Ix/V1 and V2, which may then promote von Willebrand antigen and platelet aggregation.

The prognosis for patients with HUS is variable. Children with oliguria for less than 24 hours generally have full recovery of renal function. Adults generally have a worse prognosis for recovery of renal function. Supportive therapy with dialysis may be needed.

There is at present no effective therapy for HUS. Plasma infusion and exchange have demonstrated equivocal results,[53,54] but given the overlap of HUS with TTP, both modalities are often tried. The optimal treatment is primary prevention. Meat products, especially ground beef and hot dogs, should be thoroughly cooked. Unpasteurized dairy products should be avoided.

Preeclampsia

Preeclampsia is a multisystem vascular disease of gravid women in their second and third trimester. It affects 3% to 5% of all pregnancies.[55] Preeclampsia remains a major cause of maternal-fetal morbidity and mortality. Renal involvement in preeclampsia is common and can be used to support the diagnosis.

Studies have long suggested that risk factors for the development of preeclampsia are related to decreased placental perfusion. For example, primigravidas, who tend to have a relatively small placenta, have twice the risk of developing preeclampsia than women in their second or third pregnancies.[56] Other risk factors linked to decreased placental perfusion include diabetes mellitus, preexisting hypertension, thrombophilia, and the presence of a hydatidiform mole or multiple gestations.

Recent hypotheses regarding the pathophysiology of preeclampsia have focused on immune tolerance and circulating

angiogenic factors. The placenta is a conglomerate of two genotypically different populations of cells. Preeclampsia may result from a maternal intolerance to fetal cells. Repeated low-level exposure to these antigens may promote tolerance; this would explain why studies have shown a protective effect of sperm exposure and multiple pregnancies with the same partner but an increased risk of preeclampsia with changes in partners or donor insemination.[57] Circulating soluble fms-like tyrosine kinase–1 (sFlt-1) may also have an important role in the pathogenesis of preeclampsia. This molecule typically binds vascular endothelium growth factor (VEGF), which is necessary for normal angiogenesis, including in the placenta. In normotensive pregnancies, sFlt-1 levels start to increase at 33 to 36 weeks of gestation; in preeclampsia, the levels are postulated to increase earlier in gestation. The resultant decrease in VEGF may then lead to insufficient angiogenesis of the placenta. A case-control study demonstrated that elevated sFlt-1 levels and decreased VEGF levels were associated with the development of preeclampsia.[58]

The classic presentation of preeclampsia is the triad of hypertension, edema, and proteinuria. Hypertension (blood pressure ≥ 140/90 mm Hg) is a relatively sensitive finding but is not highly specific, given that it may occur in 12% to 22% of all pregnancies.[59-61] Leakage of protein into the Bowman space and subsequently into the urine is thought to result from glomerular capillary endotheliosis (proliferation of endothelium). Urine protein excretion exceeds 0.3 g in 24 hours and can often be in the nephrotic range (≥ 3.5 g/24 hr). Edema is thought to result from the loss of oncotic pressure in the vessels and is believed to affect all organs, but it is often apparent only in the upper extremities or face. Liver edema may result in the development of epigastric and right upper quadrant abdominal pain. Capillary obstruction undoubtedly leads to a decrease in renal blood flow and GFR. Reductions in GFR may not be detected until the late stages of preeclampsia because of the normal elevation of GFR during pregnancy and the insensitivity of serum creatinine concentration measurement. An increased serum uric acid concentration, which is indicative of decreased tubular secretory function, may be a better indicator of early renal dysfunction in preeclampsia. In developed countries, acute renal failure is rare because of successful delivery of premature fetuses. Ischemic acute tubular necrosis and cortical necrosis from preeclampsia now account for less than 4% of all cases of acute renal failure associated with pregnancy.

Although preeclampsia might seem easy to diagnose, it has often been mistaken for such varied conditions as acute pancreatitis, gallbladder disease, appendicitis, nephrolithiasis, and glomerulonephritis.[62] Distinguishing between preeclampsia and glomerulonephritis may be difficult. A history of proteinuria and hematuria before the pregnancy strongly suggests an underlying glomerulonephritis. Yet, often the patient's old records are not available. Also, the severity of proteinuria and hematuria can markedly increase during pregnancy. Thus, a renal biopsy may be the only way to distinguish preeclampsia from glomerulonephritis. On biopsy, the histologic change considered specific to preeclampsia, endotheliosis, can help distinguish it from other forms of TMA. A renal biopsy is indicated in suspected preeclampsia before 28 weeks' gestation if the survival of the fetus is desired.[63] In general, the procedure is well tolerated by both mother and fetus.[64]

If left untreated, preeclampsia will progress rapidly to eclampsia, the constellation of signs that result from the development of cerebral edema. These signs include encephalopathy, papilledema, central visual loss, and seizures. Permanent neurologic sequelae, acute cortical necrosis of the kidney, and death may occur. Delivery of the fetus and placenta is considered the treatment of choice for preeclampsia.

Prevention of preeclampsia is a standard aspect of obstetric care. Dipstick assays for proteinuria and measurement of blood pressure are performed at all prenatal visits. The development of hypertension during pregnancy brings closer supervision, with more frequent urine protein screening and measurement of serum uric acid. Antihypertensive therapy is considered if the blood pressure rises above 160/110 mm Hg and is mandated when levels surpass 170/110 mm Hg.[65] The aim of therapy is to keep the mean arterial pressure at 125 to 105 mm Hg and to maintain the diastolic blood pressure at 105 to 90 mm Hg.[62] Currently, hydralazine, nifedipine, α-methyldopa, and labetalol are the preferred agents. ACE inhibitors, which are often used in other kidney diseases, are contraindicated, especially in the second and third trimesters, because of their potential to cause adverse effects in both mother and fetus. Corticosteroids are administered to promote the production of surfactant in the fetus. Magnesium sulfate is used to prevent both seizures and preterm delivery. Compared with placebo, interventions such as oral calcium, aspirin, and salt reduction have not been shown to significantly reduce the incidence of preeclampsia.

Antiphospholipid Syndrome

APS refers to hypercoagulability in the presence of autoantibodies to negatively charged phospholipids. These antibodies have traditionally been believed to cause thrombosis of large arteries and veins and to lead to miscarriages. The full spectrum of disease has now been clarified. Renal involvement is common in APS. Furthermore, when renal involvement occurs, it typically presents as a TMA.

The term antiphospholipid antibodies is actually a misnomer. With the exception of cardiolipin, the target antigens are plasma proteins bound to phospholipids. Antiphospholipid antibodies are currently grouped into four categories,[66,67] although there may be overlap, because the antibodies share antigen affinity. The four subtypes are as follows: (1) anticardiolipin antibodies (ACAs), (2) lupus anticoagulant, (3) anti–β_2-glycoprotein I antibodies, and (4) antiprothrombin antibodies. ACAs were first isolated from patients with syphilis almost 100 years ago. They were later found to be directed against a mitochondrial phospholipid, cardiolipin. Lupus anticoagulant is a group of immunoglobulins originally found in patients with systemic SLE about 50 years ago. Despite its name—which is derived from its in vitro property of slowing thrombin generation—lupus anticoagulant is an in vivo procoagulant. Lupus anticoagulant is directed against plasma proteins (primarily prothrombin, but also annexin and β_2-glycoprotein) bound to negatively charged phospholipids. β_2-glycoprotein I is an important inhibitor of the coagulation cascade. Antibodies to this protein were described nearly a decade ago. Antibodies directed against prothrombin that are different from lupus anticoagulant have been described within the past 5 years.

It is still not clear why antiphospholipid antibodies cause thrombosis. Antiphospholipid antibodies have both procoagulant and anticoagulant effects. Also, it is not known what agents incite the production of these antibodies. Given that β_2-glycoprotein I antibodies may cross-react with bacterial antigens, it may be speculated that exposure to bacterial pathogens and, possibly, viral pathogens induces the formation of these antibodies.

Antiphospholipid antibodies are quite common. They can be found in 1% to 5% of healthy persons.[68] It is not clear whether they cause APS in healthy persons, although epidemiologic studies have shown an association between these antibodies and myocardial infarction (MI), stroke, and venous thrombosis. Between 12% and 30% of patients with SLE have ACA,[69,70] and 15% to 34% have lupus anticoagulant.[69,71] APS may develop in 50% to 70% of patients with lupus and antiphospholipid antibodies.[68,72]

The clinical presentation of APS most often comprises a single thrombotic event in either the arterial or venous system. Deep vein thrombosis of the lower extremities is the most common occurrence. Arterial thrombosis occurs less frequently than venous thrombosis and primarily affects the brain (stroke) or coronary vessels (acute MI). During pregnancy, miscarriage may occur.

A second type of presentation is catastrophic APS, which is a TMA.[67] The kidney is the organ most often affected in these cases, followed by the lung, central nervous system, heart, and skin. Most patients with kidney involvement develop hypertension, and up to 25% develop progressive renal insufficiency, necessitating supportive dialysis. DIC may develop in the setting of catastrophic APS.

The prognosis of APS varies with the clinical severity, and management strategies vary accordingly. In patients who have antiphospholipid antibodies but have never experienced APS, it is not clear whether anticoagulation is indicated for the primary prevention of thrombosis. A careful review should be performed to eliminate conditions or situations that may increase the patient's risk for thrombosis. For example, oral contraceptive use should be discouraged, as should the use of venous catheters in hemodialysis patients. Prophylactic anticoagulation is indicated if the use of oral contraceptives or venous catheters cannot be eliminated.

In patients with antiphospholipid antibodies who have experienced a single episode of thrombosis, anticoagulation with warfarin is indicated for secondary prevention. The therapeutic goal is an INR of 2 to 3; therapy should probably be continued lifelong.

Mortality in patients with catastrophic APS is high—up to 50%. Death generally results from multiorgan failure. Treatment with plasmapheresis and plasma exchange, along with high-dose corticosteroid therapy at doses similar to those used for pulmonary-renal syndrome, has been used. Immunoglobulin therapy may also be tried, in an attempt to bind the offending autoantibody. Outcome data on the efficacy of these interventions for catastrophic APS remain sparse.

Renal Disease Associated with Systemic Sclerosis

Systemic sclerosis is a subtype of the scleroderma disorders, a varied group of diseases that generally share the finding of thickened, sclerotic skin lesions. Several organ systems are involved, including the heart, lungs, gastrointestinal tract, muscles, and kidneys.

Systemic sclerosis is relatively uncommon. In the United States, the prevalence ranges from four to 253 cases per one million population.[73] Females are more commonly affected than males, at a ratio of 3:1 to 8:1.[74] In the United States, there is a slightly higher occurrence in blacks than in whites.[75]

Renal involvement is common in systemic sclerosis. About 10% to 15% of patients with systemic sclerosis experience renal crisis, a syndrome marked by acute renal failure, severe hypertension, encephalopathy, and, in some cases, heart failure. More commonly, patients with systemic sclerosis have less dramatic evidence of renal involvement. Nearly all patients with systemic sclerosis have abnormal renal blood flow,[76] and two thirds have renal vascular lesions.[77] Half of all patients with systemic sclerosis may have hypertension, an elevated plasma creatinine concentration, or proteinuria.[78,79] Important risk factors for the development of renal disease in systemic sclerosis include black race, high-dose corticosteroid use (dosages above 15 mg/day), and diffuse skin involvement.

Most cases of systemic sclerosis are idiopathic. As with most connective tissue diseases, a complex interplay of host and environmental factors is undoubtedly involved. Some cases of systemic sclerosis overlap with other rheumatologic conditions, including SLE, rheumatoid arthritis, and undifferentiated mixed connective tissue disorder. Environmental exposure to organic solvents such as benzene, toluene, and trichloroethylene has been implicated as a secondary cause of systemic sclerosis. The proposed mechanisms have been that these organic solvents are metabolized to epoxy compounds that bind proteins and produce autoantigens[80] or that solvent-induced injury leads to microvascular abnormalities and target-organ fibrosis.[81] However, an extensive review found insufficient evidence to support a causal link between chemicals and systemic sclerosis.[82]

The pathognomonic findings in systemic sclerosis are the irreversible proliferation of connective tissue, mucoid thickening of vascular walls, and narrowing of the vascular lumen. In most cases, the mucoid changes are found in the renal arterioles. In renal crisis, these histopathologic changes are in the arcuate and interlobular arteries and in the glomeruli. Fibrin thrombi are seen in the arterioles and capillaries, as well as in areas of fibrinoid necrosis. Concentric onion-skin hypertrophy of the interlobular arteries results from repeated cycles of disease activity and quiescence.

The diagnosis of renal disease may not be difficult in the setting of established systemic sclerosis. The differential diagnosis includes other types of TMA. Serologic tests are helpful in confirming systemic sclerosis: anti–Scl-70 has a high specificity, albeit low sensitivity. Capillary loss and enlargement on nail-fold capillaroscopy are sensitive and specific.[83] Renal biopsy is not routinely indicated but should be done in atypical cases. For example, there is a small subset of patients who have systemic sclerosis without the typical skin findings (called scleroderma sine scleroderma). Some patients experiencing scleroderma renal crisis may not have elevated blood pressure. Biopsy of the skin or kidney may help confirm the diagnosis in these scenarios.

Scleroderma renal crisis was once deemed a uniformly fatal complication of systemic sclerosis. With the introduction of dialysis therapy, mortality decreased, and the use of pharmacologic therapy brought a further reduction of nearly 50% in morbidity and mortality.

Control of systemic arterial pressure is the cornerstone of therapy in systemic sclerosis with renal crisis. Blood pressure control is best achieved gradually, with reduction in pressure by 10 to 15 mm Hg a day until the diastolic blood pressure reaches 85 to 90 mm Hg. ACE inhibitors are the agents of choice; they have been shown to reduce the risk of ESRD[84] and the need for renal replacement therapy.[85] These benefits may flow from a unique property of this class of drugs or from greater efficacy in reduction of blood pressure. An important caveat is that the majority of studies have used captopril, and thus, the beneficial effect may not occur with other ACE inhibitors or with ARBs. The dihydropyridine calcium channel blocker nifedipine has been suggested as an additional blood pressure agent for patients who show an inadequate response to an ACE inhibitor. Recurrent renal crises are unusual if control of blood pressure is attained.

Malignant Hypertension

Malignant hypertension is defined as a syndrome of acute vascular damage to the kidneys, retina, and brain in the setting of severe elevations of blood pressure. Typically, the systolic blood pressure exceeds 180 mm Hg and the diastolic pressure exceeds 120 mm Hg.

Traditional risk factors for malignant hypertension are male gender, black race, and noncompliance with medication use. The frequency of malignant hypertension has been decreasing. This is likely the result of increased awareness, treatment of hypertension, and access to medical care for minorities.

Histologically, malignant hypertension is marked by diffuse fibrinoid necrosis and intimal thickening of small vessels (arterioles). In the kidney, these changes are indistinguishable from those seen in other TMAs, systemic sclerosis, and preeclampsia. Renal function is typically depressed and can decline rapidly over weeks to months in the absence of treatment.

Current understanding of the pathophysiology of malignant hypertension has focused on the role of the endothelium.[86] During severe elevations of blood pressure, endothelial control of vascular tone is overwhelmed, leading to end-organ hyperperfusion, arteriolar fibrinoid necrosis, and increased endothelial permeability with perivascular edema. The subsequent disruption of the endothelium leads to the loss of its fibrinolytic activity and exposes platelets to cellular adhesion molecules.

The prognosis of patients with malignant hypertension has improved tremendously over the past 40 years. This condition was originally termed malignant because it typically conferred the same abbreviation of life expectancy as an advanced malignancy. It is now known that aggressive blood pressure control with a variety of agents alleviates most of the target-organ damage [*see 1:III Hypertension*]. Initially, there were concerns that reduction of mean arterial pressure would cause progressive azotemia, but studies have instead demonstrated long-term stabilization or improvement of renal function in most cases.

Sickle Cell Disease

Sickle cell disease (SCD) is an autosomal recessive condition that results from the substitution of valine for glutamine at the sixth position of the β-hemoglobin subunit. Persons who are homozygous for sickle hemoglobin are susceptible to sickle cell crisis from capillary occlusion [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*]. Heterozygosity for sickle hemoglobin may provide protection against malaria.

SCD occurs primarily in persons of eastern Mediterranean and sub-Saharan African ancestry. In the United States, SCD predominantly affects the African-American population. SCD is a systemic vascular disease, with morbidity and mortality deriving primarily from vaso-occlusive injury to the lungs, central nervous system, spleen, and kidneys.[87]

Traditionally, hypoxemia has been considered central to the pathophysiology of sickle hemoglobin. Under low partial pressure of oxygen, polymerization of the hemoglobin molecule occurs, leading to the sickling of the erythrocyte; these sickled cells produce stasis and occlusion of capillaries. Recent investigations have also suggested that dysregulation of erythrocyte volume and interaction between the erythrocyte cell membrane and vascular endothelium may also contribute to vaso-occlusive events.[88] Finally, compensatory glomerular hypertrophy may have a role in the proteinuria and renal insufficiency seen in patients with SCD.[89]

a

b

Figure 7 **In an injection microradioangiogram of a normal kidney (*a*), vasa recta radiating into the renal papilla are visible. In contrast, in an injection microradioangiogram from a patient with sickle cell disease (*b*), vasa recta are virtually absent.[62]**

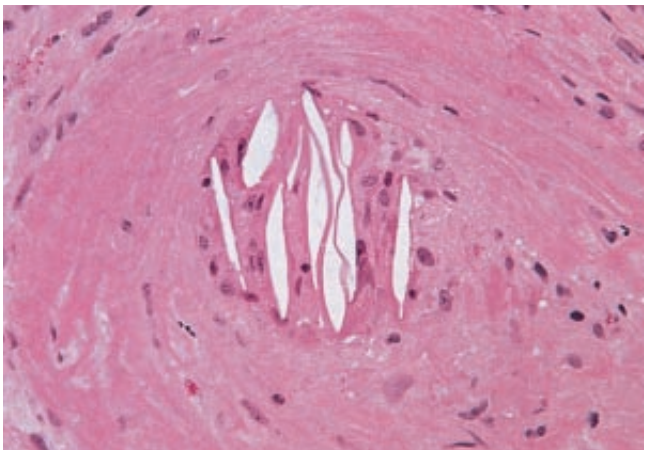


Figure 8 **Light microscopy reveals an atheroembolus lodged in a small renal artery and occlusion of the vascular lumen. The biconcave parallel slits, which represent cholesterol crystals that have dissolved during paraffin fixation, are characteristic of this disorder.**

Renal involvement is relatively common in SCD. The renal medulla, an area of chronic hypoxia, is the principal site of involvement. Stasis of erythrocytes in the vasa recta of the medulla can lead to focal areas of hemorrhage and necrosis and eventually to interstitial inflammation, fibrosis, tubular atrophy, and papillary infarction (necrosis). Over time, these vaso-occlusive events lead to an almost complete loss of the vasa recta [*see Figure 7*].

Three pathologic renal lesions have been described in SCD. Necrosis of the renal papillae is a frequent complication, occurring in 15% to 67% of patients.[90-92] It develops as a result of papillary infarction from repeated vaso-occlusive events. Patients with papillary necrosis may be asymptomatic or may present with gross hematuria and renal failure. Focal segmental glomerulosclerosis (FSGS) is the most common glomerulopathy found in patients with SCD and proteinuria.[93-95] FSGS has been proposed to result from hyperfiltration, exposure to reactive oxygen species, and exposure to filtered plasma proteins. A third pathologic lesion, type I membranoproliferative glomerulopathy, has been described in only a few patients with SCD.[89]

In addition to the pathologic findings, 26% of patients with SCD have proteinuria and up to 7% have some level of renal insufficiency[95]; a significant minority progresses to ESRD. Diminished urinary concentrating ability is almost universal in SCD. Vaso-occlusive episodes in the vasa recta are thought to interfere with countercurrent exchange, leading to a maximum urine osmolality of 400 to 450 mOsm/kg. An incomplete distal renal tubular acidosis may also develop because of impaired hydrogen and potassium secretion.

The diagnosis of renal disease associated with SCD is generally not difficult to make. Renal complications tend to occur late in the course of disease, when the primary diagnosis has already been made [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*].

The prognosis of SCD has improved tremendously over the past 30 years. Sickle cell anemia was initially considered to be a childhood disease, because the vast majority of patients did not survive into adulthood. Survival has been improving, however,[87,96] with a greater proportion of patients reaching young adulthood. Better supportive care, as well as aggressive treatment of infections, may account for the improvement. Induction of fetal hemoglobin production has been a mainstay of therapy. Hydroxyurea, a myelosuppressive antitumor drug, has been shown to reduce the occurrence of acute chest, painful crises, and need for transfusions.[88] Future directions include the development of drugs that may inhibit polymerization of the sickle hemoglobin molecule, bone marrow transplantation, and the introduction of wild type β-hemoglobin via a virus vector.

Atheroembolism

Atheroembolism is a rare condition that affects several vascular beds but, most notably, the small vessels of the kidney. It is characterized by the deposition of cholesterol crystals in glomerular capillaries and arterioles. Other terms for this condition are cholesterol crystal embolization disease, cholesterol atheroembolic disease, and cholesterol embolism.

The epidemiology of atheroembolic disease is not clear.[97] It is not possible to estimate incidence and prevalence in the general population; most information is abstracted from highly selected groups. As a rule of thumb, atheroembolic disease tends to occur in persons with risk factors for atherosclerosis and with extensive generalized atherosclerosis. Thus, risk factors include age over 55 years, hypercholesterolemia, diabetes mellitus and hypertension, peripheral vascular disease, abdominal aortic atherosclerosis, and atherosclerotic RAS. Episodes of atheroembolism almost always follow manipulation of or trauma to an atheromatous plaque. Thus, episodes may occur after arteriography, especially of the large vessels such as the aorta, renal arteries, and coronary arteries, or after thrombolytic therapy. In one study, atheroembolism occurred in 12% of patients undergoing coronary angiography and bypass surgery.[98]

Histologically, the lesion of atheroembolic disease is characterized by the occlusion of arcuate and interlobular arteries (150 to 200 μm in diameter) and glomerular capillaries by embolized cholesterol crystals. The crystals dissolve during routine preparation of tissue but leave a characteristic biconvex, needle-shaped cleft outline, referred to as a ghost [*see Figure 8*]. Initially, infiltration of polymorphonuclear leukocytes and eosinophils is seen around the occluded vessel. Over time, the crystals can attract a mononuclear inflammatory cellular infiltrate, and intimal hyperplasia and perivascular fibrosis of the affected vessels may occur. Recanalization may be observed in the occluded vessel.

Large vessel atherosclerosis is obviously an important component in the pathogenesis of atheroembolic disease. Atheromatous plaques in the aorta and renal arteries are typically the source of the cholesterol emboli to the kidney. Activation of the classic or alternative complement pathways by cholesterol crystals may help explain the inflammatory component of the disease observed on biopsy.

Atheroembolic disease is difficult to diagnose. Cases may occur spontaneously, but most are temporally related to invasive vascular procedures or thrombolytic therapy. However, the temporal relationships to these invasive vascular procedures are variable; the disease may occur immediately after the event or may develop insidiously over a period of weeks or months. Furthermore, patients typically have symptoms and signs involving multiple organs: kidneys (renal insufficiency, proteinuria, hematuria), skin (livedo reticularis, purple toes, gangrene), gastrointestinal tract (anorexia, nausea, bowel ischemia or infarction, pancreatitis, abnormal liver enzymes), nervous system (headache, amaurosis fugax, mononeuropathy, cerebrovascular disease), and eyes (Hollenhorst plaques). Furthermore, nonspecific consti-

tutional symptoms such as fever, malaise, and weight loss are often present and may further obscure the diagnosis. Laboratory findings include a hypocomplementemia (both C3 and C4) and elevated inflammatory markers (ESR, C-reactive protein level); some patients have elevated levels of serum creatinine and blood urea nitrogen, proteinuria, and hematuria.

A practical approach to diagnosis of renal atheroembolic disease is to maintain a high index of suspicion for the disorder in patients with acute renal failure, especially after invasive vascular procedures or thrombolytic therapy. For example, 80% of patients with atheroembolism and renal involvement have a serum creatinine level above 2 mg/dl and hypertension (with worsening control).[99] Cutaneous manifestations that suggest atheroembolism include livedo reticularis and purple digits (especially toes). Low complement levels further strengthen the case for atheroembolic disease. Diagnosis is made by biopsy of affected skin, muscle, or kidney.

The prognosis for patients with atheroembolic disease, especially with renal involvement, is poor. Mortality varies from 60% to 80%.[99] There is no specific therapy for the condition, so treatment usually consists of supportive care (including dialysis).

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Harding MB, Smith LR, Himmelstein SI, et al: Renal artery stenosis: prevalence and associated risk factors in patients undergoing routine cardiac catheterization. J Am Soc Nephrol 2:1608, 1992

2. Missouris CG, Buckenham T, Cappuccio FP, et al: Renal artery stenosis: a common and important problem in patients with peripheral vascular disease. Am J Med 96:10, 1994

3. Uzu T, Inoue T, Fujii T, et al: Prevalence and predictors of renal artery stenosis in patients with myocardial infarction. Am J Kidney Dis 29:733, 1997

4. Kuroda S, Nishida N, Uzu T, et al: Prevalence of renal artery stenosis in autopsy patients with stroke. Stroke 31:61, 2000

5. Svetkey LP, Kadir S, Dunnick NR, et al: Similar prevalence of renovascular hypertension in selected blacks and whites. Hypertension 17:678, 1991

6. Emovon OE, Klotman PE, Dunnick NR, et al: Renovascular hypertension in blacks. Am J Hypertens 9:18, 1996

7. Goldblatt H, Lynch J, Hanzal RF, et al: Studies on experimental hypertension: 1. The production of persistent elevation of systolic blood pressure by means of renal ischemia. J Exp Med 59:347, 1934

8. McLean AG, Hilson AJ, Scoble JE, et al: Screening for renovascular disease with captopril-enhanced renography. Nephrol Dial Transplant 7:211, 1992

9. Thornton J, O'Callaghan J, Walshe J, et al: Comparison of digital subtraction angiography with gadolinium-enhanced magnetic resonance angiography in the diagnosis of renal artery stenosis. Eur Radiol 9:930, 1999

10. Ramsay LE, Waller PC: Blood pressure response to percutaneous transluminal angioplasty for renovascular hypertension: an overview of published series. BMJ 300:569, 1990

11. Leertouwer TC, Gussenhoven EJ, Bosch JL, et al: Stent placement for renal arterial stenosis: where do we stand? A meta-analysis. Radiology 216:78, 2000

12. Isles CG, Robertson S, Hill D: Management of renovascular disease: a review of renal artery stenting in ten studies. QJM 92:159, 1999

13. Alcazar JM, Rodicio JL: Ischemic nephropathy: clinical characteristics and treatment. Am J Kidney Dis 36:883, 2000

14. Radermacher J, Chavan A, Bleck J, et al: Use of Doppler ultrasonography to predict the outcome of therapy for renal-artery stenosis. N Engl J Med 344:410, 2001

15. Bonelli FS, McKusick MA, Textor SC, et al: Renal artery angioplasty: technical results and clinical outcome in 320 patients. Mayo Clin Proc 70:1041, 1995

16. van Bockel JH, van Schilfgaarde R, Felthuis W, et al: Long-term results of in situ and extracorporeal surgery for renovascular hypertension caused by fibrodysplasia. J Vasc Surg 6:355, 1987

17. Lupi-Herrera E, Sanchez-Torres G, Marcushamer J, et al: Takayasu's arteritis: clinical study of 107 cases. Am Heart J 93:94, 1977

18. Subramanyan R, Joy J, Balakrishnan KG: Natural history of aortoarteritis (Takayasu's disease). Circulation 80:429, 1989

19. Arend WP, Michel BA, Bloch DA, et al: The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. Arthritis Rheum 33:1129, 1990

20. Kerr GS, Hallahan CW, Giordano J, et al: Takayasu arteritis. Ann Intern Med 120:919, 1994

21. Hoffman GS, Leavitt RY, Kerr GS, et al: Treatment of glucocorticoid-resistant or relapsing Takayasu arteritis with methotrexate. Arthritis Rheum 37:578, 1994

22. Valsakumar AK, Valappil UC, Jorapur V, et al: Role of immunosuppressive therapy on clinical, immunological, and angiographic outcome in active Takayasu's arteritis. J Rheumatol 30:1793, 2003

23. Ishikawa K: Natural history and classification of occlusive thromboaortopathy (Takayasu's disease). Circulation 57:27, 1978

24. Jennette JC, Falk RJ, Andrassy K, et al: Nomenclature of systemic vasculitides: proposal of an international consensus conference. Arthritis Rheum 37:187, 1994

25. Sack M, Cassidy JT, Bole GG: Prognostic factors in polyarteritis. J Rheumatol 2:411, 1975

26. Scott DG, Bacon PA, Elliott PJ, et al: Systemic vasculitis in a district general hospital 1972–1980: clinical and laboratory features, classification and prognosis of 80 cases. QJM 51:292, 1982

27. McMahon BJ, Heyward WL, Templin DW, et al: Hepatitis B-associated polyarteritis nodosa in Alaskan Eskimos: clinical and epidemiologic features and long-term follow-up. Hepatology 9:97, 1989

28. Lam KC, Lai CL, Trepo C, et al: Deleterious effect of prednisolone in HBsAg-positive chronic active hepatitis. N Engl J Med 304:380, 1981

29. Guillevin L, Lhote F, Gayraud M, et al: Prognostic factors in polyarteritis nodosa and Churg-Strauss syndrome: a prospective study in 342 patients. Medicine (Baltimore) 75:17, 1996

30. Pettersson EE, Sundelin B, Heigl Z: Incidence and outcome of pauci-immune necrotizing and crescentic glomerulonephritis in adults. Clin Nephrol 43:141, 1995

31. Savage CO, Harper L, Adu D: Primary systemic vasculitis. Lancet 349:553, 1997

32. Trimethoprim-sulfamethoxazole (co-trimoxazole) for the prevention of relapses of Wegener's granulomatosis. Dutch Co-Trimoxazole Wegener Study Group. N Engl J Med 335:16, 1996

33. Panegyres PK, Blumbergs PC, Leong AS, et al: Vasculitis of peripheral nerve and skeletal muscle: clinicopathological correlation and immunopathic mechanisms. J Neurol Sci 100:193, 1990

34. Kissel JT, Mendell JR: Vasculitic neuropathy. Neurol Clin 10:761, 1992

35. Jennette JC, Falk RJ: Small-vessel vasculitis. N Engl J Med 337:1512, 1997

36. Hogan SL, Nachman PH, Wilkman AS, et al: Prognostic markers in patients with antineutrophil cytoplasmic autoantibody-associated microscopic polyangiitis and glomerulonephritis. J Am Soc Nephrol 7:23, 1996

37. Jayne D, Rasmussen N, Andrassy K, et al: A randomized trial of maintenance therapy for vasculitis associated with antineutrophil cytoplasmic autoantibodies. European Vasculitis Study Group. N Engl J Med 349:36, 2003

38. Pusey CD: Anti-glomerular basement membrane disease. Kidney Int 64:1535, 2003

39. Hellmark T, Niles JL, Collins AB, et al: Comparison of anti-GBM antibodies in sera with or without ANCA. J Am Soc Nephrol 8:376, 1997

40. Weening JJ, D'Agati VD, Schwartz MM, et al: The classification of glomerulonephritis in systemic lupus erythematosus revisited. J Am Soc Nephrol 15:241, 2004

41. Chan TM, Li FK, Tang CS, et al: Efficacy of mycophenolate mofetil in patients with diffuse proliferative lupus nephritis. Hong Kong–Guangzhou Nephrology Study Group. N Engl J Med 343:1156, 2000

42. Contreras G, Pardo V, Leclercq B, et al: Sequential therapies for proliferative lupus nephritis. N Engl J Med 350:971, 2004

43. Moake JL: Thrombotic microangiopathies. N Engl J Med 347:589, 2001

44. Warwicker P, Goodship TH, Donne RL, et al: Genetic studies into inherited and sporadic hemolytic uremic syndrome. Kidney Int 53:836, 1998

45. Caprioli J, Bettinaglio P, Zipfel PF, et al: The molecular basis of familial hemolytic uremic syndrome: mutation analysis of factor H gene reveals a hot spot in short consensus repeat 20. Italian Registry of Familial and Recurrent HUS/TTP. J Am Soc Nephrol 12:297, 2001

46. Sanchez-Corral P, Perez-Caballero D, Huarte O, et al: Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. Am J Hum Genet 71:1285, 2002

47. Noris M, Brioschi S, Caprioli J, et al: Familial haemolytic uraemic syndrome and an MCP mutation. International Registry of Recurrent and Familial HUS/TTP. Lancet 362:1542, 2003

48. Bell WR, Braine HG, Ness PM, et al: Improved survival in thrombotic thrombocytopenic purpura-hemolytic uremic syndrome: clinical experience in 108 patients. N Engl J Med 325:398, 1991

49. Rosove MH, Ho WG, Goldfinger D: Ineffectiveness of aspirin and dipyridamole in the treatment of thrombotic thrombocytopenic purpura. Ann Intern Med 96:27, 1982

50. Karmali MA, Petric M, Lim C, et al: The association between idiopathic hemolytic uremic syndrome and infection by verotoxin-producing *Escherichia coli*. J Infect Dis 151:775, 1985

51. Cleary TG: Cytotoxin-producing *Escherichia coli* and the hemolytic uremic syndrome. Pediatr Clin North Am 35:485, 1988

52. Banatvala N, Griffin PM, Greene KD, et al: The United States National Prospective Hemolytic Uremic Syndrome Study: microbiologic, serologic, clinical, and epidemiologic findings. Hemolytic Uremic Syndrome Study Collaborators. J Infect Dis 183:1063, 2001

53. Misiani R, Appiani AC, Edefonti A, et al: Haemolytic uraemic syndrome: therapeutic effect of plasma infusion. Br Med J (Clin Res Ed) 285:1304, 1982

54. Sheth KJ, Gill JC, Hanna J, et al: Failure of fresh frozen plasma infusions to alter the course of hemolytic uremic syndrome. Child Nephrol Urol 9:38, 1988–89

55. Roberts JM, Cooper DW: Pathogenesis and genetics of pre-eclampsia. Lancet 357:53, 2001

56. Trupin LS, Simon LP, Eskenazi B: Change in paternity: a risk factor for preeclampsia in multiparas. Epidemiology 7:240, 1996

57. Dekker G, Sibai B: Primary, secondary, and tertiary prevention of pre-eclampsia. Lancet 357:209, 2001

58. Levine RJ, Maynard SE, Qian C, et al: Circulating angiogenic factors and the risk of preeclampsia. N Engl J Med 350:672, 2004

59. Sibai BM, Caritis SN, Thom E, et al: Prevention of preeclampsia with low-dose aspirin in healthy, nulliparous pregnant women. The National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. N Engl J Med 329:1213, 1993

60. Levine RJ, Hauth JC, Curet LB, et al: Trial of calcium to prevent preeclampsia. N Engl J Med 337:69, 1997

61. Knight M, Duley L, Henderson-Smart DJ, et al: Antiplatelet agents for preventing and treating pre-eclampsia. Cochrane Database Syst Rev (2):CD000492, 2000

62. Sibai BM: Treatment of hypertension in pregnant women. N Engl J Med 335:257, 1996

63. Lindheimer MD, Davison JM: Renal biopsy during pregnancy: 'to b . . . or not to b . . .?'. Br J Obstet Gynaecol 94:932, 1987

64. Packham D, Fairley KF: Renal biopsy: indications and complications in pregnancy. Br J Obstet Gynaecol 94:935, 1987

65. Walker JJ: Pre-eclampsia. Lancet 356:1260, 2000

66. Greaves M: Antiphospholipid antibodies and thrombosis. Lancet 353:1348, 1999

67. Levine JS, Branch DW, Rauch J: The antiphospholipid syndrome. N Engl J Med 346:752, 2002

68. Petri M: Epidemiology of the antiphospholipid antibody syndrome. J Autoimmun 15:145, 2000

69. Cervera R, Khamashta MA, Font J, et al: Systemic lupus erythematosus: clinical and immunologic patterns of disease expression in a cohort of 1,000 patients. The European Working Party on Systemic Lupus Erythematosus. Medicine (Baltimore) 72:113, 1993

70. Merkel PA, Chang Y, Pierangeli SS, et al: The prevalence and clinical associations of anticardiolipin antibodies in a large inception cohort of patients with connective tissue diseases. Am J Med 101:576, 1996

71. Love PE, Santoro SA: Antiphospholipid antibodies: anticardiolipin and the lupus anticoagulant in systemic lupus erythematosus (SLE) and in non-SLE disorders: prevalence and clinical significance. Ann Intern Med 112:682, 1990

72. Alarcon-Segovia D, Perez-Vazquez ME, Villa AR, et al: Preliminary classification criteria for the antiphospholipid syndrome within systemic lupus erythematosus. Semin Arthritis Rheum 21:275, 1992

73. Lawrence RC, Helmick CG, Arnett FC, et al: Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 41:778, 1998

74. Silman AJ: Epidemiology of scleroderma. Ann Rheum Dis 50:846, 1991

75. Medsger TA, Masi AT: The epidemiology of systematic sclerosis (scleroderma) Ann Intern Med 74:714, 1971

76. Livi R, Teghini L, Pignone A, et al: Renal functional reserve is impaired in patients with systemic sclerosis without clinical signs of kidney involvement. Ann Rheum Dis 61:682, 2002

77. Kovalchik MT, Guggenheim SJ, Silverman MH, et al: The kidney in progressive systemic sclerosis: a prospective study. Ann Intern Med 89:881, 1978

78. Tuffanelli DL, Winkelmann RK: Systemic scleroderma, a clinical study of 727 cases. Arch Dermatol 84:359, 1961

79. Traub YM, Shapiro AP, Rodnan GP, et al: Hypertension and renal failure (scleroderma renal crisis) in progressive systemic sclerosis: review of a 25-year experience with 68 cases. Medicine (Baltimore) 62:335, 1983

80. Silman AJ, Hochberg MC: Occupational and environmental influences on scleroderma. Rheum Dis Clin N Am 22:737, 1996

81. Haustein U-F, Herrmann K: Environmental scleroderma. Clin Dermatol 12:467, 1994

82. Garabrant DH, Dumas C: Epidemiology of organic solvents and connective tissue disease. Arthritis Res 2:5, 2000

83. Ferri C, Bernini L, Cecchetti R, et al: Cutaneous and serologic subsets of systemic sclerosis. J Rheumatol 18:1826, 1991

84. Steen VD, Medsger TA Jr: Case-control study of corticosteroids and other drugs that either precipitate or protect from the development of scleroderma renal crisis. Arthritis Rheum 41:1613, 1998

85. Steen VD, Medsger TA Jr: Outcome of scleroderma renal crisis (abstr). J Am Soc Nephrol 8:148A, 1997

86. Vaughan CJ, Delanty N: Hypertensive emergencies. Lancet 356:411, 2000

87. Platt OS, Brambilla DJ, Rosse WF, et al: Mortality in sickle cell disease: life expectancy and risk factors for early death. N Engl J Med 330:1639, 1994

88. Bunn HF: Pathogenesis and treatment of sickle cell disease. N Engl J Med 337:762, 1997

89. Wesson DE: The initiation and progression of sickle cell nephropathy. Kidney Int 61:2277, 2002

90. Addae SK: The kidney in sickle cell disease. V: Clinical manifestations. Ghana Med J 12:352, 1973

91. Flaster S, Lome LG, Presman D: Urologic complications of renal papillary necrosis. Urology 5: 331, 1975

92. McCall JW, Moule N, Desal P: Urologic findings in homozygous sickle cell disease. Radiology 126:99, 1978

93. Tejani A, Phadke K, Adamson O, et al: Renal lesions in sickle cell nephropathy in children. Nephron 39:352, 1985

94. Bhathena DB, Sondheimer JH: The glomerulopathy of homozygous sickle hemoglobin (SS) disease: morphology and pathogenesis. J Am Soc Nephrol 1:1241, 1991

95. Falk RJ, Scheinman J, Phillips G, et al: Prevalence and pathologic features of sickle cell nephropathy and response to inhibition of angiotensin-converting enzyme. N Engl J Med 326:910, 1992

96. Quinn CT, Rogers ZR, Buchanan GR: Survival of children with sickle cell disease. Blood 103:4023, 2004

97. Modi KS, Rao VK: Atheroembolic renal disease. J Am Soc Nephrol 12:1781, 2001

98. Blankenship JC, Butler M, Garbes A: Prospective assessment of cholesterol embolization in patients with acute myocardial infarction treated with thrombolytic vs conservative therapy. Chest 107:662, 1995

99. Fine MJ, Kapoor W, Falanga V: Cholesterol crystal embolization: a review of 221 cases in the English literature. Angiology 38:769, 1987

Acknowledgments

Figures 3, 5, 6, and 8 Jerome Jacobs.

Figure 4 Gregory Bishop.

Figure 7 Courtesy of Dr. P. E. de Jong, Department of Medicine, State University, the Netherlands.

VIII TUBULOINTERSTITIAL DISEASES

Gerald B. Appel, m.d.

Overview

Tubulointerstitial diseases involve the tubules and the interstitium (the space between the tubules), rather than the glomeruli and the vasculature of the kidney [*see Figure 1*].[1-3] Tubulointerstitial diseases may be acute or chronic and may present as a primary condition or may be associated with systemic disease. They can cause subtle tubular defects or fulminant renal failure and can lead to structural and functional changes that are either reversible or permanent. Although tubulointerstitial diseases make up only a small percentage of the cases of chronic renal failure and end-stage renal disease in the United States, damage to the tubulointerstitium of the kidney contributes significantly to the progression of most of the important forms of renal disease. Numerous studies have found that in many forms of chronic renal failure, including chronic renal failure of glomerular origin, glomerular filtration correlates better with the degree of tubulointerstitial damage than with the degree of glomerular injury. There are numerous potential mechanisms for tubulointerstitial damage. Some involve immunologic insults (e.g., antibodies directed against tubular basement membrane, immune complex deposition along the tubular basement, or cell-mediated immune responses), whereas other insults relate to toxic damage to tubular cells or other structural components of the interstitium. How these primary insults of tubulointerstitial damage are translated into chronic tubulointerstitial damage and fibrosis is an area of intensive investigation.[4,5]

Specific laboratory findings in patients who have tubulointerstitial diseases reflect the underlying etiology; the extent of injury, whether focal or diffuse; and the chronicity of the insult [*see Table 1*]. The urinary sediment often contains erythrocytes, leukocytes, and white blood cell casts. Proteinuria, when present, predominantly results from loss of low-molecular-weight proteins (e.g., β_2-microglobulin) rather than loss of albumin. Common defects in tubulointerstitial diseases include isolated tubular defects (e.g., renal tubular acidosis, glycosuria, aminoaciduria, potassium wasting or retention, and magnesium wasting) and urinary concentrating defects (manifested by nocturia and polyuria). Hypertension is often less severe than it is in the glomerulonephritides [*see 10:V Glomerular Diseases*]; salt wasting and a predisposition to volume depletion are common findings. Metabolic acidosis, when present, often results from defective ammoniagenesis and is associated with a normal anion gap rather than a high anion gap, which is seen in severe glomerular disease. Findings typical of glomerular disease—heavy albuminuria (> 2 g/day), red blood cell casts, and deformed erythrocytes in the urine sediment—are usually absent in tubulointerstital diseases unless there is associated glomerular pathology.

Drug-Induced Acute Interstitial Nephritis

Acute interstitial nephritis (AIN) is a form of acute tubulointerstitial damage that is usually related to drug use and is associated with acute renal failure.[6-8] AIN is found in 1% to 2% of all renal biopsies but in as many as 15% of renal biopsies in patients with acute renal failure.[9] It is important to recognize medication-related AIN because severe or irreversible renal damage is often preventable or correctable. AIN has been associated with an ever-increasing list of medications that can be classified into a number of pharmacologic or functional classes [*see Table 2*].

antibiotic-induced AIN

AIN Caused by β-Lactam Antibiotics

Virtually all β-lactam antibiotics (penicillins and cephalosporins) can produce AIN, but methicillin has been the most common offending agent in this class.[6-8] Neither simple pharmacokinetic studies nor tissue-binding studies explain this predilection. AIN from β-lactam antibiotics can affect both males and females of any age. It usually occurs after several weeks of high-dose antibiotic therapy.

Pathogenesis The pathogenesis of β-lactam–associated AIN remains unknown.[2-4,6-8] The disease is not dose related and occurs in only a small number of the millions of people taking β-lactam drugs each year. It can recur or can be exacerbated on rechallenge with a second β-lactam drug. Only rarely has evidence of circulating anti–tubular basement membrane antibodies been found. The nature of the cellular infiltrates supports a role for cell-mediated or eosinophil-mediated damage. Whether subclinical episodes of AIN contribute to chronic renal failure in some patients is unknown.

Diagnosis Classically, patients exhibit a triad of hypersensitivity reactions: rash, fever, and eosinophilia. The secondary fever associated with AIN usually occurs after defervescence from the original infectious disease and during the onset of the allergic reaction. Eosinophilia may vary from only a few percent to more than 20%. Urinary findings in patients with AIN include the nonspecific findings of sterile pyuria and mild proteinuria, as well as the more significant finding of hematuria, which may be gross in some cases. Eosinophils may be found in the urinary sediment on Wright or Hansel staining and are strongly suggestive of AIN but cannot confirm the diagnosis.[9] Although

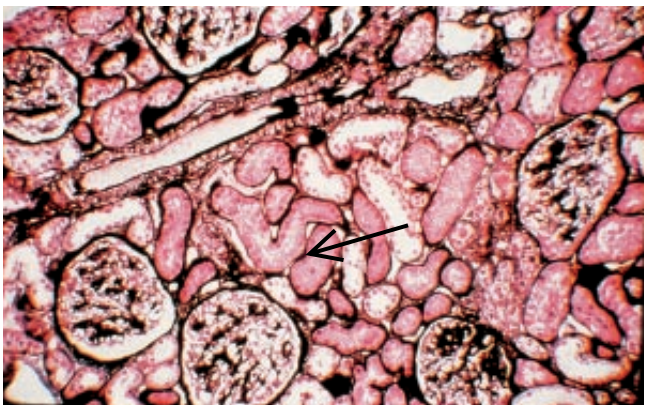


Figure 1 **Normal tubulointerstitial architecture can be observed in this micrograph. The tubules, outlined in black by silver stain, are back to back (arrow). The interstitium is the space between these tubules, the vessels, and the glomeruli.**

Table 1 Etiologies of Tubulointerstitial Diseases

- Drug-induced acute interstitial nephritis (AIN)
- Drug-induced chronic interstitial nephritis: analgesic agents, lithium, cisplatin, nitrosoureas, and cyclosporine
- Physical factors: reflux, obstruction, and radiation nephritis
- Infections: pyelonephritis and remote infections
- Metabolites and toxins: heavy metals, urate, calcium, oxalate, and hypokalemia
- Vascular conditions: hypertension, sickle cell anemia, cholesterol emboli, and renal vein thrombosis
- Dysproteinemias
- Cystic renal disease
- Other causes: transplant rejection, Balkan nephropathy, systemic diseases, and idiopathic AIN

eosinophiluria is not specific for AIN (it is also seen in chronic pyelonephritis, rapidly progressive glomerulonephritis, acute interstitial cystitis, prostatitis, and renal atheroemboli), it is highly suggestive of AIN in a patient with acute renal failure.[6-8]

Occasionally, patients experience only isolated renal tubular defects (e.g., Fanconi syndrome, renal tubular acidosis, and potassium excretory defects), but most patients have progressive acute renal failure, with rising blood urea nitrogen (BUN) and serum creatinine levels. Only a minority of patients with AIN are oliguric. Gallium scanning may show diffuse, intense bilateral uptake, whereas it is typically negative in acute tubular necrosis, which is the primary differential diagnosis for AIN.[10]

The histopathology of β-lactam–associated AIN includes edema and focal renal infiltrates that contain predominantly T cells and monocytes.[1,6,8] In cases in which biopsies are done early in the disease course, the infiltrates may include eosinophils, plasma cells, and neutrophils [*see Figures 2 and 3*], with inflammatory cells at times invading the spaces between the cells lining the tubular basement membrane (so-called tubulitis). In some cases, granuloma formation may be seen. Immunofluorescence or electron microscopy seldom reveals any changes.

Treatment β-lactam–associated AIN is treated by discontinuing the drug and avoiding other β-lactam antibiotics.[6-8,11] Most patients regain renal function, and many regain baseline renal function. Even patients requiring dialysis supportive care may regain renal function.

The use of corticosteroids to treat renal failure associated with AIN remains controversial. No randomized, controlled trials have yet proved that corticosteroid therapy has any advantages over discontinuance of medication. Some retrospective reviews suggest improved renal function and less residual interstitial damage in patients treated with a short course of corticosteroids for severe AIN. Other studies have reported no benefit with corticosteroid therapy.

AIN Caused by Other Antibiotics

Other antibiotics associated with a pattern of acute renal failure consistent with AIN include the sulfonamides, the combination of trimethoprim and sulfamethoxazole, vancomycin, and the quinolone antibiotics, including ciprofloxacin.[6-8] They can produce the classic pattern of rash, fever, and eosinophilia. Rifampin is most commonly associated with AIN when used intermittently (several times a week) or on rechallenge after a brief suspension of drug use. The clinical features are acute oliguric renal failure with back and loin pain; the triad of rash, fever, and eosinophilia; and AIN on renal biopsy.

A number of types of diuretics can produce AIN, but prerenal azotemia from volume depletion is a much more common cause of diuretic-induced renal insufficiency. AIN can also result from the use of such drugs as sulfinpyrazone, cimetidine, ranitidine, aminosalicylates (e.g., mesalamine), and allopurinol [*see Table 2*]. Allopurinol may produce a vasculitis or glomerulonephritis concurrently with AIN.

NSAID-INDUCED AIN

The nonsteroidal anti-inflammatory drugs (NSAIDs) produce a unique clinical picture of AIN.[6-8] Patients with the NSAID-induced disorder are typically older and have taken the agent for many weeks or months before the reaction occurs.

Table 2 Drugs Associated with Acute Interstitial Nephritis

β-Lactam Antibiotics
- Methicillin*
- Penicillin G
- Ampicillin
- Flucloxacillin
- Oxacillin
- Nafcillin
- Carbenicillin
- Amoxicillin
- Mezlocillin
- Piperacillin
- Cephalothin
- Cephalexin
- Cephradine
- Cefotaxime
- Cefoxitin
- Cefaclor
- Cefazolin
- Cefotetan

Other Antibiotics
- Sulfonamides*
- Trimethoprim-sulfamethoxazole*
- Rifampin*
- Polymyxin B sulfate
- Ethambutol
- Vancomycin
- Chloramphenicol
- Gentamicin?
- Isoniazid?
- Minocycline
- Aminosalicylic acid
- Ciprofloxacin
- Nitrofurantoin
- Norfloxacin
- Erythromycin
- Spiramycin
- Acyclovir
- Foscarnet

Diuretics
- Thiazides*
- Furosemide
- Chlorthalidone
- Triamterene
- Indapamide

Nonsteroidal Anti-inflammatory Drugs
- Fenoprofen*
- Indomethacin
- Naproxen
- Ibuprofen
- Mefenamic acid
- Tolmetin
- Diflunisal
- Piroxicam
- Diclofenac
- Ketoprofen
- Suprofen
- Sulindac

Other Drugs
- Phenytoin*
- Cimetidine*
- Omeprazole
- Sulfinpyrazone*
- Allopurinol*
- Aspirin
- Carbamazepine
- Clofibrate
- Azathioprine
- Phenylpropanolamine
- Methyldopa
- Phenobarbital
- Interferon alfa
- Floctafenine
- Haloperidol
- Warfarin
- Diazepam
- Valproate
- Chlorprothixene
- Captopril
- Propranolol
- Amphetamines
- Doxepin
- Quinine
- Ranitidine
- Interleukin-2
- Propylthiouracil

*Most common causative agents af AIN. ? May cause AIN.

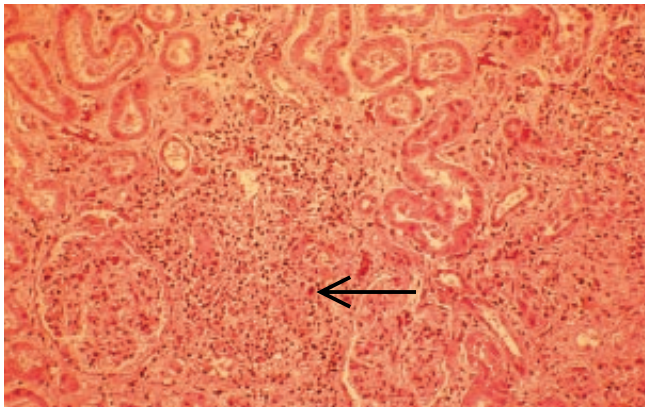


Figure 2 **Acute interstitial nephritis (AIN) caused by a β-lactam antibiotic is apparent in this low-power light micrograph. The patchy infiltrate (arrow) obscures some but not all of the tubular architecture.**

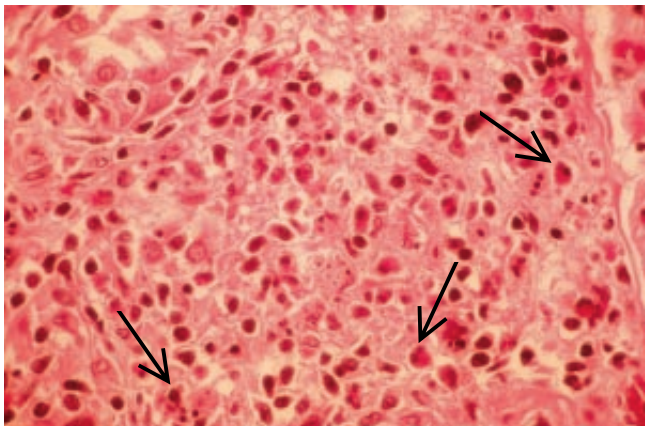


Figure 3 **The pleomorphic infiltrate, including binucleated eosinophils (arrows), which is characteristic of drug-induced AIN, can be seen in this high-power light micrograph.**

Pathogenesis

The pathogenesis of this lesion remains unknown. Studies with monoclonal antibodies have revealed no consistent differences between β-lactam–associated and NSAID-associated AIN in terms of number and types of invading lymphocytes.[12]

Diagnosis

Rash, fever, and eosinophilia may not be present, even when eosinophilic infiltrates are noted on kidney biopsy. The most striking finding is the association of AIN with the glomerular lesions of minimal-change nephrotic syndrome. Although isolated cases of this syndrome have been reported in patients taking ampicillin, rifampin, interferon, or ranitidine, NSAIDs are the only drugs frequently reported to cause AIN associated with a glomerular lesion. This has been reported most frequently with fenoprofen, but it has been seen with many structurally dissimilar NSAIDs, such as the cyclooxygenase-2 (COX-2) inhibitors.[13,14] Light microscopy shows normal glomeruli (fused foot processes are revealed by electron microscopy) and the tubulointerstitial changes of AIN.[2,15]

Treatment

Reversal of renal failure and remission of the nephrotic syndrome usually occur after discontinuance of the NSAID. Some patients with persistent renal failure or unremitting nephrotic syndrome have been treated successfully with a short course of corticosteroids. Membranous nephropathy with or without AIN has been reported in association with NSAIDs, including COX-2 inhibitors.[13,14]

Drug-Induced Chronic Interstitial Nephritis

Certain drugs have been associated with the development of chronic renal insufficiency and chronic interstitial damage. The relation between drug use and renal lesions in chronic interstitial nephritis (CIN) has often been more difficult to define than that in AIN. This difficulty arises from the slower, more insidious nature of the disease process in CIN and from the complex medication regimens of many patients with AIN or CIN who develop renal lesions. Drugs that can cause CIN include analgesic agents such as phenacetin, acetaminophen, and aspirin; lithium; the antineoplastic agents cisplatin, semustine, and carmustine; and cyclosporine.

ANALGESIC NEPHROPATHY

Epidemiology

Analgesic agents in over-the-counter preparations are extensively used, and there is considerable concern and controversy concerning their nephrotoxic potential.[16-19] Epidemiologic studies have provided conflicting reports on the incidence and causal agents of analgesic nephropathy; however, the mixed findings are largely due to the retrospective nature and design limitations of the studies.[15,17,18] The pathogenesis of chronic renal insufficiency and interstitial damage produced by analgesic agents is influenced by social, cultural, and personal factors, as well as by the type and dose of analgesic agent taken and the duration of therapy. The incidence of analgesic nephropathy varies greatly among countries and even within regions of one country. In general, countries with a higher per capita consumption of phenacetin and other analgesic compound combinations have a higher incidence of analgesic nephropathy.[16,19] Studies have documented analgesic nephropathy as a major cause of end-stage renal disease (ESRD), occurring in more than 15% of ESRD patients in some European countries and Australia. Restrictions on the sale and use of phenacetin or analgesic mixtures have led to a dramatic decrease in the incidence of analgesic nephropathy as a cause of ESRD in some European countries and a major decline in the incidence in Australia. In the United States, reports of the prevalence of analgesic nephropathy vary from more than 10% in the southeast to less than 2% elsewhere in the country.

A well-controlled study of more than 600 middle-aged Swiss working women clearly documented a higher incidence of renal insufficiency and increased mortality from urinary tract disease and cardiovascular disease in a population consuming phenacetin.[20] A retrospective, case-control study in North Carolina also found significantly more renal disease in consumers of analgesic agents than in the control population.[21] The risk of renal disease was increased with daily consumption of phenacetin and acetaminophen but not with daily use of aspirin. This study confirms the risk of renal damage with phenacetin (which has since been removed from analgesic preparations in the United States) but also suggests that acetaminophen, a major metabolite of phenacetin, is nephrotoxic as well. Other recent studies have confirmed the association between daily acetaminophen con-

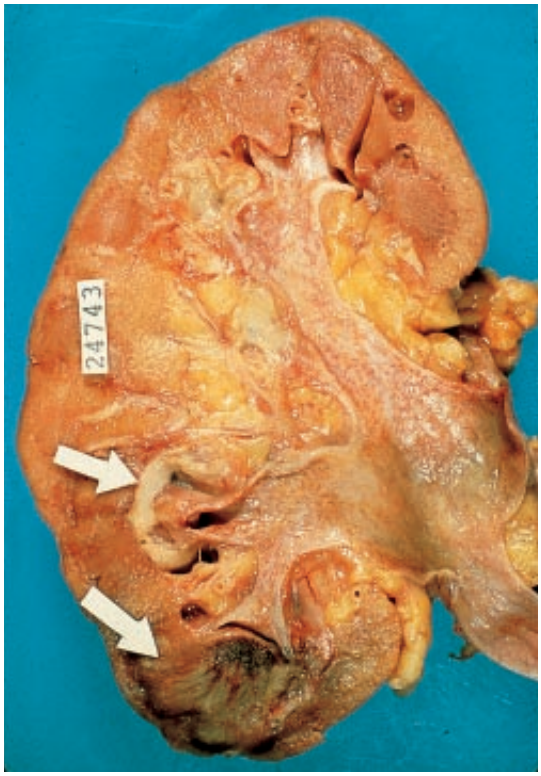


Figure 4 **Areas of renal papillary necrosis (arrows) that developed secondary to analgesic nephropathy and chronic tubulointerstitial nephritis can be observed in this cross section of kidney.**

sumption and the risk of renal failure, although evidence of renal damage from acetaminophen as part of analgesic combinations is much more convincing.[15,17-19,22] One case-control study of almost 1,000 Swedish patients with new-onset renal dysfunction documents the use of acetaminophen as a risk factor for chronic renal failure and provides evidence of increased risk associated with increased dosage.[16]

Pathogenesis

Acetaminophen can accumulate in the renal medulla, where it may be oxidized into reactive intermediates that cause cell necrosis, perhaps through lipid peroxidation. This is especially true when reduced glutathione has been depleted. By inhibiting prostaglandin synthesis and decreasing renal blood flow, some analgesic agents (e.g., aspirin) can further decrease the oxygen supply to the medulla. This decrease, along with salicylate's ability to lower concentrations of reduced glutathione, may account for the potential increased toxicity of analgesic mixtures that contain a combination of phenacetin and aspirin or their metabolites. The findings in the kidneys of patients with analgesic nephropathy reflect this predominant medullary damage. Prominent thickening of the capillaries of the vasa recta is associated with papillary tubular necrosis. Similar capillary changes are found in the renal pelvis and ureters. In later stages, papillary necrosis with associated interstitial fibrosis and chronic inflammatory cell infiltrates are noted.

Most epidemiologic and clinical data suggest that aspirin alone does not cause nephrotoxicity, as is seen with other analgesics. A position paper from an expert panel convened by the National Kidney Foundation concluded that aspirin alone in routine doses in normal persons does not impair renal function, that the habitual consumption of acetaminophen alone should be discouraged, and that the habitual use of analgesic combinations is associated with an increased prevalence of kidney injury and chronic renal failure.[23]

Diagnosis

The typical patient with chronic analgesic nephropathy is a middle-aged woman (four to seven times more women than men have this disorder) who has chronic headaches or arthritic problems and has consumed large amounts of phenacetin, acetaminophen, or aspirin compounds daily for many years. Ingestion of 1 g of analgesic preparations daily for more than 2 years is considered the minimum dosage and time required to produce clinical analgesic nephropathy. Systemic symptoms, such as malaise, weight loss, emotional and psychiatric disorders, anemia, and peptic ulcer disease, may be related in part to analgesic nephropathy and in part to the characteristics of the population that overuses these medications. Diagnosis can be difficult because most patients do not consider over-the-counter preparations to be medications and do not volunteer information regarding their use. NSAID-induced nephropathy is more easily identified than nephropathy caused by acetaminophen, aspirin, and other over-the-counter analgesic drugs; because NSAIDs are frequently prescribed for arthritis, headache, and backache, information on their use may be more readily volunteered by patients [*see* NSAID-Induced Nephropathy, *below*.]

Renal findings are related to chronic interstitial disease and may include nocturia and polyuria, sterile pyuria, urinary tract infections, acidification defects, a predisposition to volume depletion, renal colic and hematuria, and hypertension. Renal insufficiency may be present in asymptomatic patients and is often progressive if analgesic consumption is continued. In rare cases, patients will have flank pain or gross hematuria associated with sloughed or obstructing papillae. Renal papillary necrosis [*see Figure 4*] on intravenous pyelography (IVP) or computed tomography helps establish the diagnosis of analgesic nephropathy.[24]

Ultrasonography is not as sensitive as IVP or CT in the detection of renal papillary necrosis. On noncontrast CT scan of the kidneys, the characteristic findings of analgesic nephropathy are papillary calcifications, decreased renal size, and bumpy renal contours. Patients with analgesic nephropathy also have an increased incidence (up to 8%) of uroepithelial tumors and may have accelerated atherosclerosis. At a late stage of renal dysfunction in analgesic nephropathy, proteinuria and focal glomerulosclerosis can be found, resulting from secondary hemodynamic changes associated with nephron loss.[25]

Treatment

Management of patients with analgesic nephropathy includes total cessation of analgesic use, adequate hydration, and control of hypertension. These measures improve or stabilize renal function in most patients. Regular surveillance of urine cytology will detect uroepithelial tumors, which may arise after discontinuance of the analgesic agent.

NSAID-INDUCED NEPHROPATHY

A number of NSAIDs, including indomethacin, phenylbutazone, ibuprofen, naproxen, ketoprofen, mefenamic acid, and fenoprofen, have been reported to be associated with the development of CIN and renal papillary necrosis. A prospective study of 69 new cases of analgesic nephropathy with papillary

necrosis confirmed by IVP, ultrasonography, or CT found that 42% of the patients had consumed excessive amounts of NSAIDs and that an additional 13% had consumed NSAIDs with other analgesic agents.[26] In contrast to the typical patients with analgesic nephropathy (see above), most of these patients took the NSAIDs as prescribed for arthritis, headache, and backache; male patients outnumbered female patients; and patients did not have emotional or psychiatric symptoms. However, a number of epidemiologic studies have been unable to incriminate long-term use of NSAIDs in the development of CIN. It is likely that only a small percentage of patients with a history of long-term use of NSAIDs develop significant chronic renal damage, and even these may have prior predisposing factors.

LITHIUM-INDUCED CIN

Lithium salts, which are widely used to treat bipolar and affective disorders, have been associated with a number of renal abnormalities—most prominently, a syndrome of nephrogenic diabetes insipidus and polyuria. Although reductions in the glomerular filtration rate (GFR) and development of CIN have been attributed to lithium use, the relation has not been clearly established. Some studies have found that a small percentage of patients given lithium develop an abnormal plasma creatinine level, but other studies have not confirmed this finding. A composite review of studies of almost 500 patients taking lithium found that only 15% to 17% of the patients had a reduced GFR.[27] In patients with a decreased GFR, the reduction was mild; most patients had a GFR of 60 to 75 ml/min. The renal dysfunction in these patients cannot be attributed solely to lithium use, because some psychiatric patients who are not taking lithium also exhibit a reduced GFR and chronic inflammatory changes on renal biopsy. Although early studies failed to document a positive correlation between duration of lithium treatment and the reduction in GFR, studies in which the average duration of lithium treatment was longest (6.5 to 10 years) have noted a positive correlation.[28,29]

A histopathologic study of the renal findings of 24 patients with chronic lithium toxicity clearly defined the biopsy findings.[29] A chronic tubulointerstitial nephropathy was found in every case, in association with cortical and medullary tubular cysts. Distal tubules and collecting ducts were predominantly affected. Many patients had evidence of secondary glomerular focal sclerosis, presumably caused by a hyperfiltration mechanism. Such changes have been noted in both patients with a normal GFR and patients with a decreased GFR.

Prolonged use of lithium for many years is probably associated with some decline in the GFR and with interstitial damage.[28,29] Most patients present with polyuria, nocturia, polydipsia, and elevated BUN and serum creatinine levels. Although the damage is usually mild to moderate, some studies indicate that patients with serum creatinine levels greater than 2.5 mg/dl typically progress to renal failure even after discontinuing lithium.[27,29] The contribution to renal disease of other psychotropic medications or other factors associated with affective disorders is not known. Discontinuance of lithium in patients with less severe renal impairment often leads to stabilization of kidney function.

CHEMOTHERAPY-INDUCED CIN

Most patients with chemotherapy-induced CIN present with asymptomatic worsening of renal function and elevated BUN and creatinine levels. Cisplatin, an antineoplastic agent that is widely used to treat carcinomas and germ cell tumors, may cause acute renal failure and, less frequently, chronic renal insufficiency.[30,31] Cisplatin therapy results in a decrease in glomerular filtration rate in 20% to 30% of patients receiving the drug, and a long-term decrease in renal function is present in up to 20%. Interstitial fibrosis and chronic inflammatory changes have been noted on renal biopsy in patients with chronic renal damage. Only in rare instances has renal failure been reported with the newer agent carboplatin.[32] Studies have shown that the tubulointerstitial damage caused by cisplatin also causes renal salt wasting (leading to orthostatic hypotension) and magnesium wasting (leading to hypomagnesemia and hypocalcemia).[30,31] Hypertension develops in up to 25% of patients receiving cisplatin-based chemotherapy and may contribute to chronic renal damage.

Ifosfamide, a synthetic analogue of cyclophosphamide that is used in treating sarcomas and germ cell tumors, can produce acute renal dysfunction, tubular defects, and chronic renal insufficiency with tubulointerstitial damage.[33,34] The major metabolite of the drug, chloroacetaldehyde, is nephrotoxic to renal tubular cells. Although acute renal dysfunction is usually reversible, chronic renal insufficiency requiring dialysis has been documented.[33,34]

Multiple agents of the nitrosourea class of antineoplastic drugs can produce dose-related nephrotoxicity and CIN. Carmustine (BCNU), semustine (methyl-CCNU), lomustine (CCNU), and streptozotocin have all been reported to cause chronic renal damage. Renal biopsies from patients who sustained renal damage after taking these drugs have shown severe tubular atrophy, glomerulosclerosis, and interstitial fibrosis with chronic inflammatory infiltrates. In some patients, the chronic tubulointerstitial damage has led to ESRD.

Therapy, as in other forms of drug-induced CIN, involves discontinuance of the offending agent and avoidance of other nephrotoxins.

CIN CAUSED BY CYCLOSPORINE AND TACROLIMUS

Cyclosporine and tacrolimus are potent immunosuppressive agents that are widely used in transplantation. Cyclosporine can cause not only acute renal damage but also chronic tubulointerstitial fibrosis.[35,36] The fibrosis usually occurs after months of treatment and is often associated with drug-related obliterative arteriolopathy and microvascular damage to the renal arterioles. The tubulointerstitial damage may occur in a bandlike pattern in the kidney (so-called striped fibrosis). Secondary scarring of the glomeruli may also be present. This damage has been seen both in transplant populations and in patients without prior renal disease who are taking cyclosporine for autoimmune disorders.[37] In patients with autoimmune diseases, the likelihood that cyclosporine nephrotoxicity will develop increases with older age, a higher daily dose of the drug, and a greater initial rise in the serum creatinine level from baseline when drug therapy is started. The disorder is usually associated with a decreased GFR and progressive renal insufficiency.

Although many therapies have been suggested to ameliorate nephrotoxicity caused by cyclosporine, the only treatment that provides consistent results is a reduction in the dose of the drug, which often leads to some improvement in GFR.[38] Tacrolimus has been shown to have acute and chronic nephrotoxicities similar to those associated with cyclosporine.[39] Reduction in drug dose is the recommended treatment to ameliorate nephrotoxicity.

Interstitial Damage Resulting from Physical Factors

OBSTRUCTIVE NEPHROPATHY

Obstructive nephropathy results from the impaired outflow of urine.[40] Although urinary obstruction may cause acute renal failure, it may also produce chronic interstitial damage, especially when the obstruction is partial or intermittent and long-standing. Obstructive nephropathy from congenital abnormalities is common in children but in persons older than 60 years, benign prostatic hypertrophy and prostatic and gynecologic cancers are more common etiologies. Although obstruction from stones, clots, and strictures usually leads to acute renal failure, such obstructions can also produce chronic tubulointerstitial damage to the kidney. Less common causes of slow chronic obstruction (e.g., ureteropelvic junction obstruction and retroperitoneal fibrosis) are often associated with chronic tubulointerstitital damage and renal insufficiency.[41]

Pathogenesis

The pathophysiology of obstructive nephropathy involves increased intratubular pressure, local ischemia, and, often, associated infection, which combine to produce tubulointerstitial damage.[40] Infiltration by inflammatory T cells and macrophages, an autoimmune response to refluxed urinary Tamm-Horsfall mucoprotein (a protein produced by cells of the ascending limb of the loop of Henle), and vasoactive hormones may all play a role in renal damage. The pathologic findings consist of dilatation of the collecting ducts and distal tubules, along with chronic tubular atrophy with little glomerular damage.

Diagnosis

Most patients with tubulointerstitial damage caused by obstruction present with symptoms related to the decreased GFR or to tubular defects, such as loss of urinary concentrating ability (polyuria and nocturia), hyperkalemia associated with a type 4 renal tubular acidosis, or salt wasting that predisposes to volume depletion. Pain is not a prominent symptom of slowly developing obstructive nephropathy but, rather, is a feature of acute distention of the bladder, collecting system, or renal capsule. The diagnosis of obstructive nephropathy is typically based on evidence of renal dysfunction and blockage of urine outflow in the absence of other causes of renal damage. Ultrasonography usually shows dilatation of the urinary collecting system and hydronephrosis. In some patients, CT scanning or IVP will be helpful to delineate the location and nature of the obstruction.

Treatment

Treatment of obstructive nephropathy consists of relieving the obstruction. Every effort should be made to rapidly diagnose and correct any obstruction; data from animal models and human studies have demonstrated return of considerable renal function, even after prolonged partial obstruction of a single kidney.

REFLUX NEPHROPATHY

The abnormal backflow of urine from the bladder into the upper urinary tract (i.e., vesicoureteral reflux) can cause chronic tubulointerstitial damage, known as reflux nephropathy.[42-44] Various anatomic factors appear to play a role in incompetence at the vesicoureteral junction. A number of these factors have been defined: a shortened ureteral intravesicular submucosal tunnel, abnormal amounts or types of muscle fibers around the intravesicular ureter, an abnormality of the trigone, and chronic increased intravesicular pressure without other factors. In each case, the incompetent vesicoureteral junction allows urine to reflux into the ureters and collecting system of the kidneys. The reflux may be unilateral or bilateral, and the severity is variable. When infection is present, even low-pressure refluxed urine that reaches the kidney can produce chronic interstitial inflammation and scarring. However, it is unclear whether repeated episodes of high-pressure sterile reflux alone can damage the kidney. An autoimmune inflammatory response to refluxed Tamm-Horsfall mucoprotein has been proposed as one mechanism of renal damage from sterile reflux.

Reflux in young children causes irregular scars at the poles of the kidney, where compound papillae allow the higher pressures in the urinary collecting system to be directly transmitted to the renal parenchyma.[42,44] The areas of scarring are associated with tubulointerstitial fibrosis overlying deformed calyxes of the collecting system. Glomeruli show only minor changes until there has been considerable parenchymal damage. At this point, the glomeruli exhibit focal segmental glomerulosclerosis from hyperfiltration and hemodynamic changes. There is often a poor relation between the severity of a patient's symptoms, the degree of reflux, and the number of infections the patient has experienced.

Diagnosis

Presenting features of reflux nephropathy can include signs or symptoms of urinary tract infection such as back or flank pain, fever, and dysuria. Hypertension, when present, is often associated with high levels of renin, which may derive from the segmental areas of scarred parenchyma. Patients often have a urinary concentrating defect, leading to nocturia and polyuria. Proteinuria is mild unless the patient has developed advanced renal insufficiency, with glomerular hyperfiltration, and focal segmental glomerulosclerosis.[42-44]

Diagnosis of reflux nephropathy is based on evidence of the reflux of urine and on characteristic parenchymal and clinical findings. Reflux is demonstrated with a voiding cystourethrogram, either with contrast agents or by radionuclide scanning. Because urinary tract infection may be associated with reflux, it is best to wait several weeks after treating a urinary tract infection before trying to diagnose reflux nephropathy.

Treatment

The treatment of low-grade reflux (i.e., reflux without major dilatation of the collecting system) is medical: long-term antibiotic therapy is used to sterilize the urine and prevent reinfection. Many persons with such mild reflux undergo spontaneous remission with time.[45] More severe reflux (i.e., reflux with major dilatation of the collecting system) may require surgical intervention, although most comparative studies have not found an advantage with surgical intervention over medical therapy.[46] In general, patients with proteinuria (> 1.5 g/day), a very depressed GFR, and large amounts of renal scarring show little response to such interventions and go on to develop progressive renal failure.

RADIATION NEPHRITIS

Radiation-induced renal damage is a rare form of tubulointerstitial damage that occurs when the kidneys are exposed to high doses of radiation over a short period. The incidence of ra-

diation nephritis has greatly declined with the ability to precisely deliver high-dose radiation to a given abdominal port and with better shielding of the kidneys from radiation exposure. Current cases usually occur in patients who received high-dose, diffuse abdominal radiation, often in combination with radiomimetic chemotherapeutic agents.[47] The pathogenesis of radiation nephritis remains unknown.

The clinical picture produced by radiation damage has been divided into a number of distinct syndromes, including acute radiation nephritis; chronic radiation nephritis; and radiation-related hypertension (mild or severe), proteinuria, or both. There is often no clear clinical or pathologic distinction between these entities, however. Acute radiation nephritis occurs 6 to 12 months after exposure to radiation and is characterized by severe hypertension, salt retention and edema, mild hematuria and proteinuria (usually < 2 g/day), and renal dysfunction (i.e., reduced GFR). Anemia may be out of proportion to the degree of depression of the GFR. Chronic radiation damage occurs in patients who have survived acute radiation damage or in those who received smaller radiation doses (500 to 1,000 cGy) over longer periods. Clinical manifestations may appear years after radiation exposure, with the insidious onset of hypertension, renal dysfunction, a urinary concentrating defect, and mild to moderate proteinuria (< 2 g/day). Some patients manifest mild or severe hypertension or isolated proteinuria after radiation exposure. In some cases, hypertension has been documented to be caused by renin oversecretion by the shrunken, irradiated kidney; this condition has responded to uninephrectomy.

On histopathologic examination, kidneys affected by radiation nephritis typically show damage to all structural components. There may be various degrees of glomerular proliferation or sclerosis, tubular degeneration and atrophy, fibrinoid necrosis and intimal proliferation of the small arteries and arterioles, and interstitial edema and fibrosis. The only treatment for radiation nephritis entails control of hypertension. Preventing the development of radiation nephritis by appropriate delivery of radiation, with shielding of the kidneys, is crucial.

Infectious Tubulointerstitial Nephritis

Infectious agents can produce acute and chronic tubulointerstitial damage. Some organisms directly invade the normal kidney, leading to acute bacterial pyelonephritis. Other organisms invade the kidney when obstruction or urinary reflux is present, contributing to tubulointerstitial damage in the form of chronic pyelonephritis. Some systemic infections lead to an acute interstitial reaction without direct invasion of the renal parenchyma. The diagnosis and treatment of urinary tract infections are covered elsewhere [*see 7:XXIII Infections of the Urinary Tract*], but the relation between these infections and structural renal damage is important to understand.

Acute Bacterial Pyelonephritis

Acute bacterial pyelonephritis accounts for over 250,000 hospital admissions in the United States annually. It is characterized by fever, chills, sweating, and flank, back, or abdominal pain.[48] It may be associated with dysuria, frequency, urgency, and other symptoms related to lower urinary tract infection. Pyuria, hematuria, and leukocyte casts in the urine sediment are often present. Urine cultures are usually positive, and blood cultures are positive in 10% to 20% of cases. The histopathology reveals patchy areas of edema and acute inflammatory cell infiltrates (polymorphonuclear cells, lymphocytes, and other cell types) that obliterate the normal tubulointerstitial architecture. There is tubular atrophy and necrosis in these areas; tubulointerstitial areas remote from the infection may appear normal. Most cases of acute bacterial pyelonephritis that occur in previously normal kidneys resolve after antibiotic therapy without residual damage. In rare instances, chronic pyelonephritic changes result from repeated or unresolved cases of acute infection in unobstructed and nonrefluxing kidneys.

Chronic Pyelonephritis

Chronic pyelonephritis, a term misused in the past to describe kidneys with tubulointerstitial damage of many causes, properly refers to the chronic damage caused by infection of the pyelocalyceal system and renal parenchyma.[49] Chronic pyelonephritis often presents insidiously with signs of chronic infection or renal insufficiency. Thus, some patients may experience dysuria, fever, and vague flank or back pain, whereas others may exhibit hypertension, salt wasting, a urinary concentrating defect, and hyperkalemia. The urinary sediment usually contains leukocytes, leukocyte casts, and broad casts, which denote chronicity of the tubulointerstitial process. Chronic pyelonephritis is most often the result of chronic or repeated bacterial infection complicating vesicoureteral reflux, but it may also be the result of infection complicating preexisting obstruction or structural renal problems. The kidneys are often contracted, with asymmetrical cortical scars corresponding to deformity of the underlying calyceal system. The histopathology is characterized by tubular atrophy and interstitial fibrosis, with a chronic inflammatory infiltrate of lymphocytes and mononuclear cells.

Xanthogranulomatous Pyelonephritis

Xanthogranulomatous pyelonephritis refers to an unusual form of localized chronic bacterial renal infection. Characterized by inflammation and granuloma formation with lipid-laden macrophages, xanthogranulomatous pyelonephritis is typically found in obstructed kidneys containing stones. Symptoms may include those of acute bacterial pyelonephritis (see above), as well as hypertension; in many patients, a palpable mass is present.[48] The kidney, which may be nonfunctional, contains a localized abscessed area that may be mistaken for a tumor on intravenous urography. CT and magnetic resonance imaging may help clarify the diagnosis. Most patients have a positive urine culture; often, more than one organism is present. The pathogenesis of xanthogranulomatous pyelonephritis is unknown. Treatment consists of antibiotic therapy and surgical resection of the abscessed area.

Interstitial Inflammation Secondary to Systemic Infection

Infections remote from the kidney may be associated with an acute interstitial inflammatory reaction.[6] This adverse renal reaction has been associated with bacterial, viral, and parasitic infections, including streptococcal infections, diphtheria, leptospirosis, leishmaniasis, toxoplasmosis, infectious mononucleosis, Epstein-Barr virus infection, measles, brucellosis, syphilis, mycoplasmal pneumonia, HIV infections, Kawasaki disease, and Legionnaires disease. Some infectious agents, such as those responsible for leptospirosis, mycobacterial infections, cytomegalovirus infection, Rocky Mountain spotted fever, and hantavirus infection, may directly infect the kidney and cause interstitial inflammation during the course of systemic infection.

The acute inflammatory infiltrate can be composed of lymphocytes, plasma cells, monocytes, and polymorphonuclear cells, but eosinophils are not commonly found. The acute renal dysfunction is manifested by an acute decrease in the GFR or isolated tubular defects. All manifestations generally resolve if the systemic infection is successfully treated.

VIRAL-RELATED TUBULOINTERSTITIAL NEPHRITIS IN IMMUNOSUPPRESSED PATIENTS

Although viral infections can cause tubulointerstitial disease by direct invasion of the kidney or remote effects, certain forms of viral-related tubulointerstitial nephritis warrant special comment. Certain polyomaviruses (e.g., BK virus and JC virus) can cause tubulointerstitial disease in immunosuppressed patients with organ transplants or HIV infection.[50-52] Tubulointerstitial nephritis may mimic renal transplant rejection, and the increased immunosuppression used for rejection may worsen the renal dysfunction. Also, although HIV infection is classically associated with a glomerular lesion of a collapsing form of focal sclerosis, tubulointerstitial disease may be prominent in this entity.[53] Microcystic dilatation of the tubules is responsible for the large echogenic kidneys seen in HIV-associated nephropathy.

Metabolic and Toxic Tubulointerstitial Nephritis

A number of metabolic disturbances, including disorders affecting the metabolism of oxalate, calcium, uric acid, and potassium, can produce renal tubulointerstitial abnormalities. Exposure to certain heavy metals, such as lead and cadmium, can also produce chronic renal damage.

HYPEROXALURIA

Oxalic acid is a dicarboxylic end product of metabolism that is removed from the body only by renal excretion. Normal urinary excretion is between 15 and 50 mg daily. Oxalate is highly insoluble when combined with calcium, and urinary excretion of oxalate in amounts greater than normal readily exceeds the solubility product for calcium oxalate. Precipitation of calcium oxalate can produce nephrolithiasis, acute renal failure, or chronic tubulointerstitial damage. The histopathology of chronic tubulointerstitial damage from oxalate deposition shows calcium oxalate crystals (which are birefringent under polarizing light) in the tubular lumina and the interstitium, surrounded by inflammation (which may include giant cells) and interstitial fibrosis.

Excessive oxalate excretion can result from two genetic disorders: primary hyperoxaluria 1 and 2 (PHO1 and PHO2).[54-56] Both are caused by autosomal recessive enzymatic defects in the metabolic pathway of glyoxylic acid. In each disorder, the excretion of oxalate typically ranges from 100 mg to more than 250 mg daily. PHO1 is the more common disorder; a defective gene on chromosome 2 causes a deficiency of the hepatic enzyme alanine-glyoxylate aminotransferase that leads to excessive excretion of glycolate, glyoxylate, and oxalate. Calcium oxalate precipitates in the kidney, cardiac system, joints, peripheral blood vessels, and other organs. Patients may have hematuria, renal colic from oxalate stones, livedo reticularis, urinary tract infections and pyuria, hypertension, and renal tubular acidosis. Most patients develop ESRD by 20 years of age. PHO2 is much rarer; it is caused by a deficiency of D-glycerate dehydrogenase and glyoxalate reductase that leads to excessive excretion of oxalate and L-glycerate. It also leads to nephrolithiasis and renal insufficiency. The treatment of choice for many patients is combined liver-kidney transplantation.[57,58]

Metabolism of exogenous substances into oxalate or excessive absorption of oxalate from the gut may also cause oxalate-induced damage.[59] Ethylene glycol ingestion and methoxyflurane anesthesia, although usually associated with acute renal failure, may cause residual tubulointerstitial damage in patients who recover from the acute event. Patients who ingest large amounts of ascorbic acid may metabolize it to glyoxylate and oxalate, causing both stones and renal damage. Patients with steatorrhea resulting from various intestinal diseases—including celiac disease, Crohn disease, Wilson disease, and chronic pancreatitis—or from small bowel resection or bypass operations for obesity may hyperabsorb oxalate from the large bowel. The pathogenesis of oxalate hyperabsorption involves abnormal binding of intraluminal gut calcium to fats, freeing more oxalate for absorption. In addition, the solubilizing effect of bile acids on the large bowel permits greater absorption of oxalate. The result can be nephrolithiasis, acute renal insufficiency, or chronic tubulointerstitial damage.

The treatment of chronic renal damage from oxalate deposition is correction of the primary cause when possible (e.g., stopping excessive ascorbic acid ingestion or preventing steatorrhea in a gastrointestinal disorder). Other measures include following a low-oxalate diet, increasing intake of fluids to increase urinary volume, increasing oral calcium intake to bind gut oxalate, and taking pyridoxine supplements.

HYPERURICEMIA AND HYPERURICOSURIA

Hyperuricemia and hyperuricosuria can lead to uric acid nephrolithiasis and acute oliguric renal failure from urate deposition. These conditions occur after massive release of uric acid, as in tumor lysis syndrome, or chronic renal failure. Chronic urate nephropathy is characterized by urate crystal deposition in the interstitium accompanied by tubular atrophy, chronic inflammatory infiltrates, and fibrosis. It is unclear whether patients with long-standing gout or chronic asymptomatic hyperuricemia have a higher incidence of chronic tubulointerstitial damage from excretion of urate. It was thought that patients with gout or asymptomatic hyperuricemia who experienced renal dysfunction had gouty nephropathy, characterized by urate crystal deposition in the interstitium accompanied by tubular atrophy. Studies suggest, however, that much of the renal dysfunction seen in patients with gout or chronic hyperuricemia can be explained by concurrent hypertension, ischemia, diabetes, and other causes of renal damage. Some patients are shown to have excessive lead concentrations when tested by ethylenediaminetetraacetic acid (EDTA) loading. It has been suggested that these patients actually have chronic lead nephropathy (see below) rather than gouty nephropathy. Moreover, many tubulointerstitial diseases are associated with decreased urate excretion and thus with a higher incidence of hyperuricemia and gout. However, recent evidence supports a primary role for uric acid in renal damage.[60] Chronic hyperuricemia and gout are only two of a number of contributing factors to produce chronic renal damage; thus, therapy should focus on all aspects of this multifactorial disorder and not just on lowering the uric acid level.

HYPERCALCEMIA AND HYPERCALCIURIA

Hypercalcemia and hypercalciuria can also produce a number of adverse renal effects. Hypercalcemia may contribute to hypertension by causing vasoconstriction, may cause nephroli-

thiasis and vasopressin-resistant urinary concentrating defects, and may lead to chronic tubulointerstitial damage. In primary hyperparathyroidism, renal insufficiency is related to the degree and duration of the hypercalcemia. Chronic elevations of the calcium level can lead to calcium salt deposition in the tubules and interstitial regions, associated with chronic interstitial inflammation, tubular atrophy, and fibrosis. Some patients exhibit only a reduced GFR, whereas others exhibit polyuria and nocturia from the concentrating defect, salt wasting, loss of urinary magnesium or potassium, and renal tubular acidosis or hypertension.

CHRONIC HYPOKALEMIA

Persistently low levels of serum potassium, as seen in patients who abuse laxatives or diuretics or who have chronic hyperaldosteronism, have been associated with structural and physiologic changes in the kidney. In animal models, chronic hypokalemia can lead to tubulointerstitial damage. In humans, chronic hypokalemia is associated with prominent vacuolation of the proximal tubular cells caused by dilated intercellular spaces, basal infoldings, or dilated cisternae of the endoplasmic reticulum. Chronic hypokalemia is associated with impaired urinary concentration, impaired ability to conserve salt on a low-salt diet, and hypertension from impaired ability to excrete a salt load. It is unclear how often uncomplicated chronic hypokalemia leads to chronic tubulointerstitial damage. Interstitial inflammatory changes and fibrosis have been noted on kidney biopsies of patients with chronic hypokalemia, and renal medullary cysts may form. The mechanisms involved in cyst formation are unknown but may relate to excess ammoniagenesis or growth factors and cytokine production.[61,62]

NEPHROTOXICITY CAUSED BY LEAD OR CADMIUM EXPOSURE

Two heavy metals, lead and cadmium, clearly produce tubulointerstitial damage. A number of other metals and elements that cause acute tubular damage and acute renal failure (e.g., mercury or arsenic) may be associated with residual tubulointerstitial damage after the patient recovers from acute exposure.

Lead-Induced Nephrotoxicity

Etiology and pathogenesis Exposure to lead can occur from ingestion of lead-based paints, storage of alcoholic beverages in crystal decanters made with lead, production of moonshine whiskey in a lead-containing still, the manufacture or destruction of lead batteries, or ingestion of lead-containing aerosols in the workplace. Lead accumulates in the proximal tubules of the kidney and can lead to cell damage and tubular defects, such as Fanconi syndrome, aminoaciduria, and renal glycosuria. Observational studies suggest that chronic low-level exposure over years may result in a decreased GFR.[63,64] Long-term lead exposure can result in chronic tubulointerstitial nephritis and renal insufficiency, hypertension, and saturnine gout from reduced renal excretion of uric acid. Epidemiologic studies confirm an inverse relation between blood lead levels and renal function.[63,65] Long-term exposure induces progressive tubular atrophy, fibrosis, and chronic renal failure.

Diagnosis Although acute lead toxicity is associated with abdominal pain, anemia, and encephalopathy, these findings are not usually present in lead nephropathy, which presents as hypertension and insidious renal dysfunction. The histopathology of lead nephropathy resembles that of CIN. The diagnosis of lead-induced nephrotoxicity cannot be made from serum levels, but it can be made by demonstrating excessive urinary excretion of chelated lead after an intravenous or intramuscular injection of EDTA.[66] Because hypertension and hyperuricemia are common in patients with this condition, it is easy to confuse the diagnosis of lead nephropathy with hypertensive nephrosclerosis or chronic urate nephropathy.

Treatment Chelation therapy with EDTA or oral succimer (2,3-dimercaptosuccinic acid) may slow the progression of renal insufficiency in chronic lead tubulointerstitial damage.[66]

Cadmium-Induced Nephrotoxicity

Cadmium, used in metallurgy, can produce chronic tubulointerstitial damage after prolonged exposure.[67,68] Some patients have tubular defects such as Fanconi syndrome, glycosuria, aminoaciduria, and tubular proteinuria with β_2-microglobulin excretion. Other patients have a decreased GFR. The diagnosis is based on evidence of excessive urinary cadmium levels. There is no specific treatment for cadmium nephrotoxicity.

CHINESE-HERB NEPHROTOXICITY

Certain Chinese herbs, many used in weight-reduction programs, have been associated with tubulointerstitial nephritis and progression to renal failure.[69,70] Over 100 cases have been reported, and many have progressed to dialysis and transplantation.[70] Additional agents, such as appetite suppressants (e.g., fenfluramine and diethylpropion) with vasoconstrictive properties, were ingested by some of the affected patients and may have contributed to the renal interstitial damage.

The pathogenesis, although incompletely understood, may relate to a plant nephrotoxin aristolochic acid, which can produce similar interstitial fibrosis in animal models.[71]

Patients with Chinese-herb nephrotoxicity present with renal insufficiency, mild hypertension, mild proteinuria (< 1.5 g/day), and bland urinary sediment. On biopsy there is extensive cortical interstitial fibrosis with atrophy and tubular dropout and only a sparse interstitial inflammatory infiltrate. There is a high incidence of uroepithelial carcinomas of the genitourinary tract in patients with Chinese-herb nephropathy.[72]

There is no proven therapy for this form of tubulointerstitial nephropathy, but an uncontrolled report has shown that corticosteroids may slow the progression.[70]

Interstitial Disease Associated with Vascular Damage

HYPERTENSION

Hypertensive nephrosclerosis is first associated with changes predominantly in the renal interlobular and afferent arterioles. Subsequent ischemia can lead to glomerular lesions, as well as to tubular atrophy and interstitial damage characterized by patchy, chronic inflammatory changes and fibrosis in the renal cortex. The renal medulla is often spared, in contrast to what occurs in chronic pyelonephritis. Malignant or accelerated hypertension is commonly associated with such interstitial changes. Hypertension is the primary clinical finding; laboratory tests usually reveal azotemia and proteinuria that is rarely in the nephrotic range.

SICKLE CELL ANEMIA

Patients with sickle cell anemia and other hemoglobinopathies can demonstrate various physiologic, morphologic, and clinical renal manifestations.[73,74] Almost 8% of African Americans

carry the sickle cell trait, and 0.2% have sickle cell disease. Other African Americans have sickle cell–hemoglobin C disease or other hemoglobinopathies associated with altered red blood cell deformity.

Children and teenagers with sickle cell disease may present with some clinical renal abnormalities, but the spectrum of tubulointerstitial abnormalities is typically revealed in adults. Sickled erythrocytes impair blood flow in the renal microcirculation, as in other organs, leading to local ischemia and infarction. Ischemia is prominent in the hypertonic, acidotic, relatively hypoxemic environment of the vasa recta of the renal medulla. Medullary renal infarction leads to areas of tubular necrosis and interstitial fibrosis and, eventually, papillary necrosis. Patients with sickle cell disease exhibit impaired urinary concentrating ability, leading to polyuria and nocturia. Fluid and volume loss predispose to acute declines in the GFR when intercurrent illness prevents adequate fluid intake to match urinary losses. Microscopic and gross hematuria are common, resulting from microinfarcts in the renal medulla. A distal or type 4 renal tubular acidosis, with hyperchloremic metabolic acidosis and impaired potassium excretion, is often found in adults with sickle cell disease. Papillary necrosis occurs in most adults with this disease but only rarely leads to obstructive uropathy from a sloughed papilla or to major declines in the GFR. In rare instances, patients with sickle cell disease develop ESRD, but such patients usually have an associated glomerular lesion or large-vessel renal infarcts. Sickle cell nephropathy has been associated with the rare renal neoplasm of renal medullary carcinoma.[75]

The treatment of sickle cell nephropathy entails correction of volume depletion and electrolyte imbalances.[73-75] Episodes of gross hematuria should be investigated to rule out associated urinary tract pathology, but most patients require only symptomatic treatment, with hydration and transfusion to correct blood loss and raise the concentration of nonsickling erythrocytes. Nephrectomy should be avoided because the hematuria is usually self-limited and because patients may develop gross hematuria from the other kidney.

CHOLESTEROL EMBOLI

Cholesterol embolization may cause both acute and chronic renal insufficiency and is associated with damage to both glomeruli and the tubulointerstitial region of the kidney.[76-78] Atheromatous renal emboli containing cholesterol crystals may occur spontaneously in older patients with severe atherosclerotic plaques of the aorta and major vessels. Cholesterol emboli may also occur in patients undergoing angiographic or vascular surgical procedures. The number of renal cholesterol emboli found at autopsy parallels the degree of aortic atherosclerosis of the person.

Symptoms and signs of embolization include fever, myalgias, lower-extremity livedo reticularis, petechial skin lesions, digital ischemia, and amaurosis fugax and other fleeting neurologic deficits. Patients may have bland urinary sediment or microhematuria, leukocytosis, transient eosinophilia or eosinophiluria, hypocomplementemia, thrombocytopenia, and an elevated erythrocyte sedimentation rate. Most patients present with progressive renal dysfunction over months to years. When cholesterol embolization is associated with a surgical or angiographic procedure, patients typically develop hypertension and progressive renal insufficiency, with rising BUN and serum creatinine levels, several weeks after the procedure.

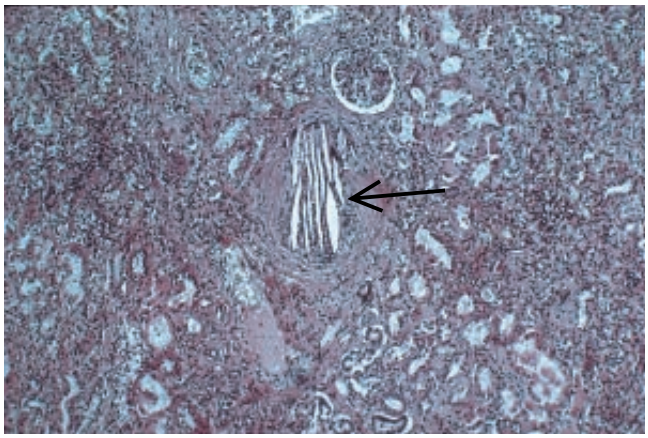


Figure 5 **Cholesterol emboli (arrow) can be seen occluding a medium-sized vessel; an intense interstitial inflammatory reaction is also apparent in the surrounding interstitial area.**

Cholesterol emboli can produce a spectrum of renal function impairments, ranging from mild, stable renal insufficiency to end-stage renal failure. The characteristic histopathologic finding is occlusion of the lumina of the small arteries and arterioles by atheromatous debris. Cholesterol crystals dissolve in the tissue-fixation process, leaving telltale biconcave, needle-shaped clefts in the center of the occluded vessels [*see Figure 5*]. Interstitial fibrosis may result from combined ischemia and inflammatory damage. The diagnosis of renal cholesterol embolization is suggested by the clinical features and the presence of retinal vessel involvement documented by ophthalmoscopy. Cholesterol embolization is confirmed histopathologically by the finding of biconcave clefts in small vessels in biopsy samples of muscle, skin, kidney, or other organs.

There is no effective therapy for renal cholesterol emboli. Management entails controlling hypertension; discontinuing anticoagulation, which may worsen the renal dysfunction; and providing supportive care.[79] In some patients who receive supportive care, gradual abatement of symptoms and even of severe renal insufficiency may occur. The progressive and long-term survival of patients with cholesterol emboli has greatly improved.

RENAL VEIN THROMBOSIS

Manifestations of renal vein thrombosis can range from asymptomatic occurrence in a nephrotic adult to acute oliguric renal failure in a volume-depleted child. Most patients with renal vein thrombosis have underlying nephrosis, renal tumors that invade the renal veins, or clotting abnormalities that predispose to coagulation (e.g., anticardiolipin syndrome or protein C or S deficiency).[80,81] Clinical manifestations and physiologic effects depend on the rapidity and completeness of renal vein occlusion. Slowly progressing renal vein thrombosis permits development of adequate collateral circulation, and despite the potential for pulmonary emboli, few renal manifestations occur. Rapid development of near-total venous occlusion can lead to flank pain, gross hematuria, a decline in the GFR, increased proteinuria, and fluid retention in the nephrotic patient. The histopathology of acute renal vein thrombosis includes interstitial edema with sparse cellular infiltrates; in contrast, chronic renal vein thrombosis is associated with tubular atrophy and interstitial fibrosis.

Tubulointerstitial Disease Caused by Dysproteinemias and Other Tumors

Multiple myeloma and other dysproteinemias, characterized by the excessive production of monoclonal immunoglobulins or their fragments, may cause both glomerular and tubulointerstitial injury to the kidney.[82-84] Whereas the glomerulopathies are predominantly associated with the glomerular precipitation of paraproteins, the tubulointerstitial damage can be caused either by a similar mechanism or by a number of metabolic and complicating features, such as hypercalciuria and hyperuricemia.[82,83]

MULTIPLE MYELOMA

Renal insufficiency and acute renal failure are common and significant contributors to morbidity and mortality in multiple myeloma.[82,83,85] Most patients present with isolated proteinuria as the first renal manifestation of dysproteinemia. Clinical indications that a patient with unexplained renal disease has multiple myeloma include (1) low-level proteinuria by urinary dipstick measurement but high-level proteinuria by 24-hour quantitative measurement (the dipstick primarily detects albumin, not Bence Jones protein), (2) a low anion gap (caused by the cationic charge on some monoclonal immunoglobulins), (3) hypercalcemia in the presence of renal failure and a high serum phosphate level, and (4) anemia out of proportion to the degree of renal insufficiency. A common manifestation of myeloma is renal insufficiency, which is present in more than 50% of patients. Patients with renal insufficiency have a shorter survival than those with normal renal function. Renal insufficiency may be indolent, chronic and progressive, or rapidly progressive. The degree of insufficiency is usually closely correlated with the excretion of light-chain proteins. A plasma creatinine level greater than 2 mg/dl is found in about 30% of patients at presentation and in 50% during the course of the disorder. A smaller percentage present with acute renal failure.

Excessive production and filtration of monoclonal light chains (Bence Jones protein) can cause direct tubular cell damage, as well as tubular obstruction by casts. When human Bence Jones proteins are injected into animal models, only some of the proteins coaggregate with Tamm-Horsfall mucoproteins (normally secreted by the tubules) to form obstructing casts in the distal tubules. In classic myeloma kidney, there are a large number of refractile, fractured-appearing casts, especially in the lumina of the distal tubules. These casts cause tubular cell damage, with peritubular inflammatory infiltrates of neutrophils and mononuclear cells; giant cells surround some casts [*see Figure 6*]. The casts may ultimately be found free in the interstitium, surrounded only by fibrous tissue and a few chronic inflammatory cells. Dysproteinemias can also be associated with tubulointerstitial precipitation of urate crystals caused by urate overproduction or lysis of plasma cells. In addition, the tubulointerstitial area can be damaged by deposition of calcium salt crystals as a result of hypercalciuria from lytic bone lesions and calcium-mobilizing humoral factors. Upper and lower urinary tract infections, resulting from suppression of normal humoral immunity, are also common in patients with myeloma and can cause acute and chronic tubulointerstitial damage. In rare instances, abnormal plasma cells infiltrate the interstitial area of the kidney, leading to renal dysfunction. Finally, many therapeutic agents cause renal toxicity (e.g., aminoglycosides and other antibiotics and contrast agents), especially when they are used in patients with myeloma who are dehydrated and volume depleted. Therapy for myeloma-induced renal damage consists of adequate hydration, treatment of hypercalcemia and other metabolic abnormalities, and methods to decrease the production and renal excretion of abnormal paraproteins. These include chemotherapy with melphalan or multidrug regimens (e.g., vincristine, doxorubicin, and dexamethasone), plasmapheresis for acute renal failure, and stem cell transplantation for selected patients.

AMYLOIDOSIS

Amyloidosis, which is caused by the extracellular deposition of fibrillary proteins in a β-pleated sheet, typically leads to glomerulopathy with heavy albuminuria. Tubular and interstitial deposition may also occur, however, leading to tubulointerstitial damage and isolated tubular defects. Proximal and distal renal tubular acidosis and vasopressin-resistant nephrogenic diabetes insipidus have been reported in patients with amyloidosis. In amyloidosis associated with the precipitation of serum amyloid A protein, especially in chronic heroin addicts, there is often a prominent tubulointerstitial inflammatory reaction.[86]

LIGHT-CHAIN DEPOSITION DISEASE

Tubulointerstitial inflammatory infiltrates and fibrosis can occur in light-chain deposition disease, in which light chains that do not have the configuration or staining properties of amyloid are deposited in the kidney and in other organs and tissues.[84] In light-chain deposition disease, tubular basement membrane deposits of light chains are found even more commonly than glomerular basement membrane deposits. Almost 50% of patients with light-chain deposition disease have been shown to have associated cast nephropathy (akin to patients with myeloma cast nephropathy) in addition to their glomerular findings.[87]

DYSPROTEINEMIAS AND OTHER TUMORS

In Waldenström macroglobulinemia, IgM is typically deposited in the glomeruli, but the disease is also associated with the presence of aggregates of atypical lymphocytes in the interstitial regions of the renal cortex and medulla.

Other tumors may cause tubulointerstitial damage through metabolic disturbances secondary to hyperuricemia and hyperuricosuria, hypokalemia, and hypercalcemia and hypercalciuria. The metabolic disturbances may be caused by the elaboration of hormonal substances or other tumor products or by chemotherapy. Direct tumor invasion of the kidney, which has

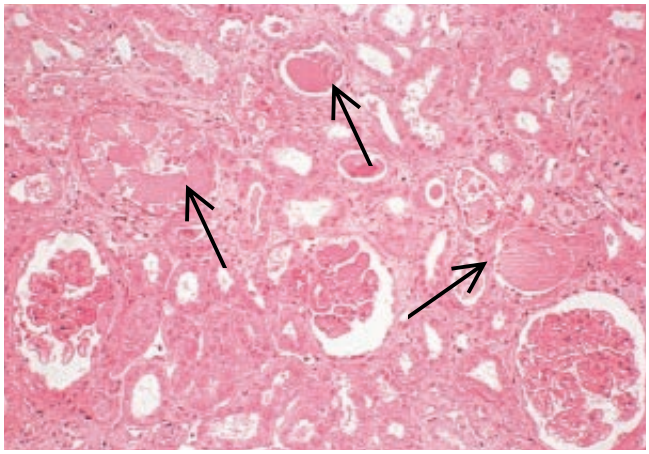


Figure 6 **Bence Jones cast nephropathy in multiple myeloma is associated with glassy, fracturing casts occluding the tubular lumina (arrows). Multinucleated giant cells surround some casts.**

been noted at autopsy, occurs often with leukemias and less often with lymphomas. However, only a few patients with leukemia or lymphoma have large kidneys or renal insufficiency.

Cystic Disease of the Kidney

As knowledge of the genetic background of cystic kidney disease has increased, the clinical spectrum and terminology have become clearer. Common forms of renal cystic involvement include simple cysts, autosomal dominant polycystic kidney disease (ADPKD), autosomal recessive polycystic kidney disease, acquired cystic disease, nephronophthisis–medullary cystic disease, and medullary sponge kidney. Each has typical clinical and histopathologic features.

SIMPLE RENAL CYSTS

Simple renal cysts, which are also known as solitary or single cysts, are fluid-filled cavities lined by epithelium.[88] These lesions are found with ultrasonogaphy or CT in approximately 25% to 30% of patients older than 50 years and in 50% of this age group at autopsy. Almost all of these cysts are asymptomatic. In rare instances, hematuria, pain, hypertension, or infection can complicate a simple renal cyst.

POLYCYSTIC KIDNEY DISEASE

ADPKD has a high prevalence in whites in the United States, occurring in one in 400 to one in 1,000, and is transmitted to 50% of the offspring of affected persons.[89] ADPKD is an inherited systemic disorder that commonly leads to progressive renal failure but that can also affect a number of other organ systems. Almost 90% of patients with ADPKD have an abnormal gene on the short arm of chromosome 16 (the ADPKD1 locus).[90] This gene encodes for polycystin, a plasma membrane protein that is expressed in most cystic epithelia from patients with PKD1.[91] The physiologic role of polycystin in PKD1 is not fully defined, although it may be involved in cell-cell or cell-matrix interactions. The remaining patients have the so-called non-PKD1 abnormality and have a different clinical course.[92] In most other patients, a locus for this abnormality has been found on chromosome 4 (the PKD2 locus); it encodes for polycystin 2, which may be involved in cell calcium signaling. As many as 75% of all ADPKD patients have a positive family history.[90,92,93] In ADPKD, the cysts arise from outpouchings of the renal tubule and Bowman space. They are lined by epithelium and contain fluid derived from the glomerular filtrate and modified by the action of the tubular epithelial lining. The underlying mechanisms of cyst formation in ADPKD are unclear.[90,91,93] Cyst formation begins in utero, and cysts increase in size and number as the patient ages. Most patients come to medical attention in middle age. Cyst enlargement and ESRD occur later in life in PKD2.[92] As the cysts enlarge, they compress adjacent normal tissue and lead to interstitial scarring. Although the exact mechanism responsible for this interstitial damage is unclear, increased cyst growth is associated with a decline in the GFR. Symptoms and signs include flank or back pain, abdominal masses, gross hematuria, urinary tract infections, and stone disease. Patients are typically less anemic than is expected for their degree of renal insufficiency. Hypertension that is related to increased activation of the renin-angiotensin-aldosterone system is found in 60% to 75% of adults with ADPKD and is an early manifestation of the disease.

Although CT may be a more sensitive technique for diagnosis of PKD, ultrasonography is usually the diagnostic procedure of choice because it is sensitive and inexpensive and does not expose the patient to radiation. In a patient younger than 30 years, diagnostic criteria for PKD1 include the finding of two cysts either unilaterally or bilaterally; in a patient between 30 and 60 years, the finding of at least two cysts in each kidney; and in patients older than 60 years, the finding of at least four cysts in each kidney. The presence of at least three cysts in each kidney of a patient with a family history of ADPKD confirms the diagnosis, and the presence of any cysts on ultrasonography in a child with a family history of ADPKD is strongly suggestive. The absence of cysts on ultrasonography at 30 years of age excludes the diagnosis in almost 100% of patients.

The kidneys typically enlarge progressively as the patient ages and the cysts grow.[93,94] The kidneys may reach massive proportions, causing abdominal distention [*see Figures 7 and 8*]. Half of affected persons with the ADPKD1 abnormality progress to

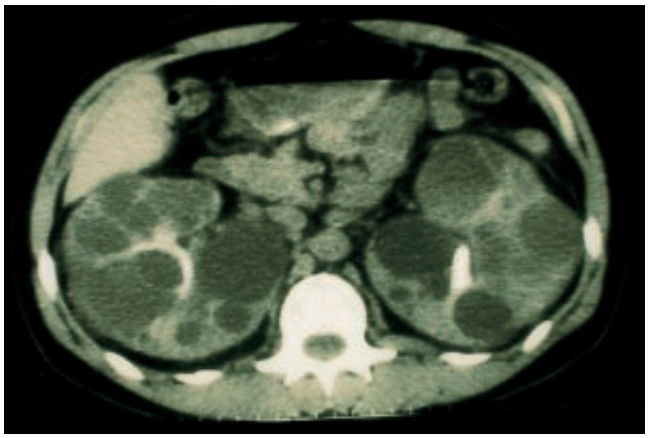


Figure 7 **Computed tomography of the abdomen demonstrates multiple bilateral renal cysts in a patient with autosomal dominant polycystic kidney disease.**

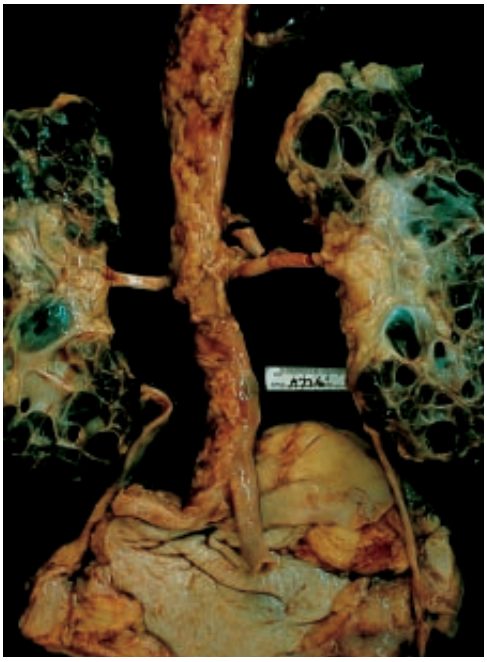


Figure 8 **Multiple cysts of all sizes can be observed in this gross section of enlarged kidneys from a patient with autosomal dominant polycystic kidney disease.**

renal failure by 60 years of age. The mean age of end-stage renal failure in patients with ADPKD1 is 55 to 60 years, compared with 70 years for non-ADPKD1 patients who die of renal failure. Other factors that presage more rapidly progressive renal disease are diagnosis at a younger age, black race, male sex, larger renal volume, and the presence of hypertension, hepatic cysts (in women), and gross hematuria. Cysts can also occur in the liver, pancreas, and spleen [*see Figure 9*].[95,96] Hepatic cysts increase with age—from 10% of patients younger than 30 years to more than 40% of patients older than 60 years—and are more common in women, especially in those who have had several pregnancies. Other systemic findings that commonly occur in patients with polycystic kidney disease are berry aneurysms of the circle of Willis [*see Figures 10 and 11*], which occur in 4% to 10% of patients; mitral valve prolapse, which is revealed by echocardiography in as many as 25% of patients; and colonic diverticula and diverticulitis and hiatal hernias.[97,98]

There is no specific treatment for ADPKD. Ascending infections can be treated with conventional antibiotics, but infected cysts must be treated with lipid-soluble antibiotics such as trimethoprim-sulfamethoxazole, chloramphenicol, and ciprofloxacin.[99] Control of infections and hypertension may slow the progression of renal dysfunction.[94] Although early studies suggested a benefit of a protein-restricted diet in PKD1, this has been only moderately supported by large trials.[100] Cyst puncture or surgical decompression may relieve pressure symptoms, although it will not alter the course of the disease.[101] Genetic counseling is useful because the chromosomal abnormalities can be detected reliably in almost all patients with ADPKD.

Autosomal recessive polycystic kidney disease, formerly called infantile polycystic disease, is a rare disorder, with an incidence of one in 40,000 persons to one in 10,000 persons.[102] It presents as severe cystic involvement in infancy or early childhood; may be associated with hepatic cysts, fibrosis, and portal hypertension; and leads to progressive severe renal failure. Often, death occurs at an early age.

ACQUIRED CYSTIC DISEASE

Acquired cystic disease is diagnosed when patients with renal insufficiency or renal failure are found to have more than five cysts per kidney in the absence of a family history.[103] In acquired cystic disease, in contrast to ADPKD, the kidneys are not enlarged but rather are small and scarred; in addition, there are no systemic manifestations. Such cysts develop in approximately 20% to 25% of dialysis patients and in as many as 50% of patients on dialysis for more than 5 years. The cysts are far more common in men than in women. The pathogenesis of these cysts is not understood, but they are limited to the kidney and arise from the renal tubules, especially from the proximal tubules, and vary in size from microscopic to several centimeters in diameter. Although most patients with acquired cystic disease do not have symptoms related to the cysts, some have hematuria, pain, fever with infection, and erythrocytosis. Renal tumors can develop within the cysts.[104] In dialysis patients with cystic disease, the incidence of adenomas is 5% to 10% and that of adenocarcinomas is 4% to 7%. The incidence of malignant tumors increases with increasing time on dialysis, and in 50% of cases, the tumors are multiple or bilateral.[105]

Multiple cysts have also been described in patients with hypokalemia and hyperaldosteronism. These cysts may disappear after successful removal of an adrenal adenoma and are usually of no major significance.

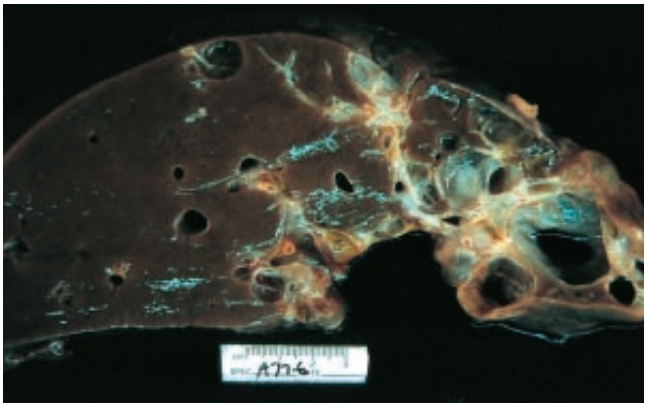


Figure 9 **Cysts are apparent in the liver of a patient with autosomal dominant polycystic kidney disease.**

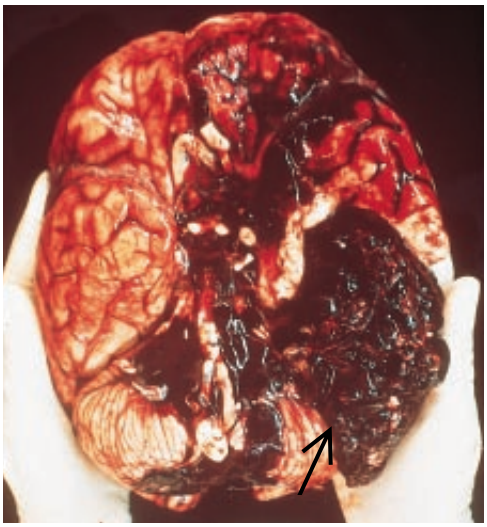


Figure 10 **A massive subarachnoid hemorrhage (arrow) caused by a ruptured berry aneurysm of the circle of Willis is apparent in this autopsy specimen from a patient with autosomal dominant polycystic kidney disease.**

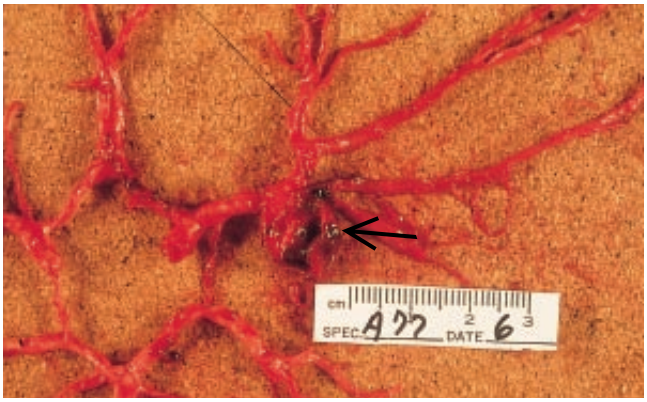


Figure 11 **A close-up of the circle of Willis shown in Figure 10 reveals the ruptured aneurysm (arrow).**

NEPHRONOPHTHISIS AND MEDULLARY CYSTIC DISEASE

Familial juvenile nephronophthisis and medullary cystic disease constitute a group of disorders characterized by cystic in-

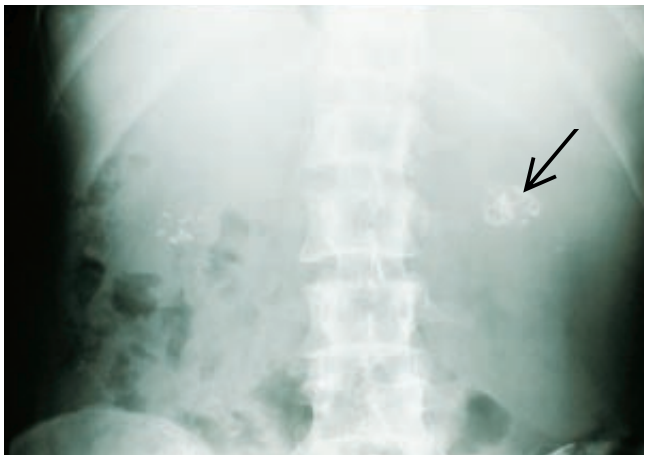


Figure 12 **Abdominal x-ray without contrast shows calcifications (arrow) in dilated collecting ducts in medullary sponge kidney.**

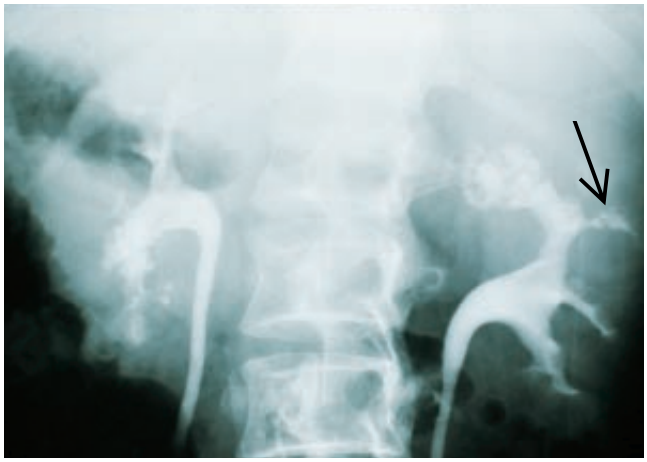


Figure 13 **Intravenous pyelogram of medullary sponge kidney reveals the so-called flower-spray or paintbrush dilatations of the collecting ducts of Bellini (arrow).**

volvement of the renal medulla and insidious progression to renal insufficiency and ESRD. Nephronophthisis typically occurs in the first 2 decades of life. It is inherited as an autosomal recessive trait, which in many families localizes to chomosome 2q13, where it encodes for the protein nephrocystin.[106,107] In some patients, it is associated with retinal pigment degeneration (i.e., the retinal-renal dysplasia syndrome). Medullary cystic disease presents most often in the third or fourth decade and is typically inherited as an autosomal dominant trait with genetic heterogeneity.[108] In both diseases, the kidneys are shrunken and have a thinned cortex. The cysts arise from the distal and collecting tubule cells and are most prominent at the corticomedullary junction, where they are surrounded by scarring and interstitial fibrosis. Patients typically present with polyuria, polydipsia, and nocturia, which result from impaired renal concentrating ability. Such patients often excrete large amounts of sodium in the urine, which can lead to volume depletion during periods of limited fluid access. These disorders invariably progress to renal failure, and the anemia is proportional to the degree of renal failure.

MEDULLARY SPONGE KIDNEY

In medullary sponge kidney, small cystic outpouchings form in the collecting ducts of the renal papillae. In some cases, families exhibit a dominant inheritance pattern, but other cases are sporadic. Asymptomatic ectasias 1 to 7 mm long are found diffusely and bilaterally in the collecting tubules by use of intravenous urography. They give a so-called paintbrush or flower-spray appearance to the renal papillae. At times, calcifications in the ducts can be seen on plain abdominal radiographs [*see Figures 12 and 13*]. Medullary sponge kidney may be associated with microscopic or gross hematuria, hypercalciuria, nephrocalcinosis and calcium stone disease, urinary tract infections, and, occasionally, isolated tubular defects (e.g., a distal renal tubular acidosis) and potassium-excreting and potassium-conserving defects.[109] The disease does not cause a major decline in the GFR and does not lead to renal failure.

Other Diseases Associated with Tubulointerstitial Nephritis

IDIOPATHIC ACUTE INTERSTITIAL NEPHRITIS

Patients with biopsy-documented AIN who are taking no medications and have no associated systemic or renal infections are classified as having idiopathic AIN.[2] In rare cases, the presence of circulating antibodies against the tubular basement membrane or the demonstration of linear immunofluorescence along the renal tubular basement membranes documents the immune nature of the renal insult. The pathogenesis is unclear in most patients who have either undefined granular immune deposits or no immune deposition. Patients usually present with either acute renal insufficiency or renal failure; on rare occasions, patients present with isolated tubular defects. Acute renal failure sometimes responds dramatically to short courses of corticosteroids or other immunosuppressive therapy.

A small group of patients, usually young women, have eosinophilic AIN associated with a variable combination of anemia, azotemia, uveitis, peripheral eosinophilia, and granulomas of the bone marrow or lymph nodes.[110] Such patients with the so-called TINU syndrome (tubulointerstitial nephritis and uveitis) may have tubular defects or acute renal failure. Although the uveitis may recur, the renal disease in TINU syndrome is often cured by treatment with corticosteroids.

SYSTEMIC LUPUS ERYTHEMATOSUS AND OTHER IMMUNE COMPLEX DISEASES

Although systemic lupus erythematosus is commonly associated with glomerular lesions, renal biopsies often show TIN or immune complex deposits along the tubular basement membrane as well.[111] Of patients with severe lupus nephritis, 60% to 70% have such interstitial involvement. In rare instances, TIN occurs in systemic lupus erythematosus without associated glomerular changes. In general, the degree of interstitial inflammation correlates with the degree of renal dysfunction, the presence of hypertension, and the severity of the glomerular damage. Such inflammation also augurs a greater decline in renal function, which is independent of glomerular histology. In comparison, immune complex deposition along the tubular basement membranes correlates with serologic activity (a high anti–DNA antibody titer and low complement level), but it does not correlate with other clinical features. Immune complex deposition is probably only one of many mechanisms that lead to tubulointerstitial damage. On rare occasions, patients with systemic lupus erythematosus have isolated tubular defects in acid-

ification or potassium excretion. In patients with mixed cryoglobulinemia or Sjögren syndrome, acute and chronic interstitial infiltrates may be found in the kidney with or without clinical features of renal dysfunction and such tubular disorders as Fanconi syndrome, distal renal tubular acidosis, nephrogenic diabetes insipidus, and hypokalemia.

SARCOIDOSIS

CIN is common in sarcoidosis.[112,113] Renal involvement typically includes diffuse interstitial noncaseating granulomas that displace and disrupt the renal parenchyma, nephrocalcinosis secondary to hypercalcemia and hypercalciuria, and acute interstitial lymphocytic inflammatory infiltrates with giant cells. The interstitial inflammation is often asymptomatic; as many as 20% of patients without clinical renal disease have interstitial granulomas at autopsy. Patients with clinically evident renal dysfunction usually have clear evidence of diffuse active sarcoidosis. Patients usually have mild proteinuria, tubular defects, sterile pyuria, and concentrating and acidifying defects. On rare occasions, patients present with acute renal failure associated with acute severe interstitial inflammation and granuloma formation, which may respond dramatically to corticosteroid therapy.

BALKAN NEPHROPATHY

Balkan nephropathy is a chronic progressive interstitial nephritis of unknown etiology that is found in long-term inhabitants of the rural areas surrounding the tributaries of the Danube River.[114] Although multiple environmental and genetic factors have been proposed as the etiology of Balkan nephropathy, none has been proved. Patients present in later life with tubular proteinuria, glycosuria, renal tubular acidosis, and progressive azotemia. The renal histopathology is characterized by an atrophic cortex, severe tubular degeneration, and interstitial fibrosis. Uroepithelial cancers of the renal pelvis and ureters develop in more than one third of patients.

ACUTE RENAL TRANSPLANT REJECTION

Acute rejection of a renal allograft is characterized by interstitial and perivascular infiltrates of lymphocytic cells, with lymphocytes invading the spaces between the cells lining the tubules (so-called lymphocytic tubulitis), and intimal arteritis.[115] Acute cellular rejection is associated with a decrease in the GFR and is a potentially treatable form of transplant rejection. If the rejection episode does not respond to corticosteroid therapy or other immunosuppressive medications, it can result in varying degrees of chronic interstitial fibrosis and scarring.

References

1. Rastegar A, Kashgarian M: The clinical spectrum of tubulointerstitial nephritis. Kidney Int 54:313, 1998

2. Finn WF: Tubulointerstitial diseases. Primer on Kidney Disease, 3rd ed. Greenberg A, Cheung AK, Coffman TM, et al, Eds. Elsevier Science and Technology, San Diego, 2001, p 319

3. Meyers CM, Neilson EG: Immunopathogenesis of tubulointerstitial disease. Massry & Glassock's Textbook of Nephrology, 3rd ed, vol 2. Massry SG, Glassock RJ, Eds. Williams & Wilkins, Baltimore, 1995, p 671

4. Neilson EG: Pathogenesis and therapy of interstitial nephritis. Kidney Int 35:1257, 1989

5. Border WA, Noble NA: TGF-beta in kidney fibrosis. Kidney Int 51:1388,1997

6. Appel GB: Acute interstitial nephritis. Immunologic Renal Diseases, 2nd ed. Neilson EG, Couser WG, Eds. Lippincott-Raven Publishers, Philadelphia, 2001, p 1269

7. Michel DM, Kelly CJ: Acute interstitial nephritis. J Am Soc Nephrol 9:506, 1998

8. Rossert J: Drug-induced acute interstitial nephritis. Kidney Int 60:804, 2001

9. Ruffing KA, Hoppes P, Blend D, et al: Eosinophils in urine revisited. Clin Nephrol 41:163, 1994

10. Haas M, Spargo BH, Wit EJ, et al: Etiologies and outcome of acute renal insufficiency in older adults: a renal biopsy study of 259 cases. Am J Kidney Dis 35:433, 2000

11. Schwarz A, Krause P-H, Kunzendorf U, et al: The outcome of acute interstitial nephritis: risk factors for the transition from acute to chronic interstitial nephritis. Clin Nephrol 54:179, 2000

12. D'Agati VD, Theise ND, Pirani CL, et al: Interstitial nephritis related to nonsteroidal anti-inflammatory agents and beta-lactam antibiotics: a comparative study of the interstitial infiltrates using monoclonal antibodies. Mod Pathol 2:390, 1989

13. Markowitz GS, Falkowitz DC, Isom R, et al: Membranous glomerulopathy and acute interstitial nephritis following treatment with Celebrex. Clin Nephrol 59:137, 2003

14. Appel GB: COX 2 inhibitors and the kidney. Clin Exp Rheumatol 19:S37, 2001

15. Feinstein AR, Heinemann LA, Curhan GC, et al: Relationship between nonphenacetin combined analgesics and nephropathy. Kidney Int 58:2259, 2000

16. Fored CM, Ejerblad E, Lindblad P, et al: Acetaminophen, aspirin, and chronic renal failure. N Engl J Med 345:1801, 2001

17. Mclaughlin JK, Lipworth L, Chow W-H, et al: Analgesic use and chronic renal failure: a critical review of the epidemiologic literature. Kidney Int 54:679, 1998

18. Delzell E, Schapiro S: A review of epidemiologic studies of nonnarcotic analgesics and chronic renal disease. Medicine 77:102, 1998

19. Bennett WM, Porter GA: Analgesic nephropathy and the use of nonsteroidal antiinflammatory drugs in renal patients: new insights. J Nephrol 11:70, 1998

20. Dubach UC, Rosner B, Sturmer T: An epidemiologic study of abuse of analgesic drugs: effects of phenacetin and salicylate on mortality and cardiovascular morbidity. N Engl J Med 324:155, 1991

21. Sandler DP, Smith JC, Weinberg CR, et al: Analgesic use and chronic renal disease. N Engl J Med 320:1238, 1989

22. Rexrode KM, Buring JE, Glynn RJ, et al: Analgesic use and renal function in men. JAMA 286:315, 2001

23. Henrich WL, Agodoa LE, Barrett B, et al: Analgesics and the kidney: summary and recommendation to the Scientific Advisory Board of the National Kidney Foundation from an ad hoc committee of the National Kidney Foundation. Am J Kidney Dis 27:162, 1996

24. Elseviers MM, De Schepper A, Cortouts R, et al: High diagnostic performance CT scan for analgesic nephropathy in patients with incipient to severe renal failure. Kidney Int 48:1316, 1995

25. Garber SL, Mirochnik Y, Arruda JA, et al: Evolution of experimentally induced papillary necrosis to focal segmental glomerulosclerosis and nephrotic proteinuria. Am J Kidney Dis 33:1033, 1999

26. Segasothy M, Samad SA, Zulfigar A, et al: Chronic renal disease and papillary necrosis associated with the long-term use of nonsteroidal anti-inflammatory drugs as the sole or predominant analgesic. Am J Kidney Dis 24:17, 1994

27. Boton R, Gaviria M, Batlle DC: Prevalence, pathogenesis and treatment of renal dysfunction associated with chronic lithium therapy. Am J Kidney Dis 10:329, 1987

28. Bendz H, Aurell M, Balldin J, et al: Kidney damage in long-term lithium patients: a cross-sectional study of patients with 15 years or more on lithium. Nephrol Dial Transplant 9:1250, 1994

29. Markowitz G, Radhakrishnan J, Khambham N, et al: Lithium nephrotoxicity: a progressive combined glomerular and tubulointerstitial nephropathy. J Am Soc Nephrol 11:1439, 2000

30. Peterson PM, Hansen SW: The course of long-term toxicity in patients treated with cisplatin-based chemotherapy for non-seminomatous germ-cell cancer. Ann Oncol 10:1475, 1999

31. Bokemeyer C, Berger CC, Kuczyk MA, et al: Evaluation of long-term toxicity after therapy for testicular cancer. J Clin Oncol 14:2923, 1996

32. Horwich A, Sleijfer DT, Fossa SD, et al. Randomized trial of bleomycin, etoposide, and cisplatin compared with bleomycin, etoposide, and carboplatin in good prognosis metastatic nonseminomatous germ cell cancer: a Multiinstitutional Medical Research Council/ European Organization for Research and Treatment Cancer Trial. J Clin Oncol 15:1844, 1997

33. Skinner R, Pearson AD, English MW, et al: Risk factors for ifosfamide nephrotoxicity in children. Lancet 348:578, 1996

34. Rossi R: Nephrotoxicity of ifosfamide: moving toward understanding the molecular mechanisms. Nephrol Dial Transplant 12:1091, 1997

35. de Mattos AM, Olyaei AJ, Bennett WM: Nephrotoxicity of immunosuppressive drugs: long-term consequences and challenges for the future. Am J Kidney Dis 35:333, 2000

36. Benigni A, Bruzzi I, Mister M, et al: Nature and mediators of renal lesions in kidney transplant patients given cyclosporine for more than one year. Kidney Int 55:674, 1999

37. Isnard Bagnis C, Tezenas du Montcel S, Beaufils H, et al: Long-term renal effects of low dose cyclosporine in uveitis-treated patients. J Am Soc Nephrol 13:2962, 2002

38. Miura K, Nakatani T, Asai T, et al: Role of hypomagnesemia in chronic cyclosporine nephropathy. Transplantation 73:340, 2002

39. Porayko MK, Textor SC, Krom RA, et al: Nephrotoxic effects of primary immunosuppression with FK-506 and cyclosporine regimens after liver transplantation. Mayo Clin Proc 69:105, 1994

40. Korbet SM: Obstructive uropathy. Primer on Kidney Diseases, 3rd ed. Greenberg A, Cheung AK, Coffman TM, et al, Eds. Elsevier Science and Technology, San Diego, 2001, p 336

41. Demko TM, Diamond JR, Groff J: Obstructive nephropathy as a result of retroperitoneal fibrosis: a review of its pathogenesis and associations. J Am Soc Nephrol 8:684, 1997

42. Dillon MJ, Goonasekera CD: Reflux nephropathy. J Am Soc Nephrol 9:2377, 1998

43. Hiraoka M, Hori C, Tsukahara H, et al: Vesicoureteral reflux in male and female neonates as detected by voiding ultrasonography. Kidney Int 55:1486, 1999

44. Kohler J, Tencer J, Thysell H, et al: Vesicoureteral reflux diagnosed in adulthood: incidence of urinary tract infections, hypertension, proteinuria, back pain and renal calculi. Nephrol Dial Transplant 12:2580, 1997

45. Smellie JM, Jodal U, Lax H, et al: Outcome at 10 years of severe vesicoureteric reflux managed medically: report of the International Reflux Study in Children. J Pediatr 139:656, 2001

46. Smellie JM, Barratt TM, Chantler C, et al: Medical versus surgical treatment in children with severe bilateral vesicoureteric reflux and bilateral nephropathy: a randomized trial. Lancet 357:1329, 2001

47. Borg M, Hughes T, Horvath N, et al: Renal toxicity after total body irradiation. Int J Radiat Oncol Biol Phys 54:1165, 2002

48. Nicolle L: Urinary tract infection. Primer on Kidney Diseases, 3rd ed. Greenberg A, Cheung AK, Coffman TM, et al, Eds., Elsevier Science and Technology, San Diego, 2001, p 354

49. Weiner ID, Wingo CS: Hypokalemia: consequences, causes, and correction. J Am Soc Nephrol 8:1179, 1997

50. Randhawa PS, Finkelstein S, Scantlebury V, et al: Human polyoma virus-associated interstitial nephritis in the allograft kidney. Transplantation 67:103, 1999

51. Nickeleit V, Klimkait T, Binet IF, et al: Testing for polyomavirus type BK DNA in plasma to identify renal-allograft recipients with viral nephropathy. N Engl J Med 342:1309, 2000

52. Ramos E, Drachenberg CB, Papadimitriou JC, et al: Clinical course of polyoma virus nephropathy in 67 renal transplant patients. J Am Soc Nephrol 13:2145, 2002

53. D'Agati V, Appel GB: Renal pathology of human immunodeficiency virus infection. Semin Nephrol 18:406, 1998

54. Milliner DS, Wilson DM, Smith LH: Phenotypic expression of primary hyperoxaluria: comparative features of types I and II. Kidney Int 59:31, 2001

55. Cochat P: Primary hyperoxaluria type 1. Kidney Int 55:2533, 1999

56. Danpure CJ: Advances in the enzymology and molecular genetics of primary hyperoxaluria type 1: prospects for gene therapy. Nephrol Dial Transplant 10:24, 1995

57. Gruessner RW: Preeemptive liver transplantation from a living related donor for primary hyperoxaluria type 1. N Engl J Med 338:1924, 1998

58. Saborio P, Scheinman JI: Transplantation for primary hyperoxaluria in the United States. Kidney Int 56:1094, 1999

59. Monico CG, Milliner DS: Hyperoxaluria. Primer on Kidney Diseases, 3rd ed. Greenberg A, Cheung AK, Coffman TM, et al, Eds. Elsevier Science and Technology, Dordrecht, 2001, p 327

60. Johnson RJ, Kivlighn SD, Kim YG, et al: Reappraisal of the pathogenesis and consequences of hyperuricemia in hypertension, cardiovascular disease, and renal disease. Am J Kidney Dis 33:225, 1999

61. Tsao T, Fawcett J, Fervenza FC, et al: Expression of insulin-like growth factor-I and transforming growth factor-beta in hypokalemic nephropathy in the rat. Kidney Int 59:96, 2001

62. Suga S, Mazzali M, Ray PE, et al: Angiotensin II type I receptor blockade ameliorates tubulointerstitial injury induced by chronic potassium deficiency. Kidney Int 61:951, 2002

63. Kim R, Rotnitsky A, Sparrow D, et al. A longitudinal study of low-level lead exposure and impairment of renal function. JAMA 275:1177, 1996

64. Fels LM, Wunsch M, Baranowski J, et al: Adverse effects of chronic low level lead exposure on kidney function—a risk group study in children. Nephrol Dial Transplant 13:2248, 1998

65. Muntner P, He J, Vupputuri S: Blood lead and chronic kidney disease in the general United States population: results from NHANES III. Kidney Int 63:1044, 2003

66. Lin JL, Ho HH, Yu CC: Chelation therapy for patients with elevated body lead burden and progressive renal insufficiency: a randomized controlled trial. Ann Intern Med 130:7, 1999

67. Hotz P, Buchet JP, Bernard A, et al: Renal effects of low-level environmental cadmium exposure: 5-year follow-up of a subcohort from the Cadmibel study. Lancet 354:1508, 1999

68. Hellstrom L, Elinder CG, Dahlberg B, et al: Cadmium exposure and end-stage renal disease. Am J Kidney Dis 38:1001, 2001

69. Yang CS, Lin CH, Chang SH, et al: Rapidly progressive fibrosing interstitial nephritis associated with Chinese herbal drugs. Am J Kidney Dis 35:313, 2000

70. Vanherweghem JL: Nephropathy and herbal medicine. Am J Kidney Dis 35:330, 2000

71. Debelle FD, Nortier JL, De Prez EG, et al: Aristolochic acids induce chronic renal failure with interstitial fibrosis in salt-depleted rats. J Am Soc Nephrol 13:431, 2002

72. Nortier JL, Martinez MC, Schmeisser HH, et al: Urothelial carcinoma associated with the use of a Chinese herb (*Aristolochia fangchi*). N Engl J Med 342:1686, 2000

73. Wesson DE: The initiation and progression of sickle cell nephropathy. Kidney Int 61:2277, 2002

74. Pham PT, Pham PC, Wilkinson AH, et al: Renal abnormalities in sickle cell disease. Kidney Int 57:1, 2000

75. Ataga KI, Orringer EP: Renal abnormalities in sickle cell disease. Am J Hematol 63:205, 2000

76. Scolari F, Tardanico R, Zani R, et al: Cholesterol crystal embolism: a recognizable cause of renal disease. Am J Kidney Dis 36:1089, 2000

77. Modi KS, Rao VK: Atheroembolic renal disease. J Am Soc Nephrol 12:1781, 2001

78. Meyrier A, Hill GS, Simon P: Ischemic renal disease: new insights into old entities. Kidney Int 54:2, 1998

79. Belenfant X, Meyrier A, Jacquot C: Supportive treatment improves survival in multivisceral cholesterol crystal embolism. Am J Kidney Dis 33:840, 1999

80. Orth SR, Ritz E: The nephrotic syndrome. N Engl J Med 338:1202, 1998

81. Appel GB: The nephrotic syndrome in glomerulonephritis. Cecil Textbook of Medicine, 22nd ed. Goldman L, Ausiell D, Eds., Elsevier Science, Philadelphia (in press)

82. Irish AB, Winearls CG, Littlewood T: Presentation and survival of patients with severe renal failure and myeloma. QJM 90:773, 1997

83. Blade J, Fernandez-Llama P, Bosch F, et al: Renal failure in multiple myeloma: presenting features and predictors of outcome in 94 patients from a single institution. Arch Intern Med 158:1889, 1998

84. Lin J, Markowitz GS, Valeri AM, et al: Renal monoclonal immunoglobulin deposition disease: the disease spectrum. J Am Soc Nephrol 12:1482, 2001

85. Bataille R, Harousseau JL: Multiple myeloma. N Engl J Med 336:1657, 1997

86. Kunis CLK, Ward H, Appel GB: Renal disease associated with illicit drug abuse. Clinical Nephrotoxins: Renal Injury from Drugs and Chemicals, 2nd ed DeBroe M, Porter G, Bennett WM, et al, Eds. Kluwer Academic Publishers, Dordrecht,, 2003, p 397

87. Pozzi C, Locatelli F: Kidney and liver involvement in monoclonal light chain disorders. Semin Nephrol 22:319, 2002

88. Slywotzky CM, Bosniak MA: Localized cystic disease of the kidney. AJR Am J Roentgenol 176:843, 2001

89. Gabow PA: Autosomal dominant polycystic kidney disease. N Engl J Med 329:332, 1993

90. Igarashi P, Somlo S: Genetics and pathogenesis of polycystic kidney disease. J Am Soc Nephrol 13:2384, 2002

91. Wilson PD: Polycystin: new aspects of structure, function, and regulation. J Am Soc Nephrol 12:834, 2001

92. Hateboer N, v Dijk MA, Bogdanova N, et al: Comparison of phenotypes of polycystic kidney disease types1 and 2: European PKD1-PKD2 Study Group. Lancet 353:103, 1999

93. Grantham JJ: The etiology, pathogenesis, and treatment of autosomal dominant polycystic kidney disease: recent advances. Am J Kidney Dis 28:788, 1996

94. Peters DJ, Breuning MH: Autosomal dominant polycystic kidney disease: modification of disease progression. Lancet 358:1439, 2001

95. Watson ML: Complications of polycystic kidney disease. Kidney Int 51:353, 1997

96. Chauveau D, Fakhouri F, Grunfeld JP: Liver involvement in autosomal-dominant polycystic kidney disease: therapeutic dilemma. J Am Soc Nephrol 11:1767, 2000

97. Pirson Y, Chauveau D, Torres V: Management of cerebral aneurysms in autosomal dominant polycystic kidney disease. J Am Soc Nephrol 13:269, 2002

98. Lumiaho A, Ikaheimo R, Miettinen R, et al: Mitral valve prolapse and mitral regurgitation are common in patients with polycystic kidney disease type 1. Am J Kidney Dis 38:1208, 2001

99. Gibson P, Watson ML: Cyst infection in polycystic kidney disease: a clinical challenge. Nephrol Dial Transplant 13:2455, 1998

100. Dietary protein restriction, blood pressure control, and the progression of polycystic kidney disease. Modification of Diet in Renal Disease Study Group. J Am Soc Nephrol 5:2037, 1995

101. Elzinga LW, Barry JM, Bennett WM: Surgical management of painful polycystic kidneys. Am J Kidney Dis 22:532, 1993

102. Zerres K, Rudnik-Schoneborn S, Senderek S: Autosomal recessive polycystic kidney disease (ARPKD). J Nephrol 16:453, 2003

103. Fick GM, Gabow PA: Hereditary and acquired cystic disease of the kidney. Kidney Int 46:951, 1994

104. Truong LD, Krishnan B, Cao JT, et al: Renal neoplasm in acquired cystic disease. Am J Kidney Dis 26:1, 1995

105. Sarasin FP, Wong JB, Levey AS, et al: Screening for acquired cystic kidney disease: a decision analytic perspective. Kidney Int 48:207, 1995

106. Hildebrandt F, Rensing C, Betz R, et al: Establishing an algorithm for molecular genetic diagnostics in 127 families with juvenile nephronophthisis. Kidney Int 59:434, 2001

107. Caridi G, Dagnino M, Gusmano R, et al: Clinical and molecular heterogeneity of juvenile nephronophthisis in Italy: insights from molecular screening. Am J Kidney Dis 35:44, 2000

108. Stavrou C, Koptides M, Tombazos C, et al: Autosomal-dominant medullary cystic disease type 1: clinical and molecular findings in six large Cypriot families. Kidney Int 62:1385, 2002

109. Yagisawa T, Kobayashi C, Hayashi T, et al: Contributory metabolic factors in the development of nephrolithiasis in patients with medullary sponge kidney. Am J Kidney Dis 37:1140, 2001

110. Takemura T, Okada M, Hino S, et al: Course and outcome of tubulointerstitial nephritis and uveitis syndrome. Am J Kidney Dis 34:1016, 1999

111. Appel GB, Radhakrishnan J, D'Agati V: Secondary glomerular disease. Brenner and Rector's The Kidney, 7th ed. Brenner BM, Ed. WB Saunders Co, Philadelphia (in press)

112. O'Riordan E, Willert RP, Reeve R, et al: Isolated sarcoid granulomatous interstitial nephritis: review of five cases at one center. Clin Nephrol 55:297, 2001

113. Brause M, Magnusson K, Degenhardt S, et al: Renal involvement in sarcoidosis: a report of six cases. Clin Nephrol 57:142, 2002

114. Balkan endemic nephropathy. Clinical Nephrotoxins: Renal Injury from Drugs and Chemicals, 2nd ed. DeBroe M, Porter G, Bennett WM, et al, Eds. Kluwer Academic Publisher, Dordrecht, 2003, p 425

115. Norman DJ: Renal transplantation: immunosuppression and postoperative management. Primer on Kidney Diseases, 3rd ed. Greenberg A, Cheung AK, Coffman TM, et al, Eds., Elsevier Science and Technology, San Diego, 2001, p 460

IX PHARMACOLOGIC APPROACH TO RENAL INSUFFICIENCY

Suzanne K. Swan, m.d.

The metabolism and elimination of many pharmacologic agents depend on normal renal function, in part because these agents have pharmacologically active metabolites that undergo renal excretion. In addition, many agents that do not undergo renal elimination from the body may nonetheless cause adverse effects in patients with renal failure. As a result, dosage adjustment of many drugs is required for patients with renal insufficiency to prevent toxicity and ensure efficacy. The clinician who prescribes therapy for patients with altered renal function must be familiar with basic pharmacologic principles because such patients must often be maintained within a narrow therapeutic window; that is, both drug accumulation and subtherapeutic dosing must be prevented.

This subsection addresses the pharmacologic parameters that are altered by renal dysfunction and presents a systematic, stepwise approach to dosage adjustment. Drug removal by dialysis and specific pharmacologic considerations in the setting of renal insufficiency are also outlined. Dosage recommendations appear in tabular form as the appendix to this subsection and represent a rough guide to drug therapy in patients with renal insufficiency. These dosing recommendations are derived from an extensive database in the medical literature that is often conflicting, that applies to varied patient populations, and that is seldom based on prospective, controlled trials. Thus, the prescribing clinician must not depend solely on dosing tables or nomograms; rather, such information should be used as a starting point in the treatment of patients with renal insufficiency.

Altered Pharmacokinetic Principles in Renal Failure

A given compound may travel through several different compartments in the body before being eliminated [*see Figure 1*]. Alterations in renal function affect several pharmacokinetic factors, including bioavailability, volume of distribution (V_D), protein binding, and biotransformation.

bioavailability

The bioavailability of a drug refers to the fraction, expressed as a percentage, of a given dose that reaches the systemic circulation. Bioavailability is primarily determined by the rate and route of administration. In general, a drug that is administered intravenously has 100% bioavailability because all of the dose reaches the systemic circulation; when it is given by the oral, intramuscular, or subcutaneous route, a smaller percentage of the dose often reaches target tissues. For example, furosemide has 100% bioavailability when given intravenously but only 50% when given orally. This finding underlies the common practice of doubling the dose of furosemide when a patient is switched from an intravenous to an oral regimen.

Drug absorption dictates bioavailability and may be altered in patients with renal failure. Diminished absorption may result from uremia-induced vomiting or gastroparesis caused by the neuropathic changes associated with diabetes mellitus or aging. The bowel wall edema commonly seen in patients with cirrhosis, nephrosis, or congestive heart failure may also impair absorption. Concomitant administration of certain medications, such as aluminum- or calcium-containing phosphate binders with antibiotics or iron supplements, can result in the formation of insoluble complexes that both limit absorption and slow gut motility. Absorption is also impaired by the concomitant administration of drugs that increase gastric pH, which include the phosphate binders and the H_2-receptor blockers, such as ranitidine and cimetidine.

volume of distribution

The V_D of a pharmacologic agent does not refer to a specific anatomic compartment but is a value derived by dividing the total amount of drug in the body by its concentration in the blood. Clinically, it can be used to calculate the dose required to achieve a desired systemic drug level. As a general rule, there is an inverse correlation between the serum concentration and the V_D. Alterations in extracellular fluid volume can affect the V_D; volume contraction corresponds to a fall in the V_D and a subsequent rise in the serum concentration, particularly of hydrophilic compounds, such as the aminoglycosides. Conversely, edema and ascites tend to increase the V_D of these agents, which results in lower serum concentrations. Both digoxin and insulin have a markedly decreased V_D in uremic patients, as evidenced by the increased serum concentrations of these agents in such patients.

protein binding

The portion of a given agent that is bound to protein can be thought of as the storage pool for that agent. Although compounds circulate in both bound and unbound forms, it is only the unbound, or free, form that is ultimately distributed and biologically active. Renal failure tends to decrease protein binding for most agents [*see Table 1*] because organic waste products block the binding sites on carrier proteins and thus displace the pharmacologic agent. As a result, a larger proportion of the drug circulates in its unbound, active form. Because standard drug as-

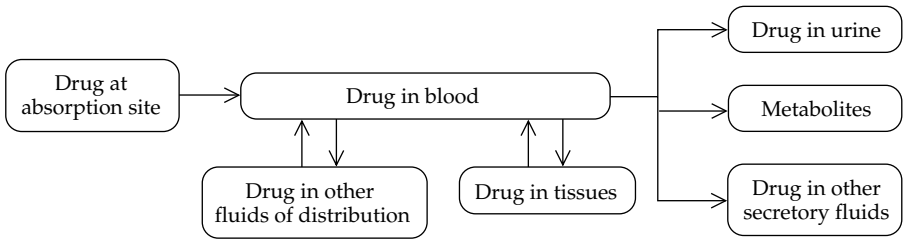


Figure 1 **Compounds may traverse several different compartments in the body before being eliminated. Changes in renal function affect pharmacokinetic factors, often necessitating dosage adjustment in patients with renal insufficiency.**

Table 1 Compounds with Decreased Protein Binding in Uremic Patients

Cephalosporins	Phenytoin
Clofibrate	Primidone
Diazepam	Salicylate
Diazoxide	Sulfonamide
Furosemide	Theophylline
Morphine	Valproic acid
Penicillins	Warfarin
Phenobarbital	

says tend to measure total drug concentrations (i.e., both bound and unbound drug levels), it may be prudent to specifically monitor unbound drug levels (e.g., in the case of a patient who is receiving phenytoin) when the therapeutic index is narrow.

BIOTRANSFORMATION

The biotransformation of a drug is its biochemical conversion from one chemical form to another. Biotransformation usually occurs via one of the hepatic metabolic pathways, such as oxidation, reduction, acetylation, or hydrolysis, and results in a more polar, less lipid-soluble, and more readily excreted metabolite. Pharmacologically active metabolites that depend on renal excretion for elimination from the body may be formed; such is the case with *N*-acetylprocainamide, the pharmacologically active metabolite of procainamide, an antiarrhythmic agent [*see Table 2*]. Toxic metabolites that depend on renal excretion may also be formed. For example, meperidine, a commonly prescribed narcotic, is metabolized to normeperidine, which undergoes renal excretion. Although this metabolite has little narcotic effect, it lowers the seizure threshold when it accumulates in uremic patients.

Stepwise Approach to Dosage Adjustment

The following five steps provide a framework for dosage adjustment in patients with renal failure [*see Figure 2*]. It must be emphasized that this stepwise approach is only a starting point from which dosage adjustment must be closely monitored and modified on a patient-by-patient basis.

STEP 1: ELICIT HISTORY AND PERFORM PHYSICAL EXAMINATION

Eliciting the patient's history and performing a physical examination are the first steps in determining the need for dosage adjustment in any patient. Renal dysfunction should be defined as acute or chronic, and if possible, the cause of the dysfunction should be ascertained. Whether the patient has a history of prior drug intolerance, allergy, or nephrotoxicity should also be determined. The physician should review the medications, both prescription and nonprescription, that the patient is taking to identify potential nephrotoxins.

At this stage, the patient's ideal body weight (IBW) should be calculated using the patient's height: for women, IBW is 45.5 kg plus 2.3 kg for every 2.5 cm over 152 cm; for men, IBW is 50 kg plus 2.3 kg for every 2.5 cm over 152 cm. The patient's volume status should be assessed because the V_D of a drug can be altered by marked shifts in extracellular fluid volume. Finally, whether the patient has hepatic dysfunction in addition to renal impairment should be determined, because concomitant hepatic dysfunction may necessitate even greater dosage adjustments.

STEP 2: ASSESS RENAL FUNCTION

Because drug elimination by the kidney correlates with the glomerular filtration rate (GFR), it is logical to use this measurement to gauge dosage adjustment. Creatinine clearance (C_{Cr}) is traditionally used to approximate the GFR because blood urea nitrogen (BUN) and serum creatinine (S_{Cr}) values are, at best, crude markers of renal function. The Cockcroft-Gault formula,[1] which includes the variables of age (in years), IBW (in kilograms), and S_{Cr} (in milligrams per deciliter), should be used to calculate the C_{Cr} (in milliliters per minute):

$$C_{Cr} = \frac{(140 - \text{age}) \times \text{IBW}}{72 \times S_{Cr}}$$

In women, the result should be multiplied by 0.85. It is important to remember that the C_{Cr} overestimates glomerular filtration. For purposes of dosage adjustment, patients with acute renal failure should be assumed to have a C_{Cr} of less than 10 ml/min.

Iohexol is a new marker of the GFR that is currently used in both research and clinical settings to accurately and efficiently quantitate renal function without exposing the patient to radiolabeled material.[2,3]

Recently, a new formula to predict renal function was reported.[4] The Modification of Diet in Renal Disease (MDRD) Study prediction equation estimates GFR rather than creatinine clearance and as such may provide a more accurate estimate of renal function. Factors contributing to greater accuracy of this equation include use of a creatinine assay (the kinetic alkaline picrate reaction) that is least subject to artifactual interference, equation validation in a cohort of patients that differed from the cohort used to derive the equation, prediction of GFR over a wide range of values, inclusion of variables for ethnicity as well as the serum albumin concentration, and no reliance on timed urine collections. The MDRD Study prediction equation can be expressed as follows:

$$\text{GFR} = 170 \times [\text{Scr}]^{-0.999} \times [\text{Age}]^{-0.176} \times [0.762 \text{ if female}] \times [1.180 \text{ if patient is black}] \times [\text{BUN}]^{-0.170} \times [\text{Albumin}]^{-0.318}$$

Whether this equation will be routinely implemented into clinical laboratory reports remains to be seen

STEP 3: DETERMINE LOADING DOSE

In patients with normal renal function, steady-state drug concentrations are achieved after approximately 3.3 half-lives. In patients with renal failure, a drug's half-life may be greatly prolonged, and if no loading dose is given, achievement of steady-state levels and therapeutic efficacy may be significantly delayed. In general, the standard loading dose that is given to patients with normal renal function should also be given to those with renal insufficiency so that therapeutic drug levels are reached rapidly. Digoxin is the major exception to this rule: only 50% to 75% of the usual loading dose of digoxin should be administered to patients with renal failure, because this drug has a marked reduction in its V_D in this setting.

The loading dose can be calculated by the following formula, in which the V_D is in liters per kilogram, IBW is in kilograms, and [Cp] is the desired plasma concentration in milligrams per liter:

$$\text{Loading dose} = V_D \times \text{IBW} \times [\text{Cp}]$$

Table 2 Drugs with Active or Toxic Metabolites Excreted by the Kidney

Acebutolol	Imipramine
Allopurinol	Meperidine
Cephalosporins	Methyldopa
Chlorpropamide	Nitrofurantoin
Clofibrate	Prednisone
Daunorubicin	Procainamide
Diazepam	Propoxyphene
Digoxin	Rifampin
Doxorubicin	Sodium nitroprusside
Enalapril	Succinylcholine
Flurazepam	Sulfonamides

STEP 4: DETERMINE MAINTENANCE DOSE

Two methods can be used to adjust the maintenance dose in patients with renal insufficiency. One method involves lengthening the dosing interval to correspond to the degree of renal impairment according to the following formula:

$$\text{Dosing interval} = \frac{\text{normal } C_{Cr}}{\text{patient's } C_{Cr}} \times \text{ normal interval}$$

Thus, if the patient's C_{Cr} is approximately 50 ml/min and the normal C_{Cr} is assumed to be approximately 100 to 120 ml/min, the dosing interval for that patient would essentially be doubled.

Alternatively, the dose can be reduced in proportion to the patient's degree of renal insufficiency while the dosing interval remains unchanged, as shown in the following formula:

$$\text{Dose} = \frac{\text{patient's } C_{Cr}}{\text{normal } C_{Cr}} \times \text{normal dose}$$

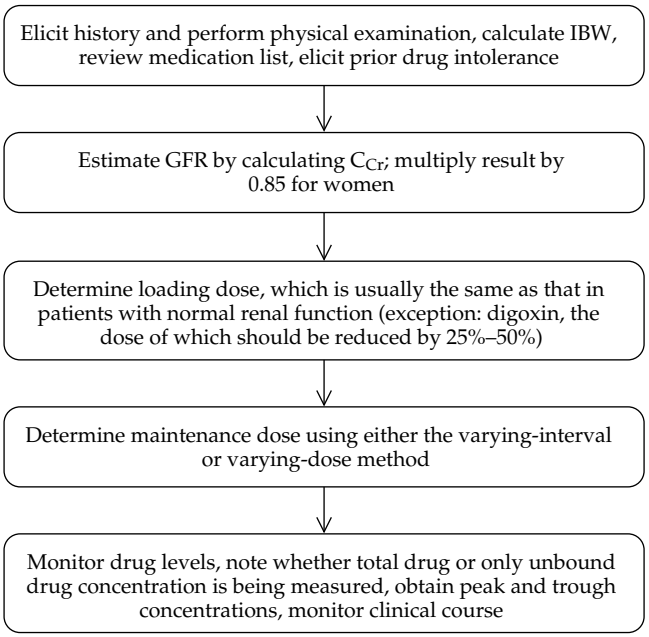


Figure 2 **A stepwise approach to dosage adjustment in patients with renal failure. (IBW—ideal body weight; GFR—glomerular filtration rate; C_{Cr}—creatinine clearance)**

In this case, assuming a C_{Cr} of 50 ml/min for the patient and a normal C_{Cr} of 100 ml/min would result in halving the standard dose but administering it at the same interval used for patients with normal renal function.

The first, or varying-interval, method may provide adequate peak concentrations of a drug while putting the patient at risk for periods of subtherapeutic levels. The second, or varying-dose, method may provide more constant drug levels but puts the patient at risk for toxicity caused by higher trough concentrations.

STEP 5: MONITOR DRUG LEVELS

Varying the dose or dosing interval for a given therapeutic agent may not be sufficient to prevent toxicity and guarantee therapeutic efficacy in patients with renal failure. Monitoring of drug levels can greatly enhance patient care, particularly in the setting of renal disease. Interpreting drug concentrations requires a knowledge of the exact dose given, the route of administration used, the time that has elapsed since the last dose was given, and the drug's elimination half-life.

Monitoring of drug levels may be performed after an appropriate loading dose or three or four maintenance doses have been administered to ensure that steady-state concentrations have been reached. Peak levels reflect the highest drug concentration achieved after an initial rapid distribution phase and tend to correlate with drug efficacy. Trough levels are generally obtained immediately before the next dose is given to ascertain the lowest concentration of drug in the body and, thus, systemic clearance. Drug trough levels are generally used as markers of toxicity.

Although monitoring of drug levels may help ensure therapeutic efficacy and safety, this service is expensive and is not always available. In addition, for some classes of drugs, such as the aminoglycoside antibiotics, monitoring of drug levels has not reduced the incidence of toxicity because of the poor correlation between serum levels and the uptake of drug into certain tissues, such as the renal cortex and inner ear. Interpretation of drug levels must be paired with ongoing clinical assessment of the patient, because toxicity can occur when drug levels are within the so-called therapeutic range. For example, digoxin intoxication can occur despite therapeutic serum levels if the patient has hypokalemia or metabolic alkalosis. Similarly, an increase in the unbound, biologically active fraction of a given pharmacologic agent may not be evident on monitoring of drug levels, because most assays measure total drug concentrations (i.e., both protein-bound and protein-unbound fractions).

Drug Clearance in Dialysis

Patients who are receiving dialysis therapy require special attention with regard to the scheduling of drug administration and may need supplemental dosing of agents that are substantially cleared by dialysis. In general, whenever possible, scheduled doses should be given after completion of a dialysis treatment. If such scheduling is not feasible and the dialytic therapy is known to increase total body clearance by 30% or more, supplemental dosing may be indicated.

The molecular weight, extent of protein binding, and water solubility of a compound are the primary determinants of dialyzability (e.g., small [< 500 daltons], hydrophilic, unbound drugs are readily cleared). The other drug properties that affect clearance by dialysis are the drug's ionic charge, V_D, nonrenal

Table 3 Drug Interference with Renal Function Testing

	Drug	*Test and Mechanism of Interference*
Increased Serum Creatinine	Ascorbic acid	Elevates total chromogen
	Levodopa Methyldopa	Interfere with autoanalyzer method
	Aspirin Cimetidine Trimethoprim	Block tubular secretion of creatinine
	Cefotetan Cefoxitin	Interfere with Jaffé reaction
Increased BUN	Acetaminophen Aminophylline Ascorbic acid Salicylates	Interfere with nonenzymatic method
	Methyldopa	Interferes with phosphotungstate method
	Levodopa	Interferes with autoanalyzer method

BUN—blood urea nitrogen

excretion, and erythrocyte partitioning. Properties of the dialysate (e.g., flow rate, temperature, solute composition, pH, and, in patients who are receiving peritoneal dialysis, volume) and of the dialyzer membrane (e.g., blood-flow rate, pore size, and surface area) also affect drug clearance.

Continuous renal replacement therapies (CRRTs), such as continuous venovenous hemofiltration and continuous arteriovenous hemofiltration, are increasingly being used in medical and surgical intensive care units and achieve C_{Cr} rates of as high as 20 to 30 ml/min. During CRRT, solutes and pharmacologic agents are removed by convective transport. Unless they are bound to protein, solutes and drugs dissolved in plasma water cross the dialysis membrane via plasma water ultrafiltration, in which the ultrafiltrate drug concentration equals the plasma concentration multiplied by the percentage of unbound drug. Pharmacologic agents may also be removed excessively from the body by membrane-drug binding. Because few data exist regarding drug removal and dosage adjustment during CRRT, close monitoring of the patient's clinical status, as well as drug levels, when available, is necessary.

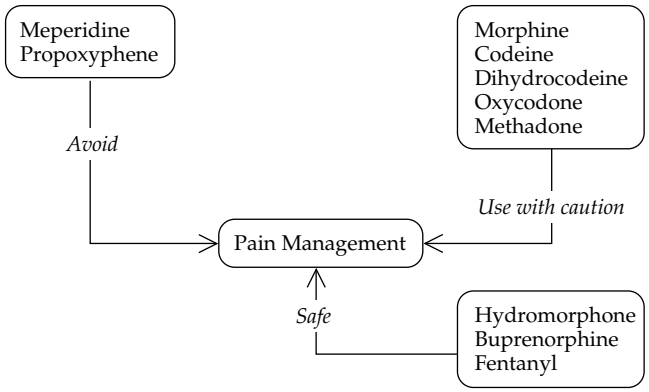


Figure 3 **Flowchart showing the relative safety of narcotic therapy for pain management in end-stage renal disease patients.**

Specific Pharmacologic Considerations

A number of pharmacologic agents can lead to aberrant results on renal function testing. These results can cause unnecessary concern and lead to costly evaluation if the clinician is unfamiliar with the causes of spurious elevations in the S_{Cr} level, the BUN level, or both.

There are several mechanisms by which various compounds interfere with laboratory markers of renal function [*see Table 3*]. Drugs can cause a rise in the S_{Cr} concentration either by interfering with the chromogen assay or by inhibiting renal tubular secretion of creatinine. Clinicians may need to discontinue such agents, at least temporarily, when S_{Cr} values rise unexpectedly. A number of drugs can also cause a rise in the BUN value by interfering with the various available assay methods.

Many patients with reduced renal function may be capable of maintaining metabolic balance as long as certain medications are avoided, because the limited nephron mass of these patients may be unable to meet the excretory demands of the agents. Some drugs may carry so-called metabolic loads of excess acid, alkali, magnesium, potassium, or sodium that overwhelm the failing kidney's excretory capacity. Other drugs may induce metabolic loads by increasing creatinine production, urea production, or both (e.g., glucocorticoids and androgens) or may impair the patient's ability to excrete free water (e.g., nonsteroidal anti-inflammatory drugs).

Thus, responses to pharmacologic therapy by patients with renal insufficiency are complex and heterogeneous and require an understanding of basic pharmacologic principles. Furthermore, such patients require close monitoring, not only of drug levels but also of their clinical course. Although reductions in the GFR should be factored in to dosage adjustment strategies, such a calculation is merely the initial step in considering the many pharmacokinetic and pharmacodynamic parameters involved in pharmacotherapy for patients with impaired renal function.

NARCOTIC USE IN PAIN MANAGEMENT

Although many options exist for pain management with narcotics, limitations in these options arise in the dialysis population. It should be noted that few, if any, data exist in the published literature regarding the use of narcotics in patients with reduced renal function.

It is possible to categorize narcotic use in renal failure on the basis of relative safety recommendations [*see Figure 3*]. Administration of medications such as meperidine, in which multiple dosing should be avoided, may still be an option as one-time therapy for procedure-related analgesia. Normeperidine, a metabolite of meperidine, lowers the seizure threshold of the central nervous system. It is eliminated by the kidneys and may accumulate in patients with reduced renal function; close monitoring is necessary, especially when repeated dosing is used. Respiratory depression or increased sedation may indicate accumulation of a parent compound or active metabolite.[4]

References

1. Cockcroft DW, Gault MH: Prediction of creatinine clearance from serum creatinine. Nephron 16:31, 1976

2. Gaspari F, Perico N, Remuzzi G: Application of newer clearance techniques for the determination of glomerular filtration rate. Curr Opin Nephrol Hypertens 7:675, 1998

3. Swan SK, Halstenson CE, Kasiske BL, et al: Determination of residual renal function with iohexol clearance in hemodialysis patients. Kidney Int 49:232, 1996

4. Levey AS, Bosch JP, Lewis JB, et al: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461, 1999

5. Foote EF, Dahl NV: More drug idiosyncracies in end-stage renal disease. Semin Dialysis 11:167, 1998

Recommended Readings

Aronoff GR, Berns JS, Brier ME, et al: Drug Prescribing in Renal Failure: Dosing Guidelines for Adults, 4th ed. American College of Physicians, Philadelphia, 1999

Bennett WM, Henrich WL, Stoff JS: The renal effects of nonsteroidal antiinflammatory drugs: summary and recommendations. Am J Kidney Dis 28:S56, 1996

Golper TA, Marx MA: Drug dosing adjustments during continuous renal replacement therapies. Kidney Int Suppl 66:S165, 1998

Keller F, Bohler J, Czock D, et al: Individualized drug dosage in patients treated with continuous hemofiltration. Kidney Int Suppl 72:S29, 1999

McCormack JP, Carleton B: A simpler approach to pharmacokinetic dosage adjustments. Pharmacotherapy 17:1349, 1997

Swan SK: Diuretic strategies in patients with renal failure. Drugs 48:380, 1994

Acknowledgment

Figures 1 through 3 Marcia Kammerer.

Appendix A Drug Therapy in Renal Disease

[Adapted in part from *Drug Prescribing in Renal Failure: Dosing Guidelines for Adults,* 4th ed., by G. R. Aronoff, J. S. Berns, M. E. Brier, et al. American College of Physicians, Philadelphia, 1999]

The various categories of drugs appear as follows:

	Drug	Pharmacokinetic Parameters				
		Elimination and Metabolism	*Half-life*		*Plasma Protein Binding (%)*	*Volume of Distribution (L/kg)*
			Normal (hr)	*ESRD (hr)*		
Antimicrobial Agents	Aminoglycosides*†					
	Gentamicin	Renal	2	24–60	<5	0.23–0.26
	Netilmicin	Renal	2.7	35–72	<5	0.22–0.26
	Tobramycin	Renal	2.5	27–60	<5	0.22–0.25
	Antifungal drugs Amphotericin B Amphotericin B lipid complex (ABLC) Amphotericin B colloidal dispersion (ABCD) Liposomal amphotericin B (AmBisome)	Nonrenal	24 (15 days)*	Unch	90–95	4
	Fluconazole	Renal	22–36	Unknown	11	0.5–0.6
	Flucytosine	Renal	3–6	75–200	<10	0.6
	Itraconazole	Hepatic	20–25	20–25	99	10
	Ketoconazole	Hepatic	1–3	3	99	0.36
	Antimycobacterial drugs					
	Ethambutol	Renal	4	7–15	10–30	1.6–3.2
	Isoniazid	Hepatic*	2–4 (slow acetylators); 0.5–1.5 (rapid acetylators)	8–17 8–17	4–30 —	0.75 —
	Rifampin	Hepatic	2–5	2–11	60–90	0.9
	Antiviral agents					
	Acyclovir	Renal	2.1–3.8	20	15–30	0.6–0.8
	Famciclovir	Renal	1.5–3	10–22	<25	1.5
	Foscarnet	Renal	3–6	75–100	17	0.3–0.6
	Ganciclovir	Renal	2.5	20–24	?	0.4–0.6
	Indinavir	Hepatic	1.8	—	60	—
	Lamivudine	Renal	5–11	20	36	0.8

ESRD—end-stage renal disease
GFR—glomerular filtration rate
I—interval extension method of dosage adjustment; data units are hours between maintenance doses
D—dose reduction method of dosage adjustment; data units are percentage of usual maintenance dose
H—hemodialysis
P—peritoneal dialysis
Unch—unchanged

Superscript symbols designate specific points under "Toxic Effects and Remarks."

Adjustment for Renal Failure					*Toxic Effects and Remarks*
Method	*GFR (ml/min)*			*Removed by Dialysis*	
	>50	*10–50*	*<10*		
					*Group toxicity: all agents in this group are nephrotoxic and ototoxic; rarely cause respiratory paralysis †Group remarks: need usual loading doses in renal failure patients; blood levels are best guide to therapy; concurrent penicillin administration may result in subtherapeutic blood levels
D* I	60–90 q. 8–12 hr or 100% q. 12–24 hr	30–70 q. 12 hr or 100% q. 24–48 hr	20–30 q. 24 hr or 100% q. 48–72 hr	Yes (H, P)†	Group toxicity; group remarks; absorption of 50% of intraperitoneal dose in 6-hr continuous ambulatory peritoneal dialysis (CAPD) exchange; poor clearance from blood to peritoneum in CAPD *D and I methods are combined †May add 5 mg to each liter of peritoneal dialysate to obtain satisfactory serum levels
D* I	60–90 q. 8–12 hr or 100% q. 12–24 hr	30–70 q. 12 hr or 100% q. 24–48 hr	20–30 q. 24 hr or 100% q. 48–72 hr	Yes (H, P)	Group toxicity; group remarks *D and I methods are combined
D* I	60–90 q. 8–12 hr or 100% q. 12–24 hr	30–70 q. 12 hr or 100% q. 24–48 hr	20–30 q. 24 hr or 100% q. 48–72 hr	Yes (H, P)†	Group toxicity; group remarks *D and I methods are combined †May add 4–5 mg to each liter of peritoneal dialysate to obtain satisfactory serum levels
I	24	24	24–36†	No (H, P)	Nephrotoxic; renal tubular acidosis; hypokalemia; renal failure; nephrogenic diabetes insipidus; toxicity lessened by saline loading *Terminal-phase half-life equals 15 days because of drug movement from a slowly equilibrating compartment †Ineffective for urinary tract infection in ESRD Less nephrotoxicity
D	Unch	Unch	Unch	Yes (H)*	May increase blood cyclosporine levels *Patients on peritoneal dialysis should receive 50–100 mg daily
I	6–12	16	24	Yes (H, P)*	Hepatic dysfunction; marrow suppression more common in azotemic patients *Dose of 20–30 mg/kg needed after hemodialysis
D	Unch	Unch	50		H (100 mg q. 12–24 hr), P (100 mg q. 12–24 hr)
D	Unch	Unch	Unch	No (H, P)	May increase blood cyclosporine levels
I	24	24–48	48	Yes (H)	Decreased visual acuity; peripheral neuritis may mimic uremia
D	Unch	Unch	Unch	Yes (H)	*Genetic variation in hepatic acetylation Dose after hemodialysis; dosing guidelines for slow acetylators; neurotoxicity reduced by pyridoxine supplementation of 100 mg/day
D	Unch	50–100	50	No (H)	May cause acute renal failure (toxic or immunologic), potassium wasting, and other tubular defects
D, I	5 mg/kg q. 8 hr	5 mg/kg q. 12–24 hr	2.5 mg/kg q. 24 hr	Yes (H)	CNS toxicity in patients with renal failure; may cause acute renal failure if injected rapidly intravenously
D, I	100%	q. 12–48 hr	50% q. 48 hr	Yes (H)	—
D	50–100	15–50	Avoid	Yes (H)	Nonoliguric acute renal dysfunction is common; may cause hypokalemia, hypocalcemia, and hypomagnesemia
I	12	12–24	24–48	Yes (H)	Neutropenia may be more common if drug accumulates
I	Unch	12–24	24	No (H, P)	Nephrolithiasis; acute renal failure caused by crystalluria; tubulointerstitial nephritis
D, I	100	50–150 mg q. 24 hr (full 1st dose)	25–50 mg q. 24 hr (50 mg 1st dose)	Yes (H)	—

Appendix A Drug Therapy in Renal Disease (*continued*)

	Drug	Pharmacokinetic Parameters				
		Elimination and Metabolism	Half-life		Plasma Protein Binding (%)	Volume of Distribution (L/kg)
			Normal (hr)	ESRD (hr)		
	Ribavirin	Nonrenal (renal)	27	Prolonged	<10	9–12
	Vidarabine	Renal (hepatic)	1.5	5	20–30	0.7
	Zidovudine (AZT)	Hepatic (renal)	1.0	3	7–38	1.0–1.6
Antimicrobial Agents (continued)	Cephalosporins*					
	Cefaclor	Renal	1	3	25	0.24–0.36
	Cefamandole	Renal	1	6–11	75	0.16–0.25
	Cefazolin	Renal	1.4–2.2	40–70	80	0.13–0.2
	Cefotetan	Renal	3.5	13–25	85	0.15
	Cefoxitin	Renal	0.7	13–22	75	0.2
	Ceftazidime	Renal	1.2	20–30	17	0.28
	Ceftriaxone	Renal (hepatic)	7.8–9.0	12–24	85–95	0.12–0.14
	Cefuroxime	Renal	1.6–2.2	17	33	0.13–1.80
	Cephalothin	Renal (hepatic)	0.5–1	3–18	65	0.26
	Clindamycin	Hepatic	2–4	3–5	60–95	0.6–1.2
	Imipenem	Renal (hepatic)	1	4	13–20	0.24–0.27
	Macrolides					
	Azithromycin	Hepatic	10–60	Unch	10–50	18
	Clarithromycin	Hepatic	2.6–6	15–20	40–70	2–3
	Erythromycin	Hepatic	1.2–2.6	4–6	70–75	0.5*
	Metronidazole*	Hepatic (renal)	6–14	7–21	20	0.6–0.8
	Monobactam					
	Aztreonam	Renal	1.7–2.9	6–8	50–60	0.5–1.0

ESRD—end-stage renal disease
GFR—glomerular filtration rate

I—interval extension method of dosage adjustment; data units are hours between maintenance doses
D—dose reduction method of dosage adjustment; data units are percentage of usual maintenance dose

H—hemodialysis
P—peritoneal dialysis
Unch—unchanged

Superscript symbols designate specific points under "Toxic Effects and Remarks."

Adjustment for Renal Failure					*Toxic Effects and Remarks*
Method	*GFR (ml/min)*			*Removed by Dialysis*	
	>50	*10–50*	*<10*		
D D D, I	Unch 100 200 mg q. 8 hr	50–100 50–100 200 mg q. 8 hr	50 50 100 mg q. 12 hr	No (H) Yes (H) Yes (H); No (P)	Accumulates in red blood cells; half-life prolonged with long-term use 50% of active hypoxanthine metabolite is excreted by the kidney Inactive metabolite accumulates in renal failure; increased clearance with concurrent rifampin therapy; may cause bone marrow suppression
					*Group toxicity: agents in this group may enhance nephrotoxicity if given in combination with aminoglycoside antibiotics, diuretics, and volume depletion; rare allergic interstitial nephritis; well absorbed from peritoneal fluid in CAPD, but transfer from blood to peritoneum is generally poor; some agents may cause bleeding in patients with renal failure from impaired prothrombin biosynthesis
D	Unch	50–100	50	Yes (H, P)	Group toxicity
I	6	6–8	12	Yes (H)	Group toxicity
I	8	12	24–48	Yes (H)	Group toxicity; ineffective for urinary tract infections when GFR <10 ml/min; for CAPD: 0.5 g q. 12 hr
D	Unch	50	25	Yes (H)	Group toxicity
I	8	8–12	24–48	Yes (H)	Group toxicity; may raise serum creatinine levels by interference with autoanalyzer methods
I	8–12	12–24	24–48	Yes (H); No (P)	Group toxicity
I	Unch	12–24	24	Yes (H, P)	Group toxicity; associated with cholelithiasis
I	8	8–12	12–24	Yes (H); No (P)	Group toxicity
I	6	6–8	8–12*	Yes (H); No (P)†	Group toxicity; nephrotoxicity rare when given alone; may spuriously elevate serum creatinine levels *Serum sickness, hematologic abnormalities more likely with decreased GFR †May be added to peritoneal dialysate at a concentration of 125 mg/L to achieve desired serum concentration or give 1 q. 12 hr
D	Unch	Unch	Unch	No (H, P)	Pseudomembranous enterocolitis may cause volume depletion
I	6	6–12	12	Yes (H)	Seizures in ESRD Combine with cilastatin, a dipeptidase inhibitor that has kinetics similar to imipenem, to prevent renal inactivation
D	Unch	Unch	Unch	No (H)	—
D	Unch	50–100	50	?	Increases theophylline serum levels; hepatotoxicity reported in animal models; reversible dose-related ototoxicity
D	Unch	Unch	50–75	No (H, P)	Ototoxicity may occur with high doses in ESRD *Doubled in ESRD
D	Unch	Unch	50	Yes (H); No (P)	Neurotoxic (vestibular); gastrointestinal symptoms may mimic uremia; on rare occasions, drug-induced lupus may occur Active metabolites have long half-life in ESRD *Recommendations pertain to bacterial infections
D	Unch	50–100	25	Yes (H); No (P)	—

Appendix A Drug Therapy in Renal Disease (*continued*)

	Drug	Pharmacokinetic Parameters				
		Elimination and Metabolism	Half-life		Plasma Protein Binding (%)	Volume of Distribution (L/kg)
			Normal (hr)	ESRD (hr)		
Antimicrobial Agents (continued)	Penicillins*					
	Amoxicillin*	Renal	0.9–2.3	5–20	15–25	0.25–0.42
	Ampicillin	Renal (hepatic)	0.8–1.5	7–20	8–20	0.17–0.31
	Dicloxacillin	Renal (hepatic)	0.5–0.9	1.0–1.6	91–98	0.13–0.19
	Nafcillin	Hepatic	0.5	1.2	80–90	0.31–0.38
	Penicillin G	Renal	0.5	6–20	40–60	0.30–0.42
	Piperacillin	Renal	0.8–1.5	3.3–5.1	30	0.18–0.30
	Ticarcillin	Renal	1.0–1.5	16	45–60	0.14–0.21
	Pentamidine	Hepatic	29*	73–120	70	3
	Quinolones					
	Ciprofloxacin	Renal (hepatic)	3–6	6–17	40	2.1
	Levofloxacin	Renal	4–8	76	24–38	1.0–1.5
	Norfloxacin	Hepatic	3.5–6.5	8	14	0.5
	Ofloxacin	Renal	3.5–9.0	24–36	30	2.5
	Trimethoprim-sulfamethoxazole	Trimethoprim Renal (hepatic)	8–15	20–49	30–70	1.0–2.2
		Sulfamethoxazole Renal	9–11*	20–50	40–60†	0.28–0.38
	Vancomycin	Renal	6–8	200–250	10	0.47–1.11

ESRD—end-stage renal disease
GFR—glomerular filtration rate
I—interval extension method of dosage adjustment; data units are hours between maintenance doses
D—dose reduction method of dosage adjustment; data units are percentage of usual maintenance dose
H—hemodialysis
P—peritoneal dialysis
Unch—unchanged

Superscript symbols designate specific points under "Toxic Effects and Remarks."

Adjustment for Renal Failure					*Toxic Effects and Remarks*
Method	*GFR (ml/min)*			*Removed by Dialysis*	
	>50	*10–50*	*<10*		
I	8	8–12	12–24[†]	Yes (H)	*Group toxicity: agents in this group may cause allergic interstitial nephritis; seizures and coagulopathy at high blood levels Group toxicity *Same data apply for amoxicillin esters (bacampicillin, hetacillin, pivampicillin, talampicillin) [†]High doses needed to treat urinary tract infections in ESRD; 250 mg q. 12 hr in CAPD
I	6	6–12	12–24*	Yes (H); No (P)	Group toxicity; contains 3 mEq Na^+/g *High doses needed to treat urinary tract infections in ESRD; 250 mg q. 12 hr in CAPD; adverse reactions more common in renal failure
I	Unch	Unch	Unch	No (H, P)*	Group toxicity *May be added to peritoneal dialysate
D	Unch	Unch	Unch	No (H)	Group toxicity; dose must be reduced in combined hepatic and renal failure to prevent high serum levels and coagulopathy
D	Unch	50–100	25–50	Yes (H); No (P)*	Group toxicity; upper limit of 4–6 million U/day in severe renal failure; false positive urine protein reactions with biuret reagent and sulfosalicylic acid; potassium salt has 1.7 mEq K^+/million U *May be added to peritoneal dialysate
I	4–6	6–8	8	Yes (H)	Group toxicity; contains 1.9 mEq Na^+/g; elimination is dose dependent
D, I	Unch	1–2 g q. 8 hr	1–2 g q. 12 hr	Yes (H)	Group toxicity; remarks same as group remarks for aminoglycosides; contains 5.2 mEq Na^+/g
I	24	24–48	48	No (H, P)	Nephrotoxic *Tissue uptake extensive
D	Unch	50–100	50	Yes (H, P)	Group remarks: poorly absorbed if taken with antacids or phosphate binders; quinolones decrease phenytoin levels, increase theophylline levels
D, I	Unch	250 mg q. 24–48 hr	250 mg q. 48 hr	No (H, P)	
D	Unch	12–24	24	No (H)	
I	Unch	24	24	Yes (H)	
I	12	12–24	24[‡§]	Yes (H); No (P)	*Half-life reduced in alkaline urine [†]Binding to plasma proteins decreased in ESRD [‡]May cause increase in serum creatinine levels in patients with creatinine >2 mg/dl; this may reflect secretory competition with creatinine or a nephrotoxic reaction [§]May achieve adequate urine concentration in patients with low GFR using normal doses; hematologic side effects caused by antifolate action; can cause hyperkalemia
D, I	1 g q. 12–24 hr	1 g q. 24–96 hr	1 g q. 96–168 hr	No (H, P)	Best guide to therapy is serum level before next dose; elimination variable in renal failure; can be used intraperitoneally in CAPD

	Drug	Pharmacokinetic Parameters				
		Elimination and Metabolism	*Half-life*		*Plasma Protein Binding (%)*	*Volume of Distribution (L/kg)*
			Normal (hr)	*ESRD (hr)*		
Cardiovascular and Antihypertensive Agents	Angiotensin-converting enzyme inhibitors*					
	Benazepril	Hepatic* (renal)	21–22	30	>95	0.15
	Captopril	Hepatic (renal)	1.9	21–32	25–30	0.7
	Enalapril	Hepatic* (renal)	11–24	40–60	50–60	?
	Lisinopril	Renal	12–36	36–48	0–10	0.13–0.15
	Angiotensin II receptor antagonist					
	Losartan	Hepatic	3	4–6	30	0.4
	Antiarrhythmic agents*					
	N-Acetylprocainamide	Renal	6–8	42–70	10	1.5–1.7
	Amiodarone	Hepatic	3–100 days	Unch	96	Variable: 1–148
	Bretylium	Renal	6 (p.o.); 13.6 (I.V.)	16–32	6	8
	Lidocaine	Hepatic*	1.2–2.2	1.3–3.0	60–66	1.3–2.2
	Procainamide	Renal* (hepatic 7%–24%)	2.5–4.9	5.3–5.9	15	2.2
	Quinidine	Hepatic*	5.0–7.2	4–14	70–95	2.0–3.5
	Beta blockers*					
	Atenolol	Renal	5–7	15–35	<5	0.7–1.1
	Carvedilol	Hepatic	5–8	5–8	96	1.9
	Esmolol	Nonrenal*	9 min	9 min	55	1.6–3.5
	Labetalol	Hepatic	3–8	3–8	50	5.6
	Metoprolol	Hepatic	2.5–4.5*	2.5–4.5	8	5–6
	Nadolol	Renal	14–24	45	25–30	2
	Pindolol	Hepatic	3–4	3–4	40–57	1.2
	Propranolol	Hepatic*	3.5–6.0	1–6†	90–96	2.8
	Sotalol	Renal	7–15	56	<1	1.3

ESRD—end-stage renal disease
GFR—glomerular filtration rate
I—interval extension method of dosage adjustment; data units are hours between maintenance doses
D—dose reduction method of dosage adjustment; data units are percentage of usual maintenance dose
H—hemodialysis
P—peritoneal dialysis
Unch—unchanged

Superscript symbols designate specific points under "Toxic Effects and Remarks."

Adjustment for Renal Failure					*Toxic Effects and Remarks*
Method	*GFR (ml/min)*			*Removed by Dialysis*	
	>50	*10–50*	*<10*		
					*Group remarks: particularly useful for treatment of malignant or unilateral renovascular hypertension; effect magnified by natriuretic agents or sodium depletion; dry cough in 10%–15%
D	Unch	50–100	25–50	No (H, P)	Group remarks; hyperkalemia
					*Active moiety, benazeprilat, is hepatic metabolite and renally excreted
D, I	100% q. 8–12 hr	50%–100% q. 12–24 hr	50% q. 24 hr	Yes (H); No (P)	Group remarks; blood pressure best guide to dose and interval; proteinuria; nephrotic syndrome; granulocytopenia; hyperkalemia
D	100	50–100	50	Yes (H); No (P)	Group remarks
					*Active moiety, enalaprilat, is formed by hepatic ester hydrolysis and is renally excreted; hyperkalemia
D	100	50–75	25–50	Yes (H); No (P)	Group remarks; hyperkalemia
D	Unch	Unch	Unch	No (H)	Group remarks same as for angiotensin-converting enzyme inhibitors
					*Group remarks: in this group, blood levels best guide to therapy; half-life may be prolonged in heart disease or with reduced hepatic blood flow, or both
D, I	100% q. 6–8 hr	50% q. 8–12 hr	25% q. 12 hr	No (H, P)	Group remarks; drug is active metabolite of procainamide (acetylator phenotype dependent)
D	Unch	Unch	Unch	No (H, P)	Group remarks; thyroid dysfunction; peripheral neuropathy; hepatic dysfunction; pulmonary fibrosis; increases digoxin and cyclosporine levels
D	Unch	25–50*	25	No (H, P)	Group remarks; poorly tolerated when used to control blood pressure
					*No specific data in ESRD
D	Unch	Unch	Unch	No (H, P)	Group remarks: half-life dependent on hepatic blood flow; active metabolite
					*Excretion enhanced in acid urine
I	4	6–12	12–24	Yes (H)†	Group remarks; may induce lupuslike syndrome
					*Renal excretion of active metabolite *N*-acetylprocainamide
					†May be able to treat poisoning with hemofiltration
D	Unch	Unch	75	Yes (H); No (P)†	Group remarks; active metabolite; may cause lupuslike syndrome; increases plasma digoxin and possibly digitoxin
					*Excretion enhanced in acid urine
					†Hemodialysis with low potassium bath may be effective for poisoning
					*Group remarks: blood pressure best guide to dose and interval
D, I	100% q. 24 hr	50% q. 48 hr	25% q. 96 hr	Yes (H); No (P)	Group remarks; significant accumulation in ESRD
D	Unch	Unch	Unch	No (H, P)	Group remarks; half-life prolonged in elderly to as much as 15 hr
D	Unch	Unch	Unch	No (H, P)	Suitable for critical care situations; rapidly hydrolyzed by red blood cell esterases; active metabolite in ESRD
D	Unch	Unch	Unch	No (H, P)	Group remarks
D	Unch	Unch	Unch	Yes (H)	Group remarks
					*Hypotension effect lasts 24 hr
D	Unch	50	25	Yes (H)	Group remarks; significant accumulation in ESRD
D	Unch	Unch	Unch	No (H, P)	Group remarks
D	Unch	Unch	Unch†	No (H)	Group remarks; metabolites may accumulate; metabolites spuriously increase bilirubin by interference with assay
					*Clearance depends on hepatic flow; p.o. dose <30 mg extracted by normal liver; decreased hepatic extraction in ESRD
					†Complex biexponential pharmacokinetics in ESRD; blood levels best guide; less frequent doses needed during chronic therapy
D	Unch	30	15–30	Yes (H)	Group remarks

Appendix A Drug Therapy in Renal Disease (*continued*)

	Drug	Pharmacokinetic Parameters				
		Elimination and Metabolism	*Half-life*		*Plasma Protein Binding (%)*	*Volume of Distribution (L/kg)*
			Normal (hr)	*ESRD (hr)*		
Cardiovascular and Antihypertensive Agents (continued)	Calcium blocking agents*					
	Amlodipine	Hepatic	35–50	50	>95	21
	Diltiazem	Hepatic	2–8	2–8	98	9–10
	Nifedipine	Hepatic	4.0–5.5	5–7	92–98	1.4
	Verapamil	Hepatic	3–7	2.4–4.0	83–93	3–6
	Cardiac glycosides*					
	Digitoxin	Hepatic* (renal)	144–200[†]	210	94[‡]	0.6[§]
	Digoxin	Renal (nonrenal 15%–40%)*	36–44	80–120[†]	20–30	5–8*
Miscellaneous Agents	Antidiabetic agents					
	Chlorpropamide	Renal (hepatic)	24–48	50–200	60–90	0.14
	Glipizide	Hepatic	3–7	?	92–99	0.13–0.16
	Glyburide	Renal	1.5–3	?	99	0.2–0.3
	Insulin	Hepatic (renal)	1–4	Prolonged*	5	0.1–0.15
	Metformin	Renal	1–5	Prolonged	Minimal	1–4
	H_2-receptor blockers					
	Cimetidine	Renal (hepatic)	1.5–2.0	3.5	20	0.8
	Famotidine	Renal	2.5–4.0	12–19	15–22	0.8–1.4
	Ranitidine	Renal (hepatic)	1.5–3.0	6–9	15	1.1–1.9

ESRD—end-stage renal disease
GFR—glomerular filtration rate
I—interval extension method of dosage adjustment; data units are hours between maintenance doses
D—dose reduction method of dosage adjustment; data units are percentage of usual maintenance dose
H—hemodialysis
P—peritoneal dialysis
Unch—unchanged

Superscript symbols designate specific points under "Toxic Effects and Remarks."

Adjustment for Renal Failure					*Toxic Effects and Remarks*
Method	*GFR (ml/min)*			*Removed by Dialysis*	
	>50	*10–50*	*<10*		
					*Group remarks: associated with headache, flushing, and dizziness in patients with renal disease; can increase digoxin and cyclosporine levels
D	Unch	Unch	Unch	No (H, P)	Group remarks
D	Unch	Unch	Unch	No (H)	Group remarks; active metabolites; acute renal dysfunction
D	Unch	Unch	Unch	No (H)	Group remarks; may cause edema; acute renal dysfunction
D	Unch	Unch	50–75	No (H)	Group remarks; active metabolites; acute renal dysfunction
					*Group remarks: agents in this group add to uremic gastrointestinal symptoms; serum level best guide to dosage; equations, nomograms, and computer programs available; usual clinical practice is to reduce size of dose when failure is present; toxicity enhanced by dialysis removal of potassium and magnesium
D	Unch	Unch	50–75	No (H, P)	Group remarks; serum level may rise if quinidine given
					*Converted to digoxin (8%); conversion increased in ESRD
					†Half-life decreased by cholestyramine and in nephrotic patients
					‡Binding decreased by dialysis to 88%
					§Volume of distribution decreased in ESRD
D, I	100% q. 24 hr	25%–75% q. 36 hr	10%–25%‡ q. 48 hr	No (H, P)‡§	Group remarks; radioimmunoassay may overestimate serum levels in ESRD; variable bioavailability; clearance decreased by hypokalemia, spironolactone, quinidine, and verapamil
					*Volume of distribution and total body clearance decreased in ESRD
					†Correlates poorly with GFR; half-life may increase in prerenal azotemia
					‡Serum level 12 hr after dose best guide in ESRD; decrease loading dose by 50% in ESRD
					§Serum level transiently decreased by dialysis; myocardial level decreased by uremia; charcoal hemoperfusion may be useful in massive overdoses
I	24	Avoid*	Avoid*	No (P)	Impairs renal water excretion
					*Prolonged hypoglycemia in azotemic patients
D	Unch	Unch	Unch	?	—
D	Unch	Avoid	Avoid	Avoid	—
D	Unch	75†	50†	?	*Therapeutic half-life significantly longer: 2 hr for regular insulin and 8–12 hr for NPH bovine-porcine insulin after subcutaneous injection
					†Dosage dependent on blood glucose level
D	50–100	25	Avoid*	?	*Lactic acidosis
D	100	50–75	25–50	No (H, P)	May increase serum creatinine level and decrease creatinine clearance (because of inhibition of renal tubular secretion) without change in GFR; can cause confusion in patients with renal or hepatic disease, or both; acute renal failure reported
D	50%	25%	25%	No (H, P)	—
D	Unch	50–75	25–50	Yes (H)	—

X CHRONIC RENAL FAILURE AND DIALYSIS

ERIC P. COHEN, M.D.

Chronic Renal Failure

DEFINITIONS

Chronic renal failure (CRF) is defined as a persistent impairment of kidney function. Clinically, it often involves a progressive loss of kidney function and may result in complete renal failure, necessitating renal replacement therapy (i.e., dialysis or transplantation). When renal replacement therapy is required, the condition is termed end-stage renal disease (ESRD). In this chapter, the term CRF is used to designate an established impairment of renal function that does not require treatment by dialysis or transplantation.

EPIDEMIOLOGY

End-stage Renal Disease

The incidence and prevalence of ESRD in the United States are very well known, because cases of ESRD that are managed with long-term dialysis or kidney transplantation are reported to the Health Care Financing Administration as part of the Medicare entitlement that is granted for ESRD. These reports are the basis for the United States Renal Data System (USRDS), a database that enables ongoing surveillance and analysis of ESRD. The most recent report of the USRDS is available on the Internet (http://www.usrds.org).

ESRD develops in almost 100,000 persons a year in the United States.[1] By comparison, there are approximately 180,000 new cases of lung cancer and more than one million new cases of myocardial infarction a year in the United States.[2,3] Of the 400,000 or more patients in the United States who have ESRD, about 300,000 are receiving long-term dialysis.

Rates of ESRD vary with race [*see Figure* 1]. Both the incidence and the prevalence of ESRD are three to four times higher in African Americans than in whites.[1] ESRD is also more common in Native Americans than in whites. In Native Americans, 60% of new cases of ESRD are the result of type 2 diabetes mellitus, whereas only 30% of new ESRD cases in whites are caused by type 2 diabetes. Also, there is a greater prevalence of both type 2 diabetes mellitus and ESRD in people of Hispanic origin than in whites.[4]

The reasons for the greater prevalence of ESRD in nonwhite populations remain unclear. The greater prevalence of hypertension in African Americans has been cited as a factor. Limited access to medical care has been suggested as a factor in African Americans,[4] but nonwhites in Great Britain also have a higher incidence of ESRD per million population than whites, despite the fact that nonwhites have equal access to medical care in that country.[5]

Chronic Renal Failure

The incidence and prevalence of CRF in the United States are uncertain. The incidence of CRF has recently been calculated at 1,700 per million population in a mixed urban and rural area of England.[6] The prevalence of CRF in the United States has been estimated to be two to three times that of ESRD. The third National Health and Nutrition Examination Survey (NHANES III) showed that one million people in the United States have a serum creatinine level of 2 mg/dl or greater.[7] By way of comparison, the estimated prevalence of CRF is comparable to that of Parkinson disease. It is likely that CRF has a racial distribution similar to that of ESRD.

CRF has long been known to be more prevalent in men than in women. Eight of 10 of Bright's original cases were in men,[8] and in a recent population survey in England, 60% more men than women had a plasma creatinine level greater than 1.6 mg/dl.[6] This gender disparity, which remains incompletely explained, extends to ESRD: data published in 2003 by the USRDS showed that 55% of cases of ESRD are in men.[1]

ETIOLOGY

Diabetes mellitus is the most frequent cause of ESRD [*see Table 1*]. Type 2 diabetes mellitus accounts for about two thirds of diabetes-related cases. Hypertension is the second most frequent cause of ESRD. Renal arterial disease—including both large- and small-vessel disease—appears to be rising in incidence, especially in the elderly.[9] The etiologic picture for CRF is similar to that for ESRD.

PATHOPHYSIOLOGY

The mechanisms by which renal failure progresses in patients with CRF have received ample attention over the past 20 years. Generally, progressive loss of renal function in CRF has been related to arterial hypertension, glomerular hypertension, proteinuria, and other metabolic derangements (e.g., hyperlipidemia).[10]

Both experimentally and clinically, arterial hypertension has been associated with more rapid decline in kidney function.[11] In the Multiple Risk Factor Intervention Trial, kidney function of-

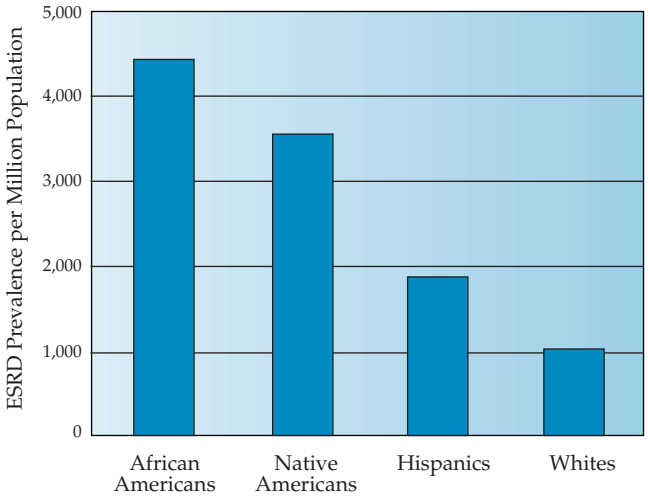


Figure 1 **Prevalence per million population of end-stage renal disease (ESRD) in the United States, according to ethnic group. The overrepresentation of ESRD in nonwhites is obvious.**

Table 1 Prevalence of End-Stage Renal Disease in the United States[1]

Primary Disease	*Number of Patients with End-Stage Renal Disease*
Diabetes	142,963
Hypertension	93,608
Glomerulonephritis	62,908
Cystic kidneys	17,489
Other known cause	53,990
Unknown	15,875

ten stabilized or even improved with a decrease in diastolic blood pressure to less than 95 mm Hg, whereas kidney function tended to decline in patients whose diastolic blood pressure remained above 95 mm Hg.[12] This relation between kidney function and arterial hypertension was also evident in a cohort study of 30,000 kidney transplant recipients, in whom increasing elevations in systolic blood pressure correlated with a progressive increase in late graft failure.[13]

Although animal studies using the remnant kidney model[14] have found that glomerular hypertension—so-called hyperfiltration—plays a role in progressive renal failure, this association has not been confirmed in humans. The proposed mechanism is that excessive filtration pressure within the glomerulus leads to scarring and, ultimately, kidney failure. Such excess pressure may lead to greater protein leakage across the glomerular filtration barrier, although the severity of proteinuria may be a function of the extent of glomerular injury rather than the degree of filtration pressure elevation.

The scarring process may also involve the tubules and interstitium, thereby altering intrarenal function and hemodynamics. Interstitial scarring may also directly impair nephron function as a result of associated peritubular fibrosis and tubular constriction.[15] Sodium excretion may diminish relative to arterial perfusion pressure, which may predispose to hypertension and loss of renal function.[16]

On the cellular level, the presence of activated fibroblasts, known as myofibroblasts, has been associated with loss of renal function in experimental and clinical kidney disease.[17] One mediator that may play a role in renal scarring, perhaps in part through the induction of myofibroblast formation, is transforming growth factor–β_1 (TGF-β_1), a polypeptide that has been implicated in chronic glomerulonephritis, chronic rejection of kidney transplants, and even diabetic nephropathy.[18] This multifunctional molecule can enhance matrix formation and reduce matrix degradation, thus promoting fibrosis. Experimental data suggest that molecular inactivation of TGF-β_1 may be therapeutically useful.

The relationship between proteinuria and loss of renal function has been addressed in both animal and human studies.[19] Proteinuria itself could cause downstream damage to the tubular epithelium; alternatively, some filtered proteins may carry iron, which can induce oxidative membrane damage. Whatever the mechanism, more severe proteinuria is clinically associated with faster loss of kidney function. Stable proteinuria of less than 2 g/day has been associated with little or no tendency to progression, whereas proteinuria of 5 g/day or more portends rapid loss of kidney function.

The early changes seen in diabetic nephropathy deserve special attention. In both type 1 and type 2 diabetes mellitus, microalbuminuria may precede gross proteinuria. Microalbuminuria is defined as the urinary excretion of albumin at a rate of 30 to 300 mg/day; this amount of albumin will not register on a dipstick or be detectable in a 24-hour urine collection. Microalbuminuria may coincide with early histologic changes, such as thickening of glomerular basement membranes and mesangial expansion.[20] If not treated, microalbuminuria tends to progress to overt proteinuria, then to loss of renal function.

DIAGNOSIS AND CLINICAL COURSE

A combination of clinical manifestations and laboratory tests often determine the diagnosis of CRF—in particular, the differentiation of CRF from acute renal failure (ARF) [*see Table 2*]. This differentiation has major implications for prognosis and treatment [*see 10:III Approach to the Patient with Renal Disease*]. Duration of illness is significant: long-standing symptoms point to chronic disease. For the most part, CRF and ARF produce the same symptoms. Itching, however, occurs in CRF but not ARF, so its presence has diagnostic value.

On laboratory testing, a cardinal sign of renal failure is azotemia—an increased concentration of nitrogenous waste products in the blood (i.e., the blood urea nitrogen [BUN] level). In contrast, the term uremia denotes symptomatic renal failure.

The symptoms and signs of chronic renal failure correlate with the degree of renal function impairment [*see Figure 2*]. However, a doubling of the serum creatinine level from 1 mg/dl to 2 mg/dl, which represents as much as a 50% loss in renal function, may result in few clinical manifestations. Proteinuria, especially proteinuria that is in the nephrotic range, predisposes to fluid retention, which, in turn, causes hypertension and edema. Thus, development of edema or hypertension points to possible kidney disease and should prompt urine and blood tests.

With increasing loss of renal function, parenchymal synthesis of calcitriol (1,25-dihydroxyvitamin D_3 [1,25-$(OH)_2D_3$]) decreases and blood concentrations of parathyroid hormone rise. When renal function drops to 50% to 75% of normal, renal concentrating capacity declines, and as a result, nocturia may occur.

Once the glomerular filtration rate (GFR) has declined to 30% or less of normal, signs and symptoms of CRF become more evident [*see Figure 2*]. Anemia is directly related to azotemia and is usually evident once the serum creatinine level exceeds 3 mg/dl. Fatigue and lack of physical endurance may result not only from anemia but also from acidosis. The acidosis associated with CRF is primarily caused by an inability of the kidneys to raise the level of ammoniagenesis (and ammonium excretion) to more than four times the baseline value. Thus,

Table 2 Differential Diagnosis of Chronic and Acute Renal Failure

Diagnostic Feature	*Acute*	*Chronic*
Duration of symptoms	Short	Long
Itching	No	Yes
Anemia	±	Yes
Echogenic kidneys (on ultrasound)	No	Yes
Bone disease (on x-ray)	No	Yes

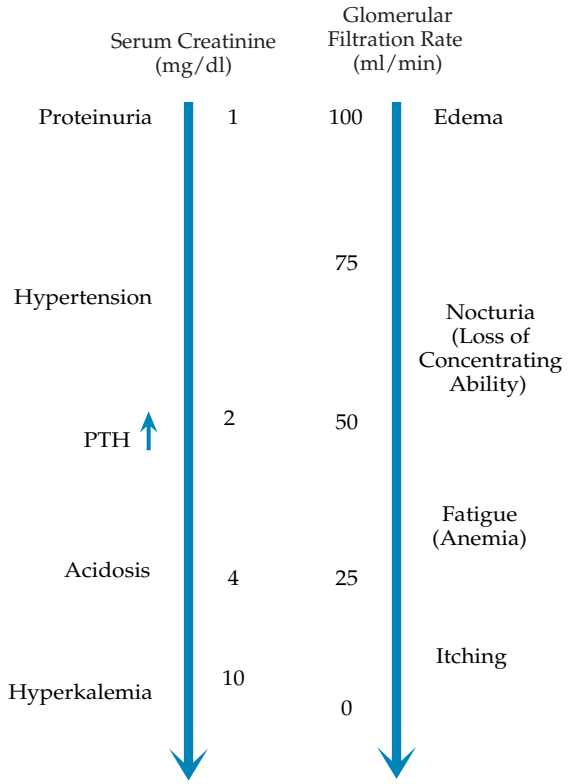


Figure 2 **Relationship between the level of kidney function and common signs and symptoms in chronic renal failure. Signs are shown on the left and symptoms on the right. Progression of chronic renal failure, which is typically linear over time, is displayed as the two downward directed arrows. The glomerular filtration rate (GFR), in ml/min, is linearly related to the reciprocal of the serum creatinine level times 100, and that relation is shown by comparison of the paired serum creatinine and GFR values of the diagram. Proteinuria and edema may occur early in the nephrotic syndrome, when kidney function is still normal, or may be absent in early chronic disease. Hypertension that is secondary to kidney disease generally does not occur until there has been some loss of function. Parathyroid hormone (PTH) levels begin to rise when GFR is 50 ml/min or less. Deficient urinary acid excretion manifests as a systemic acidosis when GFR is approximately 25 ml/min; acidosis occurs because the maximum fourfold compensatory ammoniagenesis per nephron unit can no longer keep up with the daily acid load. The hematocrit level declines linearly with the rising serum creatinine level, but there is a wide scatter in this relation [*see Figure 4*]. Fatigue and other symptoms of anemia do not usually occur until the hematocrit has dropped to 30%; at that level, the GFR is about 25 or 30 ml/min. Itching caused by renal failure occurs with a GFR of 15 ml/min or less and bears some relation to the severity of azotemia. Hyperkalemia may occur as a result of tubular disease, but the fractional excretion of potassium is enhanced as GFR falls; as a generic manifestation of renal failure, hyperkalemia does not usually occur until the very last stages.**

acidosis is the rule when kidney function drops below 25% of normal. Itching is a late symptom of CRF and generally occurs at or about the time a patient needs to begin long-term dialysis treatments.[21]

Hyperkalemia usually occurs late in the course of CRF, when the GFR is 10% of normal or less. However, it can occur when the GFR is 25% to 50% of normal, and in such cases, it may be a side effect of medication. Angiotensin-converting enzyme (ACE) inhibitors and nonsteroidal anti-inflammatory drugs (NSAIDs) have been associated with hyperkalemia.

The evolution of kidney disease from the initial sign or symptom to development of ESRD may take months or years. However, the rate of progression of kidney failure and the clinical course vary with the specific kidney disease. For example, kidney failure is rapid in diabetic nephropathy, intermediate in polycystic kidney disease, and slow in lithium nephropathy. In diabetic nephropathy, the rate of loss of GFR may be as high as 1 to 2 ml/min/mo, whereas in polycystic kidney disease, the loss rate is usually 1 ml/min/mo or less.

Certain accompanying features are specific to particular types of kidney disease. Both retinopathy and neuropathy often accompany diabetic kidney failure. Diabetic neuropathy may cause autonomic disturbances, such as gastroparesis and vomiting, that will add to uremic symptoms. In polycystic kidney disease, berry aneurysms of the circle of Willis may occur. Anemia may be less severe in patients with polycystic kidney disease than it is in patients with CRF from other causes.[22]

Calculation and Tracking of Renal Function

The urinary clearance of endogenous creatinine, measured in milliliters per minute, provides an adequate estimation of GFR [*see Figure 3*]. Because the creatinine in plasma (or serum) derives from muscle, body size must be taken into account; age is also a significant variable. The Cockcroft-Gault formula, which corrects for these factors, remains widely accepted as a way of estimating the creatinine clearance and, correspondingly, the GFR.[23] For men, the formula is as follows:

$$\text{Creatinine clearance (ml/min)} = \frac{(140 - \text{age}) \times \text{weight (kg)}}{72 \times \text{serum creatinine (mg/dl)}}$$

For women, this quotient is multiplied by 0.85, because of their lesser muscle mass. The normal GFR is approximately 95 ± 20 ml/min in women and 125 ± 25 ml/min in men.

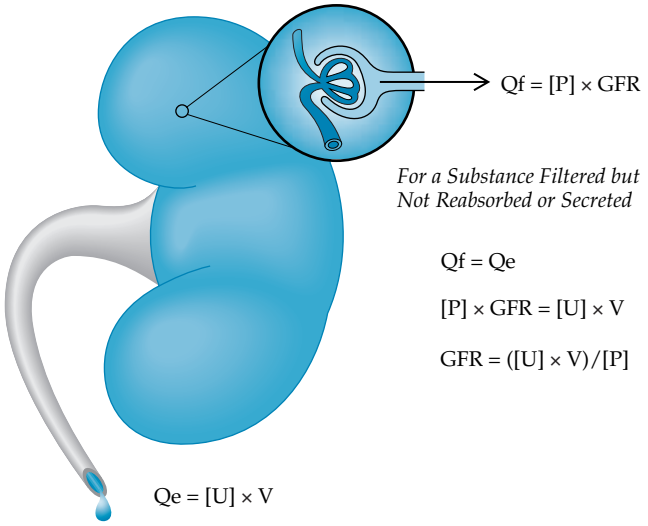


Figure 3 **Schema of the calculation of the GFR. A single kidney is shown. A magnified glomerulus represents filtration, the quantity of a filtered substance (Qf) being equal to its plasma concentration ([P]) multiplied by the GFR. An excreted substance appears in the urine in a quantity (Qe) equal to its urine concentration ([U]) multiplied by the urine volume (V). If a substance is only filtered and is neither reabsorbed nor secreted by the tubules, Qf = Qe; and the GFR may be calculated. Endogenous creatinine is freely filtered, not reabsorbed, and not secreted to a major extent. Its clearance thus approximates the GFR.**

The Cockcroft-Gault formula has been validated in persons older than 65 years,[24] as well as in diabetic patients with kidney disease.[25] The GFR can also be calculated by an equation developed by Levey and coworkers,[26] which utilizes the plasma creatinine, albumin, and urea concentrations, along with patient age and gender.

Measurement of the serum cystatin C concentration has been advocated as a more sensitive way to assess decrements in GFR.[27] The normal serum concentration of this 13 kd protein is about 0.7 mg/L, and the concentration increases when the GFR declines. It is possible that the serum cystatin C concentration may increase before the plasma creatinine concentration does; if so, the cystatin C concentration would be a more sensitive and useful test than measurement of the plasma creatinine concentration. However, a serum cystatin C test costs four times as much as a plasma creatinine test and, therefore, has not entered routine use.

Most types of chronic kidney disease tend to progress and are characterized by gradual loss of function. This loss of function is constant over time[25]—a fact that enables the clinician to quantify the rate of progression in a given case. A simple graph that plots the plasma creatinine level on the y-axis (expressed as 100 divided by the serum creatinine concentration) and the passage of time on the x-axis allows a rapid estimation of the rate at which the GFR declines [*see Figure 4*]. This estimation is possible because of the inverse (hyperbolic) relation of plasma creatinine level to GFR and the corresponding inverse relation of the plasma creatinine level to its reciprocal. Such graphs are easily constructed with spreadsheet computer programs.

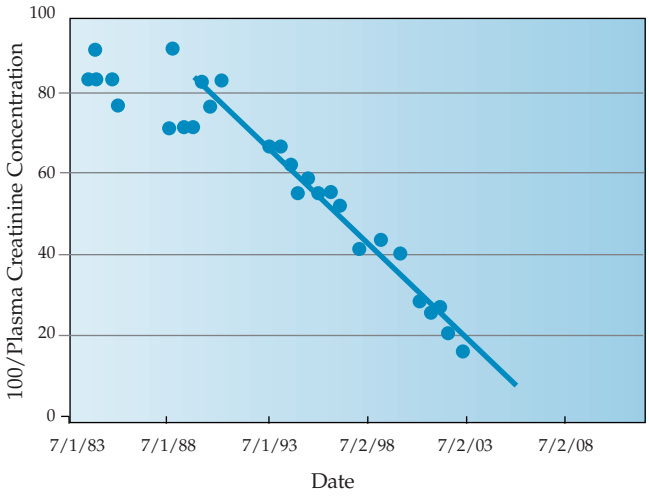


Figure 4 **Deterioration of kidney function over time in a patient with autosomal dominant polycystic kidney disease. This patient began to lose kidney function in or about 1990, when he was 35 years old; from that point onward, the glomerular filtration rate declined by approximately 3.9 ml/min/yr. He developed uremic symptoms in 2003 and underwent kidney transplantation in the fall of that year.**

TREATMENT

The goal of treatment in patients with CRF is to interrupt the progressive loss of kidney function. This goal is now genuinely possible for most glomerular diseases, including diabetic nephropathy.

In the case of type 1 diabetes mellitus, good control of the blood glucose level does diminish the frequency of diabetic complications[28]; however, even with good control, diabetic nephropathy with proteinuria and loss of renal function may still occur. The use of ACE inhibitors, in combination with other antihypertensive drugs, will slow the loss of kidney function in patients with diabetic nephropathy. For example, in a randomized, double-blind study, patients receiving captopril (25 mg p.o., t.i.d.) had a 50% lower risk of death or of ESRD than patients receiving placebo.[29] In this study, the blood pressure levels of patients in both the captopril and the placebo groups were maintained at 140/90 mm Hg or less, which is a mean arterial pressure of 107 mm Hg. There is evidence that in patients with type 1 diabetes mellitus, the greatest slowing of the loss of kidney function is associated with achievement of a mean arterial pressure as low as 90 mm Hg,[30] which corresponds to a blood pressure of 110/80 mm Hg.

In patients with type 2 diabetes and renal impairment, a randomized, prospective study showed that nondihydropyridine calcium channel blockers (diltiazem and verapamil) reduced proteinuria and slowed the progression of renal failure.[31] Calcium channel blockers were more effective than atenolol when blood pressure was maintained at similar levels (approximately 140/90 mm Hg or less). In a randomized, open-label, parallel group trial, verapamil and the ACE inhibitor trandolapril were combined for treatment of nephropathy in patients with type 2 diabetes; the combination of these agents was more effective in reducing proteinuria than either agent used alone.[32] In this study, the average blood pressure was 140/90 mm Hg or less. These and other studies have clearly demonstrated that more than one antihypertensive drug is usually needed to control the blood pressure in diabetic patients.

For early diabetic nephropathy—that is, for patients with microalbuminuria—there are similar promising data. In a placebo-controlled trial in patients with type 2 diabetes who had microalbuminuria, enalapril (10 mg daily) was more effective than placebo in preventing a worsening of albuminuria and in stabilizing renal function.[33] ACE inhibitors are thus recommended for the treatment of diabetic patients with microalbuminuria, as well as treatment of those with overt proteinuria.

The value of ACE inhibitors is not limited to patients with CRF from diabetes. In a 3-year randomized, double-blind study of more than 500 patients with diverse forms of CRF whose serum creatinine levels were between 1.5 and 4.0 mg/dl, the ACE inhibitor benazepril (10 mg/day) was compared with placebo.[34] Benazepril lowered urinary protein levels and blood pressure and slowed the loss of kidney function by about 50%, especially in those patients with type 2 diabetes or chronic glomerulonephritis. Progression of kidney failure was not slowed in patients with polycystic kidney disease. In this study, more than one antihypertensive drug was required for control of blood pressure.

As a group, ACE inhibitors appear to slow the progression of CRF in disease affecting the glomeruli. Some calcium channel blockers also appear to be effective. There is some evidence that maintenance of blood pressure at even less than 140/90 mm Hg may be beneficial. Although treatment with these medications is costly, it is less costly than dialysis or transplantation for ESRD. A cost-effectiveness study of patients with CRF showed a savings of $4,000 per patient after 3 years of use of benazepril versus placebo.[35] However, use of ACE inhibitors—even in investigational settings—appears to reduce the rate of progression of kidney failure by only 50%. The implication is that additional factors, perhaps unrelated to the renin-angiotensin system or its inhibition, may account for loss of function in CRF.

Whereas ACE inhibitors block the synthesis of angiotensin II, the newer angiotensin receptor blockers (ARBs) work by blocking the action of angiotensin I on its receptor. Compared with placebo or the calcium antagonist amlodipine, the ARB irbesartan has been shown to slow the progression of renal failure in patients with type 2 diabetes.[36]

CONSIDERATIONS IN LONG-TERM MANAGEMENT

Several general therapeutic measures can slow the progressive loss of renal function in CRF [*see Table 3*]. Use of nephrotoxins must be avoided—especially NSAIDs, which may impair renal function because of their effects on prostaglandin synthesis. Patient education on this topic is important, because NSAIDs such as ibuprofen and naproxen are available without prescription. The newer cyclooxygenase-2 (COX-2) NSAIDs also reduce kidney function.[37] Limiting dietary protein, controlling hyperphosphatemia, and treating acidosis are additional general measures to be considered for patients with CRF.

Role of Dietary Protein

Dietary protein both contributes to uremic symptoms and promotes the progressive loss of renal function in CRF. A prospective observational study confirmed that patients with CRF spontaneously reduce their intake of dietary protein as they lose renal function.[38] When the GFR is less than 20 ml/min, aversion to meat is not uncommon. At that level of renal function, the spontaneous intake of dietary protein may be 0.8 g/kg/day or lower. Historically, low-protein diets were prescribed to reduce uremic symptoms. Anecdotal evidence suggests that restriction of dietary protein may relieve specific uremic symptoms, such as itching.[39] However, adherence to a low-protein diet is difficult. There is controversy as to whether restricting the intake of daily protein to less than 1 g/kg/day slows the progression of CRF. A randomized study of 64 Australian patients with CRF showed a slowing of renal function loss in patients who consumed a diet containing 0.4 g/kg/day of protein, compared with patients receiving 0.75 g/kg/day of protein.[40] The Modification of Diet in Renal Disease[41] study evaluated 840 patients with moderate or severe CRF (i.e., with a GFR of 25 to 55 ml/min or of 12 to 25 ml/min, respectively). Patients were followed for an average of 2 years or more. When compared with a diet containing 1.3 g/kg/day of protein, a diet containing 0.6 g/kg/day of protein did not appreciably slow the loss of renal function in those patients with moderate CRF. A very low protein diet—0.3 g/kg/day—did not slow the progression of renal failure when compared with a 0.6 g/kg/day protein diet in patients with severe CRF. The average mean arterial pressure in the study patients was less than 100 mm Hg (which corresponds to a blood pressure of 130/80 mm Hg), so it is possible that the benefit from this good control of blood pressure eclipsed the benefit derived from dietary restriction of protein.

Table 3 Evidence-based General Interventions to Slow the Loss of Renal Function in Chronic Renal Failure

Intervention	*Comment*
Control of arterial blood pressure to < 140/90 mm Hg	Lower limit of blood pressure control not established
Use of ACE inhibitors or ARBs	Effective in diabetic nephropathy
Avoidance of nephrotoxins	Over-the-counter nonsteroidal anti-inflammatory drugs may be a culprit in worsening renal function

ACE—angiotensin-converting enzyme ARBs—angiotensin receptor blockers

Control of Hyperphosphatemia

Hyperphosphatemia plays a major role in the development of the secondary hyperparathyroidism seen in CRF. Yi and coworkers[42] showed that a halving of usual dietary phosphorus, independent of calcium or protein intake, reduced secondary hyperparathyroidism in rats with CRF. This effect had been previously shown in a cohort-intervention study in children whose dietary phosphorus was reduced from 1,200 mg/day to 250 mg/day.[43] Reduction of dietary phosphorus may reduce renal osteodystrophy, but long-term compliance with dietary phosphorus restriction is difficult.[44] The use of phosphate binders, such as calcium carbonate or aluminum hydroxide, may reduce the absorption of ingested phosphorus, thus lowering plasma phosphate levels and enhancing the effects of dietary restriction. In children with CRF, phosphate-binder therapy has been associated with a reduction in secondary hyperparathyroidism and improved growth velocity.[45] Dietary phosphorus restriction may slow the loss of renal function in animals, but its effect in humans has not been demonstrated consistently.[46] Nonetheless, the restriction of dietary phosphorus, either by itself or in conjunction with the use of phosphate binders, is advisable for treating or preventing hyperphosphatemia in patients with CRF.

In a review, Coburn and Elangovan[44] summarized the available data regarding use of calcitriol in patients with CRF. Calcitriol, which is the active form of vitamin D, may be deficient in patients with CRF because of reduction in functional kidney parenchyma and, consequently, diminished 1a-hydroxylation of vitamin D. In modest doses (0.25 to 1 mg daily), calcitriol may reduce secondary hyperparathyroidism and improve bone histology. However, incautious use of calcitriol may cause hypercalcemia, which can worsen kidney function. On balance, use of calcitriol should be undertaken only with appropriate monitoring and an awareness of the potential hazards.

Control of Acidosis

Metabolic acidosis contributes to bone mineral loss and is thus a likely factor in the development of the renal osteodystrophy seen in CRF.[47] Metabolic acidosis also may enhance muscle catabolism and decrease albumin synthesis.[48,49] In addition, metabolic acidosis appears to decrease thyroid hormone secretion[50]—an effect that may explain the twofold greater occurrence of hypothyroidism in patients on long-term dialysis than in persons without kidney disease.[51] Correction of acidosis in CRF may raise serum concentrations of calcitriol and lower those of parathyroid hormone.[52] In children with renal tubular acidosis, long-term alkali supplementation improves growth rates.[53] These effects and the effects on bone provide justification for the use of alkali supplements in patients with CRF who have metabolic acidosis. Sodium bicarbonate tablets (650 mg or 7.6 mEq) may be used. Alternatively, baking soda (42 mEq of bicarbonate per teaspoon) may be used. As much as 1 mEq/kg body weight of bicarbonate may be needed daily to maintain the plasma bicarbonate level over 20 mEq/L in patients with CRF.

Treatment of Concomitant Conditions

Anemia With declining renal function, anemia becomes likely. Studies in patients with CRF have shown an inverse correlation of hematocrit with azotemia[54] [*see Figure 5*]. This decline in the hematocrit is largely the result of a reduction in the production of erythropoietin by the kidney. Since 1989, exogenous erythropoietin has been used to treat anemia in patients with renal failure.

Erythropoietin will raise the hematocrit and may improve the appetite and well-being of patients with CRF.[55] This protein hormone must be given by subcutaneous or intravenous injection. The subcutaneous route is preferable, and the initial dosage is 50 to 75 U/kg/wk or, typically, 4,000 units once a week. An erythropoietin with two more carbohydrate side chains, darbepoetin, has been synthesized. It is also effective in treating anemia of CRF, and its typical once-weekly dose is 0.5 μg/kg body weight.

Iron stores must be adequate for erythropoietin therapy to be effective; they may be evaluated by measuring the serum ferritin level or the transferrin saturation. The serum ferritin level should be greater than 100 mg/L; transferrin saturation should be greater than 20%. The goal of treatment is to raise the hematocrit level to at least 30% but, generally, not higher than 40%. The increase in hematocrit may lead to an increase in blood pressure; erythropoietin itself may also contribute to an increase in blood pressure. At one time, there was concern that the use of erythropoietin would accelerate the progression of CRF, but this does not appear to be the case.

Uremic bleeding Patients with CRF may have decreased platelet function and therefore may bleed.[56] The bleeding can range from mild subepidermal bleeding to a life-threatening gastrointestinal hemorrhage. If the platelet count is normal but the bleeding time is prolonged, a defect in platelet function is likely. Such a defect may occur when the BUN level is greater than 60 mg/dl, although diminished platelet function is seen across a considerable range of BUN levels in patients with CRF.[57] Excessive bleeding that is caused by such a defect may be treated with cryoprecipitate or estrogens. Desmopressin (0.3 mg/kg I.V. every 6 hours for up to three doses) is an alternative. Uremic bleeding is probably less frequent today than it was in previous decades, perhaps because long-term dialysis is initiated at lower levels of azotemia than in the past.

Cardiovascular disease There appears to be a surfeit of cardiovascular disease in persons with impaired kidney function.[58] An analysis by the USRDS of patients older than 67 years showed that the incidence of new cardiovascular disease was more than 50% higher in those with reduced kidney function than in those with normal kidney function.[1] Typical risk factors such as hypertension and hyperlipidemia account for some of this overrepresentation, and it is possible that additional factors, such as hyperhomocysteinemia or even an inflammatory state, may play a role.

INDICATIONS FOR INITIATION OF DIALYSIS

Half of patients with a plasma creatinine concentration greater than 6 mg/dl will require dialysis within 3 months.[59] If these patients are not referred to a nephrologist, their conditions often worsen until they require emergent dialysis [*see Figure 6*].[60] Emergent dialysis costs about $50,000 more per patient than nonemergent dialysis. This increased cost stems from the need for emergent access procedures, the greater morbidity at the time dialysis is initiated, and longer hospital stays.

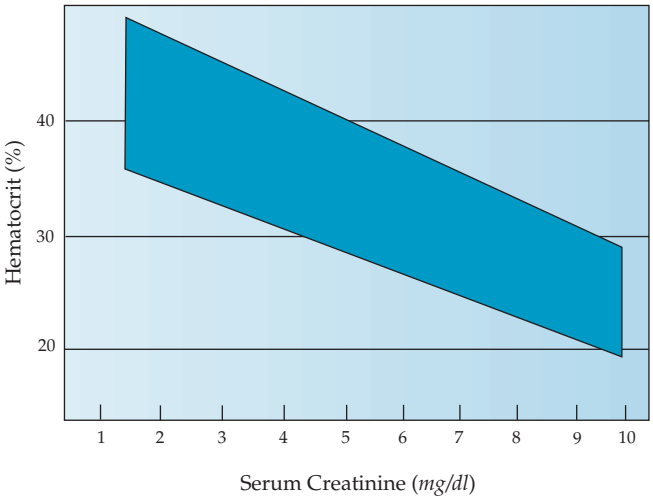


Figure 5 **Relation of hematocrit to serum creatinine in patients with chronic renal failure. Anemia (characterized by a hematocrit of less than 30%) is the rule when the serum creatinine level reaches 10 mg/dl. The polygon encompasses the area of the actual data points.[54]**

Edema, myoclonus, lethargy, and pericarditis are signs that are typically associated with advancing renal failure; typical symptoms include dyspnea, muscle cramps, and itching. On average, three or four signs or symptoms are present when dialysis is required.[61] As of 2002, the average BUN level at the start of dialysis was 85 mg/dl, and the average serum creatinine concentration was 8 mg/dl.[1] It should be emphasized that the serum creatinine level alone does not determine when dialysis should be initiated. Rather, the decision to start dialysis is based on the patient's signs and symptoms. A young, muscular patient may be uremic only with a serum creatinine level of 15 mg/dl, whereas an older diabetic patient may require dialysis when the serum creatinine level is 5 mg/dl. In diabetic patients, in particular, daily creatinine production may be reduced because of coexisting diabetic neuropathy, and the serum creatinine level may be correspondingly lower despite a GFR that is in the uremic range.

Long-term Dialysis

Despite advances in medical care, progressive loss of kidney function still occurs in CRF, with the eventual need for renal replacement therapy. Fewer than 5% of patients who experience ESRD may undergo kidney transplantation without needing long-term dialysis. The great majority of new ESRD patients—over 90,000 persons a year in the United States—require dialysis.

Long-term dialysis—either peritoneal dialysis or hemodialysis—requires access to the circulatory system. For patients receiving peritoneal dialysis therapy, that access is achieved via a peritoneal catheter; for patients receiving hemodialysis, circulatory access is through a native or artificial arteriovenous fistula. None of these forms of access are immediately usable on the day that they are placed. Typically, a fistula is not usable until 1 month after it is placed; an arteriovenous graft may be ready for use 2 weeks after placement. Thus, if a patient has severe CRF and urgently needs to start dialysis, a temporary catheter may be needed. Use of such catheters, which are placed in the large central veins, may result in hemorrhage or infection. Therefore, the need for long-term dialysis should be anticipated and appropri-

ate plans made far enough in advance to avoid temporary catheterization.

Of the 300,000 Americans who were on long-term dialysis in 2001, 275,000 were receiving hemodialysis. About 25,000 were receiving peritoneal dialysis.[1]

PERITONEAL DIALYSIS

Peritoneal dialysis is performed through a plastic catheter permanently implanted into the peritoneal cavity. Typically, 2 or 2.5 L of electrolyte solution is allowed to flow into the peritoneal cavity, where it remains for 4 to 6 hours and is then allowed to flow out. This process, which is repeated throughout the day, allows adequate removal of toxins, salt, and water. Toxin removal occurs by diffusion across the peritoneal membrane. Removal of salt and water occurs by osmosis because of high glucose concentrations in the peritoneal dialysis fluid. Peritoneal dialysis is sometimes performed at night, depending on the characteristics of peritoneal transport. Peritoneal dialysis is sometimes referred to as continuous ambulatory peritoneal dialysis (CAPD), a term that emphasizes the continuous nature of this therapy—in contrast to hemodialysis, which is administered intermittently—as well as the fact that patients on peritoneal dialysis are ambulatory. Patients can undergo peritoneal dialysis while at home or at work. Peritoneal dialysis is less popular than hemodialysis because of the sometimes lesser effectiveness of peritoneal dialysis and because of the risk of peritonitis and its complications.

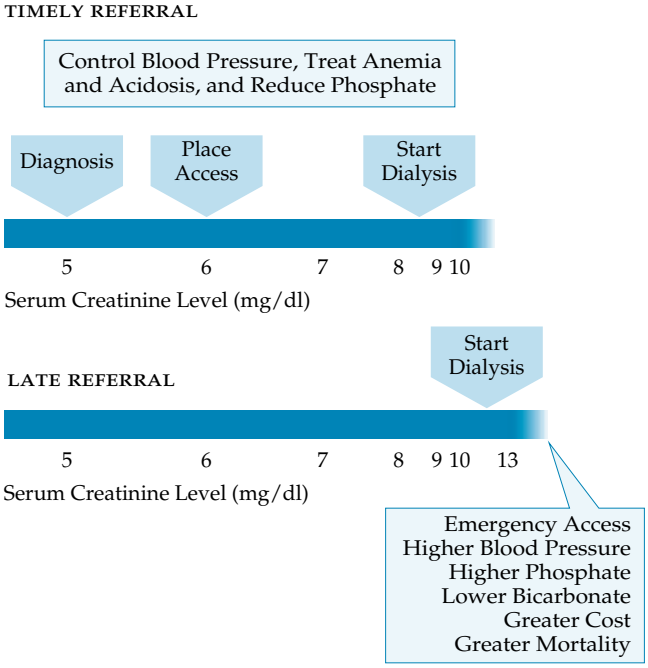


Figure 6 **Schematic comparison of timely versus late referral for dialysis. With timely referral (i.e., referral occurring before the plasma creatinine level has reached 5 mg/dl), it is possible to treat anemia and acidosis and thereby alleviate symptoms. Control of the blood pressure may exert long-term reduction in cardiovascular risk. Access to the circulatory system should be established before such access is actually needed to avoid use of temporary catheters. With late referral, the metabolic derangements at the start of dialysis are more severe, the need for emergency access creates additional risk, and the ultimate financial cost is higher. An emergent start of long-term dialysis is also more psychologically stressful.**

HEMODIALYSIS

Hemodialysis is a technique in which blood is circulated outside the body through a hemodialyzer [*see Figure 7*]. The essential principles of hemodialysis used for treatment of azotemia have not changed greatly since Kolff's pioneering work of over 50 years ago.[62] The hemodialyzer is composed of multiple tubules, each 200 mm in diameter. The tubules are hollow fibers made of semipermeable membranes through which the patient's blood flows. The outer surfaces of the tubules (i.e., the surfaces that are not in contact with the patient's blood) are bathed in a dialysate solution containing sodium (135 to 140 mmol/L), potassium (usually 2 mmol/L), chloride (100 mmol/L), a combination of bicarbonate and acetate (35 to 40 mmol/L), and calcium (1 to 1.5 mmol/L). No urea, creatinine, phosphate, or other electrolyte is present in the dialysate. Thus, diffusion of urea and other waste products takes place across the dialyzer membrane into the dialysate. Molecules larger than 1,000 daltons do not traverse the membrane well. Thus, there is virtually no loss of plasma proteins during dialysis. Uremic toxin removal is enhanced through use of bigger dialyzers, which have greater membrane surface areas. Toxin removal also increases with increasing amounts of blood circulated through the dialyzer. For these reasons, adequate blood flow and sufficient dialysis session time are essential.

Because there is no protein in the dialysate, the calcium that is present in the solution is diffusible. At 1.0 to 1.5 mmol/L, the calcium concentration in the dialysate is actually higher than the usual diffusable plasma calcium concentration of 1.0 to 1.2 mmol/L. Thus, a hemodialysis session does not remove calcium but rather delivers it to the patient. The immediate postdialysis blood calcium concentration may be transiently elevated. It is likely that such a temporary positive calcium balance is helpful in attenuating secondary hyperparathyroidism, perhaps by direct suppression of parathyroid hormone secretion.[63]

In a manner analogous to diffusive calcium delivery, there is delivery of alkali by diffusion from the dialysate to the patient. Diffusive alkali delivery is useful in reducing the metabolic acidosis that is almost always present in CRF. Alkali delivery was formerly achieved by adding acetate to the dialysate; the acetate would be metabolized by the patient's liver, and as a result, bicarbonate would be generated. Acetate may cause hypotension, which may be symptomatic. Modern dialyzers use a buffer system that is predominantly bicarbonate; this buffer system directly delivers bicarbonate to the patient's blood while avoiding the side effects of acetate. As with calcium, the blood bicarbonate concentration may be elevated immediately after a dialysis treatment session. The bicarbonate level will subsequently fall as the bicarbonate is titrated by the fixed acid of protein catabolism.

Salt and water removal is achieved by convective movement of serum water and electrolytes across the dialyzer's semipermeable membrane. The rate of this movement is a function of hydraulic pressure across the semipermeable membrane. This transmembrane pressure is determined by a negative pressure applied to the dialysate. Interdialytic weight gains are typically in the range of 2 to 3 kg. During a dialysis session, this excess fluid must be removed. Modern machines allow easy determination of fluid removal rates, which range from 0.1 to 2 L/hr.

Additional aspects of dialyzer use include the type of semipermeable membrane employed and its biocompatibility, dialyzer cost, and whether the dialyzer may be recycled (reused) for more than one dialysis treatment. Removal of other toxins of uremia, such as β_2-microglobulin—a culprit in dialysis-related

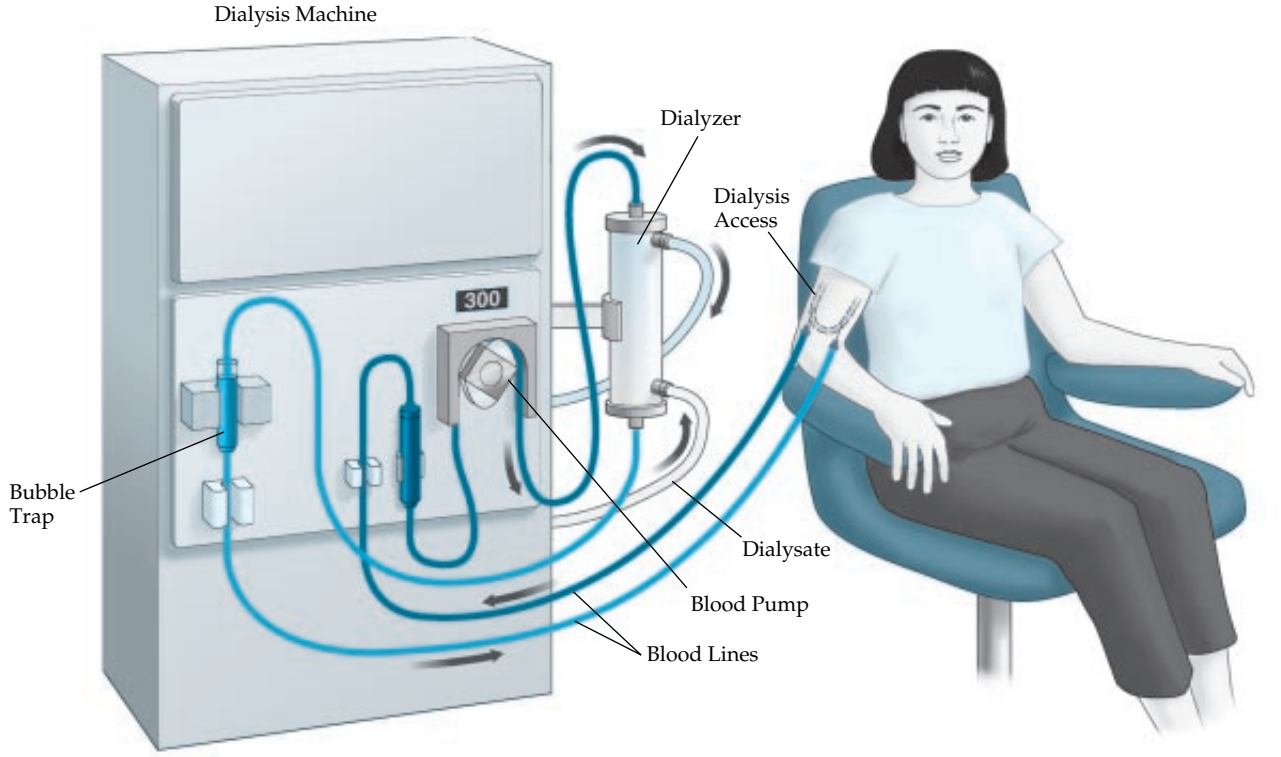


Figure 7 **The process of hemodialysis. The patient has a right arm arteriovenous graft, in which the needles connected to the blood lines are inserted. The blood lines lead to and from the dialyzer, the performance of which is controlled by the dialysis machine. In the dialyzer, blood circulates through hundreds of hollow fibers made of a semipermeable membrane. Dialysate solution flows around the fibers in the opposite direction (counterflow). Toxin removal occurs by passage of toxin from blood to dialysate by diffusion. Fluid removal occurs because of a machine-created hydraulic pressure across the dialyzer semipermeable membrane. An alternative method of accessing the peripheral bloodstream is by use of a cannula. The equipment used for hemodialysis includes reservoirs for purified water and concentrated dialysate solution, a mixing unit that measures conductivity and temperature, an ultrafiltration control unit that maintains dialysate pressure, a flowmeter, a blood pump, the dialysis cartridge (dialyzer), and an air detector.**

amyloidosis—is also related to hemodialyzer use. These issues are beyond the scope of this chapter.

MANAGEMENT AND CARE OF PATIENTS RECEIVING LONG-TERM DIALYSIS

Long-term hemodialysis is performed three times a week, with each session lasting 3 to 5 hours. Some centers are performing daily short dialysis treatments, and these may also be done at home. Because of logistical issues, however, few patients currently use daily dialysis.

Hemodialysis removes toxins, salt, and water from the blood by dialysis. Access to the blood circulation is achieved by use of a central venous catheter or a native or artificial arteriovenous fistula [*see Figure 7*]. Central venous catheters may be required for emergent dialysis, but they should otherwise be avoided. The preferred method for gaining access to the circulation for hemodialysis is through a surgically created arteriovenous fistula, which is often created at the wrist between the radial artery and the cephalic vein or at the elbow between the brachial artery and the basilic vein.[64] Preexisting venous injury, from disease or the use of indwelling catheters, may force the use of artificial arteriovenous fistulas. Artificial fistulas, which are made of polytetrafluoroethylene (PTFE), are prone to clotting, and their use increases the risk of infection. Despite these disadvantages, artificial fistulas are currently the most common way of achieving access to the circulation for the purpose of hemodialysis. The maintenance of adequate blood flow and hemodialysis function requires the ongoing and integrated efforts of nephrologists, surgeons, and interventional radiologists.[65]

Both long-term hemodialysis and long-term peritoneal dialysis usually provide no more than about 10% of the normal kidney function. Thus, persistent uremia, hypertension, and intercurrent illness are often major problems in patients receiving dialysis therapy. The National Kidney Foundation has issued quality-of-care guidelines for patients on long-term dialysis, which can be found at their Web site (http://www.kidney.org).

Adequacy of Dialysis

For patients receiving long-term dialysis, it is critical that enough dialysis be delivered. Cohort-intervention studies and comparative trials have shown that inefficient dialysis leads to buildup of uremic toxins, which is associated with greater morbidity, increased hospitalization, and decreased survival as compared with efficient dialysis.[66,67] Even with newer, more efficient dialyzers, providing sufficient time for each hemodialysis session is vital. Patients beginning hemodialysis therapy may be started at three sessions a week, with each session lasting 4 hours, but even longer session times are associated with the best outcomes.[68] Dialysis efficacy is also directly correlated with outcome in patients on peritoneal dialysis.[69] Measurements of urea or even creatinine in the peritoneal fluid allow quantification of peritoneal dialysis. The quantity of hemodialysis that is deliv-

ered may be measured by the percentage reduction in BUN that is achieved in a particular dialysis session. The BUN reduction should be at least 65%; in diabetic patients, it should be even higher—more than 70%. Changes in urea that result from dialysis may be evaluated mathematically by use of the KT/V quotient; in this expression, K is dialyzer clearance (ml/min), T is time (minutes), and V is the patient's body water volume (ml). Dialyzer clearance measures its ability to remove a substance from the blood. The National Kidney Foundation currently recommends that KT/V be delivered at 1.2 or higher for patients receiving long-term hemodialysis. A multicenter study of prevalent chronic hemodialysis patients has shown no apparent benefit for KT/V greater than 1.3.[70] However, a larger study that included all incident United States hemodialysis patients who started dialysis in 1998 suggested that an even higher KT/V, up to 1.7, was associated with improved survival.[71] Accordingly, patients must be monitored for clinical symptoms reflecting azotemia. Itching, lack of appetite, or insomnia may be evidence of inadequate dialysis, even when the urea reduction ratio seems to be adequate. In patients suspected of being inadequately dialyzed, ongoing clinical monitoring of symptoms is essential. It is worth emphasizing that uremia is more than an elevated BUN. By extension, adequacy of urea removal by dialysis may not fully correspond to adequacy of treatment for ESRD.

Control of Concomitant Conditions

Hypertension Most patients who begin long-term dialysis have hypertension, and hypertension persists in the majority of patients on long-term dialysis. Hypertension has been associated with cardiovascular morbidity and mortality in patients on dialysis.[72] Debate persists, however, as to whether the control of blood pressure has a bearing on long-term outcome in patients with ESRD. One cohort study found no relation of hypertension to 1-year mortality,[73] and another study found that a predialysis systolic blood pressure of less than 140 mm Hg was actually associated with worse outcome than higher predialysis blood pressure.[74] The best long-term survival for patients on long-term dialysis is strongly associated with maintaining the predialysis mean arterial blood pressure at 100 mm Hg or less.[68] This blood pressure is achieved by assiduous control of the volume status through removal of excess extracellular fluid during sufficiently long dialysis sessions. Intervention studies have shown regression of left ventricular hypertrophy by similar methods.[75]

Hyperphosphatemia Hyperphosphatemia promotes secondary hyperparathyroidism in dialysis patients just as it does in predialysis patients with CRF. Its control is achieved by dietary phosphorus restriction, use of phosphate binders, and ongoing dialysis. Cross-sectional data have shown a greater risk of mortality when the serum phosphorus level is greater than 6.5 mg/dl in patients on long-term dialysis, a finding that emphasizes the importance of interventions to limit hyperphosphatemia.[76]

Secondary hyperparathyroidism Virtually all patients on long-term dialysis have secondary hyperparathyroidism. If left untreated, the hyperparathyroidism may result in bone pain, fractures, or soft tissue calcification. Management of this condition includes not only the treatment of hyperphosphatemia but also the use of parenteral vitamin D preparations—mainly calcitriol and paricalcitol, which are given intravenously during dialysis sessions. They act directly on the parathyroid glands to suppress parathormone synthesis and secretion. There is evidence that medical management of secondary hyperparathyroidism has improved over the past 10 years, thus reducing the need for surgical treatment.[77] An additional tool in the management of secondary hyperparathyroidism may be the orally active calcimimetic agent cinacalcet, which acts directly on the calcium receptors of the parathyroid gland cells to suppress parathormone secretion.[78]

Anemia Anemia is managed in the same way as it would be in the predialysis patient with CRF. Thus, a patient's hematocrit may be maintained between 30% and 35% by use of erythropoietin and sufficient iron, without recourse to transfusions. The guidelines for iron metabolism in patients with ESRD and CRF are the same: levels of ferritin should be maintained at 100 mg/L, and transferrin saturation should be maintained above 20%. In dialysis patients who do not respond to erythropoietin therapy, evaluation for anemia may reveal gastrointestinal blood loss, deficiency of folate or vitamin B_{12}, or an inflammatory syndrome that is blunting the response to erythropoietin therapy.

Other concerns Although uremic bleeding is usually not a problem in the well-dialyzed ESRD patient, the use of heparin during each hemodialysis session can predispose a patient to hemorrhagic complications. For example, subdural hematomas are more common in ESRD patients than in the general population and may be fatal in ESRD patients.[56] The use of phosphate binders that contain aluminum may lead to aluminum intoxication, although this has become less of a problem since the availability of phosphate binders that contain calcium instead of aluminum and the availability of the noncalcium phosphate binder sevelamer (RenaGel).[79] Use of medication—either prescription or nonprescription—requires special care in patients with kidney disease, regardless of whether they are on dialysis. Nonsteroidal arthritis medicines such as ibuprofen can worsen kidney function or cause gastrointestinal bleeding and anemia. The normeperidine metabolite of meperidine may accumulate in patients with CRF or ESRD and cause seizures [*see 10:IX Pharmacologic Approach to Renal Insufficiency*].

Infection ESRD patients are also at risk for infection, which is second only to cardiovascular disease as a cause of death in these patients. A recent longitudinal cohort study of ESRD patients identified low serum albumin levels and use of devices such as central venous catheters or artificial arteriovenous fistulas as major risk factors for septicemia.[80] Thus, nutritional status requires ongoing attention, and the use of temporary dialysis catheters should be avoided. Early referral to a nephrologist—well before the time when dialysis is needed—may reduce the severity of these problems by enabling improved predialysis nutrition and timely placement of permanent dialysis access.

Most deaths caused by infection are related to colonization of devices used to gain temporary access to the circulatory system, such as temporary dialysis catheters; pulmonary and intra-abdominal infections may also occur.[81] Bacteremia is a constant concern; it occurs with an incidence of about 10 episodes per 1,000 patient-months. Thus, in a dialysis unit of 100 patients, one episode a month would be expected. Over 50% of cases of bacteremia in dialysis patients are related to the dialysis access. Of the devices for gaining circulatory access, indwelling catheters carry the most risk, PTFE arteriovenous grafts carry somewhat less risk, and native arteriovenous fistulas are least likely to become infected. *Staphylococcus aureus* and *S. epidermidis* are the

most commonly identified infectious agents; both may cause septicemia and complications, such as endocarditis and septic arthritis. Empirical treatment of either bacteremia or an infection at the site of circulatory access thus requires use of an agent that is effective against staphylococci. Of course, removal of the catheter or PTFE graft may be needed. The results of cultures and sensitivity testing will guide antibiotic therapy. Vancomycin (1.0 to 1.5 g I.V. every 5 to 10 days as guided by blood levels) is often used. There is some general concern that overuse of vancomycin may predispose to overgrowth of vancomycin-resistant enterococci, but vancomycin is the drug of choice for infections caused by methicillin-resistant *S. aureus*.[82] The long-acting cephalosporin cefepime (1 to 2 g I.V. after each dialysis session for the duration of antibiotic therapy) may be useful because it has good activity against staphylococci and *Pseudomonas* and *Enterobacter*. Aminoglycosides, such as gentamicin and tobramycin, may be needed for some gram-negative infections, but if they are used for more than 1 to 2 weeks, there is substantial risk of ototoxicity or vestibular toxicity.

Advisability of Screening Examinations

Some investigators have advised that screening mammography or sigmoidoscopy be routinely used in patients receiving long-term dialysis.[83] A careful cost-to-benefit analysis has invalidated that advice.[84] Such screening examinations are unwarranted, largely because the life expectancy of patients with ESRD is very much less than that of healthy persons of similar age. The greater mortality in ESRD, which mostly results from cardiovascular disease or infections, eliminates the benefit of screening programs. Chertow and coworkers[84] point out that the money spent on screening tests would be far better spent on improving the quality of the dialysis.

COST OF ESRD

Dialysis, whether it be hemodialysis or peritoneal dialysis, involves substantial cost. Dialyzers, blood lines, and peritoneal dialysis fluid are ongoing requirements. Moreover, the average long-term dialysis patient is hospitalized once or twice a year and spends an average of 15 days in the hospital annually[1]; these hospital stays contribute substantially to the overall cost. The annual cost per patient can be $50,000 or more.[1] Although the initial cost of kidney transplantation is high, after 3 years, the cost of care of a patient who has received a successful kidney transplant is less than the cost of long-term dialysis. By comparison, the average annual cost of care of a person with multiple sclerosis has been estimated to be $30,000 to $50,000.[85]

PROGNOSIS IN ESRD

Survival of patients on long-term dialysis is distinctly less than that of the overall population.[1] On average, the mortality for patients on long-term dialysis is 20% a year. A person 50 years of age who starts long-term dialysis may expect to live only 5 to 7 more years. The main causes of death in these patients are cardiovascular disease and infections; mortality also is inversely related to adequacy of dialysis. These causes of death emphasize the importance of a better understanding of cardiovascular and infectious illness in patients with ESRD and of the provision of optimal long-term dialysis for those who need it.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. U.S. Renal Data System: USRDS 2003 Annual Data Report, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, April 2003

2. Greenlee RT, Hill-Harmon MB, Murray T, et al: Cancer statistics, 2001. CA Cancer J Clin 51:15, 2001

3. Rosamond WD, Chambless LE, Folsom AR, et al: Trends in the incidence of myocardial infarction and in mortality due to coronary heart disease, 1987 to 1994. N Engl J Med 339:861, 1998

4. Feldman HI, Klag MJ, Chiapella AP, et al: End-stage renal disease in US minority groups. Am J Kidney Dis 19:397, 1992

5. Trehan A, Winterbottom J, Lane B, et al: End-stage renal disease in Indo-Asians in the Northwest of England. QJM 96:499, 2003

6. Drey N, Roderick P, Mullee M, et al: A population-based study of the incidence and outcomes of diagnosed chronic kidney disease. Am J Kidney Dis 42:677, 2003

7. Jones CA, McQuillan GM, Kusek JW, et al: Serum creatinine levels in the US population: third National Health and Nutrition Examination Survey. Am J Kidney Dis 32:992, 1998

8. Bright R: Cases and observations illustrative of renal disease. Guys Hospital Report 1:338, 1836

9. Gomez Campdera FJ, Luno J, Garcia de Vinuesa S, et al: Renal vascular disease in the elderly. Kidney Int Suppl 68:S73, 1998

10. Klahr S, Schreiner G, Ichikawa I: The progression of renal disease. N Engl J Med 318:1657, 1988

11. Raij L, Azar S, Keane W: Mesangial immune injury, hypertension, and progressive glomerular damage in Dahl rats. Kidney Int 26:137, 1984

12. Walker WG, Neaton JD, Cutler JA, et al: Renal function change in hypertensive members of the multiple risk factor intervention trial. JAMA 268:3085, 1992

13. Opelz G, Wujciak T, Ritz E, et al: Association of chronic kidney graft failure with recipient blood pressure. Kidney Int 53:217, 1998

14. Anderson S, Rennke HG, Brenner BM: Therapeutic advantage of converting enzyme inhibitors in arresting progressive renal disease associated with systemic hypertension. J Clin Invest 77:1993, 1986

15. Cohen EP, Robbins MEC, Whitehouse E, et al: Stenosis of the tubular neck: a possible mechanism for progressive renal failure. J Lab Clin Med 129:567, 1997

16. Johnson RJ, Schreiner GF: Hypothesis: the role of acquired tubulointerstitial disease in the pathogenesis of salt-dependent hypertension. Kidney Int 52:1169, 1997

17. Hewitson TD, Becker GJ: Interstitial myofibroblasts in IgA glomerulonephritis. Am J Nephrol 15:111, 1995

18. Border WA, Noble NA: TGF-b in kidney fibrosis: a target for gene therapy. Kidney Int 51:1388, 1997

19. Remuzzi G, Ruggenenti P, Benigni A: Understanding the nature of renal disease progression. Kidney Int 51:2, 1997

20. Chavers BM, Bilious RW, Ellis EN, et al: Glomerular lesions and urinary albumin excretion in type 1 diabetes without overt proteinuria. N Engl J Med 320:966, 1989

21. Cohen EP: Le prurit urémique. Néphrologie 14:215, 1993

22. McGonigle RJ, Husserl F, Wallin JD, et al: Hemodialysis and continuous ambulatory peritoneal dialysis effects on erythropoiesis in renal failure. Kidney Int 25:430, 1984

23. Cockcroft DW, Gault MH: Prediction of creatinine clearance from serum creatinine. Nephron 16:31, 1976

24. Nicoll SR, Sainsbury R, Bailey RR, et al: Assessment of creatinine clearance in healthy subjects over 65 years of age. Nephron 59:621, 1991

25. Cohen EP, Lemann J: The role of the laboratory in evaluation of kidney function. Clin Chem 37:785, 1991

26. Levey AS, Bosch JP, Lewis JB, et al: A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 130:461, 1999

27. Coll E, Botey A, Alvarez L, et al: Serum cystatin C as a new marker for noninvasive estimation of glomerular filtration rate and as a marker for early renal impairment. Am J Kidney Dis 36:29, 2000

28. The effect of intensive treatment on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control and Complications Trial Research Group. N Engl J Med 329:977, 1993

29. Lewis EJ, Hunsicker LG, Bain RP, et al: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. N Engl J Med 329:1456, 1993

30. Hovind P, Rossing P, Tarnow L, et al: Progression of diabetic nephropathy. Kidney Int 59:702, 2001

31. Bakris GL, Copley JB, Vicknair N, et al: Calcium channel blockers versus other antihypertensive therapies on progression of NIDDM associated nephropathy. Kidney Int 50:1641, 1996

32. Bakris GL, Weir MR, DeQuattro V, et al: Effects of an ACE-inhibitor/calcium antagonist combination on proteinuria in diabetic nephropathy. Kidney Int 54:1283, 1998

33. Ravid M, Savin H, Jutrin I, et al: Long-term stabilizing effect of angiotensin-converting enzyme inhibition on plasma creatinine and on proteinuria in normotensive type II diabetic patients. Ann Intern Med 118:577, 1993

34. Maschio G, Alberti D, Janin G, et al: Effect of the angiotensin-converting-enzyme inhibitor benazepril on the progression of chronic renal insufficiency. N Engl J Med 334:939, 1996

35. Van Hout BA, Simeon GP, McDonnell J, et al: Economic evaluation of benazepril in

chronic renal insufficiency. Kidney Int Suppl 63:S159, 1997

36. Lewis EJ, Hunsicker LG, Clarke WR, et al: Renoprotective effect of the angiotensin receptor antagonist irbesartan in patients with nephropathy due to type 2 diabetes. N Engl J Med 345:851, 2001

37. Swan SK, Rudy DW, Lasseter KC, et al: Effect of cyclooxygenase-2 inhibition on renal function in elderly persons receiving a low-salt diet: a randomized controlled trial. Ann Intern Med 133:1, 2000

38. Ikizler TA, Greene JH, Wingard RL, et al: Spontaneous dietary protein intake during progressive renal failure. J Am Soc Nephrol 6:1386, 1995

39. Boulton-Jones JM, Sissons JG, Harrison ER: Itching in renal failure (letter). Lancet 1:355, 1974

40. Ihle BU, Becker GJ, Whitworth JA, et al: The effect of protein restriction on the progression of renal insufficiency. N Engl J Med 321:1773, 1989

41. Klahr S, Levey AS, Beck GJ, et al: The effects of dietary protein restriction and blood pressure control on the progression of chronic renal disease. N Engl J Med 330:877, 1994

42. Yi H, Fukagawa M, Yamato H, et al: Prevention of enhanced parathyroid hormone secretion, synthesis, and hyperplasia by mild dietary phosphorus restriction in early chronic renal failure in rats: possible direct role of phosphorus. Nephron 70:242, 1995

43. Portale AA, Booth BE, Halloran BP, et al: Effect of dietary phosphorus on circulating concentrations of 1,25-dihydroxyvitamin D and immunoreactive parathyroid hormone in children with moderate renal insufficiency. J Clin Invest 73:1580, 1984

44. Coburn JW, Elangovan L: Prevention of metabolic bone disease in the pre-end-stage renal disease setting. J Am Soc Nephrol 9(suppl):S71, 1998

45. Mak RH, Turner C, Thompson T, et al: Suppression of secondary hyperparathyroidism in children with chronic renal failure by high dose phosphate binders: calcium carbonate versus aluminum hydroxide. Br Med J 291:623, 1985

46. Combe C, Morel D, de Precigout V, et al: Long-term control of hyperparathyroidism in advanced renal failure by low-phosphorus low-protein diet supplemented with calcium (without changes in plasma calcitriol). Nephron 70:287, 1995

47. Lemann J, Litzow JR, Lennon EJ: The effects of chronic acid loads in normal man: further evidence for the participation of bone mineral in the defense against chronic metabolic acidosis. J Clin Invest 45:1608, 1966

48. Ballmer PE, McNurlan MA, Hulter HN, et al: Chronic metabolic acidosis decreases albumin synthesis and induces negative nitrogen balance in humans. J Clin Invest 95:39, 1995

49. Williams B, Hattersley J, Layward E, et al: Metabolic acidosis and skeletal muscle adaptation to low protein diets in chronic uremia. Kidney Int 40:779, 1991

50. Brungger M, Hulter HN, Krapf R: Effect of chronic metabolic acidosis on thyroid hormone homeostasis in humans. Am J Physiol 272:F648, 1997

51. Kaptein EM, Quion-Verde H, Chooljian CJ, et al: The thyroid in end-stage renal disease. Medicine (Baltimore) 67:187, 1988

52. Lu KC, Shieh SD, Li BL, et al: Rapid correction of metabolic acidosis in chronic renal failure: effect on parathyroid hormone activity. Nephron 67:419, 1994

53. Nash MA, Torrado AD, Greifer I, et al: Renal tubular acidosis in infants and children. J Pediatr 80:738, 1972

54. McGonigle RJ, Wallin JD, Shadduck RK, et al: Erythropoietin deficiency and inhibition of erythropoiesis in renal insufficiency. Kidney Int 25:437, 1984

55. Valderrabano F: Erythropoietin in renal failure. Kidney Int 50:1373, 1996

56. Lohr JW, Schwab SJ: Minimizing hemorrhagic complications in dialysis patients. J Am Soc Nephrol 2:961, 1991

57. Steiner RW, Coggins C, Carvalho ACA: Bleeding time in uremia: a useful test to assess clinical bleeding. Am J Hematol 7:107, 1979

58. Baigent C, Barbury K, Wheeler D: Premature cardiovascular disease in chronic renal failure. Lancet 356:147, 2000

59. Eadington DW: Delayed referral for dialysis. Nephrol Dial Transplant 11:2124, 1996

60. Jungers P, Zingraff J, Albouze G, et al: Late referral to maintenance dialysis: detrimental consequences. Nephrol Dial Transplant 8:1089, 1993

61. Sievers S, Cohen EP: Unpublished data, 1999

62. Kolff WJ: The beginning of the artificial kidney. Artif Organs 17:293, 1993

63. Fernandez E, Borras M, Pais B, et al: Low-calcium dialysate stimulates parathormone secretion and its long-term use worsens secondary hyperparathyroidism. J Am Soc Nephrol 6:132, 1995

64. Adams MB: Hemodialysis for end-stage renal disease: options and strategies. Semin Vasc Surg 10:151, 1997

65. Besarab A, Frinak S, Zasuwa G: Prospective evaluation of vascular access function: the nephrologist's perspective. Semin Dial 9(suppl 1):S21, 1996

66. Hakim RM, Breyer J, Ismail N, et al: Effects of dose of dialysis on morbidity and mortality. Am J Kidney Dis 23:661, 1994

67. Bloembergen WE, Stannard DC, Port FK, et al: Relationship of dose of dialysis and cause-specific mortality. Kidney Int 50:557, 1996

68. Charra B, Laurent G, Chazot C, et al: Hemodialysis trends in time, 1989 to 1998, independent of dose and outcome. Am J Kidney Dis 32(6 suppl 4):S63, 1998

69. Oreopoulos DG: The optimization of continuous ambulatory peritoneal dialysis. Kidney Int 55:1131, 1999

70. Eknoyan G, Beck GJ, Cheung AK, et al: Effect of dialysis dose and membrane flux in maintenance dialysis. N Engl J Med 347:2010, 2002

71. Port FK, Ashby VB, Dhingra RK, et al: Dialysis dose and body mass index are strongly associated with survival in hemodialysis patients. J Am Soc Nephrol 13:1061, 2002

72. Foley RN, Parfrey PS, Harnett JD, et al: Impact of hypertension on cardiomyopathy, morbidity and mortality in end-stage renal disease. Kidney Int 49:1379, 1996

73. Salem MM, Bower J: Hypertension in the hemodialysis population: any relation to one-year survival? Am J Kidney Dis 28:737, 1996

74. Zager PG, Nikolic J, Brown RH, et al: "U" curve association of blood pressure and mortality in hemodialysis patients. Kidney Int 54:561, 1998

75. Ozkahya M, Ok E, Cirit M, et al: Regression of left ventricular hypertrophy in hemodialysis patients by ultrafiltration and reduced salt intake without antihypertensive drugs. Nephrol Dial Transplant 13:1489, 1998

76. Block GA, Hulbert-Shearon TE, Levin NW, et al: Association of serum phosphorus and calcium x phosphate product with mortality risk in chronic hemodialysis patients: a national study. Am J Kidney Dis 31:607, 1998

77. Cohen EP, Moulder JE: Parathyroidectomy in chronic renal failure: has medical care reduced the need for surgery? Nephron 89:271, 2001

78. Quarles LD, Sherrard DJ, Adler S, et al: The calcimimetic AMG 073 as a potential treatment for secondary hyperparathyroidism of end-stage renal disease. J Am Soc Nephrol 14:575, 2003

79. Slatopolsky EA, Burke SK, Dillon MA, et al: RenaGel, a nonabsorbed calcium- and aluminum-free phosphate binder, lowers serum phosphorus and parathyroid hormone. The RenaGel Study Group. Kidney Int 55:299, 1999

80. Powe NR, Jaar B, Furth SL, et al: Septicemia in dialysis patients: incidence, risk factors, and prognosis. Kidney Int 55:1081, 1999

81. Rinehart A, Collins AJ, Keane WF: Host defenses and infectious complications in maintenance hemodialysis patients. Replacement of Renal Function by Dialysis. Jacobs C, Kjellstrand CM, Koch KM, et al, Eds. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996

82. Lowy FD: *Staphylococcus aureus* infections. N Engl J Med 339:520, 1998

83. Mahmood K, Rasgon S, Yeoh HH: Health screening. Semin Dial 7:48, 1994

84. Chertow GM, Paltiel AD, Owen WF, et al: Cost-effectiveness of cancer screening in end-stage renal disease. Arch Intern Med 156:1345, 1996

85. Whetten-Goldstein K, Sloan FA, Goldstein LB, et al: A comprehensive assessment of the cost of multiple sclerosis in the United States. Mult Scler 4:419, 1998

Acknowledgments

Figures 2, 3, and 5 Seward Hung.

Figure 6 Tom Moore.

XI RENAL TRANSPLANTATION

David K. Klassen, M.D.
Matthew R. Weir, M.D., F.A.C.P.

Introduction

The first kidney transplantation to achieve long-term success was performed in Boston in 1954 between identical twin brothers. This case demonstrated the potential for recovery in patients with advanced kidney failure, but transplantation could be done only in highly selected cases utilizing living-donor organs. Successful transplantation utilizing cadaveric kidneys became common in the 1960s, with the development of pharmacologic immunosuppression. A marked expansion of effective care for patients with end-stage renal disease occurred in the 1970s, when Medicare began providing payment for transplantation and long-term dialysis. Successful transplantation remained limited, however, by a continued high incidence of rejection and the complications of immunosuppressive medications such as azathioprine and corticosteroids. The introduction of cyclosporine in 1983 greatly improved the success rate of kidney transplantation and allowed a wider application of this treatment for patients with renal disease. However, the majority of the more than 300,000 patients in the United States with end-stage renal disease continue to be treated with dialysis.

Dialysis Outcomes Versus Transplantation Outcomes

Continued improvement in outcomes has made transplantation the treatment of choice for patients with end-stage renal disease. Unlike dialysis patients, transplantation patients have a documented improvement in quality of life and a comparatively high rate of return to employment.[1]

A study of more than 220,000 patients receiving treatment for end-stage renal disease in the United States[2] showed that regardless of age, race, sex, cause of end-stage renal disease, or geographic region, long-term survival of patients who received a renal transplant was superior to that of patients who either remained on the transplantation waiting list or continued with long-term dialysis. Patients who underwent transplantation had an initial increase in mortality related to the surgical procedure; however, this initial risk was rapidly eclipsed by the improved long-term survival of transplant recipients. The annual mortality for patients on dialysis is 2.6 times greater than that for patients on the waiting list for transplantation, and the annual mortality for patients on the waiting list is 1.7 times greater than that for transplant recipients.[2]

Number of Potential Transplantation Recipients Versus Number of Donors

The disparity between the number of patients on the waiting list to receive transplants and the number of organs that become available each year has increased dramatically over the past decade. Currently, over 58,000 patients are on the waiting list. In 2002, 8,287 cadaveric and 6,236 living-donor renal transplantations were performed; in 1990, 7,062 cadaveric and 1,181 living-donor transplantations were performed. The increase in the number of living-donor transplantations is partially the result of the expanded use of kidneys from living nonrelated donors, such as spouses or other persons with close emotional ties to the transplant recipients.[3,4]

In the United States, cadaveric organs are distributed to potential recipients on the basis of an algorithm that is intended to optimize transplantation outcomes and maintain equity among patients on the waiting list. Patients are assigned a point score on the basis of the degree of human leukocyte antigen (HLA) match with the available kidney, the length of waiting time, and the degree of immunologic presensitization to HLA.[5]

The degree of presensitization to HLA is an important factor in the success of kidney transplantation. A patient with a high degree of presensitization resulting, for example, from a previous blood transfusion or from multiple pregnancies is difficult to match successfully with an organ. The patient is thus given additional consideration in the distribution algorithm. Children are given extra points to limit their time on the waiting list for transplantation. Available cadaveric organs are shared nationally among patients with complete HLA matches. Kidneys for which no complete-HLA-matched recipients are available are distributed on first a local, then a regional, and then a national basis to patients with the highest point totals. Under the present distribution system, the etiology or severity of the patient's kidney disease is not considered.

Expanded-Criteria Donor Kidneys

In an attempt to encourage and facilitate deceased-donor transplantation, in 2002, new policies and procedures were developed to expedite the allocation of kidneys with unfavorable donor characteristics. These expanded-criteria donor (ECD) kidneys are defined as kidneys that have a higher rate of graft loss as a result of a combination of donor factors, including increased age, a history of hypertension, death from cerebrovascular accident, and elevations in the creatinine level at the time of kidney recovery.[6,7] In 1993, only 8% of deceased-donor transplants involved ECD kidneys; in 2003, 16% of deceased-donor transplants were performed with ECD kidneys, and this rate continues to go up. Although allograft survival rates after 1, 3, and 5 years is lower for ECD kidneys than for non-ECD kidneys (81%, 67%, and 51%, respectively, for recipients of ECD kidneys, as compared with 90%, 81%, and 68% for recipients of non-ECD kidneys),[8] most patients and physicians would agree that receiving a kidney offers substantial improvement in quality of life, as compared with remaining on dialysis. The use of ECD kidneys in older patients has become much more common; over 70% of ECD kidneys go to recipients 50 years of age or older. It is uncertain whether this pattern of use is appropriate; early evidence suggests that the risk of graft loss with ECD kidneys is increased in older patients.[9]

Recipient Evaluation

Potential transplant recipients require a careful medical assessment before being added to the waiting list.[6] The goal of this evaluation is to ensure that the transplantation procedure is done as safely as possible and to maximize the long-term survival of the recipient. Most transplantation programs offer transplants to medically appropriate recipients regardless of age. Data show that older transplant recipients have excellent sur-

Table 1 Contraindications to Transplantation

Active or recent malignancy
Active infection
Severe uncorrectable nonrenal disease
Uncontrolled psychosis
Active substance abuse

vival rates after renal transplantation and may in fact have a lower incidence of episodes of acute rejection than younger recipients.[10]

IDENTIFYING RISK FACTORS FOR POSTTRANSPLANTATION MORBIDITY AND MORTALITY

There are relatively few absolute contraindications to renal transplantation. They include active or recent malignancy that shortens life expectancy; active infection; severe uncorrectable nonrenal diseases; uncontrolled psychosis; and active substance abuse [*see Table 1*]. These conditions should be identified in any patient considered a possible candidate for transplantation.

Cardiovascular Disease

The leading cause of death in transplant recipients, as in the general population, is cardiovascular disease. The pretransplantation evaluation should identify risk factors for coronary artery disease, such as hyperlipidemia, hypertension, diabetes mellitus, and a family history of coronary artery disease.[11] Operative morbidity and mortality are increased in patients who have had a myocardial infarction within 6 months before surgery. Congestive heart failure, unstable angina, ventricular arrhythmias, and electrocardiographic abnormalities are also predictive of a higher rate of surgical complications. For a pretransplantation patient, the physical examination and history should focus on careful examination for signs and symptoms suggestive of vascular disease. Evidence of diminished pulses or bruits, as determined by physical examination, should be followed up with more direct assessment of the vascular system. Stress testing is appropriate for patients with documented risk factors for coronary artery disease, particularly for patients with diabetes, who constitute the largest single subgroup of pretransplantation patients. Studies have shown that up to one third of diabetic patients will have significant coronary artery disease,[12] and silent coronary artery disease is more common in these patients. Diabetic patients or other patients found to have significant abnormalities on stress testing should undergo a diagnostic cardiac catheterization. Correction of significant coronary artery lesions by angioplasty or coronary artery bypass grafting should be done before transplantation.

Malignancy

Patients must be evaluated for occult malignancy and to ascertain any history of malignancy. Immunosuppressive therapy may promote the growth of some malignancies, and patients who have recently had malignancies may be at increased risk for recurrence and subsequent decreased life expectancy. In such patients, transplantation is not appropriate. For most patients who have had malignancies, a 2-year period without evidence of recurrence is adequate for showing they are not at high risk for recurrence. Up to 5 years of follow-up may be appropriate for patients who have had breast cancer, colorectal cancer, and melanoma. However, some malignancies require no waiting period, such as in situ bladder and cervical cancer, Dukes stage A colon cancer, Clark level I melanoma, basal cell skin cancer, and in situ lobular breast carcinoma.[13]

Infection

Active or occult infection must be identified and treated before transplantation. Potential sources of infection include dental abscesses, infected dialysis access catheters, urinary tract infections, and tuberculosis. All patients who produce significant amounts of urine should have a urine culture. Serologic studies to document prior viral infection, particularly with herpesviruses, should be done. IgG and IgM titers for herpes simplex virus types 1 and 2, as well as titers for cytomegalovirus (CMV) and Epstein-Barr virus (EBV), are required. All patients must be screened for evidence of HIV and human T cell lymphotropic virus type I (HTLV-I). Patients should also have a chest x-ray, be screened for evidence of varicella and rubella infections, and undergo purified protein derivative (PPD) testing.

HIV-positive recipients In the past, HIV-positive patients with end-stage renal disease were excluded from consideration for transplantation. With the advent of highly active antiretroviral therapy, many of these patients have very prolonged survival on dialysis. Some transplant programs have begun to cautiously offer transplantation to highly selected HIV-positive patients. Patients with AIDS and low $CD4^+$ T cell counts are not at this time considered for transplantation. Some immunosuppressant medications appear to have antiretroviral properties. Initial reports are promising, and a multicenter trial evaluating transplantation in these patients is under way.[14]

Hepatitis C The prevalence of hepatitis C is high in the dialysis patient population. Studies have shown that hepatitis C–positive recipients have an increased risk of liver disease after transplantation.[14,15] However, it is not known whether there is a decrease in recipient survival related to hepatitis C infections. Patients identified as carriers of hepatitis C should undergo a pretransplantation liver biopsy to assess the degree of histologic disease. Patients with evidence of advanced disease are not candidates for kidney transplantation, although they may be considered for combined kidney-liver transplantation.

Hepatitis B Patients should be screened for evidence of hepatitis B infection. Transplantation is not contraindicated for patients who are positive for surface antigens as long as there is no evidence of advanced liver disease; however, patients must be screened for evidence of active viral replication with hepatitis B e antigens or with hepatitis B DNA. Patients who have active replication are at higher risk for disease progression after transplantation and generally do not receive transplants. Patients with hepatitis B or C who show evidence of abnormal liver enzymes should undergo a liver biopsy before transplantation; those found to have advanced liver disease are not candidates for transplantation.

Primary renal disease Patients with end-stage renal disease who have a primary glomerulopathy are at risk for disease recurrence in the transplanted organ. If possible, the glomerulopathy should be identified before transplantation because the recurrence rates vary with specific diseases [*see Table 2*].

As a general rule, primary glomerular diseases recur relative-

Table 2 Primary Renal Disease Recurrence after Transplantation

Disease	*Recurrence Rate (%)*	*Graft Loss with Recurrence (%)*
Membranous nephropathy	3–20	Minimal
IgA nephropathy	40–60	< 10
Focal segmental glomerulosclerosis	20–30	30–40
Type I membranoproliferative glomerulonephritis	20–30	30–40
Type II membranoproliferative glomerulonephritis	80–90	10–20
Glomerular basement membrane disease	5–10	Rare

ly commonly, but graft losses secondary to recurrence are uncommon; renal transplantation is not contraindicated in patients with primary glomerulonephritis.[16] Focal segmental glomerulosclerosis has a recurrence rate of approximately 20% to 30%. Patients at highest risk are those whose original disease was aggressive and rapidly caused the loss of renal function. In this setting, recurrent disease can occur immediately after transplantation, and graft loss may be high. Living-donor transplantation may be relatively contraindicated.

Anti–glomerular basement membrane disease has a very low recurrence rate if transplantation is delayed for 6 months after documentation of negative antibody titers. Membranous nephropathy recurs with a reported incidence of 3% to 20%, but graft loss resulting from recurrence is unusual. Type II membranoproliferative glomerulosclerosis has a recurrence rate as high as 80% to 90%, and 10% to 20% of patients who experience recurrence may have graft loss. IgA nephropathy recurs in approximately 50% of patients, but significant renal impairment is uncommon. Lupus nephritis rarely recurs after transplantation. Patients with low levels of active disease have successfully received transplants.

Hemolytic-uremic syndrome may recur in as many as 20% of patients, but graft loss is uncommon. Cyclosporine, tacrolimus, and sirolimus, which are medications commonly used as primary immunosuppressants, can cause a thrombotic microangiography that resembles hemolytic-uremic syndrome; however, these drugs are not contraindicated and may be used as immunosuppressants in transplant recipients whose primary renal disease was hemolytic-uremic syndrome.

Anti–glomerular basement membrane disease can develop in the allograft of patients with Alport syndrome. The disease results from exposure to glomerular basement membrane antigens that were absent in the native kidney. However, this complication is uncommon and is not a contraindication to transplantation.

Medical Compliance and Substance Abuse

Patients on dialysis who have a history of multiple episodes of medical noncompliance can receive transplants successfully. However, these patients need extremely close follow-up. Patients with a history of drug or alcohol abuse should show evidence of participation in a rehabilitation program and successful avoidance of drugs and alcohol. A substance-abuse screen is useful to document abstinence. The experience of most transplantation programs is that 6 months of successful participation in rehabilitation is sufficient to indicate correction of the underlying problem. Neither psychiatric illness that is controlled with active treatment nor mental retardation is an absolute contraindication to transplantation.

TIMING OF TRANSPLANTATION

Ideally, the transplantation evaluation should be initiated early in the management of chronic renal disease. Preemptive transplantation—that is, transplantation before the initiation of long-term dialysis—has been shown to result in improved patient outcome.[10,17] Patients who have received a living-donor kidney before the initiation of long-term dialysis have shown a substantially reduced risk of allograft failure for the first year and subsequent years after transplantation. In addition, longer times spent on dialysis before undergoing transplantation negatively affect posttransplantation graft and patient survival.[18,19]

The average waiting time for the receipt of a cadaveric transplant continues to increase; in most areas of the United States, the average waiting time is in excess of 3 years. Therefore, preemptive transplantation is done most effectively with the use of living-donor organs. Patients should be made aware of the potential benefits in terms of improved outcome and survival with living-donor organ transplantation and should be given assistance in identifying potential donors.

Living-Donor Evaluation

Currently, almost one half of the organ transplantation procedures done in the United States are living-donor transplantations. The percentage of living-donor organ transplantations has increased substantially in recent years, in part because of the well-documented success of transplantation using kidneys from living donors who are not biologically related to the recipient and because of the widespread use of laparoscopic donor nephrectomy.[20,21] Longitudinal studies have demonstrated conclusively that in healthy individuals, kidney donation is safe and does not result in progressive loss of renal function.[22,23] Over time, donors may experience a slight rise in blood pressure, but the clinical consequences are minimal. The donor surgical procedure has a good record of safety. Studies suggest a mortality of approximately 0.03% and a complication rate of perhaps 2% to 10%. The common postsurgical problems associated with traditional open nephrectomy techniques are hemothorax, pneumonia, pneumothorax, and wound complications. Laparoscopic donor nephrectomy is very well tolerated by donors and results in a shorter hospital stay and a quicker return to work.[24] Renal function of kidneys procured by laparoscopic techniques is comparable to that of kidneys procured by traditional surgical techniques.[25]

MEDICAL SCREENING AND BLOOD TESTING

The preliminary workup of a potential kidney donor should include a general medical screen for obvious contraindications, such as hypertension and diabetes mellitus. Medical screening should be followed by the assessment of the donor and recipient blood groups and a crossmatch between donor and recipient; ABO compatibility is usually required for successful transplantation. Tissue typing is performed to determine the degree of HLA matching between donor and recipient. The degree of HLA matching has an important effect on long-term outcome and may affect donor selection when multiple potential donors

are available. The best long-term outcomes are obtained between two haplotype-matched siblings, although excellent results are possible with lesser degrees of HLA matching.[26] A crossmatch is also performed between the donor and recipient to detect preexisting antibody sensitivity against donor antigens in the recipient. A positive crossmatch generally precludes proceeding with transplantation because, in such circumstances, there is a high risk of hyperacute rejection. ABO incompatibility or a positive crossmatch between donor and recipient previously was considered a contraindication for transplantation; however, new techniques have been developed to allow successful transplantation in some cases of ABO mismatch or positive crossmatch. These techniques generally utilize plasmapheresis to remove preformed recipient antibody and are combined with pretransplant immunosuppression to inhibit antibody formation. Initial results have been good, and these techniques are being more widely applied.

LABORATORY TESTS

If ABO compatibility and a negative crossmatch are documented, evaluation of the prospective donor can proceed. Clinical testing of the potential donor should include a complete blood count, chemistry screen, urinalysis, urine culture, and coagulation studies. A 24-hour urine collection should be done to assess for creatinine clearance and any evidence of protein excretion. Serologic testing should include antibodies to HIV, CMV, EBV, and hepatitis B and C. In addition, a chest x-ray, an ECG, and a PPD skin test should be performed. Contraindications to live donation include impaired kidney function, significant nonorthostatic proteinuria (> 200 mg/24 hr), active malignancy, active substance abuse, severe chronic illness, and pregnancy.

ARTERIAL IMAGING

A careful assessment of arterial anatomy of the kidney donor is essential. In the past, this assessment required formal contrast angiographic studies. Currently, spiral computed tomography or magnetic resonance angiography has shown adequate sensitivity in detecting multiple renal arteries. These techniques are now being utilized at many transplantation centers.

SCREENING FOR INHERITED RENAL DISEASE

On occasion, the etiology of the renal disease in the recipient is unknown. In such circumstances, a detailed history must be obtained from the donor to determine whether there is evidence of familial renal disease.[27] Polycystic kidney disease is the most common hereditary kidney disease encountered. In cases in which the recipient has polycystic kidney disease, related donors must be carefully assessed for evidence of this disease. In patients older than 30 years, assessment is easily accomplished with a high-resolution CT scan for evidence of renal cysts. For patients between 21 and 30 years of age, this approach is somewhat controversial, and genetic studies may sometimes be necessary to find evidence of polycystic kidney disease. Donors with a family history of diabetes should undergo glucose tolerance testing. Patients with abnormal glucose tolerance should be excluded from donation. Patients who are normotensive but have a strong family history of essential hypertension may not be appropriate for donation. In an attempt to expand the donor pool, some transplant programs have begun to use living donors with mild and easily controlled hypertension; however, this issue is controversial, and these situations must be assessed on a case-by-case basis. Donors with a remote history of an isolated renal stone and no evidence of current stones on radiologic screening may be considered for donation. In this situation, the risk of recurrent stone disease is low. Persons with sickle trait are acceptable as donors. In most transplantation centers, there is no upper age limit for consideration of donor nephrectomy.

PSYCHOLOGICAL ASSESSMENT

Donors' motivation should be assessed to ensure that donations are voluntary and that donors have a good understanding of the potential outcomes for both themselves and the recipients. Kidney donation appears to provide psychological benefit for most donors. However, potential donors need to be aware that the outcome for the recipient cannot be guaranteed.

Pretransplantation Patient Management

IMMUNOSUPPRESSION

Clinical immunosuppression is an empirical and imprecise practice to provide prophylaxis against acute rejection without harming the patient or the graft. The reason for the imprecision is the lack of a satisfactory in vivo technique for determining the adequacy of immunosuppressive agents in a given patient. Inadequate immunosuppression results in rejection of the transplanted organ, whereas overimmunosuppression can lead to immunodeficiency, infection, and, possibly, malignancy. Nonimmune toxicity related to individual drugs, either alone or in combination with other drugs, necessitates close patient scrutiny and careful adjustments of not only immunosuppressive drugs but also drugs that can interact with them. Thus, there is a very delicate balance that needs to be achieved, and it must be determined individually for each patient. Although protocols have been developed to streamline immunosuppression strategies and provide a broad approach, these templates are limited because they require constant, careful adjustment. With the release of many new immunosuppressive medications in the past few years, immunosuppression is more complicated than ever. We have many new options, but clinical experience is needed to determine the optimal strategies for individual patients.

Standard immunosuppressive therapy is based on the premise that immediately after transplantation, the immune response is more vigorous and will require more medication for full suppression. In time, relative tolerance occurs and less immunosuppression is needed. However, our ability to determine the appropriate level of immunosuppression during this transition period from initial exposure to long-term relative tolerance remains difficult. Consequently, most immunosuppressive protocols start with high doses in the early posttransplantation period and steadily reduce the doses over time to a level that is believed to be appropriate for a given patient. The final dose is usually administered according to protocol but is adjusted on the basis of the presence or absence of infection, rejection, or nonimmune toxicity.

Immunosuppressive Drugs

The major classes of oral immunosuppressive drugs include corticosteroids, antimetabolites (e.g., azathioprine, mycophenolate mofetil, and mycophenolic acid), calcineurin inhibitors (e.g., cyclosporine and tacrolimus), and TOR (target of rapamycin) inhibitors (sirolimus). In addition to the previously mentioned

oral medications, monoclonal antibody preparations (anti-CD25 antibodies [i.e., daclizumab and basiliximab] and OKT3) and polyclonal antibody preparations (ATGAM [antithymocyte globulin] and RATG [rabbit antithymocyte globulin, also called thymoglobulin]) can be employed to prevent or treat rejection [*see Table 3*].

Corticosteroids Although corticosteroids have been important immunosuppressive drugs for many years, there is no general consensus on how best to use these agents. In the past, with few alternatives available, extremely high doses of corticosteroids were used. The nonimmune toxicity has been well described. Toxic effects include fragile skin, diabetes, hypertension, dyslipidemia, and avascular necrosis. However, corticosteroids also provide important immunosuppressive and anti-inflammatory effects by blocking the expression of several cytokines, the most important of which are interleukin-1 (IL-1) and IL-2. Current protocols employ much lower doses of corticosteroids, and it is unlikely that doses of up to 0.1 mg/kg/day will cause long-term problems. Important new evidence from clinical trials suggests that a rapid taper of steroids within 3 to 7 days after transplantation results in low rejection rates in subgroups of patients felt to be at lower risk for acute rejection (i.e., first-time transplant recipients and white recipients), particularly if immunosuppression has been induced with anti-CD25 antibody or RATG.[28] Thus, newer immunosuppression strategies may allow for near-complete avoidance of steroid use. This strategy may be safer than elimination of steroid use in the late posttransplant period, because long-term steroid use may condition the immune response in such a way as to result in an increased risk of rejection with late cessation. A meta-analysis of immunosuppression withdrawal trials indicated that withdrawal of steroids is more likely to precipitate acute rejection than is removal of calcineurin inhibitors.[29]

Antimetabolites Antimetabolites are an important part of immunosuppressive strategies, largely because they have no demonstrable nephrotoxicity and little effect on blood pressure, cholesterol levels, or glycemic control. Mycophenolate mofetil has generally replaced azathioprine because of its greater potency and specificity for affecting lymphocytes.[30,31] Gastrointestinal toxicity in the form of diarrhea and nausea and bone marrow toxicity are clinical problems that are usually resolved by dosage adjustment.

Azathioprine is a purine analogue that is incorporated into cellular DNA, where it inhibits purine nucleotide synthesis. It also interferes with the synthesis and metabolism of RNA. Mycophenolate mofetil, which is metabolized to mycophenolic acid, is a reversible inhibitor of the enzyme inosine monophosphate dehydrogenase, which is a critical rate-limiting enzyme in the de novo synthesis of purines in lymphocytes. Because lymphocytes rely exclusively on the de novo purine synthesis pathway rather than on the salvage pathway for purine nucleotide production, they are very susceptible to the effects of mycophenolate mofetil. An oral preparation of mycophenolic acid is available that has fewer gastrointestinal side effects than mycophenolate mofetil.

Table 3 Immunosuppressive Drugs

Class	*Drug*
Corticosteroids	Prednisone
Antimetabolites	Azathioprine Mycophenolate mofetil Mycophenolic acid
Calcineurin inhibitors	Cyclosporine Tacrolimus
TOR inhibitor	Sirolimus
Polyclonal antibodies	ATGAM RATG
Monoclonal antibodies	OKT3 Anti-CD25 Basiliximab Daclizumab Anti-CD20 Rituximab Campath-1H

ATGAM—antithymocyte globulin RATG—rabbit antithymocyte globulin
TOR—target of rapamycin

Calcineurin inhibitors Calcineurin inhibitors such as cyclosporine and tacrolimus have been the cornerstone of immunosuppression for over 20 years. However, the major drawback of these drugs is primarily related to their nonimmune toxicity, particularly nephrotoxicity.[32] The toxicity of these drugs is multifactorial and may involve direct and indirect effects. The major effect is intrarenal vasoconstriction leading to intrarenal ischemia and renal tubular cell toxicity. Early in the treatment, toxic effects are reversible if the drugs are discontinued. However, with long-term treatment, effects of the drugs, such as interstitial fibrosis and tubular atrophy, may be irreversible.[33] Recipients of kidneys that have preservation or reperfusion injury are particularly vulnerable to the nephrotoxic effects of calcineurin inhibitors.[34] Long-term use of these drugs may result in progressive loss of renal function, leading to dialysis or retransplantation. Clinical studies have indicated that calcineurin inhibitors can be successfully reduced or withdrawn 1 to 2 years after transplantation in patients with chronic allograft nephropathy who are receiving mycophenolate mofetil and steroids.[35] This practice is associated with stabilization of renal function, improved graft survival, and a low likelihood of acute rejection. In particular, mycophenolate mofetil and sirolimus may prove to be useful, less nephrotoxic substitutes for calcineurin inhibitors and may also provide an opportunity for improving long-term graft function.

Other nonimmune toxicities related to calcineurin inhibitors are largely metabolic. Cyclosporine commonly induces an increase in blood pressure, cholesterol levels, and, occasionally, serum glucose.[36] Tacrolimus, although it affects blood pressure and cholesterol levels to a lesser degree than cyclosporine, has a hyperglycemic effect in predisposed individuals. Concerns about these side effects are justifiable in that they affect both patient and graft survival because of increased risk of atherosclerotic cardiovascular disease and chronic allograft nephropathy. Other side effects related to cyclosporine include hirsutism, tremor, and gingival hypertrophy. Tacrolimus may also cause tremor, alopecia, and headache.

Calcineurin inhibitors primarily interfere with the immune system by binding to a cytoplasmic receptor protein, which in turn binds to an intracellular protein known as calcineurin. This action interferes with the activation of several genes that

are critical for interleukin expression—most importantly, IL-2. This lymphokine is particularly important with regard to the clonal expansion and generation of helper T cells and cytotoxic T cells.

TOR inhibitors Sirolimus is a newer immunosuppressant whose immunosuppressive activity is unique in comparison to the previously described drugs. Although sirolimus binds to a cytoplasmic binding protein similar to the one tacrolimus binds to, the drug–binding protein complex does not attach to calcineurin but, rather, targets the TOR protein and impairs the capacity of previously synthesized cytokines to activate T cells to enter the cell-division cycle.

Clinical trials indicate that sirolimus, the calcineurin inhibitor cyclosporine, and low-dose prednisone provide effective immunosuppression.[37,38] It also may be possible to combine sirolimus with mycophenolate mofetil and steroids.

The principal drug-related toxicity of sirolimus is hypertriglyceridemia and mild thrombocytopenia. There is limited evidence of nephrotoxicity unless sirolimus is combined with cyclosporine or, in some instances, tacrolimus.

Polyclonal and monoclonal antibodies Polyclonal and monoclonal antibodies are useful short-course therapies that may be helpful in providing initial or induction immunosuppression or in treating rejection.

Polyclonal antibodies have been available for many years. The currently available products are ATGAM and RATG. These agents are used primarily for either induction immunosuppression or the treatment of rejection. The primary mode of action is related to the clearance of lymphocytes in the reticuloendothelial system or through lysis after antibodies have become bound to T cells. Allergic reactions are the most common problem. In addition, because these agents are so potent, overimmunosuppression may occur. Monitoring of total lymphocyte counts is necessary to identify the optimal therapeutic dose. These agents are usually given for 1 to 2 weeks on a daily basis.

Monoclonal antibody preparations are a technological advancement over polyclonal antibody preparations. OKT3 antibody was the first monoclonal antibody approved for clinical transplantation.[39] It is specifically directed against the CD3 antigen complex found on all mature human T cells. It is a very effective and potent agent that is administered daily for 7 to 14 days. It is particularly useful for induction immunosuppression or for the treatment of rejection. Its major side effect is an acute cytokine release syndrome, which can be very serious in some patients, necessitating mechanical ventilation. Even though this side effect is uncommon, patients should be monitored very carefully, particularly after the first few doses.

Anti-CD25 monoclonal antibodies have recently been developed.[40] Two such agents, daclizumab and basiliximab, are largely humanized murine monoclonal antibodies against the IL-2 receptor site. These agents have been shown to be very effective in reducing the incidence of acute rejection if they are administered immediately before and shortly after engraftment. Side effects are negligible. Newer antibody therapies that target B cells include two anti-CD20 monoclonal antibodies: rituximab and campath-1H. Both may hold promise for treatment of vascular rejection; they may also help with induction immunosuppression, particularly in presensitized patients, although these drugs are not approved for use in transplantation.[41]

IMMUNOSUPPRESSION STRATEGIES

Induction immunosuppression is intended to minimize the risk of early rejection and to maximize the likelihood of the development of a stable form of tolerance. In addition, induction strategies can be adjusted to provide more immunosuppression for patients at higher risk, such as those undergoing a repeat kidney transplantation, those who are highly sensitized, and those who have a poor crossmatch. Likewise, induction immunosuppression can be adjusted to reduce or eliminate the exposure of calcineurin inhibitors early in the posttransplantation period in patients who have received a kidney that was partially damaged by a preservation injury such as ischemia and a reperfusion injury. The use of anti-CD25 antibodies, OKT3, or ATGAM can intensify immunosuppression and delay the need for calcineurin inhibitors until adequate graft function has been established. This strategy is based on the premise that the use of nephrotoxic immunosuppressants early in the posttransplantation period will slow the rate of renal recovery from the ischemic insult, thereby increasing the likelihood of delayed graft function, the subsequent need for dialysis and kidney biopsies, and the risk of rejection. Consequently, the use of steroids, mycophenolate mofetil, or sirolimus in conjunction with an antibody induction regimen is a popular strategy and has demonstrated important efficacy in the prophylaxis of early rejection.

The purpose of maintenance immunosuppression is to provide long-term prophylaxis against acute rejection while minimizing the risk of immune and nonimmune toxicities. The lack of in vivo or in vitro methods to determine the adequacy of immunosuppression necessitates the use of standardized protocols that are adjusted on a case-by-case basis in the presence of rejection or immune or nonimmune toxicity. Insufficient immunosuppression, which results in rejection, will necessitate intervention with powerful antirejection therapies, such as steroids in high doses or of polyclonal or monoclonal antibody preparations, and increase the risk of complications of overimmunosuppression (e.g., opportunistic infection and malignancy). Excess immunosuppression may result in opportunistic infection.

Conventional practice is to use three oral immunosuppressive agents, most commonly a corticosteroid, a calcineurin inhibitor, and either mycophenolate mofetil or sirolimus. With the availability of more therapeutic choices, fewer than three drugs may be effective; in some patients, lower doses of four drugs may be appropriate. Many centers now use anti-CD25 or RATG, followed by steroids (either low-dose steroids or standard-dose steroids that are eliminated early on) and drugs from two of three drug classes (i.e., a calcineurin inhibitor and either mycophenolate or sirolimus) to maintain immunosuppression. Thus, there is more complexity in the current clinical approach to immunosuppression.

Because immunosuppressive agents are given for a lifetime, careful scrutiny to prevent toxicity is of paramount importance. Such monitoring is particularly appropriate with regard to kidney function, because there is some concern that the inadequacy of immunosuppression can result in a chronic rejection process that leads to vasculopathy and progressive graft dysfunction. There is also concern that the calcineurin inhibitors themselves—despite their effectiveness in preventing rejection—can lead to a progressive interstitial fibrosis and scarring, either directly or indirectly, through their effects on blood pressure, cholesterol levels, and glucose metabolism. These concerns are of particular importance in patients who manifest evidence of progressive graft dysfunction and proteinuria. Adjustments in im-

munosuppression—primarily reduction or withdrawal of calcineurin inhibitors and the substitution of other agents to maintain adequate immunosuppression—are being actively studied as possible solutions to problems caused by calcineurin inhibitors. In addition, meticulous control of blood pressure, cholesterol levels, and glucose metabolism may also prove to be extremely important in reducing the likelihood of progressive graft dysfunction.

The usual treatment of acute rejection consists of high doses of corticosteroids (methylprednisolone, 500 to 1,000 mg/day for 3 days intravenously, followed by an oral steroid taper), the administration of the monoclonal antibody preparation (e.g., OKT3), or the administration of a polyclonal antibody preparation (e.g., RATG or ATGAM). The polyclonal or monoclonal antibody preparation can be given intravenously on a daily basis for 7 to 14 days. Once treatment for acute rejection has been administered, a careful evaluation of long-term maintenance immunosuppression should be performed to adjust the medications and to prevent subsequent rejection. Adjustment may require an increase in drug dose and levels, the substitution of new drugs, or even the addition of new medications.

Table 4 Causes of Renal Transplant Dysfunction

Category	*Causes*
Structural	Extrinsic compression (fluid collection) Ureteral obstruction (stone, scarring, mass) Urethral obstruction (prostate, stricture)
Vascular	Renal artery stenosis
Volume depletion	Cardiac failure Diuretics
Renal parenchymal	Acute rejection Chronic rejection Recurrent glomerulonephritis
Drugs	ACE inhibitors Angiotensin II receptor blockers Antibiotics Calcineurin inhibitors NSAIDs

ACE—angiotensin-converting enzyme NSAIDs—nonsteroidal anti-inflammatory drugs

Posttransplantation Complications

ACUTE AND CHRONIC RENAL TRANSPLANT DYSFUNCTION

Etiology

The causes of acute or chronic renal dysfunction in the long-term renal transplant recipient should be carefully evaluated. Dysfunction can be related to vascular, structural, or renal parenchymal disorders or to volume depletion [*see Table 4*].

Structural disorders The most common cause of structural dysfunction is disruption of the drainage system, usually by ureteral scarring and obstruction or, in the case of males, because of abnormalities of prostatic urethral outflow or stricture. Fluid collections can also cause clinically significant obstruction. A renal sonogram can be very helpful in elucidating the location of a potential obstruction. An appropriate radiologic or urologic intervention can be made to relieve the obstruction.

Vascular disorders Transplantation renal artery stenosis, suggested by the presence of decreased glomerular filtration rate (GFR) and hypertension, may also cause graft dysfunction. Although it is less common than ureteral abnormalities, renal artery stenosis can lead to hypertension associated with graft dysfunction that is more difficult to treat. Radiographic imaging techniques, such as duplex ultrasound and magnetic resonance arteriography, are critical to rule out this process. Renal artery stenosis of the transplanted kidney typically requires percutaneous transluminal angioplasty or surgical correction. Patients with normal GFR and hypertension who are discovered to have renal artery stenosis should be closely observed without intervention.

Volume depletion Volume depletion caused by loop diuretics leads to a reduction in renal function. Likewise, the use of nephrotoxic drugs such as the calcineurin inhibitors, nonsteroidal anti-inflammatory drugs, antibiotics, and other agents must be reviewed carefully. The use of angiotensin II receptor blockers or angiotensin-converting enzyme (ACE) inhibitors may cause transient reductions in renal function that are usually not of clinical significance. However, these changes may be magnified in the presence of excessive diuretic use and volume depletion or transplantation renal artery stenosis. In such cases, either discontinuance of the ACE inhibitor or angiotensin II receptor blocker or a dosage adjustment may be needed.

Renal parenchymal disorders Other major causes of allograft dysfunction can be related to renal parenchymal abnormalities. The three most important areas of concern are recurrent glomerular kidney disease, acute rejection, and chronic allograft nephropathy. A percutaneous allograft biopsy with careful tissue examination by light microscopy, electron microscopy, and immunofluorescence is critical to making the correct diagnosis. Although recurrent kidney disease can occur and, in selected cases, may result in progressive loss of renal function, it is much less likely to occur than acute rejection or chronic allograft nephropathy. Acute rejection occurs rarely, usually only if there has been a major change in immunosuppression. More common is chronic allograft nephropathy, which is a multifactorial process that may involve long-term alloimmune processes, metabolic processes, and direct toxicity of the calcineurin inhibitors.

Early Diagnosis of Dysfunction

Subtle increases in serum creatinine in long-term recipients of kidney transplants are not acceptable and must be explained. A high index of suspicion during early diagnostic efforts is essential to prompt the likelihood of early recognition and management of disease.[42] An ultrasound scan, a urinalysis, and a urine culture, along with a medication-list evaluation, are critical. If necessary, radiographic procedures and percutaneous graft biopsy should be undertaken.

POSTTRANSPLANTATION INFECTIONS

Although the incidence of acute rejection has decreased dramatically over the past decade, problems with infection related to the surgical procedure and immunosuppressive medications remain. Infections occur after transplantation in a relatively predictable pattern [*see Figure 1*]. This pattern is influenced by the degree of immunosuppression.

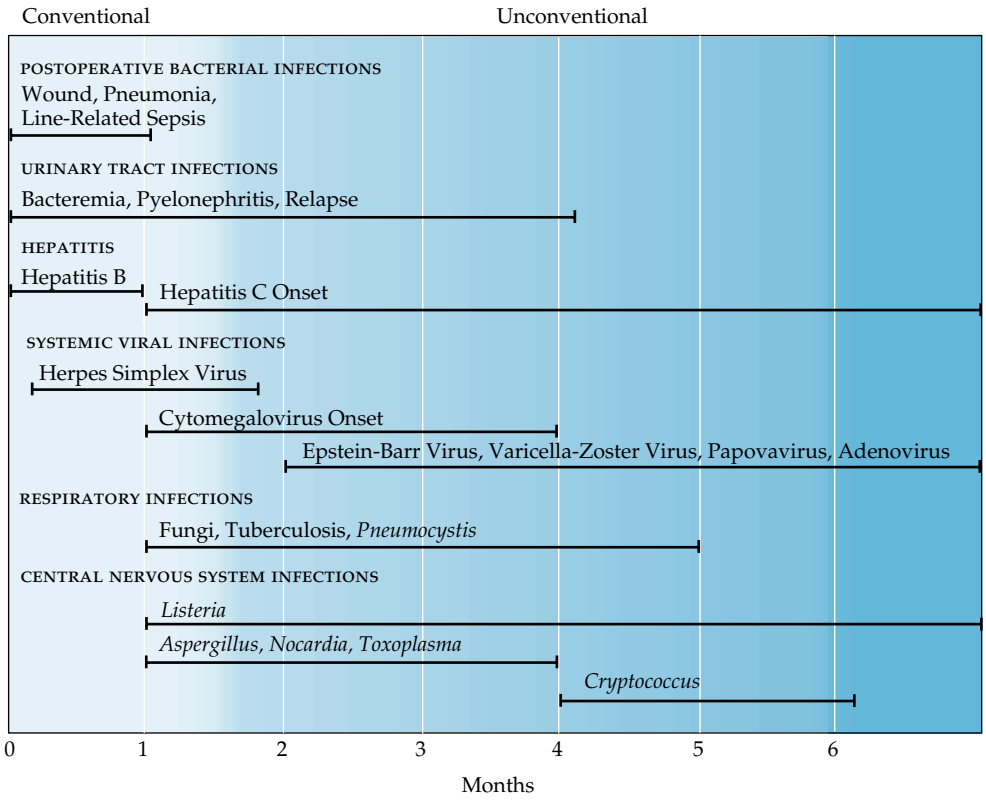


Figure 1 **Timing of conventional and unconventional infections in the renal transplant recipient.**

Fishman and Rubin have grouped the occurrence of posttransplantation infections into three periods.[43] In the first month after transplantation, the most common infections are those resulting from complications of the surgical procedure, such as wound infections, catheter-related urinary tract infections, aspiration pneumonia, and infections of hematomas or lymphoceles. Occasionally, infections may be transmitted by the allograft.

After the first month, the spectrum of infection changes. The high dose of immunosuppressive drugs used early in the transplantation procedure places the patient at risk for a variety of opportunistic bacterial, fungal, and viral infections from such organisms as *Nocardia, Aspergillus, Cryptococcus,* and *Candida*. The activation of latent viral infections from such pathogens as CMV, EBV, and varicella also pose a threat during this period.

Six months after transplantation, most immunosuppressive medications have been reduced to relatively low levels. Community-acquired respiratory infections from pneumococcal, influenzal, and other respiratory pathogens are more common than opportunistic viral or fungal infections.

Cytomegalovirus

CMV is perhaps the most important posttransplantation infection. The risk of CMV infection, which can present as a systemic viral illness, pneumonia, or gastrointestinal disease, is related to the serologic status of the donor and recipient. A primary CMV infection is transmitted by the organ of a seropositive donor to a seronegative recipient. CMV disease will occur in 60% to 75% of such cases and tends to be severe unless effective prophylactic measures are instituted. A reactivation infection occurs when latent viruses are reactivated in a seropositive recipient. In this situation, the disease is generally less severe and occurs in about 20% of such recipients. The decrease in severity of the disease may be related to the preexisting immunity of the recipient. The incidence and severity of CMV disease has been shown to be decreased by prophylactic oral ganciclovir or valganciclovir, which is started at the time of transplantation and continued for approximately 12 weeks. If CMV disease occurs, treatment with high-dose intravenous ganciclovir or valganciclovir with reduction in immunosuppressive medications provides effective control.

Polyomavirus

Most transplant recipients harbor a latent infection with the BK strain of polyomavirus. This virus is acquired in childhood by virtually all individuals and is thought to remain latent in the kidney and genitourinary tract. With chronic immunosuppression, it may reactivate and cause a progressive fibrotic and inflammatory nephropathy that results in gradual loss of renal function.[44] It may also cause ureteral obstruction. This infection results in a clinical picture that resembles chronic rejection or chronic allograft nephropathy. It has been seen with increasing frequency in association with the use of more potent immunosuppressive medications. The incidence is likely related to the overall degree of immunosuppression and not to the use of any specific drug.

Urine cytologic examination has proved to be an effective screening tool for reactivation of polyomavirus. Patients with reactivated polyomavirus will shed identifiable renal tubular cells containing viral inclusion bodies, the so-called decoy cells. Up to 20% of renal transplant recipients have been reported to shed decoy cells in their urine, although a much smaller number of patients actually have active nephropathy. A transplant biopsy is required to confirm the presence of active nephropathy. These biopsies typically show scattered inflammatory cell infiltrates

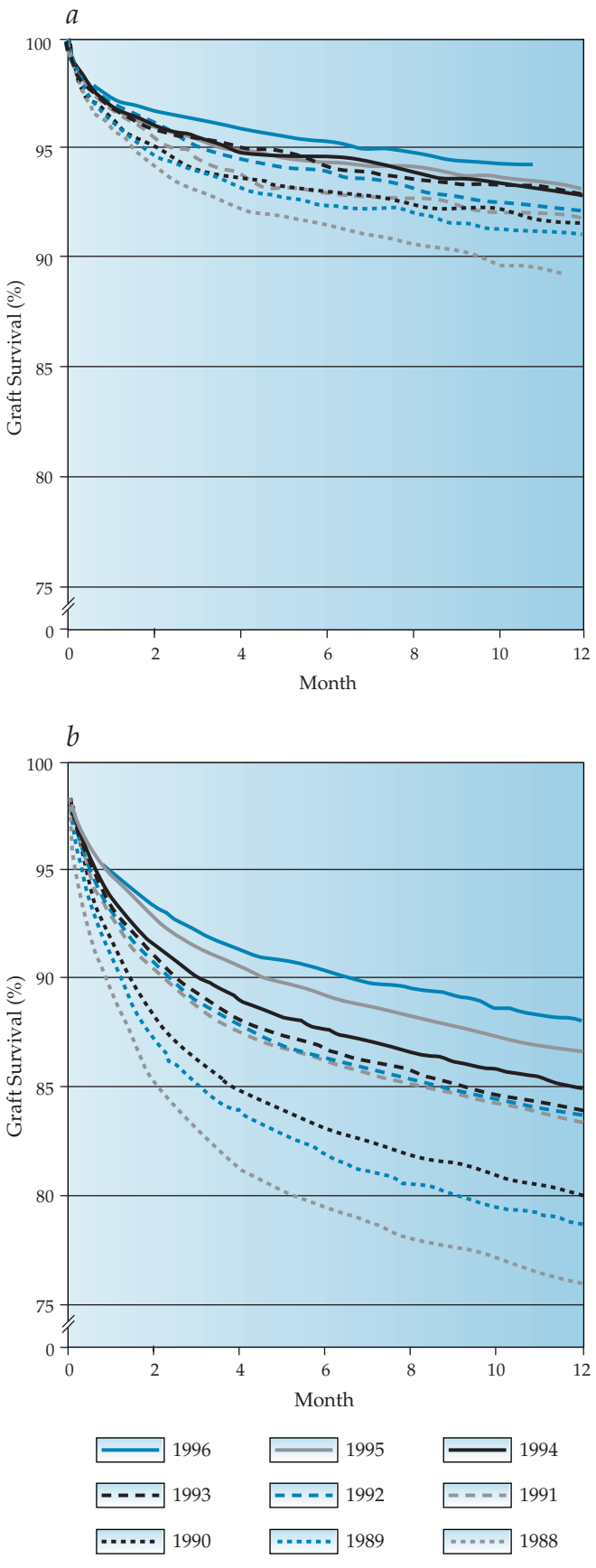


Figure 2 **Kaplan-Meier estimates of graft survival during the first year after transplantation from living donors (*a*) and cadaveric donors (*b*) from 1988 to 1996.**

and tubular cells with viral inclusion bodies. A urine cytology evaluation that does not show decoy cells virtually eliminates the possibility of active polyomavirus nephropathy.[45] There is at present no known therapeutic strategy to correct this disorder other than to reduce immunosuppression. Unfortunately, the reduction of immunosuppression occasionally results in the development of acute rejection.

Epstein-Barr Virus

Transplant recipients who are seronegative for EBV are at high risk for infection if they receive organs from seropositive donors. A primary EBV infection places patients at risk for EBV-related lymphoproliferative disease. This syndrome is a spectrum of illnesses that may range from a virally driven B cell proliferation presenting as a mononucleosis-like disorder to a high-grade malignant lymphoma. The transplantation of an allograft that is seropositive for EBV into a seronegative recipient should be avoided if possible. If such a transplantation is unavoidable, the recipient should be placed on long-term antiviral prophylaxis with acyclovir or ganciclovir.

Urinary Tract Infections

Prophylactic strategies are routinely employed to prevent postoperative urinary tract infections. Patients with structural abnormalities or functional bladder problems are at particular risk. Trimethoprim-sulfamethoxazole is commonly prescribed on a daily basis for 3 to 6 months. This regimen also provides effective prophylaxis for *Pneumocystis* pneumonia and *Nocardia* infections.

Parvovirus B19

Parvovirus B19 is the etiologic agent of the highly contagious erythema infectiosum (fifth disease). It is a common infection in childhood; half of 15-year-old adolescents have specific antiparvovirus B19 antibodies. Infection occurs at a lower rate throughout adult life.

Parvovirus B19 is known to infect the human erythroid progenitor cells in the bone marrow. Failure to develop protective antibodies following infection in an immunosuppressed patient may result in persistent B19 infection and a severe anemia resulting from a pure red cell aplasia. Assays of viral DNA are required to diagnose persistent infection. Treatment is based on the use of intravenous immunoglobulin to provide neutralizing antibodies that result in clearance of the virus and may produce a permanent cure.[46]

Long-term Management Strategies and Outcomes

The introduction of many new immunosuppressive agents, particularly cyclosporine, has dramatically improved short-term outcome of renal transplantation. The incidence of acute rejection is often less than 10% in the first year, and newer drugs have helped provide a noticeable improvement in long-term graft survival [*see Figure 2*]. Hariharan and colleagues demonstrated that from 1988 to 1996, the 1-year survival rate for kidney grafts from living donors increased from 88.8% to 93.9%, and the rate for cadaveric grafts increased from 85.7% to 87.7%.[47] More important, the projected half-life for grafts from living donors increased from 12.7 years to 21.6 years, and the projected half-life for cadaveric grafts increased from 7.9 years to 13.8 years. After censoring the data for patients who died with functioning grafts, the projected half-life for grafts from living

donors increased from 16.9 years to 35.9 years; for cadaveric grafts, half-life increased from 11.0 years to 19.5 years. Thus, the newer immunosuppressive agents have had a substantial positive effect in reducing the risk of acute rejection in the first year and on improved long-term graft survival.

Long-term graft failure is usually the result of the death of a patient with a functioning graft, chronic rejection, or recurrent kidney disease. Greater interest has been placed on strategies to improve long-term patient and graft survival.

PREVENTION OF CHRONIC REJECTION

Long-term studies show that chronic rejection remains the second most important cause of graft loss.[48] The leading cause of graft failure is patient death, which is usually caused by cardiovascular disease. The prevalence of chronic rejection ranges from 10% to 80%, depending on the duration of follow-up. Chronic rejection may be related to either past events (e.g., graft injury), current events (e.g., rejection, dyslipidemia, diabetes), or both.

The most important predictor of chronic rejection is a previous episode of acute rejection, particularly acute rejection after the first year. Other factors that may increase the risk of chronic rejection include ongoing immunologic injury, African-American ancestry, poor HLA matching, poor blood pressure control, and injury to the graft at the time of transplantation (i.e., ischemia or reperfusion injury) [*see Figure 3*]. Early factors that predispose to chronic rejection are delayed graft function and acute rejection. After 6 months, acute rejection remains an important risk factor; factors such as coexistent hypertension and diabetes increase risk.[49,50]

Because chronic rejection is not always a specific alloimmune process and may reflect a number of diseases, more attention has been focused on clinical factors that may contribute to the chronic vasculopathy that leads to progressive obsolescence of the vessels within the kidney and the progressive ischemic injury resulting in interstitial fibrosis, scarring, and nephron loss. When chronic rejection is suspected, a kidney biopsy is essential, because the kidney pathology can guide the decisions concerning treatment and changes in medication.

The critical plan is to have a high index of suspicion for kidney biopsy and undertake biopsy when there are subtle changes in kidney function or new evidence of proteinuria.

MANAGEMENT OF COEXISTING DISEASE

The long-term follow-up of renal transplant recipients should focus primarily on diseases that have been shown to be major causes of recipient morbidity and mortality and to lead to progressive injury of the allograft. The major causes of renal transplant recipient mortality are coronary artery disease, cancer, infection (sepsis), and progressive hepatic dysfunction resulting in liver failure.

The major causes of progressive cardiovascular disease in transplant recipients are primarily related to hypertension, hyperlipidemia, and diabetes mellitus.[51] Improved control of blood pressure, serum glucose level, and lipid level are some of the most important factors in preventing cardiovascular events. Kidney transplant recipients have a much greater risk of cardiovascular events compared to the general population; this increased risk is not entirely explained by traditional Framingham Heart Study cardiovascular risk factors.[52] Consequently, more attention must be placed on controlling known cardiovascular risk factors and reducing the metabolic consequences of immunosuppressive drugs.

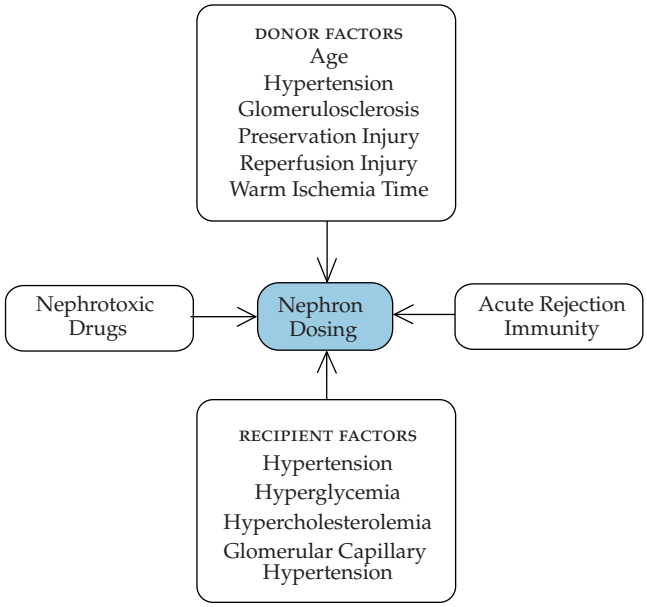


Figure 3 **Paradigm for the development of chronic allograft nephropathy.**

Hypertension

Hypertension is one of the most common complications after renal transplantation; it has multiple etiologies. Hypertension can be a pretransplant condition as well as a complication of transplantation. The prevalence of hypertension, which is frequently as high as 80% or 90%, is in large part caused by disease in the native kidneys and immunosuppressive medications such as cyclosporine and corticosteroids. Cyclosporine not only causes direct vasoconstriction but also induces preglomerular vasoconstriction, which results in a salt-sensitive, volume-dependent form of hypertension. In addition, cyclosporine may have indirect vasoconstrictive effects through an alteration of the balance of vasodilatory and vasoconstricting prostaglandins. Effects on the renin-angiotensin system, the sympathetic nervous system, or both can occur with or instead of the alteration in prostaglandin balance.[53] Graft dysfunction can also worsen the hypertensive process. Renal artery stenosis occurs infrequently but should be considered if blood pressure control is resistant to therapy or renal function worsens as blood pressure is better controlled.

Historically, high blood pressure in renal transplant recipients has not been optimally treated because of the many concomitant medications these patients require for their various medical problems. Many transplant recipients will require from three to five medications to achieve a blood pressure of 140/90 mm Hg. Epidemiologic analyses indicate that there is a direct correlation between systolic blood pressure and graft half-life.[54] A systolic blood pressure near 120 mm Hg is optimal for prolonging graft function. This level of blood pressure control is achieved in only a minority of patients in clinical practice, not only because of the many drugs required but also because of physician reluctance to accept these blood pressure goals, particularly in older hypertensive patients who have only one kidney. Current guidelines of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure recommend a blood pressure goal of less than 130/80 mm Hg for all patients with kidney disease, diabetes, or microalbuminuria.[55] This includes all kidney transplant recipients; howev-

er, observational data in transplant recipients suggest that 120 mm Hg is a preferable goal.

As in patients with native renal disease, drugs that block the renin-angiotensin system should be routinely employed as part of a multidrug strategy to improve blood pressure control. ACE inhibitors and angiotensin II receptor blockers provide better control of glomerular capillary pressure and proteinuria than do other commonly used drugs. This effect may be important in stabilizing renal function and preventing loss of graft function. These drugs may also produce blood pressure–independent benefits by attenuating glomerular scarring and interstitial fibrosis. Physicians frequently avoid these drugs because of concerns about increases in potassium or serum creatinine levels, which could be misconstrued as indicating rejection. These drugs can be used safely, especially 3 to 6 months after transplantation, when the risk of rejection is markedly reduced.

Hyperlipidemia

Hyperlipidemia is common after transplantation. It is usually characterized by increased total cholesterol. The etiology of posttransplantation dyslipidemia is multifactorial but is, in part, related to immunosuppressive drugs given in combination, particularly corticosteroids and cyclosporine. Newer medications, such as sirolimus, may also adversely affect lipid profiles. Frequently, lipid abnormalities will improve within 3 to 6 months after transplantation, as the doses of immunosuppressants are reduced. However, there is a clear association between cardiovascular disease and posttransplantation hyperlipidemia that needs to be considered.[56] More often than not, dietary measures are insufficient to control lipid abnormalities, and frequently, the use of HMG-CoA (3-hydroxy-3-methylglutaryl coenzyme A) reductase inhibitors (statins) are necessary. Although some early reports indicate that higher doses of statins have the potential to cause rhabdomyolysis, particularly in combination with cyclosporine, this problem is unusual. Occasionally, a poorly described form of myopathic pain occurs and subsequently disappears when the medication is discontinued. Other medications, such as fibrates and nicotinic acid, may be necessary to facilitate control of lipids. Azetimide is a new nonabsorbable agent that can be combined with statins to reduce serum levels of low-density lipoprotein (LDL) cholesterol, which plays a major role in the development of atherosclerosis. Caution is warranted before combinations of cholesterol-lowering agents are administered, so that drug-drug interactions are avoided. If patients have diabetes, improved glycemic control is helpful.

Control of hyperlipidemia is an important factor that may affect not only patient survival but also graft survival. Progressive vascular disease in the form of atherosclerosis can occur in the graft as chronic vasculopathy and may contribute to the development of tubulointerstitial ischemia, scarring, and fibrosis. Current adult treatment guidelines recommend maintaining LDL cholesterol levels below 100 mg/dl in patients who have coronary artery disease or who have risk factors for coronary artery disease; however, newer clinical studies illustrate the advantage of treatment that maintains LDL cholesterol levels below 70 mg/dl.[57,58] Lower treatment goals are likely to be advantageous for kidney transplant recipients, given their high cardiovascular disease burden and high cardiovascular death rates.

Diabetes Mellitus

With the increasing numbers of patients who have type 2 (non–insulin-dependent) diabetes mellitus as the cause of end-stage renal disease and older patients who are susceptible to diabetes, abnormalities of glucose tolerance are common after transplantation.

Posttransplantation diabetes mellitus occurs de novo in 4% to 20% of patients.[59] However, it is likely that this number underrepresents the amount of glucose intolerance and clinical diabetes in transplant recipients because of the common use of random glucose or fasting glucose measurements in monitoring patients. Many of the immunosuppressive agents, particularly corticosteroids and tacrolimus, cause or exacerbate posttransplantation diabetes. Cyclosporine may also be diabetogenic.

Efforts designed to improve glycemic control may be helpful in reducing the likelihood of progressive vasculopathy in the transplanted kidney. Minimizing or stopping corticosteroid therapy can markedly improve glycemic control. In patients who are more than 6 months beyond transplantation and who have not experienced rejection, corticosteroids can be completely withdrawn. Diabetogenic agents such as tacrolimus can be replaced with nondiabetogenic immunosuppressants such as sirolimus. Ideally, glycated hemoglobin (HbA_{1c}) should be reduced to less than 7%; this results in fewer cardiovascular events and a reduction of diabetic target organ damage.[60]

MANAGEMENT OF OTHER LONG-TERM MEDICAL CONDITIONS

Musculoskeletal Conditions

Other posttransplantation problems include musculoskeletal conditions, such as osteoporosis. This condition stems from the older age of the patients receiving transplants, the presence of diabetes, poor nutritional state, and concomitant use of corticosteroids. Clinical studies suggest that the majority of bone loss occurs in the first 6 months after transplantation, when steroid use is at its highest. Although newer immunosuppressive regimens employing lower doses of steroids have reduced the frequency of avascular necrosis, osteopenia and osteoporosis remain widespread. These conditions can progress and lead to pathologic fractures, vertebral collapse, and substantial morbidity. Consequently, early identification of bone loss with dual-energy x-ray absorptiometry (DEXA) scanning and interventional therapy with calcium, vitamin D, bisphosphonates, and calcitonin are important in preventing pathologic fractures. Occasionally, hyperparathyroidism can persist after transplantation despite restoration of normal renal function. Chronic hyperparathyroidism may also lead to osteopenia and may require surgical correction.

Chronic hyperparathyroidism may require 6 to 12 months to resolve after successful transplantation. Vitamin D therapy may be useful in suppressing parathyroid activity. Patients with hyperparathyroidism that persists beyond 12 months are candidates for parathyroidectomy.

Hematologic Conditions

Hematologic conditions are usually caused by immunosuppressants or excessive blood draws. Anemia and leukopenia are not uncommon. Many of these conditions can be corrected with vitamins, iron, folate, and reduced doses of antimetabolites or reduction in the frequency of blood draws. Parvovirus infection can also cause anemia, and inadequate renal function may require concomitant use of erythropoietin. Infection with parvovirus B19 can be diagnosed by changes in serology or by detection of the parvovirus genome in sera or bone marrow by polymerase chain reaction. Leukopenia can be caused either by

infection, such as CMV, or by medications, such as azathioprine or mycophenolate mofetil. Drugs that block the renin-angiotensin system, such as ACE inhibitors or angiotensin II receptor blockers, may also cause a 10% to 20% reduction in hematocrit.

Erythrocytosis may develop in 10% to 15% of renal transplant recipients. Repeated phlebotomy or use of ACE inhibitors or angiotensin II receptor blockers can normalize the hematocrit in most patients.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Valderrabano F, Jofre R, Lopez-Gomez JM: Quality of life in end-stage renal disease patients. Am J Kidney Dis 38:443, 2001

2. Wolfe RA, Ashby VB, Milford EL, et al: Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. N Engl J Med 341:1725, 1999

3. Terasaki PI, Cecka JM, Gjertson DW, et al: High survival rates of kidney transplants from spousal and living unrelated donors. N Engl J Med 333:333, 1995

4. Humar A, Durand B, Gillingham K, et al: Living unrelated donors in kidney transplants: better long-term results than with non-HLA-identical living related donors? Transplantation 69:1942, 2000

5. Annual Report of the US Scientific Registry for Transplant Recipients and the Organ Procurement and Transplantation Network: Transplant Data: 2003. United States Department of Health and Human Services, Health Resources and Services Administration, Office of Special Programs, Division of Transplantation, Rockville, Maryland, 2004

6. Metzger RA, Delmonico FL, Feng S, et al: Expanded criteria donors for kidney transplantation. Am J Transplant 3(suppl 4):114, 2003

7. Danovitch GM, Cecka JM: Allocation of deceased donor kidneys: past, present, and future. Am J Kidney Dis 42:882, 2003

8. Wynn JJ, Distant DA, Pirsch JD, et al: Kidney and pancreas transplantation. Am J Transplant 4(suppl 9):72, 2004

9. Port FK, Bragg-Gresham JL, Metzger RA, et al: Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. Transplantation 74:1281, 2002

10. Roodnat JI, Zietse R, Mulder PG, et al: The vanishing importance of age in renal transplantation. Transplantation 67:576, 1999

11. Rice M, Martin J, Hathaway D, et al: Prevalence of cardiovascular risk factors before kidney transplantation. Prog Transplant 12:299, 2002

12. Manske CL, Wilson RF, Wang Y, et al: Prevalence of, and risk factors for, angiographically determined coronary artery disease in type I-diabetic patients with nephropathy. Arch Intern Med 152:2450, 1992

13. Penn I: The effect of immunosuppression on pre-existing cancers. Transplantation 55:742, 1993

14. Abbott KC, Swanson SJ, Agodoa LY, et al: Human immunodeficiency virus infection and kidney transplantation in the era of highly active antiretroviral therapy and modern immunosuppression. J Am Soc Nephrol 15:1633, 2004

15. Pereira BJ, Levey AS: Hepatitis C virus infection in dialysis and renal transplantation. Kidney Int 51:981, 1997

16. Kotanko P, Pusey CD, Levy JB: Recurrent glomerulonephritis following renal transplantation. Transplantation 63:1045, 1997

17. Cosio FG, Alamir A, Yim S, et al: Patient survival after renal transplantation: I. The impact of dialysis pre-transplant. Kidney Int 53:767, 1998

18. Meier-Kriesche HU, Port FK, Ojo AO, et al: Effect of waiting time on renal transplant outcome. Kidney Int 58:1311, 2000

19. Mange KC, Joffe MM, Feldman HI: Effect of the use or nonuse of long-term dialysis on the subsequent survival of renal transplants from living donors. N Engl J Med 344:726, 2001

20. Sesso R, Klag MJ, Ancao MS, et al: Kidney transplantation from living unrelated donors. Ann Intern Med 117:983, 1992

21. Handschin AE, Weber M, Demartines N, et al: Laparoscopic donor nephrectomy. Br J Surg 90:1323, 2003

22. Najarian JS, Chavers BM, McHugh LE, et al: 20 years or more of follow-up of living kidney donors. Lancet 340:807, 1992

23. Narkun-Burgess DM, Nolan CR, Norman JE, et al: Forty-five year follow-up after uninephrectomy. Kidney Int 43:1110, 1993

24. Flowers JL, Jacobs S, Cho E, et al: Comparison of open and laparoscopic live donor nephrectomy. Ann Surg 226:483, 1997

25. Nogueira JM, Cangro CB, Fink JC, et al: A comparison of recipient renal outcomes with laparoscopic versus open live donor nephrectomy. Transplantation 67:722, 1999

26. Cecka MJ, Terasaki PI: The UNOS Scientific Renal Transplant registry—2000, Clinical Transplants 2000. Cecka MJ, Terasaki PI, Eds. UCLA Immunogenetics Center, Los Angeles, 2001, p 1

27. Kasiske BL, Bia MJ: The evaluation and selection of living kidney donors. Am J Kidney Dis 26:387, 1995

28. Khwaja K, Asolati M, Harmon J, et al: Outcome at 3 years with a prednisone-free maintenance regimen: a single-center experience with 349 kidney transplant recipients. Am J Transplant 4:980, 2004

29. Kasiske BL, Chakkera HA, Louis TA, et al: A meta-analysis of immunosuppression withdrawal trials in renal transplantation. J Am Soc Nephrol 11:1910, 2000

30. Mycophenolate mofetil in cadaveric renal transplantation. US Renal Transplant Mycophenolate Mofetil Study Group. Am J Kidney Dis 34:296, 1999

31. Ojo AO, Meier-Kriesche HU, Hanson JA, et al: Mycophenolate mofetil reduces late renal allograft loss independent of acute rejection. Transplantation 69:2405, 2000

32. Pirsch JD, Miller J, Deierhoi MH, et al: A comparison of tacrolimus (FK506) and cyclosporine for immunosuppression after cadaveric renal transplantation. FK506 Kidney Transplant Study Group. Transplantation 63:977, 1997

33. Andoh TF, Bennett WM: Chronic cyclosporine nephrotoxicity. Curr Opin Nephrol Hypertens 7:265, 1998

34. Burke JF Jr, Pirsch JD, Ramos EL, et al: Long-term efficacy and safety of cyclosporine in renal-transplant recipients. N Engl J Med 331:358, 1994

35. Weir MR, Ward MT, Blahut SA, et al: Long-term impact of discontinued or reduced calcineurin inhibitor in patients with chronic allograft nephropathy. Kidney Int 59:1567, 2001

36. Ligtenberg G, Hene RJ, Blankestijn PJ, et al: Cardiovascular risk factors in renal transplant patients: cyclosporin A versus tacrolimus. J Am Soc Nephrol 12:368, 2001

37. Kreis H, Cisterne JM, Land W, et al: Sirolimus in association with mycophenolate mofetil induction for the prevention of acute graft rejection in renal allograft recipients. Transplantation 69:1252, 2000

38. Kahan BD: Efficacy of sirolimus compared with azathioprine for reduction of acute renal allograft rejection: a randomised multicentre study. The Rapamune US Study Group. Lancet 356:194, 2000

39. A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. Ortho Multicenter Transplant Study Group. N Engl J Med 313:337, 1985

40. Olyaei AJ, Thi K, deMattos AM, et al: Use of basiliximab and daclizumab in kidney transplantation. Prog Transplant 11:33, 2001

41. Sawada T, Fuchinoue S, Teraoka S: Successful A1-to-O ABO-incompatible kidney transplantation after a preconditioning regimen consisting of anti-CD20 monoclonal antibody infusions, splenectomy, and double-filtration plasmapheresis. Transplantation 74:1207, 2002

42. Humar A, Payne WD, Sutherland DE, et al: Clinical determinants of multiple acute rejection episodes in kidney transplant recipients. Transplantation 69:2357, 2000

43. Fishman JA, Rubin RH: Infection in organ-transplant recipients. N Engl J Med 338:1741, 1998

44. Randhawa PS, Finkelstein S, Scantlebury V, et al: Human polyoma virus-associated interstitial nephritis in the allograft kidney. Transplantation 67:103, 1999

45. Drachenberg RC, Drachenberg CB, Papadimitriou JC, et al: Morphological spectrum of polyoma virus disease in renal allografts: diagnostic accuracy of urine cytology. Am J Transplant 1:373, 2001

46. Young NS, Brown KE: Parvovirus B19. N Engl J Med 350:586, 2004

47. Hariharan S, Johnson CP, Bresnahan BA, et al: Improved graft survival after renal transplantation in the United States, 1988 to 1996. N Engl J Med 342:605, 2000

48. Jindal RM, Hariharan S: Chronic rejection in kidney transplants: an in-depth review. Nephron 83:13, 1999

49. Prommool S, Jhangri GS, Cockfield SM, et al: Time dependency of factors affecting renal allograft survival. J Am Soc Nephrol 11:565, 2000

50. Matas AJ, Gillingham KJ, Humar A, et al: Immunologic and nonimmunologic factors: different risks for cadaver and living donor transplantation. Transplantation 69:54, 2000

51. Kasiske BL: Risk factors for cardiovascular disease after renal transplantation. Miner Electrolyte Metab 19:186, 1993

52. Kasiske BL, Chakkera HA, Roel J: Explained and unexplained ischemic heart disease risk after renal transplantation. J Am Soc Nephrol 11:1735, 2000

53. Curtis JJ, Luke RG, Jones P, et al: Hypertension in cyclosporine-treated renal transplant recipients is sodium dependent. Am J Med 85:134, 1988

54. Opelz G, Wujciak T, Ritz E: Association of chronic kidney graft failure with recipient blood pressure. Collaborative Transplant Study. Kidney Int 53:217, 1998

55. Chobanian AV, Bakris GL, Black HR, et al: The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 289:2560, 2003

56. Massy ZA, Kasiske BL: Post-transplant hyperlipidemia: mechanisms and management. J Am Soc Nephrol 7:971, 1996

57. Cannon CP, Braunwald E, McCabe CH, et al: Intensive versus moderate lipid lowering with statins after acute coronary syndromes. N Engl J Med 350:1495, 2004

58. Nissen SE, Tuzcu EM, Schoenhagen P, et al: Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. JAMA 291:1071, 2004

59. Weir MR, Fink JC: Risk for posttransplant diabetes mellitus with current immunosuppressive medications. Am J Kidney Dis 34:1, 1999

60. UK Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 352:837, 1998

XII NEPHROLITHIASIS

Fuad N. Ziyadeh, m.d.
Stanley Goldfarb, m.d.

The annual incidence of nephrolithiasis in the United States is 7 to 21 per 10,000 population (higher in the Southeast and lower in the Northwest), and the annual incidence of hospitalization for nephrolithiasis is 7 to 10 cases per 1,000 admissions.[1] As many as 5% to 10% of all cases of active stone passage may require hospitalization to manage complications such as urinary tract obstruction, infection, and severe pain.

Active Stone Passage

DIAGNOSIS

Clinical Presentation

Patients with active stone passage generally present with pain and either microscopic or macroscopic hematuria. Usually, the pain begins abruptly, is severe, is constant rather than colicky, and is either localized to the flank or radiating from the flank to the groin area. Flank pain may be associated with nausea and vomiting. Midanterior abdominal pain, however, is an uncommon clinical presentation of stone passage and should prompt a careful search for other disorders.

The site of obstruction determines the location of pain. A stone in the renal pelvis or upper part of the ureter usually causes flank pain, whereas a stone in the middle or lower ureter causes radiation of pain to the groin and genitalia. Obstruction at the terminal ureter usually results in signs and symptoms that mimic those of cystitis (urinary frequency, urgency, and dysuria), as well as pain in the lower abdomen. Occasionally, patients present with signs of extracellular volume depletion as a result of vomiting. Atypical presentation may include obstruction without pain or hematuria, progressive renal failure, urosepsis, or adynamic paralytic ileus, which may suggest a primary gastrointestinal disorder.

Passage of a sloughed papilla, clot, or fungal ball can produce renal colic and symptoms similar to those produced by the passage of a stone.

Laboratory Studies

If active stone passage is suspected, freshly voided urine should be filtered to recover stones or gravel. Microscopic examination of freshly voided, warm urine can detect diagnostically important crystalluria. Although the finding of hematuria has been considered a hallmark of acute stone passage, studies utilizing sensitive imaging techniques suggest that 20% to 25% of patients with acute stone passage may not have hematuria.[2,3] Urine specimens should be sent for culture in all cases of nephrolithiasis because the coexistence of unsuspected infection and obstruction could lead to sepsis and death. Stones are best analyzed by x-ray diffraction, but infrared spectroscopy is an acceptable alternative.

Leukocytosis may indicate concurrent infection but may also result from stress-induced demargination of leukocytes. Total serum content of CO_2 may be low in the case of acute respiratory alkalosis resulting from associated pain or anxiety or may be high in the case of metabolic alkalosis resulting from vomiting.

It is generally not useful to perform extensive biochemical screening tests of calcium metabolism during active stone passage. Serum calcium levels could be falsely elevated as a result of concurrent extracellular fluid volume depletion. Also, abnormalities in levels of urinary calcium, phosphate, and citrate, as well as pH, may be induced by acute obstruction; therefore, the values obtained would not reflect true steady-state conditions. Such metabolic evaluations should usually be delayed until at least 6 weeks after an acute stone episode.[4]

Radiologic Studies

Radiologic studies are very useful in the initial evaluation of active stone passage. A flat film of the abdomen may show the location of a radiopaque stone and may also provide evidence of ileus. Ultrasonography may detect radiolucent stones and can identify urinary tract obstruction. However, ultrasonography may fail to identify very small stones (< 5 mm in diameter) and cannot define the degree or site of obstruction. Intravenous urography (IVU) is a standard method for examining the urinary tract for obstruction in cases of renal colic [*see Figure 1*]. A meta-analysis has shown the accuracy of helical scan computed tomography to be superior to IVU in diagnosing acute nephrolithiasis.[5] This modality is now widely available and should be employed, although its higher cost must be considered.

Recommended Workup

Patients presenting with acute onset of abdominal or flank pain should be evaluated for the possibility of nephrolithiasis. If therapy is to be initiated for a presumed kidney or ureteral stone, red blood cells should be detected on urinalysis or a radiodense stone should be seen on flat abdominal film. If hematuria is absent or if a stone is not seen on a flat film of the abdomen, a helical CT scan should be performed. If a stone is not seen on helical CT scan, the diagnosis of nephrolithiasis can be ruled out.

TREATMENT

Most kidney stones less than 5 mm in diameter pass into the urine within 48 hours after the onset of an acute episode of renal colic. Up to 20% of patients with a kidney stone require some form of intervention to effect passage of the stone. Patients generally require supportive treatment for pain and for extracellular fluid volume depletion or dehydration. The majority of patients do not require hospitalization unless they have infection, obstruction, severe pain, or intractable vomiting. If a stone that causes total obstruction does not pass, active intervention is needed [*see Figure 1*]. Urgent nephrostomy relieves the obstruction and can be followed by extracorporeal shock wave lithotripsy (ESWL), percutaneous lithotripsy, retrograde basket retrieval of the stone, or, in rare instances, surgical intervention.[6]

Chronic and Recurrent Nephrolithiasis

Pursuit of extensive biochemical studies in patients who have passed a single stone is controversial, particularly when there is no evidence of continued stone formation. Although the decision to proceed with diagnostic studies and laboratory evalua-

tion should be individualized, most physicians delay evaluations until metabolic activity can be substantiated by the formation of new stones or by an increase in size in situ of a known stone. However, one study of stone-forming patients suggested that the greater the number of stones occurring before therapy is instituted, the greater the likelihood of relapse.[7]

Nephrolithiasis usually recurs; in males, there is a 40% chance of recurrence within 5 years and an almost 60% chance within 10 years after the appearance of the first stone.[6] Recurrence rates for females are about two thirds those of males.[8] A stone may remain asymptomatic and without associated metabolic activity; however, the cumulative occurrence of a symptomatic event related to stone passage over a period of 5 years is about 50%.[9]

DIAGNOSIS

Laboratory Studies

A detailed evaluation of patients should include duplicate measurements, taken at different times, of 24-hour urine levels of calcium, phosphate, uric acid, sodium, citrate, oxalate, cystine, creatinine, and urea nitrogen, as well as urine volume.[10] Patients must not deviate from their normal diet and fluid intake before testing. Modification of diet or fluid intake before the assessment may obscure the possible role of these factors in the pathogenesis of the patient's stone diathesis. Concomitant plasma determinations of calcium, phosphate, creatinine, uric acid, bicarbonate, and urea nitrogen should be obtained. Fasting urinary pH should be checked in all cases, especially in those patients with hypokalemia, hyperchloremic acidosis, or radiologic evidence of nephrocalcinosis, which are all findings characteristic of distal renal tubular acidosis. Microscopic examination of urine, which can detect crystals and signs of infection [*see Figures 2a through d*], should be performed, and specimens should be sent for urine culture to determine the type of stone [*see* Crystallographic Analysis, *below*].

Radiologic Studies

Radiologic studies can localize in situ stones, document current stone size, and assess the possible efficacy of future therapy aimed at shrinking stones and preventing new stone formation. An abdominal plain film is the initial radiologic study test because of its availability and practicality in an acute setting; however, it has limited sensitivity and specificity in the diagnosis of stones. A CT scan of the kidney is an excellent modality for identifying all radiolucent or radiopaque calculi, but cost limits its routine use. Noncontrast-enhanced spiral CT is very effective in determining the presence of urinary tract obstruction in patients with acute flank pain, and the entire test can be completed within 5 minutes.[5] Spiral CT can differentiate between a ureteral calculus and a phlebolith overlying the ureter.[11]

IVU is the previous gold standard for identifying stones and remains useful in planning therapy and confirming diagnosis. It is more sensitive and more specific than ultrasonography in identifying stones.[12] Ultrasonography is able to show a dilated renal pelvis, which indicates that obstruction is likely. IVU is able to show not only a dilated renal pelvis or dilated ureter but also the site or level of obstruction. IVU affords the additional benefit of assessing how urine flow and, consequently, kidney function are affected by the obstruction. Slow or absent urine flow proximal to the site of obstruction indicates poor kidney function.

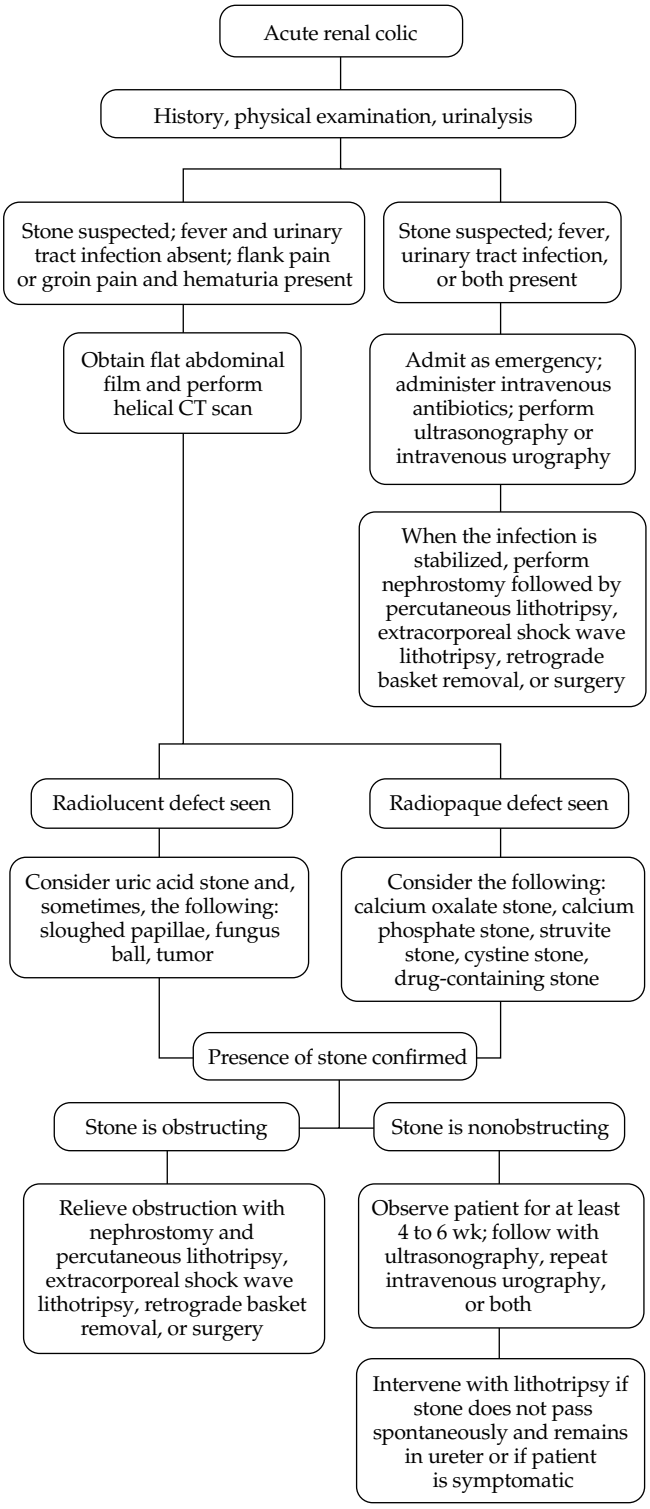


Figure 1 **This flowchart diagrams the protocol for diagnosing and treating a patient with acute renal colic.**

A follow-up flat abdominal film with or without tomography should be obtained and compared with any previous films to identify changes in the size, number, and location of stones. Quality of the plain abdominal film is improved if the patient takes a laxative the night before. Such films can be used for subsequent monitoring of in situ stones.

Crystallographic Analysis

Crystallographic analysis of renal stones serves as the basis for the diagnostic evaluation [*see Table 1*]. Stone formation is determined by the characteristics of urine, including its stone-forming constituents, volume, pH, and the various inhibitors of crystal formation, aggregation, and growth. The majority of patients with urolithiasis form calcium stones, most of which are composed of calcium oxalate or, less often, calcium phosphate. Other types of stone include uric acid, struvite, and cystine stones. In HIV patients, antiretroviral agents may induce urolithiathasis [*see* Drug-Induced Stones (Indinavir and Nelfinavir Stones), *below*]

OVERVIEW OF MANAGEMENT

The long-term treatment goal for patients with recurrent nephrolithiasis is to prevent new stone formation and growth of existing stones. Prevention of renal stone formation and growth can be accomplished by reducing the concentration of stone-forming constituents in the urine, manipulating the ionic milieu of the urine to favor solubilization of these stone-producing constituents, and correcting any mechanical or structural abnormalities that are associated with stasis in the urinary tract.

Although prevention should always be a major focus of therapy, there is an important role for stone removal in the initial evaluation. Percutaneous nephrostolithotomy and open surgery are equally effective for the management of renal stones, but percutaneous nephrostolithotomy shortens hospitalization by 60% and leads to a much more rapid return to normal activities.[13] It should be noted that percutaneous nephrostolithotomy is highly operator dependent, and published data from the highly experienced centers cannot be generalized to every urology practice. Percutaneous stone-removal procedures are currently reserved for patients with the following conditions:

1. Extremely large (> 2 cm in diameter) or complex calculi.
2. Cystine stones that are relatively resistant to ESWL.
3. Anatomic abnormalities, including horseshoe kidneys or ureteropelvic junction obstruction.
4. Stones within caliceal diverticula.

ESWL, which employs high-energy shock waves produced by an electrical discharge, is a major advance in the treatment of renal and ureteral stones, and most renal and proximal ureteral stones are currently managed by this technique. The shock waves are transmitted through water and directly focused on a renal stone, a ureteral stone, or both with the aid of biplanar fluoroscopy. The change in tissue density between the soft renal tissue and the hard stone causes a release of energy at the stone surface, which fragments the stone. Current lithotriptor technology is quite effective in fragmenting the majority of renal and ureteral calculi; data suggest that the technique is safe.[14] Recovery of stone fragments after ESWL provides important information about stone composition, which, in turn, allows the clinician to develop a management plan that minimizes recurrent stone formation.

Calcium-Containing Stones

DIAGNOSIS

Urinary calcium excretion is normally less than 300 mg/24 hr in males and less than 250 mg/24 hr in females, or less than 4 mg/kg body weight/24 hr.[15] Fasting urinary pH should be measured in all patients with calcium-containing stones to identify those patients who form hypercalciuric stones and those who form normocalciuric stones, because the management differs accordingly. If after a 12-hour fast urinary pH is greater than 6.0 in the absence of urinary tract infection (a potential cause of increased urinary pH if the infecting organism produces the ectoenzyme urease), an ammonium chloride acid loading test (100 mg/kg body weight, using a 500 mg/5 ml oral solution) should be performed. If urinary pH does not fall to less than 5.4 when serum bicarbonate falls below 20 mEq/L, classic distal renal tubular acidosis is present. In this condition, hypercalciuria occurs because acidosis enhances bone resorption and directly reduces reabsorption of calcium in the distal nephron.

RISK FACTORS FOR RECURRENT STONE FORMATION

The risk factors for calcium-containing stones include increased crystalloid concentration in the urine as a result of tubular disorders, hypercalcemia and hypercalciuria, hyperoxaluria, hyperuricosuria, reduced urine volume, abnormalities in the excretion of inhibitors of stone formation, and the presence of promoters of stone formation.[16] Calcium-containing stones may be primarily calcium oxalate or calcium phosphate [*see Table 1*]. Often, a person at risk has a family history of calcium stones.[17] Hypercalciuria is present in approximately 60% of patients with calcium-containing stones but is also detected in about 10% of people who never form stones. Studies have shown that genetic defects may underlie some forms of nephrolithiasis. Mutations in the gene encoding a chloride channel (*CLC-5*) expressed in the distal tubule may manifest themselves in a variety of phenotypic conditions, including hypercalciuria and proteinuria.[18] Various conditions increase the risk of stone formation; identification and correction of these underlying conditions is key to prevention of stone recurrence.

Specific Tubular Disorders

Distal (type 1) renal tubular acidosis contributes to stone formation in two ways: persistently high urine pH, which favors precipitation of calcium phosphate, and low levels of citrate excretion, which is normally an inhibitor of stone formation. Renal tubular acidosis should be suspected when one or more of the following are present: punctate medullary nephrocalcinosis, metabolic acidosis with a normal anion gap (hyperchloremic acidosis), alkaline urine pH, and hypokalemia. These findings, however, can be absent in so-called incomplete distal renal tubular acidosis, during which the serum bicarbonate concentration and the blood pH are within the normal range but only intracellular acidosis is present. The fasting urinary pH is not maximally acidic (pH > 5.8) and fails to fall during ammonium chloride acid loading. Intracellular acidosis is manifested by a reduction in the excretion of urinary citrate, the principal chelator of urine calcium and an inhibitor of crystallization.[19] Citrate excretion is usually reduced in states of metabolic acidosis, which stimulates the proximal reabsorption of this anion.[15]

Hypercalcemia and Hypercalciuria

Primary hyperparathyroidism accounts for 5% of calcium stones. Hypercalcemia that is caused by primary hyperparathyroidism increases the filtered and, therefore, the excreted load of calcium despite the opposing action of parathyroid hormone to enhance calcium reabsorption. Intermittent hypercalcemia may be seen in patients with mild primary hyperparathyroidism.

A particularly severe form of hypercalciuria may be caused

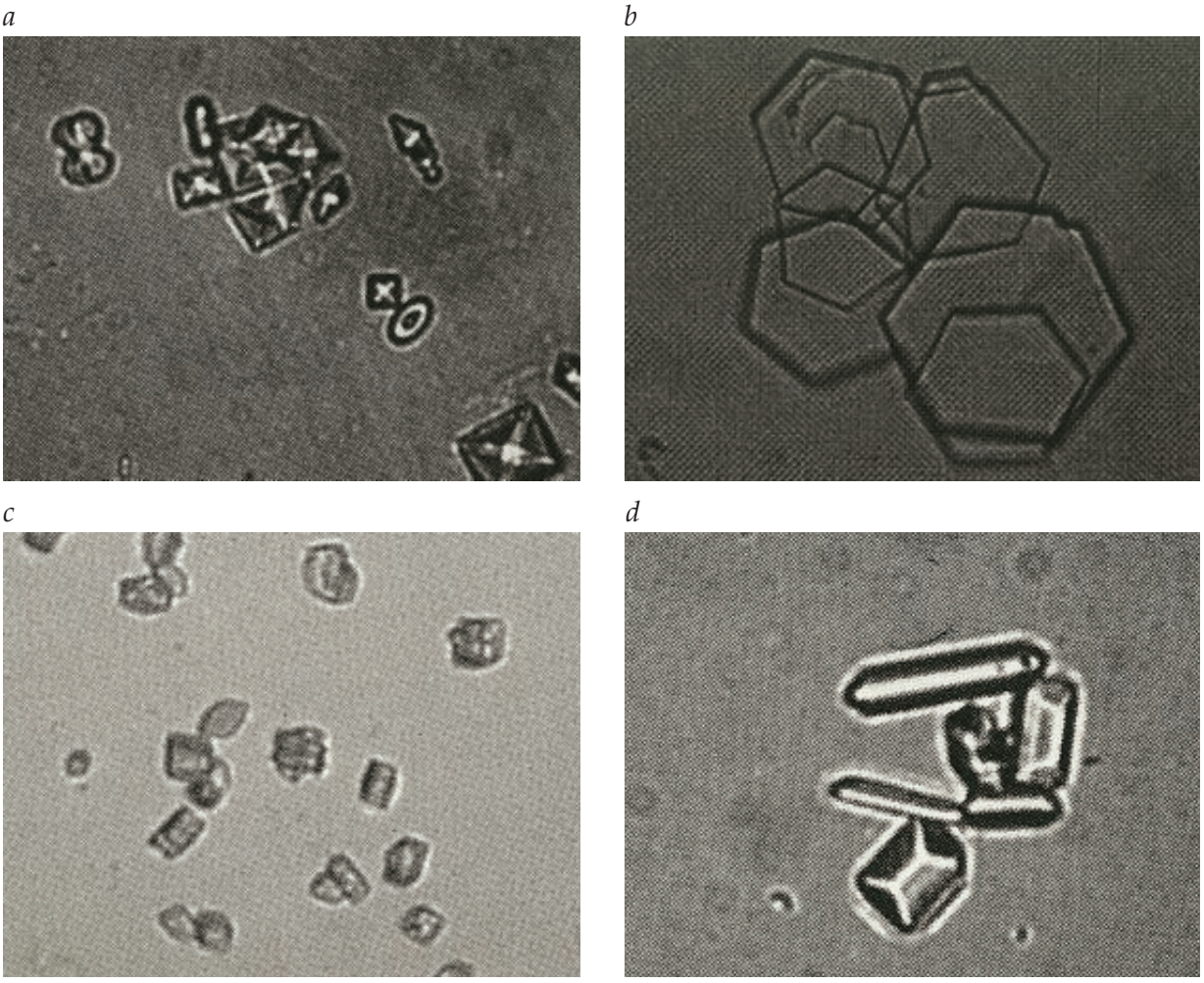


Figure 2 **In a case of chronic recurrent nephrolithiasis, microscopic examination of urine can detect and identify crystal-forming substances. Such substances include calcium oxalate (*a*), cystine (*b*), uric acid (*c*), and struvite (*d*).**

by sarcoidosis and, on rare occasions, other granulomatous conditions because a larger load of calcium is filtered and excreted (up to 600 to 700 mg/day), and secondary suppression of parathyroid hormone occurs as a result of the hypercalcemia. Circulating levels of 1,25-dihydroxyvitamin D_3 (calcitriol) may be increased in these conditions, and vitamin D intoxication may present as hypercalciuria. Patients on vitamin D therapy for osteoporosis or hypoparathyroidism may experience severe hypercalciuria and nephrolithiasis. Thyroid excess, from inappropriate hormone replacement therapy or from spontaneous disease, may also induce hypercalciuria and nephrolithiasis.

Dietary Influences

Absorptive hypercalciuria is a condition in which hypercalciuria occurs only with a high dietary intake or large oral ingestion of calcium. Renal hypercalciuria is defined by an increased urinary calcium excretion after an overnight fast. The classification of idiopathic hypercalciuria as absorptive or renal is highly dependent on the effects of dietary sodium. Renal hypercalciuria may be attributable to an intake of excess dietary sodium during the previous 24 hours in many, if not most, cases. Restriction of dietary sodium intake can produce a marked and predictable fall in urinary calcium excretion and thus cause renal hypercalciuria to be recategorized as absorptive hypercalciuria. Specific treatment of hypercalciuria on the basis of a distinction between renal and absorptive hypercalciuria is probably unwarranted.[20]

The evidence that high dietary protein intake is an important risk factor for recurrent nephrolithiasis is derived from various findings. First, population studies show a clear correlation between the societal consumption of animal protein and the prevalence of upper urinary tract stone disease.[21] Second, increased dietary protein ingestion raises urinary calcium, uric acid, and possibly oxalate excretion in normal individuals.[22] The risk has also been found in epidemiologic studies.[23] Third, high dietary protein has a hypocitraturic effect—that is, it produces a decline in the excretion of citrate, a principal factor in the chelation of urinary calcium.[24,25] Fourth, studies of dietary protein restriction have shown a beneficial effect in vegetarians and in specific therapeutic trials, particularly in persons who ingest an excessive amount of protein. Finally, some studies have found that patients with recurrent stone disease tend to be hyperresponsive to the calciuric effect of high dietary protein because they excrete larger amounts of calcium at any level of dietary protein intake than do control subjects.[26] In any event, a careful review of the literature has reaffirmed the potential role of excessive dietary protein in the pathogenesis of kidney stone formation [*see* Treatment, Dietary Modification, *below*],[27] although a

comparative trial did not find that a diet high in fiber and low in animal protein provided benefit over increased fluid intake alone.[28]

Hyperoxaluria

Normal 24-hour urinary excretion of oxalate should not exceed 40 to 45 mg/24 hr (0.7 mg/kg/24 hr) to reduce the risk of calcium oxalate stone formation. For physicochemical reasons, small increases in the urinary oxalate excretion may actually contribute more to calcium oxalate stone formation than a proportional increase in urinary calcium excretion.[29]

There are several causes of hyperoxaluria in patients with calcium oxalate stones, including hyperabsorption of oxalate, vitamin B_6 deficiency, and genetic predisposition.

Hyperabsorption of oxalate, mainly in the colon, may occur if there is a reduction in free calcium in the intestinal lumen. Calcium complexes with unabsorbed fat, which allows the dietary oxalate to be absorbed more readily; bile and long-chain fatty acids also increase oxalate permeability. The resultant increase in oxalate absorption leads to hyperoxaluria. This mechanism has been proposed to explain the hyperoxaluria commonly seen in states of GI malabsorption.[29] In some persons, marked increases in urinary oxalate excretion occur after the ingestion of 1 to 2 g of ascorbate because oxalate is an end product of ascorbate metabolism.

Vitamin B_6, or pyridoxine, is a cofactor in the interconversion of glycine and glyoxylate, which are both precursors of oxalate. A deficiency of this vitamin results in the increased formation and excretion of oxalate.[30]

Primary hyperoxaluria (type I and type II) is inherited as an autosomal recessive trait and is characterized by overproduction of oxalate. Glycolic acid excretion is increased in type I primary hyperoxaluria, whereas L-glyceric acid excretion is increased in type II primary hyperoxaluria.[31]

A deficiency of oxalate-degrading bacteria (in particular *Oxalobacter formigenes*) has been suggested as a factor in calcium oxalate stone formation.[32] By utilizing oxalate in the digestive tract, oxalate-degrading bacteria regulate oxalate homeostasis and prevent excessive enteric absorption of oxalate, thereby lowering the risk of urolithiasis; a deficiency of oxalate-degrading bacteria causes increased oxalate excretion, which in turn may increase the risk of hyperoxaluria and calcium oxalate stone formation.

Hypocitraturia

Normal 24-hour urinary excretion of citrate is approximately 300 to 900 mg, with levels substantially higher in premenopausal women, especially pregnant women, than in men. Hypocitraturia is usually profound in distal renal tubular acidosis and is perhaps the most important risk factor for calcium oxalate stone formation in patients with that disorder.[17] Hypocitraturia is caused by metabolic acidosis in distal renal tubular acidosis and by intracellular acidosis in incomplete renal tubular acidosis. Other causes of hypocitraturia include potassium depletion, presumably because of intracellular acidosis; bacteriuria, because the infecting organisms may metabolize the urinary citrate; and acidifying conditions, such as renal insufficiency and chronic diarrhea. An idiopathic variety of hypocitraturia may be associated with recurrent nephrolithiasis in many patients, partly because of excessive dietary protein intake, which leads to increased endogenous acid production and excretion. Studies have shown an important role for the enzyme citrate lyase in the hypocitraturia of metabolic acidosis. Low intracellular pH in the proximal tubule activates this enzyme and leads to increased intracellular citrate utilization, thereby enhancing citrate absorption and lowering urinary citrate in metabolic acidosis.[33]

Hyperuricosuria

Hyperuricosuria plays a role in the formation of both uric acid and calcium oxalate stones. A relation between hyperuricosuria and a particularly severe form of recurrent calcium oxalate nephrolithiasis has been observed.[34] Experimental in vitro studies using highly defined artificial solutions suggest that urate crystals, but not uric acid crystals, can act as epitaxial templates that bind calcium oxalate and promote crystal growth. Uric acid crystals, however, are the predominant form in urine; urate crystals are rarely found in urine and, furthermore, are rather poor promoters of crystal growth in the highly defined artificial urine solutions that are used in experimental studies to investigate the nature of crystal growth.

The ability of uric acid crystals to adsorb glutamic acid crystals[35] may explain why uric acid is a potent promoter of crystallization in urine, which contains various organic compounds that act as crystallization inhibitors, but not in highly defined artificial urine solutions. Hyperuricosuria in patients with calcium oxalate or uric acid stones is a function of excess dietary purine consumption rather than any intrinsic abnormality in purine metabolism.[36]

Reduced Urine Volume

Urine volume is a crucial determinant of the risk of recurrent nephrolithiasis. The risk of calcium stone formation is substantially increased if the urine output is less than 15 ml/kg/day in children[37] and less than approximately 1,500 ml/24 hr in adults.[38]

Abnormalities in the Excretion of Inhibitors of Stone Formation

Protein macromolecules account for most of the inhibition of crystal growth observed in in vitro assays of normal urine. Low-

Table 1 Crystallographic Analysis of the Stone Serves as the Basis for Diagnostic Evaluation

Stone Composition	*Incidence (%)*	*Appearance of Crystals on Urinalysis*	*Radiographic Appearance*
Calcium oxalate with or without calcium apatite	70	Octahedrons	Radiopaque
Calcium monohydrogen phosphate (brushite)	2	Amorphous crystals	Radiopaque
Ammoniomagnesium phosphate (struvite)	15–17	Coffin lids	Radiopaque
Uric acid	10	Pears and diamonds	Radiolucent
Cystine	1	Hexagons	Radiopaque

molecular-weight urinary substances, including magnesium, pyrophosphate, and citrate, also show inhibitory activity.[39] Patients who have no identifiable abnormality but who have recurrent calcium-containing stones probably lack urinary inhibitors of stone formation. Abnormalities have been reported in the structure and function of nephrocalcin isolated from the urine and stones of some patients with calcium oxalate stones.[39] Nephrocalcin is a 14 kd glycoprotein that contains γ-carboxyglutamic acid, which is a vitamin K–dependent amino acid that is also found in osteopontin (a bone glycoprotein) and its urinary analogue, uropontin.[40]

TREATMENT

The initial treatment strategy for patients with recurrent stone disease is to treat the underlying disorder. Specific therapies are obviously indicated in certain conditions implicated in the pathogenesis of calcium-containing stones; for example, patients with primary hyperparathyroidism require parathyroid surgery, those with distal renal tubular acidosis require alkali therapy, and those with sarcoidosis require steroids or other specific therapy.

Dietary Modification

Evidence supports the view that reducing the prevailing urinary concentrations of potentially insoluble constituents of kidney stones by increasing fluid intake and modifying the diet are effective in the prevention of calcium-containing stones. A prospective 5-year controlled trial of water intake in patients with a first calcium-containing stone showed that increasing urine volume to about 2.5L/day, which reduced the relative saturation of urine with respect to calcium oxalate, reduced stone recurrence to 10%; persons who did not increase water intake had a stone recurrence of 25%.[38] The type of fluid consumed may be important; grapefruit juice may enhance the risk of stone formation,[41] whereas patients who ingest one bottle of beer a day may reduce the risk of stone formation by 40%.[42]

Timed urine collections for volume, urea nitrogen, creatinine, calcium, and sodium are helpful in monitoring compliance. Patients are encouraged to reduce their calcium intake only if it is very excessive (but to maintain it at or above 1 g/day) and to reduce oxalate intake as much as possible by avoiding foods that contain high amounts of oxalate. In fact, a prospective study of 45,617 healthy men between 40 and 75 years of age who did not have a history of kidney stones led to the conclusion that a high dietary calcium intake decreases the risk of symptomatic kidney stones, presumably by reducing the gastrointestinal absorption of oxalate and hence the urinary oxalate excretion.[43,44] Curhan and colleagues showed that a higher intake of dietary calcium decreases the risk of kidney stone formation in younger women; in this patient population, supplemental calcium was not associated with increased risk of kidney stones.[23] In older women, high intake of dietary calcium appeared to decrease risk for symptomatic kidney stones; however, unlike in younger women, intake of supplemental calcium was associated with increased risk of stone formation. The reasons for this differing response to supplemental calcium are not clear.[45]

For patients with recurrent calcium-containing stones, restricted intake of animal protein and salt, combined with a normal calcium intake, reduced stone recurrence over 5 years more effectively than the traditional low-calcium diet.[29] The benefits of dietary therapy in patients with recurrent stone formation have not been rigorously tested in multiple controlled studies; in addition, there are very few clinical studies of proposed pharmacologic therapies (e.g., thiazides, phosphates, allopurinol, and citrate) for calcium-containing stone formation.[46] Evidence indicates that treatment with allopurinol, which blocks uric acid production, decreases the recurrence rate of calcium oxalate kidney stones in patients with hyperuricosuria, although allopurinol has no effect on urinary calcium or oxalate saturation.[47]

With no clear-cut advantage of any specific pharmacologic agent, it appears that many patients may benefit from dietary modification and avoid embarking on a lifelong use of medications to prevent recurrent stone formation. However, if patients are persistently hypercalciuric despite dietary modification or if they persistently form stones, pharmacologic therapy should be considered. A meta-analysis supports this concept.[48]

Treatment of Hypercalcemia and Hypercalciuria

Drugs that reduce urine calcium excretion Thiazides, most frequently hydrochlorothiazide, and longer-acting drugs such as chlorthalidone may be used to lower urine calcium excretion. Therapy with hydrochlorothiazide was highly effective in reducing new stone formation in two randomized controlled clinical trials.[49,50] Studies that failed to show a beneficial effect[51,52] were flawed because of relatively short follow-up (12 to 18 months).[53] The use of thiazide diuretics reduces the incidence of recurrent calcium nephrolithiasis by approximately 80%.[54] A total daily dose of 50 to 100 mg of hydrochlorothiazide effectively reduces calcium excretion. A daily dose of 25 mg is sometimes sufficient and avoids many of the side effects of thiazides. Failure of this therapy is usually a result of a marked increase in sodium intake (which promotes hypercalciuria) or hypokalemia. Potassium depletion leads to intracellular acidosis and causes hypocitraturia.[55]

Thiazides must reduce extracellular fluid volume to produce their full hypocalciuric action. A component of the hypocalciuric action of thiazides is also related to direct stimulation of calcium reabsorption in the distal tubule. The impact of side effects of thiazides, such as hyperlipidemia, glucose intolerance, and mild hypokalemia, has not been extensively studied in patients with nephrolithiasis. On the other hand, it has been shown that bone mineralization probably increases in persons receiving long-term thiazide therapy, thus decreasing the risk of osteoporosis.[56] Patients who have underlying disorders of calcium metabolism, such as primary hyperparathyroidism, or who are being treated with vitamin D supplements may be at risk for hypercalcemia when receiving thiazide therapy. Serum calcium levels should be monitored for 2 months after initiation of thiazide therapy.

Amiloride The hypocalciuric action of the diuretic amiloride is probably achieved through secondary stimulation of a sodium-calcium antiporter at the epithelial cell basolateral membrane that leads to stimulation of calcium transport in the distal convoluted tubule. Amiloride may be particularly useful because it retards potassium excretion and prevents thiazide-induced hypokalemia and the associated hypocitraturia. Thiazide therapy is contraindicated or ineffective in approximately 7% to 15% of patients; amiloride can be used as an adjunct or substitute.

Neutral sodium or potassium phosphate Neutral sodium or potassium phosphate (nonacidic) is an effective therapy for hypercalciuria.[57] Phosphates reduce bone resorption of calcium and promote renal calcium reabsorption. Diarrhea is the most

common short-term side effect and is a frequent problem. These agents must be administered four times a day. Acid-containing phosphate salts should not be used, because the effects of acidosis on bone and kidney (hypocitraturia and hypercalciuria) may override the beneficial effects of the phosphate salts.

Cellulose phosphate Cellulose phosphate is a nonabsorbable ion exchange resin that binds calcium in the gastrointestinal tract. Cellulose phosphate has been advocated as an effective therapy for calcium stone disease, particularly in children who have shown inadequate response to thiazide therapy, in patients with renal failure induced by sarcoid-associated nephrocalcinosis, and in patients with hypercalcemia or hypercalciuria associated with administration of calcium and vitamin D supplements after parathyroidectomy.[58,59] The increased risk of secondary hyperoxaluria may negate the beneficial action of cellulose phosphate.

Potassium citrate and potassium bicarbonate Some patients with calcium-containing stones demonstrate a marked hypocitraturia rather than hypercalciuria. This abnormality may result from a metabolic disorder or from chronic metabolic acidosis. Studies have suggested that dietary supplementation with a citrate-containing salt in a dosage of 1 mEq/kg body weight raises the level of urinary citrate excretion. This effect results from the alkali load produced by citrate supplementation rather than from the direct appearance of ingested citrate in the urine. Citrate therapy effectively reduces the frequency of new stone formation.[60,61] Equivalent sodium bicarbonate therapy may achieve the same hypocitraturic effect but is generally not well tolerated because of gastric bloating or belching. Moreover, increased intake of sodium may adversely affect response to therapy by increasing urinary calcium excretion. It has been suggested that concurrent hypercalciuria and hyperuricosuria may underlie a particularly virulent form of recurrent nephrolithiasis. The treatment of hyperuricosuria with dietary purine restriction or with allopurinol may be beneficial if the usual treatment of hypercalciuria fails to alleviate the disorder.[34]

Treatment of Hyperoxaluria

In certain patients with hyperoxaluria, particularly those with inflammatory bowel disease, additional calcium intake, as calcium carbonate (500 mg of elemental calcium four times a day), may bind intestinal oxalate and reduce its absorption. Cholestyramine, an ion exchange resin, can also reduce oxalate absorption. Cholestyramine may be administered in a dosage of 4 g two to four times a day. This resin binds bile acid, the free form of which increases colonic absorption of oxalate by increasing its permeability across the intestinal wall. Cholestyramine may also bind oxalate in the intestinal lumen. Pyridoxine alters the cellular metabolism of glycine and, in a dosage of 25 to 50 mg/day, significantly reduces the urinary excretion of oxalate.[62] A trial of pyridoxine is indicated in cases in which the dietary restriction of oxalate and increase of calcium intake are not effective in reducing hyperoxaluria. Adjunctive therapy with magnesium oxide (400 mg once or twice daily with meals) may be useful because magnesium forms a soluble complex with oxalate.

Uric Acid Stones

Although relatively uncommon, uric acid stones may recur frequently. Pure uric acid stones primarily occur in patients in whom a persistently acid urine promotes uric acid precipitation. In addition, 10% to 20% of patients with gout produce uric acid, resulting in increased risk of uric acid stone formation. Obesity resulting from dietary indiscretion and probable purine gluttony appears to have a significant role in recurrent uric acid stone formation.[63] In men, gout is a significant risk factor for uric stone formation. A large prospective study found that the prevalence of kidney stone disease was almost twofold higher in men with a history of gout, even after adjusting for age and body mass index.[64] Low urinary pH (pH < 5.5) is the most important risk factor for uric acid nephrolithiasis; in normouricosuric stone disease, the primary defect appears to be in the renal excretion of ammonia and is linked to an insulin-resistant state.[65]

DIAGNOSIS

Although typically radiolucent, uric acid stones may be slightly radiopaque when mixed with or coated by calcium crystals. Urine pH actually has greater impact on stone formation than the urinary uric acid concentration. When the urine pH decreases from 6.0 to 5.0, the concentration of undissociated uric acid (as opposed to the urate concentration) increases sixfold, leading to precipitation and uric acid stone formation. The limit of solubility for undissociated uric acid in urine is only 90 mg/L. Because total excretion values of undissociated uric acid usually range between 500 and 600 mg/day, the presence of persistently acid urine is an important cause of uric acid stone formation.

Only 10% to 20% of uric acid stones are associated with hyperuricosuria. Typically, serum uric acid levels and urinary uric acid levels are normal in patients with recurrent uric acid stones. However, the excretion of more than 1,000 mg of uric acid per 24 hours in urine that has an average pH of 5.6 is probably a risk factor for uric acid stone formation. Uric acid stones are often discrete, but they may be large and extend between minor calyxes, producing a staghorn appearance.

DIFFERENTIAL DIAGNOSIS

Xanthine and 2,8-dihydroxyadenine stones are rare forms of radiolucent stones that may be mistaken for uric acid stones. The conditions that predispose to the development of xanthine or 2,8-dihydroxyadenine stones are inherited as autosomal recessive disorders. Xanthinuria, which may lead to xanthine stone formation, may also sometimes develop in patients who receive allopurinol as therapy for excessive uric acid production. The normal urinary xanthine excretory rate is 5.1 to 8.6 mg/24 hr. An excretory rate greater than 8.6 mg/24 hr is classified as xanthinuria. The presence of 2,8-dihydroxyadenine in the urine and gross reduction of the level of adenine phosphoribosyltransferase in erythrocytes allow for the diagnosis of 2,8-dihydroxyadenine stones.[66]

These two rare forms of radiolucent stones are virtually never detected by any means other than stone analysis.

TREATMENT

In patients with increased uric acid excretion, treatment options include dietary purine restriction, urinary alkalinization, and administration of allopurinol. Urinary alkalinization, which can be achieved with sodium bicarbonate or potassium citrate (1.0 to 1.5 mEq/kg/day) or their equivalent, substantially reduces the risk of further uric acid stone formation but may increase the risk of calcium phosphate stone formation. Patients must be instructed in the measurement of urinary pH with nitrazine paper and must strive to adjust the bicarbonate dosage

to maintain the pH value between 6.0 and 6.5.

Allopurinol, given in a dosage of 200 to 300 mg/day, is the preferred treatment for reducing uric acid excretion in patients with hyperuricosuria. Allopurinol decreases the risk of further stone formation and solubilizes existent uric acid stones. The side effects include rash and other hypersensitivity reactions; allopurinol therapy must therefore be carefully monitored.

In the majority of patients with uric acid stones, uric acid excretion is quantitatively normal. Uric acid stone formation is caused by a persistently acidic pH of the urine (pH < 5.5), which fails to increase in the immediate postprandial period. Impaired production of urinary buffers (ammonia and titratable acids) may underlie this defect. In these patients, alkali therapy is more likely to be effective. Allopurinol may also be useful in such patients if alkalinization and increased fluid intake fail to dissolve the stones.

Struvite Stones

Infection of the urinary tract with organisms that secrete the ectoenzyme urease can greatly increase the urinary ammonia concentration and pH. *Proteus* is the most common cause of such infection; other organisms include *Pseudomonas, Klebsiella, Providencia, Serratia,* and staphylococci. When urea is deaminated, NH3 is formed and binds protons, thereby alkalizing the urine. These conditions dramatically increase the risk of the formation of stones composed of ammoniomagnesium phosphate (struvite).

DIAGNOSIS

Characteristically, struvite stones are very large and often have a staghorn appearance. The stones are radiopaque, but to a lesser degree than those composed of calcium. The crystals have a so-called coffin-lid shape [*see Figures 2a through d*]. Urine as well as fragments of the stone, if available, should be cultured for microorganisms. If staghorn calculi caused by struvite stones are present bilaterally, renal failure can result (about a 30% chance over 8 years).[67] Because many patients with struvite stones also have an underlying metabolic disorder, such as hypercalciuria, a complete metabolic workup should be done once these stones have been detected.

TREATMENT

Preventing recurrences of struvite stones is difficult and requires successful sterilization of the urine. Because organisms may exist in the interstices of the calculus material, all such material must be removed from the urinary tract. A combination of percutaneous lithotripsy and ESWL can achieve favorable results. The procedure involves one or more sessions of passing a nephroscope into the renal pelvis and fragmenting the stone with an ultrasonic probe or pulsed dye laser, followed 2 to 3 days later by ESWL to break up smaller fragments. All visible calculus material can be successfully removed by the combination of nonsurgical stone removal and irrigation of the urinary tract with a solution of strong organic acids, such as hemiacidrin, a 10% citric acid compound.[68] This approach, used in conjunction with organism-specific, prolonged (up to 4 weeks) antimicrobial therapy, can result in complete cure of struvite stones in approximately 90% of treated patients if all stone fragments are removed during initial therapy. CT scanning without contrast can be used to document removal of all stones and for future follow-up.

Open surgical procedures such as anatrophic nephrolithotomy are still useful in the management of struvite stones because they can achieve success rates of 90% to 100%, even with larger stones, and cure is achieved in less time than with the protracted treatment protocols of percutaneous lithotripsy plus ESWL.[69]

In rare cases in which urologic procedures cannot be attempted, the use of medical therapy is indicated as a second-line approach. Acetohydroxamic acid is an orally active bacterial urease inhibitor.[70] This agent is effective in preventing stone growth even when stone fragments and bacterial infection persist. Unfortunately, acetohydroxamic acid has several significant side effects, including hemolytic anemia, intractable headache, and, possibly, thrombophlebitis.[71]

Cystine Stones

Cystinuria is an autosomal recessive disorder in which excessive urinary excretion of the dicarboxylic amino acids (cystine, ornithine, lysine, and arginine) results from impaired transport.[72] The genetic defect in 30% of cases studied is in six missense mutations in the human *rBAT* gene, which is involved in high-affinity transport of cystine and dibasic amino acids in the kidney and intestine.[73]

DIAGNOSIS

The normal rate of cystine excretion is 30 mg/day (1.3 mmol/day). In contrast, patients with cystinuria excrete more than 400 mg/day and sometimes up to 3,600 mg/day. The sulfur atom present in cystine accounts for the moderate radiopacity of cystine stones on plain films. The stones are discrete, but staghorn calculi can occur. Urinalysis may demonstrate the pathognomonic hexagonal crystals in approximately 50% of patients, but the urine sodium nitroprusside test is more useful for screening. The nitroprusside test is positive at a cystine concentration greater than 75 mg/g creatinine. If a qualitative test is positive for cystinuria, a 24-hour quantitative urinary cystine excretion rate should be determined. Because some patients with cystinuria will not present with symptoms until their fifth or even sixth decade, a diagnosis of cystinuria should not be excluded on the basis of a patient's age.

TREATMENT

Treatment of cystine stones consists of reducing the urinary cystine concentration by hydration, increasing cystine solubility by alkalinization, and lowering the urine cystine concentration with agents such as penicillamine or α-mercaptopropionyl glycine (tiopronin). Optimal therapy not only reduces the risk of new stone formation but also causes dissolution of existing stones.[74] The limit of cystine solubility in urine is approximately 300 mg/L (1.25 mmol/L). To prevent stone formation or growth in cystinuric patients excreting more than 1,000 mg of cystine a day requires a urine volume of at least 4 L/day spread uniformly over a 24-hour interval. Fluid therapy is usually unsuccessful because patients find it difficult to drink enough fluid throughout the day and night to achieve the necessary urine volume. Adolescents and young adults, in particular, may not be conscientious in maintaining a high fluid intake. When combined with the proper fluid intake, alkali therapy is beneficial because the solubility limit of cystine can be doubled (to 600 mg/L urine) by maintaining the urinary pH at 7.5 to 7.8. Such urine pH levels are difficult to achieve but may be attempted in highly motivated patients. Unfortunately, a major side effect of alkalinization is

the increased propensity to form calcium phosphate stones. A reduction of dietary sodium intake (50 mEq/day) may be helpful by increasing renal cystine reabsorption and thereby reducing the urinary burden.[74]

In some patients, therapy with penicillamine or related drugs (e.g., tiopronin)[60] may be required. These compounds form mixed, soluble disulfides with cysteine and thereby prevent cystine stone formation. Cystine production may also decrease because of an intracellular effect of the drug. Unfortunately, penicillamine must be withdrawn in as many as 90% of patients because of an adverse reaction. A comparative study showed that serious adverse effects requiring cessation of therapy were approximately 50% less frequent with tiopronin than with penicillamine.[75] Side effects from penicillamine range from insomnia and dysgeusia to a vasculitic syndrome and Goodpasture syndrome. Insomnia and dysgeusia can be prevented with zinc supplements. Pyridoxine supplements are required to prevent pyridoxine deficiency. The therapeutic benefit attributed to captopril (also a thiol-containing compound) is based on a very limited number of patients.[76] Subsequent studies support the preliminary observations of the efficacy of captopril, but rigorous controlled trials have not yet been performed.[77] Captopril leads to the formation of captopril-cysteine complexes, but the available amounts of captopril at usual doses may not be sufficient to effectively complex large amounts of cysteine.

Fluid therapy during the day combined with penicillamine or preferably tiopronin administration at night may be an effective approach in reducing the risk of cystine stone formation while minimizing the risk of penicillamine toxicity. A lower and, thus, less toxic dose of the drug is used in this protocol.

Drug-Induced Stones (Indinavir and Nelfinavir Stones)

Components of highly active antiretroviral therapies such as nelfinavir and particularly indinavir can contribute to the formation of urinary stones. These agents are excreted as urinary crystals that may result in crystal deposition or stone formation.[78] Nephrolithiasis is an increasingly recognized complication of indinavir therapy.[79]

Indinavir urolithiasis is unique in that CT, which was once thought to be efficacious in identifying all urinary calculi, is not useful in imaging stones that are composed of pure indinavir. Ultrasonography or intravenous urography may be required to confirm the diagnosis of pure indinavir stones in patients at risk for this complication.[80]

Prevention and treatment of indinavir-induced nephrolithiasis include greatly increased water intake and possibly urinary acidification.[81] Emergency drainage may be required for patients with severe obstruction.

Fuad N. Ziyadeh, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Stanley Goldfarb, M.D., has served as an adviser to Amgen, Inc.

References

1. Consensus Conference. Prevention and treatment of kidney stones. JAMA 260:977, 1988

2. Li J, Kennedy D, Levine M, et al: Absent hematuria and expensive computerized tomography: case characteristics of emergency urolithiasis. J Urol 165:782, 2001

3. Bove P, Kaplan D, Dalrymple N, et al: Reexamining the value of hematuria testing in patients with acute flank pain (pt 1). J Urol 162:655, 1999

4. Urivetzky M, Ravalli R, Weinberg J, et al: Biochemical evaluation of calcium stone patients: how soon can it be done after stone surgery/passage? Urology 36:410, 1990

5. Worster A, Preyra I, Weaver B, et al: The accuracy of noncontrast helical computed tomography versus intravenous pyelography in the diagnosis of suspected acute urolithiasis: a meta-analysis. Ann Emerg Med 40:280, 2002

6. Murray MJ, Chandhoke PS, Berman CJ, et al: Outcome of extracorporeal shockwave lithotripsy monotherapy for large renal calculi: effect of stone and collecting system surface areas and cost-effectiveness of treatment. J Endourol 9:9, 1995

7. Parks JH, Coe FL: An increasing number of calcium oxalate stone events worsens treatment outcome. Kidney Int 45:1722, 1994

8. Uribarri J, Oh MS, Carroll HJ: The first kidney stone. Ann Intern Med 111:1006, 1989

9. Glowacki LS, Beecroft ML, Cook RJ, et al: The natural history of asymptomatic nephrolithiasis. J Urol 147:319, 1992

10. Rivers K, Shetty S, Menon M: When and how to evaluate a patient with nephrolithiasis. Urol Clin North Am 27:203, 2000

11. Heneghan JP, Dalrymple NC, Verga M, et al: Soft-tissue "rim" sign in the diagnosis of ureteral calculi with use of unenhanced helical CT. Radiology 202:709, 1997

12. Saita H, Matsukawa M, Fukushima H, et al: Ultrasound diagnosis of ureteral stones: its usefulness with subsequent excretory urography. J Urol 140:28, 1988

13. Netto NR Jr, Lemos GC, De Almeida Claro JF, et al: Comparison between percutaneous nephrolithotomy and open stone procedures. Int Urol Nephrol 20:225, 1988

14. Liou LS, Streem SB: Long-term renal functional effects of shock wave lithotripsy, percutaneous nephrolithotomy and combination therapy: a comparative study of patients with solitary kidney. J Urol 166:36, 2001

15. Bushinsky DA: Nephrolithiasis. J Am Soc Nephrol 9:917, 1998

16. Bihl G, Meyers A: Recurrent renal stone disease: advances in pathogenesis and clinical management. Lancet 358:651, 2001

17. Curhan GC, Willett WC, Rimm EB, et al: Family history and risk of kidney stones. J Am Soc Nephrol 8:1568, 1997

18. Lloyd SE, Pearce SH, Fisher SE, et al: A common molecular basis for three inherited kidney stone diseases. Nature 379:445, 1996

19. Marangella M, Bagnis C, Bruno M, et al: Crystallization inhibitors in the pathophysiology and treatment of nephrolithiasis. Urol Int 72(suppl 1):6, 2004

20. Delvecchio FC, Preminger GM: Medical management of stone disease. Curr Opin Urol 13:229, 2003

21. Robertson WG: Diet and calcium stones. Miner Electrolyte Metab 13:228, 1987

22. Lemann J Jr, Gray RW, Maierhofer WJ, et al: The importance of renal net acid excretion as a determinant of fasting urinary calcium excretion. Kidney Int 29:743, 1986

23. Curhan GC, Willett WC, Knight EL, et al: Dietary factors and the risk of incident kidney stones in younger women. Nurses' Health Study II. Arch Intern Med 164:885, 2004

24. Nicar M, Skurla C, Sakhaee K, et al: Low urinary citrate excretion in nephrolithiasis. Urology 21:8, 1983

25. Goldfarb S: Dietary factors in the pathogenesis and prophylaxis of calcium nephrolithiasis. Kidney Int 34:544, 1988

26. Wasserstein A, Stolley P, Soper K: Case-control study of risk factors for idiopathic calcium nephrolithiasis. Miner Electrolyte Metab 13:85, 1987

27. Parivar F, Low RK, Stoller ML: The influence of diet on urinary stone disease. J Urol 155:432, 1996

28. Borghi L, Schianchi T, Meschi T, et al: Comparison of two diets for the prevention of recurrent stones in idiopathic hypercalciuria. N Engl J Med 346:77, 2002

29. Larsson L, Tiselius H-G: Hyperoxaluria. Miner Electrolyte Metab 13:242, 1987

30. Curhan GC, Willett WC, Speizer FE, et al: Intake of vitamins B_6 and C and the risk of kidney stones in women. J Am Soc Nephrol 10:840, 1999

31. Danpure C, Jennings P: Peroxisomal alanine:glyoxylate aminotransferase deficiency in primary hyperoxaluria type I. FEBS Lett 201:20, 1986

32. Mittal RD, Kumar R: Gut-inhabiting bacterium *Oxalobacter formigenes*: role in calcium oxalate urolithiasis. J Endourol 18:418, 2004

33. Melnick JZ, Srere PA, Elshourbagy NA, et al: Adenosine triphosphate citrate lyase mediates hypocitraturia in rats. J Clin Invest 98:2381, 1996

34. Pak CY, Poindexter JR, Peterson RD, et al: Biochemical distinction between hyperuricosuric calcium urolithiasis and gouty diathesis. Urology 60:789, 2002

35. Sarig S: The hyperuricosuric calcium oxalate stone former. Miner Electrolyte Metab 13:251, 1987

36. Coe FL, Parks JH, Asplin JR: The pathogenesis and treatment of kidney stones. N Engl J Med 327:1141, 1992

37. Miller LA, Stapleton FB: Urinary volume in children with urolithiasis. J Urol 141:918, 1989

38. Borghi L, Meschi T, Amato F, et al: Urinary volume, water and recurrences in idiopathic calcium nephrolithiasis: a 5-year randomized prospective study. J Urol 155:839, 1996

39. Nakagawa Y, Abram V, Parks JH, et al: Urine glycoprotein crystal growth inhibitors: evidence for a molecular abnormality in calcium oxalate nephrolithiasis. J Clin Invest 76:1455, 1985

40. Lieske JC, Hammes MS, Hoyer JR, et al: Renal cell osteopontin production is stimulated by calcium oxalate monohydrate crystals. Kidney Int 51:679, 1997

41. Curhan GC, Willett WC, Speizer FE, et al: Beverage use and risk for kidney stones in women. Ann Intern Med 128:534, 1998

42. Hirvonen T, Pietinen P, Virtanen M, et al: Nutrient intake and use of beverages and the risk of kidney stones among male smokers. Am J Epidemiol 150:187, 1999

43. Curhan GC, Willett WC, Rimm EB, et al: A prospective study of dietary calcium and

other nutrients and the risk of symptomatic kidney stones. N Engl J Med 328:833, 1993

44. Lemann J Jr: Composition of the diet and calcium kidney stones. N Engl J Med 328:880, 1993

45. Curhan GC, Willett WC, Speizer FE, et al: Comparison of dietary calcium with supplemental calcium and other nutrients as factors affecting the risk for kidney stones in women. Ann Intern Med 126:497, 1997

46. Churchill DN: Medical treatment to prevent recurrent calcium urolithiasis: a guide to critical appraisal. Miner Electrolyte Metab 13:294, 1987

47. Goldfarb DS, Coe FL: Prevention of recurrent nephrolithiasis. Am Fam Physician 60:2269,1999

48. Pearle MS, Roehrborn CG, Pak CY: Meta-analysis of randomized trials for medical prevention of calcium oxalate nephrolithiasis. J Endourol 13:679, 1999

49. Ohkawa M, Tokunaga S, Nakashima T, et al: Thiazide treatment for calcium urolithiasis in patients with idiopathic hypercalciuria. Br J Urol 69:571, 1992

50. Ettinger B, Citron JT, Livermore B, et al: Chlorthalidone reduces calcium oxalate calculous recurrence but magnesium hydroxide does not. J Urol 139:679, 1988

51. Brocks P, Dahl C, Wolf H, et al: Do thiazides prevent recurrent idiopathic renal calcium stones? Lancet 2:124, 1981

52. Scholz D, Schwille PO, Sigel A: Double-blind study with thiazide in recurrent calcium lithiasis. J Urol 128:903, 1982

53. Ziyadeh FN: Results of therapy of calcium renal calculi with the thiazides. Diuretics IV: Chemistry, Pharmacology, and Clinical Applications. Puschett JB, Ed. Excerpta Medica, Amsterdam, 1993, p 79

54. Pearle MS: Prevention of nephrolithiasis. Curr Opin Nephrol Hypertens 10:203, 2001

55. Pak C, Fuller C: Idiopathic hypocitraturic calcium-oxalate nephrolithiasis successfully treated with potassium citrate. Ann Intern Med 104:33, 1986

56. Reid IR, Ames RW, Orr-Walker BJ, et al: Hydrochlorothiazide reduces loss of cortical bone in normal postmenopausal women: a randomized controlled trial. Am J Med 109:362, 2000

57. Insogna K, Ellison A, Burtis W, et al: Trichlormethiazide and oral phosphate therapy in patients with absorptive hypercalciuria. J Urol 141:269, 1989

58. Burke J, Cowley D, Mottrem B, et al: Cellulose phosphate and chlorothiazide in childhood idiopathic hypercalciuria. Aust N Z J Med 16:43, 1986

59. Waron M, Weissgarten J, Gil I, et al: Sarcoid nephrocalcinotic renal failure reversed by sodium cellulose phosphate. Am J Nephrol 6:220, 1986

60. Pak C, Fuller C, Sakhaee K, et al: Long-term treatment of calcium nephrolithiasis with potassium citrate. J Urol 134:11, 1985

61. Sakhaee K, Poindexter JR, Griffith CS, et al: Stone forming risk of calcium citrate supplementation in healthy postmenopausal women. J Urol 172:958, 2004

62. Mitwalli A, Ayiomamitis A, Grass L, et al: Control of hyperoxaluria with large doses of pyridoxine in patients with kidney stones. Int Urol Nephrol 20:353, 1988

63. Ekeruo WO, Tan YH, Young MD, et al: Metabolic risk factors and the impact of medical therapy on the management of nephrolithiasis in obese patients. J Urol 172:159, 2004

64. Kramer JH, Choi HK, Atkinson K, et al: The association between gout and nephrolithiasis in men. The Health Professionals' Follow-Up Study. Kidney Int 64:1002, 2003

65. Parmar MS: Kidney stones. BMJ 328:1420, 2004

66. Deng L, Yang M, Frund S, et al: 2,8-Dihydroxyadenine urolithiasis in a patient with considerable residual adenine phosphoribosyltransferase activity in cell extracts but with mutations in both copies of APRT. Mol Genet Metab 72:260, 2001

67. Teichman JM, Long RD, Hulbert JC: Long-term fate and prognosis after staghorn calculus management. J Urol 153:1403, 1995

68. Bernardo NO, Smith AD: Chemolysis of urinary caculi. Urol Clin North Am 27:355, 2000

69. Assimos DG, Wrenn JJ, Harrison LH, et al: A comparison of anatrophic nephrolithotomy and percutaneous nephrolithotomy with or without extracorporeal shock-wave lithotripsy for management of patients with staghorn calculi. J Urol 145:710, 1991

70. Griffith DP, Gleeson MJ, Lee H, et al: Randomized, double-blind trial of Lithostat (acetohydroxamic acid) in the palliative treatment of infection-induced urinary calculi. Eur Urol 20:243, 1991

71. Wang LP, Wong HY, Griffith DP: Treatment options in struvite stones. Urol Clin North Am 24:149, 1997

72. Sakhaee K: Pathogenesis and medical management of cystinuria. Semin Nephrol 16:435, 1996

73. Langman CB: The molecular basis of kidney stones. Curr Opin Pediatr 16:188, 2004

74. Singer A, Das S: Cystinuria: a review of the pathophysiology and management. J Urol 142:669, 1989

75. Pak C, Fuller C, Sakhaee K, et al: Management of cystine nephrolithiasis with alpha-mercaptopropionylglycine. J Urol 136:1003, 1986

76. Perazella MA, Buller GK: Successful treatment of cystinuria with captopril. Am J Kidney Dis 21:504, 1993

77. Cohen TD, Streem SB, Hall P: Clinical effect of captopril on the formation and growth of cystine calculi. J Urol 154:164, 1995

78. Engeler DS, John H, Rentsch KM, et al: Nelfinavir urinary stones. J Urol 167:1384, 2002

79. Dieleman JP, Sturkenboom MC, Jambroes M, et al: Risk factors for urological symptoms in a cohort of users of the HIV protease inhibitor sulfate: the ATHENA cohort. Arch Intern Med 162:1493, 2002

80. Wu Ds, Stoller ML: Indinavir urolithiasis. Curr Opin Urol 10:557, 2000

81. Kopp JB, Miller KD, Mican JA, et al: Crystalluria and urinary tract abnormalities associated with indinavir. Ann Intern Med 127:119, 1997

Acknowledgment

Figure 1 Janet Betries.

XIII BENIGN PROSTATIC HYPERPLASIA

Michael J. Barry, m.d.

Benign prostatic hyperplasia (BPH) is a very common cause of morbidity in older men in virtually all developed countries. BPH causes lower urinary tract symptoms and occasionally results in such complications as acute urinary retention, urinary tract infection, and even obstructive uropathy. Although increasing appreciation of the relatively benign natural history of this condition and the development of effective medical treatments have reduced the role of surgery, prostatectomy for BPH remains a widely performed procedure. The sheer prevalence of lower urinary tract symptoms attributable to BPH, the impact of these symptoms on patients' quality of life, and the availability of medication for this condition make BPH important in both primary care and urologic practice.

Epidemiology and Risk Factors

Autopsy studies indicate that the prevalence of histologic BPH ranges from about 25% for men in their 40s to about 85% for men older than 80 years [*see Figure 1*].[1] The average prostate volume in young men without histologic evidence of BPH is about 20 ml. A community-based study of men from Olmsted County, Minnesota, has indicated that mean prostate volume increases with age: 23 ml at 40 to 49 years of age, 27 ml at 50 to 59, 32 ml at 60 to 69, and 39 ml at 70 to 79.[2] These volumes were higher in every age category than for men in a Japanese community, where mean prostate volumes ranged from 17 ml for men in their 40s to 21 ml for men in their 70s. In another community-based study, from the Netherlands, the prevalence of a prostate volume greater than 40 ml ranged from 15% in men 55 to 59 years of age to 28% for those 60 to 64, 36% for those 65 to 69, and 55% for those 70 to 74.[3]

Other epidemiologic studies have examined the relationship of peak urinary flow (uroflow) rates (Q_{max}) and age. In the Olmsted County study, mean Q_{max} decreased from about 20 ml/sec in men 40 to 44 years of age to about 12 ml/sec at 75 to 79 years of age.[4] The proportion of men with a flow rate less than 10 ml/sec ranged from about 5% at 40 to 44 years of age to about 25% at 70 to 74 years of age. Another parameter sometimes linked to BPH, postvoid residual volume (PVR), does not appear to be age related.

Of course, BPH without symptoms is more a histologic curiosity than a disease, and not all lower urinary tract symptoms are attributable to BPH. What is the prevalence of symptomatic BPH in older men? The answer depends on how BPH is defined. A Dutch study of a community-based sample of men 55 to 74 years of age found that the prevalence of BPH varied from 4.3% to 19%, depending on which of eight case definitions of clinically important disease was applied.[5] By any definition, however, BPH is a common clinical complaint in older men, at least in Western countries.

Although the prevalence of histologic BPH is remarkably similar across countries with populations of varying races and ethnicities, Asian men appear to have smaller prostates at any given age and undergo less surgery for BPH. Whites and African Americans appear to have similar risks. Studies conflict on whether there may be a positive or negative association between BPH and cigarette smoking. Obesity also has an equivocal relationship with BPH. Cirrhosis of the liver appears to be negatively associated with BPH, probably because of higher circulating levels of estrogens relative to androgens. Men with a family history of BPH may have an increased risk of BPH, especially if the affected relative was diagnosed before age 60, and twin studies indicate a 3.3-fold higher risk in monozygotic twins with affected siblings. Despite the apparent contribution of family history as a risk factor, the genetics of BPH remains poorly understood.[6]

Pathophysiology

The normal prostate consists of epithelial glandular components in a stroma of connective tissue and smooth muscle [*see Figure 2*]. BPH involves hyperplasia of both the epithelial and the stromal compartments. The hyperplastic process begins in the periurethral and transition zones of the prostate; in contrast, prostate cancer preferentially develops in the peripheral zones.[7] Over time, multiple small hyperplastic nodules grow and coalesce into a large central adenoma, which can compress the more peripheral prostate tissue against the fibrous prostatic capsule. This compressed peripheral tissue can create what is in effect a surgical capsule bounding the central adenoma. The capsule serves as the plane of resection when a particularly large adenoma is removed at open prostatectomy.

Eventually, prostatic hyperplasia can result in bladder outlet obstruction, which is defined as increased bladder detrusor pressure relative to the rate of urine flow. Bladder obstruction has both a static component (from luminal narrowing secondary to prostate tissue growth) and a dynamic component (from increased neuromuscular tone in the hyperplastic prostate). Neuromuscular tone in the bladder neck, prostatic capsule, and prostatic stroma is attributable to a relatively high density of $alpha_2$-adrenergic receptors in these tissues. At least three types of $alpha_2$-adrenergic receptors have been identified in the lower urinary tract. Epithelial and vascular tissues, on the other hand, are more richly supplied with $alpha_1$-adrenergic receptors. The bladder

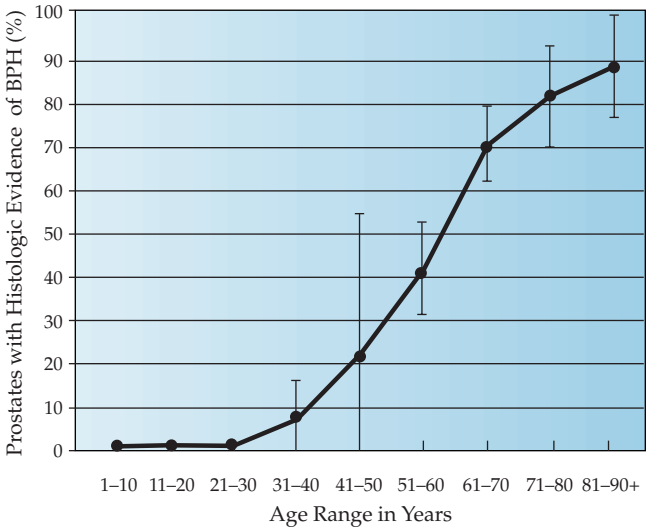


Figure 1 **Prevalence of BPH histology with age in 1,075 human autopsies.[1]**

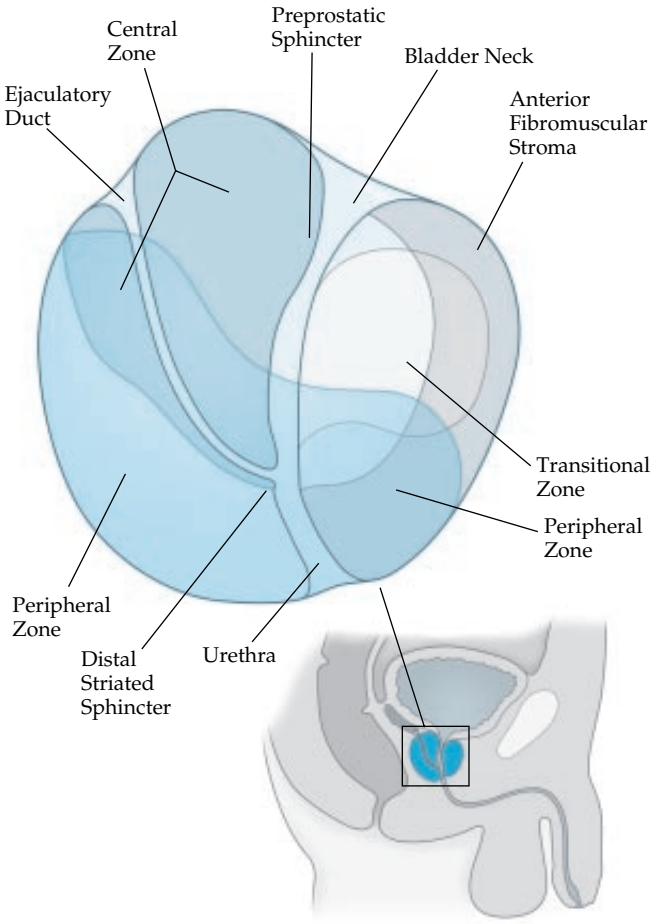


Figure 2 **Sagittal diagram of the prostate and its relationship to the anteromedial nonglandular tissues, along with a three-dimensional representation of the glandular prostate.**

may respond to outlet obstruction with hypertrophy, uninhibited contractions, and, if obstruction is long-standing, fibrosis.

These histologic, anatomic, and pathophysiologic processes may result in lower urinary tract symptoms. Traditionally, such symptoms have been grouped into two categories: bladder emptying (obstructive) symptoms, which presumably result primarily from mechanical obstruction, and bladder filling (irritative) symptoms, which presumably result primarily from detrusor instability (see below).[8]

Although this model of the link between BPH and lower urinary tract symptoms is convenient, it is undoubtedly simplistic. Researchers have documented that the severity of lower urinary tract symptoms correlates poorly with measurements of prostate size, the degree of bladder outlet obstruction, and the severity of uninhibited detrusor contractions. Moreover, improvement in symptoms with treatment does not correlate well with improvements in these ostensibly objective parameters; for some treatment modalities, symptoms may improve despite the absence of changes in these other measures. The scientific uncertainty over the mechanisms that link the histologic process of BPH with lower urinary tract symptoms contributes to clinical uncertainty over how to treat older men who present with bothersome lower urinary tract symptoms.

The pathophysiologic mechanisms underlying the development and progression of BPH are also incompletely understood. Aging and the presence of functioning testes through puberty into adulthood are absolute requirements. Clearly, BPH involves prolonged exposure of the prostate gland to androgens. The conversion of testosterone to dihydrotestosterone (DHT), which is the major intraprostatic androgen, is accomplished by the type 2 5α-reductase isoenzyme; this isoenzyme occurs largely in prostatic stromal cells, although it is also found in some peripheral tissues. In contrast, the type 1 5α-reductase isoenzyme has low activity in the prostate and is expressed predominantly in the skin and liver. DHT serves as a powerful mitogen for prostatic cells. Men who were castrated before puberty do not develop BPH, and medical or surgical castration later in life causes apoptosis of prostatic epithelial cells. Other factors extrinsic to the prostate also appear to be important in prostate growth, including estrogens, possibly a nonandrogenic testicular factor, and environmental factors such as diet and genetics. In the prostate, interactions between epithelial and stromal cells and the extracellular matrix, mediated primarily by locally produced (intrinsic) growth factors, appear important. These peptide growth factors, which include fibroblast growth factors, insulinlike growth factors, and epidermal growth factors, are felt to be the local forces that determine prostate growth.[9]

Diagnosis

The extent of the medical evaluation necessary to establish a working clinical diagnosis of lower urinary tract symptoms attributable to BPH is controversial. In most men older than 50 years, a characteristic history alone may be sufficient. However, not all older men with a typical history will have either prostatic enlargement or evidence of bladder outlet obstruction on pressure-flow studies. In some of these men, environmental factors or systemic diseases may be causing or exacerbating their symptoms. Others may have disorders of the lower urinary tract that are not associated with bladder outlet obstruction, or they may have bladder outlet obstruction that is not caused by BPH.

CLINICAL FEATURES

Slowly progressive difficulties with bladder emptying and filling are characteristic of BPH. Bladder emptying symptoms in men with BPH include straining, hesitancy, intermittency, a weak stream, terminal dribbling, and a sensation of incomplete emptying. Bladder filling symptoms include daytime frequency, nocturia, urgency, and urge incontinence.

HISTORY

When older men present with lower urinary tract symptoms, a careful history will uncover most alternative causes. The physician should look for evidence of systemic diseases that can present with lower urinary tract symptoms, particularly urinary frequency and nocturia. Examples of such diseases include diabetes, heart failure (which can cause a nocturnal diuresis separate from the effect of diuretics), and hyperparathyroidism. Neurologic diseases—including diabetic peripheral and autonomic neuropathy, stroke, and Parkinson disease—may be associated with neuropathic disorders of the bladder. A complete list of medications should be obtained, and the time at which the patient takes medications such as diuretics should be ascertained. Patients should be specifically questioned about the use of over-the-counter medications for cough, cold, and allergy, because the anticholinergic and sympathomimetic agents in these preparations may cause or exacerbate symptoms.

The urologic history should include any genitourinary malignancies (personal or family), hematuria, or acute urinary retention or urinary tract infection. Patients should be asked about urethral strictures, as well as genitourinary trauma, instrumentation, and sexually transmitted diseases, which may result in stricture.

Lower urinary tract symptoms can be quantified using the American Urological Association Symptom Index (AUASI).[10] This seven-item questionnaire, which is widely used in epidemiologic and clinical research, as well as clinical practice, assesses the frequency of seven lower urinary tract symptoms, each on a scale of 0 to 5. The individual item scores can be added to calculate an overall score ranging from 0 (best) to 35 (worst). Scores of less than 8 reflect mild symptoms; 8 to 18, moderate symptoms; and 19 to 35, severe symptoms. The International Prostate Symptom Score (IPSS)[11] contains the same symptom question set as the AUASI, but it also includes a separately scored quality-of-life question addressing the degree of bother caused by the respondent's symptoms [*see Figure 3*]. The IPSS questionnaire is designed for self-administration, but visually impaired patients or patients who cannot read may have an interviewer read the questionnaire to them. Quantifying the severity of symptoms provides objective documentation of the patient's condition and serves as a baseline against which to judge improvement, stability, or worsening over time, with or without treatment. Responses to the individual items, as well as the overall score, should be noted.

Men with symptoms attributable to BPH usually have a balance of filling and voiding symptoms; men with mostly filling symptoms, and particularly nocturia alone, are more likely to have another cause for their symptoms. Patients should also be asked about the time course of the development of their symptoms. Rapid onset of symptoms is uncharacteristic of BPH and should be a red-flag warning of other potential causes, including genitourinary malignancy. Dysuria and pelvic pain are also uncommon in men with BPH; these symptoms are more often seen with prostatitis.[12]

Finally, because BPH and its treatment can affect both sexual function and urinary continence, men should be asked about these issues before any treatment is begun. Urge incontinence may be seen with BPH as a result of uninhibited detrusor contractions; other types of incontinence may be engendered by treatment.

Some clinicians ask patients with lower urinary tract symptoms to keep a voiding diary over a few days to a week.[13] Patients should measure and record the volume of each void, noting any episodes of incontinence. The pattern of voiding reflected in the diary may provide valuable clues to the cause of symptoms, particularly when it reveals nocturnal polyuria.

Perhaps the most important aspect of assessing lower urinary tract symptoms is to ask the patient how much he is bothered by them. Although the level of bother is generally related to symptom severity, some patients tolerate severe symptoms quite well,

	Not at All	*Less Than One Time in Five*	*Less Than Half the Time*	*About Half the Time*	*More Than Half the Time*	*Almost Always*
1. Over the past month or so, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
2. Over the past month or so, how often have you had to urinate again less than 2 hours after you finished urinating?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
3. Over the past month or so, how often have you found you stopped and started again several times when you urinated?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
4. Over the past month or so, how often have you found it difficult to postpone urination?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
5. Over the past month or so, how often have you had a weak urinary stream?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
6. Over the past month or so, how often have you had to push or strain to begin urination?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐

7. Over the past month, how many times did you most typically get up to urinate from the time you went to bed at night until the time you got up in the morning?

0 ☐ none	1 ☐ one time	2 ☐ two times	3 ☐ three times	4 ☐ four times	5 ☐ five or more times

Total I-PSS Score = Sum of Questions 1–7 = __ __

Quality of Life Due to Urinary Symptoms

If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about that?

Delighted	Pleased	Mostly Satisfied	Mixed (about Equally Satisfied and Dissatisfied)	Mostly Dissatisfied	Unhappy	Terrible
0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐

Figure 3 **The International Prostate Symptom Score (IPSS).[11] The seven symptom questions are scored separately from the eighth question about symptom bother.**

whereas other patients are aggravated by moderate symptoms. Intervention may make more sense for the latter than the former, regardless of symptom score. The patient's answer to the question about bother on the IPSS may serve as a good entry into this discussion [*see Figure 3*].

PHYSICAL EXAMINATION

Patients considering alpha-blocker therapy for BPH should have their baseline blood pressure determined, both supine and standing. Patients considering operative treatment should undergo a general evaluation for risks of complications from anesthesia and surgery.

The physician should perform a targeted physical examination that includes palpation of the lower abdomen for evidence of a grossly enlarged bladder, assessment of the extremities for peripheral neuropathy, and assessment of the perineal area and inner thighs for saddle anesthesia. A patient with neuropathy or saddle anesthesia is more likely to have a neurologic disorder affecting the bladder.

A digital rectal examination (DRE) should be done, and the examiner should note the size, symmetry, and texture of the prostate. BPH characteristically produces a symmetrically enlarged, firm prostate of the consistency of the tip of the nose, but that finding is nonspecific. Asymmetry in consistency or hard nodules suggest prostate cancer, but these findings are likewise not specific.

DRE tends to produce an underestimation of prostate size.[14] When a clinician thinks a prostate is enlarged, it usually is; however, the prostate may well be enlarged even if the clinician thinks it is not. This discrepancy may be important, because for some medical therapies, efficacy is related to prostate size. Moreover, agreement between clinicians, even urologists, on the presence or absence of palpable prostate abnormalities suggesting cancer has also been shown to be only fair.[15]

LABORATORY TESTING

Routine tests performed on men with lower urinary tract symptoms should generally include a urinalysis to screen for hematuria and infection. Many clinicians order serum creatinine and prostate-specific antigen (PSA) tests, although the value of those is arguable for many patients. Less frequently performed tests in this setting include PVR measurement, uroflow measurement, and imaging of the upper urinary tract.

Urinalysis

Pyuria suggests infection, either primary or superimposed on bladder outlet obstruction. Microscopic hematuria may indicate simply that the prostate is enlarged and vascular, but it should prompt further evaluation for genitourinary malignancy.

Serum Creatinine Measurement

An elevated creatinine level may be a clue to the occasional case of obstructive uropathy. However, the rarity of this condition has recently led to the publication of practice guidelines that recommend against routine measurements of serum creatinine in men with lower urinary tract symptoms suggestive of BPH.[16]

Prostate-Specific Antigen Testing

The role of PSA testing in men with lower urinary tract symptoms has been controversial. Prostate cancer screening studies do not confirm that patients with lower urinary tract symptoms suggestive of BPH are more likely to harbor prostate cancer.[17] However, because BPH and prostate cancer are both age related, they may coexist. Therefore, looking for prostate cancer in men with lower urinary tract symptoms and a nonsuspicious DRE is most appropriately considered screening, rather than case finding. The diagnostic yield of PSA measurement as a screening test deteriorates in the setting of BPH, primarily because of lower specificity. Although a PSA level greater than 10 ng/ml is still relatively specific for prostate cancer, only about 25% of men with PSA levels in the range of 4.1 to 10 ng/ml have prostate cancer at biopsy; many of the rest have BPH.[18]

Most urologists doubt the value of early detection of prostate cancer in men whose life expectancy is less than 10 years[19] or who are in average health but older than about 75 years. Clinicians should have a conversation about the possibility of coexisting prostate cancer with patients who present with lower urinary tract symptoms, and the clinician and the patient should decide together whether to pursue such a diagnosis.

A PSA test result of over 4 ng/ml has traditionally been considered an indication for biopsy. However, a number of studies have reported that when men with so-called high-normal PSA levels (2.5 to 4.0 ng/ml) undergo biopsy, only 15% to 20% are found to have prostate cancer.[18] Many attempts have been made to find laboratory features that would indicate the likelihood of cancer in patients with an elevated PSA level, and so improve biopsy yield. In general, however, the performance of these so-called PSA derivatives has been disappointing. For example, measurement of the proportion of total PSA that is circulating in the free form has been proposed for men with total PSA levels of 4.1 to 10 ng/ml and benign DRE results, with a proportion above 25% considered to offer reassurance that cancer is absent. However, only about 20% of men with total PSA levels in this range will have a supposedly reassuring free PSA percentage; and even in those men, the probability of cancer at biopsy is still about 10%.[20]

Another potential reason to consider ordering a PSA test for a man with lower urinary tract symptoms is to help stratify his risk for future problems, such as progression to acute urinary retention or a condition warranting surgery.[21] A PSA level can also help predict the preventive effect of medical therapy with a 5α-reductase inhibitor (see below).

Residual Volume Measurement

An optional test in the setting of lower urinary tract symptoms is the assessment of PVR by catheterization or ultrasonography. Obtaining such measurements is difficult in primary care settings, and substantial day-to-day variation limits their value. Probably only consistent, gross elevations in PVR (> 350 ml) are a reliable indicator of the need for more aggressive treatment.

Urinary Flow Rate

Many urologists measure uroflow rates in their offices. About 90% of men with peak uroflow rates of less than 10 ml/sec have bladder outlet obstruction, compared with about 30% of men with peak flow rates greater than 15 ml/sec. Values between 10 and 15 ml/sec are associated with bladder outlet obstruction in about two thirds of cases.[22] Unfortunately, uroflow rate testing is also limited by within-patient variability, and multiple measurements are necessary to improve reliability. Also, a voided volume of at least 150 ml is needed for a reliable result, and a substantial proportion of patients with lower urinary tract symptoms cannot meet that requirement. Moreover, low flow rates can be seen with weak detrusor contraction in the absence of

outlet obstruction, and strong bladder contraction can maintain flow even in the presence of obstruction. For all of these reasons, uroflowmetry is an optional test.

The simultaneous measurement of bladder pressure and urine flow is considered the gold standard for the diagnosis of bladder outlet obstruction.[23] There are a number of criteria for calling a pressure-flow study positive, all of which are based, in various ways, on documenting a high bladder pressure relative to the uroflow rate. More sophisticated videourodynamics can localize the site of obstruction. These tests require special equipment and training; are somewhat invasive, requiring a urethral catheter; can be expensive; and yield equivocal results in some men. They are usually recommended for patients with atypical presentations; for patients with neurologic diseases, in whom the probability of primary bladder disorders is higher; or for men with persistent symptoms after surgical therapy for symptoms presumed to be from BPH. Few urologists would recommend a pressure-flow study for a typical patient before a trial of medical therapy, but some would routinely recommend such a study before undertaking a prostatectomy after medical therapy fails. There is not universal agreement regarding this recommendation, however, because a substantial proportion of men with lower urinary tract symptoms who do not have unequivocal evidence of obstruction on a pressure-flow study will nevertheless have good symptomatic responses to surgery.[24]

Upper Urinary Tract Imaging

Upper urinary tract imaging (by ultrasonography, computed tomography, or intravenous pyelography) and urethrocystoscopy are not indicated for routine cases of lower urinary tract symptoms attributable to BPH.[25] Cystoscopic findings in particular have not correlated well with the severity of symptoms of bladder outlet obstruction. Such studies should be reserved for selected patients with findings such as microscopic hematuria.

Management

Once the clinician has made a working clinical diagnosis of symptomatic BPH and when the patient is sufficiently bothered by his lower urinary tract symptoms to at least consider therapy, the initial choice of management is usually between a period of watchful waiting and a course of medical therapy. Medical therapies include alpha blockers and 5α-reductase inhibitors. Many patients also use alternative or complementary treatment for prostate problems, especially phytotherapies.

WATCHFUL WAITING

When the bother of a patient's lower urinary tract symptoms is not sufficient, in his mind, to justify the potential side effects, costs, and inconvenience of medication, watchful waiting is the optimal management strategy. Although virtually all patients with mild symptom scores (an IPSS from 0 to 7 points) choose this course, it is also a reasonable course for patients with moderate or severe symptoms who are not very bothered by them. Patients should be advised to take lifestyle measures to improve symptoms, such as avoiding fluids, particularly beverages containing caffeine, just before bedtime. They should also avoid offensive medications, especially nonprescription ones [*see* History, *above*]. Finally, they should report worsening symptoms, particularly an increase in their feeling bothered by the symptoms. Although, on average, symptoms can be expected to gradually increase, individual courses are quite variable, and some men note improvement over time [*see Figure 4*].[26]

MEDICAL TREATMENT

Alpha Blockers

Alpha$_1$-adrenergic blockers work primarily through relaxation of prostatic smooth muscle and relief of the dynamic component of bladder outlet obstruction. However, additional mechanisms have been proposed, including increased apoptosis of prostatic cells. Alpha blockers neither reduce prostate size nor lower PSA levels. Their onset of action is relatively rapid, although most alpha blockers require dose titration to achieve a maximal therapeutic effect while minimizing side effects. Many patients have hypertension and BPH, and both problems may be treated with an alpha blocker; however, data strongly suggest that alpha-blocker monotherapy is not optimal treatment for hypertension.[27]

Many randomized trials have documented the efficacy of alpha blockers over placebo in studies lasting up to 4 years.[28,29] For example, in the largest trial reported to date, over 2,000 men were randomized to receive either placebo or terazosin in doses

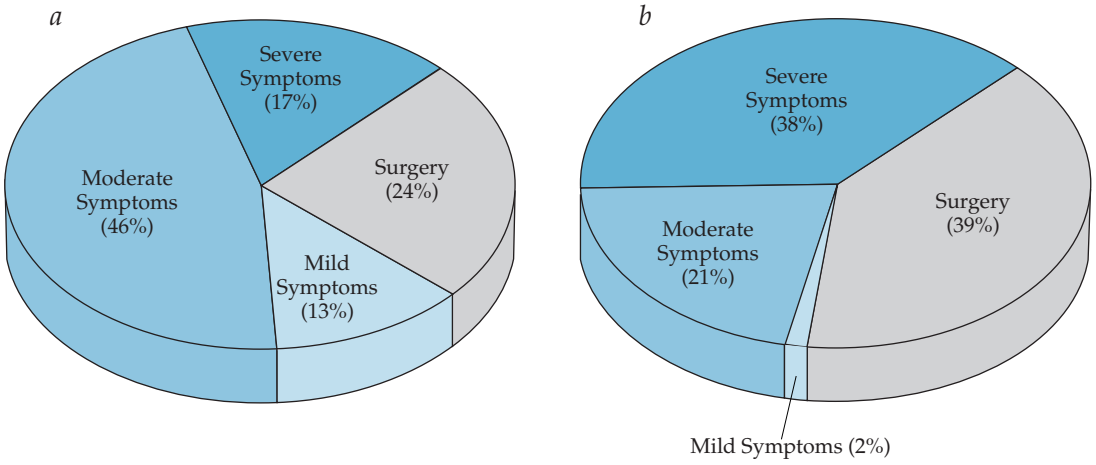


Figure 4 **Four-year outcomes in men with (*a*) moderate (N = 245) or (*b*) severe (N = 66) lower urinary tract symptoms attributed to benign prostatic hypertrophy. These men were considered reasonable surgical candidates but chose watchful waiting instead.[26]**

Table 1 Alpha Blockers Commonly Used in the Treatment of Men with Lower Urinary Tract Symptoms Attributed to BPH

Drug	*Strengths Available*	*Recommended Dose-Titration Steps*
Alfuzosin*	2.5 mg	2.5 mg t.i.d.
Alfuzosin SR*	5 mg	5 mg b.i.d.
Alfuzosin XL	10 mg	10 mg q.d.
Doxazosin	1, 2, 4, and 8 mg	1, 2, 4, and 8 mg q.d.
Doxazosin GITS*	4 and 8 mg	4 and 8 mg q.d.
Tamsulosin	0.4 mg	0.4 and 0.8 mg q.d.
Terazosin	1, 2, 5, and 10 mg	1, 2, 5, and 10 mg q.d.

*Not available in the United States.
GITS—Gastrointestinal Therapeutic System

of up to 10 mg daily. IPSS scores dropped about 8 points with the alpha blocker and about 4 points with placebo.[30] Such data highlight the impressive magnitude of the placebo effect on lower urinary tract symptoms in most trials of therapy. Some of this placebo effect may be attributable to regression toward the mean, which is seen when men with particularly marked symptoms are recruited into trials. In clinical practice, patients perceive their change in symptoms from baseline, so their improvement results from a combination of the pharmacologic benefit of the intervention with these other effects. The therapeutic effect of the alpha blockers is independent of prostate size.[31] As is true of all treatments for men with symptoms attributed to BPH, the men with the highest symptom scores can expect the greatest absolute decrease in their symptom scores with alpha blockers. Therefore, many men with symptom scores in the severe range will benefit from a therapeutic trial of alpha blockers before considering more invasive treatment.

Various alpha blockers are used in the treatment of lower urinary tract symptoms attributed to BPH. Although head-to-head trials of alpha blockers are relatively few, the therapeutic effects of different agents appear similar.[28] There is currently great interest in developing alpha blockers with selectivity for the subtypes of the $alpha_1$-adrenergic receptor to maximize the therapeutic effect while minimizing side effects. Tamsulosin and alfuzosin indeed appear more selective than doxazosin and terazosin for the $alpha_{1a}$-adrenergic receptor, which is the predominant subtype in prostatic smooth muscle. However, this difference in pharmacologic selectivity does not appear to translate into major differences in efficacy.[28,32]

Side effects of alpha blockers include orthostatic hypotension, dizziness, and asthenia. The last two side effects are not primarily mediated by low blood pressure.[33] Tamsulosin and alfuzosin do not affect blood pressure, as do the other alpha blockers, but it is unclear whether the risk of side effects other than orthostatic hypotension is appreciably reduced with these agents.[16,28] Abnormal ejaculation has been reported in about 10% of men on tamsulosin.[16] In short-term clinical trials, the proportion of men who withdraw from alpha-blocker treatment because of side effects is generally low, but this proportion is likely to be higher in clinical practice, particularly over time. Caution must be exercised in combining alpha blockers with phosphodiesterase type 5 inhibitors used for the treatment of erectile dysfunction, which commonly coexists with BPH, for fear of inducing hypotension. Sildenafil should not be taken within 4 hours of taking an alpha blocker; vardenafil should not be taken at all by men on alpha blockers; and tadalafil should be combined only with tamsulosin at the lower dose of 0.4 mg daily.

Nonselective $alpha_1$-adrenergic blockers such as doxazosin and terazosin require dose titration [*see Table 1*]. Most clinicians start with a 1 mg dose at bedtime for several days to avoid so-called first-dose hypotension, although the clinical relevance of this phenomenon is poorly documented. In most patients, the dose should be pushed to the highest level if possible; the clinician should monitor supine and standing blood pressure, side effects, and the patient's global impressions of improvement before dose escalation. Follow-up administration of the IPSS may help objectify improvements and aid in the decision whether to proceed to the highest recommended dose. Both doxazosin and terazosin can be given once daily. Many clinicians recommend taking them at bedtime, although it is unclear whether this strategy reduces side effects.

Tamsulosin, extended-release doxazosin (the doxazosin Gastrointestinal Therapeutic System [GITS]), and a new extended-release formulation of alfuzosin are designed for once-daily administration and do not require as many dose-titration steps as doxazosin and terazosin. Titration of tamsulosin and doxazosin GITS is complicated by the fact that they have two recommended dose levels. Doxazosin GITS is not yet available in the United States.

Although alpha blockers have repeatedly been shown to reduce lower urinary tract symptoms, their ability to reduce the risk of BPH complications over the long term, such as progression to surgery or the development of acute urinary retention, appears negligible.[29]

5α-Reductase Inhibitors

The 5α-reductase inhibitors currently available for the treatment of BPH are finasteride and dutasteride. Finasteride selectively and irreversibly binds with the type 2 5α-reductase isoenzyme, which predominates in the prostate, and thereby blocks conversion of testosterone to DHT, the dominant intraprostatic androgen. This agent lowers serum DHT by about 70% and intraprostatic DHT to an even greater degree. Dutasteride is a dual 5α-reductase inhibitor; it blocks both type 1 and type 2 isoenzymes and lowers serum DHT by about 90%.[34] Whether this dual blockade offers any clinical advantage is uncertain.

Men who take finasteride at the recommended dose of 5 mg daily or dutasteride at 0.5 mg daily can expect a 20% to 25% reduction in prostate size over the first year of therapy,[34,35] accompanied by about a 50% reduction in PSA level.[34,36] Despite this effect on PSA, these drugs have not been shown to impair the ability to diagnose prostate cancer in controlled trials. However, PSA levels must be interpreted differently. The most common recommendation is simply to double the measured PSA levels for men on finasteride or dutasteride, then interpret the results as usual.

Like the alpha blockers, finasteride and dutasteride have been shown in a number of clinical trials to produce greater symptom reduction than does placebo. However, the symptom reduction has generally been smaller than for the alpha blockers, averaging 2 to 4 points from baseline IPSS scores.[34,37] The efficacy of finasteride, reflected in the group mean difference between finasteride and placebo in these trials, depends on prostate size. For men with prostates less than 40 ml, the differences between fi-

nasteride and placebo are minimal, but the differences in IPSS scores between finasteride and placebo increase to 1.0 to 1.5 points for men with prostates greater than 40 ml in trials lasting up to 2 years.[37] Dutasteride's efficacy relative to prostate size has not been as well studied.

In a longer-term trial, finasteride resulted in about a 3-point decrease in symptom score from baseline over 4 years, roughly 2 points greater than the decrease shown with placebo.[35] The slow response of symptoms to finasteride was obvious, with nadir scores not reached for at least 2 years. Finasteride also significantly reduced the risk of acute urinary retention (from 6.6% to 2.8%) and of progression to prostatectomy (from 8.3% to 4.2%). In a 2-year trial, the effects of dutasteride on the relative risk of acute urinary retention and on progression to surgery were similar.[34] Data from the finasteride trial indicated that about 15 patients would have to take finasteride for 4 years to prevent the development of acute retention or the progression to surgery in one of those patients. This trial also documented that this preventive benefit of finasteride depends on prostate size and, furthermore, that serum PSA level could be used as a proxy for prostate size to estimate the expected benefit.[38] The estimated number of men who would need to be treated for 4 years to prevent one episode of acute retention or progression to surgery is about 29 for baseline PSA levels less than 1.4 ng/ml, 18 for PSA levels of 1.4 to 3.2 ng/ml, and 9 for PSA levels greater than 3.2 ng/ml.[21]

Finasteride also seems to be effective at reducing the risk of recurrent hematuria in men presenting with hematuria that appears to be from BPH.[39] This effect is probably attributable to finasteride-induced suppression of vascular endothelial growth factors.

Finasteride is relatively well tolerated. About 5% to 10% of men notice decreased libido or ejaculatory abnormalities over the first year of treatment, and fewer than 1% of men have rashes, gynecomastia, or breast tenderness. After the first year of treatment, side effects with finasteride and placebo are essentially identical.[35] The side-effect profile of dutasteride is similar.[34]

In a large cancer prevention trial, the 7-year cumulative likelihood of a prostate cancer diagnosis was 18.4% with finasteride and 24.4% with placebo.[40] The likelihood in both trial arms was higher than expected because of intensive surveillance for prostate cancer during the trial, including end-of-study biopsies (regardless of DRE findings or PSA levels) in most participants. However, the likelihood of a high-grade prostate cancer (defined as a cancer with a Gleason score of 7 to 10) increased from 5.1% with placebo to 6.4% with finasteride. These results raise doubt about whether finasteride would do more good than harm in terms of its effect on prostate cancer morbidity and mortality. This should be considered when prescribing finasteride (and, presumably, dutasteride, which works by the same mechanism) to men for BPH.

Combination Therapy

Given that alpha blockers and 5α-reductase inhibitors work through different mechanisms, combination therapy is an attractive concept. Nevertheless, in head-to-head comparative trials, both terazosin and alfuzosin used alone were found to be superior to finasteride for symptom relief over 6 to 12 months of therapy,[41,42] and the combination of finasteride with terazosin or alfuzosin was found to be no better than either alpha blocker alone in the short term. However, in a subsequent 4-year trial, doxazosin, finasteride, and combination therapy were compared with placebo in terms of their ability to reduce the rate of BPH progression, defined primarily as 4-point increases in IPSS score. The risk of progression at 4 years was 17% with placebo, 10% with either doxazosin or finasteride, and 5% with both.[29] Patients who take combination therapy are exposed to the side effects of both agents, as well as to any higher risk of more aggressive prostate cancers from the use of finasteride.

Phytotherapy

Many patients around the world take phytotherapies for lower urinary tract symptoms. They may take these preparations on their own initiative or at the recommendation of a clinician. Meta-analyses suggest that a number of these phytotherapies, including extracts of the saw palmetto plant,[43] β-sitosterol plant extracts,[44] rye grass pollen extract,[45] and *Pygeum africanum* plant extracts,[46] are effective for reducing some lower urinary tract symptoms. The effects appear to be small, and the trials are generally limited by design problems, particularly short durations of follow-up and incomplete outcome assessments. In particular, few studies used validated symptom scores as outcome measures. The mechanisms of any effects are poorly defined. Moreover, preparations of these phytotherapeutic agents vary, so patients are not assured that any given preparation will afford the same benefits seen in the trials.

SURGERY AND MINIMALLY INVASIVE TREATMENT

Transurethral and Open Prostatectomy

Transurethral prostatectomy (TURP) remains the gold standard for relieving symptoms and reducing the risk of complications for men with BPH. TURP involves resecting the central adenoma of the hyperplastic prostate transurethrally under direct visualization using a resectoscope with an electrified cutting loop. General or spinal anesthesia is usually used.[47] In the United States, patients may have as little as an overnight hospital stay after TURP and may require a urinary catheter for a short period until the initial hematuria induced by the resection begins to clear.

In a randomized trial that compared TURP and watchful waiting, mean IPSS scores fell by about 12 points from baseline with TURP versus 1 point with expectant management by 7.5 months after randomization.[48] Although there are no randomized trials comparing medical therapy with TURP, the decreases in symptom scores with TURP are substantially greater than with any medication. In the largest trial of TURP versus expectant management for men with moderate symptoms, which involved almost 600 men, treatment failures (predominantly deterioration to severe symptoms, acute retention, or the development of a very large residual volume) were significantly reduced by about 50% with TURP.[49] On the other hand, the absolute risks of treatment failure were relatively small—over 3 years, the rate of treatment failure was about 20% with watchful waiting versus 10% with TURP. However, these results must be interpreted in light of the fact that about a quarter of the patients initially assigned to expectant management crossed over to surgery within 3 years.

With the availability of other effective and less invasive treatment options, age-adjusted rates of TURP in Medicare beneficiaries fell by about half from 1984 to 1997.[50] However, the procedure got progressively safer over this period; currently, 30-day mortality is less than 1%. Moreover, TURP provides durable symptom relief, with a 5-year risk for reoperation of about 5%.[50] Although sexual dysfunction and incontinence have been traditionally considered potential side effects of TURP, the risks were

found to be similar for TURP and watchful waiting in the largest comparative trial.[49] Retrograde ejaculation, however, is a common outcome of TURP, occurring in the majority of cases. Men should be warned to expect this result when considering TURP.

A prostatic adenoma can also be resected through an open incision using either a retropubic or a suprapubic approach.[51] To avoid the incision and potential complications such as wound infection, open prostatectomy is generally reserved for men with the largest prostates, in whom a transurethral resection may be especially challenging. An open prostatectomy can also be expected to provide excellent and durable symptom relief.

Transurethral Incision and Electrovaporization

Transurethral incision of the prostate (TUIP) involves making one or two longitudinal incisions in the prostate without resecting tissue. This procedure has been considered especially well suited to younger men with smaller prostates. In a meta-analysis of four small trials, symptom relief at 12 months with TURP was similar to that with TUIP.[52] However, long-term comparative results are not available. The risk of retrograde ejaculation was 73% with TURP but only 21% with TUIP.

Transurethral electrovaporization of the prostate (TUVP) is similar to TURP, but TUVP uses a roller electrode to vaporize superficial layers of prostate tissue while coagulating deeper layers. A meta-analysis of several small trials suggested that the primary benefit of TUVP is fewer bleeding complications, but this advantage has not been determined definitively.[53] The longest trial (5 years) suggested that TURP and TUVP had similar effects on symptoms, but few patients completed 5 years of follow-up.[54] Longer-term comparative results are unknown.

Laser Coagulation and Vaporization

A number of treatment strategies using laser energy to coagulate or vaporize prostate tissue have been explored for men with symptoms attributed to BPH. An early technique of noncontact coagulation has been totally abandoned. In a meta-analysis, short-term symptom responses to a spectrum of laser procedures were somewhat less impressive than for TURP, with higher risks of reoperation, though with fewer transfusions and postoperative strictures.[55] Longer-term trials comparing visual laser ablation of the prostate (VLAP) and contact laser vaporization with TURP have also shown higher retreatment rates with the laser strategies over 3 to 5 years.[56,57] A newer, holmium laser ablation technique appeared to have results comparable to those of TURP in a 1-year trial,[58] but longer comparative studies are needed.

Thermal Therapies

In recent years, a number of minimally invasive treatments have been developed to coagulate prostate tissue with heat generated through various mechanisms. These office-based treatments are designed to reduce symptoms, with less sedation and fewer side effects than seen with surgery. No tissue is resected with thermal therapies, so their mechanism of action is uncertain. Transurethral microwave therapy (TUMT) uses a microwave antenna placed in the urethra to generate heat. The urethra, which is pain sensitive, is protected by a cooling jacket. TUMT reduced symptoms more than terazosin in an 18-month trial[59]; however, in small trials comparing TUMT and TURP, symptomatic outcomes were somewhat less impressive with TUMT.[53] There were fewer side effects related to bleeding with TUMT. Transurethral needle ablation (TUNA) uses two radiofrequency needles placed directly into the prostate via the urethra to heat and coagulate tissue. In one comparative trial, symptom reduction was somewhat less with TUNA than with TURP at 1 year. However, both bleeding and retrograde ejaculation were less common with TUNA.[60] Long-term outcomes are poorly defined for either of these heat-based treatments.

MANAGEMENT OF ACUTE URINARY RETENTION

Acute urinary retention occurs at a rate of 1% to 2% a year in men with lower urinary tract symptoms attributed to BPH. The occurrence of acute urinary retention used to be considered an absolute indication for surgery. However, small case series have documented that up to half of men with acute retention have a successful voiding trial after a period of bladder rest via catheter drainage, and most of the men who experience success will continue to void, at least over the next 6 months. The optimal duration of catheter drainage is poorly defined, with recommended periods ranging from a few days to several weeks. Studies of whether the administration of alpha blockers increases the likelihood of a successful voiding trial have provided conflicting results.[61,62]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Berry SJ, Coffey DS, Walsh PC, et al: The development of human benign prostatic hyperplasia with age. J Urol 132:474, 1984

2. Masumori N, Tsukamoto T, Kumamoto Y, et al: Japanese men have smaller prostate volumes but comparable urinary flow rates relative to American men: results of community based studies in 2 countries. J Urol 155:1324, 1996

3. Bosch J, Hop W, Niemer A, et al: Parameters of prostate volume and shape in a community based population of men 55 to 74 years old. J Urol 152:1501, 1994

4. Girman C, Panser L, Chute C, et al: Natural history of prostatism: urinary flow rates in a community-based study. J Urol 150:887, 1993

5. Bosch J, Hop WC, Kirkels WJ, et al: Natural history of benign prostatic hyperplasia: appropriate case definition and estimation of its prevalence in the community. Urology 46:34, 1995

6. Boyle P, Liu G, Jacobsen S, et al: Epidemiology and natural history. Benign Prostatic Hyperplasia (5th International Consultation on BPH). Chatelain C, Denis L, Foo K, et al, Eds. Health Publication, Ltd (distributed by Plymbridge Distributors, Ltd, Plymouth, United Kingdom), 2001, p 19

7. McNeal J: Normal histology of the prostate. Am J Surg Pathol 12:619, 1988

8. Abrams P: New words for old: lower urinary tract symptoms for "prostatism." BMJ 308:929, 1994

9. Lee C, Cockett A, Cussenot O, et al: Regulation of prostate growth. Benign Prostatic Hyperplasia (5th International Consultation on BPH). Chatelain C, Denis L, Foo K, et al, Eds. Health Publication, Ltd. (distributed by Plymbridge Distributors, Ltd, Plymouth, UK), 2001, p 81

10. Barry MJ, Fowler FJ Jr, O'Leary MP, et al: The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. J Urol 148:1549, 1992

11. Barry M, Batista J, Donovan J, et al: Measuring the symptoms and health impact of benign prostatic hyperplasia and its treatments. Benign Prostatic Hyperplasia (5th International Consultation on BPH). Chatelain C, Denis L, Foo K, et al, Eds. Health Publication, Ltd (distributed by Plymbridge Distributors, Ltd, Plymouth, United Kingdom), 2001, p 203

12. Collins MM, Stafford R, O'Leary M, et al: Distinguishing chronic prostatitis and benign prostatic hyperplasia symptoms: results of a national survey of visits. Urology 53:921, 1999

13. Abrams P, Klevmark B: Frequency volume charts: an indispensable part of lower urinary symptom assessment. Scand J Urol Nephrol Suppl 179:47, 1996

14. Roehrborn C, Girman C, Rhodes T, et al: Correlation between prostate size estimated by digital rectal examination and measured by transrectal ultrasound. Urology 49:548, 1997

15. Smith D, Catalona W: Interexaminer variability of digital rectal examination in detecting prostate cancer. Urology 45:70, 1995

16. AUA guideline on management of benign prostatic hyperplasia (2003). Chapter 1: Diagnosis and treatment recommendations. AUA Practice Guidelines Committee. J Urol 170:530, 2003

17. Catalona W, Richie J, Ahmann F, et al: Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. J Urol 151:1283, 1994

18. Barry M: Prostate-specific-antigen testing for early diagnosis of prostate cancer. N Engl J Med 344:1373, 2001

19. Fowler FJ, Bin L, McNaughton Collins M, et al: Prostate cancer screening and beliefs about treatment efficacy: a national survey of primary care physicians and urologists. Am J Med 104:526, 1998

20. Catalona W, Partin A, Slawin K, et al: Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease. JAMA 279:1542, 1998

21. Roehrborn C, McConnell J, Lieber M, et al: Serum prostate-specific antigen concentration is a powerful predictor of acute urinary retention and need for surgery in men with clinical benign prostatic hyperplasia. Urology 53:473, 1999

22. Abrams P, Griffith D, Hofner K, et al: The urodynamic assessment of lower urinary tract symptoms. Benign Prostatic Hyperplasia (5th International Consultation on BPH). Chatelain C, Denis L, Foo K, et al, Eds. Health Publication, Ltd (distributed by Plymbridge Distributors, Ltd, Plymouth, United Kingdom), 2001, p 229

23. Abrams P: In support of pressure-flow studies for evaluating men with lower urinary tract symptoms. Urology 44:153, 1994

24. McConnell J: Why pressure-flow studies should be optional and not mandatory studies for evaluating men with benign prostatic hyperplasia. Urology 44:156, 1994

25. McConnell JD, Barry MJ, Bruskewitz RC, et al: Benign prostatic hyperplasia: diagnosis and treatment. Clinical Practice Guideline, Number 8. Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, Rockville, Maryland, 1994

26. Barry MJ, Fowler FJ, Bin L, et al: The natural history of patients with benign prostatic hyperplasia as diagnosed by North American urologists. J Urol 157:10, 1997

27. Major cardiovascular events in hypertensive patients randomized to doxazosin vs chlorthalidone: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Collaborative Research Group. JAMA 283:1967, 2000

28. Djavan B, Marberger M: A meta-analysis on the efficacy and tolerability of alpha$_1$-adrenoceptor antagonists in patients with lower urinary tract symptoms suggestive of benign prostatic obstruction. Eur Urol 36:1, 1999

29. McConnell JD, Roehrborn CG, Bautista OM, et al: The long-term effect of doxazosin, finasteride, and combination therapy on the clinical progression of benign prostatic hyperplasia. N Engl J Med 349:2385, 2003

30. The Hytrin Community Assessment Trial study: a one-year study of terazosin versus placebo in the treatment of men with symptomatic benign hyperplasia. HYCAT Investigator Group. Urology 47:159, 1996

31. The impact of medical therapy on bother due to symptoms, quality of life and global outcome, and factors predicting response. Veterans Affairs Cooperative Studies Benign Prostatic Hyperplasia Study Group. J Urol 160:1358, 1998

32. Kaplan S: Uroselective alpha-blockade for benign prostatic hyperplasia: clinically significant or marketing savvy? Urology 54:776, 1999

33. Lepor H, Jones K, Williford W: The mechanism of adverse events associated with terazosin: an analysis of the Veterans Affairs Cooperative Study. J Urol 163:1134, 2000

34. Roehrborn CG, Boyle P, Nickel JC, et al: Efficacy and safety of a dual inhibitor of 5-alpha reductase types 1 and 2 (dutasteride) in men with benign prostatic hyperplasia. Urology 60:434, 2002

35. McConnell JD, Bruskewitz R, Walsh P, et al: The effect of finasteride on the risk of acute urinary retention and the need for surgical treatment among men with benign prostatic hyperplasia. N Engl J Med 338:557, 1998

36. Treatment with finasteride preserves usefulness of prostate-specific antigen in the detection of prostate cancer: results of a randomized, double-blind, placebo-controlled clinical trial. Proscar Long-term Efficacy and Safety Study. PLESS Study Group. Urology 52:195, 1998

37. Boyle P, Gould AL, Roehrborn CG: Prostate volume predicts outcome of treatment of benign prostatic hyperplasia with finasteride: meta-analysis of randomized clinical trials. Urology 48:398, 1996

38. Roehrborn C, Boyle P, Gould A, et al: Serum prostate-specific antigen as a predictor of prostate volume in men with benign prostatic hyperplasia. Urology 53:581, 1999

39. Foley S, Soloman L, Wedderburn A, et al: A prospective study of the natural history of hematuria associated with benign prostatic hyperplasia and the effect of finasteride. J Urol 163:496, 2000

40. Thompson IM, Goodman PJ, Tangen CM, et al: The influence of finasteride on the development of prostate cancer. N Engl J Med 349:215, 2003

41. Lepor H, Williford WO, Barry MJ, et al: The efficacy of terazosin, finasteride, or both in benign prostatic hyperplasia. N Engl J Med 335:533, 1996

42. Sustained-release alfuzosin, finasteride and the combination of both in the treatment of benign prostatic hyperplasia. European ALFIN Study Group. Eur Urol 34:169, 1998

43. Wilt TJ, Ishani A, Stark G, et al: Saw palmetto extracts for the treatment of benign prostatic hyperplasia: a systematic review. JAMA 280:1604, 1998

44. Wilt T, MacDonald R, Ishani A: Beta-sitosterol for the treatment of benign prostatic hyperplasia: a systematic review. BJU International 83:976, 1999

45. MacDonald R, Ishani A, Rutks I, et al: A systematic review of cernilton for the treatment of benign prostatic hyperplasia. BJU International 85:836, 2000

46. Ishani A, MacDonald R, Nelson D, et al: *Pygeum africanum* for the treatment of patients with benign prostatic hyperplasia: a systematic review and quantitative meta-analysis. Am J Med 109:654, 2000

47. Mebust W: Transurethral surgery. Campbell's Urology. Walsh P, Ed. WB Saunders Co, Philadephia, 1998, p 1511

48. Donovan J, Peters T, Neal D, et al: A randomized trial comparing transurethral resection of the prostate, laser therapy and conservative treatment of men with symptoms associated with benign prostatic enlargement: the CLasP study. J Urol 164:65, 2000

49. Wasson J, Reda D, Bruskewitz R, et al: A comparison of transurethral surgery with watchful waiting for moderate symptoms of benign prostatic hyperplasia. N Engl J Med 332:75, 1995

50. Wasson J, Bubolz T, Lu-Yao G, et al: Transurethral resection of the prostate among Medicare beneficiaries: 1984–1997. J Urol 164:1212, 2000

51. Oesterling J: Retropubic and suprapubic prostatectomy. Campbell's Urology. Walsh P, Ed. WB Saunders Co, Philadelphia, 1998, p 1529

52. Yang Q, Peters T, Donovan J, et al: Transurethral incision compared with transurethral resection of the prostate for bladder outlet obstruction: a systematic review and meta-analysis of randomized controlled trials. J Urol 165:1526, 2001

53. Wheelahan J, Scott N, Cartmill R, et al: Minimally invasive non-laser thermal techniques for prostatectomy: a systematic review. BJU International 86:977, 2000

54. Hammadeh MY, Madaan S, Hines J, et al: 5-Year outcome of a prospective, randomized trial to compare transurethral electrovaporization of the prostate and standard transurethral resection. Urology 61:1166, 2003

55. Hoffman RM, MacDonald R, Slaton JW, Wilt TJ: Laser prostatectomy versus transurethral resection for treating benign prostatic obstruction: a systematic review. J Urol 169:210, 2003

56. McAllister W, Absalom M, Mir K, et al: Does endoscopic laser ablation of the prostate stand the test of time? Five-year results from a multicentre randomized controlled trial of endoscopic laser ablation against transurethral resection of the prostate. BJU International 85:437, 2000

57. Keoghane SR, Lawrence KC, Gray A, et al: A double blind randomized controlled trial and economic evaluation of transurethral resection vs contact laser vaporization for benign prostatic enlargement: a 3-year follow-up. BJU International 85:74, 2000

58. Gilling PJ, Mackey M, Cresswell M, et al: Holmium laser versus transurethral resection of the prostate: a randomized prospective trial with 1-year followup. J Urol 162:1640, 1999

59. Djavan B, Seitz C, Roehrborn C, et al: Targeted transurethral microwave thermotherapy versus alpha-blockade in benign prostatic hyperplasia: outcomes at 18 months. Urology 57:66, 2001

60. Bruskewitz R, Issa M, Roehrborn C, et al: A prospective, randomized 1-year clinical trial comparing transurethral needle ablation to transurethral resection of the prostate for the treatment of symptomatic benign prostatic hyperplasia. J Urol 159:1588, 1998

61. McNeill S, Daruwala P, Mitchell I, et al: Sustained release of alfuzosin and trial without catheter after acute urinary retention: a prospective, placebo-controlled trial. BJU International 84:622, 1999

62. Shah T, Paylit V, Biyani S, et al: Randomised, placebo controlled, double blind study of alfuzosin SR in patients undergoing trial without catheter following acute urinary retention. European Urology 42:329, 2002

Acknowledgments

Figure 2 Tom Moore.

Figure 3 International Consultation on BPH.

I THE DIZZY PATIENT

David Solomon, m.d., ph.d.
Elliot M. Frohman, m.d., ph.d.

Introduction

Dizziness is an inherently neurologic symptom, even if a systemic cause is primarily responsible. Although dizziness is common to diverse diseases, careful evaluation of dizziness and of any accompanying signs and symptoms can provide the gateway to diagnosis, which in turn is the touchstone of successful treatment.

Vestibular Physiology and Anatomy

Dizziness and the postural instability that often accompanies dizziness typically represent abnormalities of inner-ear function. Vestibular sensation and motor responses are mediated by two relatively simple reflexes: the vestibulo-ocular reflex (VOR) and the vestibulospinal reflex (VSR). VOR maintains visual fixation by generating eye movements to counter head movement in space (e.g., when the head moves upward, the eyes move downward). VSR keeps the head and body upright through changes in axial muscle tone.

Within the vestibular labyrinth in each ear, there are three semicircular canals (SCCs) that respond to angular acceleration of the head in three dimensions [*see Figure 1*]. When the head turns, the resulting deflection of the cupula in the ampulla of each SCC generates a signal that is transmitted to the brain via the vestibular portion of cranial nerve VIII. The strength of that signal is proportional to the angular velocity of the head. Information about rotation of the head in space reaches the vestibular nuclear complex in the brain stem [*see Figure 2*], and output from this complex drives the compensatory slow phases of the angular VOR.

The otolith organs (utricle and saccule) respond to linear acceleration during translation of the head (motion in a straight line) and from gravity (uprightness). Linear accelerations cause movement of the gelatinous matrix in the otolithic macula, deflecting the otolith hair cell processes. As these hairs bend, their firing rates change, depending on the direction of acceleration [*see Figure 3*]. Input from the otoliths drives the responses to head tilt and translation for the compensatory slow phases of the linear VOR.

Transmission of sensory information from the vestibular organs (SCCs and otoliths) occurs through the transduction of physical force into changes in electrical potential by hair cells in the sensory neuroepithelium of each organ.[1] Cupular deflection in the SCCs during angular acceleration and otolithic membrane movement within the otolith organs from linear acceleration alter the membrane potential of hair cells, depending on whether the bundle is deflected toward (depolarization) or away (hyperpolarization) from the kinocilium.[2] Generator potentials of the hair cells then modulate the firing rate of primary vestibular afferent fibers in cranial nerve VIII. These fibers have a resting spontaneous firing rate and, therefore, are active even when the head is not in motion. Rotation of the head toward any given SCC causes nerves innervating that canal to increase their firing rate; rotation away from a canal causes a decrease in firing in axons from that canal.

The SCCs and the otoliths detect motion in three dimensions to provide information for the VSR, which maintains balance. Visual (optokinetic) and somatosensory pathways also provide partially redundant and complementary signals for balance, enabling recovery in response to vestibular dysfunction. Optokinetic signals supplant labyrinthine signals for the low-frequency sustained components of the VOR to head rotation. These different

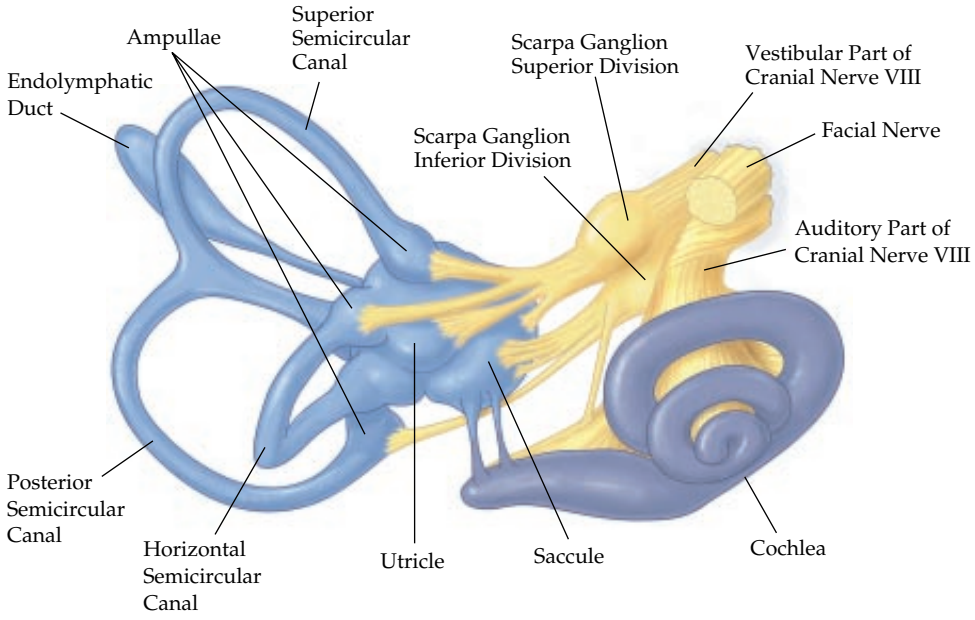


Figure 1 **Soft tissue structures within the bony labyrinth.**

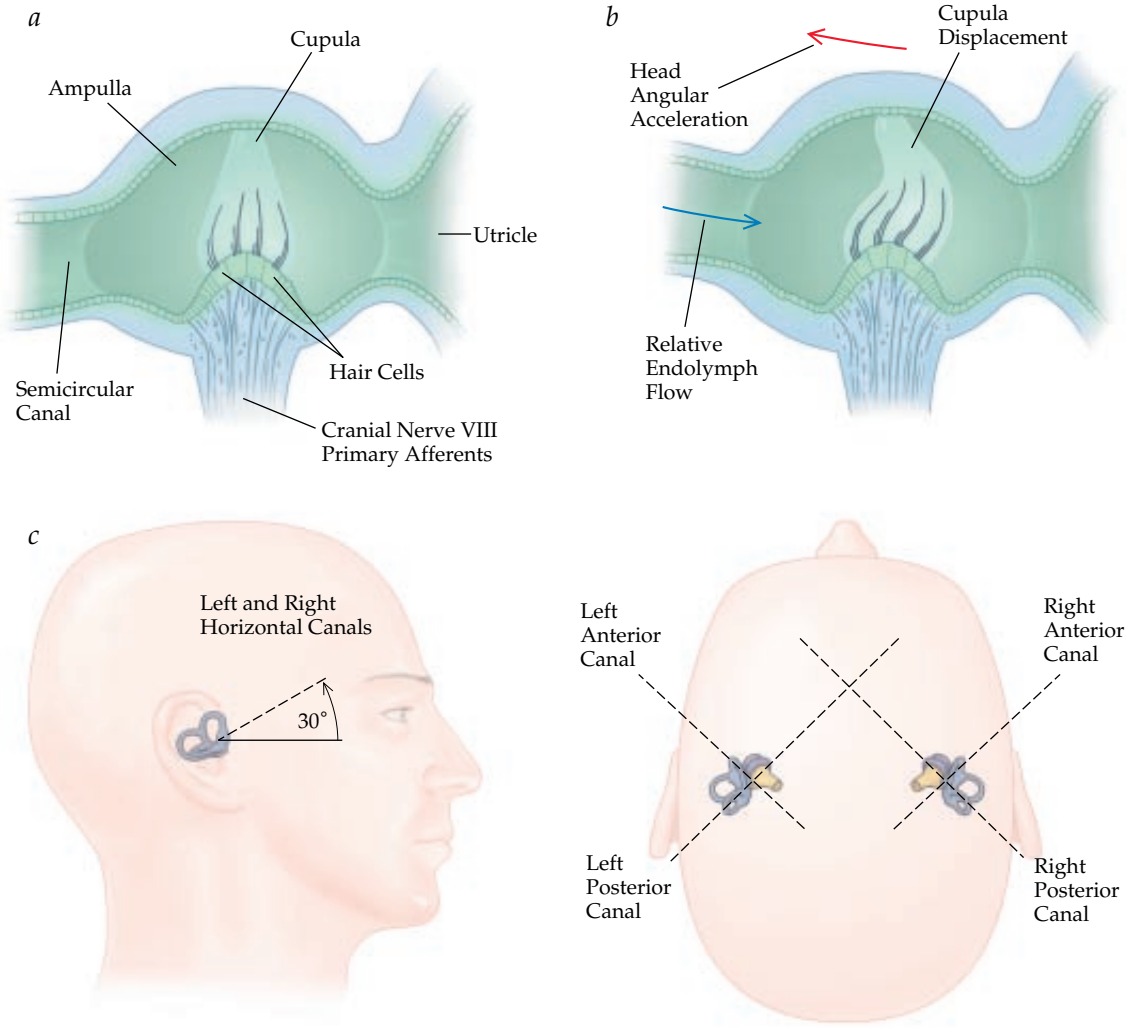


Figure 2 **(*a*) Shown is the functional organization of the cupula and the associated hair cells within the ampulla of the semicircular canals. The cupula changes position depending on the direction of angular acceleration. (*b*) The relation between the hair cell stereocilia and kinocilium is also shown. Movement of the stereocilia (cupula displacement) in a direction toward the kinocilium results in depolarization of hair cell axons, which project through cranial nerve VIII into the brain stem vestibular pathways. Movement of stereocilia in the opposite direction results in hyperpolarization. Similar anatomic arrangements are seen in the saccule and utricle. (*c*) The location of the semicircular canals in the head is shown. In the lateral semicircular canals, the kinocilium is located on the side nearest the utricle, whereas in the superior and posterior semicircular canals, the kinocilium is away from the utricle.**

somatosensory and visual signals that help stabilize posture converge on the same neurons in the vestibular nuclei. Thus, a disturbance in any of the sensory inputs to the vestibular nuclei can lead to a sensory conflict and abnormal vestibular sensations. On occasion, dizziness is produced by visual-vestibular mismatches. A classic example is car sickness: for example, while one reads in the back seat of a moving car, information processed in the visual system informs the brain that the body is not in motion; however, the vestibular system provides information to the contrary. The central neural circuitry of these reflexes is complex. Connections between the vestibular neurons and the oculomotor nuclei form the substrate for the VOR [*see Figure 4*].

For the VSR, the SCCs contribute to postural reflexes via projections to the lateral vestibular nucleus, giving rise to the lateral vestibulospinal tract. The vestibular contribution to sensations of head and body motion and spatial position likely reaches consciousness through the rostral projections of vestibular nuclei to the cerebellum, thalamus, and cerebral cortex.

Clinical Manifestations of Vestibular Dysfunction

Vestibular dysfunction can present in a variety of ways, including oscillopsia (the visual illusion of environmental movement), postural instability, and spatial disorientation. Patients often describe difficulty driving, walking in large open spaces, or walking in crowded environments such as shopping malls and supermarkets. They will avoid head movement. Sensory-rich environments provoke and intensify sensory conflicts, which can produce feelings of disorientation and vegetative symptoms such as nausea and vomiting.[3]

On examination, patients complaining of vertigo may exhibit nystagmus, as a result of abnormalities in the VOR. Vestibular nystagmus has a slow phase, in which the eyes drift during attempted steady fixation, followed by a fast (or quick) phase, in which the eyes reset on the fixation target through rapid eye movements similar to saccades (the eye movements normally made from place to place with the head still or while reading). Nystagmus may be spontaneous or induced by changes in

gaze or body position; the distinction between spontaneous and induced nystagmus, as well as the direction of the nystagmus (i.e., horizontal, torsional, or vertical), has considerable diagnostic utility [*see Tables 1 and 2*]. By convention, nystagmus is named by the direction of the quick phase.

Vestibular signs and symptoms may occur when the head is still, when the head moves, or both. Static disturbances suggest a different set of diagnostic possibilities than dynamic disturbances.

STATIC IMBALANCE

Vestibular disturbance in the absence of head motion is characterized by vertigo and spontaneous nystagmus, which results when a unilateral lesion disturbs the normal balance of tonic discharges from the two labyrinths. Other clinical features of static vestibular disturbance include diplopia (double vision); tilt of the head and body; and vegetative symptoms such as nausea, vomiting, diaphoresis, and occasionally hypotension and syncope.

Skew deviation, a vertical misalignment of the eyes, is the hallmark of an imbalance in the tonic levels of activity underlying otolith-ocular reflexes. Patients with this misalignment often complain of vertical diplopia, sometimes with the illusion of tilt of the visual world, and the head may also be tilted. Skew deviation can be detected clinically by using an ocular cover test. In this test, the physician moves a cover from one of the patient's eyes to the other while watching for vertical corrective eye movement when the cover is switched.

Skew deviation, ocular counterrolling (rotation of the eyes about the line of sight; ocular torsion), and head tilt together constitute the ocular tilt reaction.[4] The ocular tilt reaction can occur with lesions anywhere in the otolith-ocular pathway; this pathway includes the peripheral labyrinth,[5] vestibular nerve, vestibular nucleus in the medulla, medial longitudinal fasciculus in the pons or midbrain, and interstitial nucleus of Cajal [*see Figure 4*]. With peripheral and vestibular nucleus lesions (e.g., vestibular nerve section or Wallenberg syndrome), the lower eye is on the side of the lesion. With lesions above the level of otolith-ocular pathway decussation at the vestibular nucleus (e.g., an internuclear ophthalmoplegia from a lesion in the medial longitudinal fasciculus in the pons or midbrain), the higher eye is on the side of the lesion and the head is tilted toward the lower eye.[6]

Abnormalities in the VSR are assessed with the tandem walking test: the patient places one foot directly in front of the other while keeping the head relatively fixed; the test is done first with the eyes open and then with the eyes closed. Problems with tandem walking suggest an imbalance in vestibular tone—patients will generally veer or fall toward the paretic side.

Static imbalance should also be assessed with the Romberg test, in which the patient stands with feet together and closes the eyes while holding the arms and hands outward with palms up. Excessive sway on the Romberg test can indicate abnormalities in otolith-spinal reflexes.

Past-pointing of the arms or feet to previously seen targets with eyes closed may be another sign of vestibulospinal imbalance. For the arms, past-pointing is best elicited by having the patient repetitively raise both arms over the head, with the index fingers extended, and then bring them down, with eyes closed, toward the examiner's index fingers located at waist level. Errors in pointing generally are in the direction of the weak labyrinth and should be equal in the two arms.

In some cases, the characteristics of spontaneous nystagmus can help localize the lesion. When vestibular damage is peripheral, the nystagmus is characteristically diminished by visual fixation and increased when fixation is eliminated; thus, it is best observed by techniques that impair or remove visual fixation (e.g., occlusive ophthalmoscopy, Frenzel glasses).[7] The slow phases of spontaneous nystagmus generally move toward the side of a peripheral vestibular lesion. For example, in the case of a left vestibular neuritis, right vestibular function is relatively increased; consequently, in such cases, patients exhibit slow eye movements to the left, toward the neuritis, with a corresponding fast phase to the other side (right-beating nystagmus).

Nystagmus with a slow phase toward the side of the lesion is also seen with unilateral lesions in the cranial nerve VIII entry zone or the vestibular nucleus. As with plaques of demyelination from multiple sclerosis (MS), a lesion in the last few millime-

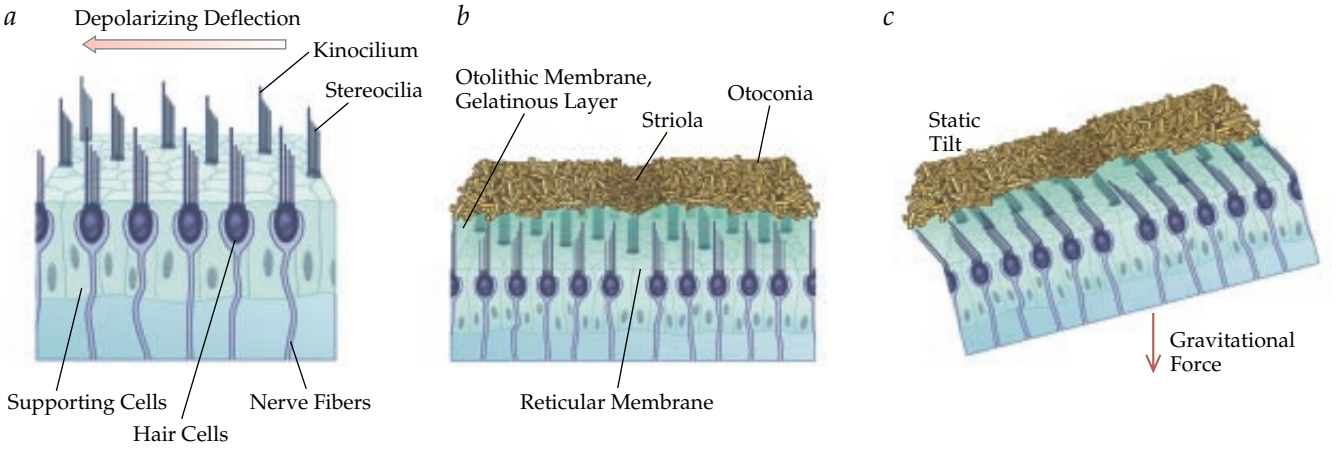


Figure 3 **(*a*) The otolith organs (the utricle and saccule) are specially adapted to detect linear accelerations, including those in the gravitational plane. (*b*) Otolith hair cell stereocilia and kinocilia project into a gelatinous membrane that is covered by crystalline particles of calcium carbonate called otoconia. (*c*) Shearing forces imposed across the membrane occur with static tilt and result in characteristic changes in the movement of stereocilia that can produce either depolarization or hyperpolarization, depending on the direction of movement.**

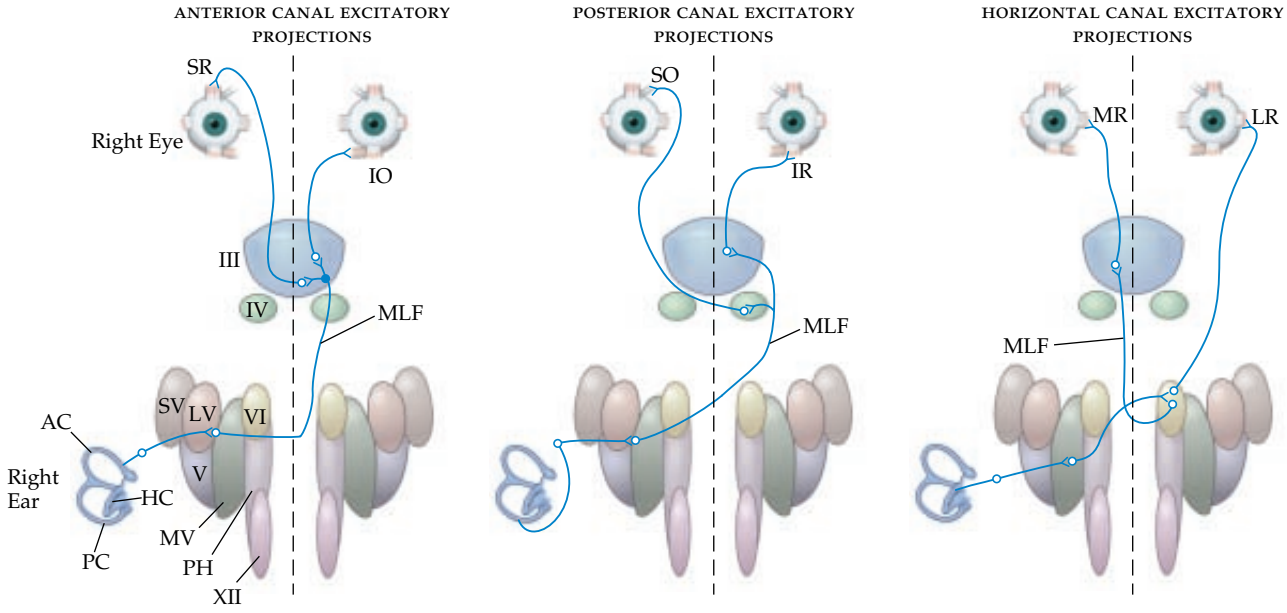


Figure 4 **Shown are the excitatory neural connections that give rise to the three-dimensional vestibular ocular reflex. (*left*) Primary afferents innervating the anterior semicircular canal (AC) synapse onto second order neurons in the vestibular nuclei. Projections cross and ascend in the medial longitudinal fasciculus (MLF) to the contralateral third cranial nerve (oculomotor) nucleus, specifically the inferior oblique (IO) and superior rectus (SR) subnuclei. These motor neurons innervate the inferior oblique muscle on the side opposite the stimulated AC and the ipsilateral superior rectus (via a second decussation) muscle to move the eyes in a vertical upward and torsional direction (upper poles away from the stimulated canal). (*center*) Posterior canal (PC) afferents project centrally to synapse in the vestibular nucleus before decussating and traveling in the MLF to innervate the contralateral third and fourth (trochlear) cranial nerve nuclei. Motor neurons ultimately contact the inferior rectus muscle (IR) on the side opposite the stimulated PC and the ipsilateral superior oblique muscle (via a second decussation). (*right*) Afferent input from the horizontal canal (HC) system projects to the medial vestibular (MV) nucleus and then crosses to innervate two populations of cells in the contralateral abducens (VI) nucleus: motor neurons project to the lateral rectus (LR) muscle, and interneurons cross back over and travel in the MLF to innervate the medial rectus (MR) subnucleus of cranial nerve III, which innervates the ipsilateral MR muscle. (LV—lateral vestibular nucleus; PH—prepositus nucleus hypoglossi; SO—superior oblique muscle; SV—superior vestibular nucleus; III—oculomotor nuclear complex; IV—trochlear nucleus; V—inferior vestibular nucleus; VI—abducens nucleus; XII—hypoglossal nucleus)**

ters of cranial nerve VIII before it enters the brain stem at the pontomedullary junction can be indistinguishable from a process affecting the peripheral labyrinth. However, oscillopsia associated with a peripheral vestibular lesion often resolves quickly as compensatory mechanisms come into play; in contrast, compensation for oscillopsia caused by a brain stem disorder generally does not occur rapidly or to the same extent.

Other central abnormalities are less localizable by the direction of nystagmus. With vestibular cerebellar lesions, the slow phase is typically to the side opposite the lesion—likely because of removal of cerebellar inhibition of the ipsilateral vestibular apparatus. Pure upbeat, pure downbeat, or pure torsional nystagmus almost always has a central origin. A mixed horizontal-torsional or vertical-torsional nystagmus, however, usually indicates a peripheral lesion involving the entire vestibular nerve or all three SCCs in one labyrinth [*see* Nystagmus, *below*].

Gaze-evoked nystagmus commonly occurs as a side effect of certain medications—especially anticonvulsants, hypnotics, and tranquilizers—and with disease of the vestibulocerebellum or its brain stem connections in the medial vestibular nucleus and the nucleus prepositus hypoglossi (both are components of the neural integrator, along with the cerebellar flocculus).[8] It is not present in straight-ahead gaze but is elicited when patients attempt to hold the eyes eccentric in the orbit. The eyes will drift back toward the center, and then a quick phase will follow to move the eyes back into the desired position. When gaze-evoked nystagmus results from toxic or metabolic conditions (e.g., Wernicke encephalopathy, alcohol intoxication), it is direction changing—left beating in left gaze, right beating in right gaze. This distinguishes it from nystagmus caused by labyrinthine weakness, which is always direction fixed.

DYNAMIC IMBALANCE

Dynamic vestibular disturbances are provoked by head motion or change in position. They may reflect unilateral or bilateral abnormalities in amplitude, direction, or timing of the VOR and VSR; or they may reflect mechanical disruptions in the labyrinth (e.g., benign paroxysmal positioning vertigo [BPPV]). Unilateral lesions sometimes cause dynamic disturbances through loss of the normal push-pull relation, whereby activity from one labyrinth increases as that from the other decreases, which results in compensatory eye movement in response to head motion. Bilateral vestibular lesions (e.g., from aminoglycoside toxicity) lead to dynamic disturbances because of an overall loss of function but rarely give rise to vegetative symptoms. Dynamic vestibulospinal function can be clinically assessed by observing postural stability while the patient makes rapid turns during ambulation.

In the case of a dynamic imbalance or bilateral vestibular weakness, angular VOR can be tested clinically by observing the effects of head rotation on visual acuity and on eye movements themselves [*see Table 3*]. For example, dynamic visual acuity is

Table 1 Nystagmus and Its Likely Causes

Type of Nystagmus	*Likely Cause*	*Comments*
Purely torsional	Brain stem disease at vestibular nuclei	Nystagmus from unilateral peripheral vestibular loss can appear purely torsional because of suppression of the horizontal component
Purely vertical Downbeat in primary position Upbeat in primary position	Diverse intrinsic brain stem abnormalities; craniocervical junction (Arnold-Chiari malformation, cerebellar lesion), pontomedullary junction, or pontomesencephalic junction lesion	Often increases in intensity in lateral gaze
Horizontal-torsional	Labyrinthine dysfunction; lesions within cranial nerve VIII or the vestibular nucleus	Horizontal component may be suppressed with vision, causing a peripheral nystagmus to appear purely torsional if viewed with fixation present
Purely horizontal Periodic alternating (jerk, changes direction every 2 min)	Lesion in the cerebellar nodulus	Majority of cases can be treated successfully with baclofen
Primarily horizontal Even during attempted up or down gaze, accented by attempted fixation, diminished by convergence or active eyelid closure, associated with a head turn, sometimes accompanied by reverse (or perverted) smooth pursuit	Congenital	Many patients with congenital nystagmus exhibit normal visual acuity; in one form, latent (occlusion) nystagmus, the slow-phase direction depends on which eye is viewing
Pendular (sinusoidal oscillation rather than unidirectional drift)	Congenital	Nystagmus often appears superficially pendular, but pendular nystagmus may be a sequel to brain stem stroke or a manifestation of various disorders, including multiple sclerosis, Whipple disease, toluene intoxication, acquired visual loss, and Pelzius-Mertzbacher disease
Convergent retraction	Midbrain lesion	Usually coexists with upgaze paralysis (Parinaud syndrome)
Seesaw (one eye ascends and intorts; the other descends and extorts)	Midbrain lesion	—
Dissociated, or disconjugate (greatest or only present in abducting eye)	Internuclear ophthalmoplegia	—

evaluated using a near vision card at 14 in. or a distance acuity chart; the patient's head is passively rotated—horizontally (as in shaking the head, "no"), then vertically ("yes"). The amplitude of the movement is not critical, but the head should pass through the straight-ahead position twice each second. Patients with impaired labyrinthine function usually lose more than two lines of acuity secondary to a reduced VOR gain (eye velocity/head velocity).[9,10]

Corrective saccades during head rotations are also a sign of an abnormal VOR. To assess the VOR clinically, the patient is requested to maintain gaze on the examiner's nose. For yaw (horizontal) and pitch (vertical) rotations, the examiner oscillates the patient's head at a rate of about once every 2 seconds, turning the head across almost the entire ocular motor range [*see Table 1*]. The patient is instructed to continue looking carefully at the examiner's nose and to avoid blinking. During this maneuver, patients can maintain gaze on the target using both the VOR and pursuit eye movements. As the speed of head movements increases, the appearance of any saccadic, rather than slow, eye movements indicates a deficit in the VOR, usually from a peripheral vestibular paresis on the side toward which the head is turning when the rapid eye movements are required. Brief high-acceleration head thrusts are then performed,[11] starting with the head turned about 15° from center and ending with the head facing directly forward.[12] If a corrective (catch-up) saccade is required to move the eyes back onto the target after the rapid head turn, then a canal paresis is present on the side toward which the head had just been turned.

Head-shaking nystagmus is elicited by turning the patient's head vigorously side to side, with the chin pitched slightly downward and the eyes closed, about 20 to 30 times.[13] Any nystagmus present after head rotation stops and eyes are opened is noted; this finding is best assessed using Frenzel glasses so that vision is occluded. Normal individuals show at most a beat or two, whereas a peripheral lesion causing a unilateral loss of labyrinthine function usually produces vigorous nystagmus after head shaking; the slow phases are initially directed toward the side of the lesion, and the fast phases are directed away.[14] Cerebellar dysfunction and other central disturbances may also lead to nystagmus after horizontal head shaking, but generally, the nystagmus is most often vertical (so-called cross-coupled nystagmus).[15] Nystagmus after head shaking can arise from mechanical disturbances in the labyrinth as well (e.g., perilymph fistula, Meniere disease, or abnormality of the cupula).

PURSUIT EYE MOVEMENTS AND VOR CANCELLATION

When the head and eyes track a moving target (pursuit) or when a target is moving at the same velocity as the head (as in

Table 2 Characteristics of Nystagmus Elicited by Eye and Head Position

Lesion Site	*Nystagmus Type (Examination Technique)*		
	Spontaneous (Patient upright; eyes straight ahead)	*Gaze-Evoked (Patient upright; eyes directed right, left, up, then down)*	*Positional (Patient supine; head turned first left, then right; Dix-Hallpike: head rotated 45° with patient sitting upright, then supine)*
Central	Pure vertical or pure torsion Not decreased with fixation Pendular or mixed waveform Alternating direction	Direction changing—quick phases in direction of gaze Unilateral and not present in straight-ahead gaze Downbeating often accentuated in lateral gaze	Nystagmus > vertigo Pure vertical Sustained
Peripheral	Mixed horizontal/torsional Direction fixed Jerk waveform Suppressed by fixation	Direction fixed; must be present in straight-ahead gaze (may need to remove fixation) Can be oblique but not purely vertical	Nystagmus = vertigo Must be in plane of stimulated canal (e.g., upbeat and right torsional in right Dix-Hallpike)

reading a book in a moving vehicle), the VOR is normally suppressed to maintain a stable image: this is VOR cancellation or suppression (VORS). The integrity of VORS can be assessed by asking patients to fix their vision on their outstretched thumbs while horizontally rotating the head and arms together. In normal individuals, the eyes do not move during this maneuver.[16] In patients with cerebellar disease,[17] however, and occasionally in patients with brain stem disturbances that affect pursuit eye movements, one can observe a jerky movement of the eyes, suggesting that VORS is incomplete. A number of drugs that affect the function of the vestibulocerebellum can produce this abnormality, including antiepileptic agents, sedative-hypnotics, and tricyclic antidepressants.

NYSTAGMUS

Nystagmus is a very helpful physical finding in both static and dynamic imbalance. Formal analysis of the slow-phase waveform, as with electronystagmography, can often help localize the causative lesion. Jerk nystagmus, or constant-velocity drifts of the eyes with corrective quick phases, is usually caused by an imbalance of vestibular, optokinetic, or pursuit signals. Lesions of the peripheral vestibular apparatus usually cause a mixed horizontal-torsional nystagmus, with slow phases directed toward the side of the lesion and central pursuit mechanisms generally preserved. Vestibular nystagmus is often suppressed during visual fixation and can easily be assessed by performing occlusive ophthalmoscopy: one eye is covered and the smallest-aperture light is directed at the uncovered eye's optic nerve head (the so-called blind spot), so that no visual stimulus is available to suppress any vestibular nystagmus. Drifts of the optic disk and retinal vessels may appear or increase in velocity if an underlying vestibular imbalance exists. The clinician should note that because the back of the eye is being viewed, an upward drift of the fundus corresponds to a downward drift of the pupil, indicating an upbeating nystagmus.

Nystagmus can also be induced by a change in head position [*see Table 2*]. Positional nystagmus frequently results from degenerative changes in the labyrinth—most commonly, otoconia-derived debris that stimulates the posterior SCC (occasionally affecting the horizontal or anterior SCC) by shifting in response to gravity.[18] In the typical clinical picture of probable benign labyrinthine disease, nystagmus is provoked by the Dix-Hallpike maneuver[19] (also known as the Nylen-Barany maneuver) and is characterized by a latency of up to 30 seconds, with predominant vertical and torsional components parallel to the affected SCC, and a duration of 10 to 60 seconds.[20] This nystagmus may transiently reappear with the slow-phase direction reversed when the patient sits up. Repeat testing with the maneuver may induce further episodes of nystagmus, but fatigability is generally observed.

Peripheral positional nystagmus (i.e., from labyrinthine disease) can be distinguished from positional nystagmus originating in the central nervous system. For positional nystagmus to be considered peripheral, it must be consistent with whichever canal is stimulated by the head position in which it is elicited.[21] Most often, central nystagmus is purely horizontal, vertical, or torsional; is conjugate; is without latency; is nonfatigable; and lasts longer than 30 seconds. Its particular characteristics usually reflect a specific lesion location or disease process [*see Table 1*].

Table 3 Clinical Examination Techniques for Dizzy Patients

Technique	*Associated Disorder or Finding*
Dix-Hallpike, Epley, and Semont maneuvers	BPPV, other benign labyrinthine disease
Ductions and versions	Extraocular muscle palsy, strabismus
Dynamic visual acuity	Bilateral labyrinthine loss
Head shaking	Vestibular asymmetry
Head thrusts (VOR gain)	Unilateral or bilateral labyrinthine loss
Hyperventilation	Compressive cerebellopontine angle lesions
Occlusive ophthalmoscopy	Spontaneous nystagmus
Ocular alignment	Skew deviation
Past-pointing	Vestibulospinal tone imbalance
Pursuit eye movements	Cerebellar dysfunction
Rapid turns during ambulation	VSR dysfunction
Saccades	Brain stem or cerebellar dysfunction
Tragal compression	Perilymphatic fistula
Valsalva maneuvers	SCD syndrome
VOR cancellation	Cerebellar function

BPPV—benign paroxysmal positioning vertigo SCD—superior canal dehiscence
VOR—vestibulo-ocular reflex VSR—vestibulospinal reflex

Diagnosis

When patients complain of dizziness, clinicians should ask themselves a series of key questions to ascertain the underlying mechanism, and they must address whether the patient's abnormality has a vestibular or a general medical basis [*see Table 4*]. Psychiatric factors should also be considered.

With dizziness that appears to be vestibular in origin, the clinician must consider whether the abnormality lies peripherally—at the level of the SCCs, the otoliths, the entire labyrinth, or cranial nerve VIII—or lies within the CNS. Central vestibular dysfunction is most commonly associated with lesions in the brain stem or cerebellum.

HISTORY

In taking the history, the clinician must elicit a clear description of the patient's condition, including the duration of vertigo and any factors that worsen or ease the vertigo. The words that patients use to describe their symptoms may be significant, because a number of seemingly synonymous terms often relate to distinct physiologic mechanisms [*see Table 5 and* Clinical Manifestations, *below*].

Obtaining an accurate account may require persistent, close questioning. For example, a patient may first indicate that the vertigo lasted all day, but further inquiry may reveal that in fact the patient had recurrent episodes throughout the day and that each episode lasted for only 10 to 15 seconds and was brought on by changes in head or body position. The duration of vertiginous episodes often points toward a diagnosis. Episodes lasting only seconds, especially in association with movement of the head, suggest BPPV. Patients who experience vertigo that lasts for minutes may be suffering from vestibular migraine or perhaps vertebrobasilar insufficiency. When vertigo lasts from hours to a day, Meniere disease should be considered. Vertigo lasting for days to weeks raises concern about vestibular neuritis (acute peripheral vestibulopathy) or labyrinthine infarction.

CLINICAL MANIFESTATIONS

Vertigo

Vertigo is generally produced when injury occurs to one vestibular apparatus while the other remains intact or when there is asymmetrical involvement. There is a perception of motion, whether of turning, tilting, or rocking, which persists with eyes closed. A classic example is vestibular neuritis, in which inflammation or infection of cranial nerve VIII on the right or left leads to diminished activity on that side relative to the intact side. This imbalance leads to vertigo and nystagmus.

Patients who complain of vertigo should be asked about the direction of perceived movement of the visual world with eyes open and the sensation of rotation with eyes closed. In the case of left vestibular neuritis, nystagmus slow phases will move toward the left, producing the perception of environmental movement to the right when the eyes are open. A head movement to the right would be necessary to produce a leftward slow phase or vestibular ocular reflex; therefore, with left vestibular neuritis, the patient's perception of self-rotation with the eyes closed will be to the right. Vertiginous patients who fall do so generally in the direction opposite their perceived direction of movement with their eyes closed; therefore, a patient with a lesion on the left side would most often fall to the left. This direction represents a vestibulospinal correction. Unfortunately, many patients are unable to describe the direction of either environmental movement or self-movement during an episode of vertigo because of disturbing vegetative symptoms, disorientation, and altered concentration and attention.

Table 4 Key Questions in the Evaluation of the Dizzy Patient

Is the complaint consistent with an abnormality of the vestibular system?
What level of the vestibular system is involved?
What is the duration of the vestibular disorder?
What factors exacerbate or mitigate the patient's symptoms?
Is there a strong positional component in precipitating symptoms?
Is there associated tinnitus, hearing loss, fullness, or pressure in the ears?
Are there other neurologic accompaniments?
Are there cerebrovascular risk factors?
Does Valsalva maneuver, coughing, sneezing, laughing, or exposure to loud noise have any effect?
Is there a history of seizures?
Is there a history of migraine, and is headache associated with symptoms?
Is there a family history of similar events?

When vestibular insults are bilateral and of similar severity, patients will generally not experience vertigo. They will, however, manifest other features of vestibulopathy, such as gait imbalance and an abnormal VOR.

Light-headedness

Some patients with complaints of dizziness describe their symptoms as light-headedness. This choice of words raises concern about presyncope. The causes of presyncope (and nonvertiginous dizziness) are manifold and include orthostatic hypotension, cardiac arrhythmia, certain metabolic derangements (e.g., hypoglycemia), use of drugs with anticholinergic activity, and autonomic dysfunction.

Disequilibrium

Disequilibrium most commonly refers to the inability to maintain a normal gait and upright posture—often arising from abnormalities in the proprioceptive pathways, as occurs in peripheral neuropathy, such as from diabetes mellitus or Guillain-Barré syndrome—and disease in the dorsal column medial lemniscal system in the upper spinal cord and brain stem, such as from cervical myelopathy or vitamin B_{12} deficiency. Disruption

Table 5 Terms That Patients Use to Describe Dizziness

Bouncing	Rolling
Falling	Spinning
Floating	Swaying
Imbalance	Swimming
Light-headedness	Tilting
Listing	Twisting
Passing out	Unsteadiness
Rocking	Vertigo

Table 6 Etiology of Vertigo

Barotrauma	Phobic postural vertigo
Benign paroxysmal positioning vertigo	Psychogenic
Drugs*	Seizure
Familial syndrome	Transient ischemic attacks and stroke
Infection/inflammation†	Tumor (e.g., acoustic neuroma)
Meniere disease	Vascular compression
Multiple sclerosis	Vestibular migraine
Otosclerosis	Vestibular neuritis
Perilymphatic fistula	

*See Table 11.
†See Table 7.

in the descending vestibulospinal tract, as well as cerebellar and other motor abnormalities (e.g., extrapyramidal parkinsonian syndromes or normal-pressure hydrocephalus), may also contribute to the development of disequilibrium.

Vegetative Symptoms

Many patients with vertigo experience vegetative symptoms such as diaphoresis, nausea, and vomiting. These symptoms are often related to instability in autonomic centers localized in the floor of the fourth ventricle, in close proximity to the central vestibular apparatus. A chemosensitive trigger zone in this region appears to be involved with a number of autonomic responses. Increased vagal tone may predominate, with pallor and hypotension; however, sympathetic responses (e.g., hyperventilation or fear of falling) may also contribute to autonomic imbalance.

Spatial Disorientation

When exposed to sensory-rich environments, patients with vertigo often experience perceptual abnormalities such as spatial disorientation, confusion, and discomfort. These abnormalities are likely to be related to faulty sensory vestibular processing in the cortical vestibular areas. During vertigo, the patient's vestibular sensation represents a conflict between labyrinthine sense and self-referred visual senses.

DIFFERENTIAL DIAGNOSIS

A number of conditions can produce vertigo with characteristic features that often allow the clinician to confirm an etiologic diagnosis [*see Table 6*].

Vestibular Disorders

BENIGN PAROXYSMAL POSITIONING VERTIGO

The most common cause of vertigo in patients presenting for medical attention is BPPV.[22] It is characterized by the vertigo and nystagmus that are associated with changes in head position. BPPV can occur after head trauma or prolonged periods of bed rest, or it can be precipitated by assuming unusual positions such as head extension in a dentist's or hairdresser's chair. Frequently, BPPV is idiopathic, especially in the elderly.

Pathophysiology

Dislocation of otoconia from the utricular macula likely underlies BPPV.[18] These otoconia can migrate into the SCC (most often the posterior SCC), leading to canalolithiasis or possibly cupulolithiasis. A change in head position that promotes movement of the otoconia away from the cupula establishes a gravity-sensitive current, whereby stereocilia in the cupula move toward the kinocilium, resulting in the activation of that SCC [*see Figure 2*]. In the posterior canal, such movement produces depolarization and an irritative response in the vestibular nerve. For example, quickly moving a patient with a right posterior canal BPPV onto his or her right side (during either the Dix-Hallpike or the Semont maneuver) will produce activation of that canal.

The posterior SCC projects centrally to cranial nerve nuclei (IV and III) that mediate downward eye movements by innervating the ipsilateral superior oblique and contralateral inferior rectus muscles. This condition produces intorsion and depression of the ipsilateral (lower) eye and produces extorsion and depression of the contralateral (upper) eye.

Diagnosis of BPPV

Patients with BPPV describe vertigo provoked by head movement—most commonly, when turning over in bed, getting out of bed (matutinal vertigo), and reaching upward (e.g., toward a high shelf) with extension of the neck (so-called top-shelf vertigo). The vertigo is characteristically short-lived, generally not lasting longer than 15 seconds to 1 minute, although the episodes can be recurrent with repeated positional changes.[23] For some patients, this form of vertigo is self-limited, whereas in others it is unrelentingly recurrent. Patients will often sleep with the affected ear up to avoid provoking the vertigo.

Positioning nystagmus can be provoked by performance of the Dix-Hallpike maneuver [*see Table 2*].[24] This maneuver is performed with the patient seated on the examining table: the examiner turns the patient's head 45° either to the right or the left, then rapidly moves the patient into the supine position with the head hanging 30° over the edge of the table.[10] In patients with BPPV, nystagmus and vertigo then develop with a latency of 3 to, in rare instances, 30 seconds. Once the nystagmus subsides, the patient is returned to the sitting position; this may produce a recurrence of the vertigo. The maneuver is then repeated with the patient's head turned to the other side. With repeated maneuvers, the vertigo tends to fatigue.

In patients with BPPV involving the posterior SCC, the Dix-Hallpike maneuver will indicate which labyrinth is affected. For example, when a patient with right posterior canal BPPV is moved into the supine position after the head is turned 45° to the right [*see Figure 5*], the slow phase of nystagmus will have a torsional component such that the upper pole of the eye moves toward the patient's left and the vertical component moves downward (relative to the patient's body). Quick phases of nystagmus will be opposite, with an upbeating vertical component and torsional quick phases beating toward the dependent right ear. (If the patient has not eliminated fixation, the vertical component of the nystagmus may be suppressed and only the torsional component observed.)

Treatment

The treatment of BPPV involves the performance of maneuvers that aim to promote the repositioning of the otoconia from the SCC system (particle repositioning).[25] One such maneuver, the Epley maneuver, begins in a fashion that is identical to the Dix-Hallpike maneuver except that the patient's head is eventually moved through a complete turn rather than limited to only a left or right downward position [*see Figure 5*]. The movement of

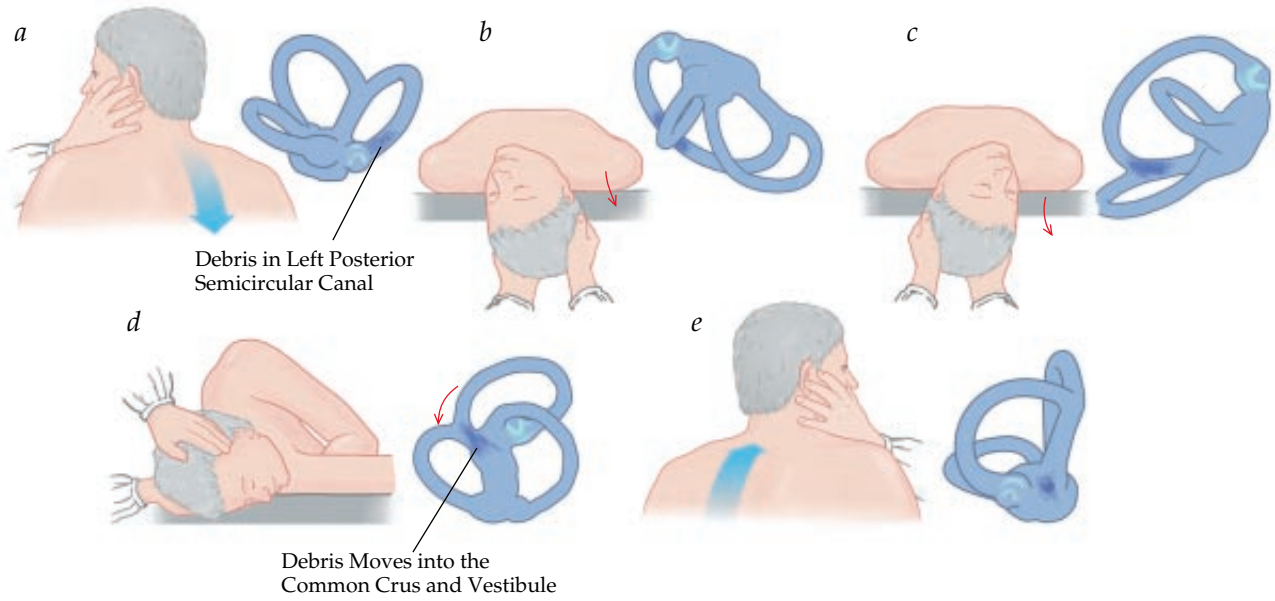


Figure 5 **With the Dix-Hallpike (*a and b*) and particle repositioning (*c through f*) maneuvers, the debris is sequentially moved further into the canal and ultimately to the vestibule, where it is effectively repositioned.**

the head gradually directs the otoconia from the long arm of the canal into the vestibule, where it no longer causes symptoms. During the modified Epley maneuver, the initial head-hanging position with the involved ear down is maintained until the observed nystagmus has fatigued. The head is then turned, with the neck extended, until the noninvolved ear is facing downward. Nystagmus may recur at this point and should be in the same direction as it was when initially elicited. This second position is held until the nystagmus fatigues or for approximately 30 to 60 seconds. The patient is then rolled onto his or her side into a lateral decubitus body position, with the nose pointed downward, for another 30 to 60 seconds. Finally, the chin is tucked toward the chest and the patient is helped up into the seated position. The examiner must maintain contact with the patient at this point, because sitting up will sometimes elicit a vestibulospinal response, which may cause the patient to fall from the table. Two aspects of the original particle repositioning procedure have been found to provide no additional benefit: instructing the patient to sleep upright for 48 hours after treatment[26] and the use of an oscillator over the mastoid during the maneuver.[27,28]

An alternative to the Epley maneuver, the Semont liberatory maneuver, involves the same theoretical elements as the Epley maneuver but utilizes a distinctly different procedure. The Semont maneuver is performed by rapidly moving the patient onto the left or right side and simultaneously rotating the head so that it is pointing toward the ceiling. This position is maintained for 30 to 60 seconds before the patient is rapidly moved into a face-down position on the opposite side of the table.[29]

These maneuvers resolve BPPV in the majority of patients.[30] When this form of therapy fails and BPPV is still the most likely diagnosis, the physician should repeat the Dix-Hallpike maneuver and apply a vibrator to the mastoid bone on the side of involvement while completing a standard particle-repositioning maneuver. The vast majority of patients with BPPV can be effectively treated in this fashion. Very rarely, BPPV is refractory to all conventional treatment interventions. In such cases, alternatives such as surgical neurectomy (singular neurectomy)[31] or SCC occlusion[32] may be necessary.

The use of a vestibular suppressant (e.g., diazepam, clonazepam, or meclizine) is inappropriate for patients with probable BPPV. In general, these agents do not have a role in the treatment of BPPV.

VESTIBULAR NEURITIS

After BPPV, the second most common cause of vertigo is vestibular neuritis, also known as acute peripheral vestibulopathy.[33] The condition was previously called vestibular neuronitis, but it is now believed that inflammation occurs in the nerve itself and not generally in the vestibular ganglion.

Diagnosis

Patients with vestibular neuritis experience severe rotational vertigo, often associated with nausea and vomiting. Typically, the examiner will observe a horizontal-torsional nystagmus beating, with the fast phase toward the normal ear and the slow phase toward the abnormal ear. Hearing is typically preserved, differentiating this condition from labyrinthitis, in which hearing is typically affected. Symptoms often abate in 48 to 72 hours, although complete recovery may take as long as 6 weeks. Vestibular neuritis is often preceded by an upper respiratory tract infection. A number of infectious agents have been associated with vestibular neuritis [*see Table 7*].

Treatment

Corticosteroids are effective for the treatment of acute vestibular neuritis [*see Table 8*].[34] No benefit with antiviral treatment has been found. Symptomatic therapy with antiemetics and vestibular suppressants such as meclizine, diazepam, or promethazine can be used initially, but such agents may prolong the recovery process by preventing CNS adaptation and can also produce excessive sedation that predisposes patients (especially the elderly) to a greater risk of disequilibrium and injury. Hence, vestibular-suppressive agents are recommended only during the initial acute period of vestibulopathy (24 to 72 hours). They can be used in combination with stimulants such as caffeine and methylphenidate (5 to 10 mg every 6 to 8

Table 7 Infections and Inflammatory Conditions Associated with Vestibular Neuritis

Adenovirus	Lyme disease
Cytomegalovirus	Otitis media/interna
Enterovirus	Sarcoidosis
Epstein-Barr virus	Syphilis
Hepatitis	Tuberculosis
Herpes simplex virus	Varicella-zoster virus (Ramsay Hunt syndrome)
HIV	
Influenza	

hours), thereby minimizing daytime sleepiness. In patients with dehydration secondary to vomiting and diaphoresis, rehydration is obviously also essential.

Vestibular neuritis is generally self-limited, but occasionally, a failure of the compensation process that is required to adapt to a residual loss of function will cause patients to have continued symptoms. If, after an initial monophasic vestibular event, patients continue to complain of dizziness provoked by head movement, then rehabilitation with daily home exercises is indicated. Such rehabilitation, which can be supervised by physical or occupational therapists or even primary care nurses or physician extenders, results in significant improvement in the majority of patients.[35] BPPV also frequently occurs weeks or months after vestibular neuritis.

MENIERE DISEASE

Pathophysiology

Meniere disease is thought to involve the accumulation of excessive endolymphatic fluid within the labyrinth. The term endolymphatic hydrops[36] is sometimes applied to Meniere disease, on the basis of the pathologic appearance of the temporal bone in postmortem examinations, although this finding is often present in temporal bones from asymptomatic persons.[36] Idiopathic periodic microruptures in the labyrinthine sac allow mixing of potassium-rich endolymph with perilymph. Cranial nerve VIII, located in the perilymph, becomes hyperpolarized when exposed to potassium-rich endolymph. The consequence of unilateral cranial nerve VIII hyperpolarization is a vestibular functional asymmetry culminating in nystagmus, vertigo, postural instability, and vegetative symptoms.

Table 8 Treatment Schedule for Acute Vestibular Neuritis

Treatment Day	*Oral Methylprednisolone, Single Daily Dose (mg)*
1–3	100
4–6	80
7–9	60
10–12	40
13–15	20
16–18	10
19	0
20	10
21	0
22	10

Diagnosis

Meniere disease involves attacks of vertigo that are typically associated with hearing loss, tinnitus, and a sensation of pressure or fullness in the affected ear. Both clinical and laboratory criteria are used for diagnosis [*see Table 9*]. Attacks are paroxysmal and generally last between 2 and 24 hours. Often, a history of high sodium intake hours before an attack can be elicited. Patients with Meniere disease may experience sudden falls, and they usually describe the perception of being thrown to the floor (otolithic crisis of Tumarkin). This probably results from abnormal excitation of the otolith organs leading to inappropriate vestibulospinal responses.[37]

The nystagmus observed in patients with Meniere disease is similar to that observed in patients with other forms of peripheral vestibulopathy, such as vestibular neuritis. As such, the nystagmus is mainly horizontal with a torsional component and will decrease in intensity with visual fixation.

Patients with Meniere disease typically lose hearing over time. Low-frequency hearing is most affected early in the disease course, whereas high-frequency hearing loss is a late manifestation. These losses, together with relative preservation of middle-frequency hearing, produce a characteristic audiogram pattern referred to as a peak audiogram [*see Figure 6*].[38] Results of transtympanic electrocochleography are also abnormal in most patients with Meniere disease, suggesting an inner-ear fluid imbalance.[39] This test is rarely used in clinical practice, however.

The differential diagnosis for patients presenting with a Meniere-type disease is varied [*see Table 10*]. A number of familial autosomal dominant conditions are associated with periodic vertigo. For example, in its chronic stages, otosclerosis leads to conductive hearing loss and can involve the labyrinth, producing sensorineural hearing loss and vestibulopathy.[40]

Treatment

Conservative treatment of Meniere disease consists of dietary-salt restriction and diuretic agents. Betahistine, a histamine analogue, is prescribed for Meniere disease in many countries, but it has not been approved for use in the United States and the balance of evidence does not support its use.[41]

For acute attacks of vertigo, vestibular suppressants are often helpful in combination with antiemetics. In cases that do not respond to medical treatment, the standard therapy is becoming intratympanic gentamicin, which results in a partial chemical labyrinthectomy, with a small risk of hearing loss.[33] Surgical resection of the vestibular nerve is also effective, albeit more invasive.[42] Endolymphatic shunting is sometimes performed but has not clearly been shown to be beneficial. Pressure devices are currently undergoing study and may become a minimally invasive treatment option.[43]

PHOBIC POSTURAL VERTIGO

Diagnosis

With phobic postural vertigo, patients report a subjective balance disturbance despite normal results on balance testing.[44] Common complaints include postural vertigo and fluctuating

Table 9 Diagnostic Criteria for Meniere Disease[82]

Possible	Episodic vertigo of the Meniere type without documented hearing loss *or* Sensorineural hearing loss, fluctuating or fixed, with disequilibrium but without definitive episodes
Probable	One definitive episode of vertigo Audiometrically documented hearing loss on at least one occasion Tinnitus or aural fullness in the treated ear
Definite	Two or more definitive spontaneous episodes of vertigo lasting 20 min or longer Audiometrically documented hearing loss on at least one occasion Tinnitus or aural fullness in the treated ear
Certain	Definite Meniere disease, plus histopathologic confirmation

Note: all classifications require that other causes are excluded [*see Table 10*].

unsteadiness, often associated with anxiety and vegetative symptoms. Attacks of vertigo are typically associated with particular perceptual stimuli, such as exposure to malls, bridges, staircases, or specific social situations.

Phobic postural vertigo is most commonly observed after the onset of a stressful event or recent illness, especially a vestibular disorder. Many patients who experience vestibulopathy initially use adaptive behavioral mechanisms (e.g., walking with assistive devices and utilizing furniture and walls for stability) to avoid falling. Unfortunately, many patients persist in utilizing these mechanisms long after the vestibulopathy or stressful event has resolved. Ultimately, most of these patients exhibit elements of anxiety and phobia that are most prominent when ambulating. This condition should not be confused with malingering; rather, it is a defensive reaction, often with associated autonomic features (e.g., hyperventilation[45] and tachycardia) that exacerbate the dizziness and perpetuate the maladaptive behavior.[46]

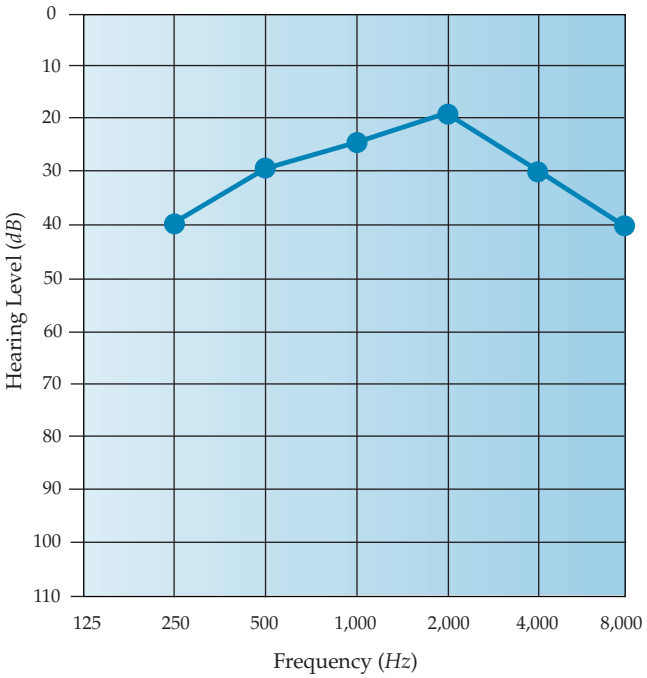


Figure 6 **The audiogram demonstrates the characteristic pattern of hearing loss observed in patients with inner-ear fluid imbalances such as Meniere disease. The reduction in low-frequency and high-frequency hearing and the relative preservation of middle-frequency hearing is referred to as a peak pattern.**

Treatment

Patients with phobic postural vertigo are highly responsive to vestibular rehabilitation, counseling, and exercise. Initially, it is explained to patients that their vertigo is related to disturbed sensory integration. Further, it is emphasized that behavioral strategies that were appropriate during a period of vestibulopathy or other stressful life event may ultimately become maladaptive if they persist. An exercise program and desensitization with psychotherapy are prescribed.

Pharmacotherapy can be a useful adjunct. The usual choice is a vestibular suppressant such as clonazepam or diazepam. These agents should be used at the lowest dose possible because they can also retard the process of vestibular compensation. For patients with a concomitant mood or anxiety disorder, selective serotonin reuptake inhibitors are often useful.[47]

VERTEBROBASILAR INSUFFICIENCY

Two thirds of patients with vertebrobasilar insufficiency will complain of dizziness at some point during the disease.[48] Of patients with completed posterior-circulation strokes, 15% to 20% will have had antecedent isolated attacks of dizziness. These attacks generally last on the order of minutes. It is important to recognize that vestibulopathy related to vertebrobasilar insuffi-

Table 10 Differential Diagnosis of Meniere Disease

Type of Disorder	*Diagnosis*
Vascular	Giant cell arteritis Migraine headache Microvascular compression Posterior-circulation stroke Vertebrobasilar transient ischemic attack
Infectious	HIV Syphilis
Inflammatory/ immune mediated	Cogan syndrome Lupus Multiple sclerosis Chronic inflammatory demyelinating polyneuropathy Sarcoidosis Susac syndrome Idiopathic progressive bilateral sensorineural hearing loss
Structural	Neoplasia Perilymphatic fistula
Other	Superior canal dehiscence Episodic ataxia type 2 Seizure Thyroid disease

ciency can be the result of ischemia in the brain stem, cerebellum, or inner-ear labyrinth, secondary to disease in the branches of the basilar artery. With completed strokes, the vertigo can persist for hours or longer.

Pathophysiology

The principal source of blood supply to the labyrinth is the anterior inferior cerebellar artery. This branch of the basilar trunk gives rise to the internal auditory artery, which ultimately divides into the anterior and posterior vestibular arteries. The anterior branch perfuses the anterior and horizontal SCCs, as well as the utricle, and the posterior branch serves the posterior SCC, the cochlea, and the saccule. Hence, an embolic event that specifically involves the anterior vestibular artery can engender a partial labyrinthine infarction, producing vertigo without hearing loss[49] or vertigo with bilateral hearing loss.[50] Transient bilateral hearing loss may herald a catastrophic basilar artery occlusion.[51] Complete labyrinthine infarction can produce profound and lasting hearing loss and vestibular loss. Even though this would be considered a peripheral rather than a central lesion, the risk of further posterior-circulation strokes is the same as that after a pontine or cerebellar infarction.

Diagnosis

Vertebrobasilar ischemia that affects the brain stem and cerebellum can produce a broad array of neurologic manifestations. These include nystagmus, diplopia, tinnitus, sensory changes, ataxia, dysarthric speech, facial pain, Horner syndrome, weakness, and postural abnormalities. The nystagmus can be highly varied and can have characteristics of either a central or a peripheral type.

Because vertebrobasilar insufficiency is potentially life threatening, it should be actively excluded as a cause of dizziness in patients with a compatible clinical picture, especially those with cerebrovascular risk factors such as hypertension and diabetes. Magnetic resonance imaging of the brain is useful to look for evidence of posterior fossa lesions suggestive of stroke. High-resolution magnetic resonance angiography is increasingly utilized to visualize the patency of the basilar and vertebral arteries. With further refinements, this imaging technique may largely replace conventional angiography as the gold standard for assessing the anatomy of the posterior cerebral vasculature. Patients with cerebellar or brain stem hemorrhagic stroke may present similarly but with a more prolonged duration of symptoms. Emergent imaging is indicated in such cases.

Treatment

In patients with established vertebrobasilar insufficiency, a critical aspect of therapy is the reduction of modifiable risk factors, such as smoking, hypertension, and hyperlipidemia. Antiplatelet or anticoagulation therapy must also be carefully considered [*see 11:IV Cerebrovascular Disorders*].

DRUG-INDUCED VERTIGO

Many drugs can produce dizziness by their effects on the vestibular system and cerebellum [*see Table 11*]. Alcohol changes the specific gravity of endolymph in the cupula and can lead to nystagmus and vestibulopathy. It also impairs cerebellar function, which can produce ataxia, speech changes, postural instability, and nystagmus. Lithium therapy can result in cerebellar dysfunction and can be a cause of downbeating nystagmus.

Table 11 Drugs That Can Cause Vertigo

Drug or Category	*Mechanism*
Alcohol	Peripheral: alteration of endolymph specific gravity Central: gaze-evoked nystagmus, cellular atrophy
Anticholinergics, antihypertensives, smooth muscle relaxants (e.g., alpha blockers)	Orthostatic hypotension
Anticonvulsants	Cerebellar dysfunction
Aminoglycoside antibiotics	Vestibular hair cell damage (causes imbalance rather than vertigo)
High-dose aspirin	Ototoxic damage to vestibular hair cells
Chemotherapeutic agents (e.g., cisplatin)	Hair cell damage

Other agents commonly associated with dizziness include tranquilizers and antihypertensive medications, which cause dizziness by producing orthostatic hypotension; anticonvulsant agents, which produce cerebellar dysfunction; and antibiotics, which can damage vestibular hair cells bilaterally, leading to gait instability and oscillopsia. Additional sources of vestibular toxicity include high-dose aspirin[52] and chemotherapeutic agents such as cisplatin.

One of the most ototoxic agents is gentamicin, which can produce profound vestibular dysfunction by causing bilateral damage to inner-ear hair cells. The symmetrical damage explains the absence of vertigo and nystagmus. In one study, vestibulopathy was found to begin as long as 1 week after gentamicin was discontinued, and onset was delayed in 32 of 36 patients.[53] Half of the affected patients did not have abnormal peak or trough antibiotic levels. Metronidazole given concurrently with gentamicin may increase the risk of vestibulotoxicity.[54] Dynamic visual-acuity testing can be helpful in early recognition of gentamicin-associated vestibular toxicity. Prompt termination of gentamicin treatment may confer a more complete recovery in such patients. There may be a genetic predisposition to gentamicin ototoxicity related to an abnormality in ribosomal messenger RNA, although most cases of ototoxicity occur in patients who do not have this mutation.[55]

VESTIBULAR MIGRAINE

Migraine headache often has neurologic accompaniments, the most common of which are visual distortions. The second most common neurologic manifestation of migraine is vestibulopathy, principally characterized by vertigo.[56] Central nystagmus is the typical type observed in patients with migraine.[57] The most extreme form of vestibular migraine, basilar artery migraine, was first described and named by Bickerstaff. This migraine occurs in adolescents and young adults; typically presents as visual loss or aura, vertigo, ataxia, dysarthria, and dysesthesia lasting up to 45 minutes; and is always followed by a throbbing headache, frequently with vomiting.[58] Much less dramatic, but more common, is so-called migrainous vertigo, which combines vestibular symptoms with characteristic features of migraine.[59]

In up to 30% of patients with vestibular migraine, vertigo oc-

curs completely independently of headache (i.e., acephalgic migraine). A family history of migraine is common, as is a personal history of car sickness in childhood. Uncommonly, sudden hearing loss is attributable to migraine.[60] On the other hand, patients considered to have Meniere syndrome who have no documented hearing loss may in fact have migraine. Patients who complain of "sinus headaches" that include vestibular symptoms most likely have vestibular migraine. Most patients with vestibular migraine have episodes of dizziness that generally last for minutes, although in some, the episodes last for hours. The vertigo may be position induced and can occur in association with central-pattern nystagmus. Some women have a greater disposition for vestibular migraine around the time of their menses (so-called catamenial migraine).

The treatment of vestibular migraine is similar to that of other forms of migraine [*see 11:VIII Headache*] and includes elimination of habitual use of caffeine and over-the-counter analgesics and prophylactic treatment with beta blockers, tricyclic antidepressants, calcium channel blockers, valproic acid, and, occasionally, acetazolamide. Triptans may play a role in therapy, but their efficacy has not yet been determined.[61,62]

PERILYMPHATIC FISTULA AND SUPERIOR CANAL DEHISCENCE SYNDROME

Perilymphatic fistulas are abnormal communications between perilymph and the outside world.[63] Generally, these fistulas occur between the middle and inner ear at the oval or round windows. Precipitating causes include head trauma, Valsalva maneuvers that increase intracranial pressure, vigorous exercise, stapes surgery, barotrauma, erosive lesions in the petrous bone (e.g., a neoplastic lesion or cholesteatoma), inflammatory conditions such as syphilis, and congenital conditions.

Superior canal dehiscence (SCD) syndrome involves the opening of a third window into the labyrinth because of the absence of bone overlying the superior canal.[64] Although the syndrome probably results from congenital thinness of the temporal bone in this area, symptoms may not appear until adulthood, often after head trauma.

Diagnosis

Patients with perilymphatic fistula show hearing loss that is generally similar to that which occurs in Meniere disease. Position-induced vertigo and nystagmus may also occur. Vertigo and nystagmus can also be induced by tragal compression or with swallowing. A useful clinical test for diagnosing fistula involves insufflation of the external ear, which in some patients can produce nystagmus (Hennebert sign). This has been incorporated into platform posturographic testing.[65]

Vertigo in patients with perilymphatic fistula is often secondary to activation of the SCC. Patients with either perilymphatic or otolithic fistula can experience vertigo that is provoked by exposure to loud noise (the Tullio phenomenon), a response that may also be seen in Meniere disease. Patients with SCD also experience the Tullio phenomenon, sometimes even from their own voice. A Valsalva maneuver performed against pinched nostrils, which results in increased middle-ear pressure, will induce a nystagmus consistent with excitation of the superior canal—namely, a mixed torsional-vertical nystagmus with fast phases directed downward and with the upper pole beating toward the affected ear. Hearing loss is often mixed, with both a conductive and a sensorineural component, predominantly in the lower frequencies. Some patients report that when a tuning fork is placed on the ankle, they can hear it in the affected ear. A useful laboratory test, although not widely available, is measurement of the vestibular evoked myogenic potential (VEMP).[66] This test measures electromyographic activity in the sternocleidomastoid muscle by modulating loud clicks in each ear. The VEMP is elicited at a lower threshold than normal in an ear with SCD.

Treatment

Patients with a fistula are generally treated conservatively, with bed rest, fluid hydration, elevation of the head, and avoidance of straining that may cause a Valsalva response. If conservative measures fail, surgical occlusion of the fistula site can be attempted. This generally involves patching the oval and round windows with fatty tissue. Success is variable.[67]

Surgical resurfacing of the arcuate eminence, the area of bone overlying the superior canal, can effectively reduce or eliminate symptoms of SCD.[63] This involves a craniotomy and middle fossa approach.

STRUCTURAL ABNORMALITIES ASSOCIATED WITH VERTIGO

A variety of structural abnormalities have been associated with the development of vertigo. Cerebellopontine angle tumors (e.g., acoustic neuromas) and Arnold-Chiari malformation may present as new-onset vertigo associated with asymmetry in sensorineural hearing thresholds or with new-onset or atypical headache. Consequently, such findings are accepted indications for brain imaging with contrast. Other structural lesions that can produce vertigo include primary brain tumor, metastatic lesion, neurofibroma, traumatic labyrinthine fracture, dermoid and epidermoid lesions, and vascular malformation.

Some patients who have stereotypical paroxysmal episodes of vertigo may suffer from a vascular compression syndrome.[68] This syndrome is produced by a vascular ectasia in one of the posterior-circulation arteries that leads to compression of the caudal brain stem or cranial nerve VIII. Such patients experience rotational or fore-and-aft vertigo. Attacks last seconds to minutes, are generally provoked by particular head positions, and can be associated with hyperacusis or tinnitus. Results of vestibular or auditory function testing may be abnormal. Hyperventilation or exercise-induced symptoms are typical of, but not specific for, microvascular compression syndromes. Most patients respond to membrane-stabilizing agents such as carbamazepine, phenytoin, clonazepam, and baclofen. Medically refractory cases may be responsive to vascular decompression surgery.[69]

CEREBELLAR ATAXIA

Patients with cerebellar disease from any cause may complain of vertigo and oscillopsia because of spontaneous or positional nystagmus. Subacute development of imbalance, dysarthria, and incoordination over weeks to months should prompt a search for paraneoplastic antibodies and an occult malignancy.[70] A family history should be sought and molecular testing for the various spinocerebellar atrophies considered under the appropriate circumstances. Episodic ataxia type 2 (EA2; familial acetazolamide-responsive episodic or periodic ataxia syndrome) is an autosomal dominant condition caused by mutations in the calcium channel gene (*CACNA1A*), located on chromosome 19[71] [*see 11:X Inherited Ataxias*]. This is the same gene that is involved in spinocerebellar ataxia type 6 and famil-

ial hemiplegic migraine. Treatment with acetazolamide is generally effective in reducing the frequency and severity of attacks. However, patients often have interictal abnormalities, including downbeating nystagmus. Fortunately, 4-aminopyridine (4AP) or the related compound 3,4-diaminopyridine have been reported to decrease downbeating nystagmus.[72] Some patients benefit from treatment with clonazepam, baclofen, or gabapentin. Bilateral vestibular loss and cerebellar ataxia can occur as part of the same syndrome and can be recognized by a failure of visual-pursuit mechanisms to enhance the VOR during head rotation in light.[73]

VESTIBULAR SEIZURE

In rare cases, patients present with stereotypical episodes of vertigo on the basis of seizure activity.[74] These episodes have been referred to as tornado epilepsy. Such attacks appear to involve discharges from primary vestibular cortical zones that produce symptoms of vertigo and, on occasion, the perception of environmental tilt. It is of interest that Penfield,[75] in his original human cortical stimulation experiments, was able to elicit vertigo by stimulation of the parietal insular regions of the brain, though frontal areas have also been implicated.[76] Fortunately, these seizures often respond to treatment with standard anticonvulsant agents [*see 11:XII Epilepsy*]. It should be noted that a normal interictal electroencephalogram does not exclude the presence of an epileptic focus. Nystagmus, however, is not uncommon during a seizure and may have some localizing value.[77]

MULTIPLE SCLEROSIS

Although vertigo is an uncommon initial presentation of MS, accounting for less than 5% of cases, up to 20% of MS patients will experience vertigo at some time over the course of the illness. Vertigo secondary to demyelination most commonly arises from lesions in the vestibular nucleus and the root entry zone of cranial nerve VIII.[78] Nystagmus from such lesions can appear to be of central or peripheral origin (e.g., mixed torsional-horizontal, pendular, or pure vertical).[79] When a patient with MS has vertigo resulting from inflammatory lesions, appropriate treatment includes corticosteroids and vestibular suppressants. BPPV must be excluded in any MS patient with vertigo, given the ease of BPPV diagnosis and the high response rate to treatment.[80,81]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hudspeth AJ: How the ear's works work. Nature 341:397, 1989

2. Sohmer H, Freeman S: Basic and clinical physiology of the inner ear receptors and their neural pathways in the brain. J Basic Clin Physiol Pharmacol 11:367, 2000

3. Balaban CD: Vestibular autonomic regulation (including motion sickness and the mechanism of vomiting). Curr Opin Neurol 12:29, 1999

4. Brandt T, Dieterich M: Vestibular syndromes in the roll plane: topographic diagnosis from brainstem to cortex. Ann Neurol 36:337, 1994

5. Halmagyi GM, Gresty MA, Gibson WP: Ocular tilt reaction with peripheral vestibular lesion. Ann Neurol 6:80, 1979

6. Halmagyi GM, Brandt T, Dieterich M, et al: Tonic contraversive ocular tilt reaction due to unilateral meso-diencephalic lesion. Neurology 40:1503, 1990

7. Maire R, Duvoisin B: Localization of static positional nystagmus with the ocular fixation test. Laryngoscope 109:606, 1999

8. Leigh RJ, Zee DS: The Neurology of Eye Movements, 3rd ed. Oxford University Press, New York, 1999

9. Longridge NS, Mallinson AI: The dynamic illegible E-test: a technique for assessing the vestibulo-ocular reflex. Acta Otolaryngol 103:273, 1987

10. Herdman SJ, Tusa RJ, Blatt P, et al: Computerized dynamic visual acuity test in the assessment of vestibular deficits. Am J Otol 19:790, 1998

11. Halmagyi GM, Curthoys IS: A clinical sign of canal paresis. Arch Neurol 45:737, 1988

12. Oliva M, Martin Garcia MA, Bartual J, et al: [The head-thrust test (HTT): physiopathological considerations and its clinical use in daily practice]. Acta Otorhinolaringol Esp 49:275, 1998

13. Hain TC, Fetter M, Zee DS: Head-shaking nystagmus in patients with unilateral peripheral vestibular lesions. Am J Otolaryngol 8:36, 1987

14. Asawavichianginda S, Fujimoto M, Mai M, et al: Prevalence of head-shaking nystagmus in patients according to their diagnostic classification in a dizziness unit. J Otolaryngol 26:20, 1997

15. Walker MF, Zee DS: Directional abnormalities of vestibular and optokinetic responses in cerebellar disease. Ann NY Acad Sci 871:205, 1999

16. Eggers SD, Zee DS: Evaluating the dizzy patient: bedside examination and laboratory assessment of the vestibular system. Semin Neurol 23:47, 2003

17. Lewis RF, Zee DS: Ocular motor disorders associated with cerebellar lesions: pathophysiology and topical localization. Rev Neurol (Paris) 149:665, 1993

18. Brandt T, Steddin S: Current view of the mechanism of benign paroxysmal positioning vertigo: cupulolithiasis or canalolithiasis? J Vestib Res 3:373, 1993

19. Baloh RW: Charles Skinner Hallpike and the beginnings of neurotology. Neurology 54:2138, 2000

20. Brandt T, Steddin S, Daroff RB: Therapy for benign paroxysmal positioning vertigo, revisited. Neurology 44:796, 1994

21. Buttner U, Helmchen C, Brandt T: Diagnostic criteria for central versus peripheral positioning nystagmus and vertigo: a review. Acta Otolaryngol 119:1, 1999

22. Waterston J: Neurology. 3: Dizziness. Med J Aust 172:506, 2000

23. Solomon D: Benign paroxysmal positional vertigo. Curr Treat Options Neurol 2:417, 2000

24. Furman JM, Cass SP: Benign paroxysmal positional vertigo. N Engl J Med 341:1590, 1999

25. Epley JM: The canalith repositioning procedure: for treatment of benign paroxysmal positional vertigo. Otolaryngol Head Neck Surg 107:399, 1992

26. Massoud EA, Ireland DJ: Post-treatment instructions in the nonsurgical management of benign paroxysmal positional vertigo. J Otolaryngol 25:121, 1996

27. Wolf JS, Boyev KP, Manokey BJ, et al: Success of the modified Epley maneuver in treating benign paroxysmal positional vertigo. Laryngoscope 109:900, 1999

28. Hain TC, Helminski JO, Reis IL, et al: Vibration does not improve results of the canalith repositioning procedure. Arch Otolaryngol Head Neck Surg 126:617, 2000

29. Semont A, Freyss G, Vitte E: Curing the BPPV with a liberatory maneuver. Adv Otorhinolaryngol 42:290, 1988

30. Froehling DA, Bowen JM, Mohr DN, et al: The canalith repositioning procedure for the treatment of benign paroxysmal positional vertigo: a randomized controlled trial. Mayo Clin Proc 75:695, 2000

31. Gacek RR: Technique and results of singular neurectomy for the management of benign paroxysmal positional vertigo. Acta Otolaryngol 115:154, 1995

32. Parnes LS, McClure JA: Posterior semicircular canal occlusion for intractable benign paroxysmal positional vertigo. Ann Otol Rhinol Laryngol 99:330, 1990

33. Hanley K, O'Dowd T, Considine N: A systematic review of vertigo in primary care. Br J Gen Pract 51:666, 2001

34. Strupp M, Zingler VC, Arbusow V, et al: Methylprednisolone, valacyclovir, or the combination for vestibular neuritis. N Engl J Med 351:354, 2004

35. Yardley L, Donovan-Hall M, Smith HE, et al: Effectiveness of primary care-based vestibular rehabilitation for chronic dizziness. Ann Intern Med 141:598, 2004

36. Rauch SD, Merchant SN, Thedinger BA: Meniere's syndrome and endolymphatic hydrops: double-blind temporal bone study. Ann Otol Rhinol Laryngol 98:873, 1989

37. Baloh RW, Jacobson K, Winder T: Drop attacks with Meniere's syndrome. Ann Neurol 28:384, 1990

38. Paparella MM, McDermott JC, de Sousa LC: Meniere's disease and the peak audiogram. Arch Otolaryngol 108:555, 1982

39. Conlon BJ, Gibson WP: Electrocochleography in the diagnosis of Meniere's disease. Acta Otolaryngol 120:480, 2000

40. Virolainen E: Vestibular disturbances in clinical otosclerosis. Acta Otolaryngol Suppl 306:1, 1972

41. James AL, Burton MJ: Betahistine for Menière's disease or syndrome. Cochrane Database Syst Rev (1):CD001873, 2001

42. Pareschi R, Destito D, Falco Raucci A, et al: Posterior fossa vestibular neurotomy as primary surgical treatment of Meniere's disease: a re-evaluation. J Laryngol Otol 116:593, 2002

43. Gates GA, Green JD Jr, Tucci DL, et al: The effects of transtympanic micropressure treatment in people with unilateral Meniere's disease. Arch Otolaryngol Head Neck Surg 130:718, 2004

44. Brandt T: Phobic postural vertigo. Neurology 46:1515, 1996

45. Evans RW: Neurologic aspects of hyperventilation syndrome. Semin Neurol 15:115, 1995

46. Solomon D, Shepard NT: Chronic dizziness. Curr Treat Options Neurol 4:281, 2002

47. Staab JP, Ruckenstein MJ, Solomon D, et al: Serotonin reuptake inhibitors for dizziness with psychiatric symptoms. Arch Otolaryngol Head Neck Surg 128:554, 2002

48. Gomez CR, Cruz-Flores S, Malkoff MD, et al: Isolated vertigo as a manifestation of

vertebrobasilar ischemia. Neurology 47:94, 1996

49. Oas JG, Baloh RW: Vertigo and the anterior inferior cerebellar artery syndrome. Neurology 42:2274, 1992

50. Lee H, Yi HA, Baloh RW: Sudden bilateral simultaneous deafness with vertigo as a sole manifestation of vertebrobasilar insufficiency. J Neurol Neurosurg Psychiatry 74:539, 2003

51. Toyoda K: Bilateral deafness as a prodromal symptom of basilar artery occlusion. J Neurol Sci 193:147, 2002

52. Strupp M, Jahn K, Brandt T: Another adverse effect of aspirin: bilateral vestibulopathy. J Neurol Neurosurg Psychiatry 74:691, 2003

53. Halmagyi GM, Fattore CM, Curthoys IS, et al: Gentamicin vestibulotoxicity. Otolaryngol Head Neck Surg 111:571, 1994

54. Riggs LC, Shofner WP, Shah AR, et al: Ototoxicity resulting from combined administration of metronidazole and gentamicin. Am J Otol 20:430, 1999

55. Chen JM, Williamson PA, Hutchin T, et al: Topical gentamicin-induced hearing loss: a mitochondrial ribosomal RNA study of genetic susceptibility. Am J Otol 17:850, 1996

56. Cutrer FM, Baloh RW: Migraine-associated dizziness. Headache 32:300, 1992

57. Wieser T, Wolff R, Hoffmann KP, et al: Persistent ocular motor disturbances in migraine without aura. Neurol Sci 25:8, 2004

58. Bickerstaff ER: Basilar artery migraine. Lancet 277:15, 1961

59. Neuhauser H, Leopold M, von Brevern M, et al: The interrelations of migraine, vertigo, and migrainous vertigo. Neurology 56:436, 2001

60. Viirre ES, Baloh RW: Migraine as a cause of sudden hearing loss. Headache 36:24, 1996

61. Bikhazi P, Jackson C, Ruckenstein MJ: Efficacy of antimigrainous therapy in the treatment of migraine-associated dizziness. Am J Otol 18:350, 1997

62. Neuhauser H, Radtke A, von Brevern M, et al: Zolmitriptan for treatment of migrainous vertigo: a pilot randomized placebo-controlled trial. Neurology 60:882, 2003

63. Minor LB: Labyrinthine fistulae: pathobiology and management. Curr Opin Otolaryngol Head Neck Surg 11:340, 2003

64. Minor LB, Solomon D, Zinreich JS, et al: Sound- and/or pressure-induced vertigo due to bone dehiscence of the superior semicircular canal. Arch Otolaryngol Head Neck Surg 124:249, 1998

65. Shepard NT, Telian SA, Niparko JK, et al: Platform pressure test in identification of perilymphatic fistula. Am J Otol 13:49, 1992

66. Halmagyi GM, McGarvie LA, Aw ST, et al: The click-evoked vestibulo-ocular reflex in superior semicircular canal dehiscence. Neurology 60:1172, 2003

67. Rizer FM, House JW: Perilymph fistulas: the House Ear Clinic experience. Otolaryngol Head Neck Surg 104:239, 1991

68. Brandt T, Dieterich M: VIIIth nerve vascular compression syndrome: vestibular paroxysmia. Baillieres Clin Neurol 3:565, 1994

69. Moller AR: The cranial nerve vascular compression syndrome: I. A review of treatment. Acta Neurochirurgica 113:18, 1991

70. Baloh RW: Paraneoplastic cerebellar disorders. Otolaryngol Head Neck Surg 112:125, 1995

71. Jen J, Kim GW, Baloh RW: Clinical spectrum of episodic ataxia type 2. Neurology 62:17, 2004

72. Strupp M, Schuler O, Krafczyk S, et al: Treatment of downbeat nystagmus with 3,4-diaminopyridine: a placebo-controlled study. Neurology 61:165, 2003

73. Migliaccio AA, Halmagyi GM, McGarvie LA, et al: Cerebellar ataxia with bilateral vestibulopathy: description of a syndrome and its characteristic clinical sign. Brain 127:280, 2004

74. Smith BH: Vestibular disturbances in epilepsy. Neurology 10:465, 1960

75. Penfield W: Vestibular sensation and the cerebral cortex. Ann Otol Rhinol Laryngol 66:691, 1957

76. Kluge M, Beyenburg S, Fernandez G, et al: Epileptic vertigo: evidence for vestibular representation in human frontal cortex. Neurology 55:1906, 2000

77. Tusa RJ, Kaplan PW, Hain TC, et al: Ipsiversive eye deviation and epileptic nystagmus. Neurology 40:662, 1990

78. Francis DA, Bronstein AM, Rudge P, et al: The site of brainstem lesions causing semicircular canal paresis: an MRI study. J Neurol Neurosurg Psychiatry 55:446, 1992

79. Fetter M: Assessing vestibular function: which tests, when? J Neurol 247:335, 2000

80. Frohman EM, Kramer PD, Dewey RB, et al: Benign paroxysmal positioning vertigo in multiple sclerosis: diagnosis, pathophysiology and therapeutic techniques. Mult Scler 9:250, 2003

81. Frohman EM, Zhang H, Dewey RB, et al: Vertigo in MS: utility of positional and particle repositioning maneuvers. Neurology 55:1566, 2000

82. Committee on Hearing and Equilibrium guidelines for the diagnosis and evaluation of therapy in Meniere's disease. American Academy of Otolaryngology–Head and Neck Foundation, Inc. Otolaryngol Head Neck Surg 113:181, 1995

Acknowledgment

Figures 1 through 5 Tom Moore.

II DISEASES OF THE PERIPHERAL NERVOUS SYSTEM

COLIN H. CHALK, M.D., C.M.

Anatomy of the Peripheral Nervous System

The peripheral nervous system (PNS) comprises all of the cranial nerves (except the first and second), the nerve roots, the spinal ganglia, the segmental nerves, the nerve plexuses, and the limb nerves. The role of the PNS is simple but vital: to link the central nervous system to the external environment and the milieu intérieur of the body.

Peripheral nerves are composed of fascicles of endoneurium, containing several thousand axons and various supporting cells [*see Figure 1*]. Each endoneurial fascicle is enveloped by the perineurium (several concentric layers of fibroblast-like cells). Nerve trunks are composed of several endoneurial fascicles bound together by the epineurium (a loose framework of collagen and fibroblasts). There are two complementary networks of blood vessels in nerves: (1) an epineurial system of arterioles and venules, which are tributaries of the major limb vessels, and (2) a series of capillary-like microvessels in the endoneurium. The anastomoses between these two networks are extensive, resulting in a robust supply system.

Each axon is a longitudinal extension of a neuron cell body. The functioning and maintenance of axons depend on the cell body's biochemical and genetic machinery. An energy-dependent bidirectional system of axonal transport shuttles structural and signaling molecules between the cell body and the axon terminal. Motor, sensory, and autonomic axons are intermingled in most nerves, but the cell bodies are grouped separately in the spinal cord and brain stem (motor), the dorsal root ganglia (sensory), and the paravertebral and somatic ganglia (autonomic).

Roughly one quarter of the axons in the PNS are myelinated. Myelin is produced and maintained by Schwann cells, the most important of the supporting cells in the endoneurium. During development, Schwann cells elaborate increased amounts of cytoplasmic membrane to form a spiral around some axons. The cytoplasmic membrane forms the lamellar wrapping of myelin. A single Schwann cell myelinates 0.2 to 1.8 mm of cytoplasmic membrane along the length of a single axon, meaning that the myelin along the entire length of a single axon is the collective product of the actions of several hundred consecutive Schwann cells. The more numerous unmyelinated axons are also intimately associated with Schwann cells. Normal development of axons and normal development of Schwann cells are interdependent, and many details of their complex interactions are known.[1]

Clinically Significant Features of the Peripheral Nervous System

To early anatomists, dividing the nervous system into central and peripheral parts was intuitively reasonable. This division is not arbitrary; several differences between the PNS and the CNS are of clinical importance. The most apparent difference between the two is the functional simplicity of the PNS. In essence, the PNS is a simple relay system carrying motor, sensory, and autonomic signals between the CNS and somatic structures. The clinical correlate is the limited repertoire of symptoms and signs produced by malfunction of the PNS. One can infer relatively little about the cause of a peripheral nerve disorder from the symptoms and signs alone.

A second distinctive feature of the PNS is its accessibility. Unlike the brain and the spinal cord, which are well protected by bony cases, most of the major peripheral nerves reside for much of their course only a short distance beneath the skin. This accessibility simplifies direct study of the electrical and histologic characteristics of the peripheral nerves. Clinical evaluation of PNS diseases has been greatly facilitated by nerve conduction studies and nerve biopsies, which can be performed with essentially no risk to the patient.

Regenerative capacities differ between the PNS and the CNS. In the mature CNS, little or no regrowth of axons occurs after injury. However, an injured peripheral nerve can often regenerate over long distances and reestablish functional connections. Thus, many peripheral nerve disorders have the potential for significant recovery of function with treatment.

The difference in regenerative capacity was long thought to be related to some intrinsic property of PNS neurons. However, it appears that the critical issue is the environment to which neurons are exposed.[1] If a bundle of CNS nerve fibers (e.g., the optic nerve or the spinal cord) is transected and then immediately reanastomosed, there is little axonal outgrowth. However, if the proximal end of the same lesion is anastomosed to a peripheral nerve graft, axonal sprouts will extend many millimeters through the peripheral nerve graft and may even establish functional connections with target CNS neurons.

The cellular and molecular basis for this difference between PNS and CNS supporting tissues is not fully understood. The Schwann cell (which is exclusive to the PNS) appears key, although other cells (e.g., fibroblasts and macrophages) are probably involved. It is thought that the Schwann cell and the extracellular milieu of the peripheral nerve provide a favorable environment for axonal growth, attracting and perhaps guiding axonal sprouts. Two major classes of molecules influence the growth and development of sprouting axons in vitro: growth factors and adhesion molecules. Some of these molecules (e.g., nerve growth factor, N-cadherin, and laminin) have been well characterized.[1]

General Approach to Peripheral Nerve Diseases

CLINICAL MANIFESTATIONS

PNS disorders produce combinations of motor, sensory, and autonomic symptoms. These symptoms are primarily determined by the class of nerve fibers affected (e.g., motor or sensory fibers) and by the location of the lesions, rather than by the etiology of the process. Motor symptoms include weakness (a common presenting complaint), muscle atrophy, fasciculations, and cramps; the last two motor symptoms may be discovered only by specific questioning of the patient. Sensory symptoms include the loss of certain types of sensation and the presence of prickling, pins-and-needles, and burning sensations. The patient's own description of the quality and distribution of sensory symptoms is often more revealing than the sensory examination.

Pain is a common symptom in peripheral nerve disorders,

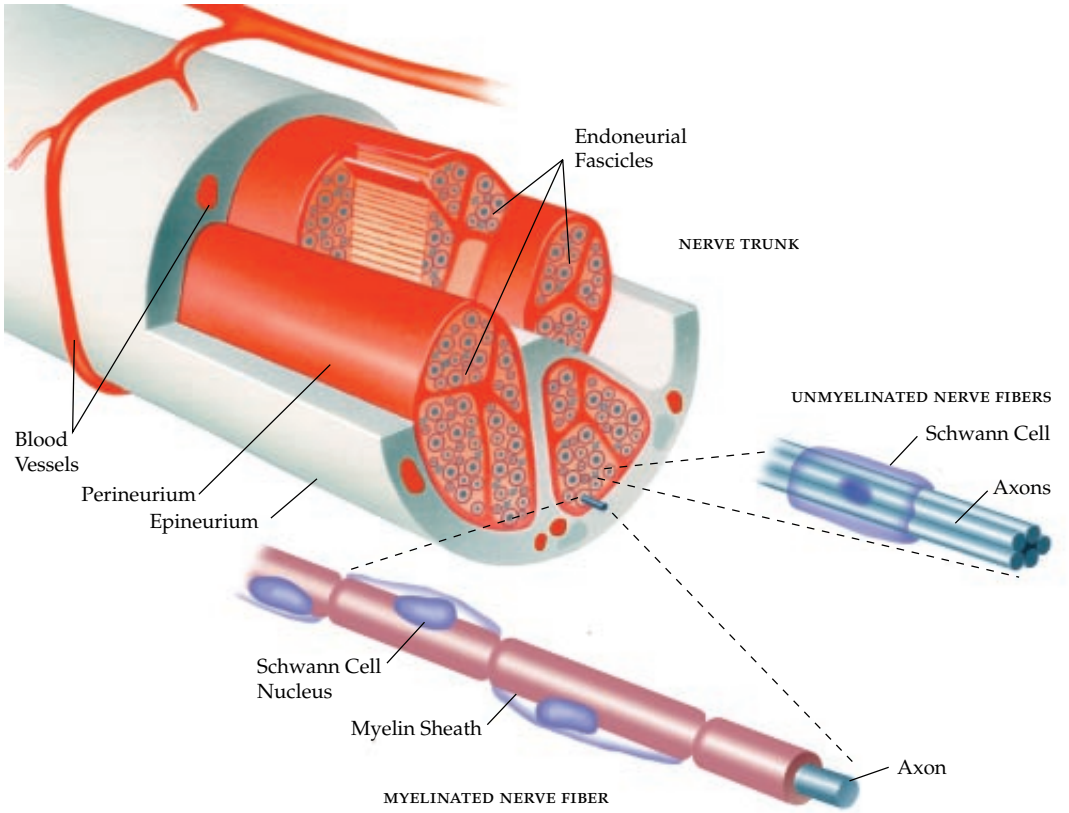


Figure 1 **Peripheral nerves are composed of several fascicles of endoneurium, each enveloped by perineurium and bound together by loose connective tissue, the epineurium. Each fascicle of endoneurium contains up to several thousand axons. Blood is supplied to the endoneurium by a network of capillary-like microvessels derived from arterioles and venules in the epineurium, which in turn are branches of major limb vessels.**

and it often occurs in association with other neuropathic symptoms. If pain is the only symptom, its cause is probably related to structures other than the peripheral nerves. Autonomic symptoms are heterogeneous and are often discovered only by direct inquiry because they may not seem especially neurologic to the patient. Common autonomic symptoms are orthostatic faintness or dizziness, excessive or diminished sweating, impotence or ejaculatory failure, and impaired gastrointestinal motility (particularly, delayed gastric emptying).

LABORATORY TESTS

Laboratory investigations are often required to clarify the nature of a peripheral nerve disorder. In addition to testing used in the evaluation of other types of neurologic diseases (e.g., lumbar puncture), techniques for the direct study of peripheral nerve physiology and histology are available. Neurophysiologic testing (especially nerve conduction studies) can be considered a routine part of the evaluation of any patient with polyneuropathy. Nerve biopsy, on the other hand, is best reserved for patients with disabling neurologic symptoms and signs.

Nerve Conduction Studies

Nerve conduction studies can provide support for a diagnosis of neuropathy and are a relatively objective way to follow the course of the disease. These studies may also allow a clinician to infer whether a disease is affecting primarily axons or their myelin sheaths—an important distinction in the differential diagnosis of polyneuropathy. Nerve conduction studies provide an indication of the numbers of functioning large myelinated motor and sensory axons in various nerves and the speed of impulse transmission in these fibers. The procedure consists of depolarizing a short segment of nerve with a brief electric shock and recording the resultant volley of action potentials as they propagate down the nerve. Both motor and sensory fibers are activated by this procedure; the motor and sensory responses are distinguished from each other by recording either from a cutaneous branch of the nerve (sensory) or from an innervated muscle (motor). If the nerve conduction studies show decreased motor and sensory amplitudes but preserved conduction velocities, the underlying pathologic process is probably loss of or destruction of axons. By contrast, if amplitudes are relatively preserved but there is marked slowing of conduction velocity or conduction block, an abnormality of the myelin sheath is likely.

Needle Electromyography

Needle electromyography is a complementary investigation that is usually performed at the same time as nerve conduction studies. In a patient with peripheral neuropathy, electromyography is helpful in detecting small degrees of axon loss that may go undetected by nerve conduction studies. Electromyography can also be helpful in pinpointing the precise location of the lesion in a mononeuropathy. The procedure entails recording electrical activity in muscle by use of a needle electrode inserted into the muscle. Muscles are examined at rest and during mild voluntary contraction. If the electrical activity produces an abnormal pattern, the electromyographer can determine whether a patient's weakness is the result of a diseased muscle or a diseased motor nerve.

Nerve Biopsy

Nerve biopsy can be diagnostically useful in carefully selected patients. Nerve biopsy can provide a specific diagnosis (e.g., necrotizing vasculitis) but more often gives less specific information that can be useful in combination with the other clinical data. Apart from alteration in the number or size of axons, the most common abnormality found by nerve biopsy is inflammation, usually related to altered immune function. On occasion, a disease that affects both the CNS and the PNS (e.g., metachromatic leukodystrophy) can be conveniently diagnosed by nerve biopsy. The sural nerve is preferred in most instances because sural nerve biopsy results in a small area of generally well-tolerated sensory loss along the lateral side of the foot.

Proper pathologic examination of a nerve biopsy specimen requires techniques that are often available only in specialized laboratories. When such facilities are not available locally, the patient should be referred to an appropriate center.

TREATMENT OF NEUROPATHIC SYMPTOMS

The symptoms of neuropathy may be treatable even if the cause of the neuropathy is untreatable or unknown. With simple measures, many patients can have meaningful relief of symptoms.

The ankle-foot orthosis is a simple device that compensates for footdrop. The main benefit offered by an ankle-foot orthosis is greater stability at the ankle, resulting in better balance and overcoming the tendency to catch the toes on edges of steps, curbs, and carpets. A lightweight, properly fitted ankle-foot orthosis that slips into the shoe is usually the most comfortable. Patients with weakness of wrist and finger extensors may be helped by a brace that keeps the wrist and fingers in a neutral position. Unfortunately, this type of splint is of little help when there is significant concomitant weakness of intrinsic hand muscles.

Pain, particularly in the feet, often accompanies the sensory disturbances of neuropathy. Paradoxically, in polyneuropathies with prominent pain, the neurologic deficits are often minor and there is a dissociation between the actual disability and the level of the patient's distress. Patients may interpret pain as a sign of a serious disorder that threatens their independence or their life. Sometimes, simple reassurance, emphasizing how little neurologic function has been lost, helps the patient cope effectively with neuropathic pain.

Nonpharmacologic measures may be as effective as medications. Careful attention should be paid to footwear. Loose-fitting, soft-soled shoes and thick socks are advisable. Neuropathic pain tends to be aggravated by extremes of temperature—especially heat—and open-toed sandals may give relief. Prolonged weight bearing often worsens neuropathic pain; patients whose work requires them to be on their feet may be helped by frequent short sitting breaks.

Foot soaking produces short-lived relief in many patients. Soaking the feet in ankle-deep cold tap water (without ice) for 15 to 20 minutes can be particularly helpful at bedtime. Although the relief is short-lived, it may be sufficient to allow the patient to fall asleep and to sleep well. For some patients, warm water is better than cold water, and others find that alternating cold and hot water (so-called contrast soaks) provides the best relief. Daily inspection of the feet for undetected injuries—an important habit that patients with neuropathy should develop—can be conveniently combined with nightly foot soaks.

Medications can be useful in the management of neuropathic pain, but the goals of therapy should be realistic. Complete pain relief is unlikely. Therefore, the aim of therapy should be to make the pain more tolerable without adding intolerable side effects of medication. Of the many drugs that can be tried for neuropathic pain, tricyclics (especially amitriptyline) and carbamazepine are still most frequently used, although gabapentin is increasingly used as a first-line agent. Lancinating, paroxysmal pains may be more likely to respond to carbamazepine (200 mg t.i.d.), whereas amitriptyline (10 to 30 mg at bedtime) is the preferred agent for the more common, continuous burning numbness. Several randomized, controlled trials in patients with neuropathic pain of various etiologies have shown a clear benefit of gabapentin over placebo in several measures of pain relief.[2] Effective dosages of gabapentin vary considerably; in most patients, a good initial dosage is 300 mg three times a day, which can be titrated upward over several weeks to 1,800 to 3,600 mg daily or higher, according to response.

At least 1 month of treatment should be undertaken before any conclusions are drawn about the usefulness of a drug for neuropathic pain. In addition to tricyclics, carbamazepine, and gabapentin, salicylates and other simple analgesics give relief to some patients. As is the case with other types of chronic pain, the usual practice has been to avoid narcotics, but a new appreciation for the potential benefits of opioids in managing chronic neuropathic pain is developing.[3] Second-line drugs for neuropathic pain include baclofen, mexiletine, valproic acid, and topiramate. Topical capsaicin ointment causes depletion of the neurotransmitter substance P in the dorsal horn of the spinal cord and helps some patients. For most patients, however, the expense and inconvenience of topical capsaicin outweigh any benefits.

In some patients, the pain remains refractory to all of these measures. These patients present difficult management problems. Concurrent depression can complicate the situation, and psychiatric referral may be worthwhile. Referral to a multidisciplinary pain management center can also be considered.

Peripheral Nerve Diseases

It is helpful to divide PNS diseases into those that affect single nerves (mononeuropathies) and those that involve the PNS diffusely (polyneuropathies). A mononeuropathy may involve any cranial or limb nerve. In general, mononeuropathies occur in otherwise healthy patients, cause troublesome symptoms but no major disability, and often improve spontaneously with time.

CRANIAL MONONEUROPATHIES

Neuropathies can arise in any individual cranial nerve [*see Table 1*]. The seventh (facial), fifth (trigeminal), and third (oculomotor) cranial nerves account for most cranial mononeuropathies encountered in clinical practice.

Facial Neuropathy

Acute, idiopathic facial neuropathy (Bell palsy) is the most common cranial neuropathy, with an annual incidence of about 25 per 100,000 persons. The main symptom is unilateral facial weakness that begins abruptly and is often preceded by or accompanied by pain behind the ear. In some patients, a mild upper respiratory tract infection may have occurred up to 2 weeks earlier. The facial weakness usually reaches its zenith within 24 hours after onset and is of the lower motor neuron type, affecting both the upper and the lower face. Some patients experience altered taste on one half of the tongue and hyperacusis. Patients often complain of numbness of the face, but sensory loss is not found.

Table 1 Common Cranial Neuropathies

Cranial Nerve	*Effects of Lesion*	*Causes*
Olfactory (I)	Loss of smell	Trauma, olfactory groove meningioma
Optic (II)	Loss of vision (monocular or bitemporal hemianopsia)	Optic neuritis, pituitary tumor, ischemic optic neuropathy, optic nerve glioma, Leber disease
Oculomotor (III)	Weakness of adduction, elevation, and depression of eye Ptosis Dilated, nonreactive pupil	Trauma, microvessel ischemia, compression by aneurysm or mass, brain stem stroke or tumor, orbital tumor
Trochlear (IV)	Weakness of depression and intorsion of eye	Trauma, microvessel ischemia, brain stem stroke or tumor, orbital tumor
Trigeminal (V)	Weakness of chewing and other jaw movements Loss of facial sensation and corneal reflex	Trigeminal neuralgia (tic douloureux), scleroderma and other connective tissue diseases, tumor of petrosal bone
Abducens (VI)	Weakness of abduction of eye	Trauma, raised intracranial pressure, microvessel ischemia, brain stem stroke or tumor, orbital tumor
Facial (VII)	Weakness of half of face Loss of taste Hyperacusis	Bell palsy, Lyme disease, sarcoidosis, herpes zoster, cerebellopontine angle tumor, pontine stroke or tumor
Vestibulocochlear (VIII)	Unilateral deafness Vertigo Nystagmus	Vestibular neuronitis, acoustic schwannoma
Glossopharyngeal (IX)	Dysphagia Weakness of elevation of palate Loss of gag reflex	Motor neuron disease, jugular foramen tumor, nasopharyngeal carcinoma, metastases to skull base, glossopharyngeal neuralgia
Vagus (X)	Dysphagia Dysarthria Weakness of elevation of palate Loss of gag reflex	Motor neuron disease, jugular foramen tumor, metastases to skull base, nasopharyngeal carcinoma
Accessory (XI)	Weakness of sternocleidomastoid and trapezius muscles	Trauma (surgical and other), jugular foramen tumor
Hypoglossal (XII)	Weakness of tongue	Motor neuron disease, skull base tumor, carotid artery dissection or trauma

When the history and examination are typical of Bell palsy and the condition is mild and nonprogressive, no tests are needed. Severe or progressing cases should be referred to a neurologist to confirm the diagnosis and decide whether intervention is needed. Occasionally, sarcoidosis, Lyme disease, and herpes zoster (shingles) present as facial neuropathy; testing to exclude these conditions may be appropriate. When the neurologic examination reveals more than a unilateral facial neuropathy or when the weakness has evolved over weeks or months, an alternative diagnosis should be sought, such as a cerebellopontine angle tumor or a brain stem lesion.

The natural history of untreated Bell palsy is generally favorable: 85% of patients recover fully within 1 year. However, some clinical features are associated with a poorer prognosis, including old age, complete facial palsy, and the presence of hyperacusis or altered taste. Hemifacial spasm (a syndrome of painless, brief, involuntary contraction of the facial muscles) occasionally develops after Bell palsy and other types of facial neuropathy.

Although there is general agreement that the pathogenesis of Bell palsy involves facial nerve inflammation, there is debate about whether the condition is a postinfectious autoimmune phenomenon or the result of a direct viral infection of the facial nerve. Herpes simplex virus type 1 has been implicated by the presence of its DNA in facial nerve specimens from 11 of 14 patients with Bell palsy but not in patients with other causes of facial neuropathy, including herpes zoster.[4]

Because of the presumed role of an inflammatory response, Bell palsy is often treated with glucocorticoids. Several studies have suggested that glucocorticoid therapy speeds recovery. However, a meta-analysis of the few placebo-controlled trials concluded that clear evidence for the efficacy of glucocorticoids had not been demonstrated.[5] Nevertheless, for patients seen within a few days after onset of Bell palsy, it is reasonable to follow the common practice of prescribing a short, tapering course of glucocorticoids (e.g., prednisone, 60 mg/day for 5 days, tapering to zero over the next 10 days). A controlled trial of acyclovir plus prednisone versus prednisone alone found a modest but significant improvement in clinical outcome in the acyclovir-treated group.[6] On the basis of this study, the combination of acyclovir and prednisone can reasonably be considered for most patients with Bell palsy, although the natural history of the condition is generally benign, regardless of treatment.

Prevention of corneal desiccation with an eye patch is important for patients who cannot fully close the eye. Facial nerve decompression was once widely used, but the procedure has been largely abandoned because its efficacy is unproved. Injections of botulinum toxin are helpful in treating hemifacial spasm.

Trigeminal Neuralgia (Tic Douloureux)

Trigeminal neuralgia, or tic douloureux, is a distinctive facial pain syndrome of middle-aged and elderly persons in which intense, paroxysmal pain occurs several times to dozens of times daily. The pain is often likened to an intense electric shock. The pain lasts for seconds and is localized to the trigeminal distribution, usually the maxillary or mandibular territories. Paroxysms of pain may be triggered by minor tactile stimuli on the affected side of the face, such as shaving, washing, or chewing. Between paroxysms, patients may have dull background pain but not numbness or paresthesia. Trigeminal neuralgia may be mistaken for dental pain if the characteristic features are not recognized. A similar but much less common syndrome, glossopharyngeal neuralgia, produces paroxysms of throat or pharyngeal pain.

Examination of the cranial nerves is often difficult in patients with trigeminal neuralgia, because paroxysms of pain may be triggered by the examination. Nevertheless, sensory loss or other signs are not found. Cranial nerve and other neurologic signs should prompt investigation for structural lesions of the trigeminal nerve or nucleus or for a brain stem lesion. Multiple sclerosis is the most likely diagnosis in this situation, but this diagnosis has usually been well established before episodes of trigeminal neuralgia occur.

The cause of trigeminal neuralgia is subject to debate, although many investigators believe that compression or distortion of the trigeminal nerve by pulsations from an ectatic superior cerebellar artery is responsible. Spontaneous remissions may occur but are seldom of lasting duration.

The mainstay of medical treatment is carbamazepine, 600 to 1,200 mg daily. Alternative therapies are baclofen, 30 to 60 mg daily, and gabapentin, 300 to 1,200 mg three times a day. If trigeminal neuralgia becomes refractory to medication, surgical approaches can be tried and are usually successful. The appropriate trigeminal branch can be injected with alcohol or phenol, or a lesion can be made percutaneously in the trigeminal ganglion. However, with these procedures, the patient is trading pain for varying degrees of anesthesia, which is not always preferable. Microvascular decompression to relieve the hypothesized effects of the superior cerebellar artery seems to be highly effective in the hands of an experienced surgeon, but this procedure requires a posterior fossa craniotomy.

Oculomotor Neuropathy

Isolated third-nerve palsies are not common, but they present a difficult differential diagnosis between life-threatening and benign causes. The onset is abrupt, diplopia and headache being the usual symptoms. The diplopia is both horizontal and vertical, and the relation between the two images changes with the direction of gaze. The headache may be retro-orbital or diffuse. Most patients are aware of ptosis, and some may note pupil dilatation.

The main considerations in a patient with an isolated third-nerve palsy are compression by an expanding mass (particularly an aneurysm of the posterior communicating artery) and a so-called microvascular lesion, in which the oculomotor palsy is assumed to be ischemic and to be caused by a disease of the nerve's blood supply. Once it is clear that the only signs are those of a unilateral third-nerve palsy, examination may help distinguish between a compressive and a microvascular lesion. If there is pupillary paralysis, a compressive lesion is more likely, and cranial imaging with CT or MRI and cerebral angiography are indicated. A pupil-sparing third-nerve palsy is probably microvascular; patients are typically elderly, diabetic, and hypertensive. However, in a young patient who is neither diabetic nor hypertensive, cranial imaging is advisable regardless of whether the pupil is spared. For vascular imaging, magnetic resonance angiography and CT angiography, which are noninvasive, are increasingly replacing conventional cerebral angiography.

MONONEUROPATHIES OF LIMB NERVES

Mononeuropathies of limb nerves are common. Mechanical compression of the nerve at a vulnerable location is often assumed to be the cause, but this assumption is not always convincingly proved. An identical clinical picture may be caused by other rare pathologies, such as a nerve tumor. The autosomal dominant syndrome of inherited susceptibility to pressure palsies should be considered in a patient with recurrent compressive mononeuropathies or when the precipitating trauma is trivial compared with the resulting deficit.[7]

Carpal Tunnel Syndrome

Carpal tunnel syndrome (CTS) is caused by compression of the distal median nerve as it traverses the tunnel in the wrist formed by carpal bones inferiorly and the carpal ligament superiorly [*see 15:XII Back Pain and Common Musculoskeletal Problems*].

Diagnosis In CTS, the patient complains of numbness in the hand (e.g., prickling sensations, sensory loss, or both) that is unpleasant or painful. The numbness is usually within the median sensory distribution, but some patients report involvement of all digits or discomfort in the forearm. The symptoms are characteristically episodic and occur at night (often waking the patient) and during activities in which the wrists are held in flexion or extension, such as knitting or holding a steering wheel. Vigorously shaking the hand often produces prompt relief, and this is very suggestive of CTS. Numbness may make it difficult or uncomfortable for the patient to use the hand properly, but true weakness is uncommon. When present, impaired prehensile grip related to weakness of the thenar muscles may interfere with activities such as holding a pen. In some patients (particularly the elderly), sensory complaints may be minimal and symptoms of thenar weakness may predominate.

In most patients, CTS probably develops because of a congenitally narrow carpal tunnel. Certain occupations that require repetitive flexion and extension of the wrist (e.g., carpentry, sewing, and use of a computer keyboard) may predispose a patient to CTS. The syndrome is common in pregnancy (presumably because of generalized fluid retention) and may occur after injuries such as a Colles fracture. On occasion, CTS occurs in the context of a systemic illness that decreases leeway in the carpal tunnel, such as rheumatoid arthritis, myxedema, amyloidosis, and acromegaly. Patients with diabetic and certain other polyneuropathies may be predisposed to superimposed CTS and other mononeuropathies. In patients with diabetes, increased stiffness and volume of the carpal ligament may be an important factor.

Examination may reveal sensory or motor deficits in a median distribution, but it is often normal, particularly in young patients with intermittent symptoms. Percussion over the median nerve at the wrist may produce transient paresthesia in median-innervated digits (Tinel sign). Holding the wrist in forced extension or flexion for 1 to 2 minutes may precipitate the patient's sensory symptoms, which are promptly relieved by hanging the arm loosely at the side without angulation at the wrist (Phalen sign). Neither the Tinel sign nor the Phalen sign is entirely specific or sensitive.

Nerve conduction studies often help support a clinical diagnosis of CTS. The essential principle is to show slowing of median nerve conduction in the carpal tunnel relative to the neighboring ulnar or radial nerves. Nerve conduction studies have a high diagnostic sensitivity in CTS but may be normal in mild cases. Study results usually improve but do not return to normal even if the patient's symptoms are completely relieved by surgery.

Treatment Patients who have mild symptoms but have no deficits may need only reassurance. A splint that holds the wrist in the neutral position is often effective for nocturnal symptoms. Some patients are helped by periodic injections of glucocorticoids into the carpal ligament. Surgical sectioning of the carpal ligament, which is a low-risk procedure, is widely used for CTS; and the procedure seems to result in substantial, long-lasting improvement in most patients. There is conflicting evidence about whether endoscopic techniques allow for faster recovery than open surgical procedures.[8] Surgery is advised for patients with troublesome symptoms that are unresponsive to nonsurgical treatments and for those patients with progressive motor or sensory deficits.

Ulnar Neuropathy

Ulnar neuropathy most often results from mechanical injury at the elbow, where the nerve lies subcutaneously and rests on the unyielding floor of the epicondylar groove. Distal to the elbow, the nerve can be compressed as it passes under the aponeurosis of the flexor carpi ulnaris into the deep muscle compartment of the forearm. At the wrist, the nerve again becomes relatively superficial and may be compressed by excessive weight bearing on the hands, as with long-distance cycling.

Diagnosis Patients with ulnar neuropathies present with varying degrees of motor and sensory symptoms. Observant patients will notice that the sensory disturbance involves the medial half of the ring finger, in addition to the little finger and medial hand. Weakness, when present, is usually confined to the hand. The hand may seem generally weak, or there may be specific complaints, such as difficulty holding a pen or spreading the fingers.

Weakness and wasting are most easily identified in the first dorsal interosseous muscle (between the thumb and index finger) and the hypothenar muscles. Complete loss of ulnar sensation is uncommon, but sensation is often altered in the ulnar distribution. Comparing the medial (ulnar) and lateral (median) sides of the ring finger may provide convincing evidence of an ulnar neuropathy. Production of numbness or electrical sensations in the hand with gentle manipulation of the nerve at the elbow favors the elbow as the site of the pathology. Electrodiagnostic testing is useful in ulnar neuropathies both to confirm the diagnosis and to localize the precise site of the lesion.

In rare instances, ulnar neuropathy can result from compression in the upper arm or in the hand. Because of the arrangement of the motor and sensory branches in the wrist and hand, ulnar compression in the hand may involve motor branches only. This condition results in hand weakness but no pain or sensory symptoms and may be mistaken for early motor neuron disease. This type of ulnar neuropathy is caused by weight bearing on the hands, as may occur with the use of a four-point walker or with prolonged cycling.

Treatment The appropriate treatment for ulnar neuropathy at the elbow is unclear. Although the condition is common, there is surprisingly little good information about its natural history. Spontaneous improvement is frequent, particularly in patients with a clear precipitating episode of compression. In patients in whom repeated minor trauma to the nerve is thought to be the cause, ulnar neuropathy often improves if they avoid leaning on their elbows.

Several surgical procedures are used in the treatment of ulnar neuropathy at the elbow.[9] In selected patients who have progressive symptoms and signs despite conservative therapy, surgery may be considered, but the relative merits of surgery versus nonoperative treatments have not been carefully studied.

Brachial Plexus Neuropathy

Brachial plexus neuropathy (also known as neuralgic amyotrophy and Parsonage-Turner syndrome) is uncommon but probably underdiagnosed. The history is distinctive. The patient experiences severe, aching, bursitislike pain in the shoulder without an apparent precipitant. After several days to 2 weeks, the pain subsides and the patient becomes aware of weakness in the arm. Most often, the shoulder girdle muscles are affected, but the hand or forearm may be involved. Sometimes, the weakness is strikingly focal or patchy. There may be sensory loss in an epaulet distribution, but it is not a prominent feature. Patients gradually recover, and 90% have little or no deficit 1 to 2 years after onset. In some patients, the two arms are affected either simultaneously or sequentially, and some patients have recurrent episodes.

Brachial plexus neuropathy is most common in young and middle-aged men but can affect both sexes and all age groups. Biopsies of the brachial plexus have shown florid lymphocytic inflammation, and the disorder is believed to be immune mediated.[10] However, immunomodulation therapies are of no proven benefit. An autosomal dominant form of brachial plexus neuropathy also exists and is characterized by recurrent episodes of shoulder pain and weakness.

Peroneal Neuropathy

Peroneal neuropathy is most often the result of injury to the peroneal nerve at the point where the nerve winds around the head of the fibula, passing from the popliteal fossa to the anterior compartment of the leg. Distal peroneal branches are superficial at the ankle, where they are sometimes compressed by tight-fitting shoes.

Diagnosis Patients with peroneal neuropathy at the knee usually present with unilateral footdrop caused by weakness of ankle dorsiflexors. However, footdrop may be caused by an L5 radiculopathy, a lumbosacral plexopathy, a proximal sciatic neuropathy, or a lesion in the contralateral cerebral motor cortex. L5 radiculopathy, the most common differential diagnosis, is suggested by the presence of back pain radiating to the leg and by weakness of ankle invertors and hip abductors in addition to the footdrop. Electrodiagnostic testing may be particularly helpful in localizing the lesion in patients with footdrop.

Trauma (e.g., fracture of the fibula) is often responsible for peroneal neuropathy at the knee. Compression during states of altered consciousness (e.g., during anesthesia and coma) is another common cause. Herniation of a Baker cyst may compromise the nerve. In patients without such precipitants, peroneal neuropathy at the knee is often attributed to habitually sitting with legs crossed—a plausible explanation, although difficult to prove.

Treatment Management of peroneal neuropathy at the knee is usually conservative. In most patients who have suffered trauma to the nerve, improvement will occur with time, particularly when there is some residual function in the distal peroneal nerve. In severe trauma, the nerve is sometimes completely transected, in which case the outcome is usually poor and surgical anastomosis offers the best hope of recovery. When no other cause of the peroneal neuropathy is apparent, patients can be advised to avoid sitting with legs crossed; slow improvement may result. Regardless of the cause, an ankle-foot orthosis should be tried to compensate for the footdrop; most patients find that this simple device greatly improves their walking [*see* Treatment of Neuropathic Symptoms, *above*].

Meralgia Paresthetica

Meralgia paresthetica is characterized by sensory symptoms in the distribution of the lateral cutaneous nerve of the thigh. Patients report some combination of sensory loss, prickling paresthesia, and hypersensitivity over the anterolateral thigh. The boundaries of the abnormal area are typically sharply defined, with no associated motor or reflex abnormalities. Meralgia paresthetica is usually attributed to compression of the lateral

Table 2 Distinctive Neuropathies with Limited Differential Diagnoses

Multiple mononeuropathy
Vasculitis
Diabetes mellitus*
Sarcoidosis
Cryoglobulinemia
Inherited susceptibility to pressure palsies
Leprosy
Lymphomatoid granulomatosis
HIV infection*
Neoplastic invasion of nerves or nerve roots
Lyme disease
Multifocal neuropathy with conduction block
Neurofibromatosis
Ataxic sensory neuropathy
Paraneoplastic sensory neuropathy (small cell lung or breast cancer)
Sjögren syndrome
Monoclonal protein–associated neuropathy
Cisplatin toxicity
Pyridoxine toxicity
Vitamin B_{12} deficiency
Tabes dorsalis
Nonmalignant inflammatory sensory polyganglionopathy
Inherited spinocerebellar ataxia
Neuropathy with prominent autonomic features
Diabetes mellitus
Amyloidosis
Hereditary sensory and autonomic neuropathy
Acute pandysautonomia
Paraneoplastic sensory autonomic polyganglionopathy
Acute-onset polyneuropathy
Guillain-Barré syndrome and variants
Vasculitic neuropathy
Porphyria
Lyme disease
Toxic substances (e.g., arsenic, lead, thallium)
Poliomyelitis
Diphtheria
Botulism

*Usually present as symmetrical polyneuropathy.

cutaneous nerve by the inguinal ligament as it passes from the retroperitoneum to the anterior thigh. The syndrome is more common in the obese, perhaps because of increased mechanical stress on the inguinal ligament. In rare instances, the lateral cutaneous nerve is injured during surgery, or it may be compromised by enlarged inguinal lymph nodes or other masses. Most patients require only reassurance, because the symptoms often subside with time. Treatment with medications, such as amitriptyline, can be tried for bothersome, persistent symptoms. Surgical exploration of the inguinal region is a last resort.

Lumbosacral Plexopathy

Lumbosacral plexopathy resembles brachial plexus neuropathy. The typical history of lumbosacral plexus involvement begins with pain in the anterior thigh, usually severe and continuous, lasting for several weeks. As the pain subsides, weakness appears, most often involving the quadriceps and hip flexors. Often, the weakness causes no symptoms until the patient falls, giving the misleading impression of an acute event. Impressive wasting of the quadriceps may occur, and the knee jerk is absent. Sensory signs are minimal. Sometimes, the initial symptoms and signs involve only the lumbar or sacral parts of the plexus, but there is a tendency for the process to spread to the whole plexus and to involve the contralateral side. Very gradual recovery is the rule.

Lumbosacral plexopathy characteristically affects middle-aged and elderly men. Patients often have diabetes (in which case the terms diabetic amyotrophy, diabetic femoral neuropathy, proximal diabetic neuropathy, and diabetic lumbosacral radiculoplexus neuropathy are sometimes used), although an identical syndrome can occur in nondiabetic patients. The pathogenesis of lumbosacral plexopathy is debated. An ischemic basis was long favored, particularly because the condition frequently occurs in patients with microvascular complications of diabetes. However, the balance of evidence from biopsy studies now favors an inflammatory process, resembling a restricted form of necrotizing vasculitis in some cases.[11]

Electrodiagnostic testing is helpful in supporting the diagnosis of lumbosacral plexopathy. Imaging studies of the pelvis to exclude malignant invasion of the lumbar plexus, retroperitoneal hematoma, and psoas abscess are often needed.

Immunomodulation has been recommended for treatment of lumbosacral plexopathy. Agents for this use have included corticosteroids, azathioprine, and intravenous immunoglobulin. Clear evidence of efficacy is lacking, but controlled trials are in progress. Pain that lasts more than a few weeks may be helped by drugs such as amitriptyline.

POLYNEUROPATHIES

Determining that a patient has a polyneuropathy is usually not difficult. However, because the hundred or so possible causes of polyneuropathy produce similar symptoms and signs, arriving at a specific diagnosis can be challenging. The differential diagnosis can be simplified by the following questions:

1. Do the symptoms and signs fit a pattern other than that of a symmetrical sensorimotor polyneuropathy? Several distinctive syndromes, unlike symmetrical sensorimotor polyneuropathy, have short differential diagnoses. Examples include multiple mononeuropathies (also known as mononeuritis multiplex), ataxic sensory neuropathy, pure or predominantly autonomic neuropathy, and acute-onset polyneuropathy [*see Table 2*].
2. Do nerve conduction studies suggest demyelination? After history and examination, nerve conduction studies are the next logical step. If marked slowing of conduction velocity or other electrophysiologic evidence suggesting demyelination is noted, the differential diagnosis is short [*see Table 3*].
3. Is the polyneuropathy acquired or inherited? This question is an important consideration in all patients with a chronic symmetrical polyneuropathy. Inherited neuropathy is the most frequently overlooked cause of undiagnosed polyneuropathy.[12] Pes cavus and hammer toes can be important clues toward a diagnosis of inherited neuropathy and should always be specifically sought. Inherited neuropathies are often mild, nondisabling, and very slowly progressive. Thus, they often do not present until middle age or later and often go unrecognized in families [*see Table 4*].

Further investigation of symmetrical, sensorimotor, axonal neuropathies can be guided by collateral information in the history. If inherited neuropathy is possible, first-degree relatives should be examined. The cause of an acquired neuropathy may

Table 3 Differential Diagnosis of Demyelinating Polyneuropathy

- Guillain-Barré syndrome (acute inflammatory demyelinating polyradiculoneuropathy)
- Chronic inflammatory demyelinating polyradiculoneuropathy
- Monoclonal protein–associated neuropathy
- Osteosclerotic myeloma
- Diphtheria
- Perhexiline maleate toxicity
- Hereditary motor and sensory neuropathy type 1
- Hereditary susceptibility to pressure palsies

be one of the patient's medications [*see Table 5*] or a concomitant medical illness, with diabetes mellitus and chronic renal failure being the most common. On occasion, neuropathy may be the presenting feature of systemic vasculitis or an occult malignancy.

For adult patients, reasonable routine laboratory tests may include a fasting blood glucose level, glycosylated hemoglobin concentration, serum creatinine level, complete blood count, chest x-ray (for smokers), erythrocyte sedimentation rate or C-reactive protein, rheumatoid factor, antinuclear antibody, and, in patients older than 50 years, immunofixation of plasma and urinary proteins. Nerve biopsy can be valuable when specific information is being sought. Biopsy is not a routine test, and it is seldom helpful when done as a last resort.

Diabetic Polyneuropathy

Peripheral neuropathy is common in patients with diabetes mellitus. Of the various types of diabetic neuropathy, by far the most common is a distal, symmetrical sensorimotor neuropathy, commonly referred to as diabetic polyneuropathy. Estimating the incidence and severity of diabetic polyneuropathy has been difficult because of (1) how the condition is defined, (2) the populations that have been studied, and (3) efforts to ensure that other causes of neuropathy have been excluded. In one prospective, population-based study of Americans mainly of northern European ancestry, diabetic polyneuropathy was found in 54% of patients with type 1 diabetes mellitus and 45% of patients with type 2 diabetes mellitus.[13] However, symptomatic polyneuropathy occurred in only 15% of the cohort, and none of the patients had disabling neurologic deficits. The severity of the polyneuropathy correlated more closely with the degree of hyperglycemia (mean glycosylated hemoglobin) than with the duration of diabetes.[14] In this study and other large studies, the prevalence of diabetic polyneuropathy increased with the duration of diabetes, and a strong correlation existed between the presence of diabetic polyneuropathy, retinopathy, and nephropathy. An important practical corollary of these observations is that a diagnosis of diabetic polyneuropathy in a patient with newly diagnosed diabetes but without other diabetic complications is likely to be incorrect.

Pathogenesis The pathogenesis of diabetic polyneuropathy continues to be debated. A popular theory is that increased activity of the polyol pathway, brought about by chronic hyperglycemia, leads to accumulation of sorbitol and depletion of *myo*-inositol in the nerve. Decreased *myo*-inositol, in turn, is postulated to produce reduced Na^+,K^+-ATPase activity, leading to altered nerve conduction and structural alterations of the axon. In diabetic animals, some of these abnormalities can be demon-

Table 4 Basic Features of Inherited Neuropathies

	Hereditary Motor and Sensory Neuropathy	*Hereditary Sensory and Autonomic Neuropathy (HSAN)*	*Hereditary Motor Neuropathy*
Clinical features	Distal muscle wasting, weakness, and sensory loss Pes cavus	Distal sensory loss Variable dysautonomia Neurogenic arthropathy Plantar ulcers	Muscle wasting and weakness, most often proximal
Site of abnormality	Axon or Schwann cell	Dorsal root and autonomic ganglia neurons	Anterior horn cells
Inheritance	Most cases are autosomal dominant Some cases are autosomal recessive or X-linked	Autosomal dominant Autosomal recessive	Autosomal dominant Autosomal recessive
Known genetic defects	Charcot-Marie-Tooth disease type 1 Duplication or point mutation of *PMP22* gene (chromosome 17) Point mutations of *Po* gene (chromosome 1) Point mutations of *Cx32* gene (X chromosome) Point mutations of early growth response gene (chromosome 10) Mutation of *LITAF/SIMPLE* gene (chromosome 16) Various genes cause autosomal recessive forms of the disease (mostly single kinships) Charcot-Marie-Tooth disease type 2 Linkage to 1p, 3q, and 7p Mutations of *KIF 1Bβ* (chromosome 1), *RAB7* (chromosome 3), *NF-L* (chromosome 8) genes, and linkage to other loci	HSAN type 1 (dominant) Mutation in serine palmitoyl transferase gene (chromosome 9) HSAN type 2 (recessive) Unknown HSAN type 3 (recessive) Mutation in 1κB kinase complex–associated protein gene (chromosome 9) HSAN types 4 and 5 (recessive) Mutations in tyrosine kinase A receptor gene (chromosome 1)	Werdnig-Hoffmann disease (and other childhood-onset types) Linkage to 5q Kennedy syndrome (X-linked bulbospinal muscular atrophy) Defect in androgen-receptor gene (*CAG* triplet repeat)

Table 5 Drugs That May Cause Polyneuropathy

- Antineoplastics
  - Cisplatin (ataxic sensory neuropathy is dose limiting)
  - Bortezomib
  - Suramin
  - Taxoids (docetaxel, paclitaxel)
  - Vinca alkaloids (vincristine, vinblastine; neuropathy is dose limiting)
- Antiretrovirals
  - Didanosine (dideoxyinosine)
  - Stavudine (d4T)
  - Zalcitabine (dideoxycytidine)
- Other antimicrobials
  - Chloramphenicol
  - Dapsone (mainly motor neuropathy; higher doses than in leprosy therapy)
  - Isoniazid (preventable by pyridoxine supplementation)
  - Metronidazole (prolonged courses)
  - Nitrofurantoin
- Drugs used in rheumatologic diseases
  - Chloroquine
  - Colchicine
  - Gold
  - Thalidomide
- Miscellaneous
  - Amiodarone (antiarrhythmic)
  - Disulfiram (Antabuse)
  - Perhexilene (demyelinating neuropathy)
  - Phenytoin (after years of use; very mild neuropathy)
  - Pyridoxine (megadoses cause ataxic sensory neuropathy)
  - Simvastatin

strated and corrected by inhibiting the polyol pathway, but in nerves in humans with diabetes, the *myo*-inositol content is normal. In clinical trials, small improvements in motor nerve conduction velocity have been observed during treatment with aldose reductase inhibitors, which suppress polyol-pathway activity. Whether such treatment is clinically useful is disputed.[15]

An ischemic-hypoxic basis for diabetic polyneuropathy is an alternative or complementary hypothesis.[16] Degeneration of pericytes and reduplication of capillary basal laminae occur in nerve and other tissues in diabetic patients, which may implicate microvessels in the pathogenesis of complications such as neuropathy. In diabetic rats, a reduction in nerve blood flow and endoneurial oxygen tension can be demonstrated, and partial correction of nerve conduction abnormalities can be achieved with oxygen supplementation. In biopsy specimens from persons with diabetic polyneuropathy, the degree of nerve fiber loss has been shown to be quite variable from fascicle to fascicle, a pattern seen in neuropathies with a clearly established ischemic basis, such as vasculitic neuropathy. At present, the hypoxic hypothesis is perhaps the most widely accepted, but this hypothesis does not necessarily exclude a role for the polyol pathway or other metabolic derangements.

Diagnosis Diabetic polyneuropathy has the classic so-called glove-and-stocking distribution of symptoms, usually a combination of sensory loss and an unpleasant feeling of numbness or burning. Sensory loss in the feet and fingers and mild weakness in the feet and ankles are typical. Diabetic polyneuropathy can be expected to worsen slowly over years,[17] although this progression may be influenced by treatment of the diabetes. Disabling neurologic deficits and rapid worsening are not expected in diabetic polyneuropathy; if present, they should prompt a search for other causes.

Treatment Data on the effect of glycemic control on diabetic polyneuropathy continue to accumulate. Of the patients with type 1 diabetes who were followed in the Diabetes Control and Complications Trial, polyneuropathy developed in 5% of those whose therapy included rigorous glycemic control and in 13% of those who were treated conventionally.[18] A follow-up study of a cohort of recipients of pancreas transplants showed significant improvement in nerve conduction parameters and perhaps a modest clinical improvement; in control patients treated with conventional insulin regimens who did not receive transplants, conditions worsened over the same period.[19] By contrast, the United Kingdom Prospective Diabetes Study, which involved nearly 4,000 patients with type 2 diabetes, failed to show a difference in the prevalence of neuropathy between patients treated conventionally and those treated with tight glycemic control during up to 10 years of follow-up.[20] Thus, optimal glycemic control appears to lower the risk of developing diabetic polyneuropathy in patients with type 1 diabetes, but no protective effect has been demonstrated in patients with type 2 diabetes. In established neuropathy, tight glycemic control may have a beneficial effect, but whether clinically significant improvement over several decades is possible remains to be shown.

The hope that additional pharmacologic interventions will prevent or alleviate diabetic neuropathy has spawned numerous clinical trials that have evaluated such drugs as aldose reductase inhibitors, acylcarnitine, aminoguanidine (decreases advanced glycation end products), evening primrose oil (an enriched source of linoleic acid), α-lipoic acid (an oxygen free-radical scavenger), and nerve growth factor. To date, some of these trials have suggested modest benefits, but problems of trial design (e.g., duration of treatment and whether results are clinically meaningful) have prevented formation of a consensus about the practical use of these agents.[21] A large, randomized, placebo-controlled trial of recombinant human nerve growth factor given over 12 months failed to show improvement in various clinical and laboratory measures of function, despite promising results of phase II trials.[22]

Patients with diabetic polyneuropathy should be counseled about good foot care, including daily careful inspection of the plantar surfaces, so that minor injuries can be detected early and so that major complications such as plantar ulcers or osteomyelitis can be prevented. Measures for the management of neuropathic pain are often needed.

Other Diabetic Neuropathies

The other varieties of diabetic neuropathy usually occur on a background of diabetic polyneuropathy. Two mononeuropathies (diabetic ophthalmoplegia and proximal diabetic neuropathy) may occur in diabetic patients, as well [*see* Oculomotor Neuropathy *and* Lumbosacral Plexopathy, *above*].

Diabetic autonomic neuropathy Some degree of diabetic autonomic neuropathy is found in most patients with diabetic polyneuropathy, although in some patients, the autonomic symptoms and signs predominate. Orthostatic hypotension, impaired gastrointestinal motility (including gastroparesis), and blunting of the sympathetically mediated warning symptoms of hypoglycemia are important management problems. Some au-

tonomic symptoms can be treated effectively; for example, gastroparesis can be treated with metoclopramide, and orthostatism can be treated with support hose, volume expansion, and vasopressors such as midodrine.

Diabetic truncal neuropathy Diabetic truncal neuropathy is a painful sensory syndrome affecting one or several thoracic dermatomes, usually unilaterally. Clinically, the process is a thoracic radiculopathy with abrupt development of tactile hypersensitivity or sensory loss in a dermatomal distribution. Motor involvement can usually be demonstrated by electromyography, and sometimes, an asymmetrical bulging of the abdomen occurs because of unilateral weakness of the abdominal muscles. The main alternative diagnosis is herpes zoster radiculopathy, usually distinguished by vesicular skin eruption. Diabetic truncal radiculopathy is self-limited, although aggravating symptoms may persist for weeks. Administration of analgesics, supplemented by amitriptyline, gabapentin, or other drugs for neuropathic pain, is the main management approach.

Other Metabolic Disorders

Polyneuropathy develops in up to 60% of patients with chronic renal failure. The risk of uremic neuropathy is related to the duration and severity of renal failure. Slowing of motor conduction velocity is seen when the creatinine clearance falls below 10% of normal, although conduction slowing is only roughly correlated with symptoms. Uremic neuropathy is sensorimotor in nature, with distal predominance of symptoms and signs. Once established, uremic neuropathy tends to worsen slowly, the main troublesome symptom being unpleasant dysesthesia in the feet. Disabling motor weakness is rare. The chief pathologic finding in uremic neuropathy is axonal degeneration, which is most abundant in the most distal parts of the PNS. Etiologically, impairment of neuronal soma or axon function probably results from the buildup of a neurotoxic substance or substances normally excreted by the kidneys, but the details are not known. Both dialysis and renal transplantation generally have a beneficial effect. Dialysis can prevent, stabilize, or often decrease the severity of neuropathy. The effects of renal transplantation may be striking, and even severe uremic neuropathy can be expected to improve in the months after transplantation. The widespread availability of these treatments has made symptomatic uremic neuropathy relatively uncommon.

Mild distal axonal polyneuropathy sometimes occurs in patients with hypothyroidism, acromegaly, or polycythemia. These neuropathies are not disabling, but sensory symptoms may be troublesome. Patients in intensive care units with multiple organ dysfunction syndrome and sepsis sometimes experience a severe axonal polyneuropathy. The pathogenesis of this condition, known as critical illness polyneuropathy, is unknown.[23] Critical illness polyneuropathy is usually suspected either when a patient in an intensive care unit cannot be weaned from the ventilator despite adequate cardiopulmonary function or when limb weakness occurs in an alert ICU patient. Because it is difficult to examine the PNS properly in critically ill patients and because critical illness polyneuropathy is a diagnosis of exclusion, clearly establishing the diagnosis can be difficult. A similar clinical picture can also be caused by a critical illness myopathy, particularly in patients treated with high-dose glucocorticoids and long-acting neuromuscular blockers (e.g., vecuronium). Limited data suggest that critical illness neuropathy improves in patients who survive the illness.[24]

Inherited Polyneuropathies

Inherited polyneuropathies are common and frequently overlooked, especially in adults.[12] Patients often remain relatively asymptomatic for many years and, therefore, may not seek medical advice until middle age or later. A very slow progression of symptoms over years is suggestive of an inherited cause. Patients with inherited polyneuropathies may complain of numbness, by which they usually mean decreased sensation. Symptoms of prickling or pins-and-needles paresthesia in a patient with polyneuropathy favors an acquired cause. However, the complaint of burning feet is nonspecific, occurring in both inherited and acquired polyneuropathies. As a rule, inherited neuropathies begin and progress symmetrically; asymmetrical progression suggests an acquired polyneuropathy.

Once inherited polyneuropathy is suspected, proving the diagnosis may require considerable effort. Simply asking the patient about a family history of neuropathy is seldom productive. Questioning should be more general, such as whether anyone in the family has foot complaints, has a foot deformity (especially high arches), or requires special shoes, canes, or braces. A negative family history does not exclude a diagnosis of inherited neuropathy. Because of its chronicity or mildness, or both, the condition often goes unrecognized in a family. In recessive cases, by definition, neither parent is affected, and siblings may be spared. Paternity may be less well established than it appears. The most effective next step is to examine first-degree relatives, particularly those with any history of foot or walking troubles. Although this procedure is time-consuming, it has a good yield and is more cost-effective than a barrage of laboratory tests that may all prove negative.

Knowledge about the specific genetic defects related to various inherited neuropathies is expanding rapidly,[25] but clinical classification remains important. Traditionally, the inherited neuropathies are divided into three main categories according to clinical features: (1) hereditary motor and sensory neuropathy, (2) hereditary sensory and autonomic neuropathy, and (3) hereditary motor neuropathy (also called spinal muscular atrophy) [*see Table 4*]. Genetic testing for some inherited neuropathies is now available (e.g., *PMP22* gene [Charcot-Marie-Tooth disease type 1], *connexin-32* gene [X-linked Charcot-Marie-Tooth disease]). Although there are still no specific therapies for inherited neuropathies, correct diagnosis remains important for prognosis, education, and genetic counseling. In general, the deficits in inherited neuropathies progress very slowly, life expectancy is normal, and most patients remain ambulatory all their lives.

Immune-Inflammatory Polyneuropathies

Guillain-Barré syndrome Guillain-Barré syndrome (GBS), or acute inflammatory demyelinating polyradiculoneuropathy, is the most common cause of acute generalized paralysis in the Western world. Most large general hospitals care for several patients with GBS every year, and the management of these patients often falls to general physicians. The natural history of GBS is favorable in most cases, but a good outcome depends on meticulous medical and nursing care during the peak of neurologic disability.

The diagnosis of GBS is not difficult once the signs are full-blown. However, in the first few days, symptoms can be vague and the signs equivocal. It is common for a patient to be discharged from the emergency room with a diagnosis of anxiety, only to return a day or two later with obvious progressive limb weakness. Most often, the first symptom is prickling paresthe-

sia, beginning in the feet and spreading proximally hour by hour. Weakness is noticed some hours to a few days later. Some patients have only motor symptoms without sensory symptoms. Classically, symptoms begin symmetrically in the distal limbs and proceed proximally (so-called ascending paralysis), and nerve conduction studies provide evidence of a demyelinating process affecting spinal roots and peripheral nerves (a demyelinating polyradiculoneuropathy). Variations on this classic presentation are common and include syndromes in which cranial nerve (e.g., Miller Fisher syndrome) or autonomic involvement predominate, as well as syndromes that are primarily axonal rather than demyelinating.[26] It is evident that Guillain-Barré syndrome must be regarded as a rather elastic category, although all clinical variants share the key features of being acute-onset, immune-mediated disorders of the PNS, with a self-limited monophasic course. In addition to motor, sensory, or autonomic symptoms, many patients have pain, most often manifested as a deep ache in the back or limbs or as dysesthetic limb pain.[27]

Autonomic dysfunction may be a significant problem in some patients with GBS, particularly those with severe weakness. Marked blood pressure fluctuation and refractory hypotension are the major problems, although hyperthermia, pupillary paralysis, and cardiac dysrhythmias may occur.

Muscle weakness can be expected to increase during the first few days and then remain stable for days to weeks. Weakness may range from mild (e.g., slight weakness of ankle dorsiflexion) to severe; flaccid quadriparesis with respiratory muscle paralysis occurs in as many as 30% of patients. It is impossible to predict within the first 2 days how much weakness will develop in a given patient. Therefore, it is advisable to observe any patient with GBS in the hospital until the severity is apparent. As a general rule, weakness reaches its maximum within 14 days. The ensuing period of stable weakness, before recovery begins, lasts days to months, with a median duration of 4 weeks. A rough correlation exists between the severity of weakness and the interval before the onset of recovery. Once recovery begins, the patient usually makes noticeable gains on a weekly basis. A year after onset, most patients have made a complete or substantial recovery, but as many as 15% remain bedridden or wheelchair bound. Recurrent episodes of GBS, sometimes spread out over many years, occur in about 3% of patients.

The fundamental pathologic event in GBS is the stripping of myelin from axons by macrophages, which occurs in a patchy fashion throughout the PNS. A cascade of events involving cell-mediated and humoral immune mechanisms is assumed to be activated, and lymphocytic inflammatory infiltrates are often found in nerve and nerve roots examined by biopsy or at autopsy. The triggers and specific molecular targets of the immune attack are unknown. There is a lingering concern about influenza vaccination as a precipitant of GBS; retrospective epidemiologic studies have found a small relative risk of GBS among vaccine recipients in the 1992 through 1994 influenza seasons.[28] Studies of the pathogenesis of GBS have focused on the potential roles of antecedent *Campylobacter jejuni* infection and the production of antiganglioside autoantibodies, both of which occur in a large number of patients with GBS.[29]

The differential diagnosis of an acute sensorimotor polyneuropathy is short [*see Table 2*]. Before the development of the polio vaccine, the major differential diagnosis was between GBS and poliomyelitis, but poliomyelitis is now rare. If signs of systemic illness accompany an acute polyneuropathy, consideration should be given to vasculitic neuropathy, lymphomatous infiltration of the nerve roots, acute intermittent porphyria, diphtheria, and arsenic poisoning. An ischemic lesion of the pons may initially produce acute flaccid quadriparesis and mimic GBS before clear upper motor neuron signs develop. With cranial nerve variants of GBS (e.g., Miller Fisher syndrome), myasthenia gravis and botulism must be excluded.

GBS is diagnosed primarily from the clinical features, but laboratory tests can help support the diagnosis and exclude other causes of acute polyneuropathy. The cerebrospinal fluid in GBS characteristically shows elevated protein levels, normal glucose levels, and no pleocytosis. If a mononuclear pleocytosis is found, meningeal infiltration by lymphoma or carcinoma must be seriously considered. An acute polyradiculopathy similar to GBS but with CSF pleocytosis may occur at the time of HIV seroconversion [*see* Polyneuropathies Caused by Infectious Diseases, *below*]. Nerve conduction studies in GBS usually demonstrate widespread slowing of conduction velocities or, more commonly, proximal conduction block and can provide some prognostic information. Neither the CSF nor the nerve conduction abnormalities are diagnostic of GBS, and both tests may be normal, particularly during the first few days of the disease.

Once a diagnosis of GBS is made, the first priority is to monitor respiratory muscle function closely and be prepared to intervene if ventilatory failure develops. Frequent bedside measurements of vital capacity or maximum inspiratory force, or both, and clinical evaluation for signs of respiratory muscle fatigue are important. A patient with a vital capacity of less than 20 ml/kg is likely to experience frank ventilatory failure, and elective intubation should be strongly considered.

Several large clinical trials have shown that two therapies—plasma exchange and intravenous immunoglobulin (IVIg)—improve the rate of recovery in patients with moderate to severe GBS (e.g., patients who are unable to walk).[30-32] Both treatments appear to modulate the inflammatory process that produces nerve injury, but they are not thought to promote remyelination or nerve regeneration. A beneficial effect can be shown only if plasma exchange or IVIg is begun within 2 weeks after onset of neuropathic symptoms. It is probably wise to begin treatment as soon as the diagnosis has been made, if the weakness is sufficiently severe, or if the signs progress while the patient is under observation. Usually, five treatments are given over 5 to 10 days; some patients may relapse and require further treatments (e.g., twice weekly for 3 weeks). Combined treatment (plasma exchange followed by IVIg) offers no clear advantage over either treatment alone,[30] and glucocorticoids conferred no additional benefit when added to plasma exchange in a large controlled trial[33] and have no benefit when used alone.[34] Plasma exchange and IVIg appear to be equally effective and have similar costs and relapse rates, but many clinicians believe IVIg is the treatment of choice because it is considerably easier to administer than plasma exchange. Because IVIg is made from pooled blood, it is difficult to completely eliminate concerns about infectious risk, even though multiple purification procedures are now used, and there have been no proven cases of IVIg-transmitted infection for over a decade. There is also uncertainty about whether the supply of IVIg is adequate to meet the demand. A new therapeutic strategy is CSF filtration; a small controlled trial found this modality to be at least as effective as plasma exchange, with fewer complications.[35]

Because many patients remain bedridden for months, meticulous nursing care is essential. Measures to prevent pressure

sores and venous thrombosis are necessary, and the patient is also in constant danger of urinary and pulmonary infections. Pain is a major problem for some patients and seems to be related partly to immobility and partly to nerve inflammation and dysfunction. Regular range-of-motion physiotherapy exercises and judicious use of analgesics and agents such as amitriptyline are beneficial. A particular challenge for all staff members is psychological support of the patient; there are several eloquent accounts of the patient's perspective in GBS.[36]

Chronic inflammatory demyelinating polyradiculoneuropathy Like GBS, chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) is an immune-mediated neuropathy. The fundamental pathologic event is the stripping of myelin from axons by macrophages, which slows or blocks nerve impulse conduction, causing weakness and sensory loss. Clinically, CIDP differs from GBS in several important respects. The onset of CIDP is insidious, with symptoms and signs developing over weeks to months, compared with the rapid initial course of GBS. Unlike the monophasic, self-limited course of GBS, the natural history of CIDP is variable. The usual pattern of CIDP is slow worsening over months, producing chronic moderate disability. Some patients have a relapsing and remitting course, and a few experience gradual spontaneous remission. Major autonomic or respiratory involvement, which is common in GBS, is unusual in CIDP.

No single clinical or laboratory finding is pathognomonic for CIDP, and the diagnosis is made from a combination of an appropriate history, signs of neuropathy, electrophysiologic evidence of demyelination, and an acellular CSF with an increased protein level. Nerve biopsy can support the diagnosis but is usually unnecessary. A similar type of chronic neuropathy can occur in patients with monoclonal gammopathy, osteosclerotic myeloma, and HIV infection.

The efficacy of a number of treatments for CIDP has been demonstrated in controlled trials. Plasma exchange[37] and IVIg[38] produce improvement in 2 to 3 weeks in most patients, but the improvement is short-lived, and patients need either repeated treatments or alternative long-term therapies. High-dose prednisone is also effective,[39] but the response is usually less impressive and slower, and the patient faces the potential complications associated with long-term glucocorticoid use. A double-blind crossover trial of interferon beta in treatment-resistant CIDP showed no benefit,[40] but there is anecdotal support for other immunomodulating drugs, including cyclosporine, azathioprine, interferon alfa, cyclophosphamide, and mycophenolate mofetil. Careful monitoring of neurologic impairment and treatment by neurologists experienced with CIDP are important.

Monoclonal protein–associated polyneuropathy Polyneuropathy and a monoclonal gammopathy may occur in some patients who have amyloidosis, multiple myeloma, osteosclerotic myeloma, Waldenström macroglobulinemia, or lymphoma. If none of these hematologic conditions is found, then the patient is considered to have monoclonal gammopathy of undetermined significance (MGUS). The relation between MGUS and neuropathy is unsettled, but epidemiologic and pathologic evidence suggests that MGUS causes the neuropathy.

MGUS neuropathy has variable clinical and electrophysiologic features. There is often electrophysiologic and pathologic evidence of prominent demyelination, and the clinical picture closely resembles that of CIDP. Some patients have a painful, axonal neuropathy with mild deficits. IgM-associated neuropathy tends to be more severe, with more ataxia and more evidence of nerve conduction slowing and dispersion, which indicates that some relation exists between the class of immunoglobulins and the features of the neuropathy.[41]

MGUS neuropathy is generally slowly progressive, though the degree of disability is variable. Therapeutic approaches aim either to suppress production of the monoclonal protein or to remove it from the circulation. Anecdotal reports exist of responses to chlorambucil, melphalan, and prednisone. A controlled trial showed plasma exchange to be beneficial in patients with IgG and IgA gammopathies but not in those with IgM gammopathy.[42] IVIg appears to be beneficial in IgG MGUS neuropathy, but a controlled trial has not been conducted.[43]

Amyloid neuropathy Polyneuropathy characterized by amyloid deposition in nerve occurs in two settings: (1) as a feature in about 15% of patients with systemic amyloidosis, in whom the amyloidogenic protein is immunoglobulin, and (2) as an autosomal dominant disorder, familial amyloid polyneuropathy (FAP), in which the amyloid most commonly consists of mutated forms of transthyretin, a normal serum protein.[44] Both types feature a symmetrical sensorimotor polyneuropathy distinguished by the frequent presence of important or predominant autonomic features. Amyloidosis is also a rare cause of carpal tunnel syndrome.

The diagnosis of amyloid neuropathy is made by nerve biopsy. Acquired amyloid neuropathy and inherited amyloid neuropathy are histopathologically similar but can be distinguished by immunohistochemical analysis. Trials of melphalan, glucocorticoids, and colchicine have had little impact on the outcome of patients with primary amyloidosis, including any improvement of the neuropathy. Unlike in primary amyloidosis, neuropathy is usually the major cause of disability in FAP, and most patients survive for a decade or more after diagnosis. Liver transplantation may halt progression of the neuropathy in FAP.[45]

Osteosclerotic myeloma Osteosclerotic myeloma is a variant of myeloma in which patients have one or sometimes several osteosclerotic bony lesions. Biopsy reveals malignant plasma cell proliferation. Commonly, these bony lesions occur within a constellation that includes polyneuropathy, organomegaly, endocrine abnormalities, monoclonal gammopathy, and skin changes (POEMS syndrome). The polyneuropathy is typically demyelinating and resembles CIDP. POEMS syndrome is important to recognize because there may be significant clinical improvement with irradiation of the bony lesion.

Paraneoplastic neuropathy Several distinctive nonmetastatic neurologic syndromes that develop in cancer patients have been described since the 1950s. Paraneoplastic syndromes are believed to have an immunologic basis, developing as a consequence of the host's attempt to mount an immune response to the cancer [*see 11:VI Neoplastic Disorders*].

Vasculitic neuropathy Neuropathy is a frequent manifestation of certain systemic vasculitides. Vasculitic neuropathy is ischemic, a consequence of involvement of nutrient vessels of nerve by the inflammatory process. Because of the robust blood supply of nerve and its relative resistance to ischemic injury, the development of neuropathy in vasculitis implies extensive vessel involvement. The vasculitis tends to be patchy, and asymme-

try of nerve involvement is common, so individual nerves are often affected while neighboring nerves are spared; this is the classic syndrome of multiple mononeuropathies. However, as the involvement of nerve blood supply advances and more nerves become involved, a pattern of multiple mononeuropathies may become more difficult to identify. About 30% of cases of vasculitic neuropathy are symmetrical polyneuropathies at initial diagnosis, 30% are asymmetrical polyneuropathies, and 40% are multiple mononeuropathies.[46]

Among patients with vasculitic neuropathy, the main associated systemic vasculitides are polyarteritis nodosa, rheumatoid vasculitis, Sjögren syndrome, Wegener granulomatosis, and allergic granulomatous angiitis (Churg-Strauss syndrome). Neuropathy is particularly common in polyarteritis nodosa, occurring in at least half of all cases. The clinical and neuropathologic features of the neuropathy in these disorders are similar, and a specific diagnosis depends on the systemic, nonneurologic features. Vasculitic neuropathy can often be suspected from the clinical setting alone, but definitive diagnosis depends on nerve biopsy. Antineutrophil cytoplasmic antibodies are often found in vasculitic neuropathy, but false positive test results limit their diagnostic usefulness.[47]

Vasculitic neuropathy is managed by treating the underlying systemic vasculitis. This therapy generally requires high-dose prednisone, cyclophosphamide, or both. In vasculitis, neuropathy is seldom a cause of death, although it may produce significant disability. If life-threatening events, such as renal or cardiorespiratory failure, can be successfully treated, the prospects for neurologic improvement are good, although improvement may take many months.[48]

Patients with clinical and neuropathologic features of vasculitic neuropathy sometimes have no evidence of systemic vasculitis. This syndrome, termed nonsystemic vasculitic neuropathy, appears to have a more benign natural history than does systemic vasculitis. Corticosteroids are often used if the patient has significant or progressive neurologic disability, although there are no controlled trials. The largest retrospective series of patients with nonsystemic vasculitic neuropathy suggested that combination immunotherapy (corticosteroids plus cyclophosphamide) was more effective than corticosteroids alone in inducing remission and improving disability.[49]

Neuropathies related to other connective tissue diseases Peripheral neuropathy may occur in other connective tissue diseases, such as systemic lupus erythematosus [*see 15:IV Systemic Lupus Erythematosus*], although it can be difficult to ascertain whether the neuropathy is a direct complication of the connective tissue disease or a secondary effect of another complication (e.g., secondary to renal failure). It is usually assumed that neuropathy in such cases has an inflammatory basis, but few pathologic studies exist to support this belief. An ataxic sensory neuronopathy, clinically very similar to the paraneoplastic syndrome [*see 11:VI Neoplastic Disorders*], has been described in patients with Sjögren syndrome and is caused by an inflammatory infiltration of dorsal root ganglia.[50] Inflammatory sensory neuropathies also occur in patients with sicca syndrome who lack extraglandular features of Sjögren syndrome.[51]

Neuropathies Caused by Toxins and Nutritional Deficiencies

Substances toxic to peripheral nerves include a variety of industrial chemicals, naturally occurring compounds, and drugs. Most toxic neuropathies begin distally, progress insidiously over weeks to months, and have electrophysiologic features of an axonal neuropathy. With some exceptions, nonspecific axonal degeneration is the main histologic feature, and nerve biopsy is seldom helpful in making the diagnosis.

Drugs Drug-induced neuropathy is a common problem, particularly in a hospital-based practice. A careful drug history is an important part of the investigation of any polyneuropathy. Peripheral neuropathy is a well-established adverse effect of many drugs [*see Table 5*]. In ideal circumstances, before a drug is labeled as potentially neurotoxic, the affected patient should be found to be free of other potential causes of neuropathy, and cessation of the drug should lead to some clinical improvement. It is important to realize that recovery may take many months. In some toxic neuropathies, the patient experiences the phenomenon known as coasting, in which the neuropathy continues to worsen for some weeks after exposure has ceased. For most well-established peripheral nerve toxins, there is experimental evidence of toxicity in animals and neuronal cell culture. With some antineoplastics (e.g., cisplatin), neuropathy may be the dose-limiting side effect. In animal models, cisplatin neuropathy may be prevented or alleviated by coadministration of neurotrophins, such as NT-3, a strategy that may prove applicable to humans.[52]

Industrial chemicals A patient's occupational history may be important because a number of industrial chemicals are known to be neurotoxic [*see Table 6*]. Most of these compounds were found to be neurotoxic after clusters of cases appeared in workers in specific industries, and subsequent awareness of the dangers of these compounds has reduced the incidence of cases. Most of these chemicals produce axonal neuropathies with nonspecific pathologic changes. Proving that a neuropathy is caused by exposure to a chemical requires careful epidemiologic work supported by animal or tissue culture studies.

Metals In addition to drugs containing gold and platinum, poisoning by other metals may produce neuropathy. Exposure is often the result of homicidal or suicidal intent, so the history may be of dubious value. Diagnosis must be made from the associated clinical features. For example, lead neuropathy is predominantly motor, with a predilection for the upper limbs, and is associated with abdominal pain, constipation, and anemia. In addition to causing neuropathy, arsenic poisoning produces abdominal pain, vomiting, diarrhea, skin and nail changes, and pancytopenia. Thallium neuropathy is distinguished by alopecia and abdominal pain. Organic and inorganic mercury compounds can produce neuropathy, but CNS effects usually predominate [*see 8:I Management of Poisoning and Drug Overdose*].

Table 6 Industrial Chemicals That May Cause Polyneuropathy

Acrylamide	Methyl bromide
Allyl chloride	Methyl butyl ketone*
Carbon disulfide	Organophosphorus esters
Dimethylaminopropionitrile	Polychlorinated biphenyls
Ethylene oxide	Trichloroethylene
Hexane*	Vacor

*Exposure to these compounds also occurs as a result of glue sniffing and other solvent abuse.

The diagnosis of metal neuropathy hinges on demonstrating increased urinary excretion of the metal or increased levels in the hair or nails. Requests for such assays should be prompted by clinical suspicion and are not part of the routine investigation of polyneuropathy.

Ethanol A distal, often painful neuropathy is common in chronic alcoholics. The main symptoms are burning, stabbing pains and numbness in the feet and sometimes in the hands. Sensory loss or painful hypersensitivity in the feet, loss of the ankle reflexes, and mild distal weakness form the typical picture. Whether the neuropathy is caused by a direct toxic effect of ethanol, malnutrition, or both remains unresolved.

Attempts to produce neuropathy with ethanol in well-fed animals have been unsuccessful, although in neuronal cell cultures, growth inhibition can be produced with moderately high concentrations of ethanol. A Danish study that carefully examined a cohort of alcoholic beer drinkers with neuropathy found no clinical, electrophysiologic, or histologic differences between those who were well nourished and those who were malnourished.[53] Danish beer is supplemented with thiamine and pyridoxine, so it is unlikely that deficiency of these vitamins was the cause of neuropathy in these patients.

Although the pathogenesis of neuropathy in alcoholics is debated, the clinical setting is characteristic in that other signs and symptoms of alcoholism are generally present, including chronic liver disease, memory impairment, and ataxia of gait. Estimates of ethanol intake are notoriously unreliable, but in the Danish study, neuropathy developed only in patients who consumed 3 L of beer or 300 ml of spirits daily for at least 3 years. An important corollary of these observations is that in a patient with neuropathy who drinks moderately, is well nourished, and is free of signs of chronic liver disease, alcoholic neuropathy is unlikely; therefore, in such a patient, other causes of neuropathy should be sought.

Treatment of alcoholic neuropathy is simple in principle but difficult in practice. Ethanol ingestion must be stopped and adequate nutrition ensured. If such measures can be achieved, gradual improvement can be expected, although it may take months and be incomplete.

Nutritional deficiencies Polyneuropathy can be a manifestation of starvation, as seen in famine victims or in prisoners of war. The precise dietary components responsible for neuropathy in starvation are unclear, although one or more B vitamins are assumed to be critical. Certain vitamin deficiencies can cause neuropathy in specific situations, but there is no physiologic basis for routinely prescribing multivitamins for neuropathy in patients with normal diets.

Thiamine (vitamin B_1) deficiency produces beriberi, the main features of which are polyneuropathy and cardiac failure. The neuropathy is distal and axonal, with painful sensory symptoms. With progression, distal weakness may develop. Cranial nerve involvement was described in older accounts of beriberi.

Pyridoxine (vitamin B_6) deficiency is responsible for the neuropathy caused by isoniazid, which increases excretion of pyridoxine. Administration of pyridoxine with isoniazid prevents the neuropathy. However, excessive doses of pyridoxine (a fad in the 1970s) produce a severe sensory neuronopathy.

A mild polyneuropathy may be part of the neurologic syndrome produced by cobalamin (vitamin B_{12}) deficiency. However, vitamin B_{12} deficiency probably does not present as polyneuropathy alone; the main clinical features are the result of myelopathy (subacute combined degeneration).

Vitamin E deficiency from malabsorption may result in an ataxic syndrome caused by degeneration of the peripheral and central processes of dorsal root ganglia neurons. Cerebellar involvement may be present in some patients.

Polyneuropathies Caused by Infectious Diseases

Leprosy Leprosy, a mycobacterial infectious disease of peripheral nerves, is probably the most common cause of polyneuropathy in the world [*see 7:II Infections Due to Mycobacteria*]. Most cases occur in tropical and subtropical regions, although there are endemic foci along the Gulf Coast of Florida and Louisiana. Most patients with leprosy in North America and Europe are immigrants from countries where leprosy is common.

Sensory loss is the cardinal symptom of leprosy. It is often discovered because of a painless injury. Because of the temperature requirements of *Mycobacterium leprae*, cutaneous sensory and mixed nerves in parts of the body with low ambient temperature are most likely to be affected. The result is a distribution of signs unlike those of any other polyneuropathy, with sensory loss over the external ears, the zygomatic arches, and extensor surfaces of joints. Major nerves are most likely to be affected where they travel close to the surface (e.g., the ulnar nerve at the elbow). Involvement of cutaneous nerves is generally sharply demarcated, especially in the tuberculoid form of leprosy, and the overlying dermis and epidermis are affected, producing the classic anesthetic macule. Motor weakness does not occur until sensory loss is well established. Weakness is usually patchy and asymmetrical and may suggest other causes of multiple mononeuropathies. The diagnosis of leprosy is made by nerve or skin biopsy, using the Fite method to stain and identify *M. leprae*.

Leprosy is a curable disease. The degree of recovery in advanced cases may be limited, so it is important that the disease be diagnosed and treated before major neuropathic deficits develop [*see 7:II Infections Due to Mycobacteria*].

HIV infection Several PNS disorders occur in patients with HIV infection, some in the early stages of infection and some only after progression to AIDS [*see 7:XXXIII HIV and AIDS*]. Distal, painful neuropathy is very common in patients with AIDS. The main symptom is continuous burning discomfort, mostly in the feet, where some degree of sensory loss is apparent. Motor involvement is usually minor, although the patients are often debilitated by concomitant infections and weight loss. The cause is sometimes identifiable—for example, vitamin B_{12} deficiency or treatment with a retroviral agent known to be neurotoxic (e.g., 2,3-dideoxycytidine or zalcitabine)[54]—but the cause in some patients remains unclear. The HIV genome has been detected in dorsal root ganglia neurons and satellite cells in patients with AIDS and neuropathy,[55] and expression of the HIV genome in transgenic mice produces peripheral nerve disease.[56] Whether neuropathy in humans with HIV is actually caused by direct infection of nerves by HIV remains to be proved.

Neuropathy is a frequent dose-limiting side effect of nucleoside analogue antiretrovirals (e.g., zalcitabine), but fortunately, neither zidovudine (AZT) nor the protease inhibitors appear to cause neuropathy. In a phase II trial, nerve growth factor, compared with placebo, reduced neuropathic pain in HIV-infected patients who had distal sensory neuropathy, although injection-site pain produced impaired masking in a substantial portion of treated patients.[57]

GBS and CIDP typically occur early in the course of HIV infection. They may be the presenting features of HIV infection, and HIV testing should be considered in patients with GBS or CIDP. The neurologic picture and response to treatment in patients with HIV are similar to those in non–HIV-infected patients, except that there is usually a lymphocytic pleocytosis in the CSF.

A syndrome of multiple mononeuropathies may occur in HIV-positive patients. Nerve biopsy may show perivascular inflammatory infiltrates, necrotizing vasculitis, or cytomegalovirus (CMV) inclusions. Patients with evidence of CMV infection may respond to ganciclovir[58] [*see 7:XXVI Herpesvirus Infections*]. Patients with HIV infection that is complicated by the diffuse infiltrative lymphocytosis syndrome may experience a subacute axonal neuropathy; nerve biopsies in these patients have shown striking perivascular $CD8^+$ lymphocyte infiltrates, and the neuropathy seems to abate with corticosteroids or zidovudine.[59]

Acute lumbosacral polyradiculopathy is a devastating syndrome of leg weakness, leg and perineal sensory loss, and urinary retention that develops over 1 to 2 weeks, usually in patients with advanced AIDS. The CSF shows a distinctive polymorphonuclear pleocytosis. In most cases, there is evidence of concomitant CMV infection, and the neurologic picture is caused by invasion of lumbosacral nerve roots by CMV. Prompt diagnosis and treatment with ganciclovir may produce improvement [*see 7:XXVI Herpesvirus Infections*]. A similar picture may be produced by lymphoma.[60]

Lyme disease Lyme disease is caused by the spirochete *Borrelia burgdorferi*, which is transmitted to humans by ixodid ticks. The disease occurs worldwide but is most common in the northeastern United States and northern Europe [*see 7:VII Leptospirosis, Relapsing Fever, Rat-Bite Fever, and Lyme Disease*]. Peripheral neuropathy may occur in early or late disseminated Lyme disease. It is unclear whether the neurologic manifestations are caused directly by spirochetal invasion of nerve tissue or by the host's immune response to the organism.

The main early neurologic features are cranial neuropathies, spinal radiculopathies, or both.[61] Headache and neck stiffness may accompany the peripheral nerve symptoms, reflecting meningeal inflammation. Facial neuropathy is the most common cranial neuropathy, and a misdiagnosis of Bell palsy is often made. However, unlike in idiopathic Bell palsy, bilateral facial neuropathy is common in Lyme disease. Other cranial neuropathies occur less often.

Spinal nerve root involvement typically begins with pain in a radicular distribution, followed by weakness in 1 to 4 weeks. The weakness is often asymmetrical and patchy, resembling multiple mononeuropathies. In addition to the predominant radicular signs, there is frequently electrophysiologic evidence of a mild widespread polyneuropathy. Both cranial and spinal neuropathies in early-stage Lyme disease have a favorable natural history, with recovery in weeks to months. Treatment with antibiotics may hasten recovery.

In late-stage Lyme disease, patients may experience either a mild distal polyneuropathy or radicular pain with sensory signs. Many of these patients will have had previous symptoms of localized or early disseminated disease. Unlike cranial neuropathy and radiculoneuropathy of early-stage disease, the peripheral neuropathy of late-stage disease tends not to resolve without treatment.

The diagnosis of Lyme neuropathy rests on a history of possible tick exposure, compatible symptoms and signs, and positive serologic testing for *B. burgdorferi*. In early-stage disease, neurologic involvement is almost always accompanied by a lymphocytic CSF pleocytosis, but in late-stage disease, pleocytosis is exceptional and the CSF may be normal. Sural nerve biopsy in early- or late-stage disease shows axonal degeneration and perivascular inflammatory infiltrates, but these findings are nonspecific. Particularly in late-stage disease, response to a course of antibiotics may provide the most convincing evidence of the diagnosis [*see 7:VII Leptospirosis, Relapsing Fever, Rat-Bite Fever, and Lyme Disease*].

Varicella-zoster virus infection Varicella-zoster virus (VZV), a member of the human herpesvirus family, is the most common viral pathogen in the PNS [*see 7:XXVI Herpesvirus Infections*]. After initial VZV infection, which usually occurs during childhood in the form of varicella (chickenpox), some VZV virions may enter cutaneous sensory axons and are carried by retrograde axonal transport to sensory neuron cell bodies in the dorsal root or cranial ganglia, where they remain in an inactive form.

Years later, the virus can return to an active, proliferative state because of alterations in immune function that are incompletely understood. The cytopathic effects of VZV replication and the ensuing immune response produce an intense inflammatory necrotic ganglionitis, resulting in dermatomal sensory alteration, pain, and the dermatomal vesicular skin rash known as herpes zoster.

Herpes zoster is most common in persons older than 60 years, although it also occurs in younger persons. The incidence is much higher in immunocompromised patients, in whom there is also a risk of disseminated VZV infection, which has considerable mortality. Herpes zoster is usually unilateral, involving one to three adjacent dermatomes. Thoracic and trigeminal (especially ophthalmic) dermatomes are most often affected.

The initial symptom is dermatomal pain, followed in 3 to 7 days by the vesicular eruption. Sensory loss is difficult to demonstrate until the skin lesions begin to heal. Motor weakness is reported in as many as 30% of patients but may not be noticed by a patient distracted by pain. The cutaneous lesions persist for 7 to 10 days and then resolve, often leaving depigmented areas and scarring. Motor weakness can be expected to improve spontaneously in most patients. Pain also improves slowly, except in patients who experience postherpetic neuralgia (PHN).

PHN is usually defined as pain persisting more than 4 to 8 weeks after healing of the skin lesions. The risk of PHN after an episode of herpes zoster increases with age, reaching as high as 45% in patients older than 65 years. Patients with PHN experience a continuous burning dermatomal pain, on which brief, lancinating pains may be superimposed. Often, there is also allodynia in the affected dermatome, so that the touch of clothes or even hair may cause excruciating pain. Some patients with PHN experience gradual improvement over weeks to months, but persistence of symptoms for years is not unusual. Prevention of PHN is probably the main rationale for using antiviral agents in an immunocompetent patient. A large placebo-controlled trial of famciclovir showed a clear reduction in the incidence of PHN, especially in the elderly.[62] It is assumed that acyclovir offers a similar benefit, but this assumption is unproved. In one study, the addition of prednisolone to 7- or 21-day courses of acyclovir did not alter the frequency of PHN,[63] although glucocorticoids may reduce acute-phase pain.

PHN presents a formidable therapeutic challenge. Amitriptyline or other tricyclic medications have been the usual first-line agents, but gabapentin is being increasingly used and was shown to be superior to placebo in a controlled trial.[64] Lancinat-

ing pain may be reduced by carbamazepine or other anticonvulsants. Topical application of local anesthetics and long-term use of opioids are advocated by some clinicians.[65]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Jacobs WB, Fehlings MG: The molecular basis of neural regeneration. Neurosurgery 53:943, 2003

2. Mendell JR, Sahenk Z: Painful sensory neuropathy. N Engl J Med 348:1243, 2003

3. Foley KM: Opioids and chronic neuropathic pain. N Engl J Med 348:1279, 2003

4. Murakami S, Mizobuchi M, Nakashiro Y, et al: Bell palsy and herpes simplex virus: identification of viral DNA in endoneurial fluid and muscle. Ann Intern Med 124:27, 1996

5. Salinas RA, Alvarez G, Alvarez MI, et al: Corticosteroids for Bell's palsy (idiopathic facial paralysis). Cochrane Database Syst Rev (1):CD001942, 2002

6. Adour KK, Ruboyiane J, Van Doersten PG, et al: Bell's palsy treatment with acyclovir and prednisone compared with prednisone alone: a double-blind, randomized, controlled trial. Ann Otol Rhinol Laryngol 105:371, 1996

7. Chance PF: Overview of hereditary neuropathy with liability to pressure palsies. Ann NY Acad Sci 883:14, 1999

8. Scholten RJ, Gerritsen AA, Uitdehaag BM, et al: Surgical treatment options for carpal tunnel syndrome. Cochrane Database Syst Rev (4):CD003905, 2002

9. Kim DH, Han K, Tiel RL, et al: Surgical outcomes of 654 ulnar nerve lesions. J Neurosurgery 98:993, 2003

10. Suarez GA, Giannini C, Bosch EP, et al: Immune brachial plexus neuropathy: suggestive evidence for an inflammatory-immune pathogenesis. Neurology 46:559, 1996

11. Dyck PJB, Windebank AJ: Diabetic and nondiabetic lumbosacral radiculoplexus neuropathies: new insights into pathophysiology and treatment. Muscle Nerve 25:477, 2002

12. Dyck PJ, Oviatt KF, Lambert EH: Intensive evaluation of referred unclassified neuropathies yields improved diagnosis. Ann Neurol 10:222, 1981

13. Dyck PJ, Kratz KM, Karnes JL, et al: The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: the Rochester Diabetic Neuropathy Study. Neurology 43:817, 1993

14. Dyck PJ, Davies JL, Wilson DM, et al: Risk factors for severity of diabetic polyneuropathy: intensive longitudinal assessment of the Rochester Diabetic Neuropathy Study cohort. Diabetes Care 22:1479, 1999

15. Airey M, Bennett C, Nicolucci A, et al: Aldose reductase inhibitors for the prevention and treatment of diabetic peripheral neuropathy. Cochrane Database Syst Rev (2):CD002182, 2000

16. Dyck PJ: Hypoxic neuropathy: does hypoxia play a role in diabetic neuropathy? Neurology 39:111, 1989

17. Dyck PJ, Davies JL, Litchy WJ, et al: Longitudinal assessment of diabetic polyneuropathy using a composite score in the Rochester Diabetic Neuropathy Study cohort. Neurology 49:229, 1997

18. The effect of intensive diabetes therapy on the development and progression of neuropathy. The Diabetes Control and Complications Trial Research Group. Ann Intern Med 122:561, 1995

19. Navarro X, Sutherland DER, Kennedy WR: Long-term effects of pancreatic transplantation on diabetic neuropathy. Ann Neurol 42:727, 1997

20. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. UK Prospective Diabetes Study Group. Lancet 352:837, 1998

21. Albers JW, Andersen H, Arezzo JC, et al: Diabetic polyneuropathy in controlled clinical trials: consensus report of the Peripheral Nerve Society. Ann Neurol 38:478, 1995

22. Apfel SC, Schwartz S, Adornato BF, et al: Efficacy and safety of recombinant human nerve growth factor in patients with diabetic polyneuropathy. JAMA 284:2215, 2000

23. Zochodne DW, Bolton CF, Wells GA, et al: Critical illness polyneuropathy: a complication of sepsis and multiple organ failure. Brain 110:819, 1987

24. De Jonghe B, Sharshar T, Lefaucheur J-P, et al: Paresis acquired in the intensive care unit: a prospective multicenter study. JAMA 288:2859, 2002

25. Berciano J, Combarros O: Hereditary neuropathies. Curr Opin Neurol 16:613, 2003

26. Hahn AF: Guillain-Barré syndrome. Lancet 352:635, 1998

27. Moulin DE, Hagen N, Feasby TE, et al: Pain in Guillain-Barré syndrome. Neurology 48:328, 1997

28. Lasky T, Terracciano GJ, Magder L, et al: The Guillain-Barré syndrome and the 1992–1993 and 1993–1994 influenza vaccines. N Engl J Med 339:1797, 1998

29. Hartung HP, Kieseier BC, Kiefer R: Progress in Guillain-Barré syndrome. Curr Opin Neurol 14:597, 2001

30. Randomised trial of plasma exchange, intravenous immunoglobulin, and combined treatments in Guillain-Barré syndrome. Plasma Exchange/Sandoglobulin Guillain-Barré Syndrome Trial Group. Lancet 349:225, 1997

31. van der Meche FGA, Schmitz PIM: A randomized trial comparing intravenous immune globulin and plasma exchange in Guillain-Barré syndrome. N Engl J Med 326:1123, 1992

32. Plasmapheresis and acute Guillain-Barré syndrome. The Guillain-Barré Syndrome Study Group. Neurology 35:1096, 1985

33. Double-blind trial of intravenous methylprednisolone in Guillain-Barré syndrome. Guillain-Barré Syndrome Steroid Trial Group. Lancet 341:586, 1993

34. Hughes RA, van Der Meche FG: Corticosteroids for treating Guillain-Barré syndrome. Cochrane Database Syst Rev (2):CD001446, 2000

35. Wollinsky KH, Hülser P-J, Brinkmeier H, et al: CSF filtration is an effective treatment of Guillain-Barré syndrome. Neurology 57:774, 2001

36. Bowes D: The doctor as patient: an encounter with Guillain-Barré syndrome. Can Med Assoc J 131:1343, 1984

37. Hahn AF, Bolton CF, Pillay N, et al: Plasma exchange therapy in chronic inflammatory demyelinating polyneuropathy: a double-blind, sham-controlled, cross-over study. Brain 119:1055, 1996

38. Mendell JR, Barohn RJ, Freimer ML, et al: Randomized controlled trial of IVIg in untreated chronic inflammatory demyelinating polyradiculoneuropathy. Neurology 56:445, 2001

39. Dyck PJ, O'Brien PC, Oviatt KF, et al: Prednisone improves chronic inflammatory demyelinating polyradiculoneuropathy more than no treatment. Ann Neurol 11:136, 1982

40. Hadden RDM, Sharrack B, Bensa S, et al: Randomized trial of interferon beta-1a in chronic inflammatory demyelinating polyradiculoneuropathy. Neurology 53:57, 1999

41. Gosselin S, Kyle RA, Dyck PJ: Neuropathy associated with monoclonal gammopathies of undetermined significance. Ann Neurol 30:54, 1991

42. Dyck PJ, Low PA, Windebank AJ, et al: Plasma exchange in polyneuropathy associated with monoclonal gammopathy of undetermined significance. N Engl J Med 325:1482, 1991

43. Gorson KC, Ropper AH, Weinberg DH, et al: Efficacy of intravenous immunoglobulin in patients with IgG monoclonal gammopathy and neuropathy. Arch Neurol 59:766, 2002

44. Plante-Bordeneuve V, Lalu T, Misrahi M, et al: Genotypic-phenotypic variations in a series of 65 patients with familial amyloid polyneuropathy. Neurology 51:708, 1998

45. Bergethon PR, Sabin TD, Lewis D, et al: Improvement in the polyneuropathy associated with familial amyloid polyneuropathy after liver transplantation. Neurology 47:944, 1996

46. Hawke SHB, Davies L, Pamphlett R, et al: Vasculitic neuropathy: a clinical and pathological study. Brain 114:2175, 1991

47. Chalk CH, Homburger HA, Dyck PJ: Anti-neutrophil cytoplasmic antibodies in vasculitic peripheral neuropathy. Neurology 43:1826, 1993

48. Chang RW, Bell CL, Hallett M: Clinical characteristics and prognosis of vasculitic mononeuropathy multiplex. Arch Neurol 41:618, 1984

49. Collins MP, Periquet, MI, Mendell JR et al: Nonsystemic vasculitic neuropathy: insights from a clinical cohort. Neurology 61:623, 2003

50. Griffin JW, Cornblath DR, Alexander E, et al: Ataxic sensory neuropathy and dorsal root ganglionitis associated with Sjögren's syndrome. Ann Neurol 27:304, 1990

51. Grant IA, Hunder GG, Homburger HA, et al: Peripheral neuropathy associated with sicca complex. Neurology 48:855, 1997

52. Gao WQ, Dybdal N, Shinsky N, et al: Neurotrophin-3 reverses experimental cisplatin-induced peripheral sensory neuropathy. Ann Neurol 38:30, 1995

53. Behse F, Buchthal F: Alcoholic neuropathy: clinical, electrophysiological, and biopsy findings. Ann Neurol 2:95, 1977

54. Simpson DM, Tagliati M: Nucleoside analogue-associated peripheral neuropathy in human immunodeficiency virus infection. J Acquir Immune Defic Syndr 9:153, 1995

55. Brannagan TH, Nuovo GJ, Hays AP, et al: Human immunodeficiency virus infection of dorsal root ganglion neurons detected by polymerase chain reaction in situ hybridization. Ann Neurol 42:368, 1997

56. Thomas FP, Chalk C, Lalonde R, et al: Expression of human immunodeficiency virus type 1 in the nervous system of transgenic mice leads to neurological disease. J Virol 68:7099, 1994

57. McArthur JC, Yiannoutsos C, Simpson DM, et al: A phase II trial of nerve growth factor for sensory neuropathy associated with HIV infection. Neurology 54:1080, 2000

58. Roullet E, Assuerus V, Gozlan J, et al: Cytomegalovirus multifocal neuropathy in AIDS: analysis of 15 consecutive cases. Neurology 44:2174, 1994

59. Berger JR, Simpson DM: The pathogenesis of diffuse infiltrative lymphocytosis syndrome: an AIDS-related peripheral neuropathy. Neurology 50:855, 1998

60. So YT, Olney RK: Acute lumbosacral polyradiculopathy in acquired immunodeficiency syndrome: experience in 23 patients. Ann Neurol 35:53, 1994

61. Logigian EL: Peripheral nervous system Lyme borreliosis. Semin Neurol 17:25, 1997

62. Tyring S, Barbarash RA, Nahlik JE, et al: Famciclovir for the treatment of acute herpes zoster: effects on acute disease and postherpetic neuralgia. Ann Intern Med 123:89, 1995

63. Wood MJ, Johnson RW, McKendrick MW, et al: A randomized trial of acyclovir for 7 days or 21 days with and without prednisolone for treatment of acute herpes zoster. N Engl J Med 330:896, 1994

64. Rowbotham M, Harden N, Stacey B, et al: Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial. JAMA 280:1837, 1998

65. Rowbotham M: Managing post-herpetic neuralgia with opioids and local anaesthetics. Ann Neurol 35:S46, 1994

Acknowledgment

Figure 1 George Kelvin.

III DISEASES OF MUSCLE AND THE NEUROMUSCULAR JUNCTION

Marinos C. Dalakas, m.d.

Approach to Patients with Neuromuscular Disease

The diseases of muscle and the neuromuscular junction constitute a heterogeneous group of acquired and hereditary disorders. At presentation, symptoms are varied and include fatigability, skeletal muscle weakness, atrophy, muscle cramps, and impaired function of respiratory, pharyngeal, facial, and ocular muscles. In general, the proximal muscles are more selectively and severely affected than the distal muscles. Thus, patients experience difficulty climbing stairs, rising from a low chair, running, combing their hair, and lifting themselves or turning in bed. Most myopathies affect only skeletal muscles, but smooth muscle and cardiac muscle may also be impaired. Patients with myopathy do not have sensory disturbances or autonomic dysfunction, because peripheral nerves and the autonomic nervous system are spared.

Clinical examination is aimed at localizing findings to muscle and excluding diseases that may cause myopathic symptoms and signs, such as motor neuron syndromes, motor neuropathies, and psychogenic diseases. A complete family history and examination of family members are often necessary to exclude a hereditary disease. Useful laboratory evaluations include (1) studies to exclude a systemic disease, exogenous factors, toxins, or viruses that may induce myopathy; (2) electromyography (EMG) to localize the lesion to the muscle or the neuromuscular junction and to exclude motor neuron or peripheral nerve disorders; (3) determination of serum levels of muscle enzymes; (4) measurement of specific autoantibodies directed against known or putative muscle antigens; (5) muscle biopsy for enzyme histochemistry, immunocytochemistry, electron microscopy, biochemical measurement of a specific muscle enzyme or protein, and genetic studies; (6) ischemic exercise test to measure the production of lactate and ammonia in metabolic myopathies; (7) genetic tests on peripheral blood lymphocytes if the gene is known; and (8) muscle imaging, as dictated by the specific clinical problem being investigated.

Most myopathies are disabling or catastrophic if left untreated. Therefore, diagnosis must be established quickly to initiate early therapy. For patients with untreatable disorders, proper supportive care, rehabilitation, and genetic counseling are critical.

In this chapter, the most common myopathies and disorders of the neuromuscular junction are described, with emphasis on the clinical picture, pathogenesis, diagnosis, and therapy.

Muscular Dystrophies

The muscular dystrophies constitute a heterogeneous group of congenital muscle diseases characterized by severe muscle weakness, atrophy, elevation of serum muscle enzyme levels, and destructive cytoarchitectural changes of muscle fibers. The traditional classification of muscular dystrophies into Duchenne, Becker, limb-girdle, and congenital has changed because genetic defects in muscle proteins responsible for most of these diseases have been identified and the deficiency of specific muscle proteins has been demonstrated as the cause of these diseases.

Molecular, biochemical, and immunocytochemical studies[1,2] have identified the dystrophin-glycoprotein complex as a key multisubunit complex of proteins linking the cytoskeleton with the extracellular matrix [*see Figure 1*]. Deficiencies in certain components of this system cause sarcolemmal instability resulting in muscle fiber necrosis and specific clinical syndromes.[1-3] The muscular dystrophies are now best classified according to the gene and the defective protein involved; the defective protein may be a component of the nucleus, the cytosol, the cytoskeleton, the sarcolemma, the extracellular matrix, or the intermediate filament. The most common muscular dystrophies, categorized in accordance with current molecular genetic analysis and mode of inheritance, are listed [*see Table 1*]. The role of each protein in supporting, reinforcing, or connecting the nucleus with the cytoskeleton, the sarcolemma, and the extracellular matrix [*see Figure 1*] is discussed in the sections that describe each specific disorder.

X-linked recessive muscular dystrophies

Dystrophinopathies

Dystrophinopathies are caused by a deficiency of dystrophin, a 427 kilodalton rod-shaped cytoskeletal protein. Dystrophin constitutes 5% of all sarcolemmal cytoskeletal proteins and serves to anchor F-actin (the filamentous form of actin) to the plasma membrane (sarcolemma) of muscle [*see Figure 1*].[2-7] Dystrophin appears to reinforce and stabilize the plasma membrane during the stress of muscle contraction by maintaining a mechanical link between the cytoskeleton and the extracellular matrix. Deficiency or absence of dystrophin is associated with various dystrophinopathies, the prototype of which is Duchenne muscular dystrophy (DMD).[4-7]

Duchenne muscular dystrophy Duchenne muscular dystrophy, an X-linked recessive disorder [*see Table 1*], is caused by mutations in the *dystrophin* gene on the short arm of the X chromosome at position Xp21. The *dystrophin* gene spans more than 2,000 kilobases of DNA. In 65% to 70% of cases, DMD results from large deletions (several kilobases) in the *dystrophin* gene and a consequent lack of muscle dystrophin. Spontaneous mutations are also frequently noted in DMD patients.[4-7] The absence of dystrophin weakens and disrupts the sarcolemmal membrane, thereby allowing calcium entry, which causes muscle fiber necrosis. The deletions, detected in the DNA extracted from peripheral blood lymphocytes, disrupt the open reading frame of the messenger RNA (mRNA) triplet codons and result in severe forms of DMD. Partial gene duplications account for 6% of the dystrophin mutations.

A multiplex polymerase chain reaction (PCR) test that examines so-called hot spots in two exons detects almost two thirds of cases of DMD by screening DNA from the blood. However, this technique does not detect small mutations (e.g., point mutations and splicing errors) that produce a truncated dystrophin protein and account for as many as 30% of DMD cases. A highly sensitive single-strand conformation polymorphism method screens all 79 exons of the dystrophin gene and detects 90% of DMD mu-

tations by DNA analysis of the peripheral blood.[7] Similar results are obtained with PCR followed by direct sequence analysis (i.e., single condition amplification internal/primer sequencing [SCAIP]).[8]

DMD occurs in one in 3,000 male births. Affected boys become symptomatic after they begin to walk, usually from 2 to 3 years of age. DMD occurs in girls only in extremely rare circumstances [*see* Female Carriers and Dystrophinopathy in Women,

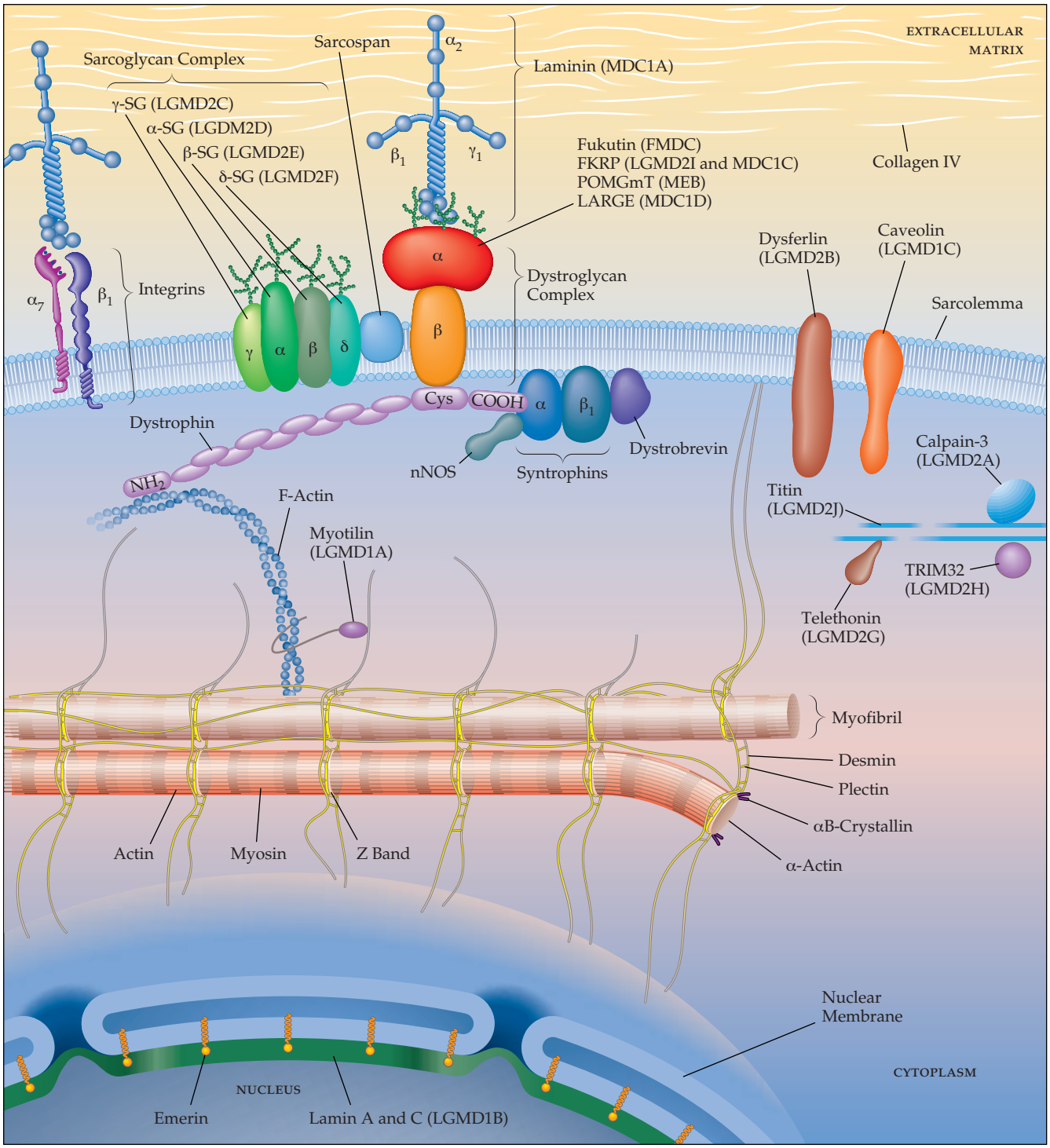


Figure 1 **The current conception of the molecular organization of the dystroglycan complex at the extrajunctional sarcolemma. The deficiency or absence of various proteins results in particular muscular dystrophies. Mutations in the gene for dystrophin produce Duchenne muscular dystrophy (DMD) or Becker muscular dystrophy (BMD); mutations in the genes for the sarcoglycan complex produce various subtypes of limb-girdle muscular dystrophies (LGMDs); and mutations in the gene for α_2-laminin (merosin) produce congenital muscular dystrophy (MDC). Other proteins in the region of the dystrophin-glycoprotein complex include caveolin, neuronal nitric oxide synthase (nNOS), dystrobrevin, the syntrophins, and actin. Collagen VI is a component of the basal lamina. (COOH—carboxyl terminus; Cys—cysteine-rich domain; NH2—amino terminus)**

below]. Clumsiness of gait, lordotic posture, calf hypertrophy, joint contractures, and toe-walking are early manifestations. These are followed by progressive muscle weakness and wasting; getting up from the floor or a low chair, climbing stairs, and arm raising become difficult. Delayed gastric emptying can cause sudden episodes of vomiting and abdominal pain. The disease is relentlessly progressive. By 12 years of age, affected children are wheelchair bound, and by 25 years of age, they die of complications of respiratory failure. Although DMD is a disease of skeletal muscle, cardiac muscle is often affected, and congestive heart failure and arrhythmia may occur later in the disease.[9] Mild, nonprogressive central nervous system involvement also occurs and manifests as irritability, hyperactivity, or cognitive dysfunction. Patients with cognitive impairment have been found to have abnormal electroretinograms owing to defects in a dystrophin isoform expressed in the retina.[7]

Routine laboratory studies of patients with DMD initially show serum creatine kinase (CK) levels as high as 20,000 IU/L, which steadily decline as the muscle mass is depleted. Muscle biopsies reveal a severe destructive myopathy; some $CD8^+$ T cells are often present and invade muscle fibers; a large number of macrophages are associated with phagocytosis of necrotic fibers; the amount of connective tissue is increased; and hypercontracted muscle fibers are common. Diagnosis is confirmed by the absence of dystrophin or by levels lower than 3% of the normal dystrophin concentration; dystrophin levels are demonstrat-

Table 1 Muscular Dystrophies with Gene Locations and Products

Disease	*Gene Locus*	*Gene Product*	*Allelic Disorders*
X-linked recessive muscular dystrophies			
Duchenne muscular dystrophy	Xp21	Dystrophin	Isolated cardiomyopathy, Becker muscular dystrophy
Becker muscular dystrophy	Xp21	Dystrophin	Isolated cardiomyopathy, Duchenne muscular dystrophy
EDMD	Xq28	Emerin	LGMD1B
Autosomal dominant LGMD			
LGMD1A	5q22-5q31	Myotilin	Myofibrillar myopathy
LGMD1B	1q11-21	Lamin A and C	Autosomal dominant EDMD
LGMD1C	3p25	Caveolin-3	
LGMD1D	7q	Unknown	
LGMD1E	7 q	Unknown	
Autosomal recessive LGMD			
LGMD2A	15q15	Calpain-3	
LGMD2B	2p13	Dysferlin	
LGMD2C	13q12	γ-Sarcoglycan	
LGMD2D	17q12-q21	α-Sarcoglycan	
LGMD2E	4q12	β-Sarcoglycan	
LGMD2F	5q33-q34	δ-Sarcoglycan	
LGMD2G	17q11-q12	Telethonin	
LGMD2H	9q3-q34	TRIM32	
LGMD2I	19q13.3	Fukutin-related protein	MDC1C, FMDC
LGMD2J	2q24.3	Titin	
MDCs			
MDC1A "classic" MDC	6q22	α_2-Laminin (merosin)	
MDC1B	12q13	α_7-Integrin	
MDC1C	19q13.3	Fukutin-related protein	LGMD2I
MDC1D		LARGE	
FMDC	9q31-q33	Fukutin	
Muscle-eye-brain disease	1p32-p34	POMGnT	
Walker-Warburg syndrome	?	POMGnT	
Other MDCs			
MDC with rigid spine myopathy	1q35-36	Selenoprotein N	
Bethlem autosomal dominant myopathy	21q22	Collagen VIα1, α2	
Ullrich myopathy	2q37	Collagen VIα3	
Intermediate filament myopathy			
Desmin	2	Desmin	
αB-Crystallin	11q21-23	Desmin, αB-Crystallin	
Epidermolysis bullosa and muscular dystrophy	8q24-qter	Plectin	
Myotilin	5q22-5q31	Myotilin	
Autosomal dominant dystrophies with a unique phenotype			
Myotonic dystrophy			
DM1	19q13	Myotonin-protein kinase	
DM2	3q21	Zinc finger protein 9	
Facioscapulohumeral muscular atrophy	4q35		
Oculopharyngeal muscular dystrophy	14q11.2-q13	Poly A binding protein 2	

EDMD—Emery-Dreifuss muscular dystrophy FMDC—Fukuyama congenital muscular dystrophy LGMD—limb-girdle muscular dystrophy MDC—congenital muscular dystrophy

ed either by immunocytochemistry on muscle biopsy sections or by immunoblots prepared from muscle biopsy specimens stained with antidystrophin antibodies.

Treatment is entirely symptomatic, with emphasis on providing systematic respiratory and physical therapy as well as psychosocial support for both the patient and the family. Genetic counseling is highly appropriate. The presence of endomysial inflammation has prompted the use of steroids to treat DMD. In a controlled study, use of steroids resulted in temporary, mild improvement and slowed disease progress.[10] The long-term use of steroids is restricted by severe side effects, especially obesity, fractures, osteoporosis, diabetes, and hypertension. Deflazacort, a steroid with fewer mineralocorticoid side effects, appears to be a little safer.[11] Creatine supplementation is often used, although its benefit is marginal.[12] In a modified form of gene therapy, human myoblasts carrying normal dystrophin were injected into the muscles of patients with DMD, but the procedure failed to change the recipients' muscle function.[13] Several prospective, randomized, placebo-controlled trials involving repeated myoblast injections to the same muscles failed to demonstrate any improvement in strength. The finding in dystrophic mice that aminoglycoside antibiotics can "read through" nonsense mutations and generate a full-length protein, thereby restoring dystrophin function,[14] led to a trial examining the efficacy of gentamicin therapy in DMD patients who had a premature stop codon. The results were disappointing, although mild reduction in the serum CK level was noted.[15] Future gene therapies may prove efficacious if proper vectors are found that can be used to effectively insert the gene into the muscle.[16]

Becker muscular dystrophy Becker muscular dystrophy (BMD) and DMD are allelic disorders [*see Table 1*], but BMD generally starts later and progresses more slowly.

About 65% of patients with BMD have in-frame deletions in the *dystrophin* gene, but the produced protein is often truncated and only semifunctional.[2-6] Instances of duplication of the *dystrophin* gene have resulted in a longer dystrophin rod. Immunocytochemistry of BMD muscle using antidystrophin antibodies demonstrates preserved but attenuated sarcolemmal staining (not absence, as in DMD) and reveals membrane fragmentation in the immunostained areas of the sarcolemma. Immunoblots detect a reduced amount of a smaller than normal or larger than normal dystrophin.

The age at onset of BMD is variable. Cases may be recognized by as early as 3 years of age or as late as 70 years of age; the mean age at onset is 12 years. The spectrum of phenotypic expression of BMD is also wide. Mild forms manifest only as muscle cramps, exercise intolerance, myoglobulinuria, asymptomatic elevation of serum CK levels, mild muscle weakness, or quadriceps myopathy.[2-6] Calf pain on exercise is often a presenting symptom, and calf enlargement is frequent. Most patients lose ambulation by the age of 40 (range, 10 to 70 years of age). The age of death also varies, from 23 to 89 years (mean age, 42 years). Patients present with proximal muscle weakness and serum CK levels as high as 20 times normal. The muscle biopsy findings are similar to those in DMD patients but are not as severe. In patients younger than 8 years, the presentation of BMD is usually indistinguishable from that of DMD. Cardiac manifestations are common, and cardiomyopathy can be severe. The severity of cardiac symptoms, however, is unrelated to the severity of the myopathy. There is no effective therapy for BMD.

Female carriers and dystrophinopathy in women Careful history and clinical examination of asymptomatic female carriers may reveal mild muscle weakness, muscle cramps, isolated calf hypertrophy, fatigue, and elevated serum CK levels.[3-7] Muscle biopsy reveals dystrophin-negative fibers. In heterozygotes, when the specific deletion-prone exons within the dystrophin gene are amplified using quantitative PCR, the deletions are recognized as a 50% reduction in the intensity of the amplified DNA band, compared with the band of the wild-type exons.[4-7] This method detects about 98% of the deletions. Not infrequently, however, the affected boy's mother does not carry his mutation in her blood. Such cases, linked to newly recognized DMD mutations, account for as many as 20% of new DMD cases and result from maternal gonadal mosaicism; that is, the mutations are found only in the oocytes.[4-7] Women with these mutations may produce affected males or carrier females. Daughters of these women should be studied to identify carriers; however, because the mutation occurs in the oocytes, a woman's sisters could not have inherited the mutations from their parents and need not be studied as potential carriers.[4-7] The manifestation of DMD in heterozygote females occurs when the normal paternal X chromosome that harbors the normal dystrophin gene is inactivated in a large proportion of embryonic cells (Lyon hypothesis). The disease in these females may be as severe as in males.

X-linked dilated cardiomyopathy X-linked dilated cardiomyopathy results from dystrophin deficiency in cardiac but not skeletal muscle. Patients present with a progressive cardiac disorder; congestive heart failure occurs in the second or third decade of life. Female carriers manifesting the disease have a slow-onset cardiomyopathy that presents by the fifth decade. Deletions near exon 1 of the dystrophin gene, which affect the expression or function of dystrophin in cardiac muscle, have been proposed to cause the disease.[4-7]

Emery-Dreifuss Muscular Dystrophy

Emery-Dreifuss muscular dystrophy (EDMD) has two genetic forms, an X-linked recessive disease mapped to Xq28 (XL-EDMD) and a less common autosomal dominant form (AD-EDMD) mapped to 1q11-q23. The two forms are clinically indistinguishable. XL-EDMD is caused by mutations in the nuclear membrane protein emerin [*see Figure 1*][17]; the AD-EDMD is caused by mutations in the *lamin A/C* gene, which encodes two proteins of the nuclear lamina, lamin A and lamin C.[18] Mutations in the *lamin A/C* gene most often cause cardiomyopathy and conduction defects.[19] The integral nuclear membrane proteins interact closely with nuclear lamins, which are intermediate filament proteins found on the nucleic side (inner side) of the nuclear membrane [*see Figure 1*]. Emerin binds lamin A, one of the lamin A/C gene products.

Patients with EDMD present with a triad of symptoms: (1) a humeroperoneal distribution of muscle involvement with prominent wasting and weakness in the biceps, triceps, anterior tibial, and peroneal muscles that progresses slowly in a scapulopelviperoneal pattern to include pectoral and pelvic muscle involvement; (2) early development of contractions at the elbows, flexors, Achilles tendon, neck, and spine, which may occur before there is significant muscle weakness; and (3) cardiac involvement, which presents as conduction defects with bradycardia and a prolonged PR interval. Isolated atrial paralysis strongly suggests EDMD.

Female carriers of XL-EDMD generally do not have skeletal muscle weakness; however, they can develop heart block.

EDMD begins within the first two decades of life. Diagnosis is made on the basis of clinical presentation and muscle biopsy and confirmed by mutation analysis; however, a slightly elevated CK level is suggestive. Muscle biopsies reveal nonspecific myopathic features, but the absence of emerin as evidenced by applying antibodies to emerin on immunocytochemistry or Western blot assay on specimens derived from muscle biopsies of patients and their families may support the diagnosis. Available genetic tests identify the carriers who may have cardiac conduction defects. Prompt recognition and cardiac pacing can prevent sudden death or syncopal attacks.

No specific treatment is available, but physical therapy may slow the development of contractions.

LIMB-GIRDLE MUSCULAR DYSTROPHIES

Limb-girdle muscular dystrophies (LGMD) comprise a heterogenous group of disorders designated as autosomal dominant (LGMD1), autosomal recessive (LGMD2), or congenital (MDC) [*see Table 1*]. [4,6,20]

Autosomal Dominant Limb-Girdle Muscular Dystrophies

The autosomal dominant forms of LGMD1 [*see Table 1*] are generally uncommon disorders representing less than 10% of all LGMDs; they present with slowly progressive proximal and distal muscle weakness and elevations of serum CK levels. They tend to be milder than other LGMD presentations. The diagnosis is made by identifying the missing protein on muscle biopsy preparation and the genetic defect on mutation analysis. There is no specific treatment for this group of disorders.

LGMD1A LGMD1A results from mutations in myotilin, a protein necessary for normal assembly and maintenance of the sarcomere.[21] In addition to proximal and distal muscle weakness, symptoms may include dysarthria, hypophonia, and a nasal voice. This disorder is allelic to the form of myofibrillar myopathy caused by myotilin mutation [*see* Myopathies Due to Mutations in the Intermediate Filament Proteins, *below*].[21]

LGMD1B LGMD1B is caused by mutations in *lamin A/C*; the phenotype for this disorder is identical to that of AD-EDMD [*see* Emery-Dreifuss Muscular Dystrophy, *above*]. In addition to having LGMD symptoms, some patients have familial partial lipodystrophy characterized by reduced subcutaneous fat, insulin resistance, increased triglyceride levels, low levels of high-density lipoproteins (HDLs), diabetes mellitus, and increased risk of atherosclerotic vascular disease.[6]

LGMD1C LGMD1C is caused by mutations in the *caveolin-3* gene.[19,22] Caveolin-3, a 21 to 24 kd internal membrane protein, may play a role in the regulation of muscle glycolysis. Patients may have different clinical phenotypes; they may present with LGMD, isolated hyperCKemia, rippling muscle disease, or distal myopathy.

LGMD1D, 1E, 1F These three disorders represent rare mutations; linkage analysis has determined the chromosomal loci of the causative mutations, but the pertinent genes have not been identified.

Autosomal Recessive Limb-Girdle Muscular Dystrophies

At present, 10 forms of LGMD2 have been identified, and more are likely to be recognized [*see Table 1*].[20,23]

LGMD2A LGMD2A is caused by mutations in the *calpain-3* gene.[24] Calpain deficiency is the most frequent form of LGMD. Calpain-3 is a calcium-activated protease that plays a role in muscle differentiation.

Disease onset occurs in patients 8 to 30 years of age. Patients with calpain-3 deficiency present with weakness in the pelvic girdle muscles, especially glutei, that spares the hip abductors. Scapular winging and posterior thigh involvement are commonly seen. The serum CK level can range from 100 IU/L to more than 9,000 IU/L. Mutations in the gene encoding calpain-3, which is located on chromosome 15q15.1–q15.3, have been identified.[24] The diagnosis is confirmed with Western blot assay. There is no specific treatment for this disorder.

LGMD2B LGMD2B is caused by mutations in the *dysferlin* gene,[25,26] which is a membrane-associated protein not part of the dystrophin-glycoprotein complex [*see Figure 1*]. Dysferlin interacts with caveolin and may be involved in membrane repair. Mutations in the *dysferlin* gene cause two types of myopathy: the Miyoshi myopathy, which is characterized by very high serum CK levels and early involvement of the gastrocnemius muscle, and LGMD2B, which may present as distal and proximal muscle weakness in patients in their late teens or early 20s. Identical mutations can cause either Miyoshi myopathy or LGMD2B.[25,26]

Some patients with dysferlinopathy present with prominent distal weakness but do not have the phenotype of Miyoshi myopathy. Dysferlinopathies are common disorders. The diagnosis is confirmed by finding absence of dysferlin in the sarcolemma of muscle biopsy specimens using immunohistochemistry and antidysferlin antibodies. The defect can also be seen in peripheral blood monocytes using Western blot assay. The absence of dysferlin could lead to perturbation of membrane resealing and interference with the repair of damaged muscle fibers, possibly resulting from defects in vesicular trafficking within the muscle fiber.[6,20,23] There is no specific treatment for this disorder.

LGMD2C, 2D, 2E, 2F These four disorders are caused by mutations in the genes encoding the four members of the sarcoglycan complex: LGMD2C (γ-sarcoglycan gene mutation mapped to chromosome 5q33), LGMD2D (α-sarcoglycan, mapped to 17q12), LGMD2E (β-sarcoglycan, mapped to 4), and LGMD2F (δ-sarcoglycan, mapped to 13q12).[20,23] In all four disorders, the three other components of the sarcoglycan complex are lost or partially absent, but the dystroglycan complex is normal.[4,6,20,23,27]

Mutations in the sarcoglycan genes have been identified in various European and American families who have autosomal recessive LGMDs; these mutations include missense, nonsense, and in-frame deletions that result in an LGMD phenotype of varying clinical severity. The mutations cause an improper assembly of the sarcoglycans with the dystroglycan complex proteins, disrupting the linkage between sarcolemma and extracellular matrix [*see Figure 1*].

The transmembrane components of the sarcoglycan complex are specific to skeletal and cardiac muscle, and their integrity is critical for normal muscle physiology.

The autosomal recessive LGMDs 2D, 2E, and 2F are similar in presentation to a severe form of DMD; they affect males and fe-

males. Patients present with mild to severe weakness of the proximal muscles (especially in the legs), an elevation of the serum CK level to about 3,000 IU/L, dystrophic changes in the muscle, and frequently calf hypertrophy. The age at onset is variable. Results of immunocytochemistry and immunoblot analysis of muscle biopsy specimens for dystrophin are normal, but sarcoglycan is absent or severely deficient. Deficiency in the sarcoglycan proteins—easily detected using the proper antibodies—is found in 20% of patients with the presentation of LGMD. Mutations in the respective genes have been reported in as many as 60% of patients.

There is a great variability in the severity of muscle weakness, which probably depends on the degree of residual sarcoglycan expression. Of note, cardiomyopathy may also be present either as an isolated manifestation of dilated cardiomyopathy or in conjunction with the skeletal myopathy. There is no specific treatment for this disorder.

LGMD2G LGMD2G is caused by mutations in the gene encoding telethonin, which is a sarcomeric protein localized to the Z disk of skeletal muscle. Mutations in the *telethonin* gene cause disruption of the sarcomeric structure.[28] LGMD2G is a rare, relatively mild autosomal recessive disorder.[28]

Disease onset occurs at 2 to 15 years of age, often with distal muscle involvement. Disease progression is variable. Muscle biopsy may show vacuoles within the muscle fibers. The serum CK levels are increased 10-fold to 30-fold above normal. The absence of telethonin as detected by immunocytochemistry confirms the diagnosis.

LGMD2H LGMD2H is a very rare disease identified only in the Manitoba Hutterite population. It is caused by mutations in the *TRIM32* gene (the Tripartite-motif-containing gene) on chromosome 9q33.1, which is probably related to an E3-ubiquitin ligase.[29]

LGMD2I LGMD2I is caused by mutations in the *fukutin-related protein (FKRP)* gene, on chromosome 19q13.32.[30] This disorder and the Fukuyama congenital muscular dystrophy (FMDC) are allelic disorders [*see* Fukuyama MDC, *below*]. Both are related to changes in the α-dystroglycan (α-DG) expression resulting from defects in glycosylation. LGMD2I is not uncommon; in some countries, such as the United Kingdom,[31] it is one of the most common forms of LGMD.

Onset of LGMD2I is variable, occurring in early childhood or in adulthood. Patients present with proximal muscle weakness similar to that found in BMD; enlarged calves; an elevated serum CK level; sometimes, an enlarged tongue; and, very commonly, involvement of the respiratory and cardiac muscles. Intelligence is normal. Secondary reduction of α-DG expression on muscle biopsy or reduced molecular weight protein on immunoblot assay, accompanied by reduction of α_2-laminin by immunocytochemistry, raises suspicion for the diagnosis of LGMD2I.

LGMD2J LGMD2J results from mutations in the *titin* gene, on chromosome 2q31. This disorder, which has been reported in Finland,[32] is allelic with the tibial muscular dystrophy, an automosmal dominant disease that has been reported in Finland.

Congenital Muscular Dystrophies

MDCs are autosomal recessive disorders characterized by muscle weakness in early infancy. Frequently, these disorders are associated with brain malformations and cognitive abnormalities.[33-35] In the neonatal period, the serum CK level is highly elevated, often into the thousands. MDCs represent disorders of neuronal migration, referred to as cobblestone cortex. They result in abnormal glycosylation of α-DG that disrupts the interaction between the membrane and the extracellular matrix in muscle and brain [*see Figure 1*]. They are often referred to as α-dystroglycanopathies [*see Table 1*].

MDC1A MDC1A is the classic form of MDC, accounting for more than 40% of cases of MDC. The disease is linked to chromosome 6q22 and is caused by a defect in merosin (α_2-laminin), which represents the structural backbone of the basement membrane [*see Figure 1*]. In contrast to other forms of MDC, there is no brain malformation or mental retardation in MDC1A. This disease occurs in children but can also occur in young adults. The patients have α_2-laminin deficiency in the skeletal muscle basal lamina.[4,6,35] Some patients have partial merosin deficiency resulting either from a mild mutation or from secondary causes, most often mutations in the genes encoding fukutin or FKRP.

MDC1A may be associated with an MRI-defined leukoencephalopathy (without overt signs of intellectual deterioration), neonatal hypotonia, elevated serum CK levels, delayed motor milestones, axonal neuropathy, and respiratory muscle weakness. The degree of the clinical phenotype varies from moderate to severe. In patients who lack α_2-laminin, the dystroglycan complex is deranged, and consequently, the muscle membrane becomes defective.[4,6,35] Because α_2-laminin, together with β_1-laminin and γ_1-laminin, are also present in Schwann cells, some MDC patients have neuropathic findings.

MDC1B This disorder is caused by mutations in the *α_7-integrin* gene on chromosome 12q13. In patients with this disorder, *α_7-integrin* is absent from the sarcolemma.[36] The clinical presentation of MDC1B is identical to that of MDC1A.

MDC1C MDC1C is a rare disease caused by mutations in *FKRP* on chromosome 19q13.3. MDC1C and LGMD2I are allelic disorders. Onset of muscle weakness occurs in the first week after birth; affected children do not achieve independent ambulation. The disease is marked by respiratory involvement; in contrast to most other MDCs, in MDC1C intelligence may be normal. The disease is suspected by the variable reduction or absence of α-DG in the muscle or reduced α-DG size on immunoblot.[33,34]

MDC1D MDC1D is caused by a mutation in the *LARGE* gene. This is a rare disorder resulting in severe mental retardation and dystrophy. Diagnosis is confirmed by the absence of α-DG on muscle biopsy.[33,34]

Fukuyama MDC This disorder is the most common congenital muscular dystrophy in persons of Japanese descent. FMDC is caused by a mutation in the *fukutin* gene that results in a deficiency of fukutin and decreased glycosylation of α-DG detected by reduced α-DG staining in muscle. Patients have brain malformations, profound mental retardation, and ophthalmologic abnormalities. The disease starts before the age of 9 months, and the patients never learn to walk. They die by the age of 20.[33,34]

Muscle-eye-brain disease Muscle-eye-brain disease (MEB) is caused by a mutation in a glycosyltransferase *POMGnT* gene, which results in a deficiency of α-DG, as confirmed by immuno-

cytochemistry. Patients with MEB present with hypotonia, weakness, mild to moderate hydrocephalus, cortical or cerebellar hypoplasia, and eye abnormalities (e.g., myopia, microphthalmia, and optic nerve hypoplasia).[33,34]

Walker-Warburg syndrome This disorder results from mutations in an O-mannosyl transferase *POMGnT$_1$* gene. It is the most severe MDC, and it shares clinical features of both FMDC and MEB. Patients usually die by age 3.[33,34]

MDC with joint contractures Three forms of MDC are characterized by joint contractions; patients with these disorders have neither a defect in glycosylation nor mental retardation. These diseases are congenital muscular dystrophy with rigid spine syndrome (RSMD), Ullrich myopathy, and Bethlem myopathy. No specific therapy is available.

RSMD results from mutations in selenoprotein N (SEPN$_1$) on chromosome 1q35-36.[37] Patients with this disorder present with muscle weakness; hypotonia; facial weakness; and a distinctive spinal rigidity that results in inability to flex the neck, scoliosis, and respiratory difficulties.

Ullrich myopathy is caused by mutations in one of the polypeptide chains that form collagen VI, which is needed to interact with the extracellular matrix.[38] Patients have excessive distal mobility combined with proximal contractures, rigid spine at birth, and muscle weakness. Muscle biopsy shows absence of collagen VI. Ullrich myopathy and Bethlem myopathy are allelic disorders.

Bethlem myopathy is an autosomal dominant disease caused by mutations in subunits of the extracellular matrix proteins, collagen VI α_1, α_2 and α_3.[39] Disease onset is in childhood or adolescence. Clinical manifestations include mild muscle weakness and contractures in multiple joints, which may be present even at birth.

MYOPATHIES DUE TO MUTATIONS IN THE INTERMEDIATE FILAMENT PROTEINS

Intermediate filaments play a critical role in providing mechanical integration for the myofibrils and in protecting the muscle fiber from repeated mechanical stress. The protein desmin is the chief intermediate filament of the skeletal and cardiac muscle; it maintains the structural and functional integrity of the myofibrils and functions as a cytoskeletal protein, linking Z bands to the plasma membrane [*see Figure 1*]. The heat shock protein αβ-crystallin acts as a chaperone protein in protecting the desmin filaments from stress-induced damage. Desmin filaments encircle the Z bands and are fastened to them and to one another by plectin filaments. Myotilin is a Z-disc protein that interacts with α-actinin and filamin C and directly binds to F-actin [*see Figure 1*].[21]

Defects in intermediate filaments cause myofibrillar myopathies; the prototypic myopathy in this group is the desmin myopathy caused by mutations in the desmin gene.[40-42] The desmin myopathy is of dominant or, rarely, sporadic inheritance; patients may present either with cardiac conduction defects or with a distal-onset skeletal myopathy that progresses to involve proximal, facial, or respiratory muscles.

Mutations in the *αB-crystallin* gene cause a skeletal myopathy similar to desmin myopathy.[43] A case of myopathy associated with mutant plectin has also been reported.[44] Mutations in the myotilin also cause myofibrillar myopathy.[21] As in other myofibrillar myopathies, patients with myotilin mutations present with distal muscle weakness; over time, they develop cardiomyopathy. The diagnosis is suspected when myofibrillar products are accumulated on muscle biopsy, as evidenced by enzyme histochemical stains; it is confirmed by mutation analysis.

There is no specific treatment for these disorders, but early recognition may lead to the identification of potential candidates for insertion of a cardiac pacemaker to prevent sudden death from arrhythmias.

AUTOSOMAL DOMINANT DYSTROPHIES WITH A UNIQUE PHENOTYPE

Myotonic Dystrophy

Myotonic dystrophy (DM) is the most common adult muscular dystrophy. It has an incidence of one per 8,000 population and a prevalence of about five per 100,000 population. The disorder comprises two subsets: DM1 (the classic myotonic dystrophy, Steinert disease) and DM2 (proximal myotonic myopathy [PROMM]) [*see Table 1*]. Both are autosomal dominant multiorgan syndromes; they exhibit striking similarities in clinical manifestations.[45,46]

Diagnosis DM has a unique distribution: (1) ptosis of the eyelids, without extraocular muscle involvement; (2) atrophy of the masseters and the temporal muscles, which results in a unique, narrow facial configuration; (3) sternocleidomastoid muscle atrophy with relatively preserved posterior neck and shoulder girdle muscles (a clinical sign that differentiates DM from facioscapulohumeral muscular dystrophy [FSHD]); (4) distal muscle group atrophy, with slight proximal involvement in the earlier stages of disease; and (5) involvement of the palatal and pharyngeal muscles, which may produce dysarthria and dysphagia.[4,6,45,46]

Myotonia, defined as the slowing of relaxation of a normal muscle contraction, is an important clinical sign. To elicit myotonia during examination, a patient is requested to make a firm hand grip and let it go rapidly; in myotonia, an inability to immediately release the grip is evident. Percussion of the thenar eminence or the extensor digitorum also shows the characteristic slow relaxation of myotonia. Systemic features include cardiac conduction defects, mild mental dysfunction (often with silly or inappropriate behavior and expressions), testicular atrophy, frontal baldness, cataracts, gastrointestinal tract involvement (with delayed motility and emptying), hypersomnia, and a diminished response to hypoxia, which leads to poor concentration and apathy. Clinically DM2 resembles adult DM1; however, the degree of expression of the systemic features may vary from that found in DM1. The most important factors distinguishing DM1 from DM2 are the preferential proximal muscle involvement in DM2, the lack of congenital DM2 forms, the rare incidence of anticipation (see below) in DM2 , and the rare incidence of cognitive dysfunction in DM2.

Children of affected mothers may have reduced fetal movements and early life symptoms of severe hypotonia, feeding difficulties, bilateral facial weakness, and respiratory distress (congenital myotonic dystrophy). Myotonic dystrophy must be distinguished from myotonia congenita, which follows recessive and dominant inheritance patterns. Patients with myotonia congenita present with myotonia and often with muscle hypertrophy. In contrast to myotonic dystrophy, however, myotonia congenita is not associated with muscle weakness, atrophy, or systemic symptoms; it results from a distinct genetic defect that affects chloride channels [*see* Ion Channelopathies, Periodic Paralyses, and Nondystrophic Myotonias, *below*].

The clinical diagnosis of myotonic dystrophy is confirmed by EMG, which shows the myotonic discharges. In difficult cases, slit-lamp examinations may show early cataract formation. The clinical expression of myotonic dystrophy is variable, and the disorder may remain undiagnosed until patients have had children. In DM1 the age at onset is progressively earlier in successive generations (a phenomenon known as anticipation). It is not unusual to see families in which a grandmother's only symptom was early cataracts, and yet her daughter, who did not know that she herself was affected, gave birth to a child with severe congenital myotonic dystrophy. Anticipation is common in DM1 but rare in DM2.

Genetic counseling and genetically targeted therapy DM1 is caused by an unstable CTG (cytosine, thymine, guanine) trinucleotide repeat sequence in the nonprotein coding region of a protein kinase gene, called *DMPK*, on chromosome 19.[45-48] In mildly affected persons, a polymorphic CTG repeat region in the protein kinase gene expands by 50 to 80 repeats; in severely affected persons, more than 2,000 repeats may be present. The size of the CTG expansion increases over generations, which accounts for the anticipation. Measurements of the CTG expansion length can be used to confirm the presence of myotonic dystrophy in family members, for prenatal diagnosis, or for effective genetic counseling of asymptomatic persons who are at risk for myotonic dystrophy.

DM2 is also caused by an expanded repeat in a nonprotein coding region involving a tetranucleotide CCTG repeat in the *ZNF9* gene (zinc finger protein 9), on chromosome 3q21, which codes a transcription factor. The pathogenic factor of DM1 and DM2 is the RNA product of the mutant gene, not the protein product. This is the first disease known to be caused by harmful RNA. The mutant RNA forms inclusions in the nucleus (ribonuclear inclusions). It has been shown that the RNA repeat binds to proteins in the *muscleblind* family and interferes with the normal function of RNA processing.[48] These new data suggest that therapy for DM should be aimed at eliminating the harmful RNA.

Treatment At present, therapy for DM is symptomatic. Emotional support and education regarding the precautions necessary to avoid falls and injuries are essential. Careful monitoring of the cardiac status, especially during the administration of anesthesia, is important. Drugs such as quinine, procainamide, mexiletine, phenytoin, and beta blockers may help relieve the myotonia but not the weakness. Of these agents, mexiletine appears to be the most effective. Testosterone has failed as therapy for DM. Creatine monohydrate may offer minimal relief from myalgia. Modafinil may reduce excessive daytime somnolence, improve mood, and decrease fatigue.

Facioscapulohumeral Muscular Dystrophy

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common form of muscular dystrophy. It usually begins during the second decade of life.

FSHD is linked to chromosome 4q. It is an autosomal dominant disease, but 25% of cases are the result of new mutations.[49] There is significant heterogeneity in this disease. Some patients with the deletion mutation do not have the typical FSHD features [*see* Diagnosis, *below*]; instead they may present with distal myopathies characteristic of LGMD. In some family members, only minimal facial muscle involvement may occur. In 10% of families, germline mosaicism occurs; this means that more than one sibling is affected in a given generation without involvement of either parent.

Diagnosis Patients present with facial muscle weakness (especially of the orbicular muscle of the eye); the extraocular and masseter muscles are spared. Early weakness of the scapular muscles produces prominent scapular winging and gives the shoulders a forward, sloped appearance. Weakness in the anterior tibial muscles, which leads to footdrop, is always present. The disease progresses slowly, and there are long periods of stability. Tongue atrophy is not unusual. Progression occurs in a descending manner: involvement of the shoulder girdle muscles is followed by involvement of the biceps, triceps, and pelvic girdle muscles. A large majority of patients have retinal capillary abnormalities, retinal detachment, and impaired hearing; these findings are more frequent in the infantile form of FSHD.

The serum CK level in FSHD is slightly elevated, and cardiac muscle is spared. Muscle biopsy findings are variable and may include the presence of inflammatory cells. The diagnosis is suspected on clinical grounds and confirmed by DNA analysis. The disease is caused by a deletion within a series of 3.3 kb repeats (D4Z4) on chromosome 4. When the number of repeats falls below a critical number (approximately 10), there is clinical expression of the gene. About 95% of patients have a deletion resulting in a DNA short fragment that is less than 35 kb in length, as determined by the use of certain restriction enzymes.[49]

Treatment In controlled trials, use of prednisone or the $beta_2$-adrenergic agonist albuterol has shown no benefit in the treatment of FSHD**,** although the latter may increase muscle mass.[50]

Oculopharyngeal Muscular Dystrophy

Oculopharyngeal muscular dystrophy (OPMD) is a rare autosomal dominant disease that manifests from the fourth to the sixth decades of life. It is characterized by ptosis and dysphagia, both of which can be severe. Mild distal muscle weakness may occur. The mutation responsible for OPMD is caused by a trinucleotide GCG expansion repeat in the first exon of the poly(A) binding protein 2 gene (*PABP2*), on chromosome 14q11.2. PABP2 is localized to the intramuscular aggregates of muscle fibers and is causally related to these inclusions.[51]

Malignant Hyperthermia

Malignant hyperthermia (MH) occurs in one in 50,000 to one in 100,000 adults during general anesthesia, especially when halothane is used alone or in conjunction with succinylcholine and other depolarizing muscle relaxants. MH is characterized by a rapid increase of aerobic and anaerobic metabolism, during which the body temperature may exceed 43° C (109.4° F). In addition, the concentrations of carbon dioxide and lactate increase (the arterial carbon dioxide tension [P_aCO_2] may exceed 100 mm Hg), and the blood pH may fall below 7.

ETIOLOGY

Triggers for MH include potent volatile anesthetics, such as halothane, enflurane, desflurane, cyclopropane, ether, and succinylcholine. Although MH occurs in some patients without a known muscle disease, patients at risk include those who have multiple congenital musculoskeletal abnormalities, isolated congenital hip dislocation, or central core disease; it has also been known to occur in some patients with DMD or BMD.

PATHOGENESIS

Susceptibility to MH is inherited as an autosomal dominant trait; in up to 50% of the pedigrees, mutant alleles have been found on the ryonidine receptor (RyR),[52] on chromosome 19q13, or on the *CACNA1S* gene on chromosome 1q, which encodes the α_1 subunit of the dihydropyridine-sensitive L-type voltage-dependent calcium channel.[52] Some other loci have been found on a second dihydropyridine receptor locus (*CACNLA*$_2$) on chromosome 7q. MH appears to be precipitated by an inability to control calcium concentrations within the muscle fibers because of a malfunctioning sarcoplasmic reticulum (SR) and mutations in the *RyR* gene. The RyR is the calcium release channel of the SR and bridges the gap between the SR and the transverse tubule. Mutations in the receptor affect communication between the SR and the transverse tubule such that accelerated calcium release from the SR occurs when depolarization of the transverse tubule takes place.

DIAGNOSIS

MH presents with tachycardia, muscle rigidity (caused by muscle contracture that may progress to rigor or death), increased muscle permeability (resulting in increased serum levels of K^+, Ca^+, and Na^+ and muscle edema), excessive release of myoglobulin from the muscle, and myoglobinuria. Trismus or masseter muscle spasm that occurs during induction of anesthesia may be indicative of MH. MH must be distinguished from (1) postanesthetic rhabdomyolysis after muscular stress; (2) toxic reaction to drugs; (3) porphyria; (4) thyroid storm precipitated by surgery and anesthesia; and (5) the neuroleptic malignant syndrome precipitated by psychoactive drugs (e.g., haloperiodol and phenothiazines) that block central dopaminergic pathways.[52]

The in vitro caffeine-halothane contraction test, which is performed on muscle biopsy specimens, is used to screen patients for MH. The sensitivity and specificity of this test, if performed properly, may be as high as 97% and 80%, respectively.[53]

TREATMENT

Dantrolene is an effective treatment for MH. It decreases calcium release from the SR without altering calcium reuptake. Dantrolene use has reduced the mortality of this condition to 7%. The acute episode of MH is treated symptomatically. Intravenous dantrolene (2 to 10 mg/kg every 5 minutes) must be given early in the episode, while there is still adequate muscle perfusion. For patients who are known to be susceptible to MH, dantrolene can be administered at a dosage of 2 mg/kg 10 to 15 minutes before anesthesia administration. The best way to prevent MH episodes in susceptible persons is to use safe anesthetic agents (e.g., nitrous oxide and thiopental) and nondepolarizing muscle relaxants.

Metabolic Myopathies

PRINCIPLES OF MUSCLE ENERGY

As a group, the metabolic myopathies are characterized by deficiency of energy production caused by disorders of glycogen, lipid or mitochondria.

The two major sources of energy for muscle are glycogen and fatty acids whose metabolic pathways converge into acetyl coenzyme A (acetyl-CoA) for final oxidation within the mitochondria through the Krebs cycle and the respiratory chain.[54,55] At rest, muscle energy is mostly derived from oxidation of free fatty acids (FFAs). During high-intensity aerobic exercise, glycogen is the main source of fuel for oxidative phosphorylation. Muscle glycogen stores are depleted after 90 minutes of exercise. If exercise is prolonged, utilization of FFAs and blood glucose increases. Because the availability of FFAs from adipose tissue is almost unlimited, a healthy person can perform moderate exercise for many hours. In patients with metabolic myopathies, symptoms become evident during activities that require increased metabolic demands, such as exercise. Metabolic myopathies result from defects in the metabolism and utilization of glycogen, glucose, or lipid by the muscle. The myopathies caused by glycogen and glucose utilization are categorized according to the sequence of the enzyme defects along the glycogenolytic or glycolytic pathways [*see Figure 2*] and characterized by two main presentations: (1) exercise intolerance, myalgia, cramps, and, finally, myoglobinuria; and (2) fixed, progressive weakness.

GLYCOGENOSES

Muscle Phosphorylase Deficiency

Muscle phosphorylase deficiency (also known as McArdle disease) is the prototypical glycogenosis: glycogen breakdown is inhibited, which leads to pyruvate shortage and impaired energy output [*see Figure 2*]. It is the second most common cause of recurrent myoglobinuria after carnitine palmitoyltransferase (CPT) deficiency.

Diagnosis This autosomal recessive disease presents as exercise intolerance and myoglobinuria in patients older than 15 years. If patients rest briefly after exercise-induced myalgia and stiffness, they can resume activity with better endurance (second-wind phenomenon), owing to increased mobilization and utilization of FFAs and glucose. Fixed muscle weakness may develop later in life. The resting serum CK level is often elevated. The inability to produce venous lactate after exercise has been traditionally examined with the ischemic forearm exercise test. This test, however, is falling out of favor as a diagnostic tool because it produces false positive results, is not specific, can be painful, and may result in focal muscle damage. A nonischemic exercise test is of equal diagnostic value and does not have the drawbacks of the ischemic test. Muscle biopsy shows an absence of phosphorylase, the presence of subsarcolemmal vacuoles, and increased glycogen accumulation. The diagnosis is confirmed by biochemical analysis of muscle and molecular analysis of blood cells. The defect is caused by mutations in the muscle isoform of phosphorylase on chromosome 11q13 and can be detected in the leukocytes in more than 90% of patients.

Treatment No treatment is available for McArdle disease, but aerobic exercise training and a high-protein diet can be helpful. Vitamin B_6 supplementation has also been reported to be helpful. The results of a controlled trial indicate that creatine supplementation may improve muscle function.[56] Most encouraging was the result from a recent study showing that 75 g of sucrose before exercise markedly improved exercise tolerance and may protect against exercise-induced rhabdomyolysis.[57] Experimentations with gene therapy are ongoing.

Phosphofructokinase Deficiency

Phosphofructokinase (PFK) is an enzyme with three genetically distinct structural subunits: M, expressed in muscle, heart, and brain; L, expressed in liver and erythrocytes; and P, expressed in platelets. PFK deficiency is an autosomal recessive disease. Distinct mutations in the M subunit, localized to chro-

mosome 1, cause myopathic symptoms and chronic hemolysis and an increased serum bilirubin level and reticulocyte count.[58]

Diagnosis Because PFK deficiency is a glycolytic defect [*see Figure 2*], the functional consequences of PFK deficiency are similar to those observed in McArdle disease. PFK deficiency should be suspected in patients who experience exercise intolerance, nausea, and myoglobinuria. Fixed muscle weakness may develop later in life. A long history of mild, compensated hemolysis, a high reticulocyte count, a high bilirubin level, and hyperuricemia also indicate PFK deficiency, especially in certain ethnic groups, such as Japanese and Ashkenazi Jews. The diagnosis is confirmed by biochemical studies on muscle and molecular analysis of blood cells.[58]

Treatment There is no specific treatment for PFK deficiency. Patients should avoid high-carbohydrate meals. Recently, a ketogenic diet has been advocated.

Phosphoglycerate Kinase Deficiency

Phosphoglycerate kinase (PGK) deficiency is a rare X-linked recessive disorder that presents with exercise intolerance, episodes of myoglobinuria, hemolytic anemia, myopathy, and, occasionally, mild mental retardation[54,55] [*see Figure 2*].

Phosphoglycerate Mutase Deficiency

Phosphoglycerate mutase (PGAM) deficiency is a very rare disease that affects only muscle. In the United States it has only been identified in African Americans [*see Figure 2*].

Acid Maltase Deficiency

Acid maltase deficiency (AMD) is an autosomal recessive glycogen storage disease caused by deficiency of α-glycosidase, an enzyme encoded in a gene localized to chromosome 17q23.[55] Mutations or small deletions that cause abnormal splicing affect α-glycosidase expression.[59,60]

There are three clinical forms of AMD: infantile, childhood, and adult.[59,60] The infantile form (Pompe disease) presents within the first few months of life as hypotonia, weakness, and enlargement of the heart, tongue, and liver; respiratory and cardiovascular changes lead to death before 2 years of age.

In the childhood form of AMD, patients present with a myopathy characterized by delayed motor milestones, proximal muscle weakness, respiratory muscle involvement, and calf enlargement. The disease leads to death by the second decade of life.

The adult form of AMD manifests in persons older than 20 years as a proximal muscle weakness that resembles polymyositis or limb-girdle dystrophy. Respiratory muscle weakness may be the presenting symptom in one third of AMD cases in adults. Accumulation of glycogen occurs predominantly in the muscle. However, the enzyme is deficient in muscle, liver, heart, and cultured fibroblasts. Patients have elevated serum CK levels; EMG shows prominent myotonic discharges (without clinical myotonia), especially in the paraspinal muscles; and the muscle biopsy shows multiple vacuoles with high glycogen concentrations that react strongly with acid phosphatase, indicative of increased lysosomal activity. Because the utilization of glycogen and glucose is not compromised [*see Figure 2*], AMD causes fixed weakness but not exercise intolerance or myoglobinuria. The muscle fiber undergoes an autophagic process because of abnormalities in the lysosomes.

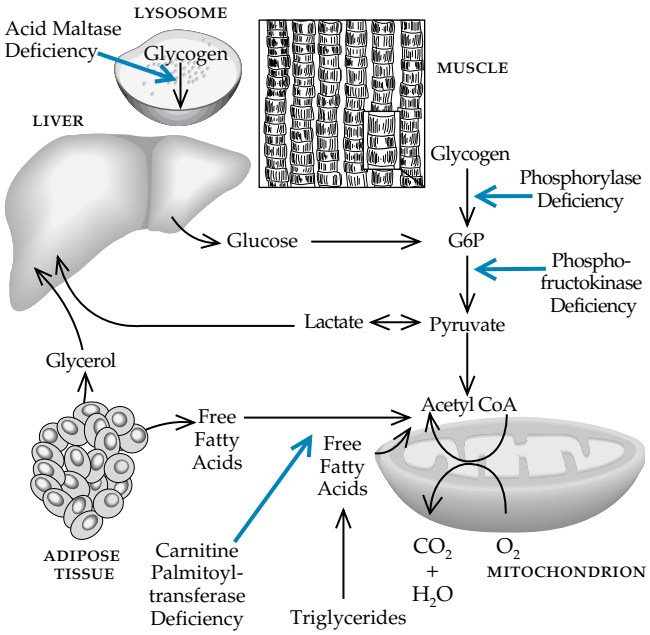


Figure 2 **Scheme of glycogen metabolism, glycolysis, and utilization of fatty acids. Some of the most common myopathies are those that result from deficiencies in the enzymes acid maltase, muscle phosphorylase, phosphofructokinase, and carnitine palmitoyltransferase.**

The diagnosis is confirmed by biochemical assay of the enzyme level in the muscle, cultured fibroblasts, lymphocytes, or urine. Molecular genetic analysis is now available. There is no specific treatment.

Other Rare Glycogenoses

Rare muscle glycogenoses include disorders caused by a deficiency of specific muscle enzymes; these include lactate dehydrogenase (LDH), β-enolase deficiency, debrancher enzyme deficiency, brancher enzyme deficiency, and aldolase deficiency.[55,56]

LIPID STORAGE MYOPATHIES

During sustained exercise, long-chain fatty acids (LCFAs) are the main energy source for the muscle. LCFAs are derived from food or, during fasting conditions, from adipose tissue. LCFAs first need to be transported to the mitochondria for oxidation; their transfer across the inner mitochondrial membrane requires L-carnitine and two enzymes, the carnitine palmitoyltransferases (CPT I and CPT II), which are located in the outer and the inner mitochondrial membranes, respectively. Inside the mitochondria, α-oxidation is facilitated first by the acyl coenzyme A (acyl-CoA) dehydrogenases and then by the transfer of electrons through flavoproteins to the respiratory chain proteins.[54,55] Lipid storage myopathies are caused by impaired fatty acid oxidation by the mitochondria, which results from defects in (1) carnitine and CPT, impairing the transport of fatty acids across the mitochondrial membrane, (2) the enzymes associated with β-oxidation, and (3) the respiratory chain proteins and electron-transferring flavoproteins.[54]

Carnitine Deficiency

Carnitine is mostly derived from the diet, but 25% is synthesized in the liver from lysine and methionine. Carnitine is crucial for the oxidation of LCFAs.[54] The burden of carnitine deficiency

is dysfunction of the liver, heart, and muscle tissues, which are highly dependent on LCFA oxidation.

Primary carnitine deficiency (PCD) is an autosomal recessive disorder caused by mutations in the *SLC22A5* gene, which encodes the sodium ion-dependent cation transporter–2 (OCTN2). These mutations cause a deficiency in the number of functional high-affinity carnitine receptors, which results in defects in carnitine transport across cell membranes.[54] PCD is an uncommon disorder seen more often in childhood.

The most common causes of carnitine deficiency are secondary. They result from (1) defective β-oxidation, which is associated with organic acidurias; (2) mitochondrial dysfunction and defects in the respiratory chain proteins; (3) renal disease, such as Fanconi syndrome, nephropathic cystinosis, or chronic hemodialysis; and (4) treatment with drugs, especially zidovudine (AZT) and valproate.[61]

Diagnosis Patients with PCD experience progressive cardiomyopathy, episodes of hypoketotic hypoglycemia (because of hepatic dysfunction), and proximal myopathic weakness. Lipids accumulate in the muscle, forming small lipid droplets.

Treatment Carnitine supplementation has produced variable treatment results.

Carnitine Palmitoyltransferase Deficiency

In infants, CPT I deficiency manifests as Reye syndrome, with hepatic encephalopathy, hypoketotic hypoglycemia, and hyperammonemia. In adults, carnitine deficiency syndrome most often results from CPT II deficiency, caused by mutations in the gene for CPT II, located on chromosome 11p11-p13.[55] CPT deficiency represents the most common cause of myoglobinuria in young adults. Patients present with attacks of muscle stiffness, cramps, myalgia, and myoglobinuria hours after prolonged or sustained exercise, especially after fasting or when muscle energy depends on utilization of LCFAs and not on the utilization of glycogen or glucose. CPT II–deficient patients do not have reduced exercise tolerance, second-wind phenomena, or warning signs of myalgia that prevent them from further exercise. Between attacks, muscle strength and the serum CK level are normal. Diagnosis is established by measuring CPT II enzyme activity in the muscle or by genetic testing.

It is unclear why CPT deficiency causes intermittent attacks of myoglobinuria and why lipid does not accumulate in the muscle. There is no therapy to prevent myoglobinuric attacks. However, a high-carbohydrate and low-fat diet, frequent meals, and extra carbohydrate intake before and during sustained exercise are recommended.

Mitochondrial Myopathies and Encephalopathies

Mitochondrial myopathies and encephalopathies constitute a diverse group of disorders that affect not only muscle and the nervous system but also other organs. These disorders are characterized by a primary defect in mitochondrial energy output. Genetic defects of mitochondrial energy enzymes may be caused by either nuclear DNA genetic mutations or mitochondrial DNA (mtDNA) mutations. Three types of mutation are responsible for this varied group of diseases: (1) sporadic mtDNA mutations, which cause large-scale deletions of mtDNA, are responsible for multisystem disorders (e.g., Kearns-Sayre syndrome) that can affect the heart, brain, endocrine system, and gastrointestinal tract; (2) maternally inherited mtDNA point mutations, which affect the brain and muscle, cause such disorders as MELAS syndrome (mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes), MERRF syndrome (myoclonic epilepsy and myopathy with ragged-red fibers), and Leber hereditary optic neuropathy; and (3) tissue-specific depletion of mtDNA causes a multiorgan syndrome that can affect muscle, liver, kidney, and brain.

PATHOGENESIS

The main function of mitochondria is to generate energy for the cell by producing adenosine triphosphate (ATP) through oxidative phosphorylation (OXPHOS). OXPHOS relies on five enzymatic complexes, which include coenzymes and transition metal components (iron, copper), located in the inner mitochondrial membrane. These complexes (designated as I, II, III, IV [cytochrome-c oxidase], and V) sequentially collect and transfer electrons, derived from the catabolism of fats, proteins, and carbohydrates, to O_2. The coupling of oxidation and phosphorylation, which occurs via a proton gradient across the inner mitochondrial membrane, enables the phosphorylation of adenosine diphosphate (ADP) to produce ATP.[62-65]

Mitochondria contain their own extrachromosomal DNA (mtDNA), which is distinct from nuclear DNA. It is a double-stranded, circular molecule that encodes 24 structural RNAs, 2 ribosomal RNAs (rRNAs), 22 transfer RNAs (tRNAs), and 13 mRNAs. The mRNAs encode several polypeptides of the respiratory chain. The remaining subunits of OXPHOS and other mitochondrial proteins are encoded by nuclear genes. The organization of mtDNA is highly compact; it has no introns. As a result, random mutations in mtDNA usually strike a coding sequence and frequently cause disease. In addition, mtDNA is susceptible to oxygen-radical damage because of its proximity to oxygen-radical production by OXPHOS and because of its minimal repair mechanisms.[62-65] The entire mtDNA of each person is exclusively maternally inherited (nonmendelian inheritance) because the sperm cell contributes only its nuclear DNA to the zygote during fertilization. Occasionally, diseases of OXPHOS may occur as the result of mutations in some of the nuclear-encoded OXPHOS genes; such diseases follow a mendelian inheritance pattern.

Muscle biopsy results of patients with OXPHOS defects are abnormal and reveal ragged-red fibers on trichrome stain, or ragged-blue fibers on succinate dehydrogenase stain. These findings result from the accumulation of mitochondria in the periphery of the muscle fibers, and accumulation of fibers that are negative for cytochrome-c oxidase. On electron microscopy, the mitochondria have paracrystalline inclusions or abnormal cristae.[44] Specific mutations, deletions, or depletions are detected by study of the mtDNA.

SPORADIC MTDNA DELETIONS

Kearns-Sayre Syndrome

Kearns-Sayre syndrome presents in patients younger than 20 years as ophthalmoplegia, ptosis, retinitis pigmentosa, and myopathic weakness. Short stature, cardiac conduction defects, elevated levels of cerebrospinal fluid protein, cerebellar syndromes, sensorineural hearing loss, and elevated serum lactate levels are common. Muscle biopsy reveals ragged-red fibers. In persons older than 20 years in whom ophthalmoplegia is the predominant phenotype, KSS is classified as chronic progressive external ophthalmoplegia (CPEO).[62-65]

KSS and the more limited CPEO are characterized by a single, large mtDNA deletion that is between nine and 50 base pairs. Most often, the deletion occurs sporadically and is rarely maternally inherited. OXPHOS is defective, and the levels of activity of complex I and complex IV are reduced. Variants of CPEO, which are characterized by multiple mtDNA deletions in the nuclear DNA–encoded genes of OXPHOS, may be transmitted in an autosomal dominant or recessive fashion.

Mitochondrial Neurogastrointestinal Encephalomyopathy

A special form of autosomal recessive CPEO is a multisystem syndrome known as MNGIE (mitochondrial neurogastrointestinal encephalomyopathy). Disease can occur in persons 20 to 60 years of age. Patients present with progressive external ophthalmoplegia accompanied by intestinal dysfunction, peripheral neuropathy, and leukoencephalopathy. Patients with MNGIE have increased serum levels of thymidine and decreased activity of thymidine phosphorylase on leukocytes; these findings, along with mutations in the thymidine phosphorylase nuclear gene, which impair the replication and repair of mtDNA, confirm the diagnosis.

MATERNALLY INHERITED MTDNA POINT MUTATIONS

MELAS Syndrome

Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes combine to form the MELAS syndrome. Muscle biopsies in affected patients reveal ragged-red fibers. Patients may also have hearing loss, short stature, cardiomyopathy, diabetes, or pigmentary retinal degenerations resembling KSS or CPEO. As many as 80% of patients have mtDNA point mutations in the tRNA leucine gene. These mutations, like the other mtDNA mutations, are heteroplasmic, meaning that normal and mutant mtDNA coexist in a cell. A cell normally contains between two and 10 mtDNA molecules, which allows an otherwise lethal mutation (i.e., lethal impairment of OXPHOS) to persist in viable organisms.

MERRF Syndrome

The MERRF syndrome consists of myoclonic epilepsy and myopathy with ragged-red fibers. In addition, ataxia, dementia, deafness, weakness, wasting, and cardiac abnormalities are common, although their expression is variable and depends on the degree of heteroplasmy. Approximately 80% of MERRF patients have a mutation in the tRNA lysine gene.[48]

Leber Hereditary Optic Neuropathy

Leber hereditary optic neuropathy (LHON) is the most common cause of blindness in young adults (the prevalence is higher in men than in women). Patients present with painless, subacute, and bilateral vision loss. In at least 90% of the families, several mtDNA mutations have been observed. Some patients with LHON and distinct mutations may have other associated conditions, such as encephalopathy, deafness, ataxia, myelopathy, or dystonia.[62-65]

MTDNA DEPLETION SYNDROME

mtDNA depletion syndrome is an autosomal recessive disorder that is usually fatal by 3 years of age. It affects the muscle, liver, kidneys, and brain; death is caused by encephalopathy or respiratory failure. The mitochondrial defect is quantitative rather than qualitative. The disorder results from imbalances in the mitochondrial nucleotide pool which, in turn, impair mtDNA replication and repair.

Treatment of all the mitochondrial disorders is supportive and symptomatic. Ubiquinone, creatine, vitamins C and K_3, carnitine, and vitamin E are often used. Moderate exercises are recommended.

Ion Channelopathies, Periodic Paralyses, and Nondystrophic Myotonias

Ion channelopathies that affect muscle-fiber excitability produce a range of disorders that include periodic paralysis, myotonia, and episodic ataxia-myokymia. These represent a rare group of disorders that usually start in childhood and typically present with attacks of paralysis. During paralytic attacks, alterations in the level of serum potassium are common. Myotonia is also common in some forms. The disorders are suspected when there is a history of similar attacks in family members, especially if attacks are provoked by rest after exercise or certain carbohydrate-enriched meals; the diagnosis is confirmed by genetic analysis of blood DNA.

PATHOGENESIS

After excitation at the neuromuscular junction, action potentials are propagated through ion fluxes across the sarcolemmal membrane, which depend on the opening (activation) and closing (inactivation) of the appropriate ion channel. In muscle and nerve cells, opening of the voltage-gated Na^+ channel causes a rapid increase in Na^+ permeability and hence membrane depolarization. For the membrane to initiate the next action potential, however, the Na^+ channels must close.[66-68] The voltage-gated K^+ channels open and K^+ ions flow out of the cell, creating a hyperpolarized voltage across the cell membrane. The Cl^- channels contribute to repolarization by stabilizing the membrane potential.

Disturbances in membrane excitability can lead to myotonia with or without periodic paralysis. Myotonia, as a symptom in nondystrophic myotonias, occurs in sodium or chloride channel disorders. Myotonia manifests as painless stiffness after a period of inactivity, which improves after continuous movements (warm-up phenomenon). Myotonia that develops after exposure to cold and worsens with exercise is called paradoxical myotonia. The other manifestation of membrane excitability is periodic paralysis, characterized by paralytic attacks associated with either hyperkalemia caused by mutations in the Na^+ channel genes or hypokalemia caused by mutations in the voltage-gated calcium channel gene. Patients with cardiodysrhythmic periodic paralysis (Anderson syndrome) and those with episodic ataxia-myokymia have mutations in the potassium channel gene.

SODIUM CHANNEL DISORDERS

In the skeletal muscle, sodium channel disorders result from mutations in *SCN4A*, the sodium channel gene. Mutations in *SCN4A* can produce the clinical phenotypes of hyperkalemic and normokalemic periodic paralysis (hyperKPP), paramyotonia congenita, and potassium-aggravated myotonia (previously called acetazolamide-responsive myotonia or myotonia fluctuans). Patients with these allelic disorders demonstrate varying degrees of myotonia of eye closure, chewing, swallowing, and gripping of the hands.

HyperKPP and paramyotonia congenita (also called paradoxical myotonia) are autosomal dominant diseases. They are characterized by attacks of weakness that begin in infancy or early childhood. HyperKPP attacks are precipitated by rest after exercise, stress, administration of K^+, cold, and certain foods. Mild myotonia may be present, and paradoxical myotonia of the eyelids is common. Patients with paramyotonia congenita develop paradoxical myotonia, which worsens with repetitive exercise; when exposed to cold, they develop stiffness, especially in the face, tongue, eyelids, and hands. Episodic attacks of weakness are also common, resembling those seen with hyperKPP, and they are accompanied by myotonia. The symptoms in patients with paramyotonia congenita, including the frequency of attacks and the interattack weakness, are treated with carbonic anhydrase inhibitors, such as acetazolamide and dichlorphenamide.[69]

POTASSIUM CHANNEL DISORDERS

Point mutations in the potassium channel gene *KCNA1,* on chromosome 12, have been associated with episodic ataxia-myokymia. Mutations in other potassium channel genes cause long QT syndrome. Andersen syndrome is caused by mutations in potassium channel gene *KCNJ2*, which encodes the inward-rectifying K^+ channel gene, *Kir21*, on chromosome 17q. Andersen syndrome is characterized by periodic paralysis, which may be accompanied by fixed weakness; long QT syndrome with cardiac ventricular arrhythmias; and dysmorphic craniofacial features, such as micrognathia, low-set ears, short stature, and syndactyly. Anderson syndrome, as well as the other potassium channel disorders, usually responds to carbonic anhydrase inhibitors.

CALCIUM CHANNEL DISORDERS

Calcium channel disorders present as hypokalemic periodic paralysis (hypoKPP). This is an autosomal dominant disease localized to chromosome 1q31-q32 and caused by mutations in the dihydropyridine-sensitive L-type calcium channel gene.[70] HypoKPP presents as episodic paralytic attacks that begin in the first or second decade of life, usually during sleep. They are precipitated by high carbohydrate consumption, rest after exercise, and excitement. During attacks, the serum K^+ concentration decreases and urinary retention of Na^+ and water occurs. Between attacks, many patients develop a permanent, slowly progressive myopathy.

The disease should be distinguished from all the causes of secondary hypokalemias that can cause weakness, including diuretic use, kidney disease, hyperaldosteronism, licorice intoxication, laxative abuse, and potassium-losing GI diseases. In secondary hypokalemias, the paralytic attacks do not occur unless the K^+ concentration falls below 3 mEq (usually below 2.5 mEq). In contrast, in hypoKPP the attacks occur even when the serum K^+ concentration is closer to the normal level.

HypoKPP can be associated with thyrotoxicosis, especially in persons of Asian descent. In such cases, hypoKPP responds to propranolol or resolves after reversion of the patient to a euthyroid state. Although hypoKPP is dominantly inherited, one third of patients may have sporadic disease. A provocation test, with insulin and glucose infusion under cardiac and respiratory monitoring, was necessary for the diagnosis until genetic identification became available. The frequency of the attacks and the interattack weakness respond to carbonic anhydrase inhibitor, acetazolamide, and dichlorphenamide.[71,72]

CHLORIDE CHANNEL DISORDERS

Chloride channel disorders include autosomal dominant myotonia congenita (Thomsen disease), which occurs in the first decade of life and is often associated with muscle hypertrophy, and autosomal recessive myotonia congenita (Becker-type myotonia) (BMD), which comes later in life and can be more severe.[67,68] Both of these diseases have been linked to distinct mutations on the chloride channel gene on chromosome 7q32, designated as the *CLCN1* locus.[66-68] Patients with these disorders do not have periodic paralysis or weakness, although in Becker-type myotonia, transient but sometimes severe weakness may occur; such weakness improves with exercise. Typical percussion myotonia and generalized myotonia, experienced by the patient as stiffness and hypertrophy of the legs and buttocks, are characteristic findings. Myotonic stiffness improves with exercise (warm-up phenomenon). The myotonic symptoms may variably respond to phenytoin, mexiletine, or tocainide.

ISOLATED NEUROMYOTONIA (ISAAC DISEASE)

Autoimmune acquired neuromyotonia occurs as an isolated neuromyotonia (also called Isaac disease) and as neuromyotonia associated with neuropathy, myasthenia gravis, and thymoma [*see* Neuromyotonia with K^+ Channel Autoantibodies (Isaac Syndrome), *below*]. The isolated form of the disease is characterized by (1) myokymia, which is a rippling muscle twitching that occurs during rest; (2) impaired muscle relaxation and stiffness at rest; (3) painful cramps; and (4) increased sweating. It is caused by hyperactivity of the peripheral motor nerve endings, and it is correctly called neuromyotonia because the continuous muscle fiber activity is abolished by curare but not by proximal nerve block. Neuromyotonia occurs sporadically, but familial cases have occurred.[73,74] Electromyography confirms the presence of neuromyotonia, and determination of K^+ channel autoantibodies identifies the autoimmune nature of the disease.

Treatment is symptomatic (phenytoin, carbamazepine, and mexiletine are used); in resistant cases, immunotherapy with intravenous immunoglobulin or plasmapheresis may be necessary.

Drug-Induced Toxic Myopathies

Although drug-induced myopathies are not uncommon, their diagnosis may be elusive in clinical practice. The clinician should suspect a toxic myopathy in a patient who does not have a preexisting muscle disease, whose myopathy develops acutely or subacutely (sometimes slowly) with muscle pain and weakness, who manifests myoglobinuria after the administration of a known or putative myotoxic agent, or who improves upon withdrawal of a putative toxic drug [*see Table 2*].[75,76] Myotoxic agents can cause myopathy by directly affecting a muscle organelle, such as the mitochondrion or thick filaments; by altering muscle antigens, thereby inducing an immunologic inflammatory reaction; or by inducing such secondary systemic effects as electrolyte disturbance, nutritional deprivation (e.g., the agent competes with vitamins), and malabsorption. Several drugs can be myotoxic either alone or in combination with other drugs. In addition to knowing which drugs are myotoxic, the clinician should always be alert to the potential myotoxicity of newly marketed drugs. The most common myotoxic drugs are discussed in the following sections.

ZIDOVUDINE MYOPATHY AND TOXICITY OF OTHER NUCLEOSIDE ANALOGUES

The clinical features of zidovudine myopathy are proximal muscle weakness, myalgia (predominantly in the thighs and

calves), fatigue, myopathic changes on EMG, and elevated serum CK levels, which often increase with exercise.[61,77-79] Weight loss and elevation of the serum lactate level may herald the onset of zidovudine myopathy. Zidovudine can cause myopathic symptoms after 1 year of administration. Symptoms resolve 4 to 6 weeks after zidovudine is discontinued.[78,79]

The unique histologic features of zidovudine myopathy are ragged-red fibers containing lipid accumulation and numerous cytochrome-c oxidase–negative fibers; both findings are suggestive of mitochondrial abnormalities.[61,78,79] Zidovudine is a DNA-chain terminator that inhibits the γ-DNA polymerase in the mitochondrial matrix; termination of mtDNA synthesis then occurs, and as much as a 78% depletion of muscle mtDNA results.[78,79] Among the other nucleoside analogues used in the treatment of AIDS, stavudine may also be myotoxic. The others, such as dideoxycytosine, dideoxyinosine, and lamivudine, cause a painful axonal neuropathy but not myopathy. A syndrome of lipodystrophy, lactic acidosis, and myopathy has been seen with highly active antiretroviral therapy consisting of one of the newly introduced protease inhibitors in combination with two nucleoside analogues, especially stavudine (d4T).[80]

MYOPATHY CAUSED BY CHOLESTEROL-LOWERING AGENTS

Several cholesterol-lowering drugs [*see Table 2*] cause an often reversible myopathy that is characterized by proximal muscle weakness, myalgia, elevation of the serum CK level, and myopathic changes on EMG. Because cholesterol is the major sterol constituent of muscle membranes, reduction of the normal cholesterol pool available for membrane synthesis can increase membrane fluidity, leading to an unstable sarcolemma, myotonic discharges, and, in advanced cases, rhabdomyolysis. Statins also inhibit the production of ubiquinone, thereby interfering with mitochondrial ATP and energy metabolism within the myocyte. EMG findings include fibrillation potentials and myotonic or complex repetitive discharges.[76,77,81,82] Gradual recovery can occur after withdrawal of the offending drug.

Although a transient increase in the serum CK level and myalgia is not unusual after treatment with statins, the drugs have been associated with clinically overt myopathic symptoms in fewer than 5% of treated patients. Coadministration of cyclosporine and lovastatin to patients with heart or kidney transplants and hyperlipidemia increased the incidence of myopathy with rhabdomyolysis. In a recent review, cerivastatin was the statin most commonly associated with myopathic symptoms; this was followed sequentially by lovastatin, fluvastatin, atorvastatin, simvastatin, and pravastatin.[81,82] The risk for statin-induced myopathy is increased with use of higher doses, statins that have a lipophilic action (which includes all except pravastatin and rosuvastatin), and concomitant therapy with drugs such as gemfibrozil, colchicine, or cyclosporine; elderly patients are at increased risk for myopathy when treated with statins.[81,82]

MYOPATHY ASSOCIATED WITH CRITICAL ILLNESS

In patients with prolonged paralysis induced by such nondepolarizing blocking agents as pancuronium, an acute myopathy may develop after mechanical ventilation is discontinued. Most of these patients received high doses of corticosteroids for status asthmaticus or another systemic illness for which artificial paralysis was induced to secure an aggressive pulmonary toilet.[77,82,83] The combination of blocking agents and corticosteroids has been consistently implicated in the development of critical illness myopathy, but sometimes patients have had only minimal exposure to one of these two agents.[77,82-85]

Table 2 Myopathy-Inducing Agents

- *Prescribed Medications and Vitamins*
  - Amiodarone
  - Amphetamines
  - ε-Aminocaproic acid
  - Chloroquine
  - Colchicine
  - Cyclosporine
  - Emetine
  - Fialuridine
  - Hypocholesterolemic drugs
    - 20,25-Diazacholesterol
    - Fibric acid derivatives
      - Bezafibrate
      - Biclofibrate
      - Clofibrate
      - Etofibrate
      - Fenofibrate
      - Gemfibrozil
    - 3-Hydroxy-3-methylglutaryl coenzyme A inhibitors
      - Lovastatin
      - Pravastatin
      - Simvastatin
    - Nicotinic acid (niacin)
  - Interferon alfa
  - Isotretinoin
  - Labetalol
  - Minocycline
  - Neuroleptics (in the absence of neuroleptic malignant syndrome)
    - Clozapine
    - Risperidone
    - Olanzapine
    - Haloperidol
    - Loxapine
    - Melperone
  - D-Penicillamine
  - Pancuronium
  - Procainamide
  - Steroids
    - Prednisone
    - Dexamethasone
  - Vincristine
  - Vitamin E
  - Zidovudine (AZT)
- *Nonprescription Drugs*
  - Cocaine
  - Ethanol
  - Heroin
  - Phencyclidine (PCP)
- *Intramuscular Injections*
  - Meperidine
  - Pentazocine
- *Other*
  - Diuretics
  - Laxatives
  - Licorice

The clinical presentation of myopathy associated with critical illness is characterized by severe generalized weakness of all extremities, failure to wean from mechanical ventilation, muscular atrophy, a normal or moderately elevated serum CK level, and myopathic changes on EMG. The weakness usually slowly improves.[77,82-85] Blocking agent–corticosteroid myopathy may coexist with an axonal neuropathy, but the consensus now is that the paralytic disease is predominantly caused by a myopathy. The differential diagnosis should include neuromuscular transmission defects, such as myasthenia gravis; Guillain-Barré syndrome; inflammatory or acute necrotizing myopathy; and periodic paralysis. Muscle biopsy shows extensive morphologic abnormalities, which result from the selective loss of thick filaments without signs of necrosis, inflammation, or phagocytosis. With adenosine triphosphatase (ATPase) staining, striking areas of central pallor are seen in many fibers.[77] Calpain expression is markedly enhanced, implicating an altered calcium homeostasis and enhanced proteolysis. Electron microscopy reveals selective and extensive loss of thick myofilaments; the thin myofilaments and Z disks are preserved.

High doses of steroids should be used with caution in patients receiving paralytic agents for prolonged periods. The myopathy improves slowly with intense rehabilitation.

VITAMIN E INTOXICATION

A necrotizing myopathy with proximal muscle weakness and elevation of the serum CK level has been reported to occur in pa-

tients self-medicated with high doses of vitamin E.[75,76,86] This condition is very rare and has not been clearly substantiated.

CHLOROQUINE MYOPATHY

The antimalarial drug chloroquine is often used by rheumatologists for the treatment of various collagen vascular diseases. It can cause macular and corneal degeneration, peripheral neuropathy, and myopathy. The myopathy is seen with long-term administration of high doses of chloroquine (500 mg daily). The myopathy has interesting morphologic features that resemble those that occur with acid maltase deficiency: multiple vacuoles with acid phosphatase-positive material, myeloid bodies within the vacuoles, and enlarged lysosomes with increased lysosomal enzyme activity.[75]

COLCHICINE MYOPATHY

Colchicine interferes with the growth of microtubules, thereby affecting mitosis. After long-term use, colchicine causes a vacuolar myopathy and an axonal neuropathy, often in patients who are between 50 and 70 years of age and who have gout and a mild, chronic renal insufficiency.[87] The concomitant use of statins increases the risk of myopathy. Symptoms include proximal muscle weakness, elevation of the serum CK level, distal sensory involvement, and areflexia. Symptoms resolve 4 to 6 weeks after discontinuance of the drug.

STEROID-INDUCED MYOPATHY

Patients with hyperadrenocorticism (i.e., Cushing syndrome) may develop weakness. Similar changes can occur during long-term administration of prednisone (usually at dosages greater than 20 mg daily) or dexamethasone, especially in poorly mobilized patients. The steroid-induced myopathic weakness is mild, spares the neck flexor muscles,[77] and may aggravate the weakness caused by the underlying immune disease or malignancy for which steroids are administered. Lowering the dose of the drug reverses the myopathy. The serum CK level is normal, the muscle biopsy shows only type II fiber atrophy, and the EMG is not informative.

MYOPATHY ASSOCIATED WITH ALCOHOLISM

Patients with alcoholism most often develop an axonal peripheral neuropathy and, occasionally, an acute or chronic myopathy. The acute myopathy presents with tightly swollen, painful, and tender muscles; it occurs in chronic alcoholics, frequently after a heavy bout of drinking. Acute myopathy can be localized (resembling thrombophlebitis) or generalized and severe (presenting as rhabdomyolysis and myoglobinuria).[75,82] It can recur if the patient resumes drinking. Acute myopathy in patients with alcoholism may also be related to hypokalemia when the serum K^+ concentration is below 2.5 mEq. This myopathy is painless, not accompanied by muscle swelling, and quickly reversible.

Proximal muscle weakness in patients with long-term alcoholism is often multifactorial and not necessarily a primary toxic effect of alcohol to the muscle; for example, poor nutrition, inactivity, or neurogenic disease may be involved.[58] Histologically, type II fiber atrophy is most common. Some long-term drinkers may experience an asymptomatic elevation of the serum CK level—as much as 20 times higher than normal levels—that is aggravated by physical activity.

DRUG-INDUCED INFLAMMATORY MYOPATHY

Patients with Wilson disease, rheumatoid arthritis, or progressive systemic sclerosis can develop polymyositis or myasthenia gravis during treatment with D-penicillamine. The disease responds to discontinuance of the drug or the administration of steroids. In single, unconfirmed reports, procainamide, propylthiouracil, and cimetidine have been reported to cause a polymyositis-like disease.

Contaminated L-tryptophan was responsible for an outbreak of a syndrome of eosinophilic fasciitis, myositis, thickening of the skin, axonal neuropathy, and other systemic manifestations. An immune process against fibroblasts in the extracellular matrix triggered by the contaminant was implicated in the cause of this syndrome, called the eosinophilia-myalgia syndrome.[88] The disease has left several patients with residual skin thickening, myalgia, and fatigue. Macrophagic myofasciitis is an entity encountered in France and resulting from aluminum-containing vaccines. Patients present with myalgia and fatigue. Biopsy at the sites of vaccines shows accumulation of macrophages in the epimysium, perimysium, and endomysium.

Interferon alfa (IFN-α), now commonly used for certain malignancies and hepatitis, causes fatigue, arthralgias, and, at times, myalgia. The exact mechanism of the muscle fatigue and myalgia is unknown, but rare cases of myositis and myasthenia-like disease have been noted. The cytokine interleukin-2, used for renal cell carcinoma, can exacerbate preexisting polymyositis or dermatomyositis.

Disorders of Neuromuscular Transmission

The main areas of the neuromuscular junction are (1) the presynaptic region, with active zones that contain parallel double rows of voltage-gated calcium channels (VGCCs) and synaptic vesicles that contain acetylcholine (ACh), (2) the synaptic space, and (3) the postsynaptic region, consisting of the junctional folds that contain the acetylcholine receptors (AChRs) [*see Figure 3*]. Functionally, neuromuscular transmission depends on the release of ACh from the motor nerve terminal, interaction of ACh with AChR, and the resulting depolarization of the muscle fiber. In a resting state, there is a random release of ACh from a single vesicle. ACh binds to AChR and opens its cation-selective ion channel, allowing Na^+ to enter the end-plate region; this creates a depolarization potential called the miniature end-plate potential.[89-91] ACh is then disassociated from AChR and rapidly hydrolyzed by acetylcholinesterase (AChE). When AChE is inhibited, as occurs during treatment with such anticholinesterase drugs as pyridostigmine, more ACh molecules bind to AChR and open multiple ion channels.

When the action potential reaches the nerve terminal, the VGCC opens and Ca^{2+} ions flow into the nerve terminal. The entry of Ca^{2+} triggers the fusion and exocytosis of the synaptic vesicles and the release of ACh molecules (each vesicle contains 6,000 to 10,000 ACh molecules) that diffuse and spread over the junctional folds. The simultaneous release of ACh by the fusion of 50 to 300 vesicles results in the end-plate potential (EPP). If the EPP exceeds a certain threshold, a muscle action potential (MAP) triggers muscle contraction. Normally, the interactions between the released ACh and ACR are more than sufficient to produce an EPP capable of triggering an MAP. The difference between the actual EPP amplitude and the EPP amplitude required to trigger an MAP is called the safety margin.

There are three main disorders affecting neuromuscular transmission: myasthenia gravis (MG), the Lambert-Eaton myasthenic syndrome (LEMS), and the various congenital myasthenic syndromes.

Myasthenia gravis (MG) is a prototypical autoimmune disease; it fulfills all the immunologic criteria [89-91]: the antigen is AChR; antibodies directed against AChR are detected and measured in the serum of affected patients; the disease can be transmitted to animals with the patient's pathogenic IgG; serum immunoglobulin binds to AChR, causing a complement-fixing destruction of the receptors, which affects neuromuscular transmission; and removal of the antibodies results in clinical improvement.

In MG, antibodies against AChR reduce the number of available receptors. Consequently, the released ACh cannot bind to enough AChR; thus, the generated EPP is insufficient to trigger an MAP, and neuromuscular transmission fails. When neuromuscular transmission failure occurs in many neuromuscular junctions, the muscle cannot be fully activated; this results in fatigue and muscle weakness. During repetitive nerve stimulation in healthy persons, the amount of ACh released is reduced after the first few impulses, because the nerve cannot sustain its original release rate; however, neuromuscular transmission does not fail, because the safety margin is sufficient. In patients with MG, the EPP is already small. During repetitive nerve stimulation, the EPP is depressed further because the decreasing amounts of ACh released per subsequent impulse are not sufficient to bind to the reduced number of AChR; this results in a further decrement of the MAP. Repetitive nerve stimulation studies are used diagnostically to demonstrate impaired neuromuscular transmission in MG patients. Because MG affects the nicotinic AChR at the neuromuscular junction, only motor systems are affected.

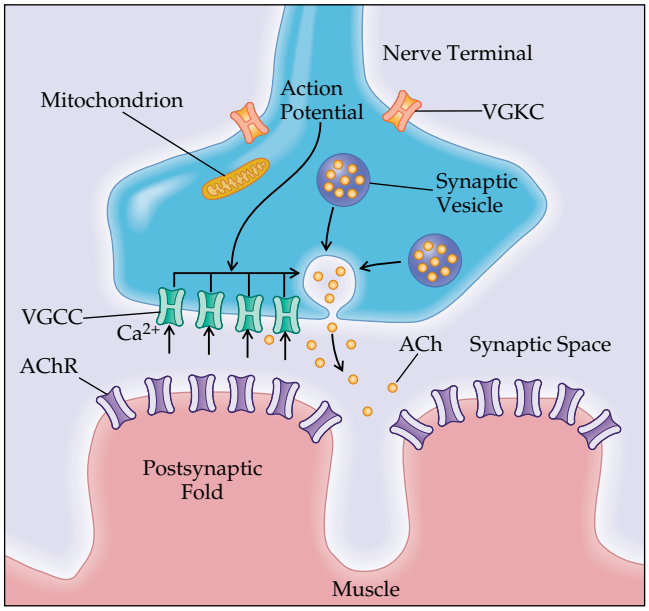


Figure 3 **Transmission from the nerve ending at the neuromuscular junction takes place at localized sites (active zones). The presynaptic region contains the voltage-gated calcium channels (VGCCs) and synaptic vesicles that hold 6,000 to 8,000 molecules of the neurotransmitter acetylcholine (ACh). When the action potential reaches the nerve terminal, the VGCCs open and Ca^{2+} flows into the nerve terminal. The entry of Ca^{2+} triggers the fusion and exocytosis of the synaptic vesicles. ACh is released into the synaptic space and binds to the acetylcholine receptor (AChR) located in the end-plate region of the muscle membrane (postsynaptic folds). Activation of AChR by ACh causes ions to flow through AChR across the membrane, initiating the electrical response of the muscle called the end-plate potential (EPP). The EPP spreads from the end-plate region to the surrounding muscle membrane and initiates the impulse response of the muscle that ultimately leads to muscle contraction.**

Epidemiology

The prevalence of MG is approximately four to six cases per 100,000 population. The annual incidence varies, but two peaks are recognized: one is associated with women in their second or third decade of life, and the other is mostly associated with men older than 60 years. The mean age at onset is 28 years in women and 42 years in men. There is recent documentation that MG is underdiagnosed in patients older than 70 years.[92] Classic autoimmune MG can also develop during childhood and responds well to thymectomy or anticholinesterases. MG can be associated with thymoma in as many as 15% of adult patients, or it may develop after removal of a thymoma. In addition, as many as 5% of patients with MG have other systemic autoimmune conditions, as many as 15% of patients have various autoantibodies, as many as 20% of patients have thyroid disease, and as many as 5% of patients have a family history of another autoimmune disease. Pregnancy has a variable effect, but for most patients, weakness increases in the postpartum period. About 12% of children born to mothers with MG develop transient neonatal MG as a result of transplacental transfer of circulating anti-AChR antibodies from the myasthenic mother to the fetus. These infants develop weakness, feeding difficulties, respiratory problems, a weak cry, and facial weakness within the first few hours after birth. The condition lasts as long as 3 weeks, coinciding with the half-life of IgG.

Immunopathogenesis

Autoantibodies MG is mediated by pathogenic autoantibodies that bind to AChR and cause a functional loss of the receptors and focal destruction of the postsynaptic junctional folds, interfering with the depolarization of the postsynaptic membrane [*see Figure 4*]. Between 10% and 15% of MG patients, especially those with mild, childhood, or localized (e.g., ocular) forms, do not have detectable antibodies. Most of these patients, although seronegative, have an antibody-mediated disorder. Among the 15% seronegative patients, about 6% have antibodies against MuSK,[93] a muscle-specific kinase needed for the development and clustering of the AChR at the neuromuscular junction [*see Figure 4*]; the other 9% of seronegative patients contain poorly characterized IgG antibodies that reversibly inhibit AChR function or IgM antibodies that indirectly inhibit AChR function.[94]

Other antibodies associated with MG include antistriational cell antibodies. These are found in as many as 80% of patients with thymoma and MG, 25% of patients with thymoma without MG, 30% of adults with MG (55% when MG begins after age 60), 6% of patients with LEMS, 3% of patients with lung cancer, and frequently in patients with autoimmune liver disease.[94] Anti-titin antibodies are also present in 55% of AChR-positive patients with late-onset MG, regardless of whether the patients have thymoma. Antibodies to thyroid microsomes and thyroglobulin are frequently present in patients with ocular MG and in patients with LEMS without cancer.

Immunoregulation The AChR is a T cell–dependent antigen. In MG, the $CD4^+$ cells are sensitized and respond to stimulation with AChR or to synthetic immunodominant peptides of AChR.[89-92] Because lymphocytes from normal, healthy persons also respond to such peptides, albeit to a lesser degree, the

AChR-specific T cells are part of the normal immune repertoire. This suggests that MG is a disease of abnormal immunoregulation. Induction of MG by microbial antigens (e.g., antigens to *Escherichia coli* and *Klebsiella*) or herpes simplex viruses have been postulated because these bacteria and viruses share sequence homology with peptides of the alpha subunit of AChR.

Approximately 75% of patients with MG have thymic abnormalities, either hyperplasia with germinal center formation or thymoma (15% of all patients with MG). Removal of the thymus results in clinical improvement or remission. Additional evidence implicating a role for the thymus in MG is based on several findings: (1) the thymus of an affected person contains a greater population of B cells that could secrete AChR antibodies than the thymus of a normal, healthy person; (2) the thymus of an affected person possesses myoid cells that contain AChR proximal to the interdigitating dendritic cells; and (3) the number of T cells sensitized to AChR is greater in the thymus of an affected person than in the thymus of a healthy person.

In MG, breaking of self-tolerance may begin in the thymus. AChR on the myoid cells may thus be recognized by the thymic dendritic cells that present the antigen to $CD4^+$ cells, which in turn stimulate the antibody-producing B cells to make anti-AChR antibodies. Subsequently, these antibodies recognize and cross-react with AChR in skeletal muscle.

MG is frequently associated with other autoimmune conditions, including collagen vascular diseases, polymyositis, pemphigus, autoimmune thyroid diseases, and graft versus host disease. The use of D-penicillamine has induced classic MG, which improves when the drug is discontinued.

Diagnosis

Clinical manifestations Patients with MG present with skeletal muscle weakness and fatigability. The weakness often affects elevation of the eyelids and movement of the extraocular muscles, which causes a characteristic asymmetrical ptosis and diplopia. The neck extensors are often weak, and the head droops. Patients with MG may have weakness of facial and bulbar muscles that causes a facial snarl when a patient smiles; nasal or dysarthric and low-volume dysphonic speech; and dysphagia, which can result in regurgitation and choking. Proximal skeletal muscle weakness and fatigue produce difficulty in walking, climbing steps, combing hair, and carrying objects. Respiratory muscle weakness can be significant. Symptoms fluctuate and are often better in the morning. Fatigability is usually worse at the end of the day or after repeated or continuous activity. If a patient is asked to keep the arms abducted, a gradual decline in arm height is noted. If a patient is asked to talk continually, the voice may become husky, nasal, slurred, and finally inaudible. Sensation and cognition are normal. Tendon reflexes, especially from nonatrophic muscles, tend to be brisk or normally active. The toes are downgoing. The symptoms worsen during or before the menstrual period and during viral or bacterial infections.

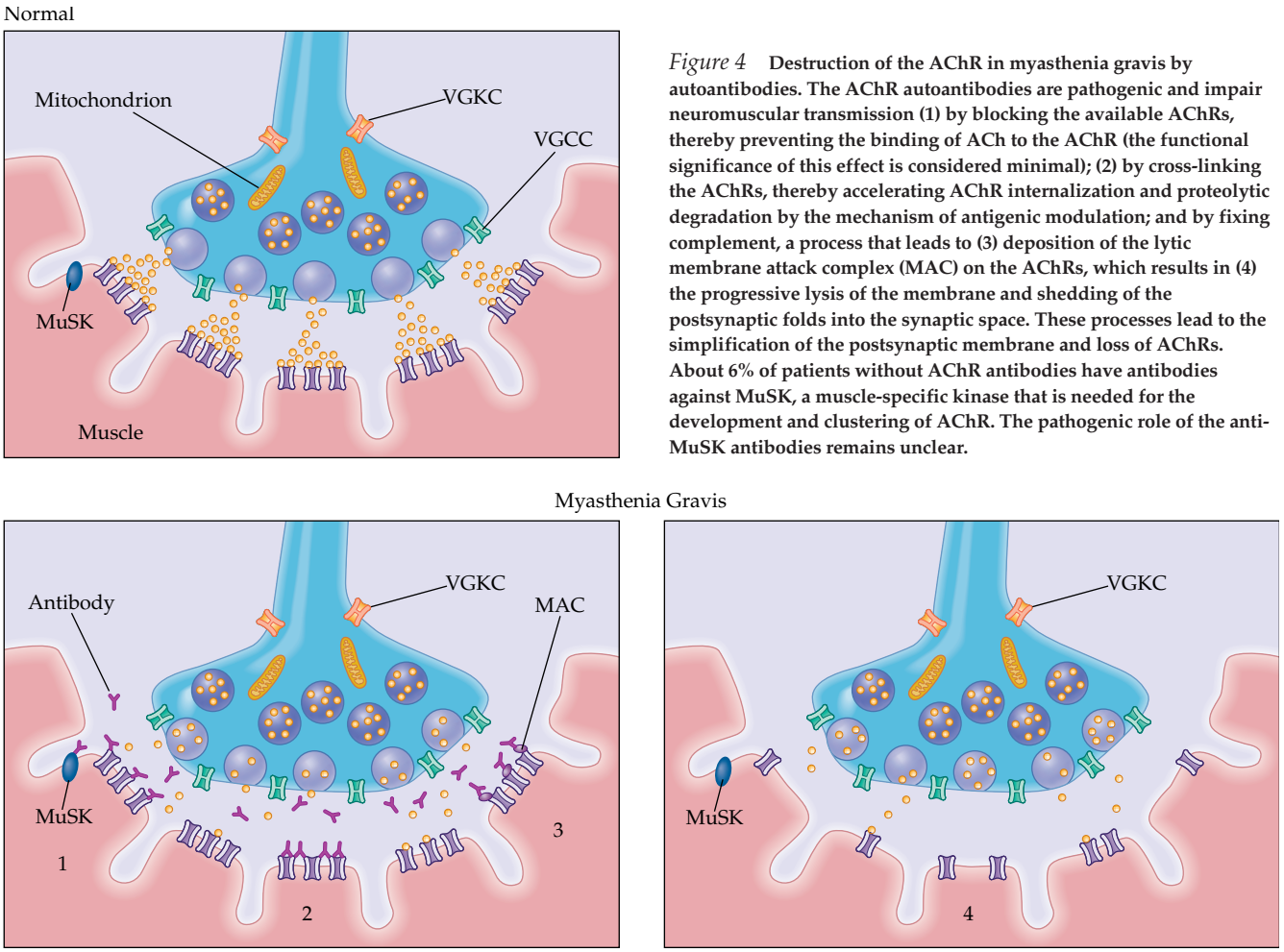


Figure 4 **Destruction of the AChR in myasthenia gravis by autoantibodies. The AChR autoantibodies are pathogenic and impair neuromuscular transmission (1) by blocking the available AChRs, thereby preventing the binding of ACh to the AChR (the functional significance of this effect is considered minimal); (2) by cross-linking the AChRs, thereby accelerating AChR internalization and proteolytic degradation by the mechanism of antigenic modulation; and by fixing complement, a process that leads to (3) deposition of the lytic membrane attack complex (MAC) on the AChRs, which results in (4) the progressive lysis of the membrane and shedding of the postsynaptic folds into the synaptic space. These processes lead to the simplification of the postsynaptic membrane and loss of AChRs. About 6% of patients without AChR antibodies have antibodies against MuSK, a muscle-specific kinase that is needed for the development and clustering of AChR. The pathogenic role of the anti-MuSK antibodies remains unclear.**

Patients with MuSK antibodies tend to be women; disease onset in these patients occurs at 20 to 60 years of age. Frequently, patients present with prominent neck, shoulder, or respiratory muscle weakness and bulbar signs.[95] Others present with predominantly oculobulbar signs.

Laboratory tests In addition to demonstration of fatigue and weakness on examination, the clinical diagnosis can be confirmed by an edrophonium or neostigmine test. These drugs inhibit AChE, which allows ACh to interact repeatedly with the remaining AChR, and as a result, muscle strength improves. Repetitive nerve stimulation studies demonstrate a rapid reduction (> 12%) in the amplitude of the MAP. Single-fiber EMG demonstrates increased jitter and blocking in more than 90% of patients. An MRI of the chest should be performed to search for thymic hyperplasia or thymoma. MG must be differentiated from diseases or exogenous agent–induced toxic disorders that have similar clinical pictures, such as botulism, organosphosphate poisoning, D-penicillamine toxic reaction, mitochondrial myopathy, compressive lesion affecting cranial nerves, LEMS, and congenital myasthenic syndromes. The serologic diagnosis of MG is made by detecting AChR or MuSK antibodies with a radioimmunoprecipitation assay.[90] The AChR antibodies are present in as many as 85% to 90% of patients with generalized MG and in as many as 70% of patients with ocular MG; they are also found in as many as 30% of patients with autoimmune liver disease, 10% of patients with pernicious anemia, and as many as 13% of patients with LEMS.[96] Generally, AChR antibody titers do not correlate with clinical severity.

Treatment

The term gravis is now a misnomer because MG responds fairly well to the available therapies. Although the sequence in which the various therapeutic modalities are applied may differ among physicians, the initial choice in mild cases is generally an anticholinesterase drug given every 4 hours while the patient is awake. Thymectomy is done in most centers through the transsternal approach in patients beyond puberty and up to 55 years of age. The beneficial effect of thymectomy, however, has been questioned and has necessitated a large multicenter study.[97] Thymoma should always be removed, followed by radiation therapy.

Prednisone is the first-line immunotherapeutic drug therapy. The preferred dosage is 60 to 80 mg as a single daily dose in the morning for an initial period of 3 to 4 weeks. Then, over an 8-week period, the dose on alternate days is gradually reduced by 10 mg a week (more often if side effects are severe) until the lowest dose that controls the disease is reached. Some physicians prefer to start with lower dosages of prednisone (i.e., 20 to 40 mg/day) and gradually increase the dosage to prevent an occasionally noted transient worsening. Plasmapheresis or intravenous immune globulin is generally reserved for crisis and severe cases and to strengthen a patient before thymectomy. Cyclosporine, mycophenolate, or azathioprine may be used as a second-line therapy to maintain remission after lowering the steroid dosage.[89,90] Azathioprine, for example, given orally in dosages of up to 3 mg/kg/day has been shown in a controlled study to be effective in maintaining remission of symptoms after more than 12 months of administration, permitting a reduction in steroid dose.

Patients with MuSK antibodies do not improve with thymectomy, and their response to cholinesterase inhibitors is variable. The response to immunotherapy is generally good but variable: some patients may develop permanent facial, pharyngeal, or tongue weakness and atrophy.

LAMBERT-EATON MYASTHENIC SYNDROME

In LEMS, specific antibodies against VGCCs cause antigenic modulation and depletion of VGCCs [*see Figure 4*]. This restricts Ca^{2+} ingress into the motor nerve terminal and reduces ACh release during a nerve impulse. The amount of ACh released can generate only a small EPP that cannot trigger an MAP; this results in neuromuscular transmission failure and muscle weakness. An important observation in LEMS is its frequent association with small cell lung cancer (SCLC), which also expresses VGCC.[90,91,98-100]

LEMS occurs more frequently in men than in women. Approximately 70% of men with LEMS have SCLC, as compared with 20% of women. LEMS may precede the detection of malignancy by up to 3 years. Malignancy is rare in persons younger than 40 years.

Patients present with proximal muscle weakness, increased fatigability, and transient ocular symptoms of ptosis or diplopia. Autonomic symptomatology, with dry mouth, dry eyes, sexual impotence, constipation, and abnormal pupillary reflexes, is frequent. Hyporeflexia is common. A typical sign is an increase in muscle strength and reflexes a few seconds after a sustained maximal effort. The diagnosis is confirmed by repetitive nerve stimulation studies that demonstrate a dramatic increase in the amplitude of the MAP upon tetanic stimulation. Single-fiber EMG shows increased jitter and blocking, which improve with faster stimulation rates. Using radiolabeled peptides from cone snail toxins, antibodies to P-type or Q-type calcium channels are detected in as many as 95% of LEMS patients.[90,91,99,100] Antibodies to N-type channels are seen in as many as 75% of LEMS patients, especially those who have lung cancer. A search for SCLC, especially in men older than 40 years, is an essential part of the patient workup. Antibodies to the P-type or Q-type VGCC are also found in patients with paraneoplastic cerebellar degeneration in both the serum and the CSF. These antibodies may interfere with calcium influx into cerebellar Purkinje cells and may be implicated in the patient's ataxia.

The presynaptic region of the motor end plates in patients with LEMS exhibits marked depletion of the active zone particles that represent the VGCCs. The VGCCs in LEMS are no longer in parallel rows but are aggregated as a result of cross-linking by the specific IgG antibodies. The patient's IgG [$F(ab')_2$ fraction] injected into mice transmits the electrophysiologic and morphologic changes associated with the disease. Complement is not involved in the process. SCLC expresses VGCC of the N, L, or P type. It is believed that tumor cells trigger autoantibodies against surface determinants, which cross-react with similar epitopes at the presynaptic regions of the neuromuscular junctions.

In LEMS associated with SCLC, the primary focus is to treat the tumor because the symptoms of LEMS improve as the cancer regresses. The first drug of choice in LEMS is 3,4-diaminopyridine. This drug prolongs the duration of the presynaptic action potential by blocking the outward K^+ currents; this allows increased Ca^{2+} entry into the nerve terminal and further release of ACh. The drug alleviates fatigue and weakness in LEMS patients, and it is well tolerated in dosages of 40 to 60 mg daily. Sometimes, pyridostigmine is added. Prednisone helps substantially. Azathioprine, plasmapheresis, and intravenous immune globulin are additional therapeutic approaches that are given as steroid-sparing agents or to improve the strength of patients not responding well to 3,4-diaminopyridine and prednisone.

NEUROMYOTONIA WITH K^+ CHANNEL AUTOANTIBODIES (ISAAC SYNDROME)

Isaac syndrome is a rare autoimmune disease presenting as isolated neuromyotonia [*see* Isolated Neuromyotonia (Isaac Disease), *above*] or in association with thymoma, MG, and CNS symptoms, such as limbic encephalitis, personality changes, hallucinations, and sleep disturbances.[74] Patients with neuromyotonia have myokymia, myotonia, and, often, muscle hypertrophy. Myokymic discharges and prolonged motor unit discharges are seen electrophysiologically. Other patients have cramp-fasciculation syndrome and peripheral neuropathy.

An antibody against *o*-dendrotoxin–sensitive K^+ channels is responsible for the disease.[73,74] Affected patients have antibodies against voltage-gated potassium channels (VGKCs). The MAP is generated by the opening of the VGCCs; inactivation of the VGKCs, followed by the opening of the potassium channels, leads to repolarization of the nerve terminal. In patients with antibodies to VGKC, the depolarization is prolonged, prompting the calcium channel to remain open longer, so that more calcium enters the nerve terminal and more quanta are released.

This autoimmune disorder responds to plasma exchange, intravenous immune globulin, or immunosuppressants.

CONGENITAL MYASTHENIC SYNDROMES

The congenital myasthenic syndromes are a heterogeneous group of syndromes that cause a failure of neuromuscular transmission that is not autoimmune, but, rather, is the result of genetic defects in various molecules, enzymes, or channels at the neuromuscular junction.

A congenital myasthenic syndrome is suspected when an infant, a child, or, occasionally, a young adult presents with fluctuating ptosis, fatigability, increased weakness on sustained exertion (common to all the neuromuscular transmission defects), mild or delayed motor milestones, absence of autoimmune MG, and a positive family history.[101]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Campbell KP: Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage. Cell 80:675, 1995

2. Cohn RD, Campbell KP: Molecular basis of muscular dystrophies. Muscle Nerve 23:1456, 2000

3. Durbeej M, Campbell KP: Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models. Curr Opin Genet Dev 12:349, 2002

4. Beggs AH, Koenig M, Boyce FM, et al: Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. Hum Genet 86:45, 1990

5. Mathews KD: Muscular dystrophy overview: genetics and diagnosis. Neurol Clin 21:795, 2003

6. Muntoni F, Torelli S, Ferlini A: Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol 2:731, 2003

7. Mendell JR, Buzin CH, Feng J, et al: Diagnosis of Duchenne dystrophy by enhanced detection of small mutations. Neurology 57:645, 2001

8. Flanigan KM, von Niederhausern A, Dunn DM, et al: Rapid direct sequence analysis of the dystrophin gene. Am J Hum Genet 72:931, 2003

9. Finsterer J, Stollberger C: The heart in human dystrophinopathies. Cardiology 99:1, 2003

10. Mendell JR, Moxley RT, Griggs RC, et al: Randomized, double-blind six-month trial of prednisone in Duchenne's muscular dystrophy. N Engl J Med 320:1592, 1989

11. Bonifati MD, Ruzza G, Bonometto P, et al: A multicenter, double-blind, randomized trial of deflazacort versus prednisone in Duchenne muscular dystrophy. Muscle Nerve 23:1344, 2000

12. Magali L, Lebacq J, Poortmans J, et al: Beneficial effects of creatine supplementation in dystrophic patients. Muscle Nerve 27:604, 2003

13. Mendell JR, Kissel JT, Amato AA, et al: Myoblast transfer in the treatment of Duchenne's muscular dystrophy. N Engl J Med 333:832, 1995

14. Barton-Davis ER, Cordier L, Shoturma DI, et al: Aminoglycoside antibiotics restore dystrophin function to skeletal muscles of mdx mice. J Clin Invest 104:367, 1999

15. Wagner KR, Hamed S, Hadley DW, et al: Gentamicin treatment of Duchenne and Becker muscular dystrophy due to nonsense mutations. Ann Neurol 49: 706, 2001

16. Gregorevic P, Chamberlain JS: Gene therapy for muscular dystrophy- a review of promising progress. Expert Opin Biol Ther 3: 803, 2003

17. Nagano A, Koga R, Ogawa M, et al: Emerin deficiency at the nuclear membrane in patients with Emery-Dreifuss muscular dystrophy. Nat Genet 12:254, 1996

18. Fatkin D, MacRae C, Sasaki T, et al: Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. N Engl J Med 341:1715, 1999

19. Minetti C, Sotgia F, Bruno C, et al: Mutations in the caveolin-3 gene cause autosomal dominant limb-girdle muscular dystrophy. Nat Genet 18:365, 1998

20. Nigro V: Molecular bases of autosomal recessive limb-girdle muscular dystrophies. Acta Myol 22: 35, 2003

21. Selcen D, Engel AG: Mutations in myotilin cause myofibrillar myopathy. Neurology 62:1363, 2004

22. Zatz M, de Paula F, Starling A, et al: The 10 autosomal recessive limb-girdle muscular dystrophies. Neuromuscul Disord 13:532, 2003

23. Betz RC, Schoser BG, Kasper D, et al: Mutations in CAV3 cause mechanical hyperirritability of skeletal muscle in rippling muscle disease. Nat Genet 28:218, 2001

24. Richard I, Broux O, Allamand V, et al: Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. Cell 81:27, 1995

25. Liu J, Aoki M, Illa I, et al: Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. Nat Genet 20:31, 1998

26. Bashir R, Britton S, Strachan T, et al: A gene related to *Caenorhabditis elegans* spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nat Genet 20:37, 1998

27. Duggan DJ, Gorospe JR, Fanin M, et al: Mutations in the sarcoglycan genes in patients with myopathy. N Engl J Med 336:618, 1997

28. Moreira ES, Wiltshire TJ, Faulkner G, et al: Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. Nat Genet 24:163, 2000

29. Frosk P, Greenberg CR, Tennese AA, et al: The most common mutation in FKRP causing limb girdle muscular dystrophy type 2I (LGMD2I) may have occurred only once and is present in Hutterites and other populations. Hum Mutat 25:38, 2005

30. Poppe M, Cree L, Bourke J, et al: The phenotype of limb-girdle muscular dystrophy type 2I. Neurology 60:1246, 2003

31. Bushby KM, Beckmann JS: The 105th ENMC sponsored workshop: pathogenesis in the non-sarcoglycan limb-girdle muscular dystrophies. Neuromuscul Disord 13:80, 2003

32. Hackman P, Vihola A, Haravuori H, et al.: Tibial muscular dystrophy is a titinopathy caused by mutations in TTN, the gene encoding the giant skeletal-muscle protein titin. Am J Hum Genet 71:492, 2002

33. Hewitt JE, Grewl PK: Glycosylation defects in inherited muscle disease. Cell Mol Life Sci 60: 251, 2003

34. Grewal PK, Hewitt JE: Glycosylation defects: a new mechanism for muscular dystrophy. Hum Mol Genet 12: R259, 2003

35. Arahata K, Ishii H, Hayashi YK: Congenital muscular dystrophies. Curr Opin Neurol 8:385, 1995

36. Hayashi YK, Chou FL, Engvall E, et al: Mutations in the integrin alpha7 gene cause congenital myopathy. Nat Genet 19:94, 1998

37. Flanigan KM, Kerr L, Bromberg MD, et al: Congenital muscular dystrophy with rigid spine syndrome: a clinical, pathological, radiological, and genetic study. Ann Neurol 47:147, 2000

38. Camacho Vanegas O, Bertini E, Zhang RZ, et al: Ullrich scleratonic muscular dystrophy is caused by recessive mutations in collagen type VI. Proc Natl Acad Sci USA 98:7516, 2001

39. Lamande SR, Bateman JF, Hutchison W, et al: Reduced collagen VI causes Bethlem myopathy: a heterozygous COL6A1 nonsense mutation results in mRNA decay and functional haploinsufficiency. Hum Mol Genet 7:981, 1998

40. Dalakas MC, Park KY, Semino-Mora C, et al: Desmin myopathy, a skeletal myopathy with cardiomyopathy caused by mutations in the desmin gene. N Engl J Med 342:770, 2000

41. Goldfarb LG, Vicart P, Goebel H, et al: Desmin myopathy. Brain 127:723, 2004

42. Selcen D, Ohno K, Engel AG: Myofibrillar myopathy: clinical, morphological and genetic studies in 63 patients. Brain 127:439, 2004

43. Vicart P, Dupret JM, Hazan J, et al: Human desmin gene: cDNA sequence, regional localization and exclusion of the locus in a familial desmin-related myopathy. Hum Genet 98:422, 1996

44. Banwell BL, Russel J, Fukudome T, et al: Myopathy, myasthenic syndrome, and epidermolysis bullosa simplex due to plectin deficiency. J Neuropathol Exp Neurol 58:832, 1999

45. Mankodi A, Thornton CA: Myotonic syndromes. Curr Opin Neurol 15:545, 2002

46. Day JW, Ricker K, Jacobsen JF, et al: Myotonic dystrophy type 2: molecular, diagnostic and clinical spectrum. Neurology 60:657, 2003

47. Mankodi A, Teng-Umnuay P, Krym M, et al: Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2. Ann Neurol 54:760, 2003

48. Kanandia RN, Johnstone KA, Mankodi et al: A muscleblind knockout model for myotonic dystrophy. Science 302:1978, 2003

49. Lemmers RJL, de Kievit P, van Geel M, et al: Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. Ann Neurol 50:816, 2001

50. Rose MR, Tawil R: Drug treatment for facioscapulohumeral muscular dystrophy. Cochrane Database Syst Rev (2):CD002276, 2004

51. Shanmugam V, Dion P, Rochefort D, et al: PABP2 polyalanine tract expansion causes intranuclear inclusions in oculopharyngeal muscular dystrophy. Ann Neurol 48:798, 2000

52. Hogan K: The anesthetic myopathies and malignant hyperthermias. Curr Opin Neurol 11:469, 1998

53. Allen GC, Larach MG, Kunselman AR: The sensitivity and specificity of the caffeine-halothane contracture test: a report from the North American Malignant Hyperthermia Registry. The North American Malignant Hyperthermia Registry of MHAUS. Anesthesiology 88:579, 1998

54. DiMauro S, Lamperti C: Muscle glycogenoses. Muscle Nerve 24: 984, 2001

55. Hirano M, DiMauro S: Metabolic myopathies. Adv Neurol 88:217, 2002

56. Quinlivan R, Beynon RJ: Pharmacological and nutritional treatment for McArdle's disease (glycogen storage disease type V). Cochrane Database Syst Rev (3):CD003458, 2004

57. Vissing J, Haller RG: The effect of oral sucrose on exercise tolerance in patients with McArdle's disease. N Engl J Med 349: 2503, 2003

58. Vasconcelos O, Sivakumar K, Dalakas MC, et al: Nonsense mutation in the human phosphofructokinase muscle subunit gene associated with retention of intron 10 in one of the isolated transcripts in Ashkenazi Jewish patients with Tarui disease. Proc Natl Acad Sci USA 92:10322, 1995

59. Engel AG, Hirschhorn R, Huie ML: Acid maltase deficiency. Myology. Engel AG, Franzini-Armstrong C, Eds. McGraw-Hill, New York, 2004, p 1559

60. Laforet P, Nicolino M, Eymard PB, et al: Juvenile and adult-onset acid maltase deficiency in France: genotype-phenotype correlation. Neurology 55:1122, 2000

61. Dalakas MC, Leon-Monzon ME, Bernardini I, et al: The zidovudine-induced mitochondrial myopathy is associated with muscle carnitine deficiency and lipid storage. Ann Neurol 35:482, 1994

62. Johns DR: Mitochondrial DNA and disease. N Engl J Med 333:638, 1995

63. DiMauro S, Schon EA: Mitochondrial respiratory-chain diseases. N Engl J Med 348:266, 2003

64. McFarland R, Taylor RW, Turnbull DM: The neurology of mitochondrial DNA disease. Lancet Neurol 6:343, 2002

65. Zeviani M, Carelli V: Mitochondrial disorders. Curr Opin Neurol 16:585, 2003

66. Lehmann-Horn F, Jurkat-Rott K: Voltage-gated ion channels and hereditary disease. Physiol Rev 79:1317, 1999

67. Celesia GG: Disorders of membrane channels or channelopathies. Clin Neurophysiol 112:2, 2001

68. Kleopa KA, Barchi RL: Genetic disorders of neuromuscular ion channels. Muscle Nerve 97:244, 2002

69. Ptacek LJ, Tawil R, Griggs RC, et al: Sodium channel mutations in acetazolamide-responsive myotonic congenital, paramyotonia congenital, and hyperkalemic periodic paralysis. Neurology 44:1500, 1994

70. Fouad G, Dalakas M, Servidei S, et al: Genotype-phenotype correlations of DHP receptor alpha 1-subunit gene mutations causing hypokalemic periodic paralysis. Neuromuscul Disord 7:33, 1997

71. Dalakas MC, Engel WK: Treatment of "permanent" muscle weakness in familial hypokalemic periodic paralysis. Muscle Nerve 6:182, 1983

72. Tawil R, McDermott MP, Brown R, et al: Randomized trial of dichlorphenamide in the periodic paralyses. Working Group on Periodic Paralysis. Ann Neurol 47:46, 2000

73. Sinha S, Newson-Davis J, Mills K, et al: Autoimmune aetiology for acquired neuromyotonia (Isaac's syndrome). Lancet 338:75, 1991

74. McConville J, Vincent A: Diseases of the neuromuscular junction. Curr Opin Pharmacol 2:296, 2002

75. Sieb JP: Myopathies due to drugs, toxins, and nutritional deficiency. Myology. Engel AG, Franzini-Armstrong C, Eds. McGraw-Hill, New York, 2004, p 1693

76. Argov Z: Drug-induced myopathies. Curr Opin Neurol 13:541, 2000

77. Dalakas MC: Inflammatory and toxic myopathies. Curr Opin Neurol Neurosurg 5:645, 1992

78. Dalakas MC, Illa I, Pezeshkpour GH, et al: Mitochondrial myopathy caused by long-term zidovudine therapy. N Engl J Med 322:1098, 1990

79. Lewis W, Dalakas MC: Mitochondrial toxicity of antiviral drugs. Nat Med 1:417, 1995

80. Brinkman K, Smeitink JA, Romijn JA, et al: Mitochondrial toxicity induced by nucleoside-analogue reverse-transcriptase inhibitors is a key factor in the pathogenesis of antiretroviral-therapy-related lipodystrophy. Lancet 354:1112, 1999

81. Rosenson RS: Current overview of statin-induced myopathy. Am J Med 116:408, 2004

82. Sieb JP, Gillessen T: Iatrogenic and toxic myopathies. Muscle and Nerve 27:142, 2003

83. Danon MJ, Carpenter S: Myopathy with thick filament loss following prolonged paralysis with vacuronium during steroid treatment. Muscle Nerve 14:1131, 1991

84. Hirano M, Ott B, Raps E, et al: Acute quadriplegic myopathy: a complication of steroids, nondepolarizing blocking agents or both. Neurology 42:2082, 1992

85. Bella I, Chad DA: Neuromuscular disorders and acute respiratory failure. Neurol Clin 16:391, 1998

86. Bardosi A, Dickmann U: Necrotizing myopathy with paracrystalline inclusion bodies in hypervitaminosis E. Acta Neuropathol 75:166, 1987

87. Kuncl RW: Colchicine myopathy and neuropathy. N Engl J Med 316:1562, 1987

88. Illa I, Dinsmore S, Dalakas MC: Immune-mediated mechanisms and immune activation of fibroblasts in the pathogenesis of eosinophilia-myalgia syndrome induced by L-tryptophan. Hum Pathol 24:702, 1993

89. Drachman DB: Myasthenia gravis. N Engl J Med 330:1797, 1994

90. Hill M: The neuromuscular junction disorders. J Neurol Neurosurg Psychiatry 74:ii32, 2003

91. Vincent V, Newsom-Davis J: Immunology of the motor nerve terminal. Immunology of Neuromuscular Disease. Hohlfeld R, Ed. Kluwer Academic Publishers, Lancaster, England, 1994, p 147

92. Aragones JM, Bolibar I, Bonfill X, et al: Myasthenia gravis: a higher than expected incidence in the elderly. Neurology 60:1024, 2003

93. Hoch W, McConville J, Helms S, et al: Auto-antibodies to the receptor tyrosine kinase MuSK in patients with myasthenia gravis without acetylcholine receptor antibodies. Nat Med 7:365, 2001

94. Sanders DB, El-Salem K, Massey JM: Clinical aspects of MuSK antibody positive seronegative MG. Neurology 60: 1978, 2003

95. Ciafaloni E, Massey JM, Tucker-Lipscomb B, et al: Mycophenolate mofetil for myasthenia gravis: an open-label pilot study. Neurology 56:97, 2001

96. Vincent A, Bowen J, Newsom-Davis J: Seronegative generalised myasthenia gravis: clinical features, antibodies, and their targets. Lancet Neurol 2:99, 2003

97. Wolfe GI, Kaminski HJ, Jaretzki A 3rd, et al: Development of a thymectomy trial in nonthymomatous myasthenia gravis patients receiving immunosuppressive therapy. Ann NY Acad Sci 998:473, 2003

98. Lennon VA, Kryzer TJ, Griesmann GE, et al: Calcium-channel antibodies in the Lambert-Eaton myasthenic syndrome and other paraneoplastic syndromes. N Engl J Med 332:1467, 1995

99. Newsom-Davis J: Lambert-Eaton myasthenic syndrome. Curr Treatment Options in Neurology 3:127, 2001

100. Maddison P, Lang B, Mills K: Long-term outcome in Lambert-Eaton myasthenic syndrome without lung cancer. J Neurol Neurosurg Psychiatry 70:212, 2001

101. Engel AG, Ohno K, Sine SM: Congenital myasthenic syndromes: A diverse array of molecular targets. J Neurocytol 32: 1017, 2003

Acknowledgments

Figure 1 Seward Hung.

Figure 2 Dana Burns-Pizer.

Figures 3 and 4 Tom Moore.

IV CEREBROVASCULAR DISORDERS

Scott E. Kasner, m.d.
Lewis B. Morgenstern, m.d.

Definitions

Stroke is a sudden neurologic deficit caused by either ischemia (80%) or hemorrhage (20%). Ischemic stroke is classified according to the area of the brain affected and the etiologic mechanism. Hemorrhagic stroke is classified as either subarachnoid (5%) or intracerebral (intraparenchymal) (15%). Transient ischemic attack (TIA) is a sudden vascular-related focal neurologic deficit that resolves completely. TIAs are classically defined as lasting less than 24 hours; they generally last less than 1 hour. A TIA should not be considered a separate entity but, rather, a herald of ischemic stroke and an opportunity to intervene.

Epidemiology

Stroke is the leading cause of disability and the third leading cause of death in the United States. Until 1998, the annual incidence of stroke was estimated to be 550,000 on the basis of studies of homogeneous, predominantly white populations. Results of the Greater Cincinnati/Northern Kentucky Stroke Study suggest that previous reports underestimated stroke incidence by almost 50%. By rigorously counting strokes in all racial and ethnic groups in Cincinnati, the yearly estimate for the United States was revised to 731,100.[1] The explanation for the previous underestimation of stroke incidence is that the incidence is higher in African Americans than in non-Hispanic whites. African Americans in Cincinnati had a twofold to fourfold higher incidence of stroke than did non-Hispanic whites in Rochester, Minnesota, during the same period. In northern Manhattan, African Americans had a 2.4-fold higher incidence of stroke than did non-Hispanic whites.[2]

African Americans also have higher stroke mortality. In Texas, African-American men 45 to 59 years of age had a 306% higher stroke mortality than did non-Hispanic whites. At 75 years of age and older, when stroke mortality is at its highest, the excess mortality for African-American men had diminished to 26% above the mortality for non-Hispanic whites. African-American women in the same age groups had a 222% and a 10% greater stroke mortality, respectively, than that of non-Hispanic whites.[3] The increased burden of stroke in African Americans may be worsening. Between 1992 and 1996, rates of almost all subtypes of stroke had increased among African Americans, and the number of stroke deaths had increased by 8%.[4]

The pathogenesis of stroke in African Americans may also differ somewhat from that in non-Hispanic whites. Extracranial carotid disease and cardioembolism more commonly cause ischemic stroke in non-Hispanic whites, whereas intracranial thromboembolic disease is more common in African Americans.[5] However, the various disease mechanisms can occur in either racial group.

Information on stroke in Hispanic Americans is less readily available. Data from northern Manhattan suggest a twofold higher incidence of stroke in Hispanics than in non-Hispanic whites.[2] In Mexican Americans, the incidence of ischemic stroke and intracerebral hemorrhage is substantially greater than in non-Hispanic whites.[6] As the Mexican-American population grows and ages, targeting of this population for stroke prevention becomes critical.

Women have lower stroke rates than men at all age ranges except 75 years and older, when stroke rates are at their highest. In Texas, 61% of all stroke-related deaths occur in women.[3] Stroke and vascular disease have traditionally been seen as male disorders; this has resulted in a shift in the focus of prevention and acute treatment regimens away from women.

Overall, the decline in stroke-related mortality has inexplicably slowed over the past several decades and almost came to a halt in the 1990s.[4] A lack of emphasis on prevention may have led to an increase in incidence, particularly in some racial and ethnic groups. As people survive other diseases, stroke may become a more common cause of death. Nonetheless, this leveling of the previous downward trend in stroke mortality is cause for alarm.

Approach to the Acute Stroke Patient

Stroke is the quintessential medical emergency. Stroke patients who receive neurologic care within 6 hours of the onset of symptoms have a fourfold greater chance of a good outcome than those treated after this acute period.[7] This observation, which was made before the thrombolytic era, illustrates the crucial link between time and outcome for stroke victims. Patients who receive early treatment to restore cerebral perfusion and to maximize protection of neurons have better outcomes. Physicians can reduce the delay in getting stroke victims to the emergency department by educating at-risk patients and their families about stroke symptoms and encouraging them to call 911 if stroke symptoms occur.[8] Once a stroke patient arrives at the emergency department, triage and treatment are critical in reducing mortality and morbidity [*see Figure 1*]. One clinical instrument for assessing stroke patients is the National Institutes of Health Stroke Scale, which is available online at http://www.ninds.nih.gov/doctors/NIH_Stroke_Scale.pd. Before specific treatment is implemented, however, neuroimaging must be performed.

Neuroimaging of Acute Stroke

The brain should first be imaged with noncontrast computed tomography, which reliably distinguishes acute intracerebral hemorrhage (ICH) from ischemia [*see Figure 2*]. This distinction is critical because the management of hemorrhagic stroke [*see* Hemorrhagic Stroke, *below*] is substantially different from that of ischemic stroke. CT is relatively insensitive to ischemic changes in the first few hours after stroke. Early CT findings may be subtle [*see Figure 2*], but hyperdense vessels (suggestive of acute thrombus), loss of boundaries between gray and white matter, and effacement of cerebral sulci are highly specific for ischemic stroke.[9]

The ability to identify at-risk cerebral tissue has considerable clinical relevance because therapy has the potential to salvage viable tissue.[10] Physiologic imaging may detect the presence of so-called ischemic penumbra (tissue with impaired blood flow but active metabolism) surrounding infarcted tissue [*see Figure 3*]. Aerobic metabolism cannot be sustained in the face of inadequate blood flow, and necrosis ultimately results. The time

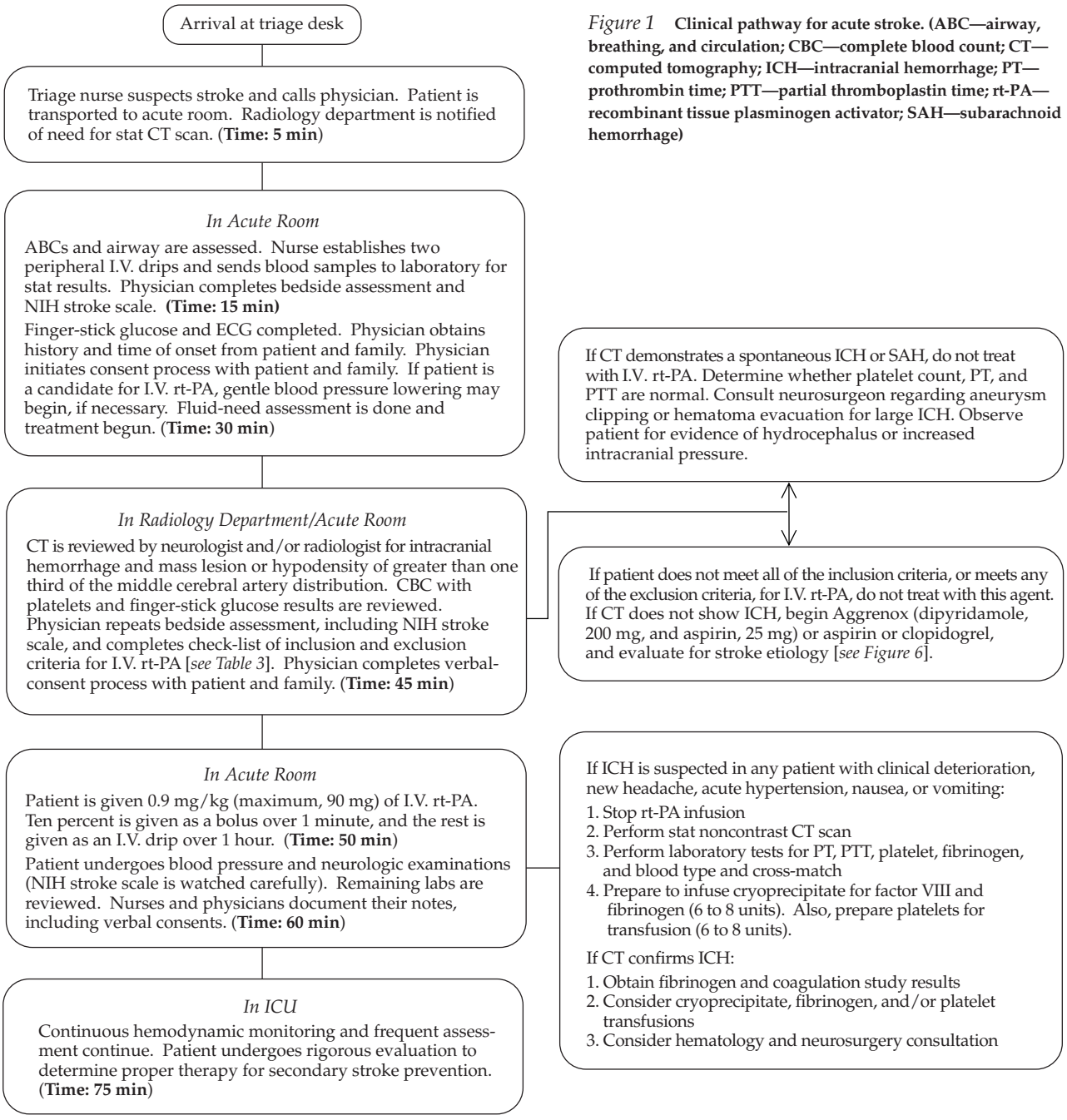


Figure 1 **Clinical pathway for acute stroke. (ABC—airway, breathing, and circulation; CBC—complete blood count; CT—computed tomography; ICH—intracranial hemorrhage; PT—prothrombin time; PTT—partial thromboplastin time; rt-PA—recombinant tissue plasminogen activator; SAH—subarachnoid hemorrhage)**

course of this process is poorly defined, although substantial volumes of brain potentially remain viable for up to 17 hours after stroke. Diffusion-weighted imaging (DWI) reveals changes in infarcted tissue hours before conventional CT or magnetic resonance imaging. Hyperintensity on DWI is predictive of the extent of infarction, and quantitative analysis of the apparent diffusion coefficient may identify areas of cytotoxic edema, suggesting permanently infarcted tissue. Perfusion-weighted imaging (PWI) demonstrates regionally reduced cerebral blood flow. The difference between the perfusion defect and the diffusion defect may represent penumbral tissue, which is the target of acute therapy [*see Figure 3*]. Further, MRI spectroscopy may provide biochemical markers of the extent of ischemic injury. However, these MRI techniques are not available emergently in many centers, which limits their clinical utility at present.

Several imaging techniques may characterize the acute vascular lesion.[10] The gold standard is conventional angiography, which may demonstrate acute occlusion or an embolus lodged at a vascular bifurcation. The vasculature can also be visualized quickly and noninvasively with CT angiography (CTA) and magnetic resonance angiography (MRA). Alternatively, transcranial Doppler ultrasonography can provide indirect evidence of major vascular occlusion and offers the advantage of real-time bedside monitoring in patients who receive thrombolytic therapy.[11]

Ischemic Stroke

Patients who present with neurologic dysfunction of sudden onset or who report neurologic signs and symptoms evolving

over a few minutes to a few hours are most likely experiencing a stroke. In most of these patients, the stroke is ischemic rather than hemorrhagic [*see Table 1*].

ETIOLOGY AND PATHOGENESIS

Classically, ischemic strokes have been categorized as either thrombotic or embolic; clinically, however, strokes resulting from thrombosis may be indistinguishable from those caused by embolism. Ischemic strokes are more reliably categorized as resulting from cardioembolism, large-vessel atherothromboembolism, small-vessel occlusive disease, or other identified mechanism or as cryptogenic (idiopathic).[12] The first three causes account for 70% to 90% of ischemic strokes. The latter two causes are pathophysiologically, diagnostically, and therapeutically distinct, and they are, therefore, are discussed separately [*see* Uncommon Causes of Ischemic Stroke, *below*].

Common Mechanisms

Cardioembolism Cardioembolism most commonly results from atrial fibrillation, mural thrombus, ventricular akinesis after myocardial infarction, dilated cardiomyopathy, and valvular disease [*see Figure 4*]. In each of these disorders, thrombus develops within the heart and embolizes to the brain. Ischemic events may be multiple and may occur in any major vessel. Therefore,

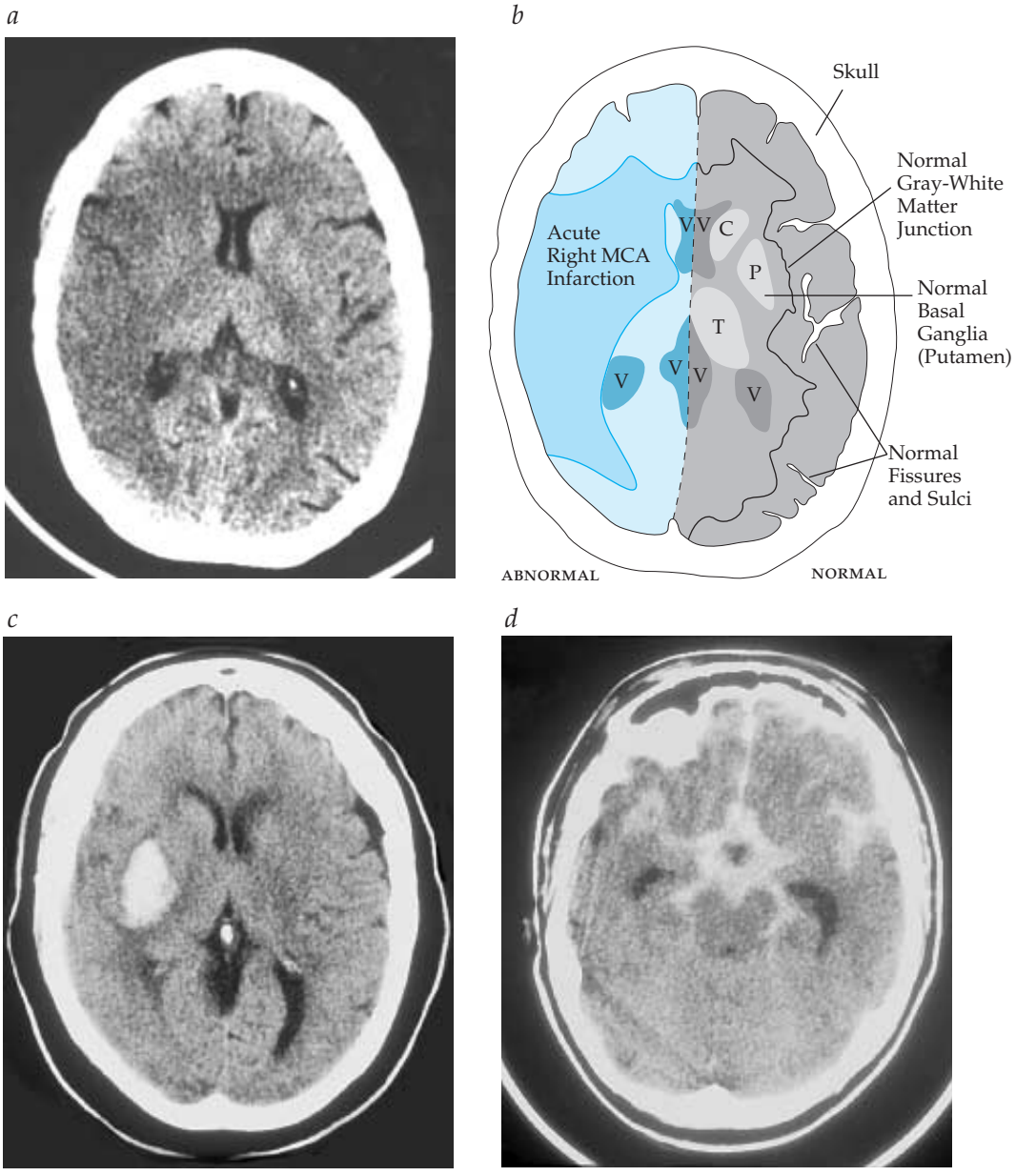


Figure 2 **(*a*) Early computed tomographic findings in acute ischemic stroke. Three hours after onset of left hemiparesis and neglect, this noncontrast CT scan reveals extensive early findings in the right hemisphere, including obscuration of the gray-white junction and the basal ganglia and effacement of the cortical sulci. (*b*) Detail of CT findings shown in panel *a*. (*c*) Noncontrast head CT scan of a right putamenal intracerebral hemorrhage. (*d*) Noncontrast head CT scan of a subarachnoid hemorrhage manifesting the classic star-shaped area of hyperdensity in the basal cisterns. (C—caudate nucleus; P—putamen; T—thalamus; V—ventricles)**

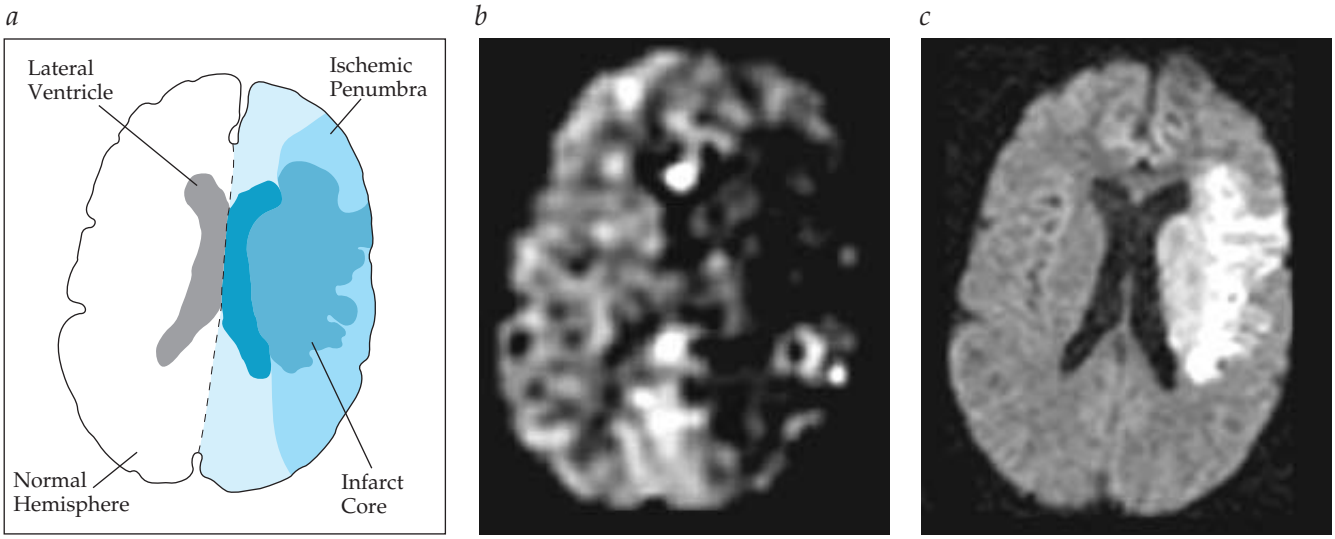


Figure 3 **Magnetic resonance imaging scan showing perfusion-diffusion mismatch in acute ischemic stroke. The mismatch between diffusion and perfusion abnormalities may represent the ischemic penumbra (*a*). This patient presented with right hemiparesis, right-visual-field impairment, right sensory loss, and aphasia. The area of decreased perfusion (*b*) is much larger than the area identified on diffusion imaging (*c*), suggesting that the surrounding tissue may still be salvageable.**

cardioembolism must be considered in nearly every ischemic stroke patient.

Large-vessel atherothromboembolism Atherosclerotic narrowing of the major extracranial or intracranial arteries may cause acute occlusion of the vessel itself, distal artery-to-artery embolic stroke, or focal hypoperfusion [*see Figure 5*]. A hallmark is the recurrence of similar clinical events, as a result of ischemia in the same vascular territory caused by involvement of a single large vessel. In general, atherosclerosis occurs at major arterial branches (e.g., the carotid bifurcation in the neck or intracranial branch points) and at vessel origins (e.g., the origin of the vertebral artery from the subclavian artery).

Small-vessel occlusive disease Small vessels are tiny terminal branches of larger vessels, such as those in the internal capsule, corona radiata, thalamus, and pons. Occlusion of these small vessels is often synonymous with so-called lacunar infarction. The mechanism of the occlusive process is uncertain, but lipohyalinosis, local atherosclerosis, and microthrombosis are possible. The process is most common in patients with long-standing diabetes or hypertension and is characterized by several specific clinical syndromes [*see* Clinical Manifestations and Lesion Localization, *below*]. The diagnosis of small-vessel disease rests on the clinical syndrome and the absence of an alternative etiology.

DIAGNOSIS OF ACUTE ISCHEMIC STROKE

Clinical Manifestations and Lesion Localization

Two pairs of vessels supply blood to the brain: the internal carotid arteries and the vertebral arteries. These vessels, which deliver 20% of the cardiac output, join on the ventral surface of the brain to form the intracranial vessels and the circle of Willis [*see Figure 5*]. The anterior cerebral artery supplies blood to the medial frontal and deep structures. Occlusion of the anterior cerebral artery is characterized by contralateral leg weakness [*see Table 2*], but isolated infarction of the anterior cerebral artery is uncommon.

The middle cerebral artery (MCA) divides into two major trunks, and each trunk divides into five to seven branches that supply blood to the lateral hemisphere. Because the MCA supplies a large territory, MCA occlusion causes a clinical syndrome that includes contralateral hemiparesis and hemisensory deficit (in which the deficit in the face and arm is greater than the deficit in the leg), aphasia (if the dominant hemisphere is affected) or neglect (if the nondominant hemisphere is affected), contralateral visual-field defect, deviation of gaze, dysarthria, and other cortical symptoms.

Table 1 Differential Diagnosis of Acute Ischemic Stroke

Possible Cause	*Comment*
Drugs or other toxins	Unlikely to cause focal neurologic symptoms; exclude by history and toxic screening
Seizure	Can mimic focal neurologic signs; exclude by history
Metabolic derangements	Abnormalities of glucose, calcium, Po_2, Pco_2, and electrolytes and liver and kidney dysfunction can all cause neurologic abnormalities; hypoglycemia and hyperglycemia are notorious for causing focal signs; exclude by laboratory tests
Migraine	Although migraine is a diagnosis of exclusion, it must be considered in patients with stroke; headache can be a prominent component of both ischemic and hemorrhagic stroke; exclude by history and physical examination
Brain tumor	Unlikely to present acutely; exclude by history and CT scan
Intracranial hemorrhage	Subdural, epidural, subarachnoid, and intracerebral hemorrhage can all mimic ischemic stroke; exclude by CT scan
Psychiatric disease	Conversion disorder and malingering can usually be discovered by a careful physical examination and history

CT—computed tomography Pco_2—carbon dioxide tension Po_2—oxygen tension

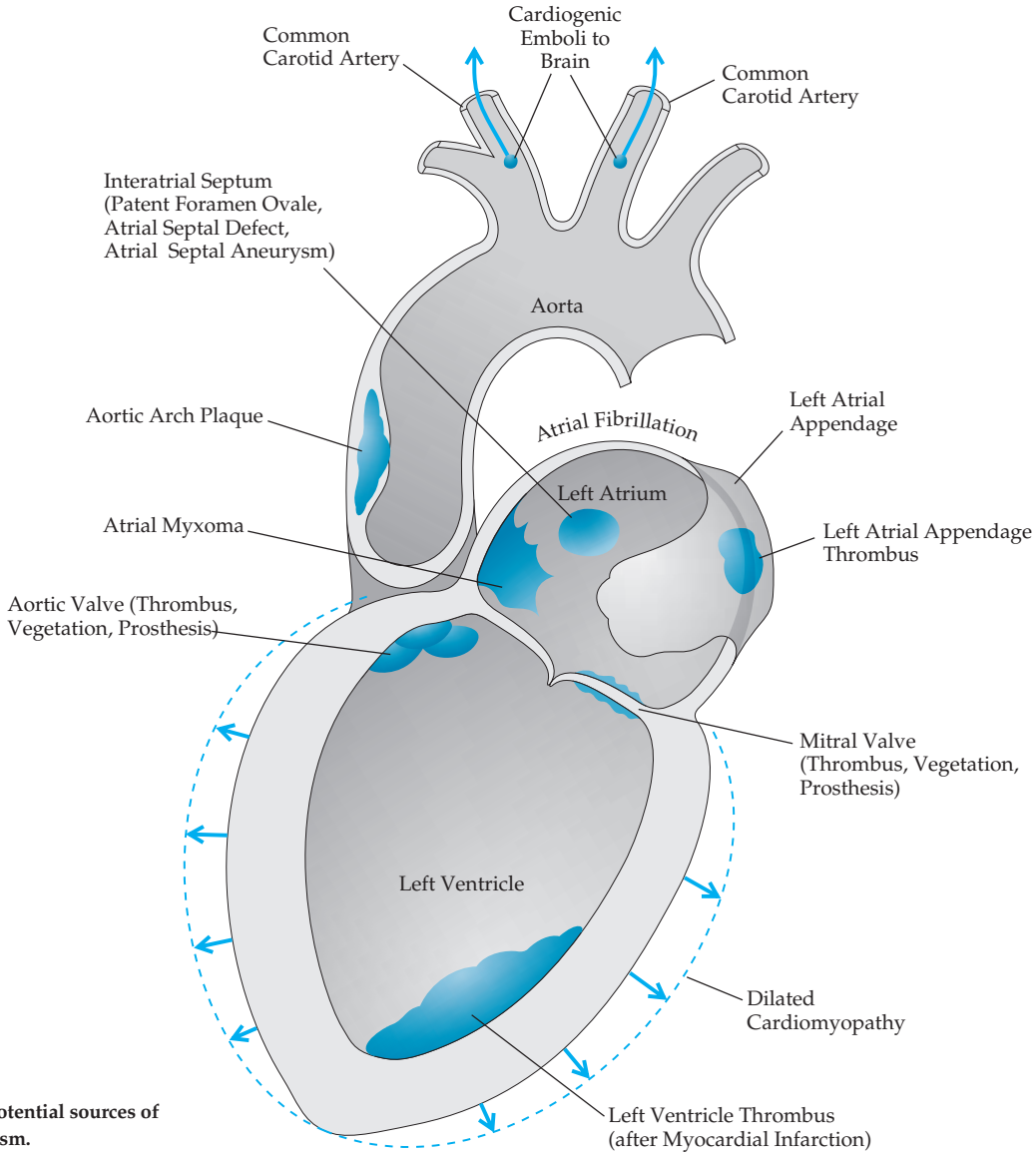


Figure 4 **Potential sources of cardioembolism.**

The two major branches of the vertebral arteries are the anterior spinal artery, which supplies the spinal cord, and the posterior inferior cerebellar artery, which leads to the inferior cerebellum and the lateral medulla. The two vertebral arteries then unite to form the basilar artery. The major branches of the basilar artery are the anterior inferior cerebellar artery and the superior cerebellar artery, which supply parts of the pons and cerebellum. Occlusion of the vertebral arteries or basilar artery leads to a combination of signs and symptoms that depend on the level and extent of infarction. These signs and symptoms include so-called crossed facial sensory and body motor signs, diplopia, facial numbness and weakness, vertigo, nausea and vomiting, tinnitus, hearing loss, ataxia, gait abnormality, hemiparesis, dysphagia, and dysarthria. The basilar artery terminates by dividing into two posterior cerebral arteries that supply the medial temporal lobe, the occipital lobe, and parts of the thalamus. Occlusion of the posterior cerebral artery results in occipital infarction and therefore contralateral visual-field loss. Such occlusion may also cause contralateral hemiparesis and behavioral changes.

After leaving the circle of Willis, the vessels branch repeatedly and ultimately become end arteries. Occlusion of these penetrating vessels typically manifests as pure motor hemiparesis, pure sensory stroke, clumsy hand–dysarthria syndrome, or ataxic hemiparesis.

Diagnostic Evaluation

The highest risk of recurrent stroke occurs within the first month after initial stroke symptoms (and within the first few days after TIA)[13,14]; therefore, an expeditious evaluation of patients presenting with TIA or suspected stroke should be undertaken and prophylactic therapy begun immediately at presentation [*see Figure 6*] (though after initial treatment with thrombolytics, if applicable [*see* Treatment of Acute Ischemic Stroke, *below*]). In a patient with recent cerebral ischemia, the first step is to localize the lesion [*see* Clinical Manifestations and Lesion Localization, *above*]. Patients with anterior circulation strokes should undergo evaluation of the heart, extracranial carotid arteries, and intracranial anterior circulation. Cardiac evaluation [*see Figure 4*] begins with an electrocardiogram, a cardiac history and examination, and either a transthoracic echocardiogram (TTE) or a transesophageal echocardiogram (TEE). If a TTE does not provide an absolute diagnosis but suggests mural throm-

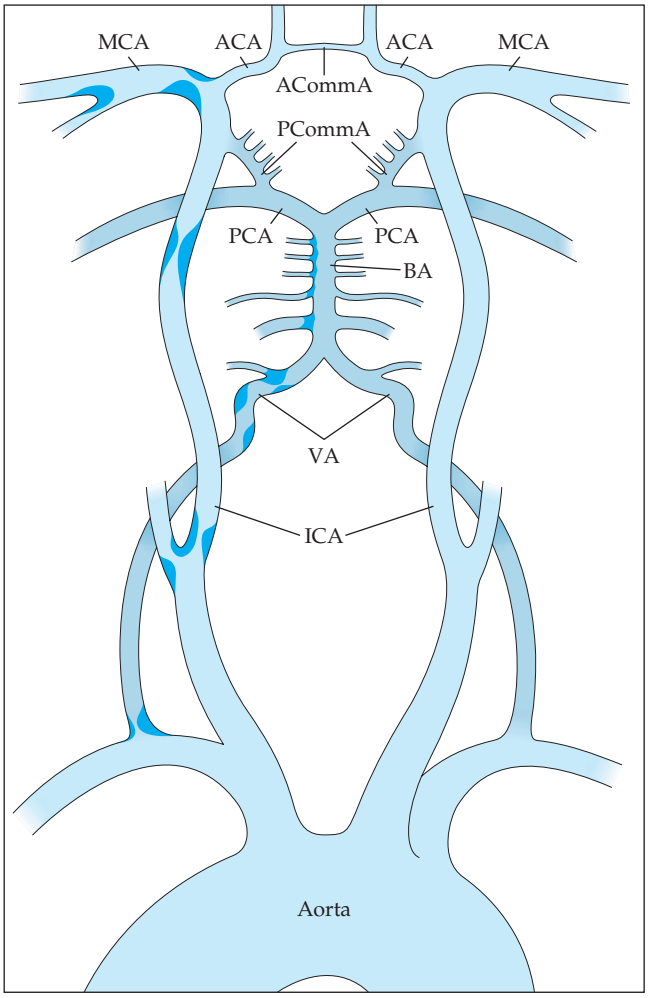


Figure 5 **Cerebrovascular anatomy and common sites of atherosclerosis are shown. The internal carotid artery (ICA) enters the skull, and its first major branch is the ophthalmic artery to the eye. Next are the anterior choroidal artery and the posterior communicating artery (PCommA). The PCommA connects the anterior circulation to the posterior circulation. The ICA then terminates as it divides into the anterior cerebral artery (ACA) and the middle cerebral artery (MCA). The vertebral arteries (VA) enter the skull and merge at the inferior border of the pons to form the basilar artery (BA). The BA then terminates as it divides into the two posterior cerebral arteries (PCA). (ACommA—anterior communicating artery)**

bus, valvular disease, or patent foramen ovale, then a TEE is ordered. A TEE allows visualization of structures not seen on a TTE, including clots in the left atrial appendage and aortic arch. The extracranial carotid circulation and intracranial anterior circulation can be visualized by MRA or CTA. In hospitals where these diagnostic tools are either not available or of poor quality, carotid ultrasonography should be performed. Transcranial Doppler imaging can help detect intracranial stenosis. The gold standard remains conventional cerebral angiography. Because this invasive test has possible complications (1% risk of stroke in most series) and is expensive, it should be reserved for cases in which its results could change treatment decisions. Posterior circulation evaluation involves the same cardiac evaluation. The posterior circulation is visualized by MRA, CTA, or conventional angiography.

Subcortical infarctions, or lacunae, that are greater than 1.5 cm in diameter are usually thromboembolic in nature. The evaluation in these patients is the same as that described above.

Laboratory investigations for all patients include measurement of the fasting lipid level within 48 hours of symptom onset, measurement of the homocysteine level, complete blood count, prothrombin time, partial thromboplastin time, and chemistry panel. If the patient is younger than 45 years or has no stroke risk factor or other identified etiology, the following tests should be considered but are of low yield: anticardiolipin antibody test, lupus anticoagulant profile, erythrocyte sedimentation rate, testing for factor V Leiden, rapid plasmin reagin test, and antinuclear antibody test. Tests for rarer causes of ischemic stroke include assays for protein C and protein S, measurement of antithrombin III, testing for prothrombin gene mutation, testing for HIV, and measurement of lactate level (for mitochondrial disease). Additionally, lumbar puncture and cerebral angiography should be considered if there is suspicion of infection or inflammation of the cerebral blood vessels.

TREATMENT OF ACUTE ISCHEMIC STROKE

Antiplatelet and Antithrombotic Treatment

Aspirin (160 to 325 mg daily) administered within 48 hours of stroke onset has been shown to significantly reduce the risk of recurrent stroke during the first 2 weeks and, possibly, to improve outcome at 6 months.[15,16] Therefore, aspirin is recommended as initial therapy for most acute stroke patients. However, aspirin should be withheld for at least 24 hours after administration of thrombolytics [*see* Intravenous Recombinant Tissue Plasminogen Activator, *below*]. For patients who have a known contraindication to aspirin, other antiplatelet agents may serve as rational, although unproven, alternatives.

Anticoagulation is commonly used in the acute setting to prevent progressive or recurrent thromboembolic events. Nevertheless, the efficacy and safety of anticoagulation for this purpose are not well established, and its role in clinical stroke management is controversial. Many neurologists formerly used heparin, although studies have demonstrated that it offers no appreciable benefit for most patients. The International Stroke Trial was a multicenter clinical trial involving 19,436 patients who were ran-

Table 2 Clinical Features of the Major Cerebrovascular Occlusive Syndromes

Artery	*Major Clinical Features*
Anterior cerebral artery	Contralateral leg weakness
Middle cerebral artery	Contralateral face + arm > leg weakness, sensory loss, visual-field cut, aphasia/neglect
Posterior cerebral artery	Contralateral visual-field cut
Basilar artery	Oculomotor deficits and/or ataxia with "crossed"sensory/motor deficits
Vertebral artery	Lower cranial nerve deficits and/or ataxia with "crossed" sensory deficits
Penetrators	Contralateral motor or sensory deficit without cortical signs*

*Cortical signs include aphasia, apraxia, neglect, and other cognitive abnormalities.

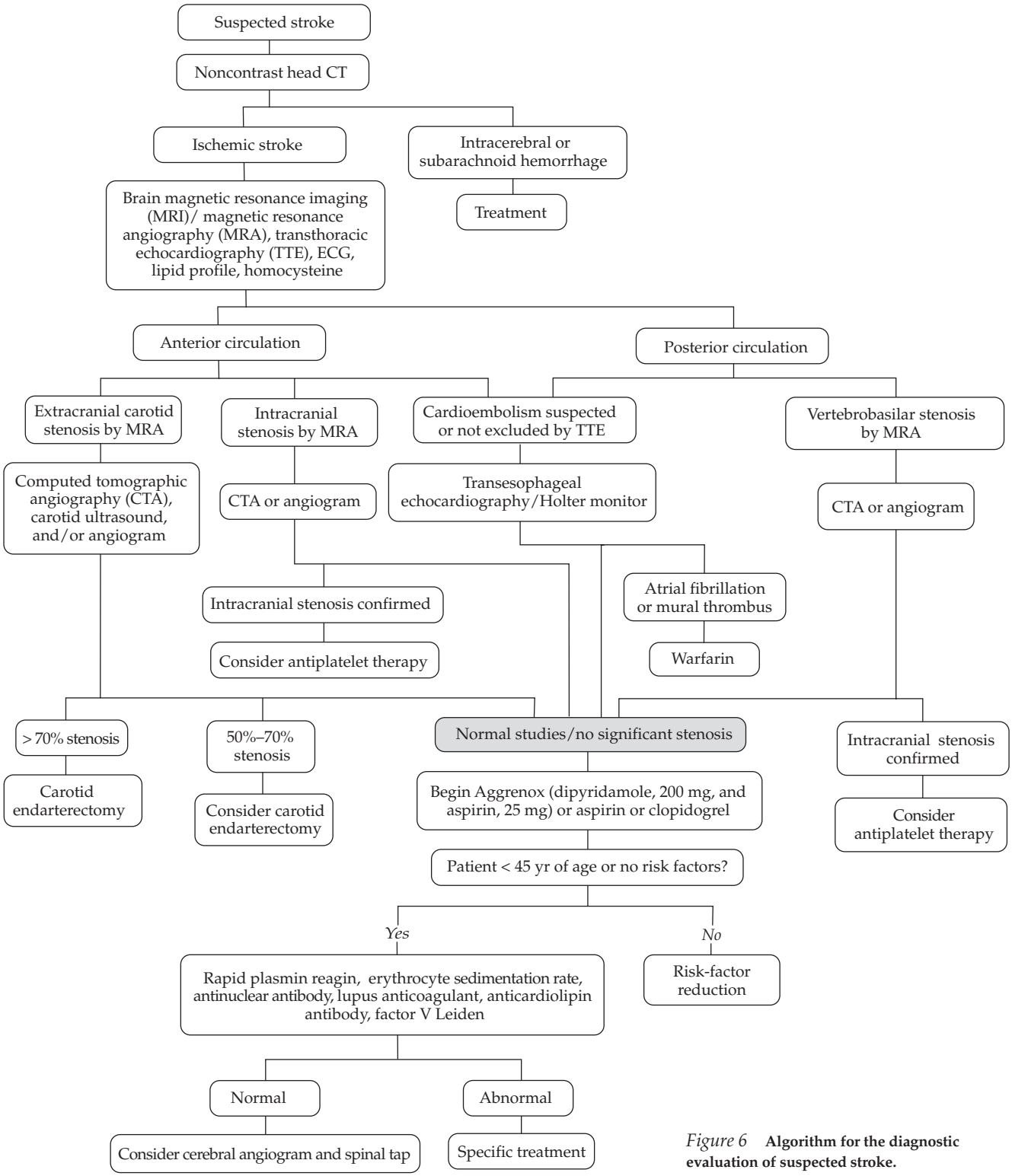


Figure 6 **Algorithm for the diagnostic evaluation of suspected stroke.**

domized within 48 hours of stroke onset into two arms. In the first arm, patients received subcutaneous heparin at a dosage of either 12,500 units or 5,000 units twice daily for 14 days. In the second arm, patients received no heparin.[15] Patients were independently randomized to receive either 300 mg of aspirin daily or no aspirin. The rate of recurrent stroke or death at 14 days was 11.7% in the patients receiving heparin and 12.0% in the nonheparin group—an insignificant difference—and there was no improvement in outcome at 6 months. The reduction in ischemic strokes by heparin was completely counterbalanced by an increase in hemorrhagic strokes. Similarly, in randomized clinical trials, use of low-molecular-weight heparin (LMWH) did not lead to a reduction in recurrent stroke or improvement in outcomes.[17] Therefore, as noted in a scientific statement developed jointly by the American Heart Association and the American Academy of Neurology, most patients with ischemic stroke

should not be treated with anticoagulation.[17] Some neurologists believe that carefully selected patients, such as those with acute basilar thrombosis, may benefit from acute anticoagulation therapy, although there is no clear evidence for or against its use.

Intravenous Recombinant Tissue Plasminogen Activator

Intravenous recombinant tissue plasminogen activator (rt-PA) was approved for use in acute stroke by the Food and Drug Administration in 1996, and consensus statements from the American Academy of Neurology and the American Heart Association support its use.[18] Although the FDA has approved the use of intravenous rt-PA for acute ischemic stroke up to 3 hours after the onset of symptoms, physicians should strive to treat patients as quickly as possible because earlier therapy is associated with better outcomes.[19] In acute stroke patients who meet the criteria for its use [*see Table 3*], rt-PA is given in a dose of 0.9 mg/kg (maximum dose, 90 mg) infused over 1 hour, with 10% of the total dose infused over the first minute. If treatment with rt-PA is suspected of inducing intracranial hemorrhage, the infusion should be suspended.

Two major trials conducted by the National Institute of Neurological Disorders and Stroke led to FDA approval of rt-PA.[19,20] The final outcome measure was return to independent function (i.e., no disability) at 3 months after onset of stroke. These studies together enrolled 624 patients at eight centers across the United States. Intravenous rt-PA was compared with placebo for patients presenting within 3 hours after onset of symptoms. The rate of symptomatic intracranial hemorrhage at 36 hours after administration of intravenous rt-PA was 6.5%, compared with 0.6% in the control group. Mortality from intracranial hemorrhage was 2.9% in the rt-PA group, compared with 0.3% in the placebo group.

Including the risk of intracranial hemorrhage, disability was significantly reduced when measured at 3 months, 6 months, and 1 year.[21] The rt-PA–treated patients were at least 30% more likely to recover to independent function than the placebo-treated patients. Mortality at 1 year was 24% in the rt-PA group and 28% in the placebo group. This benefit was seen in all stroke subtypes regardless of age or patient risk factors. Patients with large strokes were more likely to experience ICH after use of intravenous rt-PA, but this group was also more likely to have severe disability or to die if left untreated.

Only a small proportion of stroke patients are currently treated with rt-PA. Aggressive public and professional education effectively increases this proportion.[22]

Intra-arterial Thrombolysis

Direct infusion of thrombolytic agents into occluded blood vessels is a potential alternative or adjunct to therapy with intravenous rt-PA [*see Figure 7*]. The value of intra-arterial thrombolysis has been demonstrated with the experimental agent prourokinase (Prolyse), which is the highly clot-specific precursor of urokinase. The Prolyse in Acute Cerebral Thromboembolism (ProACT) trial enrolled patients with ischemic stroke caused by middle cerebral artery (MCA) occlusion within 6 hours of onset.[23] Eligibility criteria were similar to those for intravenous rt-PA treatment. Patients underwent emergent angiography; if MCA occlusion was confirmed, either intra-arterial prourokinase (9 mg) or placebo was infused for 2 hours. All patients received concomitant intravenous heparin. Arterial occlusions effectively opened in 67% of patients given prourokinase, whereas only 18% opened in patients given placebo. Three months after stroke, 40% of treated patients were functionally independent, compared with 25% of those given placebo ($P = 0.04$). The drug seemed particularly effective for moderately severe strokes. Symptomatic intracerebral hemorrhage occurred in 10.2% of treated patients, compared with 1.8% of those given placebo.

Because prourokinase is not commercially available, many centers that are capable of performing intra-arterial thrombolysis are using rt-PA, although it is not approved for this purpose. The intra-arterial rt-PA dose is uncertain, though many centers use 0.2 mg/kg (maximum, 20 mg), and all other aspects of the ProACT protocol may be followed.

Table 3 Indications and Contraindications for Intravenous rt-PA Treatment in Acute Ischemic Stroke

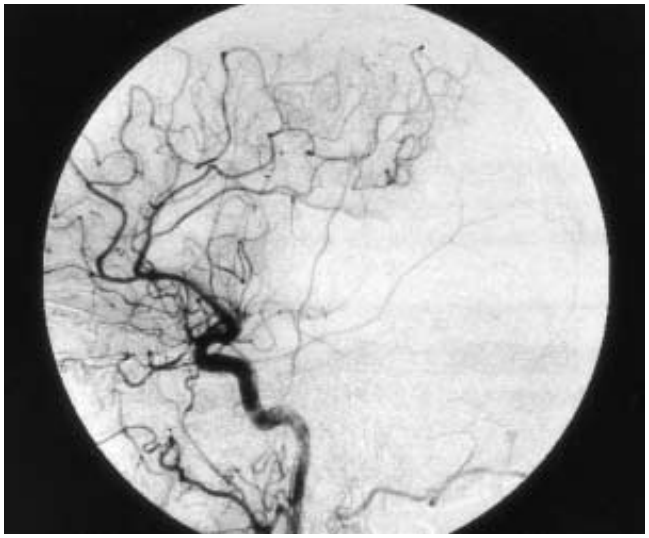
Indications	*Contraindications*	
	Absolute	*Relative*
Clinical diagnosis of disabling stroke firmly established	Onset > 3 hr ago or patient not seen normal within previous 3 hr	Glucose < 50 or > 400 mg/dl
Patient > 17 yr	Intracranial mass lesion or hemorrhage on noncontrast head CT	Seizure at stroke onset
Onset of symptoms or last time seen normal < 3 hr ago	Previous stroke or serious head trauma within previous 3 mo	Major surgery within 14 days
Previously independent functional status	Any history of intracranial hemorrhage	Arterial puncture at a noncompressible site or LP within 1 wk
	Current use of anticoagulants with PT > 15 sec or use of heparin within the past 48 hr*	Rapidly improving symptoms suggestive of TIA
	Platelets < 100,000/mm^3	GI or GU hemorrhage within 21 days
	Presenting symptoms suggestive of subarachnoid hemorrhage (worst headache of patient's life)	
	Blood pressure > 185/110 mm Hg unless minimal doses of a smooth-acting I.V. agent such as labetalol were sufficient to lower below this range†	
	Previously known cerebral aneurysm or arteriovenous malformation	

*Partial thromboplastin time and PT results are not needed before therapy unless patient is on anticoagulants.

†Caution: Do not lower blood pressure acutely by more than 10% to 15% and avoid agents that precipitously lower blood pressure. A patient who requires multiple doses should be excluded. After I.V. rt-PA is administered, blood pressure must be kept below 185/110 mm Hg for at least 24 hours.

CT—computed tomography GI—gastrointestinal GU—genitourinary LP—lumbar puncture PT—prothrombin time
rt-PA—recombinant tissue plasminogen activator TIA—transient ischemic attack

a



b

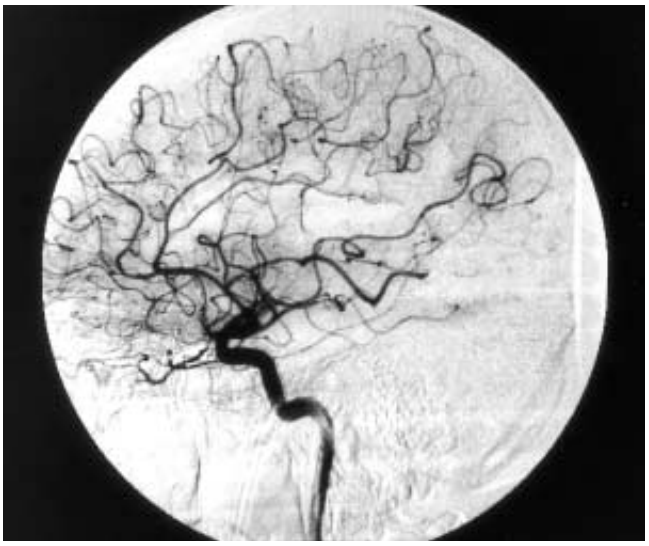


Figure 7 **Intra-arterial thrombolysis for acute ischemic stroke. Four hours after onset of left hemiparesis and neglect, conventional angiography (lateral view shown) revealed acute occlusion of the middle cerebral artery (MCA) (*a*). After treatment with intra-arterial thrombolysis, the MCA recanalized (*b*) and the patient had near-complete recovery from the deficit.**

The role of intra-arterial thrombolysis in other cerebral vessels remains uncertain. Cases of spectacular responses to thrombolytics have been reported in patients with basilar occlusion, even well beyond 6 hours. Although it is claimed that the brain stem may be relatively more tolerant of ischemia and less susceptible to hemorrhage, this hypothesis remains to be proved.

Combinations of intravenous and intra-arterial thrombolysis,[24] as well as mechanical manipulation or retrieval of an acute thrombus, are currently under investigation.

Neuroprotection

Neuroprotective strategies involve interfering with the ischemic cascade, thereby prolonging cellular viability and substantially reducing stroke size. Regrettably, a number of different compounds that seemed effective in animal studies failed in human clinical trials.[25] While several agents are being developed, researchers are attempting to determine how best to simulate stroke in animals so that animal data will be more generalizable to humans.

Supportive Medical Management

Several general medical issues are important for all stroke patients, including management of airway and oxygenation, blood pressure and hemodynamics, blood glucose, and temperature. Medical complications are also common after stroke, and such complications are associated with poor outcomes.[26] Management of these common issues and complications is critically important. Therefore, it is advisable to treat stroke patients in dedicated stroke units to reduce morbidity, mortality, and disability.[27]

Respiration Respiratory function must be evaluated immediately in all stroke patients. Ventilatory drive is usually intact except after medullary or massive hemispheric infarction or hemorrhage. The ability to protect the airway from aspiration may also be impaired, particularly in the acute setting. Intubation and mechanical ventilation may be necessary in these patients.[28] Most stroke patients do not require such aggressive maneuvers, but supplemental oxygen should be provided to maintain oxygen saturation above 95%.[29]

Blood pressure Maintenance of adequate blood pressure is vital for all patients. Cerebral blood flow to ischemic regions is dependent on cerebral perfusion pressure, which in turn is determined by the difference between mean arterial pressure and intracranial pressure (ICP). Elevated blood pressure is common at the time of initial stroke presentation, even in patients without chronic hypertension. Rapid lowering of blood pressure may further impair cerebral blood flow and worsen the ischemic injury.[30] Elevations in blood pressure will often spontaneously and gradually improve during the first few days after stroke.[30]

Antihypertensive therapy is indicated before and during thrombolysis with rt-PA; when infarction converts to hemorrhage; and in patients with myocardial ischemia, aortic dissection, or hypertensive encephalopathy. Candidates for thrombolysis should be treated only with modest measures (e.g., topical nitropaste or small intravenous boluses of labetalol) to maintain blood pressure below 185/110 mm Hg. Definitive treatment—including intravenous infusions of nicardipine, labetalol, or sodium nitroprusside—is appropriate for the other indications.

Fluid volume Careful volume replacement in patients with acute stroke improves cardiac output and cerebral perfusion.[31] Patients should receive isotonic saline to maintain euvolemia. Additional fluid administration (hypervolemic hemodilution) may increase cerebral blood flow while reducing blood viscosity without causing a reduction in oxygen delivery. However, clinical trials of hemodilution have yielded mixed results, with the largest of these trials showing no benefit.[32] Further investigation is required before hemodilution can be advocated.

Temperature Brain and body temperatures play an important role in outcome after stroke. Hypothermia improves outcome after ischemic injury in animals, after cardiac arrest, and after cardiac and neurosurgical procedures in humans, but it has not been adequately studied in acute stroke. Hypothermia may mitigate neurotransmitter toxicity, reduce neuronal metabolic

demands, and improve cerebral edema in acute stroke.[33] On the other hand, fever or even mild hyperthermia is known to be deleterious.[34] An elevation in body temperature dramatically increases the odds of severe disability or death, with risk doubling for each 1° C (1.8° F) above normal.[35] Normothermia should be maintained with antipyretics or cooling blankets, but therapeutic hypothermia must undergo more investigation before it can be recommended for focal ischemia.

Blood glucose levels Hyperglycemia appears to correlate with poor outcome in stroke.[36] Glucose may be metabolized to lactic acid, resulting in acidosis and increased tissue injury. Although the effect of correcting the blood glucose level is unknown, normalization is recommended. The administration of parenteral glucose should be minimized in patients with acute stroke.

Intracranial pressure In the most critically ill stroke patients, cerebral edema and elevations in ICP may complicate the clinical course. The expanding infarction may cause both focal and diffuse effects that typically peak at 2 to 5 days.[37,38] Large hemispheric stroke may result in malignant MCA syndrome, in which the edematous infarcted tissue compresses the anterior and posterior cerebral arteries, resulting in secondary infarctions.[37] Similarly, infarction of the cerebellum may lead to basilar artery compression and brain stem ischemia. Mortality in both malignant MCA syndrome and cerebellar infarction approaches 80% [*see* Intracerebral Hemorrhage, *below*]. Surgical decompression has a potential role in a minority of stroke patients. In acute cerebellar stroke, craniotomy with cerebellar resection is a lifesaving intervention that has become widely accepted.[39] Surgery removes the mass effect and prevents secondary brain stem and vascular compression. Malignant MCA syndrome may be similarly amenable to hemicraniectomy; this controversial approach is under investigation.

Preventive measures Prophylaxis for deep vein thrombosis should be instituted early with heparin (5,000 units given subcutaneously every 12 hours)[15] [*see 1:XVIII Venous Thromboembolism*]. For patients in whom heparin is contraindicated (e.g., patients with acute hemorrhage), pneumatic compression stockings are employed. Similarly, prevention of aspiration pneumonia should be a priority from the initial presentation at the hospital. Early intervention with physical therapy, occupational therapy, and speech therapy is important in recovery and prevention of complications.

PROGNOSIS AND RECOVERY

Initial stroke severity is one of the strongest predictors of outcome,[35] and early evidence of improvement is a good prognostic sign.[40] Recovery also depends on the size and location of the infarction or hemorrhage. Small infarctions, particularly subcortical lacunar strokes, may result in little chronic deficit, whereas large cortical infarctions may cause severe, permanent disability. Comorbid diseases, such as hypertension and diabetes, do not appear to affect recovery, but younger patients have a better prognosis than older patients.[40] Despite these predictors, the marked variability among patients makes early prognostication difficult. In general, recovery is greatest in the first 3 months after stroke.

The mechanisms of recovery after stroke remain poorly understood. Infarcted brain tissue is irreparable, so recovery of function has long been presumed to occur by recruitment of other neurons to serve new or additional roles. In rodents, neurons may be influenced to create new synapses after stroke.[41] Electrical brain mapping in monkeys has demonstrated that the cerebral cortex can be functionally reorganized during recovery after an infarction.[42] Similarly, functional MRI in humans has shown that activity in both hemispheres increases as patients improve, suggesting recruitment of neighboring cortex, as well as recruitment of the corresponding area of the contralateral cortex.[43]

Recovery may be improved by inducing these restorative mechanisms. Physical, occupational, and speech therapy are widely used, but no consensus exists regarding the optimal approach or timing of intervention. Neurotrophic growth factors and amphetamines may stimulate neuronal sprouting and also accelerate recovery by increasing the activity of uninjured neurons.[41] However, these pharmacologic approaches require extensive research before they can be advocated for routine use in stroke rehabilitation.

Table 4 Modifiable Risk Factors for Stroke

Risk Factor	*Prevention*	*Risk-Reduction Potential*	*Clinical Trial Evidence?*
Hypertension	P, S	25%–47%	Yes
Tobacco	P, S	?	No
Hyperlipidemia	P, S	24%–31%	Yes
Diabetes	P	?	No
Alcohol	P	50%	No
Exercise	P, S	?	No
Homocystine or homocysteine	P, S	?	No
Infection	P	?	No
Atrial fibrillation	P, S	68%	Yes
Symptomatic carotid stenosis > 70%*	S	65%	Yes
Asymptomatic carotid stenosis > 60%†	S	53%	Yes

*Two-year absolute risk reduction of 17%.
†Five-year absolute risk reduction of 6%.
P—primary S—secondary

ISCHEMIC STROKE PREVENTION

Reduction of Risk Factors

Numerous risk factors for stroke are modifiable. Reduction of factors such as excessive alcohol consumption, tobacco use, hypertension, and diabetes, as well as aggressive lipid control, contributes to stroke prevention [*see Table 4*].

Hypertension Hypertension has the highest population-attributable risk of any of the modifiable risk factors for stroke, and reduction in blood pressure has been shown to dramatically reduce stroke risk.[44] The reduction in risk of first stroke for those treated with antihypertensive agents is 25% to 47%.[45] Both diastolic and systolic hypertension have been linked to excess risk of stroke. Reducing isolated systolic hypertension even in the elderly has been shown to markedly lower stroke rates.[46] Patients who are undertreated with antihypertensive agents still have a higher stroke rate than those who are adequately treated.[47] Recommended actions to reduce the risk of stroke include (1) maintaining blood pressure below 140/90 mm Hg, (2) frequent checking of patients' blood pressure by physicians, and (3) at-home monitoring of blood pressure by patients.[44] Lowering diastolic blood pressure by 5 to 6 mm Hg can reduce stroke risk by 42%[44]; thus, hypertension should be assiduously diagnosed and treated. In the primary prevention of stroke and other major

vascular events, the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) showed that the thiazide diuretic chlorthalidone was effective and safe, and it is also inexpensive.[48] Comprehensive discussion about the management of hypertension for primary prevention is provided elsewhere [*see 1:III Hypertension*].

Although many antihypertensive agents can effectively reduce blood pressure, drugs that act upon the renin-angiotensin system appear to have a unique role in both primary and secondary stroke prevention. Two major studies demonstrated that angiotensin-converting enzyme (ACE) inhibitors—specifically, ramipril used alone[49] or perindopril used in combination with the diuretic indapamide[50]—reduce the risk of recurrent stroke by 25% to 30%. Furthermore, patients without a history of hypertension also seemed to benefit from the addition of an ACE inhibitor to their preventive regimen. A similar effect may exist for angiotensin receptor blockers (ARBs). In a trial comparing the beta blocker atenolol with the ARB losartan in patients with hypertension and left ventricular hypertrophy, there was a 25% reduction in the relative risk of a first stroke with the ARB.[51] In these trials, ACE inhibitors and ARBs appeared to decrease the risk of stroke more than would be expected by their relatively modest lowering of blood pressure; this suggests the possibility of an additional beneficial action via an uncertain mechanism.

Tobacco use Daily cigarette smoking has been shown to increase the risk of stroke by 250%.[52] A dose-effect response is seen in most studies. For those who smoke less than one pack a day, quitting reduces their risk to baseline within 5 years. For heavy smokers, the risk is greatly reduced but remains higher than that in individuals who never smoked. Switching from cigarettes to a pipe or cigars does not reduce stroke risk.[53]

Hyperlipidemia Evidence has emerged implicating hyperlipidemia as an independent risk factor for stroke, and studies have demonstrated impressive stroke risk reductions with statin agents. The Medical Research Council/British Heart Foundation Heart Protection Study (HPS), which included over 20,500 persons at high risk for coronary artery disease but who had characteristics that excluded them from previous statin studies, showed that long-term treatment with simvastatin (i.e., 40 mg daily for more than 5 years) reduced all strokes by 27%.[54] The study included patients with total cholesterol levels of 135 mg/dl or more; benefits were apparent even in patients with normal total cholesterol levels.

In an older study of patients who had had a previous myocardial infarction and whose cholesterol level was lower than 240 mg/dl, those given pravastatin had a 31% reduction in the risk of stroke, compared with patients given placebo.[55] In a study involving patients who had a median cholesterol level of 218 mg/dl and who received pravastatin, the relative risk reduction of stroke was 19%.[56] Meta-analyses also suggest that the use of statins reduces the risk of first stroke by 24% to 29%.[57,58] High-density lipoprotein cholesterol may be especially important in the risk of stroke. Statins may also exert their protective effects through mechanisms other than pure regulation of serum lipid concentrations. These effects include the processes involved in inflammation and thrombosis. In addition to pravastatin, simvastatin has an FDA indication for stroke prevention in patients with coronary artery disease and elevated cholesterol levels. Gemfibrozil was found to modestly reduce stroke in patients with coronary artery disease.[59]

Diabetes Although diabetes mellitus is a well-recognized risk factor for stroke, not all features of diabetes appear to contribute equally to that risk. Treatment of elevations in blood pressure and serum lipid levels appears to be more effective in reducing the risk of stroke, whereas treatment of hyperglycemia reduces the risk of microvascular complications.[60] At present, no studies directly link glucose control with reduction in risk of stroke. However, the international Action in Diabetes and Vascular Disease (ADVANCE) study is examining the effects of controlling blood pressure and blood glucose in patients with type 2 diabetes mellitus.[61]

Alcohol consumption In a quadratic model of the risk of stroke, there appears to be a J-shaped relation between alcohol intake and risk. In one report, those who consumed more than six drinks a day were at increased risk for stroke; those who consumed one to two drinks a day appeared to benefit from a protective effect of alcohol that reduced the risk of stroke by almost 50%.[62] There did not appear to be a difference associated with the type of alcoholic beverage consumed (i.e., wine, beer, or liquor). In another report, drinking as little as one alcoholic drink a week reduced the risk of stroke by 22%, compared with drinking no alcohol.[63]

Homocysteine level Elevated homocysteine levels are an independent risk factor for stroke.[64] Trials of high doses of vitamins B_6, B_{12}, and folic acid demonstrated that these agents can reduce homocysteine levels, but their use had no significant impact on stroke risk. Because vitamin supplementation is inexpensive and readily available, some clinicians use this approach empirically, though fortification of the United States food supply with folic acid may obviate this practice.[65]

Exercise Data from the Physician Health Study, a prospective cohort study of 21,823 men, demonstrated that exercise significantly reduced the risk of stroke, most likely by reducing other risk factors, including hypertension, lipid levels, and diabetes.[66]

Infection Infection and inflammation may be the single most important area of stroke research in the next decade. There have been several observations linking infection and inflammation with stroke. People with poor dentition seem to have a higher incidence of stroke. Although this increase may be related to other risk factors and access to care, it may also be related to oral infections. *Chlamydophila* (formerly *Chlamydia*) *pneumoniae* titers may be an independent risk factor for stroke.[67] *Chlamydophila* seems to promote thrombosis through effects on fibrinogen and vascular endothelium.

Therapeutic Measures to Prevent Ischemic Stroke

Management of risk of cardioembolism Disorders in many parts of the heart can potentially lead to stroke [*see Figure 4*]. Cardiac valves affected by bacterial endocarditis can give rise to septic emboli. The therapy for this condition is aggressive administration of antibiotics [*see 7:XVIII Infective Endocarditis*]. Akinetic ventricular segments can cause mural thrombi that in turn act as cardiac emboli. The treatment for this disorder is anticoagulation with warfarin to an international normalized ratio (INR) of 2.0 to 3.0. Similarly, anticoagulation after myocardial infarction is beneficial for patients with concomitant atrial fibrillation, decreased left ventricular function, or left ventricular throm-

bus.[44] Long-term warfarin therapy is necessary for patients with mechanical prosthetic valves. However, bioprosthetic valves require only brief anticoagulation, followed by antiplatelet treatment. Other conditions, such as patent foramen ovale, septal aneurysm, so-called ventricular smoke, and aortic arch atheroma, are more of a therapeutic dilemma. Although these conditions are known to increase stroke risk, there is still uncertainty as to whether antiplatelet or antithrombotic treatment is superior for stroke prevention. Indications for surgical intervention are also uncertain.

The most rigorously studied cardiac condition in terms of stroke prevention is fortunately the one most commonly related to stroke: atrial fibrillation [*see Table 4 and 1:IV Atrial Fibrillation*]. Nonvalvular atrial fibrillation is a common and readily preventable cause of stroke in the elderly.[68]

Lack of treatment of at-risk patients remains a significant public health challenge. The groups at highest risk for stroke include those with hypertension, diabetes mellitus, previous TIA or stroke, or poor left ventricular function and women older than 75 years. These patients should be treated with warfarin if they are appropriate candidates. For patients without risk factors and for those younger than 65 years, the risk of stroke is 1% a year without therapy. Thus, warfarin treatment is not necessary. For those individuals 65 to 75 years of age who are without risk factors, the yearly risk of stroke is 1.1% with warfarin therapy and 1.4% with aspirin therapy. Patients' INRs must be monitored and maintained in the 2.0 to 3.0 range so as to minimize the risk of ischemic stroke or hemorrhage from undertreatment or overtreatment.[69] An oral direct thrombin inhibitor, ximelagatran, has been shown to have efficacy and safety similar to those of warfarin in patients with atrial fibrillation and deep vein thromboembolism.[70,71] Ximelagatran offers several apparent advantages over warfarin, including minimal if any food or drug interactions, a high therapeutic index, rapid onset, and no need for INR monitoring. However, 6% of patients on this agent develop elevations in the transaminase level, which may limit its use.

Management of carotid artery disease Surgical treatment of symptomatic carotid stenosis (i.e., after a TIA or minor stroke) greatly reduces stroke risk.[72] In patients with a stenosis that is greater than 70% of the vessel diameter and who are good surgical candidates, surgery dramatically reduces the risk of stroke occurring within 2 years from 26% to 9%. For patients with 50% to 70% stenosis, the benefit is not as great. In this moderate-stenosis group, surgery reduces the risk of stroke over 5 years from 22% to 16%. This benefit is seen mostly in men, in those with recent stroke symptoms, and in those with hemispheric rather than ocular symptoms. Endarterectomy is most beneficial when it can be performed in the first 2 to 4 weeks after the initial cerebrovascular event.

Asymptomatic patients whose stenosis is less than 60% of the carotid diameter also benefit from surgery, but the risk and the results from surgery are more modest. In the Asymptomatic Carotid Atherosclerosis Study, surgery reduced stroke risk at 5 years from 11% to 5%. A surgical benefit was not seen in subgroup analysis for disabling stroke or death, and no benefit was seen for women.[73] Studies of community practice suggest that the risk of carotid endarterectomy may be higher than that seen in randomized clinical trials.[74] Many endarterectomies are done for asymptomatic patients in low-volume hospitals where perioperative morbidity and mortality are high.[74] For the beneficial effect of carotid endarterectomy in symptomatic patients to be realized, referring physicians should insist that surgeons provide objective evidence of surgical complication rates of no higher than 6%.[74] In asymptomatic carotid stenosis, surgical risk must be less than 3% because the combined risk of angiography and surgery in trials demonstrating surgical benefit was 2.4%.[74]

Carotid angioplasty and stenting (CAS) is being evaluated as an alternative to carotid endarterectomy. For patients who are at high risk for major complications during carotid endarterectomy, such as those with severe cardiac disease or prior carotid endarterectomy, CAS has been associated with better outcomes, both in the short term and long term.[75] However, for patients who would be candidates for carotid endarterectomy, there are no data from randomized clinical trials to support CAS except as part of a research protocol with appropriate informed consent from the patient.

Antiplatelet and antithrombotic treatment Inhibition of platelet activation can be achieved with several agents, including aspirin, dipyridamole, ticlopidine, and clopidogrel [*see Table 5*]. The role of aspirin in the primary prevention of stroke is uncertain, although it prevents myocardial infarction in high-risk patients.[76] However, aspirin is clearly indicated for secondary prevention of stroke in patients who have already experienced TIA or stroke. Numerous trials of antiplatelet therapy demonstrated that aspirin reduced the risk of nonfatal stroke by about 30%.[77] The ideal aspirin dose is controversial, and there is significant variability in patient responses. For the majority of patients, 50 to 325 mg a day (the dose range recommended by the FDA) appears to maximize the prophylactic effect and minimize the bleeding risk.

Other antiplatelet drugs offer modest additional preventive benefit, compared with aspirin, but at a greater cost and with more potential adverse effects. These medications are recommended for patients who are unable to tolerate aspirin or who have recurrent vascular events while on aspirin. Ticlopidine and clopidogrel reduce the risk of stroke by approximately 21%[78] and 7.3%,[79] respectively, compared with aspirin. Ticlopidine may cause significant neutropenia and thrombocytopenia and therefore requires complete blood count monitoring every 2 weeks for the first 3 months. Although early studies suggested a possible specific benefit for ticlopidine, as compared with aspirin, in African Americans, this was refuted by a large randomized trial.[80] Clopidogrel is associated with a lower frequency of neutropenia than ticlopidine, but thrombotic thrombocytopenic purpura has been reported.[81] Dipyridamole inhibits platelet phosphodiesterase activity and increases the availability of adenosine. Although early trials failed to demonstrate any benefit of dipyridamole, the European Stroke Prevention Study–2 compared regimens consisting of aspirin (50 mg), dipyridamole (extended release, 200 mg), both drugs in combination, and placebo and found a relative risk reduction with the combination of aspirin and dipyridamole of 23% when compared with aspirin alone.[82]

Oral anticoagulation with warfarin is used to inhibit the coagulation cascade and the formation of red blood cell thrombi, and it is appropriate for prevention in patients with high-risk sources of cardioembolism (see above). Warfarin was sometimes used empirically for patients with recurrent cerebrovascular events while on antiplatelet therapy, but in this setting warfarin may not be warranted. The Warfarin Aspirin Recurrent Stroke Study (WARSS) compared warfarin with aspirin in 2,000 patients with noncardioembolic strokes, most of which were lacunar strokes.

Table 5 Comparison of Antiplatelet Therapies for Prevention of Ischemic Stroke

Antiplatelet Drug	*Relative Risk Reduction**	*Major Side Effects*
Aspirin	(Reference drug)	Gastritis Peptic ulcer disease
Ticlopidine	21%	Neutropenia Diarrhea Rash
Clopidogrel	7.3%	Rash Diarrhea Thrombotic thrombocytopenic purpura
Dipyridamole + aspirin	23%	Headache

*Relative risk reduction for stroke.

The risk of recurrent stroke was the same in the two treatment arms, although the risk of hemorrhage tended to be slightly higher in the warfarin group.[83] Consequently, the role of warfarin in noncardioembolic strokes has become somewhat dubious. Further, in the randomized Warfarin versus Aspirin for Symptomatic Intracranial Disease (WASID) trial, both drugs had similar impact on the risk of recurrent stroke or death in patients with large-vessel intracranial stenosis, but the use of warfarin was associated with an increased risk of major bleeding complications.[84] Together, WARSS and WASID show that aspirin is at least as effective as warfarin, and in real-world practice, it is likely safer. Some stroke specialists use warfarin for acute internal carotid artery occlusion and arterial dissection, although controlled trials are lacking.

Contraindications to warfarin include pregnancy, poor compliance, alcohol abuse, and risk of falling. Long-term anticoagulation therapy is associated with a risk of major hemorrhage, which occurs at a rate of 1% to 3% a year.[85]

UNCOMMON CAUSES OF ISCHEMIC STROKE

Mechanisms other than cardioembolism, large-vessel atherothromboembolism, and small-vessel occlusive disease account for only a minority of all ischemic strokes, but they have specific diagnostic and therapeutic implications. Moreover, these unusual causes are disproportionately represented in young stroke victims, accounting for nearly one third of strokes in patients younger than 45 years.[86]

Atherosclerosis of the extracranial and intracranial arteries is a common cause of stroke, but several nonatherosclerotic disorders may also cause stroke. These disorders include inflammatory arteriopathies such as collagen vascular diseases, Takayasu disease, and neurovascular syphilis, as well as noninflammatory arteriopathies such as arterial dissection, fibromuscular dysplasia, moyamoya disease, CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy), radiation vasculopathy, and vasospasm after subarachnoid hemorrhage (SAH).

Arterial Dissection

Dissection of the internal carotid artery and vertebral artery can occur after head and neck trauma but may also occur spontaneously. Some connective tissue disorders may be risk factors, including fibromuscular dysplasia, Marfan syndrome, and Ehlers-Danlos syndrome. Arterial wall dissection causes vascular stenosis, occlusion, or a dissecting aneurysm. Clinical features include neck pain, headache, Horner syndrome, TIA or ischemic stroke, and tinnitus or audible bruits. Conventional angiography is the diagnostic gold standard and reveals the string sign, tapered stenosis or occlusion, dissecting aneurysm, intimal flap, distal pouch formation, and an underlying arteriopathy. Dissection may be diagnosed noninvasively with ultrasonography, CTA, MRI, or MRA, but each modality has potential limitations.

Prevention of stroke secondary to extracranial arterial dissection consists primarily of antithrombotic therapy. Heparin should be considered early because stroke risk is greatest in the first few days after the initial vascular injury. If anticoagulation is contraindicated, aspirin is recommended. Antithrombotic therapy should be continued until serial imaging demonstrates recanalization or stabilization of the dissected vessel.[87]

Inflammatory Arteriopathy

Inflammatory arteriopathies, or vasculitides, are a heterogeneous group of disorders in which vascular inflammation results in cerebral ischemia.[88] Vasculitis may be primary (e.g., isolated angiitis of the central nervous system) or secondary to infections (e.g., syphilis, tuberculosis, or varicella-zoster virus), toxins (e.g., cocaine, amphetamines, or LSD), neoplasms, or systemic inflammatory disorders (e.g., polyarteritis nodosa, Churg-Strauss angiitis, Wegener granulomatosis, giant cell arteritis, systemic lupus erythematosus, or rheumatoid arthritis). Symptoms may include headache, seizures, focal neurologic deficits, and multifocal encephalopathy. Clinical and serologic features of the vasculitides vary, but angiographic findings tend to be similar and nonspecific, showing segmental narrowing and dilatation ("beading"). Brain biopsy may confirm the diagnosis. Treatment should be directed at any underlying systemic disorder. Immunosuppressive regimens with corticosteroids and other agents are often used empirically.

Prothrombotic States

Ischemic stroke may be associated with hereditary and acquired prothrombotic states, including abnormalities of red cell or platelet function, of coagulation factors, or of endogenous fibrinolysis. Such disorders are uncommon but should be considered when no alternative etiology is identified.[86,89]

Cerebral Vein Thrombosis

Cerebral vein thrombosis (CVT) is a rare but important cause of stroke that is often missed or discovered late in diagnosis. Infection is the most common cause of CVT in children; in adults, most cases are associated with pregnancy. Infrequent etiologies include severe dehydration, sickle cell anemia, malignancy, and hypercoagulable states. Oral contraceptive agents have also been implicated. Severe headache, nausea, and vomiting are nonspecific but common symptoms. Papilledema, if present, may be the only abnormality on initial examination. Fluctuating focal neurologic deficits, such as unilateral weakness, numbness, or seizures, may appear. Lumbar puncture may demonstrate elevated protein levels, the presence of red blood cells, or xanthochromia. Noncontrast CT can determine whether acute hemorrhage or mass effect is present. Contrast CT may demonstrate a so-called empty delta sign in the sagittal sinus. MRI and MR venogram (MRV) have better sensitivity for detection of CVT; conventional cerebral angiography remains the diagnostic gold

standard. A small randomized trial demonstrated safety and efficacy of intravenous heparin in CVT, even in patients with preexisting hemorrhage.[90] Consequently, acute anticoagulation is recommended for most patients with CVT. Thrombolysis may be considered, but its role remains unconfirmed, and it should be reserved for patients whose condition declines despite adequate anticoagulation.[91,92] After the acute period, oral anticoagulation is typically used for several months until MRI or MRV demonstrates sinus patency. Mortality from CVT is estimated to be 5% to 30%, but survivors have a good prognosis with little residual deficit.[91]

Idiopathic Stroke

Cryptogenic (idiopathic) stroke is diagnosed when all indicated studies fail to identify the likely stroke mechanism. About half of strokes in young patients are diagnosed as cryptogenic.[86] In general, the risk of recurrent stroke in this context is believed to be relatively low.

Hemorrhagic Stroke

INTRACEREBRAL HEMORRHAGE

ICH accounts for 11% of stroke deaths. Prevalence among African Americans is notably higher than in the general population.[93] ICH cannot be reliably distinguished from ischemic stroke by clinical criteria alone. Noncontrast CT imaging establishes the diagnosis and is required to detect the presence of blood. The volume of ICH and the level of consciousness are the two most powerful predictors of outcome [*see Figure 2*]. Specific therapy for ICH remains largely an enigma. Patients with ICH frequently deteriorate as edema worsens over the first 24 to 48 hours. Late hematoma evacuation is ineffective in reducing mortality or improving outcome. Early surgical evacuation remains controversial. Theoretically, early hematoma evacuation may reduce surrounding ischemic injury and prevent edema formation and consequent herniation. Pilot studies of surgery within 12 hours[94] and 24 hours[95] suggested that early surgery is feasible and may be beneficial. In contrast, a trial comparing surgical evacuation with medical management for patients with ICH who appeared to be clinically stable (i.e., whose condition was not declining or who did not show signs of herniation) up to 48 hours after onset showed no difference in mortality or functional outcomes.[96]

Besides surgery, other strategies may be used to control ICP [*see Table 6*]. Intraventricular hemorrhage is a particularly bad complication of ICH. Prompt ventricular drainage should be performed if there are any signs of hydrocephalus. Ventricular drainage may reduce ICP without obvious hydrocephalus. Sedation with or without chemical neuromuscular paralysis is often helpful in controlling ICP in intubated patients. Propofol is an especially useful agent because of its short half-life. An ICP monitor should be used in all sedated patients and in patients who are obtunded (i.e., those whose Glasgow Coma Scale score is less than 9) to follow ICP. Osmotic diuretics are often useful in the short term before a definitive procedure such as hematoma evacuation is performed. The benefit of osmotic diuretics dissipates after 12 to 24 hours, but rebound elevations in ICP can occur. The same is true of hyperventilation, which lowers ICP by reducing cerebral blood flow through vasoconstriction. Although use of hyperventilation to reduce carbon dioxide tension (P_{CO_2}) to 30 to 35 mm Hg may transiently lower ICP, the effect is short-lived (6 to 12 hours), and rebound elevations in ICP are a serious problem if normocapnia is not slowly restored. Steroids are not recommended.

Table 6 Management Strategies for Elevated Intracranial Pressure

Strategy	*Comment*
Ventricular drainage	Most useful in hydrocephalus
Osmotic diuresis	Mannitol load, 0.5–1.0 g/kg I.V.; maintenance dose, 0.25–1.0 g q. 6 hr; titrate to keep serum osmolality 300 to 310 mOsm/kg H_2O
Hyperventilation	Titrate to keep P_{CO_2} 30–35 mm Hg; wean slowly
Sedation	Consider propofol or other benzodiazepine drip
Neuromuscular blockade	Always combine with sedation
Barbiturate coma	Rarely indicated

Blood pressure management in ICH also remains controversial. Observations suggest that about one third of ICHs expand during the first 24 hours.[97] Some investigators have therefore concluded that blood pressure should be lowered in patients with acute ICH. No trial has demonstrated that this action is necessary, and the concern about reducing cerebral perfusion pressure in patients with elevated ICP remains. The American Heart Association guidelines recommend only that mean arterial blood pressure be kept lower than 130 mm Hg in patients with a history of hypertension.[98]

A novel approach to preventing ICH expansion involves the use of procoagulants, and activated factor VII is currently under investigation. This approach may increase the risk of thromboembolic events, which are common complications in patients with ICH, so extreme caution is warranted until further data become available.

Further supportive care should probably be similar to those for patients with ischemic stroke [*see* Supportive Medical Management, *above*]. Premature withdrawal of these measures may worsen outcome.[99]

SUBARACHNOID HEMORRHAGE

Spontaneous SAH most commonly results from aneurysms of the circle of Willis. Aneurysms of the anterior and posterior communicating arteries are most frequently responsible [*see Figure 5*]. Hypertension and cigarette smoking are clear risk factors for aneurysmal rupture. A family history of SAH associated with ruptured intracranial aneurysms in first-degree relatives of patients with SAH is also a risk factor for aneurysm (unruptured aneurysms are detected in about 4% of such patients), but routine screening is not recommended.[100] The risk of rupture depends on aneurysm size. For patients with no history of SAH, the risk of rupture of aneurysms less than 7 mm in diameter is 0.05% a year. For aneurysms greater than 10 mm in diameter, the risk is slightly less than 1% a year. For aneurysms at least 25 mm in diameter, the risk jumps to 6% in the first year.[101]

Diagnosis

Up to 50% of patients with SAH present with a so-called warning leak or sentinel hemorrhage. Establishing the diagnosis early and consequent prompt aneurysm clipping can reduce long-term morbidity and mortality. Modern head CT imaging [*see Figure 2*] can establish the diagnosis in 97% of patients presenting to the emergency department with "the worst headache

of my life." In the remaining small percentage of patients, lumbar puncture and examination of the cerebrospinal fluid for xanthochromia are necessary.[102] In addition to severe headache, the following all suggest SAH and should prompt a thorough evaluation: rapid onset, photophobia, stiff neck, decreased level of consciousness, and focal neurologic signs.

Treatment

The treatment of SAH involves localizing the aneurysm with cerebral angiography and securing it to prevent subsequent bleeding. Traditionally, surgical clipping within 72 hours of onset has been recommended, although in patients with severe symptoms (coma), surgery is often delayed.[103] In a randomized trial involving 2,143 patients with ruptured intracranial aneurysms for whom both surgery or endovascular coiling were technically feasible, disability-free survival at 1 year was better with coiling.[104] However, these patients were very carefully selected, and there is considerable debate about how widely these results can be generalized. In practice, evaluations by both a neurosurgeon and an endovascular interventionalist are recommended.

Before aneurysm clipping or coiling, patients are kept mildly sedated in a quiet room and given stool softeners to reduce the risk of rebleeding. Anticonvulsants should be given at the first sign of seizure. Blood pressure is gently controlled. Although hypertension is related to rebleeding, some investigators believe that blood pressure works to tamponade the bleeding, and drastic reductions in blood pressure should be avoided.

Hydrocephalus is common after SAH and is very treatable with ventricular drainage. Any change in mental status should prompt the performance of an emergency CT scan to look for signs of hydrocephalus. Because it has been shown to improve outcome, nimodipine, a calcium channel blocker, is begun on the first day and continued for 21 days. After aneurysm clipping, the goal is to prevent vasospasm. Daily transcranial Doppler examinations are warranted. Patients should be well hydrated, and blood pressure should be slightly high. At the first clinical or transcranial Doppler sign of vasospasm, so-called triple H (hypertensive, hypervolemic, and hemodilution) therapy should be initiated to maximize cerebral blood flow, but only in patients with secured (i.e., clipped or coiled) aneurysms. Both colloid and crystalloid therapy are employed, and frequently, pressor support is needed.

Scott E. Kasner, M.D., participates in the speakers' bureaus for Boehringer Ingelheim, Bristol-Myers Squibb Co., and Genentech; is a consultant for Boehringer Ingelheim and AstraZeneca; and receives research grant support from AstraZeneca, Pfizer, Suntory, Ono, and NovoNordisk pharmaceutical companies.

Lewis B. Morgenstern, M.D., participates in the speakers' bureaus for Boehringer Ingelheim and Genentech.

References

1. Woo D, Gebel J, Miller R, et al: Incidence rates of first-ever ischemic stroke subtypes among blacks: a population-based study. Stroke 30:2517, 1999

2. Sacco RL, Boden-Albala B, Gan R: Stroke incidence among white, black and Hispanic residents of an urban community: the Northern Manhattan Stroke Study. Am J Epidemiol 147:259, 1998

3. Morgenstern LB, Spears WD, Goff DCJ: African Americans and women have the highest stroke mortality in Texas. Stroke 28:15, 1997

4. Gillum RF: Stroke mortality in blacks: disturbing trends. Stroke 30:1711, 1999

5. Frey JL, Jahnke HK, Bulfinch EW: Differences in stroke between white, Hispanic, and Native American patients: the Barrow Neurological Institute stroke database. Stroke 29:29, 1998

6. Morgenstern LB, Smith MA, Lisabeth LD, et al: Excess stroke in Mexican Americans compared with non Hispanic whites: the Brain Attack Surveillance in Corpus Christi Project. Am J Epidemiol 160:376, 2004

7. Davalos A, Castillo J, Martinez-Vila E, et al: Delay in neurological attention and stroke outcome. Stroke 26:2233, 1995

8. Menon SC, Pandey DK, Morgenstern LB: Critical factors determining access to acute stroke care. Neurology 51:427, 1998

9. von Kummer R, Nolte PN, Schnittger H, et al: Detectability of cerebral hemisphere ischaemic infarcts by CT within 6 h of stroke. Neuroradiology 38:31, 1996

10. Liebeskind DS: Rapid diagnostic evaluation. Prevention and Treatment of Ischemic Stroke. Kasner SE, Gorelick PB, Eds. Butterworth-Heinemann, Philadelphia, 2004, p 237

11. Alexandrov AV, Demchuk AM, Wein TH, et al: Yield of transcranial Doppler in acute cerebral ischemia. Stroke 30:1604, 1999

12. Adams HP, Bendixen BH, Kappelle LJ, et al: Classification of subtype of acute ischemic stroke: definitions for use in a multicenter clinical trial. Stroke 24:35, 1993

13. Johnston SC, Gress DR, Browner WS, et al: Short-term prognosis after emergency department diagnosis of TIA. JAMA 284:2901, 2000

14. Early risk of stroke after a transient ischemic attack in patients with internal carotid artery disease. North American Symptomatic Carotid Endarterectomy Trial Group. CMAJ 170:1105, 2004

15. The International Stroke Trial (IST): a randomised trial of aspirin, heparin, both, or neither among 19435 patients with acute ischaemic stroke. International Stroke Trial Collaborative Group. Lancet 349:1569, 1997

16. Randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischemic stroke. Chinese Acute Stroke Trial Collaborative Group. Lancet 349:1641, 1997

17. Coull BM, Williams LS, Goldstein LB, et al: Anticoagulants and antiplatelet agents in acute ischemic stroke: report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a division of the American Heart Association). Stroke 33:1934, 2002

18. Adams HP, Brott TG, Furlan AJ, et al: Guidelines for thrombolytic therapy for acute stroke: a supplement to the guidelines for the management of patients with acute ischemic stroke. Circulation 94:1167, 1996

19. Marler JR, Tilley BC, Lu M, et al: Early stroke treatment associated with better outcome: the NINDS rt-PA stroke study. Neurology 55:1649, 2000

20. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. N Engl J Med 333:1581, 1995

21. Kwiatkowski TG, Libman RB, Frankel M, et al: Effects of tissue plasminogen activator for acute ischemic stroke at one year. N Engl J Med 340:1781, 1999

22. Morgenstern LB, Bartholomew LK, Grotta JC, et al: Sustained benefit of a community and professional intervention to increase acute stroke therapy. Arch Intern Med 163:2198, 2003

23. Furlan A, Higashida R, Wechsler L, et al: Intra-arterial prourokinase for acute ischemic stroke. The PROACT II study: a randomized controlled trial. JAMA 282:2003, 1999

24. Combined intravenous and intra-arterial recanalization for acute ischemic stroke: the Interventional Management of Stroke Study. IMS Study Investigators. Stroke 35:904, 2004

25. Morgenstern LB, Pettigrew LC: Brain protection: human data and potential new therapies. New Horiz 5:397, 1997

26. Medical and neurological complications of ischemic stroke: experience from the RANTTAS trial. Randomized Trial of Tirilazad Mesylate in Acute Stroke (RANTTAS) Investigators. Stroke 29:447, 1998

27. Indredavik B, Bakke F, Slordahl SA, et al: Treatment in a combined acute and rehabilitation stroke unit: which aspects are most important? Stroke 30:917, 1999

28. Grotta JC, Pasteur W, Khwaja G, et al: Elective intubation for neurologic deterioration after stroke. Neurology 45:640, 1995

29. Vender RL: Respiratory management after stroke: what to do, what to avoid. J Crit Illness 7:1895, 1992

30. Lisk DR, Grotta JC, Lamki LM, et al: Should hypertension be treated after acute stroke? A randomized controlled trial using single photon emission computed tomography. Arch Neurol 50:855, 1993

31. Grotta JC, Pettigrew LC, Allen S, et al: Baseline hemodilution state and response to hemodilution in patients with acute cerebral ischemia. Stroke 16:790, 1985

32. Hemodilution in acute stroke: results of the Italian haemodilution trial. Italian Acute Stroke Study Group. Lancet 1:318, 1988

33. Maher J, Hachinski V: Hypothermia as a potential treatment of cerebral ischemia. Cerebrovasc Brain Metab Rev 5:277, 1993

34. Azzimondi G, Bassein L, Nonino F, et al: Fever in acute stroke worsens prognosis: a prospective study. Stroke 26:2040, 1995

35. Reith J, Jorgensen HS, Pedersen PM, et al: Body temperature in acute stroke: relation to stroke severity, infarct size, mortality, and outcome. Lancet 347:422, 1996

36. Gray CS, Hildreth AJ, Alberti GK, et al: Poststroke hyperglycemia: natural history and immediate management. Stroke 35:122, 2004

37. Hacke W, Schwab S, Horn M, et al: 'Malignant' middle cerebral artery territory infarction: clinical course and prognostic signs. Arch Neurol 53:309, 1996

38. Frank JI: Large hemispheric infarction, deterioration, and intracranial pressure. Neurology 45:1286, 1995

39. Rieke K, Schwab S, Krieger D, et al: Decompressive surgery in space-occupying hemispheric infarction: results of an open, prospective trial. Crit Care Med 23:1576, 1995

40. Jorgensen HS, Reith J, Nakayama H, et al: What determines good recovery in patients with the most severe strokes? Stroke 30:2008, 1999

41. Stroemer RP, Kent TA, Hulsebosch CE: Enhanced neocortical neural sprouting,

synaptogenesis, and behavioral recovery with d-amphetamine therapy after neocortical infarction in rats. Stroke 29:2381, 1998

42. Nudo RJ, Wise BM, SiFuentes F, et al: Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. Science 272:1791, 1996

43. Cramer SC, Nelles G, Benson RR, et al: A functional MRI study of subjects recovered from hemiparetic stroke. Stroke 28:2518, 1997

44. Gorelick PB, Sacco RL, Smith DB, et al: Prevention of a first stroke: a review of guidelines and a multidisciplinary consensus statement from the National Stroke Association. JAMA 281:1112, 1999

45. O'Brien AA, Rajkumar C, Bulpitt CJ: Blood pressure lowering for the primary and secondary prevention of stroke: treatment of hypertension reduces the risk of stroke. J Cardiovasc Risk 6:203, 1999

46. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. SHEP Cooperative Research Group. JAMA 265:3255, 1991

47. Klungel OH, Stricker BHC, Paes AH, et al: Excess stroke among hypertensive men and women attributable to undertreatment of hypertension. Stroke 30:1312, 1999

48. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). ALLHAT Collaborative Research Group. JAMA 288:2981, 2002

49. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 342:145, 2000

50. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. PROGRESS Collaborative Group. Lancet 358:1033, 2001

51. Dahlof B, Devereux RB, Kjeldsen SE, et al: Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. Lancet 359:995, 2002

52. Haheim LL, Holme I, Hjermann I, et al: Smoking habits and risk of fatal stroke: 18 years follow up of the Oslo study. J Epidemiol Community Health 50:621, 1996

53. Wannamethee SG, Shaper AG, Whincup PH, et al: Smoking cessation and the risk of stroke in middle-aged men. JAMA 274:155, 1995

54. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. Heart Protection Study Collaborative Group. Lancet 360:7, 2002

55. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. Cholesterol and Recurrent Events Trial investigators. N Engl J Med 335:1001, 1996

56. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. The Long-Term Intervention with Pravastatin in Ischaemic Disease (LIPID) Study Group. N Engl J Med 339:1349, 1998

57. Hebert PR, Gaziano JM, Chan KS, et al: Cholesterol lowering with statin drugs, risk of stroke, and total mortality. JAMA 278:313, 1997

58. Bucher HC, Griffith LE, Guyatt GH: Effect of HMG-CoA reductase inhibitors on stroke: a meta-analysis of randomized, controlled trials. Ann Intern Med 128:89, 1998

59. Rubins HB, Robins SJ, Collins D, et al: Gemfibrozil for the secondary prevention of coronary heart disease in men with low levels of high-density lipoprotein cholesterol. N Engl J Med 341:410, 1999

60. Yki-Jarvinen H: Management of type 2 diabetes mellitus and cardiovascular risk: lessons from intervention trials. Drugs 60:975, 2000

61. Study rationale and design of ADVANCE: action in diabetes and vascular disease—preterax and diamicron MR controlled evaluation. Diabetologia 44:1118, 2001

62. Sacco RL, Elkind M, Boden-Albala B, et al: The protective effect of moderate alcohol consumption on ischemic stroke. JAMA 281:53, 1999

63. Berger K, Ajani UA, Kase CS, et al: Light-to-moderate alcohol consumption and the risk of stroke among U.S. male physicians. N Engl J Med 341:1557, 1999

64. Kittner SJ, Giles WH, Macko RF, et al: Homocyst(e)ine and risk of cerebral infarction in a biracial population: the stroke prevention in young women study. Stroke 30:1554, 1999

65. Kasner SE: Folate and the risk of stroke: fortify first and ask questions later? Stroke 33:1188, 2002

66. Lee IM, Hennekens CH, Berger K, et al: Exercise and risk of stroke in male physicians. Stroke 30:1, 1999

67. Fagerberg B, Gnarpe J, Gnarpe H, et al: *Chlamydia pneumoniae* but not cytomegalovirus antibodies are associated with future risk of stroke and cardiovascular disease: a prospective study in middle-aged to elderly men with treated hypertension. Stroke 30:299, 1999

68. Hart RG, Benavente O, McBride R, et al: Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 131:492, 1999

69. Ezekowitz MD, Levine JA: Preventing stroke in patients with atrial fibrillation. JAMA 281:1830, 1999

70. Schulman S, Wahlander K, Lundstrom T, et al: Secondary prevention of venous thromboembolism with the oral direct thrombin inhibitor ximelagatran. N Engl J Med 349:1713, 2003

71. Olsson SB, Executive Steering Committee on behalf of the SPORTIF IIII: Stroke prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation (SPORTIF III): randomised controlled trial. Lancet 362:1691, 2003

72. Endarterectomy for symptomatic carotid stenosis in relation to clinical subgroups and timing of surgery. Carotid Endarterectomy Trialists Collaboration. Lancet 363:915, 2004

73. Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. JAMA 273:1421, 1995

74. Biller J, Feinberg WM, Castaldo JE, et al: Guidelines for carotid endarterectomy: a statement for healthcare professionals from a special writing group of the Stroke Council, American Heart Association. Stroke 29:554, 1998

75. Stenting and Angioplasty with Protection in Patients at High Risk for Endarterectomy (SAPPHIRE) study. (publication pending)

76. Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. N Engl J Med 321:129, 1989

77. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. Antiplatelet Trialists' Collaboration. BMJ 324:141, 2001

78. Hass WK, Easton JD, Adams HP, et al: A randomized trial comparing ticlopidine hydrochloride with aspirin for the prevention of stroke in high-risk patients. N Engl J Med 321:501, 1989

79. A randomised, blinded trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). CAPRIE Steering Committee. Lancet 348:1329, 1996

80. Aspirin and ticlopidine for prevention of recurrent stroke in black patients: a randomized trial. African American Antiplatelet Stroke Prevention Study Investigators. JAMA 289:2947, 2003

81. Nara W, Ashley I, Rosner F: Thrombotic thrombocytopenic purpura associated with clopidogrel administration: case report and brief review. Am J Med Sci 322:170, 2001

82. Diener HC, Cunha L, Forbes C, et al: European Stroke Prevention Study: 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. J Neurol Sci 143:1, 1996

83. A comparison of warfarin and aspirin for the prevention of recurrent ischemic stroke. Warfarin-Aspirin Recurrent Stroke Study Group. N Engl J Med 345:1444, 2001

84. Chimowitz M, Lynn M, Howlett-Smith H, et al: Warfarin-Aspirin Symptomatic Intracranial Disease (WASID) trial: final results (abst). Stroke 35:235, 2004

85. Bleeding during antithrombotic therapy in patients with atrial fibrillation. The Stroke Prevention in Atrial Fibrillation Investigators. Arch Intern Med 156:409, 1996

86. Bogousslavsky J, Pierre P: Ischemic stroke in patients under age 45. Neurol Clin 10:113, 1992

87. Kasner SE, Hankins LL, Bratina P, et al: Magnetic resonance angiography demonstrates vascular healing of carotid and vertebral artery dissections. Stroke 28:1993, 1997

88. Ferro JM: Vasculitis of the central nervous system. J Neurol 245:766, 1998

89. Adams HP Jr, Kappelle LJ, Biller J, et al: Ischemic stroke in young adults: experience in 329 patients enrolled in the Iowa Registry of Stroke in Young Adults. Arch Neurol 52:491, 1995

90. Einhaupl KM, Villringer A, Meister W, et al: Heparin treatment in sinus venous thrombosis. Lancet 338:597, 1991

91. Bousser M-G, Ross-Russell R: Cerebral Venous Thrombosis. WB Saunders Co, London, 1997

92. Kasner SE, Gurian JH, Grotta JC: Urokinase treatment of sagittal sinus thrombosis with venous hemorrhagic infarction. J Stroke Cerebrovasc Dis 7:421, 1998

93. Morgenstern LB, Spears WD: A triethnic comparison of intracerebral hemorrhage mortality in Texas. Ann Neurol 42:919, 1997

94. Morgenstern LB, Frankowski RF, Shedden P, et al: Surgical treatment for intracerebral hemorrhage (STICH): a single-center, randomized clinical trial. Neurology 51:1359, 1998

95. Zuccarello M, Brott T, Derex L, et al: Early surgical treatment for supratentorial intracerebral hemorrhage: a randomized feasibility study. Stroke 30:1833, 1999

96. The International Surgical Trial in Intracerebral Haemorrhage. Lancet (in press)

97. Brott T, Broderick J, Kothari R, et al: Early hemorrhage growth in patients with intracerebral hemorrhage. Stroke 28:1, 1997

98. Broderick JP, Adams HP, Barsan W, et al: Guidelines for the management of spontaneous intracerebral hemorrhage. Stroke 30:905, 1999

99. Becker KJ, Baxter AB, Cohen WA, et al: Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies. Neurology 56:766, 2001

100. Risks and benefits of screening for intracranial aneurysms in first-degree relatives of patients with sporadic subarachnoid hemorrhage. The Magnetic Resonance Angiography in Relatives of Patients with Subarachnoid Hemorrhage Study Group. N Engl J Med 341:1344, 1999

101. Unruptured intracranial aneurysms: risks of rupture and risks of surgical intervention. The International Study of Unruptured Intracranial Aneurysms Investigators. N Engl J Med 339:1725, 1998

102. Morgenstern LB, Luna-Gonzales H, Huber JC Jr, et al: Worst headache and subarachnoid hemorrhage: prospective, modern computed tomography and spinal fluid analysis. Ann Emerg Med 32:297, 1998

103. Mayberg MR, Batjer HH, Dacey R, et al: Guidelines for the management of aneurysmal subarachnoid hemorrhage. Stroke 25:2315, 1994

104. Molyneux A: International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. Lancet 360:1267, 2002

Acknowledgment

Figures 2b, 3a, 4, and 5 Seward Hung.

V TRAUMATIC BRAIN INJURY

Andres M. Salazar, m.d.

Traumatic brain injury (TBI) is the leading cause of death and disability in young adults in the United States; the total national cost of TBI is more than $39 billion a year.[1] Elucidation of the pathophysiology of brain injury is an important challenge for physicians who care for TBI patients. Equally important are prevention and a better scientific understanding of recovery and rehabilitation.

TBI is now generally viewed as a multidimensional, dynamic process. It is not unusual for a patient with TBI who initially is relatively stable and awake or in a light coma to deteriorate rapidly. Delayed hematoma or expanding contusions that are amenable to surgery account for many such cases. Others are related to uncontrolled brain swelling that may not respond to conventional management. Delayed secondary injury at the cellular level is also a major contributor to brain swelling and tissue loss after TBI. The ultimate pathologic picture thus evolves during the first few hours and days after trauma, and the physiologic, clinical, and behavioral aspects of recovery can continue for years. For these reasons, therapy should be predicated on an understanding of the multidimensional pathology of TBI and its evolution.

TBI is traditionally classified by its severity [*see Table 1*], though the current definitions are imperfect and distinctions between mild, moderate, and severe head injury can be difficult to make in the acute period. For example, acute management of an unconscious patient with a moderate head injury may differ little from that of a comatose patient with a more severe head injury; or a patient with little or no initial loss of consciousness may harbor a more serious and even life-threatening pathology, such as a delayed hematoma. Nevertheless, the distinctions are generally useful in guiding the approach to the patient.

Etiology

Brain injury may be caused by any of several types of head trauma, including the more typical closed head injury (in which rapid acceleration or deceleration causes the brain to strike the inside of the skull), direct impact to the head, or penetration by a missile or other foreign object. Although some details of the pathology of these types of trauma may differ, acute and long-term management are similar in most cases.

Pathogenesis

The pathologic changes that occur in TBI may result less from the injury itself than from an uncontrolled vicious circle of biochemical and physiologic events set in motion by the trauma. These biochemical events include changes in arachidonic acid metabolites (e.g., the prostaglandins); the formation of oxygen free radicals and lipid peroxidation; and changes in electrolytes (e.g., calcium and magnesium) in excitotoxic neurotransmitters (e.g., glutamate) and changes in various kinins and cytokines.[2] These events can result in progressive injury to otherwise viable brain tissue by altering vascular reactivity and producing further ischemia, by producing brain swelling (hyperemia, edema, or both), by injuring neurons and glia directly, or by activating macrophages that cause neuronal and glial injury. The patient with severe TBI also often has multiple systemic abnormalities—such as changes in nutrition, cardiopulmonary status,[3] circulating catecholamines, and coagulation[4]—that may be directly related to the brain injury and may have a profound impact on treatment.

At least five parallel components of the pathology of closed head injury have been identified: (1) focal hematomas and contusions, (2) diffuse axonal injury, (3) diffuse microvascular injury with loss of autoregulation and acute brain swelling, (4) hypoxia-ischemia, and (5) selective neuronal loss (especially of thalamic reticular, hippocampal, and cerebellar neurons), possibly caused by excitotoxins. In addition, recent electrophysiologic evidence suggests that a diffuse neuronal (gray matter) dysfunction may be the most subtle and sensitive measure of mild TBI and concussion [*see* Diffuse Gray Matter Dysfunction, *below*]. Each component may have a different effect on the patient, depending on the patient's premorbid status, the severity of the injury, the treatment given, and the time that has elapsed since injury.

Some of these pathologic processes, such as focal hematomas and microvascular injury with brain swelling, can result in the death of the patient soon after injury; others, such as diffuse axonal injury and excitotoxic injury, may result principally in the death of neuronal groups and thus have implications for long-term function. Evaluation of the clinical efficacy of specific therapies therefore depends on the particular pathology targeted by the treatment. For example, survival after TBI may be a good outcome measure of the efficacy of agents designed to limit subacute brain swelling, but it may not be a good measure for evaluating potentially neuron-sparing agents, such as certain glutamate antagonists and neurotrophins.

Focal Injury

Focal injuries include intracerebral and extracerebral hematomas and focal contusions. Hematomas are most common after the rapid acceleration or deceleration that occurs as

Table 1 Severity of Traumatic Brain Injury

Severity	*Admission GCS Score*	*Duration of Unconsciousness*	*Duration of Posttraumatic Amnesia*	*CT/MRI*
Uncomplicated mild	13–15	0–20 min	< 24 hr	Normal
Complicated mild	13–15	0–20 min	< 24 hr	Abnormal
Moderate	9–12	< 24 hr	> 24 hr	Usually abnormal
Severe	3–8	> 24 hr	Weeks	Abnormal

GCS—Glasgow Coma Scale

the result of a fall or another form of impact, especially in the elderly. Delayed hematomas, which can occur in patients who are initially at low risk but whose condition deteriorates rapidly, are particularly important. Small hematomas can be treated conservatively, but delaying the surgical removal of large hematomas for longer than 4 hours after injury significantly increases mortality and morbidity.

Focal contusions may occur under the site of impact, but by far the most common locations after acceleration-deceleration injury are in the orbitofrontal and anterior temporal lobes, where the brain abuts the base of the skull. The most troubling clinical sequelae are usually behavioral and cognitive abnormalities referable to these areas of the brain. Contusions can undergo secondary expansion or result in delayed hematomas. Patients with such injuries require particularly close observation in the acute period. Both hematomas and contusions are also significant risk factors for the development of posttraumatic epilepsy [*see 11:XII Epilepsy*].

DIFFUSE AXONAL INJURY

Diffuse axonal injury—a shearing injury of axons in the hemispheric white matter, corpus callosum, and brain stem[5]—is a significant cause of persistent, severe neurologic deficits in closed-head injury. When severe, the injury is manifested clinically by immediate and prolonged loss of consciousness. Petechial hemorrhages in the white matter or blurring of the gray matter–white matter junction is best seen on magnetic resonance imaging, especially in coronal slices. However, the only early abnormality may be microscopic focal cytoskeletal disruption. These changes lead to disturbance of axonal flow and the subsequent severing of axons, with the typical light microscopic picture appearing 12 to 24 hours later. If severe enough, such axonal injury can lead to wallerian degeneration and diffuse target deafferentation.[6,7] It is possible that medical treatments can prevent total axonal disruption. Certain neurotrophins, such as brain-derived neurotrophic factor and perhaps insulinlike growth factor, may attenuate it.[8]

Such axonal injury may also occur even after mild TBI (MTBI) and in the absence of morphopathologic change in any other vascular, neural, or glial element. Some of the cognitive changes seen after MTBI may relate to diffuse axonal injury.

DIFFUSE MICROVASCULAR DAMAGE

Diffuse microvascular damage is a major component of both closed and penetrating TBI. Depending on the severity of the injury, early changes may include loss of cerebrovascular autoregulation, with decreased responses to changes in carbon dioxide and perfusion pressure, and transient systemic hypertension. The loss of autoregulation makes the brain particularly susceptible to fluctuations in systemic blood pressure; otherwise tolerable hypotension can thus result in cerebral ischemic damage in the patient with TBI. In addition, altered vascular sensitivity to circulating catecholamines can lead to vasoconstriction and further focal ischemia or reperfusion injury. The microvascular pathology includes an endothelial change that probably involves oxygen free radical–induced decreases in endothelial nitric oxide, with a concomitant vasoconstriction, and an initial hyperglycolysis with a dissociation of cerebral blood flow and metabolism.[9] Positron emission tomography has demonstrated these metabolic changes even in patients with MTBI.

Free radicals, including the superoxide radical peroxynitrite, and the process of lipid peroxidation play a critical role in secondary injury, not only by their effect on the microvasculature but also by their direct effects on tissue. The pharmacologic agents used to reduce the formation of free radicals or to scavenge those already formed include steroids to inhibit lipid peroxidation; α-tocopherol (vitamin E) and its analogues; α-lipoic acid; iron chelators, such as deferoxamine; and enzymes such as superoxide dismutase. However, in phase III clinical studies, two such compounds—the 21-aminosteroid tirilazad mesylate and a polyethylene glycol–conjugated superoxide dismutase—failed to provide clear benefit to patients with acute, severe TBI.[10,11]

HYPOXIA-ISCHEMIA

The classic pathology of hypoxia-ischemia primarily involves the hippocampus and the vascular border zones of the brain. It is often superimposed on other, more specific pathologies of TBI. The traumatized brain is particularly sensitive to hypoxia-ischemia, possibly because of the metabolic demands already placed on neurons by the trauma itself[12] or by increasing vascular permeability.[13] The most significant improvements in the survival of patients with TBI have resulted from recognition of the importance of this component and its prevention, largely through training of paramedics, the development of emergency transport systems, and immediate resuscitation protocols.[14]

SELECTIVE NEURONAL VULNERABILITY, EXCITOTOXIC INJURY, AND NEURONAL ENERGY FAILURE

Selective vulnerability of certain neuronal groups, including hippocampal and thalamic reticular neurons that receive glutaminergic afferents from the orbitofrontal cortex, occurs after head injury and appears to be caused by glutamate excitotoxicity.[15] It also occurs after mild head injury in animal models and may be a cause of the fatigue, attention, and memory problems often seen in postconcussion syndrome in humans.

Excitotoxic injury may be one of the most important mechanisms of neuronal death after traumatic or ischemic injury. Excessive release of glutamate and other neurotransmitters unleashes a chain of cellular events that deplete neuronal energy stores, damage mitochondria, and result in cell death or apoptosis. Thus, glutamate antagonists or other neuroprotectants, such as dextrorphan, riluzole, memantine, and magnesium, may play a role in the acute treatment of TBI. However, because of the importance of glutamate in the brain, receptor blockade is usually accompanied by intolerable side effects; a number of clinical trials with such agents have failed to show a clear benefit.[16] Alternatively, therapies that prevent depletion of neuronal metabolic stores or that enhance neuronal stores have also been shown to protect cells in models of glutamate toxicity. One related neuroprotective strategy that reduces glutamate release and appears to protect cellular energy metabolism is moderate systemic hypothermia. A controlled pilot study showed long-term benefit from treatment with hypothermia for patients surviving severe TBI.[17] However, a larger multicenter study failed to confirm these findings.[18]

A more direct approach may be to use agents that enhance neuronal energy metabolism and mitochondrial function after injury. Because it addresses a "final common path" in neuronal dysfunction, death, or both, this approach may protect against various stressors, including excitotoxic, oxidative, or calcium-induced injury.[19] Depletion of neuronal energy can result because of the increased demands placed on neurons and their membrane pumps by the injury and because of a failure of adenosine triphosphate (ATP) production.

Mitochondrial dysfunction has been demonstrated after brain injury. One mechanism of mitochondrial dysfunction is the fail-

ure of the mitochondrial membrane, with consequent release of cytochrome-c into the cytoplasm and the subsequent activation of caspases and the apoptotic cascade. Similarly, it has been postulated that mitochondrial failure, including failure of the pyruvate dehydrogenase pathways, may underlie the demonstrated uncoupling of blood flow and metabolism after TBI.[19]

Cyclosporine has been demonstrated to have a neuroprotective effect, which is achieved through stabilization of the mitochondrial membrane.[20,21] Both creatine and the three-carbon sugar pyruvate have been shown to have marked neuroprotective effects in animal models of TBI. As with cyclosporine, this effect is achieved through stabilization of the mitochondrial membrane.[22,23] Pyruvate is not only the primary energy substrate in neuronal mitochondria but also a good scavenger of oxygen free radicals. Both pyruvate and creatine are inexpensive and nontoxic; they are in early clinical trials for the treatment of brain injury.

DIFFUSE GRAY MATTER DYSFUNCTION

In addition to selective neuronal vulnerability, recent evidence from quantitative electroencephalograhic and quantitative MRI studies suggests that a very common effect of TBI may be a diffuse gray matter dysfunction that manifests itself primarily through changes in brain electrical activity, as measured by EEG coherence, phase, and power.[24-27] These alterations may, in turn, reflect a relative loss of neuronal membrane electrical efficiency, probably as a consequence of the failure of neuronal energy metabolism and the ATP-driven neuronal ion pumps. Such failure would not be unexpected in the face of a probable diffuse excitotoxic challenge or other challenges in the early period after TBI. These alterations may be the only physiologic or pathologic evidence of MTBI or concussion. Although the pathology of this dysfunction is likely to be very subtle, there is a strong correlation between these changes and changes in the T_2-weighted MRI signal in the gray matter, which in turn is thought to reflect the functional integrity of neuronal membranes (T_2 refers to spin-spin, or transverse, relaxation time). These EEG changes have also been demonstrated to correlate with neuropsychological performance, suggesting that they could also be responsible for some of the cognitive changes that occur after MTBI. Restoration of neuronal energy metabolism might be expected to ameliorate these cognitive changes.

Mild Traumatic Brain Injury

With an incidence of 180 per 100,000 people, MTBI is more common than any other neurologic diagnosis except migraine. MTBI is variably defined as any TBI/concussion with loss of consciousness of 0 to 30 minutes, a Glasgow Coma Scale (GCS) score of 13 to 15 on admission [*see Table 2*], posttraumatic amnesia or confusion lasting less than 24 hours, and no evidence of contusion or hematoma on CT. Concussion can be further divided into grades I, II, and III. Grade I concussion is characterized by transient mental changes lasting longer than 15 minutes, with no loss of consciousness; in grade II concussion, transient mental changes last longer than 15 minutes, and there is no loss of consciousness; and grade III concussion is characterized by brief loss of consciousness. Although these distinctions, especially the distinction between grade I and grade II concussion, can be relatively subtle, they can serve as a useful guide for return to normal activity, especially in athletes or other active individuals.[28] In any case, MTBI, even without loss of consciousness, has been repeatedly associated with measurable abnormalities in cognition, attention, and behavior, as well as documented quantitative EEG and neuropathologic changes.[24,29] Abnormalities seen on assessments of cognitive task have been repeatedly documented after MTBI; these abnormalities usually include disturbances of attention, information processing, and memory.[30-33] As might be expected, MTBI also has a significant psychosocial impact.

Over 75% of MTBI patients report some somatic or cognitive symptoms over the first several weeks after injury; these can have important functional, social, and economic implications. Symptoms include headache, dizziness or vertigo, blurred vision, fatigue, sleep disturbance, irritability, depression, anxiety, and poor memory and concentration. Typically, these symptoms improve steadily and are largely cleared after the first 3 months after injury. However, some symptoms, especially the emotional symptoms, can persist longer. The term postconcussion syndrome is often applied when this complex of symptoms is persistent.[30]

There has been increasing attention paid to MTBI in sports.[34] The study of MTBI in athletes offers several advantages, including the generally high preinjury health and motivation of athletes, the ability to conduct preinjury testing, and the relative predictability of the time of the injury. In one study, college football players were examined before and after injury; significant attention deficits were found to persist for as long as 5 days after a minor "ding" that was not associated with loss of consciousness.[35] Similar findings have been reported in soccer players.[36]

PATHOGENESIS

The pathology associated with MTBI or concussion is still unclear, but some evidence suggests that these injuries may be associated with a diffuse cortical neuronal dysfunction; selective vulnerability of certain neurons and a modest amount of diffuse axonal injury may also be factors. Microvascular injury with alterations in autoregulation and uncoupling of blood flow and metabolism has also been described in MTBI, especially with repeated MTBI—the so-called second impact syndrome—in pediatric and adolescent patients.[9,37,38] Finally, hematomas occur with some frequency in patients who might otherwise be classified as having MTBI.[39-41]

Table 2 Glasgow Coma Scale[92]

Test	*Response*	*Score*
Eye opening	Spontaneous	4
	To speech	3
	To pain	2
	None	1
Best verbal response	Oriented	5
	Confused	4
	Inappropriate	3
	Incomprehensible	2
	None	1
Best motor response (arm)	Obedience to commands	6
	Localization of pain	5
	Withdrawal response to pain	4
	Flexion response to pain	3
	Extension response to pain	2
	None	1

a *b*

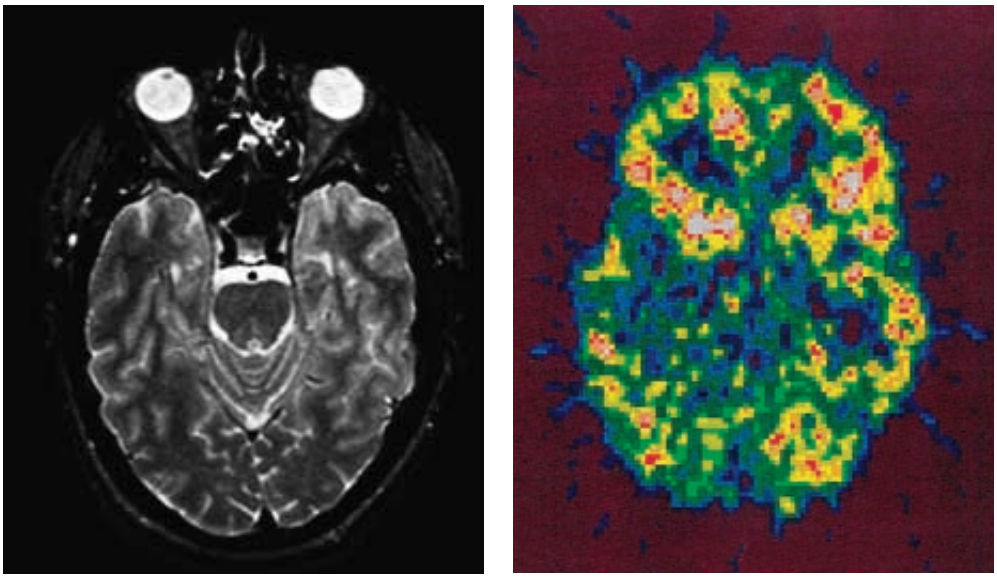


Figure 1 **(*a*) MRI in a 25-year-old man with mild to moderate traumatic brain injury and residual irritability, lability of affect, aggressivity, and occasional dyscontrol episodes initially seems normal but on closer scrutiny shows blurring of the gray matter–white matter junction in the left temporal lobe. (*b*) Positron emission tomography confirms decreased metabolism in the left temporal lobe.**

EVALUATION AND ACUTE MANAGEMENT

Although the emergency department evaluation and management of MTBI is controversial, the principal concern is with identifying evolving surgical lesions such as hematomas and contusions. The cause of injury can be a factor in management; for example, motor vehicle accidents involving large forces or impact to the head may raise the likelihood of hematomas, as do falls in older patients, especially if the patient is taking anticoagulants or aspirin. Documentation of the history of the injury, as well as the length of the period of unconsciousness, mental confusion, or both, can be very helpful in both acute and long-term management. In the past, many patients with traumatic loss of consciousness were admitted to the hospital for overnight observation, especially if there was no responsible adult to be with the patient at home. However, changes in managed care over the past several decades and the increased availability of computed tomography have led to a marked decrease in such admissions.[42]

In addition to history and examination, CT has become the mainstay of evaluation, to the exclusion of skull x-ray. Prolonged or deteriorating mental status or the presence of neurologic signs or other risk factors are still clear indications for CT scanning, observation, or both after MTBI. For example, MTBI patients whose GCS scores are 13 or 14 on admission have a much higher incidence (up to 28%) of abnormal findings on CT scanning than do patients with a GCS score of 15.[43] In addition, on rare occasions, even MTBI patients whose initial CT scans are negative may develop surgical complications after discharge; repeat CT scanning should thus be considered in patients who return with severe, persistent symptoms or new neurologic signs.

The diagnosis of contusion is also important for its longer-term prognostic value. Patients with MTBI complicated by cerebral contusion have a 6-month outcome that is more consistent with that of patients with moderate head injury[39] and are thus candidates for more intensive, longer-term observation and management.

In the postacute period, MRI, single-photon emission computed tomography (SPECT), and quantitative EEG can provide additional documentation and localization of brain injury that can be very valuable in guiding nonsurgical management [*see Figure 1*]. For example, MRI can help provide a presumptive diagnosis of diffuse axonal injury and subtle brain contusions that might have been be missed on CT. Standard neuropsychological testing is usually not indicated in the MTBI patient, but some brief, specialized cognitive batteries that have been developed for this population can be very helpful in diagnosis and follow-up. These include the Sideline Assessment of Concussion (SAC) and the Automated Neuropsychological Assessment Metric (ANAM).[28,29,44-46]

MANAGEMENT OF MTBI

Postconcussive symptoms, which include headache, dizziness, fatigue, and documented deficits in cognition, can be seen even after mild injuries without loss of consciousness. In general, the prognosis for recovery is very good, with most cognitive and somatic sequelae improving markedly by 3 months; 85% of patients experience no disabling symptoms 1 year after injury.[40,47] In the small percentage of patients who have postconcussive complaints and disability over periods exceeding 1 year, psychogenic factors can often contribute to the persistence of symptoms. Patients who have persistent symptoms of anxiety, depression, or both need appropriate diagnosis and treatment, preferably by a psychiatrist who has experience with TBI.

Probably the most important element in the longer-term management of patients with MTBI is the clinician's recognition that the postconcussive symptoms in these patients have a structural cause and that the patients usually recover. Attention deficits and fatigue appear to be especially common and troubling, and there is nothing more disconcerting to the intelligent but symptomatic MTBI patient than to be told by his physician, family, or employer that there is nothing wrong. Early symptomatic management and counseling as to what to expect can help avoid the all too common delayed emotional symptoms of anxiety and depression, especially in patients who may be at risk because of underlying psychopathology. Recognition that the concussed patient is not always able to understand oral counseling in the emergency department has led some physicians to use information booklets that expand and reinforce initial recommendations. Recent controlled studies have convincingly demonstrated that the use of such booklets significantly reduces long-term morbidity, particularly with regard to emotional symptoms.[48]

Moderate and Severe Head Injury

INITIAL EVALUATION AND RESUSCITATION

An organized team approach to the acute management of the unconscious patient with TBI is essential and includes prehospital, intensive care unit, and post-ICU care. A cornerstone of early evaluation and care is recognition of deterioration in patient status through the sequential use of a standardized measure such as the GCS score [*see Figure 2*], along with checks of lateralized deficits in neurologic function and careful attention to pupillary responses. The history should be obtained from witnesses, particularly with regard to the onset of coma. For example, if a patient who is comatose had an initial interval of lucidity or semilucidity, an expanding mass lesion may be present, and severe diffuse axonal injury is less likely.[49]

The importance of cardiopulmonary resuscitation in patients with acute TBI cannot be overstated.[50] Airway and shock management should be the first priority in all trauma patients. The loss of cerebrovascular autoregulation places the brain at increased risk for cerebral ischemia from systemic hypotension, and levels of hypercapnia tolerated by the normal brain can lead to critical marginal increases in intracranial pressure (ICP) in the patient with head injury. Most prehospital deaths after TBI are probably caused by vascular and respiratory failure. This is supported by the marked improvements in outcome achieved by emergency care systems with early prehospital intubation and resuscitation.

Shock is usually caused by hemorrhage elsewhere in the body, not in the head. Cerebral perfusion pressure should be maintained above 70 mm Hg by vigorous management of hypotension. Fluid resuscitation with normal saline or lactated Ringer solution is generally recommended, but patients with TBI should not receive excessive hydration, and central venous pressure should be monitored. Glucose administration should be avoided because it has been linked to poor outcome, possibly through increased lactic acidosis.[51]

Comatose patients with TBI are often hypoxic or hypercapnic, even though ventilation may appear to be normal. Patients who are in a coma (i.e., those with a GCS score < 8) should undergo gentle hyperventilation, via intubation if necessary, until a carbon dioxide tension (P_{CO_2}) of about 35 mm Hg is achieved. Short-term hyperventilation to levels of about 25 mm Hg can be lifesaving in the patient with impending tentorial herniation. However, the recommended standard is that chronic hyperventilation be maintained at a P_{CO_2} no lower than 25 mm Hg, because lower levels reduce cerebral blood flow and have a negative impact on outcome. Sedation or pharmacologic paralysis should be used when necessary to control acute agitation. The head should be elevated and immobilized in the plane of the body for airway maintenance and facilitation of cranial venous return.

Finally, the special nutritional requirements of the TBI patient also need particular attention, from coma through subacute recovery. Early nutritional support (often parenteral) may be associated with improved survival and decreased disability.[52]

Radiologic Examination

CT has revolutionized the diagnosis and management of mass lesions in patients with head trauma; it should be performed in all patients with a GCS score of less than 15 and in those who have focal signs or posttraumatic amnesia. Comatose patients must be accompanied by trained personnel on the way to the CT suite because patients who are assumed to be stabilized may suffer respiratory arrest or irreversible brain damage as a result of simple airway problems en route.

The principal role of CT is in the diagnosis and management of acute surgical lesions. Hemorrhage can occur in the subarachnoid, subdural, epidural, and intraventricular spaces or in the brain parenchyma. Subdural and epidural hematomas should

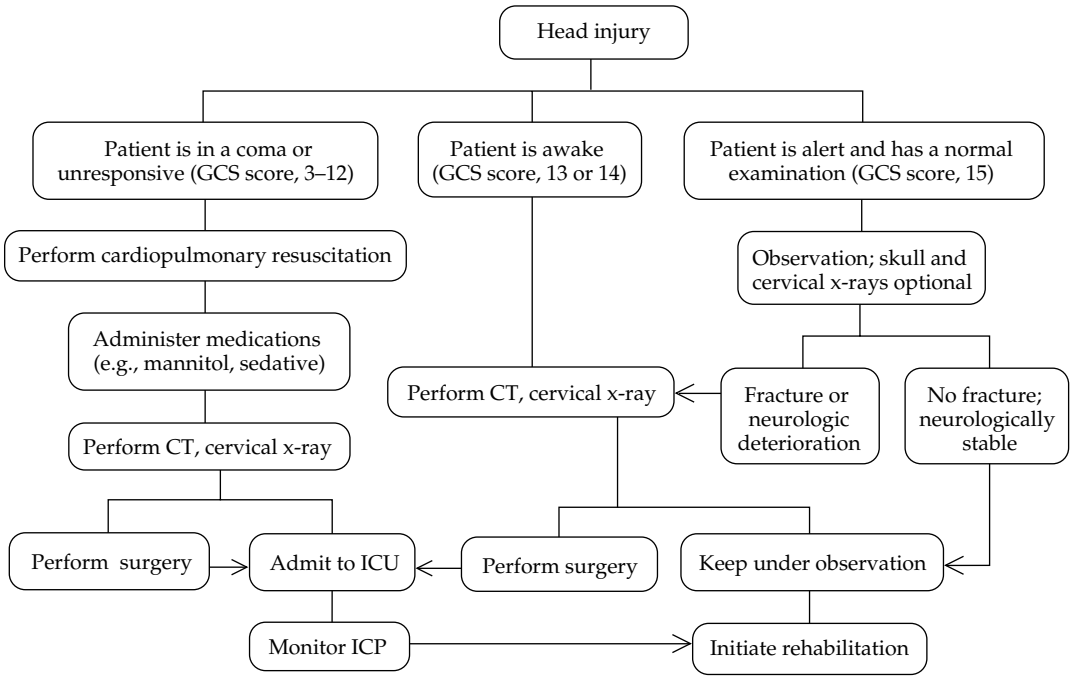


Figure 2 **Management algorithm for patients with traumatic brain injury. (GCS—Glasgow Coma Scale; ICP—intracranial pressure)**

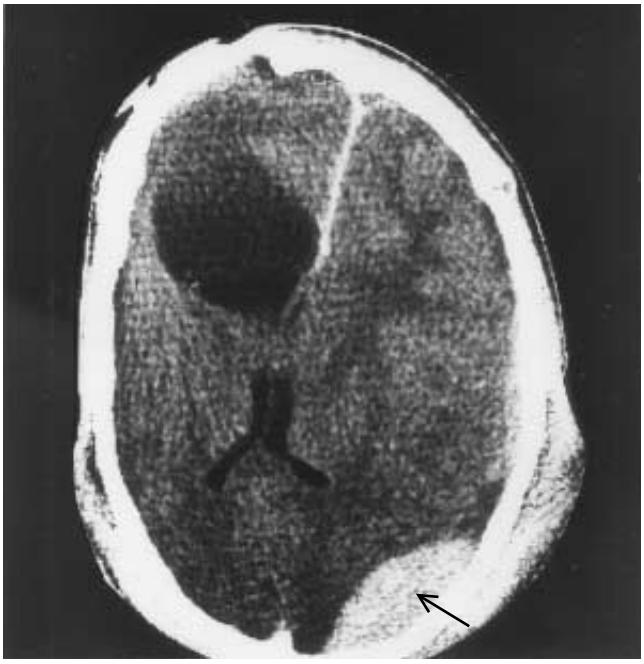


Figure 3 **CT in a 19-year-old man with a history of head injury and recent occipital injury shows the old lesion, the resulting midline distortion, and an epidural hematoma (arrow) resulting from the recent injury.**

be evacuated promptly when associated with a significant mass effect, because it has been shown that there is a significantly poorer outcome with surgical delays of greater than 4 hours.[53,54] However, surgical management of intraparenchymal hemorrhage will vary, depending on the size, mass effect, location, and neurologic status. Intraventricular hemorrhage will generally require ventricular drainage, and it has also been associated with a worse prognosis. Obliteration of the basilar cisterns from mass effect also portends a worse outcome, as does diffuse hypodensity typical of cerebral hypoxia.[49]

CT has made the skull x-ray all but obsolete, and the latter rarely affects management.[55] Normal findings on initial skull x-ray or CT should not lull the clinician into ignoring the possibility that a delayed hematoma may develop, although this is fortunately relatively rare [*see Figure 3*]. Whether in the hospital or at home with a responsible adult, observation remains a critical element of care, even for patients with MTBI.

MRI promises to be very useful in the long-term management of moderate and severe TBI as well as in the documentation of brain pathology in patients with milder injury. However, it is often impractical and not cost-effective in the acutely comatose patient and is not as good as CT for diagnosis of acute hematomas.

Laboratory tests should include a complete blood count; measurements of electrolytes, glucose, arterial blood gases, and blood alcohol; liver and kidney function tests; and a toxicology screen. In addition, because coagulopathies frequently occur after TBI, other tests may be indicated. Such tests include a platelet count; prothrombin, partial thromboplastin, and thrombin times; and an evaluation of fibrinogen and fibrinogen degradation products.

ACUTE MANAGEMENT

Acute management of severe head trauma is aimed at minimizing the progression or the effects of secondary injury. Although there are no specific treatments targeting the biochemical events discussed (see above), much progress has been made in early resuscitation and the management of elevated ICP. Evidence-based guidelines and standards for management of patients with severe acute TBI, as well as options for the care of such patients, have been updated by the Joint Section on Neurotrauma and Critical Care and the American Association of Neurological Surgeons, under sponsorship of the Brain Trauma Foundation.[56] These guidelines are also available through the Brain Trauma Foundation Web site (www.braintrauma.org). They represent a major advance in standardization of the management of severe TBI and should be consulted by anyone involved in the care of these patients.

Aspects of care described in the guidelines include early resuscitation; ICP monitoring; ICP treatment threshold and methods; and the use of mannitol, barbiturates, nutrition, hyperventilation, corticosteroids, and prophylactic anticonvulsants. It is important to note that even guidelines that are presented as options for care are very valuable, because such options represent the consensus of experts in areas where studies documenting more definitive levels of certainty are not available or are not possible [*see Table 3*]. The guidelines currently support a standard recommendation against the routine use of aggressive hyperventilation, corticosteroids, and prophylactic anticonvulsants.

The Intensive Care Unit, Intracranial Pressure Monitoring, and Cerebral Perfusion Pressure

After a mass lesion has been surgically treated or excluded, the comatose patient should be managed in the ICU. Preventing secondary insults to the brain remains the principal goal of therapy. In general, the same principles of care that are applied in earlier stages of treatment (see above) are applied at this stage, but better monitoring is available in the ICU. Organization, training, and adherence to relatively simple principles are the mainstay of care.

ICP and cerebral perfusion pressure (CPP) are probably the most sensitive measures for monitoring the patient with severe TBI, and the results of these assessments correlate significantly with outcome.[57-59] The current guideline recommends that ICP be monitored in comatose patients in whom CT yields abnormal results and in those in whom CT is normal but who have two or more of the following risk factors: age greater than 40 years, motor posturing, or a systolic blood pressure of less than 90 mm Hg. The particular monitoring technique to be used is determined by the neurosurgeon and the facilities available. An intraventricular catheter is recommended, because it can also be used for ventricular drainage as needed, though fiberoptic epidural or subdural transducers can also be used.[60]

The current standard measures for control of elevated ICP include sedation, paralysis, controlled hyperventilation, use of mannitol and other osmotics, ventricular drainage, and barbiturate coma [*see Table 4*]; these interventions are usually undertaken in that sequence to maintain an ICP of lower than 20 mm Hg.[58,61] It is recommended as a guideline that mannitol be given in intermittent boluses of 0.25 to 1.0 g/kg every 4 hours as needed, but serum osmolarity should be kept below 320 mOsm/L because of concerns about renal failure. Use of a Foley catheter is strongly recommended to monitor urine output and help maintain euvolemia through adequate fluid replacement. Barbiturate coma significantly improves outcome in patients younger than 45 years with otherwise uncontrolled ICP. This is the last step recommended as a guideline in the nonsurgical control of ICP.

Barbiturate coma is induced with pentobarbital at an initial loading dose of 10 mg/kg I.V. over 30 minutes, and serum levels should then be maintained at 3 to 4 mg/dl with dosages of about 1 mg/kg/hr. The literature supports a standard recommendation that corticosteroids not be used for neuroprotection or control of ICP in patients with severe TBI.[62] Finally, progressive elevations in ICP may be caused by lesions that require surgery, such as delayed hematoma or hydrocephalus. Similarly, seizures, hyponatremia, and airway problems will raise ICP.

MANAGEMENT OF AGITATION

Patients with TBI often experience agitation during the immediate recovery period. Nonpharmacologic interventions, including limiting environmental stimuli and providing gentle interaction with the patient, are of great importance in early management and should typically be the first line of therapy. When ICU patients require chemical restraint for their own safety and the safety of staff, propofol may be given intravenously to manage agitation; the recovery time with propofol is quicker than with benzodiazepines such as midazolam.

After the patient has left the ICU, benzodiazepines (e.g., lorazepam, 0.5 to 2.0 mg p.o. or I.M., or clonazepam, 0.5 to 2.0 mg p.o.) may be the first choice, either alone or in combination with carbamazepine or valproate. However, benzodiazepines may cause disinhibition in some patients with brain injury. Neuroleptics (e.g., molindone, 10 mg p.o., b.i.d., or haloperidol, 0.5 to 2.5 mg p.o., b.i.d.) are less desirable because they may cause extrapyramidal effects, akathisia (a subjective sense of restlessness that may prolong agitation), or both. Some newer agents, such as olanzapine, 2.5 to 5.0 mg/day orally, may have fewer side ef-

Table 3 Evidence-Based Guidelines for the Management of Severe Head Injury[62]

Subject	*Certainty Level**	*Recommendations*
Trauma systems	Guideline	All regions in the United States should have an organized trauma care system
Resuscitation	Guideline	Systolic BP < 90 mm Hg or hypoxia must be scrupulously avoided or corrected immediately
	Option	Mean arterial pressure should be kept above 90 mm Hg
Integration of brain-specific treatments into initial resuscitation	Option	When clear signs of transtentorial herniation are present, the herniation should be treated aggressively; hyperventilation should be performed rapidly (mannitol is desirable with adequate volume resuscitation); sedation and short-acting neuromuscular blockade can be used but may interfere with the neurologic examination
ICP monitoring	Guideline	Admission GCS 3–8 plus abnormal CT scan, *or* GCS 3–8 and normal CT plus age > 40 years, *or* motor posturing, *or* systolic BP < 90 mm Hg
ICP treatment threshold	Guideline	ICP > 20–25 mm Hg
	Option	ICP treatment should be corroborated by frequent clinical examination and cerebral perfusion pressure data
ICP treatment critical pathway	Option	[*See* text, Table 4, and reference 47]
CPP	Option	Maintain CPP above 70 mm Hg
Hyperventilation	Standard	In the absence of increased ICP, chronic prolonged hyperventilation (P_aco_2 < 25 mm Hg) should be avoided
	Guideline	Prophylactic hyperventilation (P_aco_2 < 35 mm Hg) should be avoided during the first 24 hr after severe TBI because it can compromise CPP when cerebral blood flow is reduced
	Option	Hyperventilation therapy may be necessary for brief periods when there is neurologic deterioration or when ICP elevations are refractory to other treatment
Mannitol	Guideline	Mannitol is effective for ICP control; intermittent boluses (0.25 to 1.0 g/kg) may be more effective than continuous infusion
	Options	Indications for mannitol before ICP monitoring are progressive neurologic deterioration or transtentorial herniation not attributable to systemic pathology
		Maintain serum osmolality < 320 mOsm
		Maintain euvolemia by adequate fluid replacement; use Foley catheter
Barbiturates	Guideline	High-dose barbiturates may be used in hemodynamically stable severe TBI patients with ICP elevations refractory to maximal medical and surgical therapy
Glucocorticoids	Standard	Glucocorticoids are not recommended for ICP control or improving outcome in severe TBI patients
Nutritional support	Guideline	Replace 140% of resting metabolic expenditure (100% in paralyzed patients) by using enteral or parenteral formulas, with at least 15% of calories as protein
	Option	Feeding by gastrojejunostomy is preferred
Antiseizure prophylaxis	Standard	Prophylactic use of phenytoin, carbamazepine, or phenobarbital is not recommended for preventing late posttraumatic seizures
	Option	Short-term (1 wk) phenytoin or carbamazepine is recommended to prevent early posttraumatic seizures in high-risk patients after head injury

Note: This table should not be used alone as a guide to therapy. Clinicians are referred to the full guideline document (see text).
*For determining certainty level, studies that are controlled and randomized rank highest, and expert opinion ranks lowest. The highest level of certainty is represented by standards; the next highest, by guidelines; and the lowest, by options for care.
BP—blood pressure CPP—cerebral perfusion pressure GCS—Glasgow Coma Scale ICP—intracranial pressure P_aco_2—arterial carbon dioxide tension TBI—traumatic brain injury

Table 4 Management of Intracranial Pressure

Treatment	*Dosage*
Sedation (with morphine)	As needed
Paralysis (with pancuronium)	As needed
Ventricular drainage	As needed
Mannitol	0.25–1.0 g/kg q. 4 hr
Hyperventilation	To a Pco_2 of 35 mm Hg (25 mm Hg only for brief periods, if needed for transtentorial herniation)
Barbiturate coma (pentobarbital)	Loading dose of 10 mg/kg, then 1 mg/kg/hr

Pco_2—carbon dioxide tension

fects. Animal models suggest that neuroleptics have a negative long-term effect on recovery.[63] Dopamine agonists such as amantadine and bromocriptine have also been successfully used for postcoma agitation caused by impairment of dopaminergic and other ascending monoaminergic pathways.

A recent survey of rehabilitation physicians suggests that those who are more experienced in caring for patients with brain injury tend to use carbamazepine and beta blockers in preference to neuroleptics for management of agitation. In all cases of prolonged confusion or agitation, however, other causes must also be considered, such as the side effects of medication, infection, electrolyte imbalance, hypoxia, and late intracranial complications.

TREATMENT OF NEUROPSYCHIATRIC SEQUELAE

The neuropsychiatric sequelae of brain injury, both socially and in the workplace, are well appreciated. Verbalizations and behavior can be striking, especially when the patient has a reduced ability to self-monitor and is unconcerned.[64,65] Neurologic abnormalities may not be as distressing to the patient and his or her family as personality changes and inappropriate behavior. Suitable treatment of neurobehavioral sequelae will often decrease patient and caregiver distress and markedly improve overall outcome.[66]

Intellectual impairments increase as the duration of posttraumatic amnesia rises from less than 1 hour to longer than 7 days. Inappropriate behavior associated with frontal, temporal, and limbic connections (e.g., poor social judgment, increased irritability, and poor impulse control) is particularly common, even in patients in whom imaging studies show no focal pathology. Frontal-thalamic reticular circuit damage may also cause fatigue and frequent sleep disturbances.

Depression

Studies consistently show a 25% to 50% incidence of depression after TBI. In one study, of the 75% of TBI patients who were not depressed at the initial interview, 25% developed depression during the first year of follow-up; the mean duration of depression was 4 to 5 months for the total sample.[67] Patients frequently complain of hopelessness, a loss of interest in usual activities, self-deprecation, a lack of energy, and a lack of self-confidence. Anxiety symptoms may be prominent, especially within 6 months after injury.

Many reports suggest that the depression associated with stroke and TBI has a neurologic basis. Acutely depressed patients with TBI frequently have left anterior lesions, whereas patients with mixed anxiety and depression are more likely to have right-hemisphere lesions, a longer duration of depression, and poorer psychosocial outcome.[67] The incidence of depression, its duration, and its associated symptoms, such as anxiety, may therefore be related to the location and laterality of cerebral pathology.

Antidepressants are indicated in the treatment of depression and of mixed anxiety and depression in TBI [*see 13:VIII Anxiety Disorders*]. The selective serotonin reuptake inhibitors (SSRIs) fluoxetine, sertraline, and paroxetine are favored because they are safe and easy to administer and do not cause unwanted anticholinergic side effects. Tricyclic antidepressants may also be used; desipramine and nortriptyline have the fewest anticholinergic and antihistaminic properties. The antidepressant venlafaxine is both serotoninergic and dopaminergic. Although stimulants such as methylphenidate and dextroamphetamine have primarily been used in TBI patients to treat attention difficulties, they can also be used to treat depression or to augment antidepressant treatment with SSRIs or tricyclic antidepressants. Bupropion should be given only with caution to patients with brain injury, because it can lower the seizure threshold.

Anxiety

Anxiety disorders, alone or in combination with depression, occur in patients with TBI and are treated with SSRIs, tricyclic antidepressants, benzodiazepines, or buspirone. As with antidepressants, buspirone has a 2- to 3-week latency period until it reaches full therapeutic effect. Patients with phobic avoidance are best treated with a combination of cognitive-behavioral therapy and benzodiazepines. Symptoms of posttraumatic stress disorder (PTSD) may occur and are more common in patients with MTBI than in patients with more severe TBI. Longer periods of posttraumatic amnesia, in which an explicit memory of the upsetting events is not established, may protect against the development of nightmares and intrusive thoughts about the trauma.[68] The development of PTSD symptoms probably entails implicit memory by a mechanism similar to negative conditioning and thus may not require explicit memory of the event.[69]

Irritability and Aggression

Irritability occurs in more than 50% of patients with moderate or severe traumatic head injury during the first 6 months after injury. Aggressive outbursts in brain-injured patients tend to be verbal and brief and are precipitated by seemingly trivial provocations. Patients may be remorseful and apologetic afterward but seem to be unable to curtail subsequent outbursts. Irritability and aggression may occur in patients without a history of such behavior; disinhibition caused by frontal system dysfunction and possible injury to limbic and hypothalamic structures is implicated.[70] Impaired serotonin transmission may also be involved, because low levels of 5-hydroxyindoleacetic acid (5-HIAA) have been noted in the cerebrospinal fluid of impulsive, violent patients. Anticonvulsants, lithium, buspirone, beta blockers, SSRIs, and stimulants have been variously reported to be of use in decreasing the amount of aggression.[71]

Attention Deficits

The stimulants dextroamphetamine and methylphenidate may improve attention and concentration and are often used in the clinic in selected patients. A randomized, controlled study of methylphenidate in adults with moderate TBI found a more

rapid rate of recovery of attention and improved scores on motor and disability tests but no change in ultimate overall outcome through the use of methylphenidate.[72] Findings of a recent controlled crossover study on the use of methylphenidate in children suggested that there is no difference in behavior, attention, memory, or processing speed between persons receiving medication and those given placebo.[73] Clear guidelines on the use of stimulants to treat attention deficits in TBI await additional studies.

TREATMENT OF POSTTRAUMATIC EPILEPSY

The risk of epilepsy in patients with closed-head injury is relatively small: 2% to 5% in all patients and about 10% to 20% in patients with severe closed-head injury.[74] Some studies have shown a higher incidence of seizures in patients with depressed skull fractures (15%), hematomas (31%), and penetrating brain wounds (50%).[75,76] In all cases, the risk decreases markedly with time. Although the relative risk of developing epilepsy 10 to 15 years after penetrating head injury is still 25 times higher than in the normal age-matched population, 95% of patients with penetrating head injury will remain seizure-free if they have no seizures during the first 3 years after injury.[77]

Because most patients who develop posttraumatic epilepsy in the first week after injury will have recurrent seizures for some time, anticonvulsant therapy is indicated in documented cases.[78] Controlled, randomized studies have shown that the use of phenytoin, phenobarbital, carbamazepine, or valproate does not prevent the development of posttraumatic epilepsy beyond the first week after injury. It is now recommended as a standard of care that these medications not be used to prevent posttraumatic epilepsy in patients who have not had a seizure.[79] In light of the sensitivity of the acutely traumatized brain to the secondary insult of a grand mal seizure, I recommend routine short-term use (for 1 to 2 weeks after injury) of phenytoin or carbamazepine in high-risk patients with acute TBI. Carbamazepine may be preferable because it helps control agitation in some patients.

Evidence suggests that iron-catalyzed lipid peroxidation may partly mediate the development of posttraumatic epilepsy; inhibitors of lipid peroxidation, such as methylprednisolone and α-tocopherol, can prevent iron-induced epilepsy in animals, but no well-controlled clinical trials have explored this avenue.[80]

Long-term Outcome

The young-adult brain has a remarkable capacity to compensate for many aspects of injury naturally. Although disabilities such as hemiparesis, seizures, and certain language disorders may initially appear more dramatic, the most devastating long-term impairments are the cognitive defects, attention deficits, and, in particular, behavioral changes that often persist after TBI.[81]

Prognosis for full recovery must be more guarded in the elderly, who have been reported to have about twice the mortality of younger TBI patients. Likewise, seemingly less severe injuries often result in worse functional outcomes in older patients.[82]

Measurement of outcome from TBI remains a challenge. Functional measurement instruments include the Glasgow Outcome Score, the Disability Rating Scale, the Rancho Los Amigos Score, the Functional Independence Measure, and various neuropsychological, behavioral, and quality-of-life measures. However, return to gainful employment is probably the best overall measure of long-term outcome.[83] About 50% of patients who survive severe TBI eventually return to work. In recent studies, return to work was also the single best correlate of perceived quality of life.[84,85]

Accurate predictors of outcome are also important to patients, their families, and caregivers in understanding recovery and planning for care. Early predictors of good outcome include a higher preinjury intelligence, youth, and a lower severity of injury.[49] Certain genotypes, such as the *ApoE4* allele, may affect recovery and outcome.[86,87] In a large multidisciplinary study of survivors of penetrating head injury, the presence of seven factors were significantly predictive of unemployment: hemiparesis, epilepsy, visual field loss, verbal memory loss, visual memory loss, psychological problems, and violent behavior. These factors represent different domains of brain function and were relatively equipotent in the model (i.e., it was the number of impaired domains, not impairment in any one particular domain, that was predictive of unemployment). The brain may thus compensate for injury by utilizing whichever functional domains are still available to it.[88]

Traumatic Brain Injury Rehabilitation

In the 1990s, the field of TBI rehabilitation blossomed. A profusion of therapies, including coma stimulation, cognitive rehabilitation, speech therapy, occupational therapy, and recreational therapy, are now available. Their use is largely empirical, and these sometimes expensive interventions have not been subjected to a high level of scientific scrutiny for efficacy and cost-effectiveness. Nevertheless, a growing body of animal and clinical literature supports the beneficial effects of training on the brain and on performance after injury.

The goals of therapy should be recovery of the patient's independence and his or her reintegration into the community. The prevention of maladaptive behaviors is an important secondary goal. However, patients with scarce economic resources can find those resources depleted in the early phases of recovery by the evaluation of and therapy for specific neurologic or cognitive deficits that may resolve even without therapy or may ultimately be of marginal importance to the patient's achieving the goal of independence. Interventions that may be more cost-effective, such as training in decision-making or other community reintegration skills and certain forms of behavioral management, may ultimately be omitted for lack of resources.

Although there is consensus about the benefits of some forms of rehabilitation for patients with TBI, the type, intensity, and duration of rehabilitation that are best for a given patient remain hotly debated.[89,90] For example, a recent large prospective, randomized, controlled trial compared an intensive in-hospital program of cognitive rehabilitation with a limited (and much less expensive) home rehabilitation program in soldiers recovering from moderate to severe TBI.[91] At 1 year after injury, there was no difference between the two groups with regard to return to work, fitness for military duty, or behavioral, neuropsychological, or quality-of-life measures. However, in a subset analysis of patients who were unconscious for longer than 1 hour after suffering TBI, there was a higher return-to-work rate for the patients who underwent hospital rehabilitation than for those who underwent home rehabilitation, suggesting the differential value of these approaches for selected patients. Thus, a fundamental challenge for rehabilitation is to distinguish the effect of the brain's natural processes of recovery from the effects of treatment in patients with varying prognostic risk factors for long-term function. Clear

resolution of these issues will require further properly designed, prospective, controlled, randomized trials.

Additional Information

Additional information on TBI can be obtained from the American Academy of Neurology (http://www.aan.com), the Brain Injury Association (http://www.biausa.org), and the Brain Trauma Foundation (http://www.braintrauma.org).

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Max W, MacKenzie E, Rice D: Head injuries: costs and consequences. J Head Trauma Rehabil 6:76, 1991

2. Faden A, Demediuk P, Panter S, et al: The role of excitatory amino acids and NMDA receptors in traumatic brain injury. Science 244:798, 1989

3. Clifton G, Robertson C, Grossman R: Management of the cardiovascular and metabolic responses to severe head injury. Central Nervous System Trauma Status Report. Becker D, Povlishock J, Eds. National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, 1985, p 139

4. Kaufman H, Mattson J: Coagulopathy in head injury. Central Nervous System Trauma Status Report. Becker D, Povlishock J, Eds. National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, 1985, p 187

5. Gennarelli T, Thibault L, Adams J, et al: Diffuse axonal injury and traumatic coma in the primate. Ann Neurol 12:564, 1982

6. Povlishock JT, Christman CW: The pathobiology of traumatically induced axonal injury in animals and humans: a review of current thoughts. J Neurotrauma 12:555, 1995

7. Maxwell WL, Povlishock JT, Graham DL: A mechanistic analysis of nondisruptive axonal injury: a review. J Neurotrauma 14:419, 1997

8. Hatton J, Rapp R, Kudsk K, et al: Intravenous insulin-like growth factor–I (IGF-I) in moderate-to- severe head injury: a phase II safety and efficacy trial. J Neurosurg 86:779, 1997

9. Bergsneider M, Hovda D, Lee S, et al: Dissociation of cerebral glucose metabolism and level of consciousness during the period of metabolic depression following human traumatic brain injury. J Neurotrauma 17:389, 2000

10. Marshall L, Marshall S: Pitfall and advances from the international tirilazad trial in moderate and severe head injury. J Neurotrauma 12:929, 1995

11. Young B, Runge J, Waxman K, et al: Effects of pegorgotein on neurologic outcome of patients with severe head injury: a multicenter, randomized controlled trial. JAMA 276:538, 1996

12. Ishige N, Pitts L, Hashimoto T, et al: Effect of hypoxia on traumatic brain injury in rats. Part 1: Changes in neurological function, electroencephalograms, and histopathology. Neurosurgery 30:848, 1987

13. Schoch HJ, Fischer S, Marti HH: Hypoxia-induced vascular endothelial growth factor expression causes vascular leakage in the brain. Brain 125(pt 11):2549, 2002

14. Biros MH, Heegaard W: Prehospital and resuscitative care of the head-injured patient. Curr Opin Crit Care 7:444, 2001

15. Ross DT, Graham DI, Adams JH: Selective loss of neurones from the thalamic reticular nucleus following severe human head injury. J Neurotrauma 10:151, 1993

16. Bullock R: Strategies for neuroprotection with glutamate antagonists: extrapolating from evidence taken from the first stroke and head injury studies. Ann N Y Acad Sci 765:272, 1995

17. Marion D, Penrod L, Kelsey S, et al: Treatment of traumatic brain injury with moderate hypothermia. N Engl J Med 336:540, 1997

18. Clifton G, Miller E, Choi S, et al: Lack of effect of induction of hypothermia after acute brain injury. N Engl J Med 344:556, 2001

19. Verma A: Opportunities for neuroprotection in traumatic brain injury. J Head Trauma Rehabilitation 15:1149, 2000

20. Sullivan P, Thompson M, Scheff S: Cyclosporin A attenuates acute mitochondrial dysfunction following traumatic brain injury. Exp Neurol 160:226, 1999

21. Buki A, Okonkwo D, Povlishock J: Postinjury cyclosporin A administration limits axonal damage and disconnection in traumatic brain injury. J Neurotrauma 16:511, 1999

22. Sullivan P, Geiger J, Mattson M, et al: Dietary supplement creatine protects against traumatic brain injury. Ann Neurol 48:723, 2000

23. Sailor K, Vemuganti R, Gilboe D, et al: Pyruvate treatment is neuroprotective after traumatic brain injury. 30th Annual Meeting of the Society for Neuroscience, New Orleans, Nov. 4th to 9th , 2000

24. Thatcher R, Walker R, Gerson I, et al: EEG discriminant analysis of mild head trauma. Electroencephalogr Clin Neurophysiol 73:93, 1989

25. Thatcher R, Biver C, McAlaster R, et al: Biophysical linkage between MRI and EEG amplitude in closed head injury. Neuroimage 7:352, 1998

26. Thatcher R, Biver C, McAlaster R, et al: Biophysical linkage between MRI and EEG coherence in closed head injury. Neuroimage 8:307, 1998

27. Thatcher R, Biver C, Gomez M, et al: Estimation of the EEG power spectrum using MRI T(2) relaxation time in traumatic brain injury. Clin Neurophysiol 112:1729, 2001

28. McCrea M, Kelly J, Kluge J, et al: Standardized assessment of concussion in football players. Neurology 48:586, 1997

29. Oppenheimer DR: Microscopic lesions in the brain following head injury. J Neurol Neurosurg Psychiatry 31:299, 1968

30. Levin HS, Mattis S, Ruff RM, et al: Neurobehavioral outcome following minor head injury: a three-center study. J Neurosurg 66:234, 1987

31. Warden D, Bleiberg J, Cameron M, et al: Persistent prolongation of simple reaction time in sports concussion. Neurology 57:524, 2001

32. McCrea M, Kelly JP, Randolph C, et al: Immediate neurocognitive effects of concussion. Neurosurgery 50:1032, 2002

33. McAllister TW, Sparling MB, Flashman LA, et al: Differential working memory load effects after mild traumatic brain injury. Neuroimage 14:1004, 2001

34. Thurman D, Branche C, Sniezeck J: The epidemiology of sports-related traumatic brain injuries in the United States: recent developments. J Head Trauma Rehabil 13:1, 1998

35. Macciocchi S, Jane J, Rimel R, et al: Neuropsychological functioning and recovery after mild head injury in collegiate athletes. Neurosurgery 39:510, 1996

36. Matser EJ, Kessels AG, Lezak MD, et al: Neuropsychological impairment in amateur soccer players. JAMA 282:971, 1999

37. Junger EC, Newell DW, Grant GA, et al: Cerebral autoregulation following minor head injury. J Neurosurg 87:485, 1997

38. McCrory PR, Berkovic SF: Second impact syndrome. Neurology 50:677, 1998

39. Williams DH, Levin HS, Eisenberg HM: Mild head injury classification. Neurosurgery 27:422, 1990

40. Levin HS, Amparo E, Eisenberg HM, et al: Magnetic resonance imaging and computerized tomography in relation to the neurobehavioral sequelae of mild and moderate head injuries. J Neurosurg 66:706, 1987

41. Shaw N: The neurophysiology of concussion. Prog Neurobiol 67:281, 2002

42. Thurman DJ, Alverson C, Dunn KA, et al: Traumatic brain injury in the United States: a public health perspective. J Head Trauma Rehabil 14:602, 1999

43. Culotta VP, Sementilli ME, Gerold K, et al: Clinicopathological heterogeneity in the classification of mild head injury. Neurosurgery 38:245, 1996

44. Bleiberg J, Halpern EL, Reeves D, et al: Future directions for the neuropsychological assessment of sports concussion. J Head Trauma Rehabil 13:36, 1998

45. McCrea M, Kelly JP, Randolph C, et al: Standardized assessment of concussion (SAC): on-site mental status evaluation of the athlete. J Head Trauma Rehabil 13:27, 1998

46. Servadei F, Teasdale G, Merry G: Defining acute mild head injury in adults: a proposal based on prognostic factors, diagnosis, and management. J Neurotrauma 18:657, 2001

47. Alexander MP: Mild traumatic brain injury: pathophysiology, natural history, and clinical management. Neurology 45:1253, 1995

48. Ponsford J, Willmott C, Rothwell A, et al: Impact of early intervention on outcome following mild head injury in adults. J Neurol Neurosurg Psychiatry 73:330, 2002

49. Chesnut R, Ghajar J, Maas A, et al: Early indicators of prognosis in severe traumatic brain injury. J Neurotrauma 17:557, 2000

50. Resuscitation of blood pressure and oxygenation. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:471, 2000

51. Robertson CS, Goodman JC, Narayan RK, et al: The effect of glucose administration on carbohydrate metabolism after head injury. J Neurosurg 74:43, 1991

52. Yanagawa T, Buff F, Roberts I, et al: Nutritional support for head-injured patients. Cochrane Database Syst Rev (3):CD001530, 2002

53. Marshall L, Toole B, Bowers S: The National Traumatic Coma Data Bank. Part II: Patients who talk and deteriorate: implications for treatment. J Neurosurg 59:285, 1983

54. Seelig JM, Becker DP, Miller JD, et al: Traumatic acute subdural hematoma: major mortality reduction in comatose patients treated within four hours. N Engl J Med 304:1511, 1981

55. Ingebrigtsen T, Romner B: Routine early CT-scan is cost saving after minor head injury. Acta Neurol Scand 93:207, 1996

56. Trauma systems. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:457, 2000

57. Indications for intracranial pressure monitoring. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:479, 2000

58. Guidelines for cerebral perfusion pressure. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:507, 2000

59. Hackbarth RM, Rzeszutko KM, Sturm G, et al: Survival and functional outcome in pediatric traumatic brain injury: a retrospective review and analysis of predictive factors. Crit Care Med 30:1630, 2002

60. Recommendations for intracranial pressure monitoring technology. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:497, 2000

61. Intracranial pressure treatment threshold. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:493, 2000

62. Bullock R, Chesnut R, Clifton G, et al: Guidelines for the management of severe head injury. J Neurotrauma 13:639, 1996

63. Feeney D, Gonzalez A, Law W: Amphetamine, haloperidol and experience interact to affect rate of recovery after motor cortex injury. Science 217:855, 1982

64. Silver J, Yudofsky S, Hales R: Neuropsychiatry of traumatic brain injury. American Psychiatric Press, Washington, DC, 1994

65. Lishman W: Head injury. Organic Psychiatry. Lishman W, Ed. Blackwell Publishers, Oxford, 1987, p 137

66. Viguier D, Dellatolas G, Gasquet I, et al: A psychological assessment of adolescent and young adult inpatients after traumatic brain injury. Brain Inj 15:263, 2001

67. Jorge R, Robinson R, Starkstein S, et al: Influence of major depression on 1-year outcome in patients with traumatic brain injury. J Neurosurg 81:726, 1994

68. Warden D, Labbate L, Salazar A, et al: Posttraumatic stress disorder in patients with traumatic brain injury and amnesia for the event? J Neuropsychiatry Clin Neurosci 9:18, 1997

69. Krikorian R, Layton B: Implicit memory in posttraumatic stress disorder with amnesia for the traumatic event. J Neuropsychiatry Clin Neurosci 10:359, 1998

70. Grafman J, Schwab K, Warden D, et al: Frontal lobe injuries, violence, and aggression: a report of the Vietnam head injury study. Neurology 46:1231, 1996

71. Mooney GF, Haas LJ: Effect of methylphenidate on brain injury-related anger. Arch Phys Med Rehabil 74:153, 1993

72. Plenger P, Dixon C, Castillo R, et al: Subacute methylphenidate treatment for moderate to moderately severe traumatic brain injury: a preliminary double-blind placebo-controlled study. Arch Phys Med Rehabil 77:536, 1996

73. Williams SE, Ris MD, Ayyangar R, et al: Recovery in pediatric brain injury: is psycho-stimulant medication beneficial? J Head Trauma Rehabil 13:73, 1998

74. Annegers JF, Hauser WA, Coan SP, et al: A population-based study of seizures after traumatic brain injuries. N Engl J Med 338:20, 1998

75. Salazar AM, Jabbari B, Vance SC, et al: Epilepsy after penetrating head injury: I. Clinical correlates: a report of the Vietnam Head Injury Study. Neurology 35:1406, 1985

76. Jennett W: Epilepsy after Non-missile Head Injuries. Yearbook Medical Publishers, Chicago, 1975

77. Weiss GH, Salazar AM, Vance SC, et al: Predicting posttraumatic epilepsy in penetrating head injury. Arch Neurol 43:771, 1986

78. Haltiner AM, Temkin NR, Dikmen SS: Risk of seizure recurrence after the first late posttraumatic seizure. Arch Phys Med Rehabil 78:835, 1997

79. Role of antiseizure prophylaxis following head injury. The Brain Trauma Foundation, the American Association of Neurological Surgeons, and the Joint Section on Neurotrauma and Critical Care. J Neurotrauma 17:549, 2000

80. Willmore LJ: Post-traumatic epilepsy: cellular mechanism and implications for treatment. Epilepsia 31(suppl 3):S-67, 1990

81. Millis SR, Rosenthal M, Novack TA, et al: Long-term neuropsychological outcome after traumatic brain injury. J Head Trauma Rehabil 16:343, 2001

82. Susman M, DiRusso SM, Sullivan T, et al: Traumatic brain injury in the elderly: increased mortality and worse functional outcome at discharge despite lower injury severity. J Trauma 53:219, 2002

83. Wehman P, Kreutzer J, West M, et al: Employment outcomes of persons following traumatic brain injury: pre-injury, post-injury, and supported employment. Brain Inj 3:397, 1989

84. Melamed S, Grosswasser Z, Stern MJ: Acceptance of disability, work involvement and subjective rehabilitation status of traumatic brain-injured (TBI) patients. Brain Inj 6:233, 1992

85. Hoofien D, Gilboa A, Vakil E, et al: Traumatic brain injury (TBI) 10–20 years later: a comprehensive outcome study of psychiatric symptomatology, cognitive abilities and psychosocial functioning. Brain Inj 15:189, 2001

86. Jordan BD, Relkin NR, Ravdin LD, et al: Apolipoprotein Eε4 associated with chronic traumatic brain injury in boxing. JAMA 278:136, 1997

87. Mayeux R, Ottoman R, Maestre G, et al: Synergistic effect of traumatic head injury and apolipoprotein-α4 in patients with Alzheimer's disease. Neurology 45:555, 1995

88. Schwab K, Grafman J, Salazar AM, et al: Residual impairments and work status 15 years after penetrating head injuries: report from the Vietnam Head Injury Study. Neurology 43:95, 1993

89. Chesnut R, Carney N, Maynard H, et al: Evidence Report: Rehabilitation for Traumatic Brain Injury. Agency for Health Care Policy and Research, Report No. AHCPR 99-EOO6, Rockville, Maryland, 1999

90. Cicerone KD, Dahlberg C, Kalmar K, et al: Evidence-based cognitive rehabilitation: recommendations for clinical practice. Arch Phys Med Rehabil 81:1596, 2000

91. Salazar AM, Warden DL, Schwab K, et al: Cognitive rehabilitation for traumatic brain injury: a randomized trial. JAMA 283:3075, 2000

92. Teasdale G, Murray G, Parker L, et al: Adding up the Glasgow Coma Scale. Acta Neurochir Suppl (Wien) 28:13, 1979

Acknowledgments

Deborah Warden, M.D., was coauthor of the initial version of this subsection, particularly the neuropsychiatric section.

Figure 1 Positron-emission tomography scan courtesy of Dr. J. C. Umhau, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland.

Figure 2 Marcia Kammerer.

VI NEOPLASTIC DISORDERS

Jerome B. Posner, m.d.

Neoplasms can damage the nervous system in three ways: cells within the central nervous system or peripheral nervous system can become neoplastic[1]; tumors located outside the nervous system (i.e., systemic cancers) can metastasize to neural structures[2,3]; and systemic cancers can cause neural dysfunction indirectly [*see Table 1*].[2,4] This chapter focuses on primary and metastatic tumors of the CNS. Strictly speaking, meningiomas, pituitary tumors, and pineal region tumors are not brain tumors; however, because they arise within the intracranial cavity, they are considered here. Some nonmetastatic effects of cancer on the nervous system are also discussed.

Primary Nervous System Tumors—General Principles

EPIDEMIOLOGY

Neoplasms can develop in the brain and spinal cord, especially from glial cells or their precursors; neoplasms can also develop from intracranial structures outside of the brain and spinal cord, such as the meninges and the pituitary and pineal glands. Brain tumors of neuronal origin are uncommon.[5] The American Cancer Society (http://www.cancer.org) estimates that in 2004, 18,400 new cases of brain cancer and other CNS cancers developed (10,540 in males and 7,860 in females) and 12,690 patients died of CNS cancer.[6]

Although the incidence of brain tumor increases with age and thus most brain tumors are encountered in the elderly, brain tumors are the most common solid tumor of childhood and are the second most common cause of cancer death, after leukemia, in males under the age of 40 and in females under the age of 20.[6] Brain tumors affect both sexes and all races. Meningiomas and pituitary adenomas are more common in women and African Americans, whereas gliomas are more common in men and whites. Meningiomas and pituitary tumors[7] are quite common. Low-grade gliomas, such as astrocytomas, are more common in the young; high-grade gliomas, such as anaplastic astrocytomas and glioblastomas multiforme, are more common in the elderly. The incidence of glioma may actually decrease in the very old (i.e., persons 75 to 80 years of age or older).[8] The apparent incidence of CNS cancers increased during the late 1970s and 1980s as a result of better diagnosis, but the overall rate now is stable.[8] However, the incidence of primary CNS lymphoma (PCNSL) in immunocompetent persons (e.g., those who are not infected with HIV) continues to increase.[9] Although once rare, PCNSL is now an important consideration in the diagnosis and treatment of brain tumors, particularly because the tumor is sometimes curable.[10]

CLASSIFICATION

The World Health Organization (http://www.who.int) classifies CNS and peripheral nervous system tumors by their presumed cell of origin [*see Table 2*].[5] Most parenchymal brain and spinal cord tumors are of neuroepithelial origin—glial cells (astrocytes or oligodendroglia) or their precursors. Oligodendroglial tumors constitute 5% to 10% of glial tumors. Glial tumors are classified as either high grade or low grade; nuclear atypia, mitotic figures, microvascular proliferation, and necrosis indicate high-grade tumors. If three or all four of these features are present, the prognosis is poor.

Although often used, the terms benign and malignant have less meaning in tumors of brain and spinal cord parenchyma than in other cancers. Brain and spinal cord tumors are seldom truly benign because surgery rarely cures; they are not truly malignant in the usual sense because they seldom metastasize to other organs. Surgery fails because even histologically benign intrinsic brain tumors (e.g., gliomas) infiltrate normal tissue, thereby preventing total resection and allowing tumor recurrence. Certain tumors, such as medulloblastoma, ependymoma, glioblastoma, and PCNSL, may seed the leptomeninges or spread to distant sites within the neuraxis, but they rarely spread outside the CNS.

ETIOLOGY

Environmental Risk Factors

Immunosuppression—either congenital or acquired—and exposure to ionizing radiation are the only well-established environmental risk factors for brain tumors.[11] Low-dose irradiation of the scalp, once a treatment of tinea capitis, has been shown to cause a 10-fold increase in the incidence of meningiomas, many of which are anaplastic or malignant, and a threefold increase in the incidence of glial tumors. High-dose irradiation for intracranial or extracranial cancers, including prophylactic irradiation for leukemia,[12] increases the incidence of glioma, meningioma, and sarcoma. Dental radiography is probably not a risk factor.

Numerous other conditions and environmental exposures have been suggested as possible risk factors for brain tumors [*see Table 3*]; however, evidence to support a causal relationship is weak.[11] Head trauma may be a risk factor for the development of

Table 1 How Neoplasms Damage the Nervous System

Means of Damage	*Examples*
Primary tumors	Brain, spinal cord, and meninges Gliomas Meningiomas Cranial and peripheral nerves Schwannomas (neurilemomas) Neurofibromas Other intracranial structures Pituitary Pineal region
Metastatic tumors	Intracranial structures (usually brain) Spinal cord (usually epidural) Leptomeninges Peripheral nerves
Nonmetastatic effects	Vascular disorders Infections Metabolic and nutritional disorders Side effects of therapy Paraneoplastic syndromes

Table 2 Classification of Primary Brain Tumors

Tumor Type	*Examples*	*% Brain Tumors*	*% Gliomas*
Neuroepithelial tumors	*Gliomas (Total)*	*29*	*100*
	High-grade Gliomas (Subtotal)	*17*	*62*
	Anaplastic astrocytoma, oligodendroglioma		
	Glioblastoma		
	Low-grade Gliomas (Subtotal)	*12*	*38*
	Astrocytoma	*7*	20
	Pilocytic astrocytoma*	1	6
	Oligodendroglioma	3†	10†
	Ependymoma	1	1
Neuronal tumors	Neurocytoma Medulloblastoma* Neuroblastoma* Dysembryonic neuroepithelial tumor	2	
Mixed neuronal and glial tumors	Ganglioglioma Plesiomorphic xanthoastrocytoma		
Meningeal tumors	Meningioma	40	
Pineal region tumors	Pineocytoma Pineoblastoma General tumors*	< 1	
Pituitary tumors	Adenoma	10	
Primary CNS lymphomas	Non-Hodgkin lymphoma	1†	
Tumors of blood vessel origin	Hemangioblastoma Capillary hemangioblastoma Vascular malformations	1	
Nerve sheath tumors	Neurofibroma (schwannoma)	4	
Malformative tumors	Craniopharyngioma Colloid cyst of third ventricle	2	

*Usually children.
†Probably underestimated.

glial and other brain tumors.[13,14] HIV infection and other immunosuppressive disorders constitute risk factors for the development of PCNSL, probably glioma, and possibly leiosarcoma.[15,16] Exposure to electromagnetic radiation (including cellular telephones),[17-19] hair dyes, aspartame, pesticides, formaldehyde, or other industrial or occupational substances have not been proved to be causes of brain tumors.[11] If these are indeed risk factors, they cause only a small percentage of brain tumors.

Genetic Risk Factors

Most CNS tumors are sporadic, but some run in families [*see Table 4*].[20,21] Gliomas develop in as many as 14% of patients with neurofibromatosis type 1 (NF-1, or von Recklinghausen disease). Schwannomas, meningiomas, and, less commonly, ependymomas are found in patients with neurofibromatosis type 2 (NF-2). Both NF-1 and NF-2 tumor suppressor genes have been cloned, as have the genes responsible for other familial CNS tumors.[22] However, the low concordance of brain tumors in monozygotic twins indicates that inherited factors are a minor cause of brain tumor.[23]

Acquired (i.e., nonfamilial) genetic abnormalities are associated with CNS tumors [*see Table 5*]. These abnormalities include the loss or mutation of a tumor suppressor gene, such as *p53* or the retinoblastoma gene, and the amplification and rearrangement of oncogenes.[22] Many oncogenes code for growth factors or growth factor receptors that can self-stimulate tumor cells (i.e., autocrine stimulation) or stimulate cells nearby (i.e., paracrine stimulation). The two most common abnormalities identified in patients with glial tumors are mutations in *p53,* which occur in 40% of patients with astrocytoma and with approximately equal frequency in patients with anaplastic astrocytoma or glioblastoma multiforme, and overexpression of the epidermal growth factor receptor, in either normal or mutated form, which occurs in approximately 40% of patients with glioblastoma multiforme. Platelet-derived growth factor and its receptor have also been implicated. A sequence of genetic abnormalities can be identified in the progression of some tumors from low grade (astrocytoma or oligodendroglioma) to high grade (anaplastic glioma or glioblastoma multiforme).[5] However, glioblastoma multiforme can also arise de novo. Mutations in *p53* are rare in patients with glioblastomas who are younger than 18 years; they are common in patients between the ages of 18 and 45 years and are less common in patients older than 45 years. This suggests that glioblastomas in the young and elderly may arise via a different pathway or pathways than those that arise in midlife.

PATHOPHYSIOLOGY

The pathophysiology of CNS tumors explains why even relatively small growths often cause symptoms, whereas small systemic tumors usually do not.[1,2] First, small, strategically located tumors (e.g., tumors of the brain stem) damage vital pathways of the brain and spinal cord, thereby causing severe dysfunction.

Similar lesions in more homogeneous organs, such as the lung or liver, must destroy large areas of these organs before symptoms appear. Second, because the brain and spinal cord are encased in unyielding dura and bone, there is little room for a neoplasm to grow without compressing normal tissue [*see Figure 1*]. Third, tumor vessels do not possess a normal blood-brain barrier. Because these neovessels leak, proteins and other potentially noxious substances can enter the tumor and diffuse into the surrounding normal tissue, causing edema and further compressing normal structures [*see Figure 1*]. Edema removal is slow because the brain itself lacks lymphatics, although lymphatics around cranial nerves empty into cervical lymph nodes.[24] Tumors create their own vascular supply by secreting angiogenesis factors such as vascular endothelial growth factor. Angiogenesis factors are now targets for brain tumor therapy.[25]

Although small tumors often cause symptoms, exceptions do occur. Slowly growing tumors, especially those in the so-called silent brain areas (e.g., the frontal pole), may not cause symptoms until they become very large. Even glioblastoma multiforme sometimes causes fewer symptoms than might be expected from the size of the lesion and the degree of shift seen on imaging studies.

CNS symptoms are caused by three mechanisms:

1. The tumor invades, irritates, and replaces normal tissue. This mechanism is characteristic of infiltrating low-grade gliomas but rarely occurs in metastatic tumors or meningiomas.
2. Both the tumor and the edema compress normal tissue and its blood vessels, causing ischemia.
3. Tumors that compress or grow in the third or fourth ventricle (e.g., ependymomas) or the leptomeninges obstruct cerebrospinal fluid pathways, causing hydrocephalus.

The effects of tumor invasion, peritumoral edema, and hydrocephalus combine to cause herniation in normal cerebral structures under the falx cerebri, through the tentorium cerebelli, or through the foramen magnum [*see Figure 1*], thus causing symptoms at a distance from the tumor site.

Table 3 Possible Risk Factors for Primary Brain Tumors of Neuroepithelial, Meningeal, or Lymphocytic Origin[11]

Hereditary syndromes [*see Table 4*]
Ionizing radiation: therapeutic, diagnostic, and other sources
Family history of brain tumors
Immunosuppression
Constitutive polymorphisms
Lymphocyte mutagen sensitivity to gamma radiation
Prior cancers
Infectious agents or immunologic response: viruses, *Toxoplasma gondii*
Allergies
Head trauma
Epilepsy, seizures, or convulsions
Drugs and medications
Diet and vitamins: nitrosamine/nitrosamide/nitrate/nitrite consumption
Tobacco smoke exposure (women)
Alcohol use
Coffee use
Hair dyes and sprays
Traffic-related air pollution
Occupations and industries: synthetic rubber manufacturing, vinyl chloride, petroleum refining/production work, licensed pesticide applicators, agricultural work, and others
Cellular telephone use
Exposure to radio waves
Exposure to electromagnetic fields associated with power lines
Left-handedness (fewer gliomas in left-handed persons)
Sociodemographic status (more low-grade tumors in affluent; more high-grade tumors in Medicaid patients)

SYMPTOMS AND SIGNS

Patients with brain tumors may have generalized symptoms caused by diffuse intracranial pressure, focal symptoms caused by ischemia and compression at the site of the brain tumor, or false localizing symptoms caused by shifts of cerebral or spinal structures.[1] Generalized or false localizing symptoms are probably caused by slowly growing tumors in silent areas, whereas focal symptoms occur with even small tumors in more functionally important areas of the brain, such as the motor cortex and brain stem, or in the spinal cord.

Generalized Symptoms and Signs

Headache, the most common symptom of increased intracranial pressure, is the first symptom in about 40% of patients with a brain tumor.[26] It is more frequently the first symptom of brain tumor in those who have a history of headache than in those who have no such history. Most of the headaches associated with brain tumors are nonspecific. A brain tumor should be suspected when headache is present on awakening from sleep but disappears within 1 hour, when headaches begin in a middle-aged or an older person who has not previously experienced them, or when the character or severity of headache suddenly changes in a chronic headache sufferer. Localized headache is a reliable indicator of the side of the head that contains the tumor but does not mark the precise location of the tumor. For example, a right frontal headache indicates that the tumor is on the right side but does not indicate that the tumor is frontal; it could be occipital or even cerebellar.

Vomiting that is or is not preceded by nausea—particularly vomiting that occurs on awakening—is a common symptom of brain tumor in children but is less common in adults. Acute headache that is immediately followed by vomiting is characteristic of a brain tumor and indicates increased intracranial pressure. In contrast, a headache that is followed by vomiting several hours later is characteristic of migraine. Papilledema is a sign of increased intracranial pressure; it commonly occurs in children but is less common in the elderly. Papilledema is usually asymptomatic but may cause transient episodes of blindness.

The mental changes associated with a brain tumor begin with irritability and progress to apathy. All activities are performed more slowly. Patients sleep longer, seem preoccupied when awake, and often fail to initiate activity, including conversation. However, if spoken to, they usually respond appropriately. Psychiatric consultation for the treatment of what is thought to be depression is often obtained before a brain tumor is suspected.

Brain tumors may be associated with episodic symptoms such as headache, visual loss, altered consciousness, and sometimes transient weakness of the extremities. These episodes are often precipitated by rising from a recumbent position, cough-

Table 4 Familial CNS Tumor Syndromes

Disorders	*CNS Tumors*	*Tumors of Other Organs and Tissues*	*Skin Lesions*	*Genes*	*Chromosomes*
Neurofibromatosis-1	Gliomas, neurofibromas	Iris hamartomas, osseous lesions, pheochromocytoma, leukemia	Café au lait spots, cutaneous axillary freckling, neurofibromas	*NF1*	17q11.2
Neurofibromatosis-2	Schwannomas, meningiomas	Posterior lens opacities, retinal hamartoma	None	*NF2*	22q12.2
von Hippel-Lindau disease	Hemangioblastoma, endolymphatic sac	Retinal hemangioblastomas, renal cell carcinoma, pheochromocytoma, visceral cysts, endolymphatic sac tumor	None	*VHL*	3p25-p26
Tuberous sclerosis	Astrocytomas	Cardiac rhabdomyomas, adenomatous polyps of the duodenum and small intestine, cysts of the lung and kidney, lymphangioleiomyomatosis, renal angiomyolipoma	Cutaneous angiofibroma ("adenoma sebaceum"), peau de chagrin, subungual fibromas	*TSC1, TSC2*	9q34 16p13.3
Li-Fraumeni syndrome	Gliomas (10%)	Breast carcinoma; bone and soft tissue sarcomas; adrenocortical, lung, and GI carcinomas; leukemia	None	*TP53*	17p13.1
Cowden disease	Cerebellar mass (Lhermitte-Duclos disease)	Hamartomatous polyps of the eye, colon, and thyroid; breast carcinoma, thyroid cancer	Multiple trichilemmomas, fibromas	*PTEN*	10q22.3 10q23.31
Turcot A syndrome Turcot B syndrome	Medulloblastoma Glioma	Colorectal polyps, colon carcinoma Colon cancer, no polyps	Café au lait spots	*APC* *MLH1* *PMS2*	5q21-22 3p21.3 7p22
Nevoid basal cell carcinoma syndrome (Gorlin syndrome)	Medulloblastoma (anaplastic)	Jaw cysts, ovarian fibromas, skeletal abnormalities	Multiple basal cell carcinomas, palmar and plantar pits	*PTCH*	9q22.3-31
Retinoblastoma	Pineal tumors	Retinal tumors, osteosarcomas and other tumors	None	*RB1*	13q14
Bloom syndrome	Medulloblastoma, meningioma	Characteristic face and voice, gonadal failure, diabetes, immunodeficiency	Sun sensitivity, patches of hyperpigmentation and hypopigmentation	*BLM*	15q26.1
Fanconi anemia	Astrocytoma, medulloblastoma	Anemia, skeletal malformations, enlarged cerebral ventricles, gastrointestinal malformations	Café au lait spots, hyperpigmentation and hypopigmentation	*FANCA*	16q24.3
Familial melanoma	Astrocytoma	None	Patches of hyperpigmentation	*MLM* *CDKN2 A/p14 ARF*	1p36 9p21
Rhabdoid predisposition syndrome	PNET, choroid plexus carcinoma	Renal tumors, extrarenal malignant rhabdoid tumors	None	*HSNFA/ 1NH1*	22q11
Multiple endocrine neoplasia (MEN-1 Carney complex)	Pituitary adenomas	Hyperparathyroidism, gastrinoma, insulinoma, thyroid/bronchial carcinoid	Facial angiofibroma, lipomas, collagenomas	*MEN1*	11q13
Ataxia-telangiectasia	Astrocytoma, medulloblastoma, cerebellar ataxia	Lymphomas, hypogonadism, radiation sensitivity, insulin resistance, premature aging, small stature	Telangiectasias	*ATM*	11q22-q23

PNET—primitive neuroectodermal tumor

ing, or sneezing and are associated with plateau waves—abrupt increases in an already elevated intracranial pressure that last for 5 to 20 minutes.[1] These symptoms are not seizures; they respond to a decrease in the intracranial pressure but do not respond to anticonvulsant therapy.[27]

Focal Symptoms and Signs

Seizures are the most common focal sign of a brain tumor. Focal seizures are particularly common in patients who have tumors that compress the cortex, such as meningiomas, or that arise in or near the motor cortex or temporal lobe (both the motor cortex and the temporal cortex are epileptogenic areas). Focal seizures caused by frontal or temporal tumors often cause behavioral or emotional symptoms that are sometimes confused with panic attacks or psychological disorders. Focal seizures may progress to generalized convulsions, or a generalized seizure may arise from an asymptomatic focal discharge. Depending on the growth rate of the tumor, seizures may be present for months to years before other symptoms develop. Any patient with focal or generalized seizures that begin in adulthood should undergo diagnostic evaluation for a brain tumor.[28]

Other focal symptoms and signs of a brain tumor depend on the site of the lesion. These focal symptoms and signs (e.g., hemiparesis, aphasia, and visual field defect) are the same as

Table 5 Molecular and Cytogenetic Abnormalities Associated with Common Primary Brain Tumors

Tumor	*Abnormalities*	*Prevalence (%)*	*Gene*
Astrocytoma	17p loss or mutation	65	*p53*
	PDGFR overexpression	60	*PDGFR*
	22q (loss)		Unknown
	13q	25	*Rb*
Anaplastic astrocytoma	9p (loss)	50	*INKA-ARF*
	13p (loss)	25	*Rb*
	19q (loss)	50	Unknown
	11q (loss)		Unknown
	CDK4 amplification	10–20	*CDK4*
Glioblastoma	10q (loss)	80	*PTEN* (*MMAC, TEP1*)*
	EGFR amplification, rearrangement	40	*EGFR*
	17p (loss)	30	*p53*
	9p (loss)	70	*INK4A*
	13q	40	*RB*
	CDK4 amplification	10–20	*CDK4*
	MDM2 amplification	10	*MDM2*
Oligodendroglioma	19q (loss)	50–80	Unknown
	1p (loss)	40–92	Unknown
	17p (loss)	10–15	*p53*
	7 (gain)		*EGFR*
Anaplastic oligodendroglioma	9p (loss)		*INK4-ARF*
	10q (loss)		*?PTEN* (*MMAC1*)
	EGFR amplification		*EGFR*
	CDK4 amplification		*CDK4*
Medulloblastoma	9q (loss)	10–20	*ptcl*
	11p (loss)	30–45	Unknown
	17p (loss)	30–50	Unknown
Meningioma	1p (loss)		Unknown
	9q (loss)		Unknown
	10q (loss) (only malignant)		Unknown
	14q (loss)		Unknown
	17p (loss)		*p53*
	22q (loss)	60	*NF2*, ?other
Ependymoma	17p (loss)		Unknown
	22q (loss)	30	Unknown
Hemangioblastoma	3p (loss)	100	*VHL*

*Only 30% of 10q loss is attributed to *PTEN*.

those of CNS infection, stroke, or other structural diseases of the brain or spinal cord.[1]

False Localizing Symptoms and Signs

False localizing symptoms are caused by shifts of cerebral structures. Diplopia may result from displacement or compression of the sixth cranial nerve at the base of the brain. Hemianopsia may be caused by tentorial herniation that compresses the posterior cerebral artery. A number of other cranial nerve palsies associated with shifts of brain stem structures may also occur.[1]

DIAGNOSIS

A CNS neoplasm should be suspected in all adults with new-onset seizures and all patients with papilledema or new focal motor or sensory signs. Such patients warrant magnetic resonance imaging with the injection of contrast material (e.g., gadolinium) to determine whether the blood-brain barrier has been disrupted and, if so, where.[29] Patients with behavioral or personality changes should be evaluated by MRI if drowsiness, apathy, or memory loss accompanies the psychiatric symptoms or if the psychiatric symptoms are atypical. Patients whose only symptom is headache require MRI only if the headache began recently, has changed in character, or fails to respond to headache therapy [*see 11:VIII Headache*]. A negative MRI almost always rules out a tumor as the cause of the patient's symptoms or signs.

MRI identifies tumors that computed tomography misses and distinguishes tumors from vascular lesions. With the exception of biopsy, other laboratory tests are unnecessary. MRI may also suggest the histology of the tumor. Low-grade gliomas often fail to exhibit contrast enhancement; their presence is indicated by a hyperintense T_2-weighted image in conjunction with a normal or hypointense T_1-weighted image. The hyperintense area on the T_2-weighted image includes both the tumor and its

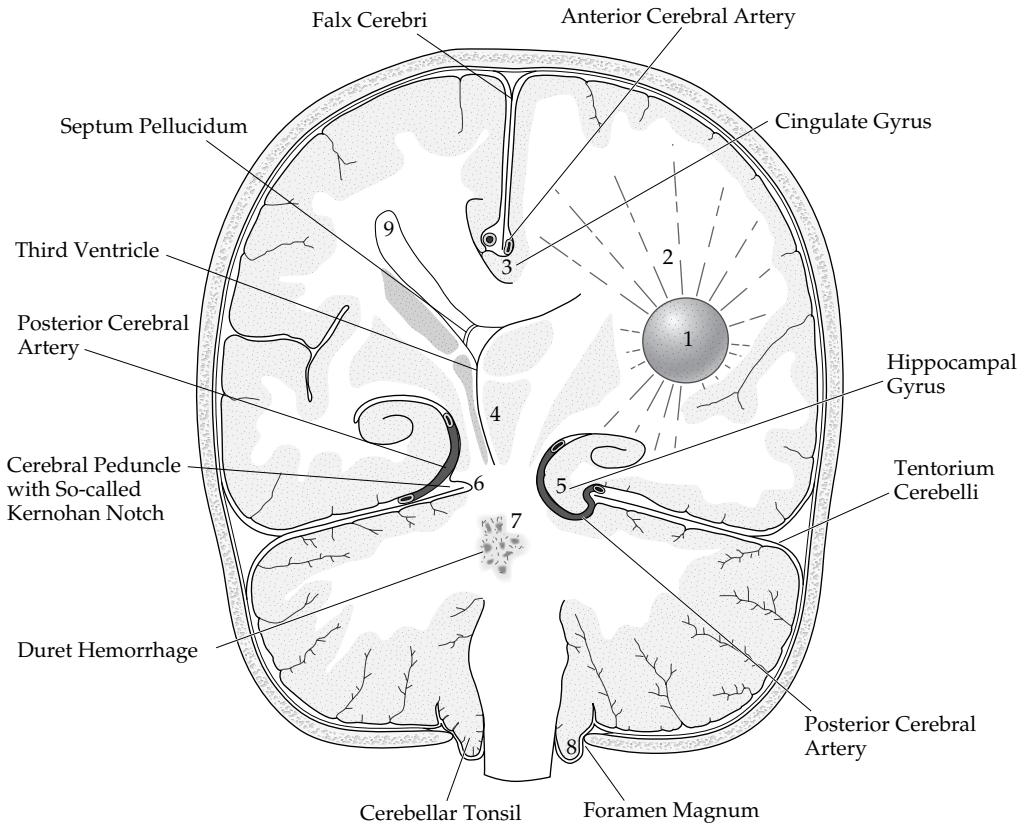


Figure 1 **This schematic drawing illustrates a brain tumor in the cerebral hemisphere (1), the surrounding edema (2), and several areas of herniation (3 through 8), which occur because of the increasing mass effect. Also illustrated is ventricular enlargement on the contralateral side caused by obstruction of the third ventricle (9).**

associated edema. High-grade gliomas usually exhibit contrast enhancement, with the T_1-weighted image showing an enhanced rim of irregular shape and thickness that surrounds a hypointense center [*see Figure 2a*]. The T_2-weighted image shows only hyperintensity. Although the enhancement does not encompass the entire infiltrating margin, it allows a clinically useful approximation of tumor volume to be made. A metastasis usually has a more regular, spherical enhancing rim than does a glioma [*see Figure 2b*]. Some low-grade tumors (e.g., meningiomas and pilocytic astrocytomas) also exhibit contrast enhancement. Metastases are more likely than gliomas to be multiple.

About 75% to 80% of patients with metastases have multiple lesions, whereas only 5% of gliomas appear at more than one site. Of patients with PCNSL, 40% have multiple tumors; these are located periventricularly, usually exhibit diffuse contrast enhancement, have poorly circumscribed margins compared with gliomas and metastases, and are usually surrounded by less edema than those tumors [*see Figure 2*].

Although MRI may suggest the histologic diagnosis, only biopsy is definitive. An exception to this rule occurs in PCNSL, in which lumbar puncture or, if the eyes are involved, vitrectomy may yield malignant cells. In other tumors, lumbar puncture poses a risk of cerebral herniation that outweighs any likely benefit.

Positron emission tomography (PET) with positron-emitting radionuclides, single-photon emission computed tomography (SPECT),[30] magnetic resonance spectroscopy (MRS), and perfusion-weighted images may prove to be useful noninvasive methods for determining the histologic type and degree of malignancy of a brain tumor.[31]

TREATMENT

The treatment of patients with CNS neoplasms depends on the nature and location of the neoplasm and the general condition of the patient. Treatments of specific tumors are detailed in the sections on individual tumors; certain principles that apply to all cases of CNS neoplasms are considered in this section.[32] Survival rates for patients with brain tumors have increased since the 1970s as a result of better diagnosis and therapy.[33]

Corticosteroids

Corticosteroids dramatically relieve the symptoms of most brain and spinal cord tumors by decreasing intracranial pressure and reducing the edema surrounding the lesions.[2] Symptomatic improvement may begin within minutes, and patients often become asymptomatic within 24 to 48 hours after administration. The mechanisms by which corticosteroids exert these effects are poorly understood. Corticosteroids decrease the flux of water-soluble agents across the disrupted blood-brain barrier, thus providing one mechanism for decreasing edema. Corticosteroids may also have the unintended effect of inhibiting the entry of chemotherapeutic agents into the tumor.

Corticosteroids are indicated in all symptomatic patients with brain and spinal cord tumors, with the exception of patients with PCNSL. Because of their lympholytic effects, cortico-

steroids may cause tumor necrosis, precluding definitive diagnosis of PCNSL. Therefore, when PCNSL is suspected, corticosteroids should be withheld until a biopsy is performed to establish the diagnosis.

Dexamethasone at a dosage of 16 mg/day is usually administered. Once treatment has begun, a corticosteroid is taken until the patient's symptoms are relieved and intracranial pressure is diminished. The drug is then tapered to the lowest dose commensurate with good neurologic function; it can often be completely discontinued.

Surgery

For intracranial tumors that arise outside the brain, such as meningiomas, schwannomas, and pituitary adenomas, surgery is often curative. Surgery for tumors that arise within the brain or spinal cord is rarely curative, although pilocytic astrocytomas and ependymomas are exceptions. Nevertheless, surgery is the most important treatment in patients with accessible tumors other than PCNSL. Surgery establishes the diagnosis, relieves intracranial pressure, and improves symptoms and seizure control.[34] With modern surgical techniques[35] and corticosteroid therapy to prevent postoperative swelling, the risk of worsening neurologic function is reduced. Even areas once thought surgically unapproachable, such as the insula and the corpus callosum, are now accessible through an operative approach.[36,37] Most studies indicate that more complete surgical removal of the gliomas improves both the duration of survival and quality of life.[38,39] If the tumor exhibits contrast enhancement before surgery, a postoperative contrast-enhanced MRI obtained within 3 days after resection accurately predicts the extent of residual tumor and thereby helps establish the prognosis; surgeons' estimates of the extent of resection are often inaccurate.

Radiation Therapy

Postoperative radiation therapy increases survival and improves the quality of life in patients with high-grade tumors (i.e., anaplastic astrocytomas or gliomas). However, its role in patients with low-grade tumors, particularly those that are asymptomatic, is uncertain.[40] After total resection of pilocytic astrocytomas, radiation therapy is not required. In patients with low-grade astrocytomas or oligodendrogliomas whose only symptoms are well-controlled seizures, radiation therapy may safely be deferred until symptoms develop.[40] High-grade gliomas should be treated with high doses of radiation (5,500 to 6,000 cGy in fractions of 180 to 200 cGy) delivered to a limited field that encompasses the tumor and its immediate surroundings.[34] New techniques of conformal field radiation planning, such as intensity-modulated radiotherapy (IMRT), decrease treatment-related morbidity and may enhance tumor control.[41] Stereotactic radiosurgery, in which high doses of ionizing radiation in multiple narrow beams are directed to a precise intracranial location by stereotaxy, is used in some centers to treat metastases, schwannomas, meningiomas, and some pituitary tumors and to boost conventional radiation for gliomas; its efficacy for gliomas has not been established.[42]

Chemotherapy

With the exception of oligodendrogliomas, germ cell tumors, and medulloblastomas, the response of most brain tumors to chemotherapy is limited; the addition of nitrosoureas or temozolomide to radiation therapy increases the survival of some patients with high-grade gliomas.[43,44] It is not possible to determine, in advance, which patients will respond to chemotherapy. Therefore, all patients with high-grade gliomas should be treated with a combination of radiation therapy and a nitrosourea or another lipid-soluble chemotherapeutic agent such as temozolomide.[45] Whether to give chemotherapy during radiation therapy, perhaps enhancing the effect of the radiation, or to begin chemotherapy after radiation therapy is open to debate. There is no evidence that chemotherapy given concurrently with radiation therapy enhances survival over chemotherapy administered after radiation therapy, nor is there any evidence that combination chemotherapy is superior to single-agent chemotherapy.

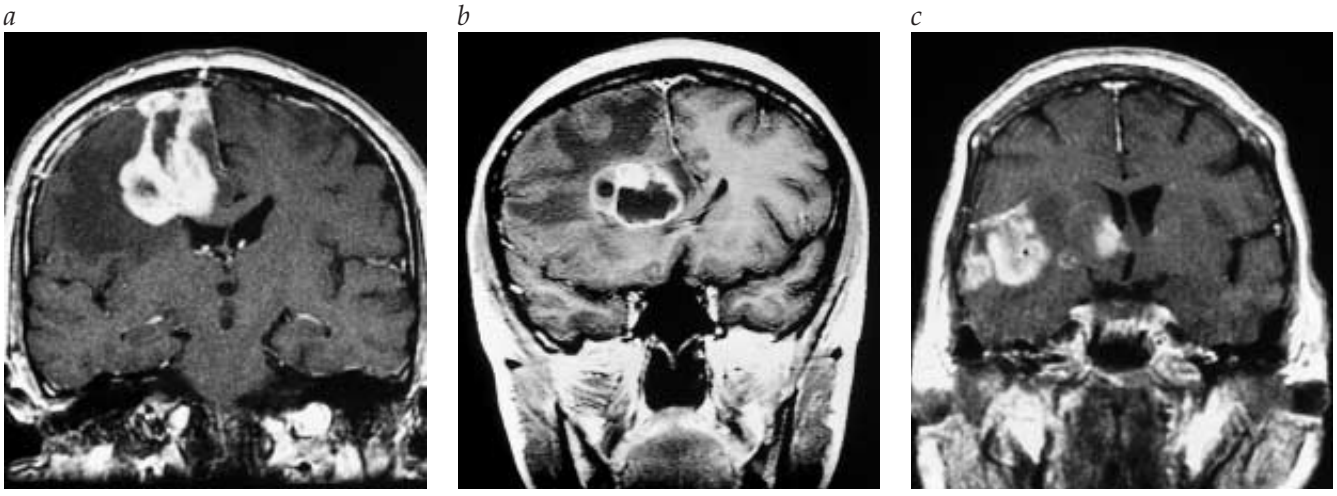


Figure 2 **(*a*) Magnetic resonance imaging of a frontal section shows the characteristics of a high-grade glioma. The tumor is irregular in shape, and the rim of contrast enhancement is thick and irregular in shape. The rim is surrounded by edema, as illustrated by the hypointensity around the area of contrast enhancement. (*b*) On MRI, the characteristics of a metastatic tumor are different from those of a high-grade glioma. The tumor is more spherical, and the contrasting edge is relatively thin and regular in shape. The thick area at the top of the tumor is a hemorrhage, not contrast enhancement. White-matter edema, as illustrated by the hypointensity surrounding the tumor, is massive. (*c*) MRI of a primary central nervous system lymphoma displays multiple lesions—with one in the periventricular area—that uniformly exhibit contrast enhancement. There is little surrounding edema. The edges of the tumor are less well defined than those of a high-grade glioma or a metastatic tumor.**

Experimental Therapies

Several new therapeutic approaches show promise in the treatment of brain tumors.[46] These include new delivery systems that use liposomes to enhance entry into the brain or that infuse or implant chemotherapy directly into the tumor bed. Other techniques include gene therapy, immunotherapy, and small molecules targeted to inhibit pathways necessary for tumor growth. None of these approaches have yet proved to be superior to conventional therapy.

PROGNOSIS

Histologic grade is an important prognostic factor for patients with primary brain tumors.[47] For example, the median survival of patients with glioblastoma multiforme is a little longer than 1 year, and the 5-year survival rate is less than 10%. Patients with anaplastic astrocytoma have a median survival of 2 to 3 years; for those with astrocytoma, median survival is 5 to 10 years. Patients with a pilocytic astrocytoma often have a normal life span. Many patients with meningiomas[48] and pituitary tumors[7] remain asymptomatic; in such patients, the tumor is often discovered incidentally when a brain image is done for another reason or at autopsy after the patient dies of another disease. Age is an important determinant of prognosis; younger patients survive longer than older patients despite having tumors of similar histology. Gene expression profiles may prove to be a better prognostic factor than histology.[49]

Primary Nervous System Tumors—Specific Diagnoses

ASTROCYTOMAS

Tumors believed to arise from astrocytes or their precursors are classified by the World Health Organization into four grades: pilocytic astrocytomas (grade I), diffuse astrocytomas (grade II), anaplastic astrocytomas (grade III), and glioblastomas multiforme (grade IV). The classification is made on histologic criteria.[5]

Pilocytic astrocytomas, which usually occur during childhood, are characterized by low cellularity and the presence of microcystic areas and eosinophilic intracytoplasmic masses called Rosenthal fibers. The tumors are well circumscribed on MRI and frequently exhibit contrast enhancement.

Diffuse astrocytomas are either isointense or hypointense on T_1-weighted MRI. They usually do not show contrast enhancement.

Anaplastic astrocytomas are characterized by nuclear atypia and mitotic activity. They may or may not show contrast enhancement.

Glioblastomas are characterized by nuclear atypia, mitosis, endothelial proliferation, and, sometimes, necrosis. They are poorly circumscribed and usually exhibit contrast enhancement.

Grade I

Pilocytic astrocytomas, which often occur in the cerebellum, are generally tumors of children. Whenever possible, they are treated surgically. Postoperative radiation therapy is usually unnecessary unless the tumor recurs. It has not been determined whether chemotherapy has any effect on pilocytic astrocytomas. The prognosis is excellent.

Grade II

Diffuse astrocytomas are tumors of young adults. Patients usually present with seizures; with time, other neurologic symptoms and signs may develop. Tumors preferentially occur in areas such as the insula and supplementary motor area, which are difficult sites to approach surgically.[50] For surgically accessible lesions, a maximally feasible resection is indicated.

Postoperative radiation therapy decreases tumor growth and may cause some shrinkage. Although radiation therapy of asymptomatic patients increases the time to development of symptoms, it does not increase survival.[51] Accordingly, it makes no difference in long-term prognosis whether the patient receives radiation therapy when the tumor is first discovered or later, after the tumor causes neurologic signs and symptoms. Because radiation therapy can cause neurotoxicity—particularly neurotoxicity affecting cognitive function—many authors recommend deferring radiation therapy in patients who have no symptoms or signs other than seizures that are controllable by anticonvulsants until such time as neurologic signs and symptoms develop. It has not been determined whether chemotherapy has any significant effect on these tumors.

Grade II astrocytomas often progress to grade III or IV tumors. The short-term prognosis is good; however, even if the tumor remains of low grade, it continues to grow at a steady rate, and the median survival is only a few years.[52]

Grades III and IV

Anaplastic astrocytomas and glioblastomas multiforme (malignant gliomas) are treated with a combination of surgery, radiation therapy, and chemotherapy.[45] A neurosurgeon skilled in the management of brain tumors should attempt to remove all grossly visible tumor. Surgical resection should be followed by radiation therapy to the tumor bed and to an area 2 to 3 cm around the tumor. There is no evidence that a radiosurgical boost to the tumor bed improves survival. Radiation therapy should be followed by chemotherapy. A number of chemotherapeutic agents have proved to have some modest effect. Most physicians use either lomustine or temozolomide.

Despite treatment, the prognosis is poor. Most such tumors recur within 1 to 2 years. The median survival for patients with glioblastomas is about 1 year; for patients with anaplastic astrocytomas, median survival is slightly longer. Age is a major prognostic factor, with younger patients having a better prognosis than older patients.

OLIGODENDROGLIOMAS

Oligodendrogliomas are believed to arise from precursors of oligodendroglial cells—cells that are responsible for the production of myelin within the CNS. Oligodendrogliomas are often difficult to distinguish histologically from astrocytomas, but they are genetically different. Most oligodendrogliomas have deletions of the short arm of chromosome 1 (1p deletion) and the long arm of chromosome 19 (19q deletion).[49] These deletions are associated with chemosensitivity and longer survival.[49] Histologically, oligodendrogliomas are characterized by cells that have round, homogeneous nuclei; on paraffin sections, cells are seen to have a swollen, clear cytoplasm (i.e., a fried-egg appearance). Lesions may be calcified and contain a dense network of delicate, branching capillaries. Unless calcification is present, oligodendrogliomas appear no different from astrocytomas on MRI. Sometimes, hemorrhage can be identified within the tumor. Hemorrhage is more likely to occur with oligodendrogliomas than with astrocytomas.

Patients who are asymptomatic, including those with large low-grade oligodendrogliomas, may not require treatment. Pa-

tients with anaplastic oligodendrogliomas and patients with low-grade tumors who are symptomatic should be treated. The treatment of choice for initial therapy is surgical resection; the resection should be as extensive as possible. Patients with a 1p/19q deletion usually respond to chemotherapy.

The established chemotherapeutic regimen is combination therapy with procarbazine, vincristine, and carmustine. Other chemotherapeutic agents, particularly temozolomide, may be equally efficacious.[53] It has not been determined, especially for oligodendrogliomas of lower grade, whether chemotherapy alone suffices or whether chemotherapy should be followed by radiation therapy. In patients without a 1p deletion, postoperative radiation therapy is probably the treatment of choice.

With oligodendrogliomas of all grades, the prognosis is substantially better than that for patients with comparable astrocytic tumors.

MENINGIOMAS

Meningiomas arise from the dura and the arachnoid villi of the intracranial and spinal spaces. They are found along the dorsal surface and base of the brain, the falx cerebri, and the sphenoid ridge, as well as in the lateral ventricles and the thoracic spinal canal. More common in women than in men, meningiomas possess progesterone receptors; some also express estrogen receptors.[54] The incidence of meningiomas is increased in women with breast cancer.

The most frequent genetic abnormality is allelic loss of chromosome 22q12. Atypical and anaplastic tumors may have losses of 1p, 9q, 10q, 14q, and 17p.[55] Meningiomas may be multiple and often grow slowly, or sometimes not at all.[56]

Meningiomas cause symptoms by brain compression, not by invasion. The common presenting symptoms of patients with brain meningiomas are seizures and cranial nerve palsies; for patients with spinal cord meningiomas, the most common presenting symptom is a gradual and sometimes painless paraparesis.

Meningiomas can be easily diagnosed with contrast-enhanced MRI; they are of the same intensity as brain tissue on T_1-weighted MRI. They exhibit uniform and intense contrast enhancement and are clearly outside the brain or spinal cord. An elongated dural attachment (i.e., a dural tail) can sometimes be identified. Occasionally, meningiomas appear en plaque, causing widespread thickening of the dura.

Patients with meningiomas can often be cured by complete surgical removal of the tumor or tumors. However, many meningiomas, particularly those in the cavernous sinus and at the base of the brain, cannot be completely resected. In addition, a few meningiomas, especially those that are anaplastic, recur despite apparently complete resection. Radiation therapy may control their regrowth.[57] Radiosurgery may control small lesions.[57] Chemotherapeutic and hormonal agents have not yet demonstrated benefit. Because meningiomas have a slow rate of growth, patients with small or asymptomatic meningiomas can usually be managed with follow-up examinations rather than resection; this is particularly the case with elderly patients.

If histology shows the meningioma to be benign, total resection is usually curative. Patients with partially resected meningiomas or those with anaplastic or malignant meningiomas may benefit from postoperative radiation therapy. In any event, patients who have undergone total resection should receive follow-up examinations because even after complete resection, some tumors recur.

SCHWANNOMAS

Schwannomas (also called Schwann cell tumors or neurilemomas) occur in the cranium or along peripheral nerves, both within and outside the spinal canal. Vestibular schwannomas, also called acoustic neuromas, arise from the vestibular portion of the eighth cranial nerve near or in the internal auditory canal where Schwann cells replace oligodendroglia as myelin producers.

Schwannomas usually grow slowly, and some do not grow at all.[58] Initially, they cause asymptomatic unilateral vestibular failure; later, symptomatic unilateral hearing loss occurs. Untreated tumors may grow large, compressing the nearby trigeminal and facial nerves and the cerebellum. The tumors are often associated with loss of the *NF2* tumor suppressor gene on chromosome 22q. Bilateral vestibular schwannomas are virtually pathognomonic for NF-2. Patients with unexplained unilateral hearing loss, tinnitus, or vertigo should undergo further evaluation by MRI. MRI is especially indicated if the caloric test (i.e., stimulation of the semicircular canal on the ipsilateral side through alteration of the temperature of the eardrum by use of warm and cold water) fails to elicit nystagmus; such a finding indicates vestibular nerve failure.

Early diagnosis with auditory testing and MRI enables the surgeon to remove the tumors by microsurgery, which preserves facial nerve function and sometimes hearing. Some investigators advocate the use of stereotactic radiosurgery in patients with small vestibular schwannomas. Patients with small, asymptomatic lesions may choose a course of watchful waiting in lieu of undergoing treatment.[59]

If the tumors are treated early with either surgery or radiosurgery, the prognosis is excellent. Although larger tumors can be treated successfully, treatment may result in complete deafness and facial paralysis. Headaches are a common complication of surgery for vestibular schwannomas[60]; however, CSF leaks are uncommon.[61]

PITUITARY ADENOMAS

When pituitary adenomas are symptomatic, patients usually present with endocrine symptoms. Headache, another common symptom (70% prevalence in one series), is unrelated to tumor size.[62] Macroadenomas (tumors > 1 cm in diameter) may breach the diaphragma sellae and compress the optic chiasm, causing bitemporal hemianopsia. The mass may extend laterally into the cavernous sinus, causing diplopia by compressing the ocular nerves. Visual loss and diplopia occur acutely when the tumor bleeds or infarction develops (a condition known as pituitary apoplexy).

Patients with pituitary tumors can be treated with transsphenoidal resection,[63] radiation therapy,[64] or radiosurgery[65]; patients with prolactinomas, and perhaps thyrotropin-secreting tumors,[66] can be effectively treated with bromocriptine or related dopamine agonists.[67] Pharmacologic therapy is also effective in some patients with growth hormone–secreting tumors.[68] Most patients with tumors that compress the optic chiasm are treated surgically. With appropriate treatment, most patients with pituitary adenomas are cured.

PINEAL REGION TUMORS

Tumors of the pineal region compress the upper brain stem, causing paresis of upward gaze (Parinaud syndrome), pupillary dysfunction, and hydrocephalus through compression of the aqueduct of Sylvius. Patients first complain of headache, and they may develop papilledema from increased intracranial pres-

sure caused by the hydrocephalus. Double vision and ataxia from compression of the upper brain stem may also be presenting symptoms.

Pineal parenchymal tumors may be low grade (pineocytoma), high grade (pineoblastoma), or intermediate grade. The more malignant varieties usually occur in children and often seed the leptomeninges.[69] Surgery often cures pineocytomas and establishes the diagnosis for patients with the more malignant varieties. For pineocytomas that are only partially resected and for the more malignant tumors, postoperative radiation with or without chemotherapy is indicated.[70-72]

Tumors of the pineal region also arise from germ cell structures and include germinomas, yolk sac tumors, and choriocarcinomas. Some germ cell tumors spill biochemical markers into the CSF, making noninvasive diagnosis possible (especially in choriocarcinomas releasing β–human chorionic gonadotropin and in yolk sac tumors releasing α-fetoprotein). Patients with germinomas can be cured by radiation therapy[73] or by a combination of chemotherapy and radiation therapy.[74,75] Most other germ cell tumors are incurable.[76]

PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA

PCNSL arises from B cells that are believed to enter the CNS from the bloodstream either by routine trafficking of B cells across the blood-brain barrier or in response to tissue injury.[77] It is not known whether the B cells that enter the brain are already malignant or are transformed in the brain. PCNSL differs from systemic lymphoma in several respects. Although PCNSL is localized, it is less likely to be cured than systemic lymphoma.[78] Furthermore, when systemic lymphoma metastasizes to the CNS, it usually seeds the leptomeninges, whereas PCNSL usually presents as a brain mass. Less commonly, PCNSL begins in the eye, spinal cord, or leptomeninges.

Headache and hydrocephalus may result from leptomeningeal tumors. If the eyes are affected, visual difficulties that suggest uveitis or vitreitis may be the first complaint. Because the mass lesions in the brain are generally deep, patients often present with behavioral and personality change rather than the seizures that characterize gliomas. PCNSL grows rapidly and spreads widely throughout the brain and spinal cord but rarely metastasizes outside the CNS. Multiple lesions are common.

Diagnosis of PCNSL can be made by vitrectomy (in cases of eye involvement), CSF examination (in cases of leptomeningeal tumor), or needle biopsy of the brain. In all cases, diagnosis is made by identification of malignant lymphocytes (usually B cells), most of which are of the cleaved, large cell, or immunoblastic type. Untreated patients die within several months after diagnosis.

Treatment consists of high-dose, methotrexate-based systemic chemotherapy with or without subsequent radiation therapy.[78] Compared with radiation therapy alone, chemotherapy increases the median time to relapse; for some patients, this treatment is curative. HIV-positive patients with PCNSL are also treated with this regimen. Treatment with highly active antiretroviral therapy (HAART) has been effective in some patients with AIDS.

MEDULLOBLASTOMAS

Medulloblastomas are highly aggressive tumors that occur most often in children[79]; they arise from primitive neuroectodermal cells in the vermis of the cerebellum. Early symptoms include truncal ataxia, headache, nausea, and vomiting. Headache, nausea, and vomiting may be secondary to early hydrocephalus. In both adults and children, the tumors often seed the leptomeninges. Treatment consists of surgical removal followed by whole neuraxis irradiation and adjuvant chemotherapy.[79,80] Cure is effected in over 70% of patients with medulloblastomas. Unfortunately, despite cure of the tumor, many patients suffer cognitive defects and growth failure as a result of the neuraxial radiation.[81]

SPINAL CORD TUMORS

Spinal cord tumors can be epidural, subdural but extramedullary, or intramedullary.[82] The site usually cannot be determined clinically but is easily established by MRI. Spinal cord tumors are characterized by myelopathy, which usually consists of paraparesis or quadriparesis, hyperactive reflexes, Babinski signs, and loss of sensation below the lesion. Local pain and sometimes radicular pain appear weeks or months before the myelopathy develops. Bladder and bowel functions are usually spared until late in the course of disease.

Epidural tumors are usually metastases. Subdural tumors are usually benign meningiomas or schwannomas. Occasionally, leptomeningeal metastases grow large enough to cause a subdural mass. The signs and symptoms of subdural meningiomas and schwannomas develop slowly; pain may be absent, particularly in patients with meningiomas. The treatment of meningiomas and schwannomas is surgery.

Intramedullary tumors include ependymomas, astrocytomas, and metastases. Patients with ependymomas of the spinal cord are often cured by surgical removal of the tumor.[83] Patients with ependymomas that arise below the spinal cord in the filum terminale can also be cured if the tumor can be totally removed by surgery; partial resections should be followed by radiation therapy.[84] Astrocytomas of the spinal cord cannot be totally resected. Patients with these tumors are treated with postoperative radiation therapy.[85] There is no evidence that chemotherapy is beneficial.

Metastases to the Nervous System

Metastases to the nervous system are far more common than primary CNS tumors.[86] Symptomatic CNS metastases often considerably decrease the quality of life of patients who are otherwise functional, despite having cancer. Early diagnosis of CNS metastases and vigorous treatment often lead to amelioration of CNS symptoms and improvement in the patient's quality of life, even though survival may not be prolonged. With the development of more effective treatment for tumors outside the nervous system, CNS metastases, which are protected by the blood-brain barrier, are becoming increasingly more common. Metastases affect the brain by hematogenous spread to brain parenchyma and affect the spinal cord by epidural compression. Cranial and spinal nerves are affected by leptomeningeal seeding or by direct compression by bone or lymph node metastases. Brain metastases may be the presenting symptom in as many as 10% of patients with lung cancer who are otherwise asymptomatic. Routine brain scans of patients with newly diagnosed lung cancer show asymptomatic metastases in 5% to 10% of these patients. Epidural spinal cord compression is the presenting sign of cancer in about 20% of patients.[87]

Tumors that commonly metastasize to the brain include lung and breast cancers and malignant melanoma. Uterine, ovarian, and prostate cancers rarely cause brain metastases. Epidural spinal cord compression is a common complication of cancers of the prostate, breast, and lung. Leptomeningeal metastases com-

plicate breast cancer, leukemia and lymphoma, melanoma, and small cell lung cancer.

BRAIN METASTASES

The signs and symptoms of brain metastases are the same as those of primary brain tumors.[2] MRI characteristically identifies one or more spherical lesions within the white matter that exhibit contrast enhancement and are surrounded by edema. With small lesions, the contrast enhancement may be uniform. With larger lesions, a symmetrical rim of contrast enhancement surrounds a central nonenhanced area. MRI cannot unequivocally distinguish these lesions from brain abscesses or primary tumors. Dura-based metastases appear similar to meningiomas but usually cause more brain edema. In certain clinical settings (e.g., in a patient with multiple lesions in the brain or in a patient with one lesion who is known to have active cancer), MRI is diagnostic. If the image is atypical or if the patient is not known to have cancer, biopsy may be required. In one series of presumed single-brain metastasis, biopsy revealed an 11% error rate; several of the lesions were benign.[88]

Use of corticosteroids in the treatment of brain metastases is the same as in the treatment of primary brain tumors. In a patient whose systemic cancer is controlled, single, surgically accessible brain metastases should be excised,[89] especially if they are greater than 3 cm in diameter and are therefore too large for radiosurgery. Randomized trials indicate that whole brain irradiation after surgical excision decreases recurrence.[90,91] Patients with multiple or surgically inaccessible metastatic tumors are treated with whole brain irradiation to eradicate both the visible tumor and micrometastases that may be present elsewhere in the brain. Stereotactic radiosurgery may control metastases smaller than 3 cm in diameter and has been used to treat multiple lesions.[92] Metastatic brain tumors that disrupt the blood-brain barrier often respond as well to chemotherapy, as does the primary tumor.[93,94] Such tumors can be identified on contrast-enhanced MRI. Thus, when dealing with small, relatively asymptomatic tumors in a patient requiring chemotherapy for tumors outside the CNS, it may be worthwhile to treat the brain metastasis with chemotherapy before initiating radiation therapy. The brain acts as a sanctuary for small, nonenhancing metastases; however, disruption of the blood-brain barrier caused by larger tumors allows even water-soluble chemotherapeutic agents to enter the brain. The brain sanctuary explains the frequent occurrence of isolated CNS relapse in patients with peripherally controlled breast cancer.

SPINAL CORD METASTASES

Metastases to vertebral bodies cause symptoms by compressing the spinal cord.[95] Epidural spinal cord compression by a metastatic tumor is a neurologic emergency. At our institution, treatment consists of high-dose dexamethasone, 100 mg by I.V. bolus, followed by 100 mg/day in divided doses for 3 days; the drug is then rapidly tapered. Other institutions use smaller doses. Definitive therapy entails irradiation with or without surgical decompression. A study by Patchell and colleagues indicated that patients treated with radical decompressive surgery plus postoperative radiation therapy regain the ability to walk more often and maintain it longer than patients treated with radiation alone.[96] Surgical procedures include laminectomy or removal of a vertebral body.[97,98] Patients who are ambulatory when treated usually remain so after treatment.[95] Those who are not ambulatory may regain the ability to walk; those who are paraplegic rarely do so.

LEPTOMENINGEAL METASTASES

The diagnosis of leptomeningeal metastases is established by cytologic identification of malignant cells in the CSF, by contrast enhancement of the brain or spinal leptomeninges on MRI,[99,100] or by the presence of tumor markers in CSF when compared with serum. Tumors that metastasize from non-Hodgkin lymphoma or leukemia usually respond to the intrathecal administration of drugs such as methotrexate or cytarabine. Administration of radiation therapy to symptomatic areas often affords significant palliation; many solid tumors also respond. Systemic chemotherapy using drugs that produce therapeutic levels in the CNS may be equally efficacious. Disruption of the blood-CSF barrier by neovascularization of the tumor may allow water-soluble chemotherapeutic agents to cross the blood-brain barrier. Nevertheless, with the exception of some patients with leptomeningeal metastases from breast cancer and patients with meningeal lymphoma or leukemia, the prognosis is poor, whatever therapy is applied.

Nonmetastatic Complications of Cancer

Cancer may profoundly affect the nervous system even when it does not metastasize [*see Table 1*]. The nonmetastatic complications of cancer discussed here include paraneoplastic syndromes and the side effects of anticancer therapy.

PARANEOPLASTIC SYNDROMES

Paraneoplastic syndromes involve remote effects of cancer on the nervous system.[4] Paraneoplastic syndromes are caused by cancer, but the symptoms cannot be ascribed to metastases or identifiable infectious, vascular, nutritional, toxic, or metabolic disorders. Most paraneoplastic syndromes are believed to result from an immune attack mounted against an antigen that is expressed by both the cancer and the nervous system.[4] The immune response often controls growth of the cancer but damages the part of the nervous system that expresses the same or similar antigens. The best example of this mechanism occurs in the Lambert-Eaton myasthenic syndrome (LEMS), in which a humoral immune response develops against calcium channel proteins in small cell lung cancer cells. The antibody binds to calcium channels on the presynaptic side of the cholinergic synapse, blocking the calcium uptake necessary for acetylcholine release, resulting in muscle weakness.

Paraneoplastic syndromes can affect any nervous system structure. They cause symptoms that mimick similar, presumably autoimmune, syndromes not associated with cancer. The neurologic symptoms usually precede identification of a cancer that may be too small to be detected. Several clinical characteristics suggest a paraneoplastic syndrome.[2,4]

1. The neurologic disorder usually has a rapid course that evolves over weeks or months and then stabilizes.
2. The incidence of certain clinical syndromes (e.g., LEMS) is substantially higher in cancer patients than in the general population. For example, in 60% of patients presenting with LEMS, cancer—usually small cell lung cancer—is the cause. In about 50% of patients presenting with a subacute onset of pancerebellar degeneration, the disorder is paraneoplastic.
3. In some patients with a given clinical syndrome, specific serum antibodies react against both the nervous system and the cancer. The presence of these autoantibodies identifies the clinical disorder as paraneoplastic and suggests the type of

neoplasm involved [*see Table 6*]. These autoantibodies include anti-Yo antibodies found in patients with the paraneoplastic cerebellar degeneration associated with gynecologic cancers; antimyelin-associated glycoprotein antibodies (anti-MAG) found in patients with the peripheral neuropathy associated with lymphoma and Waldenström disease; anti-Hu antibodies found in patients with the sensory neuropathy or encephalomyelitis associated with small cell lung cancer; and antibodies to the cancer-associated retinopathy antigen. Tests for many of these antibodies are commercially available.

Paraneoplastic Cerebellar Degeneration

Paraneoplastic cerebellar degeneration is characterized by the rapid development of bilateral, usually symmetrical, cerebellar dysfunction, including arm and leg ataxia and severe dysarthria. Vertigo, diplopia, and nystagmus are common. CSF findings include pleocytosis and elevated protein and IgG concentrations. The disease, which may be associated with any cancer, usually precedes discovery of the neoplasm by a few weeks to 3 years. MRI, particularly when performed several months after the illness begins, shows cerebellar atrophy. Anti-Purkinje cell antibodies are found in the serum of some patients, especially patients with ovarian cancer.[101] Pathologic changes are often restricted to loss of Purkinje cells throughout the cerebellum. Lymphocytic cuffs may be encountered around blood vessels, particularly in deep cerebellar nuclei.

Clinically, paraneoplastic cerebellar degeneration can be distinguished from cerebellar metastases by the symmetry of its signs and the absence of increased intracranial pressure. It differs from alcoholic cerebellar degeneration in that dysarthria and ataxia of the upper extremities are prominent features of the paraneoplastic form but are usually mild or absent in the alcoholic form. It differs from the hereditary types of cerebellar degeneration in that the latter rarely have as rapid a course.

Paraneoplastic cerebellar degeneration runs its course over weeks to months, usually rendering the patient severely disabled. In rare instances, the disorder improves with successful treatment of the tumor.

Paraneoplastic Opsoclonus

Paraneoplastic opsoclonus (spontaneous, conjugate, chaotic, and uncontrollable eye movements) is associated with cerebellar ataxia and myoclonus of the trunk and extremities. It occurs most commonly in children as a remote effect of neuroblastoma. It also occurs in adults with various cancers.[101] In children, the neurologic symptoms respond to treatment with corticosteroids and antitumor therapy.

Paraneoplastic Sensory Neuronopathy

Paraneoplastic sensory neuronopathy is marked by loss of sensation with relative preservation of motor power. The illness usually precedes the appearance of cancer—typically, small cell lung cancer (in which case it is known as the anti-Hu syndrome)[101]—and progresses over a few months, leaving the patient moderately or severely disabled. CSF pleocytosis is common at onset but resolves in a few weeks. Pathologic findings include destruction of the dorsal root ganglia and the posterior columns of the spinal cord; perivascular lymphocytic cuffing;

Table 6 Antineuronal Antibody Associated Paraneoplastic Disorders

Antibody	*Neuronal Reactivity*	*Protein Antigens*	*Cloned Genes*	*Tumor*	*Paraneoplastic Symptoms*
Anti-Hu (ANNA1)	Nucleus/cytoplasm (all neurons)	35-40 kd	*HuD, HuC Hel-N1*	SCLC, neuroblastoma, prostate,	PEM, PSN, PCD, autonomic dysfunction
Anti-Yo (PCA-1)	Cytoplasm Purkinje cells	34, 62 kd	*CDR34, CDR62*	Ovary, breast, lung	PCD
Anti-Ri	Nucleus/cytoplasm (CNS neurons)	55, 80	Nova	Breast, gyn, lung, bladder	Ataxia/opsoclonus
Anti-Tr	Cytoplasm Purkinje cells	?	—	Hodgkin	PCD
Anti-VGCC	Presynaptic NMJ	VGCC 64 kd	P/Q type *MysB*	VGCC, SCLC	LEMS
Antiretinal	Photoreceptor, ganglion cells	23, 65, 145, 205 kd	Recoverin	SCLC, melanoma, gyn	CAR, MAR
Anti-amphiphysin	Presynaptic NMJ	128 kd	Amphiphysin	Breast, SCLC	Stiff-person syndrome PEM
Anti-CRMP5 (Anti-CV2)	Oligodendrocytes, neurons, cytoplasm	66 kd	*CRMP5 (POP66)*	SCLC, thymoma	PEM, PCD chorea, sensory neuropathy
Anti-PCA2	Purkinje cytoplasm and other neurons	280 kd	—	SCLC	PEM, PCD, LEMS
Anti-Ma1	Neurons (subnucleus)	40 kd,	*Ma1*	Lung, others	Brain stem, PCD
Anti-Ma2	Neurons (subnucleus)	41.5 kd	*Ma2*	Testis	Limbic brain stem encephalitis
ANNA3	Nuclei, Purkinje cells	170 kd	—	Lung	Sensory neuropathy PEM
Anti-mGluR1	Purkinje cells, olfactory neurons, hippocampus	Metabotropic glutamite receptor	—	Hodgkin	PCD
Anti-VGKC	Peripheral nerve	VGKC	Potassium channels	Thymoma, SCLC	Neuromyotonia
Anti-MAG	Peripheral nerve	MAG	*MAG*	Waldenstrom	Peripheral neuropathy

CAR—cancer-associated retinopathy Gyn—gynecologic cancer LEMS—Lambert-Eaton myasthenic syndrome MAG—myelin-associated glycoprotein
MAR—melanoma-associated retinopathy NMJ—neuromuscular junction PCD—paraneoplastic cerebellar degeneration PEM—paraneoplastic encephalomyelitis
PSN—paraneoplastic sensory neuronopathy SCLC—small cell lung cancer VGCC—voltage-gated calcium channel VGKC—voltage-gated potassium channel

Table 7 Neurotoxicity Caused by Chemotherapeutic Agents

Acute Encephalopathy (Delirium)
Corticosteroids*
Methotrexate (high-dose I.V. or standard I.T.)*
Ifosfamide/mesna*
Interferons*
Interleukin-2*
Cyclophosphamide (high-dose)
Cisplatin
Vincristine
Asparaginase
Procarbazine
Fluorouracil (with or without levamisole)
Cytarabine (high-dose)
Nitrosoureas (high-dose or I.A.)
Cyclosporine
Tamoxifen
Etoposide (high-dose)
Paclitaxel (high-dose)

Chronic Encephalopathy (Dementia)
Methotrexate*
Carmustine (I.A.)*
Cytarabine
Fludarabine

Visual Loss
Nitrosoureas (I.A.)*
Tamoxifen
Cisplatin
Fludarabine

Cerebellar Dysfunction/Ataxia
Cytarabine (high-dose)*
Phenytoin*
Fluorouracil + allopurinol*
Procarbazine
Vincristine
Cyclosporine
Ifosfamide

Aseptic Meningitis
Methotrexate (I.T.)*
Immune globulin (I.V.)*
Metrizamide*
Trimethoprim-sulfamethoxazole
Nonsteroidal anti-inflammatory drugs
Levamisole
Cytarabine (I.T.)
Carbamazepine

Headache without Meningitis
Retinoic acid*
Corticosteroids (withdrawal)*
Trimethoprim-sulfamethoxazole
Cimetidine
Tamoxifen
Cisplatin

Seizures
Methotrexate
Etoposide (high-dose)
Cisplatin
Vincristine
Asparaginase
Carmustine
Busulfan (high-dose)
Cyclosporine
Iodinated contrast material (I.V. or I.T.)
Ifosfamide

Myelopathy (Caused by I.T. Drugs)
Methotrexate
Cytarabine
Thiotepa

Peripheral Neuropathy
Vinca alkaloids*
Cisplatin*/oxaliplatin*
Paclitaxel*/docetaxel
Suramin
Procarbazine
Teniposide
Etoposide
Cytarabine (high-dose)
Ifosfamide

*Common causes of complications.
I.V.—intravenous I.T.—intrathecal I.A.—intra-arterial

and wallerian degeneration of sensory nerves. Many patients also have inflammatory and degenerative changes in the brain and spinal cord. In some patients, sensory neuronopathy improves when the tumor is treated; in a few patients, the neuronopathy improves spontaneously or responds to immunosuppression (e.g., plasma exchange or intravenous gamma globulin). However, in most patients, the neuronopathy does not improve, even after the tumor is cured.

Lambert-Eaton Myasthenic Syndrome

LEMS is characterized by proximal muscle weakness with ptosis, dry mouth, and, in men, impotence.[101] Physical examination reveals that the patient has greater strength than would be expected given the patient's complaints, and that when the patient makes a sustained effort, strength increases over several seconds. Deep tendon reflexes at the knees and ankles are usually absent but sometimes return transiently with exercise. The diagnosis is made by electrical studies. In patients with LEMS, neuromuscular transmission increases with repetitive stimulation—the opposite of what occurs in patients with myasthenia gravis. Antibodies against P/Q-type calcium channels are found in the patient's serum. Successful treatments are plasma exchange and eradication of the underlying cancer.

NEUROTOXICITY CAUSED BY CHEMOTHERAPEUTIC AGENTS

Anticancer chemotherapy can cause a number of neurotoxic side effects [*see Table 7*], the most prominent being peripheral neuropathy and encephalopathy.[102,103] When neurologic dysfunction occurs in patients who are undergoing chemotherapy, even in those who are taking drugs that are usually not considered neurotoxic, the presence of drug-related neurotoxicity should be considered and standard references consulted for precedents.

Peripheral Neuropathy and Myopathy

The vinca alkaloids, particularly vincristine, affect the microtubules, causing a sensorimotor peripheral neuropathy.[102,103] Virtually all patients treated with vincristine complain of paresthesias in the fingertips and sometimes in the toes. Ankle reflexes disappear early. Occasionally, patients develop motor weakness that occurs in a diffuse distal distribution or in individual cranial or peripheral nerves. Autonomic neuropathy may also occur. These symptoms are reversible.

Cisplatin causes a large-fiber sensory neuropathy by damaging dorsal root ganglia. This dose-related disorder occurs in patients who receive more than 450 mg/m^2. Symptoms often begin several weeks after completion of cisplatin therapy and progress for several months. Loss of proprioceptive sensation may be severe enough to prevent the patient from walking and effectively using his or her hands. Pinprick and temperature sensations are normal, and motor power is spared; some patients recover. In addition to sensory neuropathy, oxaliplatin can cause the acute onset of reversible symptoms consisting of paresthesias, cold hypersensitivity, jaw and eye pain, ptosis, leg cramps, and vision and voice changes. Carboplatinum rarely causes neurotoxicity. Platinum-induced neurotoxicity may be somewhat ameliorated by administration of amifostine.[102]

Doses of paclitaxel greater than 250 mg/m^2 cause a predominantly sensory neuropathy, characterized by paresthesias in the

Table 8 Radiation Injury to the Nervous System

Time after Radiation Therapy	*Organ Affected*	*Clinical Findings*
Primary injury		
Acute (minutes to hours)	Brain	Acute encephalopathy
Early delayed (6–16 wk)	Brain	Somnolence, focal signs
	Spinal cord	Lhermitte sign
Late delayed (months to years)	Brain	Dementia, focal signs
	Spinal cord	Transverse myelopathy
	Peripheral nerves	Paralysis, sensory loss
Secondary injury (years)	CNS, peripheral nervous system	Brain, cranial, or peripheral nerve sheath tumors
	Arteries (atherosclerosis)	Cerebral infarction
	Endocrine organs	Hypothyroidism, hypopituitarism

hands and feet. It may be accompanied by a sensory ataxia. Proximal muscle weakness may complicate the disorder. The neuropathy may resolve even with continuation of therapy and is rarely a disabling problem. Docetaxel is less neurotoxic.

Standard dosages of corticosteroids (prednisone or its equivalent at 60 to 100 mg/day) cause myopathy characterized by weakness of neck flexors and of the proximal extremities. The first sign of this neurotoxicity is difficulty in rising from low chairs or the toilet seat without pushing down with the arms; difficulty in climbing stairs subsequently occurs. Respiratory muscles may be affected.[104] The myopathy usually resolves when the drug is discontinued.

Encephalopathy

Many chemotherapeutic agents cause encephalopathy [*see Table 7*].[102,103] High-dose cyclophosphamide is sometimes associated with acute delirium; reversible acute encephalopathy is more common with ifosfamide. High-dose cytarabine may cause acute delirium or acute cerebellar degeneration, both of which are usually reversible. High-dose intravenous methotrexate or standard doses of intrathecal drugs can cause reversible acute encephalopathy. However, prolonged use of methotrexate, particularly in conjunction with radiation therapy, can lead to chronic encephalopathy that is characterized by progressive dementia.

NEUROTOXICITY CAUSED BY RADIATION THERAPY

The neurotoxic effects of radiation therapy may involve any portion of the CNS or peripheral nervous system and may occur acutely or be delayed for weeks to years after irradiation[105] [*see Table 8*]. The extent of neurotoxicity is determined by the total dose of radiation, the size of each fraction, the total time over which the dose is received, the volume of nervous system tissue that is irradiated, and the time that has elapsed since treatment. Other factors, such as underlying nervous system disease (e.g., a brain tumor or cerebral edema), previous surgery, concomitant use of chemotherapeutic agents, and individual susceptibility (especially to methotrexate), make it impossible to determine precisely the safe dose for a given person. However, guidelines exist that allow the radiation oncologist to calculate doses that are generally safe.

Acute Encephalopathy

Acute encephalopathy occurs in patients who have increased intracranial pressure, particularly in those who are given large fractional doses of radiation without corticosteroid prophylaxis. It is believed to occur when radiation-induced alteration of the blood-brain barrier causes brain edema that increases intracranial pressure. Acute worsening of neurologic symptoms does not occur after spinal cord irradiation.

Immediately after the first treatment, susceptible patients experience headache, nausea and vomiting, somnolence, fever, and, occasionally, worsening of neurologic signs. In rare instances, cerebral herniation and death occur. Acute encephalopathy responds to corticosteroids.

Early Delayed Encephalopathy or Myelopathy

Early delayed encephalopathy or myelopathy appears 4 to 16 weeks after therapy and persists for days to weeks. These early delayed irradiation syndromes are believed to result from demyelination, which may be caused by radiation-induced damage to oligodendroglia.

In children, early delayed encephalopathy, also called the postirradiation somnolence syndrome, commonly occurs after prophylactic irradiation of the brain for leukemia. The disorder is characterized by somnolence that is often associated with headache, nausea, and vomiting and is sometimes associated with fever. Electroencephalography may show slow waves, but focal signs are absent.

In adults, the syndrome usually occurs after cranial irradiation for brain tumors and is characterized by lethargy and worsening of focal neurologic signs. MRIs may also indicate transient worsening, suggesting that the tumor is not responding. In both adults and children, the disorder usually responds to corticosteroids; if it goes untreated, it usually resolves spontaneously.

Early delayed myelopathy occurs after radiation therapy to the neck or upper thorax. It is characterized by the Lhermitte sign, an electric shock–like sensation that radiates into various parts of the body when the neck is flexed. The symptoms resolve spontaneously.

Late Delayed Radiation Injury

Late delayed radiation injury appears months to years after radiation therapy and may affect any part of the nervous system. In the brain, two clinical syndromes occur: diffuse injury and focal radiation damage. Diffuse injury may be caused by whole brain irradiation, either in patients who do not have brain tumors (i.e., in those who receive prophylactic irradiation for small cell lung cancer) or in some patients with primary or metastatic brain tumors. The disorder is characterized by dementia, sometimes with gait apraxia and incontinence. MRI shows cerebral atrophy or hydrocephalus; pathologic changes are nonspecific. Some patients respond temporarily to ventricular shunting.[106]

Focal radiation damage affects patients who receive either focal brain irradiation for the treatment of extracranial neoplasms or whole brain irradiation for the treatment of intracranial neoplasms. Neurologic signs suggest that a mass is present and include headache, focal or generalized seizures, and hemiparesis. MRI reveals a hypointense mass that sometimes exhibits contrast enhancement. Neuropathologic features are coagulative necrosis in the white matter, vascular abnormalities (i.e., telangiectasia, fibrinoid necrosis, and thrombus formation), and glial proliferation with bizarre, multinucleated astrocytes. The clinical and MRI findings in patients with focal radiation damage cannot be distinguished from those in patients with brain tumors; diagnosis can be made only by biopsy. Corticosteroids may ameliorate symptoms. If they fail, treatment consists of surgical removal of the mass.

Late Delayed Myelopathy or Neuropathy

Late delayed myelopathy often begins with the Brown-Séquard syndrome (weakness and loss of proprioception in the extremities of one side and loss of pain and temperature sensations on the other) and is subsequently characterized by progressive paralysis, sensory changes, and sometimes pain. Occasionally, patients transiently respond to corticosteroids, and the disorder may stop progressing. Generally, however, paraplegia or quadriplegia develops. Pathologic changes include necrosis of the spinal cord.

Late delayed neuropathy may affect any cranial or peripheral nerve. Common disorders are blindness, caused by optic neuropathy, and paralysis of an upper extremity, caused by brachial plexopathy after therapy for lung or breast cancer. The pathogenetic mechanisms are probably fibrosis and ischemia of the plexus. There is no therapy for patients with these disorders.

Radiation-Induced Tumors

Meningiomas,[107] sarcomas, or, less commonly, gliomas may appear years to decades after cranial irradiation and may be associated with even low-dose radiation therapy. Malignant or atypical nerve sheath tumors may develop after irradiation of the brachial, cervical, or lumbar plexus. The CNS may also be damaged when radiation alters extraneural structures.

Other CNS Syndromes

Radiation therapy accelerates atherosclerosis, and cerebral infarction associated with carotid artery occlusion may occur many years after neck irradiation. Endocrinologic dysfunction (involving the pituitary, thyroid, or parathyroid) may be caused by radiation therapy and may be associated with neurologic signs. Hypothyroidism that results from irradiation may also cause encephalopathy.

Additional Information

The National Cancer Institute of the National Institutes of Health provides continually updated information on CNS neoplasms on their Web site CancerNet (http://cancernet.nci.nih.gov). CancerNet provides information on clinical trials; summaries on cancer treatment, screening, prevention, and supportive care; and publications of the National Cancer Institute.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. DeAngelis LM, Gutin PH, Leibel SA, et al: Intracranial Tumors: Diagnosis and Treatment, Taylor & Francis Books, Ltd., London, 2002

2. Posner JB: Neurologic Complications of Cancer. FA Davis Co, Philadelphia, 1995

3. Schiff D, Wen PY: Cancer Neurology in Clinical Practice. Humana Press, Inc., Totowa, New Jersey, 2002

4. Darnell RB, Posner JB: Paraneoplastic syndromes involving the nervous system. N Engl J Med 349:1543, 2003

5. Kleihues P, Cavenee WK: Pathology and Genetics of Tumours of the Nervous System. IARC Press, Lyon, 2000

6. Jemal A, Tiwari RC, Murray T, et al: Cancer statistics, 2004. CA Cancer J Clin 54:8, 2004

7. Ezzat S, Asa SL, Couldwell WT, et al: The prevalence of pituitary adenomas. Cancer 101:613, 2004

8. Lönn S, Klaeboe L, Hall P, et al: Incidence trends of adult primary intracerebral tumors in four Nordic countries. Int J Cancer 108:450, 2004

9. Olson JE, Janney CA, Rao RD, et al: The continuing increase in the incidence of primary central nervous system non-Hodgkin lymphoma: a surveillance, epidemiology, and end results analysis. Cancer 95:1504, 2002

10. DeAngelis LM: Primary central nervous system lymphoma: a curable brain tumor. J Clin Oncol 21:4471, 2003

11. Wrensch M, Minn Y, Chew T, et al: Epidemiology of primary brain tumors: current concepts and review of the literature. Neuro-oncol 4:278, 2002

12. Relling MV, Rubnitz JE, Rivera GK, et al: High incidence of secondary brain tumours after radiotherapy and antimetabolites. Lancet 354:34, 1999

13. Wrensch M, Miike R, Lee M, et al: Are prior head injuries or diagnostic x-rays associated with glioma in adults? The effects of control selection bias. Neuroepidemiology 19:234, 2000

14. Henry PT, Rajshekhar V: Post-traumatic malignant glioma: case report and review of the literature. Br J Neurosurg 14:64, 2000

15. Blumenthal DT, Raizer JJ, Rosenblum MK, et al: Primary intracranial neoplasms in patients with HIV. Neurology 52:1648, 1999

16. Vannemreddy PSSV, Fowler M, Polin RS, et al: Glioblastoma multiforme in a case of acquired immunodeficiency syndrome: investigating a possible oncogenic influence of human immunodeficiency virus on glial cells: case report and review of the literature. J Neurosurg 92:161, 2000

17. Muscat JE, Malkin MG, Thompson S, et al: Handheld cellular telephone use and risk of brain cancer. JAMA 284:3001, 2000

18. Rothman KJ: Epidemiological evidence on health risks of cellular telephones. Lancet 356:1837, 2000

19. Hardell L, Mild KH, Carlberg M: Further aspects on cellular and cordless telephones and brain tumours. Int J Oncol 22:399, 2003

20. Grossman SA, Osman M, Hruban R, et al: Central nervous system cancers in first-degree relatives and spouses. Cancer Invest 17:299, 1999

21. Hemminki K, Li X: Familial risks in nervous system tumors. Cancer Epidemiol Biomarkers Prev 12:1137, 2003

22. Collins VP: Brain tumours: classification and genes. J Neurol Neurosurg Psychiatry 75(suppl 2):ii2, 2004

23. Lichtenstein P, Holm NV, Verkasalo PK, et al: Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78, 2000

24. Weller RO, Engelhardt B, Phillips MJ: Lymphocyte targeting of the central nervous system: a review of afferent and efferent CNS-immune pathways. Brain Pathol 6:275, 1996

25. Puduvalli VK: Inhibition of angiogenesis as a therapeutic strategy against brain tumors. Cancer Treat Res 117:307, 2004

26. Purdy RA, Kirby S: Headaches and brain tumors. Neurol Clin 22:39, 2004

27. Watling CJ, Cairncross JG: Acetazolamide therapy for symptomatic plateau waves in patients with brain tumors: report of three cases. J Neurosurg 97:224, 2002

28. Holt-Seitz A, Wirrell EC, Sundaram MB: Seizures in the elderly: etiology and prognosis. Can J Neurol Sci 26:110, 1999

29. Landy HJ, Lee TT, Potter P, et al: Early MRI findings in high grade glioma. J Neurooncol 47:65, 2000

30. Woesler B, Kuwert T, Morgenroth C, et al: Non-invasive grading of primary brain tumours: Results of a comparative study between SPET with ^{123}I-α-methyl tyrosine and PET with ^{18}F-deoxyglucose. Eur J Nucl Med 24:428, 1997

31. Nelson SJ, Cha S: Imaging glioblastoma multiforme. Cancer J 9:134, 2003

32. NCCN adult brain tumor practice guidelines. Oncology (Huntingt) 11:237, 1997

33. McKinney PA: Brain tumours: incidence, survival, and aetiology. J Neurol Neurosurg Psychiatry 75(suppl 2):ii12, 2004

34. Prados MD, Levin V: Biology and treatment of malignant glioma. Semin Oncol 27:1, 2000

35. Zimmermann M, Seifert V, Trantakis C, et al: Open MRI-guided microsurgery of intracranial tumours: preliminary experience using a vertical open MRI-scanner 22. Acta Neurochir (Wien) 142:177, 2000

36. Duffau H, Capelle L, Lopes M, et al: The insular lobe: physiopathological and surgical considerations. Neurosurgery 47:801, 2000

37. Duffau H, Khalil I, Gatignol P, et al: Surgical removal of corpus callosum infiltrated by low-grade glioma: functional outcome and oncological considerations. J Neurosurg 100:431, 2004

38. Nakamura M, Konishi N, Tsunoda S, et al: Analysis of prognostic and survival factors related to treatment of low-grade astrocytomas in adults. Oncology 58:108, 2000

39. Figarella-Branger D, Civatte M, Bouvier-Labit C, et al: Prognostic factors in intracranial ependymomas in children. J Neurosurg 93:605, 2000

40. Dropcho EJ: Low-grade gliomas in adults. Curr Treat Options Neurol 6:265, 2004

41. The BS, Mai WY, Grant WH, et al: Intensity modulated radiotherapy (IMRT) decreases treatment-related morbidity and potentially enhances tumor control. Cancer Investigation 20:437, 2002

42. Souhami L, Seiferheld W, Brachman D, et al: Randomized comparison of stereotactic radiosurgery followed by conventional radiotherapy with carmustine to conventional radiotherapy with carmustine for patients with glioblastoma multiforme: report of Radiation Therapy Oncology Group 93-05 protocol. Int J Radiat Oncol Biol Phys 60:853, 2004

43. Galanis E, Buckner JC: Chemotherapy of brain tumors. Curr Opin Neurol 13:619, 2001

44. Stupp R, Mason WP, van den Bent MJ, et al: Concomitant and adjuvant temozolomide (TMZ) and radiotherapy (RT) for newly diagnosed glioblastoma multiforme: conclusive results of a randomized phase III trial by the EORTC Brain and RT Groups and YCIC Clinical Trials Group. J Clin Oncol ASCO Annual Meeting Proceedings 22(14S):2, 2004

45. Brandes AA: State-of-the-art treatment of high-grade brain tumors. Semin Oncol 30:4, 2003

46. Lesniak MS, Brem H: Targeted therapy for brain tumours. Nat Rev Drug Discov 3:499, 2004

47. Buckner JC: Factors influencing survival in high-grade gliomas. Semin Oncol 30:10, 2003

48. Krampla W, Newrkla S, Pfisterer W, et al: Frequency and risk factors for meningioma in clinically healthy 75-year-old patients: results of the Transdanube Ageing Study (VITA). Cancer 100:1208, 2004

49. Huang H, Okamoto Y, Yokoo H, et al: Gene expression profiling and subgroup identification of oligodendrogliomas. Oncogene 23:6012, 2004

50. Duffau H, Capelle L: Preferential brain locations of low-grade gliomas. Cancer 100:2622, 2004

51. Kortmann RD, Jeremic B, Weller M, et al: Immediate postoperative radiotherapy or "watch and wait" in the management of adult low-grade glioma? Strahlenther Onkol 180:408, 2004

52. Mandonnet E, Delattre JY, Tanguy ML, et al: Continuous growth of mean tumor diameter in a subset of grade II gliomas. Ann Neurol 53:524, 2003

53. Soffietti R: Chemotherapy of anaplastic oligodendroglial tumours. Expert Opin Pharmacother 5:295, 2004

54. Carroll RS, Zhang J, Black PM: Expression of estrogen receptors alpha and beta in human meningiomas. J Neurooncol 42:109, 1999

55. Cerdá-Nicolás M, López-Gines C, Pérez-Bacete M, et al: Histopathological and cytogenetic findings in benign, atypical and anaplastic human meningiomas: a study of 60 tumors. J Neurooncol 47:99, 2000

56. Nakamura M, Roser F, Michel J, et al: The natural history of incidental meningiomas. Neurosurgery 53:62, 2003

57. D'Ambrosio AL, Bruce JN: Treatment of meningioma: an update. Curr Neurol Neurosci Rep 3:206, 2003

58. Walsh RM, Bath AP, Bance ML, et al: The role of conservative management of vestibular schwannomas. Clin Otolaryngol 25:28, 2000

59. Pogodzinski MS, Harner SG, Link MJ: Patient choice in treatment of vestibular schwannoma. Otolaryngol Head Neck Surg 130:611, 2004

60. Levo H, Pyykkö I, Blomstedt G: Postoperative headache after surgery for vestibular schwannoma. Ann Otol Rhinol Laryngol 109:853, 2000

61. Khrais TH, Falcioni M, Taibah A, et al: Cerebrospinal fluid leak prevention after translabyrinthine removal of vestibular schwannoma. Laryngoscope 114:1015, 2004

62. Levy MJ, Jäger R, Powell M, et al: Pituitary volume and headache: size is not everything. Arch Neurol 61:721, 2004

63. Kurosaki M, Lüdecke DK, Flitsch J, et al: Surgical treatment of clinically nonsecreting pituitary adenomas in elderly patients. Neurosurgery 47:843, 2000

64. Kokubo M, Sasai K, Shibamoto Y, et al: Long-term results of radiation therapy for pituitary adenoma. J Neurooncol 47:79, 2000

65. Marcou Y, Plowman PN: Stereotactic radiosurgery for pituitary adenomas. Trends Endocrinol Metab 11:132, 2000

66. Camacho MP, Mazzone MT: Thyrotropin-secreting pituitary adenoma responsive to bromocriptine therapy. Endocr Pract 5:257, 1999

67. Colao A, Di Sarno A, Landi ML, et al: Long-term and low-dose treatment with cabergoline induces macroprolactinoma shrinkage. J Clin Endocrinol Metab 82:3574, 1997

68. Muller AF, van der Lely AJ: Pharmacological therapy for acromegaly: a critical review. Drugs 64:1817, 2004

69. Fauchon F, Jouvet A, Paquis P, et al: Parenchymal pineal tumors: a clinicopathological study of 76 cases. Int J Radiat Oncol Biol Phys 46:959, 2000

70. Jackson AS, Plowman PN: Pineal parenchymal tumours: I. Pineocytoma: a tumour responsive to platinum-based chemotherapy. Clin Oncol (R Coll Radiol) 16:238, 2004

71. Plowman PN, Pizer B, Kingston JE: Pineal parenchymal tumours: II. On the aggressive behaviour of pineoblastoma in patients with an inherited mutation of the RB1 gene. Clin Oncol (R Coll Radiol) 16:244, 2004

72. Deshmukh VR, Smith KA, Rekate HL, et al: Diagnosis and management of pineocytomas. Neurosurgery 55:349, 2004

73. Bamberg M, Kortmann R, Calaminus G, et al: Radiation therapy for intracranial geminoma: results of the German Cooperative Prospective Trials MAKEI 83/86/89. J Clin Oncol 17:2585, 1999

74. Balmaceda C, Finlay J: Current advances in the diagnosis and management of intracranial germ cell tumors. Curr Neurol Neurosci Rep 4:253, 2004

75. Smith AA, Weng E, Handler M, et al: Intracranial germ cell tumors: a single institution experience and review of the literature. J Neurooncol 68:153, 2004

76. Knappe UJ, Bentele K, Horstmann M, et al: Treatment and long-term outcome of pineal nongerminomatous germ cell tumors. Pediatr Neurosurg 28:241, 1998

77. Alter A, Duddy M, Hebert S, et al: Determinants of human B cell migration across brain endothelial cells. J Immunol 170:4497, 2003

78. Gavrilovic IT, Abrey LE: Primary central nervous system lymphoma. Curr Oncol Rep 6:388, 2004

79. Gilbertson RJ: Medulloblastoma: signalling a change in treatment. Lancet Oncol 5:209, 2004

80. Perez-Martinez A, Quintero V, Vicent MG, et al: High-dose chemotherapy with autologous stem cell rescue as first line of treatment in young children with medulloblastoma and supratentorial primitive neuroectodermal tumors. J Neurooncol 67:101, 2004

81. Mulhern RK, Merchant TE, Gajjar A, et al: Late neurocognitive sequelae in survivors of brain tumours in childhood. Lancet Oncol 5:399, 2004

82. Van Goethem JW, van den Hauwe L, Ozsarlak O, et al: Spinal tumors. Eur J Radiol 50:159, 2004

83. Hanbali F, Fourney DR, Marmor E, et al: Spinal cord ependymoma: radical surgical resection and outcome. Neurosurgery 51:1162, 2002

84. Gilhuis HJ, Kappelle AC, Beute G, et al: Radiotherapy for partially resected spinal ependymomas: a retrospective study of 60 cases. Oncol Rep 10:2079, 2003

85. Nishio S, Morioka T, Fujii K, et al: Spinal cord gliomas: management and outcome with reference to adjuvant therapy. J Clin Neurosci 7:20, 2000

86. Johnson JD, Young B: Demographics of brain metastasis. Neurosurg Clin N Am 7:337, 1996

87. Schiff D, O'Neill BP, Suman VJ: Spinal epidural metastasis as the initial manifestation of malignancy: clinical features and diagnostic approach. Neurology 49:452, 1997

88. Patchell RA, Tibbs PA, Walsh JW: A randomized trial of surgery in the treatment of single metastases to the brain. N Engl J Med 322:494, 1990

89. Yoshida S, Morii K: The role of surgery in the treatment of brain metastasis: a retrospective review. Acta Neurochir (Wien) 146:767, 2004

90. Patchell RA, Tibbs PA, Regine WF, et al: Postoperative radiotherapy in the treatment of single metastases to the brain: a randomized trial. JAMA 280:1485, 1998

91. Vecht CJ, Haaxma-Reiche H, Noordijk EM, et al: Treatment of single brain metastasis: radiotherapy alone or combined with neurosurgery? Ann Neurol 33:583, 1993

92. Sheehan J, Niranjan A, Flickinger JC, et al: The expanding role of neurosurgeons in the management of brain metastases. Surg Neurol 62:32, 2004

93. Omuro AM, Abrey LE: Brain metastases. Curr Neurol Neurosci Rep 4:205, 2004

94. Cortes J, Rodriguez J, Aramendia JM, et al: Front-line paclitaxel/cisplatin-based chemotherapy in brain metastases from non–small-cell lung cancer. Oncology (Basel) 64:28, 2003

95. Byrne TN: Metastatic epidural cord compression. Curr Neurol Neurosci Rep 4:191, 2004

96. Patchell RA: A randomized trial of direct decompressive surgical resection in the treatment of spinal cord compression caused by metastasis. Plenary presentation, ASCO Annual Meeting, Chicago, May 31–June 3, 2003

97. Maranzano E, Trippa F, Chirico L, et al: Management of metastatic spinal cord compression. Tumori 89:469, 2003

98. Bilsky MH, Lis E, Raizer J, et al: The diagnosis and treatment of metastatic spinal tumor. Oncologist 4:459, 1999

99. Demopoulos A: Leptomeningeal metastases. Curr Neurol Neurosci Rep 4:196, 2004

100. Chamberlain MC: Neoplastic meningitis: a guide to diagnosis and treatment. Curr Opin Neurol 13:641, 2001

101. Bataller L, Dalmau J: Paraneoplastic neurologic syndromes. Neurol Clin 21:221, 2003

102. Verstappen CC, Heimans JJ, Hoekman K, et al: Neurotoxic complications of chemotherapy in patients with cancer: clinical signs and optimal management. Drugs 63:1549, 2003

103. Plotkin SR, Wen PY: Neurologic complications of cancer therapy. Neurol Clin 21:279, 2003

104. Batchelor TT, Taylor LP, Thaler HT, et al: Steroid myopathy in cancer patients. Neurology 48:1234, 1997

105. Cross NE, Glantz MJ: Neurologic complications of radiation therapy. Neurol Clin 21:249, 2003

106. Thiessen B, DeAngelis LM: Hydrocephalus in radiation leukoencephalopathy: results of ventriculoperitoneal shunting. Arch Neurol 55:705, 1998

107. Strojan P, Popovic M, Jereb B: Secondary intracranial meningiomas after high-dose cranial irradiation: report of five cases and review of the literature. Int J Radiat Oncol Biol Phys 48:65, 2000

VII ANOXIC, METABOLIC, AND TOXIC ENCEPHALOPATHIES

MICHAEL J. AMINOFF, M.D., D.SC.

The term encephalopathy is generally used to designate diffuse cerebral dysfunction. Such dysfunction is typically manifested by alterations in cortical function and disturbances of consciousness, ranging from mild confusional states (i.e., obtundation) to coma.[1] Abnormalities of consciousness reflect dysfunction of both cerebral hemispheres or of the reticular activating system in the brain stem. Encephalopathies may also be characterized by focal deficits, reflecting more localized cerebral dysfunction. In general, however, the cause is usually a systemic disorder that affects the brain diffusely, though some regions are more severely affected than others. A variety of mechanisms may contribute to encephalopathies, but anoxic, metabolic, and toxic factors are often significant and can lead to secondary structural abnormalities of the brain. Because of the risk of brain damage or death, diagnostic evaluation in patients with an encephalopathy of uncertain cause should proceed concurrently with stabilization and, in patients with acute coma of unknown cause, empirical treatment for common precipitating factors [*see Table 1*].

Clinical Evaluation

Metabolic and toxic encephalopathies cannot be distinguished with confidence from those caused by a mass lesion, but certain general points can be made. Onset is often insidious except when an acute event, such as cardiac arrest or drug overdose, is responsible. In general, the neurologic findings are symmetrical or multifocal in distribution, and tremor, asterixis, and myoclonus are common. Asterixis (sometimes referred to as a flapping tremor) consists of brief lapses of a sustained muscular contraction, as when the arms are held outstretched against gravity. Focal or lateralizing signs are absent or inconsistent; when present, they sometimes alternate from one side to the other. With some exceptions, preserved pupillary responses in the context of impaired brain stem function are strongly suggestive of metabolic or toxic disorders.

HISTORY

Evaluation requires an accurate history to determine the cause of the encephalopathy and the prognosis for recovery. It is important to determine whether the neurologic symptoms came on abruptly (as with vascular pathology) or gradually, whether the symptoms have progressed since their onset, and whether they were preceded by other symptoms and signs that may suggest the cause of the dysfunction. Generalized seizures occur in drug and alcohol withdrawal states, with various other toxic and metabolic encephalopathies, and in the presence of structural lesions of the brain. Partial seizures are more suggestive of focal pathology but may occur in certain metabolic disorders, especially when the disorder is superimposed on a preexisting focal structural lesion, such as an old stroke. Similarly, the past medical history should be reviewed in detail. A history or clinical features indicative of diabetes mellitus suggest that the obtundation is associated with iatrogenic hypoglycemia or a nonketotic hyperosmolar state, whereas the cause of obtundation in an alcoholic patient may be a metabolic disorder (e.g., hepatic disease or thiamine deficiency), toxicity (e.g., ethanol intoxication or withdrawal), an infection, or trauma.

PHYSICAL EXAMINATION

The general physical examination of encephalopathic patients is important. Jaundice, petechial hemorrhages, gastrointestinal bleeding, ascites, or hypothermia may indicate hepatic dysfunction. A coarse facies, dry hair, or bradycardia suggests hypothyroidism. Acne, obesity, and hypertension are common in Cushing syndrome. Needle tracks in the skin raise the possibility of a toxic encephalopathy. Hypertension suggests that the encephalopathy is caused by a metabolic disorder (e.g., a renal or endocrinologic disorder) or an ischemic disorder (e.g., a cerebrovascular or cardiovascular condition), and hypothermia suggests a metabolic or toxic cause. Individual signs may be misleading, however, and must be evaluated within their clinical context; for example, fever and tachycardia are common signs of infection but also occur in drug and alcohol withdrawal states.

NEUROLOGIC EXAMINATION

The neurologic examination should characterize the nature and severity of the encephalopathy and should exclude a primary disorder of the central nervous system. An encephalopathy associated with signs of meningeal irritation suggests meningitis or subarachnoid hemorrhage, whereas a focal neurologic deficit or evidence of increased intracranial pressure mandates exclusion of an intracranial mass lesion. In metabolic or toxic en-

Table 1 Immediate Management of Patients with an Encephalopathy of Uncertain Cause

- Maintain adequacy of respiration and circulation
- Obtain blood samples for determination of the following:
  - Serum glucose and electrolytes
  - Complete blood count and sedimentation rate
  - Liver and kidney function studies
  - Toxicity screen
- Obtain urine for toxicity screen
- For coma of acute onset and unknown cause, administer the following:
  - Dextrose, 25 g I.V. (to treat possible hypoglycemia)
  - Thiamine, 100 mg I.V. (to prevent or treat Wernicke encephalopathy)
  - Naloxone, 1 mg I.V. (to treat possible opiate overdose)
- General clinical and neurologic examination
- Computed tomographic scanning of the head (if focal intracranial lesion is suspected)
- Lumbar puncture (if meningitis or subarachnoid hemorrhage is suspected)
- Arterial blood gas determinations (to distinguish between different causes of metabolic encephalopathy)
- Chest radiography
- Further investigation and treatment, depending on results of initial studies

cephalopathies, focal or lateralizing neurologic signs are often absent, but their presence does not exclude such disorders.

In determining the nature and severity of an encephalopathy, the mental status is evaluated with particular regard to the level of consciousness as judged by the attention span and response to verbal or painful stimuli [*see Table 2*]. Orientation, behavior, language function, mood and affect, thought content, and memory should also be assessed, if possible. Brief examination of the CNS requires evaluation of the cranial nerves, especially the pupillary responses, and of sensorimotor functions in the limbs, including tendon reflexes and plantar responses.

Although pupillary responses to light are often normal in persons with metabolic and toxic encephalopathies, a variety of pupillary abnormalities may occur. For example, fixed dilated or poorly responsive pupils often result from acute cerebral anoxia or intoxication with anticholinergic or sympathomimetic agents; depending on the circumstances in which they are encountered, such pupillary responses should also raise concern about a herniating intracranial mass lesion. Pinpoint pupils are a feature of opioid toxicity, organophosphate poisoning, or the use of miotic eyedrops; they are also a common sequela of pontine damage, as from a stroke. Abnormal asymmetry of pupil size or responsiveness suggests that a structural brain stem (or cranial nerve) lesion is responsible; such symptoms are unlikely in metabolic and toxic encephalopathies. Reflex ocular movements should also be assessed. In comatose patients, loss of oculovestibular responses may occur with either a structural pontine lesion or sedative intoxication. By contrast, downward deviation of one or both eyes with unilateral cold-water stimulation strongly suggests sedative intoxication.

As an encephalopathy becomes progressively more severe, patients become comatose. The depth of coma is best characterized by the response to external stimulation [*see Table 2*]. Lateralized responses suggest a structural lesion, whereas bilaterally symmetrical responses occur with either structural or metabolic-toxic pathology. In cases of expanding or progressive structural lesions causing downward transtentorial herniation, loss of cortical function may occur in a rostrocaudal sequence [*see Table 2*].

Table 2 Evaluation of Level of Consciousness

Level of Consciousness	*Characteristics*
Confusional state	Patient is disoriented, is irritable, and has a poor attention span Hallucinations and delusions may occur
Stupor	Patient is inattentive, drowsy, and unresponsive but can be aroused by vigorous stimuli for short periods
Coma	Patient is unresponsive and unarousable In patients with downward transtentorial herniation, the level of dysfunction is further characterized (below)
Diencephalic level	Reactive pupils Preserved oculocephalic responses
Early	Purposive response to pain
Late	Decorticate response to pain
Midbrain level	Fixed and midsized pupils Abnormal oculocephalic responses Decerebrate response to pain
Lower brain stem level	Fixed and midsized pupils Abnormal oculocephalic responses No response of upper limbs to pain

Anoxic Encephalopathies

CARDIAC DISORDERS

Circulatory Arrest

Transient circulatory arrest may lead to global cerebral ischemia and thus to syncope, which is sometimes preceded by nonspecific premonitory symptoms such as paresthesias, lightheadedness, palpitations, and a graying-out of vision. Syncope is associated with pallor and loss of muscle tone, but prolonged ischemia results in tonic posturing, sometimes accompanied by irregular jerking movements that resemble seizures. If postictal confusion occurs, it clears within 1 minute. In elderly patients, syncope may present simply as unexplained falls. Syncope may be related to cardiac pathology, dysautonomia, postural hypotension, endocrinopathies, and metabolic disorders. Neurocardiogenic (vasovagal) syncope, however, is the most common variety [*see 1:1 Approach to the Cardiovascular Patient*].

Depending on its duration, ventricular fibrillation or asystole may cause irreversible anoxic-ischemic brain damage. The prognosis varies with the patient's age, the duration of circulatory arrest, and the interval before cardiopulmonary resuscitation and defibrillating procedures were undertaken. Circulatory arrest from ventricular fibrillation has a better prognosis than that from asystole. The neurologic consequences of the arrest may relate to the accumulation of intracellular calcium, increased extracellular concentrations of glutamate and aspartate, and increased levels of free radicals.

In the mature nervous system, gray matter is generally more vulnerable to ischemia than white matter, and the cerebral cortex is more sensitive than the brain stem. So-called watershed areas bordering the zones supplied by major arteries are especially vulnerable.

Circulatory arrest of less than 5 minutes' duration leads to transient confusion or temporary loss of consciousness and impaired cognitive function. Complete recovery is usual, but in rare instances, the circulatory arrest is followed after 7 to 10 days by a demyelinating encephalopathy, with increasing cognitive dysfunction and pyramidal or extrapyramidal deficits that may have a fatal outcome. In such cases, patients regain consciousness several hours after the circulatory arrest but then develop progressive neurologic deficits, such as intellectual deterioration; personality changes; seizures; cortical blindness; amnestic syndromes; or, less commonly, the locked-in syndrome (characterized by quadriplegia and mutism), extrapyramidal syndromes, bibrachial paresis, or intention (action) myoclonus. Spinal cord dysfunction may occur but is unusual.

Circulatory arrest of longer than 5 minutes' duration may cause widespread and irreversible brain damage, resulting in prolonged coma. Prognosis for survival or useful recovery is poor, especially when brain stem reflexes (most notably the pupillary responses to light) are lost. In particular, loss of pupillary reactivity for more than 24 hours or persistence of coma for more than 4 days indicates a poor prognosis. In one study, comatose survivors of cardiac arrest who continued to have nonreactive pupils, failed to open their eyes in response to pain, or had absent or reflex motor responses 3 days after onset of coma generally failed to survive or to regain useful independent function [*see Table 3*].[2] In this study, the most accurate sin-

gle predictor of poor outcome immediately after restoration of spontaneous circulation was the absence of pupillary response to light: Outcome was poor (i.e., either death or a persistent vegetative state) in 73 of 89 patients (82%) with absent pupillary responses.[2]

In the first several days after the onset of ischemic encephalopathy, diffusion-weighted magnetic resonance imaging may be useful in determining prognosis; hyperintense cortical lesions are associated with a poor outcome.[3,4]

Evoked potential studies can also have prognostic utility. In a systematic review of somatosensory evoked potentials performed early after onset of coma, rates of awakening in adults with hypoxic-ischemic encephalopathy ranged from 0% when somatosensory evoked potentials were absent to 52% when somatosensory evoked potentials were normal.[5] In a prospective study of 346 comatose patients, pupillary reflex was the strongest prognostic variable for awakening. The probability of awakening was higher when late auditory evoked potentials were also present, and it was higher still when middle-latency evoked potentials were present. No patient with cognitive potentials became permanently vegetative.[6]

Even if consciousness is regained, focal or multifocal neurologic signs may lead to significant disability from focal motor deficits, extrapyramidal disturbances (e.g., parkinsonism), sensory loss, seizures, myoclonus, and disturbances of higher cortical function from which recovery is usually delayed and incomplete. Intention (action) myoclonus is particularly characteristic in such circumstances; it is often activated by startle or various sensory stimuli and is responsive only occasionally to clonazepam, valproate, piracetam, or 5-hydroxytryptophan. The last two medications are not commercially available in the United States.

Some patients do not fully regain consciousness after circulatory arrest. These patients may be in one of three states: the minimally conscious state, the persistent vegetative state, or brain death.[7,8]

The minimally conscious state, which is characterized by inconsistent but clearly discernible behavior of consciousness, can be distinguished from coma and a vegetative state by the presence of behavioral conditions not found in either of those two conditions, including intermittent behavioral evidence of awareness of self or environment; although these patients are unable to follow simple instructions or communicate reliably, functional MRI studies suggest that some of them may retain widely distributed cortical systems with potential for cognitive and sensory function.[9] Identification of the minimally conscious state is important because these patients appear to have a better outcome than do patients in a persistent vegetative state. The minimally conscious state may be transient or permanent.[10]

The persistent vegetative state is characterized by the return of sleep-wake cycles and of various reflex activities, but wakefulness is without awareness.[11,12]

Brain death is defined as the loss of all cerebral activity, including activity of the cerebral cortex and brain stem, for at least 6 hours if confirmed by electroencephalographic evidence of electrocerebral inactivity or for 24 hours without a confirmatory electroencephalogram. A useful clinical test in patients with suspected brain death is the apnea test. This test involves evaluation of the respiratory response of the brain stem by allowing the carbon dioxide tension (P_{CO_2}) to rise to 60 mm Hg while 100% oxygen is given through the endotracheal tube. Brain-dead patients have no ventilatory response to the apnea test.

Brain death may be simulated clinically by extreme hypothermia, sedative overdose, and neuromuscular blockade. Such conditions must always be excluded, especially when no clear history of circulatory arrest can be obtained.

Table 3 Clinical Evaluation of Prognosis in Comatose Survivors of Cardiac Arrest

Sign	*Patients with Poor Outcome (%)*		
	Immediate	*Day 3*	*Day 7*
Lack of response to pain			
No opening of the eyes	69	100	100
No motor response	75	100	100
Lack of response to verbal stimuli	67	94	100
Lack of pupillary response	83	100	100

Disorders Associated with Cardiac Procedures

Cardiac catheterization or percutaneous transluminal coronary angioplasty sometimes causes cerebral emboli that may lead to focal neurologic deficits or an encephalopathy manifested by a behavioral disturbance. Encephalopathy, seizures, and cerebral infarction after cardiac surgery usually result from hypoxia or emboli. Postoperative encephalopathies may also be caused by metabolic disturbances, medication, infection (especially in immunosuppressed patients), or multiple organ dysfunction syndrome (MODS). Postoperative seizures may result from focal or generalized cerebral ischemia, electrolyte or metabolic disturbances, or MODS. Recognition of the precise cause of encephalopathy in such cases can be difficult. After cardiopulmonary bypass is performed, intracranial hemorrhage may result because of diminished platelet adhesiveness and reduced levels of coagulation factors.

Coronary angioplasty leads to cerebral emboli in approximately 1% of cases; however, when undertaken after acute myocardial infarction, it is associated with a higher risk of stroke and anoxic encephalopathy.[13]

An encephalopathy may occur soon after cardiac transplantation as a side effect of an immunosuppressive agent or as the result of an infection (e.g., meningitis, meningoencephalitis, or cerebral abscess) related to immunosuppressive therapy. Infecting organisms include *Aspergillus, Toxoplasma, Cryptococcus, Candida, Nocardia,* and viruses. In patients on long-term immunosuppressive agents, an encephalopathy may develop from a primary CNS lymphoma [*see 11: VI Neoplastic Disorders*].

The occurrence of an encephalopathy after coronary artery bypass surgery may be caused by stroke, which develops in about 2% to 4% of bypass patients[14]; it is either embolic or, less commonly, the result of watershed infarction from hypoperfusion. Risk factors include advanced age, proximal aortic atherosclerosis, hypertension, previous stroke or transient ischemic attack (TIA), and diabetes.[15] A carotid bruit or radiologic evidence of atherosclerosis of the carotid artery does not clearly increase the risk of stroke, and carotid endarterectomy before cardiac surgery is of questionable utility.[16]

In rare cases, patients do not recover consciousness after surgery, and no specific metabolic cause can be identified. This encephalopathy is probably the result of diffuse cerebral ischemia or hypoxia. Hemispheric or multifocal infarction is sometimes responsible.

RESPIRATORY DISEASES

Hypoxia and Hypercapnia

The pathogenesis of neurologic abnormalities related to hypoxia and hypercapnia is not fully understood, because hypoxia is often associated with acid-base imbalance and leads to hematologic and biochemical changes that affect cerebral function. Moreover, both hypercapnia and hypoxemia can result from impaired ventilation, and their neurologic sequelae are not easily differentiated.

Chronic pulmonary insufficiency leads to an encephalopathy characterized by headache, confusion, disorientation, and impaired cognitive function. Examination may also reveal a postural tremor, myoclonus, asterixis, and hyperreflexia; papilledema is sometimes present. These findings are not only the result of cerebral hypoxia but also the result of hypercapnia, which produces cerebral vasodilatation, increased cerebrospinal fluid pressure, and an altered pH of the CSF.

High-altitude sickness can lead to an encephalopathy characterized by headache, fatigue, anorexia, nausea, poor concentration, and sleep disturbances.[17] Symptoms of high-altitude sickness begin within hours or days of ascent to altitudes above 10,000 ft. In severe cases or at higher altitudes, consciousness is impaired and coma may occur—sometimes with a fatal outcome. Cerebral edema causes papilledema, retinal hemorrhages, cranial neuropathies, a variety of sensorimotor deficits, and behavioral disturbances. High-altitude cerebral edema arises because the low barometric pressure encountered at high altitudes causes a reduction in the partial pressure of oxygen (P_{O_2}); the condition often follows acute mountain sickness and may be an extreme form of that disorder. Acetazolamide (250 mg once or twice a day) or dexamethasone (2 mg every 6 to 8 hours) may prevent acute mountain sickness. High-altitude cerebral edema is treated with prompt descent to a lower altitude and administration of oxygen and dexamethasone (8 mg, then 4 mg every 6 hours).[17]

Hypocapnia

Hypocapnia, which results from hyperventilation, causes cerebral vasoconstriction, a decline in the peripheral availability of oxygen, and an altered ionic balance of calcium. The resultant encephalopathy leads to light-headedness, paresthesias, visual disturbances, headache, unsteadiness, tremor, nausea, palpitations, and loss of consciousness. Muscle cramps and carpopedal spasms also occur. The many causes of hyperventilation include hepatic coma, brain stem lesions, and certain cardiopulmonary diseases, but in many instances, no specific cause can be found.

SEPSIS

A diffuse encephalopathy with progressive obtundation may complicate sepsis, especially in patients with acute respiratory distress syndrome. The cause of sepsis-related encephalopathies is uncertain but may relate to cerebral edema, hypoxia, disruption of the blood-brain barrier, direct cerebral infection, toxins produced by organisms infecting other tissues, alterations in the cerebral microcirculation, metabolic disturbances, and the effects of medications.[18] Sepsis-related encephalopathy tends to be worse at night; it is associated with marked EEG abnormalities and often clears spontaneously. Overt infection should be treated vigorously; in addition, metabolic abnormalities should be corrected, and medication requirements should be reviewed.

LIVER DISEASE

Portosystemic Encephalopathy

Encephalopathy can result from chronic liver disease and sometimes precedes systemic features of hepatic dysfunction [*see 4:IX Cirrhosis of the Liver*]. It may be precipitated by GI hemorrhage, a high protein intake, use of certain sedatives and diuretics, or sepsis. Portosystemic encephalopathy is characterized by a fluctuatingly abnormal mental status, often with an insidious onset that delays clinical recognition of the disorder. Somnolence, obtundation, and agitation can occur and may progress to coma. Ocular reflexes are preserved, but disconjugate eye movements or tonic ocular deviation (downward) may be found in rare instances. A flapping tremor (asterixis) is often conspicuous; in severe cases, decerebrate or decorticate posturing, hyperreflexia, and bilateral extensor plantar responses may be present. Routine liver function tests may not correlate with the severity of the encephalopathy. The fasting arterial ammonia concentration and EEG findings of diffuse slow activity with associated triphasic waves are more helpful in determining severity. Respiratory alkalosis is commonly present. The CSF often shows nonspecific abnormalities, but an increase in the glutamine level is strongly supportive of hepatic encephalopathy. Abnormal signal intensities may be found in the basal ganglia on T_1-weighted MRI.

The pathogenesis of hepatic encephalopathy is multifactorial. A critical element is altered brain-energy metabolism and increased permeability of the blood-brain barrier, which facilitates the passage of neurotoxins.[19] Treatment consists of reduction of hyperammonemia; restriction of dietary protein intake; control of GI bleeding; management of portal hypertension; removal of blood from the GI tract; administration of lactulose or neomycin; correction of associated electrolyte, biochemical, and hematologic disturbances; and general supportive measures[20] [*see 4:XIII Enteral and Parenteral Nutrition*].

Chronic Non-Wilsonian Hepatocerebral Degeneration

Some patients with chronic liver disease develop a permanent neurologic deficit resembling that of Wilson disease, with action (intention) tremor, ataxia, dysarthria, and choreoathetosis. Severity correlates best with the fasting arterial ammonia level. Neuroimaging studies may be abnormal. There is no specific treatment.

Liver Transplantation

An encephalopathy that worsens soon after liver transplantation suggests organ rejection, cerebral anoxia, or a complication of immunosuppressive agents, especially cyclosporine. Seizures often occur, suggesting metabolic disturbances, cerebrovascular disease, infections, or medication complications. Encephalopathies occurring weeks or months after liver transplantation are usually caused by infections or malignancies involving the nervous system.

PANCREATIC ENCEPHALOPATHY

Acute pancreatitis has been associated with a transient encephalopathy, but its symptoms are nonspecific and resemble those of other metabolic encephalopathies. Diagnosis, therefore, hinges on exclusion of other metabolic causes.

GASTROINTESTINAL DISEASES

Nutritional deficiency is the usual cause of any neurologic complication of GI disorders, but it is usually impossible to determine the responsible nutrient. Neurologic complications oc-

cur in up to 15% of patients who undergo gastric resection. Vitamin B_{12} absorption is impaired because of loss of gastric intrinsic factor; impaired vitamin B_{12} absorption can lead to a variety of disturbances [*see 5:III Anemia: Production Defects*]. Gastric plication has been associated with a nonspecific encephalopathy, myelopathy, polyneuropathy, Wernicke encephalopathy, and a nutritional amblyopia, but the responsible nutritional deficiencies are unknown. Chronic gluten enteropathy causes a progressive and sometimes fatal CNS disorder, with some combination of encephalopathy, myelopathy, cerebellar disturbances, and peripheral neuropathy.

RENAL FAILURE

Encephalopathy can result from uremia or its treatment.[21] Uremic encephalopathy clinically resembles other metabolic encephalopathies, and its severity cannot be related to any single laboratory abnormality. Its pathophysiologic basis remains uncertain, but it is usually attributed to the accumulation of toxic organic acids in the CNS or to the direct toxic effects of parathyroid hormone.

Dialysis Disequilibrium Syndrome

The dialysis disequilibrium syndrome consists of an encephalopathy characterized by headache, irritability, agitation, somnolence, seizures, muscle cramps, and nausea. It occurs during or after hemodialysis or peritoneal dialysis and has been related to shifting of water to the brain. Other features of dialysis disequilibrium syndrome include exophthalmos, increased intraocular and intracranial pressure, and papilledema.

Dialysis Dementia

In some patients who have undergone dialysis for more than a year, a fatal encephalopathy called dialysis dementia has developed. The cause of this condition is uncertain, although aluminum intoxication has been suggested as the cause, for two reasons: (1) increased cerebral concentrations of aluminum are found at postmortem examination and (2) dialysis dementia has become rare since aluminum was removed from dialysates.[22] A characteristic early feature is hesitancy of speech, followed by speech arrest. As the disorder advances, intellectual function declines, and hallucination, delusions, seizures, myoclonus, asterixis, gait disturbances, and other neurologic abnormalities develop. Death usually occurs within 1 year after onset. Deferoxamine, a chelating agent that binds aluminum, is often prescribed, but the optimal duration of treatment is unknown. Deferoxamine therapy may exacerbate encephalopathy in patients with high serum aluminum levels and may provoke visual and auditory disturbances.[23,24]

Renal Transplantation

The long-term immunosuppressive treatment of patients who undergo renal transplantation can lead to encephalopathic complications.[21]

ELECTROLYTE DISTURBANCES

Sodium

Hyponatremia and hypernatremia have several causes [*see 10:I Renal Function and Disorders of Water and Sodium Balance*]. Rapid changes in serum sodium concentration can cause encephalopathy because the osmotic equilibrium between the CSF and other body fluids is altered. Disturbances of cognition and arousal occur and may lead to coma. Associated features include myoclonus, asterixis, tremulousness, and seizures. Seizures are often poorly responsive to anticonvulsant medication unless the associated metabolic disturbance has been corrected. Focal motor deficits (e.g., hemiparesis) can occur with hyponatremia in the absence of any structural lesion or with hypernatremia as a result of intracerebral or subdural hemorrhage related to osmotically caused brain shrinkage, with secondary tearing of blood vessels.

In patients with acute brain syndromes, such as subarachnoid hemorrhage, hyponatremia is often erroneously attributed to the syndrome of inappropriate antidiuretic hormone secretion (SIADH). In such patients, hyponatremia is more often caused by salt wasting than by SIADH; because plasma volume is reduced, fluid restriction exacerbates hypovolemia and can result in cerebral ischemia. Hyponatremia should be corrected at a rate not exceeding 12 mEq/L/day, because rapid correction of hyponatremia leads to central pontine myelinolysis [*see* Nutritional Deficiencies, *below*].[25] Central pontine myelinolysis may obscure or follow improvement in hyponatremic encephalopathy. When severe, it leads to obtundation, a spastic or flaccid quadriparesis, and pseudobulbar palsy; in mild cases, clinical deficits are minimal, though conspicuous abnormalities may be detectable on MRI.

Potassium

Alterations of serum potassium concentration can have several causes [*see 10:II Disorders of Acid-Base and Potassium Balance*]. Hyperkalemia usually causes disturbances of cardiac rhythm before affecting neurologic function, but occasionally the arrhythmia is accompanied by burning paresthesias, progressive flaccid paralysis, depressed tendon reflexes, and mental changes. Treatment depends on the underlying cause, the severity of the electrolyte disturbance, and the electrocardiographic findings. Hypokalemia usually causes reversible neuromuscular dysfunction rather than an encephalopathy.

Calcium

The main CNS complication of hypercalcemia is an encephalopathy characterized by an impaired level of consciousness, headache, apathy or agitation, and, in rare cases, seizures. Neuromuscular complications (e.g., muscle weakness and fatigability) result from involvement of the peripheral nervous system.

Tetany is a widely recognized manifestation of hypocalcemia, but focal or generalized seizures can also occur, as can an encephalopathy characterized by confusion, hallucinations, delusions, psychosis, disturbances of consciousness, and cognitive impairment. The seizures are resistant to anticonvulsant drugs until the hypocalcemia is corrected. Other CNS complications include parkinsonism and chorea that clear with correction of the serum calcium level; increased intracranial pressure and a myelopathy may also occur with hypocalcemia.

Magnesium

Hypomagnesemia may coexist with hypocalcemia and has similar neurologic complications. Hypermagnesemia leads to an encephalopathy with drowsiness, confusion, diminished responsiveness, and depressed or absent tendon reflexes. Hypotension, respiratory depression, and weakness from impaired neuromuscular transmission may also be present. In severe cases, coma ensues, with the possibility of a fatal outcome.

PITUITARY DISEASE

Encephalopathy is common in Cushing disease, which leads to a variety of symptoms, including anxiety, agitation, insomnia,

depression, euphoria, excitement, and psychosis. Intracranial hypertension, with its attendant effects on cerebral function, may complicate Cushing syndrome; it occurs particularly after resection of the pituitary adenoma.

Hypopituitarism leads to apathy and intellectual decline, but the specific hormonal basis of these symptoms is uncertain because several hormones are affected concurrently.

Diabetes insipidus leads to an encephalopathy that ranges in severity from irritability to somnolence to coma. Patients are hypotensive and hyperthermic. Vasopressin or a long-acting vasopressin analogue is the usual therapeutic approach.

THYROID DISEASE

Hyperthyroidism

An encephalopathy that is common in hyperthyroidism takes the form of anxiety, restlessness, tremulousness, irritability, emotional lability, poor concentration, headaches, and insomnia. Depression and lethargy may be conspicuous in elderly patients (i.e., apathetic hyperthyroidism). Seizures can occur. Examination commonly reveals a postural tremor and generalized hyperreflexia. Chorea and paroxysmal choreoathetosis have also been described.

A more severe encephalopathic disturbance characterizes thyrotoxic crisis, with confusion and agitation progressing to coma. Thyrotoxic crisis is often associated with fever, cardiac arrhythmias, and other systemic disturbances. It is treated with hydrating and cooling agents, beta blockers, glucocorticoids, and, occasionally, plasmapheresis [*see 3:I Thyroid*].

Hypothyroidism

Mental changes are often seen in cases of hypothyroidism. Apathy, somnolence, and poor concentration are often attributed to depression. Cognitive function may decline. Confusion, delirium, and psychosis also occur. Symptoms reverse with correction of the thyroid disorder. Severe hypothyroidism may lead to impairment of consciousness accompanied by hypotension, hypothermia, respiratory failure, hypoglycemia, and other metabolic abnormalities. If treatment is delayed, the encephalopathy progresses to coma, sometimes with a fatal outcome.

Hashimoto Thyroiditis

A relapsing encephalopathy manifested by confusion, alterations in consciousness, seizures, tremulousness, and myoclonus may develop in patients with Hashimoto disease and high serum antithyroid antibody concentrations; the condition is termed Hashimoto encephalopathy.[26,27] Strokelike episodes of deterioration are common. Investigations reveal a diffusely abnormal EEG and an increased CSF protein concentration without associated pleocytosis, but neuroimaging studies show no abnormalities other than a patchy uptake on isotope brain scan. The disorder is treated with glucocorticoids and is associated with a good long-term prognosis.[27]

DIABETES MELLITUS

An encephalopathy may occur in diabetic patients as a consequence of metabolic derangements directly related to the diabetes or its treatment or to complications of the disease, such as renal failure (see above). It may also be related to cerebrovascular disease, which is relatively common in diabetic patients, who have an increased incidence of hypertension and atherosclerosis.

Diabetic ketoacidosis may be the presenting feature of previously unrecognized diabetes. The condition presents as an encephalopathy, manifested by an altered state of consciousness that progresses from mild confusion to coma. If the patient has no underlying structural disease of the brain, there are usually no focal or lateralizing signs. The encephalopathy is probably multifactorial. Serum hyperosmolality, acidosis, and disseminated intravascular coagulation (DIC) are likely contributing factors of importance; other metabolic derangements, infection, vascular occlusive phenomena, and cerebral edema may also be involved. Treatment is with fluid and electrolyte replacement and I.V. insulin (10 to 20 units given as a bolus, followed by continuous infusion of 10 to 15 U/hr in normal saline, with the dose being lowered as the blood glucose level declines) [*see 9:VI Diabetes Mellitus*].

Nonketotic hyperosmolar coma typically occurs in elderly patients with mild diabetes [*see 9:VI Diabetes Mellitus*]. Progressive obtundation is the predominant clinical manifestation, sometimes accompanied by seizures and focal deficits. Hypotension and evidence of dehydration may be present. Treatment involves fluid replacement with hypotonic (one-half normal) saline, correction of hyperglycemia with I.V. insulin, and correction of other biochemical derangements.

HYPOGLYCEMIA

Hypoglycemia causes an acute metabolic encephalopathy that can lead to irreversible brain damage if treatment is delayed. The most common cause of the hypoglycemia is insulin administration in diabetic patients [*see 9:VI Diabetes Mellitus*]; other causes include hepatic disease, alcoholism, and various tumors [*see 3:IX Hypoglycemia*]. Hypoglycemia leads initially to sweating, tachycardia, dilated pupils, tremulousness, and mental changes, which, depending on the level of the hypoglycemia, may include anxiety, confusion, stupor, or coma. The warning signs of sympathetic overactivity may not be evident in patients with autonomic dysfunction. As the depth of coma increases, the plantar responses become extensor, and decorticate or decerebrate posturing occurs. Brain stem dysfunction (including abnormal pupillary responses) and transitory focal neurologic deficits—sometimes alternating from side to side—may also occur. In some patients, seizures may be the only manifestation of hypoglycemia. Depressed respiration, heart rate, and tendon reflexes presage the development of irreversible brain damage. Administration of glucose improves or reverses symptoms and signs within a few minutes and should not be delayed for laboratory confirmation of hypoglycemia. All patients presenting with an encephalopathy of uncertain cause should therefore immediately receive 50 ml of 50% dextrose I.V.

NUTRITIONAL DEFICIENCIES

Wernicke Encephalopathy

Wernicke encephalopathy is common in alcoholic patients. It also occurs in malnourished patients (especially when glucose or oral hypoglycemic agents are administered), in patients on kidney dialysis, in obese patients treated with gastroplasty, and in patients who vomit persistently. Thiamine (vitamin B_1) deficiency is responsible for the hallmark features of Wernicke encephalopathy, which include ophthalmoplegia (i.e., nystagmus, extraocular palsies, gaze palsies, and, in rare cases, internuclear ophthalmoplegia), gait ataxia, and fluctuating confusional states. Pathologic changes occur in characteristic regions of the brain stem, especially in the mamillary bodies and thalamus. Diffusion-weighted MRI may show signal changes in these character-

istic midline locations.[28] A polyneuropathy is often present, and hypothalamic involvement may lead to hypotension and hypothermia. The prognosis depends on the rapid initiation of effective therapy. Treatment is with thiamine replacement (100 mg I.V. daily for 1 week, followed by daily oral supplementation). Improvement of the ophthalmoplegia commences within a day or so, and improvement of the encephalopathy is seen in days to weeks, but residual deficits are common.

Korsakoff Encephalopathy

As with Wernicke encephalopathy, Korsakoff encephalopathy is attributed to thiamine deficiency, though the precise pathophysiology is unknown. Selective disturbance of memory is the predominant clinical abnormality, but thiamine replacement therapy rarely leads to improvement. There is marked impairment of recent memory and difficulty in incorporating new memories, though immediate recall is intact. Patients are unaware of any deficit and often confabulate. Other cognitive abnormalities are found less often. The disorder is common in persons with chronic alcoholism, often occurring in association with Wernicke encephalopathy (Wernicke-Korsakoff syndrome). The pathologic changes are similar in distribution to those in Wernicke encephalopathy (see above).

Subacute Necrotizing Encephalomyelopathy

Subacute necrotizing encephalomyelopathy occurs as an autosomal recessive disorder in children (rarely in adults). The distribution of pathologic changes of this disorder resembles that of Wernicke disease, suggesting that the disorder relates to a disturbance of thiamine metabolism. Thiamine supplementation is sometimes helpful, but the encephalomyelopathy is not caused by nutritional deficiencies alone, and its precise pathophysiologic basis is unclear. Symptoms and signs include cognitive decline, seizures, flaccid weakness, optic atrophy, nystagmus, ataxia, vomiting, and irregular respirations. Death often results within a few months. There is no specific treatment.

Pellagra

Deficiency of niacin (nicotinic acid) leads to an encephalopathy, as well as cutaneous lesions, glossitis, anemia, and GI disturbances (e.g., anorexia, nausea, and diarrhea). Poor concentration, irritability, and affective complaints are followed by confusion, hallucinations, delusions, and pyramidal and extrapyramidal deficits (e.g., tremor and rigidity). A polyneuropathy may also occur. Pathologic changes are widespread in the CNS but most commonly involve Betz cells in the motor cortex. The response to treatment with niacin is usually disappointing, which suggests that other nutritional deficiencies may be involved.[29]

Central Pontine Myelinolysis

Central pontine myelinolysis is manifested clinically by confusion or more marked impairment of consciousness; pseudobulbar palsy; pyramidal deficits in the lower extremities or all extremities; and extensor plantar responses. Symptoms and signs progress over days to weeks; locked-in syndrome or coma may occur in advanced cases. The pathologic hallmark of the disorder is a breakdown and loss of myelin in the anterior pons and other brain stem regions, which may be visualized by MRI.[30] The disorder is associated with alcoholism, electrolyte disturbances, malignant disease, and malnutrition. It relates particularly to the rapid correction of hyponatremia; for this reason, hyponatremia should generally be corrected by no more than 12 mEq/L/day [*see* Electrolyte Disturbances, *above*].

Vitamin B_{12} Deficiency

Encephalopathy is a well-recognized complication of vitamin B_{12} deficiency. It may be accompanied by myelopathy, optic neuropathy, or peripheral neuropathy, or any combination of these conditions. The neurologic complications do not reflect the presence or the severity of any associated megaloblastic anemia. Folic acid masks the hematologic abnormality and fails to prevent the neurologic complications. Although cereal grains have been fortified with folic acid since 1994 in the United States, a 2003 Veterans Affairs study found that since fortification began, there has been no evidence of an increase in the incidence of vitamin B_{12} deficiency in persons who are not anemic.[31]

Treatment with vitamin B_{12} arrests and may reverse the neurologic disorder; the extent of any residual deficit relates to the severity and duration of symptoms before initiation of treatment[32] [*see 5:III Anemia: Production Defects*].

Hyperalimentation

An encephalopathy may occur in patients receiving parenteral hyperalimentation, usually because of metabolic abnormalities such as hypophosphatemia, hyperammonemia, or hyperosmolarity. A malignant form of Wernicke encephalopathy (see above) may develop if thiamine supplementation is not provided to patients who are being supported parenterally for prolonged periods.

Toxic Encephalopathies

IATROGENIC DISORDERS

Many toxic encephalopathies are iatrogenic in origin.[33] The CNS is affected by many drugs in a variety of ways, but the resultant disturbances of CNS function generally reverse with withdrawal of the offending medication. Attention here is directed at those medications that cause diffuse disturbances of cerebral function manifested particularly by alterations in the level of consciousness. Iatrogenic encephalopathies commonly manifest themselves primarily as seizures or as focal neurologic deficits, such as extrapyramidal and cerebellar syndromes.

Coma is a common sequela of overdosage with various drugs, including hypnotics, sedatives, neuroleptics, antidepressants, anticonvulsants, and analgesics [*see 8:I Management of Poisoning and Drug Overdose*]. On clinical examination, pupillary responses are generally found to be intact, though pinpoint pupils are found in cases of opioid poisoning, and dilated, sluggish pupils occur in cases of barbiturate or glutethimide overdose. Spontaneous and reflex ocular movements are typically impaired, even at an early stage, in barbiturate and phenytoin intoxication. Depending on the depth of coma, the corneal reflex may be lost, and painful stimuli to the trunk or extremities may lead to purposive movements, decorticate or decerebrate posturing, or an absence of motor response. Flaccidity and hyporeflexia of the limbs are common, but spasticity, hyperreflexia, and extensor plantar responses are sometimes found.

A fluctuating level of consciousness—with confusion, delirium, hallucinations, and a poor attention span—may occur, especially in the elderly, in reaction to many medications, including various antimicrobials, CNS depressants, antiparkinsonian drugs, anticonvulsants, and cardiovascular agents. Nonspecific

behavioral changes—such as restlessness, irritability, sleep disturbances, affective changes, and nightmares—often occur initially, but their significance may not be recognized. Examination often reveals tremor, asterixis, and myoclonus in addition to mental disturbances. Nystagmus may also be present, particularly when CNS depressants or anticonvulsants have been taken. A similar clinical disturbance may result from the withdrawal of certain medications, such as benzodiazepines and barbiturates.

Some encephalopathies may be related to the infectious and neoplastic complications associated with immunosuppressive and chemotherapeutic agents. Such agents and other drugs causing encephalopathies include (but are not limited to) cisplatin,[34] paclitaxel,[35] valproate,[36] vigabatrin,[37] cyclosporine,[33,38] methotrexate,[39,40] and cefuroxime.[41] Encephalopathies have been associated with transcatheter embolization[42] and allogeneic bone marrow transplantation.[43] Treatment with glucocorticoids can cause encephalopathies characterized by behavioral disturbances and psychoses.[44] Among other medications prescribed for the treatment of cardiac disorders, lidocaine and related agents may cause seizures, tremor, paresthesias, and confusional states. Calcium channel blockers, beta blockers, digoxin, and thiazide diuretics can also cause encephalopathies.

ALCOHOL-RELATED DISORDERS

Alcohol intoxication leads initially to behavioral changes (e.g., disinhibition, irritability, euphoria), dysarthria, ataxia, nystagmus, tachycardia, and cutaneous flushing. Increased alcohol intake may cause obtundation, coma, hyporeflexia, respiratory depression, and death. Level of tolerance to alcohol may influence the behavioral response to a particular blood level of alcohol.

Acute withdrawal from alcohol after a period of regular consumption can lead initially to postural tremor and signs of autonomic hyperactivity. Seizures may also occur, especially within the first 48 hours after withdrawal. The seizures are usually generalized tonic-clonic convulsions that are self-limited and rarely require anticonvulsant drug treatment. Delirium tremens may develop 2 to 7 days after withdrawal; it is characterized by marked agitation, excitement, hallucinations, hyperthermia, dehydration, and hypotension. Treatment is with benzodiazepines, fluid and electrolyte replacement, and general supportive measures [*see 13:III Alcohol Abuse and Dependency*].

Various encephalopathies can occur in alcoholics as a result of nutritional deficiencies (see above).

Miscellaneous Encephalopathies

DISSEMINATED INTRAVASCULAR COAGULATION

DIC may occur in patients with diseases of the brain and other organs; septicemia; immune-mediated disorders; diabetic ketoacidosis; neoplastic disease; or obstetric complications [*see 5:XV Coagulation Disorders*]. The clinical manifestations of DIC, including the predominance of thrombosis or hemorrhage, are influenced by its cause, rate of onset, and severity. Encephalopathy is a common manifestation, ranging in severity from mild confusion to coma. Even comatose patients may recover fully; they therefore require continuing support.

CONNECTIVE TISSUE DISEASES AND VASCULITIDES

Connective tissue diseases are characterized by an autoimmune inflammatory response and vasculitis. The common direct CNS manifestations are encephalopathy with cognitive or behavioral changes and focal neurologic deficits. An encephalopathy may also result from metabolic disturbances related to the involvement of other organs or to treatment with glucocorticoids or immunosuppressive agents.

Common features of direct CNS involvement in polyarteritis nodosa, allergic granulomatous angiitis (Churg-Strauss syndrome), and overlap syndrome include headache (sometimes indicative of an aseptic meningitis) and behavioral disturbances, such as cognitive decline, acute confusion, and affective or psychotic disorders. The EEG may be diffusely slow; neuroimaging studies are sometimes abnormal. Focal CNS deficits, which are uncommon, are usually caused by infarction or hemorrhage. Angiography may not reveal the vasculopathy.

Headache is the most common initial complaint of patients with giant cell (temporal) arteritis [*see 15:VIII Systemic Vasculitis Syndromes*]. Other features of giant cell arteritis include masticatory claudication and acute unilateral or bilateral blindness, which may be permanent. Other CNS complications are rare, but encephalopathy with neuropsychiatric disturbances, strokes, seizures, and other manifestations sometimes occurs in these patients.

In Wegener granulomatosis, cerebral involvement results from vasculitis or extension of granulomas from the upper respiratory tract.[45] The resulting encephalopathy may be caused by basilar meningitis with associated cranial neuropathies, temporal lobe dysfunction, cerebral infarction, or venous sinus obstruction.

Headache, cognitive deficits, behavioral and neuropsychiatric disturbances, and focal or multifocal deficits from small infarcts are usual presenting features of isolated angiitis of the CNS. Consciousness becomes depressed as the disease advances. The CSF exhibits a lymphocytic pleocytosis and increased protein concentration. Focal ischemic changes may be detected by computed tomography or MRI; angiography sometimes shows beading of vessels. Meningeal and brain biopsies are usually necessary to make a definitive diagnosis.

Rheumatoid arthritis, the most common of the connective tissue diseases, rarely causes encephalopathy unless involvement of the upper cervical spine or atlantoaxial dislocation causes headaches or hydrocephalus or leads to brain stem deficits from direct medullary compression or vertebral artery involvement.

In most patients with systemic lupus erythematosus, neurologic complications ultimately develop, often during the first year. Episodic affective or psychotic disorders are common and are often difficult to distinguish from corticosteroid-related mental disturbances. Alterations in consciousness sometimes occur. Focal neurologic deficits may be manifestations of strokes related to cardiac valvular disease, the presence of antiphospholipid antibodies, or cerebral vasculitis.[46] Generalized or partial seizures sometimes occur; they probably result from microinfarcts, metabolic disturbances, and systemic infections.

Uncommonly, patients with Sjögren syndrome may experience encephalopathies that are characterized by behavioral and psychiatric disturbances; these encephalopathies may result from aseptic meningitis, meningoencephalitis, or focal neurologic dysfunction. About 20% of patients with Behçet syndrome develop an aseptic meningitis or meningoencephalitis that leads to an encephalopathy. Focal or multifocal deficits may also result from cerebral ischemia. The CSF is commonly abnormal, with a mild pleocytosis and increased protein concentration.

Antiphospholipid antibodies (i.e., the lupus anticoagulant and anticardiolipin antibodies) are found especially in patients with certain connective tissue diseases,[47] in patients taking various

medications, in patients with infections and obstetric complications, and as an incidental finding [*see 5:XIV Thrombotic Disorders*]. An acute ischemic encephalopathy, manifested by confusion, obtundation, quadriparesis, and bilateral pyramidal signs, has been described in patients with antiphospholipid antibodies.[48]

The pathogenesis of the thrombotic tendency associated with the presence of these antibodies is unclear. Immunosuppressive therapy is not indicated. Cerebral thrombosis is managed as it is in other contexts.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Coma and Impaired Consciousness: A Clinical Perspective. Young GB, Ropper AH, Bolton CF, Eds. McGraw-Hill, New York, 1998

2. Edgren E, Hedstrand U, Kelsey S, et al: Assessment of neurological prognosis in comatose survivors of cardiac arrest. Lancet 343:1055, 1994

3. McKinney AM, Teksam M, Felice R, et al: Diffusion-weighted imaging in the setting of diffuse cortical laminar necrosis and hypoxic-ischemic encephalopathy. AJNR Am J Neuroradiol 25:1659, 2004

4. Els T, Kassubek J, Kubalek R, et al: Diffusion-weighted MRI during early global cerebral hypoxia: a predictor for clinical outcome? Acta Neurol Scand 110:361, 2004

5. Robinson LR, Micklesen PJ, Tirschwell DL, et al: Predictive value of somatosensory evoked potentials for awakening from coma. Crit Care Med 31:960, 2003

6. Fischer C, Luaute J, Adeleine P, et al: Predictive value of sensory and cognitive evoked potentials for awakening from coma. Neurology 63:669, 2004

7. Giacino J, Whyte J: The vegetative and minimally conscious states: current knowledge and remaining questions. J Head Trauma Rehabil 20:30, 2005

8. Wijdicks EF, Cranford RE: Clinical diagnosis of prolonged states of impaired consciousness in adults. Mayo Clin Proc 80:1037, 2005

9. Schiff ND, Rodriquez-Moreno D, Kamal A, et al: fMRI reveals large-scale network activation in minimally conscious patients. Neurology 64:514, 2005

10. Giacino JT, Ashwal S, Childs N, et al: The minimally conscious state: definitions and diagnostic criteria. Neurology 58:349, 2002

11. Medical aspects of the persistent vegetative state (pt 1). Multi-Society Task Force on PVS. N Engl J Med 330:1499, 1994

12. Medical aspects of the persistent vegetative state (pt 2). Multi-Society Task Force on PVS. N Engl J Med 330:1572, 1994

13. Brodie BR, Stuckey TD, Hansen CJ, et al: Timing and mechanism of death determined clinically after primary angioplasty for acute myocardial infarction. Am J Cardiol 79:1586, 1997

14. Baker RA, Andrew MJ, Knight JL: Evaluation of neurologic assessment and outcomes in cardiac surgical patients. Semin Thorac Cardiovasc Surg 13:149, 2001

15. Nussmeier NA: A review of risk factors for adverse neurologic outcome after cardiac surgery. J Extra Corpor Technol 34:4, 2002

16. Hotson JR: Neurological complications of cardiac surgery. Neurology and General Medicine, 3rd ed. Aminoff MJ, Ed. Churchill Livingstone, New York, 2001, p 45

17. West JB: The physiologic basis of high-altitude diseases. Ann Intern Med 141:789, 2004

18. Wilson JX, Young GB: Progress in clinical neurosciences: sepsis-associated encephalopathy: evolving concepts. Can J Neurol Sci 30:98, 2003

19. Ferenci P, Lockwood A, Mullen K, et al: Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology 35:716, 2002

20. Butterworth RF: Complications of cirrhosis III: hepatic encephalopathy. J Hepatol 32(1 suppl):171, 2000

21. Brouns R, De Deyn PP: Neurological complications in renal failure: a review. Clin Neurol Neurosurg 107:1, 2004

22. Rob PM, Niederstadt C, Reusche E: Dementia in patients undergoing long-term dialysis: aetiology, differential diagnoses, epidemiology and management. CNS Drugs 15:691, 2001

23. Sherrard DJ, Walker JV, Boykin JL: Precipitation of dialysis dementia by deferoxamine treatment of aluminum-related bone disease. Am J Kidney Dis 12:126, 1988

24. Olivieri NF, Buncic JR, Chew E, et al: Visual and auditory neurotoxicity in patients receiving subcutaneous deferoxamine infusions. N Engl J Med 314:869, 1986

25. Sterns RH, Riggs JE, Schochett SS: Osmotic demyelination syndrome following correction of hyponatremia. N Engl J Med 314:1535, 1986

26. Shaw PJ, Walls TJ, Newman MK, et al: Hashimoto's encephalopathy: a steroid-responsive disorder associated with high anti-thyroid antibody titers: report of 5 cases. Neurology 41:228, 1991

27. Chong JY, Rowland LP, Utiger RD: Hashimoto encephalopathy: syndrome or myth? Arch Neurol 60:164, 2003

28. Doherty MJ, Watson NF, Uchino K, et al: Diffusion abnormalities in patients with Wernicke encephalopathy. Neurology 58:655, 2002

29. Mancall EL: Nutritional disorders of the nervous system. Neurology and General Medicine, 3rd ed. Aminoff MJ, Ed. Churchill Livingstone, New York, 2001, p 277

30. Lampl C, Yazdi K: Central pontine myelinolysis. Eur Neurol 47:3, 2002

31. Mills JL, Von Kohorn I, Conley MR, et al: Low vitamin B-12 concentrations in patients without anemia: the effect of folic acid fortification of grain. Am J Clin Nutr 77:1474, 2003

32. Healton EB, Savage DG, Brust JC, et al: Neurologic aspects of cobalamin deficiency. Medicine (Baltimore) 70:229, 1991

33. Mastaglia FL: Iatrogenic (drug-induced) disorders of the nervous system. Neurology and General Medicine, 3rd ed. Aminoff MJ, Ed. Churchill Livingstone, New York, 2001, p 593

34. Lyass O, Lossos A, Hubert A, et al: Cisplatin-induced non-convulsive encephalopathy. Anticancer Drugs 9:100, 1998

35. Chang SM, Kuhn JG, Rizzo J, et al: Phase I study of paclitaxel in patients with recurrent malignant glioma: a North American Brain Tumor Consortium report. J Clin Oncol 16:2188, 1998

36. Verrotti A, Trotta D, Morgese G, et al: Valproate-induced hyperammonemic encephalopathy. Metab Brain Dis 17:367, 2002

37. Ifergane G, Masalha R, Zigulinski R, et al: Acute encephalopathy associated with vigabatrin monotherapy in patients with mild renal failure. Neurology 51:314, 1998

38. Madan B, Schey SA: Reversible cortical blindness and convulsions with cyclosporin A toxicity in a patient undergoing allogeneic peripheral stem cell transplantation. Bone Marrow Transplant 20:793, 1997

39. Rubnitz JE, Relling MV, Harrison PL, et al: Transient encephalopathy following high-dose methotrexate treatment in childhood acute lymphoblastic leukemia. Leukemia 12:1176, 1998

40. Lovblad K, Kelkar P, Ozdoba C, et al: Pure methotrexate encephalopathy presenting with seizures: CT and MRI features. Pediatr Radiol 28:86, 1998

41. Herishanu YO, Zlotnik M, Mostoslavsky M, et al: Cefuroxime-induced encephalopathy. Neurology 50:1873, 1998

42. Katsushima S, Inokuma T, Oi H, et al: Acute hepatic failure following transcatheter arterial embolization for the treatment of hepatocellular carcinoma. Digestion 58:189, 1997

43. Antonini G, Ceschin V, Morino S, et al: Early neurologic complications following allogeneic bone marrow transplant for leukemia: a prospective study. Neurology 50:1441, 1998

44. De Angelis LM, Delattre JY, Posner JB: Neurological complications of chemotherapy and radiation therapy. Neurology and General Medicine, 3rd ed. Aminoff MJ, Ed. Churchill Livingstone, New York, 2001, p 437

45. Nishino H, Rubino FA, DeRemee RA, et al: Neurological involvement in Wegener's granulomatosis: an analysis of 324 consecutive patients at the Mayo Clinic. Ann Neurol 33:4, 1993

46. Futrell N, Schultz LR, Millikan C: Central nervous system disease in patients with systemic lupus erythematosus. Neurology 42:1649, 1992

47. Cervera R, Piette JC, Font J, et al: Antiphospholipid syndrome: clinical and immunologic manifestations and patterns of disease expression in a cohort of 1,000 patients. Arthritis Rheum 46:1019, 2002

48. Briley DP, Coull BM, Goodnight SH Jr: Neurological disease associated with antiphospholipid antibodies. Ann Neurol 25:221, 1989

Acknowledgments

Table 2 Data from Diagnosis of Stupor and Coma, 2nd ed., by F. Plum and J. B. Posner. FA Davis Co, Philadelphia, 1972.

Table 3 Data from "Assessment of Neurological Prognosis in Comatose Survivors of Cardiac Arrest," by E. Edgren, U. Hedstrand, S. Kelsey, et al, in Lancet 343:1055, 1994.

VIII HEADACHE

Randolph W. Evans, m.d.

Headaches are a near-universal experience, with a 1-year prevalence of 90% and a lifetime prevalence of 99%. Each year in the United States, 9% of adults see physicians for headaches and 83% self-medicate. Headaches are one of the most common complaints of patients seen by primary care physicians and account for 20% of outpatient visits to neurologists.

The differential diagnosis of headaches is one of the longest in medicine, with over 300 different types and causes. Although most headaches are of benign (and still poorly understood) origin, some headaches can have serious and even potentially life-threatening causes. Thus, it is critical for the physician to diagnose headaches as precisely as possible.

The International Headache Society (IHS) criteria,[1] which were introduced in 1988 and updated in 2004, are the worldwide standard for headache classification. IHS criteria categorize headaches as primary or secondary. Primary headaches—those with no other underlying cause—account for 90% of headaches. This category includes migraine, tension, cluster, and miscellaneous headaches, such as primary exertional headaches. There are a large number of secondary headaches, which are classified according to their causes [*see Table 1*].

A careful history, examination, and, in some cases, diagnostic testing will usually provide the accurate diagnosis of a headache, although a precise diagnosis is not always possible. For example, some benign headaches have both migraine and tension-type features. Chronic daily headaches may also be difficult to classify.

This chapter reviews pain-sensitive structures in the head, the history and examination in patients with headache, and many of the primary and secondary headaches. The interested reader may wish to refer to a headache textbook for more comprehensive information (see the reference list at the end of the chapter).[2-5]

Pain-Sensitive Structures

Similar headaches can have different causes because there are a limited number of pain-sensitive structures in the head. Although all pain is registered in the brain, the brain parenchyma itself is not pain sensitive. The arachnoid, ependyma, and dura (except portions near blood vessels) are also insensitive to pain. The following are sensitive to pain: cranial nerves V, VII, IX, and X; the circle of Willis and proximal continuations; meningeal arteries; large veins in the brain and dura; and structures external to the skull (including scalp and neck muscles, cutaneous nerves and skin, the mucosa of paranasal sinuses, teeth, cervical nerves and roots, and the external carotid arteries and branches).

Headache pain may be felt at its source (e.g., cheek or forehead pain from maxillary or frontal sinusitis) or be referred from another site. For example, supratentorial structures are innervated by the ophthalmic division of the trigeminal nerve, whereas infratentorial and posterior fossa structures are supplied by C_2 and C_3. Thus, a cerebellar hemisphere lesion generally refers pain posteriorly and an occipital lobe lesion refers pain anteriorly. The caudal nucleus of the trigeminal nerve, which is located from the midpons to the third cervical segment, receives pain messages from the upper cervical roots and the trigeminal nerve. Thus, pain from the upper cervical spine or posterior fossa can also be referred to the front of the head.

Headache History

The headache history is usually essential to establishing the diagnosis.[6] Key elements of the history include not only the features of the headache but also the patient's own diagnosis, past history, psychosocial history, and family history [*see Table 2*]. In gathering the key elements, both open-ended ("What are your headaches like?") and closed-ended ("Do you have nausea with the headache?") questions are necessary [*see Table 3*]. Often, it is helpful to ask about a history of mild headaches and bad headaches. Some patients are not able to clearly remember or articulate features of the headache ("It's just a headache, doc."). With patients who have chronic headaches, it may be necessary to provide a headache diary or have them record features of their headaches and then return for a later appointment.

Table 1 Major Categories of Headache Disorders[1]

Primary Headaches
Migraine
Tension-type headache
Cluster headache and chronic paroxysmal hemicrania
Miscellaneous headaches unassociated with structural lesion: idiopathic stabbing, external compression, cold stimulus, benign cough, benign exertional, associated with sexual activity
Secondary Headaches
Headache associated with head trauma
Headache associated with vascular disorder: acute ischemic cerebrovascular disorder, intracranial, hematoma, subarachnoid hemorrhage, unruptured vascular malformation, arteritis, carotid or vertebral artery pain, venous thrombosis, arterial hypertension, associated with other vascular disorder
Headache associated with nonvascular intracranial disorder: high and low cerebrospinal fluid pressure, intracranial infection, intracranial sarcoidosis and other noninfectious inflammatory disease, related to intrathecal injections, intracranial neoplasm, associated with other intracranial disorder
Headache associated with substances or their withdrawal: acute and long-term substance use or exposure, withdrawal after acute and long-term use, associated with substances with uncertain mechanism
Headache associated with noncephalic infection: viral infection, bacterial infection, other infection
Headache associated with metabolic disorder: hypoxia, hypercapnia, mixed hypoxia and hypercapnia, hypoglycemia, dialysis, other metabolic abnormality
Headache or facial pain associated with disorder of cranium, neck, eyes, ears, nose, sinuses, teeth, mouth, or other facial or cranial structures
Cranial neuralgias, nerve trunk pain, and deafferentation pain
Persistent pain of cranial nerve origin, trigeminal neuralgia, glossopharyngeal neuralgia, nervus intermedius neuralgia, superior laryngeal neuralgia, occipital neuralgia, central causes of head and facial pain other than tic douloureux

Table 2 Key Elements of the Headache History

Element	*Examples*
Temporal profile	
Age of onset	Migraine, usually ≤ 40 yr; temporal arteritis after age 50
Time to maximum intensity	Gradual in migraine; immediate in thunderclap headache
Frequency	Eight per day with cluster; a few in a lifetime with migraine
Time of day	Cluster or migraine on awakening; tension in the afternoon
Duration	Migraine, 4–72 hr; cluster, 15 min to 3 hr
Recurrence	Migraines recur about 30% of the time after relief with a triptan
Headache features	
Location	Cluster always unilateral; migraine unilateral or bilateral
Quality of pain	Migraine, throbbing; tension, pressure; cluster, boring
Severity of pain	Most severe headaches are migraine or cluster
Associated symptoms and signs	
Before headache	Migraine aura or prodrome, fever before meningitis
During headache	Nausea/vomiting in migraine, eye redness and tearing in cluster
After headache	Mental dullness after migraine
Aggravating or precipitating factors	
Trauma	Migraine, subdural hematoma
Medical conditions	Obesity in pseudotumor cerebri
Triggers	Present in 85% of migraineurs; stress in tension type
Trigger zones	Trigeminal neuralgia
Activity	Exertional headache, benign orgasmic cephalalgia
Pharmacologic	Oral contraceptives, rebound headaches
Relieving factors	
Nonpharmacologic	Sleep for migraine, relaxation for tension type
Pharmacologic	Prescription and over-the-counter drugs, herbs, and vitamins
Previous evaluation and treatment	Obtain medical records as appropriate
Psychosocial history	
Substance use	Rebound headaches from too much caffeine
Occupational and personal life	Stress, occupational toxin exposures
Psychological history	Depression, anxiety
Sleep history	Deprivation causing headaches; sleep apnea
Impact of headache	Missed school, work, family activities
Patient's own diagnosis	May incorrectly self-diagnose as brain tumor, sinus headache, aneurysm
Family history	70% of migraineurs have a first-degree relative with migraine
Complete medical and surgical history	Asthma as a contraindication to beta blockers for migraine

Physical Examination

A directed physical examination may be informative. Examples of significant abnormal findings include hypertension, fever, cervical lymphadenopathy in infectious mononucleosis, cervical trigger points in tension-type headache, and maxillary sinus tenderness in sinusitis. Every patient seen for headaches should have, at the least, a screening neurologic examination; this takes only a few minutes to perform. Although the results of this examination are usually normal, the examination can point to significant underlying disease by revealing abnormalities such as papilledema, a mild lateral rectus paresis, unequal pupils, a mild hemiparesis, or a Babinski sign.

Clinical Classification

In most cases, the findings on history and physical examination will point the clinician toward the diagnosis of primary headache. The three most common primary headaches are migraine, episodic tension type, and cluster headache [*see Table 4*]. Much less often, the clinical features suggest secondary headache [*see Table 5*].

Diagnostic Testing

The vast majority of headaches require no diagnostic testing at all; they can be diagnosed accurately on the basis of a detailed history and a physical examination.[7,8] For example, patients who meet IHS criteria for migraine rarely have abnormal neuroimaging findings to explain the headache. In patients with headache of any type and a normal neurologic examination, the yield of a computed tomographic or magnetic resonance imaging scan is only about 2% or less. However, certain clinical features, patient characteristics, and associated symptoms and signs justify neuroimaging for headaches [*see Table 6*].

Although a CT scan of the head will detect most pathologic conditions that cause headaches and it is the preferred study for acute head trauma and subarachnoid hemorrhage (SAH), an MRI scan is generally preferred for evaluation of headaches. An MRI may demonstrate pathology not seen on a standard CT scan, including abnormalities of the paranasal sinuses, pituitary, posterior fossa, cortical veins (e.g., superior sagittal sinus thrombosis), cervicomedullary junction (e.g., Chiari I malformation), intracranial aneurysms, carotid dissection, infarcts, white-matter abnormalities, congenital abnormalities, and neoplasms.

Electroencephalography (EEG) is not useful for the routine evaluation of patients with headache. However, EEG may be helpful if the patient has headaches and symptoms suggesting a seizure disorder or alteration of consciousness.

Blood tests are generally not helpful for the diagnosis of headaches, but specific tests are indicated when certain conditions are suspected. Such conditions, along with their indicated tests, include arteritis (erythrocyte sedimentation rate [ESR] or C-reactive protein [CRP] in temporal arteritis); infection (tests for infectious mononucleosis, Lyme disease, and HIV); and antiphospholipid antibody syndrome, which should be considered in patients with white-matter lesions on MRI. Complete blood counts and metabolic studies are indicated in suspected anemia, renal failure, or hypercalcemia; and endocrine studies are indicated in suspected hypothyroidism or pituitary tumors. Blood studies are also valuable as a baseline for monitoring adverse effects of certain medications for preventing headache (e.g., valproic acid and carbamazepine).

Lumbar puncture can be helpful in diagnosing meningitis, encephalitis, meningeal carcinomatosis, lymphomatosis, SAH, high cerebrospinal fluid pressure (e.g., pseudotumor cerebri), and low CSF pressure. MRI or CT scanning is always performed before a lumbar puncture for the evaluation of headaches, to rule out mass lesions, except in some cases in which acute meningitis is suspected. Lumbar puncture is often indicated for the first or worst headache, to exclude SAH [*see* First or Worst and Thunderclap Headaches, *below*]; headache with fever or other findings that suggest an infectious cause; a subacute or progressive headache in a patient with HIV infection or carcinoma; and an atypical chronic headache (e.g., to rule out pseudotumor cerebri in an obese woman without papilledema).

Migraine

EPIDEMIOLOGY

The 1-year prevalence of migraine is 18% in women, 6% in men, and 5% in children (both boys and girls). In the United States, 28 million persons have migraine each year.[9,10] The lifetime prevalence is 25% for women and 8% for men. Migraine begins before the age of 20 in 50% of cases and after the age of 50 in 2%; the highest prevalence is from 25 to 50 years of age. About 70% of migraineurs have a positive family history in a first-degree relative. The mode of transmission of susceptibility to migraine remains unclear; there is probably genetic heterogeneity. Interestingly, there is a high prevalence in neurologists (lifetime of 47% in males and 63% in females).[11]

The frequency of migraine can range from once in a lifetime to almost daily. The frequency of attacks in migraineurs is as follows: one to 12 a year, 38%; one to three a month, 37%; one a week, 11%; and two to six a week, 14%. Half of migraineurs do not know that they have migraine, with 42% of undiagnosed patients self-diagnosing themselves with so-called sinus headache and 43% having received a diagnosis of sinus headache from a physician.[10]

There are many disorders with a greater-than-coincidental association (comorbidity) with migraine, including stroke, epilepsy, systemic lupus erythematosus, Raynaud syndrome, multiple sclerosis, and essential tremor.[12] Psychiatric disorders comorbid with migraine include bipolar disorder, major depression, generalized anxiety disorder, panic disorder, and simple and social phobia. Migraine may also be associated with hypertension, mitral valve prolapse, and patent foramen ovale.[13]

Table 3 Helpful Questions to Ask for the Headache History[6]

Do you have different types of headaches or just one?
Where does the headache hurt?
When did you first start having these headaches?
What were you doing when the headache started?
How long before the headache reaches maximal intensity?
How long does the headache last?
Does the headache recur? If so, how often?
What is the pain like? Is it a pressure, throbbing, pounding, aching, or stabbing pain?
Is the pain mild, moderate, or severe?
On a scale of 1 to 10, with 10 the worst and 1 the least, how would you rate the headache?
Do you have trouble with your vision before or during the headache?
Do you have other symptoms (e.g., nausea, vomiting, light sensitivity, noise sensitivity, discomfort with eye movement) with the headache?
Are there signs present (e.g., fever, ptosis, miosis)?
Do you have triggers of your headaches (e.g., menses, stress, foods, beverages, lack of sleep, oversleeping, strong odors, trigger zones)?
What makes the headache worse (e.g., coughing, bending over, physical activity)?
What makes the headache better (e.g., sleep, lying down in a quiet room)?
Do your headaches have any impact on your life?
Do you take over-the-counter medications, vitamins, or herbs for your headaches? If so, how much and how often?
Do you drink caffeinated beverages? If so, what types and how many?
What prescription drugs have you tried and with what effect?
What doctors have you seen in the past for your headaches?
What other treatments have you tried and with what success (e.g., acupuncture, chiropractic, biofeedback, stress management, massage)?
Have you been under much stress lately?
Have you been depressed?
Do you have any parents or siblings with a history of migraines or bad headaches?

ETIOLOGY AND PATHOPHYSIOLOGY

Although the mechanism of migraine remains incompletely understood, there is growing evidence that migraine is a neurovascular disorder.[14] The aura that precedes some migraines is a slow march of visual or other neurologic symptoms associated with changes in neuronal activity that result in spreading neural depression from the occipital cortex. Excitatory changes produce increased blood flow, followed by reduced blood flow caused by neuronal inhibition.

The trigeminal nerve and the blood vessels it innervates may constitute the anatomic substrate for migraine pain. Input from the pain-sensitive cranial nerves and dura passes through the ophthalmic division of the trigeminal ganglion to the trigeminocervical complex (the trigeminal nucleus caudalis and dorsal horns of C_1 and C_2) that produces referred pain in the head (especially the ophthalmic division) and upper posterior neck. When the peripheral branches of the trigeminal nerve are activated during migraine, pain results from neurogenic inflammation that is produced by the antidromic release of calcitonin gene–related peptide by trigeminal nerve endings and that is associated with the release of other pain substances from plasma, platelets, and mast cells (e.g., histamine, prostaglandin, and serotonin). These substances induce vasodilatation and extravasation of plasma proteins and the sensitization of trigeminal nociceptive nerve endings. Throbbing pain and exacerbation by activities such as bending over, head movement, coughing, and walking may reflect mechanical hypersensitivity of meningeal C-fiber nociceptors. Nitric oxide released from blood vessels, perivascular nerve endings, or brain tissue can be a trigger for migraine pain.

Pain signals in the trigeminocervical complex undergo central processing, with second-order neurons receiving input and projecting rostrally to the contralateral thalamus (ventrobasal complex and medial nuclei) and then to the activating cortex (anterior cingulate, insular, and frontal), the periaqueductal gray matter (dorsal raphe nuclei), and the locus coeruleus. Aminergic areas in the periaqueductal gray matter and locus coeruleus influence the incoming pain and cortical blood flow. A continuous

Table 4 Features of Selected Primary Headaches[6]

Feature	*Migraine*	*Episodic Tension-Type*	*Episodic Cluster*
Female-to-male ratio	1:1 before puberty, 3:1 after	5:4	1:5
Family history	First-degree relatives affected in 80% of cases	Frequent	Rare
Typical age at onset (yr)	92% before age 40, 2% after age 50	20–40	20–40
Visual aura	20% of cases	No	Occasional
Location	Unilateral 60%, bilateral 40%	Bilateral > unilateral; anywhere on the head, posterior neck, face	Unilateral, especially orbital, periorbital, frontotemporal
Quality	Pulsatile or throbbing in 85%	Pressure, aching, tight, squeezing	Boring, burning, or stabbing
Severity	Mild to severe (moderate to severe [untreated] in 80%)	Mild to moderate	Severe
Triggers	Present in 85%; numerous	Stress, lack of sleep	Alcohol, nitrates
Duration	4–72 hr; duration > 24 hr (untreated) in 62%; may be < 1 hr in children	Hours to days	15 min to 3 hr
Frequency	Rare to frequent	Rare to frequent	1–8 a day during clusters
Periodicity	Menstrual migraine	No	Yes; average bout, 4–8 wk; average 1 or 2 bouts yearly
Associated features	Nausea in 80%, vomiting in 30%, light and noise sensitivity in 80%	Occasional nausea	Ipsilateral conjunctival injection, tearing, and nasal congestion or drainage; ptosis and miosis in 30%
Behavior during headache	Still, quiet, tries to sleep	No change	Often paces, agitated
Awakens patient from sleep	Can occur	Rare	Frequently

discharge in this pain-control system may occur from stimulation from the cortex or hypothalamus caused by stress or by excessive afferent input from the special senses or from cerebral or extracranial vessels. The migraine prodrome may originate in the hypothalamus.

CLINICAL FEATURES

Migraine can occur with or without an aura [*see* Migraine Aura, *below*]. Migraine without aura (formerly referred to as common migraine) occurs in 80% of migraineurs, and migraine with aura (formerly referred to as classic migraine) occurs in 20%. Most patients who have migraine with aura also have migraine without aura.

According to the IHS criteria for migraine without aura, the duration of untreated or unsuccessfully treated episodes ranges from 4 to 72 hours. The headaches are associated with at least two of the following pain characteristics: unilateral location; pulsating quality; moderate or severe intensity; and aggravation by, or resultant avoidance of, routine physical activity (e.g., walking or climbing stairs). The pain is accompanied by nausea, vomiting, or both, as well as by sensitivity to light (photophobia) and sound (phonophobia). Also, the patient has a history of at least five previous attacks that meet these criteria. If there are no indications that other primary etiologies may be responsible for the headaches, a diagnosis of migraine without aura can be reasonably established.[1]

Although the IHS criteria have been very useful for research purposes, most clinicians recognize migraine through familiarity with the general features [*see Table 4*]. Migraine pain is unilateral in 60% of cases and bilateral in 40%. About 15% of migraineurs report so-called side-locked headaches, with migraine always occurring on the same side. The pain will often be more intense in the frontotemporal and ocular regions before spreading to the parietal and occipital areas. Any region of the head or face may be affected, including the parietal region, the upper or lower jaw or teeth, the malar eminence, and the upper anterior neck. Throbbing pain is present in 85% of episodes of migraine, although up to 50% of patients describe nonthrobbing pain during some attacks. Along with having head pain, up to 75% of migraineurs report having unilateral or bilateral tightness, stiffness, or throbbing pain in the posterior neck. The neck pain can occur during the migraine prodrome, the attack itself, or the postdrome and is typically relieved by migraine medication such as a triptan.

Migraine persisting for more than 72 hours is termed status migrainosus. Without treatment, 80% of patients have moderate to severe pain and 20% have mild pain. The pain, which is usually increased by physical activity or movement, is associated with nausea in about 80% of episodes, vomiting in about 30%, photophobia in about 90%, and phonophobia in about 80%.[10] In children, migraine pain is bilateral in 60% and unilateral in 40%. The duration of the untreated headache in children can be 1 hour or more, much shorter than that in adults.

During an attack, 45% of migraineurs have at least one autonomic symptom (i.e., lacrimation, eye redness, ptosis, eyelid edema, nasal congestion, or rhinorrhea). These symptoms are caused by parasympathetic activation of the sphenopalantine ganglion, which innervates the tear ducts and sinuses, and these symptoms can lead to confusion of migraine with so-called sinus headaches.[15] Of patients with autonomic symptoms, 45% have both nasal and ocular symptoms, 21% have nasal symptoms only, and 34% have ocular symptoms only.

Prodromal symptoms (premonitory phenomena) may be present in about 10% of cases and precede the migraine attack by hours or by up to 1 or 2 days. Prodromal symptoms include changes in mental state such as depression, hyperactivity, euphoria, talkativeness, irritability, drowsiness, and restlessness. Neurologic symptoms may include photophobia, difficulty concentrating, phonophobia, dysphasia, hyperosmia, and yawning. General symptoms may include stiff neck, food cravings, feeling cold, anorexia, sluggishness, diarrhea or constipation, thirst, and fluid retention.

Triggers

Migraines are often triggered by environmental or other factors; 85% of migraineurs report triggers. Patients typically have multiple triggers, with a mean of three.[16] Up to 50% of migraineurs report that a change of weather is a trigger. Other environmental triggers are heat, high humidity, and high altitude. There are numerous additional triggers, including stress (reported by about 50% of patients), letdown after stress, vacations, and crying. Missing a meal (40%), lack of sleep, oversleeping, and fatigue are also commonly reported as triggers. Sensory triggers include bright lights, glare, flickering lights, loud noise, and strong smells such as perfume or cigarette

Table 5 Features of Selected Secondary Headaches[6]

Headache Type	*Epidemiology*	*Age of Onset*	*Location*	*Quality and Severity*	*Frequency*	*Associated Features*	*Comments*
Trigeminal neuralgia	4.3/100,000/yr; male-to-female ratio, 1.6:1	Usually > 40 yr; if < 40 yr, consider multiple sclerosis	Unilateral, 96%; second or third trigeminal division more often than first	Stabbing; electrical bursts; burning; lasts few seconds to < 2 min	Few to many a day	Trigger zone present in > 90% of cases	Usually due to vascular compression of CN V; scan needed to exclude occasional tumor
Brain tumor	Persons/yr: 24,000 primary, 170,000 metastatic	Any age	Often bifrontal; unilateral or bilateral; any location	Can be pressure or throbbing, mild to severe	Occasional to daily; usually progressive	Papilledema in 40%; at time of diagnosis, headache present in 30%–70%	Primaries in adults: lung, 64%; breast, 14%; unknown, 8%; melanoma, 4%; colorectal, 3%; hypernephroma, 2%
Pseudotumor cerebri	1/100,000/yr; 90% are female; 90% are obese	Mean of 30 yr	Often bifrontotemporal but can occur in other locations and unilaterally	Pulsatile; moderate to severe	Daily	Papilledema in 95%; transient visual obscurations in 70%; intracranial noises in 60%; CN VI palsy in 20%	MRI scan preferred to better exclude cortical venous thrombosis and posterior fossa lesions
Subarachnoid hemorrhage	30,000/yr caused by saccular aneurysm	Mean of 50 yr	Usually bilateral; any location	Usually severe but can be mild and gradually increasing	Paroxysmal	Often with nausea, vomiting, stiff neck, focal findings, syncope; stiff neck absent in 36%	CT scan abnormal on first day in 95%; third day, 74%; 1 wk, 50%; lumbar puncture may be essential for diagnosis
Temporal arteritis	In age > 50 yr, annual incidence of 18/100,000; male-to-female ratio, 3:1	Rare before 50 yr; mean age of 70 yr	Variable, unilateral, or bilateral; often temporofrontal	Often throbbing; may be sharp, dull, burning, or lancinating; mild to severe	Intermittent to continuous	50% have PMR; jaw claudication in 38%; 50% have absent pulse or tender STA	ESR WNL in up to 36%; CRP usually elevated; STA biopsy false negative in up to 44%
Acute paranasal sinusitis	More common in children (in whom frontal and sphenoid sinusitis are rare) than in adults	Any age	Frontal (forehead), maxillary (cheek), ethmoid (between eyes), sphenoid (variable)	Dull, aching; can be severe	Acute lasts from 1 day to 3 wk	Fever in about 50%; nasal congestion and purulent nasal drainage usually present (less often in sphenoid)	Well visualized on routine MRI but not on routine head CT scan; sinus CT is the best study
Subdural hematoma	Occurs in 1% after mild head injury; in chronic cases, up to 50% without history of head injury	Any age	Unilateral or bilateral	Mild to severe; may be aching, dull, or throbbing	Paroxysmal to constant	Normal neurologic exam in 50%; alteration in consciousness and focal findings may be present	MRI may detect the occasional isodense subdural hematoma, which can be missed on CT scan

CN—cranial nerve CRP—C-reactive protein ESR—erythrocyte sedimentation rate MRI—magnetic resonance imaging PMR—polymyalgia rheumatica STA—superficial temporal artery WNL—within normal limits

Table 6 Reasons to Consider Neuroimaging for Headaches[6]

Temporal and clinical features	First or worst headache Subacute headaches with increasing frequency or severity A progressive or new daily persistent headache Chronic daily headache Headaches always on the same side Headaches not responding to treatment
Patient characteristics	New-onset headaches in patients with cancer or HIV infection New-onset headaches after age 50
Associated symptoms and signs	Fever, stiff neck, nausea, and vomiting Aura and focal neurologic symptoms or signs in nonmigraine headache Papilledema, cognitive impairment, or personality change Seizures

smoke. Up to 50% of patients report alcohol as a trigger; this can be all forms of alcohol or only one type, such as red wine or beer. Up to 45% report food triggers such as chocolate, dairy products (particularly cheese), citrus fruits, fried foods, and nitrates and nitrites in cured meats or fish (e.g., frankfurters, bacon, and smoked salmon). Other triggers include minor head trauma, exertion, and nitroglycerin.

There are triggers unique to women. Half of women with migraine report menses as a trigger, and 14% have migraines associated only with their menses. During pregnancy, the frequency of migraines decreases (especially during the second and third trimesters) in 60%, remains the same in 20%, and increases in 20%. Migraines may occur for the first time when women start using oral contraceptives (OCs). Low-estrogen OCs usually have no effect on migraine or may even improve it, although frequency can increase. Of patients with new-onset migraine or increased frequency of migraine associated with OCs, 30% to 40% may improve when OCs are discontinued, although improvement may not occur for up to 1 year. Two thirds of women with prior migraine improve with physiologic menopause. Surgical menopause results in worsening of migraine in two thirds of cases.

Migraine Aura

The migraine aura has a total duration of usually less than 1 hour and frequently less than 30 minutes. An aura lasting more than 1 hour but less than 1 week is termed migraine with prolonged aura, or complicated migraine. The most common aura is a vision-related one, which is present in 99% of cases. There are two types: (1) positive visual phenomena, with hallucinations, and (2) negative visual phenomona, or scotomas, with either an incomplete or a complete loss of vision in a portion or all of the visual field. Most visual auras have a hemianoptic distribution. Photopsias consist of small spots, dots, stars, unformed flashes or streaks of light, or simple geometric forms and patterns that typically flicker or sparkle.

A scintillating scotoma, also called a fortification spectrum (because of its resemblance to a medieval fortified town as viewed from above) or teichopsia (seeing fortifications), is present in about 10% of cases. The scotoma, which is frequently semicircular or horseshoe shaped, usually begins in the center of the visual field and then slowly extends laterally. The scotomatous arc or band is a shimmering or glittering, bright, zigzag border. Most visual auras consist of flickering, colored or uncolored, unilateral or bilateral zigzag lines or patterns, semicircular or arcuate patterns, wavy lines, or irregular patterns. Rare visual auras include metamorphopsia (objects appear to change in size and shape), macropsia, micropsia, telescopic vision (objects appear larger than normal), teleopsia (objects appear to be far away), mosaic vision, Alice in Wonderland syndrome (distorted body image), and multiple images. Headaches, when unilateral, usually occur on the side contralateral to the visual symptoms but can occasionally be ipsilateral.

A sensory aura, which is present in about 30% of episodes of migraine with aura, consists of numbness, tingling, or a pins-and-needles sensation. The aura, which is usually unilateral, commonly affects the hand and then the face, or it may affect either one alone. Paresthesias of one side of the tongue is typical. Less often, the leg and trunk may be involved. A true motor aura is rare, but sensory ataxia or a heavy feeling is often misinterpreted as weakness.

Speech and language disturbances may occur in up to 20% of cases. Patients often report a speech disturbance when the spreading paresthesias reach the face or tongue. Slurred speech may be present. With involvement of the dominant hemisphere, paraphasic errors and other types of impaired language production and comprehension may occur. Rarely, other aura symptoms may be described, including déjà vu and olfactory and gustatory hallucinations.

Although visual symptoms frequently occur by themselves, combinations of aura symptoms can occur. Sensory, speech, and motor symptoms are usually associated with visual symptoms or with one or more other symptoms. When two or more aura symptoms are present, they almost always occur in succession rather than simultaneously.

Migraine aura can occur without headache (acephalgic migraine), often in patients whose migraine episodes typically involve headache (with or without aura). A visual aura is the most common in such cases. Another type of acephalgic migraine is episodic vertigo without a headache, auditory disturbances, or other neurologic symptoms, lasting minutes to days.[17] In older persons, the aura—termed late-life migraine accompaniment—can be confused with a transient ischemic attack [*see* Geriatric Headache, *below*]. Rarely, migraineurs have persistent visual aura. This usually consists of simple, unformed hallucinations in the entire visual field of both eyes, including innumerable dots, television static, clouds, heat waves, flashing or flickering lights, lines of ants, a rainlike or snowlike pattern, squiggles, bubbles, and grainy vision. Occasionally, palinopsia (the persistence of visual images), micropsia, or formed hallucinations occur.

Migraine Variants

Migraine variants include familial hemiplegic; basilar-type; benign paroxysmal vertigo of childhood; abdominal; confusional; so-called footballer's; benign episodic mydriasis; and retinal. Familial hemiplegic migraine is a rare variant of migraine with aura accompanied by hemiplegia or hemiparesis. Attacks may occur on the same side as previous episodes or on another side and typically feature a slow spread of paresis involving the face, arm, and leg. Alteration of consciousness, ranging from confusion to coma and aphasia, may be present. Familial hemiplegic migraine type I is caused by an autosomal dominant mutation in a brain-specific P/Q-type calcium channel subunit on chromosome 19.

Basilar-type migraine is a rare disorder that most often occurs in children and rarely occurs in patients older than 50 years.[18] According to IHS criteria, attacks are marked by two or more of the following fully reversible aura symptoms: dysarthria, vertigo, tinnitus, hypacusia, diplopia, visual symptoms simultaneously in the temporal and nasal fields of the two eyes, ataxia, decreased level of consciousness, and simultaneous bilateral paresthesias. These symptoms, which originate from the brain stem or both occipital lobes, are not accompanied by motor weakness. There is also at least one of the following: bilateral paresthesias; gradual development of at least one aura symptom over 5 minutes or longer, the occurrence of different aura symptoms in succession over 5 minutes or longer, or both; or persistence of each aura symptom for 5 to 60 minutes. Patients with basilar-type migraine may also have other types of migraine. Visual symptoms—which usually take the form of blurred vision, shimmering colored lights accompanied by blank spots in the visual field, scintillating scotoma, and graying of vision—may start in one visual field and then spread to become bilateral. Diplopia occurs in up to 16% of cases. Vertigo may be present, either alone or accompanied by various combinations of tinnitus, dysarthria, gait ataxia, and paresthesias (usually bilateral but sometimes affecting alternate sides in successive episodes). Impairment of consciousness is common and may include obtundation, amnesia, syncope, and, rarely, prolonged coma. A severe throbbing headache, typically with a bilateral occipital location, is present in 96% of cases. Nausea and vomiting typically occur, with light and noise sensitivity in up to 50% of cases.

Benign paroxysmal vertigo of childhood presents as episodes of vertigo without headache. Abdominal migraine also occurs in children and features recurring episodes of abdominal pain without headache that may be associated with nausea, vomiting, pallor, and flushing. Confusional migraine presents with a headache, which can be minimal, associated with a confusional state that can last from 10 minutes to 2 days. Agitation and impaired memory may be present. The patient may exhibit inattention, distractibility, and difficulty maintaining coherent speech or action. So-called footballer's migraine (originally described in soccer players) refers to the triggering of migraine by acute minor head trauma in children or adolescents.

Benign episodic mydriasis is a transient, isolated mydriasis. This disorder typically occurs in young adults or children. Patients have normal vision and pupillary reactivity to light that may occasionally accompany migraine headaches. The episodes last 15 minutes to 24 hours, are often associated with blurred vision, and can average two or three a month. Eyelid or ocular motility abnormalities are absent. Angle-closure glaucoma should be excluded. Dilatation of the pupil is secondary either to parasympathetic insufficiency of the iris sphincter or to sympathetic hyperactivity of the iris dilator. Retinal migraine, a rare diagnosis of exclusion, produces episodes of transient monocular visual loss lasting minutes to hours, which may or may not be associated with headache.

ACUTE TREATMENT

Certain general principles apply to the use of medications for acute (symptomatic) treatment of migraine. Early treatment, when the headache is mild, is much more effective than later treatment, when the migraine is moderate or severe in intensity. Frequent use of acute-treatment medications can lead to rebound headache; for that reason, acute therapy should be restricted to a maximum of 2 or 3 days a week. Different patients may respond to different medications at different times. Patients benefit from stratified care.[19] Treatment is based on characteristics of the patient's episodes (including peak intensity, time to peak intensity, associated symptoms, and disability) and is tailored to specific patient needs. Nasal, parenteral, or rectal administration of medication should be used in patients with significant nausea or vomiting or gastroparesis. Antinausea medications such as promethazine and prochlorperazine may help in such cases. Many migraineurs respond to over-the-counter medications for acute symptoms [*see Table 7*]. Over-the-counter drugs are often more effective if taken when the pain is mild rather than when it has become more intense.

Patients who do not respond or who respond incompletely to over-the-counter medications may require a prescription medication. The combination of isometheptene mucate, dichloralphenazone, and acetaminophen (Midrin) can be highly effective, especially for mild to moderate headache.[20] Combination medications that include butalbital (e.g., Fiorinal) may also be effective. Surprisingly, despite their common use, the butalbital combinations have not been studied in a placebo-controlled trial.[21] Oral and intranasal opiates may also be effective. Butorphanol nasal spray (Stadol) may be very effective for some patients who have severe migraine with nausea or vomiting and cannot keep an oral medication down, do not respond to triptans, or require a rescue medication when the usual medication is ineffective. Side effects include dizziness, nausea, vomiting, and drowsiness. However, frequent use of medications such as

Table 7 Efficacy of Selected Over-the-Counter Medications for Relief of Moderate to Severe Migraine Pain[75-78]

Medication and Dose	*Percentage of Responders at 2 Hr (Percentage Placebo)*		*Percentage of Responders at 6 Hr (Percentage Placebo)*	
	Mild or No Pain	*No Pain*	*Mild or No Pain*	*No Pain*
Ibuprofen, 400 mg	42 (28)	15 (8)	49 (32)	31 (20)
Acetaminophen, 1,000 mg	58 (39)	22 (11)	77 (46)	46 (28)
Acetaminophen, 500 mg, plus aspirin, 500 mg, plus caffeine, 130 mg	59 (33)	22 (7)	79 (52)	51 (23)
Aspirin, effervescent, 1,000 mg	55 (37)	29 (17)	Not assessed	

Note: Percentages are rounded.

butalbital and opiates can lead to rebound headaches and habituation.

Triptans

The introduction of the triptans has dramatically improved the acute treatment of migraine. Triptan medications are selective 5-hydroxytryptamine ($5\text{-}HT_{1B/1D}$) receptor agonists that share a basic indole ring structure with different side chains. Triptans have three potential mechanisms of action: cranial vasoconstriction, peripheral neuronal inhibition, and inhibition of transmission through second-order neurons of the trigeminocervical complex. These mechanisms inhibit the effects of activated nociceptive trigeminal afferents and control acute migraine attacks.[14]

Over the past decade, seven triptans have become available in the United States: sumatriptan, zolmitriptan, naratriptan, rizatriptan, almotriptan, frovatriptan, and eletriptan [*see Table 8*].[22] In migraineurs who take oral triptans when their pain is moderate to severe in intensity, the 2-hour response rate (i.e., no pain or mild pain) is about 45% for naratriptan and frovatriptan and about 65% to 70% for the others. With all of the triptans, the 2-hour pain-free response rates are much higher if the drug is taken when the headache is mild; depending on the drug, the response rate may exceed 70%.

The oral triptans may not be equally effective for all patients. If a patient has an unsatisfactory or inconsistent response, unpleasant side effects, or tachyphylaxis with one triptan, a different triptan may prove to be effective and tolerable. Patients who have prominent vomiting or nausea or who desire the quickest relief may benefit from subcutaneous sumatriptan (at 2 hours, 79% of patients show a response and 60% are pain free) or intranasal sumatriptan or zolmitriptan.

Patients may experience a recurrence, which is defined as the return of headache (usually of moderate or severe intensity) within 24 hours after an initial response to acute treatment. When taken for moderate to severe pain, naratriptan, frovatriptan, almotriptan, and eletriptan have the lowest recurrence rates (about 20% to 25%). The recurrence rates for the other triptans are about 31% with zolmitriptan, 33% with sumatriptan, and 40% with rizatriptan. The time to recurrence is generally about 12 hours.[23]

Triptans can stimulate $5\text{-}HT_{1B}$ receptors on coronary arteries and result in constriction, which may become clinically significant in patients with coronary artery stenosis or vasospastic disease.[24] For example, in a study of patients undergoing cardiac catheterization, 6 mg of subcutaneously administered sumatriptan resulted in a 14% reduction in coronary artery diameter.[25] Consequently, triptans as a class are contraindicated in patients with known or suspected ischemic heart disease, Prinzmetal angina, or uncontrolled hypertension. Data on cardiac risk have been derived from the use of sumatriptan because sumatriptan has been the most widely used triptan and has the largest reported studies; however, all the triptans have a similar cardiac risk. The common triptan side effects—tightness, heaviness, pressure, or pain in the chest, neck, or throat—are not associated with electrocardiogram changes and are not caused by coronary vasoconstriction.[26]

The estimated chance of a myocardial infarction occurring within 24 hours of oral or subcutaneous use of sumatriptan is small but not negligible. Between 1991 and December 1996, 19 cardiac-related deaths occurring within 24 hours after the last sumatriptan dose were reported worldwide in five million migraineurs treating more than 100 million attacks.[26] In addition, there have been seven reported cases of atrial fibrillation triggered by sumatriptan,[27] possibly caused by transient elevations in atrial pressure.[28] In a prospective study of 23,339 migraineurs who used 6 mg of sumatriptan subcutaneously for up to 12 months, a total of 185,579 migraine attacks occurred.[29] There were a total of 13 cardiac events (three myocardial infarctions, six anginal episodes, and four cases of dysrhythmia) occurring 24 hours or more after the administration of sumatriptan. Subcutaneously administered sumatriptan is eliminated from the body within 10 hours, so the late cardiac events were probably unrelated to the drug. However, only about 15% of the patients in this study were older than 50 years and only a small percentage were older than 60 years.

In view of the potential adverse cardiovascular events associated with triptans, a cardiac evaluation is recommended for patients at risk for unrecognized coronary artery disease. The evaluation can be done before or during the use of triptans. Patients at risk include men older than 40 years, women older than 50 years, and those with cardiac risk factors. There is no consensus, however, on what constitutes an appropriate cardiac evaluation.[30]

INTRACTABLE MIGRAINE AND MIGRAINE STATUS

Intravenous fluids and electrolyte replacement may be necessary for patients with intractable vomiting associated with migraine. Medication options include the following:

1. Sumatriptan, 6 mg subcutaneously.
2. Dihydroergotamine (DHE), 0.5 to 1 mg by slow intravenous push, perhaps combined with an antiemetic such as metoclopramide, because DHE may cause nausea (DHE and triptans should not be used within 24 hours of each other).
3. Prochlorperazine, 5 to 10 mg intravenously.
4. Ketorolac, 30 to 60 mg intramuscularly.
5. Corticosteroids (a single or rapidly tapering dose of prednisone, starting at 80 mg a day, or dexamethasone, 6 mg orally or I.V.).

Table 8 Triptans Available in the United States

Drug (Brand Name)	*Formulation*	*Strengths (mg)*
Almotriptan (Axert)	Tablets	12.5
Eletriptan (Relpax)	Tablets	20, 40
Frovatriptan (Frova)	Tablets	2.5
Naratriptan (Amerge)	Tablets	1, 2.5
Rizatriptan (Maxalt)	Tablets Orally disintegrating preparation* (Maxalt MLT)	5, 10
Sumatriptan (Imitrex)	Subcutaneous injection Tablets Nasal spray	6 25, 50, 100 5, 20
Zolmitriptan (Zomig)	Tablets Orally disintegrating preparation* (Zomig ZMT)	2.5, 5

*Dissolves on the tongue; can be taken without water (efficacy similar to that of tablet form).

Table 9 Preventive Medications for Migraine

Drug Class	*Agent*	*Dosage*	*Typical Side Effects*
Beta blockers	Propranolol*	40–120 mg b.i.d.	Hypotension, tiredness, exacerbation of asthma
	Propranolol long acting*	60–160 mg/day	
	Metoprolol	50–200 mg/day	
	Nadolol	40–160 mg/day	
	Atenolol	50–100 mg/day	
	Timolol	10–30 mg/day	
Antidepressants	Amitriptyline*	25–150 mg h.s.	Drowsiness, dry mouth, weight gain, constipation
	Nortriptyline	25–150 mg h.s.	
	Venlafaxine	37.5–225 mg/day	Nausea and vomiting
Anticonvulsants	Divalproex sodium*	500–1,000 mg/day	Nausea, tremor, drowsiness, weight gain, alopecia, hematologic and liver abnormalities, fetal abnormalities
	Topiramate*	25–300 mg h.s.	Weight loss, paresthesias, cognitive disturbances, kidney stones
	Gabapentin	300–800 mg t.i.d.	Dizziness, fatigue, drowsiness

*Class I evidence indicates that these are the most effective medications for migraine prevention.

6. Parenteral narcotics such as meperidine, which may be combined with promethazine.
7. Valproate sodium (500 mg diluted in 50 ml of saline, administered intravenously over 5 to 10 minutes and repeated every 8 hours, if necessary).
8. Droperidol (2.5 mg I.M. or I.V.).
9. Intravenous metoclopramide.

There is a very small risk of torsade de pointes with the use of neuroleptics such as prochlorperazine and droperidol.

PREVENTIVE TREATMENT

A number of factors may justify daily preventive medication for patients with migraines [*see Table 9*].[31] Indications for preventive treatment are as follows:

1. The headaches significantly interfere with the patient's daily routine, despite acute treatment.
2. Acute medications are contraindicated, ineffective, or overused or have intolerable side effects.
3. Frequent migraines (two or more attacks a week).
4. Uncommon migraine types (hemiplegic, basilar, prolonged aura, or migrainous infarction).
5. The cost of acute medications is significantly greater than the cost of preventive medication.
6. Patient preference (i.e., the patient is willing to risk the possibility of side effects from the preventive medication to reduce the frequency of headaches).

Several general principles apply to the use of preventive medications[31]:

1. The clinician should start with a low dose of medication and increase it slowly, depending on the response and whether side effects occur.
2. Each medication should be given a trial of 2 to 3 months at adequate doses.
3. Overused medications that may be causing rebound headache and may decrease the efficacy of preventive treatment should be discontinued or tapered (depending on the drug).
4. The patient should keep a headache diary to monitor his or her headaches.
5. The clinician should educate the patient about the rationale for treatment and possible side effects and should address the patient's expectations for treatment. Many patients want a complete cure, and although this is certainly understandable, it is usually not possible.

Coexistent or comorbid conditions should be considered. Some medications may be effective against both migraine and another disorder. Other disorders, along with the migraine medications that may be effective against them, include epilepsy (divalproex sodium, topiramate, and gabapentin), hypertension (beta blockers), depression (tricyclic antidepressants), bipolar disorder (divalproex sodium or topiramate), insomnia (tricyclic antidepressants), essential tremor (beta blockers and topirimate), and overweight or obesity (topiramate). On the other hand, coexistent diseases such as depression or asthma may be relative contraindications to the use of beta blockers. In a woman who is pregnant or may become pregnant, the potential for teratogenesis should be considered. Patients who have mild responses to one preventive agent may benefit from the addition of a second agent. Finally, if a medication does not work or has significant side effects, withdrawal of the agent may need to be done slowly, especially if the patient has been receiving moderate or high doses of the drug. This is particularly true of tricyclic antidepressants and beta blockers.

Class I evidence indicates that the beta blocker propranolol, the tricyclic antidepressant amitriptyline, and the antiseizure medications divalproex sodium and topiramate are the most effective preventive medications, reducing the frequency of migraines by more than 50% in about 50% of patients. In general, preventive medications are more effective when patients are placed on a titration schedule with a minimum target dose. Titration schedules and minimum target doses are as follows:

propranolol (either regular or long acting), 40 mg daily, increased weekly by 40 mg to a maximum daily dose of 120 to 160 mg; amitriptyline, 10 mg at bedtime, increased weekly by 10 mg to a maximum daily dose of 50 mg; divalproex sodium (either regular or extended release), 500 mg daily for 1 week and then 1,000 mg daily; and topiramate, 25 mg daily for the first week, increased by 25 mg/wk in divided doses, to a maximum daily dose of 100 mg administered at a dosage of 50 mg twice daily.[32]

In addition to propranolol, other beta blockers may be effective [*see Table 9*]. Regarding the tricyclic antidepressants, the quality of evidence for nortriptyline is not as good as that for amitriptyline, but nortriptyline has been shown to have similar efficacy with less sedation. Venlafaxine may be as effective as amitriptyline with fewer side effects.[33,34] Selective serotonin reuptake inhibitors (SSRIs) are probably not effective for migraine prevention, and verapamil and gabapentin are only modestly effective.[31]

There are natural products that may be beneficial for migraine prevention, including the herb feverfew (*Tanacetum parthenium*); extract from the butterbur plant, *Petasites hybridus* (Petadolex, 75 mg twice daily); riboflavin (400 mg a day); coenzyme Q10 (100 mg three times daily[35]); and oral magnesium supplements. Botulinum toxin injections may also be of benefit, especially in intractable cases. The relative benefit of these treatments may become clearer with additional studies, but for now, some migraineurs may prefer them because they have few if any side effects.

For many migraineurs, the avoidance of triggers may be useful. Examples include adequate sleep at set hours, routine exercise, regular meals, avoiding triggering foods and beverages, and wearing sunglasses in bright sunlight or glare. Some patients may benefit from biofeedback, relaxation training, and psychotherapy.

WOMEN AND MIGRAINE

There are issues specific to treatment of female migraineurs.[36] Menstrual migraine is treated with the same acute-treatment medications as other migraines (see above). Interval or short-term preventive treatment of menstrual migraine, starting 2 or 3 days before menses and continuing during the menses, may be helpful for some women with regular menses and migraines that are poorly responsive to symptomatic medications. Potentially effective medications include the following: amitriptyline or nortriptyline, 25 mg at bedtime; long-acting propranolol, 60 to 80 mg daily, or nadolol, 40 mg daily; nonsteroidal anti-inflammatory drugs (NSAIDS) such as naproxen sodium, 550 mg twice daily; ergotamine, 1 mg once or twice a day, or DHE, 1 mg subcutaneously or intramuscularly; naratriptan, 1 mg orally twice daily, or frovatriptan, 2.5 mg twice daily, for 6 days perimenstrually; transdermal estradiol, 100 µg applied 3 days before the expected start of menses and replaced after 3 days; continuous combined OC use, with a lower estrogen dose given during the menses; and extended-duration OC use.

Although there is controversy regarding whether low-estrogen OCs increase the risk of stroke, most women who have migraine without aura can safely take low-estrogen OCs if they have no other contraindications or risk factors. When taking low-estrogen OCs, women younger than 35 years who have migraine with aura (e.g., visual symptoms lasting less than 1 hour) have a risk of ischemic stroke of about 30 per 100,000 annually, which is twice the risk of those women who have migraine without aura.[37] An IHS task force concluded that OCs may be contraindicated in women with migraine who have additional risk factors that cannot easily be controlled, including migraine with aura, because of a possible increase in the risk of ischemic stroke, and that these risks must be assessed and evaluated on an individual basis.[38] Women with aura symptoms such as hemiparesis or aphasia or prolonged focal neurologic symptoms and signs lasting more than 1 hour should avoid starting low-estrogen OCs and should stop the medication if they are already taking it. Progestin-only OCs and the many other contraceptive options can be considered, as appropriate. Cigarette smoking should be strongly discouraged, because female migraineurs who smoke one or more packs of cigarettes a day raise their risk of ischemic stroke by a factor of about 10.

Estrogen replacement therapy has a variable effect on migraine: 45% of patients show improvement, 46% show worsening of migraine, and 9% show no effect. If migraines increase when a patient starts estrogen replacement, the following strategies may be beneficial:

1. Reduce the estrogen dose.
2. Change the estrogen type to one less likely to promote migraine. From most to least likely to promote migraine, these are, in order, conjugated estrogens (Premarin), pure estradiol (Estrace), synthetic estrogen (Estinyl), and pure estrogen (Ogen).
3. Convert from interrupted to continuous dosing in the case of estrogen-withdrawal migraine.
4. Convert from oral to parenteral administration (e.g., a transdermal patch).
5. Add androgens.

Management of migraine during pregnancy and breastfeeding[39] is beyond the scope of this chapter.

Tension-Type Headaches

The 1-year prevalence of tension-type headaches has been variably reported as being from 30% to 90%. The lifetime prevalence is 78% (63% in males and 86% in females; the male-to-female ratio is about 1:1.3). The prevalence peaks in the fourth decade of life.

CLINICAL FEATURES

Tension-type headache may be episodic or chronic. The IHS criteria for episodic tension-type headache are as follows: at least 10 previous headache episodes fulfilling the criteria; number of days with the headache being less than 180 a year or 15 a month; and headache lasting from 30 minutes to 7 days. At least two of the following pain characteristics should be present: pressing/tightening (nonpulsating quality); mild or moderate severity; bilateral location; and no aggravation of headache by walking up and down stairs or performing similar types of routine physical activity. There should be no nausea or vomiting (anorexia may occur); and either photophobia or phonophobia may be present, but not both.

The pain is variably described as pressure, soreness, tightness, a band or cap on the head, or a weight on the head. During severe episodes, a pulsating sensation is occasionally present. Although 90% of episodic tension-type headaches are bilateral, these headaches can be unilateral in patients with trigger points or oromandibular dysfunction.

In chronic tension-type headache, according to IHS criteria,

the average headache frequency is 15 days or more a month for at least 6 months, or 180 days or more a year. The pain characteristics are the same as for episodic tension-type. There should be no concomitant vomiting, and no more than one of the following features should be present: nausea, photophobia, or phonophobia. Some patients may have continuous headaches for years. Secondary causes of episodic and chronic tension-type headaches should be excluded, as appropriate.

TREATMENT

Acute headaches may respond to the following: aspirin or acetaminophen, alone or in combination with caffeine; NSAIDs; isometheptene in combination with other agents; and butalbital with other agents. Overuse of any of these medications, however, may lead to rebound headaches. Frequent butalbital use can also result in dependency. The muscle relaxants baclofen and tizanidine may also be effective and are not habituating, whereas the muscle relaxant carisoprodol can be habituating. Tizanidine is an α_2-adrenergic agonist that inhibits the release and effectiveness of norepinephrine both at central sites (e.g., the locus coeruleus) and at the spinal cord. It has central muscle relaxant and antinociceptive effects. Tizanidine can be given at a dosage of 2 mg three times a day, or it can be started as 2 mg at bedtime and titrated upward to the maximum tolerated dose or a maximum dosage of 18 mg in three divided doses daily, depending on the response. Because about 5% of patients on tizanidine develop abnormally elevated transaminase levels, which reverse after discontinuance of the drug, baseline measures and periodic monitoring of liver function for the first 6 months are recommended. Tizanidine may be effective for chronic tension-type and chronic daily headaches.

Frequent headaches may require preventive medications. Tricyclic antidepressants are generally more effective than SSRIs. Other migraine preventive agents (see above) may be helpful, especially when tension-type headache and migraine are both present.

Chronic Daily Headache

Chronic daily headache (CDH) has a frequency of 15 or more days a month. The 1-year prevalence of CDH in adults is about 3% in males and 5% in females; it is about 1% in adolescents of both sexes. Severe CHD affects 0.5% of the population of the United States.

HEADACHE TYPES IN CDH

CDH includes four different headache types: chronic, or transformed, migraine (35% of patients with CDH); chronic tension-type headache, occurring in more than 50% of patients with CDH; hemicrania continua; and new daily persistent headache.

Chronic migraine, or transformed migraine, is a complication of intermittent migraine that usually occurs by 20 to 30 years of age. It may occur with or without medication overuse. In 70% of patients, there is a gradual transformation from episodic migraine to CDH that may be associated with analgesic overuse, psychological factors (e.g., depression, anxiety, abnormal personality profile, and home or work stress), and obesity. In 30% of patients, there is a sudden transformation that may be triggered by head or neck trauma, flulike illness, aseptic meningitis, surgery, or medical illness. Migraine characteristics are present to a significant degree, intermittently or continuously.

Chronic tension-type headache, with or without medication overuse, occurs 15 days a month for at least 6 months or occurs at least 180 days a year [*see* Tension-Type Headaches, *above*]. Hemicrania continua, with or without medication overuse, is a rare entity with constant, unilateral pain of variable intensity that responds dramatically to indomethacin. Painful exacerbations are associated with ptosis, lacrimation, and nasal congestion. New daily persistent headache, with or without medication overuse, involves a fairly rapid onset of a daily persistent headache in a patient with no past history of increasingly frequent migraine or tension-type headache.[40] This is probably a heterogeneous disorder of uncertain cause, which in some cases may be triggered by a viral infection.

TREATMENT

If medication-overuse or rebound headaches (see below) are a possibility, medications that may be responsible should be tapered. Some acute-treatment medications that may be effective are longer-acting NSAIDs (e.g., naproxen sodium), baclofen, tizanidine, and hydroxyzine (50 mg p.o., t.i.d., p.r.n.), which are not associated with rebound. Triptans may be used as appropriate but should be limited to 2 or 3 days a week because of the risk of rebound.

For prevention of CDH, the same medications are used as for chronic tension-type headache and migraine. Combination therapy may be helpful in some cases. The effect of treatment may not be apparent for weeks. Treatment may not be effective until rebound is eliminated.

For detoxification or if there is significant medical or psychiatric comorbidity, inpatient treatment may be indicated if outpatient therapy fails. Options include intravenous DHE (0.5 to 1.0 mg I.V.), usually given with an antiemetic (e.g., metoclopramide, 5 to 10 mg I.V.) every 8 hours. DHE may be combined with other medications, such as NSAIDs, oral or intravenous corticosteroids, intravenous prochlorperazine, and intravenous valproate sodium [*see* Intractable Migraine and Migraine Status, *above*]. One or more of these treatments can be used in patients who cannot tolerate DHE or patients in whom DHE is contraindicated.

Behavioral therapy and psychological or psychiatric referral may be beneficial. Physical therapy may be useful if there is a myofascial contribution to the headaches. Trigger-point injections and occipital nerve blocks may be worthwhile in some cases.

Even with optimal therapy, about one third of patients who show improvement will experience recurrence of their daily headache and medication-overuse pattern. Some patients have intractable CDH that is resistant to all treatments.

Medication-Overuse Headaches

Migraineurs are particularly susceptible to medication-overuse or rebound headaches, which can occur with frequent use of symptomatic medications, including acetaminophen, aspirin, caffeine, NSAIDs with short half-lives (e.g., ibuprofen), butalbital, ergotamine, opiate agonists, and triptans.[41] Frequent use of symptomatic medications may also result in tolerance (the decreased effectiveness of the same dose of an analgesic, often leading to the use of higher doses to achieve the same degree of effectiveness) and in habituation and dependence (respectively, the psychological and physical need to repeatedly use drugs).

Rebound headache is a retrospective diagnosis made when headache frequency decreases after the patient stops or reduces

the medication suspected of causing the headache. The best evidence is from a prospective study of caffeine-withdrawal headache in persons with low to moderate caffeine intake (the equivalent of about 2.5 cups of coffee daily). In this study, 50% of persons given placebo had a headache by day 2, compared with 6% of those given caffeine.[42] Withdrawal was also associated with nausea, depression, and flulike symptoms.

Because it would be unethical to conduct prospective studies on rebound headache from medication withdrawal, only limited information is available regarding the percentage of migraineurs who are susceptible to rebound, the dosage limits, and the time required for rebound to develop. Medication overuse may occur when simple analgesics are taken 15 or more days a month for 3 months; when triptans are taken 10 or more days a month; when combination analgesics containing simple analgesics plus opioids, butalbital, or caffeine (alone or together) are taken for 10 or more days a month for 3 months; and when opioids are used 10 or more days a month.[43]

In the treatment of suspected rebound headache, the medications acetaminophen, aspirin, NSAIDs with short half-lives, and triptans can be stopped abruptly. Caffeine use should be tapered off, to avoid withdrawal symptoms. Opiates and butalbital should be tapered because of the risk of a serious withdrawal syndrome. If butalbital is abruptly discontinued, phenobarbital can be substituted to prevent withdrawal; the phenobarbital is tapered down from 60 mg to 15 mg at night over 1 week.[44] After medication withdrawal, the duration of rebound headaches from triptans is about 4 days and from other analgesics is about 9 days.[45] A migraine preventive medication can also be started, but it may not be effective when patients are overusing symptomatic medications.

Two outpatient transitional strategies have been suggested to reduce the headaches during the withdrawal period. One approach is the use of prednisone: 60 mg/day for 2 days, 40 mg/day for 2 days, and then 20 mg/day for 2 days.[46] Alternatively, the combination of tizanidine and a long-acting NSAID such as naproxen may be effective.[47] Inpatient treatment is the same as for CDH (see above).

Drug-Induced Headache

Many drugs can induce acute headache, including nitroglycerin, antihypertensive agents (beta blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, and methyldopa), dipyridamole, hydralazine, phosphodiesterase-5 inhibitors (e.g., sildenafil), histamine receptor antagonists (e.g., cimetidine and ranitidine), NSAIDs (especially indomethacin), cyclosporine, and antibiotics (especially amphotericin, griseofulvin, tetracycline, and sulfonamides).

Drug-induced aseptic meningitis, a rare occurrence, has numerous possible causes, including NSAIDs, antibiotics (e.g., trimethoprim-sulfamethoxazole, sulfasalazine, cephalosporins, ciprofloxacin, isoniazid, and penicillin), intrathecal drugs and diagnostics (e.g., antineoplastic agents such as methotrexate and cytarabine; gentamicin; corticosteroids; spinal anesthesia; baclofen; repeated iophendylate for myelography; and radiolabeled albumin); intraventricular chemotherapy; intravenous immunoglobulin; vaccines (polio; measles, mumps, and rubella; and hepatitis B); and other drugs, such as carbamazepine, muromonab-CD3, and ranitidine.[48]

The clinical presentation of drug-induced aseptic meningitis is the same as that of viral meningitis. CSF findings are the same as those in viral meningitis, except for a neutrophil predominance; however, in cases induced by intravenous immunoglobulin, eosinophils are present.

Cluster Headaches

Cluster headaches are an uncommon headache type, occurring in only about 0.4% of the general population. Cluster headaches are five times more common in men than in women. The headaches can occur at any age, including childhood and adolescence (although they are rare in children younger than 10 years), but usually begin in the third or fourth decade of life. As the name denotes, this condition is marked by periods of recurrent headaches (one to eight a day) interspersed with periods of remission. In 90% of patients, the clusters occur episodically. In the remaining 10%, the clusters are chronic, with cluster periods lasting for more than 1 year without remission or with remission lasting less than 14 days.

CLINICAL FEATURES

Cluster headaches are one sided and severe. The most common types of pain, in order of decreasing frequency, are orbital, retro-orbital, temporal, supraorbital, and infraorbital. The headache may alternate sides between cluster periods or, rarely, within the same period. The pain is described as constant, boring, pressing, burning, or stabbing; about 30% of patients describe throbbing or pulsating pain. Cluster headaches have a rapid onset, with peak intensity in 5 to 10 minutes, and usually a short duration of 30 to 45 minutes, although a minority of patients may have pain persisting up to 3 hours (rarely longer). During attacks, most patients prefer to walk, sit, kneel, stand, or jog in place. Many find it difficult to lie down, and they feel restless and agitated.

Autonomic symptoms are present in over 97% of cases. Lacrimation and conjunctival injection are each present in about 80% of cases, and ipsilateral nasal congestion or clear drainage is present in 75%.[49] A partial Horner syndrome with a slight ipsilateral ptosis or miosis or a combination of both is present in about 65% of cases and, in some patients, may persist between attacks in later stages of the disorder. Increased forehead sweating may occur in some patients during attacks. Erythema of the eyelid or a circumscribed area of the face or forehead may be present. Nausea and sensitivity to light and noise accompany the headache in some patients. An aura, usually visual, occasionally precedes the headache. Small quantities of alcohol, nitroglycerin, and histamine can trigger attacks during cluster periods but not during remission.

DIAGNOSTIC TESTING

Cluster headaches can usually be diagnosed on the basis of the clinical criteria alone. Neuroimaging, preferably MRI, may be considered in cases with the following features: a pattern of clusterlike headache that does not conform to the clinical criteria; onset of cluster headache after age 40; a progressive pattern of headaches; chronic cluster headache; and any focal neurologic deficit other than Horner syndrome.

Symptomatic or secondary cluster headache can result from head trauma or iatrogenic trauma (e.g., orbital enucleation or dental extraction). A variety of pathologic conditions have been associated with clusterlike headaches, including arteriovenous malformations, aneurysms, sphenoid sinusitis, parasellar tumors, upper cervical cord meningioma and infarction, subdur-

al hematoma, cerebral metastases, and temporal arteritis. These headaches are usually atypical in their lack of periodicity or response to medications or their accompaniment with abnormal neurologic signs.

TREATMENT

For acute attacks of cluster headache, inhalation of 100% oxygen at a rate of 7 to 10 L/min for 15 to 20 minutes with a nonrebreathing face mask is effective in about 60% of cases.[50] Sumatriptan, 6 mg subcutaneously, is effective in about 75% of all cluster headache patients, with no tachyphylaxis or rebound effect in most patients. Intranasal sumatriptan or oral triptans are less efficacious. Intravenous DHE, 1 mg, may provide relief in less than 10 minutes; onset is slower with intramuscular or intranasal administration. Triptans and DHE should not be used within 24 hours of each other. Ergotamine may also be effective. Topical lidocaine 4%, administered as nosedrops, may be effective in at least one third of patients. To administer the drops, the patient lies supine with the head tilted backward 30° and turned to the side of the headache. A nasal dropper may be used. The dose (1 ml) may be repeated once after 15 minutes. Butorphanol nasal spray may be tried if other treatments are not effective or are contraindicated, but this medication has a significant potential for habituation and addiction.[51]

Transitional treatments are medications that may induce rapid suppression of attacks before a preventive medication takes effect. Transitional treatments include prednisone, 60 mg daily for 3 days, followed by 10 mg decrements every 3 days (the drug is given in the morning to prevent interference with sleep); ergotamine tartrate, 1 mg orally twice a day, including a bedtime dose if nocturnal attacks occur (the drug is contraindicated in patients with peripheral vascular and cardiovascular disease; ergotamine and triptans should not be used within 24 hours of each other); DHE, 0.5 to 1.0 mg subcutaneously or intramuscularly every 8 to 12 hours; and a greater occipital nerve block on the side ipsilateral to the attacks, using 120 mg of methylprednisolone and 3 ml of 1% lidocaine.

A number of medications may be effective for prevention of cluster headaches. Verapamil is the drug of choice for both episodic and chronic types. It is started at 120 to 240 mg a day and slowly increased (up to 80 mg increase every 3 days) to 480 mg if necessary.[50] The drug can be given in both a regular formulation three times daily and an extended-release formulation once a day. In some cases of chronic cluster headache, a daily dose of more than 720 mg may be necessary. With daily doses of 240 mg or higher, baseline and serial electrocardiograms are indicated to monitor for the development of heart block. Methysergide may be effective for younger patients with episodic cluster headache who have no contraindications for its use. Methysergide is started at 2 mg three times a day and increased to 12 mg daily if necessary. It may be best to avoid combining methysergide with ergotamine, DHE, or triptans because of the potential for additive vasoconstrictor effect. However, methysergide has limited availability in the United States. Other medications that may be effective include topiramate (50 to 125 mg daily); divalproex sodium (500 to 2,000 mg daily); lithium carbonate (150 to 300 mg t.i.d. with monitoring of blood levels), especially for chronic rather than episodic cluster; baclofen (10 mg t.i.d.); and melatonin, 10 mg at bedtime. Topical capsaicin cream 0.025% may be effective; it is applied with a cotton-tipped applicator 0.5 in. up the nostril on the same side as the headache three times daily for 7 days. For chronic or intractable cases, combination therapy can be used.

Surgical treatment may be useful for patients with total resistance to medical treatment and strictly unilateral pain. Options include percutaneous radiofrequency retrogasserian rhizotomy, gamma-knife radiosurgery to ablate the trigeminal nerve root, and percutaneous retrogasserian glycerol rhizolysis.

Geriatric Headache

Older persons have fewer headaches than younger ones. The prevalence of headaches at different ages in women and men, respectively, is as follows: 21 to 34 years, 92% and 74%; 55 to 74 years, 66% and 53%; and after age 75, 55% and 22%.[52] Although 90% of headaches in younger patients are of the primary type, only 66% of headaches in the elderly are primary.[53] There is a decreasing prevalence of migraine with older age. Past the age of 70 years, only 5% of women and 2% of men still have migraine. There are many causes of new-onset headaches in the elderly, some of which can be particularly worrisome.[54] The risk of serious secondary disorders in persons older than 65 years is 10 times higher than that in younger persons.[55]

LATE-LIFE MIGRAINE ACCOMPANIMENTS

Late-life migraine accompaniments are transient visual, sensory, motor, or behavioral neurologic manifestations that are similar or identical to migraine aura.[56] Headache is associated with only 50% of cases and may be mild. These accompaniments occur more often in men than in women. From most to least common, migraine accompaniments consist of visual symptoms (transient blindness, homonymous hemianopsia, and blurring of vision); paresthesias (numbness, tingling, pins-and-needles sensation, or a heavy feeling of an extremity); brain stem and cerebellar dysfunction (ataxia, clumsiness, hearing loss, tinnitus, vertigo, and syncope); and disturbances of speech (dysarthria or dysphasia).

Other causes of transient cerebral ischemia should be considered, especially when the patient is seen after the first episode or if the case has unusual aspects. The usual diagnostic evaluation for transient ischemic attacks (TIAs) or seizures is performed [*see 11:IV Cerebrovascular Disorders and 11:XII Epilepsy*].

Features that help distinguish migraine accompaniments from TIAs include a gradual buildup of sensory symptoms; a march of sensory paresthesias; serial progression from one accompaniment to another; longer duration (90% of TIAs last for less than 15 minutes); and multiple stereotypical episodes.

If the episodes are frequent, preventive treatment can be considered with medications such as verapamil, topiramate, divalproex sodium, and aspirin. Beta blockers should be avoided because of the potential for worsening of vasospasm. For acute treatment, ergotamine, DHE, and triptans should be avoided because of the risk of increasing cerebral vasospasm.

CEREBROVASCULAR DISEASE

Headaches commonly accompany stroke. In a prospective study of 163 patients with stroke, headache occurred in 29% with bland infarcts, 57% with parenchymal hemorrhage, 36% with TIAs, and 17% with lacunar infarcts.[57] Women and patients with a history of recurrent throbbing headaches were more likely to have headaches associated with stroke. The headache began before the stroke in 60% of cases and at its onset in 25%. The quality, onset, and duration of stroke-associated headaches vary widely. The headaches are equally likely to be abrupt and to be gradual in onset. In patients presenting with

what they consider to be the worst headaches of their lives, SAH should be excluded.

Headache accompanying stroke is usually unilateral, focal, and of mild to moderate severity, although up to 46% of patients may have an incapacitating headache. The headache may be throbbing or nonthrobbing and, in rare cases, may be stabbing. The headache is more often ipsilateral than contralateral to the side of the cerebral ischemia. Headache is more common in ischemia of the posterior circulation than of the anterior circulation and more common in cortical than in subcortical events. The headache is of longest duration in cardioembolic infarcts and thrombotic infarcts, of medium duration in lacunar infarction, and of shortest duration in TIAs.

HEAD TRAUMA

Although there are numerous causes of head trauma, falls are of particular concern in the elderly. Approximately 30% of all persons older than 65 years fall at least once a year. Subdural hematomas follow approximately 1% of mild head injuries, even those involving no loss of consciousness, such as a bump on the head or riding a roller coaster. Chronic subdural hematomas occur more often in the elderly because of brain atrophy that causes stretching of the parasagittal bridging veins and a predisposition to tearing. The atrophy in an older person also permits hematomas to accumulate without symptoms for a longer period of time than it does in a younger person. Other risk factors include use of aspirin or warfarin[58] and alcoholism.

Headaches are present in up to 90% of patients with head trauma. The headaches are nonspecific; they can range from mild to severe and from paroxysmal to constant and can be bilateral or unilateral. They may be exacerbated by coughing, straining, or exercise and may be associated with vomiting or nausea. About 50% of patients with chronic subdural hematomas will have altered mental status. A strokelike presentation with a transient or persistent hemiparesis can also occur. Only about 50% of patients with a chronic subdural hematoma will have a history of a head injury. The history may also be inaccurate in patients with dementia.

TEMPORAL ARTERITIS

Temporal (giant cell) arteritis (TA) is a systemic panarteritis that selectively involves arterial walls with significant amounts of elastin. Approximately 50% of patients with TA have polymyalgia rheumatica, and about 15% of patients with polymyalgia rheumatica have TA. Both conditions occur almost exclusively in patients older than 50 years, with a mean age of onset of about 70. The ratio of women to men with TA is 3 to 1. The annual incidence is about 18 per 100,000 population in persons older than 50 years.

Headaches are the most common symptom of TA, reported by 60% to 90% of TA patients.[59] The pain is most often throbbing, although many patients describe a sharp, dull, burning, or lancinating pain. The pain may be intermittent or continuous and is more often severe than moderate or slight. For some patients, the pain may be worse at night when lying on a pillow, while combing the hair, or when washing the face. Tenderness or decreased pulsation of the superficial temporal arteries is present on physical examination in about half of all patients with TA. The location of the headache is variable and may be unilateral or bilateral. Intermittent jaw claudication occurs in 38% of cases.

The diagnosis of TA is based on clinical suspicion, which is usually but not always confirmed by laboratory testing.[60] The three best tests are the Westergren ESR, the CRP level, and temporal artery biopsy. For elderly patients, the ESR range of normal may vary from less than 20 mm/hr to 40 mm/hr. Elevation of the ESR is not specific for TA; elevation of the ESR can be seen in any infectious, inflammatory, or rheumatic disease. TA with a normal ESR has been reported in 10% to 36% of patients. When abnormal, the ESR averages 70 to 80 mm/hr and may reach 120 or even 130 mm/hr. If the ESR is elevated at the time of diagnosis, it can be followed to help guide the corticosteroid dosage.

CRP is an acute-phase plasma protein from the liver. As with the ESR, elevation of CRP levels is nonspecific and can be seen with numerous disorders. The CRP level is not influenced by various hematologic factors or age and is more sensitive than the ESR for the detection of TA. The combination of ESR and CRP levels gives the best specificity (97%).

The diagnosis of TA is made with certainty when a superficial temporal artery biopsy demonstrates necrotizing arteritis characterized by a predominance of mononuclear cell infiltrates or a granulomatous process with multinucleated giant cells. The false negative rate of temporal artery biopsies ranges from 5% to 44%.

In patients without contraindications, treatment of TA is typically started with prednisone at a dosage of 40 to 80 mg a day. The headache will often improve within 24 hours. The initial dose is maintained for about 4 weeks and then slowly reduced over many months, depending on the clinical effect, the ESR, and the occurrence of side effects. Long-term treatment is often required.

TRIGEMINAL NEURALGIA

Trigeminal neuralgia begins after the age of 40 in 90% of cases. About 80% of cases result from vascular compression of the trigeminal nerve at the root entry zone; most commonly, such compression is caused by a branch of the superior cerebellar artery. About 5% of cases are caused by tumors. The pain is a severe, sharp, shooting, or electric shock–like sensation lasting seconds to 2 minutes. It is usually in a unilateral maxillary or mandibular trigeminal distribution and uncommonly in the ophthalmic division.[61]

In about 90% of cases of trigeminal neuralgia, the patient has trigger zones, usually in the central part of the face around the nose and lips. Normally nonpainful stimuli in these zones can trigger pain. Stimuli can include talking, chewing, washing the face, brushing the teeth, shaving, facial movement, and cold air. After a paroxysm of pain, there is a refractory period lasting up to several minutes during which stimulation of the trigger zone will not trigger pain. Facial grimacing or spasm may accompany the pain (tic douloureux). Between painful paroxysms, the patient is usually pain free, although dull aching may persist for a few minutes after attacks of long duration or multiple clustered attacks. Multiple attacks may occur for weeks or months. About 50% of patients with trigeminal neuralgia will have spontaneous remissions for at least 6 months. Physical examination is usually normal except for trigger zones, although up to 25% of patients will have sensory loss.

Medications that may be effective against trigeminal neuralgia, alone or sometimes in combination, include carbamazepine, oxcarbazepine, baclofen, phenytoin, clonazepam, divalproex sodium, topiramate, lamotrigine, gabapentin, and pimozide. About 30% of patients do not respond to medical treatment but may respond to one of the many surgical approaches available.

POSTHERPETIC NEURALGIA

Although herpes zoster most commonly occurs in the thoracic region, the second most commonly involved area is a trigeminal distribution, usually in the ophthalmic division (herpes zoster ophthalmicus), which occurs in 23% of cases. The zoster is almost always unilateral. The incidence of postherpetic neuralgia (PHN) (i.e., the persistence of pain for more than 1 month after the initial outbreak) greatly increases with older age, to about 1,000 per 100,000 population for those who are 80 years of age or older. PHN develops in 50% of persons older than 50 years and in 80% of those older than 80 years. Zoster involving the face nearly doubles the risk of developing facial PHN, which lasts longer than PHN in other locations.

Typically, the vesicles crust, the skin heals, and the pain resolves within 3 to 4 weeks after the onset of the rash of herpes zoster. PHN involves three types of pain: a constant burning or deep aching; an intermittent spontaneous pain with a jabbing or lancinating quality; and a superficial, sharp, or radiating pain or itching provoked by light touch (allodynia), which is present in 90% of persons with PHN and often interferes with sleep.[62] The type of pain experienced varies from patient to patient.

Treatment with oral corticosteroids (e.g., prednisone, starting at 60 mg/day and tapering off over 2 weeks) may reduce acute pain in herpes zoster but does not lower the risk of PHN. One week of therapy with famciclovir (500 mg every 8 hours) or valacyclovir (1,000 mg every 8 hours), ideally started within 72 hours after onset of acute zoster, mildly reduces the risk and duration of PHN.[63] Numerous treatments of varying efficacy are available for PHN, including tricyclic antidepressants (amitriptyline, nortriptyline, and desipramine), duloxetine, gabapentin, pregabalin, topical agents (capsaicin, lidocaine, aspirin, and NSAIDs), opioids, and tramadol. Unfortunately, PHN persists for 1 year or more in over 20% of patients.

CARDIAC ISCHEMIA

In rare cases, cardiac ischemia can cause a unilateral or bilateral headache brought on by exercise and relieved by rest.[64] The headache can occur alone or can be accompanied by chest pain. Angina is generally believed to be caused by afferent impulses that traverse cervicothoracic sympathetic ganglia, enter the spinal cord via the first to the fifth thoracic dorsal roots, and produce the characteristic pain in the chest or inner aspects of the arms. Cardiac vagal afferents, which mediate anginal pain in a minority of patients, join the tractus solitarius. A potential pathway for referral of cardiac pain to the head would be convergence with craniovascular afferents.[65]

HYPNIC HEADACHE

Hypnic headache is a rare disorder that occurs more often in the elderly (but with a range of 36 to 83 years of age) and predominantly in women.[66] The headache occurs only during sleep and awakens the sufferer at a consistent time. Nausea is infrequent, and autonomic symptoms are rare. The headache can be unilateral or bilateral, throbbing or nonthrobbing, and mild to severe in intensity. The headaches can last 15 minutes to 3 hours and can occur frequently, as often as nightly, for many years. Medications reported to be effective include caffeine (one or two cups of caffeinated coffee or a 40 to 60 mg caffeine tablet before bedtime), lithium carbonate (300 mg at bedtime), indomethacin, atenolol, melatonin, cyclobenzaprine, prednisone, and flunarizine (not available in the United States).

The diagnosis of hypnic headache is one of exclusion. Secondary causes of nocturnal headaches that must be ruled out include drug withdrawal, temporal arteritis, sleep apnea, oxygen desaturation, pheochromocytoma, primary and secondary neoplasms, communicating hydrocephalus, subdural hematoma, and vascular lesions.[67] Migraine, cluster, and chronic paroxysmal hemicrania are other primary headaches that can cause awakening from sleep. Migraine typically has associated symptoms and very uncommonly occurs only during sleep. Cluster headaches have autonomic symptoms and may occur during the day as well as during sleep. Chronic paroxysmal hemicrania occurs both during the day and at night, lasts for less than 30 minutes, and occurs 10 to 30 times a day.

Other Headaches

Numerous other types of headache have been identified. The more common of these are briefly discussed.

FIRST OR WORST AND THUNDERCLAP HEADACHES

The term first or worst refers to severe headache of a type the patient has never experienced before, which may be the first episode of a primary headache such as migraine or cluster, or to the worst headache the patient has ever had, which can be caused by numerous primary and secondary disorders [*see Table 10*].[6]

Headache in Subarachnoid Hemorrhage

Headache is present in 90% of cases of subarachnoid hemorrhage, or SAH.[68] The classic headache is sudden, severe, and continuous, often with nausea, vomiting, meningismus, focal neurologic findings, and loss of consciousness [*see 11:IV Cerebrovascular Disorders*].

Thunderclap Headache

A sudden severe headache with maximal onset within 1 minute without evidence of SAH is termed a thunderclap headache.[69] A small percentage of patients with thunderclap headache will have unruptured aneurysms, cerebral vasospasm, cerebral venous thrombosis, carotid artery or vertebral artery dissections, pituitary apoplexy, occipital neuralgia, and possibly Erve virus infection. Most cases of thunderclap headache are primary disorders: primary thunderclap headache, so-called crash migraine, and primary orgasmic headache.

COUGH, EXERTIONAL, AND SEXUAL HEADACHES

Primary cough, exertional, and sexual headache have lifetime prevalence rates of 1% each. All three types occur more often in men.[70]

Primary cough headache is a bilateral headache of sudden onset that is precipitated by coughing and lasts less than 1 minute. About 25% of patients have the onset after a respiratory infection with cough. This is an infrequent type of headache, with patients having a mean age of onset of 55 years. The diagnosis is one of exclusion. Patients should undergo neuroimaging to exclude pathology such as Chiari malformation, platybasia, basilar impression, brain tumors, cerebral aneurysm, carotid stenosis, and vertebrobasilar disease. Treatments that may be effective include indomethacin, a single lumbar puncture, and methysergide. Some patients have an abrupt recovery after extraction of abscessed teeth.

Table 10 Differential Diagnosis of the Acute, Severe New-Onset Headache

- Primary headache disorders
  - Migraine
  - Cluster
  - Primary exertional headache
  - Primary orgasmic cephalalgia
- Posttraumatic
  - Associated with vascular disorders
    - Acute ischemic cerebrovascular disease
    - Subdural and epidural hematomas
    - Parenchymal hemorrhage
    - Unruptured saccular aneurysm
    - Subarachnoid hemorrhage
    - Systemic lupus erythematosus
    - Temporal arteritis
    - Internal carotid and vertebral artery dissection
    - Cerebral venous thrombosis
    - Acute hypertension
      - Pressor response
      - Pheochromocytoma
      - Preeclampsia
  - Associated with nonvascular intracranial disorders
    - Intermittent hydrocephalus
    - Benign intracranial hypertension
    - Post–lumbar puncture
    - Related to intrathecal injections
    - Intracranial neoplasm
    - Pituitary apoplexy
  - Acute intoxications
  - Associated with noncephalic infection
    - Acute febrile illness
    - Acute pyelonephritis
  - Cephalic infection
    - Meningoencephalitis
    - Acute sinusitis
  - Acute mountain sickness
  - Disorders of eyes
    - Acute optic neuritis
    - Acute glaucoma
  - Cervicogenic
    - Greater occipital neuralgia
    - Cervical myositis
  - Trigeminal neuralgia

Primary cough headache also includes headache brought on by sneezing, blowing the nose, laughing, crying, weightlifting, bending, stooping, or straining with a bowel movement. Weightlifting can also produce a benign acute bilateral nuchal-occipital or nuchal-occipital-parietal headache that can persist as a residual ache for days or weeks. SAH should be considered as a cause with the initial presentation.

Primary exertional headache is a bilateral, usually throbbing, headache brought on by physical activity and lasting from 5 minutes to 24 hours. Some of the activities that can cause this headache are running, rowing, tennis, and swimming. In some persons, the headache may be precipitated by one activity but not others. Exercise can trigger a migraine in migraineurs. Depending on the clinical scenario and number of headaches, secondary causes may need to be excluded, such as SAH, sinusitis, brain tumors, pheochromocytoma, cardiac ischemia (anginal headache), and intracranial arterial dissection. The headaches may be prevented by a warm-up period or by avoiding particular activities. Indomethacin may be preventive. Migraineurs with exertional headache may respond to migraine preventive medications.

The IHS criteria describe two types of primary headache precipitated by sexual activity in the absence of any intracranial disorder. Both are usually bilateral and may be prevented or eased by stopping sexual activity before orgasm. The preorgasmic type is a dull ache in the head or neck that intensifies as sexual excitement increases and is probably caused by muscle contraction. The orgasmic type is a sudden severe headache occurring at orgasm; the headache may remain severe for minutes to 4 hours and then fade to a milder headache lasting up to 48 hours. Forty percent of patients with the explosive type of headache also have exertional headache. A postural headache similar to a post–lumbar puncture headache can occur after sexual activity, presumably because of a dural tear and CSF leak triggered by the activity.

Sexual headache occurs more often when a person tries to have repeated orgasms in close succession. A personal or family history of migraine is common. SAH should be excluded, especially when patients present with their first sexual headache, because sexual activity is a precipitant of up to 12% of ruptured saccular aneurysms. Rarely, pheochromocytoma is a cause. Phosphodiesterase-5 inhibitors for erectile dysfunction can cause headaches in about 15% of users.

In some patients, primary orgasmic headaches can be prevented by weight loss, an exercise program, a more passive role during intercourse, variation in posture, and limitation of additional sexual activity on a single day. The headache may also be prevented by taking medication (e.g., indomethacin, ergotamine, or a triptan) 30 to 60 minutes before engaging in sexual activity.[71] Patients with frequent sexual headaches may respond to migraine preventive medications, such as a beta blocker or verapamil.

HEADACHE IN PSEUDOTUMOR CEREBRI

Pseudotumor cerebri, also known as idiopathic intracranial hypertension, is a disorder of unknown etiology, with an incidence of one per 100,000 population and an onset usually in persons between the ages of 11 and 58 (mean age, 31 years). Ninety percent of patients are young, obese women. Headache is present in 75% or more of patients, papilledema in 95%, a cranial nerve VI palsy in 25%, transient visual obscurations in 70%, visual loss in 30%, and roaring noises in 70%. The headaches, which are usually pulsatile, daily, and continuous, can be unilateral or bilateral, with a bifrontotemporal location being the most common. Nausea is present in about 60% of cases, and vomiting is present in 40%.

The diagnosis of pseudotumor cerebri is one of exclusion, because there are many other causes of papilledema [*see Table 11*]. Testing includes a scan of the brain. MRI is more sensitive than CT, and magnetic resonance venography will exclude cerebral venous thrombosis. If the brain scan is negative, a lumbar puncture should be done. The opening pressure is usually elevated and the CSF analysis is normal, except for a low CSF protein level in some cases.

Treatments include weight loss and diuretics to decrease CSF production. Diuretics used in pseudotumor cerebri include acetazolamide, starting with a dosage of 500 mg twice daily and increasing to as much as 1 g twice daily, if necessary, and furosemide, starting at 20 mg twice daily and increasing to as much as 40 mg three times daily. Patients taking furosemide

should also receive potassium supplementation. Migraine preventive medications can be useful for persistent headache. Topiramate is especially useful because weight loss is a side effect.[72] Funduscopic exam, visual acuity, and visual fields should be closely monitored to help prevent visual loss. Corticosteroids can be used for emergency treatment of impending visual loss. Surgery may be considered for patients who do not respond to medical treatment and are experiencing progressive visual loss. Surgical options are optic nerve sheath fenestration and lumboperitoneal shunting.

HEADACHE IN BRAIN TUMORS

Up to 70% of persons with brain tumors report headaches.[73] The headaches are usually similar to tension-type headaches but can mimic migraine and cluster headaches. The headaches are usually bilateral but can be unilateral. The neurologic examination can be normal. Suspicion of a brain tumor should be raised when a patient has new-onset or progressive headaches or headaches associated with other problems, such as a seizure, confusion, prolonged nausea and vomiting, hemiparesis, or other focal findings. Headaches that are worst on arising in the morning account for less than 20% of brain tumor headaches.

HEADACHE IN PARANASAL SINUSITIS

Acute sinusitis lasts from 1 day to 4 weeks, and subacute sinusitis lasts from 4 to 12 weeks. Nasal congestion, purulent nasal drainage, and facial tenderness and pain are common. Fever is present in 50% of patients. Anosmia, pain on mastication, and halitosis may also be present. Maxillary sinusitis usually causes pain in the cheek, gums, and maxillary teeth; less often, it causes pain in the periorbital, supraorbital, or temporal areas. The pain decreases when the patient is supine and increases when the head is upright. The maxillary sinus is tender to palpation. Frontal sinusitis causes severe frontal headaches with tenderness to percussion or palpation over the frontal sinus. The pain is less when the head is upright and worse when the patient is supine. Complications include brain abscess, meningitis, subdural or epidural abscess, osteomyelitis, subperiosteal abscess, orbital edema, orbital cellulitis, and orbital abscess.[74]

The headache of sphenoid sinusitis, which accounts for 3% of all cases of acute sinusitis and is usually associated with pansinusitis, may be frontal, occipital, or temporal (alone or in combination) and periorbital. The pain is less when the person is upright and increases when the person is supine, standing, walking, bending, or coughing. Nausea and vomiting are common. Photophobia and eye tearing may be present. Nasal discharge and drainage are present in 30% of cases, and fever occurs in more than 50%. Sphenoid sinusitis may be misdiagnosed as migraine, meningitis, trigeminal neuralgia, or brain tumor. Complications include bacterial meningitis, cavernous sinus thrombosis, subdural abscess, cortical vein thrombosis, ophthalmoplegia, and pituitary insufficiency. A parameningeal focus may cause an aseptic meningitis.

Ethmoid sinusitis produces pain in the periorbital, retro-orbital, temporal, or inner canthal area or between the eyes, and it is usually associated with rhinitis. Coughing, straining, or lying supine can worsen the pain, whereas keeping the head upright lessens it. Complications include meningitis, orbital cellulitis, cavernous sinus thrombosis, and cortical vein thrombosis.

Chronic sinusitis has a duration longer than 12 weeks and can produce a usually low-grade and diffuse headache often accompanied by nasal obstruction, congestion, and fullness. The symptoms often increase during the day.

Plain sinus radiographs can be used to diagnose acute maxillary or frontal sinusitis but are often inadequate for ethmoid or sphenoid disease. CT of the sinuses in the coronal plane is highly sensitive for the detection of nasal and paranasal sinus disease. However, a routine CT scan of the head may inadequately cover these areas. An MRI scan of the brain routinely visualizes the paranasal sinuses. Radiographic evidence of sinusitis is present as an incidental finding in 40% of adults without symptoms.

Treatment of paranasal sinusitis is discussed in detail elsewhere [*see 7:XIX Bacterial Infections of the Upper Respiratory Tract*].

HYPERTENSION

Although mild or moderate hypertension does not usually cause headache, severe hypertension from the following conditions can cause headache: acute pressor response to exogenous agents; pheochromocytoma; malignant hypertension; and preeclampsia and eclampsia. Headaches from severe hypertension are usually bioccipital and throbbing but can be generalized or involve frontal throbbing. The headache is often present on awakening in the morning. The diastolic blood pressure is usually elevated to 120 mm Hg or higher. Hypertensive encephalopathy can present as headache, nausea, and vomiting, which may be associated with visual symptoms. Papilledema, focal neurologic deficits, seizures, and decreased levels of consciousness may be present.

Table 11 Etiologies of Papilledema and Headache[79]

- Intracranial mass
- Obstruction or deformity of the ventricular system
- Cerebral venous thrombosis
- Extracranial venous obstruction
  - Radical neck dissection
  - Cardiac failure
  - Chronic respiratory disease
- Hypertensive encephalopathy
- Preeclampsia and eclampsia
- Meningitis/encephalitis
- Meningeal carcinomatosis
- Elevated CSF protein concentration
  - Guillain-Barré syndrome
  - Systemic lupus erythematosus
  - Spinal tumors, especially oligodendroglioma
- Large arteriovenous malformations
- Optic neuritis (usually unilateral)
- Central retinal venous thrombosis (usually unilateral)
- Lead toxicity (in children)
- Lyme disease (in children)
- Parameningeal infection (in children)
- Head trauma
- Medications
  - Vitamin A and derivatives (isotretinoin, etretinate)
  - Minocycline and tetracycline
  - Anabolic steroids
  - Steroid withdrawal
  - Nalidixic acid
- Other medical conditions
  - Renal disease
  - Hypoparathyroidism
  - Hypercoagulable states

A sudden severe headache can reflect an acute pressor response caused when patients receiving monoamine oxidase inhibitors ingest wine or foods with a high tyramine level. Illicit drugs with sympathomimetic actions, such as cocaine, methamphetamine, and methylenedioxymethamphetamine (ecstasy), can also cause acute hypertension and stroke [*see 8:I Management of Poisoning and Drug Overdose*].

OCCIPITAL NEURALGIA

The term occipital neuralgia is in some ways a misnomer, because the pain is not necessarily from the occipital nerve and does not usually have a neuralgic quality. Greater occipital neuralgia is a common type of posttraumatic headache but frequently is also seen in patients without injury. The aching, pressure, stabbing, or throbbing pain may be in a nuchal-occipital, parietal, temporal, frontal, periorbital, or retro-orbital distribution. Occasionally, a true neuralgia may be present, with paroxysmal shooting pain. The headache may last for minutes or hours to days and can be unilateral or bilateral. Lesser occipital neuralgia tends to be similar but with pain generally referred more laterally over the head.

The headache may result from an entrapment of the greater occipital nerve in the aponeurosis of the superior trapezius or semispinalis capitis muscle or may instead be referred pain without nerve compression from trigger points in these or other suboccipital muscles. Digital pressure over the greater occipital nerve at the midsuperior nuchal line (halfway between the posterior mastoid and the occipital protuberance) reproduces the headache. However, pain referred from the C2-C3 facet joint or other area of the upper cervical spine and posterior fossa pathology may produce a similar headache.

Occipital neuralgia may improve with local anesthetic nerve blocks, which can be combined with an injectable corticosteroid (e.g., 3 ml of 1% lidocaine or 2.5 ml of 1% lidocaine and 3 mg of betamethasone). Before giving the injection, the physician should perform aspiration to avoid inadvertent injection into the occipital or vertebral artery. NSAIDs and muscle relaxants may also be of benefit. If the patient has a true occipital neuralgia with paroxysmal lancinating pain, treatment with baclofen, carbamazepine, or gabapentin may help. Physical therapy and transcutaneous nerve stimulators may help some patients.

The author is a consultant or a member of the speakers' bureaus of GlaxoSmithKline, AstraZeneca, Merck & Co., Inc., Pfizer, Inc., UCB Pharma, Inc., Eisai Inc., and Ortho-McNeil Neurologics.

References

1. The International Classification of Headache Disorders: 2nd edition. Headache Classification Subcommittee of the International Headache Society. Cephalalgia 24(suppl 1):1, 2004

2. Evans RW, Mathew NT: Handbook of Headache, 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 2005

3. Lance JW, Goadsby PJ: Mechanism and Management of Headache, 7th ed. Butterworth Heinemann, Oxford, 2005

4. The Headaches, 2nd ed. Olesen J, Tfelt-Hansen P, Welch KMA, et al, Eds. Lippincott Williams & Wilkins, Philadelphia, 2006

5. Silberstein SD, Lipton RB, Dalessio SJ: Wolff's Headache and Other Head Pain, 7th ed. Oxford University Press, New York, 2001

6. Evans RW: Diagnosis of headaches. Handbook of Headache, 2nd ed. Evans RW, Mathew NT, Eds. Lippincott Williams & Wilkins, Philadelphia, 2005, p 1

7. Evans RW: The evaluation of headaches. Diagnostic Testing in Neurology. Evans RW, Ed. WB Saunders Co, Philadelphia, 1999, p 1

8. Evans RW: Diagnostic testing for headache. Med Clin North Am 85:865, 2001

9. Lipton RB, Stewart WF, Diamond S, et al: Prevalence and burden of migraine in the United States: data from the American Migraine Study II. Headache 41:6467, 2001

10. Lipton RB, Scher AI, Kolodner K, et al: Migraine in the United States. Neurology 58:885, 2002

11. Evans RW, Lipton RB, Silberstein SD: The prevalence of migraine in neurologists. Neurology 61:1271, 2003

12. Schecter AL, Lipton RB, Silberstein SD: Migraine comorbidity. Wolff's Headache and Other Head Pain. Silberstein SD, Lipton RB, Dalessio DJ, Eds. Oxford, New York, 2001, p 108

13. Scher AI, Bigal ME, Lipton RB: Comorbidity of migraine. Curr Opin Neurol 18:305, 2005

14. Goadsby PJ, Lipton RB, Ferrari MD: Migraine—current understanding and treatment. N Engl J Med 346:257, 2002

15. Barbanti P, Fabbrini G, Pesare M, et al: Neurovascular symptoms during migraine attacks. Cephalalgia 21:295, 2001

16. Martin VT, Behbehani MM: Toward a rational understanding of migraine trigger factors. Med Clin North Am 85:911, 2001

17. Lempert T, Neuhauser H: Migrainous vertigo. Neurol Clin 23:715, 2005

18. Siow HC, Welch KMA: Basilar-type migraine. MedLink Neurology (www.medlink.com). Gilman S, Ed. MedLink Corporation, San Diego, 2006

19. Lipton RB, Stewart WF, Stone AM: Stratified care vs step care strategies for migraine: the Disability in Strategies of Care (DISC) Study: a randomized trial. JAMA 2284:2599, 2000

20. Freitag FG, Cady R, DiSerio F, et al: Comparative study of a combination of isometheptene mucate, dichloralphenazone with acetaminophen and sumatriptan succinate in the treatment of migraine. Headache 41:391, 2001

21. Silberstein SD, McCrory DC: Butalbital in the treatment of headache: history, pharmacology, and efficacy. Headache 41:953, 2001

22. Silberstein SD: Migraine. Lancet 363:381, 2004

23. Geraud G, Keywood C, Senard JM: Migraine headache recurrence: relationship to clinical, pharmacological, and pharmacokinetic properties of triptans. Headache 43:376, 2003

24. Dodick D, Lipton RB, Martin V, et al: Consensus statement: cardiovascular safety profile of triptans (5-HT agonists) in the acute treatment of migraine. Headache 44:414, 2004

25. Maclntyre PD, Bhargava B, Hogg HJ, et al: Effect of subcutaneous sumatriptan, a selected 5HT agonist, on the systemic, pulmonary, and coronary circulation. Circulation 87:401, 1993

26. Dahlof CGH, Matthew NT: Cardiovascular safety of 5HT1B/1D agonists: is there a cause for concern? Cephalalgia 18:539, 1998

27. Morgan DR, Trimble M, McVeigh GE: Atrial fibrillation associated with sumatriptan. BMJ 321:275, 2000

28. Hood S: Hemodynamic changes may have caused AF. BMJ #9026 August 2, 2000 bmjjournals.com/cgi/eletters/321/7256/275

29. O'Quinn S, Davis RL, Gutterman DL, et al: Prospective large-scale study of the tolerability of subcutaneous injection for acute treatment of migraine. Cephalalgia 19:223, 1999

30. Evans RW, Martin V: Expert opinion: assessing cardiac risk prior to use of triptans. Headache 40:599, 2000

31. Ramadan NM, Silberstein SD, Freitag FG, et al: Evidence-based guidelines for migraine headache in the primary-care setting: pharmacological management for prevention of migraine. American Headache Society, Mount Royal, New Jersey, 2004 http://www.aan.com/professionals/practice/pdfs/g10090.pdf

32. Evans RW, Bigal ME, Grosberg B, et al: Target doses and titration schedules for migraine preventive medications. Headache 46:160 2006

33. Bulut S, Berilgen MS, Baran A, et al: Venlafaxine versus amitriptyline in the prophylactic treatment of migraine: randomized, double-blind, crossover study. Clin Neurol Neurosurg 107:44, 2004

34. Ozyalcin SN, Talu GK, Kiziltan E, et al: The efficacy and safety of venlafaxine in the prophylaxis of migraine. Headache 45:144, 2005

35. Sandor PS, Di Clemente L, Coppola G, et al: Efficacy of coenzyme Q10 in migraine prophylaxis: a randomized controlled trial. Neurology 64:713, 2005

36. Silberstein SD: Hormone-related headache. Med Clin North Am 85:1017, 2001

37. Allais G, De Lorenzo C, Mana O, et al: Oral contraceptives in women with migraine: balancing risks and benefits. Neurol Sci 25(suppl 3):S211, 2004

38. Recommendations on the use of oral contraceptives in women with migraine. International Headache Society Task Force. Cephalalgia 20:155, 2000

39. Evans RW: Headaches in women. Handbook of Headache, 2nd ed. Evans RW, Mathew NT, Eds. Lippincott Williams & Wilkins, Philadelphia, 2005

40. Li D, Rozen TD: The clinical characteristics of new daily persistent headache. Cephalalgia 22:66, 2002

41. Silberstein SD, Dongmei L: Drug overuse and rebound headache. Curr Pain Headache Rep 6:240, 2002

42. Silverman K, Evans SM, Strain EC, et al: Withdrawal syndrome after the double-blind cessation of caffeine consumption. N Engl J Med 327:1109, 1992

43. Silberstein SD: Chronic daily headache. J Am Osteopath Assoc 105(4 suppl 2):23S, 2005

44. Diener HC, Limmroth V: Medication-overuse headache: a worldwide problem. Lancet Neurol 3:475, 2004

45. Katsarava Z, Fritsche G, Muessig M, et al: Clinical features of withdrawal headache following overuse of triptans and other headache drugs. Neurology 57:1694, 2001

46. Krymchantowski AV, Barbosa JS: Prednisone as initial treatment of analgesic-induced daily headache. Cephalalgia 20:107, 2000

47. Smith TR: Low-dose tizanidine with nonsteroidal anti-inflammatory drugs for detoxification from analgesic rebound headache. Headache 42:175, 2002

48. Jolles S, Sewell WA, Leighton C: Drug-induced aseptic meningitis: diagnosis and management. Drug Saf 22:215, 2000

49. Bahra A, May A, Goadsby PJ: Cluster headache: a prospective clinical study with diagnostic implications. Neurology 58:354, 2002

50. May A: Cluster headache: pathogenesis, diagnosis, and management. Lancet 366:843, 2005

51. Dodick DW, Campbell JK: Cluster headache: diagnosis, management, and treatment. Wolff's Headache and Other Head Pain, 7th ed. Silberstein SD, Lipton RB, Dalessio DJ, Eds. Oxford University Press, New York, 2001, p 283

52. Waters WE: The Pontypridd headache survey. Headache 14:81, 1974

53. Solomon GD, Kunkel RS, Frame J: Demographics of headache in elderly patients. Headache 30:273, 1990

54. Evans RW: Headaches over the age of 50. Handbook of Headache, 2nd ed. Evans RW, Mathew NT, Eds. Lippincott Williams & Wilkins, Philadelphia, 2005

55. Pascual J, Berciano J: Experience in the diagnosis of headaches that start in elderly people. J Neurol Neurosurg Psychiatry 57:1255, 1994

56. Fisher CM: Late-life migraine accompaniments: further experience. Stroke 17:1033, 1986

57. Portenoy RK, Abissi CJ, Lipton RB, et al: Headache in cerebrovascular disease. Stroke 15:1009, 1984

58. O'Brien DF, Basu S, O'Donnell JR, et al: The impact of aspirin therapy and anticoagulation on the prevalence of spontaneous subdural haematoma. Ir Med J 93:244, 2000

59. Myklebust G, Gran JT: A prospective study of 287 patients with polymyalgia rheumatica and temporal arteritis: clinical and laboratory manifestations at onset of disease and at the time of diagnosis. Br J Rheumatol 35:1661, 1996

60. Rahman W, Rahman FZ: Giant cell (temporal) arteritis: an overview and update. Surv Ophthalmol 50:415, 2005

61. Rozen TD: Trigeminal neuralgia and glossopharyngeal neuralgia. Neurol Clin 22:185, 2004

62. Cluff RS, Rowbotham MC: Pain caused by herpes zoster infection. Neurol Clin16:813, 1998

63. Gnann JW, Whitley RJ: Clinical practice: herpes zoster. N Engl J Med 347:340, 2002

64. Lipton RB, Lowenkopf T, Bajwa ZH, et al: Cardiac cephalgia: a treatable form of exertional headache. Neurology 49:813, 1997

65. Lance JW, Lambros J: Unilateral exertional headache as a symptom of cardiac ischemia. Headache 38:315, 1998

66. Evers S, Goadsby PJ: Hypnic headache: clinical features, pathophysiology, and treatment. Neurology 60:905, 2003

67. Gould JD, Silberstein SD: Unilateral hypnic headache: a case study. Neurology 49:1749, 1997

68. Weir B: Headaches from aneurysms. Cephalalgia 14:79, 1994

69. Dodick DW: Thunderclap headache. J Neurol Neurosurg Psychiatry 72:6, 2002

70. Cutrer FM, Boes CJ: Cough, exertional, and sex headaches. Neurol Clin 22:133, 2004

71. Evans RW, Pascual J: Orgasmic headaches: clinical features, diagnosis, and management. Headache 40:491, 2000

72. Friedman DI: Pseudotumor cerebri. Neurol Clin 22:99, 2004

73. Purdy RA, Kirby S: Headaches and brain tumors. Neurol Clin 22:39, 2004

74. Silberstein SD: Headaches due to nasal and paranasal sinus disease. Neurol Clin 22:1, 2004

75. Codispoti JR, Prior MJ, Fu M, et al: Efficacy of nonprescription doses of ibuprofen for treating migraine headache. a randomized controlled trial. Headache 41:665, 2001

76. Lipton RB, Baggish JS, Stewart WF, et al: Efficacy and safety of acetaminophen in the treatment of migraine: results of a randomized, double-blind, placebo-controlled, population-based study. Arch Intern Med 160:3486, 2000

77. Lipton RB, Stewart WF, Ryan RE, et al: Efficacy and safety of acetaminophen, aspirin, and caffeine in alleviating migraine headache pain: three double-blind, randomized, placebo-controlled trials. Arch Neurol 55:210, 1998

78. Lange R, Schwarz JA, Hohn M: Acetylsalicylic acid effervescent 1000 mg (aspirin) in acute migraine attacks; a multicentre, randomized, double-blind, single-dose, placebo-controlled parallel group study. Cephalalgia 20:663, 2000

79. Evans RW: Headaches and neoplasms, high and low pressure, and HEENT disorders. Handbook of Headache, 2nd ed. Evans RW, Mathew NT, Eds. Lippincott Williams & Wilkins, Philadelphia, 2005, p 269

IX DEMYELINATING DISEASES

J. William Lindsey, M.D.
Jerry S. Wolinsky, M.D.

In both the central and the peripheral nervous systems, large-diameter axons are myelinated. Myelin is formed and maintained by oligodendrocytes within the central nervous system and by Schwann cells in the peripheral nervous system (PNS). Myelin both insulates the invested axons and organizes surface membrane constituents of the axon—functions that are critical for the rapid transfer of signals necessary for coordinated motor activity, proper integration and interpretation of sensory stimuli, and facile cognition. Diseases that affect the integrity of the oligodendrocyte and its ability to produce and maintain myelin and diseases that directly damage the myelin sheath disturb conduction in myelinated white-matter pathways and produce a broad array of motor, sensory, and cognitive dysfunctions [*see Figure 1*].

Demyelinating diseases disturb the integrity of myelin, but the axons are relatively spared [*see Table 1*]. These diseases primarily affect oligodendroglial survival (e.g., progressive multifocal leukoencephalopathy), oligodendroglial metabolism (e.g., vitamin B_{12} deficiency), and the myelin sheath with secondary effects on the oligodendrocytes (e.g., multiple sclerosis [MS]).

Immune-mediated demyelinating diseases include recurrent or chronically progressive demyelinating diseases (MS and its variants) and monophasic demyelinating diseases (optic neuritis, acute disseminated encephalomyelitis, and transverse myelitis). Monophasic demyelination may be the first clinical sign of MS.

Multiple Sclerosis

MS is characterized clinically by recurrent or chronically progressive neurologic dysfunction caused by lesions in the CNS. Pathologically, the lesions are multiple areas of demyelination that affect the brain, optic nerves, and spinal cord.

EPIDEMIOLOGY

Approximately 250,000 to 350,000 people in the United States have MS; the prevalence is about 85 cases per 100,000 population.[1] MS is more common in women than men, with a ratio of 2:1. MS occurs in all major racial groups but is most common in whites, less common in blacks, and rare in Asians. The onset of disease usually occurs between 20 and 50 years of age, with a peak at 30 years of age. The prevalence of MS varies widely with location; the highest prevalence is found at higher latitudes in northern Europe and northern North America. The geographic variation suggests that MS may in part be caused by the action of some environmental factor that is more common at high latitudes.[2] To some extent, this idea is supported by studies of migrants. Moving from a high-risk to a low-risk area early in life tends to lower the risk of MS, whereas moving from a low-risk to a high-risk area tends to increase the risk. Much of the state-by-state variation in MS risk in the United States, however, correlates with Scandinavian ancestry, which suggests that geographic variation may be a reflection of the geographic distribution of susceptible ethnic groups.

ETIOLOGY

Although the etiology of MS is unknown, the importance of genetic factors has been firmly established by studies of familial MS. The contribution of nongenetic factors is demonstrated by the lack of concordance for disease in most identical twins and by the effect of migration on risk of disease. The underlying disease process is unknown, but most authorities accept that MS is at least partly an autoimmune or immune-mediated disease. The evidence for this conclusion includes pathologic findings (i.e., immune components within acute and chronic plaques), an in-

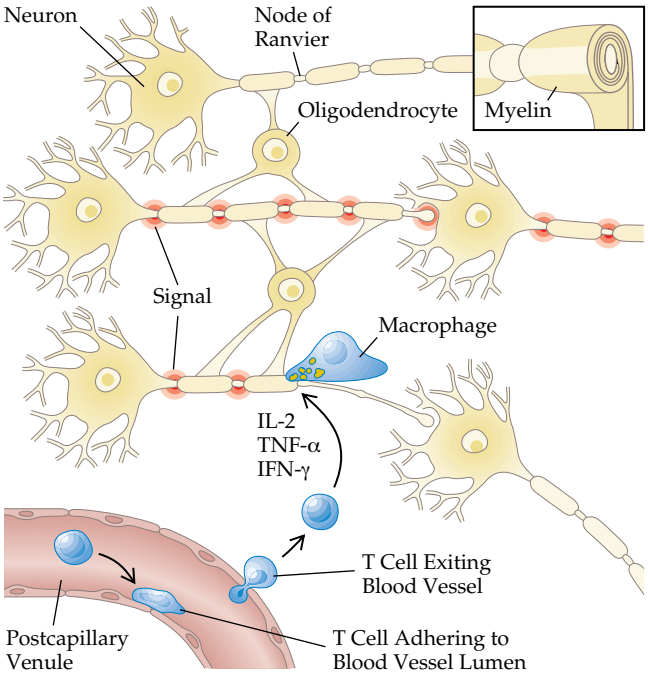


Figure 1 **A single oligodendrocyte may support myelin internodes for 60 or more neighboring axons. A major consequence of myelination is organization of the microenvironment of the larger-diameter axons. Sodium channels are clustered at segments of the axon between myelin segments (the nodes of Ranvier), and potassium channels are diffusely distributed under the myelin-invested segments of the axons. This arrangement allows myelinated axons to rapidly propagate action potentials from the axonal cell body distally to its synapse by saltatory conduction.[93] Depolarization of the nodes of Ranvier in sequential fashion results in rapid propagation of the neural impulses in the largest-diameter nerve fibers at rates in excess of 100 m/sec.**
The upper part of the figure depicts a neuron with a myelinated axon. At the right corner, the neuron is magnified and cut in cross section, demonstrating the concentric lamellar structure of myelin. The middle of the figure shows a neuron with a myelinated axon forming a synapse with another neuron. Two oligodendrocytes are shown; each cell myelinates multiple segments of more than one axon. The bottom neuron has been partially demyelinated, and it is surrounded by T cells, which secrete inflammatory cytokines (interleukin-2 [IL-2], interferon gamma [IFN-γ], and tumor necrosis factor–α [TNF-α]), and by macrophages, which are stripping myelin from the axon. The macrophages contain myelin debris in phagocytic vacuoles. Conduction in the demyelinated axon is blocked. A blood vessel in cross section shows T cells adhering to the lumen and crossing from the vessel into the brain.

crease in the synthesis of IgG in the cerebrospinal fluid, and a variety of in vitro abnormalities of leukocyte function. The autoimmune response could be either the primary cause of disease or an epiphenomenon of another disease process. The popular molecular mimicry model postulates that an infectious organism that contains a protein similar to a myelin protein precipitates the autoimmune attack on myelin. The infection elicits a vigorous immune response from lymphocytes that recognize the cross-reactive protein; in the process of eliminating the organism, the activated lymphocytes damage myelin. The plausibility of this theory has been demonstrated in elegant transgenic experiments in mice,[3] but the relevance of this model in MS remains unproved. Other processes or mechanisms, such as a persistent CNS infection or a biochemical defect in myelin, may cause MS.

GENETICS

The suspicion that susceptibility to MS is at least partly familial has been confirmed in extensive studies. The risk of MS occurring in a monozygotic twin of an MS patient has been found to be 25.3%, whereas the risk of MS occurring in a dizygotic twin has been determined to be about 5%.[4] The risk for a sibling or parent of an affected person is 2% to 5%, compared with a risk in the general population of only 0.1%.[5] Further studies in adopted siblings and half-siblings demonstrated that this increased familial risk is entirely attributable to genetic rather than environmental factors.[6] These findings imply that several interacting genes influence susceptibility to MS.[7]

The genetic factors that confer susceptibility to MS are only partly known. The only locus reproducibly shown to be associated with MS is human leukocyte antigen (HLA) class II; the DRB1*1501 extended haplotype is associated with MS in white patients.[8,9] In an attempt to identify additional loci that contribute to MS, independent research groups conducted genome-wide screens on large groups of families with multiplex MS but found no genome regions with particularly strong linkage to MS.[10]

PATHOLOGY

Plaques, which consist of varying amounts of perivenular mononuclear cell inflammation, demyelination with relative axonal sparing, loss of oligodendrocytes, and astrocyte proliferation with resultant gliosis, are the hallmark of MS.[11] Within the CNS, plaques can appear wherever there are myelinated fibers. The majority of plaques congregate along periventricular draining veins, but plaques also commonly occur within the spinal cord, optic nerves, brain stem, and white matter of the cerebral hemispheres and cerebellum. Plaques can also occur in the connecting pathways of subcortical white matter. Demyelination in gray matter may account for a large fraction of the lesion burden.[12]

The histopathologic appearance of plaques varies over time. Acute, newly formed lesions are dominated by perivenular cellular infiltrates of T cells and macrophages. Blood-brain barrier function is regionally disrupted; this disruption is associated with vasogenic edema. There is immunocytochemical and cytochemical evidence of local endothelial cell activation by cytokines, activation of T cells, and activation of macrophages.[13] Longer-standing lesions are characterized by a total loss of myelin and oligodendrocytes, an intense astrogliosis, variable degrees of axonal loss, and a scant residual infiltrate of mononuclear cells, some of which are immunoglobulin-secreting B cells. Axonal transection is common in many MS lesions, particularly those that appear as areas of persistent low-signal intensity on T_1-weighted magnetic resonance imaging, and axonal transection may be the cause of fixed neurologic deficits.[14-16] Careful pathologic studies of autopsy and biopsy material have defined different types of MS lesions; lesion types differ with regard to the pattern of demyelination, the occurrence of apoptosis in oligodendrocytes, and the activation of complement. This suggests that the underlying pathogenesis of MS may be heterogeneous.[17,18]

Table 1 Demyelinating Diseases of the Central Nervous System

Type	*Disease*
Immune-mediated	Recurrent Multiple sclerosis Monophasic Optic neuritis Transverse myelitis Acute disseminated encephalomyelitis
Inherited	Adrenoleukodystrophy Metachromatic leukodystrophy
Metabolic	Vitamin B_{12} deficiency Central pontine myelinolysis
Infectious	Progressive multifocal leukoencephalopathy Subacute sclerosing panencephalitis

DIAGNOSIS

The diagnosis of MS is made on the basis of the clinical signs and symptoms; MRI and other laboratory tests play a supporting role. The diagnosis of MS requires evidence of the dissemination of lesions in the body over time and the careful exclusion of other causes. The patient should have had more than one episode of neurologic dysfunction and should have evidence of white-matter lesions in more than one part of the CNS. Because there is no pathognomonic sign or symptom or definitive laboratory test result, diagnosis requires careful clinical judgment and should be made only by an experienced neurologist.

The most current version of the recommended diagnostic criteria for MS utilizes both clinical and laboratory data [*see Table 2*].[19] The criteria require the objective demonstration of dissemination of lesions in space and time. This can be achieved with clinical evidence alone or with a combination of clinical and laboratory evidence. The types of laboratory evidence considered are MRI, CSF examination, and evoked responses.

Clinical Features

Almost any neurologic deficit can occur in MS, but there are several signs and symptoms that are characteristic. Although no findings are pathognomonic for MS, the presence of certain signs and symptoms should suggest MS as a possible diagnosis, particularly in young adults.

The typical findings include optic neuritis, internuclear ophthalmoplegia, heat sensitivity, and Lhermitte symptom. Optic neuritis is the initial symptom in about 20% of patients with MS and ultimately develops in more than half of all MS patients. In patients with optic neuritis, vision disturbances are restricted to one eye. These disturbances characteristically occur with central scotoma. There is usually retro-orbital pain that is accentuated with movement of the globe. Results of funduscopic examination are usually normal, but there may be papillitis or pallor of the disk from previous attacks. Symptoms generally evolve over

several days before maximal deficits are reached. A second typical symptom is diplopia caused by an internuclear ophthalmoplegia. On examination, there is failure of adduction on lateral gaze but preservation of adduction with convergence. Internuclear ophthalmoplegia is caused by a lesion in the medial longitudinal fasciculus and is suggestive of MS.

Sensitivity to heat is a characteristic complaint in MS. Exercise, fever, a hot bath, or other activities that raise body temperature may cause the appearance of new symptoms or the recurrence of old symptoms. These events occur as a result of temperature-induced conduction block across partially demyelinated fibers. The symptoms resolve when body temperature returns to normal. Patients with the Lhermitte symptom experience paresthesias that radiate down the spine and into the extremities on neck flexion. This symptom may be reproduced on clinical examination and indicates the presence of a lesion in the cervical spine.

There are a number of other symptoms that are extremely common in MS but also occur in many other disease processes. These symptoms include weakness, sensory loss, ataxia, and bowel and bladder symptoms. All these symptoms develop in the majority of patients during the course of MS [*see Table 3*]. The weakness associated with MS is of the upper motor neuron type and is accompanied by spasticity and increased reflexes. Weakness may occur in any pattern, including paraparesis, hemiparesis, and monoparesis. In established disease, the lower extremities are usually more affected than the upper extremities. Sensory symptoms are also frequent and can occur in different patterns. Patients may note either paresthesias or loss of sensation. Again, the lower extremities tend to be more severely affected. The loss of vibration sense is often most prominent. Ataxia is uncommon at the onset of MS, but it occurs to some degree in most patients. Difficulties with bladder control are frequent and can be extremely distressing. Most commonly, patients have urinary urgency, frequency, and urge incontinence consistent with a spastic bladder, but hesitancy, retention, overflow incontinence, and dyssynergia also occur. Constipation is the most common bowel complaint.

Although almost any deficit can occur in MS, some are extremely rare. These include seizures, aphasia, movement disorders, and muscle atrophy, which reflect primary involvement of gray matter or peripheral nerves. Although these symptoms can occur in MS, their presence is unusual and should prompt reevaluation of the diagnosis.

Table 2 Diagnostic Criteria for Multiple Sclerosis[19]

Clinical Presentation	*Additional Data Needed for MS Diagnosis*
Two or more attacks; objective clinical evidence of two or more lesions	None
Two or more attacks; objective clinical evidence of one lesion	Dissemination in space demonstrated on MRI *or* Two or more MRI lesions plus positive cerebrospinal fluid *or* Await further clinical attack in different site
One attack; objective clinical evidence of two or more lesions	Dissemination in time demonstrated on MRI *or* Second clinical attack
One attack; objective clinical evidence of one lesion	Dissemination in space demonstrated on MRI *or* Two or more MRI lesions plus positive CSF plus dissemination in time demonstrated on MRI *or* Second clinical attack
Insidious progression suggestive of MS	Positive CSF *and* Dissemination in space demonstrated by MRI or MRI and visual evoked response (VER) *and* Dissemination in time demonstrated by MRI or continued progression for 1 yr

Note: a pocket card with current diagnostic criteria and explanation of their use can be obtained from http://www.nationalmssociety.org.

Clinical Course

The clinical course of MS varies greatly among patients. MS typically has a relapsing-remitting pattern, with acute exacerbations followed by partial or complete resolution. New neurologic deficits develop over the course of several hours or days, remain stable for a period of a few days to a few weeks, and then gradually improve. Early in the course of the disease, the symptoms may resolve with minimal residua. With repeated exacerbations, permanent neurologic deficits tend to develop. Patients usually have symptom-free intervals of months or years between attacks.

Symptoms may also occur in a progressive manner, with steady accumulation of neurologic deficits in the absence of clearly defined exacerbations. Patients who initially have relapsing-remitting disease and who then enter a progressive phase are said to have secondary progressive disease, whereas those whose symptoms are progressive from onset are said to have primary progressive disease. About 15% of patients have primary progressive disease; of those who initially have relapsing-remitting disease, 30% to 50% will experience progressive symptoms in the first 10 years.[20] Radiologic and pathologic studies suggest that the primary progressive group may be distinct from the relapsing-remitting and secondary progressive group.[21]

Laboratory Tests

Magnetic resonance imaging MRI is the single most useful laboratory test in the diagnosis of MS. Most patients with MS have abnormalities that can be seen with MRI, and the superb

Table 3 Approximate Distribution of Neurologic Deficits at Onset of MS and after 5 to 10 Years of Disease[94,95]

Deficits	*At Onset (%)*	*5 to 10 Years after Onset (%)*
Cognitive deficits	< 5	30
Visual deficits	20–30	50
Diplopia	10–20	35
Weakness	40	80
Ataxia	5–20	65
Sensory deficits	40	80
Bowel or bladder symptoms	5–10	55

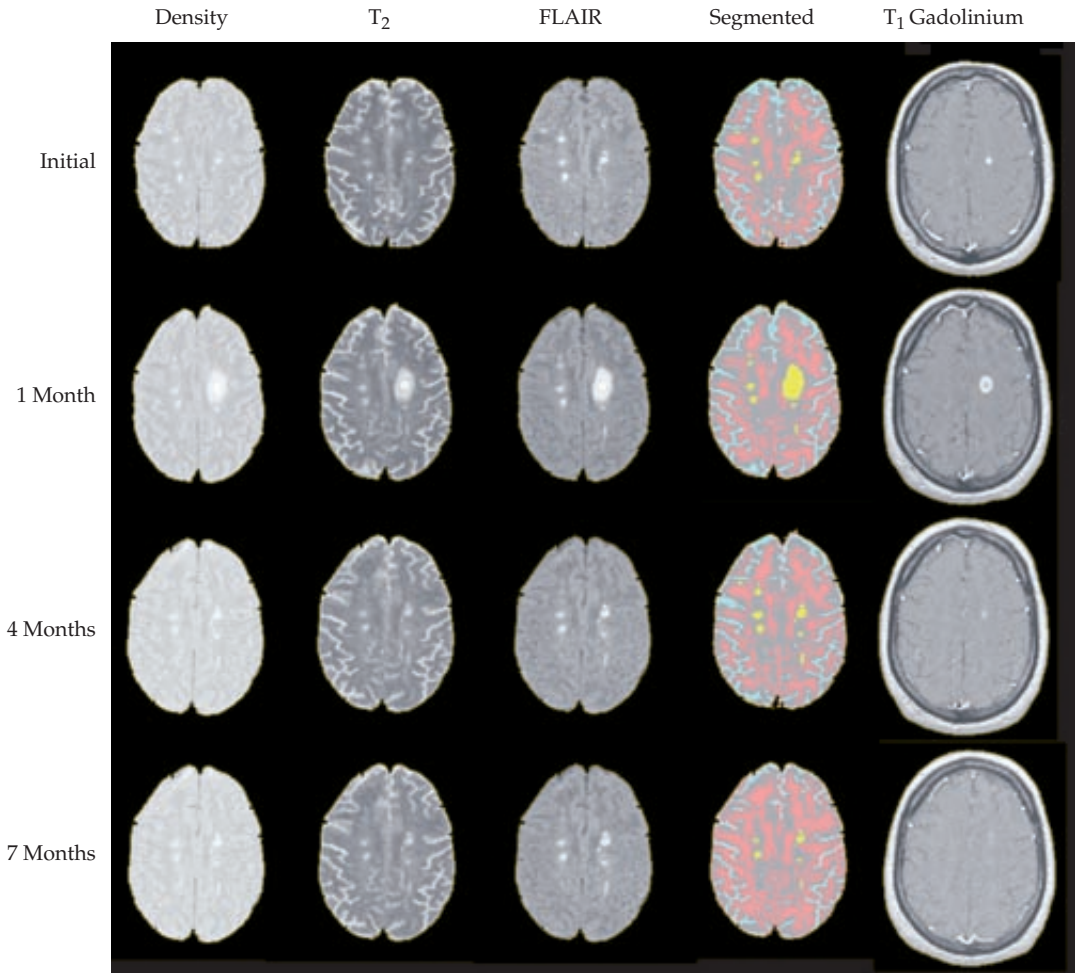


Figure 2 **The composite shows some of the changes captured by serial MRI that are characteristic of the dynamic nature of the underlying pathologic disease activity in multiple sclerosis. The patient was evaluated serially by high-resolution MRI, and the images were quantified automatically. The more conventional appearing images were made using fast spin-echo pulse sequences to produce density-weighted and T_2-weighted images and a fast FLAIR (fluid-attenuated inversion recovery) sequence with magnetization contrast pulse to improve identification of the lesions and suppress signals from the cerebrospinal fluid. The gadolinium-enhanced T_1-weighted image also incorporates a magnetization transfer contrast pulse to amplify the enhancement effects. All images are from the same region of the brain and are 3 mm thick. The segmented images were computer generated through the use of an algorithm that incorporates information from the multiple images. In the segmented images, gray matter is rendered gray, white matter is pink, cerebrospinal fluid appears blue, and the lesions are shown in yellow. The total lesion load in the entire brain varied from 16.71 to 26.98 ml to 16.98 to 21.52 ml at the intervals shown. The total enhanced tissue volume varied from 0.99 to 2.99 ml to 0.33 to 0.04 ml at the same intervals. The patient had no clinically defined change in his neurologic symptoms or findings during this 7-month interval despite the significant activity demonstrated by MRI.**

anatomic resolution of MRI permits the exclusion of many diseases that mimic MS. MS lesions are hyperintense on T_2-weighted or proton-density imaging and are hypointense or isointense on T_1-weighted imaging [*see Figure 2*]. Typical MS lesions are ovoid and periventricular, with their long axis perpendicular to the ventricle, but lesions may appear anywhere in the white matter. Some lesions may be enhanced by the administration of gadolinium chelates; enhancement by gadolinium chelates indicates a breakdown in the blood-brain barrier. Even in early MS, several lesions are usually present, although they may not produce obvious symptoms.

Although MRI is extremely sensitive in detecting white-matter lesions in patients with MS, it is not very specific. Many other diseases produce multiple white-matter lesions; thus, MRI findings should never be used as the sole basis for the diagnosis. MRI findings supportive of the diagnosis of MS include the presence of three or more white-matter lesions, lesions that abut the body of the lateral ventricles, juxtacortical lesions, infratentorial lesions, lesions that are greater than 5 mm, and lesions that show gadolinium enhancement.

The changes that occur in MS lesions over time have been investigated with serial MRI. The size and number of T_2-weighted hyperintense lesions fluctuate, but new lesions tend to accumulate, and the total lesion burden tends to increase.[22] In studies of patients with relapsing-remitting disease who were treated only with short courses of corticosteroids for clinical attacks, the total T_2-weighted lesion burden increased at a rate of 6% to 8% a year.[23,24] Gadolinium enhancement is a transient phenomenon

that usually lasts less than 8 weeks and usually occurs when a new lesion first appears. Gadolinium-enhanced lesions correspond to areas of acute inflammation on pathologic examination, whereas nonenhanced lesions correspond to more chronic disease.[25] Stable, hypointense lesions on T_1-weighted imaging that are not enhanced by gadolinium appear to reflect extensive tissue disruption, including axonal loss. These lesions are highly correlated with clinical measures of neurologic dysfunction.[26] Cerebral atrophy is often present on MRI images. Quantitative computed analysis has demonstrated that the amount of atrophy increases over time[27] and correlates well with clinical disability and disease duration.[28]

The use of serial MRI to study MS has led to fundamental changes in neurologists' concept of the disease process. Disease activity as measured by the appearance of new hyperintense lesions on T_2-weighted images or the appearance of contrast-enhanced lesions on T_1-weighted images greatly exceeds the severity of disease as evidenced by clinical symptoms. Disease activity as assessed by MRI is still episodic but occurs 10 to 20 times more often than clinical symptoms, which suggests that MS is much more of an ongoing and active process, rather than the intermittent process suggested by clinical activity.

Current clinical studies of new therapies for MS include MRI assessment of disease activity as one of the end points. To facilitate the use of MRI in large trials, several groups have developed automated methods for measuring lesion burden and other parameters of interest.[29] Newer sequences, such as magnetization transfer imaging, fluid-attenuated inversion recovery (FLAIR), diffusion tensor imaging, and combinations of these, are being developed to show more subtle changes and greater pathology [*see Figure 2*].[30] Similarly, magnetic resonance spectroscopic imaging (MRSI) [*see Figure 3*] provides additional information on the extent of axonal loss within lesions[31,32] and more directly measures lipid release during active demyelination[33] and in the cortex.[34] MRI may also provide insights into the pathogenesis of the MS lesion. Serial MRI studies demonstrate that a decrease in the magnetization transfer ratio on MRI or the presence of lipid peaks on MRSI in normal-appearing white matter may precede the development of contrast-enhanced lesions, which suggests that the inflammatory response is a secondary factor in the development of a new lesion.[35,36]

Cerebrospinal fluid Most CSF constituents are minimally affected in MS. A mild mononuclear cell pleocytosis can be found during acute relapses, but total cell counts greater than 50 cells/mm^3 are uncommon. The CSF protein level may be mildly

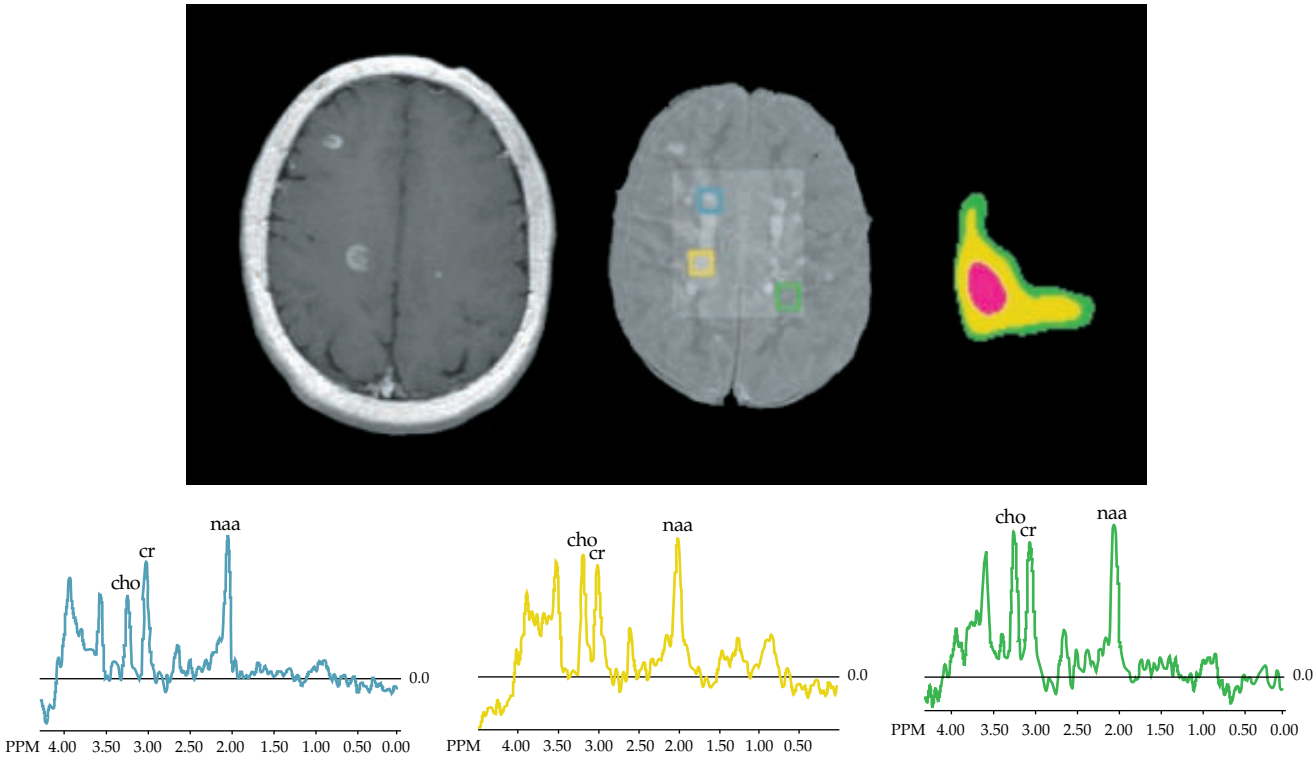


Figure 3 **Two-dimensional proton magnetic resonance spectroscopic imaging (MRSI) was used to investigate an asymptomatic woman with relapsing multiple sclerosis. The central figure is a composite of five 3 mm thick spin density–weighted images through the corpus callosum. A number of scattered lesions are seen, and the rectangular region selected for MRSI is highlighted. To the right is a composite of the same region studied by gadolinium-enhanced T_1-weighted imaging. Three enhancements are seen. The largest enhanced lesion corresponds to the lesion contained within the yellow box in the central spin density image. The blue box in this image contains an unenhanced lesion, and the green box contains an area of normal-appearing white matter. The MRSI is shown next to the spin density image. Here, the maximal regional intensity of resonances attributed to mobile lipids is depicted in red, which corresponds to the region of the enhanced lesion. Relatively less resonance intensity is shown in colors ranging from yellow to green and, finally, to black, indicating the normal lack of lipid resonances from intact white matter. Spectra from individual MRSI voxels that contain the largest enhanced lesion (yellow spectrum), the unenhanced lesion (blue spectrum), and the area of normal-appearing white matter (green spectrum) are shown for comparative detail. The major metabolites that give rise to well-defined spectral peaks include choline (cho), creatine (cr), and *N*-acetyl aspartate (naa). Characteristic paired lipid peaks are seen to the right of the naa peak only in the yellow spectrum. These findings are consistent with active myelin breakdown in association with the enhanced lesion.**

elevated but rarely exceeds 100 mg/dl. During acute attacks, especially those involving the spinal cord and brain stem, the CSF may contain measurable amounts of myelin basic protein. The most characteristic abnormality in MS is intrathecal synthesis of immunoglobulins of restricted heterogeneity. The presence of this abnormality is best determined by comparative electrophoresis of serum and concentrated CSF, which shows oligoclonal immunoglobulin bands specific to CSF. For optimal sensitivity, the paired samples should be analyzed with isoelectric focusing followed by immunofixation. Quantitative measures of immunoglobulin content in CSF, such as the IgG index and the rate of IgG synthesis, are also quite sensitive and useful in clinical practice. Oligoclonal bands or abnormal immunoglobulin synthesis is found in about 90% of patients with clinically definite MS; although not specific to MS, these findings support the diagnosis of MS in equivocal cases. CNS infections or diseases that cause chronic CNS inflammation may also stimulate abnormal immunoglobulin synthesis.

Evoked response The slowing of conduction over demyelinated segments of axons or over incompletely remyelinated pathways provides a useful marker for identifying additional subclinical lesions in sensory pathways. Conduction can be measured along visual, auditory, and somatosensory pathways by use of summated cortical evoked responses. In these tests, a time-locked recording of the electroencephalogram over the afferent cortex of interest is obtained after repeated visual, auditory, or sensory stimulation. If demyelination is significant, conduction over central pathways will be delayed. Evoked responses may be used in MS to reveal evidence of demyelination in a particular sensory pathway when none is clinically evident or to confirm symptomatic pathway involvement in the absence of convincing clinical signs. In general, visual evoked responses provide the most useful information.[37]

MS VARIANTS

A number of clinical or pathologic variants of MS have been identified. Marburg-type MS, or acute MS, is a variant with a fulminant, monophasic course. Schilder disease, or diffuse sclerosis, is a rapidly progressive variant associated with large cerebral demyelinated lesions. Baló disease is a pathologic variant in which concentric areas of demyelination occur in the cerebrum.[38]

Neuromyelitis optica, or Devic disease, is a rare but often severe syndrome that may be distinct from MS. It is characterized by the simultaneous or sequential involvement of the optic nerves and spinal cord with relative sparing of the brain, absence of oligoclonal bands, and longitudinally extensive lesions in the spinal cord on MRI. Demonstration of neuromyelitis optica IgG (NMO-IgG), a serum autoantibody that binds to a component of the cerebral vasculature, may aid in the diagnosis.[39] Treatment of neuromyelitis optica differs from that of typical MS: parenteral corticosteroids and, if necessary, rescue plasmapheresis can be used for acute attacks; systemic immunosuppression, usually with azathioprine and oral corticosteroids, may be recommended for relapse prevention; and rituximab may be beneficial.[40,41]

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of MS depends on the clinical presentation. For a classic case of relapsing-remitting symptoms in a young adult with typical MRI findings, the differential diagnosis is limited. For an older patient presenting with a progressive myelopathy, the differential diagnosis is extensive.[42] There is no standard list of alternative diagnoses that should be considered in every patient suspected of having MS. Instead, the treating physician should carefully consider the clinical and laboratory features of the particular case to generate the relevant list of possible alternative diagnoses. Frequent diagnostic considerations include structural lesions, inherited demyelinating or degenerative diseases, vasculitides, vascular disease, chronic infections (e.g., syphilis, Lyme disease, and human T cell lymphotropic virus type I), vitamin B_{12} deficiency, and neurosarcoidosis.

TREATMENT

Treatment of MS can be discussed in terms of the management of acute relapses, the prevention of relapses as modification of the disease process, and the management of symptoms and fixed neurologic deficits.

Acute Relapse

Management of relapses varies with the severity of the presenting signs and symptoms. Mild attacks that do not significantly alter the patient's ability to function require no more than a supportive physician-patient relationship. High-dose corticosteroid therapy is indicated for exacerbations that adversely affect the patient's function. Intravenous methylprednisolone in daily doses of 0.5 to 1.0 g for 3 to 7 days reduces the duration of maximal neurologic signs and symptoms and usually rapidly reverses the fatigue that frequently accompanies acute attacks.[43,44] A short, tapering course of oral corticosteroids may be given afterward. Equivalent doses of oral corticosteroids may have a similar effect,[45] but treatment with lower doses (e.g., 1 mg/kg/day) is controversial. The early benefits may be similar, but the interval between attacks may be shorter with low-dose steroids.[46,47] The important roles of reduced physical activity during an exacerbation and early institution of rehabilitation therapy should not be ignored. Relative contraindications to corticosteroids include type 1 diabetes mellitus, uncontrolled hypertension, and prior steroid-induced depression or psychoses. Although corticosteroids have a short-term beneficial effect when used for acute exacerbations, their long-term effect on the course of MS is less clear.[48]

Prevention of Relapse

Three different medications that affect the long-term clinical course of MS are available: interferon beta-1b, interferon beta-1a, and glatiramer acetate. In patients with relapsing-remitting MS, these drugs reduce the frequency of attacks, reduce the rate of MS lesion accumulation on MRI, and reduce the accumulation of disability.

The beta interferons reduce major histocompatibility complex (MHC) class II antigen expression, alter the pattern of cytokine secretion, inhibit matrix metalloproteinase activity, and increase antigen-nonspecific suppressor mechanisms in model systems.[49] In clinical trials, they reduce the frequency of clinical attacks, decrease the number of contrast-enhanced lesions, and limit fixed-lesion accumulation on MRI. In early relapsing-remitting MS, they may delay the accumulation of disability. Recombinant human interferon beta-1b and interferon beta-1a vary in their routes of administration, side-effect profiles, and apparent magnitude of effects. These differences are only partially explained by the structural properties of the two molecularly engineered molecules and the study designs of their pivotal clinical trials. Both preparations induce antibodies that may limit clinical benefits.[23]

In an extended controlled study involving patients with re-

lapsing disease of mild to moderate severity, 0.25 mg (8 million IU) of interferon beta-1b, given subcutaneously every other day, reduced the annual exacerbation rate by 30%. This reduction was maintained for 5 years.[23] The frequency of contrast-enhanced lesions was markedly reduced; the total burden of disease, as measured by MRI, increased over time in the placebo group but stabilized for those on high-dose treatment. The benefits of treatment with interferon beta-1b were dose dependent, with better results seen at the higher dose. Antibodies that may limit the effect of the drug developed in up to 38% of treated patients.[50]

The majority of patients experience flulike symptoms of varying severity on initiation of therapy with interferon beta-1b. For most patients, these symptoms can be controlled with prior administration of a nonsteroidal anti-inflammatory drug.[51] The symptoms become less severe and may disappear over time. Local injection-site reactions are usually only of cosmetic concern, but frank skin necrosis can occur. About 20% of patients discontinue treatment because of local or systemic side effects or other issues. In some of those patients who tolerate treatment, therapeutic benefit may decrease as a result of the development of neutralizing antibody. Nevertheless, patients who avoid these pitfalls of treatment stand to benefit substantially; moreover, the benefits are independent of the disability status at initiation of therapy.[52]

Interferon beta-1a is currently available in two forms. One is given intramuscularly once weekly at a dose of 30 μg, and the other is given subcutaneously three times weekly at a dose of 44 μg for a total weekly dose of 132 μg. In patients with slight to mild disability who experience relapse, 30 μg given once a week reduced the relapse rate by 18% and reduced the proportion of patients with sustained progression of neurologic disability. When the same dose of interferon beta-1a was administered within several weeks of symptom onset in patients who had experienced a single attack and were at high risk for having a second, disease-defining attack, the time to the next attack was prolonged.[52] When interferon beta-1a was given subcutaneously at 44 μg three times a week to patients with mild to moderate disability, the relapse rate decreased by 33% and the accumulation of disability slowed.[53] In patients receiving interferon beta-1a, the number of contrast-enhanced lesions decreases and there is an improvement in the MRI burden of disease. In studies of both interferon beta-1a and interferon beta-1b, the beneficial effects were greater with higher doses.[54,55]

Glatiramer is a synthetic random polymer of four amino acids (hence its original name, copolymer-1). Its mechanism of action is not definitively established, but it promiscuously binds to MHC class II antigen and induces organ-specific T helper type 2 cell responses.[56] Glatiramer reduces the frequency of relapses in MS, may reduce accumulation of disability, and does not appear to induce host responses that limit its benefit over time. Glatiramer reduces the annual attack rate by 32%, with the greatest effect seen in those patients with the least neurologic impairment.[57] The progression of disability is also slowed, and benefits may be maintained over 6 or more years.[58] Over the first 9 months of treatment, the number of contrast-enhanced lesions is reduced by 35%, which correlates well with the effect on relapse rate.[59] Side effects are minimal, but a transient systemic reaction occurs in about 15% of patients after one or more injections of the drug. Essentially all patients treated with glatiramer develop antibodies that bind to the drug, but these do not appear to limit the drug's activity.[57]

Another drug for the prevention of relapses, with a mechanism of action different from that of glatiramer, was briefly marketed in the United States from November 2004 to February 2005. Natalizumab is a monoclonal antibody against the adhesion molecule α_4 integrin. The binding of the antibody to its target inhibits the adhesion of circulating lymphocytes to the blood vessel wall and thus prevents the subsequent migration of those lymphocytes into the CNS. When administered intravenously every 4 weeks for a total of six doses, natalizumab significantly reduced the number of new lesions seen on MRI, as well as the number of clinical relapses.[60] In a subsequent, larger study, natalizumab was given every 4 weeks for a year; this regimen resulted in a reduction of the relapse rate by 68% as compared with placebo. The Food and Drug Administration granted accelerated approval on the basis of these promising results. Soon afterward, however, two cases of progressive multifocal leukoencephalopathy, one of them fatal, occurred in patients treated with the combination of natalizumab and interferon. All use of natalizumab is currently suspended, and the future of this drug is uncertain.

In summary, prophylactic treatment with either a beta interferon or glatiramer is appropriate for patients with relapsing disease and mild to moderate disability. The choice of agent depends on the particular patient. Patients who are maintained on these therapies can expect an 18% to 50% reduction in attack frequency. With any of the available drugs, best responses appear to result when treatment is initiated relatively early in the disease course.

These agents have only modest effect when used as monotherapy; consequently, there is intense interest in using these agents in combination to achieve greater benefit. Numerous combinations have been proposed, including glatiramer plus interferon, interferon plus one of the many immunosuppressive agents, and glatiramer plus other agents. Some of these combinations are under active investigation, but to date, no adequate, completed clinical trial has demonstrated increased efficacy with combination therapy. For the present, monotherapy with either interferon or glatiramer is recommended for most patients with relapsing-remitting disease. Combinations of drugs, the use of other immunomodulatory drugs, and several antigen- or disease-specific interventions are currently under active study, and additional therapeutic alternatives may be available in the near future.

Clinically Isolated Syndromes

Neurologists frequently see patients who are having a first episode of neurologic symptoms and whose MRI scans show white-matter lesions. Many of the patients who experience these clinically isolated syndromes will go on to develop MS. Early treatment can delay a second episode in these patients.[61,62] The argument in favor of early treatment is that some of the future relapses, and possibly some of the future disability, can be prevented. Arguments against early treatment are the uncertainty of the diagnosis early in the course, the limited effect of the available treatments, and the possibility of benign disease. The number of patients that are needed to be treated to prevent a single relapse of the disease varies with the number of T_2 lesions found at presentation—a fact that helps guide the physician and patient in making a decision on early initiation of therapy.[63]

Progressive MS

Treatment of progressive MS is not as well established as it is for relapsing disease. As discussed [*see* Clinical Course, *above*],

disease that is progressive from onset is called primary progressive, whereas disease that is initially characterized by relapses and remissions and then enters a progressive phase is called secondary progressive. Results of interferon-beta treatment in secondary progressive MS have been mixed and disappointing. In one trial of interferon beta-1b, disease progressed in 50% of the patients who received placebo, whereas disease progressed in 39% of patients treated with interferon beta-1b.[64] However, subsequent studies of interferon beta-1a and beta-1b have not shown an effect on the progression of disability as measured by the Expanded Disability Status Scale (EDSS), the traditional measure of disability in MS.[65,66] Relapse rates and MRI measures of disease activity were decreased in all of these studies. The available data suggest that glatiramer acetate likewise has limited or no benefit in either secondary or primary progressive disease.[67] Together, these results suggest that the pathophysiologic processes causing progressive symptoms may be different from those causing acute exacerbations and CNS inflammation detected by MRI.

In addition to interferon and glatiramer, various aggressive immunosuppressive therapies have been tried. The cytotoxic drug mitoxantrone has been approved for the treatment of aggressive relapsing and secondary progressive MS. In a 2-year, placebo-controlled study of 194 patients, 12 mg/m^2 of mitoxantrone given every 3 months for 2 years reduced progression of disability, reduced relapse frequency by 67%, and decreased the accumulation of lesions and the number of enhancing lesions on MRI. In this study, the selection criteria were that the patients have active disease with recent deterioration and that they had not had previous therapy with interferon or glatiramer. Observed short-term side effects included nausea, hair loss, menstrual irregularities, and infections. The cumulative dose was limited because of the increasing risk of cardiotoxicity at higher doses.[68]

Mitoxantrone currently has a role for selected patients with very active disease, but its significant side effects limit its use. Several important questions about mitoxantrone remain to be answered. These include the duration of benefit, its effectiveness in patients with progressive disease without relapses, and its effectiveness in patients in whom interferon or glatiramer therapy has failed. Also of concern is the fact that numerous similar cytotoxic or immunosuppressive agents were previously fashionable for a time but are currently little used because of limited benefit and significant side effects. These include cyclophosphamide, cladribine, azathioprine, methotrexate, and cyclosporine.

Symptomatic Therapy

Several of the common symptoms of MS respond to pharmacologic treatment [*see Table 4*]. Prominent among the most frequent symptoms of MS are depression and fatigue.

Depression and fatigue Depression occurs in about 20% of MS patients, probably more frequently than anticipated for other chronic illnesses. Treatment of depression in MS patients does not differ from that in others. Fatigue is present in at least half of all MS patients; often, it is not proportional to the extent of neurologic compromise. It commonly occurs during relapses and can also be present between attacks. Fatigue in MS is differentiated from normal fatigue by its severity and by the fact that it is frequently unrelated to activity. About 50% of patients with MS obtain at least partial relief by treatment with amantadine, but dosing late in the day may induce insomnia and thus should be avoided. When amantadine fails to relieve fatigue, other agents that may be useful are modafanil,[69,70] methylphenidate, and selective serotonin reuptake inhibitors.

Spasticity Spasticity is another common symptom that is amenable to treatment. The increased extensor tone in the lower extremities can to some extent compensate for associated weakness, but excessive spasticity inhibits a fluid gait and may result in painful spasms or joint contracture. Centrally active γ-aminobutyric acid (GABA) agonists, such as baclofen, often reduce spasticity but preserve functional gait or lower-limb strength, which is needed for weight transfer. Effective doses of baclofen vary from as little as 5 mg to as much as 240 mg daily in divided doses. Treatment should be initiated at a low dose and titrated upward to the level that provides maximum benefit. Other drugs that may be helpful include diazepam, clonazepam, and the alpha agonist tizanidine. Although not specifically investigated as an antispasmolytic agent, gabapentin may be useful in reducing spasticity. Some patients benefit from dantrolene. Selected patients who have severe spasticity that is unresponsive to oral treatment may benefit from intrathecal baclofen or selective botulinum toxin injections.

Bladder dysfunction Bladder dysfunction is common in MS. The pathophysiology is complex; the symptoms of detrusor-sphincter dyssynergia, hyperactive detrusor function, and flaccid bladder may occur individually or in various combinations and often fluctuate over time. Formal urodynamic studies are essential to delineate these various patterns and to determine rational approaches to treatment. Most patients will present with urgency and frequency, which may be alleviated with smooth muscle relaxants and anticholinergic agents. Periodic determination of postvoid residual urine volumes is used to identify patients who will benefit from self-catheterization programs.

It is important to recognize that bladder infections are common in women with MS. Because of the frequent loss of perineal sensation, acute bladder infections may not cause dysuria but instead may become evident as a global deterioration of neurologic function, which may be mistaken for an acute relapse. The evaluation of patients in relapse should include a search for pyuria. Asymptomatic bacteriuria may demand concomitant treatment of acute relapses with a combination of methylprednisolone and an appropriate antibiotic. Significant urinary retention with overflow incontinence may be accompanied by reflex exaggeration of lower limb spasticity. Institution of a self-catheterization program for affected patients often reduces spasticity more effectively than pharmacotherapy.

Pain Pain is not uncommon in MS. Often, it is secondary to unusual mechanical stress resulting from asymmetrical weakness or spasticity. Appropriate orthotics, antispasmolytics, and self-performed exercise routines, supplemented by simple analgesics, are helpful. Paroxysmal pain syndromes, typified by trigeminal neuralgia, usually respond to low doses of carbamazepine or other antiepileptic drugs, particularly gabapentin. Distal dysesthetic sensations may also respond to these drugs. Low doses of tricyclic antidepressants may prove useful.

Ataxia and intention tremor These manifestations are particularly difficult to manage. Use of counterweights on limbs may aid some patients. Pharmacologic interventions are generally disappointing, but clonazepam or gabapentin at maximally

Table 4 Symptomatic Therapy for MS*

Indication	*Drug*	*Dosage*
Fatigue	Amantadine	100 mg b.i.d. or t.i.d.
	Methylphenidate	10 mg b.i.d. to 20 mg t.i.d.
	Modafinil	100 mg b.i.d.
Bladder		
Urgency	Oxybutynin	5 mg b.i.d. to q.i.d.
	Tolterodine	2 mg b.i.d.
	Imipramine	25 to 75 mg q.h.s.
	Hyoscyamine	0.125 mg b.i.d. to 0.25 mg q.i.d.
	Propantheline	7.5 mg t.i.d. to 15 mg q.i.d.
Dyssynergia	Phenoxybenzamine	10 mg b.i.d. to 20 mg t.i.d.
	Clonidine	0.1 mg b.i.d. to 0.2 mg t.i.d.
	Terazosin	1 to 5 mg q.d.
Retention	Intermittent catheterization	Four or more times daily
	Bethanechol	10 mg t.i.d. to 50 mg q.i.d.
Spasticity	Baclofen	5 mg t.i.d. to 20 mg q.i.d.
	Diazepam	2 mg t.i.d. to 10 mg q.i.d.
	Tizanidine	4 mg q.d. to 12 mg t.i.d.
	Clonazepam	0.5 mg t.i.d. to 5 mg q.i.d.
	Clonidine (adjunctive to baclofen)	0.1 mg b.i.d. to 0.2 mg t.i.d.
	Dantrolene	25 mg q.d. to 100 mg q.i.d.
Ataxia	Clonazepam	0.5 mg t.i.d. to 5 mg q.i.d.
	Gabapentin	100 to 600 mg t.i.d.
Pain		
Paroxysmal	Carbamazepine	100 to 300 mg t.i.d.
	Phenytoin	300 to 400 mg q.d.
	Misoprostol (trigeminal neuralgia)	100 to 200 μg q.i.d.
Dysesthetic	Amitriptyline	50 to 150 mg q.h.s.
	Phenytoin	300 to 400 mg q.d.
	Gabapentin	100 to 600 mg t.i.d.
	Valproic acid	250 to 1,000 mg t.i.d.

*Usual adult doses for medications commonly used to treat MS syndromes. See appropriate reference for complete prescribing information, including contraindications, warnings, side effects, and initiation and termination of treatment.

tolerated doses sometimes provides symptomatic relief for patients with disabling upper extremity ataxia.

PROGNOSIS

Although the course of MS varies from patient to patient, the effect of the disease in large cohorts of MS patients has been determined. The median time from onset of disease to disability severe enough for the patient to require aids for ambulation is 15 years. MS has minimal effect on life span.[20] For perhaps 10% to 15% of patients, MS has a relatively benign course, with patients experiencing minimal or no disability 20 years after onset of symptoms. For patients with relapsing-remitting disease, the mean relapse frequency is about once every 2 years. There are no known factors that are predictive of the clinical course in an individual patient, but female sex, younger age at onset, and optic neuritis or sensory symptoms as the presenting symptoms tend to be associated with a more favorable prognosis. A normal cerebral MRI in patients presenting with optic neuritis, partial transverse myelitis, and brain stem syndromes associated with MS predicts a mild course for the first decade or more of the disease.[71]

Optic Neuritis

Optic neuritis is an acute inflammatory optic neuropathy. The cardinal symptoms are unilateral vision loss and retrobulbar pain with eye movement. MRI of the orbits is indicated to confirm the presence of optic nerve inflammation; brain MRI results are useful in counseling patients regarding their risk of developing MS. Referral to an ophthalmologist to exclude uveitis, glaucoma, or other ophthalmologic causes of visual loss is indicated. Differential diagnosis includes anterior ischemic optic neuropathy, which is usually painless and typically occurs in patients older than 50 years; hereditary diseases, such as Leber hereditary optic neuropathy; and toxic or nutritional optic neuropathies.[72] Treatment with intravenous methylprednisolone at a dosage of 1 g/day for 3 days followed by oral prednisone for 11 days hastens recovery of vision but has little residual benefit at 1 year. One study showed that prednisone at 1 mg/kg/day for 14 days had no benefit and was associated with an excess of recurrences.[73] Even without treatment, almost all patients begin to recover vision within 4 weeks.

The relation of optic neuritis to MS is controversial. Some regard optic neuritis as a distinct entity, but others consider it part of the clinical continuum of MS. More than half of all patients with MS have optic neuritis at some time during the disease. Of patients who present with optic neuritis and who have no other neurologic deficit, almost 40% have one or more ovoid or periventricular lesions visible on brain MRI; clinically definitive MS eventually develops in 60%.[45,74] Patients with completely normal results on both MRI and comprehensive CSF evaluation seldom progress to MS.[75]

Acute Disseminated Encephalomyelitis

Acute disseminated encephalomyelitis (ADEM) is a monophasic syndrome that is usually preceded by a viral exanthem, an upper respiratory infection, or vaccination. The most commonly associated viruses are measles, paramyxovirus, varicella, rubella, and Epstein-Barr virus. Onset is often rapid and is characterized by meningeal signs, headache, seizures, and altered mental status. The associated neurologic deficits are variable and may include hemiplegia, paraplegia, sensory loss, vision loss, and transverse myelitis. ADEM can be fatal, but most patients begin to recover within 2 to 4 weeks.

Acute hemorrhagic encephalomyelitis is probably a fulminant variant of ADEM. The main pathologic features of ADEM are multiple areas of perivascular inflammation and demyelination, without evidence of active viral infection. ADEM may be caused by an autoimmune response against myelin antigens elicited by cross-reactive viral proteins. Usually, multiple white-matter lesions are seen on MRI, and the majority of the lesions can be enhanced with contrast.[76] Corticosteroid treatment is often used, although the efficacy of this approach has not been proved in clinical trials. Plasmapheresis may also be useful.[77] Prognosis varies with the inciting virus, but in one case series, 89% of patients had good recovery.[78]

Transverse Myelitis

Acute transverse myelitis is a syndrome of spinal cord dysfunction[79]; it has a rapid onset. Like ADEM, it may occur after infection or vaccination, or it may occur with no discernible precipitant. It may also be the initial presentation of MS. Symptoms include paraparesis, which is initially flaccid and then spastic; loss of sensation with a sensory level on the trunk; and bowel and bladder dysfunction. Back pain precedes the neurologic symptoms, and the sensory symptoms may begin distally and ascend. The thoracic cord is most often affected. The differential diagnosis includes other causes of acute myelopathy, such as compression of the cord by an extradural structural lesion, spinal cord neoplasms, ischemia, and systemic lupus erythematosus. MRI is extremely useful for excluding structural lesions and for confirming the presence of an intramedullary lesion at the level in the spinal cord commensurate with the symptoms. The lesions of acute transverse myelitis are typically hyperintense on T_2-weighted imaging; they involve the majority of the cross-sectional area of the cord over several segments and may be enhanced with contrast. The lesions may cause swelling of the spinal cord.[80,81] No treatment has proved beneficial, but corticosteroids are often used. Prognosis is variable: one third of patients have a good outcome, one third have a fair outcome, and one third do not recover.[82] Spinal shock, back pain, and catastrophic onset are associated with poor outcome.

Inherited Demyelinating Diseases

ADRENOLEUKODYSTROPHY

Adrenoleukodystrophy is an inherited disorder that is associated with progressive demyelination and dysfunction of the adrenal cortex.[83] The inheritance pattern may be either autosomal recessive or X-linked recessive. The X-linked form is caused by the mutation of a gene encoding an integral membrane protein found in the peroxisome. Defects in this gene lead to accumulation of very long chain fatty acids (VLCFAs). The phenotypes may vary considerably, even within the same family. In the childhood form, the patient presents with cognitive deficits, and rapid neurologic deterioration ensues, with death occurring in 2 to 5 years. The adult form, called adrenomyeloneuropathy, presents in patients at a mean age of 28 years as progressive spinal cord dysfunction with spastic paraparesis, sensory loss, and bowel and bladder symptoms. Cerebral involvement may be minimal. Only half of patients with adult-onset disease have brain abnormalities on MRI; these are most often found in the corticospinal tracts.[84] Most patients have diffuse atrophy of the spinal cord. Diagnosis is made on the basis of the combination of neurologic and adrenal involvement, family history, and elevated levels of serum VLCFAs. Dietary treatment with unsaturated fatty acids lowers the level of VLCFAs but does not significantly affect the progression of symptoms.[85] Bone marrow transplantation may be effective if performed before severe symptoms develop. Prognosis is poor for patients with the childhood form of disease. Patients with adult-onset disease usually require assistance with ambulation within 10 to 15 years; in a large percentage of patients, rapidly progressive cerebral lesions develop 5 to 10 years after the onset of spinal cord symptoms.[86]

METACHROMATIC LEUKODYSTROPHY

Metachromatic leukodystrophy is an autosomal recessive disorder that results in demyelination of axons in the central and peripheral nervous systems. It is caused by mutations in the gene for arylsulfatase A that lead to an accumulation of metachromatically staining sulfatides.[87] Onset usually occurs in infancy or childhood; adult onset is rare. The symptoms of adult-onset disease are progressive behavioral abnormalities, dementia, ataxia, and neuropathy.[88] MRI or CT of the brain demonstrates atrophy and diffuse white-matter abnormalities, particularly in the frontal lobes. Diagnosis is confirmed by measurement of arylsulfatase A activity in peripheral blood leukocytes, urine, or skin fibroblasts. True arylsulfatase deficiency must be distinguished from a common pseudodeficiency state that is caused by an allele with low enzymatic activity.[87] The symptoms of metachromatic leukodystrophy are relentlessly progressive, and earlier onset is associated with more rapid progression. The mean survival for adult-onset disease is about 12 years. No effective treatment is available, but allogeneic bone marrow transplantation and gene therapy are under investigation.

Metabolic Demyelinating Diseases

CENTRAL PONTINE MYELINOLYSIS

Central pontine myelinolysis (CPM) is a syndrome in which neurologic deficits occur after rapid correction of hyponatremia [*see 10:I Renal Function and Disorders of Water and Sodium Balance*].[89] CPM usually occurs in young to middle-aged adults and is often associated with alcohol abuse or malnutrition. Signs and symptoms usually begin 3 days after the start of sodium replacement and consist of changes in mental status, dysarthria and other signs of corticobulbar dysfunction, and spastic quadriplegia. Improvement usually begins about 2 weeks after the onset of symptoms, but the degree of recovery is variable. The most striking finding on pathologic examination is the presence of symmetrical demyelinated lesions in the central pons. Demyelinated lesions may also occur in a relatively symmetrical pattern in the basal ganglia, thalamus, internal capsule, subcortical white matter, and cerebellum. T_2-weighted MRI usually

demonstrates the presence of hyperintense lesions. These lesions usually do not enhance with contrast. CPM may also occur after liver transplantation. There is no specific treatment once symptoms have developed. Long duration and rapid correction of hyponatremia increase the risk of CPM; the recommended rate for correction of hyponatremia is no faster than 10 to 12 mEq in 24 hours.

VITAMIN B_{12} DEFICIENCY

Vitamin B_{12} deficiency results in demyelination of axons in the central and peripheral nervous systems. The dorsal and lateral white-matter tracts of the spinal cord are most affected—a characteristic that has given rise to the name subacute combined degeneration of the spinal cord. The most common presenting symptoms are paresthesias, sensory loss that begins in the feet and progresses proximally, and sensory ataxia.[90] Weakness almost always begins after sensory loss. Memory difficulties, irritability, and confusion occur in a minority of patients. On examination, patients usually have decreased vibration and position sense, which is worse in the feet than in the hands, and they may have spastic paraparesis. Pathologic examination reveals symmetrical loss of myelin in the posterior and lateral columns of the spinal cord and sometimes patchy demyelination in the cerebral white matter. MRI of the spinal cord often demonstrates white-matter lesions, which resolve with treatment. Diagnosis is made on the basis of the clinical findings and a low serum cobalamin level. Macrocytosis or anemia is present in most patients but cannot be used in place of the cobalamin level as a diagnostic measure.[91] For patients who have symptoms and a low-normal cobalamin level, demonstration of elevated levels of serum methylmalonic acid and total homocysteine can confirm the presence of a functionally significant cobalamin deficiency.

If cobalamin deficiency is present, the underlying etiology should be investigated. About 80% of patients with cobalamin deficiency have pernicious anemia. Administration of cobalamin prevents progression of symptoms and produces clinical improvement in most patients [*see 5:III Anemia: Production Defects*].[90] Nitrous oxide prevents the metabolism of cobalamin and can cause similar symptoms after prolonged exposure; after a single exposure, it can unmask a subclinical cobalamin deficiency.

Virus-Induced Demyelination

PROGRESSIVE MULTIFOCAL ENCEPHALOPATHY

Progressive multifocal encephalopathy is a lethal demyelinating disease caused by an opportunistic viral infection of oligodendrocytes in immunocompromised patients. The causative agent is JC virus, a ubiquitous papovavirus that infects the majority of the population before adulthood and establishes a latent infection in the kidney. In immunocompromised hosts, the virus can reactivate and productively infect oligodendrocytes. This previously rare condition is now more common because it occurs in 4% of patients with AIDS. It appears to complicate natalizumab treatment of both MS and Crohn disease. Patients usually present with relentlessly progressive focal neurologic deficits, such as hemiparesis or visual-field deficits, or with alterations in mental status. On brain MRI, one or more white-matter lesions are present; they are hyperintense on T_2-weighted images and hypointense on T_1-weighted images. There is no mass effect, and contrast enhancement is rare. Diagnosis can be confirmed by brain biopsy, with demonstration of virus by in situ hybridization or immunocytochemistry. Polymerase chain reaction amplification of JC virus sequences from the CSF can confirm diagnosis without the need for biopsy.[92] Currently, there is no effective therapy. Survival after diagnosis is about 3 to 5 months in AIDS patients [*see 11:XVII Central Nervous System Diseases Due to Slow Viruses and Prions*].

SUBACUTE SCLEROSING PANENCEPHALITIS

Subacute sclerosing panencephalitis (SSPE) is a rare late complication of measles virus infection. It occurs most often in patients whose initial infection with measles virus occurred before 2 years of age; the mean lag time between initial infection and SSPE is 7 years. The use of measles vaccine has greatly reduced the incidence of this complication in developed countries. The earliest symptom is usually progressive cognitive deterioration, which is followed by motor dysfunction and myoclonus associated with distinctive electroencephalographic abnormalities. Pathologic examination reveals active viral infection in the brain, with measles virus protein and RNA detectable in both oligodendrocytes and neurons, and a vigorous inflammatory response. The course is progressive, with occasional temporary remissions. There is no satisfactory treatment [*see 11:XVII Central Nervous System Diseases Due to Slow Viruses and Prions*].

J. William Lindsey, M.D., has received speaking honoraria from Teva Pharmaceuticals Ltd. and Serono.

Jerry S. Wolinsky, M.D., has served as a consultant or speaker for Teva Pharmaceutical Ltd., Serono, and Schering-Plough Corp.

References

1. Noonan CW, Kathman SJ, White MC: Prevalence estimates for MS in the United States and evidence of an increasing trend for women. Neurology 58:136, 2002

2. Kurtzke JF: Epidemiology of multiple sclerosis: does this really point toward an etiology? Lectio Doctoralis. Neurol Sci 21:383, 2000

3. Evans CF, Horwitz MS, Hobbs MV, et al: Viral infection of transgenic mice expressing a viral protein in oligodendrocytes leads to chronic central nervous system autoimmune disease. J Exp Med 184:2371, 1996

4. Willer CJ, Dyment DA, Risch NJ, et al: Twin concordance and sibling recurrence rates in multiple sclerosis. Proc Natl Acad Sci USA 100:12877, 2003

5. Factors influencing sib risks for multiple sclerosis. The Canadian Collaborative Study Group. Clin Genet 58:431, 2000

6. Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group. Lancet 347:1728, 1996

7. Lindsey JW: Familial recurrence rates and genetic models of multiple sclerosis. Am J Med Genet A 135:53, 2005

8. Masterman T, Ligers A, Olsson T, et al: HLA-DR15 is associated with lower age at onset in multiple sclerosis. Ann Neurol 48:211, 2000

9. Genetic basis for clinical expression in multiple sclerosis. The Multiple Sclerosis Genetics Group. Brain 125:150, 2002

10. Dyment DA, Willer CJ, Scott B, et al: Genetic susceptibility to MS: a second stage analysis in Canadian MS families. Neurogenetics 3:145, 2001

11. Brueck W, Schmied M, Suchanek G, et al: Oligodendrocytes in the early course of multiple sclerosis. Ann Neurol 35:65, 1994

12. Bo L, Vedeler CA, Nyland HI, et al: Subpial demyelination in the cerebral cortex of multiple sclerosis patients. J Neuropathol Exp Neurol 62:723, 2003

13. Bitsch A, Kuhlmann T, Da Costa C, et al: Tumour necrosis factor alpha mRNA expression in early multiple sclerosis lesions: correlation with demyelinating activity and oligodendrocyte pathology. Glia 29:366, 2000

14. Trapp BD, Peterson J, Ransohoff RM, et al: Axonal transection in the lesions of multiple sclerosis. N Engl J Med 338:278, 1998

15. Matthews PM, De Stefano N, Narayanan S, et al: Putting magnetic resonance spectroscopy studies in context: axonal damage and disability in multiple sclerosis. Semin Neurol 18:327, 1999

16. van Waesberghe JH, Kamphorst W, De Groot CJ, et al: Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability. Ann Neurol 46:747, 1999

17. Lucchinetti C, Brueck W, Parisi J, et al: Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. Ann Neurol 47:707, 2000

18. Lucchinetti CF, Mandler RN, McGavern D, et al: A role for humoral mechanisms in the pathogenesis of Devic's neuromyelitis optica. Brain 125:1450, 2002

19. McDonald WI, Compston DAS, Edan G, et al: Recommended diagnostic criteria for

multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 50:121, 2001

20. Runmarker B, Andersen O: Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. Brain 116:117, 1993

21. Revesz T, Kidd D, Thompson AJ, et al: A comparison of the pathology of primary and secondary progressive multiple sclerosis. Brain 117:759, 1994

22. Stone LA, Albert PS, Smith ME, et al: Changes in the amount of diseased white matter over time in patients with relapsing-remitting multiple sclerosis. Neurology 45:1808, 1995

23. Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. The IFNB Multiple Sclerosis Study Group, University of British Columbia MS/MRI Analysis Group. Neurology 45:1277, 1995

24. Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. The IFNB Multiple Sclerosis Study Group, University of British Columbia MS/MRI Analysis Group. Lancet 352:1498, 1998

25. Katz D, Taubenberger JK, Cannella B, et al: Correlation between magnetic resonance imaging findings and lesion development in chronic, active multiple sclerosis. Ann Neurol 34:661, 1993

26. Truyen L, van Waesberghe JHTM, van Walderveen MAA, et al: Accumulation of hypointense lesions ("black holes") on T_1 spin-echo MRI correlates with disease progression in multiple sclerosis. Neurology 47:1469, 1996

27. Use of the brain parenchymal fraction to measure brain atrophy in relapsing-remitting MS. The Multiple Sclerosis Collaborative Research Group. Neurology 53:1698, 1999

28. Linomide in relapsing and secondary progressive multiple sclerosis: MRI results. The North American Linomide Investigators. Neurology 54:1734, 2000

29. Filippi M, Horsfield MA, Ader HJ, et al: Guidelines for using quantitative measures of brain magnetic resonance imaging abnormalities in monitoring the treatment of multiple sclerosis. Ann Neurol 43:499, 1998

30. Filippi M, Grossman RI: MRI techniques to monitor MS evolution: the present and the future. Neurology 58:1147, 2002

31. De Stefano N, Matthews PM, Antel JP, et al: Chemical pathology of acute demyelinating lesions and its correlation with disability. Ann Neurol 38:901, 1995

32. De Stefano N, Narayanan S, Matthews PM, et al: In vivo evidence for axonal dysfunction remote from focal cerebral demyelination of the type seen in multiple sclerosis. Brain 122:1933, 1999

33. Narayana PA, Wolinsky JS, Jackson EF, et al: Proton MR spectroscopy of gadolinium-enhanced multiple sclerosis plaques. J Magn Reson Imaging 2:263, 1992

34. Sharma R, Narayana PA, Wolinsky JS: Gray matter abnormalities in multiple sclerosis: proton magnetic resonance spectroscopic imaging. Mult Scler 7:221, 2001

35. Filippi M, Rocca MA, Martino G, et al: Magnetization transfer changes in the normal appearing white matter precede the appearance of enhancing lesions in patients with multiple sclerosis. Ann Neurol 43:809, 1998

36. Narayana PA, Doyle TJ, Lai DJ, et al: Serial proton magnetic resonance spectroscopic imaging, contrast-enhanced magnetic resonance imaging, and quantitative lesion volumetry in multiple sclerosis. Ann Neurol 43:56, 1998

37. Gronseth GS, Ashman EJ: Practice parameter: the usefulness of evoked potentials in identifying clinically silent lesions in patients with suspected multiple sclerosis (an evidence-based review). Neurology 54:1720, 2000

38. Noseworthy JN, Lucchinetti C, Rodriguez M, et al: Multiple sclerosis. N Engl J Med 343:938, 2000

39. Lennon VA, Wingerchuk DM, Kryzer TJ, et al: A serum autoantibody marker of neuromyelitis optica: distinction from multiple sclerosis. Lancet 364:2106, 2004

40. Wingerchuk DM, Weinshenker BG: Neuromyelitis optica. Curr Treat Options Neurol 7:173, 2005

41. Cree BA, Lamb S, Morgan K, et al: An open label study of the effects of rituximab in neuromyelitis optica. Neurology 64:1270, 2005

42. Wolinsky JS, PROMiSe Study Group: The diagnosis of primary progressive multiple sclerosis. J Neurol Sci 206:145, 2003

43. Durelli L, Cocito D, Ficcio A, et al: High-dose intravenous methylprednisolone in the treatment of multiple sclerosis: clinical-immunologic correlations. Neurology 36:238, 1986

44. Milligan NM, Newcombe R, Compston DAS: A double-blind controlled trial of high dose methylprednisolone in patients with multiple sclerosis: 1. Clinical effects. J Neurol Neurosurg Psychiatry 50:511, 1987

45. Alam SM, Kyriakides T, Lawden M, et al: Methylprednisolone in multiple sclerosis: a comparison of oral with intravenous therapy at equivalent high dose. J Neurol Neurosurg Psychiatry 56:1219, 1993

46. Barnes D, Hughes RAC, Morris RW, et al: Randomised trial of oral and intravenous methylprednisolone in acute relapses of multiple sclerosis. Lancet 349:902, 1997

47. Beck RW, Cleary PA, Trobe JD, et al: The effect of corticosteroids for acute optic neuritis on the subsequent development of multiple sclerosis. N Engl J Med 329:1764, 1993

48. Zivadinov R, Rudick RA, De Masi R, et al: Effects of IV methylprednisolone on brain atrophy in relapsing-remitting MS. Neurology 57:1239, 2001

49. Dhib-Jalbut S: Mechanisms of action of interferons and glatiramer acetate in multiple sclerosis. Neurology 58(8 suppl 4):S3, 2002

50. Duquette P, Girard M, Dubois R, et al: Neutralizing antibodies during treatment of multiple sclerosis with interferon beta-1b: experience during the first three years. Neurology 47:889, 1996

51. Lublin FD, Whitaker JN, Eidelman BH, et al: Management of patients receiving interferon beta-1b for multiple sclerosis: report of a consensus conference. Neurology 46:12, 1996

52. Jacobs LD, Beck RW, Simon JH, et al: Intramuscular interferon beta-1a therapy initiated during the first demyelinating event in multiple sclerosis. N Engl J Med 343:898, 2000

53. PRISMS-4: Long-term efficacy of interferon-beta-1a in relapsing MS. PRISMS Study Group, University of British Columbia MS/MRI Analysis Group. Neurology 56:1628, 2001

54. Evidence of interferon B-1a dose response in relapsing-remitting MS. Once Weekly Interferon for MS Study Group (OWIMS). Neurology 53:679, 1999

55. Panitch H, Goodin DS, Francis G, et al: Randomized, comparative study of interferon beta-1a treatment regimens in MS: the EVIDENCE trial. Neurology 59:1496, 2002

56. Wolinsky JS: Glatiramer acetate for the treatment of multiple sclerosis. Expert Opin Pharmacother 5:875, 2004

57. Johnson KP, Brooks BR, Cohen JA, et al: Extended use of glatiramer acetate (Copaxone) is well tolerated and maintains its clinical effect on multiple sclerosis relapse rate and degree of disability. Neurology 50:701, 1998

58. Johnson KP, Brooks BR, Ford CC, et al: Sustained clinical benefits of glatiramer acetate in relapsing multiple sclerosis patients observed for 6 years. Mult Scler 6:255, 2000

59. European/Canadian multicenter, double-blind, randomized, placebo controlled study of the effects of glatiramer acetate on magnetic resonance imaging–measured disease activity and burden in patients with relapsing multiple sclerosis. European/Canadian Glatiramer Acetate Study Group. Ann Neurol 49:290, 2001

60. Miller DH, Khan OA, Sheremata WA, et al: A controlled trial of natalizumab for relapsing multiple sclerosis. N Engl J Med 348:15, 2003

61. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. CHAMPS Study Group. N Engl J Med 343:898, 2000

62. Comi G, Filippi M, Barkhof F, et al: Effect of early interferon treatment on conversion to definite multiple sclerosis: a randomised study. Lancet 357:1576, 2001

63. Barkhof F, Rocca M, Francis G, et al: Validation of diagnostic magnetic resonance imaging criteria for multiple sclerosis and response to interferon beta1a. Ann Neurol 53:718, 2003

64. Randomized controlled trial of interferon-beta-1a in secondary progressive MS: clinical results. Secondary Progressive Efficacy Clinical Trial of Recombinant Interferon-beta-1a in MS (SPECTRIMS) Study Group. Neurology 56:1496, 2001

65. Benefit of interferon beta-1a on MSFC progression in secondary progressive MS. The IMPACT Investigators. Neurology 59:679, 2002

66. Interferon beta-1b in secondary progressive MS: clinical and MRI results of a 3-year randomized controlled trial (abstr). The North American Study Group on Interferon beta-1b in Secondary Progressive MS. Neurology 54:2352, 2000

67. Bornstein MB, Miller A, Slagle S, et al: A placebo-controlled, double-blind, randomized, two-center, pilot trial of Cop 1 in chronic progressive multiple sclerosis. Neurology 41:533, 1991

68. Ghalie RG, Edan G, Laurent M, et al: Cardiac adverse effects associated with mitoxantrone (Novantrone) therapy in patients with MS. Neurology 59:909, 2002

69. Rammohan KW, Rosenberg JH, Lynn DJ, et al: Efficacy and safety of modafinil (Provigil) for the treatment of fatigue in multiple sclerosis: a two centre phase 2 study. J Neurol Neurosurg Psychiatry 72:179, 2002

70. Stankoff B, Waubant E, Confavreux C, et al: Modafinil for fatigue in MS: a randomized placebo-controlled double-blind study. French Modafinil Study Group. Neurology 64:1139, 2005

71. Brex PA, Ciccarelli O, O'Riordan JL, et al: A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. N Engl J Med 346:158, 2002

72. Newman NJ: Optic neuropathy. Neurology 46:315, 1996

73. Beck RW, Cleary PA, Anderson MM, et al: A randomized, controlled trial of corticosteroids in the treatment of acute optic neuritis. N Engl J Med 326:581, 1992

74. Rodriguez M, Siva A, Cross SA, et al: Optic neuritis: a population-based study in Olmsted County, Minnesota. Neurology 45:244, 1995

75. Soederstroem M, Ya-Ping J, Hillert J: Optic neuritis: prognosis for multiple sclerosis from MRI, CSF, and HLA findings. Neurology 50:708, 1998

76. Mader I, Stock KW, Ettlin T, et al: Acute disseminated encephalomyelitis: MR and CT features. AJNR Am J Neuroradiol 17:104, 1996

77. Lin CH, Jeng JS, Yip PK: Plasmapheresis in acute disseminated encephalomyelitis. J Clin Apheresis 19:154, 2004

78. Tenembaum S, Chamoles N, Fejerman N: Acute disseminated encephalomyelitis: a long-term follow-up study of 84 pediatric patients. Neurology 59:1224, 2002

79. Proposed diagnostic criteria and nosology of acute transverse myelitis. Transverse Myelitis Consortium Working Group. Neurology 59:499, 2002

80. Choi KH, Lee KS, Chung SO, et al: Idiopathic transverse myelitis: MR characteristics. AJNR Am J Neuroradiol 17:1151, 1996

81. Tartaglino LM, Croul SE, Flanders AE, et al: Idiopathic acute transverse myelitis: MR imaging findings. Radiology 201:661, 1996

82. Christensen PB, Wermuth L, Hinge HH, et al: Clinical course and long-term prognosis of acute transverse myelopathy. Acta Neurol Scand 81:431, 1990

83. Moser HW: Clinical and therapeutic aspects of adrenoleukodystrophy and adrenomyeloneuropathy. J Neuropathol Exp Neurol 54:740, 1995

84. Kumar AJ, Kohler W, Kruse B, et al: MR findings in adult-onset adrenoleukodystrophy. AJNR Am J Neuroradiol 16:1227, 1995

85. Aubourg P, Adamsbaum C, Lavallard-Rousseau MC, et al: A two-year trial of oleic and erucic acids ("Lorenzo's oil") as treatment for adrenomyeloneuropathy. N Engl J Med 329:745, 1993

86. Eichler FS, Barker PB, Cox C, et al: Proton MR spectroscopic imaging predicts lesion progression on MRI in X-linked adrenoleukodystrophy. Neurology 58:901, 2002

87. Gieselmann V, Polten A, Kreysing J, et al: Molecular genetics of metachromatic leukodystrophy. J Inherit Metab Dis 17:500, 1994

88. Hageman AT, Gabreels FJ, de Jong JG, et al: Clinical symptoms of adult metachromatic leukodystrophy and arylsulfatase A pseudodeficiency. Arch Neurol 52:408, 1995

89. Pirzada NA, Ali II: Central pontine myelinolysis. Mayo Clin Proc 76:559, 2001

90. Healton EB, Savage DG, Brust JC, et al: Neurologic aspects of cobalamin deficiency. Medicine (Baltimore) 70:229, 1991

91. Lindenbaum J, Healton EB, Savage DG, et al: Neuropsychiatric disorders caused by cobalamin deficiency in the absence of anemia or macrocytosis. N Engl J Med 318:1720, 1988

92. Antinori A, Ammassari A, DeLuca A, et al: Diagnosis of AIDS-related focal brain lesions: a decision-making analysis based on clinical and neuroradiologic characteristics combined with polymerase chain reaction assays in CSF. Neurology 48:687, 1997

93. Waxman SG, Ritchie JM: Molecular dissection of the myelinated axon. Ann Neurol 33:121, 1993

94. Shepherd DI: Clinical features of multiple sclerosis in northeast Scotland. Acta Neurol Scand 60:218, 1979

95. Posner S, Wikstrom J, Bauer HJ: Clinical data and the identification of special forms of multiple sclerosis in 1271 cases studied with a standardized documentation system. J Neurol Sci 40:159, 1979

Acknowledgment

Figure 1 Seward Hung.

X INHERITED ATAXIAS

S. H. Subramony, m.d.

The inherited ataxias are disorders that cause progressive imbalance as a result of pathology in the cerebellum and its various connecting pathways. Ever since Nicholas Friedreich's initial recognition, more than 100 years ago, that ataxia may result from genetic mechanisms, clinicians and investigators have been struck by the variability and complexity of the clinical, neuropathologic, and genetic aspects of these disorders. In addition to their difficulties with balance, persons with inherited ataxias often exhibit a variety of clinical signs that indicate dysfunction in the basal ganglia, the upper motor neuron, the oculomotor and other brain stem nuclei, the anterior horn cells, and the dorsal root ganglia. In some instances, vision loss and cognitive decline may occur as well.

Within the past decade, molecular genetic studies have identified specific gene mutations in many inherited ataxias and have localized the cause of others to specific chromosomal loci. This has led to the following advances:

1. Classification of these disorders on the basis of genotype.
2. Ability to identify the precise genotype in many patients.
3. Availability of predictive and prenatal diagnostic studies.
4. An understanding of some of the phenotypic variability related to mutations in the same gene.
5. Growing knowledge about the pathogenetic mechanisms of neuronal loss and clinical disease in the setting of specific gene mutations.
6. Early attempts at specific drug therapies in some of the diseases, based on this evolving knowledge.

Genotypic Classification

Within the past dozen years, the surprising genetic heterogeneity of the inherited ataxias has been amply documented. Some autosomal recessive and autosomal dominant ataxias have been defined at a molecular level [*see Tables 1 and 2*]. However, some early-onset ataxic syndromes remain poorly understood at a genetic level [*see Table 3*]. In addition, approximately a third of all autosomal dominant ataxias still have undiscovered gene loci. There are also ataxic disorders that have either mitochondrial or X-linked inheritance patterns.

Autosomal Recessive Ataxias

Friedreich ataxia

Friedreich ataxia (FA) is the most common type of recessively inherited ataxia; it has a prevalence of 2×10^{-5}.[1]

Genetics and Pathogenesis

The mutation in FA is an unstable expansion of a GAA trinucleotide repeat in the first intron of the gene *X-25*, located on chromosome 9q [*see Figure 1*]. Normal alleles have fewer than 40 GAA repeats; over 80% of these are short normal alleles, which carry six to 10 repeats.[1] Long normal alleles, which carry more than 12 repeats, may serve as a reservoir for expansion into pathogenic alleles; they appear to be confined to Indo-Caucasian populations. Enlarged alleles carry from 66 to over 1,000 repeats.[2-5] Most patients with FA are homozygous for the trinucleotide expansion; fewer than 5% of patients with typical FA, however, carry only one expanded allele coupled with a point mutation in the second allele.[6]

The presence of the expanded GAA sequence results in reduced transcriptional and translational efficiency of the gene, which leads to a deficiency of frataxin, a nuclearly encoded mitochondrial protein. This effect appears to be related to so-called sticky DNA, an unusual DNA configuration induced by the expansion.[7] In yeast, lack of the yeast frataxin homologue (YFH) causes a variety of disturbances, including reduced respiratory efficiency of mitochondria, loss of mitochondrial

Table 1 Autosomal Recessive Ataxias with Known Gene Loci

Disease	*Gene Locus*	*Gene*	*Mutation*
Friedreich ataxia	9q13-21.1	*X25*	GAA expansion
Ataxia-telangiectasia	11q22-23	*ATM*	Point mutations/deletions
Ataxia with isolated vitamin E deficiency	8q	*α-TTP*	Point mutations
Autosomal recessive ataxia of Charlevoix-Saguenay (ARSACS)	13q11	*SACS*	Point mutations
Ataxia with oculomotor apraxia	9p13	*Aprataxin*	Point mutations/deletions/insertions
Ataxia, neuropathy, high α-fetoprotein	9q33-34	Unknown	Unknown
Infantile-onset olivopontocerebellar atrophy (IOSCA)	10q24	Unknown	Unknown
Ataxia, deafness, optic atrophy	6p21-23	Unknown	Unknown
Unverricht-Lundborg disease	21q	*Cystatin B*	Repeat expansion
Ataxia-telangiectasia–like disorder (ATLD)	11q21	*MRE11*	Point mutations
Spinocerebellar ataxia with axonal neuropathy (SCAN-1)	14q31	*TDP1*	Point mutations, insertions, deletions

Table 2 Autosomal Dominant Ataxias with Known Gene Loci

Disease	*Locus*	*Gene*	*Mutation*
SCA1	6p23	*Ataxin1*	CAG expansion
SCA2	12q23-24.1	*Ataxin2*	CAG expansion
MJD (SCA3)	14q21	*Ataxin3*	CAG expansion
SCA4	16q24ter	*	*
SCA5	11p11-q11	*	*
SCA6	19p	*CACNA1*	CAG expansion
SCA7	3p21.2-12	*Ataxin 7*	CAG expansion
SCA8	13q21	*Ataxin 8*	CTG expansion
SCA10	22q13	*Ataxin 10*	ATTCT expansion
SCA11	15q14-21.3	*	*
SCA12	5q31-33	*PPP2R2B*	CAG expansion
SCA13	10q13.3-13.4	*	*
SCA14	19q13.4	*PKCγ*	Point mutations
SCA16	8q23-24.1	*	*
SCA17	6q21ter	*TBP*	CAG expansion
SCA18	7q31-32	*	*
SCA19	1p21-q21	*	*
SCA21	7p21	*	*
SCA22	1p21-q23	*	*
SCA23	20p13	*	*
DRPLA	12p	*Atrophin*	CAG expansion
EA-1	12p	*KCNA1*	Point mutations
EA-2	19p	*CACNA1*	Point mutations

*Unknown
DRPLA—dentatorubral pallidoluysian atrophy EA—episodic ataxia
MJD—Machado-Joseph disease TBP—TATA-binding protein
SCA—spinocerebellar ataxia PKCγ—protein kinase Cγ

DNA, excess intramitochondrial iron, and increased susceptibility to oxidative stress.[8,9] A similar process may occur in human tissues. Iron accumulation has been noted in cardiac myocytes at autopsy and in the cerebellar dentate nuclei on magnetic resonance imaging.[10] Endomyocardial biopsies from patients with FA have shown a reduction in complexes I, II, and III, as well as levels of the enzyme aconitase. These enzymes share a common feature: the presence of iron-sulfur (Fe-S) clusters. Fe-S cluster proteins are especially susceptible to oxidative stress, which can be triggered by excess iron. It is also possible that the deficiency of Fe-S cluster enzymes may be related directly to frataxin deficiency.

Diagnosis

Clinical features The onset of symptoms in patients with FA is usually in the first or second decade of life but may be delayed to the third decade or later. The classic descriptions of the disease include progressive gait ataxia with onset before 25 years of age, loss of deep tendon reflexes, and proprioceptive loss in the limbs—all signs of major early pathology in the dorsal root ganglion cells and their peripheral sensory processes. Other signs include dysarthria, extensor plantar responses, and oculomotor abnormalities such as square-wave jerks. Muscle atrophy, weakness, and dysphagia occur late in the disease. Tremor, vision loss, and hearing loss may occur in a few patients. About 30% to 50% of patients develop symptomatic heart disease, including hypertrophic cardiomyopathy. Diabetes occurs in 10% of patients, and skeletal deformities, such as scoliosis, are frequent.

The FA GAA expansion has been shown to be associated with a more variable phenotype than the classic features (see above).[2-5] For example, the onset may be at a much older age (up to the early 50s). In addition, ataxia associated with preserved and, occasionally, brisk reflexes can occur. These phenotypes have been labeled late-onset Friedreich ataxia (LOFA) and Friedreich ataxia with retained reflexes (FARR), respectively. Age at onset is inversely correlated with the size of the GAA expansion, especially that of the smaller allele. Patients with heterozygous expansion coupled with a point mutation also often have atypical phenotypes with slower disease progression, especially when the point mutation occurs in the amino-terminal half of the protein.[6]

Laboratory tests One laboratory test that is of value is the FA mutation analysis. More than 95% of patients with typical FA will have a homozygous GAA expansion, but a heterozygous expansion is also diagnostic; in such cases, it can be presumed that the unexpanded allele has a point mutation. If the clinical picture is atypical, however, a heterozygous expansion may simply reflect an incidental carrier state. The FA mutation analysis is also indicated in almost all cases of recessive or sporadic ataxia of childhood or adult onset that cannot be readily categorized otherwise. Two other useful laboratory tests are nerve conduction tests, which show a predominantly sensory neuropathy, and brain imaging, which shows spinal cord atrophy with a relatively preserved cerebellum.

Differential Diagnosis

The phenotype of typical childhood-onset FA is usually not difficult to recognize. The major differential diagnoses include Charcot-Marie-Tooth disease and pure sensory neuropathies, as well as certain metabolic errors that can mimic FA disease, such as vitamin E deficiency.

Management

Laboratory studies have shown a beneficial effect of the coenzyme Q analogue idebenone on iron-induced oxidative damage. Idebenone has also been reported to have a therapeutic effect on the cardiomyopathy of FA.[11] More extended studies are in progress. A combination of coenzyme Q10 (400 mg daily) and vitamin E (2,100 units daily) has been shown to improve cardiac and skeletal muscle bioenergetics in FA patients, as assessed by magnetic resonance spectroscopy.[12]

Because the specific therapy for FA is still in an experimental stage, symptomatic management of the disease still requires attention. This includes appropriate rehabilitation

Table 3 Childhood or Young Adult–Onset Ataxias with Ill-Defined Genetic Abnormalities

Disease	*Clinical or Laboratory Feature*
Early-onset ataxia with retained deep tendon reflexes	Some have FA/ARSACS mutation; others unknown
Ataxia with hypogonadism (Holmes ataxia)	Primary or secondary hypogonadism
Ataxia with coenzyme Q10 deficiency	Low coenzyme Q10 levels in muscle
Ataxia with myoclonus	MtDNA mutations, sialidosis, ceroid lipofuscinosis

FA/ARSACS—Friedreich ataxia/autosomal recessive ataxia of Charlevoix-Saguenay

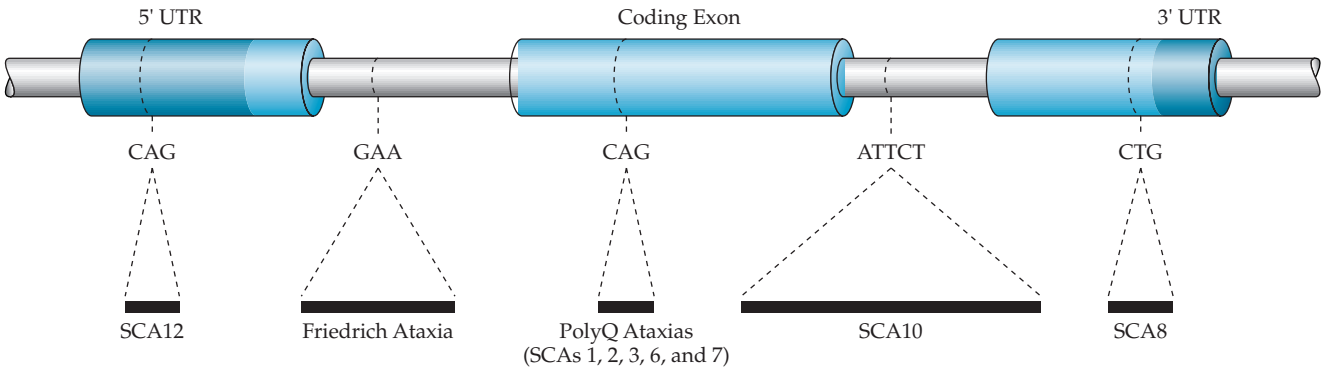


Figure 1 **Diagrammatic representation of the type of repeat expansions associated with ataxias. (SCA—spinocerebellar ataxia; UTR—untranslated region)**

measures; monitoring and treating the systemic features, such as cardiomyopathy and diabetes; surgical correction of skeletal deformity when appropriate; and other types of symptom management. In end-stage FA, appropriate palliative measures need to be taken.

FA families often need genetic counseling. The risk in siblings of patients is 25%. The carrier frequency in Indo-Caucasian populations has been estimated to be approximately one in 100.

ataxia with isolated vitamin e deficiency

Ataxia with isolated vitamin E deficiency (AVED) is a recessively inherited disorder that resembles FA, with childhood onset of ataxia, areflexia, and proprioceptive loss. Vitamin E levels are low, and vitamin E supplementation may arrest disease progression. Mutations in the α-tocopherol transfer protein gene (*TTP*) have been identified in AVED.[13] The phenotype can vary depending on the type of mutation and the degree of residual TTP activity; older age at onset and retention of tendon reflexes cannot exclude AVED. Thus, measurement of vitamin E levels is indicated in any case of sporadic or autosomal recessive ataxia with onset during childhood or the young-adult years.

ataxia-telangiectasia

Ataxia-telangiectasia (AT) typically manifests itself early in the first decade of life as increasing gait ataxia. Neurologic signs that evolve during the first decade include hypotonia, choreoathetosis, areflexia, and a characteristic oculomotor disorder that is associated with an impaired ability to generate saccades, which necessitates head thrusts to move the eyes. The typical telangiectasia appears in children at about 5 years of age and can be found over the conjunctiva, eyelids, and antecubital and popliteal fossae. These children have a high risk of malignancies, especially lymphomas. Measurement of altered radiation sensitivity and elevated α-fetoprotein levels in serum are useful to confirm the diagnosis.

The underlying gene defect in AT involves the AT mutated (*ATM*) gene, which is located on chromosome 11q. To date, AT has been linked to over 300 mutations in the *ATM* gene[14]; this knowledge has made prenatal molecular diagnosis of AT possible in some families The protein product of *ATM* appears to function as a protein kinase and is activated by ionizing radiation, with subsequent activation of cell cycle checkpoints. Defective checkpoint function may underlie some of the hematologic and immunologic phenotypes of AT. The molecular basis of Purkinje cell degeneration in AT is still poorly understood.

Because there are numerous mutations scattered over the *ATM* gene, DNA diagnosis of AT is more cumbersome than the clinical and biochemical diagnosis. Management of AT involves the care of its many systemic manifestations, such as neurologic deficits and increased susceptibility to infections and malignancies.

autosomal recessive ataxia of charlevoix-saguenay

This spastic ataxia with onset in early childhood, which was first described in a population in the Charlevoix-Saguenay province of Quebec, has recently been linked to mutations in the *Sacsin* gene, which is located on chromosome 13.[15] Ataxias in families in other parts of the world have been linked to the same locus, but the worldwide prevalence of this mutation in patients with non-FA childhood ataxia remains to be determined.

autosomal recessive ataxia with oculomotor apraxia

This childhood-onset ataxia is marked by oculomotor apraxia, areflexia, and cerebellar atrophy. The serum albumin level is low, and the serum cholesterol level may be elevated. Mutation in a gene encoding a widely expressed protein-sharing homology with the histidine-triad proteins (HIT) and with the DNA-binding zinc-finger proteins has been identified in this disease.[16,17] This gene may be involved in DNA single-strand break repair.

other disorders of dna break repair

Interestingly, additional abnormalities of DNA single- or double-strand break repair have been shown to cause autosomal recessive ataxias. These include ataxia-telangiectasia–like disorder (ATLD) related to mutations in the gene *mRE11* and spinocerebellar ataxia with axonal neuropathy (SCAN1) caused by a mutation in the gene encoding tyrosyl-DNA phosphodiesterase-1 (*Tdp1*).[18,19] Thus, the list of disorders of DNA break repair that have been noted to cause ataxia include AT, ataxia with oculomotor apraxia, ATLD, SCAN1, Cockayne syndrome, and xeroderma pigmentosum.

infantile-onset olivopontocerebellar atrophy

Infantile-onset olivopontocerebellar atrophy (IOSCA) has been described in a population isolate in Finland. Children with IOSCA present with ataxia, peripheral neuropathy,

Selected Internet Resources for Information on Ataxias

HUGO (Human Genome Organization) Nomenclature Committee
http://www.gene.ucl.ac.uk/nomenclature
Names and symbols of human genes.

OMIM (Online Mendelian Inheritance in Man)
http://www3.ncbi.nlm.nih.gov/Omim
A National Center for Biotechnology Information catalog of human genes and genetic disorders.

GeneTests
http://www.geneclinics.org
Disease reviews, international directory of genetic testing laboratories and genetics and prenatal diagnosis clinics, educational materials.

National Ataxia Foundation
http://www.ataxia.org
Hereditary ataxias patient support group.

areflexia, athetosis, and Babinski signs early in the first decade of life. Other clinical features include speech impairment, hearing loss, optic atrophy, seizures, and learning deficits. IOSCA has been mapped to chromosome 10q24.[20]

OTHER AUTOSOMAL RECESSIVE ATAXIAS

A significant proportion of childhood-onset or young-adult–onset ataxias remain unresolved at a molecular level. Koenig and coworkers have localized the gene for a syndrome comprising ataxia, neuropathy, cerebellar atrophy, and elevated creatine kinase levels to chromosome 9q33.3-34.3. They have also localized another disorder, involving ataxia, deafness, and optic atrophy, to chromosome 6p21-23.[21] Other investigators have described a childhood-onset ataxia with deficiency of coenzyme Q10 levels in muscle.[22]

The combination of youth-onset ataxia and myoclonus or myoclonic seizures can be seen under a variety of circumstances. The term Ramsay Hunt syndrome is loosely applied to this entity. One disorder that has this combination of features is Unverricht-Lundborg disease, which has been linked to mutations in the *cystatin B* gene.[23] Some patients with Ramsay Hunt syndrome may have defined diseases such as ceroid lipofuscinosis, sialidosis, and mitochondrial cytopathies.

The term early-onset ataxia with retained tendon reflexes was coined by the English neurologist Anita Harding in 1981 to describe children with progressive ataxia resembling FA, except for the retention of reflexes. Many of these patients are now known to harbor the FA mutation. In others, the genotype is still unknown. It is also known that a small minority of patients with the typical FA phenotype do not carry the FA mutation.[24]

Autosomal Dominant Ataxias

The autosomal dominant ataxias [*see Table 2*] are genetically heterogeneous but share many clinical features. Therefore, precise identification of the underlying genotype on the basis of clinical features alone can be difficult. A clinical diagnosis is often helped by examining several affected family members to assess the degree of phenotypic variability in the same family. Autosomal dominant ataxias with a progressive course are labeled SCA, for spinocerebellar ataxia, followed by a number that denotes the specific gene locus involved. The SCA nomenclature is administered by the Human Genome Organization [*see Sidebar* Selected Internet Resources for Information on Ataxias].

GENETICS AND PATHOGENESIS

Most of the autosomal dominant progressive ataxias elucidated to date are related to unstable expansions of repeated nucleotide sequences [*see Table 2 and Figure 1*]. A CAG expansion in the coding region of the gene occurs in SCA1, SCA2, SCA3, SCA6, SCA7, SCA17, and dentatorubral pallidoluysian atrophy (DRPLA).[25-28] These are all examples of polyglutamine disorders, because CAG codes for glutamine. A CAG expansion in the promoter region of the gene occurs in SCA12.[29] SCA8 has been related to a CTG expansion in an untranslated region of the gene[30] and SCA10 to a pentanucleotide repeat expansion.[31]

The CAG expansion disorders share certain features. The age at onset is typically inversely correlated with the number of CAG repeats in the expanded allele. In addition, the degree of expansion also correlates, to some extent, with rates of disease progression and some of the phenotypic features. Thus, larger expansions may result in more rapid loss of neuronal function and more widespread neuronal pathology. The normal alleles in these disorders are stable on intergenerational transmission. The expanded alleles are often unstable, and further expansion is more likely than contraction. Expansion of the repeat is more likely to occur with paternal inheritance than with maternal inheritance, and it often is of larger magnitude with paternal inheritance; this accounts for some of the anticipation (i.e., observed onset at a progressively earlier age in successive generations) in these disorders.

The genes with CAG expansions that are mutated in the autosomal dominant ataxias are widely expressed in neuronal and nonneuronal tissues. Such dominant ataxias are classified among the polyglutamine disorders because the CAG tract is translated into a polyglutamine stretch in the corresponding protein. There is increasing evidence that the polyglutamine disorders are related to the acquisition of an abnormal toxic property (gain of function) by the protein product of the

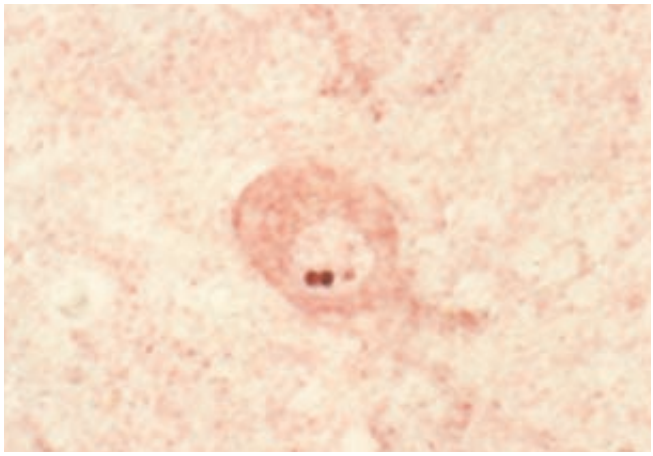


Figure 2 **Aggregates of ataxin 3 are visible within the nucleus of a pontine neuron from a patient with Machado-Joseph disease.**

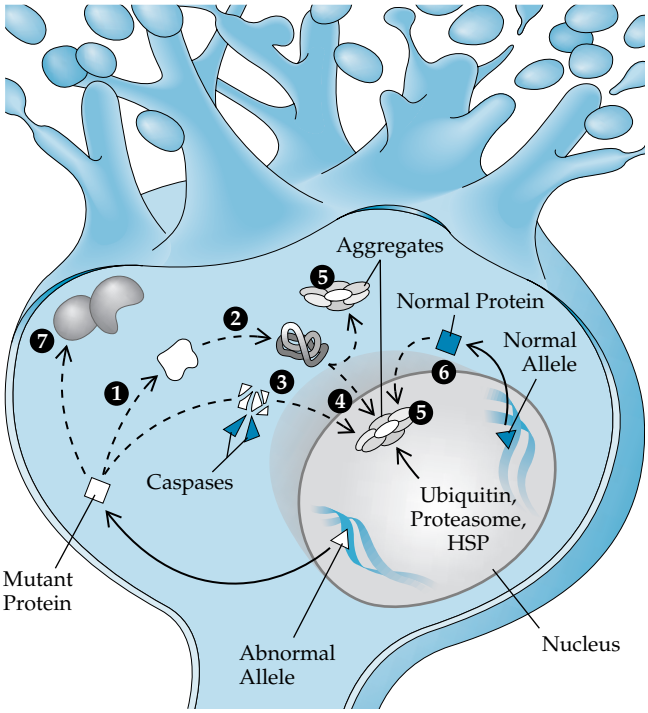


Figure 3 **Schematic picture of some of the proposed pathogenic mechanisms in the polyglutamine ataxias. The abnormal allele (Δ) is translated into a protein with abnormal molecular weight (□). The mutant protein with an excess number of glutamine molecules may have an altered conformation (1), allowing misfolding (2) and perhaps cleavage by caspases (3). The entire mutant protein or cleaved fragments may enter the nucleus and form aggregates (4) or may aggregate in the cytoplasm (5). These aggregates often recruit ubiquitin, proteasome, and components of the heat shock protein (HSP) chaperone system, suggesting misfolding and attempts at proteolysis. The product of the normal allele may also be recruited into these aggregates (6). The presence of mutant protein in the nucleus may alter expression of other genes essential for neuronal function. Abnormal interaction of mutant protein with other proteins (7) may account for some of the specificity of neuronal loss.**

mutated allele. Most of the gene products related to the CAG-expansion disorders are novel proteins of unknown function. Normally, many of these proteins are distributed diffusely in the cytoplasm and are variably present in the nucleus.

Both the messenger RNA (mRNA) and the protein corresponding to the normal and the expanded alleles can be detected in the tissues of the affected persons. In contrast to the diffuse localization of the normal protein, neurons in patients with the disease tend to show aggregates of the mutated protein, mostly within neuronal nuclei [*see Figure 2*] but also in the cytoplasm and in the neurites. The role of these nuclear inclusions (NIs) in the pathogenesis of the dominant ataxias is debated [*see Figure 3*]. However, many observations have been made, such as the following[25-27]:

1. NIs can be identified in many transgenic and transfection models of the ataxias.
2. NIs can be dissociated from the pathology; thus, genetic engineering can abolish the presence of NIs without reducing the pathology; this has been documented for SCA1.
3. In SCA1, preventing nuclear entry of mutated protein can prevent pathology.
4. Truncated constructs of the gene resulting in fragments of the protein containing the polyglutamine tract can often be more toxic than full-length protein.
5. Caspases may be involved in the generation of such truncated version of the proteins in vivo.
6. NIs usually recruit many other proteins; the presence of proteins belonging to the heat shock protein (HSP) chaperone system suggests that the mutated polyglutamine protein assumes a misfolded configuration. Also, the presence of proteins belonging to the ubiquitin-proteasome pathway in the NI indicates attempts at removing the aggregated protein. Manipulation of both the HSP and the ubiquitin paths can alter the intensity of NI and sometimes the pathogenic severity.
7. The aberrant interaction of the mutant protein in the nuclear context may alter many essential nuclear functions. Many of the mutant proteins can bind transcription factors such as the cyclic adenosine monophosphate (cAMP) response element-binding protein (CREB)–binding protein (CBP). Others have RNA-binding activity. Such interaction may result in upregulation or downregulation of other important genes that can have pathogenetic implications.

The mechanism of disease in the nonpolyglutamine ataxias is still poorly understood. In the case of SCA8, the CTG expansion in the 3′ untranslated region of the gene may result in aberrant RNA activity that could be lethal to neurons.[32] In SCA6 and the episodic ataxia (EA) syndromes, which are related to mutations in channel proteins, abnormal electrical properties of the mutated channel may have a role.[33] Lastly, recent reports have identified point mutations in genes encoding protein kinase C-γ (in SCA14), as well as fibroblast growth factor 14, as causes of progressive spinocerebellar ataxias.[34,35]

DIAGNOSIS

Clinical Features

Patients with autosomal dominant ataxia present with progressive gait problems, dysarthria, and occasionally diplopia.[36-38] In addition to having cerebellar signs, these patients exhibit a variety of other neurologic deficits. As the disease progresses, eye-movement abnormalities such as gaze paresis, lid retraction, and blepharospasm, as well as bulbar signs such as facial and temporal atrophy, facial fasciculations, and mild tongue atrophy, may occur. Many of the disorders are associated with upper motor neuron signs such as brisk deep tendon reflexes, spasticity, and extensor plantar responses. Some of the genotypes are associated with basal ganglia signs such as an akinetic rigid syndrome, dystonia [*see Figure 4*], and choreoathetosis. Involvement of peripheral nerves can be found in the form of distal sensory loss and areflexia, as well as amyotrophy, fasciculations, and cramps. Tremor resembling essential tremor and torticollis can occur in some cases. Cognitive changes, seizures, and myoclonus reflect cortical pathology. Evidence of retinal disease occurs in certain genotypes. Patients typically lose ambulation 10 to 15 years after onset and, at this stage, develop increasing dysphagia; some may have urinary incontinence.

The EA syndromes cause intermittent episodes of ataxia; these are very short-lived (minutes) in EA-1 and longer (hours) in EA-2. Interictally, EA-1 can be associated with

Figure 4 **Dystonia in a patient with Machado-Joseph disease.**

skeletal muscle myokymia and EA-2 with nystagmus. The EA-2 phenotype can be variable and may be similar to SCA6 and may be associated with migraine (both SCA6 and familial hemiplegic migraine are allelic to EA-2).

The various autosomal dominant ataxias are marked by phenotypic differences [*see Table 4*]. It should be stressed, however, that phenotypic features alone are often insufficient to permit diagnosis of the genotype in individual patients. Within specific genotypes, the phenotype may vary significantly. In many instances, this variance is related to the duration of disease; however, the variation in the number of nucleotide repeats from generation to generation can influence the phenotype. Thus, in all CAG repeat disorders, there is an inverse relationship between the age at onset and the size of the expanded allele. Other modifier genes that can influence the phenotype are currently the focus of considerable interest.

Table 4 Phenotypic Features That May Indicate a Specific Genotype in Autosomal Dominant Ataxias

Phenotypic Feature	*Disorders*
Age at onset	Young adult: SCA1, SCA2, MJD
	Older adult: SCA6
	Childhood: frequent in SCA7/DRPLA
Degree of anticipation	More in SCA7, DRPLA
Benign course	SCA6
Upper motor neuron signs	SCAs 1, 7, and 8; MJD; rare in SCA2
Akinetic-rigid/Parkinson signs	MJD, SCA2, SCA17
Chorea	Prominent in DRPLA; late in SCA2, SCA1, MJD
Action tremor	SCA12, SCA16
Very slow saccades	Early in SCA2, SCA7; late in SCA1, MJD; never in SCA6
Downbeat nystagmus	SCA6, EA2
Generalized areflexia	SCA2, SCA4, older adult–onset MJD
Visual loss	SCA7
Seizures	SCA10, early-onset DRPLA, SCA7

DRPLA—dentatorubral pallidoluysian atrophy EA—episodic ataxia
MJD—Machado-Joseph disease SCA—spinocerebellar ataxia

Table 5 Normal and Expanded Range of Various Repetitive Nucleotide Sequences in Inherited Ataxias

Disease	*Repeat*	*Normal*	*Intermediate*	*Expanded*
SCA1	CAG	6–44		39–82[1]
SCA2	CAG	14–31	32–33	33–64
MJD	CAG	12–40		54–86
SCA6	CAG	4–18		19–30
SCA7	CAG	4–27	28–36	37–200
SCA8	CTG	16–91		107–127[2]
SCA10	ATTCT	10–22		1,000–4,500
SCA12	CAG	< 29		66–78
SCA17 (*TBP* gene)	CAG	25–44		50–63

Note: Pathogenic and nonpathogenic alleles of the same size can be distinguished by the presence of CAT interruptions in the latter. The test for CTG expansion in SCA8 has to be interpreted with caution; whether the CTG expansion is the specific cause for ataxia in SCA8 has been debated, because in many instances this expansion has been found in patients with unrelated disorders. Yet at least in one large family, the repeat expansion was significantly associated with ataxia.

MJD—Machado-Joseph disease SCA—spinocerebellar ataxia
TBP—TATA-binding protein

Laboratory Studies

Imaging studies typically show either isolated cerebellar atrophy (in SCA6 and other disorders with a so-called pure cerebellar presentation) or pontocerebellar atrophy (in SCA1, SCA2, SCA3, SCA7). Individual genotypes cannot be distinguished by imaging studies alone. T_2-weighted images may show increased signal density along the pontocerebellar fibers.

Nerve conduction studies often show evidence of an axonal neuropathy in some disorders (SCA 1 through 4). Brain stem evoked potentials are abnormal in the disorders associated with brain stem pathology. Abnormal results on electroretinography can antedate visual loss in patients with SCA7.

Molecular tests are now available for SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, SCA10, SCA12, SCA17, and DRPLA [*see Table 5*]. Gene tests are possible for the EA syndromes but involve sequencing or similar tests that are technically more difficult; these diseases are better diagnosed by clinical and biochemical parameters.

Differential Diagnosis of Inherited Ataxias

In patients presenting with progressive ataxia, a number of acquired disorders need to be excluded by appropriate imaging and laboratory studies. These disorders include vascular problems such as infarcts or malformations; tumors; demyelinating disease; paraneoplastic cerebellar degenerations; hypothyroidism; and cerebellotoxic disease, such as that related to alcohol, antiepileptic drugs, and cytosine arabinoside. Other possible immune etiologies for ataxic disorders have been proposed, including antibodies against glutamic acid decarboxylase and gliadin. Lastly, in a substantial number of patients, a definitive genetic or environmental cause cannot be established. Such patients are diagnosed as having late-onset sporadic or idiopathic ataxia.

In young adults or children, genetically determined metabolic errors may have ataxia as a dominant clinical fea-

ture. These disorders need to be excluded when appropriate. Examples of such diseases are Wilson disease, cerebrotendinous xanthomatosis, abetalipoproteinemia, some of the leukodystrophies and gangliosidoses, sialidosis, ceroid lipofuscinosis, carbohydrate-glycoprotein deficiency syndromes, and giant axonal neuropathy.

Management of Inherited Ataxias

Currently, no specific interventions are available for most of the inherited ataxias. The possible treatment strategies in FA have been discussed (see above). When specific biochemical defects are identified, these can possibly be corrected—for example, with vitamin E supplements in AVED, other vitamin therapy, or dietary manipulations. The episodic ataxias (EA-1 and EA-2) often respond to acetazolamide.

The progressive ataxias are treated by rehabilitative measures such as assistive equipment, home modification, dysphagia treatment, and physical therapy. Several symptoms may need to be treated, including depression, anxiety, and pain.

Genetic counseling is important. In the dominant ataxias, offspring of affected persons have a 50% risk of developing the disease. When the molecular defect in the family is known, predictive testing can be performed. The guidelines used for Huntington disease are usually adopted for this purpose.[39]

The author has been a member of the speakers' bureau for Athena Neurodiagnostics during the past 12 months.

References

1. Pandolfo M: Friedreich's ataxia: clinical aspects and pathogenesis. Semin Neurol 19: 311, 1999

2. Durr A, Cossee M, Agid Y, et al: Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med 335:1169, 1996

3. Filla A, DeMichele G, Cavalcanti F, et al: The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. Am J Hum Genet 59:554, 1996

4. Montermini L, Richter A, Morgan K, et al: Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. Ann Neurol 41:675, 1997

5. Schols L, Amoiridis G, Przuntek H, et al: Friedreich's ataxia: revision of the phenotype according to molecular genetics. Brain 120:2131, 1997

6. Cossee M, Durr A, Schmitt M, et al: Friedreich's ataxia: point mutations and clinical presentation of compound heterozygous Friedreich's ataxia patients. Ann Neurol 45:200, 1999

7. Sakamoto N, Chastain PD, Pamiewski P, et al: Sticky DNA: self-association properties of long GAA.TTC repeats in R R.Y triplex structures from Friedreich's ataxia. Mol Cell 3:465, 1999

8. Puccio H, Koenig M: Recent advances in the molecular pathogenesis of Friedreich ataxia. Hum Mol Genet 9:887, 2000

9. Lodi R, Taylor DJ, Schapira AH: Mitochondrial dysfunction in Friedreich's ataxia. Biol Signals Recept 10:263, 2001

10. Waldvogel D, van Gelderen P, Hallett M, et al: Increased iron in the dentate nucleus of patients with Friedreich's ataxia. Ann Neurol 46:123, 1999

11. Rustin P, von Kleist-Retzow JC, Chantrel-Groussard K, et al: Effect of idebenone on cardiomyopathy in Friedreich's ataxia. Lancet 354:477, 1999

12. Lodi R, Hart PE, Rajagopalan B, et al: Antioxidant treatment improves in vivo cardiac and skeletal muscle bioenergetics in patients with Friedreich's ataxia. Ann Neurol 49:590, 2001

13. Ouahchi K, Arita M, Kayden H, et al: Ataxia with isolated vitamin E deficiency is caused by mutations in the α-tocopherol transfer protein. Nat Genet 9:141, 1995

14. Halazoineitis TD, Shiloh Y: Many faces of ATM: Eighth international workshop on ataxia telangiectasia. Biochimica et Biophysica Acta 1424:R45, 1999

15. Engert JC, Berube P, Mercier J, et al: ARSACS, a spastic ataxia common in northeastern Quebec, is caused by mutations in a new gene encoding an 11.5-kb ORF. Nat Genet 24:120, 2000

16. Moreira M-C, Barbot C, Tachi N, et al: The gene mutated in ataxia-oculomotor apraxia 1 encodes the new HIT/Zn-finger protein aprataxin. Nat Genet 29:189, 2001

17. Date H, Onodera O, Tanaka H, et al: Early-onset ataxia with ocular motor apraxia and hypoalbuminemia is caused by mutations in a new HIT superfamily gene. Nat Genet 29:184, 2001

18. Stewart GS, Maser RS, Stankovic T, et al: The DNA double-strand break repair gene *hMRE 11* is mutated in individuals with an ataxia-telangiectasia–like disorder. Cell 99:577, 1999

19. Takashima H, Boerkoel CF, John J, et al: Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. Nat Genet 32:267, 2002

20. Nikali K, Isosomppi J, Lonnqvist T, et al: Toward cloning of a novel ataxia gene: refined assignment and physical map of the IOSCA locus (SCA 8) on 10q24. Genomics 39:185, 1997

21. Bomont P, Watanabe M, Gershoni-Barush R, et al: Homozygosity mapping of spinocerebellar ataxia with cerebellar atrophy and peripheral neuropathy to 9q33-34, and with hearing impairment and optic atrophy to 6p21-23. Eur J Hum Genet 8:986, 2000

22. Musumeci O, Naini A, Slonim AE, et al: Familial cerebellar ataxia with muscle coenzyme Q10 deficiency. Neurology 56:849, 2001

23. Lehesjoki AE, Koskiniemi M: Clinical features and genetics of progressive myoclonus epilepsy of the Unverricht-Lundborg type. Ann Med 30:474, 1998

24. McCabe DJH, Ryan F, Moore DP, et al: Typical Friedreich's ataxia without GAA expansions and GAA expansions without typical Friedreich's ataxia. J Neurol 247:346, 2000

25. Orr HT: Beyond the Qs in the polyglutamine diseases. Genes Develop 15:925, 2001

26. Zoghbi HY, Orr HT: Glutamine repeats and neurodegeneration. Annu Rev Neurosci 23:217, 2000

27. Paulson HL: Protein fate in neurodegenerative proteinopathies: polyglutamine diseases join the (mis)fold. Am J Hum Genet 64:339, 1999

28. Holmes SE, O'Hearn EE, McInnis MG, et al: Expansion of a novel CAG trinucleotide repeat in the 5′ region of PPP2R2B is associated with SCA 12. Nat Genet 23:391, 1999

29. Zuhlke C, Hellenbroich Y, Dalski A, et al: Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. Eur J Hum Genet 9:260, 2001

30. Koob MD, Moseley ML, Schut LJ, et al: An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). Nat Genet 21:379, 1999

31. Matsuura T, Yamagata T, Burgess DL, et al: Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 26:191, 2000

32. Nemes JP, Benzow KA, Koob MD, et al: The SCA 8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1) Hum Mol Genet 9:1543, 2000

33. Frontali M: Spinocerebellar ataxia type 6: channelopathy or glutamine repeat disorder? Brain Res Bull 56:227, 2001

34. Chen DH, Brkanac Z, Verlinde CL, et al: Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. Am J Hum Genet 72:839, 2003

35. van Swieten JC, Brusse E, de Graaf BM, et al: A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia. Am J Hum Genet 72:191, 2003

36. Subramony SH, Vig PJS, McDaniel DO: Dominantly inherited ataxias. Sem Neurol 19:419, 1999

37. Schols L, Amoiridis G, Buttner T, et al: Autosomal dominant cerebellar ataxia: phenotypic differences in genetically determined subtypes? Ann Neurol 42:924, 1997

38. Subramony SH, Filla A: Autosomal dominant spinocerebellar ataxias ad infinitum? Neurology 56:287, 2001

39. Ethical issues policy statement on Huntington's disease molecular genetics predictive test. International Huntington Association; World Federation of Neurology. J Med Ethics 27:34, 1990

Acknowledgment

Figure 1 Provided by Dr. Henry Paulson, University of Iowa School of Medicine.

Figure 3 Seward Hung.

XI ALZHEIMER DISEASE AND OTHER MAJOR DEMENTING ILLNESSES

David S. Knopman, m.d.

Alzheimer Disease

definition

Alzheimer disease (AD) is histopathologically defined by neurofibrillary tangles and neuritic plaques in the cerebral cortex. AD is also commonly used as a clinical diagnosis for a dementia syndrome in which anterograde amnesia is a dominant symptom. The most common pathologic basis for the clinical syndrome AD is indeed the pathologically defined disease AD.

A person with the dementia of AD experiences a loss of cognitive function that interferes with social or occupational activities and generally results in a dependence on others for carrying out at least some necessary daily affairs. In the language of the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition:* "The cognitive deficits cause significant impairment in social and occupational functioning and represent a significant decline from a previous level of functioning."[1] The decline from a previously higher level of functioning distinguishes dementia from developmental delays or mental retardation. The sensorium (i.e., the person's level of arousal and alertness) is unaffected in dementia. This feature distinguishes dementia from delirium, in which cognition is impaired in the setting of an altered sensorium.

As defined, dementia is impairment of more than one area of cognitive function. In the dementia syndrome of AD, therefore, anterograde amnesia must be accompanied by impairment of at least one of the following: language, abstract reasoning, executive functioning, and visuospatial processing. The most widely used definition of AD is that based on criteria established by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association (NINCDS-ADRDA) work group[2] [*see Table 1*].

The pivotal cognitive finding in AD is anterograde amnesia, which is an inability to learn new things. Persons with anterograde amnesia typically cannot keep track of the date, remember recent conversations, or remember where they set something down; and they often repeat themselves in conversation. Although these types of difficulties are commonly described as short-term memory loss, the key difficulty in anterograde amnesia of AD is in acquiring or learning new information. Failure to learn results in difficulties recalling related information a few minutes later. AD is not a generalized disorder of recall, however, because early in the disease—especially, but even into its moderate stages—persons with AD can retrieve and recall learned information from the remote past.

epidemiology

The prevalence of AD is approximately 8% of the United States population older than 65 years.[3] Approximately two to three million persons in the United States have AD; estimates vary somewhat because of different case definitions used.[3] If very mild AD cases are included, the number of affected persons exceeds four million.[4] The prevalence is strongly age related [*see Figure 1*]: AD is uncommon in persons younger than 65 years, but prevalence doubles every 5 years from then on. As many as 30% to 40% of persons older than 85 years have AD.

The incidence of AD mirrors the prevalence, being very low before 65 years of age and then steadily increasing into the 70s and 80s. By the age of 80, the annual incidence of AD approaches two per 100 persons. There is some controversy regarding the incidence of AD in the 10th and 11th decades of life; it probably continues to rise, but the methodology and numbers of cases at the extremes of life expectancy make for uncertainty.

There is also controversy as to whether women are more commonly affected than men. Most studies of prevalence and incidence show that women are slightly overrepresented. The difference is small, however, and could be an artifact of the overall lower life expectancy of men.[5] Emerging studies of dementia in African Americans[6] and Japanese Americans[7] suggest that there are no ethnic differences in the prevalence or incidence of AD.

protective and risk factors

Several factors appear to alter the risk of developing AD. The most prominent risk factors for AD are advancing age and a family history of dementia. Cardiovascular disease[8] and diabetes mellitus[9] both appear to be associated with an increased risk of AD. Elevated levels of homocysteine also are associated with a higher risk of AD.[10]

Table 1 Criteria for the Clinical Diagnosis of Alzheimer Disease*

A. Evidence from the history and mental-status examination that indicates a disorder characterized by major impairment of learning and of retaining new information, as well as at least one of the following:
   1. Impairment in handling complex tasks
   2. Impairment in reasoning ability
   3. Impaired spatial ability and orientation
   4. Impaired language

B. The disturbances listed under A above significantly interfere with work or usual social activities or relationships with other people.

C. The disturbances represent a significant decline from a previous level of functioning.

D. The disturbances are of insidious onset and are progressive, based on evidence from the history or serial mental-status examinations.

E. The disturbances are not occurring exclusively during the course of delirium.

F. The disturbances are not better accounted for by a major psychiatric diagnosis.

G. The disturbances are not better accounted for by a systemic disease or another brain disease.

*These criteria are based on the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer Disease and Related Disorders Association work group report[2] and the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*.[1]

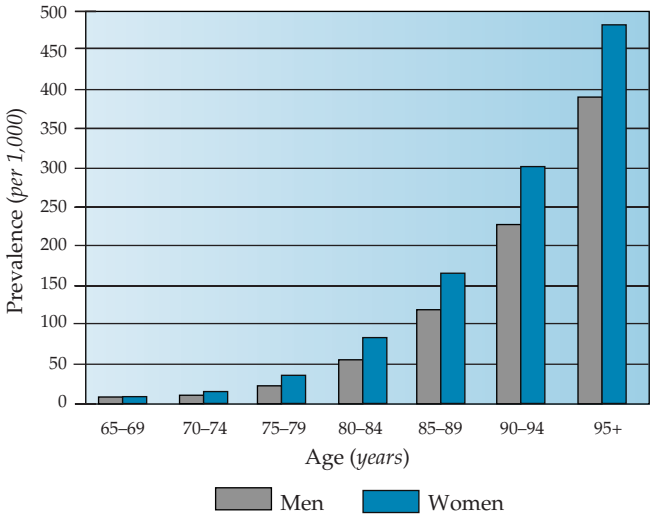


Figure 1 **Prevalence of Alzheimer disease per 1,000 population, by gender and age.**

Use of certain medications is associated with a lower risk of developing AD. These include estrogen, the statin-type cholesterol-lowering drugs, and nonsteroidal anti-inflammatory drugs (NSAIDs). For estrogen and the NSAIDs, medication use was assessed when persons were not demented, a methodological strategy that strengthened the belief that the observed reduction in risk was not an artifact of the appearance of early AD. Unfortunately, subsequent clinical trials of estrogen[11] and NSAIDs[12] in AD patients, as well as a very large prevention study using estrogen and progesterone,[13] have uniformly failed to show evidence of benefits for either class of drug. On the basis of the same types of observations,[14] clinical trials with statins in AD are under way. The disappointing study conclusion with estrogen and NSAIDs is that epidemiologic associations do not automatically translate into successful therapeutic interventions.

Level of educational attainment and intellectual performance in childhood also appear to be related to the risk of developing AD in later life. In a study of persons who took academic achievement tests at 11 years of age and then were evaluated in adulthood for dementia, there was a notable association between higher scores at age 11 and a lower likelihood of developing AD in later life.[15] Similarly, a study of elderly Catholic nuns who had written autobiographies when they entered the order, at approximately age 20, showed that the nuns with better vocabularies and syntactic abilities when young had a lower likelihood of developing AD in later life.[16] Many other studies have observed that educational attainment seems to be inversely related to the risk of AD, though questions have been raised as to whether the so-called education effect is actually of socioeconomic origin.[17]

GENETICS AND MOLECULAR BIOLOGY

The genetic basis of AD can be divided into four profiles: (1) the very rare autosomal dominant forms of AD; (2) the more common, later-onset form of AD, in which family history exerts a modestly strong risk; (3) sporadic AD, which comprises the majority of AD patients, including both early-onset and, especially, late-onset forms that appear to be nongenetic; and (4) persons with Down syndrome (trisomy 21).

The number of persons with autosomal dominant AD is infinitesimal compared with the large number of persons with late-onset sporadic AD. The autosomal dominant families have been pivotal, however, in providing insights into the mechanisms of the disease. In general, affected individuals in autosomal dominant AD families experience dementia in their 30s to 60s; the genetic disorder advances the age of onset of AD by roughly 2 to 3 decades relative to sporadic AD.

Three genes have been implicated in autosomal dominant AD. One is the amyloid precursor protein (*APP*) gene, on chromosome 21, which encodes the amyloid-β peptide [*see Figure 2*]. The amyloid-β peptide appears to be a key pathogenic molecule. Only a handful of families in the world harbor the mutations in this gene that cause AD. The other two genes involved in autosomal dominant AD are homologous; the much more common of the two is the presenilin-1 (*PS-1*) gene, located on chromosome 14. Approximately half of the few hundred autosomal dominant AD families have disease linked to mutations on chromosome 14.[18] The less common homologue, presenilin-2 (*PS-2*), is encoded by a gene on chromosome 1.[19] Families with the *PS-2* mutation are almost exclusively of one ethnic group—Germans whose ancestors emigrated to the western United States and Canada after a sojourn of a few hundred years in the Volga River area of Russia. The mutation is thought to have resulted from a founder effect, which propagated by virtue of the cultural isolation in a foreign land of this group of people.

In addition to the rare autosomal dominant families, a family history of AD, even of later onset, is a consistent risk factor for the disease.[20] One early study showed that the risk of AD in siblings of probands with autopsy-proven AD rose to nearly 50% by age 85 if two features were present: the presence of dementia in another first-degree relative and an age of onset under 70 years in the proband.[21] Subsequently, as families with late-onset AD were studied genetically, the apolipoprotein E (*APOE*) gene was found to account for a substantial fraction of the familial risk in late-onset AD.[22,23] The *APOE* gene has three allelic variations (ε*2*, ε*3*, and ε*4*) in humans, as a result of polymorphisms in two nucleotides. The *APOE* ε*4* allele is strongly associated with the development of AD at an earlier age than in non-ε*4* carriers.[24] However, the presence of an *APOE* ε*4* allele does not guarantee the development of AD in *APOE* ε*4* carriers, and its absence does not protect against AD in *APOE* non-ε*4* carriers.[25] There are no clinical differences between AD patients with and AD patients without the *APOE* ε*4* allele. No other gene has been linked definitively to risk of AD. Although there is intense interest in a region on chromosome 10, the exact gene has not been identified.

Progress in defining the molecular biology of AD has provided the framework for understanding the probable pathophysiologic pathways that lead to AD [*see Figure 2*]. In essence, the so-called amyloid hypothesis of AD asserts that the amyloid-β peptide is toxic either to neurons or to synaptic function in a dose-dependent fashion.[26,27] According to this theory, excess production or reduced clearance of the amyloid-β peptide will lead to AD, and the higher the amyloid-β peptide load, the earlier the onset of the dementia. Individuals with Down syndrome also exhibit AD pathology 3 to 4 decades earlier than that which is seen in patients with sporadic AD, presumably because of the extra chromosome 21 that contains the wild-type *APP* gene. The *APP* gene mutations lead to overproduction of the amyloid-β peptide because the mutations upset the balance of the three secretases involved in APP metabolism. The *PS-1* and *PS-2* genes appear to play an integral role in the action of one of those secretases, the γ-secretase. The mutations in *PS-1* and *PS-2* result in a

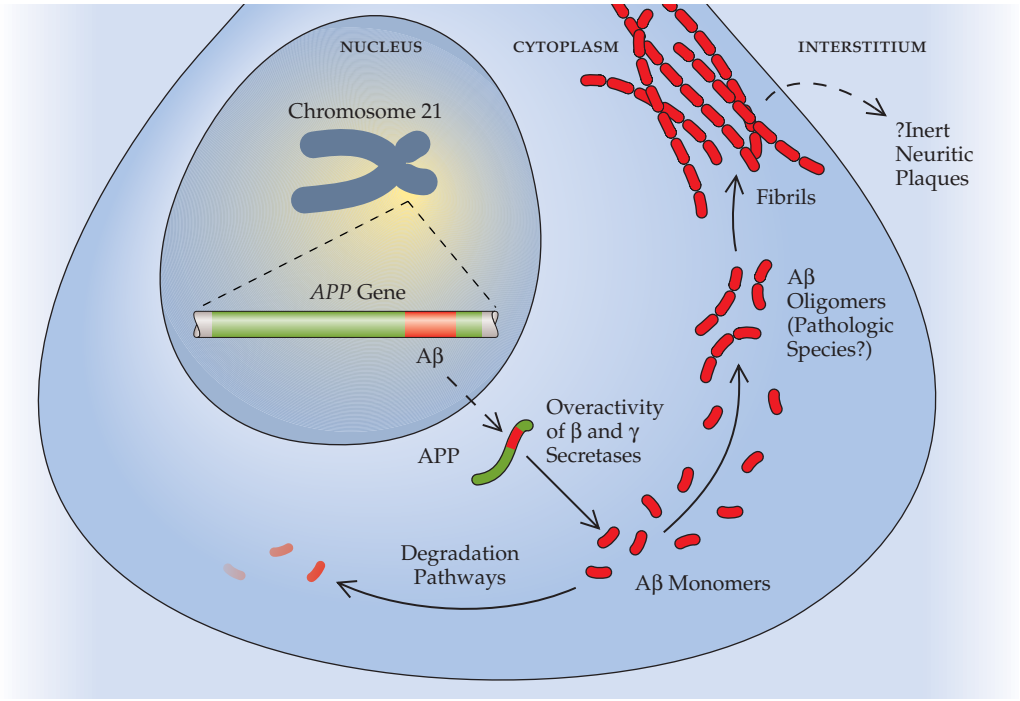


Figure 2 **Schematic overview of the amyloid-β peptide production pathway. The Alzheimer precursor protein (APP) is encoded by a gene on chromosome 21. Under certain circumstances, presumably involving relative overactivity of β and γ secretases compared with α secretase, the amyloid-β peptide is produced. It may assemble itself into dimers, trimers, or other oligomers, which are thought to be cytotoxic. Eventually, the oligomers polymerize into fibrils and, later, into plaques.[3]**

heightened activity of γ-secretase that also leads to excess production of the amyloid-β peptide. Individuals with the *APOE ε4* allele also have excess amyloid-β peptide in their brains, but the precise mechanism by which the *APOE* genotype modifies the molecular biology of AD is still a matter of debate.

PATHOPHYSIOLOGY

The brain of a patient who has died with the dementia syndrome of AD will usually show gross atrophy of the cerebral hemispheres, particularly in the posterior half of the brain. Microscopic examination of the brain is required to make a definitive diagnosis of AD, however.

There are two principal histologic findings in AD. The first is the neuritic plaque. Neuritic plaques are extracellular structures that contain a core made up of the amyloid-β peptide and surrounded by degenerated axonal and dendritic elements of many neurons [*see Figure 3*]. They can be imaged with routine hematoxylin and eosin staining, but they are best seen with silver stains. Neuritic plaques may be present in very small numbers in nondemented elderly persons, but in the typical patient with AD, neuritic plaques are numerous[28] and are located in many parts of the cerebral cortex.[29]

The genesis of neuritic plaques is better understood than it once was, but there are still gaps in the described pathway between an excess of the amyloid-β peptide and the appearance of neuritic plaques. It is currently believed that the amyloid-β peptide polymerizes and that the oligomer consisting of a few copies of the amyloid-β peptide is neurotoxic. Presumably, the amyloid-β peptide oligomers continue to aggregate and eventually form the amyloid core of the neuritic plaques. Many experts believe that the oligomeric form of the amyloid-β peptide is the key pathogen in AD[30] and that the neuritic plaques merely represent the end stage of the pathologic process.

The other principal histologic feature in AD is the intracellular neurofibrillary tangle [*see Figure 4*]. Neurofibrillary tangles are found inside neurons and consist of a protein known as tau that has been pathologically modified so that it has polymerized into an insoluble form that assembles itself into filaments. Neurofibrillary tangles may be seen in nondemented elderly persons in the entorhinal cortex and even in the hippocampus, but when neurofibrillary tangles are found in larg-

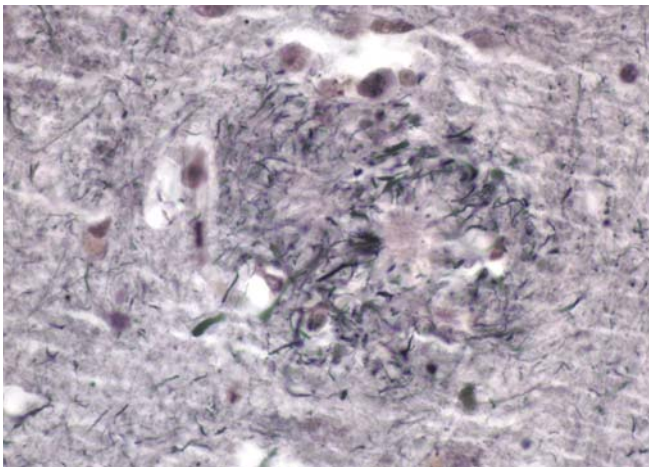


Figure 3 **Photomicrograph of a neuritic plaque in the cortex of a patient with Alzheimer disease.**

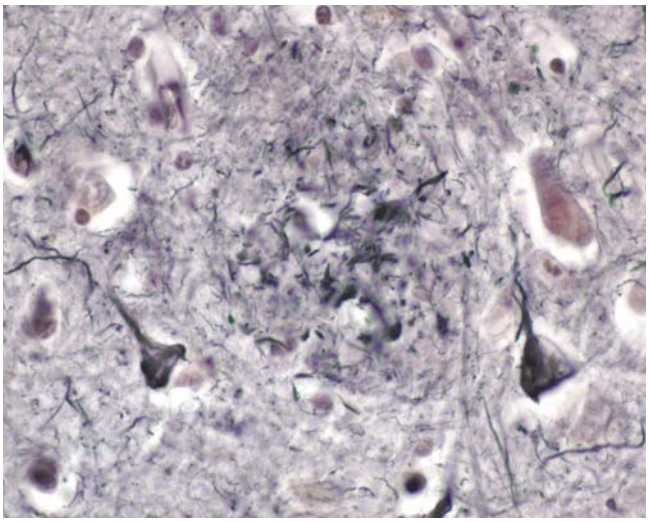


Figure 4 **Photomicrograph of several neurons containing neurofibrillary tangles from the cortex of a patient with Alzheimer disease.**

er numbers in the neocortex of the temporal or parietal lobes, the person will almost invariably have had AD dementia.[29,31]

The tau protein is the most important molecule in neurofibrillary tangles. It is ordinarily involved in stabilizing the cytoskeletal elements in neurons (i.e., the microtubules). In AD, tau becomes excessively phosphorylated and aggregated.[32] Curiously, despite the discovery of mutations in the *tau* gene that cause other forms of dementia, no *tau* gene mutations have been associated with AD. It seems most likely that formation of neurofibrillary tangles represents a technically normal but ultimately disastrous response to injury caused by the amyloid-β peptide. The amyloid-β peptide induces excessive phosphorylation of tau protein, which destabilizes microtubules and leads to aggregation of hyperphosphylated tau into fibril.[33]

DIAGNOSIS

Clinical Manifestations

The dementing illness in AD is almost always very gradual and insidious in its onset. Typically, once the condition is diagnosed, changes in memory and daily functioning will have been evident for at least a year, if not 2 or 3 years. Once the dementia is diagnosed, the clinical course usually plays out over 3 to 7 years before severe dementia becomes present.[34] The median duration from the onset of symptoms to death is approximately 7 years.[35,36] It is longer in younger patients and somewhat shorter in patients who are in their 80s and 90s.

The earliest clinical manifestations of AD typically involve problems related to the anterograde amnesia (i.e., the short-term memory difficulties). Persons with AD often experience short-term memory loss as the only difficulty for several years. This condition is referred to as mild cognitive impairment (MCI). MCI is recognized as a risk state for the subsequent development of AD.[37] Patients who eventually develop AD always pass through a stage of MCI, if only briefly, but not all patients with MCI inevitably go on to develop dementia.

One of the most frustrating and challenging aspects of the evolution of the clinical syndrome of AD is the gradually intensifying loss of self-awareness and insight.[38] This is referred to as anosognosia, a loss of awareness of one's deficits. AD is almost unique among medical conditions in that patients themselves are usually not instrumental in bringing their problem to medical attention. Affected persons will sometimes articulate concerns about their memory, but family members are the ones who force the issue in the majority of instances.

The anterograde amnesia of AD typically manifests itself in the daily affairs of the patient as a difficulty in recalling recent events and conversations, a tendency to repeat oneself in conversation, and a difficulty in keeping track of the date. At first, these behaviors may be intermittent or may appear only during times of stress or when the person is taken out of his or her usual routines. Gradually, as the condition progresses, the instances of forgetfulness become more frequent and lead to mistakes in daily affairs. These errors may include forgetting appointments, difficulties in shopping, and problems in following multistep instructions (e.g., recipes) or in following directions while traveling.

During the early evolutionary phases of the dementia, other cognitive deficits begin to emerge and intrude into daily affairs. These deficits may include a loss of facility with common words or a difficulty with remembering names of family and acquaintances. Other developments include a reduced ability to solve common household or daily problems; reduced ability to manage financial affairs; and greater difficulty in planning a meal, a family gathering, or a trip. Difficulties with planning, mental agility, problem solving, and abstract reasoning are sometimes referred to as executive dysfunction. Geographic disorientation is also a feature of emerging dementia from AD, and it manifests itself as getting lost in what was once familiar territory. A typical story is that the patient sets out by automobile to visit a friend but becomes confused about the way to get there. Patients will sometimes drive around for hours trying to reorient themselves; in other instances, they will simply give up and try to return home.

As the dementia continues to emerge, changes in behavior and mood often occur, but not always. A loss of interest in previously enjoyed pastimes and a loss of interest in family affairs are manifestations of the apathy and loss of initiative that are common in early AD. Less commonly, the person experiences substantial alterations in mood, such as tearfulness, melancholia, or anxiety. Interestingly, depression is less common than might be predicted, perhaps because of the concomitant loss of insight that accompanies early AD dementia.[39]

The progression of AD from mild to severe has been characterized by a number of different staging systems. Most systems define mild AD as that stage in which patients can care for their basic needs and still carry out some activities, such as simple hobbies and participation in social activities. Severe dementia is usually defined as the point at which the loss of ability to manage basic activities of daily living (e.g., bathing, dressing, toileting, or eating) becomes evident. The progression of the symptoms of AD varies tremendously from one person to the next.

Another way of characterizing the symptoms and the progression of AD is by assessments of mental functioning. Such assessments can be done either by a primary care physician in the routine office setting or by a neuropsychologist in a special laboratory. The two methods are concordant and rely on the same principles of assessment of cognitive abilities. The neuropsychological approach is more time consuming, but as a consequence, it provides a more detailed view of the patient's cognitive functioning.

Mental-Status Examination

The mental-status examination is a cornerstone of the diagnosis of AD. The Mini-Mental State examination (MMSE) is widely

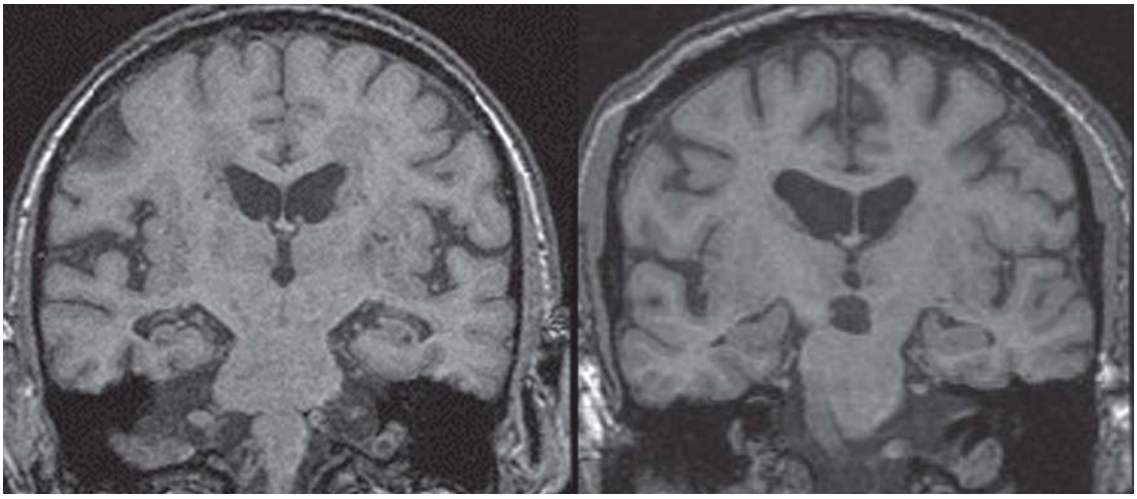


Figure 5 **Coronal MRI scan demonstrating hippocampal atrophy in a patient with clinically diagnosed Alzheimer disease (left), compared with a scan of a normal person of comparable age (right).**

used and has norms for age and education.[40] It consists of questions assessing orientation and includes evaluations of language, visuomotor function, mental agility, and delayed recall. Although the MMSE lacks sensitivity for detecting mild cognitive dysfunction, it is a very effective tool for clinicians. At the Mayo Clinic, we have been using the short test of mental status instead, because it was constructed to be more sensitive to early dementia.[41] The two examinations are highly correlated. With either test, clinicians can gain an objective view of the patient's cognitive function. Typically, patients with early AD make one or two errors on orientation to time or place and fail to recall any items specifically presented to be remembered a few minutes later, but they make no errors on reading, writing, naming simple items, or following multistep commands. As the dementia of AD progresses, performance on the mental-status examination steadily declines. On the MMSE, AD patients lose an average of three points a year. However, the range of performance from one year to the next in a particular patient can vary considerably, from no loss of points to a loss of six or seven points.

Clinical Assessment

The diagnosis of AD remains a clinical one that is based on the history, mental-status examination, remainder of the neurologic examination, and general physical examination.[42] There is no laboratory marker for AD that is sufficiently accurate to replace or supersede the judgment of an experienced physician. The physician weighs the evidence from the history of cognitive decline and judges that history in the light of the mental-status examination. To make a diagnosis of AD dementia, the history of cognitive decline and the findings on the mental-status examination should be concordant. When the history strongly suggests a cognitive disorder but the mental-status examination result is normal, it is still possible that one could be dealing with a very early case of AD, especially in a highly educated person. When the result of the mental-status examination is quite abnormal but family members say that the patient has not had difficulties with daily functioning, AD must also be considered, although delirium or a rapidly progressive disorder may be more likely. Family members sometimes understate the degree of impairment of a loved one because of lack of observation, denial, or lack of sophistication.

Laboratory Testing

Laboratory testing has a role in the assessment of dementia patients even when AD seems most likely on clinical grounds.[42] Blood tests such as a complete blood count and measurements of electrolytes, calcium, urea nitrogen, thyrotropin, and vitamin B_{12} are needed to rule out previously overlooked metabolic or hematologic derangements. Screening for neurosyphilis can be considered in regions where primary or secondary syphilis is seen with any frequency. Although none of these tests are particularly sensitive or specific for a commonly encountered dementing illness, they are simple and inexpensive tests and help exclude common general medical conditions. Examination of the cerebrospinal fluid, a urinalysis for detection of toxins, and electroencephalograms are not needed except when specific indications exist.

Imaging Studies

Brain imaging is necessary for the initial diagnostic evaluation of demented patients. A brain CT scan may be adequate, but an MRI scan without contrast enhancement will yield more clinically useful information. The fundamental purpose of a brain-imaging study is to rule out space-occupying lesions such as tumors and subdural hematomas. MRI is becoming increasingly useful for assessing the burden of cerebral infarcts, as well. Elderly persons are at increased risk for small infarcts in deep brain structures that may occur covertly.[43]

MRI can also detect hippocampal atrophy when imaging is performed in a plane perpendicular to the axis of the temporal lobe[44] [*see Figure 5*]. Hippocampal atrophy is commonly seen in AD, but this finding is suggestive rather than diagnostic of AD because it is also seen both in normal elderly patients and in those with other dementing illnesses.

Positron emission tomography (PET) that uses fluorodeoxyglucose as a tracer may show reduced metabolic activity in the temporoparietal regions in AD,[45] but it is not clear whether this technique augments the clinical diagnosis. A PET tracer known as Pittsburgh compound B (PIB) has been shown to label brain amyloid in patients with AD. Studies are under way to determine whether PET imaging using PIB is of value in the diagnosis of AD.[46]

Evaluation of proteins in the CSF for the specific diagnosis of AD has been encouraging but has not yet reached the stage of definite clinical utility. Tests for three proteins—amyloid-β peptide,[47] tau,[47] and neural thread[48]—have shown modest sensitivity and specificity for differentiating AD from other conditions. However, in almost all instances, these tests have added little to the information derived from the clinical evaluation.

Genotyping of dementia patients for diagnostic purposes is not recommended at this time. Its additive value to the history and examination is quite modest.

DIFFERENTIAL DIAGNOSIS

When considering the diagnosis of dementia from AD, the clinician should first determine whether the patient's clinical profile fits that of dementia—that is, daily functioning has declined from a previously higher level and is now impaired. If the patient was cognitively intact at some earlier point, this eliminates developmental disorders such as mental retardation. The presence of broadly normal levels of arousal and attention excludes delirium. For dementia from AD, onset and progression should be gradual. The general time frame of onset of symptoms is at least 6 months; more often, it is a year or longer. More rapid onset of dementia (i.e., over weeks or months) raises the possibility of Creutzfeldt-Jakob disease [*see 11:XVII Central Nervous System Diseases Due to Slow Viruses and Prions*].

Seizures and headaches are not part of the initial presentation of AD dementia. A patient who presents with dementia, seizures, and headaches should receive a workup for a space-occupying lesion of the brain, such as a brain tumor, subdural hematoma, or CNS infection, all of which typically progress over days to weeks.

When clinical strokes precede dementia, particularly when a stroke occurred within 3 months before the onset of dementia, vascular dementia should be seriously considered. The possibility of vascular dementia also arises when imaging studies disclose bilateral gray-matter infarcts, even if the infarcts were not associated with overt clinical stroke. In patients lacking either of these features, vascular dementia is unlikely.

The presence of parkinsonism or other prominent disorders of motor function points to a diagnosis other than AD. Dementia with Lewy bodies should be considered in a person with Parkinson disease who develops dementia or in someone who initially has dementia and then begins to display signs of parkinsonism, such as a stooped posture and impaired gait and balance. Additional features that suggest dementia with Lewy bodies rather than AD are day-to-day fluctuations in alertness and a peculiar sleep disorder [*see* Dementia with Lewy Bodies, *below*].

The cognitive deficits necessary for making a diagnosis of AD include impairment of short-term memory and of at least one other cognitive domain. However, when disturbances of speaking or language predominate over short-term memory deficits, the diagnosis of progressive aphasia or a disorder known as semantic dementia should be considered. The diagnosis of frontotemporal dementia is suggested when changes in behavior, conduct, or personality overshadow memory difficulties.

In persons older than 60 years, AD constitutes 60% to 80% of all dementias.[49-51] Thus, with patients in this age range, it is reasonable to suspect AD unless some feature in the history or examination strongly points to another specific diagnosis. In contrast, the differential diagnosis of dementia in patients younger than 60 years is quite different. In younger patients, disorders involving altered metabolic pathways are more likely; these include mitochondrial diseases, Wilson disease, and metachromatic leukodystrophy. Diseases such as multiple sclerosis should be considered. In addition, the overlap between psychiatric diseases and dementia may be a prominent diagnostic challenge in younger persons, as may the overlap of substance abuse and progressive cognitive impairment.

TREATMENT

Therapy for AD can be primary or secondary. Primary therapies are intended to stabilize the core symptoms. Secondary therapies are intended to address such problems as depression, anxiety, sleep disorders, and agitation.

Primary Therapies

The cholinesterase inhibitors donepezil,[52] galantamine,[53] and rivastigmine[54] have been approved by the Food and Drug Administration for the treatment of AD. Clinical trials with each of these agents have shown that compared with no treatment, long-term use results in modest stabilization of cognitive and functional status for approximately 6 to 12 months. A particularly illuminating study was one in which patients with mild to moderate AD were randomly assigned to receive either donepezil or placebo for up to 1 year. The end point of therapy was the loss of a predefined degree of daily functioning. Whereas 51% of placebo-treated patients had lost a critical amount of daily functional ability after 12 months of treatment, only 38% of donepezil-treated patients did so.[55]

A 3-year study of donepezil in mild cognitive impairment showed that donepezil delayed progression to AD for about 1 year[56]; the magnitude of the drug's effect in mild cognitive impairment is comparable to that reported in studies on AD.

Side effects of the cholinesterase inhibitors are primarily gastrointestinal and consist of nausea; loss of appetite; diarrhea; and, less commonly, vomiting. These side effects are strongly dose related and often occur when the drugs are initiated or the dose is raised.

Patients with mild to moderate AD are the appropriate candidates for cholinesterase-inhibitor therapy. These patients need a designated caregiver to supervise the use of the medication. Treatment with the cholinesterase inhibitors should be maintained until the patient reaches a stage of severe dementia, in which further decline is expected.

Vitamin E is often recommended for patients with AD because a 2-year study of AD patients with moderate dementia showed that vitamin E (α-tocopherol, 2,000 IU a day) delayed the progression to severe dementia by about 200 days.[57] A study of the use of vitamin E in patients with mild cognitive impairment failed to show any benefits.[56] These results call into question the rationale for vitamin E treatment in patients with mild AD.

Memantine is now also approved by the FDA for the treatment of AD. This agent, a glutamate modulator that is a noncompetitive receptor antagonist of *N*-methyl-D-aspartate, has been the subject of several clinical trials that have reported positive results in moderate to severe dementia.[58,59] It is hypothesized that glutamatergic overstimulation is a part of the pathogenic cycle in AD. In the clinical trials, memantine delayed symptom progression. There is no evidence that memantine affects the biologic course of AD, however. Studies of patients with mild to moderate AD have been completed, but no results are yet available. The results of a study that paired memantine with donepezil suggest that patients with moderate to severe AD should receive both agents.[59]

The field of therapeutics in AD is rapidly evolving, but as yet, no agents in phase II trials about to enter the pivotal phase III trials have shown strong efficacy results. The use of secretase inhibitors involved in the production of the amyloid-β peptide are in early stages of development. A trial using the amyloid-β peptide as a vaccine was halted after serious toxicity emerged, but the general approach of an immunologic attack on the amyloid-β peptide is still receiving serious consideration.

Secondary Therapies

Treatment of depression or anxiety in patients with AD should be pursued as aggressively as in patients without AD, with adherence to the best practices of geriatric pharmacology. Depression frequently coexists with AD and contributes to morbidity and loss of function.[60] The newer agents, such as citalopram, paroxetine, sertraline, and mirtazapine, are generally well tolerated in dementia patients. Dosing can begin at a lower level than might be used in a young adult, and dose increases should be spaced further apart.

Treatment of anxiety presents somewhat more of a challenge in AD patients, because the agents commonly used in younger patients, the benzodiazepines, have distinctly unwanted side effects in AD patients. Drugs such as lorazepam and alprazolam can increase confusion in AD patients. The longer-acting agent clonazepam may be a better choice. Buspirone is another alternative for the treatment of anxiety in AD patients.

Treatment of agitation—which may manifest itself as physical aggressiveness, disruptive verbal agitation, or frightening or anxiety-provoking hallucinations—generally requires antipsychotics. Quetiapine has the significant advantage of being much less likely to induce extrapyramidal effects than both newer and older agents. One of the challenges in using any antipsychotic in elderly patients with AD is that a low dose must be used at initiation of treatment. Several dose increments may be needed to reach the effective therapeutic dose. Given the sometimes catastrophic nature of physical aggressiveness or severe hallucinations in AD patients, caregivers can easily and justifiably become frustrated if control of the symptoms is delayed. Unfortunately, an overly aggressive approach to antipsychotic use can lead to oversedation and serious side effects.

Caregiver Support

Support for the caregivers of AD patients must be an integral part of management. The emotional and physical health of caregivers is critical to long-term patient outcomes. Support may take the form of providing one-on-one education and encouraging participation in support groups, involvement in the Alzheimer's Association (http://www.alz.org), and the use of day care and respite care. For primary care physicians, knowing where to find specialized dementia expertise and how to contact the Alzheimer's Association may be as important as making a correct diagnosis and initiating therapy.

GENETIC COUNSELING

Genetic counseling for individuals with a family history of late-onset AD is usually complex and often must be carried out with incomplete information about the family history. Diagnoses of AD in parents may have been made retrospectively or without autopsy confirmation, and the family history often contains many gaps. Nonetheless, studies in the literature can be useful for gaining some insights into the risks of AD in siblings or children of an AD patient.[20,21,61] Two studies suggest that as age of onset in the proband rises, the genetically mediated risk of AD decreases.[21]

Non–Alzheimer Disease Dementias

VASCULAR DEMENTIA

Vascular dementia (VaD) is the name now used to define dementing illness that results from cerebral infarcts, supplanting older, less precise terms. The definition of VaD unfortunately lacks consensus; several sets of diagnostic criteria are currently in use. The National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l'Enseignement en Neurosciences (NINDS-AIREN) criteria for probable VaD[62] are most widely used and have been adopted for clinical trials. The primary NINDS-AIREN criteria are that (1) the onset or worsening of dementia occurred within 3 months after a clinical stroke; (2) imaging studies show evidence of bilateral infarcts in cortical regions, basal ganglia, thalamus, or white matter; and (3) neurologic examination shows focal neurologic deficits. Although all three criteria are required for VaD, the third criterion is probably redundant if the first two are present; the temporal relationship between a stroke and dementia and the presence of infarcts on imaging are the key clinical features for VaD. Clinical-pathologic correlation studies have shown that this definition is quite specific, meaning that patients who meet these criteria are highly likely to have VaD pathologically. However, the NINDS-AIREN criteria are very insensitive, failing to diagnose VaD in the majority of patients who prove to have VaD at autopsy.[50] Many VaD patients have one key feature but not the other. At the price of sacrificing some specificity to increase sensitivity, the Mayo Clinic criteria for VaD[49] stipulate that either stroke and dementia must be temporally related or bilateral infarcts must be present on imaging.

Prevalence and incidence of VaD are approximately 10% to 20% of those of AD.[63,64] Like AD, VaD becomes more common with advancing age. Risk factors for VaD are identical to those for cerebrovascular disease and include hypertension, diabetes, and cardiovascular disease.

There is no genetic predilection for the typical forms of VaD, but there are very rare hereditary forms of dementia from vasculopathy, such as the CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) syndrome.[65]

The dementia of VaD often becomes apparent in the recovery period after a stroke. Because the motor and sensory deficits of stroke are often overwhelming, the cognitive impairment is often overshadowed at first. VaD may also begin insidiously, because there is a substantial percentage of VaD cases that appear to result from the accumulation of a series of so-called silent or covert infarcts.[43] Patients in this group do not meet the diagnostic criteria of dementia temporally linked to stroke, but they do have brain infarcts, which are best visualized with MRI.

Neuroimaging is essential for the proper diagnosis of VaD. MRI is preferable to CT scan because of its superior ability to detect lacunar infarcts. Typically, a patient with VaD has extensive infarcts [*see Figure 6*]. For VaD, the volume of infarction may not be as critical as its location. Bilaterality of ischemic injury is also relevant, although strategically placed unilateral infarcts in such locations as the caudate nucleus, thalamus, hippocampal formation, or parietal lobe can produce cognitive symptoms that close-

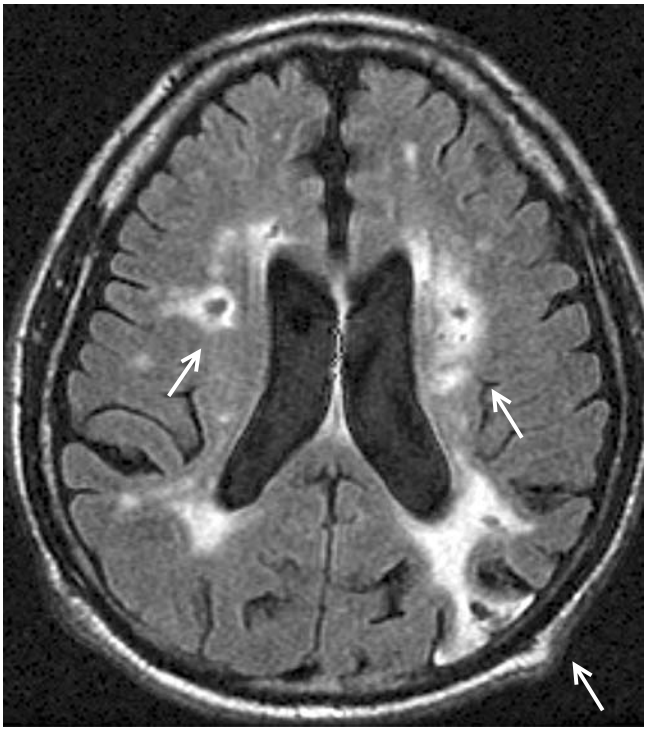


Figure 6 **MRI scan of a patient with vascular dementia showing multiple infarcts (arrows): one parieto-occipital cortical infarct and two white-matter infarcts.**

ly approximate those seen in dementia. White-matter hyperintensities, by themselves, are not sufficient evidence for infarcts.[66]

The cognitive deficits in VaD do not follow a particular pattern. Hence, there is no VaD profile of cognitive dysfunction. VaD patients tend to have more deficits in executive function because infarcts in the caudate nuclei, thalami, or white matter of the frontoparietal lobes tend to disrupt circuits that are involved in executive functions, but some VaD patients have prominent anterograde amnesia just like that of AD, without a prominent executive component.

Therapy for VaD should be primarily directed at prevention through smoking cessation and treatment of hypertension, diabetes mellitus, and known risks for cerebral embolization. Several studies have shown modest reductions in the incidence of dementia (presumably, dementia with a vascular component) by treatment of hypertension.[67,68] In patients with VaD, cholinesterase inhibitors[69,70] and memantine[71] have been effective in delaying progression of symptoms.

The natural history of VaD is variable. Some patients experience prolonged plateaus if they cease having strokes, whereas other patients decline inexorably. VaD patients whose disorder meets NINDS-AIREN criteria have considerably lower survival rates than AD patients.[35]

DEMENTIA WITH LEWY BODIES

Definition of the dementing illness associated with Parkinson disease has undergone several amplifications over the past decade. Dementia is recognized as a relatively common concomitant syndrome in Parkinson disease, and the coappearance of motor system dysfunction from Parkinson disease and dementia from AD is also recognized more commonly. In addition, dramatic alterations in daytime alertness and arousal, as well as the rapid eye movement sleep behavior disorder (RBD; see below), are coming to be recognized as associated with the dementia of Lewy body parkinsonism.

The term dementia with Lewy bodies (DLB)[72] is in vogue as a label for patients who have spontaneous (i.e., not drug-induced) parkinsonism, dementia, and often symptoms of disordered arousal. Although the consensus criteria for DLB lack precision, the importance of the disorder is not in doubt. After AD, DLB is one of the major dementing illnesses of later life,[50] occurring with approximately the same frequency as VaD. The demographics of DLB closely resemble those of AD.

The pathology of DLB includes both the typical Lewy body pathology in the brain stem seen in Parkinson disease and the appearance of cortical Lewy bodies. There is a spectrum of cortical pathology in DLB that ranges from modest numbers of cortical Lewy bodies and prominent AD pathology to profuse cortical Lewy bodies with minimal AD pathology.

The cognitive disorder of DLB can closely resemble AD, but in many patients, there are some notable differences.[73] These differences in DLB include a slightly less prominent deficit in learning and memory and more prominent difficulties with visuospatial functions, performance on timed tasks, and executive functions. The neuropsychological profile is not diagnostic, however.

The parkinsonism in DLB can range from a relatively isolated gait instability with frequent falling to a typical pattern of Parkinson disease with rest tremor, rigidity, bradykinesia, and postural instability.

Patients with DLB often experience marked fluctuations in their alertness and level of arousal from one day to the next. They often sleep excessively. RBD can often precede the dementia and the movement disorder by years and is highly specific for DLB.[74] In RBD, patients engage in dream enactment, thrashing about in bed or talking in their sleep. Another feature of DLB that may be related to the disorder of arousal is frequent daytime hallucinations that can be very vivid and detailed. It is postulated that the prominent hallucinations in DLB represent a sleep activity inappropriately intruding into wakefulness.

The diagnosis of DLB is a clinical one, based on the history and neurologic examination. The differential diagnosis of a dementia with parkinsonism in an elderly person also includes less common movement disorders such as corticobasal degeneration, progressive supranuclear palsy, and multisystem atrophy.

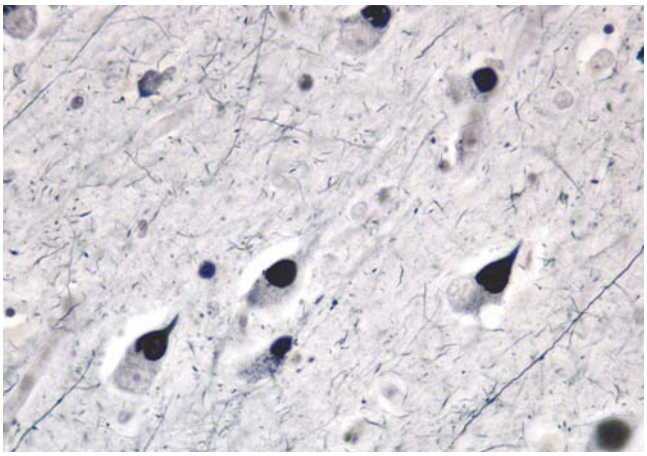
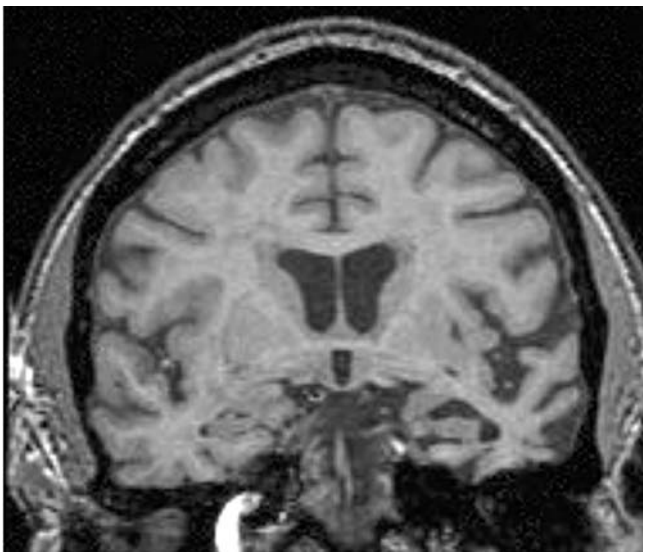


Figure 7 **Photomicrograph of Pick bodies, a feature of a frontotemporal lobar degeneration involving abnormalities of the tau protein.**

a



b

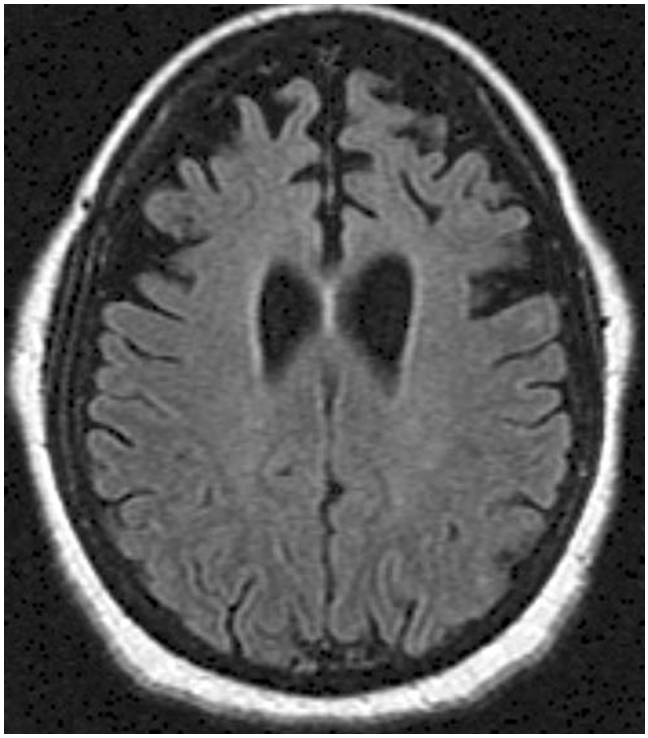


Figure 8 **MRI scan of focal temporal (*a*) or frontal atrophy (*b*) in patients with frontotemporal lobar degenerations.**

In younger patients, Wilson disease or Huntington disease should be considered.

The treatment of DLB can be complex because it involves the interplay of three symptom domains (dementia, movement disorder, and behavioral disturbances) and because medications for one domain may impact another domain. Cholinesterase inhibitors are of some benefit in DLB and do not appear to worsen the parkinsonism or behavior.[75] Levodopa may be essential to manage the gait and balance problems, but it may increase hallucinations or confusion. Antipsychotic agents such as quetiapine or clozapine will not exacerbate the parkinsonism or the confusion.

FRONTOTEMPORAL LOBAR DEGENERATIONS

The frontotemporal lobar degenerations (FTLDs) constitute a much less common subgroup of the dementias.[76] The term FTLD is currently preferred over the older eponymous name Pick disease because of the idiosyncratic and restricted meaning that Pick disease has acquired.

Recognition of the three different types of FTLD—frontotemporal dementia (FTD), progressive aphasia, and semantic dementia—is important because the principal management issues and the prognoses are quite different from those of AD, with which they are unfortunately often confused.

The most common of the three FTLDs is frontotemporal dementia. FTD presents as a disorder of behavioral control and loss of social skills, as well as a disorder of executive cognitive function.[77] Short-term memory may show impairment, but it is either quite modest or definitely less intrusive than the disorder of behavior and social conduct. Progressive aphasia may often not be a dementia, per se, but represents a difficulty with expressive language and increasing impairment in word-finding (anomia). The expressive speech of most patients with progressive aphasia becomes increasingly nonfluent, meaning that the number of words per utterance diminishes, the melody of speech vanishes, speech becomes more labored, and anomia becomes severe. Initially, comprehension and other cognitive functions are often preserved. Some patients with progressive aphasia go on to become demented, whereas others remain nondemented despite worsening expressive speech. The least common of the FTLDs is semantic dementia, a condition in which fluent speech is preserved but loss of word meaning is profound.

Some of the FTLDs have a genetic basis. Mutations in the *tau* gene, on chromosome 17, cause FTLD.[78] The mutations decrease the ability of the tau protein to bind to microtubules, which leads in turn to aggregation of tau molecules into an insoluble fibrillar form. Abnormal accumulations of tau protein are observed in the Pick bodies and the swollen neurons that are seen in some of the FTLDs. In the United States, *tau* mutations in patients with sporadic FTLD are exceedingly rare.[79]

The pathology of the FTLDs is also distinct from that of AD. The brunt of the pathology is localized to frontal and anterior temporal regions. Microscopically, there are several variants of FTLD. In one variant, cellular inclusions known as Pick bodies are seen [*see Figure 7*]; these were the basis for the diagnosis of Pick disease in the past. Pick bodies are positive for the tau protein. In another variant, an amorphous intracellular accumulation of tau protein–positive material can be seen in neurons. In these two pathologic variants of FTLD, the molecular composition of the tau proteins is different from that seen in AD. In the most commonly encountered pathology in the FTLDs, there is a loss of cortical neurons, an intense astrocytic reaction, and often ubiquitin-positive but tau-negative inclusions. The link between the tau-positive and tau-negative FTLDs has not yet been clarified.

The FTLDs can be diagnosed at the time of their initial presentation on the basis of the distinctive clinical profile, neuropsychometric testing, and neuroimaging. Neuropsychological profiles of patients with an FTLD are usually distinguishable from those of AD patients.[80] Many, although not all, patients with an FTLD have focal frontal or anterior temporal lobe atrophy, either symmetrically or asymmetrically [*see Figure 8*]. In addition, functional neuroimaging techniques such as PET and single-photon emission computed tomography are useful in demonstrating reduced metabolism or perfusion in frontal or anterior temporal

regions. There is currently no treatment for the FTLDs, other than the use of agents to control depression, anxiety, or agitation.

The author has received research grants from Elan Pharmaceuticals, Inc., and consulting honoraria from GE Healthcare, Myriad Pharmaceuticals, Inc., and Neurochem, Inc..

Cholinesterase inhibitors have not been approved by the FDA for use in the treatment of non-Alzheimer dementias.

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC, 1994

2. McKhann G, Drachman D, Folstein M, et al: Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34:939, 1984

3. Hy LX, Keller DM: Prevalence of AD among whites: a summary by levels of severity. Neurology 55:198, 2000

4. Hebert LE, Scherr PA, Bienias JL, et al: Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 60:1119, 2003

5. Hebert LE, Scherr PA, McCann JJ, et al: Is the risk of developing Alzheimer's disease greater for women than for men? Am J Epidemiol 153:132, 2001

6. Fillenbaum GG, Heyman A, Huber MS, et al: The prevalence and 3-year incidence of dementia in older black and white community residents. J Clin Epidemiol 51:587, 1998

7. Graves AB, Larson EB, Edland SD, et al: Prevalence of dementia and its subtypes in the Japanese American population of King County, Washington state. The Kame Project. Am J Epidemiol 144:760, 1996

8. Carmelli D, Swan GE, Reed T, et al: Midlife cardiovascular risk factors, ApoE, and cognitive decline in elderly male twins. Neurology 50:1580, 1998

9. Ott A, Stolk RP, van Harskamp F, et al: Diabetes mellitus and the risk of dementia: the Rotterdam Study. Neurology 53:1937, 1999

10. Seshadri S, Beiser A, Selhub J, et al: Plasma homocysteine as a risk factor for dementia and Alzheimer's disease. N Engl J Med 346:476, 2002

11. Mulnard RA, Cotman CW, Kawas C, et al: Estrogen replacement therapy for treatment of mild to moderate Alzheimer disease: a randomized controlled trial. Alzheimer's Disease Cooperative Study. JAMA 283:1007, 2000

12. Aisen PS, Schafer KA, Grundman M, et al: Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. JAMA 289:2819, 2003

13. Shumaker SA, Legault C, Thal L, et al: Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 289:2651, 2003

14. Wolozin B, Kellman W, Ruosseau P, et al: Decreased prevalence of Alzheimer disease associated with 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors. Arch Neurol 57:1439, 2000

15. Whalley LJ, Starr JM, Athawes R, et al: Childhood mental ability and dementia. Neurology 55:1455, 2000

16. Snowdon DA, Kemper SJ, Mortimer JA, et al: Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study. JAMA 275:528, 1996

17. Evans DA, Hebert LE, Beckett LA, et al: Education and other measures of socioeconomic status and risk of incident Alzheimer disease in a defined population of older persons. Arch Neurol 54:1399, 1997

18. Tanzi RE, Bertram L: New frontiers in Alzheimer's disease genetics. Neuron 32:181, 2001

19. Bird TD, Levy-Lahad E, Poorkaj P, et al: Wide range in age of onset for chromosome 1–related familial Alzheimer's disease. Ann Neurol 40:932, 1996

20. Lautenschlager NT, Cupples LA, Rao VS, et al: Risk of dementia among relatives of Alzheimer's disease patients in the MIRAGE study: what is in store for the oldest old? Neurology 46:641, 1996

21. Heston LL, Mastri AR, Anderson VE, et al: Dementia of the Alzheimer type: clinical genetics, natural history, and associated conditions. Arch Gen Psychiatry 38:1085, 1981

22. Corder EH, Saunders AM, Strittmatter WJ, et al: Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261:921, 1993

23. Poirier J, Davignon J, Bouthillier D, et al: Apolipoprotein E polymorphism and Alzheimer's disease. Lancet 342:697, 1993

24. Breitner JC, Wyse BW, Anthony JC, et al: APOE-ε4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. Neurology 53:321, 1999

25. Myers RH, Schaefer EJ, Wilson PW, et al: Apolipoprotein E ε4 association with dementia in a population-based study: the Framingham Study. Neurology 46:673, 1996

26. Hardy J, Selkoe DJ: The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. Science 297:353, 2002

27. Tanzi R: The synaptic Aβ hypothesis of Alzheimer disease. Nature Neuroscience 8:977, 2005

28. Haroutunian V, Perl DP, Purohit DP, et al: Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer disease. Arch Neurol 55:1185, 1998

29. Arnold SE, Hyman BT, Flory J, et al: The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. Cereb Cortex 1:103, 1991

30. Walsh DM, Klyubin I, Fadeeva JV, et al: Amyloid-beta oligomers: their production, toxicity and therapeutic inhibition. Biochem Soc Trans 30:552, 2002

31. Braak H, Braak E: Neuropathological staging of Alzheimer-related changes. Acta Neuropathol (Berl) 82:239, 1991

32. Selkoe D: Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's disease. Nature Cell Biology 6:1054, 2004

33. Iqbal K, Alonso AC, Gong CX, et al: Mechanisms of neurofibrillary degeneration and the formation of neurofibrillary tangles. J Neural Transm Suppl 53:169, 1998

34. Heyman A, Peterson B, Fillenbaum G, et al: Predictors of time to institutionalization of patients with Alzheimer's disease: the CERAD experience, part XVII. Neurology 48:1304, 1997

35. Knopman D, Rocca WA, Cha RH, et al: Survival study of vascular dementia in Rochester, Minnesota. Arch Neurol 60:85, 2003

36. Heyman A, Peterson B, Fillenbaum G, et al: The consortium to establish a registry for Alzheimer's disease (CERAD). Part XIV: Demographic and clinical predictors of survival in patients with Alzheimer's disease. Neurology 46:656, 1996

37. Petersen RC, Smith GE, Waring SC, et al: Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 56:303, 1999

38. Starkstein SE, Sabe L, Chemerinski E, et al: Two domains of anosognosia in Alzheimer's disease. J Neurol Neurosurg Psychiatry 61:485, 1996

39. Devanand DP, Jacobs DM, Tang MX, et al: The course of psychopathologic features in mild to moderate Alzheimer disease. Arch Gen Psychiatry 54:257, 1997

40. Crum RM, Anthony JC, Bassett SS, et al: Population-based norms for the Mini-Mental State Examination by age and educational level. JAMA 269:2386, 1993

41. Tang-Wai DF, Knopman DS, Geda YE, et al: Comparison of short test of mental status and mini mental state examination in mild cognitive impairment. Arch Neurol 60:1777, 2003

42. Knopman DS, DeKosky ST, Cummings JL, et al: Practice parameter: diagnosis of dementia (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 56:1143, 2001

43. Longstreth WT Jr, Dulberg C, Manolio TA, et al: Incidence, manifestations, and predictors of brain infarcts defined by serial cranial magnetic resonance imaging in the elderly: the Cardiovascular Health Study. Stroke 33:2376, 2002

44. Jack CR Jr, Petersen RC, Xu YC, et al: Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. Neurology 49:786, 1997

45. Silverman DH, Small GW, Chang CY, et al: Positron emission tomography in evaluation of dementia: regional brain metabolism and long-term outcome. JAMA 286:2120, 2001

46. Klunk WE, Engler H, Nordberg A, et al: Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 55:306, 2004

47. Galasko D, Chang L, Motter R, et al: High cerebrospinal fluid tau and low amyloid β42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype. Arch Neurol 55:937, 1998

48. Kahle PJ, Jakowec M, Teipel SJ, et al: Combined assessment of tau and neuronal thread protein in Alzheimer's disease CSF. Neurology 54:1498, 2000

49. Knopman D, Parisi JE, Rocca WA, et al: Vascular dementia in a community-based autopsy study. Arch Neurol 60:569, 2003

50. Holmes C, Cairns N, Lantos P, et al: Validity of current clinical criteria for Alzheimer's disease, vascular dementia and dementia with Lewy bodies. Br J Psychiatry 174:45, 1999

51. Lim A, Tsuang D, Kukull W, et al: Clinico-neuropathological correlation of Alzheimer's disease in a community-based case series. J Am Geriatr Soc 47:564, 1999

52. Rogers SL, Farlow MR, Doody RS, et al: A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology 50:136, 1998

53. Tariot PN, Solomon PR, Morris JC, et al: A 5-month, randomized, placebo-controlled trial of galantamine in AD. The Galantamine USA-10 Study Group. Neurology 54:2269, 2000

54. A randomized trial evaluating the efficacy and safety of ENA 713 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease. ENA 713 B352 Study Group. Int J Geriatric Psychopharmacol 1:55, 1998

55. Mohs RC, Doody RS, Morris JC, et al: A 1-year, placebo-controlled preservation of function survival study of donepezil in AD patients. Neurology 57:481, 2001

56. Petersen RC, Thomas RG, Grundman M, et al: Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 352:2379, 2005

57. Sano M, Ernesto C, Thomas RG, et al: A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. The Alzheimer's Disease Cooperative Study. N Engl J Med 336:1216, 1997

58. Reisberg B, Doody R, Stoffler A, et al: Memantine in moderate-to-severe Alzheimer's disease. N Engl J Med 348:1333, 2003

59. Tariot PN, Farlow MR, Grossberg GT, et al: Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. JAMA 291:317, 2004

60. Forsell Y, Winblad B: Major depression in a population of demented and nondemented older people: prevalence and correlates. J Am Geriatr Soc 46:27, 1998

61. Silverman JM, Smith CJ, Marin DB, et al: Familial patterns of risk in very late-onset Alzheimer disease. Arch Gen Psychiatry 60:190, 2003

62. Roman GC, Tatemichi TK, Erkinjuntti T, et al: Vascular dementia: diagnostic crite-

ria for research studies: report of the NINDS-AIREN International Workshop. Neurology 43:250, 1993

63. Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 54:S4, 2000

64. Incidence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 54:S10, 2000

65. Desmond DW, Moroney JT, Lynch T, et al: The natural history of CADASIL: a pooled analysis of previously published cases. Stroke 30:1230, 1999

66. Fazekas F, Kapeller P, Schmidt R, et al: The relation of cerebral magnetic resonance signal hyperintensities to Alzheimer's disease. J Neurol Sci 142:121, 1996

67. Forette F, Seux ML, Staessen JA, et al: Prevention of dementia in randomised double-blind placebo-controlled Systolic Hypertension in Europe (Syst-Eur) trial. Lancet 352:1347, 1998

68. Tzourio C, Anderson C, Chapman N, et al: Effects of blood pressure lowering with perindopril and indapamide therapy on dementia and cognitive decline in patients with cerebrovascular disease. Arch Intern Med 163:1069, 2003

69. Erkinjuntti T, Kurz A, Gauthier S, et al: Efficacy of galantamine in probable vascular dementia and Alzheimer's disease combined with cerebrovascular disease: a randomised trial. Lancet 359:1283, 2002

70. Wilkinson D, Doody R, Helme R, et al: Donepezil in vascular dementia: a randomized, placebo-controlled study. Neurology 61:479, 2003

71. Orgogozo JM, Rigaud AS, Stoffler A, et al: Efficacy and safety of memantine in patients with mild to moderate vascular dementia: a randomized, placebo-controlled trial (MMM 300). Stroke 33:1834, 2002

72. McKeith IG, Dickson DW, Lowe J, et al: Diagnosis and management of dementia with Lewy bodies: third report of the DLB consortium. Neurology October 19, 2005 [epub ahead of print]

73. Ferman TJ, Boeve BF, Smith GE, et al: Dementia with Lewy bodies may present as dementia and REM sleep behavior disorder without parkinsonism or hallucinations. J Int Neuropsychol Soc 8:907, 2002

74. Boeve BF, Silber MH, Ferman TJ, et al: REM sleep behavior disorder and degenerative dementia: an association likely reflecting Lewy body disease. Neurology 51:363, 1998

75. McKeith IG, Grace JB, Walker Z, et al: Rivastigmine in the treatment of dementia with Lewy bodies: preliminary findings from an open trial. Int J Geriatr Psychiatry 15:387, 2000

76. Knopman D, Petersen R, Edland S, et al: The incidence of frontotemporal lobar degeneration in Rochester, Minnesota, 1990–1994. Neurology 62:506, 2004

77. Neary D, Snowden JS, Gustafson L, et al: Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology 51:1546, 1998

78. Reed LA, Wszolek ZK, Hutton M: Phenotypic correlations in FTDP-17. Neurobiol Aging 22:89, 2001

79. Houlden H, Baker M, Adamson J, et al: Frequency of tau mutations in three series of non-Alzheimer's degenerative dementia. Ann Neurol 46:243, 1999

80. Rascovsky K, Salmon DP, Ho GJ, et al: Cognitive profiles differ in autopsy-confirmed frontotemporal dementia and AD. Neurology 58:1801, 2002

Acknowledgments

Figure 2 Seward Hung.

Figures 3, 4, and 8 Joseph Parisi, M.D., Mayo Clinic.

XII EPILEPSY

James W. Wheless, m.d.

Definition

Seizures are caused by transient, paroxysmal, and synchronous discharges of groups of neurons in the brain. Although considered abnormal, seizures can occur in both normal and abnormal brain tissue. Seizures are one of the most common neurologic disorders and may occur at any age. More than 10% of the population will have a seizure or a few seizures during a lifetime. The clinical manifestations of seizures depend on the location and number of the neurons involved in the seizure discharge, the spread of the discharge to other parts of the brain, and the duration of the discharge. Seizures may be caused by a transient abnormality of brain function resulting from such clinical events as hypoglycemia, hyponatremia, and drug toxicity. In these circumstances, a person usually has only a single seizure or a few seizures at most; treatment of the underlying abnormality corrects the seizure disorder.

Epilepsy is defined as recurrent seizures. Epilepsy occurs in 1% to 2% of the population. The incidence of epilepsy is high in infancy but decreases during childhood; it is lowest in adolescence and adulthood but increases greatly in the elderly.[1]

Epilepsy may result from hereditary or acquired factors. Hereditary epilepsies typically have a complex pattern of inheritance, with simultaneous involvement of multiple genes, and the identity of genes that govern susceptibility remains largely unknown.[2] An increasing number of hereditary epileptic syndromes have been traced to mutations in genes that encode for ion channels—membrane-spanning proteins that form selective pores for sodium, potassium, chloride, or calcium ions and that provide the basis for regulating excitability in the central nervous system. Diseases resulting from mutated ion channels are termed channelopathies.[3]

Acquired epilepsy is often the result of a chronic, static neurologic abnormality. Causes include head trauma, CNS infection, structural brain malformations, vascular malformations, strokes, and cerebral neoplasm.

Classification

The International League Against Epilepsy has classified epileptic seizures on the basis of clinical and electroencephalographic criteria [*see Table 1*]. Proposed in 1981, this classification is widely accepted by neurologists but is still not well known among nonneurologists. The classification divides seizures into three major categories: partial, generalized, and unclassified.[4]

Partial seizures

Partial seizures are described as either simple or complex, depending on whether consciousness remains intact or is impaired during the seizure.[5] Simple partial seizures can arise from any neocortical region. The manifestations fall into four categories: motor, sensory, autonomic, or psychic. Motor symptoms include focal myoclonus of a limb or region of the hand or face. Sensory symptoms include seeing colored spots or lines, as occurs with discharge in the primary visual cortex. A common autonomic symptom in mesial temporal lobe epilepsy is a rising epigastric sensation. Psychic symptoms may involve memory (e.g., déjà vu), affect, or other complex phenomena.[5]

In complex partial seizures, the focal discharge involves brain regions subserving awareness or spreads widely enough to reach those regions, whereby the patient loses conscious contact. Although the patient's level of consciousness is a key component of the definition of complex partial seizures, this information may be difficult to obtain. During altered conscious contact, patients are unable to respond to commands, interact with their surroundings, or recall events that occurred during the seizure. A report of behavior observed by a witness such as a family member can be important for differentiating a simple seizure from a complex one. For example, a patient may lose the ability to speak during a simple seizure originating in the dominant temporal, frontal, or parietal cortex. However, if the patient followed commands during a seizure or recalled specific verbal information, the physician can conclude that the patient retained cognition during the seizure.[6]

Complex partial seizures usually begin with arrest of motion and a blank stare. Automatisms (e.g., simple hand movements), oroalimentary behavior (e.g., tasting movements or swallowing), or verbal utterances may occur either initially or during the seizure. If the patient is engaged in a complex motor task at the beginning of a seizure, that activity may continue, but the accuracy of the behavior will deteriorate. Key historical points are the duration of the seizure, because most complex partial spells

Table 1 International Classification of Epileptic Seizures*

- *Partial (Focal, Local) Seizures*
  - Simple partial seizures (consciousness not impaired)
    - Motor signs
    - Somatosensory or special sensory symptoms
    - Autonomic symptoms or signs
    - Psychic symptoms
  - Complex partial seizures (consciousness impaired)
    - Simple partial onset followed by impaired consciousness
    - Consciousness impaired at onset
  - Partial seizures evolving to generalized seizures (tonic, clonic, or tonic-clonic)
    - Simple partial seizures evolving to generalized seizures
    - Complex partial seizures evolving to generalized seizures
    - Simple partial seizures evolving to complex partial seizures evolving to generalized seizures
- *Generalized Seizures (Convulsive or Nonconvulsive)*
  - Absence seizures
    - Typical (brief stare, eye flickering, no emotion)
    - Atypical (associated with movement)
  - Myoclonic seizures
  - Clonic seizures
  - Tonic seizures
  - Tonic-clonic seizures
  - Atonic seizures
- *Unclassified Epileptic Seizures*

*From the Commission on Classification and Terminology of the International League Against Epilepsy.[4]

last only a minute or two, and the stereotyped nature of behaviors. A seizure is stereotyped when the behaviors are the same or similar for each seizure. At the termination of the seizure, the patient may be momentarily confused, fatigued, or disoriented. This alteration is of critical diagnostic importance and must be sought with leading questions by the physician. Postictal effects differentiate a complex partial seizure of temporal lobe origin from an absence seizure; in the latter, the patient does not have any postictal symptoms.[7]

Complex partial seizures may be heralded by sensory, autonomic, or psychic symptoms that precede impairment of consciousness. These symptoms are referred to as an aura.

Simple and complex partial seizures may secondarily generalize and produce a tonic-clonic seizure or a convulsion. Spread of discharge from a localized focus to encompass the entire brain is common, but clinically, most partial seizures do not progress to a secondarily generalized tonic-clonic seizure.

Because the focal manifestation of a secondarily generalized convulsion may not be identified by observers or recalled by the patient, it is best to assume that a newly diagnosed generalized tonic-clonic seizure originated from a focus until such a process can be excluded.[8] Focal processes to consider in an adult with new-onset partial or tonic-clonic seizures include brain tumor, cerebral infarction, or infection.

GENERALIZED SEIZURES

Generalized seizures cause a spectrum of behavior from the nonconvulsive pattern of simple absence seizure through myoclonus to the fully developed generalized tonic-clonic seizure.[4] Absence seizures, which are manifestations of an epilepsy syndrome that begins in childhood, are brief, usually lasting 10 seconds or less. The seizures are not preceded by an aura or followed by postictal effects. During an absence seizure, the patient has arrest of behavior and a blank stare. Occasionally, mild clonic manifestations occur, such as subtle eyelid blinking and changes in postural tone. A specific electroencephalographic pattern of generalized spike-and-wave discharges at 3/sec occurs during absence seizures. This EEG abnormality can be activated by deliberate hyperventilation. EEGs taken between seizures (interictal) may show unrecognized absence seizures or short bursts (usually less than 3 seconds) of the generalized discharge. Patients with these abnormalities have normal intellectual and neurologic function.

Atypical absence seizures differ from simple absence seizures in that onset occurs at an earlier age in childhood, the interictal EEG background is abnormal, and the EEG discharges are slower than 3/sec. Patients with these seizures may also have atonic and myoclonic seizures. Patients with atypical absence seizures may be mentally retarded.

Myoclonus consists of brief jerks or contractions of a specific muscle or group of muscles. Hypnagogic myoclonus is both common and normal. Focal myoclonus may be caused by destructive lesions of the brain stem or spinal cord. Metabolic disease, hypoxia, toxic processes, and infectious disease may cause focal or diffuse myoclonus. Myoclonus associated with epilepsy is commonly symmetrical. Myoclonus is a component of several epilepsy syndromes and occurs before or as a part of both absence and generalized tonic-clonic seizures. One important syndrome to identify is juvenile myoclonic epilepsy (see below). Myoclonic seizures may be caused by disorders other than primary brain dysfunction; these include metabolic diseases or genetic brain disorders such as Lafora disease.

Atonic seizures involve a sudden loss of postural tone and may cause falls, with trauma to the head or body. This treatment-refractory seizure type is often associated with the Lennox-Gastaut syndrome.

Convulsions are the most common type of generalized

Table 2 International Classification of Epilepsies, Epilepsy Syndromes, and Related Seizure Disorders*

Localization-related (focal, local, partial) epilepsies†	Idiopathic Benign childhood epilepsy with centrotemporal spikes Childhood epilepsy with occipital paroxysms Primary reading epilepsy Symptomatic Temporal lobe epilepsy Frontal lobe epilepsy Parietal lobe epilepsy Occipital lobe epilepsy Chronic progressive epilepsia partialis continua Cryptogenic (presumed to be symptomatic but cause is unknown) Temporal lobe epilepsy Frontal lobe epilepsy Parietal lobe epilepsy Occipital lobe epilepsy Chronic progressive epilepsia partialis continua
Generalized epilepsies	Idiopathic Benign neonatal convulsions (familial and nonfamilial) Benign myoclonic epilepsy in infancy Childhood absence epilepsy Juvenile myoclonic epilepsy Epilepsy with generalized tonic-clonic seizures on awakening Symptomatic Nonspecific etiology Early myoclonic encephalopathy Early infantile epileptic encephalopathy with suppression burst Other symptomatic generalized epilepsies Cryptogenic West syndrome (infantile spasms) Lennox-Gastaut syndrome Epilepsy with myoclonic-astatic seizures Epilepsy with myoclonic absences Specific syndromes (disease states in which seizures are a presenting or predominant feature)
Undetermined epilepsies	Generalized and focal features Neonatal seizures Severe myoclonic epilepsy of childhood Epilepsy with continuous spike waves during slow-wave sleep Acquired epileptic aphasia (Landau-Kleffner syndrome) Other undetermined epilepsies without unequivocal generalized or focal features
Special syndromes	Situation-related seizures Febrile convulsions Isolated seizures or status epilepticus Seizures caused by an acute or toxic event, such as alcohol or drug overdose, eclampsia, or hyperglycemia

*From the Commission on Classification and Terminology of the International League Against Epilepsy.[4]

†Syndrome defined by seizure type and other clinical features, including anatomic localization and etiology.

seizures. They are characterized by loss of consciousness associated with apnea and violent contractions of the musculature of the trunk and extremities. Frequently, patients suffer mouth trauma and have bladder incontinence. Salivation often increases, and both pulse rate and blood pressure rise during the seizure. Most generalized convulsions begin with a tonic phase, in which there is sustained contraction of all muscles with extended legs and either flexed or extended arms. This phase lasts for several seconds and is followed by a clonic phase, in which there are rhythmic contractions of the limbs that begin with high-frequency, low-amplitude movements and then gradually decrease in frequency for several seconds to a few minutes. Some patients may have only tonic seizures or only clonic seizures. A sequence of alternating tonic and clonic movements may be observed in patients with primary generalized seizures. After the violent muscle contractions subside, the patient enters a postictal phase, in which breathing resumes and unresponsiveness is followed by gradual recovery of consciousness. The patient may remain confused for several minutes or longer and may subsequently complain of muscle pain and headache. This sequence of behavior is stereotyped, occurring regardless of cause. If a localized structural lesion in the brain cannot be found, the convulsion is an idiopathic or a primary generalized seizure. If the convulsion is preceded by a partial seizure, it is said to be a secondarily generalized seizure. About 75% to 85% of convulsions in adults are preceded by partial seizures.

Epilepsy Syndromes

In 1989, the International League Against Epilepsy recognized that many patients with seizures have brain abnormalities that affect their quality of life independent of the epilepsy alone. This classification[4] recognizes that an epilepsy syndrome encompasses not only the behavior during a seizure but also the EEG changes, the patient's mental and motor development, and the family history [*see Table 2*]. Because the repertoire of seizure manifestations is limited, similar seizure types may be found as components of several syndromes with widely divergent prognoses. Defining a specific epilepsy syndrome often requires repeated assessment, evaluation of development, and review of responses to treatment. Syndromes are considered to be either benign or progressive. These terms are usually applied to ultimate outcome of intellectual function and survival. Although some syndromic seizures are benign in their impact on intellectual function, lifelong treatment with antiepileptic drugs (AEDs) may be required. Five pertinent epilepsy syndromes are described in this subsection: febrile seizures, benign childhood epilepsy with centrotemporal spikes, childhood absence epilepsy, juvenile myoclonic epilepsy, and chronic progressive epilepsia partialis continua.

FEBRILE SEIZURES

Febrile seizures affect 2% to 5% of children. These seizures typically occur before 6 years of age. Although one third of children who have a febrile seizure will have recurrent febrile seizures, only a small minority develop afebrile seizures. A syndrome called febrile seizures plus is now recognized; patients with this syndrome have febrile seizures that continue beyond 6 years of age or have associated afebrile tonic-clonic seizures. Febrile seizures plus is self-limited, with seizures typically ceasing by mid-adolescence. The syndrome appears to be genetic, with an autosomal dominant inheritance pattern. In turn, febrile seizures plus appears to be one manifestation of a broader syndrome termed generalized epilepsy with febrile seizures plus, whose spectrum of clinical manifestations also includes febrile seizures. Genetic linkage studies have mapped the syndrome to multiple loci.[9,10] Patients with this syndrome may have febrile seizures in childhood, followed by a quiescent interval of many years and then the onset of seizures of a different type.

BENIGN CHILDHOOD EPILEPSY WITH CENTROTEMPORAL SPIKES

This common syndrome is characterized by hemifacial motor seizures that are brief, partial, usually nocturnal, and occasionally generalized. Patients also have somatosensory symptoms such as facial twitching and tongue numbness. These children develop normally and have no focal changes evident on brain imaging. The EEG shows sleep activation of high-voltage centrotemporal spikes, with a blunt configuration and a characteristic slow wave. Complete remission by the mid-teenage years is usual in more than 90% of cases.[11]

CHILDHOOD ABSENCE EPILEPSY

The prevalence of childhood absence epilepsy is estimated to be 2% to 8%. The associated EEG abnormality of spike-and-wave discharges at 3/sec occurs in the milieu of a normal background EEG pattern [*see Figure 1*].[12] Remissions of 80% are reported in patients with uncomplicated absence seizures. However, tonic-clonic seizures develop in some cases, and only 30% of those patients have complete resolution without antiepileptic medication. Patients who are at least 15 years old at onset of absence epilepsy also do not have a good prognosis for complete resolution without antiepileptic medication.[13] Remission is most likely with short duration of illness, which highlights the need for early identification of patients and rapid institution of treatment. Factors associated with a good prognosis are normal intelligence quotient (IQ) and no history of generalized tonic-clonic seizures. Unfortunately, tonic-clonic seizures complicate the course of absence epilepsy in about 50% of patients.[11] Risk factors for development of tonic-clonic seizures include later age at onset of absence epilepsy, difficulty in controlling absence seizures with medication, and abnormal background activity on the EEG.[12,14] Before the development of broad-spectrum drugs, some patients were given an additional drug to prevent tonic-clonic seizures. This dual-drug approach has been supplanted by the use of valproate, a broad-spectrum medication with combined efficacy against both absence and generalized tonic-clonic seizures. Discussion of treatment with patients and parents must include informing them of the risk of developing tonic-clonic seizures.[15]

JUVENILE MYOCLONIC EPILEPSY

Although the onset of juvenile myoclonic epilepsy occurs in childhood, the disorder typically persists throughout adulthood. Juvenile myoclonic epilepsy is genetically transmitted; patients typically have a family history of seizures. The genetic defect is an abnormality on the short arm of chromosome 6.

Patients with juvenile myoclonic epilepsy have characteristic myoclonic jerks, generalized tonic-clonic seizures, and photosensitivity that is revealed by EEG. Precipitants of seizures include sleep deprivation, stress, and alcohol use. The myoclonic jerks are quite prominent in the morning and may involve the large muscles of the legs. The correct diagnosis is important because these seizures are usually well controlled with valproate.

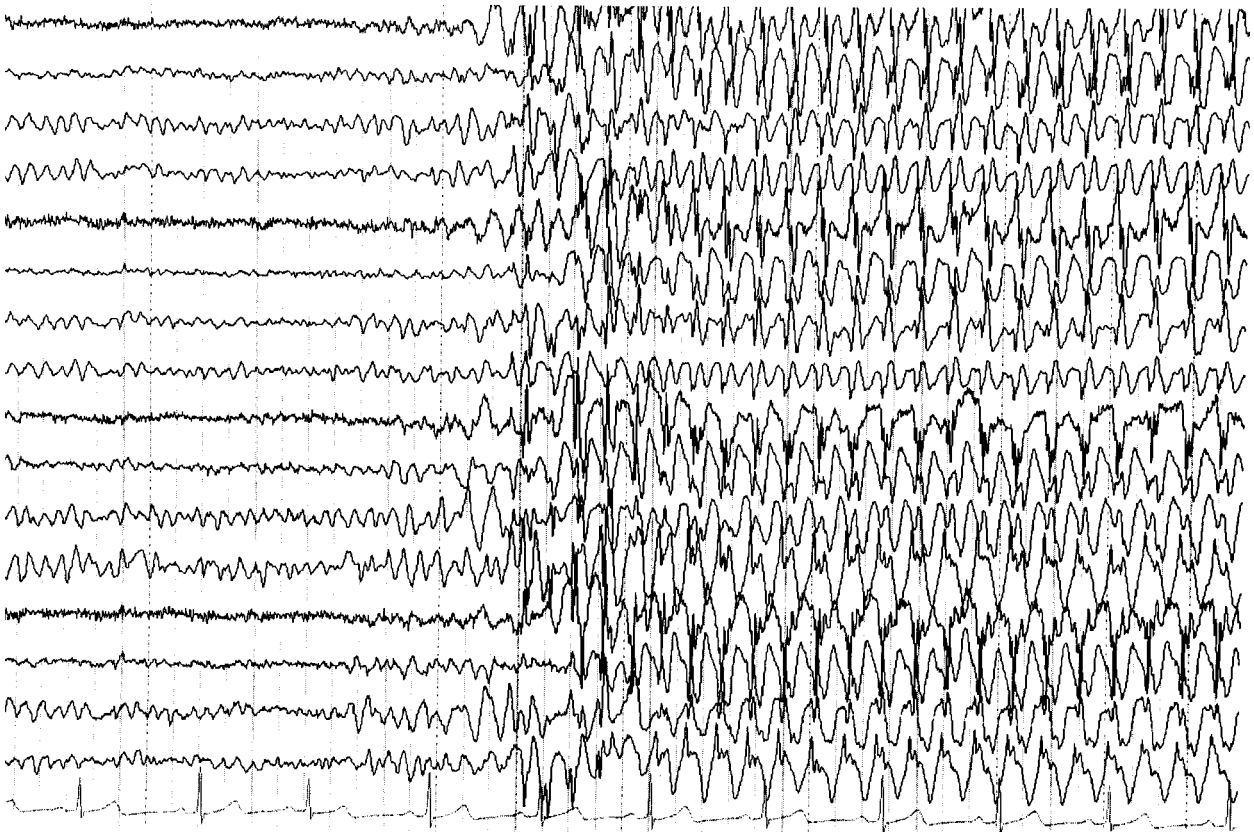


Figure 1 **Electroencephalogram obtained from electrodes affixed to the patient's scalp and reformatted with analog-to-digital conversion. The normal background patterns are interrupted by generalized polyspike and then spike and slow-wave discharges at 3/sec. This EEG is typical for that recorded from patients with childhood absence epilepsy. The dark vertical lines denote 1 second; the sensitivity is 7 μV/mm.**

The prognosis for normal intellectual function is usually good; however, the outlook for drug-free remission is poor.[13]

CHRONIC PROGRESSIVE EPILEPSIA PARTIALIS CONTINUA

Patients with partial seizures or a focal brain abnormality may experience continuous focal seizures. The EEG shows continuous focal spikes and slow waves. Prognosis is related to the underlying pathogenic lesion and is usually not associated with progression in clinical pattern. The childhood form (Rasmussen encephalitis)[16] begins with focal seizures involving motor systems and then progresses to motor deficits and mental deterioration.[11] This rare form of encephalitis also occurs in adults.

Diagnostic Evaluation

Whereas the patient's history forms the basis for characterization of seizures, several laboratory studies are required to establish the diagnosis, management regimen, and prognosis in epilepsy. Most patients with epilepsy have more than one type of seizure. Integrating the clinical and laboratory data can allow the clinician to identify the epilepsy syndrome involved; this provides a powerful tool for patient care.[5]

ELECTROENCEPHALOGRAPHY

An EEG is a graphic or electronic display of amplified physiologic brain activity recorded from electrodes attached to a patient's scalp in a standardized array. Focal or generalized seizure patterns interrupting background patterns confirm the diagnosis of epilepsy. Because sleep increases epileptiform activity, the diagnostic yield of an EEG is improved when the patient is sleep deprived, so that natural sleep will occur during the study. Patients with partial seizures arising in mesial temporal lobe structures often have a normal EEG on the initial study. Repeated EEGs are needed to obtain a confirmatory study. Diagnostic patterns include focal spikes, sharp waves, or spike-wave complexes [*see Figure 2*]. Generalized abnormalities range from momentary suppression of background rhythms to clusters of multiple rapid spikes to the classic pattern of spike-and-wave discharges at 3/sec characteristic of absence epilepsy.

Video-Electroencephalographic Monitoring

In most patients with epilepsy, routine EEG provides sufficient diagnostic information. In selected patients with recurrent unprovoked seizures that are refractory to therapy, the combination of video and EEG can help confirm the diagnosis of a seizure disorder and classify seizure type. In some cases, video-EEG monitoring discloses physiologic or psychological disorders that have been confused with epilepsy. Video-EEG monitoring can be performed on an outpatient basis or in an epilepsy monitoring unit.[17]

BRAIN IMAGING

Because epilepsy is a symptom, particularly in adults, treatable illnesses must be excluded in the initial diagnostic evalua-

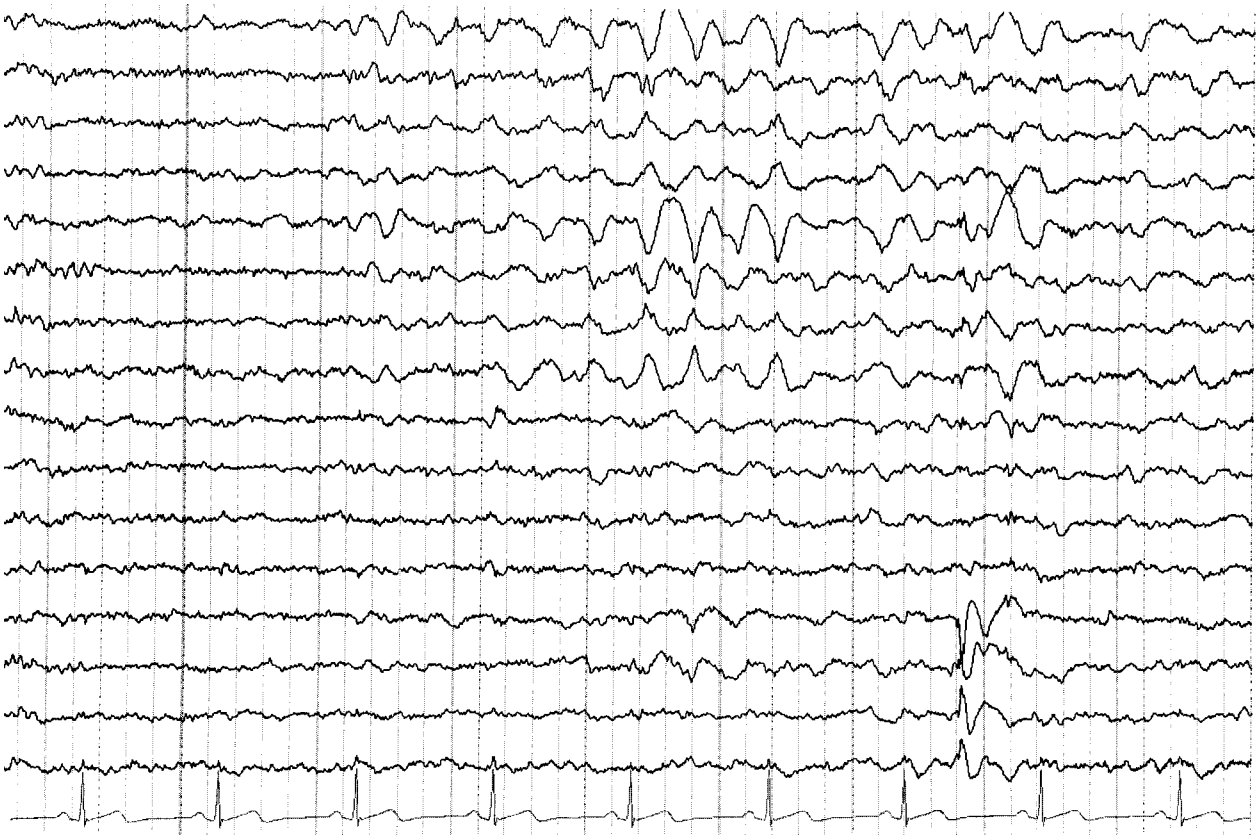


Figure 2 **Electroencephalogram obtained from electrodes affixed to the patient's scalp and reformatted with analog-to-digital conversion. An isolated focal epileptiform discharge present in the lower four traces occurs 7 seconds after the beginning of this EEG epoch. This interictal EEG pattern is typical for a patient with complex partial seizures of temporal lobe origin. The darker vertical lines denote 1 second; the sensitivity is 7 μV/mm.**

tion. A CT scan of the brain will reveal whether the patient has lesions needing urgent treatment, but magnetic resonance imaging is the preferred method for studying brain structure in a patient with epilepsy; indeed, structural MRI is a routine part of the epilepsy workup. Patients with intractable complex partial seizures arising from temporal lobe structures have mesial temporal atrophy or sclerosis as their most common focal lesion; detecting such lesions is of prognostic value.[18] MRI changes in these patients typically consist of increased signal intensity and loss of detailed hippocampal structure with atrophy [*see Figure 3*]. Quantitative MRI aids in identification of hippocampal atrophy.[19]

MRI is both sensitive and specific in patients with partial seizures, allowing noninvasive diagnosis through detection of many types of lesions.[20] These lesions include such neoplastic changes as low-grade glioma and ganglioglioma and specific lesions such as dysembryoplastic neuroepithelial tumors. Vascular changes include arteriovenous malformation, cavernous hemangioma, venous angioma, and telangiectasias. Developmental abnormalities include such neuronal migration disorders as lissencephaly, pachygyria, band or laminar heterotopia, and subependymal heterotopias. More restricted lesions include focal cortical dysplasia, polymicrogyria, focal subependymal heterotopia, and schizencephaly.[20]

Imaging techniques that show cerebral function play a growing role in the evaluation of epilepsy. Such techniques include functional MRI, magnetoencephalography (MEG), magnetic resonance spectroscopy (MRS), single photon emission computed tomography (SPECT), and positron emission tomography (PET).[21]

The use of functional MRI to evaluate epilepsy is a relatively recent development. Functional MRI can detect transient focal or regional changes in cerebral blood flow associated with seizure activity. The technique shows promise in the preoperative assessment of candidates for surgical excision of epileptogenic foci.[21,22]

MEG is a noninvasive method of imaging brain function. MEG identifies the location of activated sets of neurons, on the basis of the associated magnetic flux on the head surface, and projects the location onto an MRI for visualization of the activated region. Currently, MEG is the only imaging modality approved by the Food and Drug Administration that can provide noninvasive imaging of the motor cortex, the sensory cortex, and language function. MEG has an emerging role in the preoperative evaluation of patients with refractory epilepsy.[23]

MRS measures concentrations of a variety of substances within the brain, offering a way to obtain metabolic information in patients with epilepsy. It can be used to detect increased inorganic phosphate concentration, altered pH, and decreased levels of phosphomonoesters. Alterations in levels of *N*-acetylaspartate, choline, and creatine/phosphocreatine and increased lactic acid concentration postictally may aid identification of the location of onset of focal seizures.[24]

SPECT reveals decreased blood flow in the region of seizure

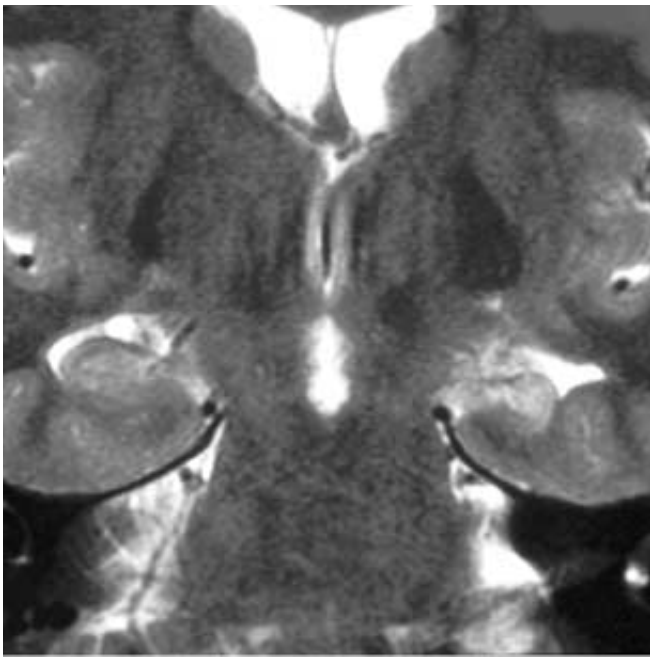


Figure 3 **Mesial temporal sclerosis, a cause of seizures that may be amenable to surgical correction, is visible only on MRI with a specific epilepsy sequence. This coronal view, obtained perpendicular to the hippocampus, demonstrates left mesial temporal sclerosis (the left hippocampus is on the right side of the figure). Note the smaller size and increased signal (brighter appearance) of the left hippocampus and the compensatory increased size of the adjacent temporal horn of the lateral ventricle. The patient, an adult, presented with new-onset complex partial seizures and a history of two prolonged febrile seizures in childhood. A CT scan of the brain was normal.**

origin in some patients.[25] Use of SPECT alone between seizures is currently of little value, but ictal SPECT may be helpful with patients (especially those with temporal lobe epilepsy) in whom other imaging techniques give normal results. In the future, the most important role for SPECT may be its use in combination with MRI for presurgical evaluation. So-called ictal difference images, which are created through coregistation of interictal and ictal SPECT images, can in turn be coregistered with MRI results to provide an accurate picture of cerebral function.[21]

In the past, PET with 18-fluorodeoxyglucose to image the cerebral metabolic rate was useful for detecting a seizure focus that might be amenable to surgery. With advances in MRI technology, however, the value of PET is now limited to selected patients with normal MRI results.[21] In patients with complex partial seizures of temporal lobe origin, PET may reveal interictal hypometabolism, depending on the duration of epilepsy.[26]

Differential Diagnosis

Although sudden alterations in neurologic function are characteristic of seizures, sudden alterations also occur when intracranial structures are deprived of glucose or oxygen. Failure to maintain adequate cerebral perfusion will cause loss of consciousness, as in syncope of any cause. Motor activity during simple syncope is uncommon. However, prolonged interruption of cerebral perfusion may cause convulsive movements or even tonic-clonic seizures.

Sensory or motor dysfunction can be caused by transient ischemia, which may result from embolic cerebrovascular disease, extracranial carotid or basilar artery disease, and migraine. Seizures may occur with embolic strokes. Metabolic disease, particularly disease related to glucose metabolism that requires regulation of blood glucose concentration with insulin therapy, may be associated with episodes of hypoglycemia. Both focal and generalized seizures can be manifestations of hypoglycemia or of hyperosmolar states.

All patients with initial presentation of a seizure must undergo a complete blood count and blood chemistry studies. The CBC results may suggest infection, a platelet abnormality, or anemia. The blood chemistry studies must include measurement of glucose, urea nitrogen, electrolyte, and liver enzyme levels.

Lumbar puncture with a cerebrospinal fluid study should be performed if the patient is febrile or has altered cognitive function, such as behavioral changes, that could be ascribed to an encephalopathic process. To prevent complications from lumbar puncture, the clinician must first exclude the presence of an intracranial mass or increased intracranial pressure.

NONEPILEPTIC SEIZURES

Approximately 20% of patients admitted to epilepsy monitoring units for diagnostic evaluation have episodic behavioral alterations that are not caused by physiologic dysfunction of the brain.[27] In the past, these alterations were described as pseudoseizures. Although that term is still used for communication among physicians, it is not appropriate for communication with patients.[28] Currently, the preferred term is nonepileptic seizures (or nonepileptic events). Use of these terms tends to help patients understand their problem and facilitates referral for behavioral therapy. Patients react angrily to the term pseudo and are less likely to believe the physician is interested in their problem.

An important clue to the diagnosis of nonepileptic seizures is that they are periodic events that tend not to be stereotyped. Both patients and observers report varied behaviors with each event. Another clue is prolonged duration. Nonepileptic seizures may last 30 minutes to several hours. Epileptic seizures, both partial and generalized, seldom continue for more than several minutes. Patients with both nonepileptic seizures and epilepsy pose a challenging clinical problem; this combination is occasionally found in patients undergoing assessment in epilepsy monitoring units. Treatment of nonepileptic seizures requires behavioral intervention. If both disorders are found, treatment of epilepsy needs to be continued in parallel with behavioral therapy.

Treatment of Seizures and Epilepsy

Historically, drug therapy with AEDs has been the mainstay of treatment for seizures. Nonpharmacologic options have gained greater prominence in recent years, however. Use of vagus nerve stimulation devices is growing rapidly. Some patients may be candidates for surgical treatment. In children, the ketogenic diet has reemerged.

DRUG THERAPY

Antiseizure medications, usually referred to as antiepileptic drugs, have been available for more than a century. However, the modern era of AED therapy began in the early 20th century with the introduction of phenobarbital. More than two dozen agents available in the United States are classified as AEDs.

Most were introduced before 1980, but a number of new drugs were licensed for epilepsy treatment in the 1990s.

AED treatment should be directed at both controlling seizures and, when possible, correcting the underlying disease or disorder. AEDs may be used only briefly, if at all, in patients with a single seizure or a few seizures resulting from a transient disorder, such as drug intoxication, withdrawal from alcohol or sedative-hypnotic drugs, hyponatremia, or hypoglycemia. Patients who have recurrent seizures secondary to a treatable neurologic disease, such as brain tumor or intracranial infection, should be treated with AEDs; the underlying problem should also be treated. In many patients, such as those with completed cerebral infarction or brain contusion secondary to head injury, recurrent seizures result from the prior insult or from an undefined process. In these patients, AEDs are the only possible therapy. Patients with chronic recurrent seizures, regardless of etiology, should be treated with AEDs.

Whether to start AED treatment after a single unprovoked first seizure is a controversial issue. Many authorities argue that because many patients will not have a second seizure during several years of follow-up, there may be no justification for beginning AED treatment and subjecting patients to possible side effects.[29] In support of this idea is a retrospective study reporting that only 60% of patients who had a single convulsion had a second one within 5 years.[30] Also, the risk of seizure recurrence diminishes progressively the longer a patient remains seizure free. Because many patients have recurrent seizures, however, not all physicians accept that justification is lacking for AED treatment after single first seizures. I recommend not treating patients after a single seizure if results of the neurologic examination, EEG, and imaging study of the brain are normal. In addition, the patient must be willing to accept the risks associated with another seizure. Treatment with an AED should be initiated for most patients after a first seizure when evaluation reveals any structural or functional brain abnormality such as epileptiform discharges on the EEG.

Treatment with AEDs should follow certain basic principles. Therapy should be started with a single suitable agent. Seizure control should be achieved, if possible, by increasing the dosage of this agent rather than by adding a second one. If seizure control cannot be achieved with the first medication, a second, alternative agent should be considered. Monotherapy—usually with the first or second agent chosen—can control seizures in about 60% of patients with newly diagnosed epilepsy. In the remainder of cases, control is often difficult from the outset.[31]

The use of two or more AEDs in combination should be avoided whenever possible, but rational drug combinations may be useful when monotherapy fails. Drug selection should be guided by the patient's seizure type and epilepsy syndrome classification, in concert with the mechanisms of action and side effects of the individual agents.[31] Changes in dosage should be guided by the patient's clinical response rather than by drug levels, with inadequate seizure control indicating the need for raising the dosage and toxicity indicating the need to lower the dosage.

Monitoring of drug levels is usually not necessary for patients who tolerate their medication well and have adequate seizure control. Under some circumstances, however, monitoring drug levels may be useful to determine prescription compliance or explain changes in seizure control or drug toxicity.

Mechanisms of Action of AEDs

AEDs appear to control seizures by decreasing neuronal excitability or enhancing inhibition of neurotransmission. This is achieved by altering intrinsic membrane currents, such as sodium, potassium, and calcium conductance through ion channels, or by affecting the activity of various neurotransmitters, such as γ-aminobutyric acid (GABA), glutamate, or other neurotransmitters that may take part in seizure activity. Although several AEDs have common mechanisms, each seems to have distinct actions. Each AED has been reported to have several molecular and cellular actions, but probably only some of these individual actions are responsible for the anticonvulsant and antiepileptic effects.

The intrinsic membrane currents affected by AEDs are primarily those involving the voltage-gated sodium channel and the calcium channel. The drugs that act at sodium channels in therapeutic concentrations—phenytoin, carbamazepine, primidone, valproate, and lamotrigine—inhibit high-frequency neuronal discharge.[32] Because they block the sodium channel gradually and in proportion to the rate of firing (i.e., they are both use dependent and voltage dependent), they have little effect on normal neuronal activity. However, during high-frequency firing, which typically occurs at seizure onset, they delay reactivation of the sodium channel and produce an increasing inhibitory effect on the action potential until firing is completely blocked.

Some AEDs act at both the sodium and the calcium channel, such as phenytoin, carbamazepine, valproate, lamotrigine, and zonisamide.[32] Other AEDs that act at calcium channels include ethosuximide and phenobarbital. Ethosuximide selectively blocks the T-type calcium current, which is not greatly affected by most of the other AEDs (except zonisamide). This current is thought to act as a pacemaker in thalamic neurons and may be important in absence seizures. Valproate and lamotrigine are also effective against absence seizures; however, it remains to be determined whether their efficacy results from action on the T-type calcium current or from some yet undefined mechanism. The blockade of other calcium currents, such as the L and N types, may be a less specific effect that affects neuronal excitability and propagation of seizure activity.[33]

Drugs that alter synaptic function act primarily by enhancing GABA-mediated neuronal inhibition, the brain's main inhibitory mechanism; these drugs include phenobarbital and benzodiazepines. Each of these drugs acts through a different mechanism to augment GABA influences in the CNS.[34] Benzodiazepines increase the frequency of GABA-mediated $GABA_A$ receptor channel openings, and barbiturates increase the duration of channel openings. A different strategy to increase extracellular GABA uses inhibition of transporters. Tiagabine delays the reuptake of GABA from the synaptic cleft, effectively enhancing the GABA effect after synaptic release. Vigabatrin, which is not available in the United States, increases GABA concentration by irreversibly binding to GABA transaminase, the enzyme that metabolizes GABA.[35]

The mechanisms of anticonvulsant and antiepileptic action of the major drugs used as therapy for seizure disorders are not completely understood. Some AEDs may affect ionic conductances other than those discussed, and some may directly or indirectly affect neurotransmitter processes other than the GABA system; these actions may be important for their clinical effects. For example, felbamate and topiramate may block glutamate receptors to interfere with excitation.[35] Additional research is needed to completely define the molecular and cellular mechanisms and the sites of action of AEDs.

Table 3 Antiepileptic Drugs

Agent (Trade Name)	*Formulations*	*Dosage*	*Principal Sites of Action*	*Targeted Seizure Type*
Carbamazepine (Tegretol, Tegretol-XR, Carbatrol)	Tablets: 200 mg Chewable tablets: 100 mg Suspension: 100 mg/5 ml Extended-release tablets: 100, 200, and 400 mg Extended-release capsules: 200 and 300 mg	10–40 mg/kg/day in three or four divided doses (regular tablets or suspension) or two divided doses (extended-release forms)	Sodium channels	Focal and secondary GTC
Clonazepam (Klonopin)	Tablets: 0.5, 1, and 2 mg	0.01–0.30 mg/kg/day in two or three divided doses or h.s.	GABA receptors	All generalized focal and secondary GTC
Diazepam for rectal administration (Diastat)	Rectal syringe: pediatric: 2.5, 5, and 10 mg; adult, 10, 15, and 20 mg	2–5 yr: 0.5 mg/kg 6–11 yr: 0.3 mg/kg > 12 yr: 0.2 mg/kg	GABA receptors	Acute repetitive seizures
Ethosuximide (Zarontin)	Capsules: 250 mg Suspension: 250 mg/5 ml	15–40 mg/kg/day in two or three divided doses	T-type calcium channels	Absence
Felbamate (Felbatol)	Tablets: 400 and 600 mg Suspension: 600 mg/5 ml	15–45 mg/kg/day (or 1,200–3,600 mg/day) in two or three divided doses	NMDA receptors Sodium channels GABA receptors	Focal and secondary GTC Lennox-Gastaut syndrome
Fosphenytoin (Cerebyx)	Injection: 50 mg/ml (phenytoin-equivalent)	10–20 mg/kg I.V. loading dose	Sodium channels	Status epilepticus
Gabapentin (Neurontin)	Tablets: 600 and 800 mg Capsules: 100, 300, and 400 mg Suspension: 250 mg/5 ml	10–60 mg/kg/day	Increases GABA release	Focal and secondary GTC
Lamotrigine (Lamictal)	Tablets: 25, 100, 150, and 200 mg Chewable tablets: 2, 5, and 25 mg	200–500 mg/day in two divided doses Slow titration required, especially if taken with valproic acid or valproate.	Sodium channels (modulating EAA release) Calcium channels	Focal and secondary GTC Absence Tonic/atonic GTC Myoclonic
Levetiracetam (Keppra)	Tablets: 250, 500, and 750 mg	1,000–3,000 mg/day in two divided doses	Potassium channels Calcium channels	Focal and secondary GTC
Lorazepam (Ativan)	Injection: 2 or 4 mg/ml	0.05–0.1 mg/kg I.V. (total dose)	GABA receptors	Status epilepticus
Oxcarbazepine (Trileptal)	Tablets: 150, 300, and 600 mg Suspension: 300 mg/5 ml	900–2,400 mg/day in two divided doses	Sodium channels	Focal and secondary GTC

AED Selection

The principal AEDs used to treat patients with epilepsy in the United States are carbamazepine, ethosuximide, gabapentin, lamotrigine, levetiracetam, oxcarbazepine, phenobarbital, phenytoin, primidone, topiramate, tiagabine, valproate, and zonisamide [*see Table 3*]. Some benzodiazepines, including clonazepam, diazepam, and lorazepam, are also used to treat seizures. With the exception of clonazepam, the benzodiazepines are used for short-term treatment of acute seizures or status epilepticus and are usually administered parenterally. Clonazepam can be used to treat epilepsy but is not recommended, because most patients develop tolerance to its antiepileptic effect. Felbamate should be reserved for use in selected patients with uncontrolled seizures. Because of the high incidence of serious adverse effects associated with felbamate, treatment with this drug should be managed by experienced clinicians in established epilepsy centers.[36] Several additional drugs, including methsuximide and acetazolamide, are used occasionally; none of these drugs will be considered in this subsection.

Partial seizures Most epileptologists agree that the drugs of choice for partial seizures are carbamazepine and phenytoin, which are highly effective and have acceptable adverse effects.[37] Valproate also has been demonstrated to be effective in the treatment of partial seizures.[38,39] Both phenobarbital and primidone are probably as effective as carbamazepine or phenytoin, but the two barbiturates are associated with a much higher incidence of adverse effects, particularly sedation and impaired cognition.[37] Gabapentin, lamotrigine, levetiracetam, oxcarbazepine, tiagabine, topiramate, and zonisamide are effective against partial seizures; each is usually used in combination with one of the older agents. However, of the newer AEDs, oxcarbazepine is approved as monotherapy, lamotrigine is approved as monotherapy after initial adjunctive use, and topiramate has a monotherapy application pending. Partial seizures that secondarily generalize respond to these same AEDs.

Generalized convulsive seizures Valproate, lamotrigine, and topiramate are the AEDs of choice for patients with primary generalized convulsions, because some of these patients may also have other generalized seizures that can be controlled with one of these agents. Carbamazepine and phenytoin were used historically to treat generalized convulsive seizures. The barbiturates are not favored in the treatment of generalized convulsive seizures, primarily because of their sedative effects. Zonisamide and levetiracetam may benefit some patients with generalized convulsive seizures.

Table 3 Antiepileptic Drugs (*continued*)

Agent (Trade Name)	*Formulations*	*Dosage*	*Principal Sites of Action*	*Targeted Seizure Type*
Phenobarbital (Solfoton, Luminal)	Tablets: 8, 15, 16, 30, 32, 60, 65, and 100 mg Injection: 30, 60, 65, and 130 mg/ml Solution: 20 mg/5 ml	2–5 mg/kg/day q.d. or in two divided doses	Chloride channels	Focal Primary and secondary GTC
Phenytoin (Dilantin, Phenytek)	Chewable tablets: 50 mg Extended-release capsules: 30, 100, and 300 mg Suspension: 125 mg/5 ml Injection: 50 mg/ml	Maintenance: 3–7 mg/kg/day in three divided doses; single dose or two divided doses for extended-release forms	Sodium channels Calcium channels NMDA receptors	Focal Primary and secondary GTC
Primidone (Mysoline)	Tablets: 50 and 250 mg Suspension: 250 mg/5 ml	500–1,500 mg/day in two or three divided doses	Chloride channels GABA uptake	Focal Primary and secondary GTC
Tiagabine (Gabitril)	Tablets: 2, 4, 12, 16, and 20 mg	32–56 mg/day in two to four divided doses	Sodium channels Potassium channels Glutamate receptors Carbonic anhydrase inhibition	Partial Focal and secondary GTC Tonic/atonic GTC Myoclonic
Topiramate (Topamax) Valproic acid/valproate sodium (Depakene/Depacon)	Tablets: 25, 100, and 200 mg Capsules: 15 and 25 mg Capsules: 125, 250, and 500 mg Extended-release capsules: 250 and 500 mg Syrup: 250 mg/ml Injection: 100 mg/ml	500–3,000 mg/day in two to four divided doses	Sodium channels Calcium channels	Focal and secondary GTC Absence Tonic/atonic GTC Myoclonic
Vigabatrin (Sabril)†	Tablets: 500 mg	2–4 g/day	GABA transaminase	Focal and secondary GTC Infantile spasms Lennox-Gastaut syndrome
Zonisamide (Zonegran)	Capsules: 100 mg	200–600 mg/day in two divided doses	Sodium channels Calcium channels Carbonic anhydrase inhibition	Focal Primary and secondary GTC Poor or none

*Rarely, lamotrigine may exacerbate myoclonic seizures
†Not available in the United States
EAA—excitatory amino acids GABA—γ-aminobutyric acid GTC—generalized tonic-clonic NMDA—*N*-methyl-D-aspartate

Generalized nonconvulsive seizures Generalized nonconvulsive seizures, particularly absence seizures, can be treated with ethosuximide, lamotrigine, or valproate. For patients who have only absence seizures, ethosuximide is satisfactory. However, for patients who have absence seizures in conjunction with other types of seizures, such as generalized convulsions or myoclonic seizures, valproate is the drug of choice; lamotrigine may also be effective. Some patients may respond to therapy with topiramate or zonisamide.

AED Pharmacokinetics

All AEDs are given orally once daily or more frequently [*see Table 3*]. Absorption of most AEDs usually occurs slowly over a period of hours and may be incomplete, especially in the case of gabapentin. Protein binding varies considerably among the drugs, ranging from 0% for gabapentin to 90% or more for phenytoin. Except for gabapentin, levetiracetam, and vigabatrin, all AEDs are metabolized in the liver before renal excretion. Additionally, zonisamide and topiramate have significant renal excretion.

AED clearance and half-life vary from hours to days. The half-lives of carbamazepine, valproate, primidone, and gabapentin are relatively short, ranging from 4 to 8 hours, which necessitates administering these agents at least three times daily (although extended-release formulations of valproate and carbamazepine allow twice-daily dosing). Phenytoin, lamotrigine, phenobarbital, and zonisamide have half-lives of a day or longer and may be administered once or twice daily. Phenytoin, carbamazepine, and the barbiturates cause enzyme induction, and long-term treatment with these drugs can affect their own rates of metabolism. Perhaps as important, these agents may affect metabolism of other medications the patient is taking. All AEDs that are metabolized in the liver can compete with other drugs that also undergo hepatic metabolism and so can delay clearance of those drugs. Because of their ability to induce or block drug metabolism, all of the AEDs except gabapentin, levetiracetam, and vigabatrin may be associated with pharmacokinetic drug interactions in which drug levels and effects can be markedly altered by the concomitant administration of two or more drugs. This kind of interaction should be anticipated in all patients. It can be detected from clinical symptoms and levels of AED in the blood, especially free drug levels, and can be corrected by adjustment of dosages. Pharmacodynamic drug interaction may also occur. In this situation, the combination of two or more drugs with similar or antagonistic mechanisms causes their clinical effects to be enhanced or diminished, and these changes may necessitate adjustment of dosages. Pharmacodynamic drug interactions can be anticipated when drug mechanisms of action are known.

Adverse Effects of AEDs

All AEDs can produce adverse effects, which are numerous

Table 4 Idiosyncratic Reactions to Antiepileptic Drugs[83,84]

Agent	*Agranulocytosis*	*Allergic Dermatitis/ Rash*	*Aplastic Anemia*	*Hepatic Failure*	*Pancreatitis*	*Serum Sickness Reaction*	*Stevens-Johnson Syndrome*	*Teratogenic Effects (FDA Pregnancy Category*)*
Carbamazepine	+	+	+	+	+	+	+	Neural tube defects, facial anomalies, microcephaly (D)
Clonazepam	+	–	–	–	–	–	–	(D)
Ethosuximide	+	+	+	–	–	+	+	(C)
Felbamate	+	+	+	+	–	–	–	(C)
Gabapentin	–	+	–	–	–	–	–	(C)
Lamotrigine	–	+	–	–	–	–	+	(C)
Levetiracetam	–	+	–	–	–	–	–	(C)
Oxcarbazepine	–	+	–	–	–	–	–	(C)
Phenobarbital	+	+	–	+	–	+	+	Facial anomalies (D)
Phenytoin	+	+	+	+	+	+	+	Facial anomalies, low IQ, IUGR (D)
Primidone	+	+	–	+	–	+	+	Facial anomalies, low IQ (D)
Tiagabine	–	+	–	–	–	–	–	(C)
Topiramate	–	+	–	–	–	–	–	(C)
Valproate	+	+	+	+	+	+	+	Neural tube defects, facial anomalies (D)
Zonisamide	–	+	–	–	–	–	+	(C)

*FDA Pregnancy Category C: Animal reproduction studies have shown an adverse effect on the fetus, and there are no adequate and well-controlled studies in humans; but potential benefits may warrant use of the drug in pregnant women despite potential risks.
FDA Pregnancy Category D: There is positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience or studies in humans, but potential benefits may warrant use of the drug in pregnant women despite potential risks.
FDA—Food and Drug Administration IQ—intelligence quotient IUGR—intrauterine growth retardation

and vary considerably from patient to patient. Most of the adverse effects caused by AEDs can be tolerated. However, rare idiosyncratic effects may be life threatening [*see Table 4*]. Good medical practice demands that physicians obtain screening studies before starting a patient on AED treatment. However, prospective studies of the effect of routine blood and urine assessment in patients receiving long-term AEDs showed no value in asymptomatic patients.[40] On the other hand, clinical monitoring—by reviewing and reporting symptoms or by physical examination—is cost-effective and practical. Routine laboratory monitoring, as currently practiced, is not effective in anticipating life-threatening effects associated with the major AEDs, including carbamazepine, phenytoin, phenobarbital, lamotrigine, and valproate. Continuing communication with the patient and family is the best strategy for monitoring for adverse events. They must be aware of potential complications and the symptoms that might herald an adverse event. An approach that combines clinical expertise with educating patients and caregivers about the risks and benefits of treatment and enlisting their active participation is becoming the standard of care.

Some of the undesirable side effects of AEDs, such as neurotoxicity, are clearly dose related. Other adverse effects, such as some forms of hepatotoxicity or aplastic anemia, seem to be idiosyncratic. Still others have other mechanisms. An efficient way to understand the adverse effects of AEDs is to define them according to the affected organ system.

Effects on the central nervous system All AEDs can cause depression of cortical function, with symptoms of sedation and lethargy. Dose-dependent neurologic toxic changes are the most common adverse effects of AEDs and limit the amount of drug that may be used. CNS symptoms are used to titrate to an acceptable dose; this is independent of AED blood levels, because the therapeutic range is statistically derived and may not apply to an individual patient. For example, carbamazepine causes diplopia or sedation; either one serves as an indication that the patient has reached the maximum tolerated dose. If AED blood levels reach the upper limit of the therapeutic range but seizures continue to be uncontrolled and the patient remains free of side effects, the dose is increased until neurotoxicity develops. This defines the maximum tolerable dose for that patient.

Although dose-related effects of phenytoin on the cerebellum are classic indicators of toxicity, the typical findings are often blunted or inapparent because of individual patient responses.[40] The expected progression of symptoms from nystagmus to ataxia and confusion or feelings of intoxication does not always occur. Residual ataxia after prolonged phenytoin intoxication can probably occur, especially in vulnerable patients.[41]

Persistent or fluctuating ataxia appears to be related to long duration of epilepsy, use of multiple drugs, and persistently high serum drug levels. Changes in intellectual function and alterations in personality or even psychiatric effects have been re-

ported with several AEDs. Altered IQ in children treated with phenytoin and primidone appears to be related to higher blood levels of the drugs. Assessment of cognitive function in patients treated in random fashion failed to reveal any differences among the major AEDs, although phenobarbital appears to have a greater effect than other medications.[42] Drugs administered to children may have unpredictable or paradoxical effects; for example, barbiturates are well known to cause hyperactivity in children. Depression and psychosis may be caused by AEDs. Most effects on behavior appear to be dose related. Psychosis after seizure control has been reported, as have other behavioral problems; in some cases, this appears to be related to normalization of the EEG and is termed forced normalization.[43]

Phenytoin may cause dose-related abnormal movements; carbamazepine seldom causes involuntary movements. Valproate may cause tremor.[43]

Mild sensory neuropathy has been reported in 8% to 15% of patients treated with older AEDs[44] and seems to require long-term exposure. Prolonged exposure to high plasma levels causes loss of large myelinated fibers and clustered nonrandom distribution of segmental demyelination that suggests a pattern of axonal neuropathy and secondary demyelination. Use of multiple drugs also is associated with chronic neuropathy. Altered peripheral nerve function is thought to be an effect of AEDs on folate metabolism, with resultant mild alteration of peripheral nerve function. Usually, the clinical effect is not severe, but loss of deep tendon responses or vibratory sensibility at the ankles may occur.

Aside from these dose-dependent effects, AEDs may induce a reversible encephalopathy and even progressive mental deterioration or delirium. On rare occasions, the initial exposure of a patient to valproate causes coma; the mechanism of this effect may be related to mitochondrial metabolism.[45]

Gastrointestinal and hepatic effects Valproate produces hepatic failure in approximately one in 10,000 patients. The highest rate of hepatic failure is observed in children younger than 2 years who are treated with multiple drugs. Older patients on monotherapy have a much lower rate: one in 45,000.[46] The mechanism of valproate-induced hepatotoxicity is unknown but may involve toxic metabolites of the drug. Hepatic damage associated with phenytoin, carbamazepine, or phenobarbital therapy is infrequent. The mechanism that causes the hepatic failure is thought to be an idiosyncratic hypersensitivity reaction.

Pancreatitis has been observed during valproate therapy, and fatalities have occurred in both children and adults.[47] Reexposure of affected patients has caused recurrent pancreatitis, but the mechanism of this adverse effect has not been established.[47]

Hematologic effects Adverse hematologic reactions associated with idiosyncratic responses to AEDs range from a mild decrease in the number of blood cells to aplastic anemia. Fortunately, serious problems are infrequent. Phenytoin may alter lymphocyte function, given that 21% to 25% of patients on long-term therapy have decreased circulating levels of IgA along with depressed lymphocyte phytohemagglutinin transformation.[48] Phenytoin hypersensitivity may cause generalized lymphadenopathy and in rare instances may be associated with lymphoma.[49] Carbamazepine treatment commonly causes hematologic changes. Dose-related leukopenia is reported to occur in 12% of treated patients.[40] In my experience, this estimate is low. The mechanisms leading to this effect on granulocytes is unknown, but the effect appears to be dose related and does not herald more serious reactions. There appears to be little need for concern until the total white blood cell count is less than 2,500/mm^3 or the total granulocyte count is less than 750/mm^3.

Aplastic anemia occurs infrequently with AEDs, and no single drug, with the exception of felbamate, seems to be more likely than another to produce this serious complication. Aplastic anemia associated with carbamazepine treatment seems to be related to dosing patterns and advanced patient age.[40] Of 65 patients who died of aplastic anemia associated with carbamazepine, only four were children.[40] Valproate causes a dose-related decrease in the level of circulating platelets that is only occasionally symptomatic.[50] Occurrence of purpura or petechial hemorrhage requires discontinuance of the drug. The mechanism is a change in platelet adhesiveness, with acceleration in the second stage of platelet consumption causing accelerated loss of circulating platelets. An IgM platelet autoantibody is occasionally detected.[51] Mild macrocytosis and decreased red blood cell folate levels occur with antiepileptic drug therapy, particularly phenytoin. Megaloblastic anemia is an occasional effect of antiepileptic drug treatment that appears to be related to alterations in metabolism of folate. Folate supplementation is required in some cases.

Dermatologic reactions Skin reactions to AEDs are not uncommon, but life-threatening dermatologic responses are rare.[52] Drug-induced eruptions range from mild erythema, with or without pruritus, to serious exfoliative reactions or development of bullae. Dose-related exanthema is the most common skin reaction. This form of rash may respond to dose reduction. The rash may be pruritic, maculopapular, or morbilliform, and about 50% of patients may have prodromal symptoms that include malaise and fever. Drug eruptions that begin as pruritic effects with morbilliform or scarlatiniform rashes may progress to severe exfoliative dermatitis. In such cases, the offending drug must be discontinued promptly and another AED substituted. The mechanism of skin reaction may be determined by pharmacogenetic factors related to the ability to metabolize arene compounds formed during drug metabolism.[52] The exfoliative dermatitides produced by AEDs include erythema multiforme, Stevens-Johnson syndrome, and Lyell syndrome.[52] Exfoliative dermatitis commonly occurs within 1 to 4 weeks after initiation of treatment and may be fatal. The drug responsible for the reaction must be discontinued and treatment with systemic steroids instituted. Erythema multiforme as a drug reaction is rapid in onset, with erythematous lesions that range from a macular pattern with varied shapes to the development of vesicles or bullae. If mucosal lesions are present, the reaction has evolved to the Stevens-Johnson syndrome. In either case, dermatologic consultation is imperative. Acne is a common minor skin problem associated with phenytoin use. Hypertrichosis is associated with both phenytoin and carbamazepine. Alopecia occurs during the first few weeks of treatment with valproate. Administration of multivitamins containing zinc appears to prevent this change in hair-shaft strength.

Connective tissue disorders AEDs may cause lupus erythematosus, scleroderma, Sjögren syndrome, and eosinophilic fasciitis. Lupus occurs in both a discoid and a systemic pattern. Criteria for diagnosis include malar erythema with discoid skin

changes, photosensitivity, oral ulceration, nonerosive arthritis, serositis, and nephropathy. Hematologic and immunologic changes are revealed by positive lupus erythematosus cell preparations, anti-DNA (double stranded) antibodies, and antinuclear antibodies.[53] Drugs may precipitate lupus, exacerbate existing lupus, or be associated with an isolated drug-related form. Drug-related lupus usually abates after discontinuance of the inciting drug. Further, the pattern of organ involvement usually spares skin, kidneys, and the CNS. The immunologic pattern usually does not involve the induction of antibodies to double-stranded DNA. Long-term phenytoin use has caused a syndrome of ecchymoses, gingival bleeding, and lupus anticoagulant antibodies associated with prothrombin deficiency. Carbamazepine and ethosuximide have also been associated with the development of drug-induced lupus.

Metabolic and endocrine effects Hyponatremia is a dose-related effect of carbamazepine that seems to occur only in adults. Oxcarbazepine is also associated with dose-related hyponatremia, primarily in adults; this occurs more commonly than with carbamazepine.[54] Central effects on antidiuretic hormone and peripheral or renal effects have been considered as mechanisms, but the exact pathogenesis remains unknown. With either drug, the hyponatremia seldom causes a clinical problem. Effects of phenytoin on pituitary-adrenal function are related to the peripheral effects of induction of cytochrome P-450 hepatic enzymes, with resultant accentuation of metabolism of endogenous hormone. Accelerated metabolism of exogenously administered steroid hormone, such as birth control pills, may cause contraceptive failure. Hypothalamic function may be affected by phenytoin. Changes include altered release of antidiuretic hormone, block of thyroid-stimulating hormone effect by thyrotropin-releasing hormone, and, in women, augmentation of secretion of follicle-stimulating hormone and luteinizing hormone.[55,56] Phenytoin can alter the results of thyroid function studies: it lowers total triiodothyronine (T_3) and thyroxine (T_4), so patients may have a decreased free T_4 level, with a normal thyroid-stimulating hormone (TSH) level; however, phenytoin also increases free T_4 and T_3 by induction of displacement from thyroxine-binding globulin, which may normalize free T_4 levels. Most patients remain clinically euthyroid.[55]

Weight gain may occur with valproate or gabapentin; weight loss may occur with topiramate or zonisamide. The exact mechanisms are unknown. Other AEDs are considered weight neutral.

Discontinuing AED Therapy

Most patients treated with AEDs require treatment for several years. When a patient has remained seizure free after several years of AED therapy and the underlying epilepsy syndrome is not one that is known to require continuous treatment (e.g., juvenile myoclonic epilepsy), the question arises as to whether to discontinue treatment. A meta-analysis[57] concluded that the likelihood of successful AED withdrawal was highest in patients meeting the following criteria: (1) they have been seizure free from 2 to 5 years on AED treatment, (2) they have a single type of partial seizure (simple partial, complex partial, or secondary generalized tonic-clonic seizure) or single type of primary generalized tonic-clonic seizure, (3) their neurologic examination results and IQ are normal, and (4) their EEG has normalized with treatment.

VAGUS NERVE STIMULATION

Electrical stimulation of the vagus nerve has both acute and chronic antiepileptic effects. The acute effects result from polysynaptic transmission and activation of key inhibitory pathways; the chronic effects may be the result of persistent changes in neurotransmitters or changes in cortical and subcortical synaptic activity.

Vagus nerve stimulation (VNS) is now the second most common treatment for epilepsy in the United States, after AEDs. VNS devices were approved by the Food and Drug Administration in 1997, and over 16,000 of these devices have now been implanted in patients worldwide.[58]

Implantation of a VNS device usually takes less than 1 hour and can be done as an outpatient procedure. The generator, which measures 6.9 by 52 mm and is made of titanium, is placed subcutaneously in the left upper chest. The lead wires are placed on the left cervical vagus nerve and tunneled to the generator. VNS devices have programmable settings; the intensity and duration of the pulses (e.g., 1.5 mA for 30 sec) are tailored to the individual patient. Once programmed, the device operates on its own. In addition, patients can activate the device themselves, through the use of a handheld magnet, which triggers a switch that initiates a separate program in the device. This ability to abort seizures gives patients an active role in the treatment of their epilepsy and is an important psychosocial benefit of VNS therapy.

Criteria for VNS placement include partial onset seizures that persist despite adequate trials of two or three AEDs (preferably, AEDs with differing mechanisms of action). Nonepileptic events must be excluded. The patient should not be a good candidate for focal resective surgery (see below); in patients who are good candidates, this surgery is preferred to VNS because it is more likely to render the patient seizure free. In the United States, the use of VNS is limited to patients older than 12 years; in the European Union, there is no age limit with VNS therapy, although the size of the device may limit its use in children younger than 3 or 4 years.[59]

SURGICAL TREATMENT

Surgical treatment is an option for patients who have failed to respond to conventional AED treatment or who have had intolerable adverse drug effects, whose seizures have a focal origin, and whose seizures originate in tissue that can be removed without causing disability.[60] Epilepsy may be associated with a structural lesion in the region of seizure onset. The presence of a seizure focus in the hemisphere dominant for language and the occurrence of complex partial seizures arising in extratemporal tissue necessitate special assessment. Detection of bilateral epileptiform discharges, development of secondarily generalized seizures, and occasional onset of a seizure from contralateral tissue tend to add to the complexity of preoperative assessment.[61]

Current standards for surgery require locating the seizure focus by several means. Although ictal and interictal scalp EEG and intraoperative electrocorticography have been used, these methods may fail to provide adequate lateralizing or localizing information. Therefore, most epilepsy centers require combined video and scalp EEG recording of at least three of a patient's typical seizures. Additional localizing information is obtained from MRI, neuropsychology, and isotope studies of blood flow or metabolism. Location of memory function and speech lateralization are also important. Insertion of intracranial electrodes is required should the scalp EEG fail to provide adequate lateral-

izing or localizing information. However, because of the small but significant morbidity associated with invasive electrode studies, functional imaging [*see* Brain Imaging, *above*] is being explored as an alternative.[21]

The most commonly performed surgical procedure for epilepsy is temporal lobectomy. After this procedure, approximately 70% of patients remain seizure free, and an additional 20% show marked improvement. Histopathologic assessment shows that in at least 60% of patients, sclerotic changes occur in the resected hippocampus. Long-term follow-up studies have shown a fixed rate of recurrent seizures after temporal lobectomy, so these patients are no longer taken off AEDs postoperatively.[62] The goal of this procedure is not to eliminate AED use; it is to eliminate seizures. Patients may be able to reduce the number of AEDs or lower the dose, but even those who remain on the same AED regimen generally enjoy significant improvement in their quality of life.

Complications associated with resective surgery remain low and generally acceptable. A superior quadrantic visual field defect may occur in patients undergoing classic temporal lobectomy. Permanent hemiparesis is reported in up to 2.4% of patients; mortality varies from 0% to 1.7%.[60]

More advanced procedures (i.e., resection of extratemporal foci, sectioning of the corpus callosum, and functional hemispherectomy) are performed in special epilepsy centers. Corpus callosotomy is a palliative procedure for patients who sustain injuries secondary to falls during seizures.[63] It has largely been supplanted by VSN device implantation, however.

THE KETOGENIC DIET

A diet high in fat and low in carbohydrates can prevent seizures by maintaining the patient in ketosis. The use of ketogenic diets for epilepsy was pioneered in the 1920s. Such diets remained a standard aspect of epilepsy treatment until the last decades of the 20th century, when their popularity was eclipsed by that of AEDs. However, the ketogenic diet has reemerged, and clinical studies have confirmed its efficacy in almost all seizure types.[59]

The classic ketogenic diet has a 4-to-1 ratio of calories derived from fats to calories derived from proteins and carbohydrates. The diet is individualized, to the extent possible, according to the patient's food preferences and eating habits. Initiation of the diet is done with the patient in hospital. Strict adherence to the diet is necessary.

Ketogenic diets are used principally in children. Although the exact antiepileptic mechanism of ketogenic diets remains unknown, it is known that during the diet, the brain utilizes ketone bodies for fuel. With maturity, the brain's ability to extract ketones from the blood decreases as much as fivefold, and this may make the ketogenic diet slightly less effective in adults. Most often, however, adults do not use the diet because they find it impractical.

Treatment of Status Epilepticus

Status epilepticus is a danger to the patient and a treatment challenge to the physician. The cardinal feature of this serious epileptic state is continuous seizures or two or more seizures occurring in sequence without recovery of consciousness between them. Status epilepticus has varied forms of clinical presentation, including repeated generalized convulsive seizures, with coma between seizures; nonconvulsive seizures, causing a change in cognitive function; and sequential focal seizures, including focal motor seizures or focal sensory complaints.

Generalized convulsive status epilepticus is the most common and most challenging form of status epilepticus. The patient has convulsive movements and is unconscious. The motor manifestations of convulsive status epilepticus may be symmetrical, with tonic and then clonic activity. If the generalized seizures arise from a focus, lateralized movements occur at the onset or during seizures.

In practical terms, any patients who are in seizure when they reach the emergency department or who are observed having a seizure for 10 minutes must be treated on the assumption that they are in status epilepticus. Treatment is divided into acute management and drug therapy [*see Figure 4*].

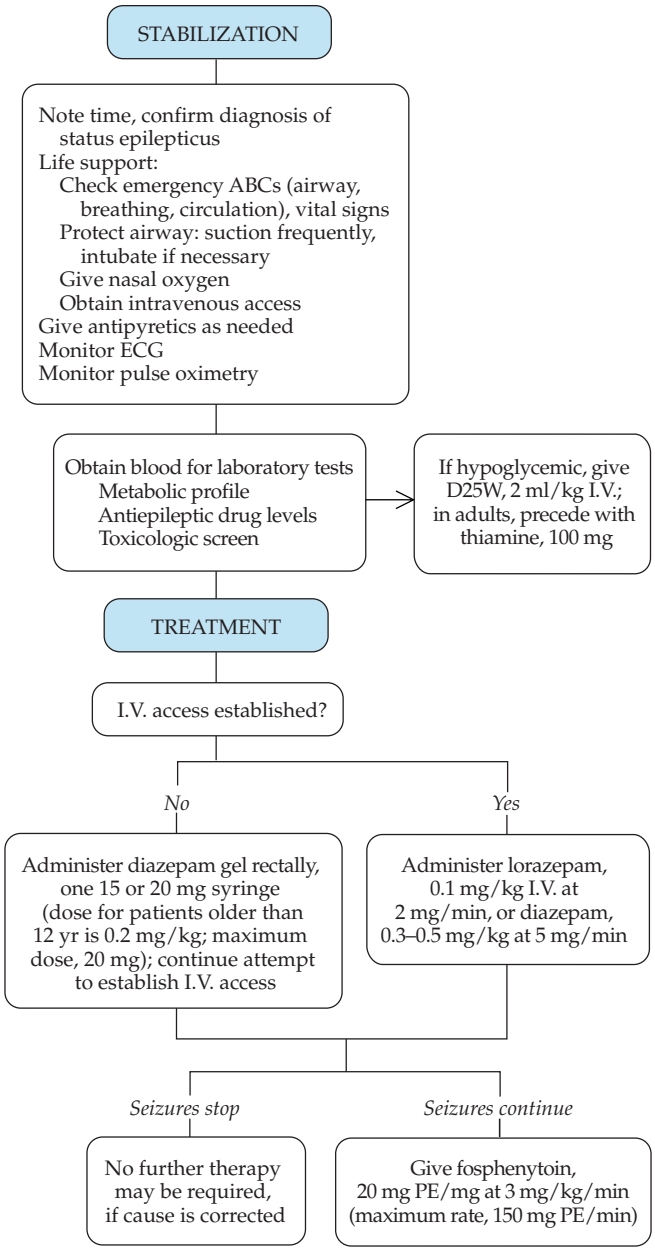


Figure 4 **Treatment of convulsive status epilepticus. (D25W—25% dextrose in water; ECG—electrocardiogram; PE—phenytoin equivalents)**

ACUTE MANAGEMENT

Treatment begins with life support, including clearing the airway and supporting ventilation, maintaining blood pressure, and establishing intravenous access. The airway must be protected and appropriate oxygenation ensured. Although intubation may be necessary, the decision to use it must be made in parallel with selection and administration of an anticonvulsant drug. Physiologic monitoring must include assessment by electrocardiography and measurements of blood pressure, levels of blood gases, levels of biochemical markers, and body temperature.

Hypoglycemia must be excluded or if present, corrected. Blood glucose levels must be measured promptly. If this is not possible, the patient should be treated empirically with intravenous glucose. Adults should first receive thiamine, 100 mg intravenously, to avoid precipitating Wernicke disease in thiamine-deficient alcoholic patients. The dose of glucose for adults is a bolus of 50 ml of 50% glucose; for children, the dose is 2 ml/kg of 25% glucose solution.

DRUG THERAPY

Clinical and electrical seizures must be terminated rapidly because total duration of convulsive status epilepticus correlates with response to treatment and with outcome.[64] Physicians treating patients with convulsive status epilepticus must be familiar with the drugs available, including methods of administration, dosages, and acute side effects. Best outcomes are associated with having a treatment plan, using effective drugs administered by the appropriate route and in adequate doses, and anticipating apnea. A protocol should be followed, and drugs should be administered intravenously (except for diazepam gel per rectum [see below]). Currently, intramuscular administration has no role in treatment. The clinical situation determines which drug should be administered. If a patient is having a seizure at the time of assessment, a benzodiazepine is needed. If the patient has a history of serial seizures but convulsive seizures have abated, administration of a long-acting AED is the best choice.[65] In the only randomized, controlled trial comparing initial intravenous treatments for status epilepticus (which was conducted at Veterans Affairs hospitals), lorazepam was found to be more effective than phenytoin and easier to use, although no more effective than phenobarbital alone or diazepam followed by phenytoin.[66]

Benzodiazepines

The benzodiazepines are effective and highly potent. Diazepam and lorazepam are commonly used. A rectal gel formulation of diazepam is currently available, packaged in predosed syringes. Because it can be administered to a patient with no intravenous access, it is suited for use in the home, in the field, and in the emergency department while intravenous access is being obtained. Diazepam is highly lipid soluble and penetrates the brain rapidly. However, redistribution into nonneural fatty tissue causes rapid decline in both brain and blood concentrations of the drug. If diazepam is used to terminate seizures, a long-acting AED such as fosphenytoin must be administered. Lorazepam has a longer duration of action than diazepam but is associated with a prolonged time to full recovery.[65] Both of these drugs cause depression of breathing and even apnea, so ventilation must be supported, and intubation may be necessary.

Fosphenytoin

Convulsive status epilepticus can be terminated by intravenous loading of fosphenytoin, a water-soluble prodrug form of phenytoin. Fosphenytoin is administered in phenytoin equivalents of up to 150 mg/min, or 3 mg/kg/min in children weighing less than 50 kg. It is rapidly converted by phosphatases into phenytoin, leading to blood levels equivalent to those achieved with phenytoin itself, but without the difficulties of administration and risk of tissue injury.

During fosphenytoin treatment, the elderly and patients with cardiac disease or with difficulty maintaining blood pressure require careful monitoring of cardiac rhythm and rate along with blood pressure. Those with hypotension require a slower rate of infusion. Electrocardiographic monitoring may show widening of the QT interval or even induction of arrhythmia. These changes signal the need to slow the rate of infusion even further.

Phenobarbital

Phenobarbital is highly effective and is known to many physicians. The adult dose is 10 to 20 mg/kg. Sedation may result, and apnea is a risk, particularly if the patient has taken benzodiazepines. Blood pressure monitoring is critical; hypotension responds to slowing the rate of administration.

TREATMENT OF REFRACTORY STATUS EPILEPTICUS

If a patient fails to regain consciousness or continues to have seizures after first-line therapy, neurologic consultation is required. Such a patient requires urgent EEG recording,[67] and anesthesia must be considered. An effective anesthetic drug is pentobarbital, although some authors suggest use of propofol or midazolam.[65,68] Midazolam (0.2 mg/kg administered by slow I.V. bolus injection followed by 0.75 to 10 μg/kg/min) and propofol (1 to 2 mg/kg with 2 to 10 mg/kg/hr) appear to be replacing pentobarbital as the drugs of choice.[65] The patient must be intubated and appropriate intensive care monitoring established. Patients are maintained in anesthetic coma for variable periods. Continued management requires gradual tapering of the maintenance dose at 4 hours, again at 8 hours, and on a regular schedule thereafter, with observation to determine whether seizures have abated and the EEG remains free of seizure discharges. Vasopressors may be needed during pentobarbital coma. Subtle convulsive status epilepticus occasionally develops. EEG assessment is required, particularly when motor changes are not obvious.[67]

AFTERCARE

After initial diagnostic studies are performed and the seizures controlled, the cause of status epilepticus must be sought. Medical and neurologic evaluations are important. When the patient is known to have epilepsy and the chart is available for review, further evaluation may be unnecessary. Brain-imaging studies are appropriate after seizures are under control. A history of head trauma, focal seizures, or signs of systemic illness should guide evaluation. If the patient is febrile, a lumbar puncture with examination of the CSF is needed, but only after mass lesions and ventricular obstruction are excluded by CT or MRI. An EEG is indicated in all patients with new-onset status epilepticus or those with treated status epilepticus who are not recovering their previous neurologic baseline. The clinical situation must determine decisions.

Prognosis of Epilepsy with Treatment

The likelihood of control or remission of epilepsy varies ac-

Internet Resources on Epilepsy

Web Sites for Patients

British Epilepsy Association
http://www.epilepsy.org.uk

Cyberonics (vagus nerve stimulation)
http://www.cyberonics.com

Epilepsy Canada (English and French)
http://www.epilepsy.ca/

Epilepsy Foundation of America
http://www.epilepsyfoundation.org

Epilepsy Information Service
http://www.bgsm.edu/neuro/disease/epilinfo.shtml

The Epilepsy Research Foundation
http://www.erf.org.uk

University of Washington Regional Epilepsy Center
http://faculty.washington.edu/chudler/epi.html

Web Sites for Physicians

American Epilepsy Society
http://www.aesnet.org

University of Washington Regional Epilepsy Center—Antiepileptic Drugs
http://elliott.hmc.washington.edu/EpiInfo/antiepileptic.htm

cording to the particular epilepsy syndrome, as well as factors specific to the individual case. Control, when applied to epilepsy, means lack of seizures with use of medication. Remission means lack of seizures without medication.

Prognosis with medical control of generalized and partial seizures has improved with better methods for assessment, introduction of new drugs, and more rational use of older agents.[69] The simultaneous use of EEG and video monitoring has improved diagnostic accuracy. Of patients with new-onset seizures, 60% to 65% achieve good control with a single drug.[70] Some studies show a 5-year absence of seizures in at least 70% of patients who were followed for 20 years. Of these patients, 50% were in remission—that is, they were not taking any medication.[29]

Control is difficult to achieve in patients with epilepsy of longer duration, with partial seizures, with more seizures before the start of treatment, with seizures that are of known cause, and with epileptiform patterns on EEG.[69] Poor seizure control also is associated with various types of seizures with abnormalities on EEG, delay in the start of treatment for more than 1 year, and frequent seizures; poor control may result in impaired social adjustment.[69]

The prognosis for control is less favorable with complex partial seizures than with generalized seizures. Of patients with complex partial seizures, the outcome tends to be better in those who have normal mentation, short duration of illness, and low seizure frequency.[29,69] However, if complex partial seizures are complicated by generalized tonic-clonic seizures, complete control is difficult to achieve. Patients at high risk for development of intractable complex partial seizures are those with clusters of seizures, more than one seizure a day, aura at the onset of seizures, and psychiatric disease.[69] In drug trials using high-dose carbamazepine, phenytoin, and barbiturates to the point of clinical toxicity in patients with intractable seizures, high-dose therapy resulted in complete control in 22% of patients; 38% had either an increase or no change in seizure frequency, and approximately 30% were not affected by any drug.[71]

Epilepsy may affect life span. Review of coroners' cases reveals common factors in sudden death in epileptic patients. Most deaths occurred when patients were in bed, with 6% to 30% occurring during sleep. At autopsy, few patients had therapeutic blood levels of prescribed AEDs; 50% had no detectable levels. Seizures are not always immediately associated with death[72]; cardiac arrhythmia has been implicated.[72,73] The prevalence of sudden death is between one in 2,000 and one in 900 patients with epilepsy, with higher rates occurring in patients with refractory epilepsy and nocturnal convulsive seizures.

Special Issues

BRAIN TUMORS

Brain tumors heighten susceptibility to seizures; by the time of diagnosis, 20% to 40% of patients with brain tumors have experienced a seizure. Treatment with anticonvulsant medications is clearly indicated for brain tumor patients who have had a seizure. Prophylactic AEDs should not be used routinely in patients with newly diagnosed brain tumors, however, because these agents are not effective in preventing first seizures in this population, and AED side effects are more frequent and more severe.[74]

PSYCHOSOCIAL DISABILITY

Chronic medical conditions causing the need for continuous medical treatment affect self-image and self-esteem. Epilepsy has the added adversity of inducing a fear of loss of control and presents situational pressures such as difficulties in gaining employment, inability to obtain insurance, and driving restrictions. All of these problems exacerbate social isolation. Local resources, such as a chapter of the Epilepsy Foundation of America [*see Sidebar* Internet Resources on Epilepsy], provide an opportunity for patients to receive help with these important problems. Each patient must be instructed about driving in accordance with state law, and this warning must be recorded on the patient's chart.

ORAL CONTRACEPTIVES

Enzyme-inducing AEDs accelerate metabolism of oral contraceptives; these include carbamazepine, oxcarbazepine, phenobarbital, phenytoin, and topiramate (in doses over 200 mg daily).[75] Birth control pills may fail; the dose of hormone needs to be adjusted, and the patient must be informed of this important drug interaction. This discussion must be recorded on the patient's chart.

PREGNANCY

Women are instructed by gynecologists to avoid medication during pregnancy, yet women with epilepsy may require long-term AED treatment to control seizures. Physicians must review issues regarding teratogenesis with all women capable of childbearing, and this discussion must be recorded on the patient's chart.[76] Maternal epilepsy is associated with risk of fetal malformations, and treatment with AEDs increases the risk from between 2% and 3% to between 4% and 6%.[77] All of the commonly used AEDs cause major malformations, including ventricular septal defects and cleft lip and palate. Valproate and carbamazepine treatment are associated with spina bifida; the

Table 5 Use of Antiepileptic Drugs in the Elderly

Generation	*Agent*	*Starting Dose (mg)*	*Proposed Dose Titration (Total mg/day)*	*Titration Interval (Weeks)*	*ProposedMaintenanceDose (Total mg/day)*
First	Carbamazepine	50–100 b.i.d.	100–200	1–2	100–400 divided b.i.d.–t.i.d.
	Clonazepam	0.25–0.5 q.d.	0.25–0.5	1	0.5–1.0 divided b.i.d.–t.i.d.
	Ethosuximide	250 q.d.	250	2	500–1,000 divided b.i.d.–t.i.d.
	Phenobarbital	15–30 q.d.	15–30	2	30–120 divided q.h.s.–b.i.d.
	Phenytoin	100–200 q.d.	50	1–2	100–300 divided q.d.–b.i.d.
	Primidone	125 q.d.–b.i.d.	125	2	250–750 divided b.i.d.–t.i.d.
Second	Felbamate*	200 b.i.d.	200–400	2	800–2,400 divided t.i.d.–q.i.d.
	Gabapentin	100 q.d. –t.i.d.	300	1	900–3,600 divided t.i.d.–q.i.d.
	Lamotrigine†	25–50 q.d.	25–50	2	100–400 divided b.i.d.
	Levetiracetam	250–500 b.i.d.	500–1,000	2	1,000–3,000 divided b.i.d.–t.i.d.
	Oxcarbazepine	150 b.i.d.	150–300	2	300–1,200 divided b.i.d.
	Tiagabine	2 q.d.–b.i.d.	2–4	1–2	24–48 divided b.i.d.
	Topiramate	25 q.d.	25–50	2	200–400 divided b.i.d.
	Valproate	125–250 b.i.d.	250–500	1–2	750–1,500 divided t.i.d.
	Zonisamide	100 q.o.d.–q.d.	100	2	200–600 divided b.i.d.

Avoid using in the elderly: felbamate may cause acute, sometimes fatal, fulminant hepatic failure or aplastic anemia.

†Lamotrigine may cause a dose-dependent rash. Dosing depends on whether patient is concurrently taking an inducer or inhibitor of hepatic cytochrome P-450 enzymes.

incidence is 1% to 2%, which is 10 to 20 times greater than that in the general population.[78,79] Minor morphologic malformations, however, are common. Midface abnormalities, digital hypoplasia, and nail-bed hypoplasia have been observed in 5% to 30% of infants of mothers treated with AEDs. Women requiring valproate and carbamazepine especially need preconception counseling.

Mechanisms of teratogenesis are unknown, but recommendations have been developed on the basis of population studies. Epileptic mothers should be treated with the lowest effective dose of a drug, and the frequency of doses should be designed to avoid high peak serum levels during each day. Use of a single drug is best, if possible. All women with epilepsy who are planning a pregnancy must be given supplemental folate before conception and daily throughout pregnancy. Vitamin K_1 must be given to mothers during the last 4 weeks of gestation to reduce the occurrence of coagulopathy in the infant. If the mother is inclined to consider therapeutic abortion should the fetus have a neural tube defect, careful assessment with ultrasonography and amniocentesis will be needed.[76,77]

ELDERLY PATIENTS

The incidence of epilepsy among the elderly is increasing as longevity increases. Given the steady growth in this population segment, the elderly are likely to become the most common source of new-onset seizures.[80] Whereas generalized epilepsies peak in the first 5 years of life, focal epilepsies predominate in patients 65 years of age and older.

Acute seizures in the elderly can result from conditions such as drug intoxication or withdrawal (including prescription drugs and ethanol) and electrolyte disorders. Causes of focal epilepsy include strokes (especially hemorrhagic stroke), cerebral neoplasm, degenerative brain disease, and closed head trauma. Only a minority of patients who experience early seizures after strokes (within the first 2 weeks, typically within 24 to 48 hours) will develop epilepsy, but the majority of those with late seizures (beyond 2 weeks afterward) develop epilepsy. Epilepsy is sixfold to eightfold more common in patients with dementia and may occur within 3 to 4 years after the onset of dementia.

Clinical features of specific epilepsy syndromes are the same in the elderly, but postictal symptoms tend to be more marked and much more prolonged; whereas postictal features typically resolve within 30 minutes in younger persons, in the elderly they may last as long as 4 to 8 days. Diagnostic testing in the elderly is complicated by the fact that in between seizures, epileptiform abnormalities on EEG become less common with advancing age. The differential diagnosis should include syncope (which may be preceded by auralike symptoms and accompanied by tonic or myoclonic jerks) and transient ischemic attacks.

The choice of drugs is similar to that for younger patients, but changes in body function associated with aging complicate treatment.[81] Decline in renal and hepatic function reduces drug protein binding and clearance. Altered body mass and structure change tissue-penetration kinetics. Consequently, introduction of AED therapy in elderly patients follows the so-called low and slow rule: treatment should start with lower doses and titrate slowly [*see Table 5*]. Because plasma levels of AEDs correlate less

well with efficacy and toxicity in the elderly, physicians should instead use clinical end points to guide dosing (e.g., seizure frequency to determine efficacy, and vigilance to detect symptoms of drug side effects). If blood levels of protein-bound AEDs are measured, measurement of unbound drug levels provides more useful information than total plasma levels, because serum protein levels decline with advancing age.[82] Finally, many elderly patients are receiving pharmacologic treatment for other concomitant illnesses, and the resulting polypharmacy can lead to adverse drug interactions.

The author has received grants for clinical research from, served as an advisor or consultant to, or been a member of the speakers' bureaus of Abbott Laboratories, Cyberonics, Inc., Elan Corporation, Novartis Pharmaceuticals Corp., Ortho-McNeil Pharmaceutical, Inc., Shire Richwood Inc., UCB Pharma, and Xcel Pharmaceuticals Inc.

References

1. Hauser WA: Seizure disorders: the changes with age. Epilepsia 33(suppl 4):S6, 1992

2. Kaneko S, Iwasa H, Okada M: Genetic identifiers of epilepsy. Epilepsia 43(suppl 9):16, 2002

3. Lerche H, Jurkat-Rott K, Lehmann-Horn F: Ion channels and epilepsy. Semin Med Genet 106:146, 2001

4. Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. Epilepsia 30:389, 1989

5. Benbadis SR: Epileptic seizures and syndromes. Neurol Clin 19:251, 2001

6. Theodore WH, Porter RJ, Penry JK: Complex partial seizures: clinical characteristics and differential diagnosis. Neurology 33:1115, 1983

7. Rocca WA, Sharbrough FW, Hauser WA, et al: Risk factors for complex partial seizures: a population-based case-control study. Ann Neurol 21:22, 1987

8. Keranen T, Sillanpaa M, Riekkinen PR: Distribution of seizure types in an epileptic population. Epilepsia 29:1, 1988

9. Berkovic SF, Scheffer IE: Febrile seizures: genetics and relationship to other epilepsy syndromes. Curr Opin Neurol 11:129, 1998

10. Singh R, Scheffer IE, Crossland K, et al: Generalized epilepsy with febrile seizures plus: a common childhood-onset genetic epilepsy syndrome. Ann Neurol 45:75, 1999

11. Dreifuss F: Pediatric epilepsy syndromes: an overview. Cleve Clin J Med 56 (suppl):S166, 1989

12. Penry JK, Porter RJ, Dreifuss FE: Simultaneous recording of absence seizures with videotape and electroencephalography: a study of 374 seizures in 48 patients. Brain 98:427, 1975

13. Delgado-Escueta AV, Enrile-Bascal F: Juvenile myoclonic epilepsy of Janz. Neurology 34:285, 1984

14. Sato S, Dreifuss FE, Penry JK, et al: Long-term follow-up of absence seizures. Neurology 33:1590, 1983

15. Wirrell EC, Camfield CS, Camfield PR, et al: Long-term prognosis of typical childhood absence epilepsy: remission or progression to juvenile myoclonic epilepsy. Neurology 47:912, 1996

16. Rasmussen TE, Olszewsik J, Lloyd-Smith D: Focal cortical seizures due to chronic localized encephalitides. Neurology 8:435, 1958

17. Cascino GD: Video-EEG monitoring in adults. Epilepsia 43(suppl 3):80, 2002

18. Radhakrishnan K, So EL, Silbert PL, et al: Predictors of outcome of anterior temporal lobectomy for intractable epilepsy. Neurology 51:465, 1998

19. Cascino GD, Jack CR Jr, Parisi JE, et al: Magnetic resonance imaging–based volume studies in temporal lobe epilepsy: pathological correlations. Ann Neurol 30:31, 1991

20. Cascino GD: Structural brain imaging. Epilepsy: A Comprehensive Textbook. Engel J Jr, Pedley TA, Eds. Lippincott-Raven, Philadelphia, 1997, p 937

21. Richardson MP: Functional imaging in epilepsy. Seizure 10:139, 2001

22. Kuzniecky R: Magnetic resonance and functional magnetic resonance imaging: tools for the study of human epilepsy. Curr Opin Neurol 10:88, 1997

23. Ishibashi H, Simos PG, Castillo EM, et al: Detection and significance of focal, interictal, slow-wave activity visualized by magnetoencephalography for localization of a primary epileptogenic region. J Neurosurg 96:724, 2002

24. Laxer KD: Clinical applications of magnetic resonance spectroscopy. Epilepsia 38(suppl):S13, 1997

25. Ryvlin P, Philippon B, Cinotti L, et al: Functional neuroimaging strategy in temporal lobe epilepsy: a comparative study of 18FDG-PET and 99mTc-HMPAO-SPECT. Ann Neurol 31:650, 1993

26. Breier JI, Mullani NA, Thomas AB, et al: Effects of duration of epilepsy on the uncoupling of metabolism and blood flow in complex partial seizures. Neurology 48:1047, 1997

27. Meierkord H, Will B, Fish D, et al: The clinical features and prognosis of pseudoseizures diagnosed using video-EEG telemetry. Neurology 41:1643, 1991

28. Shen W, Bowman ES, Markand ON: Presenting the diagnosis of pseudoseizure. Neurology 40:756, 1990

29. Hauser WA, Rich SS, Lee JR-J, et al: Rise of recurrent seizures after two unprovoked seizures. N Engl J Med 338:429, 1998

30. Hauser WA, Hesdorffer DC: Epilepsy: Frequency, Causes and Consequences. Demos, New York, 1990

31. Brodie MJ, Kwan P: Staged approach to epilepsy management. Neurology 58(suppl 5): S2, 2002

32. Moshé SL: Mechanisms of action of anticonvulsant agents. Neurology 55(suppl 1):S33, 2000

33. Stefani A, Spadoni F, Bernardi G: Voltage-activated calcium channels: targets of antiepileptic drug therapy? Epilepsia 38:959, 1997

34. Olsen RW, Avoli M: GABA and epileptogenesis. Epilepsia 38:399, 1997

35. White HS: Comparative anticonvulsant and mechanistic profile of the established and newer antiepileptic drugs. Epilepsia 40(suppl 5):S2, 1999

36. Pellock JM, Brodie MJ: Felbamate: 1997 update. Epilepsia 38:1261, 1997

37. Mattson RH, Cramer JA, Collins JF, et al: Comparison of carbamazepine, phenobarbital, phenytoin, and primidone in partial and secondarily generalized tonic-clonic seizures. N Engl J Med 313:145, 1985

38. Willmore LJ, Shu V, Wallin B, et al: Efficacy and safety of add-on divalproex sodium in the treatment of complex partial seizures. M88-194 Study Group. Neurology 46:49, 1996

39. A comparison of valproate with carbamazepine for the treatment of complex partial seizures and secondarily generalized tonic-clonic seizures in adults. Department of Veterans Affairs Epilepsy Cooperative Study, No. 264 Group. N Engl J Med 327:765, 1992

40. Pellock JM, Willmore LJ: A rational guide to routine blood monitoring in patients receiving antiepileptic drugs. Neurology 41:961, 1991

41. Plaa GI, Willmore LJ: General Principles: Toxicology. Antiepileptic Drugs, 4th ed. Levy RH, Mattson RH, Meldrum BS, Eds. Raven Press, New York, 1998, p 51

42. Meador KJ, Loring DW, Huh K, et al: Comparative cognitive effects of anticonvulsants. Neurology 40:391, 1990

43. Reynolds EH, Trimble MR: Adverse neuropsychiatric effects of anticonvulsant drugs. Drugs 29:570, 1985

44. Shorvon SD, Reynolds EH: Anticonvulsant peripheral neuropathy: a clinical and electrophysiological study of patients on single drug treatment with phenytoin, carbamazepine or barbiturates. J Neurol Neurosurg Psychiatry 45:620, 1982

45. Triggs WJ, Bohan TP, Lin S, et al: Valproate-induced coma with ketosis and carnitine insufficiency. Arch Neurol 47:1131, 1990

46. Bryant AE III, Dreifuss FE: Valproic acid hepatic fatalities: III. U.S. experience since 1986. Neurology 46:465, 1996

47. Pellock JM, Wilder SJ, Deaton R, et al: Acute pancreatitis coincident with valproate use: a critical review. Epilepsia 43:1431, 2002

48. Sorrell TC, Forbes IJ: Depression of immune competence by phenytoin and carbamazepine. Clin Exp Immunol 20:273, 1975

49. Jeng YM, Tien HF, Su IJ: Phenytoin-induced pseudolymphoma: reevaluation using modern molecular biologic techniques. Epilepsia 37:104, 1996

50. Harden CL: Therapeutic safety monitoring: what to look for and when to look for it. Epilepsia 41(suppl 8):S37, 2000

51. Sandler RM, Emberson C, Roberts GE, et al: IgM platelet autoantibody due to sodium valproate. Br Med J 2:1683, 1978

52. Roujeau J, Kelly JP, Naldi L, et al: Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. N Engl J Med 333:1600, 1995

53. Klinman DM, Steinberg AD: Systemic lupus erythematosus and overlap syndromes. Immunological Diseases, 4th ed. Samter M, Talmage DW, Frank MM, et al, Eds. Little, Brown & Co, Boston, 1988, p 1335

54. Asconape JJ: Some common issues in the use of antiepileptic drugs. Semin Neurol 22:27, 2002

55. Surks MI, DeFesi CR: Normal serum free thyroid hormone concentrations in patients treated with phenytoin or carbamazepine: a paradox resolved. JAMA 275:1495, 1996

56. Toone BK, Wheller M, Fenwick PBC: Sex hormone changes in male epileptics. Clin Endocrinol 12:391, 1980

57. Practice parameter: a guideline for discontinuing antiepileptic drugs in seizure-free patients—summary statement. Quality Standards Subcommittee of the American Academy of Neurology Neurology. 47:600, 1996

58. Schachter SC: Vagus nerve stimulation therapy summary: five years after FDA approval. Neurology 59(suppl 4):S15, 2002

59. Wheless JW, Baumgartner J, Ghanbari C: Vagus nerve stimulation and the ketogenic diet. Neurol Clin 19:371, 2001

60. Engel JJ: Surgical Treatment of the Epilepsies, 2nd ed. Raven Press, New York, 1993

61. Salanova V, Markand ON, Worth R: Clinical characteristics and predictive factors in 98 patients with complex partial seizures treated with temporal resection. Ann Neurol 51:1008, 1994

62. Jarrar RG, Buchhalter JR, Meyer FB, et al: Long-term follow-up of temporal lobectomy in children. Neurology 59:1635, 2002

63. Sorenson JM, Wheless JW, Baumgartner JE, et al: Corpus callosotomy for medically intractable seizures. Pediatr Neurosurg 27:260, 1997

64. Towne AR, Pellock JM, Ko D, et al: Determinants of mortality in status epilepticus. Epilepsia 35:27, 1994

65. Lowenstein DH, Alldredge BK: Status epilepticus. N Engl J Med 338:970, 1998

66. A comparison of four treatments for generalized convulsive status epilepticus. Veterans Affairs Status Epilepticus Cooperative Study Group. N Engl J Med 339:792, 1998

67. DeLorenzo RJ, Waterhouse EJ, Towne AR, et al: Persistent nonconvulsive status epilepticus after control of convulsive status epilepticus. Epilepsia 39:833, 1998

68. Bleck TP: Advances in the management of refractory status epilepticus. Crit Care Med 21:955, 1993

69. Willmore LJ: Prognosis of epilepsy. Prognosis in Neurological Disease. Evans RW, Baskin DS, Yatsu FM, Eds. Oxford University Press, New York, 1989, p 483

70. Beghi E, Togoni G: Prognosis of epilepsy in newly referred patients: a multicenter prospective study. Epilepsia 29:236, 1988

71. Kotagal P, Rothner AD, Erenberg G, et al: Complex partial seizures of childhood onset: a five-year follow-up study. Arch Neurol 44:1177, 1987

72. Leestma JE, Annegers JF, Brodie MJ, et al: Sudden unexplained death in epilepsy: observations from a large clinical development program. Epilepsia 38:47, 1997

73. Annegers JF, Coan SP, Hauser WA, et al: Epilepsy, vagal nerve stimulation by the NCP system, mortality, and sudden, unexpected, unexplained death. Epilepsia 39:206, 1998

74. Glantz MJ, Cole BF, Forsyth PA, et al: Practice parameter: anticonvulsant prophylaxis in patients with newly diagnosed brain tumors: report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 54:1886, 2000

75. Mattson RH, Cramer JA, Darney PD, et al: Use of oral contraceptives by women with epilepsy. JAMA 256:238, 1986

76. Delgado-Escueta AV, Janz D: Consensus guidelines: preconception counseling, management, and care of the pregnant woman with epilepsy. Neurology 42(suppl 5):149, 1992

77. Yerby MS: Pregnancy, teratogenesis, and epilepsy. Neurol Clin 12:749, 1994

78. Lindhout D, Meinardi H: Spina bifida and in-utero exposure to valproate. Lancet 2:396, 1984

79. Rosa FW: Spina bifida in infants of women treated with carbamazepine during pregnancy. N Engl J Med 324:674, 1991

80. Leppik IE: Treatment of epilepsy in the elderly. The Treatment of Epilepsy: Principles and Practice, 3rd ed. Wyllie E, Ed. Lippincott Williams & Wilkins, Philadelphia, 2001, p 787

81. Jacobson MP: Epilepsy in Aging Populations. Curr Treat Options Neurol 4:19, 2002

82. Willmore LJ: The effect of age on pharmacokinetics of antiepileptic drugs. Epilepsia 36(suppl):S14, 1995

83. Pellock JM: Standard approach to antiepileptic drug therapy in the United States. Epilepsia 35(suppl 4):S11, 1994

84. Glauser TA: Idiosyncratic reactions: new methods of identifying high-risk patients. Epilepsia 41(suppl 8):S16, 2000

Acknowledgment

The author wishes to thank Howard L. Kim, M.D., for his assistance.

XIII DISORDERS OF SLEEP

SUDHANSU CHOKROVERTY, M.D.

Sleep and wakefulness are two basic biologic processes that have independent functions and controls. Sleep has been defined on the basis of behavioral criteria (e.g., lack of mobility or slight mobility, closed eyes, a markedly reduced response to external stimulation, a characteristic sleeping posture, and a reversibly unconscious state) and physiologic criteria (e.g., readings obtained by electroencephalography [EEG], electro-oculography [EOG], and electromyography [EMG]).[1-3]

The exact biologic functions of sleep are not known, although it is known that sleep is essential and that sleep deprivation leads to impaired attention and decreased performance, in addition to sleepiness. Sleep is believed to have restorative, conservative, adaptive, thermoregulatory, and consolidative functions. Sleep is also required for maintenance of synaptic and neuronal network integrity.[1] The theory that memory reinforcement and consolidation take place during sleep has been strengthened by scientific data.[4,5] Although the exact physiologic functions of dreaming are unknown, they may include activation of the neural networks of the brain, restructuring and reinterpretation of data stored in memory, and removal of unnecessary and useless information from the brain.

Primary care physicians should have a high index of suspicion for the presence of sleep disorders. Most sleep disorders, once recognized, can be managed with limited consultation. The initial step is to treat any condition that may be secondarily responsible for hypersomnia or insomnia. However, the treatment of primary sleep disorders is best handled by a sleep specialist.

Physiology of Sleep

SLEEP ARCHITECTURE AND STAGES

Sleep is divided into two independent states: non–rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. NREM sleep is further divided into four stages [*see* Non–Rapid Eye Movement Sleep, *below*], primarily on the basis of electroencephalographic criteria [*see Figure 1a*]; together, stage III and stage IV NREM sleep constitute slow wave sleep. NREM and REM sleep alternate, with each cycle lasting about 90 to 100 minutes. Four to six such cycles are noted during a normal sleep period. The first third of a normal sleep period is dominated by slow wave sleep (see below), and the last third is dominated by REM sleep.

Non–Rapid Eye Movement Sleep

In stage I NREM sleep, the alpha wave (8 to 13 Hz) rhythms characteristic of wakefulness diminish to less than 50% in an epoch (i.e., a 30-second segment of a polysomnographic tracing with a speed of 10 mm/sec), and a mixture of slower rhythms called theta waves (4 to 7 Hz) appears; EMG activity decreases slightly, and slow, rolling eye movements may be apparent [*see Figure 1b*]. Toward the end of stage I NREM sleep, vertex sharp waves appear. In normal persons, stage II NREM sleep begins

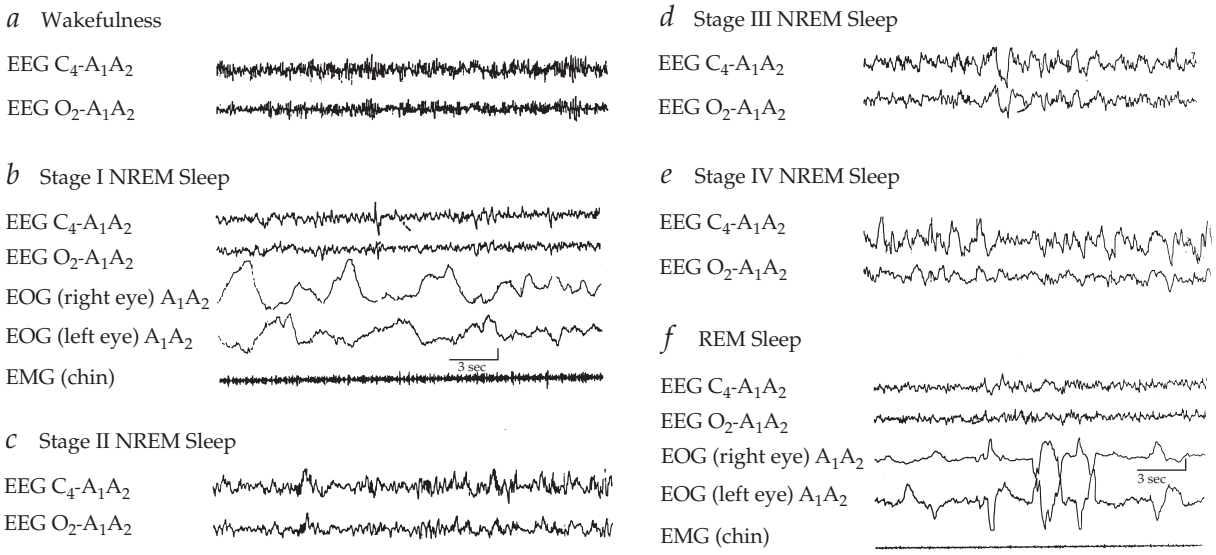


Figure 1 **Human sleep and wakefulness stages are characterized by readings obtained by electroencephalography (EEG), electro-oculography (EOG), and electromyography (EMG). (*a*) An EEG tracing shows the alpha wave (9 to 10 Hz) activity representative of wakefulness in adults. (*b*) Stage I non–rapid eye movement (NREM) sleep is characterized by low-amplitude theta wave (4 to 7 Hz) activity mixed with a small amount (< 50%) of alpha wave activity and vertex sharp waves in the EEG tracing; slow, rolling movements in right and left eye EOG recordings; and tonic chin EMG activity. (*c*) EEG tracings of stage II NREM show sleep spindles. (*d*) EEG tracings reveal that in stage III NREM sleep, 20% to 50% of an epoch is occupied by waves of 2 Hz or less with an amplitude of greater than 75 mV. (*e*) In stage IV NREM sleep, these waves make up more than 50% of an epoch. Together, stages III and IV constitute slow wave sleep. (*f*) REM sleep is characterized on EEG tracings by low-amplitude mixed-frequency theta wave and beta wave (> 13 Hz) activity, including a small amount of alpha waves without sleep spindles or K complexes. Rapid eye movements are seen in the EOG readouts. The chin EMG shows marked reduction or absence of tonic activity.**

after about 10 to 12 minutes of stage I NREM sleep. The characteristic EEG features of stage II NREM sleep are sleep spindles (12 to 18 Hz, most often 14 Hz) and K complexes intermixed with vertex sharp waves [*see Figure 1c*]. The background rhythms of stage II NREM sleep consist of theta waves and delta waves (< 4 Hz). Less than 20% of the epoch is occupied by delta waves. Stage II NREM sleep lasts for about 30 to 60 minutes. During stage III NREM sleep, the delta waves occupy 20% to 50% of the epoch [*see Figure 1d*], and during stage IV NREM sleep, the delta waves occupy more than 50% of the epoch [*see Figure 1e*].

Rapid Eye Movement Sleep

Sixty to 90 minutes after the onset of sleep, the first REM sleep episode is noted. The EEG tracings during REM sleep are characterized by fast rhythms and theta wave activities, some of which may have a sawtooth appearance [*see Figure 1f*]. The hallmark of REM sleep is the presence of rapid eye movements in all directions and the marked diminution or absence of muscle activities in the chin EMG. Tonic REM sleep is characterized by a desynchronized EEG and muscle atonia, and phasic REM sleep is characterized by REMs as well as phasic swings in blood pressure and heart rate, irregular respiration, and phasic tongue movements. A few periods of apnea or hypopnea may arise during REM sleep.

SLEEP PATTERNS ACROSS THE LIFESPAN

Sleep requirements change dramatically from infancy to old age. Newborns have a polyphasic sleep pattern, with a total of 16 hours of sleep a day. By the time a child is 3 to 5 years of age, the sleep requirement falls to about 10 hours a day. In preschool children, sleep assumes a biphasic pattern. Adults exhibit a monophasic sleep pattern, with an average duration of 7.5 to 8.0 hours a night, but the pattern reverts to biphasic in elderly persons.

In newborn infants, the amount of sleep time spent in the REM state is about 50%, but by 6 years of age, REM sleep has decreased to the normal adult pattern of 25%. By 3 months of age, the NREM-REM cyclic pattern of adult sleep is established and sleep spindles appear. K complexes are first seen in infants at about 6 months of age. Attenuation of the amplitude of delta waves and repeated awakenings—including early morning awakenings—are important features of sleep in elderly persons.

FUNCTIONAL NEUROANATOMY

The neuroanatomic substrates for REM sleep and NREM sleep are located in separate parts of the central nervous system.[2,6] REM sleep is generated by neurons in the pontomesencephalic region.[6,7] The cells that activate REM sleep are cholinergic neurons located in the pedunculopontine tegmental nucleus and the laterodorsal tegmental nucleus in the pontomesencephalic region; these cells fire maximally during REM sleep. The cells that deactivate REM sleep are aminergic neurons located in the locus coeruleus and raphe nuclei; these cells are inactive during REM sleep. A reciprocal interaction in the brain stem between REM-activating and REM-deactivating neurons is responsible for REM generation and maintenance.

The active hypnogenic neurons for NREM sleep are located primarily in the preoptic area of the hypothalamus, the basal forebrain area, and the solitary nucleus region of the medulla. The reticular nucleus of the thalamus is thought to be responsible for sleep spindle generation.

Sleep most likely results from both passive and active mechanisms.[6] Muscle hypotonia during REM sleep is thought to depend on the inhibitory postsynaptic potentials generated by dorsal pontine descending axons.[8,9] The roles of neurotransmitters and neuropeptides are not clearly delineated; they are believed to modulate various sleep stages and cycles. It has been shown that

Table 1 Behavioral and Physiologic Characteristics of States of Awareness

	Characteristics	*Awake*	*NREM Sleep*	*REM Sleep*
Behavioral	Posture	Erect, sitting, or recumbent	Recumbent	Recumbent
	Mobility	Normal	Mildly reduced to absent; postural shifts	Moderately reduced to absent; myoclonic jerks
	Response to stimulation	Normal	Mildly to markedly reduced	Moderately reduced to absent
	Alertness level	Alert	Reversibly unconscious	Reversibly unconscious
	Eye position and movements	Open	Closed	Closed
		WEMs	SEMs	REMs
Physiologic	EEG	Alpha waves Desynchronized	Synchronized	Theta or sawtooth waves Desynchronized
	EMG (muscle tone)	Normal	Mildly reduced	Moderately to severely reduced or absent
	EOG	WEMs	SEMs	REMs
	Heart rate	Normal sinus rhythm	Bradycardia	Bradytachyarrhythmia
	Blood pressure	Normal	Decreases	Variable
	Cardiac output	Normal	Decreases	Decreases further
	Respiratory rate	Normal	Decreases	Variable Apneas may occur
	Alveolar ventilation	Normal	Decreases	Decreases further
	Upper airway muscle tone	Normal	Mildly reduced	Moderately to markedly reduced or absent
	Upper airway resistance	Normal	Moderately increased	Markedly increased
	Cerebral blood flow	Normal	Decreased or normal	Markedly increased
	Parasympathetic activity	Normal	Moderately increased	Markedly increased
	Sympathetic activity	Normal	Mildly reduced	Moderately to markedly reduced or variable
	Penile or clitoral tumescence	Normal	Normal	Markedly increased
	Thermoregulation	Normal	Mildly reduced	Absent

NREM—non–rapid eye movement REM—rapid eye movement SEM—slow eye movement WEM—waking eye movement

adenosine may fulfill the major criteria for neural sleep factor, which mediates the somnogenic effects of prolonged wakefulness.[10]

The recently described hypocretin (orexin) peptidergic system, which is located in the lateral hypothalamic region and perifornical area and has widespread ascending and descending projections, is thought to play a role in the control of sleep and wakefulness.[11-13] Sleepiness may be induced by reduction of activity of the hypocretin projections to the locus coeruleus; midline raphe; and the mesopontine, posterior hypothalamic, and tuberomammillary regions.

PHYSIOLOGIC CHANGES DURING SLEEP

A variety of behavioral and physiologic changes occur during normal wakefulness, NREM sleep, and REM sleep [*see Table 1*]. These changes are most commonly noted in the somatic and autonomic nervous systems (ANS); in the respiratory, cardiovascular, and gastrointestinal systems; in endocrine, renal, and sexual function; and in thermoregulation.[2,14] The fundamental changes in the ANS consist of an increase in the parasympathetic tone and a decrease in sympathetic activity during NREM sleep, with further increase of parasympathetic tone and decrease in sympathetic activity during REM sleep. During the REM sleep phase, however, sympathetic activity increases intermittently. Sympathetic nerve activity in muscle and the vascular beds of the skin, as measured by microneurographic technique, is reduced during NREM sleep but is increased during REM sleep.[15]

During both NREM and REM sleep, the respiratory neurons in the pontomedullary region show a decreased firing rate. Muscle tone in the upper airway decreases in NREM sleep and disappears in REM sleep, resulting in an increase in upper airway resistance. Two separate systems—the metabolic (or automatic) and voluntary (or behavioral) systems—control respiration during sleep and wakefulness. Both the metabolic and the voluntary system operate during wakefulness, whereas only the metabolic system operates during NREM sleep. The wakefulness stimuli that act through the ascending reticular activating system also act as tonic stimuli to ventilation. Hypercapnic and hypoxic ventilatory responses are moderately reduced in NREM sleep but are more markedly decreased during REM sleep. Tidal volume and alveolar ventilation decrease during sleep; arterial oxygen tension is mildly decreased and arterial carbon dioxide tension is increased during both NREM and REM sleep. Thus, respiration is vulnerable during normal sleep, and a few periods of apnea may occur, especially at the onset of sleep and during REM sleep. Heart rate, blood pressure, cardiac output, and peripheral vascular resistance decrease during NREM sleep and decrease still further during REM sleep. Cerebral blood flow and cerebral metabolic rates for glucose and oxygen decrease during NREM sleep but increase to above waking values during REM sleep.

Profound changes in endocrine secretions are registered during sleep. Growth hormone secretion exhibits a pulsatile increase during NREM sleep in the first third of the normal sleep period. Prolactin secretion also rises 30 to 90 minutes after the onset of sleep. Sleep inhibits cortisol secretion. Secretion of thyroid-stimulating hormone reaches a peak in the evening and then decreases throughout the night. Testosterone levels in men increase during sleep, rising from trough levels at 8 P.M. to peak levels at 8 A.M., but no clear relation has been demonstrated between levels of gonadotropic hormones and the sleep-wake cycle in children or adults. Melatonin, which is released by the pineal gland, attains its highest secretion level between 3 A.M. and 5 A.M., then decreases to low levels during the day.

Body temperature begins to fall at the onset of sleep and reaches its lowest point during the third sleep cycle. Thermoregulation is maintained during NREM sleep but is nonexistent in REM sleep. Penile erection and clitoral tumescence occur during REM sleep.

CIRCADIAN RHYTHMS

Circadian means a period of about 24 hours; however, the human circadian rhythm cycle is about 25 hours. The existence of circadian rhythms has been known since 1731, when the French astronomer J. J. D. de Mairan noted a circadian rhythm in a plant.[16] The suprachiasmatic nucleus of the hypothalamus is thought to be the site of this biologic clock. The body temperature rhythm closely follows the circadian rhythm of sleep and wakefulness but is independent of it. Dysfunction of circadian rhythms results in some important human sleep disorders.[17]

General Approach to Disorders of Sleep

Sleep complaints are very common in the general population. According to the report of the National Center on Sleep Disorders Research, more than 40 million persons in the United States suffer from chronic disorders of sleep and wakefulness.[18]

CLASSIFICATION

The second edition of the International Classification of Sleep Disorders (ICSD-2) lists eight categories of sleep disorders.[17] In addition, two appendices in the ICSD-2 list sleep disorders associated with other conditions, as well as psychiatric and behavioral disorders frequently encountered in the differential diagnosis of sleep disorders [*see Table 2*].[17]

DIAGNOSIS

Clinical Manifestations

The four major sleep-related complaints for which patients seek medical attention are excessive daytime somnolence (EDS), insomnia, abnormal movements or behavior during sleep, and an inability to sleep at the desired time. Insomnia patients may complain of some or all of the following: difficulty initiating or maintaining sleep, repeated awakenings or early morning awakenings, nonrestorative sleep, daytime fatigue, lack of concentration, irritability, anxiety, depression, and muscle aches and pains. Insomnia may be idiopathic (i.e., no cause is found) or comorbid with other conditions [*see Table 3*]. Patients with hypersomnia may complain of EDS, a lack of relief of symptoms after additional nighttime sleep, inability to concentrate, and impaired cognition and motor skills. The most common cause of EDS is behaviorally induced insufficient sleep syndrome (sleep deprivation) [*see Table 2*]. Some patients with restless legs syndrome (RLS), periodic limb movements in sleep (PLMS), or circadian rhythm sleep disorders may also experience hypersomnolence in the daytime.

The first step in assessing a patient with a sleep disturbance is obtaining a history, which should include a family history and directed inquiries on sleeping habits, drug and alcohol consumption, and previous or current psychiatric, medical, and neurologic illnesses. Diagnostic tests must be confirmatory and subservient to the clinical history and physical examination. Electrophysiologic investigation of patients with suspected sleep disorders is resource intensive and is best planned in consultation with a sleep specialist.

Table 2 International Classification of Sleep Disorders, Second Edition[17]

Insomnia
- Acute insomnia
- Psychophysiologic insomnia
- Paradoxical insomnia (sleep-state misperception)
- Idiopathic insomnia
- Insomnia due to mental disorder
- Inadequate sleep hygiene
- Behavioral insomnia of childhood
- Insomnia due to drug or substance
- Insomnia due to medical condition
- Insomnia not due to substance or known physiologic condition, unspecified (nonorganic)
- Physiologic insomnia, unspecified (organic)

Sleep-Related Breathing Disorders
- Central sleep apnea syndromes
  - Primary central sleep apnea
  - Central sleep apnea due to Cheyne-Stokes breathing pattern
  - Central sleep apnea due to high-altitude periodic breathing
  - Central sleep apnea due to medical condition not Cheyne-Stokes
  - Central sleep apnea due to drug or substance
  - Primary sleep apnea of infancy (formerly primary sleep apnea of newborn)
- Obstructive sleep apnea syndromes
  - Obstructive sleep apnea, adult
  - Obstructive sleep apnea, pediatric
- Sleep-related hypoventilation/hypoxemic syndromes
  - Sleep-related nonobstructive alveolar hypoventilation, idiopathic
  - Congenital central alveolar hypoventilation syndrome
- Sleep-related hypoventilation/hypoxemia due to medical condition
  - Sleep-related hypoventilation/hypoxemia due to pulmonary parenchymal or vascular pathology
  - Sleep-related hypoventilation/hypoxemia due to lower airway obstruction
  - Sleep-related hypoventilation/hypoxemia due to neuromuscular and chest wall disorders
- Other sleep-related breathing disorder
  - Sleep apnea/sleep-related breathing disorder, unspecified

Hypersomnias of Central Origin Not Due to a Circadian Rhythm Sleep Disorder, Sleep-Related Breathing Disorder, or Other Cause of Disturbed Nocturnal Sleep
- Narcolepsy with cataplexy
- Narcolepsy without cataplexy
- Narcolepsy due to medical condition
- Narcolepsy, unspecified
- Recurrent hypersomnia
  - Kleine-Levin syndrome
  - Menstrual-related hypersomnia
- Idiopathic hypersomnia with long sleep time
- Idiopathic hypersomnia without long sleep time
- Behaviorally induced insufficient sleep syndrome
- Hypersomnia due to medical condition
- Hypersomnia due to drug or substance
- Hypersomnia not due to substance or known physiologic condition (nonorganic hypersomnia, not otherwise specified [NOS])
- Physiologic (organic) hypersomnia, unspecified (organic hypersomnia, NOS)

Circadian Rhythm Sleep Disorders
- Circadian rhythm sleep disorder, delayed-sleep-phase type (delayed-sleep-phase disorder)
- Circadian rhythm sleep disorder, advanced-sleep-phase type (advanced-sleep-phase disorder)
- Circadian rhythm sleep disorder, irregular sleep-wake type (irregular sleep-wake rhythm)
- Circadian rhythm sleep disorder, free-running type (nonentrained type)
- Circadian rhythm sleep disorder, jet-lag type (jet-lag disorder)
- Circadian rhythm sleep disorder, shift work type (shift-work disorder)
- Circadian rhythm sleep disorder due to medical condition
- Other circadian rhythm sleep disorder (circadian rhythm disorder, NOS)
- Other circadian rhythm sleep disorder due to drug or substance

(*continued*)

Laboratory Studies

The two most important laboratory tests for sleep disorders are the all-night polysomnography (PSG) study and the Multiple Sleep Latency Test (MSLT). All-night PSG studies simultaneously record several physiologic variables (i.e., EEG, EMG, EOG, electrocardiography, airflow at the nose and mouth, respiratory effort, and oxygen saturation) and are important in confirming a diagnosis of EDS or obstructive sleep apnea syndrome (OSAS), as well as documenting the severity of sleep apnea, hypoxemia, and sleep fragmentation. Overnight PSG determines the optimal pressure for continuous positive airway pressure (CPAP)—a treatment for OSAS—and is also helpful for supporting the diagnosis of narcolepsy and the parasomnias. Overnight PSG with simultaneous video recording can confirm REM sleep behavior disorder (RBD) and is particularly useful for the documentation of unusual movements and behavior during nighttime sleep in patients with parasomnias and nocturnal seizures.

The MSLT is essential in documenting pathologic sleepiness (e.g., sleep-onset latency of less than 5 minutes)[19] and in diagnosing narcolepsy. The presence of two sleep-onset REMs on four or five nap studies and sleep-onset latency of less than 8 minutes strongly suggest a diagnosis of narcolepsy.

Another important laboratory test for assessing sleep disorders is actigraphy.[20] This technique utilizes an actigraph (also known as an actometer) worn on the wrist or ankle to record acceleration or deceleration of body movements, which indirectly indicates sleep-wakefulness. Actigraphy for days or weeks is a useful laboratory test in patients with insomnia and circadian rhythm sleep disorders, as well as in some patients with prolonged daytime sleepiness.

Magnetic resonance imaging studies and other neuroimaging techniques should be performed to exclude structural neurologic lesions. Appropriate laboratory tests should always be performed to exclude any suspected medical disorders that may be the cause of the patient's insomnia or hypersomnia.

Specific Disorders of Sleep

OBSTRUCTIVE SLEEP APNEA SYNDROME

Apnea, or cessation of breathing during sleep, consists of three types: central, obstructive, and mixed.[2] In central apnea, both the airflow at the upper airway (e.g., pharynx and nose) and the effort by the diaphragm and other respiratory mus-

Table 2 (*continued*)

- *Parasomnias*
  - Disorders of arousal (from non–rapid eye movement [REM] sleep)
    - Confusional arousals
    - Sleepwalking
    - Sleep terrors
  - Parasomnias usually associated with REM sleep
    - REM sleep behavior disorder (including parasomnia overlap disorder and status dissociates)
    - Recurrent isolated sleep paralysis
    - Nightmare disorder
  - Other parasomnias
    - Sleep-related dissociative disorders
    - Sleep enuresis
    - Sleep-related groaning (catathrenia)
    - Exploding head syndrome
    - Sleep-related hallucinations
    - Sleep-related eating disorders
    - Parasomnia, unspecified
    - Parasomnia due to drug or substance
    - Parasomnia due to medical condition
- *Sleep-Related Movement Disorders*
  - Restless legs syndrome
  - Periodic limb movement disorder
  - Sleep-related leg cramps
  - Sleep-related bruxism
  - Sleep-related rhythmic movement disorder
  - Sleep-related movement disorder, unspecified
  - Sleep-related movement disorder due to drug or substance
  - Sleep-related movement disorder due to medical condition
- *Isolated Symptoms, Apparently Normal Variants, and Unresolved Issues*
  - Long sleeper
  - Short sleeper
  - Snoring
  - Sleep talking
  - Sleep starts (hypnic jerks)
  - Benign sleep myoclonus of infancy
  - Hypnagogic foot tremor and alternating leg muscle activation during sleep
  - Propriospinal myoclonus at sleep onset
  - Excessive fragmentary myoclonus
- *Other Sleep Disorders*
  - Other physiologic (organic) sleep disorder
  - Other sleep disorder not due to substance or known physiologic condition
  - Environmental sleep disorder
- *Sleep Disorders Associated with Conditions Classifiable Elsewhere*
  - Fatal familial insomnia
  - Fibromyalgia
  - Sleep-related epilepsy
  - Sleep-related headaches
  - Sleep-related gastroesophageal reflux disease
  - Sleep-related coronary artery ischemia
  - Sleep-related abnormal swallowing, choking, and laryngospasm
- *Other Psychiatric and Behavioral Disorders Frequently Encountered in the Differential Diagnosis of Sleep Disorders*
  - Mood disorders
  - Anxiety disorders
  - Somatoform disorders
  - Schizophrenia and other psychotic disorders
  - Disorders usually first diagnosed in infancy, childhood, or adolescence
  - Personality disorders

cles cease. By contrast, during obstructive apnea, the airflow stops while the effort continues [*see Figure 2*]. In mixed apnea, an initial period of central apnea is followed by a period of obstructive apnea. The most common type of apnea is OSAS.[21] Sleep hypopnea, defined as a decrease of breathing to half the volume measured during the preceding or following respiratory cycle accompanied by an arousal or oxygen desaturation of 3% to 4% or more, has the same significance as apnea.[2,21] To qualify as pathologic sleep apnea or hypopnea, the decreased breathing must last for at least 10 seconds and the apnea-hypopnea index or respiratory disturbance index (i.e., the number of episodes of apnea-hypopnea per hour of sleep) must be at least five.[2,21]

OSAS is common in middle-aged and elderly men. An important study indicated a prevalence of 4% in men and 2% in women between 30 and 60 years of age.[22] In women, the incidence of OSAS is greater after menopause. The pathogenesis of OSAS includes local anatomic factors (e.g., narrowing of the upper airway and excessive relaxation of the upper airway muscles during sleep, with increased upper airway resistance) and neurologic factors that may cause dysfunction of the respiratory neurons in the brain stem [*see 14:VI Ventilatory Control during Wakefulness and Sleep*].

Diagnosis

Nocturnal sleep symptoms of OSAS are loud snoring (often with a long history), choking, cessation of breathing, sitting up

Table 3 Conditions Comorbid with Insomnia

- Psychiatric disorders
  - Mood disorders (e.g., major depressive and bipolar disorders)
  - Anxiety disorders (e.g., panic and posttraumatic stress disorder)
- Stress
- Other psychophysiologic factors
- Medical disorders
  - Bronchial asthma
  - Coronary artery disease
  - Peptic ulcer disease
  - Rheumatic disorders
- Neurologic disorders
  - Stroke
  - Neurodegenerative diseases
  - Brain tumors
  - Headache syndromes
  - Neuromuscular disorders (e.g., painful peripheral neuropathies)
- Traumatic brain injury
- Fatal familial insomnia (a rare prion disease)
- Pain anywhere in the body
- Sleep-wake schedule disruptions
- Other disorders of circadian sleep rhythms
- Drug or alcohol abuse

and fighting for breath, abnormal motor activities and thrashing about in bed, gastroesophageal reflux, nocturia and nocturnal enuresis (seen mostly in children), and, occasionally, hyperhidrosis.

The major daytime symptom of OSAS is EDS; patients fall asleep during the day at inappropriate times and in inappropriate places and may be involved in driving accidents. They cannot function adequately during the day and may complain of morning headaches and forgetfulness; men may report impotence. The prolonged duration and the nonrefreshing nature of daytime sleep in EDS differentiate it from narcoleptic sleep attacks. Physical examination reveals obesity in about 70% of patients; in many cases, local anatomic abnormalities of the upper airway are found. Repeated hypoxemia associated with cessation of breathing during sleep can lead to hypertension, cardiac arrhythmias, heart failure, and stroke.[23]

Treatment

The general treatment of OSAS consists of avoiding sedatives, hypnotics, and alcohol and, in obese patients, reduction of body weight. Mild cases of OSAS may improve with these measures. Pharmacologic treatment remains unsatisfactory.[2,21]

Application of CPAP is effective in over 70% of moderate to severe cases of OSAS and is the best treatment available.[24] CPAP opens up the upper airway passages and acts as a pneumatic splint, thereby eliminating obstructive apneas, hypoxemias, snoring, and sleep fragmentation. The optimal pressure of CPAP must first be determined during overnight PSG. The patient can then purchase a home unit to use during nightly sleep. Manufacturers have devised auto-CPAP machines, which automatically titrate pressure according to detected problems. The role of these devices remains to be determined.

In the few severe cases of OSAS in which CPAP therapy fails, uvulopalatopharyngoplasty or other surgical approaches may be needed, although their roles remain uncertain. Laser or radiofrequency uvulopalatopharyngoplasty has been tried. Severe respiratory compromise or severe apnea associated with dangerous cardiac arrhythmias creates a life-threatening situation that may require emergency tracheostomy. In patients with neuromuscular disorders, intermittent positive pressure ventilation through a nasal mask may be needed to treat sleep hypoventilation or apnea. In some mild to moderate cases of OSAS, dental appliances may be tried.

NARCOLEPSY

The most important clinical manifestations of narcolepsy are sleep attacks and cataplexy, although the ICSD-2 includes one category as narcolepsy without cataplexy.[17,25] Narcoleptic sleep attacks usually begin in patients between 15 and 25 years of age but may occur earlier or later. Narcolepsy is a lifelong disorder, but it becomes less severe with old age. From 1% to 2% of the first-degree relatives of narcoleptic patients manifest the illness, compared with 0.02% to 0.18% in the general population.[26] The cause of narcolepsy is unknown and is thought to result from an interplay of genetic and environmental factors. There is a strong association between narcolepsy and the presence of the HLA DQB1*0602 subtype.[26]

The most exciting recent development in our understanding of narcolepsy is the documentation of an abnormality in the hypocretin (orexin) neurons in the lateral hypothalamus.[27-32] Human narcolepsy-cataplexy can be considered a hypocretin (orexin) deficiency syndrome. Four lines of evidence can be cited in support of hypocretin abnormality in narcolepsy: (1) induction of narcolepsy-like symptoms after mutation of the hypocretin receptor 2 gene in dogs and after preprohypocretin knockout in mice, (2) decreased hypocretin 1 levels in the cerebrospinal fluid of narcolepsy-cataplexy patients, (3) postmortem documentation of decreased numbers of hypocretin neurons in narcoleptic brains, and (4) preprohypocretin gene mutation in a child with severe narcolepsy associated with generalized absence of hypocretin peptides in the brain.

Diagnosis

The classic narcoleptic sleep attack consists of an irresistible desire to sleep in an inappropriate place and under inappropriate circumstances. Attacks generally last from a few minutes to 30 minutes, and the patient feels refreshed on awakening. The initial manifestations of narcolepsy are generally sleep attacks; after months or years, 60% to 70% of patients experience cataplexy, during which they transiently lose muscle tone (e.g., sagging of the head, drooping of the eyelids, and buckling of the knees). Cataplexy is often triggered by a sudden emotional outburst. The patient may momentarily collapse and fall to the ground or may simply slump forward for a few seconds before regaining awareness.

Additional manifestations of narcolepsy include hypnagogic or hypnopompic hallucinations (fearful or vivid dreams at sleep onset or end), sleep paralysis (feeling of loss of power or focal weakness on one side of the body or in one limb during the transition period between sleep and wakefulness), disturbance of nocturnal sleep, and automatic behaviors (during which the patient might keep doing one thing repeatedly or drive to a place without recollection of the events). In some patients, narcoleptic sleep attacks may be associated with sleep apnea, REM behavior disorder, or PLMS.

Treatment

The administration of a stimulant (e.g., modafinil, methylphenidate, dextroamphetamine, or methamphetamine) is the treatment of choice for narcoleptic sleep attacks.[25] In newly diagnosed patients, the drug most commonly used initially is modafinil, a novel wakefulness-promoting agent, or methylphenidate. Dextroamphetamine and methamphetamine are used in patients who do not respond satisfactorily to these stimulant drugs.

Tricyclic antidepressants (e.g., protriptyline, imipramine, and clomipramine) and selective serotonin reuptake inhibitors (SSRIs, such as fluoxetine) are used to treat cataplexy or other auxiliary symptoms of narcolepsy. Sodium oxybate, which is more commonly known as γ-hydroxybutyric acid (GHB) [*see 8:I Management of Poisoning and Drug Overdose*], has been approved by the Food and Drug Administration for the treatment of cataplexy. It is an endogenous hypnotic that acts by consolidating REM and slow wave sleep. The medication is administered in two divided doses at night. The first dose is taken at bedtime; the patient must awaken 2.5 to 4 hours later to take the second dose. Taking short daytime naps and joining narcolepsy support groups can be useful approaches as well.

IDIOPATHIC HYPERSOMNIA

The International Classification of Sleep Disorders defines idiopathic hypersomnia as a disorder of excessive sleepiness of presumed CNS cause that is associated with major sleep episodes.[17] The disorder is further categorized as with or without long sleep time (over 10 hours or from 6 to 10 hours, respectively). The disease develops insidiously, generally between the ages

of 15 and 30 years. It closely resembles narcolepsy. Affected patients generally sleep for hours, and the sleep is not refreshing. The patient does not give a history of cataplexy or snoring.

Idiopathic hypersomnia is a disabling and lifelong disorder. The MSLT shows a mean sleep latency of less than 8 minutes without sleep-onset REMs. The treatment of idiopathic hypersomnia is similar to the stimulant treatment of narcolepsy; however, the therapeutic response is unsatisfactory.

RESTLESS LEGS SYNDROME AND PERIODIC LIMB MOVEMENTS IN SLEEP

Restless Legs Syndrome

RLS is a lifelong sensorimotor neurologic disorder that may begin at any age.[33] RLS is most severe in middle-aged or elderly persons, in whom it has a chronic, progressive course. The prevalence of RLS has been estimated to be about 10% for all adult populations, particularly those of European descent. However, in some surveys from Asia, the prevalence is extremely low, suggesting a possible ethnic and racial difference in RLS prevalence.

Clinical features Minimal criteria for the clinical diagnosis of RLS include the following four sensorimotor features[33,34]: (1) the patient has an urge or compulsion to move the legs, usually accompanied by uncomfortable or unpleasant sensations, primarily in the legs (generally below the knees) but sometimes in the arms; (2) the urge to move or uncomfortable sensations begin or worsen during rest and repose or inactivity; (3) the urge to move or uncomfortable sensations are partially or totally relieved by movement, at least in the beginning of the illness, and at least as long as the activity continues; (4) the uncomfortable sensations or the urge to move is worse in the evening or early part of the night than during the day.

Supportive diagnostic criteria include responsiveness to dopaminergic drugs, at least early in the disease; presence of periodic limb movements; and positive family history.[34] Up to about 50% of patients have a family history of a similar condition, suggesting a dominant mode of inheritance. At least 80% of RLS patients have PLMS (see below). In addition, many RLS patients also have periodic limb movements in wakefulness (PLMW). Associated features of RLS include a progressive clinical course, nor-

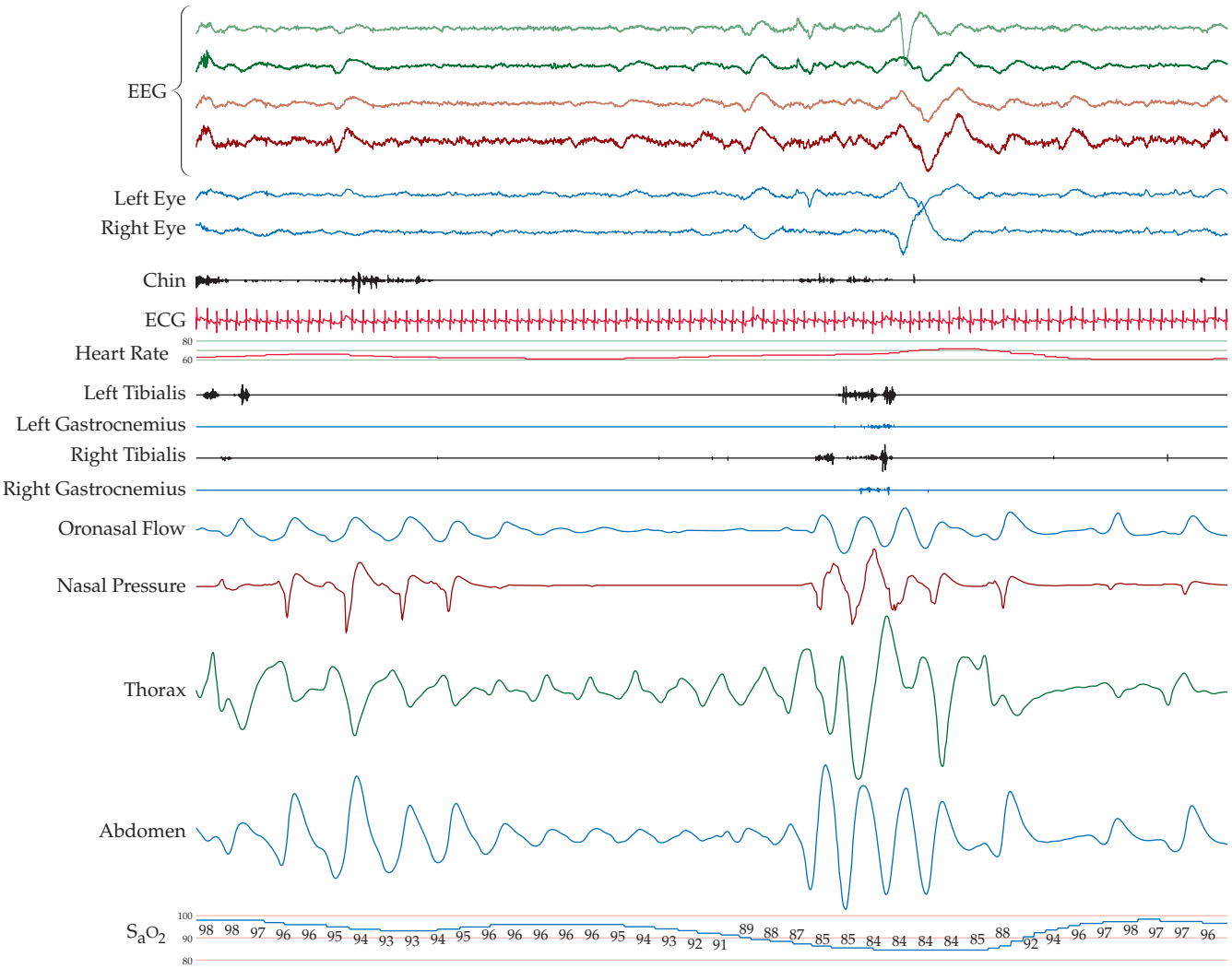


Figure 2 **Polysomnographic (PSG) recording of a patient with sleep apnea syndrome includes EEG tracings (top four channels), EOG readouts of the left and the right eyes, EMG of mentalis muscle (chin), electrocardiogram, EMG of the left tibialis and right tibialis and gastrocnemius muscles, and recordings of oronasal airflow, nasal pressure, respiratory effort (thoracic and abdominal readings), and oxygen saturation (S_aO_2). The PSG study shows episodes of obstructive sleep apnea during stage II NREM sleep, followed by arousals.**

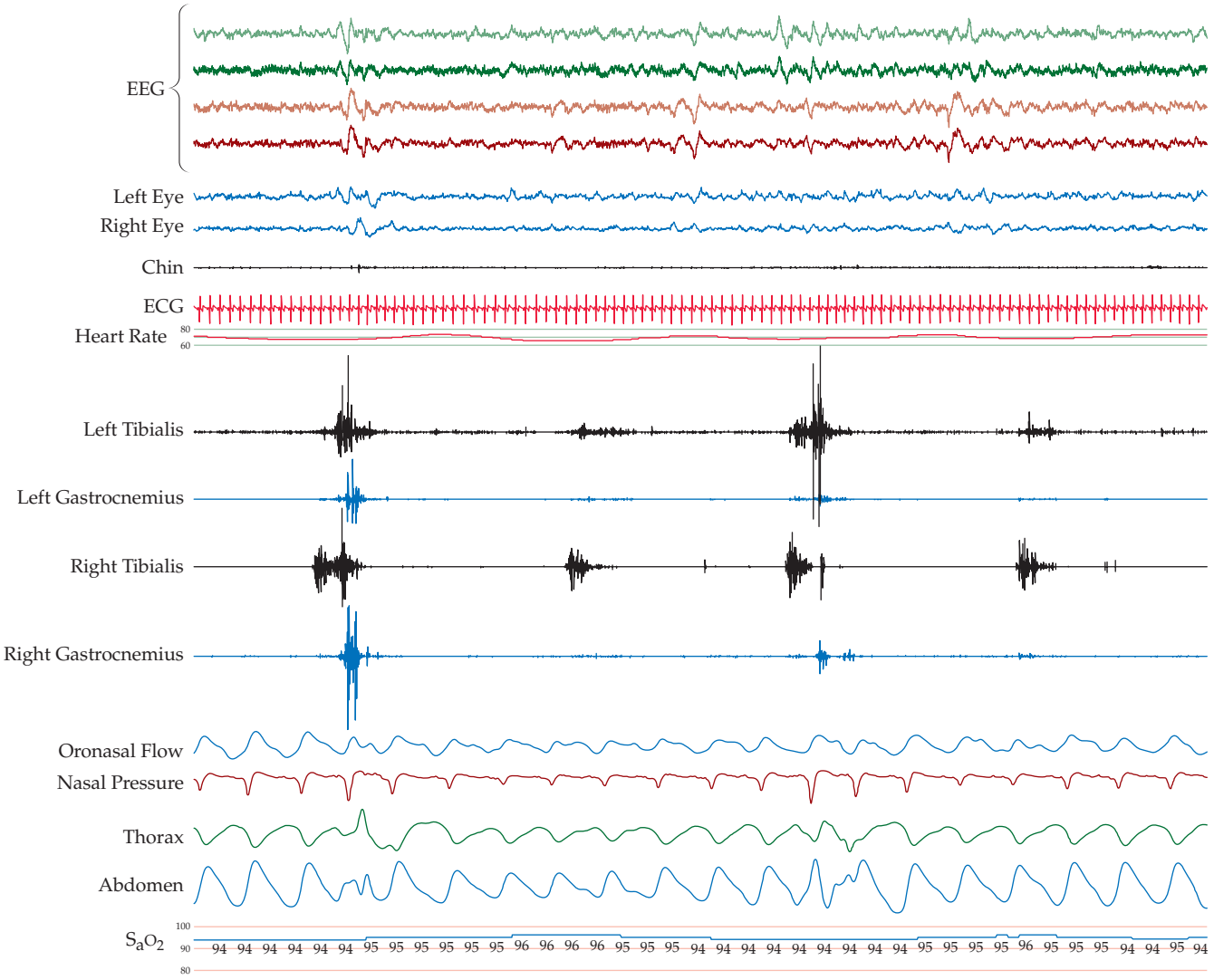


Figure 3 **Polysomnographic recording of a patient with restless legs syndrome includes EEG tracings (top four channels), EOG readouts of the left and the right eyes, EMG of the mentalis muscle (chin), electrocardiogram, EMG of the left and the right tibialis and gastrocnemius muscles, and recordings of oronasal airflow, nasal pressure, respiratory effort (thoracic and abdominal readings), and oxygen saturation (S_aO_2). The tibialis and gastrocnemius EMG recordings reveal periodic limb movements in sleep.**

mal neurologic examination in the idiopathic or primary type of RLS, and sleep disturbance.[34]

It is important to exclude potential underlying causes of RLS (e.g., iron deficiency anemia, uremia, and polyneuropathy). The condition may have a profound impact on sleep, and occasionally, patients have sleep apnea or daytime hypersomnolence.

Periodic Limb Movements in Sleep

PLMS is common in the general population, with increasing prevalence in persons of advanced age; patients may remain completely asymptomatic.[35,36] PLMS is a polysomnographic finding and is characterized by repetitive, often stereotyped limb movements during NREM sleep. The characteristic movements are usually extension of the great toe, dorsiflexion of the ankle, and flexion of the knee. Similar movements are sometimes seen in the arms. At least four consecutive movements occur on a pseudoperiodic basis, with an average interval of 20 to 40 seconds (range, 5 to 90 seconds) and a duration of 0.5 to 5.0 seconds [*see Figure 3*]. PLMS appears most commonly in RLS and may also occur in a large number of other medical, neurologic, and sleep disorders and with medication (e.g., tricyclic antidepressants, SSRIs, and levodopa). PLMS may occur with or without awakenings. Whether PLMS occurs as a primary condition unassociated with RLS causing repeated awakenings and sleep fragmentation remains controversial. In the current International Classification of Sleep Disorders, periodic limb movement disorder (PLMD) is listed as a separate entity.[17] A growing body of evidence suggests that PLMS may not have any clinical significance and is simply a polysomnographic observation noted in a variety of sleep disorders but present in the majority of patients with RLS.[35,36]

Treatment

Four groups of drugs are available to treat RLS and PLMS[33]: dopaminergic drugs (e.g., carbidopa-levodopa and dopamine agonists such as pergolide, pramipexole, ropinirole, and cabergoline); benzodiazepines (e.g., clonazepam); opioids (e.g., codeine,

propoxyphene, oxycodone, and hydrocodone); and anticonvulsants (e.g., gabapentin). The best drug for initial therapy in most cases is a dopamine agonist. The only drug currently approved by the Food and Drug Administration for use in RLS is ropinirole, but several clinical trials have proved that other agents can be used for this disorder.

CIRCADIAN RHYTHM SLEEP DISORDERS

In circadian rhythm sleep disorders, patients have difficulty sleeping as a result of desynchronization between their internal circadian rhythms and external time.[37] The insomnia of these disorders is often associated with other somatic complaints. The most common circadian rhythm sleep disorders are jet lag (associated with high-speed air travel across time zones) and shift-work sleep disorder (seen in patients who work nonstandard shifts). In delayed-sleep-phase syndrome, the patient has difficulty sleeping early in the evening, goes to sleep late (e.g., 2 A.M.), and wakes up late in the morning. Patients with advanced-sleep-phase syndrome go to sleep early in the evening and wake up early in the morning; this disorder is often seen in patients with depression and in normal elderly persons.

Appropriately timed bright-light therapy and chronotherapy may be effective in these disorders.[37] Jet lag and shift-work sleep disorder may also be treated with zolpidem. Melatonin has been found to be useful in some persons with jet lag and shift-work sleep disorders, as well as in patients with non–24-hour circadian rhythm disorders. There is no generally accepted standard dose for melatonin; in general, 0.5 to 3 mg has been used for jet lag, and 5 to 10 mg for non–24-hour syndrome. Melatonin has also been used in combination with bright-light therapy in patients with delayed-sleep-phase syndrome.

INSOMNIAS

Insomnia may occur at any age. The patient may complain of difficulty initiating or maintaining sleep or of awakening early in the morning and being unable to go back to sleep, as well as nonrestorative sleep or poor quality of sleep.[38] Insomnia may be associated with a variety of medical, psychiatric, and neurologic illnesses or may be drug- or alcohol-induced [*see Table 3*]. Insomnia is most commonly caused by psychiatric or psychophysiologic disorders, with depression and anxiety among the most important. Early morning awakening is a characteristic finding in depression. In some cases, no cause of the insomnia is found; this disorder is termed idiopathic, or primary, insomnia and is a lifelong condition.

For transient insomnia or insomnia of short duration, treatment with sedative-hypnotics (e.g., zolpidem, zaleplon, or eszopiclone) or short- or intermediate-acting benzodiazepines (e.g., temazepam) for a few nights to a few weeks is appropriate [*see Table 4*]. Hypnotic medications should not be used for chronic insomnia. The best treatment for patients with chronic insomnia consists of a combination of sleep hygiene measures (e.g., setting fixed times for retiring and awakening; avoidance of caffeinated beverages, tobacco, and alcohol before retiring; and regular exercise, preferably undertaken 4 to 6 hours before bedtime), stimulus-control therapy, sleep restriction, relaxation training, and other psychological treatments.[38] One study showed a trend toward better outcomes with a combination of cognitive-behavioral therapy (i.e., stimulus control, sleep restriction, sleep hygiene, and cognitive therapy) and pharmacotherapy.[39] Sedative-antidepressants should be used for insomnia associated with depression.

Table 4 Short-Term Drug Therapy for Insomnia

Category	*Agent*	*Dose (mg)**
Sedative-hypnotics	Zaleplon	5–10
	Zolpidem	5–10
	Zolpidem extended release	6.25–12.5
	Eszopiclone	1–3
	Ramelteon	8
Short-acting or intermediate-acting benzodiazepines	Temazepam	7.5–30
	Triazolam	0.125–0.250
	Flurazepam	15–30
	Estazolam	1–2

*All agents are given at bedtime for a few nights to a few weeks.

PARASOMNIAS

Partial Arousal Disorders

Partial arousal disorders include confusional arousals, sleepwalking, and sleep terrors.[40] Confusional arousals, which occur during slow wave sleep, mostly in children younger than 5 years, are characterized clinically by confusion and mild automatic and inappropriate behavior. Most confusional arousals are benign.

Sleepwalking, or somnambulism, is most common in children between 5 and 12 years of age. Sleepwalking begins with abrupt motor activity arising out of slow wave sleep during the first third of the sleep period. Episodes generally last less than 10 minutes. There is a high incidence of positive family history in sleepwalking. Injuries and violent actions have been reported during sleepwalking episodes. Sleep deprivation, fatigue, concurrent illness, and sedatives may act as precipitating factors.

Sleep terrors, or pavor nocturnus, also occur during slow wave sleep. Peak onset is between 5 and 7 years of age. As with sleepwalking, a high incidence of familial cases is seen in sleep terrors. Episodes of sleep terrors are characterized by intense autonomic and motor symptoms, including a loud, piercing scream. Many patients also have sleepwalking episodes. Precipitating factors are similar to those described with sleepwalking.

No special treatment is needed for most of the parasomnias. If attacks of sleepwalking or sleep terrors are frequent, treatment with an antidepressant or benzodiazepine may be tried for a short period.

Nocturnal Frontal Lobe Epilepsy

Nocturnal frontal lobe epilepsy (NFLE) was formerly known as nocturnal paroxysmal dystonia. NFLE may occur at any time from infancy to the fifth decade of life[41]; the disorder is seldom familial. The attacks occur suddenly during NREM sleep. The most common type is a spell of short duration (lasting from 15 seconds to less than 2 minutes). In rare cases, the spell is of longer duration (lasting from 2 minutes to more than 1 hour). The attacks are characterized by ballismic, choreoathetoid, or dystonic movements, which may occur in clusters. The EEG is generally normal. Carbamazepine is the treatment of choice for NFLE. Several families with autosomal dominant NFLE have been described.[42]

REM Sleep Behavior Disorder

REM sleep behavior disorder is an important REM sleep parasomnia commonly seen in elderly persons.[43] A characteristic clinical feature of RBD is intermittent loss of REM-related muscle hypotonia or atonia and the appearance of a variety of abnormal

motor activities during sleep. The patient presents with violent, dream-enacting behavior during REM sleep, often causing self-injury or injury to the patient's bed partner. Initially, it was thought that RBD was mostly idiopathic, but as more patients were described, it was realized that most cases are secondary and associated with neurodegenerative diseases. RBD occurs with increasing frequency in patients with Parkinson disease (PD), multiple system atrophy (MSA), diffuse Lewy body disease (DLBD), olivopontocerebellar atrophy (OPCA), progressive supranuclear palsy (PSP), and corticobasal ganglionic degeneration (CBD).[44] RBD has also been described in many cases of narcolepsy, which may be considered a degenerative disease of hypocretin-containing neurons in the hypothalamus. In many of these neurodegenerative diseases (e.g., PD, MSA, and DLBD), α-synuclein inclusions have been noted; some authors even propose that RBD may be an α-synucleinopathy disorder.[44] The findings of reduced striatal presynaptic dopamine transporter on positron emission tomographic scans and of a reduction of postsynaptic D_2 dopamine receptors on single-photon emission computed tomography scans suggest the linking of RBD to dopamine cell dysfunction.[45] RBD may precede many of these neurodegenerative diseases or coexist with these disorders. RBD may also be associated with structural lesions of the brain stem and is sometimes associated with alcoholism or the ingestion of drugs (e.g., sedative-hypnotics, tricyclic antidepressants, SSRIs, and anticholinergics). On polysomnographic studies, the prominent finding is REM sleep without muscle atonia. In experiments on cats, similar behavior has been produced by a bilateral perilocus coeruleus lesion. Most cases of RBD respond to a low dose of a benzodiazepine (e.g., clonazepam).

The author has been a member of the speakers' bureau for Sanofi-Aventis and has served as a consultant for Boehringer Ingelheim Corp.

References

1. Chokroverty S: An overview of sleep. Sleep and Movement Disorders. Chokroverty S, Hening W, Walters A, Eds. Butterworth-Heinemann, Boston, 2003, p 20

2. Chokroverty S: Clinical Companion to Sleep Disorders Medicine. Butterworth-Heinemann, Boston, 2000

3. Tobler I: Is sleep fundamentally different between mammalian species? Behav Brain Res 69:35, 1995

4. Walker MP, Stickgold R: Sleep, memory, and plasticity. Annu Rev Psychol 57:139, 2006

5. Peigneux P, Laureys S, Fuchs S, et al: Are spatial memories strengthened in the human hippocampus during slow wave sleep? Neuron 44:535, 2004

6. Sinton CM, McCarley RW: Neurophysiological mechanisms of sleep wakefulness: a question of balance. Semin Neurol 24:211, 2004

7. Steriade MM, McCarley RW: Brain Control of Wakefulness and Sleep, 2nd ed. Kluwer Academic/Plenum Publishers, New York, 2005

8. Chase MH, Yamuy J, Xi M-C, et al: The control of motoneurons during sleep. Sleep and Movement Disorders. Chokroverty S, Hening W, Walters A, Eds. Butterworth-Heinemann, Boston, 2003

9. Chase MH, Morales FR: Control of motoneurons during sleep. Principles and Practice of Sleep Medicine, 4th ed. Kryger MH, Roth T, Dement WC, Eds. WB Saunders Co, Philadelphia, 2005, p 154

10. Porkka-Heiskanen T, Strecker E, Thakkar M, et al: Adenosine: a mediator of the sleep-inducing effects of prolonged wakefulness. Science 276:1265, 1997

11. McCarley RW: Introductory essay: orexin, sleep and movement disorders. Sleep and Movement Disorders. Chokroverty S, Hening W, Walters R, Eds. Butterworth-Heinemann, Boston, 2003

12. De Lecea L, Kilduff TS, Peyron C, et al: The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. Proc Natl Acad Sci USA 95:322,1998

13. Sakurai T, Amemiya A, Ishii M, et al: Orexin and orexin receptors: a family of hypothalamic neuropeptides and G protein–coupled receptors that regulate feeding behavior. Cell 92:1, 1998

14. Chokroverty S: Physiological changes in sleep. Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects, 2nd ed. Chokroverty S, Ed. Butterworth-Heinemann, Boston, 1999, p 95

15. Somers VK, Dyken ME, Mark AL, et al: Sympathetic nerve activity during sleep in normal subjects. N Engl J Med 328:303, 1993

16. de Mairan JJD: Observation botanique. Histoire de l'Académie Royale des Sciences. Imprimerie Royale, Paris, 1731, p 35

17. The International Classification of Sleep Disorders, 2nd ed: Diagnostic and Coding Manual. American Academy of Sleep Medicine, Westchester, Illinois, 2005

18. National Center of Sleep Disorders Research. US Department of Health and Human Services, 1994
http://www.nhlbi.nih.gov/health/prof/sleep/sleep.txt

19. Carskadon MA, Dement W, Mitler MM, et al: Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 9:519, 1986

20. Allen RP: Actigraphy. Sleep and Movement Disorders. Chokroverty S, Hening W, Walters A, Eds. Butterworth-Heinemann, Boston, 2003

21. Flemons WW: Obstructive sleep apnea. N Engl J Med 347:498, 2002

22. Young T, Palta M, Dempsey J, et al: The occurrence of sleep-disordered breathing among middle-aged adults. N Engl J Med 328:1230, 1993

23. Yaggi HK, Concato J, Kernan WN, et al: Obstructive sleep apnea as a risk factor for stroke and death. N Engl J Med 353:2034, 2005

24. Sanders MH, Stiller RA: Positive airway pressure in the treatment of sleep-related breathing disorders. Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects, 2nd ed. Chokroverty S, Ed. Butterworth-Heinemann, Boston, 1999, p 355

25. Overeem S, Mignot E, van Dijk JG, et al: Narcolepsy: clinical features, new pathophysiologic insights and future perspectives. J Clin Neurophysiol 18:78, 2001

26. Mignot E: Genetic and familial aspects of narcolepsy. Neurology 50(suppl 1):S16, 1998

27. Lin L, Faraco J, Li R, et al: The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. Cell 98:365, 1999

28. Chemelli RM, Willie JT, Sinton CM, et al: Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. Cell 98:437, 1999

29. Nishino S, Ripley B, Overeem S, et al: Hypocretin (orexin) deficiency in human narcolepsy. Lancet 355:39, 2000

30. Nishino S, Ripley B, Overeem S, et al: Low cerebrospinal fluid hypocretin (orexin) and altered energy homeostasis in human narcolepsy. Ann Neurol 50:381, 2001

31. Scammell TE: The neurobiology, diagnosis and treatment of narcolepsy. Ann Neurol 53:154, 2003

32. Mignot E, Lammers GJ, Ripley B, et al: The role of cerebrospinal fluid hypocretin measurement in the diagnosis of narcolepsy and other hypersomnias. Arch Neurol 59:2, 2002

33. Allen RP, Earley CJ: Restless legs syndrome: a review of clinical and pathophysiologic features. J Clin Neurophysiol 18:128, 2001

34. Allen RP, Picchietti D, Hening WA, et al: Restless legs syndrome: diagnostic criteria, special considerations, and epidemiology. Sleep Med 4:101, 2003

35. Montplaisir J, Michaud M, Denesle R, et al: Periodic leg movements are not more prevalent in insomnia or hypersomnia but are specifically associated with sleep disorders involving a dopaminergic impairment. Sleep Med 1:163, 2000

36. Mahowald MW: Hope for the PLMS quagmire. Sleep Med 3:463, 2002

37. Reid KJ, Zee PC: Circadian rhythm disorders. Semin Neurol 24:315, 2004

38. Morin CM, Daley M, Ouellet MC: Insomnia in adults. Curr Treat Options Neurol 3:9, 2001

39. Morin CM, Colecchi C, Stone J, et al: Behavioral and pharmaceutical therapies for late-life insomnia: a randomized controlled trial. JAMA 281:991, 1999

40. Broughton RJ: Parasomnias. Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects, 2nd ed. Chokroverty S, Ed. Butterworth-Heinemann, Boston, 1999, p 635

41. Provini F, Plazzi G, Tinuper P, et al: Nocturnal frontal lobe epilepsy: a clinical and polygraphic overview of 100 consecutive cases. Brain 122:1017, 1999

42. Phillips HA, Scheffer IE, Crossland KM, et al: Autosomal dominant nocturnal frontal lobe epilepsy: genetic heterogeneity and evidence for a second locus at 15q24. Am J Hum Genet 63:1108, 1998

43. Schenck CH, Mahowald MW: REM sleep behavior disorder. Sleep and Movement Disorders. Chokroverty S, Hening W, Walters A, Eds. Butterworth-Heinemann, Boston, 2003

44. Boeve BF, Silber MH, Ferman TJ, et al: Association of REM sleep behavior disorder and neurodegenerative disease may reflect an underlying synucleinopathy. Mov Disord 16:622, 2001

45. Eisensehr I, Linke R, Noachtar S, et al: Reduced striatal dopamine transporters in idiopathic rapid eye movement sleep behavior disorders compared with Parkinson's disease and controls. Brain 123:1155, 2000

Acknowledgment

Figures 2 and 3 Seward Hung.

XIV PAIN

Alan Carver, m.d.

Definitions and Overview

The International Association for the Study of Pain defines pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage."[1] Although pain most often has a proximate physical cause, the current definition avoids tying pain directly to a stimulus, because pain of probable psychological origin may be reported in the absence of actual tissue damage or any other likely pathophysiologic cause.

Pain is not simply the unidirectional transmission of data along hardwired tracts from peripheral tissue to the central nervous system.[2] Indeed, pain is experienced within a complex biologic, emotional, psychological, and social context that may defy physical examination findings, diagnostic procedures, and laboratory tests.

Pain may be considered to consist of four broad components: nociception, perception, suffering, and behaviors.[3] Nociception is the detection of tissue damage by specialized receptors (nociceptors) on A-delta fibers and C fibers. This is a dynamic process in which the action of nociceptors is influenced by the local chemical environment and by neural changes set into motion by local tissue damage.[4] Perception of pain, although often triggered by local injury, may also result from lesions in the peripheral nervous system or the CNS; pain may thus be perceived in the absence of nociception. The intensity of pain may have little or no relation to the extent of objective pathology. Suffering may be defined as the state of severe distress associated with events that threaten the intactness of the person.[5] Suffering can include physical pain but is not limited to it. Pain behaviors are observable, quantifiable behaviors arising from pain and suffering; such behaviors include restricting activity, verbally complaining of pain, or seeking health care.

There are several different types of pain, with differing mechanisms, temporal courses, and management options. Transient pain results from the activation of nociceptors in the absence of tissue damage; it is not a clinically significant cause of health care utilization. Acute pain is marked by an extensive nociceptive and behavioral cascade triggered by local tissue damage. It is a normal physiologic response to adverse chemical, mechanical, or thermal stimuli. Although acute pain is not defined by time course, it will generally subside within weeks. Chronic pain most often occurs when an initial injury exceeds the body's capacity for healing or involves the nervous system itself. Pain is defined as chronic rather than acute not on the basis of its duration, per se, but on the basis of the body's inability to restore physiologic function to homeostatic levels. The intensity of chronic pain may bear little relation to initial tissue damage or subsequent quantifiable pathology; indeed, a number of chronic pain syndromes lack any identifiable tissue damage or trigger injury.[6]

The complexity of pain physiology may lead to suboptimal pain assessment and management, as occurs when patient complaints are not believed because of a lack of objective data or when clinicians lack an understanding of pain physiology. Surveys of physicians and medical students have repeatedly confirmed the need for improved education and training in pain recognition and treatment.[7] Provider surveys reveal that many practitioners do not feel comfortable with pain management. One large study found that 86% of physicians believed that most of their patients with pain were undertreated, whereas only 51% felt pain control in their practices was "good" or "very good."[8] A model of care that is oriented toward diseases rather than symptoms may incorrectly minimize the therapeutic importance of symptom management. In addition, attitudes of both physicians and patients toward opioids and the fear of patient addiction contribute to the inadequate treatment of pain associated with cancer and a variety of other chronic conditions.[9]

Epidemiology

Acute pain is the most common symptom for which patients seek medical evaluation.[10] New pain complaints result in 40 million visits to the doctor annually, and 45% of persons in the United States will visit a doctor for pain at some point in their lives.[11] The prevalence of various chronic pain syndromes in the United States is estimated to range from 2% to 40%.[12] Approximately 75 million persons in the United States live with "serious pain," and nearly 50 million are partially or totally disabled by pain.[13] In addition, recurrent acute pain is a prominent feature of a number of diseases, such as sickle cell anemia, AIDS, and malignancy. The prevalence of pain in cancer patients is estimated to range from 14% to 100%, with pain more frequent in advanced disease and certain types of malignancy.[14] Estimates of the prevalence of pain in AIDS patients vary from 30% to 90%; in one study, ambulatory AIDS patients experienced two to three concurrent pains.[15] The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT) found that approximately 50% of adults who die in hospitals experience moderate to severe pain in the period immediately before death.[16]

The financial costs of pain, especially untreated pain, to society are significant—approximately $100 billion annually in the United States.[17] The American Productivity Audit estimates that losses in productivity because of common painful conditions in active workers costs the economy $61.2 billion annually, with the majority of this cost related to reduced performance while at work rather than absence from the job.[18] One study found that 13% of the total workforce experienced a loss in productive time during a 2-week period because of a common pain condition. The cost of chronic back pain alone has been estimated to be $28 billion per year.[19] In addition, pain results in 50 million lost work days a year in the United States. Both acute and chronic pain are significant drivers of increased utilization of health care resources. Persons with chronic pain are five times as likely as those without chronic pain to use health care services.

Neurobiology of Pain

Peripheral Pain

Nociception

Nociception is the reception of signals in the CNS evoked by the activation of specialized sensory receptors (nociceptors) that provide information about tissue damage [*see Figure 1*]. Nociceptors are the free peripheral endings of primary sensory neurons;

the cell bodies of nociceptors are located in the dorsal root and trigeminal ganglia. Nociceptors are the least differentiated of cutaneous sensory receptors; unlike other somatic sensory receptors, they lack the structures for filtering peripheral stimuli.

Different classes of nociceptive afferent fibers react to different types of stimuli. A-delta fibers function as thermal and high-threshold mechanical receptors. Activation of these fibers generally results in short sensations of sharp, pricking pain. A-delta fibers are of small diameter and are thinly myelinated; they conduct impulses at the relatively fast rate of 5 to 30 m/sec. They terminate in laminae I and V of the dorsal horn. C fibers function as polymodal receptors and are activated by various high-intensity mechanical or chemical stimuli, as well as by hot (> 45° C) or cold stimuli. Activation of C fibers results in more prolonged sensations of dull pain. C fibers are small in diameter; they are unmyelinated and conduct impulses at the slow rate of 0.5 to 2 m/sec. C fibers have smaller receptive fields than the A-delta nociceptors. They terminate in lamina II. Polymodal C fibers compose a large majority of peripheral nociceptors. The cell bodies of both A-delta fibers and C fibers reside in the dorsal root ganglion. Some nociceptors, which are termed silent or sleeping, have such a high activation threshold that under normal circumstances they do not react to most stimuli; they are recruited under conditions of inflammation and tissue damage (see below). All classes of nociceptive fibers are widely distributed throughout cutaneous and deeper tissue.

Nociceptors do not fire spontaneously at rest. Their electrical action potential is triggered by transduction, which occurs when a noxious stimulus of sufficient strength depolarizes the nociceptor membrane. The specific receptive properties of nociceptors are determined by their expression of transducing ion-channel receptors. These ion channels are nonselective potassium or sodium channels gated by temperature, chemical stimuli, or mechanical shearing forces rather than by voltage. Activation of the channels by an appropriate stimulus leads to an inward current that depolarizes the receptor membrane. If this depolarizing current is sufficient to activate voltage-gated sodium channels, further depolarization of the membrane will occur, and a burst of action potentials will be initiated. The duration and frequency of this burst are determined by the duration and intensity of the noxious stimulus. Many, but not all, of these transducing receptors have been identified.

The typical noxious stimulus affects some combination of the different types of nociceptors in a given area of tissue damage. A-delta fibers provide sensory input for immediate, sharp pain, whereas C fibers are responsible for delayed, dull pain. The summation of input from both types of nociceptors provides the sensory basis for the perception of pain.

Peripheral Sensitization

The nociceptive process occurs in the context of a number of changes in the chemical environment brought about by local tissue damage and the release of a variety of inflammatory mediators. Some of these mediators act to enhance the sensation of pain in response to subsequent stimuli in the affected area by decreasing nociceptor stimulus thresholds and increasing and prolonging receptor firing and intensity in response to a suprathreshold stimulus, a process called peripheral sensitization.[20-22] Prostaglandin E_2, released from arachidonic acid–damaged cells, binds to G protein–coupled prostaglandin E receptors to sensitize nociceptors. Interleukin-1b and tumor necrosis factor–α, released by immune system cells involved in the inflammatory response, induce the release of cyclooxygenase-2 (COX-2) several hours after the start of inflammation. COX-2 converts arachidonic acid to prostaglandin H, which is subsequently converted into prostanoid species, including prostaglandin E_2. Bradykinin, a peptide derived from plasma kininogen, sensitizes primary afferent neurons via its constitutive B_2 receptor. Nerve growth factor sensitizes nociceptor terminals by binding to the G protein–coupled tyrosine kinase A receptor. Leukotrienes derived from arachidonic acid–damaged cells also alter the sensitivity of nociceptor terminals. Certain peptides released by primary afferent neurons themselves support the process of peripheral sensitization. Substance P, released from activated nociceptors, acts on local mast cells to increase histamine release.

Peripheral sensitization is thought to be one of the processes underlying the phenomenon of primary hyperalgesia, whereby injury to peripheral tissues results in an enhanced sensation of pain in response to subsequent suprathreshold stimuli in the damaged area. Experimental evidence supports the ability of heat stimuli to yield primary hyperalgesia; data concerning mechanical stimuli are not as clear. The evolutionary purpose of hyperalgesia is to discourage further contact with damaged tissue, thereby expediting the healing process. Peripheral sensitization and the consequent hyperalgesia normally resolve as tissue heals. However, chronic pain (see below) may occur if the body is unable to restore homeostasis because the initial injury has exceeded the body's capacity for recovery or the injury involves the nervous system itself. Secondary hyperalgesia—the expansion of hyperalgesia beyond the region of initial tissue damage in the absence of ongoing or recurrent injury—is caused by additional nervous system changes secondary to the initial insult [*see* Central Sensitization, *below*].

DORSAL HORN PHYSIOLOGY

Nociceptive fibers bifurcate upon entering the spinal cord laterally in the dorsal root, with branches of each axon ascending and descending for several segments in the tract of Lissauer. A-delta and C fibers terminate mainly in lamina I (marginal zone) and lamina II (substantia gelatinosa) of the dorsal horn [*see Figure 2*]. In addition, some A-delta fibers terminate in lamina V.

The primary afferent fibers communicate directly or indirectly with two major classes of neurons in the dorsal horn: projection neurons and local circuit interneurons. Projection neurons are contacted directly by A-delta fibers and relay sensory data to the brain stem, hypothalamus, and thalamus. They stretch the length of the spinal cord, with cell bodies in laminae I and V of the dorsal horn. Nociceptive-specific (NS) projection neurons relay data solely from A-delta and C fibers, whereas wide-dynamic-range (WDR) projection neurons relay both noxious data from nociceptors and a variety of innocuous stimuli from low-threshold mechanoreceptors. Both NS and WDR neurons are important in the relaying of nociceptive information; NS data signal the presence of possible tissue damage, whereas WDR data concern stimulus quality and, possibly, location. Local circuit interneurons modify the output of projection neurons. Excitatory interneurons relay sensory input from C fibers to projection neurons; inhibitory interneurons regulate and suppress the flow of nociceptive data to higher centers.

The relaying of nociceptive data in the dorsal horn involves several different neurotransmitters. Both A-delta and C fibers release glutamate, an excitatory amino acid that is the major fast excitatory neurotransmitter for all nociceptive modalities. Glutamate acts at two major classes of glutamate receptors on sec-

ondary afferent neurons: ionotropic and metabotrophic. Ionotropic glutamate receptors, which include the receptors for *N*-methyl-D-aspartate (NMDA), α-amino-3-hydroxy-5-methyl-4-isoxazoleproprionate (AMPA), and kainite, mediate fast synaptic transmission, the rapid depolarization of the secondary afferent neuron membrane. This depolarization will generate an action potential if the threshold is met. Metabotrophic glutamate receptors are coupled to a variety of intracellular second messenger systems via G proteins. These receptors are expressed presynaptically and postsynaptically in the dorsal horn and throughout the neuraxis, and they play a role in the modulation of pain. The presynaptic receptors on the central terminals of nociceptors may alter neurotransmitter release; they are a major target of analgesic therapies.

In addition to glutamate, nociceptive afferents (primarily C fibers) release a large variety of neuropeptides, including substance P, calcitonin gene–related peptide, cholecystokinin, somatostatin, galanin, gastrin-releasing peptide, and substance K. The neuropeptides act on G protein–coupled receptors and are responsible for evoking slow excitatory postsynaptic potentials in the secondary afferent neurons of the dorsal horn.

Initial modulation of the transmission of pain data from the

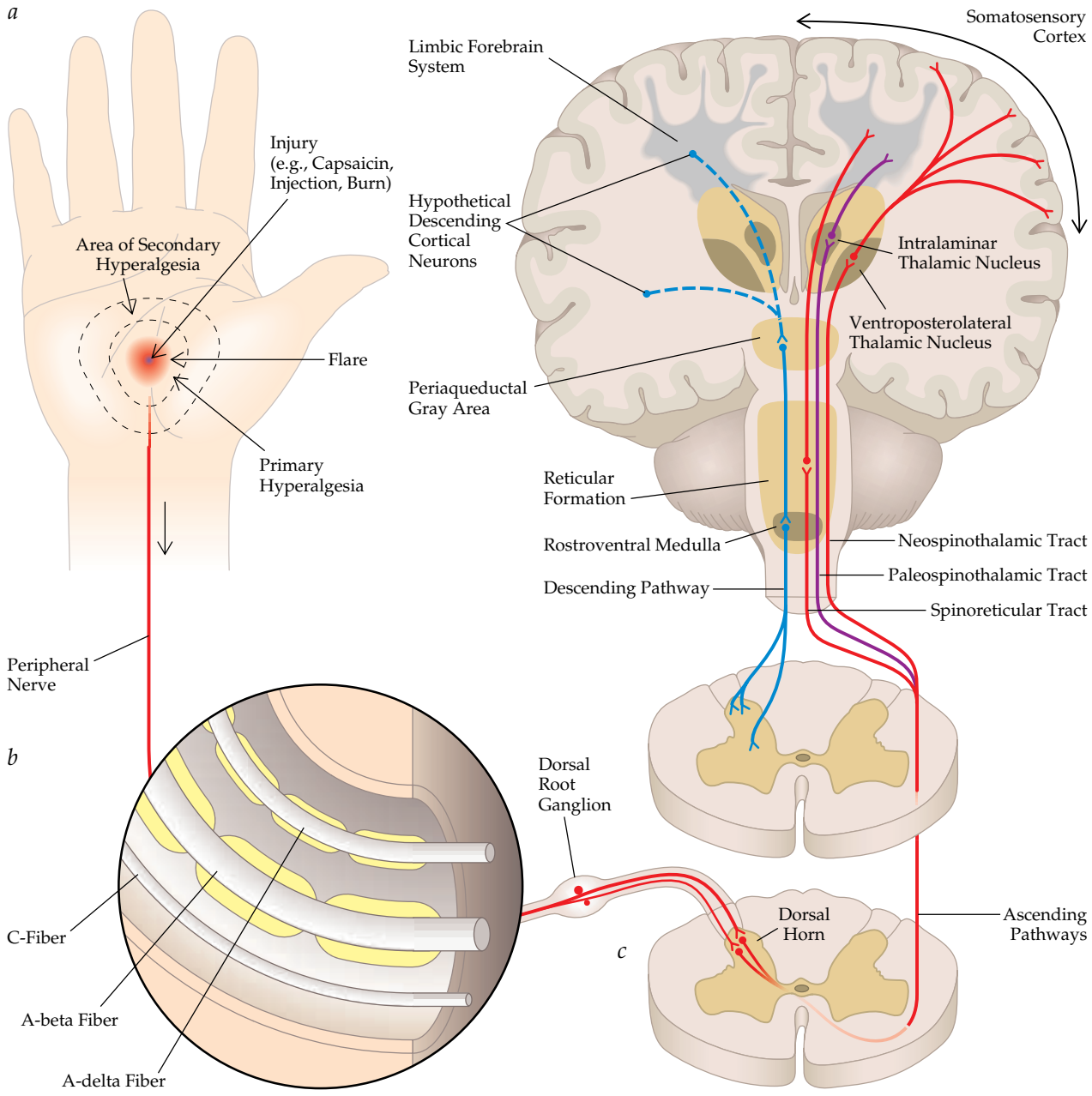


Figure 1 **(*a*) Stimulation of C fibers by intradermal injection of capsaicin or thermal injury produces spontaneous pain and pain evoked by light touch at the site of injury (primary hyperalgesia). In addition, an area of secondary hyperalgesia (increased pain after a noxious stimulus) outside of the area of primary injury is produced by the activation of *N*-methyl-D-aspartate (NMDA) receptors in the central nervous system, which occurs as a consequence of the afferent nociceptive impulse traffic. (*b*) When activated by mechanical, thermal, and chemical stimuli, nociceptors conduct afferent impulses toward the spinal cord. (*c*) When areas in the thalamus and cerebral cortex are activated, secondary projections in the spinothalamic tract, dorsal column tract, and other nociceptive pathways lead to the conscious perception of pain.**

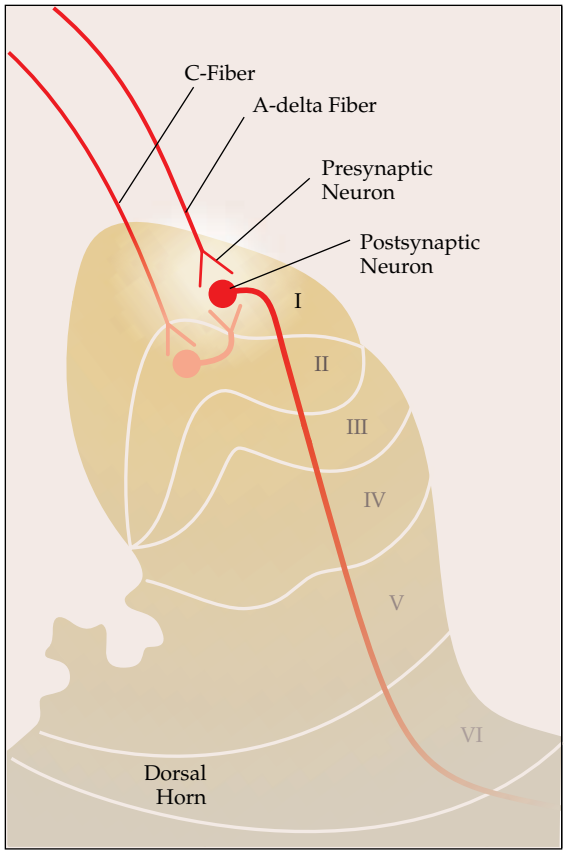


Figure 2 **Upon entering the spinal cord laterally in the dorsal root, A-delta fibers and C fibers terminate mainly in lamina I (marginal zone) and lamina II (substantia gelatinosa) of the dorsal horn. In addition, some A-delta fibers terminate in lamina V. Secondary afferent neurons of the dorsal horn are involved in determination of which action potentials trigger a reflex response to avoid further injury from noxious stimuli, as well as which action potentials are transmitted to the brain. A major role of the dorsal horn is to suppress pain. Most sensory information relayed to the secondary afferent neurons of the dorsal horn does not evoke an action potential.**

periphery to higher centers occurs in the dorsal horn. Secondary afferent neurons of the dorsal horn are involved in the determination of which action potentials trigger reflex action to avoid further injury from noxious stimuli, as well as which action potentials are transmitted to the brain. Inhibitory interneurons release the neurotransmitter γ-aminobutyric acid (GABA), which acts at receptors on secondary afferent neurons to produce inhibitory postsynaptic action potentials, modulating the transmission of excitatory action potentials received from primary afferents. A major role of the dorsal horn is to suppress pain. Most sensory information relayed to the secondary afferent neurons of the dorsal horn does not evoke an action potential.

GATE CONTROL THEORY

The gate control theory of pain, originally described by Melzack and Wall in 1965, posits that input from large-diameter primary afferent fibers in the peripheral nervous system activate inhibitory interneurons in the substantia gelatinosa (lamina II of the dorsal horn).[23] Activation of these inhibitory interneurons, in turn, reduces the effectiveness of nociceptive input from small-diameter primary afferent fibers in activating projection neurons of the spinothalamic, spinoreticular, or spinomesencephalic tracts. Thus, large-diameter input "controls the gate" by which small-diameter input transmits data on noxious stimuli to higher pain centers. This concept is the theoretical basis for the inhibition of pain by rubbing or vibration, as well as therapeutic modalities such as transcutaneous electrical nerve stimulation (TENS).[24]

CENTRAL SENSITIZATION

Central sensitization is a pathologic state in which dorsal horn excitability is increased and gatekeeping function is lost.[25-27] This facilitated responsiveness to sensory input leads to primary hyperalgesia (as is seen in peripheral sensitization). It also leads to secondary hyperalgesia, in which an exaggerated response to noxious stimuli is observed beyond the region of actual tissue damage. Allodynia, in which pain is generated by a low-intensity stimulus that would not be noxious to normal tissue, is also seen.

The pathogenesis of central sensitization may be thought of as occurring in two phases. The first stage is triggered by intense nociceptive input to the secondary afferent neurons of the dorsal horn and may thus be considered activity dependent. This excessive input may arise from persistent acute injury, surgical insult, peripheral sensitization of nociceptors during inflammation, ectopic discharge from nerve injury, a variety of chronic pain syndromes, or certain other conditions. The activity-dependent stage of central sensitization occurs rapidly; hyperresponsiveness of spinal neurons may be seen within seconds of massive sensory input triggered by an appropriate initial insult.

NMDA receptors on secondary afferent neurons of the dorsal horn are of critical importance in the pathology of central sensitization. Under normal conditions, glutamate does not have activity at the NMDA receptor during nociception. This is because the receptor has a voltage-dependent magnesium ion block on its calcium channel; glutamate is rapidly cleared from the synaptic cleft, and the neuron is not depolarized long enough to allow the magnesium ion block to be dislodged. However, excessive frequency of nociceptive input—primarily, temporally summated slow excitatory postsynaptic potentials from C fibers—lead to a persistent net increase in the amount of glutamate in the synaptic cleft and depolarization of secondary afferent neurons of sufficient duration to dislodge the magnesium ion block. The resultant activation of NMDA receptors by glutamate leads to a number of signaling cascades initiated by increased intracellular calcium influx, including G protein–coupled neurokinin receptors and receptor tyrosine kinases, as well as phosphokinases.[28] Activation of protein kinases leads to phosphorylation of the NMDA receptor ion channels, decreasing the magnesium block at resting membrane potentials and prolonging channel opening time. As a result, sensitized secondary afferent neurons respond more easily to continued signals from C fibers; in addition, spontaneous or ectopic signal discharge may occur in the context of loss of gatekeeping function.

Central sensitization is mediated by both postsynaptic (see above) and presynaptic NMDA receptors in the dorsal horn. Many small-diameter primary afferent fibers synapsing in the dorsal horn express NMDA receptors. Glutamate released from the presynaptic terminal may use these receptors to enhance its own release in a feed-forward mechanism in response to subsequent stimuli. Presynaptic NMDA receptors can facilitate the transmission of nociceptive messages secondary to the release of substance P, calcitonin gene–related peptide, and glutamate from primary afferent terminals.

The second stage of central sensitization is sustained beyond the initial insult by transcriptional changes in the cell and may be

considered transcription dependent. Two basic types of transcriptional change occur. One type is restricted to parts of the nervous system receiving sensory input from injured tissue; this process is activity driven. The other type of transcriptional change is widespread and produces a variety of functional effects. Induction of COX-2 in the central nervous system after peripheral inflammation is an example of this process; widespread production of COX-2 leads to a diffuse increase in prostaglandin E_2, with consequent diffusely increased neuronal excitability.

ASCENDING PATHWAYS

Nociceptive input to the dorsal horn is relayed to higher centers via one of a number of ascending pathways.

Spinothalamic Tract

The spinothalamic tract is the most prominent ascending pathway for nociceptive neural transmission. Axons from WDR and NS neurons in laminae I and V decussate diagonally in the anterior white commissure before ascending in the contralateral anterolateral white matter, becoming the ascending projections of the neospinothalamic and paleospinothalamic tracts. As their names imply, the neospinothalamic tract is a phylogenetically younger system than the paleospinothalamic pathway. The neospinothalamic tract relays information concerning sharp pain that may be graded in intensity and localized to a specific region of the body. Axons of this tract terminate monosynaptically in the ventroposterolateral (VPL) nucleus of the thalamus. As such, information traveling this pathway is not modulated before arriving at the thalamus.

The majority of spinothalamic tract axons belong to the paleospinothalamic pathway. This pathway produces diffuse, aching pain that is difficult to localize but lets the body know that it must take action to avoid further injury. The paleospinothalamic tract sends multiple projections to the brain stem reticular formation in the medulla and pons and the medial thalamus before terminating in the VPL nucleus. A series of nuclei in the medulla and pons act to modify ascending nociceptive data in the paleospinothalamic tract. For instance, the locus ceruleus nucleus of the pons produces glutamate and serotonin precursors in response to fear or shock; these act to reduce the transmission of pain beyond the pons in acutely fear-producing situations (a mechanism of obvious survival value). The brain stem reticular formation also has extensive connections with a number of structures in the midbrain, including the periaqueductal gray matter, deep layers of the superior colliculus, the red nucleus, anterior and posterior pretectal nuclei, the nucleus of Darkschewitsch, and the interstitial nucleus of Cajal. These structures may act in conjunction with the reticular formation to further modulate data in the paleospinothalamic pathway.

Spinoreticular Tract

The spinoreticular tract is composed of axons of nociceptive neurons in laminae VII and VIII that ascend in the ipsilateral anterolateral quadrant of the spinal cord, joining the spinothalamic tract in the medial lemniscus in the brain stem before terminating at many sites throughout the medullary, pontine, and mesencephalic reticular formation. In addition, some axons send branches that terminate in both the reticular formation and the thalamus. Of note, some spinoreticular fibers do not cross the midline before terminating. Neurons of the reticular formation project, directly or indirectly, to many areas of the brain, including the thalamus, hypothalamus, limbic system, and neocortex. Nociceptive data relayed to the reticular formation by the spinoreticular tract contribute to the affective, motivational, and aversive response components of pain. This tract predates the more direct spinothalamic pathway in vertebrate evolution.

Spinomesencephalic Tract

This tract consists of nociceptive neurons originating in lamina I and laminae IV through VI that project to the mesencephalic reticular formation, the lateral portion of the periaqueductal gray region, the nucleus cuneiformis, the superior colliculus, and other sites in the midbrain. The periaqueductal region, in turn, has reciprocal connections with the limbic system via the hypothalamus.

The spinomesencephalic tract is involved in the production of fear and related affective and aversive behaviors associated with pain. The tract's connections may also initiate orienting responses to noxious stimuli. Spinomesencephalic communication with the periaqueductal gray matter activates the system of descending pain modulation that subsequently produces endogenous analgesia.

Spinocervicothalamic Tract

Although most neurons in laminae III and IV of the dorsal horn respond solely to tactile stimuli, a minority respond to noxious stimuli as well. Spinocervicothalamic neurons project through the ipsilateral dorsolateral spinal cord to the lateral cervical nucleus in segments C1 and C2. Axons from this nucleus then decussate, ascending in the medial lemniscus in the brain stem to midbrain nuclei and to the ventroposterior lateral and posterior medial nuclei of the thalamus.

Postsynaptic Dorsal Column Fibers

Some nociceptive neurons in laminae III and IV project their axons in the dorsal column of the spinal cord to the cuneate and gracile nuclei in the medulla. Internal arcuate fibers then connect these nuclei with the contralateral VPL nucleus of the thalamus. Transmission of nociceptive data through the dorsal column in this fashion may be responsible for the persistence of painful symptoms after selective ablation of the spinothalamic tract (i.e., cordotomy). Stimulation of these fibers is a target for pain relief.

Visceral Nociceptive Tracts in Dorsal Columns

The dorsal columns of the spinal cord are primarily associated with discriminatory touch and proprioception; classically, they were not thought to be involved in pain transmission. However, a number of studies indicate that the dorsal columns have a role in the processing of visceral pain. In addition to relaying viscerosensory information, this pathway may also facilitate spinal neuronal sensitization of visceral origin; creation of dorsal column lesions reduced such sensitization in a model of chronic visceral pain. In a number of recent clinical studies, the creation of lesions in the dorsal columns resulted in good relief of visceral pain in cancer patients.

Spinohypothalamic Tract

The spinohypothalamic tract consists of a large number of neurons widely distributed throughout the spinal cord that communicate with the hypothalamus without synapse in the reticular formation. In addition, widespread collateral branching from spinohypothalamic tract neurons to a variety of targets in the medulla, pons, midbrain, and thalamus has been described. This pathway is involved in autonomic, neuroendocrine, and affective/emotional responses to noxious and other somatosensory

stimuli of cutaneous or visceral origin. Some studies suggest that the majority of spinohypothalamic tract neurons respond exclusively or preferentially to noxious stimuli.

Thalamocortical Pathways

Thalamus The ascending pathways project directly or indirectly to a number of nuclei in both the lateral and medial thalamus. In the lateral thalamus, the VPL nucleus is a main site of information relay from ascending pathways to the cortex. VPL neurons are primarily involved in sensory-discriminative aspects of pain perception, such as localization; however, the VPL may also be involved in visceral and referred pain. Neurons of the VPL nucleus, with receptive fields that are restricted in size, respond maximally to noxious mechanical stimuli; they also respond to noxious heat and C-fiber volleys. In addition, they respond weakly to nonnoxious mechanical stimuli. Most VPL neurons respond to both cutaneous and visceral stimuli; cutaneous input is somatotopic, but visceral input is not similarly arranged. Almost all VPL neurons project to the primary somatosensory cortex.

In the medial thalamus, the spinothalamic tract projects to the central lateral nucleus and other intralaminar nuclei, as well as to the ventromedial preoptic nucleus (VMpo), whereas the spinomesencephalic and spinoreticular tracts project heavily to the intralaminar nuclei and other medial thalamic nuclei via the reticular formation. The medial thalamic nuclei appear to play their most significant role in processing the affective and motivational components of pain. Primate studies suggest that neurons of the intralaminar nuclei have large, bilateral receptive fields (presumably limiting their role in sensory discrimination, or at least in localization). VMpo neurons, which have small, somatotopically organized receptive fields, project to the insula, a fact that suggests they have a role in affective and motivational responses to pain, as well as memory processing.

Cortex There is still a great deal of uncertainty concerning cortical processing of pain. There is no single discrete cortical pain center. Indeed, imaging studies support activation of a number of cortical areas in response to painful stimuli, including the anterior cingular, insular, parietal, frontal, premotor, and primary and secondary somatosensory cortices.[29] Patterns of cortical activation seen in experimentally induced pain may not match those seen in chronic pain, however.[30] Positron emission tomography studies have associated experimentally induced pain with increased regional cerebral blood flow to the bilateral anterior insular and posterior parietal cortices, bilateral inferior lateral prefrontal cortices, and cerebellar vermis and decreased blood flow to the contralateral posterior thalamus.[29]

The preponderance of imaging studies that associate activity in the anterior cingulate cortex (ACC) with both pain and behavioral drive and volition support a primary role for this region in the cortical processing of the motivational and affective components of pain. The critical role of the ACC in processing nociceptive data has raised interest in this cortex as a target of neuroablative therapy for intractable pain.[31] The anterior cingulate cortex, a major destination of outflow from the medial thalamus, is selectively activated by painful thermal stimuli and appears to be important in the processing of affective components of the pain experience.

Lesion studies suggest that large areas of damage to the somatosensory cortex do not yield impaired response to noxious stimuli. Cortical stimulation in general is only infrequently associated with the production of pain; large lesions of the somatosensory cortex may produce minimal or no pain, whereas small lesions elsewhere in the cortex may be associated with spontaneous pain or increased pain perception.[32]

DESCENDING PATHWAYS

The perception of pain at higher levels of the CNS triggers activation of descending pain suppression pathways involving numerous cortical and subcortical structures, as well as a variety of neurotransmitters.[33,34] The existence of such descending pathways was first suggested by the finding that brain stem stimulation inhibited nociceptive neurons in the dorsal horn of the spinal cord, but such suppression was negated by lesions of the dorsolateral funiculus. Neurons in the midbrain periaqueductal gray matter (PAG) and periventricular gray matter make excitatory connections to the nucleus raphe magnus (NRM) and the adjacent nucleus reticularis paragigantocellularis in the rostroventral medulla, as well as the locus ceruleus, among other nuclei. Neurons from these regions, especially the NRM and adjacent reticular formation, then make inhibitory connections in laminae I, II, and V of the dorsal horn. Thus, stimulation of PAG and subsequent stimulation of rostroventral medullary neurons lead to inhibition of dorsal horn neurons that transmit nociceptive data. Neurotransmitters involved in this pathway include endogenous opiates, serotonin, and norepinephrine. Noradrenergic projections from the dorsolateral pons to the dorsal horn of the spinal cord, relayed through the raphe magnus and pallidus, PAG, and ventrolateral reticular formation, constitute another significant antinociceptive pathway. Studies of stimulation-invoked analgesia also support roles for neurons of the reticular formation and anterior pretectal nucleus in descending antinociception. Stimulation of the VPL or ventroposteromedial thalamic nuclei leads to reduction in severity of several painful syndromes. Inhibition of spinothalamic tract neurons secondary to thalamic stimulation may result from antidromic activation of axons of spinothalamic tract neurons sending collaterals to brain stem nuclei such as the PAG. As described, sensory and descending neurons and interneurons converge in the dorsal horn of the spinal cord. These local circuits play a significant role in modulating descending pathway activity.

Endogenous opiates include enkephalins, dynorphins, and β-endorphins. Enkephalins and dynorphins are found in the PAG and rostroventral medulla, as well as in the dorsal horn of the spinal cord (especially laminae I and II). β-Endorphins are found in hypothalamic neurons that project to the PAG and noradrenergic nuclei in the brain stem.

Three major classes of opiate receptors are found in the descending pain suppression pathways and are widely distributed throughout the CNS. Enkephalins are active at the mu and delta receptors, whereas dynorphin is active at the kappa receptor. The primary classes of opiate receptors may be further subdivided into distinct subtypes. The different receptor classes may modulate the activity of different types of nociceptive inputs.

Opiate alkaloids such as morphine are potent agonists of the mu receptor, found in high concentration in the PAG and the superficial dorsal horn of the spinal cord. Opiates act at supraspinal levels (i.e., the raphe nuclei and PAG) to suppress GABA-releasing interneurons that normally inhibit descending pathway activity. This disinhibitory mechanism thus serves to activate the descending pain suppression pathways. Opiates also exhibit analgesic activity at the level of the primary afferent synapse in the dorsal horn of the spinal cord. The superficial dorsal horn has a large number of enkephalin- and dynorphin-con-

taining interneurons. Mu opiate receptors are located on the terminals of nociceptive afferents and the dendrites of postsynaptic dorsal horn neurons.

Opiate alkaloids and endogenous opioid peptides act presynaptically to suppress neurotransmitter release from sensory neurons. They also act postsynaptically to suppress the activity of the nociceptive dorsal horn neurons. That opioid receptors are ubiquitous throughout the CNS suggests that modulation of nociceptive data may occur at multiple sites, in addition to the aforementioned descending pathways and dorsal horn.

Neurobiology of Specific Pain Types

CANCER PAIN

Approximately 70% to 90% of patients with advanced cancer have chronic pain; pain is directly related to the presence of primary or metastatic cancer in up to three quarters of these patients.[35] Although cancer pain may involve any tissue, certain pain syndromes are commonly seen in the setting of malignant disease. Common mechanisms for cancer-related pain include direct effects of tumor growth, visceral involvement, bony metastases, soft tissue invasion, and nerve infiltration or compression. Exposition of the full range of cancer-related pain is beyond the scope of this chapter; some of the more common or significant syndromes are discussed.

Bone Pain

Tumor involvement of bone is the most frequent cause of cancer pain. Multiple myeloma and metastatic disease of the breast, lung, thyroid, and prostate are the cancers that most commonly involve bone. Different tumor types may produce primarily lytic or blastic metastases, but the majority result in mixed lesions. Osteoclast-mediated destruction of bone appears to be an important mechanism of pain in malignant disease. Tumors may also cause bone pain via ischemia, pressure, or secretion of chemical mediators. Pathologic fracture secondary to significant infiltration of bone may lead to an acute increase in pain severity, as well as neural impingement or compression. Metastasis to the vertebral column (see below) may lead to a wide variety of pain symptoms. Although the clinical experience of pain will vary with the mechanisms of the pain, most cancer-related bone pain is described as focal and constant, increasing in severity over time with progression of disease.

Plexopathy

A number of plexopathies are seen in cancer patients. Brachial plexopathy secondary either to metastasis or to tumor spread from lymphoma or from lung, breast, or thyroid cancer is a common cause of rapidly progressive arm or shoulder pain in cancer patients. Horner syndrome may also be seen in this setting. Brachial plexopathy may also result from radiotherapy (see below). Cervical plexopathy results from malignant invasion of the cervical plexus by head and neck tumors and produces pain in the ear or neck. Lumbosacral plexopathy may produce some combination of local, referred, or radicular pain in the low back, abdomen, perineum, or leg.

Back Pain

The prevalence of back pain in cancer is not known, but one large series found that it was the most common neurologic symptom in patients with a history of systemic cancer.[36] Back pain in this setting is typically from bone or epidural metastases; up to one third of cancer patients may develop metastases to the spine, with the thoracic vertebral column being the most frequent site of bony metastases. Acutely worsening back pain or increasing neurologic symptoms may indicate involvement of the spinal cord or one or more nerve roots. Any patient with a history of cancer who complains of back pain must be evaluated emergently to rule out epidural spinal cord compression [*see 12:XII Oncologic Emergencies*]. Less commonly, back pain may be caused by retroperitoneal tumors.

Headache

Estimates of the prevalence of headache in patients with brain tumors vary widely, from 8% to 71%.[37] Infratentorial and intraventricular tumors, multiple metastases, and leptomeningeal carcinomatosis are most commonly associated with cancer-related headache. Headache in the setting of brain tumor is generally from traction on and displacement of intracranial pain-sensitive structures, such as cranial and cervical nerves, blood vessels, and dura. Pain may also arise from inflammation or direct infiltration of those structures. It may also occur as the result of disturbance or obstruction of the flow of cerebrospinal fluid; sagittal sinus occlusion; stroke; or other causes. Cancer-related headache pain is variable in presentation but is most often described as being of mild or moderate intensity, lasting for hours, possibly developing over weeks or months, and possibly associated with nausea, vomiting, mental status changes, or focal neurologic findings. Forsyth and colleagues reported that patients with a history of headache who develop brain tumors are more likely to complain of headache as their primary presenting tumor symptom than patients without a history of headache.[38] These data underscore the critical importance of a comprehensive neurologic assessment in all patients presenting with headache.

Abdominal Pain

Primary or metastatic intra-abdominal malignancy may cause poorly localized visceral pain, sometimes with associated nausea, vomiting, or systemic symptoms. Causes of pain include direct invasion of tissue, peritoneal inflammation, and vascular or lymphatic obstruction. Pain secondary to abdominal cancers may be referred elsewhere, as when diaphragmatic irritation results in shoulder pain. Alternatively, pain from spinal malignancy may be referred to the abdomen.

Iatrogenic Pain

Many of the diagnostic and therapeutic modalities utilized in the care of cancer patients can result in pain.[35] Radiotherapy is associated with such painful conditions as osteonecrosis, acute and chronic enteritis, mucositis, and brachial plexopathy. Chemotherapeutic agents may produce painful neuropathies, as well as a variety of distressing side effects. Steroid therapy may lead to avascular necrosis, as well as to ultimately painful consequences of immunosuppression. Preparation for bone marrow or stem cell transplantation leads to mucositis in 70% of patients. Pain after cancer surgeries may be severe. In addition, chronic pain syndromes have been identified after mastectomy, thoracotomy, amputation, and radical neck dissection. Surgery for abdominal malignancies may lead to adhesions and subsequent painful bowel obstruction. Common hospital procedures may also be a significant source of pain and distress.

HIV INFECTION AND AIDS PAIN

HIV infection and AIDS can lead to pain through numerous distinct pathophysiologic mechanisms. Pain in this setting may be attributable to the direct neurotoxic effects of HIV on central or peripheral nerves, opportunistic infection or malignancy secondary to immunosuppression, or medication side effects. In one survey, 67% of AIDS patients reported pain in the previous 4 weeks[39]; many pain syndromes in ambulatory AIDS patients are directly related to infection or other consequences of immunosuppression. Most underlying causes of pain in AIDS are treatable.

HIV-associated peripheral neuropathy is the most frequent neurologic complication of AIDS, affecting almost one third of patients with the disease.[40] Distal symmetrical polyneuropathy (DSP), the most common form, is characterized by paresthesias, dysesthesias, or numbness occurring in a stocking-and-glove distribution. Distal pain, often described as sharp, stabbing, or burning, may become excruciating. DSP is seen primarily in advanced disease or secondary to antiretroviral treatment. Clinically, DSP caused by HIV infection is indistinguishable from DSP caused by antiretroviral agents, except for the latter's response to reduction or withdrawal of antiretroviral agents. Acute inflammatory demyelinating polyneuropathy (AIDP), characterized by muscle weakness, tingling, and ascending paralysis, may be the initial manifestation of HIV infection. Chronic inflammatory demyelinating polyneuropathy, a chronic form of AIDP marked by reduced reflexes and patchy weakness or numbness, may develop at seroconversion or a later stage of disease. Progressive polyneuropathy is characterized by rapidly progressive flaccid paraparesis, subacute low back and radicular pain and paresthesias, areflexia, and sphincter dysfunction. Most often, it is secondary to cytomegalovirus infection; it is observed most frequently in advanced HIV disease. Mononeuropathy multiplex occurring in early HIV disease is marked by self-limited sensory and motor deficits in the distribution of individual peripheral nerves; in advanced disease, multiple nerves in multiple extremities or cranial nerves may be involved. In mononeuropathy multiplex secondary to vasculitis, pain will precede sensory and motor deficits. Diffuse infiltrative lymphocytosis syndrome (DILS) is a subacute, often painful, distal sensorimotor neuropathy associated with parotid enlargement and sicca syndrome; there is sometimes systemic involvement (i.e., lymphadenopathy, splenomegaly, or interstitial pneumonia). DILS may occur during symptomatic HIV infection.

Pain Management

BARRIERS TO EFFECTIVE PAIN MANAGEMENT

Undertreatment of pain remains a significant clinical problem. SUPPORT found that 50% of adults dying in the hospital experienced moderate to severe pain in the period immediately before death.[16] In one large multicenter study, 86% of physicians felt that the majority of their patients with pain were undertreated.[8] The preponderance of studies suggest that most cancer patients do not receive adequate pain relief.

A number of barriers to effective pain management relate to health care delivery systems. Most health care organizations have not implemented procedures to ensure accountability for adequate assessment and treatment of pain. Many organizations fail to make pain relief a visible priority through such measures as adoption of standardized protocols or provision of sufficient time for training. Fragmentation of patient care across multiple clinical settings increases the likelihood of poor care. Use of gatekeepers and an emphasis on cost control by managed care organizations provide further systemic disincentives for the treatment of pain. Restrictive laws governing the use of controlled substances may have a significant effect on analgesic prescribing patterns. Clinician pain management practices are likely influenced by concerns about licensing, state disciplinary actions, and legal penalties arising from opioid prescribing patterns.

Table 1 Principles of Pain Assessment

1. Believe the patient's complaint of pain
2. Take a careful pain history
   - Onset
   - Temporality
   - Location
   - Quality (sensory versus affective)
   - Intensity
   - Aggravating/relieving factors
   - Associated symptoms
3. List and prioritize each pain complaint
4. Evaluate the response to current and previous treatment approaches
5. Evaluate the patient's psychological state
6. Assess patient function
7. Ask about a history of alcohol or drug dependence
8. Consider relevant past medical history
9. Perform a comprehensive medical and neurologic examination
10. Order and review appropriate diagnostic procedures
11. Make relevant referrals for multidisciplinary assessment
12. Assess patient expectations and goals

Several barriers center on physician attitudes or practices. A paucity of time devoted to the subject in the curricula of most United States medical schools and residencies has led to deficiencies in the pain management skills of many clinicians. In one study of physician attitudes regarding cancer pain, inadequate assessment skills were identified most frequently as a significant impediment to effective pain relief.[8] In another study, 70% to 80% of neurologists did not feel adequately trained to diagnose or treat patients with pain.[41] Some physicians are reluctant to prescribe opioids in any setting other than end-stage terminal illness. Clinician discomfort with opioid use is multifactorial; sources of unease include inadequate knowledge of opioid management, fear of patient addiction, and fear of professional censure or legal penalties for excessive use of these analgesics. Some clinicians do not view pain relief as central to the treatment of disease, or they may have low expectations regarding adequate pain relief. Physicians may not accept patient self-reports of pain as reliable.

Other barriers relate to patients' attitudes. Patients may be reluctant to report pain for any of a number of reasons. They may fear appearing "difficult" or "problematic," or they may have low expectations of obtaining relief. Alternatively, they may deny pain because of fear of its potential implications for their underlying illness. Patients may be reluctant to take prescribed opioids because of concerns about addiction or drug side effects. Patients with chronic pain may cease seeking medical treatment out of frustration with previous unsuccessful attempts at pain relief. In a survey of parents of children who had died of cancer, the parents reported overall satisfaction in the care their children received despite noting that pain was inadequately controlled.[42]

Financial and cultural barriers may also prevent effective pain management in some populations.

PRINCIPLES OF PAIN ASSESSMENT

Accurate pain assessment is the necessary precursor to effective pain management. At least one large multicenter study named inadequate pain assessment skills as the most frequent barrier to the provision of effective pain care.[8] Adherence to the following principles of pain assessment will greatly improve the treatment of pain [*see Table 1*].[43]

- Believe the patient's complaint of pain. Accepting that a patient's complaint is real and therefore deserving of further investigation is the logical prerequisite to further effective pain assessment. The patient's self-report is the single most accurate indicator of the existence and severity of pain. Fear of drug-seeking may predispose some physicians to disbelieve the pain complaints of some patients. However, lack of trust is associated with poor pain management and corresponding poor outcomes.
- Take a careful pain history. As with the medical workup of any complaint, a thorough history of the present illness is necessary. Specific areas to evaluate include the following.

Pain onset When did the pain begin? Was the onset spontaneous or gradual? Was there a possible precipitating event (e.g., trauma or surgery)? If so, the physician should obtain relevant details of the precipitating event (e.g., the nature and extent of trauma, any treatment at time of trauma). Did the patient ever experience similar pain?

Temporality What is the duration of the pain? Is it chronic, episodic, or intermittent? Is it acute or subacute? Is breakthrough pain experienced? Is it incidental?

Location Where is the pain? Is it in one or multiple locations? Does it radiate? Does the location of pain vary?

Quality What does the pain feel like? What word or words are used to describe the pain (e.g., sharp, dull, burning)? Adjectives chosen often fall into two broad categories: sensory, in which pain is described in essentially physical terms; and affective, in which the pain experience may have a significant emotional component. Particular words chosen may also provide important clues to the etiology of pain (e.g., neuropathic versus musculoskeletal).

Intensity Pain may be measured by use of a variety of scientifically validated assessment tools.[44] The Numeric Pain Intensity Scale, in which the patient rates his or her pain on a scale from zero to ten (0 = no pain, 10 = worst pain imaginable), allows for quantification of pain severity, evolution over time, and response to treatment. Use of the Visual Analog Scale or the Simple Descriptive Pain Intensity Scale may be better for some patient populations. In the Visual Analog Scale, a mark on a horizontal line denotes pain intensity; in the Simple Descriptive Pain Intensity Scale, the patient picks an appropriate verbal pain descriptor. The Faces Pain Scale for Adults and Children and the Wong-Baker Faces Rating Scale consist of multiple faces with various expressions; the patient may select the facial expression most consistent with his or her current level of pain. This methodology has been validated for use in the pediatric population. Several more extensive tools for assessing pain and its impact on the activities of daily living, such as the Initial Pain Assessment Tool, Brief Pain Inventory, McGill Pain Questionnaire, Memorial Pain Assessment Card, Neuropathic Pain Scale, and Memorial Symptom Assessment Scale [*see CE:X Symptom Management in Palliative Medicine*], permit valid serial assessment of pain and its functional import. For general clinical practice, however, simple assessment tools are sufficient.

Aggravating or relieving factors What exacerbates the pain? What relieves it?

Associated symptoms Is the patient experiencing any other symptoms, such as nausea, vomiting, diarrhea, or anorexia? Such symptoms may relate to the pain itself or to an underlying disease process.

- List and prioritize each pain complaint. Prioritization of multiple pain complaints should be based primarily on the patient's subjective ranking of significance.
- Evaluate the response to current and previous treatment approaches. What is being done for pain relief at present? What pharmacologic or other therapeutic approaches have been attempted in the past? Were adequate trials of medications given? Were these approaches successful? Has the patient experienced adverse events associated with pain management modalities?
- Evaluate the patient's psychological state. Consideration of psychological and psychosocial factors should start at the time of initial pain assessment and continue throughout treatment. Are there any recent sources of stress (e.g., marital or vocational problems)? Has the pain affected the patient's mood? Is the patient suffering from any psychiatric illness? Is this illness treated or untreated? Certain psychiatric conditions, such as depression, anxiety, and posttraumatic stress disorder, frequently coexist with chronic pain and may have significant somatic components. Clinical depression may have a profound influence on the patient's perception of pain and on the patient's ability to comply with subsequent treatment. Thus, simultaneous treatment of psychiatric pathology may be an integral part of the pain management plan. Depression and anxiety may lead to somatic complaints, and untreated pain may exacerbate depression and anxiety. Instruments such as the Medical Outcomes Study 36-Item Short Form (SF-36) and the Sickness Impact Profile may be used to assess the broader impact of pain on quality of life, including social, psychological, and spiritual well-being; these instruments are available through the Medical Outcomes Trust (http://www.outcomes-trust.org/instruments.htm). The patient's mental status should also be assessed. Cognitively impaired patients can sometimes give reliable reports of pain, but other assessment tools may be needed. Obtaining a history from family members, friends, or both is critical in caring for such patients.
- Assess patient function. How has the pain affected work, family, or social responsibilities or relationships? Have activities of daily living been impaired? What (if any) important activities is the patient unable to perform? Is the patient eating and sleeping as he or she did before the pain? In the case of chronic pain, what support structures does the patient have? The Wisconsin Brief Pain Questionnaire, or Brief Pain Inventory (http://www.stat.washington.edu/TALARIA/attachb1.html), is one of the instruments that may be used to evaluate the effect of pain on various domains of function.
- Evaluate for a history of alcohol or drug dependence. Previous or current substance abuse or dependence may impact

the patient's compliance with analgesic regimens, psychological state, function, and other domains. Ensuring the safety of the patient in the community when considering prescribing potent analgesic medication is of paramount importance [*see 13:VI Drug Abuse and Dependence*].

- Consider relevant past medical history. Many acute and chronic conditions may present as or eventually lead to pain; these include diabetes, coronary artery disease, malignancy, and the various chronic pain syndromes. A comprehensive medical history may provide important cues for treatment.
- Perform a comprehensive medical and neurologic examination. Performance of a physical examination will, at the very least, reassure the patient that his or her complaint is being taken seriously. The examination may also reveal the pathologic basis of pain. Often, the revealed disease process is unsuspected; in one study, pain consultants discovered a previously undiagnosed cause of pain in 64% of patients seen.[45] The examiner should look for any pain behaviors (e.g., abnormal gait or posture, guarding) and interpret their presence or absence in terms of the patient's overall presentation and social interaction. Observation for signs of systemic illness should accompany a detailed evaluation of reported pain sites. Every patient with a complaint of pain should, at minimum, undergo a neurologic examination. Chronic pain is often a sequela of neurologic pathology, and such patients are best cared for within a neurologic framework.
- Order and review appropriate diagnostic procedures. Although the diagnosis of painful conditions is usually made on the basis of the history and physical examination, and pain per se is not identifiable by any diagnostic modality, imaging and other studies such as laboratory tests, electromyelography, nerve conduction studies, or diagnostic nerve blocks may be valuable in the assessment of disease processes that can give rise to pain. Certain disease-specific laboratory measurements (e.g., tumor markers in some malignancies) are sensitive indicators of disease progression and may correlate with painful symptoms. Findings on imaging studies do not correlate, or correlate weakly, with patient pain experiences.
- Make relevant referrals for multidisciplinary assessment. Many painful conditions may be managed adequately by a patient's primary care physician. However, referral to a pain specialist or a team approach may be best for certain cases of chronic, cancer-related, or otherwise complex pain that is debilitating or refractory to treatment. Important members of the pain management team may include neurologists, psychiatrists, anesthesiologists, physiatrists, physical or occupational therapists, clergy, social workers, and counselors. Pain clinics may bring all relevant team members "under one roof." Many pain specialists believe that referrals frequently are made past the so-called golden hour when their intervention may be of maximal effectiveness, especially in cases of neuropathic and cancer pain. Referral to a pain specialist ideally should occur before significant disability or loss of function occurs; pain behaviors or the emergence of maladaptive coping strategies may serve as cues for referral.
- Assess patient expectations and goals. What does the patient expect or hope for in terms of treatment outcomes? Patient goals should be assessed in a number of domains, including pain intensity, daily function, and quality of life. Having patients set goals will keep the physician from taking a narrow focus on physical factors such as pain intensity, to the exclusion of the larger aspects of pain's impact on the patient's life.

Principles of Pain Management

Eleven principles should guide the management of pain [*see Table 2*].

- Take a detailed history of the patient's pain (see above). An accurate assessment of the patient's pain is essential for the direction of future therapy.
- Design a patient-specific diagnostic and therapeutic approach. Although there are numerous guidelines for treatment of pain in different conditions, pain is a complex subjective experience requiring an individualized approach. Management of pain should take into account such factors as the underlying cause (or causes) of pain, medical comorbidities, psychological state, response to prior therapies, functional status, compliance with or tolerance of different therapeutic regimens, and objectives of therapy.
- Ensure the availability of expertise to provide therapeutic alternatives. Pain may be managed in a variety of clinical settings. In a multidisciplinary pain treatment center, expertise on a variety of alternative therapies may be readily available. When primary management of pain occurs in other settings, access to external resources must be assured.
- Choose the simplest approach before attempting more complicated and invasive techniques. As initial therapy of a given painful condition, medical management may be superior to invasive techniques on the basis of risk/benefit assessment, cost, and other factors. For drug treatment, selection of agents can follow the stepwise approach advocated by the World Health Organization (WHO).[46]
- Maintain ongoing communication between physician and patient in defining therapeutic options and potential risk-to-benefit ratios of each approach. Patients may respond better to therapeutic regimens that they helped to select and in which they therefore feel invested.
- Titrate doses to maximize efficacy and minimize side effects. Inadequate dosing of analgesics may be the most frequent cause of undertreatment of pain. To provide optimal pain management, physicians require training in proper dosing of pain medications, guidelines for dose titration, and significant side effects of pharmacologic agents (see below). Optimal titration will vary between patients. Physicians who adhere to accepted management guidelines need not fear legal or professional repercussions from their treatment of patients in pain.
- Anticipate and treat side effects. Poor management of side effects is associated with suboptimal treatment compliance, leading to continued pain and such secondary pain-related problems as mood disturbance, inactivity, and impaired function.
- Utilize adjuvant medications for analgesia and for specific pain syndromes. Specific painful conditions may best be treated with multiple analgesics. Adjuvant medications should be employed when clinical evidence supports their use.
- Distinguish physical dependence from psychological dependence. Surveillance for and prevention of psychological dependence (i.e., addiction) is a legitimate concern when opiates are used for pain management. Nevertheless, excessive fear of addiction is a significant barrier to optimal pain management. Patient (and physician) fears and misconceptions should be dealt with in frank discussion. Patients must be counseled that in the absence of a history of substance abuse, the likelihood of developing addiction while taking opioid medications for pain has proved exceedingly rare (occurring in less than 1% of patients in many studies).[47] Physical dependence, a common phe-

nomenon, must be defined and differentiated from addiction. Especially with new patients, it is important for physicians to be familiar with the diagnostic criteria for substance abuse disorders; these disorders present a different set of management problems [*see 13:VI Drug Abuse and Addiction*].

- Provide continuity of care throughout evaluation and treatment. Fragmentation of care is a significant contributor to the suboptimal medical treatment often received by patients with chronic or complex pain. Failure to establish a durable physician-patient relationship may lead to the incorrect perception of so-called doctor-shopping and drug-seeking behavior by a patient who lacks a coordinated pain management strategy. Involvement of numerous providers in the pain management process may predispose to such fragmentation.
- Continually reassess the degree of pain relief and the impact of treatment on mood, function, and overall quality of life. Ongoing assessment of the degree to which treatment goals are being met allows for appropriate adjustment of the therapeutic regimen. Reassessment is especially critical after changes in pharmacologic agent or dosing, as well as after nonpharmacologic interventions. Much of the clinical import of pain relates to its impact on mood, function, and quality of life; assessment of the effect of therapy on these secondary variables is of special significance. Assessment of ongoing patient pain complaints may also be a valuable part of monitoring underlying disease processes.

The frequency of pain reassessment should be keyed to the nature of the pain. Reevaluation of acute pain is usually done after therapeutic interventions over a short time frame. Reassessment of chronic pain is more likely to occur at predetermined intervals over an extended period. Frequency will also vary with clinical setting. For example, inpatients and long-term care residents may undergo pain assessment as the so-called fifth vital sign on nursing rounds, whereas outpatients may have follow-up as appropriate. The specific tools with which pain is reassessed will vary with patients as necessary.

Management of Pain in Specific Conditions

CANCER

Most pain in patients with advanced cancer is caused by the cancer itself (e.g., tumor growth leading to tissue invasion, painful metastases, or involvement of the nervous system or visceral structures).[35] Therefore, treatment of the underlying malignancy with chemotherapy, radiotherapy, or surgery is a vital component of the management of cancer pain. However, symptomatic relief is often necessary as well, which is commonly achieved with drug therapy.

The WHO's three-step analgesic ladder can be used to guide the selection of pain medication [*see Figure 3*].[48] In validation studies, the WHO ladder was found to be effective in 69% to 100% of adults with cancer.[48-50] In step one, nonsteroidal anti-inflammatory drugs (NSAIDs) and acetaminophen are used for patients with mild pain. If the pain persists or increases (i.e., becomes moderate pain), treatment progresses to step two, in which nonopioids are combined with opioids such as codeine and hydrocodone or are given in the form of combination analgesics such as oxycodone with acetaminophen. For severe pain, treatment progresses to step three, which involves opioids with higher potency (e.g., morphine, methadone, hydromorphone, or fentanyl), with or without nonopioids. Adjuvant drugs may be utilized at any level of the ladder. Common adjuvants include corticosteroids (the most widely used agents for this purpose); antidepressants, anticonvulsants, and other agents for neuropathic pain; bisphosphonates and radionuclides for bone pain; and antibiotics for pain from ulcerating tumors. Nonpharmacologic adjuvant treatments such as external-beam radiation, neurosurgical ablative procedures, psychiatric therapy, and anesthetic interventions may also be employed. Management of cancer pain may be complicated significantly by the simultaneous need to manage multiple other symptoms of advancing malignancy, including fatigue, depression, dyspnea, anorexia, cachexia, chronic nausea, and anxiety.

Table 2 Principles of Pain Management

1. Take a detailed history of the patient's pain
2. Individualize the therapeutic approach
3. Ensure the availability of expertise to provide therapeutic alternatives
4. Choose the simplest approach before more complicated and invasive techniques
5. Maintain ongoing communication between physician and patient in defining therapeutic options and potential risk-to-benefit ratios of each approach
6. Titrate doses to maximize efficacy and minimize side effects
7. Anticipate and treat side effects
8. Utilize adjuvant medications for analgesia and in specific pain syndromes
9. Distinguish physical dependence from psychological dependence
10. Provide continuity of care throughout evaluation and treatment
11. Continually reassess the degree of pain relief and the impact of treatment on mood, function, and overall quality of life

HIV/AIDS

The increasing life expectancy of patients with HIV infection has led to the characterization of HIV/AIDS-related pain as a type of chronic pain. Management of pain in this setting is complicated greatly by the diversity of pathologic mechanisms un-

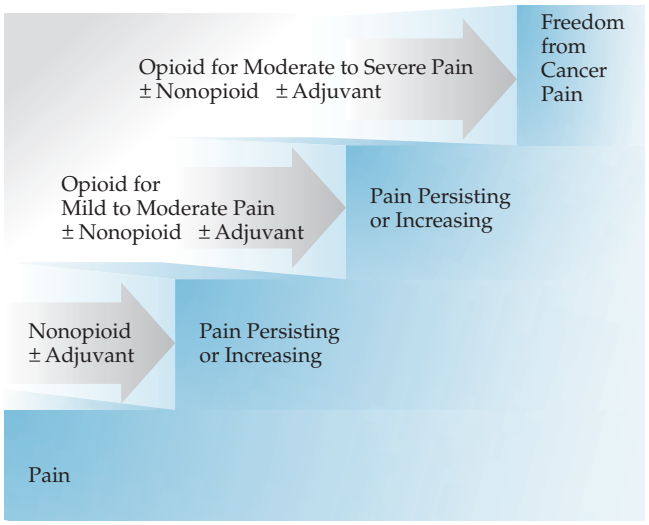


Figure 3 **The World Heath Organization's three-step analgesic ladder can be used to guide the selection of pain medication.**

derlying the variety of pains and pain syndromes seen in these patients. The Agency for Healthcare Policy and Research (AHCPR) recommends that the WHO approach to cancer pain management (see above) be used for the treatment of painful conditions in HIV/AIDS patients.

It has been suggested that as many as 85% of patients with HIV/AIDS-related pain do not receive adequate pain treatment. Women, substance abusers, and patients of low socioeconomic status are particularly vulnerable.[51]

Pharmacologic Management of Pain

The WHO, the American Pain Society, and the AHCPR have published specific guidelines on analgesic use.[45,47,52] Although the WHO analgesic ladder (see above) provides a useful construct for drug selection, pharmacotherapy ultimately must be tailored to the individual patient, taking into account the patient's medical comorbidities, response to prescribed drugs, side effects experienced, and a wide range of psychosocial and quality-of-life variables.

OPIOID ANALGESICS

Indications

Opioids play a significant role in the treatment of a number of conditions, including acute pain, trauma-related pain, postoperative pain, cancer pain, and some types of chronic noncancer pain, including osteoarthritis, low back pain, and neuropathic pain. Because responsiveness to opioids varies between patients, failure of response to an adequate trial of one opioid should be followed by an adequate trial of another opioid. Within a given potency range, there is little evidence to recommend one opioid over another as the initial prescribed agent. Selection of initial analgesic agent often is based on a combination of drug characteristics (e.g., half-life and duration of effect, speed of onset, route of administration) and patient characteristics (e.g., prior response to different opioids) [*see Tables 3 and 4*].

Table 3 Opioid Analgesics Used for Mild to Moderate Pain

Drug	*Dosage (Oral)**	*Analgesic Equivalence (mg)†*	*Duration*
Codeine	15–60 mg q. 4 hr	200	3–4 hr
Hydrocodone	5–15 mg q. 4–6 hr	200‡	3–5 hr
Meperidine	50–150 mg q. 3–4 hr	300	2–3 hr
Propoxyphene	65 mg q. 4 hr	200	3–6 hr

*These agents are typically combined with acetaminophen or aspirin, which limits dose escalation.
†Compared with 30 mg of oral morphine.
‡Equivalence unsubstantiated but thought to approximate codeine.

Neuropathic pain The role of opioids in neuropathic pain has been established since the late 1990s, when a series of well-designed randomized controlled clinical trials clearly demonstrated efficacy in neuropathic pain states such as diabetic neuropathy. Current data suggest that neuropathic pain may be managed more effectively when morphine is combined with gabapentin than when either agent is used alone [*see* Antiepileptic Drugs, *below*]. Although monotherapy is often tried initially in managing neuropathic pain (most commonly with gabapentin), many patients are best served by a so-called cocktail approach that targets multiple mechanisms.

Mechanism

Opioids achieve their analgesic effects chiefly by interacting with opioid receptors in the CNS. This binding inhibits the transmission of nociceptive impulses from periphery to spinal cord, activates descending pain suppression pathways, and modulates activity in the limbic system. Pain relief from opioids may also derive in smaller part from activity at receptors in the peripheral nervous system. Individual opioids display different affinities for, and interactions with, each of the three types of opioid receptors (mu, kappa, and delta; see above). Many of the most commonly used opioids are mu agonists. Differences in binding patterns underlie many of the differences in side effect

Table 4 Opioid Analgesics Commonly Used for Moderate to Severe Pain

Drug	*Dose*	*Half-life (hr)*	*Comment*
Morphine	10 mg I.V., 30 mg p.o.	2–4	Standard of comparison for opioid analgesics; start with a lower dose in patients with respiratory insufficiency, increased intracranial pressure, liver failure, or renal failure
Hydromorphone	1.5 mg I.V. , 7.5 mg p.o.	2–3	Slightly shorter acting
Methadone	10 mg I.V., 20 mg p.o.	15–20	Good oral potency; long plasma half-life, so repetitive dosing may result in accumulation and excessive sedation
Fentanyl	100 μg = 4 mg I.V. morphine	1–2	Short half-life; transdermal and transmucosal preparations available
Meperidine	75 mg I.V., 300 mg p.o.	2–3	Not recommended for patients with chronic cancer pain or impaired renal function or those receiving monoamine oxidase inhibitors
Levorphanol tartrate	2 mg I.V., 4 mg p.o.	12–16	—
Oxycodone	30 mg p.o.	3–4	Typically used in combination with nonopioid analgesics that limit dose escalation
Oxymorphone	1 mg I.V., 6 mg p.r.	2–3	Not available orally

profile among individual agents. Although opioid receptors are found primarily in the CNS and the gastrointestinal tract, they are widely distributed throughout the body; the potential analgesic significance of receptor networks outside the CNS is not fully understood at present.

Administration and Formulation

Opioids may be administered via a wide range of routes—oral, sublingual, rectal, intravenous, transdermal, intrathecal, epidermal, and subcutaneous. The preferred mode of administration will vary with the specific opioid, the severity of pain, and the patient's medical condition. Not all opioids are available in all dosing forms. The oral route is usually used for systemic administration. For mild to moderate pain, there is no consistent evidence that any route of systemic administration is superior to oral administration; other systemic routes may be preferred for more severe pain. Oral or transdermal administration generally is preferred for treatment of chronic conditions. Medical factors may prevent oral administration. For example, patients with end-stage dementia often lose the ability to swallow medications, and 80% of cancer patients are unable to take oral opioids for some time before death, because of intractable nausea, oral mucositis, dysphagia or aspiration concerns, bowel obstruction, or other conditions. Oral administration may be impractical for very high dose opioid therapy because the patient would have to swallow an unreasonable number of tablets.

Intravenous opioid administration has a number of other advantages. Intravenous administration provides rapid onset of effect and allows rapid titration to optimal analgesia. Avoidance of first-pass hepatic degradation yields higher bioavailability and thus an opportunity to reduce total dosage. Continuous intravenous administration may provide the patient with a consistent level of analgesia. However, the greater potency of intravenous opioids increases the risk of systemic side effects, particularly if rapid dose escalation occurs or if the patient is opioid naïve.

Subcutaneous injection may be used when small fluid volumes are sufficient to deliver the prescribed opioid dose. This route enables rapid titration and onset of analgesia; however, a shortened duration of effect necessitates frequent redosing to maintain constant pain relief. The risk of side effects is greater with subcutaneous than with oral administration. Rectal administration offers ease of use but the variation in absorption and first-pass metabolism from patient to patient can make efficacy unpredictable. Furthermore, rectal dosing cannot be used in the setting of diarrhea, transmucosal lesions, neutropenia, or severe thrombocytopenia. Transdermal administration may be useful when the oral route is unavailable; however, this dosing method is limited by the amount of available skin surface and is not effective for control of fluctuating pain levels. Intramuscular administration should, in general, be avoided because of erratic absorption and pain associated with the procedure.

Patient-controlled analgesia (PCA) devices allow patients to adjust opioid delivery according to their level of pain. PCA may be used for initiation of parenteral opioid therapy, rapid opioid titration, or treatment of incident pain. The devices are programmed for dose size, lockout interval, and cumulative dose allowable over a given period; continuous infusion may be programmed to supplement PCA doses, enabling sleep and covering baseline pain. Safety features prevent the delivery of more medicine when the patient is excessively sedated. Compared with conventionally administered opioids, PCA provides equivalent or superior analgesia with less total opioid consumption, fewer side effects, and no greater likelihood of dependence. PCA may be used in the home, allowing severely ill patients to receive adequate pain relief outside of the hospital.

Intraspinal opioid administration is useful for patients requiring such large amounts of opioids that systemic side effects prevent the maintenance of adequate drug levels. Delivery of opioids to the epidural or intrathecal space exploits the large concentration of opioid receptors in the dorsal horn of the spinal cord, achieving analgesia at significantly smaller doses of opioid than would be required with systemic administration. Compared with the intravenous route, intrathecal and epidural morphine are 100 times and 10 times more potent, respectively. The smaller opioid doses required for intraspinal administration result in effective analgesia at a greatly reduced side-effect burden. Implantable intrathecal drug delivery systems consist of an implanted programmable pump that delivers small doses of morphine directly to the spinal fluid. Although intraspinal opioid therapy has obvious benefits, it is not a first-line treatment. It is indicated only for patients who have experienced some degree of pain relief in multiple trials of different systemically administered opioids but who have suffered intolerable side-effect burdens that effectively prohibited the continued use of these regimens.

Desired duration of drug action varies with clinical circumstances. Long-acting and sustained release formulations are useful for patients in continuous pain; these drugs reduce severity of end-of-dose pain and may let patients sleep through the night. Sustained-release formulations are frequently used to treat patients with cancer pain. Most opioids may be given on an around-the-clock or an as-needed basis, but around-the-clock administration may improve patient outcomes and adherence. Long-acting formulations also may improve patient adherence. Short-acting formulations are used to manage intermittent pain and breakthrough pain.

Dosing

Titration to optimal dosage generally involves gradual adjustment of a small initial dose, on the basis of both the pain relief achieved and the severity of side effects encountered. A patient may be given an immediate-release oral formulation of an agent with a short half-life for a limited time (e.g., 24 to 48 hours), during which medication consumption, efficacy of pain relief, and side effects encountered are recorded in a patient diary. Alternatively, patients in severe pain may need rapid titration of a potent opioid via continuous intravenous infusion. Once pain control has been achieved and the optimal dose determined, the opioid may (if desired) be prescribed in long-acting or sustained-release form.

The opioid dose required to control pain in a given patient is influenced significantly by the type and severity of pain, any preexisting opioid tolerance, psychological distress, age, and genetic factors. Changes in dose may be required if pain severity increases. In general, dose changes should be made in increments of one third to one half the prior dose and should be modified according to the patient's use of breakthrough medications. This formula also applies to the reduction or cessation of opioid dose in patients whose source of pain is eliminated. Gradual reduction of the daily dose (i.e., reductions in increments of less than 25%) is necessary to prevent symptomatic withdrawal.

Patients receiving long-acting opioid therapy may experience episodes of acute pain that cannot be adequately controlled by the baseline analgesia provided by long-acting opioids. Short-acting, immediate-release opioids are used to cover this break-

through pain and are a critical component of long-term opioid therapy. There is no standard methodology for the dosing of breakthrough medications. Single rescue doses are commonly set at either 10% to 20% of the total daily dose or 25% to 30% of the single standing dose. As is the case with standing doses, breakthrough doses should be adjusted as appropriate for the individual patient. When possible, the same agent should be used for standing and breakthrough medications; this approach eases titration and eventual conversion (see below).

Weak opioids often are compounded with a nonopioid analgesic (commonly, acetaminophen). This combination may result in an opioid dose-sparing effect; however, it does not consistently reduce side effects. In addition, the potential for overdose of the nonopioid component limits escalation of the opioid dose. This may necessitate a switch to another opioid or combination if pain increases in severity.

Opioid Conversion

Although there is no theoretical upper limit to the dose of most opioids, practical ceilings for a given agent may be imposed by side effects, administration concerns (e.g., excessive injectate volume or skin surface area required for patches), or lack of apparent response to the drug. Such impediments to dose escalation may require switching of agent, route of administration, or both. When changing from one agent to another is necessary, clinicians often make use of published equianalgesic conversion charts. However, these standardized tables were constructed on the basis of limited data; therefore, they should serve as only a starting point for opioid conversion dosing.

Consideration of opioid conversion because of increasing pain should begin with a comprehensive assessment of the nature of the pain necessitating the switch. Pain caused by a worsening underlying medical condition may require a change in the treatment of that condition in addition to, or in place of, a change in analgesic. Pain attributable to inflammatory or neuropathic etiologies may be better treated with adjuvant analgesics than with a modification of opioid therapy. Pain purely attributable to opioid side effects may simply require treatment of the adverse effects (see below). Opioid conversion for refractory pain or intolerable side effect burden is justified if the pain is opioid-responsive, the current analgesic has been titrated to maximal effect, and side effects are already optimally managed.

Little evidence supports a particular selection order of opioid analgesics. The choice of a new agent may be made on the basis of such factors as the patient's prior experience with opioids of a given class or receptor profile. Once the new agent has been selected, equianalgesic conversion tables should be consulted to arrive at a starting point for dosing [*see Tables 3 and 4*]. If the patient had been receiving multiple opioids, conversion should be based on the total dose of all prior agents, expressed as morphine equivalents. The calculated dose should then be individualized according to the nature and severity of the patient's pain. If opioid conversion was necessitated by intolerable side effects in the setting of adequate pain control, the calculated dose should be reduced by approximately 25% to 50%; this will allow for the lack of cross-tolerance among agents. Alternatively, if conversion was prompted by inadequate analgesia and significantly impaired quality of life, the new agent may be started at or near an equianalgesic dose. Additional short-acting opioids should be made available to the patient during titration of the new drug in order to achieve stable analgesia. Breakthrough medication is especially important during opioid conversion. Continuous reassessment of the patient's pain and total daily dosage (extended release plus immediate release) should occur for the first 2 weeks after opioid conversion, with titration of the extended-release and breakthrough doses as appropriate. Both inadequate pain control and excessive narcosis are threats in the immediate conversion period.

Patients whose pain cannot be controlled with opioids should be considered for invasive pain control modalities. In general, however, such modalities should not be explored until adequate trials of several opioids have been completed.

Opioid Side Effects

Sedation Most patients experience sedation with initiation of therapy or an increase in dosage, but this sedation commonly resolves within 3 to 7 days; chronic sedation occurs in a small minority of patients. Sedation is a special concern in the elderly and in patients taking other sedating medications. Sedation may also reflect the relief of pain itself, as patients recoup rest previously lost to discomfort. Nonessential medications with sedating effects should be eliminated in patients experiencing opioid-related sedation. Stimulants such as caffeine or methylphenidate should be considered for use in cancer patients with significant persistent sedation, although the use of psychoactive medication in the elderly should always be approached with caution.[53] Stimulants are not recommended for patients experiencing sedation caused by opioid management of chronic pain.

Nausea Opioids may cause nausea and vomiting by both central and peripheral mechanisms. Centrally, opioids directly stimulate the chemoreceptor trigger zone in the medulla. Peripherally, opioids slow gastric motility. Additionally, many patients receiving opioid therapy will also have medical comorbidities or will be receiving treatments (e.g., chemotherapy) that contribute to nausea and vomiting. Use of a variety of antiemetics during the first 1 or 2 days of opioid administration may successfully control nausea and vomiting. Ondansetron, prochlorperazine, or hydroxyzine may alleviate centrally-induced nausea, whereas metoclopramide may reduce nausea from slowed gastric motility. Scopolamine may alleviate motion-exacerbated nausea in ambulatory patients.[54] Anticholinergic side effects of common antiemetics may be clinically significant, particularly in the elderly. As is the case with sedation, nausea and vomiting frequently resolve shortly after onset of treatment.

Constipation Constipation is a common and clinically significant side effect of all opioids. It is caused by decreased peristalsis and intestinal secretions and increased electrolyte and water resorption in the large intestine. Additional risk factors for constipation include advanced age, impaired mobility, concurrent medications with constipating side effects, and gastrointestinal comorbidities. Unlike sedation and nausea, constipation does not improve with time. This side effect is underdiagnosed and may lead to anorexia, vomiting, abdominal pain, obstruction, impaction, and perforation.

Dietary changes to prevent constipation should be implemented with opioid administration. All patients on around-the-clock opioids should undergo regular assessment for constipation and be given a stool softener and a stimulant laxative. Additional agents may be used if constipation is severe or persists; prolonged absence of bowel movements should result in assessment for fecal impaction, with treatment if necessary. Peripheral opioid antagonists are currently being tested and may prove useful.

Respiratory depression Respiratory depression, caused by direct inhibition of brain stem respiratory centers when the dose is escalated too rapidly, is a rare side effect of opioid therapy. Respiratory depression is primarily of concern in opioid-naïve patients receiving large opioid doses; it also may be seen in the setting of head injury or concurrent pulmonary disease. Sedation level and respiratory status should be monitored regularly during the first 24 hours of therapy in opioid-naïve patients. If respiratory depression occurs, the opioid may be stopped until the depression resolves, after which therapy may be continued at 75% of the previous dose. Spirometry and supplemental oxygen may be useful in these patients. Severe respiratory depression may be treated with intravenous naloxone.[54] Fear of respiratory depression should never preclude the effective delivery of pain relief.

Confusion Opioid-induced neurotoxicity is a relatively recently identified syndrome that occurs primarily in patients receiving high-dose or prolonged opioid therapy and in those with decreased renal function.[55] Patients with previous cognitive impairment are at additional risk. Delirium, agitation, myoclonus, and hyperalgesia are features of this condition. Confusion is a side effect of particular concern in the elderly or in patients with concurrent CNS disease. Nonessential medications with CNS effects should not be prescribed for elderly patients requiring opioid therapy. Neuroleptics may be useful against confusion, mental clouding, or persistent delirium.

Other side effects Pruritus secondary to histamine release is an additional opioid side effect; it is generally managed with diphenhydramine or hydroxyzine. Myoclonus may be treated with clonazepam. Urinary retention, vertigo, sweating, and hypotension are less frequently reported side effects of opioid therapy.

Tolerance, Physical Dependence, and Psychological Dependence

Tolerance Tolerance occurs when exposure to a drug changes the dose-response relationship, such that a higher dose of agent is required to maintain a given effect. There is no clinical limit to tolerance, and wide ranges of opioid requirements have been reported. Patients in the last 4 to 6 weeks of life require an average of 400 to 600 mg of morphine equivalents per 24 hours, and up to 10% of patients require more than 5,000 mg.[56] As noted, opioids do not have a ceiling effect per se and may be titrated to very large doses as required for optimal pain relief. Progression of side effects with increasing dosage may be managed with opioid rotation, adjuvant analgesic drugs, and epidural local anesthetics. Tolerance rarely develops in patients with stable disease; increasing requirements in the setting of previously controlled chronic pain should prompt a comprehensive medical evaluation.

Physical dependence Physical dependence describes a state of adaptation, often including tolerance, in which a drug class–specific withdrawal syndrome could be produced by abrupt cessation of drug administration, rapid reduction in dose, decreasing blood level of drug, or administration of an antagonist or mixed agonist-antagonist. Physical dependence is virtually universal with prolonged opioid therapy; to avoid withdrawal syndrome, all patients receiving opioids for 1 week or longer should have the drug tapered off rather than abruptly discontinued.

Psychological dependence Psychological dependence, or addiction, is a primary, chronic, neurobiologic disease whose development and manifestations are influenced by genetic, psychosocial, and environmental factors. Addiction is characterized by impaired control over drug use, compulsive use, continued use despite harm, and craving. The psychologically dependent person has a consuming involvement with drug use, which typically results in a range of social, vocational, economic, and legal difficulties.

Fear of addiction on the part of both physicians and patients is a significant cause of underutilization of opioids in the treatment of pain, yet this fear is largely unfounded. Numerous studies have shown that the prevalence of psychological dependence in patients receiving opioids for the relief of pain is extremely low—0.03%, in one study.[47]

More common in this population is pseudoaddiction, a condition in which patients with poorly managed pain mimic many of the signs of psychological dependence. Pseudoaddiction is characterized by drug seeking, or an increased focus on obtaining medications, and possibly illicit drug use or deception. The defining distinction between addiction and pseudoaddiction is that pseudoaddiction behaviors resolve with effective pain management. As such, pseudoaddiction is best managed with effective pain relief and is likely to be exacerbated by the curtailing of opioid therapy.

NONSTEROIDAL ANTI-INFLAMMATORY DRUGS AND ACETAMINOPHEN

Indications

NSAIDs, acetaminophen, or both may be employed against a variety of types of acute and chronic pain. They are especially useful for certain types of somatic pain, including muscle and joint, bone, dental, postoperative, and inflammatory pain. These drugs may provide sufficient relief for mild or moderate pain. In the setting of severe pain, they may be added to an opioid regimen for their opioid-sparing effect. Combination opioid-nonopioid therapy may allow for enhanced pain relief with reduced side effect burden.

Mechanism

NSAIDs act primarily through inhibition of cyclooxygenase, thereby decreasing prostaglandin synthesis. Nonselective NSAIDs inhibit both COX-1 (a normal constituent of blood vessels, stomach, and kidney) and COX-2 (an enzyme induced in the setting of peripheral inflammation); COX-2 inhibitors selectively inhibit the COX-2 enzyme. All NSAIDs have analgesic, anti-inflammatory, and antipyretic effects. Acetaminophen works via a central mechanism; its actions are primarily analgesic and antipyretic.

Administration Route and Formulation

Most NSAIDs and acetaminophen are taken orally, although other routes of administration exist. Numerous formulations are available without prescription. In addition, combinations of NSAIDs or acetaminophen with opioids or members of other drug classes are commonly used.

Dosing

Dosing amount and frequency vary across preparations [*see Table 5*]. NSAIDs and acetaminophen exhibit a dosage ceiling, beyond which additional side effects accrue without additional therapeutic benefit.

Side Effects

Exact side effect profiles vary with individual drugs, but non-

Table 5 Nonopioid Analgesics Commonly Used for Mild to Moderate Pain

Drug Name	*Starting Oral Dose (mg)*	*Comments**
Aspirin	650	Often used in combination with opioid analgesics; reduces platelet function; avoid in pregnant patients and those with clotting disorders; avoid concomitant use with steroids
Choline magnesium trisalicylate	1,500	Does not affect platelet function
Acetaminophen	650	Does not affect inflammation or platelet function; use with caution in patients with hepatic dysfunction
Ibuprofen	200–400	Higher analgesic potential than aspirin
Diflunisal	500–1,000	Longer duration of action than ibuprofen
Naproxen	250–300	Longer duration of action than ibuprofen
Ketorolac tromethamine	10	Do not use for more than 5 days; I.M. formulation can be given in dose of 30 mg
Tramadol	50–200	Weak mu agonist and serotoninergic and catecholamine reuptake inhibitor
Celecoxib	100–200	Cyclooxygenase-2 inhibitor; contraindicated in sulfa-allergic patients

*Except for acetaminophen, tramadol, and celecoxib, all these drugs are nonselective nonsteroidal anti-inflammatory agents and share the side effects of this class of agents [*see text*].

selective NSAIDs as a class exhibit certain characteristic side effects, including gastrointestinal disturbances, bleeding, hypersensitivity, and CNS effects. In addition, nonselective NSAIDs have nephrotoxic potential, especially in patients with renal dysfunction who require prostaglandins for maintaining renal vasodilatation. NSAID side effects are often dose dependent; use of these drugs at low doses for short periods may minimize their impact. Additionally, preventive strategies may be employed, such as coadministration of gastrointestinal tract–protective drugs (e.g., lansoprazole, esomeprazole, or misoprostol). Although the reputation of selective COX-2 inhibitors was tarnished by the discovery of life-threatening side effects with rofecoxib and valdecoxib and the subsequent withdrawal of these agents from the market, celecoxib remains a very helpful agent for some patients, particularly those with low or no risk of coronary artery or cerebrovascular disease. Nonopioids have a ceiling effect for analgesia; increasing dosage above a certain level will increase side effect burden without yielding enhanced analgesia. Neither tolerance nor physical dependence is seen with NSAID use.

Acetaminophen provides pain relief comparable to most NSAIDs without the gastrointestinal disturbances and bleeding seen with NSAIDs. Acetaminophen must be used with caution, if at all, in patients with liver disease. In patients with normal livers, however, hepatic toxicity is rare with dosages beneath the ceiling of 3 to 4 g/day.

CORTICOSTEROIDS

Indications

Corticosteroids are the most widely used general adjuvant analgesics.[57] In the setting of cancer, they are especially useful for relief of acute pain associated with spinal cord compression, increased intracranial pressure, superior vena cava syndrome, metastatic bone pain, neuropathic pain secondary to infiltration or compression by tumor, and hepatic capsule distension. High-dose corticosteroids are often used for inpatients with advanced disease who are in acute pain crisis.[58] Corticosteroids are commonly used for pain related to musculoskeletal conditions.

Mechanism

Steroids act through an anti-inflammatory mechanism. They also produce glucocorticoid and mineralocorticoid effects.

Administration Route and Dosing

Corticosteroids are commonly given in oral or injectable form. Dosing varies with agent and indication.

Side Effects

Steroids are generally well tolerated for short-term treatment; however, toxicities often arise with prolonged high-dose therapy. Common toxicities include adrenocortical insufficiency, hypertension, immune suppression, the masking of signs of infection, glaucoma, electrolyte imbalances, gastrointestinal ulceration and bleeding, osteoporosis and pathologic fracture, and psychiatric disturbance or psychosis. Care must also be taken to prevent withdrawal syndrome upon discontinuance.

ANTIEPILEPTIC DRUGS

Indications

Antiepileptic drugs (AEDs) such as gabapentin, carbamazepine, and topiramate may be useful as adjuvant drugs in the treatment of a variety of forms of neuropathic pain, including peripheral diabetic neuropathy, postherpetic neuralgia, reflex sympathetic dystrophy, trigeminal and glossopharyngeal neuralgia, HIV neuropathy, and spinal cord injury–related dysesthesias. These agents are also useful for postlaminectomy, phantom limb, and cancer pain.

Considerable clinical data suggest that gabapentin should be a first-line treatment for neuropathic pain, because of its greater clinical efficacy and a more favorable side effect profile than seen with older AEDs and other agents such as tricyclic antidepressants (TCAs).[59,60] Gabapentin does not interact with other drugs. Carbamazepine should be considered as only a second- or third-line agent for most neuropathic pain conditions. However, carbamazepine remains the first-line treatment for trigeminal neuralgia [*see 11:II Diseases of the Peripheral Nervous System*]. Topira-

mate is increasingly used off-label as an adjuvant analgesic agent for neuropathic pain.

Mechanism

The exact mechanism of the analgesic effect of AEDs is unknown. These agents reduce membrane excitability, thereby possibly suppressing discharges from pathologically altered neurons, likely primarily by the inhibition of sodium channel transmission.

Administration Route and Dosing

AEDs are administered orally. Current guidelines recommend starting gabapentin at a dosage of 100 mg every 8 hours; however, optimal control of neuropathic pain often requires titration to 3,600 mg a day in divided doses. Topiramate must be titrated slowly (i.e., no faster than 25 mg a week) to decrease the incidence of side effects. Although daily doses of up to 400 mg are used for epilepsy, 100 to 200 mg is a more typical dose for pain. Dosing of other AEDs varies with agent and indication.

Side Effects

Gabapentin is generally well tolerated; some patients experience mild sedation and gastrointestinal effects. However, elderly patients often have poor tolerance of gabapentin because of cognitive effects, particularly at higher dose levels. Topiramate can cause cognitive slowing and paresthesias, especially with overly rapid titration.

The most common side effects seen with other AEDs include sedation, nausea and vomiting, and dizziness. Carbamazepine may cause thrombocytopenia or liver damage. Phenytoin may induce dose-related CNS effects, as well as hepatotoxicity, hypersensitivity reactions, and lymphadenopathy at high serum concentrations.

ANTIDEPRESSANTS

Tricyclic Antidepressants

Indications Amitriptyline, nortriptyline, imipramine, and other TCAs are frequently used as adjuvant analgesics in the treatment of a variety of types of neuropathic pain (e.g., painful diabetic neuropathy, postherpetic neuralgia, chronic facial pain, central pain) and chronic pain (e.g., cancer pain, chronic low back pain, osteoarthritis). TCAs have efficacy equivalent to AEDs in the treatment of certain types of neuropathic pain.[61] Efficacy is comparable among most TCAs.

Mechanism The analgesic effect of antidepressants may derive from the inhibition of norepinephrine and serotonin reuptake in the CNS, thereby increasing the activity of endogenous pain suppression pathways.

Administration route and dosing TCAs are taken in oral form. Although the effect of these agents may be seen at low doses, achieving optimum analgesic efficacy may require higher dose titration. Amitriptyline, desipramine, doxepin, and nortriptyline are typically started at a dose of 10 to 25 mg and may be titrated to a maximum daily dose of 150 mg, if tolerated.

Side effects Common side effects of TCAs as a class include anticholinergic effects such as sedation, hypotension, constipation, and urinary retention; the elderly are most susceptible to these effects. Amitriptyline has the strongest anticholinergic profile and for that reason is given at bedtime; nortriptyline, which has a less prominent anticholinergic profile, is generally a better choice for older patients. At doses above those typically prescribed for pain, lethal cardiac arrhythmias may occur; thus, these medications are contraindicated in patients with conduction abnormalities.

Selective Serotonin Reuptake Inhibitors

Overall, selective serotonin reuptake inhibitors (SSRIs) have been disappointing as primary analgesic agents compared with TCAs and AEDs, despite their favorable side effect profiles. Both paroxetine and venlafaxine (which blocks reuptake of norepinephrine as well as serotonin) have been shown in randomized controlled trials to be useful in managing neuropathic pain, but both agents are generally considered to be second-line choices; they are used when refractoriness or poor tolerability has been demonstrated with other agents such as TCAs, AEDs, and opioids.

TOPICAL ANALGESIC AGENTS

There has been a growing interest in the role of topical therapy in the management of neuropathic pain. The enthusiasm has been partly driven by the discovery of the role of peripheral nociceptor modulation in dampening pain transmission and partly by the perpetual difficulty in certain populations with tolerability of systemic medications.

Transdermal Lidocaine

Approved by the Food and Drug Administration for the treatment of postherpetic neuralgia in 1999, the 5% lidocaine patch is now a first-line agent in managing postherpetic neuralgia. Patients are instructed to wear the patch for 12 consecutive hours over the area of pain and to then discard it; they are further instructed to wait to apply a new patch until another 12 hours have passed. One to three patches can be used at any one time, depending on the surface area of the pain site. There have been reports that transdermal lidocaine is effective in other painful conditions, such as low back pain, but few published data support the use of this patch in conditions other than postherpetic neuralgia.

EMLA

EMLA cream (lidocaine 2.5% and prilocaine 2.5%) has been shown to be very effective in children undergoing painful procedures, such as needle insertions and blood draws. Its role in managing neuropathic pain has yet to be demonstrated.

NEW APPROACHES

Patients with pain have benefited enormously in the past 10 to 15 years from the attention paid to the mechanism of pain, which has moved the field of pain research from empiricism to scientific validity. Several drugs have sprung from such scientific discoveries; some are now in development and some have been approved for use.

Ziconotide

Approved by the FDA in 2004, ziconotide is a synthetic version of a South Pacific marine snail toxin that blocks calcium transmission. It is 1,000 times more potent than morphine and thus may only be delivered intrathecally by an implantable pump.

Duloxetine

Duloxetine, a norepinephrine and serotonin reuptake inhibitor, was approved by the FDA in 2004. Its main indication is for burning peripheral neuropathy, although there is hope that it may be helpful as an antidepressant as well, in view of its mechanism of action.

Capsaicin

Capsaicin contains the ingredient in chili peppers that causes a burning sensation. Although a weak formulation of capsaicin has been available over the counter for many years, trials are now under way of a far more potent version, which may decrease pain more successfully by dampening pain transmission at the skin receptor level.

Pregabalin

The success of gabapentin as a first-line agent for neuropathic pain has resulted in the development of the related compound pregabalin, which is expected to be introduced in the United States in the second half of 2005. Although the mechanism of action of pregabalin is similar to that of gabapentin, there is hope that this newer agent may be useful as an alternative when patients do not respond adequately to other agents.

The author is a member of the speakers' bureaus of Pfizer/Pharmacia, GlaxoSmithKline, and Ortho-McNeil Pharmaceutical, Inc.

References

1. Merskey H, Bogduk N: Classification of Chronic Pain: Descriptions of Chronic Pain Syndromes and Definitions of Pain Terms, 2nd ed. IASP Press, Seattle, 1994, p 212

2. Towards a New Pharmacology of Pain. Basbaum AI, Besson JM, Eds. John Wiley & Sons, London, 1991

3. Loeser JD: Perspectives on pain. Proceedings of the First World Congress on Clinical Pharmacology and Therapeutics. Turner P, Ed. Macmillan, London, 1980, p 316

4. Levine J, Taiwo Y: Inflammatory pain. Textbook of Pain. Wall PD, Melzack R, Eds. Churchill Livingstone, Edinburgh, 1994, p 45

5. Cassel EJ: The nature of suffering and the goals of medicine. N Engl J Med 306:639, 1982

6. Jacobsen S, Mariano A: General considerations of chronic pain. Bonica's Management of Pain, 3rd ed. Loeser JD, Butler SH, Chapman CR, et al, Eds. Lippincott Williams & Wilkins, Baltimore, 2001, p 241

7. Drayer RA, Henderson J, Reidenberg M: Barriers to better pain control in hospitalized patients. J Pain Symptom Manage 17:434, 1999

8. Von Roenn JH, Cleeland CS, Gonin R, et al: Physician attitudes and practice in cancer pain management: a survey from the Eastern Cooperative Oncology Group. Ann Intern Med 119:121, 1993

9. Quality improvement guidelines for the treatment of acute pain and cancer pain. American Pain Society Quality of Care Committee. JAMA 274:1874, 1995

10. Coda BA, Bonica JJ: General considerations of acute pain. Bonica's Management of Pain, 3rd ed. Loeser JD, Butler SH, Chapman CR, et al, Eds. Lippincott Williams & Wilkins, Baltimore, 2001, p 222

11. Joint Commission on Accreditation of Healthcare Organizations: Pain Assessment and Management: An Organizational Approach. JCAHO, Oakbrook Terrace, Illinois, 2000

12. Glajchen M: Chronic pain: treatment barriers and strategies for clinical practice. J Am Board Fam Pract 14:211, 2001

13. FAQs about pain. American Academy of Pain Medicine, Glenview, Illinois, 2005 http://www.painmed.org/faqs/pain_faqs.html

14. Symptom management in cancer: pain, depression, and fatigue. NIH Consens State Sci Statements 19:1, 2002

15. HIV-Related Neuropsychiatric Complications and Treatments: Pain and HIV/AIDS. American Psychiatric Association, Office of HIV Psychiatry, Arlington, Virginia, 2004

16. A controlled trial to improve care for seriously ill hospitalized patients: the study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. JAMA 274:1591, 1995

17. Facts about pain. American Pain Foundation, Baltimore, 2004 http://www.painfoundation.org/page.asp?file=Media/fastfacts.htm

18. Stewart WF, Ricci JA, Chee E, et al: Lost productive time and cost due to common pain conditions in the US workforce. JAMA 290:2443, 2003

19. Rizzo JA, Abbott TA 3rd, Berger ML: The labor productivity effects of chronic backache in the United States. Med Care 36:1471, 1998

20. Woolf CJ: Recent advances in the pathophysiology of acute pain. Br J Anesthesiol 63:139, 1989

21. Neumann S, Doubell TP, Leslie T, et al: Inflammatory pain hypersensitivity mediated by phenotypic switch in myelinated primary sensory neurons. Nature 384:360, 1996

22. Fields HL, Rowbotham M, Baron R: Postherpetic neuralgia: irritable nociceptors and deafferentation. Neurobiol Dis 5:209, 1998

23. Melzack R, Wall PD: Pain mechanisms: a new theory. Science 150:971, 1965

24. Sluka KA, Walsh D: Transcutaneous electrical nerve stimulation: basic science mechanisms and clinical effectiveness. J Pain 4:109, 2003

25. Woolf CJ: Evidence for a central component of post-injury pain hypersensitivity. Nature 306:686, 1983

26. Costigan M, Woolf CJ: Pain: molecular mechanisms. J Pain 1(3 suppl 1):35, 2000

27. Covington EC: The biological basis of pain. Int Rev Psychiatry 12:128, 2000

28. Ru-Rong JI, Woolf CJ: Neuronal plasticity and signal transduction in nociceptive neurons: implications for the initiation and maintenance of pathological pain. Neurobiol Dis 8:1, 2001

29. Casey KL, Minoshima S, Berger KL, et al: Positron emission tomography analysis of cerebral structures activated specifically by repetitive noxious heat stimuli. J Neurophysiol 2:802, 1994

30. Iadarola MJ, Max MB, Berman KF, et al: Unilateral decrease in thalamic activity observed with positron emission tomography in patients with chronic neuropathic pain. Pain 63:55, 1995

31. Hassenbusch SJ, Pillay P, Barnett GH: Radiofrequency cingulotomy for intractable cancer pain using stereotaxis guided by magnetic resonance imaging. Neurosurg 27:220, 1990

32. Penfield W, Boldrey E: Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain 60:389, 1937

33. Fields HL, Besson JM: Pain modulation. Prog Brain Res 77:1, 1988

34. Basbaum AI, Fields HL: Endogenous pain control mechanisms: review and hypothesis. Ann Neurol 4:451, 1978

35. Portenoy RK, Lesage P: Management of cancer pain. Lancet 353:1695, 1999

36. Clouston PD, DeAngelis LM, Posner JB: The spectrum of neurological disease in patients with systemic cancer. Ann Neurol 31:268, 1992

37. Suwanwela N, Phanthumchinda K, Kaoropthum S: Headache in brain tumor: a cross-sectional study. Headache 34:435, 1994

38. Forsyth PA, Posner JB: Headaches in patients with brain tumors: a study of 111 patients. Neurology 43:1678, 1993

39. Dobalian A, Tsao JC, Duncan RP: Pain and the use of outpatient services among persons with HIV: results from a nationally representative survey. Med Care 42:129, 2004

40. Wulff EA, Wang AK, Simpson DM: HIV-associated peripheral neuropathy: epidemiology, pathophysiology and treatment. Drugs 59:1251, 2000

41. Galer BS, Keran C, Frisinger M: Pain medicine education among American neurologists: a need for improvement. Neurology 52:1710, 1999

42. Wolfe J, Grier HE, Klar N, et al: Symptoms and suffering at the end of life in children with cancer. N Engl J Med 342:326, 2000

43. Carver AC, Foley KM: Symptom assessment and management. Neurol Clin 19:921, 2001

44. Jacox A, Carr DB, Payne R, et al: Management of Cancer Pain: Adults Quick Reference Guide. No. 9. AHCPR Publication No. 94-0593. Agency for Health Care Policy and Research, US Dept of Health and Human Services, Public Health Service, Rockville, Maryland, 1994

45. Gonzales GR, Elliott KJ, Portenoy RK, et al: The impact of a comprehensive evaluation in the management of cancer pain. Pain 47:141, 1991

46. Report of the WHO Expert Committee on Cancer Pain Relief and Active Supportive Care: Cancer Pain Relief with a Guide to Opioid Availability, ed 2. Technical Report Series, 804, World Health Organization, Geneva, 1986

47. Porter J, Jick H: Addiction rare in patients treated with narcotics. N Engl J Med 302:123, 1980

48. Zech DF, Grond S, Lynch J, et al: Validation of WHO guidelines for cancer pain relief: a 10-year prospective study. Pain 63:65, 1995

49. Jadad AR, Bowman GP: The WHO analgesic ladder for cancer pain management: stepping up the quality of its evaluation. JAMA 274:1870, 1996

50. Ventafridda V, Tambourini M, Caraceni A, et al: A validation study of the WHO methods for cancer pain relief. Cancer 59:850, 1987

51. Breitbart W, Rosenfeld BD, Passik SD, et al: The undertreatment of pain in ambulatory AIDS patients. Pain 65:243, 1996

52. Principles of analgesic use in the treatment of acute pain and cancer pain, 4th ed. American Pain Society, Glenview, Illinois, 1999

53. Bruera E, Watanabe S: Psychostimulants as adjuvant analgesics. J Pain Symptom Manage 9:412, 1994

54. Levy MH: Pharmacologic management of cancer pain. Semin Oncol 21:718, 1994

55. Mao J, Sung B, Ji RR, et al: Neuronal apoptosis associated with morphine tolerance: evidence for an opioid-induced neurotoxic mechanism. J Neurosci 22:7650, 2002

56. Coyle N, Adelhardt J, Foley KM, et al: Character of terminal illness in the advanced cancer patient: pain and other symptoms during the last four weeks of life J Pain Symptom Manage 5:83, 1990

57. Wooldridge JE, Anderson CM, Perry MC: Corticosteroids in advanced cancer. Oncology 15:225, 2001

58. Ettinger AB, Portenoy RK: Use of corticosteroids in the treatment of symptoms associated with cancer. J Pain Symptom Manage 3:99, 1988

59. Rowbotham MC, Harden N, Stacey B, et al: Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial. JAMA 280:1837, 1998

60. Watson CP: The treatment of neuropathic pain: antidepressants and opioids. Clin J Pain 16(suppl 2):S49, 2000

61. Collins SL, Moore RA, McQuay HJ, et al: Antidepressants and anticonvulsants for diabetic neuropathy and postherpetic neuralgia: a quantitative systemic review. J Pain Symptom Manage 20:449, 2000

Acknowledgment

Figures 1 through 3 Seward Hung.

XV PARKINSON DISEASE AND OTHER MOVEMENT DISORDERS

Jorge L. Juncos, m.d.
Mahlon R. DeLong, m.d.

Most movement disorders result from diseases that affect the basal ganglia. Clinically, movement disorders can be divided into hypokinetic and hyperkinetic disorders: hypokinetic disorders are characterized by slow, impaired voluntary movements typified by parkinsonism; hyperkinetic disorders are characterized by involuntary movements, including chorea, ballismus, dystonia, tremor, tic, and myoclonus. Because of shared neural mechanisms, movement disorders with different etiologies may have similar motor manifestations. This clinical overlap, together with the lack of biologic markers for many of these disorders, can make etiologic diagnosis a challenging exercise. Nevertheless, major advances in the understanding of the genetics, molecular biology, and pathophysiology of these disorders have led to the availability of genetic testing for a selected few and to the development of effective pharmacologic and surgical treatments for others. Furthermore, the symptomatic therapies used today are often effective irrespective of etiology.

Movement disorders usually coexist with psychiatric, cognitive, and sleep disturbances that can contribute significantly to disability and that may even dominate the clinical picture. For example, depression can impair the response to otherwise adequate treatment of motor symptoms. In many instances, treatment of these comorbid conditions is an important aspect of management.

Pathophysiology

Movement disorders may arise from dysfunction in the basal ganglia, the thalamocortical motor circuits, or brain stem connections [*see Figure 1*].[1] Alterations of basal ganglia output at the level of the internal segment of the globus pallidus (GPi) and the subthalamic nucleus (STN) can lead both to impairment of voluntary movements and to involuntary movements.[2]

A pathophysiologic model of parkinsonism has been developed on the basis of animal studies [*see Figure 2*]. The biochemical hallmark of this syndrome is dopamine deficiency in the putamen and the caudate nucleus, which promotes excessive excitatory drive from the STN to the GPi and, in turn, abnormal output from the GPi.[2] Although initially, attention was focused on the increased inhibitory output from the GPi, it now appears that changes in the pattern of GPi output play a major role as well. That impairment of movement results from the excessive and abnormal GPi output is supported by studies demonstrating that treatment directed toward the GPi or STN (either surgery or deep brain stimulation) greatly improves the motor signs of Parkinson disease (PD).[3] Moreover, in patients with PD, the restoration of cortical activation during a movement task has been demonstrated by comparing positron emission tomography (PET) scans taken before and after treatment with pallidotomy (a procedure that surgically destroys the motor segment of the GPi) or with electrical stimulation of the STN.[4]

In contrast to hypokinetic disorders, hyperkinetic disorders appear to result from abnormally low and altered GPi output [*see Figure 2*].[5,6] Reduced STN output leads to a reduction in the thalamic inhibition by GPi. In hemiballismus, this stems from damage to the STN or its connections. In Huntington disease, it is the result of selective loss of striatal neurons that project to the external portion of the GPi or the substantia nigra.[7] Just as the same pathophysiologic changes in GPi and STN output are responsible for the disturbances of eye movements that occur in PD and Huntington disease, it is tempting to speculate that changes in the other basal ganglia–thalamocortical circuits contribute to behavioral, cognitive, and limbic signs and symptoms that parallel many of the motor symptoms and signs of these disorders.[8]

Hypokinetic Disorders

Parkinsonism is a clinical syndrome with multiple etiologies that is characterized by varying degrees of bradykinesia, tremor, rigidity, and postural instability.[9] The common denominator of all forms of parkinsonism is the reduction of striatal dopaminergic transmission in the nigrostriatal pathways. Parkinsonism may be subdivided into PD, which accounts for approximately 75% of cases, and atypical forms of parkinsonism, which account for the rest [*see Table 1*]. It is important to distinguish PD from these other forms of parkinsonism because prognosis and treatment differ significantly in each. Early etiologic diagnosis will become increasingly important in the coming years, with the emergence of therapies that may be specific to subtypes of PD.

PARKINSON DISEASE

Epidemiology, Etiology, and Genetics

PD is a progressive neurodegenerative disorder that afflicts more than one million persons in the United States, including 1% of the population older than 55 years.[10] A genetic contribution to familial forms of PD has now been firmly established; thus, these forms should be considered primary rather than idiopathic PD. Accordingly, PD is now considered a syndrome with many genetic and nongenetic causes. Genetic forms are associated with at least eight identified autosomal dominant and autosomal recessive mutations.[11] PD resulting from these mutations often has an earlier onset (occurring in patients younger than 50 years) and a severely progressive course compared with the typical sporadic form of PD, which has a mean age of onset of 60 years. In some of these disorders, signs of parkinsonism develop before age 20; these are termed juvenile parkinsonism. The genetic basis of most cases of early-onset disease is further supported by epidemiologic studies in monozygotic and dizygotic twins.[12]

One line of PD research has addressed the possible causative role of environmental factors (e.g., head injury, rural living, exposure to pesticides). Current thinking, which is influenced by evidence from animal models of PD, is that any role of environmental factors is almost exclusively limited to persons with a genetic predisposition.[13,14]

Autosomal dominant cases of PD are linked to two independent mutations in the α-synuclein gene.[11,15] The first α-synuclein mutation, *PARK1,* was found in a large Italian kindred and in three unrelated families of Greek origin.[15] This gene is located on chromosome 4q21 and codes for α-synuclein, the precursor protein of the non–β-amyloid component of Alzheimer disease found

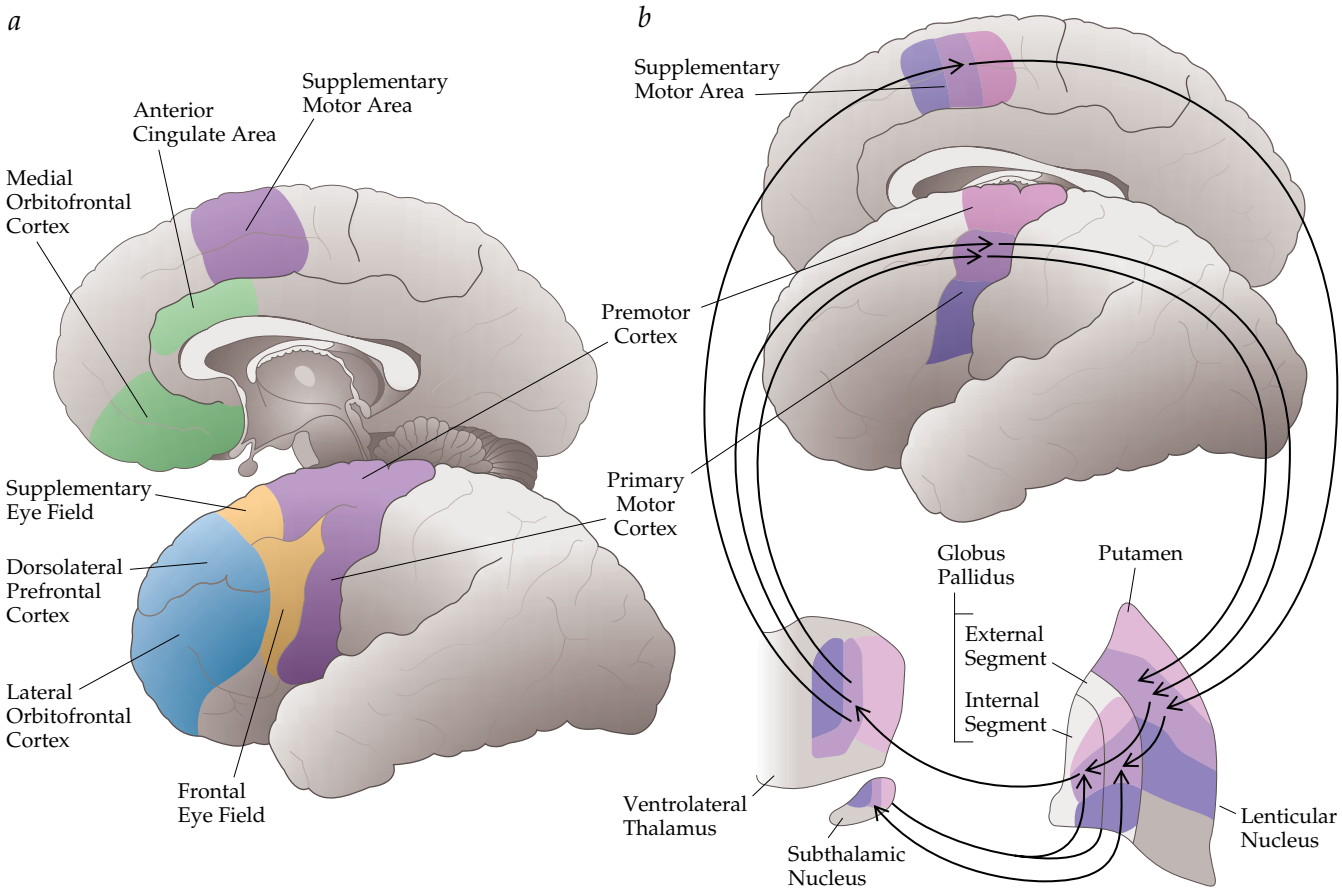


Figure 1 **Functional anatomy of the basal ganglia. (*a*) The basal ganglia are appropriately viewed as components of larger cortical-subcortical reentrant pathways that also include portions of the thalamus. These form a family of functionally segregated circuits that subserve and ultimately target the skeletomotor, oculomotor, associative (cognitive), and limbic cortices of the frontal lobes. (*b*) The basal ganglia comprise the corpus striatum (caudate nucleus and putamen), the globus pallidus, the subthalamic nucleus, and the two parts (pars compacta and pars reticularis) of the substantia nigra. These components are found deep in the cerebral hemispheres and nearby parts of the diencephalon and midbrain. The figure illustrates a simplified schema of the motor circuit. The segregated organization of the cortical-subcortical circuits permits simultaneous and independent parallel processing of diverse motor and nonmotor inputs (e.g., representation of face, arm, and leg). For each circuit, output (i.e., cortical impulses) from a specific cortical area passes through a unique portion of the striatum (e.g., putamen), the external and internal segments of the globus pallidus, the substantia nigra pars reticularis (not shown in this figure), the subthalamic nucleus, and the ventrolateral thalamus and returns to the specific frontal cortical area—in this case, the area related to motor function (i.e., premotor cortex, primary motor cortex, and supplementary motor cortex). The motor circuit originates from the precentral and postcentral sensorimotor fields, engages specific portions of the basal ganglia and motor thalamus, and ends back in the precentral motor fields of the frontal lobe. For example, voluntary movements are normally initiated in cortical areas that provide input to the basal ganglia and thalamus, which in turn modify these same cortical areas via the return projection through the motor circuit.**

in neuritic plaques. The protein is found in presynaptic terminals, and its distribution in the brain is identical to that of Lewy bodies. Its normal function is still unclear. For PD from the *PARK1* mutation, disease onset occurs before age 50; the disease generally has a rapid course and is often accompanied by dementia.

A second form of autosomal dominant, Lewy body–positive PD has been linked to the *PARK3* mutation on chromosome 2p13. *PARK3* may also play a role in sporadic PD, given that PD develops in only 40% of persons with this mutation and that the age of onset and response to levodopa therapy in these cases are indistinguishable from those in sporadic cases.[16]

There are at least six other rare autosomal dominant mutations that can cause parkinsonian syndromes similar to idiopathic PD. Clinically, these conditions resemble autosomal recessive juvenile parkinsonism (AR-JP), but they are often also associated with ataxia and amyotrophy.[16,17]

One form of AR-JP is linked to the *parkin* gene, which has been mapped to a large region of chromosome 6 (*PARK2*) and is found in Japanese and some European families.[18-20] In these patients, an initially robust response to levodopa is soon complicated by severe motor fluctuations. Pathologically, these patients do not have Lewy bodies but instead express Lewy fibrils and neurites that are immunohistochemically positive for α-synuclein. Two autosomal recessive forms, involving the *PARK2* and *PARK7* mutations, have atypical features and very early (juvenile) onset.

In the vast majority of patients with sporadic PD, no strong genetic determinant appears to play a role. It is believed, however, that these patients may share an as yet undefined genetic vulnerability that results in clinical illness only when the individual is exposed to as yet uncertain internal or external environmental factors.[12] Genetic vulnerability may be attributed to deficits in

mitochondrial oxidative phosphorylation genes (*OXPHOS* complex I), cytochrome P-450, and other polymorphisms, as well as to deficits in oxidative radical scavenging enzyme activity (e.g., glutathione transferases).

Certain epidemiologic factors have been linked to a reduced incidence of PD. These include coffee drinking, cigarette smoking, use of nonsteroidal anti-inflammatory drugs (NSAIDs), and estrogen replacement in postmenopausal women.[21-23]

Pathophysiology and Pathogenesis

The classic pathologic hallmarks of PD are degeneration of the dopaminergic cells of the substantia nigra pars compacta (SNc) and the presence of Lewy bodies in pigmented brain stem neurons, including the locus coeruleus, the pontine raphe nuclei, and the dorsal motor nucleus of the vagus nerve. The earliest pathologic changes may occur in the mesenteric plexus of the intestine and in the anterior olfactory nuclei.[24] The process spreads rostrally to the dorsal motor nucleus of the vagus and glossopharyngeal

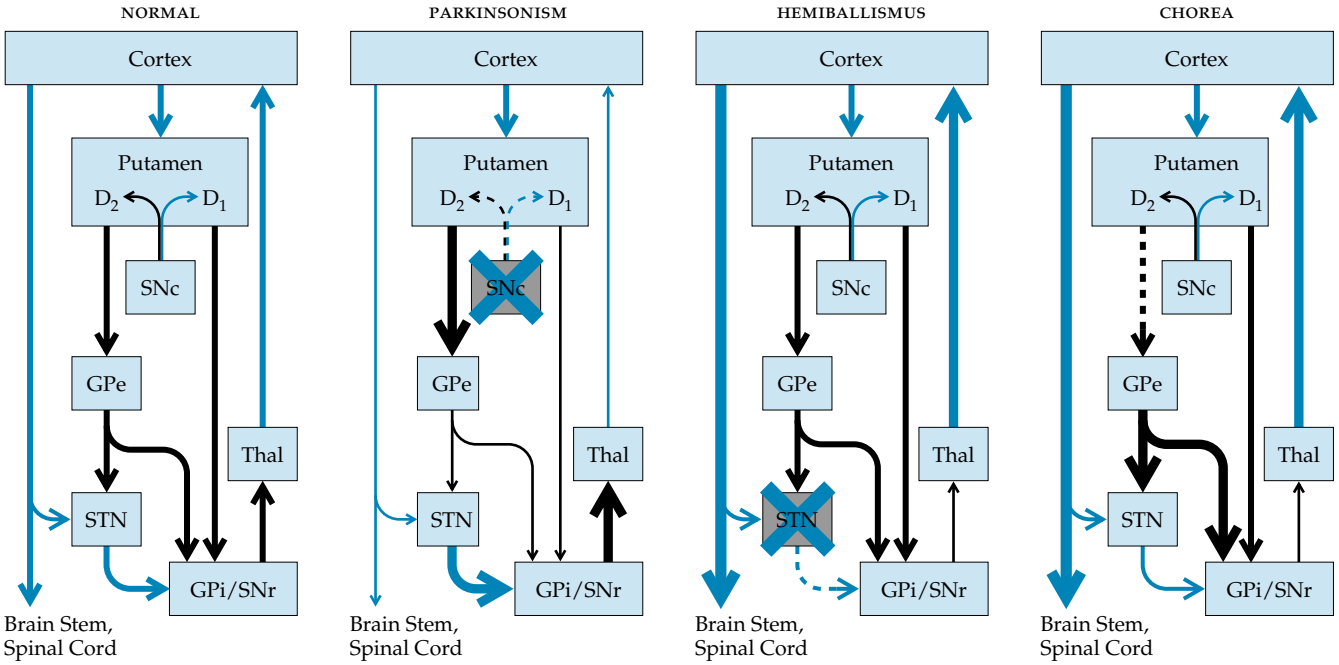


Figure 2 **This figure gives a more detailed illustration of striatal outflow and its functional significance in movement disorders. The corpus striatum serves as the input or receptive stage of the basal ganglia portion of the circuit, whereas the internal segment of the globus pallidus (GPi) and the substantia nigra reticularis (SNr) serve as output stages. The output nuclei of the basal ganglia are linked to the corpus striatum by a direct pathway between the corpus striatum and the GPi and the SNr and an indirect projection to the GPi via the external segment of the globus pallidus (GPe) and the subthalamic nucleus (STN). Other than the excitatory (glutamatergic) STN-GPi pathway (blue arrows), all projections of the direct and indirect pathways are inhibitory, or GABAergic (black arrows). (Note that the pathways from the cortex to the putamen, from the thalamus to the cortex, and from the cortex to the brain stem are also excitatory.) Considering the polarity of these connections and the fact that GPi/SNr output is tonic and inhibitory on thalamic neurons, it follows that activation of the direct pathway by cortical input (or dopamine stimulation) will reduce the activity of the output nuclei GPi and SNr and, thereby, disinhibit (by removal of inhibitory input) thalamocortical projection neurons, which will in turn facilitate movement. In contrast, activation of the indirect pathway, which has a net excitatory effect on GPi/SNr activity, will act to inhibit thalamocortical neurons or inhibit movements that may conflict with the intended movement. Too much activity of this pathway leads to movement inhibition in general and thus to parkinsonism. Dopamine is released by the terminals of the substantia nigra pars compacta (SNc) projections to the corpus striatum (putamen) and acts to modulate the activity in these circuits. Dopamine differentially influences the balance between the two pathways by inhibiting transmission in the indirect pathway via D_2 dopamine receptors on striatal neurons projecting to GPe and by facilitating transmission in the direct pathway via D_1 dopamine receptors on striatal neurons projecting to GPi. Thus, release of dopamine in the corpus striatum reduces basal ganglia output to the thalamus (Thal), whereas loss of dopamine increases output.**

Disturbances of the proper balance between the two circuits result in a variety of clinical syndromes. For example, in Parkinson disease, which is a hypokinetic disorder, dopamine deficiency increases the activity of the indirect pathway and thus the excitatory drive from STN to GPi/SNr. This condition stimulates inhibitory output from GPi and results in the increased inhibition of thalamocortical neurons, which renders cortical projection areas less responsive to inputs normally involved in the initiation and execution of movement. This process is illustrated by the reduced thickness of the excitatory arrows from the thalamus to the cortex and from the cortex to the brain stem and spinal cord. Changes in the pattern of GPi discharge are also a major factor in Parkinson disease.

In the hyperkinetic disorder hemiballismus, a lesion in the STN knocks out the excitatory drive from the STN to the GPi, which leads to loss of inhibition, or stimulation of the thalamocortical neurons, rendering cortical projection areas highly responsive to inputs involved in the initiation and execution of movement (illustrated by the thick excitatory arrows from the thalamus to the cortex and from the cortex to the brain stem and spinal cord). In the case of chorea, early selective loss of striatal neurons projecting to GPe via the indirect pathway (e.g., Huntington disease or proposed neuroleptic-induced toxicity in tardive dyskinesia) leads to the disinhibition of the GPe and thus excessive inhibition of striatal outflow (GPi/SNr). The result is a loss of inhibition of the excitatory thalamocortical pathway that consequently leads to the multiple, poorly synchronized movements of chorea.

Table 1 Differential Diagnosis of Parkinsonism

Primary	Idiopathic (sporadic) Parkinson disease Most common form; clinical expression may be influenced by so-called vulnerability genes Genetically mediated (primary) Parkinson disease Identified mutations (rare) PARK1 (α-synuclein; two mutations) PARK2 (*Parkin* gene mutation) with multiple polymorphisms PARK3
	Other neurodegenerative disorders α-Synuclein disorders Multiple-system atrophies (glial and neuronal inclusions) Striatonigral degeneration Olivopontocerebellar atrophy Shy-Drager syndrome Motor neuron disease with parkinsonian features Lewy body dementia (cortical and brain stem neuronal inclusions) Tauopathies (disorders with primary tau pathology) Progressive supranuclear palsy Corticobasal degeneration Frontotemporal dementia Amyloidopathies (disorders with primary amyloid pathology and secondary tau pathology and dementia) Alzheimer disease with parkinsonism (sporadic, amyloid precursor protein–related, presenilin 1– and presenilin 2– related)
	Genetically mediated disorders with occasional parkinsonian features Wilson disease Hallervorden-Spatz disease Chédiack-Higashi syndrome Spinocerebellar ataxia type 3 X-linked dystonia parkinsonism (DYT3) Huntington disease (Westphal variant) Prion disease
Secondary	Miscellaneous acquired conditions Vascular parkinsonism Normal-pressure hydrocephalus Catatonia Cerebral palsy Repeated head trauma (dementia pugilistica) Infectious and postinfectious diseases Postencephalitic Parkinson disease Neurosyphilis
	Toxins Carbon disulfide Carbon monoxide Cyanide Manganese Methanol MPTP (1-methyl-4-phenyl-1,2,4,6-tetrahydropyridine) *n*-Hexane?
	Drugs Neuroleptics (typical antipsychotics) Selected atypical antipsychotics [*see text*] Antiemetics (e.g., prochlorperazine, metoclopramide) Dopamine-depleting agents (e.g., reserpine, tetrabenazine) α-Methyldopa Lithium carbonate Valproic acid Fluoxetine
	Metabolic Conditions Hypoparathyroidism or pseudohypoparathyroidism with basal ganglia calcifications

nerves, with progressive ascending involvement of anatomically connected brain stem nuclei, the basal forebrain nuclei, and the SNc. Adding to the appeal of this hypothesis is the concept that the intestine and its plexus may serve as the portal of entry for environmental factors that may influence the expression of PD.

Ultimately, cell depletion occurs gradually in the SNc and other areas; parkinsonian signs begin to develop after 50% to 80% of midbrain dopamine neurons are lost and compensatory mechanisms fail [*see Figure 3*]. Involvement of the other pathologically involved nuclei may play a major role in the development of associated nonmotor aspects of PD (e.g., autonomic dysfunction, sleep disorders, depression).

Considerable attention has focused on the role of oxidative stress in PD and the damage that may be caused by free radicals produced by the metabolism of dopamine.[25,26] Mitochondrial dysfunction is another possible factor that may contribute to vulnerability to oxidative stresses. There is now solid evidence of mitochondrial dysfunction in the brain and platelets of patients with PD.[27,28] Defects in energy metabolism may interact with the protein products of mutations, leading to protein aggregation, low-grade inflammation, microglial activation, and further escalation of oxidative stress.[29] Other factors, such as accumulation of protein aggregates as a result of proteosomal dysfunction, may play a role as well in some cases of PD.[30,31]

For example, it has been postulated that the mutated α-synuclein protein of *PARK1* has an abnormal tertiary structure that makes it prone to self-aggregation and thus formation of the amyloidlike cores of Lewy bodies.[15] Mice that hyperexpress this gene develop granular intracytoplasmic inclusions similar to Lewy bodies.[15] In vitro, low concentrations of aggregated fibrillogenic fragments of α-synuclein are toxic to dopamine neurons,[32] and mutations that alter the expression of the α-synuclein protein make dopamine neurons highly vulnerable to oxidative stress.[33,34]

Clinical Features

The cardinal clinical features of PD are bradykinesia, muscular rigidity, and a rest tremor of 4 to 6 cycles per second (hertz [Hz]). The presence of two of the three cardinal signs and a robust clinical response to levodopa, the amino acid precursor of dopamine, are sufficient to make a probable diagnosis of PD.[9] Onset of symptoms on one side of the body and the presence of rest tremor further support a diagnosis of PD. Masklike facies [*see Figure 4*], decreased blinking, stooped posture, decreased arm swing when walking, and micrographia are often present early in the disease.

Diagnosis of PD in its early stages may be difficult unless the characteristic rest tremor is present. In the absence of tremor, the most common complaints are nonspecific; patients note weakness, fatigue, and muscle aches.

Parkinsonian tremor typically begins as a so-called pill-rolling motion of the fingers of one hand. The tremor usually spreads proximally in the arm and sometimes to the leg before crossing to the other side of the body over the next 1 to 2 years. The face, tongue, and jaw may be involved, but typically, the patient does not have voice tremor (lingual titubation). In contrast to essential and cerebellar tremors, parkinsonian tremor is usually present at rest and attenuates or disappears during movement. It may reappear after the movement is completed, even when a posture is sustained. However, postural or essential tremor may sometimes precede and coexist with parkinsonian rest tremor.

Bradykinesia (slowness of movement) and akinesia (paucity of movement) are among the most disabling features of PD be-

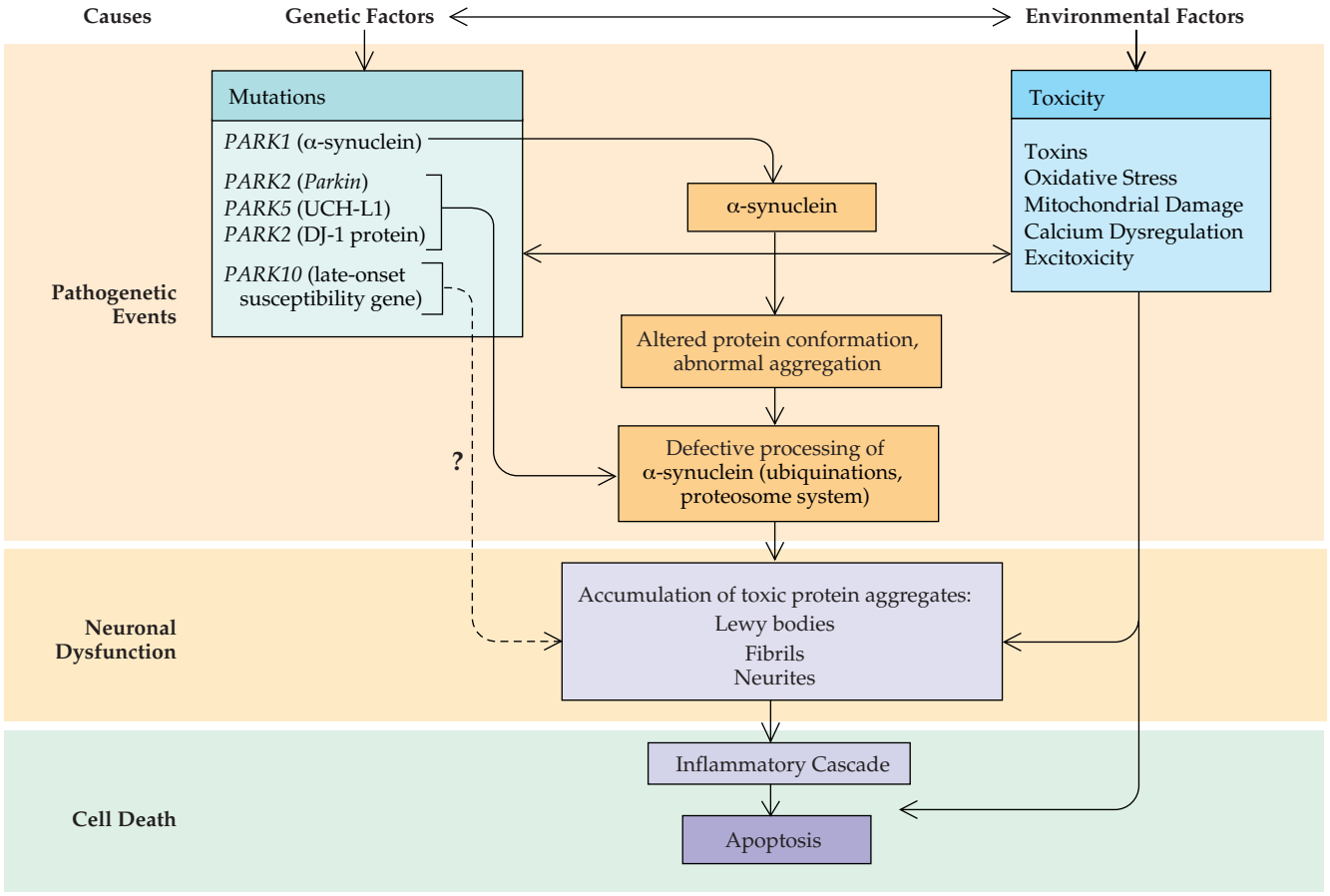


Figure 3 **Proposed cascade of pathogenic events leading to the death of neurons in Parkinson disease. Note the possible interaction between genetic and environmental factors.**

cause they interfere with all aspects of daily living, especially ambulation. Fine motor control is also affected, as manifested by impaired finger dexterity and micrographia. Hypophonia and monotonic speech are manifestations of bradykinesia affecting the cranial musculature.

Rigidity in PD represents a uniform increase in resistance to passive movement throughout the full range of motion of all muscle groups acting around a joint. Brief, regular interruptions of resistance during passive movement, corresponding to subclinical tremor, may give rise to a cogwheeling sensation. Combinations of rigidity and dystonia typically result in neck flexion, stooping, and a tendency to hold the arms in a flexed, adducted position.

Parkinsonian gait is usually associated with shuffling, short steps and a tendency to turn en bloc. In advanced stages of disease, gait initiation and turning become increasingly difficult. Advanced PD may be marked by festinating gait, which is characterized by stiff flexion at the knees and hips, an absence of arm swinging, and short steps (often on tiptoe) that accelerate in an apparent effort to catch up with the body's center of gravity. Postural instability is one of the most disabling and refractory features of parkinsonism and a major contributor to falls. Although decreased postural reflexes may be noted early in PD, significant postural instability is rarely a problem until the disease progresses to more advanced stages. Indeed, significant postural instability early in the course of the illness suggests a diagnosis other than PD.

Freezing, a feature of more advanced PD, occurs most commonly at the initiation of locomotion (start hesitation), when attempting to change direction, or upon entering a crowded or

Figure 4 **Face of a patient with Parkinson disease. Note the lack of expressivity and poor definition of the nasolabial fold. Scaling seborrhea is also common.**

narrow space (e.g., a doorway). Freezing is most commonly a manifestation of inadequate dosing or wearing off of medication but may occasionally occur in well-medicated patients (so-called on freezing). The presence of an ataxic or otherwise atypical gait, however, suggests an alternative diagnosis, such as normal-pressure hydrocephalus, multiple system atrophy with cerebellar involvement, or spinal stenosis.

Nonmotor disturbances In addition to motor symptoms, many patients with PD experience disturbances of cognition, mood, sensation, sleep, and autonomic function. Changes in cognition, behavior, and mood may result in part from involvement of nuclei in the brain stem and elsewhere, as well as disruption of the nonmotor circuits of the basal ganglia [*see Figure 1*].

Dementia is not a presenting or early feature of idiopathic PD, but dementia occurs in up to 30% of patients in advanced stages of disease.[35,36] It is more common in patients with the akinetic-rigid subtype of parkinsonism. Dementia is frequently preceded by depression or drug-induced hallucinations. Patients with parkinsonism and dementia typically have mixed pathologies, including senile dementia of the Alzheimer type and diffuse cortical Lewy bodies (i.e., Lewy body dementia).[36]

Depression affects as many as 50% of patients with PD and may develop at any stage of the illness, irrespective of disease severity.[37] Depression may be difficult to recognize because overt emotionality may be lacking and the psychomotor retardation of depression may closely resemble the hypokinetic aspects of parkinsonism. Anxiety often accompanies depression. Anxiety and dysphoria can occur episodically in relation to the wearing off of levodopa. Depression can be induced or aggravated iatrogenically by polypharmacy and poorly managed motor symptoms. Conversely, emerging depression may adversely affect motor symptoms. Personality changes in the form of apathy, lack of assertiveness, and indecisiveness are also common. It is important to rule out hypothyroidism and hypogonadism with low testosterone levels in men, which may not only contribute to the negative symptoms of PD but may make depression resistant to treatment.

Sleep disorders are common in PD.[38] A variety of problems, including insomnia and sleep fragmentation, can interfere with both the duration and quality of sleep. In addition, rapid eye movement sleep–related behavioral disorders (e.g., nocturnal vocalizations and bursts of motor activity) and restless legs syndrome contribute to sleep disruption. Nighttime reemergence of parkinsonian motor signs and involuntary movements (e.g., myoclonic jerks and periodic leg movements) may also disrupt sleep.[38] Vivid dreams and hallucinations, which may be side effects of antiparkinsonian therapy, may also contribute to sleep disruption.[39] In addition, sleep apnea and other sleep disturbances often occur in older PD patients. Daytime drowsiness and frequent napping may be typical signs of sleep disruption or an adverse effect of dopaminomimetic therapy, particularly dopamine agonists.

Other nonmotor disturbances in PD include orthostatic hypotension, constipation, sexual and urinary dysfunction, and seborrhea. All of these result from autonomic dysfunction.

Treatment

The goals of treatment in PD are to maintain function by controlling primary disability with symptomatic therapy and to minimize secondary disability (e.g., deconditioning, contractures, accelerated arthritis) with a consistent program of physical exercise.[40] Other goals include maintaining drug efficacy, avoiding drug-induced dyskinesias, preventing psychiatric complications, and slowing the progression of the disease. Goals still on the horizon include addressing currently unresponsive symptoms such as speech, balance disturbances, and autonomic dysfunction and eliminating common side effects such as nausea, orthostatic hypotension, and daytime sedation.

Levodopa Since 1969, when Cotzias demonstrated that oral administration of the dopamine precursor levodopa could alleviate parkinsonism,[41] levodopa therapy has been the mainstay of treatment. It greatly improves the quality of life and the lifespan of patients with PD. Although levodopa is toxic to dopamine neurons in vitro, it has not been shown to be toxic in vivo.[42] In a randomized, double-blind, placebo-controlled study of carbidopa-levodopa in early PD, clinical data suggested that treatment with levodopa for 40 weeks either slows the progression of PD or at least has a prolonged effect on disease symptoms; whereas neuroimaging studies suggested that either levodopa accelerates the loss of nigrostriatal dopamine nerve terminals or its pharmacologic effects modify the dopamine transporter.[43] Long-term levodopa administration clearly can kindle motor fluctuations and dyskinesias in PD.[44,45]

Dopamine agonists Concerns about side effects with long-term levodopa use have spurred interest in treating PD with dopamine agonists (see below). Several double-blind placebo-controlled studies have now established that compared with carbidopa-levodopa treatment, starting PD treatment with

Table 2 Carbidopa-Levodopa Preparations

Preparation	*Levodopa Dose Equivalency*	*Available Strengths (mg)*	*Initial Dosage*	*Comments*
Carbidopa-levodopa IR	100 mg (anchor dose)	10/100 25/100 25/250	25/100; 0.5–1 tablet t.i.d.	—
Carbidopa-levodopa CR	150 mg	25/100 50/200	50/200; one tablet b.i.d or t.i.d	Increased bioavailability with food; splitting the tablet negates the CR properties
Carbidopa-levodopa-entacapone	120 mg	12.5/50/200 25/100/200 37.5/150/200	25/100/200; one tablet b.i.d or t.i.d.	Do not split tablets

CR—controlled release IR—immediate release

dopamine-agonist monotherapy will delay the onset and reduce the severity of motor fluctuations.[46-48] A randomized, controlled comparison of initial PD treatment with pramipexole or levodopa found that initial treatment with pramipexole resulted in lower incidences of dyskinesias and wearing-off (i.e., shortened duration of therapeutic effect of individual drug doses), whereas initial treatment with levodopa resulted in lower incidences of freezing, somnolence, and edema and provided for better symptomatic control.[49] Of interest, a comparison of controlled-release carbidopa-levodopa (e.g., Sinemet CR) with regular (immediate-release) carbidopa-levodopa found that the incidence of dyskinesia was not significantly lower with the controlled-release formulation, which suggests that there are fundamental differences in the long-term consequences of dopamine receptor stimulation between levodopa and dopamine agonists.[50]

Initiating treatment The major determinants in the choice of initial therapy in PD are age of onset, the presence of dementia, the level of existing polypharmacy, and cost.[45,51] In cognitively intact patients with no significant medical problems, we favor initiation of dopamine-agonist monotherapy in sufficient doses to adequately control motor symptoms. This typically means giving higher doses than when dopamine agonists are used in combination with levodopa. Although symptoms such as orthostasis, dizziness, and nausea are more common with dopamine agonists than with levodopa, these can be avoided in most cases by slow titration. A common mistake with dopamine-agonist monotherapy is the failure to push the dose high enough to maintain adequate symptom control, leading to the premature addition of levodopa.

Most, if not all, patients with PD will need levodopa therapy at some point. When to introduce levodopa depends on the patient's ability to maintain function on dopamine-agonist monotherapy. Some patients resist starting levodopa for fear of its side effects. These patients should be advised against compromising their quality of life for the sake of avoiding levodopa therapy. In addition, they should be reassured that when levodopa is added, its dose is limited to that necessary to complement the effect of the dopamine-agonist, which remains the dominant component of therapy.

Still controversial is whether dopamine agonists provide a neuroprotective effect.[45] This hypothesis is currently being tested in a number of clinical trials in Europe and the United States.

Demented patients and those on multiple other medications are particularly prone to developing sedation, orthostasis, and drug-induced psychiatric symptoms. Initiation of therapy in this setting is aimed at reducing the incidence of central nervous system symptoms. To this end, in these patients we favor starting treatment with levodopa, with or without levodopa augmentation strategies (see below). Dopamine-agonist augmentation is introduced later as needed. In demented patients, anticholinergics and amantadine are to be avoided as much as possible to minimize CNS side effects. In cognitively intact patients, these agents may be given to minimize the use of levodopa and to control tremor and dystonic symptoms.

Levodopa formulations Levodopa is customarily given in combination with carbidopa, a peripheral dopa decarboxylase (DDC) inhibitor that blocks peripheral breakdown of levodopa, thus augmenting the bioavailability of levodopa while minimizing the side effects associated with peripheral conversion of levodopa to dopamine (notably, nausea). Carbidopa-levodopa is available in immediate-release formulations (10 mg carbidopa/100 mg levodopa, 25 mg/100 mg, and 25 mg/250 mg) and in controlled-release formulations (25 mg/100 mg, 50 mg/200 mg) [*see Table 2*]. Immediate-release carbidopa-levodopa is also available in an orally disintegrating form (Parcopa) that can be taken without water.

Carbidopa supplements (25 to 200 mg/day) are occasionally used to pretreat selected patients who have persistent nausea, presumably in response to inadequate plasma and carbidopa levels. For patients with persistent nausea, antiemetics such as trimethobenzamide and domperidone (available in Canada and Europe) may be well-tolerated and effective temporary measures. Because of its high cost, ondansetron is not a practical choice for long-term outpatient management. Other antiemetics, such as the dopamine receptor antagonists prochlorperazine and metoclopramide, should be avoided because they may significantly worsen parkinsonian symptoms.

Initiating treatment with either a regular or a controlled-release formulation of carbidopa-levodopa is acceptable. Although regular carbidopa-levodopa is better absorbed on an empty stomach, many experts recommend that early in treatment it be taken with meals to minimize nausea. In contrast, the absorption and bioavailability of the controlled-release formulations are increased when taken with meals.[52] The side effects of carbidopa-levodopa can be controlled by starting with a small dose (e.g., one-half tablet of carbidopa-levodopa, 25 mg/100 mg b.i.d.) and then increasing the dose as tolerance to the nausea develops. Dose adjustments should be made slowly because the full response to each increase in dose evolves over weeks. In the early stages of the disease, a typical daily dose should not exceed 300 to 400 mg/day of levodopa (as carbidopa-levodopa). Higher and more frequent doses may be necessary in patients with more advanced disease who experience wearing-off and other motor fluctuations.

Compared with immediate-release carbidopa-levodopa, controlled-release carbidopa-levodopa provides the convenience of one or two fewer daily doses and smaller oscillations in plasma levodopa levels, which results in a decrease in motor fluctuations in patients who experience mild to moderate wearing-off.[50,52] The bioavailability of levodopa in the controlled-release formulations is about 70% of that in the immediate-release formulations. The bioavailability of carbidopa in controlled-release formulations is about 50% that of immediate-release formulations. If controlled-release carbidopa-levodopa is selected as initial therapy, the initial dose is half of a 50 mg/200 mg tablet taken twice daily. Splitting the tablet negates the controlled-release properties; we do not recommend initiating therapy with controlled-release 25 mg/100 mg formulations because of the high incidence of nausea compared with immediate-release therapy. We prefer that patients start taking doses at breakfast and lunch only, maintaining 5- to 6-hour intervals between doses, rather than adopt a strict twice-daily schedule. The goal is to build a plateau of plasma levodopa levels early in the day and let the patient "ride" this plateau the rest of the day. Eventually, the patient will need late-afternoon and evening doses. Typical schedules of the various formulations of carbidopa-levodopa vary at different stages of disease and with different types of motor response abnormalities [*see Table 3*]. As the disease progresses and the level of motor-response abnormality increases, we rely less on controlled-release carbidopa-levodopa formulations and more on immediate-release formulations or on dopamine-agonist–dominant treatment.

Table 3 Strategies for the Use of Carbidopa-Levodopa Formulations

Time	Dosage (Tablets)				
	Monotherapy (CR 50 mg/200 mg)	Early Fluctuations		Late Fluctuations	
		CR 50 mg/200 mg	IR 25 mg/100 mg*	CR 50 mg/200 mg	IR 25 mg/100 mg
0700 hr	1	1	1	1	1
1100 hr	—	1	—	1	1
1200 hr	1	—	—	—	—
1500 hr	—	1	½	½	1½
1900 hr	½	½	—	—	1
2300 hr	—	1	—	—	1

*Booster doses.
CR—controlled release IR—immediate release

Levodopa augmentation One strategy for levodopa augmentation is the use of monoamine oxidase (MAO) inhibitors. MAO-B enzyme activity is one of the major catabolic pathways for dopamine in the CNS. Blocking its activity increases the intrasynaptic half-life of dopamine, leading to reduced motor fluctuations and tremor. Selegiline is a selective and irreversible MAO-B inhibitor that has a weak antiparkinsonian effect when used alone and a moderate effect when used as an adjunct to carbidopa-levodopa.[53] Selegiline is available as a 5 mg capsule or tablet. The usual dose is 5 mg with breakfast and lunch. In contrast to the nonselective MAO inhibitors (MAO-AB) and the MAO-A inhibitors, dietary tyramine restrictions are not necessary with selegiline. Insomnia is a significant side effect of selegiline, particularly when the drug is taken after midday. Older persons and those with significant but stable cardiac disease may benefit from daily doses as low as 2.5 mg. The dose should be reduced or the drug withdrawn if refractory hyperdopaminergic side effects occur (e.g., worsening dyskinesias, hallucinosis, confusion). A single European report citing increased mortality in elderly patients using selegiline[54] is inconsistent with all previous studies and with the long-term experience with selegiline in the Parkinson Study Group in North America.

Another levodopa augmentation strategy consists of blocking the activity of the enzyme catechol-O-methyltransferase (COMT). Together, COMT and DDC activity account for approximately 85% (10% and 75%, respectively) of the peripheral breakdown of levodopa.[55] Inhibition of COMT activity with tolcapone or entacapone can decrease oscillations in plasma levodopa levels. Both short-term and long-term administration of these agents slow the elimination of carbidopa-levodopa, thus increasing the area under the curve of plasma levodopa by 10% to 15% without changing the peak plasma level or the time to reach this peak.[55]

A number of double-blind, placebo-controlled studies have shown that this leads to a clinically important reduction in motor fluctuations and an increase in the time that the medication dose is effective ("on" time) of 1 to 2 hours a day.[56] In patients treated with tolcapone, the daily dose of carbidopa-levodopa may need to be decreased by 10% to 30% to avoid dyskinesias or other hyperdopaminergic side effects. In patients treated with entacapone, we generally wait until after the first few weeks of treatment to see whether any reduction in the dose of carbidopa-levodopa becomes necessary. Tolcapone is available in 100 and 200 mg tablets and is taken three times a day. Entacapone is available in 200 mg tablets and is administered as a single tablet with each dose of carbidopa-levodopa. Entacapone is now available in combination with carbidopa-levodopa (Stalevo 50 mg/100 mg/150 mg), which offers some convenience in reducing the number of pills taken.

COMT inhibitors can worsen levodopa-induced dyskinesias and cause nausea and diarrhea. Tolcapone carries a small risk of hepatotoxicity (including fatal hepatic necrosis) that has prompted a black-box warning from the Food and Drug Administration; for that reason, it is necessary to obtain informed consent before initiating therapy and to monitor liver function during therapy.[57] Rare cases of rhabdomyolysis associated with tolcapone use have also been described. Entacapone causes orange urine discoloration. Both agents can cause piloerection. Although the clinical antiparkinsonian effect of tolcapone appears to be more robust than that of entacapone, we favor initiating treatment with entacapone because of its greater safety.

There is little information comparing COMT inhibitors with dopamine agonists. Trials comparing tolcapone with the dopamine agonists bromocriptine and pergolide have been conducted, but those trials did not have sufficient power to detect clinically relevant differences between the agents.[58] Unlike dopamine agonists, COMT inhibitors are not effective for monotherapy, nor have they been shown to decrease the incidence of dyskinesias when introduced early in the course of treatment.

Dopamine agonists Dopamine agonists require neither transformation nor facilitated transport across the blood-brain barrier; they act directly on postsynaptic dopamine receptors. The available agents are the ergot alkaloids bromocriptine and pergolide and the newer nonergot alkaloids pramipexole and ropinirole [*see Table 4*].[46,48] The nonergot alkaloids are approved for early PD monotherapy and as adjuncts to carbidopa-levodopa in patients with advancing disease. Accumulating data indicate that pergolide may also be an effective monotherapeutic agent in PD.[48] However, pergolide has been associated with rare cases of cardiac valve dysfunction, most commonly tricuspid regurgitation.[59]

In a few head-to-head comparisons, the newer agents proved to be slightly superior to bromocriptine. There is no obvious advantage among the other three agents. Nonergot alkaloids are thought to cause less nausea and orthostatic hypotension than the ergots, but this difference appears to be clinically marginal. The new-

er agents offer no advantage over the old agents in terms of CNS side effects (e.g., confusion, hallucinations, and sleep disturbances). In fact, pramipexole has been associated with episodes of somnolence during the daytime—so-called sleep attacks—that have led to automobile accidents in patients.[60] This has led to the FDA's issuing a warning. There have been a few similar reports with ropinirole. Although the ergot alkaloids appear to be less likely to cause such episodes, we warn all our patients about this potential danger.

Amantadine and anticholinergics Some physicians prefer to treat early, mild PD with less potent agents, such as amantadine or anticholinergic agents, to delay the introduction of levodopa. There are no data to indicate that this levodopa-sparing maneuver prevents the emergence of dyskinesias.

Amantadine is particularly helpful in the treatment of drug-induced dyskinesias and thus may be especially valuable in later stages of the disease. Amantadine is available in 100 mg tablets and capsules and in 50 mg/5 ml syrup; the usual dosage is 200 to 400 mg/day in two or three divided doses. The mechanism of action of amantadine is unclear. In vitro, it can enhance dopamine transmission, it has anticholinergic effects, and it acts as a weak glutamate antagonist.[61] The side effects of amantadine are edema, erythema, and livedo reticularis. In older patients, it may aggravate confusion and psychosis. Amantadine needs to be used with caution and in smaller doses in patients with renal insufficiency.

Anticholinergics have long been used in PD, particularly for their effects on tremor, rigidity, and dystonia. They have limited effects on akinesia, bradykinesia, posture, and gait. Their major drawbacks, particularly in elderly patients, are memory impairment, hallucinations, visual blurring, aggravation of glaucoma, constipation, urinary hesitancy, and dry mouth. The last side effect can be exploited as a treatment for the drooling (sialorrhea) that affects many patients with PD. Specifically, patients with significant sialorrhea may benefit from glycopyrrolate (1 to 2 mg b.i.d.), which is an anticholinergic with low CNS penetrability. Other available anticholinergic agents include ethopropazine (50 to 200 mg/day), trihexyphenidyl (1 to 10 mg/day) and benztropine (1 to 4 mg/day). Tolerance of these agents decreases with advancing age. Of these agents, ethopropazine is a less potent antiparkinsonian agent but is more likely to be tolerated by elderly patients. In the United States, ethopropazine can be obtained through compounding pharmacies.

Treatment of advanced disease and its complications Clinical management of PD becomes increasingly complex as the disease advances and patients experience worsening gait and postural instability, increased freezing, falls, and the emergence of motor fluctuations and dyskinesias. Equally disabling are drug-induced confusional spells, hallucinations, and psychosis. Referral to a movement-disorder specialist is indicated when disease progresses to this point.[62]

Motor fluctuations Wearing off of levodopa's clinical antiparkinsonian benefit after 3 to 4 hours is the earliest sign of motor fluctuations. Giving antiparkinsonian drugs before rather than with meals or in a slurry (liquid levodopa) helps to bypass the erratic and partial drug absorption that contributes to this problem. Similarly, first-pass metabolism is avoided through the use of the orally disintegrating formulation of carbidopa-levodopa. Immediate-release carbidopa-levodopa can be combined with the controlled-release formulation to provide a better bal-

Table 4 Dopamine Agonists for Parkinson Disease

Class	*Agent*	*Dose Equivalent to Levodopa Anchor Dose**	*Starting Dosage*	*Titration Schedule†*	*Typical Maintenance Dosage for Single-Agent Therapy*	*Typical Maintenance Dosage As Adjunct to Levodopa*	*Comments*
Nonergot alkaloids	Pramipexole	1 mg	0.125 mg t.i.d.	Increase to 0.25 mg t.i.d. after 1 wk, then increase weekly in increments of 0.25 mg t.i.d.	Maximum dosage 1.5 mg t.i.d.	Maximum dosage 1.5 mg t.i.d.	Renal metabolism; dose adjustments needed in renal insufficiency; occasionally associated with sleep attacks
	Ropinirole	5 mg	0.25 mg t.i.d.	Weeks 1–4, increase weekly in increments of 0.25 mg t.i.d. to a dosage of 1 mg t.i.d.; after week 4, increase in weekly increments of 1.5 mg/day to a dosage of 9 mg/day, then of 3 mg/day to maximum dosage of 24 mg/day	12–24 mg/day	3–16 mg/day	Hepatic metabolism; potential drug-drug interactions; occasionally associated with sleep attacks
Ergot alkaloids	Bromocriptine	2 mg	1.25 mg b.i.d. or t.i.d.	Increase by 2.5 mg/day every 2–4 wk	7.5–15 mg/day	3.75–7.5 mg/day	Rare reports of pulmonary and retroperitoneal fibrosis; relative incidence of sleep attacks not well studied
	Pergolide	1 mg	0.05 mg/day, t.i.d. at mealtimes	Every 2–5 days	3–4.5 mg/day	1.5–3 mg/day	Rare reports of valvular heart disease; fewer reports of sleep attacks compared with nonergot agents

*100 mg immediate-release levodopa; see Table 2.
† Because of lower tolerance, some patients will require slower titration.

ance between prompt and sustained antiparkinsonian responses [*see Table 2*]. Dopamine agonists and selegiline have a longer half-life than carbidopa-levodopa, and combination therapy that includes one of these agents produces a smoother clinical response than carbidopa-levodopa alone in patients who experience motor fluctuations.

Another strategy involves shortening the dosing intervals. However, regimens that involve dosing more frequently than every 4 hours are difficult to manage clinically; such regimens should probably be undertaken in specialty movement-disorder clinics.[62]

Neuropsychiatric symptoms Depression in patients with PD typically responds well to conventional antidepressants (e.g., tricyclics, selective serotonin reuptake inhibitors [SSRIs], venlafaxine, and bupropion). Short-acting rather than long-acting SSRIs are preferred, because there are reports that fluoxetine, a long-acting SSRI, may produce extrapyramidal symptoms (EPS). Although all SSRIs are probably effective in PD, we find that citalopram (10 to 30 mg/day), sertraline (50 to 150 mg/day), and paroxetine (10 to 30 mg/day) are particularly well tolerated. Limited experience with bupropion (50 to 200 mg/day) and sustained-release venlafaxine (37.5 to 150 mg/day) suggests that these agents are also well tolerated and may be less sedating. Concerns about the potential for hyperserotoninergic reaction (delirium with myoclonus and hyperpyrexia) stemming from the combination of selegiline and SSRIs in PD appear to have been grossly exaggerated.[62]

In patients who fail to respond to antidepressant medication, it is important to rule out other causes of refractory depression, such as hypothyroidism. Patients with PD who suffer from refractory depression or are intolerant of oral antidepressants are candidates for electroconvulsive therapy (ECT). In addition to alleviating depression, ECT may also improve parkinsonian motor symptoms.[63]

Patients with PD may undergo slow but profound drug-induced personality changes, which may be expressed as erratic, temperamental, unreasonable, and demanding behaviors; self-centeredness; and apparent disregard for the needs of others. These may be a prelude to depression or psychotic symptoms that are either drug induced or secondary to emerging dementia. Drug-induced psychosis is typified by visual hallucinations with retention of insight. Auditory hallucinations suggest coexisting psychotic depression or dementia, but they may be a side effect of anticholinergic medications.[64] In many instances, disturbing cognitive and psychiatric symptoms will cease with elimination of anticholinergics or amantadine. Some patients, however, require reduction in doses or elimination of drugs, typically in the following order: selegiline, nocturnal doses of dopamine agonists or controlled-release carbidopa-levodopa, and regular carbidopa-levodopa. If the patient improves after only a modest adjustment in polypharmacy, the impact on the parkinsonian motor symptoms will range from negligible to positive. If the patient requires a drastic reduction in antiparkinsonian therapy, the resulting aggravation of motor symptoms is likely to be intolerable.[64] Levodopa drug holidays are no longer recommended, because several reports indicate that they are potentially dangerous and provide no significant long-term benefit.

When the reduction in antiparkinsonian polypharmacy has intolerable consequences, an antipsychotic drug may be necessary. Conventional antipsychotics are poorly tolerated because of associated EPS and worsening of parkinsonism. Atypical antipsychotics have a higher ratio of serotonin 5-HT_{2a} to dopamine D_2 blockade and, compared with conventional antipsychotics, a lower incidence of EPS. Of the atypical antipsychotics, clozapine has been shown to be more effective than placebo in the treatment of drug-induced hallucinations in PD in two double-blind, placebo-controlled studies.[65] Clozapine may have additional beneficial effects on tremor, dystonia, and dyskinesias. Associated side effects include dizziness, orthostatic hypotension, sialorrhea, and confusion. Although most PD patients tolerate small doses of clozapine (12.5 to 75 mg in the evening), its use has been tempered by the 1% risk of agranulocytosis, which requires frequent monitoring of the leukocyte count.

The atypical antipsychotic quetiapine appears to be the first-line drug for the treatment of all psychotic symptoms in PD. It is not associated with hematologic or other serious toxicity and has a low incidence of EPS across the therapeutic spectrum. In a series of open-label studies, it was shown to be effective and well tolerated in PD.[65] The median range of effective doses is 50 to 75 mg/day, with most of the dose administered at night. This dose is significantly less than that typically used in schizophrenia (400 to 600 mg/day). The side effects are similar to those of clozapine except for the lack of hematologic toxicity or sialorrhea. If neuropsychiatric symptoms are not well controlled with quetiapine, the patient is switched to clozapine.

Risperidone is the second oldest atypical antipsychotic. In small doses (0.25 to 1 mg/day), it is an effective antipsychotic in PD patients who have drug-induced hallucinations. In a few open-label studies, however, it appeared to aggravate motor symptoms in many patients over a period of weeks to months.[65] This worsening was typified by increasing akinesia, decreased walking, and other signs of EPS. Olanzapine is not well tolerated by patients with PD.[66] All atypical antipsychotics carry warnings of potential hyperglycemia; monitoring of blood glucose levels is recommended.

Treatment of dementia in PD with the dual cholinesterase inhibitor rivastigmine has been studied. In a 24-week, placebo-controlled study, rivastigmine use was associated with moderate improvements in PD-associated dementia; however, it was also associated with higher rates of nausea, vomiting, and tremor.[67]

Neuroprotective therapy Slowing the progression of PD through neuroprotective therapy is a major focus of current research.[68] Epidemiologic studies suggest that the long-term use of cyclooxygenase inhibitors or of estrogen replacement in postmenopausal women may delay or prevent the onset of PD through yet unclear mechanisms. Coffee drinking has also been associated with a reduced incidence of PD, as has cigarette smoking. Current treatment strategies involve interrupting the cascade of biochemical events that leads to death of dopaminergic cells [*see Table 5*]. The first such clinical trial in PD was the large multicenter Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP) study, in which selegiline monotherapy delayed the need for levodopa therapy by 9 to 12 months in newly diagnosed patients.[69] Most evidence indicates that this delay resulted from a mild symptomatic effect of selegiline. Long-term follow-up of the DATATOP cohort revealed that patients who remained on selegiline for 7 years experienced slower motor decline compared with those who were changed to placebo after 5 years. Interestingly, patients in the 7-year patient group were more likely to develop dyskinesias but less likely to develop freezing gait. A metabolite of selegiline, desmethylselegiline, has been shown in laboratory studies to have power-

Table 5 Selected Neuroprotective Treatments for Parkinson Disease

Pathogenetic Factor	*Intervention*	*Specific Agents*
Oxidative stress	Antioxidants	Vitamin E, vitamin C, iron chelators
	MAO-B inhibitors	Selegiline, rasagiline
Mitochondrial dysfunction	Bioenergetic agents	Coenzyme Q10
Excitotoxicity	Antiglutamatergic agents	NMDA receptor antagonists
Inflammation	Anti-inflammatory agents	COX-2 inhibitors
Neuronal dysfunction	Trophic factors	Glial cell line–derived neurotrophic factor
Apoptosis	Antiapoptotic agents	Dopamine agonists, caspase inhibitors, minocycline, propargylamines

COX—cyclooxygenase MAO—monoamine oxidase NMDA—*N*-methyl-D-aspartate

ful neuroprotective effects. Clinical trials to test this agent are now in progress.

In another pilot study, coenzyme Q10, an antioxidant and a cofactor of complex I of the mitochondrial oxidative chain, appeared to delay progression of early disability in PD.[70] A large multicenter trial is now under way. Other potentially neuroprotective agents under investigation are acetyl-levo-carnitine and creatine monohydrate. A large controlled study of the antiglutamatergic agent riluzole was prematurely discontinued after a futility analysis revealed little effect on progression of symptoms.[71] As noted, dopamine agonists are also being studied for a neuroprotective role, based on their ability in vitro to decrease dopamine turnover, scavenge free radicals, and interfere with proapoptotic cell signals. Other promising agents include nitric oxide synthase inhibitors, antiapoptotic agents such as Jun N-terminal kinase inhibitors, and the antibiotic minocycline.[71] Minocycline can inhibit microglial activation in vitro and interrupt apoptosis by inhibiting caspase 1 and 3, which are involved in the enzymatic processing of α-synuclein.

Surgical treatment The past decade has witnessed a renaissance in the surgical treatment of PD and other movement disorders. This has been motivated first by the fact that after 5 or more years of pharmacologic treatment, many patients develop significant drug-induced motor fluctuations and dyskinesias. Second, major advances in understanding the pathophysiologic basis of parkinsonism have provided a clearer rationale for the effectiveness of surgical procedures and guidance for targeting specific structures, such as the GPi and the STN [*see Figure 2*].

The selection of suitable patients for surgery is critical, because in general, patients with atypical PD do not benefit from surgery. Candidates for surgery must have the following: (1) a clear diagnosis of idiopathic PD, (2) a good initial response to levodopa, and (3) significant intractable symptoms of PD or (4) drug-induced dyskinesias and wearing-off. The major contraindications to surgery are atypical PD, dementia, major psychiatric illness, and substantial medical comorbidities. Age is not a contraindication, but older patients derive less benefit from surgery. Patients with clinical manifestations such as unresponsive features, postural instability and falling, hypophonia, micrographia, drooling, and autonomic dysfunction are unlikely to benefit from surgery. In general, the benefits of surgery are unlikely to exceed the best benefits of antiparkinson medication. The major benefit of surgery is the elimination of dyskinesias and "off" periods (i.e., periods when medication is ineffective) and production of a stable "on" state. The decision for surgery should be made by a movement disorder neurologist who is part of a team that includes a neurosurgeon with fellowship training in functional neurosurgery, a psychiatrist, a neuropsychologist, and trained technicians.

Use of procedures that involve the creation of surgical lesions (e.g., pallidotomy or thalamotomy) has decreased greatly since the introduction of deep brain stimulation. In deep brain stimulation, a lead wire is surgically implanted in the GPi or STN and connected to a pulse generator that is implanted subcutaneously, generally on the chest wall near the clavicle. The patient can activate the system by passing a small handheld magnet over the generator. This system was approved by the FDA in August 2004 for patients with PD.

The major advantages of deep brain stimulation are that it is less invasive than the creation of surgical lesions and that the system may be adjusted to best effect after implantation. Although the choice between STN and GPi as targets for deep brain stimulation has shifted toward the STN, the available evidence suggests that both sites are effective for all the cardinal features of PD, as well as for dyskinesias and motor fluctuations. Several clinical trials are now under way to compare these two targets. Although bilateral surgery is generally necessary for patients with advanced disease and those with significant bilateral manifestations, unilateral stimulation is appropriate for patients with asymmetrical disease. Postoperative reductions in drug dosages appear to be easier with STN than GPi approaches.

The mechanism of action of deep brain stimulation remains controversial. Because both ablation and stimulation of a given target have a similar clinical effect, it was assumed that stimulation caused a functional blockade. With both approaches, it is probable that the remaining motor systems in the brain stem, thalamus, and cortex are able to compensate more effectively for the abnormal activity associated with the parkinsonian state. Whatever the mechanism, it is clear that these approaches can offer impressive results in properly selected patients with PD, as well as in patients with other movement disorders.

Although deep brain stimulation can be very effective, serious complications occur in 1% to 2% of patients who receive the device.[72] Even in the best hands, approximately 15% of patients require a second operation to treat complications or improve efficacy.

Although neurotransplantation of dopamine-producing fetal cells once seemed promising, the results from two large con-

trolled clinical trials have proved considerably disappointing.[73] The first study showed only modest benefit in patients younger than 60 years, and it showed no benefit in those older than 60 years. Moreover, a number of patients developed dyskinesias, which occurred when the patients were off medication. The second study showed similar findings with regard to benefit and the development of dyskinesias. Because of this, as well as the considerable obstacles to obtaining sufficient fetal tissue and the opposition to the use of fetal tissue on ethical grounds, this approach is now viewed as purely investigational. Conceivably, other sources of replacement dopaminergic cells (e.g., carotid body cells, stem cells, or encapsulated, genetically engineered cells capable of producing levodopa, dopamine, or trophic factors) may prove more successful. An open-label trial of direct infusion of glial cell–derived neurotrophic factor (GDNF) to the putamen in a small number of patients with PD raised hopes for this approach; however, in a subsequent randomized, double-blind trial of GDNF administered through an implanted intracerebroventricular catheter, patients showed no improvements in parkinsonism.[74]

OTHER FORMS OF PARKINSONISM

Pathologic studies reveal that the error rate in the clinical diagnosis of PD can be as high as 25%, which suggests that a number of other entities can closely mimic this syndrome [*see Table 1*]. Most important in the differential diagnosis are drug-induced parkinsonism, degenerative disorders with parkinsonlike features, vascular parkinsonism, and normal-pressure hydrocephalus. Scenarios under which an alternative diagnosis should be considered include the following: recent exposure to drugs or toxins; a history of, or significant risk factors for, cerebrovascular disease; symmetrical symptoms or abrupt onset of symptoms; early dementia; prominent postural instability from the outset; and cerebellar or pyramidal signs or autonomic dysfunction.

Drug-Induced Parkinsonism

Drug-induced parkinsonism (DIP) closely resembles clinical PD, except that the rest tremor is generally less prominent and more symmetrical in DIP than in PD. Postural tremor is also more prominent in DIP. DIP is commonly caused by the use of dopamine blockers such as neuroleptics and antiemetics.[75] Other drugs with different or unknown mechanisms include lithium carbonate and, to a lesser extent, SSRIs, divalproex, and selected calcium channel blockers used in Europe. Among the dopamine blockers, the most common offending agents are conventional antipsychotics and metoclopramide.

Atypical antipsychotics can also cause parkinsonism in a dose-dependent manner, with risperidone and olanzapine being more frequent offenders than clozapine and quetiapine. In parkinsonism induced by atypical antipsychotics, onset is delayed and there is less tremor than in DIP caused by typical antipsychotics. The most prominent features of DIP caused by atypical antipsychotics are increased akinesia and ataxia. This condition is often missed in the setting of chronic illnesses such as dementia and PD with dementia, because the delayed deterioration in the patient's condition is assumed to be secondary to the progression of the underlying disease.

DIP can also be induced by the long-term administration of such antihypertensive agents as reserpine, which depletes presynaptic dopamine stores, and α-methyldopa, which can inhibit the synthesis of dopamine by blocking the transport of levodopa through the blood-brain barrier.

Exposure to certain toxins can also lead to a parkinsonian state [*see Table 1*]. Any young adult presenting with parkinsonism should be questioned about the use of street drugs, because mistakes in the illicit synthesis of certain narcotics can result in their contamination with MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), which may cause an irreversible parkinsonian syndrome.

DIP is treated by stopping the offending agent. Symptoms usually resolve within days or weeks. When DIP persists despite withdrawal of a drug that is not known to produce permanent neuronal damage, presumably the drug exposure simply exposed a parkinsonian disorder that was in the process of developing. In DIP precipitated by atypical antipsychotics, neuroleptics can be substituted for the offending drug, if necessary. In patients who need to continue using the offending agent and whose symptoms are not incapacitating, DIP can be alleviated by anticholinergics, amantadine, and dopaminomimetics.

Vascular Parkinsonism

Patients with vascular parkinsonism exhibit an akinetic-rigid lower-body parkinsonism with gait and balance problems that include a short-stepping gait (marche à petits pas).[76] Upper body rigidity and tremor are less prominent than in idiopathic PD, and action tremor is more common than rest tremor. Most patients report an acute to subacute onset of symptoms and have a history of cerebrovascular disease, diabetes mellitus, or hyperlipidemia. Accompanying neurologic findings in vascular parkinsonism that are typically absent in idiopathic PD include upper motor neuron signs, frontal release signs, pseudobulbar palsy, and dementia. Other diseases that occasionally produce a similar picture include normal-pressure hydrocephalus, chronic subdural hematoma, inflammatory vasculitides, the antiphospholipid syndromes, and hyperhomocysteinemia. Rare entities such as leukodystrophies and demyelinating disorders can present similarly. The diagnosis is made on the basis of the clinical picture, the presence of risk factors, and evidence on magnetic resonance imaging of multiple strokes, a hypertensive lacunar state, or a multifocal or diffuse microangiopathy [*see Figures 5a and c*]. A postmortem examination is necessary for definitive diagnosis, given the absence of definite disease markers.

No satisfactory therapeutic options are available for vascular parkinsonism. Response to dopaminomimetic agents is limited; anticholinergics and amantadine frequently cause confusion and delirium.

Multiple-System Atrophy

The multiple-system atrophies (MSAs) are a clinically diverse group of neurodegenerative disorders of unknown etiology characterized by varying degrees of parkinsonism and corticospinal, cerebellar, and autonomic dysfunction[77] [*see Figures 5a and b*]. Widespread degeneration is typically found in the basal ganglia, spinal cord, brain stem, and cerebellum. A unique and distinctive pathologic feature of the MSAs is the presence of α-synuclein–positive glial cytoplasmic (and to a lesser extent neuronal) inclusions in the absence of Lewy bodies. The role of α-synuclein–positive (and ubiquitin-positive) inclusions in the pathophysiology of these illnesses remains unclear.

Clinical features The average age of onset of MSAs is 50 years (range, 33 to 76 years). Median survival is 6 years. Men are affected more often than women. The presentation is highly varied and may begin with signs of parkinsonism (46%), cerebellar dysfunction (< 25%), or autonomic dysfunction. The signs of au-

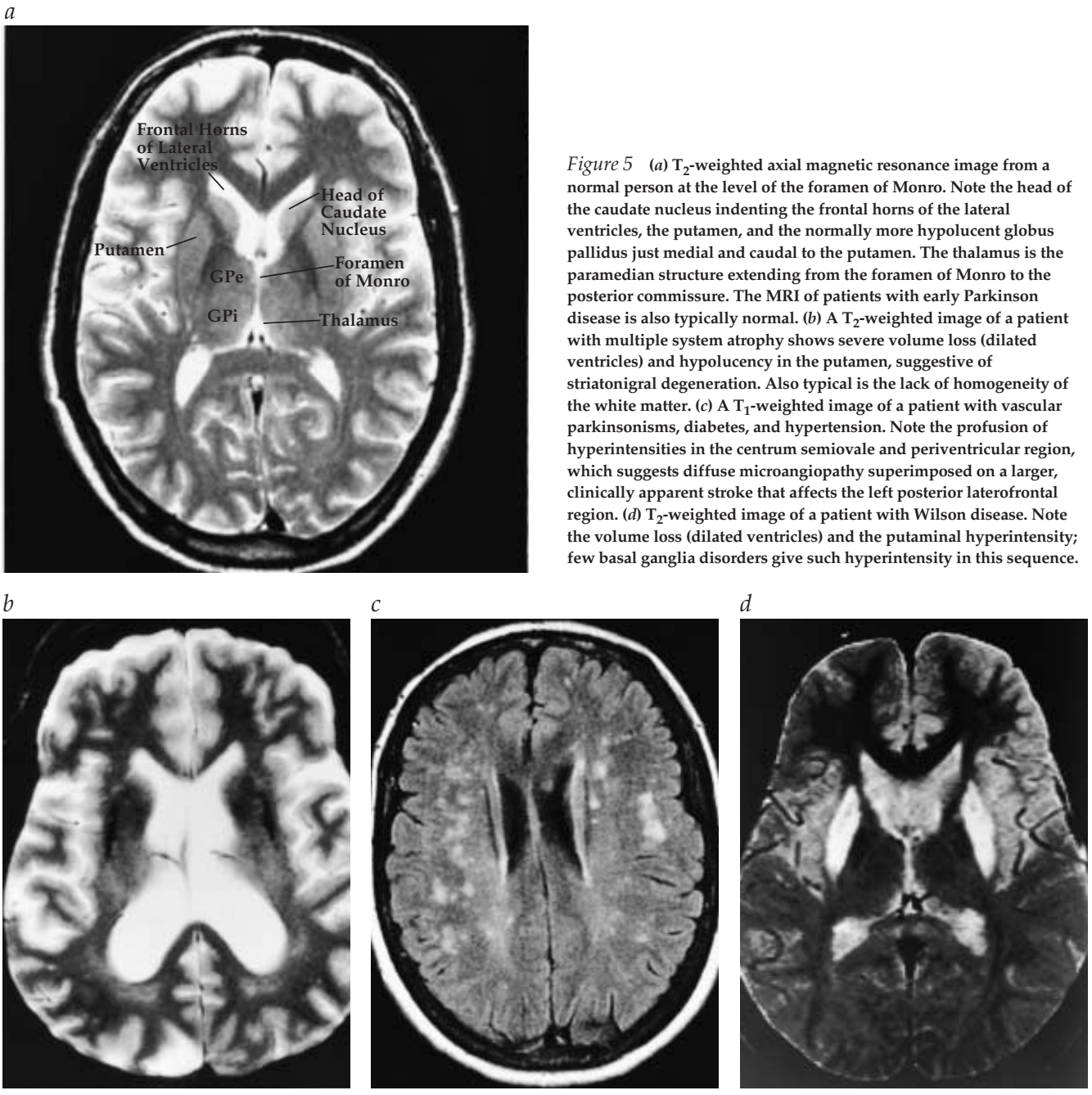


Figure 5 **(*a*) T_2-weighted axial magnetic resonance image from a normal person at the level of the foramen of Monro. Note the head of the caudate nucleus indenting the frontal horns of the lateral ventricles, the putamen, and the normally more hypolucent globus pallidus just medial and caudal to the putamen. The thalamus is the paramedian structure extending from the foramen of Monro to the posterior commissure. The MRI of patients with early Parkinson disease is also typically normal. (*b*) A T_2-weighted image of a patient with multiple system atrophy shows severe volume loss (dilated ventricles) and hypolucency in the putamen, suggestive of striatonigral degeneration. Also typical is the lack of homogeneity of the white matter. (*c*) A T_1-weighted image of a patient with vascular parkinsonisms, diabetes, and hypertension. Note the profusion of hyperintensities in the centrum semiovale and periventricular region, which suggests diffuse microangiopathy superimposed on a larger, clinically apparent stroke that affects the left posterior laterofrontal region. (*d*) T_2-weighted image of a patient with Wilson disease. Note the volume loss (dilated ventricles) and the putaminal hyperintensity; few basal ganglia disorders give such hyperintensity in this sequence.**

tonomic failure include orthostatic hypotension, urinary symptoms, constipation, impotence, and loss of sweating. Parkinsonism occurs in about 90% of patients at some time during the illness.[78] Autonomic symptoms are the initial feature in less than half of patients, but they develop later in virtually all patients. A high percentage of patients with MSA have upper motor neuron signs (e.g., spasticity and Babinski signs).[78]

MSAs have been grouped into three broad categories: striatonigral degeneration (MSA-P), olivopontocerebellar atrophy (MSA-C or OPCA), and Shy-Drager syndrome. Patients with relatively pure parkinsonism are designated as having striatonigral degeneration. Those who express a combination of parkinsonism and other signs, such as ataxia, upper motor neuron and corticobulbar involvement, myoclonus, oculomotor abnormalities, peripheral neuropathy, and deafness, fit into the grouping of OPCA. The OPCAs are a markedly heterogeneous group that includes both hereditary and sporadic forms of the disease. If parkinsonian features are associated with prominent signs of autonomic failure, a diagnosis of Shy-Drager syndrome is justified.

Treatment In general, the parkinsonian features of the MSAs respond only modestly to antiparkinsonian medications. The efficacy of dopaminomimetic therapy is often limited by orthostatic hypotension. Treatment of orthostatic hypotension in patients with MSAs (as well as in those with PD and other disorders associated with symptomatic orthostatic hypotension) includes simple measures such as changing position slowly, increasing salt and fluid intake, and avoiding overheated rooms, large meals, and alcohol. Other measures include wearing support stockings, sleeping in the reverse Trendelenburg position,

raising the hematocrit if it is below 32%, and adjusting other drug therapies that may affect blood pressure. Because blood pressure tends to be higher in the afternoon, patients can learn to conduct most activities of daily living in the afternoon. If these measures fail, pharmacologic strategies to deal with the autonomic insufficiency may be effective, including the use of vasopressors such as midodrine.

Other symptoms, such as impotence, urinary urgency and incontinence, and autonomic storms with profound diaphoresis, are more difficult and sometimes impossible to treat pharmacologically. Impotence occasionally responds to yohimbine or to the more contemporary agents sildenafil, vardenafil, and tadalafil. Penile implants may help but should not be pursued until orthostatic changes are well controlled. Constipation can be treated with high-dose fiber, hydration, bisacodyl, senna concentrate, or lactulose. Urinary urgency and frequency may respond to oxybutynin, tolterodine tartrate, and hyoscyamine sulfate; incontinence requires intermittent catheterization, adult diapers, or both. Gastroparesis can be improved with metoclopramide—which may, however, exacerbate parkinsonian symptoms—or with erythromycin ethylsuccinate, erythromycin lactobionate, or domperidone (available in Canada and Europe); all are taken before meals.

Disorders Associated with Primary Tau Pathology

Tau is a microtubule-associated phosphoprotein that promotes tubulin polymerization and the stabilization of microtubules in neurons. Disorders of tau protein lead to neurodegeneration by causing the intraglial and intraneuronal accumulation of tau protein–positive neuropil threads and secondary neuronal damage in the form of achromatic neurons and tangles. These tau protein inclusions are thought to cause neurotoxic damage to gray and white matter, particularly in basal ganglia and the frontotemporal cortex. Several mutations in the gene that codes for tau protein on chromosome 17 have been identified. Most of these are associated with familial frontotemporal dementia [*see 11:XI Alzheimer Disease and Other Dementing Illnesses*]. Linkage disequilibrium studies suggest that progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) may be autosomal recessive conditions associated with polymorphisms that map to the *tau* gene in chromosome 17.[79] For instance, a missense mutation in exon 13 of the *tau* allele R406W has been linked to a dominantly inherited form of PSP in Europe but not to 31 unrelated individuals with PSP or CBD in the United States.[80,81] Pathologically, these conditions share overexpression of the same tau protein haplotype in the affected regions of brain.

Progressive supranuclear palsy PSP is perhaps the most common of the so-called Parkinson-plus syndromes. Like PD, PSP begins in the sixth or seventh decade of life; unlike PD, it progresses to death in 5 to 10 years, mostly from complications of aspiration.[82] Early on, nonspecific complaints of visual impairment, dizziness, and unsteadiness are common. Later in the disease course, more specific complaints of slowness, falls, and deterioration of speech and swallowing begin to appear. Tremor is distinctly uncommon. A staring look (because of partial lid retraction) with a furrowed brow, deep facial lines, and a hyperextended head provides an image distinct from the masked facies and flexed posture of PD [*see Figure 6*]. Supranuclear control of voluntary vertical gaze is impaired first, followed by difficulties with accommodation (blurring of vision) and, eventually, impairment of horizontal eye movements. Parkinsonian features can be prominent, whereas the characteristic oculomotor findings may be absent early on, leading to diagnostic confusion. The rest of the examination reveals prominent axial rigidity and dystonia, dysarthria, and upper motor neuron signs with loss of fine motor dexterity and gait imbalance. Patients complain of a tendency to "fall like a tree" without warning. In more advanced stages, appendicular dystonia and cerebellar signs develop. A significant number of patients progress to dementia similar to that seen in patients with CBD. Widespread degeneration in the brain stem, basal ganglia, and cerebellum is present in varying degrees. Although most patients respond to high doses of antiparkinsonian medications initially, response to treatment is mediocre at best compared with that in PD.

Corticobasal degeneration CBD is almost as common as PSP, but it may tend to be missed because it often presents as an atypical dementia with few of the characteristic movement abnormalities that may be associated with PSP.[82,83] The syndrome typically begins in the sixth decade of life and presents as a varied but striking unilateral syndrome with rigidity, dystonia, slowness, and apraxia, with or without cortical sensory loss.[84] Symptoms worsen until the patient becomes totally incapacitated. MRI and gross examination of biopsy specimens reveal prominent focal cortical loss in the contralateral parietal and frontal areas. Histologic examination reveals gliosis and swollen (ballooned) achromatic neurons and neuronal loss. There are also white-matter abnormalities consisting of swollen and demyelinated axons, and the neuropil has a spongiform appearance. In addition to affecting the cortex, these changes affect the subthalamic nucleus, globus pallidus, striatum, thalamus, and several cerebellar nuclei. Treatments are generally ineffective.

Parkinsonism and Dementia

A number of disorders (e.g., Alzheimer disease, Creutzfeldt-Jakob disease, and frontotemporal dementia [Pick disease]) present with dementia and variable degrees of extrapyramidal features. For example, 20% of patients with Alzheimer disease exhibit mild rigidity and bradykinesia. As many as 40% of PD patients have some Alzheimer disease pathology.[36]

Patients with parkinsonism and dementia typically present with an akinetic-rigid form of parkinsonism without tremor. The progression of disability is more rapid, and management is more difficult than in nondemented patients because of the high incidence of drug-induced cognitive side effects. In these patients, central dopaminomimetic toxicity can present in many ways, including sleep disruption (with daytime sleepiness), personality changes, depression and mental dullness, episodic confusion, hallucinations, and disruptive behaviors.[64]

Diffuse Lewy body disease is a form of dementia with prominent parkinsonian features that is recognized with increasing frequency.[85] In some families with diffuse Lewy body disease, patients with pure Lewy body PD are found. Pathologically, Lewy bodies are found in the cortex and brain stem. Many of these patients also exhibit Alzheimer disease pathology, which makes it difficult to distinguish these patients from those with a PD–Alzheimer disease overlap syndrome. Clinically, an action tremor may precede other parkinsonian features, and dementia is often heralded by levodopa-induced sedation, myoclonus, and hallucinations. Early on, the response to levodopa can be substantial, making it difficult to differentiate diffuse Lewy body

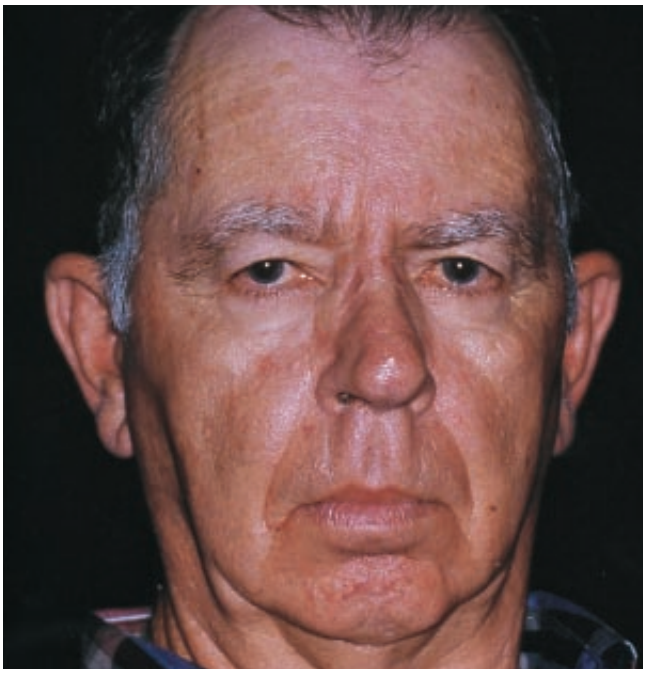


Figure 6 **Face of a patient with progressive supranuclear palsy. Note the intense leonine expression and deep facial furrows, in contrast to the passivity of expression in Parkinson disease. Patients with progressive supranuclear palsy typically have a surprised or staring expression.**

disease from PD. The progression of symptoms of diffuse Lewy body disease appears to be intermediate, as compared with the progression of symptoms in the PD and PD–Alzheimer disease subgroups.

Hyperkinetic Disorders

Hyperkinetic movement disorders encompass a variety of involuntary movements (dyskinesias) that may occur in isolation, in combination, and even in the setting of parkinsonism. Hyperkinesias have a wide spectrum of severity, ranging from barely detectable (e.g., restlessness or fidgetiness) to violent (e.g., hemiballismus) and from straightforward to complex and emotionally laden (e.g., focal tics and coprolalia in Tourette syndrome). When dyskinesias occur in combination, it is important to determine which one is predominant, because this will help guide the diagnostic workup and suggest potential therapies.

Chorea is perhaps the most common form of dyskinesia.[83] The term chorea, which means dance, refers to arrhythmic, involuntary movements that are typically sudden and brief and seem to flow from one part of the body to another. When combined with slower, writhing movements or dystonic posturing, the dyskinesia is sometimes referred to as choreoathetosis. Hemiballismus is a unilateral, large-amplitude, violent form of chorea that affects the proximal limbs more than the distal limbs. The most common cause of hemiballismus is a vascular lesion of the subthalamic nucleus.

Examples of other quick, involuntary movements that may be confused with chorea include myoclonus, tics, and dystonic tremors. Myoclonic jerks are lightning fast but lack the rhythmic flow of activity seen in chorea. Although patients with myoclonic jerks commonly lose motor control and drop objects, this rarely happens to patients with chorea or tics. Unlike chorea and myoclonus, motor tics can be readily suppressed.

Dystonic movements tend to be slower and the contractions are more sustained (lasting seconds or longer), as compared with chorea and tics. At first, dystonia may be present only with voluntary movement (action dystonia); later, dystonia may occur spontaneously. The dynamic contractions often lead to postural deformities such as those in spasmodic torticollis. Early on, dystonic movements may respond to sensory tricks, such as holding the chin lightly to prevent the neck from moving. Tremor associated with dystonia tends to have a jerking, irregular quality that is different from the oscillatory quality of most other tremors. Chorea, myoclonus, tremor, and dystonic contractions are perceived by patients as involuntary. In contrast, tics are often perceived as voluntary responses to underlying urges.

CHOREA

Etiology

Chorea has been described in at least 150 somatic illnesses and as a reaction to numerous drugs and toxins. From a medical standpoint, the more important causes of chorea are postinfectious, immunologic, vascular, metabolic, endocrine, and drug-induced processes [*see Tables 6 and 7*]. Systemic illnesses that may cause chorea include inflammatory disorders such as systemic lupus erythematosus and periarteritis nodosa, as well as infectious or parainfectious processes such as streptococcal infection, diphtheria, pertussis, toxoplasmosis, and AIDS. Cerebrovascular disorders that cause chorea are associated with asymmetrical neurologic signs and a stepwise or stuttering course that is consistent with the lacunar state.

The most important pediatric entities associated with chorea are Sydenham chorea, Wilson disease, torsion dystonia, Tourette syndrome, and, in rare instances, Huntington disease. Rare disorders of intermediary metabolism can also lead to chorea, but their associated findings, such as myoclonus, encephalopathy, and mental subnormality, tend to be more prominent than the chorea [*see Table 8*]. Disorders associated with chorea that may also present in adulthood include Wilson disease, Huntington disease, and familial calcification of the basal ganglia (Fahr disease).

Treatment

Regardless of etiology, chorea can usually be treated with so-called dopamine depletors (e.g., reserpine, tetrabenazine) and dopamine blockers. Dopamine depletors are preferred because they tend to produce fewer extrapyramidal side effects than dopamine blockers and do not carry the risk of tardive dyskinesia.[86] Potentially disabling side effects of reserpine include parkinsonism, acute dystonia, and postural hypotension. Reserpine is typically started at low doses (0.10 to 0.25 mg/day); doses are increased as tolerated (usually weekly) until the chorea is controlled or side effects supervene.

HUNTINGTON DISEASE

Epidemiology, Etiology, and Pathogenesis

Huntington disease is a uniformly fatal autosomal dominant disorder with high penetrance that is characterized by progressive motor, emotional, and cognitive dysfunction. Onset is typically in midlife, although it ranges from 3 to 70 years of age. Huntington disease is recognized worldwide and has a prevalence of 10 cases per 100,000 population.

Table 6 Selected Causes of Chorea

Developmental and Age-Related
- Cerebral palsy
- Kernicterus
- Senile chorea
- Oral-lingual-buccal dyskinesias of aging

Hereditary
- Disorders of intermediary metabolism
- Lesch-Nyhan syndrome and its variants
- Wilson disease
- Hallervorden-Spatz disease
- Ataxia-telangiectasia
- Tuberous sclerosis
- Huntington disease
- Choreoacanthocytosis
- Familial calcification of the basal ganglia

Cerebrovascular
- Basal ganglia infarction or hemorrhage
- Arteriovenous malformation
- Polycythemia vera
- Hemoglobin sickle cell disease

Immunologic
- Systemic lupus erythematosus
- Antiphospholipid syndrome
- Periarteritis
- Behçet syndrome
- Henoch-Schönlein purpura

Nutritional
- Beriberi, pellagra
- B_{12} deficiency in infants

Infectious
- Sydenham chorea (poststreptococcal)
- Scarlet fever
- Diphtheria
- Pertussis
- Typhoid fever
- Viral encephalitis
- Postvaccinial
- Neurosyphilis
- Mononucleosis
- Legionnaires disease
- Bacterial endocarditis

Metabolic
- Hyponatremia
- Hypocalcemia
- Hypoglycemia
- Hypomagnesemia
- Hepatic encephalopathy
- Renal encephalopathy
- Porphyria

Endocrine
- Hyperthyroidism
- Hypoparathyroidism and hyperparathyroidism
- Pseudohypoparathyroidism
- Chorea gravidarum (in pregnancy)
- Addison disease

Miscellaneous
- Multiple sclerosis
- Sarcoidosis

Huntington disease is one of a group of neurologic disorders caused by an increase in the number of repeats of the DNA trinucleotide, CAG, that encodes glutamine; thus, these disorders are characterized by expanded numbers of glutamine residues [*see 11:X Inherited Ataxias*].[87] Like the other trinucleotide repeat disorders, Huntington disease exhibits anticipation (a tendency toward earlier onset in successive generations), a preponderance of unstable repeats from the paternal chromosome, and a correlation between the number of repeats and the severity of disease and age of onset.[87]

Huntington disease is caused by the *HD* gene, an expanding and unstable polyglutamine repeat in the short arm of chromosome 4.[88] Normal persons have up to 34 CAG repeats, whereas nearly all patients with Huntington disease have more than 40 repeats. Individuals with 34 to 39 repeats are thought to have unstable and dynamic so-called premutations that may lead to higher numbers of repeats—and, hence, clinical disease—in their progeny.[87]

The *HD* gene encodes a protein named huntingtin that is widely distributed in neural and nonneural tissue. The role of this protein in the pathogenesis of the illness is unknown, but patients with Huntington disease, as well as experimental mice expressing the mutation, have intranuclear huntingtin-positive and ubiquitin-positive inclusions in the affected brain regions.[89] This has led to speculation that abnormalities in the intracellular degradation of these proteins mediate the process of apoptotic neuronal death in patients with Huntington disease.[90,91]

Clinical Features

A diagnosis of Huntington disease can be made on clinical grounds alone in patients who present with chorea and have a positive family history. Dementia and emotional symptoms (e.g., depression, irritability) may precede and typically overshadow the motor symptoms, accounting for the greatest disability and hardship. The clinical course can be as long as 20 years. In the early stages of the disease, physical examination usually reveals focal choreatic movements (e.g., increased blinking or grimacing) that progress and eventually involve multiple body parts.[92] The chorea typically peaks within 10 years and is gradually replaced by muscle restlessness, hypertonia and dystonia, and tremor or myoclonic jerks.

Other movement abnormalities in Huntington disease include bradykinesia and slow, saccadic eye movements. Gross motor coordination and gait become impaired, clumsiness increases, and patients move with a wide-based, slow gait and irregular, lurching steps. Fine motor coordination and, eventually, speech become severely affected. Swallowing abnormalities can lead to aspiration pneumonia. Patients progress to a nearly mute, akinetic state. In 6% to 10% of cases, Huntington disease may present as parkinsonian features rather than as chorea (Westphal variant). The Westphal variant typically presents early (i.e., onset occurs in persons younger than 20 years) and is more likely to exhibit paternal inheritance.

Laboratory Tests

The diagnosis of Huntington disease can be confirmed with genetic testing. For issues related to genetic counseling and presymptomatic testing, the reader is referred to a review of these important issues.[93] It is not necessary to test persons for whom there is genetic or pathologic confirmation of the disease in other family members. Ancillary diagnostic measures include MRI to look for selective atrophy of the caudate nucleus, a typical finding that may be hard to detect.

Treatment

Treatment of Huntington disease should involve a multidisciplinary team that can provide social, medical, neuropsychiatric, and genetic guidance to patients and families throughout the course of the illness. Symptomatic treatment can alleviate chorea, although this is not typically the most disabling aspect of the illness; indications to treat chorea include interference with activities of daily living and social embarrassment. Dopamine blockers are at best moderately effective, and they may aggravate the nonchoreatic movement abnormalities of Huntington disease, such as bradykinesia and dystonia. Atypical antipsychotic agents such as risperidone and olanzapine are better tolerated and may also help in controlling chorea. The treatment of depression in Huntington disease is similar to that in PD.

Experimental treatments include the use of glutamate antagonists (e.g., lamotrigine) as neuroprotective agents meant to block neuronal apoptosis.[94] A more novel strategy is inhibition of the protease caspase-1, with the intention of modulating the presumably abnormal intracellular degradation process that leads to the accumulation of intracellular inclusions.[90]

Table 7 Selected Drug- and Toxin-Induced Movement Disorders*

Chorea
- Antipsychotic neuroleptic drugs
- Antiparkinsonian drugs
- Noradrenergic stimulants
- Anabolic steroids
- Oral contraceptives
- Tricyclic antidepressants

Myoclonus
- Psychiatric medications
  - Selective serotonin reuptake inhibitors
  - Tricyclic antidepressants
  - Monoamine oxidase inhibitors
  - Lithium carbonate
  - Clozapine
  - Dopamine blockers (tardive syndrome)
  - Methaqualone
  - Buspirone
  - Benzodiazepine withdrawal
- Narcotics
  - Morphine and its derivatives
  - Fentanyl
- Toxins
  - Bismuth
  - Mercury
  - Methyl bromide
  - Strychnine
  - Marijuana
  - Aluminum
  - Tetraethyl lead
  - Oven cleaner
  - Tetanus toxoid
- Anticonvulsants
  - Phenytoin
  - Valproic acid
  - Carbamazepine
- Anesthetics
  - Etomidate
  - Enflurane
  - Isoflurane
  - Midazolam
- Cardiac medications
  - Calcium channel blockers
  - Antiarrhythmics (flecainide, propafenone)
- Antibacterial and antiviral agents
  - Penicillins
  - Monobactams
  - Isoniazid
  - Acyclovir
  - Vidarabine
- Miscellaneous
  - Dopamine agonists
  - Metoclopramide
  - Physostigmine
  - Antihistamines
  - Cimetidine

*Drug-induced parkinsonism, tremor, and dystonia are discussed in the text.

DYSTONIA

Dystonia is a term used to describe both a constellation of signs and a series of hyperkinetic disorders. Among movement disorders, dystonia is one of the more common and also one of the most frequently misdiagnosed and untreated. The worldwide prevalence is not certain because of underreporting, but taking all forms of dystonia into account, it probably exceeds 300 million, a prevalence equal to that of multiple sclerosis.[95]

Classification

The dystonias can be classified according to the age at onset, the region of the body involved, and the etiology. Using an etio-

Table 8 Differential Diagnosis of Myoclonus

Medical Causes of Myoclonus
- Metabolic disorders
  - Hyponatremia
  - Hypoglycemia
  - Nonketotic hyperglycemia
  - Hepatic failure
  - Renal failure
  - Dialysis disequilibrium syndrome
  - Mitochondrial defects
  - Multiple carboxylase deficiency
- Viral encephalopathies
  - Herpes simplex encephalitis
  - Arbovirus encephalitis
  - Human immunodeficiency virus
  - Encephalitis lethargica
  - Subacute sclerosing panencephalitis
  - Postinfectious encephalitis
- Malabsorption syndromes
  - Celiac disease
  - Whipple disease
- Paraneoplastic syndromes

Physical Encephalopathies
- Posthypoxia (Lance-Adams)
- Heatstroke
- Electric shock

Neurologic Causes of Myoclonus
- Focal/diffuse CNS damage
  - Stroke
  - Thalamotomy
  - Hypoxia (Lance-Adams)
  - Tumors
- Essential myoclonus
  - Hereditary (autosomal dominant)
  - Sporadic
  - Palatal myoclonus
- Epileptic myoclonus (childhood seizures without encephalopathy)
  - Epilepsia partialis continua
  - Idiopathic stimulus-sensitive myoclonus
  - Myoclonic absences in petit mal epilepsy
  - Childhood myoclonic epilepsies
- Progressive myoclonic epilepsy (seizures with encephalopathy; onset age variable)
  - Lennox-Gastaut syndrome
  - Baltic myoclonic epilepsy (Unverricht-Lundborg disease)
  - Lafora myoclonic epilepsy
  - Adult ceroid lipofuscinosis (Kuf disease)
  - Neuronal ceroid lipofuscinosis (Batten disease)
  - Storage diseases
    - Lipidoses
      - GM_2 gangliosidosis
      - Tay-Sachs disease
      - Krabbe disease
    - Sialidosis (cherry-red spot myoclonus)
- Defects of oxidative phosphorylation
  - Myoclonic epilepsy and myopathy with ragged red fibers
  - Mitochondrial encephalopathy lactic acidosis strokelike episodes
- Progressive myoclonic ataxias (prominent ataxia; onset age, encephalopathy, and seizures are variable)
  - Friedreich ataxia
  - Ataxia-telangiectasia
  - Dyssynergia cerebellaris progressiva
  - Spinocerebellar ataxias
  - Dentatorubral-pallidoluysian atrophy
- Basal ganglia degenerations
  - Wilson disease
  - Torsion dystonia
  - Hallervorden-Spatz disease
  - Huntington disease
  - Multiple system atrophy
  - Cortical basal ganglia degeneration
- Primary dementias
  - Creutzfeldt-Jakob disease
  - Alzheimer disease
  - Diffuse Lewy body disease

logic scheme similar to that used for PD, the dystonias may be divided into primary, secondary, and dystonia-plus syndromes and heredodegenerative disorders with dystonic manifestation.

Primary dystonias Primary dystonias include syndromes in which dystonia is the principal clinical manifestation of the disease. Although no pathology has been identified, the basal ganglia are strongly implicated. Fluorodeoxyglucose PET scans in patients with primary dystonias reveal increased metabolism in the GPi, the cerebellum, and the supplementary motor area. These abnormalities have been interpreted as representing increased (or abnormal) drive in the direct striatopallidal pathway that inhibits GPi. This results in decreased inhibition (or modulation) of thalamocortical input, which leads to overactivity in the frontal lobe planning centers (e.g., supplemental motor area, dorsolateral prefrontal cortex) and underactivity in the motor executive areas such as the primary motor cortex.[95]

The major primary dystonia is dystonia type 1 (DYT1), an autosomal dominant hereditary disorder that can present as either generalized or focal dystonia.[96] A high proportion of DYT1 occurs in Ashkenazi Jewish families, but the disorder is also found in non-Jewish families. The disease has a penetrance of about 30% and is caused by a gene located on chromosome 9q34. The *DYT1* gene codes for the protein torsin A and is adjacent to the gene *DQ1,* which encodes for the related protein torsin B.[96] Torsin A and torsin B form part of a larger group of adenosine triphosphatase (ATPase)–associated proteins (AAA family) that either have ATPase activity or that bind to adenosine triphosphate (ATP), forming oligomeric complexes with other proteins. These complexes are involved in a variety of processes, including cell signaling, folding and unfolding of other proteins, and recovery and protection from stress by penetration of degraded proteins.[96] How these proteins relate to dystonia itself remains unclear, but the role of these complexes in providing recovery and protection from stress opens the possibility that these genes confer susceptibility to a so-called second hit; such a two-stage process might account for the 30% penetrance of the gene and perhaps some of the secondary (e.g., posttraumatic) forms of dystonia.

Dystonia-plus syndromes Dystonia-plus syndromes refer to those in which the patient exhibits neurologic findings in addition to dystonia. Perhaps the most interesting and treatable of these is dopamine-responsive dystonia (DRD), which is dominantly inherited but is more common in women. DRD is associated with multiple mutations in the gene for guanine triphosphate cyclohydrolase I, the rate-limiting enzyme in the synthesis of the tyrosine hydroxylase cofactor tetrahydrobiopterin.[97] Tyrosine hydroxylase is the rate-limiting enzyme for dopamine synthesis. DRD typically presents in childhood, beginning in the legs and then spreading to the arms. Marked diurnal fluctuations are common. Some patients present later in life with parkinsonian symptoms instead of dystonia. The hallmarks of DRD are an excellent response to low doses of levodopa and, in many patients, slowly evolving features of PD. DRD may be misdiagnosed as athetoid cerebral palsy, spastic paraplegia, or parkinsonism. Although rare, DRD and its other phenotypic forms are so responsive to levodopa that a trial of levodopa is warranted in all suspected cases. DRD appears to be a purely biochemical disorder with no degenerative change in the brain. An interesting form of dystonia-plus syndrome is myoclonic dystonia, which has been linked to polymorphism in the D_2 dopamine receptors.[98]

Heredodegenerative dystonias The heredodegenerative dystonias (dystonias with demonstrable and progressive neuropathology) include DYT3, which is an X-linked form of dystonia-parkinsonism found in men in certain Philippine families; women carriers are mildly affected. The disease is sometimes called by its Filipino name, Lubag.[99] Other examples of heredodegenerative dystonias are rapid-onset dystonia parkinsonism,[100] dystonia-associated trinucleotide repeat disorders (e.g., Huntington disease), Machado-Joseph disease (spinocerebellar ataxia type 3 [SCA 3]), and dentatorubropallidoluysian atrophy (DRPLA) [*see 11:X Inherited Ataxias*]. A few metabolic disorders associated with prominent dystonia include Wilson disease, GM1 and GM2 gangliosidosis, glutaric acidemia, and neuroacanthocytosis. Mitochondrial disorders with dystonic manifestations include Leigh disease and Leber disease.

Secondary dystonias Secondary dystonias are caused largely by drugs, trauma, or other environmental factors. They include levodopa-induced dystonia, acute and tardive dystonia associated with dopamine receptor blockers, dystonia associated with cerebral palsy (athetoid cerebral palsy), cerebral trauma, cerebral hypoxia, and peripheral nerve injury. Dystonia can also be associated with infectious and postinfectious states and with toxic exposure to manganese, cyanide, and 3-nitropropionic acid.

Clinical Features

Dystonic signs include involuntary muscle contractions, which typically occur first with voluntary movement (action dystonia) and later at rest and which result in twisting of body parts and abnormal postures. Dystonia is sometimes accompanied by a coarse, irregular, jerking tremor (dystonic tremor) or by a typical postural tremor. Cocontraction of antagonist muscles is a fundamental feature of dystonia that distinguishes it from chorea, tics, and other dyskinesias. Dystonia is also associated with so-called overflow phenomena, which involve the abnormal spread of activation to muscles other than those required for the intended movement. The most common forms of dystonia are the focal dystonias, which may affect (1) the eyelids (blepharospasm), causing them to close involuntarily; (2) the neck and shoulders (cervical dystonia), causing the neck to twist to the side (torticollis), forward (anterocollis), or backward (retrocollis); (3) the lower face and jaw (oromandibular dystonia); and (4) the larynx (spasmodic dysphonia), causing the voice to have a strained and discontinuous quality as a result of involuntary closure of the vocal cords. Focal dystonia may affect the hand and forearm during the performance of specific activities, such as handwriting (writer's cramp), typing, or playing a musical instrument (musician's cramp), or during almost any other activity that involves repetitive actions of the hands (occupational dystonia). The focal dystonias that occur in adults are often misdiagnosed as psychiatric or orthopedic problems, as compared with generalized dystonia, which begins in childhood. In children, generalized dystonia begins as an action dystonia of the foot. Later, dystonia occurs at rest; eventually, it causes postural abnormalities as it spreads first ipsilaterally and then to the limbs contralateral to the initially affected foot.

Treatment

Treatment of dystonia is strictly symptomatic and consists of medical and surgical therapies. In general, high doses of anticholinergics, sometimes in combination with neuroleptics or

dopamine depletors, are used to treat generalized dystonia.[101] Stereotactic surgery (pallidotomy and thalamotomy) has been used successfully to treat medically refractory cases.[102] The best therapeutic outcomes in idiopathic torsion dystonia (especially DYT1) appear to be in patients treated with deep brain stimulation of the GPi; this procedure has been approved by the FDA for this purpose. The neurosurgery literature shows substantial response to deep brain stimulation in generalized dystonias—a response that is much greater than that achieved by medical management.[103] Early referral of these patients to a movement disorder specialist is indicated so that this intervention can be considered.

The focal dystonias generally respond poorly to drugs, but they respond dramatically to botulinum toxin injections of the affected muscle group. Botulinum toxin acts by blocking the release of acetylcholine at the neuromuscular junction, resulting in dose-dependent weakness, which alleviates the spasmodic contraction without producing paralysis. Repeat injections are required every 2 to 5 months.

WILSON DISEASE

Epidemiology and Etiology

Wilson disease, or hepatolenticular degeneration, is an autosomal recessive disorder of copper metabolism with a prevalence of one per 30,000 population.[104] Although relatively rare, it is clinically important because it is treatable and mimics other movement disorders. The gene for Wilson disease, localized in chromosome 13, encodes a 7.5 kb protein transcript that is highly expressed in the liver and brain.[105] The protein belongs to a class of ATP-dependent copper-transporting ATPases. Multiple mutations of the Wilson disease gene give rise to the same phenotype. Most victims are compound heterozygotes who have different mutations in each allele. This complicates the molecular diagnosis of Wilson disease and limits screening to family members of known patients.[106,107]

Pathogenesis

In Wilson disease, reduced hepatic incorporation of copper into ceruloplasmin and reduced biliary copper excretion eventually lead to exhaustion of hepatic storage sites; spillover of copper into plasma; and pathologic deposition of copper in the brain (primarily in the basal ganglia), cornea (producing Kayser-Fleischer rings), kidney (causing tubular dysfunction), bone marrow (causing anemia and thrombocytopenia), and musculoskeletal system (causing arthritis and osteoporosis).[104] Acute crises (acute encephalopathy with seizures, hemolytic crisis, and acute tubular damage) are caused by the sudden dumping of large quantities of copper into the bloodstream iatrogenically (see below) or after liver infarction or acute hepatitis.

Figure 7 **Kayser-Fleischer rings in a young boy with Wilson disease.**

Clinical Features

The age at presentation in Wilson disease ranges from 3 to 60 years, with a peak onset at around 16 years. In one large study, 54% of patients presented with neurologic manifestations, 31% presented with hepatic dysfunction, and as many as 33% had psychiatric manifestations.[104] Most patients who present with liver abnormalities are female and do so typically between 8 and 16 years of age; males present more commonly with neurologic manifestations. The presentation can vary from acute fulminant hepatic failure to low-grade insidious cirrhosis. Wilson disease may be the most common cause of chronic liver dysfunction in children; all children with that presentation should be investigated for the disorder.

Important hematologic manifestations of Wilson disease include hemolytic anemia, thrombocytopenia (found in 50% of all patients with hematologic presentations), and neutropenia (30% of all patients with hematologic presentations). Other systemic manifestations include pancreatic exocrine abnormalities, hypoparathyroidism, polyarthritis, renal stones, and cardiomyopathy from myocardial copper accumulation.

The neurologic manifestations of Wilson disease are myriad and include a variety of movement disorders, such as dystonia (65% of patients), rigidity or parkinsonism (52%), postural and gait abnormalities (42%), and tremor (32%). The most common neurologic presentation, however, is dysarthria (97%), which is sometimes accompanied by abnormal facial expressions (the classic risus sardonicus).

The pleomorphic psychiatric and behavioral manifestations of Wilson disease include irritability, impaired work performance, depression, and mood changes. Affective disorders that mimic bipolar illness and psychotic symptoms also occur. Cognitive problems may start with poor school performance and culminate in dementia. Given the extreme variability of the neuropsychiatric manifestations of Wilson disease, it is important to consider this diagnosis in virtually all patients with movement disorders that do not have a readily demonstrable etiology, especially younger patients who manifest psychiatric disturbances.

Laboratory Findings

In about 80% of patients with Wilson disease, the diagnosis can be made by demonstrating a decrease in serum ceruloplasmin (Wilson disease, 0 to 200 mg/L; normal, 200 to 400 mg/L). Patients with normal ceruloplasmin levels are presumed to be heterozygotes. The serum copper level (bound plus unbound) is also decreased (Wilson disease, 3 to 10 μmol/L; normal, 11 to 24 μmol/L). Measurement of 24-hour urinary copper excretion may be more sensitive than measurement of serum copper levels for detecting Wilson disease. In patients with Wilson disease, 24-hour urinary copper excretion is 100 to 1,000 μg/24 hr (normal, ≤ 40 μg/24 hr), but this result is also found in other disorders that produce cholestasis. When measurement of ceruloplasmin or copper is not conclusive, demonstration of increased hepatic copper content on liver biopsy is the definitive test (Wilson disease, 200 to 3,000 μg/g dry liver; normal, ≤ 50 μg/g dry liver).

Almost all patients who have neurologic involvement have the classic ophthalmologic findings of Kayser-Fleischer rings, or sunflower cataracts [*see Figure 7*]. Kayser-Fleischer rings represent copper deposits in the basement membrane of the cornea, the detection of which may require a slit-lamp examination. Neuroimaging studies reveal loss of brain volume, with ventricular dilatation in more than 95% of neurologically affected persons, as well as in some who are neuropsychiatrically intact. MRI shows a unique hyperintense signal on T_2-weighted images of the lenticular nucleus, as well as on images of the thalamus, midbrain pons, and dentate gyrus of patients with more advanced disease [*see Figure 5d*]. T_1-weighted images may reveal white-matter lesions.

Treatment

Treatment of Wilson disease has to be undertaken cautiously to prevent acute encephalopathy or hepatic toxicity induced by a rapid mobilization of copper. We favor obtaining an MRI before initiating treatment, even in asymptomatic patients. Treatment consists of lowering dietary copper and reducing copper absorption with zinc acetate or zinc sulfate (150 mg/day) or potassium sulfide (20 mg t.i.d.).[108] To mobilize tissue copper deposits, it is necessary to use copper chelation with penicillamine (1 to 3 g/day orally). Chelation therapy needs close monitoring for patients to achieve an initial 24-hour copper urinary excretion level of 1.2 to 2.0 g/day. Penicillamine therapy is poorly tolerated in as many as 30% of patients because of idiosyncratic allergic side effects, including autoimmune conditions. In contrast to penicillamine, zinc therapy is well tolerated. Zinc can increase plasma amylase, lipase, and alkaline phosphatase levels; reports of immunosuppressive effects are rare. In patients who cannot tolerate penicillamine or in those who worsen as therapy is initiated, alternative copper-chelating agents are trientine (with or without lower doses of penicillamine), dimercaprol, and cobalt chloride (now rarely used).[109] Other strategies to minimize tissue damage from copper deposition include free-radical scavengers such as mercaptodextran. When there is evidence of progressive disease despite optimal dietary and pharmacologic therapy or when hepatic failure occurs, autologous liver transplantation is the most definitive therapy.[110]

The neurologic symptoms of Wilson disease generally respond poorly to therapy. Anticholinergics may be given to treat dystonia and tremor; dopaminomimetics may be given to treat rigidity. Psychiatric symptoms (i.e., anxiety and depression) are treated with standard therapies. Improvement in neuropsychiatric and other symptoms may take months to as long as 2 years after initiation of chelation therapy.

TOURETTE SYNDROME

Etiology

Tourette syndrome is an autosomal dominant disorder for which sib pair studies have identified at least two regions in the genome with high probability (LOD scores > 2) of containing the gene or genes responsible for this complex illness.[111] Tourette syndrome has a broad range of clinical manifestations that is influenced by gender as well as variable penetrance. The diagnostic criteria include multiple motor and vocal tics that appear before age 21 and last longer than 1 year. The symptoms must be of sufficient severity to warrant a medical diagnosis, and the disturbance cannot be associated with other neurologic or drug-induced conditions. A family history is important and may include individuals affected by a single motor tic, obsessive-compulsive symptoms (OCS), and perhaps attention-deficit disorder (ADD). The male-to-female ratio for Tourette syndrome is more than 4:1.[112] Learning disabilities, sleep disorders, and a variety of poorly defined behavioral syndromes may also be present.[113]

Pathogenesis

The pathogenesis of Tourette syndrome is unknown, but clinical pharmacologic evidence and limited autopsy studies suggest that it may be mediated by hyperdopaminergic innervation of the corpus striatum. Morphometric studies using brain MRI scans and metabolic studies using PET scans suggest the presence of anatomic and functional abnormalities in cortico-striato-pallido-thalamic circuits that project to the prefrontal cortex and cingulate gyrus. Surgical interruption of these circuits (e.g., anterior tractotomy and cingulotomies) may be of value in treating OCS and, to a lesser extent, tics, in carefully selected cases.

A still undetermined subgroup of Tourette syndrome (and OCS) cases have been linked to a poststreptococcal syndrome associated with circulating antibodies that presumably cross-react with neurons in the striatum and cause dysfunction in these circuits. This syndrome forms part of a broader category of pediatric autoimmune neurologic disorders associated with streptococcal infections (PANDAS). The onset of symptoms is often temporally linked with a history of pharyngitis and tends to be more abrupt than in regular Tourette syndrome. Symptoms tend to fluctuate wildly. There is no associated carditis or nephritis. Successful treatment with plasma exchange has been reported, but this and other proposed treatments are still experimental.

Clinical Features

Tics, both motor and vocal, are a prominent feature of Tourette syndrome. A tic is a brief, rapid, repetitive, and seemingly purposeless stereotyped action that may involve a single or several muscle groups. Both motor and vocal tics can be simple (e.g., nose twitching, coughing, and grunting) or complex (self-mutilation and coprolalia). Tics can be associated with brief focal sensory experiences (sensory tics), which may precede, accompany, or follow the movements or vocalization. In general, tics are the least disabling aspect of Tourette syndrome and may not require treatment. During school years, ADD and its behavioral accompaniments (impulsivity and irritability) and learning deficits tend to be the most disabling and disruptive features. OCS begins during early adolescence and can be quite disabling. It responds less reliably to clomipramine and standard SSRIs than does obsessive-compulsive disorder without Tourette syndrome. Sleep disorders are frequent but poorly understood.

Treatment

The treatment of Tourette syndrome is challenging and should be undertaken by specialists. When the symptoms are complex (e.g., tics, OCS, and ADD), pediatricians and neurologists may need to work in concert with psychiatrists who can provide guidance and treatment of the behavioral symptoms. Tics generally respond to dopamine blockers (e.g., fluphenazine, haloperidol, pimozide, olanzapine).[114] Some children may be particularly intolerant of the side effects of these drugs, which range from acute dystonia and motor restlessness (akathisia) to daytime sedation and school phobias. The atypical neuroleptic risperidone may also be effective. Other treatments for tics include clonidine (0.1 to 0.4 mg/day, either as a pill or as a transdermal patch) and clonazepam (0.5 to 4.0 mg/day).

Tremor

The term tremor describes a number of unrelated movements that share one major characteristic: oscillation of a limb or body part. Parkinsonian rest tremor and cerebellar tremor are examples.

ENHANCED PHYSIOLOGIC TREMOR

Physiologic tremor is typically not clinically apparent, except when exaggerated by stress, anxiety, or stimulants. Stage fright and other forms of situational anxiety are well-recognized and readily treatable examples of stress-enhanced physiologic tremor. Other causes of endogenous epinephrine release that may enhance physiologic tremor are pheochromocytoma and hypoglycemia. Drugs that may enhance physiologic tremor include beta-adrenergic agonists (e.g., exogenous epinephrine, isoproterenol, terbutaline, and theophylline), lithium, tricyclic antidepressants, stimulants, valproate, steroids, thyroid hormones, and caffeine. Withdrawal of certain drugs, such as alcohol, opiates, and benzodiazepines, and muscle fatigue from excessive exercise can cause enhanced physiologic tremor.

Treatment

Treatment includes eliminating or reducing tremor-enhancing drugs when possible, using beta-adrenergic blockers, controlling anxiety, and treating any associated endocrinopathy. Situational anxiety can be controlled by taking propranolol (20 to 60 mg) 1 hour before the triggering event or situation.

ESSENTIAL TREMOR

Essential tremor has a frequency of 6 to 8 Hz, is of central origin, and commonly follows an autosomal dominant pattern of inheritance.[115] Unlike patients with PD, patients with essential tremor may lack other parkinsonian features and may exhibit isolated head titubation and voice tremor. The latter gives speech a quivering, shaky quality. Essential tremor is exacerbated by the same factors that enhance physiologic tremor.

Treatment

Characteristically, essential tremor responds rapidly, albeit briefly, to alcohol (as blood alcohol level falls, however, the rebound effect may worsen the tremor). More practically, it also responds to beta-adrenergic blockers. Propranolol is usually an effective long-term treatment; other helpful beta blockers include metoprolol, timolol, atenolol, and sotalol. Primidone can be effective at doses of 50 to 500 mg/day, which are lower than those required for anticonvulsant therapy. Other drugs that may be beneficial include benzodiazepines (especially clonazepam), phenobarbital, and clonidine. Anticonvulsants such as gabapentin and topiramate have also proved to be helpful. In particularly refractory cases, clozapine can be considered. Stereotactic neurosurgery (ventrolateral thalamotomy or deep brain stimulation) is very effective in severe, medically refractory essential tremor. It is less effective in cerebellar tremor.

MYOCLONUS

Clinical Features

Myoclonus is a sudden, brief, shocklike, involuntary jerk. Myoclonic jerks may be considered fragments of epilepsy; they have electroencephalographic characteristics that are similar to focal epileptic discharges. Positive myoclonus consists of an active muscle contraction; negative myoclonus consists of a sudden lapse in contraction, which is clinically indistinguishable from asterixis. Specialized electrophysiologic testing can localize myoclonic discharges to the somatosensory cortex, the brain stem (the reticular activating system), and, in rare instances, the spinal cord. Like seizures, myoclonic discharges represent a nonspecific sign of CNS irritation with a broad differential diagnosis [*see Table 8*].[116]

Myoclonic jerks can be focal, segmental, or generalized. When focal, they suggest a focal pathology, which can range from epilepsy to focal encephalitis [*see Table 8*]. They can be spontaneous, stimulus sensitive (reflex), or related to ongoing activities (action). When generalized, they can be associated with epilepsy or with a host of static and progressive encephalopathies. To narrow the differential diagnosis, it is important to look for signs of encephalopathy, dementia, retinal pathology, hearing loss, myopathy, neuropathy, and ataxia.

Differential Diagnosis

In the differential diagnosis of myoclonus, it is important to first consider toxic etiologies and intercurrent medical illness before considering a primary disorder of the CNS.[116,117] An important element of the differential diagnosis is whether the myoclonus is epileptic or nonepileptic.[118,119] The age at onset is also extremely important, because the more common progressive forms of epilepsy that result from degenerative processes tend to occur in the pediatric age group.

Laboratory Tests

Basic laboratory screening should include a routine hemogram, blood chemistries, and a toxicologic screen. Brain imaging and electroencephalography are indicated when the cause of the myoclonus is not immediately apparent. When clinically indicated, specialized metabolic studies should be considered to rule out systemic inflammatory disorders, endocrinopathies, disorders of copper metabolism (e.g., Wilson disease and Menke disease), selected enzyme deficiencies, and lysosomal storage disorders [*see Table 8*]. A subgroup of these disorders, such as some disorders of mitochondrial oxidative phosphorylation (see below) and some of the SCAs, can now be diagnosed with genetic tests.

Treatment

Treatment of myoclonus is strictly symptomatic. Commonly used drugs include clonazepam (2 to 12 mg/day), sodium valproate (600 to 3,000 mg/day), primidone (500 to 1,500 mg/day), and, in Europe, piracetam (6 to 20 mg/day). SSRIs and 5-hydroxytryptophan (with carbidopa) may help specific forms of myoclonus (reticular reflex myoclonus) but tend to aggravate most other forms of myoclonus. Another special case is palatal myoclonus, which responds to trihexyphenidyl (6 mg/day). Treatment of epileptic syndromes associated with the above myoclonic disorders is beyond the scope of this chapter.

MYOCLONUS IN GENERAL MEDICAL PRACTICE

In general medical practice, the most common form of myoclonus results from metabolic disorders, in which negative myoclonus, or asterixis, tends to dominate the clinical picture [*see Table 8*]. Etiologies of myoclonus also include viral encephalitis, which typically presents as positive and negative myoclonus. Examples include encephalitides associated with herpes simplex virus and arbovirus, as well as those associated with postinfectious syndromes (e.g., subacute sclerosing panencephalitis and encephalitis lethargica).

Other common forms of myoclonus observed in general medical practice are drug-induced myoclonus and myoclonus sec-

ondary to drug withdrawal. Drugs that induce myoclonus include lithium carbonate, tricyclic antidepressants, and SSRIs. Coadministration of SSRIs and MAO-A or MAO-AB inhibitors can cause a hyperserotoninergic syndrome that is typically heralded by myoclonus and changes in mental status. Antimicrobial agents occasionally associated with myoclonus include penicillins and cephalosporins [*see Table 7*]. In anesthesia, ultra–short-acting nonbarbiturate hypnotics (e.g., etomidate) can also cause myoclonus. Patients with PD who are taking levodopa occasionally develop myoclonus.

Overdose or intoxication with anticonvulsants, antihistamines, or hypnosedatives can cause myoclonus, with or without generalized seizures. Other drugs include antineoplastics, such as chlorambucil, NSAIDs in patients with renal failure, and bismuth. Extended exposures to the following toxins are also associated with myoclonus: methyl bromide, organic mercury, tetraethyl lead (gasoline), strychnine, and DDT. Withdrawal from alcohol, benzodiazepines, or opiates is also commonly associated with myoclonus. Finally, myoclonus can be a sequela of such global cerebral insults as trauma, hypoxia-ischemia (post-hypoxic myoclonus), heatstroke, and electrocution.[120]

Dementing disorders associated with myoclonus include prion-associated disorders, such as Creutzfeldt-Jakob disease. In addition, a small percentage of patients with Alzheimer disease (10% to 15%) and a larger percentage of patients with diffuse Lewy body disease experience myoclonus. In these disorders, myoclonus may evolve from a focal to a more generalized form.

Several additional disorders represent uncommon primary genetic disorders with complex associated symptom clusters [*see Table 8*]. Beyond controlling myoclonic jerks and seizures, therapy for these conditions is very limited.

Drug-Induced Movement Disorders

Drug-induced movement disorders are an important group of iatrogenic and mostly reversible conditions that are occasionally encountered in general clinical practice. The frequency of drug-induced movement disorders depends on the nature of the compound, the age of the affected person, and other risk factors.

Most drug-related movement disorders are associated with agents that directly or indirectly affect central dopaminergic transmission. These include CNS stimulants, the dopamine precursor levodopa, direct-acting dopamine agonists, and central dopamine receptor blockers. Movement disorders produced by these compounds can be acute; subacute; or delayed, as in the case of tardive dyskinesia or tardive dystonia. The mechanisms of acute and subacute movement disorders appear to be idiosyncratic extensions of the intended action of the compound. The mechanism of tardive syndromes is more obscure.

Approximately one third of patients with tardive dyskinesia experience remission within 3 months of stopping neuroleptic therapy. In most patients, the movements gradually remit within 5 years. Patients at risk for permanent tardive dyskinesia include the elderly, the edentulous, and those with underlying organic cerebral dysfunction. Patients with affective illnesses appear more likely to develop tardive dyskinesia than patients with schizophrenia.

ACUTE DRUG-INDUCED MOVEMENT DISORDERS

Acute dystonia may be provoked by central dopamine blockers such as lithium, calcium channel blockers, and SSRIs. Young men are most vulnerable to these reactions. Acute dystonia can be readily treated with the parenteral administration of anticholinergics (benztropine or diphenhydramine) or benzodiazepines (lorazepam or diazepam). Other acute drug-induced movement disorders include dyskinesias; stereotyped behaviors in response to neuroleptics; and tics, which may occur in response to such CNS stimulants as methylphenidate, dextroamphetamine, and pemoline.

SUBACUTE DRUG-INDUCED MOVEMENT DISORDERS

Probably the most common subacute drug-induced movement disorder is neuroleptic-induced akathisia.[119] Akathisia is a state of motor restlessness that involves an irresistible need to move. Movement tends to alleviate the symptoms temporarily. The restlessness is thought to be mediated by mesolimbic dopamine receptor blockade. Symptomatically, this may be related to restless legs syndrome [*see 11:XIII Disorders of Sleep*]. Therapy consists of removing the offending agent. When this is not possible, symptoms can be alleviated with benzodiazepines, anticholinergics, beta blockers, or, in some cases, dopamine agonists.

TARDIVE MOVEMENT DISORDERS

Clinical Features

Tardive movement disorders are primarily caused by long-term exposure (longer than 3 months) to central dopamine blockers.[121] The movements are most often choreatic and are referred to as tardive dyskinesia. Initially, the cranial musculature is affected (in particular, the mouth, lips, and tongue); later, the trunk and limbs are affected. In a fully developed case, a person may experience head nodding, pelvic rocking, and fine movements of the fingers and toes. In rare instances, tardive dyskinesia can affect the diaphragm and cause respiratory distress.

In tardive dystonia, dystonic contractions are more often axial than appendicular. Other tardive syndromes include tardive akathisia, tardive tics, and even tardive tremor.[122]

Treatment

The best treatment of tardive dyskinesia is prevention. Neuroleptics should be used judiciously, for as short a period as possible. Atypical neuroleptics pose a significantly lower risk of causing tardive dyskinesia than do conventional neuroleptics, for which the risk ranges from 13% to 20%. The lower risk associated with atypical neuroleptics may be attributed to their higher affinity for mesolimbic D_4 and D_3 dopamine receptors rather than striatal D_2 dopamine receptors. Atypical neuroleptics include clozapine, risperidone, olanzapine, quetiapine, and aripiprazole.

When withdrawal of the agent that causes tardive dyskinesia is not possible, replacing traditional neuroleptics with atypical neuroleptics should be considered.[122] If the patient is also taking stimulants and anticholinergics, eliminating them may alleviate dyskinesias.

Reserpine is the treatment of choice for choreatic tardive dyskinesia. The initial dosage of reserpine should be 0.125 mg/day; the dosage can be increased slowly up to 6 mg/day, as needed. Elderly patients are less likely to tolerate reserpine because of orthostatic hypotension. Long-term use of reserpine is associated with an approximate 15% incidence of depression.

Another treatment strategy consists of stimulating γ-aminobutyric acid (GABA) release with baclofen (40 to 80 mg/day),

clonazepam (1 to 8 mg/day), or valproic acid (1 to 3 g/day). GABAergic strategies are particularly helpful in patients with tardive dystonia, who can also benefit from anticholinergic therapy and botulinum toxin injections that target the muscles most affected. Buspirone (40 to 60 mg/day) has proved helpful in some cases. Vitamin E was not helpful in tardive dyskinesia in a multicenter, double-blind, placebo-controlled study.[122]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Alexander G, Crutcher MD: Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13:266, 1990

2. DeLong M: Primate models of movement disorders of basal ganglia origin. Trends Neurosci 13:281, 1990

3. Limousin P, Krack P, Pollak P, et al: Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. N Engl J Med 339:1105, 1998

4. Eidelberg D, Moeller JR, Ishikawa T, et al: Regional metabolic correlates of surgical outcome following unilateral pallidotomy for Parkinson's disease. Ann Neurol 39:450, 1996

5. Chesselet MF, Delfs JM: Basal ganglia and movement disorders: an update. Trends Neurosci 19:417, 1996

6. Albin RL, Young AB, Penney JB: The functional anatomy of disorders of the basal ganglia. Trends Neurosci 18:63, 1995

7. Albin RL, Reiner A, Anderson KD, et al: Preferential loss of striato-external pallidal projection neurons in presymptomatic Huntington's disease. Ann Neurol 31:425, 1992

8. Alexander G, Crutcher MD, DeLong MR: Basal ganglia-thalamocortical circuits: parallel substrates for motor, oculomotor, "prefrontal" and "limbic" functions. Prog Brain Res 85:119, 1990

9. Gelb DJ, Oliver E, Gilman S: Diagnostic criteria for Parkinson disease. Arch Neurol 56:33, 1999

10. Tanner CM, Aston DA: Epidemiology of Parkinson's disease and akinetic syndromes. Curr Opin Neurol 13:427, 2000

11. Feany MB: New genetic insights into Parkinson's disease. N Engl J Med 351:1937, 2004

12. Tanner CM, Ottman R, Goldman SM, et al: Parkinson disease in twins: an etiologic study. JAMA 281:341, 1999

13. Ritz B, Yu F: Parkinson's disease mortality and pesticide exposure in California 1984–1994. Int J Epidemiol 29:323, 2000

14. Hicks AA, Petursson H, Jonsson T, et al: A susceptibility gene for late-onset idiopathic Parkinson's disease. Ann Neurol 52:549, 2002

15. Polymeropoulos MH, Lavedan C, Leroy E, et al: Mutation in the α-synuclein gene identified in families with Parkinson's disease. Science 276:2045, 1997

16. Gasser T: Genetics of Parkinson's disease. Clin Genet 54:259, 1998

17. Tan EK, Khajavi M, Thornby JI, et al: Variability and validity of polymorphism association studies in Parkinson's disease. Neurology 55:533, 2000

18. Hattori N, Kitada T, Matsumine H, et al: Molecular genetic analysis of a novel *Parkin* gene in Japanese families with autosomal recessive juvenile parkinsonism. Ann Neurol 44:935, 1998

19. Lucking CB, Durr A, Bonifati V, et al: Association between early-onset Parkinson's disease and mutations in the *parkin* gene. N Engl J Med 342:1560, 2000

20. Klein C, Pramstaller PP, Kis B, et al: Parkin deletions in a family with adult-onset, tremor-dominant parkinsonism: expanding the phenotype. Ann Neurol 48:65, 2000

21. Hernan MA, Takkouche B, Caamano-Isorna F, et al: A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. Ann Neurol 52:276, 2002

22. Chen H, Zhang SM, Hernan MA, et al: Nonsteroidal anti-inflammatory drugs and the risk of Parkinson disease. Arch Neurol 60:1059, 2003

23. Currie LJ, Harrison MB, Trugman JM, et al: Postmenopausal estrogen use affects risk for Parkinson disease. Arch Neurol 61:886, 2004

24. Braak H, Rub U, Gai WP, et al: Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. J Neural Transm 110:517, 2003

25. Halliwell B: Role of free radicals in the neurodegenerative diseases: therapeutic implications for antioxidant treatment. Drugs Aging 18:685, 2001

26. Jenner P: Oxidative stress in Parkinson's disease. Ann Neurol 53(suppl 3):S26, 2003

27. Schapira AH: Mitochondrial involvement in Parkinson's disease, Huntington's disease, hereditary spastic paraplegia and Friedreich's ataxia. Biochim Biophys Acta 1410:159, 1999

28. Sherer TB, Betarbet R, Stout AK, et al: An in vitro model of Parkinson's disease: linking mitochondrial impairment to altered alpha-synuclein metabolism and oxidative damage. J Neurosci 22:7006, 2002

29. Orr CF, Rowe DB, Halliday GM: An inflammatory review of Parkinson's disease. Prog Neurobiol 68:325, 2002

30. Zhou Y, Shie FS, Piccardo P, et al: Proteasomal inhibition induced by manganese ethylene-bis-dithiocarbamate: relevance to Parkinson's disease. Neuroscience 128:281, 2004

31. Bennett JP: Free radicals, oxidative stress and the origin of Parkinson's disease. J Neurol Sci 170:75, 1999

32. Forloni G, Bertani I, Calella AM, et al: Alpha-synuclein and Parkinson's disease: selective neurodegenerative effect of alpha-synuclein fragment on dopaminergic neurons in vitro and in vivo. Ann Neurol 47:632, 2000

33. Abeliovich A, Schmitz Y, Farinas I, et al: Mice lacking α-synuclein display functional deficits in the nigrostriatal system. Neuron 25:239, 2000

34. Mouradian MM: Recent advances in the genetics and pathogenesis of Parkinson disease. Neurology 58:179, 2002

35. Rajput A: Prevalence of dementia in Parkinson's disease. Parkinson's Disease: Neurobehavioral Aspects. Huber S, Cummings JL, Eds. Oxford University Press, New York, 1992, p 119

36. Ditter S, Mirra SS: Neuropathologic and clinical features of Parkinson's disease in Alzheimer's disease patients. Neurology 37:754, 1987

37. Cummings JL, Masterman DL: Depression in patients with Parkinson's disease. Int J Geriatr Psychiatry 14:711, 1999

38. Bliwise D, Watts RL, Watts N, et al: Disruptive nocturnal behavior in Parkinson's disease and Alzheimer's disease. J Geriatr Psychiatry Neurol 8:107, 1995

39. Pappert EJ, Goetz CG, Niederman FG, et al: Hallucinations, sleep fragmentation, and altered dream phenomena in Parkinson's disease. Mov Disord 14:117, 1999

40. Reuter I, Engelhardt M, Stecker K, et al: Therapeutic value of exercise training in Parkinson's disease. Med Sci Sports Exerc 31:1544, 1999

41. Cotzias GC, Papavasiliou PS, Gellene R: Modification of Parkinsonism: chronic treatment with L-dopa. N Engl J Med 280:337, 1969

42. Agid Y: Levodopa: is toxicity a myth? Neurology 50:858, 1998

43. Levodopa and the progression of Parkinson's disease. The Parkinson Study Group. N Engl J Med 351:2498, 2004

44. Caraceni T, Scigliano G, Musicco M: The occurrence of motor fluctuations in parkinsonian patients treated long term with levodopa: role of early treatment and disease progression. Neurology 41:380, 1994

45. Ahlskog JE: Treatment of early Parkinson's disease: are complicated strategies justified? Mayo Clin Proc 71:659, 1996

46. Rascol O, Brooks DJ, Korczyn AD, et al: A five-year study of the incidence of dyskinesia in patients with early Parkinson's disease who were treated with ropinirole or levodopa. N Engl J Med 342:1484, 2000

47. Holloway R: Pramipexole versus levodopa in early Parkinson's disease: a randomized clinical trial. Parkinson Study Group. JAMA 284:1931, 2000

48. Barone P, Bravi D, Bermejo-Parejo F, et al: Pergolide monotherapy in the treatment of early PD: a randomized, controlled study. Pergolide Monotherapy Study Group. Neurology 53:573, 1999

49. Pramipexole vs levodopa as initial treatment for Parkinson disease: a 4-year randomized controlled trial. Parkinson Study Group. Arch Neurol 61:1044, 2004

50. Koller WC, Hutton JT, Tolosa E, et al: Immediate-release and controlled-release carbidopa/levodopa in PD. Carbidopa/Levodopa Study Group. Neurology 53:1012, 1999

51. Goldman SM, Tanner C: Etiology of Parkinson's disease. Parkinson's Disease and Movement Disorders. Jankovic J, Tolosa E, Eds. Williams & Wilkins, Baltimore, 1998, p 133

52. Block G, Liss C, Reines S, et al: Comparison of immediate-release and controlled-release carbidopa/levodopa in Parkinson's disease. Eur Neurol 37:23, 1997

53. Youdim MB, Riederer PF: A review of the mechanisms and role of monoamine oxidase inhibitors in Parkinson's disease. Neurology 63(7 suppl 2):S32, 2004

54. Lees A: Comparison of therapeutic effects and mortality data of levodopa and levodopa combined with selegiline in patients with early, mild Parkinson's disease. BMJ 311:1602, 1995

55. Mannisto PT, Kaakkola S: Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. Pharmacol Rev 51:593, 1999

56. Deane K, Spieker S, Clarke C: Catechol-O-methyltransferase inhibitors for levodopa-induced complications in Parkinson's disease. Cochrane Database Syst Rev (4):CD004554, 2004

57. Olanow CW: Tolcapone and hepatotoxic effects. Tasmar Advisory Panel. Arch Neurol 57:263, 2000

58. Deane K, Spieker S, Clarke C: Catechol-O-methyltransferase inhibitors versus active comparators for levodopa-induced complications in Parkinson's disease. Cochrane Database Syst Rev (4):CD004553, 2004

59. Baseman DG, O'Suilleabhain PE, Reimold SC, et al: Pergolide use in Parkinson disease is associated with cardiac valve regurgitation. Neurology 63:301, 2004

60. Hauser RA, Gauger L, Anderson M, et al: Pramipexole-induced somnolence and episodes of daytime sleep. Mov Disord 15:658, 2000

61. Verhagen ML, Del Dotto P, van den Munckof P, et al: Amantadine as treatment for dyskinesias and motor fluctuations in Parkinson's disease. Neurology 50:1323, 1998

62. Ahlskog JE: Medical treatment of later-stage motor problems in Parkinson's disease. Mayo Clinic Proc 74:1239, 1999

63. Kennedy R, Mittal D, O'Jile J: Electroconvulsive therapy in movement disorders: an update. J Neuropsychiatry Clin Neurosci 15:407, 2003

64. Juncos JL: Management of psychotic aspects of Parkinson's disease. J Clin Psychiatry 60:42, 1999

65. Friedman H, Factor SA: Atypical antipsychotics in the treatment of drug-induced psychosis in Parkinson's disease. Mov Disord 15:201, 2000

66. Goetz CG, Blasucci LM, Leurgans S, et al: Olanzapine and clozapine: comparative effects on motor function in hallucinating PD patients. Neurology 55:789, 2000

67. Emre M, Aarsland D, Alanese A, et al: Rivastigmine for dementia associated with Parkinson's disease. N Engl J Med 351:2509, 2004

68. Schapira AH, Olanow CW: Neuroprotection in Parkinson disease: mysteries, myths, and misconceptions. JAMA 291:358, 2004

69. Impact of deprenyl and tocopherol treatment on Parkinson's disease in DATATOP subjects not requiring levodopa. Parkinson Study Group. Ann Neurol 39:29, 1996

70. Effects of coenzyme Q10 in early Parkinson disease: evidence of slowing of the functional decline. Parkinson Study Group. Arch Neurol 59:1541, 2002

71. Hart RG, Ravina BM: Randomized clinical trials of neuroprotective agents in Parkinson's disease: the NINDS registry. National Institute of Neurological Disorders and Stroke, Bethesda, Maryland, 2003
http://www.ninds.nih.gov/funding/research/parkinsonsweb/neuroprotection_trials.pdf

72. Lyons KE, Pahwa R: Deep brain stimulation in Parkinson's disease. Curr Neurol Neurosci Rep 4:290, 2004

73. Linazasoro G: Recent failures of new potential symptomatic treatments for Parkinson's disease: causes and solutions. Mov Disord 19:743, 2004

74. Nutt JG, Burchiel KJ, Comella CL, et al: Randomized, double-blind trial of glial cell line–derived neurotrophic factor (GDNF) in PD. Neurology 60:69, 2003

75. Hardie R, Lees AJ: Neuroleptic-induced Parkinson's syndrome: clinical features and results of treatment with levodopa. J Neurol Neurosurg Psychiatry 51:850, 1988

76. Thompson P, Marsden CD: Gait disorder of subcortical arteriosclerotic encephalopathy: Binswanger's disease. Mov Disord 2:1, 1987

77. Gilman S, Low PA, Quinn N, et al: Consensus statement on the diagnosis of multiple system atrophy. J Autonom Nerv Syst 11:74:189, 1998

78. Wenning GK, Ben Shlomo Y, Magalhaes M, et al: Clinical features and natural history of multiple system atrophy: an analysis of 100 cases. Brain 117:835, 1994

79. Conrad C, Andreadis A, Trojanowski JQ, et al: Genetic evidence for the involvement of tau in progressive supranuclear palsy. Ann Neurol 41:277, 1997

80. Bennett P, Bonifati V, Bonuccelli U, et al: Direct genetic evidence for involvement of tau in progressive supranuclear palsy. European Study Group on Atypical Parkinsonism Consortium. Neurology 51:982, 1998

81. Higgins JJ, Litvan I, Nee LE, et al: A lack of the R406W tau mutation in progressive supranuclear palsy and corticobasal degeneration. Neurology 52:404, 1999

82. Tolosa E, Duvoisin R, Cruz-Sanchez FF: Progressive Supranuclear Palsy: Diagnosis, Pathology, and Therapy. Springer-Verlag, New York, 1994

83. Grimes DA, Lang AE, Bergeron CB: Dementia as the most common presentation of cortical-basal ganglionic degeneration. Neurology 53:1969, 1999

84. Litvan I, Goetz CG, Lang AE: Corticobasal degeneration and related disorders. Adv Neurol 82:1, 2000

85. Mega M, Masterman DL, Benson F, et al: Dementia with Lewy bodies: reliability and validity of clinical and pathologic criteria. Neurology 47:1403, 1996

86. Ondo WG, Hanna PA, Jankovic J: Tetrabenazine treatment for tardive dyskinesia: assessment by randomized videotape protocol. Am J Psychiatry 156:1279, 1999

87. Martin JB: Molecular basis of neurodegenerative disorders. N Engl J Med 340:1970, 1999

88. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Huntington's Disease Collaborative Research Group. Cell 72:971, 1993

89. Scherzinger E, Lurz R, Turmaine M, et al: Huntingtin-encoded polyglutamine expansions form amyloid-like aggregates in vitro and in vivo. Cell 90:549, 1997

90. Ona VO, Li M, Vonsattel JPG, et al: Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. Nature 399:263, 1999

91. Yamamoto A, Lucas JJ, Hen R: Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. Cell 101:57, 2000

92. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. Mov Disord 11:136, 1996

93. Hersch S, Jones R, Koroshetz W, et al: The neurogenetics genie: testing for the Huntington's disease mutation. Neurology 44:1369, 1994

94. Kremer B, Clark CM, Almqvist EW, et al: Influence of lamotrigine on progression of early Huntington disease. Neurology 53:1000, 1999

95. Fahn S, Marsden CD, DeLong MR: Dystonia 3. Adv Neurol 78:1, 1998

96. Ozelius LJ, Hewett JW, Page CE, et al: The gene (DYT1) for early-onset torsion dystonia encodes a novel protein related to the Clp protease/heat shock family. Adv Neurol 78:93, 1998

97. Furukama Y, Shimadzu M, Rajput AH, et al: GTP-cyclohydrolase-1 gene mutations in hereditary progressive and dopa-responsive dystonia. Ann Neurol 39:609, 1996

98. Klein C, Gurvich N, Sena-Esteves M, et al: Evaluation of the role of the D2 dopamine receptor in myoclonus dystonia. Ann Neurol 47:369, 2000

99. Nemeth AH, Nolte D, Dunne E, et al: Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). Genomics 60:320, 1999

100. Kramer PL, Mineta M, Klein C, et al: Rapid-onset dystonia-parkinsonism: linkage to chromosome 19q13. Ann Neurol 46:176, 1999

101. Karp BI, Goldstein SR, Chen R, et al: An open trial of clozapine for dystonia. Mov Disord 14:652, 1999

102. Lai T, Lai JM, Grossman RG: Functional recovery after bilateral pallidotomy for the treatment of early-onset primary generalized dystonia. Arch Phys Med Rehabil 80:1340, 1999

103. Coubes P, Cif L, El Fertit H, et al: Electrical stimulation of the globus pallidus internus in patients with generalized dystonia: long-term results. J Neurosurg 101:189, 2004

104. Gitlin JD: Wilson disease. Gastroenterology 125:1868, 2003

105. Petrukhin K, Fischer SG, Pirastu M, et al: Mapping, cloning and genetic characterization of the region containing the Wilson disease gene. Nat Genet 5:338, 1993

106. Thomas G, Forbes JR, Roberts EA, et al: The Wilson disease gene: spectrum of mutations and their consequences. Nat Genet 9:210, 1995

107. Petrukhin K, Lutsenko S, Chernov I, et al: Characterization of the Wilson disease gene encoding a P-type copper transporting ATPase: genomic organization, alternative splicing, and structure/function predictions. Hum Mol Genet 3:1647, 1994

108. Brewer GJ: Penicillamine should not be used as initial therapy in Wilson's disease. Mov Disord 14:551, 1999

109. Walshe JM: Penicillamine: The treatment of first choice for patients with Wilson's disease. Mov Disord 14:542, 1999

110. Sutcliffe RP, Maguire DD, Muiesan P, et al: Liver transplantation for Wilson's disease: long-term results and quality-of-life assessment. Transplantation 75:1003, 2003

111. Seuchter SA, Hebebrand J, Klug B, et al: Complex segregation analysis of families ascertained through Gilles de la Tourette syndrome. Genet Epidemiol 18:33, 2000

112. Freeman RD, Fast DK, Burd L, et al: An international perspective on Tourette syndrome: selected findings from 3,500 individuals in 22 countries. Dev Med Child Neurol 42:436, 2000

113. Jankovic J: Tourette's syndrome. N Engl J Med 345:1184, 2001

114. Stamenkovic M, Schindler SD, Aschauer HN, et al: Effective open-label treatment of Tourette's disorder with olanzapine. Int Clin Psychopharmacol 15:23, 2000

115. Louis ED: Essential tremor. Arch Neurol 57:1522, 2000

116. Fahn S: Overview, history, and classification of myoclonus. Adv Neurol 89:13, 2002

117. Caviness JN, Alving LI, Maraganore DM, et al: The incidence and prevalence of myoclonus in Olmsted county, Minnesota. Mayo Clin Proc 74:565, 1999

118. Berkovic SF, Andermann F, Carpenter S, et al: Progressive myoclonus epilepsies: specific causes and diagnosis. N Engl J Med 315:296, 1986

119. Drug-Induced Movement Disorders. Lang EA, Weiner WJ, Eds. Futura Publishing, New York, 1992

120. Elble R, Koller WC: Tremor. Johns Hopkins University Press, Baltimore, 1990

121. Cunningham-Owens DG: A Guide to the Extrapyramidal Side-effects of Antipsychotic Drugs. Cambridge University Press, Cambridge, England, 1999

122. Adler LA, Rotrosen J, Edson R, et al: Vitamin E treatment for tardive dyskinesia. Veterans Affairs Cooperative Study #394 Study Group. Arch Gen Psychiatry 56:836, 1999

Acknowledgments

This work was supported in part by a Center of Excellence grant to Emory University School of Medicine from the American Parkinson's Disease Association, New York. Dr. Juncos is supported by NIH RO1 AT00612-01 and the Fauver Family Foundation.

Figures 1 and 2 Seward Hung.

Figure 4 Courtesy of Thomas M. Aaberg, Sr., M.D., Department of Ophthalmology, Emory University School of Medicine.

XVI ACUTE VIRAL CENTRAL NERVOUS SYSTEM DISEASES

Donald H. Gilden, m.d.

Viral infections of the central nervous system produce aseptic meningitis, encephalitis, and sometimes vasculopathy. Although nearly all the viruses discussed in this chapter are capable of producing meningitis and encephalitis, each virus typically causes one of these diseases. Viruses that most often cause acute aseptic meningitis include the enteroviruses (i.e., coxsackievirus, echovirus, and nonparalytic polioviruses) and the influenza and parainfluenza viruses (especially mumps virus). Herpes simplex virus type 2 (HSV-2) and, rarely, HSV type 1 (HSV-1) cause recurrent meningitis. Viruses that primarily produce acute encephalitis include the herpesviruses (i.e., HSV-1, varicella-zoster virus [VZV], and Epstein-Barr virus [EBV]), the togaviruses (i.e., St. Louis encephalitis virus, the California encephalitis virus group, the equine encephalitis viruses, Powassan virus, and Colorado tick fever virus), the arenaviruses (lymphocytic choriomeningitis [LCM] virus), and the adenoviruses.

The time of year that an infection manifests itself may be a clue for identifying the causative virus [*see Table 1*]. For example, enterovirus infections occur in the summer and account for most cases of aseptic meningitis. Mumps and other parainfluenza virus infections, as well as influenza virus infection, most often present in the winter, although they may occur at any time of year. Togavirus infection occurs in the late summer and autumn (the mosquito and tick season); LCM infection presents in the early winter (when mice come indoors); and chickenpox (varicella) occurs predominantly in the spring. All HSV infections occur year-round.

Approach to Viral Aseptic Meningitis and Encephalitis

DIAGNOSIS

Clinical Features

Aseptic meningitis is associated with headache, fever, and stiff neck. Irritability and alterations in state of consciousness may occur, although these symptoms are more common in encephalitis. Patients may suffer sore throat, photophobia, abdominal discomfort, focal paresthesias, diplopia, and nausea and vomiting. An erythematous maculopapular rash may result from enterovirus (echovirus or coxsackievirus) infection; a vesicular eruption suggests coxsackievirus or a herpesvirus infection. Stretching the irritated nerve roots may elicit pain. For example, in the recumbent patient, the Kernig sign is pain produced by attempts to extend the leg after flexion of the thigh. The Brudzinski sign is flexion of the thighs when the head is passively flexed onto the chest. These signs of meningeal irritation appear more often in nonviral and bacterial meningitis, however.

Clinical symptoms and signs of viral encephalitis include insomnia, lethargy, mental status changes, seizures, aphasia, and hemiplegia. Focal deficit is most common in HSV and VZV encephalitis and less common in cytomegalovirus (CMV) and EBV encephalitis. Brain imaging is usually normal but may reveal effacement [*see Figure 1a*] and an increase in water content [*see Figure 1b*], which is indicative of brain swelling.

Diagnostic Tests

Clinical symptoms and signs alone usually cannot establish a specific diagnosis of viral meningitis or encephalitis, and nonviral causes of meningitis and encephalitis can be confused clinically with viral infections [*see Table 2*].

Careful examination of the cerebrospinal fluid is the mainstay of diagnosis of viral meningitis or encephalitis. Characteristically, the CSF is clear and features a predominantly mononuclear pleocytosis and normal glucose content. Initially, the CSF may contain polymorphonuclear leukocytes. The CSF cell count is usually below 100/mm^3; it may be higher with enteroviral infections, however, and the CSF may contain thousands of mononuclear cells after mumps and LCM virus infections. The CSF protein concentration is usually normal or mildly elevated. Bacteria are not found on Gram stain, and CSF cultures are sterile. A mild depression in the CSF glucose content develops in approximately one third of patients with meningitis caused by mumps or LCM virus; this drop in CSF glucose level occurs less often after enterovirus infection. In rare instances, the CSF glucose content is depressed in aseptic meningitis caused by HSV or VZV.

Most important is that a mononuclear pleocytosis with hypoglycorrhachia (low CSF glucose content) occurs in some nonviral forms of meningitis and requires prompt treatment. Such nonviral forms include tuberculous (TB) or fungal menin-

Table 1 Important Clinical Features of the Major Virus Infections of the Nervous System

Virus	*Seasons*	*Major Acute Presentation*	*Brain Imaging*
Herpes simplex virus type 1	All	Encephalitis	Medial temporal lobe
Varicella-zoster virus	All*	Localized herpes zoster, postherpetic neuralgia, encephalitis with focal deficit	Multifocal pale and hemorrhagic infarction
Enterovirus	Summer	Aseptic meningitis	Normal; possible effacement
HIV	All	Aseptic meningitis, subacute encephalitis	Atrophy; diffuse white matter attenuation
Arthropod-borne encephalitis	Summer and fall	Encephalitis	Normal; possible effacement

*Although chickenpox (varicella) occurs mostly in the spring, zoster develops at any time of the year.

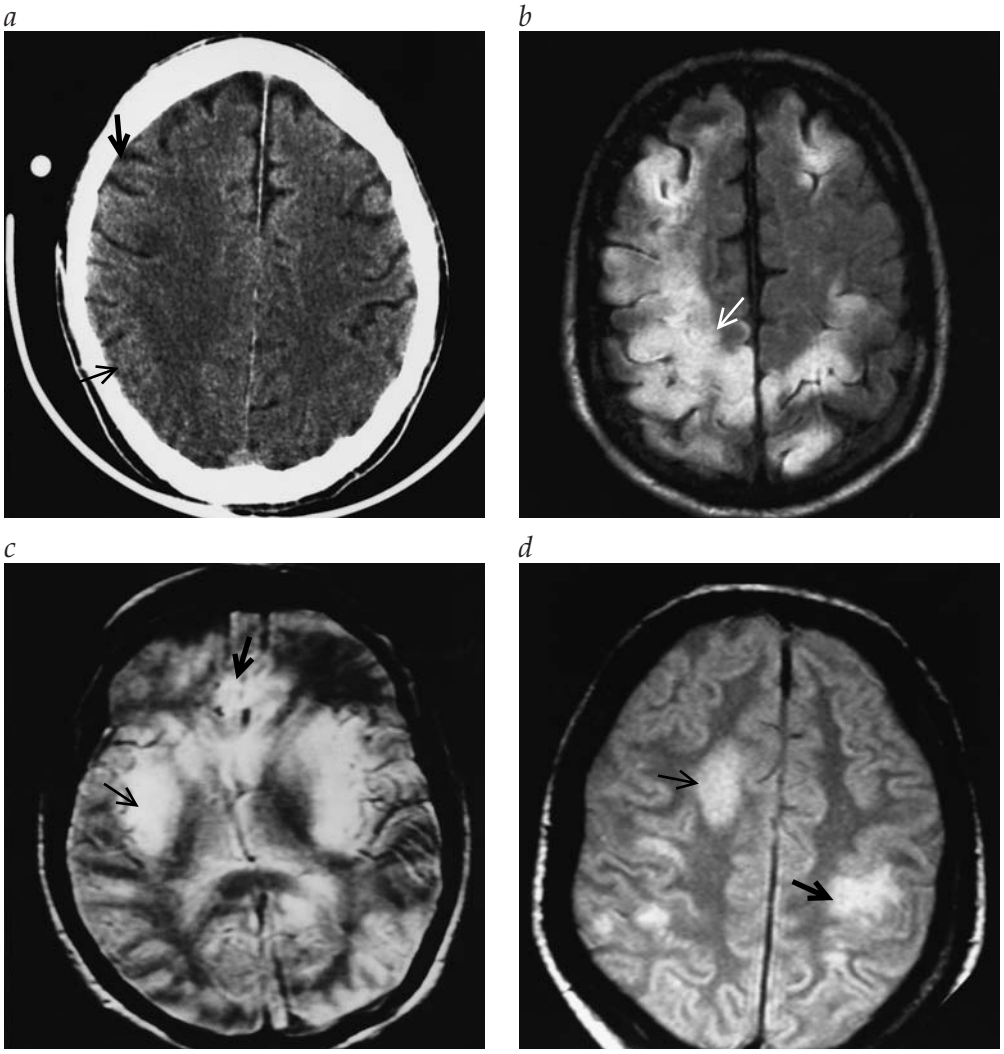


Figure 1 **(*a*) CT changes that occur in most viral encephalitides. Computed tomography scan of the brain demonstrates relative effacement of sulci posteriorly in both hemispheres (thin arrow), compared with normal sulcal spaces anteriorly (thick arrow). (*b*) MRI changes that occur in most viral encephalitides. T_2-weighted inversion recovery (fluid-attenuated inversion recovery [FLAIR]) MRI brain scan of the same patient demonstrates areas of increased signal in both hemispheres, with the increased signal being greater on the right and even more so posteriorly (arrow), reflecting increased water content in mildly swollen brain. (*c*) Herpes simplex virus encephalitis. T_2-weighted MRI brain scan demonstrates bilateral involvement of temporal lobes. The exaggerated signal does not extend beyond the insular cortex (thin arrow) but does involve the cingulate gyrus (thick arrow). (*d*) Varicella-zoster virus encephalitis. Proton-density brain MRI scan shows multiple areas of infarction in both hemispheres, particularly involving white matter (thin arrow) and extending to gray-white–matter junctions (thick arrow).**

gitis and carcinomatous or lymphomatous meningitis; those CSF findings also occur in about 30% of cases of sarcoid meningitis. Furthermore, viral aseptic meningitis may be mimicked by granulomatous meningitis (i.e., TB, fungal, sarcoid, and toxoplasmic meningitides), leptospiral infection (e.g., Lyme disease), *Brucella* infection, *Mycoplasma* pneumonia, sulfonamide or nonsteroidal anti-inflammatory drug toxicity, and various vasculitides.

Laboratory methods that are used to identify the causative agent of viral meningitis and encephalitis include polymerase chain reaction (PCR) detection of viral RNA or DNA in the CSF, the culturing of virus from the CSF, the detection of antigen or antibody in the CSF, serologic tests (e.g., enzyme immunoassays and Western blot assay), and brain biopsy [*see Table 3*].

NONSPECIFIC TREATMENT OF VIRAL CNS INFECTIONS

Bed rest is indicated for patients with viral CNS infection. If a patient has recently traveled to an area endemic for highly contagious agents that are spread by direct patient contact (e.g., Ebola virus), isolation procedures should be followed until the cause of the illness is established. Most other agents that cause viral meningitis and encephalitis present no special dangers to medical personnel, provided antiviral precautions for blood products are followed, or to others. The patient's airway can be protected by repeated suctioning and, if necessary, by endotracheal intubation or tracheostomy. Nutritional support is important, but only conscious patients with intact brain stem function should receive oral feedings. All others should be fed by parenteral or nasogastric techniques. Patients should be monitored closely for

secondary infections, especially of the urinary tract and lungs. Passive range-of-motion exercises and use of foot boards or Styrofoam boots will minimize contractures. Frequent turning decreases the risk of bedsores. Patients with encephalitis may develop the syndrome of inappropriate antidiuretic hormone secretion and become water intoxicated, which can lead to deepening coma and, sometimes, seizures. Serum and urine electrolyte levels and urine output should be monitored frequently. Encephalitis is treated by restricting fluid intake and, in rare instances, with hypertonic saline solution. Headache is a frequent problem and can be treated with acetaminophen, ibuprofen, and codeine; occasionally, meperidine (50 mg every 4 to 5 hours) is used. Most viruses are thermolabile, and modest temperature elevation may serve as a natural defense mechanism; but fever can be treated with acetaminophen if necessary. Specific viruses that cause aseptic meningitis, encephalitis, or both may also require particular treatment regimens.

Table 3 Viral Diagnosis of Aseptic Meningitis Syndrome and Encephalitis

Virus	*Tissue Culture*	*Antigen/Antibody Detection in CSF**	*PCR*
Adenovirus	×		
Cytomegalovirus	×	×	×[69,70]
Enteroviruses	×	×	×[5]
Epstein-Barr virus		×	×[42]
Herpes simplex virus type 1	×	×	×[19,20]
Herpes simplex virus type 2	×	×	×[16,17]
HIV		×	
Influenza		×	
Measles		×	
Mumps		×	
Poliovirus	×	×	
Varicella-zoster virus	×	×	×[71]

*Results are not usually available during acute disease.

Viruses That Cause Meningitis or Encephalitis

ENTEROVIRUSES

Enteroviruses are the most common cause of viral aseptic meningitis. Most, but not all, cases of enterovirus infection occur in children. Epidemics are common. The Picornaviridae family comprises nearly 70 distinct serotypes. Enterovirus infections are acquired primarily by the fecal-oral route and occasionally by the respiratory route. Fever is often biphasic. The first phase is associated with constitutional and gastroenterologic symptoms. After resolution of these symptoms, the fever may reappear with other signs and symptoms of meningeal irritation. Almost all patients with enterovirus meningitis are immunocompetent. In agammaglobulinemic persons, however, enterovirus infection may result in a chronic meningitis or meningoencephalitis that lasts years, and the outcome is often fatal.[1] The CSF contains between 50 and 1,000 cells, which are predominantly mononuclear. Elevated CSF protein levels are common. The CSF glucose content is usually normal, but hypoglycorrhachia has been reported.[2] The optimal time to culture virus from the CSF is 8 to 18 days after infection; PCR amplification of the conserved region of enterovirus is possible.[3] Enterovirus antibodies appear in the CSF at the end of the second week and persist longer than 1 month.

In rare instances, enteroviruses cause focal encephalitis. In addition, enterovirus infections have been associated with ataxia, opsoclonus and myoclonus,[4] and parkinsonism. In the 1998 enterovirus 71 epidemic in Taiwan, the main neurologic complication was rhombencephalitis. Myoclonic jerks were the most common initial symptoms, and magnetic resonance imaging usually showed evidence of brain stem involvement.[5] Enteroviral infections may lead to myelitis and, occasionally, neuropathy (e.g., cranial nerve palsies). Since the eradication of poliovirus in Brazil, enterovirus 71 has emerged as a cause of persistent flaccid paralysis[6]; acute pulmonary edema is common in enterovirus 71 encephalomyelitis.[7] Chronic progressive echovirus and coxsackievirus meningoencephalitides in hypogammaglobulinemic and agammaglobulinemic persons have been treated successfully with immunoglobulin.[8]

Table 2 Nonviral Causes of Meningitis and Encephalitis

Behçet syndrome
Brucella infection
Cryptococcosis
Cysticercosis
Fungus
Granulomatous angiitis
Histoplasmosis
Leptospirosis (Weil disease)
Lyme disease
Mollaret meningitis*
Mycoplasma infection
Rocky Mountain spotted fever
Sarcoidosis
Syphilis
Systemic lupus erythematosus
Toxoplasmosis
Tuberculosis
Uveomeningoencephalitis (Vogt-Koyanagi-Harada syndrome)†
Whipple disease
Sulfonamides and nonsteroidal anti-inflammatory drugs‡

*Many cases are found to be herpesvirus infections when studied by polymerase chain reaction.
†This may be viral if not autoimmune.
‡These agents mimic granulomatous meningitis.

POLIOVIRUS

Although immunization has nearly eradicated encephalomyelitis caused by type 1 poliovirus, a case of encephalomyelitis with oral-facial dyskinesias and quadriplegia was attributed to type 3 poliovirus on the basis of positive stool cultures and high neutralizing serum titers.[9]

In the past, immunization against polio was accomplished with an oral vaccine, which contained attenuated live virus. Rare reports of oral vaccine–associated polio led to a recommendation by the Advisory Committee on Immunization Practices that as of January 1, 2000, only inactivated poliovirus vaccine (IPV) should be used for routine childhood polio vaccination in the United States. The current schedule involves a total of four doses of IPV: the first two doses are administered at 2 months and 4 months of age, the third dose is administered at 6 to 18 months of age, and the final dose is administered at 4 to 6 years of age.[10]

MUMPS

Aseptic meningitis is a common complication of mumps virus infection. Widespread vaccination against mumps has dramatically reduced the incidence of mumps and, consequently, of mumps meningitis. Mumps vaccine contains attenuated virus, however, and outbreaks of mumps meningitis have been traced to the use of incompletely attenuated mumps virus strains.

Mumps virus is a paramyxovirus that is transmitted by respiratory droplets. After the initial mucosal infection, viremia can lead to involvement of many organs, including the heart and, in particular, the glands. The association of aseptic meningitis with parotitis, pancreatitis (abdominal pain and increased serum amylase levels), or oophoritis (abdominal pain and ovarian tenderness on pelvic examination) is a clue that mumps virus is the causative agent. One study revealed that aseptic meningitis may develop in as many as 23% of patients with mumps, whereas encephalitis occurs in only one in 6,000 cases. Months of persistent pleocytosis[11] and a low CSF glucose content or chronic encephalomyelitis[12] may follow mumps meningitis. Mumps meningoencephalitis is rarely fatal. Because mumps virus replicates in ependymal cells, aqueductal stenosis and hydrocephalus may ensue months to years after an otherwise benign case of aseptic meningitis.

INFLUENZA VIRUS

Influenza is another myxovirus that causes neurologic disease. Acute encephalitis develops during the height of influenza epidemics. Mental and motor deficits (including ataxia) and a CSF lymphocytic pleocytosis, in association with rising complement-fixing and hemagglutination-inhibiting antibody titers to influenza virus, provide presumptive evidence of influenza encephalitis. In seven patients with CNS disease who had recently been infected with influenza virus, PCR assay of the CSF for influenza virus was helpful in making the diagnosis.[13] During the winter of 1998 to 1999, an outbreak of encephalitis/encephalopathy in Japan was associated with influenza. Of 202 cases analyzed, 148 were confirmed virologically.[14] Seizures were the most frequent presenting feature; mental status changes were common. Many patients experienced multiple-organ failure; the mortality was 31.8%. Thrombocytopenia and severely elevated transaminase levels indicated a poor prognosis.

NIPAH VIRUS

Nipah virus is a paramyxovirus that is found in the respiratory tract of Malaysian pigs and that can cause human disease [*see 7:XXXI Viral Zoonoses*]. Outbreaks of encephalitis from Nipah virus have occurred in Malaysia and Singapore.[15]

HERPES SIMPLEX VIRUS TYPE 2

HSV-2 Aseptic Meningitis

HSV-2 aseptic meningitis is the main neurologic complication of HSV-2 infection. HSV-2 causes genital herpes. In the United States, HSV-2 is the third most common cause of aseptic meningitis, accounting for approximately 5% of all cases. Unlike viral meningitides that have a seasonal association, HSV-2 meningitis occurs at any time of year. The typical symptoms and signs are headache, fever, stiff neck, and a marked lymphocytic pleocytosis of the CSF. Meningitis may be preceded by genital or pelvic pain, and the astute clinician who suspects HSV-2 meningitis will ask about recent symptoms of pelvic inflammatory disease or penile lesions. The workup for suspected HSV-2 meningitis should include a careful search for vesicular lesions over the external genitalia and a pelvic examination for lesions in the vagina or on the cervix. The cause of aseptic meningitis in one patient with recurrent dermatomal skin lesions was shown to be HSV-2.[16] PCR has also revealed that the primary agent causing benign recurrent lymphocytic meningitis is HSV-2.[17] Occasionally, HSV-1 is the culprit, as evidenced by the detection by PCR of HSV-1 DNA in the CSF of patients with benign recurrent lymphocytic meningitis.[18] HSV-2 meningitis is self-limited; antiviral treatment is not required.

HSV-2 Encephalitis

HSV-2 encephalitis is a rare disease; it occurs most often in newborns and, in rare instances, in immunocompromised adults. In both of these populations, CNS infection by HSV-2 is diffuse, unlike the focal medial-temporal and orbital-frontal lesions observed in cases of HSV-1 infection. Nevertheless, seizures, alterations in state of consciousness, and focal neurologic deficits—common clinical features of HSV-1 encephalitis—also characterize HSV-2 encephalitis. AIDS predisposes a person to HSV-2 encephalitis, which may occur in association with CNS infection by such other opportunistic agents as cytomegalovirus. HSV-2 encephalitis is treated with 15 to 30 mg/kg/day of acyclovir. After 10 to 14 days of intravenous acyclovir treatment, AIDS patients with HSV encephalitis may require indefinite maintenance on famciclovir or valacyclovir to prevent further reactivation of the disease.

HERPES SIMPLEX VIRUS TYPE 1

HSV-1 Encephalitis

Unlike most viral encephalitides, HSV-1 encephalitis is focal. HSV replication in the medial temporal lobe and orbital surface of the frontal lobe, with accompanying inflammation, produces the characteristic clinical picture.

Clinical features Fever, headache, lethargy, irritability, and confusion are typical. Seizures (major motor, complex partial, focal, and even absence attacks) affect approximately 40% of patients. If the dominant temporal lobe is involved, aphasia and focal motor or sensory deficits develop. Progressive hemorrhagic necrosis and temporal lobe edema may lead to uncal herniation, the most common features of which are tachycardia, hyperventilation, flexor (and later extensor) posturing, and a dilated pupil (usually on the side of the herniated temporal lobe).

Because HSV infection becomes latent and periodically reactivates to produce recurrent herpes labialis, the misconception exists that HSV encephalitis is usually protracted or chronic. Although survivors of HSV encephalitis may have a permanent seizure disorder, mental changes, aphasia, or motor deficit, the onset of neurologic disease is usually acute or subacute, as in other viral encephalitides, and early treatment is crucial to a favorable outcome. Before acyclovir became available for treatment, approximately two thirds of patients with HSV encephalitis died.

CSF findings The CSF is usually abnormal in HSV encephalitis. The CSF opening pressure is often elevated and may be very high if there is brain swelling and impending temporal lobe herniation. CSF examination is usually performed in the first few days of illness, before significant brain swelling develops, to decrease the potential for herniation after lumbar puncture. The

electroencephalogram and imaging studies may demonstrate features highly suggestive of HSV encephalitis, often obviating subsequent lumbar punctures. CSF pleocytosis is observed in more than 90% of patients, although its absence at initial evaluation does not rule out HSV encephalitis. The CSF cell count ranges from 4 to 755/mm^3, and more than 200/mm^3 may be present weeks after the onset of disease. The predominant cell type is mononuclear. Although red blood cells (RBCs) are unusual in other viral encephalitides, in HSV encephalitis they are often present in the CSF, which may also be xanthochromic; this presumably reflects the hemorrhagic nature of brain lesions. Instead of attributing the presence of RBCs in CSF to a so-called traumatic tap, the astute clinician may use this finding to support the presumptive diagnosis of HSV encephalitis.

The majority of patients with HSV encephalitis have elevated CSF protein. In rare instances, hypoglycorrhachia occurs. Increased levels of antibody to HSV, suggestive of recent infection, may be found in serum and CSF; increased CSF-to-serum ratios of anti-HSV antibody may help in making the diagnosis of HSV encephalitis. Unfortunately, increased antibody titers are not usually detected until 2 weeks or longer after the onset of disease; thus, their practical value lies more in retrospective presumptive diagnosis than in identifying acute encephalitis. Although HSV can often be isolated from cerebral biopsy or autopsy material, the isolation of HSV from CSF during acute disease is unusual. PCR detection of HSV-1 DNA in the CSF is both sensitive and specific and has become the gold standard in suspected cases of HSV encephalitis.[19] Nonetheless, clinicians should be aware that PCR may be negative for HSV in the first few days of illness with HSV encephalitis.[20]

Electroencephalography Early in the course of the disease, background disorganization with generalized or focal slowing occurs, predominantly over the involved temporal region. Within days, widespread, periodic, stereotyped sharp-wave and slow-wave complexes develop, usually at regular intervals of 2 to 3 seconds.[21] Bilateral periodic complexes appear if both sides of the brain are involved. Although these features can be seen in other CNS disorders (e.g., tumor, abscess, syphilis, infarct), their presence in the clinical setting of fever and rapidly progressive neurologic disease provides strong presumptive evidence for HSV encephalitis.

Imaging studies Computed tomographic scanning shows hypodense lesions involving the medial temporal regions. An important diagnostic clue is a sharp transition from the hypodense temporal lesion to the normal basal ganglia.[22] Edema and mass effect occur in 80% of cases, and contrast enhancement appears in more than 50% of cases.

MRI reveals a decrease in T_1 and an increase in T_2 signal [*see Figure 1c*], and the signal abnormality includes a larger area of brain than is usually found with CT scanning. Unlike with CT scanning, MRI of the temporal lobes is not subject to artifact from the petrous and sphenoid bones, which often obscure the temporal fossa. A comparison between CT and MRI in four cases of HSV encephalitis showed that temporal lobe inflammation was obvious on MRI days before changes were visible on CT.[23]

Differential diagnosis Other viral and bacterial infections, cerebral abscess, tumor, and stroke can mimic the clinical features of HSV encephalitis. A focal area of decreased density and increased vascularity that extends into subcortical areas is consistent with the cerebritis that heralds frank abscess formation and helps distinguish evolving abscess from HSV encephalitis. Ring-enhancing lesions characteristic of cerebral abscess or tumor are not seen in the first few days of HSV encephalitis. Glioma and infarct are not likely to be associated with fever, and brain imaging reveals that they are not restricted to the medial temporal lobe.

Treatment Patients with HSV encephalitis should be treated with intravenous acyclovir, 15 to 30 mg/kg/day in three divided doses for at least 10 days. Early institution of acyclovir has reduced the mortality—which ranged between 60% and 70%—to slightly lower than 30%.[24] Generally, early treatment (before coma ensues) is associated with a more favorable outcome. Acyclovir appears to be safe; only mild hematologic, hepatic, and renal function abnormalities have been reported, although relapses after acyclovir therapy have been documented.[25]

Cerebral edema may be the most frequent cause of death resulting from HSV encephalitis. Treatment involves reducing the total fluid intake to between one half and two thirds of that required for maintenance and elevating the patient's head to 30°. Edema is treated by intubation and hyperventilation to bring the partial pressure of carbon dioxide to 25 mmol/L. This reduces increased intracranial pressure (ICP) by constricting the intracranial vasculature. After 24 hours, hyperventilation is less effective. Steroids and osmotic agents may also control brain swelling. The benefits of a short course of steroids in instances of cerebral edema and impending herniation probably outweigh the small risk of potentiating HSV infection. Dexamethasone is administered intravenously at an initial dose of 10 mg, then 4 to 8 mg every 4 to 5 hours for the next 3 days. Incipient herniation may be managed with osmotic diuretics. Mannitol is given in repeated doses of 0.25 to 2.0 g/kg. The serum osmolality should be monitored closely and kept below 310 mOsm/L. The effectiveness of mannitol decreases with repeated use, and rebound increases in ICP may occur.

Anticonvulsants are used to treat seizures. For initial seizures, diazepam may be administered intravenously in doses of 2 mg/min for a maximum total dose of 15 to 20 mg; alternatively, lorazepam may be administered intravenously in doses of 2 mg/min for a maximum total dose of 5 to 10 mg. These treatments are followed by fosphenytoin, which is administered in a loading dose of 20 mg/kg at 150 mg/min. Blood pressure and electrocardiographic monitoring should be employed during treatment with fosphenytoin. Maintenance doses of phenytoin average between 300 and 400 mg/day. Therapy should be guided in accordance with blood levels, side effects, and the presence or absence of seizures. Anticonvulsants should be continued for several months after the acute illness. There are no controlled studies of the prophylactic use of steroids or anticonvulsants in HSV encephalitis.

HSV Neuropathy (Zosteriform Eruption)

Pain and sensory loss in a dermatomal distribution may accompany a zosteriform eruption. The syndrome is usually characterized by a prodrome of diffuse neuralgia, often with malaise and fever, followed within a few days by vesicular eruption.[26] Most reports describe lesions on the face in one or more areas served by the trigeminal nerve, but zosteriform eruptions attributed to HSV have also been described on the trunk, extremities,

and genitalia. A long-standing report of recurrent sciatica associated with HSV is particularly interesting because it identifies a form of sciatica that can be treated with antiviral agents.[27] A first episode may be confused with herpes zoster, but recurrent episodes of dermatomal neuralgic pain and zosteriform eruptions are usually caused by HSV-2.[16]

Although HSV neuropathy is well documented as a clinical entity, more information is needed to know the exact type of HSV responsible for each disease. Clinicians attribute herpes lesions above the neck to the reactivation of HSV-1, whereas lesions below the waist are attributed to the reactivation of HSV-2. Future DNA analysis of herpes isolates by PCR with primers specific for HSV-1 or HSV-2 will identify whether herpes neuropathy is caused primarily by HSV-1 or HSV-2.

Most accounts of HSV neuropathy were made before the development of acyclovir or famciclovir. Because antiviral treatment will probably reduce the number of days that patients have pain and rash, treatment with famciclovir (500 mg three times daily) for 7 to 10 days is reasonable. Neither HSV-1 nor HSV-2 can be eradicated by antiviral therapy. Rather, resolution of acute disease is followed by a return of the herpesvirus to the latent state with the potential for future reactivation.

VARICELLA-ZOSTER VIRUS

VZV causes chickenpox (varicella) in childhood and becomes latent in the ganglia of the cranial nerves, the dorsal root, and the autonomic nervous system. Reactivation decades later produces shingles (herpes zoster), the most common cause of virus-induced neurologic disease in adults.

Herpes Zoster

Herpes zoster is characterized by pain and vesicular rash on an erythematous base in one to three dermatomes (localized zoster). Most herpes zoster cases reflect VZV reactivation from ganglia. In rare instances, cases of herpes zoster are clustered in time, supporting the notion that herpes zoster can be acquired by exogenous reinfection. More than a half-million persons in the United States develop herpes zoster each year. Fewer than 10% of cases of herpes zoster are recurrent. Although chickenpox occurs primarily in the spring, herpes zoster occurs at any time of year. There is no sex bias. Herpes zoster is eight to 10 times more common in persons older than 60 years, and approximately 50% of persons 80 years of age or older have had herpes zoster once. A continuing increase in the number of elderly persons will result in greater herpes zoster–associated morbidity and mortality. Herpes zoster and its attendant complications [*see Figure 2*] are also more common in immunocompromised persons, especially persons with AIDS. Before the AIDS era, herpes zoster occurred most often in bone marrow transplant recipients and patients with leukemia, particularly those who had received radiation therapy; herpes zoster developed at the irradiated site. In my experience, long-term, low-dose steroid therapy also predisposes patients to the serious complications of VZV reactivation, including postherpetic neuralgia (PHN), disseminated herpes zoster, and herpes zoster encephalitis.

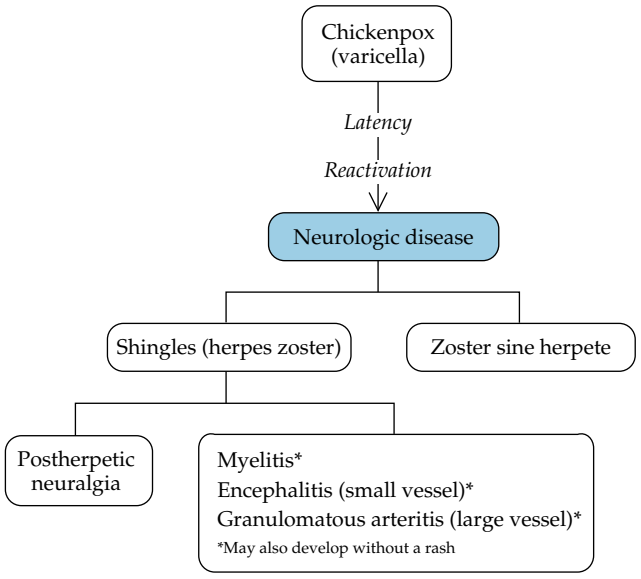


Figure 2 **The neurologic complications of varicella-zoster virus reactivation.**

Clinical features Herpes zoster pain is severe, deep, burning, and jabbing and usually lasts 4 to 6 weeks. Examination reveals hyperesthesia (allodynia) in a dermatomal distribution with mixed hypopigmentation and hyperpigmentation. Except for its shorter duration, acute herpes zoster pain does not differ from the pain of PHN. Herpes zoster may involve any dermatome. The trunk, represented by 12 pairs of thoracic ganglia, is the most frequently involved site. The face, which is supplied by the trigeminal nerve, is involved more often than any single thoracic dermatome, however.

Trigeminal distribution of herpes zoster most frequently occurs in the ophthalmic branch. Cranial neuropathies often occur weeks or months after cutaneous signs, perhaps because of microinfarction of cranial nerves from occlusion of the vasa vasorum. The blood supply of the cranial nerves is part of the carotid circulation. Because trigeminal afferent nerves innervate the large extracranial and intracranial blood vessels, virus can also spread along ganglionic afferent fibers to small vessels supplying cranial nerves, in a manner similar to that proposed to explain the pathogenesis of VZV-induced granulomatous arteritis. Peripheral facial weakness, an abnormality of the cranial nerves, is often associated with vesicles in the ear (herpes zoster oticus); the combination of these conditions constitutes the Ramsay Hunt syndrome.

Herpes zoster in cervical or lumbosacral dermatomes may be associated with muscle weakness. The incidence of herpes zoster paresis in patients with zoster varies from as low as 0.5% to as high as 31%. Bladder and bowel dysfunction are often associated with a sacral distribution of herpes zoster.

Treatment of acute herpes zoster Valacyclovir (1,000 mg three times daily for 7 days) and famciclovir (500 mg three times daily for 7 days) appear to be equally effective in clearing skin lesions and reducing pain. Both drugs are safe and well tolerated. Because it is easier for patients to take medication three times daily instead of five times daily, valacyclovir and famciclovir are preferred over acyclovir. In combination with antiviral therapy, steroids may reduce the use of analgesics and the time to uninterrupted sleep and to resumption of usual activities.[28]

Complications of herpes zoster infection PHN, defined as pain persisting longer than 4 to 6 weeks after the appearance of herpes zoster rash, is the most common complication of herpes

zoster. Age is the most important factor in predicting the development of PHN. The incidence of PHN is 46.9% in patients 60 years of age and older who have contracted acute herpes zoster[29] and varies from zero to 15% in patients younger than 60 years; most of these cases occur in patients between 50 and 60 years of age. PHN is slightly more common in women than in men, but in both sexes, it occurs more often after ophthalmic-distribution zoster.

Analgesics are used to reduce the pain of PHN; if possible, narcotics should be avoided. Gabapentin, beginning with 300 mg daily and gradually increased to a maximum of 3,600 mg daily, provides some relief. Starting with 25 mg at night, amitriptyline can be gradually increased to 150 mg daily. Side effects include postural hypotension, confusion, and urinary retention. Carbamazepine may help, but doses in excess of 600 mg daily are often required, and side effects of sleepiness, confusion, and unsteadiness, particularly in the elderly, may be intolerable. Carbamazepine should be started at doses of 100 mg daily and gradually increased by 100 mg every third day until pain decreases. Lidocaine skin patches are often helpful, as is topical application of triethanolamine salicylate (Aspercreme) or Flex-All 454 if the agent is applied alternatively throughout the day and before sleep. Transcutaneous electrical nerve stimulation units, nerve blocks, and rhizotomy do not usually provide long-term relief.

Studies linking PHN with VZV persistence provide a rationale for treatment with high-dose, perhaps intravenous, acyclovir for more than 10 days to reduce the virus burden.[30] Further studies are needed to determine whether antiviral agents, with or without steroids, initiated at the onset of herpes zoster prevents PHN and whether short-term, high-dose antiviral agents effectively treat PHN.

Zoster Sine Herpete

Prolonged radicular pain without rash may indicate zoster sine herpete. This nosologic entity was verified in two men, aged 62 and 66 years, who developed radicular pain that persisted for years. The pain did not differ from that of PHN. VZV DNA, but not HSV DNA, was found in the blood mononuclear cells of one man and in the CSF of both men 5 to 8 months after the onset of pain.[31] Treatment with intravenous acyclovir, 10 to 15 mg/kg every 8 hours for 14 days, followed by oral acyclovir, relieved pain in both patients. A remarkable case of zoster sine herpete characterized by prolonged pain in a trigeminal distribution, sensory loss, and tumor was shown to be caused by an active VZV ganglionitis.[32]

VZV Vasculopathy

CNS disease is the most alarming complication of herpes zoster. Two conditions predominate: multifocal (small-vessel) vasculopathy, which occurs primarily in immunocompromised patients; and unifocal (large-vessel) vasculopathy, which occurs in immunocompetent patients.

It is important to recognize that syndromes of large-vessel and small-vessel vasculopathy are not always distinct. Both types of syndromes may be involved in a given patient, and together they can produce waxing and waning neurologic symptoms and signs. Therefore, the designation VZV unifocal or multifocal vasculopathy rather than VZV encephalitis seems appropriate because the symptoms and signs seen clinically and on brain imaging indicate vasculopathy rather than encephalitis. Prompt recognition of a unifocal or multifocal vasculopathy after zoster or varicella should lead to improved diagnosis and rapid treatment.

VZV multifocal vasculopathy Multifocal vasculopathy is the most common form of CNS involvement caused by VZV. Disease develops in the setting of cancer, immunosuppression, or AIDS. Neurologic disease is subacute, and death is common. Herpes zoster small-vessel multifocal vasculopathy presents as headache, fever, vomiting, mental changes, seizures, and focal deficit. Brain imaging reveals large and small ischemic or hemorrhagic infarcts—often both—of cortex and subcortical gray and white matter [*see Figure 1d*]. Deep-seated white-matter lesions often predominate and are ischemic or demyelinative. The demyelinative lesions are smaller and less coalescent than those seen in progressive multifocal leukoencephalopathy. The CSF shows a mild to moderate pleocytosis (predominantly mononuclear), normal or mildly elevated concentrations of CSF protein, and a normal CSF glucose content—findings that do not differ significantly from those associated with herpes zoster without vasculopathy. Treatment of herpes zoster vasculopathy includes intravenous acyclovir at a dosage of 15 to 30 mg/kg/day for 10 days. Longer treatment may be necessary in severely immunocompromised patients.

VZV unifocal vasculopathy The salient feature of this form of VZV-induced CNS disease is an acute focal deficit that occurs weeks or months after an episode of herpes zoster in a contralateral trigeminal distribution. Stroke results from a necrotizing arteritis, primarily of large cerebral arteries. One comprehensive review showed that most patients with large-vessel vasculopathy were older than 60 years and that there was no sex bias.[33] The mean onset of neurologic disease was 7 weeks, and the longest interval between the onset of herpes zoster and the onset of neurologic disease was 6 months. Transient ischemic attacks and mental symptoms were common. Twenty-five percent of patients died. The majority of patients had CSF pleocytosis, usually fewer than 100/mm^3 (predominantly mononuclear), oligoclonal bands, and increased CSF IgG. Besides contralateral hemiplegia, ipsilateral central retinal artery occlusion and posterior circulation involvement have been described. Angiographic examination has revealed focal constriction and segmental narrowing, primarily in middle cerebral, internal carotid, and anterior cerebral arteries. Microscopic examination has revealed a necrotizing arteritis (primarily involving the intima and adventitia), inflammation with multinucleated giant cells, VZV antigen, Cowdry type A inclusions, and herpesvirus particles. Most herpes zoster–associated granulomatous angiitis infarcts are pale,[34] but hemorrhagic infarction also occurs. Afferent trigeminal ganglionic fibers to both intracranial and extracranial blood vessels provide an anatomic pathway for the spread of virus.

No definitive treatment is available for large-vessel herpes zoster vasculopathy. Nevertheless, because productive virus infection is found in arteries, patients should receive intravenous acyclovir (to kill persistent replicating virus) and steroids (for their anti-inflammatory effect).

The neurologic features of VZV vasculopathy are protean. Neurologic disease often occurs months after zoster and sometimes without any history of zoster rash. MRI scanning, cerebral angiography, and examination of CSF with virologic analysis are needed to confirm the diagnosis. VZV DNA is not always present in the CSF of patients with VZV vasculopathy, but the diagnosis can be confirmed by finding anti-VZV antibody in CSF in

conjunction with reduced serum-to-CSF ratios of VZV IgG compared with the normally high serum-to-CSF ratios of albumin or total IgG. When VZV vasculopathy develops months after zoster, antiviral treatment is often effective.

In addition to causing vasculopathy, VZV may cause disease at multiple levels of the neuraxis, involving the central and peripheral nervous systems in the absence of rash. Cases include encephalomyeloradiculoneuropathy; mixed small-vessel and large-vessel vasculopathy; acute, chronic, and recurrent neuropathy[35]; and myelopathy.[36] In some of these cases, patients have responded well to treatment with intravenous acyclovir, underscoring the value of aggressive testing for VZV in unusual cases of CNS disease or neuropathy.

Overall, VZV is a neurotropic virus that infects CNS blood vessels of various calibers, producing protean multifocal or unifocal neurologic disorders. Diagnosis requires recognition of the diversity of syndromes that can be caused by VZV, a high index of suspicion, and extensive testing for VZV DNA and antibody in CSF. Finally, nearly all CNS disease caused by VZV infection entails stroke, and infarction may follow zoster or varicella. Disease is not a primary encephalitis but is instead the result of unifocal or multifocal infarction that develops secondary to productive virus infection within large and small cerebral arteries. Rarely, VZV is associated with disseminated encephalomyelitis, a CNS complication of chickenpox.[37]

EPSTEIN-BARR VIRUS

EBV causes infectious mononucleosis and has been implicated as a cause of nasopharyngeal carcinoma and Burkitt lymphoma. EBV is ubiquitous. Primary infection of most humans usually occurs in childhood or adolescence. By 30 years of age, more than 90% of adults have antibody to EBV. Despite the prevalence of EBV infection, neurologic disease is rare. The most common presentation is meningoencephalitis, which is often associated with acute cerebellar ataxia. More serious meningoencephalopathy, presenting as athetosis and chorea or as stupor and coma, has been described. In a fascinating case of acute demyelinating encephalopathy after EBV infection, a patient presented with behavioral abnormalities, visual illusions, and a seizure.[38] EBV mononucleosis may also be followed by recurrent aseptic meningitis. Chronic active EBV infection has been associated with calcification in the basal ganglia.[39] EBV-associated cranial neuropathies, including acute autonomic neuropathy, have been described.[40,41] In a comprehensive study of four patients with EBV myeloradiculitis and encephalomyeloradiculitis, none of the patients died or had brain swelling, but residual neurologic deficits were evident.[42] A case of myeloradiculitis in an immunocompetent 72-year-old woman who was seronegative for EBV was proved to be caused by EBV by the detection of EBV DNA in both serum and CSF; quantitative PCR revealed an EBV DNA load not detected in healthy subjects who had latent infection or were not infected.[43] Detection of EBV DNA in CSF and reductions in the ratio of EBV antibody in serum to EBV antibody in CSF have been used to diagnose infection in the peripheral nervous system and CNS.[41,42,44] There is no specific treatment for EBV infection.

CYTOMEGALOVIRUS

CMV produces neurologic disease, primarily in infants, as part of the clinical spectrum of congenital CMV infection. Although most congenital CMV infections are asymptomatic, many carriers develop sensorineural hearing loss and intellectual handicaps. Other neurologic complications of congenital infections include microcephaly, seizures, hypotonia, and spasticity. Lethargy and coma may follow a severe case of CMV meningoencephalitis.

In immunocompetent adults, the most common neurologic complication of CMV infection is the Guillain-Barré syndrome. The association was made on the basis of the following findings: the isolation of CMV from buffy coat of blood, the development of CMV IgM antibodies in serum,[45] and the demonstration of a fourfold or greater alteration in complement-fixing antibody to CMV in 21 of 92 patients with acute polyneuritis.[46] CMV infection has also been associated with acute brachial plexopathy.[47]

Until the AIDS era, CMV infection of the CNS was rare. There were a few cases of CMV encephalitis in immunocompetent adults, characterized by mental status changes and pyramidal signs. Because the natural history of CMV encephalitis in immunocompetent persons is not known, no treatment protocols have been established. However, two patients with presumptive CMV encephalitis, verified by the isolation of CMV from CSF and urine, were treated with vidarabine and responded well.[48] In contrast to the rare occurrences of CMV infection in the immunocompetent adult, CMV infections in the immunocompromised population, particularly AIDS patients, are common. CMV has emerged as an important cause of encephalitis, myelitis, and polyradiculitis. CMV encephalitis was first documented in renal transplant[49] and bone marrow transplant recipients; an interesting case of CMV vasculitis and retinitis in a lymphoma patient receiving multiple cytotoxic and anti-inflammatory drugs has been reported. In AIDS patients, CMV is the most common opportunistic infectious agent found in the nervous system. Complications include retinitis, encephalitis,[50] progressive myelitis, and polyradiculopathy. CMV encephalitis is usually subacute and is characterized by headache, seizures, progressive dementia, and diffuse weakness. Thus, neurologic deficit may be diffuse, focal, or both.

Like other herpesviruses, CMV can establish a latent infection in persons with intact immune systems. The lymphoid system is the most likely site for viral latency. The source of CMV encephalitis may be circulating lymphocytes. Perhaps because of the ill-defined clinical picture, the diagnosis of CMV encephalitis has been made at autopsy in most cases. The CSF may show a neutrophilic or mononuclear pleocytosis, elevated protein concentration, and depressed glucose content, but these findings are hardly specific, particularly in immunosuppressed patients. MRI may show enhancement in the ventricular ependyma, a finding that is highly suggestive of CMV ventriculitis or encephalitis. Focal disease has been attributed to CMV vasculitis or foci of demyelination. Characteristic owl-eyed cytomegalic inclusions and CMV-specific antigens have been found in the CNS and blood vessels of AIDS patients with subacute encephalopathy. Complicating the picture, however, is the additional presence of HIV or HSV-2. Thus, the attribution of specific symptoms and signs to CMV is difficult. The organism can occasionally be isolated from the CSF of AIDS patients.

Of particular interest is the occurrence in AIDS patients of a CMV polyradiculopathy. Disease often begins insidiously as a cauda equina syndrome with paresthesias and distal weakness (which is usually asymmetrical), incontinence, and sacral-distribution sensory loss. CSF pleocytosis is usually present; either polymorphonuclear leukocytes or mononuclear cells may predominate; CSF protein concentration may be elevated; hypoglycorrhachia may be present; and, occasionally, CMV is isolated

from CSF. Postmortem examination reveals inflammation, necrosis, and focal vasculitis of nerve roots, with typical CMV intranuclear and intracytoplasmic inclusions.[51]

Ganciclovir and foscarnet demonstrate high activity against CMV, although formal trials have not been performed to establish dosage or treatment length. Measurement of viral load in the CSF by PCR is a potential way to assess therapeutic efficacy. CMV infections of the nervous system are currently best treated with intravenous ganciclovir, 5 mg/kg every 12 hours for 2 weeks, followed by maintenance therapy of 5 mg/kg of ganciclovir daily. Patients may need maintenance therapy for life to prevent relapse. Oral ganciclovir has been approved by the Food and Drug Administration for maintenance therapy in AIDS patients. The recommended maintenance dosage is 1,000 mg three times daily with food or 500 mg six times daily (every 3 hours) with food. The main side effect of ganciclovir is neutropenia. The drug should be stopped if the total white blood cell count falls below 750/mm^3.

HUMAN HERPESVIRUS TYPE 6

Human herpesvirus type 6 (HHV-6) causes encephalitis primarily in bone marrow and stem cell transplant recipients[52]; it may present as limbic encephalitis.[53] Diagnosis is confirmed by PCR amplification of HHV-6 DNA in CSF. Not enough cases have been reported to know whether antiviral therapy is effective.

PAPOVAVIRUS BK

The papovavirus BK virus is a relative of JC virus, the cause of progressive multifocal leukoencephalopathy. Like JC virus, BK virus is most commonly found in immunosuppressed patients.[54] On the basis of the development of IgM antibody and PCR results that are positive for viral DNA in CSF, papovavirus BK has also been shown to cause encephalitis in an immunocompetent patient.[55]

HIV

HIV infects the meninges early and persists in the CNS after primary infection. Aseptic meningitis affects 5% to 10% of HIV-infected patients just before seroconversion or after a mononucleosis-like syndrome that is characterized by headache, fever, altered mental status, and focal or generalized seizures. HIV infection is verified by the presence of p24 antigen in serum or the CSF; proof of infection is aided by evidence of seroconversion 1 or 2 months later. Acute meningoencephalitis is usually self-limited, although slowly progressive HIV encephalitis with mental status changes and motor deficit may follow [*see 7:XXXIII HIV and AIDS*].

ARTHROPOD-BORNE ENCEPHALITIS VIRUSES

Arthropod-borne viruses include the virus families Togaviridae (alphaviruses and flaviviruses), Bunyaviridae, and Orbiviridae—all RNA viruses. Humans are infected through mosquito or tick bites. Mosquitoes are infected by biting infected horses or birds. Human infection occurs in the summer and early fall. Human epidemics often follow epidemics in horses, the migration of infected birds, and the infection of rodents. Typical symptoms and signs of infection include headache, photophobia, myalgia, fever, lethargy, confusion, and seizures. Tremor and focal deficit may develop. The CSF pattern is the same as in aseptic meningitis. A specific diagnosis can be verified by the following means: virus isolation from CSF, the production of encephalitis in suckling mice inoculated with infected CSF, the development of a cytopathic effect in tissue culture cells inoculated with infected CSF, the detection of virus-specific antibody in CSF, the presence of virus-specific IgM during acute disease, or a rising titer of virus-specific IgG in sera during the acute and convalescent stages of the disease.

In the United States, the most common vector-borne viral (arbovirus) infection is St. Louis encephalitis (SLE), followed by Western equine encephalitis (WEE), California virus encephalitis, and Eastern equine encephalitis (EEE) [*see 7:XXXI Viral Zoonoses*]. SLE is transmitted mostly by mosquitoes. Nervous system infections usually produce encephalitis, although aseptic meningitis occurs in about 15% of cases. Prominent opsoclonus and tremulousness[56] and brain stem involvement have been described. Nearly all SLE cases are self-limited, although approximately 3% to 4% of cases are fatal.

In 1999, an outbreak of encephalitis accompanied by severe weakness and axonal neuropathy occurred in New York. Initial serologic analysis attributed disease to SLE virus in five of six patients. More extensive molecular analysis identified the virus responsible as a Kunjin/West Nile–like flavivirus, part of the Japanese encephalitis virus serogroup.[57] West Nile virus has now spread to the central and western United States. Although most human cases of West Nile virus infection are acquired through bites from infected mosquitoes, infection may also be transmitted by infected blood products or transplanted organs.[58]

The neurologic complications of West Nile virus infection are protean. Patients have presented not only with meningoencephalitis but also with rhombencephalitis,[59] opsoclonus-myoclonus cerebellar ataxia,[60] unilateral brachial plexopathy with meningoencephalitis,[61] and Guillain-Barré syndrome.[62] As with other arbovirus infections, deep-seated lesions in the basal ganglia and thalamus are often seen on brain MRI.[63] PCR examination of the CSF often reveals West Nile virus RNA.[64]

The bunyaviruses are transmitted by mosquitoes and cause California encephalitis. La Crosse virus is the most common and causes encephalitis primarily in school-age children; hyponatremia is common.[65] Although encephalitis is usually mild, 50% of victims have seizures, and focal deficits and focal EEG abnormalities are seen in 20% to 40% of patients. EEE is more serious; more than 50% of patients die, and mental, visual, auditory, speech, and motor deficits remain in more than 80% of survivors. MRI often reveals focal increases in signal intensity in the basal ganglia and thalamus[66] that distinguish EEE from HSV encephalitis. WEE has a case-fatality rate of approximately 10%; it is more severe in infants and children and may produce transient parkinsonism.[67] Powassan encephalitis is caused by a flavivirus transported by ticks. Almost total ophthalmoplegia in a virologically verified case of Powassan encephalitis has been reported.[68] Orbivirus infection causes Colorado tick fever. There is no specific treatment for arbovirus encephalitis.

ADENOVIRUSES

Adenoviruses are large DNA viruses that usually cause outbreaks of acute respiratory illness, keratoconjunctivitis, and, occasionally, pneumonia. Different adenovirus serotypes produce meningoencephalitis. Although the incidence of CNS disease after adenovirus infection is unknown, the mortality after CNS infection ranges from 26% to 38%.[69] A fascinating case of subacute focal encephalitis in an immunocompromised man from whom type 32 adenovirus was isolated from the brain has been described.[70]

ARENAVIRUSES

The arenaviruses are RNA viruses endemic in rodents. Human infection, which occurs mostly in South America and Africa, produces various hemorrhagic fevers that are often fatal [*see 7:XXXI Viral Zoonoses*]. In the United States, the most common arenavirus infection is produced by LCM virus. Infection is acquired from mice or pet hamsters or through occupational exposure in laboratory personnel working with the virus. Most cases present as aseptic meningitis, but fatal meningoencephalitis also occurs. LCM preceded by a lupuslike syndrome that included rash and the presence of a circulating anticoagulant has been described.[71] In 2005, a cluster of fatal cases of meningoencephalitis in solid organ transplant recipients was associated with LCM infection.[72] As with mumps, orchitis and parotitis may develop concurrently with CNS disease[73]; profound mononuclear pleocytosis and hypoglycorrhachia are also produced. A specific diagnosis can be made by observing the development of choriomeningitis in weanling or adult mice inoculated intracerebrally with infected CSF or the development of LCM-specific antibody in the serum or CSF of infected humans.

HEPATITIS C VIRUS

Acute disseminated encephalomyelitis has been reported in a patient who had received multiple blood transfusions.[74] T_2-weighted MRI revealed symmetrical multifocal changes in the gray matter and white matter of the brain, with some gadolinium enhancement. Hepatitis C virus (HCV) infection was verified by a high titer of IgM antibody to HCV and a strongly positive reaction for HCV RNA. Treatment with high-dose dexamethasone led to a dramatic improvement.

GLOBAL MOBILITY AND DISEASE SPREAD

Although some new viruses that cause encephalitis have been found only in the Far East or South America, it is worth emphasizing that the globe can be circled in 24 hours. Thus, every viral infection in the world could travel anywhere in the world within one incubation period.[75] Over 500 million people cross international boundaries on commercial aircraft every year, about 70 million work legally or illegally in other countries, and 50 million are refugees or displaced persons.[76] This global mobility represents an enormous and steadily increasing opportunity to transport infectious agents or their vectors.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. McKinney RE Jr, Katz SL, Wilfert CM: Chronic enteroviral meningoencephalitis in agammaglobulinemic patients. Rev Infect Dis 9:334, 1987

2. Malcom BS, Eiden JJ, Hendley JO: ECHO virus type 9 meningitis simulating tuberculous meningitis. Pediatrics 65:725, 1980

3. Rotbart HA, Levin MJ, Villarreal LP: Use of subgenomic poliovirus DNA hybridization probes to detect the major subgroups of enteroviruses. J Clin Microbiol 20:1105, 1984

4. Kuban KC, Ephros MA, Freeman FL, et al: Syndrome of opsoclonus-myoclonus caused by coxsackie B3 infection. Ann Neurol 13:69, 1983

5. Huang C-C, Liu C-C, Chang Y-C, et al: Neurologic complications in children with enterovirus 71 infection. N Engl J Med 341:936, 1999

6. da Silva EE, Winkler MT, Pallansch MA: Role of enterovirus 71 in acute flaccid paralysis after the eradication of poliovirus in Brazil. Emerg Infect Dis 2:231, 1996

7. Nolan MA, Craig ME, Lahra MM, et al: Survival after pulmonary edema due to enterovirus 71 encephalitis. Neurology 60:1651, 2003

8. Geller TJ, Condie D: A case of protracted coxsackie virus meningoencephalitis in a marginally immunodeficient child treated successfully with intravenous immunoglobulin. J Neurol Sci 129:131, 1995

9. Rosenberg RN, Chadwick DL: Oral-facial dyskinesia and quadriplegia associated with poliomyelitis virus type 3. Am J Dis Child 126:699, 1973

10. Poliomyelitis prevention in the United States: updated recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Morb Mortal Wkly Rep 49(RR-5):1, 2000

11. Azimi PH, Shaban S, Hilty MD, et al: Mumps meningoencephalitis: prolonged abnormality of cerebrospinal fluid. JAMA 234:1161, 1975

12. Vaheri A, Julkunen I, Koskiniemi M-L: Chronic encephalomyelitis with specific increase in intrathecal mumps antibody. Lancet 2:685, 1982

13. Fujimoto S, Kobayashi M, Uemura O, et al: PCR on cerebrospinal fluid to show influenza-associated acute encephalopathy or encephalitis. Lancet 352:873, 1998

14. Morishima T, Togashi T, Yokota S, et al: Encephalitis and encephalopathy associated with an influenza epidemic in Japan. Clin Infect Dis 35:512, 2002

15. Chua KB, Bellini WJ, Rota PA, et al: Nipah virus: a recently emergent deadly paramyxovirus. Science 288:1432, 2000

16. Gonzales N, Tyler KL, Gilden DH: Recurrent dermatomal vesicular skin lesions: a clue to diagnosis of type 2 herpes simplex virus meningitis. (submitted for publication)

17. Tedder DG, Ashley R, Tyler KL, et al: Herpes simplex virus infection as a cause of benign recurrent lymphocytic meningitis. Ann Intern Med 121:334, 1994

18. Yamamoto LJ, Tedder DG, Ashley R, et al: Herpes simplex virus type 1 DNA in cerebrospinal fluid of a patient with Mollaret's meningitis. N Engl J Med 325:1082, 1991

19. Lakeman FD, Whitley RJ: Diagnosis of herpes simplex encephalitis: application of polymerase chain reaction to cerebrospinal fluid from brain-biopsied patients and correlation with disease. National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group. J Infect Dis 171:857, 1995

20. Weil AA, Glaser CA, Amad Z, et al: Patients with suspected herpes simplex encephalitis: rethinking an initial negative polymerase chain reaction result. Clin Infect Dis 34:1154, 2002

21. Smith JB, Westmoreland BF, Reagan TJ, et al: A distinctive clinical EEG profile in herpes simplex encephalitis. Mayo Clin Proc 50:469, 1975

22. Zimmerman RD, Russell EJ, Leeds NE: CT in the early diagnosis of herpes simplex encephalitis. Am J Radiol 134:61, 1980

23. Schroth G, Gawehn J, Thron A, et al: Early diagnosis of herpes simplex encephalitis by MRI. Neurology 37:179, 1987

24. Whitley RJ, Alford CA, Hirsch MS, et al: Vidarabine versus acyclovir therapy in herpes simplex encephalitis. N Engl J Med 314:144, 1986

25. Knezevic W, Carrol WM: Relapse of herpes simplex encephalitis after acyclovir therapy. Aust N Z J Med 13:625, 1983

26. Krohel GB, Richardson JR, Farrell DF: Herpes simplex neuropathy. Neurology 26:596, 1976

27. Morris HH III, Peters BH: Recurrent sciatica associated with herpes simplex: case report. J Neurosurg 41:97, 1974

28. Whitley RJ, Weiss H, Gnann JW, et al: Acyclovir with and without prednisone for the treatment of herpes zoster: a randomized, placebo-controlled trial. Ann Intern Med 125:376, 1996

29. Rogers RS, Tindall JP: Geriatric herpes zoster. J Am Geriatr Soc 19:495, 1971

30. Mahalingam R, Wellish M, Brucklier J, et al: Persistence of varicella-zoster virus DNA in elderly patients with postherpetic neuralgia. J Neurovirol 1:130, 1995

31. Gilden DH, Wright RR, Schneck SA, et al: Zoster sine herpete, a clinical variant. Ann Neurol 35:530, 1994

32. Hevner R, Vilela M, Rostomily R, et al: An unusual cause of trigeminal-distribution pain and tumour. Lancet Neurol 2:567 2003

33. Hilt DC, Buchholz D, Krumholz A, et al: Herpes zoster ophthalmicus and delayed contralateral hemiparesis caused by cerebral angiitis: diagnosis and management approaches. Ann Neurol 14:543, 1983

34. Kuroiwa Y, Furukawa T: Hemispheric infarction after herpes zoster ophthalmicus: computed tomography and angiography. Neurology 31:1030, 1981

35. Fox RJ, Galetta SL, Mahalingam R, et al: Acute, chronic, and recurrent varicella zoster virus neuropathy without zoster rash. Neurology 57:351, 2001

36. Gilden DH, Beinlich BR, Rubinstein EM, et al: Varicella zoster virus myelitis: an expanding spectrum. Neurology 44:1818, 1994

37. Gilden DH: Varicella zoster virus vasculopathy and disseminated encephalomyelitis. J Neurol Sci 195:99, 2002

38. Paskavitz JF, Anderson CA, Filley CM, et al: Acute arcuate fiber demyelinating encephalopathy following Epstein-Barr virus infection. Ann Neurol 38:127, 1995

39. Morita M, Tsuge I, Matsuoka H, et al: Calcification in the basal ganglia with chronic active Epstein-Barr virus infection. Neurology 50:1485, 1998

40. Salazar A, Martinez H, Sotelo J: Ophthalmoplegic polyneuropathy associated with infectious mononucleosis (letter). Ann Neurol 13:219, 1983

41. Bennett JL, Mahalingam R, Wellish MC, et al: Epstein-Barr virus–associated acute autonomic neuropathy. Ann Neurol 40:453, 1996

42. Majid A, Galetta SL, Sweeney CJ, et al: Epstein-Barr virus myeloradiculitis and encephalomyeloradiculitis. Brain 125:159, 2002

43. Muhlau M, Bulow S, Stimmer H, et al: Seronegative Epstein-Barr virus myeloradiculitis in an immunocompetent 72-year-old woman. Neurology 65:1329, 2005

44. Tselis A, Duman R, Storch GA, et al: Epstein-Barr virus encephalomyelitis diagnosed by polymerase chain reaction: detection of the genome in the CSF. Neurology 48:1351, 1997

45. Kabins S, Keller R, Peitchel RT, et al: Acute idiopathic polyneuritis caused by cytomegalovirus. Arch Intern Med 136:100, 1976

46. Dowling P, Menonna J, Cook S: Cytomegalovirus complement fixation antibody in

Guillain-Barré syndrome. Neurology 27:1153, 1977

47. Duchowny M, Caplan L, Siber G: Cytomegalovirus infection of the adult nervous system. Ann Neurol 5:458, 1979

48. Phillips CA, Fanning WL, Gump DW, et al: Cytomegalovirus encephalitis in imunologically normal adults: successful treatment with vidarabine. JAMA 238:2299, 1977

49. Schneck SA: Neuropathological features of human organ transplantation: I. Probable cytomegalovirus infection. J Neuropathol Exp Neurol 24:415, 1965

50. Laskin OL, Stahl-Bayliss CM, Morgello S: Concomitant herpes simplex virus type 1 and cytomegalovirus ventriculoencephalitis in acquired immunodeficiency syndrome. Arch Neurol 44:843, 1987

51. Eidelberg D, Sotrel A, Vogel H, et al: Progressive polyradiculopathy in acquired immune deficiency syndrome. Neurology 36:912, 1986

52. Zerr DM, Gooley TA, Yeung L, et al: Human herpesvirus 6 reactivation and encephalitis in allogeneic bone marrow transplant recipients. Clin Infect Dis 33:763, 2001

53. Wainwright MS, Martin PL, Morse RP, et al: Human herpesvirus 6 limbic encephalitis after stem cell transplantation. Ann Neurol 50:612, 2001

54. Reploeg MD, Storch GA, Clifford DB: BK virus: a clinical review. Clin Infect Dis 33:191, 2001

55. Voltz R, Jäger G, Seelos K, et al: BK virus encephalitis in an immunocompetent patient. Arch Neurol 53:101, 1996

56. Estrin WJ: The serological diagnosis of St Louis encephalitis in a patient with the syndrome of opsoclonia, body tremulousness, and benign encephalitis. Ann Neurol 1:596, 1977

57. Briese T, Xi-Yu J, Huang C, et al: Identification of a Kunjin/West Nile–like flavivirus in brains of patients with New York encephalitis. Lancet 354:1261, 1999

58. Hong DS, Jacobson KL, Raad II, et al: West Nile encephalitis in 2 hematopoietic stem cell transplant recipients: case series and literature review. Clin Infect Dis 37:1044, 2003

59. Nichter CA, Pavlakis SG, Shaikh U, et al: Rhombencephalitis caused by West Nile fever virus. Neurology 55:153, 2000

60. Khosla JS, Edelman MJ, Kennedy N, et al: West Nile virus presenting as opsoclonus-myoclonus cerebellar ataxia. Neurology 64:1095, 2005

61. Almhanna K, Palanichamy N, Sharma M, et al: Unilateral brachial plexopathy associated with West Nile virus meningoencephalitis. Clin Infect Dis 36:1629, 2003

62. Ahmed S, Libman R, Wesson K, et al: Guillain-Barré syndrome: an unusual presentation of West Nile virus infection. Neurology 55:144, 2000

63. Rosas H, Wippold FJ 2nd: West Nile virus: case report with MR imaging findings. AJNR Am J Neuroradiol 24:1376, 2003

64. Nash D, Mostashari F, Fine A, et al: The outbreak of West Nile virus infection in the New York City area in 1999. N Engl J Med 344:1807, 2001

65. McJunkin JE, de los Reyes EC, Irazuzta JE, et al: La Crosse encephalitis in children. N Engl J Med 344:801, 2001

66. Deresiewicz RL, Thaler SJ, Hsu L, et al: Clinical and neuroradiographic manifestations of eastern equine encephalitis. N Engl J Med 336:1867, 1997

67. Schultz DR, Barthal JS, Garrett C: Western equine encephalitis with rapid onset of parkinsonism. Neurology 27:1095, 1977

68. Lessell S, Collins TE: Ophthalmoplegia in Powassan encephalitis. Neurology 60:1726, 2003

69. Similä S, Jouppila R, Salmi A, et al: Encephalomeningitis in children associated with an adenovirus type 7 epidemic. Acta Paediatr Scand 59:310, 1970

70. Chou SM, Roos R, Burrell R, et al: Subacute focal adenovirus encephalitis. J Neuropathol Exp Neurol 32:34, 1973

71. Schanen A, Gallou G, Hincky J-M, et al: A rash, circulating anticoagulant, then meningitis. Lancet 351:1956, 1998

72. Lymphocytic choriomeningitis virus infection in organ transplant recipients—Massachusetts, Rhode Island, 2005. MMWR Morb Mortal Wkly Rep 54:537, 2005

73. Lewis JM, Utz JP: Orchitis, parotitis and meningoencephalitis due to lymphocytic-choriomeningitis virus. N Engl J Med 265:776, 1961

74. Sacconi S, Salviati L, Merelli E: Acute disseminated encephalomyelitis associated with hepatitis C virus infection. Arch Neurol 58:1679, 2001

75. Johnson RT: Emerging viral infections of the nervous system. J Neurovirol 9:140, 2003

76. Wilson ME: Infectious diseases: an ecological perspective. BMJ 311:1681, 1995

Acknowledgment

Figure 1a Courtesy of Dr. Elliott Sandberg, Department of Radiology, the University of Colorado School of Medicine, Denver.

XVII CENTRAL NERVOUS SYSTEM DISEASES DUE TO SLOW VIRUSES AND PRIONS

Francisco González-Scarano, m.d.

There are a number of central nervous system diseases whose common elements include a long incubation period and a progressive clinical course leading to severe neurologic dysfunction or death. The term slow virus infection was at one time used to describe all these diseases, and most of them are in fact caused by viruses. One group, however, is now believed to be caused by abnormally configured proteins known as prions. This chapter reviews the more common of these diseases: HIV-associated dementia (HAD or HIVD), human T cell lymphotropic virus type I (HTLV-I)–associated myelopathy, Creutzfeldt-Jakob disease (CJD), progressive multifocal leukoencephalopathy (PML), and subacute sclerosing encephalitis (SSPE), which is associated with a variant of measles virus.

HIV-Associated Dementia

The primary neurologic syndromes associated with HIV infection are among the more complex and enigmatic complications of HAD. HIV can involve both the peripheral and the central nervous systems. Dementia is the most common of a number of primary neurologic consequences of HIV infection of the CNS; motor dysfunction (e.g., rigidity, spasticity, and ataxia) is also frequently encountered. HAD (also known as AIDS dementia complex [ADC]) is associated with cognitive and motor dysfunction and, in untreated patients, it follows a progressive course that can lead to a nearly vegetative state.[1] HIV has also been associated with spinal cord degeneration (vacuolar myelopathy), which results in the severe motor and sensory symptomatology characteristic of spinal cord involvement. HAD is more common than vacuolar myelopathy and was clinically diagnosed in up to 30% of HIV-infected patients before the development of highly active antiretroviral therapy (HAART). Aggressive treatment of the systemic disease has decreased the incidence of this HIV complication, but the prevalence may be rising as patients with HIV infection and AIDS live longer. In addition, a more subtle form of this complication—namely, minor cognitive and motor disorder (MCMD), which has a similar neuropathology—has been identified in patients whose systemic disease is well controlled.[2] Autopsy studies have detected neuropathologic changes in a greater percentage of patients, but frequently, there is discordance between the clinical syndrome and the pathologic findings (see below).

The peripheral nervous system is also a target for HIV, and distal sensory polyneuropathy (DSPN) is now more common than HAD. Some antiretroviral drugs, chiefly the dideoxynucleosides, are also associated with polyneuropathy; the correct diagnosis of the underlying cause of the polyneuropathy is often dependent on the patient's response to medication withdrawal. There is no generally used test that definitively distinguishes between DSPN and drug-induced neuropathy.[3,4]

PATHOGENESIS

The enigmatic pathogenesis of HAD [*see Figure 1*] may represent a new paradigm for the development of neurologic problems from a viral infection. First, there is discordance between the extensive clinical findings and the neuropathologic substrate. Second, the most prominently infected cells are brain macrophages and microglia rather than neurons or neuroglia.[5] Third, neurons appear to undergo apoptotic cell death in spite of the absence of viral infection.

Most investigators believe that the abnormalities associated with HAD result from the secretion of neurotoxins from chronically infected microglia.[3,5] Among the putative toxins are proinflammatory cytokines such as tumor necrosis factor–α (TNF-α) and platelet-activating factor.[6] Some investigators have also proposed that certain viral proteins, such as gp120 (one of the external or envelope proteins), tat (a viral transactivator), and vpr, can directly injure neurons when secreted by infected cells.[7,8] All these substances have been shown to mediate deleterious effects in cultured neurons of several mammalian species, but there is no conclusive proof that any of them are present in the HIV-infected CNS in levels sufficiently high to lead to comparable damage.

Alternatively, a low-level infection of neuroglia and neurons could mediate HAD. Conventional techniques (e.g., immunohistochemistry and in situ hybridization) have yielded mostly negative results regarding infection of oligodendrocytes and neurons. Astrocytes could be responsible for some of the viral burden, but infection of these cells is generally restricted (i.e., it results in little production of new virus).[9,10]

An interesting finding is that chemokine receptors, which are among the components of the HIV cellular receptor complex, are present in many cells besides microglia (where they would be expected to be found because these cells can be infected with HIV). Chemokine receptors may provide a bridge through which the virus can interact with neurons and astrocytes, although this function has to be considered speculative at this point.

DIAGNOSIS

Clinical Features

Neurologic problems—including cranial neuropathies, meningoencephalitis, and, rarely, coma—develop in some patients who are clinically ill with acute HIV infection [*see 7:XXXIII HIV and AIDS*]. Although the primary HIV syndrome is self-limited, seeding of the CNS during that period of high viremia may result in chronic asymptomatic mononuclear meningitis in some patients.

HAD generally occurs in the later stages of HIV infection, when low levels of $CD4^+$ T cells and a compromised immune system have led to opportunistic infections involving other organ systems. Initially, HAD may present as mild cognitive impairment that is detectable only by the administration of a battery of neuropsychological tests by psychologists who have experience with this particular condition.[11,12] Such mild cognitive

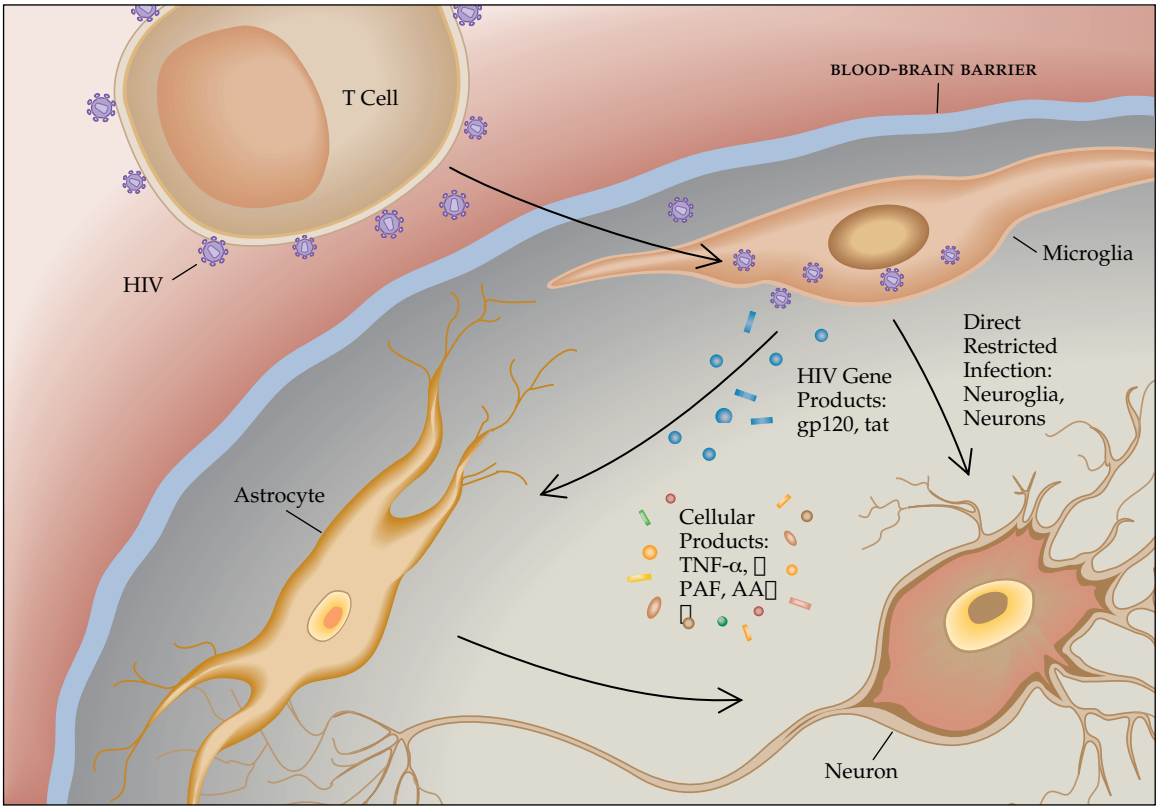


Figure 1 **The enigmatic pathogenesis of HIV-associated dementia (HAD). HIV enters the nervous system early in infection, carried in the plasma, monocytes, or T cells from the peripheral blood. HIV can then infect perivascular macrophages and perhaps also microglia, both of which express the primary HIV receptor, the $CD4^+$ molecule. The infected macrophages may cause disease by releasing cytokines (e.g., tumor necrosis factor–α [TNF-α], platelet-activating factor [PAF]), other inflammatory mediators (e.g., arachidonic acid [AA]), and viral proteins (e.g., gp120, tat), some of which may be toxic to neural cells (astrocytes, oligodendrocytes, neurons). Infected and uninfected macrophages and microglia fuse together to form multinucleated giant cells, the hallmark neuropathology of HAD [*see Figure 2*].**

deterioration may not interfere with job performance or activities of daily living.

HAD is a subcortical dementia, and psychomotor slowing, apathy, and such motor symptoms as ataxia and paralysis can precede memory loss and the deterioration of language function. In its end stages [*see Table 1*], persons with HAD may be nearly vegetative. Antiretroviral therapy can arrest or, in some instances, reverse the symptoms of HAD.

MCMD has been recognized more recently and is particularly worrisome because it may occur in patients who have been adequately treated with antiretroviral therapy. Although its pathologic substrate (see below) is similar to that of HAD, the symptoms of MCMD are more subtle and evidently less progressive than those of HAD.

The diagnosis of HAD depends on the exclusion of other causes of dementia in an HIV-infected patient. Cerebrospinal fluid abnormalities (i.e., elevated protein and decreased cell count) are characteristic of HIV infection but are not specific for HAD. The presence of HIV in the CSF is also not diagnostic of HAD, because both cell-free and cell-associated virus can be cultured from the CSF of persons who do not manifest CNS symptoms. However, there may be a rough correlation between the level of viral RNA in the CSF and the incidence of HAD. Magnetic resonance imaging studies have indicated that HAD is associated with global atrophy of the brain, but a high degree of brain atrophy is present in many persons with AIDS. Furthermore, metabolic abnormalities detectable with magnetic resonance spectroscopy point to neuronal destruction, but these cannot be used diagnostically.

Pathologic Features

The major neuropathologic findings associated with HAD are (1) multinucleated giant cell encephalitis (MGCE) [*see Figure 2*], (2) HIV leukoencephalopathy (myelin pallor and other white-matter changes), and (3) astrocytosis and perhaps neuronal dropout. MGCE, which is the most specific neuropathologic finding in HAD, is present in approximately 25% of persons who have been diagnosed clinically and is characterized by macrophage/microglial syncytia that may contain HIV-specific nucleic acid and antigens. These neuropathologic findings of encephalitis are characteristically more prominent in the perivascular regions, the basal ganglia, and other subcortical areas. Myelin pallor—which, as the name implies, is defined by decreased uptake of histochemicals by the CNS white matter—is present in 33% of patients with HAD; however, it is a less specific finding than MGCE and may be caused by alterations of the blood-brain barrier rather than by oligodendrocyte pathology.

The most striking feature of the neuropathology of HAD is the discordance between the histologic findings and the clinical disease. For example, in one study, neither MGCE nor diffuse pallor was present in 50% of the patients, and even in cases with morphologic changes, the abnormalities were frequently rela-

tively unimpressive.[13] This implies that the symptoms associated with HAD, although clearly associated with viral infection of macrophages and microglia, are ultimately the result of a process in which a mild infection of one cell type is amplified by unknown mechanisms.

TREATMENT

There is increasing evidence that successful treatment of HIV infection brings improvement in HAD,[2] although some studies suggest that HAART has less of an effect on HAD than it has on other conditions associated with AIDS.[12] Recent findings have suggested that antiretroviral regimens that include CSF-penetrating drugs are more effective at inducing clinical improvements in patients with MCMD or HAD.[14] These findings imply that the CNS viral load is directly responsible for cognitive dysfunction in patients with AIDS. The CNS may also be a reservoir for virus that has not been eliminated by HAART.

HTLV-I Myelopathy/Tropical Spastic Paraparesis

The entity known as HTLV-I–associated myelopathy (HAM), which was initially described in Japanese patients, is the same disease as tropical spastic paraparesis (TSP), which occurs in some Caribbean islands. Epidemiologic studies have firmly established the association between HTLV-I infection and the development of this slowly progressive neurologic syndrome. However, HAM/TSP is a relatively rare complication of HTLV-I, with an incidence roughly comparable to that of adult T cell leukemia, another rare condition associated with HTLV-I.

DIAGNOSIS

HAM/TSP is characterized by the development of signs of spinal cord dysfunction, including paraparesis and urinary incontinence. Sensory changes are less common. The diagnosis is based on the exclusion of space-occupying lesions by MRI of the spinal cord in the setting of HTLV-1 seropositivity. Additionally, the T_2-weighted images (T_2 refers to the spin-spin, or transverse, relaxation time) may demonstrate areas of increased signal intensity in the spinal cord and, occasionally, above the foramen magnum; these findings may resemble the radiologic findings associated with multiple sclerosis, although these entities are clearly distinct. The CSF of patients with HAM may demonstrate high levels of protein, and most patients have oligoclonal bands and an elevated IgG index, which are indicative of intrathecal antibody synthesis. Cells resembling those of adult T cell leukemia may also be present in the CSF. Patients with HAM/TSP will usually have a progressive course, leading to lower limb paralysis and, possibly, arm involvement. HAM/TSP is rarely fatal, however.

Table 1 Stages of HIV Dementia (AIDS Dementia Complex)

Stage	*Characteristics*
0 (normal)	Normal mental and motor function
0.5 (subclinical)	Minimal or equivocal symptoms; no impairment of work; can perform tasks of daily life; few, if any, neurologic signs; and normal motor function
1 (mild)	Mild but unequivocal evidence of functional and intellectual impairment, although able to perform all but the most demanding tasks of work or daily life; ambulatory without assistance
2 (moderate)	Incapacitated for work but able to perform basic tasks of daily life; may need unilateral assistance for walking (cane)
3 (severe)	Major intellectual dysfunction, with slowing of mental processes or severe motor disability, requiring bilateral assistance for ambulation
4 (end stage)	Nearly vegetative, with minimal intellectual and social comprehension; nearly mute and paraparetic or paraplegic

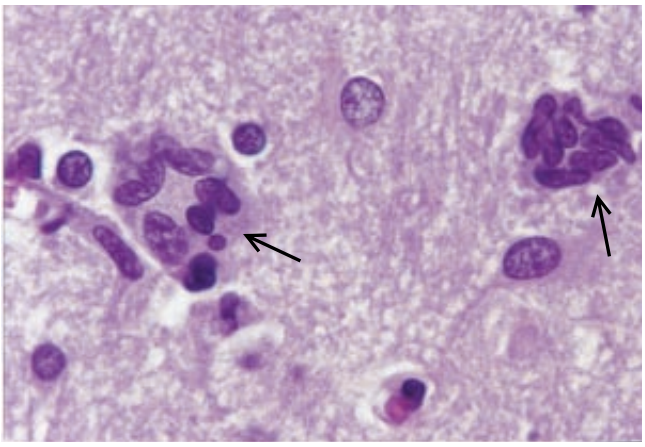


Figure 2 **Pathology of HIV-associated dementia (HAD). Multinucleated giant cells (arrows) formed by the fusion of HIV-infected microglia or macrophages are the most specific finding in HIV dementia. Fusion occurs through the interaction of the HIV glycoproteins with the cell receptors for the virus. Other pathologic changes (not shown) include astrocytosis and myelin pallor.**

Autopsy studies have revealed low levels of HTLV-I in the spinal cord of HAM/TSP patients, indicating that direct cytolysis by virus infection is probably not the cause of the myelopathy.[15] The pathogenesis is generally thought to be, at the very least, associated with infiltration of the spinal cord by $CD8^+$ T cells, as well as possibly with antibodies against the viral protein tax.[16,17] In one series of studies, these antibodies were linked to molecular mimicry of a neuronal protein.[17]

TREATMENT

Because of the lack of evidence that direct viral infection mediates HAM, most therapeutic trials have been directed toward reduction of the immune attack on the CNS. To date, there is no established therapy for HAM/TSP, but some patients show improved neurologic function when treated with interferon alfa (IFN-α).[18]

Transmissible Spongiform Encephalopathies

The spongiform encephalopathies, a group of neurologic disorders of humans and other mammals, are characterized by (1) subacute, progressive deterioration of neurologic function involving several regions of the neuraxis (cognitive and motor dysfunction are most common, although other systems may also be involved); (2) spongiform neuropathologic changes in the affected areas of the CNS; (3) experimental transmissibility to either the same or a related species, with long incubation periods,

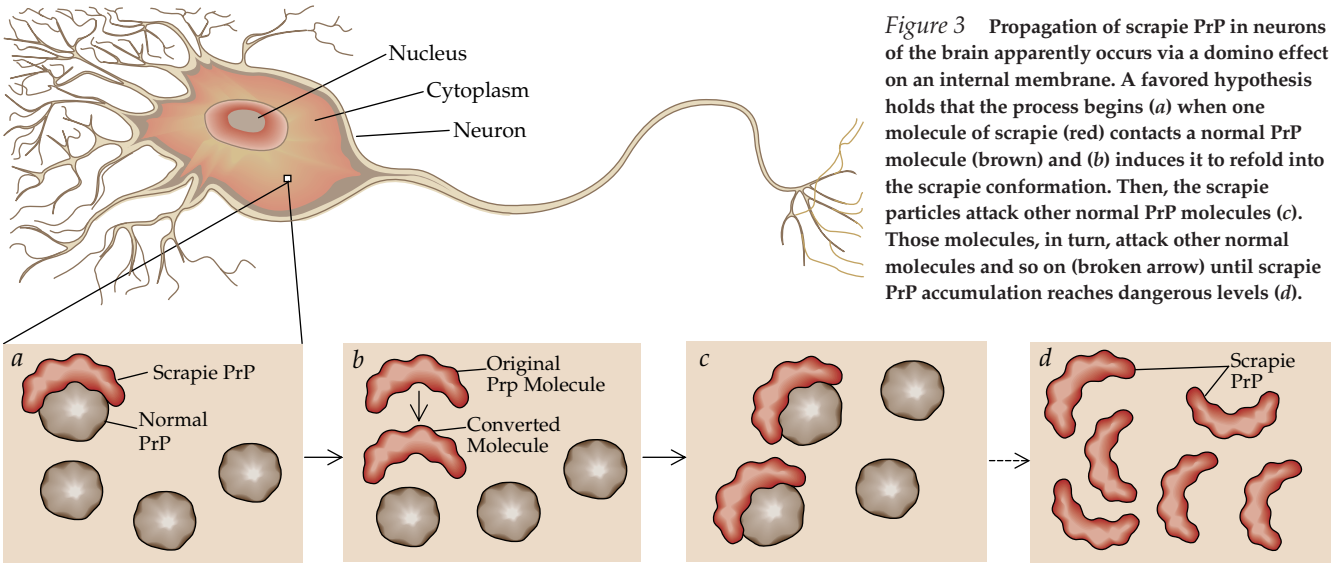


Figure 3 **Propagation of scrapie PrP in neurons of the brain apparently occurs via a domino effect on an internal membrane. A favored hypothesis holds that the process begins (*a*) when one molecule of scrapie (red) contacts a normal PrP molecule (brown) and (*b*) induces it to refold into the scrapie conformation. Then, the scrapie particles attack other normal PrP molecules (*c*). Those molecules, in turn, attack other normal molecules and so on (broken arrow) until scrapie PrP accumulation reaches dangerous levels (*d*).**

typically many months or years; and (4) no evidence of conventional transmissible agents such as viruses.

In humans, the spongiform encephalopathies include both sporadic and genetic forms of CJD, Gerstmann-Sträussler syndrome (GSS), kuru disease (a spongiform encephalopathy associated with ritual endocannibalism of the Fore tribe in the remote highlands of New Guinea), and familial fatal insomnia (FFI), as well as others. Parallel diseases in other species are sheep scrapie, transmissible mink encephalopathy, and chronic wasting disease in deer and elk in the United States.[19,20] Spongiform diseases have been experimentally transmitted to mice and hamsters from sheep tissues, to monkeys and chimpanzees from human tissues, and accidentally to dairy cows (bovine spongiform encephalopathy [BSE], or mad-cow disease) perhaps by dietary supplementation with processed organs from infected sheep. Although there are species barriers to infection, the spongiform diseases are generally considered to be caused by similar processes.

CJD and scrapie had been described for many years when transmission of the human form of the disease was first accomplished by intracerebral inoculation of primates with CNS tissues from patients dying of kuru.[19,21] Kuru disappeared with the cessation of cannibalistic practices, and human-to-human transmission of CJD and other spongiform encephalopathies is now limited to rare cases of accidental transplantation of an organ from a diseased person and parenteral exposure to CJD tissues through contaminated instruments; in new-variant CJD (nvCJD), transmission may be possible by blood transfusion (see below).[22]

PRION HYPOTHESIS OF PATHOGENESIS

Initially championed by Prusiner and colleagues and now widely accepted, the prion hypothesis provides a model for the transmission of CJD and other spongiform encephalopathies. The hypothesis is based on the unquestionable presence of a protease-resistant pathogenic form of an endogenous protein (prion protein, or PrP) in the brains of all affected species.[23,24] The natural form of the protein, PrP^{c}, is a glycosylated protein of unknown function containing lengthy coils (α-helices). In the brains of affected mammals, PrP^{c} appears to misfold into a structure with a high concentration of β-pleated sheets, a protein conformation that promotes the formation of crystals [*see Figure 3*]. The misfolded protein (PrP^{Sc}, or scrapie PrP) is deposited in amyloid fibrils, or plaques, in the diseased brain. Together with the spongy appearance of the diseased brain [*see Figure 4*], these proteinaceous deposits are characteristic of the neuropathology of all the spongiform encephalopathies, although the extent and regional location of the deposit varies from condition to condition.

Evidence supporting PrP involvement in the spongiform diseases includes experiments in which mice were genetically engineered to delete (knock out) the *PrP* gene. These mice became resistant to scrapie-infected brain and did not develop any spongiform changes, even after high concentrations of scrapie-infected tissue were injected intracerebrally. In contrast, transgenic mice that were engineered to express the *PrP* gene from other species, including humans, became particularly susceptible to the diseased brain harvested from the species supplying the transgene.[25] Because there are species barriers to the transmission of spongiform change, these latter experiments suggest that the transgene is responsible for promoting the disease process. A synthetic form of the PrP^{Sc} has been used to induce scrapie in susceptible transgenic mice that overexpressed the endogenous PrP.[26]

How does a protein devoid of nucleic acid function act as an infectious agent and encode its own replication in diseased brain? One scheme proposes that a misfolded PrP^{Sc} serves as a seed, which combines with PrP^{c} from normal tissue [*see Figure 3*]. This interaction induces PrP^{c} to change at least a portion of its α-helical structure into a structure composed primarily of β-pleated sheets. Because the β-pleated sheet–rich PrP^{Sc} is less soluble, the reaction is driven in a single direction. The initial aggregate then expands through the recruitment of additional PrP^{c} molecules, producing an insoluble deposition that is presumably injurious to neurons and neuroglia. This theory has been reviewed recently.[25]

In this mechanism, the source of the seed PrP^{Sc} in spontaneously occurring spongiform encephalopathy is unknown. In a condition such as BSE, in which the disease appears to be transmitted by feeding diseased sheep tissues to cows, the seed protein would have to have been ingested and transmitted to the

brain from peripheral tissue. The spleen and other lymphoid organs are known to harbor scrapie infectivity; therefore, circulating lymphoid cells in the CNS are prime candidates for this neuroinvasion. Extensive epidemiologic surveys have not supported a role for blood transfusion in the transmission of sporadic CJD; however, a case of a patient who developed nvCJD after a blood transfusion has raised the suspicion that blood transfusion may serve as a mechanism of transmission in certain instances.[22]

CREUTZFELDT-JAKOB DISEASE

Diagnosis

The diagnosis of CJD is often based on clinical findings. The most important feature is a dementia that progresses over weeks to months, rather than years, which is more typical of the dementia in Alzheimer disease. Therefore, rapid cognitive deterioration should alert a clinician to the potential diagnosis of CJD. In addition, patients may complain of insomnia, fatigue, and visual problems. In some instances, persons with CJD may be cortically blind. Neurologic examination shows evidence of pyramidal tract involvement and occasional muscle atrophy. Myoclonus (a jerky reaction, particularly prominent after a startle) is an important physical finding, but its absence does not rule out CJD. Other findings may include cerebellar ataxia and seizures.

No single test other than brain biopsy can confirm the diagnosis of CJD, but this is cumbersome to perform because of the need for special treatment of instruments. A combination of laboratory tests and imaging have made the diagnosis less dependent on biopsy. Specifically, examination of the CSF for an abnormal protein (14-3-3 protein), particularly in laboratories with specific expertise, can be useful.[27,28] The electroencephalogram may demonstrate characteristic triphasic forms. Importantly, new imaging modalities such as diffusion-weighted MRI are being used with increasing frequency to demonstrate the persistent edema and cortical involvement characteristic of CJD. Typical findings are a well-defined gyral pattern of restricted diffusion.[29]

Treatment

CJD is uniformly fatal, usually within 2 years of onset. In vitro studies suggest a possible therapeutic role for antimalarial agents and phenothiazines,[30] but clinical studies have not been conclusive. A recent large study reviewed the parameters associated with longer and shorter survival.[31]

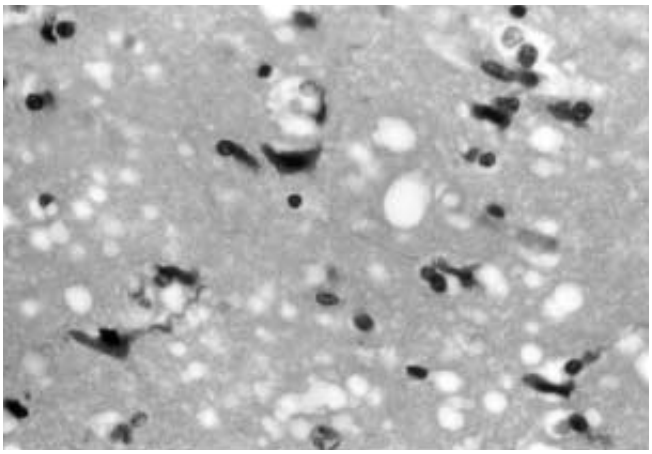


Figure 4 **Spongiform change is characteristic in the brain of a patient with Creutzfeldt-Jakob disease.**

NEW-VARIANT CJD

Bovine spongiform encephalopathy, a spongiform encephalopathy that occurred primarily in the United Kingdom, is associated with the consumption of protein supplements derived from ruminant tissue by cows. Aggressive management of potentially infected herds has reduced the incidence of BSE. However, nvCJD, appearing in young adults and with characteristic neuropathologic findings, has been temporally associated with the appearance of BSE in the United Kingdom. Reliable diagnostic criteria for nvCJD have been formulated.[32] Because there are insufficient epidemiologic data to rule out potential transmission of nvCJD through blood products, the Food and Drug Administration recommends that persons who lived in the United Kingdom for more than 6 months between 1980 and 1996 and persons suspected of nvCJD not donate blood or blood products. Current epidemiologic estimates suggest that the United Kingdom epidemic will be less widespread than initially predicted by some experts; the number of new cases appears to have reached a plateau in 2003 and 2004.[33] No cases of nvCJD have been reported in individuals who have continuously resided in the United States[34]; however, the CDC reported in 2002 the occurrence of nvCJD in a 22-year-old Florida resident who was born and raised in the United Kingdom.[35]

The nvCJD is distinguished from sporadic CJD by the wide difference in patients' ages (i.e., patients with nvCJD are much younger), the clinical presentation with neuropsychiatric abnormalities, and MRI changes in the basal ganglia. The nvCJD cases are also exclusively homozygous for methionine at a specific codon (129) of the PrP protein, whereas the cases of spontaneous CJD are more reflective of the population,[36] although homozygosity at codon 129 is associated with a worse prognosis.[31]

GENETIC DISEASES ASSOCIATED WITH PRP MUTATIONS

In humans, the spongiform diseases include, in addition to spontaneous or accidentally transmitted CJD, genetic forms of CJD and even rarer conditions, such as GSS and FFI.[37,38] In GSS, which is transmitted as an autosomal dominant trait, the affected persons develop ataxia, followed by dementia; patients with FFI are unable to sleep. All these conditions can be transmitted to susceptible nonhuman primates by inoculation of diseased tissues. Remarkably, these conditions have been associated with mutations in the *PrP* gene, and in one series of experiments, mice engineered with the *GSS* gene mutation developed spongiform changes.[37] The mutations on the *PrP* gene are thought to be involved in promoting the conversion to a misfolded, insoluble state.

The existence of diseases that are simultaneously genetic and infectious and the association of mutations in the *PrP* gene with these diseases provide strong evidence in support of the prion hypothesis, although the hypothesis initially met with considerable skepticism because it violated the central dogma of molecular biology.[39] Recent experiments with PrP in transgenic animals, as well as other prion models in primitive eukaryotes, have established that protein-only transmission of phenotypes occurs in nature.

Progressive Multifocal Leukoencephalopathy

PML is a disease of the brain white matter—characterized by dementia and progressive motor deterioration—that is caused by

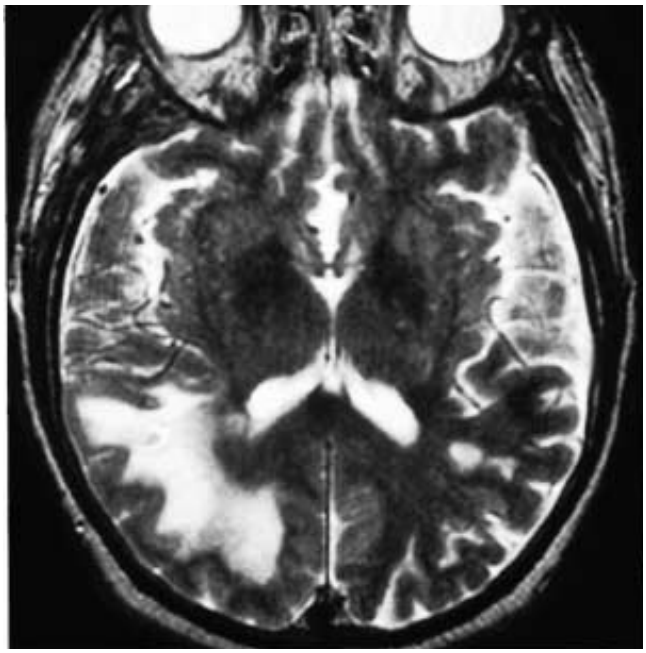


Figure 5 **Magnetic resonance imaging in progressive multifocal leukoencephalopathy. T_2-weighted MRI of a patient with progressive multifocal leukoencephalopathy (PML) demonstrates characteristic changes in the subcortical white matter of the right occipitoparietal region (increased intensity). As is typical of this disease, the abnormalities respect the gray matter–white matter junction, which accounts for the outline of the cortical sulci.**

a papovavirus-related DNA virus, referred to as JC virus. Other symptoms of PML include loss of vision and incontinence. PML was first described in patients with immunodeficiency caused by hematologic malignancies or by therapy for the prevention of transplant rejection. PML remained a very rare condition, however, until the appearance of AIDS. Before the widespread use of HAART, epidemiologic studies indicated that between 1% and 5% of HIV-infected persons would eventually develop PML. In many HIV-infected persons, the development of PML heralds the onset of immunodeficiency and is the AIDS-defining illness; in others, it is a complication associated with other opportunistic infections. Since the widespread availability of HAART, the incidence of PML has dropped considerably in HIV patients.

PATHOGENESIS

Oligodendrocytes, which are the cells responsible for the manufacture and maintenance of myelin, are the principal target of JC virus. Therefore, oligodendroglial cytolysis and apoptosis are the principal pathogenic mechanisms in PML, and they lead to the characteristic severe demyelination. The initial event in the development of PML appears to be the rise of a microscopic focus of oligodendroglial infection, from which virus spreads centrifugally to involve large segments of the white matter. Astrocytes, as well as some oligodendrocytes, demonstrate bizarre morphologic changes similar to those associated with some CNS tumors. There is little involvement of the gray matter or neurons in PML, and any involvement that does occur may reflect death of adjacent oligodendrocytes, perhaps through the recruitment of microglia and macrophages and concomitant inflammatory reaction.[40]

JC Virus and Its Role in the Pathogenesis of PML

JC virus is related to, but clearly distinct from, other polyomaviruses.[41] A large percentage of the adult population in the United States (70% to 90%) have antibodies against JC virus, indicating that this is a common infection, but no diseases have been attributed to the primary infection.

The evidence of widespread infection and the isolation of JC virus from the urine of some immunosuppressed patients are consistent with the hypothesis that PML represents the reactivation of a latent DNA virus. The reservoir organ has not been identified, however. The absence of JC virus DNA in the brains of normal persons suggests that JC virus remains latent in other tissue. Lymphocytes, lymphoid organs, and kidneys have been considered prime candidates because JC virus has been detected in these tissues by PCR or has been isolated from these tissues.[42,43] However, the JC virus isolated from the peripheral tissue of nonimmunosuppressed persons (the archetypal virus) differs from that isolated from patients with PML in the regions responsible for the regulation of viral expression.[44] One intriguing hypothesis, which may explain these findings, is that the archetypal virus is modified by deletion or duplication and is thus converted into a neurotropic form that is carried into the brain in infected B cells. The development of PML in two patients with multiple sclerosis who were treated with a combination of immunomodulators reinforced the hypothesis that JC virus is normally latent in many individuals.[45]

DIAGNOSIS

The diagnosis of PML should be considered in any person with a subacute, progressive illness involving motor function and cognition in the setting of known or presumed immunodeficiency (e.g., AIDS or malignancy). Most cases exhibit features of white-matter involvement, including ataxia, visual defects, and weakness. In rare cases, PML occurs in otherwise healthy persons. PML should be strongly suspected if MRIs show confluent areas of increased signal intensity in the T_2-weighted images [*see Figure 5*]. These areas, which respect the boundaries between gray matter and white matter, may be hypointense in the T_1-weighted images (T_1 refers to the spin-lattice, or longitudinal, relaxation time). The CSF is usually normal in PML; CSF pleocytosis, when present, is usually mild.

Pathologic examination of PML-diseased brain shows dramatic changes in the subcortical white matter. Traditionally, the definitive diagnosis of PML required a brain biopsy. However, biopsy has been largely superseded by the use of PCR to detect JC virus in the CSF.[46,47] The frequency of false positive reactions is low, and it has been estimated that CSF PCR has a sensitivity and specificity greater than 90%.[47]

TREATMENT

Improved outcomes in PML have been reported in AIDS patients treated with HAART.[48] Adjunctive treatment with cidofovir may also be beneficial, although the evidence for this is anecdotal.[49]

Subacute Sclerosing Panencephalitis

Subacute sclerosing panencephalitis (SSPE) is a rare sequela of measles infection (one per 100,000 cases). SSPE can also result from vaccination with a live, attenuated strain of measles, but the rate of development is at least 10 times less than that associat-

ed with wild-type measles infection. A similar disorder, progressive rubella paraencephalitis, is caused by rubella virus.

SSPE presents as a progressive encephalitis that develops several years after measles infection. The condition appears to result from a mutation in the measles virus that affects virus packaging, so the virus can spread only by fusion between adjacent infected and uninfected cells. This may explain viral persistence despite a strong humoral response.

Clinical features of SSPE include cognitive changes, loss of vision, seizures, and, characteristically, myoclonus. MRI may show hyperintensity of white and gray matter, and the CSF may show elevated titers of antimeasle antibody.[50]

A recent report suggested that because of the rarity of measles in the United States, SSPE often goes unrecognized, particularly in immigrant children.[51] There is no established therapy for SSPE, and its outcome is nearly always fatal

---

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. McArthur JC: HIV-associated dementia: an evolving disease. J Neuroimmunol 157:3, 2004

2. Sacktor N, Lyles RH, Skolasky R, et al: HIV-associated neurologic disease incidence changes: Multicenter AIDS Cohort Study, 1990-1998. Neurology 56:257, 2001

3. McArthur JC, Haughey N, Gartner S, et al: Human immunodeficiency virus-associated dementia: an evolving disease. J Neurovirol 9:205, 2003

4. Luciano CA, Pardo CA, McArthur JC: Recent developments in the HIV neuropathies. Curr Opin Neurol 16:403, 2003

5. Gonzalez-Scarano F, Martin-Garcia J: The neuropathogenesis of AIDS. Nat Rev Immunol 5:69, 2005

6. Albright AV, Soldan SS, Gonzalez-Scarano: Pathogenesis of human immunodeficiency virus-induced neurological disease. J Neurovirol 9:222, 2003

7. Pomerantz RJ: Effects of HIV-1 Vpr on neuroinvasion and neuropathogenesis. DNA Cell Biol 23:227, 2004

8. Nath A: Human immunodeficiency virus (HIV) proteins in neuropathogenesis of HIV dementia. J Infect Dis 186(suppl 2):S193, 2002

9. Trillo-Pazos G, Diamanturos A, Rislove L, et al: Detection of HIV-1 DNA in microglia/macrophages, astrocytes and neurons isolated from brain tissue with HIV-1 encephalitis by laser capture microdissection. Brain Pathol 13:144, 2003

10. Takahashi K, Wesselingh SL, Griffin DE, et al: Localization of HIV-1 in human brain using polymerase chain reaction/in situ hybridization and immunohistochemistry. Ann Neurol 39:705, 1996

11. McArthur JC, Cohen BA, Selnes OA, et al: Low prevalence of neurological and neuropsychological abnormalities in otherwise healthy HIV-1 infected persons: results from the multicenter AIDS cohort study. Ann Neurol 26:601, 1989

12. Tozzi V, Balestra P, Galgani S, et al: Neurocognitive performance and quality of life in patients with HIV infection. AIDS Res Hum Retroviruses 19:643, 2003

13. Glass JD, Fedor H, Wesselingh SL, et al: Immunocytochemical quantitation of human immunodeficiency virus in the brain: correlations with dementia. Ann Neurol 38:755, 1995

14. Letendre SL, McCutchan JA, Childers ME, et al: Enhancing antiretroviral therapy for human immunodeficiency virus cognitive disorders. Ann Neurol 56:416, 2004

15. Kuroda Y, Matsui M, Kikuchi M, et al: In situ demonstration of the HTLV-I genome in the spinal cord of a patient with HTLV-I myelopathy. Neurology 44:2295, 1994

16. Jacobson S: Human T-lymphotropic virus, type-I myelopathy: an immunopathologically mediated chronic progressive disease of the central nervous system. Curr Opin Neurol 8:179, 1995

17. Levin MC, Lee SM, Kalume F, et al: Autoimmunity due to molecular mimicry as a cause of neurological disease. Nat Med 8:509, 2002

18. Saito M, Nakagawa M, Kaseda S, et al: Decreased human T lymphotropic virus type I (HTLV-I) provirus load and alteration in T cell phenotype after interferon-alpha therapy for HTLV-I-associated myelopathy/tropical spastic paraparesis. J Infect Dis 189:29, 2004

19. Brown P, Gibbs CJ, Rodgers-Johnson P, et al: Human spongiform encephalopathy: the NIH series of 300 cases of experimentally transmitted disease. Ann Neurol 35:513, 1994

20. Belay ED, Gambetti P, Schonberger LB, et al: Creutzfeld-Jakob disease in unusually young patients who consumed venison. Arch Neurol 58:1673, 2001

21. Brown P, Cervenakova L, Goldfarb LG, et al: Iatrogenic Creutzfeldt-Jakob disease: an example of the interplay between ancient genes and modern medicine. Neurology 44:291, 1994

22. Llewelyn CA, Hewitt PE, Knight RS, et al: Possible transmission of variant Creutzfeldt-Jakob disease by blood transfusion. Lancet 363:417, 2004

23. Soto C, Castilla J: The controversial protein-only hypothesis of prion propagation. Nat Med 10(suppl):S63, 2004

24. Prusiner SB: Molecular biology of prion disease. Science 252:1515, 1991

25. Telling GC, Scott M, Hsiao KK, et al: Transmission of Creutzfeldt-Jakob disease from humans to transgenic mice expressing chimeric human-mouse prion protein. Proc Natl Acad Sci USA 91:9936, 1994

26. Legname G, Baskadov IV, Nguyen HO, et al: Synthetic mammalian prions. Science 305:673, 2004

27. Zerr I, Bodemer M, Gefeller O, et al: Detection of 14-3-3 protein in the cerebrospinal fluid supports the diagnosis of Creutzfeldt-Jakob disease. Ann Neurol 43:32, 1998

28. Zerr I, Schulz-Schaeffer WJ, Giese A, et al: Current clinical diagnosis in Creutzfeld-Jakob disease: identification of uncommon variants. Ann Neurol 48:323, 2000

29. Demaerel P, Sciot R, Robberecht W, et al: Accuracy of diffusion-weighted MR imaging in the diagnosis of sporadic Creutzfeldt-Jakob disease. J Neurol 250:222, 2003

30. Korth C, May BC, Cohen FE, et al: Acridine and phenothiazine derivatives as pharmacotherapeutics for prion disease. Proc Natl Acad Sci USA 98:9836, 2001

31. Pocchiari M, Puopolo M, Croes EA, et al: Predictors of survival in sporadic Creutzfeldt-Jakob disease and other human transmissible spongiform encephalopathies. Brain 127:2348, 2004

32. Will RG, Zeidler MN, Stewart GE, et al: Diagnosis of new variant Creutzfeldt-Jakob disease. Ann Neurol 47:575, 2000

33. Tan L, Williams MA, Khan MK, et al: Risk of transmission of bovine spongiform encephalopathy to humans in the United States: a report of the Council on Scientific Affairs. American Medical Association. JAMA 281:2330, 1999

34. Department of Health, United Kingdom. Monthly Creutzfeldt-Jakob disease statistics, March 2005
http://www.dh.gov.uk/PolicyAndGuidance/HealthAndSocialCareTopics/CJD/CJDGeneralInformation/CJDGeneralArticle/fs/en?CONTENT_ID=4032396&chk=5shT1Z

35. Probable variant Creutzfeldt-Jakob disease in a U.S. resident—Florida, 2002. MMWR Morb Mortal Wkly Rep 51:927, 2002

36. Valleron AJ, Boelle PY, Will R: Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. Science 294:1726, 2001

37. Hsiao K, Baker HF, Crow TJ, et al: Linkage of a prion protein missense variant to Gerstmann-Sträussler syndrome. Nature 338:342, 1989

38. Chapman J, Arlazoroff A, Goldfarb LG, et al: Fatal insomnia in a case of familial Creutzfeldt-Jakob disease with the codon 200Lys mutation. Neurology 46:758, 1996

39. Parchi P, Castellani R, Capellari S, et al: Molecular basis of phenotypic variability in sporadic Creutzfeldt-Jakob disease. Ann Neurol 39:767, 1996

40. Major EO, Amemiya K, Tornatore CS, et al: Pathogenesis and molecular biology of progressive multifocal leukoencephalopathy, the JC virus-induced demyelinating disease of the human brain. Clin Microbiol Rev 5:49, 1992

41. ZuRhein GM, Chou S-M: Particles resembling papova viruses in human cerebral demyelination disease. Science 148:1477, 1965

42. Tornatore C, Berger JR, Houff SA, et al: Detection of JC virus DNA in peripheral lymphocytes from patients with and without progressive multifocal leukoencephalopathy. Ann Neurol 31:454, 1992

43. Dubois V, Lafon ME, Ragnaud JM, et al: Detection of JC virus DNA in the peripheral blood leukocytes of HIV-infected patients. AIDS 10:353, 1996

44. Daniel AM, Swenson JJ, Mayreddy RPR, et al: Sequences within the early and late promoters of archetype JC virus restrict viral DNA replication and infectivity. Virology 216:90, 1996

45. Suspended marketing of Tysabri (natalizumab). FDA Public Health Advisory. Center for Drug Evaluation and Research. February 28, 2005
http://www.fda.gov/cder/drug/advisory/natalizumab.htm

46. Weber T, Turner RW, Frye S, et al: Progressive multifocal leukoencephalopathy diagnosed by amplification of JC virus-specific DNA from cerebrospinal fluid. AIDS 8:49, 1994

47. McGuire D, Barhite S, Hollander H, et al: Virus DNA in cerebrospinal fluid of human immunodeficiency virus-infected patients: predictive value for progressive multifocal leukoencephalopathy. Ann Neurol 37:395, 1995

48. Albrecht H, Hoffmann C, Degen O, et al: Highly active antiretroviral therapy significantly improves the prognosis of patients with HIV-associated progressive multifocal leukoencephalopathy. AIDS 12:1149, 1998

49. De Luca A, Giancola ML, Ammassari A, et al: Potent anti-retroviral therapy with or without cidofovir for AIDS-associated progressive multifocal leukoencephalopathy: extended followup of an observational study. J Neurovirol 7:364, 2001

50. Khadilkar SV, Patil SG, Kulkarni KS: A study of SSPE: early clinical findings. J Pediatr Neurol 2:73, 2004

51. Honarmand S, Glaser CA, Chow E, et al: SSPE in the differential diagnosis of encephalitis. Neurology 63:1489, 2004

Acknowledgments

Figure 2 Dimitry Schidlovsky.

Figure 3 Dimitry Schidlovsky. Adapted from "The Prion Diseases," by S. Prusiner, in Scientific American 272(1):48, 1995.

Figure 4 Dr. E. Lavi, Department of Pathology and Laboratory Medicine, University of Pennsylvania.

I CANCER EPIDEMIOLOGY AND PREVENTION

Alfred I. Neugut, m.d., ph.d., f.a.c.p.
Frederick P. Li, m.d.

An estimated 1,368,000 new cases of invasive cancer and 553,000 cancer deaths will occur in the United States in 2004.[1] An additional 1,000,000 cases of superficial squamous cell and basal cell carcinomas will also be diagnosed, but these rarely result in death. After heart disease, cancer is the leading cause of death in the United States, accounting for one in four deaths.[1] Overall age-adjusted cancer mortality has declined since the early 1990s, reversing the rising cancer mortality of preceding decades.[2] The number of cancer deaths has also reached a plateau despite a growing elderly population, who are at higher risk for cancer. Similar trends have been reported in Europe.[3] Worldwide, cancer control will also become increasingly important as life expectancy improves because of lower infant mortality and fewer deaths from infectious diseases.

Morbidity and mortality from many forms of cancer can be controlled through primary or secondary prevention. Primary prevention can be defined as risk modification to lower cancer occurrence. Secondary prevention refers to the use of screening tests to detect cancers at earlier stages. Chemoprevention—the use of medications to inhibit or reverse the carcinogenic process—is being studied as another approach to reduce cancer rates. Chemoprevention has been effective in reducing the incidence of breast cancer in women, the occurrence of second cancers in patients with oral cancers, and colorectal cancers in susceptible patients.[4-6]

Environmental Carcinogens

Up to 90% of all cancers in the United States may be primarily the result of environmental factors.[7] This conclusion is based on evidence that incidence rates of specific cancers change over time, differ substantially among populations, and can be altered by migration. Environment encompasses a person's way of life, including occupation, diet, tobacco use, alcohol consumption, and sun exposure. Alterations of the environment may be sufficient to prevent most cancers, but heredity also exerts an influence on the multistage development of human cancers.

Conclusive evidence regarding the etiologic role of specific environmental factors is difficult to obtain in studies of humans because (1) the time between a carcinogenic exposure and diagnosis of a resultant cancer is usually years or decades, (2) the amount of exposure to a carcinogen can rarely be quantified accurately in retrospect, and (3) most carcinogens induce cancer in only a small proportion of exposed persons.[8] The role of complex dietary factors in cancer development is particularly difficult to study for these reasons. However, data are available from large epidemiologic studies that employ a prospective study design, careful measurement of dietary intake, and bioassays as dosimeters of nutrients consumed. Evidence suggests that obesity and diets characterized by high caloric intake, increased consumption of red meat and animal fat, and decreased consumption of fruits and vegetables increase the risk of cancers of the breast, the colon, lung, pancreas, oral cavity, and other organs. However, the specific ingredients in the diet that elevate or lower cancer risk have been difficult to identify.[8]

Clinicians often discover the link between an unusual carcinogenic exposure and exceptional risk of a rare cancer. Carcinogens that have been identified in this way include the hormone diethylstilbestrol and the industrial chemicals bis(chloromethyl)-ether and vinyl chloride. Most cancers, however, arise as a result of fairly common exposures that may produce only moderate elevations in risk. To detect these associations, epidemiologists strive for large study populations, accurate exposure measurement, informed judgments regarding the relevant temporal relation between exposure and cancer diagnosis, and recognition of confounding (extraneous) causes of the cancer under study.[9] Despite inherent difficulties in studying human populations, epidemiology remains the most useful method for studying carcinogenic effects in humans. A causal relation between a suspected carcinogen and cancer is strengthened by corroborating studies that provide plausible biologic mechanisms.

Certain physical, chemical, and biologic agents have been identified as carcinogens.[8] These agents may act on a single organ or at multiple sites [*see Table 1*]. Ultraviolet radiation from sunlight can cause cutaneous melanoma, basal cell carcinoma, and squamous cell carcinoma, whereas ionizing radiation can induce leukemias, sarcomas, and carcinomas at various sites.[10] Several dozen chemicals are known to be carcinogenic in humans; they induce leukemias, lymphomas, and carcinomas of diverse organs, particularly of the lung and upper airway. Many more chemical compounds can induce cancer in laboratory animals, but evidence linking these compounds to cancer in humans is insufficient or inconsistent. Strong evidence of viral oncogenesis in humans has been found for hepatitis B and C viruses in patients with liver cancer, human papillomavirus in cervical and anal cancers, Epstein-Barr virus in nasopharyngeal cancer and in some lymphomas, and human herpesvirus type 8 in Kaposi sarcoma.[11,12] The discovery that *Helicobacter pylori* causes gastric adenocarcinoma and certain lymphomas represents the first time a bacterium has been linked to the etiology of malignancy, and this is significant because it suggests that antibiotics may be used for the prevention and treatment of certain malignancies.[13,14]

During the past 10 years, biomarkers have been used increasingly to enhance exposure assessment, gain insights into disease mechanisms, and identify host susceptibility. Epidemiology has traditionally employed a quantitative approach rather than mechanistic steps to the understanding of disease causation. With the advent of modern molecular techniques, molecular epidemiologists can investigate the molecular pathogenesis of cancer. Biomarkers used in cancer epidemiology may be broadly classified into categories according to study design and may include markers of internal dose, biologically effective dose, early biologic effects, susceptibility, and disease.[15] Examples of common biomarkers include DNA adducts, polymorphisms in metabolic genes, and cytogenetic alterations and mutations in cancer cells.[16-19] Validated biomarkers can be useful for exposure assessment when traditional tools are not effective for identifying susceptible and exposed persons. For example, insulinlike growth factor–1 (IGF-1) and its main binding protein, IGFBP-3, have been investigated as potential risk markers for a variety of common cancers.[20-23] Studies show that high circulating IGF-1 and low IGFBP-3 concentrations are correlated with an increased risk of cancers of the

Table 1 Established Causes of Human Cancer

	Carcinogens	*Cancers*
Chemicals and Naturally Occurring Compounds	Aflatoxins, naturally occurring	Liver
	4-Aminobiphenyl	Bladder
	Arsenic	Lung, skin, bladder
	Asbestos	Lung, mesothelioma of pleura and peritoneum, larynx
	Benzene	Acute myelocytic leukemia
	Benzidine	Bladder
	Beryllium	Lung
	Bis(chloromethyl)ether	Lung
	Cadmium	Prostate, respiratory tract, genitourinary tract
	Chromium compounds	Lung
	Erionite	Mesothelioma
	Ethylene oxide	Leukemia
	Mustard gas	Pharynx, lung
	2-Naphthylamine	Bladder
	Nickel compounds	Lung, nasal sinuses
	Radon	Lung
	Silica	Lung
	Sunlight (UV)	Skin, melanoma, eye
	Tetrachlorodibenzo-*p*-dioxin	Lymphoma, sarcoma
	Vinyl chloride	Liver (angiosarcoma)
Medicines and Hormones	Azathioprine	Lymphoma, skin, liver
	Busulphan, treosulfan	Leukemia
	Chlorambucil	Leukemia
	Cyclophosphamide	Bladder, leukemia
	Cyclosporine	Lymphoma
	Estrogens	Endometrium, liver
	Ionizing radiation	Myelogenous leukemia, breast, thyroid, lung, other
	Melphalan	Leukemia
	Methoxsalen with psoralen plus ultraviolet A therapy (PUVA)	Basal cell and squamous cell skin cancers
	Nitrosourea	Leukemia
	Phenacetin	Renal pelvis, bladder
	Tamoxifen	Endometrium
	Thiotepa	Leukemia
Infectious Agents	Epstein-Barr virus	Nasopharynx, lymphomas (Burkitt, Hodgkin, non-Hodgkin)
	Helicobacter pylori	Stomach
	Hepatitis B and C viruses	Liver
	Human herpesvirus type 8	Kaposi sarcoma, MALT lymphoma
	Human T cell lymphotropic virus type I	Adult T cell leukemia/lymphoma
	Human papillomavirus types 16 and 18	Cervix, anus
	Opisthorchis viverrini	Liver
	Schistosoma haematobium	Bladder
Mixtures	Alcohol	Liver, esophagus, pharynx, larynx
	Betel quid	Mouth
	Coal tars	Skin, lung
	Involuntary smoking	Lung
	Mineral oils	Skin
	Salted fish	Nasopharynx
	Shale oils	Skin
	Soot	Skin, lung
	Tobacco	Lung, mouth, pharynx, larynx, esophagus, stomach, pancreas, bladder, kidney, renal pelvis, cervix
	Wood dust	Nasal sinuses

Source: International Agency for Research on Cancer (http://193.51.164.110)

colon, prostate, and lung, as well as premenopausal breast cancer. IGF-1 appears to promote tissue growth and prevent apoptosis, whereas IGFBP-3 modulates IGF-1 availability and induces apoptosis via an IGF-independent mechanism.[24]

TOBACCO

Tobacco is by far the major cause of lung cancer worldwide, particularly in developing nations.[25,26] In the United States, 30% of all cancer deaths are attributed to tobacco use.[27] Tobacco use is associated primarily with squamous cell and small cell carcinomas of the lung, as well as with pulmonary adenocarcinomas. Increased tobacco consumption in the United States and elsewhere has been followed by a rise in lung cancer rates several decades later. A series of reports from the Surgeon General have summarized data from individual epidemiologic studies.[28] At least nine large prospective studies of up to 1,000,000 persons each and 50 case-control studies worldwide have consistently shown that cigarette smoking increases the risk of lung cancer. This increased risk is related to the dose of smoking as measured by years of tobacco use, amount of tobacco consumed daily, tar and nicotine content of the cigarettes, and depth of inhalation of cigarette smoke. Programs that promote reductions in tobacco use among adolescents and young adults, if addressed to these age groups, can prove effective because most smokers become addicted to cigarettes at an early age.[29,30] With smoking cessation, the risk decreases, although more than a decade is required to return to the low cancer risk level of nonsmokers; thus, both prevention and early cessation of tobacco use are important.[31]

Cigarette smoking and alternative methods of tobacco use increase the risk of laryngeal and oral cancers.[28] Fewer lung cancers develop in pipe and cigar smokers than in cigarette smokers, but pipe and cigar smokers are at high risk for oral mucosal cancers, including lip cancer in pipe users.[32] Use of smokeless tobacco also places individuals at high risk for cancer of the oral mucosa.[28] Long-term exposure to environmental tobacco smoke (passive smoking) has been associated with a 30% increased risk of lung cancer in nonsmokers and is estimated to account for 3,000 cases of lung cancer in the United States each year. The Third National Health and Nutrition Examination Survey reported in 1996[33] that 88% of a large population sample of nonsmokers had detectable levels of serum cotinine, a surrogate measure of tobacco exposure. Furthermore, the effects of passive smoking have been demonstrated in pet dogs that are at increased risk of certain malignancies only when their owners smoke.[34]

The causal effect between tobacco use and cancers of the lung and the head and neck is well established; however, tobacco also plays a major carcinogenic role in other organs of the upper aerodigestive tract. Notable among these is the esophagus, for which cigarette smoking is the major risk factor for cancer, particularly squamous cell carcinoma of the esophagus. Over the past three decades, there has been an increasing incidence of adenocarcinoma in the lowest third of the esophagus that is associated with a condition known as Barrett esophagus. Currently, this malignancy constitutes approximately 50% of esophageal cancers. Although associated with tobacco use, Barrett esophagus generally has been more closely linked to gastroesophageal reflux disease (GERD).[35-37]

Alcohol consumption increases the risk of cancers of the oral cavity and the head and neck, as well as cancers of the esophagus, but the risk is greatly increased when alcohol consumption is associated with tobacco use. Cigarette smoking and alcohol consumption are independent risk factors for oral cancer, and

their combined use results in a synergistic effect that has been associated with a 13-fold increase in risk in some studies.[38,39]

Tobacco also plays a significant etiologic role in renal cell carcinoma, bladder cancer, pancreatic cancer, cervical cancer, and gastric cancer. In each of these, the relative risk ranges between 1.5 and 3. Some studies have suggested that cigarette smoking is an etiologic factor in cancer of the colon and rectum, although this remains controversial.[40-42] Also controversial is the role of tobacco use in the development of breast cancer in women.[43]

Reduction in tobacco use is clearly the most important approach to cancer prevention. Physicians see approximately 85% of smokers in their offices at least once a year and can effectively help their patients to stop smoking by using structured, short-term counseling techniques administered by the physicians themselves or their office staff. Smoking-cessation efforts employ behavior modification techniques that have been successful with other addictive disorders; nicotine substitutes (e.g., patch or gum) have also been effective for selected patients. Relapse rates within the first year after cessation are high, and often many attempts to quit are required.[27] In men, the prevalence of smoking decreased from 52% to 27% between 1965 and 1995, and the incidence of lung cancer has declined since rates peaked in 1984.[2] In women, smoking prevalence has decreased from 34% to 23%, whereas lung cancer incidence has yet to decline.[2]

Policy approaches to tobacco control include restrictions on advertising that targets young people, aggressive antismoking education at the elementary school level, limits on government subsidies to tobacco growers, and taxation to raise prices of tobacco and tobacco products. Many communities have legislation requiring smoke-free workplaces, restaurants, and other facilities.[44] In 1998, a master settlement agreement was signed by 46 state attorneys general, a coalition of trial lawyers, and representatives of the five largest tobacco manufacturers in the United States. The settlement requires the tobacco companies to reimburse the states over $200 billion over 25 years for the cost of providing health care services to people with tobacco-related diseases.[45] It also prohibits targeting of youths by banning cartoon characters in advertising, sponsorship of certain events, and use of certain promotional tactics, such as the targeting of adolescents. Additional lawsuits at the state level, including a successful suit in Florida, have found tobacco companies liable for cancer and other diseases in smokers.[46]

In 1997, the Food and Drug Administration proposed extending its jurisdiction to the advertising and sale of tobacco products.[47] However, in 1998 the United States Court of Appeals ruled that the FDA does not have jurisdiction to regulate tobacco, and in 2000 the United States Supreme Court upheld that decision.[48]

DIET

Unusual geographic clustering has been observed for certain cancers of the digestive tract. Exceptionally high mortality rates are found for esophageal cancer in southern Africa and in parts of China and for stomach cancer in eastern Asia and eastern Europe. In the United States, nearly one fifth of all cancers arise in the digestive tract, most often in the colon or rectum. For unknown reasons, cancers of the right colon have become more common, whereas sigmoid colon and rectal cancers have diminished. Death rates from stomach cancer in the United States have declined sharply over the past 60 years, possibly because of improvements in food preservation and decreased consumption of salted, pickled, and smoked foods.[49] Incidence rates of certain digestive tract cancers change when people migrate and, as a result, modify their diets. For example, Japanese who migrate to the United States have a decline in stomach cancer rates, but their colon cancer rates rise within one generation to the levels found in the American-born white population.[49,50] Within the United States, colon cancer rates differ among subpopulations. Mormons and Seventh-Day Adventists, who tend to consume less meat, have lower rates of large bowel cancer. Mexican Americans have low rates of bowel cancer but high rates of stomach cancer.

Despite epidemiologic evidence implicating diet in digestive tract cancer, the roles of specific dietary elements remain uncertain. Suspected dietary factors in large bowel cancer include excess consumption of animal fat and red meat and excess total caloric intake.[51-53] Consumption of multivitamins containing folic acid may reduce colon cancer risk.[54,55] Although low fiber consumption has been suggested to increase colorectal cancer risk, a randomized trial and other recent studies did not show any benefit of a high-fiber, low-fat diet on the recurrence of colorectal adenomas.[56-58]

Recommended dietary changes to reduce cancer risk include controlling caloric intake, decreasing fat consumption to less than 30% of total daily caloric intake, and increasing consumption of fresh fruits and vegetables.[52] These dietary modifications may also reduce the risk of cardiovascular disease.

In general, despite widespread public interest and concern regarding the role of diet in cancer etiology, specific recommendations regarding appropriate dietary recommendations remain elusive. A major contribution in this regard is a monograph published by the World Health Organization that summarizes much of the current knowledge regarding the association between diet and cancer prevention.[59] Generally, a diet high in fresh fruits and vegetables and low in red meat seems most prudent. Much interest currently focuses on caloric consumption, physical activity, and energy utilization and how these factors contribute to obesity and cancer risk [*see* Obesity and Physical Inactivity, *below*]. Very recently, attention has focused on early-life patterns of food and caloric consumption and their role in establishing future cancer risk.

Numerous randomized trials have been undertaken to clarify the relationship between diet and cancer. Antioxidant vitamins, minerals such as selenium and calcium, low-fat diets, and other dietary patterns have been assessed as possible risk-lowering factors. Results have differed across these studies, and many studies are ongoing. The Alpha-Tocopherol, Beta-Carotene Cancer Prevention Trial randomized 29,133 Finnish male smokers to receive vitamin E, β-carotene, both, or a placebo to determine whether these supplements lowered the risk of lung cancer. Ironically, the study showed an 18% increased risk of lung cancer in patients who received β-carotene.[60] A postintervention follow-up study concluded that smokers should avoid β-carotene supplements.[61] These results suggest that dietary interventions may be harmful, and careful investigation is needed to determine their role in cancer prevention.[62]

OBESITY AND PHYSICAL INACTIVITY

Two risk factors for cancer that have come under increased study in recent years have been obesity and physical inactivity, which are related to energy intake and consumption. A study by Calle and colleagues demonstrated that obesity, which has been associated most notably with breast cancer, may be linked to a larger number of cancers than had previously been suspected, including colon, rectum, prostate, pancreas, kidney, and others.[63] Although the physiologic mechanisms linking obesity to hor-

monally related cancers (e.g., breast cancer and prostate cancer) are well elucidated, the biologic underpinnings between obesity and other cancers are not obvious.

Physical inactivity, which frequently accompanies obesity, has been implicated as a risk factor in many chronic diseases and is currently being assessed for its contribution to cancer risk. A Canadian study showed that physical inactivity, high energy intake, and obesity are associated with an increased risk of rectal cancer; the study further suggested that these three factors may have a synergistic effect that further increases cancer risk if they are all present.[64]

Increased physical activity, either on an occupational or on a recreational basis, has been shown to be a protective factor for a number of malignancies, including colorectal and breast cancer.[65,66] Whether this represents a surrogate for a healthy lifestyle, a confounder for obesity, or other physiologic mechanisms again remains unclear. Nonetheless, randomized trials are under way to assess its efficacy as a modifiable intervention for cancer prevention in persons at high risk for developing certain cancers.

INFECTIOUS AGENTS

Infectious agents have been known to cause cancer since 1911, when Peyton Rous discovered the Rous sarcoma virus. However, recent findings have made it clear that infectious agents are responsible for a larger portion of the cancer burden in the United States and around the world than was previously suspected. This is a cause for optimism because it suggests the preventive and management strategies developed for controlling infectious diseases, including sanitation, vaccination, antibiotics, and antivirals, may be applicable to cancer prevention.

Hepatitis B virus (HBV) infection is an established risk factor for the development of liver cancer in certain patient populations. Cancers of the liver, primarily hepatocellular carcinoma (HCC), are relatively rare in the United States, but the rates are rising.[67] HCC occurs largely in patients with cirrhosis caused by chronic hepatitis, hemochromatosis, and use of exogenous androgenic steroids. It also may occur as a result of occupational exposure to vinyl chloride and the subsequent development of angiosarcoma of the liver. Liver cancer is a leading cause of cancer death in certain developing nations, particularly in Africa and Asia. The major risk factor for liver cancer in these patients is chronic HBV infection, which is often acquired during infancy or early childhood. The risk of liver cancer is increased as much as 200-fold in HBV carriers. Hepatitis B vaccination decreases the risk of infection and subsequent occurrence of HCC. Since initiation of widescale hepatitis B vaccination in Taiwan, HCC rates have fallen by 25% or more.[68] In the United States, however, the rising incidence of hepatitis C virus (HCV), which causes hepatoma and for which a vaccine does not yet exist, poses an increasing problem.

Squamous cell carcinoma of the uterine cervix has been linked to human papillomavirus (HPV)[69-71] and has been studied in relation to sexual behavior. Incidence of the disease is increased in prostitutes and in other women who have multiple sexual partners, in women who have first coitus at an early age, and in women with a history of venereal disease. Cohort and case-control studies have demonstrated that HPV is associated with both in situ and invasive cervical carcinoma. HPV DNA has been found in up to 95% of cervical cancers. Furthermore, progression of cervical intraepithelial neoplasia is associated with high levels of HPV DNA within the cervical tissues. This finding has prompted two advances in the prevention and early detection of cervical cancer. First, the presence of HPV in the genital mucosa of women is now being used as a means for cervical cancer screening.[72] This technique has proved to be inexpensive and effective and has a high enough sensitivity and specificity for use in poor countries where the more commonly used Papanicolaou smear is not routinely available. Women are able to swab themselves; the swab is tested for the presence of HPV DNA, and its presence is then used to select those women for further testing where routine gynecologic care is not feasible.[73] A second major advance linked to the viral etiology of cervical cancer is the development of a vaccine for HPV. A randomized, multicenter trial found the vaccine to be effective; however, the vaccine must undergo further testing before it can be widely distributed.[74]

Another virus of interest as a carcinogen is Epstein-Barr virus (EBV). This DNA virus has been associated with a number of different malignancies, including nasopharyngeal carcinoma, Burkitt lymphoma, and possibly Hodgkin disease.

Particularly since the early 1980s, with the onset of AIDS and the HIV era, interest has focused on finding retroviruses or RNA viruses associated with cancer. This has been of particular interest because many of the oncogenic viruses found in rodent and animal studies (e.g., SV40 or Rous sarcoma virus) are RNA viruses. A major breakthrough in this regard was the discovery that human herpesvirus type 8 (HHV-8), a retrovirus, was the etiologic agent for Kaposi sarcoma.[75] Kaposi sarcoma, which classically occurs in elderly men of Mediterranean descent, became more widespread as a cancer commonly found in AIDS patients. Chang and Moore found HHV-8 to be the virus responsible for Kaposi sarcoma in both the AIDS-associated and the classic forms of the disease.[72] HHV-8 has been found to be responsible for other lymphomas as well.

Human T cell lymphotropic virus type I (HTLV-I) infection has been linked to the development of an aggressive form of leukemia, adult T cell leukemia/lymphoma.[76] This disease has an unusual geographic distribution, with clusters in Japan and the Caribbean. Spatial and temporal clusters of leukemia and Hodgkin disease have also been reported, but no etiologic agent has been found, and these clusters may be the result of chance. HIV type 1 (HIV-1) has been identified as the cause of AIDS. Patients with AIDS are at high risk for Kaposi sarcoma, Hodgkin disease, non-Hodgkin lymphoma (NHL), and other cancers.

In recent years, the incidence of NHL has increased in adolescents and adults. Both inherited and acquired immunodeficiency appear to be associated with an elevated risk for NHL. Organ transplant recipients are at high risk for NHL and carcinomas of the skin.[77] These neoplasms may appear within several weeks after renal or cardiac transplantation and therefore differ from most environmentally induced cancers that arise many years after exposure to carcinogens. The transplant-associated lymphomas have a predilection for the central nervous system. Immunosuppressive therapy with azathioprine and cyclosporine has been implicated as a risk factor in transplant recipients, although the transplanted cells per se may also have a carcinogenic influence. Epstein-Barr virus has been associated with nasopharyngeal cancer, Burkitt lymphoma, NHL, and Hodgkin disease.[11]

Evidence accrued over the past 20 years has established *H. pylori* as a causal agent in gastric cancer. This discovery is significant because it represents the first time a bacterium has been found to be associated with cancer etiology. *H. pylori* is a widely prevalent infection that increases the risk of gastric adenocarcinoma and gastric lymphoma of mucosa-associated lymphoid tissue (MALT) lymphomas.[13,14] The treatment of *H. pylori* infection with antibiotics

has resulted in cures of MALT tumors. Furthermore, a recent randomized trial from China has demonstrated that early treatment of *H. pylori* with antibiotics can decrease the subsequent incidence of gastric cancer in persons at high risk for that disease.[78]

In some parts of the world, such as the Middle East, a large number of cases of bladder cancer are attributable to chronic schistosomiasis, a parasite.

OCCUPATIONAL CARCINOGENS

Asbestos induces lung cancer and mesothelioma and is a major cause of cancers of the respiratory tract. Asbestos is estimated to cause several thousand cancer cases, primarily lung cancer, each year in the United States.[8] Asbestos is virtually indestructible and is widely used throughout industrialized environments. Mesothelioma has been reported to develop after a single identifiable exposure to asbestos. In contrast, lung cancer develops chiefly in workers who have been heavily exposed to asbestos, such as asbestos manufacturers, pipe fitters, and shipyard workers. Concerns have also been raised regarding low-level asbestos exposure associated with decaying insulation in older homes and buildings.[79] The carcinogenic effect of asbestos appears to be related to the long, needlelike physical configuration of the fiber. There are concerns that other fibers, such as certain fiberglass insulation materials, might pose a cancer risk.[80] Smoking and exposure to asbestos appear to act synergistically in producing cancer. The combination of smoking and occupational exposure to asbestos increases lung cancer risk about 60-fold.

Bladder cancer has been reported in workers exposed to certain aromatic amines in the dye, rubber, leather, tanning, and organic chemical industries. This has been found to be true for kidney cancers, as well.

Other occupational carcinogens include polycyclic aromatic hydrocarbons, such as those found in cigarette smoking. These are also derived from combustion of petroleum and its by-products in diesel engines or in other similar settings.

Certain miners, such as uranium miners, can be exposed to unusual carcinogens such as radon. Radon is inhaled and increases the risk of lung cancer. Studies have shown that there is a dose-response relationship associated with mining and that an interactive effect with cigarette smoking increases the risk of lung cancer in these miners.[81] Other types of miners, such as coal miners, may also have an increased risk of lung cancer.

CARCINOGENS AFFECTING THE REPRODUCTIVE SYSTEM

Breast cancer is the most common neoplasm in women in the United States and accounts for 30% of all new cancers diagnosed in women in 2004.[1] The cumulative incidence of breast cancer in American women is 9% by 74 years of age. The highest rates worldwide are observed in the industrialized countries of North America and Europe. Rising breast cancer rates in Japan and other newly industrialized nations, as well as in Asian immigrants to the United States, suggest environmental influences.[7,82]

Studies have focused on hormonal factors and the influence of obesity, exercise, diet, and parity in promoting carcinogenesis in the breast and female reproductive organs.[83-89] Risk of breast cancer increases with early menarche, nulliparity, older age at first birth, and late menopause. Available data suggest a slight increase in breast cancer risk with prolonged use of oral contraceptives in women younger than 35 years but not in older women.[83] Prolonged use of postmenopausal hormone replacement therapy has been shown to increase the risk of breast cancer, and the addition of progesterone appears to exacerbate this effect.[84,90]

In addition, oophorectomy reduces the incidence of breast cancer by about 50%. Taken together, these observations support a unified hypothesis that breast cancer development is promoted by prolonged exposure to cyclic ovarian secretion of estrogens and progesterone.

A higher risk of breast cancer has been found in women with benign lesions having histologic evidence of proliferation and atypia. Because most of the known risk factors for breast cancer cannot be modified easily, chemopreventive strategies are being explored for women at high risk for cancer. A large randomized trial has determined that tamoxifen substantially reduces the incidence of breast cancer in women at high risk and reduces the rate of contralateral breast cancer in breast cancer patients.[4]

Use of conjugated estrogens during menopause appears to retard the progression of osteoporosis and reduce menopausal symptoms.[90] However, these agents may promote endometrial cancer, and newer drugs, such as raloxifene, are under study in clinical trials.[91] Little is known about environmental causes of carcinoma of the ovary and endometrium. These malignant tumors have some risk factors in common with breast carcinoma, including increased risk with nulliparity and perhaps with obesity and high socioeconomic class.

Multiple primary adenocarcinomas of the breast, ovary, and endometrium have been reported in individuals and in families.[8] The risk of both ovarian and endometrial cancer may be lower in women who have used oral contraceptives. For women who were exposed in utero to diethylstilbestrol, the risk that adenocarcinoma of the vagina will develop in their early adult years is 1 in 1,000. Otherwise, this cancer is extraordinarily rare.[8]

In the past 25 years, age-adjusted incidence rates for prostate cancer in the United States have more than doubled.[1] Prostate cancer is estimated to account for 29% of all incident cancers in males in the United States. Much of the increase in incidence can be explained by the widespread use of prostate-specific antigen (PSA) for early cancer detection.[92] Clinical prostate cancer is much more common in the United States than in Japan. In both countries, however, a large proportion of elderly men have in situ malignant changes in the prostate. Prostate cancer incidence and mortality have risen more rapidly in African Americans, whose age-adjusted mortality for prostate cancer is more than double that of whites (54.8 versus 23.7 per 100,000 from 1990 to 1996).[1] Benign prostatic hyperplasia, a common condition in men older than 60 years, does not appear to be strongly associated with prostate cancer risk.

Environmental influences have been examined as an explanation for the rising incidence of prostate cancer and the higher frequency of advanced prostate cancer in the United States, even in the Asian-American population. Dietary and hormonal influences have been examined in relation to prostate cancer, but their role is still unclear. Evidence suggests that a slightly increased risk of prostate cancer is associated with increased consumption of saturated fat and red meat; vitamins A and D and β-carotene may decrease the risk.[92]

IATROGENIC CAUSES

In more than 50 patients treated with cyclophosphamide for cancer and other serious diseases, bladder cancer developed after cyclophosphamide-induced cystitis.[93] A second drug, phenacetin, is associated with cancer of the renal pelvis and bladder.[94]

Ionizing radiation can induce all major forms of leukemia except chronic lymphocytic leukemia.[8] Studies have provided evidence for and against the association between electromagnetic radiation and risk for leukemia; current evidence tends not to

support this association.[95] A high incidence of acute myeloid leukemia has been reported in cancer patients treated with alkylating agents (i.e., melphalan, cyclophosphamide, chlorambucil, and the nitrosoureas) or epipodophyllotoxins.[96] In several large studies, the cumulative leukemia risk at 10 years of follow-up was approximately 5%; the leukemogenic effect of these drugs diminishes with longer follow-up. Clinically, the risk of secondary leukemia needs to be weighed against the benefit of treatment. Alternative nonleukemogenic therapies should be sought; antimetabolites, such as cytarabine, fluorouracil, and methotrexate, do not appear to be carcinogens. Patients who undergo bone marrow transplantation are at higher risk for new solid cancers for at least 10 years after transplantation.[97]

MISCELLANEOUS ENVIRONMENTAL CAUSES

Ionizing radiation has been recognized as a human carcinogen for a century. This has been a major etiologic factor for cancer following the atomic bomb in Japan, as well as in various occupational settings. Approximately 3% to 5% of all cancers can be attributed to ionizing radiation, including high-dose therapy for cancer and other diseases.[8] Radiation can induce brain tumors, cancers of the skin and thyroid, sarcomas of bone and soft tissue, and cancers of other sites.[10] Cancer of the skin of the scrotum occasionally develops in workers topically exposed to soot and mineral oils. Chemical exposures and viruses have been studied as etiologic agents in brain tumors, Hodgkin disease, and multiple myeloma, but definitive results are scanty. Although several reports have suggested an association between electromagnetic radiation and leukemia and brain cancer, more definitive studies have demonstrated no evidence of an association.[95] Studies have examined the role of herbicides and defoliants, including Agent Orange, which was heavily used during the war in Vietnam. Agent Orange contains dioxins, which are potent animal carcinogens, although no definitive evidence has been found for increased cancer risk in Vietnam War veterans.[98] Excesses of soft tissue sarcoma and NHL have been reported in workers exposed to dioxins through their manufacture or use in farming for long periods.[8]

Inherited Factors That Predispose to Cancer

Until recently, hereditary factors received little attention in epidemiologic studies of cancer. Interest in hereditary influences emerged with the understanding that genetic alterations underlie the process of transformation of a normal cell into a cancer cell. The basis of cancer is the loss of normal genetic control of cellular processes. At the molecular level, cancer is a disorder of genes, particularly oncogenes and tumor suppressor genes. Among the estimated 50,000 genes in the human genome, only a small fraction seem to be essential for cancer development. Genetic alterations can alter regulation of cell replication, DNA repair, apoptosis, and immune surveillance against tumor cells. Inherited traits can interact with environmental carcinogens. For example, sunlight increases skin cancers in genetically susceptible Celts and other light-skinned populations. Patients with albinism or xeroderma pigmentosum have a defect in excision repair of UV-damaged DNA and develop multiple skin cancers in exposed areas. Epidemiologic studies can help define the influences of both environmental carcinogens and host factors in cancer development.[99]

Hereditary factors in cancer development have been identified in studies of families with a history of cancer. Virtually every form of cancer manifests a tendency to aggregate in families. Close relatives of a cancer patient are at increased risk for the same form of cancer and perhaps other cancers. The excess site-specific cancer risk is usually two to three times above the age- and sex-specific population rate. However, germline (inherited) mutations in some cancer genes can increase the likelihood of cancer development to nearly 100%. These potent cancer genes are rare in the population but serve as important models for studies of carcinogenesis.[100]

HIGH-PENETRANCE GENETIC FACTORS

Risk Factors for Retinoblastoma

A paradigm of hereditary cancers in humans is retinoblastoma. Approximately one third of this childhood cancer occurs in an autosomal dominant pattern with high penetrance. Carriers of a mutated retinoblastoma gene (*Rb-1*) have more than a 10,000-fold excess risk of this eye tumor, as well as a marked increase in risk of second cancers (sarcomas, melanoma, and brain tumors).[101] The *Rb-1* gene is a tumor suppressor. Loss or inactivation of both alleles of the gene abolishes its tumor suppressor function and leads to tumor development. In hereditary retinoblastoma, one abnormal allele has been inherited from a parent. However, this eye tumor develops only after the second normal allele has been inactivated through a somatic (acquired) mutation or other mechanism. Although inherited mutations in the *Rb-1* gene are rare, somatic mutations in the gene are active in the genesis of many forms of cancer, including carcinomas of the breast, the lung, and other sites. The Rb-1 protein interacts with the transcription factors and cyclins that regulate progression through the cell cycle. The *Rb-1* gene can also be inactivated through binding by the transforming proteins of several oncogenic viruses; this protein-protein interaction is one molecular mechanism of viral oncogenesis.

Risk Factors for Familial Neoplasms

An increasing number of genetic diseases are associated with a high risk of cancer [*see Table 2*].[100] Neoplasia is the primary manifestation of some cancer genes, such as the *Rb-1* gene. Carriers of these genes can often be identified by a family history of the same cancer in multiple relatives affected at unusually early ages. In other genetic disorders, such as neurofibromatosis, cancer is a less common manifestation of the underlying genetic disease. The finding of a predisposing genetic disorder should alert the physician to the possibility of early diagnosis of associated cancers through periodic surveillance.

Risk Factors for Colorectal Cancer

Colorectal cancer tends to develop within certain families. In the dominantly inherited disorder adenomatous polyposis coli (APC), colonic polyps arise in adolescence, and the lifetime risk of colon cancer is nearly 100%. The APC tumor-suppressor gene was isolated in 1991 and shown to be the inherited mutation in all polyposis families. In contrast, familial colon cancer without multiple polyposis may be caused by germline mutations in one of the DNA repair genes: *MSH2*, *MLH1*, *MSH6*, *PMS1*, or *PMS2*.[102,103] Their phenotypes include familial colorectal cancer and hereditary nonpolyposis colorectal cancer, which accounts for 5% to 10% of these cancers in the United States.[104]

Risk Factors for Reproductive Cancers

A family history of breast cancer is one of the most consistent risk factors of the neoplasm, particularly in women who have multiple relatives with bilateral premenopausal disease. Heredi-

Table 2 Common Hereditary Cancers and Syndromes Due to Germline Mutations in Predisposing Genes

Gene	*Type*	*Locus*	*Familial Neoplasms and Syndromes*
BRCA1	Tumor suppressor	17q	Hereditary breast and ovarian cancer
BRCA2	Tumor suppressor	13q	Hereditary breast and ovarian cancer
hMSH2	Mismatch repair	2p	Hereditary nonpolyposis colon cancer (also endometrium, brain, urinary tract, other)
hMLH1	Mismatch repair	3p	Hereditary nonpolyposis colon cancer (also endometrium, brain, urinary tract, other)
hPMS1	Mismatch repair	2q	Hereditary nonpolyposis colon cancer (also endometrium, brain, urinary tract, other)
hPMS2	Mismatch repair	7p	Hereditary nonpolyposis colon cancer (also endometrium, brain, urinary tract, other)
NF1	Tumor suppressor	17q	Neurofibromatosis-1 (neurofibroma, sarcoma)
NF2	Tumor suppressor	22q	Neurofibromatosis-2 (acoustic neuroma, brain)
RB1	Tumor suppressor	13q	Hereditary retinoblastoma and other second cancers
APC	Tumor suppressor	5q	Adenomatous polyposis coli (colon, brain)
p53	Tumor suppressor	17p	Li-Fraumeni syndrome (sarcoma, breast, brain, leukemia)
MEN1	Tumor suppressor	11q	Multiple endocrine neoplasia type I (pituitary, parathyroid, pancreas)
MEN2A	Oncogene	10q	Multiple endocrine neoplasia type IIA/B (thyroid, parathyroid, pheochromocytoma)
WT1	Tumor suppressor	11p	Hereditary Wilms tumor
VHL	Tumor suppressor	3p	Von Hippel–Lindau syndrome (hemangioblastoma, renal cell carcinoma)
TSC1	Tumor suppressor	9q	Tuberous sclerosis 2 (kidney, brain)
TSC2	Tumor suppressor	16p	Tuberous sclerosis 2 (kidney, brain)
CDKN2	Tumor suppressor	9p	Hereditary melanoma
ATM	DNA repair	11q	Ataxia-telangiectasia (breast in heterozygote, lymphoma in homozygote)
PTEN	Tumor suppressor	10q	Cowden disease (breast, thyroid, skin)

tary breast cancer accounts for approximately 5% of all breast cancers in the United States. Studies have identified at least five genes predisposing to inherited breast cancer. Clinically, the most important is the *BRCA1* gene, located on chromosome 17q.[105] A second breast cancer gene, *BRCA2*, has been found on chromosome 13q.[106] *BRCA1* and *BRCA2* account for most of the hereditary breast cancers in young women; carriers of *BRCA1* are also predisposed to ovarian cancer of early onset. Women who are carriers of the *BRCA1* mutation have a 50% to 85% probability of developing breast cancer by 70 years of age, and their risk for ovarian cancer is 20% to 50%.[107] Corresponding figures for breast and ovarian cancers associated with inherited *BRCA2* mutations may be lower. The frequency of inherited *BRCA1* and *BRCA2* mutations varies widely among populations; it appears to be higher among Ashkenazi Jews, approximately 2% of whom carry mutations of *BRCA1* or *BRCA2*.[107]

Another susceptibility gene for breast cancer is the *p53* tumor-suppressor gene, the inherited defect in most families with dominantly inherited Li-Fraumeni syndrome. The *PTEN* gene, which is associated with Cowden disease, also confers an increased risk for benign and malignant breast tumors, as well as brain, prostate, and thyroid neoplasms.[108] Breast cancer genes might include the ataxia-telangiectasia gene, on chromosome 11q. Ataxia-telangiectasia is a rare autosomal recessive disease in which homozygotes develop neurologic, neoplastic, and other disorders in childhood.

Prostate cancer has a hereditary component, with male relatives of prostate cancer patients exhibiting a twofold increased risk. Hereditary prostate cancer, which accounts for 5% to 10% of all cases, is primarily associated with early onset disease. Major susceptibility loci for hereditary prostate cancer were recently mapped on chromosome 1 and the X chromosome.[109,110]

Other High-Penetrance Genetic Factors

Inherited susceptibility plays a role in other common forms of human cancers, including endocrine and brain tumors, skin cancer, kidney cancer, melanoma, and the hematologic neoplasms. Familial forms of these cancers account for a small fraction of incident cases. Within affected families, however, the inherited cancer gene is an exceptionally potent oncogenic influence that can lead to cancer among multiple relatives.

LOW-PENETRANCE GENETIC FACTORS

In addition to highly penetrant cancer-susceptibility genes, many low-penetrance genetic variants (polymorphisms) may interact with environmental agents and other genes to modify cancer risk.[18,111] These polymorphisms are associated with only moderate increases in risk, but they can occur at high frequency in the population and contribute to the development of substantial numbers of cancers.[16,17,19] Unfortunately, the effects of these variants are often small and difficult to measure in heterogeneous populations. For example, rare *HRAS1* alleles in the H-*ras*-1 oncogene are reportedly associated with increased breast cancer risk, but new results based on improved analytical methods have failed to support the finding.[111,112] Inconsistent results in studies of low-penetrance cancer genes might also be the result of small sample sizes and misclassification of carcinogenic exposures.[113] Characterization of the role of genetic variants in cancer development can help identify susceptible populations to individualize cancer-prevention efforts.

Polymorphisms in genes that encode for proteins involved in metabolism of steroid hormones or carcinogens might alter cancer risks. Variants in the cytochrome *P-450* genes are associated with increased risk of lung, esophageal, and head and neck cancers. The genes *GSTM1* and *GSTT1* produce glutathione transferases that are involved in the deactivation of tobacco carcinogens.[16] Certain variants in these genes are associated with increased risk of bladder cancer and lower survival rate for patients with lung cancer.[16] A combination of genetic variants in both cytochrome *P-450* genes and *GSTM1* may produce gene-gene interactions that further increase cancer risks.[17] Polymorphisms of the *NAT1* and *NAT2* genes may affect susceptibility to cancers of the urinary bladder, colon and rectum, breast, head and neck, and lung.[18,19] Individuals with the slow *NAT1* and *NAT2* acetylator phenotypes may be prone to bladder cancer, whereas those with fast acetylator phenotypes may be predisposed to colorectal cancer.[19]

Before familial cancers are attributed to genetic susceptibility, chance association and shared exposures to environmental carcinogens should be excluded. Inherited predisposition can be identified through detection of laboratory markers of host susceptibility and by segregation and linkage analysis of the pedigree. With the increasing identification of cancer-susceptibility genes, genetic testing of individuals is becoming more widespread. Analysis of the *Rb-1* gene has been used for a decade to detect carriers in families with retinoblastoma. In affected families, surveillance of newborns for early cancer can reduce loss of vision. As genes for more common cancers have been found, ethical and social issues have become more complex [*see 3:VIII Genetic Diagnosing and Counseling*]. Careful consideration must be given to the costs and benefits of genetic testing.[114]

Screening and Early Detection

The recommendations of the American Cancer Society, the National Cancer Institute, and the United States Preventive Services Task Force for the prevention and early detection of cancer have received wide attention. These guidelines are intended primarily for asymptomatic patients at average risk of cancer. They do not apply to symptomatic patients, who should be managed by the usual standards of medical practice. Patients considered to be at increased risk because of family history or environmental exposures should seek a physician's recommendations for establishing an appropriate early-detection program.

Data are incomplete regarding the costs, risks, and benefits of cancer screening in the general population. A few large randomized studies have been completed for specific cancers, such as breast cancer. These studies have found that periodic mammography reduces breast cancer mortality by 25% to 30% in women 50 to 70 years of age [*see CE:V Adult Preventive Health Care*]. Data are scanty and uncertain for younger and older women. To be useful, a screening test must detect preclinical cancers that are less likely to be lethal after treatment than if they are allowed to progress to clinical detection. False positive results of screening tests lead to unnecessary workups and treatment for indolent tumors and increase medical, psychological, and financial costs.[115] Identifying high-risk populations through risk evaluation, including genetic testing, can help channel scarce resources to susceptible persons.

The American Cancer Society has made specific recommendations for the early detection of asymptomatic carcinoma of the colon and rectum, cervix and other pelvic organs, breast, and prostate [*see Table 3*].[116] However, because of the lack of convincing evidence of benefit from screening for lung cancer, periodic chest x-rays and sputum cytology are not recommended, nor are CT scans. PSA is being increasingly used for prostate cancer detection. Serum PSA levels correlate with the clinical stage of

Table 3 American Cancer Society Recommendations for Early Cancer Detection[113]

Tissue	*Test or Procedure*	*Sex*	*Age (years)*	*Frequency*
Breast	Breast self-examination (BSE)	Female	Early 20s	Should be informed about benefits and limitations and then may choose to do BSE or not
	Breast clinical physical examination	Female	20–39	Every 3 yr
			40	Every year
	Mammography	Female	40	Every year
Colorectal*	Fecal occult blood test	Male, female	50	Every year
	and/or			
	Flexible sigmoidoscopy	Male, female	50	Every 5 yr
	or Colonoscopy	Male, female	50	Every 10 yr
	or Double-contrast barium enema	Male, female	50	Every 10 yr
Cervix	Papanicolaou smear Pelvic examination	Female	Papanicolaou smear and pelvic examination every year for women who are or have been sexually active for 3 yr or have reached 21 yr of age; after 30 yr of age, after three or more consecutive satisfactory normal annual exams, the Pap test may be performed every 2 to 3 yr at the discretion of the physician	
Endometrium	Endometrial tissue sample	Female	Women at very high risk beginning at 35 yr of age	Should be informed about potential benefits and limitations of screening
Prostate	PSA/DRE	Male	50	Every year Annual PSA/DRE testing should be limited to men with life expectancy of at least 10 yr and men at high risk (those of African descent or with familial risk should initiate testing at 45 yr of age); the benefits and limitations should be discussed with the clinician, but if the clinician is asked to make the decision, the ACS recommends testing
Skin	Skin	Male	20–40	Every 3 yr
		Female	> 40	Every year
General	Health counseling and cancer checkup	Male	20–40	Every 3 yr
		Female	> 40	Every year

*These guidelines are intended for individuals at standard risk. Patients with a family history or other factors that indicate elevated risk should consult their physicians for recommended screening guidelines.

DRE—digital rectal examination PSA—prostate-specific antigen

prostate cancer and the volume of the cancer in the gland[117]; evidence of its efficacy is minimal, and the National Cancer Institute currently recommends informed decision making between the physician and the patient. A large randomized trial, the Prostate, Lung, Colorectal, Ovary (PLCO) trial, is currently in progress to assess screening tests for each of these cancers.[118]

The current screening guidelines can be expected to evolve with the accrual of knowledge and technological advances. For example, the technical quality of mammograms and their interpretation are critical factors in their usefulness. Mammographic detection of cancer is more problematic in the dense breast tissue of younger women, whose tumor growth rates are generally more rapid. After analyzing the same body of evidence from randomized mammographic studies, expert committees have reached different conclusions regarding routine mammographic screening of asymptomatic women 40 to 49 years of age. The American Cancer Society, the National Cancer Institute, and other groups have recommended that women begin receiving annual mammograms once they reach 40 years of age, but the National Institutes of Health consensus panel did not make this recommendation.[119] Of several methods available for colorectal cancer screening, colonoscopy was found to be superior to both double-contrast barium enema and sigmoidoscopy in two studies.[120,121] New screening methods, such as magnetic resonance imaging of the breast and spiral computed tomography of the lungs, may further enhance the sensitivity and specificity of screening tests.[122]

Clinicians can help prevent and detect cancer at early stages. They can prevent environmental cancers by counseling patients to avoid tobacco use, asbestos exposure, and unnecessary exposure to ionizing radiation. A brief medical and family history can reveal an unusual predisposition to cancer and a need for closer medical surveillance. In the course of the physical examination, attention to signs of early cancer can lead to curative treatment.

Alfred I. Neugut, M.D., Ph.D., has served as an advisor for Pfizer, Inc.

Frederick P. Li, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Jemal A, Tiwari RC, Murray T, et al: Cancer statistics, 2004. CA Cancer J Clin 54:8, 2004

2. Weir HK, Thun MJ, Hankey BF, et al: Annual report to the nation on the status of cancer, 1975-2000, featuring the uses of surveillance data for cancer prevention and control. J Natl Cancer Inst 95:1276, 2003

3. Boyle P, d'Onofrio A, Maisonneuve P, et al: Measuring progress against cancer in Europe: has the 15% decline targeted for 2000 come about? Ann Oncol 14:1312, 2003

4. Fisher B, Costantino JP, Wickerham DL, et al: Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 90:1371, 1998

5. Benner SE, Pajak TF, Lippman SM, et al: Prevention of second primary tumors with isotretinoin in patients with squamous cell carcinoma of the head and neck: long-term follow-up. J Natl Cancer Inst 86:140, 1994

6. Hawk ET, Limburg PJ, Viner JL: Epidemiology and prevention of colorectal cancer. Surg Clin N Am 82:905, 2002

7. Doll R: Epidemiological evidence of the effects of behaviour and the environment on the risk of human cancer. Recent Results Cancer Res 154:3, 1998

8. Cancer Epidemiology and Prevention, 2nd ed. Schottenfeld D, Fraumeni JF, Eds. Oxford University Press, New York, 1996

9. Kelsey J: Methods in Observational Epidemiology. Oxford University Press, New York, 1996

10. Boice JD Jr, Lubin JH: Occupational and environmental radiation and cancer. Cancer Causes Control 8:309, 1997

11. Mueller N: Overview: viral agents and cancer. Environ Health Perspect 103:259, 1995

12. Goedert JJ: The epidemiology of acquired immunodeficiency syndrome malignancies. Semin Oncol 27:390, 2000

13. Peek RM Jr, Blaser MJ: *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2:28, 2002

14. Kahl DS: Update: gastric MALT lymphoma. Curr Opinion Oncol 15:347, 2003

15. Rothman N, Stewart WF, Schulte PA: Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev 4:301, 1995

16. Rebbeck TR: Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. Cancer Epidemiol Biomarkers Prev 6:733, 1997

17. Bartsch H, Nair U, Risch A, et al: Genetic polymorphism of CYP genes, alone or in combination, as a risk modifier of tobacco-related cancers. Cancer Epidemiol Biomarkers Prev 9:3, 2000

18. Dunning AM, Healey CS, Pharoah PD, et al: A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 8:843, 1999

19. Hein DW, Doll MA, Fretland AJ, et al: Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms. Cancer Epidemiol Biomarkers Prev 9:29, 2000

20. Hankinson SE, Willett WC, Colditz GA, et al: Circulating concentrations of insulin-like growth factor-1 and risk of breast cancer. Lancet 351:1393, 1998

21. Ma J, Pollak MN, Giovannucci E, et al: Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-1 and IGF-binding protein-3. J Natl Cancer Inst 91:620, 1999

22. Chan JM, Stampfer MJ, Giovannucci E, et al: Plasma insulin-like growth factor-1 and prostate cancer risk: a prospective study. Science 279:563, 1998

23. Giovannucci E, Pollak MN, Platz EA, et al: A prospective study of plasma insulin-like growth factor-1 and binding protein-3 and risk of colorectal neoplasia in women. Cancer Epidemiol Biomarkers Prev 9:345, 2000

24. Grimberg A, Cohen P: Role of insulin-like growth factors and their binding proteins in growth control and carcinogenesis. J Cell Physiol 183:1, 2000

25. Brundtland GH: Achieving worldwide tobacco control. JAMA 284:750, 2000

26. Pisani P, Parkin DM, Bray F, et al: Estimates of the worldwide mortality from 25 cancers in 1990. Int J Cancer 83:18, 1999

27. Harvard report on cancer prevention. Vol. 1: Causes of human cancer. Cancer Causes Control 7:S5, 1996

28. Smoking and Health: A Report of the Surgeon General. U.S. Department of Health Education and Welfare. Government Printing Office, Washington, DC, 1979

29. Rigotti NA, Lee JE, Wechsler H: U.S. college students' use of tobacco products: results of a national survey. JAMA 284:699, 2000

30. Bauer UE, Johnson TM, Hopkins RS, et al: Changes in youth cigarette use and intentions following implementation of a tobacco control program: findings from the Florida Youth Tobacco Survey, 1999-2000. JAMA 284:723, 2000

31. Cinciripini PM, Hecht SS, Henningfield JE, et al: Tobacco addiction: implications for treatment and cancer prevention. J Natl Cancer Inst 89:1852, 1997

32. Baker F, Ainsworth SR, Dye JT, et al: Health risks associated with cigar smoking. JAMA 284:735, 2000

33. Dirkle JL, Flegal KM, Bernert JT, et al: Exposure of the U.S. population to environmental tobacco smoke: the Third National Health and Nutrition Examination Survey, 1998-1991. JAMA 275:1233, 1996

34. Kelsey JL, Moore AS, Glickman LT: Epidemiologic studies of risk factors for cancers in pet dogs. Epidemiol Rev 20:204, 1998

35. Engel LS, Chow WH, Vaughan TL, et al: Population attributable risks of esophageal and gastric cancers. J Natl Cancer Inst 95:1404, 2003

36. Koch TR: The changing face of esophageal malignancy. Current Gastroenterol Rpts 5:187, 2003

37. Blot WJ, McLaughlin JK: The changing epidemiology of esophageal cancer. Semin Oncol 26:2, 1999

38. Castellsague X, Quintana MJ, Martinez MC, et al: The role of type of tobacco and type of alcoholic beverage in oral carcinogenesis. Int J Cancer 108:741, 2004

39. Mashberg A, Boffetta P, Winkelman R, et al: Tobacco smoking, alcohol drinking, and cancer of the oral cavity and oropharynx among U.S. veterans. Cancer 72:1369, 1993

40. Slattery ML, Potter JD, Friedman GD, et al: Tobacco use and colon cancer. Int J Cancer 70:259, 1997

41. Heineman EF, Zahm SH, McLaughlin JK, et al: Increased risk of colorectal cancer among smokers: results of a 26-year follow-up of U.S. veterans and a review. Int J Cancer 59:728, 1994

42. Giovannucci E, Colditz GA, Stampfer MJ, et al: A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. women. J Natl Cancer Inst 86:192, 1994

43. Terry PD, Rohan TE: Cigarette smoking and the risk of breast cancer in women: a review of the literature. Cancer Epidemiol Biomarkers Prev 11:953, 2002

44. Lantz PM, Jacobson PD, Warner KE, et al: Investing in youth tobacco control: a review of smoking prevention and control strategies. Tob Control 9:47, 2000

45. Tobacco use: United States, 1900-1999. MMWR Morb Mortal Wkly Rep 48:986, 1999

46. Charatan F: Florida jury finds tobacco companies guilty of fraud. BMJ 319:143, 1999

47. Kessler DA, Barnett PS, Witt A, et al: The legal and scientific basis for FDA's assertion of jurisdiction over cigarettes and smokeless tobacco. JAMA 277:405, 1997

48. Gottleib S: Supreme Court rules that FDA cannot regulate tobacco industry. BMJ 320:894, 2000

49. Howson CP, Hiyama T, Wynder EL: The decline in gastric cancer: epidemiology of an unplanned triumph. Epidemiol Rev 8:1, 1986

50. Shimizu H, Mack TM, Ross RK, et al: Cancer of the gastrointestinal tract among Japanese and white immigrants in Los Angeles County. J Natl Cancer Inst 78:223, 1987

51. Willett WC, Stampfer MJ, Colditz GA, et al: Relation of meat, fat and fiber intake to the risk of colon cancer in a prospective study among women. N Engl J Med 323:1664, 1990

52. Tomeo CA, Colditz GA, Willett WC, et al: Harvard report on cancer prevention, Vol 3. Prevention of colon cancer in the United States. Cancer Causes Control 10:167, 1999

53. Howe GR, Aronson KJ, Benito E, et al: The relationship between dietary fat intake and risk of colorectal cancer: evidence from the combined analysis of 13 case-control studies. Cancer Causes Control 8:215, 1997

54. La Vecchia C, Negri E, Pelucchi C, et al: Dietary folate and colorectal cancer. Int J Cancer 102:545, 2002

55. Choi SW, Mason JB: Folate status: effects on pathways for colorectal carcinogenesis. J Nutr 132 (8 suppl): 2413S, 2002

56. Schatzkin A, Lanza E, Corle D, et al: Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. Polyp Prevention Trial Study Group. N Engl J Med 342:1149, 2000

57. Fuchs CS, Giovannucci EL, Colditz GA, et al: Dietary fiber and the risk of colorectal cancer and adenoma in women. N Engl J Med 340:169, 1999

58. Alberts DS, Martinez ME, Roe DJ, et al: Lack of effects of a high-fiber cereal supplement on the recurrence of colorectal adenomas. Phoenix Colon Cancer Physicians Network. N Engl J Med 342:1156, 2000

59. Food, nutrition and the prevention of cancer: a global perspective. American Institute for Cancer Research/ World Cancer Research Fund, 1996

60. The effect of Vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. N Engl J Med 330:1029, 1994

61. Virtamo J, Pietinen P, Huttunen JK, et al: Incidence of cancer and mortality following alpha-tocopherol and beta carotene supplementation: a postintervention follow-up. JAMA 290:476, 2003

62. Vainio H: Chemoprevention of cancer: lessons to be learned from beta-carotene trials. Toxicol Lett 112:513, 2000

63. Calle EE, Rodriguez C, Walter-Thurmond K, et al: Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348:1625, 2003

64. Mao Y, Pan S, Wen SW, et al: Physical inactivity, energy intake, obesity and the risk of rectal cancer in Canada. Int J Cancer 105:831, 2003

65. John EM, Horn-Ross PL, Koo J: Lifetime physical activity and breast cancer risk in a multiethnic population: the San Francisco Bay area breast cancer study. Cancer Epidemiol Biomarkers Prev 12:1143, 2003

66. Lee IM: Physical activity and cancer prevention: data from epidemiologic studies. Med Sci Sports Exerc 35:1823, 2003

67. El-Serag HB, Mason AC: Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 340:745, 1999

68. Chang MH: Decreasing incidence of hepatocellular carcinoma among children following universal hepatitis B immunization. Liver Int 23:309, 2003

69. zur Hausen H: Papillomaviruses in human cancers. Proc Assoc Am Physicians 111:581, 1999

70. Ho GY, Burk RD, Klein S, et al: Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst 87:1365, 1995

71. Franco EL: Cancer causes revisited: human papillomavirus and cervical neoplasia. J Natl Cancer Inst 87:779, 1995

72. Wright TC Jr, Schiffman M, Solomon D, et al: Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening. Obstet Gynecol 103:304, 2004

73. Goldie SJ, Kuhn L, Denny L, et al: Policy analysis of cervical cancer screening strategies in low-resource settings: clinical benefits and cost effectiveness. JAMA 285:3107, 2001

74. Koutsky LA, Ault KA, Wheeler CM, et al: A controlled trial of a human papillomavirus type 16 vaccine. N Engl J Med 347:1645, 2002

75. Moore PS, Chang Y: Kaposi's sarcoma (KS), KS-associated herpes virus, and the criteria for causality in the age of molecular biology. Am J Epidemiol 147:217, 1998

76. Yasunaga J, Matsuoka M: Leukemogenesis of adult T-cell leukemia. Int J Hematol 78:312, 2003

77. Mueller N: Overview of the epidemiology of malignancy in immune deficiency. J Acquir Immune Defic Syndr 21:S5, 1999

78. Wong BC, Lam SK, Wong WM, et al: *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. JAMA 291:187, 2004

79. Mossman BT, Kamp DW, Weitzman SA: Mechanisms of carcinogenesis and clinical features of asbestos-associated cancers. Cancer Invest 14:466, 1996

80. Pohlabeln H, Jockel KH, Bruske-Hohlfeld I, et al: Lung cancer and exposure to man-made vitreous fibers: results from a pooled case-control study in Germany. Am J Ind Med 37:469, 2000

81. Samet JM: Diseases of uranium miners and other underground miners exposed to radon. Occup Med 6:629, 1991

82. Cancer Incidence in Five Continents. Parkin D, Whelan S, Ferlay J, et al, Eds. International Agency for Research on Cancer, Lyons, France, 1997

83. Brinton LA, Daling JR, Liff JM, et al: Oral contraceptives and breast cancer risk among younger women. J Natl Cancer Inst 87:827, 1995

84. Schairer C, Lubin J, Troisi R, et al: Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk. JAMA 283:485, 2000

85. Thune I, Brenn T, Lund E, et al: Physical activity and the risk of breast cancer. N Engl J Med 336:1269, 1997

86. Rockhill B, Willett WC, Hunter DJ, et al: Physical activity and breast cancer risk in a cohort of young women. J Natl Cancer Inst 90:1155, 1998

87. Garland M, Hunter DJ, Colditz GA, et al: Alcohol consumption in relation to breast cancer risk in a cohort of United States women 25-42 years of age. Cancer Epidemiol Biomarkers Prev 8:1017, 1999

88. Hunter DJ, Spiegelman D, Adami HO, et al: Cohort studies of fat intake and the risk of breast cancer: a pooled analysis. N Engl J Med 334:356, 1996

89. Trentham-Dietz A, Newcomb PA, Storer BE, et al: Body size and risk of breast cancer. Am J Epidemiol 145:1011, 1997

90. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 350:1047, 1997

91. Delmas PD, Bjarnason NH, Mitlak BH, et al: Effects of raloxifene on bone mineral density, serum cholesterol concentrations, and uterine endometrium in postmenopausal women. N Engl J Med 337:1641, 1997

92. Haas G P, Sakr WA: Epidemiology of prostate cancer. CA Cancer J Clin 47:273, 1997

93. Ross RK, Jones PA, Yu MC: Bladder cancer epidemiology and pathogenesis. Semin Oncol 23:536, 1996

94. Piper JM, Tonascia J, Matanoski GM: Heavy phenacetin use and bladder cancer in women aged 20 to 49 years. N Engl J Med 313:292, 1985

95. Miller RD, Neuberger JS, Gerald KB: Brain cancer and leukemia and exposure to power-frequency (50- to 60-Hz) electric and magnetic fields. Epidemiol Rev 19:273, 1997

96. Pui CH, Ribeiro RC, Hancock ML, et al: Acute myeloid leukemia in children treated with epipodophyllotoxins for acute lymphoblastic leukemia. N Engl J Med 325:1682, 1991

97. Curtis R, Rowlings P, Deeg H, et al: Solid cancers after bone marrow transplantation. N Engl J Med 336:897, 1997

98. Ketchum NS, Michalek JE, Burton JE: Serum dioxin and cancer in veterans of Operation Ranch Hand. Am J Epidemiol 149:630, 1999

99. Perera FP: Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention. J Natl Cancer Inst 88:496, 1996

100. Li FP: Cancer control in susceptible groups. J Clin Oncol 17:719, 1999

101. Wong F, Boice J, Abramson D, et al: Cancer incidence after retinoblastoma: radiation dose and sarcoma risk. JAMA 278:1262, 1997

102. Marra G, Boland CR: Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. J Natl Cancer Inst 87:1114, 1995

103. Kolodner RD, Tytell JD, Schmeits JL, et al: Germ-line MSH6 mutations in colorectal cancer families. Cancer Res 59:5068, 1999

104. Lynch HT, Smyrk TC, Watson P: Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 104:1535, 1993

105. Miki Y, Swensen J, Shattuck-Edens D, et al: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66, 1994

106. Rebbeck TR, Couch FJ, Kant J, et al: Genetic heterogeneity in hereditary breast cancer: role of BRCA1 and BRCA2. Am J Hum Genet 59:547, 1996

107. Struewing J, Hartge P, Wacholder S, et al: The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336:1401, 1997

108. Liaw D, Marsh D, Li J, et al: Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nature Genet 16:64, 1997

109. Smith JR, Freije D, Carpten JD, et al: Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371, 1996

110. Xu J, Meyers D, Freije D, et al: Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175, 1998

111. Firgaira FA, Seshadri R, McEvoy CR, et al: HRAS1 rare minisatellite alleles and breast cancer in Australian women under age forty years. J Natl Cancer Inst 91:2107, 1999

112. Krontiris TG, Devlin B, Karp DD, et al: An association between the risk of cancer and mutations in the HRAS1 minisatellite locus. N Engl J Med 329:517, 1993

113. Garcia-Closas M, Rothman N, Lubin J: Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. Cancer Epidemiol Biomarkers Prev 8:1043, 1999

114. Ponder B: Genetic testing for cancer risk. Science 278:1050, 1997

115. Smith RA: Screening fundamentals. J Natl Cancer Inst Monogr 22:15, 1997

116. Smith RA, Cokkinides V, Eyre HJ: American Cancer Society guidelines for the early detection of cancer, 2004. CA Cancer J Clin 54:41, 2004

117. Brawer MK: Screening for prostate cancer. Semin Surg Oncol 18:29, 2000

118. Gohagan JK, Prorok BC, Hayes RB, et al: The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. Control Clin Trials 21(6 suppl): 251S, 2000

119. National Institutes of Health Consensus Conference on Breast Cancer Screening for Women Ages 40-49. Proceedings. Bethesda, Maryland. January 21-23, 1997. J Natl Cancer Inst Monogr 22:1, 1997

120. Winawer SJ, Stewart ET, Zauber AC, et al: A comparison of colonoscopy and double-contrast barium enema for surveillance after polypectomy. National Polyp Study Work Group. N Engl J Med 342:1766, 2000

121. Lieberman DA, Weiss DG, Bond JH, et al: Use of colonoscopy to screen asymptomatic adults for colorectal cancer. N Engl J Med 343:162, 2000

122. Henschke CI, McCauley DI, Yankelevitz DF, et al: Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet 354:99, 1999

II MOLECULAR GENETICS OF CANCER

Daniel A. Haber, m.d., ph.d.

Cancer as a Genetic Disease

The uncontrolled clonal expansion of a cell, which often leads to invasion of surrounding tissues and metastatic spread, produces cancer. The initial genetic alterations in this cell that trigger its aberrant proliferation are followed by the accumulation of additional mutations among its progeny. Finally, a selection process occurs by which subclones with enhanced growth properties become dominant within the tumor—so-called tumor progression. A clear histologic and molecular genetic evolution from precancerous lesions to frankly malignant and invasive cancer has been defined for some tumors (e.g., colon and bladder cancers); in many cancers, however, this process may not be clinically evident. The genes that are altered in the development and progression of cancer are still being intensively investigated. In many cases, identification of these genes has led to major insight into the normal physiologic mechanisms that control cell proliferation. The products of these genes are engaged in basic cell cycle regulation, transmission of growth signals, regulation of cellular differentiation and cell death, and establishment of cellular immortality. Disruption of these genes typically occurs in somatic cells destined to become cancerous. However, in rare cases, mutations may be passed on in the germline, resulting in genetic predisposition to cancer (i.e., familial cancer syndromes).

Environmental factors are also thought to contribute to the development of cancer. Interactions between environmental factors and subtle genetic variations that distinguish individuals may conceivably constitute an important determinant of cancer risk within the general population.[1] In some cases, there is a direct link between the effect of the carcinogen on DNA and specific mutations in the *p53* tumor suppressor gene (e.g., exposure to tobacco or to the fungal toxin aflatoxin). In other cases, more indirect effects may increase the risk of developing cancer (e.g., exposure to such potent carcinogens as asbestos or more subtle exposure, such as occurs with prolonged treatment with estrogen). Exposure to ionizing radiation is another cause of cancer. The incidence of leukemias and various solid tumors in atomic bomb survivors provided data for classic studies of radiation dosage and its consequences. More recent data derive from cases of late complications of radiation therapy or cases involving the administration of radiomimetic chemotherapy drugs to patients as treatment for an initial malignancy.

Finally, viral infection has been linked to the development of specific cancers. Cancer-virus linkages include cervical carcinoma with specific subtypes of human papillomavirus; hepatocellular carcinoma with chronic hepatitis B virus infection; nasopharyngeal carcinoma and lymphomas with Epstein-Barr virus in immunosuppressed hosts; the rare acute T cell leukemia with acute transforming human T cell lymphotrophic virus type I; and, potentially, Kaposi sarcoma with human herpesvirus 8. Such cases, it should be noted, represent exceptions to the rule; most human cancer does not result from viral infection. However, much of our understanding about the human genes involved in cancer was originally derived from the study of viruses that cause tumors in chickens and rodents. The misappropriation by these viruses of cellular genes involved in cellular proliferation led to the identification of oncogenes and provided the first clue to the genetic events that cause cancer in humans.

Oncogenes and Proto-oncogenes

ACTIVATING OR GAIN-OF-FUNCTION MUTATIONS

Cancer-causing genes, or oncogenes, were discovered when researchers found that specific genes of chicken and rodent retroviruses can transform normal mammalian cells in culture. These viral transforming genes proved to be activated homologues of mammalian genes (proto-oncogenes) that had been stolen from the host cell during viral evolution for their ability to stimulate cellular proliferation.[2] Primary human cancers, although not caused by viral infection, were found to harbor similar activated alleles of proto-oncogenes. Among the first oncogenes discovered were those that encoded proteins directly involved in the transmission of cellular proliferation signals. These include receptors for growth factors that are constitutively activated, as though responding to the continuous presence of a growth factor (e.g., platelet-derived growth factor receptor or epidermal growth factor receptor), and downstream signaling molecules (i.e., molecules that are activated in response to growth factor signaling) that normally shuttle between an on and off state but are mutated in cancer cells to a permanently on position (e.g., H-*ras*, K-*ras*, and N-*ras*). The mechanisms by which these normal cellular proto-oncogenes are activated in human cancers include point mutations, gene amplification, and chromosomal translocations [*see Table 1 and Figure 1*]. Because they result in novel or altered functional properties for the encoded protein and are genetically dominant over the second normal allele, these mutations are commonly referred to as gain-of-function mutations.

Point Mutations

Point mutations that are capable of activating a gene product are unusual and reflect specific functional properties of the protein itself. For instance, only specific alterations in three codons of ras family proteins lead to constitutive activation of ras signaling.[3] Presumably, other mutations are functionally silent or lead to inactive proteins and thus are not selected during malignant transformation. Similarly, specific amino acid changes in growth factor receptors or other signaling molecules confer an active, growth-inducing effect or prevent downregulation by appropriate physiologic signals. Of particular interest are mutations in the *ret* gene, which codes for a growth factor receptor and is the only proto-oncogene found to be mutated in the germline of patients predisposed to cancer. Amino acid substitutions within distinct functional domains of the protein cause multiple endocrine neoplasia type IIA or type IIB or medullary thyroid cancer, whereas inactivating mutations cause Hirschsprung disease, a developmental defect affecting colonic innervation.[4] This striking link between different mutations in the *ret* gene and distinct cancer and developmental syndromes presumably results from different functional properties that are mediated by the various domains of the ret protein, as well as from different ret-associated pathways that may be activated in different tissue types.

Table 1 Selected Oncogene Mutations in Human Cancer[95]

Gene	*Activation Mechanism*	*Protein Properties*	*Cancer Types*	*Germline Mutations*
K-*ras*	Point mutation	p21 GTPase	Pancreatic, colorectal, lung (adenocarcinoma), endometrial, other carcinomas	ND
N-*ras*	Point mutation	p21 GTPase	Myeloid leukemia	ND
H-*ras*	Point mutation	p21 GTPase	Bladder	ND
EGFR (*erb B)*	Amplification	Growth factor (EGF) receptor	Gliomas, squamous and other carcinomas	ND
NEU (erb B2)	Amplification	Growth factor receptor	Breast, ovarian, gastric, other carcinomas	ND
c-*myc*	Chromosome translocation, amplification	Transcription factor	Burkitt lymphoma, SCLC, other carcinomas	ND
N-*myc*	Amplification	Transcription factor	Neuroblastoma, SCLC	ND
L-*myc*	Amplification	Transcription factor	SCLC	ND
bcl-2	Chromosome translocation	Antiapoptosis protein	B cell lymphoma (follicular type)	ND
CYCD1	Amplification, chromosome translocation	Cyclin D, cell cycle control	Breast and other carcinomas, B cell lymphoma, parathyroid adenomas	ND
bcr-abl	Chromosome translocation	Chimeric nonreceptor tyrosine kinase	CML, ALL (T cell)	ND
ret	Chromosome translocation, point mutation	GNDF-receptor tyrosine kinase	Thyroid (papillary type) Thyroid (medullary type)	Yes (MEN2)
CDK4	Amplification, point mutation	Cyclin-dependent kinase	Sarcoma	Yes (familial melanoma)
MET	Point mutation	HGF-receptor tyrosine kinase		Yes (hereditary renal cancer, papillary type)
SMO	Point mutations	Transmembrane signaling molecule in sonic hedgehog pathway	Basal cell skin	ND
β-CAT	Point mutation, in-frame deletion	Transcriptional coactivator, links E-cadherin to cytoskeleton	Melanoma, colorectal	ND
HST	Amplification	Growth factor (FGF-like)	Gastric	ND
PML-RAR-α	Chromosome translocation	Chimeric transcription factor	APL	ND
W2A-PBX1	Chromosome translocation	Chimeric transcription factor	Pre-B ALL	ND
MDM-2	Amplification	p53-binding protein	Sarcoma	ND
gli	Amplification	Transcription factor	Sarcoma, glioma	ND
TTG	Chromosome translocation	Transcription factor	T cell ALL	ND

ALL—acute lymphocytic leukemia APL—acute promyelocytic leukemia CML—chronic myelogenous leukemia EGF—epidermal growth factor FGF—fibroblast growth factor GDNF—glial cell line–derived neurotrophic factor GTPase—guanine trinucleotide phosphatase HGF—hepatocyte growth factor ND—not determined SCLC—small cell carcinoma of the lung

Gene Amplification and Chromosomal Translocations

In addition to amino acid substitutions resulting in an activated gene product, proto-oncogenes may be converted to their oncogenic forms by gross chromosomal alterations. Gene amplification—the overreplication of large chromosomal fragments—is a common mechanism for aberrant overexpression of specific genes in tumors.[5] Amplified genes, which occasionally may number as many as hundreds of copies of normally diploid genes, may reside on small extrachromosomal elements (double minute chromosomes) or may be integrated within a single chromosomal location, detectable as a large homogeneously staining region. In cancer cells, the loss of cellular genes that normally contribute to the maintenance of genomic integrity (e.g., *p53*) may facilitate aberrant chromosomal events, including gene amplification, and thus the accumulation of additional genetic lesions. Among the best-known oncogenes that are amplified in cancer cells is N-*myc*. This gene codes for a transcription factor that plays a physiologic role in stimulating cellular proliferation and is commonly amplified in the pediatric cancer neuroblastoma; patients have a poor clinical prognosis.[6] Overexpression of a proto-oncogene may also result from a chromosomal translocation or rearrangement, which removes the cellular gene from its physiologically regulated promoter and brings it under the control of a foreign and more active promoter.[7] In Burkitt lymphoma, for instance, the chromosomal locus containing the c-*myc* gene is rearranged such that the upstream (i.e., a region that is located to the 5′ side of a gene) negative regulatory regions of c-*myc* are lost; expression of the gene is directed by the strong immunoglobulin heavy-chain enhancer, which is constitutively active in B lymphocytes.[8,9] Deregulation of c-*myc* expression in these cells is thus a potent force driving cellular proliferation.

The first specific chromosomal translocation identified in human cancer was the Philadelphia chromosome, which underlies chronic myeloid leukemia (CML). The fusion of chromosomes 9 and 22 leads to the joining of two unrelated genes, the c-*abl* gene, which encodes a tyrosine kinase and is located on chromosome 9, and the gene *bcr* (for breakpoint recombination), located on chromosome 22.[10] A chimeric protein with novel transforming properties is formed from this specific chromosomal rearrangement. Uncommon variants of CML, in which a Philadelphia chromosome is not evident by cytogenetic analysis, may have complex genetic rearrangements, resulting in a *bcr-abl* translocation that can be detected only with molecular markers. The accelerated or blast phase of CML is often associated with duplication of the Philadelphia chromosome, suggesting that increased copies of this aberrant gene confer a dose-dependent transforming effect. The *bcr-abl* fusion also demonstrates a striking cell-type specificity: the classic chromosomal breakpoint is associated with myeloid proliferation, whereas a variant breakpoint, resulting in a subtle alteration in the chimeric protein, leads to a pediatric lymphoid leukemia that has a poor prognosis.[11] The importance

of the *bcr-abl* translocation as the initiating genetic event in CML is supported by a mouse model, in which transgenic expression of a bcr-abl chimeric protein in hematopoietic cells is sufficient to trigger both myeloid and lymphoid proliferation.[12] The recent discovery of an effective inhibitor of the bcr-abl kinase, imatinib mesylate (formerly STI571), has led to dramatic responses in CML and has revolutionized treatment of this leukemia[13] [*see* Rational Drug Design and Small Molecules, *below*].

Chromosomal translocations that generate novel chimeric proteins have been linked to other types of leukemia.[14] In acute promyelocytic leukemia (APL), a chromosomal rearrangement joins a novel gene, *PML*, to the gene for the retinoic acid receptor–α (*RAR*-α). Although the precise functional properties of the chimeric molecule, PML-RAR-α, are unknown, this translocation underlies the dramatic responsiveness of this leukemia to treatment with all-*trans*-retinoic acid, which has revolutionized the clinical care of patients with APL.[15] A novel chromosomal translocation product, TEL-AML1, that underlies a common and treatment-responsive form of pediatric acute lymphoblastic leukemia has also been identified.[16]

The ability to grow leukemic cells in culture long enough to allow cytogenetic analysis has facilitated the characterization of chromosomal translocations in leukemias. However, specific chromosomal translocations have also been observed in solid tumors. Among the most noteworthy are Ewing sarcoma and the family of peripheral neuroepithelial tumors (PNETs). These once disparate tumors are now defined by a chromosomal translocation fusing the *EWS* gene to a number of transcription factors

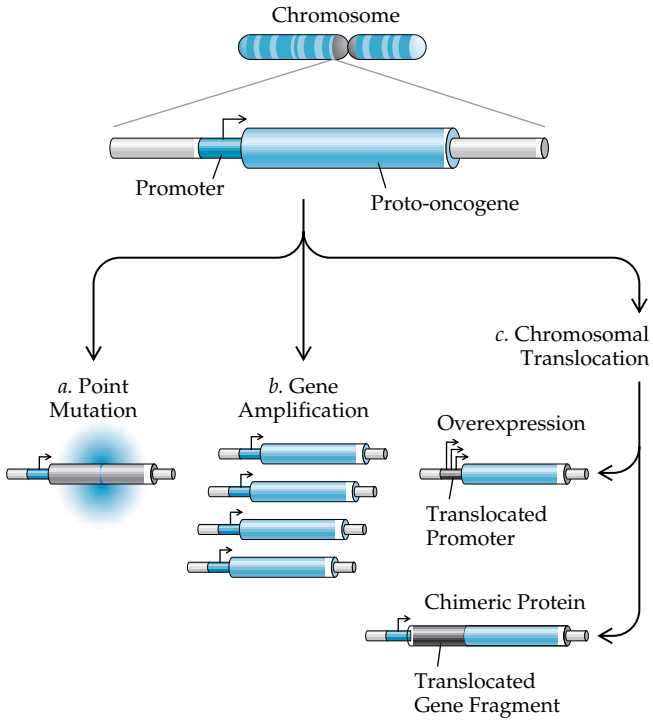


Figure 1 **Mechanisms of activation of proto-oncogenes. Cellular genes may be activated in cancer as a result of (*a*) point mutations that alter the amino acid sequence, encoding a protein that may be constitutively activated; (*b*) amplification of the cellular gene, resulting in higher levels of protein expression; or (*c*) chromosomal translocations that lead to the juxtaposition of a strong promoter and cause increased protein expression or that produce a novel chimeric protein, derived from gene fragments normally present on different chromosomes.**

of the *ETS* gene family (the most common chimeric protein is EWS-FLI1).[17] This chimeric product presumably acts directly on target promoters to direct the expression of genes that induce cellular proliferation. Identification of the *EWS* translocations allowed the molecular grouping of a class of tumors whose proliferation is driven by similar genetic alterations and that respond to similar chemotherapeutic regimens.

IN VITRO AND MOUSE MODELS OF ONCOGENE FUNCTION

The detection of gain-of-function mutations in proto-oncogenes provides strong evidence that these genetic alterations contribute to malignant transformation. However, the timing of these mutations during the initiation and progression of cancer may be variable. With the single exception of c-*ret*, mutations in proto-oncogenes are not observed in the germline of patients with familial cancer syndromes, presumably because they are likely to have adverse consequences on normal development. Chromosomal translocations appear to constitute the initiating event for malignant transformation in some leukemias and solid tumors; in other cancers, mutations in proto-oncogenes are thought to contribute to the progression of cancer once it has developed. Two model systems, transgenic mice and primary cell culture, have proved to be of great value in investigating the functional properties of oncogenes. Transgenic mice are generated by the injection of recombinant DNA plasmids into a fertilized egg.[18] These plasmids may contain an oncogene whose expression is directed by a tissue-specific promoter. A fraction of the mouse progeny will demonstrate the predicted expression of the transgenic construct. Although expression of some oncogenes in appropriate tissues may trigger malignant growth, in other cases, cancer develops only after the mating of mice expressing different oncogenes. The simultaneous expression of two different oncogenes may induce malignant growth.[19] The function of dominant oncogenes is also being tested via transformation of primary rodent cells in vitro. Whereas established cell lines often require little stimulus to become tumorigenic, primary cultured cells, such as those derived from baby rat kidneys or mouse embryo fibroblasts, are transformed only after the introduction of combinations of potent oncogenes.[20] The combination of c-*myc* and activated H-*ras*, for instance, is effective in transforming primary cells, but neither oncogene is sufficient by itself.

In addition to their role in validating the transforming potential of activated proto-oncogenes, these functional assays in the study of DNA tumor viruses have provided important insight into the link between oncogenes and another class of cancer genes, tumor suppressors. Unlike retroviruses that have activated versions of cellular proto-oncogenes incorporated into their own genomes, DNA tumor viruses, such as adenovirus, SV40, and papillomavirus, encode original viral proteins capable of transforming rodent, monkey, or human cells. Introduction of SV40 T antigen into transgenic mice leads to the development of tumors in virtually any tissue to which its expression is directed. Similarly, the combination of adenoviral proteins E1A and E1B has transforming potential, as shown in primary cell transformation assays. A major discovery regarding the etiology of cancer was the finding that the transforming genes derived from these three unrelated DNA tumor viruses have the same function—namely, the inactivation of the cellular tumor suppressor genes *p53* and *RB1*.[21] Cancer thus arises from both the gain of proliferative signals and the loss of genes that inhibit cellular proliferation.

Tumor Suppressor Genes

REVERSION OF TUMORIGENICITY

Formerly referred to as anti-oncogenes, tumor suppressors constitute a family of cellular genes whose inactivation during tumorigenesis promotes malignant growth [*see Table 2*]. Their initial discovery arose from three convergent lines of investigation: analysis of the functional properties of normal versus cancer cells, epidemiologic studies of pediatric cancer, and molecular genetic analysis of chromosomal losses in tumors. Among the earliest discoveries was that fusion of a malignant cell to a nonmalignant cell resulted in a hybrid cell that had lost its malignant properties [*see Figure 2*].[22] This unexpected finding indicated that the malignant state was recessive and suggested that genes in the nonmalignant cell could restore normal growth control to a malignant cell that had presumably lost genetic information. These experiments were followed by the demonstration that individual chromosomes isolated from normal cells were capable of reverting the tumorigenicity of cancer cells.[23] The molecular cloning of specific tumor suppressor genes then made it possible to show that recombinant DNA plasmids containing these genes were capable of suppressing malignant growth.

THE KNUDSON MODEL AND LOSS-OF-FUNCTION MUTATIONS

Most cancers arise during adulthood; presumably, the fact that the incidence of cancer rises with increasing age is a reflection of the accumulation of genetic damage in aging stem cells. Pediatric cancers, although rare, typically develop in early stem cells, such as the renal stem cells that give rise to Wilms tumor, the primitive neuroectodermal cells that are transformed in neuroblastoma, and the retinoblasts that may produce the eye tumor retinoblastoma. A striking feature of these pediatric tumors is the existence of families in which approximately half the children develop specific types of tumors. Within these cancer-prone families, the tumors that arise are frequently bilateral or multicentric at presentation and develop earlier than in corresponding cases arising sporadically within the general population. By using Poisson distribution to determine the probability of developing cancer as a function of age in familial versus sporadic pediatric cancers, Alfred Knudson proposed the model that now forms the foundation of human cancer genetics [*see Figure 3*].[24]

The Knudson model predicts that children with familial tumors have inherited an initial genetic hit (i.e., a distinct genetic alteration) and require only one additional, rate-limiting genetic hit to initiate tumorigenesis. In contrast, children with sporadic tumors need to acquire two independent genetic hits within the

Table 2 Mutations in Selected Tumor Suppressor Genes and Candidate Tumor Suppressor Genes in Cancer[95]

Gene	*Function of Protein Product*	*Cancers with Somatic Mutations*	*Germline Mutations*
RB1	Transcriptional regulator, E2F1 binding	Retinoblastoma, osteosarcoma, SCLC, breast, prostate, and bladder	Familial retinoblastoma
TP53	Transcription factor	~50% of all cancers (rare in some cancer types—e.g., prostate carcinoma, neuroblastoma)	Li-Fraumeni syndrome
p16	Cyclin-dependent kinase inhibitor	~20%–25% of many different cancer types (e.g., breast, lung, pancreatic, bladder)	Familial melanoma, familial pancreatic carcinoma
APC	Regulates β-catenin function, ?microtubule binding	Colorectal mutations, rare or absent in most other cancers	Familial adenomatous polyposis, Gardner syndrome, Turcot syndrome, familial desmoid disease
MSH2, MLH1, PMS1, PMS2	DNA-mismatch repair	Colorectal, endometrial, gastric	Hereditary nonpolyposis colorectal cancer
WT-1	Transcription factor	Wilms tumor	WAGR and Denys-Drash syndromes
NF-1	p21ras-GTPase, ?microtubule binding	Melanoma, neuroblastoma	Neurofibromatosis type 1
NF-2	Juxtamembrane link to cytoskeleton	Schwannomas, meningiomas, ependymomas	Neurofibromatosis type 2
VHL	?Regulator of protein stability	Renal (clear-cell type), hemangioblastoma	von Hippel–Lindau disease
BRCA1	?DNA repair, complexes with Rad51, transcriptional regulation	Ovarian (~10%), rare in breast	Inherited breast and ovarian carcinoma
BRCA2	DNA repair, binds to Rad51	Not known	Inherited breast (female and male), pancreatic, and ?ovarian carcinoma
MEN1	Not known	Parathyroid adenomas, pituitary adenomas, endocrine tumors of pancreas	Multiple endocrine neoplasia type I
PTCH basal	Transmembrane receptor for sonic hedgehog; negative regulator of smoothened protein	Basal cell skin, medulloblastoma	Gorlin syndrome, hereditary cell syndrome
PTEN	Tyrosine phosphatase	Glioma, breast, prostate, head and neck, squamous cell, follicular thyroid cancer	Cowden disease
DPC4	Downstream signaling in TGF-β pathway	Pancreatic, mutations rare in others (e.g., colon, gastric)	Not known
E-CAD	Transmembrane cell-cell adhesion molecule	Diffuse type gastric and lobular breast; rare mutations in others (e.g., endometrial and ovarian)	Not known
α-CAT	Links E-cadherin to the cytoskeleton	Some prostate and lung, ?others	Not known
DCC	?Transmembrane receptor for netrin cell guidance molecule	Mutations in some colorectal and other carcinomas, gliomas, neuroblastomas, germ cell tumors, AML	Not known
TGF-βII R	Transmembrane receptor for TGF-β	Colorectal and gastric, ?others	Not known

AML—acute myeloid leukemia SCLC—small cell lung cancer TGF-β—transforming growth factor–β WAGR—Wilms tumor, aniridia, genitourinary abnormalities, mental retardation syndrome

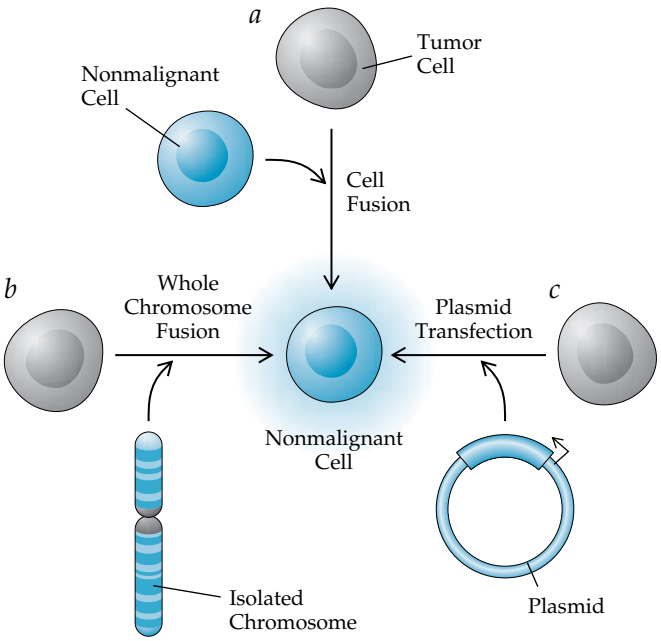


Figure 2 **Reversion of tumorigenicity by tumor suppressor genes. The observation that malignant properties of cultured cells can be reversed by introduction of normal DNA indicates that in many cases, these properties result from the loss of genetic material. These experiments include (*a*) cellular fusion of a malignant cell with a nonmalignant cell, resulting in loss of malignant growth properties; (*b*) introduction of a specific chromosome into the malignant cell, with the same result; and (*c*) insertion of a cloned individual gene (tumor suppressor) into the malignant cell, leading to loss of the malignant phenotype.**

same cell, an unlikely event that explains the less frequent, unilateral presentation and later onset of sporadic cancers. Subsequent genetic studies in two of the tumors studied by Knudson identified these so-called genetic hits as the sequential inactivation of the two alleles of a critical tumor suppressor gene: *RB1* in retinoblastoma[25] and *WT1* in Wilms tumor.[26] The Knudson model also explains the paradox that tumor suppressor gene mutations are loss-of-function or recessive mutations [*see Table 2*], yet familial cancer presents as an autosomal dominant trait. Although loss of a single allele of a tumor suppressor gene may be functionally silent in the presence of a normal second allele, the frequency of spontaneous mutations is sufficiently high to ensure that at least one cell within the target tissue is likely to lose the second allele and initiate malignant transformation [*see Figure 3*].

Tumor suppressor genes were identified by the application of cytogenetic and molecular genetic tools to these functional and epidemiologic observations. Whereas the initial mutation that inactivates one allele of a tumor suppressor gene is typically a point mutation within the gene itself, loss of the second allele commonly results from a gross chromosomal deletion or rearrangement.[27] This second hit may be readily identified by karyotype analysis, but more commonly it requires molecular analyses to detect the loss of polymorphic markers closely linked to the targeted gene. Polymorphic markers (such as restriction fragment length polymorphisms) are genetic variations that may be used to identify a chromosome inherited from either parent [*see Figure 4*]. Loss of a polymorphism derived from either parent (so-called loss of heterozygosity [LOH]) is an indication that a neighboring tumor suppressor gene has been inactivated in a cancer. Such allelic losses, which may be mapped across all chromosomes for a given tumor type, have been most effective for isolating individual tumor suppressor genes, typically with positional cloning strategies.

CELL CYCLE PROGRESSION AND THE RB1 PATHWAY

The first tumor suppressor identified was the *RB1* gene, which is involved in the pediatric eye tumor retinoblastoma.[25] In many ways, *RB1* remains the prototype for this class of genes, and its intimate connection to the basic mechanism of cell cycle progression illustrates the close connection between normal cellular proliferation and malignant transformation. As predicted by the

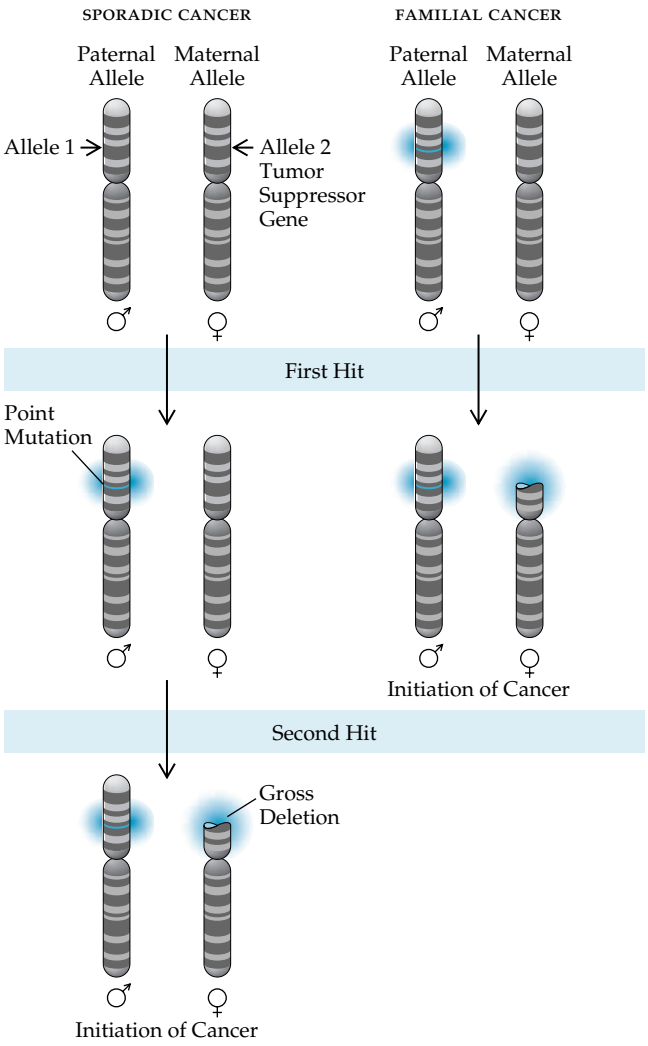


Figure 3 **The Knudson two-hit model of tumor initiation. As currently formulated, the model predicts that inactivation of both alleles of a tumor suppressor gene are required and rate-limiting for the initiation of cancer. In sporadic cancer, the inactivation of these two alleles within the same cell depends on two independent rare genetic events. In contrast, persons who have one mutated tumor suppressor gene allele in their germline, either inherited from one parent or as a result of a de novo germline mutation, require only one genetic event (i.e., loss of the second allele) for tumor initiation. This single mutation may occur in any cell within the target organ, which explains the high frequency of cancer, its earlier onset, and frequent multicentric presentation in persons with genetic predisposition to cancer.**

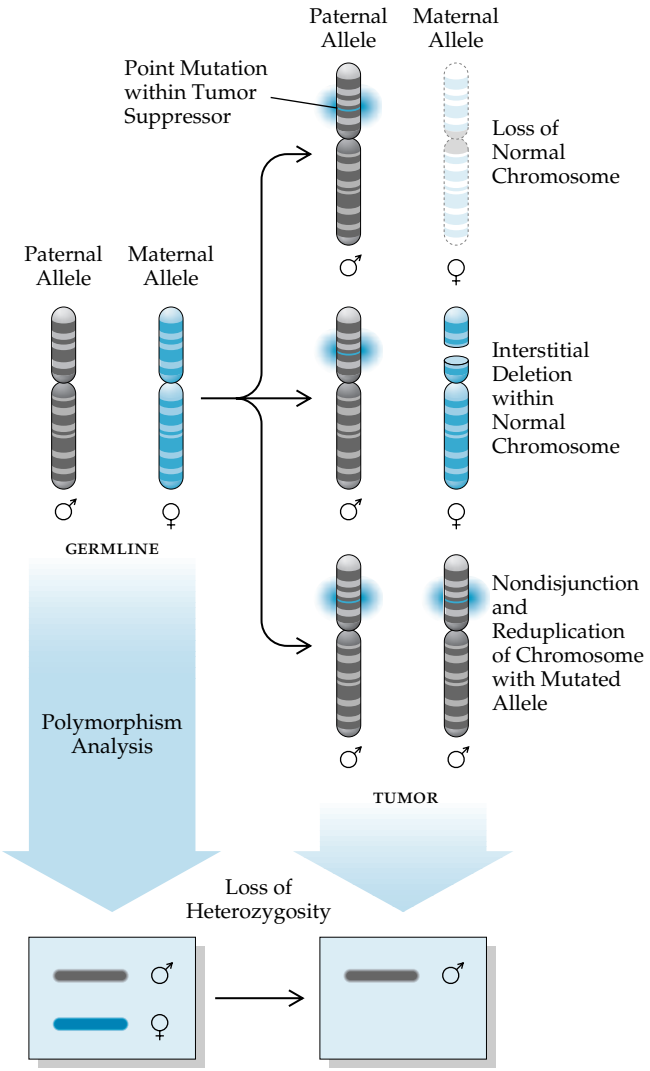


Figure 4 **Allelic losses in tumors and loss of heterozygosity (LOH). The loss of genetic material that accompanies tumor development has been used to map the chromosomal location of tumor suppressor genes. Whereas the initial genetic hit typically results from a point mutation within a tumor suppressor gene, the second gene inactivation event is commonly the result of a gross chromosomal loss. This may include loss of the entire normal chromosome, a small deletion that removes a chromosomal segment containing the gene, or a duplication of the chromosome carrying the mutant allele, with loss of the normal chromosome. These chromosomal events can be mapped using restriction fragment length polymorphisms or other molecular methods that distinguish maternally inherited alleles from paternally inherited alleles. The identification of chromosomal loci that are subject to LOH in specific cancers has formed the starting point for the cloning of cancer predisposition genes.**

Knudson model,[24] one allele of *RB1* is inactivated in the germline of children with familial retinoblastoma; their tumors demonstrate LOH, indicating somatic loss of the remaining allele. Children with sporadic retinoblastoma constitute 90% of cases; in these children, both *RB1* alleles are inactivated within a single somatic cell. Children harboring a mutation in one *RB1* allele are also prone to osteosarcomas, although they do not show increased susceptibility to other types of cancers in which somatic loss of *RB1* has been demonstrated (e.g., lung cancer). This illustrates a central paradox in cancer genetics: a gene expressed in all normal tissues is rate limiting for the initiation of cancer in only selected organs; in other organs, it appears to be just one of many genetic alterations that contribute to the malignant state. Current explanations for this paradox include potential differences in the genetic wiring of different cell types, as well as differences in compensatory mechanisms that may be activated after the loss of a tumor suppressor gene (e.g., the triggering of cell death).

The protein encoded by *RB1* is a nuclear phosphoprotein that binds to products of a gene family called *E2F*, which in turn associate with proteins of the *DP* gene family.[28] The E2F-DP complexes play a critical role in activating transcription of genes required for DNA synthesis; binding to RB1 suppresses this effect, essentially blocking cells in the G_1 phase of the cell cycle. During normal cellular proliferation, a cell cycle–regulated kinase complex containing cyclin D and cyclin-dependent kinase 4 (CDK4) phosphorylates RB1, causing its release from E2F-DP; this in turn allows cells to enter the DNA synthesis (S) phase of the cell cycle. Although the physiologic inactivation of RB1 through hyperphosphorylation is transient and reversible, mutational inactivation of RB1, common to many different tumor types, ensures a continued drive for cellular proliferation. Similarly, oncogenic DNA viruses encode proteins that specifically and irreversibly inactivate the *RB1* gene product, including the E1A protein of adenovirus, T antigen of SV40, and E7 protein of papillomavirus. *RB1* itself is a member of a gene family, with two closely related genes called *p107* and *p130*. Why *RB1* is frequently targeted by mutations in human cancers, whereas these similar genes are not mutated, is still unknown. A better understanding of subtle differences in the physiologic function of these gene family members may explain this striking specificity.

Understanding the normal regulation of *RB1* has provided important insight into other cancer genes that play a role in cell cycle regulation [*see Figure 5*]. Cyclin D, a gene whose expression leads to the hyperphosphorylation and inactivation of RB1, is commonly amplified and overexpressed in breast cancer. In contrast, *p16-INK4a*, the gene for an inhibitor of CDK4 that favors the active, hypophosphorylated form of RB1, is deleted in a wide variety of tumors. Germline mutations of *p16-INK4a* are a cause of familial melanoma.[29] Finally, mutations in the kinase gene *CDK4* that make the encoded protein insensitive to inhibition by p16-INK4a, favoring inactivation of RB1, have also been demonstrated in human cancers and in melanoma kindreds. A striking feature of mutations in these different components of the cell cycle pathway is that they appear to be mutually exclusive. A single tumor typically has an abnormality in only one of these genes—evidence that these mutations are functionally equivalent and that no further growth advantage is conferred by the accumulation of additional mutations in components of the same pathway.[30]

GENOMIC STABILITY GENES: *P53* AND *BRCA1*

Although the *RB1* tumor suppressor gene is a central component of cell cycle regulation, *p53* plays a critical role in the maintenance of genomic integrity—hence its popular designation as "guardian of the genome."[31] p53 is normally expressed at low levels in all cells. However, genetic injuries, such as those that occur through ionizing radiation, trigger the stabilization and activation of p53 protein. p53 functions as a transcription factor, directing expression of p21, an inhibitor of the cyclin-dependent kinases that regulate the cell cycle [*see Figure 6*].[32] Activation of p53 leads to arrest in the G_1 phase of the cell cycle, enabling cells to repair DNA damage before proceeding into S phase and

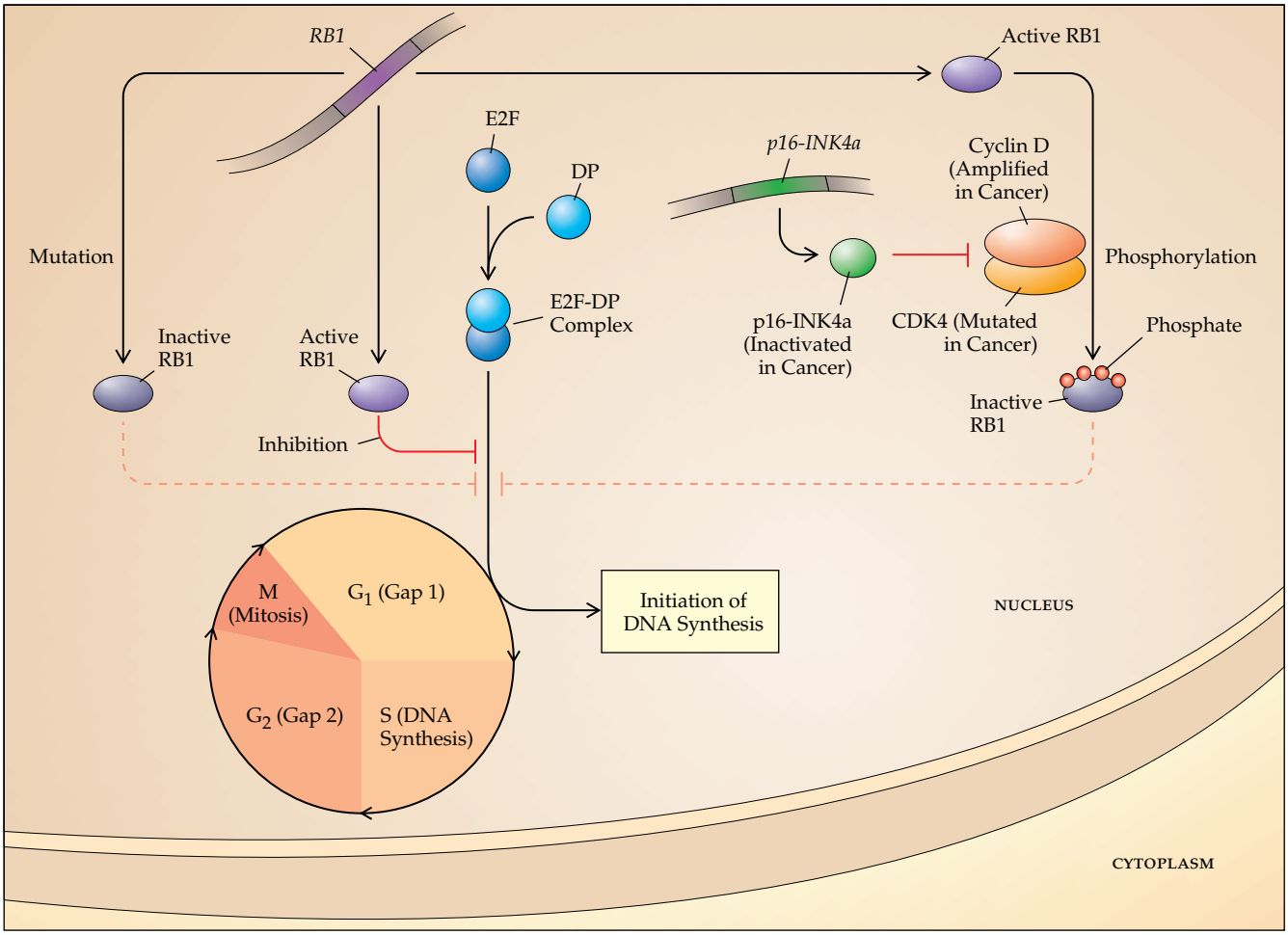


Figure 5 **The retinoblastoma gene (*RB1*) cell cycle pathway. Cellular division in mammalian cells is triggered by a critical checkpoint between G_1 phase and DNA synthesis (S phase). The E2F-DP protein complex, which activates genes required for DNA synthesis, is directly inhibited by the product of the *RB1* tumor suppressor gene. Normal cell cycle progression requires the reversible inactivation of RB1 protein by phosphorylation induced through cyclin D and CDK4. Mutations in RB1 prevent its normal inhibition of cell cycle progression, contributing to deregulated cellular division. Amplification of cyclin D, activating mutations in CDK4, and inactivation of the CDK4 inhibitor p16-INK4a are other mechanisms by which the RB1 pathway is disrupted in human cancers.**

DNA replication. In other cells, activation of p53 causes activation of multiple effectors, leading to apoptosis—a suicide program in cells whose DNA may have been irreparably damaged. Not surprisingly, mutations of p53 are common in human cancers, being demonstrable in about 50% of cases.[33] Most mutations are amino acid substitutions within the DNA-binding domain of p53, resulting in its misfolding and binding to heat shock proteins. The rate of protein turnover is greatly slowed for these mutant p53 molecules. This explains the paradox that high levels of p53 protein in tumor specimens are commonly taken as evidence of a mutation in p53.

The constellations of *p53* mutations in tumors have also provided an important correlation with the presumed mechanism of carcinogenesis. For example, specific *p53* mutations in liver cancers have been found in a population living in a region of China where there is exposure to high levels of the potent fungal liver carcinogen aflatoxin, and the panel of mutations observed in lung cancer are correlated with those induced by the polycyclic carcinogens in tobacco. In contrast, *p53* mutations detected in colorectal carcinoma are frequently seen at nucleotides that are susceptible to methylation, which appears to be a common mechanism for spontaneous mutations.[34]

Deletion of *p53* in the mouse has minimal effect on normal development but leads to greatly enhanced tumor formation.[35] This effect is more subtle in mice lacking one *p53* allele; mice missing both alleles rapidly succumb to tumors. Overexpression of some *p53* mutants may disrupt the normal formation of p53 tetramers, a so-called dominant-negative or dysfunctional phenotype, resulting in an intermediate level of tumor formation. In humans, loss of one *p53* germline allele is responsible for the multicancer phenotype known as Li-Fraumeni syndrome.[36] Families affected by this rare autosomal dominant trait demonstrate a highly penetrant predisposition to different types of cancer, including sarcomas, breast cancer, brain tumors, and leukemia. The mechanism whereby inactivation of p53 function leads to the development of cancer is thought to be loss of a checkpoint that monitors integrity of the genetic material before DNA replication. p53 also plays a critical role in triggering cellular suicide in response to inappropriate growth signals, such as those induced by the loss of other tumor suppressors or activation of proto-oncogenes. In some cancers, such as those arising in Li-Fraumeni syndrome, loss of *p53* initiates tumorigenesis; in others, such as colorectal

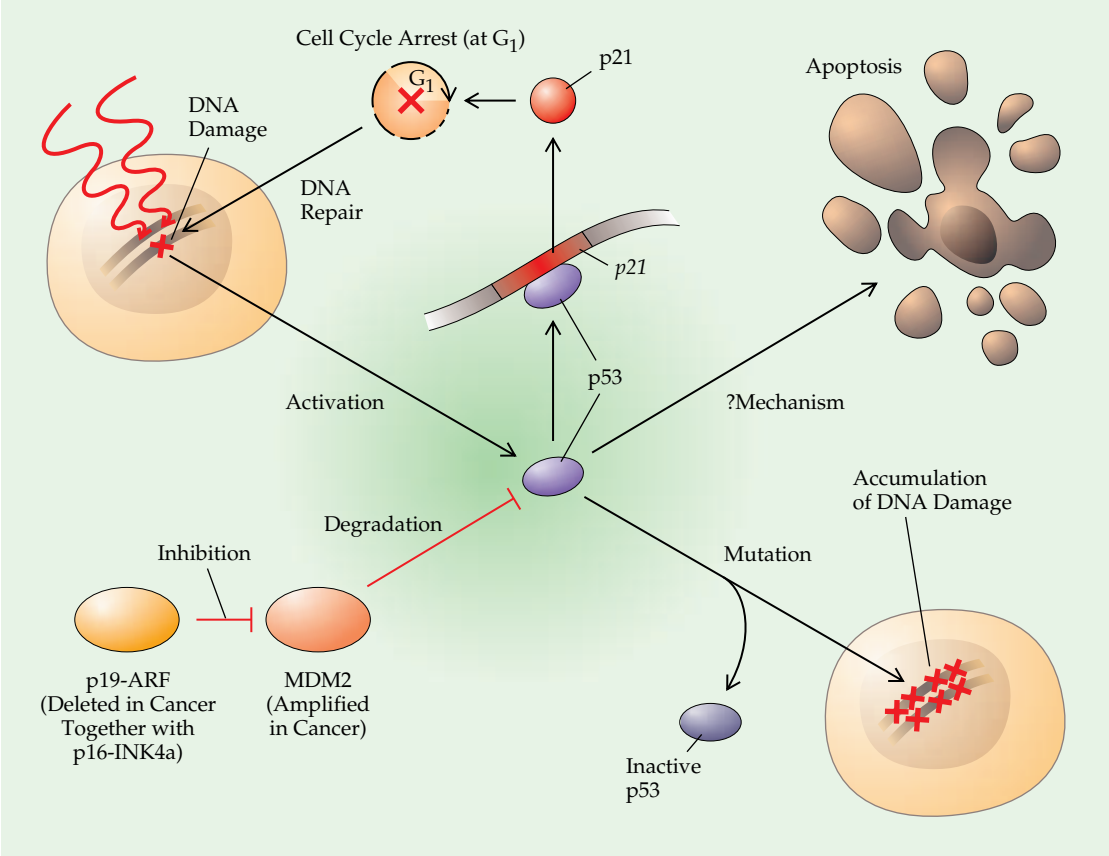


Figure 6 **The p53-DNA damage response pathway. p53 is activated after DNA damage, resulting in either of two cellular responses: induction of the CDK inhibitor p21, which leads to a G_1-phase cell cycle arrest, allowing time to repair DNA damage before cellular division; and programmed cell death (apoptosis), whereby the damaged cell undergoes suicide. In tumors with mutations of p53, DNA damage is not adequately repaired, leading to the accumulation of mutations and genetic rearrangements. Two cellular genes regulate p53 activation: *MDM2* and *p19-ARF*. MDM2 induces the normal degradation of p53 protein, and amplification of *MDM2* in some tumors results in loss of p53 function. p19-ARF inhibits the effects of MDM2 and contributes to the activation of p53 in response to DNA damage. *p19-ARF* is commonly deleted in human cancers, along with its neighboring gene, *p16-INK4a*.**

cancer, inactivation of *p53* is a late genetic event important in the progression of the malignant phenotype.

Like RB1, p53 is specifically targeted by viral oncogene products, including the E1B protein of adenovirus, T antigen of SV40, and the E6 protein of papillomavirus. Thus, DNA tumor viruses combine a dual strategy: inducing cellular proliferation by inactivating RB1 while disrupting the ability of p53 to trigger cell death in response to this aberrant proliferation signal.[37] Cellular proteins have also been found to regulate *p53* function. Among them is MDM2, a protein that is induced by p53 and that itself enhances p53 degradation, providing a critical negative feedback loop [*see Figure 6*]. The fact that p53 and MDM2 are part of a common cellular pathway is supported by mutually exclusive alterations in human cancers: osteosarcomas result from either a mutation inactivating p53 or amplification of the *MDM2* gene, leading to overproduction of the gene product.[38] Mice lacking both alleles of *MDM2* die in very early embryonic development, unless they also lack *p53*, in which case they develop normally to adulthood—evidence that MDM2 is critical to prevent unopposed activity by p53 during normal development.[39]

Activation of p53 in response to ionizing radiation occurs through the DNA damage response kinase ATM (ataxia-telangiectasia mutated) and its downstream kinases, CHK1 and CHK2.[40] Phosphorylation of p53 appears to reduce binding by MDM2 and lead to increased p53 protein stability and enhanced activity. A second important regulator of p53 is the *p19-ARF* gene. This gene is unusual in that it overlaps with the *p16-INK4a* gene but is encoded by a different codon reading frame.[41] p19-ARF interacts with MDM2 to regulate the turnover of p53; its role in human cancer is suggested by the frequency with which tumors contain genomic deletions that inactivate both p16-INK4a and p19-ARF. Activation of *p53* through the p19-ARF pathway is thought to be triggered primarily by aberrant proliferation signals associated with induction of E2F, a transcriptional regulator of p19-ARF.[42] Thus, DNA damage and aberrant oncogenic proliferation signals may activate p53 through distinct but complementary pathways. Of note, two long-overlooked gene family members of *p53*, *p73* and *p63*, have some of the same functional properties of *p53* in vitro; their possible role in human cancer is still under investigation.[43,44]

Although *p53* has been the most intensively studied tumor suppressor gene implicated in genomic stability, additional disease-associated genes appear to function in similar or parallel pathways. The ataxia-telangiectasia gene, *ATM*, is responsible for an autosomal recessive trait characterized by cerebellar degeneration, immunologic dysfunction, and cancer predisposi-

tion; as noted above, it functions upstream of *p53* and is required for *p53* to be activated after ionizing radiation.[45,46] Two other genes that may function in genomic stability are *BRCA1* and *BRCA2*.[47,48] The products of these genes, responsible for predisposition to breast and ovarian cancer, appear to interact with other proteins active in chromosomal recombination and repair.[49] Although the precise functional properties of the *BRCA1* and *BRCA2* gene products are currently unknown, mutations in these genes may trigger malignant transformation by preventing cells from repairing DNA damage. Other genes involved in the repair of chromosomal damage are associated with tumor-prone syndromes, including Fanconi anemia and the Nijmegen breakage syndrome.[50]

TUMOR SUPPRESSORS INVOLVED IN SIGNALING AND DIFFERENTIATION

The identification of tumor suppressor genes implicated in cancer predisposition syndromes led to the discovery of key components of cellular differentiation pathways. The *WT1* gene encodes a transcription regulator that is specifically expressed in podocytes of the developing glomerulus. Mutations in *WT1* cause Wilms tumor, an embryonic kidney cancer, and germline mutations lead to genitourinary developmental defects. The genes that are normally regulated by WT1 appear to be involved in proliferation and differentiation.[51] The von Hippel–Lindau gene (*VHL*) is frequently mutated in adult renal cell cancers and in the germline of persons with a syndrome that includes both benign and malignant vascular tumors. The function of the VHL protein appears to involve regulation of protein degradation pathways, particularly that of the transcription factor hypoxia inducible factor.[52] The *APC* gene is a key target in colorectal cancer: germline mutations cause familial polyposis coli, a syndrome characterized by the development of numerous colonic polyps with a very high risk of malignant transformation; somatic mutations constitute the earliest step in the development of colorectal cancer.[53] APC normally binds β-catenin, leading to its inactivation. β-Catenin is an important cofactor for the LEF/TCF family of transcription factors, which mediate signaling by the *WNT* gene family; the latter is involved in key developmental pathways. Inactivation of APC therefore disrupts WNT function, resulting in abnormal developmental signals in the colonic epithelium. The *SMAD* genes, active in signaling by transforming growth factor–β (TGF-β), are mutated in pancreatic tumors.[54] In the basal cell nevus syndrome, PTCH, a receptor for another growth factor called hedgehog, is mutated, conferring predisposition to skin cancer.[55] Other tumor suppressor genes linked to cellular signaling pathways are the *NF1* gene, which is responsible for neurofibromatosis (von Recklinghausen disease) and whose gene product normally downregulates the proto-oncogene *ras*[56]; the *NF2* gene, which is frequently mutated in mesotheliomas and schwannomas and which encodes a structural protein that may be involved in cellular adhesion and proliferation[57]; and the *PTEN* gene, which is implicated in a large number of different cancers and in the breast cancer predisposition syndrome known as Cowden disease. *PTEN* encodes a phosphatase thought to catalyze the removal of key phosphate moieties from specific target proteins.[58] PTEN negatively regulates the phosphoinositide 3-kinase pathway implicated in the regulation of proliferation and apoptosis. In sum, the genes whose disruption leads to human cancer are critical links in the pathways for normal cell differentiation and proliferation.

Microsatellite Instability and DNA Mismatch Repair

The use of short, repetitive stretches of nucleotides (microsatellites)—which are highly variable from person to person—to trace different alleles within tumor specimens led to a startling discovery: in some cases, microsatellites differ in size between tumors and normal cells of the same patient. This observation suggested that the tumors had inactivated genes that are required to correct errors in the replication of these short, repetitive stretches of DNA.[59] Inactivation of DNA mismatch repair genes, which are highly conserved from bacteria to humans, leads to the accumulation of errors during DNA replication that can be detected with randomly positioned microsatellite markers. When replication errors occur within critical genes, however, they may be detrimental and may trigger malignant proliferation. The members of the mismatch repair gene family that are most commonly targeted by mutations in human cancers are called *MSH2* and *MLH1*. Mutations in these genes result in loss of function; however, mismatch repair genes differ from classic tumor suppressors in that their inactivation initiates cancer indirectly (i.e., by increasing the frequency of mutations in other genes). Reintroduction of a normal mismatch repair gene copy after secondary mutations have arisen would presumably occur too late to reverse the malignant phenotype. Potential targets for mismatch repair genes have been identified. Among the most striking is the TGF-β receptor, which contains a short, repetitive stretch within its coding sequence.[60] Colorectal tumors, in which mutation of a mismatch repair gene triggers microsatellite instability, may have mutations within this specific sequence of the TGF-β receptor gene, thus disrupting an important pathway for the regulation of cellular proliferation in colonic epithelium. Germline mutations in mismatch repair genes are associated with a multicancer phenotype, called Lynch syndrome, or hereditary nonpolyposis colorectal cancer, which includes predisposition to cancers of the colon, ovary, and endometrium.[61] In addition, approximately 10% of sporadic colorectal cancers demonstrate microsatellite instability (so-called RER^+), suggesting that disruption of mismatch repair also contributes to colorectal cancer in the absence of familial predisposition [*see Figure 7*].[62]

Tumor Progression

THE ACCUMULATION OF GENETIC LESIONS

A genetic lesion within a single rate-limiting gene is required to initiate malignant transformation. However, the progression of the malignant phenotype requires the acquisition of additional mutations that confer an added growth advantage and are therefore selected during expansion of the malignant clone [*see Figure 8*]. A given gene may be rate limiting for transformation in one cell type but may play a secondary role in another. The most intensely studied model of tumor progression is colorectal cancer, for which Kinzler and Vogelstein correlated histologic progression from polyp to carcinoma with the accumulation of genetic events.[63] The inactivating mutation of the *APC* gene is associated with the development of epithelial hyperplasia; activating mutations in H-*ras* and changes in DNA methylation are correlated with progression to adenomatous polyps; changes in one or more genes residing on chromosome 18q denote the transition to high-grade adenomas; and finally, inactivation of p53 accompanies the evolution to malignant carcinoma [*see Figure 9*]. The preneoplastic lesions that lead to colorectal cancer are readily defined because they occur within the colonic mu-

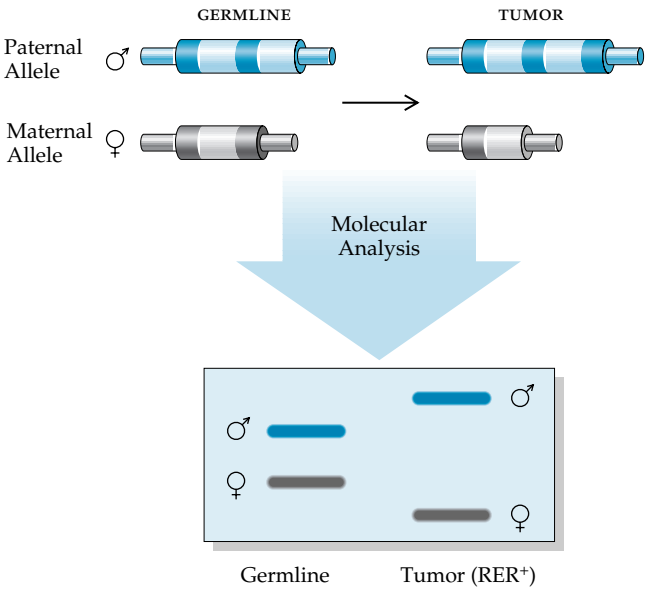


Figure 7 **Microsatellite instability and DNA mismatch repair. Microsatellites are short repeated sequences of DNA that are distributed throughout the genome and are highly variable, facilitating the distinction between maternally inherited alleles and paternally inherited alleles. Unlike tumors showing LOH, which have loss of one allele compared with the germline, tumors that have microsatellite instability (RER$^+$) demonstrate altered lengths for microsatellite markers. This results from errors in DNA mismatch repair. When microsatellite sequences are present within the coding region of a gene, such errors lead to mutations and loss of gene function.**

cosa and are accessible to colonoscopic biopsy. Other models of tumor progression are under study, including esophageal, head and neck, and bladder cancers.

CELL DEATH AND IMMORTALITY

The initial cellular proliferation drive that occurs in a transforming event is often accompanied by an increase in cell death or apoptosis, a compensatory mechanism that prevents the rapid growth of a cancer. Inactivation of the *p53* tumor suppressor is the most common genetic lesion that abrogates the cell death response, allowing cancer cells to rapidly increase in number. Another gene involved in the regulation of apoptosis is *BCL2,* which encodes a mitochondrial protein that prevents triggering of the protease cascade required for cellular suicide.[64] The role of *BCL2* in human cancer is best illustrated in follicular B cell lymphoma, an indolent tumor in which a chromosomal translocation brings the *BCL2* gene under the control of the immunoglobulin gene enhancer.[65] The increased expression of *BCL2* in these lymphoid cells prevents their programmed cell death, resulting in the increased cell mass that characterizes this slow-growing lymphoma. Mouse models that recreate this chromosomal rearrangement also demonstrate a striking increase in lymphocyte numbers, analogous to the human lymphoma. *BCL2* is part of a large gene family, with some members exhibiting proapoptotic effects and others antagonizing programmed cell death.[64] A proapoptotic member of the *BCL2* family, *BAX,* contains a repetitive nucleotide sequence that is targeted by mutations in tumors with microsatellite instability.[65,66] Thus, overexpression of antiapoptotic *BCL2* gene family members or inactivation of proapoptotic genes may contribute to the malignant phenotype.

Whereas proliferation and apoptosis constitute immediate and opposing cellular responses to mutations in proto-oncogenes and tumor suppressors, cellular immortality refers to the unlimited cellular lifespan required for tumor formation. Somatic cells are programmed to undergo only a limited number of cell divisions, after which they enter a state called senescence. This is most readily apparent in vitro, where cultured primary cells reach the end of their programmed lifespan, enlarge, and permanently cease to divide.[67] The molecular mechanisms underlying cellular senescence are not well understood, but roles for the *p16-INK4a, p53,* and *p19-ARF* tumor suppressor genes have been suggested in mouse models. Primary cells that are driven to proliferate by oncogenic transformation bypass senescence and enter crisis, or massive cell death, as they approach the end of their lifespan. The small number of cells that survive crisis are capable of an infinite number of cell divisions, essentially becoming immortal cell lines. These in vitro studies have provided a model to study the biological clock of normal somatic cells, one of the most potent safeguards against malignant transformation and one that must be overcome by all cancer cells destined to grow past their prescribed lifespan.

The cellular biological clock is linked to telomeres, the ends of chromosomes that are necessary to maintain chromosomal integrity. Telomeres are composed of a stretch of repeated nucleotides.[68] Because DNA polymerase fails to copy the extreme ends of chromosomes—a problem caused by its requirement for annealed primers to initiate DNA replication—telomeres shrink

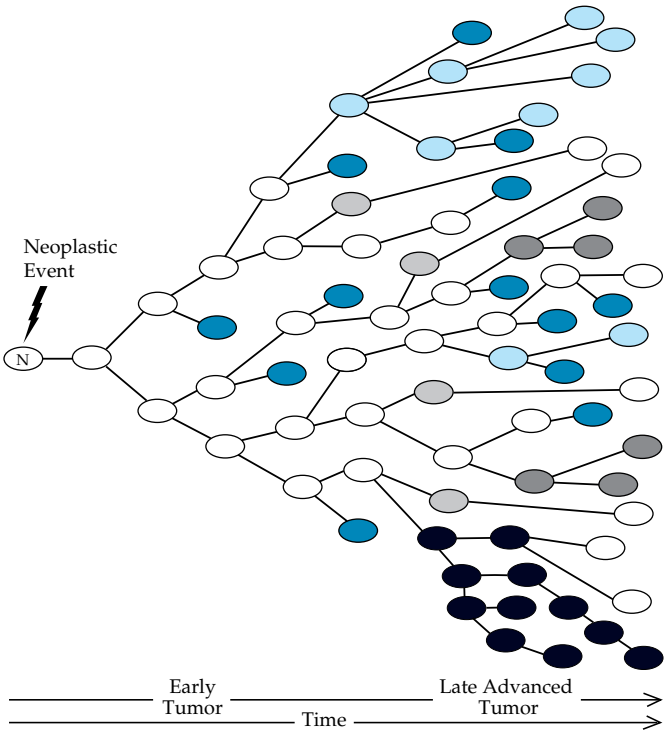
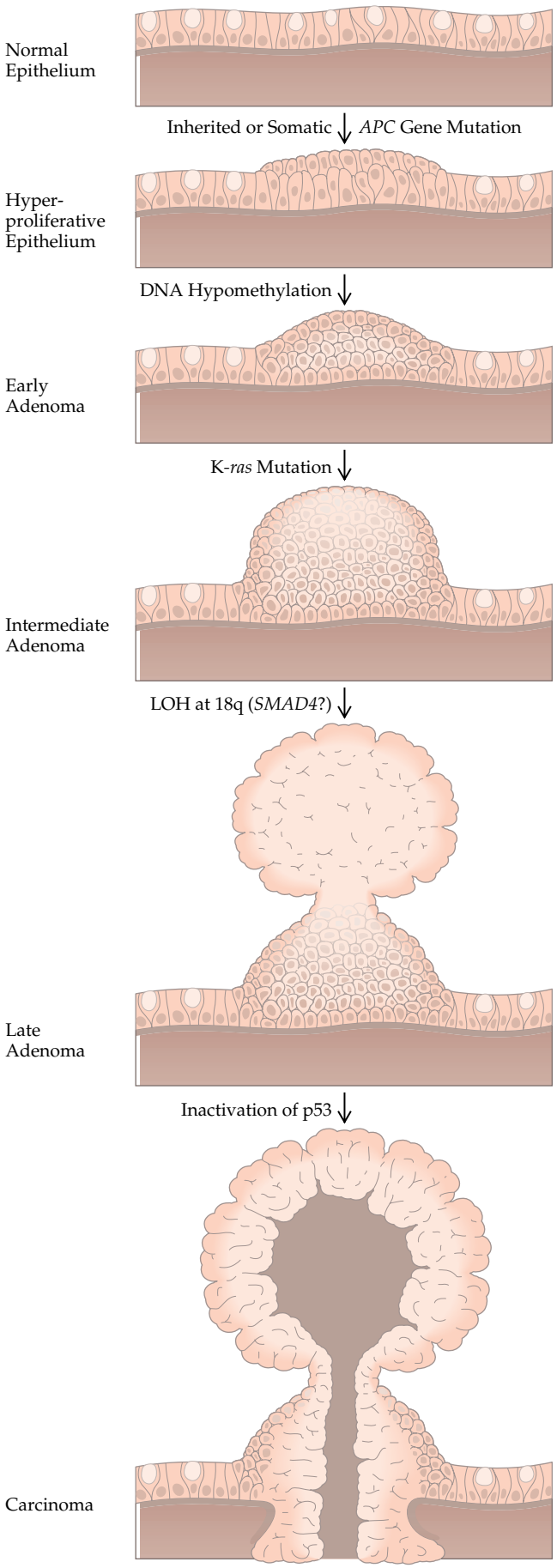


Figure 8 **After an initiating, rate-limiting, neoplastic genetic event, the proliferation of cancer cells is driven by selective growth advantages conferred by cumulative mutations. Continued cellular division and loss of genes required to maintain genomic stability facilitates the occurrence of mutations. Properties such as loss of cell death, invasiveness, angiogenesis, and drug resistance characterize the progression in the phenotype of tumor cells. The colors indicate the clonal descendants of individual cells.**



with every cell division. Cellular senescence is well correlated with the shortening of telomeres to less than the size required to maintain the integrity of chromosomes. In contrast to somatic cells that have a finite lifespan, germ cells and hematopoietic stem cells preserve the length of their telomeres with an enzyme called telomerase. A ribonucleoprotein with reverse transcriptase activity, telomerase uses an RNA template to add the appropriate nucleotide repeats to the ends of telomeres, thus matching the progressive shortening that accompanies cell division [*see Figure 10*].[69,70] This physiologic mechanism for circumventing senescence has been adapted by human tumors, most of which express high levels of telomerase. By analogy, cultured cells that emerge from crisis and have become immortal also express telomerase, validating this in vitro model of malignant transformation. The isolation of the catalytic subunit of telomerase, called hTERT, has allowed analysis of its expression pattern at the single-cell level; this analysis has shown that hTERT is commonly induced at the earliest stages of malignant transformation.[71] The fact that hTERT encodes a unique reverse transcriptase has also stimulated interest in the design of specific inhibitory agents.

METASTASIS AND ANGIOGENESIS

The accumulation of mutations driving cellular proliferation and the acquisition of an infinite lifespan provide the essential components for malignant transformation, but it is tissue invasion and metastasis that constitute the most visible and life-threatening properties of cancer. Cancer cells invade normal neighboring tissues, penetrating the epithelial basement membrane and the lining of endothelial cells and traveling through either the bloodstream or lymphatic chains to distant organs, where they exit the vasculature and establish independent sites of growth. Unfortunately, the genetic events responsible for these complex characteristics of human cancer are not well defined, and current animal models have been of only limited relevance. Cancer cells are known to secrete enzymes, such as matrix metalloproteinase-2 (MMP2), that are capable of digesting the basement membrane, facilitating tissue invasion and seeding of the bloodstream.[72] Expression of specific membrane-bound proteins, such as avb3αvβ3 integrin, appears to anchor MMP2 to the surface of tumor cells. Systemic administration of metalloproteinase inhibitors in combination with chemotherapeutic drugs is being investigated in clinical trials. Cell adhesion molecules have also been implicated in metastasis, possibly explaining in part the apparent predilection of certain tumors for specific metastatic sites. Among the best-characterized molecules is the cell surface protein CD44; alternative pre-mRNA splicing produces variants of CD44 transcripts, yielding different protein isoforms. Specific CD44 variants are present at high levels in metastatic tumor cells and may play a role in facilitating invasiveness.[73] Reduced ex-

Figure 9 **Genetic alterations during colorectal cancer progression. This model, proposed by Bert Vogelstein and commonly known as the "Vogelgram," links histologic features in the progression of colorectal cancer to specific genetic lesions. Mutation of the familial polyposis gene *APC* constitutes the initiating event for familial colorectal cancer and most cases of sporadic colorectal cancer. Additional genetic lesions are associated with polyps of increasing severity, whereas loss of *p53* commonly marks the transition to invasive colorectal cancer. The progression of other tumor types may be driven by the loss of different tumor suppressor genes or by a different cumulative order in the inactivation of the same genes, including *p53*.**

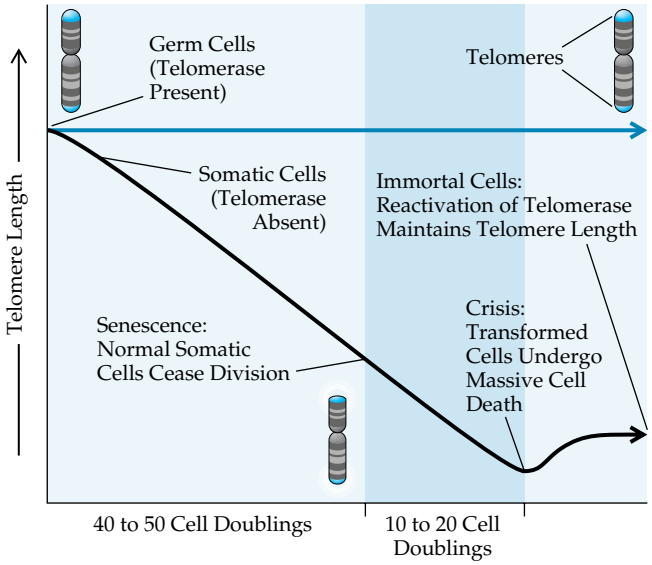


Figure 10 **Cellular senescence and the activation of telomerase. Normal cellular division requires the maintenance of telomeres at the ends of chromosomes. In germ cells programmed for an eternal lifespan, this telomere maintenance results from the activity of the enzyme telomerase. Growth of somatic cells in culture results in progressive shortening of telomeres, until such time as cellular division ceases and senescence is reached. Cells driven to proliferate past that point by the expression of transforming genes continue to shrink their telomeres until they reach crisis, a point at which chromosomes become unstable and massive cell death results. In small numbers of cells, telomerase may be activated at the time of crisis, producing immortal cell lines capable of indefinite growth in vitro. The majority of human cancers express high levels of telomerase, suggesting that a similar selection pressure in vivo contributes to their unlimited growth potential.**

pression of E-cadherin and α5β1 integrin in tumor cells has also been correlated with increased tumor cell migration and tumorigenicity. Additional comparisons between primary and metastatic tumors will probably reveal genetic changes that underlie the progression of the malignant phenotype.

A critical property required for the growth of cancers beyond a minimal size is the recruitment of blood vessels.[74] Tumors vary in their degree of vascularization; and even within a given tumor, certain areas may become hypoxic and frankly necrotic as they outgrow their blood supply. Some of the angiogenic peptides secreted by tumors are the same as certain peptides released during physiologic wound healing, including vascular epidermal growth factor and platelet-derived growth factor. However, in mouse models, certain primary tumors may also release substances that are thought to prevent their distant metastases from recruiting blood vessels—a phenomenon that might be explained as essentially preventing competition for nutrients. This discovery led Folkman and colleagues to isolate two antiangiogenic peptides called angiostatin and endostatin.[75-77] Both of these substances are unusual in that they appear to be breakdown products of other proteins—namely, plasminogen and type XVIII collagen. Their effect in mouse models is striking—they cause the necrosis of established tumors [*see Figure 11*]. In human trials to date, however, angiostatin and endostatin have not demonstrated the effects they show in mouse studies. Recent studies have focused on inhibition of other angiogenic growth factors, including vascular endothelial growth factor. The identification of short peptides that recognize specific markers within tumor endothelial cells has fostered another novel therapeutic approach that entails the designing of chemotherapeutic agents directly targeted at the tumor vascular supply.[78]

RESISTANCE TO CHEMOTHERAPEUTIC DRUGS

Surgical resection constitutes the primary treatment modality for most solid tumors. Chemotherapy and radiation therapy are effective in the treatment of tumors with known metastatic spread; in the so-called adjuvant setting, where residual microscopic tumor deposits are suspected on the basis of the clinical stage of the tumor; and, occasionally, in the preoperative setting, where initial systemic treatment may be required to make a large tumor surgically resectable. Many chemotherapeutic agents are derived from naturally occurring compounds; they include alkylating agents that chemically cross-link DNA, nucleotide analogues that inhibit enzymes required for DNA synthesis, and inhibitors of the mitotic spindle. Chemotherapeutic agents that directly damage DNA are typically administered by short-term infusion, whereas nucleotide analogues that are thought to be effective only during the S phase are given by continuous infusion to maximize the potential exposure of cycling tumor cells. In general, tumors with a high mitotic rate are thought to be more responsive to chemotherapy than those with a smaller fraction of proliferating cells. Chemotherapeutic agents are commonly administered in combination; the specific drugs to be combined are chosen on the basis of their pharmacologic properties, the spectrum of drug sensitivity displayed by different tumor types, and the need to avoid overlapping toxicities to such organs and tissues as bone marrow, intestinal epithelium, kidney, and nervous system. Attempts at in vitro testing of tumors for drug sensitivity have been limited by the fact that only a small fraction of cells from a tumor can be cultured, and these may not adequately represent the heterogeneity of the primary tumor. Radiation therapy entails targeting ionizing radiation, which causes double-stranded breaks in DNA, to specific anatomic sites infiltrated by cancer. Exposure of normal tissues is minimized by the use of multiple converging beams of radiation, together with protective blocks. In some cancers, such as localized rectal and anal cancers, the combined use of chemotherapy and radiation therapy has been particularly effective.

Resistance of cancer to the effects of chemotherapy may be intrinsic, as commonly seen in melanoma, hepatocellular carcinoma, and renal cell carcinoma, which are typically refractory to many agents. In these cases, drug resistance may reflect properties of the tissue from which the cancers arose, such as the need of normal renal tubular cells to be protected from naturally occurring toxins to which they may be exposed. In other cases, cancers that are initially highly sensitive to chemotherapy may become resistant, either during a course of treatment or, more commonly, at the time of disease recurrence. This acquired form of drug resistance presumably results from mutations that confer a selective advantage to a subset of tumor cells, analogous to other genetic events that contribute to tumor progression. A number of molecular mechanisms conferring resistance to specific chemotherapeutic agents have been defined using in vitro models. A well-studied example is the drug methotrexate, which specifically inhibits the enzyme dihydrofolate reductase (DHFR), required for DNA synthesis. Tumors may become drug resistant by acquiring defects in the intracellular transport of methotrexate; by mutations in the *DHFR* gene, resulting in an enzyme that fails to bind methotrexate; or by amplification of the *DHFR* gene itself, which increases the amount

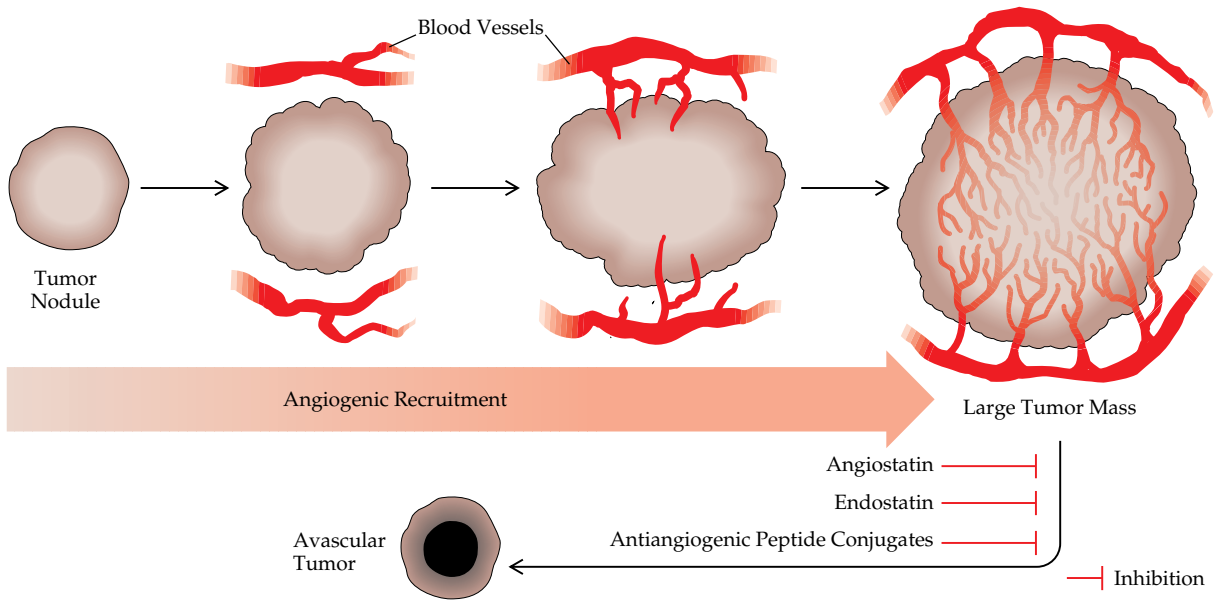


Figure 11 **Tumor angiogenesis and its inhibitors. Among the most important properties of tumors is their ability to recruit blood vessels, providing nutrients that allow their growth beyond a minimal size. Angiogenic recruitment depends on the secretion by tumors of endothelial growth factors. It may be inhibited by peptides, such as angiostatin and endostatin, as well as by drug conjugates using small peptides that target endothelial membrane proteins.**

of protein made and thus acts to titrate and overcome the intracellular drug concentration.[79]

Although drug-specific resistance mechanisms may contribute to some cases of chemotherapeutic drug resistance, most cancers appear to acquire resistance to a broad range of agents, including some to which they have had no prior exposure. This multidrug-resistance phenotype was reproduced in cell culture, and this led to the discovery of the multidrug-resistance (*mdr*) gene family. These genes encode adenosine triphosphate (ATP)–dependent pumps that span the cell membrane and appear to be involved in the transport of large organic compounds.[80] Other gene family members include the gene for cystic fibrosis and a prokaryotic gene that may contribute to quinine resistance in malaria. The most intensively studied gene implicated in resistance to cancer chemotherapy is *mdr*-1, whose gene product is capable of exporting a wide range of organic compounds, including anthracyclines and vinca alkaloids. The molecular mechanisms underlying drug specificity by mdr-related transporters is unknown. In addition, other drugs, such as cyclosporine and verapamil, are capable of noncompetitively blocking the export of chemotherapeutic agents by mdr-1 and are being included in chemotherapeutic regimens undergoing testing in clinical trials. The strong link between *mdr*-1 overexpression and resistance to chemotherapy that is evident in vitro has not been generally confirmed in the clinic, and the contribution of this gene to most cases of clinical drug resistance is still unknown. However, in some cancers, such as multiple myeloma, repeated exposure to anthracyclines and vinca alkaloids does appear to be correlated with increased expression of *mdr*-1 and resistance to chemotherapy,[81] and *mdr*-1 expression appears to be a marker for poor-prognosis acute myelogenous leukemia.[82]

New insight into resistance to chemotherapeutic agents comes from a better understanding of the mechanism of action of these agents. DNA injury mediated by chemotherapeutic agents and ionizing radiation triggers activation of p53, leading to programmed cell death, or apoptosis [*see Figure 12*]. The realization that exposure to chemotherapeutic agents initiates an active cellular suicide program, rather than simply causing

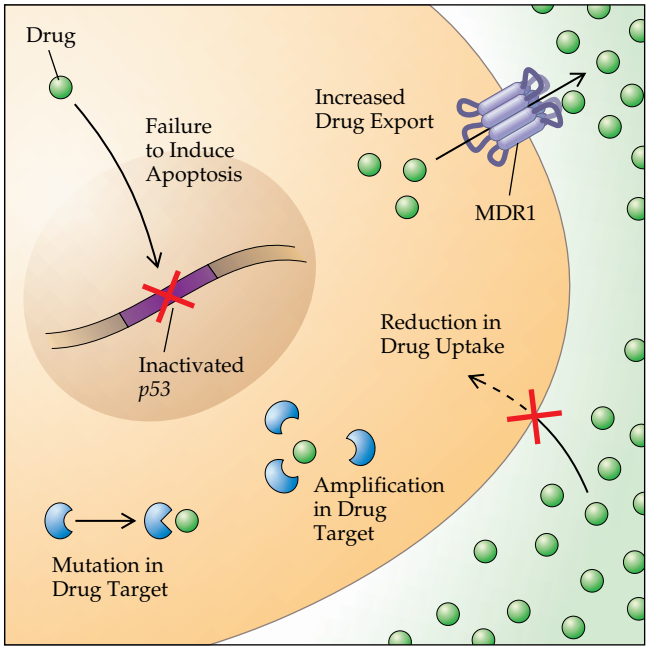


Figure 12 **Mechanisms of resistance to chemotherapeutic agents. The resistance of cancers to chemotherapeutic drugs may be intrinsic, reflecting properties of their tissue of origin, or it may be acquired, a consequence of mutations selected during tumor progression and previous drug treatments. Mechanisms include mutations in the *p53* gene, which preclude the physiologic cellular apoptotic response to DNA injury mediated by these drugs; decreased drug uptake; increased drug export by MDR1-related proteins; a compensatory amplification in the amounts of the cellular drug target; or a mutation in the target that renders it resistant to the effects of the drug.**

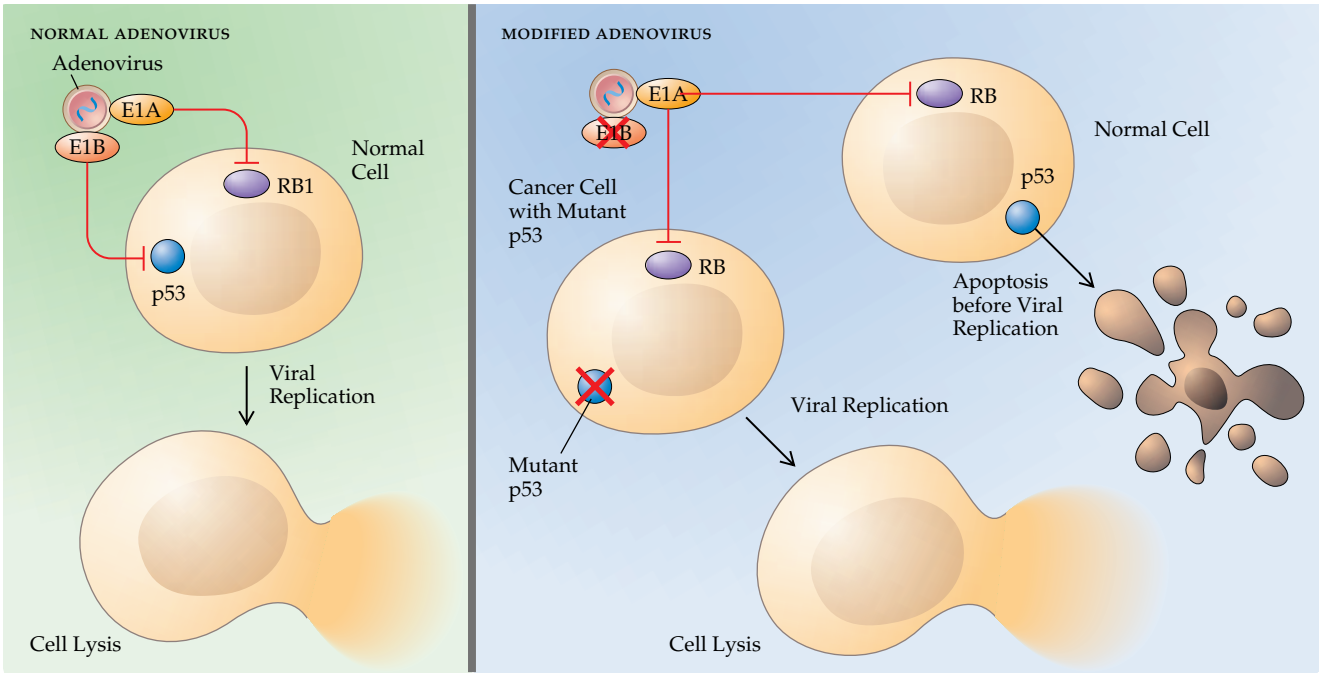


Figure 13 **Specific killing of cells with mutant p53 using a modified adenovirus. The common cold virus adenovirus encodes two genes, *E1A* and *E1B*, that inactivate the cellular tumor suppressor genes *RB1* and *p53*. Loss of RB1 leads to cellular replication, whereas loss of p53 prevents the cellular apoptotic response to viral infection and aberrant cellular replication. Cellular lysis results from unimpeded viral replication. The modified adenovirus produced by Frank McCormick and collaborators lacks the *E1B* gene and therefore can only lyse cancer cells in which the cellular *p53* gene has been mutated. The presence of intact p53 in surrounding normal cells leads to their apoptosis following viral infection, preventing widespread infection of normal tissue.**

cell death through massive DNA damage, raised the possibility that these apoptotic pathways may be disrupted in drug-resistant cancer cells. In a rodent model system, cells driven to proliferate by expression of oncogenes undergo apoptosis after treatment with chemotherapeutic agents or radiation.[82,83] However, inactivation of *p53* in these cells is associated with resistance to these therapeutic agents. Loss of *p53* is a common feature of human cancers, and a number of studies have suggested a correlation between inactivation of *p53* and resistance to chemotherapeutic agents, as well as a more aggressive clinical course.[34] The possibility of using genetic lesions in tumors to enhance, rather than diminish, the selective killing of human cancer cells is being intensively investigated. A promising approach has been the development of a modified adenovirus, capable of replicating and inducing lysis only in cells whose *p53* genes have been inactivated, leading to the specific killing of cancer cells [*see Figure 13*].[84] Selective tumor cell lysis by this modified adenovirus may also reflect other factors, such as viral infectivity[85] and the host immune response to viral infection. A second adenovirus that selectively replicates in human tumor (retinoblastoma) cells, but not normal cells, has been reported.[86]

Expression Profile Analysis and New Therapeutic Targets

MICROARRAYS

The development of high-density microarray technologies has made it possible to scan all expressed human genes simultaneously in an unbiased search for differences between clinical specimens. Whereas traditional experimental approaches have favored the study of individual genes within well-characterized cellular pathways, expression profiling aims at a global analysis of all the mRNA transcripts expressed in a tissue or tumor [*see Figure 14*]. A number of technical platforms have been developed, including oligonucleotide arrays, in which multiple short stretches of nucleotides representing each gene are attached to a chip,[87] and cDNA arrays, in which longer clones representing each gene may be fixed onto a glass slide.[88] A single chip or slide can carry markers for as many as 10,000 genes or more, making it possible to measure their relative levels of expression in a single hybridization, using labeled RNA derived from clinical material. Laser-directed scanning is then used to identify the hybridization signal for each gene. Complex algorithms are required to quantify the signal and to cluster expression patterns that may be correlated with a clinical phenotype. A different strategy, known as serial analysis of gene expression, involves direct nucleotide sequencing of multiple expressed tags derived from cellular transcripts.[89]

An important clinical application of expression profile analysis has involved its ability to distinguish tumors that are histologically similar but differ in their clinical prognoses. For instance, a cluster of gene expression patterns has been used to distinguish large-cell lymphomas with favorable prognoses from those with adverse prognoses,[90] and to identify primary breast cancers with different clinical outcomes.[91] As these studies are confirmed prospectively and as the assays themselves become more broadly available, molecular profiling may eventually become critical in making appropriate treatment decisions. Furthermore, such studies may lead to a new appreciation of similarities and differences between tumor types that is based on shared genetic pathways and mutations, rather than tissue type or histologic appearance—so-called molecular tax-

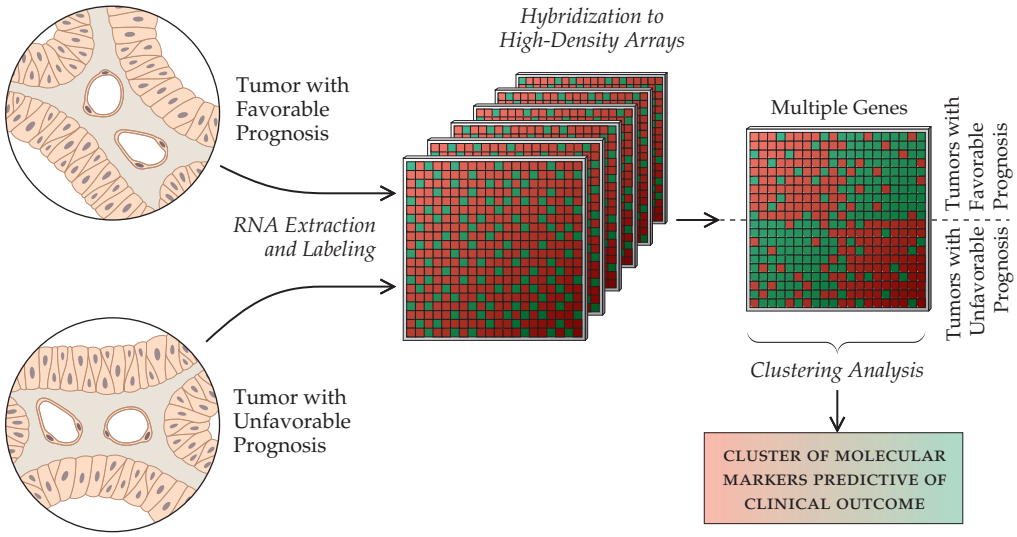


Figure 14 **Analysis of expression profiles using high-density microarrays. This strategy is designed to allow simultaneous quantitation of all messenger RNA transcripts within a tissue. For instance, tumor specimens representing different types of cancer or cancers with different clinical behavior can be compared. Total cellular mRNA is isolated from frozen tumor specimens. The RNA is labeled as a probe and hybridized to high-density microarrays carrying markers for approximately 10,000 genes. The hybridization signal for each gene is detected and compared with different clinical specimens. Complex data mining and bioinformatics approaches are required to identify clusters of expression patterns that may be correlated with different clinical measures.**

onomy.[92] Finally, expression profile analysis may lead to the discovery of novel markers expressed by different types of cancer. Some of these may encode secreted proteins that may prove important in early detection, whereas others may constitute novel chemotherapeutic drug targets.

RATIONAL DRUG DESIGN AND SMALL MOLECULES

The concept of therapeutic approaches based on common drug targets is best illustrated by the success of imatinib mesylate (Gleevec), a small-molecule inhibitor of specific protein kinases [*see Figure 15*].[13] Imatinib was initially identified by its ability to inhibit the binding of ATP to the kinase domain of the platelet-derived growth factor receptor (PDGFR); it was subsequently found to also inhibit the Abelson kinase (ABL) and the c-Kit kinase. The effectiveness of imatinib in treating CML is derived from the underlying chromosomal translocation that results in the chimeric protein bcr-abl [*see* Gene Amplification and Chromosomal Translocations, *above*]. Constitutive activation of the abl kinase by this rearrangement is essential for malignancy, and CML cells are extremely sensitive to disruption in this signaling pathway. bcr-abl thus provides an ideal drug target, because its function is essential for leukemic cell survival, whereas inhibition of abl in normal cells has few consequences. A very high frequency of complete remission induced by imatinib in chronic-phase CML has been reported, with very few side effects.[13] A remarkable coincidence is the effectiveness of imatinib in a histologically unrelated solid tumor, which is also dependent on a protein targeted by this drug. Gastrointestinal stromal tumors (GISTs) are rare tumors characterized by c-*kit* mutations that cause constitutive activation of the c-Kit kinase. Although GIST is typically refractory to standard chemotherapy, dramatic responses have been observed in patients treated with imatinib.[93] Similarly, cases of chronic myelomonocytic leukemia, in which the underlying genetic lesion involves rearrangement and activation of PDGFR, are also responsive to imatinib.[94] Thus, the rational design of small-molecule inhibitors directed at proteins that are critical components of cellular pathways may lead to novel and effective treatments of different types of cancer.

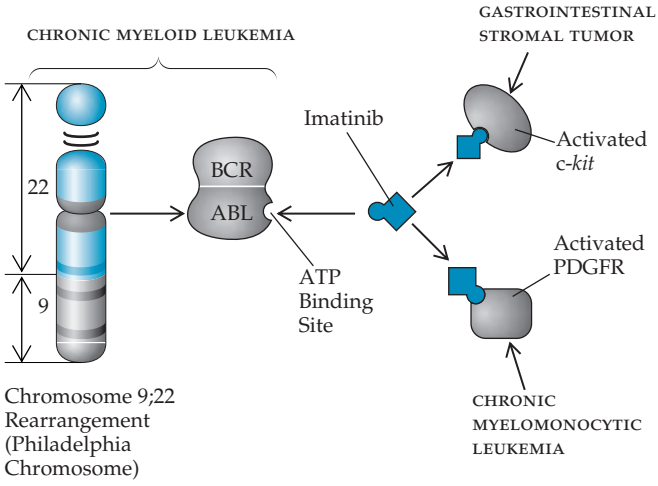


Figure 15 **Effectiveness of imatinib mesylate in targeting oncogenic proteins in diverse tumor types. Imatinib disrupts the binding of adenosine triphosphate (ATP) to the catalytic domain of three kinase proteins: platelet-derived growth factor receptor (PDGFR), ABL, and c-Kit. Tumor types in which activation of these kinase proteins is critical to tumor growth include chronic myeloid leukemia, in which the BCR-ABL chimeric fusion protein is derived from the rearrangement of chromosomes 9 and 22 (Philadelphia chromosome); gastrointestinal stromal tumor (GIST), containing activating mutations of c-*kit*; and chronic myelomonocytic leukemia (CMML), in which rearrangement of the PDGFR gene leads to its activation. In these three different malignancies, treatment with imatinib leads to rapid and sustained clinical responses.**

Clinical Implications of Genetic Analysis

The molecular genetic analysis of human cancer has elucidated mechanisms that underlie malignant transformation and the progression from early premalignant lesions to invasive carcinomas. The pace of discovery will probably accelerate as novel, high-throughput techniques facilitate the identification of genetic differences between normal cells and cancer cells. The emerging challenge will be the application of this molecular information to the clinical management of human cancer.[95] Although this field is still in its infancy, a number of approaches are under active investigation. Early diagnosis of malignancy currently relies on radiographic monitoring, such as mammography and chest x-ray; testing for secreted markers, such as prostate-specific antigen and CA125; and invasive procedures, including cystoscopy and colonoscopy. Molecular markers in cells that are sloughed in sputum, urine, or stool may eventually provide accurate and effective screening approaches.[96-99] Genetic lesions in malignant cells have also been used to facilitate diagnosis in cases in which histologic analysis is not definitive and to detect minimal residual disease that is undetectable by standard methods.[100,101] Analysis of tumors for genetic lesions that confer prognostic significance has already become standard in the management of many leukemias.[14-16] Molecular markers associated with response to chemotherapy, as well as the likelihood of metastatic spread, may eventually guide the treatment of solid tumors. Eventually, an understanding of the aberrant molecular pathways that drive cellular proliferation, angiogenic recruitment, and metastatic spread in different types of cancer may lead to the development of more specific and effective treatment regimens.

The appropriate use of germline mutation analysis may identify persons at increased risk for specific cancers who may benefit most from close monitoring.[102] In some cases, patients at high risk for developing cancer may benefit from prophylactic surgery. This is well established in carriers of *APC* mutations with familial polyposis coli, in whom prophylactic colectomy prevents the development of colorectal cancer.[103] Similarly, prophylactic thyroidectomy is commonly recommended in patients with multiple endocrine neoplasia.[104] More recently, prophylactic oophorectomy after completion of childbearing has been recommended for *BRCA1* and *BRCA2* mutation carriers, because it reduces the risk of both ovarian and breast cancers in these patients.[105,106] However, genetic predisposition to cancer in the general population is thought to result primarily from common genetic variations associated with a moderate increase in cancer risk (so-called low penetrance mutations), rather than from rare mutations that are associated with an extremely high risk of cancer.[107] The appropriate clinical options are less clear for individuals who carry such low-penetrance mutations in cancer susceptibility genes; although the effects of these mutations are apparent in population studies, their consequences are difficult to predict for any individual carrier. For instance, the *APC* I1307K mutation, which is present in 6% of the Ashkenazi Jewish population, appears to be associated with a twofold increase in risk of colorectal cancer.[108] The 1100delC mutation in the cell cycle checkpoint gene *CHK2*, which is present in 1% of the general population, has been linked to a similar increase in breast cancer risk.[109] Taken together, these and other genetic variants that will emerge from population genetics studies are likely to be important in targeting screening and chemopreventive options to individuals at increased risk for developing cancer.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Lichtenstein P, Holm NV, Verkasalo PK, et al: Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78, 2000

2. Stehelin D, Varmus HE, Bishop JM, et al: DNA related to transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature 260:170, 1976

3. Lowy DR, Willumsen BM: Function and regulation of ras. Annu Rev Biochem 62:851, 1993

4. Eng C, Clayton D, Schuffenecker I, et al: The relationship between specific net proto-oncogene mutations and disease phenotype in multiple endocrine neoplasia type 2: International RET Mutation Consortium analysis. JAMA 276:1575, 1996

5. Schimke RT: Gene amplification, drug resistance, and cancer. Cancer Res 44:1735, 1984

6. Brodeur GM, Seeger RC, Schwab M, et al: Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. Science 224:1121, 1984

7. Rabbitts TH: Chromosomal translocations in human cancer. Nature 372:143, 1994

8. Dalla-Favera R, Bregni M, Erikson J, et al: Human c-myc oncogene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc Natl Acad Sci USA 79:7824, 1982

9. Taub R, Kirsch I, Morton C, et al: Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cell. Proc Natl Acad Sci USA 79:7837, 1982

10. Shtivelman E, Lifshitz B, Gale RP, et al: Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature 315:550, 1985

11. Rowley JD: Biological implications of consistent chromosome rearrangements in leukemia and lymphoma. Cancer Res 44:3159, 1984

12. Daley GQ, Van-Etten RA, Baltimore D: Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science 247:824, 1990

13. Druker BJ, Sawyers CL, Kantarjian H, et al: Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. N Engl J Med 344:1038, 2001

14. Look AT: Oncogenic transcription factors in the human acute leukemias. Science 278:1059, 1997

15. Warrell RP Jr, Frankel SR, Miller WH Jr, et al: Differentiation therapy of acute promyelocytic leukemia with tretinoin (all-*trans*-retinoic acid). N Engl J Med 324:1385, 1991

16. Golub TR, Barker GF, Bohlander SK, et al: Fusion of the TEL gene on 12p13 to the AMLI gene on 21q22 in acute lymphoblastic leukemia. Proc Natl Acad Sci USA 92:4917, 1995

17. Delattre O, Zucman J, Melot T, et al: The Ewing family of tumors: a subgroup of small-round-cell tumors defined by specific chimeric transcripts. N Engl J Med 331:294, 1994

18. Adams JM, Cory S: Transgenic models of tumor development. Science 254:1161, 1991

19. Sinn E, Muller W, Pattengale PK, et al: Coexpression of MMTV/V-Ha-ras and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes in vivo. Cell 49:465, 1987

20. Land H, Parada LF, Weinberg RA: Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature 304:596, 1983

21. Whyte P, Buchkovich KJ, Horowitz JM, et al: Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. Nature 334:124, 1988

22. Ephrussi B, Davidson RL, Weiss MC, et al: Malignancy of somatic cell hybrids. Nature 224:1314, 1969

23. Saxon PJ, Srivastan ES, Stanbridge EJ: Introduction of human chromosome 11 via microcell transfer controls tumorigenic expression of HeLa cells. EMBO J 5:3461, 1986

24. Knudson AG Jr: Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820, 1971

25. Friend SH, Bernards R, Rogelij S, et al: A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature 323:643, 1986

26. Call KM, Glaser T, Ito CY, et al: Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell 60:509, 1990

27. Cavenee WK, Dryja TP, Phillips RA, et al: Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature 305:779, 1983

28. Weinberg RA: The retinoblastoma protein and cell cycle control. Cell 81:323, 1995

29. Hussussian CJ, Struewing JP, Goldstein AM, et al: Germline p16 mutations in familial melanoma. Nat Genet 8:15, 1994

30. Sherr CJ: Cancer cell cycles. Science 274:1672, 1996

31. Lane DP: Cancer: p53, guardian of the genome. Nature 358:15, 1992

32. El-Deiry WS, Tokino T, Velculescu VE, et al: WAF1, a potential mediator of p53 tumor suppression. Cell 75:817, 1993

33. Hollstein M, Sidransky D, Vogelstein B, et al: p53 mutations in human cancers. Science 253:49, 1991

34. Harris CC, Hollstein M: Clinical implications of the p53 tumor suppressor gene. N Engl J Med 329:1318, 1993

35. Donehower LA, Harvey M, Slagle BL, et al: Mice deficient of p53 are developmentally normal but susceptible to spontaneous tumors. Nature 356:215, 1992

36. Malkin D, Li FP, Strong LC, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 250:1233, 1990

37. Debbas M, White E: Wild-type p53 mediates apoptosis by E1A, which is inhibited by E1B. Genes Dev 7:546, 1993

38. Oliner JD, Kinzler KW, Meltzer PS, et al: Amplifications of a gene encoding a p53-as-

sociated protein in human sarcoma. Nature 358:80, 1992

39. Montes de Oca Luna R, Wagner DS, Lozano G: Rescue of early embryonic lethality in mdm2-deficient mice by deletion of p53. Nature 378:203, 1995

40. Kastan MB, Lim DS: The many substrates and functions of ATM. Nat Rev Mol Cell Biol 1:79, 2000

41. Quelle DE, Zindy F, Ashmun RA, et al: Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. Cell 83:993, 1995

42. Sherr CJ: Tumor surveillance via the ARF-p53 pathway. Genes Dev 12:2984, 1998

43. Kaghad M, Bonnet H, Yang A, et al: Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. Cell 90:809, 1997

44. Osada M, Ohba M, Kawahara C, et al: Cloning and functional analysis of human p51, which structurally and functionally resembles p53. Nat Med 4:839, 1998

45. Kastan MB, Zhan Q, el-Deiry WS, et al: A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. Cell 71:587, 1992

46. Savitsky K, Bar-Shira A, Gilad S, et al: A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268:1749, 1995

47. Miki Y, Swensew J, Shattuck-Eidens D, et al: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66, 1994

48. Wooster R, Bignell G, Lancaster J, et al: Identification of breast cancer susceptibility gene BRCA2. Nature 378:789, 1995

49. Scully R, Chen J, Plug A, et al: Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell 88:265, 1997

50. Joenje H, Patel KJ: The emerging genetic and molecular basis of Fanconi anaemia. Nat Rev Genet 2:446, 2001

51. Haber DA: Wilms tumor. The Genetic Basis of Human Cancer. Vogelstein B, Kinzler K, Eds. McGraw Hill Book Co, New York, 2002, p 403

52. Iliopoulos O, Kaelin WG Jr: The molecular basis of von Hippel-Lindau disease. Mol Med 3:289, 1997

53. Polakis P: The adenomatous polyposis coli (APC) tumor suppressor. Biochim Biophys Acta 1332:F127, 1997

54. Hahn SA, Schutte M, Hoque AT, et al: DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science 271:350, 1996

55. Hahn H, Wicking C, Zaphiropoulous PG, et al: Mutations of the human homolog of *Drosophila* patched in the nevoid basal cell carcinoma syndrome. Cell 85:841, 1996

56. McCormick F: Ras signaling and NF1. Curr Opin Genet Dev 5:51, 1995

57. Gusella JF, Ramesh V, MacCollin M, et al: Neurofibromatosis 2: loss of Merlin's protective spell. Curr Opin Genet Dev 6:87, 1996

58. Parsons R: Phosphatases and tumorigenesis. Curr Opin Oncol 10:88, 1998

59. Kolodner RD: Mismatch repair: mechanisms and relationship to cancer susceptibility. Trends Biochem Sci 20:397, 1995

60. Markowitz S, Wang J, Myeroff L, et al: Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 268:1336, 1995

61. Lynch HT, Smyrk T, Lynch JF: Molecular genetics and clinical pathology features of hereditary nonpolyposis colorectal carcinoma (Lynch syndrome): historical journey from pedigree anecdote to molecular genetic confirmation. Oncology 55:103, 1998

62. Aaltonen LA, Salovaara R, Kristo P, et al: Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 338:1481, 1998

63. Kinzler KW, Vogelstein B: Lessons from hereditary colorectal cancer. Cell 87:159, 1996

64. Korsmeyer SJ: Regulators of cell death. Trends Genet 11:101, 1995

65. Cleary ML, Smith SD, Sklar J: Cloning and structural analysis of cDNAs for bcl-2 and a hybrid bcl-2/immunoglobulin transcript resulting from the t(14;18) translocation. Cell 47:19, 1986

66. Rampino N, Yamamoto H, Ionov Y, et al: Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science 275:967, 1997

67. Hayflick L, Moorhead PS: The serial cultivation of human diploid cell strains. Exp Cell Res 25:586, 1961

68. Greider CW: Telomeres and senescence: the history, the experiment, the future. Curr Biol 8:R178, 1998

69. Meyerson M, Counter CM, Eaton EN, et al: hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell 90:785, 1997

70. Nakamura TM, Morin GB, Chapman KB, et al: Telomerase catalytic subunit homologs from fission yeast and human. Science 277:955, 1997

71. Kolquist KA, Ellisen LW, Counter CM, et al: Expression of TERT in early premalignant lesions and a subset of cells in normal tissues. Nat Genet 19:182, 1998

72. Liotta LA, Tryggvason K, Garbisa S, et al: Metastatic potential correlates with enzymatic degradation of basement membrane collagen. Nature 284:67, 1980

73. Gunthert U, Hofmann M, Rudy W, et al: A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. Cell 65:13, 1991

74. Folkman J: Fighting cancer by attacking its blood supply. Sci Am 275:150, 1996

75. O'Reilly MS, Holmgren L, Shing Y, et al: Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. Cell 79:315, 1994

76. O'Reilly MS, Boehm T, Shing Y, et al: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88:277, 1997

77. Boehm T, Folkman J, Browdar T, et al: Antiangiogenic therapy of experimental cancer does not induce acquired drug resistance. Nature 390:404, 1997

78. Arap W, Pasqualini R, Ruoslahti E: Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. Science 279:377, 1998

79. Kinsella AR, Smith D: Tumor resistance to antimetabolites. Gen Pharmacol 30:623, 1998

80. Ling V: Multidrug resistance: molecular mechanisms and clinical relevance. Cancer Chemother Pharmacol 40:S3, 1997

81. Grogan TM, Spier CM, Slamon SE, et al: P-glycoprotein expression in human plasma cell myeloma: correlation with prior chemotherapy. Blood 81:490, 1993

82. Leith CP, Kopecky KJ, Godwin J, et al: Acute myeloid leukemia in the elderly: assessment of multidrug resistance (MDR1) and cytogenetics distinguishes biologic subgroups with remarkably distinct responses to standard chemotherapy. A Southwest Oncology Group Study. Blood 89:3323, 1997

83. Lowe S, Bodis S, McClatchey A, et al: Status and the efficacy of cancer therapy in vivo. Science 266:807, 1994

84. Bischoff JR, Kirn DH, Williams A, et al: An adenovirus mutant that replicates selectively in p53-deficient human tumor cells. Science 274:373, 1996

85. Steegenga WT, Riteco N, Bos JL: Infectivity and expression of the early adenovirus proteins are important regulators of wild-type and DeltaE1B adenovirus replication in human cells. Oncogene 18:5032, 1999

86. Johnson L, Shen A, Boyle L, et al: Selectively replicating adenoviruses targeting deregulated E2F activity are potent, systemic antitumor agents. Cancer Cell 1:325, 2002

87. Lipshutz RJ, Fodor SP, Gingeras TR, et al: High density synthetic oligonucleotide arrays. Nat Genet 21(1 suppl):20, 1999

88. Brown PO, Botstein D: Exploring the new world of the genome with DNA microarrays. Nat Genet 21(1 suppl):33, 1999

89. Velculescu VE, Zhang L, Vogelstein B, et al: Serial analysis of gene expression. Science 270:484, 1995

90. Rosenwald A, Wright G, Chan WC, et al: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937, 2002

91. Van't Veer LJ, Dai H, van de Vijver MJ, et al: Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530, 2002

92. Lakhani SR, Ashworth A: Microarray and histopathological analysis of tumours: the future and the past? Nat Rev Cancer 1:151, 2001

93. Demetri GD, von Mehren M, Blanke CD, et al: Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347:472, 2002

94. Apperley JF, Gardembas M, Melo JV, et al: Response to imatinib mesylate in patients with chronic myeloproliferative diseases with rearrangements of the platelet-derived growth factor receptor beta. N Engl J Med 347:481, 2002

95. Haber D, Fearon E: The promise of cancer genetics. Lancet 351:SII1, 1998

96. Sidransky D, Tokino T, Hamilton SR, et al: Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. Science 256:102, 1992

97. Mao L, Schoenberg MP, Scicchitano M, et al: Molecular detection of primary bladder cancer by microsatellite analysis. Science 271:659, 1996

98. Steiner G, Schoenberg MP, Linn JF, et al: Detection of bladder cancer recurrence by microsatellite analysis of urine. Nat Med 3:621, 1997

99. Schneider A, Borgnat S, Lang H, et al: Evaluation of microsatellite analysis in urine sediment for diagnosis of bladder cancer. Cancer Res 60:4617, 2000

100. Brennan JA, Mao L, Hruban RH, et al: Molecular assessment of histopathological staging in squamous-cell carcinoma of the head and neck. N Engl J Med 332:429, 1995

101. Hayashi N, Arakawa H, Nagase H, et al: Genetic diagnosis identifies occult lymph node metastases undetectable by the histopathological method. Cancer Res 54:3853, 1994

102. Ponder B: Genetic testing for cancer risk. Science 278:1050, 1997

103. Kinzler KW, Vogelstein B: Colorectal tumors. The Genetic Basis of Human Cancer. Vogelstein B, Kinzler K, Eds. McGraw Hill Book Co, New York, 2002, p 583

104. Brandi ML, Gagel RF, Angeli A, et al: Guidelines for diagnosis and therapy of MEN type 1 and type 2. J Clin Endocrinol Metab 86:5658, 2001

105. Kauff ND, Satagopan JM, Robson ME, et al: Risk-reducing salpingo-oophorectomy in women with BRCA1 or BRCA2 mutations. N Engl J Med 346:1609, 2002

106. Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. Prevention and Observation of Surgical End Points Study Group. N Engl J Med 346:1616, 2002

107. Ponder BA: Cancer genetics. Nature 411:336, 2001

108. Laken SJ, Petersen GM, Gruber SB, et al: Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat Genet 17:79, 1997

109. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al: Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 31:55, 2002

Acknowledgments

Figures 1 through 7 and 9 through 15 Seward Hung.

Figure 8 Janet Betries.

III TUMOR IMMUNOLOGY

Robert F. Todd III, m.d., ph.d.
Bruce G. Redman, d.o.

The development of cancer represents a failure of immune surveillance, because the immune system has the capacity to recognize tumor-associated antigens and develop specific T cell responses to those antigens. The ability to intervene and enhance the immune system to achieve a beneficial antitumor response remains an area of intense clinical research. Considerable progress has been made in expanding our knowledge of the targets for an immune response and about the full repertoire of cellular and humoral constituents involved in the generation of an effective antitumor response. Tumor cells display a variety of mechanisms by which they evade immune detection and destruction and render the immune response ineffective. With a more complete understanding of these escape mechanisms, clinical investigators are devising strategies to enhance the development of a robust immune response in the tumor-bearing host (active tumor immunity) or, alternatively, by the adoptive transfer of activated effector cells or tumor-specific antibodies into the tumor-bearing host (passive tumor immunity). Several strategies hold promise for substantial therapeutic benefit. Finally, antibodies that recognize tumor-associated antigens can aid in the pathologic diagnosis of cancer and facilitate the staging of cancer in vivo and the detection of recurrent cancer.

Overview of the Immune Response

The cellular and humoral arms of the immune response feature lymphocytes and antibodies, respectively, but extend beyond those elements. The cellular response consists of multiple components, including different subsets of lymphocytes (helper T cells, cytotoxic T cells, B cells, and natural killer [NK] cells) and antigen-presenting cells (APCs) that include blood monocytes, tissue macrophages, and dendritic cells. Different components of the immune system communicate with one another by direct cell-to-cell interaction via various adhesion molecules and receptors on the cell surface or through secreted chemicals (usually proteins) called cytokines that circulate and bind to specific receptors on effector cells. Cytokines can be either immune stimulatory or immune inhibitory; examples of cytokines include interleukins and interferons.

The cellular and soluble constituents of the immune response work in concert to recognize foreign (nonself) antigens (e.g., proteins, glycoproteins, and glycolipids) that are expressed by or secreted from tumor cells. The afferent limb of the immune response (immune recognition) subsequently triggers a highly specific response by the effector limb, in which other immune cells and their soluble products attack tumor cells bearing the same antigens recognized by the afferent limb. The success or failure of the afferent limb in recognizing tumor cells and of the effector limb in attacking them is influenced by a variety of mitigating factors.

The identification of the cellular constituents of the immune response and knowledge of their functions have been aided by the development of monoclonal antibodies created by immunization of mice against human immune cells. Each monoclonal antibody recognizes a single glycoprotein antigen that reflects the expression of a unique cell surface receptor. By convention, these receptors and the cells that express them have been assigned a cluster designation (CD) number (e.g., CD3, CD4, CD8).

Host Immune Response to Cancer

The existence of a host immune response to cancer is supported by the following observations:

1. Histologic analysis of excised human and animal tumors has demonstrated varying degrees of immune cellular infiltration (lymphocytes and APCs), which suggests the recruitment of these cells in response to neoplastic proliferation. When analyzed, the activity of these tumor-infiltrating lymphocytes is often found to be specific for the autologous tumor, with little to no activity against unrelated tumor targets. Accompanying this mononuclear infiltration of tumors is the elaboration of various cytokines that are associated with an ongoing immune response. T cells that exhibit specific reactivity against autologous tumor have been isolated from patients with melanoma, breast cancer, ovarian cancer, and colorectal cancer.[1]
2. Long-term remissions are induced in small numbers of patients who receive some form of immunologic therapy for their advanced cancers.
3. There are documented (albeit infrequent) reports of spontaneous remissions in patients with melanoma and renal cell carcinoma that are believed to be immune mediated.

The possible existence of immune surveillance mechanisms that prevent the development of cancer is supported by the finding that immunodeficient individuals and patients undergoing long-term treatment with immunosuppressive drugs are at greater risk for cancer than the general population. Experimental evidence that directly links immune mechanisms to the defense against cancer comes from classic experiments in which immunized mice rejected chemically induced syngeneic tumors.[2] Naive syngeneic mice (i.e., mice not previously exposed to tumor) were protected against tumor growth by immunization with killed tumor cells administered before challenge with viable tumor cells [*see Figure 1*]. The specificity of this immune response was demonstrated by the lack of protective effect of prior immunization with killed cells from an unrelated tumor. The important role of lymphoid cells in this immune response was demonstrated by the protective effect of spleen cells from a tumor-bearing mouse administered to a naive host (adoptive transfer of immune cells) before tumor challenge with live tumor cells. In these experiments, the cytotoxic T cell population ($CD8^+$ T cells) within the spleens of tumor-bearing mice demonstrated unique protective activity upon transfer. Transfer of helper T cells ($CD4^+$ T cells), B cells, or serum-derived antibodies did not confer protection in this model. The results of these classic experiments, in which syngeneic tumors could be rejected in naive hosts receiving prior immunization or by adoptive transfer of immune cells, support the existence of tumor-associated antigens in animal models and form the basis of much of the experimental work in human tumor immunology.

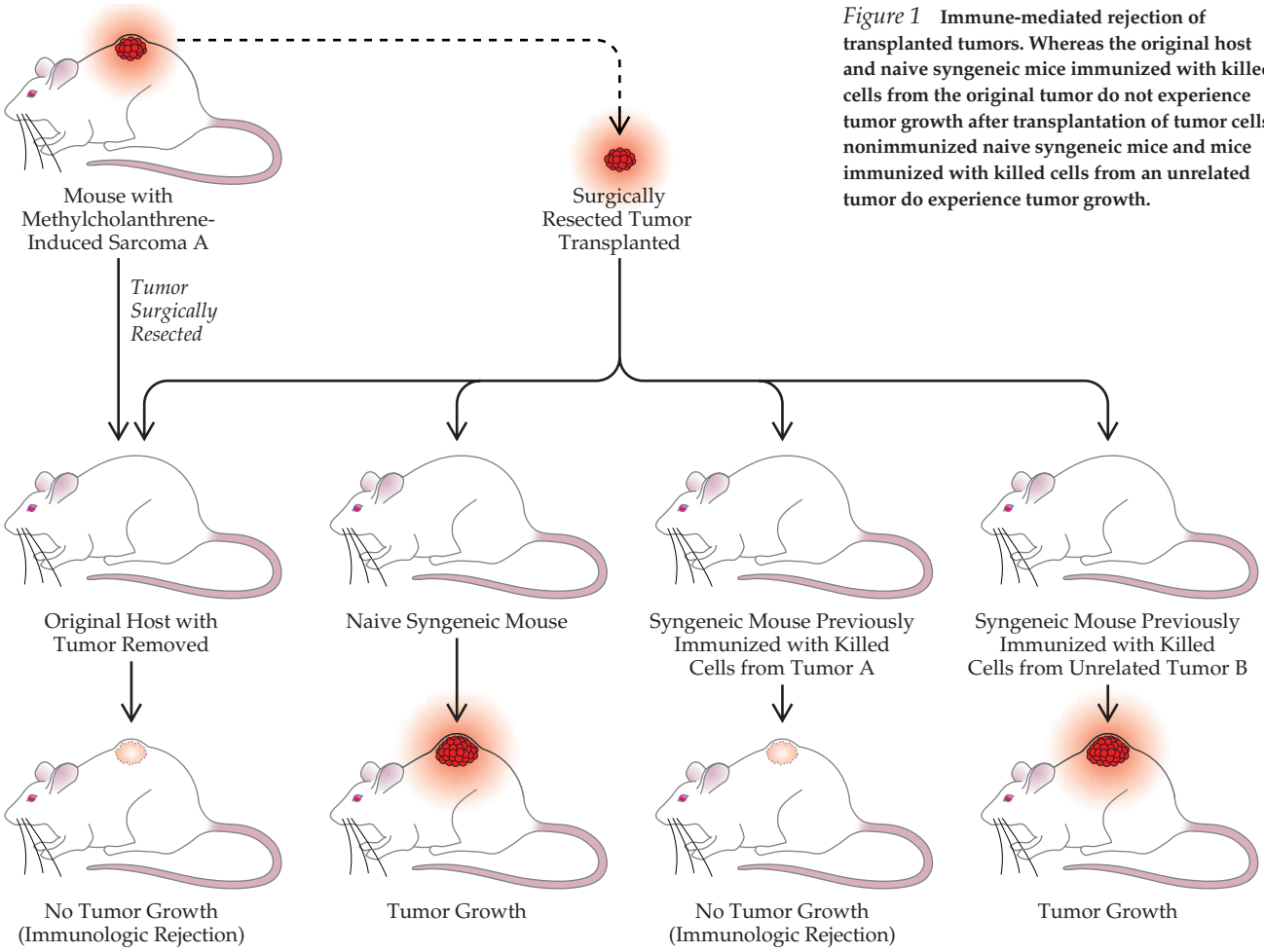


Figure 1 **Immune-mediated rejection of transplanted tumors. Whereas the original host and naive syngeneic mice immunized with killed cells from the original tumor do not experience tumor growth after transplantation of tumor cells, nonimmunized naive syngeneic mice and mice immunized with killed cells from an unrelated tumor do experience tumor growth.**

Targets for Immune Response to Cancer

The ability to immunize naive animals against cancer led to the concept of tumor-specific or tumor-associated antigens that serve as the target of an inducible effector response. The development of strategies to enhance this immune response against cancer has depended upon the further identification and characterization of these tumor-related antigens.[3] The work of many investigators over the past decade has confirmed the existence of at least five classes of potential tumor-associated antigens that are recognized by and stimulate T cells: oncoviral proteins, tumor-associated antigens, mutated or overexpressed oncogene or tumor-suppressor gene antigens, differentiation (or lineage-specific) antigens, and abnormal posttranslational modification of self-proteins.

ONCOVIRAL PROTEINS

Oncoviral proteins represent cellular antigens encoded by the genomes of oncogenic viruses. Examples of oncoviral proteins include the human papillomavirus E6 (HPV-E6) and HPV-E7 antigens, found in cervical carcinoma; and the Epstein-Barr virus (EBV) EBNA-1 antigen, found in Burkitt lymphoma and nasopharyngeal carcinoma.

TUMOR-ASSOCIATED ANTIGENS

Tumor-associated antigens (also referred to as tumor-testis antigens) are proteins that are normally expressed during the course of embryonic development and in the human adult testis; they become abnormally expressed by the cancer cells of adult individuals. Examples of tumor-associated antigens include the MAGE-1 and MAGE-3 proteins, which are expressed by a variety of tumor cells, including melanoma, glioma, and breast carcinoma.

MUTATED OR OVEREXPRESSED ONCOGENE OR TUMOR-SUPPRESSOR GENE ANTIGENS

Mutated or overexpressed oncogene or tumor-suppressor gene antigens represent the protein products of mutated or overexpressed cellular oncogenes or tumor-suppressor genes found in a variety of tumor cells. Examples of this class of tumor-associated antigens are $p21^{ras}$ (expressed in a number of carcinomas), the p210 product of the *bcr-abl* translocation found in chronic myelogenous leukemia, the cyclin-dependent kinase 4 (CdK4) and β-catenin proteins found in melanoma, the HER-2/neu protein found in breast carcinoma and other cancers, the caspase-8 protein expressed in certain squamous cell carcinomas, and the p53 tumor-suppressor gene product found in multiple tumors.

DIFFERENTIATION ANTIGENS

Differentiation (or lineage-specific) antigens are proteins that are normally expressed in a tissue-specific fashion by normal cells but are coexpressed by tumor cells derived from the normal

host tissue. Examples of differentiation antigens are tyrosinase, GP100, MART-1 antigens coexpressed by normal melanocytes and melanoma cells, and cell membrane immunoglobulin in a specific B cell clone.

ABNORMAL POSTTRANSLATIONAL MODIFICATION OF SELF-PROTEINS

Self-proteins that have undergone abnormal posttranslational modification represent mutated forms of normal protein products; these modified self-proteins give rise to unique tumor-associated carbohydrate epitopes. An example of this class of tumor antigens is the MUC-1 antigen (featuring underglycosylated mucin), which is expressed by breast and pancreatic carcinomas.

In addition to the antigens that may be the targets of a T cell–directed immune response, there are other antigens that may be recognized by T cells as well as by antibodies resulting from deliberate immunization against these antigens. These targets of an antibody-mediated immune response include (1) tissue-specific differentiation antigens that, like the lineage-specific antigens, represent protein products shared between tumor cells and the tissues from which they are derived (e.g., CD20 and the surface immunoglobulin [Ig] idiotype, which are expressed by B cell lymphomas; and the prostate-specific antigen [PSA], expressed by prostate carcinoma cells); (2) oncofetal antigens that, like the tumor-associated antigens, represent self-proteins normally expressed during embryonic development but that are found on tumor cells (e.g., the carcinoembryonic antigen [CEA], expressed by multiple carcinomas, and α-fetoprotein [AFP], expressed by hepatocellular carcinoma and germ cell tumors); and (3) altered or overexpressed glycolipid and glycoprotein antigens. These include the gangliosides GM_2 and GD_2 found on melanomas and neuroblastomas; and the mucin antigens CA125, CA19-9, and MUC-1, which are predominantly expressed by ovarian, pancreatic, and breast carcinomas, respectively.

Effector Mechanisms in the Anticancer Immune Response

The immune response against cancer includes both the cellular response and the humoral response.

CELLULAR IMMUNE RESPONSE

The cellular effector response involves activity by five major cellular constituents. (1) $CD8^+$ cytotoxic T cells recognize tumor-associated antigens that are presented in association with major histocompatibility (MHC) class I molecules [*see Figure 2*]. (2) $CD4^+$ helper T cells recognize tumor-associated antigens that are presented in association with MHC class II molecules leading to cytokine release (help signals) for the generation and activation of cytotoxic T cells. (3) NK cells may kill tumor cells in a non–MHC-dependent fashion. Rather than binding directly to tumor-associated antigens, NK cells may be targeted to antibody-coated tumor cells via recognition and attachment to the Fc portion of the antibody. (4) Mononuclear phagocytes (macrophages), like NK cells, may be targeted to antibody-coated tumor cells, where their cytotoxic activity may depend upon the release of destructive proteases, cytokines (e.g., tumor necrosis factor [TNF]), and reactive oxidative intermediates (O_2^-). (5) Dendritic cells (DCs) are important APCs that can present antigen to both $CD4^+$ and $CD8^+$ T cells and are able to stimulate a naive T cell response (i.e., stimulation of a T cell response to an antigen to which the T cell has not been previously exposed).

The recognition of tumor target cells or tumor APCs depends on a direct binding interaction mediated by several receptor species [*see Figure 2*].[4] The specificity of binding is conferred by the T cell receptor (TCR), in association with CD8 or CD4 surface glycoproteins, which recognizes antigen as pre-

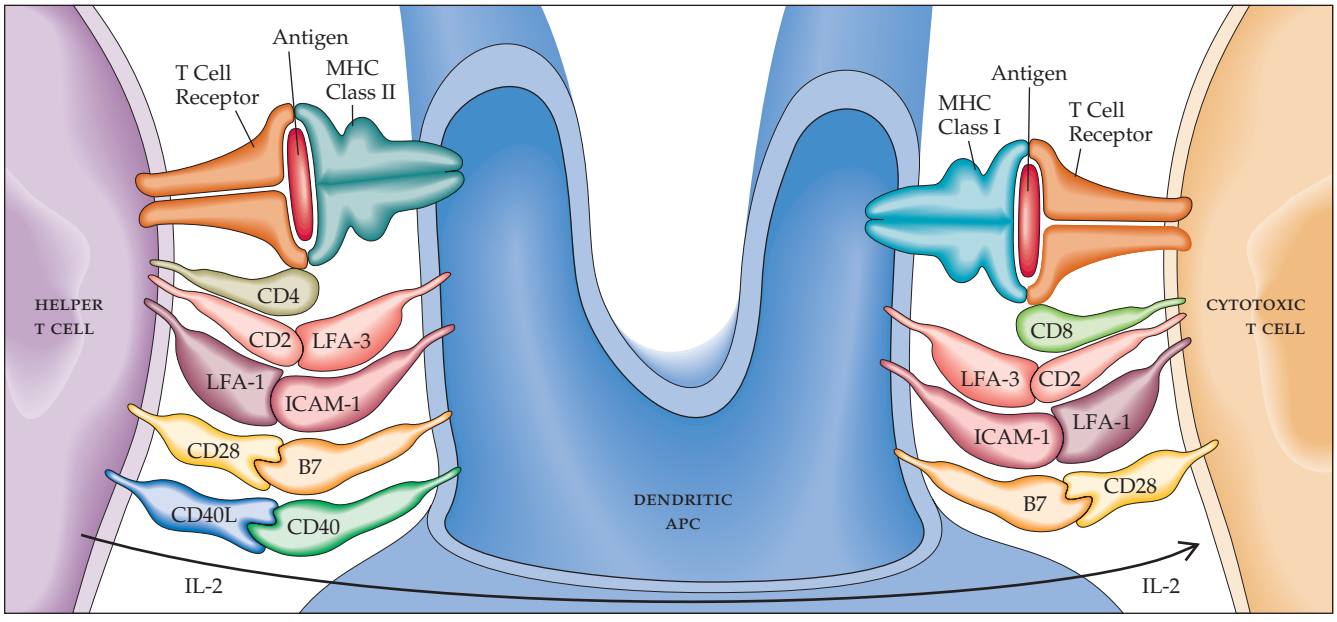


Figure 2 **T cell interaction with a dendritic antigen-presenting cell (APC). Helper T cells recognize MHC class II–associated antigens. Cytotoxic T cells recognize MHC class I–associated antigens. Release of interleukin-2 (IL-2) by the helper T cell provides a so-called help signal to the cytotoxic T cells. (ICAM—intercellular adhesion molecule; LFA—leukocyte function–associated antigen)**

sented by either class I (recognized by $CD8^+$ T cells) or class II (recognized by $CD4^+$ T cells) MHC antigens. Antigen recognition and concomitant T cell activation also depend on the binding of pairs of receptor and costimulatory molecules, including leukocyte function–associated antigen–1/intercellular adhesion molecule–1, CD28/B7, and CD2/leukocyte function–associated antigen–3. Helper T cell activation results in secretion of interleukin-2 (IL-2), which provides a necessary help signal to cytotoxic T cells. T cell binding to antigen on a target cell that does not express the requisite accessory molecules can lead to T cell apoptosis or anergy, thus preventing an immune response.

After effective membrane binding between cytotoxic T cells and tumor target cells, actual tumor cell cytolysis is accomplished by one of two general mechanisms[5]: (1) a perforin-dependent mechanism, in which killer cell enzymes (granzymes) are released from cytotoxic T cells and gain access to target cells via perforin pores, endocytosis, or granzyme receptor–mediated uptake, leading to target cell membrane damage with subsequent necrosis and apoptosis; or (2) a newly recognized mechanism of target cell cytolysis that depends on the recognition of the membrane protein Fas (expressed by certain tumor cells) by the Fas ligand expressed by killer T cells. The binding of Fas and its ligand leads to the generation of proapoptotic death signals to the target cell nucleus, resulting in DNA fragmentation and apoptosis.

HUMORAL IMMUNE RESPONSE

The recognition of tumor-associated antigens by $CD4^+$ helper T cells may also elicit a B cell humoral response leading to the development of antibodies that recognize tumor-associated antigens. Unfortunately, despite this mechanism of antitumor immunity, the results of animal experiments suggest that naturally occurring antibodies play little role in effective antitumor response. However, the administration of manufactured antibodies targeted to tumor-associated antigens may be quite useful in eliciting an effective antitumor response. Specifically, these antibodies may facilitate antibody-dependent cellular cytotoxicity (ADCC) by NK cells and macrophages, or they may be employed to kill tumor cells by direct mechanisms (e.g., induction of apoptosis or fixation of complement) or indirect mechanisms. In the latter case, antibodies may be conjugated (chemically linked) with cellular toxins (e.g., ricin or diphtheria toxin) or gamma radiation–emitting radionuclides; binding of the antibodies to tumor cells results in the death of those cells from internalization of the toxin or exposure to ionizing radiation.

Evading Antitumor Immune Responses

Despite the fact that the immune system has the potential to recognize tumor-associated antigens and to marshall an effective cytotoxic response, this response often fails to prevent the local growth and distant spread of cancers. A better understanding of the mechanisms by which tumor cells can evade a host response is emerging. Tumor cells may evade detection by multiple mechanisms, all of which are probably operational to some extent.[6] Tumor cells may express only low levels of tumor-associated antigens in cryptic sites that are covered by glycocalyx molecules, or they may undergo antigenic modulation by shedding tumor antigens. Tumor cells have been shown to express low levels or nondetectable amounts of MHC molecules or appropriate costimulatory molecules (e.g., B7) that are required for an effective immune response. In other circumstances, tumor cells that can stimulate an effective immune response may have already been eliminated, leaving only those cells that can evade an immune response. Finally, tumor cells have been found to secrete various soluble factors (e.g., transforming growth factor–β [TGF-β], IL-10, and Fas ligand) that have been found to be immunosuppressive.

Enhancement of Active Host Immune Responses

Two major approaches have been undertaken to counter the mechanisms by which tumor cells evade immune recognition: (1) more effective stimulation of a host immune response in tumor-bearing patients (active immunity) and (2) the adoptive transfer of cellular or humoral immunity to the tumor-bearing host (passive immunity).

STIMULATION OF A HOST IMMUNE RESPONSE

Nonspecific Immune Stimulation

Previous attempts at nonspecific stimulation of the immune response, such as direct intratumor injections of bacillus Calmette-Guérin (BCG) or other immune adjuvant compounds, often provoked local immune responses (with occasional regressions of injected tumors) but without consistent systemic effects on noninjected sites of disease. The local administration of BCG has been successful in treating and preventing the recurrence of noninvasive bladder cancer [*see 12:XIV Bladder, Renal, and Testicular Cancer*].[7]

Vaccine Therapy

A major strategy for generating a systemic immune response against cancer is the development of tumor vaccine therapy.[8] The goal of such therapy is to present tumor-associated antigens in association with appropriate costimulatory molecules in a way that circumvents the tumor-evasive mechanisms that would ordinarily lead to tolerance (a state of immune paralysis). Approaches to immunizing a tumor-bearing host with tumor-associated antigens have ranged from vaccination with intact tumor cells to administration of selected immunogenic peptide vaccines.

Vaccination with intact unmodified tumor cells Even with adjuvant compounds, vaccination with intact unmodified tumor cells has generally been ineffective in triggering a clinically relevant immune response. If tumor cells escape immune recognition in situ, immunization with cells from the tumor carries the same limitations.

Vaccination with genetically modified tumor cells In this strategy, host-derived tumor cells are transfected with genes encoding for allogeneic MHC genes, costimulatory molecules (e.g., B7), or immunomodulatory cytokines. These genetically modified tumor cells bearing MHC or costimulatory molecules compensate for the lack of these molecules in vivo. Tumor cells transfected with genes encoding for immunomodulatory cytokines may result in the release of these soluble factors (e.g., granulocyte-macrophage colony-stimulating factor [GM-CSF], IL-4, and IL-2), which recruit dendritic cell precursors or helper T cells into vaccination sites to facilitate the immune recognition of tumor antigen and the subsequent generation of an effector cytotoxic T lymphocyte (CTL) response. Rather than relying on the cumbersome technique of transfecting autologous tumor cells in vitro, a more practical approach is to admix autologous tumor cells with

generic transduced bystander cells or biopolymer microspheres containing cytokines or to use in situ intratumoral injection of genetic material. An alternative approach is to employ standardized (allogeneic) gene-transduced tumor cell lines that display broadly expressed antigens as a source of tumor vaccine.

Vaccination with tumor antigens or peptides The goal of this strategy is to immunize patients with tumor antigens or immunodominant peptides derived from those antigens. An example is the direct immunization with a peptide derived from the MART-1 antigen as a means of provoking an immune response from patients with melanoma. Several class I (HLA-A2)–restricted, highly immunogenic peptides of MART-1 and GP-100 have been created.[9,10] In one clinical trial, 42 patients with metastatic melanoma were vaccinated with a synthetic peptide derived from GP-100.[11] Eleven patients receiving the GP-100–derived peptide without IL-2 demonstrated no objective responses, but 13 of 31 patients (42%) receiving the peptide plus systemic IL-2 achieved a clinically relevant tumor shrinkage. The results of this early investigation are consistent with the concept that vaccines for treatment of cancer will need to be used along with adjuvants to enhance immunologic response (i.e., IL-2 or other signaling molecules).

A similar vaccination strategy was evaluated in 10 patients with multiple myeloma.[12] The vaccine consisted of autologous immunoglobulin idiotype (Id) protein purified from serum or urine, coupled to keyhole limpet hemocyanin (KLH), a protein commonly utilized as an adjuvant to augment an immune response. The patients received subcutaneous injections of Id-KLH conjugates along with low doses of GM-CSF or IL-2. Although the clinical impact of this immunization remains to be determined, 8 of 10 patients developed an Id-specific delayed-type hypersensitivity (DTH) reaction, indicating the immunocompetence of multiple myeloma patients for specific immunization.

Vaccination with recombinant viral or bacterial vaccines Another strategy for stimulating an immune response to a specific tumor antigen is through the use of recombinant viral or bacterial vaccines. In this approach, the viral or bacterial genome is modified to include a gene encoding the relevant tumor antigen with or without genes encoding for costimulatory molecules (e.g., B7). The recombinant viral or bacterial agents then act as vectors with the capacity of infecting APCs or tumor cells. This strategy also takes advantage of the natural adjuvant effect and intrinsic immunogenicity of viruses and bacteria.

Naked DNA vaccines Yet another strategy involves vaccination with so-called naked DNA encoding for a tumor-associated antigen. These nucleic acid vaccines may be less potent than vaccines comprising recombinant virus or bacteria, because unlike the latter, they have no mechanism for amplification in the host individual.

Dendritic cell vaccines The use of DCs as a major constituent of a tumor vaccine helps ensure effective presentation of antigen to $CD4^+$ and $CD8^+$ T cells.[13] Various approaches have been devised to load DCs with tumor-associated antigen, including (1) ex vivo incubation of autologous DCs with tumor antigen or tumor cell lysates before vaccination, (2) fusion of DCs with whole autologous tumor cells, and (3) ex vivo transduction with RNA or replication-defective recombinant viral vectors carrying tumor-associated antigens within their genome. An example of DC-based vaccination strategy has been evaluated in B cell non-Hodgkin lymphoma.[14] Autologous DCs were cocultured with idiotypic protein from the patient's malignant B cells and then utilized as a vaccine. Of the 35 patients vaccinated, the majority developed T cell or humoral anti-Id responses. Among the 28 patients with measurable disease at the time of vaccination (or subsequent booster injections), 11 experienced clinically significant tumor regressions, including four with complete responses (i.e., no clinically detectable tumor).

Systemic Administration of Immunostimulatory Cytokines

Another approach to stimulating an immune response in a patient with cancer is to administer recombinant cytokines that may enhance an otherwise ineffective immune response. An example of this strategy is the use of high-dose IL-2, which enhances the immune response by stimulating T cell and NK cell activity and increases the serum concentration of other immunostimulatory cytokines, such as TNF, IL-1, and interferon gamma. Systemic high-dose IL-2 has been extensively evaluated in the treatment of metastatic melanoma and renal cell carcinoma and is approved by the Food and Drug Administration for the treatment of these two diseases. The approval of high-dose IL-2 as therapy for metastatic melanoma is based in part on the collective experience of 270 patients who were entered into eight clinical trials in the United States. Doses of 600,000 to 720,000 IU/kg I.V. were administered every 8 hours over two 5-day periods, with a 6- to 9-day rest in between the 5-day cycles. Partial tumor shrinkage occurred in 10% of the treated patients, with another 6% demonstrating complete tumor shrinkage.[15] For the patients who experienced complete response, the median response duration was not reached, and 59% of these patients remained progression free. Similar response rates and durations of response were seen in patients with metastatic renal cell carcinoma.[16] The major drawback to this form of therapy is its associated toxicity, which includes the development of a capillary leak/shock syndrome.

Interferon alfa, which has direct antiproliferative effects and causes increased MHC class I expression, has been effective in inducing tumor regressions in a minority of patients with renal cell carcinoma, melanoma, hairy-cell leukemia, and lymphoma; it has also been effective in the treatment of chronic myelogenous leukemia. Other immunomodulatory cytokines that have been evaluated in clinical trials include TNF-α and interferon gamma, both of which have shown antitumor activity in vitro but are poorly tolerated or minimally active in vivo.

PASSIVE IMMUNITY AGAINST CANCER BY ADOPTIVE TRANSFER OF CELLS OR ANTIBODY

The strategy of adoptive transfer of preimmune cells or antibody to a tumor-bearing patient circumvents the potential limitation of stimulating an effective immune response in favor of boosting the host with previously activated elements of an immune effector response.

Adoptive Cellular Immunotherapy

Adoptive cellular immunotherapy is the transfer of ex vivo activated immune cells with antitumor activity into a tumor-bearing host. Examples of this strategy include the adoptive transfer of lymphokine-activated killer (LAK) cells. In this approach, peripheral blood lymphoid cells (primarily NK cells) are taken from patients with cancer and are subsequently acti-

vated and expanded in vitro with IL-2; they are then administered back to the patient (usually along with systemic IL-2). Initial results of the use of LAK cells with high-dose IL-2 in patients with renal cell carcinoma and melanoma were encouraging, but subsequent randomized trials failed to show any benefit of LAK cell administration over the modest response rate seen with high-dose IL-2 alone.[17] As an alternative to peripheral blood lymphocytes, investigators have evaluated the adoptive transfer of tumor-infiltrating lymphoid (TIL) cells isolated from inflammatory infiltrates of excised tumor nodules. Once expanded in IL-2, these cells may be enriched for tumor-specific cytotoxic T cells and NK cells. Of 86 patients with stage IV melanoma receiving TIL treatment at the National Cancer Institute, 34% had objective tumor responses (i.e., partial or complete tumor regressions).[18] Unfortunately, it was not possible to isolate sufficient numbers of TIL cells from all tumor specimens to accomplish adoptive transfer. Other patients experienced disease progression while their TIL cells were being cultured and therefore never received treatment. Despite the modest success of TIL therapy for melanoma, comparable results have not been observed in the treatment of patients with renal cell carcinoma. One of the side benefits of TIL technology has been its use in isolating and defining the tumor antigens recognized by T cells from TIL cell cultures, such as MART-1, GP-100, tyrosinase, p15, TRP-1, and β-catenin, which are all expressed by melanoma cells.[19]

Other sources for adoptively transferred immune effector cells include the ex vivo activation and expansion of lymphoid cells derived from draining lymph nodes at tumor vaccination sites or in situ genetic modification of tumor nodules (direct injection of genetic material that encodes for proteins into tumors, triggering an immune response).

Genetic-modification strategies have been used to enhance the cytolytic activity of adoptively transferred immune effector cells. This technique involves the transduction of autologous CTLs (or their bone marrow precursors) with genes encoding a TCR specific for a tumor antigen (e.g., a TCR that recognizes MART-1 antigen in melanoma) or involves a chimeric receptor (i.e., an antigen-binding domain linked to a signal-transducing domain that initiates cellular activation after cross-linking by antigen) recognizing a tumor-associated antigen such as the mucin antigen TAG-72 expressed by most adenocarcinomas.[20,21]

Mononuclear phagocytes have also been evaluated as a source of immune effector cells for adoptive cellular immunotherapy. This strategy has involved the ex vivo activation of peripheral blood monocyte–derived macrophages by interferon gamma or lipopolysaccharide, with the subsequent administration of these activated cells back into the tumor-bearing host. However, as with other adoptive cellular therapy strategies, there has been minimal therapeutic response to date.[22]

Another form of adoptive cellular therapy that has received considerable attention is the use of allogeneic donor lymphocyte infusions. The rationale for this strategy is the fact that in allogeneic bone marrow/stem cell transplantation, donor lymphocytes appear to exert a graft-versus-leukemia effect. This approach is supported by data showing a higher risk of leukemic relapse in patients receiving syngeneic bone marrow transplantation than in patients receiving allogeneic bone marrow transplantation, as well as a higher risk of relapse in patients receiving T cell–depleted bone marrow than in those receiving unmanipulated allogeneic bone marrow.[23] The greatest benefit of allogeneic lymphocyte infusions was seen in patients with chronic myeloid leukemia (CML) who had relapses after allogeneic bone marrow transplantation; of these patients, 70% had complete responses to donor lymphocyte infusions.[24] This donor-lymphocyte graft-versus-leukemia effect may reflect immune reactivity toward minor histocompatibility antigens shared between recipient tissues (resulting in graft-versus-host disease) and leukemic cells, as well as reactivity toward leukemia-associated antigens. The efficacy of donor-lymphocyte infusions seen in patients with CML has led to the use of such infusions in other diseases, and investigators are currently assessing the value of the graft-versus-tumor effect as a primary form of antineoplastic therapy. In this strategy, a nonmyeloablative (i.e., not causing sustained neutropenia) immunosuppressive preparative regimen is used to achieve engraftment of allogeneic peripheral blood or bone marrow–derived stem cells, with the express purpose of developing a graft-versus-tumor effect. After allogeneic stem cell transplant engraftment, additional infusions of donor lymphocytes are often used to enhance the graft-versus-tumor effect. When this approach was expanded to include patients with chronic lymphocytic leukemia (CLL) or indolent non-Hodgkin lymphoma (NHL), promising results were observed.[25] In one study involving 15 patients with CLL or indolent NHL, 11 patients were engrafted, and all 11 had objective responses; eight patients had complete responses. The encouraging results seen in patients with hematologic malignancies have precipitated clinical trials to determine the feasibility and efficacy of this approach in patients with solid tumors. Preliminary results indicate that this approach has some efficacy in renal cell carcinoma.[26]

Antitumor Antibody Therapy

The administration of antitumor antibodies to patients with cancer may result in tumor regression by a variety of mechanisms, including ADCC caused by NK cells and macrophages; cytotoxicity resulting from complement fixation; induction of apoptosis; and the delivery of toxins, cytotoxic pharmaceutical agents, or ionizing radiation to the tumor. In the treatment of lymphoma, promising results have been observed with the use of monoclonal antibodies that are specific for surface idiotype and for the CD20 differentiation antigen. An FDA-approved therapy for NHL employs a monoclonal antibody targeting an 85 kd B cell differentiation antigen (B1, CD20), which is expressed by most normal and malignant B cells but not by stem cells or plasma cells. The advantage of CD20 as a target of immunotherapy is that it is not shed, internalized, or otherwise modulated as a result of antibody binding. Its mechanisms of B cell cytotoxicity may include complement fixation, ADCC, and the transmission of apoptotic signals. Intravenous administration of the chimeric human/mouse CD20 antibody rituximab, alone or in combination with chemotherapy, has produced significant disease regression in patients with low-grade or follicular NHL. In a phase 3 trial of 166 patients who received four doses of rituximab alone after having relapses of indolent lymphoma,[27] 48% of the treated patients achieved a clinically relevant response, including 6% who experienced complete tumor shrinkage. The median time to progression was 13 months. In another study, in which rituximab was used to treat 28 evaluable patients with bulky relapsed NHL (the patients had a large tumor burden),[28] the response rate was 43%, and the median time to progression was 8.1 months. Rituximab may show its greatest clinical utility when used in conjunction with chemotherapy. Of 35 evaluable patients with NHL who were not previously treated, 100% of the patients experienced a re-

sponse to combination therapy; 63% experienced complete response.[29] Seven of eight evaluable patients had a so-called molecular complete response, as evidenced by the loss of the *bcl-2* oncogene on polymerase chain reaction analysis.

An alternative approach for the use of anti-CD20 antibody in the treatment of patients with low-grade or follicular NHL employs murine antibodies radioconjugated to either yttrium-90 ibritumomab (^{90}Y-ibritumomab) or iodine-131 tositumomab (^{131}I-tositumomab). In two therapeutic trials,[30,31] a single intravenous dose of ^{90}Y-ibritumomab given to patients with relapsed or refractory low-grade or follicular NHL produced overall response rates of 74% to 80% (15% to 30% had complete responses); the median time to progression was 6.8 to 11.2 months. Similarly, a single intravenous dose of ^{131}I-tositumomab given to a comparable group of NHL patients caused significant clinical responses in 57% to 65%; 17% to 32% of these patients achieved complete shrinkage of their visible tumor. The median duration of response was 20 months.[32,33] In previously untreated patients with follicular lymphoma, treatment with ^{131}I-tositumomab produced an overall response rate of 97%; 63% of these patients achieved a complete response. The rate of progression-free survival was 68% at 3 years.[34] These promising findings recently led to FDA approval of ^{90}Y-ibritumomab for the treatment of patients with relapsed or refractory low-grade or follicular NHL; FDA approval of ^{131}I-tositumomab is pending.

Antitumor antibody therapy has also proved effective against breast cancer. The antigenic target of an FDA-approved breast cancer treatment is HER-2/neu, the 185 kd transmembrane growth factor receptor that is overexpressed in 25% of human breast cancers (and in a variety of other malignancies). Preclinical studies demonstrated that anti–HER-2/neu monoclonal antibody exposure of cells overexpressing HER-2/neu had antiproliferative activity and may facilitate apoptotic cell death. On the basis of these promising preclinical findings, trastuzumab (Herceptin), a form of anti–HER-2/neu monoclonal antibody, was evaluated alone and in combination with chemotherapy in patients with metastatic breast carcinoma.[35,36] In 222 patients receiving weekly intravenous trastuzumab alone, the overall response rate was 14%, with 2% having complete response; the median duration of response was 9.1 months. In a separate clinical trial, 469 patients were randomized to receive either trastuzumab plus chemotherapy (with doxorubicin-cyclophosphamide or paclitaxel) or chemotherapy alone. The combination of trastuzumab and chemotherapy was superior to chemotherapy alone with regard to time to progression, overall response rate, and 1-year survival. These promising findings led the FDA to approve trastuzumab as an adjunct to chemotherapy in patients with HER-2/neu-positive breast cancers.

Antitumor antibody therapy has been applied to the treatment of patients with colorectal cancer. The antigenic target in this disease is a nonsecreted 40 kd glycoprotein recognized by the investigational 17-1A monoclonal antibody. The 17-1A antigen is overexpressed by most epithelioid tumors, including the majority of colorectal carcinomas. Although the 17-1A antibody displayed only minor to modest clinical activity when administered intravenously to patients with advanced stage IV disease, it may be more active as adjuvant therapy for patients who are at high risk for recurrence after surgical resection. In a study of 189 patients with stage III colorectal carcinoma (i.e., patients found to have local lymph node metastases) who were randomly assigned either to observation alone or to four postoperative infusions of 17-1A monoclonal antibody,[37] mortality in the treatment arm was reduced by 32%, and recurrence was decreased by 23%.

Other strategies of antitumor-antibody therapy include the use of toxin or drug-conjugated monoclonal antibodies targeted to tumor-associated antigens (in addition to radionuclide-conjugated antibody treatment). Another approach involves the administration of bispecific (heteroconjugate) antibodies consisting of two covalently linked antibodies, with one antigen recognition site targeted to the target tumor–associated antigen and the other recognition site targeted to a trigger molecule (expressed on the effector cell).[38] Antitumor antibodies may also be useful as a means of depleting tumor cells contaminating preparations of autologous bone marrow cells before transplantation.[39]

Immunotherapeutic approaches to the treatment of cancer is an area of active investigation. Most of these approaches are investigational and continue to be evaluated in clinical trials. However, several of the therapeutic agents (IL-2, trastuzumab, BCG, and rituximab) have been shown to be effective and are approved by the FDA.

Diagnostic Role of Monoclonal Antibodies

DIAGNOSIS OF LEUKEMIAS AND LYMPHOMAS

In addition to their potential role as therapeutic agents, monoclonal antibodies that recognize tumor-associated antigens are being used to facilitate the diagnosis of hematologic malignancies.[40,41] The rationale for their use as diagnostic tools stems from the fact that leukemias and lymphomas are thought to be neoplastic counterparts of subpopulations of normal lymphoid and myeloid cells. The identification of cell surface antigens common to both normal and neoplastic lymphoid and myeloid cells has made it possible to assign the normal cell lineage to the malignant counterparts. This lineage information provides an additional means of discriminating among the various forms of leukemia and lymphoma that may influence treatment decisions. An example of this concept is T cell differentiation [*see Figure 3*]. Research over the past 20 years has led to the identification of multiple cell surface antigens expressed by cells of the T cell lineage that are differentially expressed along the path of differentiation. Monoclonal antibodies have been developed for these so-called differentiation antigens and have been utilized in the diagnosis of leukemia and lymphoma. CD34 and MHC class II gene products, for example, are expressed by the prethymocyte but are lost by thymic T cell precursors. Conversely, in the thymus, CD4 and CD8 accessory molecules are coexpressed by a biphenotypic T cell precursor that gives rise to separate CD4 and CD8 lineages in the thymic medulla, with subsequent maturation of these lineages in peripheral lymphoid tissues. T cell acute lymphoblastic leukemia/lymphoma shares the surface marker phenotypic characteristics of the bone marrow prethymocyte and is therefore believed to be a malignant transformation of this normal stem cell. The T cell lymphomas and chronic leukemias, on the other hand, share the cell surface phenotype characteristic of mature CD4 or CD8 lymphoid cells found in the peripheral lymphoid tissues. They, in turn, are thought to represent malignant transformation of these more mature T cells.

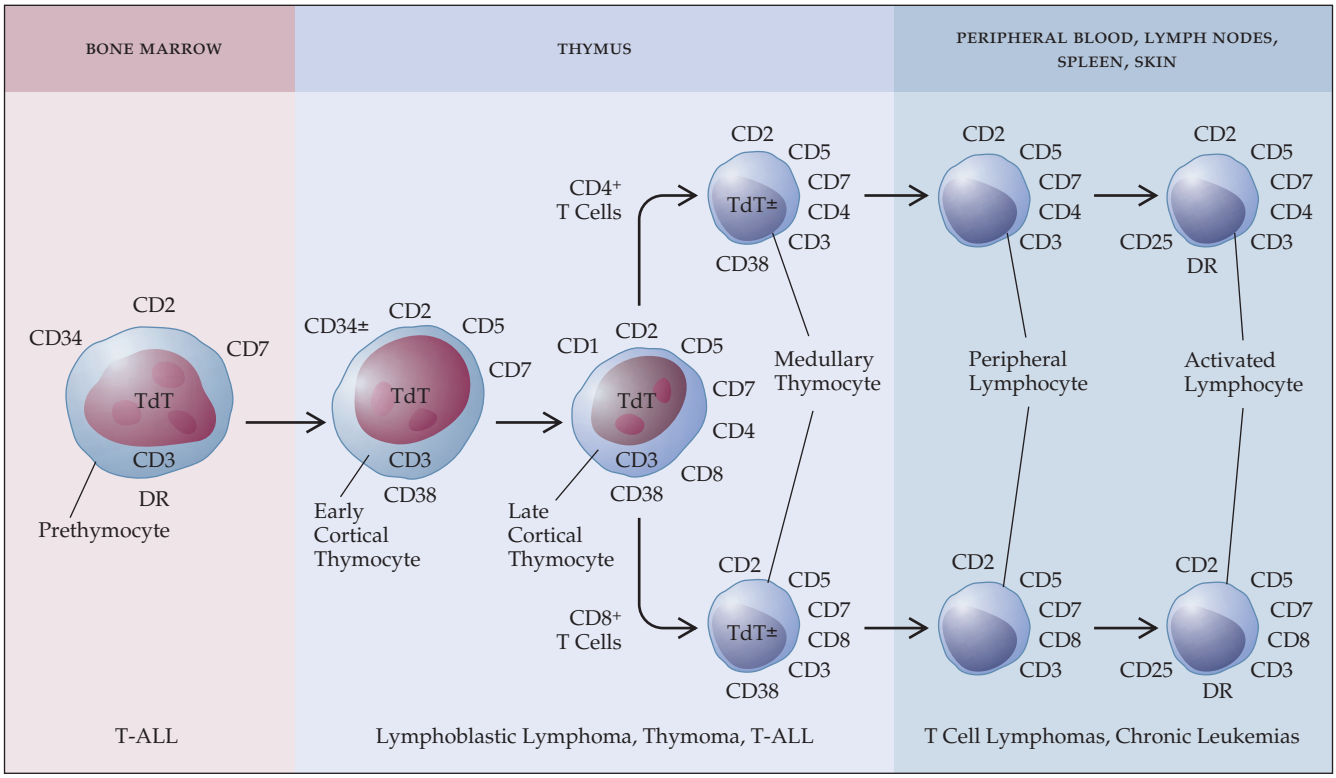


Figure 3 **Sequential expression of selected antigens during T cell development. Depicted is the pathway of T cell differentiation from the bone marrow stem cell to thymocyte maturation in the thymus, leading to the diversion of the $CD4^+$ helper and $CD8^+$ cytotoxic suppressor cell sublineages, each of which undergoes further differentiation in peripheral lymphoid tissues. Certain disorders are shown under the associated phenotypes. (DR—HLA class II; T-ALL —T cell acute lymphoblastic leukemia; TdT— terminal deoxynucleotidyl transferase)**

The expression of cell surface markers of leukemias and lymphomas can have diagnostic and prognostic significance:

1. Patients with T cell acute lymphoblastic leukemia whose cells express markers characteristic of the earlier stage of thymocyte development have a higher rate of treatment failure during induction than patients whose cells express a more mature phenotype.[42]
2. One third of patients with acute lymphoblastic leukemia (ALL) have leukemic cells expressing myeloid markers (biphenotypic ALL). Adult patients with biphenotypic ALL have a poor response to therapy and a shorter survival.[43]
3. Acute leukemias that are not classified by morphology or cytochemical characteristics (approximately 1%) can often be classified as belonging to a lymphoid lineage, as opposed to a nonlymphoid lineage, through the use of cell surface markers. In classifying myeloid leukemias into subgroups, analysis of cell surface markers can supplement morphology and cytochemical staining.
4. Of patients with CML in blast crisis, approximately one third express lymphocyte markers (e.g., CD10) and preferentially respond to the therapy for ALL.[40]

DIAGNOSIS AND STAGING OF CANCER IN VIVO

In addition to their potential utility in the pathologic diagnosis of leukemias and lymphomas, monoclonal antibodies have been employed in the diagnosis and staging of solid tumors.[44] Monoclonal antibodies specific for tumor-associated antigens have been useful in the diagnosis of carcinomas of unknown origin. Using a so-called cocktail of antibodies, pathologists are better able to determine the organ of origin of a metastatic carcinoma; such determinations may affect treatment decisions. In the setting of colon cancer, the CYT-103 (OncoScint) antibody has been approved by the FDA for use as a supplemental tool for the staging of colon cancer. The CYT-103 scan incorporates Indium-111–labeled CYT-103 antibody that is specific for a tumor-associated glycoprotein found on many mucin-containing adenocarcinomas, including many colon cancers. When used as a diagnostic tool, the CYT-103 antibody has a reported sensitivity of 92% and a specificity of 67%; its use results in a change in patient management in 33% of patients evaluated.[45] Its use may be superior to CT scanning in the detection of pelvic tumors and extrahepatic abdominal metastases. In addition, the CYT-103 antibody and other antibodies specific for the oncofetal protein CEA have been employed to identify the occult sites of metastatic disease in patients in whom the CEA level is rising, in the absence of radiographic evidence of disease recurrence. The CYT-103 antibody has also been used to detect recurrent ovarian carcinoma in the peritoneal cavity. A similarly tagged monoclonal antibody specific for prostate-specific membrane antigen is being utilized in the staging of prostate cancer.[46] Radiolabeled monoclonal antibodies specific for other cancers are being developed and evaluated for their utility in the management of cancer.

Robert F. Todd, III, M.D., Ph.D., owns stock, stock options, or bonds in Abbott Laboratories, Bristol Myers Squibb Company, Johnson & Johnson, Merck & Company, Inc., and Pfizer Inc. Bruce G. Redman, D.O., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Rosenberg SA: Principles of cancer management: biologic therapy. Cancer: Principles and Practice of Oncology. DeVita VT, Hellman S, Rosenberg SA, Eds. JB Lippincott Co, Philadelphia, 1997, p 349

2. Abbas AK, Lichtman AH, Pober JS: Immunity to tumors. Cellular and Molecular Immunology. Abbas AK, Lichtman AH, Pober JS, Eds. WB Saunders Co, Philadelphia, 1997, p 382

3. Wang RF, Rosenberg SA: Human tumor antigens for cancer vaccine development. Immunol Rev 170:85, 1999

4. Carter LL, Zhang X, Dubey C, et al: Regulation of T cell subsets from naive to memory. J Immunotherapy 21:181, 1998

5. Barry M, Bleackley RC: Cytotoxic T lymphocytes: all roads lead to death. Nature Rev Immunol 2:401, 2002

6. Gilboa E: How tumors escape immune destruction and what we can do about it. Cancer Immunol Immunother 48:382, 1999

7. Hudson MA, Gerber AJ, Ratliff TL: Management of superficial bladder cancer: Immunotherapy. Comprehensive Textbook of Genitourinary Oncology. Vogelzang NJ, Scardino PT, Shipley WU, et al, Eds. Williams & Wilkins, Baltimore, 1996, p 427

8. Pardoll DM: Cancer vaccines. Nat Med 4(suppl 5):525, 1998

9. Kawakami Y, Eliyahu S, Jennings C, et al: Recognition of multiple epitopes in the human melanoma antigen gp100 by tumor-infiltrating T lymphocytes associated with in vivo tumor regression. J Immunol 154:3961, 1995

10. Kawakami Y, Eliyahu S, Sakaguchi K, et al: Identification of the immunodominant peptides of the MART-1 human melanoma antigen recognized by the majority of HLA-A2 restricted tumor infiltrating lymphocytes. J Exp Med 180:347, 1994

11. Rosenberg SA, Yang JC, Schwartzentruber DJ, et al: Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. Nat Med 4:321, 1998

12. Massaia M, Borrione P, Battaglio S, et al: Idiotype vaccination in human myeloma: generation of tumor-specific immune responses after high-dose chemotherapy. Blood 94:673, 1999

13. Banchereau J, Steinman RM: Dendritic cells and the control of immunity. Nature 392:245, 1998

14. Timmerman JM, Czerwinski DK, Davis TA, et al: Idiotype-pulsed dendritic cell vaccination for B-cell lymphoma: clinical and immune responses in 35 patients. Blood 99:1517, 2002

15. Atkins MB, Lotze MT, Dutcher JP, et al: High-dose recombinant interleukin 2 therapy for patients with metastatic melanoma: analysis of 270 patients treated between 1985 and 1993. J Clin Oncol 17:2105, 1999

16. Bukowski RM: Immunotherapy in renal cell carcinoma. Oncology 12:801, 1999

17. Rosenberg SA, Lotze MT, Yang JC, et al: Prospective randomized trial of high dose interleukin-2 alone or in conjunction with lymphokine-activated killer cells for the treatment of patients with advanced cancer. J Natl Cancer Inst 85:622, 1993

18. Rosenberg SA, Yannelli JR, Yang JC, et al: Treatment of patients with metastatic melanoma with autologous tumor-infiltrating lymphocytes and interleukin 2. J Natl Cancer Inst 86:1159, 1994

19. Kawakami Y, Dang N, Wang X, et al: Recognition of shared melanoma antigens in association with major HLA-A alleles by tumor infiltrating T lymphocytes from 123 patients with melanoma. J Immunother 23:17, 2000

20. Cole DJ, Weil DP, Shilyansky J, et al: Characterization of the functional specificity of a cloned T cell receptor heterodimer recognizing the MART-1 melanoma antigen. Cancer Res 55:748, 1995

21. McGuinness RP, Ge Y, Patel SD, et al: Anti-tumor activity of human T cells expressing the CC49-zeta chimeric immune receptor. Hum Gene Ther 102:165, 1999

22. Andreesen R, Hennemann B, Krause SW: Adoptive immunotherapy of cancer using monocyte-derived macrophages: rationale, current status, and perspectives (review). J Leukoc Biol 64:419, 1998

23. Horowitz MM, Gale RP, Sondel PM, et al: Graft-versus-leukemia reactions after bone marrow transplantation. Blood 75:555, 1990

24. Kolb HJ, Schattenberg A, Goldman JM, et al: Graft-versus-leukemia effect of donor lymphocyte transfusions in marrow grafted patients: European Group for Blood and Marrow Transplantation Working Party Chronic Leukemia. Blood 86:2041, 1995

25. Champlin R, Khouri I, Anderlini P, et al: Nonmyeloablative preparative regimens for allogeneic hematopoietic transplantation. Bone Marrow Transplant 27:S13, 2001

26. Rini BI, Zimmerman T, Stadler WM, et al: Allogeneic stem-cell transplantation of renal cell cancer after nonmyeloablative chemotherapy: feasibility, engraftment, and clinical results. J Clin Oncol 20:2017, 2002

27. McLaughlin P, Grillo-Lopez AJ, Link BK, et al: Rituximab chimeric anti-CD20 monoclonal antibody therapy for relapsed indolent lymphoma: half of patients respond to a four-dose treatment program. J Clin Oncol 16:2825, 1998

28. Davis TA, White CA, Grillo-Lopez AJ, et al: Single agent monoclonal antibody efficacy in bulky non-Hodgkin's lymphoma: results of a phase II trial of rituximab. J Clin Oncol 17:1851, 1999

29. McLaughlin P, White CA, Grillo-Lopez AJ, et al: Clinical status and optimal use of rituximab for B-cell lymphomas. Oncology 12:1763, 1998

30. Witzig TE, White CA, Wiseman GA, et al: Phase I/II trial of IDEC-Y2B8 radioimmunotherapy for treatment of relapsed refractory CD20 positive B-cell non-Hodgkin's lymphoma. J Clin Oncol 17:3793, 1999

31. Witzig TE, White CA, Gordon LI, et al: Prospective, randomized control study of Zevalin (IDEC-Y2B8) radioimmunotherapy compared to rituximab immunotherapy for B-cell NHL: report of interim results. Blood 94:631a, 1999

32. Kaminski MS, Zelentz AD, Press O, et al: Pivotal study of iodine-131 tositumomab for chemotherapy-refractory low-grade or transformed low-grade B-cell non-Hodgkin's lymphomas. J Clin Oncol 19:3918, 2001

33. Vose JM, Wahl RL, Saleh M, et al: Multicenter phase II study of iodine-131 tositumomab for chemotherapy-relapsed/refractory low-grade and transformed low-grade B-cell non-Hodgkin's lymphoma. J Clin Oncol 18:1316, 2000

34. Kaminski MS, Estes J, Tuck M, et al: Iodine I-131 tositumomab therapy for previously untreated follicular lymphoma. Proc Am Soc Clin Oncol 19:5a, 2000

35. Cobleigh MA, Vogel CL, Tripathy NJ, et al: Efficacy and safety of herceptin as a single agent in 222 women with Her2 overexpression who relapsed following chemotherapy for metastatic breast cancer. Proc Am Soc Clin Oncol 17:97a, 1998

36. Slamon DJ, Leyland-Jones B, Shak S, et al: Use of chemotherapy plus a monoclonal antibody against HER-2 for metastatic breast cancer that overexpresses HER-2. N Engl J Med 344:783, 2001

37. Riethmüller G, Holz E, Schlimok G, et al: Monoclonal antibody therapy for resected Dukes' C colorectal cancer: seven-year outcome of a multicenter randomized trial. J Clin Oncol 16:1788, 1998

38. van de Winkel JG, Bast B, de Gast GC: Immunotherapeutic potential of bispecific antibodies. Immunol Today 18:562, 1997

39. Gribben JG, Freedman AS, Neuberg D, et al: Immunologic purging of marrow assessed by PCR before autologous bone marrow for B cell lymphoma. N Engl J Med 325:1525, 1991

40. Freedman AS: Cell surface antigens in leukemia and lymphomas. Cancer Invest 14:252, 1996

41. Jennings CD, Foon KA: Flow cytometry: recent advances in diagnosis and monitoring of leukemia. Cancer Invest 15:384, 1997

42. Crist WM, Shuster JJ, Falletta J, et al: Clinical features and outcome in childhood T-cell leukemia-lymphoma according to stage of thymocyte differentiation: a Pediatric Oncology Group study. Blood 72:1891, 1988

43. Urbano-Ispizua A, Matutes E, Villamor N, et al: Clinical significance of the presence of myeloid associated antigens in acute lymphoblastic leukemia. Br J Haematol 75:202, 1990

44. Goldenberg DM: Perspectives on oncologic imaging with radiolabeled antibodies. Cancer 80(suppl 12):2431, 1997

45. Petersen BM, Bass BL, Bates RH, et al: Use of radiolabeled murine monoclonal antibody, indium-111-CYT-103, in the management of colon cancer. Am J Surg 165:137, 1993

46. Kahn D, Williams RD, Manyak MJ, et al: 111Indium-capromab pendetide in the evaluation of patients with residual or recurrent prostate cancer after radical prostatectomy. J Urol 159:2041, 1998

Acknowledgment

Figures 1 through 3 Seward Hung.

IV PRINCIPLES OF CANCER TREATMENT

Eric H. Rubin, m.d.
William N. Hait, m.d., ph.d.

Management of Cancer Cases

DIAGNOSIS

Proper care of cancer patients begins with an accurate diagnosis. Unlike other specialties, in which a diagnosis can often be deduced from symptoms, signs, and laboratory tests, the practice of oncology depends on obtaining tissue for accurate histology. For example, although a cigarette smoker presenting with cough, hemoptysis, and weight loss is likely to have lung cancer, the selection of treatment for such a patient requires obtaining tissue to confirm the diagnosis and to specify the histologic subtype (e.g., small cell versus non–small cell carcinoma).

Both immunohistochemical and molecular analyses enhance the precision of the diagnosis. For example, the subtyping of breast cancer requires assay of hormone receptors (estrogen and progesterone) and an analysis of chromosome number and proliferative index of the tumor with a DNA histogram, which evaluates the percentage of cells in the various stages of the cell cycle. Furthermore, the expression of cancer-causing agents such as the *HER-2/neu* oncogene is becoming a standard part of breast cancer evaluation.[1] Similarly, the modern diagnosis of hematologic malignancies requires a panel of cell surface markers, as well as an evaluation of chromosomal rearrangements. An understanding of the molecular pathology of individual cancers has led to more precise prescriptions for systemic therapy.

STAGING

Knowledge of the extent of the disease at the time of diagnosis (stage) is required to properly manage patients with cancer. Staging is based on three components: the size or depth of penetration of the tumor (T), the involvement of lymph nodes (N), and the presence or absence of metastases (M). The TNM staging system is now standard, and its use is required not only for the management of cancer patients but also for the reporting of cancer cases to cooperative groups and many tumor registries.

Within each TNM category, the extent of involvement correlates with prognosis. In breast cancer, for example, small (< 1 cm) node-negative tumors are associated with a less than 10% chance of recurrence, whereas 3 cm node-negative tumors have a 50% chance of recurrence.[1] In most types of cancer, it is worse to have extensive lymph node involvement than involvement of a single lymph node. Lymph node involvement is particularly relevant in breast cancer: prognosis worsens with the increase in the number of axillary lymph nodes involved. The site of lymph node involvement may also be significant: in lung cancer, involvement of ipsilateral mediastinal lymph nodes (N2) indicates a better prognosis than involvement of lymph nodes in the contralateral mediastinum (N3).[2] Finally, the presence of metastases at the time of diagnosis predicts a poor outcome in most malignancies. For even the most treatable tumors, including testicular cancer and lymphomas, the extent or bulk of metastatic disease measured by imaging or serum markers determines the ultimate chances of survival.

The staging of Hodgkin disease and non-Hodgkin lymphoma follows similar principles as the TNM system. Although the staging systems for these cancers do not include a true T classification, the number of involved lymph node regions may be considered analogous to T and N, and visceral involvement may be considered analogous to M1 disease. In the widely used Ann Arbor classification of Hodgkin disease, stage is proportional to extent or bulk of disease, as well as involvement of viscera. As with solid tumors, overall prognosis is highly dependent on precise histologic classification, which has been one of the most challenging aspects for the hematopathologist. Thus, a large cell lymphoma may be cured in 50% of the cases regardless of the stage at diagnosis, whereas a small cell or indolent lymphoma is most often incurable regardless of stage.[3] Furthermore, certain histologic subsets have particularly aggressive behaviors and may be most amenable to early use of experimental therapies or bone marrow transplantation.

PERFORMANCE STATUS

Performance status is a measure of the functional capacity of a patient and adds significantly to the staging evaluation; for any given stage, performance status appears to be an independent predictor of outcome. Although the Karnofsky score is the most rigorous method of determining performance status, it is cumbersome and does not add appreciably to the simple scoring schema adopted by the Eastern Cooperative Oncology Group (ECOG) and World Health Organization [*see Table 1*]. Suffice it to say, knowing whether a patient is confined to bed or living a normal life is the mainstay of these classifications, which have proved to be a valuable means of prognostication.

TREATMENT

The treatment of patients with cancer requires coordination between medical, surgical, and radiation oncologists who work closely with primary care physicians to plan and implement treatment. In addition, social workers, research nurses, and pharmacists are often needed to formulate a treatment plan and make adjustments once treatment has begun.

This team must decide, in consultation with the patient, whether to treat with curative or palliative intent. In a patient

Table 1 World Health Organization Performance Scale

0	Able to carry out all normal activity without restrictions
1	Restricted in physically strenuous activity but ambulatory and able to carry out light work
2	Ambulatory and capable of all self-care but unable to carry out work; up and about more than 50% of waking hours
3	Capable only of limited self-care; confined to bed more than 50% of waking hours
4	Completely disabled; cannot carry out any self-care; totally confined to bed or chair

with advanced-stage, non–small cell lung cancer, for instance, aggressive treatment with chemotherapy, surgery, or radiation may lead to a more rapid demise than may no treatment at all. The classic example is that of a heavy smoker with potentially resectable lung cancer and chronic obstructive pulmonary disease who has limited pulmonary reserve (e.g., an FEV_1 [forced expiratory volume in 1 second] of < 800 ml). Even if this patient's tumor could be removed surgically (e.g., stage I or II disease), the surgery could leave the patient with insufficient pulmonary function to sustain life.

Curative treatment is undertaken in a patient with a good performance status who, on the basis of stage (e.g., M0) or histology (e.g., large cell lymphoma), has a reasonable probability of entering a meaningful remission and has a reasonable chance of cure.

Palliative treatment refers to the care provided to cancer patients for whom there is no possibility of cure. This care may sometimes include the same modalities as those used in curative treatment. For example, a patient with metastatic colon cancer may require surgery or radiation to control gastrointestinal bleeding despite an overall poor prognosis. Similarly, radiation and chemotherapy offer effective palliation for most patients.

The management of cancer cases also requires the use of analgesics to relieve pain, antiemetics to combat nausea and vomiting, and antidepressants to control depression. Highly emetogenic chemotherapy can be given safely and comfortably with the proper use of effective medications such as the serotonin antagonists (e.g., ondansetron), dopamine receptor antagonists (e.g., metoclopramide), phenothiazines (e.g., prochlorperazine), cannabinoids (e.g., nabilone, dronabinol), and glucocorticoids (e.g., dexamethasone). Cancer pain, which is often chronic and debilitating, usually requires the use of both long-acting and short-acting narcotic preparations such as morphine and oxycodone, supplemented with nonsteroidal anti-inflammatory drugs (NSAIDS) and antidepressants. Finally, depression, which is underdiagnosed in cancer patients, may be treated with tricyclic antidepressants or selective serotonin reuptake inhibitors (e.g., fluoxetine or paroxetine), in addition to psychosocial counseling.

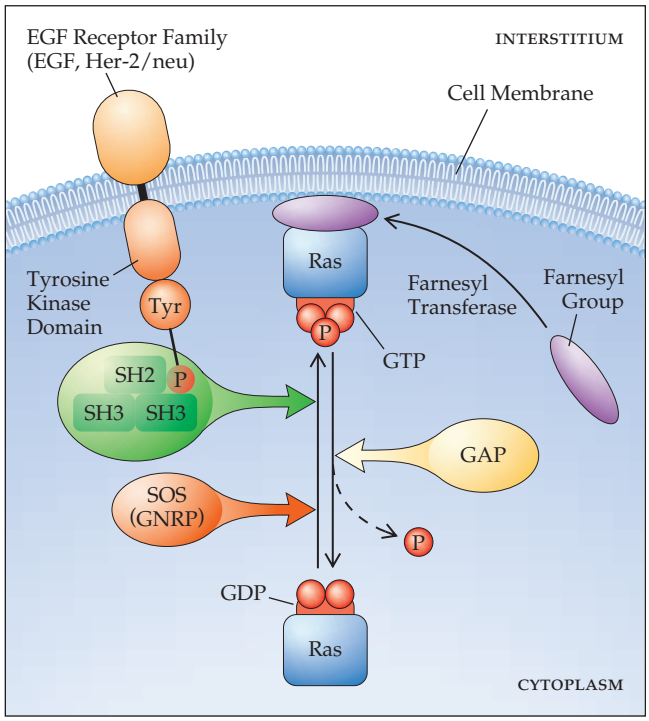


Figure 1 **Shown are growth factors, receptors, and signal transduction pathways that are targets for new cancer therapeutics. Receptor tyrosine kinases, such as the epidermal growth factor (EGF) receptor family, transmit cell proliferation signals through activation of the Ras oncoprotein through the src homology domain (SH) of the cytoplasmic portion of the receptor. Activation of Ras is enhanced by the son-of-sevenless (SOS) transducer. SOS functions as a guanine nucleotide–releasing protein (GNRP) in converting inactive Ras–guanosine diphosphate (GDP) to the active guanosine triphosphate (GTP)–bound form by a nucleotide exchange. Ras requires bound GTP for activity and is inactivated by a GTPase activating protein (GAP). Modification of Ras by farnesyl transferase allows Ras to embed in the plasma membrane, which is required for Ras activity.**

Scientific Basis of Cancer Treatment

CANCER BIOLOGY

Transformation and Proliferation

The origin of cancer in a person involves complex interactions between genes and environment, complicated further by the person's behavior and by aging. It is believed that the fundamental lesion is genetic—that is, changes in DNA lead to functional alterations in proteins that result in cellular transformation. Several of the genes and oncogenes involved in malignant transformation have been identified, including *p53*, *ras*, and *DCC* (deleted in colon cancer gene) [*see 12:II Molecular Genetics of Cancer*].[4] The expression of altered messenger RNAs (mRNAs) and proteins forms the basis of molecular diagnostic tests that will allow early detection of cells on their way to malignant transformation; such tests are under development.

Gene products involved in the development of cancer include growth factors, growth factor receptors, signal transduction pathways, and DNA transcription factors. In addition, telomerase, an enzyme that maintains the integrity of ends of chromosomes, must be active for tumors to continuously proliferate.[5] The identification of these gene products and the elucidation of their function have led to new therapeutic opportunities. For example, the *HER-2/neu* oncogene is a member of the epidermal growth factor receptor family[6] and is overexpressed in 20% to 30% of invasive breast cancers.[7] Trastuzumab, an antibody against the HER-2/neu protein, has therapeutic activity against breast cancer when used alone or in combination with chemotherapy.

The *ras* oncogene (*p21*) is mutated in 30% of human cancers, including colon cancer, pancreatic cancer, and breast cancer.[8] Ras activation leads to cell division by stimulating signal transduction pathways and transcription factors [*see Figure 1*]. Drugs that inhibit farnesyl transferase, a protein necessary for Ras to function properly, have shown promise in preclinical models and have now reached clinical trials.[9] Although inhibition of farnesylation is likely to affect other proteins besides Ras and therefore may not be as selective as originally hoped, this approach to anticancer treatment holds promise.

New therapeutic agents are being developed that inhibit the activity of both tyrosine and serine-threonine kinases.[10,11] Although members of this family of protein kinases share structural similarities, it appears possible to identify highly specific inhibitors that may have therapeutic value. The Bcr-Abl tyrosine kinase is an excellent example of a compelling target for drug

discovery. Bcr-Abl is a cytoplasmic fusion protein formed as a consequence of the t(9;22) translocation of the Philadelphia chromosome. Therefore, it does not exist in normal cells, making the potential for drug selectivity extremely high. Imatinib mesylate (Gleevec) inhibits the tyrosine kinase activity of Abl with high specificity and has been shown to have substantial activity with few side effects in the treatment of chronic myelogenous leukemia.[12] The removal of phosphates by protein phosphatases is also critical to growth control, and phosphatases are now also receiving attention as potential drug targets.

Currently available chemotherapeutic drugs are characterized as being cell-cycle or phase specific, referring to the observation that certain drugs are more active against cells traversing a specific phase of the cell cycle. A good example of the role of oncogenes in cell-cycle regulation comes from the elegant work done to confirm the Knudson hypothesis [*see 12:II Molecular Genetics of Cancer*], which predicts complete loss (both alleles) of tumor suppressor gene (*Rb* gene) function in the genesis of retinoblastoma—the so-called two-hit hypothesis.[13]

The Rb protein prevents cells from moving out of G_1 into the S phase of the cell cycle. Phosphorylation of Rb at the G_1/S interface releases bound E_2F and DRTF, two factors known to activate transcription of genes involved in DNA synthesis and cell division. These genes include the dihydrofolate reductase (*DHFR*) gene and the thymidylate synthase (*TS*) gene and may provide a molecular explanation for the cell-cycle specificity (S phase) of methotrexate and 5-fluorouracil (5-FU), drugs that target DHFR and TS proteins, respectively.

p53 is the gene most commonly mutated in human cancers.[14,15] p53 protein can activate or repress gene transcription. Cells with *p53* mutations acquire sensitivity to taxanes yet retain resistance to vinca alkaloids, both of which are chemotherapeutic drugs that target microtubules.[16] An explanation for this effect comes from the observation that wild-type p53 protein represses the transcriptional activation of a microtubule-associated protein, MAP4, which in turn regulates the stability of microtubules. Cells with mutant *p53* have increased MAP4 and stabilized microtubules, the target for taxanes such as paclitaxel and docetaxel. Conversely, cells with wild-type p53 are sensitive to vinca alkaloids and have decreased MAP4 and destabilized microtubules, the target for vinca alkaloids.

Cell Viability and Cell Death

Loss of telomeric chromosomal ends is believed to be the cause of the finite replicative potential (senescence) of nonmalignant cells. In malignant cells, this telomeric problem is solved by the reactivation of telomerase, a reverse transcriptase that systematically adds length to the shortened ends of chromosomes that occurs with each cell division.[17] Telomerase, therefore, is an attractive target for anticancer therapy and may be an early marker of cell transformation useful in early detection and chemoprevention.[18]

The accumulation of tumor cells represents a disequilibrium between cell proliferation and cell death. Programmed cell death (or apoptosis, from the Greek *apoptosis*, a falling off) involves the activation of energy-dependent suicide pathways within the cell.[19] Whereas necrotic cell death is characterized by cell swelling and disruption of cellular organelles and the cell membrane, apoptotic cells show cell shrinkage, blebbing of the plasma membrane, condensation of nuclear chromatin, and degradation of the genome into fragments that correspond to the size of nucleosomes. These nucleosomal fragments produce distinctive ladder-shaped patterns on electrophoretic gels. Apoptotic cell death is critical in normal embryonic development, during which time organs are formed and remodeled. Similarly, normal adult tissue is characterized by a steady state in which cell proliferation is balanced by cell death. The loss of this balance may result in cellular transformation [*see 12:II Molecular Genetics of Cancer*].

Invasion and Metastasis

With malignant transformation, cells acquire the ability to invade and metastasize. This process includes detachment from the primary site, invasion through the basement membrane, access to the blood or lymphatic vessels, entry to distant organs through adherence to visceral capillaries, and escape from a variety of other mechanisms designed to protect the host.

Normal cells survive and proliferate by adhering to each other and to the extracellular matrix (anchorage dependence).[20] Loss of these attachments causes normal cells to die. Therefore, cancer cells must lose their anchorage dependence—that is, they must acquire the ability to detach from the extracellular matrix without dying. Molecules such as E-cadherins and integrins mediate these attachments. E-cadherins provide the cell-cell connections essential for maintaining an organized epithelium.[21] The loss of functional E-cadherins is characteristic of malignant cells, and restoration of E-cadherin function blocks the ability of transformed cells to form tumors.

A critical moment in the life of a malignancy comes when it must acquire a blood supply of its own—a process known as angiogenesis.[22] To accomplish angiogenesis, tumor cells produce or attract angiogenic substances. Tumors produce both proangiogenic and antiangiogenic substances. Once the balance favors proangiogenic substances, new capillaries proliferate and become surrounded by rapidly dividing cancer cells. These cells secrete factors that promote the growth of both the tumor and its new blood supply. For example, capillary endothelial cells recruited to breast cancer cells produce interleukin-6, which can increase the migration of breast cancer cells into the bloodstream.[23] Interference with angiogenesis forms the basis of an exciting new approach to anticancer therapy.

PRINCIPLES OF CANCER PHARMACOLOGY

Dose Response and Schedule

A relation between dose and response exists for all drugs. The dose-response curve not only establishes effective doses but also defines the dose at which further increases produce no greater effect. The therapeutic index of a drug helps define its ultimate utility in patients. The therapeutic index is defined as the ratio of the dose that produces the desired effect 50% of the time (ED_{50}) to the dose that produces a toxic effect 50% of the time (TD_{50}). Because cancer cells originate from normal cells, it has been difficult to identify and exploit differences in sensitivity to drugs between neoplastic and nonneoplastic tissues. Moreover, many of the antineoplastic drugs in current use target cellular processes (e.g., DNA replication) that are present in both malignant and nonmalignant cells. Thus, there is often a small therapeutic index for cancer therapeutics.

In models of dose-response relations for several antineoplastic drugs, the percentage of tumor cells killed by a drug decreases as the number of cells increases. This phenomenon is related to a reduction in the percentage of proliferating cells (growth fraction) in larger tumors, defined by the Gompertz equation.[24] The decrease in growth fraction in larger tumors occurs as a result of

relative hypoperfusion and hypoxia in regions of tumor that are distant from blood vessels. The Gompertzian model correctly predicts several clinical observations: (1) increasing the drug dose may not improve antitumor activity; (2) treatment of minimal disease may be more effective than treatment of large tumor masses; and (3) small, frequent drug dosing (to decrease cell number in a large tumor mass) may be more effective than large, infrequent dosing.

With rare exception, optimal antineoplastic therapy requires repetitive dosing because even if all cells in a given tumor are sensitive to a drug, a single dose of the drug is not sufficient to kill the typically hundreds of millions or billions of cells that are present in a patient with cancer. The choice of schedule for repetitive drug administration usually relates to recovery of normal tissues. Although proper scheduling is critical for effective therapy, the drug schedule is often chosen empirically in initial drug development. The use of animal and pharmacokinetic models to optimize scheduling offers a rational alternative to empirical scheduling.[25]

Drug Resistance

For an anticancer drug to kill a cancer cell, the drug must enter the bloodstream, be activated or escape inactivation by drug-metabolizing enzymes, and reach the target in its active form. The drug–drug-target interaction must then produce cell death. Resistance to anticancer therapies results from interference with one or more of these critical steps [*see Figure 2*].[26]

To reach the target in an active form and at a therapeutic concentration, drugs must be absorbed or delivered into the body. Oral absorption requires a drug or drug formulation that resists the hostile environment of the stomach, traverses the epithelial lining of the gut, and escapes drug-metabolizing enzymes in the liver. Changes in drug metabolism can be inherited or acquired. Inherited drug resistance occurs in certain persons who rapidly metabolize drugs, so that less than therapeutic concentrations reach the pharmacologic target; or in the case of prodrugs such as cyclophosphamide, there is inadequate conversion to the active molecule. Acquired drug resistance occurs through induction of drug-metabolizing enzymes in the liver, other organs, or the tumor cells.

Finally, the tenuous blood supply to many tumor masses prevents adequate concentrations of drugs from reaching the target, or the drugs may find cells that are resistant because of relative hypoxia. Hypoxia causes resistance to both radiation and most chemotherapeutic drugs.[27] The mitomycins are an exception to this rule because these drugs are more active in hypoxic environments.

Resistance to drugs that act on extracellular receptors can occur if the receptor is modified. Many examples exist of receptors that are mutated during the process of oncogenesis. Therefore, effective therapeutic agents must recognize the altered form of the receptor protein.

A drug not only must reach the tumor in therapeutic concentrations but also must interact with targets that are often intracellular. Cancer cells acquire mechanisms to block the intracellular accumulation of drugs by producing proteins that transport drugs to the extracellular space, where they are removed by the circulation. Alternatively, cancer cells can neutralize the effects of drugs within the intracellular milieu.

Multidrug resistance (MDR) refers to the clinical and laboratory circumstance in which a tumor is no longer susceptible to several chemotherapeutic drugs having different mechanisms or targets. In the early 1980s, the *mdr-1* gene product P-glycoprotein was discovered in cells that were selected for resistance to antimicrotubule drugs.[28,29] P-glycoprotein is a member of the adenosine triphosphate (ATP)–binding cassette family that includes both eukaryotic and prokaryotic energy-dependent transporters. P-glycoprotein spans the plasma membrane and recognizes a broad spectrum of anticancer drugs, many of which are naturally occurring substances. In the presence of ATP, P-glycoprotein pumps the drugs to the extracellular space, so that an effective concentration at the intracellular target is never achieved. Examples of drugs recognized and transported by P-glycoprotein include anthracyclines, vinca alkaloids, taxanes, epidophyllotoxins, and other hydrophobic compounds. In contrast, many alkylating agents, platinating agents, and antimetabolites are not affected.

P-glycoprotein is a tantalizing target for new drug discovery. The structural features of drugs that block drug transport mediated by P-glycoprotein were defined in the mid-1980s,[30] and verapamil, an early prototype, soon reached clinical trial. The observation that cyclosporins could also inhibit P-glycoprotein led to the realization that nonimmunosuppressive analogues of cyclosporine and other immunosuppressives could be developed as modulators of MDR. Several of these compounds have been tested, and some have shown early clinical promise. In particular, it appears that PSC-833 (valspodar), a nonimmunosuppressive cyclosporine, in combination with mitoxantrone, etoposide, and cytosine arabinoside, can provide a therapeutic advantage for patients with refractory acute myelocytic leukemia.[31] Ongoing studies in myeloma, non-Hodgkin lymphoma, breast cancer, and other malignancies will define the utility of this new class of anticancer agent.

The multidrug resistance protein (MRP) is a second member of the ATP-binding cassette family that confers MDR. MRP was discovered when investigators found that certain cells with long-term exposure to anthracyclines or anthracenes behaved like MDR cell lines but did not express P-glycoprotein.[32,33] MRP is located on both the plasma membrane and the endoplasmic reticulum and removes drugs from the cell after conjugation to or association with glutathione. The spectrum of drugs affected by MRP is similar to that affected by P-glycoprotein, with the exception of taxanes, which are poorly transported. MRP avidly transports conjugated leukotrienes, which are proinflammatory substances that mediate numerous biologic and pathologic responses. Several substances that interfere with leukotriene transport inhibit the transport by MRP and may represent new classes of drug-resistance modulators. In addition, several drugs that block the function of P-glycoprotein also block the function of MRP. These broad-spectrum MDR modulators are also undergoing clinical trials. Finally, disruption of glutathione conjugation through depletion of glutathione or inhibition of glutathione transferase inhibits MRP-mediated drug transport and may also sensitize drugs to alkylating agents.

The breast cancer resistance protein (BCRP, also known as ABCP [placenta ABC protein] and MXR [mitoxantrone-resistance protein) is a new member of the ATP-binding cassette family and is implicated in resistance to anthracyclines, mitoxantrone, and camptothecins.[34,35] Expression of BCRP is associated with a poor response to induction therapy in childhood acute leukemia.[36]

Once a drug reaches its therapeutic target, the target must be susceptible to the anticancer agent. For example, the topoisomerases are the target for many anticancer drugs, including the anthracyclines, anthracenes, and epidophyllotoxins (topoiso-

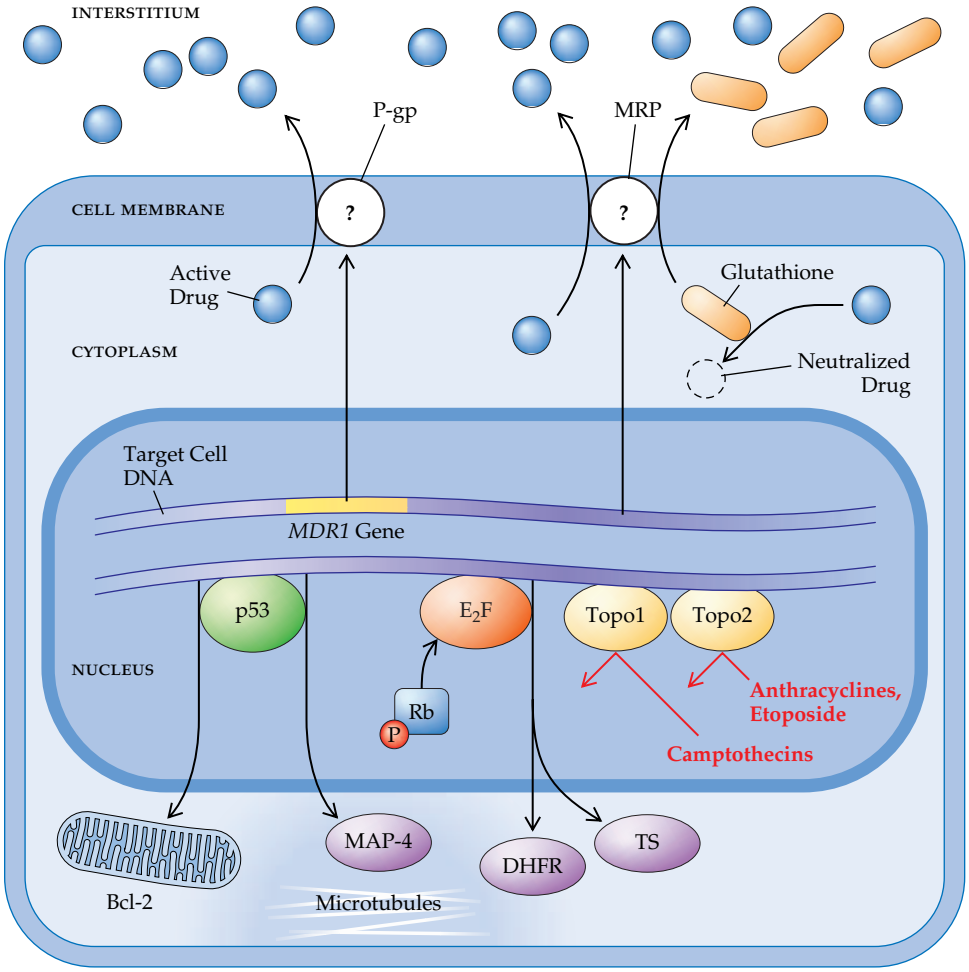


Figure 2 **Drug resistance can occur through failure of the drug to reach its target because of drug efflux mechanisms such as P-glycoprotein (P-gp), the *MDR1* gene product, or the multidrug resistance protein (MRP). MRP cotransports substances with glutathione, a molecule that can itself render cells resistant to a variety of alkylating agents through neutralization. p53 can regulate the sensitivity to drugs by changing the expression of proapoptotic and antiapoptotic molecules, such as Bcl-2, and by regulating the polymerization state of microtubules through microtubule-associated protein 4 (MAP4). Phosphorylation of the retinoblastoma protein (Rb) releases the transcription factor E_2F, which in turn regulates the sensitivity to methotrexate and fluorouracil through the production of dihydrofolate reductase (DHFR) and thymidylate synthetase (TS), respectively. Downregulation or mutations in topoisomerases (topo) produce resistance to camptothecins and topo II poisons, such as anthracyclines and etoposide.**

merase II), as well as the camptothecins (topoisomerase I).[37] Resistance to these drugs occurs through mutations in the drug-binding domain of the enzyme, the ATP-binding domain; through a decreased expression of the protein; or through failure of the enzyme to localize properly in the nucleus.[38,39] Binding of topoisomerase drugs to the enzyme activates a ubiquitin-mediated pathway of enzyme degradation.[40] Downregulation of one of the topoisomerases leaves the cell susceptible to poisoning by an inhibitor of the other. This discovery led to the sequential use of topoisomerase I and II inhibitors as a potentially more effective means of administering these chemotherapeutic agents.[41]

Resistance to methotrexate was one of the earliest understood phenomena in cancer pharmacology. Resistance can occur through mutations in the drug target, dihydrofolate reductase, which maintains cellular reduced folate pools that are required for DNA synthesis. Resistance to methotrexate can also occur by amplification or overexpression of the dihydrofolate reductase enzyme.[42] Similarly, resistance to 5-FU can occur by virtue of alteration or overexpression of the enyzme that generates thymidine for DNA synthesis, thymidylate synthase.

Resistance to antimicrotubule drugs occurs through MDR and through alterations in the drug targets. Vinca alkaloids bind depolymerized tubulin and prevent microtubule polymerization, whereas taxanes bind to and stabilize polymerized microtubules. Changes in the dynamic equilibrium of microtubules through alterations in microtubule-associated proteins or tubulin mutations can also alter the sensitivity to antimicrotubule drugs.[16]

Resistance to alkylating agents occurs through overexpression of both drug-neutralizing substances and metabolizing proteins such as glutathione and glutathione transferase. Depletion of glutathione with buthionine sulfoxamine sensitizes resistant cells to alkylating agents and is under clinical investigation.[43]

We now more fully understand the molecular mechanisms that account for changes in several drug targets. For example, inactivation of Rb through phosphorylation or mutation releases bound E_2F, which in turn leads to the transcriptional activation of several enzymes involved in cell proliferation, including dihydrofolate reductase and thymidylate synthase. Changes in expression of these drug targets affect the sensitivity to methotrexate and 5-FU, respectively. Finally, the presence of antiapoptotic proteins such as Bcl-2 may produce resistance to both chemotherapy and radiation by interfering with the cell death signal generated after an interaction between a drug and its therapeutic target.[44]

Combination Chemotherapy

Many cancers are incurable as a result of either intrinsic or acquired resistance to antineoplastic drugs. The Goldie-Coldman model relates the probability of drug resistance to cell number and the spontaneous mutation rate of cancer cells.[45] Assuming a spontaneous mutation rate of 1×10^{-6}, this model predicts that a 1 mm^3 tumor nodule (below the level of clinical detection) has a greater than 60% chance of harboring a cell with a resistance-conferring mutation. The Goldie-Coldman model also predicts that antitumor activity will be enhanced by combining drugs for which resistance mechanisms are independent. For example, in a 1 mm^3 tumor nodule, the likelihood of a cell possessing simultaneous resistance to two non–cross-resistant drugs is less than 0.000001. These predictions are consistent with the clinical observation that with few exceptions, cancers are cured only by combination chemotherapy.

The successful application of combination chemotherapy may thus depend on the availability of non–cross-resistant drugs with nonoverlapping toxicities. Although the existence of a wide variety of mechanistically distinct antineoplastic drugs suggests that finding non–cross-resistant drugs may be accomplished, cancer cells are known to possess a variety of mechanisms that confer multidrug resistance. These mechanisms of resistance contribute to the failure of combination chemotherapy to cure patients with common malignancies, such as metastatic colon or prostate cancer.

Common Toxicities

Most antineoplastic drugs target proteins or nucleic acids that are common to both malignant and nonmalignant cells and thus have a narrow therapeutic index. In addition, antineoplastic drugs are usually administered at the maximum tolerated doses identified in phase I trials. Coupled with interpatient variability in drug metabolism, this dosing strategy results in a relatively high incidence of toxicity. Although toxicities may be unique for a particular drug, several toxicities are shared by many antineoplastic drugs. These include nausea and vomiting, mucositis and diarrhea, myelosuppression, alopecia, infertility, and increased risk of secondary malignancy.

Nausea and vomiting result from local gastrointestinal effects, as well as activation of the chemoreceptor trigger zone in the central nervous system.[46] The identification of neurotransmitter pathways involved in chemotherapy-induced nausea and vomiting led to development of potent antiemetics such as the serotonin antagonists ondansetron and granisetron. The use of these drugs has significantly improved administration of highly emetogenic drugs such as cisplatin.

Mucositis and diarrhea are other common gastrointestinal toxicities that are attributable to the relatively high proliferative rate of normal gastrointestinal tissues, which makes these tissues more susceptible to the cytotoxic effects of many chemotherapeutics. Similar mechanisms result in chemotherapy-induced myelosuppression, alopecia, and infertility.

Many chemotherapeutics are found to be mutagenic when tested in vitro. Moreover, epidemiologic studies of patients cured of a pediatric cancer or lymphoma have demonstrated that certain therapies are associated with a higher risk of secondary malignancy than others.[47] Agents that damage DNA, such as alkylating agents and topoisomerase poisons, are believed to be responsible for these secondary malignancies. In vitro models suggest that even non–DNA-damaging agents, such as paclitaxel, may be carcinogenic.[48]

Pharmacokinetics and Pharmacogenetics

Given the narrow therapeutic index of most antineoplastic drugs, it is critical to administer an effective but tolerable dose to each patient. Individual dosage is based on a patient's body surface area (BSA); however, BSA is a poor predictor of clearance for most drugs.[49] In some cases, alternative strategies of individualized dosing are available. For example, because carboplatin is eliminated predominantly by the kidney, drug clearance correlates with renal function. Therefore, carboplatin dosing is based on a patient's creatinine clearance, which can be measured either directly or indirectly. The importance of individualized dosing is evident from a study of childhood leukemia, in which patients with B cell lineage leukemia randomized to receive individualized drug dosing (based on determination of patient clearances for each of the administered drugs) had significantly greater survival than patients who were treated with BSA-based dosing.[50]

Advances in understanding the molecular and genetic determinants of drug metabolism should lead to more precise, individualized dosing for other antineoplastic drugs. For example, certain polymorphisms in the dihydropyrimidine dehydrogenase gene, which degrades fluoropyrimidines, and the uridine diphosphate glucuronosyltransferase *1A1* gene, which when mutated results in Gilbert syndrome, may alter the metabolism of 5-FU and irinotecan, respectively.[51,52] These genetic polymorphisms result in increased toxicity in patients receiving these drugs. In the future, this problem may be addressed by customizing drug administration for each patient by use of genetic determinants of drug metabolism and toxicity.

In addition to allowing individualized dosing of drugs, genetic studies may soon allow physicians to predict response to a particular drug or drug combination. For example, certain polymorphisms in the thymidylate synthase gene are associated with reduced survival in patients treated with methotrexate for leukemia[53] or treated with 5-FU for colorectal cancer.[54] Conversely, a polymorphism in the glutathione S-transferase P1 gene is associated with improvement in patients with colorectal cancer treated with a combination of 5-FU and oxaliplatin.[55]

Specific Chemotherapeutic Agents

DRUGS THAT ALTER NUCLEIC ACID SYNTHESIS OR FUNCTION

DNA Alkylating and Cross-linking Agents

This category of drugs contains the oldest and some of the most widely used antineoplastic agents. They form covalent alkyl bonds with nucleic acid bases, resulting in intrastrand or interstrand DNA cross-links that are particularly toxic to cells

undergoing division. Examples include the nitrogen mustards (e.g., mechlorethamine, cyclophosphamide, and ifosfamide), the alkyl sulfonates (e.g., busulfan), the triazenes (e.g., dacarbazine), and the nitrosoureas (e.g., carmustine [BCNU]) [*see Table 2*]. By altering the structure of DNA, these drugs inhibit DNA replication and transcription. DNA damage produced by alkylators may be more likely to kill malignant cells than nonmalignant cells because of rates of proliferation and an altered balance of proapoptotic and antiapoptotic pathways. For example, testicular cancer, which is one of the few nonhematologic malignancies that are curable even when metastatic, is composed of cells that undergo apoptotic cell death more readily than normal cells after DNA damage.[56] Differences in the cellular mechanisms responsible for cell-cycle arrest and DNA repair (e.g., p53, the DNA mismatch repair pathway, and O^6-alkyltransferase) between malignant cells and normal cells may also contribute to the increased sensitivity of cancer cells to alkylating agents.

Alkylating agents such as cyclophosphamide are used in the treatment of most hematologic and nonhematologic malignancies. Because dose-response curves of alkylating agents in cell culture systems do not reach a plateau at high doses, cyclophosphamide is commonly used in high-dose chemotherapy regimens that involve bone marrow or stem cell transplantation. Certain alkylating agents are also useful in the treatment of brain tumors, including the nitrosourea BCNU and the triazene temozolomide.[57]

Myelosuppression is typically the most prominent toxicity observed with alkylating agents, such as busulfan, which is associated with prolonged bone marrow impairment. Treatment with alkylating agents is often associated with gonadal dysfunction, including oligospermia and amenorrhea. Bladder toxicity, such as hemorrhagic cystitis, can result from irritation by the acrolein metabolite of cyclophosphamide or ifosfamide, particularly when these drugs are administered at high doses. Not unexpectedly, alkylating agents can be carcinogenic, with secondary malignancies reported after therapy with several of these drugs.

Cisplatin, a simple compound containing a platinum atom, two amines, and two chlorides, produces effects similar to those of the alkylating agents by cross-linking adjacent or opposing guanine bases. Cisplatin or its analogue carboplatin is an important part of the treatment of many nonhematologic malignancies, including testicular, lung, bladder, and ovarian cancers. Oxaliplatin is a cisplatin analogue that is approved for the treatment of colorectal cancer. Toxicities associated with platinum agents include nausea and vomiting, alopecia, and myelosuppression. Cisplatin can produce severe renal toxicity and is therefore administered in conjunction with hydration and often an osmotic diuretic such as mannitol. Neurotoxicity may also be observed with cisplatin therapy, commonly consisting of peripheral neuropathies and auditory impairment.

Inhibitors of Nucleic Acid Synthesis

Many antineoplastic drugs, including folate analogues and nucleoside derivatives, inhibit the synthesis of nucleic acids and are often referred to as antimetabolites. These drugs are particularly effective in destroying cells during the S phase of the cell cycle, which is when DNA is synthesized. Thus, the selective toxicity these drugs exhibit toward cancer cells may relate to a greater rate of DNA replication in cancer cells than in normal cells. Nevertheless, these drugs share a common toxicity profile (e.g., myelosuppression and mucositis) that results from targeting of proliferating but nonmalignant tissues [*see Table 2*].

Methotrexate is a folate analogue that is used in the treatment of many different hematologic and nonhematologic malignancies. This drug inhibits dihydrofolate reductase, an enzyme that uses reduced folate as a methyl donor in the synthesis of pyrimidine and purine nucleosides. Similarly, the pyrimidine analogue 5-FU blocks production of thymine nucleotides by inhibiting thymidylate synthase. Other antineoplastic nucleosides, such as cytosine arabinoside and gemcitabine, are incorporated into DNA or RNA strands and either block further strand synthesis or cause structural perturbations that result in damage to DNA. 6-Mercaptopurine is a purine analogue that functions similarly. Hydroxyurea also belongs in this category, because it inhibits ribonucleotide reductase, an enzyme that converts ribonucleoside precursors into deoxyribonucleosides that are eventually used by DNA polymerases to replicate DNA.

Although both hydroxyurea and nucleoside derivatives destroy cancer cells by inhibiting nucleic acid synthesis, these agents exhibit different clinical antitumor activities. For example, 5-FU is effective in the treatment of gastrointestinal malignancies, whereas cytosine arabinoside is not. The converse is true for leukemia. Similarly, gemcitabine and cytosine arabinoside are very similar structurally [*see Table 2*], yet gemcitabine is active in several nonhematologic malignancies, whereas cytosine arabinoside is not. The reasons for these specificities are not understood, but they likely relate to additional unique effects of nucleoside derivatives on DNA and RNA metabolism, as well as differences in cellular and extracellular metabolism.

The success of folate and nucleoside analogues encouraged development of potentially more effective derivatives. For example, capecitabine is an orally administered drug that is metabolized to fluorouracil after ingestion and is approved for the treatment of breast cancer.[58] Capecitabine is converted to 5´-deoxy-5-fluorocytidine (5´-DFCR) by carboxylesterases in the liver and then to 5´-deoxy-5-fluorouridine (5´-DFUR) by cytidine deaminase in the liver and tumor tissue. DFUR is ultimately converted to fluorouracil by thymidine phosphorylase. Because there is more activity of thymidine phosphorylase in tumors than in certain normal tissues, capecitabine has the potential to be more selective than 5-FU.

Pemetrexed (Alimta) is a new folate antagonist that is active in the treatment of lung cancer and mesothelioma. This drug inhibits multiple enzymes involved in the folate pathway, including thymidylate synthase, glycinamide ribonucleotide formyltransferase, and dihydrofolate reductase.

DNA Topoisomerase Inhibitors

Topoisomerases correct alterations in DNA topology that occur during replication and transcription. These enzymes make transient single-strand (topoisomerase I) or double-strand (topoisomerase II) nicks in DNA. Several antineoplastic drugs inhibit either topoisomerase I or II. The effect of these drugs on topoisomerases is unique relative to other anticancer drugs that are enzyme inhibitors. They poison the enzymes and inhibit religation of the DNA nicks produced during topoisomerase catalysis.[59] Because cancer cells possess more topoisomerase activity than normal cells, there is more drug-induced DNA damage and resultant cell death. Toxicities observed with these drugs relate to targeting of topoisomerases in normal proliferating tissues and include myelosuppression and mucositis.

Topoisomerase II inhibitors such as doxorubicin, daunorubicin, etoposide, and teniposide are part of most combination chemotherapy regimens. Topoisomerase I inhibitors were iden-

Table 2 Chemotherapeutic Agents

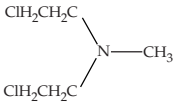
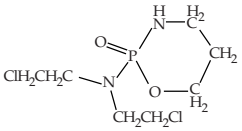
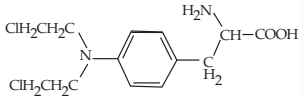
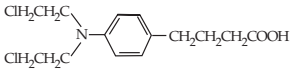
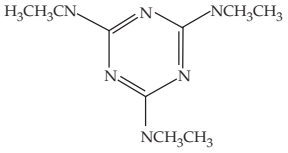
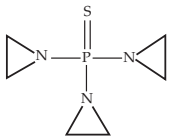
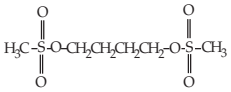
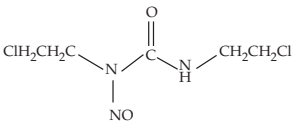
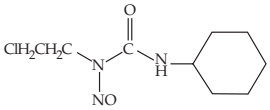
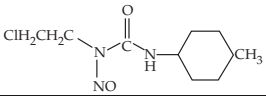
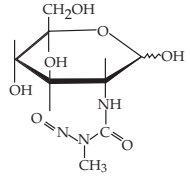
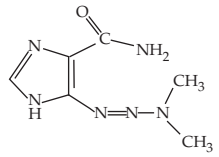
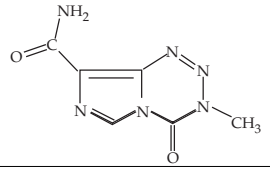
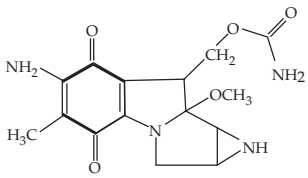
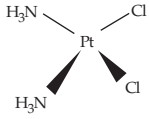
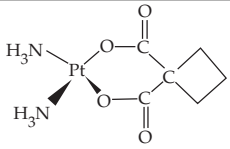
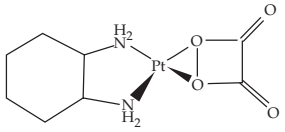
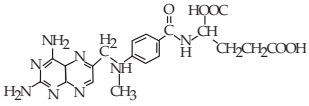
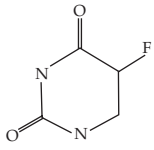
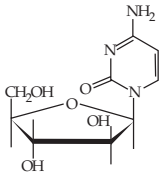
	Class	*Structure*	*Route*	*Target*	*Uses*	*Toxicities*
Alkylating Agents	*Nitrogen Mustards* Mechlorethamine		I.V.	N-7 of guanine; carbamoylation of proteins	Hodgkin disease; non-Hodgkin lymphoma	Nausea and vomiting, lacrimation, myelosuppression, mucositis, alopecia, vesication
	Cyclophosphamide; ifosfamide		I.V., p.o.		Acute lymphocytic leukemia (ALL) and chronic lymphocytic leukemia (CLL); lymphomas; myeloma; neuroblastoma; breast, ovarian, lung, cervical, and testicular cancers; sarcoma; Wilms tumor	Myelosuppression, mucositis, alopecia, emesis, dizziness, hemorrhagic cystitis, skin pigmentation, cardiac, renal, central nervous system (particularly caused by ifosfamide, including seizures, altered mental status, coma, and paralysis at high doses), syndrome of inappropriate secretion of antidiuretic hormone
	Melphalan		p.o.		Myeloma and breast cancers	Myelosuppression
	Chlorambucil		p.o.		CLL, primary macroglobulinemia, Hodgkin disease, non-Hodgkin lymphoma	Myelosuppression, secondary leukemias and other tumors
	Ethylenimines and Methylmelamines Hexamethylmelamine		I.V.		Ovarian cancer	Nausea and vomiting, diarrhea, myelosuppression
	Thiotepa		I.V.		Bladder, breast, and ovarian cancers	Myelosuppression, mucositis, nausea, and vomiting
	Alkyl Sulfonates Busulfan		p.o.		Chronic myelogenous leukemia; acute myelogenous leukemia (in bone marrow transplantation cases)	Myelosuppression, prolonged thrombocytopenia
	Nitrosoureas Carmustine (BCNU)		I.V.	O-6 of guanine; carbamoylation of lysine ε-amino groups	Hodgkin disease, non-Hodgkin lymphoma, astrocytomas, myeloma, melanoma	Delayed myelosuppression (nadir at 4–6 wk); hepatic necrosis at high doses, nausea and vomiting, flushing; interstitial pulmonary fibrosis and renal failure at doses > 1,000 mg/m^2
	Lomustine (CCNU)		p.o.		Hodgkin disease, non-Hodgkin lymphoma, astrocytomas, small cell lung cancer	Myelosuppression
	Semustine (methyl-CCNU)		p.o.		Colon cancer	Myelosuppression
	Streptozotocin		I.V.		Insulinoma, carcinoid tumor	Nausea, renal (proximal tubule effects), hepatic, mild to moderate myelosuppression

Table 2 Chemotherapeutic Agents (*continued*)

	Class	Structure	Route	Target	Uses	Toxicities
Alkylating Agents	*Triazenes* Dacarbazine (DTIC)		I.V.	O-6 of guanine; N-7 of guanine	Melanoma, Hodgkin disease, sarcomas	Myelosuppression, nausea, vomiting, flulike syndrome
	Temozolomide		p.o.		Astrocytoma, melanoma	Lymphocytopenia, hepatic toxicity, nausea and vomiting, hyperglycemia, anemia, and thrombocytopenia
	Bioreductive Alkylating Agents Mitomycin-C		I.V.	N-6 of adenine; N-7 of guanine; and O-6 of guanine	Colon, rectal, breast, gastric, head and neck, lung, and cervical cancers	Myelosuppression with delayed recovery (6–8 wk), nausea, vomiting, diarrhea, mucositis, dermatitis, asthenia, interstitial fibrosis, hemolytic-uremic syndrome, congestive heart failure
Platinating Agents	*Platinum Compounds* Cisplatin		I.V.	Interstrand and intrastrand cross-linking of DNA	Testicular, ovarian, bladder, head and neck, lung, thyroid, cervical, and endometrial cancers; neuroblastoma, osteogenic sarcoma	Renal toxicity, nausea and vomiting, peripheral neuropathy, ototoxicity, myelosuppression, electrolyte abnormalities (hypomagnesemia, hypocalcemia, hypophosphatemia), anaphylactoid reactions
	Carboplatin		I.V.			Myelosuppression (thrombocytopenia)
	Oxaliplatin		p.o.		Colon cancer	Peripheral neuropathy, nausea and vomiting, myelosuppression
Nucleic Acid Synthesis Inhibitors	*Folate Analogues* Methotrexate		I.V., p.o.	Inhibition of dihydrofolate reductase; inhibition of folate-dependent synthesis of purines and thymidylate	ALL, choriocarcinoma, non-Hodgkin lymphoma; breast, head and neck, and lung cancers; osteogenic sarcoma	Myelosuppression, mucositis, renal failure, reversible pneumonitis, hepatic fibrosis (low-dose oral administration)
	Pyrimidine Analogues Fluorouracil (5-FU); floxuridine (fluorodeoxyuridine, FUDR), capecitabine		I.V., p.o.	Inhibition of thymidylate synthase; inhibition of RNA processing, incorporation into DNA	Breast, colon, stomach, pancreatic, ovarian, head and neck, bladder, cervical, and prostate cancers; hepatoma	Myelosuppression, mucositis, nausea, diarrhea, hair loss, nail changes, pigmentation, chest pain, hand-foot syndrome
	Cytarabine (cytosine arabinoside)		I.V.	Analogue of 2′-deoxycytidine; inhibition of DNA chain elongation	Acute myeloid leukemia (AML) and ALL, non-Hodgkin lymphoma	Myelosuppression, mucositis, conjunctivitis, reversible hepatic dysfunction

(*continued*)

Table 2 Chemotherapeutic Agents (*continued*)

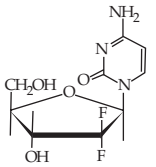
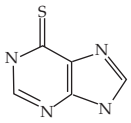
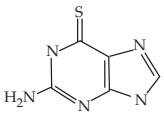
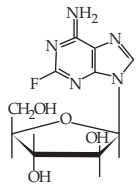
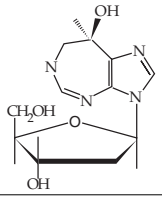
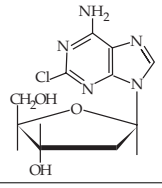
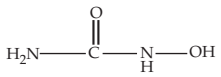
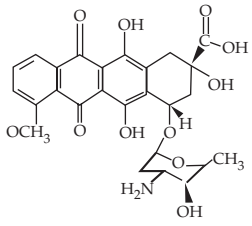
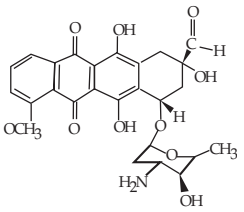
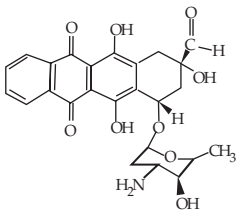
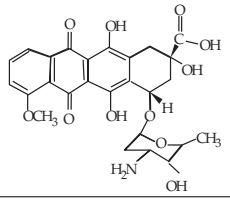
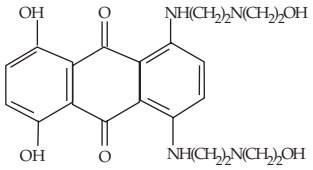
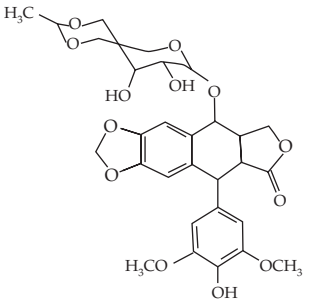
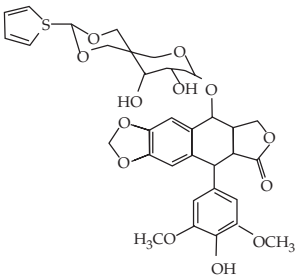
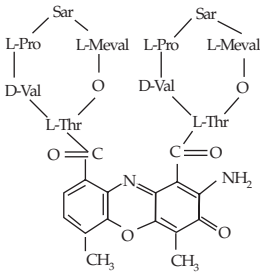
	Class	*Structure*	*Route*	*Target*	*Uses*	*Toxicities*
Nucleic Acid Synthesis Inhibitors (continued)	Gemcitabine		I.V.	Analogue of 2′-deoxycytidine; inhibits DNA synthesis; incorporated into DNA; inhibits ribonucleotide reductase	Pancreatic, bladder, breast, and lung cancers	Myelosuppression, flulike syndrome, nausea
	Purine Analogues Mercaptopurine		p.o.	Inhibits de novo purine production; incorporates into DNA	ALL, AML, chronic myeloid leukemia (CML)	Myelosuppression, anorexia, nausea and vomiting, jaundice
	Thioguanine			Inhibits de novo purine production; incorporates into DNA	ALL, AML, CML	Myelosuppression, anorexia, nausea and vomiting
	Fludarabine		I.V.	Inhibits DNA polymerase, primase, ribonucleotide reductase DNA, and RNA	CLL, low-grade non-Hodgkin lymphoma	Myelosuppression, fever, chills, asthenia, anorexia; depletion of CD4 cells can lead to opportunistic infections; peripheral neuropathy, altered mental status, seizure, optic neuritis, and coma at higher doses
	Pentostatin (2-deoxycoformycin)		I.V.	Inhibits adenosine deaminase and blocks DNA synthesis	Hairy-cell leukemia, cutaneous T cell lymphoma, CLL	Myelosuppression, nausea, vomiting, skin rashes, abnormal liver function; depletion of T cells can lead to opportunistic infections
	Cladribine (2-chlorodeoxyadenosine)		I.V.	Incorporated into DNA and leads to strand breakage	Hairy-cell leukemia, CLL, cutaneous T cell lymphoma, low-grade lymphoma, Waldenstrom macroglobulinemia	Myelosuppression, nausea, infections, fever, headache, asthenia, skin rashes, tumor lysis syndrome
	Hydroxyurea		p.o.	Blocks conversion of ribonucleotides to deoxyribonucleotides; inhibits DNA synthesis	CML, polycythemia vera, essential thrombocytosis, melanoma	Myelosuppression, dermatologic changes
DNA Topoisomerase Inhibitors	*Anthracyclines* Doxorubicin		I.V.	DNA intercalation; free radical generation after metabolism by cytochrome P-450 reductase and reduced nicotinamide adenine dinucleotide phosphate, peroxidation of cell membrane lipids	Sarcomas; Hodgkin disease, non-Hodgkin lymphoma, ALL; bladder, thyroid, lung, breast, and gastric cancers; neuroblastoma	Cardiomyopathy, myelosuppression, mucositis, radiation recall, local flare reactions, vesication
	Daunomycin		I.V.		AML, ALL	

Table 2 Chemotherapeutic Agents (*continued*)

	Class	*Structure*	*Route*	*Target*	*Uses*	*Toxicities*
DNA Topoisomerase Inhibitors (continued)	Idarubicin		I.V.		AML, ALL	
	Epirubicin		I.V.		Sarcomas; Hodgkin disease, non-Hodgkin lymphoma, ALL; bladder, lung, gastric, and breast cancers	Less cardiac toxicity, alopecia, phlebitis; myelosuppression is dose-limiting
	Anthracenediones Mitoxantrone		I.V.	Less ability to form free radicals	AML, breast cancer	Myelosuppression, nausea and vomiting, mucositis
	Epipodophyllotoxins Etoposide		I.V., p.o.	Form ternary complexes with topoisomerase II and DNA	Testicular, lung, and breast cancers; Hodgkin disease, non-Hodgkin lymphoma, AML, Kaposi sarcoma	Myelosuppression, nausea and vomiting, mucositis, alopecia, hypotension, hepatic toxicity after high doses
	Teniposide		I.V., p.o.	Form ternary complexes with topoisomerase II and DNA	Pediatric ALL and AML	Myelosuppression, nausea and vomiting, hypotension
	Dactinomycin (actinomycin-D)		I.V.	DNA intercalation, forms ternary complexes with DNA and topoisomerase II	Wilms tumor, rhabdomyosarcoma, testicular cancer, choriocarcinoma, Kaposi and Ewing sarcoma	Nausea and vomiting, anorexia, myelosuppression, mucositis, cheilitis, glossitis, alopecia, erythema and desquamation of skin, radiation recall, vesication

(*continued*)

Table 2 Chemotherapeutic Agents (*continued*)

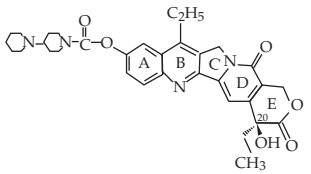
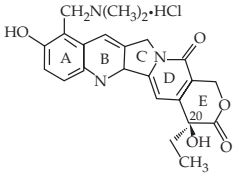
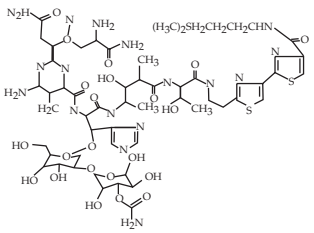
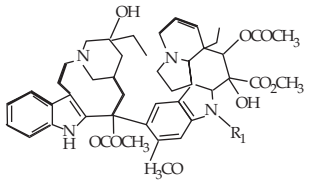
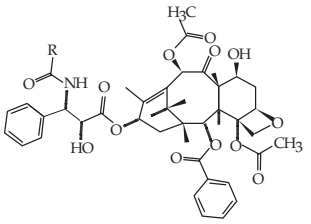
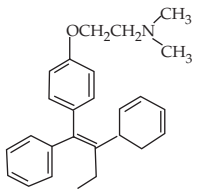
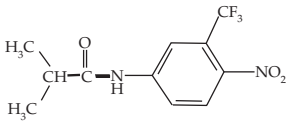
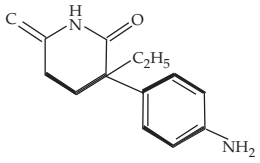
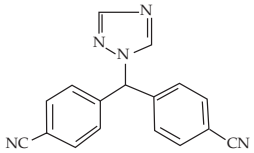
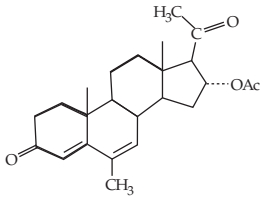
	Class	Structure	Route	Target	Uses	Toxicities
DNA Topoisomerase Inhibitors (continued)	*Camptothecins* Irinotecan		I.V.	Form ternary complexes with topoisomerase I and DNA	Colon and ovarian cancers	Myelosuppression, diarrhea, alopecia
	Topotecan		I.V.		Ovarian and lung cancers	Myelosuppression, diarrhea, alopecia
Other DNA-Damaging Agents	*Antibiotic* Bleomycin		I.V.	DNA cleavage caused by generation of reactive oxygen	Non-Hodgkin lymphoma, Hodgkin disease, testicular and head and neck cancers	Hypersensitivity reactions most prominent in patients with lymphomas, skin pigmentation and ulceration, interstitial pulmonary fibrosis
Antimicrotubule Agents	*Vinca Alkaloids* Vinblastine $R_1 = CH_3$		I.V.	Inhibit tubulin polymerization	Hodgkin disease, non-Hodgkin lymphoma; breast and testicular cancers	Myelosuppression, peripheral neuropathy, obstipation
	Vincristine $R_1 = CHO$				ALL, neuroblastoma, Wilms tumor, rhabdomyosarcoma, Hodgkin disease, non-Hodgkin lymphoma, small cell lung cancer	Peripheral neuropathy, myelosuppression
	Vinorelbine (structure not shown)				Breast and lung cancers	Myelosuppression, peripheral neuropathy, pain
	Taxanes Paclitaxel R = 		I.V.	Inhibit tubulin depolymerization	Ovarian, breast, lung, and bladder cancers	Myelosuppression, alopecia totalis, peripheral neuropathy, pain, hypersensitivity reactions
	Docetaxel $R = OC(CH_3)_3$				Ovarian, breast, lung, and bladder cancers	Myelosuppression, alopecia totalis, peripheral neuropathy, pain, capillary leakage, cardiac arrhythmias, hypersensitivity reactions
Signal Transduction Modulators	*Antiestrogens* Tamoxifen		p.o.	Block activation of estrogen receptor by estrogens; produce TGF-β; inhibit signal transduction enzymes	Breast cancer	Amenorrhea, hot flashes, nausea, weight gain, hypercalcemia, venous thrombosis, endometrial cancer
	Toremifene (structure not shown)				Breast cancer	Hot flashes, nausea, venous thrombosis
	Raloxifene (structure not shown)				Breast cancer	Hot flashes, nausea, venous thrombosis

Table 2 Chemotherapeutic Agents (*continued*)

	Class	*Structure*	*Route*	*Target*	*Uses*	*Toxicities*
Signal Transduction Modulators (continued)	*Antiandrogens* Flutamide		p.o.	Block activation of the androgen receptor by dihydrotestosterone	Prostate cancer	Nausea, diarrhea, constipation, mastodynia gynecomastia, galactorrhea, hot flashes, loss of facial hair
	Bicalutamide (structure not shown)					Similar to flutamide
	Nilutamide (structure not shown)					Similar to flutamide; also visual disturbances (delayed adaptation to dark)
	Monoclonal Antibodies Rituximab (structure not shown)		I.V.	Binds to CD20 and activates humoral and cellular immunity	B cell non-Hodgkin lymphoma, CLL	Infusion-related fever, chills, vomiting, urticaria, rash, rigors; non–infusion-related (uncommon) arrhythmias, pain, myalgias, angioedema, bronchospasm; myelosuppression; major toxicities associated with high numbers of circulating CD20-positive cells
	Trastuzumab (structure not shown)		I.V.	Binds to HER-2/neu receptor; blocks cell proliferation; elicits immune response	Breast cancer	Fever, chills, pain at tumor site, diarrhea
	Aromatase Inhibitors Aminoglutethimide		p.o.	Aromatase	Breast cancer	Glucocorticoid deficiency, skin rash, lethargy, orthostatic hypotension
	Anastrazole					Nausea, diarrhea, asthenia, headache, hot flashes, pain
	Letrozole					Headache, nausea, vomiting, constipation, heartburn
	Gonadotropin-Releasing Agents Leuprolide	pyro-Glu-His-Trp-Ser-Tyr-D-Leu-Leu-Arg-Pro-ethylamide	I.M.	Activate gonadotropin-releasing hormone receptor and block subsequent activation by receptor desensitization	Prostate and breast cancers	Loss of libido, impotence, hot flashes, diarrhea; increased pain at sites of bony metastases (tumor flare) resulting from initial gonadotropin stimulation
	Buserelin (structure not shown)					
	Progestin Megestrol acetate		p.o.		Breast and endometrial cancers	Weight gain

tified more recently and also exhibit a broad spectrum of antineoplastic activity. Currently approved topoisomerase I inhibitors include the camptothecin derivatives topotecan and irinotecan. These drugs are active in the treatment of leukemia, ovarian cancer, small cell lung cancer, and colon cancer.

Other Drugs That Impair Nucleic Acid Function

Bleomycin is a fungal product commonly used in the treatment of testicular cancer and has a relatively unique mechanism of damaging DNA. It possesses a tripeptide component that binds DNA and a metal-binding region. In the presence of oxygen and either iron or copper, bleomycin produces free radicals that create single-strand and double-strand DNA breaks.

ANTIMICROTUBULE DRUGS

Microtubules are subcellular structures that are essential for the normal function of cells. They form the mitotic spindle; maintain cell shape; organize the location of organelles; and mediate intracellular transport and secretion, neurotransmission, axonemal flow, and cell motility.[60] Microtubules are composed of α-tubulin and β-tubulin dimers arrayed in bundles of 13 protofilaments that form hollow cylinders. In living cells, microtubules grow in spurts or may disappear altogether. This process, termed dynamic equilibrium, is an essential feature of microtubule physiology. For microtubules to elongate, they require the addition of both α-tubulin and β-tubulin bound to guanosine triphosphate (GTP). Once bound to GTP, β-tubulin forms a GTP cap at the elongating end. Rapid growth requires that GTP be bound, which increases the affinity for other tubulin molecules. During depolymerization, GTP is hydrolyzed more rapidly than it can be added, resulting in weakening of the bonds that hold the tubulin molecules together.

Microtubule-associated proteins (MAPs) bind to microtubules, promote polymerization, and mediate interactions with other cellular components. The binding of MAPs to microtubules promotes polymerization by accelerating nucleation and inhibiting the dissociation of tubulin from microtubule ends, thereby stabilizing microtubules once they have formed.

Antimitotic drugs act by interfering with the normal dynamic equilibrium of microtubules, thereby disrupting the normal function of the mitotic apparatus. In addition, by affecting microtubules in interphase cells as well as in mitotic populations, they have the potential to inhibit cell motility and normal subcellular organization.

The two major classes of antimitotic drugs used to treat cancer are the vinca alkaloids and the taxanes. Vinca alkaloids bind to depolymerized microtubules and inhibit microtubule assembly. Taxanes bind to polymerized microtubules and inhibit disassembly. The result of these interactions is the failure of the cell to undergo a normal mitosis. This leads to a death signal mediated by yet-to-be-discovered pathways.

Vinca Alkaloids

Vinca alkaloids represent the active medicinal ingredients from the pink periwinkle plant (*Catharanthus roseus*, formerly *Vinca rosea*). The observation that plant extracts produced granulocytopenia in rats led to the isolation of four active alkaloids, of which two, vincristine and vinblastine, became active therapeutic agents.[61] Today, this class includes the original two alkaloids plus vinorelbine and vindesine.

Although the vinca alkaloids share structural similarities, their spectrum of activity and toxicities are somewhat different. For example, vincristine is highly effective against non-Hodgkin lymphoma, Hodgkin disease, and pediatric solid tumors, yet it has little activity against adult solid tumors. Vinorelbine, in contrast, is active against breast and lung cancer. Vinblastine is most frequently used in the treatment of testicular cancer and non-Hodgkin lymphoma and is a third-line agent in the treatment of breast cancer.

The mechanism of action of vinca alkaloids is related to concentration.[62] At substoichiometric concentrations, they bind to high-affinity sites at the ends of microtubules ($K_a = 5.3 \times 10^{-5}$ M) and prevent microtubule polymerization. At higher concentrations, vinca alkaloids bind to low-affinity, high-capacity sites ($K_a \sim 3 \times 10^{-3}$ M), which leads to the disintegration of formed microtubules.

The major toxicities of the vinca alkaloids are predictable on the basis of their mechanism of action and include dose-limiting myelosuppression and neurotoxicity. Vinblastine and vinorelbine produce far greater neutropenia than vincristine. Nadirs occur at 4 to 10 days, with recovery seen in most patients by 7 to 21 days. All three agents cause mild alopecia and are severe vesicants. Vinorelbine may cause chest pain and other deep-seated pain of unspecified origin. Respiratory reactions, including acute bronchospasm, subacute cough, dyspnea, and pulmonary infiltrates, have also been reported.

The most frequent neurotoxicities from the vinca alkaloids are numbness and tingling of the extremities, loss of deep tendon reflexes, and distal muscle weakness. Although sensory changes are bothersome, they usually reverse over time and do not require discontinuance of the drug. Loss of motor function is a later and more ominous side effect, requiring discontinuance of the medication or a search for other contributing factors.

Taxanes

The screening of plant extracts led to the isolation of an active moiety (paclitaxel) from the Pacific yew tree, *Taxus brevifolia*, in 1971.[63] However, the difficulty in formulation of this insoluble compound and its toxicities led to an initial lack of interest. When the unique mechanism of action was identified,[64] renewed interest in taxanes resulted in the development of one of the most active chemotherapeutic agents. Docetaxel is a semisynthetic derivative that is somewhat more water soluble and more potent in vitro.

Paclitaxel and docetaxel are the only taxanes approved for clinical use. They share a broad spectrum of similar antitumor activity in breast, lung, ovarian, and bladder cancers. Both drugs are also active against lymphoid malignancies.

The taxanes block the function of the mitotic apparatus by impeding the normal function of microtubules. Unlike vinca alkaloids, which affect the rates of tubulin polymerization, the taxanes inhibit microtubule depolymerization. The mechanism of action of the taxanes is dose dependent.[65] At nanomolar concentrations, taxanes induce a mitotic block without a demonstrable increase in microtubule polymer mass. These effects are associated with abnormalities in the metaphase plate and mitotic asters. At stoichiometric concentrations (1 M drug/1 M tubulin dimer), taxanes induce polymerization of stable microtubules in the absence of GTP or MAPs and are resistant to depolymerization by calcium or low temperature.

The main difference between paclitaxel and docetaxel is the spectrum of untoward side effects. Whereas both produce peripheral neuropathy, dose-limiting bone marrow suppression,

and alopecia, docetaxel appears to have unique vascular permeability properties that can lead to peripheral edema, pleural effusions, and ascites. Fluid retention is not usually evident at cumulative doses of less than 400 mg/m^2 and may be decreased by lower single doses (< 60 to 75 mg/m^2) or premedication with dexamethasone. Docetaxel also produces skin toxicities, including an erythematous maculopapular rash of the forearms and hands. Less common effects include superficial desquamation of the hands and feet and nail changes, including brown discoloration, ridging, onycholysis, and soreness and brittleness of the fingernails. In addition, the so-called hand-foot syndrome and asthenia have also been reported. Cardiac effects, including arrhythmias and ischemia, may be higher with paclitaxel, whereas stomatitis may be greater with docetaxel. Nausea, vomiting, and diarrhea occur with both drugs but are rarely severe. Both paclitaxel and docetaxel cause type I hypersensitivity reactions, characterized by flushing, bronchospasm, dyspnea, and hypotension.

Estramustine

Estramustine is a unique anticancer agent that has received increased attention. Although originally designed to target its alkylating moiety to cells with steroid hormone receptors, its activity is not dependent on the hormone-receptor interaction. It is now known that estramustine exerts at least part of its effects by binding to MAP4, leading to dissociation of MAP4 from microtubules, inhibition of microtubule assembly, and depolymerization.[66] Estramustine also binds to an estramustine-binding protein originally isolated from prostate tissue. Estramustine is part of an active combination against hormone-refractory prostate cancer.

DRUGS THAT AFFECT GROWTH FACTORS, RECEPTORS, AND SIGNAL TRANSDUCTION PATHWAYS

Normal cell division results from the interaction of growth factors with specific receptors. This interaction initiates a series of enzyme reactions (signal transduction), culminating in activation of nuclear transcription factors that produce cell proliferation molecules. It is not surprising that cancer cells usurp these normal pathways during their evolution. For example, mutations in the *HER-2/neu* oncogene result in an encoded receptor protein that does not require an extracellular ligand to activate intracellular signaling.[6] Similarly, mutations in the *ras* gene produce a form of the molecule that is locked in the on position by virtue of resistance to GTPases.[8] These activated signaling pathways may serve as critical viability factors for tumors and therefore may represent a so-called biologic Achilles heel. To date, both of these pathways have received considerable attention in the development of anticancer drugs. For example, trastuzumab is an antibody that binds to the HER-2/neu protein. Trastuzumab alone produces cell death in susceptible populations, and chemotherapeutic agents enhance its activity.

Similarly, imatinib mesylate is a 2-phenylaminopyrimidine tyrosine kinase inhibitor that specifically inhibits the Abl, platelet-derived growth factor receptor, and c-KIT tyrosine kinases. This drug is active in chronic myelogenous leukemia and in gastrointestinal stromal tumors, which are diseases dependent on the activity of the Bcr-Abl and c-KIT tyrosine kinases, respectively.

Other examples of disruption of growth factor–receptor interactions are found in the hormonal treatment of cancer. Antiestrogens, such as tamoxifen, are useful for the treatment of breast cancers that express hormone receptors (either estrogen or progesterone receptors).[67] The estrogen receptor resides in the cytosol and, upon occupation by estradiol, is transported to the nucleus, where it activates genes (including those encoding proliferation molecules) containing estrogen-response elements.

Tamoxifen binds to the estrogen receptor and disrupts this interaction in some but not all estrogen-responsive tissues. For example, tamoxifen is antiestrogenic in the breast and ovary but estrogenic in the uterus, liver, and bone. As a result, tamoxifen is an effective agent in the treatment of breast cancer but produces untoward estrogenic side effects, including blood clots (in < 1% of patients), endometrial cancer (in ~ 0.3% of patients), and early menopausal symptoms (e.g., hot flashes). In addition, tamoxifen lowers cholesterol levels and increases bone density.

The response to tamoxifen is directly proportional to the degree of expression of the hormone receptors with which it interacts. In cells that are strongly hormone receptor–positive, the response to tamoxifen is greater than 60%. In contrast, tamoxifen is of little benefit in women whose tumors do not express hormone receptors, despite the other potential therapeutic actions of this medication (e.g., stimulation of growth inhibitory factors, inhibition of growth stimulatory enzymes).

Tamoxifen is valuable in preventing and treating breast cancer. After surgical removal of the primary tumor, adjuvant treatment with tamoxifen in women whose tumors are hormone receptor–positive reduces the absolute odds of recurrence of this disease by more than 30%. For example, a woman with a node-negative cancer harbors a 10-year risk of recurrence of approximately 30%. Tamoxifen decreases the risk of recurrence by 30%. Thus, the absolute reduction is 9% (0.3 × 0.3 = 0.09), bringing the estimated risk of recurrence down to 21%. Tamoxifen may also reduce the appearance of breast cancer over a 3- to 5-year period by as much as 50% in women who do not have breast cancer but who are at high risk for the disease (on the basis of a model that defines high risk as the risk for an average 60-year-old woman in the United States). Additional antiestrogens that may have greater selectivity for the breast include raloxifene, toremifene, and fulvestrant, the last being a steroidal antiestrogen that binds to its receptor and induces receptor degradation.[68,69] Early studies of raloxifene for the treatment of osteoporosis found a decreased occurrence of breast cancer in those patients. Future studies will be required to define the long-term risks and therapeutic benefits of these newer drugs.

Antiandrogens such as cyproterone acetate and flutamide are competitive antagonists of the interaction between androstenedione and the androgen receptor. Flutamide is a nonsteroidal compound that has potent, pure antiandrogenic activity when metabolized to its hydroxylated derivative. When given in combination with drugs that decrease androgen production, flutamide is an effective treatment for hormone-dependent prostate cancer. Currently, there are three nonsteroidal antiandrogens approved by the Food and Drug Administration for the treatment of prostate cancer: bicalutamide, flutamide, and nilutamide. These drugs appear similar in regard to efficacy. The most common side effects resulting in discontinuance of nonsteroidal antiandrogens are nausea, diarrhea, and constipation. Androgenic blockade results in feminizing side effects in men, including mastodynia, gynecomastia, galactorrhea, hot flashes, and loss of facial hair. Nilutamide is associated with idiosyncratic reactions, including visual disturbances (delayed adaptation from light to dark in up to 50% of patients). Interstitial pneumonitis and alcohol intolerance have also been reported.

Indirect interference with the hormone–hormone receptor interaction can be accomplished by drugs that produce a medical

adrenalectomy or hypophysectomy or that inhibit the peripheral conversion of hormones to their active forms. For example, inhibition of aromatase is an effective treatment for postmenopausal women with breast cancer, in which the greatest source of estrogen comes from the conversion of androstenedione to estrone in liver, muscle, and fat.

Aromatase is an enzyme complex made up of two proteins, aromatase cytochrome P-450 (CYP19) and NADPH-cytochrome P-450 reductase. Inhibition of aromatase blocks the conversion of androgens to estrone in peripheral tissues, including fat, liver, muscle, and breast. Whereas treatment with aminoglutethimide was complicated by a loss of circulating glucocorticoids as well as skin rash, lethargy, and orthostatic hypotension, the high affinity of anastrozole and letrozole for CYP19 makes them far more selective and less toxic.[70] Exemestane is a steroidal aromatase inhibitor that binds to the enzyme complex and promotes enzyme degradation.

A medical hypophysectomy is produced by luteinizing hormone–releasing hormone agonists, such as leuprolide, buserelin, and goserelin, which inhibit secretion of follicle-stimulating hormone/luteinizing hormone (FSH/LH) by downregulating receptors for FSH/LH–releasing hormone in the pituitary. The end result is castrate levels of sex hormones and palliation of both breast and prostate cancers. An acute exacerbation of disease can occur when these drugs are started because of an initial surge of LH and FSH. This condition can be ameliorated by pretreatment with an androgen receptor antagonist such as flutamide.

DRUGS THAT INHIBIT METASTASES OR ANGIOGENESIS

By understanding the mechanisms of tumor growth and metastasis, it has been possible to design drugs to inhibit these processes. Tumor cells secrete a variety of proteins that facilitate angiogenesis and tissue invasion,[71] such as vascular endothelial growth factor, fibroblast growth factors, and matrix metalloproteinases (MMPs). Moreover, several intracellular tyrosine kinases are activated in endothelial cells after binding of angiogenic growth factors and are potential drug targets. Because antiangiogenesis drugs do not directly target malignant cells, drug-induced selection of resistant cells among a genetically diverse population is less likely to occur.[72]

Several drugs currently in various stages of clinical testing are designed to inhibit angiogenic growth factors. Although the mechanism of action of thalidomide is poorly understood, this drug possesses anti-angiogenic activity and is active in the treatment of multiple myeloma.[73]

GENE-BASED THERAPIES

Alterations in certain genes are known to be important in many cancers. Gene-targeting methods that are amenable to clinical application are now available, including gene-replacement or antisense therapies. The major difficulty with the clinical use of nucleic acid–based therapies is delivery of the gene vector or antisense oligonucleotide to its target.[74] Current approaches use modified viruses or liposomes. Earlier difficulties with the instability of injected antisense oligonucleotides have been lessened by modifying their structure without changing base specificity, such as replacement of oxygen with sulfur in the phosphodiester backbone, creating a more stable phosphorothioate.

Gene-based cytotoxic therapies are also under investigation, such as the use of viral vectors to deliver specific gene promoters to produce tissue-specific expression of a toxic protein or a prodrug-converting enzyme (such as herpes virus thymidine kinase, which converts ganciclovir to a cytotoxic agent). Although a gene-based therapy for cancer has not been approved, the recent approval of an antisense oligonucleotide drug (fomivirsen) for the treatment of cytomegalovirus retinitis suggests that similar approvals for cancer will be forthcoming.

There are several antisense strategies that are currently in clinical trials, including oligonucleotides that target *bcl-2* or the *bcr-abl* fusion gene. With regard to gene replacement, the frequent loss of *p53* function in malignancies has made this gene an attractive target for viral-based replacement therapies. For example, adenoviruses must inactivate *p53* to successfully invade and proliferate in host cells. In theory, tumor cells with mutant *p53* may be a better host than normal cells for adenoviruses lacking the *p53* inactivation molecule, E1A. These genetically modified viruses are currently being tested as potentially selective anticancer therapies. Alternatively, vectors can deliver wild-type *p53* to cells with mutant *p53*, with the hope that reestablishing the role of wild-type *p53* in tumor cells will initiate programmed cell death.

IMMUNOTHERAPIES

Cancer cells use several mechanisms to evade destruction by immunologic defenses. This knowledge led to methods of recruiting the immune system for the destruction of cancer cells and has resulted in agents that either activate immune effector systems (such as interferon alfa and interleukin-2) or activate a response that is specific for a particular tumor antigen.[75] The latter approach may involve either identifying or targeting a particular tumor antigen or using tumor specimens to generate vaccines.

Antibody-based therapies that have been approved for the treatment of malignancy include trastuzumab, a monoclonal antibody that targets the HER-2/neu protein; alemtuzumab, which targets CD52 on the surface of normal and malignant B and T cells; and rituxumab, which targets the CD20 antigen present on the surface of B cell lymphoma cells and chronic lymphocytic leukemia cells. Radiolabeled antibodies recognizing CD20 offer the potential of delivering a radioactive payload directly to the lymphoma. Examples include iodine-131 tositumomab and yttrium-90 ibritumomab tiuxetan, both of which are active in the treatment of patients with B cell non-Hodgkin lymphoma.[76]

Vaccine therapies are currently investigational but appear particularly promising in attempts either to prevent cancer in high-risk individuals or to treat microscopic diseases that may remain after a surgical resection. Vaccine therapy can incorporate genetic manipulation of tumors to make them more antigenic, such as transfecting tumor cells with a DNA construct that expresses cytokines to augment the immune response.

Eric. H. Rubin, M.D., has received grants for clinical research from Novartis Pharmaceuticals Corp., Pharmacia, Aventis Pharmaceuticals, Inc., AstraZeneca, Bristol-Myers Squibb Co., and Merck & Co., Inc.

References

1. Harris JR: Diseases of the Breast. Lippincott-Raven Publishers, Philadelphia, 1996

2. Aisner J: Comprehensive Textbook of Thoracic Oncology. Williams & Wilkins, Baltimore, 1996

3. Armitage JO, Weisenburger DD: New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. J Clin Oncol 16:2780, 1998

4. Hahn WC, Counter CM, Lundberg AS, et al: Creation of human tumour cells with defined genetic elements. Nature 400:464, 1999

5. Hahn WC, Stewart SA, Brooks MW, et al: Inhibition of telomerase limits the growth of human cancer cells. Nat Med 5:1164, 1999

6. Bargmann CI, Hung MC, Weinberg RA: The neu oncogene encodes an epidermal growth factor receptor–related protein. Nature 319:226, 1986

7. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244:707, 1989

8. Bos JL: ras oncogenes in human cancer: a review. Cancer Res 49:4682, 1989

9. Rowinsky EK, Windle JJ, Von Hoff DD: Ras protein farnesyltransferase: a strategic target for anticancer therapeutic development. J Clin Oncol 17:3631, 1999

10. Sausville EA, Zaharevitz D, Gussio R, et al: Cyclin-dependent kinases: initial approaches to exploit a novel therapeutic target. Pharmacol Ther 82:285, 1999

11. Woodburn JR: The epidermal growth factor receptor and its inhibition in cancer therapy. Pharmacol Ther 82:241, 1999

12. Drucker BJ, Lydon NB: Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. J Clin Invest 205:3, 2000

13. Knudson AG Jr: Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 68:820, 1971

14. Levine AJ, Momand J, Finlay CA: The p53 tumour suppressor gene. Nature 351:453, 1991

15. Hollstein M, Sidransky D, Vogelstein B, et al: p53 mutations in human cancers. Science 253:49, 1991

16. Zhang CC, Yang JM, White E, et al: The role of MAP4 expression in the sensitivity to paclitaxel and resistance to vinca alkaloids in p53 mutant cells. Oncogene 16:1617, 1998

17. Greider CW: Telomerase activation: one step on the road to cancer? Trends Genet 15:109, 1999

18. Holt SE, Shay JW: Role of telomerase in cellular proliferation and cancer. J Cell Physiol 180:10, 1999

19. Kerr JF, Wyllie AH, Currie AR: Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 26:239, 1972

20. Ruoslahti E, Reed JC: Anchorage dependence, integrins, and apoptosis. Cell 77:477, 1994

21. Christofori G, Semb H: The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. Trends Biochem Sci 24:73, 1999

22. Risau W: Mechanisms of angiogenesis. Nature 386:671, 1997

23. Tamm I, Cardinale I, Krueger J, et al: Interleukin 6 decreases cell-cell association and increases motility of ductal breast carcinoma cells. J Exp Med 170:1649, 1989

24. Norton L, Simon R, Brereton HD, et al: Predicting the course of gompertzian growth. Nature 264:542, 1976

25. Thompson J, Stewart CF, Houghton PJ: Animal models for studying the action of topoisomerase I targeted drugs. Biochim Biophys Acta 1400:301, 1998

26. Hait WN: Drug Resistance. Kluwer Academic Publishers, Boston, 1996

27. Teicher BA: Physiologic mechanisms of therapeutic resistance: blood flow and hypoxia. Hematol Oncol Clin North Am 9:475, 1995

28. Biedler JL, Riehm H: Cellular resistance to actinomycin D in Chinese hamster cells in vitro: cross-resistance, radioautographic, and cytogenetic studies. Cancer Res 30:1174, 1970

29. Kartner N, Riordan JR, Ling V: Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines. Science 221:1285, 1983

30. Ford JM, Prozialeck WC, Hait WN: Structural features determining activity of phenothiazines and related drugs for inhibition of cell growth and reversal of multidrug resistance. Mol Pharmacol 35:105, 1989

31. Advani R, Saba HI, Tallman MS, et al: Treatment of refractory and relapsed acute myelogenous leukemia with combination chemotherapy plus the multidrug resistance modulator PSC 833 (Valspodar). Blood 93:787, 1999

32. McGrath T, Center MS: Adriamycin resistance in HL60 cells in the absence of detectable P-glycoprotein. Biochem Biophys Res Commun 145:1171, 1987

33. Cole SP, Bhardwaj G, Gerlach JH, et al: Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. Science 258:1650, 1992

34. Doyle LA, Yang W, Abruzzo LV, et al: A multidrug resistance transporter from human MCF-7 breast cancer cells. Proc Natl Acad Sci USA 95:15665, 1998

35. Allen JD, Brinkuis RF, Wijnholds J, et al: The mouse Bcrp1/Mxr/Abcp gene: amplification and overexpression in cell lines selected for resistance to topotecan, mitoxantrone, or doxorubicin. Cancer Res 59:4237, 1999

36. Steinbach D, Sell W, Voigt A, et al: BCRP gene expression is associated with a poor response to remission induction therapy in childhood acute myeloid leukemia. Leukemia 16:1443, 2002

37. Rubin EH, Li T, Duan P, et al: Cellular resistance to topoisomerase poisons. Drug Resistance. Hait WN, Ed. Kluwer Academic Publishers, Norwell, Massachusetts, 1996, p 243

38. Mirski SEL, Cole SPC: Cytoplasmic localization of a mutant Mr 160,000 topoisomerase IIa is associated with the loss of putative bipartite nuclear localization signals in a drug-resistant human lung cancer cell line. Cancer Res 55:2129, 1995

39. Saleem A, Edwards TK, Shaman JA, et al: Role for Nsr1 in the cellular localization of topo 1p and in sensitivity to camptothecin. Proc Am Assoc Cancer Res 41:A1344, 2000

40. Desai SD, Liu LF, Vazquez-Abad D, et al: Ubiquitin-dependent destruction of topoisomerase I is stimulated by the antitumor drug camptothecin. J Biol Chem 272:24159, 1997

41. Gupta E, Toppmeyer D, Zamek R, et al: Clinical evaluation of sequential topoisomerase targeting in the treatment of advanced malignancy. Cancer Therapeutics 1:292, 1998

42. Schimke RT, Alt FW, Kellems RE, et al: Amplification of dihydrofolate reductase genes in methotrexate-resistant cultured mouse cells. Cold Spring Harb Symp Quant Biol 42:649, 1978

43. Bailey HH: L-S,R-buthionine sulfoximine: historical development and clinical issues. Chem Biol Interact 111:239, 1998

44. Reed JC: Regulation of apoptosis by bcl-2 family proteins and its role in cancer and chemoresistance. Curr Opin Oncol 7:541, 1995

45. Goldie JH, Coldman AJ: A mathematic model for relating the drug sensitivity of tumors to their spontaneous mutation rate. Cancer Treat Rep 63:1727, 1979

46. Borison HL, McCarthy LE: Neuropharmacology of chemotherapy-induced emesis. Drugs 25(suppl 1):8, 1983

47. Socie G: Secondary malignancies. Curr Opin Hematol 3:466, 1996

48. Tinwell H, Ashby J: Genetic toxicity and potential carcinogenicity of taxol. Carcinogenesis 15:1499, 1994

49. Gurney H: Dose calculation of anticancer drugs: a review of the current practice and introduction of an alternative. J Clin Oncol 14:2590, 1996

50. Evans WE, Relling MV, Rodman JH: Conventional compared with individualized chemotherapy for childhood acute lymphoblastic leukemia. N Engl J Med 338:499, 1998

51. Ridge SA, Sludden J, Wei X, et al: Dihydropyrimidine dehydrogenase pharmacogenetics in patients with colorectal cancer. Br J Cancer 77:497, 1998

52. Wasserman E, Myara A, Lokiec F, et al: Severe CPT-11 toxicity in patients with Gilbert's syndrome: two case reports. Ann Oncol 8:1049, 1997

53. Krajinovic M, Costea I, Chiasson S: Polymorphism of the thymidylate synthase gene and outcome of acute lymphoblastic leukaemia. Lancet 359:1033, 2002

54. Iacopetta B, Grieu F, Joseph D, et al: A polymorphism in the enhancer region of the thymidylate synthase promoter influences the survival of colorectal cancer patients treated with 5-fluorouracil. Br J Cancer 85:827, 2001

55. Stoehlmacher J, Park DJ, Zhang W, et al: Association between glutathione S-transferase P1, T1, and M1 genetic polymorphism and survival of patients with metastatic colorectal cancer. J Natl Cancer Inst 94:936, 2002

56. Lutzker SG, Levine AJ: A functionally inactive p53 protein in teratocarcinoma cells is activated by either DNA damage or cellular differentiation. Nat Med 2:804, 1996

57. Bower M, Newlands ES, Bleehen NM, et al: Multicentre CRC phase II trial of temozolomide in recurrent or progressive high-grade glioma. Cancer Chemother Pharmacol 40:484, 1997

58. Blum JL, Jones SE, Buzdar AU, et al: Multicenter phase II study of capecitabine in paclitaxel-refractory metastatic breast cancer. J Clin Oncol 17:485, 1999

59. Liu LF: DNA topoisomerase poisons as antitumor drugs. Annu Rev Biochem 58:351, 1989

60. Wilson L, Jordan MA: Microtubule dynamics: taking aim at a moving target. Chem Biol 2:569, 1995

61. Johnson IS: Historical background of vinca alkaloid research and areas of future interest. Cancer Chemother Rep 52:455, 1968

62. Himes RH: Interactions of the *Catharanthus* (vinca) alkaloids with tubulin and microtubules. Pharmacol Ther 51:257, 1991

63. Wall ME, Wani MC: Camptothecin and taxol: discovery to clinic—Thirteenth Bruce F. Cain Memorial Award lecture. Cancer Res 55:753, 1995

64. Schiff PB, Fant J, Horwitz SB: Promotion of microtubule assembly in vitro by taxol. Nature 277:665, 1979

65. Dumontet C, Sikic BI: Mechanisms of action of and resistance to antitubulin agents: microtubule dynamics, drug transport, and cell death. J Clin Oncol 17:1061, 1999

66. Speicher LA, Laing N, Barone LR, et al: Interaction of an estramustine photoaffinity analogue with cytoskeletal proteins in prostate carcinoma cells. Mol Pharmacol 46:866, 1994

67. Lippman ME: Antiestrogen therapy of breast cancer. Semin Oncol 10:11, 1983

68. Osborne CK, Pippen J, Jones SE, et al: Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. J Clin Oncol 20:3386, 2002

69. Howell A, Robertson JF, Quaresma Albano J, et al: Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. J Clin Oncol 20:3396, 2002

70. Buzdar AU, Plourde PV, Hortobagyi GN: Aromatase inhibitors in metastatic breast cancer. Semin Oncol 23:28, 1996

71. Boyd D: Invasion and metastasis. Cancer Metastasis Rev 15:77, 1996

72. Bergers G, Javaherian K, Lo KM, et al: Effects of angiogenesis inhibitors on multistage carcinogenesis in mice. Science 284:808, 1999

73. Singhal S, Mehta J, Desikan R, et al: Antitumor activity of thalidomide in refractory multiple myeloma. N Engl J Med 341:1565, 1999

74. Gewirtz AM, Sokol DL, Ratajczak MZ: Nucleic acid therapeutics: state of the art and future prospects. Blood 92:712, 1998

75. Rosenberg SA: A new era for cancer immunotherapy based on the genes that encode cancer antigens. Immunity 10:281, 1999

76. Dillman RO: Radiolabeled anti-CD20 monoclonal antibodies for the treatment of B-cell lymphoma. J Clin Oncol 20:3345, 2002

Acknowledgment

Figures 1 and 2 Seward Hung.

V COLORECTAL CANCER

Bernard Levin, m.d., f.a.c.p.

Epidemiology

Colorectal cancer is the second most common cancer and the second leading cause of cancer death in the United States.[1] An estimated 145,000 new cases and 56,000 deaths from colorectal cancer were expected to occur in 2005. Worldwide, the annual mortality from colorectal cancer is estimated to be about 500,000. The incidence and mortality of colorectal cancer increase with age, especially after 60 years of age. The lifetime probability (magnitude of absolute risk) for the development of colorectal cancer in the United States is about 6%. In the United States, the overall mortality declined by 1.8% a year from 1992 through 1998, although the mortality for African Americans remains higher than that for other ethnic and racial groups.[2]

Pathogenesis

adenomatous polyps (adenomas)

It is thought that most colorectal cancers arise from preexisting adenomas. Such potentially premalignant lesions should be distinguished from juvenile polyps, hamartomas, and inflammatory polyps, which are not thought to progress to colorectal cancer. Serrated adenomas, hyperplastic polyps, and admixed polyps may arise through a pathway different from that of conventional adenomatous polyps—that is, through abnormalities in mismatch repair.[3]

Adenomatous polyps are grossly visible, gland-forming mucosal protrusions that may be pedunculated (attached by a narrow base and a long stalk) or sessile (attached across a broad, flat base with no stalk)[4] [*see Figure 1*]. Histologically, adenomatous polyps may be tubular (composed of tubular glands extending downward from the outer surface of the polyp), villous (composed of fingerlike epithelial projections extending outward from the surface of the bowel mucosa), or both (tubulovillous).[4] The larger the adenoma, the greater the likelihood that a villous component will be present. Villous polyps are more likely to contain invasive carcinoma than are tubular polyps of the same size.[5,6] Regardless of histologic class, large polyps—especially those larger than 1.0 cm in diameter—are more likely to contain invasive carcinoma.[5,7]

Five lines of evidence suggest that colon cancers develop from premalignant adenomatous polyps:

1. Countries in which colon cancer is prevalent also have a high prevalence of colonic adenomatous polyps; conversely, a low prevalence of colon cancer correlates with a low prevalence of such adenomas.
2. Patients who have undergone resection of adenomatous polyps are at increased risk for the subsequent development of colorectal cancer.[8]
3. Adenomatous polyps occur in younger persons than do carcinomas,[9] a finding that is consistent with the suggestion that adenomas are precursor lesions.
4. Susceptibility to the development of adenomatous polyps and colorectal cancer is commonly inherited.
5. Colonoscopic polypectomy reduces the expected incidence of colorectal cancer.[10]

Adenomatous polyps are common: autopsy studies have demonstrated that such lesions are present in more than 30% of persons older than 50 years and that their prevalence increases with age.[11] However, fewer than 1% of adenomatous polyps ever become malignant.[9] After an adenomatous polyp is detected, the entire large bowel should be visualized endoscopically because synchronous lesions are found in approximately 33% of cases. Thereafter, colonoscopy should be repeated periodically, even in the absence of a documented cancer, because patients in whom one adenomatous polyp is detected have a 30% to 50% risk of developing another adenoma and are at higher than average risk for colorectal cancer.[12] The risk of subsequent colon cancer appears to depend on the histologic type, size, and number of adenomas found at the time of initial examination.[13,14] It is thought that most adenomatous polyps require more than 5 years of growth before they become clinically significant.

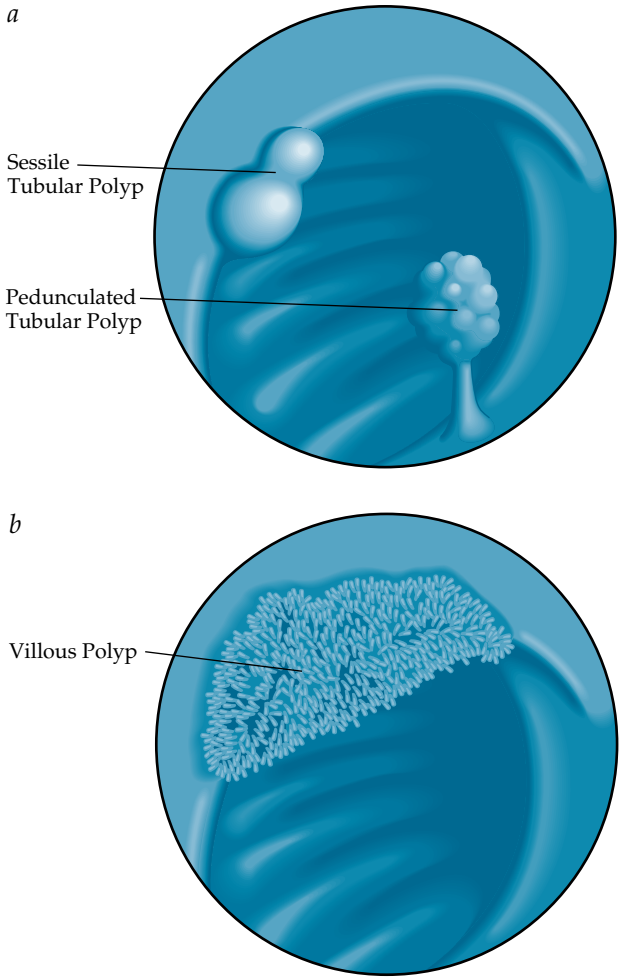


Figure 1 **It is thought that most colorectal cancers arise from adenomatous polyps, which may be tubular (pedunculated or sessile [*a*]), villous (*b*), or tubulovillous.**

GENETIC ALTERATIONS

Colorectal cancer is a heterogeneous disease arising from a complex series of molecular changes.[15] The successive evolution of normal colonic mucosa to a benign adenoma, then to an adenomatous polyp containing cancer, and then to a potentially life-threatening invasive cancer is associated with a series of genetic events occurring over a long period. The original model that was proposed over a decade ago involves key derangements in several genes, including *APC* (in patients with familial adenomatous polyposis [FAP]), K-*ras*, *DCC*, and *p53*. *APC* gene mutations alter cell adhesion by affecting the binding of β-catenin; they also affect the WnT signaling pathway. C-myc, a transcription factor, is activated by abnormal WnT signaling, skewing the balance between proliferation and apoptosis. Loss of APC influences tumor initiation, whereas inactivation of E-cadherin, which complexes with β-catenin, plays a role in tumor progression.

K-*ras* gene mutations induce proliferation via the EGFR-RAS-RAF-ERK-JUN/FOS pathway and inhibit apoptosis by phosphorylating procaspase. Other genes may be affected, including *SMAD4* and *SMAD2,* which in turn may interfere with transforming growth factor–β signaling. Additional genetic changes (e.g., DNA aneuploidy, *p53* overexpression), as well as other abnormalities, have been described in subsequent malignant transformation. Cancers arising through these mechanisms are termed microsatellite stable.

A different pathway involves deficiency of mismatch repair, as exemplified by hereditary nonpolyposis colorectal cancer. Dominant inheritance of a germline mutation involving *hMLH1, hMLH2, or hMSH6* coupled with somatic loss of the second allele results in a deficiency of mismatch repair. This mutator phenotype is characterized by the presence of multiple somatic mutations involving microsatellite repeats, termed microsatellite instability (MSI). MSI is also found in about 15% of sporadic colorectal cancers and is usually the result of acquired loss of *hMLH1* through hypermethylation of the gene promoter. Dense aggregates of the cytosine-guanine dinucleotide sequence (CpG dinucleotides) may occur in the promoter regions of several genes; these aggregates are termed CpG islands. Extensive methylation of cytosine bases is associated with promoter silencing. Cancers demonstrating methylation of multiple genes are described as CIMP (CpG island methylation phenotype) positive.

Etiology

As with many other cancers, the development of colorectal cancer typically results from a complex interaction between genetic and environmental influences.

HEREDITARY SYNDROMES AND PREDISPOSING CONDITIONS

As many as 25% of patients with colorectal cancer have a family history of the disease, which suggests the involvement of a genetic factor. Such inherited colon cancers can be divided into two main types: the well-studied but rare FAP syndrome, which accounts for approximately 1% of cases of colon cancer annually, and the increasingly well-characterized, more common hereditary nonpolyposis colorectal cancer (HNPCC), which accounts for 5% to 10% of cases.[16]

Familial Adenomatous Polyposis

FAP is an autosomal dominant disorder characterized by the appearance of thousands of adenomatous polyps, each typically smaller than 1 cm in diameter, throughout the entire large bowel. Such polyps may occasionally be accompanied by extracolonic findings, such as osteomas, desmoid tumors, epidermoid and sebaceous cysts, pigmented retinal lesions, upper gastrointestinal tract polyps, and periampullary cancers (Gardner syndrome) or brain tumors (Turcot syndrome).[16] Persons with FAP are born with normal-appearing colonic mucosa; polyps develop during the second and third decades of life. If surgical treatment is not performed, colorectal cancer is almost certain to develop by 40 years of age.

FAP is associated with a deletion of chromosome 5q21 (known as the *APC* gene)[17] in neoplastic cells (somatic mutation) and normal cells (germline mutation); this deletion apparently leads to abnormal proliferative patterns in the colonic mucosa.[18] Mutations at the 5′ end (first five exons), in exon 9, and in the distal 3′ end of the *APC* gene result in an attenuated form of FAP characterized by fewer adenomas, a proximal colonic distribution of polyps, a somewhat delayed development of adenomas and cancer, and a decreased colon cancer risk.[19]

Genetic testing is now the standard of care for FAP. Despite the detailed genetic knowledge of FAP that is now available, genetic testing is often poorly interpreted. Consequently, genetic counseling is an integral part of management and should precede genetic testing.[20] Testing for FAP in a family is most informative when it begins with the affected family member, to identify the mutation responsible for FAP within that family. Once a causal mutation has been identified in an affected person, predictive testing can be done to identify other family members at risk.[21] DNA testing for *APC* gene mutations has a sensitivity of 70% to 90% and a specificity of 100%. If the test result is positive or the test is not available, flexible sigmoidoscopy is performed at 10 to 12 years of age. During the procedure, mucosal biopsy specimens are taken to identify subtle adenomatous changes. Colonoscopy with mucosal biopsies is advisable at 18 to 20 years of age. If adenomas are detected, surgical prophylaxis should be considered. Routine gastroduodenoscopic surveillance is also recommended for patients with FAP, because these patients are at high risk for potentially precancerous gastric and duodenal adenomas.[22]

MYH is a base excision repair gene that is involved in repairing oxidative damage to DNA. Loss of *MYH* is associated with G:C to T:A transversions in APC, leading to polyp formation and carcinoma. The two most common *MYH* mutations, G382D and Y165C, account for approximately 85% of MYH-associated polyposis. Appropriate candidates for mutational analysis testing are patients with multiple adenomas or FAP who have a family history compatible with a recessive pattern of inheritance (i.e., colon cancers or multiple adenomas in only one generation or in skipped generations) and in whom testing has failed to show a germline *APC* mutation.

Surgical prophylaxis in FAP consists of resection of the entire large bowel to prevent malignant transformation. In the past, surgical alternatives included total colectomy with a permanent ileostomy and subtotal colectomy with an ileorectal anastomosis; the latter procedure is complicated by the frequent appearance of rectal polyps, which often necessitates a subsequent proctectomy. Currently, total proctocolectomy with J-pouch ileoanal anastomosis is advocated as surgical prophylaxis.

In a prospective, randomized trial of 83 FAP patients, a 6-month regimen of celecoxib was shown to reduce the number of colorectal adenomas by an average of 28%, compared with a 5% reduction with placebo ($P = 0.003$).[23] On the basis of this study, the Food and Drug Administration approved celecoxib as oral adjunctive

therapy for adults with FAP. Nevertheless, endoscopic surveillance and colectomy as indicated remain the standard of care.

Hereditary Nonpolyposis Colorectal Cancer

HNPCC, like FAP, is an autosomal dominant disorder. The median age at which adenocarcinomas appear in HNPCC is less than 50 years, which is 10 to 15 years younger than the median age at which they appear in the general population.[24,25] In contrast to FAP, HNPCC is associated with an unusually high frequency of cancers in the proximal large bowel. Also, families with HNPCC often include persons with multiple primary cancers; in women, an association between colorectal cancer and either endometrial or ovarian carcinoma is especially prominent.

Several sets of selection criteria have been developed for identifying patients with this syndrome. The Amsterdam-2 criteria comprise the following: histologically documented colorectal cancer (or other HNPCC-related tumor) in at least three relatives, one of whom is a first-degree relative of the other two; a family history of one or more cases of colorectal cancer diagnosed before 50 years of age; and cases of colorectal cancer in at least two successive generations of the family. Affected relatives should be on the same side of the family (maternal or paternal), FAP must be excluded in colorectal cancer cases, and tumors must be pathologically verified. Another selection set, the Bethesda criteria, is more sensitive than the Amsterdam criteria but is less specific. These selection criteria were developed to identify patients whose tumors should be tested for features consistent with HNPCC, such as MSI.[26,27] For tumors that display the MSI-H phenotype, *hMLH1* and *hMSH2* germline mutational analysis is indicated.

In addition to testing for MSI, genetic testing is available for three mismatch repair genes, *hMSH2, hMLH1,* and *hMLH6*, which account for about 90% of all HNPCC cases. Unfortunately, tests for mutations of these genes are not perfectly sensitive, and genetic sequencing is expensive. The discovery of a mutation in a family provides a rationale for predictive testing for at-risk family members, although such testing may cause the unaffected individuals unnecessary concern and lead to unnecessary screening procedures.[28]

If HNPCC is confirmed, affected family members should undergo colonoscopy between the ages of 20 and 25 or at the age that is 10 years less than the earliest age at diagnosis in the family, whichever is earlier. This procedure should be repeated every 1 to 2 years. Recommended screening for women includes an annual transvaginal ultrasound or endometrial aspiration, beginning at age 25 to 35 years.[29]

If an adenoma or adenocarcinoma of the colon is identified, total abdominal colectomy with an ileorectal anastomosis is recommended. In women, total abdominal hysterectomy and bilateral salpingo-oophorectomy are often considered, particularly if the patient has no intention of having children in the future because of the increased risk of ovarian and endometrial carcinoma.

Inflammatory Bowel Disease

Long-standing, extensive inflammatory bowel disease, including both ulcerative colitis and Crohn colitis, increases the risk of colon cancer. Surveillance colonoscopy with multiple biopsies of the entire colon should be considered every 1 to 2 years after 8 years of disease in patients with pancolitis or after 12 to 15 years of disease in those with left-sided colitis. Colectomy must be carefully considered in each case, depending on the biopsy results.[29]

ENVIRONMENTAL FACTORS

Specific factors that increase the risk of colorectal cancer have been identified, as have factors that reduce risk [*see Table 1*].

Diet

Diet plays a complex role in the etiology of colorectal cancer [*see Table 2*].[30] Both the total energy intake and individual components of the diet have been implicated.

The relationship between total energy intake and colorectal cancer risk is not simple; possible biologic mechanisms include increased levels of endogenous hormones (e.g., sex steroids, insulin, and insulin growth factor).[30] Nevertheless, a prospective study of a large United States cohort suggested that obesity (i.e., high body mass index [BMI]) is a significant risk factor for the development of colorectal cancer.[31] The relative risks of BMIs of 25 to 29.9, 30 to 34.9, and 35 to 35.9 were 1.20, 1.47, and 1.84, respectively. These data are consistent with previous case-control and cohort studies.

Dietary fat Many experimental studies have shown that intestinal tumorigenesis is enhanced as the fat content of the diet is increased. Dietary fat is thought to increase the concentration of bile acid in the bowel or to promote the formation of excess intraluminal diacylglycerol as the result of the interaction of fat, bile acids, and bacteria.[32] The effect of diacylglycerol may be to amplify cell-replication signals.[33] Early studies supported the relationship of dietary fat to colorectal cancer; these studies showed that colon cancer rates were high in populations with high total fat intake, with odds ratios of 1.3 to 2.2, and were lower in those consuming less fat.[34] However, many of these early studies failed to adjust for total energy intake. Cohort studies and a meta-analysis of 13 case-control studies of colorectal cancer failed to find clear evidence of the association of colorectal cancer with dietary fat intake.[32] In the aggregate, the epidemiologic and experimental evidence suggests that diets high in total fat may increase the risk of colorectal cancer.[30] A review of the data concerning

Table 1 Risk Factors for Colorectal Cancer

Average risk	Age 50 years and older
Decreased risk	High vegetable consumption Oral contraceptive use Estrogen replacement Multivitamins containing folic acid Long-term use of aspirin and other NSAIDs

Increased risk	Family history
	  Colorectal cancer
	  Colorectal adenomas
	Personal history
	  Colorectal adenomas
	  Ovarian, uterine cancer
	Familial adenomatous polyposis
	Hereditary nonpolyposis colorectal cancer
	Peutz-Jegher syndrome
	Juvenile polyposis
	Inflammatory bowel disease of long standing
	Physical inactivity (< 3 hr of exercise a week)
	Obesity
	Smoking
	Alcohol (> 1 drink/day)

NSAIDs—nonsteroidal anti-inflammatory drugs

Table 2 Food and Nutrition and the Risk of Colorectal Cancer

Strength of Evidence	*Decreases Risk*	*No Relationship*	*Increases Risk*
Convincing	Vegetables (not fruit)	—	—
Probable	—	—	Red meat Alcohol
Possible	Fiber Starch Carotenoids	Calcium Selenium Fish	High body mass* Greater adult height Frequent eating Sugar Total fat Saturated/animal fat Processed meat Eggs Heavily cooked meat
Insufficient	Resistant starch Vitamin C Vitamin D Vitamin E Folate Methionine Cereals Coffee	—	Iron

*Colon cancer only.

saturated fat concluded that diets high in fat possibly increased the risk of colorectal cancer but that the evidence relating risk to intake of monounsaturated fat and polyunsaturated fat was inconsistent.[30] High fat intake has also been found to increase the risk of adenoma recurrence after polypectomy.[35]

Meat An authoritative review concluded that red meat intake is associated with increased risk of colorectal cancer and that processed meat possibly increases the risk.[30] However, the data concerning red meat are not entirely consistent. The Nurses' Health Study reported that persons who consumed red meat frequently had an increased risk of colon cancer (relative risk, 2.5), compared with those who rarely consumed red meat.[36] No increase in risk with meat or fat consumption was seen, however, in two other large prospective studies: the American Cancer Society's Cancer Prevention Study II and the Iowa Women's Health Study.[37,38] It has been hypothesized that heterocyclic amines that are formed when fish or meat is cooked at high temperature may increase the risk, but not all of the mechanisms are well understood.

Dietary fiber The term fiber is used to describe a complex mixture of compounds that include insoluble fiber (typified by wheat bran) and soluble fiber (oat bran). Ingestion of fiber could modify carcinogenesis in the large bowel by a number of potential mechanisms.[39] A meta-analysis of 13 case-control studies from nine countries concluded that intake of fiber-rich foods is inversely related to cancers of both the colon and the rectum.[40] Despite the evidence of a protective effect from case-control studies, results from a large prospective trial in women (the Nurses' Health Study) found no difference in risk of colorectal cancer between the highest and the lowest quintile groups with respect to dietary fiber.[41] In a multicenter, randomized, controlled trial, a diet low in fat (20% of total calories) and high in fiber, fruits, and vegetables did not reduce the risk of recurrence of colorectal adenomas.[42] High-fiber cereal supplements also failed to influence the rate of recurrence of colorectal adenomas.[43]

Vegetables and fruit Many epidemiologic studies have examined the relationship between fruit and vegetable intake and the incidence of colon or rectal cancer, with considerable variation in the findings. A prospective study utilized food-frequency questionnaires to study dietary intake in 88,764 women and 47,325 men; no association was found in men or women between overall fruit and vegetable consumption and risk of colon or rectal cancer.[44]

Calcium Calcium may indirectly inhibit colorectal cancer by binding bile acids into insoluble soaps, thereby blocking contact with the luminal epithelium. It may also modulate protein kinase C and fatty acid–induced destabilization of cellular membranes. Experimental studies and epidemiologic studies[45] have reported an inverse relationship between calcium intake and cancer risk. A randomized, placebo-controlled trial tested the effect of calcium supplementation (3 g of calcium carbonate daily, which is equivalent to 1,200 mg of elemental calcium) on the risk of recurrent adenoma.[46] The effect was modest; supplemental calcium reduced the risk of recurrence by 19%. The effect may be dependent on dose. It is possible to safely administer up to 2,000 mg of calcium daily.

Antioxidants It has been postulated that antioxidants such as retinoids, carotenoids, ascorbic acid, α-tocopherol, and selenium prevent carcinogen formation by neutralizing free radicals. Epidemiologic evidence of their benefit is difficult to confirm, however, because antioxidants and other putative cancer-prevention agents are comingled in common foods such as fruits and vegetables. β-Carotene may reduce the risk of adenoma recurrence in nonsmokers but increases the risk in those who both smoke and drink.

Folate and methionine Fresh fruit and leafy green vegeta-

bles are rich in folate, whereas red meat, chicken, and fish have relatively high concentrations of methionine. Both folate and methionine supply the methyl groups necessary for essential cellular processes such as nucleotide synthesis and gene regulation. Retrospective and prospective studies support an inverse association between dietary folate or methionine intake and the risk of colorectal adenomas and carcinomas.[47] Both the level and duration of intake are important. For persons in the highest folate and methionine quintiles, the risk of distal colorectal adenomas was approximately 35% lower in participants in the Health Professionals Follow-Up Study and 25% lower in the Nurses' Health Study.[48] In women, the long-term (≥ 15 years) use of folate-containing multivitamins had a striking preventive benefit (75% risk reduction), whereas use of such multivitamins for 4 years had no preventive benefit.[49] Prospective studies of folate in patients who underwent resection of adenomas are in progress.

Medications

The ability of specific chemical compounds to reduce the incidence of colorectal cancer has been demonstrated in epidemiologic studies and is currently undergoing clinical testing [*see Table 3*]. Technological advances, including genomics and proteomics, will facilitate the identification of new molecular targets for chemoprevention.

Postmenopausal hormone therapy Many epidemiologic studies have examined the possible associations between exogenous estrogens and colorectal neoplasia risk. In a meta-analysis of 18 epidemiologic studies, postmenopausal hormone therapy (HT) was associated with a 33% reduction in the risk of colon cancer in recent users; the relative risk was 0.67, compared with a relative risk of 0.92 in women who had used HT more than 1 year ago.[50] Similarly, HT may protect against adenoma formation. In the Women's Health Initiative trial, which included over 16,000 postmenopausal women, the combination of estrogen and progesterone was associated with a reduction in the risk of colorectal cancer (hazard ratio, 0.56; 95% confidence interval, 0.38–0.81; $P = 0.003$).[51] Invasive colorectal cancers in the hormone-treated group were similar in histologic features and grade to those in the group receiving placebo. However, in patients who received HT, the number of positive lymph nodes was slightly greater, and the cancer was more advanced regionally or metastatically.

Table 3 Selected Agents under Study for Prevention of Colorectal Cancer in Humans[124]

Nonspecific NSAIDs	Miscellaneous
Aspirin	Folic acid
Piroxicam	Sulindac sulfone
Sulindac	Ursodeoxycholic acid
Cyclooxygenase-2–selective NSAID	DFMO
Celecoxib	Combinations
	Sulindac + DFMO
	Aspirin + folic acid
	Piroxicam + calcium
	Celecoxib + DFMO
	Celecoxib + selenium

DFMO—difluoromethylornithine NSAIDs—nonsteroidal anti-inflammatory drugs

Nonsteroidal anti-inflammatory drugs Most epidemiologic studies have reported reductions in the incidence of colorectal adenomas, colorectal cancer, and colon cancer mortality associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs), including aspirin. For example, a prospective study in over 660,000 adults found that mortality from cancers of the colon and rectum was about 40% lower in regular users of aspirin.[52] The Health Professionals Follow-Up Study of 47,000 men found that the regular use of aspirin (at least twice a week) was associated with a 30% overall reduction in colorectal cancer.[53] However, in the Physicians' Health Study, there was no reduction in invasive cancers or adenomas at a median follow-up of 4.5 years when aspirin was taken at a dosage of 325 mg every other day.[54] A study of patients with prior colorectal cancer who had undergone curative resection found that compared with placebo, 325 mg of aspirin a day over 13 months was associated with a 35% reduction in the risk of recurrent adenoma, as well as a prolongation in the time to detection of a first adenoma.[55] In a study of patients with a recent history of adenomas who had undergone resection by colonoscopic polypectomy, ingestion of 81 mg of aspirin a day for an average of nearly 3 years reduced the risk of recurrence of any adenoma by 19%.[56] For advanced neoplasms (i.e., adenomas measuring at least 1 cm in diameter or with tubulovillous or villous features, severe dysplasia, or invasive cancer), the reduction was about 41%. Interestingly, 325 mg of aspirin appeared to have minimal effect. The reason for this discrepancy in the dose-effect relationship is unknown. Side effects of treatment included upper gastrointestinal hemorrhage and hemorrhagic stroke. The potential value of other NSAIDs for primary prevention of colorectal neoplasia is being studied. Although rofecoxib was associated with a reduction in sporadic polyp recurrence in a randomized, placebo-controlled trial, this drug was removed from the market because of an increase in the risk of serious cardiovascular events with long-term use.[57] Similar trials with celecoxib have closed because of concern over increased cardiovascular risk.[58] Efficacy analyses are still being conducted to determine whether polyp recurrence is diminished.

Drugs in familial adenomatous polyposis Several medications may reduce colorectal cancer risk in FAP. Sulindac reduces the size and number of adenomas.[59] A randomized, double-blind, placebo-controlled study of 83 patients with FAP found that patients who received 400 mg of celecoxib twice a day had a 28% reduction in the mean number of colorectal adenomas and a 30.7% reduction in polyp burden (i.e., the sum of polyp diameters), as compared with a group receiving placebo; a lower dose (100 mg twice daily) was associated with a non–statistically significant reduction.[23]

Lifestyle Factors

Physical activity Most studies have shown an inverse relationship between physical activity and colon cancer incidence.[60] The average relative risk reduction provided by regular physical activity is 40%. In men, physical activity for 2 hours or more a week was more strongly associated with a reduction in risk of advanced adenomas than in a reduction in risk of nonadvanced

Table 4 Evidence Supporting the Effectiveness of Colorectal Cancer Screening Tests[125]

Test	*Quality of Evidence**	*Comments*
Fecal occult blood test (FOBT)	I	33% reduction in colorectal cancer mortality with annual rehydrated FOBT; 15%–18% reduction in colorectal cancer mortality with biennial, unrehydrated FOBT
Sigmoidoscopy	I	80% reduction in colorectal cancer incidence with flexible sigmoidoscopy
	II	60%–95% reduction in mortality for cases of distal colorectal cancer
Colonoscopy	III	Sensitivity 27%–47% greater than that of 60-cm flexible sigmoidoscopy for advanced adenomas
Double-contrast barium enema	III	Sensitivity for adenomatous polyps lower than that of colonoscopy but higher than that of FOBT

*I = randomized controlled trial; II = controlled observational study (case-control or cohort); III = descriptive study.

adenomas.[61] The mechanism by which physical activity provides protection is unknown but may be linked to effects on colonic mucosal prostaglandins.

Cigarette smoking Most case-control studies of cigarette exposure and adenomas have found an elevated risk for smokers.[62] In addition, a significantly increased risk of adenoma recurrence after polypectomy has been associated with smoking in both men and women.[63] In the Nurses' Health Study, the minimum induction period for cancer appears to be at least 35 years.[64] In the Cancer Prevention Study II, a large national cohort study, mortality was highest in current smokers, intermediate in former smokers, and lowest in persons who never smoked; an increased risk was observed in both men and women after 20 or more years of smoking.[63]

Screening

Screening and early detection (secondary prevention) are important in influencing the outcome in patients with colorectal neoplasia. Many deaths from colorectal cancers could probably be averted by appropriate use of screening.[65] Despite the acknowledged benefits of screening, most average-risk persons in the United States do not undergo screening for colorectal cancer. Data from the 1999 Behavioral Risk Factor Surveillance System Survey indicate that only 21% of respondents had undergone a fecal occult blood test (FOBT) within the previous year, and approximately 34% had undergone sigmoidoscopy or colonoscopy within the previous 5 years.[66] Improving screening rates for these persons remains a significant challenge.

A screening test is intended to distinguish those most likely to have a neoplastic lesion from those least likely. Those with abnormal results are advised to undergo diagnostic tests to confirm the presence or absence of cancer. The rationale for screening for colorectal neoplasia is twofold: first, detection of adenomas and their removal will prevent subsequent development of colorectal cancer; second, detection of localized tumors in asymptomatic individuals will increase the surgical cure rate.

Of the four screening tests currently in routine use, FOBT is supported by the strongest evidence. Intermediate-level evidence is available for flexible sigmoidoscopy. Only indirect evidence supports the use of colonoscopy and double-contrast barium enema (DCBE) [*see Table 4*].

FECAL OCCULT BLOOD TESTING

The rationale for screening for the presence of blood in the stool is that large adenomas and most cancers bleed intermittently. Small adenomas rarely bleed. Annual testing is recommended with FOBTs because randomized trials have demonstrated that testing every 2 years is less effective. Annual testing may allow detection of disease that, although undetected on previous occasions, has not yet reached an advanced and perhaps incurable stage. Meta-analysis of mortality results from randomized, controlled trials shows that in patients who underwent screening with FOBTs, colorectal cancer mortality decreased by 16% (relative risk, 0.84). When adjusted for screening attendance in the individual studies, the mortality reduction was 23% (relative risk, 0.77).[67]

Estimates of the sensitivity of FOBT have ranged from 25% to over 90%. In studies that show a higher sensitivity, the figure usually refers to so-called program sensitivity, which is the effectiveness of repeated FOBTs over several years. The major concern over low sensitivity is that patients who have a falsely negative test may be falsely reassured. A high false positive rate has also been a concern, because it results in persons who are free of colorectal neoplasia undergoing follow-up screening colonoscopy, with its associated risks and costs. It is estimated that if 10,000 people were included in a biennial FOBT screening program and two thirds underwent at least one test, there would be 8.5 deaths from colorectal cancer prevented over 10 years. Data from the Minnesota trial suggest that the screening process would also result in 2,800 participants having at least one colonoscopy, with 3.4 colonoscopy complications (perforation or hemorrhage).[67] Compared with endoscopic tests, FOBT detects relatively few adenomas; the principal benefit of an FOBT program is to increase detection of early-stage cancers.

Two types of FOBTs are used: chemical (guaiac-based) and immunochemical [*see Tables 5 and 6*]. Guaiac-based tests, which detect peroxidase activity, have been the most commonly used in population screening; typically, patients take test cards home and collect two samples from three consecutive specimens. These tests are subject to false positive results caused by dietary substances or drugs. Newer FOBTs have enhanced guaiac reagents, which improve sensitivity with little loss of specificity if recommended dietary measures are followed. The sensitivity of immunochemical FOBTs is better than that of guaiac-based tests, without an unacceptable decline in specificity.[68] Newer immunochemical FOBTs are more user friendly and can be read by automated techniques.[69] Immunochemical tests do not require any alteration in diet or medication intake.

ENDOSCOPIC SCREENING TESTS

Examination of the large bowel by flexible sigmoidoscopy or colonoscopy permits direct visualization of the mucosa and allows photodocumentation, biopsy of suspicious lesions, and endoscopic polypectomy.

Table 5 Main Features of Different Types of Fecal Occult Blood Tests (FOBTs)[65]

Type of FOBT	*Basis*	*Stool-Sampling Method*	*End Point*
Chemical	Guaiac; detects peroxidase	Wooden spatula and fecal smear for most	Blue color on paper card
Immunochemical	Antihemoglobin antibody	Wooden spatula, spoon, or brush	Latex or red cell agglutination Solid-phase immunochromatography Enzyme-linked immunosorbent assay

Flexible Sigmoidoscopy

A case-control study demonstrated a risk reduction of 70% for death from cancers within reach of the sigmoidoscope; the data suggested that the benefit may last as long as 10 years.[70] A prospective study showed that screening (primarily with flexible sigmoidoscopy) was associated with a 60% reduction in the incidence of distal colorectal cancer.[71] In a prospective, randomized trial in Norway, 400 patients underwent screening with flexible sigmoidoscopy; those patients in whom polyps were detected underwent colonoscopy. This protocol led to an 80% reduction in the incidence of colorectal cancer, compared with results in unscreened control subjects.[72] Two large-scale, prospective, randomized, controlled studies are examining the efficacy of flexible sigmoidoscopy. In the United States, the Prostate, Lung, Colorectal, Ovary (PLCO) trial has enrolled 154,000 patients 55 to 74 years of age. The final results will not be available for several years.[73] In the United Kingdom, the baseline findings of a multicenter trial showed that of approximately 40,000 patients screened, distal adenomas were detected in 12%, and distal cancers were detected in 0.3%.[74] Proximal adenomas were detected in 19% of those undergoing colonoscopy, and proximal cancer was detected in 0.4%; 62% of cancers were of early stage (Dukes stage A). There was one perforation after flexible sigmoidoscopy, and four perforations after colonoscopy.[75]

Sigmoidoscopy detects 70% to 85% of advanced lesions in the entire colon.[76] Patients with an advanced distal adenoma have a 6% to 10% chance of having an advanced proximal adenoma. When a nonadvanced adenoma is found during sigmoidoscopy, the chance of a proximal advanced lesion is lower—4.7%. Studies of screening colonoscopy have suggested that patients with an apparently normal sigmoidoscopy have a 1% to 2% risk of having an advanced proximal lesion.[77] In contrast, in patients with advanced distal polyps, the prevalence of advanced proximal neoplasia was 11.5%.

Obstacles to more widespread use of flexible sigmoidoscopy include lack of training and relatively low reimbursement rates. Training of nonphysicians to perform flexible sigmoidoscopy may facilitate more widespread use of this technique, especially in high-volume centers.

Combining FOBT with flexible sigmoidoscopy is a recognized approach to screening, but the data regarding the impact on mortality are limited. In a study of 2,885 veterans (97% men; mean age, 63 years), the prevalence of advanced adenoma at colonoscopy was 10.6%. Using examination of the rectum and sigmoid during colonoscopy as a surrogate for sigmoidoscopy, these researchers estimated that combined screening with one-time FOBT and sigmoidoscopy would detect 75.8% of advanced neoplasms. This represented a statistically insignificant increase in detection rate of advanced neoplasia when compared with flexible sigmoidoscopy alone (70.3%).[77] In assessing this study, it is important to note that in calculating the detection rate, the researchers assumed that all patients with an adenoma in the distal colon would undergo colonoscopy; it is also important to note that the use of one-time FOBT in this study differs significantly from the annual or biennial method used in large-scale trials.

Colonoscopy

The effectiveness of colonoscopy has been demonstrated by several studies. Observational, case-control, and prospective, randomized trials have shown that colonoscopic polypectomy lowers the incidence of colorectal cancers by 50% to 90%.[77-81]

Three nonrandomized studies have reported the baseline results of screening colonoscopy. In a study of 3,212 United States veterans (almost all men) with a mean age of 63 years, an adenoma incidence rate of 37% was reported.[76] The incidence of advanced adenoma (defined as adenomas with a diameter of at least 10 mm, villous features, and high-grade dysplasia or cancer) was 10.5%. A study in 1,322 women reported an adenoma incidence of 21% and an advanced adenoma incidence of 3%.[81] A randomized, controlled trial of screening colonoscopy to examine efficacy

Table 6 Usage Issues with Different Types of Fecal Occult Blood Tests[65]

Type of FOBT	*Diet Restrictions*	*Drug Interference*	*Site of Bleeding Detectable*	*End Point for Test Result*
Chemical	Must avoid red meats; possibly avoid certain raw plant foods*	Vitamin C; possibly NSAIDs†	Rectum > colon > stomach (in decreasing order of sensitivity)	Subjective and transient‡
Immunochemical	None required	None required	Colon and rectum	Agglutination tests‡—can be difficult to read Immunochromatography—easy to read ELISA—machine-read

*Delaying development for 72 hr minimizes interference from plant foods and avoids the need for their restriction.
†Low-dose aspirin is not a problem, but therapeutic doses for rheumatic disorders may be so.
‡The tests generally provide a qualitative result, but newer immunochromatographic tests may be quantifiable.
ELISA—enzyme-linked immunosorbent assay FOBT—fecal occult blood test NSAIDs—nonsteroidal anti-inflammatory drugs

Table 7 American Cancer Society Recommendations for the Early Detection of Colorectal Cancer in Average-Risk, Asymptomatic Men and Women[126]

Annual fecal occult blood test (FOBT), starting at age 50
or
Flexible sigmoidoscopy every 5 yr, starting at age 50
or
Annual FOBT* and flexible sigmoidoscopy every 5 yr, starting at age 50
or
Double-contrast barium enema every 5 yr, starting at age 50
and
Colonoscopy every 10 yr, starting at age 50

*FOBT, as it is sometimes done in physicians' offices, with the single stool sample collected on the fingertip during a digital rectal examination, is not an adequate substitute for the recommended at-home procedure of collecting two samples from three consecutive specimens. Toilet-bowl FOBT tests also are not recommended. In comparison with guaiac-based tests for detection of occult blood, immunochemical tests are more patient-friendly and are likely to have equal or better sensitivity or specificity. There is no justification for repeating FOBT in response to an initial positive finding. Flexible sigmoidoscopy together with FOBT is preferable to FOBT or flexible sigmoidoscopy alone.

has not been performed, although a pilot study is in progress.[82]

The American Cancer Society currently recommends colonoscopy every 10 years, starting at age 50, for asymptomatic adults at average risk for colorectal cancer. Repeat examinations at more frequent intervals are indicated for patients at increased or high risk [*see* Implementation of Screening, *below*].

DOUBLE-CONTRAST BARIUM ENEMA

There has not been a formal trial of DCBE as a screening test for colorectal neoplasia in a general population. A comparison study in patients who have undergone colonoscopic polypectomy found colonoscopy to be a more effective method of surveillance than DCBE.[83] The proportion of examinations in which adenomatous polyps were detected by DCBE, compared with colonoscopy, was significantly related to the size of the adenomas: the rate for DCBE was 53% for detection of lesions from 0.6 to 1.0 cm in size and 48% for lesions exceeding 1.0 cm. Although current screening recommendations for colorectal cancer include DCBE, the steady decrease in training of new radiologists in this technique and the development of other imaging techniques, such as virtual colonoscopy, is likely to limit its use in the future.

EMERGING TECHNOLOGIES FOR SCREENING

Molecular Detection Methods

Detection of gene mutations in the stool has been possible for over a decade.[84] It is technically feasible to detect *APC* and *p53* mutations, long DNA, and K-*ras* mutations.[85-87] In addition, right-sided lesions can be detected by the identification of *BAT-26* mutations. In a 2004 study of over 2,500 persons, a fecal DNA mutation panel detected 29 of 71 (41%) invasive cancers and adenomas with high-grade dysplasia. Specificity was 94%. This is a rapidly advancing field, and improvements in technology are likely to enhance sensitivity of detection.[88] The cost of such techniques is high, however, and it remains to be seen whether their use will be cost-effective relative to other techniques, including newer immunochemical tests for fecal occult blood.

Virtual Colonoscopy (Computed Tomography Colonography)

CT colonography relies on sophisticated graphic software to assemble, from a fast CT scan, an endoluminal image that includes surface and volume characteristics.[89] Data from specialized centers suggest a sensitivity of 90% for lesions larger than 1 cm. Advances in this technique are occurring rapidly and include the possibility that patients will not need extensive bowel preparation. The effectiveness of the technique in screening remains to be established; large-scale multicenter trials comparing conventional colonoscopy with virtual colonoscopy are currently under way. One cost analysis has suggested that CT colonography is unlikely to be cost-effective relative to colonoscopy or other screening modalities.[90]

IMPLEMENTATION OF SCREENING

The process of screening starts with targeting patients at risk for colorectal cancer. Adults 50 years of age and older who have no other risk factors are considered at average risk, and published screening recommendations endorse regular screening in this population [*see Table 7*].[28,91,92] Screening schedules and methods vary for patients at increased or higher risk [*see Table 8*]. Follow-up is an important part of the screening process [*see Table 9*].

COST-EFFECTIVENESS OF COLORECTAL CANCER SCREENING

Analyses of cost-effectiveness of colorectal cancer screening programs have been carried out to provide a basis for legislative decision making and to construct health plans. The cost-effectiveness of colorectal cancer screening is estimated to be approximately $20,000 to $40,000 per year of life gained. This compares favorably with the cost of other usually accepted preventive services, such as end-stage renal dialysis or mammography.[93]

Diagnosis

CLINICAL MANIFESTATIONS

The presenting symptoms that lead patients with colorectal cancer to seek medical attention vary with the anatomic location of the lesion. Because stool is relatively liquid as it passes into the right side of the colon through the ileocecal valve, tumors in the cecum and ascending colon can become large and can markedly narrow the bowel lumen without causing any obstructive symptoms or noticeably altering bowel habits. Lesions in the ascending colon frequently ulcerate, which leads to chronic blood loss in the stool; however, the stool retains a grossly normal color. Therefore, patients experience symptoms related to anemia, such as fatigue, palpitations, and even angina pectoris, and are found to have a hypochromic, microcytic anemia indicative of iron deficiency. Because the tumor may bleed only intermittently, a test for fecal occult blood (see above) may not always reveal the presence of occult blood in the stool. For that reason, colonoscopy should be performed in any adult who develops iron deficiency anemia, perhaps with the exception of a young menstruating woman.

Stool becomes more concentrated as it passes into the transverse colon. Cancers arising in this section of the large bowel cause abdominal cramping, occasional obstruction, and even

Table 8 American Cancer Society Guidelines on Screening and Surveillance for the Early Detection of Colorectal Adenomas and Cancer for Women and Men at Increased Risk or High Risk[126]

Risk Category	*Description*	*Age to Begin Screening*	*Recommended Screening Method*	*Comments*
Increased risk	Persons with a single, small (< 1 cm) adenoma	3–6 yr after the initial polypectomy	Colonoscopy*	If examination is normal, the patient can thereafter be screened as per average-risk guidelines
	Persons with a large (≥ 1 cm) adenoma, multiple adenomas, or adenomas with high-grade dysplasia or villous change	Within 3 yr after the initial polypectomy	Colonoscopy*	If normal, repeat examination in 3 yr; if normal then, the patient can thereafter be screened as per average-risk guidelines
	Personal history of curative-intent resection of colorectal cancer	Within 1 yr after cancer resection	Colonoscopy*	If normal, repeat examination in 3 yr; if normal then, repeat examination every 5 yr
	Colorectal cancer or polyps in any-degree relative before age 60 or in two or more first-degree relatives at any age (if not a hereditary syndrome)	Age 40, or 10 yr before the age of the youngest patient in the immediate family	Colonoscopy*	Every 5–10 yr; colorectal cancer in relatives more distant than first-degree does not increase risk substantially above the average-risk group
High risk	Family history of familial adenomatous polyposis (FAP)	Puberty	Early surveillance with endoscopy, and counseling to consider genetic testing	If the genetic test is positive, colectomy is indicated; these patients are best referred to a center with experience in the management of FAP
	Family history of hereditary nonpolyposis colon cancer (HNPCC)	Age 21	Colonoscopy and counseling to consider genetic testing	If the genetic test is positive or if the patient has not had genetic testing, every 1–2 yr until age 40, then annually; these patients are best referred to a center with experience in the management of HNPCC
	Inflammatory bowel disease (chronic ulcerative colitis, Crohn disease)	Cancer risk begins to be significant 8 yr after the onset of pancolitis, or 12–15 yr after the onset of left-sided colitis	Colonoscopy with biopsies for dysplasia Expert pathologic evaluation is essential	Every 1–2 yr; these patients are best referred to a center with experience in the surveillance and management of inflammatory bowel disease

*If colonoscopy is unavailable, not feasible, or not desired by the patient, double-contrast barium enema (DCBE) alone or the combination of flexible sigmoidoscopy and DCBE are acceptable alternatives. Adding flexible sigmoidoscopy to DCBE may provide a more comprehensive diagnostic evaluation than DCBE alone in finding significant lesions. A supplementary DCBE may be needed if a colonoscopic examination fails to reach the cecum, and a supplementary colonoscopy may be needed if a DCBE identifies a possible lesion or does not adequately visualize the entire colorectum.

perforation. Tumors that markedly narrow the bowel lumen (so-called napkin-ring lesions) in the transverse colon are often smaller than those in the more proximal bowel because the short duration and localized pattern of the symptoms lead to a somewhat earlier diagnosis.

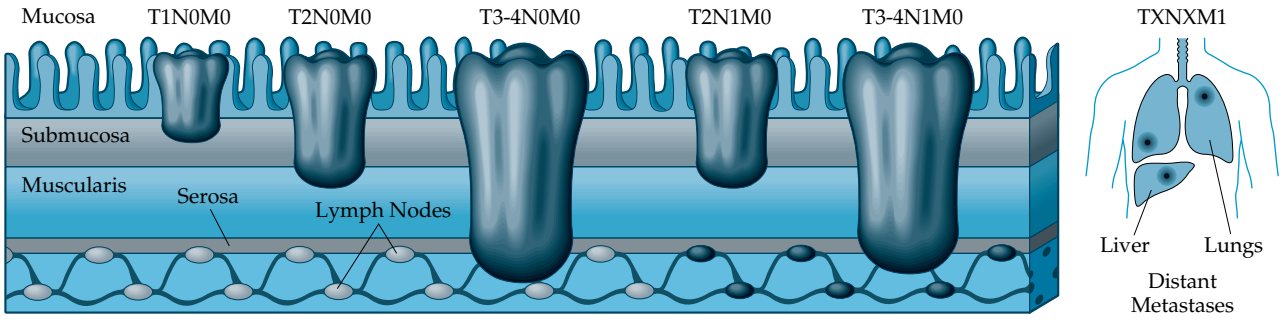
Cancers developing in the rectosigmoid are associated with tenesmus, narrowing of the stool, and hematochezia. Anemia is unusual, despite the passage of what appears to the patient to be copious quantities of bright-red blood from the rectum. Patients and physicians may attribute these symptoms to hemorrhoids. Development of altered bowel habits, rectal bleeding, or both mandates a digital rectal examination and colonoscopy.

STAGING AND PROGNOSIS

The prognosis for patients with adenocarcinoma of the colorectum is closely associated with the depth of tumor penetration into the bowel wall and the presence or absence of regional lymph node involvement and distant metastases. The most frequently used staging system, which incorporates these prognostic variables, was introduced by Dukes and later modified by Kirklin, Astler, and Coller, as well as others. More recently, the Dukes system has been applied to the TNM classification method, in which T represents the depth of tumor penetration; N, the presence or absence of lymph node involvement; and M, the presence or absence of distant metastases. In addition, the American Joint Committee on Cancer and the Union Interna-

Table 9 Screening Process for Colorectal Cancer[65]

- Target patients at risk for colorectal cancer
- Invite participation in screening program
- Identify special circumstances (e.g., symptoms, family history) early in the process so that appropriate measures can be advised
- Perform the screening test
- Use the result to identify patients who should undertake the diagnostic process (colonoscopy preferred)
- Ensure compliance with the appropriate diagnostic follow-up
- Provide adequate subsequent treatment
- Offer rescreening at appropriate intervals
- Monitor the outcomes of the program



TNM Stage	*Description*
T1N0M0	Infiltration no deeper than submucosa
T2N0M0	Infiltration of muscularis; no penetration through colonic wall; no lymph node involvement
T3-4N0M0	Extension through colonic wall; no lymph node involvement
T2N1M0	Infiltration of muscularis; no penetration through colonic wall; lymph node involvement
T3-4N1M0	Extension through colonic wall; lymph node involvement
TXNXM1	Distant metastases

Figure 2 **In the TNM (tumor, node, metastasis) staging system for colorectal cancer, a tumor is classified according to the extent of infiltration of the bowel wall and whether it has spread to lymph nodes or to distant organs such as the liver or lungs.**

tionale Contre le Cancer (AJCC/UICC) have developed a system that roughly parallels the modified Astler-Coller system but designates the stages as I through IV rather than as A through D [*see Figure 2 and Table 10*]. These staging systems divide colorectal cancer into the following categories[94]:

1. Stage A (T1N0M0). These are superficial lesions that do not penetrate the muscularis and do not involve regional lymph nodes.
2. Stage B. These are tumors that penetrate more deeply into the bowel wall without lymph node involvement; stage B is subdivided into stage B1 (T2N0M0), in which the tumor is restricted to the muscularis; stage B2 (T3NOMO); and stage B3 (T4NOMO), in which the tumor penetrates into or through the serosa.
3. Stage C. These are tumors that involve regional nodes; in the AJCC/UICC system, they are designated as stage III and are subdivided into stage IIIA (T1/2N1), stage IIIB (T3/4N1), and stage IIIC (any T, N2).
4. Stage D. These are tumors that have metastasized to liver, lung, bone, or other anatomically distant sites (TXNXM1).

In the absence of obvious evidence of metastatic disease, the stage of the disease can be accurately determined only after resection and histopathologic analysis of the specimen.

Because most recurrences after resection occur within 3 to 4 years, the cure rate is reasonably estimated by 5-year survival rates. Five-year survival is closely linked to the stage of the disease [*see Figure 3*]. For uncertain reasons, 5-year survival rates

Table 10 Staging of Colorectal Cancer

AJCC/UICC	*TNM*			*Modified Astler-Coller*	*Dukes*
Stage I	T1	N0	M0	Stage A	A
	T2	N0	M0	Stage B1	A
Stage IIA	T3	N0	M0	Stage B2	B
Stage IIB	T4	N0	M0	Stage B3	B
Stage IIIA	T1, T2	N1*	M0*	Stage C1	C
Stage IIIB	T3, T4	N1*	M0*	Stage C2, C3	C
Stage IIIC	Any T	N2†	M0†	Stage C1, C2, C3	C
Stage IV	Any T	Any N	M1	Stage D	—

*N1 = involvement of one to three lymph nodes.
†N2 = involvement of four or more lymph nodes.
AJCC/UICC—American Joint Committee on Cancer/Union Internationale Contre le Cancer TNM—tumor penetration, lymph node involvement, distant metastasis [*see Figure 2*]

have improved for patients at almost every stage during the past several decades. This seeming improvement has been reported from single institutions[95] and multi-institutional groups[96] and represents the experience of community and university-affiliated hospitals.[96] It seems unlikely that changes in surgical technique or in the etiologic factors that lead to tumor development can explain this improvement. Rather, this change may best be understood as resulting from more thorough staging procedures, which involve careful intraoperative inspection of the liver and peritoneal cavity and more meticulous dissection of the resected specimen, including identification and examination of each of the lymph nodes surrounding the tumor. Improvements in adjuvant therapy also help account for the improvement in survival.

More exacting attention to pathologic detail has revealed that prognosis after resection not only is related to regional lymph node involvement but also may be more precisely defined by the number of involved nodes. The prognosis for patients with colorectal cancer is more favorable when tumor specimens reveal involvement of one to three lymph nodes than when they reveal four or more involved nodes. The impact of degree of local penetration and the number of lymph nodes is illustrated by the 5-year survival rates for three subcategories: 59.8% for stage IIIA, 42% for stage IIIB, and 27.3% for stage IIIC. Other indicators of poor prognosis in patients who have undergone a complete resection include poorly differentiated histologic type, tumor adherence to adjacent organs, bowel perforation, colonic obstruction at the time of diagnosis, and venous invasion by the tumor. Preoperative elevation of the carcinoembryonic antigen (CEA) level (see below) suggests that the tumor will recur, regardless of the clinicopathologic stage of the resected specimen.[97,98] The prognosis is also less favorable, particularly for patients who have undergone resection of stage B2 tumors, when the DNA content of the malignant cells (i.e., ploidy) and the percentage of proliferating cells are increased[99] or when there is allelic loss of chromosome 18q.[100] In contrast to the prognosis for patients with most other solid tumors, the prognosis for patients with colorectal cancer is not influenced by the size of the primary lesion when corrected for nodal involvement and histologic differentiation.[101]

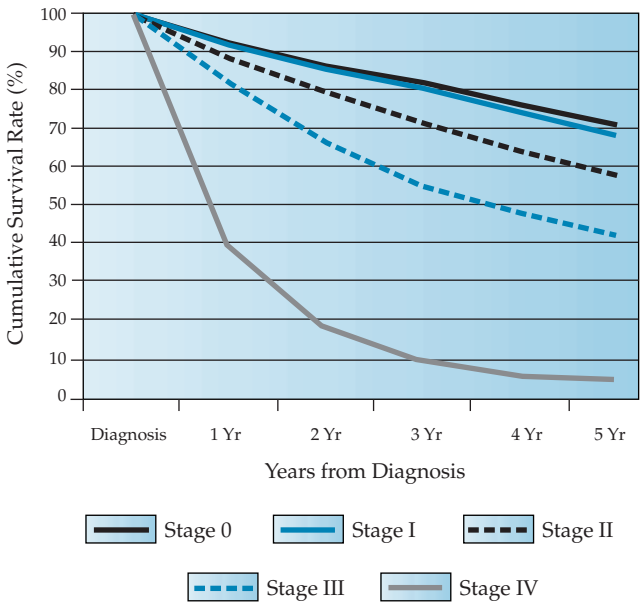


Figure 3 **Relationship between TNM stage and survival in colorectal carcinoma.[127]**

Table 11 Indicators of Poor Prognosis for Colorectal Cancer

Regional lymph node involvement
≥ 4 involved regional lymph nodes
Tumor penetration through the bowel wall
Poorly differentiated histologic findings
Tumor adherence to adjacent organs
Bowel perforation
Obstruction
Venous invasion
Preoperative elevation of carcinoembryonic antigen level to > 5.0 ng/ml
Increased DNA content (aneuploidy) of malignant cells
Allelic loss of chromosome 18q

Colorectal cancer initially spreads to regional lymph nodes and then through the portal venous circulation to the liver, which represents the most common visceral site of metastasis: the liver is the initial site of distant spread in one third of patients with recurrences, and liver involvement is seen in more than two thirds of patients by the time of death. As a rule, colorectal cancer rarely spreads to supradiaphragmatic sites, such as the lungs or supraclavicular nodes, or to less common areas, such as bone or the brain, without previous involvement of the liver. The main exception to this generalization occurs in patients whose primary tumor is in the distal rectum: tumor cells that are shed from these lesions may escape the portal venous system and spread through the paravertebral venous plexus to the lungs or supraclavicular nodes.[102] The median survival time for patients after the detection of distant metastases (stage D) is 6 to 12 months if untreated.

Carcinoembryonic Antigen

Although CEA is an imperfect tumor marker, it can provide useful information for the management of colorectal cancer patients if its limitations and attributes are understood[103] [*see Table 11*]. CEA is a glycoprotein that was initially found in association only with colorectal cancer and embryonic and fetal gut tissues. Plasma levels of CEA, which can be measured by radioimmunoassay, have subsequently been shown to be elevated in patients with many other malignant diseases (e.g., cancers of the stomach, pancreas, breast, and lung) and with various nonmalignant conditions (e.g., alcoholic liver disease, inflammatory bowel disease, heavy cigarette smoking, chronic bronchitis, and pancreatitis). Plasma CEA levels generally rise only when colorectal tumor cells have penetrated through the bowel wall (i.e., with cancer of at least stage B2). Therefore, the CEA assay is not useful as a screening test, even when applied to patients with gastrointestinal signs or symptoms. The CEA test is erratic as a quantitative measure of tumor bulk because the CEA level is highest when the liver is involved, even to only a minor degree, and may be barely elevated in patients with a bulky intra-abdominal recurrence [*see Table 12*].

Despite these shortcomings, there are several defined roles for the CEA assay:

1. Preoperatively, the CEA level is related to the stage of disease and may serve as a predictor of surgical incurability: preoperative CEA values greater than 5 ng/ml have been associated with a poor prognosis, independent of surgical stage.[97,98]
2. Postoperatively, the CEA level may serve as a measure of the completeness of tumor resection. If a preoperatively elevated CEA value does not fall to normal levels within 4 weeks (a period that is twice the plasma half-life of CEA) after surgery, the resection was probably incomplete or occult metastases are present.
3. The CEA level may serve as a useful monitor of tumor recurrence.
4. The CEA assay may serve as a monitor of response to treatment of metastatic disease. Serial CEA values parallel either tumor regression or tumor progression.[104] A rising CEA level is incompatible with tumor regression, whereas CEA values decrease in most patients who have responded to treatment.

Until better reagents become clinically available, the CEA assay will remain a source of information about tumor status that is unavailable through other means. Thus, this assay continues to merit a role in the management of selected colorectal cancer patients.

Table 12 Carcinoembryonic Antigen Level as a Marker for Colorectal Cancer

Uses	
	Preoperative prognostic indicator
	Postoperative measure of completeness of tumor resection
	Monitor of response of metastatic disease to treatment
	Most sensitive monitor of recurrence
Limitations	
	Nonspecific: elevated levels in many malignant and benign diseases
	Unsatisfactory screening test
	Erratic monitor of tumor bulk
	Not produced by poorly differentiated tumors

Treatment

SURGICAL MANAGEMENT

Curative resection offers the greatest potential for cure in patients with invasive colorectal cancer. Patients considered for such surgery are often elderly and should be evaluated preoperatively for metastatic disease by thorough physical examination, biochemical studies, and imaging of the chest and pelvis (when appropriate). These studies are particularly appropriate for patients with significant comorbid disease who might represent poor operative candidates and for patients who have symptoms suggestive of hepatic spread, such as weight loss, anorexia, and fever. The identification of metastases does not constitute an absolute contraindication to surgery in patients experiencing tumor-induced gastrointestinal bleeding or obstruction, but it frequently results in a more conservative operative procedure designed primarily to relieve symptoms. Before surgery, the CEA titer should be determined and, if possible, the entire bowel mucosa should be visualized by colonoscopy to detect synchronous polyps or neoplasms.

The most useful adjunct for the preoperative assessment of rectal lesions (in addition to a digital rectal examination) is endorectal ultrasonography. Endorectal ultrasound allows clear visualization of the layers of the rectal wall and thus enables the depth of invasion to be precisely determined. Endorectal ultrasound has an accuracy of 82% to 93% with respect to depth of invasion. Assessment of lymph node involvement is less reliable, with reported accuracy of 65% to 81%.

Surgery for colorectal cancer is based on the pattern of local disease spread and on the vascular anatomy of the bowel. Limited, or wedge, resections are inadequate; the regional lymph nodes draining a given segment of large bowel should be removed, along with associated blood vessels, and surgical margins of at least 5 cm should be obtained. During laparotomy, the surgeon should thoroughly examine the entire abdomen—including the liver, hemidiaphragms, and pelvis—and carefully palpate the full length of the large bowel.

The surgical management of cancers that arise in the distal rectum presents a particular problem because the traditional operative procedure for these lesions—abdominal-perineal resection—requires that patients receive a permanent sigmoid colostomy. Although such an operation remains unavoidable in most cases in which the cancer occurs within 5 to 6 cm of the anal verge, staple devices permit the construction of end-to-end anastomoses by experienced surgical oncologists for many patients with midrectal lesions. These anastomoses, which do not increase the risk of complications or tumor recurrence, allow the anal sphincter to be preserved. Preservation of the anal sphincter has also been achieved by the use of transanal or transcoccygeal resection in selected patients who have superficial, nonulcerated tumors that are too close to the anal verge for a stapled anastomosis.[105] The ideal margin for rectal cancer resection is 2 cm or more distally and 5 cm or more proximally. Lymphovascular resection of the rectum should include a wide anatomic resection of the mesorectum.[106]

Laparoscopic Colectomy

Laparoscopic techniques have become widely used in the management of colorectal cancer. These techniques can be safe and successful in the hands of an experienced surgeon. A slightly shortened hospital stay and more rapid postoperative recovery also are potential benefits. In a multicenter trial that excluded transverse colon and rectal cancers, rates of recurrence and of overall survival were equivalent, respectively, for patients undergoing open colectomy and for those undergoing laparoscopic colectomy.[107] Wound recurrence rates were lower than 1% in both groups. In a single-hospital study, laparoscopic resection was associated with lower cancer-related mortality, lower overall mortality, and a lower recurrence rate.[108]

Postoperative Surveillance

Colonoscopy Before a planned curative resection of a colorectal cancer, patients should undergo a preoperative colonoscopy to exclude the presence of synchronous neoplasia. If the colon is obstructed preoperatively, colonoscopy can be performed approximately 3 to 6 months after surgery. If this or a complete preoperative examination is normal, subsequent colonoscopy should be offered 1 year later; if the results of that colonoscopy are normal, colonoscopies should be performed every 3 to 5 years thereafter. The intent is to detect recurrent adenomatous polyps and new primary cancers.[28] Anastomotic recurrences, which occur in only about 2% of colon cancers, are common-

ly associated with intra-abdominal recurrence that cannot be surgically resected. Recurrent tumors may be missed, as demonstrated by a study in which the incidence of secondary colorectal cancers was increased despite intensive surveillance; the cumulative incidence of recurrence was 1.5% at 5 years in patients treated for localized colon cancer.[109]

History and physical examination There are no data that directly address the contribution of the history and physical examination to outcomes of colorectal cancer surveillance. However, a clinical history and pertinent physical examination should be performed every 3 to 6 months for the first 3 years and annually thereafter.[110]

Complete blood count Routine complete blood counts are not recommended.

Carcinoembryonic antigen If resection of liver metastases is clinically indicated and serious comorbid conditions are absent, it is recommended that postoperative serum CEA testing be performed every 3 months in patients with stage II or III disease for 2 years after diagnosis. An elevated CEA level, if confirmed by retesting, warrants further evaluation for metastatic disease but does not justify the institution of systemic therapy for presumed metastatic disease.

Liver enzymes Regular monitoring of liver enzyme levels after primary therapy for colon and rectal cancer is not recommended.

Fecal occult blood test Periodic FOBTs are not recommended.

Chest x-ray Routine yearly chest x-rays are not recommended. Chest x-rays may be ordered for patients with elevated CEA levels or symptoms suggestive of a pulmonary metastasis.

Computed tomography Routine CT scans are not recommended in the follow-up of patients with colorectal cancer.

Positron emission tomography Fluorine-18-deoxyglucose–labeled positron emission tomography (PET) is a complementary staging method that improves the management of patients with liver metastases, especially by detecting unsuspected extrahepatic disease.[111] The combination of PET and CT is now in use and can provide complementary information in the management of patients who are undergoing chemotherapy or radiation therapy.[112]

Flexible proctosigmoidoscopy Surveillance with flexible proctosigmoidoscopy is appropriate for selected patients with rectal cancer. A combination of chemotherapy and pelvic radiation therapy represents the standard treatment of rectal cancer of stage II or III (see below). For patients who have not received pelvic radiation, direct visualization of the rectum at periodic intervals is suggested. It is the custom of some surgeons to inspect the rectum in all patients who have undergone low anterior resection by performing proctosigmoidoscopy 6 and 12 months after resection.

RADIATION THERAPY

Radiation therapy plays an important role in the treatment of patients with stage B2 or stage C rectal tumors. Cancer recurs locally or regionally in 20% to 40% of these patients after complete resection.[113] This unusually high frequency of recurrence is presumably the result of two factors: the loss of integrity of the serosa of the large bowel as it enters the pelvis facilitates the infiltration of tumor, and the rich lymphatic supply of the pelvic side wall immediately adjacent to the rectum enhances the early spread of malignant cells into surgically inaccessible tissue. Therefore, adjuvant radiation therapy was introduced to remove tumor cells from perirectal tissue and to increase the chance of cure.

The rationale for using adjuvant radiation therapy to decrease pelvic recurrence appears sound, but it is uncertain whether such treatment should be administered before or after surgery. Patients with large, potentially unresectable rectal cancers may require preoperative irradiation to shrink the tumor sufficiently to allow its resection.

Prospective, randomized trials have demonstrated that neither preoperative[114] nor postoperative[115] irradiation alone significantly affects overall survival, although both forms significantly reduce the local recurrence rate. Survival is prolonged, however, when such adjuvant radiation therapy is combined with concomitant chemotherapy (see below).

Postoperative radiation therapy, when administered in doses of 4,500 to 5,000 cGy, causes transient diarrhea, cystitis, and perianal skin irritation; long-term damage to the small bowel or bladder is uncommon.[114] For patients whose tumor has spread through the bowel wall or who have lymph node metastases demonstrated by endorectal ultrasonography, preoperative chemoradiation has resulted in an increase in the percentage of sphincter-saving procedures that can be performed, with greater patient satisfaction. Postoperative lymph node status has been found to be an important predictor of recurrence and survival.[116]

CHEMOTHERAPY

After decades in which fluorouracil (5-FU) was the only chemotherapeutic agent available, the advent of new and more effective agents has completely changed the management of colorectal cancer. Although 5-FU remains the backbone of most regimens, new chemotherapeutic agents (i.e., irinotecan, oxaliplatin, and capecitabine) and new biologic agents (i.e., the monoclonal antibodies cetuximab and bevacizumab) offer additional benefit. Multiagent regimens that include oxaliplatin or irinotecan in combination with these monoclonal antibodies have entered widespread use.

Fluorouracil

Synthesized in 1952, 5-FU remains an important drug in the treatment of advanced colon cancer. 5-FU may be administered as a bolus injection either weekly or daily for 5 days every 4 weeks. Partial response rates with these regimens have been approximately 10% to 15%. The development of permanent venous access devices and portable infusion pumps now permits the continuous infusion of 5-FU on an outpatient basis. The use of continuous infusion enhances the likelihood that 5-FU will be present during the S phase of the tumor cell cycle, when this agent is most effective.

5-FU is modulated by leucovorin, which raises the level of 5,10-methylenetetrahydrofolate and results in the formation of a stable ternary complex of the folate coenzyme thymidylate synthase with 5-FU in the form of its principal metabolite, fluorodeoxyuridine. The use of 5-FU with leucovorin results in a higher response rate than with 5-FU alone.

Capecitabine

Capecitabine is a fluoropyrimidine—specifically, a prodrug of 5-FU—that mimics continuous-infusion 5-FU. It has the advantage of oral administration. In two studies that compared capecitabine with 5-FU in patients with advanced disease, capecitabine therapy was associated with an improved response rate (26% versus 17%) and with less toxicity, but there was no significant benefit in survival.[117] Diarrhea and microcutaneous toxicity (e.g., hand-foot syndrome) are common side effects.

Irinotecan

Irinotecan is a novel topoisomerase inhibitor that has significant therapeutic activity in metastatic colorectal cancer. It is used when the tumor has recurred or spread after standard chemotherapy. The primary toxicities are diarrhea and neutropenia, which can be life threatening if they are not treated promptly and aggressively.

Oxaliplatin

Oxaliplatin differs in its preclinical activity profile from cisplatin and also has a distinct toxicity profile. It causes no renal toxicity and minimal hematologic toxicity, but it is associated with both a reversible, acute, cold-related dysesthesia and a dose-limiting, cumulative, peripheral sensory neuropathy.

Cetuximab

Cetuximab is a chimeric monoclonal antibody that bonds to the external growth factor receptor, which is commonly expressed on colorectal carcinoma cells. Dysregulation of cell signaling through this receptor is thought to be an important factor in the growth of many epithelial malignancies. Patients treated with combinations of cetuximab and irinotecan have shown a higher response rate and a longer time to worsening of the metastatic disease than patients treated with either drug alone.[118] Side effects of cetuximab include an acneiform rash, malaise, and diarrhea.

Bevacizumab

Bevacizumab is a human chimeric antibody to vascular endothelial growth factor that stimulates angiogenesis. New blood vessel formation is important in tumor growth and invasion. The addition of bevacizumab to combination therapy with irinotecan, leucovorin, and 5-FU has been shown to increase therapeutic efficacy: patients receiving the four-drug combination survived for approximately 20 months—about 5 months longer than those receiving the three-drug combination.[119] Combinations of bevacizumab and cetuximab are also being assessed because of preclinical and early clinical data indicating synergistic activity.

Adjuvant Therapy

The use of 5-FU plus leucovorin is modestly effective in reducing tumor recurrence, with disease-free survival of about 60% and overall survival of about 65%. Capecitabine has been shown to be as effective as 5-FU plus leucovorin, with less associated toxicity and greater patient convenience. In patients with stage III colon cancer, the addition of oxaliplatin to 5-FU plus leucovorin has prolonged the 3-year disease-free survival from 73% to 77%.[120]

Management of Advanced Colorectal Cancer

Numerous combinations of chemotherapeutic drugs plus monoclonal antibodies are in clinical trials. The best choice for initial therapy remains to be determined. It is encouraging that median survival of patients with advanced disease has increased from about 13 months with 5-FU plus folinic acid, usually administered intravenously, to about 22 months with infusional 5-FU, leucovorin, and oxaliplatin. Outside of clinical trials, oncologists will need to consider not only the various toxicities caused by these new agents and the risk-to-benefit ratio but also the financial costs for each patient.

Other Molecular Targets

Molecular targets for therapy are becoming more widely known and include epidermal growth factors, tyrosine kinases, vascular endothelial growth factor, and intracellular signaling pathways. Many trials of compounds targeting these pathways in tumor cells and the surrounding stroma are in progress.

CHEMORADIOTHERAPY FOR RECTAL CANCER

Concomitant radiotherapy and chemotherapy decrease the probability of local or regional recurrences after complete resection of stage B2 or stage C rectal cancer and prolong survival.[121-123] In this setting, the chemotherapy is thought to make the radiation therapy more effective (a phenomenon known as radiosensitization). This approach produced its most successful outcome when a continuous intravenous infusion of 5-FU was administered through a portable pump for the entire 4- to 5-week period of radiation therapy.[123] Capecitabine is now being tested in combination with postoperative radiation therapy.

PALLIATION

Laser photoablation or stenting of obstructing rectosigmoid or rectal cancers should be considered if surgical decompression is not possible or advisable because of extensive metastatic disease or comorbidity.

The author has received grants for clinical research from Pfizer/Pharmacia.

References

1. Jemal A, Murray T, Ward E, et al: Cancer Statistics, 2005. CA Cancer J Clin 55:10, 2005

2. Howe HL, Wingo PA, Thun MJ, et al: Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J Natl Cancer Inst 93:824, 2001

3. Jass JR, Whitehall VLJ, Young J, et al: Emerging concepts in colorectal neoplasia. Gastroenterology 123:862, 2002

4. Fenoglio-Preiser CM, Hutter RVP: Colorectal polyps: pathologic diagnosis and clinical significance. CA Cancer J Clin 35:322, 1985

5. O'Brien MJ, Winawer SJ, Zauber AG, et al: The National Polyp Study: patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 98:371, 1990

6. Coutsoftides T, Lavery I, Benjamin SP, et al: Malignant polyps of the colon and rectum: a clinicopathologic study. Dis Colon Rectum 22:82, 1979

7. Shinya H, Wolff WI: Morphology, anatomic distribution and cancer potential of colonic polyps. Ann Surg 190:679, 1979

8. Kune GA, Kune S, Watson LF: History of colorectal polypectomy and risk of subsequent colorectal cancer. Br J Surg 74:1064, 1987

9. Simons BD, Morrison AS, Lev R, et al: Relationship of polyps to cancer of the large intestine. J Natl Cancer Inst 84:962, 1992

10. Cannon-Albright LA, Skolnick MH, Bishop T, et al: Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. N Engl J Med 319:533, 1988

11. Rickert RR, Auerbach O, Garfinkel L, et al: Adenomatous lesions of the large bowel: an autopsy survey. Cancer 43:1847, 1979

12. Blackstone MO: Colorectal adenomas on follow-up colonoscopy: is the cancer risk unchanged? Lancet 345:1126, 1995

13. Atkin WS, Morson BC, Cuzick J: Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. N Engl J Med 326:658, 1992

14. Neugut AI, Jacobson JS, Ahsan H, et al: Incidence and recurrence rates of colorectal adenomas: a prospective study. Gastroenterology 108:402, 1995

15. Jass JR, Whitehall VL, Young J, et al: Emerging concepts in colorectal neoplasia. Gastroenterology 123:862, 2002

16. Rustgi AK: Hereditary gastrointestinal polyposis and nonpolyposis syndromes. N Engl J Med 331:1694, 1994

17. Cottrell S, Bicknell D, Kaklamanis L, et al: Molecular analysis of *APC* mutations in familial adenomatous polyposis and sporadic colon carcinomas. Lancet 340:626, 1992

18. Lipkin M, Blattner WA, Gardner EJ, et al: Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [^{3}H]thymidine labeling patterns in colonic epithelial cells. Cancer Res 44:4201, 1984

19. Soravia C, Berk T, Maklensky L, et al: Genotype phenotype correlations in attenuated adenomatous polyposis coli. Am J Hum Genet 62:1290, 1998

20. Giardiello FM, Brensinger JD, Petersen GM, et al: The use and interpretation of commercial *APC* gene testing for familial adenomatous polyposis. N Engl J Med 336:823, 1997

21. Colorectal cancer screening. National Comprehensive Cancer Network. J Natl Compr Canc Netw 1:72, 2003

22. Sarre RG, Frost AG, Jagelman DG, et al: Gastric and duodenal polyps in familial adenomatous polyposis: a prospective study of the nature and prevalence of upper gastrointestinal polyps. Gut 28:306, 1987

23. Steinbach G, Lynch PM, Phillips RK, et al: The effect of celecoxib, a cyclo-oxygenase inhibitor, in familial adenomatous polyposis. N Engl J Med 342:1946, 2000

24. Vasen HFA, Mecklin JP, Khan PM, et al: The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 34:424, 1991

25. Vasen HF, Watson P, Mecklin JP, et al: New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. Gastroenterology 116:1453, 1999

26. Rodriguez-Bigas MA, Boland CR, Hamilton SR: A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 89:1758, 1997

27. Halford S, Sasieni P, Rowan A, et al: Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait. Cancer Res 62:53, 2002

28. Syngal S, Fox E, Li C: Interpretation of genetic test results for hereditary non-polyposis colorectal cancer: implications for clinical predisposition testing. JAMA 282:253, 1999

29. Winawer S, Fletcher RW, Rex D, et al: Colorectal cancer screening and surveillance: clinical guidelines and rationale—update based on new evidence. Gastroenterology 124:544, 2003

30. Colon and Rectum. Food, Nutrition and the Prevention of Cancer: A Global Perspective. World Cancer Research Fund/American Institute for Cancer Research, Washington, DC, 1997, p 216

31. Calle EE, Rodriguez C, Walker-Thurmond K, et al: Overweight, obesity and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 348:1625, 2003

32. Howe GR, Aaronson KJ, Benito E, et al: The relationship between dietary fat intake and risk of colorectal cancer: evidence from the combined analysis of 13 case control studies. Cancer Causes Control 8:215, 1997

33. Atillasoy E, Fein B, Weinstein IB, et al: Fecal diacylglycerol concentrations and calcium supplementation. Cancer Epidemiol Biomarkers Prev 4:795, 1997

34. Slattery ML, Schemacker MC, Smith KR, et al: Physical activity, diet and risk of colon cancer in Utah. Am J Epidemiol 128:989, 1988

35. Neugut AI, Jacobson JS, De Vivo I: Epidemiology of colorectal adenomatous polyps. Cancer Epidemiol Biomarkers Prev 2:159, 1993

36. Willett SC, Stampler MJ, Colditz GA, et al: Relation of meat fat and fiber intake to the risk of colon cancer in a prospective study among women. N Engl J Med 323:1664, 1990

37. Thun MJ, Calle EE, Namboodiri MM, et al: Risk factors for fatal colon cancer in a large prospective study. J Natl Cancer Inst 84:1491, 1992

38. Bostick RM, Potter JD, Sellers TA, et al: Relation of calcium, vitamin D and dairy food intake to incidence of colon cancer among older women: the Iowa Women's Health Study. Am J Epidemiol 137:1302, 1993

39. Steinmetz KA, Potter JD: Vegetables, fruit and cancer. II: mechanisms. Cancer Causes Control 2:427, 1991

40. Howe GR, Benito E, Castelleto R, et al: Dietary intake of fiber and decreased risk of cancers of the colon and rectum: evidence from the combined analysis of 13 case-control studies. J Natl Cancer Inst 84:1887, 1992

41. Fuchs CS, Giovannucci EL, Colditz GA, et al: Dietary fiber and the risk of colorectal cancer and adenoma in women. N Engl J Med 340:169, 1999

42. Schatzkin A, Lanza E, Corle D, et al: Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. N Engl J Med 342:1149, 2000

43. Alberts DS, Martinez ME, Roe DJ, et al: Lack of effect of a high-fiber cereal supplement on the recurrence of colorectal adenomas. N Engl J Med 342:1146, 2000

44. Terry P, Giovannucci E, Michels KB, et al: Fruit, vegetables, dietary fiber, and risk of colorectal cancer. J Natl Cancer Inst 93:525, 2001

45. Martinez ME, Willett WC: Calcium, vitamin D and colorectal cancer: a review of the epidemiologic evidence. Cancer Epidemiol Biomarkers Prev 7:163, 1998

46. Calcium supplements for the prevention of colorectal adenomas. Calcium Polyp Prevention Study Group. N Engl J Med 340:101, 1999

47. Choi SW, Mason JB: Folate and carcinogenesis: an integrated scheme. J Nutr 130:129, 2000

48. Giovannucci E, Stampfer MJ, Colditz GA, et al: Folate, methionine and alcohol intake and risk of colorectal adenoma. J Natl Cancer Inst 85:875, 1993

49. Giovannucci E, Stampfer MJ, Colditz GA, et al: Multivitamin use, folate and colon cancer in women in the Nurses' Health Study. Ann Intern Med 129:517, 1998

50. Nanda K, Bastian L, Hasselblad V, et al: Hormone replacement therapy and the risk of colorectal cancer: a meta-analysis. Obstet Gynecol 93:880, 1999

51. Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, et al: Estrogen plus progestin and colorectal cancer in postmenopausal women. Women's Health Initiative Investigators. N Engl J Med 350:991, 2004

52. Thurn MJ, Namboodiri MM, Heath CW, et al: Aspirin use and reduced risk of fatal colon cancer. N Engl J Med 325:1593, 1991

53. Giovannucci E, Rimm EB, Stampfer MJ, et al: Aspirin use and the risk for colorectal cancer and adenoma in male health professionals. Ann Intern Med 121:241, 1994

54. Gann PH, Manson E, Glynn RJ, et al: Low-dose aspirin and incidence of colorectal tumors in a randomized trial. J Natl Cancer Inst 85:1220, 1993

55. Sandler RS, Halabi S, Baron JA, et al: A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. N Engl J Med 348:883, 2003

56. Baron JA, Cole BF, Sandler RS, et al: A randomized trial of aspirin to prevent colorectal adenoma. N Engl J Med 348:891, 2003

57. Bresalier RS, Sandler RS, Quan H, et al: Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. N Engl J Med 352:1092, 2005

58. Solomon SD, McMurry JJ, Pfeffer MA, et al: Cardiovascular risk associated with celecoxib in a clinical trial for colorectal adenoma prevention. N Engl J Med 352:1071, 2005

59. Giardiello FM, Hamilton SR, Koush AJ, et al: Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. N Engl J Med 328:1313, 1993

60. White E, Jacobs EJ, Daling JR: Physical activity in relation to colon cancer in middle-aged men and women. Am J Epidemiol 144:42, 1996

61. Terry MB, Neugut AL, Bostick R, et al: Risk factors for advanced colorectal adenomas: a pooled analysis. Cancer Epidemiol Biomarkers Prev 11:622, 2002

62. Giovannucci E: An updated review of the epidemiologic evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 10:225, 2001

63. Chao A, Thun MJ, Jacobs EJ, et al: Cigarette smoking and colorectal cancer mortality in the Cancer Prevention Study II. J Natl Cancer Inst 92:1888, 2000

64. Giovannucci E, Colditz GA, Stampfer MJ, et al: A prospective study of cigarette smoking and risk of colorectal cancer in U.S. women. J Natl Cancer Inst 86:192, 1994

65. Young GP, Rozen P, Levin B: How should we screen for early colorectal neoplasia? Colorectal Cancer in Clinical Practice: Prevention, Early Detection and Management. Rozen P, Young GP, Levin B, et al, Eds. Martin Dunitz, Ltd, London, 2002, p 77

66. Trends in screening for colorectal cancer—United States, 1997 and 1999. MMWR Morb Mortal Wkly Rep 50:162, 2001

67. Towler BP, Irwig L, Glasziou P, et al: Screening for colorectal cancer using the faecal occult blood test, Hemoccult. Cochrane Database Syst Rev (2):CD001216, 2002

68. Hardcastle JD, Chamberlain JD, Robinson NHI, et al: Randomized controlled trial of faecal occult blood screening for colorectal cancer. Lancet 348:1472, 1996

69. Levin B, Brooks D, Smith RA, et al: Emerging technologies in screening for colorectal cancer: CT colonography, immunochemical fecal occult blood tests and stool screening using molecular markers. CA Cancer J Clin 53:44, 2003

70. Selby JV, Fredman GD, Quesenbery CP Jr, et al: A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 326:653, 1992

71. Kavanagh AM, Giovannucci E, Fuchs CS, et al: Screening endoscopy and risk of colorectal cancer in United States men. Cancer Causes Control 9:455, 1998

72. Thiis-Evenson E, Hoff GS, Sauar J, et al: Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer. Telemark Polyp Study I. Scand J Gastroenterol 34:414, 1999

73. Prorok PC, Andriole GL, Bresalier RS, et al: Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Controlled Clin Trials 21:2735, 2000

74. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a U.K. multicenter randomized trial. U.K. Flexible Sigmoidoscopy Screening Trial Investigations. Lancet 359:1291, 2002

75. Imperiale TF, Wagner DR, Lin CY, et al: Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. N Engl J Med 334:169, 2000

76. Lieberman DA, Weiss DG, Bond JH, et al: Use of colonoscopy to screen asymptomatic adults for colorectal cancer. VA Cooperative Study Group 380. N Engl J Med 343:162, 2000

77. One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon. Veterans Affairs Cooperative Study Group 380. N Engl J Med 345:555, 2001

78. Muller AD, Sonnenberg A: Prevention of colorectal cancer by flexible endoscopy and polypectomy. Ann Intern Med 23:904, 1995

79. Citarda F, Tomaselli G, Capocaccia R, et al: Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. Gut 48:812, 2001

80. Orgensen OD, Kronborg O, Fenger C: The Funen adenoma follow-up study: incidence and death from colorectal carcinoma in an adenoma surveillance program. Scand J Gastroenterol 28:869, 1993

81. Cash B, Schoenfeld PS, Dobhan R, et al: Colorectal neoplasia screening with colonoscopy in asymptomatic women at regional naval medical centers. Gastroenterology 120:A509, 2001

82. Winawer SJ, Zauber AG, Church TR, et al: National Colonoscopy Study (NCS) preliminary results: randomized controlled trial of general population screening colonoscopy. Gastroenterology 122:T1560, 2002

83. A comparison of colonoscopy and double-contrast barium enema for surveillance after polypectomy. National Polyp Study Work Group. N Engl J Med 342:1766, 2000

84. Sidransky D, Tokino T, Hamilton SR, et al: Identification of *ras* oncogene mutations in the stool of patients with curable colorectal tumors. Science 256:102, 1992

85. Ahlquist DA, Shuber AP: Stool screening for colorectal cancer: evolution from occult blood to molecular markers. Clin Chim Acta 315:157, 2002

86. Traverso G, Shuber A, Levin B, et al: Detection of *APC* mutations in fecal DNA of patients with colorectal tumors. N Engl J Med 346:311, 2002

87. Traverso G, Shuber A, Olsson L, et al: Detection of proximal colorectal cancer through the analysis of faecal DNA. Lancet 359:403, 2002

88. Imperiale TE, Ransohoff DF, Itzkowitz SH, et al: Fecal DNA versus fecal occult blood for colorectal cancer screening in average risk population. N Engl J Med 351:2704, 2004

89. Pickhardt PJ, Choi JR, Hwang I, et al: Computed tomographic colonography (virtual colonoscopy): a multicenter comparison with standard colonoscopy for detection of colorectal neoplasia. N Engl J Med 349:2191, 2003

90. Sonnenberg A, Delco F, Inadomi JM, et al: Cost-effectiveness of colonoscopy in screening for colorectal cancer. Ann Intern Med 133:573, 2000

91. Smith RA, Cokkinides V, Von Eschenbach AC, et al: American Cancer Society guidelines for the early detection of cancer. CA Cancer J Clin 52:8, 2002

92. Screening for colorectal cancer: recommendation and rationale. U.S. Preventive Services Task Force. Ann Intern Med 137:129, 2002

93. Spann S, Rozen P, Levin B, et al: The Pros and Cons of Population-Based Colorectal Cancer Preventative Strategies. Colorectal Cancer in Clinical Practice: Prevention, Early Detection and Management. Rozen P, Young G, Levin B, et al, Eds. Martin Dunitz, Ltd, London, 2002, p 115

94. Cohen AM, Minsky BD, Schilsky RL: Colon cancer. Cancer: Principles & Practice of Oncology, 4th ed. DeVita VT Jr, Hellman S, Rosenberg SA, Eds. JB Lippincott Co, Philadelphia, 1993, p 929

95. Adjuvant therapy of colon cancer: results of a prospectively randomized trial. The Gastrointestinal Tumor Study Group. N Engl J Med 310:737, 1984

96. Mandel JS, Bond JH, Church TR, et al: Reducing mortality from colorectal cancer by screening for fecal occult blood. N Engl J Med 328:1365, 1993

97. Wanebo HJ, Rao B, Pinsky CM, et al: Preoperative carcinoembryonic antigen level as a prognostic indicator in colorectal cancer. N Engl J Med 229:448, 1978

98. Wolmark N, Fisher B, Wieand HS, et al: The prognostic significance of preoperative carcinoembryonic antigen levels in colorectal cancer: results from NSABP clinical trials. Ann Surg 199:375, 1984

99. Witzig TE, Loprinzi CL, Gonchoroff NJ, et al: DNA ploidy and cell kinetic measurements as predictors of recurrence and survival in stages B2 and C colorectal adenocarcinoma. Cancer 68:879, 1991

100. Jen J, Kim H, Piantadosi S, et al: Allelic loss of chromosome 18q and prognosis in colorectal cancer. N Engl J Med 331:213, 1994

101. Wolmark N, Cruz I, Redmond CK, et al: Tumor size and regional lymph node metastasis in colorectal cancer: a preliminary analysis from the NSABP clinical trials. Cancer 51:1315, 1983

102. Batson OV: The role of vertebral veins in metastatic processes. Ann Intern Med 16:38, 1942

103. Posner MR, Mayer RJ: The use of serologic tumor markers in gastrointestinal malignancies. Hematol Oncol Clin North Am 8:533, 1994

104. Mayer RJ, Garnick MB, Steele GD, et al: Carcinoembryonic antigen (CEA) as a monitor of chemotherapy in disseminated colorectal cancer. Cancer 42:1428, 1978

105. Steele G Jr, Busse P, Huberman MS, et al: A pilot study of sphincter sparing management of adenocarcinoma of the rectum. Arch Surg 126:696, 1991

106. Nelson H, Petrelli N, Carlin A, et al: Guidelines 2000 for colon and rectal cancer surgery. J Natl Cancer Inst 93:583, 2001

107. A comparison of laparoscopically assisted and open colectomy for colon cancer. Clinical Outcomes of Surgical Therapy Study Group. N Engl J Med 350:2050, 2004

108. Lacy AM, Garcia-Valdecaas JC, Delgado S, et al: Laparoscopically-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomized trial. Lancet 359:2224, 2002

109. Green RJ, Metlay JP, Propert K, et al: Surveillance for second primary colorectal cancer after adjuvant chemotherapy: an analysis of Intergroup 0089. Ann Intern Med 136:261, 2002

110. Desch CE, Benson AB, Smith TJ: Recommended colorectal cancer surveillance guidelines by the American Society of Clinical Oncology. J Clin Oncol 17:1312, 1999

111. Ruers TJ, Langenhoff BS, Neelemen N, et al: Value of positron emission tomography with [F-18]fluorodeoxyglucose in patients with colorectal liver metastases: a prospective study. J Clin Oncol 20:388, 2002

112. Kamel IR, Cohade C, Neyman E, et al: Incremental value of CT in PET/CT of patients with colorectal carcinoma. Abdom Imaging 29:663, 2004

113. Mohiuddin M, Marks G: Adjuvant radiation therapy for colon and rectal cancer. Semin Oncol 18:411, 1991

114. Gerard A, Buyse M, Nordlinger B, et al: Preoperative radiotherapy as adjuvant treatment in rectal cancer: final results of a randomized study of the European Organization for Research and Treatment of Cancer (EORTC). Ann Surg 208:606, 1988

115. Bentzen SM, Balslev I, Pederson M, et al: A regression analysis of prognostic factors after resection of Dukes' B and C carcinoma of the rectum and rectosigmoid: does postoperative radiotherapy change the prognosis? Br J Cancer 58:195, 1988

116. Onaitis NW, Noone RB, Hartwig M, et al: Neoadjuvant chemoradiation for rectal cancer: analysis of clinical outcomes from a 13-year-institutional experience. Ann Surg 233:778, 2001

117. Capecitabine as first-line treatment in colorectal cancer: pooled data from two large, phase III trials. Xeloda Colorectal Cancer Group. Eur J Cancer 38(suppl 2):15, 2002

118. Cunningham D, Humblet Y, Siena S, et al: Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351:337, 2004

119. Hurwitz H, Fehrenbacher L, Novotny W, et al: Bevacizumab plus irinotecan, fluorouracil and leucovorin for metastatic colorectal cancer. N Engl J Med 350:2335, 2004

120. Wolmark N, Wieand S, Kuebler JP, et al: A phase III trial comparing FULV to FULV + oxaliplatin in stage II or III carcinoma of the colon: results of NSABP protocol C-07. J Clin Oncol 23:LBA3500, 2005

121. Prolongation of the disease-free interval in surgically treated rectal carcinoma. Gastrointestinal Tumor Study Group. N Engl J Med 312:1465, 1985

122. Krook JE, Moertel CG, Gunderson LL, et al: Effective surgical adjuvant therapy for high-risk rectal carcinoma. N Engl J Med 324:709, 1991

123. O'Connell MJ, Martenson JA, Wieand HS, et al: Improving adjuvant therapy for rectal cancer by combining protracted-infusion fluorouracil with radiation therapy after curative surgery. N Engl J Med 331:502, 1994

124. O'Shaughnessy JA, Kelloff GJ, Gordon GB, et al: Treatment and prevention of intra-epithelial neoplasia: an important target for accelerated new agent development. Clin Cancer Res 8:314, 2002

125. Selby JV: Colorectal cancer screening: keeping the options open. Eff Clin Pract 4:39, 2001

126. American Cancer Society guidelines for the early detection of cancer, 2003. American Cancer Society. CA Cancer J Clin 53:27, 2003

127. Compton CC, Greene FL: The staging of colorectal cancer: 2004 and beyond. CA Cancer J Clin 54:295, 2004

VI PANCREATIC, GASTRIC, AND OTHER GASTROINTESTINAL CANCERS

Weijing Sun, m.d.
Daniel Haller, m.d.

On the basis of 2003 estimates from the American Cancer Society, cancers originating in the gastrointestinal tract rank second in both incidence and cancer-related deaths. One in four deaths in the United States is caused by cancer, and 25% of cancer-related deaths are caused by gastrointestinal malignancies; more than 40% of these deaths are caused by cancer of the pancreas, stomach, and esophagus [*see Table 1*].[1] In the past 10 years, advances in molecular biology, medical genetics, imaging and endoscopic techniques, and the development of antitumor agents have significantly altered approaches to the prevention, diagnosis, and treatment of GI cancers.

Esophageal Cancer

The esophagus extends from the cricopharyngeal sphincter to the gastroesophageal (GE) junction and is commonly divided into three portions: the cervical, upper to midthoracic, and thoracic. Histology and optimal treatment approaches may vary considerably because of the locations of the cancers.

EPIDEMIOLOGY

Esophageal cancer, which includes squamous cell carcinoma (SCC) and adenocarcinoma, is the ninth most common cancer worldwide. In the 1970s, approximately 90% of esophageal cancers were SCCs. The incidence of adenocarcinoma has increased dramatically and currently accounts for about 50% of new cases of esophageal cancer, a rate of acceleration greater than that of any other cancer in the United States.[2] The reason for the rapid increase in the incidence of esophageal adenocarcinoma is unknown. Esophageal cancer occurs about two to four times more frequently in men than in women. It is curable if the diagnosis is made in its early stages, but the disease usually presents in an advanced stage.

There is great geographical variation in the incidence of esophageal SCC. High-incidence regions, where rates may exceed 100 per 100,000 population, include northern China, India, parts of Iran, areas north and east of the Caspian Sea, and the Transkei area of South Africa.[3] Recent data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) indicate that the annual rate of SCC of the esophagus per 100,000 population declined from 3.4 during the period 1974 through 1976 to 2.2 during the period 1992 through 1994[4]; however, the disease is more common in black males (16.8 per 100,000 population) than white males (1.0 per 100,000 population).[2] The rate of SCC in black males continued to decline into the 1990s.

The incidence of esophageal adenocarcinoma is rapidly increasing in the United States and other countries.[5] It has been documented in population-based cohort studies that the incidence has increased 300% to 500% in the past 40 years.[6] Approximately 50% of new cases of esophageal cancer in the United States in 2002 were adenocarcinomas. Distal esophageal adenocarcinomas occur eight times more frequently in white males than in white females and are five times more common in white men than in black men, making white males the group at highest risk for the disease. The rate of esophageal adenocarcinoma in black women is too low to yield a reliable estimate of trends.[4]

ETIOLOGY AND RISK FACTORS

Tobacco Smoking and Alcohol Consumption

The major causes of SCC of the esophagus in the United States and other Western countries are cigarette smoking and alcohol consumption. Both have been found to be independent risk factors in men and women, and their combined use substantially increases risk.[7] It has been observed that the dose-response relationships for alcohol and tobacco are different. For tobacco smoking, the risk depends mainly on the duration of tobacco consumption (odds ratio is 6.4 for more than 35 years of smok-

Table 1 Estimated New Gastrointestinal Cancers and Deaths in the United States in 2003[1]

Site	New Cases			Deaths		
	Male	Female	All	Male	Female	All
Colon	49,000	56,500	105,500	28,300	28,800	57,100
Rectum	23,800	18,200	42,000			
Esophagus	10,600	3,300	13,900	9,900	3,100	13,000
Stomach	13,400	9,000	22,400	7,000	5,100	12,100
Pancreas	14,900	15,800	30,700	14,700	15,300	30,000
Hepatobiliary tract	14,800	9,300	24,100	10,500	7,400	17,900
Small intestine	2,700	2,600	5,300	600	500	1,100
Anus	1,700	2,300	4,000	200	300	500
Other sites	1,400	3,100	4,500	700	1,200	1,900
Total	132,300	120,100	252,400	71,900	61,700	133,600

ing); however, number of cigarettes a day also increases risk.[8] The risk associated with smoking black tobacco is two times higher than that with smoking blond or mixed tobacco. For alcohol consumption, the risk of esophageal SCC is based primarily on mean daily intake, not on duration of the drinking habit. The linear relationship between overall daily ethanol consumption and esophageal cancer is more marked for strong drinkers than for lighter drinkers. The risk of esophageal cancer also depends on the type of alcoholic beverage; spirits and beer carry the highest risk. A statistically significant synergistic interaction has been found between the type and amount of smoking and heavy alcohol drinking.[8] Those with simultaneous heavy black tobacco smoking and heavy alcohol intake represent the subgroup with the highest risk for SCC of the esophagus (odds ratio, > 100).

Cessation of smoking decreases the risk of esophageal SCC, and a 5-year cessation reduces the risk by 50%. In contrast, only after a 10-year abstention from alcohol consumption does the risk of esophageal cancer decline to the upper risk level of abstainers.[9] After 10 or more years' cessation of both habits, the risk of development of esophageal cancer is decreased to one tenth that of current smokers and drinkers.

It has been suggested that tobacco smoking has a strong role both in the initiation and ongoing development of esophageal SCC, whereas alcohol consumption appears more important as a carcinogenic factor in the later stages of development.[10]

Dietary Factors

Case-control studies indicate that dietary factors may play an important role in the development of esophageal SCC, and vitamin and trace mineral deficiencies are crucial as risk factors for esophageal SCC.[11] In particular, important vitamins and minerals include A, C, folic acid, E, B_{12}, and riboflavin.

Foods that lower risk Diets containing high amounts of green and yellow vegetables that are rich in β-carotene decrease the incidence of esophageal SCC. Citrus fruits that are high in vitamin C may also have a protective effect. Lower plasma levels of folic acid have been documented in patients with esophageal SCC. In the Linxian area of China, where esophageal cancer is endemic, riboflavin deficiency, manifested by cheilosis and glossitis, is common.[12] It appears that vitamins A, C, and E exert antioxidant effects; the last two vitamins are known to influence the formation of nitrosamines.

Foods that increase risk An inverse association has been shown between the levels of certain trace elements (e.g., selenium and zinc) and mortality associated with esophageal cancer in high-incidence areas in the world. Achalasia has also been proposed as a predisposing risk factor, with a prevalence of about 5% in individuals with chronic achalasia.[13]

Predisposing Medical Conditions

Risk factors for SCC Not surprisingly, patients with head and neck SCC are at increased risk for the development of esophageal cancer, and this risk is increased by alcohol and tobacco consumption.[14] Either coexisting or sequential esophageal cancers may occur at an annual rate of 3% to 7%. Approximately 10% of second primary cancers in patients with primary lung and oropharyngeal carcinomas arise in the esophagus.[15] As a result, endoscopic surveillance of the esophagus has been advocated in this setting. Factors that increase the risk of esophageal SCC include lye ingestion, ionizing radiation, celiac sprue, human papillomavirus infection, Plummer-Vinson syndrome, esophageal diverticula, and radiation treatment.

Risk factors for adenocarcinoma Gastroesophageal reflux disease (GERD) is a risk factor for esophageal adenocarcinoma. Barrett esophagus, a metaplastic change of the lining of the esophagus in which the normal squamous cell epithelium is replaced by columnar intestinal-type epithelium, is a complication of chronic reflux disease, and it is associated with an increased risk of adenocarcinoma of the esophagus. Whether Barrett esophagus is a necessary precursor to all cases of esophageal adenocarcinoma is unknown. There are no prospective cohort studies of patients with reflux to assess cancer risk; however, in population-based, case-control studies examining the relationship between reflux and adenocarcinoma of the esophagus, patients with esophageal adenocarcinoma were eight times more likely to have at least weekly symptoms of reflux or regurgitation than were control subjects.[16] An individual with long-term (> 20 years) and severe reflux symptoms has an increased risk for esophageal adenocarcinoma, with an odds ratio of 43.3. On the other hand, 40% of those who develop cancer do not have frequent symptoms. Although less than 1% of the general population has Barrett esophagus, 5% to 15% of those with long-term reflux symptoms will have Barrett esophagus.[17] High-grade dysplasia in Barrett mucosa is a marker indicating high probability of invasive adenocarcinoma[18]; therefore, Barrett esophagus is a likely precursor lesion that can evolve into esophageal adenocarcinoma. Individuals with Barrett esophagus have a risk of developing esophageal adenocarcinoma that is 40 to 125 times higher than the risk of the general population.[10] The annual risk of an individual with Barrett esophagus developing esophageal adenocarcinoma is approximately one in 200[19,20]; patients with high-grade dysplasia demonstrate a risk of subsequent adenocarcinoma exceeding 25%.[21] Evidence suggests that increasing body mass index is also associated with the risk of adenocarcinoma of the esophagus.[22]

SCREENING AND PREVENTION

Squamous Cell Carcinoma

High-incidence populations, such as those found in northern China, may merit endoscopic screening in the general population. The efficacy of a screening program has been documented in China, where 5-year survival is nearly 90% after operative resection of tumors detected through screening.[23]

Mass screening in the United States is not feasible for esophageal SCC, given the low incidence.[24] However, institution of screening and surveillance endoscopic programs may be applied to certain clinical scenarios, such as in patients with tylosis, achalasia, and a history of other upper aerodigestive tract cancers. The decreasing incidence of SCC of the esophagus in the United States appears secondary to improvement of socioeconomic status and behavioral and dietary changes, including cessation of smoking and a decrease in alcohol intake in high-risk populations.

Adenocarcinoma

Because GERD is a risk factor for esophageal adenocarcinoma and because Barrett esophagus is highly associated with the disease, there is increased clinical interest in pharmacologic, surgical, or endoscopic therapy to decrease the risk, as well as prevent the development, of adenocarcinoma of the esophagus. There is no evidence to suggest that proton pump inhibitors either stop

Table 2 American College of Gastroenterology Guideline for the Diagnosis and Surveillance of Barrett Esophagus[33]

Patient Characteristics	*Suggested Protocol*
Long-standing reflux symptoms, especially in individuals 50 years of age or older	A single upper endoscopy to rule out Barrett esophagus
Barrett esophagus, no dysplasia	After a second confirming endoscopy, surveillance every 2–3 yr
Barrett esophagus, low-grade dysplasia	Endoscopy every 6 mo for the first year, then annually
Barrett esophagus, high-grade dysplasia	Expert confirmation of the histology; if confirmed, consideration of esophageal resection or endoscopy every 3 mo

the progression of Barrett esophagus to adenocarcinoma or lead to regression in the presence of metaplastic tissue.[25] It is unclear whether surgical antireflux procedures would diminish the risk of developing esophageal adenocarcinoma for those with Barrett esophagus.[26] To reduce the risk of adenocarcinoma, several therapeutic endoscopic ablation modalities (e.g., lasers, argon plasma coagulation, multipolar electrocoagulation, and photodynamic therapy [PDT]) have been used to destroy areas of Barrett esophagus.[27-30] Each of these techniques may successfully destroy the metaplastic epithelium, but it is unclear whether treatment with these ablative therapies reduces risk of adenocarcinoma in the treated areas.[31,32] Studies show that reversion to squamous epithelium might be incomplete; Barrett mucosa may underlie what appears to be normal squamous cell epithelium, and adenocarcinoma may develop beneath the treated squamous cell epithelium. These therapeutic modalities are considered experimental at present, and their long-term utility remains uncertain.

Another issue is the periodicity of upper endoscopy screening appropriate for patients with GERD—especially those who show a change in symptoms—to rule out cancer or to assess for complications of reflux disease (e.g, stricture, erosive esophagitis, or Barrett esophagus). The American College of Gastroenterology has published surveillance guidelines for those patients with long-term reflux symptoms and Barrett esophagus [*see Table 2*].[33] However, there are no prospective trials demonstrating the prevention of esophageal adenocarcinoma or a survival benefit by use of screening endoscopy in patients with long-term GERD. A large population-based study that used data from SEER and Medicare health claims supports the role of screening endoscopy and surveillance for Barrett esophagus and highlights the underdiagnosis of the population at risk.[34] The study showed a shift toward earlier diagnosis of esophageal adenocarcinoma in patients who had a previous endoscopy or a diagnosis of Barrett esophagus and showed a reduced risk of death from esophageal cancer in those who underwent screening endoscopy.

MOLECULAR MUTATIONS AND PATHOGENESIS

With advances in molecular biology and genetics, it has been recognized that the accumulation of alterations in oncogenes, tumor suppressor genes, and DNA mismatch repair genes plays a significant role in the development of esophageal carcinomas. Perhaps the most critical oncogene in esophageal SCC pathogenesis is cyclin D1. Cyclin D1 is overexpressed in as many as 50% of esophageal squamous cell cancers and is associated with poor prognosis. Alteration of tumor suppressor genes *p53* and *p16* and loss of heterozygosity in other tumor suppressor gene loci, such as *pRb* and *APC*, have been described in the development of esophageal carcinomas.[35] The transition of Barrett metaplasia to adenocarcinoma involves progression through low-grade dysplasia, high-grade dysplasia, and carcinoma in situ. It has been demonstrated that abnormal DNA content, as reflected by aneuploidy measured through flow cytometry, is associated with dysplasia and adenocarcinoma. Increased expression of epidermal growth factor receptor and its ligands epidermal growth factor and transforming growth factor–α (TGF-α) are found in Barrett-associated adenocarcinoma when studied by flow cytometry.[36]

DIAGNOSIS

Clinical Manifestations

Few esophageal cancers are diagnosed at an early stage. In more than 90% of patients, dysphagia is the most common presenting complaint, which usually is not noticed until the esophageal lumen is narrowed to less than 40% of normal. The disease most commonly spreads to adjacent structures and lymph nodes. Patients may present with hoarseness, if the recurrent laryngeal nerve is involved, or cough, suggesting local extension into the trachea with resultant tracheoesophageal fistula. Supraclavicular or cervical lymph node metastases may also occur. The disease may also distantly metastasize to liver and lungs, which may present as malignant pleural effusion or ascites. Bone metastases can be identified by pain or by associated hypercalcemia. Other clinical presentations include hematemesis or melena, hemoptysis, and superior vena cava syndrome.

Diagnostic Evaluation

Upper gastrointestinal endoscopy (esophagoscopy), for both visualization and histopathologic diagnosis, should be performed in all patients suspected of having an esophageal abnormality. The combination of endoscopic biopsies and brush cytology has an accuracy of virtually 100% in obtaining tissue diagnosis of esophageal cancer. Bronchoscopy should also be performed to examine the larynx, trachea, and bronchi in patients with suspected esophageal cancer, especially if tumors involve the upper two thirds of the esophagus, to exclude invasion of the posterior membranous trachea or tracheoesophageal fistula and to diagnose associated lung cancers and cancers of the head and neck.

Barium contrast radiography may identify large esophageal lesions and can document contour and motility abnormalities, as well as unexpected airway fistulae [*see Figure 1*]. It does not facilitate biopsies and may not visualize small lesions; however, barium contrast radiography is useful for patients in whom upper gastrointestinal endoscopy is unsuitable (e.g., in patients at risk for sedation or injury from endoscopy) and in cases in which the entire esophagus has not been visualized endoscopically (e.g., in patients whose disease or complications have contraindicated use of endoscopy).

STAGING

Once a diagnosis has been established and careful physical examination and routine blood tests have been performed, a CT scan of the chest, abdomen, and pelvis should be obtained to assess tumor extent, nodal involvement, and metastatic disease.

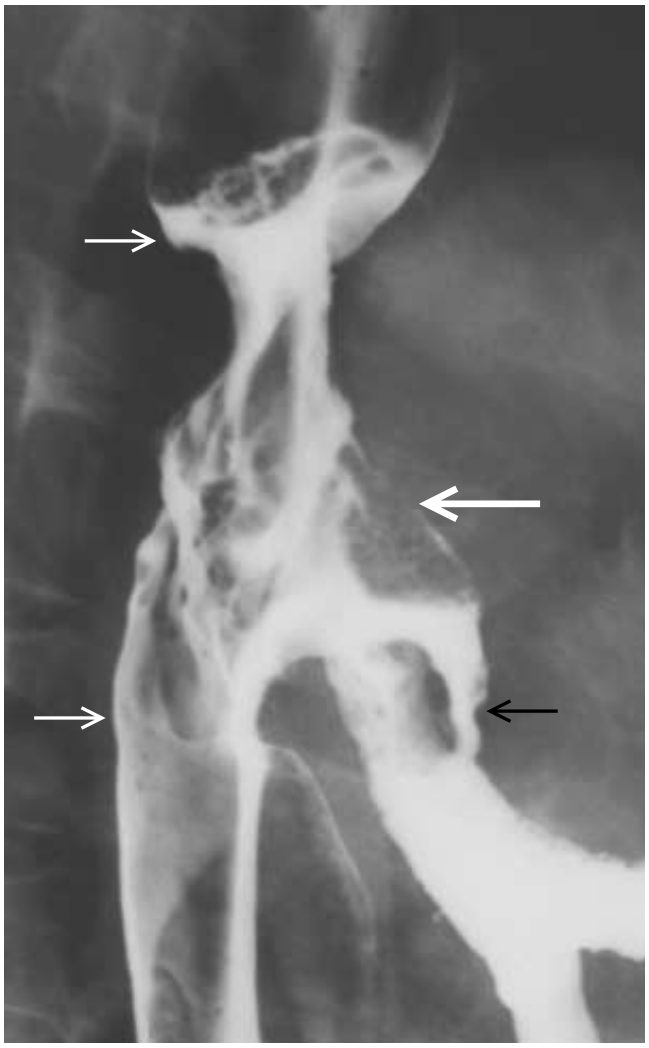


Figure 1 **Barium esophagogram of squamous cell carcinoma. The barium esophagogram shows a squamous cell carcinoma of the esophagus with esophagotracheal fistula. A 5 cm annular mass (between small, white arrows) is located in the midesophagus. The mucosa is nodular. The tumor has invaded the mediastinum, forming a cavity (large arrow). A fistula (black arrow) extends into the left mainstem bronchus.**

However, CT scanning may underestimate the depth of tumor invasion and periesophageal lymph node involvement in up to 50% of cases. Endoscopic ultrasonography (EUS) has the advantage of being able to image distinct wall layers, thereby providing a representation of the depth of tumor invasion with an accuracy of up to 90% and detecting regional lymph node involvement with an accuracy of 75% [*see Figure 2*]. EUS also can detect local tumor recurrence at an early stage. EUS should be considered as a mandatory procedure for staging workup, especially for patients who are being considered for preoperative treatments. Positron emission tomography (PET) scanning has become widely available and may be an important tool for staging, with both a sensitivity and a specificity of approximately 90%.[37] PET scanning is considered to be superior to CT scanning in the evaluation of distant metastases.

TREATMENT

Depending on the depth and degree of invasion of the primary tumor, on extension into surrounding lymph nodes and nearby vital structures, and on metastasis involving distant visceral organs, esophageal cancer can be classified as local and resectable, locally advanced and unresectable, or metastatic. Treatment options for esophageal cancer are based on the stage of the disease at presentation. Surgery remains the mainstay of treatment of esophageal cancer, and it can be curative in persons with resectable local and locoregional disease. Advances in surgical therapy, staging techniques, patient selection, and supportive care in the past several decades have led to a marked improvement in surgery-related mortality and morbidity. However, only about 40% to 60% of patients present with localized disease, and many are not candidates for curative resection. The overall 5-year survival in patients amenable to surgery ranges from 5% to 20%.

Local Disease

Surgery For patients with T1 and T2 lesions without any evidence of nodal involvement, surgery alone may be curative in more than 60% of patients. The two most common approaches for surgical resection of esophageal cancer are transhiatal esophagectomy (THE) and transthoracic esophagectomy (TTE); there are no obvious survival differences between the two procedures.[38] Factors that may affect the selection of one technique over the other include tumor location, depth of tumor invasion, status of lymphadenopathy, overall performance status and body habitus of the patient, previous treatment (i.e., radiation or chemoradiation), and the preference and biases of the surgeon. Comparisons between surgical data from Eastern and Western populations are confounded by the differences in the predominant type of cancer and in the extent of surgical resection.[39]

Surgery combined with other modalities Both local and distal recurrence rates of esophageal cancer are high, which may be based on the anatomic features of the esophagus, including the lack of a fibrous serosa and a rich lymphatic network in the sub-

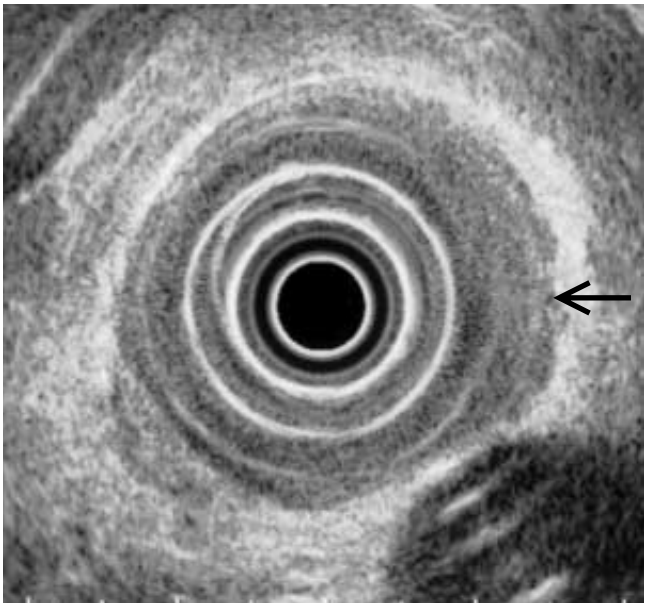


Figure 2 **Endoscopic ultrasonography (EUS) of esophageal cancer. An adenocarcinoma penetrates the full thickness of muscularis propria (arrow) of the esophageal wall.**

mucosa and muscularis. Multimodality treatment of esophageal cancer has been investigated extensively to improve treatment efficacy and survival.

A survival benefit for preoperative chemotherapy followed by surgery, compared with surgery alone, was shown in one study from the United Kingdom,[40] but this result was not confirmed in another large intergroup study from the United States.[41] Preoperative chemoradiation is the most intensely studied treatment strategy. Although the results from three published randomized phase III studies vary and controversies still exist, all favor preoperative chemoradiation.[42-44] Large randomized trials are needed to further address this approach. The benefit of postoperative chemoradiation in esophageal adenocarcinoma is suggested from a large intergroup gastric cancer trial, which included some patients with GE junction adenocarcinomas.[45]

Primary radiation therapy and primary chemoradiation Radiation as a single modality is limited to patients who are not candidates for surgical resection or who are medically unable to tolerate multimodality treatment.[46] Definitive primary chemoradiation therapy for esophageal carcinoma is considered an alternative to surgical resection, although there are no large prospective, randomized trials in which chemoradiation alone has been compared with surgery alone. The Radiation Therapy Oncology Group (RTOG) trial demonstrated the superiority of chemoradiation over radiation alone.[47] A total dose of radiation of 5,000 cGy (in fractions of 200 cGy) is given concurrently with intravenous 5-fluorouracil (5-FU) (1,000 mg/m^2/24 hr for 4 consecutive days) and cisplatin (75 mg/m^2 on day 1 in weeks 1, 5, 8, and 11). In the latest updated analysis, the 5-year survival rate was 26% in the chemoradiation group, compared with 0% in the group that received only radiation; 22% survived in the chemoradiation group at least 8 years after the treatment.[47] No deaths that occurred after 5 years were the result of esophageal cancer, suggesting that cure may be achieved with chemoradiation and indicating that chemoradiation may be considered a viable alternative for nonsurgical candidates.

Symptom management Intraluminal radiation (brachytherapy) or intraluminal stents can ameliorate symptoms resulting from obstructing cancers. Complications of brachytherapy include ulcers, strictures, and tracheoesophageal fistula in up to 25% of patients.[48]

Metastatic Disease

Chemotherapy For distant metastatic disease, systemic chemotherapy may provide palliation. 5-FU and cisplatin have been considered standard treatments, alone and in combination, but response rates (30% to 40%) and median survival (6 to 9 months) are modest.[49] Many new chemotherapy agents developed in the past several years (e.g., paclitaxel, docetaxel, gemcitabine, irinotecan, and oxaliplatin) have been associated with improved response rates and median survival rates in phase II trials but with uncertain advantages over standard therapy.[50-52] Other novel target-oriented agents are being tested in esophageal cancer, including antiangiogenesis agents, farnesyl tranferase inhibitors, vascular endothelial growth factor inhibitors, epidermal growth factor receptor tyrosine kinase (EGFR-TK) inhibitors, anti-EGFR antibodies, and antisense nucleotides.

Symptom management It should be emphasized that improvement in the quality of life, including management of primary symptoms and complications (e.g., dysphagia, malnutrition, pain, and tracheoesophageal fistulas), is extremely important for esophageal cancer patients, especially for those with unresectable and incurable tumors. Endoscopic dilatation can be employed to relieve strictures and is associated with a low complication rate; however, repeated dilations may be necessary.[53] Use of expandable metal stents is a good option to maintain patency; additionally, laser recannulization of the esophageal lumen and bipolar electrocoagulation therapy may be helpful in reducing dysphagia. Endoscopic stents can also control tracheoesophageal fistulas to prevent aspiration pneumonia.[54] Photodynamic therapy with porfimer sodium and an argon-pumped dye laser can provide effective palliation of dysphagia.[55] Surgical placement of a gastrostomy or jejunostomy feeding tube for nutritional support is necessary for many esophageal cancer patients, especially for those receiving perioperative chemoradiation treatments, and may relieve symptoms and improve quality of life in patients with unresectable disease.

Gastric Cancer

EPIDEMIOLOGY

Gastric cancer is an aggressive neoplasm that has a marked variation in both incidence and mortality between different populations. Although the incidence of gastric cancer has decreased in the past several decades in Western countries, gastric cancer is still one of the most common cancers worldwide. Overall prognosis has not changed: only 40% of patients are eligible to undergo potentially curative surgery. Gastric cancer represented the leading cause of cancer-related mortality in the United States until the 1930s and worldwide until the late 1980s, when it was surpassed by lung malignancies. In 2003, an estimated 22,400 persons in the United States will have developed gastric cancer and 12,100 will have died of the disease.[1] Global estimates for 1999 suggest that 789,000 persons were diagnosed with gastric cancer in that year and that 628,000 persons died from the disease.[56] A high incidence of gastric cancer is observed in Asia, South America (Chile and Costa Rica), Eastern Europe, and the Middle East. Although the etiology of gastric cancer has become clearer, there has been an inexplicable worldwide increase in cancers of the proximal stomach and GE junction. In the United States, the 4.3% annual increase in the rate of GE junction cancer is higher than the increase in the rate of lung cancer.

ETIOLOGY AND RISK FACTORS

Extensive research has identified factors and events that influence the initiation, promotion, and progression of stomach cancer. Gastric cancer is a disease of complex etiology involving multiple risk factors, including dietary, infectious, occupational, genetic, and preneoplastic factors. The pattern of decreasing incidence and mortality of gastric cancer in both high- and low-incidence regions [*see Table* 3] may reflect important changes in lifestyle linked to socioeconomic factors, such as refrigeration of stored foods, improved water supply, and a decline in infections and poor dietary habits.[57,58]

Dietary Factors

Salted, smoked, and dried foods contain high concentrations of nitrates, which convert into carcinogenic nitrosamines and nitrites by anaerobic bacteria, and diets rich in such foods may be associated with the development of gastric cancer.

Table 3 Decline in Age-Adjusted Incidence and Mortality of Gastric Cancer in Japan, Slovenia, and the United States, 1975–1995[57]

	Japan		*Slovenia*		*United States*	
	Men	*Women*	*Men*	*Women*	*Men*	*Women*
Decrease in incidence (per 100,000)	76.0 → 53.0	38.4 → 21.3	40.2 → 24.1	16.6 → 10.8	9.5 → 6.9	4.3 → 2.9
Decrease in mortality (per 100,000)	60.2 → 34.2	30.5 → 14.1	37.7 → 21.2	13.8 → 9.0	5.6 → 4.7	2.5 → 2.3

Infections

Infection from *Helicobacter pylori*, a gram-negative spiral bacterium, is a primary risk factor for gastric cancer. The first strong data came from three separate, nested case-control studies.[59-61] The statistically significant relative risk ranged from 2.8 in a British population to 6.0 in a cohort of Japanese males living in Hawaii.

H. pylori gastritis causes cell proliferation with increased risk of DNA damage, leading to inadequate repair and malignant transformation.[62] *H. pylori* infection is very common, with the prevalence ranging from 40% or less in developed countries to 70% or higher in developing countries. Several cofactors or modifiers have been examined to associate *H. pylori* with gastric cancer.[63] Infection at an earlier age (i.e., 29 years or younger) may increase the risk of gastric cancer. Those infected with *cagA*-positive strains (associated with cytotoxin expression) have a risk of developing gastric cancer that is two to three times higher than those with *cagA*-negative strains.[64]

Other gastric cancer risk factors, such as diet (including low intake of ascorbic acid, β-carotene, and vitamin E) and cigarette smoking, have been examined as possible modifiers of the association between *H. pylori* and gastric cancer.[65] Atrophic gastritis secondary to *H. pylori* infection consumption of excessively salty food, achlorhydria, pernicious anemia, or previous partial gastrectomy cause the loss of gastric acidity, which leads to the process of intestinal metaplasia, atypia, and neoplasia. Although reports suggest that patients who underwent gastric resection for benign disease have an increased risk for the development of gastric cancer, this association has not been definitely proved.[66] The relationship of pernicious anemia to the development of gastric carcinoma is also questionable.

Familial Syndromes

Although most cases of gastric cancer appear to be sporadic, approximately 10% of cases involve familial clustering.[67] Hereditary diffuse gastric cancer is a familial syndrome with autosomal dominant inheritance [*see Figure 3*], high penetration (60% to 80%), and high mortality; hereditary diffuse gastric cancer develops at a young age.[68,69] Inactivating mutations in the E-cadherin gene (*CDH1*) have been identified in hereditary diffuse gastric cancer kindreds, which account for approximately 25% of cases of hereditary diffuse gastric cancer worldwide.[70] For those with suspicious familial clustering, genetic tests can clearly identify persons at high risk and allow for targeted intervention, from aggressive surveillance to prophylactic gastrectomy.[69]

Other hereditary cancer syndromes in which gastric cancer may occur include hereditary nonpolyposis colorectal cancer, familial adenomatous polyposis, the Li-Fraumeni syndrome, and the Peutz-Jeghers syndrome. The unique patterns of cancer that define each syndrome must be taken into account in assessing the genetic risk of cancer.[71] For sporadic gastric cancer, molecular analyses also show critical, acquired alterations, including aneuploidy, loss of heterozygosity (LOH) of 17p and 18q, microsatellite instability (MSI), and mutations in *p53*, *p16*, and E-cadherin.[72]

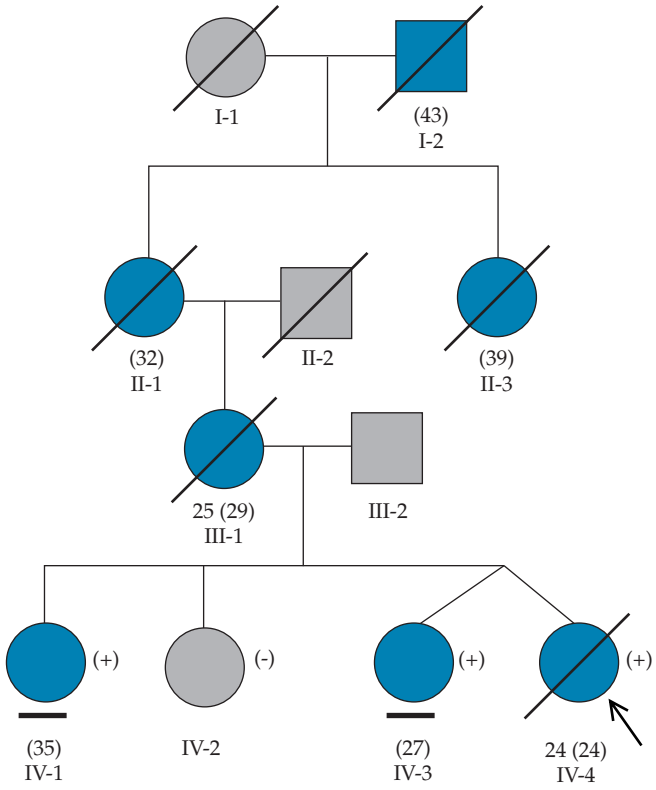


Figure 3 **Pedigree of a family with germline mutation inactivation of E-cadherin and associated cases of hereditary diffuse gastric cancer. Squares represent male family members, and circles represent female family members. Gray symbols indicate unaffected persons, and blue symbols indicate affected persons. A slash over the symbol represents death, and a line under the symbol signifies a family member who received prophylatic gastrectomy. A plus sign to the right of the symbol indicates E-cadherin mutation positive, whereas a minus sign indicates E-cadherin mutation negative. The arrow identifies the proband. The ages at diagnosis and at death (in parentheses) of each individual are marked below each symbol.**

Overexpression of EGF and EGFR, including c-*erbB*-2, TGF-α, and cyclooxygenase–2 (COX2) in gastric cancer are involved with tumor proliferation, invasion, and migration. Recent application of cyclic DNA (cDNA) array technology in gastric cancer has shown at least 40 genes with significantly altered expression in primary gastric cancer.

SCREENING AND PREVENTION

Screening general populations for gastric cancer is not practical in Western countries because of the disease's low incidence. However, mass screening is more appropriate in high-risk areas, such as Japan. Japanese screening guidelines include initial endoscopy at 50 years of age, with periodic follow-up. Routine screening should also be considered in individuals with known genetic predisposition [*see* Familial Syndromes, *above*].

Table 4 TNM Staging of Gastric Cancer[74]

Primary Tumor (T)		
TX		Primary tumor cannot be assessed
T0		No evidence of primary tumor
Tis		Carcinoma in situ: intraepithelial tumor without invasion of the lamina propria
T1		Tumor invades lamina propria or submucosa
T2		Tumor invades muscularis propria or subserosa*
	T2a	Tumor invades muscularis propria
	T2b	Tumor invades subserosa
T3		Tumor penetrates serosa (visceral peritoneum) without invasion of adjacent structures†‡
T4		Tumor invades adjacent structures†‡
Regional Lymph Nodes (N)		
NX		Regional lymph nodes cannot be assessed
N0		No regional lymph node metastasis
N1		Metastasis in 1 to 6 regional lymph nodes
N2		Metastasis in 7 to 15 regional lymph nodes
N3		Metastasis in more than 15 regional lymph nodes
Distant Metastasis (M)		
MX		Distant metastasis cannot be assessed
M1		No distant metastasis
M2		Distant metastasis

Stage Grouping			
Stage 0	Tis	N0	M0
Stage 1A	T1	N0	M0
Stage 1B	T1	N1	M0
	T2a/b	N0	M0
Stage II	T1	N2	M0
	T2a/b	N1	M0
	T3	N0	M0
Stage IIIA	T2a/b	N2	M0
	T3	N1	M0
	T4	N0	M0
Stage IIIB	T3	N2	M0
	T3	N1	M0
Stage IV	T4	N1–3	M0
	T1–3	N3	M0
	Any T	Any N	M1

*A tumor may penetrate the muscularis propria, with extension into the gastrocolic or gastrohepatic ligaments, or into the greater or lesser omentum, without perforation of the visceral peritoneum covering these structures. In this case, the tumor is classified T2. If there is perforation of the visceral peritoneum covering the gastric ligaments or the omentum, the tumor should be classified T3.

†Including spleen, transverse colon, liver, diaphragm, pancreas, abdominal wall, adrenal gland, kidney, small intestine, and retroperitoneum.

‡Intramural extension to the duodenum or esophagus is classified by the depth of the greatest invasion in any of these sites, including the stomach.

PATHOPHYSIOLOGY

Approximately 95% of gastric cancers are adenocarcinomas, with two distinct histologic types: intestinal and diffuse. Other malignancies are very rare, including SCC, carcinoid tumors, adenoacanthoma, leiomyosarcoma, and gastric lymphoma.

Cohesive neoplastic cells that form glandlike tubular structures characterize the intestinal form of gastric cancer, whereas the diffuse type lacks cell cohesion. The intestinal type is more differentiated and is associated with precancerous lesions, gastric atrophy, and intestinal metaplasia. The intestinal form accounts for the majority of distal cancers and is associated with *H. pylori* infection and environmental factors.

The diffuse type of gastric cancer exhibits undifferentiated signet-ring histology and is more common in younger patients. In the diffuse type, there is a predilection for submucosal spread (because of the disease's characteristic lack of cell adhesion), which results in linitis plastica appearance and contiguous spread to the peritoneum. Patients with genetic predisposition are more likely to develop the diffuse type of gastric cancer. The association between *H. pylori* infection and diffuse-type gastric cancer is less clear than the *H. pylori* association with the intestinal type. The prognosis of diffuse gastric cancer is worse than that of the intestinal type, stage for stage.[73]

STAGING AND PROGNOSIS

The overall prognosis of gastric cancer is based primarily on the depth of penetration of the primary tumor, the extent of involvement of local regional lymph nodes, and distant metastasis. To eliminate the likelihood of stage migration (i.e., inappropriate staging) , the American Joint Committee on Cancer (AJCC) sixth-edition staging system changed the nodal staging from anatomic location to the number of nodes involved [*see Table 4*].[74] Stage migration mainly results from variable surgical techniques and extent of lymphadenectomy in different countries and whether node analysis is optimal, as noted from different prognoses in patients with similarly staged gastric cancer in Japan and the United States.[75,76] The AJCC and the International Union Against Cancer (UICC) both have recommended analysis of at least 15 lymph nodes for accurate staging of gastric cancer.

Another important and unique prognostic factor in gastric cancer is the location of the primary tumor. Stage for stage, proximal gastric cancers have worse prognoses, which suggests that the biologies of proximal and distal gastric cancers are different. Proximal gastric cancer in Japan accounts for fewer than 10% of gastric cancer cases, in contrast to much higher percentages in Europe and the United States.[77,78] Other factors, including age and overall performance status, also influence the prognosis of patients with gastric cancer. The prognostic implications of tumor suppressor genes, oncogenes, and factors altering proliferation and apoptosis are areas of active investigation.

DIAGNOSIS

Clinical Manifestations

Because initial clinical symptoms are nonspecific, gastric cancer frequently is diagnosed at an advanced stage. The most com-

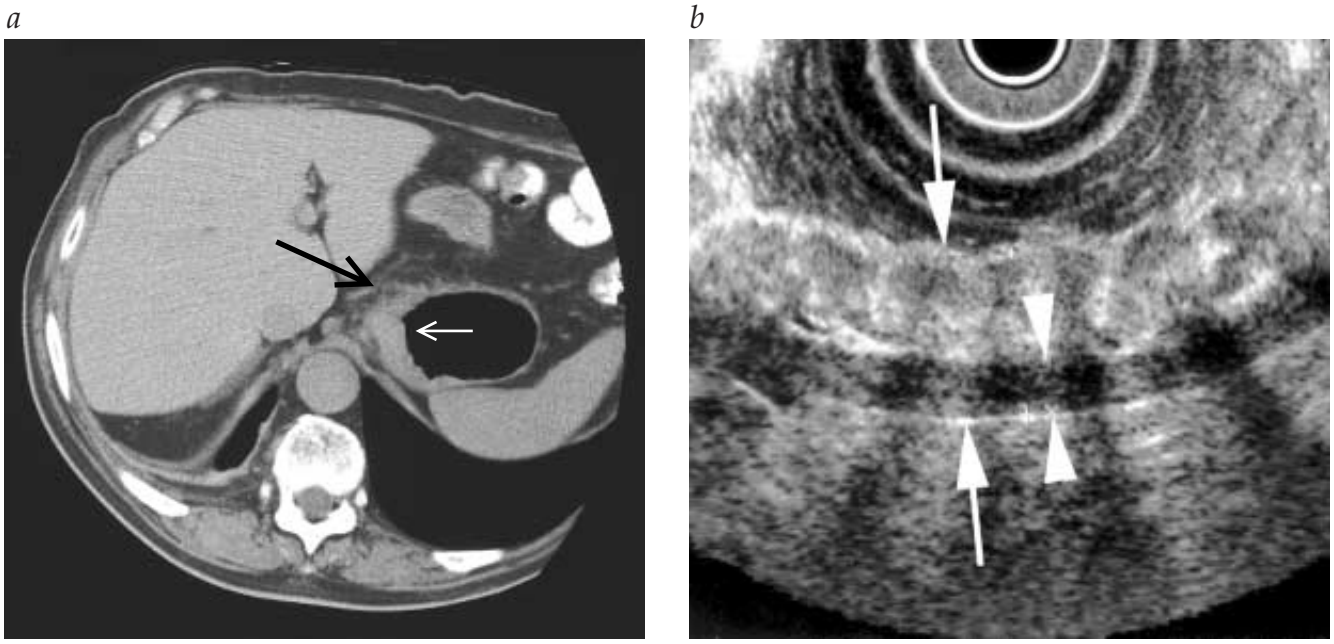


Figure 4 **(*a*) CT scan of gastric cancer. A gastric adenocarcinoma is located at the posterior wall of the gastric fundus (white arrow), and invasion of the gastrohepatic ligament (black arrow) is evident. (*b*) EUS of gastric cancer. The arrows show a linitis gastric cancer with diffused invasion of all levels of the stomach wall and a linitis plastica appearance.**

mon presenting symptoms are abdominal or epigastric pain and weight loss. Early satiety may be the first symptom, especially in those with the linitis plastica form of diffuse gastric cancer. Other symptoms are dysphagia, nausea and vomiting, gastrointestinal bleeding, reflux, fatigue, abdominal wall masses or subcutaneous nodules, and lymphadenopathy. Classic but common physical examination findings are those secondary to metastatic involvement of lymph nodes and peritoneal membrane–namely, periumbilical lymph nodes (Sister Mary Joseph node), supraclavicular lymph node (Virchow node), anterior axillary lymph node (Irish node), and the peritoneal cul-de-sac (Blumer shelf). Two important but infrequent signs of gastric cancer are thrombophlebitis and microangiopathic hemolytic anemia. Other paraneoplastic syndromes that may be associated with gastric cancer include acanthosis nigricans, dermatomyositis, circinate erythemas, pemphigoid, and seborrheic keratoses. Clinical findings include anemia, weight loss, and abnormalities of electrolytes and liver enzymes. About 40% to 50% of gastric cancer cases have increased carcinoembryonic antigen (CEA), and 30% have elevated α-fetoprotein (AFP) or serum carbohydrate antigen 19-9 (CA19-9).

Diagnostic Evaluation

Upper gastrointestinal endoscopy is routinely performed as an initial procedure to identify suggestive lesions and has a diagnostic accuracy of 95%. Double-contrast barium swallow is only 75% accurate and should be used only when upper GI endoscopy is unavailable or patients are unable to tolerate the procedure. Although CT scanning is useful for assessing local extension, lymph node involvement, and the presence of metastasis, understaging is a major issue. EUS, a relatively new staging technique, can assess the depth of tumor invasion and perigastric regional nodes more accurately than CT and should be mandatory for those considered for surgical resection or neoadjuvant treatment [*see Figures 4a and b*].[79]

TREATMENT

Local Disease

Surgery Surgical resection with extended lymphadenectomy and negative surgical margins remains the only therapeutic modality capable of cure.[80] With the application of screening and the awareness of the high risk of gastric cancer in Japan's general population, many early-stage, noninvasive cancers are detected,[81] allowing for more complete surgical resection than is generally accomplished in Western countries.

The location and extent of the primary cancer dictate the type of surgical procedure. With distal lesions, subtotal gastrectomy is preferred over total gastrectomy if the fundus or GE junction is not involved, because the former provides comparable survival and lower morbidity. Proximal gastrectomy may be performed for tumors in the GE junction or the proximal third of the stomach, but this procedure is associated with increased postoperative morbidity and mortality. Total gastrectomy is more appropriate if tumor involvement is diffuse or if it arises in the body of the stomach and has extension to within 6 cm of the cardia. Transection of the distal esophagus and proximal duodenum with an omentectomy is required when a total gastrectomy is performed.

Complete (R0) resections along with depth of invasion and adequate nodal staging remain the most important prognostic factors, shown by the results of the National Cancer Data Base report that provides stage-stratified survival rates [*see Figure 5*]. Considerable controversy has surrounded the notion of what is defined as an adequate lymphadenectomy for the potentially curative treatment of gastric cancer. In a D1 dissection, only the perigastric lymph nodes of the stomach (N1 lymph node group: right and left cardia, lesser and greater curvature, and suprapyloric and infrapyloric lymph nodes) are removed. Extended D2 dissection involves more extensive N2 lymph node group resection (left gastric, common hepatic, splenic, and celiac axis artery lymph nodes). For patients with N2 lymph node

disease, complete resection of diseased tissue can be achieved only by D2 dissection.[82,83] This operation has become standard practice in Japan and other high-incidence countries because of the apparent improvement in disease-free and overall survival in nonrandomized comparisons; however, it has been difficult to separate the effect of stage migration from the therapeutic effects of lymphadenectomy. Controversy still exists over the benefit of D2 dissection because of the lack of evidence for improved survival as demonstrated in randomized, controlled trials.[84,85] Because of the excess morbidity and mortality associated with the D2 dissection, many surgeons now advocate the D1+ procedure, which entails extended lymphadenectomy but no splenectomy.

There is no standard postsurgical follow-up, but options include periodic clinic visits for liver function tests and radiologic imaging. Unlike colon cancer, for which a small but clinically important group of patients with metastatic disease can be cured with surgery, there is no effective salvage therapy for recurrent gastric cancer, limiting the effectiveness of routine surveillance. If subtotal gastrectomy is performed, periodic upper endoscopy is recommended to detect second primary tumors. Nutritional support, including vitamin B_{12} supplementation, should be delivered postsurgically.

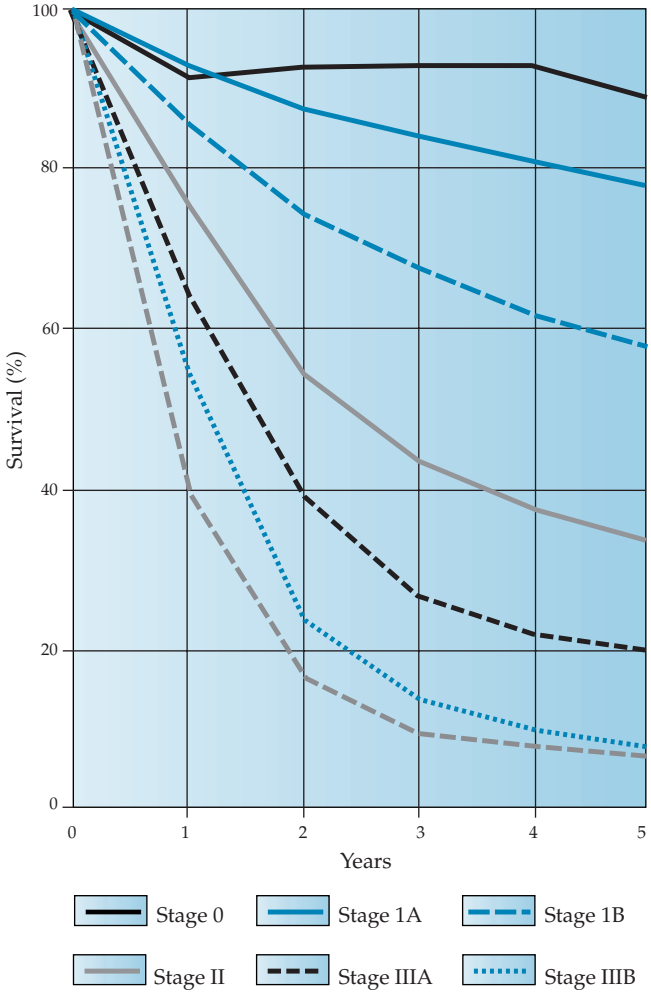


Figure 5 **Gastric cancer survival rate (years 1985 through 1996) after gastrectomy by stage.**

The 5-year survival rate after curative resection for gastric cancer is only about 40%. Treatment failure results from a combination of local or regional recurrence and distant metastases.

Surgery combined with other modalities Multidisciplinary approaches using postoperative (adjuvant) and preoperative (neoadjuvant) therapeutic principles have received increasing attention as strategies to increase the cure rate from surgery. Benefits of early studies involving chemotherapy or radiation alone as adjuvant treatment are not clear because of inadequate clinical trial design and the lack of effective regimens.[86-88] However, a survival benefit from adjuvant chemoradiation therapy in patients who had gastric cancer resection was demonstrated in a large, well-controlled prospective study in the United States.[45] Patients with complete resections were randomized to receive either surgery alone or surgery plus postoperative chemotherapy with 5-FU and leukovorin (LV) and concurrent radiotherapy.[45] Patients in the treatment group were treated over 6 to 7 months with regimens of chemotherapy (intravenous 5-FU and LV), followed by chemoradiation (4,500 cGy [fractions of 180 cGy/day] with 5-FU/LV), and more chemotherapy (5-FU/LV) after the chemoradiation treatment. Benefit in both disease-free survival and overall survival was evident in the treatment group (3-year disease-free survival of 49% versus 32%, and 3-year overall survival of 52% versus 41%, $P = 0.005$). The median duration of relapse-free survival was 30 months in the chemoradiation group and 19 months in the surgery-alone group ($P < 0.001$). This combination of adjuvant chemotherapy and chemoradiation is now considered the standard treatment in the United States. However, two important issues were identified in the study. One is the precision and accuracy of the radiation field, which is based on the result of locoregional patterns of recurrence of gastric cancer.[89] All radiation treatment plans were centrally prereviewed, and 35% were found to contain potential major or minor deviations (because of either potential critical-organ toxicity or failure to treat protocol-defined target volumes) requiring correction.

Another issue is the significance of nutrition for this population, in which chemoradiation might do harm without proper support. J-tube placement is encouraged for these patients before treatment. The results of this trial, therefore, can be generalized only if an experienced group of surgeons, radiation oncologists, and medical oncologists, together with nutritional support personnel, are involved in patient selection and treatment. This trial has also been criticized because few patients underwent D2 dissection, and the assumption has been made that the benefit from treatment was simply improvement in local-control outcome after inadequate surgery. Without a drastic change in surgical practice, postoperative treatment appears advisable; however, additional trials of chemoradiation are warranted in patients who have undergone D1+ or D2 dissections.

Preoperative chemoradiation therapy can reduce the size of the tumor, effectively downstaging gastric cancer (including pathologic complete response [pCR]) in some patients, to the extent that an initially unresectable cancer may be converted to resectable status.[90,91] Large phase III randomized trials are needed to further evaluate preoperative chemoradiation in terms of potential improvements in treatment tolerability, tumor resectability, and relapse or survival outcomes.

For patients with locally advanced unresectable gastric cancer, a combination of chemotherapy and radiotherapy is also the treatment of choice.

Metastatic Disease

When possible, all newly diagnosed patients with metastatic gastric cancer should be considered for clinical trials. Although many chemotherapy agents show activity in advanced gastric cancer, the overall benefit is still disappointing; on average, trials show a median survival of 8 to 9 months. Single agents include 5-FU, cisplatin, mitomycin, anthracyclines (doxorubicin and epirubicin), etoposide, and methotrexate. Combination treatments (e.g., cisplatin and 5-FU) may increase response rate but with only a modest impact on overall survival.[92,93] There are several new agents that have demonstrated activity in gastric cancer, including taxanes (paclitaxel and docetaxel), oxaliplatin, and irinotecan. The response rates for combinations of these new agents are 50% or higher in selected patients from phase II trials, but the confirmation that these regimens are superior to standard regimens (e.g., 5-FU and cisplatin) requires phase III evaluation.[94-96] Another treatment option available for highly selected patients with gastric cancer is intraoperative hyperthermic intraperitoneal chemotherapy.[97] Many novel target-oriented agents are currently under study for gastric cancer treatment.

Pancreatic Cancer

EPIDEMIOLOGY

Pancreatic cancer is the fourth leading cause of death from cancer in both males and females in the United States.[1] From 1930 to 1970, the incidence increased approximately twofold, to 10 cases per 100,000 men and 7.2 cases per 100,000 women, but the incidence has stabilized and even decreased modestly. The yearly incidence and mortality are roughly equivalent, indicating the lethal nature of this disease. In 2002, an estimated 30,300 persons in the United States and 200,000 persons worldwide were diagnosed with pancreatic cancer. Approximately 95% of pancreatic malignant tumors are exocrine pancreatic cancers, two thirds of which occur in the pancreatic head and one third in the pancreatic body and tail; the remaining 5% of malignant lesions are mostly islet cell tumors.

The mean age of patients at onset of pancreatic cancer is about 65 years; the disease is extremely unusual in patients younger than 30 years, and it rarely occurs before age of 50. The incidence of pancreatic cancer is higher in males than in females and is higher in blacks than in whites.[98] An association between diabetes and pancreatic cancer has been recognized; however, a causal relationship has not been fully elucidated.[99,100]

ETIOLOGY AND RISK FACTORS

Tobacco Smoking

Tobacco smoking has been the most consistently demonstrated risk factor, implicated as a cause in roughly 30% of cases of pancreatic cancer.[101-103] Most studies show a dose response, and the relative risk ranges from 1.5 in light smokers to 10 for two-pack-a-day smokers. Smoking remains the strongest risk factor amenable to preventive intervention.

Age

Age is also an extremely important determinant of risk. With increasing age, the risk of pancreatic cancer increases exponentially.[98]

Table 5 Familial Pancreatic Carcinoma[109]

Syndrome	*Affected Gene*	*Frequency (%)*
Familial pancreatic cancer	Unknown	< 1
Familial breast/ovarian cancer	*BRCA2*	5–7
Hereditary nonpolyposis colorectal cancer	MMR genes	< 3
Peutz-Jeghers syndrome	*LKB1/STK11*	< 1
Familial atypical multiple mole melanoma	*p16*	< 1
Hereditary pancreatitis	*PRSS1*	< 1
von Hippel–Lindau disease	*VHL*	< 1

MMR—mismatch repair

Dietary Factors

Dietary factors have been linked to pancreatic cancer[104]; however, more convincing evidence is needed to confirm such associations. High levels of folate seem to reduce the risk of pancreatic cancer in smokers.[105] Coffee and alcohol consumption do not seem to increase the risk of pancreatic cancer.[106]

Predisposing Medical Conditions

The abrupt onset of diabetes in a previously healthy, nonobese middle-aged or older patient with no family history of diabetes can be the first sign of pancreatic cancer. Long-standing chronic pancreatitis is associated with pancreatic cancer, and the increased risk is independent of the cause of pancreatitis, gender, and nation of origin.

Medications

A large prospective study examined the association between the use of asprin and other nonsteroidal anti-inflammatory drugs (NSAIDs) and the incidence of pancreatic cancer in 28,283 postmenopausal women.[107] The data from the study showed a trend of decreasing risk of pancreatic cancer with increasing frequency of aspirin use. There was no association shown between nonaspirin NSAID use and the incidence of pancreatic cancer.

Genetic Factors and Familial Syndromes

Inherited genetic predisposition is observed in patients diagnosed with pancreatic cancer. Up to 10% of patients report pancreatic cancer in a first-degree relative.[108] The data from the National Familial Pancreatic Tumor Registry (NFPTR) demonstrated an 18-fold risk of pancreatic carcinoma in first-degree relatives of familial pancreatic carcinoma kindreds. When three or more members in a kindred are affected, the risk is increased 57-fold.

A variety of familial syndromes have been described [*see Table 5*]; however, some remain without a known genetic basis.[109] Meticulous attention to family history is required to detect these genetic syndromes, because a familial syndrome may present as a clustering of pancreatic cancers among several first- and second-degree relatives, or it may be even more subtly distributed among extended family members.

SCREENING AND PREVENTION

Pancreatic cancer is a poor candidate for population-based screening because of relatively low incidence, lack of cost-effec-

tive or clinically proven screening tools, and limited options for effective management, even with early tumors. For patients at high risk for pancreatic cancer, such as those with a familial syndrome, screening may have true value, although the efficacy of screening has not been demonstrated. CA19-9, a serum tumor marker, may have value for following the therapeutic response of patients with pancreatic cancer[110]; however, it has not proved to be useful in screening, because (1) approximately 10% to 15% of individuals do not secrete CA19-9; (2) the CA19-9 level may be within normal range in the early stage of pancreatic cancer; and (3) the CA19-9 may be elevated in benign conditions (e.g., chronic pancreatitis or acute cholangitis).[111,112] Given that K-*ras* mutations are present in approximately 90% of pancreatic cancers and as relatively early transforming events, they have been sought in pancreatic juice, duodenal juice, feces, and serum.[113-115] However, K-*ras* mutations can be present in pancreatitis or dysplastic lesions that may not necessarily progress to pancreatic cancers, and so these mutations lack sensitivity and specificity as markers for pancreatic cancer. Improved screening and early detection may derive from increased knowledge of molecular events in pancreatic cancer and new techniques (e.g., measurement of telomerase activity and comprehensive cellular protein analysis).

MOLECULAR MUTATIONS AND PATHOGENESIS

In the past 10 years, the understanding of the molecular progression of pancreatic cancer has increased, making this malignancy one of the best genetically characterized tumors and leading to the formation of a progression model based on histopathologic, clinical, and genetic evidence.

Genetic analysis of pancreatic carcinomas has proved that the so-called multi-hit model of carcinogenesis applies to this tumor type [*see Table 6*].[116] Precursor lesions of pancreatic cancer have been recognized in careful morphologic studies; however, a variety of terms were previously used to refer to these precursor lesions. The nomenclature for the precursor lesions was recently standardized to four categories of pancreatic intraepithelial neoplasia (PanIN) [*see Table 7*].[117] Examination of pancreatic duct precursor lesions and pancreatic adenocarcinomas by modern molecular techniques has led to the identification of certain key genetic alterations that are expressed at different points in the progression of the lesions. Point mutations in K-*ras* and overexpression of HER-2/*neu* are identified in the earliest pancreatic ducts lesions (PanIN-1A and PanIN-1B) and are thought to be early genetic events in the development of pancreatic cancer. Inactivation of the *p16* tumor suppressor gene occurs in the higher-grade duct lesions (PanIN-2 and PanIN-3). The tumor suppressor genes *p53, DPC4* (deleted in pancreatic carcinoma locus 4), and *BRCA2* are lost even later in the development of pancreatic neoplasia (PanIN-3).[118] These morphologic and genetic data support the progression model for the development of pancreatic neoplasm from low-grade PanIN to carcinoma in situ (PanIN-3) and, subsequently, to infiltrating carcinoma [*see Figure 6*]. This progression model may be very important in the development of screening tests to detect early, potentially curable pancreatic neoplasms and may also help find targets for prevention.

Table 6 Mutations Involved in Pancreatic Cancer[116]

Gene	*Chromosomal Locus*	*Frequency (%)*
K-*ras*	12p13	*> 90*
p16	9p21	*> 80*
p53	17p13	*50–75*
DPC	18q21	*50*
HER-2/neu	17q21	*20–25*
BRCA2	13q12	*7*

Table 7 Pancreatic Intraepithelial Neoplasia (PanIN)[117]

Nomenclature	*Description*
PanIN-1A	A flat mucinous epithelium, no sign of atypia
PanIN-1B	With a papillary architecture, no sign of atypia
PanIN-2	A prevalence of papillary architecture, sign of atypia, low- to moderate-grade dysplasia
PanIN-3	A prevalence of papillary architecture, sign of atypia, high-grade dysplasia

DIAGNOSIS

Clinical Manifestations

Initial symptoms experienced by pancreatic cancer patients are insidious and relatively nonspecific (e.g., weight loss, anorexia, abdominal discomfort or pain, and nausea), which may delay the diagnosis for several months. Specific symptoms usually relate to localized invasion of peripancreatic structures. Pain can be a presenting symptom and is usually associated with localized invasion of peripancreatic structures (e.g., splanchnic plexus and retroperitoneum), particularly from lesions located in the body or tail of the pancreas. Pain is typically described as gnawing and severe, radiating to the back and worsening in the supine position. Jaundice as a presenting symptom is most commonly related to compression of the common bile duct by tumors in the head of the pancreas. Anorexia and weight loss partly result from cytokines released by tumor cells, although decreased caloric intake and malabsorption may also play a role. Invasion of the duodenum or gastric outlet may give rise to nausea or vomiting as a presenting symptom. The Courvoisier sign, a palpable gallbladder in the absence of cholecystitis or cholangitis, presents in about 25% of all patients with pancreatic cancer. Trousseau syndrome (migratory superficial phlebitis) and venous thrombosis may present as initial symptoms of the disease.

Diagnostic Evaluation

The early diagnosis of a potentially resectable pancreatic cancer is extremely difficult because of nonspecific initial symptoms and poor sensitivity of noninvasive techniques such as CT and ultrasonography. Compared to ultrasound, CT provides better definition of the tumor and surrounding structures and is less operator dependent. CT correctly predicts unresectable tumors in 85% of patients and resectable tumors in 70% of patients. Spiral CT increases the accuracy of staging pancreatic carcinoma and visualization of vessel encasement. MRI improves the visualization of pancreatic tumor and may detect distant metastatic disease not seen on CT scan [*see Figure 7*]. Endoscopic retrograde cholangiopancreatography (ERCP) may prove useful as a complement to CT or MRI in diagnosing and staging ambiguous cas-

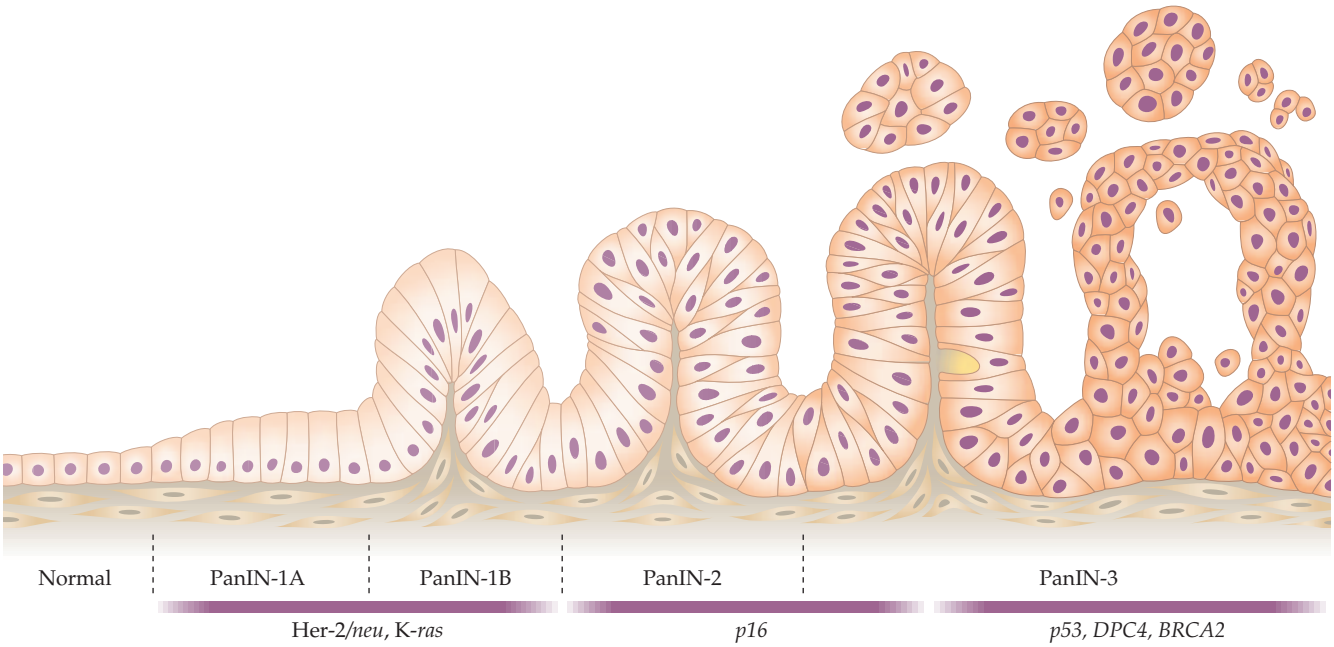


Figure 6 **Progression model for pancreatic cancer.**

es; ERCP may also supply cytologic information. Endoscopic ultrasonography (EUS) is the single most accurate test for imaging and staging pancreatic carcinoma [*see Figure 8*][119] and can clearly evaluate pancreatic mucosal, vascular, ductal, and parenchymal abnormalities, as well as lymph node metastases.[120]

Patients with clinical symptoms that may represent pancreatic cancer should have an initial standard CT scan or an abdominal ultrasound. If a pancreatic mass is suspected on one of these initial tests, further evaluation is necessary. If the tumor appears to be larger than 4 cm or appears unresectable, spiral CT with intravenous contrast and ERCP with fine-needle aspiration (FNA) should be considered. On the basis of size alone, masses greater than 4 cm have less than a 10% chance of being resectable and nonmetastatic.[121] If a patient has only a dilated duct and no obvious mass or a tumor smaller than 4 cm with no other signs of unresectability (e.g., metastasis), EUS is performed for further evaluation. EUS improves the prediction of resectability for tumors, which may prevent unnecessary laparotomy.

STAGING

Because surgical resection is the only curative modality for pancreatic cancer and because only 10% to 15% of patients present with resectable disease, the diagnosis, stage, and management are based on resectability. The sixth edition of the American Joint Committee on Cancer TNM classification contains changes in both clinical and pathologic staging [*see Table 8*].[74]

TREATMENT

Local, Resectable Disease

Surgery If the tumor is deemed to be resectable after an extensive workup, a standard pancreaticoduodenectomy (Whipple procedure) or pylorus-preserving Whipple procedure is performed. Total pancreatectomy is also performed occasionally in selected patients. Experienced surgeons should perform these operations, because greater surgical experience and a high case volume are associated with better and safer outcomes.[122] Operative results have improved dramatically in the past decade, owing to better staging of the disease, improved surgical techniques and perioperative care, and the identification of high-volume centers. Although the mortality has decreased to below 5%, morbidity rates have remained at 30% to 50%.[123] The pylorus-preserving Whipple procedure theoretically eliminates nutritional problems caused by a reduced gastric reservoir and gastric dumping. Tumor recurrence and long-term survival appear unaffected by the choice of procedure; instead, prognosis is related to the size of the primary tumor, status of the resected lymph nodes, and margins.[124] The median survival for those patients who have had complete resection is approximately 20 months.

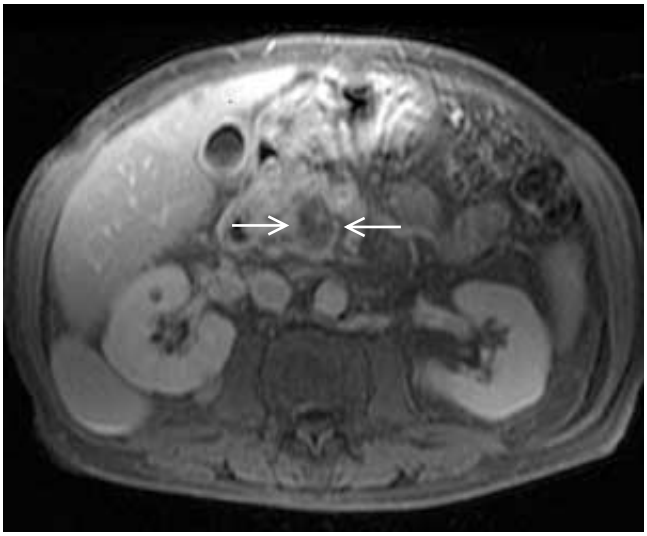


Figure 7 **Magnetic resonance image of pancreatic cancer. An axial T_1-weighted MRI shows the low-signal intensity of a mass located at the pancreatic head (arrows) .**

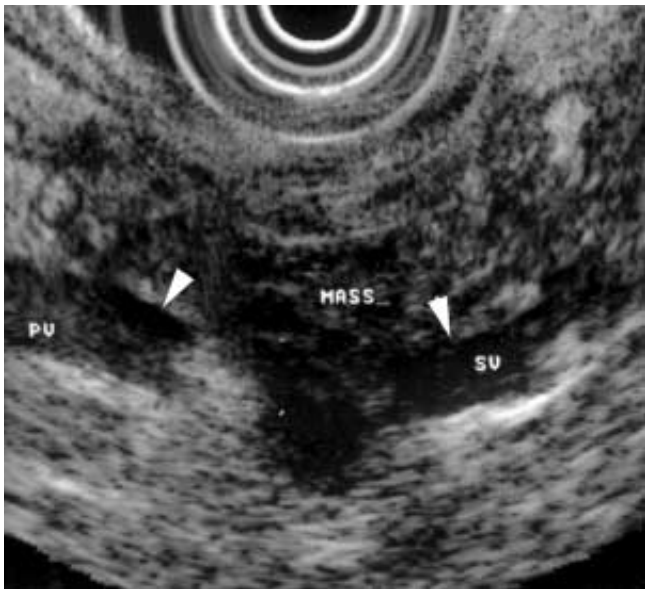


Figure 8 **An EUS of pancreatic cancer shows direct invasion of the superior mesenteric vein (SV). Arrows show extent of invasive mass. (PV—portal vein)**

Surgery combined with other modalities The issue of adjuvant chemoradiation therapy after complete resection for pancreatic cancer remains somewhat unsettled, although it is frequently recommended in the United States. The benefit of adjuvant chemoradiation therapy was suggested from a small, randomized study of the Gastrointestinal Tumor Study Group (GITSG). Twenty-one patients who received a combination of an intravenous bolus of 5-FU and radiotherapy after surgery had a median survival of 20 months, compared to a survival of 11 months for 22 patients treated with surgery alone.[125] Similar results were demonstrated in a larger, single institutional nonrandomized study, which showed a median survival of 19.5 months in patients who received adjuvant chemoradiation therapy and 13.5 months in patients who received no further treatment.[126] However, a large study from the European Study Group for Pancreatic Cancer (ESPAC) showed no survival benefit for adjuvant chemoradiation therapy (median survival of 15.5 months in the treatment group, 16.1 months in the control group) but suggested a potential benefit for adjuvant chemotherapy (median survival, 19.7 months in the treatment group, 14.0 months in the control group; $P = 0.0005$).[127] There have been no large, definitive, well-designed studies assessing the true benefit of adjuvant treatment in cases of pancreatic cancer, although adjuvant therapy is recommended for most patients in the United States on the basis of the GITSG trial. The most recently completed United States intergroup trial compared two chemoradiation therapies without a surgery-alone control arm.

Locally Advanced, Unresectable Disease

For locally advanced, unresectable pancreatic cancer, either chemoradiation or chemotherapy alone is a reasonable treatment option. External-beam irradiation (3,750 to 5,500 cGy) may relieve local symptoms by reducing the tumor size.

Chemoradiation Chemoradiation in regimens including 5-FU has become the accepted standard of care.[128] In an attempt to improve local control and prevent radiation toxicity to the normal tissue, intraoperative radiotherapy (IORT) has been evaluated but has failed to show superiority to external-beam irradiation.[129] Gemcitabine, which is the standard agent in chemotherapy for pancreatic cancer, is a potent radiation sensitizer and is tolerated concomitantly with radiotherapy only at low doses.[130] Studies of this combination are under way to establish whether there is superiority of gemcitabine with radiation (concurrently and sequentially) over the current standard of radiation with 5-FU. A current United States trial is comparing gemcitabine alone with chemoradiation to assess the contribution of radiation in patients with locally advanced disease. It is reasonable to assume that radiotherapy may not be required for a subset of patients who have locally advanced disease and no local symptoms or that initiation of radiotherapy can be delayed, in the absence of metastatic disease, until local control is needed.

Chemotherapy Gemcitabine, as compared with 5-FU, is now considered the standard chemotherapeutic agent for advanced and metastatic pancreatic cancer on the basis of clinical-benefit response (reduction in pain intensity, analgesic use, and weight gain and improvement of performance status) and survival benefit.[131] The overall results of chemotherapy for patients with advanced metastatic pancreatic cancer are still disappointing, and there are many ongoing studies of new molecular targeting agents or combinations of cytotoxic agents,[132,133] but as yet, none has been reported to have clinical benefit superior to that of gemcitabine alone.

Table 8 TNM Staging of Pancreatic Cancer[74]

Primary Tumor (T)	
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor limited to the pancreas, 2 cm or less in greatest dimension
T2	Tumor limited to the pancreas, more than 2 cm in greatest dimension
T3	Tumor extends beyond the pancreas but without involvement of the celiac axis or the superior mesenteric artery
T4	Tumor involves the celiac axis or the superior mesenteric artery (unresectable primary tumor)
Regional Lymph Nodes (N)	
NX	Regional lymph nodes cannot be assessed
N1	No regional lymph node metastasis
N2	Regional lymph node metastasis
Distant Metastasis (M)	
MX	Distant metastasis cannot be assessed
M1	No distant metastasis
M2	Distant metastasis

Stage Grouping			
Stage 0	Tis	N0	M0
Stage 1A	T1	N0	M0
Stage 1B	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1	N1	M0
	T2	N1	M0
	T3	N1	M0
Stage III	T4	Any N	M0
Stage IV	Any T	Any N	M1

Symptom management It should be emphasized that supportive care and symptom relief are important for patients with advanced and metastatic pancreatic cancer; pain control may be accomplished with analgesics or celiac plexus neurolysis, and biliary drainage may be achieved either by surgical bypass procedures or by use of metal stents placed endoscopically or percutaneously.

Hepatocellular Carcinoma

EPIDEMIOLOGY AND ETIOLOGY

Hepatocellular carcinoma (HCC) is the most common primary malignant tumor of the liver. It is the fifth most common malignancy in the world (564,000 cases a year) and the third cause of cancer-related deaths worldwide.[134]

HCC is most often a complication of liver cirrhosis caused by chronic infection by the hepatitis B virus (HBV), hepatitis C virus (HCV), or alcohol.[135] Its incidence varies widely internationally, with more than 85% of cases occurring in countries with a high prevalence of HBV infection. HCC is considerably less common in Europe and North America (two to seven per 100,000 population) than in Taiwan and southeastern China (30 per 100,000 population).[136] However, the incidence of HCC in the United States has increased from 1.4 per 100,000 population for the period from 1976 through 1980 to 2.4 per 100,000 for the period from 1991 through 1995. This increase is considered to be primarily related to an increase in HCV infection.[137] It has been reported that in the United States, approximately 1.8% of the population has anti-HCV antibody, and four million persons are infected with HCV, which is a risk factor for chronic liver disease. The incidence of HCC in patients with cirrhosis is about 3% to 5 % a year.[138] It is estimated that 7% to 14% of HCV-infected patients with cirrhosis in the United States will develop HCC within 10 years.[139]

Concomitant with the rising rates of HCC, there has been a shift of incidence from elderly to relatively younger populations (40 to 60 years).[140] A Canadian epidemiologic study also showed an increased incidence of HCC in Canada since 1991, which is consistent with the reported increase in the prevalence of HCV and HBV infection.[141] In Japan, because of passive and active immunoprophylaxis of HBV, approximately 16% of cases of HCC are caused by HBV, whereas 80% of cases are associated with HCV infection.[142]

Hereditary hemochromatosis is also a risk factor for the development of HCC.[143] Diabetes mellitus may be associated with HCC in patients with chronic HCV infection,[144] and a significant synergy exists between heavy alcohol consumption, hepatitis virus infection (both HBV and HCV), and diabetes mellitus and the development of HCC.[145] Patients with other metabolic disorders or conditions that may lead to cirrhosis (e.g., α_1-antitrypsin deficiency, type I glycogen storage disease, tyrosinemia, and even biliary atresia) are also at risk of developing HCC.[146] Other risk factors include long-time ingestion of food contaminated with aflatoxins, metabolites of the mold *Aspergillus flavus*, and long-term use of oral contraceptives and exogenous androgens.[147-149]

SCREENING AND PREVENTION

In some Asian countries in which there is a high prevalence of HBV and HCC, screening programs have been developed to detect small tumors and to allow treatment at an early stage of the disease and have resulted in favorable outcomes.[150] Mother-to-baby transmission of HBV has been prevented by a combined passive and active immunoprophylaxis with hepatitis B immune globulin and vaccine, which has reduced the chronic carrier state by 10-fold,[151] suggesting that the incidence of HBV-associated HCC will decline gradually over the next several decades.

Because aflatoxin contamination of food and excessive alcohol consumption are important risk factors for HCC development, improvement of public health standards and avoidance of known carcinogens may also be important. Antiviral therapy in patients with chronic HBV infection may increase survival and decrease HCC development; however, the data are limited and still controversial.[152] Prevention of initial HCV infection is the key to preventing HCV-related HCC. Evidence suggests that effective control of HCV infection or HCV-related inflammation with interferon may decrease the incidence, or at least delay the development, of HCC.[153,154]

Surveillance and screening for HCC in patients with recognized risk factors remain controversial.[155] There are no randomized, controlled trials confirming that surveillance for HCC improves disease-related or overall survival. AFP is not an accurate surveillance test because of low sensitivity and specificity. Periodic ultrasound examination may be the best choice for surveillance; however, the optimal screening interval has not been defined, and there are no adequate analyses of cost efficacy.

DIAGNOSIS

Clinical Manifestations

The most common clinical feature of HCC is right upper quadrant or epigastric pain. Patients may initially present with ascites, weight loss, anorexia, malaise, vomiting, and jaundice. Some patients may present with an acute abdomen resulting from a tumor rupture. The most common physical signs are hepatomegaly and the stigmata of cirrhosis, but abdominal bruits secondary to increased vascularity, splenomegaly, Budd-Chiari syndrome, and Virchow node may also occur. Abrupt worsening of portal vein hypertension may also be the first sign of HCC. A variety of paraneoplastic syndromes may occur, including hypoglycemia, hypercalcemia, erythrocytosis, and porphyria cutanea tarda.

Common laboratory abnormalities include elevated levels of liver transaminases, alkaline phosphatase, and bilirubin; but normal liver tests do not exclude HCC. A low albumin level may reflect poor synthetic function of the liver. The AFP level is elevated in 85% of patients with HCC. An AFP level greater than 400 ng/ml (normal range: 0 to 20 ng/ml) is considered diagnostic for HCC. False positive results may be due to acute or chronic hepatitis, germ cell tumors, or pregnancy. The *des*-γ-carboxy prothrombin protein induced by vitamin K is elevated in more than 90% of HCC patients, but it is not specific for the disease.[156]

Diagnostic Evaluation

A variety of imaging studies are available to assist in the diagnosis, staging, and evaluation of HCC. Ultrasonography is inexpensive, widely available, and useful for screening high-risk populations. CT can provide information on the extent of hepatic involvement, invasion or thrombosis of hepatic and portal veins, regional lymph nodes, splenomegaly, and ascites. MRI with intravenous gadolinium and angiography can provide details about vascular anatomy and can detect small hepatic tumors, which is important in determining resectability of lesions and guiding local-regional treatment.

Table 9 TNM Staging of Hepatocellular Carcinoma[74]

Primary Tumor (T)			
TX	Cannot be assessed		
T0	No evidence of primary tumor		
T1	Solitary tumor without vascular invasion		
T2	Solitary tumor with vascular invasion or multiple tumors ≤ 5 cm		
T3	Multiple tumors ≤ 5 cm or tumor involving a major branch of the portal of hepatic vein		
T4	Tumor(s) with direct invasion of adjacent organs other than the gallbladder or with perforation of visceral peritoneum		
Regional Lymph Nodes (N)			
NX	Cannot be assessed		
N0	No regional lymph node metastasis		
N1	Regional lymph node metastasis		
Distant Metastasis (M)			
MX	Cannot be assessed		
M0	No distant metastasis		
M1	Distant metastasis		
Stage Grouping			
Stage I	T1	N0	M0
Stage II	T2	N0	M0
Stage IIIA	T3	N0	M0
IIIB	T4	N0	M0
IIIC	Any T	N1	M0
Stage IV	Any T	Any N	M1

STAGING AND PROGNOSIS

HCC arises from epithelial cells and histologically is an adenocarcinoma. HCC may present as a nodular, diffuse, or massive form. The spectrum includes typical HCC, undifferentiated HCC, mixed HCC with cholangiocarcinoma, and the fibrolamellar variant of HCC. The fibrolamellar variant of HCC has a more favorable prognosis, occurring in young females without cirrhosis. Invasion of the hepatic vein is common with HCC, and advanced-stage HCC may be complicated by thrombosis in the hepatic vein, vena cava, and portal vein.

Treatment options and prognosis in most HCC patients take into account the stage and etiology of the underlying cirrhosis. Besides the TNM staging system [*see Table 9*], the Child-Pugh classification for staging of cirrhosis should always be considered in the evaluation of treatment options for HCC patients. The Chinese University Prognostic Index (CUPI) assigns points for TNM stage, presence or absence of symptoms and ascites, AFP, total bilirubin, and alkaline phosphatase [*see Table 10*].[157]

TREATMENT

Local, Resectable Disease

Surgical resection Surgery is the only treatment modality with curative intent. Asymptomatic patients with a single HCC tumor, no significant portal hypertension, and preserved liver function are optimal candidates.[158] Ultrasound examination should be performed intraoperatively to detect additional lesions and to guide anatomic resection. Partial hepatectomy is considered safe and carries a perioperative mortality of less than 5% in specialized centers.[159]

The high rate of recurrence of HCC (up to 70% at 5 years) and underlying diseases limit the long-term outcome after surgical resection.[160] Factors associated with poor prognosis are the presence of vascular invasion, multifocal or satellite lesions, tumors larger than 5 cm, and poor histologic differentiation.[161,162]

Preoperative therapies Preoperative (neoadjuvant) therapies may be given to reduce tumor bulk and to eradicate microscopic disease, thus facilitating resection and possibly improving survival. It has been shown that transarterial chemoembolization (TACE) may downstage previously unresectable HCC, but this has not translated into long-term improved survival.[163] In a single trial, biochemotherapy with 5-FU, cisplatin, doxorubicin, and interferon-alfa (PIAF) may convert 25% of unresectable HCC to resectable HCC and may induce complete pathologic remission.[164]

Postoperative therapies Various treatments have been used as adjuvant therapy in HCC, but the benefit of postoperative treatment is uncertain. A potential benefit from adjuvant therapy has been shown by some studies[165,166] but not by others.[167] The development of adjuvant-therapy programs is hampered by the inability to define significantly active regional or systemic therapy for patients with HCC. Recent studies suggest that postoperative interferon-alfa may improve the outcome after resection of HCV-related HCC.[168]

Total hepatectomy and adjuvant chemotherapy Total hepatectomy with orthotopic liver transplantation (OLT) may be the optimal treatment for HCC because it removes the primary tumor and the abnormal liver, both of which reduce life expectancy. OLT results in overall survival of 60% to 70% at 5 years for highly selected patients with HCC, but the recurrence rate is also high (30% in 5 years).[169] Recurrence of the disease is associated with the presence and grade of the vascular invasion, bilobar and multifocal tumor distribution, tumor size greater than 5 cm, positive surgical margins, and male sex.[170] Adjuvant chemotherapy has been associated with increased survival for patients

Table 10 Chinese University Prognostic Index (CUPI)[157]

Characteristic	*Weight*
TNM Stage	
I and II	–3
IIIa and IIIb	–1
IVa and IVb	0
Asymptomatic	–4
Ascites	3
α-Fetoprotein ≥ 500 ng/ml	2
Total bilirubin	
< 34 μmol/L (< 2.0 mg/dl)	0
34–51 μmol/L (2.0–3.0 mg/dl)	3
> 52 μmol/L (> 3.0 mg/dl)	4
Alkaline phosphatase ≥ 200 IU/L	3
Stage	*Score*
Low-risk group	≤ 1
Intermediate-risk group	2–7
High-risk group	> 8

with a high risk of recurrence of HCC, but there have been no randomized trials to prove such a benefit.[171,172]

Advanced, Unresectable Disease

Unfortunately, most HCC patients are not candidates for surgical treatment. Multiple approaches to locoregional nonsurgical therapy have been investigated to control symptoms and improve quality of life. Because the hepatic artery supplies most of the tumor, TACE will devascularize the tumor and induce tumor necrosis. Various agents may be used to embolize the hepatic artery (e.g., gelatin sponge, polyvinyl alcohol foam, collagen, lipiodol, and angiotensin II), and one or more embolization agents are frequently given in conjunction with chemotherapeutic agents.[173]

TACE is contraindicated in patients who have portal hypertension, portal vein occlusion, or decompensated liver cirrhosis. In a single trial, TACE demonstrated a survival benefit over conservative treatment.[174] Hepatic arterial infusion of yttrium-90–labeled glass microspheres (TheraSphere) appears to also be a viable option for unresectable HCC.[175] Radiofrequency ablation (RFA), which is performed with a needle electrode and a high–frequency alternating current, also induces necrosis of tumor cells. RFA is associated with few side effects and can be performed either percutaneously or intraoperatively. The procedure is most successful in the treatment of tumors 3 cm or smaller, and complete necrosis can usually be achieved after two treatments. It has been reported that the success rate of ablation can reach 100%, with 1- and 2-year survival rates in highly selected patients; however, these studies have a short follow-up, and the degree of benefit may be based as much on patient selection as on tumor ablation.[176] One recent study showed that RFA achieved significantly higher disease-free survival than that achieved by percutaneous ethanol injection (PEI).[177] Other local-regional methods include cryosurgery, microwave coagulation therapy or laser-induced thermotherapy, and intra-arterial chemotherapy.

Patients with advanced disease that is not amenable to local treatment may be considered for systemic therapy. Because many patients have underlying chronic liver disease, which may affect the efficacy or toxicity of chemotherapeutic agents, survival may also be influenced by severity of the underlying liver disease. Doxorubicin and 5-FU are the agents most widely used in single-agent chemotherapy regimens, and both have response rates of approximately 20%.[178] Combination chemotherapy may improve the response rate, but response duration is short, and overall survival is not improved.[179] Over the past several years, biologic agents have also been tested. A small prospective, randomized study demonstrated a significant improvement in survival with octreotide, a somatostatin analogue, compared with no treatment.[180] A combination of interferon-alfa with doxorubicin, cisplatin, and 5-FU (PIAF) has shown encouraging results.[164] Benefit has also been suggested in treatment with other agents, including pravastatin, a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase inhibitor[181]; thalidomide, a sedative agent with anti-angiogenic properties that inhibit tumor necrosis factor–α (TNF-α) activity[182]; and vaccination with dendritic cells, which induces cellular immunity.[183]

Biliary Tract Cancer

Biliary tract cancer is the second most common primary hepatobiliary cancer. Biliary tract cancers include gallbladder cancer, extrahepatic bile duct cancer, ampullary cancer, and intrahepatic bile duct cancers. Cholangiocarcinoma was initially used only to describe intrahepatic bile duct cancer, but it now refers to the entire spectrum of tumors arising from the bile ducts, including intrahepatic, perihilar, and distal extrahepatic carcinomas.[184] The designation Klatskin tumor is reserved for perihilar bile duct cancer involving the bifurcation of the hepatic duct.[185]

Gallbladder cancer is relatively rare in the United States, compared with South America, Japan, and Eastern Europe.[186] It occurs more frequently in women than in men, at a ratio of 1.5:1 to 2:1.

ETIOLOGY AND RISK FACTORS

Cholelithiasis has been widely recognized as a risk factor for gallbladder cancer, especially if the gallstones are large and cause symptoms.[187] Inflammation and bacterial infection have also been associated with gallbladder cancer, particularly with *Salmonella typhi*, *Escherichia coli*, and *Helicobactor* species.[186] Other risk factors include advanced age, obesity, and an anomalous pancreaticobiliary ductal junction.[188]

The risk factors of cholangiocarcinoma are less well described than those of gallbladder cancer, but inflammation (e.g., HCV-related cirrhosis and parasitic infestation) appears to be important.[189] Patients with primary sclerosing cholangitis (PSC) have at least a 10% lifetime risk for cholangiocarcinoma.[190] Patients with ulcerative colitis, in the absence of PSC, also have an increased risk for cholangiocarcinoma.

A variety of genetic mutations and factors related to antiapoptosis have been investigated in biliary tract cancers (i.e., K-*ras*, c-*myc*, c-*erb*-B2, *p53*, and *bcl*-2); these may help in designing future chemobiologic therapy for the disease.[188]

DIAGNOSIS

Clinical Manifestations

The most common clinical presentation of gallbladder cancer is right upper quadrant abdominal pain, suggestive of gallstones, but is complicated by the fact that 50% to 100% of patients have concurrent stones.[189] The incidental pathologic finding of gallbladder cancer in patients undergoing cholecystectomy for suspected gallstones is often reported. Other common symptoms are nausea, vomiting, fatty-food intolerance, and anorexia with weight loss. Jaundice is a late sign, often suggesting advanced disease. On the other hand, cholangiocarcinomas rarely present symptomatically before biliary tract obstruction.

Diagnostic Evalulation

Serum carbohydrate antigen 19-9, or CA19-9, is widely used for detecting cholangiocarcinoma, especially for those patients with PSC. However, neither the sensitivity nor the specificity of the test is optimal, and it may be more useful in the assessment of the effect of treatment and the detection of tumor recurrence. Ultrasonography and CT are the most common initial imaging procedures. Cholangiography (either ERCP or percutaneous cholangiography [PTC]) is an important procedure for assessing the resectability of a tumor. MRI and magnetic resonance cholangiopancreatography (MRCP) permit excellent visualization of hepatic parenchymal abnormalities for the evaluation of resectability [*see Figure 9*].

TREATMENT

Surgery

Radical surgical removal is the only modality associated with improved 5-year survival. Unfortunately, most patients present

with advanced disease, and the overall survival is poor. Both location (intrahepatic, extrahepatic, perihilar, or gallbladder) and extent of disease are important for both the choice of surgical procedure and overall outcome. Liver transplantation may achieve long-term survival for selected patients with early-stage intrahepatic cholangiocarcinoma and possibly for carefully selected patients with unresectable hilar cholangiocarcinoma.[191]

Palliative symptomatic control of pain and jaundice with plastic or metal stents, narcotics, and celiac plexus block may improve quality of life.

Adjuvant Therapy

Although there are no randomized studies demonstrating a survival benefit for adjuvant treatment for biliary tract cancer, radiation (external or transcatheter brachytherapy) in combination with chemotherapy is frequently utilized, and survival benefit is suggested by the findings of nonrandomized studies.[192,193] Significant palliation and occasional long-term survival can be achieved with combined-modality treatment with chemoradiation for locally advanced unresectable biliary tract cancers. Many chemotherapeutic agents, such as 5-FU, mitomycin, methotrexate, etoposide, and cisplatin, have been tested for treatment of biliary tract cancer and have resulted in objective response rates of 10% to 20%. Other potential agents include gemcitabine, irinotecan, docetaxel, and oral fluoropyrimidine (capecitabine).[194-197]

Anal Cancer

Anal cancer is a rare malignancy and accounts for only 1% to 2% of all large bowel cancers. The majority of anal cancers are SCCs. Other histologic types include transitional cell (cloacogenic) carcinoma and mucinous adenocarcinoma.

ETIOLOGY AND RISK FACTORS

Environmental factors are implicated in the carcinogenesis of anal cancer, and human papillomavirus (HPV) infection and condylomata acuminata have been identified as preneoplastic lesions.[198] HIV is also associated with both anal cancer and its precursor lesion, anal squamous intraepithelial lesions (SILs), both of which are associated with immunosuppression and anal-receptive intercourse.[199,200] The relative risk for development of anal cancer is 37-fold higher for HIV-positive men than for the general population, and HIV-positive men who have had anal-receptive intercourse have a 60-fold higher risk of developing anal cancer.

Diseases that cause chronic anal irritation are risk factors for anal cancer and include anal fistulae, fissures, chronic local inflammation, chronic hemorrhoids, Crohn disease, and history of carcinoma of the cervix and the vulva.

Previous radiation exposure and immunosuppression (e.g., immunosuppression in preparation for organ transplantation) may increase the incidence of anal cancer. In the general population, anal cancers develop more often in women than in men (ratio, 2:1) and are more common in middle-aged persons. There are no known racial, dietary, or genetic risk factors.

Regular surveillance with sigmoidoscopy or anoscopy may be indicated for high-risk groups.

DIAGNOSIS

Anal cancers usually present with local symptoms (i.e., bleeding, perianal pruritus, pain, and the sensation of a perianal mass), and the diagnosis of the disease is delayed in most cases because the presentation is mistaken for more common benign conditions (e.g., hemorrhoids). An incisional biopsy is preferred to confirm the diagnosis. Inguinal lymph nodes should be examined because they are frequently involved at early stages of the disease. Endoanal ultrasonography may accurately determine the depth of infiltration of cancer into the sphincter muscle.[201]

TREATMENT

Combined-modality chemoradiation is the treatment of choice for anal cancer, except for very small, localized, superficial tumors (carcinoma in situ or tumor < 2 cm). Multiagent chemotherapy (5-FU with mitomycin-C) used concurrently with radiation for SCC of the anal canal was initially administered as preoperative therapy in patients selected for abdominoperineal resection (APR). The sensitivity of the tumors was such that many patients were found to have complete pathologic respons-

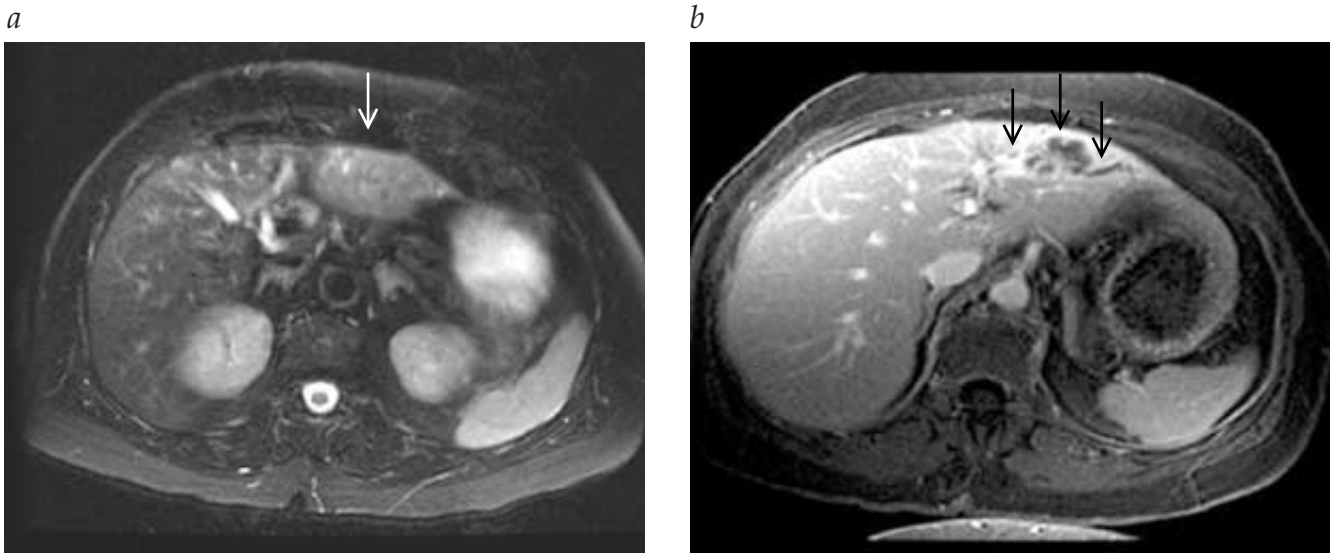


Figure 9 **MRI views of intrahepatic bile duct carcinoma (cholangiocarcinoma). MRI images (T_1- and T_2-weighted) show (*a*) an infiltrative mass (white arrow) of the left lobe of the liver and (*b*) resulting peripheral duct dilatation (black arrows).**

es with treatment, and surgery as part of primary therapy was deleted, resulting in equivalent, if not superior, local control and survival compared with radical surgery.[202] Chemoradiation therapy produces 5-year survival of up to 85% and results in sphincter preservation in in up to 70% of cases. Trials have demonstrated the superiority of chemoradiation over radiation alone and the benefit of mitomycin-C added to infusional 5-FU during radiation.[203] A trial is exploring whether cisplatin is superior to mitomycin-C in this setting. APR may still play an important role in the salvage of treatment failures, as well as provide benefit to patients with preexisting incontinence or patients who are unlikely to tolerate chemoradiation treatment.[204] Recurrences of the disease after initial chemoradiation or surgery are most often local or in inguinal nodes and may be treated with additional surgery, radiation, or chemotherapy, with an average 5-year survival of 40% to 50%.

Other Gastrointestinal Cancers

GASTROINTESTINAL STROMAL TUMOR

Gastrointestinal stromal tumor (GIST) has received tremendous attention in recent years because of successful treatment of unresectable and advanced GISTs with imatinib mesylate (STI571, signal transduction inhibitor), which has proved the concept that selective inhibition of aberrant signal transduction can provide anticancer activity if signaling pathways are identified and blocked.[205]

GISTs are mesenchymal neoplasms of the gastrointestinal tract, related to the intestinal cell of Cajal, with similar differentiation markers.[206,207] The frequency of malignant GIST is estimated to be 20% to 30% of all gastrointestinal soft tissue tumors [*see 12:XIII Sarcoma of Soft Tissue and Bone*], but small benign GISTs are often found incidentally during unrelated surgery or autopsy.[208] GISTs have been documented in all parts of the GI tract, most often in the stomach (60% to 70%) and small intestine (25% to 35%).

Diagnosis

The most common symptom at presentation is gastrointestinal bleeding. The prediction of the malignant potential and prognosis of GISTs is mainly based on histologic characteristics, and it is often very difficult. Tumors that have mitotic activity counts exceeding 5 per 50 high-power fields (HPF) or those larger than 5 cm have a high frequency of intra-abdominal recurrence and liver metastasis. In contrast, tumors smaller than 2 cm and those with mitotic activity counts less than 5 per 50 HPF are likely to be benign,[209] leading to controversy in distinguishing benign from malignant tumors.

Treatment

GISTs express the cell-surface transmembrane receptor KIT, a protein production of the *c-KIT* proto-oncogene that has tyrosine kinase activity.[210,211] KIT is universally phosphorylated in GISTs. A high frequency of mutations leads to constitutive activation of KIT tyrosine kinase without stimulation from its ligand, which subsequently causes uncontrolled stimulation of downstream signaling cascades with aberrant cellular proliferation and resistance to apoptosis. Imatinib mesylate is a selective inhibitor of certain protein tyrosine kinases that was initially designed to target the chimeric BCR-ABL fusion protein, which is responsible for the pathogenesis of chronic myelogenous leukemia (CML).[212] Subsequently, it was shown that imatinib blocks KIT tyrosine kinase as well, and it arrests proliferation and induces apoptotic cell death of GIST cells.[212-214]

A multicenter clinical trial has demonstrated that imatinib is safe and effective in the treatment of patients with advanced GISTs. More than half of patients with advanced unresectable or metastatic GISTs achieve sustained objective response.[205] PET has been shown to be a useful method for the assessment of tumor response to treatment, as compared with more traditional imaging studies (i.e., CT or MRI).[215]

GASTRIC LYMPHOMA

Epidemiology

Primary extranodal lymphoma constitutes 25% of non-Hodgkin lymphoma (NHL) cases in North America.[216] The most common site of extranodal NHL is the stomach, which represents approximately 24% of all primary extranodal lymphoma, according to the End Results Groups Cancer Registries in the United States. Compared with carcinoma incidence, gastric lymphoma is rare, representing 2% to 8% of all gastric malignancies in the United States, but its incidence is increasing.[217] Most patients with gastric lymphoma are older than 50 years. Men are affected more often than women, and it is relatively more common in whites than blacks.

Etiology and Risk Factors

New insights into gastric lymphomas and their etiology and pathogenesis continue to be gained. The multistep pathogenesis by which chronic inflammation of *H. pylori* gastritis converts to low-grade gastric mucosa–associated lymphoid tissue (MALT) lymphoma and progresses to a tumor of higher stage and grade is characterized by multiple molecular biologic events.[218]

Diagnosis

The initial symptom is vague, nonspecific upper gastric discomfort, resembling peptic ulcer disease or gastritis, for many months before diagnosis.[219] The most common complaints are epigastric pain, weight loss, and nausea and vomiting.[220] Unlike in nodal lymphoma, night sweats are not a common feature. Occasionally, an abdominal mass is palpable. Lymphadenopathy is rare, and patients often have no physical signs. Gastric perforation, bleeding, and obstruction are uncommon.

The most common endoscopic finding is a diffuse infiltrative process with thick, rigid folds; and multiple biopsies are often needed to improve the diagnostic accuracy, because the primary lesion is submucosal.[221] Abdominal CT scans may show the extent of the lesion and rule out other metastatic disease, but they cannot reliably differentiate between metastatic lymphadenopathy and reactive lymphoid hyperplasia. Endoscopic ultrasonography is particularly important and considered mandatory in detection of the depth of invasion and the presence of perigastric lymphadenopathy.[222]

Staging and Prognosis

The most frequent site of gastric lymphoma is the distal portion of the stomach, as is the epidemiologically related adenocarcinoma of the stomach. The primary lesion is submucosal, originating from the lymphoid tissue in the lamina propria. It usually invades outwardly through the serosa, and the mucosa is involved later in the disease process. Lymph node involvement usually precedes distant metastases.

The prognosis of patients with gastric lymphoma is better overall than that of patients with gastric carcinoma and is based on

grade, depth of invasion in the gastric wall, and lymph node involvement.[223] Younger patients may have a better prognosis than older patients at the same stage. T cell lymphoma is less common but more aggressive than its B cell counterpart. Other poor prognostic factors include aneuploid tumors, a high proliferation index by antibody Ki-67, and histiocytic histologic subtype. Lymph node involvement usually precedes distant metastases.

Treatment

Eradication of *H. pylori* is considered to be first-line treatment in patients with low-grade superficial gastric MALT lymphoma, and the benefit of treatment may be sustained[224]; however, there is still a lack of long-term follow-up. Surgical resection, radiotherapy, chemotherapy, and the combination of these modalities have proved to be effective in the management of gastric lymphoma, but controversy remains as to whether operative or conservative therapeutic strategies should be favored as the primary treatment. Large prospective, randomized studies are needed to more accurately define the optimal therapeutic options in gastric lymphoma.

Weijing Sun, M.D., has served as an advisor for Roche Pharmaceuticals, Pfizer, Inc., and Aventis Pharmaceuticals and has received grants for clinical research from Pfizer, Inc.

Daniel Haller, M.D., has received grants for clinical research and has served as advisor or consultant to Sanofi Pharmaceuticals, Inc., Roche Pharmaceuticals, Pfizer, Inc., Aventis Pharmaceuticals, and Eli Lilly and Company.

References

1. Jemal A, Murray T, Samuels A, et al: Cancer statistics, 2003. CA Cancer J Clin 53:5, 2003

2. Blot WJ, Devesa SS, Kneller RW, et al: Rising incidence of adenocarcinoma of the esophagus and gastric cardia. JAMA 265:1287, 1991

3. Silber W: Carcinoma of the oesophagus: aspects of epidemiology and aetiology. Proc Nutr Soc 44:101,1985

4. Devesa SS, Blot WJ, Fraumeni JF Jr: Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. Cancer 83:2049, 1998

5. Daly JM, Karnell LH, Menck HR: National Cancer Data Base report on esophageal carcinoma. Cancer 78:1820, 1996

6. Bytzer P, Christensen PB, Damkier P, et al: Adenocarcinoma of the esophagus and Barrett's esophagus: a population-based study. Am J Gastroenterol 94:86, 1998

7. Castellsague X, Munoz N, De Stefani E, et al: Independent and joint effects of tobacco smoking and alcohol drinking on the risk of esophageal cancer in men and women. Int J Cancer 82:657, 1999

8. Zambon P, Talamini R, La Vecchia C, et al: Smoking, type of alcoholic beverage and squamous-cell oesophageal cancer in northern Italy. Int J Cancer 86:144, 2000

9. Bosetti C, Franceschi S, Levi F, et al: Smoking and drinking cessation and the risk of oesophageal cancer. Br J Cancer 83:689, 2000

10. Pera M, Pera M: Recent changes in the epidemiology of esophageal cancer. Surg Oncol 10:81, 2001

11. Brown LM, Swanson CA, Gridley G: Dietary factors and the risk of squamous cell esophageal cancer among black and white men in the United States. Cancer Causes Control 9:467, 1998

12. Thurnham DI, Zheng SF, Munoz N, et al: Comparison of riboflavin, vitamin A, and zinc status of Chinese populations at high and low risk for esophageal cancer. Nutr Cancer 7:131, 1985

13. Sandler RS, Nyren O, Ekbom A, et al: The risk of esophageal cancer in patients with achalasia: a population-based study. JAMA 274:1359, 1995

14. Leon X, Quer M, Diez S, et al: Second neoplasm in patients with head and neck cancer. Head Neck 21:204, 1999

15. Levi R, Randimbison L, Te VC, et al: Second primary cancers in patients with lung carcinoma. Cancer 86:186, 1999

16. Lagergren J, Bergstrom R, Lindgren A, et al: Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med 340:825, 1999

17. Winters C Jr, Spurling TJ, Chobanian SJ: Barrett's esophagus: a prevalent, occult complication of gastroesophageal reflux disease. Gastroenterology 92:118, 1987

18. Hamilton SR, Smith RR: The relationship between columnar epithelial dysplasia and invasive adenocarcinoma arising in Barrett's esophagus. Am J Clin Pathol 87:301, 1987

19. O'Connor JB, Falk GW, Richter JE: The incidence of adenocarcinoma and dysplasia in Barrett's esophagus: report on the Cleveland Clinic Barrett's Esophagus Registry. Am J Gastroenterol 94:2037, 1999

20. Shaheen NJ, Crosby MA, Bozymski EM, et al: Is there publication bias in the reporting of cancer risk in Barrett's esophagus? Gastroenterology 119:333, 2000

21. Reid BJ, Levine DS, Longton GL, et al: Predictors of progression to cancer in Barrett's esophagus: baseline histology and flow cytometry identify low- and high-risk patient subsets. Am J Gastroenterol 95:1669, 2000

22. Lagergren J, Bergstrom R, Nyren O: Association between body mass and adenocarcinoma of the esophagus and gastric cardia. Ann Intern Med 130:883, 1999

23. Huang GJ, Shao LF, Zhang DW, et al: Diagnosis and surgical treatment of early esophageal carcinoma. Chin Med J 94:229, 1981

24. Aste H, Saccomanno S, Munizzi F: Blind pan-esophageal brush cytology: diagnostic accuracy. Endoscopy 16:165, 1984

25. Sharma P, Sampliner RE, Camargo E: Normalization of esophageal pH with high-dose proton pump inhibitor therapy does not result in regression of Barrett's esophagus. Am J Gastroenterol 92:582, 1997

26. McDonald ML, Trastek VF, Allen MS, et al: Barrett's esophagus: does an antireflux procedure reduce the need for endoscopic surveillance? J Thorac Cardiovasc Surg 111:1135, 1996

27. McCarthy M, Wilkinson ML: Treatment of Barrett's esophagus by endoscopic laser ablation and antireflux surgery. Gastrointest Endosc 49:129, 1999

28. Schulz H, Miehlke S, Antos D, et al: Ablation of Barrett's epithelium by endoscopic argon plasma coagulation in combination with high-dose omeprazole. Gastrointest Endosc 51:659, 2000

29. Sampliner RE, Faigel D, Fennerty MB, et al: Effective and safe endoscopic reversal of nondysplastic Barrett's esophagus with thermal electrocoagulation combined with high-dose acid inhibition: a multi-center study. Gastrointest Endosc 53:554, 2001

30. Panjehpour M, Overholt BF, Haydek JM, et al: Results of photodynamic therapy for ablation of dysplasia and early cancer in Barrett's esophagus and effect of oral steroids on stricture formation. Am J Gastroenterol 95:2177, 2000

31. Van Laethem JL, Peny MO, Salmon I, et al: Intramucosal adenocarcinoma arising under squamous re-epithelialisation of Barrett's oesophagus. Gut 46:574, 2000

32. van den Boogert J, van Hillegersberg R, Siersema PD, et al: Endoscopic ablation therapy for Barrett's esophagus with high-grade dysplasia: a review. Am J Gastroenterol 94:1153, 1999

33. Sampliner RE: Practice guidelines on the diagnosis, surveillance, and therapy of Barrett's esophagus. The Practice Parameters Committee of the American College of Gastroenterology. Am J Gastroenterol 93:1028, 1998

34. Cooper GS, Yuan Z, Chak A, et al: Association of prediagnosis endoscopy with stage and survival in adenocarcinoma of the esophagus and gastric cardia. Cancer 95:32, 2002

35. Meltzer SJ: The molecular biology of esophageal carcinoma. Recent Results Cancer Res 142:1, 1996

36. Yacoub L, Goldman H, Odze RD: Transforming growth factor-alpha, epidermal growth factor receptor, and MiB-1 expression in Barrett's-associated neoplasia: correlation with prognosis. Mod Pathol 10:105, 1997

37. Block MI, Patterson GA, Sundaresan RS, et al: Improvement in staging of esophageal cancer with the addition of positron emission tomography. Ann Thorac Surg 64:770, 1997

38. Pompili MF, Marks JB: The history of surgery for carcinoma of the esophagus. Chest Surg Clin N Am 10:145, 2000

39. Law S, Wong J: What is appropriate treatment for carcinoma of the thoracic esophagus? World J Surg 25:189, 2001

40. Clark P: Surgical resection with or without pre-operative chemotherapy in oesophageal cancer: an updated analysis of a randomized controlled trial conducted by the UK Medical Research Council upper GI tract cancer group. Pro Am Clin Oncol 20:A502, 2001

41. Kelsen DP, Ginsberg R, Pajak TF, et al: Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. N Engl J Med 339:1979, 1998

42. Walsh TN, Noonan N, Hollywood D, et al: A comparison of multimodal therapy and surgery for esophageal adenocarcinoma. N Engl J Med 335:462, 1996

43. Bosset JF, Gignoux M, Triboulet JP, et al: Chemoradiotherapy followed by surgery compared with surgery alone in squamous cell cancer of the esophagus. N Engl J Med 337:161, 1997

44. Urba SG, Orringer MB, Turrisi A, et al: Randomized trial of preoperative chemoradiation versus surgery alone in patients with locoregional esophageal carcinoma. J Clin Oncol 19:305, 2001

45. Macdonald JS, Smalley SR, Benedetti J, et al: Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. N Engl J Med 345:725, 2001

46. Herskovic A, Martz K, Al-Sarraf M, et al: Combined chemotherapy and radiotherapy compared with radiotherapy alone in patients with cancer of the esophagus. N Engl J Med 326:1593, 1992

47. Cooper JS, Guo MD, Herskovic A, et al: Chemoradiotherapy of locally advanced esophageal cancer: long-term follow-up of a prospective randomized trial (RTOG 85-01). JAMA 281:1623, 1999

48. Hishikawa Y, Kurisu K, Tanigushi M, et al: High-dose-rate intraluminal brachytherapy (HDRIBT) for esophageal cancer. Int J Radiat Oncol Biol Phys 21:1133, 1991

49. Bleiberg H, Conroy T, Paillot B, et al: Randomized phase II trial of cisplatin 5-fluorouracil versus cisplatin alone in advanced squamous cell oesophageal cancer. Eur J Cancer 33:1216, 1997

50. Ilson DH, Saltz L, Enzinger P, et al: Phase II trial of weekly irion tecan plus cisplatin in advanced esophageal cancer. J Clin Oncol 17:3270, 1999

51. Ilson DH, Ajani J, Bhalla K, et al: Phase II trial of paclitaxel, fluorouracil, and cisplatin in patients with advanced carcinoma of the esophagus. J Clin Oncol 16:1826, 1998

52. Mauer AM, Kraut EH, Rudin CM, et al: Phase II study of oxaliplatin, fluorouracil, and leucovorin in metastatic carcinoma of the esophagus/gastric cardia. Pro Am Clin Oncol 20:A650, 2001

53. Knyrim K, Wagner HJ, Bethge N, et al: A controlled trial of an expansile metal stent for palliation of esophageal obstruction due to inoperable cancer. N Engl J Med 329:1302, 1993

54. Raijman I, Siddique I Ajani J, et al: Palliation of malignant dysphagia and fistulae with coated expandable metal stents: experience with 101 patients. Gastrointest Endosc 48:172, 1998

55. Boyce HW: Palliation of dysphagia of esophageal cancer by endoscopic lumen restoration techniques. Cancer Control 6:73, 1999

56. Parkin DM, Pisani P, Ferlay J: Global cancer statistics. CA Cancer J Clin 49:33, 1999

57. Lambert R, Guilloux A, Oshima A, et al: Incidence and mortality from stomach cancer in Japan, Slovenia and the USA. Int J Cancer 97:811, 2002

58. Corella D, Guillen M: Dietary habits and epidemiology of gastric cancer. Hepatogastroenterology 48:1537, 2001

59. Forman D, Newell DG, Fullerton F, et al: Association between infection with *Helicobacter pylori* and risk of gastric cancer: evidence from a prospective investigation. BMJ 302:1302, 1991

60. Nomura A, Stemmermann GN, Chyou PH, et al: Helicobacter pylori infection and gastric carcinoma among Japanese Americans in Hawaii. N Engl J Med 325:1132, 1991

61. Parsonnet J, Friedman GD, Vandersteen DP, et al: Helicobacter pylori infection and the risk of gastric carcinoma. N Engl J Med 325:1127, 1991

62. Stadtlander CT, Waterbor JW: Molecular epidemiology, pathology and prevention of gastric cancer. Carcinogenesis 20:2195, 1999

63. Brown L: Helicobacter pylori: epidemiology and routes of transmission. Epidemiol Rev 22:283, 2000

64. Parsonnet J, Friedman G, Orentreich N, et al: Risk for gastric cancer in people with Cag A positive or Cag A negative Helicobacter pylori infection. Gut 40:297, 1997

65. Graham D: Helicobacter pylori infection is the primary cause of gastric cancer. J Gastroenterol 35:90, 2000

66. Newman E, Brennan MF, Hochwald SN, et al: Gastric remnant carcinoma: just another proximal gastric cancer or a unique entity? Am J Surg 173:292,1997

67. Weitzel JN: Genetic cancer risk assessment: putting it all together. Cancer 86:2483, 1999

68. Caldas C, Carneiro F, Lynch HT, et al: Familial gastric cancer: overview and guidelines for management. J Med Genet 36:873, 1999

69. Huntsman DG, Carneiro F, Lewis FL, et al: Early gastric cancer in young, asymptomatic carriers of germ-line E-cadherin mutations. N Engl J Med 344:1904, 2001

70. Guilford P, Hopkins J, Harraway J, et al: E-cadherin germline mutations in familial gastric cancer. Nature 392:402, 1998

71. Lindor NM, Green MH: The concise handbook of family cancer syndromes. J Natl Cancer Inst 90:1039, 1998

72. El-Rifai W, Powell SM.: Molecular biology of gastric cancer. Semin Radiat Oncol 12:128, 2002

73. Lauren PA, Nevalainen TJ: Epidemiology of intestinal and diffuse types of gastric carcinoma: a time-trend study in Finland with comparison between studies from high- and low-risk areas. Cancer 71:2926, 1993

74. American Joint Committee on Cancer Staging Manual, 6th ed. Greene FL, Page DL, Balch CM, et al, Eds. Springer-Verlag Incorporated, New York, 2002

75. Miwa K: Evalation of the TNM classification of stomach cancer and proposal for its rational stage grouping. Jpn J Clin Oncol 14:385, 1984

76. Hundahl SA, Menck HR, Mansour EG, et al: The National Cancer Data Base report on gastric carcinoma. Cancer 80:2333, 1997

77. Siewert JR, Bottcher K, Stein HJ, et al: Problem of proximal third gastric carcinoma. World J Surg 19:523, 1995

78. Hundahl SA, Phillips JL, Menck HR: The National Cancer Data Base Report on poor survival of U.S. gastric carcinoma patients treated with gastrectomy: fifth edition American Joint Committee on Cancer staging, proximal disease, and the "different disease" hypothesis. Cancer 88:921, 2000

79. Chen CH, Yang CC, Yeh YH: Preoperative staging of gastric cancer by endoscopic ultrasound: the prognostic usefulness of ascites detected by endoscopic ultrasound. J Clin Gastroenterol 35:321, 2002

80. Brennan MF, Karpeh MS: Surgery for gastric cancer: the American view. Semin Oncol 23:352, 1996

81. Sasako M: What is reasonable treatment for gastric adenocarcinoma? J Gastroenterol 35:116, 2000

82. Bonenkamp JJ, Songun I, Hermans J, et al: Randomized comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. Lancet 345:745, 1995

83. Roukos DH: Current status and future perspectives in gastric cancer management. Cancer Treat Rev 26:243, 2000

84. Bonenkamp JJ, Hermans J, Sasako M, et al: Extended lymph-node dissection for gastric cancer. N Engl J Med 340:908, 1999

85. Cuschieri A, Weeden S, Fielding J, et al: Patients survival after D1 and D2 resections for gastric cancer: long-term results of the MRC randomized surgical trial. Br J Cancer 79:1522, 1999

86. Earle CC, Maroun JA: Adjuvant chemotherapy after curative resection for gastric cancer in non-Asian patients: revisiting a meta-analysis of randomized trials. Eur J Cancer 35:1059, 1999

87. Mari E, Floriani I, Tinazzi A, et al: Efficacy of adjuvant chemotherapy after curative resection for gastric cancer: a meta-analysis of published randomized trials. A study of the GISCAD. Ann Oncol 11:837, 2000

88. Hallissey MT, Dunn JA, Ward LC, et al: The second British Stomach Cancer Group trial of adjuvant radiotherapy or chemotherapy in resectable gastric cancer: five-year follow-up. Lancet 343:1309, 1994

89. Gunderson LL, Sosin H: Adenocarcinoma of the stomach: areas of failure in a re-operation series (second or symptomatic look) clinicopathologic correlation and implications for adjuvant therapy. Int J Radiat Oncol Biol Phys 8:1, 1982

90. Lowy AM, Mansfield PF, Leach SD, et al: Response to neoadjuvant chemotherapy best predicts survival after curative resection of gastric cancer. Ann Surg 229: 303, 1999

91. Stein HJ, Sendler A, Fink U, et al: Multidisciplinary approach to esophageal and gastric cancer. Surg Clin North Am 80:659, 2000

92. Webb A, Cunningham D, Scarffe JH, et al: Randomized trial comparing epirubucin, cisplatin, and fluorouracil versus fluorouracil, doxorubicin, and methotrexate in advanced esophagogastric cancer. J Clin Oncol 15:261, 1997

93. Waters JS, Norman A, Cunningham D, et al: Long-term survival after epirubicin, cisplatin and fluorouracil for gastric cancer: results of a randomized trial. Br J Cancer 80:269, 1999

94. Ajani J, Baker J, Pisters P, et al: Irinotecan plus cisplatin in advanced gastric or gastroesophageal junction carcinoma. Oncology (Huntingt) 3:52, 2001

95. Louvet C, Andre T, Tigaud JM, et al: Phase II study of oxaliplatin, fluorouracil, and folinic acid in locally advanced or metastatic gastric cancer patients. J Clin Oncol 20:4543, 2002

96. Van Cutsem E, Ajani J, Tjulandin S, et al: Docetaxel (T) in combination with cisplatin (C) with or without 5-fluorouracil (F) in patients (pts) with advanced gastric or GE junction adenocarcinoma (AGC): preliminary results. Ann Oncol 11:63a, 2000

97. Witkamp AJ, de Bree E, Van Goethem AR, et al: rationale and techniques of intraoperative hyperthermic intraperitoneal chemotherapy. Cancer Treat Rev 27:365, 2001

98. Howe HL, Wingo PA, Thun MJ, et al: Annual report to the nation on the status of cancer (1973-1998) featuring cancers with recent increasing trends. J Natl Cancer Inst 93:1656, 2001

99. Silverman DT, Schiffman M Everhart J, et al: Diabetes mellitus, other medical conditions and familial history of cancer as risk factors for pancreatic cancer. Br J Cancer 80:1830, 1999

100. Hjalgrim H, Frisch M, Ekbom A, et al: Cancer and diabetes: a follow-up study of two population-based cohorts of diabetic patients. J Intern Med 241:471, 1997

101. Silverman DT, Dunn JA, Hoover RN, et al: Cigarette smoking and pancreas cancer: a case-control study based on direct interviews. J Natl Cancer Inst 86:1510, 1994

102. Bueno de Mesquita HB, Maisonneuve P, Moerman CJ, et al: Life-time history of smoking and exocrine carcinoma of the pancreas: a population-based case-control study in the Netherlands. Int J Cancer 49:816, 1991

103. Harnack LJ, Anderson KE, Zheng W, et al: Smoking, alcohol, coffee, and tea intake and incidence of cancer of the exocrine pancreas. The Iowa Women's Health Study. Cancer Epidemiol Biomarkers Prev 6:1081, 1997

104. Baghurst PA, McMichael AJ, Slavotinek AH, et al: A case-control study of diet and cancer of the pancreas. Am J Epidemiol 134:167, 1991

105. Stolzenberg-Solomon RZ, Pirtnen P, Barrett MJ, et al: Dietary and other methyl-group availability factors and pancreatic cancer risk in a cohort of male smokers. Am J Epidemiol 153:680, 2001

106. Lowenfels AB, Maisonneuve P: Epimemiology and etiologic factors of pancreatic cancer. Hematol Oncol Clin N Am 16:1, 2002

107. Anderson KE, Johnson TW, Lazovich D, et al: Association between nonsteroidal anti-inflammatory drug use and the incidence of pancreatic cancer. J Natl Cancer Inst 94:1168, 2002

108. Tersmette AC, Petersen GM, Offerhaus GJ, et al: Increased risk of incident pancreatic cancer among first-degree relatives of patients with familial pancreatic cancer. Clin Cancer Res 7:738, 2001

109. Hruban RH, Petersen GM, Goggins M, et al: Familial pancreatic cancer. Ann Oncol 10:69, 1999

110. Ritts RE, Pitt HA: CA19-9 in pancreatic cancer. Surg Oncol Clin N Am 18:622, 1997

111. Albert MB, Steinberg WM, Henry JP: Elevated serum levels of tumor marker CA19-9 in acute cholangitis. Dig Dis Sci 33:1223, 1988

112. Furuya N, Kawa S, Hasebe O, et al: Comparative study of CA242 and CA19-9 in chronic pancreatitis. Br J Cancer 73:372, 1996

113. Berthelemy P, Bouisson M, Escourrou J, et al: Identification of K-ras mutations in pancreatic juice in the early diagnosis of pancreatic cancer. Ann Intern Med 123:188, 1995

114. Tada M, Omata M, Kawai S, et al: Detection of ras gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. Cancer Res 53:2472, 1993

115. Caldas C, Hahn SA, Hruban RH, et al: Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. Cancer Res 54:3568, 1994

116. Konner J, O'Reilly E: Pancreatic cancer: epidemiology, genetics, and approaches to screening. Oncology (Huntingt) 12:1615, 2002

117. Hruban RH, Adsay NV, Albores-Saavedra J, et al: Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for pancreatic duct lesions. Am J Surg Pathol 25:579, 2001

118. Hruban RH, Goggins M, Parsons J, et al: Progression model for pancreatic cancer. Clin Cancer Res 6:2969, 2000

119. Metz HR, Sechopoulos P, Delbeke D, et al: EUS, PET, and CT scanning for evaluation of pancreatic adenocarcinoma. Gastrointest Endosc 52:367, 2000

120. Snady H: The role of endoscopic ultrasonography in diagnosis, staging, and outcomes of gastrointestinal diseases. Gastroenterologist 2:91, 1994

121. Nix G, Schmitz P, Wilson J, et al: Carcinoma of the head of the pancreas: therapeutic implications of endoscopic retrograde cholangiopancreatography findings. Gastroenterology 87:37, 1984

122. Lieberman MD, Kilburn H, Lindsey M, et al: Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy. Ann Surg 222:638, 1995

123. Schafer M, Mullhaupt B, Clavien PA: Evidence-based pancreatic head resection for pancreatic cancer and chronic pancreatitis. Ann Surg 236:137, 2002

124. Yeo CJ, Cameron JL: Improving results of pancreaticoduodenectomy for pancreatic cancer. World J Surg 23:907, 1999

125. Further evidence of effective adjuvant combined radiation and chemotherapy following curative resection of pancreatic cancer. Cancer 59:2006, 1987

126. Yeo CJ, Abrams RA, Grochow LB, et al: Pancreaticoduodenectomy for pancreatic adenocarcinoma: postoperative adjuvant chemoradiation improves survival: a prospective, single institution experience. Ann Surg 225:621, 1997

127. Neoptolemos JP, Dunn JA, Stocken DD, et al: Adjuvant chemoradiation and chemotherapy in resectable pancreatic cancer: a randomized controlled trial. Lancet 358:1576, 2001

128. Klaassen DJ, MacIntyre JM, Catton GE, et al: Treatment of locally unresectable cancer of the stomach and pancreas: a randomized comparing 5-fluorouracil alone with radiation plus concurrent and maintenance 5-fluorouracil: an Eastern Cooperative Oncology Group study. J Clin Oncol 3:373, 1985

129. Mohiuddin M, Regine WF, Stevens J, et al: Combined intraoperative radiation and perioperative chemotherapy for unresectable pancreatic cancers of the pancreas. J Clin Oncol 13:2764, 1995

130. Wolff RA, Evans DB, Gravel DM, et al: Phase I trial of gemcitabine combined with radiation for the treatment of locally advanced pancreatic adenocarcinoma. Clin Cancer Res 7:2246, 2001

131. Burris HA 3rd, Moore MJ, Andersen J, et al: Improvements in survival and clinic benefit with gemcitabine as first-line therapy for patients with advanced pancreatic cancer: a randomized trial. J Clin Oncol 15:2403, 1997

132. Rocha Lima CMS, Savarese D, Bruckner H, et al: Irinotecan plus gemcitabine induces both radiographic and CA19-9 tumor marker responses in patients with previously untreated advanced pancreatic cancer. J Clin Oncol 20:1182, 2002

133. Abbruzzese JL, Rosenberg A, Xiong Q, et al: Phase II study of anti-epidermal growth factor receptor (EGFR) antibody cetuximab (IMC-C225) in combination with gemcitabine in patients with advanced pancreatic cancer. Pro Am Soc Clin Oncol 20:130a, 2001

134. Parkin DM, Bray F, Fealay J, et al. Estimating the world cancer burden: GLOBOCAN 2000. Int J Cancer 94:153, 2001

135. Bruix J, Sherman M, Llovet JM, et al: Clinical management of hepatocellular carcinoma: conclusions of Barcelona-2000 EASL conference. J Hepatol 35:421, 2001

136. Monto A, Wright T: The epidemiology and prevention of hepatocellular carcinoma. Semin Oncol 28:441, 2001

137. El-Serag HB, Mason AC: Rising incidence of hepatocellular carcinoma in the United States. N Engl J Med 340:745, 1999

138. Cottone M, Turri M, Caltagirone M, et al: Screening for hepatocellular carcinoma in patients with Child's A cirrhosis: an 8-year prospective study by ultrasound and alphafetoprotein. J Hepatol 21:1029, 1994

139. Fattovich G, Giustina G, Degos F, et al: Morbidity and mortality in compensated cirrhosis type C: a retrospective follow-up study of 384 patients. Gastroenterology 112:463, 1997

140. El-Serag HB: Hepatocellular carcinoma: an epidemiologic view. J Clin Gastroenterol 35:S72, 2002

141. El-Saadany S, Tepper M, Mao Y, et al: An epidemiologic study of hepatocellular carcinoma in Canada. Can J Public Health 93:443, 2002

142. Yoshizawa H: Hepatocellular carcinoma associated with hepatitis C virus infection in Japan: projection to other countries in the foreseeable future. Oncology 62:8, 2002

143. Deugnier YM, Guyader D, Crantock L, et al: Primary liver cancer in genetic hemochromatosis: a clinical, pathological, and pathogenetic study of 54 cases. Gastroenterology 104:228, 1993

144. Tazawa J, Maeda M, Nakagawa M, et al: Diabetes mellitus may be associated with hepatocarcinogenesis in patients with chronic hepatitis C. Dig Dis Sci 47:710, 2002

145. Hassan MM, Hwang LY, Hatten CJ, et al: Risk factors for hepatocellular carcinoma: synergism of alcohol with viral hepatitis and diabetes mellitus. Hepatology 36:1206, 2002

146. Esquivel CO, Gutierrez C, Cox KL, et al: Hepatocellular carcinoma and liver cell dysplasia in children with chronic liver disease. J Pediatr Surg 29:1465, 1994

147. Ozturk M: p53 mutation in hepatocellular carcinoma after aflatoxin exposure. Lancet 338:1356, 1991

148. Forman D, Vincent TJ, Doll R: Cancer of the liver and the use of oral contraceptives. Br Med J 292:1357, 1986

149. Farrell GC, Joshua DE, Uren RF, et al: Androgen-induced hepatoma. Lancet 1:430, 1975

150. Strong RW: Transplantation for liver and biliary cancer. Semin Surg Oncol 19:189, 2000

151. Chang MH, Chen CJ, Lai MS, et al: Universal hepatitis B vaccination in Taiwan and the incidence of hepatocellular carcinoma in children. N Engl J Med 336:1855, 1997

152. Monto A, Wright TL: The epidemiology and prevention of HCC. Semin Oncol 28:441, 2001

153. Kasahara A, Hayashi N, Mochizuki K, et al: Risk factors for hepatocellular carcinoma and its incidence after interferon treatment in patients with chronic hepatitis C. Hepatology 27:1394, 1998

154. Shiratori Y, Shiina S, Teratani T, et al: Interferon therapy after tumor ablation improves prognosis in patients with hepatocellular carcinoma associated with hepatitis C virus. Ann Intern Med 138:299, 2003

155. Sherman M: Surveillance for hepatocellular carcinoma. Semin Oncol 28:450, 2001

156. Ikoma J, Kaito M, Ishihara T, et al: Early diagnosis of hepatocellular carcinoma using a sensitive assay for serum des-gamma-carboxy prothrombin: a prospective study. Hepatogastroenterology 49:235, 2002

157. Leung TW, Tang AM, Zee B, et al: Construction of the Chinese University Prognosis Index (CUPI) for hepatocellular carcinoma and comparison with TNM, Okuda and CLIP staging systems: a study based on 926 patients. Cancer 94:1760, 2002

158. Llovet JM, Fuster J, Bruix J: Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. Hepatology 30:1434, 1999

159. Little SA, Fong Y: Hepatocellular carcinoma: current surgical management. Semin Oncol 28:474, 2001

160. Poon RT, Fan ST, Lo CM, et al: Improving survival results after resection of hepatocellular carcinoma: a prospective study of 377 patients over 10 years. Ann Surg 234:63, 2001

161. Poon RT, Ng IO, Fan ST, et al: Clinicopoathologic features of long-term survivors and disease-free survivals after resection of hepatocellular carcinoma: a study of prospective cohort. J Clin Oncol 19: 3037, 2001

162. Arii S, Tanaka J, Yamazoe Y, et al: Predictive factors for intrahepatic recurrence of hepatocellular carcinoma after partial hepatectomy. Cancer 69:913, 1992

163. Harada T, Matsuo K, Inoue T: Is Preoperative hepatic arterial chemoembolization safe and effective for hepatocellular carcinoma? Ann Surg 224:4, 1996

164. Leung TW, Patt YZ, Lau WY, et al: Complete pathological remission is possible with systemic combination chemotherapy for inoperable hepatocellular carcinoma. Clin Cancer Res 5:1676, 1999

165. Muto Y, Moriwaki H, Ninomiya M, et al: Prevention of second primary tumors by an acyclic retinoid, polyprenoid acid, in patients with hepatocellular carcinoma. N Engl J Med 334:1561, 1996

166. Lau WY, Leung TW, Ho SK, et al: Adjuvant intraarterial iodine-131 labelled lipiodol for respectable hepatocellular carcinoma: a prospective randomized trial. Lancet 353:797, 1999

167. Ono T, Yamanoi A, El Assal ON, et al: Adjuvant chemotherapy after resection of hepatocellular carcinoma deterioration of long-term prognosis in cirrhotic patients. Cancer 91:2378, 2001

168. Kubo S, Nishiguchi S, Hirohashi K, et al: Randomized clinical trial of long-term outcome after resection of hepatitis C virus–related hepatocellular carcinoma by postoperative interferon therapy. Br J Surg 89:418, 2002

169. Penn I: Hepatic transplantation for primary and metastatic cancers of the liver. Surgery 110:726, 1991

170. Llovet JM, Fuster J, Bruix J: Intention-to-treat analysis of surgical treatment of early hepatocellular carcinoma: resection versus transplantation. Hepatology 39:1434, 1999

171. Marsh JW, Dvorchik I, Subotin M, et al: The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation: a pilot study. Hepatology 26:444, 1997

172. Olthoff KM, Rosove MH, Shackleton CR, et al: Adjuvant chemotherapy improves survival after liver transplantation for hepatocellular carcinoma. Ann Surg 221:734, 1995

173. Aguayo A, Patt YZ: Nonsurgical treatment of hepatocellular carcinoma. Semin Oncol 28:503, 2001

174. Llovet JM, Real MI, Vilana R, et al: Chemoembolization vs. conservative treatment in patients with unresectable hepatocellular carcinoma. Pro Am Soc Clin Oncol 21:132a, 2002

175. Carr BI, Sheetz M, Brown F, et al: Hepatic arterial 90 Yttrium-labeled glass microspheres (TheraSphere) as treatment for unresectable HCC in forty-three patients. Pro Am Soc Clin Oncol 21:139a, 2002

176. Seidenfeld J Korn A, Aronson N: Radiofrequency ablation of unresectable primary liver cancer. J Am Coll Surg 194: 813, 2002

177. Olschewski M, Lencioni, R, Allgaier H, et al: A randomized comparison of radiofrequency thermal ablation and percutaneous ethanol injection for the treatment of small hepatocellular carcinoma. Pro Am Soc Clin Oncol 20:126a, 2001

178. Lin DY, Lin SM, Liaw YF: Non-surgical treatment of hepatocellular carcinoma. J Gastroenterol Hepatol 12:S319, 1997

179. Simonetti RG, Liberati A, Angiolini C, et al: Treatment of hepatocellular carcinoma: a systematic review of randomized controlled trials. Ann Oncol 8:117, 1997

180. Kouroumalis E, Skordilis P, Thermos K, et al: Treatment of hepatocellular carcinoma with octreotide: a randomized controlled study. Gut 42:442, 1998

181. Kawata S, Yamasaki E, Nagase T, et al: Effect of pravastatin on survival in patients with advanced hepatocellular carcinoma: a randomized control trial. Br J Cancer 84: 886, 2001

182. Kong HL, Boyer MJ, Lim R, et al: Phase II trial of thalidomide in unresectable hepatocellular carcinoma: a Cancer Therapeutics Research Group Study. Pro Am Soc Clin Oncol 20:A2282, 2001

183. Midgley R, Kerr DJ, Mirza N, et al: A phase II dendritic cell (DC) immunotherapy trial for hepatocellular carcinoma (HCC). Pro Am Soc Clin Oncol 21:23b, 2002

184. Nakeeb A, Pitt HA, Sohn TA, et al: Cholangiocarcinoma: a spectum of intrahepatic, perihilar, and distal tumors. Ann Surg 224:463, 1996

185. Klatskin G: Adenocarcinoma of the hepatic duct at its bifurcation within the porta hepatis: an unusual tumor with distinctive clinical and pathological features. Am J Med 38:241, 1965

186. Sheth S, Bedford A, Chopra S: Primary gallbladder cancer: recognition of risk factors and the role of prophylactic cholecystectomy. Am J Gastroenterol 95:1402, 2000

187. Khan ZR, Neugut AI, Ahsan H, et al: Risk factors for biliary tract cancers. Am J Gastroenterol 94:149, 1999

188. de Groen PC, Gores GJ, LaRusso NF, et al: Biliary tract cancers. N Engl J Med 341:1368, 1999

189. Yee K, Sheppard B, Domreis J, et al: Cancers of the gallbladder and biliary ducts. Oncology (Huntingt) 16:939, 2002

190. Kornfeld D, Ekbom A, Ihre T: Survival and risk of cholangiocarcinoma in patients with primary sclerosing cholangitis: a population-based study. Scand J Gastroenterol 32:1042, 1997

191. Hassoun Z, Gores GJ, Rosen CB: Preliminary experience with liver transplantation in selected patients with unresectable hilar cholangiocarcinoma. Surg Oncol Clin N Am 11:909, 2002

192. Hejna M, Pruckmayer M, Raderer M: The role of chemotherapy and radiation in the management of biliary cancer: a review of the literature. Eur J Cancer 34:977, 1998

193. Kresl JJ, Schild SE, Henning GT, et al: Adjuvant external beam radiation therapy with concurrent chemotherapy in the management of gallbladder carcinoma. Int J Rad Oncol Biol Phys 52:167, 2002

194. Duck L, Humblet Y, Gigot J-Flonneux M, et al: Gemcitabine in advanced cholangiocarcinoma: a single-center retrospective study. Pro Am Soc Clin Oncol 21:125b, 2002

195. Fishikin P, Alberts S, Nahoney M, et al: Irinotecan (CPT-11) in patients with advanced gallbladder carcinoma: a North Central Cancer Treatment Group phase II study. Pro Am Soc Clin Oncol 20:1551, 2001

196. Papakostas P, Kouroussis C, Adroulakis N: First-line chemotherapy with docetaxel for unresectable or metastatic carcinoma of the biliary tract: a multicenter phase II study. Pro Am Soc Clin Oncol 20:152a, 2001

197. Lozano RD, Patt YZ, Hassan MM, et al: Oral capecitabine (Xeloda) for the treatment of hepatobiliary cancers (hepatocellular carcinoma, cholangiocarcinoma, and gallbladder cancer). Pro Am Soc Clin Oncol 19:264a, 2000

198. Palefsky JM, Holly EA, Gonzales J, et al: Detection of human papillomavirus DNA in anal intraepithelial neoplasia and anal cancer. Cancer Res 51:1014, 1991

199. Frisch M, Biggar RJ, Goedert JJ: Human papillomavirus-associated cancers in patients with human immunodeficiency virus infection and acquired immunodeficiency syndrome. J Natl Cancer Inst 92:1500, 2000

200. Piketty C, Darragh TM, Costa MD, et al: High prevalence of anal human papillomavirus infection and anal cancer precursors among HIV-infected persons in the absence of anal intercourse. Ann Intern Med 183:453, 2003

201. Tarantino E, Bernstein MA: Endoanal ultrasound in the staging and management of squamous-cell carcinoma of the anal canal: potential implications of a new ultrasound staging system. Dis Colon Rectum 45:16, 2002

202. Flam M, John M, Pajak TF, et al: Role of mitomycin in combination with fluorouracil and radiotherapy, and of salvage chemoradiation in the definitive nonsurgical treatment of epidermoid cancer of the anal canal: results of phase III randomized intergroup study. J Clin Oncol 14:2527, 1996

203. Gerard JP, Ayzac L, Hun D, et al: Treatment of anal canal carcinoma with high dose radiation therapy and concomitant fluorouracil-cisplatinum: long-term results in 95 patients. Radiother Oncol 46:249, 1998

204. Moore HG, Guillem JG: Anal neoplasms. Surg Clin North Am 82:1233, 2002

205. Demetri GD, von Mehren, M, Blanke CD, et al: Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347:472, 2002

206. Miettinen M, Sarlomo-Rikala M, Lasota J: Gastrointestinal stromal tumors: recent advances in understanding of their biology. Hum Pathol 30:1213, 1999

207. Rubin BP, Fletcher JA, Fletcher CD: Molecular insights into the histogenesis and pathogenesis of gastrointestinal stromal tumors. Int J Surg Pathol 8:5, 2000

208. Miettinen M, Majidi M, Lasota J: Pathology and diagnostic criteria of gastrointestinal stromal tumors (GISTs): a review. Eur J Cancer 38:S39, 2002

209. Fletcher CD, Berman JJ, Corless C, et al: Diagnosis of gastrointestinal stromal tumors: a consensus approach. Hum Pathol 33:459, 2000

210. Lux M, Rubin BP, Biase TL, et al: KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. Am J Pathol 156:792, 2000

211. Rubin BP, Singer S, Tsao C, et al: KIT activation is a ubiquitous feature of gastrointestinal stromal tumors. Cancer Res 61:8118, 2001

212. Druker BJ, Talpaz M, Resta DJ, et al: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344:1031, 2001

213. Heinrich MC, Griffith DJ, Druker BJ, et al: Inhibition of c-kit receptor tyrosine kinase activity of STI 571, a selective tyrosine kinase inhibitor. Blood 96:925, 2000

214. Tuveson DA, Willis NA, Jacks T, et al: STI571 inactivation of the gastrointestinal stoma tumor c-KIT oncoprotein: biological and clinical implications. Oncogene 20:5054, 2001

215. Van den Abbeele AD, Badawi RD: Use of positron emission tomography in oncology and its potential role to assess response to imatinib mesylate therapy in gastrointestinal stromal tumor (GISTs) Eur J Cancer 38(suppl 5):S60, 2002

216. Aisenberg AC: Coherent view of non-Hodgkin's lymphoma. J Clin Oncol 13:2656, 1995

217. Severson RK, Davis S: Increasing incidence of primary gastric lymphoma. Cancer 66:1283, 1990

218. Stolte M, Bayerdorffer E, Morgner A, et al: *Helicobacter* and gastric MALT lymphoma. Gut 50:19, 2002

219. Weingrad DN, Decosse JJ, Sherlock P, et al: Primary gastrointestinal lymphoma: a 30-year review. Cancer 49:1258, 1982

220. d'Amore F, Brincker H, Gronbaek K, et al: Non-Hodgkin's lymphoma of the gastrointestinal tract: a population-based analysis of incidence, geographic distribution, clinicopathologic presentation features, and prognosis. Danish Lymphoma Study Group. J Clin Oncol 12:1673, 1994

221. Sandler RS: Primary gastric lymphoma: a review. Am J Gastroenterol 79:21, 1984

222. Morgner A, Bayerdorffer E, Neubauer A, et al: Malignant tumors of the stomach: gastric mucosa-associated lymphoid tissue lymphoma and *Helicobacter pylori*. Gastroenterol Clin North Am 29:593, 2000

223. Radaszkiewicz T, Dragosics B, Bauer P: Gastrointestinal malignant lymphomas of the mucosa-associated lymphoid tissue: factors relevant to prognosis. Gastroenterology 102:1628, 1992

224. Fischbach W, Goebeler-Kolve M, Starostik P, et al: Minimal residual low-grade gastric MALT-type lymphoma after eradication of *Helicobacter pylori*. Lancet 360:547, 2002

Acknowledgments

Figures 1 and 4a courtesy of Stephen E. Rubesin, M.D., University of Pennsylvania.

Figures 2, 4b, and 8 courtesy of Michael Kochman, M.D., University of Pennsylvania.

Figure 3 Adapted from "Early gastric cancer in young, asymptomatic carriers of germ-line E-cadherin mutations," by D. G. Huntsman, F. Carneiro, and F. L. Lewis, et al., in *New England Journal of Medicine* 344:1904, 2001.

Figure 5 Adapted from "The National Cancer Data Base Report on poor survival of U.S. gastric carcinoma patients treated with gastrectomy: Fifth Edition American Joint Committee on Cancer staging, proximal disease, and the "different disease" hypothesis. *Cancer* 88:921, 2000.

Figure 6 Seward Hung.

Figures 7 and 9 courtesy of Evan Siegelman, M.D., University of Pennsylvania.

VII BREAST CANCER

Nancy E. Davidson, m.d.

Invasive breast cancer, the most common nonskin cancer in women in the United States, will be diagnosed in approximately 216,000 women in this country in 2004 and is expected to result in approximately 40,000 deaths.[1] Incidence and mortality have reached a plateau and appear to be dropping in both the United States[2] and parts of western Europe.[2] This decline has been attributed to several factors, such as early detection through the use of screening mammography and appropriate use of systemic adjuvant therapy. Breast cancer and its treatment have been comprehensively reviewed.[3]

Etiology and Epidemiology

The incidence of breast cancer in the United States has increased steadily since World War II. Multiple risk factors for its development have been identified. Breast cancer is predominantly a disease of women, although it does occur in men, with an incidence of approximately 1% of that seen in women. Chief among the risk factors are advancing age and family history. Incidence increases with age, and about 75% of breast cancer cases in the United States are diagnosed in women older than 50 years. The diagnosis of breast cancer in first-degree relatives younger than 50 years is associated with a threefold to fourfold increased risk. Approximately 5% to 8% of cases of breast cancer occur in high-risk families. Several familial breast cancer syndromes and their associated molecular abnormalities have been identified. These include the breast-ovarian cancer syndrome, which is attributed to germline mutations in either of two breast cancer–susceptibility genes, *BRCA1* and *BRCA2*. These mutations are inherited as an autosomal dominant trait and can therefore be transmitted through both the maternal and the paternal line. Initial linkage studies suggested that a germline mutation in either of these genes was associated with a lifetime risk of breast cancer of up to 85%. However, more recent population-based studies suggest that breast cancer develops in 50% to 60% of carriers.[4] Other hereditary breast cancer syndromes are the Li-Fraumeni syndrome, which is associated with germline mutations in the *p53* tumor suppressor gene,[5] and Cowden disease, which is linked to inherited mutations in the *PTEN* gene.[6]

Reproductive risk factors include early menarche, late menopause, late first pregnancy, and nulliparity. All are felt to lead to a condition of prolonged estrogen exposure to the breast. Careful pathologic study has demonstrated that certain types of breast pathology, such as atypical hyperplasia and lobular carcinoma in situ, are also associated with increased risk. Furthermore, certain environmental factors increase the risk of breast cancer. These factors include exposure to ionizing radiation during adolescence,[7] prolonged use of hormone replacement therapy,[8] ongoing use of oral contraceptives,[9] and consumption of alcohol.[10] Large-scale epidemiologic studies have failed to show any convincing linkage between exposure to certain pesticides[11] or a high-fat diet[12,13] and development of breast cancer.

Prevention

CHEMOPREVENTION

There has been much interest in the development of breast cancer prevention strategies. Pharmacologic approaches are the most developed at this time. The nonsteroidal selective estrogen receptor modulator (SERM) tamoxifen, a mainstay in the management of breast cancer, has also been tested as a chemopreventive agent. A large United States trial randomized 13,388 high-risk women to receive either tamoxifen or placebo for 5 years.[14] Eligibility for this trial was based on increased risk of breast cancer; risk factors used to determine eligibility were age greater than 60 years, a diagnosis of lobular carcinoma in situ, or age of 35 to 59 years with a constellation of risk factors that, when combined, resulted in a 1.67% or greater risk of breast cancer developing within 5 years. Risk was calculated for the last group of women by a modification of a previously validated risk model. Use of tamoxifen was associated with a reduction in the diagnosis of breast cancer of about 50%; benefit was noted in all age groups. Other effects of tamoxifen are summarized [*see Table 1*]. As a general rule, an increase in serious adverse events such as endometrial cancer or thromboembolic events was confined largely to women older than 50 years. On the basis of these data, the Food and Drug Administration (FDA) has approved the use of tamoxifen to reduce the occurrence of breast cancer in high-risk women, as defined by the eligibility requirements for this prevention trial. Two other tamoxifen prevention trials, conducted in the United Kingdom[15] and Italy,[16] did not show any advantage with tamoxifen; this finding has been attributed to the smaller size and variations in eligibility criteria of the two studies. In particular, the small sample size and lower baseline breast cancer risk may have limited the ability of these trials to detect a tamoxifen effect. A fourth trial, the International Breast Cancer Intervention Study, also showed a reduction in diagnosis of invasive breast cancer in tamoxifen recipients.[17] These four trials were analyzed in a meta-analysis that confirmed a 38% reduction in invasive breast cancer with tamoxifen.[18] Salient features of reported trials are summarized [*see Table 2*].

Other selective estrogen receptor modulators are under investigation. One such agent, raloxifene, is approved for use in the prevention of osteoporosis in postmenopausal women. Early results from the osteoporosis trials suggest that raloxifene may also reduce the risk of breast cancer.[19] However, its cancer preventive effects have not been examined directly. Raloxifene's potential for preventing breast cancer is being compared with that of tamoxifen in the current United States breast cancer prevention trial, which is targeting postmenopausal women at increased risk for breast cancer. In addition, on the basis of preclinical models, the vitamin A derivative fenretinide was assessed as a breast cancer chemopreventive agent in an Italian study of about 3,000 breast cancer survivors and showed no overall impact after 5 years of treatment.[20]

LIFESTYLE MODIFICATION

Prevention strategies that involve alterations in lifestyle have been suggested. The Women's Health Initiative is examining the

Table 1 Selected Outcomes from the NSABP P-1 Tamoxifen Chemoprevention Trial[14]

Outcome	*Placebo (N = 6,599)*	*Tamoxifen (N = 6,576)*	*Risk Ratio*	*95% Confidence Interval*
Invasive breast cancer	175	89	0.51	0.39–0.66
Endometrial cancer	15	36	2.53	1.35–4.97
Other cancers	97	97	1.00	0.75–1.35
Myocardial infarction	28	31	1.11	0.65–1.92
Stroke	24	38	1.59	0.93–2.77
Pulmonary embolism	6	18	3.01	1.15–9.27
Deep vein thrombosis	22	35	1.60	0.91–2.86
Fractures	137	111	0.81	0.63–1.05
Cataract development	507	574	1.14	1.01–1.29

NSABP P-1—National Surgical Adjuvant Breast and Bowel Project

value of a low-fat diet as a means of breast cancer prevention. Regular exercise, especially during adolescence, may be associated with diminished breast cancer risk. By inference from the epidemiologic studies mentioned above, abstinence from alcohol may slightly reduce the risk of breast cancer.

PROPHYLACTIC MASTECTOMY

Prophylactic mastectomy has been considered as a means of decreasing breast cancer risk, particularly for women at high risk because of a strong family history or because the women are carriers of mutant breast cancer susceptibility genes. Database studies have suggested that prophylactic bilateral mastectomy may decrease development of breast cancer by up to 90%.[21] It is critical that women contemplating such an approach be aware of reports of cancer development in remnants of breast tissue after the prophylactic surgery. It is not known whether prophylactic surgery is superior to aggressive screening coupled with appropriate management of any breast cancer that is diagnosed.

Screening

Screening strategies for breast cancer include the triad of breast self-examination (BSE), clinical breast examination by a health care professional, and screening mammography. Although widely touted as an important component of early detection, BSE is of uncertain value. A large randomized trial that compared conventional BSE with observation in over 260,000 female Chinese textile workers failed to show any clinical advantage with BSE.[22] As a consequence, some experts, including the

Table 2 Tamoxifen Chemoprevention Trials for Breast Cancer

Trial	*Participant Number*	*Breast Cancer Events*	*Median Follow-up*	*Participant Characteristics*	*Results*
NSABP P-1	13,388	358	55 mo	≥ 1.67% 5-year risk of developing breast cancer No concurrent HRT 76% with first-degree relative 39% < 50 yr old	49% overall reduction in breast cancer diagnosis in tamoxifen group
Italy	5,408	79	81 mo	No requirement for increased risk Concurrent HRT allowed 12% with first-degree relative 38% < 50 yr old Hysterectomy required	No difference in breast cancer diagnosis except in HRT recipients
Royal Marsden	2,471	70	70 mo	Increased risk because of family history Concurrent HRT in 26% 96% with first-degree relative 62% < 50 yr old	No difference in breast cancer
IBIS-1	7,152	170	50 mo	Increased risk because of family history or LCIS 48% had a first-degree relative who developed breast cancer before age 50 Median age = 51	25% reduction in breast cancer in tamoxifen group

HRT—hormone replacement therapy LCIS—lobular carcinoma in situ

American Cancer Society, now promote breast awareness rather than regular BSE.

In contrast, regular screening by mammography[23] and clinical breast examination appear to decrease mortality from breast cancer by 25% to 30% in women older than 50 years. Considerable controversy continues over the value of screening mammography and clinical breast examination in women 40 to 50 years of age. Currently, the American Cancer Society and the National Cancer Institute recommend annual screening mammography for women older than 40 years who are at standard risk for breast cancer. A second area of controversy is the age at which screening mammography may be ceased. No randomized trial has assessed the role of screening mammography in women older than 70 years. However, it would seem reasonable to continue mammography in older women whose life expectancy exceeds 5 years; women whose survival is limited because of other medical conditions are not likely to benefit. There have been no trials of mammographic strategies directed explicitly at high-risk women, particularly those with *BRCA1* and *BRCA2* mutations. In the absence of relevant data, a reasonable approach is to begin mammographic screening at 25 years of age or 5 years earlier than the age of the person with the earliest diagnosis of breast cancer in the immediate or extended family.

Much work is focused on the value of ultrasonography, magnetic resonance imaging, and technetium-99m sestamibi imaging in screening. Currently, these modalities are used largely as ancillary diagnostic tools and have no defined place in the screening of asymptomatic persons. Further testing of these modalities as screening tools is driven in part by the knowledge that 10% to 15% of breast cancers are not detected by mammography. Evaluation of these modalities as screening tools is of particular interest in women who are at high risk, such as carriers of *BRCA1* and *BRCA2* gene mutations.

Staging and Prognosis

The TNM classification of clinical staging rests on the clinical assessment of tumor size, nodal status, and evidence of metastatic disease. However, pathologic staging is preferable because it provides the most accurate estimate of tumor involvement and prognosis. The staging system for breast cancer was recently revised.[24] Most patients present with stage I or II breast cancer. In these patients, extensive laboratory evaluation is of little value. Studies in the asymptomatic patient with apparent stage I or II breast cancer can be limited to a hemogram, chemistry panel, and chest x-ray. More sophisticated radiologic or laboratory staging is not warranted because of low yield. However, patients with symptoms suggestive of metastatic disease or women with clinical evidence of stage III or IV breast cancer should undergo more intensive evaluation of common sites of metastasis, such as bone, liver, and lung, by use of radionuclide scanning and computed tomography. Current staging guidelines are presented [*see Tables 3 and 4*].

Staging is the most important component in establishing prognosis. Indeed, axillary lymph node status and tumor size are the two most important determinants of outcome for patients with early breast cancer. Other established factors that help define prognosis are estrogen receptor–alpha (ER-α) and progesterone receptor (PR) content of the primary tumor mass and histologic grade of the tumor. Biologic factors that are still under evaluation for determining prognosis include expression of the *HER-2/neu* (also known as *c-erb-B2* or *neu*) and *p53* oncogenes and various measures of cellular proliferation, such as S-phase fraction or Ki67. Factors associated with poorer prognosis are lymph node involvement, increasing tumor size, high histologic grade, and absence of ER and PR expression. In addition, overexpression of the oncogenes, as well as increased cellular proliferative measures, may be associated with adverse clinical outcome. A new area of research is predictive factors—that is, identification of biologic features correlated with sensitivity or resistance of a tumor to a particular therapy. There are three established predictive factors for breast cancer: ER, PR, and HER-2/neu. The majority of tumors that express ER, PR, or both are initially sensitive to endocrine therapies, whereas tumors that lack ER and PR rarely respond to such treatment. Similarly, overexpression of the HER-2/neu protein or amplification of the *HER-2/neu* gene is associated with response to the HER-2/neu-targeted monoclonal antibody trastuzumab. Investigators are examining the possibility that overexpression of the HER-2/neu protein may indicate a need for doxorubicin-based chemotherapy; it may also predict relative resistance to certain hormone therapies, such as tamoxifen.

Early Breast Cancer (Stages I and II)

DIAGNOSIS

Many cases of early, clinically occult breast cancer are diagnosed on the basis of architectural changes or microcalcification seen on a mammogram. Women with clinically evident breast cancer generally present with breast-specific complaints, such as a palpable mass, a change in breast contour, or skin or nipple changes. For both clinically occult and clinically apparent cancer, pathologic evaluation is mandatory to establish a diagnosis. In the past, incisional or excisional biopsies were routinely employed for this purpose. Today, fine-needle aspiration and core-needle biopsy are the standard diagnostic modalities. These procedures can be performed in the office in patients with suspicious palpable lesions. For women with nonpalpable lesions, biopsies guided by mammography, ultrasonography, or MRI are now routine. These technologies permit an accurate diagnosis, which can be followed by definitive treatment planning; consequently, only a single surgical procedure is required. Alternatively, women whose diagnoses are unequivocally negative can be spared an open surgical biopsy. However, it is axiomatic that further evaluation be undertaken for suspicious lesions that yield an equivocal diagnosis after needle aspiration or core biopsy. In addition, bilateral breast imaging is required to identify any unsuspected lesions in the contralateral breast that may also mandate further evaluation.

LOCAL TREATMENT OF EARLY-STAGE BREAST CANCER

In Situ Carcinoma

Because of increased use of screening mammography and heightened breast cancer awareness, in situ carcinomas now account for about 20% of newly diagnosed cases of breast cancer.[25] The majority of these carcinomas are ductal carcinoma in situ (DCIS). Such lesions are associated with an approximately 30% risk of subsequent invasive breast cancer in the ipsilateral breast. The risk of metastatic breast cancer is small with DCIS; as a consequence, axillary lymph node evaluation is not routinely performed. Thus, management decisions are centered largely on the involved breast. Total mastectomy, the tradition-

Table 3 TNM Staging System for Breast Cancer[24]

Primary Tumor (T)	TX Primary tumor cannot be assessed T0 No evidence of tumor Tis Carcinoma in situ Tis (DCIS) Ductal carcinoma in situ Tis (LCIS) Lobular carcinoma in situ Tis (Paget) Paget disease of the nipple with no tumor* T1 Tumor ≤ 2 cm T1mic ≤ 0.1 cm T1a > 0.1 cm – 0.5 cm T1b > 0.5 cm – 1 cm T1c > 1 cm – 2 cm T2 Tumor > 2 cm – 5 cm T3 Tumor > 5 cm T4 Tumor of any size with direct extension to (a) chest wall or (b) skin, only as described below T4a Extension to chest wall, not including pectoralis muscle T4b Edema (including peau d'orange) or ulceration of the skin of the breast, or satellite skin nodules confined to the same breast T4c Both T4a and T4b T4d Inflammatory carcinoma
Regional Lymph Nodes (N)	NX Regional lymph nodes cannot be assessed (e.g., previously removed) N0 No regional lymph node metastasis N1 Metastasis in movable ipsilateral axillary lymph node or nodes N2 Metastasis in ipsilateral axillary lymph nodes or in clinically apparent† ipsilateral internal mammary nodes in the absence of clinically evident lymph node metastasis N2a Metastasis in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures N2b Metastasis only in clinically apparent† ipsilateral internal mammary nodes and in the absence of clinically evident axillary lymph node metastasis N3 Metastasis in ipsilateral infraclavicular lymph node or nodes or in clinically apparent† ipsilateral internal mammary lymph node or nodes and in the presence of clinically evident axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node or nodes with or without axillary or internal mammary lymph node involvement N3a Metastasis in ipsilateral infraclavicular lymph node or nodes and axillary lymph node or nodes N3b Metastasis in ipsilateral internal mammary lymph node or nodes and axillary lymph node or nodes N3c Metastasis in ipsilateral supraclavicular lymph node or nodes
Regional Lymph Nodes (pN) ‡	pNX Regional lymph nodes cannot be assessed (e.g., previously removed) pN0 No regional lymph node metastasis histologically, no additional examination for isolated tumor cells§ pN0(i-) No regional lymph node metastasis histologically, negative IHC pN0(i+) No regional lymph node metastasis histologically, positive IHC, no IHC cluster > 0.2 mm pN0(mol-) No regional lymph node metastasis histologically, negative molecular findings (RT-PCR) pN0(mol+) No regional lymph node metastasis histologically, positive molecular findings (RT-PCR) pN1mi Micrometastases (> 0.2 mm, none > 2.0 mm) pN1 Metastasis in one to three axillary lymph nodes and/or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent‖¶ pN1a Metastasis in one to three axillary lymph nodes pN1b Metastasis in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent‖ pN1c Metastasis in one to three axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent ‖¶ pN2 Metastasis in four to nine lymph nodes or in clinically apparent† internal mammary lymph nodes in the absence of axillary lymph node metatasis pN2a Metastasis in four to nine axillary lymph nodes (at least one tumor deposit > 2.0 mm) pN2b Metastasis in clinically apparent† internal mammary lymph nodes in the absence of axillary lymph node metastasis pN3 Metastasis in 10 or more axillary lymph nodes, in infraclavicular lymph nodes, or in clinically apparent† ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes with clinically negative microscopic metastasis in internal mammary lymph nodes; or in ipsilateral lymph nodes pN3a Metastasis in 10 or more axillary lymph nodes (at least one tumor deposit > 2.0 mm), or metastasis to the infraclavicular lymph nodes pN3b Metastasis in clinically apparent† ipsilateral internal mammary lymph nodes in the presence of one or more positive axillary lymph nodes; or in more than three axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not clinically apparent‖ pN3c Metastasis in ipsilateral supraclavicular lymph nodes
Metastasis (M)	MX Distant metastasis cannot be assessed M0 No distant metastases M1 Distant metastasis

IHC—immunohistochemistry RT-PCR—reverse transcriptase polymerase chain reaction

*Paget disease associated with a tumor is classified according to the size of the tumor.

†Clinically apparent is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination.

‡Classification is based on axillary lymph node dissection with or without sentinel lymph node dissection. Classification based solely on sentinel lymph node dissection without subsequent axillary lymph node dissection is designated as "sn," for sentinel node (e.g., pN0[i+][sn]).

§Isolated tumor cells are defined as single tumor cells or small cell clusters not greater than 0.2 mm, usually detected only by immunohistochemical or molecular methods but which may be verified by hematoxylin and eosin stains.

‖Not clinically apparent is defined as not detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination.

¶If associated with more than three positive axillary lymph nodes, the internal mammary nodes are classified as pN3b to reflect increased tumor burden.

Table 4 TNM Stage Grouping for Breast Cancer[24]

	T	*N*	*M*
Stage 0	Tis	N0	M0
Stage 1	T1	N0	M0
Stage IIA	T0 T1 T2	N1 N1 N0	M0
Stage IIB	T2 T3	N1 N0	M0
Stage IIIA	T0 T1 T2 T3	N2 N2 N2 N1, N2	M0
Stage IIIB	T4 Any T	Any N N3	M0
Stage IIIC	Any T	N3	M0
Stage IV	Any T	Any N	M1

al therapy, has a high likelihood of cure. Studies suggest that breast conservation is appropriate for many women with DCIS. Contraindications are poor cosmesis or multifocal disease. Several models using pathologic factors to predict outcome have been devised to guide decisions about local therapy; key factors are size and grade of lesion and status of surgical margins. If breast conservation is the goal, then an adequate excision to obtain tumor-free margins is important. Careful mammographic examination of the specimen and postexcision mammography of the breast are crucial to ascertain whether the lesion has been fully excised. A large randomized trial has demonstrated that radiotherapy along with lumpectomy, compared with lumpectomy alone, decreases the likelihood of recurrence of invasive or in situ breast cancer.[26] Nonetheless, some women with favorable lesions who are willing to undergo close surveillance are candidates for local excision alone. In one trial, a small benefit was demonstrated with the use of a 5-year regimen of tamoxifen in women with DCIS who underwent excision and radiotherapy.[26] Both ipsilateral breast cancer recurrence and contralateral breast cancer development were reduced by about 50%. A second trial did not show any advantage for the use of tamoxifen.[27]

Controversy exists as to whether lobular carcinoma in situ (LCIS) is a truly malignant lesion. Diagnosis of LCIS is usually an incidental finding on breast biopsy, and it appears to be associated with a 30% risk of development of invasive breast cancer in either breast. Women with LCIS are generally regarded as candidates for careful surveillance with regular breast examination and mammography. Bilateral total mastectomy may be considered for women with LCIS who have other risk factors or who are extremely concerned. The United States Breast Cancer Prevention Trial suggested that a 5-year regimen of tamoxifen therapy decreases the risk of invasive breast cancer by one half in women with LCIS; thus, tamoxifen chemoprevention can also be considered for these women.[14]

Invasive Breast Cancer

Surgical therapy Although radical mastectomy (removal of the involved breast, axillary contents, and underlying chest wall musculature) was the mainstay for breast cancer treatment for many decades, it is seldom performed today. Multiple randomized trials have uniformly and unequivocally shown that breast conservation therapy (BCT) that involves lumpectomy with radiotherapy and modified radical mastectomy that involves removal of the breast and axillary nodes provide identical survival rates for women with stage I or II breast cancer.[28] Patient preference (rather than physician preference) should guide decision making regarding these two options. Medical contraindications for BCT are multifocal disease, ongoing pregnancy that precludes timely administration of radiotherapy, and previous radiotherapy. Although the number of patients being managed with BCT has risen substantially over the past 10 years, wide geographic differences exist in the use of breast conservation in the United States. Patients who undergo mastectomy should be counseled about the availability of breast-reconstruction alternatives.

Because the likelihood of distant micrometastatic spread is highly correlated with the number of involved lymph nodes, axillary dissection has traditionally been used to provide prognostic information about the extent of pathologic lymph node involvement. Therapeutic benefit from axillary dissection is minimal and is likely limited to the minority of women who have lymph node–positive disease. The morbidity associated with breast cancer surgery is largely related to side effects from axillary dissection. Unfortunately, use of prognostic factors derived from the primary tumor has not proved to be an acceptable alternative to assessment of axillary lymph node status. In recent years sentinel lymph node mapping has become the method of choice to evaluate the ipsilateral axillary nodes.

Sentinel lymph node mapping involves injection of a radioactive tracer, vital blue dye, or both into the area around the primary breast tumor. The injected substance tracks rapidly to the dominant axillary lymph node—the so-called sentinel lymph node. This node can be located by use of a small axillary incision and visual inspection or by use of a handheld counter. If the sentinel node is tumor free, the remaining lymph nodes are likely to be tumor free as well, and further axillary surgery can be avoided. Women in whom the sentinel node is found to contain a tumor are currently advised to undergo standard axillary dissection. Initial studies suggest that the surgeon's experience is a key factor in identifying the sentinel lymph node. A small, randomized trial of sentinel node management versus standard axillary lymph node dissection has suggested similar outcomes with the two strategies.[29] Several extremely large randomized trials are expected to definitively resolve the role of sentinel node approaches. Sentinel node mapping is indicated only for women with invasive breast cancer without palpable axillary lymph nodes. Women with palpable nodes should undergo standard axillary dissection.

Adjuvant radiotherapy Radiotherapy is a critical component of BCT; women who undergo lumpectomy by itself have a breast cancer recurrence rate of up to 40%, whereas the rate of recurrence is less than 10% with appropriate radiotherapy. Attempts to identify women whose tumors are so favorable that radiotherapy can be safely withheld have not been successful. In addition, efforts to routinely omit radiotherapy from treatment algorithms for elderly women cannot be condoned, because local failure rates are unacceptably high. Such a strategy of mini-

mal local therapy is appropriate only for patients with significant comorbidities that limit their life expectancy. Current research is focused on the possibility that radiotherapy can be safely and effectively delivered to a smaller field (partial-breast radiotherapy) and over shorter time periods.

A continuing controversy concerns the value of postmastectomy radiotherapy. Once routine, this practice was largely abandoned because a number of randomized trials and a meta-analysis failed to show any benefit in survival from use of this practice. In these trials, the incidence of locoregional recurrence was reduced by 50% to 75%, although cardiac toxicity increased.[30] As a result, postmastectomy radiotherapy was limited to women with extensive local disease, such as skin or chest wall invasion or extensive lymph node involvement. Interest in this field has been revived by the publication of two randomized trials of mastectomy with or without subsequent radiotherapy.[31,32] These trials showed a survival advantage for women with node-positive breast cancer who received radiation. Unlike many of the older trials, these two trials incorporated appropriate adjuvant systemic therapy and modern radiotherapy techniques. Many radiation oncologists now recommend postmastectomy radiotherapy for women with more than three involved lymph nodes and discuss its use with younger women who have one to three positive lymph nodes. Skeptics have argued that the local recurrence rate in the women treated with mastectomy alone in these two trials is higher than that observed in the United States today; thus, the seemingly positive results from the addition of radiotherapy may simply reflect salvage of patients who received inadequate surgery.

SYSTEMIC TREATMENT OF EARLY-STAGE BREAST CANCER

A driving force behind the use of systemic therapy for early-stage breast cancer has been the understanding that many women with primary breast cancer already have distant micrometastases at the time of diagnosis. Over time, overt metastatic disease develops in most of these women despite treatment with state-of-the-art surgery and radiotherapy. Thus, systemic therapy is frequently used in these women to prevent or delay recurrence of disease. The treatment algorithms that are currently in use are the result of over 40 years of clinical trials; the results of these trials have been compiled in sequential overview analyses that have evaluated the worldwide experience with use of ovarian ablation, tamoxifen therapy, and chemotherapy. The most recent results of the meta-analysis for each modality of therapy are summarized below.[33-35] In addition, a new class of agents, the aromatase inhibitors, is entering the adjuvant arena for postmenopausal women with steroid receptor–positive breast cancer.

Ovarian Ablation

Ovarian ablation by way of surgery or radiotherapy is the oldest form of systemic therapy for advanced breast cancer. It was also the first adjuvant systemic therapy studied in a systematic fashion. The 1995 meta-analysis for ovarian ablation focused on 2,102 women younger than 50 years who were participants in one of 12 randomized trials begun before 1990.[33] Young age was used as a surrogate marker for premenopausal status. Some, but not all, of these studies of ovarian ablation involved chemotherapy. In these studies, positive hormone receptor status was generally not an entry criterion. Overall, these studies showed a 25% reduction in the annual odds of recurrence and a 24% reduction in the annual odds of death for women who underwent ovarian ablation in the absence of chemotherapy. The benefit was more modest for women treated with both ablation and chemotherapy—an 8% to 10% reduction. This finding is not surprising, because adjuvant chemotherapy can lead to the onset of menopause, and it is likely that the benefit of ovarian ablation is limited to women who are not menopausal. Benefits were enjoyed by both node-negative and node-positive women, but the absolute benefit was greatest in node-positive women who had a higher baseline risk of recurrence. For example, the 15-year recurrence-free survival rate for node-positive women was 37% for those who underwent ovarian ablation without adjuvant chemotherapy and 24% for those who received neither ablation nor chemotherapy; for node-negative women, the 15-year survival rates were 75% for those who underwent ablation and 66% for those who had neither ablation nor chemotherapy. Thus, the absolute improvement in survival was 13% for node-positive women and 9% for node-negative women. The meta-analysis also included 1,354 women 50 years of age or older who participated in these trials. Not surprisingly, there was no significant effect of ovarian ablation in these women, who were presumably already postmenopausal for the most part.

Interest in this general approach has been renewed by the advent of the luteinizing hormone–releasing hormone (LHRH) agonists. These agents effect a temporary so-called medical castration through their effects on the hypothalamic-pituitary axis; their use ultimately results in a decrease in the production of ovarian steroids. Randomized trials in metastatic breast cancer patients suggest that these drugs, which include goserelin and leuprolide, are as effective as oophorectomy. Several trials have compared the efficacy of LHRH therapy with that of chemotherapy alone, as well as the efficacy of a combination of LHRH therapy and chemotherapy with that of either LHRH therapy alone or chemotherapy alone.[36] In these trials, the regimens are given for 2 to 5 years. These trials are largely directed toward women whose tumors express hormone receptors and who therefore have a greater likelihood of benefiting from an endocrinologic approach. The trials suggest that in certain circumstances, ovarian suppression represents a viable adjuvant strategy for premenopausal women with steroid receptor–positive breast cancer.

Tamoxifen

The mixed estrogen agonist-antagonist tamoxifen has become the most commonly prescribed antineoplastic drug. Its clinical application was recently comprehensively reviewed.[37] After initial studies documented its efficacy in advanced breast cancer, a number of trials investigated its usefulness in women with early-stage breast cancer. Information from most of the randomized trials is collected in the 1995 overview analysis of tamoxifen, which included 37,000 women enrolled in 55 trials.[34] About 8,000 women in these studies were found to have low expression or no expression of ER in their tumors. Tamoxifen had minimal effect on outcome. These findings support preclinical and clinical data suggesting that tamoxifen's major mode of action is through its interaction with the ER. Thus, the major analysis was restricted to 18,000 women with steroid receptor–positive tumors and 12,000 women with untested tumors, 8,000 of whom were predicted to have steroid receptor–positive tumors. In these patients, the benefit of tamoxifen increased with the duration of treatment; the proportional reductions in 10-year recurrence and mortality were 47% and 26%, respectively, with 5-year regimens of tamoxifen therapy. Benefit appeared to extend be-

yond the 5-year period of drug administration. Again, the proportional reductions were similar for node-negative and node-positive women, but absolute improvements were greater in node-positive women. For node-positive women, the 10-year survival rates were 61% with tamoxifen and 50% with placebo; for node-negative women, the 10-year survival rates were 79% and 73%, respectively. Benefit was observed regardless of age, menopausal status, tamoxifen dose, and use of chemotherapy. This overview also provided information on non–breast cancer outcomes, documenting a small increase in uterine cancer and a decrease in contralateral breast cancer. Tamoxifen had no effect on the incidence of other cancers or on the incidence of death from other causes.

Aromatase Inhibitors

The aromatase inhibitors represent a new treatment strategy for postmenopausal women with steroid receptor–positive breast cancer. In older women, the primary source of circulating estrogen is the conversion of androgens (synthesized by the adrenal glands) to estrogen via the action of the enzyme aromatase, which is found in certain tissues, including adipose and mammary tissues. The aromatase inhibitors specifically inhibit this conversion leading to further estrogen deprivation in older women. Randomized trials have shown that the aromatase inhibitors (e.g., anastrozole, letrozole, and exemestane) provide efficacy similar or superior to that of tamoxifen and an acceptable side-effect profile for postmenopausal women with metastatic breast cancer.[38] One of the aromatase inhibitors, anastrozole, has been approved by the FDA as a second form of adjuvant endocrine therapy for postmenopausal women with steroid receptor–positive breast cancer.[39] Two new studies suggest that the aromatase inhibitors letrozole and exemestane can reduce breast cancer recurrence when used after several years of tamoxifen therapy, raising the possibility of sequential endocrine therapy as an effective strategy.[40,41]

Because information about long-term effects of aromatase inhibitors is sparse, some experts have urged a cautious approach to their general use.[42] Studies suggest that these agents are associated with postmenopausal symptoms, arthralgias, osteoporosis, and fractures. Unlike tamoxifen, aromatase inhibitors do not appear to be associated with uterine cancer risk or thromboembolic events. There is uncertainty about long-term side effects such as cardiac risk or effects on cognition. Optimal duration of aromatase-inhibitor therapy is not known, nor is it clear whether one agent offers certain advantages over another. Given their mechanism of action, aromatase inhibitors should not be used in treating premenopausal women.

Chemotherapy

The 1995 overview summarized results of 40,000 women who participated in 69 chemotherapy trials.[35] These trials addressed three general issues: polychemotherapy compared with no chemotherapy; chemotherapy of long duration compared with that of short duration; and anthracycline-containing regimens compared with the classic combination of cyclophosphamide, methotrexate, and fluorouracil (CMF). Combination chemotherapy reduced recurrence by 35% in women younger than 50 years and by 20% in women 50 to 69 years of age. Mortality was reduced by 27% in the younger women and 11% in the older women. As with the other modalities, the proportional reductions were similar in node-negative and node-positive patients. The absolute benefit was greatest for node-positive women younger than 50 years, in whom the 10-year survival rate increased from 43% to 53%. In contrast, node-negative women 50 to 69 years of age had an absolute gain in 10-year survival of only 2%—from 67% to 69%. In this analysis, 3 to 6 months of therapy was as effective as longer durations of therapy. Anthracyclines such as doxorubicin and epirubicin resulted in a slightly greater improvement in 5-year survival than that shown with CMF (72% for anthracycline and 69% for CMF). Chemotherapy outcome was not affected by estrogen receptor status or use of tamoxifen, nor was chemotherapy associated with an increase in deaths from causes other than breast cancer.

Other Chemotherapy Considerations

Selection of agents Initial studies of adjuvant chemotherapy used single agents such as melphalan. Because combination chemotherapy was found to effect higher response rates in metastatic disease, the use of multiagent chemotherapy was introduced into the adjuvant setting. The landmark trial of Bonadonna and coworkers established CMF as a beneficial regimen for women with node-positive breast cancer. The survival advantage of CMF persisted after 20 years of follow-up.[43] Given the efficacy of the anthracyclines (e.g., doxorubicin and epirubicin) in metastatic disease, combination regimens containing anthracyclines were also explored. Trials that compared CMF with combination regimens of cyclophosphamide, doxorubicin (or epirubicin), and fluorouracil (CAF or FEC), as well as similar combination regimens, have shown anthracycline to provide a small benefit . One large trial that compared CAF with CMF in the management of node-negative breast cancer showed 5-year recurrence-free survival rates of 85% and 82%, respectively, thereby demonstrating a modest advantage with the use of CAF.[44] Selection of a regimen containing doxorubicin or epirubicin should be guided by patient and physician preference about relative toxicity and benefit.

Current trials are evaluating the role of the taxanes (i.e., paclitaxel and docetaxel) as adjuvant therapy. Two trials comparing the value of four cycles of doxorubicin and cyclophosphamide (AC) with that of four cycles of AC followed by four cycles of paclitaxel for node-positive breast cancer demonstrated a small early improvement in recurrence-free survival with the addition of paclitaxel.[45,46] A combination of docetaxel, doxorubicin, and cyclophosphamide provided results superior to those achieved with CAF in women with node-positive breast cancer.[47] Other trials of taxanes, either in sequence or in combination, are in progress. Some standard adjuvant chemotherapy regimens are described [*see Table 5*].

Duration Early adjuvant chemotherapy trials investigated therapies lasting up to 2 years. Sequential trials demonstrated that therapy lasting 2 years was no better than therapy lasting 1 year; it was subsequently found that 6 months of therapy was equivalent to 1 year of therapy. Toxicity was greater with the prolonged regimens, and compliance diminished with longer treatment duration. Other trials compared a single cycle of chemotherapy in the perioperative period with longer periods of administration and invariably showed improved outcome with prolonged therapy. Because of these trials, current adjuvant chemotherapy is generally administered for 3 to 6 months, depending on the regimen chosen.[35]

As noted, sequential trials have shown that a 5-year regimen of tamoxifen therapy gives better results than regimens of shorter duration. However, studies of regimens lasting longer

Table 5 Some Commonly Used Adjuvant Chemotherapy Regimens

Acronym	*Drugs*	*Dose*	*Schedule*
CMF	Cyclophosphamide Methotrexate Fluorouracil	100 mg/m^2/day p.o. × 14 days 40 mg/m^2 I.V. days 1 and 8 600 mg/m^2 I.V. days 1 and 8	Repeated every 28 days for six cycles
CAF	Cyclophosphamide Doxorubicin Fluorouracil	100 mg/m^2/day p.o. × 14 days 30 mg/m^2 I.V. days 1 and 8 500 mg/m^2 I.V. days 1 and 8	Repeated every 28 days for six cycles
FAC	Fluorouracil Doxorubicin Cyclophosphamide	500 mg/m^2 I.V. days 1 and 8 50 mg/m^2 I.V. day 1 500 mg/m^2 I.V. day 1	Repeated every 21 days for six cycles
AC	Doxorubicin Cyclophosphamide	60 mg/m^2 I.V. day 1 600 mg/m^2 I.V. day 1	Repeated every 21 days for four cycles
AC → T	Doxorubicin Cyclophosphamide *followed by* Paclitaxel	60 mg/m^2 I.V. day 1 600 mg/m^2 I.V. day 1 175 mg/m^2 I.V. day 1	Repeated every 21 days for four cycles Repeated every 21 days for four cycles
Dose-dense AC → T	Same as AC → T	Same as AC → T	Repeated every 14 days for four cycles with G-CSF support
AC → docetaxel	Doxorubicin Cyclophosphamide *followed by* Docetaxel	60 mg/m^2 I.V. day 1 600 mg/m^2 I.V. day 1 100 mg/m^2 I.V. day 1	Repeated every 21 days for four cycles Repeated every 21 days for four cycles
TAC	Docetaxel Doxorubicin Cyclophosphamide	75 mg/m^2 I.V. day 1 50 mg/m^2 I.V. day 1 500 mg/m^2 I.V. day 1	Repeated every 21 days for six cycles
FEC	Fluorouracil Epirubicin Cyclophosphamide	Various doses	Various schedules

G-CSF—granulocyte colony-stimulating factor

than 5 years have failed to show an advantage beyond 5 years of therapy.[48] The 5-year tamoxifen regimen is currently regarded as standard.

Dose Preclinical data support the hypothesis that the chemotherapy dose is an important determinant of cell kill. As a consequence, the value of dose escalation has been studied extensively. Several trials now provide convincing evidence that reducing the dose to below a standard level is associated with inferior outcome. In a pivotal trial addressing this issue, 1,550 women with node-positive breast cancer were randomly assigned to receive low, medium, or high doses of CAF.[49] Low-dose CAF was clearly associated with higher recurrence rates and poorer survival after 9 years of median follow-up. The high dose of CAF in this study is the same as the standard dose used in adjuvant therapy today. Preliminary studies suggest that the high-dose regimen was superior only in patients whose tumors had poor prognostic biologic factors. In contrast, further dose escalation, made possible through the use of colony-stimulating factors, has not been beneficial. Three large randomized trials failed to show any value of a fourfold increase in the dose of cyclophosphamide[50,51] or a 50% increase in the dose of doxorubicin.[45] Thus, routine administration of high-dose chemotherapy is not warranted, because it can increase toxicity without improving outcome. In addition, several randomized trials have shown that the administration of very high dose chemotherapy combined with autologous bone marrow transplantation or peripheral progenitor cell support does not improve outcomes when compared with standard-dose therapy.[52] Once popular for women with high-risk breast cancer, this approach has been largely abandoned.

Timing Conventional adjuvant chemotherapy usually begins within a few months after surgery. Unnecessary delay is not advisable. Use of primary chemotherapy (i.e., chemotherapy administered before surgery or radiotherapy) has also been advanced and has been tested in several randomized trials. The combined results of these trials show that primary chemotherapy can increase the rate of breast conservation because of its ability to reduce tumor size. However, no reproducible advantage in other clinical outcomes has emerged. In the largest trial, about 1,500 women with palpable breast cancer were randomized to receive four cycles of AC chemotherapy either before or after surgery.[53] After 5 years, disease-free survival and overall survival were identical. Therefore, primary chemotherapy is a safe alternative to traditional adjuvant chemotherapy, particularly for women who desire breast conservation and who are not considered appropriate candidates for breast-conserving surgery at the time of diagnosis.

Given the prevalence of BCT, a second issue concerns the timing of chemotherapy in relation to radiotherapy. A randomized trial compared the outcome of women treated with AC

chemotherapy followed by breast radiotherapy with the outcome of women who received radiotherapy first, followed by identical AC chemotherapy. This trial showed a trend toward improved outcome for the women who received chemotherapy first.[54] Other strategies include concomitant chemotherapy and radiotherapy and so-called sandwich therapy, in which several cycles of chemotherapy are administered, followed by radiotherapy and then the remaining cycles of chemotherapy. Concurrent chemotherapy and radiotherapy requires that certain drugs, such as doxorubicin and methotrexate, be omitted during the period of radiotherapy to prevent toxicity. The available data are not sufficient to definitively judge the effects of concurrent or sandwich therapy on long-term outcome.

Chemoendocrine Therapy

Chemotherapy and tamoxifen or an aromatase inhibitor In light of the benefits of adjuvant chemotherapy and endocrine therapy when administered singly, a logical question concerns the role of combination therapy. Several pivotal trials have examined the value of chemotherapy combined with tamoxifen in node-positive and node-negative breast cancer. One such trial analyzed the value of tamoxifen alone, the combination of tamoxifen with CMF, and the combination of tamoxifen with methotrexate and fluorouracil (MF) for women with node-negative, steroid receptor–positive breast cancer.[55] Although reduced recurrence rates and improved survival rates were noted for all women who received combination therapy, the magnitude of the benefit of combination therapy was greatest in younger women or those with larger tumors.

The efficacy of chemotherapy in combination with tamoxifen for postmenopausal women with node-positive, steroid receptor–positive breast cancer has been addressed.[56,57] Several recent trials suggest that the addition of tamoxifen to chemotherapy reduces recurrence rates by 5% to 7% at 5 years, making the combination of chemotherapy and tamoxifen a reasonable choice for these women. Similarly, the addition of tamoxifen to chemotherapy for premenopausal women with node-positive, steroid receptor–positive breast cancer leads to a 10% improvement in 5-year recurrence-free survival.[58] Not surprisingly, the addition of tamoxifen to chemotherapy has not improved clinical outcome for women with steroid receptor–negative breast cancer in two randomized trials.[44,59]

Together, these trials suggest that chemohormonal therapy is a reasonable consideration for many women with steroid receptor–positive breast cancers. However, an explicit discussion of the magnitude of the potential benefits and side effects is critical, because some women may find the benefit of combined therapy to be too small to warrant its toxicities.

If chemotherapy and tamoxifen are both planned, a sequence of chemotherapy followed by tamoxifen is advisable. An Intergroup trial comparing the efficacy of combined CAF and tamoxifen with the efficacy of CAF followed by tamoxifen in postmenopausal women with node-positive, steroid receptor–positive breast cancer showed an advantage to the sequential approach.[57]

No information is available concerning how to integrate chemotherapy with the use of an aromatase inhibitor. Extrapolating from the information on chemotherapy and tamoxifen, a prudent strategy is to use the sequence of chemotherapy followed by an aromatase inhibitor.

Chemotherapy and ovarian ablation The combination of chemotherapy and ovarian ablation has been assessed in trials using surgical oophorectomy or the administration of LHRH analogues for 2 to 5 years. One large trial showed no overall advantage after 5 years for CAF chemotherapy followed by 5 years of goserelin in 1,500 premenopausal women with node-positive, steroid receptor–positive breast cancer.[60] Follow-up continues; the overview analysis suggests that the benefits of ovarian ablation may not emerge for many years. One side effect of adjuvant chemotherapy in young women is induction of menopause; therefore, any beneficial effects of ovarian ablation likely would be limited to women who remained premenopausal after adjuvant chemotherapy. This hypothesis is under study in a large randomized trial.

Combination endocrine therapy Because different endocrine agents have different mechanisms of action, combination hormone therapy may have better results than either approach used alone. Indeed, combined endocrine therapy in metastatic breast cancer can increase response rate, albeit with higher toxicity and no effect on long-term outcome. Ovarian ablation combined with tamoxifen or an aromatase inhibitor is being studied in premenopausal women. The combination of tamoxifen and anastrozole was shown to be no better than tamoxifen alone in a large randomized trial in postmenopausal women.[39]

Guidelines for Adjuvant-Therapy Administration

An important factor in the decision to use adjuvant therapy is an accurate assessment of the likelihood of breast cancer relapse and mortality with and without therapy. Several tools can help guide this discussion.[61,62] In addition, several groups have devised guidelines for use of adjuvant therapy. Those from the 2003 St. Gallen Breast Cancer Conference, a gathering of international breast cancer experts, are outlined [*see Table 6*].[63] Recommendations from the 2000 NIH Consensus Conference[64] and an algorithm derived from an evidence-based analysis from the National Cancer Center Network (NCCN) are also available.[65] The latter group comprises breast cancer experts from a consortium of cancer centers designated by the National Cancer Institute. These are guidelines, not mandates; they represent a framework for individualized patient counseling about prognosis and therapy. Finally, adjuvant therapy remains imperfect. Numerous clinical trials continue, and participation in clinical trials represents an excellent treatment choice for many patients.

Toxicity of Adjuvant Chemotherapy

Increased use of adjuvant chemotherapy and longer survival have led to concerns about toxicity.[66] Acute side effects of therapy are nausea and vomiting, bone marrow suppression, and hair loss; all are reversible. Induction of menopause is a common concern for younger women. Its likelihood is related to the type of chemotherapy and the age of the patient. CMF and similar regimens are more likely to induce permanent menopause than AC therapy. With either type of regimen, the incidence of menopause is greatest for women 40 years of age and older, the majority of whom will suffer a drug-induced menopause.

Given its association with breast cancer, hormone replacement therapy (HRT) is not generally recommended in breast cancer survivors.[8] Doxorubicin-related cardiomyopathy is another long-term consequence; clinical evidence of congestive heart failure is noted in about 1% of women who receive doxorubicin-containing adjuvant chemotherapy at standard doses. Use of common adjuvant-chemotherapy regimens results in a very small incidence of acute leukemia, but there is no evidence

Table 6 Adjuvant Treatment for Patients with Operable Breast Cancer[63]

Risk Group	*Endocrine-Responsive Disease*		*Endocrine-Nonresponsive Disease*	
	Premenopausal	*Postmenopausal*	*Premenopausal*	*Postmenopausal*
Node-negative disease, minimal risk*	Tamoxifen or none	Tamoxifen or none	NA	NA
Node-negative disease, average risk†	GnRH analogue (or ovarian ablation) + tamoxifen [± chemotherapy], *or* Chemotherapy → tamoxifen [± GnRH analogue (or ovarian ablation)], *or* Tamoxifen, *or* GnRH analogue (or ovarian ablation)	Tamoxifen *or* Chemotherapy → tamoxifen	Chemotherapy	Chemotherapy
Node-positive disease	Chemotherapy → tamoxifen [± GnRH analogue (or ovarian ablation)], *or* GnRH analogue (or ovarian ablation) + tamoxifen [± chemotherapy]	Chemotherapy → tamoxifen *or* Tamoxifen	Chemotherapy	Chemotherapy

*Minimal risk: estrogen and progesterone receptor **expressed**, and **all** of the following features: pathologic tumor size ≤ 2 cm, tumor grade 1, and patient age ≥ 35 yr.
†Average risk: estrogen and progesterone receptor **expressed**, and **at least one** of the following features: pathologic tumor size ≥ 2 cm or tumor grade 2 through 3, or patient age ≤ 35 yr.

of increased incidence of other second tumors. Concerns about cognitive impairment are under investigation.

FOLLOW-UP OF EARLY BREAST CANCER SURVIVORS

Most women present with stage I or II breast cancer and receive appropriate local and systemic therapy. A critical issue is how longitudinal follow-up should be conducted in these patients. Two randomized trials have addressed this issue.[67,68] Both compared a schedule of physician visits and regular laboratory testing (i.e., chest x-ray, bone scanning, and blood studies) with a program of physician visits and laboratory testing that was restricted to evaluation of symptoms. Together, they showed that routine laboratory screening did not enhance survival or quality of life when compared with a program of careful clinical examination with testing tailored for symptoms and physical findings. About 70% of metastases were first detected by the patients themselves, even those patients undergoing physician and laboratory evaluation every 3 months. On the basis of these and other studies, the American Society of Clinical Oncology has published evidence-based guidelines for follow-up of asymptomatic survivors of early-stage breast cancer.[69] These guidelines are summarized [*see Table 7*].

Stage III Breast Cancer

Stage III breast cancer accounts for about 10% of all breast cancers; it is characterized by a primary tumor measuring more than 5 cm, neoplastic invasion of the skin or chest wall, or a fixed tumor or lymph nodes. Inflammatory breast cancer falls into this category. Inflammatory breast cancer has a clinical presentation of breast swelling, erythema, warmth, and a peau d'orange appearance (characterized by a dimpled appearance caused by tumor infiltration of intradermal lymphatics); it may or may not be associated with a mass. These lesions are associated with a high risk of local disease recurrence and distant metastases. Because as many as one third of women with clinical stage III breast cancer have metastases at the time of diagnosis, many oncologists perform a metastatic evaluation at that time, even in asymptomatic patients. Diagnosis is usually established by fine-needle aspiration or core-needle biopsy. Combined-modality therapy is preferred for this stage of breast cancer. Several months of preoperative chemotherapy or hormone therapy result in partial tumor regression in most patients, thereby allowing mastectomy or breast-conserving surgery to be undertaken. Definitive surgery at the time of diagnosis should be avoided because of the high risk of subsequent chest wall recurrence. Radiotherapy is then employed to enhance local control. Some studies suggest that administration of further chemotherapy, hormone therapy, or both is then desirable. Multimodality therapy results in a 5-year disease-free survival rate of about 50%. Follow-up algorithms are similar to those recommended for women with stage I or II disease.

Stage IV or Metastatic Breast Cancer

Although seldom curable, advanced breast cancer is a highly treatable disease. Palliation or prevention of symptoms is the primary goal of treatment. The median survival after diagnosis of metastatic disease is about 2 years, although the range is great. Longitudinal studies have documented a few long-term survivors, most of whom were patients with indolent disease.[70] Several recent clinical trials have documented small improvements in survival with some of the newer therapies.

DIAGNOSIS

As noted, most women with advanced disease present with symptoms or abnormal physical findings. Common sites of relapse are bone, local soft tissues, lung, and liver. If metastasis is suspected, relevant imaging studies (e.g., nuclear medicine, CT scanning, or both) and routine hematologic and biochemical blood studies to assess location and severity of involvement are warranted. Because of the gravity of the diagnosis, pathologic confirmation is preferred. This permits verification of recurrent disease and exclusion of other diagnoses, such as a second primary cancer that would warrant different therapy. Elevated tumor markers such as CA2729 or carcinoembryonic antigen (CEA) are not pathognomonic for recurrent disease, although they may be useful adjuncts in the assessment of the effects of therapy.

TREATMENT

Unlike in early-stage breast cancer, the role of surgery in

metastatic disease is limited. It may be appropriate in some cases, such as excision of a solitary chest wall nodule, removal of a solitary brain metastasis, or orthopedic stabilization to prevent or treat a long-bone fracture. Radiotherapy is a mainstay in the management of advanced disease. It may be used at any time during the disease course to treat localized disease, such as brain metastases or painful bony metastases. In the end, however, systemic treatment with endocrine therapy and chemotherapy is the primary mode of management of disseminated disease. Guiding principles for selection of therapy include maximal palliation of symptoms, prevention of disease-related complications, and the minimizing of therapy-related toxicity. For achieving these goals, endocrine therapy is preferred wherever feasible. An algorithm for systemic treatment selection in stage IV breast cancer is presented [*see Figure 1*].

Endocrine Therapy

Factors that support the use of hormone therapy are the expression of hormone receptors, a long disease-free interval, nonvisceral disease, and the absence of symptoms. Over half of the women who meet these criteria respond to their initial course of endocrine therapy. This response lasts for 9 to 12 months on average; the length of response is a predictor of the likelihood of response to a second course of hormone therapy when the first choice fails. A second course of hormone therapy is less likely to be effective, and the duration of response is shorter; again, the period of response provides an indication of the response to a third course of endocrine therapy. In this way, some women can receive serial endocrine therapy with good disease control for several years.

Numerous types of hormone therapy are now available. Surgical therapy (with the exception of oophorectomy) and first-generation therapies, such as high-dose estrogen, progesterone, and aminoglutethimide therapy, have been largely supplanted by agents with specific mechanisms of action, such as antiestrogens, aromatase inhibitors, and LHRH agonists. Selection is usually made on the basis of efficacy, menopausal status, and toxicity. An algorithm for the selection of endocrine therapy is presented [*see Figure 2*].

Several months of therapy are necessary before the efficacy of a newly introduced hormone regimen can be judged. In addition, hormone-related tumor flare—a syndrome of worsening symptoms and rising tumor markers occurring within the first weeks of treatment—may occur in a few patients. This paradoxical response is usually of short duration and should not be confused with disease progression.

Chemotherapy

Eventually, most women with breast cancer experience hormone-refractory disease and become candidates for chemotherapy. Serial chemotherapy is again the rule. Patients receive two to four cycles of therapy and are then evaluated for evidence of disease stabilization or improvement. Duration of therapy is not fixed. Several trials have compared the strategy of continuing therapy until time of disease progression with the strategy of employing several cycles of therapy followed by a cessation of therapy, with reintroduction of therapy at the time of disease progression.[71,72] In aggregate, these studies have shown survival to be the same with these two approaches, but quality of life as judged by the patients themselves is sometimes better with continuing therapy. Thus, decisions about continuation or termination of a particular therapy will be driven by patient and physician perception of side effects and benefits.

Several alternative first-line therapies now exist. Traditionally, AC, CAF, or CMF was used as initial chemotherapy, with a 40% to 80% response rate, depending on the characteristics of the patient. A meta-analysis of randomized trials suggests that anthracycline-containing regimens may be slightly more effective, but toxicity is also greater.[73] Combinations of taxanes and anthracycline result in response rates and durations that are at least as good as those seen with AC and CAF. Caution should be exercised about their continued use, however, because the doxorubicin-paclitaxel combination is associated with high rates of congestive heart failure. In addition, in a randomized trial, the use of paclitaxel alone and the use of CMF had similar response rates, response duration, and survival rates for women with newly diagnosed stage IV breast cancer.[74] Of note, patient-assessed quality-of-life scores improved during paclitaxel therapy but decreased during CMF therapy.

The scope of chemotherapy has changed dramatically over the past few years, with the availability of new agents such as paclitaxel, docetaxel, vinorelbine, capecitabine, gemcitabine, and carboplatin. All these agents are active individually and in combination. There is debate about the value of combination therapy compared with sequential single-agent therapy for metastatic breast cancer. The taxanes in particular have been shown to improve outcome for women with anthracycline-resistant breast cancer when compared with older regimens such as mitomycin and vinblastine therapy.[75] Thus, paclitaxel and docetaxel represent excellent choices for therapy in this setting. Trials evaluating their optimal dose and schedule of administration, as well as their relative efficacy, are in progress. As with hormone therapy, response rates and durations diminish with each successive change in therapy. A difficult question is when to stop chemotherapy altogether. Although no fixed rules exist, many patients and physicians opt for a program of supportive care if two successive chemotherapy regimens fail to elicit a tumor response or to delay tumor progression. A summary of newer agents available for the treatment of advanced breast cancer is provided [*see Table 8*]. As with early-stage breast cancer, the usefulness of high-dose chemotherapy combined with autologous bone marrow transplantation or stem cell support has not shown benefit and should be viewed as investigational.

Table 7 Guidelines for Surveillance of Asymptomatic Early Breast Cancer Survivors from the American Society of Clinical Oncology[69]

Recommended	Patient education about signs and symptoms of recurrence History and physical examination every 3–6 mo for first 3 yr, every 6–12 mo for next 2 yr, and annually thereafter Monthly breast self-examination Annual mammography Age-appropriate screening for other cancers
Not recommended	Complete blood count Automated chemistry panel Tumor markers (e.g., CEA, CA2729, CA15-3) Bone scan Chest x-ray CT of chest, abdomen, pelvis, and brain

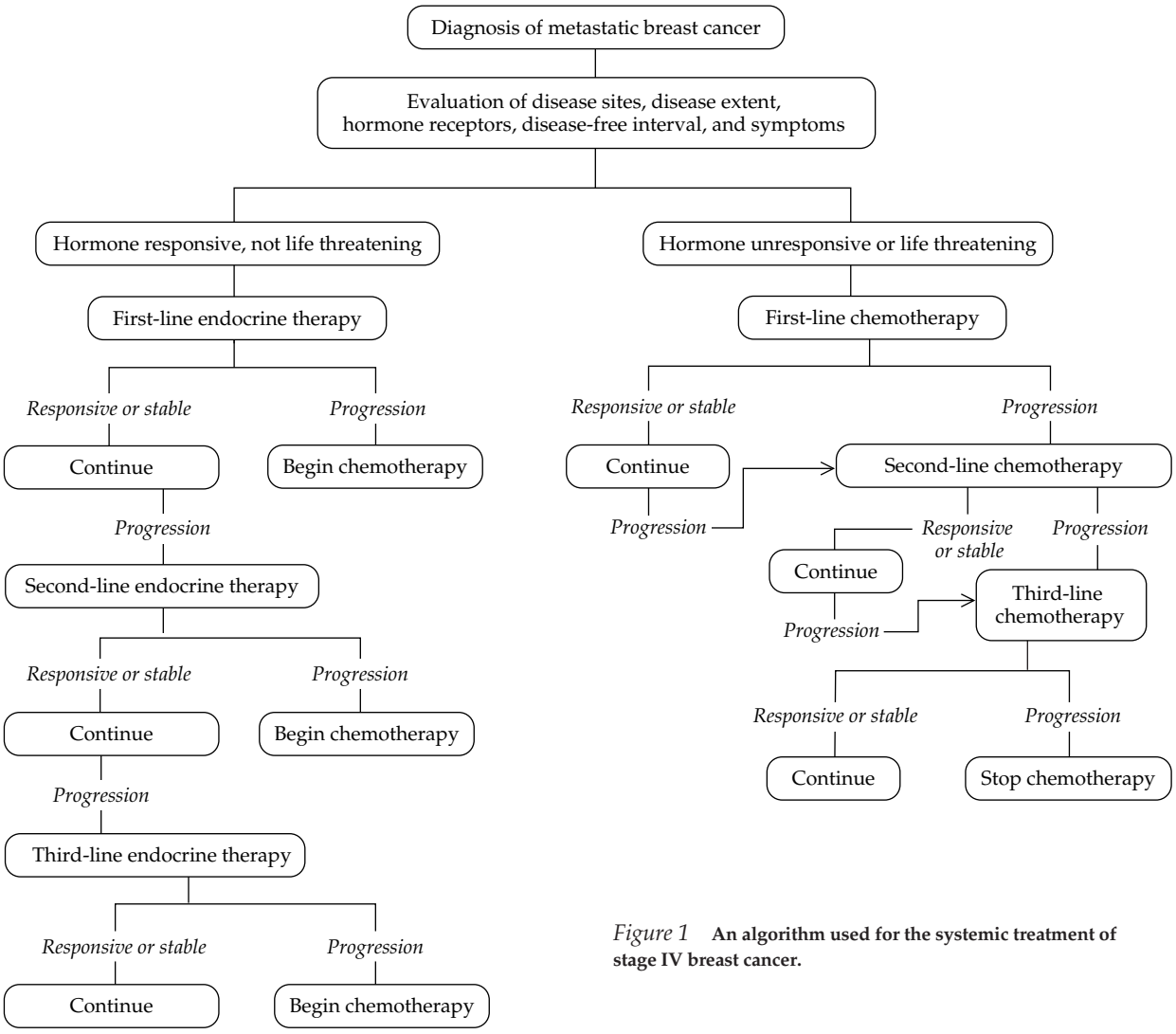


Figure 1 **An algorithm used for the systemic treatment of stage IV breast cancer.**

Biology-Based Therapy

An increased understanding of breast cancer biology has resulted in the identification of new targets for therapy. The first of these to be approved by the FDA is the monoclonal antibody trastuzumab (Herceptin), which is active against the transmembrane HER-2/neu protein. HER-2/neu protein is overexpressed in 20% to 30% of breast cancers; such overexpression may be associated with poorer prognosis and resistance to certain therapies. Weekly administration of trastuzumab resulted in partial or complete tumor regression in 15% of heavily pretreated women whose tumor overexpressed HER-2/neu[76] and in more than 30% of women with untreated HER-2/neu–overexpressing breast cancer.[77] Concomitant administration of trastuzumab with certain cytotoxic agents has also been tested. In a recent randomized trial, concurrent treatment with trastuzumab and paclitaxel increased response rate, response duration, and survival, compared with paclitaxel alone, for women with newly diagnosed metastatic breast cancer.[78] These promising results have led to four randomized trials that are testing the value of adjuvant use of trastuzumab in chemotherapy regimens. Trastuzumab also enhanced the antitumor effects of AC chemotherapy, but the combination was associated with a 20% incidence of congestive heart failure. This unexpected finding highlights the need for careful evaluation of new biology-based therapies as they are introduced into the clinic.

Other molecular targets under investigation include the epidermal growth factor receptor. One such agent is gefitinib, a small molecule that targets the epidermal growth factor receptor tyrosinase kinase. Initial studies of gefitinib in women with very advanced breast cancer showed little activity; however, this agent is the subject of ongoing study. In particular, preclinical models suggest that gefitinib may be useful in subverting some types of hormone resistance. Molecules that might block tumor angiogenesis are also under investigation. One such agent is bevacizumab, a monoclonal antibody that targets vascular endothelial growth factor; this agent has modest activity as a single agent in advanced breast cancer, and its use with chemotherapy is being explored.[79] Finally, new selective estrogen receptor modulators and pure antiestrogens are in development. The drug fulvestrant, which downregulates ER expression, is now being used in clinical practice. The effects of fulvestrant appear to be similar to those of anastrozole for postmenopausal women with advanced breast cancer who have previously received tamoxifen.[80]

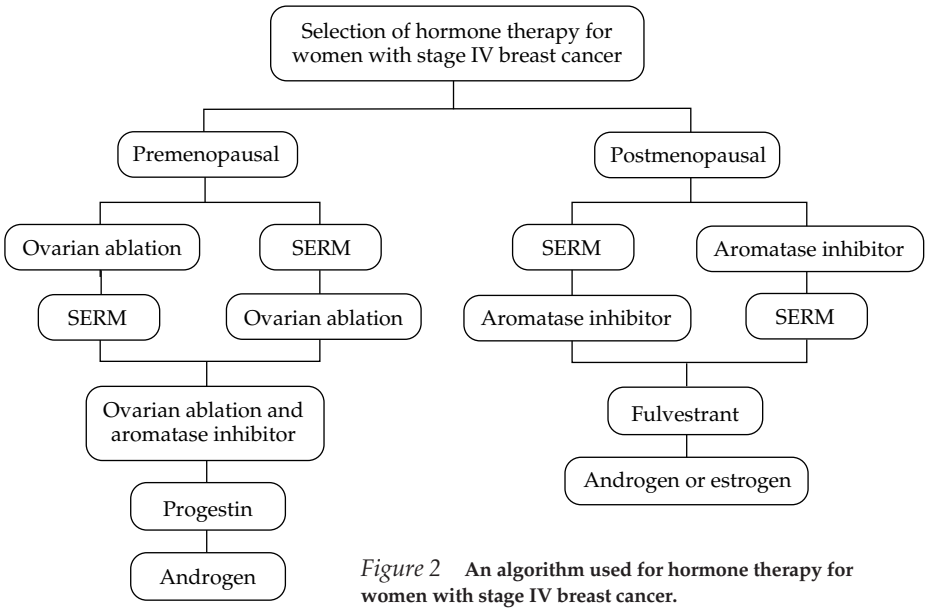


Figure 2 **An algorithm used for hormone therapy for women with stage IV breast cancer.**

Supportive Care

Because palliation of symptoms and prevention of disease complications are the primary goals of treatment of stage IV breast cancer, meticulous attention to supportive care is critical. Bone is the most common site of metastasis in breast cancer, and bony involvement can lead to significant morbidity. Several trials have demonstrated that regular administration of a bisphosphonate (zolendronate, pamidronate, or clodronate), in addition to hormone therapy or chemotherapy, can reduce pain and lower the incidence of several types of bony complications. These findings are derived from the ability of this class of drugs to inhibit osteoclast activity rather than from a specific antineoplastic effect. Thus, the FDA has approved the use of zolendronate and pamidronate as adjunctive therapy for women with lytic bony metastasis from breast cancer. A number of issues remain unanswered, including the optimal treatment interval and duration of therapy. Small trials have raised the possibility that clodronate may help prevent or delay the development of metastasis in women with early-stage breast cancer; randomized adjuvant trials are testing this possibility. Currently, evidence-based guidelines from the American Society of Clinical Oncology suggest that bisphosphonates be used only as part of a palliative program for women with metastatic bone disease.[81] Cardiotoxicity of doxorubicin may be reduced through the use of the cardioprotective agent dexrazoxane[82] or liposomal preparations. Judicious application of colony-stimulating factors or erythropoietin may help diminish symptoms of chemotherapy-related bone marrow hypoplasia. Whether these strategies are superior to reduction in chemotherapy dose or red blood cell transfusion remains to be determined. Adequate pain control should now be achievable in most patients through the use of sustained-release oral and transdermal narcotic preparations. Excellent antiemetics, such as the serotonin receptor inhibitors and a neurokinin-1 antagonist, are available to reduce chemotherapy-related emesis. Skillful application of these measures is critical in reducing morbidity from disease and its treatment.

Table 8 Newer Agents for Metastatic Breast Cancer Commercially Available in the United States[3]

Drug	*Category*	*Response Rate (%)*
Hormonal		
Toremifene (Fareston)	SERM	19–54
Anastrozole (Arimidex)	Aromatase inhibitor	12–39
Letrozole (Femara)	Aromatase inhibitor	12–39
Exemestane (Aromasin)	Aromatase inhibitor	44
Fulvestrant (Faslodex)	SERM	20
Cytotoxic		
Docetaxel (Taxotere)	Taxane	13–68
Paclitaxel (Taxol)	Taxane	13–68
Vinorelbine (Navelbine)	Vinca alkaloid	18–52
Capecitabine (Xeloda)	Thymidylate synthase inhibitor	20–36
Gemcitabine (Gemzar)	Purine analogue	25–46
Monoclonal antibody		
Trastuzumab (Herceptin)	Anti-HER-2	15–35

SERM—selective estrogen-receptor modulator

The author serves on the speaker board of AstraZeneca and is a member of the Data and Safety Monitoring Board for Eli Lilly and Company.

The drugs exemestane, letrozole, goserelin, leuprolide, and gefitinib have not been approved by the FDA for uses described in this chapter.

References

1. Jemal A, Tiwari RC, Murray T, et al: Cancer Statistics, 2004. CA Cancer J Clin 54:8, 2004

2. Peto R, Boreham J, Clarke M, et al: UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. Lancet 355:1822, 2000

3. Hortobagyi GN: Treatment of breast cancer. N Engl J Med 339:974, 1998

4. Struewing JP, Hartge P, Wacholder S, et al: The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336:1401, 1997

5. Malkin D, Li FP, Strong LC, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 250:1233, 1990

6. Liaw D, Marsh DJ, Li J, et al: Germline mutations of the PTEN gene in Cowden dis-

ease, an inherited breast and thyroid cancer syndrome. Nat Genet 16:64, 1997

7. Bhatia S, Robison LL, Oberlin O, et al: Breast cancer and other second neoplasms after childhood Hodgkin's disease. N Engl J Med 334:745, 1996

8. Chlebowski RT, Hendrix SL, Langer RD, et al: Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. JAMA 289:3243, 2003

9. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53,297 women with breast cancer and 100,239 women without breast cancer from 54 epidemiological studies. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 347:1713, 1996

10. Smith-Warner SA, Spiegelman D, Yaun SS, et al: Alcohol and breast cancer in women: a pooled analysis of cohort studies. JAMA 279:535, 1998

11. Laden R, Collman G, Iwamoto K, et al: 1,1-dichloro-2,2-bis(p-chlorophenyl)ethylene and polychlorinated biphenyls and breast cancer: combined analysis of five U.S. studies. J Natl Cancer Inst 93:768, 2001

12. Greenwald P: Role of dietary fat in the causation of breast cancer: point. Cancer Epidemiol Biomarkers Prev 8:3, 1999

13. Hunter DJ: Role of dietary fat in the causation of breast cancer: counterpoint. Cancer Epidemiol Biomarkers Prev 8:9, 1999

14. Fisher B, Costantino JP, Wickerham DL, et al: Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst 90:1371, 1998

15. Powles T, Eeles R, Ashley E, et al: Interim analysis of the incidence of breast cancer in the Royal Marsden Hospital tamoxifen randomised chemoprevention trial. Lancet 352:98, 1998

16. Veronesi U, Maisonneuve P, Rotmensz N, et al: Italian randomized trial among women with hysterectomy: tamoxifen and hormone-dependent breast cancer in high-risk women. J Natl Cancer Inst 95:160, 2003

17. Cuzick J, Forbes J, Edwards R, et al: First results from the International Breast Cancer Intervention Study (IBIS-1): a randomized prevention trial. Lancet 360:817, 2002

18. Cuzick J, Powles T, Veronesi U, et al: Overview of the main outcomes in breast cancer prevention trials. Lancet 361:296, 2003

19. Cauley JA, Norton L, Lippman ME, et al: Continued breast cancer risk reduction in postmenopausal women treated with raloxifene: 4-year results from the MOFRE trial. Multiple outcomes of raloxifene evaluation. Breast Cancer Res Treat 65:125, 2001

20. Veronesi U, De Palo G, Marubini E, et al: Randomized trial of fenretinide to prevent second breast malignancy in women with early breast cancer. J Natl Cancer Inst 91:1847, 1999

21. Meijers-Heijboer H, van Geel B, van Putten WLJ, et al: Breast cancer after prophylactic mastectomy in women with a BRCA1 or BRCA2 mutation. N Engl J Med 345:159, 2001

22. Thomas DB, Gao DL, Ray RM, et al: Randomized trial of breast self-examination in Shanghai: final results. J Natl Cancer Inst 94:1445, 2002

23. Kerlikowske K, Grady D, Rubin SM, et al: Efficacy of screening mammography: a meta-analysis. JAMA 273:149, 1995

24. Singletary SE, Allred C, Ashely P, et al: Revision of the American Joint Committee on Cancer staging system for breast cancer. J Clin Oncol 20:3628, 2002

25. Burstein HJ, Polyak K, Wong JS, et al: Ductal carcinoma in situ of the breast. N Engl J Med 350:1430, 2004

26. Fisher B, Land S, Mamounas E, et al: Prevention of invasive breast cancer in women with ductal carcinoma in situ: an update of the National Surgical Adjuvant Breast and Bowel Project experience. Semin Oncol 28:400, 2001

27. Houghton J, George WD, Cuzick J, et al: Radiotherapy and tamoxifen in women with completely excised ductal carcinoma in situ of the breast in UK, Australia, and New Zealand: randomised controlled trial. Lancet 362:95, 2003

28. Fisher R, Anderson S, Bryant J, et al: Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 347:1233, 2002

29. Veronesi U, Paganelli G, Viale G, et al: A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. N Engl J Med 349:546, 2003

30. Favourable and unfavourable effect on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 355:1757, 2000

31. Ragaz J, Jackson SM, Le N, et al: Adjuvant radiotherapy and chemotherapy in node-positive premenopausal women with breast cancer. N Engl J Med 337:956, 1997

32. Overgaard M, Hansen PS, Overgaard J, et al: Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. Danish Breast Cancer Cooperative Group 82b Trial. N Engl J Med 337:949, 1997

33. Ovarian ablation in early breast cancer: overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 348:1189, 1996

34. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 351:1451, 1998

35. Polychemotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet 352:930, 1998

36. Emens L, Davidson NE: Adjuvant hormonal therapy for premenopausal women with breast cancer. Clin Cancer Res 9:486S, 2003

37. Riggs BL, Hartmann LC: Selective estrogen-receptor modulators—mechanisms of action and application to clinical practice. N Engl J Med 348:618, 2003

38. Smith I, Dowsett M: Aromatase inhibitors in breast cancer. N Engl J Med 348:2431, 2003

39. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early breast cancer: first results of the ATAC (Arimidex, Tamoxifen Alone or in Combination) randomized trial. Lancet 359:2131, 2002

40. Goss PE, Ingle JN, Martino S, et al: A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. N Engl J Med 349:1793, 2003

41. Coombes RC, Hall E, Gibson LJ, et al: A randomized trial of exemestane after two to three years of tamoxifen therapy in postmenopausal women with primary breast cancer. N Engl J Med 350:1081, 2004

42. Winer EP, Hudis C, Burstein HJ, et al: American Society of Clinical Oncology Technology Assessment Working Group Update: use of aromatase inhibitors in the adjuvant setting. J Clin Oncol 21:2597, 2003

43. Bonadonna G, Valagussa P, Moliterni A, et al: Adjuvant cyclophosphamide, methotrexate, and fluorouracil in node-positive breast cancer: the results of 20 years of follow-up. N Engl J Med 332:901, 1998

44. Hutchins L, Green S, Ravdin P, et al: CMF versus CAF with and without tamoxifen in high-risk node-negative breast cancer patients and a natural history follow-up study in low-risk node-negative patients: first results of intergroup trial INT 0102 (abstr). Proc Am Soc Clin Oncol 17:1A, 1998

45. Henderson IC, Berry D, Demetri G, et al: Improved outcomes from adding sequential paclitaxel but not from escalating doxorubicin dose in an adjuvant chemotherapy regimen for patients with node-positive primary breast cancer. J Clin Oncol 21:976, 2003

46. Mamounas EP, Bryant J, Lembersky BC et al: Paclitaxel (T) following doxorubicin/cyclophosphamide (AC) as adjuvant chemotherapy for node-positive breast cancer: results from NSABP-B28 (abstr 12). Proc Am Soc Clin Oncol 22:4, 2003

47. Nabholtz JM, Pienkowski T, Mackey J, et al: Phase III trial comparing TAC (docetaxel, doxorubicin, cyclophosphamide) with FAC (5-fluorouracil, doxorubicin, cyclophosphamide) in the adjuvant treatment of node-positive breast cancer (BC) patients: interim analysis of the BCIRG 001 study (abstr 141). Proc Am Soc Clin Oncol 21:36a, 2002

48. Fisher B, Dignam J, Bryant J, et al: Five versus more than five years of tamoxifen therapy for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. J Natl Cancer Inst 93:682, 2001

49. Budman DR, Berry DA, Cirrincione CT, et al: Dose and dose intensity as determinants of outcome in the adjuvant treatment of breast cancer. The Cancer and Leukemia Group B. J Natl Cancer Inst 90:1205, 1998

50. Fisher B, Anderson S, Wickerham DL, et al: Increased intensification and total dose of cyclophosphamide in a doxorubicin-cyclophosphamide regimen for the treatment of primary breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-22. J Clin Oncol 15:1858, 1997

51. Fisher B, Anderson S, DeCillis A, et al: Further evaluation of intensified and increased total dose of cyclophosphamide for the treatment of primary breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-25. J Clin Oncol 17:3374, 1999

52. Tallman MS, Gray R, Robert NJ, et al: Conventional adjuvant chemotherapy with or without high-dose chemotherapy and autologous stem cell transplantation in high-risk breast cancer. N Engl J Med 349:17, 2003

53. Fisher B, Bryant J, Wolmark N, et al: Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 16:2672, 1998

54. Recht A, Come SE, Henderson IC, et al: The sequencing of chemotherapy and radiation therapy after conservative surgery for early-stage breast cancer. N Engl J Med 334:1356, 1996

55. Fisher B, Dignam J, Wolmark N, et al: Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. J Natl Cancer Inst 89:1673, 1997

56. Fisher B, Redmond C, Legault-Poisson S, et al: Postoperative chemotherapy and tamoxifen compared with tamoxifen alone in the treatment of positive-node breast cancer patients aged 50 years and older with tumors responsive to tamoxifen: results from the National Surgical Adjuvant Breast and Bowel Project B-16. J Clin Oncol 8:1005, 1990

57. Albain K, Green S, Osborne K, et al: Tamoxifen (T) versus cyclophosphamide, Adriamycin and 5-FU plus either concurrent or sequential T in postmenopausal, receptor(+), node(+) breast cancer: a Southwest Oncology Group phase III intergroup trial (SWOG-8814, INT-0100) (abstr). Proc Am Soc Clin Oncol 16:128A, 1997

58. Tormey DC, Gray DC, Abeloff MD, et al: Adjuvant therapy with a doxorubicin regimen and long-term tamoxifen in premenopausal breast cancer patients: an Eastern Cooperative Oncology Group trial. J Clin Oncol 10:1848, 1992

59. Fisher B, Anderson S, Tan-Chiu E, et al: Tamoxifen and chemotherapy for axillary node-negative breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-23. J Clin Oncol 19:931, 2001

60. Davidson NE, O'Neill AO, Vukov A, et al: Effect of chemohormonal therapy in premenopausal node (+), receptor (+) breast cancer: an Eastern Cooperative Oncology Group phase III Intergroup trial (E5188, INT-0101) (abstr). Proc Am Soc Clin Oncol 18:67a, 1999

61. AdjuvantOnline.com. Adjuvant! Inc., 2003
http://www.adjuvantonline.com

62. Adjuvant therapy for breast cancer. Mayo Foundation for Medical Science and Research, MayoClinic.com, Rochester, Minnesota, 2004
http://www.mayoclinic.com/takecharge/healthdecisionguides/avt/index.cfm

63. Goldhirsch A, Wood WC, Gelber RD, et al: Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. J Clin Oncol 21:3357, 2003

64. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. National Institutes of Health Consensus Development Panel. J Natl Cancer Inst Monogr 30:5, 2002

65. Practice guidelines in oncology: guidelines for treatment of cancer by site. National Comprehensive Cancer Center Network, Jenkintown, Pennsylvania, 2004 http://www.nccn.org/physician_gls/f_guidelines.html

66. Shapiro CL, Recht A: Side effects of adjuvant treatment of breast cancer. N Engl J Med 344:1997, 2001

67. Impact of follow-up testing of survival and health-related quality of life in breast cancer patients: a multicenter randomized controlled trial. The GIVIO Investigators. JAMA 271:1587, 1994

68. Rosselli Del Turco M, Palli D, Cariddi A, et al: Intensive diagnostic follow-up after treatment of primary breast cancer: a randomized trial. National Research Council Project on Breast Cancer follow-up. JAMA 271:1593, 1994

69. Recommended breast cancer surveillance guidelines. American Society of Clinical Oncology. J Clin Oncol 15:2149, 1997

70. Greenberg PA, Hortobagyi GN, Smith TL, et al: Long-term follow-up of patients with complete remission following combination chemotherapy for metastatic breast cancer. J Clin Oncol 14:2197, 1996

71. Muss HB, Case LD, Richards F 2nd, et al: Interrupted versus continuous chemotherapy in patients with metastatic breast cancer. The Piedmont Oncology Association. N Engl J Med 325:1342, 1991

72. Coates A, Gebski V, Bishop JF, et al: Improving the quality of life during chemotherapy for advanced breast cancer: a comparison of intermittent and continuous treatment strategies. N Engl J Med 317:1490, 1987

73. Fossati R, Confalonieri C, Torri V, et al: Cytotoxic and hormonal treatment for metastatic breast cancer: a systematic review of published randomized trials involving 31,510 women. J Clin Oncol 16:3439, 1998

74. Bishop JH, Dewar J, Toner GC, et al: Initial paclitaxel improves outcome compared with CMFP combination chemotherapy as front-line therapy in untreated metastatic breast cancer. J Clin Oncol 17:2355, 1999

75. Nabholtz JM, Senn HJ, Bezwoda WR, et al: Prospective randomized trial of docetaxel versus mitomycin plus vinblastine in patients with metastatic breast cancer progressing despite previous anthracycline-containing chemotherapy. 304 Study Group. J Clin Oncol 17:1413, 1999

76. Cobleigh MA, Vogel CL, Tripathy D, et al: Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. J Clin Oncol 17:2639, 1999

77. Vogel CL, Cobleigh MA, Tripathy D, et al: Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER-2 overexpressing metastatic breast cancer. J Clin Oncol 20:719, 2003

78. Slamon D, Leyland-Jones B, Shak S, et al: Use of chemotherapy plus a monoclonal antibody against HER-2 for metastatic breast cancer that overexpresses HER-2. N Engl J Med 344:783, 2001

79. Cobleigh MA, Langmuir VK, Sledge GW et al: A phase I/II dose-escalation trial of bevacizumab in previously treated metastatic breast cancer. Semin Oncol 30:117, 2003

80. Robertson JF, Osborne CK, Howell A, et al: Fulvestrant versus anastrozole for the treatment of advanced breast carcinoma in postmenopausal women: a prospective combined analysis of two multicenter trials. Cancer 98:229, 2003

81. Hillner BE, Ingle JN, Chelbowski RT, et al: American Society of Clinical Oncology 2003 update on the role of bisphosphonates and bone health issues in women with breast cancer. J Clin Oncol 21:4042, 2003

82. Schuchter LM, Hensley ML, Meropol NJ, et al: 2002 update of recommendations for the use of chemotherapy and radiotherapy protectants: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol 20:2895, 2002

VIII LUNG CANCER

Jeffrey Crawford, m.d.

Definition and Classifications

Bronchogenic carcinoma of the lung—lung cancer—comprises a group of malignant neoplasms that arise from bronchial epithelium. The four major pathologic cell types of lung cancer are small cell carcinoma, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma. Because they have overlapping clinical behavior and response to treatment, adenocarcinoma, squamous cell carcinoma, and large cell carcinoma are generally grouped together in the category of non–small cell lung cancer (NSCLC). NSCLC represents 75% to 80% of all cases of lung cancer. Classification systems for the four major types of lung cancer have been formulated by the World Health Organization, the Armed Forces Institute of Pathology, and the Working Party for Lung Cancer [*see Table 1*].

Epidemiology and Etiology

In the United States, lung cancer is the second most common cancer in both men and women, surpassed only by prostate cancer in men and breast cancer in women. For 2003, a total of 169,400 new cases were predicted [*see Table* 2]. Lung cancer was expected to constitute 14% of all cancer diagnoses in men and 12% of those in women. However, lung cancer is the leading cause of cancer deaths, accounting for 31% and 25% of all cancer-related deaths in men and women, respectively. For 2003, expected deaths from lung cancer were 154,900 [*see Table 2*].[1]

The epidemiology of lung cancer in the United States directly reflects patterns in cigarette smoking, albeit with a 10- to 15-year lag time.[2] Over recent decades, the prevalence of cigarette smoking in men has decreased from nearly 50% to approximately 25%, and the incidence of lung cancer in men has declined somewhat. During that same period, the prevalence of cigarette smoking in women has declined only 11%, to approximately 22%, and the incidence of lung cancer in women is only now leveling off.

In men, the incidence of lung cancer peaked in 1984, at 86.5 per 100,000 population, and by 1996 had declined to 70 per 100,000 population. For women, the incidence in 1996 was 42.3 per 100,000 population. Since 1987, more women have died each year from lung cancer than from breast cancer, and the margin between the two diseases continues to widen. Estimates suggest that in 2003, over 50% more women died of lung cancer than of breast cancer.[1]

Unfortunately, cigarette smoking became increasingly popular in teenagers in the 1990s. In the United States, the prevalence of cigarette smoking in high-school students increased during the 1990s, peaking during 1996 to 1997, then began a gradual decline.[3] The popularity of smoking varied by ethnicity and race. In a 1999 survey of high-school students, smoking rates were 15.8% in blacks, 25.8% in Hispanics, and 32.8% in whites.[4]

smoking cessation and lung cancer

Cigarette smoking continues to contribute to the risk of lung cancer long after a person has stopped smoking. The American Cancer Society evaluated this relationship in a 6-year prospective study involving more than 900,000 persons.[5] This study included persons who had never smoked, current smokers, and former smokers. As expected, the risk of dying of lung cancer was lower in patients who had quit smoking early in life than in those who quit later on, and the risk was significantly lower in those who quit than in those who did not. In a person who smoked 26 cigarettes a day starting at 17 years of age and stopped smoking between the ages of 30 and 49, the risk of death from lung cancer is slightly greater than that for persons who never smoked. For a person quitting smoking between the ages of 50 and 64, the risk of death from lung cancer plateaus at the risk level at the time of quitting and remains level until about the age of 75, when the risk appears to increase further. In this model, the annual lung cancer mortality for current smokers at age 75 is 1% for men and 0.5% for women, which is approximately 20 times higher than that of nonsmokers. Nonsmokers (i.e., persons with a lifetime exposure of less than 100 cigarettes) have a rela-

Table 1 Major Classifications of Lung Cancer

System	*Non–Small Cell Lung Cancer*			*Small Cell† Cancer*
	Squamous Cell Carcinoma*	*Adenocarcinoma*	*Large Cell Carcinoma*	
World Health Organization, No. 2 (WHO–No. 2)	Spindle cell variant	Acinar Papillary Bronchioloalveolar Solid carcinoma with mucin	Giant cell Clear cell	Oat cell Intermediate cell Combined
Armed Forces Institute of Pathology (AFIP)	Well differentiated Moderately differentiated Poorly differentiated	Well differentiated Moderately differentiated Poorly differentiated Bronchioloalveolar	Undifferentiated Giant cell Clear cell	Lymphocyte-like (oat cell) Polygonal (intermediate) Combined (usually squamous)
Working Party for Lung Cancer (WPLC)	Well differentiated Moderately differentiated Poorly differentiated	Well differentiated Moderately differentiated Poorly differentiated Bronchioloalveolar/papillary	With mucin production With stratification Giant cell Clear cell	Lymphocyte-like (oat cell) Intermediate cells (fusiform, polygonal, others)

Note: both the WHO–No. 2 and AFIP systems have a fifth category, adenosquamous cell carcinoma; benign lesions, dysplasia, carcinoma in situ, carcinoid tumors, soft tissue sarcomas, and other respiratory tract lesions, which account for only a few percent of all lung cancers, are not included in this table.
*For the WPLC system, the classification is epidermoid.
†For the WPLC system, the classification is small cell anaplastic.

Table 2 Epidemiology of Lung Cancer in the United States, 2003[1]

	New Cases	*Deaths*
Men	90,200	89,200
Women	79,200	65,700
Total	169,400	154,900

tive risk of lung cancer of 0.05 or less as compared with current smokers. For former smokers, the relative risk of lung cancer death depends on the age of smoking cessation. The risk was 0.45 for smokers who quit in their early 60s, 0.2 for those who stopped smoking in their early 50s, and 0.1 for those who stopped smoking in their 30s. All available data indicate that the lung cancer risk for former smokers is still consistently greater than for those who never smoked. Stopping smoking at any age can reduce lung cancer mortality, but the risk reduction is much greater for smokers who quit at a younger age.

In addition to age effects, there is a dose-response relationship for smoking and lung cancer. The risk for lung cancer increases with the duration of smoking and the number of cigarettes smoked. Earlier age of starting to smoke, deeper inhalation, and use of cigarettes that are unfiltered or have a high tar and nicotine content also increase the risk of lung cancer. In the current United States population over the age of 50, 23% are current smokers and 35% are former smokers. Because both groups remain at elevated risk for lung cancer for their lifetimes, clinicians should take an accurate quantitative smoking history in all patients.

GENETIC SUSCEPTIBILITY AND MOLECULAR MECHANISMS

The risk of lung cancer is affected by genetic susceptibility. Women smokers may be at higher risk for the development of lung cancer than men with a similar smoking history. Furthermore, lung cancer mortality appears to be higher in African Americans.

Mechanisms for genetic susceptibility to lung cancer include genes that govern smoking behavior, which affect dopamine reward mechanisms related to nicotine and nicotine metabolism, as well as gender[6]; individual capacity for carcinogen metabolism; germline mutations coding for dysfunctional genes; and capacity to repair DNA damage from carcinogens. Several genetic abnormalities have been associated with lung cancers [*see Table 3*].

LUNG CANCER IN NONSMOKERS

Given the dominant role of cigarette smoking in the etiology of lung cancer, determining the risk posed by other substances is difficult.[7] As many as 25% of cases of lung cancer in nonsmokers may result from second-hand tobacco smoke. A small percentage of lung cancers result from occupational exposure to carcinogens, including asbestos, arsenic, cadmium, chromium, radiation, radon, and chemicals such as chromoethyl ether. Heavy residential exposure to radon may be synergistic with cigarette smoking in promoting lung cancer, but the risk from residential radon for nonsmokers remains unclear.[8]

Pathophysiology and Pathogenesis

The prevalences of histologic subtypes of lung cancer in men and women have changed in ways that mirror the changes in smoking habits. In the early studies that established the association between smoking and lung cancer, cigarettes were unfiltered, most of the participants were men, and squamous cell carcinoma was the most common cell type. Now, with filtered cigarettes widely popular and larger numbers of women smoking, adenocarcinoma is the most common type of lung cancer in both young men and women. This changing pattern of histology correlates temporally with the change from unfiltered to filtered cigarettes and with reductions in the tar and nicotine content of cigarettes. Those changes in cigarette manufacturing have led to deeper inhalation of smoke into the lungs, which exposes the distal airways more heavily to the carcinogenic influences of tobacco smoke. Other factors likely play a part as well. In nonsmokers, adenocarcinomas are the most common histologic type of lung cancer.

The initiation of carcinogenesis from cigarette smoke is related to a complex mixture of carcinogens and tumor promoters combined with the delivery vehicle of inhalation. Serial studies of bronchial epithelium in smokers demonstrate an evolution from dysplasia to metaplasia to neoplastic changes.[9-12] Each stage has been associated with a number of genetic alterations, and the pivotal mechanisms are a topic of intense investigation. Factors associated with genetic susceptibility have yet to be identified and may emerge from studies of lung cancer in nonsmokers. Thus, the main clinical criterion for susceptibility remains a history of current or former smoking.

Table 3 Selected Molecular Genetic Abnormalities Associated with Lung Cancer

Abnormal Genes	*Mutation*	*Frequency of Abnormal Expression (%)*	
		NSCLC	*SCLC*
Oncogenes			
K-*ras*	Point mutation (codon 12)	30	Not reported
myc family	DNA amplification/overexpression	10	10–40
HER-2/*neu*	Increased expression of p185*neu*	25	Not reported
Tumor suppressor genes			
p53	Deletion Point mutation Overexpression	50	80
Rb	Deletion	15	> 90
3p	Deletion	50	90

NSCLC—non–small cell lung cancer SCLC—small cell lung cancer

Prevention

PRIMARY PREVENTION

Given that 87% of cases of lung cancer occur in smokers and that the risk of lung cancer is lower by at least 20-fold in persons who have never smoked, the obvious strategy for primary prevention is to keep young persons from starting to smoke and to promote smoking cessation in smokers of all ages [*see CE:III Reducing Risk of Injury and Disease*]. Although public health measures that discourage smoking in public places and in the workplace, as well as the development of negative societal attitudes toward smoking, are helpful in reducing the prevalence of smoking in adults, progress against smoking has been slow and teenage smoking rates remain unacceptably high.

SECONDARY PREVENTION

The use of nutritional supplements by smokers as a strategy to reduce lung cancer was suggested by an epidemiologic association of lower serum levels of β-carotene, vitamin E, and retinoids with a higher risk of lung cancer.[13] Unfortunately, in clinical trials, these agents did not reduce lung cancer risk.

One of the best known trials, the CARET (β-Carotene and Retinol Efficacy Trial), comprised over 18,000 smokers of both sexes randomized to receive a retinoid drug, retinol palmitate, in combination with β-carotene or placebo.[14] In this trial, patients who received β-carotene and retinol palmitate had a higher rate of development of lung cancer (relative risk = 1.36) and higher lung cancer mortality (relative risk = 1.59). In another placebo-controlled trial, from Finland, that studied the effects of vitamin E and β-carotene, smokers who received β-carotene were more likely to develop lung cancer (relative risk = 1.16). Vitamin E produced no effect. The risk of harm from β-carotene in this trial was more pronounced in heavy smokers. In a placebo-controlled trial in patients with resected stage I NSCLC, the use of 13-*cis*-retinoic acid increased the rate of lung cancer recurrence and mortality in patients who continued to smoke. Thus, at present, no evidence supports recommending vitamins to prevent lung cancer, and there is some evidence that β-carotene and retinoids may have harmful effects in smokers, as well as in persons with occupational exposure to asbestos.[15]

Diagnosis

SCREENING

Most patients with lung cancer present with advanced inoperable disease. Screening for detection of lung cancer at an earlier stage is therefore an attractive idea, especially because persons at high risk for lung cancer can be readily identified by a smoking history.

Early studies of screening produced disappointing results. Randomized trials of screening, conducted in the United States and in the former Czechoslovakia, suggested that chest x-ray alone was not a satisfactory screening tool to detect early lung cancer tumors.[16] Curable tumors are often too small or indistinct to be detected on a standard chest x-ray.

Spiral CT scanning may be a more sensitive technique for lung cancer screening. With this technique, radiologists obtain a low-resolution image of the entire thorax in a single breathhold, with low radiation exposure and relatively rapid throughput compared with standard CT scans. A number of studies have demonstrated the feasibility of spiral CT scanning in screening for lung cancer. In the Early Lung Cancer Action Project (ELCAP), 1,000 asymptomatic persons older than 60 years with a smoking history of 10 or more pack-years underwent both spiral CT and chest x-ray.[17] CT detected malignant nodules in 2.7% of the patients, compared with 0.6% by chest x-ray. Benign nodules were detected at a rate of 20.6% by CT versus 6.1% by chest x-ray, so careful follow-up is critical for avoiding unnecessary biopsy. A Mayo Clinic study of spiral CT[18] also demonstrated enhanced detection of malignant nodules, most of which were early-stage lung cancer, but an even higher yield of benign nodules (60%), which emphasizes the potential drawback of this technique.

At present, no data from randomized trials exist to allow an evidence-based recommendation either for or against lung cancer screening. Despite encouraging results from nonrandomized trials, several issues remain to be addressed, including lead-time bias, generalization to a broader population, application to younger patients at lower risk of lung cancer, and long-term benefit in terms of lower lung cancer mortality. Furthermore, a decision and cost-effectiveness analysis has suggested that the cost of implementing such a strategy would be substantial.[19] Currently, spiral CT screening cannot be recommended except in the context of a clinical trial. Other new technologies that deserve consideration as potential screening methods include analysis of sputum cytology by molecular markers and localization of tumors by fluorescence bronchoscopy.[20]

The National Cancer Institute is currently enrolling patients in the National Lung Screening Trial (NLST), a randomized, controlled trial that will compare standard chest x-rays with spiral CT as a screening method for lung cancer.[21] The NLST will enroll 50,000 current or former smokers between the ages of 55 and 74 years at clinical trial sites throughout the United States. Study participants will receive either a chest x-ray or a spiral CT once a year for 3 years and will then undergo monitoring until 2009. The researchers will be looking for a reduction in mortality of 20% or more with either modality. In addition to the screenings, some NLST centers will test for biologic markers that may have potential for screening.

CLINICAL MANIFESTATIONS AND LABORATORY STUDIES

The signs and symptoms of lung cancer vary with the anatomic location of the tumor, its extension into surrounding structures, metastatic spread, and the systemic effects of paraneoplastic syndromes. Unfortunately, only 6% of patients with lung cancer are asymptomatic at the time of diagnosis. The remainder of the patients present with symptoms resulting from regional spread of the tumor, mediastinal lymph node involvement, or distant metastases.

Pulmonary Manifestations

The most common manifestation of the primary tumor is cough, which results from endobronchial erosion and irritation. Others are, in decreasing order of frequency, dyspnea, chest pain, hemoptysis, and postobstructive pneumonia or pneumonitis [*see Table 4*]. Centrally located tumors also typically cause stridor, wheezing, hemoptysis, dyspnea, or chest pain, often central in location. Occlusion of the airway by a tumor can lead to a postobstructive infiltrate or pneumonia. Large tumors may cavitate and present as a lung abscess.

Manifestations of Intrathoracic Disease

Intrathoracic extension of the tumor or spread to mediastinal

Table 4 Common Signs and Symptoms of Lung Cancer at Diagnosis

Site of Tumor Involvement	*Signs or Symptoms*	*Percentage of Patients Affected*
Pulmonary	Cough	50–75
	Dyspnea	30–40
	Chest pain	25–40
	Hemoptysis	15–30
	Pneumonia/pneumonitis	10–25
Intrathoracic	Hoarseness	< 10
	Dysphagia	< 10
	Facial/arm swelling	< 10
	Shoulder/arm pain	< 10
	Pleural/chest wall pain	< 10
	Pleural/pericardial effusion	< 10
	Paraneoplastic syndromes	< 10
Extrathoracic	Anorexia/weight loss	30–50
	Generalized weakness	20–40
	Bone pain	20–30
	Liver abnormalities	10–20
	Headache/CNS abnormalities	5–15
	Flank pain	< 10
	Other (e.g., subcutaneous nodule, distant lymph nodes)	< 10

lymph nodes may produce a variety of symptoms [*see Table 4*]. Although individually these symptoms occur in fewer than 10% of patients with lung cancer, collectively they represent significant complications of locally advanced NSCLC, either at diagnosis or during the subsequent disease course. Hoarseness may result from invasion of the recurrent laryngeal nerve and resultant vocal cord paralysis. Dysphagia may be a sign of compression of the esophagus. Extensive tumor involvement of the right mediastinal lymph nodes often results in the superior vena cava syndrome, which is characterized by plethoric appearance; distention of the venous drainage of the arm and neck; and edema of the face, neck, and arms. Vena caval obstruction usually progresses gradually, allowing the development of collateral venous drainage that may be detected on physical examination.

Shoulder and arm pain from superior sulcus (Pancoast) tumor syndrome is a commonly misdiagnosed sign of lung cancer. The pain results from local extension of a tumor in the apex of the lung, with involvement of the eighth cervical and first thoracic nerves. Unfortunately, this condition is often mistaken for arthritis. In many cases, careful physical examination will identify ipsilateral Horner syndrome, which is characterized by ptosis, meiosis, and anhydrosis. The Horner syndrome is related to paravertebral extension and sympathetic nerve involvement of the tumors.

Pleuritic pain and chest wall pain occur most commonly in patients with primary tumors in the lung periphery that spread to the pleura and, in some cases, extend directly to the chest wall. Associated pleural effusion may occur in such cases; large effusions may cause dyspnea. Malignant pericardial effusions may also develop and can cause cardiac tamponade.

Paraneoplastic syndromes A minority of lung cancer patients present with paraneoplastic manifestations. The biology of these syndromes remains poorly characterized, but the syndromes appear to be cytokine-mediated responses to antigens from the intrathoracic lung tumor, rather than the result of distant spread of cancer.

The most common paraneoplastic feature associated with lung cancer is clubbing of the fingers from periosteal swelling of the distal phalanges, which may occur in 5% to 15% of patients. In a small percentage of patients, clubbing may be part of a symptomatic hypertrophic osteoarthropathy. These patients often complain of a distal symmetrical arthritis that most commonly involves the ankles or knees but can also involve the wrists, elbows, and other joints. Misdiagnosis of this condition as a strictly rheumatologic phenomenon often results in delayed recognition of the underlying neoplasm.

Although weight loss and fatigue are commonly an indication of distant metastasis, they can also represent a paraneoplastic phenomenon that occasionally occurs even with early-stage tumors. Especially in patients with small cell lung cancer (SCLC), paraneoplastic manifestations can also take the form of specific neurologic syndromes, such as the Lambert-Eaton syndrome. These patients present with muscle weakness, a variety of peripheral neuropathies, and central nervous system involvement such as subacute cerebellar degeneration or limbic encephalitis.

Another category of neoplastic syndromes relates to aberrant hormone or peptide production by lung cancer tumor cells. The most common of these is hyponatremia secondary to production of antidiuretic hormone (SIADH). Hypercalcemia can result from tumors that secrete parathyroid hormone; and Cushing syndrome, from tumors that secrete adrenocorticotropic hormone. In general, these hormonal syndromes are more common in SCLC than in NSCLC, because of the neuroendocrine nature of SCLC. However, hypercalcemia can have a range of causes—including both remote effects and direct interactions between tumor and bone—and is much more common in NSCLC than in SCLC.

Manifestations of Extrathoracic Disease

Extrathoracic manifestations of lung cancer relate to the extent and site of distant spread [*see Table 4*]. The most common of these are anorexia, weight loss, and fatigue. Bone pain commonly accompanies metastasis to bone, but with the increased use of imaging, asymptomatic bony metastases are commonly found. Liver abnormalities may be detected on clinical examination or on laboratory or imaging studies, but they are generally asymptomatic. The frequency of CNS involvement varies with the extent of other known disease, with a low incidence in patients who have no nodal spread of cancer. However, in patients with other signs of mediastinal or distant involvement, the incidence of occult brain metastases is in the range of 5% to 15%, even in NSCLC.

Occasionally, flank pain will be a presenting feature of adrenal metastases. Although flank pain occurs in fewer than 10% of patients, the adrenal gland is the most frequent site of distant metastatic spread of lung cancer, as detected by CT imaging. Adrenal insufficiency is an unusual but potentially fatal complication of adrenal metastasis from lung cancer, and it is often overlooked because the weight loss and fatigue it causes are common features in lung cancer patients. In selected cases, an adrenal stimulation test may identify patients with limited reserve who may benefit from steroid-replacement therapy.

Patients who have bronchial carcinoid tumors metastatic to liver or other sites may experience the carcinoid syndrome. This dramatic but rare syndrome is characterized by episodic

Table 5 Common Causes of a Solitary Pulmonary Nodule

Malignant	Bronchogenic carcinoma Adenocarcinoma Squamous cell Large cell Metastatic cancers
Benign	Noninfectious granuloma Sarcoidosis Wegener granulomatosis Infectious granuloma Tuberculosis Histoplasmosis Coccidioidomycosis Nontuberculous mycobacteria Benign tumors Hamartoma Lipoma Fibroma
Congenital	Arteriovenous malformation Bronchogenic cyst
Miscellaneous	Rheumatoid nodule Amyloidosis Pulmonary infarction

flushing that may be associated with abdominal pain, diarrhea, and wheezing.

CLINICAL STAGING

When the results of the clinical examination and chest x-ray indicate early-stage lung cancer, imaging studies may be limited to a chest CT. However, in patients who have clinical, laboratory, or radiologic signs of regional tumor spread, a search for occult bone and CNS metastases is warranted. For patients with suspected metastatic disease, the standard imaging evaluation should include a chest CT with images through the adrenal glands, a bone scan, and a CT or MRI scan of the brain.

The role of PET scanning in the evaluation of lung cancer patients is currently under study. F-18 fluorodeoxyglucose (FDG) uptake is greater in malignant cells than in normal, benign cells. Several series have suggested that FDG-PET imaging can be very useful in determining whether abnormalities seen on CT—particularly in the adrenal gland and bone—likely represent metastatic disease.[22] The sensitivity and specificity for mediastinal lymph node metastases is still being clarified. PET scans are also useful for evaluation of solitary pulmonary nodules, with a sensitivity of 90% to 95% and specificity of 80% to 100% for the detection of cancer. Because PET imaging can detect unsuspected metastatic disease in 11% to 14% of patients and thus help avoid futile surgery in these cases, Medicare in the United States provides coverage for FDG-PET for the staging of NSCLC.[23] Meanwhile, PET technology is evolving rapidly, improving its sensitivity for the detection of smaller lesions. Although PET scanning can detect lesions between 0.5 and 1.0 cm, most series have limited the analysis to lesions greater than 1.0 cm. In addition, techniques that incorporate simultaneous CT and PET image analysis are currently being developed.

The diagnostic approach used to confirm the presence of lung cancer and determine the subtype depends on the clinical stage at presentation. In patients with advanced disease, a needle biopsy of a metastatic site (e.g., liver, bone, or a subcutaneous nodule) is often the best choice, providing both confirmation of the diagnosis and identification of the disease stage. In patients with no extrathoracic signs of cancer, the choice of initial diagnostic procedure often depends on whether the patient is likely to be a candidate for surgery. For the surgical population, the primary diagnostic procedure in most patients should be a bronchoscopy and mediastinoscopy by the thoracic surgeon, to determine the type and stage of cancer with respect to mediastinal lymph node involvement, as well as to determine resectability. For patients with more peripheral lung masses or solitary pulmonary nodules, the procedure of choice for confirming the presence of cancer and the prospects for definitive surgery is an initial needle biopsy performed under radiologic guidance or resection by video-assisted thoracoscopic surgery (VATS). In patients with solitary pulmonary nodules, biopsy may show that the cause is not cancer but rather a benign tumor or an inflammatory, infectious, or congenital disorder [*see Table 5*].

For patients who have evidence of bulky intrathoracic disease but who are not likely to be surgical candidates, the preferred method of evaluation is bronchoscopy. During the bronchoscopy, the surgeon may perform brushings, washings, or transbronchial biopsies of the primary lesion or any associated central mediastinal lymph nodes.

Patients presenting with pleural effusions can be evaluated by diagnostic thoracentesis. In some cases, VATS can provide both definitive diagnosis and management of pleural effusions.

In addition to the clinical stage, the so-called physiologic stage of the patient is also important for determining which diagnostic strategy is best. In patients who are not candidates for surgery because of constraints such as severe comorbid disease or limited pulmonary reserve, transthoracic needle biopsy or bronchoscopy alone may suffice.

Improvements in needle-biopsy techniques have reduced the complications of these procedures, and improvements in cytology have enhanced its diagnostic power. Although these cytologic exams often cannot differentiate subtypes of NSCLC, they are 95% accurate in distinguishing SCLC from NSCLC. Definitive staging is particularly important in patients with NSCLC because of the evolution in treatment strategies for both operable (stage I to IIIA) and inoperable (stage IIIB) cases. Definitive surgical staging with bronchoscopy and mediastinoscopy remains the preferred approach for most patients with apparent early-stage lung cancer who would be candidates for surgery. If cervical mediastinoscopy is performed, nodal sampling should include the upper paratracheal (level 2), lower paratracheal (level 4), and subcarinal (level 7) stations [*see* Surgical Staging, *below*]. For patients with a left upper lobe tumor, an anterior mediastinal approach may also be indicated to sample the AP window lymph nodes (level 5).

SURGICAL STAGING

Cancer stage is by far the most important prognostic factor in lung cancer. Histology (i.e., SCLC versus NSCLC) may influence choice of treatment options. Survival rates for patients with the same stage of lung cancer are quite similar, regardless of whether they have SCLC or NSCLC. Other characteristics that can affect outcome are patient characteristics such as performance status, recent weight loss, and significant comorbid conditions. In addition, studies suggest that stage for stage, outcome with both SCLC and NSCLC is better for women than

for men. As with other cancers, advanced age may have an adverse effect on outcome, but age per se seems to be less important than the comorbid conditions that are more common in the elderly.

Staging of lung cancer is by the TNM (tumor, node, metastases) classification [*see Figure 1*].[24,25] It is based on the size, location, and regional extension of the primary tumor; on the location of regional malignant lymph nodes that drain the region; and on the absence or presence of distant metastases. T1 and T2 tumors are operable tumors differentiated predominantly by size. T1 tumors are 3 cm or less in their greatest dimension, surrounded by lung or visceral pleura, and without bronchoscopic evidence of invasion more proximal than the lobar bronchus. T2 tumors have any one of the following characteristics: size greater than 3 cm, main bronchus involvement, location 2 cm or more distal to the carina, invasion of the visceral pleura, or association with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung.

T3 tumors are tumors of any size that directly invade the chest wall, diaphragm, mediastinal pleura, parietal pleura, or pericardium; are located in the main bronchus less than 2 cm distal to the carina but do not involve the carina; or are associated with atelectasis or obstructive pneumonitis of the entire lung. T3 tumors can be considered marginally operable but require a more extensive operation that may involve removal of the chest wall or pericardium or, for more proximal tumors, a sleeve resection.

T4 tumors are grossly inoperable because they invade the mediastinum, heart, great vessels, trachea, esophagus, a vertebral body, or the carina. Tumors are also classified as T4 if they are associated with a malignant pleural or pericardial effusion or with satellite tumor nodules within the same lobe as the primary tumor.

Lymph node status is determined as N0 (no lymph node involvement), N1 (metastases to the lymph nodes within the confines of the lung), and N2 or N3 (extrapulmonary metastases). N2 represents involvement of ipsilateral mediastinal lymph nodes, whereas N3 represents involvement of contralateral lymph nodes or more distant nodes, including hilar or supraclavicular nodes.

N1 and N2 nodes are further denoted by specific location (station) [*see Figure 2*].[26] Other than level-10 hilar nodes, which may be enlarged on CT, N1 nodal involvement is generally not suspected until it is discovered at the time of surgery. Although N2 and N3 nodes can be suspiciously enlarged on CT, 40% of nodes greater than 2 cm are enlarged because of inflammation, and 10% of normal-sized nodes contain malignancy. Thus, mediastinoscopy is essential for providing pathologic definition of nodal involvement so that treatment options can be finalized.

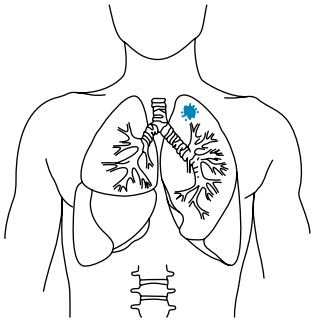
Treatment

NSCLC

The treatment of NSCLC is based on the stage of disease, as determined by the TNM staging system. For stage I or stage II disease, surgical resection is the standard treatment. Stage III disease is treated with definitive radiation therapy and chemotherapy; in addition, a subset of patients with stage IIIA disease have been shown to have improved outcome with the addition of sur-

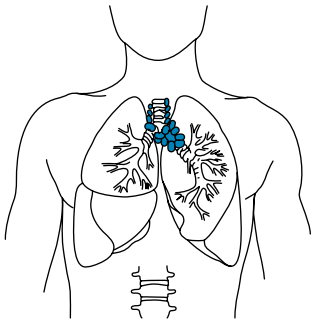
Tumor (T)

T1 ≤ 3 cm
T2 > 3 cm; visceral pleura invasion
T3 Direct extension to chest wall, mediastinal pleura, or pericardium
T4 Malignant pleural effusion, superior vena cava syndrome, or involvement of the heart, great vessels, trachea, esophagus, or vertebral bodies



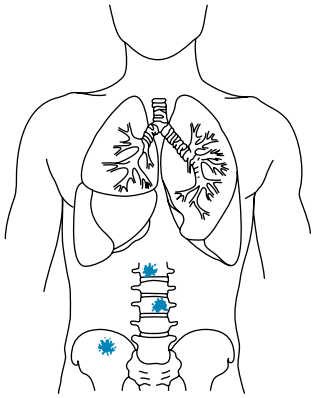
Nodes (N)

N0 Negative regional lymph nodes
N1 Peribronchial or ipsilateral hilar nodes
N2 Ipsilateral mediastinal nodes
N3 Contralateral hilar mediastinal nodes; any supraclavicular nodes



Metastasis (M)

M0 No distant metastases
M1 Distant metastases



International System for Staging Lung Cancer

Stage	*TNM*	*Clinical Staging*		*Surgical Staging*	
		Median Survival (Months)	*Five-Year Survival (%)*	*Median Survival (Months)*	*Five-Year Survival (%)*
I	T1 N0 M0 T2 N0 M0	48	48	> 60	63
II	T1 N1 M0 T2 N1 M0 T3 N0 M0	20	28	39	43
IIIA	T1–3 N2 M0 T3 N1 M0	12	12	22	30
IIIB	N3 (any T) M0 T4 (any N) M0	9	3	9	3
IV	M1 (any T or N)	5	2	5	2

Figure 1 **An international TNM four-stage system is used in the clinical and surgical evaluation of lung cancer. Definitions of TNM categories are simplified.[24,25]**

gical resection. Stage IV disease is treated with chemotherapy, palliative radiation, and supportive care [*see Table 6*].

Stages I and II Disease

Surgery Before a patient with stage I or II lung cancer undergoes surgery, the physician must undertake a determination of operability, which includes assessment of the medical risk of thoracotomy, as well as the risk of removal of the requisite pulmonary parenchyma. Cardiopulmonary disease, which is usually a consequence of tobacco use, is the major cause of postoperative morbidity and mortality in patients with stage I or II disease and, consequently, is the most significant medical factor in determining operability.

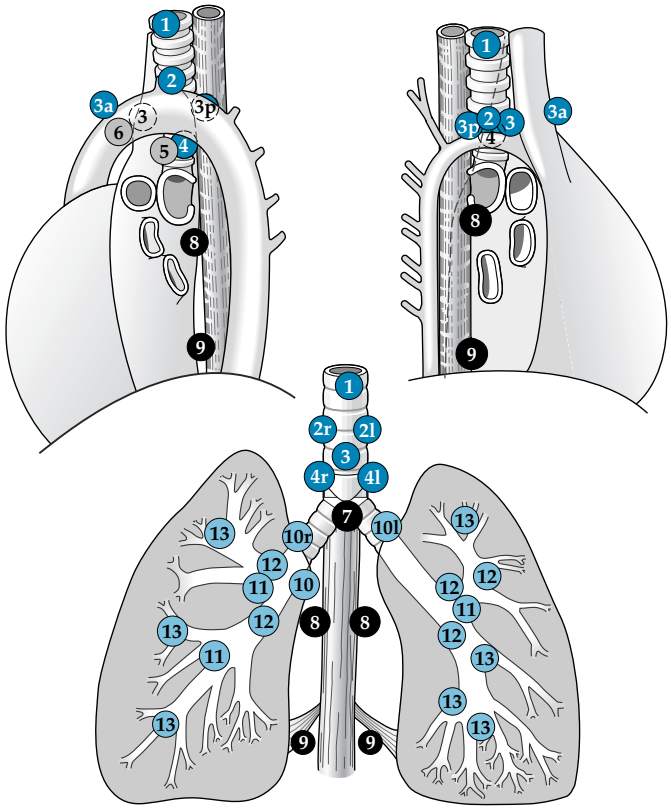
Pulmonary function testing and arterial blood gas analysis are used to determine the feasibility of pulmonary resection. Postoperative pulmonary function is estimated on the basis of the patient's preoperative function and the projected resection of pulmonary parenchyma. Resection is generally contraindicated when the predicted postoperative forced expiratory volume at 1 second (FEV_1) and forced vital capacity are less than 30% of predicted values. In patients with marginal results on preoperative pulmonary function studies, ventilation-perfusion scanning may be required to determine resectability. Postoperative FEV_1 may be predicted after assessing the contribution to overall pulmonary function made by each lung and by specific pulmonary segments.

In patients who have a history of angina or whose preoperative electrocardiogram shows ischemia or arrhythmia, radionuclide evaluation of myocardial perfusion or function is indicated. Normal results with these studies reliably exclude significant coronary artery disease; patients with positive results should undergo coronary arteriography. Recent myocardial infarction, uncontrolled heart failure, or uncontrollable arrhythmia precludes thoracotomy for pulmonary resection.

The final determination of resectability is made at thoracotomy. Contraindications to pulmonary resection at the time of thoracotomy include pleural metastases, extensive mediastinal lymph node involvement (N3 disease), or direct extension of the tumor to critical structures (T4 disease). In addition, pulmonary resection is aborted if the extent of resection required would leave the patient with inadequate pulmonary reserve, as determined by preoperative pulmonary function studies.

Four main oncologic principles guide resection for lung cancer: (1) removal of the entire tumor with an anatomically complete portion of lung (lobectomy or pneumonectomy), to ensure removal of all intraparenchymal lymphatic drainage; (2) en bloc resection of adjacent structures, if technically possible, including the chest wall, diaphragm, and pericardium, without transgressing the tumor; (3) assessment of questionable resection margins by frozen-section analysis to optimize the potential for complete resection; and (4) sampling or complete dissection of all accessible mediastinal lymph nodes to improve staging.

In patients with small (< 3 cm) peripheral nodules and no mediastinal lymphadenopathy by CT criteria (i.e., no lymph nodes > 1 cm in diameter), the procedure of choice is lobectomy and mediastinal lymph node dissection. Less extensive resection, such as wedge resection or segmentectomy, has been shown to be associated with significantly greater risk of local recurrence and cancer-specific death.[27] In patients with T2 or T3 tumors or with mediastinal adenopathy on chest CT, cervical mediastinoscopy should be performed before exploration for pulmonary resection.



N2 Nodes

Superior Mediastinal Nodes
1. Highest mediastinal
2. Upper paratracheal (2r, 2l)
3. Pretracheal and retrotracheal (3a, 3p)
4. Lower paratracheal (including azygos nodes) (4r, 4l)

Anterior Mediastinal Aortic Nodes
5. Subaortic (aortopulmonary window)
6. Para-aortic (ascending aorta or phrenic)

Inferior Mediastinal Nodes
7. Subcarinal
8. Paraesophageal (below carina)
9. Pulmonary ligament

N1 Nodes

10. Hilar, peribronchial (10r, 10l)
11. Interlobar
12. Lobar
13. Segmental

Figure 2 **Draining lymph node sites (nodal stations) in the chest that can be involved by lung cancer are noted. Clinical staging of cancer of the mediastinum is carried out by CT scanning; surgical staging is performed by mediastinoscopy, mediastinotomy, thoracoscopy, or, sometimes, thoracotomy. A lymph node larger than 1 cm on CT scan is considered abnormal, but cancer involvement must be proved by biopsy. The upper paratracheal nodal stations are designated as 2r (*right*) and 2l (*left*); the lower paratracheal nodal stations are designated as 4r and 4l. Stations 8r, 8l, 9r, and 9l are contiguous with the mediastinum, but positive nodes in these sites are not common. Station 10 nodes, when confirmed as positive by mediastinoscopy, are classified as tracheobronchial angle nodes (10r and 10l) and signify mediastinal involvement. In the drawing, r is right; l, left; a, anterior; and p, posterior.**

Table 6 Therapy and Prognosis for Non–Small Cell Lung Cancer

Stage (%)	*Surgery*	*Radiotherapy*	*Chemotherapy*	*Five-Year Survival (%)**
I (10)	T1N0 (coin lesion): lobectomy (poor pulmonary function, segmental resection) T2N0: lobectomy	None	Research	T1N0: 45–80 T2N0: 35–65
II (10)	T1N1 or T2N1: lobectomy; pneumonectomy usually required when hilar nodes are grossly involved	May reduce local recurrence but does not affect survival	Research	T1N1: 20–52 T2N1: 20–40
IIIA (20)	T3 (potentially resectable): radical resection of chest wall lesions; used after RT of Pancoast tumors	Used after surgery; may reduce local recurrence; used preoperatively for Pancoast tumors	Combined with RT and surgery†	T3 (chest wall): 30–55 T3 (Pancoast tumors): 20–40
	N2 (potentially resectable): radical resection of early intranodal disease; not indicated for extranodal or fixed, matted nodes	Used after surgery; may reduce local recurrence; used preoperatively in some patients with early intranodal disease	Combined with RT and surgery†	N2: 10–50
IIIB (20)	T4, N3, or both (unresectable)	Standard treatment for palliation of pain, hemoptysis, atelectasis, hoarseness, SVC syndrome	Combined with RT*	Median, 30 wk
IV (40)	Used rarely for isolated metastases	Useful for palliation of pain or other local problems	Response rates of 30%–40%; prolongation of survival	Median, 13–18 wk

*Survival is higher in patients staged by surgery than in patients staged clinically.
†Randomized trials show prolonged survival when chemotherapy is added to radiotherapy, surgery, or both.
RT—radiotherapy SVC—superior vena cava

Chemotherapy The benefit of chemotherapy for patients with stage I or II NSCLC is controversial. In clinical practice, treatment recommendations for such cases do not routinely include chemotherapy. Nevertheless, in reviewing treatment options with these patients, it is important to discuss the evolving data from clinical trials of chemotherapy.

A meta-analysis found that adjuvant treatment with alkylating agents in this setting resulted in a 5% decrease in survival, compared with surgery alone ($P = 0.005$).[28] Cisplatin-based regimens were associated with a 5% improvement in 5-year survival, but this effect did not reach statistical significance ($P = 0.08$). On the other hand, the clinical trials included in this meta-analysis were performed between 1965 and 1991, and both chemotherapy and supportive care have improved significantly since that time. Therefore, randomized trials of adjuvant chemotherapy versus supportive care alone in stage I and II NSCLC are currently under way. An intergroup trial sponsored by the Cancer and Leukemia Group B (CALGB) is comparing carboplatin and paclitaxel with supportive care in patients with stage IB disease. An intergroup trial sponsored by the National Cancer Institute of Canada (NCI-C) has compared cisplatin and vinorelbine with supportive care alone in patients with stage IB, IIA, and IIB disease; results of this trial are pending. The International Adjuvant Lung Cancer Trial (IALT), in which patients with stages I through IIIA resected NSCLC were randomized to cisplatin-based chemotherapy versus observation, found a 4% absolute improvement in survival with chemotherapy.[29] It is hoped that ongoing studies will confirm this benefit. In the meantime, the option of adjuvant chemotherapy should be discussed with patients after definitive surgical resection.

Another treatment strategy for stage I and II disease is the use of induction (preoperative) chemotherapy. Because of encouraging results from a phase II trial in patients with completely resected stage IB, IIA, or IIB NSCLC,[30] the Southwest Oncology Group is leading a prospective, randomized trial of induction chemotherapy with three cycles of paclitaxel and carboplatin followed by surgery, compared with surgery alone. Other randomized clinical trials of both induction and adjuvant chemotherapy are currently being conducted for patients with stage IB, IIA, or IIB NSCLC. Eligible patients should be encouraged to enroll in these clinical trials, so that oncologists can determine whether chemotherapy is beneficial in this setting.

Radiation therapy Surgery is the treatment of choice for stage I NSCLC, but patients with medical contraindications to surgery can be treated with radiation therapy alone. Retrospective studies of such cases have shown 5-year survival rates ranging from 10% to 30%.[31] Better local control was found in patients with smaller tumors (< 3 cm) and in those treated with higher doses of radiation (> 65 Gy). Consequently, recommended radiation doses range from 65 to 70 Gy; the total dose is typically given in 2-Gy fractions. Omission of regional nodal areas from the treatment fields has been found to reduce morbidity and has resulted in a nodal failure rate of only 4% to 9%. Therefore, in most cases, the primary tumor is treated with a standard margin of 1.5 to 2 cm. It is important to take into account any movement of the tumor from respiration, and this is best done under fluoroscopy. Unfortunately, most patients treated with radiation therapy succumb to recurrent lung cancer, and at least 60% experience local failure.

Stage III–Operable Patients

Patients with stage IIIA disease who appear to be candidates for surgical resection but in whom mediastinoscopy shows ipsilateral mediastinal lymph node involvement are evaluated for induction therapy (chemotherapy alone or chemotherapy and radiation therapy). Induction therapy with systemic chemotherapy has the potential to treat occult metastatic disease, which is common in patients with stage IIIA disease, even when organ-specific scans are negative. Three randomized, prospective trials that compared induction chemotherapy before surgery with surgery alone in patients with operable stage IIIA NSCLC were small in sample size, but all demonstrated

benefit from induction chemotherapy, with at least a doubling in 3-year survival.[32-34]

The addition of radiation therapy to induction therapy may improve local control in conjunction with surgical resection and may also decrease distant metastatic spread during therapy.[27] A prospective study has suggested that chemoradiation before surgery is beneficial in patients with stage IIIA NSCLC.[33] A randomized intergroup comparison of chemoradiation alone with chemoradiation followed by surgery has been conducted[35]; preliminary results suggest longer disease-free survival in the surgical arm but higher initial mortality, which complicates the analysis. Important questions remain about induction therapy, including the following: What are the optimum agents for chemotherapy? Should chemotherapy be used alone or in combination with radiation treatment? Does radiation therapy or surgery provide better local control? Should all three modalities of therapy be utilized? To answer these questions, enrollment of patients with stage IIIA NSCLC in clinical trials is critical.

After induction therapy, staging studies are repeated. Repeat mediastinoscopy is useful for reassessing the mediastinal lymph nodes, although this is more difficult than the primary procedure. Alternatively, the ipsilateral mediastinal lymph nodes may be assessed at exploratory thoracotomy. Pulmonary resection is not recommended if the involved lymph nodes have not responded to induction therapy or if there is evidence of disease progression, because the prognosis for extended survival is dismal in such patients.

Patients with involvement of the chest wall, diaphragm, or pericardium may be surgical candidates but only if the tumor can be completely resected. Incomplete resection of NSCLC provides no curative or palliative benefit.

Postoperative radiation therapy The treatment of patients found to be in stage II or III after resection is somewhat controversial. These patients are at a high risk for local and regional recurrences after surgery alone; however, they also have a very high likelihood of distant disease.[36] In a study of patients with stage II or III disease who had undergone a complete resection and were randomized to receive radiation therapy or no further treatment, the patients who received radiation therapy were found to have a significantly lower rate of local failure (3% versus 21% for patients who did not receive postoperative radiation).[37] However, there was no evidence of a survival benefit for the patients receiving postoperative radiation. Two caveats regarding this study are that it included only patients with squamous cell carcinoma and that most of the patients in the study had N1 nodal disease, precluding a valid subgroup analysis of the relationship between nodal status and survival.

A meta-analysis of nine published and unpublished randomized trials of postoperative radiation therapy—which included 2,128 patients with stage I, II or III lung cancer treated from 1966 to 1994, largely with cobalt radiation techniques—found that overall, mortality was approximately 7% higher for patients who received postoperative radiation therapy.[38] On subgroup analysis, the adverse effect was most apparent in patients with N0 and N1 disease; survival of patients with N2 disease was the same in the two groups. The results of this study indicate that radiation therapy is detrimental to patients with early stage (I and II) lung cancer that has been completely resected; the question of whether postoperative radiation therapy benefits patients with N2 disease remains unanswered. This meta-analysis has been critiqued for its inclusion of patients treated with a wide variety of radiation doses and techniques, many of them now outdated, which may have skewed the data from showing a survival benefit with postoperative radiation therapy.

In summary, the use of postoperative radiation therapy in patients with stage II or III NSCLC yields a significant increase in local control, which may be particularly important in patients with positive surgical margins. However, because of the high frequency of metastatic disease in these patients, postoperative radiation therapy appears to provide no survival benefit. Patients offered postoperative radiation therapy should clearly understand that its goal is improved local control. Meanwhile, the possible role of combination radiation therapy and chemotherapy as an adjuvant to surgery is the subject of ongoing clinical trials.

Adjuvant chemotherapy As in stage I and II NSCLC, the role of adjuvant chemotherapy in resected stage III disease is not well supported by the results of randomized clinical trials, although the IALT results (see above) may provide such support.[28,29] On the other hand, a trial comparing adjuvant cisplatin and etoposide plus radiation with radiation alone in resected stage II and IIIA disease found no survival advantage for the group receiving adjuvant chemotherapy.[39]

Stage IIIB disease A small subgroup of patients with stage IIIB NSCLC may be candidates for surgical resection. In general, T4 tumors are considered unresectable; however, there are two exceptions to this generalization. First, patients with a single satellite nodule within the same pulmonary lobe as the primary tumor are offered resection if the disease is apparently resectable by lobectomy and the results of both mediastinoscopy and organ-specific staging studies are negative. Second, in rare cases of very limited involvement of the vena cava, main pulmonary artery, or aorta by the primary tumor, en bloc resection and vascular reconstruction may be offered to selected patients; long-term survival in such cases ranges from 10% to 20%.

Some patients with contralateral mediastinal lymph node involvement (N3) are treated with induction therapy followed by surgical resection. However, the standard of care for these cases is chemotherapy and radiation therapy.

Stage III–Inoperable Patients

Radiation therapy Without treatment, most patients with stage IIIB NSCLC will succumb to their disease within 1 year. Radiation therapy does result in an improved outcome, with up to 20% of patients surviving 2 years and up to 5% surviving 5 years, but there is a high likelihood of local recurrence, ranging from 25% to 50% in some studies.[40] Distant recurrence is also common. An analysis of several studies reveals that patients with weight loss greater than 5%, performance status of less than 80%, and higher T and N stage have the worst prognosis.[41]

In an attempt to increase the efficacy of radiation therapy, fractionation schemes have been tested, including treatments given two or three times daily. A randomized trial, performed in Europe, is comparing the effectiveness of continuous hyperfractionated accelerated radiation therapy (HART), given three times a day, with conventional radiation treatment. Results of this protocol reveals a statistically significant survival benefit of 9% at 3 years for patients treated in the continuous HART arm.[42] There is also a significant increase in local control, with 7% fewer failures at 3 years in the CHART arm. These benefits

are most prominent in patients with squamous cell carcinoma. The results of several trials indicate that the addition of chemotherapy to radiation therapy leads to an improved survival. A multigroup, randomized study found that patients with unresectable cancer who received chemotherapy and radiation therapy in combination had a statistically improved overall survival compared with those who received only radiation either once or twice daily. For chemoradiation, standard radiation, and hyperfractionation, 3-year survival rates were 17%, 11%, and 9%, respectively, and median survival rates were 13.2 months, 11.4 months, and 12 months.[43] In a comparison of HART with once-daily radiation therapy after induction chemotherapy, the 2-year survival was 40% for the HART arm, compared with 33% for the standard radiation therapy group; toxicities, particularly esophagitis, were also increased.[44]

Clinical trials are currently evaluating the use of three-dimensional treatment planning systems to increase the dose of radiation therapy delivered to the primary tumor. Preliminary results show that dose escalation is feasible and does not lead to increased toxicity and that outcomes are comparable to or better than historical controls.[45]

Chemoradiotherapy The standard treatment for inoperable stage III NSCLC is a combination of chemotherapy and radiation therapy. The chemotherapy should be a platinum-based combination regimen; the radiation should be given at conventional doses, generally 66 Gy. The use of chemoradiotherapy in these cases is supported by level I-A evidence.[46] For unresectable stage III disease, chemotherapy plus radiation therapy is appropriate for patients with a good performance status (an Eastern Cooperative Oncology Group [ECOG] score of 0 to 1 or, possibly, 2).

The optimal strategy for coordinating chemotherapy with radiation therapy is evolving. Possibilities include chemotherapy before, during, or after radiation treatment. Promising results have been reported from a recent randomized phase II trial of induction chemotherapy followed by concurrent chemotherapy and radiation treatments.[47] Median survival in this trial was approximately 18 months, which compares favorably with the 13 to 14 months reported in previous trials. This trial utilized cisplatin-based chemotherapy along with vinorelbine, paclitaxel, or gemcitabine, which are all agents with documented benefit in stage IV NSCLC (see below). A randomized phase III trial is needed to determine whether the apparent improvement in median survival stems from the sequencing of chemotherapy and radiation treatment, the use of the new agents, or both.

Concurrent administration of chemotherapy and radiation therapy has also been found to be beneficial. In a randomized trial of chemotherapy with mitomycin, vinblastine, and cisplatin given either before or along with thoracic radiation, the 2-year survival rates with sequential and concurrent treatment were 27% and 35%, respectively.[48] Subsequent trials point to a survival benefit from concurrent chemotherapy and radiation treatment compared with sequential therapy, as well as an increase in toxicity, particularly esophagitis and pneumonitis.

Treatment Strategies in Stage III Disease

The optimal approach to management of stage III NSCLC remains undefined. At present, it is clear that chemotherapy, when used in combination with surgery or radiation treatment, can improve patient survival in both operable and inoperable disease. In patients with inoperable stage III NSCLC, long-term survival is better with platinum-based combination chemotherapy and radiation therapy than with either modality alone. Every attempt should be made to enroll patients in clinical trials to further clarify the optimal strategy.

Because the benefits of combination therapy have been largely demonstrated in only younger patients with higher performance status, physicians should use caution in applying these approaches to elderly patients or those with poor performance status. In the absence of data from elderly and poor-performance patient populations, low-dose chemotherapy and concurrent radiation can be considered. An alternative strategy that may result in less toxicity from esophagitis would be the use of combination chemotherapy followed by radiation treatment. This strategy allows individualization of treatment based on the patient's tolerance of induction chemotherapy. However, older patients with good performance status should not be denied the potential benefit of combined-modality therapy. In this setting, it would appear that the most appropriate choice for chemotherapy is a cisplatin-based or carboplatin-based regimen with one of the newer agents used in the management of stage IV disease.

Stage IV Disease

For patients with stage IV NSCLC, chemotherapy plus supportive care improves both survival and quality of life, compared with supportive care alone. Because 5-year survival in stage IV disease is 1% or less, discussions of outcomes in the literature often describe median survival, which can be measured in months. For a physician who is speaking with an individual patient, however, it is more meaningful to discuss the probability of living 1 or 2 years. A statement based on median survival in a large population of patients, such as "You have 6 months to live," does not help that patient understand the range of survival that occurs even in stage IV NSCLC. It gives the patient a better idea of the probabilities if the physician instead specifies the percentage of patients with advanced NSCLC who are alive at 1 year after diagnosis. According to a National Cancer Center database, untreated patients with stage IV disease had a 1-year survival of 9% to 11%, whereas patients receiving chemotherapy had a 1-year survival of 20% to 25%.[49] These data are robust, because the population includes more than 700,000 patients with lung cancer, diagnosed between 1985 and 1995, and includes all stages and treatment categories for lung cancer. However, because this represents a database rather than a randomized comparison of groups, the survival data do not reflect the clinical factors that would guide the decision to forgo treatment in some patients. Such factors might include low performance status, comorbid disease, and advanced age, all of which may adversely affect survival.[50]

Which chemotherapeutic agents are best for stage IV disease? A meta-analysis of trials of supportive care alone versus supportive care with chemotherapy for advanced-stage disease demonstrated a 6% decrease in 1-year survival with the use of alkylating agents alone and a 4% improvement in 1-year survival with the use of vinca alkaloids or etoposide. Neither of these results reached statistical significance, however. By contrast, randomized trials of cisplatin-based combination chemotherapy versus supportive care showed an absolute increase of 10% in 1-year survival ($P < 0.0001$).[28] These studies generally restricted eligibility to patients with higher performance status and enrolled a disproportionate number of younger patients

with less comorbid disease than commonly seen in the community. Nevertheless, it is of interest that the magnitude of benefit in this trial is similar to that documented in the National Cancer Center database.

On the basis of this meta-analysis and additional data, an expert panel for the American Society of Clinical Oncology has concluded that in stage IV disease, platinum-based combination chemotherapy prolongs survival and is most appropriate for patients with good performance status, including an ECOG score of 0 or 1 or, possibly, 2.[46] Although randomized trials of platinum-based chemotherapy have almost all involved cisplatin, carboplatin has a more favorable safety profile and lower toxicity. Furthermore, randomized trials comparing cisplatin and carboplatin with etoposide in NSCLC have shown comparable efficacy.[51] Although a European study has suggested a small survival benefit for cisplatin therapy, compared with carboplatin-based therapy,[52] in the United States carboplatin continues to be the most widely used agent in the palliative management of patients with advanced lung cancer.

The 1990s brought the advent of newer agents in the treatment of NSCLC, including vinorelbine, paclitaxel, and gemcitabine, all three of which have received Food and Drug Administration approval for use (in combination with cisplatin) in the treatment of advanced disease. The approval of these newer agents was based on the results of randomized clinical trials that compared them, in combination with cisplatin, with either cisplatin alone or cisplatin in combination with older agents. In these trials, vinorelbine and cisplatin were associated with 1-year survival of 40%[53] and 36%.[54] These differences were statistically superior to vinorelbine alone, cisplatin and vindesine, or cisplatin alone; 1-year survival with cisplatin as a single agent was only 20%. Similarly, the combination of cisplatin and gemcitabine provided a 1-year survival of 39%, compared with 26% for cisplatin alone.[55] In the study that led to the approval of paclitaxel for NSCLC, 1-year survival was 32% in the control groups that received cisplatin and etoposide, compared with 37% to 40% for patients receiving paclitaxel. Survival with paclitaxel depended on the dose used; patients receiving a higher dose required supportive therapy with granulocyte colony-stimulating factor.[56] A randomized trial has demonstrated a better survival rate with cisplatin and docetaxel than with cisplatin and vinorelbine.[57] In this study, survival for a carboplatin and docetaxel group was similar to that for a cisplatin and vinorelbine group, but the former had a more favorable side-effect profile.

Overall, comparative trials have suggested that several different combination regimens may be equally effective in advanced disease, although toxicities vary.[58-60] The optimal regimen should comprise two chemotherapy drugs, including a platinum agent and one of the newer chemotherapeutic agents. Of the newer agents, paclitaxel, vinorelbine, gemcitabine, or docetaxel would all be reasonable choices.

Palliative radiation therapy Radiation therapy is used for palliation of symptoms caused by metastatic NSCLC. These include obstructive symptoms, bone pain, and neurologic compromise from spinal cord compression or brain metastasis. In randomized trials, palliative radiation has been shown to produce some pain relief in 75% to 90% of patients and complete pain relief in at least 50%.[61] Several fractionation schemes seem equally effective, but there is some evidence that prolonged treatment provides longer-lasting pain relief.

Superior vena cava syndrome responds to radiation treatment in approximately 50% of cases, but a substantial number of patients do not respond. Approximately 75% of patients have resolution of hemoptysis, and 50% will have cessation of cough after palliative radiation therapy. About 50% to 75% of patients with brain metastasis have a symptomatic response.

Spinal cord compression can also be treated with radiation therapy. Of patients who have only pain, 75% remain ambulatory, but only 30% to 35% of patients with muscle weakness improve.

Quality-of-life considerations Studies of chemotherapy for advanced NSCLC have largely focused on length of survival. In view of the modest benefits of treatment, however, the impact of chemotherapy on quality of life is a critical consideration. Although earlier trials often did not assess quality of life, virtually all of the current randomized clinical trials have quality-of-life measures as a significant component. These studies suggest that combination chemotherapy often results in the improvement of symptoms such as cough, dyspnea, chest pain, and hemoptysis, often even when there is minimal evidence of tumor response.[62] In general, these studies have shown that chemotherapy produces symptomatic improvement in more than 50% of patients. This figure is significantly higher than the objective response rate, which generally varies between 20% and 40% for combination regimens. The seeming discrepancy between these figures likely reflects the fact that even minor responses, or simply stabilizing the growth of the cancer, may bring at least short-term improvement in symptoms. With the improved toxicity profiles of the newer agents, improvements in quality-of-life differences may be easier to demonstrate.[63]

The elderly and other special populations Now that the benefits of chemotherapy have been established in younger patients with good performance status, researchers are evaluating its benefits in older patients and those with lower performance status. The first randomized trial of chemotherapy versus supportive care in the elderly (≥ 70 years), the Elderly Lung Cancer Vinorelbine Italian Study group (ELVIS) trial, assessed the effects of vinorelbine, as compared with supportive care alone, on both quality of life and survival of patients with advanced NSCLC.[64] In this study, 1-year survival was 32% for the vinorelbine-treated group, compared with 14% for best supportive care alone. On the basis of this trial, the FDA has approved vinorelbine for single-agent use in NSCLC. Because of the superior outcome with the combination of cisplatin and vinorelbine, single-agent treatment has not been widely used in high-performance-status populations. However, the ELVIS trial suggests that single-agent chemotherapy is beneficial in elderly patients, many of whom may not be eligible for combination chemotherapy. However, for the fit elderly, a subset analysis of a randomized trial of carboplatin and paclitaxel compared with paclitaxel alone demonstrated a survival advantage for the combination regimen in patients older than 70 years; this result was comparable to that in younger patients.[60]

Relapse of disease For patients in whom initial chemotherapy has failed, two randomized trials have now demonstrated the benefit of salvage chemotherapy. A trial comparing docetaxel with either ifosfamide or vinorelbine showed a survival advantage for patients receiving docetaxel, particularly at the dose of 75 mg/m^2.[65] A trial comparing docetaxel with best

Table 7 Therapy and Prognosis for Small Cell Lung Cancer

Stage	*Surgery*	*Radiotherapy*	*Chemotherapy*	*Median Survival (months)*	*Two-Year Disease-Free Survival (%)*
Limited I	T1N0, T2N0: Lobectomy	None	T1N0, T2N0: Research	> 48 > 14	T1N0 > 60 T2N0 > 30
Limited II	T1N1, T2N1: Lobectomy; pneumonectomy rarely used	Rarely used	Several regimens available	> 14	T1N1, T2N1: 15–30
Limited IIIA IIIB	T3, N2, or both: research T4 or N3: not indicated	Decreases local recurrences; used prophylactically for CNS	Several regimens available; high-dose programs with autologous marrow rescue are being researched Complete response: 50%–80%	> 12	15–20
Extensive IV	Not indicated	Useful for palliation of pain, atelectasis, SVC syndrome	Several regimens available Complete response: 30%–50%	> 8	≤ 5

SVC—superior vena cava

supportive care showed that docetaxel produced improvement in both survival and quality of life.[66] On the basis of these trials, the FDA has approved docetaxel for salvage chemotherapy in this setting.

Gefinitib, a tyrosine kinase inhibitor, has received FDA approval as monotherapy for patients with locally advanced or metastatic NSCLC that has failed to respond to both platinum-based therapy and docetaxel. Approval was granted under accelerated-approval regulations; although no controlled trials have shown an improvement in symptoms or survival with the agent, gefinitib has produced objective responses and symptomatic improvement in this heavily pretreated population. Toxicities include a skin rash, which is common, and interstitial lung disease, which is a rare but potentially fatal complication of gefinitib.[67] There is reason to hope that this agent, which targets epidermal growth factor receptors, will usher in an era of more precise therapy for lung cancer.

SMALL CELL CARCINOMA OF THE LUNG

Whereas the management of NSCLC changed substantially during the past decade, the management of SCLC has evolved more slowly. Important refinements in therapy have occurred, however. Currently, new directions in therapy are being developed, on the basis of better understanding of the biology of this disease.

As in NSCLC, the outcome in patients with SCLC is determined largely by stage at presentation [*see Table 7*]. The difference between the two diseases lies in the stage distribution. In NSCLC, 25% to 30% of patients present with stage I or II disease; in SCLC, that figure is well below 5%. Several factors may contribute to this difference, but a principal one appears to be that small cell carcinoma arises from neuroendocrine cells, which normally reside below the bronchial epithelium in the mucosa and submucosa. This region is much more heavily supplied with lymphatics, which would facilitate the earlier spread of these cancer cells to regional and distant lymph nodes, as well as hematogenous spread. In addition, in the 1960s it was learned that SCLC was more sensitive to chemotherapy than NSCLC. Because of the advanced stage at presentation and this chemosensitivity, SCLC fell largely into the domain of medical oncologists and radiation therapists rather than that of thoracic surgeons.

Early-Stage Disease

Early-stage solitary pulmonary nodules have been referred to as peripheral small cell carcinoma. In patients without mediastinal nodal involvement, this disease is associated with a 5-year survival in the 30% to 60% range, depending on the series. Long-term follow-up of these patients suggests a high tendency for distant relapse, both systemically and in the CNS. Surgery should be the therapy of choice for patients with no evidence of distant disease or of mediastinal nodal involvement. Because the number of patients presenting at this stage is small, no definitive randomized trials of adjuvant therapy have been performed and there is no consensus on adjuvant therapy. In the United States, it has been standard to recommend adjuvant chemotherapy, either alone or in combination with prophylactic cranial radiation or chest radiation.

Surgery with mediastinal lymph node dissection, followed by four cycles of adjuvant platinum and etoposide, was the subject of a phase II trial in patients with stage I to IIIA SCLC conducted by the Japanese Collaborative Oncology Group.[68] In patients with stage I disease, who constituted the bulk of the study population, 5-year survival was 69%. In the limited number of patients with stage II or IIIA disease, survival was 38% to 40%. No prophylactic cranial irradiation was done, and the rate of CNS relapse was 17%, even in the stage I population, which suggests that prophylactic cranial radiation should have a role in this population.

Limited-Stage or Extensive-Stage Disease

At least 95% of patients with SCLC present with either limited-stage or extensive-stage disease. Limited-stage SCLC corresponds to stage IIIA or IIIB NSCLC without malignant pleural effusion but, generally, with supraclavicular lymph node involvement. Simplistically, limited-stage SCLC patients include those in whom the primary tumor and involved nodes, including supraclavicular nodes, can be encompassed in a standard radiation port. By contrast, patients with extensive disease have a malignant pleural effusion, contralateral lung involvement, or metastases at more distant sites. This staging separation has been important for selecting treatment and determining prognosis. Patients with limited-stage disease have a 5-year survival of 15% to 20% when treated with the combination of

chemotherapy and radiation. In contrast, patients with extensive-stage disease generally have a 5-year survival of 1% to 2% and have been treated primarily with chemotherapy, plus adjunctive radiation as needed. The distinction between limited- and extensive-stage disease has largely been based on clinical examination, as well as on the standard imaging studies of chest CT through the adrenal glands, bone scan, and brain CT or MRI. Of patients with no evidence of distant metastases on these studies, an additional 5% will be found to have distant disease by bone marrow evaluation. Currently, approximately 25% to 30% of patients with SCLC present with limited-stage disease, and the remainder present with extensive-stage disease; however, with the development of more sophisticated imaging techniques and the increased use of PET imaging, some degree of stage migration is likely to occur.

Limited-stage disease Chemotherapy has been the mainstay of treatment for limited-stage disease for the past 30 years. For at least the past 10 years, the standard treatment has been based on etoposide, either as a single agent or combined with either cisplatin or carboplatin, generally for four to six cycles. Most of the advances in the understanding of treatment of limited-stage SCLC have come from trials evaluating the role of radiation therapy.[69] These trials have shown that cure in limited-stage SCLC is associated with the use of full-dose combination chemotherapy, along with thoracic and cranial radiation therapy. Thoracic radiation appears to be more beneficial if it is done concurrently with chemotherapy, rather than sequentially. Secondly, starting radiation therapy during the first few cycles of chemotherapy appears to provide a survival advantage over beginning radiation therapy at the end of chemotherapy.[70] The explanation for this is that drug resistance may evolve rapidly in small cell carcinoma, so in some patients, delaying thoracic radiation even for a few months may result in escape of drug-resistant cells from the lung and mediastinum to distant sites. Early intervention with radiation may result in the destruction of these cells and an increase in the rate of cure. However, as with NSCLC, the benefit of concurrent chemotherapy and radiation must be balanced against its higher toxicity (e.g., esophagitis).

Although clinical practice varies, most oncologists initiate radiation therapy with chemotherapy cycle one or with cycles two or three. The choice of timing depends on the clinical situation. The advantage of delaying radiation for one or two cycles is that during this time the tumor mass may be significantly reduced, therefore easing the patient's symptoms and often improving performance status, so that the patient is better able to cope with combination treatment. Most initial trials of concurrent therapy have used pretreatment tumor volumes to determine radiation fields. However, newer strategies of radiation therapy will take advantage of chemotherapy-induced tumor reduction and use higher doses of radiation delivered to a smaller tumor volume. In one major trial, twice-daily radiation, compared with once-daily treatment, has been associated with improved surgical outcome,[71] but twice-daily treatment is also associated with significantly higher toxicity rates. The use of twice-daily radiation can be considered in some patients with high performance status. The standard dose of daily radiation therapy that is given concurrently with chemotherapy has increased from 45 Gy to 60 Gy.

Another significant advance in the understanding of SCLC involves the role of prophylactic cranial irradiation. Early trials that used prophylactic cranial irradiation along with chemotherapy reported a high incidence of both immediate and delayed neurologic toxicity. Studies have used prophylactic cranial irradiation at lower dose fractions and have delayed its initiation until chemotherapy has been completed. These adjustments seem to be associated with substantially diminished immediate and delayed neurotoxicity. Moreover, a meta-analysis of the use of prophylactic cranial irradiation from randomized clinical trials (largely composed of patients with limited-stage disease) has demonstrated not only a reduction in CNS relapse but also an increase in long-term survival by approximately 5%.[72]

Several issues involving prophylactic cranial irradiation remain to be settled. One is its role in elderly patients, who appear to be at higher risk for neurotoxicity. Also, although trials of SCLC have often included patients in complete remission from either limited or extensive disease, most of the patients have had limited disease. Therefore, the strength of the recommendation for prophylactic cranial irradiation in patients in complete remission from extensive-stage SCLC is less certain.

A number of trials have evaluated maintenance chemotherapy or the addition of other treatments after chemoradiotherapy for SCLC. Phase II studies of high-dose chemotherapy for patients in remission have suggested an encouraging long-term survival rate,[73] and a potential long-term benefit has also been reported in a small number of patients who received a vaccine against a tumor ganglioside (BEC-2)[74]; the latter approach is currently the subject of a randomized phase III trial. No benefit for these approaches has yet been proved, however, so maintenance therapy after completion of standard therapy cannot be recommended for SCLC patients, except in the context of a clinical trial.

Extensive-stage disease Progress in the treatment of extensive-stage SCLC over the past 10 years has been slower than that of limited-stage SCLC.[69] As with limited-stage disease, treatment centers on chemotherapy. Although several different chemotherapy combinations have been demonstrated to have equivalent survival outcomes, etoposide-based regimens have been shown to be superior to older regimens such as cyclophosphamide, doxorubicin, and vincristine. Secondly, most clinical trials in SCLC have used cisplatin, but as noted above, carboplatin has a better therapeutic ratio, is equally effective, and causes significantly less nausea, vomiting, and neurotoxicity.[75,76] Currently, however, cisplatin and carboplatin are used with roughly equal frequency for SCLC in community practice in the United States. Both agents are generally administered every 3 to 4 weeks, depending on hematologic recovery. Although the combination of platinum and etoposide remains the most commonly used regimen in the United States, cisplatin and irinotecan has been approved for treatment of SCLC in Japan, and randomized trials of this combination are ongoing in the United States.

Special populations Adaptation of treatment to special populations is another very important principle in the management of SCLC. Standard combination chemotherapy regimens, which were developed for younger, healthier patients, pose a greater risk of severe myelosuppression and life-threatening complications in elderly patients with significant comorbid disease and in patients with poor performance status. Modified regimens should be considered in such patients. If combination therapy is employed, the use of carboplatin (dosed according

to renal function) rather than cisplatin may produce less toxicity in this population. Although monotherapy with low-dose etoposide is certainly a consideration, most elderly patients should be considered for combination carboplatin and etoposide therapy if permitted by their level of functional illness and degree of comorbid illness. Whether these patients should receive significantly reduced doses or near-standard doses plus hematopoietic growth factor has not been well studied.

Relapse of disease Although SCLC typically responds to therapy, the majority of patients with limited-stage disease and nearly all patients with extensive-stage disease will experience relapses. For limited-stage patients, the duration of remission may range from months to years. For most patients with extensive-stage disease, remission generally lasts only weeks to months after completion of chemotherapy. In both groups of patients, a major issue is the emergence of drug resistance, which has severely hampered the benefit of second-line or so-called salvage chemotherapy.

In this setting, a critical factor in choosing therapy is the time to relapse. For patients who have had a prolonged remission (longer than 1year), retreatment with the same regimen will often produce a response. On the other hand, in patients who fail to respond to initial therapy (primary drug resistance), no therapy has been convincingly shown to be beneficial, so these patients are best enrolled in experimental trials.

For patients who have had a remission of at least 2 months after initial therapy, topotecan has been shown to produce a significant improvement in disease-related symptoms—particularly dyspnea, but also anorexia, hoarseness, and fatigue—and has been approved for second-line therapy for patients with SCLC.[77] The other commonly used approach in relapses of SCLC has been prolonged courses of oral etoposide for 10 to 21 days per cycle. Other agents that have shown promise for salvage therapy include paclitaxel, irinotecan, vinorelbine, and gemcitabine.[69,78] These agents, as well as more novel approaches, warrant further investigation.[79]

SUPPORTIVE CARE

Supportive care is important across all stages of both NSCLC and SCLC. In patients with advanced or refractory disease, it often becomes the principal form of therapy. As the ability of oncologists to deliver both chemotherapy and radiation safely has improved, it is clear that one of the best supportive care approaches is to palliate disease-related symptoms with chemotherapy, radiation, or both.

Over the past 10 years, management of pain, depression, nausea, weight loss, and constipation in cancer patients has been improved by the development of pharmacologic agents with better efficacy and less toxicity [*see CE:X Symptom Management in Palliative Medicine*]. In patients who are receiving myelosuppressive chemotherapy and so are at high risk for neutropenic complications, the use of colony-stimulating factors has been demonstrated to substantially reduce the risk of fever, neutropenia, and infection.[80] How to identify the populations who are at highest risk and are thus candidates for primary prophylaxis with colony-stimulating factors is an area that needs further investigation.

Anemia and fatigue are now appreciated as especially important concerns in patients with lung cancer.[81] Analysis of the 1,748 lung cancer patients enrolled in the prospective trials of recombinant human erythropoietin for cancer patients with anemia has shown that amelioration of anemia led to significant improvements in quality of life; these improvements applied to patients with SCLC or NSCLC receiving a variety of chemotherapy regimens.[82] Although all anemic cancer patients may benefit from higher hemoglobin levels, those with lung cancer may reap particular benefits because of dyspnea and fatigue related to the cancer itself, as well as comorbid diseases and the treatments used. In addition to substantially enhancing quality of life, normalization of hemoglobin levels may also improve therapeutic outcome, perhaps by improving tolerance to treatment or by reducing tumor hypoxia. Studies evaluating the role of hemoglobin maintenance in reducing cognitive dysfunction during treatment are ongoing.

TREATMENT UPDATES

The National Cancer Center Network (NCCN) provides evidence-based, expert-panel guidelines for the management of cancer, as well as supportive care guidelines. Clinicians are encouraged to review the NCCN Web site (http://www.nccn.org) for annual updates of lung cancer management guidelines.

Prognosis

At present, the 5-year survival rate for lung cancer is approximately 14%. The chance of survival is heavily dependent on stage at presentation, however. For example, 5-year survival in patients with stage I NSCLC is approximately 63%. Unfortunately, more than 40% of all patients with NSCLC have stage IV disease at diagnosis, and in this population the survival at 5 years is 1% or less. In patients with SCLC, 5-year survival ranges from 30% to 60% for peripheral disease (stage I/II) to 1% to 3% for patients with extensive disease (stage IIIB/IV). Unfortunately, approximately 65% to 70% of patients with small cell lung cancer present with extensive disease.

The author has received grants or research support or has been a consultant or a member of the speakers' bureaus of Abgenix, Inc.; Amgen, Inc.; AstraZeneca Pharmaceuticals LP; Bristol-Myers Squibb Co; GlaxoSmithKline; ImClone Systems Incorporated; Immunex Corporation; Eli Lilly and Company; Ortho Biotech Products, LP; QLT Inc.; and Schering-Plough Corporation during the past 12 months.

References

1. Jemal A, Murray T, Samuels A, et al: Cancer Statistics, 2003. CA Cancer J Clin 53:5, 2003

2. Wynder El: The past, present, and future of the prevention of lung cancer. Ca Epidemiol Biomarkers Prevent 7:735, 1998

3. Youth tobacco surveillance—United States, 2000. Centers for Disease Control. MMWR CDC Surveill Summ 50(4):1, 2001

4. Healton C, Messeri P, Reynolds J, et al: Tobacco use among middle and high school students—United States, 1999. MMWR Morb Mortal Wkly Rep 49(3):49, 2000

5. Halpern MT, Gillespie BW, Warner KE: Patterns of absolute risk of lung cancer mortality in former smokers. J Natl Cancer Inst 85:457, 1993

6. Shields PG: Molecular epidemiology of smoking and lung cancer. Oncogene 21:6870, 2002

7. Vallyathan V, Green F, Ductaman B, et al: Roles of epidemiology, pathology, molecular biology, and biomarkers in the investigation of occupational lung cancer. J Toxicol Environ Health B Crit Med 1:91, 1998

8. Neuberger JS, Gesell TF: Residential radon exposure and lung cancer: risk in nonsmokers. Health Phys 83:1, 2002

9. Fong KM, Sekido Y, Minna JD: Molecular pathogenesis of lung cancer. J Thoracic Cardiovasc Surg 118:1136, 1999

10. Carbone DP: The biology of lung cancer. Semin Oncol 24:388, 1997

11. Rozengurt E: Autocrine loops, signal transduction, and cell cycle abnormalities in the molecular biology of lung cancer. Curr Opin Oncol 11:116, 1999

12. Schreiber G, Pitterle D, Young-Chul K, et al: Molecular genetic analysis of primary lung cancer and cancer metastatic to the lung. Anticancer Res 19:1109, 1999

13. Toma S, Raffo P, Isnardi L, et al: Retinoids in lung cancer chemoprevention and treatment. Ann Oncol 10(suppl 5):S95, 1999

14. Omenn GS, Goodman GE, Thornquist MD, et al: Risk factors for lung cancer and for intervention effects in CARET, the Beta-Carotene and Rational Efficacy trial. J Natl Cancer Inst 88:1550, 1996

15. Caraballoso M, Sacristan M, Serra C, et al: Drugs for preventing lung cancer in healthy people. Cochrane Database Syst Rev (2):CD002141, 2003

16. Strauss GM: Randomized population trials and screening for lung cancer: breaking the cure barrier. Cancer 89(11 suppl):2399, 2000

17. Henschke CI, Naidich DP, Yankelevitz DF, et al: Early Lung Cancer Action Project: initial findings on repeating screenings. Cancer 92:153, 2001

18. Swensen SJ, Jett JR, Sloan JA, et al: Screening for lung cancer with low dose spiral computed tomography. Am J Respir Crit Care Med 165:508, 2002

19. Mahadevia PJ, Fleisher LA, Frick KD, et al: Lung cancer screening with helical computed tomography in older adult smokers: a decision and cost-effectiveness analysis. JAMA 289:313, 2003

20. Lam S, Shibuya H: Early diagnosis of lung cancer. Clin Chest Med 20:53, 1999

21. NLST: National Lung Screening Trial. National Cancer Institute, 2003. http://www.cancer.gov/nlst

22. Patz EF, Erasmus JJ: Positron emission tomography imaging in lung cancer. Clin Lung Cancer 1:42, 1999

23. van Tinteren H, Hoekstra OS, Smit EF, et al: Effectiveness of positron emission tomography in the preoperative assessment of patients with suspected non-small-cell lung cancer: the PLUS multicentre randomised trial. Lancet 359:1388, 2002

24. Mountain CF: A new international staging system for lung cancer. Chest 89(suppl 4):225S, 1986

25. Mountain CF: Revisions in the international system for staging lung cancer. Chest 111:1710, 1997

26. Mountain CF, Dresler CM: Regional lymph node classification for lung cancer staging. Chest 111:1718, 1997

27. Ginsberg RJ: Multimodality treatment of resectable non–small cell lung cancer. Clin Lung Cancer 1:194, 2000

28. Chemotherapy in non–small cell lung cancer: a meta-analysis using updated data on individual patients from 52 randomized clinical trials. Non–Small Cell Collaborative Group. BMJ 311:899, 1995

29. Results of the randomized International Adjuvant Lung Cancer Trial (IALT): cisplatin-based chemotherapy (CT) vs. no CT in 1867 patients with resected non–small cell lung cancer (NSCLC) (abstr 6). The IALT Investigators. Proc Am Soc Clin Oncol 22:2, 2003

30. Pisters KMW, Ginsberg RJ, Putnam JB, et al: Induction paclitaxel & carboplatin (PC) in early stage non–small cell lung cancer (NSCLC): early results of a completed phase II trial. Proc Am Soc Clin Oncol 18:1800, 1999

31. Sibley GS, Jamieson TA, Marks LB, et al: Radiotherapy alone for medically inoperable stage I non–small-cell lung cancer: the Duke experience. Int J Radiat Oncol Biol Phys 40:149, 1998

32. Pass HI, Pogrebniak, Steinberg HW, et al: Randomized trial of neoadjuvant therapy for lung cancer: interim analysis. Ann Thorac Surg 53:992, 1992

33. Roth JA, Fossella F, Komaki R, et al: A randomized trial comparing perioperative chemotherapy and surgery with surgery alone in resectable stage III non–small cell lung cancer. J Natl Cancer Inst 86:673, 1994

34. Rosell R, Gomez-Codina J, Camps C, et al: A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non–small cell lung cancer. N Engl J Med 330:153, 1994

35. Albain KS, Scott CB, Rusch VR, et al: Phase III comparison of concurrent chemotherapy plus radiotherapy (CT/RT) and CT/RT followed by surgical resection of stage IIIA (pN2) non–small cell lung cancer (NSCLC): initial results from intergroup trial 0139 (abstr 2497). Proc Am Soc Clin Oncol 22:621, 2003

36. Effects of postoperative mediastinal radiation on completely resected stage II and stage III epidermoid cancer of the lung. The Lung Cancer Study Group. N Engl J Med 315:1377, 1986

37. Martini N, Flehinger BJ, Zaman MB, et al: Results of resection in non–oat cell carcinoma of the lung with mediastinal lymph node metastases. Ann Surg 198:386, 1983

38. Postoperative radiotherapy in non–small-cell lung cancer: systematic review and meta-analysis of individual patient data from nine randomized controlled trials. PORT Meta-analysis Trialists Group. Lancet 352:257, 1998

39. Keller SM, Adak S, Wagner H, et al: A randomized trial of postoperative adjuvant therapy in patients with completely resected stage II or IIIa non–small cell lung cancer. N Engl J Med 343:1217, 2000

40. Perez CA, Pajak TF, Rubin P, et al: Long-term observations of the patterns of failure in patients with unresectable non–oat cell carcinoma of the lung treated with definitive radiotherapy: report by the Radiation Therapy Oncology Group. Cancer 59:1874, 1987

41. Komaki R, Scott CB, Byhardt R, et al: Failure patterns by prognostic group determined by recursive partitioning analysis (RPA) of 1547 patients on four radiation therapy oncology group (RTOG) studies in inoperable non–small cell lung cancer (NSCLC). Int J Radiat Oncol Biol Phys 42:263, 1998

42. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non–small cell lung cancer: mature data from the randomized multicentre trial. CHART Steering Committee. Radiother Oncol 52:137, 1999

43. Sause W, Kolesar P, Taylor S IV, et al: Final results of phase III trial in regionally advanced unresectable non-small cell lung cancer: radiation therapy oncology group, Eastern Cooperative Oncology Group, and Southwest Oncology Group. Chest 117:358, 2000

44. Belani CP, Wang W, Johnson DH, et al: Induction chemotherapy followed by standard thoracic radiotherapy vs. hyperfractionated accelerated radiotherapy (HART) for patients with unresectable stage IIIA and B non-small cell lung cancer: phase III study of the Eastern Cooperative Oncology Group (abstr 2500). Proc Am Soc Clin Oncol 22:622, 2003

45. Marks LB, Sibley G: The rationale and use of three-dimensional radiation treatment planning for lung cancer. Chest 116(6 suppl):539S, 1999

46. Clinical practice guidelines for the treatment of unresectable non-small cell lung cancer. American Society of Clinical Oncology. J Clin Oncol 15:2996, 1997

47. Vokes EE, Herndon JE II, Crawford J, et al: A randomized phase II study of gemcitabine or paclitaxel or vinorelbine with cisplatin and induction chemotherapy (Ind CT) followed by concomitant chemoradiotherapy (XRT) for unresectable stage III non–small cell lung cancer (NSCLC) (CALGB study 9431). J Clin Oncol 20:4191, 2002

48. Furuse K, Fukuoka M, Kawahara M, et al: Phase III study of concurrent versus sequential thoracic radiotherapy in combination with mitomycin, vindesine, and cisplatin in unresectable stage III non-small cell lung cancer. J Clin Oncol 17:2692, 1999

49. Fry WA, Phillips JL, Menck HR: Ten year survey of lung cancer treatment and survival in hospitals in the United States: a national cancer data base report. Cancer 86:1867, 1999

50. Johnson DH: Treatment strategies for metastatic non–small cell lung cancer. Clin Lung Cancer 1:34, 1999

51. Klastersky J, Sculier JP, Lacroix H, et al: A randomized study comparing cisplatin or carboplatin with etoposide in patients with advanced non-small cell lung cancer: European Organization for Research and Treatment of Cancer Protocol 007861. J Clin Oncol 8:1556, 1990

52. Rosell R, Gatzemeier U, Betticher DC, et al: Phase III randomised trial comparing paclitaxel/carboplatin with paclitaxel/cisplatin in patients with advanced non-small-cell lung cancer: a cooperative multinational trial. Ann Oncol 13:1539, 2002

53. Le Chevalier T, Brisgand D, Douillard JY, et al: Randomized study of vinorelbine and cisplatin versus vindesine and cisplatin versus vinorelbine alone in advanced non–small cell lung cancer: results of a European multicenter trial including 612 patients. J Clin Oncol 12:360, 1994

54. Wozniak AJ, Crowley JJ, Balcerzak SP, et al: Randomized trial comparing cisplatin with cisplatin plus vinorelbine in the treatment of advanced non-small cell lung cancer: a Southwest Oncology Group study. J Clin Oncol 16:2495, 1998

55. Sandler A, Nemunaitis J, Dehnam C, et al: Phase III study of cisplatin (C) with or without gemcitabine (G) in patients with advanced non–small cell lung cancer (NSCLC) (abstr). Proc Am Soc Clin Onc 17:454a 1999

56. Bonomi PD, Kim K, Fairclough D, et al: Comparison of survival and quality of life in advanced non–small cell lung cancer patients treated with two dose levels of paclitaxel combined with cisplatin versus etoposide with cisplatin: results of an Eastern Cooperative Oncology Group trial. J Clin Oncol 18:623, 2000

57. Fossella F, Pereira JR, von Pawel J, et al: Randomized, multinational, phase III study of docetaxel plus platinum combinations versus vinorelbine plus cisplatin for advanced non-small-cell lung cancer: the TAX 326 study group. J Clin Oncol 21:3016, 2003

58. Kelly K, Crowley J, Bunn P, et al: A randomized phase III trial of paclitaxel plus carboplatin (PC) versus vinorelbine plus cisplatin (VC) in untreated advanced non–small cell lung cancer (NSCLC): a Southwest Oncology Group (SWOG) trial. J Clin Oncol 19:3210, 2001

59. Schiller JH, Harrington D, Belani CP, et al: Comparison of four chemotherapy regimens for advanced non–small cell lung cancer. N Engl J Med 346:92, 2002

60. Lilenbaum RC, Herndon J, List M, et al: Single-agent versus combination chemotherapy in advanced non-small cell lung cancer: a meta-analysis and the Cancer and Leukemia Group B randomized trial of efficacy, quality of life, and cost efffectiveness (abstr 2). Proc Am Soc Clin Oncol 21:71a, 2002

61. Tong D, Gillick L, Hendrickson FR: The palliation of symptomatic osseous metastases: final results of the Study by the Radiation Therapy Oncology Group. Cancer 50:893, 1982

62. Ellis PA, Smith IE, Hardy JR, et al: Symptom relief with MVP (mitomycin C, vinblastine and cisplatin) chemotherapy in advanced nonsmall cell lung cancer. Br J Cancer 71:366, 1995

63. Thatcher N, Hopwood P, Anderson H: Improving quality of life in patients with nonsmall cell lung cancer: research experience with gemcitabine. Eur J Cancer 33(suppl 1):S8, 1997

64. Effects of vinorelbine on quality of life and survival of elderly patients with advanced non–small cell lung cancer. The Elderly Lung Cancer Vinorelbine Italian Study Group. J Natl Cancer Inst 91:66, 1999

65. Randomized phase III trial of docetaxel versus vinorelbine or ifosfamide in patients with advanced non-small-cell lung cancer previously treated with platinum-containing chemotherapy regimens. The TAX 320 Non–Small Cell Lung Cancer Study Group. J Clin Oncol 18:2354, 2000

66. Shepherd FA, Dancey J, Ramlau R, et al: Prospective randomized trial of docetaxel versus best supportive care in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy. J Clin Oncol 18:2095, 2000

67. Cohen MH, Williams GA, Sridhara R, et al: FDA drug approval summary: gefinitib (ZD1839) (Iressa) tablets. Oncologist 8:303, 2003

68. Oda M, Tsuchiya R, Suzuki K, et al: Patterns of failure in phase II trial of postoperative adjuvant cisplatin/etoposide (PE) in patients with completely resected stage I-IIIA small cell lung cancer (SCLC): the Japan Clinical Oncology Lung Cancer Study Group Trial (JCOG9101) (abstr). Proc Am Soc Clin Onc 20:316a, 2001

69. Murray N: Small cell lung cancer at the millennium: radiotherapy innovations improve survival while new chemotherapy treatments remain unproven. Clin Lung Cancer 1:181, 2000

70. Murray N, Coy P, Peter JL, et al: Importance of timing for thoracic irradiation in the combined modality treatment of limited-stage small cell lung cancer. J Clin Oncol 11:336, 1993

71. Turrisi AT III, Kim K, Blum R, et al: Twice-daily compared with once-daily thoracic radiotherapy in limited small-cell lung cancer treated concurrently with cisplatin and etoposide. N Engl J Med 340:2002, 1999

72. Auperin A, Arriagada R, Pignon JP, et al: Prophylactic cranial irradiation for patients with small-cell lung cancer in complete remission. N Engl J Med 341:476, 1999

73. Elias A: Hematopoietic stem cell transplantation for small cell lung cancer. Chest 116(6 suppl), 531S, 1999

74. Grant SC, Kris MG, Houghton AN, et al: Long survival of patients with small cell lung cancer after adjuvant treatment with the anti-idiotypic antibody BEC2 plus bacillus Calmette-Guerin. Clin Cancer Res 5:1319, 1999

75. Ettinger DS: Carboplatin in the treatment of small cell lung cancer (SCLC). Lung Cancer 18(suppl 2):132, 1997

76. Kosmidis PA, Samantas E, Fountzilas G, et al: Cisplatin/etoposide versus carboplatin/etoposide chemotherapy and irradiation in small cell lung cancer: a randomized phase III study. Semin Oncol 21(suppl 6):23, 1994

77. Von Pawel J, Schiller JH, Shepherd FA, et al: Topotecan versus cyclophosphamide, doxorubicin, and vincristine for the treatment of recurrent small-cell lung cancer. J Clin Oncol 17:658, 1999

78. Noda K, Nishiwaki Y, Kawahara M, et al: Irinotecan plus cisplatin compared with etoposide plus cisplatin for extensive small-cell lung cancer. N Engl J Med 346:85, 2002

79. Zangemeister-Wittke U, Stahel RA: Novel approaches to the treatment of small-cell lung cancer. Cell Mol Life Sci 55:1585, 1999

80. Johnston E, Crawford J: Hematopoietic growth factors in the reduction of chemotherapy toxicity. Semin Oncol 25:552, 1998

81. Groopman JE, Itri LM: Chemotherapy-induced anemia in adults: incidence and treatment. J Natl Cancer Inst 91:1616, 1999

82. Crawford J, Demetri GD, Gabrilove JL, et al: Clinical benefits of epoetin alfa therapy in patients with lung cancer. Clin Lung Cancer 3:180, 2002

Acknowledgment

The author would like to acknowledge the Ruth and Herman Albert Thoracic Oncology Program at Duke Comprehensive Cancer Center for support of this work and, in particular, Thomas A. D'Amico, M.D., Joseph A. Govert, M.D., Michael J. Kelly, M.D., and Timothy Shafman, M.D., for their participation in the multidisciplinary development of this chapter.

IX PROSTATE CANCER

Philip W. Kantoff, m.d.

Epidemiology

Prostate cancer is the most commonly diagnosed noncutaneous malignancy in men in the United States. It is expected that by the end of 2004, 230,110 men in the United States will have been diagnosed with the disease and that 29,900 men will have died of it.[1] Incidence varies greatly, with African Americans having the highest incidence in the world (224 cases per 100,000 population). The incidence of prostate cancer in African Americans is remarkably higher than the incidence in white Americans (150 per 100,000), Western Europeans (39.6 per 100,000), Japanese (8.5 per 100,000), and Chinese (1.1 per 100,000).[2]

Throughout the latter half of the 20th century, the annual incidence of prostate cancer in the United States increased slowly but steadily, probably because of an increase in life expectancy and therefore an increase in the population of men in the susceptible age range (> 50 years). However, other factors likely contributed as well—in particular, the introduction and widespread use of the prostate-specific antigen (PSA) blood test. PSA testing identified a large number of asymptomatic men with prostate cancer, or so-called prevalence cases. Since peaking in 1993 at more than 350,000 cases, the rate of prostate cancer diagnosis initially decreased, presumably because of prior identification of prevalence cases, but is now increasing once more, for unknown reasons.[3] Despite the dramatic changes in the rates of prostate cancer diagnosis, mortality from prostate cancer in the United States declined 2.6% a year from 1990 to 1998.[4]

Risk Factors

Although the etiology of prostate cancer is unknown, several factors increase the risk of prostate cancer. These include age, race, family history, diet, and hormone levels.

age and latent disease versus clinical disease

Advancing age is the most obvious risk factor for prostate cancer; perhaps no other cancer is as age dependent. At the cellular level, however, the onset of prostate cancer appears to occur at an early age: prostate cancer has been detected at autopsy in the prostates of men in their 20s and 30s. Latent, or autopsy, prostate cancer (i.e., disease identified as an incidental finding at postmortem examination) is defined as microscopic quantities of low-grade cancer incidentally found in the prostate. In one study, 20% of men younger than 40 years had histologic evidence of prostate cancer.[5] In contrast, clinically detected prostate cancer is rare in men younger than 40 years. The prevalence of both latent and clinically detected prostate cancer increases dramatically with age: 20% to 30% of men older than 50 years and, perhaps, 50% of men older than 80 years may harbor latent prostate cancer. Most clinically detected prostate cancers are detected in the seventh and eighth decades of life. The relationship between latent and clinically detected prostate cancer is uncertain. Whether they represent two distinct diseases or latent prostate cancer is a precursor of clinically detected prostate cancer remains an unanswered question. If the latter is the case, the distinction may blur with earlier detection of prostate cancer.

racial and environmental variability

Other risk determinants of prostate cancer are race and geographic origin. African Americans have the highest incidence of prostate cancer. The lowest incidence rates are in Japan and China. When members of a racial group move from geographic regions of low incidence of prostate cancer to regions of higher incidence, the incidence of prostate cancer increases in that group. These migration studies demonstrate the probable importance of environmental factors. Genetic factors, however, may be operative as well. For example, the androgen receptor gene (which is coded on the X chromosome) contains a polymorphic cystosine-adenosine-guanine (CAG) repeat sequence. Several studies indicate that individuals with fewer CAG repeats in androgen receptor genes have a higher risk of prostate cancer.[6-9] African Americans have been found to have fewer CAG repeats than do Hispanic and non-Hispanic whites and Asians. Part of the increased risk of prostate cancer observed in African Americans may be attributed to this factor. Other, undefined genetic factors are probably involved as well. African Americans may develop a more virulent form of prostate cancer, leading to a higher mortality. A variety of factors, including genetics, diet, socioeconomics, and health care access, have been proposed to explain this observation.[10,11]

family history

The incidence of prostate cancer is higher in men who have relatives with prostate cancer. The risk increases with the number of affected relatives. First-degree relatives of men with prostate cancer have a twofold to threefold higher risk; men with two or more relatives with prostate cancer have a fivefold greater risk.[12] Overall, men who have relatives with other forms of cancer, including breast cancer, do not appear to be at significantly higher risk for prostate cancer.

diet

The relation between diet and prostate cancer is probably very important, although it is difficult to prove because most of the supporting data are from epidemiologic studies rather than prospective intervention studies. Nonetheless, the dramatic differences between the Asian and Western diets may contribute to the significant difference in risk. Data from large cohort studies and case-control studies support the contentions that red meat, animal fat, and higher total fat consumption increase risk.[13] Other dietary factors may also be important. Higher consumption of lycopene, a carotenoid present in tomatoes and beets, may decrease risk.[14] Micronutrients have been implicated: selenium, vitamin E, soy (isoflavones), and vitamin D consumption may lower risk, whereas a high dietary calcium intake may increase risk, perhaps by lowering levels of the vitamin D_3 metabolite 1,25-dihydroxyvitamin D[1,25$(OH)_2$D].

hormones and growth factors

Serum hormone and growth factor levels influence risk of prostate cancer. The relation between serum androgen level and prostate cancer risk is controversial. The prostate requires exposure to androgens both to increase in size and to develop cancer. Furthermore, individuals with very low androgen levels, such as

eunuchs, rarely develop prostate cancer. However, whether variations in the serum levels of androgen within the normal range are associated with differences in risk is debated. In the largest prospective study (conducted as part of the Physicians' Health Study), the higher the level of free testosterone, the greater the risk of subsequent prostate cancer.[15] In another analysis of the Physicians' Health Study, higher serum levels of insulinlike growth factor–1 (IGF-1) correlated with a greater risk of prostate cancer.[16] In the Health Professionals Follow-up Study, men with lower testosterone levels who subsequently developed prostate cancer were more likely to develop higher-grade prostate cancers, whereas men with higher testosterone levels tended to develop lower-grade prostate cancers.[17]

Pathogenesis

Prostate cancer is a complex, heterogeneous disease that likely results from a series of genetic events in the prostatic epithelial cell. These genetic events and many of the environmental factors that promote the development or progression of the cancer are as yet poorly defined.

A small subset of prostate cancers may result from the inheritance of one of several predisposition genes that confer a high risk of prostate cancer. Familial prostate cancer constitutes an estimated 5% to 10% of all prostate cancers and perhaps as many as 50% of prostate cancers in men younger than 55 years.[18] Six candidate loci have been identified in familial prostate cancer; the gene at one of these loci, hereditary prostate cancer–1 (*HPC-1*), has been found to be ribonuclease L (*RNase L*). *HPC-1* linkage may be associated with more severe prostate cancer.[19]

Several genes that participate in cell cycle regulation and in growth factor signaling (*Rb, p53*, and *PTEN*) have been implicated in the development and progression of prostate cancer. The prevalence of mutations of these genes has varied from study to study, and the role that the genes play in the development and progression of prostate cancer has yet to be defined. Similarly, a variety of proto-oncogenes have been implicated in the pathogenesis of prostate cancer. Although there is strong evidence of dysregulation of a variety of growth factors (e.g., IGF-1, epidermal growth factor [EGF], and platelet-derived growth factor), as well as growth factor receptors (e.g., Her-2/neu, EGF receptor, and transforming growth factor–β receptor), their exact roles are not yet understood.

Screening for Prostate Cancer

The goal of screening for prostate cancer is to detect organ-confined prostate cancer, which is potentially curable with definitive local therapy. Historically, the digital rectal examination (DRE) was used to screen for prostate cancer. DRE is inadequate, however, because its interpretation is highly variable, many cancers are not palpable, and most of the cancers detectable by DRE are not organ confined and therefore are not curable. Measurement of the serum acid phosphatase level also is a poor screening test, for similar reasons.

PROSTATE-SPECIFIC ANTIGEN TESTING

Screening by PSA testing greatly enhances detection of organ-confined prostate cancer. A member of the kallikrein family, PSA is a glycoprotein with serine protease activity. It is abundant in semen,[20] where it dissolves the seminal coagulum. PSA is produced by both normal and malignant prostatic epithelial cells; production may actually be higher in normal cells than in malignant cells. Thus, conditions such as benign prostatic hyperplasia (BPH), acute prostatitis, seminal ejaculation, and genitourinary instrumentation all increase the serum PSA level. When prostate cancer develops, the serum PSA level is frequently elevated because of the absolute increase in volume of epithelial cells and possibly because of leakage of PSA from tissue to serum.

Optimal screening for prostate cancer combines use of the PSA test and the DRE,[21] because most cancers are nonpalpable and some cancers do not produce enough PSA to increase the serum PSA level. Although no randomized studies have confirmed that screening with PSA decreases prostate cancer mortality, some evidence supports its use. In population-based studies from before the 1990s, when PSA screening became widespread, fewer than 50% of diagnosed cancers were organ confined. In contrast, most cancers that are currently detected by combined PSA and DRE screening are organ confined.[22] The PSA blood test can detect prostate cancer an average of 5.5 years before clinical detection becomes possible.[23] The majority of cancers now detected by PSA-based screening are nonpalpable, asymptomatic cancers that are potentially curable. Although these observations may represent no more than lead-time and length-time bias, perhaps the most compelling evidence to support PSA-based screening is the estimated 25% decrease in annual mortality from prostate cancer over the past 5 years.

The American Cancer Society currently recommends annual DRE and PSA tests for men older than 50 years who have a life expectancy of more than 10 years. For men at high risk (i.e., African Americans and men with a family history of prostate cancer), screening is recommended to begin at 40 years of age. Annual screening has been challenged, however; in one study, screening all men at ages 40 and 45 and then screening every 2 years starting at age 50 appeared to be more efficient.[24]

Troubling issues exist regarding PSA-based screening. The detection of clinically insignificant cancers may lead to unnecessary treatment for many men. Also, an elevated PSA level lacks specificity. Despite the increased likelihood of prostate cancer in men with a moderately elevated serum PSA level (4 to 10 ng/ml), biopsy in such cases usually reveals BPH rather than prostate cancer.

Although most men with serum PSA levels above 10 ng/ml have prostate cancer, their cancers are highly likely to be non–organ confined. The lack of specificity of a serum PSA between 4 and 10 ng/ml has led to intense research over the past 10 years. Different strategies have been explored to distinguish malignancy from BPH.

Because serum PSA levels normally increase with age, the use of age-specific PSA levels can enhance the sensitivity of screening for prostate cancer in young men and increase the specificity of an elevated serum PSA in older men [*see Table 1*].[25] This strate-

Table 1 Age-Specific PSA Reference Ranges

Age	*PSA Level*
< 50 yr	≤ 2.5 ng/ml
50–59 yr	≤ 3.5 ng/ml
60–70 yr	≤ 4.5 ng/ml
> 70 yr	≤ 6.5 ng/ml

Table 2 Free PSA and Probability of Cancer in Patients with PSA Levels between 4 and 10 ng/ml[28]

Free PSA (%)	*Probability of Prostate Cancer (%)*
0–10	56
10–15	28
15–20	20
20–25	16
> 25	8

gy is particularly important in younger men, in whom prostate cancer is more likely to be fatal. This strategy has been challenged for older men, however, because sensitivity is sacrificed for a small increase in specificity.

Race-specific serum PSA ranges have also been proposed, because healthy African Americans have mean serum PSA levels that are slightly higher than those of whites.[26] Serum PSA levels increase faster in men with prostate cancer than in men without the disease, but the use of this principle (so-called PSA velocity) in determining who should undergo biopsies is limited because of the normal variation in PSA levels.

Another strategy is determination of the ratio of the serum PSA to the volume of the prostate gland (so-called PSA density [PSAD]).[27] In general, patients with cancer have a higher PSAD. A major limitation with this approach has been the difficulty in objectively assessing prostate volume through transrectal ultrasonography (TRUS).

Determination of the free PSA (i.e., the percentage of PSA that is unbound to serum proteins) is also a potential means of distinguishing malignancy from benign hyperplasia.[28] For unclear reasons, PSA derived from malignant epithelial cells tends to bind more avidly to serum proteins. Thus, in men with an elevated serum PSA level, cancer is more likely to be present when the percentage of free PSA is low [*see Table 2*]. Because sensitivity and specificity vary with respect to each other, depending on the free PSA cutoff, the optimal cutoff value for free PSA is still under debate.

The insensitivity of PSA levels below 4.0 ng/ml is also an issue. In one large study, 15.2% of men with a PSA level of less than 4.0 ng/ml who underwent biopsy had prostate cancer; 14.9% of these cancers were high grade.[29] Whether it would be prudent to lower the PSA threshold for biopsy is controversial. Although the number of cancers detected would increase, the rate of overdiagnosis and overtreatment would increase as well.

Diagnosis

A prostate biopsy is generally performed when cancer is suspected because of an elevated serum PSA level, an abnormal DRE, or both. Biopsies are performed transrectally: a TRUS probe is inserted into the rectum to visualize the prostate during biopsy sampling. TRUS itself is neither an adequate screening test nor a means of reliably visualizing nonpalpable cancer; its major contribution is in ensuring the accurate spatial sampling of the prostate gland. TRUS also is helpful in the estimation of gland volume, information that may be useful in making decisions about treatment. Traditionally, biopsy samples have been taken in a sextant pattern (bilateral base, midgland, and apex). This approach often missed cancer, particularly in the apex (the portion nearer the anal verge) and, less frequently, in the base (the portion nearer the bladder). Increasing the number of samples per session improves cancer detection,[30,31] but at the risk of added morbidity from multiple samples and the increased possibility of detecting clinically insignificant cancer.

BIOPSY FINDINGS

The vast majority of prostate cancers are adenocarcinomas; small cell carcinomas and sarcomas are occasionally seen. Adenocarcinomas arise from the glandular epithelium of the prostate and are often multifocal. The diagnosis is often made on the basis of a limited amount of gland pathology because of the limitations of sampling the gland. The subtlest pathologic change associated with carcinoma is loss of the basal layer within the glandular epithelium. Invasion into the surrounding stromal tissue is common and makes the diagnosis easier.

HISTOLOGIC GRADING

The most commonly used grading system is the Gleason grading system, in which tumors are classified by the degree of disorganization of glandular structures. Tumors are graded from 1 (least malignant) to 5 (most malignant) on the basis of the architectural patterns in the examined tissue. The grades reflecting the two most common architectural patterns are then added together to create the Gleason score (e.g., 4 + 3 = 7). The Gleason score correlates with clinical prognosis and is therefore used in stratifying patients. However, the majority of tumors are classified as either Gleason 6 or 7, which limits the clinical utility of the Gleason score.

A condition known as prostate intraepithelial neoplasia (PIN), which represents a premalignant state,[32] is felt to predate true carcinoma and often coexists with carcinoma in the prostate gland. When biopsy reveals PIN but no actual cancer, further biopsies are warranted.

Staging

Staging in prostate cancer attempts to determine which cancers are organ confined (and thus curable with local therapy alone) and which are not organ confined. Prostate cancer has a propensity to spread to lymph nodes and bone. Clinically significant spread to visceral organs, such as the liver or lung, is unusual.

CLINICAL STAGING

The clinical stage is based on the extent of disease assessed by palpation (i.e., DRE). The Marshall Jewett classification (in which stages are denoted as A, B, C, or D) has largely been replaced with the TNM classification [*see Table 3*].

Clinical staging of prostate tumors distinguishes between localized (T1 and T2) tumors and locally advanced (T3 and T4) tumors. T1a and T1b prostate tumors (stage A tumors in the Marshall Jewett system) are detected by transurethral resection. These tumors, which probably originate in the transition zone of the prostate, are generally indolent. Currently, most of the prostate cancers that are detected are nonpalpable (T1c) cancers. Although T1c and T2 tumors generally behave in a similar fashion, their distinction from T3 and T4 tumors is critical because T3 and T4 tumors will rarely be cured by surgery or radiation therapy alone. Clinical staging is otherwise of limited value in judging curability and extent of disease and in determining appropriate treatment.

Table 3 Clinical Staging Definitions

Malignancy	*Stage*	*Characteristics*
Primary Tumor (T)	TX	Primary tumor cannot be assessed
	T0	No evidence of primary tumor
	T1	Clinically inapparent tumor not palpable or visible by imaging
	T1a	Tumor incidental histologic finding in ≤ 5% of tissue resected
	T1b	Tumor incidental histologic finding in > 5% of tissue resected
	T1c	Tumor identified by needle biopsy (e.g., because of elevated PSA)
	T2	Palpable tumor confined within prostate*
	T2a	Tumor involves one lobe
	T2b	Tumor involves both lobes
	T3	Tumor extends through the prostatic capsule†
	T3a	Extracapsular extension (unilateral or bilateral)
	T3b	Tumor invades seminal vesicle(s)
	T4	Tumor is fixed or invades adjacent structures other than seminal vesicle, bladder neck, external sphincter, rectum levator muscle, and/or pelvic wall
Pathologic Primary Tumor (pT)	pT2‡	Organ confined
	pT2a	Unilateral
	pT2b	Bilateral
	pT3	Extraprostatic extension
	pT3a	Extraprostatic extension
	pT3b	Seminal vesicle infusion
	pT4	Invasion of bladder, rectum
Regional Lymph Node Metastasis	NX	Regional lymph node cannot be assessed
	N0	No regional lymph node metastasis
	N1	Metastasis in regional lymph node or nodes
Distant Metastasis§	MX	Distant metastasis cannot be assessed
	M0	No distant metastasis
	M1	Distant metastasis
	M1a	Nonregional lymph nodes
	M1b	Bone(s)
	M1c	Other site(s)

*A tumor that is found in one or both lobes by needle biopsy but is not palpable or reliably visible by imaging is classified as T1a.
†Invasion into the prostatic apex or into (but not beyond) the prostatic capsule is classified not as T3 but as T2.
‡There is no pathologic T1 classification.
§When more than one site of metastasis is present, the most advanced category is used.

MULTIFACTORIAL STAGING

The most clinically useful means of stratifying patients according to prognosis is through multifactorial staging—that is, using the clinical stage, the serum PSA level, and the Gleason score together.[33] Combined-modality staging can determine whether a patient is at low, intermediate, or high risk of biochemical relapse (i.e., relapse as evidenced on PSA testing) after local therapy [*see Table 4*].[34] In addition, refinements have been made in stratifying patients by use of the proportion of biopsy samples that contain cancer.[35] Although such classifications have proved helpful in distinguishing patients with organ-confined cancer from those with non–organ-confined cancer, thus facilitating clinical decision-making, they have not allowed discrimination of indolent cancers (which do not need to be treated) from clinically relevant tumors.

Although the use of PSA measurement to detect and stage prostate cancer has declined, changes in PSA before diagnosis may have prognostic significance. In a recent study, time to death—from prostate cancer and from any cause—was significantly shorter in men whose PSA level increased by 2 ng/ml or more in the year before the diagnosis of prostate cancer, compared with men whose PSA levels increased at a lower rate; 10-year mortality was 15% in the former group and negligible in the latter.[36]

STAGING BASED ON IMAGING STUDIES

Currently, prostate cancer is almost always diagnosed in men who have no radiographic evidence of metastases; few men come to medical attention as a result of symptoms of metastatic disease, which, most frequently, is pain resulting from osteoblastic bone metastases. For that reason, radiographic and scintigraphic methods of staging are not useful for the vast majority of patients.[37] However, bone scans to assess skeletal involvement and computed tomographic scans or magnetic resonance imaging scans to assess regional lymph nodes may be useful in patients with prostate cancers whose serum PSA level is above 10 ng/ml, whose Gleason score is 8 to 10, or who have locally advanced cancer (T3 or T4). Endorectal coil MRI scans (erMRIs) offer better definition of the prostate gland than does TRUS but are useful in staging in only a small fraction of patients (i.e., those with a serum PSA level between 10 and 20 ng/ml, a Gleason score of less than 8, and more than 50% of biopsy cores positive). In such patients who appear clinically to have organ-confined disease, erMRI is 78% accurate in predicting extracapsular disease and 90% accurate in predicting seminal vesical involvement.[38] Scintigraphy with ProstaScint (indium-111 capromab pendetide), a radiolabeled murine monoclonal antibody to a prostate-specific membrane antigen, is also of limited value in initial staging.[39]

Table 4 Combined-Modality Staging Approach

Risk	*5-Year PSA Failure-Free Survival*	*Characteristics*
Low	> 85%	T1c, T2a *and* PSA ≤ 10 ng/ml *and* Gleason score ≤ 6
Intermediate	50%	T2b *or* PSA > 10 and ≤ 20 ng/ml *or* Gleason score = 7
High	33%	T2c *or* PSA > 20 ng/ml *or* Gleason score ≥ 8

Natural History of Untreated Prostate Cancer

Data regarding the outcome of patients who receive no therapy for their prostate cancer are available from a series of over 3,000 men identified in the Connecticut Tumor Registry who, after diagnosis, either chose not to be treated or were advised by their physicians to receive no treatment.[40] These patients were followed for a minimum of 15 years or until death. The two major determinants of death from prostate cancer were tumor grade and age at diagnosis. According to the study findings, men with prostate cancer whose Gleason scores are 7 or higher and who are younger than 74 years face significant prostate cancer mortality; conversely, men of all ages whose Gleason scores are 6 or lower are at minimal to modest risk of death from prostate cancer. This study resulted in two important observations: Gleason 7 tumors should be regarded as high-grade tumors (along with Gleason 8 to 10 tumors), and age at diagnosis appears to have little impact on cancer-specific mortality after 15 years in patients with low-grade or high-grade tumors. Age is more a determinant in men with moderately differentiated tumors.

Treatment of Localized Prostate Cancer

When prostate cancer is confined to the organ, surgery or radiation therapy is commonly used with great success. Reducing the morbidity associated with treatment remains a challenge in such patients, however. In men with non–organ-confined cancer, who very frequently harbor occult metastatic disease, surgery or radiation therapy provides benefits only with respect to local disease control.

A prospective trial in which patients with localized prostate cancer were randomized to undergo either radical prostatectomy or observation found no difference in overall survival between the two arms of the study over a relatively brief follow-up period of 6.2 years.[41] However, early indicators of benefit were seen with surgery—a reduction in cancer-specific mortality and in metastasis rate. Single-institution reports have alleged that different treatment modalities are effective in the management of early prostate cancer,[42-44] but comparative data are hard to obtain because of potential bias in patient selection. Information from pooled hospital-based datasets[45,46] or from population-based databases, such as the Surveillance, Epidemiology, and End Results (SEER) database,[47] suggest only modest short-term benefits, with those benefits observed only in high-grade tumors. The 10-year survival probabilities in these studies are similar to the nontreatment data from the Connecticut Tumor Registry, largely because of the excellent short-term prognosis for patients with untreated prostate cancer. It is therefore not surprising that in studies in which data from SEER and hospital-based treatment series were analyzed, the outcomes in patients receiving radical prostatectomy and external-beam radiation therapy—the most commonly used treatment modalities—were similar 10 years after treatment [*see Table 5*].[48,49]

In the absence of randomized studies showing a therapeutic superiority of surgery over radiation therapy, treatment decisions are largely made by patients themselves after physician consultation. A mixture of physicians' bias, patient priorities and fears regarding cancer, attitudes toward treatment-related side effects, and comorbidity comes into play.

RADICAL PROSTATECTOMY

In the United States, radical prostatectomy has been the standard treatment and may offer the greatest chance of cancer control for patients with organ-confined prostate cancer. The procedure is most commonly performed in a retropubic fashion, although a transperineal approach is sometimes used. The transperineal approach has the potential advantage of a shorter postoperative hospitalization, whereas the retropubic approach allows lymph node dissection, preservation of the nerves that control erectile function, and wider resection margins.

Table 5 10-Year Disease-Specific Survival (DSS) of Patients with Localized Prostate Cancer by Grade and Treatment

Grade	*Prostatectomy*		*Radiotherapy*		*Conservative*	
	N	*% DSS (95% CI)*	*N*	*% DSS (95% CI)*	*N*	*% DSS (95% CI)*
I	3,854	94 (91–95)	4,065	90 (87–92)	9,804	93 (91–94)
II	14,287	87 (85–89)	7,939	76 (72–79)	6,198	77 (74–80)
III	5,133	67 (62–71)	2,596	53 (47–58)	2,236	45 (40–51)

CI—confidence interval N—number of patients

RADIATION THERAPY

Computer-aided planning systems, such as CT simulation and so-called conformal three-dimensional planning (in which radiation conforms to the contours of the gland), allow radiation therapy to be administered with fewer side effects and at doses higher than the traditional 7,000 cGy. Brachytherapy (i.e., implantation of radioactive iodine 125 or palladium 103 seeds directly into the tumor) is a very popular alternative therapeutic strategy because of the relative ease of administration, the apparently favorable toxicity profile, and the promising cancer control rates at 5 to 10 years of follow-up.

The efficacy of radiation therapy remains a major concern. With external-beam radiation therapy, the dose applied (7,000 cGy) may be insufficient to eradicate all cancer. With brachytherapy, implant placement may leave so-called cold spots where cancer can persist. These concerns probably apply more to patients with intermediate-risk and high-risk prostate cancer, in which tumor volumes are on average higher, but these concerns are also relevant to low-risk patients, particularly younger men.

TREATMENT-RELATED SIDE EFFECTS

Radical prostatectomy can result in urinary incontinence and erectile dysfunction; the frequency and severity of these side effects are a subject of debate.[50,51] Reports of low frequencies of these side effects may reflect patient selection, better surgical expertise, or both, although inaccuracy in obtaining symptom data may be a more realistic explanation.[52,53] A review of the SEER database suggests that the rates of postoperative and late urinary complications in men who undergo radical prostatectomy are significantly lower when the procedure is performed in a high-volume hospital and by a surgeon who performs a high number of these procedures.[54]

Stress urinary incontinence is reported in 5% to 35% of patients after radical prostatectomy.[50] Erectile dysfunction occurs frequently after radical prostatectomy, rendering most patients impotent, even when the nerves that control erectile function are spared.[50,51] In selected series, however, potency was reportedly maintained in more than half of the patients.[52,53]

External-beam radiation therapy is generally well tolerated, but erectile dysfunction occurs in the majority of patients, and rectal injury resulting in altered bowel habits or minor rectal bleeding occurs in a minority of patients.[50] Insufficient data are available to make any assessment of the long-term sequelae of brachytherapy.

Comparison of the side effects of radical prostatectomy and external-beam radiation therapy has shown that men who undergo radical prostatectomy are more likely to have urinary incontinence or impotence (although significant declines in sexual function are seen with both treatments). On the other hand, men who receive external-beam radiation therapy are more likely to suffer changes in bowel function.[55,56] Because concern about particular side effects may vary with individual patients, this information may be useful in guiding treatment decisions.

TREATMENT SELECTION BY RISK OF RECURRENCE

Low-Risk Patients

Low-risk patients have a high likelihood of organ-confined cancer [*see Table 4*]. Treatment options include radical prostatectomy, brachytherapy, external-beam radiation therapy, or no treatment. Outcomes are generally excellent, regardless of therapy, in this subset of patients. Although radical prostatectomy remains the standard approach, nonrandomized data suggest equivalence among the treatment options for this subgroup.[49,57] A significant subset of patients with low-risk prostate cancer probably do not require treatment, but the accurate identification of such patients remains uncertain.

Intermediate-Risk Patients

Intermediate-risk patients usually have a higher cancer volume and a higher likelihood of occult metastatic disease. Radical prostatectomy can be offered to these patients, but many will require further therapy, either because of recurrent or persistent local cancer or because of metastatic cancer at distant sites, as indicated by recurrent or persistent PSA elevation after treatment. Nonetheless, surgery is curative in about 50% of these cases. Efforts at improving upon the results of surgery by reducing tumor volume through androgen ablation (i.e., chemical or surgical castration) before surgery have thus far been unsuccessful.[58-60] Whether external-beam radiation therapy can eradicate all tumor cells, even when organ confined, is not established. One strategy to address this concern has been to increase the dose of radiation by use of conformal radiation, which allows higher-intensity treatment without excess tissue damage.[61] Alternative approaches combine external-beam radiation therapy either with brachytherapy[62] or with the administration of protons (generated by a cyclotron).[63] Although these approaches appear promising, none have been evaluated in a randomized trial. Several randomized trials suggest that combining external-beam radiation therapy with androgen ablation (to shrink the prostate gland) in intermediate-risk patients improves cancer control or survival.[64-66] However, the optimal timing and duration of androgen ablation are still under investigation. One randomized study found that 6 months of androgen ablation, used concomitantly with radiation, provided a survival benefit in this patient population.[67]

High-Risk Patients

High-risk patients and those with locally advanced prostate cancer (T3 and T4) have the highest likelihood of occult metastatic disease. Local therapy alone is rarely curative, making the morbidity of radical prostatectomy difficult to justify. Many such patients are now treated with androgen ablation followed by external-beam radiation therapy, although even with this multimodality approach, long-term results are poor.

ADJUVANT THERAPY FOR LOCALIZED PROSTATE CANCER

Efforts to improve survival in high-risk cases include giving additional (adjuvant) treatment once local treatment has been completed. After surgery, radiation therapy may be offered to patients with such adverse pathologic findings as resection margins involved by tumor, capsular penetration, or seminal vesical involvement. No controlled studies have indicated whether this approach influences outcome. Androgen ablation has also been used after either surgery or radiation therapy. Three years of androgen-deprivation therapy resulted in a statistically significant increase in survival in a randomized European study of 415 men with locally advanced or high-grade prostate cancer that was treated with external beam radiation.[65] No studies have yet examined the role of adjuvant cytotoxic chemotherapy after local therapy.

Treatment of Advanced Prostate Cancer

The spectrum of advanced prostate cancer has changed dramatically since the late 1980s, largely because of the use of the

PSA test. Most cases are diagnosed earlier, and treatment failure is recognized sooner. As a result, patients with advanced prostate cancer now include many men who undergo surgery or radiation therapy and subsequently experience serologic progression (i.e., a rising serum PSA level). A few of these patients can be cured with further local treatment to the prostate, but most will ultimately require systemic therapy.

SALVAGE TREATMENT

Patients who undergo radical prostatectomy and who subsequently relapse, with detectable and rising serum PSA levels, can be successfully treated with radiation therapy in some cases. Factors that appear to predict curability in this setting include organ-confined prostate cancer at the time of initial treatment, longer time to disease recurrence (> 2 years for a rise in PSA level), and a serum PSA level less than 1 ng/ml at the time of salvage radiation therapy. Salvage radiation therapy can generally be given with minimal morbidity. In contrast, disease recurrence after initial radiation therapy poses a much more difficult problem because salvage radical prostatectomy is associated with a very high morbidity. If either salvage treatment is planned, it is appropriate to restage the cancer with bone scan, CT scan, MRI, and possibly scintigraphy with ProstaScint [*see* Staging, *above*].

ANDROGEN ABLATION

The standard initial treatment for patients with advanced prostate cancer is androgen ablation, a therapeutic strategy that involves either lowering the production of testosterone or blocking its binding to the androgen receptor. Prostate cancer cells require androgen for growth and possibly survival, so removal of androgen will result in cell dormancy or apoptosis. In turn, the serum PSA level generally decreases promptly and dramatically. For the vast majority of patients, symptoms from metastases will lessen or disappear. Radiographically, metastases may diminish or remain the same; complete radiographic responses are unusual. For asymptomatic patients, the only evidence of a response to androgen ablation is a decrease in the serum PSA level.

Androgen ablation is achieved by a variety of strategies. Castration or diminishing testosterone production can be achieved surgically (i.e., with orchiectomy, which removes the testosterone-secreting organs) or chemically (i.e., through use of luteinizing hormone–releasing hormone [LHRH] agonists, which diminish luteinizing hormone and thereby testosterone production by the testicles). The permanency of orchiectomy, as well as the emotional stigma associated with surgical castration, has made the use of LHRH agonists (e.g., leuprolide and goserelin) popular. These agents are administered either monthly or every 3 or 4 months. Other modalities include antiandrogens (e.g., flutamide, bicalutamide, and nilutamide), which are drugs that bind to the androgen receptor competitively. They have mostly been used in conjunction with chemical or surgical castration in an effort to provide maximal or combined androgen blockade. The utility of antiandrogens as sole agents is not completely established.

Androgen ablation is associated with significant side effects, including hot flashes, loss of libido, impotence, gain of adipose tissue, and loss of muscle and bone mass. Despite these side effects, most men tolerate androgen ablation reasonably well. Disabling hot flashes may respond to treatment with clonidine, venlafaxine, or gabapentin.[68]

Prostate cancer typically recurs after androgen ablation. A rise in the serum PSA level is generally the first sign of recurrence, usually predating other evidence of recurrence by months or sometimes years. The duration of response to androgen ablation is highly variable, reflecting the degree of androgen sensitivity of the tumor. In patients with overt metastatic disease, the duration of response is about 12 to 18 months; in patients without evidence of metastasis, the duration is much longer.

Estrogen (e.g., diethylstilbestrol) also lowers LH levels and thereby lowers testosterone levels, and it may have direct effects on prostate cancer cells. The use of estrogens was popular in the past but fell into disfavor because of such side effects as gynecomastia and thromboembolic phenomena.

Timing of Androgen Ablation

Is it better to start androgen ablation therapy along with primary treatment or better to wait until symptomatic metastatic disease appears? This question is controversial, because androgen ablation therapy is palliative rather than curative, and its effect on survival is uncertain.

In three randomized studies, early androgen ablation therapy slowed progression of disease and prolonged survival in patients who received primary therapy with either radiation or prostatectomy for locally advanced or metastatic prostate cancer.[65,69,70] Similarly, a meta-analysis of older studies found that early androgen suppression reduces disease progression and related complications and may slightly improve 10-year survival.[71]

Optimal Androgen-Deprivation Therapy

Many studies have helped define the optimal regimen or regimens of androgen-deprivation therapy. Randomized comparisons of monotherapies (i.e., orchiectomy, LHRH agonists, and estrogens) have shown that although these modalities have differing side effects, they appear to have an equivalent impact on survival.[72,73] The possible exception is antiandrogen monotherapy, which appears to be inferior to other modalities in patients with established metastases.

The adrenal androgens are relatively weaker than testosterone, and their role in supporting prostate cancer cell growth is uncertain. Nevertheless, regimens have been devised to suppress production of both testicular and adrenal androgens to provide maximal androgen deprivation. The concept of combined androgen blockade (CAB) was objectively assessed in a multi-institutional trial in which 603 men with metastatic prostate cancer were randomized to receive either leuprolide alone or leuprolide plus flutamide.[74] Compared with leuprolide alone, CAB resulted in a significant improvement in time to disease progression and median survival. However, a larger and more recent study, which enrolled 1,387 men with metastatic prostate cancer and compared orchiectomy alone with orchiectomy plus flutamide, found no difference in overall survival between the two groups.[75] Three meta-analyses on this topic have concluded that if CAB does benefit patients with established metastases, that benefit is extremely modest.

CHEMOTHERAPY

Chemotherapy has a clear role in patients with hormone-refractory prostate cancer. Two randomized trials of mitoxantrone plus a corticosteroid documented improvements in quality of life (albeit no improvement in survival) with the use of mitoxantrone.[76,77] These two studies led to the approval of mitoxantrone by the Food and Drug Administration for the treatment of hormone-refractory prostate cancer.

An important recent finding is that taxanes (i.e., paclitaxel and docetaxel) are active against prostate cancer. These drugs inhibit mitosis in prostate epithelial cells by interfering with microtubular assembly in the nucleus. Two randomized studies demonstrated longer survival for men treated with docetaxel than for those treated with mitoxantrone plus prednisone.[78,79] Docetaxel plus prednisone is now the standard chemotherapy for men with metastatic prostate cancer.

PALLIATION OF BONE PAIN

Because prostate cancer tends to spread to bone, bone pain is a major source of morbidity in patients with advanced disease. The mainstays of treatment for bone pain are analgesics (i.e., nonsteroidal anti-inflammatory drugs and narcotics) and external-beam radiation therapy. Radiation therapy is effective at reducing or eliminating pain at specific sites in over 90% of patients within 2 weeks of treatment. Radiopharmaceuticals such as strontium-89 and samarium-153–labeled ethylenediaminetetramethylenephosphonate (EDTMP) are sometimes useful in patients with multiple sites of pain. These injectable isotopes concentrate in osseous tissue. Pain relief usually occurs within a few weeks after treatment, although 15% of patients experience a flare in pain within the first 2 weeks. These agents are generally well tolerated, but progressive bone marrow suppression may occur with repeat dosing. Their mechanism of activity is uncertain, but pain relief in patients who receive these agents is rarely coupled with a decline in serum PSA levels.

Prevention of Prostate Cancer

Risk reduction and chemoprevention of prostate cancer are important objectives. The 5α-reductase inhibitor finasteride may interfere with the development or progression of prostate cancer. Currently approved for use in BPH, finasteride reduces the conversion of testosterone to dihydrotestosterone, a more potent androgen, and decreases prostate size [*see 10:XIII Benign Prostatic Hyperplasia*]. Although a large, double-blind, placebo-controlled trial (the Prostate Cancer Prevention Trial) demonstrated a 24.4% reduction in primary prostate cancer with finasteride, study patients who developed prostate cancer while taking finasteride were more likely to have higher-grade cancer.[80] This puzzling finding has stymied the use of finasteride as a preventive agent.

Selenium and vitamin E may lower the risk of prostate cancer. That possibility is being tested in the Selenium and Vitamin E Cancer Prevention Trial (SELECT), a randomized, prospective, double-blind study.[81] Along with high lycopene consumption [*see* Diet, *above*], high consumption of vegetables (particularly cruciferous vegetables)[82] and consumption of fish (especially fatty fish such as salmon)[83] may reduce the risk of prostate cancer. Large studies have been inconclusive on possible links between obesity and prostate cancer—but whether a link exists or not, reduction of excess weight is still indicated, given the many other health risks of obesity.[84]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Cancer Facts & Figures 2004. American Cancer Society, 2004 http://www.cancer.org/downloads/STT/CAFF_finalPWSecured.pdf

2. Parkin DM, Bray FI, Devesa SS: Cancer burden in the year 2000: the global picture. Eur J Cancer 37(suppl 8):S4, 2001

3. Jemal A, Clegg LX, Ward E, et al: Annual report to the nation on the status of cancer, 1975–2001, with a special feature regarding survival. Cancer 101:3, 2004

4. Recent trends in mortality rates for four major cancers, by sex and race/ethnicity—United States, 1990–1998. MMWR Morb Mortal Wkly Rep 51(3):49, 2002

5. Sakr WA, Haas GP, Cassin BF, et al: The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients [see comments]. J Urol 150(pt 1):379, 1993

6. Giovannucci E, Stampfer MJ, Krithivas K, et al: The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 94:3320, 1997

7. Stanford JL, Just JJ, Gibbs M, et al: Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res 57:1194, 1997

8. Ingles SA, Ross RK, Yu MC, et al: Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. J Natl Cancer Inst 89:166, 1997

9. Hakimi JM, Schoenberg MP, Rondinelli RH, et al: Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. Clin Cancer Res 3:1599, 1997

10. Myers RE: African American men, prostate cancer early detection examination use, and informed decision-making. Semin Oncol 26:375, 1999

11. Robbins AS, Whittemore AS, Thom DH: Differences in socioeconomic status and survival among white and black men with prostate cancer. Am J Epidemiol 151:409, 2000

12. Steinberg GD, Carter BS, Beaty TH, et al: Family history and the risk of prostate cancer. Prostate 17:337, 1990

13. Bosland MC, Oakley-Girvan I, Whittemore AS: Dietary fat, calories, and prostate cancer risk. J Natl Cancer Inst 91:489, 1999

14. Giovannucci E, Rimm EB, Liu Y, et el: A prospective study of tomato products, lycopene, and prostate cancer risk. J Natl Cancer Inst 94:391, 2002

15. Gann PH, Hennekens CH, Ma J, et al: Prospective study of sex hormone levels and risk of prostate cancer. J Natl Cancer Inst 88:1118, 1996

16. Chan JM, Stampfer MJ, Giovannucci E, et al: Plasma insulin-like growth factor–I and prostate cancer risk: a prospective study. Science 279:563, 1998

17. Platz EA , Leitzmann MF, Rifai N, et al: Sex steroid hormones and the androgen receptor gene CAG repeat and subsequent risk of prostate cancer in the PSA era. Cancer Epidemiol Biomarkers Prev (in press)

18. Carter BS, Beaty TH, Steinberg GD, et al: Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 89:3367, 1992

19. Goode EL, Stanford JL, Peters MA, et al: Clinical characteristics of prostate cancer in an analysis of linkage to four putative susceptibility loci. Clin Cancer Res 7:2739, 2001

20. Brawer MK: Prostate-specific antigen. Semin Surg Oncol 18:3, 2000

21. Catalona WJ, Smith DS, Ratliff TL, et al: Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 324:1156, 1991

22. Crawford ED, DeAntoni EP, Etzioni R, et al: Serum prostate-specific antigen and digital rectal examination for early detection of prostate cancer in a national community-based program. The Prostate Cancer Education Council. Urology 47:863, 1996

23. Gann PH, Hennekens CH, Stampfer MJ: A prospective evaluation of plasma prostate-specific antigen for detection of prostate cancer. JAMA 273:289, 1995

24. Ross KS, Carter HB, Pearson JD, et al: Comparative efficiency of prostate-specific antigen screening strategies for prostate cancer detection. JAMA 284:1399, 2000

25. Dalkin BL, Ahmann FR, Kopp JB: Prostate specific antigen levels in men older than 50 years without clinical evidence of prostatic carcinoma. J Urol 150:1837, 1993

26. Morgan TO, Jacobsen SJ, McCarthy WF, et al: Age-specific reference ranges for prostate-specific antigen in black men. N Engl J Med 335:304, 1996

27. Benson MC, Olsson CA: Prostate specific antigen and prostate specific antigen density: roles in patient evaluation and management. Cancer 74:1667, 1994

28. Catalona WJ, Partin AW, Slawin KM, et al: Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. JAMA 279:1542, 1998

29. Thompson IM, Pauler DK, Goodman PJ, et al: Prevalence of prostate cancer among men with a prostate-specific antigen level of 4.0 ng per milliliter. N Engl J Med 350:2239, 2004

30. Chang JJ, Shinohara K, Bhargava V, et al: Prospective evaluation of lateral biopsies of the peripheral zone for prostate cancer detection. J Urol 160(pt 1):2111, 1998

31. Presti JC Jr, Chang JJ, Bhargava V, et al: The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: results of a prospective clinical trial. J Urol 163: 163, 2000

32. Bostwick DG: Prostatic intraepithelial neoplasia is a risk factor for cancer. Semin Urol Oncol 17:187, 1999

33. Partin AW, Kattan MW, Subong EN, et al: Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer: a multi-institutional update. JAMA 277:1445, 1997

34. D'Amico AV, Whittington R, Malkowicz SB, et al: Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. J Clin Oncol 17:168, 1999

35. D'Amico AV, Whittington R, Malkowicz SB, et al: Clinical utility of the percentage of positive prostate biopsies in defining localized prostate cancer. J Clin Oncol 18:1164, 2000

36. D'Amico AV, Chen M-H, Roehl KA, et al: Preoperative PSA velocity and the risk of death from prostate cancer after radical prostatectomy. N Engl J Med 351:125, 2004

37. Lin K, Szabo Z, Chin BB, et al: The value of a baseline bone scan in patients with newly diagnosed prostate cancer. Clin Nucl Med 24:579, 1999

38. D'Amico AV, Whittington R, Maldowicz SB, et al: Critical analysis of the ability of the endorectal coil magnetic resonance imaging scan to predict pathologic stage, margin status, and postoperative prostate-specific antigen failure in patients with clinically organ-confined prostate cancer. J Clin Oncol 14:1770, 1996

39. Murphy GP, Snow PB, Brandt J, et al: Evaluation of prostate cancer patients receiving multiple staging tests, including ProstaScint scintiscans. Prostate 42:145, 2000

40. Albertsen PC, Hanley JA, Gleason DF, et al: Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. JAMA 280:975, 1998

41. Holmberg L, Bill-Axelson A, Helgesen F, et al: A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. N Engl J Med 347:781, 2002

42. Catalona WJ, Smith DS: 5-Year tumor recurrence rates after anatomical radical retropubic prostatectomy for prostate cancer. J Urol 152(pt 2):1837, 1994

43. Hanks GE, Hanlon AL, Schultheiss TE, et al: Dose escalation with 3D conformal treatment: five year outcomes, treatment optimization, and future directions. Int J Radiat Oncol Biol Phys 41:501, 1998

44. Ragde H, Blasko JC, Grimm PD, et al: Brachytherapy for clinically localized prostate cancer: results at 7- and 8-year follow-up. Semin Surg Oncol 13:438, 1997

45. Gerber GS, Thisted RA, Scardino PT, et al: Results of radical prostatectomy in men with clinically localized prostate cancer. JAMA 276:615, 1996

46. Shipley WU, Thames HD, Sandler HM, et al: Radiation therapy for clinically localized prostate cancer: a multi-institutional pooled analysis. JAMA 281:1598, 1999

47. Lu-Yao GL, Yao SL: Population-based study of long-term survival in patients with clinically localized prostate cancer. Lancet 349:906, 1997

48. Keyser D, Kupelian PA, Zippe CD, et al: Stage T1-2 prostate cancer with pretreatment prostate-specific antigen level < or = 10 ng/ml: radiation therapy or surgery? Int J Radiat Oncol Biol Phys 38:723, 1997

49. D'Amico AV, Whittington R, Malkowicz SB, et al: Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA 280:969, 1998

50. Stanford JL, Feng Z, Hamilton AS, et al: Urinary and sexual function after radical prostatectomy for clinically localized prostate cancer: the Prostate Cancer Outcomes study. JAMA 283:354, 2000

51. Talcott JA, Rieker P, Clark JA, et al: Patient-reported symptoms after primary therapy for early prostate cancer: results of a prospective cohort study. J Clin Oncol 16:275, 1998

52. Catalona WJ, Basier JW: Return of erections and urinary continence following nerve sparing radical retropubic prostatectomy. J Urol 150:905, 1993

53. Walsh PC, Marschke P, Ricker D, et al: Patient-reported urinary continence and sexual function after anatomic radical prostatectomy. Urology 55:58, 2000

54. Begg CB, Riedel ER, Bach PB, et al: Variations in morbidity after radical prostatectomy. N Engl J Med 346:1138, 2002

55. Potosky AL, Legler J, Albertsen PC, et al: Health outcomes after prostatectomy or radiotherapy for prostate cancer: results from the Prostate Cancer Outcomes Study. J Natl Cancer Inst 92:1582, 2000

56. Madalinska JB, Essink-Bot ML, de Koning HJ, et al: Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer. J Clin Oncol 19:1619, 2001

57. Wallner K, Roy J, Harrison L: Tumor control and morbidity following transperineal iodine 125 implantation for stage T1/T2 prostatic carcinoma. J Clin Oncol 14:449, 1996

58. Soloway MS, Sharifi R, Wajsman Z, et al: Randomized prospective study comparing radical prostatectomy alone versus radical prostatectomy preceded by androgen blockade in clinical stage B2 (T2bNxM0) prostate cancer. The Lupron Depot Neoadjuvant Prostate Cancer Study Group. J Urol 154(pt 1):424, 1995

59. Vailancourt L, Ttu B, Fradet Y, et al: Effect of neoadjuvant endocrine therapy (combined androgen blockade) on normal prostate and prostatic carcinoma: a randomized study. Am J Surg Pathol 20:86, 1996

60. Witjes WP, Schulman CC, Debruyne FM: Preliminary results of a prospective randomized study comparing radical prostatectomy versus radical prostatectomy associated with neoadjuvant hormonal combination therapy in T2-3 N0 M0 prostatic carcinoma. The European Study Group on Neoadjuvant Treatment of Prostate Cancer. Urology 49(suppl):65, 1997

61. Zelefsky MJ, Cowen D, Fuks Z, et al: Long-term tolerance of high dose three-dimensional conformal radiotherapy in patients with localized prostate carcinoma. Cancer 85:2460, 1999

62. Critz FA, Tarlton RS, Holladay DA: Prostate specific antigen–monitored combination radiotherapy for patients with prostate cancer. I-125 implant followed by external-beam radiation. Cancer 75:2383, 1995

63. Shipley WU, Verhey LJ, Munzenrider JE, et al: Advanced prostate cancer: the results of a randomized comparative trial of high dose irradiation boosting with conformal protons compared with conventional dose irradiation using photons alone. Int J Radiat Oncol Biol Phys 32:3, 1995

64. Pilepich MV, Krall JM, al-Sarraf M, et al: Androgen deprivation with radiation therapy compared with radiation therapy alone for locally advanced prostatic carcinoma: a randomized comparative trial of the Radiation Therapy Oncology Group. Urology 45: 616, 1995

65. Bolla M, Gonzalez D, Warde P, et al: Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin. N Engl J Med 337:295, 1997

66. Pilepich MV, Caplan R, Byhardt RW, et al: Phase III trial of androgen suppression using goserelin in unfavorable-prognosis carcinoma of the prostate treated with definitive radiotherapy: a report of Radiation Therapy Oncology Group Protocol 85-31. J Clin Oncol 15:1013, 1997

67. D'Amico AV, Manola J, Loffredo M, et al: 6-Month androgen suppression plus radiation therapy vs radiation therapy alone for patients with clinically localized prostate cancer: a randomized controlled trial. JAMA 292:821, 2004

68. Jeffery SM, Pepe JJ, Popovich LM, et al: Gabapentin for hot flashes in prostate cancer. Ann Pharmacother 36:433, 2002

69. Immediate versus deferred treatment for advanced prostatic cancer: initial results of the Medical Research Council Trial. The Medical Research Council Prostate Cancer Working Party Investigators Group. Br J Urol 79:235, 1997

70. Messing EM, Manola J, Sarosdy M, et al: Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer. N Engl J Med 341:1781, 1999

71. Nair B, Wilt T, MacDonal R, et al: Early versus deferred androgen suppression in the treatment of advanced prostatic cancer. Cochrane Database Syst Rev (1):CD003506, 2002

72. Leuprolide versus diethylstilbestrol for metastatic prostate cancer. The Leuprolide Study Group. N Engl J Med 311:1281, 1984

73. Soloway MS, Chodak G, Vogelzang NJ, et al: Zoladex versus orchiectomy in treatment of advanced prostate cancer: a randomized trial. Urology 37:46, 1991

74. Crawford ED, Eisenberger MA, McLeod DG, et al: A controlled trial of leuprolide with or without flutamide in prostatic carcinoma. N Engl J Med 321:419, 1989

75. Eisenberger MA, Blumenstein BA, Crawford ED, et al: Bilateral orchiectomy with or without flutamide for metastatic prostate cancer. N Engl J Med 339:1036, 1998

76. Tannock I, Osoba D, Stockler M, et al: Chemotherapy with mitoxantrone plus prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized study with palliative end points. J Clin Oncol 14:1756, 1996

77. Kantoff P, Halabi S, Conaway M, et al: Hydrocortisone with or without mitoxantrone in men with hormone-refractory prostate cancer: the results of CALGB 91-82. J Clin Oncol 17:2506, 1999

78. Tannock IF, de Wit R, Berry WR, et al: Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cacner. N Engl J Med 351:1502, 2004

79. Petrylak DP, Tangen CM, Hussain MH, et al: Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. N Engl J Med 351:1513, 2004

80. Thompson IM, Goodman PJ, Tangen CM., et al: The influence of finasteride on the development of prostate cancer. N Engl J Med 349:215, 2003

81. Klein EA, Thompson IM, Lippman SM, et al: SELECT: the next prostate cancer prevention trial. Selenium and Vitamin E Cancer Prevention Trial. J Urol 166:1311, 2001

82. Cohen JH, Kristal AR, Stanford JL: Fruit and vegetable intakes and prostate cancer risk. J Natl Cancer Inst 92:61, 2000

83. Terry P, Lichtenstein P, Feychting M, et al: Fatty fish consumption and risk of prostate cancer. Lancet 357:1764, 2001

84. Moyad MA: Is obesity a risk factor for prostate cancer, and does it even matter? A hypothesis and different perspective. Urology 59(4 suppl 1):41, 2002

X GYNECOLOGIC CANCER

Stephen A. Cannistra, m.d.

Epithelial Cancer of the Ovary

The ovary contains three distinct cell types: germ cells, stromal cells, and epithelial cells. All of these cell types can give rise to tumors that vary widely in their presentation, natural history, and treatment approach. Ovarian germ cell tumors are derived from the oocyte and often occur in women younger than 20 years. The typical presentation is pelvic pain, which is caused by rapid tumor growth. Ovarian stromal cell tumors are derived from supporting elements such as granulosa or theca cells, which are normally responsible for sex steroid production. Stromal cell tumors can occur at any age, although they often affect postmenopausal women. These tumors are characterized by their ability to secrete estradiol. The type of ovarian tumor that most commonly affects adult women, however, is derived from the epithelial cells that cover the ovarian surface. Epithelial ovarian cancer can be divided into four common histologic subtypes: papillary serous cystadenocarcinoma, endometrioid, clear cell, and mucinous [*see Table 1*].

This chapter focuses on the diagnosis and management of epithelial ovarian cancer, with special emphasis on increased survival resulting from the use of taxanes as first-line therapy. Unless otherwise stated, the term ovarian cancer refers to an invasive neoplasm of epithelial origin, as opposed to germ cell and stromal cell neoplasms.

epidemiology and risk factors

Epithelial ovarian cancer occurs at a mean age of 60 years in the United States and is the most lethal of gynecologic tract tumors. In approximately 70% of affected women, the tumor has spread beyond the pelvis by the time of diagnosis.[1] An estimated 25,400 new cases of ovarian cancer and approximately 14,300 deaths from the disease occurred in the United States in 2003.[2]

Temporary suppression of menstrual function is associated with a decreased risk of epithelial ovarian cancer. Thus, oral-contraceptive use (any history), pregnancy, and lactation are each associated with a significant reduction in risk.[1] Conversely, nulliparity is associated with an increased risk of ovarian cancer; it is not uncommon to diagnose this tumor during evaluation for and treatment of infertility. The use of talc has been associated with increased ovarian cancer risk, although this finding is somewhat controversial.[3,4] The association between the use of such fertility drugs as clomiphene citrate and increased risk of ovarian cancer is also controversial, with some studies suggesting a higher risk and others showing no increased risk.[5-7] Interpretation of these studies is confounded by the fact that nulliparity (and therefore perhaps infertility) itself is a risk factor for ovarian cancer.

A large population study confirmed nulliparous women are at a slightly increased risk for ovarian cancer; however, the risk may be higher in nulliparous women who have a family history of breast cancer or ovarian cancer in first-degree relatives [8] The most important risk factor for ovarian cancer is family history, with two or more affected first-degree relatives conferring an approximately 10-fold increased risk.[1] Two familial syndromes that are associated with an increased risk of ovarian cancer are the breast-ovarian cancer syndrome (involving mutations in either the *BRCA1* gene, located on chromosome 17q, or the *BRCA2* gene, located on chromosome 13q) and the Lynch syndrome II (usually involving mutations in the *MSHII* or *MLHI* gene).

Familial Ovarian Cancer

Breast-ovarian cancer syndrome A family history suggestive of a genetic predisposition for ovarian cancer may be found in as many as 10% of patients affected with the disease. The breast-ovarian cancer syndrome, as the name implies, occurs in families whose members may have breast cancer, ovarian cancer, or both.[9,10] Mutations in the *BRCA1* or *BRCA2* genes are responsible for this syndrome. BRCA1 and BRCA2 are nuclear proteins that associate with RAD51, a molecule responsible for maintaining genomic integrity.[11] Loss of heterozygosity for the *BRCA1* or *BRCA2* locus (meaning that only one of the two alleles remains, usually representing the mutated gene) is a common feature of tumors occurring in the setting of an inherited germline mutation. Whether loss of *BRCA1* or *BRCA2* function increases cancer risk by predisposing to genomic instability is currently being investigated.

Because the mutated *BRCA1* and *BRCA2* genes can be inherited from either parent, it is imperative to obtain a complete history from both sides of the family. In addition, there are subtle but important differences between the phenotype of *BRCA1* mutation carriers and the phenotype of *BRCA2* mutation carriers. *BRCA1* mutations are often associated with early onset breast cancers and an overall lifetime breast cancer risk of 50% to 80%. There may also be a slightly increased risk of colorectal and prostate cancer in *BRCA1* mutation carriers. The lifetime

Table 1 Common Histologic Types of Epithelial Ovarian Cancer

Histology	*Clinical Features*
Papillary serous cystadenocarcinoma	The most common type of epithelial ovarian cancer; may also be observed in primary peritoneal serous cancers; associated with psammoma bodies
Endometrioid	Histologically resembles endometrial cancer and is sometimes associated with endometriosis; tends to occur in slightly younger women and at an earlier stage than papillary serous carcinoma; may be associated with a separate endometrial (uterine) cancer in 15% of patients
Clear cell	Distinguished by cleared-out cytoplasm (due to glycogen) and hobnail nuclei; also may be associated with endometriosis; humorally mediated hypercalcemia may occur in rare instances; the most chemoresistant of all epithelial ovarian cancer histologies
Mucinous	The only histology not convincingly associated with *BRCA1* mutations; often associated with normal or only slightly elevated CA125 level; the presence of *Pseudomyxoma peritonei* and bilaterality should suggest an appendiceal primary tumor with metastases to the ovaries

ovarian cancer risk for *BRCA1* mutation carriers is 16% to 40%.[9,10] Some investigators have suggested that patients with ovarian cancer caused by the *BRCA1* mutation have a better prognosis than patients with the sporadic variety of ovarian cancer,[12,13] but this finding is not supported by other studies.[14] In addition to developing ovarian cancer, patients with the *BRCA1* mutation may develop primary peritoneal serous cancers, which are histologically identical to ovarian cancer and are presumed to represent malignant transformation of the peritoneal mesothelial surface.[15]

BRCA2 mutations are not as likely as *BRCA1* mutations to be associated with the early onset of breast cancer, although the lifetime risks for breast cancer are similar for the two mutations.[16,17] In addition, *BRCA2* mutations are associated with a higher risk of male breast cancer and possibly with a higher risk of pancreatic cancer. The ovarian cancer risk associated with *BRCA2* mutations is about 10%, which is lower than the risk associated with *BRCA1* mutations.[10]

Specific *BRCA1* and *BRCA2* mutations have been noted to occur with a higher than expected frequency in certain ethnic groups. It is thought that this phenomenon represents a founder effect, in which, over many generations of inbreeding, an ancestral gene has been selected. The best-studied group are Ashkenazi Jewish women of eastern European descent, in whom three distinct mutations occur with a relatively high frequency: the *BRCA1* 185 del AG mutation, the *BRCA1* 5382 ins C mutation, and the *BRCA2* 6174 del T mutation.[9] In healthy Ashkenazi Jewish women, the prevalence of these mutations is approximately 2.5%, and they are found in about 30% to 40% of early onset breast and ovarian cancers. It is important to note that these and other *BRCA1* and *BRCA2* mutations are not specific for Ashkenazi Jewish women and may occur in persons of other ethnic groups.

The value of screening patients who harbor germline mutations in *BRCA1* or *BRCA2* genes is controversial, although the strategy of screening with transvaginal ultrasonography and measurement of the CA125 level is under investigation.[18] Prophylactic oophorectomy after the completion of childbearing in high-risk women older than 35 years is a reasonable consideration, but the value of this procedure in prolonging survival is unproved.[19] Patients who consider prophylactic oophorectomy should be informed that it is still possible to develop primary peritoneal serous cancer at other sites in the abdominal cavity.[15,20]

Lynch syndrome II Lynch syndrome II, also known as hereditary nonpolyposis colon cancer (HNPCC), is characterized by a familial predisposition for ovarian, endometrial, and colon cancers. Affected women are often younger than 50 years.[21] Although ovarian cancer is a component of this entity, the HNPCC syndrome more frequently occurs simply as a cancer in the right side of the colon.[22] In this regard, a patient with ovarian cancer in the setting of HNPCC will probably describe a significant family history of colon cancer, without recalling anyone else in her family having had ovarian cancer. This genetic predisposition is the result of germline mutations in genes that encode proteins responsible for mismatch repair of DNA, including *MSH2* (on chromosome 2p), *MLH1* (on chromosome 3), *PMS1* (on chromosome 2), and *PMS2* (on chromosome 7).[21]

CLINICAL FEATURES

Early onset ovarian cancer that is restricted to the pelvis usually produces no signs or symptoms. Occasionally, a premenopausal woman with cancer limited to the ovary (stage I) [*see Table 2*] may present with a pelvic mass on routine physical examination or will come to medical attention because of intermittent pelvic pain caused by ovarian torsion. A postmenopausal woman with early onset ovarian cancer may have a palpable ovary, which is not a normal finding for a woman in this age group and raises suspicion of malignancy. Unfortunately, approximately 70% of women present with advanced disease that has extended beyond the pelvis to involve other areas such as the upper abdomen (stage III) and the pleural space (stage IV) [*see Table 2*]. In these cases, the tumor cells have shed from the primary ovarian mass and have spread throughout the peritoneal cavity, resulting in diffuse tumor nodules that involve the omentum, undersides of the diaphragm, and serosal surfaces of the abdominal cavity and bowel [*see Figure 1*]. Women with advanced disease often note a progressive increase in abdominal girth and bloating for several months before they are diagnosed. These symptoms are caused by malignant ascites. The tumor can spread via the lymphatics to involve the para-aortic lymph node chain, which is the primary drainage site for the ovaries (as with the testes). In rare instances, patients may present with inguinal adenopathy as the first sign of disease. A large omental tumor cake can cause early satiety and weight loss as a result of gastric compression; however, weight loss is more commonly offset by the development of ascites. Because of the nonspecific nature of the abdominal complaints, which are related to the presence of ascites and omental disease, many patients initially undergo an upper gastrointestinal tract evaluation for a possible ulcer before the true nature of the illness is recognized. Finally, ovarian

Table 2 FIGO Surgical Staging of Ovarian Cancer*

Stage I:	Tumor confined to the ovaries
Stage IA:	Limited to one ovary, without capsular spread, tumor rupture, positive peritoneal washings, or malignant ascites
Stage IB:	Bilateral ovarian involvement, without capsular spread, tumor rupture, positive washings, or malignant ascites
Stage IC:	Capsular spread, tumor rupture, positive washings, or malignant ascites
Stage II:	Tumor extends into the pelvis
Stage IIA:	Involvement of the uterus or fallopian tubes
Stage IIB:	Involvement of other pelvic organs (bladder, cul-de-sac implants, vagina, rectum)
Stage IIC:	Pelvic extension, plus findings as indicated for stage IC
Stage III:	Tumor extends to the upper abdomen or involves retroperitoneal lymph nodes
Stage IIIA:	Microscopic seeding outside of the true pelvis (e.g., serosa of the small bowel, undersides of diaphragm, or omental involvement)
Stage IIIB:	Gross implants about 2 cm before debulking
Stage IIIC:	Gross implants > 2 cm before debulking, or the presence of retroperitoneal lymph node involvement (usually para-aortic)
Stage IV:	Distant organ involvement, including pleural space (most common) or hepatic-splenic parenchyma (unusual)

*The staging system for ovarian cancer is established by the International Federation of Gynecology and Obstetrics (FIGO) and is based on the results obtained at the time of surgery.

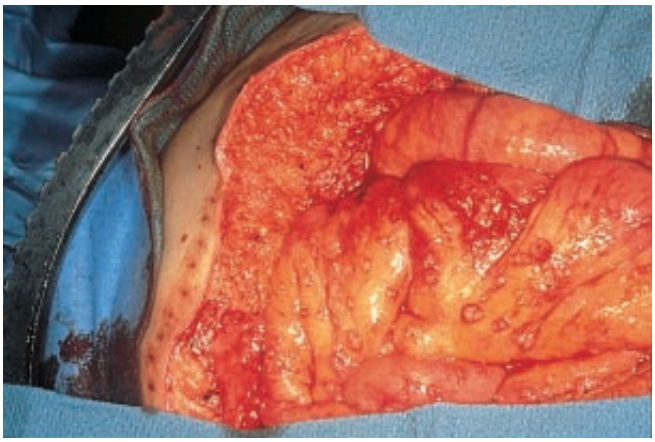


Figure 1 **Intraoperative appearance of stage III epithelial ovarian cancer. Multiple peritoneal implants are widely scattered throughout the upper abdomen in a patient with stage IIIC disease.**

cancer can sometimes be preceded by a paraneoplastic syndrome, such as cerebellar degeneration (which is often not reversible despite treatment of the tumor), migratory superficial thrombophlebitis and hypercoagulability (Trousseau syndrome), or the sign of Leser-Trélat (sudden appearance of multiple seborrheic keratoses).[1]

EVALUATION AND DIFFERENTIAL DIAGNOSIS

Women suspected of having an intra-abdominal malignancy such as ovarian cancer on the basis of findings of a pelvic mass, ascites, or both are often evaluated by transvaginal ultrasonography. A transvaginal approach is preferred for the performance of sonography because the ultrasound probe is closer to the ovaries, and the patient is not required to have a full bladder for adequate definition of pelvic anatomy.

Sonography is more sensitive than computed tomography for the detection of pelvic masses and is capable of yielding qualitative information about a mass that can help determine whether it is malignant. An ovarian mass composed of both solid and cystic components, especially if it is septated (multicystic) and echogenic (implying free-floating tumor debris), is referred to as complex and suggests a malignancy. Conversely, a so-called simple cyst is entirely fluid filled, is nonechogenic, and is not associated with a solid component. Simple cysts are usually benign.

If a complex mass is observed on a sonogram, the next step in the diagnostic evaluation is usually an exploratory laparotomy, which is required for tissue diagnosis, staging, and debulking. The performance of CT or paracentesis for preoperative confirmation of malignancy is not usually helpful once a complex mass has been discovered in a patient who is a good surgical candidate. Such a patient generally requires an operation for tissue diagnosis and tumor resection, regardless of the results of these additional studies. A tissue diagnosis (as opposed to cytologic evaluation) is often required for the distinction between primary ovarian cancer and tumors from other primary sites that have metastasized to the ovary. Metastatic involvement of the ovary is known as a Krukenberg tumor; it is often (but not always) bilateral and typically represents spread from an unsuspected gastric, colorectal, or breast cancer. Because early-stage ovarian cancer is potentially curable, it is especially important to proceed to surgical exploration in patients whose sonograms show only a pelvic mass and whose examination reveals no evidence of advanced disease. Such pelvic masses should generally not be diagnosed by percutaneous biopsy, because tumor dissemination may occur if the cyst is inadvertently ruptured.

There are occasional instances in which surgery is not required in the evaluation and management of a suspected intra-abdominal malignancy such as ovarian cancer. For instance, an elderly patient with a cytologically confirmed malignant pleural effusion containing a papillary carcinoma with psammoma bodies (concentric rings of calcification observed on microscopic evaluation) and obvious intra-abdominal disease on examination or sonogram probably has stage IV ovarian cancer (or primary peritoneal serous cancer, which is treated similarly). Because surgery is not without risk in the elderly population and because the benefits of surgical cytoreduction in the setting of stage IV disease are controversial, it is reasonable to forgo the procedure in such a patient and proceed directly with appropriate chemotherapy (see below). However, initial surgery may be appropriate in an elderly patient if imminent bowel obstruction is a concern. Also, a patient with disproportionate complaints of upper abdominal discomfort, left-sided supraclavicular adenopathy (Virchow node), blood in the stool, and iron deficiency anemia would be better served by upper endoscopy to exclude gastric cancer, which can metastasize to the ovaries and cause a Krukenberg tumor. Finally, patients with suspected ovarian cancer require a breast examination, because breast cancer can metastasize to the abdominal cavity, producing ascites and ovarian masses (a more likely presentation with lobular, as opposed to ductal, carcinoma of the breast).

Serum tumor markers such as CA125 and carcinoembryonic antigen (CEA) are generally not helpful in the differential diagnosis of a patient with presumed intra-abdominal malignancy. Although a very high CA125 value in the setting of a normal or slightly elevated CEA level is suggestive of ovarian cancer, most patients require surgery for definitive diagnosis of ovarian cancer and debulking. The CA125 level is a useful parameter for assessment of the response of ovarian cancer patients to chemotherapy and for detection of relapse.[23,24] However, measurement of the CA125 level by itself is not useful as a screening test; the CA125 level is elevated in only 50% of patients with early-stage disease and may be elevated in other cancers, as well as in nonmalignant conditions such as pregnancy and endometriosis. The strategy of both measuring the CA125 level and performing transvaginal sonography for screening has resulted in greater specificity, although its value in the early detection of ovarian cancer in the general population has not been proved and is under investigation as a screening tool in high-risk patients.[10]

STAGING AND SURGICAL MANAGEMENT

The staging system for ovarian cancer is derived from the International Federation of Gynecology and Obstetrics (FIGO) [*see Table 2*]. Early-stage disease is localized to one or both ovaries (stage I) or extends beyond the ovary to involve other pelvic structures (stage II). Advanced disease involves the upper abdomen or lymph nodes (stage III) or areas outside of the abdomen, such as a cytologically documented malignant pleural effusion (stage IV).

Exploratory laparotomy for evaluation of suspected ovarian cancer is typically performed with a vertical midline incision to provide adequate visualization of the upper abdomen. Peri-

toneal washings are obtained before removal of the ovarian mass to avoid contamination of the cytologic specimen with tumor cells inadvertently released by ovarian rupture during resection. If the suspicion of epithelial ovarian cancer is confirmed by frozen section, a bilateral salpingo-oophorectomy (BSO) and total abdominal hysterectomy (TAH) are usually performed, along with a partial omentectomy. Other sites of tumor involvement are carefully evaluated with palpation and biopsy of the undersides of the diaphragm, the serosal surfaces of the bowel, and the paracolic gutters. The para-aortic lymph nodes are typically assessed when information about lymph node status would change patient management or when precise surgical staging is required to determine eligibility for protocol therapy. Finally, an attempt is made to remove as much tumor as possible at the time of initial surgery (debulking) because patients with residual tumor measuring less than 1 cm in diameter are more likely to respond to chemotherapy and have an improved survival rate.[1] It is not known whether this is the result of an intrinsic therapeutic effect of debulking, as opposed to a selection phenomenon, whereby the ability to debulk identifies a patient population with more favorable biologic features. For patients whose tumor cannot be optimally debulked at the time of initial surgery, a repeat attempt may be considered after three cycles of chemotherapy have been administered. This approach is referred to as an interval cytoreductive procedure and was associated with an increase in the median duration of survival by approximately 6 months in a randomized trial performed by the European Organization for the Research and Treatment of Cancer (EORTC).[25]

POSTOPERATIVE TREATMENT

Early-Stage Disease (Stages I and II)

Some women with ovarian cancer have low-risk features that confer a 5-year survival rate of about 95%. Members of this low-risk group have stage IA or stage IB disease that is well or moderately well differentiated (i.e., grade I or II). These patients do not generally derive benefit from the use of postoperative adjuvant therapy.[26] In contrast, approximately 20% of women with early-stage disease have risk factors that are associated with an overall relapse rate of between 20% and 40%.[27,28] The risk factors include any or all of the following: involvement of the ovarian capsule, tumor rupture, tumor present in peritoneal washings, grade III histology, and stage II disease. Patients with early-stage disease who have any of these risk factors are commonly treated with postoperative adjuvant chemotherapy with a platinum-based regimen, because this class of agents provides effective treatment for patients with more advanced disease (see below). Adjuvant treatment of this high-risk group has been shown to improve the disease-free survival rate, although its effect on overall survival is still controversial.[27,28] More recent randomized data suggest that platinum-based adjuvant therapy may confer an overall survival advantage in high-risk, early stage disease, although the benefit appears to be restricted to those patients who were inadequately staged.[29] The overall 5-year survival rate for patients with high-risk, early-stage ovarian cancer who are treated with adjuvant chemotherapy is about 80%.

Advanced-Stage Disease (Stages III and IV)

Approximately 70% of patients with ovarian cancer have advanced disease that cannot be completely resected at the time of exploratory laparotomy. Eradication of residual disease in these patients generally requires some form of postoperative treatment, most often platinum-based chemotherapy. Postoperative therapy for advanced-stage ovarian cancer leads to long-term disease-free survival in 10% to 30% of patients, depending on the exact stage and residual disease status. For instance, a patient with stage IV, suboptimally debulked disease (residual nodules measuring more than 1 cm in diameter) has an expected long-term survival rate of less than 10% despite treatment with platinum-based chemotherapy, whereas a patient with stage III, optimally debulked disease has an approximately 30% chance of long-term survival.

The platinum analogues are the most effective agents in the treatment of epithelial ovarian cancer, with single-agent response rates in the range of 50% to 60%. Platinum compounds, such as carboplatin and cisplatin, exert their cytotoxic effects by inducing DNA damage, primarily through the formation of intrastrand cross-links. The taxanes, such as paclitaxel and docetaxel, are also highly active drugs, with single-agent response rates in newly diagnosed advanced disease of approximately 40%. Paclitaxel was originally isolated from the bark of the Pacific yew tree *Taxus brevifolia* and exerts its cytotoxic effects in part by binding to and stabilizing the tubulin polymer during mitotic spindle formation.[30] The mitotic spindle apparatus cannot readily dissociate in the presence of paclitaxel, resulting in M-phase arrest and subsequent cell death through apoptosis.[31] The distinct mechanisms of action of platinum compounds and taxanes suggest that when they are used in combination, they may act independently (i.e., there is no cross-resistance). In this regard, the activity of combined paclitaxel and cisplatin chemotherapy has been evaluated in two large randomized trials of patients with newly diagnosed ovarian cancer.[32,33] Compared with cyclophosphamide and cisplatin, paclitaxel and cisplatin demonstrated superior activity and led to statistically significant improvement in the overall survival rate. The combination of paclitaxel and carboplatin has been shown to be therapeutically equivalent to the original paclitaxel-cisplatin regimen but is associated with less emesis, neurotoxicity, and nephrotoxicity.[34,35] Paclitaxel and carboplatin are easier to administer on an outpatient basis because prehydration and posthydration are not required (cisplatin requires adequate hydration to prevent nephrotoxicity). For these reasons, the combination of paclitaxel (3-hour infusion) and carboplatin has become the standard regimen for the postoperative management of patients with advanced epithelial ovarian cancer. It is important to note that this chemotherapy is administered intravenously, although there are some investigators who believe that intraperitoneal (IP) administration of agents such as cisplatin may be beneficial.[36,37] The value of IP chemotherapy as first-line treatment of patients with optimally debulked, advanced disease is uncertain. Finally, it has been recently shown that administration of single-agent paclitaxel for 1 year is capable of prolonging median progression-free survival by 7 months in patients with advanced disease who achieve a clinical complete remission after first-line therapy.[38] Unfortunately, this strategy does not appear to prolong overall survival, and it is associated with side effects such as neuropathy and continued alopecia.

MANAGEMENT OF RELAPSE

The most common adverse prognostic factors for ovarian cancer are advanced-stage, high-grade, and suboptimally debulked disease [*see Table 3*].[1,39-45] These prognostic factors are

predictive of a poor response to chemotherapy and poor durability of complete remission. Thus, approximately 30% of patients with advanced disease have residual, active tumor after treatment with paclitaxel and carboplatin. Conversely, about 70% of newly diagnosed patients achieve a complete response to surgery followed by treatment with paclitaxel and carboplatin, although most of these women eventually experience relapse. Relapse of ovarian cancer is almost always incurable; it is often detected by the finding of an elevated CA125 level in the absence of signs, symptoms, or CT evidence of tumor.[1]

The goal of treatment of ovarian cancer relapse is palliation of symptoms. For patients with relapse detected only by the finding of an elevated CA125 level, there is no convincing evidence that second-line, cytotoxic chemotherapy improves survival rate; however, it can compromise quality of life at a time when patients are feeling well. In this situation, treatment with tamoxifen or an aromatose inhibitor is a reasonable consideration because this antiestrogen is well tolerated, is conveniently administered by mouth, and has a response rate of approximately 17% in patients experiencing relapse.[46,47] Cytotoxic chemotherapy is generally reserved for patients who progress through tamoxifen therapy or who are becoming symptomatic (often complaining of mild abdominal cramping) in the setting of CT findings such as peritoneal implants or ascites.[42,47] When chemotherapy becomes necessary, retreatment with either platinum compounds or taxanes is often a reasonable first step if the interval between the last chemotherapy regimen and relapse is longer than 6 months.[42,47] Recent data suggest that a combination of paclitaxel and a platinum compound may confer a survival advantage over single-agent platinum in patients with relapsed disease who are potentially platinum sensitive.[48] However, because of methodologic concerns, additional studies are required to confirm these findings. For cases of rapid relapse after completion of first-line therapy and cases of development of platinum-resistant disease, potentially useful second-line agents include liposomal doxorubicin, topotecan, gemcitabine, and etoposide.[47,49] For selected patients who experience a prolonged remission after first-line chemotherapy (i.e., remission lasting longer than 12 months) and who experience recurrence of localized, potentially resectable disease, a secondary cytoreductive procedure is sometimes considered before the institution of chemotherapy.[50] In patients experiencing relapse, this procedure is almost never curative, and the value of this approach has not been tested in a prospective, randomized trial. Although patients who undergo successful secondary cytoreduction tend to survive longer than those whose cancer cannot be resected, it is unclear whether the surgery itself confers the survival advantage or whether it is simply the fact that the patients who are selected to undergo cytoreduction have more indolent, chemoresponsive disease.

Table 3 Selected Adverse Prognostic Factors in Epithelial Ovarian Cancer

Advanced disease (stages III/IV)[1,39]
Suboptimal cytoreduction (> 1 cm residual)[1,39]
High grade[1,39]
Clear cell histology[1,39]
Older age (> 65 yr)[40]
Persistent CA125 elevation after three cycles of chemotherapy[41]
Rapid recurrence (within 6 mo of completing first-line therapy)[42]
p53 mutation[43]
Underexpression of the BAX protein (a death-promoting member of the BCL2 family)[44]
Amplification of the *HER-2/neu* gene (a member of the EGF tyrosine kinase receptor family)[45]

MANAGEMENT OF OTHER TYPES OF OVARIAN TUMORS

Borderline ovarian tumors, also known as tumors of low malignant potential, are epithelial tumors that show no histologic evidence of invasion and often appear at an early stage during patients' fourth or fifth decade of life. These tumors are generally associated with excellent survival rates (about 90% at 10 years).[51-55] Even advanced borderline tumors involving the upper abdomen may have a very indolent natural history, characterized by serial tumor recurrences that are managed by repeated surgical resection. In contrast to more typical invasive epithelial ovarian cancer, borderline tumors are generally not sensitive to chemotherapy, and there is no well-defined role for the use of postoperative adjuvant therapy.

Malignant ovarian germ cell tumors may be histologically classified as either dysgerminoma or nondysgerminoma. Dysgerminoma of the ovary is to women what testicular seminoma is to men; patients with dysgerminoma often present with stage I disease (although dysgerminoma may be bilateral in about 10% of cases).[56] Staging of this tumor ideally includes para-aortic lymph node dissection because dysgerminoma has a predilection for lymph node spread. After resection, dysgerminoma that is limited to the ovary (stage IA) is often managed by careful follow-up; assuming that adequate staging has been performed, postoperative adjuvant chemotherapy is not necessary. More advanced disease is usually treated with postoperative platinum-based chemotherapy containing bleomycin, etoposide, and cisplatin (BEP). Immature teratoma is a form of malignant ovarian germ cell tumor that is occasionally treated with surgery alone if the tumor is confined to the ovary and characterized as grade I or, sometimes, grade II.[57,58] Patients with stage I, grade III immature teratoma or with more advanced disease are often treated with BEP. Other types of malignant nondysgerminomas of the ovary (e.g., embryonal cell, endodermal sinus, and choriocarcinoma) generally require postoperative treatment with BEP, regardless of stage, in view of the high risk of recurrence.[59,60] Malignant ovarian germ cell tumors are often chemoresponsive and highly curable; therefore, to preserve fertility, the initial surgical management of these tumors is often a unilateral salpingo-oophorectomy (with staging) as opposed to a TAH/BSO.[61,62]

Finally, women with ovarian stromal neoplasms such as granulosa cell tumors may present with signs and symptoms related to estradiol production.[63,64] A young girl may develop precocious puberty, or a postmenopausal woman may develop breast tenderness or vaginal bleeding as a result of either endometrial proliferation or the presence of a separate uterine cancer caused by unopposed estrogen production by the granulosa cell tumor. Because of its highly vascular nature, a ruptured granulosa cell tumor may sometimes mimic an ectopic pregnancy, with hypotension and abdominal pain caused by intraperitoneal bleeding. Treatment of stromal cell ovarian tumors is primarily surgical; the role of adjuvant chemotherapy or whole abdominal radiotherapy for high-risk patients (i.e., with large primary tumors, capsular rupture, or advanced-stage disease) is undefined. Recurrence of ovarian stromal cell

tumors may occur many years after the original diagnosis, and patients therefore require long-term follow-up. Because these tumors often produce estradiol, inhibin, and müllerian inhibitory substance (MIS), these serum markers may prove to be valuable for postoperative surveillance.

Cancer of the Uterine Cervix

Invasive cervical cancer is uncommon in developed countries, partly because of the effectiveness of Papanicolaou (Pap) smear screening. Nevertheless, it is estimated that in the United States in 2003, cervical cancer affected 12,200 women and caused approximately 4,100 deaths.[2] Part of the success of Pap smear screening is due to the fact that this approach typically detects premalignant lesions, as opposed to invasive cancer. This unique feature makes it possible to eradicate precursor lesions before the development of frankly invasive cancers. In addition, the interval of time between the development of a precursor lesion and the occurrence of invasive disease may be several years, thus allowing many opportunities for the detection and eradication of premalignant disease. In developing countries, however, cervical cancer is a major cause of death in patients with gynecologic cancer; Pap smear screening on a routine basis is difficult because of a lack of resources and poor patient compliance.

EPIDEMIOLOGY AND RISK FACTORS

Cervical cancer is typically a disease of women in their fifth and sixth decades, whereas premalignant cervical lesions (see below) are often discovered in women younger than 40 years.[65] This rather large gap in the age distribution between precursor lesions and invasive cancer is indicative of a long latency period for malignant transformation. Infection with human papillomavirus (HPV)—most commonly, subtypes 16, 18, 31, 33, and 35—is largely responsible for the development of precursor lesions and subsequent transformation to invasive disease.[65-67] Not surprisingly, factors that predispose to transmission of this virus are associated with an increased risk of the development of cervical cancer. These high-risk factors are sexual intercourse at an early age, multiple male sexual partners, and male sexual partners who themselves have multiple partners.[65] A history of smoking also confers a higher risk. Immunosuppression associated with either an underlying lymphoproliferative disorder such as Hodgkin disease or immunosuppressive drugs used in the prevention of allograft rejection also confers a higher risk of cervical cancer.[68] In women with HIV infection, the immunosuppressive state associated with the infection increases the risk of development of cervical precursor lesions, although it is not clear whether the development of such lesions results in a higher incidence of invasive cervical disease.[69]

EARLY DETECTION OF PREINVASIVE DISEASE

The Pap smear detects cytologic changes that indicate the possible presence of precursor lesions that could give rise to invasive disease if not removed. Thus, the main role of this screening test is to identify patients who require further evaluation by a gynecologist and possible cervical biopsy. If a subsequent cervical biopsy reveals preinvasive disease, it may be appropriate to eradicate the lesion with conservative outpatient techniques, such as laser vaporization or the loop electric excision procedure (LEEP), or with more aggressive surgical techniques, such as cervical conization (also known as cone biopsy, in which a conical portion of cervix is removed along with a portion of the endocervical canal). Pap smear screening should begin when a female becomes sexually active or reaches 18 years of age and should continue annually for at least 3 years. If three or more annual Pap smear evaluations are negative and the patient's risk of cervical cancer is average, further testing can be performed at the physician's discretion, perhaps every 2 to 3 years, although the optimal frequency in this situation has not been established.[65] Continued annual Pap smears are reasonable for women at high risk of developing cervical cancer.

DIAGNOSIS

Cervical cancer is often asymptomatic and is commonly detected as a gross cervical lesion at the time of a routine pelvic examination. The presence of a grossly visible cervical lesion warrants biopsy, even if the Pap smear is normal. Occasionally, microscopic invasive disease is found during the evaluation of an abnormal Pap smear. It should be remembered, however, that abnormal Pap smears do not usually signify the presence of invasive disease but rather indicate the possibility of precursor lesions that could give rise to invasive disease if not removed. Some patients with cervical cancer present with vaginal discharge or postcoital bleeding. Patients with more advanced disease may present with lower extremity edema, caused by involvement of pelvic lymph nodes; such edema can be painful and can cause ureteral obstruction.

Biopsy of a grossly visible lesion or discovery of invasive cancer as part of the evaluation of an abnormal Pap smear may reveal squamous cell carcinoma, the most common histologic variant of cervical cancer. Adenocarcinoma is the second most common histologic finding; it is clinically similar to squamous cell carcinoma, with the exception of its propensity to be located within the endocervical canal and to sometimes display an endophytic growth pattern, which results in a barrel-shaped cervix. In rare instances, small cell carcinoma of the cervix is discovered on biopsy, usually in a young woman; small cell carcinoma of the cervix may be associated with rapid metastases and the syndrome of ectopic adrenocorticotropic hormone secretion.

STAGING

The FIGO staging system for cervical cancer is based entirely on clinical criteria. Stage I disease is limited to the cervix and may be either microscopic (stage IA) or grossly visible (stage IB). In stage II, the cancer has extended beyond the cervix, to involve either the upper two thirds of the vagina (stage IIA) or the parametrial tissue (stage IIB). Stage III disease is more locally extensive, involving the lower one third of the vagina (stage IIIA) or extending to the pelvic side wall or causing hydronephrosis (stage IIIB). Stage IV disease is the most advanced, representing either extensive infiltration of local pelvic structures (involvement of the bladder or rectal mucosa—stage IVA) or disease outside of the pelvis (stage IVB).

Accurate staging is necessary to help determine whether the disease is amenable to a surgical approach (appropriate for stages IA, IB, and IIA) or primary radiotherapy (appropriate for stages IIB, III, and IVA). To determine whether disease involves the parametrial tissue, it is often necessary to perform a pelvic examination with the patient under anesthesia. In addition, radiographic visualization of the ureters is often required to exclude hydronephrosis and to define the ureteral anatomy if a surgical approach is deemed appropriate. For more locally

advanced lesions, consideration of cystoscopy and proctoscopy is reasonable to exclude stage IVA disease. Finally, a chest x-ray is obtained to exclude hematogenous spread to the lungs, which is a common site of metastasis in this disease.

TREATMENT

Stage IA

Patients with stage IA squamous cell carcinoma that has extended to a depth of greater than 3 mm (stage IA2) or who have involvement of the lymphovascular space or surgical margins, as revealed by cone biopsy, are appropriate candidates for a radical hysterectomy with pelvic lymph node dissection. Radical hysterectomy is a procedure in which the uterine corpus and cervix are removed en bloc along with parametrial tissue, the ureterosacral ligaments, and a 2 to 3 cm cuff of vagina. This procedure is designed to remove lymphatic channels, which may harbor microscopic tumor cells. Pelvic lymph node dissection is typically performed as part of the radical hysterectomy procedure because the finding of metastatic disease within this nodal chain will alter postoperative treatment [*see* Special Management Considerations, *below*]. The ovaries are typically not removed; they are seldom the site of cervical cancer spread, although they may be transposed to the outside of the pelvis in the event that postoperative radiotherapy is required (see below). In contrast, a simple hysterectomy (removal of the uterus and cervix, leaving intact the other associated structures mentioned above) is a reasonable procedure for patients with stage IA cervical cancer with less than 3 mm of invasion and an absence of lymphatic channel or margin involvement on cone biopsy (stage IA1). Alternatively, selected patients with stage IA1 disease are sometimes treated with cervical conization only, especially if future childbearing is desired. The survival rate of patients with stage IA cervical cancer is about 95%.

Stages IB and IIA

Local disease that is technically resectable with tumor-free margins, such as stage IB or IIA disease [*see Figure 2*], is often treated with a radical hysterectomy and pelvic lymph node dissection. For patients with comorbid disease that precludes surgical resection, primary radiotherapy is an option that produces equivalent survival rates for patients with stage IB or IIA disease, although possible toxicities are vaginal stenosis, bladder and bowel enteritis, and cessation of ovarian function. The overall survival rate of patients with cervical cancer of stage IB or IIA is in the range of 80% to 90%.[65]

Stages IIB to IVA

Once the tumor involves the parametrial tissues and beyond, a surgical approach is not technically feasible, because the chance of obtaining tumor-free resection margins is very low. In the past, patients with this extent of disease progression were treated with primary radiotherapy consisting of external-beam pelvic radiation (to encompass the pelvic lymph nodes) followed by intracavitary treatment, which provides high local doses of radiation to control areas of tumor bulk. However, recent data suggest that these patients are best served by treatment with concomitant chemotherapy and radiation (see below). Intracavitary radiotherapy is typically provided by either cesium-137 or radium-226, which is temporarily inserted into the uterine cavity and vaginal fornices. A common reference point used in the assessment of total radiation dose is known as point A, which is located 2 cm lateral and 2 cm superior to the cervical os. An optimal target dose to point A is approximately 85 Gy delivered over 8 weeks (total external beam plus intracavity dose).

The overall survival rate of patients with disease at stages IIB to IVA treated with primary radiotherapy is about 30%; patients with stage IVA disease fare less well (about 10% survival), and patients with stage IIB disease have survival rates of about 70%.[65] Many patients with tumors in stages IIB to IVA experience local failure, suggesting that radiation by itself is often incapable of sterilizing the original site of bulk disease. In this regard, three large randomized studies have recently been performed to investigate the value of concomitant platinum-based chemotherapy and radiation in patients with cervical cancer at stages IIB to IVA.[70-72] The theoretical basis for this concept stems from in vitro observations that suggest that cisplatin can function as a radiosensitizer, presumably by virtue of inhibiting DNA repair. All three studies were performed in patients with disease at stages IIB to IVA without para-aortic lymph node involvement. The studies differed with respect to the type of chemotherapy used and inclusion of other high-risk patient subsets. Nevertheless, these studies convincingly demonstrated that platinum-based chemotherapy administered concomitantly with radiation can reduce the chance of relapse by 30% to 50% and can improve the survival rate of patients with locally advanced cervical cancer. In addition, the toxicity profile of platinum-based chemotherapy is acceptable, and treatment delays are infrequent. Although the optimal chemotherapy regimen has not been completely defined, it is reasonable to conclude that radiosensitization with single-agent cisplatin is a new standard with which other chemotherapy regimens will be compared in the treatment of patients with cervical cancer at stages IIB to IVA.

SPECIAL MANAGEMENT CONSIDERATIONS

Early-Stage Disease with High-Risk Features

Patients with disease at stages IA2, IB, or IIA are reasonable candidates for radical hysterectomy and pelvic lymph node dissection. However, the unanticipated discovery of disease in pelvic lymph nodes, parametria, or surgical resection margins places such patients at a higher risk of local relapse. In the past, these high-risk features indicated the need for postoperative external-beam radiotherapy to a pelvic field, which yields overall survival rates in the range of 40% to 60%. However, a recent randomized intergroup study conducted by the Southwest Oncology Group (SWOG) demonstrated that treatment with concomitant platinum-based chemotherapy and radiation produced a significant survival advantage.[73]

Bulky Stage IB Disease

Bulky stage IB tumors that are at least 4 cm or greater (stage IB2) have a high probability of local relapse if managed with radical hysterectomy. Patients with such tumors are reasonable candidates for primary radiotherapy and are treated similarly to those with more advanced local disease; the expected survival rate is about 70%. In an approach similar to that demonstrated in patients with disease at stages IIB to IVA, platinum-induced radiosensitization has recently been tested in a large randomized study involving 369 patients with stage IB2 disease.[74] Patients in this trial were randomized to receive pelvic

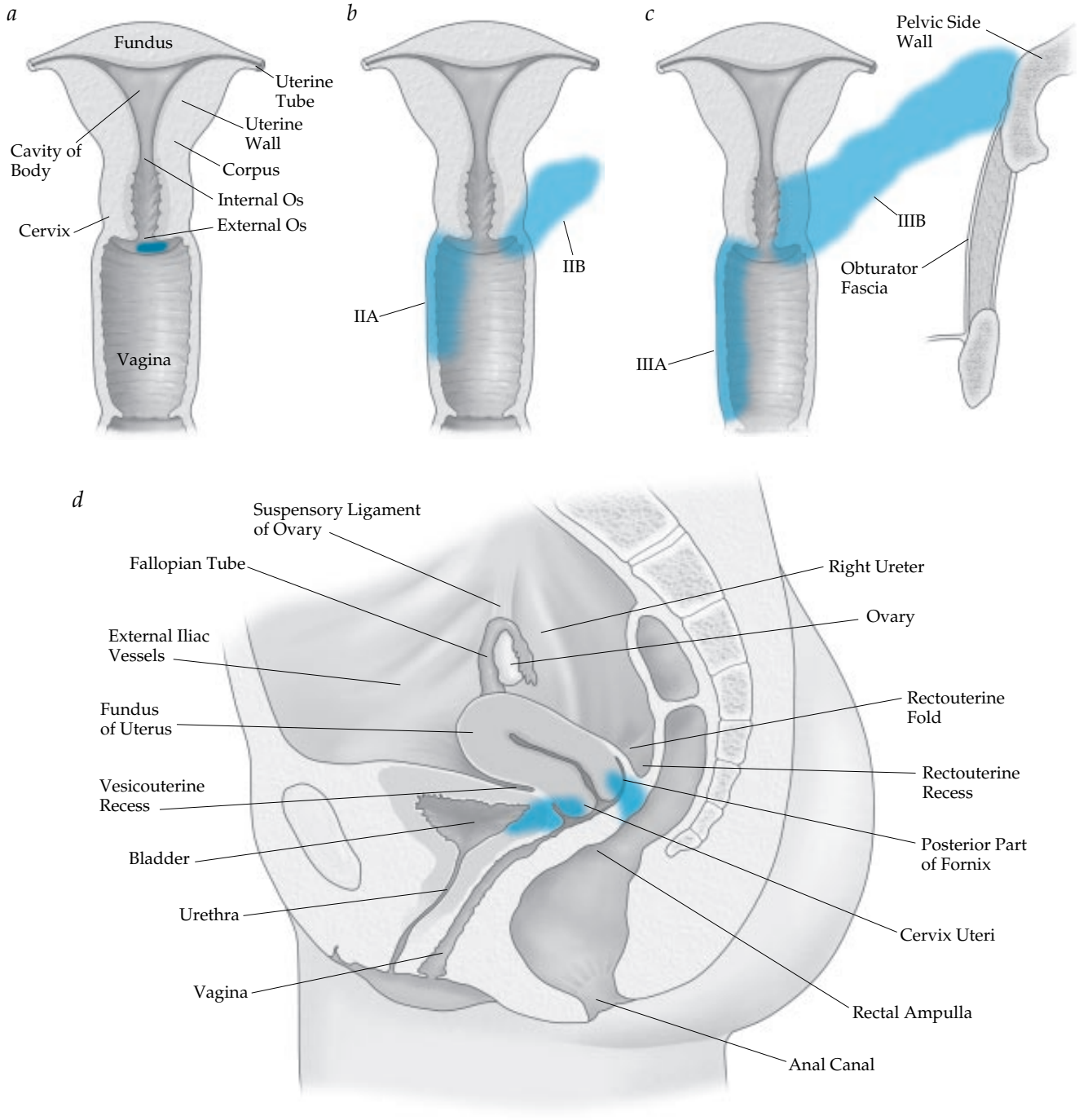


Figure 2 **International Federation of Gynecology and Obstetrics (FIGO) clinical staging of cervical cancer. (*a*) The clinical staging system for cervical cancer is defined by FIGO. Staging may include information obtained from an examination under anesthesia, an intravenous pyelogram, cystoscopy, and proctoscopy. The staging assignment for cervical cancer is determined on the basis of clinical assessment and does not change on the basis of information obtained at the time of surgery. Stage 1 disease is limited to the cervix. Stage IA disease is a microscopic lesion not exceeding 5 mm in depth from the basement membrane and no wider than 7 mm in lateral extent. Stage I is divided into stage IA1 (invasion ≤ 3 mm, no wider than 7 mm) and stage IA2 (invasion > 3 mm but ≤ 5 mm, and no wider than 7 mm). The Society of Gynecologic Oncologists (SGO) definition of microinvasion is a lesion with invasion ≤ 3 mm beneath the basement membrane, without evidence of lymphovascular space involvement. Stage IB lesions are often grossly visible tumors limited to the cervix, although they may also include microscopic lesions that are more extensive than stage IA. Stage IB1 are lesions ≤ 4 cm in diameter, and stage IB2 are lesions > 4 cm in diameter. (*b*) Stage II disease extends to the upper vagina or parametrium. Stage IIA tumors involve the upper two thirds of the vagina and are often amenable to surgical resection. In contrast, stage IIB lesions (i.e., lateral extension into the parametrium; the pelvis side wall is not involved), as well as lesions of stage III and IVA, are not technically resectable and require primary treatment with concomitant radiation and chemotherapy. (*c*) Stage IIIA disease involves the lower one third of the vagina, and stage IIIB involves lateral extension to the pelvic side wall or hydronephrosis. (*d*) Stage IVA disease is characterized by extensive locoregional involvement and has infiltrated the mucosa of the bladder or rectum. Stage IVB disease involves distant sites and is primarily treated with systemic chemotherapy, although pelvic radiotherapy may also be needed for local control.**

radiation with or without weekly cisplatin. All patients in both treatment groups subsequently underwent adjuvant hysterectomy, although it has been recently recognized that performance of surgery after primary radiotherapy confers no added survival advantage in this patient group. Weekly treatment with platinum reduced the risk of relapse and death by approximately 50% and conferred a statistically significant survival advantage (83% of patients who received cisplatin plus radiation were alive, compared with 74% of patients treated with radiation alone; $P = 0.008$). On the basis of this study, it is reasonable to consider concurrent chemoradiation with a platinum-based regimen for the primary management of patients with bulky stage I disease.

MANAGEMENT OF RELAPSE

Patients who are experiencing the relapse of cervical cancer may present with pelvic pain or bleeding caused by a pelvic mass, back pain caused by para-aortic adenopathy or bony metastasis, or shortness of breath and cough caused by pulmonary parenchymal disease. In rare instances, resection of the tumor along with the bladder and rectum (pelvic exenteration) may be considered if the recurrence is localized to the pelvis and does not involve the pelvic side walls or lymph nodes. Unfortunately, the chance is small that this procedure will be curative, and the morbidity is significant. A number of chemotherapy agents show modest results with this tumor; these agents include cisplatin, paclitaxel, ifosfamide, and vinorelbine.[65,75,76] Response rates for recurrent disease treated with single-agent cisplatin are generally about 20% and are typically observed for metastatic lesions located in sites outside of the pelvis. However, a recent study suggests that the response rate may be as high as 36% if cisplatin is combined with paclitaxel, although these data are preliminary.[77] Finally, it is important to note that patients with end-stage cervical cancer often develop renal failure as a result of bilateral hydronephrosis. Given the small chance of obtaining a meaningful response to chemotherapy, the quality of life of such patients may be best served by providing careful attention to fluid balance and pain control rather than by exposing the patient to chemotherapy-induced morbidity.

Cancer of the Endometrium (Uterine Cancer)

Cancer of the endometrium is a malignancy that begins in the inner, glandular lining (endometrium) of the uterus. Because it is usually associated with postmenopausal bleeding, the tumor is typically detected at an early stage, when it is highly curable. Although the uterus may be involved with other tumor types, such as primary uterine sarcomas, over 97% of uterine cancers are derived from the endometrium. This section is restricted to a discussion of the endometrioid histology of uterine cancer, which is the most common type of adenocarcinoma to affect the uterus.

EPIDEMIOLOGY AND RISK FACTORS

Endometrial cancer is the most frequent tumor of the gynecologic tract; it is estimated that in 2003 it affected 40,100 women and caused 6,800 deaths in the United States.[2]

Endometrial proliferation is estrogen dependent. Any factor associated with increased exposure to estrogen over time, especially if unopposed by the antiproliferative effects of progesterone, leads to an increased risk of endometrial cancer.[78] Obesity is one of the most common risk factors, an effect related to the peripheral conversion of androstenedione to estrone by aromatase present in fat cells. Early menarche, late menopause, chronic anovulation (e.g., polycystic ovary disease), and nulliparity are each associated with an increased risk as a result of either prolonged duration of estrogen exposure or absence of a progesterone effect (as is the case with chronic anovulation or nulliparity).[79] Malignancies such as granulose cell tumors of the ovary can secrete estradiol and are consequently associated with a synchronous uterine cancer in some women.[63] A synchronous uterine cancer (usually at an early stage and well differentiated) is sometimes found in association with a limited-stage endometrioid cancer of the ovary.[1] The two tumors most likely arise independently (i.e., they are clonally distinct) and tend to have a good prognosis.[80] Estrogen replacement therapy with unopposed estrogen (i.e., in the absence of progestin) is associated with a higher risk of uterine cancer development (usually well-differentiated tumors) and is not prescribed, unless the uterus has been surgically removed. Current replacement therapy entails daily administration of an estrogen and a progestin and is associated with a reduction in risk of uterine cancer development compared with matched control subjects.[79]

The use of tamoxifen, a drug traditionally thought of as an estrogen antagonist, is also associated with an increased uterine cancer risk.[81-83] This is in part caused by the tissue-specific action of tamoxifen, which has antagonistic effects on proliferation of breast epithelium but agonistic effects on bone mineral density, lipid metabolism, and endometrial proliferation. In a randomized trial conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP), tamoxifen (20 mg daily) was compared with placebo as adjuvant breast cancer therapy. The relative risk of endometrial cancer development after 5 years of tamoxifen use was 2.2; this is equivalent to an annual hazard rate of approximately two cases per 1,000 women.[84] It is important to note that most women who develop uterine cancer during tamoxifen therapy present with vaginal bleeding, have early-stage disease, and usually (but not always) have well-differentiated tumors. Furthermore, the benefits of tamoxifen in the adjuvant breast cancer setting far outweigh the small risk of uterine cancer development. Serial endometrial biopsy and ultrasonography in this patient population are associated with a high rate of false positive results; thus, their value as screening tools is controversial.[83] Currently, a reasonable approach for these patients is a routine pelvic examination and prompt evaluation of abnormal vaginal bleeding by endometrial biopsy. Finally, uterine cancer can be observed as part of the Lynch syndrome II.

CLINICAL FEATURES AND DIAGNOSTIC EVALUATION

All patients with postmenopausal vaginal bleeding require evaluation to exclude endometrial cancer, although the chance that endometrial cancer is the cause of the bleeding is age dependent. Postmenopausal bleeding in a 50-year-old woman is associated with a 9% chance of endometrial cancer, whereas the likelihood increases to 28% by age 70 and 60% by age 80.[79] A pelvic examination to exclude another cause of bleeding, such as cervical cancer, and an outpatient endometrial biopsy are common next steps in management. If uterine tissue is obtained through outpatient endometrial sampling, the chance of diagnosing endometrial cancer is excellent. If endometrial tissue is not obtained, a dilatation and curettage under general anesthesia is often required to obtain a sample suitable for fur-

Table 4 FIGO Surgical Staging of Endometrial Cancer*

Stage I: Tumor confined to the uterine corpus
- Stage IA: Limited to the endometrium (no myometrial invasion)
- Stage IB: Invading < 50% of myometrial thickness
- Stage IC: Invading > 50% of myometrial thickness

Stage II: Tumor extends into the uterine cervix
- Stage IIA: Endocervical gland involvement
- Stage IIB: Cervical stroma involvement

Stage III: More extensive locoregional disease
- Stage IIIA: Deep extension to uterine serosa or adnexal involvement, or both, or positive cytology
- Stage IIIB: Local extension into the vagina
- Stage IIIC: Retroperitoneal node involvement (pelvic or para-aortic)

Stage IV: Extensive locoregional disease or distant spread
- Stage IVA: Involvement of bladder or rectal mucosa
- Stage IVB: Distant disease, including involvement of the upper abdomen or hematogenous spread to organs outside of the abdominal cavity

*The staging system for endometrial cancer is established by the International Federation of Gynecology and Obstetrics (FIGO) and is based on the results obtained at the time of surgery. Although not specifically listed, assignment of grade is considered to be a formal part of stage designation for endometrial cancer. For example, the stage of a patient with a well-differentiated, superficial tumor would be stage IA, grade I. The stage of a patient with a high-grade, deep lesion extending out to the uterine serosa would be stage IIIA, grade III.

ther evaluation. In the setting of postmenopausal bleeding, transvaginal ultrasonography is sometimes performed to determine whether the endometrium is thickened and, therefore, possibly involved with cancer. A normal transvaginal ultrasound under these circumstances does not guarantee the absence of cancer, however, and a suggestive study will not change management in women who would require biopsy in any event.

STAGING AND SURGICAL MANAGEMENT

If the endometrial biopsy establishes the presence of cancer, surgery for definitive resection and staging is the next most common step in management. For patients who are not optimal surgical candidates, primary radiotherapy is an option that can produce long-term survival in selected patients. Surgery for uterine cancer entails performance of peritoneal washings, TAH/BSO, evaluation of pelvic and para-aortic lymph nodes (palpation and biopsy if indicated; see below), and omental biopsy. An assessment of the depth of uterine muscle (myometrial) invasion is made intraoperatively, along with frozen-section assessment of grade, as these two factors have independent predictive value for the presence of lymph node involvement.[85] Specifically, a patient with a superficial, noninvasive, grade I endometrioid uterine cancer has a negligible chance of having involvement of either pelvic or para-aortic nodes; in such patients, it may be acceptable to omit a formal biopsy of these lymph node chains. Conversely, the risk of nodal involvement increases to at least 20% in patients with high-grade lesions that are deeply invasive (greater than two thirds of the myometrial thickness). Under these circumstances, it is reasonable to consider obtaining a biopsy sample of at least the para-aortic lymph node chain so as to detect occult involvement that would require modifying the field of postoperative radiotherapy (see below).[86]

The FIGO staging system for uterine cancer is determined by the findings at surgery and histologic review [*see Table 4*]. Stage I disease is limited to the uterine corpus; stage II disease involves the uterus and cervix; stage III disease often involves

Table 5 Postoperative Management Considerations for Patients with Uterine Cancer (Endometrioid Histology)

*Stage**	*Possible Treatment Options*	*Comments*
IA, grade I or II	No further therapy	Long-term survival in over 90% of patients
IA, grade III	No further therapy	Risk of pelvic nodal involvement is low; benefit of pelvic XRT is controversial
IB, grade I or II	No further therapy, or vaginal apex brachytherapy	Benefit of vaginal apex brachytherapy for stage IB, grade I or II lesions is uncertain
IB, grade III	Pelvic XRT	Role of vaginal apex brachytherapy in addition to pelvic external-beam treatment is uncertain
IC	Pelvic XRT	Role of vaginal apex brachytherapy in addition to pelvic external-beam treatment is uncertain
II	Pelvic XRT ± vaginal apex brachytherapy	Cervical involvement is associated with a higher risk of vaginal apex recurrence
IIIA or B	Pelvic XRT or WAXRT	Pelvic XRT is considered for high-risk features, including deep myometrial invasion, grade III, or adnexal-vaginal involvement. If cancer is stage IIIA on the basis of positive peritoneal washings, risk of upper abdominal relapse is increased if the tumor extends to the uterine serosa or involves other pelvic structures (e.g., adnexa or parametria); in this case, consideration of WAXRT is reasonable; adjuvant chemotherapy in addition to XRT for treatment of high-risk stage III patients is sometimes considered, although its value is uncertain
IIIC	Pelvic and para-aortic XRT (if para-aortic nodes involved)	Morbidity of radiotherapy is increased through the use of a para-aortic field (increased nausea/myelosuppression); WAXRT is also a consideration if washings are positive
IV	Possible WAXRT; possible adjuvant chemotherapy	Risk of relapse is high with WAXRT; consideration of adjuvant chemotherapy is reasonable in view of activity of platinum-based regimens, although value is unproven; hormonal or cytotoxic chemotherapy is reasonable for distant disease

*Assumes complete surgical staging as described in text.
WAXRT—whole abdominal external-beam radiotherapy (XRT)

other adnexal structures, such as the ovaries, and may involve lymph nodes; and stage IV disease involves areas outside of the pelvis, such as the upper abdomen, or distant sites, such as lung or bone. Because the ovaries can be involved by endometrial cancer, it is important that a BSO be performed as part of routine surgical resection; this also removes a potential source of estrogen production in premenopausal women. In contrast to the staging systems for ovarian and cervical cancers, endometrial cancer staging incorporates the use of grade in its descriptions because of the independent prognostic relevance of this feature in predicting the likelihood of lymph node involvement.

PROGNOSIS AND POSTOPERATIVE TREATMENT

Approximately 80% of patients with endometrial cancer present with stage I disease and have a 5-year survival rate of greater than 80%. The 5-year survival rates of patients with stage II, III, and IV tumors are 73%, 52%, and 27%, respectively.[79] In addition to advanced stage, other adverse prognostic factors are deep myometrial invasion,[87] high grade,[87] race (African Americans have a lower rate of survival, stage for stage),[88] overexpression of the HER-2/neu protein (a member of the EGF [epidermal growth factor] receptor family of tyrosine kinases),[89] *p53* mutation,[90] and aneuploidy (having either more or less than the normal number of 46 chromosomes).[79] Postoperative pelvic radiotherapy is considered when certain high-risk features are present that confer an increased risk of local pelvic failure. These features include (1) deeply invasive, high-grade, early-stage lesions (e.g., stage IC, grade III), especially with lymphovascular involvement (LVI), as these features are associated with an approximately 20% risk of occult pelvic lymph node involvement, and (2) cervical involvement (i.e., stage II), which confers an increased risk of vaginal vault recurrence. When para-aortic lymph node involvement is documented (i.e., stage IIIC), the radiation field is often extended to encompass both the pelvic and the para-aortic lymph node chains. Whole abdominal radiotherapy is reserved for the management of positive peritoneal cytology in the setting of adverse local features, such as tumor extension to the uterine serosa or adnexal involvement (stage IIIA)[91]; under these circumstances, patients are at risk for relapse to occur in the upper abdomen.[92,93] The use of postoperative pelvic radiotherapy probably does not confer a survival advantage, although it significantly reduces the chance of local relapse, which can be highly morbid if associated with pain and bleeding. There is no proven survival advantage associated with the use of adjuvant hormone therapy (e.g., progestational agents) or chemotherapy,[94-97] although these modalities may be useful for the management of systemic relapse.[98] The postoperative management of patients with the common endometrioid variety of uterine cancer entails treatment that depends on the disease stage [*see Table 5*].

MANAGEMENT OF RELAPSED OR METASTATIC DISEASE

Patients who experience an isolated vaginal vault relapse have an approximately 30% long-term survival rate if managed with resection followed by radiotherapy, when the latter is technically possible, as determined on the basis of the patient's previous radiation exposure. For all other patients, the chance of long-term survival after relapse is very low, and the goal of treatment is palliation. Likely sites of recurrent disease are the parenchyma of the lung, bone, liver, and local pelvic and para-aortic nodes. For local disease (e.g., a painful bone metastasis or a vertebral metastasis causing cord compression), radiotherapy is a reasonable option. For more extensive systemic disease that is not associated with significant symptoms, a trial of progestational therapy (e.g., medroxyprogesterone acetate) is reasonable, although response rates are generally low (about 10% to 20%) and are related to the differentiation status of the tumor.[98] Grade I lesions, which tend to be estrogen and progesterone receptor positive, are most likely to respond, compared with poorly differentiated grade III tumors. For rapidly progressive, symptomatic recurrence, platinum-based chemotherapy is a reasonable next step. Combinations of doxorubicin and cisplatin are associated with response rates of approximately 40%,[99] and paclitaxel is also effective for recurrent disease.[100-102] The combination of paclitaxel and carboplatin has not been extensively tested but may prove to be a useful option, especially in view of the generally improved toxicity profile of carboplatin compared with cisplatin.[103] However, the increased myelosuppression associated with carboplatin may limit its usefulness in patients who have previously received pelvic radiotherapy.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Cannistra SA: Cancer of the ovary. N Engl J Med 329:1550, 1993

2. Jemal A, Murry T, Samuels A, et al: Cancer statistics, 2003. CA Cancer J Clin 53:5, 2003

3. Huncharek M, Geschwind JF, Kupelnick B: Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. Anticancer Res 23:1955, 2003

4. Wong C, Hempling RE, Piver MS, et al: Perineal talc exposure and subsequent epithelial ovarian cancer: a case-control study. Obstet Gynecol 93:372, 1999

5. Rossing MA, Daling JR, Weiss NS, et al: Ovarian tumors in a cohort of infertile women. N Engl J Med 331:771, 1994

6. Ness RB, Cramer DW, Goodman MT, et al: Infertility, fertility drugs, and ovarian cancer: a pooled analysis of case-control studies. Am J Epidemiol 155:217, 2002

7. Venn A, Watson L, Bruinsma F, et al: Risk of cancer after use of fertility drugs with in-vitro fertilisation. Lancet 354:1586, 1999

8. Vachon CM, Mink PJ, Janney CA, et al: Association of parity and ovarian cancer risk by family history of breast or ovarian cancer in a population-based study of postmenopausal women. Epidemiology 13:66, 2002

9. Struewing JP, Hartge P, Wacholder S, et al: The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336:1401, 1997

10. Burke W, Daly M, Garber J, et al: Recommendations for follow-up care of individuals with an inherited predisposition to cancer. II. BRCA1 and BRCA2. Cancer Genetics Studies Consortium. JAMA 277:997, 1997

11. Scully R, Chen J, Plug A, et al: Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell 88:265, 1997

12. Rubin SC, Benjamin I, Behbakht K, et al: Clinical and pathological features of ovarian cancer in women with germ-line mutations of BRCA1. N Engl J Med 335:1413, 1996

13. Boyd J, Sonoda Y, Federici MG, et al: Clinicopathologic features of BRCA-linked and sporadic ovarian cancer. JAMA 283:2260, 2000

14. Pharoah PD, Easton DF, Stockton DL, et al: Survival in familial, BRCA1-associated, and BRCA2-associated epithelial ovarian cancer. United Kingdom Coordinating Committee for Cancer Research (UKCCCR) Familial Cancer Study Group. Cancer Res 59:868, 1999

15. Piver MS, Jishi MF, Tsukada YNG: Primary peritoneal carcinoma after prophylactic oophorectomy in women with a family history of ovarian cancer. Cancer 71:2751, 1993

16. Krainer M, Silva-Arrieta S, Fitzgerald MG, et al: Differential contributions of BRCA1 and BRCA2 to early onset breast cancer. N Engl J Med 336:1416, 1997

17. Risch HA, McLaughlin JR, Cole DE, et al: Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population of 649 women with ovarian cancer. Am J Hum Genet 68:700, 2001

18. Jacobs IJ, Skates SJ, MacDonald N, et al: Screening for ovarian cancer: a pilot randomised controlled trial. Lancet 353:1207, 1999

19. Schrag D, Kuntz KM, Garber JE, et al: Decision analysis: effects of prophylactic mastectomy and oophorectomy on life expectancy among women with BRCA1 or BRCA2 mutations. N Engl J Med 336:1465, 1997

20. Rebbeck TR, Lynch HT, Neuhausen SL, et al: Prophylactic oophorectomy in carriers of BRCA1 or BRCA2 mutations. N Engl J Med 346:1660, 2002

21. Rustgi AK: Hereditary gastrointestinal polyposis and nonpolyposis syndromes. N

Engl J Med 331:1694, 1994

22. Peel DJ, Ziogas A, Fox EA, et al: Characterization of hereditary nonpolyposis colorectal cancer families from a population-based series of cases. J Natl Cancer Inst 92:1517, 2000

23. Bast RC, Klug TL, John ES, et al: A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 309:883, 1983

24. Niloff JM, Knapp RC, Lavin PT, et al: The CA125 assay as a predictor of clinical recurrence in epithelial ovarian cancer. Am J Obstet Gynecol 155:56, 1986

25. Van der Burg ME, van Lent M, Buyse M, et al: The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. N Engl J Med 332:629, 1995

26. Young RC, Walton LA, Ellenberg SS, et al: Adjuvant therapy in stage I and stage II epithelial ovarian cancer. N Engl J Med 322:1021, 1990

27. Bolis G, Colombo N, Pecorelli S, et al: Adjuvant treatment for early epithelial ovarian cancer: results of two randomized clinical trials comparing cisplatin to no further treatment or chromic phosphate. Ann Oncol 6:887, 1995

28. Young RC, Brady MF, Nieberg RM, et al: Randomized clinical trial of adjuvant treatment of women with early FIGO I-IIA high-risk ovarian cancer-GOG #95 (abstr). Proc Am Soc Clin Oncol 18:357a, 1999

29. Trimbos JB, Parmar M, Vergote I, et al: International Collaborative Ovarian Neoplasm trial 1 and Adjuvant ChemoTherapy In Ovarian Neoplasm trial: two parallel randomized phase III trials of adjuvant chemotherapy in patients with early-stage ovarian carcinoma. J Natl Cancer Inst 95:105, 2003

30. Rowinsky EK, Donehower RC: Paclitaxel (Taxol). N Engl J Med 332:1004, 1995

31. Strobel T, Swanson L, Korsmeyer S, et al: BAX enhances paclitaxel-induced apoptosis through a p53-independent pathway. Proc Natl Acad Sci USA 93:14094, 1996

32. McGuire WP, Hoskins WJ, Brady MF, et al: Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. N Engl J Med 334:1, 1996

33. Stuart G, Bertelsen K, Mangiono C, et al: Updated analysis shows a highly significant improved overall survival for cisplatin-paclitaxel as first line treatment of advanced ovarian cancer: mature results of the EORTC-GCCG, NOCOVA, NCIC CTG, and Scottish intergroup trial (abstr). Proc Am Soc Clin Oncol 17:361a, 1998

34. Ozols RF, Bundy BN, Fowler J, et al: Randomized phase III study of cisplatin/paclitaxel versus carboplatin/paclitaxel in optimal stage III epithelial ovarian cancer: a Gynecologic Oncology Group (GOG) trial (abstr). Proc Am Soc Clin Oncol 18:356a, 1999

35. du Bois A, Lueck HJ, Meier W, et al: Cisplatin/paclitaxel vs carboplatin/paclitaxel in ovarian cancer: update of an Arbeitsgemeinschaft Gynaekologische Onkologie (AGO) study group trial (abstr). Proc Am Soc Clin Oncol 18:356a, 1999

36. Alberts DS, Liu PY, Hannigan EV, et al: Intraperitoneal cisplatin plus intravenous cyclophosphamide versus intravenous cisplatin plus intravenous cyclophosphamide for stage III ovarian cancer. N Engl J Med 335:1950, 1996

37. Markman M, Bundy BN, Alberts DS, et al: Phase III trial of standard-dose intravenous cisplatin plus paclitaxel versus moderately high-dose carboplatin followed by intravenous paclitaxel and intraperitoneal cisplatin in small-volume stage III ovarian carcinoma: an intergroup study of the Gynecologic Oncology Group, Southwestern Oncology Group, and Eastern Cooperative Oncology Group. J Clin Oncol 19:1001, 2001

38. Markman M, Liu PY, Wilczynski S, et al: Southwest Oncology Group. Gynecologic Oncology Group. Phase III randomized trial of 12 versus 3 months of maintenance n patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest Oncology Group and Gynecologic Oncology Group trial. J Clin Oncol 21:2460, 2003

39. Omura GA, Brady MF, Homesley HD, et al: Long-term follow-up and prognostic factor analysis in advanced ovarian carcinoma: the Gynecologic Oncology Group experience. J Clin Oncol 9:1138, 1991

40. Markman M, Lewis JL Jr, Saigo P, et al: Epithelial ovarian cancer in the elderly: the Memorial Sloan-Kettering Cancer Center experience. Cancer 71:634, 1993

41. Lavin PT, Knapp RC, Malkasian G, et al: CA 125 for the monitoring of ovarian carcinoma during primary therapy. Obstet Gynecol 69:223, 1987

42. Markman M, Rothman R, Hakes T, et al: Second-line platinum therapy in patients with ovarian cancer previously treated with cisplatin. J Clin Oncol 9:389, 1991

43. Righetti SC, Della Torre G, Pilotti S, et al: A comparative study of p53 mutations, protein accumulation, and response to cisplatin-based chemotherapy in advanced ovarian carcinoma. Cancer Res 56:689, 1996

44. Tai YT, Lee S, Niloff E, et al: BAX protein expression and clinical outcome in epithelial ovarian cancer. J Clin Oncol 16:2583, 1998

45. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244:707, 1989

46. Ahlgren JD, Ellison NM, Gottlieb RJ, et al: Hormonal palliation of chemoresistant ovarian cancer: three consecutive phase II trials of the Mid-Atlantic Oncology Program. J Clin Oncol 11:1957, 1993

47. Ozols RF: Treatment of recurrent ovarian cancer: increasing options—"recurrent" results. J Clin Oncol 15:2177, 1997

48. Parmar MK, Ledermann JA, Colombo N, et al: Trope C. ICON4/AGO-OVAR-2.2 trial. Paclitaxel plus platinum-based chemotherapy versus conventional platinum-based chemotherapy in women with relapsed ovarian cancer: the ICON4/AGO-OVAR-2.2 trial. Lancet 361:2099, 2003

49. Bowman A, Gabra, Langdon SP, et al: CA125 response is associated with estrogen receptor expression in a phase II trial of leytrozole in ovarian cancer: identification of an endocrine-sensitive subgroup. Clin Cancer Res 8:2233, 2002

50. Segna RA, Dottino PR, Mandeli JP, et al: Secondary cytoreduction for ovarian cancer following cisplatin therapy. J Clin Oncol 11:434, 1993

51. Rice LW, Berkowitz RS, Mark SD, et al: Epithelial ovarian tumors of borderline malignancy. Gynecol Oncol 39:195, 1990

52. Leake JF: Tumors of low malignant potential. Curr Opin Obstet Gynecol 4:81, 1992

53. Leake JF, Currie JL, Rosenshein NB, et al: Long-term follow-up of serous ovarian tumors of low malignant potential. Gynecol Oncol 47:150, 1992

54. Trope C, Kaern J: Management of borderline tumors of the ovary: state of the art. Semin Oncol 25:372, 1998

55. Trimble CL, Kosary C, Trimble EL: Long-term survival and patterns of care in women with ovarian tumors of low malignant potential. Gynecol Oncol 86:34, 2002

56. Williams SD, Blessing JA, Hatch KD, et al: Chemotherapy of advanced dysgerminoma: trials of the Gynecologic Oncology Group. J Clin Oncol 9:1950, 1991

57. Norris HJ, Zirkin HJ, Benson WL: Immature (malignant) teratoma of the ovary: a clinical and pathologic study of 58 cases. Cancer 37:2359, 1976

58. Singh P, Yordan EL, Graham JR, et al: Combination chemotherapy for immature (malignant) teratomas of the ovary. Asia Oceania J Obstet Gynaecol 13:455, 1987

59. Gershenson DM, Kavanagh JJ, Copeland LJ, et al: Treatment of malignant nondysgerminomatous germ cell tumors of the ovary with vinblastine, bleomycin, and cisplatin. Cancer 57:1731, 1986

60. Mitchell PL, Al-Nasiri N, A'Hern R, et al: Treatment of nondysgerminomatous ovarian germ cell tumors: an analysis of 69 cases. Cancer 85:2232, 1999

61. Low JJ, Perrin LC, Crandon AJ, et al: Conservative surgery to preserve ovarian function in patients with malignant ovarian germ cell tumors: a review of 74 cases. Cancer 89:391, 2000

62. Zanetta G, Bonazzi C, Cantu M, et al: Survival and reproductive function after treatment of malignant germ cell ovarian tumors. J Clin Oncol 19:1015, 2001

63. Gershenson DM: Management of early ovarian cancer: germ cell and sex cord-stromal tumors. Gynecol Oncol 55:S62, 1994

64. Schumer ST, Cannistra SA: Granulosa cell tumor of the ovary. J Clin Oncol 21:1180, 2003

65. Cannistra SA, Niloff JM: Cancer of the uterine cervix. N Engl J Med 334:1030, 1996

66. Koutsky LA, Holmes KK, Critchlow CW, et al: A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. N Engl J Med 327:1272, 1992

67. Munoz N, Bosch FX, de Sanjose S, et al: Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348:518, 2003

68. Schiffman M, Castle PE: Human papillomavirus: epidemiology and public health. Arch Pathol Lab Med 127:930, 2003

69. Klein RS, Ho GYF, Vermund SH, et al: Risk factors for squamous intraepithelial lesions on Pap smear in women at risk for human immunodeficiency virus infection. J Infect Dis 170:1404, 1994

70. Morris M, Eifel PJ, Lu J, et al: Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. N Engl J Med 340:1137, 1999

71. Rose PG, Bundy BN, Watkins EB, et al: Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. N Engl J Med 340:1144, 1999

72. Whitney CW, Sause W, Bundy BN, et al: A randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stages IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes. J Clin Oncol 17:1339, 1999

73. Peters WA III, Liu PY, Barrett RJ II, et al: Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. J Clin Oncol 18:1606, 2000

74. Keys HM, Bundy BN, Stehman FB, et al: Cisplatin, radiation, and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. N Engl J Med 340:1154, 1999

75. McGuire WP, Blessing JM, Lentz SS, et al: Paclitaxel has moderate activity in squamous cervix cancer: a Gynecologic Oncology Group study. J Clin Oncol 14:792, 1996

76. Morris M, Brader KR, Levenback C, et al: Phase II study of vinorelbine in advanced and recurrent squamous cell carcinoma of the cervix. J Clin Oncol 16:1094, 1998

77. Moore DH, McQuellon RP, Blessing JA, et al: A randomized phase III study of cisplatin versus cisplatin plus paclitaxel in stage IVB, recurrent or persistent squamous cell carcinoma of the cervix: a Gynecologic Oncology Group study. Proc Am Soc Clin Oncol 20:201a, 2001

78. Anderson GL, Judd HL, Kauntiz AM, et al: Effects of estrogen plus progestin on gynecologic cancers and associated diagnostic procedures: the Women's Health Initiative randomized trial. JAMA 290:1739, 2003

79. Rose PG: Endometrial carcinoma. N Engl J Med 335:640, 1996

80. Shenson DL, Gallion HH, Powell DE, et al: Loss of heterozygosity and genomic instability in synchronous endometrioid tumors of the ovary and endometrium. Cancer 76: 650, 1995

81. Fornander T, Cedermark B, Mattsson A, et al: Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. Lancet 1:117, 1989

82. Bernstein L, Deapen D, Cerhan JR, et al: Tamoxifen therapy for breast cancer and endometrial cancer risk. J Natl Cancer Inst 91:1654, 1999

83. Bergman L, Beelen ML, Gallee MP, et al: Risk and prognosis of endometrial cancer after tamoxifen for breast cancer. Comprehensive Cancer Centres' ALERT Group. Assessment of Liver and Endometrial Cancer Risk following Tamoxifen. Lancet 356:881, 2000

84. Fisher B, Costantino JP, Redmond CK, et al: Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. J Natl Cancer Inst 86:527, 1994

85. Creasman WT, Morrow CP, Bundy BN, et al: Surgical pathologic spread patterns of

endometrial cancer: a Gynecologic Oncology Group study. Cancer 60:2035, 1987

86. Kim YB, Niloff JM: Endometrial carcinoma: analysis of recurrence in patients treated with a strategy of minimizing lymph node sampling and radiation therapy. Obstet Gynecol 82:175, 1993

87. Zaino RJ: Pathologic indicators of prognosis in endometrial adenocarcinoma. Selected aspects emphasizing the GOG experience. Gynecologic Oncology Group. Pathol Annu 30:1, 1995

88. Hill HA, Coates RJ, Austin H, et al: Racial differences in tumor grade among women with endometrial cancer. Gynecol Oncol 56:154, 1995

89. Hetzel DJ, Wilson TO, Keeney GL, et al: HER-2/neu expression: a major prognostic factor in endometrial cancer. Gynecol Oncol 47:179, 1992

90. Kohler MF, Berchuck A, Davidoff AM, et al: Overexpression and mutation of p53 in endometrial carcinoma. Cancer Res 52:1622, 1992

91. Gibbons S, Martinez A, Schray M, et al: Adjuvant whole abdominopelvic irradiation for high risk endometrial carcinoma. Int J Radiat Oncol Biol Phys 21:1019, 1991

92. Takeshima N, Nishida H, Tabata T, et al: Positive peritoneal cytology in endometrial cancer: enhancement of other prognostic factors. Gynecol Oncol 82:470, 2001

93. Kadar N, Homesley HD, Malfetano JH: Positive peritoneal cytology is an adverse factor in endometrial carcinoma only if there is other evidence of extrauterine disease. Gynecol Oncol 46:145, 1992

94. Morrow CP, Bundy BN, Homesley HD, et al: Doxorubicin as an adjuvant following surgery and radiation therapy in patients with high-risk endometrial carcinoma, stage I and occult stage II: a Gynecologic Oncology Group study. Gynecol Oncol 36:166, 1990

95. Lewis GC, Slack NH, Mortel R, et al: Adjuvant progestin therapy in the primary definitive treatment of endometrial cancer. Gynecol Oncol 2:368, 1974

96. Vergote I, Kjorstad K, Abeler V, et al: A randomized trial of adjuvant progestagen in early endometrial cancer. Cancer 64:1011, 1989

97. Martin-Hirsch PL, Jarvis G, Kitchener H, et al: Progestagens for endometrial cancer. Cochrane Database Syst Rev (2):CD001040, 2000

98. Elit L, Hirte H: Current status and future innovations of hormonal agents, chemotherapy, and investigational agents in endometrial cancer. Curr Opin Obstet Gynecol 14:67, 2002

99. Thigpen T, Blessing J, Homesley H, et al: Phase III trial of doxorubicin with or without cisplatin in advanced or recurrent endometrial carcinoma: a Gynecologic Oncology Group (GOG) study (abstr). Proc Am Soc Clin Oncol 12:261, 1993

100. Lissoni A, Zanetta G, Losa G, et al: Phase II study of paclitaxel as salvage treatment in advanced endometrial cancer. Ann Oncol 7:861, 1996

101. Fleming GF, Fowler JM, Waggoner SE, et al: Phase I trial of escalating doses of paclitaxel combined with fixed doses of cisplatin and doxorubicin in patients with advanced gynecologic malignancies: a Gynecologic Oncology Group study (abstr). Proc Am Soc Clin Oncol 17:361a, 1998

102. Lincoln S, Blessing JA, Lee RB, et al: Activity of paclitaxel as second-line chemotherapy in endometrial carcinoma: a Gynecologic Oncology Group study. Gynecol Oncol 88:277, 2003

103. Hoskins PJ, Swenerton KD, Pike JA, et al: Paclitaxel and carboplatin, alone or with irradiation, in advanced or recurrent endometrial cancer: a phase II study. J Clin Oncol 19:4048, 2001

Acknowledgment

Figure 2 Tom Moore.

XI LYMPHOMAS

Kieron Dunleavy, m.d.
Wyndham H. Wilson, m.d., ph.d.

Lymphoma is the sixth most common type of cancer in the United States, with an estimated incidence of over 62,000 cases annually.[1] The incidence of lymphoma has risen steadily over recent decades. Approximately 85% of lymphoma cases are non-Hodgkin lymphoma (NHL) and 15% are Hodgkin lymphoma.

Hodgkin Lymphoma

epidemiology and etiology

In the United States and Western Europe, Hodgkin lymphoma has a higher incidence in men than in women and occurs in a bimodal age distribution, with the greatest peak in the third decade and a lesser peak in the seventh decade. In developing countries, the disease is likewise more common in males, but the incidence peaks in the second decade.[2] Family members of patients with a history of Hodgkin lymphoma are at increased risk for the disease: the risk may be increased by a factor of 10 in same-sex siblings and by as much as a factor of 100 in identical twins.[3,4] There is significant human leukocyte antigen (HLA) concordance in affected family members. Specific HLA class II antigens are associated with an increased risk of Hodgkin lymphoma, and they may be associated with particular histologic subtypes of the disease.[5,6]

Epstein-Barr virus (EBV) infection is associated with Hodgkin lymphoma. There is an increased incidence of Hodgkin lymphoma in persons with a history of infectious mononucleosis. A prospective, controlled study found elevated EBV titers in patients before the diagnosis of Hodgkin lymphoma.[7,8] In addition, the EBV genome can be detected in the malignant cells of about 50% of patients with Hodgkin lymphoma.[9] The EBV genome is more commonly detected in Hodgkin lymphoma tissues from young children, in patients from developing countries, and in classic mixed cellularity–type Hodgkin lymphoma.

histologic classification

Although a variety of histologic classification schemes have been used for Hodgkin lymphoma in the past, the current World Health Organization (WHO) classification defines two biologically distinct types: classic and nodular lymphocyte predominant.[10] In turn, classic Hodgkin lymphoma comprises four subtypes: nodular sclerosis, mixed cellularity, lymphocyte rich, and lymphocyte depleted. Nodular sclerosis is the most common subtype, especially in patients younger than 40 years, followed by mixed cellularity. The subtype names describe the histologic milieu in which Reed-Sternberg cells and variants of these cells appear, as well as the extent of normal lymphocytic infiltration. Cytokines may play a role in the histologic milieu of Hodgkin lymphoma, which is characterized by plasma cells, eosinophils, neutrophils, lymphocytes, and, in the nodular sclerosis subtype, broad bands of collagen.

Immunohistochemistry is important both in the subclassification of Hodgkin lymphoma and in differentiating Hodgkin lymphoma from NHL. In classic Hodgkin lymphoma, the tumor cells are usually positive for CD30 (a marker of activated cells) and for CD15 (a monocyte-macrophage marker) and negative for CD45 and pan–B cell and pan–T cell antigens. In contrast, the tumor cells of nodular lymphocyte-predominant Hodgkin lymphoma subtype are negative for CD30 and CD15 but positive for CD20, BCL6, and CD45. Unfortunately, molecular investigation of the Reed-Sternberg cell and its morphologic variants has been hindered by the fact that they constitute only a small proportion of the total number of cells in Hodgkin lymphomas; inflammatory cells predominate.

biology

The neoplastic cell of Hodgkin lymphoma is almost always a B cell—either the Reed-Sternberg cell or one of its mononuclear variants [*see Figure 1*]. These malignant cells originate from germinal center B cells that have undergone a transforming event of some kind.[11] The critical transforming events associated with Hodgkin lymphoma have yet to be defined. However, mutations of the *IκBα* gene, causing upregulation of the nuclear factor–κB pathway, have been identified in both EBV-negative and EBV-positive Hodgkin lymphomas, and mutations in the *FAS* gene have also been identified; mutations of the *p53* gene have been found in a subset of EBV-negative cases.[12-14] The role of other genes in the etiology of Hodgkin lymphoma is under investigation, and the advent of microarray technology, which can generate distinct gene-expression signatures of different lymphomas, is enhancing oncologists' understanding of the molecular events integral to this disease. The very different appearance, presentation, and natural history of the histologic subtypes of Hodgkin lymphoma suggest variable pathways of lymphomagenesis, each involving a complex interaction between genetic factors, environmental factors, and the immune system of affected persons.

diagnosis

Clinical Features

The most common presenting feature of Hodgkin lymphoma is painless lymph node enlargement, usually of one or more lymph nodes above the diaphragm. Mediastinal lymphadenopathy is common at presentation, especially in patients 15 to 35 years of age with classic nodular sclerosis. Subdiaphragmatic involvement is more common with the classic mixed-cellularity subtype. Nodular lymphocyte-predominant Hodgkin lymphoma rarely involves the mediastinum or marrow. Occasionally, Hodgkin lymphoma presents as cough, chest discomfort, or superior vena cava syndrome.

Hodgkin lymphoma is characterized by the orderly spread from one lymph node region to contiguous nodal sites. The course of untreated Hodgkin lymphoma is highly variable. In many young patients, the disease can be quite indolent, with waxing-and-waning lymphadenopathy; in other patients, however, it may follow a more virulent course. The spleen and the lymph nodes in the celiac axis are often the first sites of subdiaphragmatic disease, which is found in 20% to 30% of patients with disease otherwise confined to the neck, axillae, or chest. Typically, the para-aortic and pelvic lymph nodes are the next to become involved. Hepatic and marrow disease are associated with splenic involvement. Contiguous intrathoracic disease of-

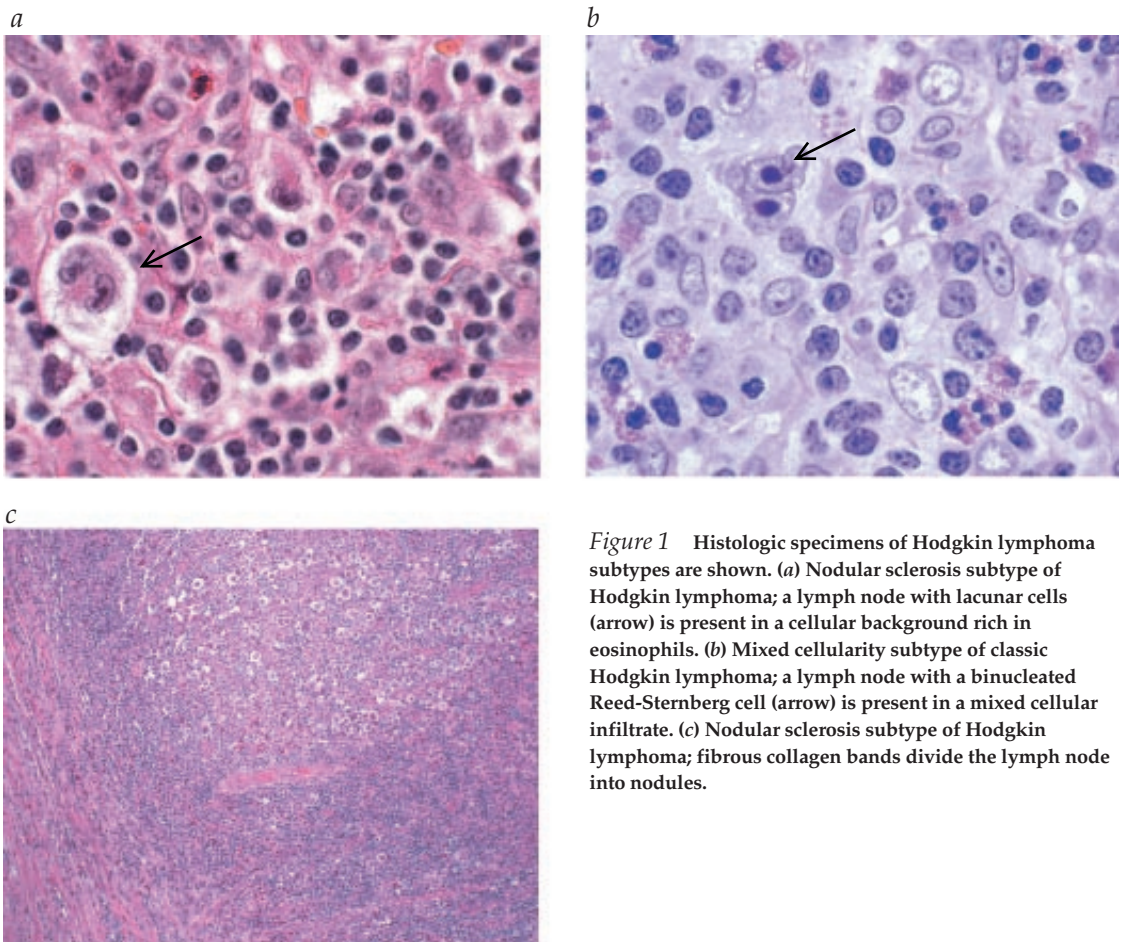


Figure 1 **Histologic specimens of Hodgkin lymphoma subtypes are shown. (*a*) Nodular sclerosis subtype of Hodgkin lymphoma; a lymph node with lacunar cells (arrow) is present in a cellular background rich in eosinophils. (*b*) Mixed cellularity subtype of classic Hodgkin lymphoma; a lymph node with a binucleated Reed-Sternberg cell (arrow) is present in a mixed cellular infiltrate. (*c*) Nodular sclerosis subtype of Hodgkin lymphoma; fibrous collagen bands divide the lymph node into nodules.**

ten occurs as a result of spread to the pulmonary parenchyma, pericardium, and chest wall from bulky mediastinal lymphadenopathy. Osseous lesions, which frequently appear osteoblastic on radiography, may develop by contiguous spread from enlarged lymph nodes or by hematogenous spread. Spinal cord compression may be caused by direct extension from enlarged nodes or by vertebral disease.

Some patients have prominent systemic symptoms known as B symptoms, which include drenching sweats at night, fever, and unexplained weight loss. Occasionally, patients experience Pel-Ebstein fevers, which are intermittent episodes of evening fevers that last for several days and alternate with afebrile periods. Total body pruritus, which is usually quite uncomfortable and refractory to antipruritic agents, may accompany Hodgkin lymphoma. A unique feature is pain at sites of lymphadenopathy immediately after ingestion of alcohol.

The clinical course of Hodgkin lymphoma differs according to histologic subtype, patient age, geography, and the integrity of the immune system. Nodular lymphocyte-predominant Hodgkin lymphoma follows a particularly indolent course: constitutional symptoms are rare and relapses are frequent but usually respond to therapy. In general, patients who are elderly, who live in developing countries, and who are infected with HIV are more likely to have widespread disease with systemic symptoms at diagnosis. The classic mixed-cellularity subtype is the most common variant observed in this group. Young-adult patients in the United States and Western Europe most often present with limited disease of the nodular sclerosis subtype.

History and Physical Examination

The history of current illness focuses on systemic symptoms, if any. The history should include questions regarding prior malignancy, chemotherapy or radiation treatment, immunosuppressive illness, or immunosuppressive therapy. Any family history of lymphoproliferative disorder should be elicited.

A detailed physical examination should include a careful assessment of all lymph node regions for extent of peripheral lymphadenopathy, as well as evaluation of the chest wall. Abdominal examination is performed to detect hepatosplenomegaly.

Laboratory Studies

The purpose of the diagnostic evaluation is to assess the extent (stage) of Hodgkin lymphoma, which determines prognosis and treatment. Toward this end, the single most useful diagnostic test for a patient with suspected lymphoma is a properly evaluated, technically adequate excisional lymph node biopsy. Optimal pathologic evaluation usually requires a whole lymph node.

Additional laboratory studies include a complete blood count; erythrocyte sedimentation rate (ESR); and serum chemistries, including the lactate dehydrogenase (LDH) level. The most common hematologic abnormality is normocytic, normochromic anemia, which is often found in patients with symptomatic, extensive disease. Cytopenias secondary to hypersplenism or marrow disease are rare. Modest elevations of alkaline phosphatase can occur in patients with limited disease; higher levels are associated with hepatic, osseous, or marrow involvement. Hypercalcemia is a rare paraneoplastic manifestation. The ESR

is commonly elevated in Hodgkin disease, and the degree of elevation has prognostic significance in patients with early-stage disease.

Imaging studies for the initial evaluation of Hodgkin lymphoma should include a screening chest radiograph and computed tomography of the chest, abdomen, and pelvis (with intravenous contrast if possible). CT is the gold-standard imaging modality. Positron emission tomography (PET) is an important adjunct to CT for initial staging. Its more important role is in the assessment of response to therapy by helping to distinguish active lymphoma from benign tissue in residual masses; in this setting, a positive PET scan is highly predictive of treatment failure.[15-17] PET findings, however, require careful interpretation, particularly with borderline specific uptake values. In general, patients with a positive PET scan after treatment should undergo biopsy to confirm the presence of active lymphoma.

One shortcoming of CT and other imaging modalities is a lack of reliability in diagnosing occult splenic and nodal disease. In one study, the overall accuracy of CT in detecting splenic disease was 58%.[18] Staging laparotomy, the only definitive modality for detecting occult splenic disease, was often included in the initial assessment of Hodgkin lymphoma in the past. Currently, its use is controversial; it is indicated only in patients for whom optimal management requires the assessment of occult splenic disease (i.e., patients for whom definitive radiotherapy is planned). Because chemotherapy is so commonly incorporated into the initial treatment of Hodgkin lymphoma, laparotomy is now rarely done. In one prospective study, laparotomy had no significant impact on survival or relapse rate in patients with a favorable prognosis who were randomized to staging with laparotomy or to clinical staging.[19] PET scanning, which may more accurately detect occult splenic disease in Hodgkin lymphoma, is very useful. Magnetic resonance imaging and isotope bone scanning are not routinely indicated in the initial assessment of Hodgkin lymphoma but may be useful in patients presenting with bone disease.

The incidence of marrow infiltration in Hodgkin lymphoma varies from series to series, with a range in adults of 2% to 32%.[20] In a review of 955 cases, the incidence of marrow involvement at diagnosis was 5.2%.[21] Despite its low yield, marrow biopsy is still recommended in all patients with newly diagnosed Hodgkin lymphoma.

Staging and Prognosis

The Ann Arbor staging system for Hodgkin lymphoma, developed over 30 years ago, continues to be used for the staging of NHL. A four-stage Cotswold modification, which accounts for such factors as bulky mediastinal disease, is currently used for staging Hodgkin lymphoma [*see Table 1*].[22]

Analysis of a large international dataset of patients with advanced Hodgkin lymphoma identified seven adverse factors, in addition to stage, that predicted outcome.[23] These are age greater than 45 years, stage IV disease, male sex, anemia, hypoalbuminemia, lymphopenia, and neutrophilia. The International Prognostic Score (IPS) incorporates these adverse factors and may be used to identify patients with poorer prognoses for more intensive therapies [*see Table 2*].

TREATMENT

An improved histopathologic classification, accurate staging, improved radiotherapy, and effective chemotherapeutic agents contribute to the high cure rate for Hodgkin lymphoma. For therapeutic success, care should be given by a multidisciplinary team with expertise in histopathology, diagnostic radiology, medical oncology, and radiation therapy.

Historically, radiation therapy was the preferred treatment for Hodgkin lymphoma, with chemotherapy reserved for advanced disease or radiotherapy failures. However, appreciation of the late effects of radiation therapy, including secondary malignancies and cardiac disease, has led to new approaches aimed at achieving high cure rates with minimal long-term toxicities.

Initial Therapy for Hodgkin Lymphoma, Stages I and II

Although stage of disease is the primary determinant of prognosis and therapy in Hodgkin lymphoma, the European Organization for Research and Treatment of Cancer (EORTC) has subdivided patients with early-stage (I and II) disease into favorable and unfavorable prognostic groups, on the basis of the presence (of at least one) or absence of the following adverse fac-

Table 1 Staging Classification of Lymphoma

*Stage**	*Ann Arbor Classification*	*Cotswold Modification*
I	Involvement of a single lymph node region (I) or of a single extralymphatic organ or site (I_E)	Involvement of a single lymph node region or lymphoid structure
II	Involvement of two or more lymph node regions on the same side of the diaphragm alone (II) or with involvement of limited, contiguous extralymphatic organ or tissue (II_E)	Involvement of two or more lymph node regions on the same side of the diaphragm (the mediastinum is considered a single site, whereas the hilar lymph nodes are considered bilaterally); the number of anatomic sites should be indicated by a subscript (e.g., II_3)
III	Involvement of lymph node regions on both sides of the diaphragm (III), which may include the spleen (III_S); a limited, contiguous extralymphatic organ or site (III_E); or both (III_{ES})	Involvement of lymph node regions on both sides of the diaphragm: III_1 (with or without involvement of splenic hilar, celiac, or portal nodes) and III_2 (with involvement of para-aortic, iliac, and mesenteric nodes)
IV	Multiple or disseminated foci of involvement of one or more extralymphatic organs or tissues, with or without lymphatic involvement	Involvement of one or more extranodal sites in addition to a site for which the designation E has been used

*All cases are subclassified to indicate the absence (A) or presence (B) of the systemic symptoms of significant fever (> 38.0° C [100.4° F]), night sweats, and unexplained weight loss exceeding 10% of normal body weight within previous 6 mo. The clinical stage (CS) denotes the stage as determined by all diagnostic examinations and a single diagnostic biopsy only. In the Ann Arbor classification, the term pathologic stage (PS) is used if a second biopsy of any kind has been obtained, whether negative or positive. In the Cotswold modification, the PS is determined by laparotomy; X designates bulky disease (widening of the mediastinum by more than one third or the presence of a nodal mass > 10 cm), and E designates involvement of a single extranodal site that is contiguous or proximal to the known nodal site.

Table 2 International Prognostic Score for Advanced Hodgkin Lymphoma[88]

*Prognostic Score**	*Frequency (%)*	*Progression-Free Survival at 5 Years (%)†*	*5-Year Survival (%)†*
0	7	84	89
1	22	77	90
2	29	67	81
3	23	60	78
4	12	51	61
≥ 5	7	42	56

*One point is given for the presence of each of the following prognostic factors: age > 45 yr, stage IV disease, male sex, albumin level < 4 g/dl, hemoglobin level < 10.5 g/dl, white blood cell count (WBC) ≥ 15,000/µl, and lymphocyte count < 600/µl or < 8% of WBC.
†Based on data from 5,141 patients.[23]

tors: large mediastinal mass, age 50 years or older, elevated ESR, and involvement of four or more lymph node regions.[24]

Early-stage favorable disease For decades, extended-field radiation therapy without chemotherapy was the treatment of choice for early-stage favorable disease. Although complete response rates were high, 25% to 30% of patients experienced relapse and required chemotherapy. Combined chemotherapy and radiation is favored by most groups; it results in superior cure rates, with potentially less long-term toxicity because of lower radiation doses.[15,25,26] A recommended combination consists of short-duration chemotherapy (e.g., two cycles of doxorubicin [Adriamycin], bleomycin, vinblastine, and dacarbazine [ABVD]) with radiation to the involved field (20 to 30 Gy).[26] For low-risk stage I neck disease, some groups still recommend radiotherapy alone.[27] Some advocate chemotherapy alone in early-stage favorable disease, but whether chemotherapy alone is equivalent to combination therapy is still under investigation.[28]

Early-stage unfavorable disease Combined-modality therapy (typically, four cycles of ABVD with 20 to 30 Gy of radiation to the involved field) is an accepted treatment for early-stage unfavorable disease. However, because early relapse occurs in 15% of patients who have received ABVD, more intensive chemotherapy regimens are under study, and the use of chemotherapy alone is increasing.[15]

Initial Therapy for Hodgkin Lymphoma, Stages III and IV

Combination chemotherapy is the treatment of choice for patients with advanced-stage disease. Although MOPP (mechlorethamine, vincristine [Oncovin], procarbazine, and prednisone) was the first multiagent regimen to achieve a high cure rate in this group, ABVD is currently the standard of care. In randomized comparisons, ABVD was more effective than MOPP and less toxic than MOPP hybrids.[29,30] Patients usually receive two cycles of therapy after complete remission, for a minimum of six and a maximum of eight cycles.

Patients with advanced-stage Hodgkin lymphoma who are treated with ABVD have close to 70% event-free survival. Nevertheless, investigators are attempting to improve outcome with more intensive regimens, many of which contain etoposide, a very active agent in Hodgkin lymphoma. The Stanford V regimen, a brief intensive regimen of chemotherapy (mechlorethamine, doxorubicin, vinblastine, vincristine, bleomycin, etoposide, and prednisone) combined with consolidative radiation to bulky lymph node sites, has produced a progression-free survival of 89% with an overall survival of 96% at a median follow-up of 5.4 years.[31] However, a randomized study suggested that it is not superior to ABVD.[32] The German Hodgkin Lymphoma Study Group has compared BEACOPP (bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, and prednisone) and increased-dose BEACOPP with the combination of COPP (cyclophosphamide, vincristine, procarbazine, and prednisone) and ABVD in patients with advanced Hodgkin lymphoma.[33] Radiation therapy, used for bulky disease or for residual disease after eight cycles of chemotherapy, was given to two thirds of patients. Although increased-dose BEACOPP had greater toxicity, it produced fewer induction failures and had a higher rate of freedom from treatment failure at 5 years and a better overall survival compared with COPP/ABVD. The advantage of increased-dose BEACOPP was greatest for patients with a poor prognosis, as defined by the IPS. Thus, although ABVD remains the standard for advanced-stage Hodgkin lymphoma, more intensive regimens may become standard for patients with poorer prognoses.

Initial Therapy in Nodular Lymphocyte-Predominant Hodgkin Lymphoma

This rare subtype of Hodgkin lymphoma behaves very differently from classic Hodgkin lymphoma and is biologically similar to low-grade B cell lymphoma. The 10-year survival for early-stage disease exceeds 80%, and the role of therapy for early-stage favorable disease is unclear. The EORTC recommends involved field radiation.[15] Patients with advanced-stage disease have an unfavorable prognosis and should probably be treated in the same way as patients with an unfavorable prognosis.[34] In a phase II trial, the anti-CD20 antibody rituximab produced a complete response rate of 46%, but follow-up was short; the long-term impact of this treatment requires further study.[35]

Relapsed or Refractory Hodgkin Lymphoma

In patients with early-stage disease who experience relapse after radiation alone, subsequent treatment with regimens such as ABVD results in relapse-free survival of up to 70% at 10 years.[36] Relapses after combination chemotherapy or combined-modality therapy present a greater challenge. If the duration of initial remission exceeds 12 months, retreatment with chemotherapy (either the same or a different combination) yields up to 50% relapse-free survival at 5 years.[37-39]

Patients who fail to achieve complete remission with initial chemotherapy or who experience relapse within 1 year after initial chemotherapy have the least favorable prognosis and a poor chance of long-term survival with standard-dose chemotherapy.[37,40,41] For both of these groups, high-dose chemotherapy with or without radiotherapy, followed by autologous hematopoietic stem cell transplantation, is a standard approach. Although few data on long-term outcome are available, early results demonstrate 5-year progression-free survival of up to 61%.[42]

A number of salvage regimens are active in relapsed Hodgkin lymphoma and may be used for disease control before autologous transplant. These include ESHAP (etoposide, methylprednisolone, high-dose cytarabine, and cisplatin), ICE (ifosfamide, carboplatin, and etoposide), and EPOCH (etoposide,

prednisone, vincristine, cyclophosphamide, and doxorubicin).[43-45] Relapse after transplantation presents a challenge. Low-dose, long-term treatment with agents such as vinblastine or gemcitabine may be used in such patients.[46,47]

Several experimental therapies for Hodgkin lymphoma are under investigation. These include monoclonal antibodies, radioimmunoconjugates, immunotherapy, tumor vaccines,[48,49] and allogeneic stem cell transplantation.[50]

Long-term Effects of Therapy

Because of the high cure rate of Hodgkin lymphoma, long-term toxicity of therapy is becoming increasingly important as the follow-up time for patients increases. The full toxicity profile of radiation is only now becoming apparent. The risk of several types of solid tumors is dramatically increased in patients who receive radiation therapy for Hodgkin lymphoma.[51,52] Patients who were treated when young are at greatest risk. In long-term survivors who received mediastinal radiation, clinically significant cardiovascular abnormalities also occur.[53] In addition, hypothyroidism is a frequently encountered late complication of radiotherapy.[54] Although the limited fields, lower doses, and improved techniques used in current radiation therapy are likely to ameliorate these toxicities, long-term follow-up is needed to confirm these benefits.

Late effects of chemotherapy include the development of leukemias and myelodysplasia in patients who have received alkylating agents such as nitrogen mustard and procarbazine in the MOPP regimen.[55] Sterility occurs in approximately 90% of men and 80% of women older than 25 years after treatment with a full course of MOPP; the incidence is lower in younger women.[56,57] In contrast, acute leukemia or myelodysplasia have rarely been reported in patients receiving ABVD, and the incidence of infertility is 10% to 20%.[58] The bleomycin in ABVD can produce long-term pulmonary toxicity.[59]

Non-Hodgkin Lymphoma

EPIDEMIOLOGY AND ETIOLOGY

NHL, with an estimated 54,370 new cases each year, accounts for 4% of all new cancer cases in the United States.[1] The incidence of NHL is higher in males and with increasing age. NHL is more common in Western countries, and different subtypes have particular geographic predilections. Over the past 20 years, the incidence of NHL has doubled in the United States.[60] Factors contributing to this trend include an increasing population at risk of AIDS-related lymphoma and an increase in the reporting and detection of lymphomas. However, these various factors account for only 50% of the observed increase; other reasons for the rise in incidence of NHL are unknown or poorly understood.[60,61] Environmental toxins are likely to be important etiologic factors.

NHL is also more commonly observed in patients with immunodeficiency states or autoimmune diseases and in patients receiving long-term immunosuppressive therapy.[60,62] EBV infection is involved in most of the lymphoproliferative diseases in these settings. Other associated viruses or pathogens, in addition to HIV, include human T cell lymphotropic virus type I (HTLV-I), hepatitis C virus, *Helicobacter pylori,* human herpesvirus type 8, *Borrelia burgdorferi,*[63-67] *Chlamydia psittaci,* and *Campylobacter jejuni*.[68,69]

HISTOLOGIC CLASSIFICATION

Various classification systems have been used for NHL, including Rappaport, Kiel, Working Formulation, and Revised European-American Classification of Lymphoid Neoplasms (REAL). The current WHO classification [*see Table 3*] evolved from an international collaboration of pathologists and clinicians and is based on insights into the molecular pathogenesis of lymphoma, including the identification of so-called hallmark genet-

Table 3 World Health Organization Classification of Hematopoietic Neoplasms[10]

B Cell Neoplasms	*T Cell and Natural Killer (NK) Cell Neoplasms*
Precursor B cell neoplasms	Precursor T cell neoplasms
Precursor B lymphoblastic leukemia/lymphoma	Precursor T lymphoblastic leukemia/lymphoma
	Blastic NK cell lymphoma
Mature B cell neoplasms	Mature T cell and NK cell neoplasms
Chronic lymphocytic leukemia/small lymphocytic lymphoma	Adult T cell leukemia/lymphoma
Lymphoplasmacytic lymphoma	Extranodal NK/T cell lymphoma, nasal type
Splenic marginal-zone lymphoma	Enteropathy-type T cell lymphoma
Extranodal marginal-zone B cell lymphoma of mucosa-associated lymphoid tissue (MALT lymphoma)	Hepatosplenic T cell lymphoma
Nodal marginal-zone B cell lymphoma	Subcutaneous panniculitis-like T cell lymphoma
Follicular lymphoma	Mycosis fungoides
Mantle cell lymphoma	Sézary syndrome
Diffuse large B cell lymphoma	Primary cutaneous anaplastic large cell lymphoma
Mediastinal (thymic) large B cell lymphoma	Peripheral T cell lymphoma, unspecified
Intravascular large B cell lymphoma	Angioimmunoblastic T cell lymphoma
Primary effusion lymphoma	Anaplastic large cell lymphoma
Burkitt lymphoma/leukemia	
B cell proliferations of uncertain malignant potential	
Lymphomatoid granulomatosis	
Posttransplant lymphoproliferative disorder, polymorphic	

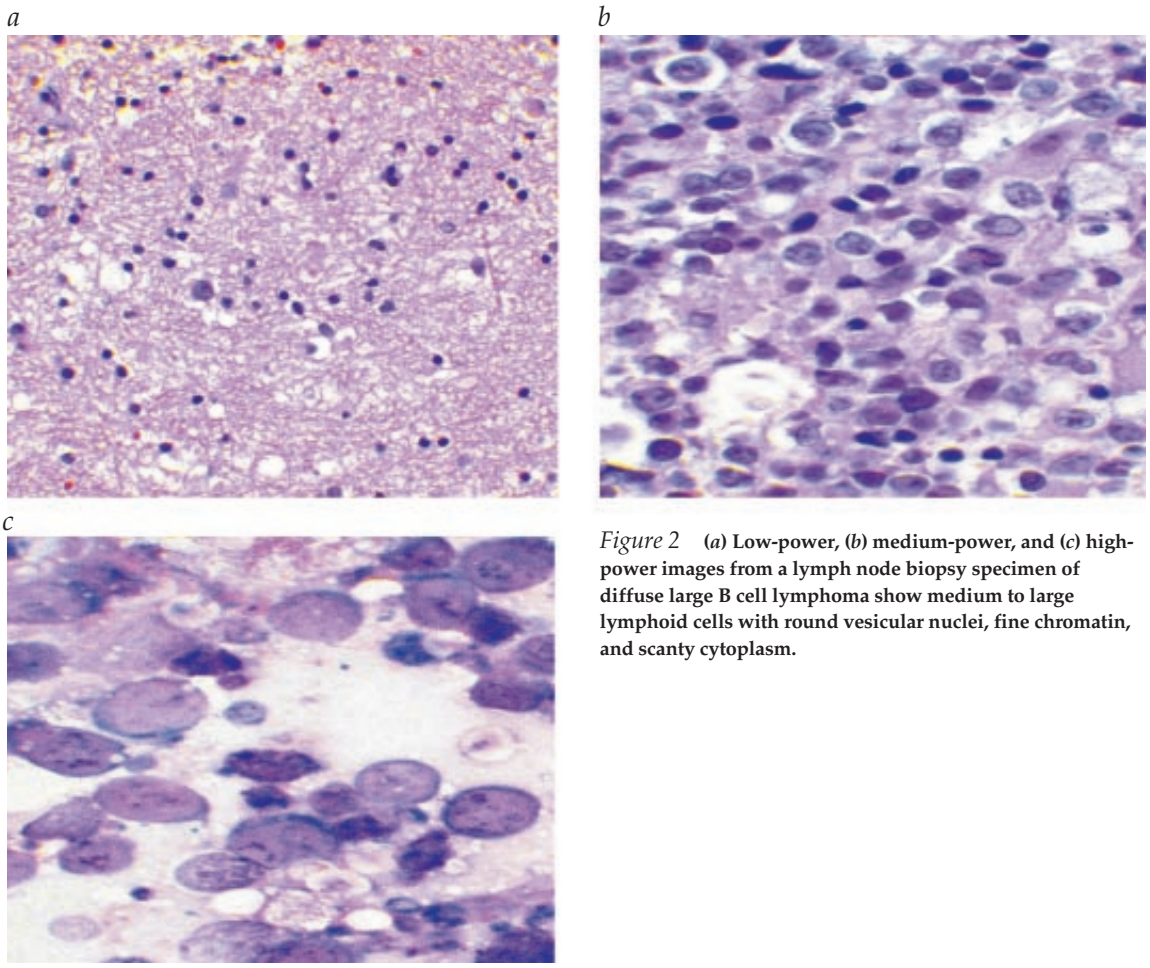


Figure 2 **(*a*) Low-power, (*b*) medium-power, and (*c*) high-power images from a lymph node biopsy specimen of diffuse large B cell lymphoma show medium to large lymphoid cells with round vesicular nuclei, fine chromatin, and scanty cytoplasm.**

ic abnormalities. Hematologic malignancies are stratified according to their lineage or cell of origin (B cell or T cell/natural killer [NK] cell).[10] Further, within each lineage or category of cell of origin, malignancies are described and defined by morphologic features, immunophenotype, genotype, and clinical behavior.

The pathologic assessment of lymphoma is based on the histologic features of the neoplasm and the immunophenotype [*see Figure 2*]. Morphology alone is often inadequate for classification, because NHLs with immunophenotypic and molecular heterogeneity may have a similar morphologic appearance. Immunophenotyping provides valuable information for classifying lymphomas. T cells and B cells express different cluster of differentiation (CD) antigens, and these antigens change according to the degree of maturation. By using monoclonal antibodies directed against these cell surface antigens, immunophenotypic analysis can define a lymphoma's cell of origin—B cell or T cell—and stage of differentiation. In addition, immunohistochemistry can detect gene products that are hallmarks of specific lymphoma subtypes, such as bcl-2, a protein that inhibits apoptosis, and bcl-1, a protein that promotes progression through the cell cycle [*see Figure 3*]. Many of these cell surface markers correlate with outcome and prognosis.[70-72]

Molecular diagnosis is very useful when morphology and immunohistochemistry fail to establish a definitive diagnosis.[73] All lymphomas have abnormal genes, a number of which are associated with cellular proliferation or apoptosis. Gene abnormalities (usually translocations) interfere with lymphoid cellular growth and differentiation, thereby causing malignant transformation. Abnormalities often involve genes that code for immunoglobulin chains. Thus, surface light-chain immunoglobulin and T cell receptor gene rearrangement, which can be detected by immunohistochemistry or DNA hybridization, can be useful in the assessment and diagnosis of lymphoid malignancies.

BIOLOGY

Lymphomas are derived from B cells, T cells, or NK cells at various stages of differentiation. For example, precursor B and precursor T lymphoblastic lymphomas are neoplasms of early lymphocyte differentiation (lymphoblasts committed to B cell and T cell lineages, respectively). In contrast, follicular lymphoma is a neoplasm of more mature lymphocytes (germinal center B cells), as is peripheral T cell lymphoma.

Karyotypic abnormalities characterize the majority of B cell NHLs [*see Table 4*]. Among these are the t(14;18)(q32;q21) translocation with rearrangement of the *bcl-2* gene, which is present in 80% to 90% of follicular lymphomas.[74,75] Inhibition of apoptosis by bcl-2 appears to play an important role in the pathogenesis of lymphomas.[75] *bcl-2* overexpression is also present in diffuse large B cell lymphoma (DLBCL) but less frequently (24% to 55% in two studies), and it is usually not associated with a *bcl-2* translocation.[70,76] Virtually all cases of mantle cell lymphoma contain the t(11;14)(q13;q32) translocation; this leads to overexpression of the *cyclin-D1* gene, which promotes progression from the G1 to S stages of the cell cycle.[77] Burkitt lymphoma is associated with a *MYC* translocation, which most commonly results from a t(8;14) translocation.[10]

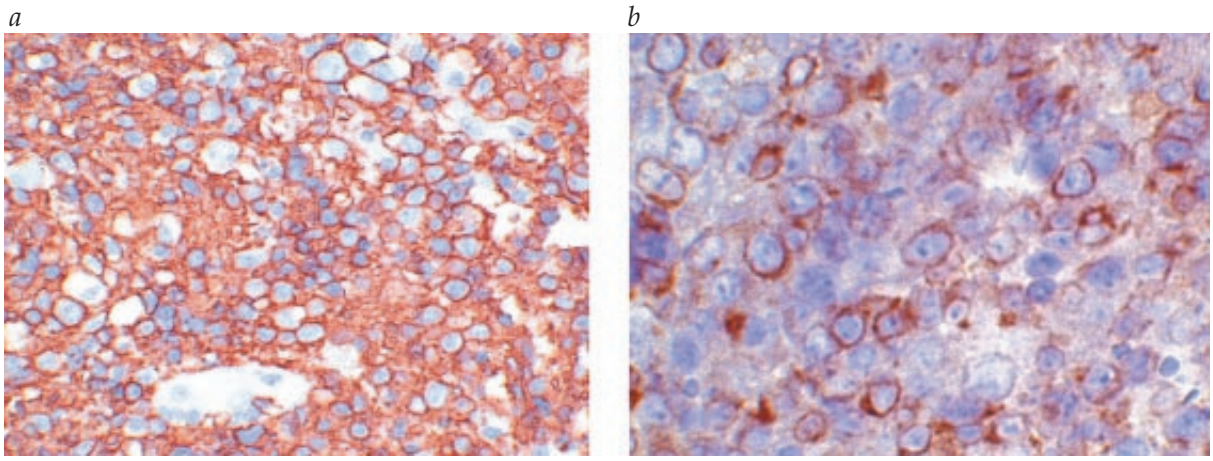


Figure 3 **Immunophenotypic analysis of a lymph node biopsy specimen from a patient with diffuse large B cell lymphoma shows that (*a*) large tumor cells express CD20 and (*b*) staining for the apoptosis-inhibiting protein bcl-2 is positive.**

Despite the increased precision in diagnosis made possible by advances in molecular techniques, response to therapy varies widely among specific lymphoma subtypes. In DLBCL, for example, only one third of patients are cured with CHOP (cyclophosphamide, hydroxydaunorubicin, vincristine, and prednisone), and about half are cured with rituximab-CHOP (R-CHOP) treatment.[78,79] This marked heterogeneity in therapeutic response implies that distinct diseases exist within a morphologic category. Molecular profiling has led to greater insights into the biologic diversity within morphologically defined categories of lymphoma.[80-82] RNA from a tumor sample can be analyzed by DNA microarrays, which can measure the expression of thousands of genes, thereby generating a distinct fingerprint of gene expression.

Table 4 Chromosomal Translocations in Non-Hodgkin Lymphoma[70-77,166-170]

*Lymphoma Subtype**	*Translocation*	*Percentage of Cases Affected*	*Proto-oncogene*	*Putative Function*	*Clinical Correlate*
Follicular center cell	t(14;18)(q32;q21)	80–90	*bcl-2*	Inhibits apoptosis	No association with prognosis
	17p13		*p53*	Confers resistance to chemotherapy	Large cell transformation and shorter time to disease progression
Small lymphocytic (CLL)	Trisomy 12	10–30	—	Loss of tumor suppressor function; associated with Ig-unmutated type	Advanced disease, atypical morphology, and resistant disease
	13q14	50	—	Putative tumor suppression gene unknown	Improved survival
	11q22-23	10–20	—	Unknown	Aggressive disease and worse prognosis
	17p13	6	*p53*	Confers resistance to chemotherapy	Shorter survival
Diffuse large B cell (DLBCL)	del(3)(q27)	35	*bcl-6*	Transcription factor for germinal center formation	Extranodal DLBCL; impact on treatment outcome controversial
	t(14;18)(q32;q21)	24–55	*bcl-2*	Inhibition of apoptosis	Poor prognosis
	17p13	20	*p53*	Confers resistance to chemotherapy	Poor prognosis
Mantle cell	t(11;14)(q13;q32)	—	*bcl-1/cyclin D1*	Cell cycle regulator; promotes G_1 to S progression	Seen in 70% of cases with cytogenetics and virtually 100% with FISH
	17p13	8	*p53*	Confers resistance to chemotherapy	Worse prognosis and blastic morphology
Burkitt	t(8;14)(q24;q32)	100	c-*myc*	Transcriptional activator regulating cell proliferation and growth	Present in 100% of cases; not entirely specific for Burkitt lymphoma
Marginal zone	t(11;18)(q21;q21)	60	*MLT-1*	Confers resistance to therapy through inhibition of apoptosis	Resistant to *Helicobacter*-eradication therapy
Anaplastic large cell	t(2;5)(p23;35)	50–80	*NPM/ALK*	Activates tyrosine kinase	Associated with a better prognosis

CLL—chronic lymphocytic leukemia FISH—fluorescence in situ hybridization

Clinical Features

A common presenting feature of NHL is painless peripheral lymphadenopathy. In contrast to Hodgkin lymphoma, mediastinal lymphadenopathy is uncommon at presentation. However, involvement of abdominal and pelvic nodes and extranodal sites is more frequent than with Hodgkin lymphoma. Extranodal disease can include any organ; the most common sites are the gastrointestinal tract, skin, and bone. Indolent lymphoma is typically characterized by waxing-and-waning lymphadenopathy, which can remain clinically quiescent over several years without therapy. In contrast, aggressive lymphoma is usually characterized by rapidly progressive lymphadenopathy that requires immediate therapy. Over a third of patients presenting with NHL have constitutional manifestations (so-called B symptoms), which include temperature higher than 38° C (100.4° F), unexplained loss of more than 10% of body weight, and night sweats; B symptoms are more common with aggressive subtypes. Marrow involvement is much more common in indolent lymphoma than in aggressive subtypes such as DLBCL.[79,83]

The clinical course of NHL depends on the histologic subtype, along with patient age and immune system status. Some histologic subtypes, such as Burkitt lymphoma and lymphoblastic lymphoma, are very aggressive, whereas follicular lymphoma is typically indolent. AIDS-related lymphomas, which are virtually always aggressive, are typically more advanced at presentation, with a high incidence of stage IV disease and a higher frequency of central nervous system involvement than similar lymphomas in HIV-negative patients.[84]

History and Physical Examination

Patients should be queried for systemic symptoms; any history of prior malignancy, chemotherapy or radiation treatment, and autoimmune or immunosuppressive illness; and any family history of malignancy. A history of infection with, or exposure to, pathogens such as HIV, hepatitis C virus, or HTLV-I should be sought. A detailed physical examination should include special attention to lymph node regions.

Laboratory Studies

The most important diagnostic test is a properly evaluated and technically adequate lymph node biopsy. Laboratory studies should include a complete blood count, serum chemistries (including LDH level), and HIV and hepatitis serologies. An elevated LDH has important prognostic implications for many lymphoma subtypes.[85] Potential etiologic pathogens or viruses suspected on the basis of the clinical evaluation should be investigated. Patients with chronic lymphatic leukemia or small lymphocytic lymphoma (SLL) should undergo serum protein electrophoresis with immunoglobulin quantification. Lymphoplas-

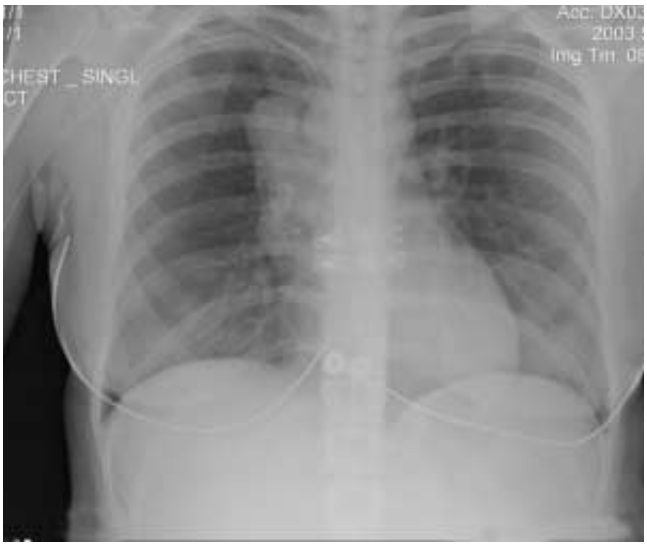


Figure 5 **A plain chest radiograph reveals a widened mediastinum in a patient with primary mediastinal B cell lymphoma.**

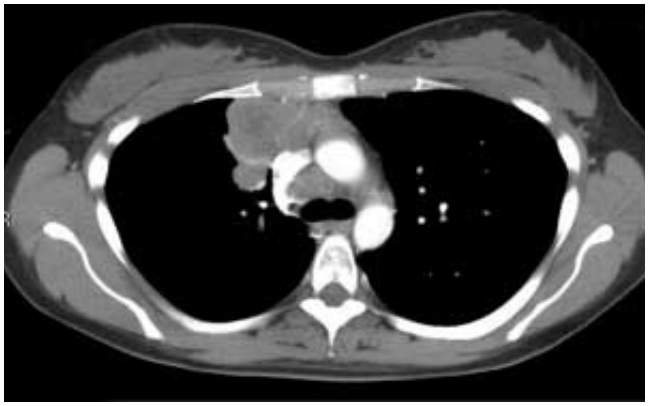


Figure 4 **CT scan of the chest showing an anterior right mediastinal mass and paratracheal lymphadenopathy. A biopsy of the mass was consistent with mediastinal (thymic) large B cell lymphoma.**

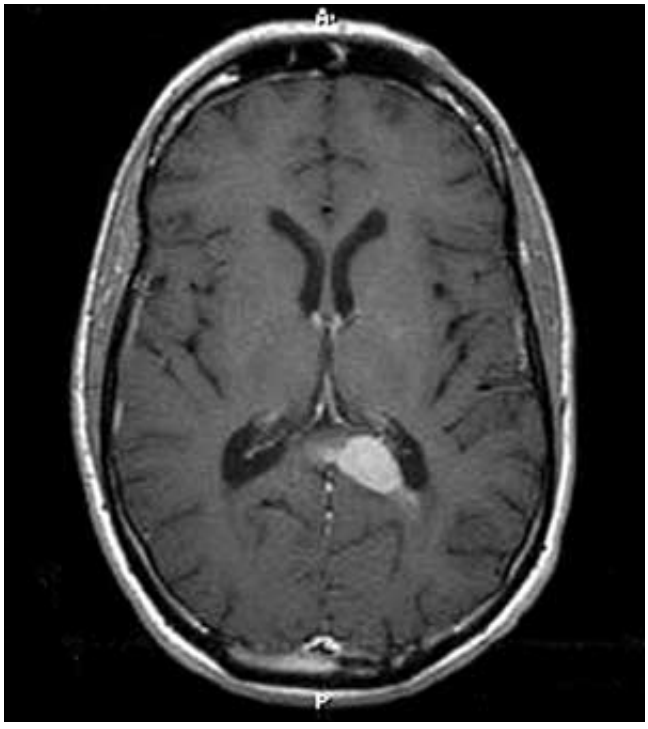


Figure 6 **An MRI scan of the brain with gadolinium contrast reveals an enhancing infiltrative mass in the major forceps of the corpus callosum. A biopsy revealed diffuse large B cell lymphoma.**

Table 5 Eastern Cooperative Oncology Group Performance Scale

Performance Status	*Definition*
0	Asymptomatic
1	Symptomatic but fully ambulatory
2	Symptomatic and in bed < 50% of the day
3	Symptomatic and in bed > 50% of the day
4	Bedridden

macytic lymphomas, such as Waldenström macroglobulinemia, are often associated with paraproteinemia. Monoclonal paraprotein can be monitored; immunoglobulin deficiency, which frequently accompanies these disease entities, can predispose to certain opportunistic infections.

Imaging studies should include a chest radiograph and CT of the chest, abdomen, and pelvis, with intravenous contrast if possible [*see Figures 4 and 5*]. The need for additional imaging studies, including MRI and PET scanning, depends on the clinical presentation and sites of disease. CT or MRI of the head is undertaken if CNS disease is strongly suspected [*see Figure 6*]. Certain aggressive histologic subtypes and clinical settings are associated with a higher risk of CNS disease.[86,87] Patients at high risk for CNS disease should undergo a lumbar puncture at diagnosis, with flow cytometric analysis of the cerebrospinal fluid in addition to cytologic evaluation.[87]

The incidence of marrow involvement in NHL varies according to histology but can approach 70% in indolent B cell lymphoma.[83] Consequently, marrow biopsy is a standard investigation at diagnosis.

Staging and Prognosis

The Ann Arbor staging system, originally designed for Hodgkin lymphoma, is now widely applied in the staging of NHL [*see Table 1*]. However, because of the tremendous heterogeneity of NHL and the lack of contiguous orderly spread from one lymph node region to another in these diseases, the Ann Arbor staging system is of limited prognostic value.

To identify prognostic factors in NHL, an international project correlated clinical variables and outcome in 2,031 patients with untreated aggressive lymphoma.[88] The following features were associated with a worse outcome: age greater than 60 years, Ann Arbor stage III or IV, elevated LDH, Eastern Oncology Group performance status of 2 or higher [*see Table 5*], and involvement of two or more extranodal sites. Project members created the International Prognostic Index (IPI) as a tool for predicting outcome. In the IPI, 1 point is allocated for each feature that is present. In the study cohort, 5-year survival proved to be directly related to the IPI score (survival of 73%, 51%, 43%, and 26% for scores of 0 to 1, 2, 3, and 4 to 5, respectively) [*see Table 6*].[88] Thus, the IPI is useful in predicting outcome for patients with aggressive lymphoma.

A validated clinical prognostic index is now available for patients with follicular lymphoma as well.[89] The Follicular Lymphoma International Prognostic Index (FLIPI), which is based on patient age, disease stage, serum LDH level, hemoglobin level, and number of nodal areas, has been found to reliably predict survival [*see Table 7*].

The clinical characteristics included in the IPI and FLIPI are probably surrogates for biologic features (see above). Expression of the proteins bcl-2, bcl-6, and MUM-1 has been evaluated and correlated with prognosis, particularly in DLBCL.[71,84,90-92] Several studies have identified bcl-2 positivity as a marker of poor outcome. A randomized study by a French group has validated the observation that the monoclonal antibody rituximab may overcome bcl-2–associated chemotherapy resistance in untreated DLBCL.[93,94]

Gene expression profiling is emerging as an important prognostic tool. In DLBCL, for example, morphologically identical tumors may show marked heterogeneity in gene expression. These genes can be classified into signatures associated with stages of differentiation.[95] On the basis of these signatures, DLBCL can be divided into two subtypes: DLBCL derived from a germinal center B cell, termed germinal center B cell–like (GCB) DLBCL; and DLBCL derived from a post–germinal center or activated B cell, termed activated B cell–like (ABC) DLBCL [*see Figure 7*].[96] Overall survival has proved to be significantly better in patients with the GCB subtype than in those with the ABC subtype (5-year survival of 60% versus 35%).[80] On the basis of these and other expression signatures, a molecular predictor of clinical outcome has been developed that can stratify patients with DLBCL into four prognostic quartiles [*see Figure 7*].[80]

The expression of six distinct genes has been found to be predictive of survival in DLBCL.[97] Gene expression profiling has also generated survival predictors in other NHLs, including follicular lymphoma, chronic lymphocytic leukemia (CLL), primary mediastinal B cell lymphoma, and mantle cell lymphoma.[81,82,98-100] Although molecular profiling is largely an experimental technique and is not widely available, it promises to improve pathologic diagnostic accuracy, predict outcome with

Table 6 International Prognostic Index for Aggressive Non-Hodgkin Lymphoma[88]

*IPI Score**	*Risk Group*	*CR Rate (%)*	*5-Year Overall Survival (%)*
0, 1	Low	87	73
2	Low intermediate	67	51
3	High intermediate	55	43
4, 5	High	44	26

*One point is given for the presence of each of the following characteristics: age > 60 yr, elevated serum LDH level, ECOG performance status ≥ 2, Ann Arbor stage III or IV, and more than two extranodal sites.

CR—complete response ECOG—Eastern Cooperative Oncology Group IPI—International Prognostic Index LDH—lactate dehydrogenase

Table 7 Follicular Lymphoma International Prognostic Index[89]

Risk Group	*Number of Factors**	*Distribution of Patients (%)*	*5-Year OS (SE)*	*10-Year OS (SE)*	*RR (95% CI)*
Low	0–1	36	90.6 (1.2)	70.7 (2.7)	1.0 (NA)
Intermediate	2	37	77.6 (1.6)	50.9 (2.7)	2.3 (1.9–2.8)
High	≥ 3	27	52.5 (2.3)	35.5 (2.8)	4.3 (3.5–5.3)

*Factors adversely affecting survival include age > 60 yr, Ann Arbor stage III or IV, more than four nodal sites, serum LDH level above the upper limit of normal, and hemoglobin < 12.0 g/dl.
CI—confidence interval LDH— lactate dehydrogenase NA—not applicable OS—overall survival percentage
RR—relative risk of death SE—standard error

greater precision, and elucidate pathways of lymphomagenesis, thus identifying pertinent cellular targets and paving the way for more individualized therapy. Integration of immunostaining and gene microarray into large clinical trials, as well as evaluation of their prognostic value, should facilitate development of new and very useful prognostic models that may guide therapeutic choices.

TREATMENT

Indolent B Cell Lymphoma

Indolent lymphomas comprise several different subtypes, of which follicular lymphoma is the most common.[101] Although new and experimental therapies may offer the possibility of cure in the future, these lymphomas are considered incurable with conventional therapy.[102] Because these diseases are characterized by a waxing-and-waning course and patients can live for years, even without therapy, they are often managed with a "watch and wait" approach.[103] Indications for treatment include symptomatic or aggressive disease, bulky lymphadenopathy or cytopenias from marrow involvement, or hypersplenism.

Follicular lymphoma After DLBCL, follicular lymphoma is the second most prevalent histologic subtype, constituting 35% of NHLs in adults in the United States.[10] Patients typically are elderly and have disseminated disease at diagnosis. During the course of the illness, transformation from follicular to large cell lymphoma occurs in up to 35% of patients.[10] Median survival is approximately 10 years, and the clinical course is characterized by treatment response followed by disease progression.

Histologically, follicular lymphoma is categorized into grades 1, 2, and 3, according to the number of centroblasts per high-power field. Histologic grade correlates with prognosis: grades 1 and 2 are indolent, whereas grade 3A is aggressive but may be curable with systemic chemotherapy. Grade 3B is considered a variant of DLBCL for the purposes of treatment, and it is potentially curable.

Because follicular lymphomas of grades 1 and 2 are considered incurable, several differing approaches to therapy are reasonable. For selected patients with Ann Arbor stage I and II disease, deferred therapy is an acceptable option. In a study of early-stage follicular lymphoma, more than half the patients did not need treatment during 6 years of follow-up, and survival in untreated patients was comparable to that in a series of treated patients.[103]

Radiation alone has been evaluated in early-stage follicular lymphoma. No prospective, randomized trial has compared radiation with another treatment modality, but radiation alone in selected patients has produced 10-year disease-free survival rates of greater than 60%.[104] Fludarabine is a very effective drug in follicular lymphoma, with high response rates in both untreated and relapsed disease.[105] Alkylating agents such as cyclophosphamide and prednisone also achieve high response rates and can be combined with fludarabine.[106] Combined-modality therapy in early-stage disease also produces good response rates.

Rituximab, the monoclonal antibody against CD20, is very effective in follicular lymphoma, with response rates of up to 73% in untreated patients and 43% in patients with relapsed or refractory disease.[107,108] For patients who previously responded to rituximab, the response rate with retreatment is 40%.[109] Chemotherapy regimens such as CHOP or fludarabine with rituximab are highly active and may produce durable responses.[110,111] In a randomized study of CHOP with or without rituximab for treatment of follicular lymphoma, rituximab was associated with a significantly higher response rate and longer treatment benefit.[112] Rituximab also has a potential role in maintenance therapy, where it may improve progression-free survival, and in combination with systemic chemotherapy in the initial treatment of follicular lymphoma.[110,113] These results suggest that rituximab is likely to become a standard component of initial chemotherapy.

Radioimmunotherapy of follicular lymphoma involves the delivery of targeted radiotherapy to tumor tissue by conjugating an anti-CD20 antibody to either yttrium-90 or iodine-131. The anti-CD20 radioimmunoconjugates ibritumomab tiuxetan and tositumomab are approved for relapsed or refractory follicular lymphoma. In a randomized trial of relapsed and refractory follicular or transformed lymphoma, the overall response rate was better with ibritumomab tiuxetan than with rituximab (80% versus 56%).[114] Radioimmunotherapy is also being evaluated in combination with systemic chemotherapy.[115] In a study that used tositumomab as initial treatment for follicular lymphoma, 75% of patients had a complete response, and the median progression-free survival was 6.1 years.[116] The exact role of radioimmunotherapy in follicular lymphoma remains to be defined, but the results thus far are very promising.

Both autologous and allogeneic stem cell transplantation can induce durable remissions in follicular lymphoma, and allotransplant may be curable in selected patients.[117] Promising experimental approaches include the use of idiotype vaccine. These vaccines are prepared by isolating tumor-specific antigen—the idiotype protein—from the neoplasm and then conjugating and immunologically manipulating it in vitro. Idiotype vaccination has proved to be effective against several tumor types.[118] Two randomized studies of idiotype vaccination after chemotherapy in follicular lymphoma are under way.[119] Another experimental treatment modality in follicular lymphoma is

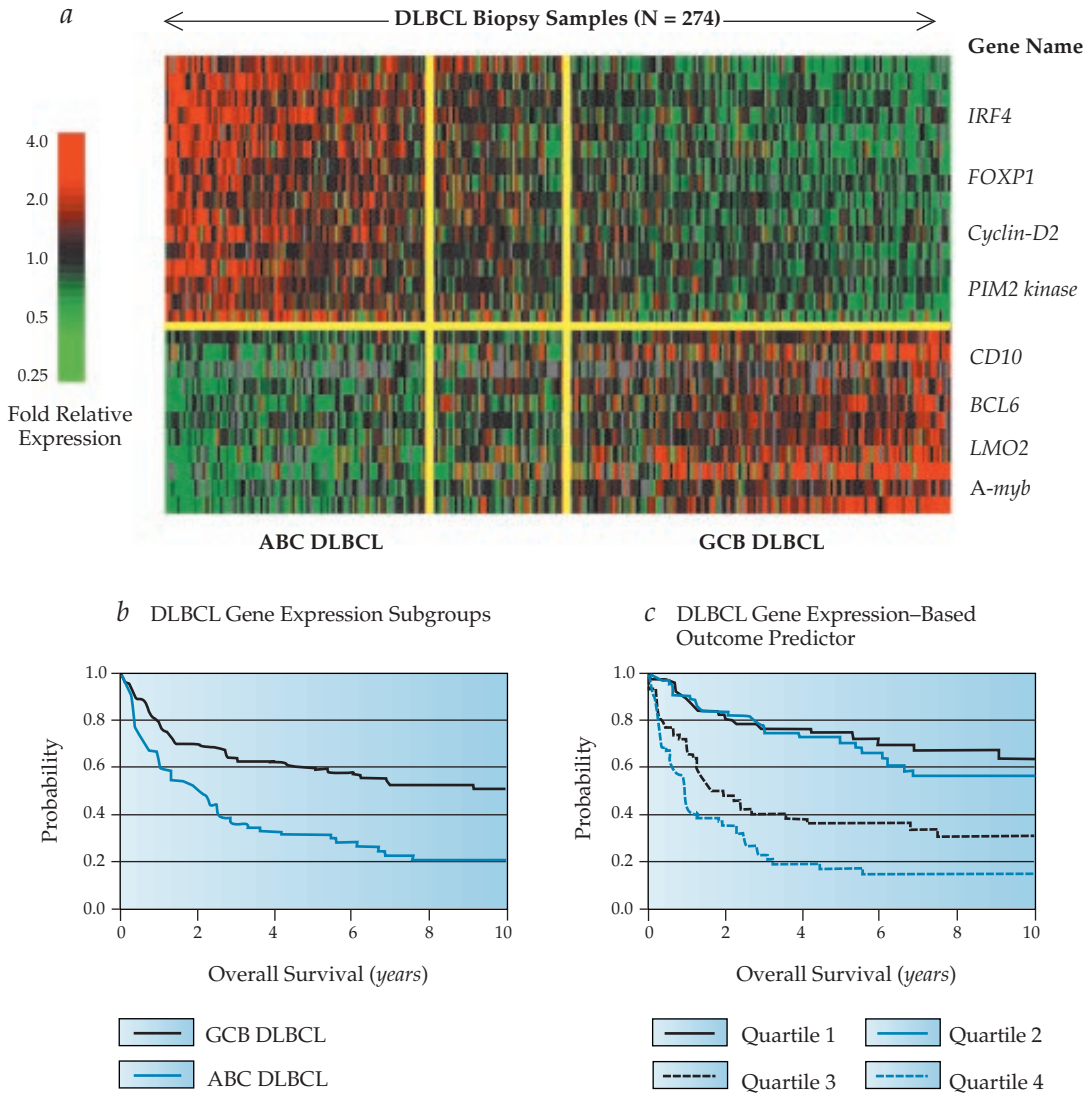


Figure 7 **Gene expression and development of a molecular predictor make it possible to diagnose disease subtypes in patients with diffuse large B cell lymphoma (DLBCL) after initial chemotherapy. (*a*) The expression levels of 27 genes from the subgroup predictor in 274 DLBCL samples are shown according to the color scale at the left. Eight named genes that showed increased expression in either the activated B cell (ABC) or the germinal center B cell (GCB) subgroups are shown at the right. (*b*) Kaplan-Meier estimates of overall survival are shown according to GCB or ABC DLBCL subtype; 5-year survival rates were 59% for patients with the GCB subtype and 31% for patients with the ABC subtype. (*c*) Kaplan-Meier estimates of overall survival according to the molecular outcome predictor are shown for each quartile. The 5-year survival rates were 73% for quartile 1, 71% for quartile 2, 36% for quartile 3, and 15% for quartile 4.**

bcl-2 antisense therapy, which putatively targets the bcl-2 oncoprotein.[120]

Small lymphocytic lymphoma and chronic lymphocytic leukemia SLL represents the solid tumor component of CLL; the two entities are considered to be biologically related [*see 12:XV Chronic Lymphoid Leukemias and Plasma Cell Disorders*]. Most patients are elderly. All patients with CLL have peripheral blood and marrow involvement at diagnosis, whereas SLL can be diagnosed in the absence of marrow or peripheral blood involvement. Treatment of CLL and SLL is similar to that of follicular lymphoma, with fludarabine and rituximab having important therapeutic roles.

Marginal-zone lymphoma Extranodal marginal-zone B cell lymphomas of mucosa-associated lymphoid tissue (MALT lymphoma) constitute 7% to 8% of all B cell lymphomas and a large proportion of primary gastric lymphomas.[10] They tend to be found in the gastrointestinal tract, salivary glands, thyroid gland, lungs, ocular adnexae, breast, and other extranodal sites. Most patients present with early-stage disease. A history of a chronic inflammatory disorder such as *Helicobacter*-associated chronic gastritis or Sjögren syndrome is common. In a small percentage of patients who have superficial, node-negative gastric MALT lymphomas, sustained remissions are possible after *Helicobacter* eradication with antibiotic therapy.[121] Otherwise, these tumors are sensitive to chemotherapy, rituximab, or radiothera-

py, alone or in combination; treatment of early-stage disease can result in long remissions.[122] MALT lymphomas tend to run an indolent course, remaining localized for long periods.

Splenic marginal-zone lymphoma, considered a separate entity in the WHO classification, involves the spleen, splenic hilar lymph nodes, marrow, and peripheral blood.[10] It characteristically has an indolent course and is classically treated with splenectomy; however, treatment with rituximab may ultimately replace this approach. Rituximab has been used in the treatment of extranodal marginal-zone lymphomas.[123] Nodal marginal-zone lymphoma is a rare, indolent, primary nodal B cell neoplasm that is considered the nodal counterpart of MALT lymphoma.

Aggressive B Cell Lymphomas

The aggressive B cell lymphomas include mantle cell lymphoma, DLBCL, Burkitt lymphoma, and lymphoblastic lymphoma. DLBCL is the most common aggressive B cell lymphoma, constituting approximately 30% to 40% of all adult NHLs in Western countries.[10] Both DLBCL and Burkitt lymphoma occur in the setting of HIV infection. Unlike indolent B cell lymphomas, the aggressive subtypes are, as their name implies, more rapidly progressive; and with the exception of mantle cell lymphoma, they are potentially curable. Hence, these tumors require expeditious evaluation and prompt, appropriate treatment.

Mantle cell lymphoma Mantle cell lymphoma is a B cell neoplasm that is composed of small to medium-size lymphoid cells with irregular nuclear outlines. It occurs at a median age of approximately 60 years and has a male predominance.[10] Most patients present with advanced-stage disease; the median survival is 3 years.[124] Morphologic variants of mantle cell lymphoma include two blastoid variants that are associated with a worse prognosis. The t(11;14) translocation occurs in almost all cases. Virtually all these tumors express cyclin-D1 (bcl-1), and a monoclonal antibody against cyclin-D1 is routinely used to confirm the diagnosis. Mantle cell lymphomas may respond to aggressive chemotherapy, but responses are generally short and the disease is considered incurable. Depending on the clinical circumstances, treatment strategies range from observation to aggressive treatment.

Although the overall response rate in mantle cell lymphoma is 89% with CHOP, the median overall survival is merely 37 months.[125] However, the hyper-CVAD regimen (hyperfractionated cyclophosphamide, doxorubicin, vincristine, and dexamethasone alternating with methotrexate and cytarabine) is effective, and adding rituximab enhances its effect.[126] In one study of untreated mantle cell lymphoma, DA-EPOCH-R (dose-adjusted etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin-rituximab) produced a complete response rate of 93%.[127] Trials of high-dose chemotherapy with autologous stem cell transplantation or allogeneic transplantation are ongoing but to date have yielded variable results.[128] Hence, transplantation should currently be considered experimental for the treatment of mantle cell lymphoma. The proteasome inhibitor bortezomib has proved active in relapsed mantle cell lymphoma and may become an important therapeutic agent for mantle cell lymphoma.[129,130]

Diffuse large B cell lymphoma DLBCL is the most prevalent histologic subtype of NHL. Histologically, it consists of a diffuse proliferation of large neoplastic B cells. Although the median age at diagnosis is in the seventh decade, DLBCL affects children and adults of any age. Patients may present with nodal or extranodal disease and localized or disseminated disease. The most common extranodal site is the gastrointestinal tract, but any extranodal tissue can be affected. DLBCL can arise de novo or result from the histologic transformation of an indolent B cell lymphoma. Morphologic variants within the category of DLBCL include centroblastic, immunoblastic, T cell rich/histiocyte rich, and anaplastic.[10] DLBCL can arise in any organ. Thus, different clinical behaviors and natural histories are seen in certain subtypes, such as primary mediastinal B cell lymphoma, which arises in the mediastinum, or primary CNS lymphoma (see below).

Radiation alone is inadequate for the treatment of DLBCL, with high recurrence rates both locally and distally.[131] Hence, all stages of disease require at least systemic chemotherapy. For early-stage disease (I or II), whether chemotherapy alone is adequate remains somewhat controversial. The benefit of combined-modality treatment was demonstrated in a randomized study comparing full-course CHOP with limited-course CHOP plus involved-field radiation in early-stage aggressive lymphoma.[132] In this study, combined-modality treatment was associated with superior overall and progression-free survival rates of 82% and 77%, respectively. However, a follow-up analysis showed a convergence of the overall survival curves because of an increase in late systemic relapses in the combined-modality arm.[133] Furthermore, an Eastern Cooperative Oncology Group study showed no survival benefit for combined-modality treatment over chemotherapy alone.[134] These results suggest that involved-field radiation may not be necessary with the use of more effective systemic treatments, such as R-CHOP or ACVBP (doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone).[79,135]

Some patients with early-stage primary mediastinal DLBCL (PMBCL), however, may require involved-field radiation after chemotherapy. In a study of 50 patients with untreated PMBCL who received MACOP-B (methotrexate with leucovorin rescue, doxorubicin, cyclophosphamide, vincristine, prednisone, and bleomycin) followed by radiation, 66% had a persistently positive gallium scan after chemotherapy alone, suggesting active disease.[136] However, treatment of PMBCL with other regimens that do not include involved-field radiation, such as ACVBP and DA-EPOCH, have yielded good results in PMBCL.[71,135] Furthermore, rituximab has improved the outcome in DLBCL.[79] Hence, it is reasonable to limit involved-field radiation to those patients with persistently positive PET scans after such treatments.

Advanced disease is treated with systemic chemotherapy. CHOP, developed in the 1970s, produces long-term remissions in approximately 35% of patients with DLBCL.[78] A number of newer regimens that improve on CHOP have been developed.[79,137-141] In a randomized trial, the addition of rituximab to CHOP resulted in a better response rate and survival in patients older than 60 years; thus, R-CHOP is emerging as the de facto standard for the treatment of DLBCL.[79] In poor-prognosis aggressive lymphoma, the regimen ACVBP improved event-free and overall survivals when compared with CHOP alone in patients older than 60 years.[138] The success of DA-EPOCH in untreated DLBCL led to the investigation of DA-EPOCH-R; this regimen has demonstrated progression-free and overall survival rates of approximately 80% at 2 years' median follow-up.[71,139] Unlike CHOP, DA-EPOCH appears to be effective in DLBCL with a high proliferative index.[71] Dose-dense CHOP

(given every 2 weeks) has been compared with CHOP given every 3 weeks, with or without the addition of etoposide (CHOP versus CHOEP), in young and elderly patients.[140,141] In young patients with aggressive lymphomas whose prognoses were good, survival was found to be better with CHOEP. In elderly patients, better results were seen with dose-dense CHOP than with CHOEP.

The initial treatment of aggressive DLBCL with high-dose chemotherapy and autologous transplantation continues to be studied but has not been definitively proved to be more effective than conventional chemotherapy.[142,143] One study showed that this approach provided better results than CHOP, but this conclusion is confounded by the study's inclusion of multiple lymphoma subtypes.[144] In addition, R-CHOP is now the standard for DLBCL, thus diminishing the value of studies that evaluate CHOP alone.

Treatment options for relapsed DLBCL include salvage chemotherapy and high-dose chemotherapy with autologous or allogeneic transplantation. For patients with a chemosensitive relapse who do not have marrow involvement, high-dose therapy and autologous transplant generally offer the best chance of response and remission.[142] With a higher percentage of current patients receiving regimens that are more effective than CHOP, the relative benefit of high-dose therapy will diminish, because current patients will have more resistant disease. Limited benefit can be obtained from conventional salvage regimens, and the choice between either salvage chemotherapy or high-dose therapy with autologous transplantation is influenced by a number of factors. Patients who experience relapse after an autologous transplantation may benefit from allogeneic transplantation.[145]

Primary central nervous system lymphoma Primary CNS lymphoma (PCNSL) is a rare lymphoma confined to the CNS; histologically, it is usually a DLBCL. Treatment of PCNSL is different from that of systemic NHL, because many of the chemotherapy agents used in the treatment of systemic NHL do not cross the blood-brain barrier. However, there may be a rationale for their use in patients in whom PCNSL disrupts this barrier. PCNSL is responsive to radiation, but responses tend to be short-lived and relapse rates high. Methotrexate is very effective in PCNSL, but when used alone, it produces a low median progression-free survival of only 12.8 months.[146] Methotrexate in combination with other chemotherapy agents is potentially curative, particularly in patients younger than 60 years, with an overall progression-free survival of 34% at 5 years.[147]

The role of consolidation radiotherapy after chemotherapy in PCNSL is controversial. Whole-brain radiation therapy is associated with significant neurotoxicity, particularly in elderly patients. Therefore, if results are equivalent with chemotherapy alone, radiation should be deferred until relapse. The Bonn group and others have reported good results with chemotherapy and deferred radiation.[148] This may become the standard approach for management of PCNSL in the future.[146,148]

Burkitt lymphoma Burkitt lymphoma is a highly aggressive lymphoma with three clinical settings: endemic, sporadic, and immunodeficiency related. Endemic Burkitt lymphoma occurs in young children in equatorial Africa; clinically, these patients often present with jaw and facial bone involvement. Sporadic Burkitt lymphoma, which is more commonly seen in children and young adults, often presents as abdominal disease or as other extranodal-site involvement. Immunodeficiency-related Burkitt lymphoma is usually seen in HIV-infected patients.

There are varying degrees of positivity for EBV infection with Burkitt lymphoma. All patients have a translocation of *MYC*.

Burkitt lymphoma is readily curable with systemic chemotherapy. Very intensive regimens that include intrathecal prophylactic therapy have traditionally been employed and continue to be the standard of care, despite significant toxicities.[149,150]

HIV-Related Lymphoma

For several reasons, including disordered immunosurveillance and chronic antigenic stimulation, the incidence of lymphoma is significantly increased in patients with HIV infection. Approximately 4% of patients with AIDS have NHL at the time of AIDS diagnosis. Although the incidence of all subtypes of lymphoma is increased in patients with HIV infection, the risk is increased 1,000-fold for Burkitt lymphoma and up to 400-fold for aggressive lymphoma.[135,151] Primary effusion lymphoma and plasmablastic lymphoma are almost exclusively seen in patients with HIV infection; both carry a very poor prognosis. Lymphomas are also observed with increased frequency after solid-organ transplantation; in certain autoimmune and inherited immunodeficiency syndromes; and with certain therapies, such as methotrexate.

Low-dose treatment strategies were standard in the early years of treating AIDS-related lymphomas, but highly active antiretroviral therapy (HAART) has improved survival and has thus allowed more aggressive treatment of lymphoma, with curative rather than palliative intent.[152] Systemic chemotherapy approaches for AIDS-related DLBCL and Burkitt lymphoma have included CHOP, but the efficacy of CHOP has been somewhat lower and its toxicity higher than in HIV-negative patients with similar lymphomas.[153] One major controversy in the treatment of AIDS-related lymphoma has been whether to suspend HAART during chemotherapy. HAART may enhance the toxicity of chemotherapy drugs, which may necessitate chemotherapy drug-dose reduction and thus compromise therapeutic efficacy. One approach is to suspend HAART during chemotherapy, with prompt resumption on completion of all cycles. The infusional regimen EPOCH, with HAART suspension, produced a 74% complete response rate and 72% overall survival at a median follow-up of 53 months.[89] These results were significantly superior to those for CHOP. The AIDS Malignancy Consortium is currently testing EPOCH in a randomized trial.

Epstein-Barr Virus–Associated Lymphoproliferative Disorders

Several lymphoproliferative disorders are associated with EBV infection. Posttransplant lymphoproliferative disorder (PTLD), which usually occurs after solid-organ transplantation, encompasses a broad spectrum of diseases with varying degrees of clinical aggressiveness. It is generally treated by withdrawal of chronic immunosuppression, but chemotherapy may be required for patients in whom PTLD behaves like an aggressive lymphoma or is resistant to immunosuppression withdrawal. Methotrexate-associated lymphoproliferative disorders can also be associated with EBV. Finally, lymphomatoid granulomatosis, a lymphoproliferative disorder involving extranodal sites, is EBV driven and is treated with interferon or chemotherapy, depending on the disease grade.[154]

Lymphoblastic Lymphoma

Precursor lymphoblastic lymphomas are cytologically identical to acute lymphoblastic leukemia. They are usually T cell lym-

phomas and are more common in adolescent boys or young men. Patients often present with a high white blood cell count and a large mediastinal mass. Treatment is with regimens that are appropriate for acute leukemia and that include CNS prophylaxis.

Mature T Cell and Natural Killer Cell Lymphomas

Lymphomas deriving from mature T cells and NK cells are uncommon, accounting for approximately 12% of all NHLs.[10] Their incidence varies significantly by geographic location, with an increased frequency in Asia. In lymphoma classifications that predated the current WHO classification, some of these lymphomas were considered unclassifiable or were not recognized. The natural history of mature T cell and NK cell lymphomas ranges from indolent to aggressive. Unfortunately, unlike aggressive B cell lymphomas, the majority of aggressive T/NK cell lymphomas are incurable. Important exceptions, however, include systemic anaplastic large cell lymphoma (ALCL), which is highly curable when it expresses the ALK fusion protein, and early-stage T/NK nasal angiocentric lymphoma (NAL), which is curable with high-dose radiation treatment.[155,156] Although a full description of these relatively rare diseases is beyond the scope of this chapter, several important subtypes are highlighted below.

Cutaneous T Cell Lymphomas

Cutaneous T cell lymphomas include mycosis fungoides, Sézary syndrome, primary cutaneous ALCL, and lymphomatoid papulosis. Mycosis fungoides presents in the skin with patches and plaques. The disease runs an indolent course and has a long natural history. General erythroderma may develop. Generalized disease dissemination is usually a late event in mycosis fungoides. A number of topical and local therapies, including ultraviolet radiation and skin electron-beam therapy, are very effective in mycosis fungoides. Systemic agents include the novel synthetic retinoid X receptor–selective retinoid bexarotene; the fusion toxin denileukin diftitox; and, experimentally, the histone deacetylase inhibitors.[157,158] Allogeneic transplantation has been investigated as a potentially curative treatment for advanced mycosis fungoides.[159]

Sézary syndrome consists of circulating neoplastic T cells, erythroderma, and lymphadenopathy. Unlike mycosis fungoides, Sézary syndrome is very aggressive.

Primary cutaneous ALCL affects the skin; this disease is distinct from systemic ALCL. In cases of localized skin involvement, skin-directed therapies are appropriate; extracutaneous disease requires more aggressive therapies. Lymphomatoid papulosis usually runs a benign clinical course and sometimes requires therapies such as low-dose methotrexate.

Anaplastic Large Cell Lymphoma

ALCL, a rare T cell lymphoma comprising 3% of all NHLs in adults, typically occurs in young men. ALCL is very curable; patients have a 70% long-term survival rate after systemic chemotherapy. The histologic hallmark of ALCL is CD30 positivity; in addition, most of these tumors express epithelial membrane antigen.[160] ALCL is associated with the t(2;5)(p23;q35) translocation, which results in the expression of the nucleophosmin anaplastic lymphoma kinase (ALK). ALK positivity, present in more than 50% of ALCLs, is associated with a better prognosis than ALK-negative disease.[155] ALK-positive ALCL often occurs in children or young adults. Most patients present with advanced disease and may have involvement of extranodal sites. ALCL responds very well to systemic chemotherapy, particularly if the disease is ALK positive.

Angioimmunoblastic T Cell Lymphoma

Angioimmunoblastic T cell lymphoma is a distinct type of peripheral T cell lymphoma that is associated with scattered EBV-positive B cells in biopsy specimens. Clinically, this disease is characterized by systemic symptoms, skin rash, organomegaly, hypergammaglobulinemia, and hemolytic anemia. Patients exhibit immunodeficiency and are particularly susceptible to infections. Angioimmunoblastic T cell lymphoma generally has an aggressive clinical course and is usually incurable with conventional chemotherapy.[161]

Peripheral T Cell Lymphoma, Unspecified

A large number of predominantly nodal T cell lymphomas are not included in other T cell lymphoma categories in the WHO classification. This group of lymphomas is very diverse, both morphologically and clinically. Patients generally have a poor long-term response to conventional chemotherapy, with a high rate of relapse and low rates of overall survival.[162] High-dose chemotherapy followed by autologous stem cell transplantation has been used with some success in patients with relapsed disease and offers some chance of long-term remission.[163,164] Early results with allogeneic transplantation in relapsed peripheral T cell lymphoma is promising, with some apparent graft versus lymphoma effect.[165]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Jemal A, Tiwari RC, Murray T, et al: Cancer statistics, 2004. CA Cancer J Clin 54:8, 2004

2. Correa P, O'Conor GT: Epidemiologic patterns of Hodgkin disease. Int J Cancer 8:192, 1971

3. Grufferman S, Cole P, Smith PG, et al: Hodgkin disease in siblings. N Engl J Med 296:248, 1977

4. Mack TM, Cozen W, Shibata DK, et al: Concordance for Hodgkin disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. N Engl J Med 332:413, 1995

5. Klitz W, Aldrich CL, Fildes N, et al: Localization of predisposition to Hodgkin disease in the HLA class II region. Am J Hum Genet 54:497, 1994

6. Harty LC, Lin AY, Goldstein AM, et al: HLA-DR, HLA-DQ, and TAP genes in familial Hodgkin disease. Blood 99:690, 2002

7. Mueller N, Evans A, Harris NL, et al: Hodgkin disease and Epstein-Barr virus: altered antibody pattern before diagnosis. N Engl J Med 320:689, 1989

8. Alexander FE, Jarrett RF, Lawrence D, et al: Risk factors for Hodgkin disease by Epstein-Barr virus (EBV) status: prior infection by EBV and other agents. Br J Cancer 82:1117, 2000

9. Weiss LM, Chen YY, Liu XF, et al: Epstein-Barr virus and Hodgkin disease: a correlative in situ hybridization and polymerase chain reaction study. Am J Pathol 139:1259, 1991

10. World Health Organization Classification of Tumors: Pathology and Genetics of Tumors of Hematopoietic and Lymphoid Tissues. Jaffe ES, Harris NL, Stein H, Eds. IARC Press, Lyon, France, 2001

11. Kuppers R, Rajewsky K: The origin of Hodgkin and Reed/Sternberg cells in Hodgkin disease. Annu Rev Immunol 16:471, 1998

12. Emmerich F, Meiser M, Hummel M, et al: Overexpression of I kappa B alpha without inhibition of NF-kappaB activity and mutations in the I kappa B alpha gene in Reed-Sternberg cells. Blood 94:3129, 1999

13. Jungnickel B, Staratschek-Jox A, Brauninger A, et al: Clonal deleterious mutations in the IkappaBalpha gene in the malignant cells in Hodgkin lymphoma. J Exp Med 191:395, 2000

14. Thomas RK, Re D, Wolf J, et al: Part I: Hodgkin lymphoma: molecular biology of Hodgkin and Reed-Sternberg cells. Lancet Oncol 5:11, 2004

15. Diehl V, Thomas RK, Re D: Part II: Hodgkin lymphoma: diagnosis and treatment. Lancet Oncol 5:19, 2004

16. Staak JO, Dietlein M, Engert A, et al. [Hodgkin lymphoma in nuclear medicine: diagnostic and therapeutic aspects]. Nuklearmedizin 42:19, 2003

17. Zinzani PL, Fanti S, Battista G, et al: Predictive role of positron emission tomography (PET) in the outcome of lymphoma patients. Br J Cancer 91:850, 2004

18. Castellino RA, Hoppe RT, Blank N, et al: Computed tomography, lymphography, and staging laparotomy: correlations in initial staging of Hodgkin disease. AJR Am J Roentgenol 143:37, 1984

19. Carde P, Hagenbeek A, Hayat M, et al: Clinical staging versus laparotomy and combined modality with MOPP versus ABVD in early-stage Hodgkin disease: the H6 twin randomized trials from the European Organization for Research and Treatment of Cancer Lymphoma Cooperative Group. J Clin Oncol 11:2258, 1993

20. Franco V, Tripodo C, Rizzo A, et al: Bone marrow biopsy in Hodgkin lymphoma. Eur J Haematol 73:149, 2004

21. Howell SJ, Grey M, Chang J, et al: The value of bone marrow examination in the staging of Hodgkin lymphoma: a review of 955 cases seen in a regional cancer centre. Br J Haematol 119:408, 2002

22. Lister TA, Crowther D, Sutcliffe SB, et al: Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin disease: Cotswolds meeting. J Clin Oncol 7:1630, 1989

23. Hasenclever D, Diehl V: A prognostic score for advanced Hodgkin disease. International Prognostic Factors Project on Advanced Hodgkin Disease. N Engl J Med 339:1506, 1998

24. Cosset JM, Henry-Amar M, Meerwaldt JH, et al: The EORTC trials for limited stage Hodgkin's disease. The EORTC Lymphoma Cooperative Group. Eur J Cancer 28A:1847, 1992

25. Karmiris TD, Grigoriou E, Tsantekidou M, et al: Treatment of early clinically staged Hodgkin disease with a combination of ABVD chemotherapy plus limited field radiotherapy. Leuk Lymphoma 44:1523, 2003

26. Bonadonna G, Bonfante V, Viviani S, et al: ABVD plus subtotal nodal versus involved-field radiotherapy in early-stage Hodgkin disease: long-term results. J Clin Oncol 22:2835, 2004

27. Bessell EM, MacLennan KA, Toghill PJ, et al: Suprahyoid Hodgkin's disease stage IA. Radiother Oncol 22:190, 1991

28. Straus DJ, Portlock CS, Qin J, et al: Results of a prospective randomized clinical trial of doxorubicin, bleomycin, vinblastine, and dacarbazine (ABVD) followed by radiation therapy (RT) versus ABVD alone for stages I, II, and IIIA nonbulky Hodgkin disease. Blood 104:3483, 2004

29. Canellos GP, Niedzwiecki D: Long-term follow-up of Hodgkin disease trial. N Engl J Med 346:1417, 2002

30. Duggan DB, Petroni GR, Johnson JL, et al: Randomized comparison of ABVD and MOPP/ABV hybrid for the treatment of advanced Hodgkin disease: report of an intergroup trial. J Clin Oncol 21:607, 2003

31. Horning SJ, Hoppe RT, Breslin S, et al: Stanford V and radiotherapy for locally extensive and advanced Hodgkin disease: mature results of a prospective clinical trial. J Clin Oncol 20:630, 2002

32. Chisesi T, Federico M, Levis A, et al: ABVD versus stanford V versus MEC in unfavourable Hodgkin's lymphoma: results of a randomised trial. Ann Oncol 13(suppl 1):102, 2002

33. Diehl V, Franklin J, Pfreundschuh M, et al: Standard and increased-dose BEACOPP chemotherapy compared with COPP-ABVD for advanced Hodgkin disease. N Engl J Med 348:2386, 2003

34. Diehl V, Sextro M, Franklin J, et al: Clinical presentation, course, and prognostic factors in lymphocyte-predominant Hodgkin disease and lymphocyte-rich classical Hodgkin disease: report from the European Task Force on Lymphoma Project on Lymphocyte-Predominant Hodgkin Disease. J Clin Oncol 17:776, 1999

35. Ekstrand BC, Lucas JB, Horwitz SM, et al: Rituximab in lymphocyte-predominant Hodgkin disease: results of a phase 2 trial. Blood 101:4285, 2003

36. Ng AK, Li S, Neuberg D, et al: Comparison of MOPP versus ABVD as salvage therapy in patients who relapse after radiation therapy alone for Hodgkin disease. Ann Oncol 15:270, 2004

37. Lohri A, Barnett M, Fairey RN, et al: Outcome of treatment of first relapse of Hodgkin disease after primary chemotherapy: identification of risk factors from the British Columbia experience 1970 to 1988. Blood 77:2292, 1991

38. Viviani S, Santoro A, Negretti E, et al: Salvage chemotherapy in Hodgkin disease: results in patients relapsing more than twelve months after first complete remission. Ann Oncol 1:123, 1990

39. Vassilakopoulos TP, Angelopoulou MK, Siakantaris MP, et al: Hodgkin lymphoma in first relapse following chemotherapy or combined modality therapy: analysis of outcome and prognostic factors after conventional salvage therapy. Eur J Haematol 68:289, 2002

40. Yuen AR, Rosenberg SA, Hoppe RT, et al: Comparison between conventional salvage therapy and high-dose therapy with autografting for recurrent or refractory Hodgkin disease. Blood 89:814, 1997

41. Bonfante V, Santoro A, Viviani S, et al: Outcome of patients with Hodgkin disease failing after primary MOPP-ABVD. J Clin Oncol 15:528, 1997

42. Diehl V, Stein H, Hummel M, et al: Hodgkin's lymphoma: biology and treatment strategies for primary, refractory, and relapsed disease. Hematology (Am Soc Hematol Educ Program), p 225, 2003

43. Aparicio J, Segura A, Garcera S, et al: ESHAP is an active regimen for relapsing Hodgkin's disease. Ann Oncol 10:593, 1999

44. Moskowitz CH, Bertino JR, Glassman JR, et al: Ifosfamide, carboplatin, and etoposide: a highly effective cytoreduction and peripheral-blood progenitor-cell mobilization regimen for transplant-eligible patients with non-Hodgkin's lymphoma. J Clin Oncol 17:3776, 1999

45. Dunleavy K, Butrynski J, Steinberg S, et al: Phase II study of EPOCH infusional chemotherapy in relapsed or refractory Hodgkin's lymphoma (HL): a report on toxicity, efficacy, and prognostic indicators of outcome. J Clin Oncol 22:6598, 2004

46. Little R, Wittes RE, Longo DL, et al: Vinblastine for recurrent Hodgkin's disease following autologous bone marrow transplant. J Clin Oncol 16:584, 1998

47. Santoro A, Bredenfeld H, Devizzi L, et al: Gemcitabine in the treatment of refractory Hodgkin's disease: results of a multicenter phase II study. J Clin Oncol 18:2615, 2000

48. Schnell R, Borchmann P, Schulz H, et al: Current strategies of antibody-based treatment in Hodgkin disease. Ann Oncol 13(suppl 1):57, 2002

49. Landgren O, Bjorkholm M, Konradsen HB, et al: A prospective study on antibody response to repeated vaccinations with pneumococcal capsular polysaccharide in splenectomized individuals with special reference to Hodgkin lymphoma. J Intern Med 255:664, 2004

50. Dean RM, Bishop MR: Allogeneic hematopoietic stem cell transplantation for lymphoma. Clin Lymphoma 4:238, 2004

51. Gustavsson A, Osterman B, Cavallin-Stahl E: A systematic overview of radiation therapy effects in Hodgkin lymphoma. Acta Oncol 42:589, 2003

52. Chronowski GM, Wilder RB, Levy LB, et al: Second malignancies after chemotherapy and radiotherapy for Hodgkin disease. Am J Clin Oncol 27:73, 2004

53. Adams MJ, Lipsitz SR, Colan SD, et al: Cardiovascular status in long-term survivors of Hodgkin disease treated with chest radiotherapy. J Clin Oncol 22:3139, 2004

54. Illes A, Biro E, Miltenyi Z, et al: Hypothyroidism and thyroiditis after therapy for Hodgkin's disease. Acta Haematol 109:11, 2003

55. Kaldor JM, Day NE, Clarke EA, et al: Leukemia following Hodgkin disease. N Engl J Med 322:7, 1990

56. Horning SJ, Hoppe RT, Kaplan HS, et al: Female reproductive potential after treatment for Hodgkin disease. N Engl J Med 304:1377, 1981

57. Chapman RM, Sutcliffe SB, Malpas JS: Male gonadal dysfunction in Hodgkin disease: a prospective study. JAMA 245:1323, 1981

58. Anselmo AP, Cartoni C, Bellantuono P, et al: Risk of infertility in patients with Hodgkin disease treated with ABVD vs MOPP vs ABVD/MOPP. Haematologica 75:155, 1990

59. Putterman C, Polliack A: Late cardiovascular and pulmonary complications of therapy in Hodgkin disease: report of three unusual cases, with a review of relevant literature. Leuk Lymphoma 7:109, 1992

60. Fisher SG, Fisher RI: The epidemiology of non-Hodgkin lymphoma. Oncogene 23:6524, 2004

61. Chiu BC, Weisenburger DD: An update of the epidemiology of non-Hodgkin lymphoma. Clin Lymphoma 4:161, 2003

62. Loren AW, Porter DL, Stadtmauer EA, et al: Post-transplant lymphoproliferative disorder: a review. Bone Marrow Transplant 31:145, 2003

63. Poiesz BJ, Papsidero LD, Ehrlich G, et al: Prevalence of HTLV-I–associated T-cell lymphoma. Am J Hematol 66:32, 2001

64. Engels EA, Chatterjee N, Cerhan JR, et al: Hepatitis C virus infection and non-Hodgkin lymphoma: results of the NCI-SEER multi-center case-control study. Int J Cancer 111:76, 2004

65. Wotherspoon AC: Gastric lymphoma of mucosa-associated lymphoid tissue and *Helicobacter pylori*. Annu Rev Med 49:289, 1998

66. Staudt MR, Kanan Y, Jeong JH, et al: The tumor microenvironment controls primary effusion lymphoma growth in vivo. Cancer Res 64:4790, 2004

67. Cerroni L, Zochling N, Putz B, et al: Infection by *Borrelia burgdorferi* and cutaneous B-cell lymphoma. J Cutan Pathol 24:457, 1997

68. Ferreri AJ, Guidoboni M, Ponzoni M, et al: Evidence for an association between *Chlamydia psittaci* and ocular adnexal lymphomas. J Natl Cancer Inst 96:586, 2004

69. Lecuit M, Abachin E, Martin A, et al: Immunoproliferative small intestinal disease associated with *Campylobacter jejuni*. N Engl J Med 350:239, 2004

70. Gascoyne RD, Adomat SA, Krajewski S, et al: Prognostic significance of Bcl-2 protein expression and Bcl-2 gene rearrangement in diffuse aggressive non-Hodgkin lymphoma. Blood 90:244, 1997

71. Wilson WH, Grossbard ML, Pittaluga S, et al: Dose-adjusted EPOCH chemotherapy for untreated large B-cell lymphomas: a pharmacodynamic approach with high efficacy. Blood 99:2685, 2002

72. Ichikawa A, Kinoshita T, Watanabe T, et al: Mutations of the p53 gene as a prognostic factor in aggressive B-cell lymphoma. N Engl J Med 337:529, 1997

73. Krause JR, Shahidi-Asl M: Molecular pathology in the diagnosis and treatment of non-Hodgkin lymphomas. J Cell Mol Med 7:494, 2003

74. Weiss LM, Warnke RA, Sklar J, et al: Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. N Engl J Med 317:1185, 1987

75. Cleary ML, Sklar J: Nucleotide sequence of a t(14;18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. Proc Natl Acad Sci USA 82:7439, 1985

76. Hill ME, MacLennan KA, Cunningham DC, et al: Prognostic significance of BCL-2 expression and bcl-2 major breakpoint region rearrangement in diffuse large cell non-Hodgkin lymphoma: a British National Lymphoma Investigation study. Blood 88:1046, 1996

77. Bosch F, Jares P, Campo E, et al: PRAD-1/cyclin D1 gene overexpression in chronic lymphoproliferative disorders: a highly specific marker of mantle cell lymphoma. Blood 84:2726, 1994

78. Fisher RI, Gaynor ER, Dahlberg S, et al: Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin lymphoma. N

Engl J Med 328:1002, 1993

79. Coiffier B, Lepage E, Briere J, et al: CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med 346:235, 2002

80. Rosenwald A, Wright G, Chan WC, et al: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937, 2002

81. Rosenwald A, Wright G, Leroy K, et al: Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med 198:851, 2003

82. Rosenwald A, Wright G, Wiestner A, et al: The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. Cancer Cell 3:185, 2003

83. Canioni D, Brice P, Lepage E, et al: Bone marrow histological patterns can predict survival of patients with grade 1 or 2 follicular lymphoma: a study from the Groupe d'Etude des Lymphomes Folliculaires. Br J Haematol 126:364, 2004

84. Little RF, Pittaluga S, Grant N, et al: Highly effective treatment of acquired immunodeficiency syndrome–related lymphoma with dose-adjusted EPOCH: impact of antiretroviral therapy suspension and tumor biology. Blood 101:4653, 2003

85. Dumontet C, Drai J, Bienvenu J, et al: Profiles and prognostic values of LDH isoenzymes in patients with non-Hodgkin lymphoma. Leukemia 13:811, 1999

86. van Besien K, Ha CS, Murphy S, et al: Risk factors, treatment, and outcome of central nervous system recurrence in adults with intermediate-grade and immunoblastic lymphoma. Blood 91:1178, 1998

87. Hegde U, Filie A, Little RF, et al: High incidence of occult leptomeningeal disease detected by flow cytometry in newly diagnosed aggressive B-cell lymphomas at risk of central nervous system involvement: the role of flow cytometry versus cytology. Blood 105:496, 2005

88. A predictive model for aggressive non-Hodgkin lymphoma. The International Non-Hodgkin Lymphoma Prognostic Factors Project. N Engl J Med 329:987, 1993

89. Solal-Celigny P, Roy P, Colombat P, et al: Follicular lymphoma international prognostic index. Blood 104:1258, 2004

90. Hans CP, Weisenburger DD, Greiner TC, et al: Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103:275, 2004

91. Sohn SK, Jung JT, Kim DH, et al: Prognostic significance of bcl-2, bax, and p53 expression in diffuse large B-cell lymphoma. Am J Hematol 73:101, 2003

92. Jerkeman M, Anderson H, Dictor M, et al: Assessment of biological prognostic factors provides clinically relevant information in patients with diffuse large B-cell lymphoma: a Nordic Lymphoma Group study. Ann Hematol 83:414, 2004

93. Wilson W, Pittaluga S, O'Connor P: Rituximab may overcome BCL-2–associated chemotherapy resistance in untreated diffuse large B-cell lymphomas (abstr). Blood 98:A-1447, 2001

94. Mounier N, Briere J, Gisselbrecht C, et al: Rituximab plus CHOP (R-CHOP) overcomes bcl-2–associated resistance to chemotherapy in elderly patients with diffuse large B-cell lymphoma (DLBCL). Blood 101:4279, 2003

95. Staudt LM: Molecular diagnosis of the hematologic cancers. N Engl J Med 348:1777, 2003

96. Alizadeh AA, Eisen MB, Davis RE, et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503, 2000

97. Lossos IS, Czerwinski DK, Alizadeh AA, et al: Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 350:1828, 2004

98. Dave S, Wright G, Tan B: A molecular predictor of survival following diagnosis of follicular lymphoma. Blood 102:617a, 2003

99. Rosenwald A, Alizadeh AA, Widhopf G, et al: Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. J Exp Med 194:1639, 2001

100. Wiestner A, Rosenwald A, Barry TS, et al: ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood 101:4944, 2003

101. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. Blood 89:3909, 1997

102. Seymour JF: New treatment approaches to indolent non-Hodgkin lymphoma. Semin Oncol 31:27, 2004

103. Advani R, Rosenberg SA, Horning SJ: Stage I and II follicular non-Hodgkin lymphoma: long-term follow-up of no initial therapy. J Clin Oncol 22:1454, 2004

104. Ott OJ, Rodel C, Gramatzki M, et al: Radiotherapy for stage I–III nodal low-grade non-Hodgkin lymphoma. Strahlenther Onkol 179:694, 2003

105. Lenz G, Hiddemann W, Dreyling M: The role of fludarabine in the treatment of follicular and mantle cell lymphoma. Cancer 101:883, 2004

106. Santini G, Chisesi T, Nati S, et al: Fludarabine, cyclophosphamide and mitoxantrone for untreated follicular lymphoma: a report from the non-Hodgkin lymphoma co-operative study group. Leuk Lymphoma 45:1141, 2004

107. Colombat P, Salles G, Brousse N, et al: Rituximab (anti-CD20 monoclonal antibody) as single first-line therapy for patients with follicular lymphoma with a low tumor burden: clinical and molecular evaluation. Blood 97:101, 2001

108. Davis TA, White CA, Grillo-Lopez AJ, et al: Single-agent monoclonal antibody efficacy in bulky non-Hodgkin lymphoma: results of a phase II trial of rituximab. J Clin Oncol 17:1851, 1999

109. Davis TA, Grillo-Lopez AJ, White CA, et al: Rituximab anti-CD20 monoclonal antibody therapy in non-Hodgkin lymphoma: safety and efficacy of re-treatment. J Clin Oncol 18:3135, 2000

110. Czuczman MS, Weaver R, Alkuzweny B, et al: Prolonged clinical and molecular remission in patients with low-grade or follicular non-Hodgkin lymphoma treated with rituximab plus CHOP chemotherapy: 9-year follow-up. J Clin Oncol 22:4659, 2004

111. Czuczman MS, Fallon A, Mohr A, et al: Rituximab in combination with CHOP or fludarabine in low-grade lymphoma. Semin Oncol 29:36, 2002

112. Hiddemann W, Forstpointer R, Kneba M, et al: The addition of rituximab to combination chemotherapy CHOP has a long lasting impact on subsequent treatment in remission in follicular lymphoma but not in mantle cell lymphoma: results of two prospective randomized studies of the German Low Grade Lymphoma Study Group (GLSG). Blood 104:161, 2004

113. Hainsworth JD, Litchy S, Burris HA 3rd, et al: Rituximab as first-line and maintenance therapy for patients with indolent non-Hodgkin lymphoma. J Clin Oncol 20:4261, 2002

114. Witzig TE, Gordon LI, Cabanillas F, et al: Randomized controlled trial of yttrium-90–labeled ibritumomab tiuxetan radioimmunotherapy versus rituximab immunotherapy for patients with relapsed or refractory low-grade, follicular, or transformed B-cell non-Hodgkin lymphoma. J Clin Oncol 20:2453, 2002

115. Gordon LI, Witzig T, Molina A, et al: Yttrium 90–labeled ibritumomab tiuxetan radioimmunotherapy produces high response rates and durable remissions in patients with previously treated B-cell lymphoma. Clin Lymphoma 5:98, 2004

116. Kaminski MS, Tuck M, Estes J, et al: ^{131}I-tositumomab therapy as initial treatment for follicular lymphoma. N Engl J Med 352:441, 2005

117. van Besien K, Loberiza FR Jr, Bajorunaite R, et al: Comparison of autologous and allogeneic hematopoietic stem cell transplantation for follicular lymphoma. Blood 102:3521, 2003

118. McCarthy H, Ottensmeier CH, Hamblin TJ, et al: Anti-idiotype vaccines. Br J Haematol 123:770, 2003

119. Kwak LW: Translational development of active immunotherapy for hematologic malignancies. Semin Oncol 30:17, 2003

120. Chanan-Khan A, Czuczman MS: Bcl-2 antisense therapy in B-cell malignant proliferative disorders. Curr Treat Options Oncol 5:261, 2004

121. Wotherspoon AC, Doglioni C, Diss TC, et al: Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue type after eradication of *Helicobacter pylori*. Lancet 342:575, 1993

122. Cavalli F, Isaacson PG, Gascoyne RD, et al: MALT Lymphomas. Hematology (Am Soc Hematol Educ Program):241, 2001

123. Nuckel H, Meller D, Steuhl KP, et al: Anti-CD20 monoclonal antibody therapy in relapsed MALT lymphoma of the conjunctiva. Eur J Haematol 73:258, 2004

124. Lenz G, Dreyling M, Hiddemann W: Mantle cell lymphoma: established therapeutic options and future directions. Ann Hematol 83:71, 2004

125. Meusers P, Engelhard M, Bartels H, et al: Multicentre randomized therapeutic trial for advanced centrocytic lymphoma: anthracycline does not improve the prognosis. Hematol Oncol 7:365, 1989

126. Hagemeister FB: Mantle cell lymphoma: non-myeloablative versus dose-intensive therapy. Leuk Lymphoma 44(suppl 3):S69, 2003

127. Wilson W, Neelapu S, Rosenwald A, et al: Idiotype vaccine and dose-adjusted EPOCH-rituximab treatment in untreated mantle cell lymphoma: preliminary report on clinical outcome and analysis of immune response. Blood 102:358, 2003

128. Jacobsen E, Freedman A: An update on the role of high-dose therapy with autologous or allogeneic stem cell transplantation in mantle cell lymphoma. Curr Opin Oncol 16:106, 2004

129. Goy A, Younes A, McLaughlin P, et al: Phase II study of proteasome inhibitor bortezomib in relapsed or refractory B-cell non-Hodgkin's lymphoma. J Clin Oncol 23:667, 2004

130. O'Connor OA, Wright J, Moskowitz C, et al: Phase II clinical experience with the novel proteasome inhibitor bortezomib in patients with indolent non-Hodgkin's lymphoma and mantle cell lymphoma. J Clin Oncol 23:676, 2005

131. Spicer J, Smith P, Maclennan K, et al: Long-term follow-up of patients treated with radiotherapy alone for early-stage histologically aggressive non-Hodgkin lymphoma. Br J Cancer 90:1151, 2004

132. Miller TP, Dahlberg S, Cassady JR, et al: Chemotherapy alone compared with chemotherapy plus radiotherapy for localized intermediate- and high-grade non-Hodgkin lymphoma. N Engl J Med 339:21, 1998

133. Miller T, Sahlberg S, Cassady JR, et al: CHOP alone compared to CHOP plus radiotherapy for early stage aggressive non-Hodgkin's lymphomas: update of the Southwest Oncology Group (SWOG) randomized trial. Blood 98:724, 2001

134. Horning SJ, Weller E, Kim K, et al: Chemotherapy with or without radiotherapy in limited-stage diffuse aggressive non-Hodgkin lymphoma: Eastern Cooperative Oncology Group study 1484. J Clin Oncol 22:3032, 2004

135. Reyes FL, Munck JN, Morel P, et al: Superiority of chemotherapy alone with the ACVBP regimen over treatment with 3 cycles of CHOP plus radiation in low risk localized aggressive lymphoma: the LNH93-1 GELA study. Proc Am Soc Hematol 2002, Abstract 343

136. Zinzani PL, Martelli M, Magagnoli M, et al: Treatment and clinical management of primary mediastinal large B-cell lymphoma with sclerosis: MACOP-B regimen and mediastinal radiotherapy monitored by (67)gallium scan in 50 patients. Blood 94:3289, 1999

137. Messori A, Vaiani M, Trippoli S, et al: Survival in patients with intermediate or high grade non-Hodgkin lymphoma: meta-analysis of randomized studies comparing third generation regimens with CHOP. Br J Cancer 84:303, 2001

138. Tilly H, Lepage E, Coiffier B, et al: Intensive conventional chemotherapy (ACVBP

regimen) compared with standard CHOP for poor-prognosis aggressive non-Hodgkin lymphoma. Blood 102:4284, 2003

139. Wilson WH, Pittaluga S, Gutierrez M, et al: Dose-adjusted EPOCH-rituximab in untreated diffuse large B-cell lymphoma: benefit of rituximab appears restricted to tumors harboring anti-apoptotic mechanisms. Blood 102:11, 2003

140. Pfreundschuh M, Trumper L, Kloess M, et al: Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of young patients with good-prognosis (normal LDH) aggressive lymphomas: results of the NHL-B1 trial of the DSHNHL. Blood 104:626, 2004

141. Pfreundschuh M, Trumper L, Kloess M, et al: Two-weekly or 3-weekly CHOP chemotherapy with or without etoposide for the treatment of elderly patients with aggressive lymphomas: results of the NHL-B2 trial of the DSHNHL. Blood 104:634, 2004

142. Shipp MA, Abeloff MD, Antman KH, et al: International consensus conference on high-dose therapy with hematopoietic stem cell transplantation in aggressive non-Hodgkin lymphomas: report of the jury. J Clin Oncol 17:423, 1999

143. Gisselbrecht C, Lepage E, Molina T, et al: Shortened first-line high-dose chemotherapy for patients with poor-prognosis aggressive lymphoma. J Clin Oncol 20:2472, 2002

144. Milpied N, Deconinck E, Gaillard F, et al: Initial treatment of aggressive lymphoma with high-dose chemotherapy and autologous stem-cell support. N Engl J Med 350:1287, 2004

145. Seropian S, Bahceci E, Cooper DL: Allogeneic peripheral blood stem cell transplantation for high-risk non-Hodgkin lymphoma. Bone Marrow Transplant 32:763, 2003

146. Batchelor T, Carson K, O'Neill A, et al: Treatment of primary CNS lymphoma with methotrexate and deferred radiotherapy: a report of NABTT 96-07. J Clin Oncol 21:1044, 2003

147. Sandor V, Stark-Vancs V, Pearson D, et al: Phase II trial of chemotherapy alone for primary CNS and intraocular lymphoma. J Clin Oncol 16:3000, 1998

148. Pels H, Schmidt-Wolf IG, Glasmacher A, et al: Primary central nervous system lymphoma: results of a pilot and phase II study of systemic and intraventricular chemotherapy with deferred radiotherapy. J Clin Oncol 21:4489, 2003

149. Magrath I, Adde M, Shad A, et al: Adults and children with small non-cleaved-cell lymphoma have a similar excellent outcome when treated with the same chemotherapy regimen. J Clin Oncol 14:925, 1996

150. Reiter A, Schrappe M, Parwaresch R, et al: Non-Hodgkin lymphomas of childhood and adolescence: results of a treatment stratified for biologic subtypes and stage: a report of the Berlin-Frankfurt-Munster Group. J Clin Oncol 13:359, 1995

151. Dal Maso L, Franceschi S: Epidemiology of non-Hodgkin lymphomas and other haemolymphopoietic neoplasms in people with AIDS. Lancet Oncol 4:110, 2003

152. Little RF, Wilson WH: Update on the pathogenesis, diagnosis, and therapy of AIDS-related lymphoma. Curr Infect Dis Rep 5:176, 2003

153. Ratner L, Lee J, Tang S, et al: Chemotherapy for human immunodeficiency virus–associated non-Hodgkin lymphoma in combination with highly active antiretroviral therapy. J Clin Oncol 19:2171, 2001

154. Wilson WH, Kingma DW, Raffeld M, et al: Association of lymphomatoid granulomatosis with Epstein-Barr viral infection of B lymphocytes and response to interferon-alpha 2b. Blood 87:4531, 1996

155. Gascoyne RD, Aoun P, Wu D, et al: Prognostic significance of anaplastic lymphoma kinase (ALK) protein expression in adults with anaplastic large cell lymphoma. Blood 93:3913, 1999

156. Ribrag V, Ell Hajj M, Janot F, et al: Early locoregional high-dose radiotherapy is associated with long-term disease control in localized primary angiocentric lymphoma of the nose and nasopharynx. Leukemia 15:1123, 2001

157. Foss F: Mycosis fungoides and the Sezary syndrome. Curr Opin Oncol 16:421, 2004

158. Piekarz RL, Robey RW, Zhan Z, et al: T-cell lymphoma as a model for the use of histone deacetylase inhibitors in cancer therapy: impact of depsipeptide on molecular markers, therapeutic targets, and mechanisms of resistance. Blood 103:4636, 2004

159. Soligo D, Ibatici A, Berti E, et al: Treatment of advanced mycosis fungoides by allogeneic stem-cell transplantation with a nonmyeloablative regimen. Bone Marrow Transplant 31:663, 2003

160. Falini B, Pileri S, Zinzani PL, et al: ALK+ lymphoma: clinico-pathological findings and outcome. Blood 93:2697, 1999

161. Dogan A, Attygalle AD, Kyriakou C: Angioimmunoblastic T-cell lymphoma. Br J Haematol 121:681, 2003

162. Coiffier B, Brousse N, Peuchmaur M, et al: Peripheral T-cell lymphomas have a worse prognosis than B-cell lymphomas: a prospective study of 361 immunophenotyped patients treated with the LNH-84 regimen. The GELA (Groupe d'Etude des Lymphomes Agressives). Ann Oncol 1:45, 1990

163. Rodriguez J, Caballero MD, Gutierrez A, et al: High-dose chemotherapy and autologous stem cell transplantation in peripheral T-cell lymphoma: the GEL-TAMO experience. Ann Oncol 14:1768, 2003

164. Song KW, Mollee P, Keating A, et al: Autologous stem cell transplant for relapsed and refractory peripheral T-cell lymphoma: variable outcome according to pathological subtype. Br J Haematol 120:978, 2003

165. Corradini P, Dodero A, Zallio F, et al: Graft-versus-lymphoma effect in relapsed peripheral T-cell non-Hodgkin lymphomas after reduced-intensity conditioning followed by allogeneic transplantation of hematopoietic cells. J Clin Oncol 22:2172, 2004

166. Tilly H, Rossi A, Stamatoullas A, et al: Prognostic value of chromosomal abnormalities in follicular lymphoma. Blood 84:1043, 1994

167. Knauf WU, Knuutila S, Zeigmeister B, et al: Trisomy 12 in B-cell chronic lymphocytic leukemia: correlation with advanced disease, atypical morphology, high levels of sCD25, and with refractoriness to treatment. Leuk Lymphoma 19:289, 1995

168. Dohner H, Stilgenbauer S, Dohner K, et al: Chromosome aberrations in B-cell chronic lymphocytic leukemia: reassessment based on molecular cytogenetic analysis. J Mol Med 77:266, 1999

169. Raty R, Franssila K, Jansson SE, et al: Predictive factors for blastoid transformation in the common variant of mantle cell lymphoma. Eur J Cancer 39:321, 2003

170. Ueda C, Akasaka T, Ohno H: Non-immunoglobulin/BCL6 gene fusion in diffuse large B-cell lymphoma: prognostic implications. Leuk Lymphoma 43:1375, 2002

XII ONCOLOGIC EMERGENCIES

Robert W. Carlson, m.d., f.a.c.p.

Malignant disease and its treatment may produce a variety of complications. Many of these complications are relatively specific to the disease (e.g., leukostasis in acute myeloid leukemia) or to a class of chemotherapeutic agents (e.g., chronic cardiomyopathy with the anthracyclines). However, life-threatening complications associated with malignancy are common to a variety of tumor types; in addition, many cytotoxic chemotherapeutic agents can produce potentially fatal toxicities, depending on the agent and its dosing schedule.

Cardiovascular Emergencies

pericardial disease and tamponade

Neoplastic pericardial disease is usually associated with advanced lung cancer, breast cancer, leukemia, or lymphoma, with pericardial involvement occurring either by direct extension or by spread via the mediastinal lymphatic vessels.[1,2] Pericardial disease with or without tamponade may also be caused by chemotherapeutic or radiotherapeutic toxicity. Patients who have neoplastic involvement of the visceral pericardium and those who receive mediastinal radiation therapy are at risk for effusive-constrictive pericarditis, a syndrome that is caused by pericardial effusion combined with constriction of the heart by the visceral pericardium. Patients who receive chemotherapy with anthracyclines or cyclophosphamide are also at risk for pericardial disease with tamponade, which may present as acute pericarditis or myocarditis syndromes. These syndromes occur primarily with very high dose chemotherapy regimens.

Diagnosis

The most common presenting symptoms of neoplastic pericardial disease with tamponade are progressive dyspnea, nonspecific chest discomfort, and cough.[1] Arterial blood pressure and heart sounds may be normal. Pericardial friction rubs frequently are not present. Although central venous pressure is always elevated in patients with tamponade, venous hypertension may be detected clinically in only about one half of cases because of anatomic variations in vasculature and body habitus and failure to observe venous pulsations.

A paradoxical pulse is the single most specific sign of pericardial tamponade. Because the signs and symptoms of pericardial tamponade in cancer patients are nonspecific, pericardial tamponade should be considered in any cancer patient who has dyspnea, nonspecific chest pain, elevations in jugular venous pressure, new and unexplained cardiomegaly on chest x-ray, or a paradoxical pulse.[3] Electrocardiography is usually not helpful in the diagnosis of malignant pericardial disease.

Echocardiography is the single most valuable noninvasive method of evaluating pericardial disease or tamponade in cancer patients. Pericardial effusions secondary to neoplastic disease usually are large and do not produce internal echoes. The presence of tamponade is indicated by early diastolic collapse of the right atrial or ventricular wall or by Doppler echocardiography–documented changes in the respiratory variation of blood flow velocity through the tricuspid or mitral valves.

Measurements of pericardial fluid pressure during pericardiocentesis may help differentiate pericardial tamponade from superior vena cava (SVC) syndrome. In tamponade, pericardial fluid pressure is elevated initially but returns to normal as the fluid is withdrawn, whereas in SVC syndrome, pericardial fluid pressure is normal and jugular venous pressure is unaffected by pericardiocentesis. In patients with effusive-constrictive pericarditis, pericardiocentesis produces only partial improvement of the hemodynamic abnormality.

Treatment

Treatment of pericardial tamponade includes pericardiocentesis, pericardial window formation, or pericardiectomy. Echocardiographically guided pericardiocentesis with concurrent catheterization of the right side of the heart allows for definitive diagnosis of cardiac tamponade, removal of pericardial fluid to relieve the tamponade, cytologic and microbiologic analysis of the fluid, and placement of an intrapericardial catheter to prevent fluid reaccumulation. Frequently, catheter drainage for several days results in a decreased rate of pericardial fluid collection and may provide definitive long-term benefit.[4] Intrapericardial instillation of bleomycin or tetracycline-doxycycline is a safe and effective method of sclerosing the pericardial space to prevent the recurrence of tamponade.[5,6] Patients in whom effusions rapidly recur may benefit from a pericardial window,[7] and those who have effusive-constrictive pericarditis may require partial pericardiectomy.[8]

Systemic chemotherapy or hormone therapy may be beneficial. Limited cardiac radiation therapy may also provide palliation, especially in patients with radiosensitive tumors. However, the survival of patients with cancer and secondary pericardial involvement is usually short because of the systemic nature of their disease [*see 1:XIII Diseases of the Pericardium, Cardiac Tumors, and Cardiac Trauma*].

superior vena cava syndrome

Extrinsic compression or intrinsic obstruction of the SVC [*see Figure 1*] may result in elevated venous pressures in the upper extremities, head, and neck and in increased intracranial pressure, soft tissue edema, venous distention, and venous collateral formation. Variations in the clinical presentation of SVC syndrome may occur if compression or obstruction involves a large upper mediastinal vein or a low cervical vein. The syndrome may develop rapidly over a period of days or slowly over a period of many months.

The most common cause of SVC syndrome is extrinsic compression of the thin-walled, low-pressure SVC by a malignant mediastinal mass, such as a bronchogenic carcinoma—especially small cell lung cancer [*see Figure 2*]—or non-Hodgkin lymphoma. Other malignant causes of SVC syndrome are thymic tumors, mediastinal germ cell tumors, and metastatic carcinoma. The differential diagnosis includes SVC thrombosis (especially in patients with central venous catheters), goiter, mediastinal fibrosis, tuberculous mediastinitis, histoplasmosis, and ascending aortic aneurysm.

In the past, SVC syndrome was considered an oncologic emergency that required the immediate initiation of mediastinal radiation therapy. Radiation therapy was thought to be neces-

sary for three reasons: to alleviate increased intracranial pressures because lung cancer was presumed to be the most likely diagnosis, and because it was erroneously believed that increased venous pressures would make diagnostic procedures hazardous. It is now recognized that in adults, SVC syndrome is usually not a true emergency and that a histologic diagnosis should be quickly established and treatment promptly initiated. Emergency treatment with endovascular stenting or mediastinal radiation therapy before a histologic diagnosis is established is warranted in children and in adults who have mental-status alteration, other life-threatening manifestations of increased intracranial pressure, cardiovascular collapse, or upper airway obstruction.

Diagnosis

SVC syndrome should be suspected in any patient who has a sensation of fullness in the face, plethora, dyspnea, facial swelling, distended neck veins, or venous collaterals in the neck or chest. In addition, many patients with SVC syndrome will acknowledge, if asked, that they have been sleeping fitfully and have been having nightmares. Headache, disturbance of vision, mental-status alteration, cough, and chest pain may also be present. In almost all patients with SVC syndrome, a chest x-ray demonstrates widening of the superior mediastinum, a right hilar or an upper lobe mass, or pleural effusion. Computed tomography of the chest is the single most useful diagnostic tool. It confirms the clinical diagnosis, localizes the abnormality, provides information regarding biopsy sites, and identifies coexisting thrombosis. The diagnosis of SVC syndrome should be assumed if the patient has clinical manifestations of the syndrome and there is radiographic evidence of mediastinal disease.

In a patient without a history of cancer, the evaluation should focus on identifing any abnormality for which a biopsy can be performed to provide a histologic diagnosis. Procedures include sputum cytology, fine-needle aspiration of pathologic lymphadenopathy, cytologic examination of pleural fluid, and bone marrow biopsy and aspiration. Because bronchogenic carcinoma is the most common cause of SVC syndrome, bronchoscopy is a high-yield procedure. Bronchoscopy can be performed safely in most patients with SVC syndrome. The combination of sputum cytology, bronchoscopy, and biopsy of palpable lymph nodes provides a definitive histologic diagnosis in approximately 70% of patients.[9] Although the safety of biopsy procedures in patients with SVC syndrome has been well established,[10] it is prudent to take biopsy samples from sites of disease outside the region of elevated venous pressures whenever possible.

Treatment

Treatment of SVC syndrome caused by a malignancy is evolving, with the increasing use of endovascular stents to rapidly reverse the hemodynamic abnormality followed by combination chemotherapy or radiation therapy to treat the underlying malignancy. Endovascular stenting provides rapid relief of symptoms and has a high success rate and low morbidity.[11-14] Combination chemotherapy is often effective as the primary therapy for SVC syndrome in patients with small cell lung cancer or lymphoma; in most patients, this treatment produces a positive disease response and improvement in the SVC syndrome.[15] Because there is increased risk of extravasation from veins with elevated pressure, chemotherapeutic agents should be infused into a vein in the lower extremity or a femoral vein. In patients with non–small cell lung cancer or other tumors that are relatively resistant to chemotherapy, endovascular stenting is generally followed with mediastinal radiation therapy.

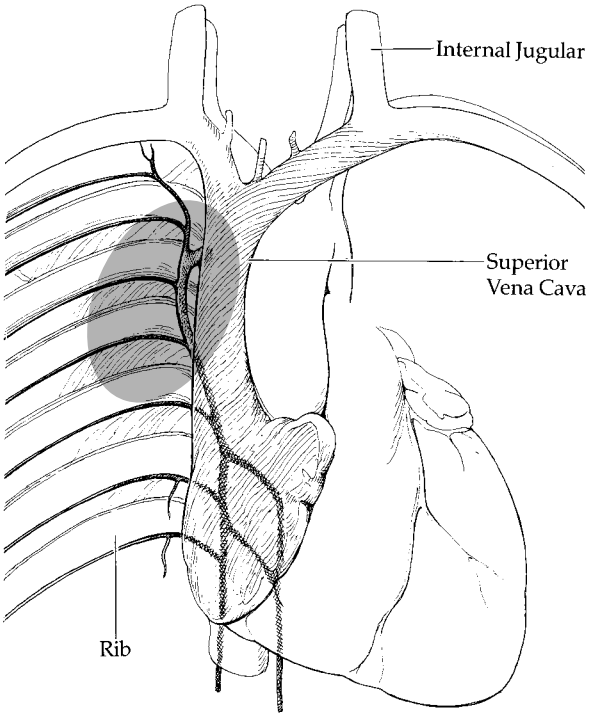


Figure 1 **Obstruction of the superior vena cava in patients with malignant disease is caused by compression of the thin-walled vessel by tumorous lymph nodes or by direct invasion of the wall of the vessel by the tumor. The shaded region represents the most common site of obstruction.**

SVC syndrome often results in SVC thrombosis, and thus, some clinicians routinely prescribe anticoagulants for these patients; however, except in patients with demonstrable thrombosis or endovascular stents in place, there are no convincing data to support routine use of these agents, and anticoagulation has been associated with a 10% fatality rate secondary to intracranial hemorrhage.[16] SVC thrombosis should be suspected in the rare patient who fails to respond to chemotherapy or radiation therapy. Treatment with anticoagulants or fibrinolytics is probably appropriate in patients with documented thrombosis. Patients with stridor or significant airway compromise should be treated with glucocorticoids. Endotracheal intubation should be considered for acute airway management.

In 10% to 20% of patients, SVC syndrome will recur after radiation therapy and chemotherapy. If the patient does not already have an intravascular stent, insertion of such a device will often provide palliation.[17] The overall prognosis in patients with SVC syndrome secondary to malignancy is determined by the malignancy, not by the presence of the syndrome per se.

Hematologic Emergencies

DISSEMINATED INTRAVASCULAR COAGULATION

Disseminated intravascular coagulation (DIC) involves systemic activation of the coagulation process by tissue factor, impaired fibrinolysis, and deficiency in the physiologic anticoagu-

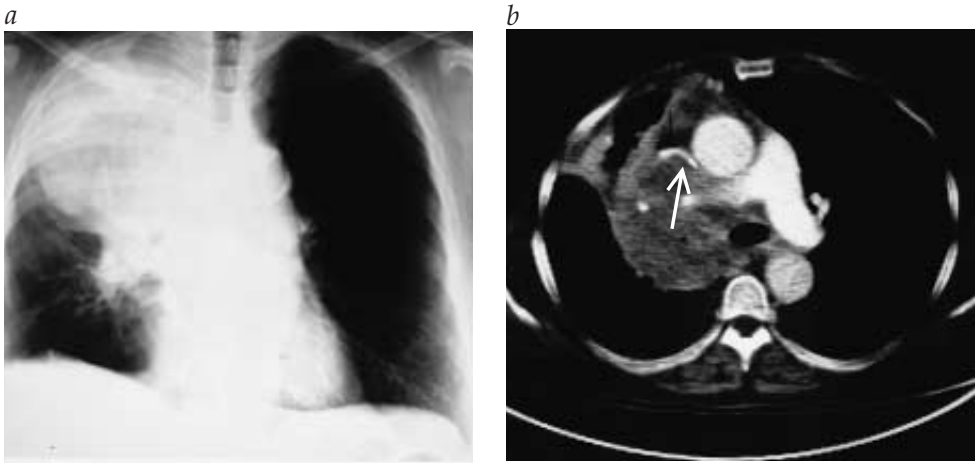


Figure 2 **(*a*) Shown is a chest x-ray of a patient with small cell lung cancer and superior vena cava syndrome. (*b*) Contrast-enhanced CT scan in the same patient is also shown.[103] Note that the thin-walled, contrast-enhanced superior vena cava is compressed by the tumor mass to a crescentic shape (arrow).**

lant system.[18,19] This leads to consumption of platelets and coagulation factors, which causes bleeding, and to deposition of fibrin, which causes microvascular thrombosis. In addition to malignancy, common causes of DIC include infection, obstetric complications, and trauma.[18] The fibrinolytic system usually is activated secondarily; it is rarely activated primarily (primary fibrinolysis) without associated DIC. Patients with malignancy often have asymptomatic abnormalities of the coagulation system. These abnormalities include increased levels of factor VIII, fibrinogen, and platelets or evidence of increased fibrinolysis, as measured by increased levels of fibrin monomer, fibrinopeptide A, and fibrin degradation products.[20] Patients with adenocarcinoma (especially of the pancreas or prostate) or promyelocytic leukemia are particularly at risk. DIC may be either acute or chronic, and the coagulopathy may vary in severity, ranging from difficult to detect—even in the laboratory—to a fulminant bleeding diathesis. Thromboembolic manifestations are common.

Chronic DIC

Low-grade chronic DIC occurs in most patients with disseminated malignancy. The body typically compensates well, and laboratory measurements of coagulation may be normal or near normal. Some patients with low-grade chronic DIC, however, experience deep vein thrombosis, Trousseau syndrome, nonbacterial thrombotic endocarditis, or embolism without bleeding.[20] Patients who have thrombosis and no evidence of bleeding should immediately be started on long-term anticoagulant therapy with heparin (either unfractionated or low-molecular-weight heparin). Warfarin provides inadequate protection against recurrent thrombosis in patients with chronic DIC.

Acute DIC

Diagnosis Patients with acute DIC usually have clinical evidence of consumption coagulopathy, a low platelet count, and activation of clotting factors, which cause prolongation of the prothrombin, partial thromboplastin, and thrombin times; and elevated levels of fibrin degradation products. Fibrinogen levels may be decreased because of consumption or may be normal or elevated as an acute-phase reactant despite consumption. Common signs of acute DIC include bleeding from surgical wounds, venipuncture sites, the nose, the mouth, and the gastrointestinal tract; ecchymoses; petechiae; purpura; and stroke. In patients with malignancy, simultaneous blood loss from three or more sites usually is an indication of life-threatening acute DIC. The bleeding associated with DIC may be fulminant and may cause shock, end-organ dysfunction, and death.

Treatment Treatment of acute DIC includes elimination of the underlying cause, such as infection or malignancy, whenever possible. DIC in patients with promyelocytic leukemia may resolve with treatment with all-*trans*-retinoic acid.[18,21-23] Replacement of consumed platelets and clotting factors usually is appropriate, and red cell transfusions should be given to patients who have significant blood loss. Despite evidence suggesting a role for supraphysiologic doses of antithrombin III, a randomized trial of high-dose antithrombin III in patients with severe sepsis did not document a reduction in mortality.[22,24,25] The use of heparin remains controversial.[19,23,24] Heparin may be valuable in inhibiting DIC and thereby in maintaining appropriate levels of clotting factors, but it carries a significant risk for increasing bleeding. When heparin is used, the lowest dosage of unfractionated heparin (300 to 500 U/hr) or low-molecular-weight heparin that maintains the platelet count (> 50,000 U/μl) and clotting factor level (a fibrinogen level > 150 mg/dl) should be given. In patients with primary fibrinolysis, treatment consists of the vigorous replacement of consumed clotting factors with cryoprecipitate or fresh frozen plasma. Early evidence suggests that the use of activated protein C may also be of benefit.[26]

BLEEDING WITH THROMBOCYTOPENIA

In patients with cancer, thrombocytopenia is most commonly caused by the underproduction of platelets secondary either to bone marrow infiltration by tumor or to the toxicity of chemotherapy or radiation therapy. Increased destruction of peripheral blood elements, as occurs in hypersplenism, DIC, and immune-mediated disorders, is also sometimes observed. In patients who are receiving myelosuppressive chemotherapy and in those with thrombocytopenia, the use of agents that inhibit platelet function (e.g., aspirin) should be specifically avoided.

Bleeding time increases when platelet counts fall below 100,000/μl, and the risk of spontaneous bleeding significantly increases when platelet counts fall below 10,000/μl.[27] The American Society of Clinical Oncology (ASCO) has developed guidelines for the use of prophylactic platelet transfusions.[28] In patients with leukemia, hematopoietic cell transplantation, or solid tumors, the ASCO guidelines recommend prophylactic platelet transfusions for chemotherapy-related thrombocytopenia at a platelet count threshold of 10,000/μl. Higher thresholds for

transfusion are appropriate for patients with hemorrhage, high fever, or coagulation abnormalities or when invasive procedures are anticipated. Alloimmunization may be reduced or prevented by the use of blood products leukoreduced by filtration. To prevent allosensitization, prophylactic platelet transfusions should not be performed in patients with increased destruction of peripheral blood elements.

Treatment of bleeding in patients with thrombocytopenia from underproduction of platelets consists of platelet and red cell transfusions and local control at the site of bleeding. Patients who become sensitized to transfused platelets may benefit from HLA-matched, single-donor platelet transfusions. Underlying DIC should also be considered and, if identified, should be treated as noted in the discussion of DIC (see above). In patients undergoing chemotherapy for solid tumors who are at high risk for thrombocytopenia, the use of recombinant human interleukin-11 (oprelvekin) has been shown to reduce thrombocytopenia and the need for platelet transfusions but is associated with substantial toxicity.[29,30]

THROMBOTIC MICROANGIOPATHY

Thrombotic microangiopathy is a rare syndrome that includes microangiopathic hemolytic anemia, thrombocytopenia, uremia, neurologic dysfunction, and fever. The syndrome may present as thrombotic thrombocytopenic purpura (TTP) or the hemolytic-uremic syndrome (HUS)[31]; TTP is typically diagnosed if neurologic symptoms predominate, whereas HUS is diagnosed if uremia predominates. Evidence suggests that TTP and HUS represent distinct histopathologic entities.[32] Both TTP and HUS involve the intravascular hemolysis of red cells. TTP is usually fatal if not treated. Thrombotic microangiopathy occurs in a variety of situations, including malignancy (e.g., adenocarcinomas and squamous cell carcinomas) and chemotherapy.

The diagnosis of thrombotic microangiography is suggested by anemia and the presence of schistocytes on peripheral blood smears. The severity of disease varies widely, ranging from minimal anemia with few schistocytes to profound anemia with rapid hemolysis and large numbers of schistocytes. The differential diagnosis includes sepsis with DIC; many patients also have associated DIC. Patients with gastric carcinoma are particularly at risk.

HUS may occur in patients who have no evidence of active disease. HUS is most commonly observed after chemotherapy with mitomycin but is also seen after treatment with bleomycin, cisplatin, dacarbazine, fluorouracil, lomustine, vinca alkaloids, gemcitabine, and high-dose chemotherapy with autologous stem cell support.[31,33] In one series, the syndrome developed in approximately 4% of mitomycin-treated patients.[33]

Mortality from thrombotic microangiopathy is high, and no consistently effective therapy is available. The mainstay of therapy is immediate plasma exchange with either fresh frozen plasma or cryosupernatant plasma.[31,34-36] Prolonged courses of plasma exchange of one calculated plasma volume a day may be required. The use of extracorporeal immunoadsorption of plasma has produced improvement in 45% of treated patients.[37] Other therapies that have been utilized in patients with refractory or recurrent thrombotic microangiopathy include vincristine, cyclophosphamide, cyclosporine, prednisone, and rituximab.[38-40]

FEVER AND NEUTROPENIA

Patients with cancer may experience transient or protracted periods of neutropenia related to the disease or its treatment. Patients with neutropenia (defined as counts of polymorphonuclear leukocytes plus bands of less than 500/μl) experience an increase in the frequency and severity of a variety of infections. The increased susceptibility to infection is related both to the level and to the duration of the neutropenia. Patients are at particular risk for life-threatening infection if neutropenia is profound (counts of polymorphonuclear leukocytes plus bands below 100/μl). Other factors that place patients with cancer at increased risk for infection include deficient phagocytic function secondary to cytotoxic therapy, the breakdown of skin and mucosal barriers secondary to chemotherapy, the presence of vascular access devices, and the high frequency of invasive procedures. Although some studies have demonstrated that patients with fever and neutropenia who are at low risk for infection may be managed in the outpatient setting, most patients should be managed in the hospital until further confirmatory studies are available.[41-43]

Prevention

The optimal dosages and schedules of antitumor chemotherapy for a number of malignancies produce neutropenia. Methods of minimizing infection in patients with protracted neutropenia include the maintenance of good hygiene, with aggressive treatment of abrasions and mucositis; careful hand washing by medical personnel between patient contacts; not placing fresh flowers or plants in the patient's room; the use of low-microbial diets; the use of prophylactic antibiotics; and the use of colony-stimulating factors. However, prophylactic antibiotic treatment is of limited usefulness, is expensive, and increases the risk of drug-resistant infection. In addition, granulocyte colony-stimulating factor (G-CSF) (e.g., filgrastim, pegfilgrastim) and granulocyte-macrophage colony-stimulating factor (GM-CSF) can decrease the duration of chemotherapy-induced neutropenia [*see* Colony-Stimulating Factors, *below*].[44-47]

Diagnosis

In patients with neutropenia, serious infection should be presumed after a single temperature measurement of 38.3° C (101 °F) or higher or after recurrent temperature measurements of 38.0° C (100.4° F) or higher.[48] In a neutropenic patient, even life-threatening infections may not be apparent on initial evaluation. Localizing signs and symptoms of infection may be absent. When present, the manifestations of infection may be subtle because of an inadequate inflammatory response; repeated daily examinations are therefore essential. Physical examination should focus on common sites of infection, including the sinuses, ears, mouth, oropharynx, skin, chest, abdomen, perianal region, and catheter sites. Cultures should be obtained from the blood and urine; in patients undergoing long-term venous catheterization, blood cultures from each catheter lumen should be obtained at the same time as peripheral venous blood cultures. Cultures of sputum, the pharynx, stool, pleural or peritoneal fluid, cerebrospinal fluid, and catheter sites should be obtained if signs or symptoms of localizing infection are present. A chest radiograph should be obtained, although it may be normal in neutropenic patients who have pneumonia, because the development of an infiltrate requires the presence of neutrophils.

Despite careful and thorough evaluation, an infectious source is identified in only 30% to 40% of patients with fever and neutropenia. Gram-positive organisms are now the most commonly identified organisms, particularly in patients undergoing long-term catheterization; the most common organisms are *Staphylo-*

coccus epidermidis, *Streptococcus* species, and *S. aureus*. Gram-negative organisms, especially *Escherichia coli*, *Klebsiella* species, and *Pseudomonas aeruginosa*, are also commonly identified.[48] Neutropenic hosts also have increased susceptibility to fungal infection, most commonly with *Candida*, *Aspergillus,* and *Zygomycetes* species [*see 7:XXXVIII Mycotic Infections in the Compromised Host*].

Treatment

Antibiotic selection Antibiotic therapy in the neutropenic patient who has fever or other manifestation of infection should be initiated promptly and at full dosages (after adjustments are made for renal and hepatic function). If a specific infectious organism is identified, the antibiotic regimen should be modified to ensure coverage of that organism, ideally with two different antibiotics. However, the spectrum of coverage should not be narrowed. Neutropenic patients who have an identifiable source of infection should receive treatment for at least as long as nonneutropenic patients with a similar infection. In the absence of an identifiable source of infection, antibiotics should be continued until the patient has a granulocyte count of greater than 500/μl and has been afebrile for 48 hours or longer. The duration of antibiotic therapy in the patient with prolonged neutropenia who becomes promptly afebrile is controversial. One reasonable approach is to continue intravenous antibiotics in high-risk patients and to change to an oral antibiotic regimen in low-risk patients after 48 hours without fever (e.g., ciprofloxacin plus amoxicillin-clavulanate).[48]

The antimicrobial agents that are initially selected for patients with fever and neutropenia should provide broad-spectrum coverage against both gram-positive and gram-negative organisms (including *P. aeruginosa*) with consideration for the frequency of antibiotic resistance in the local population. A number of intravenous and oral antibiotic regimens are effective.[48-53] Reasonable dual combination regimens include an antipseudomonal aminoglycoside combined with either an antipseudomonal penicillin or an extended-spectrum antipseudomonal cephalosporin; ciprofloxacin plus an antipseudomonal penicillin; or a double β-lactam.[53] Alternatively, monotherapy with a broad-spectrum agent (e.g., a carbapenem or an extended-spectrum antipseudomonal cephalosporin) may be used[53] [*see 7:XIV Chemotherapy of Infection*]. The addition of vancomycin to the regimen should be considered in patients with a long-term vascular access device, in those with severe mucositis, and in patients at centers with high rates of methicillin-resistant *S. aureus* infection.[48,54] Randomized studies demonstrate that patients at low risk for infection may be treated effectively with oral antibiotics (ciprofloxacin plus amoxicillin–clavulanic acid).[55,56] Limited studies also suggest that carefully selected low-risk patients may be managed on an outpatient basis with oral or parenteral antibiotics.[41-43]

Secondary infection, antimicrobial resistance, or inadequate initial coverage commonly necessitates ongoing modification of the antimicrobial regimen.[57] To minimize toxicity, serum levels of the aminoglycosides and vancomycin should be monitored and the doses adjusted if necessary. Nephrotoxicity from aminoglycosides may be increased in the elderly and in patients who have been or are being treated with cisplatin or amphotericin B.

Neutropenic patients who have persistent fever despite receiving broad-spectrum antibiotics for 5 to 7 days or who experience recurrence of fever after an initial response to antibiotics should be considered for empirical antifungal therapy with liposomal or conventional amphotericin B.[57,58]

Indwelling catheters Patients with indwelling catheters who experience fever with or without neutropenia present special problems. Indwelling catheters are associated with infections of the exit site, the catheter tunnel, and the catheter lumen. A variety of organisms cause catheter-related infections, the most common of which is *S. epidermidis*. Catheter lumen–related or exit-site infections caused by *S. epidermidis* may be effectively treated by administering antibiotics through the catheter lumen without removing the catheter. In patients with fungal infections, however, the catheter should be removed. Infections of the subcutaneous catheter tunnel usually necessitate catheter removal regardless of the causative organism. Antibiotics should be rotated through each lumen of multilumen indwelling catheters to ensure eradication of all organisms.

Colony-stimulating factors Colony-stimulating factors are hematopoietic growth factors that stimulate the growth and maturation of committed bone marrow progenitor cells. Two of these colony-stimulating factors are currently available: G-CSF and GM-CSF. Treatment with G-CSF before neutropenia develops reduces the duration of the neutropenia and decreases the frequency of infectious episodes in patients who are receiving highly myelosuppressive cytotoxic chemotherapy. However, the high cost of G-CSF and GM-CSF precludes their prophylactic use except in patients who are receiving highly myelosuppressive therapy or in those who have a history of fever and neutropenia.[48] Prospective, randomized trials of both G-CSF and GM-CSF have demonstrated limited value in the treatment of fever and neutropenia. G-CSF and GM-CSF decrease the duration of neutropenia and hospitalization but not days with fever or mortality.[47,59-62]

FEVER AND ACTUAL OR FUNCTIONAL SPLENECTOMY

The spleen is an important organ in antibody production and in the destruction of nonopsonized or poorly opsonized bacteria. Patients who have undergone surgical splenectomy or who are functionally asplenic after splenic radiation therapy are at increased risk for infection with encapsulated bacteria, including *S. pneumoniae*, *Haemophilus influenzae*, *Neisseria meningitidis*, group A streptococci, and *Capnocytophaga canimorsus*.[49] Such patients who become infected with these organisms may experience rapidly overwhelming sepsis. Therefore, asplenic or functionally asplenic patients should have in their possession antibiotics that are active against the encapsulated bacteria. Patients should start taking these antibiotics at the onset of any febrile illness and should promptly seek medical attention. The threshold for hospitalization should be low, and use of an antibiotic such as cefotaxime or ceftriaxone should be urgently initiated. Vancomycin should be considered if penicillin or cephalosporin resistance is locally common or is suspected.[63]

Metabolic Emergencies

HYPERCALCEMIA OF MALIGNANCY

Hypercalcemia of malignancy occurs in 10% to 20% of patients with cancer at some time during their illness and may range from a mildly increased calcium level in asymptomatic patients to a life-threatening emergency. Hypercalcemia is usually a manifestation of advanced disease and is observed in patients with hematologic malignancies, solid tumors without bone metastasis, and solid tumors with bone metastasis.[64]

Non–small cell lung cancer, breast cancer, head and neck cancer, renal cell cancer, myeloma, and T cell lymphoma are the tumors most commonly associated with hypercalcemia. Patients with breast cancer metastatic to bone may experience hypercalcemia during the initiation of treatment with estrogens, androgens, progestin, or antiestrogens.

Pathophysiology

Normal calcium homeostasis is a balance of intestinal absorption of calcium, bone resorption and formation, and renal excretion of calcium.[64] In healthy adults, bone resorption and formation are closely balanced, so that renal elimination of calcium is primarily determined by intestinal calcium absorption. Intestinal calcium absorption occurs both actively and passively. Active absorption of calcium is subject to saturation and is regulated primarily by 1,25-dihydroxyvitamin D_3 [1,25-$(OH)_2D_3$]. Other hormones, such as parathyroid hormone (PTH) and glucocorticoids, participate indirectly in regulation of the active intestinal absorption of calcium by modulating renal production of 1,25-$(OH)_2D_3$. Renal calcium reabsorption is increased by PTH.

Both trabecular bone and cortical bone undergo a dynamic process: bone resorption by osteoclasts, followed by bone formation mediated by osteoblasts. Bone resorption is stimulated by PTH and inhibited by calcitonin. Under normal conditions, the temporal and spatial balance of bone resorption and formation limits the ability of bone to participate in calcium regulation. The final pathway common to the mechanisms of hypercalcemia in malignancy is an uncoupling of the spatial and temporal balance of bone resorption and formation; as a result, bone resorption occurs more quickly than bone formation. Intestinal calcium absorption decreases, the extracellular volume contracts, and urinary excretion of calcium declines, producing hypercalcemia.

The uncoupling of bone resorption and formation is mediated by humoral factors released by tumors without bone metastasis and by paracrine factors released by metastatic deposits in bone. The humoral mediators that have direct effects on the uncoupling of bone include PTH-related protein (PTHrP), transforming growth factor–α (TGF-α), colony-stimulating factors, interleukin-1 (IL-1) and IL-6, tumor necrosis factor, and prostaglandins.[65,66]

PTHrP is a peptide that has homology to PTH at eight of 13 amino acids at the PTH receptor–binding domain. PTHrP is primarily associated with epithelium and is found in very high concentrations in milk. In humoral hypercalcemia of malignancy, PTHrP appears to activate the PTH receptor, which results in increased renal reabsorption of calcium and increased bone resorption. Unlike PTH, however, PTHrP does not stimulate new bone formation.[65-67]

A syndrome of increased levels of 1,25-$(OH)_2D_3$ may occur in patients with Hodgkin or non-Hodgkin lymphoma (especially patients with T cell lymphoma or leukemia who test positive for human T cell lymphotropic virus type I). In affected patients, elevations in the level of 1,25-$(OH)_2D_3$ result in increased intestinal absorption of calcium and increased bone resorption, leading to hypercalcemia.

Many patients with hypercalcemia of malignancy have widespread bone destruction secondary to metastatic lesions, a development that is especially common in patients with breast cancer and multiple myeloma. In these patients, local bone destruction appears to be the cause of the hypercalcemia. The metastatic cells either release directly or induce the surrounding normal cells to release paracrine factors such as TGF-α and prostaglandins, which uncouple bone resorption and bone formation by activating local osteoclasts.[65]

Hypercalcemia may arise in association with widespread osseous metastases, as occur with breast and prostate cancer; with the release of osteoclast-activating factor in such B cell malignancies as multiple myeloma and (less often) non-Hodgkin lymphoma; and with the release of a cytokine that interacts with the PTH receptor, as seen with any squamous cell carcinoma (e.g., cancer of the lung, cervix, anus, head and neck, and esophagus), renal cell carcinoma, or hepatocellular carcinoma. In patients with hepatocellular carcinoma, hypercalcemia may occur in the total absence of bony involvement; by contrast, in patients with breast or prostate cancer, hypercalcemia almost always occurs in the setting of bony involvement.

Diagnosis

Patients with hypercalcemia of malignancy have nonspecific symptoms. Early symptoms include polydipsia, polyuria, anorexia, fatigue, and constipation. Abdominal pain and bloating, nausea, and change in mental status are also seen. Late manifestations include coma and cardiac arrhythmia. Bone pain may or may not be present. Documentation of an elevated serum calcium level, either corrected serum calcium or ionized calcium, is required for diagnosis.

In the differential diagnosis of hypercalcemia of malignancy, the most common competing diagnosis is primary hyperparathyroidism. Primary hyperparathyroidism is associated with a slightly elevated serum calcium level (i.e., 11 to 12 mg/dl) and a decreased serum phosphate level (i.e., 2 to 3 mg/dl). In contrast, the hypercalcemia of malignancy is usually associated with a more dramatic increase in the serum calcium level (often to greater than 14 to 15 mg/dl) with a normal serum phosphate value. Immunoradiometric assays for intact PTH appear to distinguish between hypercalcemic patients with elevated PTH levels (hyperparathyroidism) and those with depressed PTH levels (hypercalcemia of malignancy).

Treatment

Hypercalcemia of malignancy is usually a manifestation of advanced cancer. In patients who have symptomatic hypercalcemia and for whom there are no effective anticancer treatment options, median survival is approximately 35 days; for patients with anticancer treatment options, median survival is less than 90 days.[68] For this reason, in patients for whom there are no anticancer treatment options, withholding active treatment of the hypercalcemia may be the most humane, compassionate, and appropriate course of action.[65] In the patient for whom effective anticancer therapy is available, treatment of the hypercalcemia is appropriate and includes volume and electrolyte repletion, inhibition of bone resorption, and treatment of the underlying malignancy.[69]

Extracellular volume deficits exist in all patients with symptomatic hypercalcemia of malignancy. The single most important and urgent therapeutic measure is the infusion of normal saline to correct the extracellular volume deficit, increase glomerular filtration rate (GFR), and, secondarily, increase renal calcium excretion. Hypokalemic metabolic alkalosis commonly is observed in patients with hypercalcemia of malignancy despite the coexisting decrease in GFR. Therefore, attention should also be directed to the correction of any coexisting potassium deficit, especially in patients who are taking digitalis preparations. Phosphate levels should be monitored, with oral phosphate

replacement as necessary. Care should be taken to maintain the calcium-phosphate product below 30 to 40.

Diuretics should not be used until the volume deficit has been fully corrected. Loop diuretics cause calciuresis and therefore may be effective in acutely decreasing calcium levels—but again, only after volume repletion. Thiazide diuretics decrease renal calcium excretion and should be specifically avoided.

The bisphosphonates offer an improved and simplified treatment of hypercalcemia of malignancy. The bisphosphonates have a high affinity for areas of high bone turnover, such as the areas of bony involvement with malignancy, where they block osteoclast attachment to bone matrix and osteoclast recruitment and differentiation.[65] Zoledronate (4 mg given in a 15-minute intravenous infusion) is the preferred bisphosphonate in the treatment of hypercalcemia of malignancy, with a complete response rate of 88% at day 10.[70] The toxic effects of the bisphosphonates include transient fever, local reactions at the infusion site, hypomagnesemia, hypophosphatemia, decreased renal function, and occasionally a flulike syndrome.

Other agents used in the second-line treatment of hypercalcemia of malignancy include the antineoplastic agent plicamycin, calcitonin, and glucocorticoids.

Immobilization leads to a rapid increase in bone resorption. Mobilization is therefore an important but frequently neglected component in the treatment of hypercalcemia of malignancy. Agents that increase calcium levels, such as thiazide diuretics, vitamin D, and lithium, should be discontinued. Except in the rare case of hypercalcemia of malignancy secondary to increased levels of $1,25\text{-}(OH)_2D_3$, intestinal absorption of calcium is depressed in patients with hypercalcemia of malignancy. Therefore, efforts to decrease intestinal absorption of calcium by dietary restriction are of limited therapeutic value.

Overall, a reasonable approach to the treatment of hypercalcemia of malignancy includes rapid saline rehydration, with correction of any coexisting electrolyte deficiency; inhibition of bone resorption with zoledronate; ambulation, if feasible; and active treatment of the underlying malignancy.

TUMOR LYSIS SYNDROME

Patients with a large tumor burden or rapidly proliferating tumors may experience a spontaneous or treatment-related tumor lysis syndrome that includes the rapid discharge of intracellular electrolytes and nucleic acids. The syndrome usually occurs within 6 to 72 hours after the initiation of therapy and is characterized by hyperkalemia, hyperuricemia, and hyperphosphatemia with secondary hypocalcemia. Tumor lysis syndrome is most often seen in patients with lymphoma or leukemia but also occurs in patients with a variety of solid tumors. A large tumor burden, a high growth fraction, an elevated pretreatment lactate dehydrogenase level, an elevated pretreatment uric acid level, or preexisting renal insufficiency increases the risk of tumor lysis syndrome. High levels of uric acid, xanthine, and phosphate may result in precipitation of these substances in the kidney. Renal sludging and acute renal insufficiency or failure further aggravates the metabolic abnormality.

The hyperkalemia associated with tumor lysis syndrome may be accentuated by associated renal insufficiency or renal failure and may cause electrocardiographic alterations and potentially fatal cardiac arrhythmia. The major manifestation of hyperphosphatemia is secondary hypocalcemia caused by precipitation of calcium phosphate in the soft tissues and the kidney. Hypocalcemia may lead to alterations in mental status, neuromuscular irritability, carpopedal spasm, and seizures.

Treatment

Anticipation and controlled management of tumor lysis are the keys to preventing the syndrome. Patients who are at risk should be hospitalized and fully hydrated, with ongoing diuresis before, during, and after treatment, and they should undergo frequent electrolyte monitoring. Diuresis minimizes the renal sludging caused by high urinary loads of uric acid, xanthine, and phosphate. Pretreatment with allopurinol blocks the conversion of hypoxanthine and xanthine to uric acid and minimizes the uric acid sludging in the kidney caused by xanthine crystals, but typically, it takes days to lower uric acid levels. Alkalinization of the urine with sodium bicarbonate infusion increases the solubility of urinary uric acid and decreases the risk of urate nephropathy but may increase risk of calcium phosphate precipitation. Recombinant urate oxidase (rasburicase) rapidly converts uric acid to allantoin, which is more soluble than uric acid and is readily cleared by the kidney.[71] Rasburicase rapidly lowers uric acid levels, and it has been approved by the Food and Drug Administration for the treatment of tumor lysis syndrome in children[71,72]; this agent appears to have similar activity in adults.[73,74] Thus, rasburicase should be considered an alternative to allopurinol in patients with very high uric acid levels or fulminant tumor lysis syndrome.

Patients at risk for tumor lysis syndrome should not receive supplemental potassium and should be monitored for development of hyperkalemia. Life-threatening hyperkalemia should be treated aggressively with diuresis, potassium-binding salts, and renal dialysis. Patients who have either symptomatic hypocalcemia or the electrocardiographic changes associated with hypocalcemia should be treated with an infusion of calcium gluconate. The hypocalcemia may persist beyond the period of observed hyperphosphatemia.

In its most severe form, tumor lysis syndrome may result in the rapid onset of profound and life-threatening fluid and electrolyte abnormalities and acute renal failure. The rapid lysis of tumor usually is iatrogenic and transient; therefore, aggressive supportive care is appropriate. Such care includes meticulous monitoring of fluid and electrolyte balances, cardiac monitoring, and hemodialysis as necessary.

Neurologic Emergencies

BRAIN METASTASIS

Intracranial metastasis occurs in 20% to 30% of patients with systemic cancer. The most common primary cancers that result in intracranial metastasis are lung cancer, breast cancer, GI cancer, genitourinary cancer, and melanoma. Most metastases are to the cortical-medullary junction, are associated with vasogenic edema, and occur with approximately equal frequency as a single metastasis or multiple metastases.

Diagnosis

Intracranial metastases compress the adjacent brain parenchyma and increase intracranial pressure. The increased intracranial pressure is associated with nonspecific symptoms, including headache that is frequently retro-orbital, nausea, and vomiting, all of which may be most severe in the morning. Cranial nerve abnormalities, including blurred vision, diplopia, and

visual-field defects, also are common. Finally, localized mass effects and edema may produce localized neurologic signs and symptoms, including motor and sensory abnormalities, dysphasia, ataxia, personality change, and seizures.

CT with contrast enhancement and magnetic resonance imaging with contrast enhancement are both relatively sensitive methods of diagnostic imaging in patients with suspected intracranial metastasis. Contrast-enhanced MRI is generally preferred, because it is more sensitive.

When multiple intraparenchymal metastases develop in a patient who has a known primary neoplasm, histologic diagnosis of the metastases usually is not warranted. However, intraparenchymal metastasis may be the presenting manifestation in some cases of neoplastic disease. In such cases, a focused evaluation for the primary tumor is appropriate, with special consideration of lung cancer, breast cancer, and melanoma. If the primary tumor is identified, it should be biopsied. Should no primary tumor be identified, biopsy of the intracranial disease is appropriate. Biopsy or excisional biopsy is particularly appropriate to help differentiate between a primary CNS tumor and a metastatic tumor in patients with no known primary malignancy who have a solitary intracranial lesion.

Treatment

Factors to consider when deciding on a treatment regimen for intracranial metastasis are the patient's age, whether the disease is controlled at other systemic sites, the patient's performance status, and the number of intraparenchymal metastases. The survival of patients with untreated intraparenchymal metastases is approximately 1 month.[75] Treatment usually provides substantial palliation and may prolong survival.

Glucocorticoids should be given to patients with newly diagnosed intraparenchymal metastases. In most patients, dexamethasone (4 to 16 mg/day in divided doses) lessens symptoms within hours to days.[76] Higher doses may improve the rates of response, but toxicity is also increased. The optimal duration of glucocorticoid administration is not known, but doses should be tapered gradually as therapy is completed.

Patients who have seizures should receive antiepileptic drugs. Erythema multiforme major (Stevens-Johnson syndrome) may occur in patients treated with phenytoin, cranial radiation therapy, and glucocorticoids.[77]

Randomized studies document the superiority of surgical resection followed by radiation therapy over the use of radiation therapy alone in the treatment of patients who have surgically resectable solitary intraparenchymal brain metastasis.[78,79] However, patients with uncontrolled systemic disease experience no benefit from the addition of initial surgery. Thus, patients who have a solitary brain metastasis and controlled systemic disease should undergo initial surgical resection when possible, followed by whole brain radiation therapy. Patients who have an unresectable solitary brain metastasis or uncontrolled systemic disease should undergo whole brain radiation therapy.

Whole brain radiation therapy provides substantial palliation and increased survival in patients who have multiple intraparenchymal metastases. A few studies have suggested that initial surgery provides some benefit in patients with multiple surgically resectable intraparenchymal metastases, but other studies have found no advantage.

Techniques now exist for the delivery of high doses of stereotactically directed ionizing radiation to defined tumor volumes with relative sparing of surrounding normal tissue.[80] These techniques have been used successfully to treat solitary and multiple brain metastases with or without whole brain radiation therapy.[81-85] In the primary treatment of intraparenchymal metastasis, the benefit of whole brain radiation therapy after stereotactically directed radiation therapy is still being defined.[81-85]

EPIDURAL SPINAL CORD COMPRESSION

Epidural compression of the spinal cord or cauda equina occurs in 5% to 10% of patients with malignancy. Patients with breast cancer, lung cancer, prostate cancer, lymphoma, renal carcinoma, or sarcoma are at particular risk. The site of epidural compression is thoracic in 70% of patients, lumbar in 20%, and cervical in 10%. More than 30% of patients experience multiple levels of epidural compression.

Pathophysiology

The epidural space lies between the dura mater of the spinal cord and the bony spinal canal. Mass lesions in the epidural space may cause injury to the spinal cord or cauda equina by direct mechanical distortion or by vascular compromise with edema and ischemia or infarction.

About 80% of epidural masses occur as the result of extension of metastasis from the bony spine, especially from the vertebral bodies. Involvement of the epidural space also may occur by extension through the intervertebral foramina by paravertebral adenopathy, particularly in lymphomas; or through the Batson venous plexus in patients with increased intra-abdominal pressure [*see Figure 3*].

Diagnosis

Because the prognosis for recovery of neurologic deficits from spinal cord compression is related to the duration and the severity of the deficits at the start of treatment, early diagnosis is crucial. More than 95% of patients with epidural compression caused by malignancy have pain, either local or radicular, both as the first symptom and at the time of diagnosis [*see Table 1*]. The pain is usually constant and progressive and increases with the Valsalva maneuver or straight leg raising. Unlike the pain of spinal disk disease, the pain of epidural compression typically is worsened by recumbency. Local vertebral tenderness to percussion often is present. Sensory loss in a distribution distal to the site of epidural compression is common and may be rapidly progressive, but this symptom is of limited value in localizing the level of compression. Hyperreflexia and Babinski reflexes may be seen with spinal cord compression above T10, whereas lesions below T10 that compress the conus or cauda equina produce hyporeflexia.

Weakness, usually bilateral and symmetrical, is present in more than 75% of patients at the time of diagnosis and may be rapidly progressive. Autonomic dysfunction of the bladder or bowels is a late sign of epidural compression. The duration, severity, and rapidity of onset of neurologic dysfunction before the initiation of treatment are strong predictors of whether neurologic function can be maintained or restored.[86-91]

Radiographs of the spine and radionuclide bone scans are usually abnormal in patients with epidural compression but are neither sensitive enough nor specific enough for definitive diagnosis or localization.[80] The current recommendation for the radiographic evaluation of patients with possible epidural compression is gadolinium-enhanced MRI of the entire spinal axis. Although myelography is sensitive and specific for detecting epidural compression, it is invasive; it may be uncomfortable for

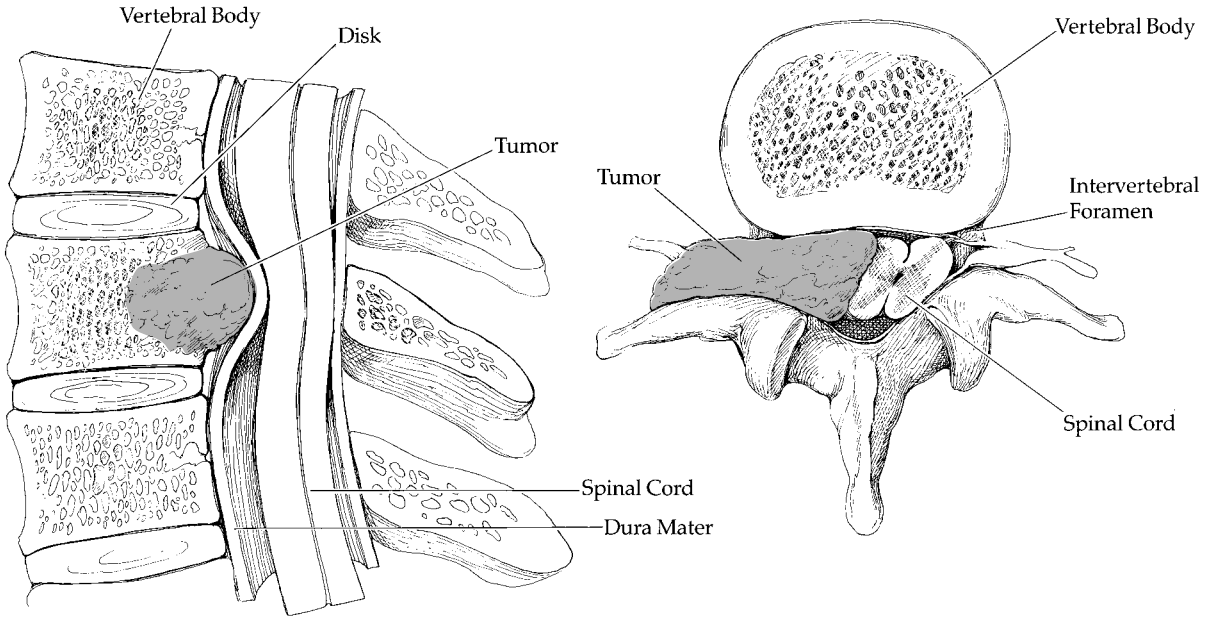


Figure 3 **Spinal cord compression as a complication of neoplastic disease can be caused by extension of a metastatic lesion located in a vertebral body (left, medial section) or by extension of a tumor through an intervertebral foramen (right, transverse section).**

the patient with severe bone pain; and radiologists experienced in its interpretation may not be available. MRI of the spine, especially with gadolinium enhancement, is highly sensitive, noninvasive, and more sensitive than myelography in detecting intramedullary metastasis. MRI of the spine should encompass the entire spinal axis and thus may be time consuming and resource intensive. CT of the spine should not be performed, because its ability to scan the entire spinal axis efficiently and its sensitivity in identifying epidural disease are inferior to those of gadolinium-enhanced MRI.

Treatment

Treatment of epidural cord compression should be initiated immediately, especially in patients with recent onset of neurologic dysfunction or whose neurologic dysfunction is progressive. Ideally, cancer patients with back pain should be evaluated and treated before neurologic dysfunction develops. Nonmalignant spinal disorders (e.g., osteomyelitis, tuberculosis) must be excluded.

The typical patient who has a known malignancy and epidural cord compression should receive immediate treatment with dexamethasone and local radiation therapy. An initial surgical approach is considered in patients who do not have a histologic diagnosis of malignancy, in patients who have an unstable bony spine, in patients whose symptoms progress during radiation therapy, or in patients whose symptoms progress after a maximally tolerated dose of radiation to the spinal cord. Dexamethasone provides pain relief for many patients and may decrease edema that is localized to the area of compression. One evidence-based guideline recommends 24 mg of dexamethasone every 6 hours.[92] Regardless of the dexamethasone dosage, it should be tapered to minimize toxicity.

Patients with known metastatic cancer sensitive to radiation and with documented spinal cord compression should undergo immediate radiation therapy to the level of epidural compression and to a margin of normal tissue above and below it, or they should have immediate anterior surgical resection of the area of metastatic tumor causing the cord compression followed by radiation therapy.[89,92,93] In a randomized clinical trial of immediate surgical resection followed by radiation therapy versus immediate radiation therapy without surgery in 101 patients with spinal cord compression,[93] patients treated with initial surgical resection had superior preservation of the ability to walk (126 days versus 35 days; $P = 0.006$) and of continence, but there was no difference in overall survival (129 days versus 100 days; $P = 0.08$). Spinal cord tolerance to radiation is related both to the fraction size and to the cumulative dose, and multiple dose and fractionation schedules have been used.[94,95] The frequent occurrence of multiple synchronous or metachronous levels of epidural compression necessitates careful planning to minimize the need for serial treatment of adjoining areas of the spine. With

Table 1 Signs and Symptoms Associated with Epidural Spinal Cord Compression[104]

Sign or Symptom	*First Symptom (%)*	*Symptoms at Diagnosis (%)*
Pain	96	96
Weakness	2	76
Autonomic dysfunction	0	57
Sensory loss	0	51
Ataxia	2	3
Herpes zoster	0	2
Flexor spasms	0	1

serial treatment, it is particularly difficult to match adjoining radiation therapy fields to avoid overlapping the fields.

A surgical approach to epidural compression may provide pain relief, halt the progression of the neurologic deficits, allow for stabilization of the spine, and provide a histologic diagnosis of malignancy in patients without a known primary tumor.[92,96] However, most patients with epidural compression have widespread disease and are at increased operative risk. Because the rate of local recurrence after surgery alone is high, local radiation therapy is still warranted. Aggressive surgery for epidural cord compression may produce encouraging results, especially when the spine is unstable, pathologic compression fracture has caused impingement of the cord by bone fragments, or the tumor is resistant to radiation therapy. However, most studies of surgery plus radiation therapy versus radiation therapy alone have not demonstrated meaningful differences in the return of neurologic function and survival.

INTRAMEDULLARY METASTASIS

Intramedullary spinal metastasis is unusual and occurs primarily in patients with lung cancer, breast cancer, colon cancer, or lymphoma. Presenting manifestations are similar to those of epidural spinal cord compression except that the associated motor weakness is commonly unilateral. Intramedullary metastasis must be distinguished from epidural cord compression, leptomeningeal metastasis, radiation myelopathy, primary intramedullary tumors, and necrotizing myelopathy. Myelography may reveal a fusiform swelling of the spinal cord, but myelographic results are frequently normal. High-resolution CT or MRI with gadolinium enhancement is superior to myelography in identifying intramedullary metastasis. Treatment is similar to that of epidural spinal cord compression.

LEPTOMENINGEAL METASTASIS

Leptomeningeal metastasis is uncommon. It is observed chiefly in patients with lung cancer, breast cancer, melanoma, or lymphoma.

Diagnosis

The signs and symptoms of leptomeningeal metastasis may be referable to the brain, cranial nerves, or spine.[97-99] In patients with malignancy, the presence of signs or symptoms that are referable to more than one location within the craniospinal axis always raises the possibility of leptomeningeal metastasis.[97-99] Headache, changes in mental status, ataxia, nausea, vomiting, diplopia, facial weakness, lower extremity weakness, paresthesia, reflex asymmetry, and spinal pain are particularly common. Results of cerebrospinal fluid examination are almost always abnormal, with elevated protein levels and positive cytology being the most common abnormalities. However, excluding leptomeningeal metastasis generally requires three lumbar punctures. Pleocytosis in the CSF may be seen as a manifestation of leptomeningeal involvement in patients with leukemia or lymphoma. Results of contrast-enhanced CT or gadolinium-enhanced MRI may be abnormal, but false negative results are common.[99]

Treatment

Systemic administration of chemotherapy usually results in very low levels of drug in the CSF. However, therapeutic levels of methotrexate, cytarabine, and thiotepa may be safely achieved in the CSF by lumbar puncture or by intraventricular instillation and provide substantial palliation.[99,100] The use of subcutaneous reservoirs attached to a catheter that is inserted into the lateral ventricle (Ommaya reservoir) is a safe and convenient method of delivering intrathecal chemotherapy and provides a uniform distribution of drug throughout the CSF in many patients.[101] A sustained-release formulation of cytarabine is commercially available. In one study comparing sustained-release cytarabine with methotrexate for intrathecal treatment of neoplastic meningitis, onset of neurologic progression was slower with cytarabine (given once every 14 days) than with methotrexate.[102] Radiation therapy may be administered to patients with leptomeningeal metastasis who have particularly severe symptoms, bulky disease, sites of involvement threatening neurologic function, or disease refractory to intrathecal chemotherapy.

The response to therapy for leptomeningeal metastasis is strongly dependent on the underlying tumor type, previous therapy, the extent of disease, and the presence or absence of blockage of CSF flow. Patients who have leptomeningeal involvement from leukemia, lymphoma, or breast cancer may experience substantial palliation and increased survival after prompt and aggressive therapy.[100]

As is the case with the other CNS emergencies, the extent of neurologic dysfunction when therapy is initiated predicts the degree of neurologic function that will be achieved after therapy. In general, however, patients with leptomeningeal metastasis have a poor prognosis.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hancock EW: Neoplastic pericardial disease. Cardiol Clin 8:673, 1990

2. Garcia-Riego A, Cuinas C, Vilanova JJ: Malignant pericardial effusion. Acta Cytol 45:561, 2001

3. Shepherd FA: Malignant pericardial effusion. Curr Opin Oncol 9:170, 1997

4. Tsang TS, Enriquez-Sarano M, Freeman WK, et al: Consecutive 1127 therapeutic echocardiographically guided pericardiocenteses: clinical profile, practice patterns, and outcomes spanning 21 years. Mayo Clin Proc 77:429, 2002

5. Liu G, Crump M, Goss PE, et al: Prospective comparison of the sclerosing agents doxycycline and bleomycin for the primary management of malignant pericardial effusion and cardiac tamponade. J Clin Oncol 14:3141, 1996

6. Shepherd FA, Morgan C, Evans WK, et al.: Medical management of malignant pericardial effusion by tetracycline sclerosis. Am J Cardiol 60:1161, 1987

7. Maher EA, Shepherd FA, Todd TJ: Pericardial sclerosis as the primary management of malignant pericardial effusion and cardiac tamponade. J Thorac Cardiovasc Surg 112:637, 1996

8. Mann T, Brodie BR, Grossman W, et al: Effusive-constrictive hemodynamic pattern due to neoplastic involvement of the pericardium. Am J Cardiol 41:781, 1978

9. Perez CA, Presant CA, Van Amburg AL III: Management of superior vena cava syndrome. Semin Oncol 5:123, 1978

10. Mineo TC, Ambrogi V, Nofroni I, et al: Mediastinoscopy in superior vena cava obstruction: analysis of 80 consecutive patients. Ann Thorac Surg 68:223, 1999

11. Lanciego C, Chacon JL, Julian A, et al: Stenting as first option for endovascular treatment of malignant superior vena cava syndrome. AJR Am J Roentgenol 177:585, 2001

12. Shah R, Sabanathan S, Lowe RA, et al: Stenting in malignant obstruction of superior vena cava. J Thorac Cardiovasc Surg 112:335, 1996

13. Rowell NP, Gleeson FV: Steroids, radiotherapy, chemotherapy and stents for superior vena caval obstruction in carcinoma of the bronchus: a systematic review. Clin Oncol (R Coll Radiol) 14:338, 2002

14. Urruticoechea A, Mesia R, Dominguez J, et al: Treatment of malignant superior vena cava syndrome by endovascular stent insertion: experience on 52 patients with lung cancer. Lung Cancer 43:209, 2004

15. Chan RH, Dar AR, Yu E, et al: Superior vena cava obstruction in small-cell lung cancer. Int J Radiat Oncol Biol Phys 38:513, 1997

16. Adelstein DJ, Hines JD, Carter SF, et al: Thromboembolic events in patients with malignant superior vena cava syndrome and the role of anticoagulation. Cancer 62:2258, 1988

17. Marcy PY, Magne N, Bentolila F, et al: Superior vena cava obstruction: is stenting necessary? Support Care Cancer 9:103, 2001

18. Levi M, de Jonge E, van der Poll T, et al: Advances in the understanding of the

pathogenetic pathways of disseminated intravascular coagulation result in more insight in the clinical picture and better management strategies. Semin Thromb Hemost 27:569, 2001

19. Levi M, de Jonge E, van der Poll T: New treatment strategies for disseminated intravascular coagulation based on current understanding of the pathophysiology. Ann Med 36:41, 2004

20. Bick RL, Strauss JF, Frenkel EP: Thrombosis and hemorrhage in oncology patients. Hematol Oncol Clin North Am 10:875, 1996

21. Higuchi T, Shimizu T, Mori H, et al: Coagulation patterns of disseminated intravascular coagulation in acute promyelocytic leukemia. Hematol Oncol 15:209, 1997

22. Barbui T, Finazzi G, Falanga A: The impact of all-trans-retinoic acid on the coagulopathy of acute promyelocytic leukemia. Blood 91:3093, 1998

23. Levi M, de Jonge E, van der Poll T, et al: Disseminated intravascular coagulation. Thromb Haemost 82:695, 1999

24. Pasquini E, Gianni L, Aitini E, et al: Acute disseminated intravascular coagulation syndrome in cancer patients. Oncology 52:505, 1995

25. Warren BL, Eid A, Singer P, et al: Caring for the critically ill patient: high-dose antithrombin III in severe sepsis: a randomized controlled trial. JAMA 286:1869, 2001

26. de Jonge E, van der Poll T, Kesecioglu J, et al: Anticoagulant factor concentrates in disseminated intravascular coagulation: rationale for use and clinical experience. Semin Thromb Hemost 27:667, 2001

27. Elting LS, Rubenstein EB, Martin CG, et al: Incidence, cost, and outcomes of bleeding and chemotherapy dose modification among solid tumor patients with chemotherapy-induced thrombocytopenia. J Clin Oncol 19:1137, 2001

28. Schiffer CA, Anderson KC, Bennett CL, et al: Platelet transfusion for patients with cancer: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol 19:1519, 2001

29. Isaacs C, Robert N, Bailey F, et al: Randomized placebo-controlled study of recombinant human interleukin-11 to prevent chemotherapy-induced thrombocytopenia in patients with breast cancer receiving dose-intensive cyclophosphamide and doxorubicin. J Clin Oncol 15:3368, 1997

30. Reynolds CH: Clinical efficacy of rhIL-11. Oncology (Huntingt) 14:32, 2000

31. Elliott MA, Nichols WL: Thrombotic thrombocytopenic purpura and hemolytic uremic syndrome. Mayo Clin Proc 76:1154, 2001

32. Hosler GA, Cusumano AM, Hutchins GM: Thrombotic thrombocytopenic purpura and hemolytic uremic syndrome are distinct pathologic entities: a review of 56 autopsy cases. Arch Pathol Lab Med 127:834, 2003

33. Moake JL, Byrnes JJ: Thrombotic microangiopathies associated with drugs and bone marrow transplantation. Hematol Oncol Clin North Am 10:485, 1996

34. Rock GA, Shumak KH, Buskard NA, et al: Comparison of plasma exchange with plasma infusion in the treatment of thrombotic thrombocytopenic purpura. Canadian Apheresis Study Group. N Engl J Med 325:393, 1991

35. Zeigler ZR, Shadduck RK, Gryn JF, et al: Cryoprecipitate poor plasma does not improve early response in primary adult thrombotic thrombocytopenic purpura (TTP). J Clin Apheresis 16:19, 2001

36. Chang JC, Kathula SK: Various clinical manifestations in patients with thrombotic microangiopathy. J Investig Med 50:201, 2002

37. Snyder HW Jr, Mittelman A, Oral A, et al: Treatment of cancer chemotherapy-associated thrombotic thrombocytopenic purpura/hemolytic uremic syndrome by protein A immunoadsorption of plasma. Cancer 71:1882, 1993

38. Itala M, Remes K: Excellent response of refractory life-threatening thrombotic thrombocytopenic purpura to cyclosporine treatment. Clin Lab Haematol 26:65, 2004

39. Allford SL, Hunt BJ, Rose P, et al: Guidelines on the diagnosis and management of the thrombotic microangiopathic haemolytic anaemias. Br J Haematol 120:556, 2003

40. Ahmad A, Aggarwal A, Sharma D, et al: Rituximab for treatment of refractory/relapsing thrombotic thrombocytopenic purpura (TTP). Am J Hematol 77:171, 2004

41. Malik IA, Khan WA, Karim M, et al: Feasibility of outpatient management of fever in cancer patients with low-risk neutropenia: results of a prospective randomized trial. Am J Med 98:224, 1995

42. Rubenstein EB, Rolston KV: Outpatient treatment of febrile neutropenic patients with cancer. Eur J Cancer 31A:2, 1995

43. Rolston KV: New trends in patient management: risk-based therapy for febrile patients with neutropenia. Clin Infect Dis 29:515, 1999

44. Morstyn GM, Campbell L, Lieschke G, et al: Treatment of chemotherapy-induced neutropenia by subcutaneously administered granulocyte colony-stimulating factor with optimization of dose and duration of therapy. J Clin Oncol 7:1554, 1989

45. Gianni AM, Bregni M, Siena S, et al: Recombinant human granulocyte-macrophage colony-stimulating factor reduces hematologic toxicity and widens clinical applicability of high-dose cyclophosphamide treatment in breast cancer and non-Hodgkin's lymphoma. J Clin Oncol 8:768, 1990

46. Holmes FA, O'Shaughnessy JA, Vukelja S, et al: Blinded, randomized, multicenter study to evaluate single administration pegfilgrastim once per cycle versus daily filgrastim as an adjunct to chemotherapy in patients with high-risk stage II or stage III/IV breast cancer. J Clin Oncol 20:727, 2002

47. Ozer H, Armitage JO, Bennett CL, et al: 2000 update of recommendations for the use of hematopoietic colony-stimulating factors: evidence-based, clinical practice guidelines. J Clin Oncol 18:3558, 2000

48. Hughes WT, Armstrong D, Bodey GP, et al: 2002 guidelines for the use of antimicrobial agents in neutropenic patients with cancer. Infectious Diseases Society of America. Clin Infect Dis 34:730, 2002

49. Pizzo PA: Fever in immunocompromised patients. N Engl J Med 341:893, 1999

50. Jones RN: Contemporary antimicrobial susceptibility patterns of bacterial pathogens commonly associated with febrile patients with neutropenia. Clin Infect Dis 29:495, 1999

51. Viscoli C: The evolution of the empirical management of fever and neutropenia in cancer patients. J Antimicrob Chemother 41(suppl D):65, 1998

52. Ramphal R: Is monotherapy for febrile neutropenia still a viable alternative? Clin Infect Dis 29:508, 1999

53. Fever and Neutropenia, Version 1.2004. National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. National Comprehensive Cancer Network, November 1, 2004 http://www.nccn.org/professionals/physician_gls/PDF/fever.pdf

54. Feld R: Criteria for response in patients in clinical trials of empiric antibiotic regimens for febrile neutropenia: is there agreement? Support Care Cancer 6:444, 1998

55. Freifeld A, Marchigiani D, Walsh T, et al: A double-blind comparison of empirical oral and intravenous antibiotic therapy for low-risk febrile patients with neutropenia during cancer chemotherapy. N Engl J Med 341:305, 1999

56. Kern WV, Cometta A, De Bock R, et al: Oral versus intravenous empirical antimicrobial therapy for fever in patients with granulocytopenia who are receiving cancer chemotherapy. International Antimicrobial Therapy Cooperative Group of the European Organization for Research and Treatment of Cancer. N Engl J Med 341:312, 1999

57. Hughes WT, Armstrong D, Bodey GP, et al: Guidelines for the use of antimicrobial agents in neutropenic patients with unexplained fever. J Infect Dis 161:381, 1990

58. Cagnoni PJ, Walsh TJ, Prendergast MM, et al: Pharmacoeconomic analysis of liposomal amphotericin B versus conventional amphotericin B in the empirical treatment of persistently febrile neutropenic patients. J Clin Oncol 18:2476, 2000

59. Mayordomo JI, Rivera F, Diaz-Puente MT, et al: Improving treatment of chemotherapy-induced neutropenic fever by administration of colony-stimulating factors. J Natl Cancer Inst 87:803, 1995

60. Ravaud A, Chevreau C, Cany L, et al: Granulocyte-macrophage colony-stimulating factor in patients with neutropenic fever is potent after low-risk but not after high-risk neutropenic chemotherapy regimens: results of a randomized phase III trial. J Clin Oncol 16:2930, 1998

61. Garcia-Carbonero R, Mayordomo JI, Tornamira MV, et al: Granulocyte colony-stimulating factor in the treatment of high-risk febrile neutropenia: a multicenter randomized trial. J Natl Cancer Inst 93:31, 2001

62. Clark OA, Lyman G, Castro AA, et al: Colony stimulating factors for chemotherapy induced febrile neutropenia. Cochrane Database Syst Rev (3):CD003039, 2003

63. Brigden ML, Pattullo AL: Prevention and management of overwhelming postsplenectomy infection—an update. Crit Care Med 27:836, 1999

64. Theriault RL: Hypercalcemia of malignancy: pathophysiology and implications for treatment. Oncology 7:47, 1993

65. Walls J, Bundred N, Howell A: Hypercalcemia and bone resorption in malignancy. Clin Orthop (312):51, 1995

66. Odell WD: Endocrine/metabolic syndromes of cancer. Semin Oncol 24:299, 1997

67. Barri YM, Knochel JP: Hypercalcemia and electrolyte disturbances in malignancy. Hematol Oncol Clin North Am 10:775, 1996

68. Ling PJ, A'Hern RP, Hardy JR: Analysis of survival following treatment of tumour-induced hypercalcaemia with intravenous pamidronate (APD). Br J Cancer 72:206, 1995

69. Stewart AF: Clinical practice: hypercalcemia associated with cancer. N Engl J Med 352:373, 2005

70. Major P, Lortholary A, Hon J, et al: Zoledronic acid is superior to pamidronate in the treatment of hypercalcemia of malignancy: a pooled analysis of two randomized, controlled clinical trials. J Clin Oncol 19:558, 2001

71. Navolanic PM, Pui CH, Larson RA, et al: Elitek-rasburicase: an effective means to prevent and treat hyperuricemia associated with tumor lysis syndrome, a Meeting Report, Dallas, Texas, January 2002. Leukemia 17:499, 2003

72. Pui C-H, Mahmoud HH, Wiley JM, et al: Recombinant urate oxidase for the prophylaxis or treatment of hyperuricemia in patients with leukemia or lymphoma. J Clin Oncol 19:697, 2001

73. Coiffier B, Mounier N, Bologna S, et al: Efficacy and safety of rasburicase (recombinant urate oxidase) for the prevention and treatment of hyperuricemia during induction chemotherapy of aggressive non-Hodgkin's lymphoma: results of the GRAAL1 (Groupe d'Etude des Lymphomes de l'Adulte Trial on Rasburicase Activity in Adult Lymphoma) study. J Clin Oncol 21:4402, 2003

74. Bosly A, Sonet A, Pinkerton CR, et al: Rasburicase (recombinant urate oxidase) for the management of hyperuricemia in patients with cancer: report of an international compassionate use study. Cancer 98:1048, 2003

75. Wen PY, Loeffler JS: Brain metastases. Curr Treat Options Oncol 1:447, 2000

76. Vecht CJ: Clinical management of brain metastasis. J Neurol 245:127, 1998

77. Micali G, Linthicum K, Han N, et al: Increased risk of erythema multiforme major with combination anticonvulsant and radiation therapies. Pharmacotherapy 19:223, 1999

78. Patchell RA, Tibbs PA, Walsh JW, et al: A randomized trial of surgery in the treatment of single metastases to the brain. N Engl J Med 322:494, 1990

79. Vecht CJ, Haaxma-Reiche H, Noordijk EM, et al: Treatment of single brain metastasis: radiotherapy alone or combined with neurosurgery. Ann Neurol 33:583, 1993

80. Boyd TS, Mehta MP: Stereotactic radiosurgery for brain metastases. Oncology (Huntingt) 13:1397, 1999

81. Sneed PK, Lamborn KR, Forstner JM, et al: Radiosurgery for brain metastases: is whole brain radiotherapy necessary? Int J Radiat Oncol Biol Phys 43:549, 1999

82. Regine WF, Huhn JL, Patchell RA, et al: Risk of symptomatic brain tumor recurrence and neurologic deficit after radiosurgery alone in patients with newly diagnosed brain metastases: results and implications. Int J Radiat Oncol Biol Phys 52:333, 2002

83. Shehata MK, Young B, Reid B, et al: Stereotactic radiosurgery of 468 brain metastases < or = 2 cm: implications for SRS dose and whole brain radiation therapy. Int J Radiat Oncol Biol Phys 59:87, 2004

84. Petrovich Z, Yu C, Giannotta SL, et al: Survival and pattern of failure in brain metastasis treated with stereotactic gamma knife radiosurgery. J Neurosurg 97:499, 2002

85. Andrews DW, Scott CB, Sperduto PW, et al: Whole brain radiation therapy with or without stereotactic radiosurgery boost for patients with one to three brain metastases: phase III results of the RTOG 9508 randomised trial. Lancet 363:1665, 2004

86. Byrne TN: Spinal cord compression from epidural metastases. N Engl J Med 327:614, 1992

87. Husband DJ: Malignant spinal cord compression: prospective study of delays in referral and treatment. BMJ 317:18, 1998

88. Solberg A, Bremnes RM: Metastatic spinal cord compression: diagnostic delay, treatment, and outcome. Anticancer Res 19:677, 1999

89. Kovner F, Spigel S, Rider I, et al: Radiation therapy of metastatic spinal cord compression. Multidisciplinary team diagnosis and treatment. J Neurooncol 42:85, 1999

90. Helweg-Larsen S, Sorensen PS, Kreiner S: Prognostic factors in metastatic spinal cord compression: a prospective study using multivariate analysis of variables influencing survival and gait function in 153 patients. Int J Radiat Oncol Biol Phys 46:1163, 2000

91. Rades D, Heidenreich F, Karstens JH: Final results of a prospective study of the prognostic value of the time to develop motor deficits before irradiation in metastatic spinal cord compression. Int J Radiat Oncol Biol Phys 53:975, 2002

92. Loblaw DA, Laperriere NJ: Emergency treatment of malignant extradural spinal cord compression: an evidence-based guideline. J Clin Oncol 16:1613, 1998

93. Patchell RA, Tibbs WF, Regine WF, et al: A randomized trial of direct decompressive surgical resection in the treatment of spinal cord compression caused by metastasis. Proc Am Soc Clin Oncol 22:1, 2003, Abstract 2
http://www.asco.org/ac/1,1003,_12-002643-00_18-0023-00_19-00101474,00.asp

94. Hoskin PJ, Grover A, Bhana R: Metastatic spinal cord compression: radiotherapy outcome and dose fractionation. Radiother Oncol 68:175, 2003

95. Rades D, Fehlauer F, Hartmann A, et al: Reducing the overall treatment time for radiotherapy of metastatic spinal cord compression (MSCC): 3-year results of a prospective observational multi-center study. J Neurooncol 70:77, 2004

96. Sundaresan N, Sachdev VP, Holland JF, et al: Surgical treatment of spinal cord compression from epidural metastasis. J Clin Oncol 13:2330, 1995

97. Wolfgang G, Marcus D, Ulrike S: LC: clinical syndrome in different primaries. J Neurooncol 38:103, 1998

98. Chamberlain MC: Leptomeningeal metastases: a review of evaluation and treatment. J Neurooncol 37:271, 1998

99. DeAngelis LM: Current diagnosis and treatment of leptomeningeal metastasis. J Neurooncol 38:245, 1998

100. Herrlinger U, Forschler H, Kuker W, et al: Leptomeningeal metastasis: survival and prognostic factors in 155 patients. J Neurol Sci 223:167, 2004

101. Chamberlain MC, Kormanik PA, Barba D: Complications associated with intraventricular chemotherapy in patients with leptomeningeal metastases. J Neurosurg 87:694, 1997

102. Glantz MJ, Jaeckle KA, Chamberlain MC, et al: A randomized controlled trial comparing intrathecal sustained-release cytarabine (DepoCyt) to intrathecal methotrexate in patients with neoplastic meningitis from solid tumors. Clin Cancer Res 5:3394, 1999

103. Carlson RW: Palliation of lung cancer: Part I: thoracic complications and pain control. Primary Care & Cancer April 19, 1987

104. Gilbert RW, Kim JH, Posner JB: Epidural spinal cord compression from metastatic tumor: diagnosis and treatment. Ann Neurol 3:40, 1978

Acknowledgment

Figures 1 and 3 Carol Donner.

XIII SARCOMAS OF SOFT TISSUE AND BONE

Haralambos Raftopoulos, m.d.
Karen H. Antman, m.d.

Sarcoma, which was distinguished from carcinoma (crablike tumors) during Hippocrates' era, is derived from the Greek root *sarc*, for "flesh." Most sarcomas develop in tissues of mesodermal embryologic origin and are thus related embryologically to leukemias and lymphomas. Neurosarcomas, peripheral neuroectodermal tumors, and probably Ewing sarcomas, however, arise from tissues of ectodermal embryologic origin.

Overview of Sarcomas

CLASSIFICATION

Sarcomas originating in bone are distinguished from those that occur in soft tissue. Soft tissue sarcomas, which can present in muscles, fat, tendons, synovial tissue, vessels, and nerves, are further subdivided into the classic soft tissue sarcomas of the trunk and extremities and into visceral sarcomas of the gastrointestinal tract[1] and gynecologic tract.[2] Miscellaneous sarcomas include Kaposi sarcoma (KS)[3] and mesotheliomas.

EPIDEMIOLOGY

Bone and soft tissue sarcomas currently represent 1% of adult malignancies and 15% of pediatric malignancies. In 2003, an estimated 10,700 persons in the United States will have developed sarcoma, and 5,200 will have died of the disease.[4]

The predominant histology of sarcoma varies with age [*see Table 1*]. Several histopathologic variants of sarcoma occur predominantly in persons older than 50 years (e.g., liposarcoma, malignant fibrous histiocytoma, and leiomyosarcoma). Embryonal rhabdomyosarcoma develops most frequently in small children (usually in the orbit) and in adolescents (often in the urinary tract).[5] Ewing sarcoma and synovial sarcoma typically develop in individuals between 15 and 30 years of age. Osteosarcoma has a biphasic pattern of incidence, most commonly presenting in adolescents and persons older than 50 years.[6,7] Chondrosarcoma occurs most frequently in persons in the sixth and seventh decades of life.

Whether males or females are primarily affected is dependent upon the histologic variant of the disease [*see Table 1*]. For example, 70% of patients with KS are men, as are 58% to 62% of patients with liposarcoma and embryonal rhabdomyosarcomas; however, 65% of leiomyosarcomas occur in women. Although the incidence of osteosarcoma by age in whites and that in nonwhites are virtually identical, Ewing sarcoma is predominantly a disease of whites.[6,8,9]

GENETIC FACTORS

Familial Sarcoma

Although sarcomas in the general population are rare, certain families have multiple members who are affected. In these families, the risk for sarcoma is increased. Germline *p53* mutations have been implicated in families in which childhood sarcomas, breast cancer, and other tumors (e.g., Li-Fraumeni syndrome) have occurred [*see Table 2*].[10-12]

Patients with neurofibromatosis have a 7% to 10% lifetime risk of developing a malignant sarcoma, most commonly a neurosarcoma or fibrosarcoma.[13,14] About half of all neurofibrosarcomas occur in patients with neurofibromatosis.

About 7% of children with familial retinoblastoma develop osteosarcoma,[15,16] both in radiation ports used for treatment of the initial retinoblastoma and in long bones that clearly are unaffected by the radiotherapy ports, thus suggesting a hereditary susceptibility to osteosarcoma. Sporadic osteosarcomas often carry a 13q chromosomal deletion associated with the lack of a retinoblastoma tumor suppressor gene (*RB*).

Persons with familial or sporadic osteochondromas or fibrous dysplasia of bone may develop osteosarcoma or chondrosarcoma in the preexisting lesion.[17,18]

Translocations Diagnostic of Sarcoma

Translocations involving chromosome 22 characterize both Ewing sarcoma and peripheral neuroectodermal tumors (PNETs). The most common translocation is between chromosome 11 and chromosome 22.[19] Variant translocations affecting t(21;22) and t(7;22) are also characteristic of Ewing sarcoma or PNET.[20]

Other sarcomas that have characteristic cytogenetic abnormalities include synovial sarcomas [t(X;18)],[21] alveolar rhabdomyosarcomas [t(2;13) and t(1;13)], and myxoid liposarcomas [t(12;16)].[22]

ETIOLOGY

Radiotherapy, chemotherapy, and exposure to certain chemicals have been identified as risk factors for the development of sarcoma. About 5% of sarcomas are associated with prior radiotherapy, generally delivered 4 to 20 years before the diagnosis of sarcoma.[23-25] Most radiation-associated sarcomas are osteosarcomas, but mixed mesodermal sarcomas of the uterus, mesotheliomas, angiosarcomas, and other soft tissue sarcomas have been reported. The risk of development of a secondary sarcoma after radiation exposure is substantially less than 1%, except for patients receiving radiation therapy for retinoblastoma or Ewing sarcoma and for radium-dial painters.[25] In a review of the Swedish Cancer Registry, Karlsson and colleagues determined an odds ratio of 2.4 for development of soft tissue sarcoma after breast cancer irradiation.[26] The incidence of angiosarcoma was strongly related to the presence of lymphedema (odds ratio, 9.5). It has been shown that sarcomas arising in radiation ports may be more resistant to chemotherapy. Alkylating agents (particularly melphalan, procarbazine, nitrosoureas, and chlorambucil) given for prior childhood malignancies are causally associated with development of sarcoma.[27]

Environmental or occupational exposures to certain chemicals have been established as risk factors for sarcoma. Asbestos clearly causes mesothelioma.[28] Polyvinyl chloride (plastic industries),[29] androgens, arsenic,[30] and iron overload (hemochromatosis) are associated with angiosarcoma of the liver. Exposure to dioxin (Agent Orange), chlorophenols, or agricultural herbicide has been linked to the development of soft tissue sarcomas in some studies.[31]

Immunosuppressed patients (e.g., renal transplantation patients and patients with AIDS, chronic lymphocytic leukemia,[32]

Table 1 Summary of Sarcomas

	Histologic Variant	Tissue of Origin	Common Age Distribution	Sex Distribution (male:female)	Comments
Soft Tissue Sarcomas	Liposarcoma	Fat	Fifth to seventh decades	3:2	Lesions may be multicentric; rarely arise from lipomas
	Fibrosarcoma	Fibrous tissue			
	Synovial sarcoma	Intramuscular tissue	Second to fourth decades	5:4	Lesions arise adjacent to joints; can be calcified; locally invasive; requires wide surgical margins
	Neurosarcoma and malignant schwannoma	Nerve sheaths	Unknown	Unknown	50% are associated with von Recklinghausen neurofibromatosis (NF1)
	Malignant fibrous histiocytoma (MFH)	Controversial	Fifth to seventh decades	2:1	Incidence decreasing; improved immunohistologic stains able to identify specific tissue markers in poorly differentiated tumors previously classified as MFH
	Peripheral neural ectodermal tumors (PNET)	Neural ectoderm	Unknown	Unknown	Similar to Ewing sarcoma in cytogenetics and treatment
	Leiomyosarcoma	Smooth muscle	Fifth and sixth decades	3:6	Lesions arise in uterus, large blood vessels, GI tract, and other smooth muscle; associated with Epstein-Barr virus
	Rhabdomyosarcoma	Striated muscle	First and second decades	6:4	Chemotherapy is established treatment; in children 4 yr of age or younger, it commonly presents in the orbit; in the early teens, it presents in the genitourinary and gynecologic tracts; in older adolescents, it presents most commonly in the extremities
	Angiosarcoma	Endothelium	Elderly men, young women		Most common primary cardiac malignancy; develops in edematous extremities (Stewart-Treves syndrome); in young women, it presents in the breast; in elderly men, in the scalp; associated with arsenic (insecticides) and vinyl chloride (plastics)
	Hemangioendothelioma	Endothelium	Fifth decade	5:5	Less aggressive than angiosarcoma
	Hemangiopericytoma	Vascular pericytes	Fifth decade		Lesions arise commonly in trunk and extremities but also intracranially
	Kaposi sarcoma (KS)	Endothelium	Differs with variant	7:3	Four variants: classic KS commonly affects elderly men of Mediterranean origin; lymphadenopathic KS affects children in endemic areas of Africa; KS occurring in transplant patients affects patients of various ages; KS occurring in AIDS patients most commonly affects men who have sex with men and affects patients of various ages; KS is associated with human herpesvirus type 8 (HHV-8)
	Mesothelioma	Mesothelium	Sixth and seventh decades	5:1	Associated with asbestos exposure
	Gastrointestinal stromal tumors	Interstitial cells of Cajal	Fourth to seventh decades		Responds to imatinib
Sarcomas of Bone	Osteosarcoma	Bone	Second and third decades; also > 50 yr	~2:1	Surgery and adjunct chemotherapy are established treatments; can be radiation associated; resistant to radiation therapy
	Ewing sarcoma	Bone	Second decade		Chemotherapy, radiation, and possibly surgery are established treatments; rare in nonwhites
	Chondrosarcoma	Cartilage	Sixth and seventh decades		Resistant to radiation and chemotherapy except for mesenchymal chondrosarcomas

or autoimmune hemolytic anemia) have a higher than normal risk of developing soft tissue sarcomas, particularly KS. A viral etiology of sarcoma is well documented in other species (murine, simian, and avian sarcoma viruses), and an infectious agent most likely plays a role in the etiology of two types of sarcoma in humans. In immunosuppressed patients, KS and leiomyosarcoma are associated with human herpesvirus type 8 (HHV-8)[33] and Epstein-Barr virus, respectively.[34,35]

Patients with Paget disease of bone have a 0.2% risk of developing osteosarcoma, particularly in families with Paget disease.[7,8,36] Fibrosarcomas have occasionally developed in scars of burns or major trauma, often after 30 years or more.

Sarcomas of Soft Tissue

Soft tissue sarcomas currently represent 1% of adult malignancies and 15% of pediatric malignancies.[4] Until recently, malignant fibrous histiocytoma (MFH) and leiomyosarcoma (LMS) were the most common diagnoses; however, the incidence of MFH and the incidence of LMS are declining as more specialized diagnostic methods (e.g., immunocytochemistry and virtually diagnostic translocations) lead to reclassification of these tumors into other subsets.[37] Most sarcomas of the GI tract previously thought to be LMS are now diagnosed as gastrointestinal stromal tumors (GISTs).

DIAGNOSIS

Clinical Manifestations

Patients with soft tissue sarcomas usually present with a painless, palpable mass on the extremities or the trunk. Forty percent of soft tissue sarcomas develop in the lower extremity, about 30% in the trunk and retroperitoneum, 15% in the upper extremities, and 15% in the head and neck. Intra-abdominal and retroperitoneal primary lesions may cause pain and weight loss from invasion or displacement of organs. The duration of symptoms before diagnosis ranges from a few weeks to decades, with a median of 1 to 3 months.

Preoperative Evaluation

Preoperative evaluation generally requires computed tomographic studies of the lungs and either a CT scan or magnetic resonance imaging of the primary lesion to establish extent of tumor involvement. Uptake in adjacent bone as evidenced on bone scan may suggest invasion or merely periosteal reaction to the tumor. An incisional biopsy for definitive diagnosis facilitates optimal management of the tumor and should be performed by the surgeon who will undertake the definitive surgical resection. The surgeon is supported by a sarcoma team that usually includes surgical, radiation, orthopedic, and medical oncologists, as well as a pathologist. The sarcoma team's involvement in the planning of the incisional biopsy reduces the risk of an improperly placed biopsy compromising the surgical resection or subsequent radiation.

GRADING AND STAGING

The American Joint Committee staging system for soft tissue sarcomas [*see Table 3*] is largely dependent on tumor size (< 5 cm or > 5 cm) and grade (low, intermediate, or highly aggressive). Stages I through III are based largely on grade, which is determined by the number of mitoses per 10 high-power fields, tumor size, and tumor depth [*see Table 3*].[38-40] The 5-year survival rates for soft tissue sarcomas arising in different anatomic sites are similar when corrected for grade, size, and depth, except for intra-abdominal and retroperitoneal tumors, which tend to be large and to invade vital organs, even if they are low grade.[41] Patients with low-grade, superficial tumors tend to have a good prognosis if the tumors are adequately resected. Even high-grade tumors are associated with a good prognosis if they are less than 5 cm in diameter.

Table 2 Familial Syndromes Associated with Increased Risk of Sarcoma

- Neurofibromatosis type 1 and type 2
- Hereditary retinoblastoma
- Li-Fraumeni syndrome
- Familial Paget disease of bone
- Multiple hereditary exostosis

Table 3 Staging of Soft Tissue Sarcomas*

Stage	*Characteristics*
Stage I	The tumor is low grade, < 5 cm in greatest diameter, and has not spread to lymph nodes or to other organs
Stage II	The tumor is high grade, < 5 cm and superficial, and has not spread to lymph nodes or to other organs
Stage III	The tumor is high grade, > 5 cm and deep, but has not spread to lymph nodes or to other organs
Stage IV	A tumor of any size or grade has spread to lymph nodes or other organs

*Staging system excludes Kaposi sarcoma, dermatofibrosarcoma, desmoid type of grade 1 fibrosarcoma, and sarcomas that begin in the dura mater (e.g., brain), parenchymatous organs, and hollow viscera.

PATHOPHYSIOLOGIC FINDINGS AND IMPLICATIONS FOR MANAGEMENT

The behavior and treatment of most of the histologic variants of soft tissue sarcoma are generally similar grade for grade. Exceptions include osteosarcomas, rhabdomyosarcomas, and Ewing sarcoma, which are treated routinely with multimodality therapy, including chemotherapy, and must be distinguished from other sarcomas. The treatment of mesothelioma and KS also differs from that of the classic sarcomas. Irrespective of the histologic variant of the tumor, the goals of treatment for all histologic variants are the same: (1) local and systemic control of the sarcoma; (2) preservation of the extremity or the vital organ, along with its function, when possible; and (3) maintenance of quality of life. Local control of a soft tissue sarcoma is generally achieved by surgical resection, which is often combined with radiotherapy [*see* Treatment of Soft Tissue Sarcomas, *below*].

Low-grade tumors push aside contiguous structures, whereas high-grade tumors invade adjacent organs and have large areas of necrosis. Soft tissue sarcomas grow along histologic planes and are usually pseudoencapsulated (i.e., microscopic projections of tumor extend beyond the apparent tumor capsule). Any excision that merely "shells out" the apparently encapsulated tumor generally leaves behind microscopic residual tumor and results in local regrowth in about 80% of cases.[40] Wide excision of 3 to 6 cm of normal tissue is thus required for local control if radiation is not planned. Pathologically documented uninvolved margins of 1 mm are acceptable to save a limb or vital organ if adequate doses of radiation can be given. Obtaining local control of intra-abdominal lesions, however, is complicated by the lesion's proximity to vital organs that cannot tolerate the high doses of radiation necessary for control of sarcoma.

HISTOLOGIC VARIANTS

Liposarcoma

Liposarcomas range from low-grade, well-differentiated myxoid lesions to high-grade, round-cell, and pleomorphic lesions. Most liposarcomas develop in the thigh or retroperitoneum, but they may be multicentric, particularly in the abdomen and the retroperitoneum. They rarely, if ever, arise from

benign lipomas. Up to 60% develop in men; patients are generally older than 50 years [*see Table 1*].[42,43]

Fibrosarcoma

Fibrosarcomas, which originate in intermuscular fibrous tissue, fascia tendons, and aponeuroses, range from borderline locally invasive fibromatosis and Dupuytren contracture to highly malignant sarcomas.

Synovial Sarcoma

Characteristically, synovial sarcoma presents as a mass adjacent to a joint and involving the joint's capsule, bursae, and tendon sheath but not the joint itself. The mass may be present for months to years before the diagnosis is made. On radiograph, one third of these lesions may contain fine stippled calcifications or even radiopaque tumor. Histologically, synovial sarcomas are monophasic (epithelial and spindle cell) or biphasic. Wide excision of synovial sarcomas is particularly important but difficult to achieve because of their proximity to large joints.

Synovial sarcoma develops predominantly in adolescents and young adults (median age, 27 years) and somewhat more frequently in men.[21,44-46] Ninety percent of synovial sarcomas contain a characteristic chromosomal translocation involving a nuclear transcription factor on chromosome 18 called SYT. Patients with the t(X;18) SYT-SSX2 translocation may have significantly better metastasis-free survival than patients with the t(X;18) SYT-SSX1 translocation.[21]

Neurosarcomas and Malignant Schwannomas

Neurosarcomas and malignant schwannomas, which constitute 5% of soft tissue sarcomas, arise from tissues of ectodermal embryologic origin. About half of these lesions develop in patients with von Recklinghausen neurofibromatosis, and biopsy should be performed promptly on any mass that suddenly enlarges or causes pain in these patients. Neurosarcomas tend to produce a large fusiform mass extending for some distance within the nerve sheath in the proximal extremity or trunk. Large surgical and radiotherapy margins are required to ensure local control.[13,14]

Malignant Fibrous Histiocytoma

Malignant fibrous histiocytoma (MFH) was previously the most common soft tissue sarcoma in patients older than 50 years [*see Table 1*],[3] but its incidence has fallen as better immunohistologic stains identify poorly differentiated fibrosarcomas and liposarcomas.[37,47] Twice as many men as women are affected by this disease. Lesions deep in muscle and in the abdomen often recur locally and metastasize, but histologically identical lesions in skin are associated with an excellent prognosis.[46,48] A rare inflammatory MFH variant may be accompanied by fever, a high sedimentation rate, and a neutrophilia or eosinophilia that results from a tumor-produced growth factor.[46] Inflammatory symptoms typically resolve if the tumor is resected. A myxoid variant grows more slowly than the typical MFH lesion but tends to recur locally, thus justifying wide local excision and radiotherapy or even amputation.[46-48]

Leiomyosarcoma

Leiomyosarcomas, which arise in smooth muscle, can develop in the extremities, the trunk, and the abdomen [*see Table 1*]. Leiomyosarcoma is the most common uterine sarcoma and until recently was considered the most common soft tissue sarcoma of the gastrointestinal tract; however, immunohistochemical stains suggest that most GI sarcomas are GISTs [*see* Gastrointestinal Stromal Tumors, *below*].

Patients with uterine leiomyosarcomas present with bleeding or pain.[49,50] Because fibroid tumors are common, sarcoma is often not suspected until the diagnosis is established by pathology. If the uterus is particularly large and a CT scan is obtained for better visualization, a tumor with large areas of necrosis may raise suspicion of sarcoma. Whether leiomyosarcomas of the uterus arise within existing fibroid tumors or develop de novo remains controversial, but most experts believe that no more than a small fraction of 1% of fibroids become malignant.

Leiomyosarcomas of the retroperitoneum and the vena cava occur most commonly in women and may grow rapidly during pregnancy. Tumors originating in other large veins affect women and men equally.[47,51,52]

Rhabdomyosarcomas

Embryonal rhabdomyosarcomas originate in striated muscle and occur most commonly in the orbit in small children and in the genitourinary tract in adolescents [*see Table 1*]. Alveolar rhabdomyosarcomas arise most commonly in the extremities of adolescents and young adults and carry a more serious prognosis.[5,53-55] Pleomorphic rhabdomyosarcomas are occasionally diagnosed, generally in older patients.

Angiosarcomas

Angiosarcomas include tumors previously called hemangiosarcoma, lymphangiosarcoma, and hemangioblastoma. Immunoperoxidase stains from these tumors suggest the same cell of origin.

Angiosarcomas, like other vascular lesions (e.g., benign angiomas), may produce microangiopathic hemolytic anemia (Kasabach-Merritt syndrome), presumably as a result of traumatic injury of red blood cells traversing the tumor bed.[56,57] Angiosarcomas most commonly present as multicentric scalp lesions in older men[58] and as lesions of edematous extremities, typically after mastectomy and lymph node dissection for breast cancer in women.[59] The remainder of the angiosarcomas develop in soft tissue and in the breast, liver, heart, and lungs.[30,57,60-64] Angiosarcomas of the breast occur in young and middle-aged women. Angiosarcomas of the liver arise in adults previously exposed to thorium dioxide (for angiography), arsenic (insecticides), and vinyl chloride (plastics). Angiosarcomas have also arisen in sites previously exposed to radiotherapy (e.g., for treatment of breast cancer, Hodgkin disease, and cervical cancer) and are the most common primary malignant tumors of the myocardium.

Angiosarcomas of the skin resemble bruises and may appear as a single lesion or in clusters. Angiosarcomas in deeper locations have the consistency of a sponge rather than being a distinct mass, and tumors infiltrate substantially beyond their apparent extent. Even well-differentiated angiosarcomas can be difficult to control locally because of their characteristic multifocality and extensive invasion. The prognosis is poor.

Hemangioendothelioma Hemangioendotheliomas are rare vascular tumors of intermediate malignancy (between benign hemangiomas and conventional angiosarcomas) that affect men and women about equally and rarely develop in childhood. A spindle cell variant occurs in younger, predominantly male, patients and principally involves the skin of the hand; it tends to be locally and regionally recurrent.[3]

Hemangiopericytoma Hemangiopericytoma is a malignancy of vascular pericytes (normally found adjacent to the blood vessel endothelium, which is the cell of origin for angiosarcomas and hemangioendotheliomas). Generally diagnosed in patients in their fifth decade of life,[65] 35% of lesions are found in the thigh and 25% in the retroperitoneum; the remaining 40% occur throughout the body. Large tumors can produce arterial venous shunting, bleeding, and even hypoglycemia. Intracranial hemangiopericytomas, formerly designated as angioblastic meningiomas, grow along the sinuses, recur locally, and may metastasize.[3]

Kaposi Sarcoma

Although Kaposi sarcoma is a vascular lesion, its distinct etiology and sensitivity both to radiation and to a wide variety of chemotherapeutic agents distinguish it from other soft tissue sarcomas. Four variants of KS have been identified: classic, transplant associated, lymphadenopathic, and AIDS-KS.

Classification The classic variant presents as multiple blue-red flat skin lesions that indolently progress to nodules and plaques on the lower legs. This variant presents most commonly in elderly men, particularly those of Mediterranean origin. A second variant occurs in patients who receive organ transplants. KS develops an average of 16 months after transplantation. Transplant-associated KS develops most commonly in men, particularly if either the recipient or the organ donor is of Mediterranean extraction. A third aggressive variant is a lymphadenopathic KS that occurs in young children and is endemic to some areas of Africa.

The fourth variant is an aggressive epidemic KS that occurs in AIDS patients. This variant is 20 times more common in men who acquire HIV by sexual contact with men than in those who acquire it by parenteral means.The percentage of AIDS patients with KS ranges from 1% in men with hemophilia to 21% in men who have sex with men.[66] The incidence of KS in men with AIDS in the United States decreased from 40% in 1981 to less than 20% in 1992,[3,67,68] and it is expected to continue to decline with more effective antiretroviral combination therapy.[67]

In 1993, Chang and colleagues identified in an AIDS-KS skin lesion DNA fragments of a previously unrecognized herpesvirus that has alternatively been called Kaposi sarcoma–associated herpesvirus (KSHV) and human herpesvirus type 8 (HHV-8).[33] Over 95% of KS lesions, regardless of source or clinical subtype, have been found to be KSHV-associated. Immunosuppression of the host appears to be an important cofactor in the clinical expression of KS.[3]

KSHV, together with Epstein-Barr virus (EBV), belongs to the gamma herpesviruses.[3] Gamma herpesviruses cause tumors, particularly lymphoproliferative disorders and lymphomas, in humans and animals [*see Table 4*]. KSHV infection rates parallel the global incidence of KS, with low rates in the United States, Europe, and Asia; intermediate rates in Mediterranean countries; and high rates (> 50%) in some locations in Central Africa (e.g., Uganda, Zambia, and South Africa).[69]

Independent of their HIV serostatus, men who have sex with men have a higher rate of KS than the general male population and can have antibodies associated with KS in rates approaching 40%.[70,71]

Diagnosis Clinical manifestations of KS may differ with the variants of the disease. In classic KS, faint reddish-purple macules or patches appear on the feet. HIV-associated KS presents most commonly as cutaneous lesions but also may appear in the oral mucosa or lymph nodes. In contrast to classic KS lesions, HIV-associated KS lesions often begin on the upper body (i.e., face, trunk, or arms). Skin biopsy should be obtained in patients with suspected KS. The histopathologies of all KS variants are similar: spindle-shaped cells in the dermis, with extravasated red blood cells present in slits between irregular vascular spaces [*see 2:X Malignant Cutaneous Tumors*].

Table 4 Herpesviruses Found in Human and Nonhuman Hosts

Gamma herpesviruses	Rhadinovirus Epstein-Barr virus Equine herpesvirus type 2 Herpesvirus saimiri Kaposi sarcoma–associated herpesvirus Lymphocryptovirus
Alpha herpesviruses	Herpes simplex virus type 1 Herpes simplex virus type 2 Equine herpesvirus type 1 Pseudorabies virus Varicella-zoster virus
Beta herpesviruses	Human cytomegalovirus Human herpesvirus type 6 Human herpesvirus type 7

Peripheral Neuroectodermal Tumors and Extraosseus Ewing Sarcoma

Peripheral neuroectodermal tumors (PNETs) and Ewing sarcoma of bone and soft tissue have similar cytogenetic mutations [*see* Translocations Diagnostic of Sarcomas, *above*]. PNETs and extraosseus Ewing sarcomas present as masses most commonly in soft tissue of the chest wall and extremities. They are generally treated with resection, radiation, and chemotherapy in regimens used for Ewing sarcoma of bone [*see* Sarcomas of Bone and Cartilage, Ewing Sarcoma, *below*].

Gastrointestinal Stromal Tumors

GISTs occur predominantly in middle-aged patients but may rarely occur in younger individuals. Approximately 70% of GISTs occur in the stomach, 20% in the small intestine, and less than 10% in the esophagus, colon, and rectum. Survival correlates with tumor location (i.e., patients with GISTs located in the esophagus and stomach survive longer than those with tumors located in the small bowel). Age, mitotic index, and size less than 5 cm are also independent prognostic factors.

Classification Until 1990, most gastrointestinal sarcomas were thought to be leiomyosarcomas because they histologically resembled smooth muscle. However, clinical oncologists observed that GI leiomyosarcomas had a distinctly lower response rate to standard doxorubicin-based chemotherapy than did leiomyosarcomas of the uterus, trunk, and extremities. Subsequent immunocytochemical studies showed that GI sarcomas frequently lacked muscle markers that were typical of leiomyosarcomas found at other anatomic sites. The subset without muscle or Schwann cell (S-100) markers were eventually termed GISTs. These tumors almost universally expressed CD117 (c-*kit*) and often CD34, which are also expressed on hematopoietic progenitor cells.

Genetic factors Most GISTs express c-*kit*, a growth factor receptor with tyrosine kinase activity. c-*kit* is a proto-oncogene located on the long arm of chromosome 4 (4q11-q12) that encodes a transmembrane tyrosine kinase receptor. Its ligand is a stem cell factor. c-*kit* regulates normal hematopoiesis, as well as migration of germ cells, and is also expressed in normal human mast cells, immature myeloid cells, melanocytes, epithelial breast cells, and interstitial cells of Cajal (the gastrointestinal pacemaker cells). Interstitial cells of Cajal have immunohistochemical profiles similar to those of GISTs (i.e., c-*kit* and CD34 positive; desmin and S-100 negative) and thus are currently thought to be the cell of origin of GISTs.

c-*kit* mutations[72] in the juxtamembrane domain of GISTs include in-frame deletions and point mutations in exon 11 in approximately 60% of cases.[73] Mutations in most of the remaining cases involve exon 13 and exon 9. The mutated receptor is activated constitutively without its ligand, driving neoplastic transformation.

Diagnosis The most common symptoms of GIST at initial presentation are abdominal pain and gastrointestinal bleeding. CT and MRI scans may suggest the diagnosis and define the extent of the lesion. Diagnosis is typically confirmed histologically. The Carney triad includes the diagnosis of gastric stromal tumors (formerly called epithelioid leiomyosarcoma), functioning extra-adrenal paragangliomas, and pulmonary hamartomas (chondromas).[74-76] Patients with one or two components of the triad should be evaluated for the other components. Pulmonary lesions detected on radiographs may be either metastases or benign hamartomas; biopsy must be performed before chemotherapy is administered for presumed metastatic disease.[3]

Mesothelioma

The annual incidence of mesothelioma is not known with certainty. Mesotheliomas are difficult to diagnose, even by expert pathologists. Because cancer deaths are coded by primary site of the neoplasm (primary neoplasms of pleura and peritoneum) and not by histologic diagnosis (mesothelioma), data from death certificates are unreliable for estimating disease frequency despite the rapid progression to fatal outcome that is typical of malignant mesothelioma. A reasonable estimate is that about 2,000 new cases of mesothelioma occur annually in the United States (range, 1,000 to > 3,000 cases).[77] Reported rates have increased in the past decade, perhaps by as much as 50%.[78] In the United States, mesothelioma is approximately threefold to fivefold more common in men than in women.[78-80]

Etiology Persons at high risk for mesothelioma are those exposed to the processing and commercial uses of asbestos. The mineral is mined, milled, and used in a wide range of industrial and commercial products, including insulation, textiles, heat protectors, filters, and construction materials (e.g., spackling, roofing, siding, and floor and ceiling tiles).[80] Workers who are exposed to high levels of asbestos include miners, millers, producers of asbestos products, and laborers who install plumbing, boilers, and heating equipment. Workers who may not handle asbestos directly but are in proximity to others working with the material (e.g., carpenters, electricians, and welders) are also at risk.

Malignant mesothelioma is rarely curable; therefore, screening of asbestos workers for mesothelioma is inappropriate.[81] However, smoking greatly increases the risk of lung cancer (but not mesothelioma) in asbestos workers, and smoking cessation efforts are needed in this high-risk group.[81,82] A detailed exposure history should emphasize the period 20 to 50 years before diagnosis and should include possible household-contact exposure. Brief exposures may be long forgotten, and asbestos exposure can be documented in about 30% to 70% of patients with mesothelioma.[77]

Malignant pleural mesothelioma most commonly develops in the fifth to seventh decades of life (median age, 60 years). A significant proportion of patients with mesothelioma in whom the diagnosis was made between 20 and 40 years of age report household or neighborhood asbestos exposure during childhood. Children who present with the disease generally have no apparent asbestos exposure.[80]

Diagnosis Dyspnea, chest pain, or both prompt patients to seek medical attention. Physical examination is remarkable for dullness at one lung base, and chest x-rays typically reveal a large, freely movable unilateral pleural effusion. Occasionally, patients are asymptomatic, and the pleural effusion is found incidentally on chest x-ray. Sixty percent of patients have right-sided lesions, and fewer than 5% have bilateral involvement at the time of diagnosis.

A CT or MRI scan of the chest to assess the extent of disease is indicated if any treatment is contemplated. Loss of lung volume is evident early on CT. In advanced disease, scoliosis with contracture of the ipsilateral hemithorax is visible even on chest x-ray. Despite a history of asbestos contact in 50% to 70% of patients, pleural plaques or interstitial fibrosis is apparent on chest x-ray in only about 20% of patients; however, pleural calcifications are evident on CT scan in nearly 50% of patients and at autopsy in 87%.

Initial misdiagnosis is common, and pathologic opinion appears particularly diverse in cases in which litigation is involved. Special immunoperoxidase staining has made the diagnosis of mesothelioma more reliable. Because a substantial percentage of mesotheliomas develop in patients with no known asbestos exposure and because other malignancies are common in asbestos workers, asbestos exposure alone should not influence the diagnosis of mesothelioma.

TREATMENT OF SOFT TISSUE SARCOMAS

Patients with possible sarcoma should be referred to an institution with an experienced sarcoma service. If preoperative evaluation reveals no evidence of metastasis on chest CT, the goal of treatment is the control of local disease. Primary therapy for most relatively small localized soft tissue sarcomas (except for rhabdomyosarcoma and Kaposi sarcoma) is surgery. Wide excision (3 to 6 cm of pathologically documented normal tissue) is optimal to avoid the need for radiotherapy, particularly for abdominal or low-grade lesions. Careful pathologic examination of the margins is essential to document adequate resection and identify any involved margins that require further resection. If lesions are deep or adjacent to a major nerve, visceral organ, or bone, surgical margins may be necessarily narrow (e.g., 1 mm or less) or even involved in tumor. When pathologic margins are small (< 3 to 6 cm), radiation may be used to decrease the risk of local failure. Adequate radiotherapy (6,600 cGy or more) can usually be delivered to lesions located in the extremities; however, the risk of radiation damage to visceral organs generally precludes these doses for treatment of tumors located in the retroperitoneum or abdomen. Low-grade lesions, if ade-

Table 5 Meta-analysis of Doxorubicin-Based Adjuvant Chemotherapy for Localized Resectable Soft Tissue Sarcoma*

	Hazard Rate	*P-Value*	*Absolute Benefit at 10 Years (%)*
Local recurrence-free intervals	0.73	0.016	6 (from 75 to 81)
Distant recurrence-free intervals	0.70	0.0003	10 (from 60 to 70)
Overall survival	0.89	0.12	4 (from 50 to 54)

*The findings are based on 14 randomized trials involving 1,568 patients.

quately resected, are generally curable; thus, wide surgical margins are especially important to ensure local control. Very large lesions are often initially treated with radiation and chemotherapy to facilitate ultimate resection.

Superficial trunk lesions are frequently resectable with wide margins. For extremity lesions, a combined modality of conservative (often limb-sparing) surgery and radiotherapy may achieve local control without increasing the risk of locoregional recurrence.[83-85] The current standard of limb-sparing surgery combined with radiotherapy or radical resection (often amputation), as practiced at experienced centers, has improved local control from approximately 20% in 1950 to 90% or higher today.[83] Local control rates remain 50% to 75% for trunk lesions but only 30% to 50% for primary retroperitoneal lesions.[86-91] Many primary retroperitoneal and head and neck lesions involve vital structures and are unresectable at presentation.

Adjuvant Chemotherapy

Adjuvant chemotherapy is standard practice for the treatment of rhabdomyosarcomas, osteosarcomas, and Ewing sarcomas, but it remains controversial for other adult soft tissue sarcomas. Nevertheless, most experts agree that the risks of adjuvant chemotherapy are not warranted for low-grade lesions (given their low probability of metastatic spread) and for lesions less than 5 cm, even if high grade, because of their excellent prognosis.

Initial pilot studies investigating the effectiveness of adjuvant chemotherapy in adult soft tissue sarcomas reported significant survival benefit compared to survival reported in historical control subjects. However, subsequent randomized studies found that survival with no adjuvant chemotherapy to be higher than previously reported. Because surgical resection of recurrent local disease and even pulmonary metastases is sometimes curative, disease-free survival may be a less meaningful end point than overall survival.

A meta-analysis of individual patient data from all randomized doxorubicin-based adjuvant trials in patients with sarcoma included data from 1,568 patients [*see Table 5*]. Adjuvant chemotherapy improved the time to local and distant recurrence-free survival (hazard ratios, 0.73 and 0.70; $P = 0.016$ and 0.0003, respectively). However, overall survival in the group receiving chemotherapy showed only a 4% advantage and a hazard ratio of 0.89 ($P = 0.12$).[92] Adjuvant trials of doxorubicin and ifosfamide regimens have produced mixed results.[93,94]

Therapy for Advanced Soft Tissue Sarcomas

Resection Resection of subpleural metastases in carefully selected patients results in disease-free survival in about 20% of patients.[95-97] Patients with a disease-free interval longer than 12 months, fewer than five nodules, and a tumor doubling time of more than 20 days have the best prognosis. In one large study, the median disease-free survival for all patients undergoing repeat resections was 42.8 months, with an estimated 5-year survival of 36%.[96] Prolonged survival was associated with fewer than three nodules at the time of the second resection. A combined-modality approach (chemotherapy before resection) has yet to be evaluated in a randomized trial.

Single-agent chemotherapy Single-agent chemotherapy is generally chosen for palliation, particularly in older patients and in those with low-grade lesions. Doxorubicin is the most active single agent in soft tissue sarcomas, with a response rate of 15% to 35% in various studies.[98-100]

A dose-response relationship has been observed for doxorubicin in randomized trials.[101-103] Dose rates of 60 to 70 mg/m^2 every 3 weeks are generally superior to dose rates of 50 mg/m^2 or less.

Dacarbazine has a single-agent response rate of about 16% for most sarcomas and response rates of up to 44% when used in combination with doxorubicin in the treatment of leiomyosarcomas.[104,105] Nausea, which is a common side effect, may be improved by continuous infusion administration.[106]

Ifosfamide, an analogue of cyclophosphamide, is effective against sarcoma alone or in combination with doxorubicin.[107,108] Mesna must be administered in conjunction with ifosfamide to prevent bladder injury.

Combination chemotherapy Combination chemotherapy is generally used (1) for patients for whom a response would facilitate curative resection of a primary lesion and (2) for patients with unresectable or metastatic sarcoma. Combination chemotherapy is also used as palliation, particularly in patients younger than 50 years who have high-grade lesions. A combination of doxorubicin and ifosfamide produces response rates of 17% to 25%.[98,109,110] The median survival for patients with unresectable or metastatic sarcoma is 12 months, with no significant differences in survival.

Treatment of Nonclassic Soft Tissue Sarcomas

Rhabdomyosarcomas Children with rhabdomyosarcoma are generally treated with multiagent chemotherapy in regimens combining vincristine, actinomycin, and either cyclophosphamide or ifosfamide.[5,53-55] Doxorubicin-based regimens are generally used in adults.[111] Lymph node metastases occur more commonly than in other sarcomas. Although children with metastatic disease sometimes can be cured, the prognosis in adults is poor despite excellent responses to primary aggressive chemotherapy.[111] Chemotherapy is followed by resection of residual rhabdomyosarcoma and radiation therapy.

Kaposi sarcoma Lesions of the classic variant of KS generally respond to resection, radiotherapy, or low doses of vinblastine, doxorubicin, or interferon alfa.[111] Secondary tumors, espe-

cially lymphomas, are common. Transplantation-associated KS may be controlled by modification or cessation of immunosuppression therapy. Although HIV-associated KS responds to radiation therapy and chemotherapy, the response is less durable than in classic KS. HIV-positive patients with KS lesions involving limited areas of the skin or oral mucosa are often most easily treated with radiation. Advanced or visceral disease may be treated with cytotoxic chemotherapy using doxorubicin, vinblastine, or paclitaxel in single-agent or combination regimens; interferon alfa is usually used as a single agent.[111]

Mesothelioma No therapy for mesothelioma is considered standard. Cure is elusive even with resection, radiation, and combination chemotherapy.[112,113] Tumor volume shrinkage of at least 50% is achieved in 15% to 25% of selected patients treated with gemcitabine- and cisplatin-based regimens.[114,115]

Gastrointestinal Stromal Tumors

Treatment of GIST has changed dramatically with improved understanding of its biology. Complete en bloc surgical resection, when possible, remains the cornerstone of curative therapy. Recurrent or metastatic disease carries a poor prognosis because of GIST resistance to chemotherapy and radiotherapy.

Imatinib is an oral, relatively specific inhibitor of three tyrosine kinase receptors: (1) the protein produced by the Philadelphia chromosome translocation of chronic myelogenous leukemia, (2) the mutated c-*kit* receptor in GISTs, and (3) the platelet-derived growth factor receptor.[116] In a European trial evaluating doses of imatinib ranging from 300 to 1,000 mg orally daily, tumor growth was inhibited in 32 of 36 patients, and in 19 patients tumor volume decreased more than 50%.[117] In a study in the United States, 36 patients with unresectable or metastatic GIST were randomized to receive either 400 or 600 mg of imatinib daily; of these, 19 (54%) had a partial response of at least a 50% decrease in the size of the lesion; no advantage was shown for either dose. Disease progression occurred in four patients (11%).[118] In the European dose-escalation study, nausea and vomiting, edema, and dyspnea precluded administration of doses of more than 500 mg twice daily. The most common side effects (generally resolving after the first 8 weeks) were nausea and vomiting, rash, edema (particularly periorbital), and conjunctivitis (rarely with bleeding sclera). Myelosuppression was infrequent, although anemia did occur. Intratumoral and gastrointestinal bleeding developed in fewer than 5% of patients.

Sarcomas of Bone and Cartilage

Sarcomas are the second most common primary tumor of bone, after myeloma. The most common sarcomas of bone are osteosarcoma, Ewing sarcoma, and chondrosarcoma. Bone sarcomas currently represent 1% of adult malignancies and 15% of pediatric malignancies. In 2003, an estimated 4,000 persons in the United States will have developed sarcoma of bone, and 2,000 will have died of this disease.[4]

Primary malignant lesions of bone (arising de novo) and secondary malignancies (arising from a prior benign tumor) must be distinguished from the vastly more common metastases to bone. Radiographs of a lesion can help distinguish benign from malignant lesions and often suggest the histologic origin of the tumor. Osteosarcoma characteristically produces irregular cortical surfaces, whereas Ewing sarcoma has a layered appearance like that of an onion skin.

OSTEOSARCOMA

Osteosarcoma, defined as a primary malignancy that produces osteoid, is the most common sarcoma of bone. The incidence ratio of men to women is 1.5:1 to 2:1. The age distribution is bimodal, with the first peak in the second and third decades of life and the second in the sixth decade.[6] Osteosarcoma classically arises in growth plates (i.e., around the epiphyses) of long bones during the adolescent growth spurt.

Osteosarcoma in older adults often develops in previously irradiated sites or in existing benign bone lesions, such as pagetoid bone, solitary osteochondromas, or multiple enchondromatosis (Ollier disease). The risk of sarcoma is 0.2% in patients with Paget disease. Axial lesions are more common in adults than in children, occurring in fewer than 10% of pediatric patients but in 30% to 50% of adult patients. Extraosseous presentations are rare and occur in older adults.

Diagnosis

Clinical manifestations Patients with osteosarcoma present with severe pain of relatively short duration, often after a sports injury. Osteosarcoma may present as an obvious firm-to-hard mass covered by stretched, shiny skin, with prominent vascular markings, or may be evident only on radiograph. Multicentric osteosarcomas have developed in watch-dial painters who are subject to chronic radium ingestion and occasionally in persons without risk factors, most commonly in children younger than 10 years.

Imaging studies Radiographs and CT scans reveal osteolysis [*see Figure 1*], periosteal new bone formation, and late cortical destruction. Periosteal elevation may produce the classic Codman triangle on radiograph. Radiologically, the differential diagnosis includes osteochondroma and myositis ossificans.

Pathology Superficial, low-grade juxtacortical or parosteal osteosarcomas constitute only 3% to 4% of osteosarcomas, arise equally in males and females, and tend to affect adults with no prior risk factors. Low-grade osteosarcomas typically present in the third and fourth decades of life, whereas high-grade osteosarcomas appear in the second decade. Grossly, these bulky tumors encircle the cortex of bone (generally, the distal femur and, less commonly, the proximal humerus).

Staging Evaluation

Staging evaluation includes radiographs and CT scans of the primary lesion and the lungs, which is the most likely site of metastases. Bone scans generally reveal intense uptake in the lesion and may detect additional lesions more proximal than the dominant lesion in the same bone (so-called skip lesions), metastases to other bones, or multicentric primary tumors; pulmonary uptake suggests lung metastases. Levels of alkaline phosphatase at diagnosis, which are generally elevated except in very undifferentiated lesions, correlate with prognosis, and elevated values after amputation predict residual or relapsing disease.[119]

Patients in whom osteosarcoma is possible should be referred to an institution with an experienced sarcoma service. Expert multimodal consultation is essential before biopsy. Incisional biopsy should be performed by the surgeon who will do the definitive resection, because an improperly placed biopsy, particularly of proximal tibial lesions, may render limb-sparing surgery impossible and significantly compromise local control.

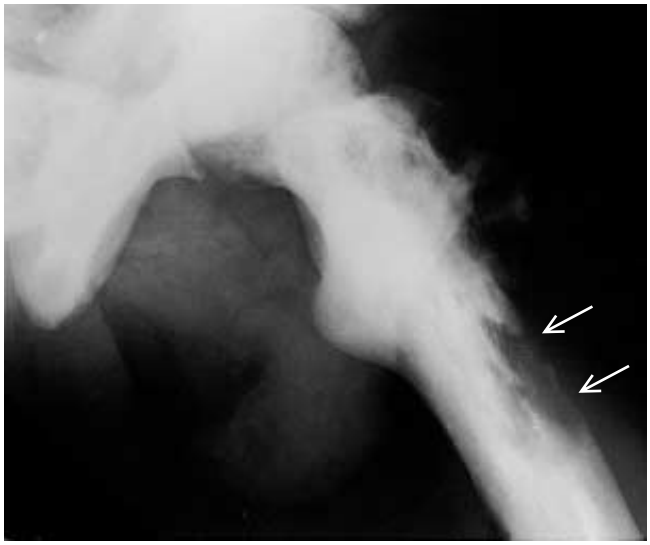


Figure 1 **Radiography reveals osteosarcoma in a 62-year-old man with Paget disease of bone. The patient had pain associated with a soft tissue mass in the area of the greater trochanter of the left femur. Note the destruction of the cortex by the osteosarcoma in pagetic bone. The position of the original cortex is indicated by the arrows.**

If the radiologic picture is virtually diagnostic, a fine-needle biopsy may be adequate to confirm the clinical diagnosis.

Treatment

Adjuvant chemotherapy On the basis of two small, randomized trials in which combination regimens were compared with observation, doxorubicin- and cisplatin-based adjuvant chemotherapy is the established treatment for osteosarcoma. The Pediatric Oncology Group and UCLA studies showed a significantly improved 2-year disease-free survival; the UCLA study also demonstrated a significantly prolonged overall survival.[120,121] The optimal combination of agents and schedule has not yet been established. Of six randomized trials evaluating high-dose methotrexate, one demonstrated an advantage for high-dose over standard-dose methotrexate, whereas five showed no advantage for high-dose methotrexate.[122,123] In a study by the European Osteosarcoma Intergroup, short intensive chemotherapy using doxorubicin and cisplatin was compared with a complex multiagent protocol in patients with operable osteosarcoma. Progression-free survival at 5 years was 44% in both groups, and the two groups had comparable overall survival.[124,125] In a follow-up study, the addition of ifosfamide did not appear to improve these results.[126]

Preoperative chemotherapy Preoperative chemotherapy has many theoretical advantages over postoperative adjuvant chemotherapy.

Early systemic treatment may more effectively eradicate microscopic metastatic deposits. The response to preoperative chemotherapy can be evaluated histologically, and the regimen can be modified if the response is suboptimal. Patients receiving preoperative chemotherapy whose resected tumor specimen is more than 90% necrotic may have a significantly higher survival rate than patients receiving only postoperative chemotherapy.[127] Patients who have less than 90% necrosis in the resected tumor after preoperative chemotherapy may have increased disease-free survival if switched to other effective agents postoperatively. Because resolution of tumor vascularity on angiography appears to correlate with histologic necrosis, monitoring angiographic findings may maximize tumor response to chemotherapy before surgery.[128]

Preoperative chemotherapy may facilitate limb-sparing surgery, preservation of muscle groups, and limb function. The delay in surgery also allows procurement of a correctly sized internal prosthesis for replacement of excised bone. However, patients receiving preoperative chemotherapy must be regularly evaluated and should undergo prompt resection if the tumor increases in size during chemotherapy.

Radiotherapy Osteosarcoma tends to be markedly radioresistant. Radiotherapy alone has not been shown to improve the rate of successful limb-sparing surgery or decrease the risk of local recurrence in tumors with close or positive surgical margins. However, in a randomized, controlled trial, higher doses of radiotherapy (given preoperatively with intra-arterial doxorubicin) significantly increased the rate of local complications.[120]

Surgical treatment Either amputation or limb-sparing surgery may be appropriate as long as lesions are widely excised with several centimeters of pathologically documented uninvolved margins. Lesions with a significant soft tissue component or neurovascular involvement generally necessitate amputation. Osteosarcomas arising in sites of prior Paget disease or irradiation tend to be vascular and centrally located and thus are difficult to resect for cure.

Although functional results after limb-sparing procedures are excellent in 60% to 75% of patients, relapse or complications from treatment may later necessitate amputation. In appropriately selected patients, local control rates (i.e., 90% to 97%) and survival rates for limb-sparing procedures are similar to those for amputation.[120,128] Limb-sparing resection tends to be more successful for upper-extremity lesions than for lower-extremity lesions because upper-extremity lesions are associated with fewer complications from weight bearing and dependent edema. In children, an internal prosthesis whose length can be adjusted as the child grows can reduce limb-length disparity. Fractures remain an important complication in active individuals, and thus, amputations are actually preferable for individuals who wish to pursue contact sports.[120]

Treatment of metastatic disease Relapses tend to occur in the first 2 years after completion of primary therapy. Major prognostic variables include grade and invasion through the bone cortex to involve soft tissue. Metastases most frequently develop in lung and, less commonly, in bone. Patients who have already received adjuvant chemotherapy may respond to the same drugs if metastases develop more than 6 months after the last adjuvant treatment. Between 20% and 40% of selected patients with five or fewer pulmonary lesions and a disease-free interval of more than 12 months can be cured with combination chemotherapy and surgical resection.

CHONDROSARCOMA

Chondrosarcoma, which represents about 20% of bone sarcomas, is the most common bone sarcoma after osteosarcoma. Extraosseous chondrosarcomas arising in soft tissue are rare.[129] Primary chondrosarcomas occur in previously normal bone; secondary chondrosarcomas arise in prior benign lesions, generally

enchondromas.[130] Chondrosarcomas in patients with multiple enchondromatosis (Ollier disease) are generally low grade, whereas chondrosarcomas associated with soft tissue hemangiomas (Maffucci disease) are frequently high grade. Chondrosarcoma represents about 10% of radiation-associated sarcomas and may arise in pagetoid bone.

The incidence of chondrosarcomas increases steadily with age.[6] The most common sites of involvement are the pelvis (31%), femur (21%), shoulder (13%), face (9%), and ribs (9%).[6] Lesions can be painless but tend to be painful if they grow rapidly.

Diagnosis

On radiograph, central chondrosarcomas appear as popcorn-shaped calcifications. Peripheral chondrosarcomas may appear as long, slightly calcified spicules radiating from the cortex to a flattened outer surface; there may be little cortical or medullary involvement and a faint Codman triangle caused by a disruption of the smooth periosteal surface. Bone scans generally demonstrate intense uptake in the lesion.

Chondrosarcomas are defined as malignant stromal tumors of bone that produce cartilage but no osteoid. Histologically, the tumors have the appearance of cartilage with malignant chondrocytes. Low histologic grade is less reliable than patient age, site of origin, lesion size, and radiographic appearance in determining biologic behavior.

Mesenchymal chondrosarcoma is a histologic variant that tends to arise in the ribs, mandible, maxilla, skull, and extraosseous sites; it is characterized by a small round cell histologic component and has some sensitivity to chemotherapy.

Treatment

Chondrosarcomas, which are relatively resistant to both radiotherapy and chemotherapy, must be adequately resected at the time of initial diagnosis. Aggressive resection is particularly appropriate for eminently curable low-grade chondrosarcomas. Local recurrences tend to increase with histologic aggressiveness.

EWING SARCOMA

Ewing sarcoma constitutes 10% to 14% of primary malignant bone tumors in whites but is rare in other races.[8] The peak incidence occurs between 10 and 25 years of age (range, 2 to 65 years), and the risk in men is twice that in women [*see Table 1*].[6] Ewing sarcoma and osteosarcoma occur in the same sex and age groups, but they can usually be distinguished radiographically and histologically.[5]

Diagnosis

Patients with Ewing sarcoma present with fever, weight loss, malaise, poorly localized bone pain, and a rapidly enlarging mass. Leukocytosis, fever, and an elevated erythrocyte sedimentation rate mimic osteomyelitis. The most common sites of involvement are femur (27%), pelvis (18%), and tibia and fibula (17%).[6] Diagnosis of pelvic Ewing sarcoma is frequently delayed because pain is poorly localized and the mass is not clinically apparent.

Histologically, Ewing sarcoma on reticulin stain appears as clusters of small blue cells bordered by fibrous septa (in contrast to lymphoma cells, which have no septa).[83] Cytogenetics generally reveal the t(11;22) translocation, which is identical to that found in PNETs, suggesting a possible neurogenic origin of Ewing sarcoma.[19]

Radiographically, Ewing sarcoma forms a fusiform enlargement of the long bones, with onion-skin layering of the periosteum and central mottling (so-called cracked-ice appearance). A pulmonary CT scan, bone scan, and marrow aspirate and biopsy are necessary for staging.

Clinically detectable metastases are present in about one third of patients at diagnosis. The most frequent sites of metastasis are the lung, bone, and marrow. Vertebral metastases commonly lead to spinal cord compression. Ewing sarcoma results in 90% mortality in patients who undergo surgical resection alone, which suggests that systemic disease may be present in a substantially higher percentage of patients than those who have detectable metastases at presentation.

Treatment

Initial treatment consists of a multiagent chemotherapy regimen of doxorubicin, vincristine, and cyclophosphamide, alternating with etoposide and ifosfamide administered before and during radiotherapy to the involved bone.[111,131] Radiotherapy concentrating on the primary lesion, if begun during the fourth or fifth cycle of chemotherapy, controls local disease in 70% of patients and provides good functional results. Because of a high risk of local failure and secondary radiation-associated osteosarcomas, expendable bones (e.g., rib, tibia) are often resected, particularly if biopsy after radiotherapy and chemotherapy reveals residual viable tumor.

Survival rates correlate inversely with age (i.e., 70% for patients younger than 10 years and 46% for those older than 16 years).[132] Lower survival rates are associated with primary lesions of the pelvis, humerus, and rib; high levels of lactate dehydrogenase; and extensive soft tissue involvement. One third of pediatric patients with metastases are cured by current therapy, but few adults with metastases survive.

OTHER BONE SARCOMAS

Fibrosarcoma, chordoma, angiosarcomas, and MFHs of bone are usually treated with resection. Radiation to the site of the lesion and adjuvant chemotherapy with doxorubicin and either cisplatin or ifosfamide may also be used.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter. Drugs discussed in this chapter do not have FDA approval for use in the treatment of sarcoma.

References

1. Keohan ML, Hibshoosh H, Antman KH: Gastrointestinal sarcomas. Gastrointestinal Cancers: A Companion to Sleisenger and Fordtran's Gastrointestinal and Liver Disease. Saunders Elsevier Science Limited, Philadelphia, 2003, p 721

2. Antman KH: Uterine sarcomas. Gynecologic Oncology. Knapp R, Bercowitz R, Eds. Macmillan, New York, 1986, p 297

3. Antman K, Chang Y: Kaposi's sarcoma. N Engl J Med 342:1027, 2000

4. Jemal A, Murray T, Samuels A, et al: Cancer statistics, 2003. CA Cancer J Clin 53:5, 2003

5. Crist WM, Anderson JR, Meza JL, et al: Intergroup rhabdomyosarcoma study-IV: results for patients with nonmetastatic disease. J Clin Oncol 19:3091, 2001

6. Fraumeni JF Jr, Boice JD Jr: Bone. Cancer Epidemiology and Prevention. Schottenfeld D, Fraumeni JF Jr, Eds. WB Saunders Co, Philadelphia, 1982, p 814

7. Grimer RJ, Cannon SR, Taminiau AM, et al: Osteosarcoma over the age of forty. Eur J Cancer 39:157, 2003

8. Guo W, Xu W, Huvos AG, et al: Comparative frequency of bone sarcomas among different racial groups. Chin Med J (Engl) 112:1101, 1999

9. Ozaki T, Schaefer KL, Wai D, et al: Population-based genetic alterations in Ewing's tumors from Japanese and European Caucasian patients. Ann Oncol 13:1656, 2002

10. Li FP, Fraumeni JF Jr: Soft-tissue sarcomas, breast cancer, and other neoplasms: a familial syndrome? Ann Intern Med 71:747, 1969

11. Malkin D, Li F, Strong L, et al: Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 250:1233, 1990

12. Evans DG, Birch JM, Thorneycroft M, et al: Low rate of TP53 germline mutations in breast cancer/sarcoma families not fulfilling classical criteria for Li-Fraumeni syn-

drome. J Med Genet 39:941, 2002

13. Huson SM: What level of care of the neurofibromatoses? Lancet 353:1114, 1999

14. Reynolds RM, Browning GG, Nawroz I, et al: Von Recklinghausen's neurofibromatosis: neurofibromatosis type 1. Lancet 361:1552, 2003

15. Francois J: Retinoblastoma and osteogenic sarcoma. Ophthalmologica 175:185, 1977

16. Wong FL, Boice JD Jr, Abramson DH, et al: Cancer incidence after retinoblastoma: radiation dose and sarcoma risk. JAMA 278:1262, 1997

17. Bovee JV, Sakkers RJ, Geirnaerdt MJ, et al: Intermediate grade osteosarcoma and chondrosarcoma arising in an osteochondroma: a case report of a patient with hereditary multiple exostoses. J Clin Pathol 55:226, 2002

18. Kaushik S, Smoker WR, Frable WJ: Malignant transformation of fibrous dysplasia into chondroblastic osteosarcoma. Skeletal Radiol 31:103, 2002

19. Whang-Peng J, Triche TJ, Knutsen T, et al: Chromosome translocation in peripheral neuroepithelioma. N Engl J Med 311:584, 1984

20. Ginsberg JP, de Alava E, Ladanyi M, et al: EWS-FLI1 and EWS-ERG gene fusions are associated with similar clinical phenotypes in Ewing's sarcoma. J Clin Oncol 17:1809, 1999

21. Kawai A, Woodruff J, Healey JH, et al: SYT-SSX gene fusion as a determinant of morphology and prognosis in synovial sarcoma. N Engl J Med 338:153, 1998

22. Hibshoosh H, Lattes R: Immunohistochemical and molecular genetic approaches to soft tissue tumor diagnosis: a primer. Semin Oncol 24:515, 1997

23. Lagrange JL, Ramaioli A, Chateau MC, et al: Sarcoma after radiation therapy: retrospective multiinstitutional study of 80 histologically confirmed cases. Radiation Therapist and Pathologist Groups of the Federation Nationale des Centres de Lutte Contre le Cancer. Radiology 216:197, 2000

24. Blanchard DK, Reynolds C, Grant CS, et al: Radiation-induced breast sarcoma. Am J Surg 184:356, 2002

25. Patel SR: Radiation-induced sarcoma. Curr Treat Options Oncol 1:258, 2000

26. Karlsson P, Holmberg E, Samuelsson A, et al: Soft tissue sarcoma after treatment for breast cancer: a Swedish population-based study. Eur J Cancer 34:2068, 1998

27. Tucker M, D'Angio G, Boice JJ, et al: Bone sarcomas linked to radiotherapy and chemotherapy in children. N Engl J Med 317:588, 1987

28. Roggli VL, Sharma A, Butnor KJ, et al: Malignant mesothelioma and occupational exposure to asbestos: a clinicopathological correlation of 1445 cases. Ultrastruct Pathol 26:55, 2002

29. Ward E, Boffetta P, Andersen A, et al: Update of the follow-up of mortality and cancer incidence among European workers employed in the vinyl chloride industry. Epidemiology 12:710, 2001

30. Centeno JA, Mullick FG, Martinez L, et al: Pathology related to chronic arsenic exposure. Environ Health Perspect 110(suppl 5):883, 2002

31. Kogevinas M, Kauppinen T, Winkelmann R, et al: Soft tissue sarcoma and non-Hodgkin's lymphoma in workers exposed to phenoxy herbicides, chlorophenols, and dioxins: two nested case-control studies. Epidemiology 6:396, 1995

32. Hisada M, Biggar RJ, Greene MH, et al: Solid tumors after chronic lymphocytic leukemia. Blood 98:1979, 2001

33. Chang Y, Cesarman E, Pessin MS, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma [see comments]. Science 266:1865, 1994

34. Lee E, Locker J, Nalesnik M, et al:: The association of Epstein-Barr virus with smooth-muscle tumors occurring after organ transplantation. N Engl J Med 332:19, 1995

35. McClain KL, Leach CT, Jenson HB, et al: Association of Epstein-Barr virus with leiomyosarcomas in children with AIDS [see comments]. N Engl J Med 332:12, 1995

36. McNairn JD, Damron TA, Landas SK, et al: Inheritance of osteosarcoma and Paget's disease of bone: a familial loss of heterozygosity study. J Mol Diagn 3:171, 2001

37. Coindre JM, Mariani O, Chibon F, et al: Most malignant fibrous histiocytomas developed in the retroperitoneum are dedifferentiated liposarcomas: a review of 25 cases initially diagnosed as malignant fibrous histiocytoma. Mod Pathol 16:256, 2003

38. Spillane A, Thomas JM, Pisters PW, et al: Misconceptions with staging of soft tissue sarcoma. J Clin Oncol 18:1800, 2000

39. Wunder JS, Healey JH, Davis AM, et al: A comparison of staging systems for localized extremity soft tissue sarcoma. Cancer 88:2721, 2000

40. Greene FL, Page DL, Fleming ID: American Joint Committee on Cancer (AJCC) Cancer Staging Manual, 6th ed. Springer-Verlag, New York, 2002

41. Vraa S, Keller JO, Nielsen OS, et al: Prognostic factors in soft tissue sarcomas: the Aarhus experience. Eur J Cancer 34:1876, 1998

42. Barile A, Zugaro L, Catalucci A, et al: Soft tissue liposarcoma: histological subtypes, MRI and CT findings. Radiol Med (Torino)104:140, 2002

43. Pisters PW, Sondack VK: Metastatic patterns of extremity liposarcoma and their outcome. J Surg Oncol 80:94, 2002

44. Okcu MF, Munsell M, Treuner J, et al: Synovial sarcoma of childhood and adolescence: a multicenter, multivariate analysis of outcome. J Clin Oncol 21:1602, 2003

45. Ladanyi M, Antonescu CR, Leung DH, et al: Impact of SYT-SSX fusion type on the clinical behavior of synovial sarcoma: a multi-institutional retrospective study of 243 patients. Cancer Res 62:135, 2002

46. Kartha SS, Bumpous JM: Synovial cell sarcoma: diagnosis, treatment, and outcomes. Laryngoscope 112:1979, 2002

47. Belal A, Kandil A, Allam A, et al: Malignant fibrous histiocytoma: a retrospective study of 109 cases. Am J Clin Oncol 25:16, 2002

48. Daw NC, Billups CA, Pappo AS, et al: Malignant fibrous histiocytoma and other fibrohistiocytic tumors in pediatric patients: the St. Jude Children's Research Hospital experience. Cancer 97:2839, 2003

49. Giuntoli RL II, Metzinger DS, DiMarco CS, et al: Retrospective review of 208 patients with leiomyosarcoma of the uterus: prognostic indicators, surgical management, and adjuvant therapy. Gynecol Oncol 89:460, 2003

50. Hsieh CH, Lin H, Huang CC, et al: Leiomyosarcoma of the uterus: a clinicopathologic study of 21 cases. Acta Obstet Gynecol Scand 82:74, 2003

51. Clary BM, DeMatteo RP, Lewis JJ, et al: Gastrointestinal stromal tumors and leiomyosarcoma of the abdomen and retroperitoneum: a clinical comparison. Ann Surg Oncol 8:290, 2001

52. Hartman DS, Hayes WS, Choyke PL, et al: From the archives of the AFIP: leiomyosarcoma of the retroperitoneum and inferior vena cava: radiologic-pathologic correlation. Radiographics 12:1203, 1992

53. Esnaola NF, Rubin BP, Baldini EH, et al: Response to chemotherapy and predictors of survival in adult rhabdomyosarcoma. Ann Surg 234:215, 2001

54. Hawkins WG, Hoos A, Antonescu CR, et al: Clinicopathologic analysis of patients with adult rhabdomyosarcoma. Cancer 91:794, 2001

55. Little DJ, Ballo MT, Zagars GK, et al: Adult rhabdomyosarcoma: outcome following multimodality treatment. Cancer 95:377, 2002

56. Alliot C, Tribout B, Barrios M, et al: Angiosarcoma variant of Kasabach-Merritt syndrome. Eur J Gastroenterol Hepatol 13:731, 2001

57. Moussa SH, Oliveira AL, de Amorim AP, et al: Angiosarcoma of the breast associated with Kasabach-Merritt syndrome. Arch Gynecol Obstet 267:43, 2002

58. Grobmyer SR, Daly JM, Glotzbach RE, et al: Role of surgery in the management of postmastectomy extremity angiosarcoma (Stewart-Treves syndrome). J Surg Oncol 73:182, 2000

59. Fata F, O'Reilly E, Ilson D, et al: Paclitaxel in the treatment of patients with angiosarcoma of the scalp or face. Cancer 86:2034, 1999

60. Fant J, Grant M, May S, et al: Angiosarcoma of the breast: mammographic, clinical, and pathologic correlation. Breast J 9:252, 2003

61. Georgiannos SN, Sheaff M: Angiosarcoma of the breast: a 30 year perspective with an optimistic outlook. Br J Plast Surg 56:129, 2003

62. Lewis R, Rempala G: A case-cohort study of angiosarcoma of the liver and brain cancer at a polymer production plant. J Occup Environ Med 45:538, 2003

63. Molina E, Hernandez A: Clinical manifestations of primary hepatic angiosarcoma. Dig Dis Sci 48:677, 2003

64. Sinatra R, Brancaccio G, di Gioia CR, et al: Integrated approach for cardiac angiosarcoma. Int J Cardiol 88:301, 2003

65. Hogle WP: Malignant hemangiopericytoma: a clinical overview and case study. Clin J Oncol Nurs 7:57, 2003

66. Beral V, Peterman TA, Berkelman RL, et al: Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 335:123, 1990

67. Eltom MA, Jemal A, Mbulaiteye SM, et al: Trends in Kaposi's sarcoma and non-Hodgkin's lymphoma incidence in the United States from 1973 through 1998. J Natl Cancer Inst 94:1204, 2002

68. Biggar RJ: AIDS-related cancers in the era of highly active antiretroviral therapy. Oncology (Huntingt) 15:439, 2001

69. Serraino D, Toma L, Andreoni M, et al: A seroprevalence study of human herpesvirus type 8 (HHV8) in eastern and Central Africa and in the Mediterranean area. Eur J Epidemiol 17:871, 2001

70. Cannon MJ, Dollard SC, Black JB, et al: Risk factors for Kaposi's sarcoma in men seropositive for both human herpesvirus 8 and human immunodeficiency virus. AIDS 17:215, 2003

71. Martin JN, Ganem DE, Osmond DH, et al: Sexual transmission and the natural history of human herpesvirus 8 infection. N Engl J Med 338:948, 1998

72. Hirota S, Isozaki K, Moriyama Y, et al: Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. Science 279:577, 1998

73. Miettinen M, Sarlomo-Rikala M, Lasota J: Gastrointestinal stromal tumors: recent advances in understanding of their biology. Hum Pathol 30:1213, 1999

74. Carney JA: Gastric stromal sarcoma, pulmonary chondroma, and extra-adrenal paraganglioma (Carney triad): natural history, adrenocortical component, and possible familial occurrence. Mayo Clin Proc 74:543, 1999

75. Carney JA, Stratakis CA: Familial paraganglioma and gastric stromal sarcoma: a new syndrome distinct from the Carney triad. Am J Med Genet 108:132, 2002

76. Kiryu T, Kawaguchi S, Matsui E, et al: Multiple chondromatous hamartomas of the lung: a case report and review of the literature with special reference to Carney syndrome. Cancer 85:2557, 1999

77. Price B: Analysis of current trends in United States mesothelioma incidence. Am J Epidemiol 145:211, 1997

78. Connelly RR, Spirtas R, Myers MH, et al: Demographic patterns for mesothelioma in the United States. J Natl Cancer Inst 78:1053, 1987

79. Enterline PE, Henderson VL: Geographic patterns for female pleural mesothelioma deaths in the United States, 1968–1981. J Natl Cancer Inst 79:31, 1987

80. McDonald AD, McDonald JC: Epidemiology of malignant mesothelioma: asbestos-related malignancy. Asbestos-Related Malignancy. Antman K, Aisner J, Eds. Grune & Stratton, Orlando, 1987, p 31

81. McNeil BJ, Eddy DM: The costs and effects of screening for cancer among asbestos-exposed workers. J Chron Dis 35:351, 1982

82. Selikoff IJ, Churg J, Hammond EC: Asbestos exposure and neoplasia. J Am Med Assoc 188:22, 1964

83. Zagars GK, Ballo MT, Pisters PW, et al: Prognostic factors for patients with localized soft-tissue sarcoma treated with conservation surgery and radiation therapy: an analysis of 225 patients. Cancer 97:2530, 2003

84. Zagars GK, Ballo MT, Pisters PW, et al: Surgical margins and reresection in the management of patients with soft tissue sarcoma using conservative surgery and radiation therapy. Cancer 97:2544, 2003

85. Zagars GK, Ballo MT, Pisters PW, et al: Preoperative vs postoperative radiation therapy for soft tissue sarcoma: a retrospective comparative evaluation of disease outcome. Int J Radiat Oncol Biol Phys 56:482, 2003

86. Rydholm A, Gustafson P: Should tumor depth be included in prognostication of soft tissue sarcoma? BMC Cancer 3:17, 2003

87. Sakata K, Johnson FE, Beitler AL, et al: Extremity soft tissue sarcoma patient follow-up: tumor grade and size affect surveillance strategies after potentially curative surgery. Int J Oncol 22:1335, 2003

88. Koea JB, Leung D, Lewis JJ, et al: Histopathologic type: an independent prognostic factor in primary soft tissue sarcoma of the extremity? Ann Surg Oncol 10:432, 2003

89. Gerrand CH, Bell RS, Wunder JS, et al: The influence of anatomic location on outcome in patients with soft tissue sarcoma of the extremity. Cancer 97:485, 2003

90. Youssef E, Fontanesi J, Mott M, et al: Long-term outcome of combined modality therapy in retroperitoneal and deep-trunk soft-tissue sarcoma: analysis of prognostic factors. Int J Radiat Oncol Biol Phys 54:514, 2002

91. Gilbeau L, Kantor G, Stoeckle E, et al: Surgical resection and radiotherapy for primary retroperitoneal soft tissue sarcoma. Radiother Oncol 65:137, 2002

92. Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: meta-analysis of individual data. Sarcoma Meta-analysis Collaboration. Lancet 350:1647, 1997

93. Frustaci S, Gherlinzoni F, De Paoli A, et al: Adjuvant chemotherapy for adult soft tissue sarcomas of the extremities and girdles: results of the Italian randomized cooperative trial. J Clin Oncol 19:1238, 2001

94. Gortzak E, Azzarelli A, Buesa J, et al: A randomised phase II study on neo-adjuvant chemotherapy for 'high- risk' adult soft-tissue sarcoma. Eur J Cancer 37:1096, 2001

95. Billingsley KG, Burt ME, Jara E, et al: Pulmonary metastases from soft tissue sarcoma: analysis of patterns of diseases and postmetastasis survival. Ann Surg 229:602, 1999

96. Weiser MR, Downey RJ, Leung DH, et al: Repeat resection of pulmonary metastases in patients with soft-tissue sarcoma. J Am Coll Surg 191:184, 2000

97. Abdalla EK, Pisters PW: Metastasectomy for limited metastases from soft tissue sarcoma. Curr Treat Options Oncol 3:497, 2002

98. Santoro A, Tursz T, Mouridsen H, et al: Doxorubicin versus CYVADIC versus doxorubicin plus ifosfamide in first-line treatment of advanced soft tissue sarcomas: a randomized study of the European Organization for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group. J Clin Oncol 13:1537, 1995

99. Verweij J, Lee SM, Ruka W, et al: Randomized phase II study of docetaxel versus doxorubicin in first- and second-line chemotherapy for locally advanced or metastatic soft tissue sarcomas in adults: a study of the European organization for research and treatment of cancer soft tissue and bone sarcoma group. J Clin Oncol 18:2081, 2000

100. Nielsen OS, Dombernowsky P, Mouridsen H, et al: High-dose epirubicin is not an alternative to standard-dose doxorubicin in the treatment of advanced soft tissue sarcomas: a study of the EORTC soft tissue and bone sarcoma group. Br J Cancer 78:1634, 1998

101. Schoenfeld D, Rosenbaum C, Horton J, et al: A comparison of adriamycin versus vincristine and adriamycin, and cyclophosphamide for advanced sarcoma. Cancer 50:2757, 1982

102. Pinedo HM, Bramwell VHC, Mouridsen HT, et al: CYVADIC in advanced soft tissue sarcoma: a randomized study comparing two schedules. A study of the EORTC Soft Tissue and Bone Sarcoma Group. Cancer 53:1825, 1984

103. Bodey GP, Rodriquez V, Murphy WK, et al: Protected environment-prophylactic antibiotic program for malignant sarcoma: randomized trial during remission induction chemotherapy. Cancer 47:2422, 1981

104. Benjamin RS, Gottlieb JA, Baker LH: CYVADIC vs CYVADACT: a randomized trial of cyclophosphamide, vincristine and adriamycin, plus either dacarbazine or actinomycin D in metastatic sarcomas. Proc Am Assoc Cancer Res 17:256, 1976

105. Borden EC, Amato D, Rosenbaum C, et al: Randomized comparison of adriamycin regimens for treatment of metastatic soft tissue sarcomas. J Clin Oncol 5:840, 1987

106. Zalupski M, Metch B, Balcerzak S, et al: Phase III comparison of doxorubicin and dacarbazine given by bolus versus infusion in patients with soft-tissue sarcomas: a Southwest Oncology Group Study. J Natl Cancer Inst 83:926, 1991

107. Antman KH, Ryan L, Elias A, et al: Response to ifosfamide and mesna: 124 previously treated patients with metastatic or unresectable sarcoma. J Clin Oncol 7:126, 1989

108. Bramwell V, Mouridsen HT, Santoro A, et al: Cyclophosphamide vs ifosfamide: final report of a randomized phase II trial in adult soft tissue sarcoma. Eur J Cancer Clin Oncol 23:311, 1987

109. Antman K, Crowley J, Balcerzak SP, et al: An intergroup phase III randomized study of doxorubicin and dacarbazine with or without ifosfamide and mesna in advanced tissue and bone sarcomas. J Clin Oncol 11:1276, 1993

110. Edmonson JH, Ryan L, Blum RH, et al: Randomized comparison of doxorubicin alone vs ifosphamide and doxorubicin or mitomycin, doxorubicin, cisplatin against advanced soft tissue sarcomas. J Clin Oncol 11:1269, 1993

111. Antman K, Crowley J, Balcerzak SP, et al: A Southwest Oncology Group and Cancer and Leukemia Group B phase II study of doxorubicin, dacarbazine, ifosfamide, and mesna in adults with advanced osteosarcoma, Ewing's sarcoma, and rhabdomyosarcoma. Cancer 82:1288, 1998

112. Sugarbaker DJ, Heher EC, Lee TH, et al: Extrapleural pneumonectomy, chemotherapy, and radiotherapy in the treatment of diffuse malignant pleural mesothelioma. J Thorac Cardiovasc Surg 102:10, 1991

113. Sebbag G, Yan H, Shmookler BM, et al: Results of treatment of 33 patients with peritoneal mesothelioma. Br J Surg 87:1587, 2000

114. Parra HS, Tixi L, Latteri F, et al: Combined regimen of cisplatin, doxorubicin, and alpha-2b interferon in the treatment of advanced malignant pleural mesothelioma: a phase II multicenter trial of the Italian Group on Rare Tumors (GITR) and the Italian Lung Cancer Task Force (FONICAP). Cancer 92:650, 2001

115. Byrne MJ, Davidson JA, Musk AW, et al: Cisplatin and gemcitabine treatment for malignant mesothelioma: a phase II study. J Clin Oncol 17:25, 1999

116. Savage DG, Antman KH: Imatinib mesylate—a new oral targeted therapy. N Engl J Med 346:683, 2002

117. van Oosterom AT, Judson I, Verweij J, et al: Safety and efficacy of imatinib (STI571) in metastatic gastrointestinal stromal tumours: a phase I study. Lancet 358:1421, 2001

118. Demetri GD, von Mehren M, Blanke CD, et al: Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. N Engl J Med 347:472, 2002

119. Tomer G, Cohen IJ, Kidron D, et al: Prognostic factors in nonmetastatic limb osteosarcoma: a 20-year experience in one center. Int J Oncol 15:179, 1999

120. Eilber F, Giuliano A, Eckardt J, et al: Adjuvant chemotherapy for osteosarcoma: a randomized prospective trial. J Clin Oncol 5:21, 1987

121. Link MP, Goorin AM, Horowitz M, et al: Adjuvant chemotherapy of high-grade osteosarcoma of the extremity: updated results of the Multi-Institutional Osteosarcoma Study. Clin Orthop 270:8, 1991

122. Krailo M, Ertel I, Makley J, et al: A randomized study comparing high-dose methotrexate with moderate-dose methotrexate as components of adjuvant chemotherapy in childhood nonmetastatic osteosarcoma: a report from the Childrens Cancer Study Group. Med Pediatr Oncol 15:69, 1987

123. Bacci G, Gherlinzoni F, Picci P, et al: Adriamycin-methotrexate high dose versus adriamycin-methotrexate moderate dose as adjuvant chemotherapy for osteosarcoma of the extremities: a randomized study. Eur J Cancer Clin Oncol 22:1337, 1986

124. Bramwell VH, Burgers M, Sneath R, et al: A comparison of two short intensive adjuvant chemotherapy regimens in operable osteosarcoma of limbs in children and young adults: the first study of the European Osteosarcoma Intergroup. J Clin Oncol 10:1579, 1992

125. Souhami RL, Craft AW, Van der Eijken JW, et al: Randomised trial of two regimens of chemotherapy in operable osteosarcoma: a study of the European Osteosarcoma Intergroup. Lancet 350:911, 1997

126. Voute PA, Souhami RL, Nooij M, et al: A phase II study of cisplatin, ifosfamide and doxorubicin in operable primary, axial skeletal and metastatic osteosarcoma. European Osteosarcoma Intergroup (EOI). Ann Oncol 10:1211, 1999

127. Willen H: Tumor necrosis and prognosis in osteosarcoma. Acta Orthop Scand Suppl 273:126, 1997

128. Murray J, Jessup K, Romsdahl M, et al. Limb-salvage study in osteosarcoma: early experience at M.D. Anderson Hospital and Tumor Institute. Cancer Treat Symp 3:131, 1985

129. Antonescu CR, Argani P, Erlandson RA, et al: Skeletal and extraskeletal myxoid chondrosarcoma: a comparative clinicopathologic, ultrastructural, and molecular study. Cancer 83:1504, 1998

130. Ahmed AR, Tan TS, Unni KK, et al: Secondary chondrosarcoma in osteochondroma: report of 107 patients. Clin Orthop 411:193, 2003

131. Martin RC 2nd, Brennan MF: Adult soft tissue Ewing sarcoma or primitive neuroectodermal tumors: predictors of survival? Arch Surg 138:281, 2003

132. Dorfman HD, Czerniak B: Bone cancers. Cancer 75:203, 1995

XIV BLADDER, RENAL, AND TESTICULAR CANCER

Derek Raghavan, m.d., ph.d.

Cancer of the Bladder

Bladder cancer accounts for more than 90% of urinary tract malignancies. It is the sixth most common cancer in the United States, with 50,000 new cases a year, and it causes over 10,000 deaths a year. The management of this disease has become a paradigm of the approach to solid tumor malignancy,[1] because bladder cancer exhibits stem cell function (with the ability to differentiate along different histologic pathways); is associated with expression of several common oncogenes; and responds to surgery, radiation, and chemotherapy.

Epidemiology and Etiology

The incidence of bladder cancer increases with age and is fourfold higher in men than in women. The disease is seen more often in whites than in African or Asian populations. The incidence is higher in urban areas, suggesting a possible role of environmental factors in carcinogenesis. Cigarette smoking is the primary risk factor, with a confirmed dose- and time-response relationship for both sexes.[2] Occupational exposure to carcinogenic compounds found in dye, rubber, paint, plastics, metal, and motor vehicle exhaust significantly raises the risk of bladder cancer. Other established risk factors include chronic infection of the lower urinary tract, history of external-beam radiation to the pelvis, long-term indwelling urinary catheter, the use of cyclophosphamide, high-fat diet, and low daily fluid intake. In the Mediterranean basin, schistosomiasis is still the main causative agent and is classically associated with squamous cell carcinoma.

More than 90% of bladder cancers are transitional cell carcinomas. The less common cell types are squamous cell carcinoma, adenocarcinoma, and small cell carcinoma; rare histologic types such as sarcoma, lymphoma, and melanoma comprise less than 1% of cases.[3] Although the histogenesis is not fully understood, bladder cancer appears to originate from a stem cell tumor[4] and, therefore, has the capacity to evolve different histologic patterns that are intermixed; for example, dominant transitional cell carcinoma may coexist with squamous and squamoid differentiation. Bladder cancer is often associated with a field defect of the urinary mucosa, implying an antecedent level of precancerous change that can affect multiple sites, with the result that the entire lining of the urinary tract may be at risk for developing malignancy.

Pathobiology

Pathologically, bladder cancer is categorized as superficial or invasive [*see Figure 1*]. Superficial bladder cancer, which constitutes about 80% of incident cases, is restricted to the level of the bladder mucosa and lamina propria. The seminal event in the development of superficial bladder cancer appears to be a loss of heterozygosity of chromosome 9 (i.e., that there is a potential oncogene located at that site).[5] Superficial bladder cancer is usually associated with a high level of histologic differentiation and with long patient survival.

Invasive bladder tumors are those that penetrate beyond the lamina propria. They are aggressive and tend to metastasize early. About 20% of patients have invasive cancer at presentation, although some superficial tumors recur as invasive disease. The

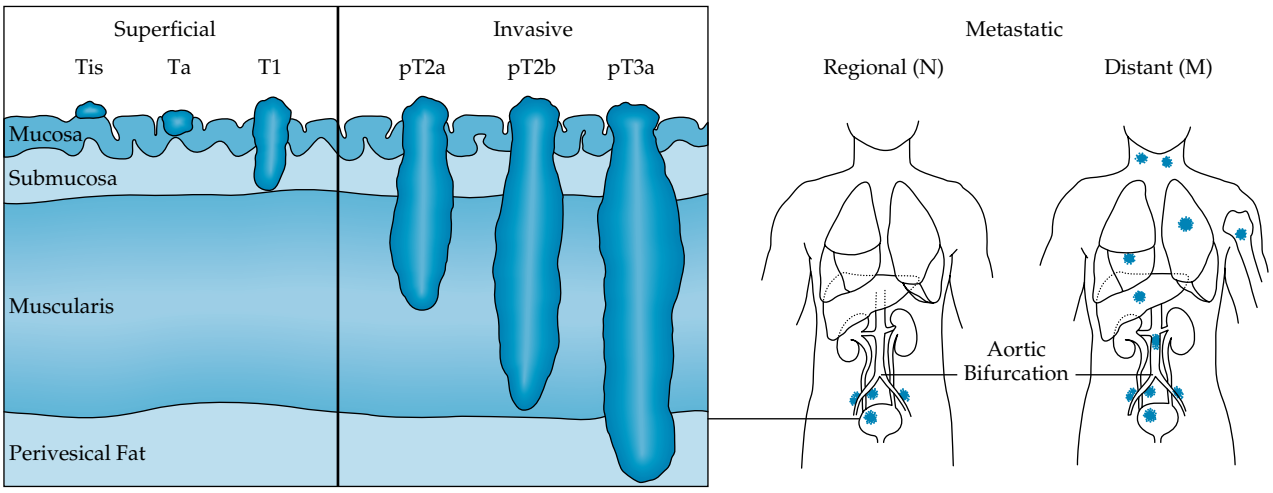


Figure 1 **Carcinoma of the bladder can be staged according to the depth of penetration of the tumor into the bladder wall and, for advanced disease, by the location of metastases. Superficial disease comprises noninvasive papillary carcinoma (Ta); carcinoma in situ, or so-called flat tumor (Tis); and tumors that extend only into the submucosa (stage T1). Invasive disease is characterized by extension past the submucosa. Invasion into the muscle layer is designated stage T2: pT2a denotes tumor that invades superficial muscle (the inner half); and pT2b, tumor that invades deep muscle (the outer half). Stage T3 tumors are those that invade perivesical fat: pT3a tumors are microscopic, and pT3b tumors are macroscopic (i.e., there is an extravesical mass). Stage T4 tumor invades adjacent organs: T4a invades the prostate, uterus, or vagina; and T4b invades the pelvic or abdominal wall. Metastatic disease represents spread to regional lymph nodes (i.e., those below the aortic bifurcation [stages N1 through N3]); involvement of lymph nodes above the aortic bifurcation, or involvement of any bone or soft tissue (stage M1).**

most important prognostic factor is the depth of tumor invasion (stage). T2 tumors, which extend into the muscularis propria, carry a better prognosis than T3 and T4 tumors, which extend beyond the muscularis propria. The level of differentiation (grade) is also an important factor. Unfortunately, most invasive tumors are high grade. Vascular or lymphatic invasion predicts an increased risk of soft tissue invasion and of distant metastasis. Other adverse prognostic factors include the absence of expression of blood group antigens on the tumor cell surface, DNA ploidy, expression of epidermal growth factor receptor (EGFR), and presence of mutations of the *p53* suppressor gene. Poorer prognosis has also been correlated with expression of other genes, including *p16*, *p21*, and *Rb;* these genes may work in concert with *p53* to regulate tumor growth.[6]

DIAGNOSIS

Clinical Findings

Most bladder cancers present as gross or microscopic hematuria that is painless and occurs suddenly or intermittently.[7] Urinary frequency, nocturia, and urgency can be the result of bladder wall irritation or volume loss from space-occupying lesions and are often misdiagnosed as chronic interstitial cystitis. Abdominal discomfort, flank pain, pelvic pain, altered bowel habit, or a palpable mass can occasionally be the first clinical evidence of invasive bladder cancer. Obstruction of the ureteral orifice with subsequent hydronephrosis and renal insufficiency may occur with an invasive tumor, although these tumors can remain clinically silent until locally advanced. Rarely, pneumaturia will suggest the presence of a fistula to the bowel.

Invasive bladder cancer may extend locally into the prostate, seminal vesicles, rectum, uterus or vagina, sacral vertebra, and the retroperitoneal soft tissue. It spreads via lymphatics and blood vessels to distant lymph nodes, the lungs, the liver, skin, peritoneum, and bones[1] and may even cause brain metastases or carcinomatous meningitis.[8] Depending on the practice setting, 5% to 20% of all patients with bladder cancer present with symptoms from metastatic lesions. Constitutional symptoms may occur with disseminated disease, but clinical paraneoplastic syndromes are relatively rare in transitional cell cancers.

Laboratory Findings

Routine urinalysis in patients with bladder cancer will usually show hematuria, the degree of which does not correlate with the extent of the lesion. Intravenous urography may reveal an intravesical filling defect and also provides anatomic information about the urinary tract, such as the presence of hydronephrosis or hydroureter. Ultrasound is sometimes used to assess the bladder wall and to evaluate the kidneys and ureters. Urine cytology is a convenient and inexpensive method to obtain a tissue diagnosis and has a specificity and sensitivity of 80% in grade III tumors, but it has a relatively low sensitivity for grade I or II tumors (about 10% and 50%, respectively). It has been suggested that flow cytometric studies of urine specimens may reveal occult tumors via the demonstration of aneuploidy. Direct visualization and biopsy of the tumor is usually achieved by cystoscopy, which has been facilitated in recent years by the introduction of flexible instrumentation.

Staging

The staging process for bladder cancer requires a thorough history and physical examination, along with laboratory studies that include a complete blood count and tests of hepatic and renal function. Renal function may be compromised by factors that caused the cancer, such as analgesic nephropathy, or by obstruction of renal outflow by the tumor mass. Occasionally, a patient will have a markedly elevated leukocyte count secondary to the production of granulocyte colony-stimulating factor (G-CSF) by the bladder cancer. A chest x-ray should be obtained in all patients. In some centers, CT scan of the chest is routinely performed instead of initial chest x-ray as part of routine staging. A CT scan of the abdomen and pelvis allows the assessment of lymphadenopathy and other organ involvement, although the yield of pelvic CT is relatively low. Radionuclide bone scans may be used for staging, but the yield in asymptomatic patients with normal serum alkaline phosphatase levels is low. MRI scanning may be helpful in defining the extent of abdominal or pelvic disease, especially in a patient with renal dysfunction who is not suitable for contrast-enhanced CT scanning. The role of positron emission tomography (PET) scanning has not yet been clearly defined in bladder cancer, although it is clear that false negative results limit utility.

It seems likely that in the future, more sophisticated molecular studies will become routine in the diagnosis and management of bladder cancer. For example, mutation of the *p53* gene appears to be an important predictor of outcome in invasive bladder cancer.[9] It is also clear that expression of the *Rb* gene, *p16*, *p21*, the epidermal growth factor receptor, thrombospondin-1, multidrug-resistance phenotype, and glutathione may be predictive of natural history or of response to therapy,[6,9-12] and it seems likely that these assays will become a part of a routine staging protocol.

MANAGEMENT

Superficial Bladder Cancer

For superficial papillary bladder cancer, the initial treatment is careful and thorough endoscopic resection of the tumor or tumors.[13] In patients at high risk for recurrence, bacillus Calmette Guérin (BCG) may be infused into the bladder through a urinary catheter as an adjuvant to transurethral resection. A common schedule is weekly administration for 6 weeks followed by monthly doses for 6 to 12 months, although the optimal approach has not been defined. The mechanism of action is incompletely understood, but it appears to be based on local immunologic response. Although there is great variability of published data, intravesical BCG appears to prevent recurrence in up to 30% of cases and delays recurrence in around 50%, compared with patients who do not receive adjuvant therapy.

Side effects of intravesical BCG include dysuria, frequency, hematuria, and a flulike syndrome. More significantly, because BCG is an attenuated mycobacterium, it can produce local, regional, and systemic infections. Granulomatous infections can occur at extravesical sites, including the prostate, epididymis, testes, kidney, liver, and lungs. BCG sepsis is the most serious complication and can be life-threatening. Systemic involvement (so-called BCGosis) is treated with triple-antibiotic antituberculous therapy for 6 months.

Cytotoxic agents can also be used for intravesical treatment, although the superiority of BCG to such agents has been demonstrated in randomized clinical trials.[13] Agents that may have a role for patients who refuse cystectomy after BCG failure include doxorubicin, thiotepa, and mitomycin-C. Because of their high molecular weight, doxorubicin and mitomycin-C have a lesser tendency to be absorbed systemically from the bladder than

Table 1 Studies of Chemotherapy plus Radiation in Invasive Bladder Cancer

Series	*Number of Patients*	*Clinical T Stage*	*Chemotherapeutic Regimen*	*Bladder Preservation (%)*	*Overall Survival*	*Median Follow-up (months)*
Coppin[18]	42	T2-T4	Cisplatin	Not reported	61% at 2yr	50
Dunst[132]	139	T1-T3	Cisplatin or carboplatin	79	40% at 7 yr	Minimum 12
Housset[133]	54	T2-T4	5-FU, cisplatin	78	59% at 3 yr	18–58
Kaufmann[134]	53	T2-T4	Cisplatin, methotrexate, vinblastine	85	53% at 4 yr	48
Sauer[135]	67	T1-T4	Carboplatin	84	66% at 3 yr	Not available
Tester[136]	46	T2-T4	Cisplatin	83	66% at 3 yr	36

does thiotepa, which is a small molecule. Clinical trials are in progress to assess the utility of some of the newer cytotoxic agents, such as the taxanes and gemcitabine, for intravesical chemotherapy. These agents are relatively large molecules, so significant systemic absorption seems unlikely.

Invasive Bladder Cancer

In North America and Europe, cystectomy is the standard treatment for organ-confined invasive bladder cancer.[14] Radical cystectomy involves the en bloc removal of the anterior pelvic organs, which include the bladder, prostate, and seminal vesicles in men and the bladder, urethra, uterus, ovaries, and vaginal cuff in women.[14,15] Bilateral pelvic lymph node dissection is often performed. The ureters are reconnected to an intestinal conduit as a urinary diversion. Traditionally, the conduit has been fashioned to drain into an external collecting bag attached to the abdominal wall. However, continent reservoirs such as the Koch pouch (using the ileum) and the Indiana and T pouches (using other segments of bowel) have become popular, because they afford greater degrees of continence and a return to more normal lifestyles. These procedures involve the creation of an internal conduit with an antireflux mechanism that is either brought to the abdominal wall or sutured to the urethra, thus allowing patients to self-catheterize or void in the normal position, which greatly improves their self-image and increases their acceptance of the procedure.

When performed in routine clinical practice, radical cystectomy results in 60% to 75% 5-year survival rates in patients with T2 disease and 20% to 40% survival in those with T3 or T4 disease. Centers of excellence have reported superior long-term results, with 10-year survival as high as 85% for patients with organ-confined disease.[15]

For patients with localized invasive disease who are not surgical candidates, radiation is the alternative definitive therapy in North America, although it should be noted that radical radiotherapy is the standard definitive treatment in some British, Canadian, and European centers.[16,17] To date, there have been no well-designed, randomized studies comparing radiation with surgery in patients with similar characteristics.

A relatively standard radiotherapy regimen is more than 65 to 70 Gy given over 6 to 7 weeks, with the major component of dosage focused on the tumor and surrounding areas, as defined by CT scanning in the prone position. Toxicities of radiation include dermatitis, proctitis that is occasionally complicated by bleeding and obstruction, cystitis or bladder fibrosis, impotence, incontinence, and development of secondary malignancies in the region surrounding the radiation field.

Combined-modality approaches, incorporating systemic chemotherapy with definitive local modalities, have been studied extensively in the past few years in the hope of sparing the bladder or improving overall survival.[1] The rationale is that systemic chemotherapy may reduce the extent of local tumor while controlling occult metastases. Trials of concurrent chemotherapy and radiation have suggested that a high rate of bladder preservation is possible with this approach, compared with radiation alone [*see Table 1*].[18] However, the one reported randomized trial did not show a statistically significant survival benefit, although it did not have sufficient statistical power to assess survival. When chemoradiation fails with local progression or relapse, salvage surgery becomes extremely difficult because of the formation of dense adhesions.

It is clear that neoadjuvant chemotherapy can often reduce the local tumor volume within the bladder, and complete remission is sometimes achieved. However, most randomized trials of neoadjuvant systemic chemotherapy followed by definitive radiotherapy or cystectomy have shown no significant survival benefit from the combined modality approaches [*see Table 2*].[19-21] It should be noted that combination chemotherapy has been more effective, and the North American Intergroup study of the combination of methotrexate-vinblastine-doxorubicin (Adriamycin)-cisplatin (MVAC) plus cystectomy versus cystectomy alone showed a statistically significant survival benefit.

Chemotherapy administered after radical cystectomy for patients with lymph node involvement may improve disease-free survival [*see Table 2*].[22,23] However, in the extant trials, which have been flawed by poor design and inadequate sample size, overall improvement in survival has not been demonstrated with statistical significance. These flaws have been addressed in a large international randomized trial that is currently under way.

Metastatic Bladder Cancer

Chemotherapy is the treatment of choice for patients with metastatic bladder cancer. The first major step in the modern era of chemotherapy was the development of MVAC by Sternberg and colleagues.[24] Several series demonstrated median survival of about 12 months in patients with advanced and metastatic bladder cancer treated with this regimen. At the time of the completion of that trial, median survival without chemotherapy was only 4 to 6 months.[1] A landmark study conducted by an international consortium compared the MVAC regimen to cisplatin alone.[25] The MVAC regimen produced a response rate of 39% with median survival of 12.5 months, which was statistically superior to the response rate of 12% and median survival of 8 months in the group that received cisplatin alone. The survival benefit persisted after a minimum follow-up of 6 years, although

Table 2 Results of Clinical Trials of Neoadjuvant Chemotherapy for Invasive Bladder Cancer, Stages T1–T4

Series	*Neoadjuvant Regimen*	*Definitive Therapy*	*Median Survival with/without Neoadjuvant Therapy (months)*	*Actuarial Long-term Survival with/without Neoadjuvant Therapy*
Shipley[19]	CMV	RT/C	36/36	48%/49% at 5 yr
MRC-EORTC[21]	CMV	RT/cystectomy	44/37.5	55%/50% at 3 yr
Intergroup[20]	MVDC	Cystectomy	72/45	55%/45% at 6 yr
Nordic 1 trial[137]	DC	Cystectomy	Not reached/72	59%/51% at 5 yr

C—cisplatin D—doxorubicin M—methotrexate MRC-EORTC—Medical Research Council/European Organization for Research and Treatment of Cancer RT—radiotherapy V—vinblastine

the vast majority of patients in both randomization arms had died by that time.[26] The clear limitations of the MVAC regimen that were evident from this analysis opened the way for investigation of novel agents.[27]

Novel cytotoxic agents are being actively studied; and in recent years, single-agent therapy with paclitaxel, gemcitabine, docetaxel, or ifosfamide has been shown to produce response rates between 20% and 30% [*see Figure 2*]. The combination of these agents with other standard or investigational drugs has resulted in response rates of 50% to 80%, with apparently less toxicity than the conventional-combination MVAC regimen.[28-30]

A phase III trial comparing the regimen of gemcitabine and cisplatin with the MVAC regimen has been completed.[31] Although it was not designed to be an equivalence trial, this study suggests that there is no great difference in survival between the gemcitabine-cisplatin regimen and the MVAC regimen but does suggest that the gemcitabine-cisplatin combination is substantially less toxic. As a consequence, an international consortium (the International Intergroup) is now comparing gemcitabine-cisplatin with gemcitabine-cisplatin-paclitaxel for patients with previously untreated metastatic transitional cell carcinomas. These investigators have indicated that it is probably time to leave the MVAC regimen behind as the standard of care.

Of particular importance, there is emerging evidence that stage migration has occurred in the management of advanced bladder cancer (presumably because of increased use of chemotherapy for patients with asymptomatic or small volume metastatic disease detected by postoperative screening techniques, such as CT, MRI, and PET scans). At Memorial Sloan-Kettering Cancer Center, the initial experience with the MVAC regimen in the 1980s produced a median survival of about 12 months,[24] whereas recent data from that institution showed a median survival of 18 months with a variant of MVAC that has a relatively minor dose escalation. This should be borne in mind when considering the utility of novel combinations, such as the ITP regimen (ifosfamide, paclitaxel, and cisplatin), which also yields a median survival of about 18 months. Before novel regimens are accepted into routine clinical practice, their safety and efficacy should be defined in randomized trials against accepted current standards.

In addition to chemotherapy, novel agents that target cell regulatory proteins may have application to the management of bladder cancer. Clinical trials are currently assessing the efficacy of agents that modulate the function of EGFR and other tyrosine kinase inhibitors. These agents are being tested both as monotherapy and in combination with chemotherapy. The ability to

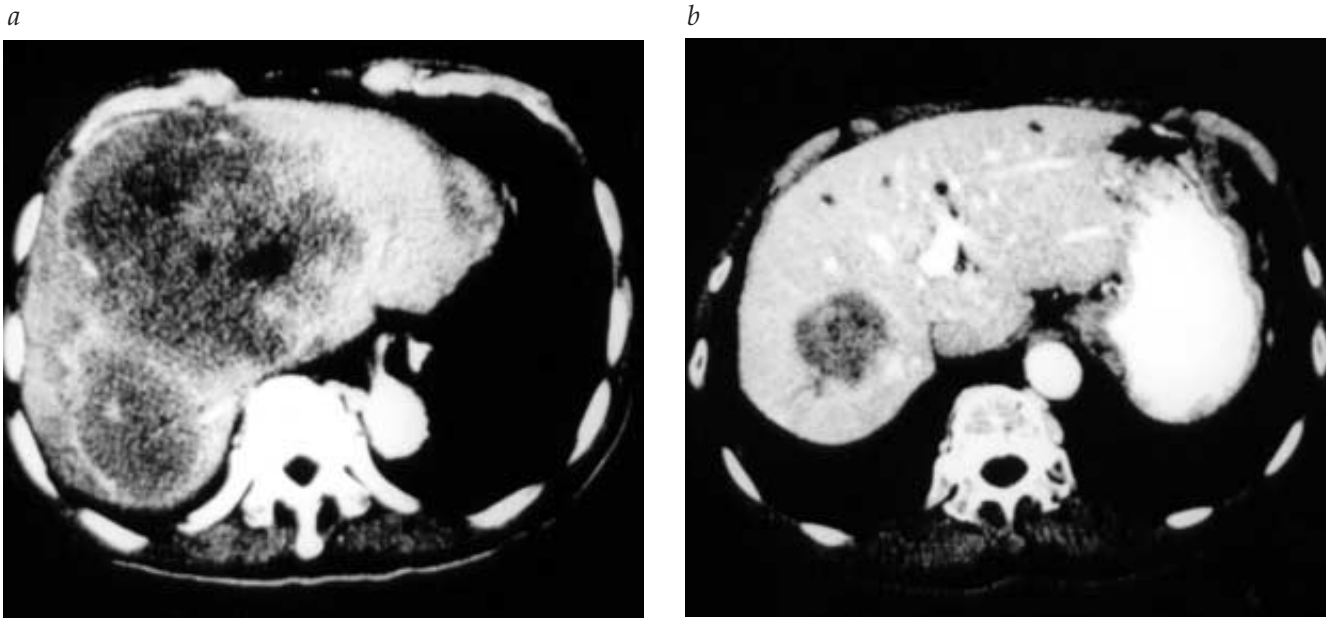


Figure 2 **CT scans of a patient with hepatic metastases from bladder cancer before (*a*) and after (*b*) treatment with gemcitabine-cisplatin.**

identify expression of the *HER-2/neu* oncogene, EGFR, and other molecular predictors of response to treatment may allow tailoring of more specific therapeutic strategies.

Cancer of the Kidney

Renal neoplasms are uncommon, accounting for 3% of all adult malignancies in Western society. The majority of these cancers are renal cell carcinomas (RCC), although other rare tumors of the kidney have been described, including transitional cell carcinomas, sarcomas, adult Wilms tumors, peripheral neuroectodermal tumors, lymphomas, and germ cell tumors.[32,33] It should not be forgotten that tumors can metastasize to the kidney, commonly from lung, breast, and gastrointestinal tract malignancies. Transitional carcinomas occur predominantly in the renal pelvis and are managed in similar fashion to transitional cell carcinomas that arise in the bladder and other sites.

EPIDEMIOLOGY AND ETIOLOGY

RCC affects 30,600 people in the United States annually, and the incidence is steadily rising, with more than 10,000 related deaths occurring each year.[34] This disease usually presents during the fifth to seventh decades of life, with a median age at diagnosis of 60 years. The incidence in men is twice that in women. RCC occurs in all ethnic groups, with no racial predilection. Most RCCs occur sporadically, but about 4% of cases present in an inherited pattern. Such familial renal cancers include von Hippel–Lindau (VHL) disease[35] and familial papillary renal cell cancers. Approximately 1.6% of RCCs are part of the autosomal dominant VHL disease, which is also characterized by retinal and central nervous system hemangioblastoma, pheochromocytoma, and pancreatic cyst. Compared with sporadic cases, RCCs in the VHL syndrome tend to be multifocal and bilateral and appear at a younger age. RCC and malignant angiomyolipoma are also associated with tuberous sclerosis complex, an autosomal dominant disorder of unknown etiology characterized by seizures, mental retardation, and hamartomas.

The true etiology of RCC remains largely undefined.[34] Documented associations include smoking, diet, obesity, and hypertension. The relationship between RCC and the use of antihypertensive agents remains controversial. Patients with autosomal dominant or recessive polycystic kidney disease are not at higher risk than the general population. However a threefold to sixfold higher incidence of RCC has been found in patients undergoing long-term dialysis and in renal transplant recipients, presumably because of the development of acquired cystic kidney disease. Occupational exposure to chemicals appears to have little consistent significance, although a range of different occupations has been implicated in isolated case-control studies.

The isolation of the *VHL* gene in 1993 on chromosome 3p25 was a critical step in the understanding of the molecular genetics of renal cell carcinoma.[35] Subsequently, the gene was found to encode for a tumor suppressor protein that prevents the formation of a transcriptional elongation complex and inhibits the transcription of certain target genes in cell proliferation. At the molecular level, both copies of the *VHL* gene must be inactivated for tumors to develop; therefore, the first mutation in the *VHL* gene is inherited in the germline, and inactivation of the remaining wild-type allele occurs as a somatic event. Inactivation of the *VHL* gene could also be responsible for the nonhereditary forms of RCC, given that somatic mutation of the *VHL* gene has been found in 75% to 80% of sporadic RCC cases. To date, the *VHL* gene does not appear to be involved in the development of papillary renal carcinomas.

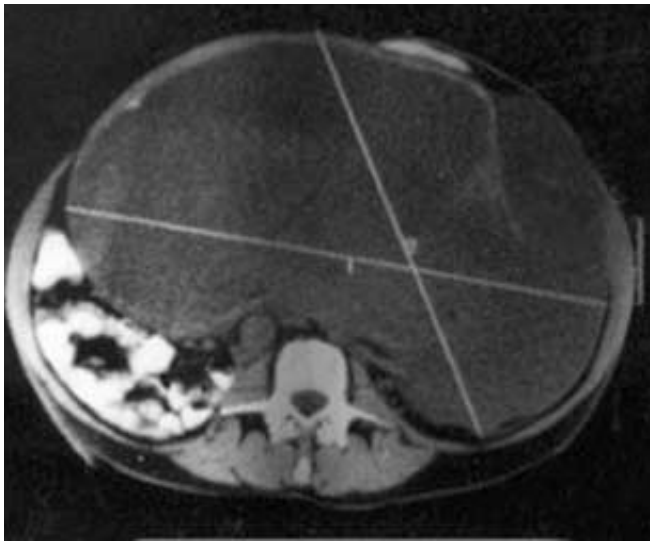


Figure 3 **Renal cell carcinoma. Figure shows a huge primary renal cell carcinoma that occupied a substantial part of the abdominal cavity but that was surgically resected without tumor spillage.**

PATHOBIOLOGY

Recent immunohistochemical studies suggest that RCCs may arise from the proximal convoluted tubular cells. Grossly, the tumors are often well demarcated, round masses protruding from the cortex. RCCs of the clear cell type are usually yellow because of high lipid content. On cross section, areas of necrosis, cystic degeneration, hemorrhage, and calcification are commonly present.[36] Histologically, RCCs are classified into five types: clear cell (75% to 85% of cases), chromophilic (12% to 14%), chromophobic (5%), oncocytic (2% to 4%), and collecting duct (1%).

DIAGNOSIS

Clinical and Laboratory Findings

With the increasing application of ultrasound and CT scans to the diagnosis of nonspecific abdominal symptoms and other conditions, 20% to 30% of RCCs are discovered incidentally as small tumors in a clinically asymptomatic stage. The classic triad of hematuria, flank pain, and a palpable mass has become a less common presentation and usually portends an unfavorable outcome.[37] However, presentation of one or two of these three symptoms is common: about 70% of patients have gross or microscopic hematuria, and 50% have abdominal or flank pain. By contrast, only 40% have a palpable mass. Constitutional symptoms (e.g., fever, night sweats, anorexia, and weight loss) are common. Exfoliative cytology of the urine may indicate the presence of carcinoma cells. Initial blood tests may reveal elevation of serum creatinine or uric acid levels; in more advanced disease, increased serum alkaline phosphatase concentrations or manifestations of paraneoplastic syndromes (see below) may be found.

Because the renal bed is a clinically silent area, some tumors are huge at first presentation [*see Figure 3*]. For the same reason, only 40% of patients have disease confined to the kidney at diagnosis, and nearly one quarter of patients present with symptoms of metastatic lesions. As the tumor grows, it may extend into and

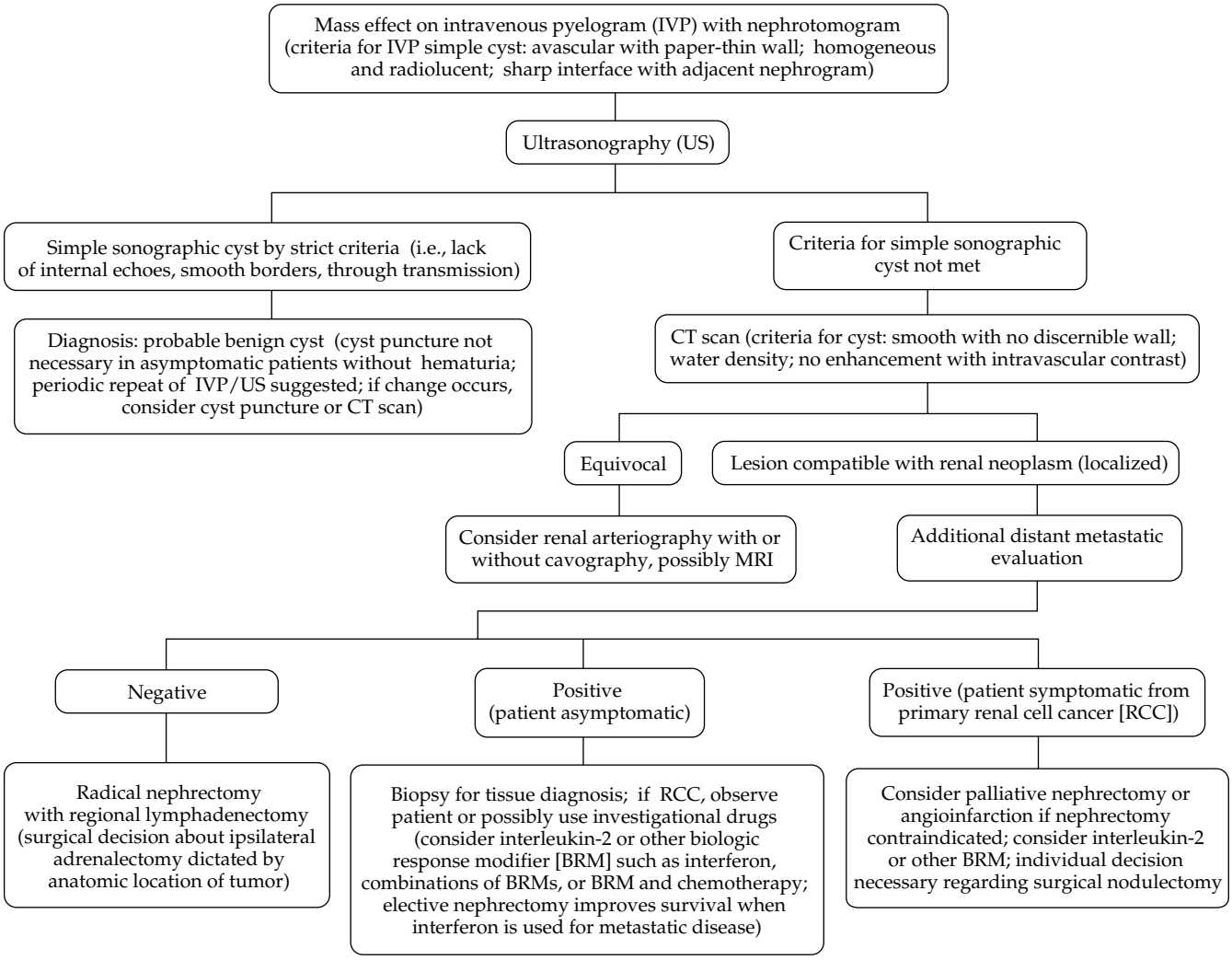


Figure 4 **Algorithm for the diagnostic evaluation of a renal mass.**

invade surrounding structures, such as the perinephric fat, posterior abdominal wall, inferior vena cava, ureter, adrenal gland, spleen, liver, or pancreas.[38] Hypertension may develop if the tumor compresses the renal parenchymal vasculature, causing increased renin production. Metastatic spread occurs via either the blood or the lymphatics. Common sites of distant metastasis include the lungs, lymph nodes, liver, bone, and brain, although metastatic lesions have been described in virtually every organ of the body.

Paraneoplastic syndromes are relatively common features in RCC, the most notable ones being hypercalcemia and erythrocytosis—from the ectopic secretion of parathyroid hormone (PTH)–like immunoreactive protein and erythropoietin, respectively. Anemia is often present. Hyponatremia from inappropriate secretion of antidiuretic hormone (ADH) and Cushing syndrome from ectopic production of adrenocorticotropic hormone (ACTH) also occur occasionally. Hypercoagulable states presenting as venous thrombosis or pulmonary embolism have been described. Other rare paraneoplastic syndromes that have been reported include amyloidosis, nonmetastatic hepatic dysfunction, limbic encephalitis, erythema gyratum repens, myopathy, and polymyalgia rheumatica.

Imaging Studies

In a patient with symptoms or signs that suggest a renal malignancy, the initial demonstration of a mass lesion is usually achieved with an intravenous pyelogram (IVP) or ultrasonography [*see Figure 4*]. An IVP may differentiate between renal cysts and malignant lesions and may reveal complications such as ureteral obstruction and hydronephrosis. Ultrasonography is a convenient method of evaluating a suspicious renal mass and is especially useful in differentiating between solid lesions and cystic lesions. When combined with an IVP, ultrasonography differentiates benign cysts from solid masses with an accuracy of more than 90%. It can also be used to evaluate the size of the contralateral kidney and to rule out obstruction and hydronephrosis, but it is not reliable for assessment of the renal vein, intrahepatic inferior vena cava, and retroperitoneal lymph nodes.

CT scanning is the technique of choice for diagnosis and staging. Performed with intravenous contrast, CT accurately defines the renal topographic anatomy, the size and location of the tumor, and the relationship of tumor to the surrounding vessels (renal vein and inferior vena cava). It also evaluates possible extracapsular spread, involvement of blood vessels or regional lymph nodes, and invasion of adjacent organs such as the

adrenals, spleen, pancreas, and liver. However, CT is less sensitive in detecting lesions smaller than 2 cm or tumor extension to the perinephric fat. Furthermore, its utility is reduced in patients in renal failure because intravenous contrast cannot be used safely in such cases. MRI may be superior to CT in detecting tumors that are less than 2 cm in diameter, in detecting extracapsular extension of tumor, and in delineating local anatomy. However, because of its expense and limited availability, MRI is often reserved for patients with known contraindications to iodine contrast or for when CT results are equivocal. The role of PET scanning has not been defined in this disease. Because 70% to 80% of RCCs are hypervascular, angiography is sometimes used to depict the lesion and provide a vascular map for the surgeon. Angiography is expensive and invasive and, therefore, should be used only for tumors in a solitary kidney, for vascular mapping before surgery, or for tumors in which angioinfarction is planned.

Staging

Staging of RCC is extremely important but sometimes difficult. The staging process should include a detailed history and physical examination, basic hematologic and biochemical studies, and a chest roentgenogram. A chest CT scan is sometimes necessary to rule out occult pulmonary and mediastinal metastasis. A bone scan should be obtained to rule out skeletal metastasis, although the yield is particularly low in asymptomatic patients with normal blood levels of alkaline phosphatase. Imaging of the brain with CT or MRI should be done in patients who have neurologic symptoms or signs. Fine-needle aspiration (FNA) of the tumor is usually not necessary, because if the FNA result is normal, the lesion will still need to be excised, given the possibility of a false negative result. On the other hand, if there is apparent extrarenal involvement of tumor, a CT-guided or ultrasound-guided biopsy of the primary lesion may be indicated if definitive surgical resection is not planned.

Tumor staging follows the standard TNM system and is predicated on the extent of the primary tumor (T), involvement of lymph nodes (N), and presence or absence of metastases (M).[39] Traditionally, T1 tumors have been defined as those limited to the kidney, with a diameter of 7 cm or less; T2 tumors, those limited to the kidney but greater than 7 cm in diameter; T3 tumors, those extending into major veins or invading adrenal or perinephric tissues but not beyond the Gerota fascia; and T4 tumors, those extending beyond the Gerota fascia. A recent modification of the staging classification divides T1 tumors into T1a (tumors 4 cm or less in diameter) and T1b (tumors greater than 4 cm in diameter), on the basis of reports showing lower 5- to 10-year survival for patients with larger T1 tumors.[39]

The pathologic stage of RCC is the most important prognostic factor. Stage I or II (tumor confined within the kidney capsule) is associated with a 5-year survival of 50% to 85%; stage III (tumor invading the renal vein, inferior vena cava, or regional lymph nodes), with 15% to 35% survival; and stage IV (T4 or N2 or metastatic disease), with 0% to 15% survival. It is still generally held that asymptomatic patients have a better prognosis than those with significant tumor-related symptoms, and normal performance status (i.e., 0) is being incorporated into many current prognostic algorithms as a favorable determinant. Important adverse prognostic factors include microscopic vascular invasion, high nuclear grade, and aneuploid DNA content. Mathematical models are being developed to incorporate these prognostic determinants into a single, predictive algorithm for individual patients,[40] but these algorithms have not yet been incorporated into standard practice internationally and still require extramural validation.

Non–clear cell histology has been investigated as a prognostic determinant but has proved difficult to assess because of the relatively small numbers of cases and the heterogeneity of histologic types (including chromophobe, collecting duct, and papillary carcinomas). However, a recent study suggests that this histologic grouping is associated with resistance to systemic therapy and poor survival, compared with the more common clear cell variants.[41] Similarly, Zisman and colleagues[42] have assessed the prognostic implications of the so-called unclassified RCC variant, comparing 29 cases of unclassified histology against a series of 264 clear cell cancers. This study also showed a more aggressive biologic behavior and worse clinical outcome, but it did suggest that immunotherapy in addition to nephrectomy may lead to an improved outcome, compared with nephrectomy alone or no active treatment. By contrast, another variant of renal malignancy, the so-called oncocytoma, characterized by dense infiltration of mitochondria within the cytoplasm, rarely metastasizes, and outcome is excellent with surgical management alone.

MANAGEMENT

Stages I, II, and III Renal Cell Carcinoma

The standard treatment for RCC in stages I through III is surgical resection. The conventional procedure is radical nephrectomy, which involves the en bloc removal of Gerota fascia with its contents and, usually, ipsilateral adrenalectomy.[43] Regional lymph nodes are resected during surgery. However, in patients with bilateral tumors, tumors in a solitary kidney, or a poorly functioning contralateral kidney, nephron-sparing surgery (partial nephrectomy or enucleation) may be considered. Recent studies have shown that for tumors 4 cm or less in size, nephron-sparing surgery provides the same rate of survival as radical nephrectomy.[44-46] However, nephron-sparing surgery does carry some risk of postoperative complications, such as hemorrhage and fistula formation, so the selection of candidates is important. In some centers, laparoscopic nephrectomy is being performed routinely for lesions of 4 to 5 cm or less, irrespective of whether nephron preservation is planned. Proponents suggest that this approach produces less morbidity because the scar is smaller and there is less disruption of the abdominal contents. Caution must also be exercised in case selection, to reduce the risk of inadequate or incomplete tumor clearance. Preoperative or postoperative radiation therapy does not reduce local recurrence or increase survival.

Stage IV Renal Cell Carcinoma

The outcome for patients with metastatic stage IV RCC is generally very poor, because RCCs are usually resistant to chemotherapy.[38] Most older clinical studies have shown response rates below 6%. Continuous infusion of floxuridine has reportedly produced response rates between 14% and 27%, although the responses are usually partial and short-lived. Investigators at the University of Chicago have published provocative data suggesting that the combination of weekly gemcitabine plus continuous-infusion fluorouracil (5-FU) may yield sustained remissions in up to 10% of cases,[47] but these data remain to be confirmed. In many cases of reported remission, the existence of true metastatic disease had not been confirmed histologically before the use of chemotherapy. This is important, because

granulomatous reactions may be found in the lungs and lymph nodes of patients with purportedly metastatic RCCs.

Surgical intervention is recommended in patients with solitary metastatic lesions that are resectable. Radiation therapy can be used to palliate symptoms from metastases to the bone or brain or from those causing spinal cord compression.

Spontaneous regression of metastatic RCC lesions has been reported in the literature, with an incidence rate of 0.8% to 3%. This has been reported for metastatic sites, including the brain, bone, regional and distant lymph nodes, liver, and, most commonly, the lungs. Complete and durable disappearance of tumors has been documented. The mechanism for this phenomenon is unknown, but it has been attributed to immunologic factors associated with reduction of tumor bulk. In many instances, spontaneous regression has occurred in supposed metastases that have not been confirmed by biopsy, and it is possible that these may have been reactive granulomata. Many of the so-called spontaneous regressions have followed surgical removal of the primary tumor, but in view of the relative rarity of this phenomenon, nephrectomy should not be routinely recommended in patients with stage IV RCC for the sole purpose of achieving spontaneous regression. Rather, nephrectomy should be considered for palliation of local symptoms such as pain and hematuria.

In a recent multicenter study of patients with advanced RCC, nephrectomy plus systemic immunotherapy with interferon was found to provide a statistically significant survival benefit over immunotherapy alone (12 months versus 8.5 months).[48] Although this difference is not very large, the order of magnitude is similar to the difference between MVAC and single-agent cisplatin in the Intergroup Advanced Bladder Cancer Study,[25] a trial that led to the establishment of MVAC as a standard of care. In addition, the tails of the survival curves were statistically different. However, critics have suggested that the results of the nephrectomy-plus-immunotherapy arm are equivalent to those reported in trials of single-modality therapy in Europe and that caution should be used in interpreting these data at this early stage. Such criticism is reasonable, but the fact is that this was a well-constructed, randomized trial with a comparison reflecting randomized cohorts. As such, this trial was less likely to be subject to the various selection biases of historical or nonrandomized comparisons. Consequently, it seems reasonable to view the combination of nephrectomy and immunotherapy as a new standard of care. However, it is not yet clear whether this applies to combinations that involve interleukin-2 (IL-2) and other, more novel cytokines.

IL-2 works through the activation of cytotoxic T cell subgroups and stimulation of cytokine release. A multicenter trial of IL-2 in patients with metastatic RCC showed therapeutic benefit (complete response rate, 4%; partial response rate, 8%; median duration of response, 23 months) but also displayed substantial toxicity, mainly from a capillary leak syndrome, which can cause cardiovascular ischemia, renal failure, and shock. Supportive measures such as use of vasopressors and intensive care unit monitoring are often needed.[49] Interferon alfa has been reported to be effective, with an overall response rate of 12%, although complete responses are less common.[50] A recent randomized study comparing IL-2 alone, interferon alfa alone, and the two agents together found a significantly higher response rate in the combination group (18.6%, versus 6.5% for IL-2 and 7.5% for interferon alfa).[50] Event-free survival at 1 year was also higher in the combination group, although there was no difference in overall survival.[50] However, other studies using similar combinations have failed to show any benefit over IL-2 used alone.[51] In general, it is believed that improved outcomes appear to be associated with the use of cytokines,[52] although most of the available data appear to be subject to the limitations of major case selection bias.

Immunologic factors seem to influence long-term survival in patients with advanced RCC. On the basis of studies showing that allogeneic stem-cell transplantation can induce curative graft-versus-leukemia reactions in some hematologic malignancies, transplantation of nonmyeloablative allogeneic peripheral blood stem cells has been used in advanced RCC, with some success.[53] Very heavy case selection occurs in the application of such strategies, however, and graft versus host disease (GVHD) is a very significant problem.[54]

Germ Cell Tumors of the Testis

The management of metastatic testicular germ cell tumors has become one of the paradigms of cancer treatment, reflecting progress in the development of chemotherapy and an improved understanding of the principles of tumor biology and of the importance of multidisciplinary management. An increased understanding of risk factors has allowed oncologists to tailor treatment to the level of risk. Because it has been determined that some treatment strategies are associated with significant late effects, modifications of the standard treatment regimens are cautiously being explored in an effort to reduce these late effects.

EPIDEMIOLOGY AND ETIOLOGY

Testicular cancer appears to be increasing in frequency in Western society. At the beginning of the 20th century, about two to three new cases per 100,000 males were identified each year. At present, the annual incidence appears to be approximately five to seven new cases per 100,000 males. This illness is found predominantly in white men and is uncommon in black and Asian populations. The etiology is unknown, although specific associations have been identified with testicular maldescent, infertility, family history of testicular cancer, carcinoma in situ (or atypical germ cells), the presence of a specific abnormality of the short arm of chromosome 12, and multiple atypical nevi.[55-61]

PATHOGENESIS

Most germ cell tumors arise from tissues derived from primordial cells that originate within the genital ridges and that usually migrate in the midline to the testicles. Less frequently, these cells migrate to the retroperitoneum, mediastinum, and pineal region; these are the source of extragonadal germ cell tumors. Two major groups of tumors, seminomas and nonseminomatous germ cell tumors (NSGCTs), form at these sites; the two groups arise from a common origin.[55,56] In the testicles, the tumors arise from precursor cells, atypical germ cells, or carcinoma in situ.[57] Less than 5% of testicular cancers consist of lymphomas and other non–germ cell tumors.[55,62] The different development pathways of these tumors are of particular importance with regard to late relapse (see below).

The most common type of seminoma, the classic variant, consists of uniform round or polygonal cells with abundant cytoplasm and a centrally placed nucleolus.[55] Less commonly, spermatocytic and anaplastic variants have been described, although it should be noted that the anaplastic variant sometimes represents a misdiagnosed NSGCT.[63] About 30% of seminomas include syncytiotrophoblastic giant cells; these may produce a

marker protein, human chorionic gonadotropin (hCG). A subgroup of seminomas has been identified that morphologically resemble a solid variant of yolk-sac carcinoma, with a worse prognosis than other seminomas when treated with conventional radiotherapy or standard chemotherapy regimens.[63] New biochemical determinants of outcome identified in studies of the human genome may explain the differences in the natural history of these tumors—for example, aberrations of expression of c-*kit* or other genes that regulate the cell cycle.

NSGCTs have several histologic subtypes: embryonal carcinoma, mature and immature teratoma, endodermal sinus tumor, and choriocarcinoma.[55] Many NSGCTs include several of these elements: histologic analysis may show undifferentiated cancer; trophoblastic tissue; and varying components of cartilage, glandular tissue, or hair.[55,56] Some elements of NSGCT may produce hCG, whereas endodermal sinus tumor classically produces α-fetoprotein (AFP). Embryonal carcinoma may also produce AFP. The presence of AFP in the circulation signifies that a tumor has nonseminomatous elements, even if the primary tumor has been diagnosed as a pure seminoma, and this constitutes an indication to treat the patient for nonseminomatous germ cell malignancy.

Testicular seminomas and NSGCT have many features in common, including age at presentation; an orderly pattern of spread from the testis to the surrounding supportive tissues or up the spermatic cord to regional and distant lymphatic channels and, sometimes, to visceral sites via blood-borne metastasis; and the elaboration of tumor markers. [55,56] Both have a characteristic marker on the short arm of chromosome 12[58,59] and are associated with testicular maldescent, carcinoma in situ of the testis,[60] and an unexplained association with the syndrome of multiple atypical nevi.[60] Both histologic patterns have been linked to a susceptibility gene localized to Xq27,[61] and tumors of both types are highly sensitive to chemotherapy.[64,65]

There are also important differences between testicular seminomas and NSGCTs. With seminoma, the age range at presentation is older, the prevalence of subsequent second primary tumors is slightly higher, and the patterns of metastasis are different. Seminomas have marked radiosensitivity, whereas NSGCTs have marked radioresistance. The c-*kit* ligand is expressed more heavily on seminomas than on NSGCTs.[66,67] It was thought that seminomas would be more sensitive than NSGCTs to treatment with modulators of these ligands, such as imatinib mesylate (Gleevec), but this does not seem to be true of the currently available modulators. Extragonadal seminomas (dysgerminomas) are sensitive to chemotherapy or radiotherapy and are relatively easily cured. In contrast, extragonadal NSGCTs have much higher relapse rates and have cure rates that are far inferior to their testicular counterparts.[68-70] The reasons for these differences are unknown, given that testicular and extragonadal tumors share virtually identical histologies and etiologic associations and have similar patterns of spread.

DIAGNOSIS

Clinical Findings

Primary testicular cancer commonly presents as a painless enlargement of the testis, often noted on self-examination.[71] (Despite popular advocacy of testicular self-examination [TSE], however, there are no randomized trial data to prove that TSE programs improve testicular cancer survival.) There may be local pain, which is often associated with hemorrhage within the tumor.[72] Occasionally, a large tumor will drag on the spermatic cord, causing referred pain in the region of the flank; therefore, this symptom does not necessarily indicate the presence of retroperitoneal lymph node metastases. Unless the patient has previously undergone surgery for testicular maldescent or has had scrotal violation as part of the initial approach to the primary tumor (a cardinal error of management), it is uncommon for inguinal lymph nodes to be involved. Lymphatic drainage of the testis is classically to the nodes at the level of the ipsilateral renal hilum,[73] although contralateral spread will sometimes occur unexpectedly.

The symptoms of metastatic germ cell tumors are protean and depend on the sites of involvement.[74,75] Most commonly, the retroperitoneal lymph nodes are involved early, which can result in lumbar backache or in central or lower abdominal pain. Occasionally, renal colic occurs from ureteral obstruction by advanced lymph node metastases. Pulmonary metastases may be asymptomatic or, if extensive, may be associated with dyspnea, cough, hemoptysis, or chest pain. Brain metastases, although uncommon, may present as headache, confusion, dementia, or focal neurologic syndromes and, occasionally, may be detected in the routine staging workup.[76,77] Liver and bone involvement, although seen frequently in the past, are now distinctly uncommon at presentation but may be associated with relapse. When these features are seen at presentation, choriocarcinoma should be considered to be the most likely histologic subtype. Seminoma will occasionally metastasize to bone, however.

Germ cell tumors may be associated with nonspecific or constitutional symptoms, including weakness, sweats, fevers, malaise, and asthenia, especially in the case of advanced disease. Gynecomastia may signal the presence of testicular cancer and is often associated with raised circulating levels of hCG; curiously, this condition sometimes occurs after the completion of therapy and reflects only changes in the hormonal milieu and does not necessarily signal a relapse.

The presentation of extragonadal germ cell tumors is similar to that of gonadal tumors but is more heavily influenced by the site of origin. For example, pineal germ cell tumors may be associated with headache, confusion, visual changes, strokelike syndromes, and other neurologic presentations that reflect the site of intracranial involvement.[68] The presentation of retroperitoneal germ cell tumors is identical to that of testicular tumors with retroperitoneal metastases, but retroperitoneal germ cell tumors do not have an obvious testicular primary tumor. Mediastinal germ cell tumors are commonly associated with dyspnea, cough, hemoptysis, chest or back pain, and, occasionally, superior vena cava syndrome. Mediastinal NSGCT is commonly associated with metastases at presentation, and the symptoms will reflect the sites of involvement.

Laboratory Tests

Imaging techniques are used both in the initial assessment of the extent of disease and serially (along with repeated physical examination) to monitor efficacy of treatment. Plain radiographs and CT scanning constitute the standard modalities; CT has almost completely replaced lymphangiography.[78] MRI scans are superior to CT for the study of brain metastases. Although there is preliminary evidence that PET scanning with labeled fluorodeoxyglucose may be useful in assessing for residual viable cancer after chemotherapy for germ cell tumors, this technology has not yet secured a defined place in management because of conflicting reports regarding specificity and sensitivity.[79,80]

Table 3 Stage Classification of Testicular Cancer

Clinical Involvement	*Classification System*		
	Royal Marsden[65]*	*Conventional United States*	*AJCC 2002[39]*
Testis only	I	A	I
Abdominal nodes	II	B	IIA-IIC
Mediastinum/ supradiaphragmatic	III	C	IIIA
Visceral metastases	IV	C	IIIA-C

*The Royal Marsden system also contains subclassifications (A, B, C), which depend on the extent of disease.
AJCC—American Joint Commission on Cancer

The circulating tumor markers AFP and hCG should be measured as part of the initial diagnostic workup and in the monitoring of therapy. The first specimens should be obtained before primary surgery. AFP has a normal half-life in the circulation of 5 to 7 days, and hCG has a half-life of 24 to 36 hours.[81] When levels of a circulating marker fail to decline according to normal half-life gradients after orchiectomy, this suggests continuing release of the marker into the blood by occult metastatic disease. Such cases require further investigation and treatment.[82] In patients with metastatic disease, the absolute levels of tumor markers in the blood constitute independent prognostic determinants (see below). During chemotherapy, the release of markers from dying cancer cells may result in a transient elevation of blood levels, after which they will decline according to normal half-life gradients. Thus, serial measurements should be taken to determine whether the patient is responding adequately to treatment. If these repeated measurements are not taken during chemotherapy, marker release will not be identified and the calculation of tumor marker half-life will be incorrect, suggesting a prolonged tumor-marker decline and incorrectly implying the presence of resistant disease.

Another important, but nonspecific, tumor marker is lactate dehydrogenase (LDH). Although LDH levels are elevated in many conditions, both benign and malignant, levels 10 or more times higher than the upper limit of normal are associated with a significantly worse outcome with standard chemotherapy (see below).

Staging

The extent of germ cell cancer is classified by standard staging systems. The staging regimen includes both noninvasive imaging methods and, in some cases of NSGCT, retroperitoneal lymph node dissection. In recent years, the European and United States classifications for testicular cancer have become quite similar [*see Table 3*].

MANAGEMENT

Stage I or A Disease

One of the most controversial issues in the management of germ cell tumors is the treatment of patients without evidence of metastatic disease (stage I or A disease).[83] Traditionally, stage I seminomas were treated with inguinal orchiectomy and adjuvant radiotherapy (originally with doses of around 35 to 40 Gy; later, with 25 to 30 Gy). With this approach, cure rates approached 90% to 95%, especially after the introduction of CT scanning and gallium scanning as part of the diagnostic workup. Today, many institutions use active surveillance, an approach involving close observation after inguinal orchiectomy, with serial measurement of tumor markers and repeated CT scans. Although this approach was initially developed for NSGCT, recent studies have suggested that it may have safe application to the management of seminomas, if it is implemented meticulously.[84,85]

Traditionally, stage I (A) NSGCT has been managed with inguinal orchiectomy, followed by a radical retroperitoneal lymph node dissection (RPLND). This has led to cure in more than 90% of cases and has even cured some patients with histologic evidence of lymphatic micrometastases without the need for chemotherapy. If extensive lymph node metastasis is identified at surgery, chemotherapy is usually required to achieve cure. The major disadvantage of this surgical approach has been the risk of surgical complications, including perioperative pain, pulmonary emboli, accumulations of lymphatic fluid, and hemorrhage. Of particular concern to men in the reproductive years is incompetent ejaculation secondary to transection of retroperitoneal nerves; such patients can achieve orgasm, but the interrupted nerve pathways preclude successful ejaculation of semen.

The potential for recurrence of stage I testicular cancer is indicated by adverse prognostic factors, including local extension to spermatic cord and surrounding structures, undifferentiated histology (embryonal carcinoma), and lymphatic-vascular invasion.[82,86] In addition, failure of tumor markers to normalize appropriately is associated with an increased risk of relapse[82] and constitutes an indication for chemotherapy. Choriocarcinoma has been suggested as an adverse prognostic determinant, but multivariate analysis has failed to support this because of both small case numbers and the significant association with other factors, such as cord involvement and failure of hCG to normalize.

Adjuvant radiotherapy does not increase survival after orchiectomy for stage I NSGCT[65,87] and is therefore no longer used. Instead, at many centers orchiectomy for stage I NSGCT is now followed with active surveillance rather than radiotherapy.[85,88] This approach, predicated on the high cure rates for good-risk metastatic disease, requires a meticulous initial diagnostic workup, with the exclusion of patients with established lymph node or visceral, small-volume metastases. For patients without major risk factors and with confirmed normal results on imaging studies, the relapse rate after orchiectomy alone is less than 25%. Thus, active surveillance should spare 75% of patients from unnecessary adjuvant therapy (RPLND or radiotherapy).[89] The trade-off is that because the risk of relapse is higher with surveillance than after RPLND, fewer patients undergo extensive abdominal surgery but, in exchange, more patients need subsequent chemotherapy.[83]

The most important risk associated with active surveillance involves lack of compliance, either through patients who do not adhere to the follow-up schedules or through physicians who are inexperienced and insufficiently familiar with the requirements of a rigorous program of follow-up. In either instance, patients who experience recurrences may present with far-advanced disease, with a lower potential for cure.

There is no single follow-up schedule that can be considered optimal.[90] However, the successful approaches all include meticulous initial diagnosis and workup and a structured approach to

routine follow-up, with serial assessment of clinical status and tumor markers and repeated chest radiographs and CT scans [*see Table 4*]. Over time, less frequent CT scans are performed, but frequent tumor-marker measurements and chest x-rays remain necessary. In centers with clinicians who are experienced in the conduct of surveillance programs and where a clear policy of patient selection and follow-up has been defined, cure rates of 90% to 95% or greater have been reported with active surveillance.[83] Occasionally, patients have relapses that prove intractable to therapy, and rarely, patients develop chemotherapy-related complications, such as leukemia.[91,92] Nevertheless, cure rates with adjuvant RPLND or radiotherapy protocols also are not 100%, so active surveillance remains a reasonable option. In choosing between active surveillance and adjuvant therapy, it is important for the clinician to explain the various options in detail and to involve the patient in the decision.

The appropriate management of stage I disease in patients with adverse risk factors is also controversial. Adjuvant chemotherapy after orchiectomy can prevent early relapse in patients who have locally extensive tumors with involvement of spermatic cord and other high-risk factors.[93] However, in patients who would not have suffered relapses, this approach results in unnecessary exposure to the toxicity of chemotherapy. Given that up to 50% of patients with high-risk stage I disease do not have relapses,[59,82,90] close observation may be a reasonable approach in the hands of an experienced practitioner. Another appropriate option is RPLND, which can define lymph node involvement, thus identifying patients who are likely to require chemotherapy. The drawback of RPLND is that some patients without evidence of retroperitoneal lymph node metastasis experience relapse in the lungs or other visceral sites. Ideally, a patient with high-risk stage I testicular cancer should be referred to a center of excellence for a second opinion before a final management plan is determined.

Lymph Node Metastases of 5 cm or Less in Diameter

In patients with seminoma who have evidence of small-volume lymph node involvement (stage IIA or IIB; stage B), treatment is usually with radiotherapy to the ipsilateral lymph nodes in the pelvis, with extension to the para-aortic chain (including the involved nodes).[94] Although there has been a recent tendency toward dose reduction to a level of 25 Gy, many clinicians use a more traditional dose of 30 to 35 Gy to ensure local tumor control. There is clear evidence that prophylactic mediastinal lymph node irradiation does not improve outcome and may, in fact, compromise subsequent chemotherapy, should that be needed. The preliminary experience with doses of 20 to 25 Gy suggests no increase in the pattern of early relapse, but longer follow-up is required to determine whether these lower doses are truly safe.

In NSGCT, the optimal management of early stage II disease is controversial. There is general agreement that radiotherapy has no role. However, the advocates for RPLND believe that surgery offers both diagnosis and definitive treatment; they cite surgical cure rates of up to 50%, particularly in patients with only microscopic evidence of lymph node involvement.[95] With this approach, chemotherapy can be used for salvage in most cases of relapse.[96] In many centers, adjuvant chemotherapy is administered routinely for patients with more than three to five nodes containing tumor.

The alternative approach is to offer first-line cytotoxic chemotherapy to such patients on the basis of cure rates approaching 100% for patients with small-volume metastases limited to the retroperitoneal nodes. The disadvantages of each approach are the morbidity from surgery and, with chemotherapy, the acute toxicity and the potential for late complications. No randomized trial has compared surgery with chemotherapy for small-volume stage II (B) disease. A randomized trial of adjuvant chemotherapy versus observation with salvage chemotherapy at the time of relapse in patients with stage II NSGCT showed no statistically significant difference in outcome.[97] Given the lack of data supporting any one approach, the choice of treatment is usually determined by the biases of the clinician and the preferences of the patient.

Another controversial issue is the management of patients with marker-only disease (i.e., without evidence of specific lymph node or other metastases but with persistent elevation of serum marker levels after orchiectomy). RPLND may reveal occult lymphatic involvement in some cases, achieving cure in up to half of these; however, 50% of the patients who undergo

Table 4 Active Surveillance Schedules for Stage A (or I) Nonseminomatous Germ Cell Tumors[138,139]

Year after Inguinal Orchiectomy	*Test*	*Annual Schedule (Number of Tests) by Institution*				
		Royal Marsden Hospital	*Memorial Sloan-Kettering*	*University of Southern California*	*Charing Cross Hospital*	*Princess Margaret Hospital*
1	H&P	12	12	8-12	12	12
	Tumor markers	12	12	8-12	25	12
	Chest x-ray (CT scan)	6 (4)	12(*)	6 (1†)	12 (4)	12 (1)
	CT of abdomen and pelvis	4	5	4-5	4	4
2–3	H&P	6	6,4	4	6,4	6,4
	Tumor markers	6	6,4	4	12,6	6,4
	Chest x-ray	6	6,4	4	6,4	6,4
	CT of abdomen and pelvis	1	3,2	2–3	6,6 (seminoma only)	3,0
4–5	H&P	2?	3,2	2–3	2–4	3,2
	Tumor markers	2?	3,2	2–3	2–4	3,2
	Chest x-ray/CT scans of abdomen and pelvis	2?/0	3,2/2	2–3/1–2	2–3	3,2/0

Note: More frequent follow-up and CT scans are indicated for patients with seminoma after 3 yr because of the risk for late relapse.
*Indicated if abnormality detected on chest x-ray.
†Chest CT scan repeated if baseline scan is abnormal.
H&P—history and physical examination

surgery will eventually require chemotherapy. As with small-volume nodal metastases, the cure rates with cytotoxic chemotherapy are close to 100%, although there is a risk of additional late toxicity compared with surgery alone.

Advanced Lymph Node Involvement and Distant Metastases

For patients with lymph node metastases measuring more than 5 cm in diameter (stage IIC or stage C disease) and for those with visceral metastases (e.g., lung, liver, bone), the treatment of choice is systemic chemotherapy.[96,98] The early combination of cisplatin, vinblastine, and bleomycin (the PVB regimen) produced cures in up to 70% of cases.[64,65] Subsequently, a less toxic regimen, in which vinblastine was replaced by etoposide, was tested and shown to be equivalent to the PVB regimen.

Certain prognostic features identify patients who are likely to experience inferior outcomes with standard cisplatin-based chemotherapy and, instead, require more intensive front-line chemotherapeutic regimens.[99-101] Adverse prognostic factors include greatly elevated serum tumor marker levels,[96,99,100,102,103] the volume of disease,[100,101] the number of pulmonary metastases,[102,103] and certain sites of involvement.[96,100,102,103] In the last instance, metastatic disease in the brain, liver, and possibly bone is associated with a worse prognosis with standard PVB therapy.[103]At the Memorial Sloan-Kettering Cancer Center, a numerical algorithm for attribution of prognosis, predicated on tumor marker levels and extent of disease,[99] has been used to identify patients requiring specific dose-intensive treatment strategies.[104]

Although most systems classify these tumors into similar prognostic groups, comparing results was difficult in the past because of subtle differences in classifications that nevertheless had substantial effects on the results of treatment.[105] As a result, an international collaboration led to a common system of classification to ensure uniform reporting of treatment results [*see Table 5*].[106] This classification, based on a cumulative experience with 5,202 NSGCTs and 660 seminomas in 10 countries, includes three prognostic groupings that are based on 5-year survival: good (91% 5-year survival), intermediate (79%), and poor (48%). However, it must be emphasized that cure is possible in patients with tumors in the worst-prognosis group; even patients with brain metastases may be cured by aggressive multimodality treatment.[76,77]

For patients with good- or intermediate-risk disease, replacement of vinblastine with etoposide in combination chemotherapy (PEB chemotherapy) has reduced acute toxicity without lowering high cure rates.[107] However, recent studies have shown occasional cases of acute leukemia in patients who have received etoposide-containing regimens, so this strategy will require further reassessment. Deletion of bleomycin in an attempt to reduce toxicity has resulted in lower cure rates[108-110] and has been abandoned in most centers. However, investigators at the Memorial Sloan-Kettering Cancer Center believe that bleomycin can be safely omitted, provided that the dose of etoposide is maintained at 500 mg/m^2.[111]

Three cycles of PEB chemotherapy have been shown to be equivalent to four cycles with respect to cure rate and are associated with less toxicity.[112] In the study demonstrating this equivalence, however, the patient numbers were not large enough to exclude a substantial difference in survival for patients with intermediate-prognosis disease. It will be important to maintain caution in the development of new trials that have the reduction of side effects of therapy as a primary end point, because there is the risk that ameliorating toxicity will be achieved at the price of reducing cure rates, as was found with bleomycin. In addition, randomized trials have shown that cure rates are reduced if cisplatin is replaced by carboplatin in combination-chemotherapy schedules for germ cell tumors.[113]

Table 5 International Germ Cell Tumor Group Classification of Metastatic Disease

*Criteria for Poor Risk Nonseminomatous Germ Cell Tumors**
Mediastinal primary site
Serum α-fetoprotein (AFP) level > 10,000 ng/ml
Serum human chorionic gonadotropin (HCG) level > 50,000 ng/ml
Serum lactate dehydrogenase (LDH) level > 10 times the upper limit of normal
Presence of nonpulmonary visceral metastases

*No seminomas are classified as poor risk.

Poor-Risk and Previously Treated Metastatic Germ Cell Tumors

Early studies of poor-risk metastatic disease were confounded by differences in systems of risk classification. For example, it was initially reported that a high-dose combination of cisplatin, etoposide, vinblastine, and bleomycin (PVeBV) yielded unsurpassed objective response rates and apparently increased survival in patients with so-called poor-risk disease. However, a randomized trial comparing this regimen with a standard-dose regimen failed to show any evidence of a significant benefit from PVeBV; these investigators suggested that the original difference in outcomes related to differences in risk attribution.[114]

Several recent reports have suggested that high-dose therapy with autologous bone marrow transfusion or peripheral stem cell support may improve outcome in patients with truly poor-risk disease (consistent with the international consensus classification)—in particular, those who have experienced relapses of germ cell tumors after initial chemotherapy and those with mediastinal NSGCT.[38,115-117] However, more extensive information is required from randomized clinical trials before a definitive statement on this topic can be made. At present, the North American Intergroup is conducting a study in which patients with poor-risk metastatic disease are randomly allocated to treatment with a standardized PEB regimen or to a similar regimen followed by high-dose therapy with autologous rescue.

One approach to poor-risk disease involves rapid cycling, dose-intense therapy (within the conventional dose range), with alternating delivery of cisplatin-vincristine-methotrexate-bleomycin (POMB regimen) and actinomycin D–cyclophosphamide-etoposide (ACE).[118] At Charing Cross Hospital, long-term survival with POMB/ACE treatment was recorded in 75% of poor-risk cases. The value of this regimen remains controversial, however.[118,119] Recently, the novel cytotoxic agents paclitaxel[120] and gemcitabine[121,122] have demonstrated activity in previously treated patients. The utility of these agents is currently being defined in a series of phase II and phase III clinical trials that are not yet complete. There are no final data reflecting their activity in previously untreated cases, but there may be an objective single-agent response rate of up to 20% in patients who have previously received cisplatin-containing regimens, depending on whether they have secured initial remissions in response to first-line cis-

Table 6 Late Toxicity of Treatment for Testis Cancer

Type	*Toxicity*
Cardiovascular	Hypertension Coronary artery disease Peripheral vascular disease Cerebrovascular disease Raynaud phenomenon Hyperlipidemia
Gastrointestinal	Peptic ulceration Abdominal pain and diarrhea
Neurologic	Peripheral neuropathy Autonomic neuropathy Hearing loss
Psychosocial	Marital problems Infertility Employment problems Legal trouble and sociopathic behavior Depression
Pulmonary	Pneumonitis and fibrosis
Renal	Renal failure and nephritis Hypomagnesemia Hyperuricemia
Second malignancies	Leukemia Melanoma and dysplastic nevi Other solid tumors

platin-containing regimens. These novel agents have been incorporated into combination regimens. For example, treatment with paclitaxel, ifosfamide, and cisplatin has produced an objective response rate of 77% in patients with relapses of germ cell tumors.[52] In heavily treated patients, the combination of paclitaxel and gemcitabine has yielded a response rate of 20%, with occasional long-term survival.[123] It is likely that these agents will form the basis of many future trials, especially trials that focus on poor-risk previously untreated cases.

Investigational protocols have been developed that incorporate some of the concepts enumerated above—for example, rapid cycling of dose-intense regimens of paclitaxel-ifosfamide and carboplatin-etoposide, which appear to yield sustained remission beyond 30 months in around 50% of patients with an unfavorable prognosis for salvage response.[52] Although there is no uniform classification of unfavorable prognosis for salvage treatment, a reasonable approach has been proposed by investigators at the Memorial Sloan-Kettering Cancer Center, which includes extragonadal primary site, progressive disease after an incomplete response to first-line platinum therapy, or poor response or lack of response to prior treatment with cisplatin plus ifosfamide-containing conventional-dose therapy.

Similar data have been reported by Rick and colleagues,[124] with a 30% long-term survival in patients who have experienced relapses or whose cancers have been refractory to treatment with cisplatin-based regimens. In this study, conventional-dose salvage therapy with cisplatin, ifosfamide, and paclitaxel was followed by one cycle of high-dose chemotherapy with thiotepa, carboplatin, and etoposide. In most of these studies, the impact of case-selection bias and improved supportive care versus that of high-dose therapy have not been clearly defined.

Treatment of Patients with Late Relapse

Late failure of treatment—that is, failure that occurs well beyond the period usually expected for relapse—is an issue of increasing concern. Centers that manage large numbers of patients with germ cell tumors are beginning to occasionally see patients with relapse after more than 10 years, characterized by malignant teratoma or the evolution of non–germ cell elements associated with germ cell tumors (including adenocarcinoma, soft tissue sarcoma, or neuroendocrine carcinomas). These cases are proving to be difficult to treat effectively and are associated with a high mortality. As yet, there is no defined optimal approach to treatment. At the University of Southern California, our management generally depends on the histology of the late relapse. Our management is usually multidisciplinary, and it particularly involves the combination of cytoreductive surgery and potentially non–cross-resistant chemotherapy. Gemcitabine and the taxanes appear to be quite active in this context, especially for adenocarcinomas that arise from germ cell tumors, and ifosfamide-based regimens seem to be effective against tumors with sarcomatous differentiation. Because optimal management of late relapses has not yet been defined, we believe that such cases should be treated in tertiary referral centers.

TOXICITY OF CHEMOTHERAPY

The acute toxicity of chemotherapy has been well defined and includes nausea and vomiting, myelosuppression, alopecia, allergic phenomena, pneumonitis, infection, anorexia, and a range of relatively uncommon complications.[125,126] Most of these effects can be controlled by modern supportive techniques [*see 12:IV Principles of Cancer Treatment*].

The long-term or delayed side effects of treatment are now becoming increasingly recognized, especially because the medical community has become used to the concept of germ cell tumors as a curable entity and because the focus is now shifting to the avoidable costs of such cure [*see Table 6*]. Of particular concern is an apparent increase in the prevalence of cardiovascular and cerebrovascular diseases, hypercholesterolemia, and a range of subtle metabolic abnormalities after cisplatin-based chemotherapy.[127-129] Perhaps more sinister is the prevalence of second malignancies, including leukemia, soft tissue sarcoma, malignant melanoma, and other solid tumors.[125,130,131]

Given that some of these delayed effects have been identified in patients 5 to 10 years after treatment, it is possible that the reported prevalence figures are low and will increase with the duration of follow-up or that other unsuspected problems will emerge. For that reason, it is essential for careful and focused follow-up to be continued in such cases, despite the efforts of many health maintenance organizations to reduce structured follow-up by specialists by returning the care of these patients to primary care physicians.

The author has received research grant support from Eli Lilly and Company, Bristol-Myers Squibb Co., and Aventis Pharmaceuticals, Inc., and has acted as a consultant for these companies.

References

1. Raghavan D, Shipley WU, Garnick MB, et al: Biology and management of bladder cancer. N Engl J Med 322:1129, 1990

2. Zhang, Z-F, Steineck G: Epidemiology and etiology of bladder cancer. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 215

3. Sternberg CN, Swanson DA: Non–transitional cell bladder cancer. Principles and

Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 315

4. Russell PJ, Wills EJ, Philips J, et al: Features of squamous and adenocarcinoma in the same cell in a xenografted human transitional cell carcinoma: evidence of a common histogenesis? Urol Res 16:79, 1988

5. Spruck CH III, Ohneseit PF, Gonzalez-Zulueta M, et al: Two molecular pathways to transitional cell carcinoma of the bladder. Cancer Res 54:784, 1994

6. Stein JP, Ginsberg DA, Grossfeld GD, et al: Effect of p21 expression on tumor progression in bladder cancer. J Natl Cancer Inst 90:1072, 1998

7. Bi J, Raghavan D: Tumors of kidney, ureter and bladder. Kelley's Textbook of Internal Medicine. Humes HD, Ed. Williams & Wilkins, Baltimore, 2000, p 1248

8. Raghavan D, Chye RWM: Treatment of carcinomatous meningitis from transitional cell carcinoma of the bladder. Br J Urol 67:438, 1991

9. Esrig D, Elmajian D, Groshen S, et al: Accumulation of nuclear p53 and tumor progression in bladder cancer. N Engl J Med 331:1259, 1994

10. Petrylak DP, Scher HI, Reuter V, et al: P-glycoprotein expression in primary and metastatic transitional cell carcinoma of the bladder. Ann Oncol 3:835, 1994

11. Grossfeld GD, Ginsberg DA, Stein JP, et al: Thrombospondin-1 expression in bladder cancer: association with p53 alterations, tumor angiogenesis and tumor progression. J Natl Cancer Inst 89:219, 1997

12. Pendyala L, Velagapudi S, Toth K, et al: Translational studies of glutathione in bladder cancer cell lines and human specimens. Clin Cancer Res 3:793, 1997

13. Herr HW, Lamm DL, Denis L: Management of superficial bladder cancer. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 273

14. Richie JP: Surgery for invasive bladder cancer. Hematol Oncol Clin North Am 6:129, 1992

15. Stein JP, Lieskovsky G, Cote R, et al: Radical cystectomy in the treatment of invasive bladder cancer: long-term results in 1054 patients. J Clin Oncol 19:666, 2001

16. Gospodarowicz MK, Hawkins NV, Rawlings GA, et al: Radical radiotherapy for muscle invasive transitional cell carcinoma of the bladder: failure analysis. J Urol 142:1448, 1989

17. Mameghan H, Fisher RJ, Watt WH, et al: The management of invasive transitional cell carcinoma of the bladder: results of definitive and preoperative radiation therapy in 390 patients treated at the Prince of Wales Hospital, Sydney, Australia. Cancer 69:2771, 1992

18. Coppin C, Gospodarowicz M, James K, et al: Improved local control of invasive bladder cancer by concurrent cisplatin and preoperative or definitive radiation. J Clin Oncol 14:2901, 1996

19. Shipley WU, Winter KA, Kaufman DS, et al: Phase III trial of neoadjuvant chemotherapy in patients with invasive bladder cancer treated with selective bladder preservation by combined radiation therapy and chemotherapy: initial results of Radiation Therapy Oncology Group 89-03. J Clin Oncol 16:3576, 1998

20. Natale RB, Grossman HB, Blumenstein B, et al: SWOG 8710 (INT-0080): randomized phase III trial of neoadjuvant MVAC + cystectomy versus cystectomy alone in patients with locally advanced bladder cancer. Proc Am Soc Clin Oncol 20:abst 3, 2001

21. Neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: a randomised controlled trial. International collaboration of trialists. Lancet 354:533, 1999

22. Stockle M, Meyenburg W, Wellek S, et al: Advanced bladder cancer (stages pT3b, pT4a, pN1 and pN2): improved survival after radical cystectomy and 3 adjuvant cycles of chemotherapy: results of a controlled prospective study. J Urol 148:302, 1992

23. Freiha F, Reese J, Torti FM: A randomized trial of radical cystectomy plus cisplatin, vinblastine and methotrexate chemotherapy for muscle invasive bladder cancer. J Urol 155:495, 1996

24. Sternberg CN, Yagoda A, Scher HI, et al: M-VAC (methotrexate, vinblastine, doxorubicin and cisplatin) for advanced transitional cell carcinoma of the urothelium. J Urol 139:461, 1988

25. Loehrer PJ, Einhorn LH, Elson PJ, et al: A randomized comparison of cisplatin alone or in combination with methotrexate, vinblastine, and doxorubicin in patients with metastatic urothelial carcinoma: a cooperative group study. J Clin Oncol 10:1066, 1992

26. Saxman SB, Propert K, Einhorn LH, et al: Long-term follow up of phase III intergroup study of cisplatin alone or in combination with methotrexate, vinblastine, and doxorubicin in patients with metastatic urothelial carcinoma: a cooperative group study. J Clin Oncol 15:2564, 1997

27. Levine E, Raghavan D: MVAC for bladder cancer: time to move forward again. J Clin Oncol 11:387, 1993

28. Redman B, Smith DC, Flaherty L, et al: Phase II trial of paclitaxel and carboplatin in the treatment of advanced urothelial carcinoma. J Clin Oncol 16:1844, 1998

29. Moore MJ, Winquist E, Murray N, et al: Gemcitabine plus cisplatin: an active regimen in advanced urothelial cancer: a phase II trial of the NCIC Clinical Trials Group. J Clin Oncol 17:2876, 1999

30. Kaufman DS, Raghavan D, Carducci M, et al: Phase II trial of gemcitabine plus cisplatin in patients with metastatic urothelial cancer. J Clin Oncol 18:1921, 2000

31. Von der Maase H, Hansen SW, Roberts JT, et al: Gemcitabine and cisplatin versus methotrexate, vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large, randomized, multinational, multicenter phase III study. J Clin Oncol 17:3068, 2000

32. Walther PJ: Uncommon neoplasms of the kidney. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, Eds. Lippincott-Raven, Philadelphia, 1997, p 897

33. Huben RP, Velagapudi S: Uncommon tumors of renal parenchyma. Textbook of Uncommon Cancer, 2nd ed. Raghavan D, Johnson D, Meropol N, et al, Eds. Wiley-Liss, Chichester, New York, 1999, p 3

34. Ellis WJ: Epidemiology and etiology of renal cell carcinoma. Principles and Practice of Genitourinary Oncology. Ragavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 795

35. Linehan WM, Lerman MI, Zbar B: Identification of the von Hippel–Lindau (VHL) gene: its role in renal cancer. JAMA 273:564, 1995

36. True LD, Grignon D: Pathology of renal cancers. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 799

37. Wood DP Jr, Montie JE: Renal cell carcinoma: clinical diagnosis and prognosis. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 823

38. Motzer RJ, Bander NH, Nanus DM: Renal cell carcinoma. N Engl J Med 335:865, 1996

39. Genitourinary sites. AJCC Cancer Staging Manual, 6th ed. American Joint Committee on Cancer. Springer, New York, 2002, p 301

40. Zisman A, Pantuck AJ, Dorey F, et al: Mathematical model to predict individual survival for patients with renal cell carcinoma. J Clin Oncol 20:1368, 2002

41. Motzer RJ, Bacik J, Mariani T, et al: Treatment outcome and survival associated with metastatic renal cell carcinoma of non–clear-cell histology. J Clin Oncol 20:2376, 2002

42. Zisman A, Chao DH, Pantuck AJ, et al: Unclassified renal cell carcinoma: clinical features and prognostic impact of a new histological subtype. J Urol 168:950, 2002

43. Marsh CL: Classical surgical approaches to renal cell carcinoma. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 843

44. Novick AC, Streem SB, Montie JE, et al: Conservative surgery for renal cell carcinoma: a single-center experience with 100 patients. J Urol 141:835, 1989

45. Novick AC: Nephron-sparing surgery for renal cell carcinoma. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, Eds. Lippincott-Raven, Philadelphia, 1997, p 861

46. Eschwege P, Saussine C, Steichen G, et al: Radical nephrectomy for renal cell carcinoma 30 mm or less: long term follow-up results. J Urol 155:1196, 1996

47. Rini BI, Vogelzang NJ, Dumas MC, et al: Phase II trial of weekly intravenous gemcitabine with continuous infusion fluorouracil in patients with metastatic renal cell cancer. J Clin Oncol 18:2419, 2000

48. Flanigan RC, Salmon SE, Blumenstein BA, et al: Nephrectomy followed by interferon alfa-2b compared with interferon alfa-2b alone for metastatic renal-cell cancer. N Engl J Med 345:1655, 2001

49. Fyfe G, Fisher RI, Rosenberg SA, et al: Results of treatment of 255 patients with metastatic renal cell carcinoma who received high-dose recombinant interleukin-2 therapy. J Clin Oncol 15:688, 1995

50. Recombinant human interleukin-2, recombinant human interferon alfa-2a, or both in metastatic renal cell carcinoma. Groupe Francais d'Immunotherapie. N Engl J Med 338:1272, 1998

51. Boccardo F, Rubagotti A, Canobbio L, et al: Interleukin-2, interferon-alpha and interleukin-2 plus interferon-alpha in renal cell carcinoma: a randomized phase II trial. Tumori 84:534, 1998

52. Motzer RJ, Mazumdar M, Bacik J, et al: Effect of cytokine therapy on survival for patients with advanced renal cell carcinoma. J Clin Oncol 18:1928, 2000

53. Childs R, Chernoff A, Contentin N, et al: Regression of metastatic renal-cell carcinoma after nonmyeloablative allogeneic peripheral-blood stem-cell transplantation. N Engl J Med 343:750, 2000

54. Rini BI, Zimmerman T, Stadler WM, et al: Allogeneic stem-cell transplantation of renal cell cancer after nonmyeloablative chemotherapy: feasibility, engraftment, and clinical results. J Clin Oncol 20:2017, 2002

55. Mostofi FK, Price EB Jr: Tumors of the Male Genital System. Atlas of Tumor Pathology, 2nd series, Fascicle 8. Armed Forces Institute of Pathology, Washington, DC, 1979, p 1

56. Raghavan D, Neville AM: The biology of testicular tumours. Scientific Foundations of Urology, 2nd ed. Innes-Williams D, Chisholm G, Eds. William Heinemann Medical Books, London, 1982, p 785

57. Skakkebaek NE: Carcinoma-in-situ of the testis: frequency and relationship to invasive germ cell tumours in infertile men. Histopathology 2:157, 1978

58. Atkin NB, Baker MC: i(12p): specific chromosomal marker in seminoma and malignant teratoma of the testis? Cancer Genet Cytogenet 10:199, 1983

59. Rodriguez E, Mathew S, Reuter V, et al: Cytogenetic analysis of 124 prospectively ascertained male germ cell tumors. Cancer Res 52:2285, 1992

60. Raghavan D, Zalcberg J, Grygiel JJ, et al: Multiple atypical nevi: a cutaneous marker of germ cell tumors. J Clin Oncol 12:2284, 1994

61. Rapley EA, Crockford GP, Teare D, et al: Localization to *Xq27* of a susceptibility gene for testicular germ cell tumors. Nature Genetics 24:197, 2000

62. Horwich A, Hamilton CR, Fisher C: Rare tumours of the testis and paratesticular tissues. Textbook of Uncommon Cancer, 2nd ed. Raghavan D, Brecher M, Johnson D, et al, Eds. Wiley-Liss, Chichester, New York, 1998, p 95

63. Raghavan D, Sullivan AL, Peckham MJ, et al: Elevated serum alphafetoprotein and seminoma: clinical evidence for a histologic continuum? Cancer 50:982, 1982

64. Einhorn LH: Testicular cancer as a model for a curable neoplasm: The Richard and Hinda Rosenthal Foundation Award Lecture. Cancer Res 41:3274, 1981

65. Peckham MJ, Barrett A, McElwain TJ, et al: Non-seminoma germ cell tumours (malignant teratoma) of the testis: results of treatment and an analysis of prognostic factors. Br J Urol 53:162, 1981

66. Strohmeyer T, Reese D, Press M, et al: Expression of the c-*kit* proto-oncogene and its ligand stem cell factor (SCF) in normal and malignant human testicular tissue. J Urol 153:511, 1995

67. Bokemeyer C, Kuczyk MA, Dunn T: Expression of stem-cell factor and its receptor c-*kit* proto-oncogene in germ cell tumours. J Cancer Res Clin Oncol 122:301, 1996

68. Boyer MJ, Raghavan D: Extragonadal germ cell tumors. Oxford Textbook of Oncology. Peckham MJ, Pinedo H, Veronesi U, Eds. Oxford University Press, Oxford, 1995, p 2169

69. Liebowitz S, Shivdasani S, Kantoff P: Extragonadal germ cell tumors. American Cancer Society Atlas of Clinical Oncology: Germ Cell Tumors. Raghavan D, Ed. BC Decker, Inc, Hamilton, Ontario (in press)

70. Bokemeyer C, Nichols CR, Droz J-P, et al: Extragonadal germ cell tumors of the mediastinum and retroperitoneum: results from an international analysis. J Clin Oncol 20:1864, 2002

71. Stephen RA: The clinical presentation of testicular tumours. Br J Urol 34:448, 1962

72. Sandeman TF: Symptoms and early management of germinal tumours of the testis. Med J Aust 2:281, 1979

73. Donohue JP, Zachary JM, Maynard BR: Distribution of nodal metastases in nonseminomatous testis cancer. J Urol 128:315, 1982

74. Horwich A, Bajorin D: Testicular cancer: presentation, assessment and prognosis. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel S, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 671

75. Bosl GJ, Motzer RJ: Testicular germ-cell cancer. N Engl J Med 337:242, 1997

76. Raghavan D, Mackintosh JF, Fox RM, et al: Improved survival after brain metastases in non-seminomatous germ cell tumours with combined modality treatment. Br J Urol 60:364, 1987

77. Bokemeyer C, Nowak P, Haupt A, et al: Treatment of brain metastases in patients with testicular cancer. J Clin Oncol 15:1449, 1997

78. Hilton S, Schwartz LH: Radiologic imaging of genitourinary neoplasms. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel SA, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 33

79. Stephens AW, Gonin R, Hutchins GD, et al: Positron emission tomography evaluation of residual radiographic abnormalities in postchemotherapy germ cell tumor patients. J Clin Oncol 14:1637, 1996

80. Hain SF, O'Doherty MJ, Timothy AR, et al: Fluorodeoxyglucose positron emission tomography in the evaluation of germ cell tumours at relapse. Br J Cancer 83:863, 2000

81. Lange PH, Raghavan D: Clinical applications of tumor markers in testicular cancer. The Management of Testicular Cancer. Donohue JP, Ed. Williams & Wilkins, Baltimore, 1983, p 111

82. Raghavan D, Peckham MJ, Heyderman E, et al: Prognostic factors in clinical stage I non-seminomatous germ-cell tumours of the testis. Br J Cancer 45:167, 1982

83. Jewett MAS, Lange PH, Raghavan D: Controversies in the treatment of early stage nonseminomatous testicular cancer. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel S, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 703

84. Warde PR, Gospodarowicz MK, Goodman PJ, et al: Results of a policy of surveillance in stage I testicular seminoma. Int J Radiat Oncol Biol Phys 27:11, 1993

85. Francis R, Bower M, Brunstrom G, et al: Surveillance for stage I testicular germ cell tumours: results and cost benefit analysis of management options. Eur J Cancer 36:1925, 2000

86. Raghavan D, Vogelzang NJ, Bosl GJ, et al: Tumor classification and size in germ-cell testicular cancer: influence on the occurrence of metastases. Cancer 50:1591, 1982

87. Rorth M, Jacobsen GK, von der Maase H, et al: Surveillance alone versus radiotherapy after orchiectomy for clinical stage I nonseminomatous testicular cancer. J Clin Oncol 9:1543, 1991

88. Peckham MJ, Barrett A, Husband JE, et al: Orchiectomy alone in testicular stage I non-seminomatous germ-cell tumours. Lancet 2:678, 1982

89. Raghavan D: 'Active surveillance' for stage I testis cancer: attaining maturity at 21 years. Eur J Cancer 36:1891, 2000

90. Raghavan D, Colls B, Levi J, et al: Surveillance for stage I non-seminomatous germ cell tumours of the testis: the optimal protocol has not yet been defined. Br J Urol 61:522, 1988

91. Segelov E, Raghavan D, Kronenberg H: Acute leukemia following chemotherapy including etoposide for testicular carcinoma. Aust NZ J Med 23:718, 1993

92. Pedersen-Bjergaard J, Daugaard G, Hansen SW, et al: Increased risk of myelodysplasia and leukemia after etoposide, cisplatin and bleomycin for germ-cell tumours. Lancet 338:359, 1991

93. Cullen MH, Stenning SP, Parkinson MC: Short-course adjuvant chemotherapy in high-risk stage I nonseminomatous germ cell tumors of the testis: a Medical Research Council report. J Clin Oncol 14:1106, 1996

94. Gospodarowicz MK, Warde PR: Management of seminoma. American Cancer Society Atlas of Clinical Oncology: Germ Cell Tumors. Raghavan D, Ed. BC Decker, Inc, Hamilton, Ontario (in press)

95. Leisinger HJ, Donohue JP: The role of retroperitoneal surgery in testis cancer. Crit Rev Oncol Hematol 44:71, 2002

96. Stoter G, Loehrer PJ Sr, Levi J: Management of stage II and good risk stage III and IV nonseminomatous germ cell tumors. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel S, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 715

97. Williams SD, Stablein DM, Einhorn LH, et al: Immediate adjuvant chemotherapy versus observation with treatment at relapse in pathological stage II testicular cancer. N Engl J Med 317:1433, 1987

98. Toner GC: Chemotherapy for good-risk metastatic germ cell tumors. American Cancer Society Atlas of Clinical Oncology: Germ Cell Tumors. Raghavan D, Ed. BC Decker, Inc, Hamilton, Ontario (in press)

99. Bosl GJ, Geller NL, Cirrincione C, et al: Multivariate analysis of prognostic variables in patients with metastatic testicular cancer. Cancer Res 43:3404, 1983

100. Bajorin D, Katz A, Chan E, et al: Comparison of criteria for assigning germ cell tumor patients to "good risk" and "poor risk" studies. J Clin Oncol 6:786, 1986

101. Birch R, Williams S, Cone A, et al: Prognostic factors for favorable outcome in disseminated germ cell tumors. J Clin Oncol 4:400, 1986

102. Levi JA, Thomson D, Sandeman T, et al: A prospective study of cisplatin-based combination chemotherapy in advanced germ cell malignancy: role of maintenance and long-term follow-up. J Clin Oncol 6:1154, 1988

103. Mead GM, Stenning SP, Parkinson MC, et al: The second Medical Research Council study of prognostic factors in nonseminomatous germ cell tumors. J Clin Oncol 10:85, 1992

104. Bosl GJ, Geller NL, Vogelzang NJ, et al: Alternating cycles of etoposide plus cisplatin and VAB-6 in the treatment of poor-risk patients with germ cell tumors. J Clin Oncol 5:436, 1987

105. Levine EG, Raghavan D: Treatment of refractory testis cancer: salvage or savage chemotherapy? Eur J Cancer 27:932, 1991

106. International germ cell consensus classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. J Clin Oncol 15:594, 1997

107. Williams SD, Birch R, Einhorn LH, et al: Treatment of disseminated germ-cell tumors with cisplatin, bleomycin, and either vinblastine or etoposide. N Engl J Med 316:1435, 1987

108. Levi JA, Raghavan D, Harvey V, et al: The importance of bleomycin in combination chemotherapy for good prognosis germ cell carcinoma. J Clin Oncol 11:1300, 1993

109. Loehrer PJ Sr, Johnson D, Elson P, et al: Importance of bleomycin in favorable-prognosis disseminated germ cell tumors: an Eastern Cooperative Oncology Group Trial. J Clin Oncol 13:470, 1995

110. Importance of bleomycin in combination chemotherapy for good-prognosis testicular nonseminoma: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group. J Clin Oncol 15:1837, 1997

111. Bajorin DF, Bosl GJ: Bleomycin in germ cell tumor therapy: not all regimens are created equal. J Clin Oncol 15:1717, 1997

112. Einhorn LH, Williams SD, Loehrer PJ, et al: Evaluation of optimal duration of chemotherapy in favorable prognosis disseminated germ cell tumors: a Southeastern Cancer Study Group protocol. J Clin Oncol 7:387, 1989

113. Horwich A, Sleijfer DT, Fossa SD, et al: Randomized trial of bleomycin, etoposide, and cisplatin compared with bleomycin, etoposide, and carboplatin in good-prognosis metastatic nonseminomatous germ cell cancer: a Multiinstitutional Medical Research Council/European Organization for Research and Treatment of Cancer Trial. J Clin Oncol 15:1844, 1997

114. Nicholls CR, Williams SD, Loehrer PJ, et al: Randomized study of cisplatin dose intensity in poor-risk germ cell tumors: a Southeastern Cancer Study Group and Southwest Oncology Group protocol. J Clin Oncol 9:1163, 1991

115. Margolin BK, Doroshow JH, Ahn C, et al: Treatment of germ cell cancer with two cycles of high-dose ifosfamide, carboplatin, and etoposide with autologous stem-cell support. J Clin Oncol 14:2631, 1996

116. Margolin K: High-dose chemotherapy and stem-cell support in the treatment of poor-risk germ cell cancer. American Cancer Society Atlas of Clinical Oncology: Germ Cell Tumors. Raghavan D, Ed. BC Decker, Inc, Hamilton, Ontario (in press)

117. Motzer RJ, Mazumdar M, Bajorin DF, et al: High-dose carboplatin, etoposide, and cyclophosphamide with autologous bone marrow transplantation in first-line therapy for patients with poor-risk germ cell tumors. J Clin Oncol 15:2546, 1997

118. Newlands ES, Begent RH, Gustin GJ, et al: Further advances in the management of malignant teratomas of the testis and other sites. Lancet 1:948, 1983

119. Bower M, Newlands ES, Holden L, et al: Treatment of men with metastatic nonseminomatous germ cell tumours with cyclical POMB/ACE chemotherapy. Ann Oncol 8:477, 1997

120. Motzer RJ, Bajorin DF, Schwartz LH, et al: Phase II trial of paclitaxel shows antitumor activity in patients with previously treated germ cell tumors. J Clin Oncol 12: 2277, 1994

121. Bokemeyer C, Gerl A, Schoffski P, et al: Gemcitabine in patients with relapsed or cisplatin-refractory testicular cancer. J Clin Oncol 17:512, 1999

122. Einhorn LH, Stender MJ, Williams SD: Phase II trial of gemcitabine in refractory germ cell tumors. J Clin Oncol 17:509, 1999

123. Hinton S, Catalano P, Einhorn LH, et al: Phase II study of paclitaxel plus gemcitabine in refractory germ cell tumors (E9897): a trial of the Eastern Cooperative Oncology Group. J Clin Oncol 20:1859, 2002

124. Rick O, Bokemeyer C, Beyer J, et al: Salvage treatment with paclitaxel, ifosfamide, and cisplatin plus high-dose carboplatin, etoposide, and thiotepa followed by autologous stem-cell rescue in patients with relapsed or refractory germ cell cancer. J Clin Oncol 19:81, 2001

125. Boyer M, Raghavan D: Toxicity of treatment of germ cell tumors. Semin Oncol 19:128, 1992

126. Boyer MJ, Roth BJ: Toxicity of treatment of germ cell tumors. Principles and Practice of Genitourinary Oncology. Raghavan D, Scher HI, Leibel S, et al, Eds. Lippincott-Raven, Philadelphia, 1997, p 754

127. Meinardi MT, Gietema JA, van der Graaf WTA, et al: Cardiovascular morbidity in long-term survivors of metastatic testicular cancer. J Clin Oncol 18:1725, 2000

128. Raghavan D, Cox K, Childs A, et al: Hypercholesterolemia after chemotherapy for testis cancer. J Clin Oncol 10:1386, 1992

129. Boyer M, Raghavan D, Harris PJ, et al: Lack of late toxicity in patients treated with cisplatin-containing combination chemotherapy for metastatic testicular cancer. J Clin Oncol 8:21, 1990

130. Van Leeuwen FE, Stiggelbout AM, van den Belt-Dusebout A, et al: Second cancer risk following testicular cancer: a follow-up study of 1909 patients. J Clin Oncol 11:415, 1993

131. Travis LB, Curtis RE, Storm H, et al: Risk of second malignant neoplasms among long-term survivors of testicular cancer. J Natl Cancer Inst 89:1429, 1997

132. Dunst J, Sauer R, Schrott KM, et al: Organ-sparing treatment of advanced bladder cancer: a ten-year experience. Int J Radiat Oncol Biol Phys 30:261, 1994

133. Housset M, Maulard C, Chretien Y, et al: Combined radiation and chemotherapy for invasive transitional cell carcinoma of the bladder: a prospective study. J Clin Oncol 11:2150, 1993

134. Kaufman DS, Shipley WU, Griffin PP, et al: Selective bladder preservation by combination treatment in invasive bladder cancer. N Engl J Med 329:1377, 1993

135. Sauer R, Dunst J, Altendorf-Hofmann A, et al: Radiotherapy with and without cisplatin in bladder cancer. Int J Radiat Oncol Biol Phys 19:687, 1990

136. Tester W, Porter A, Asbel S, et al: Combined modality program with possible organ preservation for invasive bladder carcinoma: results of RTOG protocol 85-12. Int J Radiation Oncol Biol Phys 25:783, 1993

137. Five-year followup of a prospective trial of radical cystectomy and neoadjuvant chemotherapy: Nordic Cystectomy Trial I. The Nordic Cooperative Bladder Cancer Study Group. J Urol 155:1903, 1996

138. Raghavan D: Follow up of the patient with testicular cancer. Cancer Patient Follow Up. Johnson FS, Virgo K, Eds. Mosby-Year Book, St. Louis, 1995

139. Raghavan D: Germ cell tumors: cure and consequences. Oncology (in press)

XV CHRONIC LYMPHOID LEUKEMIAS AND PLASMA CELL DISORDERS

Bruce D. Cheson, m.d.

The chronic lymphoid leukemias are a group of relatively indolent clonal lymphoid disorders, primarily of B cell lineage, that includes chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy-cell leukemia (HCL), and the leukemic phase of non-Hodgkin lymphoma (NHL).

The plasma cell dyscrasias are characterized by an accumulation of malignant plasma cells in the bone marrow, bone, or soft tissues. The name reflects the finding on bone biopsy of multiple tumors of plasma cells, rather than a single discrete mass. These diseases include multiple myeloma (MM), extramedullary plasmacytoma, and amyloidosis. Waldenström macroglobulinemia, characterized by IgM gammopathy, lymphadenopathy, and hepatosplenomegaly, was formerly included with these disorders but is now more appropriately classified as a lymphoplasmacytic, indolent non-Hodgkin lymphoma.[1]

Chronic Lymphocytic Leukemia

EPIDEMIOLOGY

CLL is the most common form of leukemia in adults in Western countries. The annual incidence of this disease appears to be decreasing in the United States, from 10,000 new cases a year a decade ago to an estimated 7,300 in 1998.[2] This decline may reflect, in part, a more appropriate classification of other lymphoid malignancies previously misdiagnosed as CLL. The median age at diagnosis is 62 years, with only 10% to 15% of cases diagnosed before age 50. CLL appears to be more common in males than in females and more common in Jewish people of Russian or Eastern European ancestry than in the general population. It occurs with similar frequency in blacks and whites. CLL is uncommon in Japan, China, and other Asian countries.

ETIOLOGY

The etiology of CLL is unknown, and there are no known causative risk factors. Studies have indicated no relationship between chemical exposure, ionizing radiation, and drug intake and the development of CLL.[3-5] Families with multiple cases of CLL have been reported, and there have been several sets of affected twins, with an increased frequency of lymphoproliferative disorders in first-degree relatives of CLL patients.[6] Spouses of patients with CLL who themselves have CLL have also been reported.[6]

PATHOGENESIS

CLL is a clonal expansion of mature-appearing lymphocytes. Although the morphologies of normal and CLL B cells are similar, there are marked molecular and immunologic differences between the two cell populations. Both CLL and normal cells express B-lineage antigens (e.g., CD19 and CD20); however, CLL cells also express activation antigens (e.g., CD23) and CD5.

$CD5^+$ B cells are not unique to CLL; they can also be found in small numbers in the germinal center of normal lymph nodes and in the peripheral blood. The early transforming events are unknown; however, CLL cells are thought to be anergic B cells with a defect in apoptosis. Nevertheless, there are numerous functional, immunologic, and molecular differences between the CLL B cell and the benign $CD5^+$ B cell.[7,8]

Cytogenetics

Acquired cytogenetic abnormalities can be detected in approximately 50% of patients with CLL by use of conventional banding techniques[9]; these abnormalities are more frequent in the advanced stages of the disease. However, more sensitive assays, such as fluorescence in situ hybridization (FISH), have revealed defects in more than 80% of cases,[10] with deletions of 13q detected in more than half of cases, deletions of 11q detected in 15% to 20% of cases, and trisomy 12 detected in 15% to 20% of cases. More than a third of patients have complex abnormalities.

Molecular Biology

No single genetic aberrancy has been implicated in the pathogenesis of CLL. However, CLL cells are characterized by a defect in apoptosis, or programmed cell death. Overexpression of the *bcl-2* gene is present in more than 70% of cases, even in the absence of the t(14;18) chromosome rearrangement.[11] The ratio of the antiapoptotic gene *bcl-2* to the proapoptotic gene *BAX* is increased in CLL cells, which favors cell survival. Deletions involving 13q34 have been identified in as many as 70% of cases by conventional cytogenetics; this defect, previously thought to be at the site of the retinoblastoma suppressor gene, has since been shown to be telomeric to that region.[12] This region of loss contains a novel suppressor gene called *DBM* (disrupted in B cell malignancy). There is an apparent correlation between the antiapoptotic protein Mcl-1 and resistance to chemotherapy.[13]

DIAGNOSIS OF CHRONIC LYMPHOCYTIC LEUKEMIA

Clinical Presentation

Patients with CLL are generally asymptomatic at presentation, and the diagnosis is often made incidentally when lymphocytosis is noted at the time of a routine evaluation. The results of physical examination are normal in 20% to 30% of patients at initial presentation, and lymphadenopathy, hepatosplenomegaly, or both are present in 40% to 50% of patients. However, as the disease progresses, generalized adenopathy and splenomegaly become common features.

Infections are a frequent complication and may be recurrent, with the most common pathogens being *Staphylococcus, Streptococcus,* other bacteria that require opsonization, and the herpesvirus.

As many as 20% of patients with CLL may have a positive result on the Coombs antiglobulin test, although clinical hemolysis is apparent in only half of those patients. The frequency of immune thrombocytopenia appears to be about 2%.

Pure red cell aplasia occurs in fewer than 5% of cases of CLL. Treatment options include corticosteroids, cyclosporine with or without concurrent corticosteroids, and systemic chemotherapy for CLL.[14]

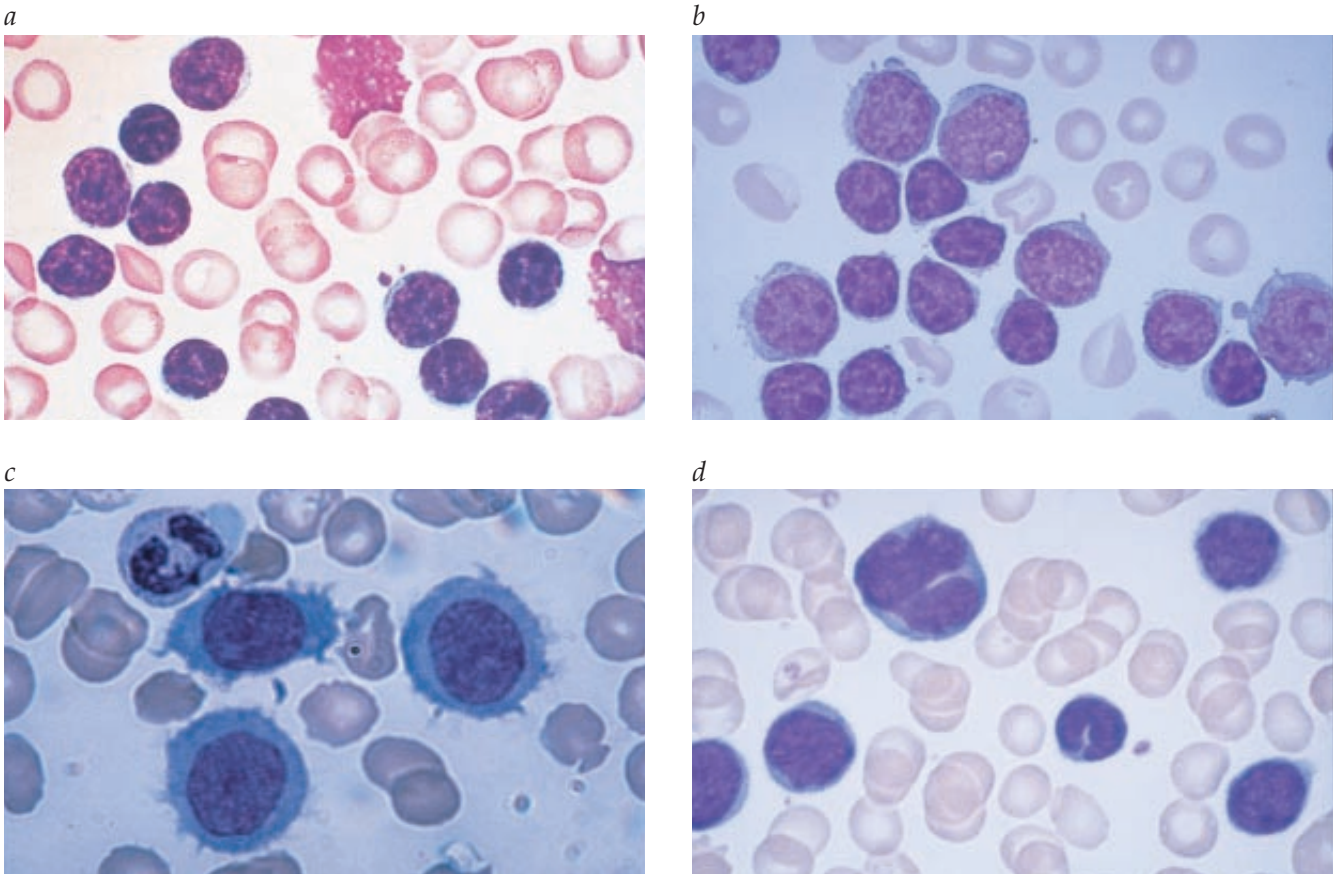


Figure 1 **Chronic lymphocytic leukemia (*a*), prolymphocytic leukemia (*b*), hairy-cell leukemia (*c*), and mantle cell lymphoma (*d*).**

Laboratory Findings

The diagnosis of CLL should be considered if there is a sustained increase in the number of small, mature-appearing lymphocytes (i.e., to > 5,000/μl) circulating in the peripheral blood and if this increase is unexplained by other medical conditions. The bone marrow aspirate and biopsy reveal infiltration by at least 30% lymphocytes. Evaluation of the bone marrow is rarely required to confirm the diagnosis of CLL, but it does provide important prognostic information and is needed to assess response to therapy.[15]

Immunophenotyping helps distinguish CLL from other chronic lymphoid leukemias, such as PLL, HCL and HCL variant, NHL in a leukemic phase (e.g., lymphoplasmacytic lymphomas; marginal-zone NHL, including splenic lymphoma with villous lymphocytes; and mantle cell NHL), and plasma cell leukemia [*see Figure 1 and Table 1*].[1,15] CLL lymphocytes are monoclonal B cells that express low levels of surface immunoglobulins and the CD5 antigen along with normal B cell antigens CD19, CD20, and CD23. Mantle cell NHL, another major $CD5^+$ B cell disorder, can be distinguished from CLL by its strong expression of surface immunoglobulin, a lack of CD23, and expression of cyclin D1.

STAGING AND PROGNOSTIC FACTORS

CLL is a clinically heterogeneous disorder; some patients live for decades without requiring treatment, whereas others die within months after diagnosis. The first staging system to be widely adopted was the five-stage Rai classification [*see Table 2*].[16] This system was subsequently simplified to three levels of risk (four stages): low risk (stage 0), intermediate risk (stages I and II), and high risk (stages III and IV). In some parts of Europe, the Binet classification is more often used[17] [*see Table 3*]. Although the Binet system fails to identify Rai stage 0 patients, the two systems nevertheless have similar predictive value. Between 60% and 70% of patients with CLL present in a low- or intermediate-risk group.

Although CLL tends to occur in patients older than 60 years, 10% to 15% of patients are younger than 50 years. Stage for stage, younger patients have a median survival comparable to that of older patients when corrected for non–CLL-related causes of death.[18]

Table 1 Chronic Lymphoid Leukemias

Leukemia	*Morphology*	*Immunophenotype*
Chronic lymphocytic leukemia	Small, compact nucleus; no visible nucleoli; homogeneous cells	sIg faint, $CD19^+$, $CD23^+$, $CD5^+$, $CD103^-$, $CD10^-$
Prolymphocytic leukemia	Larger, prominent nucleoli; mixed population	sIg^+, $CD19^+$, $CD20^+$, $CD23^{+/-}$, $CD103^-$, $CD10^-$
Hairy-cell leukemia	Filamentous projections	sIg^+, $CD19^+$, $CD20^+$, $CD23^-$, $CD5^-$, $CD103^+$, $CD10^-$
Mantle cell lymphoma	Small to medium cells; cleaved nuclei	sIg^+, $CD19^+$, $CD20^+$, $CD23^-$, $CD5^-$, $CD103^-$, $CD10^-$

sIg—surface immunoglobulin

Table 2 The Modified Rai Staging System for Chronic Lymphocytic Leukemia

Rai Stages	*Three-Stage System*	*Clinical Features*	*Median Survival (Years)*
0	Low risk	Lymphocytosis in blood and marrow only	> 10
I and II	Intermediate risk	Lymphocytosis + lymphadenopathy + splenomegaly ± hepatomegaly	7
III and IV	High risk	Lymphocytosis + anemia + thrombocytopenia	1.5

Prognostic factors include male sex, black race, poor performance status, and older age.[19] More predictive than the absolute number of circulating lymphocytes is the lymphocyte doubling time.[20] Serum β_2-microglobulin, soluble CD23,[21,22] and soluble interleukin-2 (IL-2) receptors[23] are also predictive of outcome.

Simple cytogenetic abnormalities are associated with a longer survival than complex cytogenetic abnormalities. Patients with 14q abnormalities have a particularly poor survival, whereas the outlook for those with 13q abnormalities appears to be similar to that for patients with a normal karyotype. Patients with abnormalities of 11q tend to have adenopathy that is out of proportion to the lymphocyte count; these patients have a poor prognosis.[9] Cytogenetic studies are not recommended as part of the routine evaluation of the patient with CLL, because they are expensive, are difficult to perform, and, most important, do not yet direct treatment decisions.

The ratio of bcl-2 proteins to BAX proteins in CLL cells favors bcl-2 and correlates with clinical resistance to chemotherapy and clinical outcome.[24,25] *p53* deletions are predictive of a poor response to therapy with fludarabine or pentostatin.[26]

Patients whose cells have high levels of CD38 expression and unmutated Ig genes have a significantly poorer outcome. This suggests that there may be two distinct types of CLL.[27,28]

CLINICAL COURSE AND COMPLICATIONS

Progressive Conditions

In approximately 10% to 15% of patients with CLL, a more aggressive lymphoproliferative disease develops, most commonly Richter syndrome, which is an aggressive large cell lymphoma that occurs in about 5% of patients with CLL.[29] Patients characteristically present with increasing lymphadenopathy, hepatosplenomegaly, fever, abdominal pain, weight loss, progressive anemia, and thrombocytopenia with a rapid rise in the peripheral blood lymphocyte count. Patients with Richter syndrome respond poorly to systemic therapy, and they have a short survival.

The second most common transformation is PLL. Others include acute lymphocytic leukemia, plasma cell leukemia, MM, and Hodgkin disease.

Secondary Malignancies

Secondary malignancies appear to occur with increased frequency in patients with CLL and are related both to the immune defects associated with this disease and to the consequences of therapy. These disorders primarily involve the skin, prostate, and gastrointestinal tract.[30-33]

Immune Abnormalities

A spectrum of immune defects has been reported in CLL. CLL B cells produce reduced amounts of normal immunoglobulins in response to antigenic stimuli. In addition, there are quantitative and qualitative abnormalities of normal B cells, T cells, and natural-killer cells; a reduction in the number and function of normal T cells; and impaired complement activation.[34]

THERAPY FOR CHRONIC LYMPHOCYTIC LEUKEMIA

The prognosis of patients with CLL is related to the stage of the disease. Patients with early stage disease may live a normal life span, although some may experience disease progression and die of their disease. These two groups may be distinguished by CD38 expression and mutation status of the *Ih* heavy-chain genes.[27,28] The outlook for patients with advanced disease may be improving as a result of newer treatments. Because currently available therapies are not curative, the appropriate time to initiate therapy is important. There is no survival advantage for patients with early stage disease who are treated at diagnosis. The French Cooperative Group on CLL conducted two studies of patients with Binet stage A disease. In the first trial, patients were randomized to receive either daily oral chlorambucil therapy or to undergo observation. In the second trial, patients received either intermittent chlorambucil plus prednisone or no initial treatment. Neither study detected an advantage of early intervention; moreover, in the first trial, there was a greater number of fatal, secondary solid tumors,[33] presumably related to the use of chlorambucil.

Therapy should be initiated only when indicated by one or more disease-related symptoms (e.g., fevers, chills, weight loss, or pronounced fatigue); progressive bone marrow failure with anemia, thrombocytopenia, or both; autoimmune anemia or thrombocytopenia that is unresponsive to other therapies (e.g., corticosteroids); massive or progressive hepatosplenomegaly or lymphadenopathy; or recurrent infections. Although there is no absolute number of lymphocytes that can be used as a benchmark to initiate therapy, a rapid lymphocyte doubling time (< 6 months) may support the decision to treat.[20]

Single Agents

The standard agent for CLL has been chlorambucil. When this drug is administered either at a dosage of 6 to 14 mg orally daily for 4 to 8 weeks or in higher-dose, intermittent pulses of 20 mg/m^2 orally every 3 weeks or 30 to 40 mg/m^2 once a month, disease regression in lymph nodes, liver, and spleen and normalization of blood counts occur in about half of previously un-

Table 3 The Binet Staging System for Chronic Lymphocytic Leukemia

Binet Stages	*Clinical Features*	*Median Survival (Years)*
A	Fewer than 3 areas of clinical lymphadenopathy; no anemia or thrombocytopenia	12
B	Three or more involved node areas; no anemia or thrombocytopenia	7
C	Hemoglobin ≤ 10 g/dl and/or platelets < 100,000 µl	2

treated patients. However, complete remissions are rare.[35] Cyclophosphamide appears to have activity similar to that of chlorambucil but is generally used only when chlorambucil is poorly tolerated.

Corticosteroids are less active than alkylating agents in CLL and should be reserved for patients with autoimmune complications because of the risks of bacterial, viral, and fungal infections; diabetes; and osteoporosis.

Purine Analogues

Fludarabine Fludarabine has been shown to be the most active agent in the treatment of CLL. Fludarabine was initially studied in patients who either did not respond to alkylating agents or experienced relapse during therapy with the agents.[36-39] Because of difficulties in comparing various treatment results, the National Cancer Institute–sponsored Working Group on CLL[15] standardized eligibility, response, and toxicity criteria. With the use of these guidelines, complete clinical and hematologic remissions were achieved in 3% to 13% of patients, with an overall response rate of 40% to 50%. Results varied with factors such as patient age, stage, and performance status. The median time to progression was about 18 months for patients whose conditions were refractory to alkylating agents and 17 months for patients who experienced relapse after having previously responded to therapy; the median survival was 29 months for the patients who experienced relapse and 9 months for the patients whose conditions were refractory.

When fludarabine is used as the initial treatment, the complete-response rate is about 30%, and the overall response rate is 70%.[35,40,41] Randomized trials have demonstrated higher response rates and longer time to disease progression with fludarabine than with alkylating agents and anthracycline, although a survival advantage is not yet apparent.[35,40] The median time to disease progression in untreated patients is approximately 32 months; the median survival is longer than 5 years. Fludarabine is an effective front-line option for CLL and has now replaced alkylating agents as the preferred initial therapy for most patients with CLL.[35]

When fludarabine is used in accordance with the currently recommended regimen of 25 mg/m^2 I.V. for 5 consecutive days once a month, the major toxicities encountered are moderate myelosuppression and severe immunosuppression, with occasional neurotoxicity, particularly at higher than recommended doses.[42,43] Autoimmune hemolysis and thrombocytopenia have also been reported.[44,45] These conditions generally respond to therapy with corticosteroids. If these conditions are unresponsive to corticosteroids, alternative approaches include use of high-dose intravenous immunoglobulin, splenectomy, or rituximab.[45,46] Lymphocyte counts, particularly CD4$^+$ T cell counts, decrease within weeks and do not return to normal for a year or longer after treatment has been discontinued.[41,43] The use of prophylactic antibiotics is not recommended in most patients, because it would be impossible to cover the broad spectrum of potential pathogens, the use of additional drugs is costly, and there is an increased potential for toxicity.[47]

Other agents Both cladribine and 2′-deoxycoformycin (pentostatin) are active in CLL but are less commonly used than fludarabine because of their relatively lower response rates, less durable remissions, and greater expense.[48,49]

Combination Chemotherapy

The most frequently used combination chemotherapy regimens in CLL are chlorambucil and prednisone (CP) and cyclophosphamide, vincristine, and prednisone (CVP). Reported response rates with CP and CVP in previously untreated patients with advanced-stage disease vary from less than 10% to over 60%, depending to a large extent on the criteria used for determining response. It is uncommon to achieve disappearance of evidence of disease. The median survival is less than 2 years.[50,51]

Combination therapy is not clearly superior to single-agent therapy. Early studies, including studies involving small numbers of cases, suggest that prednisone may increase the response rate to chlorambucil but does not prolong survival.[50] More intensive regimens, such as CHOP (cyclophosphamide, doxorubucin [hydroxydaunomycin], vincristine [Oncovin], and prednisone), are not clearly better than less intensive ones.[51-53] Combinations of fludarabine and cyclophosphamide are currently being compared with fludarabine alone.[54,55]

TREATMENT OF RELAPSED/REFRACTORY DISEASE

Standard Approaches

Patients with an asymptomatic recurrence are generally not treated until therapy is indicated by use of the same criteria as for front-line therapy, with palliative intent. Fludarabine is the standard salvage agent for patients who have not responded to alkylating agents. Patients whose conditions are refractory to fludarabine have a low likelihood of responding to alkylating agents.

High-Dose Therapy with Stem Cell Support

The experience with allogeneic bone marrow transplantation (BMT) in CLL is limited, in part because the median age of patients at the time of diagnosis is over 60 years.[56-58] About half of the patients who undergo BMT may remain disease free for prolonged periods; however, it is unclear whether they are cured. Unfortunately, the treatment-related mortality is 25% to 50%. Submyeloablative transplantation appears to be effective in CLL; when compared with standard allogeneic BMT, it is associated with a reduction in the rate of acute graft versus host disease (GVHD), but it is associated with significant chronic GVHD. Longer follow-up is required to determine whether there is durable clinical benefit.[59,60]

Autologous stem cell transplantation has a questionable role for patients with CLL.[58,61]

Biologic Therapy

Responses to interferon alfa tend to be brief, and interferon alfa has no influence on outcome. The CAMPATH anti-CD52 monoclonal antibody has recently been approved for use in relapsed and refractory CLL, based on a 33% response rate and acceptable toxicity.[62-64] The anti-CD20 antibody C2B8 (rituximab), which has considerable activity in follicular lymphomas, does not appear to have major activity when used as monotherapy at standard doses; however, preliminary data suggest higher response rates when rituximab is used in more intensive regimens or when it is combined with fludarabine.[65-70] These approaches are being evaluated in clinical trials.

Supportive Care

Myeloid growth factors Whether myeloid growth factors can protect against chemotherapy-induced myelosuppression has not been demonstrated, and their use is not part of standard practice.

Intravenous immunoglobulins The majority of patients with CLL experience hypogammaglobulinemia, which is more common in the advanced stages of CLL and which worsens as the disease progresses. Historically, the most common pathogens in CLL have been those that require opsonization before being killed. Unfortunately, high-dose intravenous immunoglobulins fail to protect patients from severe bacterial, viral, or fungal infections. They have no influence on overall survival and are not cost-effective.[71,72] This therapy should be reserved for patients with documented, repeated bacterial infections.

Erythropoietin Anemia in CLL is not the consequence of deficient production of erythropoietin but, more commonly, is the consequence of bone marrow replacement, splenic sequestration, or hemolysis. Nevertheless, erythropoietin may reduce transfusion requirements in selected patients, and a trial of this agent can be considered in anemic CLL patients who have no other obvious correctable cause of anemia.[73]

Splenectomy

Splenectomy may play an important role in the palliative treatment of patients with CLL who have not benefited from systemic treatment.[46] Thrombocytopenia responds in varying degrees to splenectomy in almost all patients. When this procedure is performed by an experienced surgeon, related mortality is under 10%.

Prolymphocytic Leukemia

Most cases of PLL arise de novo; nevertheless, PLL represents the most common form of aggressive transformation in patients with CLL. Patients may present with clinical deterioration, often with a wastinglike disorder. Compared with patients whose PLL has evolved from CLL, those with de novo PLL tend to be older (in their 70s and beyond) than those with transformed disease (in their 60s); have a more marked lymphocytosis; have massive, rather than moderate, splenomegaly; and generally do not have prominent lymphadenopathy. The peripheral blood smear often reveals a dual population of cells, some being characteristic small lymphocytes of CLL and others being larger cells with prominent nucleoli. The phenotype of de novo PLL differs from that of CLL in that de novo PLL has stronger CD79b and weaker CD5. In patients with primary PLL, the t(6;12) chromosome rearrangement has been reported. *p53* abnormalities, which occur in more than 50% of patients, may result in resistance to chemotherapy.

Promising preliminary results have been published regarding the use of purine analogues[74] and the CAMPATH monoclonal antibody[75] for the treatment of PLL.

Hairy-Cell Leukemia

HCL is an uncommon B cell malignancy that occurs in about 500 new patients each year in the United States. The disease most often occurs in older persons, with a high male-to-female ratio.

DIAGNOSIS

Clinical Presentation

Patients with HCL are occasionally asymptomatic at presentation, but more often, they present with symptoms of impaired bone marrow production, such as infections, bleeding, and weakness. In rare cases, patients experience a vasculitis-like disorder. The physical examination is notable for splenomegaly that is often massive and accompanied by hepatomegaly; less commonly, lymphadenopathy occurs with splenomegaly.

Laboratory Findings

Laboratory evaluation reveals pancytopenia, which raises the possibility of aplastic anemia, acute leukemia, myelodysplastic syndrome, or HCL. A bone marrow biopsy may be required to make the diagnosis, because the aspirate is often packed with hairy cells; fibrosis is also present. The malignant cells are of B cell origin, expressing CD19, CD20, and the monocyte antigen CD11c. Perhaps the most specific marker is Bly-7 (CD103).

TREATMENT

HCL is generally an indolent disorder, and 10% of patients may never require treatment. Therapy is indicated if there is massive or progressive splenomegaly, worsening blood counts, recurrent infections, more than 20,000 hairy cells/μl of peripheral blood, or bulky lymphadenopathy. Until about 1983, splenectomy was the standard approach to HCL because it improves symptoms related to splenomegaly and, in the majority of cases, improves the peripheral blood counts; however, this procedure does not affect the disease itself. Splenectomy is now less often performed because of the availability of effective systemic therapies. Interferon alfa was the first agent to achieve meaningful responses in patients with HCL. Most patients experience clinical and hematologic improvement; however, in only 10% of these patients is the response complete, and the disease invariably recurs after therapy is discontinued.

The purine analogues have revolutionized the treatment of patients with HCL. In a North American trial, previously untreated patients with HCL were randomized to receive either interferon alfa or pentostatin; the complete response rate was approximately 11% for patients given interferon alfa, compared with 76% for patients given pentostatin, with a significant advantage to pentostatin in the durability of response.[76] Cladribine induces responses in 80% to more than 90% of patients, with 65% to 85% having complete remissions.[77] Only 25% to 35% of patients experience relapse after prolonged follow-up. In many cases, relapse is characterized only by an increase in bone marrow hairy cells, with no indication for treatment. Most patients who experience relapse and again require treatment achieve a second durable response. Results with pentostatin and cladribine are equivalent; the choice of which agent to use is based on physician preference. Rituximab and an anti-CD22 immunotoxin have been reported to be active in cases involving nucleoside analogue failures.[78,79]

Multiple Myeloma

EPIDEMIOLOGY

MM is uncommon in persons younger than 40 years; the incidence increases rapidly after age 50. Men and women are equally affected. Although blacks are affected more commonly than whites, the outcomes in the two groups are similar, stage for stage.[80] Approximately 14,000 new cases were projected for 2001 in the United States.[2] No apparent correlation has been identified between MM and socioeconomic status. The incidence of this disease is apparently lower in Chinese and other Asians. The etiology of MM is

unclear. Exposure to ionizing radiation, paints and solvents, and pesticides has been implicated but never firmly established.

PATHOGENESIS

MM is characterized by an accumulation of plasma cells in the bone marrow and, less often, soft tissues or visceral organs. The malignant cell is not the mature plasma cell but a cell that occurs earlier in B cell development, after the rearrangement of the heavy-chain genes. These cells express a distinct variety of adhesion molecules (e.g., CD11b),[81] which may permit the circulating B cells to attach to nonmalignant stromal cells in the bone marrow that produce cytokines capable of supporting myeloma cell growth, such as IL-6. IL-6 acts as both an autocrine and a paracrine growth factor in MM, thereby both stimulating cell growth and preventing apoptosis. A role in pathogenesis has been suggested for a herpesvirus (Kaposi sarcoma–associated herpesvirus or herpes simplex virus–8).[82]

Cytogenetic analyses have been difficult to perform in MM. However, recent studies using FISH have demonstrated aneuploidy in about 70% of cases.[83] The most frequent abnormalities involve chromosomes 13 ($13q^-$) and 14 ($14q^+$). These defects may result in impaired apoptosis and subsequent resistance to therapy. Mutations of *p53* have been identified in patients who experience MM relapse; mutations appear to correlate with poor outcome.

IMMUNOLOGY

MM is characterized by the overproduction of a monoclonal immunoglobulin. The frequency of the various immunoglobulin types of MM parallels the serum concentrations of those Igs: IgG in 60% to 70%, IgA in 20%, and light chain in only 15%; few cases of IgD and IgE have been reported, and approximately 1% of patients are nonsecretors. A secondary finding is depressed levels of normal residual immunoglobulins.

Bence Jones proteins (urinary light chains) are detected in the urine with antisera against light chains but without an anti–heavy-chain component.

DIAGNOSIS

The diagnosis of MM is often made incidentally when an increased serum protein concentration is detected on routine evaluation or in the evaluation of anemia, renal dysfunction, or asymptomatic hypercalcemia.

Monoclonal Gammopathy of Undetermined Significance

Monoclonal gammopathy of undetermined significance (MGUS) occurs in about 1% of the general population and 3% of otherwise healthy persons older than 70 years; IgG is most commonly increased. Features of MM that distinguish it from MGUS include greater concentrations of serum and urine M proteins, increased percentage of bone marrow plasma cells, a higher plasma cell labeling index, the presence of lytic bone lesions, and the presence of circulating plasma cells in the peripheral blood. After prolonged follow-up, four groups of patients with MGUS can be identified: (1) those with a contained monoclonal gammopathy (19%); (2) those whose serum immunoglobulin level increases to more than 3 g/dl but who do not require any therapy (10%); (3) those who die without developing MM or another related disease (47%); and (4) those in whom MM, amyloidosis, macroglobulinemia, or another lymphoproliferative disorder develops (24%, two thirds of which are MM). Progression to a malignancy occurs at 8 to 10 years. Unfortunately, there are no laboratory studies that reliably predict which patients with MGUS will progress to MM.

Table 4 Sites of Involvement with Plasma Cell Dyscrasias

Sites	*Disorders*
Skull	Lytic lesions
Soft tissues	Plasmacytoma
Bone marrow	Anemia
	Thrombocytopenia
	Neutropenia
Nerves	Peripheral neuropathy
	Spinal cord compression
Tongue	Amyloid
Heart	Amyloid
Bone	Compression fractures
	Osteopenia
	Lytic lesions
Kidneys	Proteinuria
	Amyloid
	Renal failure
Long bones	Fractures

Clinical Presentation

The diagnosis of MM requires at least one of the following major criteria: plasmacytoma; bone marrow with at least 30% plasma cells; and M protein on serum electrophoresis showing levels of IgG greater than 3.5 g/dl, IgA greater than 2.0 g/dl, or urine light-chain excretion greater than 1 g/24 hr. One or more of the following minor criteria are also required: bone marrow plasmacytosis greater than 10%; M protein in serum or urine but less than the levels listed above under major criteria; lytic bone lesions; and decreased residual immunoglobulins.

MM can affect many organs, including the bones, kidneys, nerves, heart, and liver [*see Table 4*].

Plasmacytomas Plasmacytomas are solid tumors consisting of plasma cells that may occur in soft tissues (extramedullary plasmacytomas); they may present as palpable masses but more commonly present as spinal cord compression. These tumors may be cured with local radiation therapy. Patients with plasmacytomas that occur in bone may initially respond to radiation therapy, but half will experience progression to overt MM, two thirds within 3 years after diagnosis. Persistence of the presence of M protein after radiation therapy is associated with an increased likelihood of evolution to MM.

Bone disease Bone disease is one of the most common and diagnostic features of MM. Although lytic bone lesions are most typical [*see Figure 2*], osteopenia is the most common form of bone disease in myeloma. Osteosclerotic disease has also been reported. Increased osteoclast activity is notable in the affected bones and is believed to be related to so-called osteoclast activating factor, which is a manifestation of the effects of tumor necrosis factor, IL-1, and IL-6.

Radiation therapy for painful bone lesions in a patient with newly diagnosed MM should be deferred if possible until the effectiveness of systemic therapy can be assessed, because radiation therapy may impair the ability to administer adequate doses of chemotherapy. Exceptions include spinal cord compression and potential fractures in weight-bearing bones that require im-

mediate intervention. For patients in relapse, local radiation therapy can provide substantial palliative benefit.

Vertebral compression fractures during the course of systemic therapy should not be the sole reason for use of radiation therapy, because these lesions are often the result of osteopenia; they do not necessarily represent disease progression and are generally not responsive to radiation.

Pamidronate is a second-generation bisphosphonate that has been studied in a placebo-controlled clinical trial in previously untreated patients and patients in relapse with stage III MM.[84] The drug provided significant protection against skeletal complications (measured by time to first skeletal event), appeared to improve quality of life, and possibly provides a survival advantage.[85] As a result, the current recommendation is to administer bisphosphonates to all patients with evidence of bone involvement, although some clinicians recommend this agent for all patients with MM.

Anemia Anemia occurs in most patients with MM, especially as the disease progresses. Anemia is a consequence of bone marrow replacement, renal failure, and chronic disease. In patients with MM who have inappropriately low serum levels of erythropoietin, erythropoietin may reduce red cell transfusion requirements.[86]

Hypercalcemia About 10% to 20% of patients with MM present with hypercalcemia. Serum calcium levels, corrected for albumin, can be used to identify patients at risk for symptoms related to hypercalcemia. Paraproteins may occasionally bind calcium, resulting in a falsely high calcium level; therefore, a free (ionized) calcium determination should be obtained in the asymptomatic patient with an elevated serum calcium level.

Previously untreated patients who either are asymptomatic or have only mild symptoms related to their mild or moderate hypercalcemia should be treated with systemic chemotherapy, including corticosteroids. Patients with high serum calcium levels or symptoms, such as lethargy, confusion, and constipation, should receive hydration with saline diuresis. Bisphosphonates are useful as a means of reducing the serum calcium concentration, as are mithramycin and calcitonin.

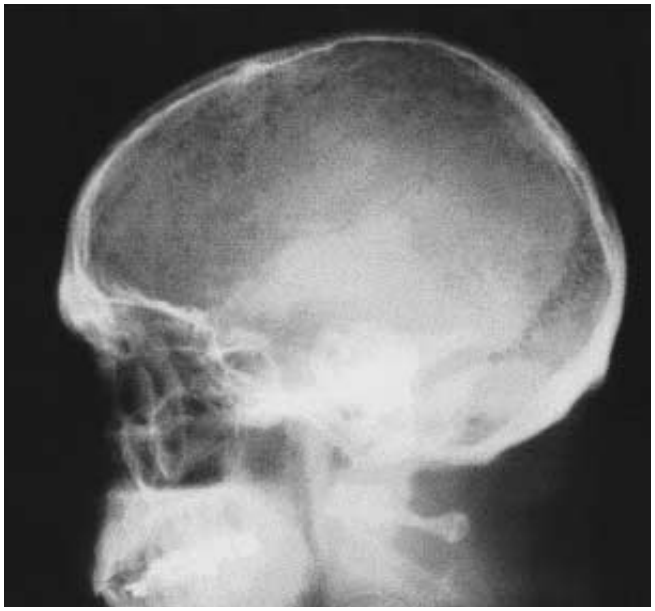


Figure 2 **Radiograph showing lytic bone lesions of the skull.**

Renal insufficiency Renal insufficiency is common in MM and may result from a variety of mechanisms. The tubules may become obstructed by myeloma protein (myeloma kidney), or the kidneys may be infiltrated by amyloid protein. The increased risk of infections creates a predisposition to pyelonephritis. Other causes include hypercalcemia, urate nephropathy, nephrotic syndrome, hyperviscosity, infiltration of the kidneys by plasma cells, and obstructive uropathy. Patients with suspected MM and abnormal renal function should be kept well hydrated and should not be subjected to dehydrating procedures such as a computed tomographic scan with contrast or an intravenous pyelogram, which can precipitate acute renal failure in the patient with occult renal impairment. Plasmapheresis may alleviate acute renal failure. Dialysis and even renal transplantation have been used successfully.

Hyperviscosity Signs and symptoms of hyperviscosity, such as shortness of breath, confusion, and chest pain, are uncommon in patients with IgG MM but may occur in patients with extremely high serum protein concentrations. Hyperviscosity occurs more often in patients with IgA MM because of greater polymerization of the paraprotein. Management should be directed at treating the underlying MM. Plasmapheresis may be considered if a rapid decrease in the paraprotein is required to alleviate symptoms.

Infectious diseases Patients with myeloma are at increased risk for life-threatening bacterial infections, even in the presence of normal numbers of circulating granulocytes. This abnormality is related to impaired synthesis of normal immunoglobulins and defective complement activation. Early in the course of their disease, patients are more likely to become infected with encapsulated organisms that require opsonization (e.g., *Streptococcus pneumoniae, Haemophilus influenzae,* and *Staphylococcus*), whereas the use of chemotherapy increases the likelihood of gram-negative infections.

Polyvalent pneumococcal vaccine is unlikely to be effective, because patients respond poorly to bacterial antigens, and other bactericidal defects are not corrected by vaccination. Neither intravenous immunoglobulins nor prophylactic antibiotics are cost-effective or uniformly protective and should be reserved for patients with documented, recurrent bacterial infections.

Neurologic symptoms Epidermal cord compression is a common presentation in patients with plasmacytoma; it is an oncologic emergency requiring decompression and radiation therapy. Paraproteinemias may be associated with other neurologic symptoms, such as POEMS syndrome. POEMS is the acronym for a collection of findings that include polyneuropathy (generally sensorimotor), organomegaly (e.g., hepatomegaly), endocrinopathy (e.g., diabetes, amenorrhea, and gynecomastia), M protein, and skin changes (e.g., hyperpigmentation). Other associated features include lymphadenopathy and osteosclerotic bone lesions.

STAGING AND PROGNOSTIC FACTORS

Once the diagnosis of MM has been made, the initial evaluation should include a complete blood count, as well as a bone

marrow aspiration and biopsy. Serum electrophoresis and quantitative immunoglobulins determine the concentration of paraprotein, which correlates with tumor mass, and an immunoelectrophoresis defines the immunoglobulin subtype. A urine electrophoresis will detect urinary light chains, if present. A radiographic bone survey is performed to identify sites of involvement and potential fractures. Radionuclide bone scans are of little value in assessing MM.

The clinical pictures of the subtypes of MM vary somewhat. For example, patients who secrete only light chains often present with progressive renal failure, initially without overt myeloma. Patients with IgD myeloma have a more aggressive clinical course characterized by a high frequency of renal failure, plasmacytomas, and amyloidosis and have a shorter survival than patients with IgG MM despite having low paraprotein concentrations.[87] IgE myeloma, which is rare, appears to be accompanied by an increased frequency of plasma cell leukemia.

Patients have most often been staged according to the classification of Durie and Salmon[88] [*see Table 5*], which correlates well with tumor mass and prognosis. Stage is determined by the level of M protein, the number of lytic bone lesions, hemoglobin concentration, and serum calcium (corrected for albumin). Patients are further subdivided into classes A and B on the basis of the serum creatinine level.

The median survival is approximately 5 years for patients with stage IA disease but 15 months for patients with IIIA MM. However, this staging system is not uniformly reliable in predicting outcome and is being replaced by more biologically relevant markers, notably the β_2-microglobulin, plasma cell labeling index, and cytogenetic abnormalities, as well as lactate dehydrogenase, Ki-67, serum levels of IL-6, C-reactive protein, and the presence of circulating plasma cells.[83,89-93] Cytogenetics are not part of standard evaluation, because such testing is expensive and does not influence therapy. Magnetic resonance imaging may prove to be valuable in the evaluation of bone marrow involvement and as a prognostic indicator.[94]

THERAPY FOR MULTIPLE MYELOMA

Indications for Treatment

A small proportion of patients are classified as having smoldering MM on the basis of the following features: M protein in serum less than 3.0 g/dl, less than 10% plasma cell infiltration in the bone marrow, no anemia, normal serum calcium and creatinine levels, no lytic bone lesions, and a low plasma cell labeling index. These patients may remain stable for several years without treatment.

Treatment of stage I and II MM should be considered for patients with progressive disease, which includes an increasing M protein level in the serum, urine, or both; decreasing serum hemoglobin level; progressive bone marrow failure with neutropenia, thrombocytopenia, or both; increasing serum calcium, creatinine, or uric acid concentrations; multiple lytic bone lesions; bone pain; recurrent infections; and extramedullary plasmacytoma. Those patients with stage III disease are generally symptomatic or have other hematologic or chemical abnormalities at presentation that warrant therapy. There are no known prophylactic therapies that can reduce the risk of developing MM.

Patients with MGUS require regular monitoring of serum immunoglobulin levels.

Table 5 Staging System for Multiple Myeloma

Stage I	
All of the following:	
Hemoglobin	> 10 g/dl
Serum calcium	< 12 mg/dl
Normal bone structure or solitary plasmacytoma on radiographs	
Low M component	
IgG	< 5 g/dl
IgA	< 3 g/dl
Urine light chains	< 4 g/24 hr
Stage II	
Fitting neither stage I nor stage III	
Stage III	
One or more of the following:	
Hemoglobin	< 8.5 g/dl
Serum calcium	>12 mg/dl
Advanced lytic bone lesions	
Hyper M component	
IgG	> 7 g/dl
IgA	> 5 g/dl
Urinary light-chain excretion	> 12 g/24 hr
Subclassification	
A: serum creatinine	< 2.0 mg/dl
B: serum creatinine	≥ 2.0 mg/dl

Standard Therapy

Once it has been decided that treatment is indicated, there are limited effective therapeutic options. Agents for use as monotherapy primarily include corticosteroids, alkylating agents, and interferon alfa. Melphalan with prednisone (MP) is most often used because of higher response rates when the two drugs are administered together. MP is generally administered every 3 to 4 weeks for 6 to 9 months, with treatment subsequently continued if a response is ongoing. Most often, the paraprotein concentration decreases to a stable value (plateau phase); continuing therapy beyond this point does not prolong survival but increases complications (e.g., infections). Other alkylating agents are also active but have no advantage over melphalan. Additional drugs used to treat MM patients, such as anthracyclines, have limited effectiveness when used as single agents.

Corticosteroids (either prednisone or dexamethasone) may decrease the M protein in as many as 40% of previously untreated patients and in 20% to 40% of patients who have not responded to primary treatment. Steroids also offer a treatment option for elderly patients and for those with impaired performance status.

Clinical response often results in a reduction in bone pain and normalization of the serum creatinine, calcium, and hemoglobin levels. The median survival of patients with MM who undergo treatment with melphalan and prednisone is 2.5 to 3 years; only 5% to 10% of patients survive 10 or more years.

Combination chemotherapy regimens (e.g., vincristine, carmustine, melphalan, cyclophosphamide, and prednisone; vincristine, carmustine, doxorubicin, and prednisone; or vincristine, cyclophosphamide, doxorubicin, and prednisone) have not been shown to be more efficacious than MP. The VAD (vincristine, doxorubicin [Adriamycin], and dexamethasone) regimen as initial therapy is less toxic to the stem cell population than MP and may be preferable for patients who are being con-

sidered for autologous stem cell transplantation. Interferon alfa has been the most widely studied of the biologic agents. Interferon alfa has activity as a single agent, but whether it adds to the effect of chemotherapy during induction or as a maintenance therapy is controversial. Most published trials and meta-analyses have failed to demonstrate a role for interferon alfa. Considering the associated toxicities and expense of the agent and its limited clinical benefit, interferon alfa cannot be recommended as a standard component of MM therapy. Use of corticosteroids as maintenance therapy after chemotherapy may prolong survival.

Relapsed/Refractory Disease

Relapse after a response to initial treatment of MM is universal. Further response to MP is less frequent and of shorter duration than after initial treatment, and therapy may be associated with more myelosuppression. The standard salvage regimen is VAD—a 4-day continuous infusion of vincristine and doxorubicin through an indwelling catheter and oral dexamethasone for 4 days, repeated on days 9 and 17 of each 21- to 28-day cycle. Responses are achieved in a third of resistant patients and in almost half of relapsed patients.

Stem cell transplantation Allogeneic BMT has not been widely used in MM, because patients tend to be elderly. BMT induces complete remissions in 40% to 50% of patients, and 25% to 30% remain alive after approximately 7 years. However, relapses may occur after several years, raising the question of whether any patients are actually cured. Because of the high rate of treatment-related morbidity and mortality, BMT is generally reserved for younger patients (< 50 years) with unresponsive or progressive disease. Submyeloablative transplantation is an encouraging new approach to the treatment of patients with MM.

Initial studies of high-dose chemotherapy and autologous stem cell transplantation were conducted in relapsed and refractory cases of MM. The frequency of response and the duration of response can be predicted by the serum lactate dehydrogenase level, β_2-microglobulin level, performance status, age (> 50 years), and responsiveness to previous therapy. For good-risk patients (i.e., those with no poor-risk factors or just one), the complete response rate is 20%, with a median relapse-free survival of 1.5 years and an overall survival of 3.5 years. With two or three poor-risk factors, the results are substantially worse, with a median relapse-free survival of 7 months and an overall survival of 20 months. With four or more poor-risk factors, the median duration of response is 3 months, with an overall survival of 5 months.

Recently, patients received this treatment earlier in the course of their disease. In a French study,[95] previously untreated patients younger than 65 years received either standard-dose chemotherapy or high-dose chemotherapy with autologous stem cell transplantation. There was an 81% response rate in the high-dose arm of the study, including 22% complete remissions and 16% very good partial remissions, compared with 57%, 5%, and 9%, respectively, in the standard-dose arm. These differences were highly significant. However, the 6-year event-free survival in the transplant arm was only 24%. A large North American randomized trial has been completed, and when the results are available, they should better characterize the role of high-dose chemotherapy with autologous or allogeneic stem cell support.

Thalidomide has emerged as an active agent for the treatment of MM, with responses in more than 30% of heavily pretreated patients. Its role in this disease is being more clearly defined.[96]

Waldenström Macroglobulinemia

Waldenström macroglobulinemia (WM) was formerly considered to be related to MM because of the presence of a paraproteinemia and infiltration of the bone marrow and other organs by plasmacytoid lymphocytes. However, several clinical features of WM are more similar to those of indolent non-Hodgkin lymphoma. In the World Health Organization Classification, WM is appropriately classified as a lymphoplasmacytic lymphoma.[97] About 1,500 persons are diagnosed with WM each year in the United States, making WM only 10% to 20% as common as MM. The median age at diagnosis is 65 years. The median survival from the time of diagnosis is about 5 years.

The diagnosis of WM is made on the basis of the presence of an IgM paraprotein and depressed levels of uninvolved immunoglobulins (i.e., IgG and IgA), with or without lymphadenopathy or hepatosplenomegaly. The bone marrow is generally infiltrated by at least 30% lymphoplasmacytoid lymphocytes. Clinical and laboratory features of WM include hyperviscosity, anemia, polyneuropathy, amyloidosis, renal abnormalities, skin deposits, and gastrointestinal involvement that leads to malabsorption and diarrhea.

Treatment of WM is directed at the complications related to the presence of either IgM or non-Hodgkin lymphoma. If patients are not symptomatic from the increase in IgM, a period of careful observation may be appropriate. If the IgM results in symptoms, treatment may involve plasmapheresis for hyperviscosity, transfusions for anemia, or treatment of hemolysis with steroids or rituximab. The lymphoma is managed as any other case of indolent non-Hodgkin lymphoma, with an alkylating agent, a purine analogue (e.g., fludarabine or cladribine), or a biologic therapy such as rituximab. Thalidomide has limited activity against this disorder.

Amyloidosis

Amyloidosis is characterized by the deposition of amyloid fibrils in various tissues.[98] Amyloidosis may occur either secondary to a chronic inflammatory condition (amyloid A, or AA) or as a primary condition (amyloid light chain–related, or AL) in the setting of a B cell malignancy. In AL, the fibrils are made up of fragments of light chains of immunoglobulins, with a clonal proliferation of bone marrow plasma cells, and Bence Jones proteinuria is present. The incidence of primary amyloidosis in the United States is 5.1 to 12.8 cases per million person-years. This disease primarily involves the heart in 90% of patients and also affects the liver, spleen, kidneys, tongue, and nervous system. The diagnosis is made by a biopsy of the affected organ or a urine sample stained with Congo red.

The treatment of amyloid AL is directed at both the affected organ and the underlying disease. Few drugs are active against amyloidosis. A randomized trial in which MP was compared with colchicine showed longer survival in patients who received the MP regimen.[99] Median survival is 18 months. High-dose therapy with stem cell support has no clear benefit.[100]

The author participates in the speakers' bureaus for Berlex Laboratories, Genentech, Inc., and IDEC Pharmaceuticals, Inc.

References

1. Harris NL, Jaffe ES, Diebold J, et al: World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the clinical advisory committee meeting — Airlie House, Virginia, November 1997. J Clin Oncol 17:3835, 1999

2. Greenlee RT, Hill-Harmon MB, Murray T, et al: Cancer statistics, 2001. CA Cancer J Clin 51:15, 2001

3. Malone KE, Koepsell TD, Daling JR, et al: Chronic lymphocytic leukemia in relation to chemical exposures. Am J Epidemiol 130:1152, 1989

4. Zahm SH, Weisenburger DD, Babbitt PA, et al: Use of hair coloring products and the risk of lymphoma, multiple myeloma, and chronic lymphocytic leukemia. Am J Public Health 82:990, 1992

5. Inskip PD, Kleinerman RA, Stovall M, et al: Leukemia, lymphoma, and multiple myeloma after pelvic radiotherapy for benign disease. Radiat Res 135:108, 1993

6. Cuttner J: Increased incidence of hematologic malignancies in first-degree relatives of patients with chronic lymphocytic leukemia. Cancer Invest 10:103, 1992

7. Caligaris-Cappio F, Hamblin TJ: B-cell chronic lymphocytic leukemia: a bird of a different feather. J Clin Oncol 17:399, 1999

8. Johnson TA, Rassenti LZ, Kipps TJ: Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol 158:235, 1997

9. Döhner H, Stilgenbauer S, Benner A, et al: Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med 343:1910, 2000

10. Stilgenbauer S, Döhner K, Bentz M, et al: Molecular cytogenetic analysis of B-cell chronic lymphocytic leukemia. Ann Hematol 76:101, 1998

11. Hanada M, Delia D, Aiello A, et al: *bcl-2* gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. Blood 82:1820, 1993

12. Bullrich F, Veronese ML, Kitada S, et al: Minimal region of loss at 13q14 in B-cell chronic lymphocytic leukemia. Blood 88:3109, 1996

13. Kitada S, Andersen J, Akar S, et al: Expression of apoptosis-regulating proteins in chronic lymphocytic leukemia: correlations with in vitro and in vivo chemoresponses. Blood 91:3379, 1998

14. Chikkappa G, Pasquale D, Zarrabi MH, et al: Cyclosporine and prednisone therapy for pure red cell aplasia in patients with chronic lymphocytic leukemia. Am J Hematol 41:5, 1992

15. Cheson BD, Bennett JM, Grever M, et al: National Cancer Institute–sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. Blood 87:4990, 1996

16. Rai KR, Sawitsky A, Cronkite EP, et al: Clinical staging of chronic lymphocytic leukemia. Blood 46:219, 1975

17. Binet JL, Auquier A, Dighiero G, et al: A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. Cancer 48:198, 1981

18. Montserrat E, Gomis F, Vallespí T, et al: Presenting features and prognosis of chronic lymphocytic leukemia in younger adults. Blood 78:1545, 1991

19. Catovsky D, Fooks J, Richards S: Prognostic factors in chronic lymphocytic leukaemia: the importance of age, sex and response to treatment in survival: a report from the MRC CLL 1 trial. Br J Haematol 72:141, 1989

20. Montserrat E, Sanchez-Bisono J, Viñolas N, et al: Lymphocyte doubling time in chronic lymphocytic leukaemia: analysis of its prognostic significance. Br J Haematol 62:567, 1986

21. Sarfati M, Chevret S, Chastang C, et al: Prognostic importance of serum soluble CD23 level in chronic lymphocytic leukemia. Blood 88:4259, 1996

22. Molica S, Levato D, Dell'Olio M, et al: Cellular expression and serum circulating levels of CD23 in B-cell chronic lymphocytic leukemia: implications for prognosis. Haematologica 81:428, 1996

23. Semenzato G, Foa R, Agostini C, et al: High serum levels of soluble interleukin 2 receptor in patients with B chronic lymphocytic leukemia. Blood 70:396, 1987

24. Robertson LE, Plunkett W, McConnell K, et al: *bcl-2* expression in chronic lymphocytic leukemia and its correlation with the induction of apoptosis and clinical outcome. Leukemia 10:456, 1996

25. Pepper C, Hoy T, Bentley DP: *Bcl-2/Bax* ratios in chronic lymphocytic leukaemia and their correlation with in vitro apoptosis and clinical resistance. Br J Cancer 76:935, 1997

26. Döhner H, Fischer K, Bentz M, et al: *p53* gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood 85:1580, 1995

27. Damle RN, Wasil T, Fais F, et al: Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94:1840, 1999

28. Hamblin T, Davis Z, Gardiner A, et al: Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94:1848, 1999

29. Giles FJ, O'Brien SM, Keating MJ: Chronic lymphocytic leukemia in (Richter's) transformation. Semin Oncol 25:117, 1998

30. Molica S, Alberti A: Second neoplasms in chronic lymphocytic leukemia: analysis of incidence as a function of the length of follow-up. Haematologica 74:481, 1989

31. Travis LB, Curtis RE, Hankey BF, et al: Second cancers in patients with chronic lymphocytic leukemia. J Natl Cancer Inst 84:1422, 1992

32. Cheson BD, Vena D, Barrett J, et al: Second malignancies as a consequence of nucleoside analog therapy of chronic lymphoid leukemias. J Clin Oncol 17:2454, 1999

33. Dighiero G, Maloum K, Desablens B, et al: Chlorambucil in indolent chronic lymphocytic leukemia. French Cooperative Group on Chronic Lymphocytic Leukemia. N Engl J Med 338:1506, 1998

34. Heath ME, Cheson BD: Defective complement activity in chronic lymphocytic leukemia. Am J Hematol 19:63, 1985

35. Rai KR, Peterson BL, Kolitz J, et al: Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia. N Engl J Med 343:1750, 2000

36. Grever MR, Kopecky KJ, Coltman CA, et al: Fludarabine monophosphate: a potentially useful agent in chronic lymphocytic leukemia. Nouv Rev Fr Hematol 30:457, 1988

37. Keating MJ, Kantarjian H, Talpaz M, et al: Fludarabine: a new agent with major activity against chronic lymphocytic leukemia. Blood 74:19, 1989

38. O'Brien S, Kantarjian H, Beran M, et al: Results of fludarabine and prednisone therapy in 264 patients with chronic lymphocytic leukemia with multivariate analysis-derived prognostic model for response to treatment. Blood 82:1695, 1993

39. Sorensen JM, Vena D, Fallavollita A, et al: Treatment of refractory chronic lymphocytic leukemia with fludarabine phosphate via the group C mechanism of the National Cancer Institute: 5-year follow-up report. J Clin Oncol 15:458, 1997

40. Multicentre prospective randomised trial of fludarabine versus cyclophosphamide, doxorubicin, and prednisone (CAP) for treatment of advanced-stage chronic lymphocytic leukaemia. The French Cooperative Group on CLL. Lancet 347:1432, 1996

41. Keating MJ, O'Brien S, Lerner S, et al: Long-term follow-up of patients with chronic lymphocytic leukemia (CLL) receiving fludarabine regimens as initial therapy. Blood 92:1165,1998

42. Cheson BD, Vena D, Foss F, et al: Neurotoxicity of purine analogs: a review. J Clin Oncol 12:2216, 1994

43. Cheson BD: Immunologic and immunosuppressive complications of purine analogue therapy. J Clin Oncol 13:2431, 1995

44. Weiss RB, Freiman J, Kweder SL, et al: Hemolytic anemia after fludarabine therapy for chronic lymphocytic leukemia. J Clin Oncol 16:1885, 1998

45. Hegde UP, Wilson WH, White T, et al: Rituximab for the treatment of fludarabine (FAMP)-associated autoimmune thrombocytopenia in chronic lymphocytic leukemia (CLL): report of three cases from a single institution. Proc ASCO (in press)

46. Seymour JF, Cusack JD, Lerner SA, et al: Case/control study of the role of splenectomy in chronic lymphocytic leukemia. J Clin Oncol 15:52, 1997

47. Anaissie EJ, Kontoyiannis DP, O'Brien S, et al: Infections in patients with chronic lymphocytic leukemia treated with fludarabine. Ann Intern Med 129:559, 1998

48. Saven A, Carrera CJ, Carson DA, et al: 2-Chlorodeoxyadenosine treatment of refractory chronic lymphocytic leukemia. Leuk Lymphoma 5:133, 1991

49. Dillman RO, Mick R, McIntyre OR: Pentostatin in chronic lymphocytic leukemia: a phase II trial of Cancer and Leukemia Group B. J Clin Oncol 7:433, 1989

50. Raphael B, Andersen JW, Silber R, et al: Comparison of chlorambucil and prednisone versus cyclophosphamide, vincristine, and prednisone as initial treatment for chronic lymphocytic leukemia: long-term follow-up of an Eastern Cooperative Oncology Group randomized clinical trial. J Clin Oncol 9:770, 1991

51. Hansen MM, Andersen E, Birgens H, et al: CHOP versus chlorambucil + prednisolone in chronic lymphocytic leukemia. Leuk Lymphoma 5:97, 1991

52. Montserrat E, Alcala A, Alonso C, et al: A randomized trial comparing chlorambucil plus prednisone vs cyclophosphamide, melphalan, and prednisone in the treatment of chronic lymphocytic leukemia stages B and C. Nouv Rev Fr Hematol 30:429, 1988

53. Kimby E, Millstedt H: Chlorambucil/prednisone versus CHOP in symptomatic chronic lymphocytic leukemias of B-cell type: a randomized trial. Leuk Lymphoma 5:93, 1991

54. Flinn IW, Byrd JC, Morrison C, et al: Fludarabine and cyclophosphamide with filgrastim support in patients with previously untreated indolent lymphoid malignancies. Blood 96:71, 2000

55. O'Brien SM, Kantarjian HM, Cortes J, et al: Results of the fludarabine and cyclophosphamide combination regimen in chronic lymphocytic leukemia. J Clin Oncol 19:1414, 2001

56. Rabinowe SN, Soiffer RJ, Gribben JG, et al: Autologous and allogeneic bone marrow transplantation for poor prognosis patients with B-cell chronic lymphocytic leukemia. Blood 82:1366, 1993

57. Michallet M, Archimbaud E, Bandini G, et al: HLA-identical sibling bone marrow transplantation in younger patients with chronic lymphocytic leukemia. Ann Intern Med 124:311, 1996

58. Khouri IF, Przepiorka D, van Besien K, et al: Allogeneic blood or marrow transplantation for chronic lymphocytic leukaemia: timing of transplantation and potential effect of fludarabine on acute graft-versus-host disease. Br J Haematol 97:466, 1997

59. Slavin S, Nagler A, Naparstek E, et al: Nonmyeloablative stem cell transplantation and cell therapy as an alternative to conventional bone marrow transplantation with lethal cytoreduction for the treatment of malignant and nonmalignant hematologic diseases. Blood 91:756, 1998

60. Khouri IF, Keating M, Körbling M, et al: Transplant-lite: induction of graft-versus malignancy using fludarabine-based nonablative chemotherapy and allogeneic blood progenitor-cell transplantation as treatment for lymphoid malignancies. J Clin Oncol 16:2817, 1998

61. Khouri IF, Keating MJ, Vriesendorp HM, et al: Autologous and allogeneic bone marrow transplantation for chronic lymphocytic leukemia: preliminary results. J Clin Oncol 12:748, 1994

62. Pawson R, Dyer MJS, Barge R, et al: Treatment of T-cell prolymphocytic leukemia with human CD52 antibody. J Clin Oncol 15:2667, 1997

63. Österborg A, Dyer MJS, Bunjes D, et al: Phase II multicenter study of human CD52 antibody in previously treated chronic lymphocytic leukemia. J Clin Oncol 15:1567, 1997

64. Keating MJ, Byrd J, Rai K, et al: Multicenter study of CAMPATH-1H in patients with chronic lymphocytic leukemia refractory to fludarabine. Blood 94:705a (abstr 3118), 1999

65. Nguyen DT, Amess JA, Doughty H, et al: IDEC-C2B8 anti-CD20 (rituximab) immunotherapy in patients with low-grade non-Hodgkin's lymphoma and lymphoproliferative disorders: evaluation of response on 48 patients. Eur J Haematol 62:76, 1999

66. Maloney DG, Grillo-López AJ, White CA, et al: IDEC-C2B8 (rituximab) anti-CD20-monoclonal antibody therapy in patients with relapsed low-grade non-Hodgkin's lymphoma. Blood 90:2188, 1997

67. Piro LD, White CA, Grillo-Lopez AJ, et al: Extended rituximab (anti-CD20 monoclonal antibody) therapy for relapsed or refractory low-grade or follicular non-Hodgkin's lymphoma. Ann Oncol 10:655, 1999

68. Byrd JC, Murphy T, Howard RS, et al: Rituximab using a thrice weekly dosing schedule in B-cell chronic lymphocytic leukemia and small lymphocytic lymphoma demonstrates clinical activity and acceptable toxicity. J Clin Oncol 19:2153, 2001

69. O'Brien SM, Kantarjian H, Thomas DA, et al: Rituximab dose-escalation trial in chronic lymphocytic leukemia. J Clin Oncol 19:2165, 2001

70. Byrd JC, Peterson B, Park K, et al: Rituximab added to fludarabine improves response in previously untreated chronic lymphocytic leukemia: preliminary results from CALGB 9712. Proc ASCO 20:280a (abstr 1116), 2001

71. Intravenous immunoglobulin for the prevention of infection in chronic lymphocytic leukemia: a randomized, controlled clinical trial. Cooperative Group for the Study of Immunoglobulin in Chronic Lymphocytic Leukemia. N Engl J Med 319:902, 1988

72. Weeks JC, Tierney MR, Weinstein MC: Cost effectiveness of prophylactic intravenous immune globulin in chronic lymphocytic leukemia. N Engl J Med 325:81, 1991

73. Rose E, Rai K, Revicki D, et al: Clinical and health status assessments in anemic chronic lymphocytic leukemia (CLL) patients treated with epoietin alfa (EPO). Blood 84:526a (abstr 2091), 1994

74. Döhner H, Ho AD, Thaler J, et al: Pentostatin in prolymphocytic leukemia: phase II trial of the European Organization for Research and Treatment of Cancer Leukemia Cooperative Study Group. J Natl Cancer Inst 85:658, 1993

75. Pawson R, Dyer MJ, Barge R, et al: Treatment of T-cell prolymphocytic leukemia with human CD52 antibody. J Clin Oncol 15:2667, 1997

76. Grever M, Kopecky K, Foucar MK, et al: A randomized comparison of pentostatin vs alpha-interferon in previously untreated patients with hairy cell leukemia: an intergroup study. J Clin Oncol 13:974, 1995

77. Saven A, Burian C, Koziol JA, et al: Long-term follow-up of patients with hairy cell leukemia after cladribine treatment. Blood 92:1918, 1998

78. Thomas DA, O'Brien S, Cortes J, et al: Pilot study of rituximab in refractory or relapsed hairy cell leukemia. Blood 94:705a (abstr 3116), 1999

79. Kreitman RJ, Wilson WH, Stetler-Stevenson MA, et al: High complete remission rate in chemotherapy-refractory classic or variant hairy-cell leukemia induced by the anti-CD22 recombinant immunotoxin RFB4(dsFv)PE38 (BL22). Blood 96:577a (abstr 2477), 2000

80. Modiano MR, Villar-Werstler P, Crowley J, et al: Evaluation of race as a prognostic factor in multiple myeloma: an ancillary of Southwest Oncology Group study 8229. J Clin Oncol 14:974, 1996

81. Vidriales MB, Anderson KC: Adhesion of multiple myeloma cells to the bone marrow microenvironment: implications for future therapeutic strategies. Mol Med Today 2:425, 1996

82. Rettig MB, Ma HJ, Vescio RA, et al: Kaposi's sarcoma-associated herpesvirus infection of bone marrow dendritic cells from multiple myeloma patients. Science 276:1851, 1997

83. Perez-Simón JA, Garcia-Sanz R, Tabernero MD, et al: Prognostic value of numerical chromosome aberrations in multiple myeloma: a FISH analysis of 15 different chromosomes. Blood 91:3366, 1998

84. Berenson JR, Lichtenstein A, Porter L, et al: Efficacy of pamidronate in reducing skeletal events in patients with advanced multiple myeloma. N Engl J Med 334:488, 1996

85. Berenson JR, Lichtenstein A, Porter L, et al: Long-term pamidronate treatment of advanced multiple myeloma patients reduces skeletal events. Myeloma Aredia Study Group. J Clin Oncol 16:593, 1998

86. Österborg A, Boogaerts MA, Cimino R, et al: Recombinant human erythropoietin in transfusion-dependent anemic patients with multiple myeloma and non-Hodgkin's lymphoma: a randomized multicenter study. Blood 87:2675, 1996

87. Bladé J, Lust JA, Kyle RA: Immunoglobulin D multiple myeloma: presenting features, response to therapy, and survival in a series of 53 cases. J Clin Oncol 12:2398, 1994

88. Durie BGM, Salmon SE: A clinical staging system for multiple myeloma. Cancer 36: 842, 1975

89. Dimopoulos MA, Barlogie B, Smith TL, et al: High serum lactate dehydrogenase level as a marker for drug resistance and short survival in multiple myeloma. Ann Intern Med 115:931, 1991

90. Greipp PR, Lust JA, O'Fallon M, et al: Plasma cell labeling index and β_2-microglobulin predict survival independent of thymidine kinase and C-reactive protein in multiple myeloma. Blood 81:3382, 1993

91. Ballester OF, Moscinski LC, Lyman GH, et al: High levels of interleukin-6 are associated with low tumor burden and low growth fraction in multiple myeloma. Blood 83:1903, 1994

92. Witzig TE, Gertz MA, Lust JA, et al: Peripheral blood monoclonal plasma cells as a predictor of survival in patients with multiple myeloma. Blood 88:1780, 1996

93. Kay NE, Lwong T, Kyle RA, et al: Circulating blood B cells in multiple myeloma: analysis and relationship to circulating clonal cells and clinical parameters in a cohort of patients entered on the Eastern Cooperative Oncology Group phase III E9486 clinical trial. Blood 90:340, 1997

94. Moulopoulos LA, Dimopoulos MA, Smith TL, et al: Prognostic significance of magnetic resonance imaging in patients with asymptomatic multiple myeloma. J Clin Oncol 13:251, 1995

95. Attal M, Harousseau JL, Stoppa AM, et al: A prospective, randomized trial of autologous bone marrow transplantation and chemotherapy in multiple myeloma. N Engl J Med 335:91, 1996

96. Singhal S, Mehta J, Desikan R, et al: Antitumor activity of thalidomide in refractory multiple myeloma. N Engl J Med 341:1565, 1999

97. Harris NL, Jaffe ES, Diebold J, et al: The World Health Organization classification of neoplastic diseases of the haematopoietic and lymphoid tissues: Report of the Clinical Advisory Committee Meeting, Airlie House, Virginia, November 1997. Histopathology 36:69, 2000

98. Falk RH, Comenzo RL, Skinner M: The systemic amyloidoses. N Engl J Med 337:898, 1997

99. Kyle RA, Gertz MA, Greipp PR, et al: A trial of three regimens for primary amyloidosis: colchicine alone, melphalan and prednisone, and melphalan, prednisone, and colchicine. N Engl J Med 336:1202, 1997

100. Comenzo RL, Vosburgh E, Falk RH, et al: Dose-intensive melphalan with blood stem-cell support for the treatment of AL (amyloid light-chain) amyloidosis: survival and responses in 25 patients. Blood 91:3662, 1998

XVI ACUTE LEUKEMIA

Richard A. Larson, m.d.

The acute leukemias are characterized by aberrant differentiation and proliferation of malignantly transformed hematopoietic progenitor cells. These cells accumulate within the bone marrow and, once a substantial burden of leukemia cells is present, lead to suppression of the growth and differentiation of normal blood cells. Symptoms result from varying degrees of anemia, neutropenia, and thrombocytopenia or from infiltration into tissues. Although virtually any organ system may become involved once leukemia cells enter the peripheral blood, the most common sites detected clinically are the lymph nodes, liver, spleen, central nervous system, and skin.

The first recognition of leukemia as a distinctive entity is usually accorded independently to Virchow in Berlin and Bennett in Scotland in 1845. In 1847, Virchow coined the term leukemia (from the Greek *leuk*, white cells, and *emia*, in the blood) to replace "weisshäme," a German term. Leukemic disorders are currently classified by their presumed cell of origin. Acute myeloid leukemia (AML) results from the malignant transformation of a bone marrow (myeloid) progenitor cell or stem cell, which is the normal precursor for granulocytes, erythrocytes, and megakaryocytes.[1] Acute lymphoblastic leukemia (ALL) is a malignant disease of early precursor cells of the B cell and T cell lymphocytic lineages. Myelodysplastic syndrome (MDS) is any one of a heterogeneous group of malignant hematologic disorders that often involve all three myeloid cell lineages and probably arise from a very primitive or totipotential stem cell.[2] MDS combines bone marrow failure with many features of a chronic or so-called smoldering leukemia that can evolve into AML.

Epidemiology

Although leukemia can occur at any age, its incidence is strongly related to increasing age. Across the entire age spectrum, ALL and AML are nearly equal in overall incidence, but ALL predominates in children, whereas AML accounts for the majority of cases in adults. The median age at diagnosis of AML is about 60 years. The peak incidence of ALL occurs at ages 3 to 5 years, and childhood cases are distinctly different from ALL in adults. Although the median age for adults with ALL who are entered into clinical trials is 35 to 40 years, it is very likely that older patients are underrepresented in these reports. There are marked differences between the various subtypes of leukemia that occur in children, young adults, and older adults. MDS is rarely observed in adults younger than 50 years and becomes increasingly common in older age groups. There are many biologic and clinical similarities between MDS and the more common subtypes of AML that are seen in older adults, and some patients with AML may have disease that evolved from an unrecognized myelodysplastic syndrome.

Acute leukemia is more common in whites than blacks at all ages.[3] Jews are more commonly affected than non-Jews, but it is unclear whether this disparity is the result of genetic or environmental factors. Acute leukemia is more common in men than in women. This difference is caused in part by a higher incidence in young boys and older men. Factors such as occupational and environmental exposure may account for some gender disparity.

Geographic variations in incidence are likely related to a number of factors, including socioeconomics, ethnicity, and an urban or rural setting, and may account for the higher frequency reported in industrialized countries and urban areas. Some cytogenetic abnormalities have been reported more frequently in certain countries.

Etiology

Although the cause of acute leukemia in humans is unknown, a variety of hereditary and environmental factors appear to play an etiologic role. The pathogenesis of acute leukemia involves complex interactions between host susceptibility, chromosomal damage secondary to physical or chemical exposure, and possibly the incorporation of genetic information transmitted virally into susceptible progenitor cells.[4] Much of the evidence is indirect and has been inferred from epidemiologic studies. Investigators have relied heavily on animal models of leukemogenesis. In certain nonhuman species, for instance, exposure to an agent such as ionizing radiation or a chemical carcinogen increases the incidence of leukemia.

Characteristic leukemic disorders can be generated in animal models by transfecting fusion genes such as *bcr-abl* into germ cells. The *bcr-abl* fusion gene results from the translocation or juxtaposition of the *abl* gene from chromosome 9 with the *bcr* gene on chromosome 22. This novel fusion gene produces a protein that is not present in normal cells. The t(9;22) gives rise to an abnormal chromosome 22 that is called the Philadelphia chromosome (Ph). This rearrangement is found in about one third of adults with ALL, as well as in all patients with chronic myeloid leukemia (CML).

RADIATION

The leukemogenic potential of ionizing radiation has been well recognized for a number of years. Leukemias associated with radiation generally derive from the myeloid lineage but often have trilineage features characteristic of multipotential stem cells. Much of the epidemiologic evidence emanates from observations after human exposures to radiation from nuclear explosions, from therapeutic or diagnostic radiation sources, and from occupational radiation sources. After the atomic-bomb attacks on Hiroshima and Nagasaki, exposed survivors experienced a marked increase in the incidence of both acute and chronic leukemia.[5] The chance of leukemia developing in a person was related to the intensity of the person's radiation exposure, with relatively little risk observed in those exposed to less than 100 cGy.[6]

Patients who were treated with external-beam irradiation for ankylosing spondylitis or menorrhagia or with radioactive phosphorus (^{32}P) for polycythemia vera have experienced an increased incidence of acute leukemia.[7,8] The risk of therapy-related AML (t-AML) after the use of radiation therapy alone to treat Hodgkin disease or breast cancer appears low.[9] Current use of diagnostic x-ray imaging does not appear to result in any increased leukemia risk to patients. However, fetal exposure to intrauterine x-rays increases the risk of subsequent childhood leukemia.[10]

The risk of leukemia developing in persons exposed to radia-

tion at their workplace was documented in early studies of women employed as radium wristwatch dial painters and by the increased incidence of leukemia in radiologists during the early part of the 20th century.[11,12] Current radiation protection practices appear to have markedly diminished this risk.

CHEMICALS

There are substantial epidemiologic data linking the development of acute leukemia to either therapeutic or industrial chemicals. Exposure to alkylating agents such as melphalan or nitrogen mustard is associated with the development of t-AML, after a latency of 5 to 7 years.[9] This disease is characterized by pancytopenia, trilineage myelodysplasia, and complex cytogenetic abnormalities, most often involving losses of part or all of chromosomes 5 or 7. t-AML is a well-recognized complication resulting from treatment of Hodgkin disease, multiple myeloma, and ovarian cancer.[13] The risk is greater in patients receiving both chemotherapy and radiotherapy. Another subtype of t-AML has been observed in patients treated for certain hematologic and solid neoplasms with the epipodophyllotoxin drugs etoposide and teniposide and has also been observed with other agents that inhibit topoisomerase II activity, such as doxorubicin.[9,14] This leukemia has a much shorter latency, often only 1 to 2 years, and a more acute presentation, with a high leukocyte count, monoblastic morphology, and usually a cytogenetic abnormality of the long arm of chromosome 11 at band q23. The *MLL* (mixed lineage leukemia) gene at this locus is frequently rearranged.[15]

In the early part of the 20th century, acute leukemia was noted in workers exposed to benzene. For example, the incidence of acute leukemia in Turkish cobblers with long-term exposure to benzene was two to three times greater than the incidence of acute leukemia in the general population.[16] With the advent of stronger occupational safety laws, benzene-associated leukemia has nearly disappeared in Western countries but is still present in developing countries.[17]

VIRUSES

A number of viruses are known to cause acute leukemia in several nonhuman species. The potential role of viruses in human leukemogenesis has been explored for decades. There are no conclusive data that any of the common types of acute leukemia in humans have a viral origin. The Epstein-Barr virus (EBV) genome is present in the majority of Burkitt leukemia/lymphoma cells. Adult T cell leukemia, an uncommon form of human leukemia that is largely restricted to southern Japan and the Caribbean, is closely linked to infection by a leukemogenic virus, human T cell lymphotropic virus type I (HTLV-I).[18] However, most persons who express antibodies to this virus never exhibit evidence of malignancy.

HEREDITY AND GENETICS

Observations of the increased incidence of acute leukemia in members of certain families with high susceptibility, the high frequency of concordant leukemia in monozygotic twins, and the association of acute leukemia in persons with genetic disorders have established that hereditary factors play a role in the development of leukemia.[4] In identical twins, if one twin has leukemia, the other is also at risk for leukemia, often before 8 years of age.[19] The disease usually develops in the second twin within a year of the first twin's diagnosis. Nonidentical siblings of those affected by leukemia have a lesser risk, but the risk is still greater than that observed in the general population.[20] Molecular evidence of the *MLL-AF4* fusion gene has been detected in newborns who later developed ALL with t(4;11) during infancy or childhood.

The incidence of both AML and ALL is increased in patients with Down syndrome, a disease characterized by trisomy 21 resulting from chromosomal nondisjunction. Approximately 5% to 10% of childhood AML occurs in children with Down syndrome. Their prognosis is more favorable than that for children of the same age with AML but without Down syndrome.[21] Further evidence of the relation between leukemia and chromosomal abnormalities is the increased incidence in patients with Bloom syndrome and Fanconi anemia. Both disorders are hereditary conditions characterized by increased chromosomal breakage. Other genetic conditions not normally associated with chromosomal abnormalities, such as ataxia telangiectasia and congenital agammaglobulinemia, are associated with acute leukemia. The deficiencies of cellular and humoral immunity that mark these conditions may increase susceptibility to the development of leukemia.

Morphologic Classification of Leukemia

The traditional classification of the acute leukemias has relied on morphologic description, reflecting the predominant cell type present within the bone marrow population and relating that cell to its normal hematopoietic counterpart. In 1976, the French-American-British (FAB) group of hematopathologists established a subclassification system for the acute leukemias that separated AML and ALL into distinct disorders.[22] This system was based solely on light-microscopic evaluation of routinely stained blood and marrow smears, supplemented by a limited number of cytochemical procedures. Because of its ease of use

Table 1 World Health Organization Classification of Acute Myeloid Leukemia

- AML with recurrent genetic abnormalities
  - AML with t(8;21)(q22;q22); *AML/ETO*
  - AML with abnormal bone marrow eosinophils and inv(16)(p13q22) or t(16;16)(q22;p13); *CBFβ/MYH11*
  - Acute promyelocytic leukemia with t(15;17)(q22;q12); *PML/RARα* and variants
  - AML with 11q23 (*MLL*) abnormalities
- AML with multilineage dysplasia
  - Following MDS or MDS/myeloproliferative disorder
  - Without antecedent MDS
- t-AML and t-MDS
  - Alkylating agent–related
  - Topoisomerase type II inhibitor–related
  - Other types
- AML not otherwise categorized
  - AML minimally differentiated
  - AML without maturation
  - AML with maturation
  - Acute myelomonocytic leukemia
  - Acute monoblastic and monocytic leukemia
  - Acute erythroid leukemia
  - Acute megakaryoblastic leukemia
  - Acute basophilic leukemia
  - Acute panmyelosis with myelofibrosis
  - Myeloid sarcoma

AML—acute myeloid leukemia MDS—myelodysplastic syndrome t-AML—therapy-related acute myeloid leukemia t-MDS—therapy-related myelodysplastic syndrome

Table 2 French-American-British (FAB) Classification of Acute Myeloid Leukemia

FAB Classification	*Subtype*	*Morphologic and Cytochemical Features*	*Cytogenetics*	*Frequency (%)*
M0	Minimally differentiated acute myeloblastic leukemia	Large, agranular myeloblasts, sometimes resembling lymphoblasts of FAB subtype L2; stain negative for myeloperoxidase and Sudan black; express CD13 or CD33 antigens on cell surface	—	2–3
M1	Acute myeloblastic leukemia without maturation	Large, poorly differentiated myeloblasts represent 90% or more of the nonerythroid cells; at least 3% of the myeloblasts stain positive for myeloperoxidase	—	20
M2	Acute myeloblastic leukemia with maturation	Between 30% and 89% of the nonerythroid cells are myeloblasts having abundant cytoplasm with moderate to many granules; Auer rods are often visible	t(8;21)(q22;q22)	25–30
M3	Acute promyelocytic leukemia (APL)	Leukemia cells usually contain heavy azurophilic granulation; nuclear size varies greatly; nuclei are often bilobed or kidney shaped; some cells contain bundles of Auer rods; leukemia cells stain strongly positive for myeloperoxidase; usually HLA-DR negative; there is a microgranular variant (M3V)	t(15;17)(q22;q11-12); t(11;17)	8–15
M4	Acute myelomonocytic leukemia (AMMoL)	Myeloblasts constitute over 30% of the nonerythroid cells, but total granulocytic precursors do not exceed 80%; monocytic cells account for > 20% of the nonerythroid cells; nonspecific esterase and chloroacetate stains are often positive; Auer rods may be present	—	20–25
M4Eo	AMMoL with abnormal eosinophils	Myelomonoblasts plus morphologically and cytochemically abnormal eosinophils	inv(16)(p13;q22); t(16;16)(p13;q22)	5
M5	Acute monoblastic leukemia (AMoL)	Monoblasts, promonocytes, or monocytes constitute 80% or more of the nonerythroid cells; in one subtype (M5A), 80% or more of all the monocytic cells are monoblasts; in the well-differentiated subtype (M5B), less than 80% are monoblasts; α-naphthyl acetate positivity is extinguished by NaF	t(9;11)(p22;q23); t(11q23)	10
M6	Acute erythroleukemia (AEL)	Greater than 50% of the nucleated marrow cells are erythroid; erythroblasts are usually strongly PAS positive; myeloblasts represent 30% or more of the nonerythroid cells	—	5
M7	Acute megakaryoblastic leukemia (AMegaL)	Large and small megakaryoblasts with high nuclear/cytoplasm ratio; cytoplasm is pale and agranular; standard cytochemical stains are not definitive; platelet peroxidase and platelet-specific antibodies are often positive	t(1;22)(p13;q13)	1–2

and applicability when comparing treatment results among institutions, this system has been widely adopted. In addition, the FAB system was shown to have independent prognostic significance regarding survival. The system was revised in 1985 to provide clarification and extend to new diagnostic techniques.[23] Most recently, a committee of the World Health Organization (WHO) described a comprehensive classification scheme that utilizes morphology, immunophenotyping, etiology, and cytogenetics and more clearly distinguishes between AML, MDS, and chronic myeloproliferative disorders [*see Table 1*].[24]

ACUTE MYELOID LEUKEMIA

The blast cells of patients with AML are most often larger than lymphoblasts and display a greater heterogeneity in size and shape. AML blasts have abundant cytoplasm and often contain cytoplasmic granules. Auer rods, azurophilic crystalline-like accumulations of abnormal lysosomal granules visible in the cytoplasm with Wright staining, are detected in about 10% of patients with AML.

A diagnosis of AML is now established when 20% or more of all the nucleated marrow cells are blast cells.[24] The FAB group defined eight variants of AML [*see Table 2*], including three types with predominantly granulocytic differentiation (M1, M2, and M3), two with at least 20% monocytic precursors (M4 and M5), one with a high proportion of erythroblasts (M6), and a more recently recognized and rarely occurring variant with predominance of megakaryoblasts (M7).[23] In addition, the FAB group described a form of AML with minimal myeloid differentiation, designated M0, which cannot be diagnosed solely on morphologic or cytochemical grounds but requires the added use of immunohistochemical staining.[25] These cases express myeloid antigens on the blast cell surface but lack myeloperoxidase reactivity.

MYELODYSPLASTIC SYNDROME

MDS is any one of a heterogeneous group of hematopoietic stem cell disorders characterized by cytologic dysplasia in the bone marrow and blood and by various combinations of anemia, neutropenia, and thrombocytopenia.[2,24,26] These disorders have in common the progressive evolution of a monoclonal population of hematopoietic cells, usually involving multiple lineages, generally with accompanying suppression of normal hematopoiesis. The natural history of these syndromes varies widely, ranging from chronic anemias with a low propensity for leukemic conversion to syndromes with severe hematologic disturbances and a high risk of progression to AML. The examination of well-prepared peripheral blood smears, bone marrow aspirate smears, and bone marrow biopsy specimens can usually confirm the diagnosis of MDS. The marrow should be at least normocellular for the patient's age, and cytologic evaluation of the blood and aspirate smear should show dysplasia in at least one cell line. It should be emphasized that these abnormalities

are not specific to MDS but can be seen in a variety of other inherited and acquired hematologic disorders. Care must be taken to exclude these other processes before making a diagnosis of MDS.

A variety of terms, including preleukemia, smoldering or subacute leukemia, dysmyelopoietic syndrome, and myelodysplasia, have all been used to describe MDS. There is considerable evidence that these disorders are clonal and neoplastic from their earliest detection. Thus, the term preleukemia seems inappropriate, because it implies a premalignant condition. Rather, these disorders are better considered as a chronic or smoldering leukemia. Unfortunately, our diagnostic abilities and nomenclature are not yet adequate to discriminate between clinically distinct subsets within these syndromes, with rare exceptions. There are considerable overlaps with other morphologically determined diseases, particularly those categorized as chronic myeloproliferative disorders (CMPDs). Whereas differentiation of hematopoiesis is qualitatively unimpaired in CMPDs, at least in their early stages, MDS is characterized chiefly by impaired differentiation. Thus, one can speculate that MDS results from alterations of genes that control transcription and cellular differentiation rather than those that regulate cell proliferation or survival.

In 1982, the FAB group attempted to standardize the classification of MDS patients by use of criteria based on cytology and the number of blast cells in the marrow and peripheral blood.[26] Later, another important subset was described: refractory cytopenia with multilineage dysplasia (RCMD).[27] The WHO classification for MDS is now widely accepted [*see Table 3*].[24] In some patients, there is a natural progression of disease between categories as cellular maturation becomes more arrested and blast cells accumulate. In other patients, however, the diagnostic category does not change during the patient's lifespan.

Increased use of cytogenetic analysis and the development of new techniques for assessing clonality are proving useful not only for diagnosis but also for providing insights into pathogenesis and patterns of responsiveness to growth factors and possible differentiation-inducing agents.[28-30] Considerable data suggest that MDS results from combined defects of both stroma and hematopoietic stem cells. Several clinical syndromes that may have a more predictable natural history can now be defined. For example, a deletion of the long arm of chromosome 5 can be detected in some older patients, especially women, with a macrocytic, refractory anemia (RA). The platelet count is typically normal or elevated. The bone marrow picture in the RA with $5q^-$ syndrome is characterized by the presence of monolobulated and bilobulated micromegakaryocytes. Two thirds of these patients have RA or RA with ringed sideroblasts (RARS), and the remainder has RAEB (RA with excess of blasts). In those patients who have a del(5q) as their sole cytogenetic abnormality, MDS tends to follow a more benign course, although progression to AML may occur.[30] Although MDS is not commonly seen in patients younger than 50 years, there are two distinct pediatric syndromes of importance: juvenile chronic myelomonocytic leukemia and the monosomy 7 syndrome.

ACUTE LYMPHOBLASTIC LEUKEMIA

The FAB classification system described three subtypes of ALL (L1, L2, and L3), defined by individual cytologic features such as cell size, nuclear chromatin pattern, nuclear shape, nucleoli, and amount of basophilia in the cytoplasm.[31] The subtype L1 accounts for over 80% of the ALL cases in children and consists of predominantly small cells, up to twice the diameter of a small lymphocyte. The majority of adult cases of ALL are of subtype L2. These cells are larger than cells in L1 and are often heterogeneous in size. This distinction, however, is often arbitrary and provides little guidance for the management of individual patients. Consequently, it has been abandoned. The

Table 3 World Health Organization Classification of Primary Myelodysplastic Syndrome (MDS)

	Subtype	*Peripheral Blood*	*Bone Marrow*
RA	Refractory anemia	< 1% blasts; reticulocytopenia; macrocytic or normochromic/normocytic	Usually erythroid hyperplasia with dyserythropoiesis; < 5% blasts
RA del(5q)	RA with isolated del(5q)	Anemia; < 5% blasts	< 5% blasts; isolated del(5q); no Auer rods
RARS	RA with ringed sideroblasts	No blasts; dimorphic red cell morphology	As in RA, but sideroblasts ≥ 15% of erythroid precursors
RCMD	Refractory cytopenia with multilineage dysplasia	< 1% blasts; cytopenia; multilineage dysplasia; no Auer rods	< 5% blasts; multilineage dysplasia (≥ 10% of cells in ≥ two myeloid cell lines)
RCMD-RS	Refractory cytopenia with multilineage dysplasia and ringed sideroblasts	< 1% blasts; cytopenia; multilineage dysplasia; no Auer rods	< 5% blasts; multilineage dysplasia (≥ 10% of cells in ≥ two myeloid cell lines); ≥ 15% ringed sideroblasts
RAEB-1	RA with excess of blasts–1	< 5% blasts; cytopenias in two or three cell lines; no Auer rods	5%–9% blasts; unilineage or multilineage dyspoiesis
RAEB-2	RA with excess of blasts–2	5%–19% blasts; cytopenias in two or three cell lines; may have Auer rods	10%–19% blasts; unilineage or multilineage dyspoiesis
MDS-U	MDS-unclassified	< 1% blasts; cytopenias	< 5% blasts; unilineage dysplasia; no Auer rods
CMMoL	Chronic myelomonocytic leukemia	< 20% blasts or promonocytes; > 1×10^9/l monocytes; dysplasia in ≥ one myeloid lineage	< 20% blasts or promonocytes; no *BCR/ABL* fusion gene
Atypical CML	Atypical chronic myeloid leukemia	Neutrophilic leukocytosis; prominent dysgranulopoiesis; basophils < 2%	< 20% blasts; no *BCR/ABL* fusion gene; hypercellular

WHO classification recognizes only precursor B cell and precursor T cell subgroups.[24] The least common form of ALL, seen in approximately 3% to 4% of both children and adults, is termed L3. In this subtype, the cells are morphologically identical to the neoplastic cells in Burkitt lymphoma. It is important to recognize L3 because initial treatment of this variant is markedly different from the treatment of other subtypes of ALL. In L3, the cells are large and uniform, with finely stippled chromatin and regular nuclear shape. Nucleoli are often prominent. Cytoplasm is moderately abundant and deeply basophilic.

Cytochemical evaluations of blast cells in ALL reveal characteristic patterns. By definition, stains for lysosomal enzymes such as myeloperoxidase or the Sudan black reaction must be negative to support a diagnosis of ALL. The periodic acid–Schiff reaction will reveal clumpy positivity caused by glycogen in ALL blasts (except L3, which reacts negatively) but is a poor discriminator of cell lineage because many AML cells will also react positively. Chloroacetate esterase and lysozyme stains are negative in ALL, but α-naphthyl acetate esterase may be positive in T lymphoblasts. ALL blast cells contain the enzyme terminal deoxynucleotidyl transferase (TdT), which, when present in the great majority of cells, is a fairly reliable marker for ALL. L3 leukemia cells often stain positively with oil red O because of the neutral lipid within the cytoplasmic vacuoles.

Immunophenotyping of Acute Leukemia

Morphologic examination of blood or bone marrow smears sometimes fails to provide an unequivocal diagnosis. However, identification of various differentiation antigens on the surface of the abnormal cells by flow cytometry studies can rapidly provide this critical information.[32]

ACUTE MYELOID LEUKEMIA

As yet, no AML-specific surface antigens have been identified. Rather, AML cells often coexpress antigens on the cell surface that are not normally coexpressed by their normal myeloid counterparts.[33] Efforts to correlate immunologic phenotyping with the FAB classification of AML or with clinical outcome have in general been imprecise. However, the expression of certain antigens, such as CD34, an early stem cell marker, may carry important prognostic information.[34] In addition, CD33, which is expressed on most myeloblasts, has become an important target for monoclonal antibody–directed therapy.

ACUTE LYMPHOBLASTIC LEUKEMIA

Approximately 80% of cases of ALL arise from the B cell lineage and express B cell differentiation antigens (CD19, CD20, or both) and have undergone heavy-chain and light-chain immunoglobulin gene rearrangement.[35] The leukemia cells in many of these cases also express the common ALL antigen, or CALLA, which is designated CD10. Lymphoblasts from patients with progenitor B cell ALL do not express CD10. CD10-bearing cells can be further divided on the basis of the presence of cytoplasmic μ heavy-chain immunoglobulin (cμ). Most cases of ALL do not express cμ on their blast cells and are termed progenitor-B ALL (if CD10-negative) or common ALL (if CD10-positive). Approximately 20% of CD10-positive cases express cμ and are designated pre-B ALL. Although the immunoglobulin genes are always clonally rearranged in B-lineage ALL, surface immunoglobulin expression occurs in only 2% to 5% of ALL cases. These disorders are termed mature B cell ALL or Burkitt-type ALL and typically display the FAB L3 morphology.

Between 15% and 20% of ALL cases arise from the T cell lineage.[32,35] These cells express T cell antigens such as CD2, CD5, and CD7. CD10 may also be present, but CD19 and CD20 are not. In the majority of cases, one or more of the T cell receptor genes are rearranged. Further subclassification of T cell ALL into early, intermediate, and mature thymocyte types is based on the expression of various T cell differentiation antigens.

Some antigens normally present only on myeloid cells can be expressed by malignant lymphoblasts of B or T cell origin. This disorder is called myeloid antigen–positive ALL.[35,36] A small percentage of ALL cases lack either B cell or T cell features and are termed acute undifferentiated leukemia.

HYBRID LEUKEMIAS

Application of immunophenotyping and molecular probes has uncovered cases in which leukemia cells display characteristics of myeloid and lymphoid cells. In such instances, a single neoplastic cell may coexpress features of distinct lineages (biphenotypic), or two distinct subpopulations of leukemia cells may express either myeloid or lymphoid features separately (bilineal). Various theories have been put forward to explain the occurrence of these hybrid leukemias. One hypothesis, termed lineage infidelity, states that the leukemia cell displays aberrant gene expression by virtue of its neoplastic transformation. Another theory, lineage promiscuity, proposes that normally differentiating cells may express characteristics of more than one distinct lineage and that the leukemia cell is merely reflecting that particular phase in a cell's development. Finally, a theory termed lineage switching proposes that in some cases leukemia is a malignancy of a pluripotent stem cell capable of differentiation along either a myeloid or a lymphoid lineage. Thus, any individual case may express one or both phenotypes. The prognostic significance of such lineage infidelity is uncertain, but these patients often have poor outcomes regardless of whether AML or ALL therapies are used.

Cytogenetics

Cytogenetic abnormalities occur commonly in the acute leukemias and are often used to identify and define the malignant hematopoietic clone.[15] Since the application of chromosome banding techniques in the early 1970s, cytogenetic analysis has provided the most clinically useful approach to subclassifying the acute leukemias. Particular nonrandom chromosomal abnormalities correlate with specific morphologic subtypes and clinical profiles.[15,37-39] These cytogenetic abnormalities are somatic (rather than germline) mutations that frequently result from translocations of chromosomal DNA, resulting in new (abnormal) protein products from the resultant fusion genes. It is assumed that the protein products from these fusion genes are responsible for the cellular dysregulation that leads to the malignant state. Such recurring chromosomal abnormalities are critical in determining therapeutic strategy and have provided important independent information regarding response to therapy and overall prognosis [*see Tables 4 and 5*].

Cytogenetic analysis has furthered our understanding of leukemogenesis.[15] Careful analyses of the chromosomal breakpoints associated with leukemia-specific cytogenetic abnormalities have permitted identification of a number of genes that appear to play an integral role in leukemogenesis.[40,41] The chromosomal location of a large number of these so-called oncogenes

Table 4 Cytogenetic Subsets in Acute Myeloid Leukemia, Treatment, and Outcomes

Karyotype	*Complete Remission Rate*	*Remission Duration*	*Treatment Approach*
t(8;21)(q22;q22)	High	Long	Standard induction with cytarabine and an anthracycline; intensive consolidation with several courses of high-dose cytarabine
inv(16)(p13;q22) or t(16;16)(p13;q22)	High	Long	Standard induction with cytarabine and an anthracycline; intensive consolidation with several courses of high-dose cytarabine
t(15;17)(q22;q11-12)	High	Long	All-*trans*-retinoic acid together with an anthracycline; arsenic trioxide to treat relapse
t(9;11)(p22;q23)	High	Intermediate	Standard induction and intensive consolidation with high-dose cytarabine; reserve stem cell transplantation for second remission for most t(9;11) patients
del(5q), +13, +8, -7, inv 3, del(12p), t(9;22), other t(11q23), or complex abnormalities	Low	Short	New induction regimens, including use of growth factors during or after chemotherapy, or modulators of drug resistance; perform stem cell transplantation in first complete remission

has been identified. The functions of these oncogenes are being investigated, but many are involved in intracellular signaling pathways and the control of cellular proliferation and differentiation. Structural chromosomal changes may lead to activation or perturbation of oncogene expression resulting in disturbance of cellular regulation and eventually malignant transformation. Deletions or loss of DNA may eliminate genes that have tumor suppressor functions.

General Principles of Therapy

The goal of remission induction chemotherapy is the rapid restoration of normal bone marrow function. The term complete remission (CR) is reserved for patients who have full recovery of normal peripheral blood counts and bone marrow cellularity with less than 5% residual blast cells [*see Table 6*]. Induction therapy aims to reduce the total-body leukemia cell population from approximately 10^{12} cells to below the cytologically detectable level of about 10^{9} cells. The leukemia cells in some patients have high levels of primary drug resistance and will be refractory to courses of remission induction chemotherapy. It is assumed, however, that even in CR a substantial burden of leukemia cells persists undetected, leading to relapse within a few weeks or months if no further therapy is administered. Postinduction or remission consolidation therapy, usually comprising several additional courses of chemotherapy, is designed to eradicate residual leukemia, allowing the possibility of cure. Multiple chemotherapy drugs in high doses are typically used to prevent emergence of resistant subclones and to limit cumulative and overlapping toxicities. Lower doses of prolonged remission maintenance therapy lasting 1 to 3 years have been used with some success in ALL, but this adjunctive therapy has uncertain value in AML.

Complications of cytotoxic therapy include tumor lysis syndrome with urate nephropathy and electrolyte imbalance (hyperkalemia, hypocalcemia, and hyperphosphatemia), gastrointestinal injury (mucositis and diarrhea), thrombocytopenic bleeding, and neutropenic infections. Prophylactic measures and supportive care are critically important during the treatment period. Patients must receive adequate intravenous hydration. Hyperuricemia responds rapidly to rasburicase. Allopurinol, parenteral hydration, and alkalinization of the urine reduce the likelihood of uric acid precipitation in the renal tubules. Blood products are transfused to maintain a platelet count greater than 10,000/μl in a nonbleeding patient and a hematocrit greater than 25%. Broad-spectrum antibiotics are be-

Table 5 Cytogenetic and Molecular Subtypes of Acute Lymphoblastic Leukemia

Subtype	*Karyotype*	*Frequency (%)*		*Disease-Free Survival**	
		Children	*Adults*	*Children*	*Adults*
Burkitt cell	t(8;14)	2	3–5	75–85	60–70
Hyperdiploid	> 50 chromosomes	25	2–5	80–90	40–50
TEL/AML1	t(12;21)	20–25	1–3	85–90	?
E2A/PBX1	t(1;19)	5–6	1–3	70–80	10–50
MLL/AF4	t(4;11)	2–3	3–6	10–35	10–40
bcr/abl	t(9;22)	3–4	25–30	20–40	< 10
Hypodiploid	< 45 chromosomes	7	4–5	25–40	10
T cell ALL	t(14q11) and others	15	20–25	65–75	60

*Percent at 3 to 5 yr.

Table 6 Terminology Used in Leukemia Treatment

Term	*Definition*
Remission induction therapy	Initial chemotherapy treatment aimed at achieving a complete remission (CR)
Remission consolidation (intensification) therapy	Postremission therapy aimed at destroying clinically occult disease
Maintenance (continuation) therapy	Lower dose of chemotherapy aimed at preventing reemergence of leukemia
CR	Disappearance of leukemia after treatment, with full regeneration of normal hematopoiesis
Cytogenetic (or molecular) CR	Inability to detect residual leukemia by use of genetic methods
Minimal residual disease	Persistent leukemia, not detectable by light microscopy
Refractory disease	Leukemia that does not enter CR
Relapsed disease	Clinically overt recurrence of leukemia after CR
Immunophenotype	Pattern of expression of cell surface or cytoplasmic markers
Cytogenetics	Analysis of chromosomes in metaphase cells for additions, losses, or rearrangements
Multidrug resistance	Biologic feature that protects cells from certain classes of chemotherapy drugs through different membrane mechanisms
Stem cell transplantation	Transfusion of hematopoietic stem cells from the patient (autologous) or a normal donor (allogeneic) after high-dose chemoradiotherapy

gun empirically whenever the temperature exceeds 38.5° C (101.3° F) in a neutropenic patient (neutrophils < 500/μl), and antifungal therapy is added for persistently febrile patients.

Myeloablative therapy followed by transplantation of bone marrow or hematopoietic stem cells (SCT) using an HLA-identical sibling donor is an established treatment modality in acute leukemia and is indicated for suitable high-risk patients in first remission or for any young or middle-aged patient in first relapse or second remission.[42] Allogeneic SCT (alloSCT) has two therapeutic components. Intensive myeloablative therapy is used to eradicate all tumor cells, if possible. Secondly, T cells in the donor marrow can produce a graft versus leukemia immune response that can destroy remaining leukemia cells; this effect has been correlated with improved disease-free survival. Unfortunately, this beneficial immune response is closely associated with acute and chronic graft versus host disease (GVHD), a major cause of morbidity and mortality after alloSCT. GVHD can be reduced by T cell depletion from the donor marrow but only at the cost of increased rates of graft failure and leukemia relapse. Because the risk of treatment-related mortality increases with age, most centers restrict SCT to patients younger than 60 years. The use of alloSCT is also limited in part by donor availability. A patient has only a 25% to 30% chance that a sibling will have inherited the identical HLA alleles.

Autologous SCT allows the use of myeloablative therapy in the absence of an allogeneic marrow donor as well as in older patients.[43] The appropriate role for this treatment modality is controversial. Treatment-related morbidity and mortality (< 5%) are relatively low, thus allowing its use in older patients, but relapse rates are high, and overall outcomes are not clearly better than in patients who receive intensive but nonablative therapy. Peripheral blood stem cell transplantation appears to be as good as, if not better than, SCT and is being used more frequently in the treatment of hematologic cancer. A recent study indicated that peripheral stem cells restore blood counts faster than marrow cells without increasing the risk of GVHD.[44]

Acute Myeloid Leukemia

DIAGNOSIS

Clinical Features

Typically, antecedent symptoms such as fatigue are brief in patients with AML. About one third of patients will present with bruising or hemorrhage. One quarter of patients will have a serious infection involving the lungs, soft tissues, or skin. Splenomegaly and hepatomegaly are not common, occurring in less than 25% of patients. Lymphadenopathy is even less common. Gingival hypertrophy or skin infiltration by leukemia (called leukemia cutis) occurs in 50% of patients with monocytic leukemia. Patients with acute promyelocytic leukemia (APL) frequently present with more severe bleeding from disseminated intravascular coagulation.

The leukocyte count is elevated above the normal range in 50% of AML patients. Counts above 100,000/μl are reported in fewer than 10% of patients, but these patients may have severe CNS effects or respiratory distress from leukostasis; emergency leukapheresis and chemotherapy may be lifesaving. The peripheral blood contains some leukemic blast cells in 85% to 90% of cases. The absolute neutrophil count is almost always depressed in AML and is less than 1,500/μl in 50% of patients at diagnosis. A moderate degree of anemia is common, and the platelet count is typically less than 100,000/μl and often less than 20,000/μl.

Mild to moderate elevations of serum uric acid levels are common and typically reflect increased cell turnover. The serum lactate dehydrogenase (LDH) level may be elevated but not as commonly as in ALL. Muramidase (lysozyme) is elevated in the serum or urine of patients with acute myelomonocytic or monoblastic leukemia and may contribute to renal dysfunction.

AML is not limited to the bone marrow and peripheral blood. Abnormalities in one or more organ systems may result from leukemia cell infiltration or metabolic complications related to leukemia. Rarely, a patient with AML develops a solid mass of leukemia cells called a granulocytic or myeloid sarcoma (chloroma). The skin and bones (particularly the sternum, ribs, and orbit) are most commonly involved, but myeloid sarcomas can occur in any organ. Respiratory distress in patients with AML is most often caused by infection. However, patients with very high numbers of circulating blast cells (> 100,000/μl) may experience severe dyspnea and hypoxemia because of leukostasis within the pulmonary vasculature. Cardiac dysfunction, including murmurs, heart failure, and dysrhythmias, is most often secondary to anemia.

Retinal hemorrhages are most often caused by thrombocytopenia. However, patients with extreme hyperleukocytosis may develop so-called cotton-wool spots as a result of retinal ischemia. Frank CNS involvement and cranial neuropathies are unusual at the time of initial diagnosis of AML.

Cytogenetic Abnormalities

Clonal chromosomal abnormalities can be detected in most cases of AML.[1,15] These abnormalities include gains or losses of whole chromosomes or losses of the long (q) or short (p) arms of chromosomes (deletions), as well as a variety of structural re-

arrangements (translocations, inversions, or insertions).

It is strongly recommended that cytogenetic analysis be performed before initiation of therapy on every newly diagnosed patient because studies of the prognostic significance of recurring cytogenetic abnormalities in AML have yielded consistently similar results.[1,38] Cytogenetic characterization has become the strongest predictor of both response to therapy and remission duration.[37] Thus, in many centers, plans for postremission therapy rely heavily on cytogenetic analysis at diagnosis [*see Table 3*].

Cytogenetic data have been used to map chromosomal breakpoints at a molecular level, allowing the use of probes for fluorescence in situ hybridization (FISH) and of primers for reverse transcriptase polymerase chain reaction (RT-PCR) methods for the detection of tumor cells. FISH and RT-PCR can detect molecular genetic rearrangements not visible when examining chromosome bands by conventional methods. However, both of the newer methods test only for specific, defined genetic mutations and cannot be used initially for general screening or for a comprehensive evaluation. FISH analysis is more sensitive than conventional karyotype analysis and can be performed on both metaphase and interphase cells. The morphology of the positive cells can be determined concurrently, and the proportional involvement by leukemia of all of the hematopoietic cells can be evaluated.

RT-PCR is the most sensitive method for detecting occult leukemia cells (about 1 in 10^5 cells). Newer methods are quantitative, and the clinical significance of a positive result after treatment is being investigated. A positive assay confirms the presence of cells with the specific genetic abnormality but does not necessarily indicate the neoplastic growth potential of these cells. For example, a positive RT-PCR assay after treatment appears to predict leukemia relapse reliably in patients with APL and a t(15;17), but not in patients with AML-M2 and a t(8;21).[45,46]

TREATMENT

Remission Induction Therapy

The most common remission induction regimen used for patients with AML is cytarabine given by continuous intravenous infusion daily for 7 days plus daunorubicin given daily for 3 days (the 7 + 3 regimen). Depending on age and patient selection, 50% to 80% of patients achieve CR.[47,48] The outcome in general has not been improved by the substitution of other anthracyclines such as idarubicin or mitoxantrone for daunorubicin, by increasing the dose of cytarabine, or by adding a third or fourth cytotoxic drug.

Postremission Therapy

Additional chemotherapy after a successful remission induction is mandatory to cure AML. The median disease-free survival for patients who receive no additional therapy is only about 4 months. When several courses of consolidation chemotherapy are given, survival at 4 years is 40% to 50% for young and middle-aged adults. For patients younger than 60 years, consolidation therapy results in significantly longer survival than maintenance therapy alone.[47-49] For consolidation, the same induction chemotherapy may be repeated for one or more cycles, with or without dose intensification, or non–cross-resistant drugs can be used. There is increasing evidence that high-dose cytarabine (HiDAC) provides the best survival for good- and intermediate-prognosis patients.[50-52]

Maintenance therapy with relatively nonmyelosuppressive doses of cytotoxic drugs has no proven benefit in the management of AML. Currently, clinical trials are testing the benefit of immunomodulatory therapies such as interleukin-2 for AML patients in remission.

Most studies reporting on allogeneic or autologous SCT for AML patients in first CR are nonrandomized, and many are retrospective.[53] Considerable selection bias is generated by the delay between remission induction and transplantation and by the entry requirements for good performance status for most trials. Prospective, randomized studies comparing intensive consolidation chemotherapy with SCT have failed to show a clear survival advantage[54-57] [*see Table 7*].

Treatment of Patients with Relapsed or Refractory AML

A limited number of agents are effective in the treatment of AML, and management of patients with resistant or relapsed disease is difficult. Patients with long initial remissions (> 1 year) have a 50% to 60% reinduction rate with daunorubicin and cytarabine or with HiDAC, but the duration of the second remission is usually shorter than that of the first. Gemtuzumab ozogamicin, an anti-CD33 immunoconjugate, is active as a single agent in relapsed AML.[58] SCT should be considered for any patient who has undergone relapse after an intensive initial treatment program.

ACUTE PROMYELOCYTIC LEUKEMIA

APL (FAB M3) is a biologically distinct disease with characteristic clinical, morphologic, and cytogenetic features.[45,59,60] It constitutes about 8% to 15% of cases of AML. The malignant cells have a t(15;17) that juxtaposes the *RARα* gene on chromosome 17 with the *PML* gene on chromosome 15. Disseminated intravascular coagulation at presentation or soon after the initiation of cytotoxic chemotherapy can cause pulmonary or cerebrovascular hemorrhage in up to 40% of patients and is associated with a high mortality. The cytoplasmic granules in the leukemic blasts contain factors with procoagulant as well as fibrinolytic activity.[61]

All-*trans*-retinoic acid (ATRA, or tretinoin) was first used in the treatment of APL in China in 1986 and has proved to be a highly effective remission-induction agent.[59,60] ATRA accelerates the terminal differentiation of malignant promyelocytes to mature neutrophils, leading to apoptosis and CR without bone marrow hypoplasia. This effect is a unique consequence of the rearranged *PML/RARα* gene resulting from the t(15;17) that defines APL.[45] ATRA induction therapy produces CR rates of 80% to 95% in both previously untreated patients and patients in relapse. The best results have been obtained when ATRA was combined with daunorubicin and cytarabine or with idarubicin alone.[59,62] Arsenic trioxide is an effective treatment for patients with relapsed APL.[63]

ATRA does not have the usual toxicities associated with cytotoxic chemotherapy. Intrinsic drug toxicity is generally minor. ATRA is neither immunosuppressive nor myelosuppressive. However, two serious and specific complications may occur with ATRA treatment of APL.[64] In 25% to 40% of patients, the so-called retinoic acid syndrome develops within 2 to 21 days after initiation of treatment and is characterized by fever, peripheral edema, pulmonary infiltrates and respiratory distress, hypotension, renal and hepatic dysfunction, and serositis leading to pleural and pericardial effusions. The syndrome is possibly the result of the infiltration of tissue by maturing malignant promyelocytes and the systemic effects of cytokine release. Ear-

Table 7 Results from Randomized Trials of Stem Cell Transplantation and Intensive Chemotherapy for Acute Myeloid Leukemia in First Complete Remission*

Study	No. of Patients			Disease-Free Survival (%)				Overall Survival (%)				Treatment Mortality (%)		
	Allo	Auto	ICC	Allo	Auto	ICC	P	Allo	Auto	ICC	P	Allo	Auto	ICC
Zittoun et al[54]	168	128	126	55	48	30	< 0.05	59	56	46	NS	17	9	7
Harousseau et al[55]	88	86	78	44	44	40	NS	53	50	55	NS	22	7	3
Cassileth et al[56]	113	116	117	43	35	35	NS	46	43	52	< 0.05	25	14	3
Burnett et al[57]	—	190	191	—	54	40	< 0.05	—	57	45	NS	—	12	4
Mayer et al[52]	—	—	156	—	—	44	—	—	—	52	—	—	—	5

*Results are analyzed by intention to treat.
Allo—allogeneic stem cell transplantation Auto—autologous stem cell transplantation ICC—intensive consolidation chemotherapy using high-dose cytarabine

ly recognition and aggressive treatment with dexamethasone has been effective. Hyperleukocytosis occurs in up to 50% of patients treated with ATRA and is probably secondary to the induction of cellular maturation.

Management of the coagulopathy associated with APL may be difficult and should be done expectantly.[59] Coagulation parameters, including fibrinogen, D-dimer, and platelet levels, should be monitored closely. Platelet transfusions and cryoprecipitate or fresh frozen plasma are used to maintain the fibrinogen level above 100 mg/dl and the platelet count above 20,000/μl. The role of heparin is controversial; a continuous low-dose infusion (5 to 10 U/kg/hr) has been reported to be effective at stopping the consumption of clotting factors. Inhibitors of fibrinolysis should be considered only for life-threatening hemorrhage. The coagulopathy of APL typically improves rapidly after treatment with ATRA is begun.

MYELODYSPLASTIC SYNDROME

No specific, uniformly effective therapy currently exists for MDS. Clinical prognostic factors, including bone marrow morphology and cytogenetics, should be evaluated at diagnosis.[30] A period of observation with close follow-up of serial blood counts is recommended for initial management, and a second bone marrow examination after several months is useful to assess the rapidity of disease progression.

RA and RARS often follow a clinically indolent course; patients may need no treatment for variable lengths of time. However, patients who are severely anemic (hemoglobin < 7 to 8 g/dl) should receive red blood cell (RBC) transfusions. Washed RBC units are rarely necessary. Commonly, the recurrence of fatigue or angina is a more useful indicator of the need for the next RBC transfusion than any particular hemoglobin level. Transfusion requirements increase as MDS progress, and iron overload and hemochromatosis may occur.

Prophylactic platelet transfusions are not routinely administered to thrombocytopenic patients with MDS who are not bleeding, because of the expense, inconvenience, and risk of alloimmunization. This risk increases with the frequency of platelet transfusion and may limit future treatment options. Platelet support should be reserved for treatment of acute hemorrhage or for prophylaxis before surgery. MDS patients with thrombocytopenia should be carefully instructed to avoid aspirin-containing products and other nonsteroidal anti-inflammatory agents, which interfere with platelet function.

Infections are a common problem in patients with MDS and can be life-threatening. The utility of vaccines (pneumococcal, influenza, and hepatitis B) should not be overlooked. Fever, localized infections, or even general malaise should be carefully evaluated in patients who have both quantitative and functional neutropenia. Prompt institution of broad-spectrum antibiotic therapy while trying to locate the source of infection is critical; therapy should be continued for 7 to 10 days even in the absence of an obvious source of infection or positive blood cultures. If a neutropenic patient fails to respond to antibacterial agents or has repeated episodes of fever in the absence of an obvious source, a fungal infection should be suspected, and treatment with amphotericin B, fluconazole, or itraconazole should be initiated.

The role of cytotoxic drugs in the treatment of MDS is uncertain for the majority of patients but should be considered for younger and healthier patients with more aggressive forms of disease, such as those with RCMD, RAEB, and therapy-related MDS (t-MDS). In general, cytotoxic therapy with standard regimens used in the treatment of AML (e.g., containing daunorubicin and standard or high doses of cytarabine) has had limited success in extending survival for the majority of MDS patients. This poor outcome is not surprising, for a number of reasons: (1) most MDS patients are elderly and tolerate aggressive treatment poorly, (2) the number of normal hematopoietic stem cells available to regenerate the marrow after therapy-induced hypoplasia is probably reduced, and (3) the majority of effective antileukemia drugs may be relatively ineffective in the low proliferative states seen in MDS because these drugs act primarily during cell division. For these reasons, intensive chemotherapy has been associated with CR rates of only 10% to 40%, with duration of remissions generally lasting less than 1 year. The toxicity of the therapy and the prolonged marrow hypoplasia and cytopenia that it produces may actually shorten the lives of older patients. It is therefore important to direct this treatment option primarily to younger MDS patients, who are better able to tolerate it.

Allogeneic SCT has been tested as curative therapy in the small group of MDS patients who are younger than 60 years (approximately 10% to 20% of all MDS patients) and for whom a histocompatible donor is available.[65] Approximately 45% of patients who undergo transplantation have disease-free survival. Patients with favorable or intermediate cytogenetic factors have better outcomes than those with poor risk factors. These data suggest that appropriately chosen patients can probably be cured with SCT.

Because the vast majority of patients with MDS are not candidates for SCT and because the outcome with conventional cytotoxic therapy has generally been poor, attention has been focused on a variety of agents that can induce differentiation of acute leukemia cell lines in vitro, slowing proliferation and restoring apparently normal maturation and function. Several such so-called differentiating agents (e.g., low-dose cytarabine, azacitidine, thalidomide, amifostine, and growth factors) have been tested in clinical trials of MDS.[66-69] Clinical trials have not demonstrated any benefit from the long-term administration of myeloid colony-stimulating factors (such as granulocyte colony-stimulating factor [G-CSF] or granulocyte-macrophage CSF [GM-CSF]), although they may have a role as a short-term adjunct with antibiotics in the treatment of infections. Erythropoietin therapy has been beneficial only for anemic patients with inappropriately low endogenous erythropoietin levels in the blood.

TREATMENT OF OLDER PATIENTS WITH ACUTE MYELOID LEUKEMIA

Two factors combine to explain in large part the poor outcome of older patients with leukemia.[70] Most evident is the inability of many of these patients to withstand the rigors of intensive chemotherapy and its expected complications. Patients with age-related chronic cardiac, pulmonary, or renal disorders suffer greater acute toxicity from chemotherapy. Older patients may also have less bone marrow regenerative capacity, even after successful leukemia cytoreduction. Inability to tolerate long periods of pancytopenia and malnutrition and the nephrotoxicity of aminoglycosides or amphotericin remain major barriers to successful treatment.

Less obvious is the fact that the genetic mutations most often associated with treatment failure in young patients (e.g., abnormalities of chromosome 5 or 7 in AML or t(9;22) in ALL) are more common in older patients. Conversely, all of the so-called favorable cytogenetic abnormalities, such as t(8;21), t(15;17), and inv(16) in AML, are more common in younger adults and are responsible in part for the better disease-free survival of young and middle-aged adults.[15,37] The multidrug-resistant phenotype that results in part from the overexpression of the P-glycoprotein responsible for drug efflux from AML cells is also more common in older patients.[71]

MDS is more resistant to therapy and is also more common in older patients. Many cases of AML with dysplasia in older patients have presumably evolved through a myelodysplastic phase. MDS is characterized by the stepwise accumulation of genetic abnormalities, analogous to the evolution of new chromosomal abnormalities that occurs as chronic myeloid leukemia accelerates into the terminal blast phase. The multidrug-resistant phenotype may also emerge during this evolutionary process. At the same time, normal hematopoiesis is increasingly inhibited, and the normal stem cell compartment may be lost. The net result is ineffective hematopoiesis and dysfunctional blood cells. By the time AML emerges, these patients often have infections from pathogenic flora; in addition, these patients often experience recurrent bleeding episodes, and they are often dependent on transfusions. The outcome after chemotherapy in these patients has been poor.

Primum Non Nocere

Not every patient benefits from intensive chemotherapy; this is particularly true of older patients. Well-meaning attempts to induce remission may actually shorten survival. Patients unlikely to survive treatment can be identified by their poor performance status or comorbid disorders. Case series from large referral institutions suggest that 25% to 50% of AML patients older than 60 years are not offered remission-induction chemotherapy.

In a few older patients with acute leukemia (determined by the usual quantitative criteria of > 20% bone marrow blast cells), the disease has a much more smoldering course, especially when the patients have had MDS. These patients suffer from bone marrow failure and pancytopenia more than from hyperleukocytosis. Their survival may be equally long and their quality of life better if transfusion support and antibiotics are used rather than intensive chemotherapy. This finding may be particularly true for patients with hypoplastic AML.

Otherwise healthy, older patients with acute leukemia, especially those with favorable cytogenetic features, should be offered curative chemotherapy. With a standard 7 + 3 regimen of cytarabine and an anthracycline, approximately half of patients older than 60 years will achieve a CR. Unfortunately, even with postremission consolidation chemotherapy, the overall survival for this group is less than 10% after 4 years.[52,70,72]

Hematopoietic Growth Factors

Several large controlled trials that studied the use of GM-CSF or G-CSF mostly in older patients with AML demonstrated that the duration of neutropenia was only minimally decreased when growth factors were given after remission-induction chemotherapy.[72,73] Even though a more rapid recovery of neutrophils has been observed in some trials, the nadir has not been affected, and thus, the incidence of severe infection remains high. At the same time, stimulation of leukemia regrowth by myeloid growth factors appears to be uncommon in vivo. There may be greater benefit from using growth factors after consolidation chemotherapy, when patients are already in remission, than there is from using growth factors earlier, during remission induction.[74] As yet, growth factors have not had a marked impact on survival or remission duration for patients with AML.

Acute Lymphoblastic Leukemia

ALL accounts for approximately 20% of adult acute leukemias. In contrast, ALL is by far the most common malignant disease in childhood.

DIAGNOSIS

Clinical Features

The clinical presentation in adults is most often acute. Symptoms are usually present for only a few weeks before diagnosis. Malaise, lethargy, weight loss, fevers, and night sweats may be present but are typically not severe. Bone pain and arthralgias occur occasionally but much less frequently in adults than in children. Infection and hemorrhage are present in one third of patients at diagnosis but are most often not as severe as in AML. Lymphadenopathy, splenomegaly, and hepatomegaly are more common than in AML, affecting half of adults with ALL. Chest radiographs may reveal a thymic mass in 10% to 15% of adults. The majority of these patients have T cell ALL.

CNS involvement by leukemia occurs in 5% to 10% of adult cases but is uncommon at diagnosis. Cranial nerve palsies most often involve the sixth and seventh cranial nerves. Headache and papilledema may be present, resulting from meningeal in-

filtration and obstruction of the outflow of CSF, leading to increased intracranial pressure. Retinal hemorrhages may be the result of thrombocytopenia.

Varying degrees of neutropenia, anemia, and thrombocytopenia are detected on the peripheral blood examination. In one series of over 1,200 adult cases, the granulocyte count was below 1,500/μl in only one fifth of patients.[75] Mild to moderate reductions in hemoglobin level were typical, but in almost one third of patients, the hemoglobin level was below 8 g/dl. Thrombocytopenia was frequent, and in over one half of patients, the platelet count was below 50,000/μl. The total leukocyte count was diminished in about one third of patients and normal or moderately elevated in close to one half. Characteristic lymphoblasts can be identified in the peripheral blood in over 90% of cases. Marked leukocytosis (> 100,000/μl) was present at diagnosis in 16% of cases, but symptomatic leukostasis is uncommon in ALL, even at these levels.

Cytogenetic Abnormalities

It is now possible to characterize subsets of ALL by distinctive genetic features. These more homogeneous subgroups of ALL have shown important clinical differences in response to specific treatment regimens.[39,76] Subgroups include precursor B cell and Burkitt cell leukemia and precursor T cell ALL cases [*see Table 5*]. However, the precursor B cell group of cases can be subdivided according to cytogenetic and molecular features. Two important subgroups are associated with a more favorable outcome. Hyperdiploid ALL with less than 50 chromosomes accounts for 25% of ALL in childhood; these cases have an 80% to 90% cure rate. Similarly, in another 20% to 25% of childhood ALL cases, patients have the *TEL/AML1* fusion gene (also called *ETV6/CBFA2*), which results from the cryptic (12;21) translocation; these cases also have an 85% to 90% cure rate. Unfortunately, these genetic subtypes are rarely seen in adult series. Instead, the single largest subgroup of adult patients with ALL are those with the *BCR-ABL* fusion gene, which results from the (9;22) translocation. The resulting subtype of ALL, also known as Philadelphia chromosome–positive (Ph^+) ALL, makes up 25% to 30% of cases of adult ALL and may not be curable with conventional chemotherapy alone. In contrast, patients with a translocation involving chromosome band 14q11 or overexpression of the *HOX11* gene usually have T cell ALL and have a good prognosis with conventional multiagent regimens. Patients with Burkitt cell ALL usually have an 8;14 translocation (*MYC/IGH* fusion gene) and also have a good prognosis, but they require quite different chemotherapy. The other recurring chromosomal abnormalities and fusion genes [*see Table 3*] are less common but again appear to identify patients that may have very different prognoses.

It is important to note that the considerable differences in long-term outcome between these various cytogenetic subgroups are not so apparent when examining only the initial CR rate. That is, with current intensive induction programs, 80% of the patients with poor-risk cytogenetics nevertheless achieve a remission comparable to 90% of those with normal karyotypes or other abnormalities. The clinical importance of this fact is that these poor-risk patients respond sufficiently well to initial treatment to become candidates for an early allogeneic SCT.

TREATMENT

The aims of modern ALL treatment regimens are the rapid restoration of bone marrow function by use of multiple chemotherapy drugs at acceptable toxicities to prevent the emergence of resistant subclones; the use of adequate prophylactic treatment of sanctuary sites, such as the CNS; and postremission consolidation therapies to eliminate minimal (undetectable) residual disease. Postremission therapy has traditionally been categorized as intensification or consolidation treatment followed by prolonged maintenance.

Four or five drugs (e.g., vincristine, prednisone, daunorubicin, L-asparaginase, and cyclophosphamide) are typically used for remission induction, followed by similar agents plus antimetabolites for remission consolidation treatment.[36,76-79] The vast majority of adults (80% to 90%) do enter CR. Approximately 30% to 50% will be alive and free of disease 3 years later.

Important adverse prognostic factors in ALL have a major influence on complete remission rates and on remission duration and survival.[36,76,77] In multivariate analyses, patients presenting with leukocyte counts greater than 30,000/μl have had significantly shorter durations of remission than patients with lower leukocyte counts. However, extreme leukocytosis does not negatively affect outcome in patients with T cell ALL.[36] Older age (> 60 years) is another adverse characteristic. Karyotype at diagnosis is an independent predictor of outcome and the key to successful treatment planning.[39,77] Minor factors, or factors that have had some significance with certain treatment regimens, are the percentage of circulating blast cells; the degree of bone marrow involvement; the presence of hepatomegaly, splenomegaly, or lymphadenopathy; LDH levels; CNS involvement at presentation; and the time required to achieve complete remission (e.g., > 4 to 6 weeks).

Remission Consolidation Treatment

Remission consolidation treatment is designed to eradicate the rapidly proliferating neoplastic cells that are thought to be responsible for early relapses. In general, drugs given during this period are cell-cycle phase–specific antimetabolites. Unlike remission-induction therapy, whose importance is securely established, intensive consolidation therapy is not indisputably needed for achieving cure. The relative benefit of any particular consolidation therapy is likely to be inversely proportional to the intensity of the initial induction therapy and its efficacy in rapidly reducing the leukemia cell mass. Transplantation has not proved to be better than intensive chemotherapy for ALL in first CR, in general, but it clearly leads to cures in patients with chemotherapy-resistant subtypes of ALL.[80-82]

Maintenance Therapy

A prolonged period of treatment with low doses of chemotherapy drugs, called remission-maintenance or continuation therapy, is still standard in ALL. This approach stands in marked contrast to treatment of most other so-called curable cancers, such as Hodgkin disease, large cell lymphoma, or testicular cancer, in which cure follows the initial intensive cytoreductive therapy and low-dose maintenance chemotherapy provides no additional benefit. The necessity of prolonged maintenance therapy for adults with ALL may also be a function of the intensity and the success of initial chemotherapy. Traditionally, 1 to 3 years of treatment with 6-mercaptopurine and methotrexate, often with monthly pulses of vincristine and prednisone, have been given. As yet, the need for maintenance therapy has not been proved in adults, but it is likely to be important at least for some types of ALL.

CNS Prophylaxis

The CNS is an important site of involvement by ALL.[83] Although not commonly found at diagnosis, CNS involvement is common at the time of relapse. The meninges may harbor leukemia cells, and the blood-brain barrier may shelter them from systemic chemotherapy. Recurrence within the CNS usually coincides with systemic relapse. The likelihood of an isolated CNS relapse for adults with ALL appears to be about 5%. Preventive treatment of the CNS during postremission therapy, termed CNS prophylaxis, has become an integral part of virtually all current adult ALL treatment protocols. Although the true value of CNS prophylaxis in adults is controversial, studies have shown that adult patients who either refused or could not receive CNS prophylaxis had a higher rate of CNS relapse than patients receiving prophylaxis. CNS leukemia is more easily prevented than treated; once overt CNS leukemia has developed, there is a high likelihood of subsequent CNS relapse despite treatment.

CNS prophylaxis typically consists of cranial irradiation plus intrathecal methotrexate. Cytarabine and hydrocortisone are sometimes added to methotrexate for triple intrathecal therapy. In lieu of cranial irradiation, some investigators have substituted high-dose systemic chemotherapy with methotrexate or cytarabine because therapeutic levels of these drugs can be achieved in the CSF when the drugs are administered intravenously in high doses.[79] Overall, the superiority of any one prophylactic therapy has not been established.

BURKITT CELL ALL

The first of the high-risk subsets that warrants special attention is Burkitt cell ALL, also known as FAB L3 or mature B cell ALL. This subset makes up 3% to 5% of adult ALL cases. The ubiquitous biologic features are the presence of monoclonal surface immunoglobulin, the 8;14 translocation or one of its two variants—t(2;8) or t(8;22)—and the constitutive expression of the oncogene *MYC*. It is relatively easily recognized at diagnosis from the characteristic clinical findings of hepatosplenomegaly and lymphadenopathy. The LDH and uric acid levels are usually markedly elevated, and there is often leptomeningeal involvement. The lymphoblasts usually lack TdT reactivity. In the past, few if any of these patients survived after standard ALL treatment regimens. Now, however, short, intensive chemotherapy programs for B cell ALL yield a high CR rate and a survival plateau in the range of 50%.[84] These regimens, which may require as few as 16 to 18 weeks of treatment, use high doses of methotrexate, cytarabine, and cyclophosphamide or ifosfamide together with other ALL drugs.

PHILADELPHIA CHROMOSOME–POSITIVE ALL

Ph^+ ALL is identified by the t(9;22)(q34;q22) or the *BCR-ABL* fusion gene. It is currently the major challenge in curing ALL because it makes up 25% to 30% of all adult cases and perhaps one half of all B-lineage ALL.[39,85] Some progress has been made. Approximately 70% of patients achieve a CR, but the remission durations are markedly shorter (median, 7 months) for Ph^+ cases than for those without a Ph chromosome (remission of almost 3 years).[36,39,79,82,86] As yet, no chemotherapy regimen alone appears to have the potential to cure this group of patients. In contrast, allogeneic SCT cures about one third of patients with Ph^+ ALL.[80,82] The probability of relapse after transplantation is approximately 30% to 50%, further attesting to the therapy-resistant nature of this disease.

Table 8 Treatment Outcome for Adults with Acute Leukemia

Disease Subset	*Complete Remission (%)*	*Disease-Free Survival (%)*
AML, overall	60–70	20–40
Favorable karyotype	90	60–70
APL, t(15;17)	90	70–75
< 60 yr	70–80	40
≥ 60 yr	45–55	5–10
ALL, overall	80–90	30–40
T-ALL	90–95	60–65
Precursor B cell ALL	75–85	30–40
Ph^+ ALL	70–75	0–10
Burkitt cell ALL	75	70
≥ 60 yr	75–80	10–25

ALL—acute lymphoblastic leukemia AML—acute myeloid leukemia APL—acute promyelocytic leukemia Ph^+—Philadelphia chromosome–positive T-ALL—T cell acute lymphoblastic leukemia

Thus, at this time, the treatment for Ph^+ ALL should include an intensive remission-induction chemotherapy program, followed by allogeneic SCT in the first CR if a donor is available. Considerable interest exists in investigating new agents, especially the tyrosine kinase inhibitor imatinib mesylate, in this high-risk group of patients.[87]

ALL patients with a t(4;11)(q21;q23) are another cytogenetic subset that has had a poor outcome after conventional therapy and should undergo allogeneic transplantation, if possible. The involved gene on chromosome 11 was named *MLL*, for mixed-lineage leukemia.[15] The incidence is 3% to 6% in adult cytogenetic series.[39] The immunophenotype is progenitor B cell (CD19, CD22, and HLA-DR positive and CD10 negative).[88] In adults, these patients are often older women (more than half are older than 50 years) with high leukocyte counts.

Despite major advances in the treatment of adults with acute leukemia in the past decade, many patients continue to die of their disease or of complications of its treatment [*see Table 8*]. However, a number of novel experimental and clinical approaches hold promise for improving cure rates. Application of modern molecular technologies designed to detect minimal residual leukemia may aid clinicians in monitoring disease during and after chemotherapy.[89] It could, for example, lead to the early detection of patients likely to experience relapse, for whom further therapy may be necessary. Novel methods of circumventing multidrug resistance, exploiting immune mechanisms, or altering the control of malignant cell growth need to be investigated.

The author has received grants for clinical research or educational activities or has served as advisor or consultant during the past year for Novartis AG, Sanofi-Synthelabo Inc., and Wyeth.

References

1. Lowenberg B, Downing JR, Burnett A: Medical progress: acute myeloid leukemia. N Engl J Med 341:1051, 1999

2. Heaney ML, Golde DW: Medical progress: myelodysplasia. N Engl J Med 340:1649, 1999

3. Linet MS: The Leukemias: Epidemiologic Aspects. Oxford University Press, New York, 1985

4. Sandler DP, Ross JA: Epidemiology of acute leukemia in children and adults. Semin Oncol 24:3, 1997

5. Brill AB, Tomonoga M, Heyssel RM: Leukemia in man following exposure to ionizing irradiation: summary of findings in Hiroshima and Nagasaki and comparison to other human experience. Ann Intern Med 56:590, 1962

6. Kato H, Schull WJ: Studies of the mortality of A-bomb survivors. 7. Mortality, 1950–1978: Part I. Cancer mortality. Radiat Res 90:395, 1982

7. Smith PG, Doll R: Mortality among patients with ankylosing spondylitis after a single treatment course with x-rays. Br Med J 284:449, 1982

8. Modan B, Lilienfeld AM: Polycythemia vera and leukemia—the role of radiation treatment. Medicine (Baltimore) 44:305, 1965

9. Thirman MJ, Larson RA: Therapy-related myeloid leukemia. Hematol/Oncol Clin North Am 10:293, 1996

10. Graham S, Levin ML, Lilienfeld AM, et al: Preconception, intrauterine, and postnatal irradiation as related to leukemia. Monogr Natl Cancer Inst 19:347, 1966

11. Polednak AP, Stehney AF, Rowland RE: Mortality among women first employed before 1930 in the U.S. radium dial painting industry. Am J Epidemiol 107:179, 1978

12. March HC: Leukemia in radiologists. Radiology 43:275, 1944

13. Tucker MA, Coleman CN, Cox RS, et al: Risk of second cancers after treatment for Hodgkin's disease. N Engl J Med 318:76, 1988

14. Pui CH, Ribeiro RC, Hancock ML, et al: Acute myeloid leukemia in children treated with epipodophyllotoxins for acute lymphoblastic leukemia. N Engl J Med 325:1682, 1991

15. Le Beau MM, Larson RA: Cytogenetics and neoplasia. Hematology: Basic Principles and Practice, 3rd ed. Hoffman R, Benz EJ Jr, Shattil SJ, et al, Eds. Churchill Livingstone, New York, 2000, p 848

16. Aksoy M, Erdem S, Din Col G: Leukemia in shoeworkers chronically exposed to benzene. Blood 44:837, 1974

17. Hayes RB, Yin SN, Dosemeci M, et al: Benzene and the dose-related incidence of hematologic neoplasms in China. J Natl Cancer Inst 89:1065, 1997

18. Robert-Guroff M, Nakao Y, Natake K, et al: Natural antibodies to human retrovirus HTLV in a cluster of Japanese patients with adult T-cell leukemia. Science 215:975, 1982

19. Miller RW: Deaths from childhood leukemia and solid tumors among twins and other sibs in the United States, 1960–67. J Natl Cancer Inst 46:203, 1971

20. Draper GF, Heaf MM, Kinnier Wilson LM: Occurrence of childhood cancers among sibs and estimation of familial risks. J Med Genet 14:81, 1977

21. Lange BJ, Kobrinsky N, Barnard DR, et al: Distinctive demography, biology, and outcome of acute myeloid leukemia and myelodysplastic syndrome in children with Down's syndrome: Children's Cancer Group studies 2861 and 2891. Blood 91:608, 1998

22. Bennett JM, Catovsky D, Daniel MT, et al: Proposals for the classification of the acute leukemias. Br J Haematol 33:451, 1976

23. Bennett JM, Catovsky D, Daniel MT, et al: Proposed revised criteria for the classification of acute myeloid leukemia: a report of the French-American-British cooperative group. Ann Intern Med 103:620, 1985

24. Jaffe ES, Harris NL, Stein H, et al: World Health Organization Classification of Tumours: Pathology and Genetics: Tumours of Haematopoietic and Lymphoid Tissues. IARC Press, Lyon, 2001

25. Bennett JM, Catovsky D, Daniel MT, et al: Proposal for the recognition of minimally differentiated acute myeloid leukemia (AML-M0). Br J Haematol 78:325, 1991

26. Bennett JM, Catovsky D, Daniel MT, et al: Proposals for the classification of the myelodysplastic syndromes. Br J Haematol 51:189, 1982

27. Rosati S, Mick R, Xu F, et al: Refractory anemia with multilineage dysplasia: further characterization of an "unclassifiable" myelodysplastic syndrome. Leukemia 10:20, 1996

28. The Myelodysplastic Syndromes: Pathobiology and Clinical Management. Bennett JM, Ed. Marcel Dekker, New York, 2002

29. List AF, Garewal HS, Sandberg AA: The myelodysplastic syndromes: biology and implications for management. J Clin Oncol 8:1424, 1990

30. Greenberg P, Cox C, Le Beau MM, et al: International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 89:2079, 1996

31. Bennett JM, Catovsky D, Daniel MT, et al: The morphologic classification of acute lymphoblastic leukemia: concordance among observers and clinical correlations. Br J Haematol 47:553, 1981

32. Terstappen LWMM: Cell differentiation and maturation in normal bone marrow and acute leukemia. Flow Cytometry Clinical Applications. Macey MG, Ed. Blackwell Scientific Publications, Oxford, England, 1995, p 101

33. Ball ED, Davis RB, Griffin JD, et al: Prognostic value of lymphocyte surface markers in acute myeloid leukemia. Blood 77:2242, 1991

34. Geller RB, Zahurak M, Hurwitz CA, et al: Prognostic importance of immunophenotyping in adults with acute myelocytic leukemia: the significance of the stem-cell glycoprotein CD34 (My10). Br J Haematol 76:340, 1990

35. Czuczman MS, Dodge RK, Stewart CC, et al: Value of immunophenotype in intensively treated adult acute lymphoblastic leukemia: Cancer and Leukemia Group B Study 8364. Blood 93:3931, 1999

36. Larson RA, Dodge RK, Burns CP, et al: A five-drug remission induction regimen with intensive consolidation for adults with acute lymphoblastic leukemia: Cancer and Leukemia Group B Study 8811. Blood 85:2025, 1995

37. Byrd JC, Mrozek K, Dodge RK, et al: Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood 100:4325, 2002

38. Mrozek K, Heinonen K, de la Chapelle A, et al: Clinical significance of cytogenetics in acute myeloid leukemia. Semin Oncol 24:17, 1997

39. Wetzler M, Dodge RK, Mrozek K, et al: Prospective karyotype analysis in adult acute leukemia: the Cancer and Leukemia Group B experience. Blood 93:3983, 1999

40. Caligiuri MA, Strout MP, Gilliland DG: Molecular biology of acute myeloid leukemia. Semin Oncol 24:32, 1997

41. Thandla S, Aplan PD: Molecular biology of acute lymphocytic leukemia. Semin Oncol 24:45, 1997

42. Appelbaum FR: Allogeneic hematopoietic stem cell transplantation for acute leukemia. Semin Oncol 24:114, 1997

43. Linker CA, Damon LE, Reis CA, et al: Autologous stem cell transplantation for advanced acute myeloid leukemia. Bone Marrow Transplant 29:297, 2002

44. Bensinger WI, Martin PJ, Storer B, et al: Transplantation of bone marrow as compared with peripheral blood cells from HLA-identical relatives in patients with hematologic cancers. N Engl J Med 344:175, 2001

45. Lo Coco F, Diverio D, Falini B, et al: Genetic diagnosis and molecular monitoring in the management of acute promyelocytic leukemia. Blood 94:12, 1999

46. Nucifora G, Larson RA, Rowley JD: Persistence of the 8;21 translocation in patients with acute myeloid leukemia type M2 in long-term remission. Blood 82:712, 1993

47. Burnett AK: Acute myeloid leukemia: treatment of adults under 60 years. Rev Clin Exp Hematol 6:26, 2002

48. Stone RM: Postremission therapy in adults with acute myeloid leukemia. Semin Hematol 38:17, 2001

49. Cassileth PA, Lynch E, Hines JD, et al: Varying intensity of postremission therapy in acute myeloid leukemia. Blood 79:1924, 1992

50. Buchner T, Hiddemann W, Wormann B, et al: Double induction strategy for acute myeloid leukemia: the effect of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanine: a randomized trial by the German AML Cooperative Group. Blood 93:4116, 1999

51. Bishop JF, Matthews JP, Young GA, et al: A randomized study of high dose cytarabine in induction in acute myeloid leukemia. Blood 87:1710, 1996

52. Mayer RJ, Davis RB, Schiffer CA, et al: Intensive postremission chemotherapy in adults with acute myeloid leukemia. N Engl J Med 331:896, 1994

53. Edenfield WJ, Gore SD: Stage-specific application of allogeneic and autologous marrow transplantation in the management of acute myeloid leukemia. Semin Oncol 26:21, 1999

54. Zittoun RA, Mandelli F, Willemze R, et al: Autologous or allogeneic bone marrow transplantation compared with intensive chemotherapy in acute myelogenous leukemia. N Engl J Med 332:217, 1995

55. Harousseau JL, Cahn JY, Pignon B, et al: Comparison of autologous bone marrow transplantation and intensive chemotherapy as postremission therapy in adult acute myeloid leukemia. Blood 90:2978, 1997

56. Cassileth PA, Harrington DP, Appelbaum FR, et al: Chemotherapy compared with autologous or allogeneic bone marrow transplantation in the management of acute myeloid leukemia in first remission. N Engl J Med 339:1649, 1998

57. Burnett AK, Goldstone AH, Stevens RMF, et al: Randomised comparison of addition of autologous bone-marrow transplantation to intensive chemotherapy for acute myeloid leukaemia in first remission: results of the MRC AML10 trial. Lancet 351:700, 1998

58. Sievers EL, Larson RA, Stadtmauer EA, et al: Efficacy and safety of gemtuzumab ozogamicin in patients with CD33-postive acute myeloid leukemia in first relapse. J Clin Oncol 19:3244, 2001

59. Fenaux P, Chomienne C, Degos L: Acute promyelocytic leukemia: biology and treatment. Semin Oncol 24:92, 1997

60. Warrell RP, De The H, Wang Z, et al: Acute promyelocytic leukemia. N Engl J Med 329:177, 1993

61. Menell JS, Cesarman GM, Jacovina AT, et al: Annexin II and bleeding in acute promyelocytic leukemia. N Engl J Med 340:994, 1999

62. Tallman MS, Nabhan C, Feusner JH, et al: Acute promyelocytic leukemia: evolving therapeutic strategies. Blood 99:759, 2002

63. Soignet SL, Frankel SR, Douer D, et al: United States multicenter study of arsenic trioxide in relapsed acute promyelocytic leukemia. J Clin Oncol 19:3852, 2001

64. Vahdat L, Maslak P, Miller WH Jr, et al: Early mortality and the retinoic acid syndrome in acute promyelocytic leukemia: impact of leukocytosis, low-dose chemotherapy, PML/RAR-α isoform, and CD13 expression in patients treated with all-*trans* retinoic acid. Blood 84:3843, 1994

65. de Witte T, Suciu S, Verhoef G, et al: Intensive chemotherapy followed by allogeneic or autologous stem cell transplantation for patients with myelodysplastic syndromes and acute myeloid leukemia following MDS. Blood 98:2326, 2001

66. Cheson BD, Zwiebel JA, Dancey J, et al: Novel therapeutic agents for the treatment of myelodysplastic syndromes. Semin Oncol 27:560, 2000

67. Silverman LR, Demakos EP, Peterson BL, et al: Randomized controlled trial of azacitidine in patients with the myelodysplastic syndrome: a study of the Cancer and Leukemia Group B. J Clin Oncol 20:2429, 2002

68. Raza A, Meyer P, Dutt D, et al: Thalidomide produces transfusion independence in long-standing refractory anemias of patients with myelodysplastic syndromes. Blood 98:958, 2001

69. List AF: Use of amifostine in hematologic malignancies, myelodysplastic syndrome, and acute leukemia. Semin Oncol 26:61, 1999

70. Sekeres MA, Stone RM: Older adults with acute myeloid leukemia. Curr Oncol Rep 4:403, 2002

71. Leith CP, Kopecky KJ, Chen IM, et al: Frequency and clinical significance of the expression of the multidrug resistance proteins MDR1/P-glycoprotein, MRP1, and LRP in acute myeloid leukemia: a Southwest Oncology Group Study. Blood 94:1086, 1999

72. Stone RM, Berg DT, George SL, et al: Granulocyte-macrophage colony stimulating factor following initial chemotherapy for patients at least 60 years old with primary acute myeloid leukemia: a double-blinded, randomized study. N Engl J Med 332:1671, 1995

73. Schiffer CA: Hematopoietic growth factors as adjuncts to the treatment of acute myeloid leukemia. Blood 88:3675, 1996

74. Heil G, Hoelzer D, Sanz MA et al: A randomized, double-blind, placebo-controlled, phase III study of filgrastim in remission induction and consolidation therapy for adults with de novo acute myeloid leukemia. Blood 90:4710, 1997

75. Hoelzer DF: Diagnosis and treatment of adult acute lymphocytic leukemia. Neoplastic Diseases of the Blood, 3rd ed. Wiernik PH, Canellos GP, Kyle RA, et al, Eds. Churchill Livingstone, New York, 1996, p 295

76. Laport GF, Larson RA: Treatment of adult acute lymphoblastic leukemia. Semin Oncol 24:70, 1997

77. Pui CH, Evans WE: Acute lymphocytic leukemia. N Engl J Med 339:605, 1998

78. Linker C, Damon L, Ries C, et al: Intensified and shortened cyclical chemotherapy for adult acute lymphoblastic leukemia. J Clin Oncol 20:2464, 2002

79. Kantarjian HM, O'Brien S, Smith TL, et al: Results of the HYPER-CVAD program, a dose intensive regimen, in adult acute lymphocytic leukemia. J Clin Oncol 18:543, 2000

80. Finiewicz KJ, Larson RA: Dose-intensive therapy for adult acute lymphoblastic leukemia. Semin Oncol 26:6, 1999

81. Thiebaut A, Vernant JP, Degos L, et al: Adult acute lymphoblastic leukemia study testing chemotherapy and autologous and allogeneic transplantation: a follow-up report of the French protocol LALA 87. Hematol Oncol Clin North Am 14:1353, 2000

82. Arico M, Valsecchi MG, Camitta B, et al: Outcome of treatment in children with Philadelphia chromosome–positive acute lymphoblastic leukemia. N Engl J Med 342:998, 2000

83. Cataland SR, Larson RA: Treatment of adult acute lymphoblastic leukemia: Perspective 1. Treatment of Acute Leukemias: New Directions for Clinical Research, Pui C-H, Ed. Humana Press, Totowa, NJ, 2003, p 131

84. Hoelzer D, Ludwig WD, Thiel D, et al: Improved outcome in adult B-cell acute lymphoblastic leukemia. Blood 87:495, 1996

85. Gleissner B, Gokbuget N, Bartram CR, et al: Leading prognostic relevance of the BCR-ABL translocation in adult acute B-lineage lymphoblastic leukemia: a prospective study of the German Multicenter Trial Group and confirmed polymerase chain reaction analysis. Blood 99:1536, 2002

86. Larson RA, Dodge RK, Linker CA, et al: A randomized controlled trial of filgrastim during remission induction and consolidation chemotherapy for adults with acute lymphoblastic leukemia: CALGB study 9111. Blood 92:1556, 1999

87. Ottmann OG, Druker BJ, Sawyers CL, et al: A phase II study of imatinib mesylate (Glivec) in patients with relapsed or refractory Philadelphia chromosome-positive acute lymphoid leukemias. Blood 100:1965, 2002

88. Ludwig W-D, Rieder H, Bartram CR, et al: Immunophenotypic and genotypic features, clinical characteristics, and treatment outcome of adult pro-B acute lymphoblastic leukemia: results of the German multicenter trials GMALL 03/87 and 04/89. Blood 92:1898, 1998

89. Stock W, Estrov Z: Studies of minimal residual disease in acute lymphoblastic leukemia. Hematol Oncol Clin North Am 14:1289, 2000

XVII CHRONIC MYELOGENOUS LEUKEMIA AND OTHER MYELOPROLIFERATIVE DISORDERS

Stefan Faderl, m.d.
Hagop M. Kantarjian, m.d.

The myeloproliferative disorders (MPDs) represent clonal disorders of the pluripotential hematopoietic stem cell and include chronic myelogenous leukemia (CML), polycythemia vera (PV), essential thrombocythemia (ET), myeloid metaplasia, and idiopathic myelofibrosis (IM). Patients with hybrid, or mixed, MPD may sometimes present with features of several of these entities (e.g., elevated hemoglobin and platelet counts), or features may shift from one to another.

Although each of these conditions represents a predominant expansion of one cell type (such as neutrophils in CML, erythrocytes in PV, and megakaryocytes in ET), cytogenetic studies have demonstrated that these cells are all clonally identifiable and are therefore derived from the same parental cell. Although myelofibrosis is recognized as a primary disorder, it frequently accompanies the other MPDs to various degrees. In all these circumstances, the bone marrow fibrosis is polyclonal, suggesting that this phenomenon is secondary to the activity of an abnormal stem cell clone, a theory supported by the fact that myelofibrosis can disappear if the abnormal stem cell clone is ablated. Clonal evolution follows at various frequencies, occurring in 5% to 90% of patients; it results in cell transformation and disease progression that frequently is manifested as acute leukemia–type clinical syndromes or so-called burnt-out fibrotic bone marrows.

Although the MPDs have similarities in the proliferation of cell lineages, peculiar molecular-pathogenetic differences distinguish the MPDs from one another in clinical presentation and clinical course, rate of transformation to acute leukemia, choice and response to therapy, and prognosis [*see Table 1*]. The molecular-cytogenetic pathophysiology has been extensively described in CML but is less well understood in other MPDs. Unraveling the molecular events involved in the initiation and transformation of CML has provided novel therapies that have changed the natural history of CML and made the disease a paradigm for our understanding of leukemogenesis and drug development.

Chronic Myelogenous Leukemia

EPIDEMIOLOGY

An estimated 4,300 patients are diagnosed with CML each year in the United States; CML accounts for 14% of all cases of leukemia in adults.[1] CML occurs at an incidence of 1 to 2 per 100,000 population; this incidence is remarkably constant worldwide.[1] Males are affected more often than females. The median age at presentation is 45 to 55 years. A third of the patients are older than 60 years. Age is important when considering therapeutic options such as stem cell transplantation (SCT), which is associated with a high procedure-related mortality; interferon alfa, which has more side effects in older patients than in younger patients; and, more recently, imatinib, which is active and well tolerated even in older patients. CML is uncommon in children and adolescents, accounting for less than 5% of the leukemias.[2]

PATHOPHYSIOLOGY AND CYTOGENETICS

CML is caused by the transforming capability of the protein products resulting from the Philadelphia translocation, t(9;22)(q34;q11), and cytogenetic-molecular changes acquired

Table 1 Distinguishing Characteristics of Myeloproliferative Disorders

Disorder	*Incidence*	*Age at Onset (years)*	*Cytogenetic Molecular Markers*	*Risk of Leukemic Transformation*	*Treatment Options*
CML	1–2 per 100,000 population	45–55	Philadelphia translocation (9;22)(q34;ql1) in up to 95% of patients	Invariable transformation to accelerated and blastic phases (85%–90%)	Stem cell transplantation, imatinib mesylate (single agent, high dose in combination with other agents [e.g., interferon alfa]), other targeted therapies (e.g., FTI, SRCI, decitabine, homoharringtonine), phlebotomy
PV	0.5–1.0 per 100,000 population	60	Cytogenetic abnormalities in up to 20% of patients at diagnosis (trisomy 8, trisomy 9, and deletion [20q$^-$] are most frequent)	10% at 15 years after diagnosis; up to 50% 20 yr after diagnosis	Phlebotomy, hydroxyurea, anagrelide, low-dose aspirin, interferon alfa
ET	Probably lower than in PV	40–50	< 5%; no consistent cytogenetic abnormalities	< 5%	Hydroxyurea, anagrelide, interferon alfa, low-dose aspirin
IM	Scarce epidemiologic data	60	No consistent cytogenetic abnormalities	5%–20%	Symptomatic therapy, hydroxyurea, interferon alfa, stem cell transplantation, thalidomide, antiangiogenesis agents

CML—chronic myelogenous leukemia ET—essential thrombocythemia FTI—farnesyl transferase inhibitors IM—idiopathic myelofibrosis PV—polycythemia vera SRCI—Src kinase inhibitors

during clonal evolution and progression of the disease. The history of the Philadelphia translocation dates back to 1960, when Nowell and Hungerford discovered an abnormal G-group chromosome in bone marrow cells of two patients with CML.[3] However, not until a decade later was it possible to prove through chromosome banding techniques that the previously identified abnormality was a shortened chromosome 22.[4] Rowley then demonstrated that the abnormality was a result of a translocation of segments between the long arms of chromosomes 9 and 22.[5] Further work in the 1980s unraveled much of the molecular details underlying the cytogenetic events.[6] The Philadelphia translocation was thus the first chromosomal abnormality to be associated with a malignant disease and has become the hallmark of CML.

That the Philadelphia translocation and its molecular sequelae represent more than just a diagnostic marker has been demonstrated by in vivo animal models.[7] Transfection of the *bcr-abl* gene into mice can generate a syndrome that closely resembles the clinical picture of human CML. Conversely, inhibition of the protein product of the *bcr-abl* gene appears to reverse the disease in humans.

Lack of concordance in monozygotic twins and the demonstration of the Philadelphia chromosome (Ph) in hematopoietic progenitor cells only suggest that CML is an acquired disorder. Predisposing factors, however, remain mostly unidentified. The incidence of CML is significantly higher for survivors of the atomic-bomb explosions in Hiroshima and Nagasaki. Effects of therapeutic doses of radiation on the development of CML are more disputed.[8] No association has been established with infectious, toxic, or environmental factors.

Up to 95% of patients with CML express Ph, which results from a reciprocal translocation between the long arms of chromosomes 9 and 22. This process involves exchanges of genetic material from a segment of the *abl* (abelson) gene on chromosome 9q34 with part of the *bcr* (breakpoint cluster region) gene on chromosome 22q11, creating a fusion *bcr-abl* gene on 22q11. The segments from the *abl* gene can be transposed into different regions on the *bcr* gene. According to the location of the breakpoint cluster region on *bcr,* messenger RNAs of various lengths are created that are then translated into fusion proteins of various sizes ($p190^{bcr\text{-}abl}$, $p210^{bcr\text{-}abl}$, and $p230^{bcr\text{-}abl}$). Most patients with CML express $p210^{bcr\text{-}abl}$. Expression of the smaller protein $p190^{bcr\text{-}abl}$ is more frequently found in children and adults with acute lymphoblastic leukemias, and $p190^{bcr\text{-}abl}$ may be associated with monocytosis in CML. Expression of the larger $p230^{bcr\text{-}abl}$ variant has been associated with chronic neutrophilic leukemia and with thrombocytosis.[9,10]

The Bcr-Abl proteins trigger molecular transformations through their increased tyrosine kinase activity and their interactions with other cellular proteins. The *abl* gene encodes a nonreceptor tyrosine kinase ($p145^{abl}$) that is involved in signal transduction and the regulation of cell growth; the activity of $p145^{abl}$ is rigorously controlled. By contrast, $p210^{bcr\text{-}abl}$ is a pleiotropic molecule that exhibits higher and uncontrolled kinase activity. Uncontrolled activation of the kinases initiates downstream signaling pathways that upregulate transcription of gene complexes mediating proliferation and transformation of CML hematopoietic progenitor cells. A central element of Bcr-Abl signaling is Ras. Activation of Ras is mediated through a series of adapter proteins that in turn connect $p210^{bcr\text{-}abl}$ to other kinases and messenger systems.[9,11] Other cytotoxic effects of $p210^{bcr\text{-}abl}$ include alterations in the cytoskeletal structure of CML stem cells and progenitor cells, as well as perturbations of adhesion molecule expression, suppression of programmed cell death (apoptosis), abrogation of growth factor dependence, and alteration of DNA repair mechanisms; these effects lead to subtle genetic errors that may result in clonal evolution and progression to the blast phase.[12] However, the multiple signaling pathways dysregulated by $p210^{bcr\text{-}abl}$ may be sufficient to explain the initial manifestations of the chronic phase of CML.

Table 2 Criteria for Defining Accelerated-Phase and Blast-Phase CML

Accelerated Phase	*Blast Phase**
Peripheral blood blasts 15%	≥ 30% blasts in blood and/or marrow
Peripheral blood blasts and promyelocytes ≥ 30% Peripheral blood basophils ≥ 20% Platelet count < 100×10^9/L unrelated to therapy Cytogenetic evolution	Extramedullary infiltrates of leukemic cells

*Lymphoid in one third of patients (TdT+, CD10+, CD19+, CD20+; frequent coexpression of myeloid markers), myeloblastic or undifferentiated in two thirds of patients.

DIAGNOSIS

CML is characterized by the expansion of myeloid progenitor cells at various stages of their maturation, their premature release into the circulation, and their tendency to home to extramedullary sites. In most patients, CML develops in two or three phases [*see Table 2*]. At diagnosis, more than 90% of patients are in the stable or chronic phase.

Clinical Presentation

Symptoms at presentation reflect the increase in mass and turnover of the leukemic cells. Patients may complain of lethargy and weakness, night sweats, and weight loss. Occasionally, increase in abdominal girth and abdominal discomfort are caused by an enlarged spleen. Less frequently, easy bruising and bleeding resulting from platelet dysfunction are recorded. In older series, 10% to 20% of patients displayed no symptoms, but in recent studies, as many as 50% of patients have shown no symptoms and have been diagnosed through routine blood tests, indicating a trend toward earlier diagnosis of CML. Presentations in accelerated or blastic phases are rare and occur in 5% to 10% of patients. Generalized lymphadenopathy and fever, which are rare in patients with chronic-phase CML, may indicate an accelerated disease course.[13] An enlarged spleen is the most common finding on physical examination and can be observed in up to 50% of patients. Hepatomegaly and lymphadenopathy are infrequent.

Laboratory Findings

Peripheral blood and bone marrow The typical laboratory manifestations in chronic-phase CML are myeloid hyperplasia in the marrow and thrombocytosis, neutrophilic leukocytosis, and basophilia in the peripheral blood. During the chronic phase, the peripheral blood is characterized by the presence of myeloid cells in all stages of their maturation, from normal neutrophils to circulating blasts that resemble the cells seen in a bone marrow aspirate smear. Peripheral blood leukocytosis of more than $100 \infty 10^9$/L is seen in 70% to 90% of patients. Dys-

plastic changes of neutrophils are not typical. An absolute monocytosis may occur. Anemia of varying degrees is frequent. The marrow is hypercellular with a myeloid-to-erythroid ratio of between 9:1 and 15:1. Myeloid cells usually show all stages of maturation, with a preponderance of immature precursors such as myelocytes and promyelocytes. The level of megakaryocytes is frequently increased, especially in accelerated phases; megakaryocyte proliferation may be marked. Marrow fibrosis may be focal in early disease stages; it progresses to a more diffuse pattern with disease evolution.

Cytogenetic analysis and molecular assays Examination of metaphase spreads from marrow samples is the definitive test for demonstration of the Philadelphia translocation. It may also reveal additional karyotypic abnormalities, which are frequently observed with disease progression (clonal evolution). Cytogenetic analysis is, however, tedious and time-consuming, and only a limited number of metaphases (< 20 to 25 per sample) are examined. In about 10% of patients with CML, no Philadelphia translocation can be identified by cytogenetic analysis. In up to one third of these patients, molecular studies will detect *bcr-abl* rearrangements that are otherwise missed by routine cytogenetics. The remaining patients test negative for Ph and for the *bcr-abl* gene and are referred to as having atypical CML, Ph-negative CML, or, in some cases, chronic myelomonocytic leukemia (CMML). The prognosis for these patients is significantly worse than the prognosis for patients who test positive for the *bcr-abl* gene.

Molecular techniques have become a mainstay in the diagnosis of CML and in the assessment of response to therapy. Polymerase chain reaction, Southern and Northern blot testing, and immunoprecipitation can determine the exact breakpoints of the fusion gene, detect *bcr-abl* transcripts at the RNA level, and demonstrate the $p210^{bcr\text{-}abl}$ protein by use of antibodies against parts of the Bcr-Abl protein. Patients who are receiving therapy are frequently followed by use of PCR and fluorescence in situ hybridization (FISH).[14] Both techniques provide a higher degree of sensitivity than can be achieved with routine cytogenetics. A practical approach to the monitoring of patients with CML during therapy involves a combination of laboratory assays [*see Table 3*].

Table 3 Use of Laboratory Assays in the Monitoring of Patients with CML on Imatinib

Time	*Assay*
Before therapy	Cytogenetics FISH (peripheral blood) Quantitative PCR (peripheral blood)
During therapy	Cytogenetics every 6 to 12 mo (when < 10% positive, confirm complete response with FISH) FISH every 2–3 mo
During complete cytogenetic response	Cytogenetics every 6–12 mo Quantitative PCR every 2–3 mo

FISH—fluorescence in situ hybridization PCR—polymerase chain reaction

Differential Diagnosis

CML can usually be diagnosed by careful examination of peripheral blood and bone marrow slides in conjunction with the presenting clinical picture. Either detection of the Philadelphia translocation on metaphase spreads or detection of the *bcr-abl* rearrangement by molecular assays is definitive for the diagnosis of CML. Uncertainty may arise in cases where either no cytogenetic-molecular studies are available or test results are negative [*see Table 4*].

Table 4 Differential Diagnosis of Philadelphia Chromosome and *bcr-abl*–Negative Chronic Myelogenous Leukemia

Condition	*Characteristics*
Leukemoid reaction	Total white blood cell count rarely $< 100 \times 10^9$/L Left shift on peripheral blood differential does not usually involve promyelocytes/myelocytes No basophilia Cytoplasmic inclusions (toxic granulation, Döhle bodies) Usually elevated leukocyte alkaline phosphatase Usually underlying etiology (infection, shock, hemolysis, chronic inflammation, other underlying malignancy)
Other myeloproliferative disorders (myelofibrosis, polycythemia vera, essential thrombocythemia)	Lesser degree of leukocytosis Larger spleen Leukoerythroblastic peripheral blood smear (teardrop cells) Established criteria for diagnosis
Atypical chronic myelogenous leukemia	Dysplastic changes in leukocytes of peripheral blood/bone marrow No basophilia
Chronic neutrophilic leukemia	Leukocytosis usually consisting of mature neutrophils Splenomegaly High leukocyte alkaline phosphatase test score
Chronic myelomonocytic leukemia (myeloproliferative)	Increase in absolute and relative monocyte counts Minimal dysplastic changes in leukocytes No basophilia Immature cells of neutrophil series present
Juvenile chronic myelogenous leukemia	Disease in children < 4 yr Hepatomegaly and splenomegaly common at presentation Markedly increased fetal hemoglobin levels

CLINICAL COURSE

At the time of diagnosis of CML, most patients are in the chronic phase. However, the natural history of CML leads invariably either to a blastic transformation directly (i.e., conversion to an acute leukemia–type disease) or, more frequently, to a blastic transformation with an intervening accelerated phase. Whereas the median time from diagnosis of the chronic phase to blastic transformation is about 36 to 40 months, blastic transformation may result in death within 3 to 6 months. Transformation is frequently heralded by refractoriness to treatment, leukocytosis with significant increases in blood and marrow blasts, increases or decreases in platelet counts unrelated to therapy, and clinical signs and symptoms such as unexplained fever, lymph-

Table 5 Drug Dosages and Side Effects in the Management of Patients with Chronic Myelogenous Leukemia

Drug	*Dosage*	*Side Effects*
Imatinib*	Chronic phase: 400 mg p.o. daily Advanced phases: 600 mg p.o. daily Avoid doses < 300 mg p.o. daily	Nausea and vomiting, bone and muscle aches, diarrhea, fatigue, fluid retention, rash
Busulfan	Administer 0.1 mg/kg daily until white blood cell (WBC) count decreases by 50%, then reduce dose by 50%; maintain WBC count between 20 and 50 × 10^9/L	Severe, prolonged myelosuppression (10%); idiosyncratic pulmonary fibrosis (busulfan lung); fibrosis of the endocardium; Addison disease–like syndrome
Hydroxyurea	40 mg/kg p.o. daily; maintain WBC count between 2 and 10 × 10^9/L	Nausea and vomiting, skin changes (scaling, atrophy, alopecia, ulcerations)
Interferon alfa	Administer 5 million units (MIU)/m^2 S.C. daily in cases of high WBC counts: cytoreduction with hydroxyurea (1 to 5 g/day) to decrease WBC count to 10 × 10^9/L; increase interferon dose gradually (3 MIU/day × 3 to 7 days, then 5 MIU/day × 3 to 7 days, then 5 MIU/m^2 or maximum tolerated dose)	*Early effects:* flulike symptoms, such as fever, chills, postnasal drip, anorexia *Late effects:* fatigue, depression, insomnia, weight loss, alopecia, reduced libido, impotence, autoimmune phenomena (hemolytic anemia and thrombocytopenia, Raynaud disease, collagen vascular disease, hypothyroidism), nephrotic syndrome, cardiac arrhythmias

*First-line therapy in chronic-phase CML (alone or in combination with other agents).

adenopathy, progressive splenomegaly, extramedullary disease, weight loss, and bone and joint pain. One third of the leukemias that follow blastic transformation are morphologically and immunophenotypically consistent with acute lymphoblastic leukemia (ALL) and respond to ALL therapy. Two thirds of the patients have an acute myeloblastic or undifferentiated leukemialike phenotype and may respond to anti–acute myeloid leukemia (AML)-type regimens. The possibility of development of myeloid or lymphoid blast crisis provides further evidence that an early stem cell clonal abnormality is the basis for this disease. Criteria for the definition of accelerated-phase and blastic-phase disease have been proposed [*see Table 2*].[15]

The progression of CML is also characterized by further abnormalities at the cytogenetic-molecular level. Ph predominates throughout the chronic phase, but additional cytogenetic abnormalities develop in up to 80% of patients as the disease progresses (clonal evolution). These findings may even precede the clinical and hematologic manifestations of the transformation. Commonly observed changes during clonal evolution of CML are trisomy 8, isochromosome i(17q), trisomy 19, and an additional Ph. Trisomy 8 and isochromosome i(17q) are common during myeloid transformation.[16] Molecular markers that undergo alterations during clonal evolution include *p53, Rb1, c-myc, $p16^{INK4a}$, ras,* and AML/EVI-1, a fusion protein resulting from translocation t(3;21)(q26;q22). Abnormalities of *p53* occur in 20% to 30% of patients and are mainly associated with myeloid transformation, whereas abnormalities of *Rb1* are mainly seen with lymphoid transformation.[17]

PROGNOSIS

The prognosis for patients with CML has changed significantly in the past 2 decades. Patients who are diagnosed with chronic-phase CML can expect a median survival of 5 to 7 years and up to 9 years if they have a good prognosis (i.e., age younger than 50 years; asymptomatic at diagnosis; no organomegaly; absence of marrow fibrosis, basophilia, anemia, and peripheral blood or bone marrow blastosis; short time to achieve hematologic remission on therapy; long remission duration; and good initial hematologic and cytogenetic response to interferon-alfa therapy). Five-year and 10-year survival rates are 60% to 70% and 30% to 40%, respectively.[15] Earlier diagnosis, better supportive care, and more effective anti-CML therapies such as allogeneic SCT, imatinib, and interferon alfa have contributed to this progress. Therapies targeting distinct molecular pathways of CML leukemogenesis, better understanding of risk stratification of patients, and risk-oriented therapies will further improve outcomes.

TREATMENT

The development of imatinib has revolutionized the therapeutic approach to patients with CML. In the few years since the first clinical trials, imatinib has been confirmed as an active agent in the treatment of both chronic and transformed phases of CML. Although the long-term impact of imatinib therapy on CML cure cannot yet be assessed, if the early clinical experience of high cytogenetic response rates, low rates of disease progression, and safe toxicity profile persists, imatinib will soon be established as the most effective therapy in CML. Currently, SCT remains the only cure for CML; however, the morbidity associated with this form of therapy has limited its use to certain patient groups. On the basis of current evidence, imatinib is supplanting interferon alfa as the treatment of choice for patients whose age and disease stage make them poor candidates for SCT [*see* Imatinib versus Stem Cell Transplantation, *below*]. Hydroxyurea and busulfan are the most commonly used agents in the treatment of CML; however, treatment with either drug is not curative and only rarely results in cytogenetic response. Thus, therapy with these agents must be considered palliative.

Conventional Agents and Adjuvant Therapy

Busulfan, an alkylating agent, was introduced as a treatment for CML in the 1950s and was the first chemotherapy agent found to be effective for this disease. Hydroxyurea, a cell cycle–specific inhibitor of DNA synthesis, has been available since the early 1970s. Busulfan and hydroxyurea can be given orally and are inexpensive [*see Table 5*]. Although both drugs can result in complete hematologic remission (CHR) (i.e., normalization of peripheral blood counts and disappearance of palpable splenomegaly and signs and symptoms of the disease) in up to 60% to 80% of patients with chronic-phase CML, cytogenetic remissions are rare. In other words, even though peripheral blood counts may return to normal, the identifiable cells in the blood

and bone marrow most often still contain Ph, suggesting that proliferation has been suppressed but the abnormal clone has not been eliminated. Therefore, both agents have little or no effect on the natural course of the disease, and patients will inevitably experience progression to the blastic phase and die of its complications.[15]

Hydroxyurea is very effective in initial cytoreduction as an adjunct to more definitive therapies (e.g., before allogeneic SCT). It is occasionally used in patients in whom interferon-based regimens, imatinib therapy, or allogeneic SCT has failed and who are not candidates for or who are between investigational therapies.

Splenectomy does not improve survival if it is performed during the chronic phase of CML.[18] It may benefit some patients who have persistent massive or symptomatic splenomegaly and refractory cytopenias. If performed before bone marrow transplantation, splenectomy reduces the time to marrow recovery but does not influence long-term prognosis.

Stem Cell Transplantation

In carefully selected patients, allogeneic SCT can cure a substantial percentage of patients who have suitable donors. Novel transplant modalities are being developed that use unrelated donors, donor lymphocytes, autologous stem cells that are subjected to various purging procedures, or peripheral blood as the stem cell source.[19] Peripheral blood stem cells appear to be as good as if not better than bone marrow stem cells and are being used more frequently in the treatment of hematologic cancers. A randomized, controlled trial indicated that peripheral stem cells restore blood counts faster than bone marrow stem cells without increasing the risk of graft versus host disease.[20]

Matched related allogeneic stem cell transplantation Allogeneic SCT results in long-term survival in 50% to 80% of patients with chronic-phase CML; disease-free survival rates are 30% to 70% for such patients. Applicability of allogeneic SCT is, however, limited by availability of matched siblings and age restrictions. Fewer than 30% of patients in Europe and North America receive stem cells from matched sibling donors. Outcome of allogeneic SCT is best in young patients who receive transplants early in chronic-phase CML, ideally within the first 2 years after the diagnosis of the disease. Studies from the Fred Hutchinson Cancer Center indicate favorable outcomes for patients 50 years of age and older, but results from most other transplant centers are worse.[21-23]

Relapses after transplantation occur in 10% to 70% of patients. The two most significant factors influencing transplant outcome are age and disease stage. The rate of relapse is lowest in younger patients and in those who receive transplants during the chronic phase; it is highest in patients who receive transplants during the blastic phase.[24,25] The prognosis of patients who have relapsed after allogeneic SCT is not necessarily dismal and depends mostly on the time interval between transplantation and relapse. Second transplants from HLA-identical sibling donors, infusion of donor lymphocytes to stimulate the so-called graft-versus-leukemia effect, and treatment with imatinib or interferon alfa have been associated with disease-free survival rates of 30% to 80%.[26] Donor lymphocyte infusions (DLIs) induce cytogenetic and complete hematologic response in 60% to 80% of patients, and those responses are durable. Responses are, however, far less frequent and short-lived in transformed CML phases.[26] Nevertheless, the impressive clinical responses to infusion of donor lymphocytes provide evidence of the role of the immune system in controlling the disease. Treatment with imatinib has resulted in a complete hematologic response rate of 74% and a cytogenetic response rate of 58% and may soon supplant DLI as the treatment of choice for patients who relapse after allogeneic SCT, especially in patients with graft versus host disease.[27] Combinations of DLI and imatinib are currently being explored.

Matched unrelated donor transplantation Matched unrelated donor transplants have high rates of treatment-related morbidity and mortality that can exceed 50%, depending on age and degree of matching. In carefully selected patients, 5-year survival rates may be as high as 70%, relapse rates less than 10%, rates of graft failure below 10%, and rates of severe acute graft versus host disease below 50%.[28] Until a high level of experience with matched unrelated donor transplantation has been achieved, the procedure should be offered preferentially to younger patients with chronic-phase CML who are resistant to imatinib and who have a fully matched donor available.[29]

Reduced-intensity transplantation Reduced-intensity transplantation (also called nonmyeloablative SCT or "mini" SCT) is designed to induce immunosuppression and allow engraftment of donor cells rather than to ablate the marrow. The rationale is twofold: (1) to induce an optimal graft-versus-leukemia effect by eliminating leukemic cells through the action of donor alloreactive immunocompetent cells and (2) to avoid standard high-dose myeloablative chemoradiotherapy, thereby reducing transplant-related morbidity and mortality and extending treatment options to older patients. Early results of reduced-intensity transplants show rapid engraftment, complete eradication of host hematopoietic cells, and minimal procedure-related toxicities.[30]

Imatinib Mesylate (Imatinib, STI571)

The Bcr-Abl proteins require phosphorylation for their increased tyrosine kinase activity, a process that can be blocked by small molecules such as imatinib. Imatinib has been shown to selectively inhibit both $p210^{bcr\text{-}abl}$ and $p190^{bcr\text{-}abl}$ disease, suggesting a role not only in CML but also in Ph-positive acute lymphoid leukemia.[31] In a phase I study of patients with blast-phase CML and patients with chronic-phase CML who were resistant to or intolerant of interferon alfa–based therapy, doses of imatinib were escalated from 25 mg to 1,000 mg daily. No dose-limiting toxicities were established, and significant clinical benefits were observed at daily oral doses of 300 mg or higher. Complete hematologic response rates occurred in 53 of 54 chronic-phase patients treated with imatinib at daily doses of 300 mg or higher; treatment responses were durable in most of these patients. Major cytogenetic responses occurred in 31% of the patients.[32] In patients with blast-phase CML, responses to imatinib therapy occurred in 55% of patients who experienced a myeloid blast crisis and 70% of patients who experienced a lymphoid blast crisis. These responses were, however, transient.[33]

Phase II studies confirmed high response rates with imatinib therapy in patients with chronic-phase CML. These studies demonstrated that imatinib was the most active single agent in accelerated-phase CML and that imatinib had single-agent biologic activity in blast-phase CML [*see Table 6*].[34-38] A multinational, randomized study compared imatinib (400 mg daily) with a regimen of interferon alfa (5 mIU/m^2 daily) and cytarabine (20 mg/m^2 daily for 10 days/mo) in patients with newly diagnosed, early chronic-phase CML.[39] A total of 1,106 patients were

enrolled and randomized equally to either treatment arm. Patients were allowed to cross over to the other treatment arm if they experienced disease progression or intolerance to therapy or if they failed to achieve a major cytogenetic response after 24 months of treatment. At a median follow-up of 19 months, imatinib proved superior to a regimen of interferon alfa and cytarabine as first-line therapy for patients with newly diagnosed chronic-phase CML in terms of major (87.1% versus 34.7%) and complete (76.2% versus 14.5%) cytogenetic responses, less likelihood of progression to advanced-stage CML (1.5% versus 7%), and increased tolerability of treatment (intolerance of 0.7% for imatinib versus 7% for interferon alfa and cytarabine).

Imatinib versus Stem Cell Transplantation

Stem cell transplantation can be curative, but treatment-related mortality and disabling morbidity may be substantial. Therapy with imatinib achieves high rates of major and complete cytogenetic responses in patients with chronic-phase CML. On the basis of experience with interferon alfa, the major cytogenetic response rate achieved with imatinib may be viewed as a surrogate end point for prolonged survival. If the results from the interferon alfa studies can be extrapolated, the median survival of patients with CML may exceed 10 years. However, more patients and longer follow-up are needed to fully appreciate the impact of imatinib on the prognosis of patients with CML. The decision regarding how best to treat a patient with CML remains, therefore, controversial, and treatment algorithms need to be continuously updated in the light of emerging data [*see Figure 1*].[40,41]

Novel Therapeutic Approaches in CML

Although rare, resistance to imatinib poses further challenges. Combination treatments with imatinib and interferon alfa–based therapy are currently in clinical trials. High-dose imatinib (up to 800 mg daily) is being evaluated in newly diagnosed patients with chronic-phase CML and has produced positive results that will require further confirmation.[42] Interferon attached to polyethylene glycol (PEG-interferon) increases the serum half-life of interferon alfa, can be given subcutaneously once a week, may be less toxic and more effective than regular interferon alfa therapy, and when combined with imatinib may prove an effective treatment for patients resistant to imatinib in single-agent regimens.[43] Other combinations of imatinib with homoharringtonine, decitabine, farnesyltransferase, and Src kinase inhibitors are being explored. Their value in the overall treatment algorithm of CML is as yet undefined.[44]

Polycythemia Vera

Polycythemia vera (PV) is a clonal disorder of hematopoietic stem cells. Unlike CML, however, no clear causative cytogenetic-molecular lesion has been identified. PV occurs with an incidence of about 1 to 2 per 100,000 population a year in Western Europe and the United States and may be more frequent than previously thought.[45,46] The median age at diagnosis is 65 years, and a slight male predominance exists. PV is extremely rare in patients younger than 30 years.[45,46]

PATHOPHYSIOLOGY

PV is characterized by an absolute increase in total red blood cell mass that may be accompanied by thrombocytosis or leukocytosis. The chromium-51 (^{51}Cr) dilution technique is simple and is well established for measurements of red blood cell mass. This expansion of red cell mass is caused by increased production by hypercellular bone marrow and is not dependent on serum levels of erythropoietin. Indeed, erythropoietin levels are typically low in PV. This distinguishes PV from the secondary polycythemia associated with certain tumors (e.g., renal cell carcinoma and hepatocellular carcinoma) and with pulmonary, cardiac, and renal disorders.[47] The pathogenesis of erythrocytosis in PV is under investigation. In vitro hypersensitivity to various cytokines caused by defects of cellular receptors or intracellular signal transduction pathways has been suggested.[48,49]

DIAGNOSIS

The clinical manifestations of PV are a consequence of the excessive proliferation of hematopoietic cell lines and are mainly characterized by microvascular and macrovascular thrombotic events. On physical examination, the most common findings in patients with PV include ruddy cyanosis, hepatosplenomegaly, conjunctival plethora, and hypertension. About 20% of patients present with thrombotic complications, and such complications will develop in 30% more patients during the course of the disease. Microvascular symptoms such as acroparesthesias, erythromelalgia, peripheral gangrene, and ischemic neurologic and visual disturbances are frequent. Particularly serious thrombotic events involve the cerebral, coronary, hepatic and inferior vena cava (Budd-Chiari syndrome), and mesenteric vessels.

The diagnosis of PV requires exclusion of secondary causes of increased red blood cell mass and blood volume. Independent determination of red cell mass and plasma volume by isotope dilution is mandatory. Identification of cytogenetic abnormalities such as trisomy of chromosomes 8 and 9 and deletions of the long arm of chromosome 20 (20q–) are firm evidence of the clonality of the process, but these findings can be demonstrated in only a minority of patients. The Polycythemia Vera Study Group (PVSG) has established diagnostic criteria that are continuously being modified [*see Table 7*].[47,50,51]

DIFFERENTIAL DIAGNOSIS

The diagnosis of PV can be masked by concurrent iron defi-

Table 6 Summary of Results of Treatment with Imatinib in Late Chronic-Phase, Accelerated-Phase, and Blast-Phase CML

Response to Imatinib (%)	*Late Chronic Phase (n = 454)*[33]	*Accelerated Phase (n = 181)*[34]	*Blast Phase (n = 229)*[35]
Hematologic response	430 (95)	125 (69)	66 (29)
Major cytogenetic response	272 (60)	43 (24)	36 (16)
Complete	188 (41)	30 (17)	15 (7)
Partial	84 (19)	13 (7)	21 (9)

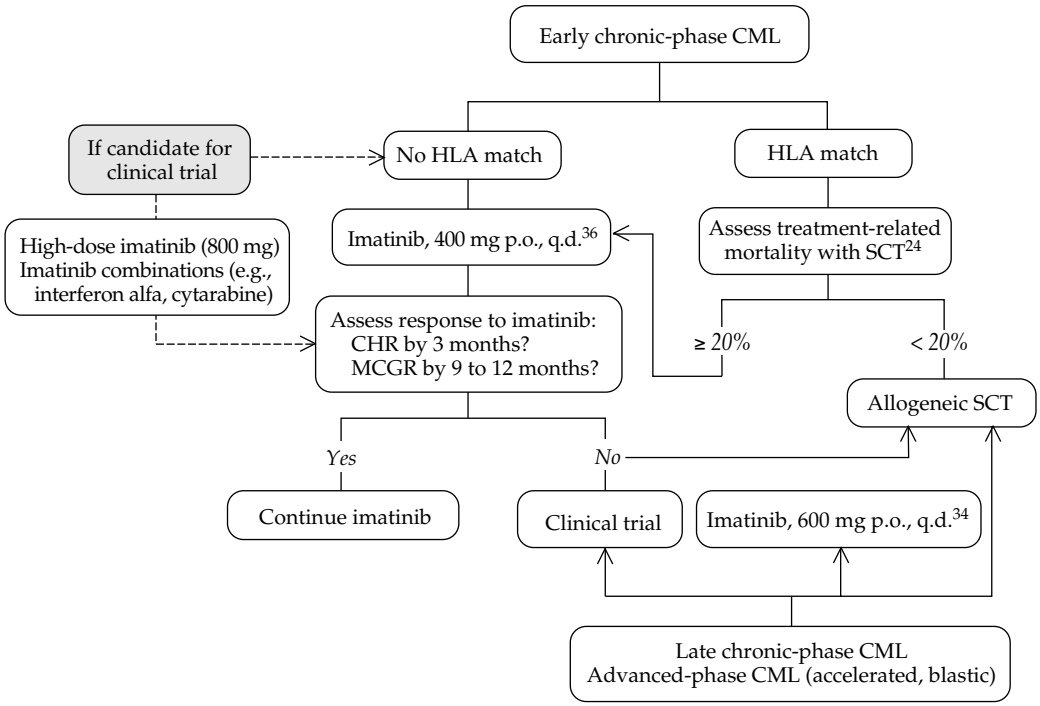


Figure 1 **For patients in early chronic-phase chronic myelogenous leukemia (CML), allogeneic stem cell transplantation and imatinib are possible first-line therapies. The treatment decision depends on several factors: (1) availability of a suitable donor; (2) the risk-benefit ratio of SCT versus imatinib; (3) the socioeconomic situation; and (4) patient preference. For patients who have a matched sibling donor and whose physical condition allows stem cell transplantation with an acceptable treatment-related morbidity and mortality (< 20%), allogeneic stem cell transplantation may still be the first choice. Although stem cell transplantation is still considered the only curative modality and concerns exist about delaying transplantation for other treatment modalities, numerous arguments favor a more conservative approach (i.e, chemotherapy using imatinib): (1) delaying transplantation for up to 24 months in chronic-phase CML does not inadvertently influence transplant outcome; (2) on the basis of experience with interferon alfa, complete and major cytogenetic response rates achieved with imatinib are valid surrogate markers for survival; (3) allogeneic transplantation is associated with mortality and considerable morbidity (e.g., cataracts, sterility, hip necrosis, second cancers, graft versus host disease); and (4) experience with imatinib has not shown any unusual or unexpected side effects, and resistance during the chronic phase is low. If a conservative approach to therapy is preferred, the patient may be considered for clinical trials utilizing high-dose imatinib (800 mg p.o. daily), combinations of imatinib with other agents (e.g., interferon alfa, PEG–interferon alfa, cytarabine), other targeted therapies (e.g., farnesyltransferase inhibitors), or other active agents (e.g., decitabine, histone deacetylase inhibitors, homoharringtonine). Patients in late chronic phase or accelerated phase should be offered investigational therapies (see above) or should undergo allogeneic transplantation. (CHR—complete hematologic remission; MCGR—major cytogenetic response; SCT—stem cell transplantation)**

ciency, resulting in a clinical picture resembling essential thrombocythemia. Unlike ET, venous and arterial thromboses occur with equal frequency in PV. The main risk factors for vascular complications include advanced age, a history of vascular events, and consistently high hematocrit values. Whereas thrombotic complications are the cause of death in 30% to 40% of patients, bleeding complications are far less frequent and are seen in about 5% to 10% of patients.[49,52]

CLINICAL COURSE

The natural history of PV in the absence of therapeutic intervention has been difficult to define. PV may transform into a myelofibrosis-type syndrome (spent phase, postpolycythemic myeloid metaplasia) or progress to a leukemic phase; however, it is difficult to exclude the impact of therapy (such as with radiophosphorus ^{31}P or alkylating agents) in all cases. The PVSG estimates that the incidence of leukemia among PV patients is 10% at 15 years after diagnosis and 50% at 20 years after diagnosis, regardless of the treatment utilized.[47] Untreated patients with persistently high hematocrit levels do poorly and may die

Table 7 Diagnostic Criteria for Polycythemia Vera*

Major Criteria	*Minor Criteria*
Increased red cell mass (in males, ≥ 36 ml/kg; in females, ≥ 32 ml/kg)	Platelet count > 400 × 10^9/L
Normal oxygen saturation, ≥ 92%	Leukocytosis > 12 × 10^9/L (no fever or infection)
Splenomegaly	Leukocyte alkaline phosphatase score > 100
	Serum B_{12} > 900 pg/ml
	Unbound B_{12} binding capacity > 2,200 pg/ml

*A diagnosis of polycythemia vera can be considered when all major criteria are present or in the presence of the first two major criteria and any two minor criteria.

from a thrombotic complication, which is the cause of death in 20% to 40% of patients.[53]

TREATMENT

Phlebotomy, low-dose aspirin, and possibly hydroxyurea represent the best approach to treatment. Low-dose aspirin (50 to 100 mg/day) is particularly effective in controlling the microvascular symptoms associated with thrombocytosis when used in combination with either phlebotomy or myelosuppressive treatments.

Reduction of the red cell mass to a safe level (i.e., reducing the hematocrit level to less than 45% in men and less than 42% in women) by phlebotomy should be a primary treatment goal.[49] Patients treated with phlebotomy alone have a low risk of leukemic transformation; in addition, patients who are maintained at a stable hematocrit level are at low risk for incurring thrombotic complications.[54] Hydroxyurea is efficient in improving thrombocytosis and erythrocytosis.[55] Because of uncertainty as to whether hydroxyurea is a potential cause of leukemic transformation, treatment with hydroxyurea should be restricted to situations where other remedies have failed. Anagrelide and interferon alfa are used in patients with persistent thrombocytosis and thrombosis.[56-58] Anecdotal evidence suggests that the tyrosine kinase inhibitor imatinib may be active against PV.[59]

Chlorambucil, an alkylating agent, and radiophosphorus ^{32}P are mainly of historical interest and should be considered obsolete as treatment of PV because of their association with the development of secondary leukemias and lymphomas.

Essential Thrombocythemia

ET, or thrombocytosis, is a clonal stem cell disorder mainly characterized by sustained proliferation of megakaryocytes, resulting in peripheral blood thrombocytosis.

EPIDEMIOLOGY

The incidence of ET is similar to that of PV. The median age at presentation is 40 to 50 years. Unlike in PV, up to 20% of patients with ET are younger than 40 years.[46]

PATHOPHYSIOLOGY

The elevated platelet count in ET is caused by increased production by megakaryocytes with normal platelet survival. Increased platelet production appears to depend on the effect of several cytokines. Lack of the growth-inhibiting effect of platelet factor 4 and transforming growth factor–β on megakaryocyte colony formation has been demonstrated. Unlike the diminished serum levels of erythropoietin in PV, thrombopoietin levels do not correlate with platelet count or megakaryocyte mass and can even be elevated in ET.[60] However, ET has been associated with dysregulation of the relationship of thrombopoietin to its receptor, c-Mpl.[61] The presence of a heterogenous pattern of c-Mpl expression in bone marrow megakaryocytes has recently been found to be a useful marker in the differential diagnosis of thrombocytosis.[62]

DIAGNOSIS

ET is diagnosed after secondary causes of elevated platelet counts (e.g., iron deficiency, malignancies, inflammatory conditions, or infections) have been excluded. In addition, criteria have been established by the PVSG: (1) although subject to question,[63] a platelet count greater than $600 \infty 10^9$/L on two occasions at least 1 month apart; (2) normal red cell mass; (3) the absence of significant fibrosis of the marrow; (4) the presence of splenomegaly; and (5) no Ph or *bcr-abl* gene rearrangements.

Two thirds of the patients present with thrombotic or hemorrhagic events. Major thrombotic complications occur in 20% to 30%. Thrombotic complications frequently manifest as deep vein thrombosis and pulmonary embolism. Thrombosis of hepatic veins leads to Budd-Chiari syndrome; thrombosis of renal veins can cause nephrotic syndrome. Erythromelalgia and digital ischemia constitute microvascular forms of arterial thrombosis in ET. Hemorrhagic events occur in up to 40% of patients, with the gastrointestinal tract, urinary tract, skin, eyes, or brain being possible bleeding sites.[64] Individual patients can suffer from both thrombotic and hemorrhagic episodes.[65]

ET is increasingly being diagnosed in asymptomatic patients as an incidental finding.

CLINICAL COURSE

Unlike PV, ET rarely progresses to leukemia, which develops in only 3% to 4% of patients.[66] The correlation between the degree of thrombocytosis and the risk of thrombosis is poor. Untreated asymptomatic patients and those who are at low risk for thrombohemorrhagic complications (no history of thrombohemorrhagic episodes, age younger than 60 years, and shorter duration of thrombocytosis) may have the same life expectancy as an age-matched and sex-matched control group without ET.[66]

TREATMENT

The indications for therapeutic intervention have to be considered carefully according to the risk of vascular complications. In general, treatment should be considered only in patients at high risk for thrombohemorrhagic events. Anagrelide, an oral imidazoquinazolin derivative, effectively reduces platelet counts at a dose of 2.5 mg/day. However, side effects (e.g., fluid retention, edema, cardiac arrhythmias, headache, dizziness, nausea, vomiting, bloating, diarrhea) are frequent and necessitate discontinuance of the drug in about 15% of patients.[67] Interferon alfa is also effective in reducing platelet counts, but its use is restricted, especially in elderly patients, by its toxicity profile.[68] The role of PEG–interferon alfa in ET has yet to be assessed.[69] Low-dose aspirin is indicated in patients with a high cardiovascular risk, previous thrombotic complications, and a history of digital or cerebrovascular ischemia.[70] Use of hydroxyurea reduces the risk of thrombosis[71] but should be used only in patients in whom anagrelide or interferon alfa have proved ineffective or cannot be tolerated.

Idiopathic Myelofibrosis

Idiopathic myelofibrosis (agnogenic myeloid metaplasia with myelofibrosis) is a clonal hematopoietic stem cell disorder predominantly affecting the megakaryocyte lineage in the bone marrow. It is characterized by splenomegaly, extramedullary hematopoiesis, and bone marrow fibrosis. The fibrosis is caused by deposition of abnormal collagen by polyclonal fibroblasts that in turn are stimulated by growth factors such as transforming growth factor–β secreted by adjacent megakaryocytes. In some cases, it may be difficult to recognize IM as a distinct entity because many different conditions can cause bone marrow fibrosis and each of the myeloproliferative disorders can terminate in a fibrotic phase.[72,73] Myeloid metaplasia often refers to the earlier phases of IM, when fibrosis is less prominent and

extramedullary hematopoiesis is evident (e.g., in the spleen or liver).

DIAGNOSIS

More than half of the patients with IM present with anemia and thrombocytosis. Most commonly, symptoms such as weakness, fatigue, or dyspnea on exertion are caused by anemia. Bleeding problems and nonspecific constitutional symptoms may accompany the clinical course. Splenomegaly and hepatomegaly occur in up to 70% of patients. Splenomegaly may be massive and reflects, at least partly, the amount of extramedullary hematopoiesis taking place in IM. Leukoerythroblastosis and the presence of teardrop cells in the peripheral blood are typical laboratory features. Bone marrow aspiration will frequently reveal a so-called dry tap, whereas biopsy demonstrates extensive fibrosis and sometimes osteosclerosis.[72]

TREATMENT

No definitive therapy exists for IM. Median survival of patients with IM is about 5 years, and the incidence of transformation to acute leukemia is estimated to be between 5% and 20%.[74] However, half of the patients studied had been exposed to alkylating agents or radiotherapy, which causes leukemic transformation that does not reflect the natural course of IM. These modalities are now rarely used. Treatment is usually reserved for patients who are symptomatic because of anemia or massive splenomegaly. Steroids, androgens, and erythropoietin have been used successfully in some cases for alleviation of anemia. Splenectomy may be indicated in patients with massive splenomegaly and may provide significant relief. However, it can cause substantial perioperative morbidity and mortality and is contraindicated in the presence of thrombocytosis. Also, a high risk of blast transformation has been reported in patients who have undergone splenectomy.[75] Hydroxyurea can improve leukocytosis and constitutional symptoms. Its effect on the natural course of the disease is, however, uncertain. Experience with interferon alfa is still limited.[72,73] Antiangiogenesis inhibitors such as thalidomide are under investigation. Intensive anti-AML–like regimens are investigational and are associated with high occurrences of morbidity and mortality. Allogeneic SCT has been attempted with various degrees of success and may reverse fibrosis.[76]

Stefan Faderl has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Hagop Kantarjian has received grants for clinical research from Novartis AG.

References

1. Jemal A, Murray T, Samuels A, et al: Cancer statistics, 2003. CA Cancer J Clin 53:5, 2003

2. Xie Y, Davis SM, Xiang Y, et al: Trends in leukemia incidence and survival in the United States (1973–1998). Cancer 97:2229, 2003

3. Nowell PC, Hungerford DA: A minute chromosome in human chronic granulocytic leukemia. Science 132:1497, 1960

4. Caspersson T, Gahrton G, Lindsten J, et al: Identification of the Philadelphia chromosome as a number 22 by quinacrine mustard fluorescence analysis. Exp Cell Res 63:238, 1970

5. Rowley JD: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine, fluorescence and Giemsa staining. Nature 243:290, 1970

6. De Klein A, Van Kessel G, Grosveld G, et al: A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelogenous leukemia. Nature 300:765, 1982

7. Ren R: Dissecting the molecular mechanisms of chronic myelogenous leukemia using murine models. Leuk Lymphoma 43:1549, 2002

8. Aguiar RC: Therapy-related chronic myeloid leukemia: an epidemiological, clinical and pathogenetic appraisal. Leuk Lymphoma 29:17, 1998

9. Faderl S, Talpaz M, Estrov Z, et al: The biology of chronic myelogenous leukemia. N Engl J Med 341:164, 1999

10. Li S, Ilaria RL Jr, Million RP, et al: The P190 , P210, and P230 forms of the BCR/ABL oncogene induce a similar chronic myeloid leukemia–like syndrome in mice but have different leukemogenic activity. J Exp Med 189:1399, 1999

11. Deininger MW, Goldman JM, Melo JV: The molecular biology of chronic myeloid leukemia. Blood 96:3343, 2000

12. Kurzrock R, Kantarjian HM, Druker BJ, Talpaz M: Philadelphia chromosome–positive leukemias: from basic mechanism to molecular therapeutics. Ann Intern Med 138:819, 2003

13. Savage DG, Szydlo RM, Goldman JM: Clinical features at diagnosis in 430 patients with chronic myeloid leukemia seen at a referral center over a 16-year period. Br J Haematol 96:111, 1997

14. Dewald GW: Interphase FISH studies of chronic myeloid leukemia. Methods Mol Biol 204:311, 2002

15. Kantarjian HM, Giles FJ, O'Brien S, et al: Clinical course and therapy of chronic myelogenous leukemia with interferon-alpha and chemotherapy. Hematol Oncol Clin North Am 12:31, 1998

16. O'Brien S, Thall PF, Siciliano MJ: Cytogenetics of chronic myelogenous leukemia. Baillieres Clin Haematol 10:259, 1997

17. Shet AS, Jahagirdar BN, Verfaillie CM: Chronic myelogenous leukemia: mechanisms underlying disease progression. Leukemia 16:1402, 2002

18. Bouvet M, Babiera GV, Termuhlen PM: Splenectomy in the accelerated or blastic phase of chronic myelogenous leukemia: a single-institution, 25-year experience. Surgery 122:20, 1997

19. Apperley JF: Hematopoietic stem cell transplantation in chronic myeloid leukemia. Curr Opin Hematol 5:445, 1998

20. Bensinger WI, Martin PJ, Storer B, et al: Transplantation of bone marrow as compared with peripheral blood cells from HLA-identical relatives in patients with hematologic cancers. N Engl J Med 344:175, 2001

21. Van Rhee F, Szydlo RM, Hermans J, et al: Long-term results after allogeneic bone marrow transplantation for chronic myelogenous leukemia in chronic phase: a report from the Chronic Leukemia Working Party of the European Group for Blood and Marrow Transplantation. Bone Marrow Transplant 20:553, 1997

22. Horowitz MM, Rowlings PA, Passweg JR: Allogeneic bone marrow transplantation for CML: a report from the International Bone Marrow Transplant Registry. Bone Marrow Transplant 17(suppl 3):S5, 1996

23. Clift RA, Storb R: Marrow transplantation for CML: the Seattle experience. Bone Marrow Transplant 17(suppl 3):S1, 1996

24. Gratwohl A, Hermans J, Goldman JM, et al: Risk assessment for patients with chronic myeloid leukemia before allogeneic blood or marrow transplantation. Chronic Leukemia Working Party of the European Group for Blood and Marrow Transplantation. Lancet 352:1087, 1998

25. Qazilbash MH, Devetten MP, Abraham J, et al: Utility of a prognostic scoring system for allogeneic stem cell transplantation in patients with chronic myeloid leukemia. Acta Haematol 109:119, 2003

26. Dazzi F, Szydlo RM, Goldman JM: Donor lymphocyte infusions for relapse of chronic myeloid leukemia after allogeneic stem cell transplantation: where we now stand. Exp Hematol 27:1477, 1999

27. Kantarjian HM, O'Brien S, Cortes J, et al: Imatinib mesylate therapy for relapse after allogeneic stem cell transplantation for chronic myelogenous leukemia. Blood 100:1590, 2002

28. McGlave P: Unrelated donor transplant therapy for chronic myelogenous leukemia. Hematol Oncol Clin North Am 12:93, 1998

29. Hansen JA, Gooley TA, Martin PJ, et al: Bone marrow transplants from unrelated donors for patients with chronic myeloid leukemia. N Engl J Med 338:962, 1998

30. Or R, Shapira MY, Resnick I, et al: Nonmyeloablative allogeneic stem cell transplantation for the treatment of chronic myeloid leukemia in first chronic phase. Blood 101:441, 2003

31. Mauro M, O'Dwyer M, Heinreich MC, et al: STI571: a paradigm of new agents for cancer therapeutics. J Clin Oncol 20:325, 2002

32. Druker BJ, Talpaz M, Resta DJ, et al: Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. N Engl J Med 344:1031, 2001

33. Druker BJ, Sawyers CL, Kantarjian H, et al: Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. N Engl J Med 344:1038, 2001

34. Kantarjian H, Sawyers C, Hochhaus A, et al: Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. N Engl J Med 346:645, 2002

35. Talpaz M, Silver RT, Druker BJ, et al: Imatinib induces durable hematologic and cytogenetic responses in patients with accelerated phase chronic myeloid leukemia: results of a phase 2 study. Blood 99:1928, 2002

36. Sawyers CL, Hochhaus A, Feldman E, et al: Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase II study. Blood 99:3530, 2002

37. Marin D, Marktel S, Szydlo R, et al: Survival of patients with chronic-phase chronic myeloid leukaemia on imatinib after failure on interferon alfa. Lancet 362:617, 2003

38. Kantarjian H, O'Brien S, Cortes J, et al: Analysis of the impact of imatinib mesylate therapy on the prognosis of patients with Philadelphia chromosome-positive chronic myelogenous leukemia treated with interferon-alpha regimens for early chronic phase. Cancer 98:1430, 2003

39. O'Brien SG, Guilhot F, Larson RA, et al: Imatinib compared with interferon and low-

dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N Engl J Med 348:994, 2003

40. Goldman JM, Marin D, Olavarria E, et al: Clinical decisions for chronic myeloid leukemia in the imatinib era. Semin Hematol 40:98, 2003

41. Hehlmann R: Current CML therapy: progress and dilemma. Leukemia 17:1010, 2003

42. Kantarjian HM, Talpaz M, O'Brien S, et al: Dose escalation of imatinib mesylate can overcome resistance to standard-dose therapy in patients with chronic myelogenous leukemia. Blood 101:473, 2003

43. Talpaz M: Interferon-alfa–based treatment of chronic myeloid leukemia and implications of signal transduction inhibition. Semin Hematol 38:22, 2001

44. Faderl S, Kantarjian HM, Talpaz M, et al: New treatment approaches for chronic myelogenous leukemia. Semin Oncol 27:578, 2000

45. Najean Y, Rain JD, Billotey C: Epidemiological data in polycythemia vera: a study of 842 cases. Hematol Cell Ther 40:159, 1998

46. Kutti J, Ridell B: Epidemiology of the myeloproliferative disorders: essential thrombocythaemia, polycythemia vera and idiopathic myelofibrosis. Pathol Biol 49:164, 2001

47. Michiels JJ, Juvonen E: Proposal for revised diagnostic criteria of essential thrombocythemia and polycythemia vera by the Polycythemia Vera Study Group. Semin Thromb Hemost 23:339, 1997

48. Hinshelwood S, Bench AJ, Green AR: Pathogenesis of polycythemia vera. Blood Rev 11:224, 1997

49. Spivak JL: Polycythemia vera: myths, mechanisms, and management. Blood 100:4272, 2002

50. Pearson TC, Messinezy M: The diagnostic criteria of polycythemia rubra vera. Leuk Lyphoma 22(suppl 1):87, 1996

51. Pearson TC: Evaluation of diagnostic criteria in polycythemia vera. Semin Hematol 38:21, 2001

52. Michiels JJ: Erythromelalgia and vascular complications in polycythemia vera. Semin Thromb Hemost 23:441, 1997

53. Polycythemia vera: the natural history of 1213 patients followed for 20 years. Gruppo Italiano Studio Policitemi. Ann Intern Med 123:656, 1995

54. Streiff MB, Smith B, Spivak JL: The diagnosis and management of polycythemia vera in the era since the Polycythemia Vera Study Group: a survey of American Society of Hematology members' practice pattern. Blood 99:1144, 2002

55. Kaplan ME, Mack K, Goldberg JD, et al: Long-term management of polycythemia vera with hydroxyurea: a progress report. Semin Hematol 23:167, 1986

56. Tatarsky I, Sharon R: Management of polycythemia vera with hydroxyurea. Semin Hematol 34:24, 1997

57. Fruchtman SM, Mack K, Kaplan ME, et al: From efficacy to safety: a Polycythemia Vera Study Group report on hydroxyurea in patients with polycythemia vera. Semin Hematol 34:17, 1997

58. Elliott MA, Tefferi A: Interferon-α therapy in polycythemia vera and essential thrombocythemia. Semin Thromb Hemost 23:463, 1997

59. Jones CM, Dickinson TM: Polycythemia vera responds to imatinib mesylate. Am J Med Sci 325:149, 2003

60. Cerutti A, Custodi P, Duranti M, et al: Thrombopoietin levels in patients with primary and reactive thrombocytosis. Br J Haematol 99:281, 1997

61. Mesa RA, Hanson CA, Li C-Y, et al: Diagnostic and prognostic value of bone marrow angiogenesis and megakaryocyte c-Mpl expression in essential thrombocythemia. Blood 99:4131, 2002

62. Teofili L, Pierconti F, Di Febo A, et al: The expression pattern of c-Mpl in megakaryocytes correlates with thrombotic risk in essential thrombocythemia. Blood 100:714, 2002

63. Sacchi S, Vinci G, Gugliotta L, et al: Diagnosis of essential thrombocythemia at platelet counts between 400 and 600 ∞ 10^9/L. Haematologica 85:492, 2000

64. Jantunen R, Juvonen E, Ikkala E, et al: The predictive value of vascular risk factors and gender for the development of thrombotic complications in essential thrombocytopenia. Ann Hematol 80:74, 2001

65. Ravandi-Kashani F, Schafer AI: Microvascular disturbances, thrombosis and bleeding in thrombocythemia: current concepts and perspectives. Semin Thromb Hemost 23:479, 1997

66. Murphy S, Peterson P, Iland H, et al: Experience of the Polycythemia Vera Study Group with essential thrombocythemia: a final report on diagnostic criteria, survival, and leukemic transition by treatment. Semin Hematol 34:29, 1997

67. Andersson BS: Essential thrombocythemia: diagnosis and treatment, with special emphasis on the use of anagrelide. Hematology 7:173, 2002

68. Radin AI, Kim HT, Grant BW, et al: Phase II study of alpha2 interferon in the treatment of the chronic myeloproliferative disorders (E5487): a trial of the Eastern Cooperative Oncology Group. Cancer 98:100, 2003

69. Alvarado Y, Cortes J, Verstovsek S, et al: Pilot study of pegylated interferon-alpha 2b in patients with essential thromobcythemia. Cancer Chemother Pharmacol 51:81, 2003

70. Tefferi A, Elliott MA, Solberg LA Jr, et al: New drugs in essential thrombocythemia and polycythemia vera. Blood Rev 11:1, 1997

71. Cortelazzo S, Finazzi G, Ruggeri M, et al: Hydroxyurea for patients with essential thrombocythemia and a high risk of thrombosis. N Engl J Med 332:1132, 1995

72. Groopman JE: The pathogenesis of myelofibrosis in myeloproliferative disorders. Ann Intern Med 92:857, 1980

73. Thiele J, Kvasnicka HM, Werden C, et al: Idiopathic primary osteo-myelofibrosis: a clinico-pathological study on 208 patients with special emphasis on evolution of disease features, differentiation from essential thrombocythemia and variables of prognostic impact. Leuk Lymphoma 22:303, 1996

74. Spivak JL: Diagnosis of the myeloproliferative disorders: resolving phenotype mimicry. Semin Hematol 40:1, 2003

75. Barosi G, Ambrosetti A, Centra A, et al: Splenectomy and risk of blast transformation in myelofibrosis with myeloid metaplasia. Italian Cooperative Study Group on Myeloid with Myeloid Metaplasia. Blood 91:3630, 1998

76. Guardiola P, Anderson JE, Bandini G, et al: Allogeneic stem cell transplantation for agnogenic myeloid metaplasia: a European Group for Blood and Marrow Transplantation, Societe Francaise de Greffe de Moelle, Gruppo Italiano per il Trapianto del Midollo Osseo, and Fred Hutchinson Cancer Research Center Collaborative Study. Blood 93:2831, 1999

XVIII HEAD AND NECK CANCER

EVERETT E. VOKES, M.D., F.A.C.P.

Head and neck cancer occurs in the United States at an annual incidence of approximately 45,000 cases.[1] The vast majority of these tumors arise from the mucosa of the upper aerodigestive tract and are squamous cell carcinomas (SCCs) [*see* Squamous Cell Carcinoma of the Head and Neck, *below*]. Anatomically, head and neck cancers are heterogeneous—arising from the pharynx (including the nasopharynx, oropharynx, and hypopharynx), the oral cavity, the larynx, and the cervical esophagus [*see Figure 1*]. Salivary gland tumors and thyroid cancers differ from other head and neck tumors in their etiology, histology, and standard therapy [*see* Salivary Gland Tumors, *below*].

Overview of Head and Neck Cancer

ETIOLOGY AND RISK FACTORS

Tobacco and Alcohol Use

The primary risk factors for development of SCC of the head and neck are the use of tobacco and alcohol. Tobacco use has clearly been demonstrated to be an independent risk factor. The likelihood of a malignancy increases with the duration and extent of exposure. Cigarette smoking, in particular, is an important risk factor for laryngeal cancer. Cigar smoking has been associated primarily with cancers of the lip and oral cavity, presumably because cigar smoke is not typically inhaled. When inhaled, cigar smoke places the smoker at risk for other head and neck malignancies. There is also concern about the use of smokeless tobacco as a risk factor for malignancies of the oral cavity. After discontinuance of the use of tobacco, the risk of SCC of the head and neck diminishes gradually over time. Evidence, however, suggests that the risk does not decline to the level associated with persons who never smoked.[2]

Alcohol use is also associated with SCC of the head and neck. By itself, alcohol use is a smaller risk factor than tobacco use and is of particular importance for malignancies of the oral cavity, oral pharynx, and hypopharynx. However, the use of alcohol and tobacco in combination results in a multiplicative risk of SCC; in persons who are both heavy smokers and heavy drinkers, the risk of SCC may be 200 times that of persons who neither smoke nor drink.[3]

Viral Infection

Chronic viral infections may be associated with SCC of the head and neck. The strongest association between a virus and head and neck cancer is that of Epstein-Barr virus (EBV) and na-

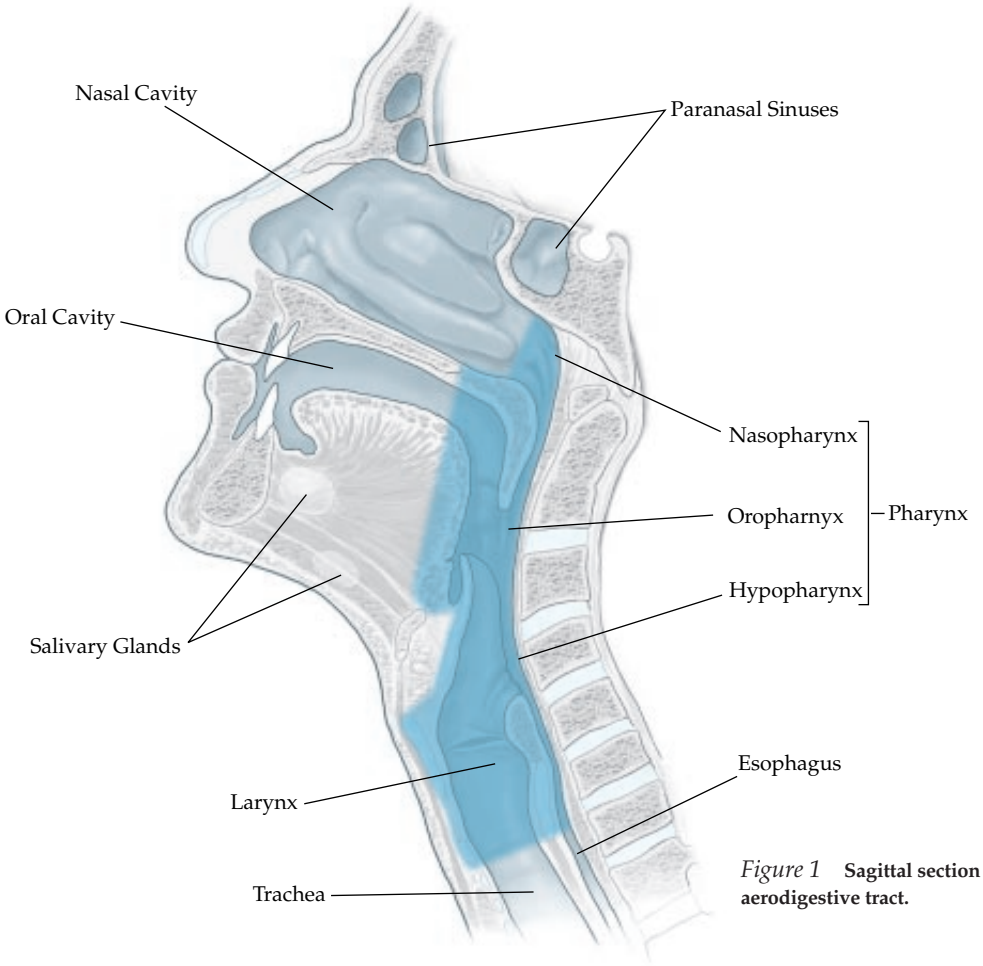


Figure 1 **Sagittal section of the upper aerodigestive tract.**

sopharyngeal carcinoma (NPC). Accumulating evidence suggests that EBV is a primary etiologic agent in the development of NPC.[4,5] In one series evaluating 11 patient samples of nasopharyngeal dysplasia or carcinoma in situ, EBV infection was detected in all cases.[4] Invasive NPC eventually developed in five of the 11 patients, strengthening the concept that EBV infection is an early event in the development of NPC.

Human papillomavirus (HPV), primarily type 16, has been detected in 8% to 36% of SCC cases of the head and neck. Infection with high-risk HPV subtypes (e.g., HPV-16) may contribute to SCC both by direct effects on proliferation and by increasing mutational frequency in the host cells. Studies have found high-risk HPV to be an independent risk factor for SCC. One study found a statistically significant synergistic effect between high-risk HPV and heavy alcohol consumption; an additive effect was found between high-risk HPV and tobacco use.[6]

Additional Risk Factors

Other risk factors that may play a role in the development of head and neck cancer include occupational exposures,[7-9] radiation,[10] dietary factors,[11] and genetic susceptibility [*see* Genetic Alterations, *below*]. In occupational exposure, the relative risk of head and neck cancer increases with increasing intensity and duration of exposure to a variety of occupational chemicals and chemical by-products.

GENETIC ALTERATIONS

Molecular characterization of head and neck cancer includes consistent chromosomal abnormalities, the activation of specific oncogenes, and deletion of tumor suppressor genes. The most frequent mutation is inactivation of the *p53* tumor suppressor gene. This abnormality is found in 50% to 70% of established tumors, but it can also be demonstrated in premalignant tissue and in normal-appearing mucosa in smokers.[12]

PATHOGENESIS

An important concept in understanding the pathogenesis of head and neck cancer is so-called field carcinogenesis, which recognizes that tobacco and alcohol are toxic not only to one specific site in the upper aerodigestive tract but to the entire exposed mucosa. Premalignant lesions such as leukoplakia or erythroplakia may develop wherever mucosa is exposed, and these lesions can progress to invasive carcinoma. Although a specific anatomic site may be characterized as malignant and another as normal or premalignant (e.g., dysplastic or hyperplastic), patients with head and neck cancer are at risk for second malignancies after treatment or may present at the time of diagnosis with multiple areas of abnormality.[13,14] Grossly normal epithelium adjacent to the lesion frequently contains dyplasia, carcinoma in situ, or invasive carcinoma, suggesting that the entire mucosal field has been damaged by a carcinogen. The morphologic equivalent of premalignant lesions are white or red patches that cannot be scraped off (leukoplakia or erythroplakia). Histopathologically, they are usually hyperplastic or dysplastic lesions. However, they may also represent carcinoma in situ or early-stage SCC. These premalignant lesions occur in high-risk patients and frequently have genetic alterations that can also be detected in fully developed head and neck cancers. They may progress to invasive cancer at an estimated rate of 3% to 5% a year.[15]

Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma is a distinct clinicopathologic entity. NPC occurs in two forms: (1) an endemic type, which is observed in Mediterranean countries, in Southeast Asia, and in the native population of Alaska, and (2) a sporadic type, which is observed in Europe and in the United States.[16] The endemic form is linked to EBV exposure. Histologically, NPC is frequently referred to as lymphoepithelioma because it is composed of lymphoid and epidermoid cells. The lymphoid cells, however, are not transformed malignant cells but normal T cells infiltrating the tumor area. In the United States, NPC has been diagnosed in younger patients as lymphoepithelioma. In adults, NPC most commonly occurs as SCC. Although integration of EBV into the cancer cell genome can be frequently documented, the SCC variant appears to be linked to smoking and alcohol use. Thus, the etiology for adult NPC in the United States is similar to that for other SCCs of the head and neck.

Squamous Cell Carcinoma of the Head and Neck

DIAGNOSIS

Clinical Features

Early disease The signs and symptoms of early-stage head and neck cancer are usually vague and differ with the specific site of the primary disease. Patients whose tumors arise from the mucosa of the oral cavity can present with nonhealing ulcers, ill-fitting dentures, nonspecific pain, and slurred speech. Patients with tumors in the oropharynx can present with vague symptoms of unilateral pain and otalgia. Patients with tumors of the larynx can present with hoarseness even at an early stage. Thus, nonresolving hoarseness is one of the potential early warning signs of cancer and should be pursued if it persists longer than 2 to 4 weeks.

Patients with nasopharyngeal cancer present with classic symptoms of otalgia (caused by blockage of the eustachian tube), otitis media, and nasal obstruction with or without nosebleeding. Patients in more advanced stages will present with lymphadenopathy of the neck, which is usually bulky and bilateral, and cranial neuropathies. Patients with tumors of the paranasal sinuses can present with vague symptoms of pain and pressure that will advance to cranial neuropathies, ocular symptoms, or both.

Advanced disease Advanced tumors of any sites involved in head and neck cancer can cause massive unilateral or bilateral lymphadenopathy. Slurred speech, bleeding, severe pain, otalgia, trismus, dysphagia, odynophagia, ocular symptoms, cranial neuropathies, stridor, and airway obstruction can all be observed in these patients.

Examination

The evaluation of a patient for possible head and neck cancer should be guided by a high index of suspicion, particularly if the patient has a known history of tobacco or alcohol use. A thorough inspection of the oral cavity and all visible mucosa should be part of any general examination. If leukoplakia, erythroplakia, or any masses are identified, they should be biopsied. As part of the history, the patient should be questioned about pain, ill-fitting dentures, bleeding, or hoarseness. A symptomatic patient should undergo further evaluation by an otolaryngologist, who uses a mirror or an endoscope to visualize the pharynx and nasopharynx, oral cavity, larynx, and esophagus. If lesions are identified, further evaluation proceeds with the patient under anesthesia, and biopsy samples are obtained from all suspicious

lesions. It is advisable to examine the esophagus and respiratory tract because of the high incidence of second primary tumors (panendoscopy) occurring in this region. CT or MRI scanning establishes the extent of locoregional involvement.

A somewhat common clinical situation is that of a patient presenting with lymphadenopathy of the head and neck area but with no detectable primary site on initial examination. Evaluation of such patients should proceed with a fine-needle aspiration of the neck node, which frequently yields a diagnosis of SCC. Further evaluation should include a panendoscopy with the patient under anesthesia to permit biopsy of sites where primary tumors frequently occur—namely, the tonsils (the biopsy entailing a tonsillectomy), the base of the tongue, the nasopharynx, and the piriform sinus. If evaluation reveals no primary site, the cancer is staged as TX, and further management is performed according to the exact nodal stage. The evaluative process then focuses on locating primary tumors outside of the head and neck region (e.g.., lung and gastrointestinal tract).

STAGING

Head and neck cancers are staged by the TNM staging classification.[17] In general, T1 to T3 describes an increasing size of the primary site, and T4 involves invasion of an additional structure, such as bone, cartilage, or the base of the tongue. Neck lymph node involvement is classified as N1, N2, or N3 on the basis of the number, location, and size of involved lymph nodes. Staging is consistent for all anatomic sites except nasopharyngeal cancer, where location of the node is taken into consideration**.**

Head and neck cancer can be categorized into three major groups: (1) early disease (stages I and II), which is characterized by small primary T1 or T2 lesions with no clinically detectable lymph node involvement; (2) locoregionally advanced disease (stages III and IV), which is characterized by large primary tumors that may or may not be resectable and possibly by significant clinical lymph node involvement, but without detectable distant metastases; and (3) recurrent or metastatic disease, which is characterized by local or regional tumor recurrence after surgery or radiation and by the identification of metastases distant from the primary site. The 2002 American Joint Committee on Cancer (AJCC) Cancer Staging Manual contains important changes in staging classifications—notably, it includes a uniform description of advanced tumors (T4 divided into stage T4a, or advanced resectable disease, and stage T4b, or advanced unresectable disease) and metastatic disease (stage T4c).[17] The newer classifications are intended to identify patients who are candidates for curative therapy.

Unlike many other malignancies, head and neck cancer is predominantly a local and regional disease. Even patients with clinically advanced stages are rarely found to have distant metastases at the time of initial diagnosis. Approximately 30% of patients present with early disease, 65% with locoregionally advanced disease, and 5% with distant metastases. However, autopsy studies have suggested that a much higher percentage of patients have distant metastases.[18] The most common organs to be involved with metastases are the lungs, bones, and liver.

Recurrent disease usually occurs within the head and neck region. Thus, most therapeutic interventions in head and neck cancer are aimed at eradication of locoregional disease.

TREATMENT

Early Disease

Early-stage head and neck cancer is treated with curative intent. Approximately 60% to 90% of patients can be expected to be cured.[19,20]

In general, patients undergo either surgery or radiation therapy, and patients who develop subsequent locoregional disease may be treated with the alternative therapeutic modality. The exact choice of primary modality depends on the location of the disease. For early-stage laryngeal or nasopharyngeal cancer, the use of radiotherapy is generally preferred. Radiotherapy will result in few side effects and allows for preservation of a functional voice. Surgery can be used for disease recurrence. On the other hand, a small lesion in the oral cavity might be easily excised with no loss of function, whereas radiotherapy to the oral cavity would require a 6-week course of treatment and carry long-term sequelae of xerostomia, loss of taste, and accelerated dental decay.

Patients cured of head and neck cancer (all stages) are known to be at high risk for second malignancies.[21] Thus, such patients remain at significant risk (estimated at about 3% to 5% a year) of a second smoking-related or alcohol-related malignancy in the upper aerodigestive tract. Finally, because most patients are elderly, with the median age of onset being 60 years, they are at risk for significant medical comorbidity and require supervision by an internist or medical oncologist.[22]

Locoregionally Advanced Disease

The therapeutic goal in locoregionally advanced head and neck cancer is to control and cure disease while preserving organs and normal function.

Historically, standard treatment for locoregionally advanced disease consisted of surgery followed by postoperative radiotherapy (or radiotherapy alone for unresectable disease, such as some advanced tongue tumors or nasopharyngeal tumors). By using this approach, cure rates were consistently less than 30% for stage IV patients.[23] In addition, the treatment of these tumors required extensive surgical procedures resulting in significant subsequent organ dysfunction. In particular, total laryngectomy and total or partial glossectomy resulted in significant alteration of function; in addition, the cosmetic and psychological sequelae resulting from these procedures or the frequently utilized radical neck dissection were of great significance. Thus, two treatment goals could readily be defined: increased cure rates and decreased treatment-related morbidity.

To improve treatment outcomes for locoregionally advanced head and neck cancer, chemotherapy has been added to the treatment regimen of surgery and radiation, most commonly in the form of induction chemotherapy.[24] Concomitant (concurrent) chemoradiotherapy is another widely used therapeutic approach that has been examined in numerous randomized trials. As studies have evolved over the past two decades, it has become evident that there are advantages to both induction chemotherapy and concomitant chemoradiotherapy.[25] This perception has led to the combination of the two approaches.

Induction chemotherapy The administration of two or three cycles of chemotherapy before surgery and radiotherapy, usually utilizing cisplatin and fluorouracil, has been shown to result in tumor shrinkage in as many as 90% of patients.[24] Nevertheless, randomized trials have demonstrated that overall survival rates are not increased through the use of induction chemotherapy, because local and regional control is not better than that achieved with surgery and radiation alone.[26] The addition of a taxane to induction regimens has resulted in increased 2-year

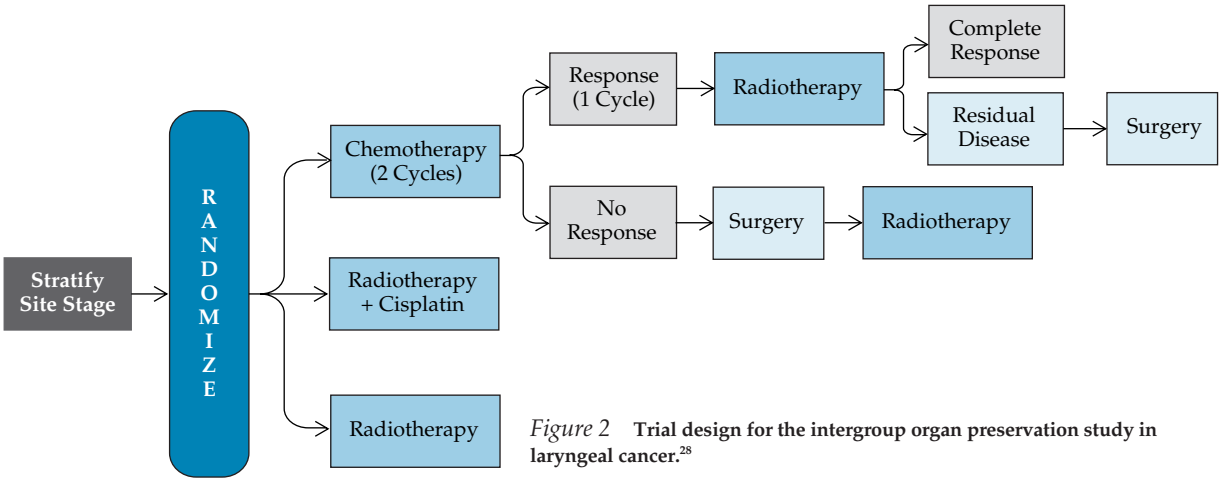


Figure 2 **Trial design for the intergroup organ preservation study in laryngeal cancer.[28]**

survival rates; however, a meta-analysis of six trials could not determine whether improved survival rates were the result of the pharmacologic effect of the taxane or that of uncontrolled prognostic factors across the various trials.[27]

Two findings of interest have been consistently reported. First, patients receiving induction chemotherapy have a lower incidence of disease progression outside of the head and neck area.[24] Thus, induction chemotherapy is able to successfully eradicate micrometastatic systemic disease. However, this result does not translate into increased survival rates, because local and regional failure is not influenced. Second, organ preservation of the larynx can be achieved. Randomized studies have specifically evaluated the use of induction chemotherapy with the goal of larynx preservation.[28-30] In these studies, patients received either surgery followed by radiotherapy or chemotherapy followed by radiotherapy; the latter patients underwent larynx resection only if they failed to respond to initial chemotherapy or had persistent or recurrent disease after completion of radiotherapy [*see Figure 2*]. Interestingly, induction chemotherapy did not increase survival rates but did allow for larynx preservation in approximately two thirds of surviving patients. In addition, patients with a preserved larynx reported less pain and better quality of life than patients who underwent surgery.[29] On the basis of these data, induction chemotherapy can be considered a standard treatment option for patients with locoregionally advanced laryngeal or hypopharyngeal cancer. Until recently, induction chemotherapy followed by radiation therapy was the standard of care for organ preservation[31]; however, the findings of a large randomized multicenter trial indicate that concomitant chemoradiotherapy is superior to induction chemotherapy for organ preservation.[32]

Concomitant chemoradiotherapy Concomitant (concurrent) chemoradiotherapy is an attractive concept for treatment of head and neck cancer. In such cases, chemotherapy is utilized primarily to sensitize tumor cells to radiotherapy and increase locoregional control. In addition, it is possible that chemotherapy can decrease systemic failure rates, although the doses and schedules of chemotherapy administered during radiotherapy may not be sufficient to achieve that goal.[24] Several randomized trials have been published in which patients received either radiotherapy alone or concomitant chemoradiotherapy (usually radiation plus cisplatin and fluorouracil). The patients treated with concomitant chemoradiotherapy had a statistically significant increase in disease-free survival and, in some studies, overall survival. These studies suggest that 3-year survival rates may be increased by as much as 20% (i.e., from 30% to 50%) through the use of concomitant chemoradiotherapy[33-37]; and they are also supported by large meta-analyses that evaluated approximately 15 trials investigating concomitant chemoradiotherapy during the 1970s and 1980s that typically involved administration of a single chemotherapeutic agent, such as bleomycin, methotrexate, mitomycin-C, or cisplatin. The meta-analysis of these older trials show a highly positive response rate, suggesting an 8% gain in absolute survival at 5 years.[38,39]

Randomized trials and a meta-analysis of three phase II trials also support concomitant chemoradiotherapy for advanced nasopharyngeal cancer.[36,40,41] On the basis of findings of the randomized trials and meta-analysis, concomitant chemoradiotherapy can be considered as a standard treatment option for the majority of patients with head and neck cancer. In addition, evidence indicates that concomitant chemoradiotherapy is viable as an organ-preserving strategy.[32] It is clear that patients achieving a complete response with concomitant chemoradiotherapy do not require surgery.[42] Therefore, this treatment modality may provide both an increased survival rate and an improved organ-preservation rate.

The findings of a 9-year multicenter study suggested that locoregional and distant recurrence rates may be dependent on treatment strategy. In that study, intensive induction chemotherapy followed by split-course chemoradiotherapy had a 13% distant failure rate, as compared with a 22% rate with chemoradiotherapy alone.[43] The study concluded that randomized clinical trials combining induction chemotherapy with concurrent chemoradiotherapy are warranted to determine whether this sequential approach is more effective in decreasing distant metastases and thus increasing overall survival rates.

Gene therapy Because head and neck cancer is a locoregional disease for most of its natural history, gene therapy has been applied to local tumor masses in an effort to reconstitute normal *p53* cellular status. Pilot data indicate that when a normal *p53* gene linked to an adenovirus is injected into locoregional tumors, tumor shrinkage can be observed, at least in some patients.[44]

Recurrent and Metastatic Disease

Patients with recurrent or metastatic disease are treated with palliative intent.[45] Curative surgery and radiotherapy are rarely

options. Thus, chemotherapy is considered standard for this group of patients. Overall response rates to current combination-chemotherapy regimens such as cisplatin and fluorouracil, cisplatin and paclitaxel, and cisplatin and docetaxel are approximately 30% to 40%.[46,47] In addition, quality of life is improved in patients receiving chemotherapy. However, the overall survival rate remains poor, with a median survival time of approximately 6 months and a 1-year survival rate of less than 30%.[48] Thus, the evaluation of new chemotherapeutic agents and other novel treatment approaches should have high priority.[49]

Salivary Gland Tumors

Salivary gland tumors arise either from one of the three major salivary glands or from the minor salivary glands lining the mucosa of the upper aerodigestive tract. Histologically, they represent very heterogeneous tumors, including mucoepidermoid cancers, pleomorphic adenoma, and adenoid cystic carcinomas, which are the more frequent specific tumor types. In a 35-year study of 2,807 patients with salivary gland tumors, 54% had benign tumors, the majority of which were pleomorphic adenomas.[50] The clinical manifestations of salivary gland tumors and diagnostic considerations are determined in part by the location of the tumor. Parotid tumors and submandibular gland tumors generally present as a solitary, asymptomatic mass. Minor salivary gland tumors arising in the oral cavity frequently present as a painless submucosal mass in the palate, lips, or buccal mucosa. Fine-needle aspiration biopsy can be used to distinguish nonneoplastic from malignant parotid masses. Its accuracy varies depending on the skill of the clinician; in experienced hands, accurate cytologic diagnosis approaches 100%.[51]

Treatment of salivary gland tumors is predominantly surgical, which leads to a proper definition of local and regional tumor stage. Surgery combined with radiotherapy, either adjuvant or postoperative, appears to increase survival.[52,53] For unresectable salivary gland tumors, neutron irradiation can be used instead of conventional radiotherapy. Chemotherapy is reserved for patients with recurrent or metastatic disease.

Everett E. Vokes, M.D., is a consultant to Bristol-Meyers Squibb., Eli Lilly and Co., Aventis Pharmaceuticals, Inc., Genentech, Inc., AstraZeneca PLC, and OSI Pharmaceuticals, Inc.

References

1. Jemal A, Tiwari RC, Murray T, et al: Cancer statistics, 2004. CA Cancer J Clin 54:8, 2004

2. Pytynia KB, Grant JR, Etzel CJ, et al: Matched-pair analysis of survival of never smokers and ever smokers with squamous cell carcinoma of the head and neck. J Clin Oncol 22:3981, 2004

3. Talamini R, Bosetti C, La Vecchia C, et al: Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. Cancer Causes Control 13:957, 2002

4. Pathmanathan R, Prasad U, Sadler R, et al: Clonal proliferations of cells infected with Epstein-Barr virus in preinvasive lesions related to nasopharyngeal carcinoma. N Engl J Med 333:693, 1995

5. Chien YC, Chen JY, Liu MY, et al: Serologic markers of Epstein-Barr virus infection and nasopharyngeal carcinoma in Taiwanese men. N Engl J Med 345:1877, 2001

6. Smith EM, Ritchie JM, Summersgill KF, et al: Human papillomavirus in oral exfoliated cells and risk of head and neck cancer. J Natl Cancer Inst 96:449, 2004

7. Dietz A, Ramroth H, Urban T, et al: Exposure to cement dust, related occupational groups and laryngeal cancer risk: results of a population based case-control study. Int J Cancer 108:907, 2004

8. Parent ME, Siemiatycki J, Fritschi L: Workplace exposures and oesophageal cancer. Occup Environ Med 57:325, 2000

9. Vaughan TL: Occupation and squamous cell cancers of the pharynx and sinonasal cavity. Am J Ind Med 16:493, 1989

10. Schneider AB, Lubin J, Ron E, et al: Salivary gland tumors after childhood radiation treatment for benign conditions of the head and neck: dose-response relationships. Radiat Res 149:625, 1998

11. Key TJ, Schatzkin A, Willett WC, et al: Diet, nutrition and the prevention of cancer. Public Health Nutr 7:187, 2004

12. Hasina R, Lingen MW: Head and neck cancer: the pursuit of molecular diagnostic markers. Semin Oncol 31:718, 2004

13. Haughey BH, Gates GA, Arfken CL, et al: Meta-analysis of second malignant tumors in head and neck neck cancer: the case for an endoscopic screening protocol. An Otol Rhinol Laryngol 101:105, 1992

14. Prout MN, Sidari JN, Witzburg RA, et al: Head and neck cancer screening among 4611 tobacco users older than forty years. Otolaryngol Head Neck Surg 116:201, 1997

15. Vokes EE, Weichselbaum RR, Lippman S, et al: Head and neck cancer. N Engl J Med 328:184, 1993

16. Vokes EE, Liebowitz DN, Weichselbaum RR: Nasopharyngeal carcinoma. Lancet 350:1087, 1997

17. AJCC Cancer Staging Handbook from the AJCC Cancer Staging Manual, 6th ed. Greene FL, Page DL, Fleming ID, et al, Eds. American Joint Committee on Cancer. Springer-Verlag, New York, 2002

18. Papac RJ: Distant metastases from head and neck cancer. Cancer 15:342, 1984

19. Spector JG, Sessions DG, Chao KS, et al: Stage I (T1 N0 M0) squamous cell carcinoma of the laryngeal glottis: therapeutic results and voice preservation. Head Neck 21:707, 1999

20. Spector JG, Sessions DG, Chao KS, et al: Management of stage II (T2N0M0) glottic carcinoma by radiotherapy and conservation surgery. Head Neck 21:116, 1999

21. Spector JG, Sessions DG, Haughey BH, et al: Delayed regional metastases, distant metastases, and second primary malignancies in squamous cell carcinomas of the larynx and hypopharynx. Laryngoscope 111:1079, 2001

22. Chen AY, Matson LK, Roberts D, et al: The significance of comorbidity in advanced laryngeal cancer. Head Neck 23:566, 2001

23. Al-Sarraf M: Treatment of locally advanced head and neck cancer: historical and critical review. Cancer Control 9:387, 2002

24. Vokes EE, Athanasiadis I: Chemotherapy for head and neck cancers: the future is now. Ann Oncol 7:15, 1996

25. Posner MR, Haddad RI, Wirth L, et al: Induction chemotherapy in locally advanced squamous cell cancer of the head and neck: evolution of the sequential treatment approach. Semin Oncol 31:778, 2004

26. Munro AJ: An overview of randomized controlled trials of adjuvant chemotherapy in head and neck cancer. Br J Cancer 71:83, 1995

27. Pignon JP, Syz N, Posner M, et al: Adjusting for patient selection suggests the addition of docetaxel to 5-fluorouracil-cisplatin induction therapy may offer survival in squamous cell cancer of the head and neck. Anticancer Drugs 15:331, 2004

28. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. Department of Veterans Affairs Laryngeal Cancer Study Group. N Engl J Med 324:1685, 1991

29. Terrell JE, Fisher SG, Wolf GT: Long-term quality of life after treatment of laryngeal cancer. The Veterans Affairs Laryngeal Cancer Study Group. Arch Otolaryngol Head Neck Surg 124:964, 1998

30. Lefebvre JL, Chevalier D, Luboinski B, et al: Larynx preservation in pyriform sinus cancer: preliminary results of a European organization for research and treatment of cancer phase III trial. J Natl Cancer Inst 88:890, 1996

31. Vokes EE, Stenson KM: Therapeutic options for laryngeal cancer. N Engl J Med 349:2087, 2003

32. Forastiere AA, Goepfert H, Maor M, et al: Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. N Engl J Med 349:2091, 2003

33. Merlano M, Benasso M, Corvo R, et al: Five-year update of a randomized trial of alternating radiotherapy and chemotherapy compared with radiotherapy alone in treatment of unresectable squamous cell carcinoma of the head and neck. J Natl Cancer Inst 88:583, 1996

34. Wendt TG, Grabenbauer GG, Rodel CM, et al: Simultaneous radiochemotherapy versus radiotherapy alone in advanced head and neck cancer: a randomized multicenter study. J Clin Oncol 16:1318, 1998

35. Brizel DM, Albers ME, Fisher SR, et al: Hyperfractionated irradiation with or without concurrent chemotherapy for locally advanced head and neck cancer. N Engl J Med 338:1798, 1998

36. Al-Sarraf M, LeBlanc M, Giri PGS, et al: Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized intergroup study 0099. J Clin Oncol 16:1310, 1998

37. Calais G, Alfonsi M, Bardet E, et al: Randomized trial of radiation therapy versus concomitant chemotherapy and radiation therapy for advanced-stage oropharynx carcinoma. J Natl Cancer Inst 91:2081, 1999

38. El-Sayed S, Nelson N: Adjuvant and adjunctive chemotherapy in the management of squamous cell carcinoma of the head and neck region: a meta-analysis of prospective and randomized trials. J Clin Oncol 14:838, 1996

39. Pignon JP, Bourhis J, Domenge C, et al: Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: three meta-analyses of updated individual data. MACH-NC Collaborative Group. Meta-Analysis of Chemotherapy on Head and Neck Cancer. Lancet 355:949, 2000

40. Chua DT, Ma J, Sham JS, et al: Long-term survival after cisplatin-based induction chemotherapy and radiotherapy for nasopharyngeal carcinoma: a pooled data analysis of two phase III trials. J Clin Oncol 23:1118, 2005

41. Argiris A, Haraf DJ, Kies MS, et al: Intensive concurrent chemoradiotherapy for

head and neck cancer with 5-fluorouracil- and hydroxyurea-based regimens: reversing a pattern of failure. Oncologist 8:350, 2003

42. Vokes EE, Kies MS, Haraf DJ, et al: Concomitant chemoradiotherapy as primary therapy for locoregionally advanced head and neck cancer. J Clin Oncol 18:1652, 2000

43. Brockstein B, Haraf DJ, Rademaker AW, et al: Patterns of failure, prognostic factors and survival in locoregionally advanced head and neck cancer treated with concomitant chemoradiotherapy: a 9-year, 337-patient, multi-institutional experience. Ann Oncol 15:1179, 2004

44. Clayman GL, El-Naggar AK, Lippman SM, et al: Adenovirus-mediated *p53* gene transfer in patients with advanced recurrent head and neck squamous cell carcinoma. J Clin Oncol 16:2221, 1998

45. Ganly I, Kaye SB: Recurrent squamous-cell carcinoma of the head and neck: overview of current therapy and future prospects. Ann Oncol 11:11, 2000

46. Schrijvers D, Johnson J, Jiminez U, et al: Phase III trial of modulation of cisplatin/fluorouracil chemotherapy by interferon alfa-2b in patients with recurrent or metastatic head and neck cancer. J Clin Oncol 16:1054, 1998

47. Forastiere AA, Neuberg D, Taylor SG IV, et al: Phase II evaluation of Taxol in advanced head and neck cancer: an Eastern Cooperative Oncology Group trial. Monogr NCI 15:181, 1993

48. Argiris A, Li Y, Forastiere A: Prognostic factors and long-term survivorship in patients with recurrent or metastatic carcinoma of the head and neck. Cancer 101:2222, 2004

49. Brockstein B, Haraf D, Stenson K, et al: A phase I study of concomitant chemoradiotherapy with paclitaxel, 5-FU, and hydroxyurea with granulocyte colony stimulating factor support for patients with poor prognosis cancer of the head and neck. J Clin Oncol 16:735, 1998

50. Spiro RH: Salivary neoplasms: overview of a 35-year experience with 2,807 patients. Head Neck Surg 8:177, 1986

51. Stewart CF, MacKenzie K, McGarry GW, et al: Fine-needle aspiration cytology of salivary gland: a review of 341 cases. Diagn Cytopathol 22:139, 2000

52. Terhaard CH, Lubsen H, Van der Tweel I, et al: Salivary gland carcinoma: independent factors for locoregional control, distant metastases, and overall survival: results of the Dutch Head and Neck Oncology Cooperative Group. Head Neck 26:681, 2004

53. Mendenhall WM, Morris CG, Amdur RJ, et al: Radiotherapy alone or combined with surgery for salivary gland carcinoma.

Acknowledgments

Figure 1 Tom Moore.

Figure 2 Seward Hung.

II DEPRESSION AND BIPOLAR DISORDER

Michael T. Compton, m.d., m.p.h.
Charles B. Nemeroff, m.d., ph.d.

Definition

Depression and bipolar disorder are two forms of psychiatric illness that lie at opposite ends of the spectrum of affect [*see Figure 1*]. Both are characterized by disturbances in feelings, thoughts, and behaviors and, consequently, by considerable impairment in functioning.

Mood disorders are divided on the basis of polarity. Unipolar depression is characterized by depressive episodes only, whereas bipolar disorder (formerly termed manic-depressive illness) is marked by hypomanic, manic, or mixed episodes, often in addition to depressive episodes. The most commonly utilized classifications rely on this division, including the two gold standards in the field, the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th edition of the *International Classification of Diseases and Related Health Problems* (ICD-10).

Mood disorders (also known as affective disorders) are syndromes consisting of signs and symptoms that are clearly abnormal for the individual experiencing them. Episodes tend to occur in periodic fashion, with varying degrees of residual symptoms between them. Mood disorders are distinguished from other major mental illnesses, such as schizophrenia, on the basis of the presence or absence of specific clusters of symptoms and the longitudinal course of the illness. Many patients with mood disorders fulfill the diagnostic criteria for other psychiatric disorders, such as anxiety disorders, substance-use disorders, and personality disorders.

Depressive Disorders

EPIDEMIOLOGY

The most commonly cited prevalence estimates for psychiatric disorders in the United States are derived from the 1980–1985 Epidemiologic Catchment Area (ECA) study and the 1990–1992 National Comorbidity Survey (NCS).[1,2] The ECA found the 1-year prevalence of major depression and dysthymia (mild, chronic depression) in adults to be 5.0% and 5.4%, respectively. NCS 1-year prevalence rates were 10.3% for major depressive episode and 2.5% for dysthymia. Of note, 1-year prevalence rates for major depressive episode and dysthymia were 12.9% and 3.0%, respectively, in females and 7.7% and 2.1% in males.[2] Subsyndromal symptomatic depression is defined as a depressive state characterized by two or more symptoms of depression of the same quality as that which occurs in major depression, excluding depressed mood or anhedonia.[3] The 1-year prevalence of subsyndromal symptomatic depression in the general population has been estimated to be 11.8%.[4] In a nationally representative sample of adolescents and young adults 15 to 24 years of age from the NCS, the 1-year prevalence of major depression was found to be 12.4%.[5] Of considerable concern was the observation that 21.9% of those adolescents and young adults with major depression reported a suicide attempt at some point in their lives. In children, the prevalence rates of major depression are approximately the same in girls and boys, whereas in adolescents, the female-to-male ratio is approximately 2:1,[6] which is consistent with the adult gender ratio. The NCS found that the lifetime prevalence rates for a major depressive episode were 12.7% and 21.3% for men and women, respectively; the lifetime prevalence rates for dysthymia were 4.8% in men and 8.0% in women.

Epidemiologic studies reveal that mood disorders are more prevalent in persons younger than 45 years; the average age of onset is between 20 and 40 years. The age of onset of unipolar depression has decreased dramatically over the past several decades. The incidence of depressive episodes, but not bipolar disorder, is about twice as high in women as in men—a sex difference that begins in early adolescence and persists into midlife. The prevalence of affective disorders does not vary significantly by race or ethnicity. Major depression appears to be more prevalent in urban residents; in single, divorced, or widowed people; in patients seen in primary care settings; in patients with comorbid medical illness; in unemployed people; in those with a history of traumatic early childhood experiences (e.g., loss of parents at an early age, sexual or physical abuse); and in those with perceived social stresses. Affective disorders in general are also more prevalent in persons with a family history of affective disorders, suicide, or alcoholism.

Depression is widespread in the elderly. Depression in late life is a serious public health concern; comorbidity of depression with other illnesses, both medical and psychiatric, is particularly problematic in older persons.[7] The prevalence of depressive symptoms in those 65 years of age and older has been estimated to be 16.9%.[8]

According to some estimates, depression ranks as the number one cause of disability worldwide, causing twice as much disability as the second leading cause, which is iron deficiency anemia. Disability from mood disorders often exceeds that from other chronic and debilitating diseases, such as hypertension, arthritis, and diabetes, causing more days in bed and higher numbers of missed days from work.

ETIOLOGY AND GENETICS

The strongest known risk factors for the development of depression are family history and previous episodes of depression. The risk of depressive disorders in first-degree relatives of patients with depression is two to three times that of the general population. If one parent has a mood disorder, a child's risk of a mood disorder is 10% to 25%; if both parents are affected, the risk roughly doubles. The importance of genetic factors in pathogenesis is also highlighted by the higher concordance rate of depression in monozygotic twins than in dizygotic twins. The likelihood that a person will develop a mood disorder is influenced by a number of factors, including gender (women are at greater risk than men), early parental loss, inadequate rearing by parents (e.g., neglect), a history of prepubertal traumatic events, certain personality characteristics, personal or family history of mood disorders, the extent of social support, and recent stressful life events.

Psychological models for the etiology of mood disorders, especially depressive disorders, have also been proposed. Some of these theories have focused on the influence of early life ex-

periences—a view supported by considerable recent neurobiologic research. Other schools of thought have advanced behavioral and cognitive factors as preeminent in the etiology of mood disorders.

Contemporary psychiatry views the etiology of depression as a multifactorial process with a definite biologic basis, though social and psychological influences are clearly important. The genetics of unipolar depression and many other psychiatric illnesses is an area of active research. Depression is undoubtedly a genetically complex disease and one that, like many medical disorders, results from an interaction between genetic susceptibility and environmental factors.[9] For example, current research suggests that the likelihood of developing depression in response to stressful life events is moderated by genotype at a functional polymorphism in the promoter region of the serotonin transporter gene.[10]

PATHOGENESIS

No single physiologic mechanism for mood disorders has been identified. Nonetheless, considerable evidence suggests the involvement of several neurotransmitter systems, especially the monoamines (noradrenergic, serotoninergic, and perhaps dopaminergic pathways), as well as abnormalities in the hypothalamic-pituitary-adrenal (HPA) and hypothalamic-pituitary-thyroid (HPT) axes. Noradrenergic systems are involved in mood, arousal, appetite, reward, and drives; serotoninergic systems are involved in mood, sleep, and impulsivity; and dopaminergic systems are important for pleasure and reward, as well as for psychomotor activity. These monoamine systems arise in cell groups in the brain stem and are widely dispersed throughout the forebrain. Chronic hyperactivity of the HPA axis is arguably the biologic finding that is most often reported in severe and psychotic depression, and it normalizes with clinical recovery. The pathogenesis of depression in relation to biogenic amines and the HPA axis is an area of intensive research.[11]

Biogenic Amine Hypotheses

It has been posited that mood disorders are caused by relative deficiencies in the availability of the catecholamine norepinephrine (NE) or the indoleamine serotonin (5-hydroxytryptamine [5-HT]) within the central nervous system, perhaps involving alterations in receptor function, signal transduction, or both.[12,13] Findings related to the adrenergic system include alterations in cerebrospinal fluid concentrations of 3-methoxy-4-hydroxyphenylglycol (MHPG), which is the major metabolite of CNS NE, and alterations in NE availability. Several subtypes of adrenergic receptors have also been studied, and several alterations have been noted. These include an increase in the binding of alpha$_2$-adrenergic receptors in platelets; a blunting of the growth hormone response to clonidine (an alpha$_2$ receptor agonist); and an increase in the density of beta-adrenergic receptors in the brains of suicide victims post mortem. Furthermore, a multitude of studies have documented downregulation of beta-adrenergic receptors after long-term treatment with antidepressants—particularly monoamine oxidase inhibitors (MAOIs) and tricyclic antidepressants (TCAs)—and after electroconvulsive therapy (ECT). During recent decades, close scrutiny of the serotoninergic system in the brain has revealed alterations in various measures of serotoninergic neural activity and integrity, as well as 5-HT receptor alterations, in patients with depression. The CSF concentration of the serotonin metabolite 5-hydoxyindoleacetic acid (5-HIAA) has repeatedly been reported to be decreased in depressed patients, especially those who are suicidal. Alterations in 5-HT transporters and receptors have also been studied; such alterations include an increase in the density of postsynaptic 5-HT$_2$ receptors and a decrease in the density of 5-HT transporter sites. The response to fenfluramine, an agent that releases 5-HT from the presynaptic terminal, is diminished in some patients with depression.

Neuroendocrine and Neuropeptide Hypotheses

Other hypotheses have focused on the multiple endocrine alterations reported to be present in patients with mood disorders, most prominently in the HPA[14] and HPT axes. Several of the hierarchical components of neuroendocrine axes may be involved in the pathophysiology of mood disorders. In the HPA axis, corticotropin-releasing factor (CRF) from the hypothalamus stimulates the release of adrenocorticotropic hormone (ACTH) from the anterior pituitary, which in turn regulates the secretion of glucocorticoids from the adrenal cortex (cortisol being the most important hormone released from the adrenal gland). Corticosteroid receptor signaling is impaired in patients with major depression,[15] and studies document marked CRF neuronal hyperactivity in depressed patients.[16,17] Several measures of HPA-axis hyperactivity have been reported in depressed patients. These include (1) increased concentrations of CRF in CSF, (2) a blunted ACTH response to exogenous CRF administration, (3) decreased CRF receptor density in the frontal cortex of suicide victims, (4) pituitary and adrenal gland enlargement, (5) increased ACTH and cortisol production, and (6) nonsuppression of plasma cortisol levels after the administration of dexamethasone (the dexamethasone suppression test).

Hypothyroidism has long been known to be frequently associated with a markedly depressed mood. The HPT axis involves the secretion of thyrotropin-releasing hormone (TRH) by the hypothalamus, which stimulates the anterior pituitary gland to release thyroid-stimulating hormone (TSH). This in turn stimulates the secretion of thyroxine and triiodothyronine from the thyroid gland. A multitude of alterations in the activity of the HPT axis have been observed in patients with depression. Such abnormalities include increased TRH concentrations in CSF, blunted or exaggerated TSH response to TRH stimulation, decreased nocturnal plasma TSH concentrations, and the presence of antimicrosomal thyroid or antithyroglobulin antibodies. Mood disorders have also been reported to be associated with alterations in the activity of the growth hormone axis and the hypothalamic-pituitary-gonadal (HPG) axis. The burgeoning field of psychoneuroimmunology has provided evidence of abnormal immune function, including alterations in cytokine secretion, in depressed patients.

Neuroimaging and Sleep Studies

Brain imaging techniques, including both structural and functional magnetic resonance imaging and positron emission tomography, are being increasingly used to study CNS pathophysiology in patients with affective disorders.[18,19] These neuroimaging studies show clear abnormalities in the activity of several circuits in depressed patients, including decreased metabolism in the frontal cortex and increased metabolism in certain limbic areas.[20] Sleep abnormalities in depression include increased time to sleep (initial insomnia), earlier onset of rapid eye movement (REM) sleep (decreased REM latency),[21] reduced sleep stages 3 and 4, changes in phasic REM activity, disturbances in the continuity of sleep, and early morning awakenings.

Clinical Manifestations

A major depressive episode consists of at least a 2-week period during which five or more specific symptoms are experienced, representing a change from previous functioning. According to DSM-IV, at least one of these symptoms must be either depressed mood or loss of interest or pleasure (anhedonia). Other potential symptoms include significant change in appetite or body weight, insomnia or excessive sleeping, psychomotor agitation or retardation, fatigue or loss of energy, feelings of worthlessness or inappropriate guilt, poor concentration or indecisiveness, and thoughts of death or suicidal ideation. Patients with a major depressive episode often experience negative cognitive patterns, including helplessness, hopelessness, and preoccupation with inadequacy. Low self-esteem and overwhelming guilt or self-reproach are commonly present. When the episode is accompanied by psychotic features, the psychotic symptoms tend to be mood-congruent—for example, the patient may experience nihilistic delusions or auditory hallucinations of derogatory commentary. Distinguishing between psychotic depression and schizoaffective disorder is important, because the two entities confer different prognoses[22] and have different treatments.

Patients experiencing depressive episodes sometimes have features described as melancholic; these include early morning awakenings, loss of appetite, marked psychomotor changes, diurnal mood variation (i.e., the mood is worse in the morning than later in the day), and unremitting anhedonia. Melancholic depression may be more severe and more likely to include biologic abnormalities; it may be most appropriately treated with pharmacotherapy or ECT. The term atypical depression describes a syndrome of prominent mood reactivity, reversed neurovegetative symptoms (increased appetite and excessive sleep rather than the typical pattern of neurovegetative symptoms that includes decreased appetite and insomnia), considerable anxiety, reversed diurnal mood variation (i.e., symptoms are more severe in the evening), prominent fatigue, and extreme sensitivity to rejection.

Other subtypes of depression include those characterized by a seasonal pattern of recurrence and, in women, worsening of depression during the premenstrual phase. In as many as 10% of patients with depression, episodes recur in a seasonal pattern, and seasonal affective disorder is an increasingly recognized subtype of depression.[23,24] Premenstrual worsening of depression has been a phenomenon of increasing interest. Some women experience a severe syndrome of mood and physiologic symptoms during the luteal phase of the menstrual cycle that generally resolves shortly after the onset of menstruation. This clinical entity is known as premenstrual dysphoric disorder (PMDD).[25,26]

Unipolar depressive disorders, as defined by DSM-IV, include major depressive disorder (characterized by either a single episode or recurrent episodes) and dysthymic disorder (chronic depressive symptoms that do not meet the full criteria for a major depressive episode) [*see Figure 1*]. Dysthymic disorder is a mild form of depression that often has an early onset (i.e., in childhood, adolescence, or early adult life) and usually lasts for long periods (i.e., for at least 2 years in adults, according to DSM-IV). Dysthymic disorder is an undertreated condition,[27] despite the significant psychosocial impairment associated with the disorder. In patients with a personality disorder, the symptoms of dysthymic disorder can be difficult to disentangle from the chronic affective symptoms of the personality disorder. Major depressive disorder may arise in persons with dysthymic disorder (so-called double depression). Subsyndromal affective symptoms are also recognizable in some individuals. A thorough history and mental status examination are necessary in differentiating between clinically significant affective symptoms and normal fluctuations in mood.

Mental Status Examination

The general appearance and attitude of the patient should be noted. When depressive symptoms become severe, the patient may become quite socially withdrawn and may begin to neglect personal hygiene and grooming. The patient's attitude may reflect sadness, irritability, or restlessness. Notable behaviors may include diminished eye contact, stooped posture, hand-wringing (a sign of anxiety or agitation), and psychomotor retardation or agitation. An assessment of speech may reveal increased response latency and decreased rate and rhythm of speech. Patients may describe their mood using such terms as down, depressed, sad, blue, unhappy, empty, helpless, or hopeless. The affect (which can be thought of as an objective assessment of the patient's internal emotional state) is often sad, tearful, or anxious. Thought processes are often slower than normal. Thought content may be characterized by inappropriate guilt and self-reproach; thoughts of worthlessness, hopelessness, or helplessness; mood-congruent psychotic features (e.g., hallucinations or delusions); and suicidal thoughts or plans. Insight into the illness, as well as judgment in general, may range from impaired to good. Impaired insight may have important implications for compliance with treatment. A cognitive assessment may reveal subtle or obvious impairments in memory and concentration.

The initial manifestations of depression may include somatic complaints such as fatigue, gastrointestinal problems, pain, or sleep disturbances. A thorough medical and psychiatric review of systems, as well as a substance-use history, should be conducted to exclude other causes of the mental status examination findings or somatic complaints. A suicide risk assessment will help determine the severity of depression and the appropriate locus of care for the initiation of treatment.[28] Having patients complete a brief, self-administered screening tool may facilitate the diagnosis of depression. Such scales include the Beck Depression Inventory, the Carroll Rating Scale for Depression, and the Zung Depression Self-Report Rating Scale.

Assessment of suicide risk A vital aspect of the mental status examination is a thorough assessment of suicidality, which includes suicidal thoughts, plans, means, intent, and attempts. Practical interviewing strategies have been developed for efficiently eliciting suicidal ideation, such as the Chronological Assessment of Suicide Events approach.[29] Evaluation of current suicidality should include extent and details of planning, lethality of such plans, intent to act on them, availability of means, and preparatory plans, such as drafting a will or securing means.[30] Because a previous suicide attempt is an important predictor of future attempts, the clinician should thoroughly evaluate details of any past suicidality. Gathering collateral history from family and friends is an important aspect of a thorough assessment. Potential risk factors for suicide include family history of suicide, concomitant anxiety or panic attacks, hopelessness, low self-esteem, psychosis, aggressive or impulsive personality characteristics, comorbid substance abuse, comorbid medical conditions, inadequate treatment of depression, and poor social supports. It is very important for clinicians to document a competent suicide

risk assessment that guides subsequent clinical interventions. Specialty psychiatric care should be obtained for patients with depression complicated by suicidal ideation.

Laboratory Tests

At present, there are no validated routine laboratory tests to diagnose depression. The dexamethasone suppression test, a measure of HPA-axis activity, is not sufficiently sensitive or specific for diagnosis or for monitoring response.[31] Instead, diagnosis is made according to the descriptive clinical criteria set forth in DSM-IV.

Several laboratory tests should be ordered to help exclude other potential causes of depressive symptoms. Medical conditions that may be physiologically associated with depression include endocrinopathies (e.g., hypothyroidism, hyperthyroidism, parathyroid disorders, Cushing syndrome, or Addison disease), neurologic disorders (e.g., stroke, subcortical dementias, neurosyphilis, multiple sclerosis, Parkinson disease, neurosarcoidosis, CNS vasculitis, or HIV-associated CNS pathology), and other disorders (e.g., vitamin deficiencies, anemia, hypoglycemia or hyperglycemia, hypoxia, end-stage renal disease, lupus, or occult malignancies such as pancreatic cancer).

DIFFERENTIAL DIAGNOSIS

The initial presentation of a patient with clinically significant depression can include a broad range of symptoms, such as somatic complaints (e.g., fatigue, insomnia, anorexia, or various nonspecific somatic symptoms), emotional concerns (e.g., sadness, anxiety, or feelings of guilt), and other problems (e.g., difficulty at work, marital problems, irritability, or memory impairment). Before making a diagnosis of major depressive disorder or dysthymic disorder, the clinician should conduct a search for medical causes of the depressive disorder. Several medical conditions (see above) can mimic the symptoms of depression. Additionally, a variety of substances of abuse can be causally implicated in depression. These include alcohol, sedative-hypnotics, and cocaine. Iatrogenic causes of depression include antihypertensives that reduce central catecholamine neurotransmission (e.g., beta blockers, reserpine, methyldopa, guanethidine, and clonidine), corticosteroids, antineoplastic agents, interferon alfa, and, possibly, isotretinoin.[32,33] Whether the depression is substance induced or caused by a general medical condition, the underlying cause should be treated; if improvement in depressive symptoms fails to occur after approximately 4 to 6 weeks, the depression should be independently diagnosed and treated.[34]

After exclusion of medical causes or substance-induced depression, the clinician should consider primary psychiatric conditions associated with depressive symptoms. The distinction between unipolar depression (i.e., a major depressive episode with no history of hypomania or mania) and bipolar depression (i.e., a major depressive episode in a patient with a history of hypomania or mania) is of vital importance and significantly affects treatment decisions. In bipolar disorder, it is not uncommon for patients to experience several episodes of depression before their first manic episode; bipolar disorder should therefore always be a part of the differential diagnosis of depression. Any history of hypomanic or manic episodes should be reviewed. Patients with bipolar depression are more likely to experience a switch to hypomania or mania or an acceleration in mood cycling if they are treated with an antidepressant in the absence of concomitant treatment with a mood stabilizer.[35] In women, the incidence of depression is increased during times of hormonal change, including pregnancy and the postpartum period. In fact, postpartum depression occurs in about 13% of women after delivery.[36] Depression after miscarriage is a common but underrecognized condition.

Other psychiatric disorders must also be excluded. Primary anxiety disorders (e.g., generalized anxiety disorder, posttraumatic stress disorder, and social anxiety disorder) should be considered in the differential diagnosis of depression [*see 13:VIII Anxiety Disorders*]. In the elderly, depression with cognitive impairment that is reversed by antidepressant treatment may be an early predictor of the development of an irreversible dementia.[7] Patients with late-onset depression are more likely to have cognitive impairment and to have more deep white-matter lesions on imaging studies.[37] However, the so-called pseudodementia of depression is common and should not be mistaken for a primary dementia. The term double depression refers to a major depressive disorder superimposed on chronic dysthymia. Many patients with personality disorders can present with symptoms of depression, and an assessment of the longitudinal history of symptoms is vital in establishing the correct diagnosis. Many patients actually have two or more comorbid psychiatric disorders, such as major depression and generalized anxiety disorder.

MANAGEMENT

Depression is often undiagnosed or inadequately treated.[38] This is extraordinarily unfortunate because a number of effective treatments are available. Incorporating depression treatment programs into primary care settings improves mental health and social outcomes,[39] and the United States Preventive Services Task Force recommends screening adults for depression in clinical practices that have systems in place to ensure accurate diagnosis, effective treatment, and follow-up. These recommendations are available on the Internet (http://www.ahrq.gov/clinic/uspstf/uspsdepr.htm).[40,41]

Pharmacologic Therapy

A broad array of antidepressants is available for the treatment of depression, allowing clinicians to select a regimen on the basis of potential side-effect profiles, subtype of depression, personal or family history of response to specific medications, and comorbid psychiatric or medical conditions.

Monoamine oxidase inhibitors MAOIs, one of the first classes of antidepressants developed, block the enzyme that metabolizes biogenic amines, increasing the availability of these neurotransmitters. In the United States, three MAOIs are currently available: phenelzine, isocarboxazid, and tranylcypromine. They are rarely used because their use is complicated by side effects (including hypotension), lethality in overdose, and lack of simplicity in dosing. Patients treated with MAOIs must follow a specific tyramine-free diet because of the potential for a pharmacodynamic interaction with tyramine that can result in a hypertensive crisis. Drugs that have been reported to interact with MAOIs include carbamazepine, cyclobenzaprine, dextromethorphan, fenfluramine, certain hypoglycemics, L-tryptophan, meperidine, selective serotonin reuptake inhibitors (SSRIs), stimulants, sympathomimetics, and TCAs. Nonetheless, the MAOIs appear to be more effective than other antidepressants in the treatment of atypical depression, though their use is usually limited to psychiatrists who have experience with these agents. MAOIs are also often effective in patients with depression that is refractory to treatment with other antidepressants.

Selective MAOIs (e.g., moclobemide), which are available outside of the United States, are easier to use because of the lack of dietary constraints and drug-drug interactions. A transdermal system for delivery of selegiline, an MAOI approved by the Food and Drug Administration for the treatment of Parkinson disease, may be effective in the treatment of depression.[42,43]

Tricyclic antidepressants TCAs act at several transporters and receptors, but their antidepressant effect is likely produced by the blocking of the reuptake of NE, 5-HT, or both at their presynaptic terminals, increasing the availability of these neurotransmitters. The TCAs can be subdivided into the tertiary amines, which are dual 5-HT/NE reuptake inhibitors (e.g., amitriptyline, imipramine, and clomipramine) and the secondary amines, which are primarily NE reuptake inhibitors (e.g., desipramine and nortriptyline).

The use of TCAs is limited by their unfavorable side-effect profile (largely resulting from their anticholinergic, antiadrenergic, and antihistaminic properties); these side effects include blurred vision, dry mouth, tachycardia, constipation, urinary retention, cognitive dysfunction, postural hypotension, dizziness, sedation, weight gain, and sexual dysfunction. TCAs also have a narrow therapeutic index and are lethal in overdose (resulting in part from an inhibition of sodium channels that causes a slowing of cardiac conduction and potentially fatal arrhythmias). Efficient use of TCAs is also limited by the need to slowly titrate the dose and the need for therapeutic drug monitoring to avoid toxicity. Although these medications are still occasionally used in refractory cases, their use should be limited to experienced practitioners and to patients who can be relied on to comply with the regular electrocardiographic and blood-drug-level monitoring that is required.

Selective serotonin reuptake inhibitors and newer antidepressants The first SSRI, fluoxetine, was introduced in the United States in 1988. There are currently six SSRIs approved by the FDA: fluoxetine, paroxetine, sertraline, fluvoxamine, citalopram, and escitalopram. These agents all share the property of blocking the reuptake of 5-HT, but there is increasing evidence that they produce other effects as well.

The SSRIs are effective in treating depression and many primary anxiety disorders. They offer several important advantages over the older medications. They require minimal dose titration, and the starting dose is sometimes an effective and even optimal dose [*see Table 1*]. Perhaps most important, these medications are safe in overdose.[44] Generic formulations of fluoxetine, paroxetine, fluvoxamine, and citalopram are currently available.

The SSRIs have a relatively mild side-effect burden. The sexual side effects of these medications (i.e., delayed ejaculation, anorgasmia, and decreased libido) are being increasingly recognized and studied; sexual side effects are one of the leading causes of discontinuance or noncompliance.[45,46] Although the correlation between SSRI treatment and increased suicide risk remains controversial, the potential for increased risk has prompted the FDA to require that manufacturers of all antidepressants include in their labeling a boxed warning (http://www.fda.gov/cder/drug/antidepressants/default.htm).[47-49]

Other, newer agents are believed to act on serotoninergic and noradrenergic pathways in the brain by a variety of mechanisms, including dual 5-HT/NE reuptake blockade (venlafaxine), 5-HT_2 blockade (trazodone, nefazodone,[50] and mirtazapine[51]), and alpha$_2$ autoreceptor blockade (mirtazapine). Bupropion was originally thought to act as a dopamine and NE reuptake inhibitor, but its mechanism of action remains somewhat obscure.[52] In addition to having FDA approval for depression, bupropion is approved for smoking cessation,[53] and venlafaxine is approved for generalized anxiety disorder.[54] Venlafaxine may be particularly effective in refractory and severe depression.[55,56] The noradrenergic effects of venlafaxine emerge at higher dosages; at lower dose ranges, it is essentially an SSRI.[57] However, venlafaxine has been reported to be similar to TCAs with regard to potential overdose and in having proconvulsant and cardiac side effects.[58] Duloxetine is a 5-HT/NE reuptake inhibitor approved by the FDA for the treatment of depression and diabetic peripheral neuropathic pain.[59,60]

Table 1 Usual Starting Doses and Dose Ranges for First-Line Antidepressant Agents

Drug (Trade Name)	*Typical Daily Starting Dose*	*Typical Daily Dose Range*
Fluoxetine (Prozac)	20 mg	20–80 mg
Paroxetine (Paxil)	20 mg	20–60 mg
Paroxetine, continuous release (Paxil CR)	12.5–25 mg	50–75 mg
Sertraline (Zoloft)	50 mg	50–200 mg
Fluvoxamine (Luvox)	50 mg	100–300 mg
Citalopram (Celexa)	20 mg	20–60 mg
Escitalopram (Lexapro)	10 mg	10–20 mg
Venlafaxine, extended release (Effexor XR)	37.5 mg	75–225 mg
Bupropion, sustained release (Wellbutrin SR)	100 mg	100–200 mg b.i.d.
Nefazodone (Serzone)	100 mg	200–600 mg
Mirtazapine (Remeron)	15 mg	15–45 mg
Duloxetine (Cymbalta)	20–30 mg	40–60 mg

Clinicians using SSRIs and newer antidepressants should be familiar with their potential for pharmacokinetic drug-drug interactions resulting from the inhibition of various cytochrome P-450 microsomal hepatic enzymes.[61] Different SSRIs and some of the other, newer antidepressants inhibit different isoenzymes of the hepatic cytochrome P-450 system; this can lead to elevation in blood levels of other medications. Nefazodone inhibits cytochrome P-450 3A4 activity and therefore increases the levels of benzodiazepines and of protease inhibitors used in the treatment of HIV infection. Fluoxetine inhibits cytochrome P-450 2D6 activity and therefore increases plasma levels of propranolol and risperidone, as well as other 2D6 substrates.

Various symptoms have been associated with the abrupt discontinuance of certain SSRIs with short half-lives and of newer agents, particularly paroxetine and venlafaxine; such symptoms include dizziness, paresthesias, asthenia, myalgias, nausea, loose stools, visual disturbances, irritability, insomnia, mood worsening, electric-shock–like sensations in the upper extremities, and headache.[62,63] With all antidepressants, tapering before discontinuance is recommended, especially for agents with short half-lives.

Research targets Development of antidepressant drugs continues to be a high priority in psychiatry, as exemplified by the efforts to develop CRF_1 and neurokinin-1 (NK_1) receptor antagonists for the treatment of depression, as well as the use of glucocorticoid receptor antagonists for the treatment of psychotic depression. The future of the treatment of depression will probably include greater knowledge of features that predict treatment response, likely through genomics and functional brain imaging.

Medication selection The vast majority of patients with depression can be treated with SSRIs and other, newer antidepressants. The patient's medical history and personal or family history of response often help predict response and side effects. Like major depression, dysthymic disorder and PMDD are effectively treated with antidepressants.[64] PMDD is especially responsive to SSRIs; the SSRIs can be used intermittently in patients with PMDD, with administration of medication limited to the luteal phase (14 days premenstrually).[65-68] Research suggests that low-dose oral contraceptive pills containing drospirenone may alleviate some symptoms of PMDD in some patients, although placebo response rates are relatively high (as is true of most randomized studies of depressive symptoms).[69]

In patients who present with clinically significant anxiety as an integral component of their depressive illness, treatment often begins with an antidepressant and an anxiolytic (e.g., clonazepam or diazepam, which are long-acting benzodiazepines). Many of the newer antidepressants, including the SSRIs, are effective anxiolytics; but their anxiolytic effects, like their antidepressant effects, do not appear until several weeks after the initiation of treatment. For this reason, it is judicious in such cases to use a brief course of a benzodiazepine anxiolytic along with the antidepressant.

Depression with associated psychotic features must be treated with an antidepressant and an antipsychotic medication.[70] Atypical antipsychotic agents are currently the first-line choice; such drugs include risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole. Severe depression is most effectively treated with ECT.[71] Patients with bipolar depression should be treated with an optimal dose of a mood-stabilizing medication (e.g., lithium, valproate, carbamazepine, or an atypical antipsychotic agent) before starting therapy with an antidepressant.[72] Lamotrigine, an anticonvulsant, may be particularly effective in the treatment of bipolar depression.[73] An olanzapine-fluoxetine combination capsule (Symbyax) also has FDA approval for the treatment of bipolar depression.[74,75]

Herbal remedies and dietary supplements A number of agents have been reported to have an effect on depressive symptoms. Most herbal remedies and dietary supplements have not yet been adequately studied. Initial enthusiasm about the potential effectiveness of *Hypericum perforatum* (St. John's wort) was diminished by two double-blind, randomized multicenter trials that failed to support the efficacy of St. John's wort in major depression when compared with placebo[76] or with placebo and sertraline.[77] However, in the latter study, there was no statistical difference between sertraline and placebo in terms of main outcome measures, which is a reminder of the relatively high placebo response rate in randomized, controlled trials for depression.[78] Other herbal remedies that may hold promise for the treatment of depression, such as S-adenosyl-methionine (SAM-e) and kava, require further study before their use can be recommended.[79] Dietary supplements, including B vitamins and omega-3 fatty acids, also require additional study to substantiate reports.

Managing therapeutic response Treatment of depression is recommended for at least 4 to 6 weeks before a decision regarding efficacy can be made. Measurement of depression severity by an observer or through use of a self-report rating scale is essential for monitoring the patient's response to antidepressant therapy. Improvement may be variable, with some symptoms improving before others, and the patient may experience occasional days of increased symptoms during the course of recovery. Complete failure to respond (which may be defined as less than 25% decrease on an accepted symptom-rating scale in a patient who has received an adequate dosage for 4 weeks) may best be addressed by switching to an agent with a different mode of action or to a dual-action agent[80] (e.g., switching from an SSRI to venlafaxine, bupropion, nefazodone, mirtazapine, or duloxetine). Considerable data suggest that many patients who do not respond to one SSRI do respond to a different SSRI.

Patients who experience a partial response (i.e., a 25% to 50% decrease on rating scales after 4 to 6 weeks at an adequate dosage) may be treated with dose escalation. If full remission still does not ensue, an augmentation or combination strategy can be tried.[81] Reassessment of the diagnosis may reveal psychiatric comorbidity, the presence of depressive subtypes, or the contribution of a previously undiagnosed medical disorder or iatrogenic cause. Augmentation agents include thyroid hormone, lithium, psychostimulants (e.g., methylphenidate, dextroamphetamine, or modafinil), buspirone, atypical antipsychotics, and other antidepressants.

Nonresponse or partial response despite adequate antidepressant treatment occurs in a substantial portion of patients. The most common cause of initial treatment failure, however, is undertreatment, which can result from an insufficient duration of treatment, a subtherapeutic dosage of medication, or poor adherence to treatment.[82] Adherence should be encouraged and monitored. Factors that may affect a patient's adherence to the medication regimen include the complexity and side-effect burden of the antidepressant regimen; concurrent substance abuse; and poor or incomplete understanding of symptoms, illness, and treatment. Sexual side effects have been widely recognized as a major cause of poor compliance with SSRI regimens, particularly after improvement of depressed mood. Several of the newer antidepressants (including bupropion, nefazodone, and mirtazapine) may be much less likely to induce these troubling side effects.

Improvement of the illness should not be interpreted by the patient or the physician as an indicator that antidepressants are no longer necessary. The continuation phase of treatment consists of 16 to 20 weeks of continued treatment (using the same antidepressant that was used in the acute phase) after remission, with the goal of preventing the relapses that typically occur in untreated patients. Maintenance-phase treatment should be considered for the prevention of recurrence; the decision whether to prescribe maintenance treatment should be based on the patient's longitudinal history and the severity of illness. Maintaining antidepressants at the dosage used to treat the acute episode is an effective means of preventing recurrence,[83] the risk of which increases with each successive episode. The optimal duration of maintenance-phase treatment has not yet been determined.[28] Clinicians should base such decisions on the estimated level of the risk of recurrence. Patients who have experienced three or

more past episodes of severe major depression should probably be treated with an antidepressant indefinitely, in view of the high risk of further relapse.

Nonpharmacologic Somatic Therapy

ECT is a highly effective treatment for depression, although its use is limited by the following factors: the need for general anesthesia; the potential for short-term memory impairment; the relatively limited number of sites at which ECT is available (generally in high-population areas); and the social stigma attached to the procedure, which is related to its historical development. Thus, despite its effectiveness, ECT is often reserved for severe or refractory depression.

Patients with a recurrent seasonal pattern of depression may be effectively treated with daily exposure to bright artificial light.[84] Potential new somatic treatments are being studied, including rapid transcranial magnetic stimulation.[85] Vagus nerve stimulation, which uses a surgically implanted device, has FDA approval for the adjunctive long-term treatment of chronic or recurrent/refractory depression.[86] Stimulation-related side effects may include hoarseness, cough, neck pain, or shortness of breath.

Psychotherapy

Psychotherapy is often used in conjunction with medications for major depression or dysthymic disorder. Only two forms of psychotherapy have been shown in controlled trials to be effective in the treatment of depression: interpersonal psychotherapy (IPT) and cognitive-behavioral therapy (CBT).

IPT explores at least three interacting components: symptom formation, social and interpersonal experiences, and enduring personality patterns. IPT is a short-term structured therapy consisting of 12 to 16 sessions; the aims of therapy are to reduce the patient's depressive symptoms, to improve self-esteem, and to help the patient develop more effective strategies for dealing with social and interpersonal relations.

CBT is a highly structured, effective short-term psychotherapy that aims to correct negative thought patterns, specific dysfunctional schemas, and cognitive distortions. Basic components of CBT include didactic aspects, cognitive techniques, and behavioral techniques.

Some studies have concluded that CBT is as effective as pharmacotherapy. IPT has also repeatedly been demonstrated in controlled trials to be effective in the treatment of depression. Some studies have shown that a combination of an antidepressant and IPT or CBT is more effective than either treatment alone, both in relieving depression and in preventing relapse.[28,87-89]

Inpatient Treatment

Hospitalization is sometimes necessary for the treatment of depression. This strategy is indicated for patients with psychotic features or major comorbid medical conditions or for those who are exhibiting suicidality. Hospitalization is also indicated for patients who are unable to meet their basic needs because of functional impairment.

COMPLICATIONS

Mood disorders are present in 50% to 70% of all cases of suicide, and patients with recurrent, serious depression (i.e., depression requiring hospitalization) have an 8% to 15% suicide rate.[90] Depressive disorders are frequently associated with significant and pervasive impairments in social functioning, causing enormous personal, social, and economic burden.[91] Furthermore, depression is associated with high levels of health care system utilization and increased morbidity and mortality related to comorbid medical conditions. For example, depression is a major risk factor for the development of cardiovascular disease and for death after an index myocardial infarction.[92-94] The mortality 6 months after a myocardial infarction has been reported to be more than five times higher in depressed patients than in those without depression.[95]

PROGNOSIS

The chance of a recurrence of depression after one episode is about 50%; the chance increases to approximately 90% after three episodes. The frequency of recurrence increases as the number of past episodes increases. Recurrence is also more likely in patients in whom the onset of depression occurred at an early age.[96] Poor prognosis is predicted by chronicity and severity of symptoms. In patients who experience a first episode of a mood disorder in later life, a medical cause can often be diagnosed. Depression arising in later life is often followed by the development of dementia. Indeed, later-life depression is now considered by many to be a risk factor for the development of Alzheimer disease. Because of the often chronic nature of affective disorders, treatment beyond the acute phase is almost always necessary to prevent recurrent episodes or to decrease their intensity.

Bipolar Disorder

EPIDEMIOLOGY

The 1-year prevalence of any bipolar disorder in adults in the ECA study was 1.2%.[1] The NCS found the 1-year prevalence of a manic episode to be 1.3%.[2] Bipolar I disorder affects men and women equally.[97] Rates of bipolar disorder are consistent across diverse cultural and ethnic groups.[98] Onset usually occurs during adolescence or young adulthood. The NCS found the lifetime prevalence rate for a manic episode to be 1.6% and 1.7% for men and women, respectively.[2]

ETIOLOGY AND GENETICS

Affective disorders are considered complex genetic disorders, with multiple genes (genetic heterogeneity) likely accounting for at least 50% to 70% of the etiology. The genes that transmit the increased diathesis for affective disorders remain obscure. However, several groups have identified chromosome 18 as a site for several genes associated with bipolar disorder.[99,100] Twin, adoption, and family studies have established a clear genetic etiology for bipolar disorder. The most frequent mood disorder identified in family members of patients with bipolar disorder is major depression; this suggests some common genetic underpinnings between the two forms of affective disorder. The genetic nature of bipolar disorder is even stronger than that of unipolar depression, with a fivefold to 10-fold greater risk of illness in first-degree relatives than in the general population (in which the risk is approximately 1%).[101] The concordance rate for affective disorders in monozygotic twins is approximately 80% to 90%, many times that of dizygotic twins (approximately 15%), indicating a significant genetic factor.

PATHOGENESIS

The pathophysiology of bipolar disorder in general—and mania in particular—is far less understood than the pathophysiology of unipolar depression, in part because of the inherent diffi-

culty in studying acute mania. The available studies suggest that in patients with mania, abnormalities in the HPA axis and in neurotransmission are actually quite similar to those seen in patients with depression. This is particularly the case for patients who experience mixed episodes [*see* Diagnosis, Clinical Manifestations, *below*].

Sleep deprivation or antidepressants can induce a manic episode in persons with bipolar disorder but not in persons without bipolar disorder; these may be the first manic episodes in a case of hitherto undiagnosed bipolar disorder. Stressful life events, changes in sleep-wake cycles, and substance abuse may affect the course of illness and lengthen time to recovery.[102] Ongoing neuropsychological and neuroimaging research will surely advance the elucidation of this complex disorder.[103,104]

DIAGNOSIS

Clinical Manifestations

The diagnosis of bipolar disorder is made whenever an episode of mania occurs. A manic episode is a distinct period of abnormally elevated, expansive, or irritable mood lasting at least a week (or less if hospitalization is required). During this period, three or more specific symptoms (four or more if the mood is only irritable) must be present: inflated self-esteem or grandiosity; a decreased need for sleep; being more talkative than usual or feeling a need to keep talking; having flight of ideas or the subjective experience that one's thoughts are racing; distractibility; experiencing an increase in goal-directed activity or psychomotor agitation; and excessive involvement in pleasurable activities that have a high potential for painful consequences. Manic symptoms are often the opposite of those of depression (e.g., euphoric mood rather than depressed mood; psychomotor agitation rather than psychomotor retardation; a speeding of speech and thoughts rather than a slowing of thoughts and an increased response latency). Mood-congruent psychotic features occur in approximately 50% of manic episodes and often consist of grandiose delusions or auditory hallucinations. About a quarter of persons with mania develop hallucinations. Mania often confers special creativity and productivity; but these characteristics are usually curtailed as the episode escalates and manifestations (which may include psychotic symptoms) become severe and psychosocial impairment ensues.

A hypomanic episode, which is less severe than mania, is a distinct period of elevated, expansive, or irritable mood lasting at least 4 days, during which time three or more symptoms of a manic episode are experienced (four if the mood is only irritable). The symptoms represent an unequivocal change in functioning and are observable by others but are not severe enough to cause marked impairment or to necessitate hospitalization. No psychotic features are present. Thus, a hypomanic episode can be thought of as a mild manic episode. Research indicates that 5% to 15% of persons with hypomania will ultimately have a manic episode.

A mixed episode is a period of at least 1 week during which the patient meets the criteria for having a major depressive episode and a manic episode simultaneously. As with both major depressive episodes and manic episodes, a mixed episode is a mood disturbance sufficiently severe to cause marked impairment in functioning (such as necessitating hospitalization to prevent harm to self or others or having associated psychotic features). Patients in the mixed state are dysphoric but hyperactive, agitated, and unable to sleep.

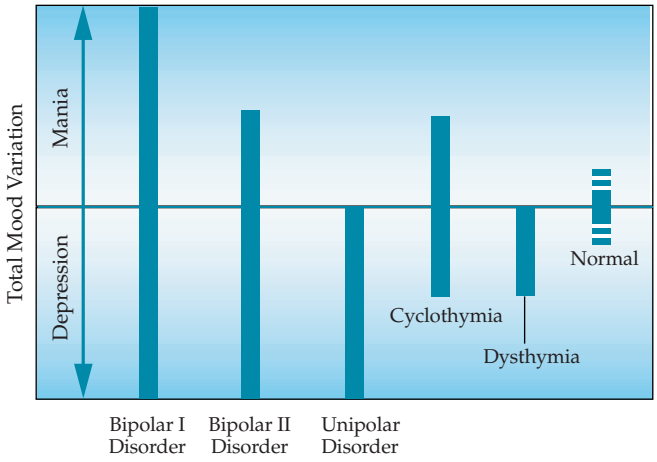


Figure 1 **The spectrum of mood disorders.**

For all four of the mood states described—mania, hypomania, depression, and mixed episode—DSM-IV specifies that the symptoms must not result from the direct physiologic effect of a substance (e.g., a drug of abuse or a medication) or a general medical condition.

Bipolar disorders include bipolar I disorder (involving the presence of one or more manic or mixed episodes, often alternating with major depressive episodes), bipolar II disorder (recurrent major depressive episodes with hypomanic episodes), and cyclothymic disorder [*see Figure 1*]. Cyclothymic disorder consists of at least 2 years of fluctuating mood disturbance involving numerous periods of hypomanic symptoms and numerous periods of depressive symptoms. Bipolar disorder is specified as rapid cycling if four or more episodes occur within a year.[105] Depressive episodes in persons with bipolar disorder are commonly marked by reversed neurovegetative symptoms and anergia (lack of energy)—so-called atypical depression. Although individual cases of unipolar mania have been reported, they are extraordinarily rare.

When psychotic symptoms are present during a manic episode, it is important to distinguish bipolar disorder from a primary psychotic disorder (e.g., schizophrenia) because the two disorders have different treatments. The diagnosis is based on the specific symptoms present and the course of the illness.

Mental Status Examination

The general appearance and attitude of the patient should be noted. During a manic episode, the patient may dress in an eccentric fashion, as exemplified by brightly colored clothing, excessive makeup, or multiple pieces of jewelry. Conversely, because of distractibility and hyperactivity, the patient's appearance may be disheveled or unkempt. The patient's attitude may reflect euphoria, excitement, intrusiveness, or irritability. Notable behaviors may include psychomotor agitation and an increase in goal-directed activity. An assessment of speech may reveal increased rate and rhythm of speech. Manic patients are commonly hyperverbal, and speech may be very rapid or even pressured. Patients may describe their mood as high or elated. The mood may be characterized as irritable in some cases. The affect is often elevated or euphoric. Thought processes may be faster than normal. Flight of ideas (described by the patient as racing thoughts) may be present, and thought disorganization

sometimes results in puns, clang associations, and loosening of associations. Thought content may be characterized by elevated self-worth, grandiosity, mood-congruent psychotic features (e.g., auditory hallucinations or delusions), and sometimes suicidal thoughts or plans. Insight into the illness—and judgment in general—may range from impaired to good but is often quite poor. As with depression, impaired insight may have important implications for compliance with treatment. A cognitive assessment may reveal subtle or obvious impairments in memory and frequent difficulties with concentration. Patients with bipolar disorder are at serious risk for suicide, both in the manic and in the depressive phases of their illness.

A thorough medical and psychiatric review of systems, as well as a substance-use history, should be conducted to exclude other causes of the mental status examination findings. A risk assessment, including assessment of the ability to adequately care for one's health and safety, excessive spending, and psychotic processes, will help determine the severity of the episode and the appropriate locus of care for the initiation of treatment. As in major depressive disorder, a thorough suicide risk assessment should be conducted [*see* Assessment of Suicide Risk, *above*].

Laboratory Tests

As is true for major depression, there are no diagnostic tests for bipolar disorder. The diagnosis is based on the presence of specific symptoms and the longitudinal course, as described in DSM-IV. During the evaluation for manic symptoms, a drug history and urine drug screen are important. Other commonly ordered tests include basic serum chemistries, thyroid function studies, and syphilis serology. Before lithium is started, baseline thyroid and kidney function studies are necessary. Blood counts and liver function are assessed before starting valproic acid–divalproex sodium or carbamazepine.

DIFFERENTIAL DIAGNOSIS

Bipolar disorder must be distinguished from other psychiatric conditions, including schizophrenia and schizoaffective disorder, personality disorders, and the abuse of certain drugs (e.g., alcohol, amphetamines, hallucinogens, and cocaine). Schizoaffective disorder is classified as a primary psychotic disorder in which a major depressive episode, a manic episode, or a mixed episode occurs and psychotic symptoms are present for at least 2 weeks in the absence of prominent mood symptoms. Certain personality disorders are characterized by chronic and pervasive rapidly fluctuating affective lability. However, many patients with cluster B personality disorders (e.g., borderline personality disorder) may also meet DSM-IV criteria for dysthymic disorder, bipolar II disorder, or cyclothymic disorder. Urine or blood toxicologic screens must be performed to differentiate primary mania from a substance-induced mood disorder with manic features. However, there is a very high rate of comorbidity between bipolar disorder and substance-use disorders, and a positive urine drug screen does not exclude the possibility of a primary bipolar disorder. In an elderly patient, the diagnosis of mania is sometimes made difficult by the overlapping of manic symptoms with other syndromes that are common in the elderly, including delirium, dementia, and medical illnesses[106] [*see CE:X Symptom Management in Palliative Medicine*].

MANAGEMENT

Immediate management of acute mania often requires hospitalization. Patients may be at risk for self-harm from erratic behavior, excessive spending, and delusional thought content. Patients with mania, a mixed episode, or bipolar depression must be evaluated thoroughly for evidence of suicidal intent.

Because bipolar disorder is a recurrent, often severely impairing psychiatric disorder, the search for effective pharmacologic treatments continues to be in the forefront of psychiatric research. The mainstays of treatment have been the traditional mood stabilizers: lithium, valproic acid–divalproex sodium, and carbamazepine [*see Table 2*]. However, the efficacy of newer agents, particularly lamotrigine and the atypical antipsychotic agents, has now been demonstrated. The precise mechanisms of action of all these mood-stabilizing medications are still unknown.

Lithium

Lithium was discovered to have antimanic properties in 1949. It is now considered the gold standard for treatment of bipolar disorder. Remarkably, the mechanism of action of lithium remains unknown, though some evidence suggests that it acts via signal transduction pathways such as the phosphatidylinositol second-messenger system. As with other mood stabilizers, full improvement with lithium may take several weeks. Lithium is most effective in classic bipolar disorder, which consists of discrete episodes of mania and depression, with symptom-free periods between episodes.[107] Continued medication use reduces the number and intensity of episodes of illness.

Renal and thyroid function should be assessed before initiating treatment with lithium, and an electrocardiogram should be obtained. Lithium is a natural element and is distributed with total body water. It is excreted unchanged into the urine. Diuretic therapy, sodium restriction, and sodium wasting increase reabsorption of lithium and thus increase serum levels. Patients taking lithium should therefore avoid nonsteroidal anti-inflammatory drugs and diuretics, and they should maintain uniform salt intake. Administering lithium with meals may minimize gastrointestinal side effects.

Table 2 Usual Starting Doses, Dose Ranges, and Therapeutic Blood Level Targets for Traditional Mood-Stabilizing Agents

Drug (Trade Name)	*Typical Daily Starting Dose*	*Typical Daily Dose Range*	*Therapeutic Blood Level*
Lithium (Lithobid, Escalith)	300 mg t.i.d. or 450 mg b.i.d.	900–1,800 mg	0.8–1.2 mEq/L
Valproic acid or divalproex sodium (Depakene, Depakote, Depakote ER)	250 mg b.i.d.	750–2,000 mg	100–125 mg/L
Carbamazepine (Tegretol, Equetro)	100–200 mg b.i.d.	400–600 mg	6–10 µg/ml

Serum levels of lithium must be monitored. The goal for the treatment of acute mania is a lithium level of 1.0 to 1.2 mEq/L [*see Table 2*]. Serum levels should be measured 10 to 12 hours after the last oral dose, and levels should be drawn 4 to 5 days after the latest change in dosage. Lithium levels should be monitored periodically, and renal function indicators (i.e., blood urea nitrogen and serum creatinine) and TSH should be measured every 6 months. Common side effects include cognitive slowing, nausea, diarrhea, polyuria, polydipsia, weight gain, tremor, and a metallic taste in the mouth. Lithium is usually administered as lithium carbonate ($LiCO_3$) three times daily, but controlled-release forms may be administered twice daily.

Anticonvulsant Medications

Several anticonvulsant medications are also effective as mood stabilizers, including valproic acid–divalproex sodium and carbamazepine. These agents may act by decreasing neuronal membrane excitability, perhaps by effects on γ-aminobutyric acid (GABA) systems. Valproic acid–divalproex sodium appears to be more effective for manias that are dysphoric or mixed and for rapid cycling.[108] Gastrointestinal side effects may be minimized by administering the medication with food. Carbamazepine may be associated with a number of pharmacokinetic drug-drug interactions because of its induction of microsomal P-450 enzymes. Carbamazepine also induces its own metabolism, so-called autoinduction, resulting in a decrease in plasma levels several weeks to months after initiation of treatment. An extended-release form of carbamazepine is approved by the FDA for acute manic and mixed episodes.

Several newer anticonvulsant medications have been studied as potential treatments of bipolar disorders.[109-111] Lamotrigine has been shown to be effective in the acute treatment of bipolar depression and as maintenance therapy for bipolar disorder, but it does not appear to possess acute antimanic properties.[112] The olanzapine-fluoxetine combination capsule has FDA approval for bipolar depression. Early evidence suggested that topiramate may be an effective mood stabilizer,[113] but three randomized, controlled clinical trials in acute mania failed to demonstrate any efficacy. Whether or not these agents or tiagabine,[114] another anticonvulsant, may be helpful as adjunctive treatments in bipolar disorder remains unclear.

Atypical Antipsychotics

The atypical antipsychotic agents are effective for the treatment of acute mania. Olanzapine, risperidone, quetiapine, ziprasidone, and aripiprazole have FDA indications for bipolar mania. The risks of weight gain and metabolic disturbances are increasingly being recognized with some of these agents, and therefore, clinical monitoring is necessary.[115,116]

Additional Therapeutic Strategies

Other strategies are frequently used in addition to mood stabilizers, specific anticonvulsants, and atypical antipsychotics, including psychotherapy, mood charting, bipolar support groups, optimizing sleep patterns, eliminating the use of substances of abuse, and the use of other medications to augment response. During episodes of depression in patients with bipolar disorder, antidepressants must be used cautiously because of their potential for inducing hypomania or mania or accelerating cycling. Limited research suggests that adjunctive omega-3 fatty acid supplements may enhance outcomes in patients with bipolar disorder.[117]

Contemporary treatment of bipolar disorder focuses almost exclusively on psychopharmacologic treatment; most often, combination and even triple-drug therapy are essential. Nonetheless, psychotherapy is useful as adjunctive treatment; it can help relieve social and interpersonal dysfunction secondary to the disorder, improve medication compliance, teach stress management, and provide psychoeducation. Although psychotherapy alone is not recommended for the treatment of bipolar disorder, evidence suggests that cognitive therapy in conjunction with mood-stabilizing medications may be beneficial in preventing relapses, alleviating symptoms, and promoting social functioning.[118]

The treatment of depression associated with bipolar disorder is somewhat distinct from that of unipolar depression. For mild cases of bipolar depression, a mood stabilizer, such as lithium, may be sufficient.[119] When an antidepressant is required, adequate mood stabilization is recommended before the initiation of treatment with an antidepressant.[72] Alternatively, lamotrigine or the olanzapine-fluoxetine combination may be utilized.

Psychotic symptoms during a manic episode should be treated with antipsychotic agents. The atypical antipsychotic agents are first-line treatment; these agents are also effective for the treatment of acute mania itself, aside from their effects on psychotic symptoms. The short-term use of a benzodiazepine may also be required for anxiety or agitation.

COMPLICATIONS

Suicide rates are high in patients with bipolar disorder. For example, completed suicide occurs in approximately 10% to 15% of those with bipolar I disorder.[102] The illness also causes substantial psychosocial impairment, involving multiple aspects of the patient's life, including interference with interpersonal relationships and occupational and financial difficulties.

PROGNOSIS

In many patients with bipolar disorder, early episodes occur years apart. Episodes tend to occur more frequently with the passage of time, with associated worsening of functioning.[107] Although some persons may experience only a single manic episode, most patients with bipolar disorder experience recurrent episodes. Between episodes, patients may be symptom free, although many patients have some residual symptoms. Women with bipolar disorder tend to have more episodes of depression, whereas men tend to have more episodes of mania. Rapid cycling in bipolar disorder may develop over time and tends to be associated with poorer outcome.

Michael T. Compton, M.D., M.P.H., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Charles B. Nemeroff, M.D., Ph.D., holds stock in Corcept Therapeutics, Inc., and Neurocrine Biosciences, Inc., and has received grants for clinical research from, served as consultant for, or been a member of the speakers' bureaus of the following companies and organizations: Abbott Laboratories; Acadia Pharmaceuticals, Inc.; the American Foundation for Suicide Prevention; AstraZeneca Pharmaceuticals LP; Bristol-Myers Squibb Co.; Corcept Therapeutics, Inc.; Cypress Bioscience, Inc.; Cyberonics, Inc.; Eli Lilly and Company; Forest Pharmaceuticals, Inc.; GlaxoSmithKline; Janssen Pharmaceutica Products, LP; Merck & Co., Inc.; the National Alliance for Research on Schizophrenia and Depression; the National Institute of Mental Health; Neurocrine Biosciences, Inc.; Novartis Pharmaceuticals Corp.; Organon International, Inc.; Otsuka America, Ltd.; Pfizer, Inc.; Sanofi-Aventis; SCIREX; Somerset Pharmaceuticals, Inc.; Stanley Foundation/NAMI Research Institute; and Wyeth.

The following medications have not received FDA approval for the treatment of depression: isocarboxazid, moclobemide, selegiline, clomipramine, fluvoxa-

mine, reboxetine, gepirone, thyroid hormone, lithium, methylphenidate, dextroamphetamine, modafinil, buspirone, olanzapine, risperidone, St. John's wort, S-adenosyl-methionine, kava, B vitamins, and omega-3 fatty acids.

References

1. Regier DA, Narrow WE, Rae DS, et al: The de facto US mental and addictive disorders service system: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. Arch Gen Psychiatry 50:85, 1993

2. Kessler RC, McGonagle KA, Zhao S, et al: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 51:8, 1994

3. Sadek N, Bona J: Subsyndromal symptomatic depression: a new concept. Depress Anxiety 12:30, 2000

4. Judd LJ, Rapaport MH, Paulus MP, et al: Subsyndromal symptomatic depression: a new mood disorder? J Clin Psychiatry 55(suppl 4):18, 1994

5. Kessler RC, Walters EE: Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. Depress Anxiety 7:3, 1998

6. Birmaher B, Ryan ND, Williamson DE, et al: Childhood and adolescent depression: a review of the past 10 years: part I. J Am Acad Child Adolesc Psychiatry 35:1427, 1996

7. Lebowitz BD, Pearson JL, Schneider LS, et al: Diagnosis and treatment of depression in late life: consensus statement update. JAMA 278:1186, 1997

8. Kennedy GJ, Kelman HR, Thomas C: The emergence of depressive symptoms in late life: the importance of declining health and increasing disability. J Community Health 15:93, 1990

9. Wong ML, Licinio J: Research and treatment approaches to depression. Nature Rev Neurosci 2:343, 2001

10. Caspi A, Sugden K, Moffitt TE, et al: Influence of life stress on depression: moderation by a polymorphism in the *5-HTT* gene. Science 301:386, 2003

11. Nemeroff CB: The neurobiology of depression. Sci Am 278:42, 1998

12. Ressler KJ, Nemeroff CB: Role of serotonergic and noradrenergic systems in the pathophysiology of depression and anxiety disorders. Depress Anxiety 12(suppl 1):2, 2000

13. Nutt DJ: The neuropharmacology of serotonin and noradrenaline in depression. Int Clin Psychopharmacol 17(suppl 1):S1, 2002

14. Arborelius L, Owens MJ, Plotsky PM, et al: The role of corticotropin-releasing factor in depression and anxiety disorders. J Endocrinol 160:1, 1999

15. Holsboer F: The corticosteroid receptor hypothesis of depression. Neuropsychopharmacol 23:477, 2000

16. Nemeroff CB: The corticotropin-releasing factor (CRF) hypothesis of depression: new findings and new directions. Mol Psychiatry 1:336, 1996

17. Austin MC, Janosky JE, Murphy HA: Increased corticotropin-releasing hormone immunoreactivity in monoamine-containing pontine nuclei of depressed suicidal men. Mol Psychiatry 8:324, 2003

18. Mayberg H: Modulating limbic-cortical circuits in depression: targets of antidepressant treatments. Sem Clin Neuropsychiatry 7:255, 2002

19. Beyer JL, Krishnan KR: Volumetric brain imaging findings in mood disorders. Bipolar Disorders 4:89, 2002

20. Mayberg HS: Modulating dysfunctional limbic-cortical circuits in depression: towards development of brain-based algorithms for diagnosis and optimized treatment. Br Med Bull 65:193, 2003

21. Kupfer DJ: REM latency: a psychobiologic marker for primary depressive disease. Biol Psychiatry 11:159, 1976

22. Tsuang D, Coryell W: An 8-year follow-up of patients with DSM-III-R psychotic depression, schizoaffective disorder, and schizophrenia. Am J Psychiatry 150:1182, 1993

23. Faedda GL, Tondo L, Teicher MH, et al: Seasonal mood disorders: patterns of seasonal recurrence of mania and depression. Arch Gen Psychiatry 50:17, 1993

24. Terman M, Terman JS, Williams JBW: Seasonal affective disorder and its treatments. J Pract Psychiatry Behav Health 5:287, 1998

25. Gold JH: Premenstrual dysphoric disorder: an update. J Pract Psychiatry Behav Health 5:209, 1999

26. Ling FW: Recognizing and treating premenstrual dysphoric disorder in the obstetric, gynecologic, and primary care practices. J Clin Psychiatry 61(suppl 12):9, 2000

27. Shelton RC, Davidson J, Yonkers KA, et al: The undertreatment of dysthymia. J Clin Psychiatry 58:59, 1997

28. American Psychiatric Association: Practice guideline for the treatment of patients with major depressive disorder (revision). Am J Psychiatry 157(suppl):1, 2000

29. Shea SC: The chronological assessment of suicide events: a practical interviewing strategy for the elicitation of suicidal ideation. J Clin Psychiatry 59(suppl 20):58, 1998

30. Oquendo MA, Malone KM, Mann JJ: Suicide: risk factors and prevention in refractory major depression. Depress Anxiety 5:202, 1997

31. Kin NM, Nair NP, Amin M, et al: The dexamethasone suppression test and treatment outcome in elderly depressed patients participating in a placebo-controlled multicenter trial involving moclobemide and nortriptyline. Biol Psychiatry 42:925, 1997

32. Hull PR, D'Arcy C: Isotretinoin use and subsequent depression and suicide: presenting the evidence. Am J Clin Dermatol 4:493, 2003

33. Magin P, Pond D, Smith W: Isotretinoin, depression and suicide: a review of the evidence. Br J Gen Pract 55:134, 2005

34. Bhatia SC, Bhatia SK: Major depression: selecting safe and effective treatment. Am Fam Physician 55:1683, 1997

35. Altshuler LL, Post RM, Leverich GS, et al: Antidepressant-induced mania and cycle acceleration: a controversy revisited. Am J Psychiatry 152:1130, 1995

36. Wisner KL, Parry BL, Piontek KM: Postpartum depression: clinical practice. N Engl J Med 347:194, 2002

37. Schweitzer I, Tuckwell V, O'Brien J, et al: Is late onset depression a prodrome to dementia? Int J Geriatr Psychiatry 17:997, 2002

38. Hirschfeld RMA, Keller MB, Panico S, et al: The National Depressive and Manic Depressive Association consensus statement on the undertreatment of depression. JAMA 277:333, 1997

39. Wells KB, Sherbourne C, Schoenbaum M, et al: Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. JAMA 283:212, 2000

40. Screening for Depression: Recommendations and Rationale. May 2002. U.S. Preventive Services Task Force. Agency for Healthcare Research and Quality, Rockville, Maryland
http://www.ahrq.gov/clinic/3rduspstf/depression/depressrr.htm

41. Pignone MP, Gaynes BN, Rushton JL, et al: Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 136:765, 2002

42. Bodkin JA, Amsterdam JD: Transdermal selegiline in major depression: a double-blind, placebo-controlled, parallel-group study in outpatients. Am J Psychiatry 159:1869, 2002

43. Amsterdam JD: A double-blind, placebo-controlled trial of the safety and efficacy of selegiline transdermal system without dietary restrictions in patients with major depressive disorder. J Clin Psychiatry 64:208, 2003

44. Buckley NA, McManus PR: Fatal toxicity of serotonergic and other antidepressant drugs: analysis of United Kingdom mortality data. BMJ 325:1332, 2002

45. Rosen RC, Lane RM, Menza M: Effects of SSRIs on sexual function: a critical review. J Clin Psychopharmacol 19:67, 1999

46. Hirschfeld RM: Management of sexual side effects of antidepressant therapy. J Clin Psychiatry 60(suppl 14):27, 1999

47. Muller-Oerlinghausen B, Berghofer A: Antidepressants and suicidal risk. J Clin Psychiatry 60(suppl 2):94, 1999

48. Khan A, Khan S, Kolts R, et al: Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. Am J Psychiatry 160:790, 2003

49. Grunebaum MF, Ellis SP, Li S, et al: Antidepressants and suicide risk in the United States, 1985–1999. J Clin Psychiatry 65:1456, 2004

50. Sheehan DV: Venlafaxine extended release (XR) in the treatment of generalized anxiety disorder. J Clin Psychiatry 60(suppl 22):23, 1999

51. Harvey AT, Rudolph RL, Preskorn SH: Evidence of the dual mechanism of action of venlafaxine. Arch Gen Psychiatry 57:503, 2000

52. Whyte IM, Dawson AH, Buckley NA: Relative toxicity of venlafaxine and selective serotonin reuptake inhibitors in overdose compared to tricyclic antidepressants. QJM 96:369, 2003

53. Richelson E: Pharmacology of antidepressants. Mayo Clin Proc 76:511, 2001

54. Wellington K, Perry CM: Venlafaxine extended-release: a review of its use in the management of major depression. CNS Drugs 15:643, 2001

55. Schatzberg AF, Prather MR, Keller MB, et al: Clinical use of nefazodone in major depression: a 6-year perspective. J Clin Psychiatry 63:18, 2002

56. Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry 153:311, 1996

57. Coupland NJ, Bell CJ, Potokar JP: Serotonin reuptake inhibitor withdrawal. J Clin Psychopharmacol 16:356, 1996

58. Davidson JR, Connor KM: Bupropion sustained release: a therapeutic overview. J Clin Psychiatry 59(suppl 4):25, 1998

59. Karpa KD, Cavanaugh JE, Lakoski JM: Duloxetine pharmacology: profile of a dual monoamine modulator. CNS Drug Rev 8:361, 2002

60. Cowen PJ, Ogilvie AD, Gama J: Efficacy, safety and tolerability of duloxetine 60 mg once daily in major depression. Curr Med Res Opin 21:345, 2005

61. Rudolph RL, Feiger AD: A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. J Affect Disord 56:171, 1999

62. Benkert O, Muller M, Szegedi A: An overview of the clinical efficacy of mirtazapine. Human Psychopharmacol 17(suppl 1):S23, 2002

63. Jorenby D: Clinical efficacy of bupropion in the management of smoking cessation. Drugs 62(suppl 2):25, 2002

64. Silva de Lima M, Hotopf M: A comparison of active drugs for the treatment of dysthymia. Cochrane Database Syst Rev (3):CD004047, 2003

65. Yonkers KA, Halbreich U, Freeman E, et al: Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment: a randomized controlled trial. JAMA 278:983, 1997

66. Pearlstein TB, Stone AB, Lund SA, et al: Comparison of fluoxetine, bupropion, and placebo in the treatment of premenstrual dysphoric disorder. J Clin Psychopharmacol 17:261, 1997

67. Steiner M, Korzekwa M, Lamont J, et al: Intermittent fluoxetine dosing in the treatment of women with premenstrual dysphoria. Psychopharmacol Bull 33:771, 1997

68. Freeman EW, Rickels K, Arredondo F, et al: Full- or half-cycle treatment of severe premenstrual syndrome with a serotonergic antidepressant. J Clin Psychopharmacol 19:3, 1999

69. Yonkers KA, Brown C, Pearlstein TB, et al: Efficacy of a new low-dose oral contraceptive with drospirenone in premenstrual dysphoric disorder. Obstet Gynecol 106:492, 2005

70. Schatzberg AF: New approaches to managing psychotic depression. J Clin Psychiatry 64(suppl 1):19, 2003

71. Minter RE, Mandel MR: The treatment of psychotic major depressive disorder with drugs and electroconvulsive therapy. J Nerv Ment Dis 167:726, 1979

72. Compton MT, Nemeroff CB: The treatment of bipolar depression. J Clin Psychiatry 61(suppl 9):97, 2000

73. Bowden CL, Calabrese JR, Sachs G, et al: A placebo-controlled 18-month trial of lamotrigine and lithium maintenance treatment in recently manic or hypomanic patients with bipolar I disorder. Arch Gen Psychiatry 60:392, 2003

74. Tohen M, Vieta E, Calabrese J, et al: Efficacy of olanzapine and olanzapine-fluoxetine combination in the treatment of bipolar I depression. Arch Gen Psychiatry 60:1079, 2003

75. Bowden CL: Treatment options for bipolar depression. J Clin Psychiatry 66(suppl 1):3, 2005

76. Shelton RC, Keller MB, Gelenberg A, et al: Effectiveness of St. John's wort in major depression: a randomized controlled trial. JAMA 285:1978, 2001

77. Effect of *Hypericum perforatum* (St. John's wort) in major depressive disorder: a randomized controlled trial. Hypericum Depression Trial Study Group. JAMA 287:1807, 2002

78. Walsh BT, Seidman SN, Sysko R, et al: Placebo response in studies of major depression: variable, substantial, and growing. JAMA 287:1840, 2002

79. Brown RP, Gerbarg PL: Herbs and nutrients in the treatment of depression, anxiety, insomnia, migraine, and obesity. J Psychiatr Pract 7:75, 2001

80. Hirschfeld RMA, Montgomery SA, Aguglia E, et al: Partial response and nonresponse to antidepressant therapy: current approaches and treatment options. J Clin Psychiatry 63:826, 2002

81. Nemeroff CB: Augmentation strategies in patients with refractory depression. Depress Anxiety 4:169, 1996–97

82. Thase ME: New approaches to managing difficult-to-treat depressions. J Clin Psychiatry 64(suppl 1):3, 2003

83. Kupfer DJ, Frank E, Perel JM, et al: Five-year outcome for maintenance therapies in recurrent depression. Arch Gen Psychiatry 49:769, 1992

84. Terman M, Terman JS, Ross DC: A controlled trial of timed bright light and negative air ionization for treatment of winter depression. Arch Gen Psychiatry 55:875, 1998

85. Martin JL, Barbanoj MJ, Schlaepfer TE, et al: Transcranial magnetic stimulation for treating depression. Cochrane Database Syst Rev (2):CD003493, 2002

86. Nahas Z, Marangell LB, Husain MM, et al: Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. J Clin Psychiatry 66:1097, 2005

87. Thase ME, Greenhouse JB, Frank E, et al: Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. Arch Gen Psychiatry 54:1009, 1997

88. Reynolds CF, Frank E, Perel JM, et al: Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: a randomized controlled trial in patients older than 59 years. JAMA 281:39, 1999

89. Keller MB, McCullough JP, Klein DN, et al: A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. N Engl J Med 342:1462, 2000

90. Guze SB, Robins E: Suicide and primary affective disorders. Br J Psychiatry 117:437, 1970

91. Hirschfeld RMA, Montgomery SA, Keller MB, et al: Social functioning in depression: a review. J Clin Psychiatry 61:268, 2000

92. Musselman DL, Evans DL, Nemeroff CB: The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. Arch Gen Psychiatry 55:580, 1998

93. Hippisley-Cox J, Fielding K, Pringle M: Depression as a risk factor for ischaemic heart disease in men: population based case-control study. BMJ 316:1714, 1998

94. Ford DE, Mead LA, Chang PP, et al: Depression is a risk factor for coronary artery disease in men: the precursors study. Arch Intern Med 158:1422, 1998

95. Frasure-Smith N, Lesperance F, Talajic M: Depression following myocardial infarction: impact on 6-month survival. JAMA 270:1819, 1993

96. Giles DE, Jarrett RB, Biggs MM, et al: Clinical predictors of recurrence in depression. Am J Psychiatry 146:764, 1989

97. Regier DA, Boyd JH, Burke JD Jr, et al: One-month prevalence of mental disorders in the United States: based on five Epidemiologic Catchment Area sites. Arch Gen Psychiatry 45:977, 1988

98. Weissman MM, Bland RC, Canino GJ, et al: Cross-national epidemiology of major depression and bipolar disorder. JAMA 276:293, 1996

99. Berrettini WH: Molecular linkage studies of bipolar disorders. Bipolar Disorders 3:276, 2001

100. Blackwood DH, Visscher PM, Muir WJ: Genetic studies of bipolar affective disorder in large families. Br J Psychiatry 41(suppl):s134, 2001

101. Craddock N, Jones I: Genetics of bipolar disorder. J Med Genet 365:585, 1999

102. American Psychiatric Association: Practice guideline for the treatment of patients with bipolar disorder (revision). Am J Psychiatry 159(suppl):1, 2002

103. Bearden CE, Hoffman KM, Cannon TD: The neuropsychology and neuroanatomy of bipolar affective disorder: a critical review. Bipolar Disorders 3:106, 2001

104. Nemeroff CB, Kilts CD, Berns GS: Functional brain imaging: twenty-first century phrenology or psychobiological advance for the millennium? Am J Psychiatry 156:671, 1999

105. Kilzieh N, Akiskal HS: Rapid-cycling bipolar disorder: an overview of research and clinical experience. Psychiatr Clin North Am 22:585, 1999

106. McDonald WM: Epidemiology, etiology, and treatment of geriatric mania. J Clin Psychiatry 61(suppl 13):3, 2000

107. Goodwin FK, Jamison KR: Manic-Depressive Illness. Oxford University Press, New York, 1990

108. Bowden CL, Brugger AM, Swann AC, et al: Efficacy of divalproex vs lithium and placebo in the treatment of mania. JAMA 271:918, 1994

109. De Leon OA: Antiepileptic drugs for the acute and maintenance treatment of bipolar disorder. Harvard Rev Psychiatry 9:209, 2001

110. Yatham LN, Kusumakar V, Calabrese JR, et al: Third generation anticonvulsants in bipolar disorder: a review of efficacy and summary of clinical recommendations. J Clin Psychiatry 63:275, 2002

111. Macdonald KJ, Young LT: Newer antiepileptic drugs in bipolar disorder: rationale for use and role in therapy. CNS Drugs 16:549, 2002

112. Bowden CL, Karren NU: Lamotrigine in the treatment of bipolar disorder. Exp Opin Pharmacother 3:1513, 2002

113. Chengappa KN, Gershon S, Levine J: The evolving role of topiramate among other mood stabilizers in the management of bipolar disorder. Bipolar Disorders 3:215, 2001

114. Carta MG, Hardoy MC, Grunze H, et al: The use of tiagabine in affective disorders. Pharmacopsychiatry 35:33, 2002

115. Consensus development conference on antipsychotic drugs and obesity and diabetes. American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Diabetes Care 27:596, 2004

116. Casey DE, Haupt DW, Newcomer JW, et al: Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. J Clin Psychiatry 65(suppl 7):4, 2004

117. Stoll AL, Severus WE, Freeman MP, et al: Omega 3 fatty acids in bipolar disorder: a preliminary double-blind, placebo-controlled trial. Arch Gen Psychiatry 56:407, 1999

118. Lam DH, Watkins ER, Hayward P, et al: A randomized controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: outcome of the first year. Arch Gen Psychiatry 60:145, 2003

119. Sachs GS, Printz DJ, Kahn DA, et al: The expert consensus guideline series: medication treatment of bipolar disorder 2000. Postgrad Med Spec No:1, 2000

III ALCOHOL ABUSE AND DEPENDENCY

PATRICK G. O'CONNOR, M.D., M.P.H.

Alcohol use is associated with a wide range of behavioral and medical disorders.[1] An estimated 4% to 40% of medical and surgical patients experience problems related to alcohol.[2] In the United States, more than 85,000 deaths a year are directly attributed to alcohol use.[3] The annual economic cost of alcohol use in the United States is estimated to be over $185 billion.[3] Thus, alcohol and its associated health problems have a major impact on the practices of generalists and other physicians.[4]

Patients with alcohol-related problems present several unique challenges for physicians.[5] As with other chronic diseases, alcohol-use disorders range from relatively asymptomatic to severe. To manage patients effectively, physicians should recognize all forms of alcohol-related problems, from the earliest to the most advanced stages. They should be able to manage patients along the entire spectrum of the disease process.[6]

Available from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) are guidelines for physicians on the management of patients with alcohol-related problems; these guidelines are available on the Web, at http://www.niaaa.nih.gov.[7,8] The NIAAA guidelines address behavioral and social problems, as well as the medical manifestations of alcohol use. This chapter provides an overview of the major clinical features and recent developments in the identification and management of alcohol-related problems in clinical practice.

Definitions and Classification

When considering patients who may have alcohol-related problems, it is helpful to distinguish between three types of alcohol use: nonproblematic, moderate drinking; hazardous, or at-risk drinking, which places patients at risk for alcohol-related problems; and harmful drinking, which directly causes specific alcohol-related problems. Patients with more advanced alcohol-related problems may meet the criteria for alcohol dependence or abuse.[9] One study suggests that in patients seen in primary care settings, the prevalence of hazardous drinking ranges from 4% to 29% and that of harmful drinking, from 1% to 10%.[10] These definitions are based on epidemiologic studies that demonstrate an association between specific levels of alcohol consumption and increasing health problems and mortality.[11,12]

MODERATE DRINKING

The NIAAA has defined moderate drinking in terms of the average number of drinks consumed a day that places an adult at relatively low risk for alcohol-related health problems.[7] For men younger than 65 years, moderate drinking is defined as drinking an average of, at most, two drinks a day. For men older than 65 years and for all women, moderate drinking is defined as drinking less than two drinks a day. Although some evidence suggests that low levels of alcohol intake may be beneficial,[13,14] the degree to which moderate drinking may confer health benefits and may be associated with reduced mortality—specifically, reduction of cardiovascular mortality—is controversial. When it exists, this benefit likely occurs at low levels of alcohol consumption (i.e., less than one drink a day).[15]

AT-RISK DRINKING

At-risk drinking (which corresponds to the category of hazardous use in the schema developed by the World Health Organization [WHO][16]) occurs when those moderate drinking levels are exceeded or when the number of drinks consumed during a single occasion exceeds a specified amount (four drinks per occasion for men and three drinks per occasion for women).

HARMFUL DRINKING

Harmful drinking is defined as alcohol consumption that results in physical or psychological harm. This disorder is also recognized by the WHO and is defined by criteria of the International Classification of Diseases, 10th Revision (ICD-10), which include (1) clear evidence that alcohol is responsible for physical or psychological harm, (2) the nature of the harm is identifiable, (3) alcohol consumption has persisted for at least 1 month or has occurred repeatedly over the previous 12-month period, and (4) the individual does not meet the criteria for alcohol dependence. The ICD-10 criteria are found on the Web, at http://www.who.int/classifications/icd/en.

ALCOHOL ABUSE

Specific criteria for diagnosing alcohol abuse have been developed by the American Psychiatric Association (APA).[9] In the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV), alcohol abuse is defined as a maladaptive pattern of alcohol use leading to clinically significant impairment or distress, manifested in a 12-month period by one or more of the following problems[9]: (1) failure to fulfill role obligations at work, school, or home; (2) recurrent use of alcohol in hazardous situations; (3) legal problems related to alcohol; and (4) continued use despite alcohol-related social problems.

ALCOHOL DEPENDENCE

By APA criteria, alcohol dependence is manifested by a maladaptive pattern of use over a 12-month period that includes three or more of the following problems[9]: (1) physiologic tolerance, characterized either by an increase in the amount of alcohol consumed or by a decrease in the effects of the amount of alcohol customarily consumed; (2) symptoms of withdrawal; (3) use of greater amounts of alcohol over a longer period than intended; (4) a persistent desire or unsuccessful attempts to control use; (5) a great deal of time spent obtaining alcohol, using alcohol, or recovering from use; (6) a reduction in the level of important social, occupational, and recreational activities; and (7) continued use despite knowledge of physical or psychological problems.

ALCOHOLISM

The term alcoholism, which is perhaps the most widely used term to describe patients with alcohol problems, has lost much of its usefulness because of the imprecision in its definition and the stigma associated with the term. A panel of 23 experts convened by the National Council on Alcoholism and Drug Dependence and the American Society of Addiction Medicine defined alcoholism as a primary, chronic disease with genetic, psychosocial, and environmental factors that is often progressive and fatal and is characterized by impaired control over drinking, pre-

occupation with the drug alcohol, use of alcohol despite adverse consequences, and distortions in thinking, most notably denial.[17] However, precise terminology for specific alcohol problems is more clinically useful.

Epidemiology

In a nationwide survey, 62% of adults in the United States reported that they drink alcohol.[18] Alcohol-related problems are common in the general population; however, prevalence of these problems has varied with different studies. The Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System reported that in 2003, 16.4% of adults were at risk for binge drinking (having five or more drinks on one occasion), and 5.7% were at risk for heavier drinking (having two or more drinks a day).[19]A nationwide survey based on face-to-face interviews reported that the 12-month prevalence rates of alcohol abuse and dependence, as defined by DSM-IV, in 2001 through 2002 were 4.65% and 3.81%, respectively.[20] Abuse and dependence were more common in males and in younger persons. Studies have demonstrated that alcohol-related problems are highly prevalent in primary care patients. For example, in two intervention studies, at-risk or problem drinking was identified in 41% of men and 28% of women from 47 general practices in the United Kingdom[21] and in one in six patients in 17 general practices in Wisconsin.[22] With regard to more severe alcohol-related problems, one study found that 20% of inpatients screened positive for alcoholism.[23] In various studies of outpatients, the lifetime prevalence of alcohol abuse and dependence ranged from 13% to 22%.[24-26]

Genetics

Evidence from family, twin, and adoption studies supports a strong genetic component for the risk of alcohol dependence.[27] In general, most studies of twins have demonstrated a higher concordance of alcohol dependence in monozygotic twins than in dizygotic twins. Adoption studies generally document that adopted children with a biologic parent who is alcoholic have a twofold to threefold greater risk of alcoholism than adopted children whose biologic parents are not alcoholic. Studies have shown that there are a variety of electroencephalographic differences and differences in behavioral and physiologic responses to test doses of alcohol between nonalcoholic adolescents and young adults with a family history of alcoholism and those without such a history. Overall, individuals with a family history of alcohol dependence have a twofold greater risk of this condition than those without such a family history.[28]

Several genetic factors may influence the risk of alcohol dependence. A lower inherent level of response to alcohol is genetically influenced and may predict the development of alcohol dependence.[27] The genes responsible for determining the level of response to alcohol have not been identified, although the γ-aminobutyric acid (GABA) receptor and serotonin transporter gene may be involved.[29] Other putative alcoholism-vulnerability genes include alcohol dehydrogenase–2 (*ADH2*), aldehyde dehydrogenase–2 (*ALDH2*), and D_2 dopamine receptor (*DRD2*).[27] Although genetic predisposition is clearly a major risk factor for alcohol-related problems, environmental influences may also impart risk. Such environmental influences include negative life events, occupational stress, expectancies about alcohol, personality factors (e.g., problem-prone behavior during adolescence), and interpersonal influences (e.g., the behaviors of family members or peers).[1,27,30]

Common Problems Associated with Alcohol Abuse

Familiarity with the wide variety of medical, behavioral, and psychiatric complications of heavy drinking or alcohol dependence facilitates detection and management of alcohol-related problems in patients.

MEDICAL PROBLEMS

The numerous medical complications associated with excessive alcohol consumption have been well documented.[1,2] Patients may present to physicians with acute and chronic clinical signs and symptoms that are the direct or indirect results of alcohol use. The most common medical effects of alcohol are seen in the central and peripheral nervous systems[31]; these effects are intoxication, withdrawal, seizures, delirium and dementia, stroke, and peripheral neuropathy. Similarly, alcohol has a wide variety of effects on the gastrointestinal system[32]; these effects include esophageal diseases (e.g., Mallory-Weiss tears and carcinoma), gastritis, and peptic ulcer disease.

Liver diseases, including acute alcoholic hepatitis and cirrhosis, are also highly prevalent.[33] Along with having the typical manifestations of liver disease, patients with alcoholic liver disease are at increased risk for toxicity from acetaminophen, even when used in therapeutic doses.[34] In addition, continued alcohol use may worsen the clinical outcomes in patients with hepatitis C.[35] Acute and chronic pancreatic disease is also a common manifestation of alcohol use.[36]

Finally, patients with alcohol-related problems may suffer a wide range of cardiovascular effects, such as hypertension, left ventricular hypertrophy and cardiomyopathy, arrhythmias, and sudden death.[37] Increased levels of alcohol consumption are associated with increased risks of both heart failure[38] and atrial fibrillation.[39] Cessation of alcohol use may result in improvement of some alcohol-related conditions. For example, alcohol users with dilated cardiomyopathy who stop using alcohol appear to have improved cardiac function over time.[40]

Along with causing major neurologic, digestive system, and cardiovascular effects, alcohol use is associated with a variety of other organ-based effects. Heavy drinking is associated with cancer of the upper digestive and respiratory tracts, the liver, and, in at least one study, the prostate, pleura, and cervix.[41] Data linking alcohol to cancers of other organs (e.g., pancreas, colon, and breast) have been less convincing.[42]

Alcohol use may present as bleeding from various sites because of dysfunction of hepatic synthesis or thrombocytopenia. Alcohol use has also been associated with metabolic and endocrine abnormalities, such as osteoporosis, menstrual dysfunction, male hypogonadism, and thyroid and adrenal dysfunction.[43,44] Gout is associated with alcoholism and may occur at serum urate levels lower than those seen in nonalcoholic patients.[45] Toxic effects of alcohol on the kidney are generally subclinical or secondary to other alcohol-related effects. Alcohol has been related to important dermatologic problems, such as psoriasis and the dermatologic findings associated with chronic liver diseases.[46] Finally, high levels of dental and periodontal disease have been documented in populations of alcoholic patients.[47]

PSYCHIATRIC PROBLEMS

Epidemiologic surveys have demonstrated high rates of psy-

chiatric illness in persons diagnosed with alcohol abuse or dependence, and comorbid psychiatric disease is associated with increased health care utilization in patients with alcohol-related problems.[48] Data from the Epidemiologic Catchment Area Study demonstrated that 45% of such persons had a lifetime psychiatric diagnosis.[49] The most common disorders were anxiety and affective and antisocial personality disorders.[49] The National Comorbidity Study found a high 12-month prevalence of anxiety disorders (36.9%) and affective disorders (29.2%) in alcohol-dependent patients.[50]

These data indicate that patients with alcohol problems require careful evaluation for comorbid psychiatric symptoms and problems. Treatment of comorbid psychiatric conditions such as depression is effective if the underlying alcohol-use disorder is treated concurrently.[51]

Similarly, patients with psychiatric disorders are at high risk for comorbid substance-use disorders.[52] A review of the medical literature revealed that the prevalence of current alcohol problems in patients with major depression is 16%, whereas that of the general population is 7%.[53] Studies have shown that the presence of an alcohol-use disorder in adolescence is predictive of both continued alcohol problems and psychiatric disease in adulthood.[54,55]

OTHER BEHAVIOR-RELATED PROBLEMS

Alcohol misuse can harm persons other than the user; it has been linked to increased rates of accidents and injuries, violence against others, and spousal abuse. Alcohol misuse is also linked to high-risk behaviors (e.g., drug and tobacco use and high-risk sexual activity) that carry increased risk of disease.

Accidents and Injuries

Alcohol use is the leading cause of accidents (most notably, automobile accidents), injuries, and trauma (e.g., drownings, head injuries, burns, and spinal cord injuries).[3,56] Alcohol use increases the risk of injury and is associated with more severe injury in trauma patients[57,58] Emergency departments are important intervention sites for patients with alcohol-related problems.[59] Intervention with these patients has been shown to reduce both alcohol consumption and subsequent use of emergency department services.[60] Despite these data, one study showed that emergency physicians did not refer any patients with an elevated blood alcohol level to alcohol treatment.[61] Another study, however, indicated that screening and brief interventions are becoming more routine in emergency departments. More widespread use of brief-intervention therapy would likely reduce recurrent alcohol-related injury.[62]

Violence and Abuse

Alcohol use is associated with injuries and trauma related to acts of violence; such acts include assault and homicide, as well as domestic violence toward children and spouses.[63,64] In one study of patients in urban trauma centers, victims of violence were more likely to use alcohol, to be male, and to possess a knife or a gun.[59] A history of childhood victimization has been identified as a predictor of the development of alcohol-related problems in women.[63] Thus, a patient's history as a victim of violence or psychological abuse should prompt a careful assessment of the patient's drinking behavior. The drinking behavior of the victim's partner should be assessed, as well. In one large study examining risk factors for injury in women related to domestic violence, alcohol abuse and drug abuse by the victims' partners were identified as the most important risk factors.[64]

Risk of HIV Infection

HIV seroprevalence may be higher in patients with more severe impairment from alcohol, and women may be at especially increased risk.[65] In two cross-sectional surveys, HIV seroprevalence in heterosexual, non–drug-injecting patients was 5%.[66] Heavy drinkers are more likely to engage in high-risk sexual behaviors, such as having many sexual partners and engaging in sex without condoms.[56,67] Counseling patients about the association between alcohol use and high-risk sexual behaviors may result in safer sexual practices. The use of alcohol by patients who have HIV infection may increase their risk of HIV-related problems, such as cognitive dysfunction.[68]

Tobacco and Other Drug Abuse

Alcohol-dependent persons are more likely to smoke tobacco than the general population. Alcohol acts as a cocarcinogen with tobacco, substantially increasing the risk of head, neck, and lung cancers.[42] Alcohol-dependent patients who stop drinking may be up to 60% more likely to quit smoking than those who continue to drink.[69] Treatment of tobacco abuse in patients with alcohol-related problems is critical to the improvement of long-term health outcomes in these patients.[70] Alcohol abuse and dependence are commonly seen in association with abuse of and dependence on prescription drugs[71] and illicit substances.[72] Careful screening is often required to identify a patient with coexisting alcohol and drug use, and treatment of such a patient must address both alcohol- and drug-related problems.

Screening and Diagnosis of Alcohol-Related Problems

Despite the prevalence of alcohol-related problems and their impact on health, most studies demonstrate that alcohol-related problems are not routinely detected in primary care settings.[73] Rates of failure to detect alcohol-related problems vary by setting and level of practitioner training. Alcohol abuse or dependence is most likely to be identified in patients experiencing severe medical complications, such as alcoholic hepatitis or cirrhosis,[23] and is less likely to be detected in women.[74] Earlier detection of medically hazardous levels of drinking before the onset of organ damage or of alcohol abuse or dependence may be critical in preventing late sequelae. Simple, brief interventions are effective in reducing heavy drinking in high-risk drinkers.[1,7,22] The most recent U.S. Preventive Health Task Force (USPHTF) recommendations concerning alcohol screening strongly endorse routine screening in primary care[75]; however, screening techniques are used inconsistently by primary care physicians.[73,76]

Easy-to-use techniques for screening patients for alcohol-use disorders are currently available. Although these techniques are used inconsistently by primary care physicians, screening for alcohol-related problems should be incorporated into the routine care of all patients[76] One such screening technique involves a four-step process for identifying and diagnosing alcohol-related problems [*see Table 1*].[3,5,7] Step 1 is to inquire about current and past alcohol use in all patients. Patients with alcohol-related problems may be somewhat sensitive about these and other questions about their drinking behavior. It is therefore critical that the patient be approached in a nonjudgmental manner. In addition, these questions may need to be asked on multiple occasions to obtain an accurate history of the patient's alcohol use. The questioner needs to be specific about current and past alco-

Table 1 Screening and Diagnosis of Alcohol Problems

General Questions to Be Asked of All Patients
Do you drink alcohol (ever or currently)?
Do you have a family history of alcoholism?

Questions Concerning Quantity and Frequency of Alcohol Use
What type(s) of alcohol (beer, wine, spirits) do you use?
How often do you drink?
How much do you usually drink on a typical drinking day?
Do you ever drink more (if so, how much) than your usual amount?

Screening Questionnaires
CAGE
- Have you ever felt you should ***c***ut down on your drinking?
- Have people ***a***nnoyed you by criticizing your drinking?
- Have you ever felt bad or ***g***uilty about your drinking?
- Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (***e***ye opener)?

Other Questionnaires
Alcohol Use Disorder Identification Test (AUDIT)
TWEAK (*T*olerance, *W*orry, *E*ye opener, *A*mnesia, *K*ut down)
Short Michigan Alcoholism Screening Test (S-MAST)
Rapid Alcohol Problems Screen 4 (RAPS4)

Specific Areas to Assess in Patients with Potential Alcohol Problems
Criteria for alcohol abuse and dependence [*see text*]
Evidence of medical and psychiatric problems
Evidence of behavioral and social problems
History of use of other substances
- Tobacco
- Mood-altering prescription drugs
- Illicit drugs (e.g., heroin, cocaine)

Previous treatment for alcohol or substance use

hol use. Because genetic predisposition and family environment are important risk factors for alcohol-related problems,[2] inquiries into family history should routinely include questions about relatives with alcohol-related problems.

Steps 2 through 4 apply to all patients who report a history of alcohol use. In step 2, a more detailed history regarding quantity and frequency of alcohol use is obtained. The type of alcohol consumed is critical. Some patients may consider beer a safer form of alcohol and may not report its use unless specifically asked. Questions can help establish the frequency of alcohol use (e.g., daily as opposed to less frequent use) and a baseline of the amount of alcohol usually consumed. It is also important to establish other patterns of alcohol consumption; for example, a patient may have only two drinks a day from Monday through Thursday but may consume 10 to 15 drinks on Friday and Saturday nights. Questions about the quantity and frequency of alcohol use may help distinguish moderate drinking from at-risk and problem drinking [*see Table 1*] and may help identify specific patterns of use, such as binge drinking.

In step 3, a standardized questionnaire is used to detect possible alcohol-related problems.[77] The most commonly studied screening instrument in primary care settings is the CAGE questionnaire [*see Table 1*]. This questionnaire screens for lifetime alcohol-related problems. It consists of four questions designed to identify ongoing alcohol use and withdrawal phenomena that demonstrate alcohol dependence. In scoring the CAGE questionnaire, each "yes" response counts as 1 point. A score of 2 or more is generally considered to be a positive result. The CAGE questionnaire is useful in identifying patients with alcohol-related problems who are seen in primary care settings; its sensitivity ranges from 43% to 95%, and its specificity ranges from 70% to 95%.[24,76,78] Supplementing the four questions of the CAGE questionnaire with additional questions about tolerance (e.g., "How many drinks can you hold?") and substituting a question about whether any of the patient's friends or family members have expressed worry or concern about the patient's alcohol use may improve the usefulness of the CAGE questionnaire.[79]

The CAGE questionnaire is a reliable screening instrument in a variety of health care settings; however, it may be less useful in specific patient groups. In one study of patients older than 60 years, the CAGE questionnaire performed poorly in identifying heavy drinkers or binge drinkers, among whom fewer than 50% had a positive CAGE score.[80] The CAGE questionnaire may also perform less well in women and persons of specific ethnic groups.[81] The sensitivity of the CAGE questionnaire and other screening instruments may vary by gender or ethnicity; thus, it may be necessary to use a combination of screening techniques to identify alcohol-related problems in some patients.

The Alcohol Use Disorder Identification Test (AUDIT) may provide additional useful information when patients are being screened for alcohol-related problems in primary health care settings.[10,77] AUDIT was developed by the WHO and was designed to identify hazardous and harmful drinking, as well as alcohol dependence. Among the 10 AUDIT questions are three quantity and frequency questions regarding current drinking behavior and seven questions regarding past drinking behavior. Each question is scored on a scale of 0 to 4; a total score of 8 or greater is considered to be a positive result. AUDIT has demonstrated a sensitivity of 92% to 96% and a specificity of 94% to 96%.[82] A three-question version of the AUDIT (AUDIT-C) has been demonstrated to be effective in men and women.[83] Other screening instruments (e.g., TWEAK [*t*olerance, *w*orried, *e*ye-openers, *a*mnesia, *k*/cut down], S-MAST [Short Michigan Alcoholism Screening Test], and RAPS4 [Rapid Alcohol Problems Screen]) are available for use in clinical practice [*see Table 1*]; however, evaluation of the effectiveness of each test in specific patient groups is ongoing.[84,85]

Step 4 involves asking further questions with regard to potential alcohol-related problems. Step 4 is applied to those patients who were identified in steps 1 through 3 as having potential alcohol-related problems. A diagnosis of alcohol abuse or dependence is made on the basis of the criteria described in DSM-IV.[9] The examination for evidence of specific alcohol-related medical and psychiatric problems should include a thorough history and physical examination. Although not useful as screening tests for alcohol-use disorders,[78] laboratory tests, such as liver enzyme assay, may be useful in identifying undiagnosed alcohol-related medical problems. Questioning patients about alcohol-related behavioral and social problems is also essential when evaluating patients suspected of having alcohol-related problems. In addition, knowledge of previous treatment of alcohol-related problems is essential when referring patients for treatment.

Treatment Approaches for Patients with Alcohol-Related Problems

Screening and diagnosis represents a basic initial step in the management of patients with alcohol-related problems. Factors such as patient denial may make the initiation of a treatment

Table 2 Management of Patients with Alcohol Problems

For all patients
- Screen for alcohol problems
- Educate about risky drinking and its consequences
- Perform a detailed alcohol history for all patients who may have problem drinking [*see Table 1*]

For at-risk drinkers
- Assess patients for alcohol-related medical, psychiatric, and behavioral problems
- Advise nondependent at-risk problem drinkers to decrease their alcohol consumption to an amount below at-risk levels (e.g., brief intervention therapy) [*see Table 4*]
- Advise those at-risk drinkers who cannot follow suggestions to decrease their alcohol use to recommended levels to abstain from alcohol
- Monitor and assess drinking behavior over time

For alcohol-dependent patients
- Advise alcohol-dependent drinkers to abstain from alcohol, and refer them to appropriate alcohol treatment services for detoxification (if needed) and prevention of relapse
- Identify and manage alcohol-related medical, psychiatric, and behavioral problems
- Monitor patients in recovery to promote abstinence and assess for relapse

plan very challenging. In addition, as with all chronic health problems and diseases, treatment is no guarantee of cure. Patients with more severe alcohol-related problems, such as alcohol dependence, may require treatment from alcohol treatment specialists, whereas those in the at-risk and problem-drinking categories may be managed in primary care settings. The primary care physician can play a critical role in the management of all patients with alcohol-related problems, regardless of the severity of those problems. For patients with more severe alcohol dependence, the primary care physician's role includes identifying patients with severe dependence, referring such patients to treatment specialists, evaluating those patients for medical complications, providing support for ongoing treatment, and monitoring for signs of relapse [*see Table 2*].

A six-stage model has been proposed for assessing patients' readiness for behavior change and treatment [*see Table 3*].[83] These stages represent a continuum, and patients may move from one stage to another. Physician assessment of readiness for behavior change may be helpful in tailoring the advice given to patients with alcohol-related problems.[82] Being in the precontemplation stage, in which a patient is unaware of the alcohol problem, is associated with poorer treatment outcomes.[86] The overall goal for physician counseling is to move patients from the precontemplation stage to the action stage so that maintenance can be achieved.[87]

MANAGEMENT OF AT-RISK AND PROBLEM DRINKERS WITH BRIEF INTERVENTION THERAPY

Brief intervention therapy typically involves providing patients with feedback about the problems associated with their drinking habits and advising them to reduce their alcohol consumption to levels considered medically safe (i.e., below the level of at-risk drinking). Brief interventions typically include counseling sessions lasting 5 to 20 minutes that are conducted on one or more occasions. The FRAMES acronym (*f*eedback, *r*esponsibility, *a*dvice, *m*enu, *e*mpathy, *s*elf-efficacy) summarizes a counseling strategy commonly used in brief interventions [*see Table 4*].[88] This counseling strategy, as well as others commonly employed in brief intervention therapy, are useful for primary care physicians, and their use in primary care settings has been advocated by government agencies and experts in the field.[7]

Three meta-analyses have examined the efficacy of brief (under 1 hour) intervention therapy.[89-91] One such study concerned over 40 controlled trials that enrolled over 6,000 patients; it concluded that brief intervention therapy is more effective than no counseling at all and is often as effective as more extensive treatment.[1,89] Another meta-analysis, of eight studies, demonstrated that patients who received brief intervention therapy were almost twice as likely to decrease their drinking as those who did not receive such therapy.[90] Finally, a third meta-analysis, which examined data from 12 controlled trials of adult primary care patients, indicated that patients who received brief intervention therapy experienced a decrease of 13% to 34% in drinks consumed per week.[91]

One study illustrates the process of brief intervention therapy.[22] In this study, a sample of 723 persons were recruited from 17 primary care practices in Wisconsin. The group receiving brief intervention therapy was given a health booklet on general health issues and also received structured counseling about their drinking behaviors during two visits with a physician [*see Table 4*]. After each visit to the physician, the patients received a telephone call from a nurse who reinforced the advice given by the physician. The control group received only the health booklet. Patients were followed for 1 year; the main outcomes used to evaluate the brief intervention therapy were measures of alcohol use (i.e., drinks per week and episodes of binge drinking), the

Table 3 Stages of Readiness to Change Addictive Behaviors and Stage-Specific Approaches

Stage	*Patient Features*	*Possible Approaches*
Precontemplation	Unaware of alcohol problems	Express concern about health problems and their link to alcohol use
Contemplation	Recognizes alcohol problems	Reinforce links between alcohol use and problems; promote behavior change
Determination	Decides to change behavior	Support decision; provide advice about short-term and long-term actions to change behavior
Action	Changes behavior	Monitor compliance with advice and outcomes
Maintenance	Continues new behavior	Continue follow-up of alcohol problems; support behavior change and treatment efforts; monitor for relapse
Relapse	Recurrence of alcohol problems	Support return to behavior change or reentry into or continuation of treatment; monitor for alcohol problems

Table 4 Components of a Brief Intervention

FRAMES Counseling Strategy for Brief Interventions

Feedback: review the problems experienced by the patient because of alcohol use

Responsibility: emphasize that changing patterns of alcohol use is the patient's choice and responsibility

Advice: advise the patient to cut down or abstain from alcohol

Menu: provide menu of options and strategies for changing behavior

Empathy: use a warm, empathic, understanding approach with the patient

Self-efficacy: encourage optimism about likelihood and benefits of changing behavior

Specific Brief Intervention Therapeutic Techniques and Visit Characteristics[21]

Interventions

- Workbook with feedback on health behavior
- Review of the prevalence of problem drinking and adverse effects of alcohol
- Worksheet on drinking cues
- Drinking agreement "prescription"
- Drinking diary

Visit characteristics

- Two 15-min visits with physician
- Phone call from nurse 2 wk after visit with physician

number of alcohol-related emergency department visits, and the number of days spent in the hospital. A 12-month follow-up evaluation showed that the mean number of drinks consumed per week decreased by more than 7.5 drinks in the group receiving brief intervention therapy, compared with a decrease of just over three drinks a week in the control group. The treatment group also had significantly fewer episodes of binge drinking over a 30-day period and experienced fewer episodes of excessive drinking. In addition, men in the treatment group had a significantly shorter length of hospitalization than men in the control group.[22] This study, along with many of the others reviewed in the three meta-analyses, demonstrates that relatively simple clinical interventions that take a minimum amount of time and are well within the skill level of most primary care physicians can effectively improve patients' drinking behaviors.

Important questions remain to be studied. The effectiveness of brief intervention therapy over time is unknown. In addition, the optimal number of brief interventions, the optimal length of time over which brief interventions should be made, and the optimal frequency of repeated brief interventions are uncertain. Other aspects of brief intervention therapy that have not yet been determined are the long-term benefits of such therapy and the effect that brief interventions have on progression to more severe alcohol-related problems.

SELF-HELP GROUPS

Alcoholics Anonymous (AA) and similar groups, such as Rational Recovery and Narcotics Anonymous, are based on a 12-step recovery model. In this model, 12 steps describe specific attitudes, beliefs, and actions that are regarded as critical to the recovery process. Meetings are held 7 days a week in multiple locations throughout the country and include both open meetings, which are open to everyone, and closed meetings, which are restricted to group members.

Physicians who are interested in encouraging their patients to attend AA meetings should obtain an AA meeting schedule from their local AA organization. In addition, it is often recommended that physicians attend at least one open meeting so that they can better counsel their patients on how AA works. AA chapters and other 12-step groups are widely available and are free of charge.

Attention has been paid to the question of the effectiveness of 12-step approaches. Studies have demonstrated that attendance at AA meetings correlates with positive drinking outcomes.[92,93] In addition to attendance, greater involvement in AA (e.g., whether the patient has a program sponsor and the degree to which the patient participates in meetings) appears to correlate with decreased alcohol consumption.[94] In one study, participants in an employee-assistance program were given compulsory treatment on an inpatient basis, were required to attend AA meetings, or could choose just one of these two options. Patients in all three groups showed improvement in alcohol consumption, and there were no significant differences in job-related outcome between the three groups.[95] A significant number of patients in AA were subsequently referred for inpatient treatment; those who received inpatient treatment had the best outcomes. Future research on AA that is methodologically rigorous and attends to important clinically oriented outcomes is needed.[92]

MANAGEMENT OF ALCOHOL ABUSE AND DEPENDENCE

In addition to providing office-based evaluation, management, and referral to self-help groups, primary care physicians will often need to provide more intensive services to patients who meet criteria for alcohol abuse and dependence. These services can include management of the alcohol withdrawal syndrome and referral to an alcohol treatment program.

Alcohol Withdrawal Syndrome

Signs and symptoms of alcohol withdrawal, which can occur in alcohol-dependent persons who stop drinking alcohol or who reduce their alcohol intake, include abnormalities in vital signs (e.g., tachycardia, hypertension, and fever), other symptoms of autonomic hyperactivity (e.g., tremor, diaphoresis, and insomnia), GI symptoms (e.g., nausea, vomiting, and diarrhea), and central nervous system effects (e.g., anxiety, agitation, hallucinations, seizures, and delirium). Withdrawal severity and response to treatment may be assessed using the revised Clinical Institute Withdrawal Assessment for Alcohol, revised (CIWA-Ar).[96] This instrument describes 10 clinical features of withdrawal that are rated by the clinician on the basis of observation of the patient. Although many patients presenting with mild withdrawal can be managed as outpatients, those with more severe withdrawal or with significant comorbid medical or psychiatric problems may require inpatient care.

When managing alcohol withdrawal, physicians should keep the following specific goals in mind.[97] Patients' symptoms should be monitored closely and treated in a manner that minimizes discomfort. Inadequate treatment of withdrawal symptoms may be a major stimulus to return to drinking. Patients should be monitored for major withdrawal complications such as seizures or delirium and treated accordingly. Clinical features such as elevated baseline blood pressure, medical comorbidities, and prior complicated withdrawal may indicate an increased risk of a severe withdrawal episode (e.g., delirium tremens).[98] Withdrawal treatments are designed to decrease the occurrence of these complications. Patients should be fully evaluated for other alcohol-related problems, and their general health status should be assessed. It is critical to immediately plan for post-

withdrawal treatment of alcohol-related problems to help patients maintain abstinence. All patients who are treated for alcohol withdrawal should be referred for ongoing treatment.

Pharmacologic therapies for alcohol withdrawal syndrome have been the focus of much research over the past 40 years. A variety of drugs, including the barbiturates, phenothiazines, carbamazepine, and alcohol itself, have been used. The benzodiazepines (e.g., chlordiazepoxide, diazepam, lorazepam, and oxazepam) are the safest and most effective medications for this purpose.[99] In addition to preventing or alleviating withdrawal symptoms, benzodiazepines may also decrease the incidence of seizures and possibly delirium tremens.[99,100] Control of agitation can be achieved using a parenteral, rapid-acting benzodiazepine that is cross-tolerant with alcohol.[100] Adequate doses should be used to maintain light somnolence for the duration of the delirium. A working group assembled by the American Society of Addiction Medicine reviewed the world literature on pharmacologic therapy for alcohol withdrawal syndrome to provide an evidence-based guideline for clinicians.[99] Their review of over 130 papers yielded 65 prospective, controlled trials, which examined over 40 drugs. This review provided strong evidence in favor of the benzodiazepines over placebo and all other drugs and suggested that the longer-acting benzodiazepines provide a smoother withdrawal and may be more effective in preventing seizures.[99] Generally, however, shorter-acting benzodiazepines are considered to be safer in the elderly and in patients with severe liver disease. Older benzodiazepines, such as chlordiazepoxide and diazepam, are the best studied and are also the least expensive.[99] Research has suggested that clonidine, beta blockers, and carbamazepine are effective in decreasing the severity of certain withdrawal symptoms but are not as effective as the benzodiazepines and presumably do not protect against seizures, as do benzodiazepines.[99] Thus, these alternative treatments are generally considered to be adjuvants to benzodiazepines. In addition, when benzodiazepines are used, a so-called symptom-triggered approach, in which benzodiazepines are administered in response to signs and symptoms assessed using careful monitoring, may have advantages over more traditional dosing practices.[101]

Alcohol Treatment Programs

Problem drinkers and drinkers who are at risk for alcohol-related problems who do not respond to brief intervention therapy, as well as patients who meet criteria for alcohol abuse and dependence, may require referral to specialists and formal alcohol treatment programs. Such referrals are particularly necessary for patients who are suffering significant medical, psychiatric, or social comorbidity that is related to their alcohol use or for patients who are alcohol dependent. The referral process can be more successful if physicians are familiar with the structure and types of treatment used in their local programs. It is important to communicate effectively with caregivers of alcohol treatment programs and to reinforce their treatment strategies when patients present for follow-up medical care.

Most patients can be treated safely and effectively in an outpatient treatment environment. Criteria have been developed by the American Society of Addiction Medicine[102] that are designed to aid in patient placement. Clinical variables that are important in determining level of service needed include the presence of medical or psychiatric comorbidity, the risk of withdrawal, the level of social support available, and previous treatment experience.

Psychotherapeutic Approaches to the Treatment of Alcohol Dependence

The psychotherapeutic approaches used in alcohol treatment programs may vary from one program to another. Common approaches may be in the form of individual or group therapy. Common objectives of alcohol treatment programs are to (1) motivate patients to change their behavior and lifestyles; (2) teach patients coping skills to avoid alcohol use; (3) encourage patients to develop activities that do not reinforce drinking and that reward abstinence; (4) help patients to improve interpersonal interactions; and (5) promote compliance with pharmacotherapy and medical care.[103] Treatment typically takes a long time and becomes less intensive as patients begin demonstrating achievement of prolonged abstinence. Along with individual and group therapy, other approaches, such as family therapy and network therapy, may have a role in the treatment of selected patients.[104]

Three commonly used psychotherapeutic approaches were evaluated in a randomized clinical trial, Project MATCH.[105] The three approaches studied were cognitive-behavioral coping-skills therapy, motivational-enhancement therapy, and 12-step facilitation. In each program, therapy was administered for 12 weeks. Two parallel but independent randomized clinical trials assessed these approaches in inpatients and outpatients. Both inpatients and outpatients experienced significant and sustained improvements in drinking outcomes, as measured 1 year after treatment. The groups did not differ in their degree of improvement. In the inpatient arm of the study, patients experienced a 70% increase in the percentage of days they were abstinent (from 20% to 90%); 35% remained completely abstinent. In the outpatient arm, patients experienced an 80% increase in the percentage of days they were abstinent; 19% maintained complete abstinence. Of note was the large percentage of both inpatients and outpatients (25% and 35%, respectively) who had a slip in abstinence but not a complete relapse (with relapse being defined as 3 consecutive days of heavy drinking).[105] For the majority of patients enrolled in this study, treatment resulted in significant improvement. In a follow-up evaluation of a subset of individuals enrolled in Project MATCH, their alcohol problems on enrollment in the program were found to be at least as severe as, if not more severe than, those of people who did not participate in the study, suggesting that the positive results of the therapeutic interventions evaluated in the original study could be achieved among larger populations of inpatients and outpatients.[106]

Pharmacologic Treatments to Prevent Relapse

Pharmacotherapy is a useful adjunct to psychotherapy in helping patients drink less alcohol [*see Table 5*].[107] Three medications have been studied extensively for this purpose: disulfiram, naltrexone, and acamprosate. Other medications, including topiramate, nalmefene, lithium, and serotoninergic agents, have also been or are being investigated.[107] Disulfiram was the first of these medications to become available. It works through the inhibition of the enzyme alcohol dehydrogenase. Patients who are receiving disulfiram and who subsequently drink alcohol are at risk for a severe adverse reaction, which includes flushing, nausea, vomiting, and diarrhea. Patients on disulfiram must avoid unintended alcohol consumption, which may occur through the consumption of alcohol-containing foods or medications. Disulfiram has demonstrated limited effectiveness in decreasing alcohol use.[107] It appears to be most effective in combination with behavioral therapies and in highly motivated patients.[107,108] Disulfi-

Table 5 Pharmacotherapeutic Agents

Drug	*Starting Dose*	*Maintenance Dose*	*Interval*	*Comments*
Disulfiram	500 mg	125–500 mg	Every morning for 1–2 wk	Requires careful patient education about disulfiram-alcohol interaction
Naltrexone	50 mg	50 mg	Once a day	Contraindicated in patients with severe liver disease; side effects generally infrequent, mild, and self-limited (e.g., nausea)
Acamprosate	1.3–2.0 g	—	Every day in three divided doses	Not approved by the FDA; side effects generally infrequent, mild, and self-limited (e.g., diarrhea)

ram has also been shown to be helpful in the management of patients who use cocaine and alcohol together.[109]

Naltrexone is an opioid antagonist that was originally developed for the treatment of opioid dependence. Subsequently, naltrexone was shown to decrease the pleasurable effects and craving associated with alcohol use.[110] A meta-analysis of 27 studies indicated that short-term treatment with naltrexone decreases the chance of alcohol relapse for 35% of alcohol-dependent patients and lowers the risk of treatment withdrawal for 28%.[111] Larger randomized, controlled trials are needed to determine how long alcohol-dependent patients should continue naltrexone treatment.

Although most of the naltrexone studies have focused on patients enrolled in formal alcohol treatment programs, one small study found that naltrexone can be effectively administered in primary care settings.[112] In this study, 29 alcohol-dependent persons received naltrexone at a dosage of 50 mg a day for 10 weeks. The majority of patients (72%) completed treatment; 35% relapsed to heavy drinking. When compared with baseline values, all drinking behaviors improved significantly in these individuals, including the percentage of days in which the patients were abstinent, which increased from 36.6% to 88.8%, and the mean number of drinks consumed per occasion, which decreased from 9.5 drinks to 2.5 drinks. Managing both side effects and patients' expectations about the effectiveness of naltrexone seems to be an important factor in determining compliance with medication.[113] Research has demonstrated that naltrexone treatment by primary care providers may be as effective as naltrexone provided in alcohol treatment programs and that extended treatment (6 months or more) with naltrexone may be especially effective in primary care.[114]

Acamprosate, which was approved by the Food and Drug Administration for the treatment of alcohol dependence in 2004,[115] has been studied extensively in Europe and more recently in the United States. In an analysis of 11 randomized, placebo-controlled trials involving a total of 3,338 patients with alcohol dependence, patients who received acamprosate demonstrated superior abstinence rates and duration of abstinence during a 6- to 12-month posttreatment follow-up period.[116] In clinical trials of acamprosate, dosages have ranged from 1.3 to 2.0 g/day in divided doses, and side effects (most commonly diarrhea) have been minimal. Research suggests that the use of acamprosate may be associated with economic as well as clinical benefits.[117]

Concomitant administration of disulfiram has improved the effectiveness of acamprosate without adverse interactions between the drugs.[118] Research is also under way to evaluate the effectiveness of acamprosate in combination with naltrexone.[107] Early research has suggested that this combination results in enhanced treatment outcomes[119]; however, more data are needed to understand the potential benefits of combination therapy. Further study is needed to establish the relative efficacies of these three drugs.

Additional Information

Additional information on the treatment of alcohol problems may be obtained from the National Institute on Alcohol Abuse and Alcoholism (http://www.niaaa.nih.gov), the National Clearinghouse for Alcohol and Drug Information (http://www.health.org), and Alcoholics Anonymous (http://www.alcoholics-anonymous.org).

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Ninth Special Report to the U.S. Congress on Alcohol and Health. U.S. Department of Health and Human Services, Washington, DC, 1997

2. Magruder-Habib K, Durand AM, Frey KA: Alcohol abuse and alcoholism in primary health care settings. J Fam Pract 32:406, 1991

3. Harwood HJ: Updating estimates of the economic costs of alcohol abuse in the United States: estimates, update methods, and data. National Institute on Alcohol Abuse and Alcoholism. Bethesda, Maryland, 2000

4. Saitz R: Clinical practice: unhealthy alcohol use. N Engl J Med 352:596, 2005

5. O'Connor PG, Schottenfeld RS: Patients with alcohol problems. N Engl J Med 338:592, 1998

6. Fiellin DA, Reid M, O'Connor PG: Outpatient management of patients with alcohol problems. Ann Intern Med 133:815, 2000

7. The Physicians' Guide to Helping Patients with Alcohol Problems. National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland, 1995

8. Helping patients with alcohol problems: a health practitioner's guide. National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health. NIH publication No. 03-3769. Rockville, Maryland, 2004

9. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. American Psychiatric Association, Washington, DC, 1994

10. Reid MC, Fiellin DA, O'Connor PG: Hazardous and harmful alcohol consumption in primary care. Arch Intern Med 159:1681, 1999

11. Camargo CA Jr, Hennekens CH, Gaziano JM, et al: Prospective study of moderate alcohol consumption and mortality in U.S. male physicians. Arch Intern Med 157:79, 1997

12. Fuchs CS, Stampfer MJ, Colditz GA, et al: Alcohol consumption and mortality among women. N Engl J Med 332:1245, 1995

13. Mukamal KJ, Conigrave KM, Mittleman MA, et al: Roles of drinking pattern and type of alcohol consumed in coronary heart disease in men. N Engl J Med 348:109, 2003

14. Knoops KT, de Groot LC, Kromhout D, et al: Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. JAMA 292:1433, 2004

15. White IR, Altmann DR, Nanchahal K: Mortality in England and Wales attributable to any drinking, drinking above sensible limits and drinking above lowest-risk level. Addiction 99:749, 2004

16. Saunders JB, Aasland OG, Babor TF, et al: Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. Addiction 88:791, 1993

17. Morse RM, Flavin DK: The definition of alcoholism. The Joint Committee of the National Council on Alcoholism and Drug Dependence and the American Society of Addiction Medicine to Study the Definition and Criteria for the Diagnosis of Alcoholism. JAMA 268:1012, 1992

18. Percent who drink beverage alcohol, by gender, 1939–2003. Online database. National Institute on Alcohol Abuse and Alcoholism. National Institutes of Health http://www.niaaa.nih.gov/databases/dkpat1.htm

19. Behavioral Risk Factor Surveillance System Online Prevalence Data, 1995–2003. Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention http://apps.ncce.cdc.gov/brfss/index.asp

20. Grant BF, Dawson DA, Stinson FS, et al: The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. Drug Alcohol Depend 74:223, 2004

21. Wallace P, Cutler S, Haines A: Randomized controlled trial of general practitioner intervention in patients with excessive alcohol consumption. BMJ 297:663, 1988

22. Fleming MF, Barry KL, Manwell LB, et al: Brief physician advice for problem alcohol drinkers: a randomized controlled trial in community-based primary care practices. JAMA 277:1039, 1997

23. Moore RD, Bone LR, Geller G, et al: Prevalence, detection, and treatment of alcoholism in hospitalized patients. JAMA 261:403, 1989

24. Liskow B, Campbell J, Nickel EJ, et al: Validity of the CAGE questionnaire in screening for alcohol dependence in a walk-in (triage) clinic. J Stud Alcohol 56:277, 1995

25. McQuade WH, Levy SM, Yanek LR, et al: Detecting symptoms of alcohol abuse in primary care settings Arch Fam Med 9:814, 2000

26. Friedmann PD, Saitz R, Gogineni A, et al: Validation of the screening strategy in the NIAAA "Physicians Guide to Helping Patients with Alcohol Problems." J Stud Alcohol 62:234, 2001

27. Schuckit MA: Genetics of the risk for alcoholism. Am J Addict 9:103, 2000

28. Nurnberger JI Jr, Wiegand R, Bucholz K, et al: A family study of alcohol dependence: coaggregation of multiple disorders in relatives of alcohol-dependent probands. Arch Gen Psychiatry 61:1246, 2004

29. Schuckit MA, Mazzanti C, Smith TL, et al: Selection genotyping for the role of 5-HT2A, 5-HT2C and GABA alpha 6 receptors and the serotonin transporter in the level of response to alcohol: a pilot study. Biol Psychiatry 45:647, 1999

30. Light JM, Irvine KM, Kjerulf L: Estimating genetic and environmental effects of alcohol use and dependence from a national survey: a "quasi-adoption" study. J Stud Alcohol 57:507, 1996

31. Charness ME, Simon RP, Greenberg DA: Ethanol and the nervous system. N Engl J Med 321:442, 1989

32. Fields JZ, Turk A, Durkin M, et al: Increased gastrointestinal symptoms in chronic alcoholics. Am J Gastroenterol 89:382, 1994

33. Lieber CS: Alcohol and the liver: 1994 update. Gastroenterology 106:1085, 1994

34. Tanaka E, Yamazaki K, Misawa S: Update: the clinical importance of acetaminophen hepatotoxicity in non-alcoholic and alcoholic subjects. J Clin Pharm Ther 25:325, 2000

35. Bhattacharya R, Shuhart MC: Hepatitis C and alcohol: interactions, outcomes, and implications. J Clin Gastroenterol 36:242, 2003

36. Jaakkola M, Sillanaukee P, Lof K, et al: Amount of alcohol is an important determinant of the severity of acute alcoholic pancreatitis. Surgery 115:31, 1994

37. Preedy VR, Richardson PJ: Ethanol induced cardiovascular disease. Br Med Bull 50:152, 1994

38. Kenchaiah S, Narula J, Vasan RS: Risk factors for heart failure. Med Clin North Am 88:1145, 2004

39. Djousse L, Levy D, Benjamin EJ, et al: Long-term alcohol consumption and the risk of atrial fibrillation in the Framingham Study. Am J Cardiol 93:710, 2004

40. Gavazzi A, De Maria R, Parolini M, et al: Alcohol abuse and dilated cardiomyopathy in men. Am J Cardiol 85:1114, 2000

41. Adami HO, McLaughlin JK, Hsing AW, et al: Alcoholism and cancer risk: a population-based cohort study. Cancer Causes Control 3:419, 1992

42. Tonnesen H, Moller H, Andersen JR, et al: Cancer morbidity in alcohol abusers. Br J Cancer 69:327, 1994

43. Turner RT: Skeletal response to alcohol. Alcohol Cin Esp Res 24:1693, 2000

44. Ellingboe J: Acute effects of ethanol on sex hormones in non-alcoholic men and women. Alcohol Alcohol 1:109 1987

45. Vandenberg MK, Moxley G, Breitbach SA, et al: Gout attacks in chronic alcoholics occur at lower serum urate levels than in nonalcoholics. J Rheumatol 21:700, 1994

46. Farber EM, Nall L: Psoriasis and alcohol. Cutis 53:21, 1994

47. Niquille M, Burnand B, Magnenet P, et al: Dental disease among alcoholic individuals: a comparative study of hospitalized patients. J Gen Intern Med 8:470, 1993

48. Wu LT, Kouzis AC, Leaf PJ: Influence of comorbid alcohol and psychiatric disorders on utilization of mental health services in the National Comorbidity Survey. Am J Psychiatry 156:1230, 1999

49. Regier DA, Farmer ME, Rae DS, et al: Comorbidity of mental disorders with alcohol and other drug abuse: results from the Epidemiologic Catchment Area (ECA) Study. JAMA 264:2511, 1990

50. Kessler RC, Nelson CB, McGonagle KA, et al: The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization. Am J Orthopsychiatry 66:17, 1996

51. Nunes EV, Levin FR: Treatment of depression in patients with alcohol or other drug dependence: a meta-analysis. JAMA 291:1887, 2004

52. Regier DA, Farmer ME, Rae DS, et al: Comorbidity of mental disorders with alcohol and other drug abuse: results from the Epidemiologic Catchment Area (ECA) Study. JAMA 264:2511, 1990

53. Sullivan LE, Fiellin DA, O'Connor PG: The prevalence and impact of alcohol problems in major depression: a systematic review. Am J Med 118:330, 2005

54. Rohde P, Lewinshon PM, Kahler CW, et al: Natural course of alcohol use disorders from adolescence to young adulthood. J Am Acad Child Adolesc Psychiatry 40:83, 2001

55. Guo J, Collins LM, Hill KG, et al: Developmental pathways to alcohol abuse and dependence in young adulthood. J Stud Alcohol 61:799, 2000

56. Rivara FP, Koepsell TD, Jurkovich GJ, et al: The effects of alcohol abuse on readmission for trauma. JAMA 270:1962, 1993

57. Jurkovich GJ, Rivara FP, Gurney JG, et al: The effect of acute alcohol intoxication and chronic alcohol abuse on outcome from trauma. JAMA 270:51, 1993

58. Cherpitel CJ, Bond J, Ye Y, et al: A cross-sectional meta-analysis of alcohol and injury: data from the Emergency Room Collaborative Alcohol Analysis Project. Addiction 98:1277, 2003

59. Whiteman PJ, Hoffman RS, Goldfrank LR: Alcoholism in the emergency department: an epidemiologic study. Acad Emerg Med 7:14, 2000

60. Crawford MJ, Patton R, Touquet R, et al: Screening and referral for brief intervention of alcohol-misusing patients in an emergency department: a pragmatic randomised controlled trial. Lancet 364:1334, 2004

61. Chang G, Astrachan BM: The emergency department surveillance of alcohol intoxication after motor vehicle accidents. JAMA 260:2533, 1988

62. Schermer CR, Gentilello LM, Hoyt DB, et al: National survey of trauma surgeons' use of alcohol screening and brief intervention. J Trauma 55:849, 2003

63. Miller BA, Downs WR, Testa M: Interrelationships between victimization experiences and women's alcohol use. J Stud Alcohol 11(suppl):109, 1993

64. Kyriacou DN, Anglin D, Taliaferro E, et al: Risk factors for injury to women from domestic violence. N Engl J Med 341:1892, 1999

65. Schleifer SJ, Keller SE, LaFarge S, et al: Human immunodeficiency virus–type 1 infection in an inner-city alcohol treatment program. Alcohol Clin Exp Res 20:75, 1996

66. Woods WJ, Lindan CP, Hudes ES, et al: HIV infection and risk behaviors in two cross-sectional surveys of heterosexuals in alcoholism treatment. J Stud Alcohol 61:262, 2000

67. Shillington AM, Cottler LB, Compton WM, et al: Is there a relationship between "heavy drinking" and HIV high risk sexual behaviors among general population subjects? Int J Addict 30:1453, 1995

68. Green JE, Saveanu RV, Bornstein RA: The effect of previous alcohol abuse on cognitive function in HIV infection. Am J Psychiatry 161:249, 2004

69. Breslau N, Peterson E, Schultz L, et al: Are smokers with alcohol use disorders less likely to quit? Am J Public Health 86:985, 1996

70. Hurt RD, Offord KP, Croghan IT, et al: Mortality following inpatient addictions treatment: role of tobacco use in a community-based cohort. JAMA 275:1097, 1996

71. Compton WM, Cottler LB, Phelps DL, et al: Psychiatric disorders among drug dependent subjects: are they primary or secondary? Am J Addict 9:126, 2000

72. Hser YI, Hoffman V, Grella CE, et al: a 33-year follow-up of narcotics addicts. Arch Gen Psychiatry 58:503, 2001

73. D'Amico EJ, Paddock SM, Burnam A, et al: Identification of and guidance for problem drinking by general medical providers: results from a national survey. Med Care 43:229, 2005

74. Amodei N, Williams JF, Seale JP, et al: Gender differences in medical presentation and detection of patients with a history of alcohol use or dependence. J Addict Dis 15:19, 1996

75. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. U.S. Preventive Services Task Force. Ann Intern Med 140:554, 2004

76. Fiellin DA, Reid MC, O'Connor PG: Screening for alcohol problems in primary care: a systematic review. Arch Intern Med 160:1977, 2000

77. Allen JP, Maisto SA, Connors GJ: Self-report screening tests for alcohol problems in primary care. Arch Intern Med 155:1726, 1995

78. Beresford TP, Blow FC, Hill E, et al: Comparison of CAGE questionnaire and computer-assisted laboratory profiles in screening for covert alcoholism. Lancet 336:482, 1990

79. Chan AW, Pristach EA, Welte JW, et al: Use of the TWEAK Test in screening for alcoholism/heavy drinking in three populations. Alcohol Clin Exp Res 17:1188, 1993

80. Adams WL, Barry KL, Fleming MF: Screening for problem drinking in older primary care patients. JAMA 276:1964, 1996

81. Steinbauer JR, Cantor SB, Holzer CE, et al: Ethnic and sex bias in primary care screening tests for alcohol use disorders. Ann Intern Med 129:353, 1998

82. Prochaska JO, DiClemente CC, Norcross JC: In search of how people change: applications to addictive behaviors. Am Psychol 47:1102, 1992

83. Bradley KA, Bush KR, Epler AJ, et al: Two brief alcohol-screening tests from the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. Arch Intern Med 163:821, 2003

84. Cherpitel CJ, Bazargan S: Screening for alcohol problems: comparison of the audit, RAPS4 and RAPS4-QF among African American and Hispanic patients in an inner city emergency department. Drug Alcohol Depend 71:275, 2003

85. Knight JR, Sherritt L, Harris SK, et al: Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcohol Clin Exp Res 27:67, 2003

86. Edens JF, Willoughby FW: Motivational patterns of alcohol dependent patients: a replication. Psychol Addict Behav 14:397, 2000

87. Samet JH, Rollnick S, Barnes H: Beyond CAGE: a brief clinical approach after detection of substance abuse. Arch Intern Med 156:2287, 1996

88. Miller WR, Rollnick S: Motivational Interviewing: Preparing People to Change Addictive Behavior. Guilford Press, New York, 1991

89. Bien TH, Miller WR, Tonigan JS: Brief interventions for alcohol problems: a review. Addiction 88:315, 1993

90. Wilk AI, Jensen NM, Havighurst TC: Meta-analysis of randomized control trials ad-

dressing brief interventions in heavy alcohol drinkers. J Gen Intern Med 12:274, 1997

91. Whitlock EP, Polen MR, Green CA, et al: Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 140:557, 2004

92. Gossop M, Harris J, Best D, et al: Is attendance at Alcoholics Anonymous meetings after inpatient treatment related to improved outcomes? A 6-month follow-up study. Alcohol Alcohol 38:421, 2003

93. Watson CG, Hancock M, Gearhart LP, et al: A comparative outcome study of frequent, moderate, occasional, and nonattenders of Alcoholics Anonymous. J Clin Psychol 53:209, 1997

94. Tonigan JS, Toscova R, Miller WR: Meta-analysis of the literature on Alcoholics Anonymous: sample and study characteristics moderate findings. J Stud Alcohol 57:65, 1996

95. Walsh DC, Hingson RW, Merrigan DM, et al: A randomized trial of treatment options for alcohol-abusing workers. N Engl J Med 325:775, 1991

96. Sullivan JT, Sykora K, Schneiderman J, et al: Assessment of alcohol withdrawal: the revised clinical institute withdrawal assessment for alcohol scale (CIWA-Ar). Br J Addict 84:1353, 1989

97. Kosten TR, O'Connor PG: Management of drug and alcohol withdrawal. N Engl J Med 348:1786, 2003

98. Fiellin DA, O'Connor PG, Holmboe ES, et al: Risk for delirium tremens in patients with alcohol withdrawal syndrome. Subst Abus 23:83, 2002

99. Mayo-Smith MF: Pharmacological management of alcohol withdrawal: a meta-analysis and evidence-based practice guideline. American Society of Addiction Medicine Working Group on Pharmacological Management of Alcohol Withdrawal. JAMA 278:144, 1997

100. Mayo-Smith MF, Beecher LH, Fischer TL, et al: Management of alcohol withdrawal delirium: an evidence-based practice guideline. Arch Intern Med 164:1405, 2004

101. Reoux JP, Miller K: Routine hospital alcohol detoxification practice compared to symptom triggered management with an Objective Withdrawal Scale (CIWA-Ar). Am J Addict 9:135, 2000

102. Patient Placement Criteria for the Treatment of Substance-Related Disorders, 2nd ed. (revised) American Society of Addiction Medicine, Chevy Chase, Maryland, 2001 http://www.asam.org

103. Carroll KM, Schottenfeld R: Nonpharmacologic approaches to substance abuse treatment. Med Clin North Am 81:927, 1997

104. Galanter M, Brook D: Network therapy for addiction: bringing family and peer support into office practice. Int J Group Psychother 51:101, 2001

105. Matching Alcoholism Treatments to Client Heterogeneity: Project MATCH post-treatment drinking outcomes. Project MATCH Research Group. J Stud Alcohol 58:7, 1997

106. Velasquez MM, DiClemente CC, Addy RC: Generalizability of Project Match: a comparison of clients enrolled to those not enrolled in the study at one aftercare site. Drug Alcohol Depend 59:177, 2000

107. Kenna GA, McGeary JE, Swift RM: Pharmacotherapy, pharmacogenomics, and the future of alcohol dependence treatment, part 1. Am J Health Syst Pharm 61:2272, 2004

108. Saitz R, O'Malley SS: Pharmacotherapies for alcohol abuse: withdrawal and treatment. Med Clin North Am 81:881, 1997

109. Carroll KM, Nich C, Ball SA, et al: One-year follow-up of disulfiram and psychotherapy for cocaine-alcohol users: sustained effects of treatment. Addiction 95:1335, 2000

110. Volpicelli JR: Naltrexone in alcohol dependence. Lancet 346:456, 1995

111. Srisurapanont M, Jarusuraisin N: Opioid antagonists for alcohol dependence. Cochrane Database Syst Rev (1):CD001867, 2005

112. O'Connor PG, Farren CK, Rounsaville BJ, et al: A preliminary investigation of the management of alcohol dependence with naltrexone by primary care providers. Am J Med 103:477, 1997

113. Rohsenow DJ, Colby SM, Monti PM, et al: Predictors of compliance with naltrexone among alcoholics. Alcohol Clin Exp Res 24:1542, 2000

114. O'Malley SS, Rounsaville BJ, Farren C, et al: Initial and maintenance naltrexone treatment for alcohol dependence using primary care vs specialty care: a nested sequence of 3 randomized trials. Arch Intern Med 163:1695, 2003

115. Acamprosate campral for alcoholism. Med Lett Drugs Ther 47:1, 2005

116. Wilde MI, Wagstaff AJ: Acamprosate: a review of its pharmacology and clinical potential in the management of alcohol dependence after detoxification. Drugs 53:1038, 1997

117. Palmer AJ, Neeser K, Weiss C, et al: The long-term cost-effectiveness of improving alcohol abstinence with adjuvant acamprosate. Alcohol Alcohol 35:478, 2000

118. Besson J, Aeby F, Kasas A, et al: Combined efficacy of acamprosate and disulfiram in the treatment of alcoholism: a controlled study. Alcohol Clin Exp Res 22:573, 1998

119. Kiefer F, Jahn H, Tarnaske T, et al: Comparing and combining naltrexone and acamprosate in relapse prevention of alcoholism: a double-blind, placebo-controlled study. Arch Gen Psychiatry 60:92, 2003

VII SCHIZOPHRENIA

William T. Carpenter, Jr., m.d.
Gunvant K. Thaker, m.d.

As recently as the mid-20th century, schizophrenia was one of the most devastating psychiatric illnesses. No effective treatment was available, and most patients suffered lifelong debilitating symptoms and required institutionalization. Research during the latter part of the 20th century better defined the syndrome, provided clues to the causes, clarified the pathophysiology, provided evidence for effective treatments and against potentially harmful ones, and permitted the transfer of many schizophrenic patients from inpatient institutions to less restrictive outpatient facilities. However, challenges remain: specific causes and causal pathways are unknown, pathophysiology is ill defined, treatments are effective only for the psychotic aspect of the syndrome, and homelessness has replaced institutionalization.

Definition

Schizophrenia is a clinical syndrome probably comprising several as-yet undefined disease entities. There is substantial heterogeneity between cases, which presumably reflects multiple overlapping etiologic factors, including contributions from several genes. The presence of schizophrenia is indicated by chronic psychotic symptoms, especially hallucinations and delusions. Disorganization of thought and behavior are common and distinguish schizophrenia from the many other causes of reality-distortion symptoms (e.g., psychotomimetic drugs). Diminished emotional experience and expression, low drive, and reduced speech are observed in a subgroup of patients. Most patients have subtle impairments in cognition. Cognitive impairments and the emotional and social aspects of the disease often appear early in life; the psychotic symptoms typically begin in late adolescence or early adulthood in men and a little later in women.

Subclassification

At the end of the 19th century, Emil Kraeplin conceptualized dementia praecox as comprising seemingly discrete diseases known as hebephrenia, catatonia, and paranoia. Eugen Bleuler suggested renaming dementia praecox as *schizophrenia*, a term he coined to capture the split within thought and between thought and emotion. Bleuler also added simple schizophrenia as a fourth subtype of the disease.

Throughout most of the 20th century, the clinical presentation of schizophrenia was often described in terms of specific subtypes. These subtypes included hebephrenic, paranoid, catatonic, and simple schizophrenia. Hebephrenic schizophrenia (now termed disorganized schizophrenia) is characterized by shallow and incongruent affect and disorganized thought and behavior. Paranoid schizophrenia is characterized by predominance in males; a later age of onset; intact cognition and affect; and hallucinations and delusions, often persecutory and bizarre. Catatonic schizophrenia is characterized by extreme psychomotor manifestations, including stupor, prolonged posturing, or excitement; it must be differentiated from periodic catatonia, which is a separate syndrome unrelated to schizophrenia. For unknown reasons, catatonic schizophrenia is now rarely seen in developed countries. Simple schizophrenia denotes a milder psychosis (i.e., fewer hallucinations and delusions and less disorganization), with cases typically featuring a schizoid lifestyle of reduced emotional experience and expression and reduced social drive and engagement. A later term, schizoaffective schizophrenia, described the combination of mood disturbances—mania or depression—with schizophrenic psychotic manifestations. Other subtypes were suggested but did not gain wide acceptance.

In general, however, schizophrenia has been approached as a diagnostic class rather than in terms of these subtypes. The traditional subtypes appear not to differ substantially in etiology, pathophysiology, or treatment response. Moreover, a patient's subtype may change over time: a patient who presents with paranoid schizophrenia may later be seen to have disorganized schizophrenia; a patient who presents with schizoaffective psychosis may eventually be seen to have a disorganized or undifferentiated subtype.

Another approach to subclassification categorizes cases on the basis of three distinct domains of psychopathology: reality distortion (i.e., hallucinations and delusions), disorganization of thought and behavior, and negative symptoms such as reduced drive and affect. Kraeplin described two "maladies"—dissociation and avolition—as the central features of dementia praecox, but in the late 20th century, psychiatrists placed diagnostic emphasis on reality distortion and disorganization of thought. It turns out that many patients do not manifest negative symptoms, whereas others have chronic reduction in drive, reduced expression of emotion, and poverty of speech. We have divided schizophrenia into two subgroups: deficit schizophrenia, in which schizophrenia causes negative symptoms and psychosis; and nondeficit schizophrenia, in which negative symptoms are absent. Clinical features, family history, neuroanatomy, treatment response, course of illness, and postmortem pathology suggest that deficit schizophrenia is a separate disease entity within the schizophrenia syndrome. Deficit schizophrenia constitutes about 15% to 20% of all schizophrenia cases and appears to occur predominantly in males.[1]

Epidemiology

The annual incidence of schizophrenia approximates one in 10,000 adults younger than 60 years, with a lifetime risk of 0.8. The onset of psychosis (i.e., reality distortion or disorganization of thought) usually occurs at about 17 to 27 years of age in males and at 20 to 37 years in females. Population-based estimates suggest that the lifetime incidence of the disease is roughly equal in males and females; in clinical populations, however, males often constitute the majority of patients. Females have often been reported to have a more favorable course of illness. One explanation is that males are more likely to have deficit schizophrenia, which tends to have a chronic course with poor functional oucome.[2]

Race does not appear to influence the distribution of schizophrenia. Some United States studies have suggested a racial bias in diagnostic practice, with the term schizophrenia being used more in African-American patients and the term affective psychosis used more in white Americans.[3]

Geographically, schizophrenia appears to be evenly distributed throughout the world.[4] A country's stage of economic development, degree of urbanization, or climate seem to have little influence on schizophrenia prevalence. However, the course of the illness has been more favorable in developing countries than in developed countries. This may reflect the tendency for developing countries to have sociocentric cultures and for developed countries to have egocentric cultures. Sociocentric cultures may be more supportive of marginal function than egocentric cultures, in which demands for autonomous function may be particularly challenging for persons with schizophrenia. The difference may also reflect the inclusion of cases with a better prognosis in the syndrome definition; patients with acute onset of bizarre psychosis tend to have a more favorable course of illness.

Decreases in rates of schizophrenia were reported in the second half of the 20th century.[5] However, these decreases have not been substantiated and may simply reflect increasingly stringent criteria for the diagnosis.

Schizophrenia is the fourth leading cause of disability in adults worldwide. In the United States, about 80% of persons with schizophrenia are unemployed, and many of the remainder are underemployed. Perhaps a third of homeless persons have schizophrenia.[6]

More than half of schizophrenic patients abuse substances, especially alcohol.[7] Over 80% of schizophrenics smoke cigarettes, and nicotine intake in schizophrenics who smoke is well above that of nonschizophrenic smokers.[8] Smoking represents a substantial financial burden to this generally impoverished population and combines with other lifestyle risk factors (e.g., inadequate diet and exercise) to place schizophrenic patients at increased risk for diabetes, cardiovascular disease, and cancer.

Etiology and Genetics

Considerable data suggest that schizophrenia is a heritable illness. The risk of the disease rises with the closeness of kinship to a schizophrenic proband. Compared with the risk in the general population, the risk is fivefold higher in second-degree relatives of schizophrenia patients, 10-fold to 15-fold higher in first-degree family members and dizygotic twins, and 40-fold to 50-fold higher in monozygotic twins or in persons whose parents both have schizophrenia.[9] There is also higher risk in adopted biologic relatives of schizophrenia probands than in their adoptive families.

Other data strongly suggest that environmental factors, either on their own or interacting with genetic factors, play a critical role in causing or modifying the clinical presentation of the illness.[10] Most of the environmental factors associated with schizophrenia are active early in development, particularly during the intrauterine period. These include maternal malnutrition; infections during critical periods of fetal development; fetal hypoxia; and other birth and obstetric complications. Further supporting an environmental role is the finding that overall, patients with schizophrenia were more likely to have been born during the winter, whereas the subgroup of patients with deficit schizophrenia were more likely to have been born during the summer.[11] How season of birth affects risk is unclear. Environmental insults that occur later in life, such as use of psychoactive drugs, also play a role.

The initial search for genes that might cause schizophrenia produced disappointing results. The first significant linkage of a broadly defined schizophrenia phenotype to a specific locus, chromosome 5q11-q13, stirred excitement, but subsequent studies failed to replicate this finding.[12] Ensuing linkage studies identified several chromosomal loci, some of which were likely false positive findings, given that replications were not forthcoming. Significant linkage findings that were subsequently replicated by other studies implicated regions on chromosomes 1, 5, 6, 8, 13, 15, and 22.[13,14] Most of these findings were not robust, however; in addition, some studies failed to identify linkages at the same locus, and many replicating studies only suggested linkage (i.e., log odds ratio scores ranged from 2 to 3). Further investigation of these loci has identified several candidate genes: *RGS4,* the dysbindin gene (*DTNBP1*), neuregulin 1, the alpha-7 nicotinic receptor gene, *COMT, G72,* and *GAD 65/67*.[15-21] Further work is needed to identify other candidate genes using other genetic approaches (see below) and to confirm the role of candidate genes in the etiology, pathophysiology, and treatment of schizophrenia.

The search for genes in schizophrenia has proved particularly challenging because of the complex and multifactorial nature of the disorder. It appears that many genes may cause vulnerability to schizophrenia, each with a modest effect, and that these genes probably interact with one another and with environmental factors. Thus, the clinical phenotype of schizophrenia may well be the common expression of a variety of genotypes; this may explain failures to replicate identified disease loci.

Arguing that schizophrenia encompasses multiple phenotypes of different genetic vulnerability profiles, several investigators have focused on a search for the genes underlying specific deficits seen in schizophrenia. Freedman and colleagues, using an electrophysiologic measure of sensory gating, demonstrated that a deficit in sensory gating in familes with schizophrenia is a result of an autosomal dominant effect of a single gene, the alpha-7 nicotinic receptor gene, on chromosome 15q14.[22] Other phenotypes that are associated with schizophrenia and that have been used in the search for schizophrenia-related genes are smooth-pursuit eye movement deficit; cognitive impairments in dimensions of attention, language, and memory; deficits in levels of the neuronal marker *N*-acetyl-aspartate in the hippocampal region; and abnormal membrane phospholipid metabolism in the prefrontal cortex, as measured by phosphorus-31 magnetic resonance spectroscopy.[23,24] Preliminary findings have suggested linkage of the smooth-pursuit phenotype to a locus on chromosome 6p and an association of prefrontal executive functional deficit with the *val* allele of the *COMT* gene.[25,26]

Pathophysiology

The pathophysiology of schizophrenia is not well understood. That schizophrenia results from an excess of dopamine has been the dominant hypothesis for decades, with several modifications and expansion of the hypothesis made during the 1990s.[27-29] Brain imaging studies have shown evidence of increased synaptic dopamine release in response to amphetamine administration in the striatum of schizophrenic patients.[30-32] Excess dopamine may explain the positive symptoms but not the negative symptoms or other aspects of the disease. The pathophysiologic mechanisms are presumably complex—even a deficiency of dopamine is a viable explanation for negative symptoms. The clinical observations that long-term phencyclidine (PCP) use produces a broader cluster of schizophrenia-like symptoms than seen with use of dopamine agonists has led to

speculation that a lack of glutamine may be involved. In contrast to the predominantly paranoid symptoms associated with long-term use of dopamine agonists, long-term PCP use produces negative symptoms and cognitive impairment in addition to reality distortion.[28] These schizophrenia-like symptoms from PCP are hypothetically initiated by blockade of the ion channel in the *N*-methyl-D-aspartate (NMDA) receptor complex, resulting in diminished glutamatergic neurotransmission. Both dopaminergic and glutamatergic terminals converge on the spines of pyramidal neurons in cortex. This suggests a shared mechanism by which glutamate and dopamine play a role in schizophrenia.[28,29] Although the two neurotransmitter systems share complex presynaptic and postsynaptic interactions,[33,34] modulation of glutamate release by D_1 dopamine receptors may mediate some of the effects of dopamine on psychosis.

Postmortem findings in schizophrenia have not been consistent. Furthermore, interpretation of postmortem findings is difficult because of the confounding effects of long-term drug treatment received by most patients. Massive cell loss or gliosis is generally not found, thus ruling out degenerative changes in the brain. Earlier postmortem histopathologic examination of the brains of patients with schizophrenia showed slight reductions in neocortical gray matter volume, decreased neuronal size, and neuronal disarray in several corticolimbic structures.[35-37] Subsequent investigations have focused on changes of neuronal connectivity or microcircuitry within the cortical layers and have noted an increase in cell packing density without a change in the number of neurons. These findings suggest a decrease in neuropil and a reduction in dendritic spine density on pyramidal neurons in the prefrontal cortex.[38,39] Using complementary DNA microarrays, Mirnics and colleagues noted a decrease in the expression of several genes involved in glutamate and γ-aminobutyric acid (GABA) transmission and in the regulation of presynaptic function and signal termination in the prefrontal cortex of patients with schizophrenia.[40] Together, these findings indicate a decrease in cortical or thalamic excitatory synaptic inputs to the pyramidal neurons.

In addition to drug probe and postmortem studies exploring the pathophysiology of schizophrenia, in vivo structural imaging studies of the brain show cortical volume reductions of one or more of the constituent structures of the medial temporal lobe (comprising the hippocampus, the amygdala, and the parahippocampal gyrus). Some studies note smaller prefrontal cortical volume, but this finding is not consistent across studies and may be limited to a subgroup of schizophrenic patients.[41] Reductions in thalamic volume, particularly of the thalamic nuclei forming part of the temporal-thalamic and prefrontal-thalamic circuitry, are also noted.[42] Functional imaging studies of neuronal circuitry during performance of cognitive tasks have found reduced neuronal activation in several cortical regions of schizophrenic patients compared with normal control subjects. However, interpretation of these data is somewhat confounded by the fact that schizophrenic patients perform poorly in such tasks. Several studies have examined brain activation during the performance of tasks that schizophrenic patients performed as well as the control subjects. These studies show that to perform equally well on a working memory task, schizophrenic patients show increased activation of the prefrontal cortical network. On other tasks, patients activated different cortical networks than were activated by healthy control subjects, suggesting that the patients used a different strategy or required compensatory mechanisms [*see Figure 1*].[43]

Clinical Course

Persons with schizophrenia often come to medical attention because of a public manifestation of bizarre behavior, disorganized thought, and reality-distortion symptoms. However, the history taken at first presentation often reveals a prolonged period of more subtle manifestations of the psychotic state, averaging perhaps 2 years; cognitive impairments and negative symptoms may have been present for years, if not throughout life. Although the course of schizophrenia varies considerably between individuals, a brief case description will illustrate several characteristic features of the disease.

A young man had a declining academic performance during the last 2 years of high school. He seemed to lose friends and motivation and gradually became a loner with strange ideas. He relinquished plans to attend college, and after high-school graduation he took a job. Soon after he began working, he acted in a bizarre and hostile manner in a confrontation; recurrence of this behavior caused dismissal from the job. He spent the next year working odd jobs, living at home, and displaying a growing indifference to his life goals. His parents suspected substance abuse and were concerned that he was not interested in having a girlfriend. On his 20th birthday, he was found sitting nude in the backyard; he claimed that he was receiving messages through solar radiation. He was taken to the emergency department.

As this case illustrates, the opportunity for intervention often occurs before the conspicuous onset of illness. Although the development of overt psychotic symptoms defines the onset of the illness, subtle impairments occur early in some patients and persist through the peaks and troughs of the psychotic symptoms. Some schizophrenic patients appear to be born with subtle motor and emotional impairments. Early in childhood and adolescence, many persons who later develop schizophrenia manifest a range of cognitive impairments on psychological tests of attention, memory, and executive function and on physiologic tests of information processing.

There is a variable course of illness after the initial psychotic break. Later in life, some patients even show clinical improvement in psychotic symptoms. In contrast, a smaller subgroup may show rapid cognitive decline. The manifestation of psychosis only modestly defines long-term morbidity, which instead is best understood in terms of functional capabilities, social and occupational roles, and quality of life. Cognitive impairments and negative symptoms are more robust determinants of these outcomes.

Diagnosis

CLINICAL MANIFESTATIONS

Diagnosis of schizophrenia is based on clinical manifestations and the clinical course. Internationally accepted criteria have been published in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders—Text Revision* [*see Table 1*] and in the 10th edition of the *International Classification of Diseases and Related Health Problems*.

The diagnosis of schizophrenia should be considered in a patient who presents with hallucinations and delusions. The presence of disorganized thought and behavior increases the likelihood of schizophrenia. In the absence of other known causes of such symptoms (e.g., substance abuse, temporal lobe epilepsy), the principal diagnostic task is to discriminate between severe

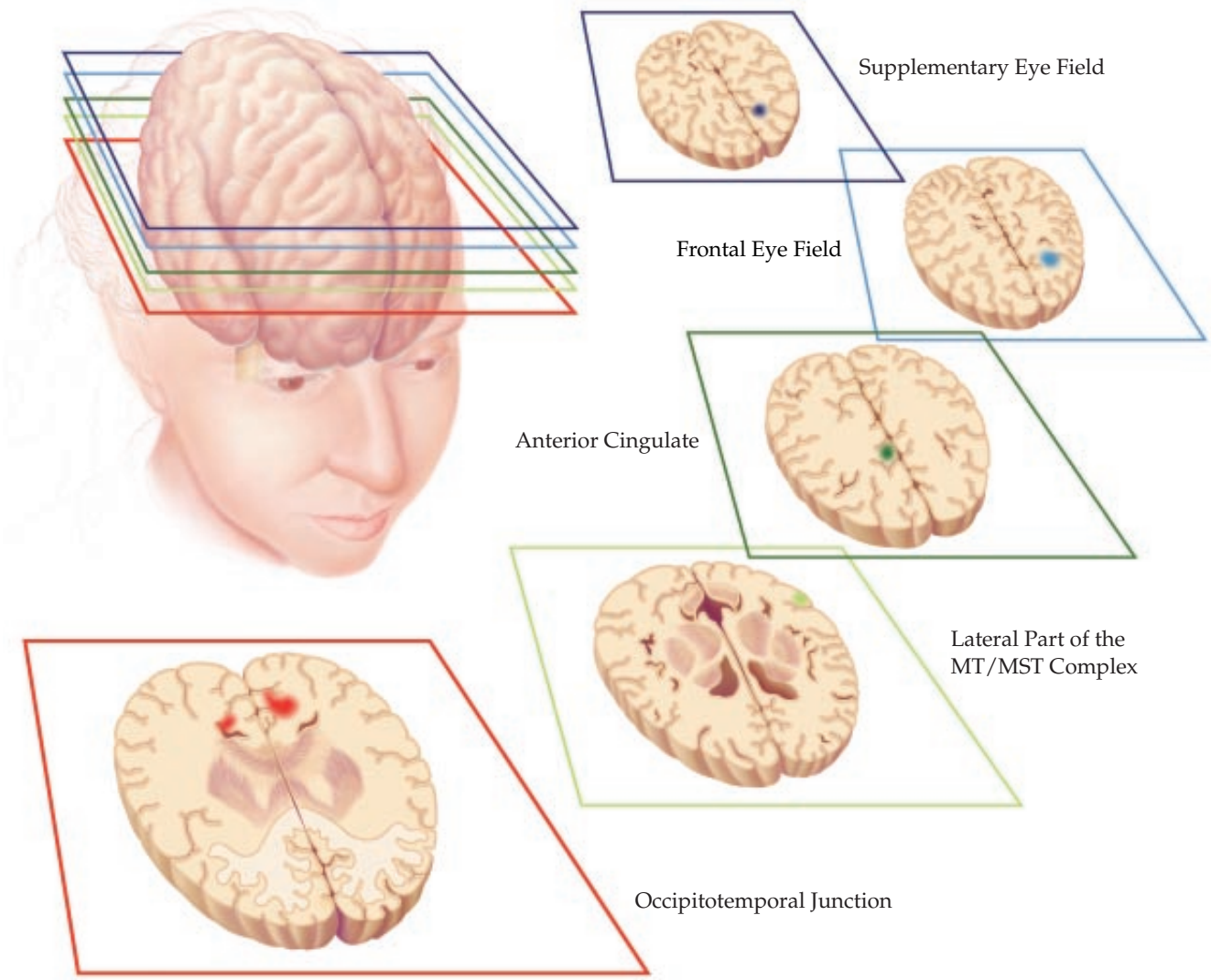


Figure 1 **In a functional magnetic resonance imaging (fMRI) study, schizophrenic patients and healthy comparison subjects were matched on the basis of how well they were able to keep their eyes on a moving target (so-called smooth-pursuit eye movement). Earlier data suggested that persons at risk for schizophrenia use a strategy different from that of healthy control subjects to maintain normal smooth pursuit. Whereas healthy persons depend more on the internal representation of target velocity (so-called extraretinal motion signals) and less on the immediate visual feedback from the retina (retinal motion signals), persons at risk for schizophrenia (i.e., relatives of schizophrenic patients) depended more on the retinal signals.[53] In schizophrenic patients who were able to maintain smooth pursuit, as well as in healthy control subjects, the fMRI analyses showed increased activation in some regions and decreased activation in others. A comparison of related brain activation in control subjects and relatives of schizophrenic patients showed that the latter had reduced activities in left frontal eye field, left supplementary eye field, anterior cingulate, and left medial superior temporal lobe; many of these areas have previously been identified as sites of extraretinal motion processing in nonhuman primates. In addition, relatives of schizophrenic patients showed increased activation in the occipitotemporal junction, an area at the medial part of the middle temporal (MT) lobe/medial superior temporal (MST) lobe that is commonly associated with retinal motion information processing.**

mental illnesses. If the patient does not have a mood disorder and the psychotic symptoms are accompanied by restricted affect, low drive, and speech disturbance [*see Table 2*], the probability of schizophrenia is high. The diagnosis can be made with greatest confidence, however, on the basis of the longitudinal pattern of the disorder, which includes the occurrence of prodromal symptoms before the initial episode, residual symptoms between psychotic episodes, and psychotic episodes that cannot be attributed to mood disturbance (e.g., manic or depressive psychosis) or other known causes of psychotic behavior [*see* Differential Diagnosis, *below*].

PHYSICAL EXAMINATION

Physical examination is unlikely to provide evidence for the diagnosis of schizophrenia but may help rule out rare cases of psychosis caused by physical ailments. Evidence of self-mutilation can alert the physician to potential risks and affect the choice of inpatient or outpatient treatment.

Although physical findings are not very informative for diagnosis, they are critical in guiding selection of drugs for treatment. About 12% to 15% of schizophrenic patients manifest motor abnormalities before antipsychotic drug treatment is initiated. The finding of involuntary movements during physical

examination should direct the clinician to select a drug with a low potential for causing extrapyramidal symptoms or tardive dyskinesia. The presence of obesity dictates avoidance of drugs that increase weight.

LABORATORY TESTS

Toxicology screens are important in the workup of schizophrenia. Positive results may provide an alternative explanation for psychosis or may confirm comorbid substance abuse. Lipid profiles may identify patients at risk for hyperlipidemia associated with the use of some antipsychotic drugs, and a metabolic profile may identify patients at risk for diabetes. An electrocardiogram is relevant as a baseline measure for patients who are subsequently prescribed antipsychotic medications that can prolong the QT interval. Electroencephalography may be used to exclude the diagnosis of temporal lobe epilepsy. Rarely, unusual presentations (e.g., the emergence of symptoms of schizophrenia after head trauma) may require magnetic resonance imaging of the brain to rule out other causes of psychosis.

There are a number of physiologic and neuroimaging findings that are helpful in distinguishing groups of schizophrenic patients from groups of healthy control subjects. However, these findings may not be unique to patients with schizophrenia, and their accuracy is inadequate for diagnostic purposes. Such findings include impairment of sensory gating as measured by evoked-potential response of P50 and by oculomotor physiologic dysfunction and prefrontal cortical dysfunction as assessed by brain imaging studies. Differences found on testing groups of schizophrenic patients have not led to the development of tests with adequate sensitivity and specificity and positive or negative predictive power for diagnosis in individual cases.

Table 1 Diagnostic Criteria for Schizophrenia[54]

All six criteria must be met for a patient to fit the diagnosis of schizophrenia.

1. *Presence of psychosis for a significant length of time*
   Any two or more of the following symptoms must be present for a significant duration during a 1-month period (less if treated):
   1. Delusions
   2. Hallucinations
   3. Disorganized speech
   4. Negative symptoms (e.g., lack of emotion and social drive and engagement)
   5. Grossly disorganized behavior
2. *Deterioration in social/occupational functioning*
   Functioning in at least one of the following areas is markedly below the level achieved before the illness:
   1. Work
   2. Interpersonal relationships
   3. Self-care
3. *Continuous signs of the disturbance for 6 months*
   The 6 months must include at least 1 month of positive symptoms. However, during prodromal or residual periods, the disturbance may be manifested by negative symptoms only.
4. *Exclusion of schizoaffective and mood disorders*
   No major depressive, manic, or mixed episodes have occurred concurrently with the active-phase symptoms; if such episodes have occurred, their total duration has been much briefer than the periods of active and residual psychosis.
5. *Exclusion of substance-induced or medical disorders*
   The disturbance is not the result of direct physiologic effects of a substance or caused by a medical condition.
6. *Relationship to pervasive developmental disorders*
   If the patient has a history of autistic disorder or another pervasive developmental disorder, a diagnosis of schizophrenia is made only if prominent delusions or hallucinations are also present for at least 1 month.

Table 2 Cognitive Disturbances Seen in Schizophrenic Speech

Abstract Thought Impairment
For example, when asked to interpret a proverb, the patient may provide a concrete description

Blocking
A halt in the flow of speech; the person is unable to continue the train of thought even with cueing

Clang Associations
Speech connections based on similarities in the sounds of words rather than on their meanings

Delayed Response
Long latency before responding to questions; at the extreme, the patient may be mute

Poverty of Content
The speech conveys little information even when it is more or less coherent

Tangentiality
Response to questions such that the answers become progressively less related to the original question

Loosening of Associations
Speech consisting of ideas that have no discernible connection to one another

Word Salad
Incoherent speech from extreme loosening of associations

Genetic Testing

Rapidly advancing knowledge concerning genotypic vulnerability to schizophrenia may lead to better susceptibility testing. However, it appears unlikely that genotypes unique to schizophrenia will be defined. At present, genetic information mainly suggests that a number of genes that individually make minor contributions to vulnerability may combine in persons with schizophrenia and may interact with certain environmental etiologic factors. It is also suggested that genes conferring vulnerability to schizophrenia may overlap with those that predispose to depression and bipolar disorder. If these suggestions prove true, specific genetics-based diagnoses of schizophrenia are unlikely to be forthcoming in the near future. However, genetic testing may eventually help guide therapy in individual patients by indicating their likelihood of response and risk of side effects with specific drugs.

Differential Diagnosis

Because schizophrenia is defined as a psychotic illness with functional impairments, distinguishing schizophrenia from normalcy is usually not difficult. However, because the clinician depends on information provided by patients regarding their inner experiences, recognition of mild psychosis is difficult in patients who are guarded and who are not forthcoming with information.

Schizotypal personality disorder shares some of the clinical characteristics of schizophrenia, such as social and physical an-

hedonia, suspiciousness, magical thinking, blunting of affect and emotional experience, and poor functioning. However, schizotypal patients do not experience overt and persistent psychotic symptoms, although rare and brief psychotic symptoms may occur.

In a patient with persistent psychosis, the differential diagnosis consists mainly of affective disorders with psychosis, substance abuse, and delusional disorders. Psychosis that coincides with depression is typically associated with such affective features as delusions of poverty or accusatory voices. Similarly, delusions of grandeur are common during manic episodes. Psychotic symptoms in affective disorders typically follow the emergence of depression or mania and fade once the affective symptoms recede. Depressive symptoms often occur in schizophrenia, particularly in the first episode. Alternatively, excitement and hyperactivity in patients with schizophrenia may mimic manic symptoms. Differential diagnosis depends on the timing of psychosis in relation to the occurrence of affective symptoms and the content of delusions and hallucinations. Psychotic symptoms can be caused by many medical conditions [*see Table 3*], as well as by alcohol, illicit drugs, and over-the-counter and prescription medications [*see Table 4*]. The history and toxicology screen can rule out psychosis caused by drug abuse, such as use of PCP or the long-term abuse of steroids. Delusional disorder is diagnosed on the basis of persistent and circumscribed delusions (more realistic and less bizarre than those associated with schizophrenia) in the absence of disorganization and negative psychopathology.

Occasionally, neurologic conditions such as brain tumor or temporal lobe epilepsy may be misdiagnosed as schizophrenia. When such conditions are suspected, MRI and EEG can help with the diagnosis.

Table 3 Drugs and Medications Commonly Associated with Psychotic Symptoms

Abused Drugs
- Alcohol (withdrawal and alcoholic hallucinosis)
- Amphetamines
- Cocaine (especially free-base or crack)
- Marijuana (panic reaction)
- Phencyclidine (PCP) and hallucinogens (acute intoxication and flashbacks)
- Sedative-hypnotics (withdrawal)

Analgesics
- Meperidine (toxic metabolite)
- Nonsteroidal anti-inflammatory drugs
- Pentazocine and other opiate mixed agonist-antagonists

Antibiotics
- Antituberculous drugs (e.g., cycloserine, isoniazid, rifampin)
- Many others

Anticholinergics
- Anticholinergic antiparkinsonians
- Antihistamines (e.g., diphenhydramine)
- Atropine and derivatives
- Cyclic antidepressants
- Low-potency antipsychotic drugs (e.g., thioridazine, clozapine)
- Meclizine
- Scopolamine

Cardiovascular Drugs
- Antiarrhythmics (e.g., amiodarone, digitalis, procainamide)
- Captopril

Dopamine Agonists
- Amantadine
- D_2 dopamine receptor agonists (e.g., bromocriptine, pergolide)
- Levodopa

Endocrine Drugs
- Estrogens, testosterone
- Glucocorticoids and adrenocorticotropic hormone (ACTH)
- Thyroid hormone (overdose)

H_2 Receptor Antagonists
- Cimetidine, ranitidine

Stimulants and Sympathomimetics
- Dextroamphetamine
- Methylphenidate
- Over-the-counter decongestants (e.g., pseudoephedrine), diet pills, and pep pills

Table 4 Medical Conditions Associated with Psychotic Symptoms

Cancer and Hematologic Malignancies
- CNS neoplasm
- Hyperviscosity syndromes (resulting from hematologic malignancy)
- Paraneoplastic syndromes (e.g., limbic encephalitis)

Cardiovascular
- Anoxia and postanoxic encephalopathy
- Hypertensive encephalopathy

Infections and Sequelae
- AIDS encephalopathy
- Encephalitis, meningitis, brain abscess
- Lyme disease
- Neurosyphilis

Metabolic and Endocrine Disorders
- Acute intermittent porphyria
- Addison disease
- Cushing disease
- Hepatic encephalopathy
- Hypocalcemia and hypercalcemia
- Hypoglycemia
- Hypothyroidism and hyperthyroidism

Neurologic Disorders
- Alzheimer disease
- Complex partial seizures
- Huntington chorea (Huntington disease)
- Multiple sclerosis
- Stroke
- Wilson disease

Nutritional Deficiencies
- Folic acid deficiency
- Pellagra (niacin deficiency)
- Vitamin B_{12} deficiency

Rheumatologic Disorders
- Lupus cerebritis

Management

Patients with schizophrenia require acute treatment for psychotic exacerbations; long-term treatment, involving mainte-

nance and prophylactic strategies to sustain symptom improvement and prevent recurrence of psychosis; and support for living. Unfortunately, the avolitional (negative) symptoms and impairments in cognition, which account for poor functional outcomes, show little response to current treatments. The future may offer treatment for these nonpsychotic aspects of the illness, as well as preventive measures for persons at high risk for schizophrenia.

ACUTE TREATMENT

The first-line therapeutic approach for an acute exacerbation of schizophrenia—whether it is the first episode or a relapse—involves antipsychotic medication (see below). To determine whether the patient requires hospitalization, the physician must evaluate patient safety and cooperation. This evaluation is clinically based and often includes discussion with family members or other informants. The patient's ability to care for himself or herself, the setting most likely to ensure adherence to therapy, and the risk of suicidal or aggressive behavior are at the forefront of this evaluation.

MAINTENANCE THERAPY

Most patients show at least a partial therapeutic response to the initiation of antipsychotic medication. Longer-term care involves maintenance therapy to minimize the expression of psychosis and prevent or delay psychotic exacerbation. Some patients will have a full remission of psychotic symptoms, but even in such cases, the continuous administration of antipsychotic drugs will delay recurrence of a psychotic exacerbation. Psychosocial treatments and pharmacologic treatments have proved effective in reducing exacerbation rates over the long term. The clinical relationship involved in these treatments also provides for ongoing evaluation of life needs, suicide risk, medication adherence, and success in coping with the many practical problems facing individuals with a chronic debilitating disease. Many patients need assistance in obtaining shelter, in developing an occupational niche or receiving vocational rehabilitation, and in participating in substance abuse programs.

Involvement of the people in the patient's life is critically important. Family members and significant others need to be educated about the disease, its treatment, and the approach to changes in clinical status.

If the patient recovers from psychosis, remains free of psychotic symptoms, and is stable for several years, the gradual discontinuance of medication may be warranted, provided that close clinical observation can be ensured. However, most patients will experience subsequent exacerbations. Early detection and intervention may be effective. Because a few patients have only a single episode of psychosis and may not require continuous antipsychotic treatment and because the risk of medication side effects is substantial, a medication-free trial is justified in selected cases.

Pharmacologic and psychosocial management need to be integrated. The two approaches appear to be synergistic in their effects, and each is crucial to the effectiveness of the other. Early detection of exacerbation is most likely when continuous clinical observation is provided in the context of psychosocial treatment; successful participation in skills training, vocational rehabilitation, living arrangements, and substance abuse treatment is most likely to be successful in the context of continuous antipsychotic drug treatment. For patients who have an established pattern of exacerbations and remission, continuous antipsychotic drug treatment is recommended for relapse prevention.

Table 5 Antipsychotic Drugs Used in the Treatment of Schizophrenia

Antipsychotic Category	*Drug (Trade Name)*
First-generation (neuroleptic)	Chlorpromazine Fluphenazine Haloperidol (Haldol) Loxapine (Loxitane) Mesoridazine (Serentil) Molindone (Moban) Perphenazine (Trilafon) Pimozide (Orap) Thioridazine Thiothixene (Navane) Aripiprazole (Abilify)
Second-generation (atypical)	Clozapine (Clozaril) Olanzapine (Zyprexa) Quetiapine (Seroquel) Risperidone (Risperdal) Ziprasidone (Geodon)

PHARMACOLOGIC TREATMENT

Antipsychotic Drugs

The antipsychotic drugs used to treat schizophrenia have a wide variety of pharmacologic properties, but all are antagonists acting at the postsynaptic D_2 dopamine receptors in the brain. Conventional, or first-generation, antipsychotic agents are often referred to as neuroleptics because of their neurologic side effects. Second-generation antipsychotic drugs also act at the dopamine receptors but are less likely to exhibit neuroleptic effects; these agents have been termed atypical antipsychotics [*see Table 5*].

Antipsychotic drug therapy is usually initiated at low doses (e.g., 1 to 5 mg of haloperidol, 5 mg of olanzapine, or 2 mg of risperidone administered once or twice a day) and titrated upward. Although the onset of therapeutic effect occurs promptly, symptom reduction is gradual, occurring over many weeks. Optimal dosing parameters have not been established for most of these drugs; an increase in dosage often results in an increase in adverse effects without a commensurate increase in therapeutic response. Blood levels have not proved effective in guiding dosing decisions; clinical response is the best guide for titrating the dose of an antipsychotic agent. The clinical effect of antipsychotics is to diminish positive psychotic symptom expression and reduce relapse rates. Although sedation may be a side effect and diminished anxiety may be a clinical effect, the primary value of these drugs is their remedial effect on positive psychotic symptoms, not their sedating or tranquilizing properties. The antipsychotic efficacy of these drugs extends beyond schizophrenia to include reducing positive psychotic symptoms associated with mental illnesses other than schizophrenia. In contrast to their effect on positive psychotic symptoms, first-generation antipsychotic agents have not been shown to be effective for either primary, enduring negative (deficit) symptoms or the cognitive impairments observed in schizophrenia.

The first atypical antipsychotic was clozapine. During the 1970s, clozapine was shown to be effective in some cases in which conventional antipsychotics did not work. However, cessation of white blood cell production occurs in approximately

1% of patients receiving clozapine; after a series of deaths from agranulocytosis in Finland during the mid-1970s, the use of clozapine declined in Europe and the drug was not marketed in the United States. Interest in clozapine was rekindled by the results of a large-scale multicenter study that yielded convincing evidence of the superior efficacy of clozapine for ameliorating positive psychotic symptoms in treatment-resistant patients with schizophrenia.[44] Consistent with the worldwide experience in the late 1970s and early 1980s, the study also showed that clozapine can be used with relative safety, provided that patients are carefully monitored for agranulocytosis. Clozapine has many pharmacologic actions and weaker affinity for the D_2 dopamine receptor; however, the basis for its superior antipsychotic action in treatment-resistant cases is not known.

The success of clozapine spawned considerable interest in the development of other antipsychotics for the treatment of schizophrenia. In the 1990s, five new antipsychotics were introduced in the United States: risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole [*see Table 5*]. It was hoped that these new drugs would have the efficacy of clozapine but without the risk of agranulocytosis and the other side effects that have limited clozapine's use. However, these new medications appear to be no more effective than the conventional antipsychotics for reducing positive psychotic symptoms. This is not surprising, given that both first- and second-generation antipsychotic drugs act as dopamine antagonists (although aripiprazole, as a partial agonist, has a different pharmacologic mechanism). These new drugs' main advantage over the conventional antipsychotics is their substantially lower incidence of extrapyramidal motor side effects. This decrease in side-effect burden may account for their apparent greater effectiveness, and it appears to have reduced the incidence of persistent dystonia and tardive dyskinesia in patients with chronic schizophrenia. Several of the second-generation antipsychotic drugs have shown greater benefit than the conventional antipsychotic agents for the treatment of depressive symptoms and the prevention of relapse and rehospitalization. These considerations have led to the second-generation antipsychotics' replacing conventional antipsychotics as first-line pharmacologic treatment for first-episode and chronic schizophrenia; this shift is controversial, however, in part because the new drugs are substantially more expensive than conventional antipsychotics. A recent study from the Department of Veterans Affairs has rekindled this debate: Rosenheck and colleagues found little difference in clinical response between patients given one of the newer drugs and those given moderate doses of haloperidol; moreover, the cost advantage of all aspects of care favored the first-generation drug.[45]

The second-generation antipsychotics are not without their limitations. Several of these agents have been associated with the development of clinically significant metabolic disturbances, including weight gain, hyperlipidemia, and new-onset type 2 diabetes mellitus. These side effects are of particular concern because patients with schizophrenia are already at increased risk for cardiovascular disease as a result of their lifestyle. Long-term studies will eventually clarify whether the decreased risk of extrapyramidal side effects, including tardive dyskinesia, warrants the increased risk of these metabolic side effects.

Do second-generation antipsychotic agents reduce the negative symptoms and cognitive impairment of schizophrenia? A number of studies have indicated that second-generation antipsychotic drugs are more effective than the first-generation drugs for negative symptoms, but the differences are usually related to concurrent changes in extrapyramidal or depressive symptoms, or excessive dosages of the first-generation drug used for comparison. In studies that have controlled for these potential sources of artifact, the apparent superior efficacy of the new-generation antipsychotics disappears. A similar story is emerging for the comparative efficacy of second-generation and conventional antipsychotics with regard to cognitive impairments. First-generation antipsychotic agents produce little improvement in cognitive function, even when they significantly relieve positive psychotic symptoms; moreover, at higher doses, they may impair cognitive function. Second-generation agents improve performance on neuropsychological measures of cognitive function, but the effect is relatively modest, and patients continue to exhibit considerable cognitive impairments compared with normal control subjects. The difference in cognitive effect between first- and second-generation antipsychotic agents is less pronounced when second-generation drugs are compared with conventional antipsychotics given at lower doses.[46]

Other Agents

The limited efficacy of conventional and atypical antipsychotic drugs against the negative symptoms and cognitive impairment of schizophrenia has led researchers to investigate drugs in other categories. Glutamatergic agents that bind to the glycine site of the NMDA glutamatergic receptor have shown potential utility for the treatment of primary, enduring negative (deficit) symptoms. Preliminary controlled clinical trials of glycine, D-cycloserine, and D-serine have produced encouraging results. In part because of the genetic linkage findings implicating the alpha-7 nicotinic receptor gene in the physiologic deficits of schizophrenia, several trials have examined the acute effects of nicotine on cognitive deficits in schizophrenia. These studies have demonstrated significant but transient reversal of impairments in sensory gating, attention, and aspects of eye tracking.[47,48] These results suggest that heavy smoking in schizophrenia may partly be self-medication on the part of the patients and have encouraged the search for nicotinic agonist drugs that have long-term efficacy. The delineation of the physiology of normal cognition has also led researchers to consider treating cognitive impairment in schizophrenia by augmenting antipsychotic therapy with drugs from other classes. Dopaminergic, adrenergic, and other pharmacologic agents are currently under study.

Augmentation strategies have also been used to treat positive psychotic symptoms that fail to respond to antipsychotic agents. However, lithium, antiepileptic drugs, antidepressants, and antianxiety agents have not been shown to substantially reduce these symptoms. It is theoretically possible that some small patient subgroups may respond to a class of drugs other than antipsychotics, but there is no way to identify such patients in advance, so it is difficult to prove or disprove this proposition. On the other hand, drugs from other classes may be effective for specific symptom targets. For example, antianxiety drugs reduce anxiety in schizophrenic patients, and drugs used for depression and mania in affective disorders are sometimes useful for these symptoms in schizophrenia.

Aggression is a special problem in schizophrenia. When aggression is the result of psychosis, antipsychotic drugs are the best treatment. Aggression may also be caused by akathisia, an intense inner restlessness and tension caused by some antipsychotic drugs; this is treated with a change in drug or the addition of another drug to counteract this side effect. Otherwise,

hostility with the threat of violence is a difficult clinical issue that is only partly addressed with concomitant administration of beta blockers such as propranolol. New-generation antipsychotic drugs appear advantageous for this aspect of the illness. Close supervision, restraints, and the use of a safe room continue to be necessary in the care of a minority of patients.

It is now common practice to add a second or third antipsychotic drug to the regimen in patients with persistent positive psychotic symptoms. There is little theoretical basis for this practice, and there is no research evidence to support this empirical strategy. The probable effects are an increase in side effects and poorer patient adherence to therapeutic recommendations.

ELECTROCONVULSIVE THERAPY

Before the introduction of antipsychotic drugs, electroconvulsive therapy (ECT) was frequently used in the treatment of patients with schizophrenia. ECT is particularly effective for catatonic stupor, excitement, and acute psychotic conditions. Results are similar to those obtained with antipsychotics—that is, positive symptoms are reduced, but long-term functional impairments are not reversed.[49] Although ECT is safe and painless, it has several limitations: it is not widely available, litigation and societal attitudes restrict its use, and any therapeutic advantage gained in an initial series of treatments is not easily maintained. Also, there is no compelling evidence that electroconvulsive treatments are effective in patients who are resistant to antipsychotic drugs. For these reasons, drug treatment is generally preferred.

PSYCHOSOCIAL INTERVENTIONS

The debate over whether schizophrenic patients should be treated with pharmacologic or psychosocial treatments has given way to the search for how these treatments should be optimally integrated. Controlled clinical trials have conclusively demonstrated that intensive psychotherapy is less effective than pharmacologic treatment; that it is not superior to cheaper, less ambitious psychosocial forms of psychotherapy; and that it should not be considered as an alternative to the use of antipsychotic drugs. On the other hand, studies have repeatedly demonstrated that supportive forms of psychosocial treatment are entirely compatible with drug treatment and can increase the effectiveness of overall treatment, reduce the amount of medication needed, enhance patient participation in the full range of treatment, and optimize social and occupational functioning. Especially impressive are studies documenting the considerable additional benefit achieved in reducing relapse and hospitalization rates when family therapy and education programs are added to maintenance pharmacologic treatment. These studies make it clear that psychosocial and rehabilitative interventions have become essential components of the comprehensive treatment of patients with schizophrenia.[50-52]

Psychosocial and rehabilitation interventions include cognitive-behavioral therapy for treatment-resistant positive psychotic symptoms; supportive, problem-solving, educationally oriented psychotherapy; family therapy and education programs aimed at helping patients and their families understand the patient's illness, reduce stress, and enhance coping capabilities; social and living skills training; supported employment programs; and the provision of supervised residential living arrangements. The development and increased utilization of psychosocial services has been complemented by the evolution of services designed to decrease the utilization of inpatient hospital services and maintain the patient in the community. Assertive community treatment teams provide intensive outreach services to patients who are unable to be maintained in the community with traditional outpatient clinical treatment. Crisis management services, including 24-hour crisis beds and partial hospitalization programs, represent alternatives to hospitalization during periods of symptom exacerbation.

The development of these services reflects the ongoing shift in the treatment of schizophrenic patients from a hospital-based to a community-based system of care. When optimal treatment with these services is provided, the rewards of therapeutic accomplishment, reduction in morbidity, and cost benefits are profound and rival therapeutic accomplishments found anywhere in medicine. The demonstrated benefits of these services challenge the field to establish an adequate community-based treatment approach that is prepared to meet the challenge and demands of broad-based integrated treatment. Society has failed to meet the challenge of providing evidence-based treatment to most persons who suffer from schizophrenia.

FUTURE TREATMENT

Intervention in the prepsychotic and early psychotic phases of schizophrenia is an unmet challenge in schizophrenia care. Early identification and treatment would clearly be preferable, but the earliest indicators of schizophrenia are usually not the psychotic features. Psychosocial treatment for nonpsychotic manifestations needs to be developed, just as pharmacologic treatments for cognitive impairments and negative psychopathology are required to address these features in patients at high risk for psychosis. It is hoped that gene discovery will lead to the development of new drugs targeting each domain of psychopathology.

The advancing knowledge of human gene and protein sequences is altering the course of schizophrenia research. These technologies promise to help delineate the molecular pathophysiology of the disease and define novel molecular targets that will provide a basis for drug development. Clinical and postmortem studies are presenting more compelling definitions of disease entities and phenotypes within the schizophrenia syndrome, which should permit the effective application of advances in genomics and proteomics.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Kirkpatrick B, Buchanan RW, Ross DE, et al: A separate disease within the syndrome of schizophrenia. Arch Gen Psychiatry 58:165, 2001

2. Kendler KS, Gallagher TJ, Abelson JM, et al: Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. The National Comorbidity Survey. Arch Gen Psychiatry 53:1022, 1996

3. Neighbors HW, Trierweiler SJ, Ford BC, et al: Racial differences in DSM diagnosis using a semi-structured instrument: the importance of clinical judgment in the diagnosis of African Americans. J Health Soc Behav 44:237, 2003

4. Jablensky A, Sartorius N, Ernberg G, et al: Schizophrenia: manifestations, incidence and course in different cultures: a World Health Organization ten-country study. Psychol Med Monogr Suppl 20:1, 1992

5. Harrison G, Mason P: Schizophrenia: falling incidence and better outcome? Br J Psychiatry 163:535, 1993

6. Bachrach LL: What we know about homelessness among mentally ill persons: an analytical review and commentary. Treating the Homeless Mentally Ill. Lamb HR, Bachrach LL, Kass FI, Eds. American Psychiatric Association, Washington, DC, 1992, p 13

7. Cantor-Graae E, Nordstrom LG, McNeil TF: Substance abuse in schizophrenia: a review of the literature and a study of correlates in Sweden. Schizophr Res Mar 48:69, 2001

8. de Leon J, Diaz FJ, Rogers T, et al: Initiation of daily smoking and nicotine dependence in schizophrenia and mood disorders. Schizophr Res 56:47, 2002

9. Kendler KS, Diehl SR: Schizophrenia: genetics. Comprehensive Textbook of Psychiatry, 6th ed. Kaplan HI, Sadock BJ, Eds. Williams & Wilkins, Baltimore, 1995, p 942

10. Karlsson H, Bachmann S, Schroder J, et al: Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci USA 98:4634, 2001

11. Messias E, Kirkpatrick B, Bromet E, et al: Summer birth and deficit schizophrenia: a pooled analysis from six countries. Arch Gen Psychiatry (in press)

12. Sherrington R, Brynjolfsson J, Petursson H, et al: Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336:164, 1988

13. Tsuang MT, Stone WS, Faraone SV: Schizophrenia: a review of genetic studies. Harv Rev Psychiatry 7:185, 1999

14. Thaker GK, Carpenter WT Jr: Advances in schizophrenia. Nat Med 7:667, 2001

15. Chowdari KV, Mirnics K, Semwal P, et al: Association and linkage analyses of RGS4 polymorphisms in schizophrenia. Hum Mol Genet 11:1373, 2002

16. Straub RE, Jiang Y, MacLean CJ, et al: Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am J Hum Genet 71:337, 2002

17. Morris DW, McGhee KA, Schwaiger S, et al: No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. Schizophr Res 60:167, 2003

18. Stefansson H, Thorgeirsson TE, Gulcher JR, et al: Neuregulin 1 in schizophrenia: out of Iceland. Mol Psychiatry 8:639, 2003

19. Chumakov I, Blumenfeld M, Guerassimenko O, et al: Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Natl Acad Sci USA 99:13675, 2002

20. Leonard S, Gault J, Hopkins J, et al: Association of promoter variants in the alpha7 nicotine acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. Arch Gen Psychiatry 12:1085, 2002

21. Gogos JA, Santha M, Takacs Z, et al: The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat Genet 21:434, 1999

22. Freedman R, Coon H, Myles-Worsley M, et al: Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA 94:587, 1997

23. Rzanny R, Klemm S, Reichenbach JR, et al: ^{31}P-MR spectroscopy in children and adolescents with a familial risk of schizophrenia. Eur Radiol 13:763, 2003

24. Keshavan MS, Stanley JA, Montrose DM, et al: Prefrontal membrane phospholipid metabolism of child and adolescent offspring at risk for schizophrenia or schizoaffective disorder: an in vivo ^{31}P MRS study. Mol Psychiatry 8:316, 2003

25. Arolt V, Lencer R, Nolte A, et al: Eye tracking dysfunction is a putative phenotypic susceptibility marker of schizophrenia and maps to a locus on chromosome 6p in families with multiple occurrence of the disease. Am J Med Genet 67:564, 1996

26. Egan MF, Goldberg TE, Kolachana BS, et al: Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci USA 98:6917, 2001

27. Davis KL, Kahn RS, Ko G, et al: Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 148:1474, 1991

28. Lipska BK, Jaskiw GE, Weinberger DR: Postpubertal emergence of hyperresponsiveness to stress and to amphetamine after neonatal excitotoxic hippocampal damage: a potential animal model of schizophrenia. Neuropsychopharmacology 9:67, 1993

29. Carlsson A, Hansson LO, Waters N, et al: Neurotransmitter aberrations in schizophrenia: new perspectives and therapeutic implications. Life Sci 61:75, 1997

30. Laruelle M, Abi-Dargham A, van Dyck CH, et al: Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Nat Acad Sci USA 93:235, 1996

31. Breier A, Su TP, Saunders R, et al: Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. Proc Natl Acad Sci USA 94:2569, 1997

32. Abi-Dargham A, Gil R, Krystal J, et al: Increased striatal dopamine transmission in schizophrenia: confirmation in a second cohort. Am J Psychiatry 155:761, 1998

33. Otani S, Auclair N, Desce JM, et al: Dopamine receptors and groups I and II mGluRs cooperate for long-term depression induction in rat prefrontal cortex through converging postsynaptic activation of MAP kinases. J Neurosci 19:9788, 1999

34. Seamans JK, Durstewitz D, Christie BR, et al: Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. Proc Natl Acad Sci USA 98:301, 2001

35. Benes FM: Emerging principles of altered neural circuitry in schizophrenia. Brain Res Rev 31:251, 2000

36. Bogerts B: The neuropathology of schizophrenic diseases: historical aspects and present knowledge. Eur Arch Psychiatry Clin Neurosci 249(suppl 4):2, 1999

37. Sanfilipo M, Lafargue T, Rusinek H, et al: Volumetric measure of the frontal and temporal lobe regions in schizophrenia: relationship to negative symptoms. Arch Gen Psychiatry 57:471, 2000

38. Selemon LD, Lidow MS, Goldman-Rakic PS: Increased volume and glial density in primate prefrontal cortex associated with chronic antipsychotic drug exposure. Biol Psychiatry 46:161, 1999

39. Glantz LA, Lewis DA: Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. Arch Gen Psychiatry 57:65, 2000

40. Mirnics K, Middleton FA, Marquez A, et al: Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 28:53, 2000

41. Buchanan RW, Breier A, Kirkpatrick B, et al: Structural abnormalities in deficit and nondeficit schizophrenia. Am J Psychiatry 150:59, 1993

42. Kemether EM, Buchsbaum MS, Byne W, et al: Magnetic resonance imaging of mediodorsal, pulvinar, and centromedian nuclei of the thalamus in patients with schizophrenia. Arch Gen Psychiatry 60:983, 2003

43. Holcomb HH, Lahti AC, Medoff DR, et al: Brain activation patterns in schizophrenic and comparison volunteers during a matched-performance auditory recognition task. Am J Psychiatry 157:1634, 2000

44. Kane J, Honigfeld G, Singer J: Clozapine for the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. Arch Gen Psychiatry 45:789, 1988

45. Rosenheck R: Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. Department of Veterans Affairs Cooperative Study Group on the Cost-Effectiveness of Olanzapine. JAMA 290:2693, 2003

46. Carpenter WT, Gold JM: Another view of therapy for cognition in schizophrenia (editorial). Biol Psychiatry 52:969, 2002

47. Adler LE, Hoffer LD, Wiser A, et al: Normalization of auditory physiology by cigarette smoking in schizophrenic patients. Am J Psychiatry 150:1856, 1993

48. Sherr JD, Myers C, Avila MT, et al: The effects of nicotine on specific eye tracking measures in schizophrenia. Biol Psychiatry 52:721, 2002

49. Tharyan P, Adams CE: Electroconvulsive therapy for schizophrenia. Cochrane Database Syst Rev (2):CD000076, 2002

50. Lehman AF, Steinwachs DM: Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) project. Schizophr Bull 24:1, 1998

51. Fenton WS, Schooler NR: Evidence-based psychosocial treatment for schizophrenia. Schizophr Bull 26:1, 2000

52. Ley A, Jeffery DP, McLaren S, Siegfried N: Treatment programmes for people with both severe mental illness and substance misuse. Cochrane Database Syst Rev (4):CD001088, 2000

53. Thaker GK, Avila MT, Hong E, et al: A model of smooth pursuit eye movement deficit associated with the schizophrenia phenotype. Psychophysiology 40:277, 2003

54. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed Text Revision. American Psychiatry Association, Washington, DC, 2000, p 312

Acknowledgment

Figure 1 courtesy of L. Elliot Hong, Malle Tagamets, and Gunvant Thaker, Maryland Psychiatric Research Center, Baltimore, Maryland. Artist: Alice Y. Chen.

VII SCHIZOPHRENIA

William T. Carpenter, Jr., m.d.
Gunvant K. Thaker, m.d.

As recently as the mid-20th century, schizophrenia was one of the most devastating psychiatric illnesses. No effective treatment was available, and most patients suffered lifelong debilitating symptoms and required institutionalization. Research during the latter part of the 20th century better defined the syndrome, provided clues to the causes, clarified the pathophysiology, provided evidence for effective treatments and against potentially harmful ones, and permitted the transfer of many schizophrenic patients from inpatient institutions to less restrictive outpatient facilities. However, challenges remain: specific causes and causal pathways are unknown, pathophysiology is ill defined, treatments are effective only for the psychotic aspect of the syndrome, and homelessness has replaced institutionalization.

Definition

Schizophrenia is a clinical syndrome probably comprising several as-yet undefined disease entities. There is substantial heterogeneity between cases, which presumably reflects multiple overlapping etiologic factors, including contributions from several genes. The presence of schizophrenia is indicated by chronic psychotic symptoms, especially hallucinations and delusions. Disorganization of thought and behavior are common and distinguish schizophrenia from the many other causes of reality-distortion symptoms (e.g., psychotomimetic drugs). Diminished emotional experience and expression, low drive, and reduced speech are observed in a subgroup of patients. Most patients have subtle impairments in cognition. Cognitive impairments and the emotional and social aspects of the disease often appear early in life; the psychotic symptoms typically begin in late adolescence or early adulthood in men and a little later in women.

Subclassification

At the end of the 19th century, Emil Kraeplin conceptualized dementia praecox as comprising seemingly discrete diseases known as hebephrenia, catatonia, and paranoia. Eugen Bleuler suggested renaming dementia praecox as *schizophrenia*, a term he coined to capture the split within thought and between thought and emotion. Bleuler also added simple schizophrenia as a fourth subtype of the disease.

Throughout most of the 20th century, the clinical presentation of schizophrenia was often described in terms of specific subtypes. These subtypes included hebephrenic, paranoid, catatonic, and simple schizophrenia. Hebephrenic schizophrenia (now termed disorganized schizophrenia) is characterized by shallow and incongruent affect and disorganized thought and behavior. Paranoid schizophrenia is characterized by predominance in males; a later age of onset; intact cognition and affect; and hallucinations and delusions, often persecutory and bizarre. Catatonic schizophrenia is characterized by extreme psychomotor manifestations, including stupor, prolonged posturing, or excitement; it must be differentiated from periodic catatonia, which is a separate syndrome unrelated to schizophrenia. For unknown reasons, catatonic schizophrenia is now rarely seen in developed countries. Simple schizophrenia denotes a milder psychosis (i.e., fewer hallucinations and delusions and less disorganization), with cases typically featuring a schizoid lifestyle of reduced emotional experience and expression and reduced social drive and engagement. A later term, schizoaffective schizophrenia, described the combination of mood disturbances—mania or depression—with schizophrenic psychotic manifestations. Other subtypes were suggested but did not gain wide acceptance.

In general, however, schizophrenia has been approached as a diagnostic class rather than in terms of these subtypes. The traditional subtypes appear not to differ substantially in etiology, pathophysiology, or treatment response. Moreover, a patient's subtype may change over time: a patient who presents with paranoid schizophrenia may later be seen to have disorganized schizophrenia; a patient who presents with schizoaffective psychosis may eventually be seen to have a disorganized or undifferentiated subtype.

Another approach to subclassification categorizes cases on the basis of three distinct domains of psychopathology: reality distortion (i.e., hallucinations and delusions), disorganization of thought and behavior, and negative symptoms such as reduced drive and affect. Kraeplin described two "maladies"—dissociation and avolition—as the central features of dementia praecox, but in the late 20th century, psychiatrists placed diagnostic emphasis on reality distortion and disorganization of thought. It turns out that many patients do not manifest negative symptoms, whereas others have chronic reduction in drive, reduced expression of emotion, and poverty of speech. We have divided schizophrenia into two subgroups: deficit schizophrenia, in which schizophrenia causes negative symptoms and psychosis; and nondeficit schizophrenia, in which negative symptoms are absent. Clinical features, family history, neuroanatomy, treatment response, course of illness, and postmortem pathology suggest that deficit schizophrenia is a separate disease entity within the schizophrenia syndrome. Deficit schizophrenia constitutes about 15% to 20% of all schizophrenia cases and appears to occur predominantly in males.[1]

Epidemiology

The annual incidence of schizophrenia approximates one in 10,000 adults younger than 60 years, with a lifetime risk of 0.8. The onset of psychosis (i.e., reality distortion or disorganization of thought) usually occurs at about 17 to 27 years of age in males and at 20 to 37 years in females. Population-based estimates suggest that the lifetime incidence of the disease is roughly equal in males and females; in clinical populations, however, males often constitute the majority of patients. Females have often been reported to have a more favorable course of illness. One explanation is that males are more likely to have deficit schizophrenia, which tends to have a chronic course with poor functional oucome.[2]

Race does not appear to influence the distribution of schizophrenia. Some United States studies have suggested a racial bias in diagnostic practice, with the term schizophrenia being used more in African-American patients and the term affective psychosis used more in white Americans.[3]

Geographically, schizophrenia appears to be evenly distributed throughout the world.[4] A country's stage of economic development, degree of urbanization, or climate seem to have little influence on schizophrenia prevalence. However, the course of the illness has been more favorable in developing countries than in developed countries. This may reflect the tendency for developing countries to have sociocentric cultures and for developed countries to have egocentric cultures. Sociocentric cultures may be more supportive of marginal function than egocentric cultures, in which demands for autonomous function may be particularly challenging for persons with schizophrenia. The difference may also reflect the inclusion of cases with a better prognosis in the syndrome definition; patients with acute onset of bizarre psychosis tend to have a more favorable course of illness.

Decreases in rates of schizophrenia were reported in the second half of the 20th century.[5] However, these decreases have not been substantiated and may simply reflect increasingly stringent criteria for the diagnosis.

Schizophrenia is the fourth leading cause of disability in adults worldwide. In the United States, about 80% of persons with schizophrenia are unemployed, and many of the remainder are underemployed. Perhaps a third of homeless persons have schizophrenia.[6]

More than half of schizophrenic patients abuse substances, especially alcohol.[7] Over 80% of schizophrenics smoke cigarettes, and nicotine intake in schizophrenics who smoke is well above that of nonschizophrenic smokers.[8] Smoking represents a substantial financial burden to this generally impoverished population and combines with other lifestyle risk factors (e.g., inadequate diet and exercise) to place schizophrenic patients at increased risk for diabetes, cardiovascular disease, and cancer.

Etiology and Genetics

Considerable data suggest that schizophrenia is a heritable illness. The risk of the disease rises with the closeness of kinship to a schizophrenic proband. Compared with the risk in the general population, the risk is fivefold higher in second-degree relatives of schizophrenia patients, 10-fold to 15-fold higher in first-degree family members and dizygotic twins, and 40-fold to 50-fold higher in monozygotic twins or in persons whose parents both have schizophrenia.[9] There is also higher risk in adopted biologic relatives of schizophrenia probands than in their adoptive families.

Other data strongly suggest that environmental factors, either on their own or interacting with genetic factors, play a critical role in causing or modifying the clinical presentation of the illness.[10] Most of the environmental factors associated with schizophrenia are active early in development, particularly during the intrauterine period. These include maternal malnutrition; infections during critical periods of fetal development; fetal hypoxia; and other birth and obstetric complications. Further supporting an environmental role is the finding that overall, patients with schizophrenia were more likely to have been born during the winter, whereas the subgroup of patients with deficit schizophrenia were more likely to have been born during the summer.[11] How season of birth affects risk is unclear. Environmental insults that occur later in life, such as use of psychoactive drugs, also play a role.

The initial search for genes that might cause schizophrenia produced disappointing results. The first significant linkage of a broadly defined schizophrenia phenotype to a specific locus, chromosome 5q11-q13, stirred excitement, but subsequent studies failed to replicate this finding.[12] Ensuing linkage studies identified several chromosomal loci, some of which were likely false positive findings, given that replications were not forthcoming. Significant linkage findings that were subsequently replicated by other studies implicated regions on chromosomes 1, 5, 6, 8, 13, 15, and 22.[13,14] Most of these findings were not robust, however; in addition, some studies failed to identify linkages at the same locus, and many replicating studies only suggested linkage (i.e., log odds ratio scores ranged from 2 to 3). Further investigation of these loci has identified several candidate genes: *RGS4,* the dysbindin gene (*DTNBP1*), neuregulin 1, the alpha-7 nicotinic receptor gene, *COMT, G72,* and *GAD 65/67*.[15-21] Further work is needed to identify other candidate genes using other genetic approaches (see below) and to confirm the role of candidate genes in the etiology, pathophysiology, and treatment of schizophrenia.

The search for genes in schizophrenia has proved particularly challenging because of the complex and multifactorial nature of the disorder. It appears that many genes may cause vulnerability to schizophrenia, each with a modest effect, and that these genes probably interact with one another and with environmental factors. Thus, the clinical phenotype of schizophrenia may well be the common expression of a variety of genotypes; this may explain failures to replicate identified disease loci.

Arguing that schizophrenia encompasses multiple phenotypes of different genetic vulnerability profiles, several investigators have focused on a search for the genes underlying specific deficits seen in schizophrenia. Freedman and colleagues, using an electrophysiologic measure of sensory gating, demonstrated that a deficit in sensory gating in familes with schizophrenia is a result of an autosomal dominant effect of a single gene, the alpha-7 nicotinic receptor gene, on chromosome 15q14.[22] Other phenotypes that are associated with schizophrenia and that have been used in the search for schizophrenia-related genes are smooth-pursuit eye movement deficit; cognitive impairments in dimensions of attention, language, and memory; deficits in levels of the neuronal marker *N*-acetyl-aspartate in the hippocampal region; and abnormal membrane phospholipid metabolism in the prefrontal cortex, as measured by phosphorus-31 magnetic resonance spectroscopy.[23,24] Preliminary findings have suggested linkage of the smooth-pursuit phenotype to a locus on chromosome 6p and an association of prefrontal executive functional deficit with the *val* allele of the *COMT* gene.[25,26]

Pathophysiology

The pathophysiology of schizophrenia is not well understood. That schizophrenia results from an excess of dopamine has been the dominant hypothesis for decades, with several modifications and expansion of the hypothesis made during the 1990s.[27-29] Brain imaging studies have shown evidence of increased synaptic dopamine release in response to amphetamine administration in the striatum of schizophrenic patients.[30-32] Excess dopamine may explain the positive symptoms but not the negative symptoms or other aspects of the disease. The pathophysiologic mechanisms are presumably complex—even a deficiency of dopamine is a viable explanation for negative symptoms. The clinical observations that long-term phencyclidine (PCP) use produces a broader cluster of schizophrenia-like symptoms than seen with use of dopamine agonists has led to

speculation that a lack of glutamine may be involved. In contrast to the predominantly paranoid symptoms associated with long-term use of dopamine agonists, long-term PCP use produces negative symptoms and cognitive impairment in addition to reality distortion.[28] These schizophrenia-like symptoms from PCP are hypothetically initiated by blockade of the ion channel in the *N*-methyl-D-aspartate (NMDA) receptor complex, resulting in diminished glutamatergic neurotransmission. Both dopaminergic and glutamatergic terminals converge on the spines of pyramidal neurons in cortex. This suggests a shared mechanism by which glutamate and dopamine play a role in schizophrenia.[28,29] Although the two neurotransmitter systems share complex presynaptic and postsynaptic interactions,[33,34] modulation of glutamate release by D_1 dopamine receptors may mediate some of the effects of dopamine on psychosis.

Postmortem findings in schizophrenia have not been consistent. Furthermore, interpretation of postmortem findings is difficult because of the confounding effects of long-term drug treatment received by most patients. Massive cell loss or gliosis is generally not found, thus ruling out degenerative changes in the brain. Earlier postmortem histopathologic examination of the brains of patients with schizophrenia showed slight reductions in neocortical gray matter volume, decreased neuronal size, and neuronal disarray in several corticolimbic structures.[35-37] Subsequent investigations have focused on changes of neuronal connectivity or microcircuitry within the cortical layers and have noted an increase in cell packing density without a change in the number of neurons. These findings suggest a decrease in neuropil and a reduction in dendritic spine density on pyramidal neurons in the prefrontal cortex.[38,39] Using complementary DNA microarrays, Mirnics and colleagues noted a decrease in the expression of several genes involved in glutamate and γ-aminobutyric acid (GABA) transmission and in the regulation of presynaptic function and signal termination in the prefrontal cortex of patients with schizophrenia.[40] Together, these findings indicate a decrease in cortical or thalamic excitatory synaptic inputs to the pyramidal neurons.

In addition to drug probe and postmortem studies exploring the pathophysiology of schizophrenia, in vivo structural imaging studies of the brain show cortical volume reductions of one or more of the constituent structures of the medial temporal lobe (comprising the hippocampus, the amygdala, and the parahippocampal gyrus). Some studies note smaller prefrontal cortical volume, but this finding is not consistent across studies and may be limited to a subgroup of schizophrenic patients.[41] Reductions in thalamic volume, particularly of the thalamic nuclei forming part of the temporal-thalamic and prefrontal-thalamic circuitry, are also noted.[42] Functional imaging studies of neuronal circuitry during performance of cognitive tasks have found reduced neuronal activation in several cortical regions of schizophrenic patients compared with normal control subjects. However, interpretation of these data is somewhat confounded by the fact that schizophrenic patients perform poorly in such tasks. Several studies have examined brain activation during the performance of tasks that schizophrenic patients performed as well as the control subjects. These studies show that to perform equally well on a working memory task, schizophrenic patients show increased activation of the prefrontal cortical network. On other tasks, patients activated different cortical networks than were activated by healthy control subjects, suggesting that the patients used a different strategy or required compensatory mechanisms [*see Figure 1*].[43]

Clinical Course

Persons with schizophrenia often come to medical attention because of a public manifestation of bizarre behavior, disorganized thought, and reality-distortion symptoms. However, the history taken at first presentation often reveals a prolonged period of more subtle manifestations of the psychotic state, averaging perhaps 2 years; cognitive impairments and negative symptoms may have been present for years, if not throughout life. Although the course of schizophrenia varies considerably between individuals, a brief case description will illustrate several characteristic features of the disease.

A young man had a declining academic performance during the last 2 years of high school. He seemed to lose friends and motivation and gradually became a loner with strange ideas. He relinquished plans to attend college, and after high-school graduation he took a job. Soon after he began working, he acted in a bizarre and hostile manner in a confrontation; recurrence of this behavior caused dismissal from the job. He spent the next year working odd jobs, living at home, and displaying a growing indifference to his life goals. His parents suspected substance abuse and were concerned that he was not interested in having a girlfriend. On his 20th birthday, he was found sitting nude in the backyard; he claimed that he was receiving messages through solar radiation. He was taken to the emergency department.

As this case illustrates, the opportunity for intervention often occurs before the conspicuous onset of illness. Although the development of overt psychotic symptoms defines the onset of the illness, subtle impairments occur early in some patients and persist through the peaks and troughs of the psychotic symptoms. Some schizophrenic patients appear to be born with subtle motor and emotional impairments. Early in childhood and adolescence, many persons who later develop schizophrenia manifest a range of cognitive impairments on psychological tests of attention, memory, and executive function and on physiologic tests of information processing.

There is a variable course of illness after the initial psychotic break. Later in life, some patients even show clinical improvement in psychotic symptoms. In contrast, a smaller subgroup may show rapid cognitive decline. The manifestation of psychosis only modestly defines long-term morbidity, which instead is best understood in terms of functional capabilities, social and occupational roles, and quality of life. Cognitive impairments and negative symptoms are more robust determinants of these outcomes.

Diagnosis

CLINICAL MANIFESTATIONS

Diagnosis of schizophrenia is based on clinical manifestations and the clinical course. Internationally accepted criteria have been published in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders—Text Revision* [*see Table 1*] and in the 10th edition of the *International Classification of Diseases and Related Health Problems*.

The diagnosis of schizophrenia should be considered in a patient who presents with hallucinations and delusions. The presence of disorganized thought and behavior increases the likelihood of schizophrenia. In the absence of other known causes of such symptoms (e.g., substance abuse, temporal lobe epilepsy), the principal diagnostic task is to discriminate between severe

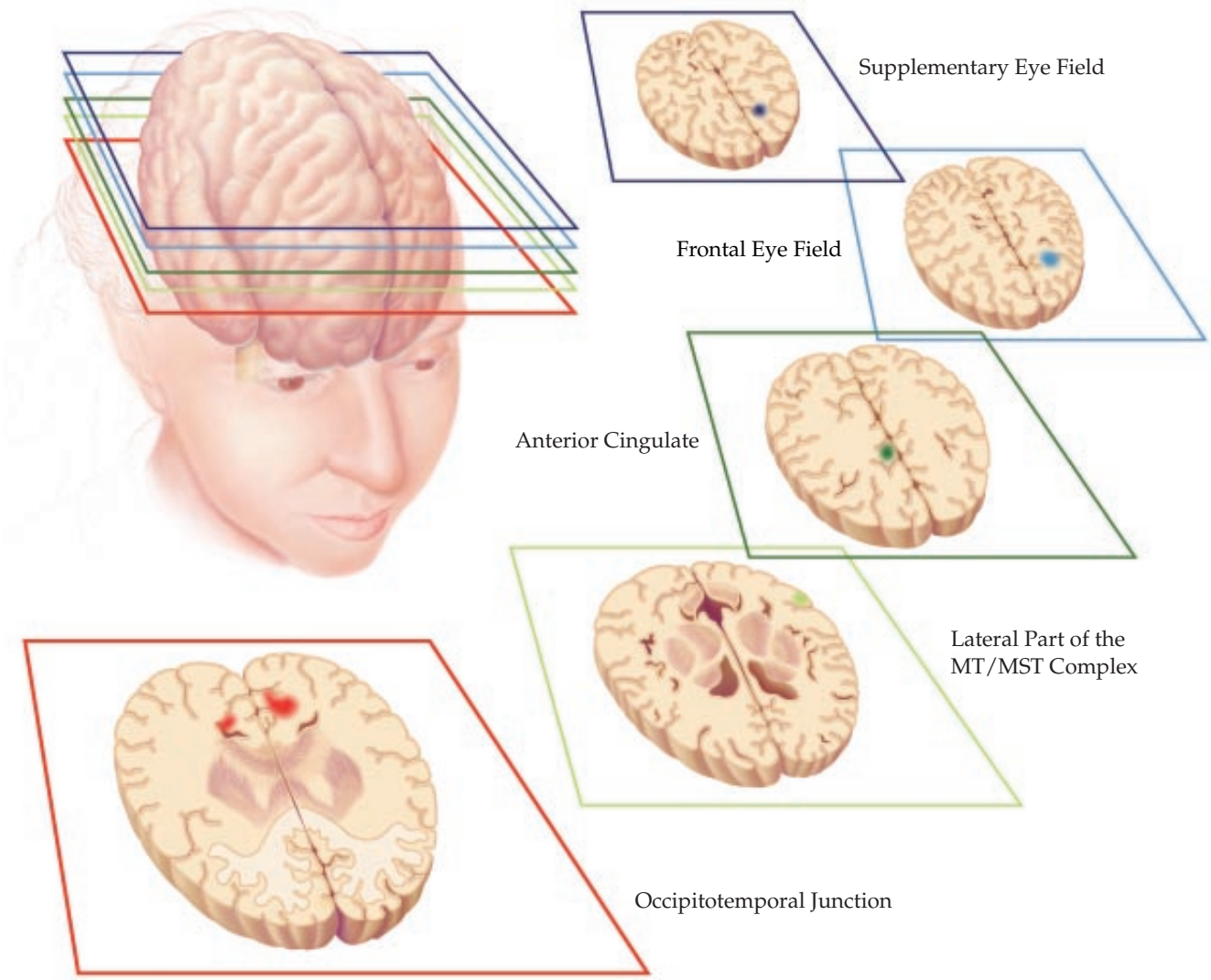


Figure 1 **In a functional magnetic resonance imaging (fMRI) study, schizophrenic patients and healthy comparison subjects were matched on the basis of how well they were able to keep their eyes on a moving target (so-called smooth-pursuit eye movement). Earlier data suggested that persons at risk for schizophrenia use a strategy different from that of healthy control subjects to maintain normal smooth pursuit. Whereas healthy persons depend more on the internal representation of target velocity (so-called extraretinal motion signals) and less on the immediate visual feedback from the retina (retinal motion signals), persons at risk for schizophrenia (i.e., relatives of schizophrenic patients) depended more on the retinal signals.[53] In schizophrenic patients who were able to maintain smooth pursuit, as well as in healthy control subjects, the fMRI analyses showed increased activation in some regions and decreased activation in others. A comparison of related brain activation in control subjects and relatives of schizophrenic patients showed that the latter had reduced activities in left frontal eye field, left supplementary eye field, anterior cingulate, and left medial superior temporal lobe; many of these areas have previously been identified as sites of extraretinal motion processing in nonhuman primates. In addition, relatives of schizophrenic patients showed increased activation in the occipitotemporal junction, an area at the medial part of the middle temporal (MT) lobe/medial superior temporal (MST) lobe that is commonly associated with retinal motion information processing.**

mental illnesses. If the patient does not have a mood disorder and the psychotic symptoms are accompanied by restricted affect, low drive, and speech disturbance [*see Table 2*], the probability of schizophrenia is high. The diagnosis can be made with greatest confidence, however, on the basis of the longitudinal pattern of the disorder, which includes the occurrence of prodromal symptoms before the initial episode, residual symptoms between psychotic episodes, and psychotic episodes that cannot be attributed to mood disturbance (e.g., manic or depressive psychosis) or other known causes of psychotic behavior [*see* Differential Diagnosis, *below*].

PHYSICAL EXAMINATION

Physical examination is unlikely to provide evidence for the diagnosis of schizophrenia but may help rule out rare cases of psychosis caused by physical ailments. Evidence of self-mutilation can alert the physician to potential risks and affect the choice of inpatient or outpatient treatment.

Although physical findings are not very informative for diagnosis, they are critical in guiding selection of drugs for treatment. About 12% to 15% of schizophrenic patients manifest motor abnormalities before antipsychotic drug treatment is initiated. The finding of involuntary movements during physical

examination should direct the clinician to select a drug with a low potential for causing extrapyramidal symptoms or tardive dyskinesia. The presence of obesity dictates avoidance of drugs that increase weight.

LABORATORY TESTS

Toxicology screens are important in the workup of schizophrenia. Positive results may provide an alternative explanation for psychosis or may confirm comorbid substance abuse. Lipid profiles may identify patients at risk for hyperlipidemia associated with the use of some antipsychotic drugs, and a metabolic profile may identify patients at risk for diabetes. An electrocardiogram is relevant as a baseline measure for patients who are subsequently prescribed antipsychotic medications that can prolong the QT interval. Electroencephalography may be used to exclude the diagnosis of temporal lobe epilepsy. Rarely, unusual presentations (e.g., the emergence of symptoms of schizophrenia after head trauma) may require magnetic resonance imaging of the brain to rule out other causes of psychosis.

There are a number of physiologic and neuroimaging findings that are helpful in distinguishing groups of schizophrenic patients from groups of healthy control subjects. However, these findings may not be unique to patients with schizophrenia, and their accuracy is inadequate for diagnostic purposes. Such findings include impairment of sensory gating as measured by evoked-potential response of P50 and by oculomotor physiologic dysfunction and prefrontal cortical dysfunction as assessed by brain imaging studies. Differences found on testing groups of schizophrenic patients have not led to the development of tests with adequate sensitivity and specificity and positive or negative predictive power for diagnosis in individual cases.

Table 1 Diagnostic Criteria for Schizophrenia[54]

All six criteria must be met for a patient to fit the diagnosis of schizophrenia.

1. *Presence of psychosis for a significant length of time*
   Any two or more of the following symptoms must be present for a significant duration during a 1-month period (less if treated):
   1. Delusions
   2. Hallucinations
   3. Disorganized speech
   4. Negative symptoms (e.g., lack of emotion and social drive and engagement)
   5. Grossly disorganized behavior
2. *Deterioration in social/occupational functioning*
   Functioning in at least one of the following areas is markedly below the level achieved before the illness:
   1. Work
   2. Interpersonal relationships
   3. Self-care
3. *Continuous signs of the disturbance for 6 months*
   The 6 months must include at least 1 month of positive symptoms. However, during prodromal or residual periods, the disturbance may be manifested by negative symptoms only.
4. *Exclusion of schizoaffective and mood disorders*
   No major depressive, manic, or mixed episodes have occurred concurrently with the active-phase symptoms; if such episodes have occurred, their total duration has been much briefer than the periods of active and residual psychosis.
5. *Exclusion of substance-induced or medical disorders*
   The disturbance is not the result of direct physiologic effects of a substance or caused by a medical condition.
6. *Relationship to pervasive developmental disorders*
   If the patient has a history of autistic disorder or another pervasive developmental disorder, a diagnosis of schizophrenia is made only if prominent delusions or hallucinations are also present for at least 1 month.

Table 2 Cognitive Disturbances Seen in Schizophrenic Speech

Abstract Thought Impairment
For example, when asked to interpret a proverb, the patient may provide a concrete description

Blocking
A halt in the flow of speech; the person is unable to continue the train of thought even with cueing

Clang Associations
Speech connections based on similarities in the sounds of words rather than on their meanings

Delayed Response
Long latency before responding to questions; at the extreme, the patient may be mute

Poverty of Content
The speech conveys little information even when it is more or less coherent

Tangentiality
Response to questions such that the answers become progressively less related to the original question

Loosening of Associations
Speech consisting of ideas that have no discernible connection to one another

Word Salad
Incoherent speech from extreme loosening of associations

Genetic Testing

Rapidly advancing knowledge concerning genotypic vulnerability to schizophrenia may lead to better susceptibility testing. However, it appears unlikely that genotypes unique to schizophrenia will be defined. At present, genetic information mainly suggests that a number of genes that individually make minor contributions to vulnerability may combine in persons with schizophrenia and may interact with certain environmental etiologic factors. It is also suggested that genes conferring vulnerability to schizophrenia may overlap with those that predispose to depression and bipolar disorder. If these suggestions prove true, specific genetics-based diagnoses of schizophrenia are unlikely to be forthcoming in the near future. However, genetic testing may eventually help guide therapy in individual patients by indicating their likelihood of response and risk of side effects with specific drugs.

Differential Diagnosis

Because schizophrenia is defined as a psychotic illness with functional impairments, distinguishing schizophrenia from normalcy is usually not difficult. However, because the clinician depends on information provided by patients regarding their inner experiences, recognition of mild psychosis is difficult in patients who are guarded and who are not forthcoming with information.

Schizotypal personality disorder shares some of the clinical characteristics of schizophrenia, such as social and physical an-

hedonia, suspiciousness, magical thinking, blunting of affect and emotional experience, and poor functioning. However, schizotypal patients do not experience overt and persistent psychotic symptoms, although rare and brief psychotic symptoms may occur.

In a patient with persistent psychosis, the differential diagnosis consists mainly of affective disorders with psychosis, substance abuse, and delusional disorders. Psychosis that coincides with depression is typically associated with such affective features as delusions of poverty or accusatory voices. Similarly, delusions of grandeur are common during manic episodes. Psychotic symptoms in affective disorders typically follow the emergence of depression or mania and fade once the affective symptoms recede. Depressive symptoms often occur in schizophrenia, particularly in the first episode. Alternatively, excitement and hyperactivity in patients with schizophrenia may mimic manic symptoms. Differential diagnosis depends on the timing of psychosis in relation to the occurrence of affective symptoms and the content of delusions and hallucinations. Psychotic symptoms can be caused by many medical conditions [*see Table 3*], as well as by alcohol, illicit drugs, and over-the-counter and prescription medications [*see Table 4*]. The history and toxicology screen can rule out psychosis caused by drug abuse, such as use of PCP or the long-term abuse of steroids. Delusional disorder is diagnosed on the basis of persistent and circumscribed delusions (more realistic and less bizarre than those associated with schizophrenia) in the absence of disorganization and negative psychopathology.

Occasionally, neurologic conditions such as brain tumor or temporal lobe epilepsy may be misdiagnosed as schizophrenia. When such conditions are suspected, MRI and EEG can help with the diagnosis.

Table 3 Drugs and Medications Commonly Associated with Psychotic Symptoms

Abused Drugs
- Alcohol (withdrawal and alcoholic hallucinosis)
- Amphetamines
- Cocaine (especially free-base or crack)
- Marijuana (panic reaction)
- Phencyclidine (PCP) and hallucinogens (acute intoxication and flashbacks)
- Sedative-hypnotics (withdrawal)

Analgesics
- Meperidine (toxic metabolite)
- Nonsteroidal anti-inflammatory drugs
- Pentazocine and other opiate mixed agonist-antagonists

Antibiotics
- Antituberculous drugs (e.g., cycloserine, isoniazid, rifampin)
- Many others

Anticholinergics
- Anticholinergic antiparkinsonians
- Antihistamines (e.g., diphenhydramine)
- Atropine and derivatives
- Cyclic antidepressants
- Low-potency antipsychotic drugs (e.g., thioridazine, clozapine)
- Meclizine
- Scopolamine

Cardiovascular Drugs
- Antiarrhythmics (e.g., amiodarone, digitalis, procainamide)
- Captopril

Dopamine Agonists
- Amantadine
- D_2 dopamine receptor agonists (e.g., bromocriptine, pergolide)
- Levodopa

Endocrine Drugs
- Estrogens, testosterone
- Glucocorticoids and adrenocorticotropic hormone (ACTH)
- Thyroid hormone (overdose)

H_2 Receptor Antagonists
- Cimetidine, ranitidine

Stimulants and Sympathomimetics
- Dextroamphetamine
- Methylphenidate
- Over-the-counter decongestants (e.g., pseudoephedrine), diet pills, and pep pills

Table 4 Medical Conditions Associated with Psychotic Symptoms

Cancer and Hematologic Malignancies
- CNS neoplasm
- Hyperviscosity syndromes (resulting from hematologic malignancy)
- Paraneoplastic syndromes (e.g., limbic encephalitis)

Cardiovascular
- Anoxia and postanoxic encephalopathy
- Hypertensive encephalopathy

Infections and Sequelae
- AIDS encephalopathy
- Encephalitis, meningitis, brain abscess
- Lyme disease
- Neurosyphilis

Metabolic and Endocrine Disorders
- Acute intermittent porphyria
- Addison disease
- Cushing disease
- Hepatic encephalopathy
- Hypocalcemia and hypercalcemia
- Hypoglycemia
- Hypothyroidism and hyperthyroidism

Neurologic Disorders
- Alzheimer disease
- Complex partial seizures
- Huntington chorea (Huntington disease)
- Multiple sclerosis
- Stroke
- Wilson disease

Nutritional Deficiencies
- Folic acid deficiency
- Pellagra (niacin deficiency)
- Vitamin B_{12} deficiency

Rheumatologic Disorders
- Lupus cerebritis

Management

Patients with schizophrenia require acute treatment for psychotic exacerbations; long-term treatment, involving mainte-

nance and prophylactic strategies to sustain symptom improvement and prevent recurrence of psychosis; and support for living. Unfortunately, the avolitional (negative) symptoms and impairments in cognition, which account for poor functional outcomes, show little response to current treatments. The future may offer treatment for these nonpsychotic aspects of the illness, as well as preventive measures for persons at high risk for schizophrenia.

ACUTE TREATMENT

The first-line therapeutic approach for an acute exacerbation of schizophrenia—whether it is the first episode or a relapse—involves antipsychotic medication (see below). To determine whether the patient requires hospitalization, the physician must evaluate patient safety and cooperation. This evaluation is clinically based and often includes discussion with family members or other informants. The patient's ability to care for himself or herself, the setting most likely to ensure adherence to therapy, and the risk of suicidal or aggressive behavior are at the forefront of this evaluation.

MAINTENANCE THERAPY

Most patients show at least a partial therapeutic response to the initiation of antipsychotic medication. Longer-term care involves maintenance therapy to minimize the expression of psychosis and prevent or delay psychotic exacerbation. Some patients will have a full remission of psychotic symptoms, but even in such cases, the continuous administration of antipsychotic drugs will delay recurrence of a psychotic exacerbation. Psychosocial treatments and pharmacologic treatments have proved effective in reducing exacerbation rates over the long term. The clinical relationship involved in these treatments also provides for ongoing evaluation of life needs, suicide risk, medication adherence, and success in coping with the many practical problems facing individuals with a chronic debilitating disease. Many patients need assistance in obtaining shelter, in developing an occupational niche or receiving vocational rehabilitation, and in participating in substance abuse programs.

Involvement of the people in the patient's life is critically important. Family members and significant others need to be educated about the disease, its treatment, and the approach to changes in clinical status.

If the patient recovers from psychosis, remains free of psychotic symptoms, and is stable for several years, the gradual discontinuance of medication may be warranted, provided that close clinical observation can be ensured. However, most patients will experience subsequent exacerbations. Early detection and intervention may be effective. Because a few patients have only a single episode of psychosis and may not require continuous antipsychotic treatment and because the risk of medication side effects is substantial, a medication-free trial is justified in selected cases.

Pharmacologic and psychosocial management need to be integrated. The two approaches appear to be synergistic in their effects, and each is crucial to the effectiveness of the other. Early detection of exacerbation is most likely when continuous clinical observation is provided in the context of psychosocial treatment; successful participation in skills training, vocational rehabilitation, living arrangements, and substance abuse treatment is most likely to be successful in the context of continuous antipsychotic drug treatment. For patients who have an established pattern of exacerbations and remission, continuous antipsychotic drug treatment is recommended for relapse prevention.

Table 5 Antipsychotic Drugs Used in the Treatment of Schizophrenia

Antipsychotic Category	*Drug (Trade Name)*
First-generation (neuroleptic)	Chlorpromazine Fluphenazine Haloperidol (Haldol) Loxapine (Loxitane) Mesoridazine (Serentil) Molindone (Moban) Perphenazine (Trilafon) Pimozide (Orap) Thioridazine Thiothixene (Navane) Aripiprazole (Abilify)
Second-generation (atypical)	Clozapine (Clozaril) Olanzapine (Zyprexa) Quetiapine (Seroquel) Risperidone (Risperdal) Ziprasidone (Geodon)

PHARMACOLOGIC TREATMENT

Antipsychotic Drugs

The antipsychotic drugs used to treat schizophrenia have a wide variety of pharmacologic properties, but all are antagonists acting at the postsynaptic D_2 dopamine receptors in the brain. Conventional, or first-generation, antipsychotic agents are often referred to as neuroleptics because of their neurologic side effects. Second-generation antipsychotic drugs also act at the dopamine receptors but are less likely to exhibit neuroleptic effects; these agents have been termed atypical antipsychotics [*see Table 5*].

Antipsychotic drug therapy is usually initiated at low doses (e.g., 1 to 5 mg of haloperidol, 5 mg of olanzapine, or 2 mg of risperidone administered once or twice a day) and titrated upward. Although the onset of therapeutic effect occurs promptly, symptom reduction is gradual, occurring over many weeks. Optimal dosing parameters have not been established for most of these drugs; an increase in dosage often results in an increase in adverse effects without a commensurate increase in therapeutic response. Blood levels have not proved effective in guiding dosing decisions; clinical response is the best guide for titrating the dose of an antipsychotic agent. The clinical effect of antipsychotics is to diminish positive psychotic symptom expression and reduce relapse rates. Although sedation may be a side effect and diminished anxiety may be a clinical effect, the primary value of these drugs is their remedial effect on positive psychotic symptoms, not their sedating or tranquilizing properties. The antipsychotic efficacy of these drugs extends beyond schizophrenia to include reducing positive psychotic symptoms associated with mental illnesses other than schizophrenia. In contrast to their effect on positive psychotic symptoms, first-generation antipsychotic agents have not been shown to be effective for either primary, enduring negative (deficit) symptoms or the cognitive impairments observed in schizophrenia.

The first atypical antipsychotic was clozapine. During the 1970s, clozapine was shown to be effective in some cases in which conventional antipsychotics did not work. However, cessation of white blood cell production occurs in approximately

1% of patients receiving clozapine; after a series of deaths from agranulocytosis in Finland during the mid-1970s, the use of clozapine declined in Europe and the drug was not marketed in the United States. Interest in clozapine was rekindled by the results of a large-scale multicenter study that yielded convincing evidence of the superior efficacy of clozapine for ameliorating positive psychotic symptoms in treatment-resistant patients with schizophrenia.[44] Consistent with the worldwide experience in the late 1970s and early 1980s, the study also showed that clozapine can be used with relative safety, provided that patients are carefully monitored for agranulocytosis. Clozapine has many pharmacologic actions and weaker affinity for the D_2 dopamine receptor; however, the basis for its superior antipsychotic action in treatment-resistant cases is not known.

The success of clozapine spawned considerable interest in the development of other antipsychotics for the treatment of schizophrenia. In the 1990s, five new antipsychotics were introduced in the United States: risperidone, olanzapine, quetiapine, ziprasidone, and aripiprazole [*see Table 5*]. It was hoped that these new drugs would have the efficacy of clozapine but without the risk of agranulocytosis and the other side effects that have limited clozapine's use. However, these new medications appear to be no more effective than the conventional antipsychotics for reducing positive psychotic symptoms. This is not surprising, given that both first- and second-generation antipsychotic drugs act as dopamine antagonists (although aripiprazole, as a partial agonist, has a different pharmacologic mechanism). These new drugs' main advantage over the conventional antipsychotics is their substantially lower incidence of extrapyramidal motor side effects. This decrease in side-effect burden may account for their apparent greater effectiveness, and it appears to have reduced the incidence of persistent dystonia and tardive dyskinesia in patients with chronic schizophrenia. Several of the second-generation antipsychotic drugs have shown greater benefit than the conventional antipsychotic agents for the treatment of depressive symptoms and the prevention of relapse and rehospitalization. These considerations have led to the second-generation antipsychotics' replacing conventional antipsychotics as first-line pharmacologic treatment for first-episode and chronic schizophrenia; this shift is controversial, however, in part because the new drugs are substantially more expensive than conventional antipsychotics. A recent study from the Department of Veterans Affairs has rekindled this debate: Rosenheck and colleagues found little difference in clinical response between patients given one of the newer drugs and those given moderate doses of haloperidol; moreover, the cost advantage of all aspects of care favored the first-generation drug.[45]

The second-generation antipsychotics are not without their limitations. Several of these agents have been associated with the development of clinically significant metabolic disturbances, including weight gain, hyperlipidemia, and new-onset type 2 diabetes mellitus. These side effects are of particular concern because patients with schizophrenia are already at increased risk for cardiovascular disease as a result of their lifestyle. Long-term studies will eventually clarify whether the decreased risk of extrapyramidal side effects, including tardive dyskinesia, warrants the increased risk of these metabolic side effects.

Do second-generation antipsychotic agents reduce the negative symptoms and cognitive impairment of schizophrenia? A number of studies have indicated that second-generation antipsychotic drugs are more effective than the first-generation drugs for negative symptoms, but the differences are usually related to concurrent changes in extrapyramidal or depressive symptoms, or excessive dosages of the first-generation drug used for comparison. In studies that have controlled for these potential sources of artifact, the apparent superior efficacy of the new-generation antipsychotics disappears. A similar story is emerging for the comparative efficacy of second-generation and conventional antipsychotics with regard to cognitive impairments. First-generation antipsychotic agents produce little improvement in cognitive function, even when they significantly relieve positive psychotic symptoms; moreover, at higher doses, they may impair cognitive function. Second-generation agents improve performance on neuropsychological measures of cognitive function, but the effect is relatively modest, and patients continue to exhibit considerable cognitive impairments compared with normal control subjects. The difference in cognitive effect between first- and second-generation antipsychotic agents is less pronounced when second-generation drugs are compared with conventional antipsychotics given at lower doses.[46]

Other Agents

The limited efficacy of conventional and atypical antipsychotic drugs against the negative symptoms and cognitive impairment of schizophrenia has led researchers to investigate drugs in other categories. Glutamatergic agents that bind to the glycine site of the NMDA glutamatergic receptor have shown potential utility for the treatment of primary, enduring negative (deficit) symptoms. Preliminary controlled clinical trials of glycine, D-cycloserine, and D-serine have produced encouraging results. In part because of the genetic linkage findings implicating the alpha-7 nicotinic receptor gene in the physiologic deficits of schizophrenia, several trials have examined the acute effects of nicotine on cognitive deficits in schizophrenia. These studies have demonstrated significant but transient reversal of impairments in sensory gating, attention, and aspects of eye tracking.[47,48] These results suggest that heavy smoking in schizophrenia may partly be self-medication on the part of the patients and have encouraged the search for nicotinic agonist drugs that have long-term efficacy. The delineation of the physiology of normal cognition has also led researchers to consider treating cognitive impairment in schizophrenia by augmenting antipsychotic therapy with drugs from other classes. Dopaminergic, adrenergic, and other pharmacologic agents are currently under study.

Augmentation strategies have also been used to treat positive psychotic symptoms that fail to respond to antipsychotic agents. However, lithium, antiepileptic drugs, antidepressants, and antianxiety agents have not been shown to substantially reduce these symptoms. It is theoretically possible that some small patient subgroups may respond to a class of drugs other than antipsychotics, but there is no way to identify such patients in advance, so it is difficult to prove or disprove this proposition. On the other hand, drugs from other classes may be effective for specific symptom targets. For example, antianxiety drugs reduce anxiety in schizophrenic patients, and drugs used for depression and mania in affective disorders are sometimes useful for these symptoms in schizophrenia.

Aggression is a special problem in schizophrenia. When aggression is the result of psychosis, antipsychotic drugs are the best treatment. Aggression may also be caused by akathisia, an intense inner restlessness and tension caused by some antipsychotic drugs; this is treated with a change in drug or the addition of another drug to counteract this side effect. Otherwise,

hostility with the threat of violence is a difficult clinical issue that is only partly addressed with concomitant administration of beta blockers such as propranolol. New-generation antipsychotic drugs appear advantageous for this aspect of the illness. Close supervision, restraints, and the use of a safe room continue to be necessary in the care of a minority of patients.

It is now common practice to add a second or third antipsychotic drug to the regimen in patients with persistent positive psychotic symptoms. There is little theoretical basis for this practice, and there is no research evidence to support this empirical strategy. The probable effects are an increase in side effects and poorer patient adherence to therapeutic recommendations.

ELECTROCONVULSIVE THERAPY

Before the introduction of antipsychotic drugs, electroconvulsive therapy (ECT) was frequently used in the treatment of patients with schizophrenia. ECT is particularly effective for catatonic stupor, excitement, and acute psychotic conditions. Results are similar to those obtained with antipsychotics—that is, positive symptoms are reduced, but long-term functional impairments are not reversed.[49] Although ECT is safe and painless, it has several limitations: it is not widely available, litigation and societal attitudes restrict its use, and any therapeutic advantage gained in an initial series of treatments is not easily maintained. Also, there is no compelling evidence that electroconvulsive treatments are effective in patients who are resistant to antipsychotic drugs. For these reasons, drug treatment is generally preferred.

PSYCHOSOCIAL INTERVENTIONS

The debate over whether schizophrenic patients should be treated with pharmacologic or psychosocial treatments has given way to the search for how these treatments should be optimally integrated. Controlled clinical trials have conclusively demonstrated that intensive psychotherapy is less effective than pharmacologic treatment; that it is not superior to cheaper, less ambitious psychosocial forms of psychotherapy; and that it should not be considered as an alternative to the use of antipsychotic drugs. On the other hand, studies have repeatedly demonstrated that supportive forms of psychosocial treatment are entirely compatible with drug treatment and can increase the effectiveness of overall treatment, reduce the amount of medication needed, enhance patient participation in the full range of treatment, and optimize social and occupational functioning. Especially impressive are studies documenting the considerable additional benefit achieved in reducing relapse and hospitalization rates when family therapy and education programs are added to maintenance pharmacologic treatment. These studies make it clear that psychosocial and rehabilitative interventions have become essential components of the comprehensive treatment of patients with schizophrenia.[50-52]

Psychosocial and rehabilitation interventions include cognitive-behavioral therapy for treatment-resistant positive psychotic symptoms; supportive, problem-solving, educationally oriented psychotherapy; family therapy and education programs aimed at helping patients and their families understand the patient's illness, reduce stress, and enhance coping capabilities; social and living skills training; supported employment programs; and the provision of supervised residential living arrangements. The development and increased utilization of psychosocial services has been complemented by the evolution of services designed to decrease the utilization of inpatient hospital services and maintain the patient in the community. Assertive community treatment teams provide intensive outreach services to patients who are unable to be maintained in the community with traditional outpatient clinical treatment. Crisis management services, including 24-hour crisis beds and partial hospitalization programs, represent alternatives to hospitalization during periods of symptom exacerbation.

The development of these services reflects the ongoing shift in the treatment of schizophrenic patients from a hospital-based to a community-based system of care. When optimal treatment with these services is provided, the rewards of therapeutic accomplishment, reduction in morbidity, and cost benefits are profound and rival therapeutic accomplishments found anywhere in medicine. The demonstrated benefits of these services challenge the field to establish an adequate community-based treatment approach that is prepared to meet the challenge and demands of broad-based integrated treatment. Society has failed to meet the challenge of providing evidence-based treatment to most persons who suffer from schizophrenia.

FUTURE TREATMENT

Intervention in the prepsychotic and early psychotic phases of schizophrenia is an unmet challenge in schizophrenia care. Early identification and treatment would clearly be preferable, but the earliest indicators of schizophrenia are usually not the psychotic features. Psychosocial treatment for nonpsychotic manifestations needs to be developed, just as pharmacologic treatments for cognitive impairments and negative psychopathology are required to address these features in patients at high risk for psychosis. It is hoped that gene discovery will lead to the development of new drugs targeting each domain of psychopathology.

The advancing knowledge of human gene and protein sequences is altering the course of schizophrenia research. These technologies promise to help delineate the molecular pathophysiology of the disease and define novel molecular targets that will provide a basis for drug development. Clinical and postmortem studies are presenting more compelling definitions of disease entities and phenotypes within the schizophrenia syndrome, which should permit the effective application of advances in genomics and proteomics.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Kirkpatrick B, Buchanan RW, Ross DE, et al: A separate disease within the syndrome of schizophrenia. Arch Gen Psychiatry 58:165, 2001

2. Kendler KS, Gallagher TJ, Abelson JM, et al: Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. The National Comorbidity Survey. Arch Gen Psychiatry 53:1022, 1996

3. Neighbors HW, Trierweiler SJ, Ford BC, et al: Racial differences in DSM diagnosis using a semi-structured instrument: the importance of clinical judgment in the diagnosis of African Americans. J Health Soc Behav 44:237, 2003

4. Jablensky A, Sartorius N, Ernberg G, et al: Schizophrenia: manifestations, incidence and course in different cultures: a World Health Organization ten-country study. Psychol Med Monogr Suppl 20:1, 1992

5. Harrison G, Mason P: Schizophrenia: falling incidence and better outcome? Br J Psychiatry 163:535, 1993

6. Bachrach LL: What we know about homelessness among mentally ill persons: an analytical review and commentary. Treating the Homeless Mentally Ill. Lamb HR, Bachrach LL, Kass FI, Eds. American Psychiatric Association, Washington, DC, 1992, p 13

7. Cantor-Graae E, Nordstrom LG, McNeil TF: Substance abuse in schizophrenia: a review of the literature and a study of correlates in Sweden. Schizophr Res Mar 48:69, 2001

8. de Leon J, Diaz FJ, Rogers T, et al: Initiation of daily smoking and nicotine dependence in schizophrenia and mood disorders. Schizophr Res 56:47, 2002

9. Kendler KS, Diehl SR: Schizophrenia: genetics. Comprehensive Textbook of Psychiatry, 6th ed. Kaplan HI, Sadock BJ, Eds. Williams & Wilkins, Baltimore, 1995, p 942

10. Karlsson H, Bachmann S, Schroder J, et al: Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. Proc Natl Acad Sci USA 98:4634, 2001

11. Messias E, Kirkpatrick B, Bromet E, et al: Summer birth and deficit schizophrenia: a pooled analysis from six countries. Arch Gen Psychiatry (in press)

12. Sherrington R, Brynjolfsson J, Petursson H, et al: Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336:164, 1988

13. Tsuang MT, Stone WS, Faraone SV: Schizophrenia: a review of genetic studies. Harv Rev Psychiatry 7:185, 1999

14. Thaker GK, Carpenter WT Jr: Advances in schizophrenia. Nat Med 7:667, 2001

15. Chowdari KV, Mirnics K, Semwal P, et al: Association and linkage analyses of RGS4 polymorphisms in schizophrenia. Hum Mol Genet 11:1373, 2002

16. Straub RE, Jiang Y, MacLean CJ, et al: Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. Am J Hum Genet 71:337, 2002

17. Morris DW, McGhee KA, Schwaiger S, et al: No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. Schizophr Res 60:167, 2003

18. Stefansson H, Thorgeirsson TE, Gulcher JR, et al: Neuregulin 1 in schizophrenia: out of Iceland. Mol Psychiatry 8:639, 2003

19. Chumakov I, Blumenfeld M, Guerassimenko O, et al: Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proc Natl Acad Sci USA 99:13675, 2002

20. Leonard S, Gault J, Hopkins J, et al: Association of promoter variants in the alpha7 nicotine acetylcholine receptor subunit gene with an inhibitory deficit found in schizophrenia. Arch Gen Psychiatry 12:1085, 2002

21. Gogos JA, Santha M, Takacs Z, et al: The gene encoding proline dehydrogenase modulates sensorimotor gating in mice. Nat Genet 21:434, 1999

22. Freedman R, Coon H, Myles-Worsley M, et al: Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA 94:587, 1997

23. Rzanny R, Klemm S, Reichenbach JR, et al: ^{31}P-MR spectroscopy in children and adolescents with a familial risk of schizophrenia. Eur Radiol 13:763, 2003

24. Keshavan MS, Stanley JA, Montrose DM, et al: Prefrontal membrane phospholipid metabolism of child and adolescent offspring at risk for schizophrenia or schizoaffective disorder: an in vivo ^{31}P MRS study. Mol Psychiatry 8:316, 2003

25. Arolt V, Lencer R, Nolte A, et al: Eye tracking dysfunction is a putative phenotypic susceptibility marker of schizophrenia and maps to a locus on chromosome 6p in families with multiple occurrence of the disease. Am J Med Genet 67:564, 1996

26. Egan MF, Goldberg TE, Kolachana BS, et al: Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. Proc Natl Acad Sci USA 98:6917, 2001

27. Davis KL, Kahn RS, Ko G, et al: Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 148:1474, 1991

28. Lipska BK, Jaskiw GE, Weinberger DR: Postpubertal emergence of hyperresponsiveness to stress and to amphetamine after neonatal excitotoxic hippocampal damage: a potential animal model of schizophrenia. Neuropsychopharmacology 9:67, 1993

29. Carlsson A, Hansson LO, Waters N, et al: Neurotransmitter aberrations in schizophrenia: new perspectives and therapeutic implications. Life Sci 61:75, 1997

30. Laruelle M, Abi-Dargham A, van Dyck CH, et al: Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Nat Acad Sci USA 93:235, 1996

31. Breier A, Su TP, Saunders R, et al: Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission tomography method. Proc Natl Acad Sci USA 94:2569, 1997

32. Abi-Dargham A, Gil R, Krystal J, et al: Increased striatal dopamine transmission in schizophrenia: confirmation in a second cohort. Am J Psychiatry 155:761, 1998

33. Otani S, Auclair N, Desce JM, et al: Dopamine receptors and groups I and II mGluRs cooperate for long-term depression induction in rat prefrontal cortex through converging postsynaptic activation of MAP kinases. J Neurosci 19:9788, 1999

34. Seamans JK, Durstewitz D, Christie BR, et al: Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. Proc Natl Acad Sci USA 98:301, 2001

35. Benes FM: Emerging principles of altered neural circuitry in schizophrenia. Brain Res Rev 31:251, 2000

36. Bogerts B: The neuropathology of schizophrenic diseases: historical aspects and present knowledge. Eur Arch Psychiatry Clin Neurosci 249(suppl 4):2, 1999

37. Sanfilipo M, Lafargue T, Rusinek H, et al: Volumetric measure of the frontal and temporal lobe regions in schizophrenia: relationship to negative symptoms. Arch Gen Psychiatry 57:471, 2000

38. Selemon LD, Lidow MS, Goldman-Rakic PS: Increased volume and glial density in primate prefrontal cortex associated with chronic antipsychotic drug exposure. Biol Psychiatry 46:161, 1999

39. Glantz LA, Lewis DA: Decreased dendritic spine density on prefrontal cortical pyramidal neurons in schizophrenia. Arch Gen Psychiatry 57:65, 2000

40. Mirnics K, Middleton FA, Marquez A, et al: Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. Neuron 28:53, 2000

41. Buchanan RW, Breier A, Kirkpatrick B, et al: Structural abnormalities in deficit and nondeficit schizophrenia. Am J Psychiatry 150:59, 1993

42. Kemether EM, Buchsbaum MS, Byne W, et al: Magnetic resonance imaging of mediodorsal, pulvinar, and centromedian nuclei of the thalamus in patients with schizophrenia. Arch Gen Psychiatry 60:983, 2003

43. Holcomb HH, Lahti AC, Medoff DR, et al: Brain activation patterns in schizophrenic and comparison volunteers during a matched-performance auditory recognition task. Am J Psychiatry 157:1634, 2000

44. Kane J, Honigfeld G, Singer J: Clozapine for the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. Arch Gen Psychiatry 45:789, 1988

45. Rosenheck R: Effectiveness and cost of olanzapine and haloperidol in the treatment of schizophrenia: a randomized controlled trial. Department of Veterans Affairs Cooperative Study Group on the Cost-Effectiveness of Olanzapine. JAMA 290:2693, 2003

46. Carpenter WT, Gold JM: Another view of therapy for cognition in schizophrenia (editorial). Biol Psychiatry 52:969, 2002

47. Adler LE, Hoffer LD, Wiser A, et al: Normalization of auditory physiology by cigarette smoking in schizophrenic patients. Am J Psychiatry 150:1856, 1993

48. Sherr JD, Myers C, Avila MT, et al: The effects of nicotine on specific eye tracking measures in schizophrenia. Biol Psychiatry 52:721, 2002

49. Tharyan P, Adams CE: Electroconvulsive therapy for schizophrenia. Cochrane Database Syst Rev (2):CD000076, 2002

50. Lehman AF, Steinwachs DM: Translating research into practice: the Schizophrenia Patient Outcomes Research Team (PORT) project. Schizophr Bull 24:1, 1998

51. Fenton WS, Schooler NR: Evidence-based psychosocial treatment for schizophrenia. Schizophr Bull 26:1, 2000

52. Ley A, Jeffery DP, McLaren S, Siegfried N: Treatment programmes for people with both severe mental illness and substance misuse. Cochrane Database Syst Rev (4):CD001088, 2000

53. Thaker GK, Avila MT, Hong E, et al: A model of smooth pursuit eye movement deficit associated with the schizophrenia phenotype. Psychophysiology 40:277, 2003

54. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed Text Revision. American Psychiatry Association, Washington, DC, 2000, p 312

Acknowledgment

Figure 1 courtesy of L. Elliot Hong, Malle Tagamets, and Gunvant Thaker, Maryland Psychiatric Research Center, Baltimore, Maryland. Artist: Alice Y. Chen.

VIII ANXIETY DISORDERS

M. Katherine Shear, m.d.

Anxiety disorders are the most common mental disorders in the community. Anxiety disorders include panic disorder, agoraphobia, specific phobia, social phobia, generalized anxiety disorder, obsessive-compulsive disorder, and acute and posttraumatic stress disorder. Shared features of these disorders are prominent somatic symptoms, panic attacks, anticipatory worry or fear, and avoidance or compulsive rituals. Persistent nagging health concerns are common and frequently cause patients with these disorders to seek medical treatment, where their anxiety disorders may go unrecognized. In addition to producing serious impairment in functioning, anxiety disorders can worsen the course of medical illnesses. Moreover, untreated anxiety symptoms lead to ineffective utilization of health care resources. Anxiety disorders frequently occur together with depression, a condition known to presage poorer medical outcomes. In addition to their similarities, the anxiety disorders have distinct clinical and neurobiologic features and typical presentations [*see Table 1*].

Approach to Management

Anxiety disorders are best managed through a systematic approach to diagnosis, treatment, and, when necessary, referral. To facilitate accurate detection of anxiety symptoms, physicians should be vigilant for the behavioral, physiologic, and cognitive features of anxiety and know the key criteria for different disorders. Anxiety disorders should be included in the differential diagnosis in many cases of apparently medical illness.

When taking the history in a patient who may have an anxiety disorder, direct and sensitive questioning is required. Phobic responses can be cognitive as well as behavioral, leading patients to neglect to mention anxiety. Fear of stigma may also lead to patient reticence.

Therapy for patients with anxiety can be initiated by the primary care physician or, in some cases, by a consulting physician. It is a truism in medicine that the first step in treatment is accurate diagnosis. Thus, upon recognition of anxiety disorder symptoms, the physician should identify the condition and provide the patient with information about its treatment and course. Such information is readily available on the Internet [*see Sidebar* Internet Resources for Anxiety Disorders]. In addition, information on panic disorders can be obtained from the Institutes of Mental Health by calling 1-800-64-PANIC.

Treatment by physicians usually entails the use of medication. Although drug treatment is certainly an appropriate intervention, the majority of patients with anxiety disorder prefer nonpharmacologic treatment. Highly effective cognitive-behavioral treatments are available. It is best if physicians discuss treatment preferences with their patients and are prepared to provide appropriate referrals if the patient so wishes.

When medications are used, effective management techniques include monitoring patients' responses to treatment by using one or more target symptoms. General rules for pharmacologic treatment of anxiety include the following: (1) serotonin-active medications are effective for every disorder except specific phobias, (2) patients with panic disorder are highly sensitive to medication, so drugs should be started at a very low dose and gradually increased to a therapeutic level, (3) barring adverse effects, a medication trial should be continued for 4 weeks before one declares the agent ineffective and changes medications, and (4) failure to respond to an initial trial of a selective serotonin reuptake inhibitor (SSRI) does not mean that another agent of this class will not work.

Referral to a mental health specialist is indicated for the following reasons: (1) uncertainty about the diagnosis, (2) presence of multiple concomitant psychiatric conditions, (3) acute suicidal ideation or a past history of serious suicide attempt or other serious behavioral disturbances, and (4) failure to respond to two trials of normally effective medications. Referral information is available at many Web sites devoted to anxiety disorders [*see Sidebar* Internet Resources for Anxiety Disorders].

Panic Disorder with or without Agoraphobia

The pathognomonic feature of panic disorder is unexpected panic attacks, which are characterized by sudden onset and rapid escalation of somatic symptoms referable to the autonomic nervous system (e.g., chest pain, shortness of breath, heart palpitations, and dizziness) along with fear or apprehension [*see Tables 1 and 2*]. Panic attacks typically last 10 to 20 minutes, but they can be shorter. Persons with panic disorder have recurrent panic attacks; between attacks, they are beset by fear of

Table 1 Key Symptoms of Anxiety Disorders

Disorder	*Symptoms*
Panic disorder	Recurrent unexpected panic attacks; fear of panic or its consequences
Agoraphobia	Fear and avoidance of places where it may be difficult to get help or easily leave (e.g., bridges, tunnels, restaurants, supermarkets, public transportation)
Posttraumatic stress disorder	Exposure to a life-threatening or physically threatening event, followed by (1) disturbing images, flashbacks, nightmares; (2) avoidance of trauma reminders; (3) hyperarousal (e.g., easy startle, irritability, difficulty concentrating, insomnia)
Generalized anxiety disorder	Excessive uncontrollable worries about everyday things; fatigue, muscle tension, insomnia; restlessness, irritability, difficulty concentrating
Social phobia	Fear and avoidance of specific social situations (e.g., fear of urinating in public bathrooms, fear of eating in restaurants, fear of public speaking) or social situations in general
Obsessive-compulsive disorder	Repeated intrusive thoughts, impulses, or images, recognized as irrational; compulsive ritualistic behaviors such as cleaning or checking
Specific phobias	Fear and avoidance of a specific place, activity, or situation (e.g., fear of blood or needles, fear of dental procedures, fear of childbirth)

Internet Resources For Anxiety Disorders

Multipurpose Sites

Anxiety Disorders Association of America
http://www.adaa.org
http://www.anxieties.com
General information, referral information, patient materials, screening measures, and medication information

Anxiety Disorders Clinic, McMaster University Medical Centre
http://www.macanxiety.com
General information, patient materials, and Canadian referral information

Freedom From Fear
http://www.freedomfromfear.com
General information, patient materials, screening measures, and medication information

National Institute of Mental Health–Anxiety
www.nimh.nih.gov/anxiety/anxietymenu.cfm
General information, referral information, patient materials in English and Spanish, screening measures, and medication information

General Information and Patient Materials

Lifeline Anxiety Disorder Newsletter
http://www.designandcopy.ca/lifeline

Mental Health: A Report of the Surgeon General
http://www.surgeongeneral.gov/library/mentalhealth/chapter4/sec2.html

Mental Health Net
http://mentalhelp.net/poc/center_index.php?id=1

Practice Guidelines

American Psychiatric Association
http://www.psych.org/clin_res/prac_guide.cfm

Psychotherapeutic Referral Sources

Academy of Cognitive Therapy
http://www.academyofct.org

American Board of Professional Psychology
http://www.abpp.org

American Psychiatric Association
www.psych.org/public_info/APA~1.HTM

Screening for Anxiety

New York University School of Medicine–Psychiatry
www.med.nyu.edu/Psych/screens/anx.html

Generalized Anxiety and Other Anxiety Disorders

Generalized Anxiety Disorder (GAD) Study
http://www.bu.edu/anxiety/gad.html

Obsessive-Compulsive Disorder

Obsessive-Compulsive Foundation
http://www.ocfoundation.org

Posttraumatic Stress Disorder and Other Stress-Related Conditions

International Society for Traumatic Stress Studies (ISTSS)
http://www.istss.org

Social Anxiety Disorder

Madison Institute of Medicine
www.socialfear.com

The Social Anxiety Network
http://www.social-anxiety-network.com

Social Phobia/Social Anxiety Disorder: Effective Treatment 2003
http://www.socialfear.com

the attacks and their consequences or implications. For example, the patient may believe that the panic attacks reflect underlying heart disease or schizophrenia.

Agoraphobia is a fear of situations in which the person would feel trapped or alone should a panic episode occur. Such situations typically include travel far away from home, on public transportation, or via bridges or tunnels; crowded places such as supermarkets, restaurants, theaters, churches, or shopping malls; standing in line; and being alone. Agoraphobia is characterized by avoidance of these situations. In its most severe form, sufferers can be housebound.

EPIDEMIOLOGY

The lifetime prevalence rates of panic disorder are about 3.5%.[1] Up to 10% of the population experience sporadic panic attacks. Panic disorder is more common in women and has an average-age-group onset of early adulthood. Persons with panic disorder are at risk for developing major depression, with over half meeting lifetime criteria for major depression.

Lifetime prevalence rates for agoraphobia are estimated at 6.7%, with a mean age of onset of 29 years.[2] Agoraphobia can occur when the panic attacks include frightening physical symptoms, such as those suggesting a medical disorder.

PATHOPHYSIOLOGY

In laboratory studies, paniclike symptoms have been provoked with pharmacologic agents such as carbon dioxide, sodium lactate, and cholecystokinin (CCK). Several neurotransmitter systems have been implicated in these symptoms, including serotonin (5-HT), norepinephrine, and γ-aminobutyric acid (GABA), along with the neuromodulatory CCK system. Theories involving 5-HT in panic disorder focus on opposing actions in the dorsal raphe nucleus (DRN) and the median raphe nucleus (MRN).[3] The MRN modulates fear and anticipatory anxiety, along with autonomic symptoms of panic, whereas the DRN modulates the behavioral aspects of panic, such as the fight-or-flight response. Both systems are involved in a feedback loop that also includes the locus coeruleus and lateral hypothalamus.[4]

Yohimbine, an α_2-adrenergic antagonist, can trigger anxiety or panic in susceptible patients,[5] which suggests that norepinephrine also has a role in panic disorder. The locus coeruleus has projections to brain structures thought to play a role in anxiety, including the amygdala, the bed nucleus of the stria terminalis (BNST), the periaqueductal gray area (PAG), the paraventricular nucleus of the hypothalamus, the lateral hypothalamus, and the nucleus of the tractus solitarius (NTS). Norepinephrine projections from the locus coeruleus also interact with corticotropin-releasing factor (CRF), which may play a role in anxious responses to stress.[6]

GABA has been implicated in the pathophysiology of panic disorder, perhaps as a nonspecific mechanism related to anxiety. Persons with panic disorder demonstrate decreased GABA ligand binding in the right orbitofrontal cortex and right insula.[7] Decreased platelet benzodiazepine receptor binding has also been observed in patients with panic disorders and other anxiety disorders. On magnetic resonance spectroscopy, patients with panic disorder have been shown to have decreased GABA function and lower levels of GABA in the occipital cortex.[8] GABA interacts with CCK in the amygdala, cortex, and hippocampus, and this interaction has been postulated to play a role in mediating anxious states.

Other models focus on the ethologic roots of panic disorder.

Table 2 Conditions Commonly Confused with Anxiety Disorders

Disorder	*Conditions*
Panic disorder	Heart palpitations, mitral valve prolapse, cardiac arrhythmias, hyperthyroidism, irritable bowel syndrome, migraine
Posttraumatic stress disorder	Insomnia, substance use disorders, unexplained fears of medical procedures, or unexplained worsening of course of medical illness
Generalized anxiety disorder	Insomnia, migraine or tension headaches, hypertension, peptic ulcer disease, irritable bowel syndrome
Social phobia	Unexplained urinary retention, academic or occupational underachievement, mood disorders, substance use disorders
Obsessive-compulsive disorder	Tourette syndrome, other tic disorders, pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS), excoriative dermatitis, unexplained fears of illness

In his false-suffocation alarm theory, Klein hypothesized a brain stem abnormality causing dysregulation and heightened sensitivity to carbon dioxide. Respiratory abnormalities have been consistently documented in patients with panic disorder[9]; these patients exhibit unstable respiratory patterns and high respiratory variability even when not experiencing panic.[10] The respiratory stimulant doxapram induces panic in many patients with panic disorder, as well as in some control subjects; a simple cognitive intervention can reduce the panic response to doxapram. Decreased heart rate variability related to decreased vagal tone has been documented in panic disorder.[11] Decreased vagal tone may play a role in the migraine and irritable bowel syndrome associated with some cases of panic disorder.[12]

DIAGNOSIS

Clinical Features

Panic attack is diagnosed when at least four of 13 symptomatic criteria occur unexpectedly and peak within 10 minutes [*see Table 3*]. The diagnosis of panic disorder requires recurrent panic attacks accompanied by significant worry or concern about panic, its consequences or implications, or a change in usual behavior.[13] Agoraphobia is diagnosed when the patient fears and avoids situations in which escape would be difficult or help would be unavailable if a panic attack were to occur.

Fear and hyperawareness of bodily sensations are hallmarks of panic disorder.[14] These patients misinterpret benign bodily sensations (e.g., dizziness, palpitation) as indications of serious medical illness.

Presentation in Medical Settings

Persons with panic disorder are more likely to present for medical treatment than for mental health treatment. The majority of cases of panic disorder seen in primary care practices and emergency departments go unrecognized.[15,16]

Panic disorder has four common clinical presentations in general medicine: (1) physical symptoms (e.g., heart palpitations, chest pain, shortness of breath, gastrointestinal complaints, headache, dizziness), (2) anxiety and tension, (3) hypochondriacal concerns, and (4) certain diagnosed medical conditions, such as migraine, asthma, chronic obstructive pulmonary disease (COPD), and labile hypertension.[17]

Panic disorder is often seen in patients referred for cardiologic complaints, especially atypical chest pain or chest pain despite angiographically normal coronary arteries. Patients with microvascular angina may have panic disorder. Because of prominent cardiovascular symptoms during panic attacks, this disorder should be considered in patients with intractable cardiac symptoms. Several studies have shown that men with phobic anxiety are at higher risk of cardiovascular mortality from fatal myocardial infarction and sudden cardiac death.[18] Idiopathic clinical and subclinical cardiomyopathies have been associated with panic disorder,[19] possibly related to increased adrenergic activity. Panic attacks are associated with transient increased blood pressure,[20] so this diagnosis should be considered in patients with labile hypertension. Panic disorder has been found in about 16% of patients with implantable cardioverter defibrillators.[21]

There is an association of respiratory illness with panic disorder. Patients with asthma have a higher incidence of panic disorder, and the presence of panic disorder predicts a poorer clinical course in asthma. A history of childhood respiratory illness is more common in patients with panic disorder than in those with other psychiatric disorders.[22] In adults, COPD is associated with panic disorder.

Panic disorder is frequently associated with other mood and anxiety disorders. Prognosis is poorer in such cases.

TREATMENT

Pharmacotherapy

Drug therapy can be markedly effective for patients with panic disorder [*see Table 4*]. However, the fear of bodily sensations in these patients may lead to difficulty taking medications, medication refusal, or poor adherence to prescribed medicine. Many patients with panic disorder believe they are allergic to medication. Physicians should convey understanding of this medication sensitivity. Whenever possible, medication for any purpose should be started at a very low dosage and then

Table 3 Diagnostic Criteria for Panic Attack*

Palpitations, pounding heart, or accelerated heart rate
Sweating
Trembling or shaking
Sensations of shortness of breath or smothering
Feeling of choking
Chest pain or discomfort
Nausea or abdominal distress
Feeling dizzy, unsteady, light-headed, or faint
Derealization (feelings of unreality) or depersonalization (being detached from oneself)
Fear of losing control or going crazy
Fear of dying
Paresthesias
Chills or hot flushes

*Four or more of these symptoms must develop abruptly and peak within 10 min.

Table 4 FDA-Approved Drugs for Anxiety Disorders

Disorder	*Drugs*	*Daily Dosages*	
		Starting	*Target*
Panic disorder	Imipramine	10–25 mg	150–200 mg
	Paroxetine	10 mg	10–60 mg
	Sertraline	25–50 mg	50–200 mg
	Alprazolam	0.25–0.50 mg b.i.d or t.i.d.	0.5–2.0 mg b.i.d. or t.i.d.
Posttraumatic stress disorder	Sertraline	50 mg	50–200 mg
	Paroxetine	10 mg	10–60 mg
Generalized anxiety disorder	Venlafaxine	75 mg in two or three divided doses	150–300 mg in two or three divided doses
Social anxiety disorder (social phobia)	Paroxetine	10–20 mg	10–60 mg
	Sertraline	50 mg	50–200 mg
Obsessive-compulsive disorder	Clomipramine	25–50 mg	150–300 mg
	Fluoxetine	5–10 mg	20–80 mg
	Fluvoxamine	50 mg h.s.	100–300 mg in two divided doses
	Paroxetine	10–20 mg	40–60 mg

gradually increased. This is especially important when using medication to treat the panic disorder itself. If the initial dose is too high, the patient may not tolerate the medication and a potentially effective treatment will be rendered useless. There is a psychological intervention of proven efficacy for fear of bodily sensations, so referral should be considered for patients in whom this fear complicates drug therapy.

Imipramine A tricyclic antidepressant with serotoninergic activity, imipramine was the first medication used for panic disorder. This drug has well-documented efficacy in controlling panic disorder and can prevent relapse with continued use.[23] The efficacy and tolerability of imipramine are similar to those of SSRIs, so patients who are responding well to imipramine need not necessarily be switched to an SSRI. However, most experts now start drug therapy for panic disorder with an SSRI.

Selective serotonin reuptake inhibitors SSRIs are the first-line pharmacotherapy for panic disorder. Every medication in this class has been proved effective. The Food and Drug Administration has approved paroxetine and sertraline for use in panic disorder, but citalopram, fluoxetine, and fluvoxamine also have well-documented beneficial effects[24] that are often lasting, even after discontinuance of the medication. However, several studies with different SSRIs show that continuing the medication for up to 80 weeks lowers relapse rates. There is a tendency for side effects to decrease over time.[25] Different SSRIs have different chemical structures and somewhat different side-effect profiles, so patients who do not respond to one agent or who fail to tolerate it may do well on a different drug in this class. Standard practice dictates trying several SSRIs before moving to another class of medication.

Other antidepressants Antidepressants with mixed neurotransmitter effects, such as venlafaxine or mirtazapine, can be used to treat panic disorder. It is less clear whether selective noradrenergically active medications are effective for panic disorder. Studies of nortriptyline suggest it may be therapeutic. Studies of bupropion have showed mixed results. Reboxetine has demonstrated efficacy in the treatment of panic disorder.

Benzodiazepines Although benzodiazepines are not first-line agents for the treatment of panic disorder, they provide relief for some patients who do not respond to other medications. Alprazolam is approved by the FDA for the treatment of panic disorder; clonazepam and lorazepam also have proven efficacy. When used, benzodiazepines should be prescribed on a daily basis rather than on an as-needed basis.

Many physicians are wary of prescribing benzodiazepines because these agents are potentially addictive. Certainly, benzodiazepines should be avoided if possible in patients with a personal or family history of alcohol or substance-abuse disorders. Patients without such a history are unlikely to abuse benzodiazepines, however.

Discontinuance of a benzodiazepine should be accomplished by very slow tapering. In patients who have difficulty being weaned from the drug, cognitive-behavioral therapy can be used to promote discontinuance.[26]

Psychological Treatment

Targeted cognitive-behavioral psychotherapy is as effective as medication in treating panic disorder.[27] Gains are maintained at 1-year follow-up, and a naturalistic follow-up study found that in some cases, the effects endured for as long as 14 years after treatment. A brief psychological intervention in the emergency department can reduce future visits by panic disorder patients.[28] Simple instructions advising patients not to avoid situations in which they experienced panic or feared the onset of panic, along with educational support about panic disorder, is helpful in an emergency setting.

When to Refer

Most cases of panic disorder, with or without agoraphobia, respond readily to drug therapy or targeted cognitive-behavioral treatment. Referral to a psychiatrist is indicated for patients whose panic disorder is relatively treatment resistant, those with comorbid depression (especially bipolar illness) or other psychological disorders, and those who are suicidal.

COMPLICATIONS

In part because clinicians fail to recognize the condition, per-

sons with panic disorder are heavy utilizers of health care services and have significant disability, with high rates of unemployment and substance abuse.[29] Panic disorder is associated with suicide risk, even in the absence of concomitant depression.

Acute and Posttraumatic Stress Disorders

Acute and posttraumatic stress disorders (ASD and PTSD) are conditions that follow exposure to violence. ASD is diagnosed up to 4 weeks after a traumatic episode. PTSD is a highly debilitating condition that is not apparent until at least 1 month after trauma exposure.

EPIDEMIOLOGY

Lifetime exposure to violence is common in the general population. In a recent large United States survey, 51% of women and 61% of men reported having a traumatic experience at some time in their lives.[30] Lifetime prevalence rates for PTSD are estimated at 7.8%, with women twice as likely to have the disorder as men (10.4% versus 5.0%). Approximately 9% to 25% of persons who experience a trauma eventually develop posttraumatic stress symptoms. Nearly 80% of those with ASD subsequently develop PTSD, compared with only 4.3% of persons who experience a trauma but do not qualify for an ASD diagnosis.

PATHOPHYSIOLOGY

Acute stress stimulates the hypothalamic-pituitary-adrenal (HPA) axis, leading to secretion of CRF and adrenocorticotropic hormone (ACTH). Increased levels of CRF have been found in the cerebrospinal fluid of patients with PTSD, as compared with CSF levels in control subjects.[31] Chronic stress has been shown to suppress development of granule neurons in the dentate gyrus of the hippocampus,[32] and patients with chronic PTSD have lower hippocampal volumes than control subjects[33]; this may result in reduced hippocampal inhibition of CRF. Increased CRF secretion then leads to greater CRF interaction with noradrenergic projections from the locus coeruleus, producing further increases in both CRF and norepinephine. Locus coeruleus activity increases with chronic uncontrollable stress. Patients with PTSD have higher urinary norepinephrine and epinephrine levels than nonanxious control subjects, and urinary catecholamine levels correlate with intrusive recollection of trauma in combat veterans.[34] Catecholamines are responsible for alertness and vigilance, and sympathetic arousal enhances memory encoding.

Alterations in benzodiazepine receptors may contribute to PTSD. Reduction in peripheral benzodiazepine receptor density has been documented in combat veterans with PTSD, and a single-photon emission computed tomography (SPECT) study showed reduced affinity for benzodiazepine receptor binding in the prefrontal cortex. On functional MRI studies, PTSD patients have an exaggerated amygdala response to threatening stimuli.[35]

DIAGNOSIS

Clinical Features

Diagnostic criteria for PTSD include exposure to an event that posed a risk of death or serious physical injury, along with subsequent symptoms that cluster in three areas: (1) reexperiencing the event (e.g., recurrent thoughts, images, dreams, illusions, and flashback episodes), (2) avoidance of trauma reminders, and (3) arousal (e.g., restlessness, insomnia, hypervigilance, difficulty concentrating, and irritability). Dissociative symptoms (e.g., numbing or detachment, reduced awareness of surroundings, depersonalization, derealization, and dissociative amnesia) may also be present, especially in the early period after a trauma. These reactions are predictive of PTSD, as is increased heart rate during the acute-trauma phase.

Presentation in Medical Settings

Evidence suggests that a patient with PTSD is more likely to present for treatment in a primary care setting than in a mental health setting. However, primary care physicians recognize anxious symptoms in these patients only about 50% of the time and often fail to make the diagnosis.[36]

Patients who present to a trauma or emergency service are at risk for PTSD, as are combat veterans and victims of rape or domestic violence. Frightening or painful medical illness or procedures also can trigger PTSD. Assisted ventilation for acute respiratory distress syndrome, cardiac defibrillation, cardiac and lung transplantation,[37] cancer diagnosis or treatment,[38] stillbirth,[39] and patient awareness during surgical procedures because of inadequate anesthesia[40] have all reportedly triggered PTSD. It is likely that any life-threatening medical event (e.g., cardiac arrest, diabetic coma, myocardial infarction, pulmonary embolism, or massive organ failure) could trigger these symptoms, especially in patients with a history of anxiety or depressive illness.

Physicians and other medical staff in the emergency department should provide patients with support and information about PTSD in the early aftermath of trauma; such measures can help with quick recognition of the disorder if it should develop. Early recognition may prevent the development of full-blown PTSD and its negative consequences for illness course.

TREATMENT

Acute Stress Disorder

No pharmacologic treatment has proved effective in acute stress disorder, although beta blockers are currently under investigation. Studies of cognitive-behavioral therapy in patients with acute stress disorder suggest that it might prevent PTSD in patients who have experienced trauma and seem to be struggling with their reaction to the event.[41]

Posttraumatic Stress Disorder

Pharmacotherapy SSRIs are effective in the treatment of PTSD. Sertraline and paroxetine have FDA approval for this indication: studies of sertraline show about a 50% reduction in symptoms in each symptom cluster, and response rates with paroxetine range from 54% to 62%. As with other anxiety disorders, however, every SSRI tested has been found to work. At least one study showed nefazodone to be helpful.

Psychological treatment An intensive form of cognitive-behavioral therapy that requires patients to imagine themselves being exposed to the traumatic event is highly effective for rape-related PTSD.[41] This technique produces significant reduction in all symptom clusters of the disorder in most patients. The limitations of the treatment are that it requires a skilled therapist and courage on the part of the patient. However, if a qualified therapist is available and the patient is willing to participate, cognitive-behavioral therapy is the treatment of choice.

COMPLICATIONS

Traumatic stress reactions are associated with debilitating psychosocial and physical impairment,[42] including increased length of hospital stay for medical conditions, independent of severity of medical illness or medical comorbidity. PTSD also contributes to prolongation of recovery from medical illness. PTSD is associated with nonadherence to medication regimens after myocardial infarction[43] and independently predicts mortality in heart transplant patients. PTSD frequently coexists with substance-abuse disorder or mood disorder.

Generalized Anxiety Disorder

The defining characteristic of generalized anxiety disorder (GAD) is persistent, excessive, and uncontrollable worry about everyday life situations. GAD can be highly debilitating and may predispose to the development of other mood or anxiety disorders.[44]

EPIDEMIOLOGY

Lifetime prevalence rates of GAD range from 4% to 7%, with nearly a 10% prevalence in women older than 40 years.[45] Onset is sometimes early in life but is also common in middle to late adulthood, with a large increase in incidence at 35 to 45 years of age. Once established, GAD often becomes a chronic condition lasting 20 years or longer. GAD is the most common anxiety disorder in the elderly.

PATHOPHYSIOLOGY

Patients with GAD exhibit reductions in heart rate variability. Heart rate variability—as measured by beat-to-beat changes in the R-R interval on ECG—is considered a measure of autonomic nervous system function, reflecting the net effect of sympathetic versus parasympathetic (vagal) stimulation. In studies of heart rate variability during experimentally induced periods of worry and periods of calmness, normal control subjects had lower heart rate variability during periods of worry than during periods of calmness, whereas patients with GAD showed little change in variability between the two states. Although worrying is potentially adaptive if it is specific and limited to problem solving, chronic emotional restraint and inhibition could lead to a reduced range in physiologic response in persons with GAD.[46]

Muscle tension is a criterion symptom for GAD. Electromyographic activity correlates with right hemispheric brain activity in patients with GAD, suggesting that right hemisphere activity may control muscle tension symptoms in this population.[47] The muscle tension in GAD patients does not reflect autonomic arousal, however; rather, GAD is characterized by decreased autonomic arousal and activity.[46] GAD patients react to stressors with less autonomic flexibility than nonanxious control patients do and have inaccurate perceptions of physiologic changes.

SPECT and MRI studies show decreased benzodiazepine receptor binding in the left temporal pole in patients with GAD.[48] Vagally mediated heart rate variability is positively correlated with blood flow to the medial prefrontal cortex during emotional processing tasks.[49]

DIAGNOSIS

Clinical Features

Worry is the key pathognomonic feature of GAD. Other diagnostic features include restlessness, fatigue, irritability, difficulties with concentration, muscle tension, and insomnia. The diagnosis of GAD is made when patients report at least 6 months of persistent worry accompanied by at least three of these associated clinical features.

Presentation in Medical Settings

GAD is the most common anxiety disorder seen in primary care settings,[50] often presenting as sleep disturbance or somatic symptoms such as muscle aches and tension headaches.[51] GAD is associated with chronic medical conditions, poor general health, frequent hospital admissions, and high health care costs and utilization.[52] Like other anxiety disorders, GAD often goes undiagnosed and untreated, especially when accompanied by physical illness. GAD patients are as likely to seek treatment from a cardiologist as are patients with panic disorder.[53]

TREATMENT

Pharmacotherapy

The FDA has approved venlafaxine for use in GAD, and most experts consider this agent to be the first-line treatment. Venlafaxine also relieves anxious and depressive symptoms in GAD patients with comorbid major depressive disorder. SSRIs have also been found to be effective in GAD.[54]

Benzodiazepines have been used to treat GAD, and at least one study suggests that their use does not lead to tolerance.[55] However, these drugs are generally not used as first-line treatment. Many patients with GAD also have depression, which benzodiazepines do not relieve.

Psychological Treatment

Cognitive-behavioral therapy for GAD has been less well studied than that for other anxiety disorders.[56] However, this approach appears promising. An intervention that targets worry has shown good results that were maintained at 6-month and 12-month follow-up. Because of the chronic nature of GAD, the efficacy of combining psychodynamic therapy with cognitive-behavioral techniques has been suggested, and a preliminary evaluation of psychodynamic treatment of GAD found positive results. A randomized trial of integrated interpersonal and cognitive-behavioral therapy, which is currently under way, suggests improved efficacy over either therapy alone.

Social Anxiety Disorder (Social Phobia)

Social phobia, or social anxiety disorder, is characterized by fear or avoidance of negative evaluation by others. Patients with this condition expect embarrassment or humiliation in a wide range of social encounters. Social phobia may occur in a specific form, with fear of one situation (most often, public speaking), or in a generalized form that is debilitating and chronic.

EPIDEMIOLOGY

In large, community-based studies, the lifetime prevalence of social phobia ranges from 7.2% to 9.5%.[57] The disorder is more common in women, normally begins in adolescence or early adulthood, and is characterized by a chronic course. One study suggests that only 30% of patients experience remission.[58]

PATHOPHYSIOLOGY

A recent review posits some interesting possibilities for amygdala involvement in social behavior and social anxiety,

suggesting that study of social primates, such as the macaque, may be informative for human anxiety.[59] Hypothetically, the amygdala may act as a behavioral brake mechanism, functioning to support adaptive evaluation of a novel situation. Social or other anxiety disorders may be associated with enhanced braking and, consequently, too little exposure and learning. A recent study of functional MRI in patients with social phobia has extended this idea by showing that these patients are hypersensitive to human facial displays of emotion, particularly anger, threat, or fear, and exhibit exaggerated amygdala responses to such facial expressions.[60] Patients with generalized social phobia have also been found to have low density of platelet peripheral benzodiazepine receptors, as occurs in other anxiety disorders.

DIAGNOSIS

Clinical Features

Persons with social phobia have persistent fear or avoidance of situations in which they expect to be scrutinized by others or to do something that is humiliating or embarrassing. Common examples are public speaking, eating in restaurants, writing in public (e.g., signing checks or sales slips), urinating in public bathrooms, talking with people in authority, or conversing in other situations of scrutiny, such as parties or social gatherings. Generalized social phobia can be very debilitating, because the person has persistent fear of almost any social situation, including speaking to a supervisor or teacher or even a physician, going to parties, or asking for assistance.

Persons with social phobia have a tendency toward negative interpretation of social events, self-focused attention, and negative-outcome–based cognitive processes (e.g., rumination and anticipatory anxiety). They have a memory bias for negative emotions in others' faces and expressions.[61] Normal individuals tend to blame negative social interactions on others while taking credit for positive experiences. Persons with social phobia often reverse this so-called self-serving bias, blaming themselves for negative experiences and crediting others when interactions are positive.[62]

Presentation in Medical Settings

A person with social phobia who fears urinating in a public bathroom can present with urinary retention. I have treated such a patient, who was serving in the military and thus had minimal privacy, and whose urinary retention was so severe that it resulted in hospitalization. Although there are few specific presentations of social phobia in medical settings, physicians should be alert to this condition. As is the case with other anxiety disorders, persons with social phobia frequent primary care settings[63] and are often not diagnosed.[64] A recent study found that only 2% of persons with social phobia who present to primary care clinicians are diagnosed and that fewer are effectively treated.[65]

TREATMENT

Pharmacotherapy

SSRIs are first-line pharmacotherapy for social phobia. Paroxetine and sertraline have FDA approval, but other SSRIs are also effective. Paroxetine response rates range from 55% to 70%. Fluvoxamine and sertraline response rates are similar.

High-potency benzodiazepines, especially clonazepam, are effective in the treatment of social phobia. Although benzodiazepines are not first-line treatment for social phobia, they can be a highly useful tool in the treatment of this disorder. Monoamine oxidase inhibitors are effective but are rarely used because of the risk of hypertensive crisis and the availability of other, less toxic choices.

Psychological Treatment

Cognitive-behavioral treatments are effective in social phobia,[66] although the disorder presents some special challenges for this approach. Several successful strategies have been developed and tested.[67] As with all the anxiety disorders, delivery of this treatment requires special training and expertise.

COMPLICATIONS

Social phobia is one of the most disabling anxiety disorders, because it can affect virtually every area of potential achievement. Patients with social phobia tend to have less education, lower occupational functioning, and lower health-related quality of life; and they utilize health care resources significantly more than nonanxious persons.[68] Comorbid mood disorders, other anxiety disorders, and substance-abuse disorders are common. Patients with social phobia are as likely to commit or attempt suicide as patients with major depressive disorder.

Obsessive-Compulsive Disorder

Obsessive-compulsive disorder (OCD) is characterized by repeated intrusive thoughts, ideas, or images (obsessions) and by repeated ritualistic behaviors (compulsions). Afflicted persons recognize the irrationality of their thoughts but are powerless to control them. Similarly, compulsive rituals are seen as senseless but frequently take many hours a day to perform. Obsessions have stereotyped content involving fear of contamination or aggression, persistent doubting, increased religiosity, or the need for symmetry. Associated compulsions often entail cleaning, checking, touching, or counting or other mental activities.

EPIDEMIOLOGY

OCD is relatively rare in comparison with other anxiety disorders, but it can be one of the most debilitating of psychiatric disorders. The estimated lifetime prevalence is approximately 1% to 3%, with monthly prevalence ranging from 0.6% to 3.3%. Mean age of onset is 24 years. The disorder is often chronic, lasting up to 40 years.[69]

PATHOPHYSIOLOGY

There is converging evidence for abnormal activity in the orbitofrontal basal-ganglia thalamocortical circuit in patients with OCD.[70] Positron emission tomography (PET) in these patients has shown that improvement after treatment with an SSRI is associated with decreased metabolism in the right caudate and putamen, the bilateral orbitofrontal cortex and thalamus, and parts of the prefrontal cortex. Functional MRI scans show increased metabolism in frontal subcortical regions after clinical improvement.

A possible autoimmune etiology of OCD has recently been suggested. The antibody D8/17, which is found in rheumatic fever and Sydenham chorea, is elevated in patients with OCD and Tourette syndrome; this elevation predicts larger basal ganglia volume. In childhood, OCD sometimes occurs as a poststreptococcal syndrome known as PANDAS (pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections).[71]

DIAGNOSIS

Clinical Features

OCD is diagnosed when obsessions, compulsions, or both are present for at least 1 hour a day or present at a level that interferes with functioning. Chronicity is the rule, although symptoms may change in focus over time.

Presentation in Medical Settings

A patient with OCD may present to any of several types of specialists, depending on the form the disorder takes. Approximately half of OCD patients partake in washing compulsions[72] in response to obsessive fears relating to germs and contamination. Repetitive hand washing is associated with increased risk for nonspecific dermatitis, which can result in these patients seeking dermatologic treatment.[73] OCD patients may also be seen by dermatologists for excoriation, or excessive picking and scratching of the skin. Those patients with contamination obsessions may present for treatment at an infectious disease practice as a result of their irrational fears.[73] OCD patients may be seen by rheumatologists for physical symptoms resulting from somatization. Many of those with contamination fears are hypochondriacal, fearing a wide range of illnesses. The result is increased utilization of health care resources and time spent with physicians. Primary care physicians may encounter patients with OCD. However, many patients with OCD feel considerable shame about their disorder and may be reluctant to reveal its existence. Like other psychiatric conditions, OCD is clinically recognized by primary care physicians in only 10% to 30% of cases.

TREATMENT

Pharmacotherapy

Serotonin-active antidepressants are the clear first-line pharmacotherapy for OCD. Clomipramine, a serotonin-active tricyclic antidepressant, was the first drug to receive FDA approval for treatment of OCD. Nearly 60% of patients improve substantially with clomipramine, compared with only 3% of those persons who receive placebo. Fluoxetine, sertraline, fluvoxamine, paroxetine, and citalopram have demonstrated comparable degrees of efficacy and tolerability in the treatment of OCD.[74]

Psychological Treatment

Cognitive-behavioral therapy focusing on exposure and response prevention is highly effective for OCD,[75] with over 85% of participants demonstrating clinical improvement in the hands of experts. Adding clomipramine therapy to cognitive-behavioral therapy does not result in improved outcome, whereas the addition of cognitive-behavioral therapy to pharmacotherapy does improve outcome. However, OCD patients with significant comorbid depression might benefit from SSRI treatment before beginning cognitive-behavioral therapy.

COMPLICATIONS

OCD is associated with significant social and occupational impairment, unemployment, and burden to family members.[76]

Specific Phobia in Medical Settings

Specific phobias are irrational fears, usually accompanied by avoidance of the feared stimulus. Physicians should be especially aware of phobias regarding blood and injections. Patients with such phobias may have syncopal episodes on exposure to blood or even to hypodermic needles. Because of their phobias, these patients may avoid obtaining needed tests or even seeking proper medical care; this can result in serious illnesses going undiagnosed and untreated.

Patients with blood-injury phobia may experience significant emotional distress if the medical staff is dismissive or unsympathetic about this problem. Blood-injury phobia differs from other anxiety and phobic reactions in that heart rate slows, rather than accelerates, on exposure to the phobic stimulus. These patients often have a family history of this condition, and it is considered to be a neurobiologic illness rather than a deficit in self-control. Studies document that short-term treatment for blood-injury phobia can be accomplished by trained therapists. Patients who receive effective treatment experience tremendous relief.

Other medically relevant phobias include dental phobia and phobia of childbirth (also called tokophobia). Persons with dental phobia experience extreme apprehension and may avoid dental treatment completely. Psychological intervention in the form of one-session exposure and stress management can be effective. Fear of childbirth can also be a clinically significant condition. At times, it may be so intense that a woman terminates a pregnancy, despite a desire to have children. Even in its milder form, this condition causes considerable emotional distress during pregnancy, and it is desirable to reduce such distress for the benefit of both mother and fetus. Again, referral to a trained therapist is indicated in such cases. No pharmacotherapy has proved effective, but cognitive-behavioral treatments are simple and beneficial.

The author has received grants for clinical research from Pfizer, Inc., and Forest Pharmaceuticals, Inc., and has served as a consultant for and member of the speakers' bureaus of Pfizer, Inc., and GlaxoSmithKline.

References

1. Kessler RC, McGonagle KA, Zhao S, et al: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. Arch Gen Psychiatry 51:8, 1994

2. Magee WJ, Eaton WW, Wittchen HU, et al: Agoraphobia, simple phobia, and social phobia in the National Comorbidity Survey. Arch Gen Psychiatry 53:159, 1996

3. Bourin M, Baker GB, Bradwejn J: Neurobiology of panic disorder. J Psychosom Res 44:163, 1998

4. Coplan JD, Lydiard RB: Brain circuits in panic disorder. Biol Psychiatry 44:1264, 1998

5. Sullivan GM, Coplan JD, Kent JM, et al: The noradrenergic system in pathological anxiety: a focus on panic with relevance to generalized anxiety disorder and phobias. Biol Psychiatry 46:1205, 1999

6. Abelson JL, Glitz D, Cameron OG, et al: Endocrine, cardiovascular, and behavioral responses to clonidine in patients with panic disorder. Biol Psychiatry 32:18, 1992

7. Malizia AL, Cunningham VJ, Bell CJ, et al: Decreased brain GABA(A)-benzodiazepine receptor binding in panic disorder: preliminary results from a quantitative PET study. Arch Gen Psychiatry 55:715, 1998

8. Marazziti D, Rotondo A, Martini C, et al: Changes in peripheral benzodiazepine receptors in patients with panic disorder and obsessive-compulsive disorder. Neuropsychobiology 29:8, 1994

9. Sinha S, Papp LA, Gorman JM: How study of respiratory physiology aided our understanding of abnormal brain function in panic disorder. J Affective Disord 61:191, 2000

10. Abelson JL, Weg JG, Nesse RM, et al: Persistent respiratory irregularity in patients with panic disorders. Biol Psychiatry 49:588, 2001

11. McCraty R, Atkinson M, Tomasino D, et al: Analysis of twenty-four hour heart rate variability in patients with panic disorder. Biol Psychol 56:131, 2001

12. Walker EA, Gelfand AN, Gelfand MD, et al: Psychiatric diagnoses, sexual and physical victimization, and disability in patients with irritable bowel syndrome or inflammatory bowel disease. Psychol Med 25:1259, 1995

13. Diagnostic and Statistical Manual of Mental Disorders (DSM IV). American Psychiatric Association, Washington, DC, 1994

14. Schmidt NB, Lerew DR, Jackson RJ: Prospective evaluation of anxiety sensitivity in the pathogenesis of panic: replication and extension. J Abnorm Psychol 108:532, 1999

15. Ballenger JC, Davidson JR, Lecrubier Y, et al: Consensus statement on panic disorder from the International Consensus Group on Depression and Anxiety. J Clin Psychiatry 59(suppl 8):47, 1998

16. Yingling KW, Wulsin LR, Arnold LM, et al: Estimated prevalences of panic disorder and depression among consecutive patients seen in an emergency department with acute chest pain. J Gen Intern Med 8:231, 1993

17. Zaubler TS, Katon W: Panic disorder in the general medical setting. J Psychosom Res 44:25, 1998

18. Kawachi I, Colditz GA, Ascherio A, et al: Prospective study of phobic anxiety and risk of coronary heart disease in men. Circulation 89:1992, 1994

19. Kahn JP, Drusin RE, Klein DF: Idiopathic cardiomyopathy and panic disorder: clinical association in cardiac transplant candidates. Am J Psychiatry 144:1327, 1987

20. Shear MK, Kligfield P, Harshfield G, et al: Cardiac rate and rhythm in panic patients. Am J Psychiatry 144:633, 1987

21. Godemann F, Ahrens B, Behrens S, et al: Classic conditioning and dysfunctional cognitions in patients with panic disorder and agoraphobia treated with an implantable cardioverter/defibrillator. Psychosom Med 63:231, 2001

22. Zandbergen J, Bright M, Pols H, et al: Higher lifetime prevalence of respiratory disease in panic disorder? Am J Psychiatry 148:1583, 1991

23. Mavissakalian MR, Perel JM: 2nd year maintenance and discontinuation of imipramine therapy in panic disorder with agoraphobia. Ann Clin Psychiatry 13:63, 2001

24. Otto MW, Tuby KS, Gould RA, et al: An effect-size analysis of the relative efficacy and tolerability of serotonin selective reuptake inhibitors for panic disorder. Am J Psychiatry 158:1989, 2001

25. Rapaport MH, Wolkow R, Rubin A, et al: Sertraline treatment of panic disorder: results of a long-term study. Acta Psychiatr Scand 104:289, 2001

26. Bruce TJ, Spiegel DA, Hegel MT: Cognitive-behavioral therapy helps prevent relapse and recurrence of panic disorder following alprazolam discontinuation: a long-term follow-up of the Peoria and Dartmouth studies. J Consult Clin Psychol 67:151, 1999

27. Barlow DH, Gorman JM, Shear MK, et al: Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. JAMA 283:2529, 2000

28. Dyckman JM, Rosenbaum RL, Hartmeyer RJ, et al: Effects of psychological intervention on panic attack patients in the emergency department. Psychosomatics 40:422, 1999

29. Leon AC, Portera L, Weissman MM: The social costs of anxiety disorders. Br J Psychiatry Suppl (27):19,1995

30. Kessler RC, Sonnega A, Bromet E, et al: Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 52:1048, 1995

31. Bremner JD, Licinio J, Darnell A, et al: Elevated CSF corticotropin-releasing factor concentrations in posttraumatic stress disorder. Am J Psychiatry 154:624, 1997

32. Gould E, Tanapat P: Stress and hippocampal neurogenesis. Biol Psychiatry 46:1472, 1999

33. Bremner JD, Randall P, Vermetten E, et al: Magnetic resonance imaging–based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. Biol Psychiatry 41:23, 1997

34. Yehuda R, Southwick S, Giller EL, et al: Urinary catecholamine excretion and severity of PTSD symptoms in Vietnam combat veterans. J Nerv Ment Dis 180:321, 1992

35. Rauch SL, Whalen PJ, Shin LM, et al: Exaggerated amygdala response to masked facial stimuli in posttraumatic stress disorder: a functional MRI study. Biol Psychiatry 47:769, 2000

36. Taubman-Ben-Ari O, Rabinowitz J, Feldman D, et al: Post-traumatic stress disorder in primary-care settings: prevalence and physicians' detection. Psychol Med 31:555, 2001

37. Dew MA, Kormos RL, DiMartini AF, et al: Prevalence and risk of depression and anxiety-related disorders during the first three years after heart transplantation. Psychosomatics 42:300, 2001

38. Mundy EA, Blanchard EB, Cirenza E, et al: Posttraumatic stress disorder in breast cancer patients following autologous bone marrow transplantation or conventional cancer treatments. Behav Res Ther 38:1015, 2000

39. Turton P, Hughes P, Evans CD, et al: Incidence, correlates and predictors of posttraumatic stress disorder in the pregnancy after stillbirth. Br J Psychiatry 178:556, 2001

40. Osterman JE, Hooper J, Heran WJ, et al: Awareness under anesthesia and the development of posttraumatic stress disorder. Gen Hosp Psychiatry 23:198, 2001

41. Foa EB, Dancu CV, Hembree EA, et al: A comparison of exposure therapy, stress inoculation training, and their combination for reducing post traumatic stress disorder in female assault victims. J Consult Clin Psychol 67:194, 1999

42. Zayfert C, Dums AR, Ferguson RJ, et al: Health functioning impairments associated with posttraumatic stress disorder, anxiety disorders, and depression. J Nerv Ment Dis 190:233, 2002

43. Shemesh E, Rudnick A, Kaluski E, et al: A prospective study of posttraumatic stress symptoms and nonadherence in survivors of a myocardial infarction (MI). Gen Hosp Psychiatry 23:215, 2001

44. Kessler RC, Keller MB, Wittchen HU: The epidemiology of generalized anxiety disorder. Psychiatr Clin North Am 24:19, 2001

45. Wittchen HU, Hoyer J: Generalized anxiety disorder: nature and course. J Clin Psychiatry 62(suppl 11):15, 2001

46. Thayer JF, Friedman BH, Borkovec TD: Autonomic characteristics of generalized anxiety disorder and worry. Biol Psychiatry 39:255, 1996

47. Hoehn-Saric R, Hazlett RL, Pourmotabbed T, et al: Does muscle tension reflect arousal? Relationship between electromyographic and electroencephalographic recordings. Psychiatry Res 71:49, 1997

48. Tiihonen J, Kuikka J, Rasanen P, et al: Cerebral benzodiazepine receptor binding and distribution in generalized anxiety disorder: a fractal anaylsis. Mol Psychiatry 2:463, 1997

49. Thayer JF, Lane RD: Perseverative thinking and health: neurovisceral concomitants. Psychol Health 17:685, 2002

50. Goldberg DP, Lecrubier Y: Mental Illness in General Health Care: An International Study. John Wiley & Sons, New York, 1995, p 323

51. Hidalgo RB, Davidson JR: Generalized anxiety disorder: an important clinical concern. Med Clin North Am 85:691, 2001

52. Newman MG: Recommendations for a cost-offset model of psychotherapy allocation using generalized anxiety disorder as an example. J Consult Clin Psychol 68:549, 2000

53. Logue MB, Thomas AM, Barbee JG, et al: Generalized anxiety disorder patients seek evaluation for cardiological symptoms at the same frequency as patients with panic disorder. J Psychiatric Res 27:55, 1993

54. Davidson JR: Pharmacotherapy of generalized anxiety disorder. J Clin Psychiatry 62(suppl 11):46, 2001

55. Rickels K, Schweizer E: Panic disorder: long-term pharmacotherapy and discontinuation. J Clin Psychopharmacol 18(6 suppl 2):12S, 1998

56. Borkovec TD, Ruscio AM: Psychotherapy for generalized anxiety disorder. J Clin Psychiatry 62(suppl 11):37, 2001

57. Stein MB, Torgrud LJ, Walker JR: Social phobia symptoms, subtypes, and severity: findings from a community survey. Arch Gen Psychiatry 57:1046, 2000

58. Yonkers KA, Dyck IR, Keller MB: An eight-year longitudinal comparison of clinical course and characteristics of social phobia among men and women. Psychiatric Services 52:637, 2001

59. Amaral DG: The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. Biol Psychiatry 51:11, 2002

60. Birbaumer N, Grodd W, Diedrich O, et al: fMRI reveals amygdala activation to human faces in social phobics. Neuroreport 9:1223, 1998

61. Foa EB, Gilboa-Schechtman E, Amir N, et al: Memory bias in generalized social phobia: remembering negative emotional expressions. J Anxiety Disord 14:501, 2000

62. Hope DA, Gansler DA, Heimberg RG: Attentional focus and causal attributions in social phobia: implications from social psychology. Clin Psychol Rev 9:49, 1989

63. Stein MB, McQuaid JR, Laffaye C, et al: Social phobia in the primary care medical setting. J Fam Pract 48:514, 1999

64. Olfson M, Guardino M, Struening E, et al: Barriers to the treatment of social anxiety. Am J Psychiatry 157:521, 2000

65. Ballenger JCL: Current treatment of the anxiety disorders in adults. Biol Psychiatry 46:1579, 1999

66. Heimberg RG: Cognitive-behavioral therapy for social anxiety disorder: current status and future directions. Bio Psychiatry 51:101, 2002

67. Shear MK, Beidel DC: Psychotherapy in the overall management strategy for social anxiety disorder. J Clin Psychiatry 59(suppl 17):39, 1998

68. Katzelnick DJ, Stein MB, Helstad CP: Impact of generalized social anxiety disorder in managed care. Psychosomatics 158:1999, 1999

69. Skoog G, Skoog I: A 40-year follow-up of patients with obsessive-compulsive disorder. Arch Gen Psychiatry 56:121, 1999

70. Micallef J, Blin O: Neurobiology and clinical pharmacology of obsessive-compulsive disorder. Clin Neuropharmacol 24:191, 2001

71. Swedo SE, Leonard HL, Garvey M, et al: Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections: clinical description of the first 50 cases. Am J Psychiatry 155:264, 1998

72. Rasmussen SA, Eisen JL: The epidemiology and differential diagnosis of obsessive compulsive disorder. J Clin Psychiatry 55(suppl):5, 1994

73. Folks DG, Warnock JK: Psychocutaneous disorders. Curr Psychiatry Rep 3:219, 2001

74. Mundo E, Bianchi L, Bellodi L: Efficacy of fluvoxamine, paroxetine, and citalopram in the treatment of obsessive-compulsive disorder: a single-blind study. J Clin Psychopharmacol 17:267, 1997

75. Kozak MJ, Liebowitz MR, Foa EB: Obsessive-Compulsive Disorder: Contemporary Issues in Treatment. Goodman WK, Rudorfer MV, Maser JD, Eds. Lawrence Erlbaum Assoc., Inc. Mahwah, New Jersey, 2002, p 501

76. Koran LM, Thienemann ML, Davenport R: Quality of life for patients with obsessive-compulsive disorder. Am J Psychiatry 153:783, 1996

IX THE EATING DISORDERS

W. Stewart Agras, m.d.

The eating disorders—anorexia nervosa, bulimia nervosa, and binge eating—tend to be chronic conditions and are accompanied by an increased risk of psychiatric illness. In addition, serious medical problems are associated with anorexia nervosa because of chronic starvation and with binge-eating disorder because of obesity. Eating disorders may also be associated with marked life impairment, faulty interpersonal interactions, and social withdrawal.

Overview of the Eating Disorders

EPIDEMIOLOGY

In women, the lifetime prevalence of eating disorders meeting full diagnostic criteria is about 4%. About 10% of cases of eating disorders occur in males.[1] Bulimia nervosa and binge-eating disorder, either of which can occur as a full or partial syndrome, are the most commonly encountered eating disorders in clinical practice. Anorexia nervosa is seen less frequently. Evidence suggests that these disorders, particularly anorexia nervosa, may be increasing in prevalence.[2] If subclinical cases are included, the overall prevalence of eating disorders in the female population is probably between 6% and 8%. In adolescents, the prevalence of subclinical eating disorders may be substantially higher.[3]

Treatment for the eating disorders is expensive. A study using data from a large insurance database[2] found that the average annual treatment costs per patient were $6,045 for anorexia nervosa, $2,962 for bulimia nervosa, and $3,207 for an eating disorder not otherwise specified (EDNOS).[4] It is likely that most cases of EDNOS were in patients with binge-eating disorder. Furthermore, the expense of treating chronic eating disorders increases as the frequency of hospitalization increases—thus the need for early diagnosis and treatment.

MANAGEMENT

Outpatient Management

Most patients with an eating disorder are treated as outpatients [*see Figure 1*]. Treatment involves both the primary care physician and a mental health professional. In difficult cases, such as a patient with anorexia nervosa or a diabetic adolescent with bulimia nervosa, a close working relationship must be established between the treatment professionals. It is important to select a mental health professional with experience and expertise in the treatment of eating disorders. When the diagnosis of an eating disorder is made, the primary care physician should provide the patient with accurate information concerning the nature, complications, and treatment of the disorder; it may also be appropriate to give this information to the patient's family members. In the adolescent, early recognition and management of anorexia nervosa is particularly important because prompt treatment may prevent the disorder from becoming chronic. The primary care physician may offer a self-help manual, an antidepressant, or both for some patients with bulimia nervosa and binge-eating disorder. Should such treatment fail, the patient should be referred to a mental health professional for treatment.

Inpatient Management

The principal criterion for hospitalization of a patient with an eating disorder is weight less than 75% of ideal body weight; other criteria include dehydration, vascular instability, hypothermia, and, in the anorectic who purges and the patient with bulimia nervosa, depletion of potassium and other electrolytes. Thus, the medical treatment of anorexia nervosa includes rehydration and correction of electrolyte disturbances. The patient should be examined frequently, and the levels of electrolytes (sodium, potassium, chloride, bicarbonate, phosphorus, and magnesium) should be monitored frequently. It is important to warn the patient that temporary edema may occur because of rehydration and electrolyte correction; this information helps allay the patient's concerns about weight and shape. Congestive heart failure occasionally occurs early in the refeeding period; hence, careful evaluation of cardiovascular status during the early stages of refeeding is necessary.[5] Hypophosphatemia may also develop at this time, when reserves are depleted.[6] Inpatient care is best provided in specialized units and should be administered by professionals who have had considerable experience in the treatment of anorexia nervosa. After gaining some weight, the patient may best be treated in a partial hospitalization program aimed at restoration of weight to at least 90% of ideal body weight. Controlled studies indicate that outpatient psychotherapy may be as effective as inpatient treatment for anorectic patients,[7] which suggests that outpatient therapy for anorexia nervosa may be underutilized. Thus, for patients who weigh more than 75% of their ideal body weight and who do not have significant medical complications, outpatient treatment should be attempted before hospitalization.

Pregnancy in Patients with an Eating Disorder

Patients with anorexia and bulimia nervosa who become pregnant during the course of their illness should receive more intensive prenatal care because of increased risk of low-birthweight infants and other perinatal complications.[8] Women who have eating disorders may have difficulty feeding their children because of concerns about their children's weight and shape,[9,10] and inadequate feeding may result in lagging growth and development. Careful monitoring of the nutritional status and development of such children is necessary.

The National Institute for Health and Clinical Excellence (NICE) provides an excellent guide to the management of the eating disorders.[11]

Anorexia Nervosa

Anorexia nervosa is characterized by a reduction in ideal body weight of at least 15%, accompanied in females by amenorrhea. It is also characterized by an intense fear of gaining weight, a preoccupation with food, obsessions and rituals concerning food, a distorted body image (the patient feels fat despite evidence to the contrary), and excessive exercise. Marked and unexplained weight loss in an adolescent female should alert the physician to the possibility of anorexia nervosa.

Long-term follow-up studies have revealed that anorexia nervosa runs a protracted course; it has a mortality of 5% to 15%, depending on the length of follow-up.[12] About half the

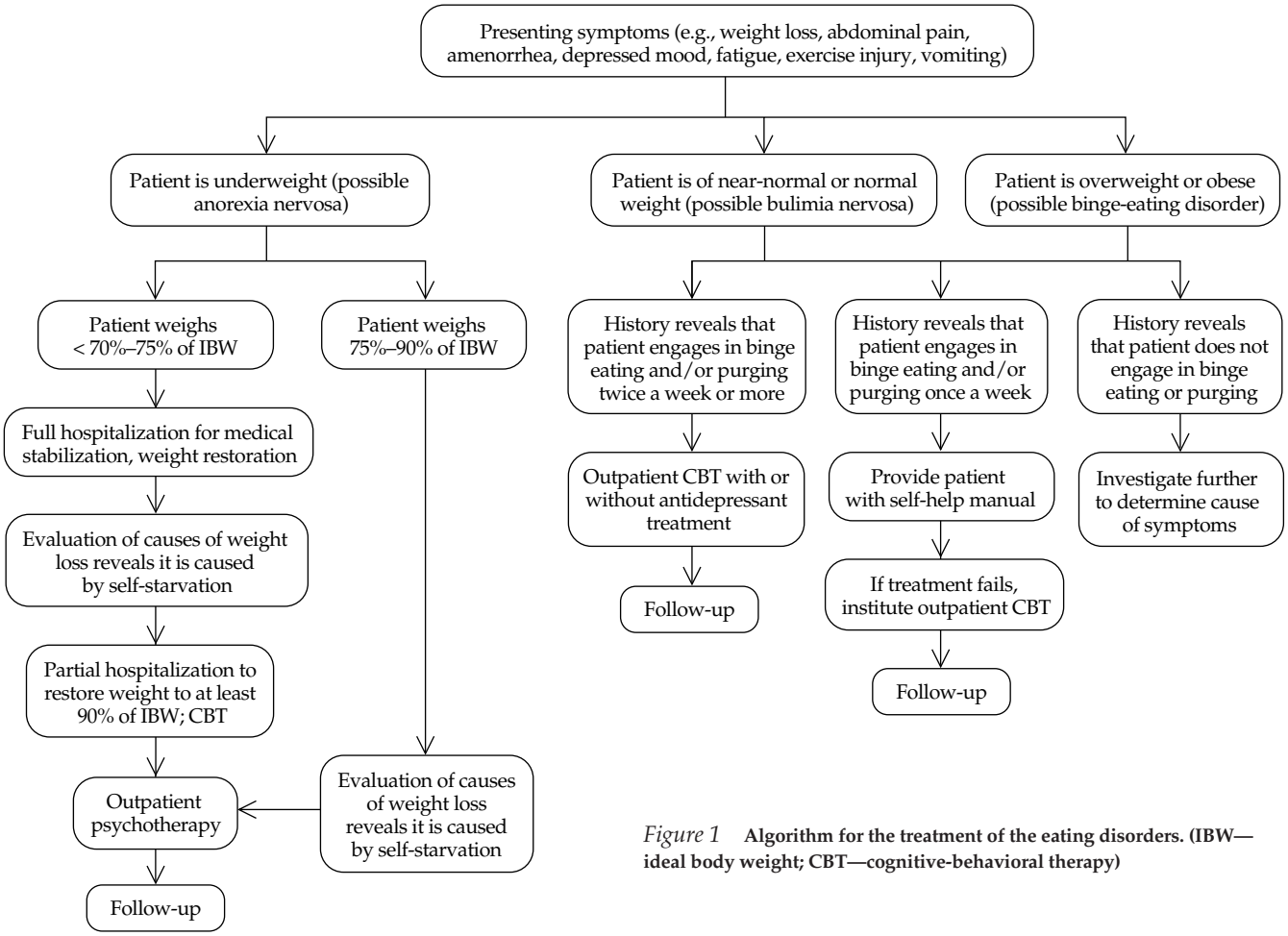


Figure 1 **Algorithm for the treatment of the eating disorders. (IBW—ideal body weight; CBT—cognitive-behavioral therapy)**

deaths are suicides associated with major depression; the other deaths result from medical complications associated with malnutrition. Both mental and physical status should be monitored throughout the course of the illness. Between 40% and 50% of anorectics recover within 10 years; the others continue to have an eating disorder, which is most frequently low-weight clinical or subclinical bulimia nervosa.[13]

EPIDEMIOLOGY

Anorexia nervosa is a relatively rare disorder, with an incidence of five to 10 cases per 100,000 population and a lifetime prevalence of 0.25% to 0.50% in the female population.[5] The disorder may begin in childhood but most frequently emerges in adolescence, with fewer cases beginning in adulthood.

ETIOLOGY/GENETICS

Family studies of anorexia nervosa have revealed that the disorder occurs more frequently in relatives of affected persons than in relatives of persons who do not have an eating disorder.[14] This finding suggests that genetic transmission, familial influences, or both play a role in etiology. The case for genetic transmission is strengthened by twin studies, which have shown that an eating disorder is more likely to develop in an identical twin than in a nonidentical twin of a person with anorexia nervosa.[15] Moreover, these studies suggest that heredity accounts for as much as 50% of the causal factors of anorexia nervosa. Family influences also undoubtedly play a part in the etiology of anorexia nervosa, although few adequate risk-factor studies exist for anorexia nervosa. Risk factors for the development of anorexia nervosa include picky eating in childhood and perfectionism. There also appears to be an overlap between obsessive-compulsive disorder and anorexia nervosa.

PATHOGENESIS

Disturbances of serotonin (5-hydroxytryptamine [5-HT]) pathways probably play a role in the pathogenesis of anorexia nervosa.[16] Serotonin pathways contribute to the modulation of a number of behaviors commonly associated with this disorder, such as behavioral inhibition, obsessionality, anxiety and fear, depression, and appetite regulation. Many studies show disturbances of 5-HT activity in persons who currently have or who have recovered from anorexia nervosa. In addition, cerebrospinal fluid concentrations of 5-hydroxyindoleacetic acid (5-HIAA), a metabolite of 5-HT that tends to inhibit eating, have been shown to be elevated in anorectics in long-term recovery. Accumulating evidence indicates that serotonin abnormalities affect the development of anorexia nervosa; however, appetite regulation is extremely complex, and it is unlikely that only the serotoninergic system would be implicated.

Genetic findings, although preliminary, also confirm abnormalities in serotonin functioning in patients with anorexia nervosa. The strongest genetic finding to date involves a functional polymorphism in the promoter region of the human 5-HT transporter (*5-HTT*) gene.[17] Several lines of evidence suggest

that the genotype that has two copies of the long allele has greater 5-HT reuptake than the genotypes that have either one or two copies of the short allele.[18] Because *5-HTT* regulates 5-HT concentrations by recycling released 5-HT, persons with the short allele are likely to have higher extracellular 5-HT concentrations. A meta-analysis suggested that the short allele appeared to increase the risk for anorexia nervosa.[19]

DIAGNOSIS

Clinical Manifestations

The primary symptoms of anorexia nervosa are severe dietary restriction and marked weight loss, which often occur after a period of mild weight gain. Sometimes, dietary restriction is triggered by negative comments concerning the weight gain. In other cases, the disorder emerges following a stressful life event. Amenorrhea is often seen soon after the onset of the disorder. Distortion of body-image perception is another signal feature: the patient is emaciated but believes that certain parts of the body (e.g., the stomach, thighs, and upper arms) are fat. Excessive exercise is also common. The patient often develops obsessive thinking regarding food and may have many irrational food rules (e.g., eat only green foods or always leave food on the plate). The patient may also perform rituals during food consumption (e.g., chewing food a certain number of times). All patients restrict their food intake; however, half of the patients also binge and purge, although the binges may be quite small. Depressive symptoms are common. A relatively high percentage of patients have an associated anxiety disorder, often a social phobia, or obsessive-compulsive disorder.

Physical Examination

Physical examination of a patient with anorexia nervosa reveals marked emaciation; dry skin that may be covered with fine body hair (lanugo); and cold, cyanotic extremities sometimes accompanied by peripheral edema, bradycardia, and hypotension. Pubic and axillary hair tend to be preserved. Bone density is reduced as a result of the poor nutrition and low estrogen levels that accompany the disorder. Over half of all patients with anorexia nervosa suffer from osteoporosis, and the risk of fractures is increased.[20] Gastric emptying is often slowed, leading to complaints of gastric distress after eating.[21] The starvation may impair cognitive function, leading to poor concentration, memory, and judgment. Cognitive impairment may not be evident on examination but may be demonstrated on psychological testing.

Laboratory Tests

Usually, few abnormalities are noted in electrolyte, enzyme, or protein levels, although potassium levels may be low in anorectics who are also bulimic, and hepatic function may occasionally be impaired. Both anemia and leukopenia are commonly present,[22] and cholesterol and β-carotene levels may be elevated. Levels of various hormones are abnormal, suggesting hypothalamic dysfunction. Plasma gonadotropin levels are significantly decreased, and the 24-hour luteinizing hormone pattern resembles that found in prepubertal girls. Follicle-stimulating hormone and estrogen levels are also low, and growth hormone levels are elevated. Total serum thyroxine and triiodothyronine levels are usually low, but the free thyroxine level is usually normal. Resting plasma cortisol levels are in the high-normal or elevated ranges, and the diurnal secretion pattern is diminished. These hormonal abnormalities are caused by starvation and are also found in other syndromes characterized by malnutrition and significant weight loss.[23]

DIFFERENTIAL DIAGNOSIS

The diagnosis of anorexia nervosa is usually straightforward because the presentation, history, and physical findings are typical of the condition. Other causes of marked weight loss in adolescence include malignancy, malabsorption syndromes, diabetes mellitus, and hyperthyroidism. The differential diagnosis in cases in which weight loss has not yet reached the diagnostic criterion (15% less than ideal body weight) may pose a problem, because amenorrhea may occur early in the disorder. In such cases, hypopituitarism may have to be ruled out, although a careful elicitation of the history usually reveals the patient's dietary restriction, preoccupation with weight and shape, and perhaps binge eating and purging. Occasionally, a patient with schizophrenia has delusions that center on food consumption (e.g., the delusion of being poisoned) and may lose considerable weight. However, the absence of weight and shape concerns and the presence of delusions usually make the diagnosis evident. Major depression accompanied by marked weight loss may also mimic anorexia nervosa, but the other features of anorexia nervosa are not present.

MANAGEMENT

A major problem in treating anorectic patients is their denial of the severity of their condition and refusal to cooperate with the treatment regimen. Because uncooperative patients will gain little from treatment, it is worth spending considerable time with patients and their families to at least partly overcome this resistance before referral for specialty care. In general, the earlier treatment begins after onset of the disorder, the more likely it is that the patient will benefit. This fact underlines the importance of early recognition and treatment of the syndrome during childhood and adolescence. Once the disorder enters the chronic phase, treatment is less likely to succeed, and rehabilitation and quality-of-life improvement become treatment goals.

Family Therapy

Small, controlled trials suggest that family therapy specifically aimed at the eating disorder should be the treatment of choice for the adolescent anorexic.[24] About two thirds of patients respond well to this form of outpatient therapy; thus, the costs associated with hospitalization are reduced. Therapy essentially consists of helping the parents take back control of feeding from the anorectic patient and initiate a program of refeeding. In later sessions, more general family problems relating to the anorectic patient are addressed.

Cognitive-Behavioral Therapy

Cognitive-behavioral therapy for anorexia nervosa, usually targeted at adult patients, has emerged as a treatment strategy. This psychotherapeutic approach focuses on giving nutritional advice, monitoring the effects of that advice, and counseling patients to overcome cognitive distortions and avoidant patterns present in their interpersonal relationships. No adequately controlled trials address the efficacy of this approach. Treatment of adults, who usually are in the chronic phase of anorexia nervosa, is difficult and may require long-term psychotherapy and rehabilitation. Such patients should be followed medically at regular intervals.

Pharmacologic Therapy

Pharmacologic therapy may be considered as adjunctive to family therapy or cognitive-behavorial therapy in patients with chronic anorexia nervosa; it should not be the sole treatment offered to the patient. The main use of pharmacologic agents in chronic anorexia nervosa is to treat associated psychopathologies, such as panic disorder, social phobia, and obsessive-compulsive disorder.

Various psychopharmacologic approaches to treatment of anorectic patients have been evaluated in controlled trials and have included the use of tricyclic antidepressants, fluoxetine, lithium, and antipsychotic compounds. There are various rationales for the use of these compounds—for instance, the observation that patients who take antipsychotic agents such as chlorpromazine tend to gain weight and the finding that anorectics often suffer from depressive symptoms. The results of these trials have been disappointing: no evidence of clinically useful effectiveness for any of these medications in the acute phase of treatment has been established.[25] Moreover, patients with anorexia nervosa are often reluctant to take medication, particularly if the medication might lead to weight gain. Symptoms such as depression often disappear as weight is gained.

Because obsessive-compulsive disorder is associated with anorexia nervosa and because the genetic findings implicate serotonin metabolism, interest has focused on the use of the selective serotonin reuptake inhibitors (SSRIs) that are effective in patients with obsessive-compulsive disorder. Small controlled and uncontrolled trials suggest that fluoxetine may be useful in helping patients maintain their weight gain.[26] Open-label trials also suggest that the newer antipsychotics, such as olanzepine, may be effective in reducing depression, anxiety, and core eating-disorder symptoms (e.g., food obsessions and distorted body image) and in enhancing weight gain.[27]

Hospitalization

The principal aim of hospitalization of the patient with anorexia nervosa is to enhance weight gain and attain physiologic stability. The refeeding syndrome is of particular concern in the early phase of hospitalization in severely malnourished patients. This syndrome involves major shifts in fluid and electrolyte levels, particularly phosphorus and potassium. The syndrome can be prevented by slow refeeding and daily monitoring of serum phosphorus and potassium levels for 7 to 10 days after admission to hospital. Studies conducted in hospital settings have shown that when caloric intake and promotion of weight gain are accompanied by some form of reinforcement, weight gain accelerates. Reinforcement strategies may be as simple as allowing the patient to leave the hospital room, to watch television, and later to leave the hospital for short periods. Such privileges are contingent on small and continued weight gains. When the patient's condition has stabilized, either partial hospitalization (i.e., day hospitalization) or outpatient treatment can begin. However, the chance of relapse is increased if the patient is discharged with a body mass index of less than 19.

Medical Complications

Cardiovascular complications Among the more serious medical complications seen in patients with anorexia nervosa are those that affect the cardiovascular system.[23] Electrocardiography typically shows a low-voltage pattern and bradycardia. However, arrhythmias, including ventricular arrhythmias, may be caused by both starvation and hypokalemia and are implicated in the sudden deaths that occasionally occur in low-weight anorectics. Cardiomyopathy secondary to starvation and to the abuse of ipecac (used to induce vomiting) may also be seen.

Renal complications Long-term dehydration accompanied by hypokalemia may cause renal complications, including irreversible damage to renal tubules. Renal calculi have also been reported to occur as a consequence of dehydration.

Osteopenia and osteoporosis Osteopenia and osteoporosis are seen in most cases of anorexia nervosa. One cross-sectional community-based study of women with anorexia nervosa found that 51.7% had osteopenia and 34.6% had osteoporosis.[22] The treatment of this important problem is uncertain. The best evidence to date suggests that adequate nutrition and weight gain, together with calcium intake of 1,500 mg/day, and vitamin D intake of 400 IU/day will result in increased bone density, but full recovery is uncommon. Hormone replacement therapy (estrogen/progestin combination) has not been shown to be helpful in patients with anorexia nervosa in a number of controlled clinical trials.[20]

PROGNOSIS

Anorexia nervosa has one of the highest mortality rates of any psychiatric disorder. In long-term follow-up studies, fewer than 50% of patients with anorexia nervosa have been shown to have recovered, and in patients who have recovered, social and work functioning are often poor. Recovered patients often suffer from anxiety and depression. The shorter the duration of the illness and the higher the weight at which the patient is initially treated, the better the prognosis.

Bulimia Nervosa

Bulimia nervosa is characterized by binge eating followed by compensatory purging. Binge eating consists of a sense of loss of control over eating and usually the consumption of large amounts of food that is typically high in simple carbohydrates and fat. Purging usually consists of self-induced vomiting and abuse of laxatives; less commonly, purging involves the use of diuretics or enemas or the practice of chewing food and spitting it out. Occasionally, patients with bulimia nervosa engage in episodes of fasting that last for several days. Bulimics also frequently engage in excessive exercise to compensate for binge eating. Concerns about weight and shape frequently dominate the thinking of these patients.

EPIDEMIOLOGY

Bulimia nervosa has a lifetime prevalence of 1.0% to 1.5% in females. The disorder rarely affects males.[1]

Cultural pressure for thinness, engendered by media, peer, or parental pressure, has increased in the United States over the past 25 years. Research has shown that concerns about weight and physique cross gender and racial and ethnic boundaries. The results of a comprehensive self-report health survey completed by 17,571 adolescent girls suggest that no relation exists between socioeconomic status and eating disorders.[28]

ETIOLOGY/GENETICS

Prospective studies have delineated several risk factors for the development of bulimic symptoms. Studies of adolescent girls have found that weight concerns and thin-body preoccupation,

together with social pressure to achieve a thinner body, strongly predict the onset of bulimic symptoms.[29] In addition, higher body weight gives rise to body dissatisfaction and to dieting. Negative affect (depression and anxiety) is also a risk factor for the development of bulimic symptoms. Physicians should be alert for the presence of such behaviors, particularly in female adolescents and preadolescents, and should monitor such patients for the development of binge eating and purging.

Family studies reveal that the relatives of probands with pure forms of anorexia nervosa or bulimia nervosa have an equal chance of having either disorder, suggesting that there may be a common genetic predisposition to both anorexia nervosa and bulimia nervosa.[24] Evidence demonstrates a partial overlap in the genetic predisposition to eating disorders and major depression, suggesting that a common genetic factor may increase the risk of both illnesses.[11] Family influences also appear to be important.

DIAGNOSIS

Clinical Manifestations

On presentation, patients with bulimia nervosa frequently complain of fatigue, decreased energy, abdominal pain, amenorrhea, recurrent vomiting (without revealing that it is self-induced), and depression; less frequently, they are found to spit up blood from a small esophageal tear. Patients may also have joint pain or other trauma after excessive exercise. The diagnosis of bulimia nervosa should also be considered when routine laboratory tests reveal an unexpectedly low serum potassium level, when there is unexplained enlargement of the parotid or submaxillary salivary glands, or when the patient has tooth decay and periodontal disease, which can result from bingeing on sweets and from the presence of acid in the mouth after purging.

Patients who meet the minimal clinical criteria for the diagnosis of bulimia nervosa (purging type) engage in binge eating and purging on an average of two or more times a week for 3 months. Alternatively, patients may not purge (nonpurging type) but may engage in compensatory behaviors such as fasting or excessive exercise.

Patients who do not meet the clinical criteria for a diagnosis of a full-fledged eating disorder or who have an atypical presentation are classified as having an EDNOS. However, evidence suggests that partial eating disorders and full-fledged eating disorders cause similar levels of disability and have similar related psychopathologies; therefore, partial cases should also be considered for referral for treatment.

Once the diagnosis of bulimia nervosa has been made, it is important to review the patient's history and current mental status for the presence of comorbid psychopathology. Over half of all patients with bulimia nervosa have a history of major depression, and about 20% have major depression at the time of diagnosis. Anxiety disorders are also fairly common. About one third of patients also have a history of alcohol or drug dependency or abuse, and a fairly high percentage are alcohol or drug dependent at the time of diagnosis.

Patient History

Bulimia nervosa usually begins in adolescence. Like patients with anorexia nervosa, patients with bulimia nervosa typically have a history of dieting after a modest weight gain; however, in bulimics, the weight loss associated with dieting is relatively small, and dieting is followed by binge eating and purging or other compensatory behaviors. The size of binges varies from small (no more than 100 kcal) to very large (several thousand kilocalories); the key feature of a binge is the sense of loss of control over eating, not the amount of food consumed. The average binge consists of about 1,500 kcal (i.e., about the size of two meals).

Binges may be precipitated by dietary restriction; however, the most commonly reported trigger is a negative emotion arising from interpersonal problems. During the binge, negative emotions may decline; as the binge progresses, however, guilt feelings increase. Although the weight of the typical bulimic is close to normal, some patients with the disorder are nearly anorectic; others are overweight and even obese. Many bulimics engage in excessive exercise and exhibit concerns about weight and shape. However, these concerns are not as extreme as the distortions of body image perceived by the anorectic.

Physical Examination

Making the diagnosis of bulimia nervosa can be difficult because patients tend to hide their symptoms from their families, friends, and physicians. Such hiding of symptoms results from the shame that these patients feel. It is essential for the clinician to establish a good relationship with the patient who is suspected of having an eating disorder and to take a detailed history using gentle, open-ended questions. Patients with eating disorders have been found to visit their family practitioner more frequently than other patients during the years before the eating disorders are diagnosed. It is important to remember that not all eating disorders are found in thin, young women. Approximately 15% of bulimics are overweight, and many are middle aged. Simple questions such as "Do you think you have an eating problem?" or "Do you worry excessively about your weight?" may be useful for general screening.[11] Although it may be difficult to obtain, a detailed history of food intake is helpful in making the diagnosis. This history can be obtained by determining the patient's food intake during the past 24 hours and then ascertaining how typical such intake is for the patient, noting differences between weekdays and weekends. The bulimic patient tends to restrict food intake early in the day but gradually loses control of eating, which leads to binge eating later in the day and in the evening.

The only findings on physical examination that suggest the presence of bulimia nervosa are dry skin, evidence of dehydration, and fluctuant hypertrophy of the salivary glands, which is seen fairly frequently. Resting bradycardia may also be present.

Laboratory Tests

The most common laboratory finding is an elevated serum amylase level that is usually of parotid origin. Metabolic alkalosis secondary to purging is also frequently found, with raised serum bicarbonate levels, hypochloremia, and, occasionally, hypokalemia, although this last disorder occurs in only 5% of patients with bulimia nervosa.

DIFFERENTIAL DIAGNOSIS

There are few conditions with symptoms similar to those of bulimia nervosa, although patients with spontaneous vomiting do need to be distinguished from those with self-induced vomiting.

MANAGEMENT

Psychotherapy

It is generally agreed that the most effective treatment for bulimia nervosa is an outpatient trial of cognitive-behavioral

therapy. A large number of controlled trials have demonstrated that such therapy is more effective than no treatment, other forms of psychotherapy, antidepressant medication, and placebo.[30,31] The usual course of treatment is 20 sessions over 6 months. Approximately 50% of patients who complete treatment with cognitive-behavioral therapy recover, and an additional 20% to 30% show significant improvement. A 5-year follow-up study of patients treated with cognitive-behavioral therapy found that the gains made during therapy are well maintained.[32]

Cognitive-behavioral therapy Cognitive-behavioral therapy is aimed at reversing the dietary restriction underlying the binge eating and purging. In the first phase of therapy, the patient monitors dietary intake, binge eating, and purging, and the therapist helps the patient move toward a pattern of eating three meals and two snacks daily, gradually increasing the amount and variety of foods that were previously avoided. In the second phase, attention shifts to overcoming the patient's distorted thinking about weight, shape, and food rules. Finally, attention is directed toward relapse prevention. These principles have recently been incorporated into self-help manuals.[33] Such manuals, combined with brief cognitive-behavioral therapy, have been helpful for about 25% of patients in several controlled trials.[31] It is unclear, however, whether brief cognitive-behavioral therapy, with or without pharmacologic therapy, is useful in nonspecialty settings. One controlled trial found very high rates of patient dropout and concluded that brief cognitive-behavioral therapy was ineffective.[34] A British study suggested that brief cognitive-behavioral therapy was as effective as specialist care; however, in that study, the number of visits to both the family practitioner and the psychiatrist or psychologist (4.9 and 4.8, respectively) were inadequate, and the improvement in bulimic symptoms was small.[35]

Interpersonal psychotherapy Interpersonal psychotherapy appears to be useful in the treatment of bulimia nervosa, although the beneficial effects of interpersonal psychotherapy are slower to manifest than those of cognitive-behavioral therapy. One large clinical trial found that cognitive-behavioral therapy was more effective than interpersonal therapy at the end of 20 weeks of treatment[36]; however, no differences between treatments were found at the 1-year follow-up. Interpersonal psychotherapy focuses directly on current interpersonal problems and may be effective because negative affect associated with faulty interpersonal interactions are reduced.[8]

Pharmacologic Therapy

Both the tricyclic antidepressants (e.g., desipramine, 150 to 300 mg/day) and the SSRIs (e.g., fluoxetine, 60 mg/day) have been shown to be effective in the treatment of bulimia nervosa.[37] However, only 20% to 30% of patients who complete a trial of medication stop binge eating and purging. A controlled trial used a sequence of two medications (i.e., desipramine followed by fluoxetine).[38] When desipramine did not work, it was discontinued and fluoxetine begun. This sequential use of medication was as effective as cognitive-behavioral therapy. Another controlled study found that fluoxetine was more effective than placebo in the treatment of patients who failed to respond to cognitive-behavioral therapy or who experienced relapses.[39] The anticonvulsant topiramate in doses up to 400 mg also appeared to be effective in reducing binge eating and purging.[40]

Treatment with antidepressants appears to be the most cost-effective approach (i.e., cost per recovered patient).[41] It may be appropriate to begin treatment with antidepressant medication, supplemented by the use of a self-help manual, and reserve cognitive-behavioral therapy for those patients who fail to improve. Medication should probably be continued for at least 6 months; it is important to taper medication gradually and monitor for symptom recurrence. If symptoms recur, medication should be restarted.

Medical treatment also includes potassium supplementation for patients with low serum potassium levels and the correction of laxative abuse. Potassium supplementation is important because as with the patient with anorexia nervosa, the low-weight bulimic patient with low potassium levels is at risk for the development of fatal arrhythmias. The best way to treat patients who abuse laxatives is to persuade the patient to abruptly stop using the laxatives or to replace stimulant laxatives with bulk laxatives; however, education and support must be provided to the patient with regard to the pain and discomfort that inevitably will occur as the gut readjusts to normal function.

Medical Complications

Dental complications Dental complications frequently occur and include erosion of enamel, which sometimes causes fillings to protrude, and periodontal disease; thus, patients should be advised not to brush their teeth after purging, to use a nonacidic mouthwash, and to seek regular dental care.[42] There are few effects on other major organ systems. Rarely seen complications are esophageal tears caused by purging and a ruptured stomach and acute dilatation of the stomach, both of which occur after very large binges. Abuse of stimulant laxatives may lead to chronic physiologic dependence on them. About one quarter of bulimics have had an episode of anorexia nervosa. In such patients, osteopenia or osteoporosis may be present and may require treatment.

Diabetes mellitus Subclinical eating disorders or bulimia nervosa may occur in a substantial number of adolescents with type 1 diabetes mellitus.[43] Adolescents with diabetes often lose weight before their diabetes is diagnosed. When treatment begins, they often gain weight rapidly. This weight gain may heighten concerns about weight and shape in adolescent females, leading to dieting and then to binge eating and purging. Unfortunately, modern intensive management of diabetes with restriction of carbohydrates, consistent eating habits, and careful titration of insulin may, because of the restrictions involved, pose a risk of an eating disorder. Adolescent diabetics often discover that they can decrease caloric intake by omitting insulin doses, despite awareness about the importance of compliance with the medical regimen. The presence of bulimia nervosa or a subclinical variant should be suspected in female adolescent diabetics who have poorly controlled diabetes. In one study, 29% of adolescent diabetics had highly or moderately disturbed eating patterns.[44] Hemoglobin A_{1c} levels were significantly higher in patients with disturbed eating behavior than in patients with normal eating patterns. At follow-up 4 years later, patients with disturbed eating patterns at baseline were more likely to demonstrate retinopathy. To prevent such early complications of diabetes, it is important to obtain a thorough history of dieting and eating habits from all adolescent female patients who are diabetic.

PROGNOSIS

The prognosis for patients with bulimia nervosa is more favorable than that for patients with anorexia nervosa; in long-term follow-up studies of patients with bulimia nervosa, approximately 70% have been shown to have recovered. After recovery, however, many patients have residual psychopathology, such as anxiety and depression. Unexpected deaths may occur in patients with bulimia nervosa, usually in the low-weight bulimic who has low blood potassium levels and cardiac arrhythmia.

Binge-Eating Disorder

Binge-eating disorder is characterized by episodes of binge eating without compensatory behaviors, such as purging, and it is often accompanied by overweight or obesity.

EPIDEMIOLOGY

Epidemiologic studies of binge-eating disorder suggest a lifetime prevalence of 2% for females,[45] and some studies suggest that the ratio of females to males with binge-eating disorder is 3 to 2, making it the only eating disorder with a substantial proportion of males. The prevalence of binge-eating disorder in the obese population is between 25% and 33%. The majority of patients with binge-eating disorder are overweight or obese; however, in adolescents and young adults, weight may be within normal limits.

ETIOLOGY

Little is known about the etiology of binge-eating disorder, although the similarity to bulimia nervosa would suggest that it shares many of the same risk factors.

DIAGNOSIS

Clinical Manifestations

Patients who meet the minimal clinical criteria for the diagnosis of binge-eating disorder engage in binge eating at least 2 days a week for 6 months. Whereas in bulimia nervosa binge eating is measured by the number of episodes, in binge-eating disorder it is measured by the number of days in which binge eating occurs, because the absence of purging in patients with binge-eating disorder makes it more difficult to delineate episodes accurately. Although patients with binge-eating disorder are often overweight, they exhibit greater dissatisfaction with their weight and shape than do overweight persons who do not binge.

Patient History

The diagnosis of binge-eating disorder should be considered for all overweight patients, particularly those with continued sizable weight gain or marked weight cycling. Binge-eating disorder usually begins in late adolescence. As with bulimic patients, patients who have binge-eating disorder have a history of weight gain followed by dieting and the onset of binge eating. Many patients report having attempted purging early in the course of the disorder but found such behavior repulsive or were unable to perform it easily or regularly. The disorder appears to be chronic and is usually accompanied by an inexorable weight gain. Patients do not usually seek treatment for binge-eating disorder; they are often diagnosed later in life when they enter a weight-loss program.

Binge-eating disorder is characterized by periods of dietary restriction that alternate with binges. The periods of dietary restriction are less striking than those of bulimic patients, and the binges are less clearly defined; one binge often merges into another. Although binges vary considerably in size, they are smaller than those seen in patients with bulimia nervosa, (1,000 kcal for binge eating, versus 1,500 kcal for bulimia nervosa). Like anorexia nervosa and bulimia nervosa, binge-eating disorder is often accompanied by major depression, anxiety disorders, and personality disorders. Because these conditions are more commonly found in patients with binge-eating disorder than in weight-matched obese persons without binge-eating disorder,[46] it appears that the psychopathology is related to the eating disorder, not to excess weight or obesity.

DIFFERENTIAL DIAGNOSIS

Apart from rare cases of brain injury associated with binge eating, there are few conditions with similar presentations.

MANAGEMENT

Psychotherapy

As with bulimia nervosa, controlled studies of binge-eating disorders suggest that cognitive-behavioral therapy is an effective approach.[46] However, cognitive-behavioral therapy has little effect on weight; therefore, concurrent treatment of overweight or obesity should be considered. The use of a self-help manual covering cognitive-behavioral principles, with brief therapist supervision, is also effective in reducing binge eating.[31] Controlled studies have also found that interpersonal psychotherapy is as effective as cognitive-behavioral therapy, both at the end of treatment and at follow-up, in the treatment of binge-eating disorder.

Because weight-loss treatment entails dietary restriction, some researchers have questioned whether such treatment leads to increased binge eating. Studies have provided little evidence to support this hypothesis. Indeed, studies have shown that weight-loss treatment reduces the rate of binge eating.[47] This reduction may occur because the chaotic eating habits exhibited by patients with binge-eating disorder are replaced by more regular eating habits once they enter weight-control programs.

Pharmacologic Therapy

As with bulimia nervosa, antidepressant medication (e.g., fluoxetine, 60 mg/day) is useful in the treatment of binge eating.[48] Although there is no evidence demonstrating that antidepressants add to the effectiveness of cognitive-behavioral therapy in alleviating the eating disorder, it has been shown that antidepressants with cognitive-behavioral therapy may lead to superior weight losses. Therefore, combining psychological and psychopharmacologic treatments should be considered. Topiramate (up to 600 mg/day) has been shown to reduce binge eating and is also associated with substantial weight loss. In one longer-term, controlled trial, participants on topiramate averaged a weight loss of 15 kg.[49] Sibutramine may also be helpful in reducing both binge eating and weight.[50]

As with weight-loss treatment in obese patients, the primary care physician can play an important role in the treatment of patients with binge-eating disorder by following up at regular intervals, monitoring both binge eating and weight

fluctuations, and reinstating appropriate treatment when necessary.

Medical Complications

The principal complication of binge-eating disorder is weight gain, which may be accompanied by the health problems that are usually caused by being overweight. Thus, for patients suspected of having binge-eating disorder, the physician should concentrate on detecting and treating problems such as hypertension, type 2 diabetes mellitus, hypercholesterolemia, and joint disorders. In addition, because of the high rates of comorbid psychopathology, a careful history should be taken for conditions such as depression, anxiety, and alcohol dependence or abuse.

PROGNOSIS

Long-term follow-up studies suggest that although most individuals recover from binge-eating disorder, about one third continue to have a subthreshold eating disorder. Binge-eating disorder appears to be more cyclical than the other eating disorders—that is, episodes of binge eating may alternate with subthreshold status or remission.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

The drugs topiramate and sibutramine have not been approved by the Food and Drug Administration for uses described in this chapter.

References

1. Muise AM, Stein D, Arbess G: Eating disorders in adolescent boys: a review of the literature. J Adolesc Health 33:427, 2003

2. Prevalence of eating disorders. Academy of Eating Disorders, Northbrook, Illinois, 2005
www.aedweb.org/eating_disorders/prevalence.cfm

3. Croll J, Newmark-Sztainer D, Story M, et al: Prevalence and risk and protective factors related to disordered eating behaviors among adolescents: relationship to gender and ethnicity. J Adolesc Health 31:166, 2002

4. Striegel-Moore RH, Leslie D, Petrill SA, et al: One-year use and cost of inpatient and outpatient services among female and male patients with an eating disorder: evidence from a national database of health insurance claims. Int J Eat Disord 27:381, 2000

5. Kohn MR, Golden NH, Shenker IR: Cardiac arrest and delirium: presentations of the refeeding syndrome in severely malnourished adolescents with anorexia nervosa. J Adolesc Health 22:239, 1998

6. Ornstein RM, Golden NH, Jacobson MS, et al: Hypophosphatemia during nutritional rehabilitation in anorexia nervosa. Int J Eat Disord 32:83, 2003

7. Crisp AH, Norton K, Gowers S, et al: A controlled study of the effect of therapies aimed at adolescent and family psychopathology in anorexia nervosa. Br J Psychiatry 159:325, 1991

8. Kouba S, Hallstrom T, Lindholm C, et al: Pregnancy and neonatal outcomes in women with eating disorders. Obstet Gynecol 105:255, 2005

9. Stein A, Murray L, Cooper P, et al: Infant growth in the context of maternal eating disorders and maternal depression. Psychol Med 26:569, 1996

10. Agras WS, Hammer L, McNicholas F: A prospective study of the influence of eating disordered mothers on their children. Int J Eat Disord 25:253, 1999

11. Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders. National Institute for Health and Clinical Excellence (NICE), London, 2004
http://www.nice.org.uk

12. Steinhausen HC: The outcome of anorexia nervosa in the 20th century. Am J Psychiatry 159:1284, 2002

13. Lowe B, Zipfel S, Buchholtz C, et al: Long-term outcome of anorexia nervosa in a prospective 21-year follow-up study. Psychol Med 31:881, 2002

14. Strober M, Freeman R, Lampert C, et al: Controlled family study of anorexia nervosa and bulimia nervosa: evidence of shared liability and transmission of partial syndromes. Am J Psychiatry 157:393, 2000

15. Gorwood P, Kipman A, Foulon C: The human genetics of anorexia nervosa. Eur J Pharmacol 7:163, 2003

16. Kaye W, Strober M, Klump K: Neurobiology of eating disorders. Pediatric Psychopharmacology: Principles and Practice. Martin A, Scahill L, Charney C, Eds. Oxford University Press, New York, 2003

17. Tozzi F, Bulik CM: Candidate genes in eating disorders. Curr Drug Targets CNS Neurol Disord 2:31, 2003

18. Murphy D: Serotonin transporter: gene, genetic disorders, and pharmacogenetics. Mol Interv 4:109, 2004

19. Gorwood P: Eating disorders, serotonin transporter polymorphisms, and potential treatment response. Am J Pharmacogenomics 4:9, 2004

20. Katzman DK: Osteoporosis in anorexia nervosa: a brittle future? Curr Drug Targets CNS Neurol Disord 2:11,2003

21. Benini L, Todesco T, Dalle Grave R, et al: Gastric emptying in patients with restricting and binge/purging subtypes of anorexia nervosa. Am J Gastroenterol 99:1448, 2004

22. Miller KK, Grinspoon SK, Ciampa J, et al: Medical findings in outpatients with anorexia nervosa. Arch Intern Med 165:561, 2005

23. Katzman DK: Medical complications in adolescents with anorexia nervosa: a review of the literature. Int J Eat Disord 37:S52, 2005

24. Lock J: Treating adolescents with eating disorders in the family context: empirical and theoretical considerations. Child Adolesc Psychiatr Clin N Am 11:331, 2002

25. Powers PS, Santana C: Available pharmacological treatments for anorexia nervosa. Expert Opin Pharmcother 5:2287, 2004

26. Kaye WH, Weltzin TE, Nagata T, et al: Double-blind placebo-controlled administration of fluoxetine in restricting and restricting-purging anorexia nervosa. Biol Psychiatry 49:644, 2002

27. Barbarich NC, McConaha CW, Gaskill J, et al: An open trial of olanzepine in anorexia nervosa. J Clin Psychiatry 65:1480, 2004

28. Rogers L, Resnick MD, Mitchell JE, et al: The relationship between socioeconomic status and eating-disordered behaviors in a community sample of adolescent girls. Int J Eat Disord 22:15, 1997

29. Stice E: Risk and maintenance factors for eating pathology: a meta-analytic study. Psych Bull 128:825, 2002

30. Fichter MM, Quadflieg N: Twelve-year course and outcome of bulimia nervosa. Psychol Med 34:1395, 2004

31. Hay PJ, Bacultchuk J, Stefano S: Psychotherapy for bulimia nervosa and binge eating. Cochrane Database Syst Rev (3):CD000562, 2004

32. Fairburn CG, Jones R, Peveler RC, et al: Psychotherapy and bulimia nervosa: longer-term effects of interpersonal psychotherapy, behavior therapy, and cognitive behavior therapy. Arch Gen Psychiatry 50:419, 1993

33. Fairburn CB: Overcoming Binge Eating. Guildford Press, New York, 1995

34. Walsh BT, Fairburn CG, Mickley D, et al: Treatment of bulimia nervosa in a primary care setting. Am J Psychiatry 161:556, 2004

35. Durand MA, King M: Specialist treatment versus self-help for bulimia nervosa: a randomized clinical trial in general practice. Br J Gen Pract 53:371, 2003

36. Agras WS, Walsh T, Fairburn DB, et al: A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. Arch Gen Psychiatry 57:459, 2000

37. Mitchell JE, de Zwaan M, Roerig JL: Drug therapy for patients with eating disorders. Curr Drug Targets CNS Neurol Disord 2:17, 2003

38. Walsh BT, Wilson GT, Loeb KL, et al: Medication and psychotherapy in the treatment of bulimia nervosa. Am J Psychiatry 154:523, 1997

39. Walsh BT, Agras WS, Devlin MJ, et al: Fluoxetine for bulimia nervosa following poor response to psychotherapy. Am J Psychiatry 157:1332, 2000

40. Hoopes SP, Reimherr FW, Hedges DW, et al: Treatment of bulimia nervosa with topiramate in a randomized, double-blind, placebo-controlled trial, part 1: improvement in binge and purge measures. J Clin Psychiatry 64:1335, 2003

41. Koran LM, Agras WS, Rossiter EM, et al: Comparing the cost-effectiveness of psychiatric treatments: bulimia nervosa. Psychiatry Res 58:13, 1995

42. Rytomaa I, Jarvinen V, Kanerva R, et al: Bulimia and tooth erosion. Acta Odontol Scand 56:36, 1998

43. Jones JM, Lawson ML, Daneman D, et al: Eating disorders in adolescent females with and without type 1 diabetes: cross sectional study. BMJ 320:1563, 2000

44. Rydall AC, Rodin GM, Olmsted MP, et al: Disordered eating behavior and microvascular complications in young women with insulin-dependent diabetes mellitus. N Engl J Med 336:1849, 1997

45. Striegel-Moore RH, Franko DL: Epidemiology of binge eating disorder. Int J Eat Disord 34:S19, 2003

46. Yanovski SZ, Nelson JE, Dubbert BK, et al: Association of binge eating disorder and psychiatric comorbidity in obese subjects. Am J Psychiatry 150:1472, 1993

47. Wonderlich SA, de Zwaan M, Mitchell JE, et al: Psychological and dietary treatments of binge eating disorder: conceptual implications. Int J Eat Disord 34:558, 2003

48. Hudson JI, McElroy SI, Raymond NC, et al: Fluvoxamine in the treatment of binge-eating disorder: a multicenter placebo-controlled, double-blind trial. Am J Psychiatry 155:1756, 1998

49. McElroy SL, Shapira NA, Arnold LM, et al: Topiramate in the long-term treatment of binge eating disorder associated with obesity. J Clin Psychiatry 65:1463, 2004

50. Appolinario JC, Bacaltchuk J, Sichieri R, et al: A randomized double-blind, placebo-controlled study of sibutramine in the treatment of binge eating disorder. Arch Gen Psychiatry 60:1109, 2003

I LUNG FUNCTION ASSESSMENT AND THORACIC DIAGNOSTIC TECHNIQUES

ROLAND H. INGRAM, JR., M.D.

Although the history and physical examination are essential to the diagnostic process, pulmonary signs and symptoms often lack sufficient specificity to allow a definitive conclusion. Further information is often required, such as that acquired from assessments of physiologic function, imaging studies, and sampling procedures. Singly and in combination, these studies are important components of the clinical approach to respiratory disorders.

Assessment of Gas Exchange

The ultimate function of the lungs is to replenish the supply of oxygen in the blood and to eliminate the carbon dioxide produced by metabolic activity. Measurement of the partial pressure of oxygen and of carbon dioxide in arterial blood is central to the assessment of respiratory function. Arterial blood gases are obtained from a peripheral artery in a heparinized syringe. Care should be taken to avoid or expel air bubbles, and samples should be delivered to the laboratory on ice and should be processed promptly before cellular metabolism leads to an artifactual decrease in arterial oxygen tension (P_aO_2) and pH and an increase in arterial carbon dioxide tension (P_aCO_2).

OXYGEN

Abnormalities in oxygen exchange are commonly described in relation to the alveolar PO_2 (P_AO_2) or inspired oxygen concentration. Using the alveolar gas equation and assuming a normal resting respiratory exchange ratio (R) of 0.8 and that P_aCO_2 equals P_ACO_2, P_AO_2 can be estimated as follows:

$$P_AO_2 = P_IO_2 - (P_aCO_2 \times 1.2)$$

where P_IO_2 represents the inspired oxygen tension at body temperature, saturated with water.

The measured value of P_aO_2 is subtracted from the calculated value of P_AO_2 to give the alveolar-arterial difference in oxygen (A-aDO_2). Normal values for A-aDO_2 increase linearly with age because of a fall in P_aO_2 with essentially unchanged P_AO_2. Average values for A-aDO_2 range from approximately 9 mm Hg at 20 years of age to 15 mm Hg at 70 years of age. The A-aDO_2 is a commonly employed measure of the efficiency of oxygen exchange. However, even in the absence of changes in pulmonary gas exchanging function, it will change as a function of inspired oxygen fraction (F_IO_2), and it varies directly with cardiac output. The arterial-alveolar oxygen ratio, P_aO_2/P_AO_2, is somewhat more stable over varying inspired oxygen concentrations. The arterial-inspired oxygen ratio, P_aO_2/F_IO_2, is now widely used to quantitate abnormalities of oxygenation in critical care unit patients. A P_aO_2/F_IO_2 ratio of less than 250 indicates the presence of mild acute lung injury, and a P_aO_2/F_IO_2 ratio of less than 100 indicates a severe disorder.

Abnormalities of oxygen exchange [*see Table 1*] are most commonly caused by mismatching of pulmonary ventilation ($\dot{V}$) and perfusion ($\dot{Q}$) or by shunting. Impaired diffusion across the alveolar-capillary membrane generally does not cause abnormalities in oxygenation at rest but can cause abnormalities during exercise and at high altitudes. Abnormalities caused by diffusion impairment and $\dot{V}/\dot{Q}$ mismatching can be corrected by increasing the F_IO_2 and can be completely abolished by 100% inspired oxygen. Hypoxemia caused by shunts, which may be pulmonary or intracardiac, is not corrected by the administration of 100% inspired oxygen. By contrast, other causes of hypoxemia, such as hypoventilation and low inspired O_2 concentration, do not cause an increase in A-aDO_2.

Systemic Oxygen Transport

The total amount of O_2 delivered to the systemic circulation is the product of the cardiac output and the O_2 content per unit of arterial blood (C_aO_2). The C_aO_2 is determined by the concentration and characteristics of hemoglobin and the arterial oxygen saturation (S_aO_2), as indicated by the following equation:

$$C_aO_2 = \text{Hgb (g/dl)} \times S_aO_2\ (\%) \times 1.39\ \text{ml } O_2/\text{g Hgb}$$

The last term in the equation reflects the amount of oxygen that is normally bound to fully saturated hemoglobin, the so-called carrying capacity of normal hemoglobin. The arterial O_2 saturation refers to the percentage of the total O_2 binding sites on hemoglobin that is actually occupied by O_2. The S_aO_2 is in turn determined by the P_aO_2 and the physicochemical properties of hemoglobin, as reflected by the oxygen-hemoglobin dissociation curve [*see Figure 1*]. In the presence of acidemia, fever, elevated concentrations of 2,3-diphosphoglycerate, and certain abnormal hemoglobin types (e.g., hemoglobin Kansas), the oxy-

Table 1 Categorization of Hypoxemia

Cause	*Example*	*P_aO_2*	*P_aCO_2*	*A-aDO_2*	*Response to ↑ F_IO_2*
Low inspired oxygen tension	Mountaineering	↓	Normal or ↓	Normal	Good
Hypoventilation	Drug overdose	↓	↑	Normal	Good
Diffusion impairment	Pulmonary fibrosis plus exercise	↓	Normal or ↓	↑	Good
Ventilation-perfusion imbalance	Chronic obstructive pulmonary disease; pneumonia	↓	Normal, ↓, or ↑	↑	Good
Right-to-left shunt	Pulmonary edema	↓	Normal or ↓	↑	Poor or none

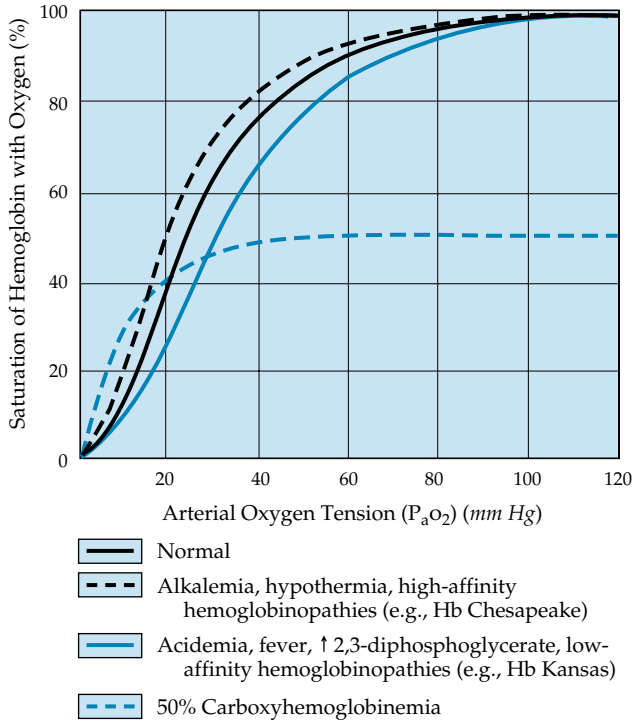


Figure 1 **The relation between arterial oxygen tension (P_aO_2) and the percentage of oxygen binding sites on hemoglobin that are saturated by oxygen is shown by the oxygen-hemoglobin dissociation curve. The sigmoidal shape of the normal curve reflects the cooperative binding of oxygen by hemoglobin. An alteration in pH or body temperature, a change in the concentration of 2,3-diphosphoglycerate, or the presence of certain abnormal hemoglobin types shifts the curve to the left or right, so that at any given P_aO_2, the hemoglobin saturation will be correspondingly increased or decreased. Carboxyhemoglobinemia, in which the oxygen binding sites of hemoglobin are occupied by carbon monoxide, results in both a deformation of the shape of the dissociation curve and a reduction in the maximum number of binding sites that are available to oxygen.**

gen-hemoglobin dissociation curve shifts to the right, which gives a decreased affinity of hemoglobin for oxygen and increased availability of oxygen to tissues. Alkalemia, hypothermia, and other abnormal hemoglobin types (e.g., hemoglobin Chesapeake) have the opposite effect and shift the position of the curve to the left, reflecting an increased affinity of hemoglobin for oxygen and reduced availability of oxygen to tissues.

Anemia does not alter P_aO_2 or S_aO_2. Anemia does, however, reduce the value for C_aO_2 and will decrease O_2 delivery to the tissues if there is not a commensurate increase in cardiac output. A fixed change in P_aO_2 causes a considerably larger change in S_aO_2 over the steep portion of the oxygen-hemoglobin dissociation curve than it does over flatter portions of the curve. For instance, with a fall in P_aO_2 from 100 mm Hg to 90 mm Hg, S_aO_2 decreases from 97.4% to 96.8%, assuming that hemoglobin type is normal and physiologic conditions are standard—pH of 7.40 and body temperature of 37° C (98.6° F). In contrast, a 10 mm Hg decrement in P_aO_2 from 55 mm Hg to 45 mm Hg causes the S_aO_2 to decrease from 88.2% to 80.5%.

A profound reduction in arterial O_2 content may be observed without a significant change in P_aO_2 if O_2 binding to hemoglobin is acutely altered, as occurs in carbon monoxide intoxication. Because of the strong affinity of carbon monoxide for hemoglobin, an arterial carbon monoxide tension (P_aCO) of less than 1 mm Hg is sufficient to cause a 50% saturation of hemoglobin with CO. Under these conditions, P_aO_2 may be 100 mm Hg, but severe tissue hypoxia may be present because S_aO_2 and C_aO_2 values have been reduced by half.[1]

Oximetry

Oxygenation can be monitored noninvasively by pulse oximetry. A pulse oximeter, placed on either a finger or an earlobe, measures the absorption of red (660 nM) and near infrared (940 nM) light through these tissue beds to estimate the ratio:

$$O_2\,Hb/O_2\,Hb + \text{reduced Hb}$$

By assuming the pulsatile portion of the signal represents arterial blood and by comparing the ratio of the pulsatile portion to the nonpulsatile component to a calibration curve of known mixtures of oxyhemoglobin and reduced hemoglobin, the pulse oximeter is capable of estimating S_aO_2. Oximetric estimates are accurate to within ± 1% to 2% for true saturations above 90%. Accuracy deteriorates at saturations below 80%, with errors of ± 5% to 8%. Pulse oximetry is commonly used in emergency and critical care settings, for in-hospital patient transportation, and for assessing oxygenation during sleep. Medicare guidelines allow the use of resting pulse oximetry for determination of qualification for long-term oxygen therapy, but arterial blood gas measurements are still necessary for exercise evaluations for adjudication of disability under Social Security Administration regulations. Accuracy of pulse oximetry is reduced by carboxyhemoglobin, methemoglobin, anemia, motion, bright ambient light, poor perfusion, nail polish, and darkly pigmented skin.[2]

CARBON DIOXIDE

Physiologic mechanisms normally act to maintain the P_ACO_2 level within a narrow range (35 to 45 mm Hg) despite large changes in metabolic CO_2 production. Elevated P_ACO_2 levels (> 45 mm Hg) are termed hypoventilation, and low P_ACO_2 values (< 35 mm Hg) are termed hyperventilation. Corresponding alterations in P_aCO_2 are termed hypercapnia and hypocapnia, respectively. The P_aCO_2 level is directly proportional to the ratio of carbon dioxide production to alveolar ventilation:

$$P_aCO_2 = K(\dot{V}CO_2/\dot{V}_A)$$

where $\dot{V}CO_2$ equals $\dot{V}_E \times F_ECO_2$ and represents the amount of CO_2 (in ml/min) produced by the body's metabolism; and K is a constant (equal to 0.863) that reflects the fact that gas exchange occurs at normal body temperature under conditions of full saturation with water, assuming P_ACO_2 is equal to P_aCO_2. Thus, for any given level of $\dot{V}CO_2$, the alveolar (and arterial) PCO_2 is determined by the level of alveolar ventilation. Hypercapnia is categorized according to causes [*see Table 2*].

VENTILATION

Alveolar ventilation ($\dot{V}_A$) is that portion of the minute ventilation ($\dot{V}_E$) that comes into equilibrium with alveolar gas and represents the difference between $\dot{V}_E$ and dead space ventilation ($\dot{V}_D$). A decrease in $\dot{V}_A$ results from either a reduction in $\dot{V}_E$ or an increase in $\dot{V}_D$.

Minute ventilation is the volume of gas that moves in and out of the lung. The minute ventilation can be calculated by multiplying the respiratory frequency (f) and the tidal volume (V_T), which is the volume of air expired with each breath. Typical resting values in the adult are as follows: respiratory frequency,

Table 2 Categorization of Hypercapnia

Cause	*Example*	P_aCO_2	$\dot{V}_E$	V_D/V_T	*A-a*DO_2
Defective central control of breathing	Drug overdose	↑	↓	Normal	Normal
Neuromuscular disease	Amyotrophic lateral sclerosis	↑	↓	Normal or ↑	Normal or ↑
Chest wall disease	Kyphoscoliosis	↑	↓	Normal or ↑	Normal or ↑
Primary lung disease	Chronic obstructive pulmonary disease	↑	Normal or ↑	↑	↑

14 breaths/min; tidal volume, 400 ml/breath; and minute ventilation, 5.6 L/min. However, there is considerable variability in these values among normal persons and in the same person throughout the day.

Dead space consists of two types: anatomic and functional (also called physiologic). Anatomic dead space is the portion of inspired tidal volume that does not communicate with perfused alveoli and therefore does not participate in gas exchange. Anatomic dead space consists of the tracheobronchial tree from the oropharynx and nasopharynx down through the nonalveolated terminal bronchioles. The volume of the anatomic dead space in milliliters is roughly equal to a person's lean body weight in pounds. Functional dead space ventilation is the sum of the excess ventilation relative to perfusion for all lung units where the ratio of ventilation to perfusion ($\dot{V}/\dot{Q}$) is significantly greater than 1.

A fixed volume of gas from each breath is required to fill the anatomic dead space. For any $\dot{V}_E$, fast and shallow breathing wastes a greater percentage of the minute ventilation because the anatomic dead space is proportionately greater as tidal volume decreases. Conversely, the alveolar ventilation at any $\dot{V}_E$ is greatest when breathing is slow and deep [*see Figure 2*]. Additionally, dead space ventilation can be caused by inequalities in the distribution of alveolar ventilation and perfusion. Certain lung regions have greater ventilation than perfusion. The excess of regional ventilation to blood flow can be thought of as wasted, although in most disease states, the mismatch is caused by regional deficits in perfusion rather than excessive ventilation.

Dead space can be assessed by measuring the fractional concentration of CO_2 in exhaled gas (F_ECO_2) and by estimating the fractional concentration of CO_2 in alveolar gas (F_ACO_2). The F_ACO_2 is estimated by assuming ideal alveolar air conditions, in which P_aCO_2 values are equal to P_ACO_2 values. V_D/V_T, the so-called dead space fraction, is calculated as follows:

$$\frac{(F_ACO_2 - F_ECO_2)}{F_ACO_2}$$

The normal value for V_D/V_T is less than 0.4. The increased work demands imposed by a high V_D/V_T contribute significantly to dyspnea and often to respiratory failure in clinical lung disease.

Pulmonary Function Tests

Pulmonary function testing is used to categorize the nature and severity of pathophysiologic disturbances, to follow the progression of known cardiopulmonary disorders, and to measure the response to therapy. Pulmonary function is often used as the basis for the definition of disability for insurance purposes. Lung resection is the only indication for preoperative pulmonary function testing for which a consensus currently exists. Otherwise, such testing offers little additional benefit over clinical parameters in assessing the risk for postoperative pulmonary complications, and prohibitive thresholds cannot be reliably established. A review reported that abnormal pulmonary function predicted a significantly increased relative risk for postoperative pulmonary complications in only four of 11 studies.[3]

The most commonly used pulmonary function tests are based on the forced vital capacity maneuver, measurements of lung volumes, and pulmonary diffusion capacity. Recording of the forced vital capacity maneuver produces a record of volume versus time, flow versus volume, or both during a forced exhalation from total lung capacity (TLC) to residual volume (RV). Forced vital capacity (FVC) is measured with a spirometer and is the most basic and useful lung function test. The test is simple and highly reproducible. However, valid values require maximal efforts by the patient, and these efforts may be compromised by pain or debilitation. The forced vital capacity maneuver shows whether obstruction is present and, if present, quantitates the severity. If there is no obstruction, a reduced vital capacity indicates restriction. In the absence of obstruction, the severity of restriction is defined by comparing vital capacity with the value predicted according to the patient's height, sex, age, and race.

LUNG VOLUMES

The balance between the physical properties of the lung and chest wall and the action of inspiratory and expiratory muscles determines lung volume. TLC is determined by the action of inspiratory muscles, chiefly the diaphragm, against the elastic recoil of the lung and of the chest wall. RV is determined by expiratory muscles opposing the outward recoil of the lung and chest wall at low lung volumes (in older persons, RV may be increased by airway closure). At the end of each tidal breath (expiration), the respiratory muscles are quiet and the lungs are expanded approximately one third of the way from RV. Functional residual capacity (FRC) is the mechanical resting position of the respiratory system and is determined by the balance of elastic recoil of the lung, which reduces respiratory system volume, and elastic recoil of the chest wall, which increases respiratory system volume. In obstructive lung diseases, FRC may be elevated above the lung–chest wall elastic balance volume if inspiration is initiated before that volume is reached. This is so-called dynamic hyperinflation (or air trapping). The vital capacity (VC) is the largest breath a person can take. The tidal volume at rest is the volume of air that is inspired and expired during normal quiet breathing. A person who has just exhaled a tidal breath (to reach FRC) can either inhale to the top of his or her lung volume (the TLC) or exhale further to the bottom of his or her lung volume (the RV). The inspiratory capacity (IC) is the volume from FRC up to TLC, and the expiratory reserve volume (ERV) is the volume from FRC down to RV. The inspirato-

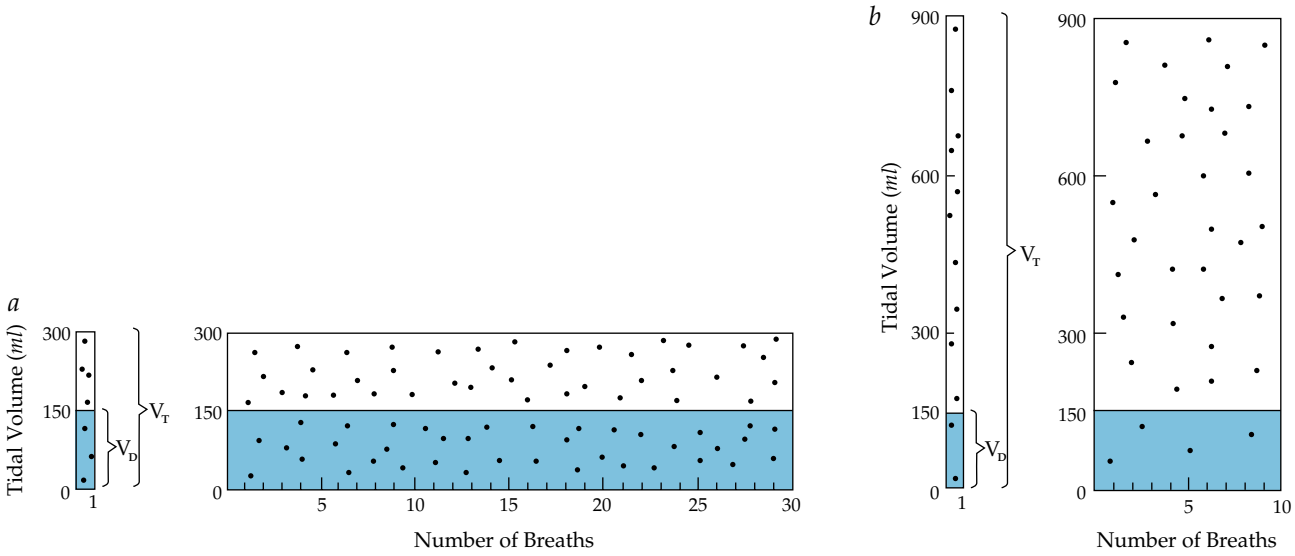
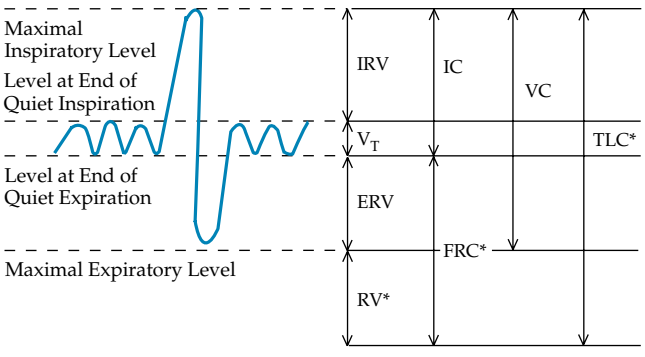


Figure 2 **The effect of varying the respiratory frequency (f) and the tidal volume (V_T) on the alveolar ventilation ($\dot{V}_A$) and the dead space ventilation ($\dot{V}_D$) is shown. Two different breathing patterns (*a* and *b*) in a normal person weighing 150 lb are shown; both patterns achieve a minute ventilation ($\dot{V}_E$) of 9 L/min. For each breath, the dead space volume (V_D) remains constant; its value in milliliters is estimated to be equal to the lean body weight in pounds, or 150 ml. In frame *a*, the person breathes rapidly and shallowly at a frequency of 30 breaths/min and a V_T of 300 ml. With each breath, 150 ml of the inspired air goes to filling the dead space; the remaining 150 ml equilibrates with the gas exchange areas of the lung. Thus, $\dot{V}_D$ equals 150 ml/breath × 30 breaths/min, or 4.5 L/min; and $\dot{V}_A$ equals 150 ml/breath × 30 breaths/min, or 4.5 L/min. The ratio of dead space volume to tidal volume, V_D/V_T, which represents the portion of the total ventilation that is not involved in gas exchange, is 0.50. In contrast, in frame *b*, breathing is slower and deeper, with a V_T of 900 ml and a frequency of 10 breaths/min. $\dot{V}_D$ remains constant at 150 ml; however, because 900 ml of air is now being taken in with each breath, the remaining 750 ml of air is available for gas exchange. Therefore, $\dot{V}_D$ equals 150 ml/breath × 10 breaths/min, or 1.5 L/min; and $\dot{V}_A$ equals 750 ml/breath × 10 breaths/min, or 7.5 L/min. The value of V_D/V_T in this example is 0.17.**

ry reserve volume (IRV) is the volume from the end of an inspired breath up to TLC.[4]

Most of these lung volumes are measured with a spirometer. In the forced expiratory maneuver, the patient breathes quietly until the tidal volume is stable, inhales to TLC, and then performs a maximal expiration followed immediately by a full inspiration to TLC. The lung volumes and capacities are determined from a mechanical or electronic record of the volume changes [*see Figure 3*]. The RV cannot be measured by spirometry. Instead, the FRC is measured indirectly by one of several methods. Once the FRC is known, the ERV is subtracted from the FRC to give the RV.



*Not determined by spirometry.

TLC= Total Lung Capacity
FRC= Functional Residual Capacity
RV= Residual Volume
VC= Vital Capacity
V_T = Tidal Volume
IC= Inspiratory Capacity
ERV= Expiratory Reserve Volume
IRV= Inspiratory Reserve Volume

Figure 3 **Most lung volumes and capacities can be measured by spirometry. Inspired and expired volumes during normal quiet breathing are shown here.**

Methods for measuring FRC include plethysmography (the so-called body box) and gas dilution techniques, such as helium equilibration and nitrogen washout.

Plethysmography

In the body plethysmograph, the patient breathes quietly at normal tidal volumes. At the end of a tidal breath (at FRC), a shutter occludes the airway. The individual pants against the occluded airway and, in so doing, expands the volume of gas in the chest with each inspiratory effort and compresses the volume of gas with each expiratory effort. The initial volume (FRC) can be calculated by the use of Boyle's law ($V_1P_1 = V_2P_2$).[5]

Gas Dilution

There are two gas dilution methods for measuring FRC: the open-circuit nitrogen washout technique and the helium dilution method.

With the open-circuit nitrogen washout technique, the patient breathes quietly at normal tidal volumes. At the end of a tidal breath, the inhaled gas is changed from air to 100% O_2. The amount of nitrogen in each subsequent exhalation decreases as O_2 replaces the resident air in the lung until the exhaled gas contains only O_2 and CO_2. The nitrogen in each breath is measured, and the nitrogen measurements of all the exhalations are then summed to give the total volume of nitrogen exhaled after the switch to pure O_2. The initial volume of gas required to provide

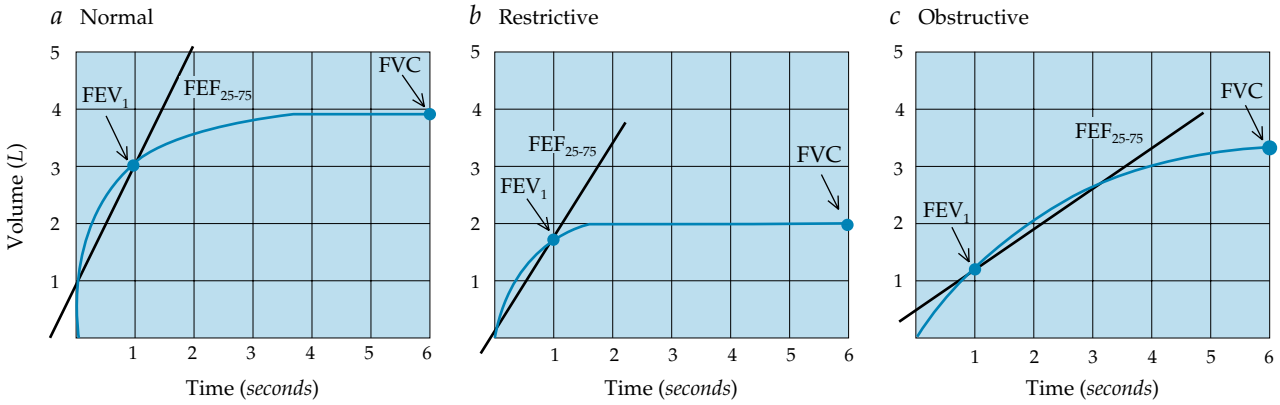


Figure 4 **The volume exhaled is plotted on the ordinate against time on the abscissa. Spirometry illustrates distinctive patterns for normal breathing (*a*), obstructive diseases (*b*), and restrictive diseases (*c*). FEV_1 is forced expiratory volume in 1 second, FVC is forced vital capacity, and FEF_{25-75} is forced expiratory flow between 25% and 75% of vital capacity.**

this much nitrogen (the FRC) can then be calculated, with the knowledge that 79% of room air is composed of nitrogen.

The helium dilution method uses a closed circuit. The patient breathes quietly at normal tidal volumes. At the end of a breath, the air is changed to a mixture that contains 10% helium. The person breathes until the helium concentration reaches a new steady state at a lower level. CO_2 is absorbed by soda lime, and O_2 is added to the system to replace that consumed. The initial and final concentrations of helium are used to determine the FRC by calculating how much helium-free gas had to be added to the volume of the closed system to cause the observed drop in helium concentration.[6]

FORCED EXPIRATORY VITAL CAPACITY (SPIROMETRY AND FLOW-VOLUME LOOPS)

The FVC maneuver is useful in assessing the dynamic mechanical properties of the lung.[7] From TLC to RV, given sufficient effort, flow rates are determined by the mechanical properties of the lung. The most likely mechanism for expiratory flow limitation is described by the wave-speed theory. During a forced exhalation, alveolar pressure exceeds pleural pressure by an amount equal to the elastic recoil pressure of the lung. As flow occurs, there is a decrease in pressure along the length of the airways because of frictional (resistive) losses, so that at some point or points in the airways, pleural (outside) pressure equals airway pressure. In points downstream (mouthward) from that point, there is compression of airways (negative transmural pressures). Flow becomes limited (i.e., a choke point forms) when and where gas velocity within the airway equals the speed of wave propagation in the airway wall (i.e., wave speed). Wave speed is directly proportional to airway area and stiffness. Larger, stiffer airways have higher wave speed and therefore allow higher maximal flow. Smaller, floppier airways have a lower wave speed and therefore allow lower maximal flows. All airway segments have a wave speed at all transmural pressures; choke points occur only when and where local intraluminal gas velocity equals the local wave speed.[8] Data from the FVC maneuver are displayed either as volume-time plots (spirometry) or flow-volume loops: they contain the same information, but the information is depicted in different formats.

Spirometric Patterns

Obstructive and restrictive disorders produce distinctive patterns on spirometry [*see Figure 4*]. A normal spirogram rises rapidly, with 95% of the vital capacity exhaled within the first 3 seconds. In obstruction, the rise is slower, and a considerable portion of the vital capacity is exhaled after 3 seconds; the spirogram fails to reach a plateau. Obstruction is identified also by an absolute reduction in the forced expiratory volume in 1 second (FEV_1) as well as the ratio with the forced vital capacity (FEV_1/FVC). A low FVC with a high FEV_1/FVC is typical of restrictive processes, such as pulmonary fibrosis, neuromuscular diseases, and chest wall deformities (e.g., kyphoscoliosis).

Flow-Volume Loop Patterns

Obstructive lung diseases reduce maximum expiratory airflow by decreasing elastic recoil pressure (e.g., emphysema), narrowing peripheral airways (e.g., chronic bronchitis and asthma), or both. Flow is reduced when airway obstruction is caused by parenchymal lung disease. As the lung empties during forced exhalation, flow rates decrease and become progressively lower than normal as volume decreases. Consequently, the curve acquires a scooped-out appearance [*see Figure 5*]. Flow rates become extremely low at the end of the FVC maneuver. In restrictive diseases, the peak flow is somewhat decreased. The total exhaled volume is small, and the flow-volume curve is horizontally compressed [*see Figure 5*]. Flow from the top of the vital capacity to the bottom decreases in a linear manner, as it does in normal lungs, and does not have the scooped-out appearance seen in airway obstruction caused by parenchymal lung disease. Variable intrathoracic airway obstruction, such as that caused by tracheomalacia, limits expiratory flow, so that the expiratory limb of the flow-volume loop is flat. During inspiration, the posterior membranous portion of the trachea billows outward in response to negative intrathoracic pressure, permitting greater flow. During expiration, pleural pressure is greater than the pressure in the intrathoracic airway, which causes a narrowing of the airway at the site of the obstruction.

The term variable indicates that the relation of intraluminal pressure to surrounding pressures [*see Figure 5*] determines flow. Variable extrathoracic airway obstruction is most common at the larynx and occurs as a result of vocal cord paralysis or tracheomalacia of the extrathoracic trachea. It causes inspiratory limitation. Because of the Bernoulli effect from laryngeal narrowing, the pressure in the extrathoracic upper airway during inspiration is less than the surrounding tissue pressure. As a

result, the airway tends to collapse, which aggravates any structural narrowing.[9]

Maximal inspiratory flow rates less than 2 L/sec suggest severe upper airway obstruction. A fixed airway obstruction such as stenosis of either the intrathoracic or the extrathoracic trachea causes equal and symmetrical flattening of the limbs of the curve that represent severe flow during inspiration and expiration [*see Figure 5*]. Lesions in either intrathoracic or extrathoracic locations may produce a fixed obstructive pattern.

DIFFUSING CAPACITY

The pulmonary diffusing capacity is physiologically determined by the surface area and thickness of the alveolar capillary membrane, by the volume of blood circulating in the alveolar capillary bed, and by the reaction rate of the test gas with hemoglobin. In actuality, the pulmonary diffusing capacity reflects the overall efficiency of gas transfer at each step from the mouth to pulmonary capillary hemoglobin. Because the process involves much more than just diffusion, the term transfer factor is more appropriate, but the term diffusing capacity remains in common use in North America.

The transfer factor of any gas is calculated by dividing the volume of gas taken up by the difference between the alveolar concentration and the mixed capillary concentration ($P_{C}O_2$). Carbon monoxide is usually used for measuring the transfer factor. The affinity of CO for hemoglobin is so high that the hemoglobin takes up almost all the CO entering the blood. The partial pressure in plasma remains so low that the value for mixed capillary PCO can be omitted from the calculation. Thus, the diffusing capacity for CO is expressed as the volume of CO taken up per minute per mm Hg of P_ACO and can be calculated by using the following equation:

$$D_{L_{CO}} = V_{CO} / P_A CO$$

Diffusing capacity is most commonly measured by the single-breath technique. The patient takes a vital capacity breath of a gas mixture containing 0.3% CO and 10% helium. After a 10-second breath hold, the patient exhales. The first portion of exhaled gas, which is contaminated with dead space ventilation, is discarded. The next liter is collected and analyzed. The helium is needed to calculate the amount of dilution of the inspired sample so that the initial alveolar CO can be calculated

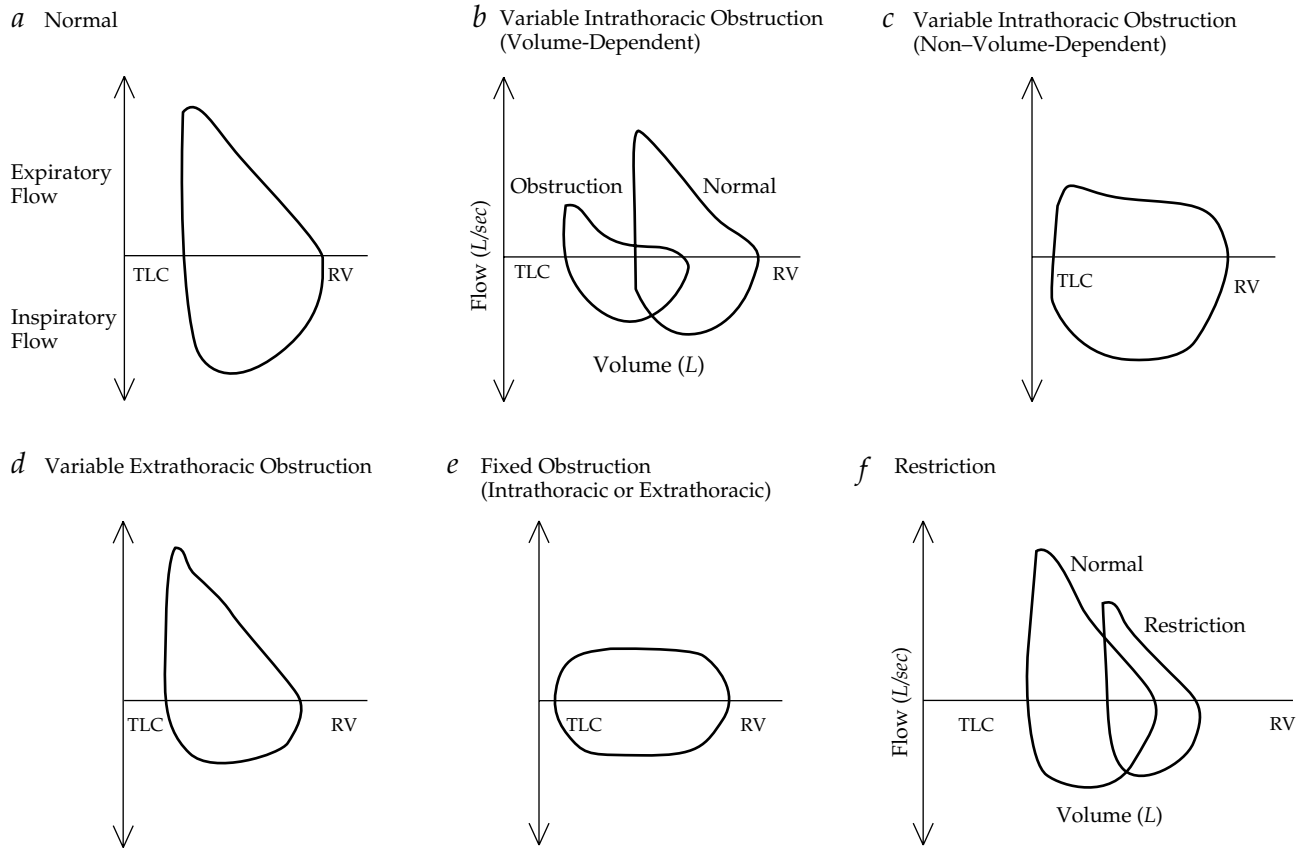


Figure 5 **Flow is depicted on the ordinate with expiration above and inspiration below the intercept. Volume is shown on the abscissa going left to right from TLC to RV. (*a*) A normal expiratory flow-volume curve is shown. (*b*) In variable intrathoracic obstruction, the expiratory flow-volume curve shows a scooped-out appearance as a result of progressive decreases in flow as lung volume becomes smaller. Obstruction is volume dependent. Flow rates at any given lung volume (isovolumic flow) are reduced. Because of air trapping, the entire curve may be shifted to a higher lung volume (leftward). This pattern is typical of chronic obstructive pulmonary disease or asthma. (*c*) The expiratory flow-volume curve shows a decreased flow that is the same at all lung volumes. Obstruction, which is not dependent on lung volume, is caused by upper airway obstruction, not by disease of the lung parenchyma. (*d*) Disproportionate reduction of inspiratory airflow is indicative of variable extrathoracic upper airway obstruction. (*e*) Fixed airway obstruction, the site of which is undetermined, shows volume-independent reduction of flow in both inspiration and expiration. (*f*) An expiratory flow-volume curve in a patient with a restrictive disorder is shown. Isovolumic flow rates are increased, whereas the volume axis is compressed and shifted toward lower volume (rightward).**

and is also used to calculate alveolar volume (VA). The slope of the natural logarithm of initial-final alveolar CO concentration versus time is the volume-normalized diffusing capacity ($D_{L_{CO}}$/VA). Multiplication by lung volume yields $D_{L_{CO}}$, the reported value. Other ways of measuring $D_{L_{CO}}$ include the steady state and rebreathing methods. Rapidly responding CO analyzers and computer technology have made it possible to measure diffusing capacity throughout a single exhalation, allowing the study of the effects of nonhomogeneities in ventilation and perfusion on diffusion, the so-called intrabreath method.

Interpretation of Pulmonary Function Tests

The results of pulmonary function tests are interpreted in comparison with predicted normal values that have been defined for various populations. Pulmonary function studies should be performed with carefully calibrated equipment that meets uniform standards, with the use of approved techniques in patients who are well coached to ensure a maximal effort and reproducible results. Interpretive errors can be minimized by adhering to the standards published by the American Thoracic Society (ATS) in 1994.[7] Newer standards are now being prepared jointly by the ATS and European Respiratory Society (http://www.thoracic.org). Results that are within the 95% confidence limits for the reference population are considered to be normal.

Pulmonary function tests seldom confirm a specific diagnosis. Rather, they show certain patterns, each of which may be consistent with a number of different diseases. Once a specific diagnosis has been made by combining the pulmonary function results with other clinical information, the quantitative results help in assessing the severity of the physiologic impairment caused by the disease. Vital capacity, FEV_1, and FEV_1/FVC are the basic spirometric parameters used for pulmonary function interpretation to make an assessment of normality or to define patterns of abnormality. TLC and $D_{L_{CO}}$ provide important additional information and may provide independent evidence of a pattern of physiologic abnormality.

OBSTRUCTION

An obstructive defect is defined as reduction in maximal expiratory flows and FEV_1/FVC. Although the earliest evidence of obstruction can be found in a reduction in the instantaneous flow after 75% of a vital capacity has been exhaled (FEF_{75}) or in average midflow rates ($FEF_{25–75}$), these are nonspecific findings of uncertain clinical significance.

When the presence of obstruction has been determined, its severity is best assessed with the FEV_1 as a percentage of the predicted value. Although there is some variability in interpretative schemes, an FEV_1 less than 50% of predicted is generally considered to be severe, a value between 50% and 70% of predicted is moderate impairment, and higher values of FEV_1 with a reduced FEV_1/FVC are considered to be mild disturbance. Obstruction with normal lung volumes and $D_{L_{co}}$ suggests airway disease with preservation of pulmonary parenchyma. These findings are consistent with mild asthma or chronic bronchitis.

Hyperinflation is defined by an elevated TLC, and air trapping is defined by an increased RV/TLC ratio. Obstruction with hyperinflation and air trapping is characteristic of more severe airway disease. Vital capacity may be reduced; the concave shape of the expiratory flow-volume curve and the flattened spirogram may provide indirect evidence of air trapping as the cause of the reduction of vital capacity, but lung volume measurements are often necessary to confirm the impression. Hyperinflation in the presence of a normal diffusing capacity for carbon monoxide indicates preservation of the alveolar capillary bed and may be suggestive of chronic obstructive bronchitis or chronic severe asthma. A low $D_{L_{CO}}$ in a patient with obstruction and hyperinflation suggests that emphysema is the cause of the obstruction. In this disorder, the obstruction occurs as a result of the loss of lung parenchyma, which ordinarily attaches to the outside of the small airways and, by radial traction, helps maintain the airway diameter during expiration. Loss of this supporting tissue allows the airways to collapse during expiration. Loss of elastic recoil also reduces expiratory flow. As the lung parenchyma is destroyed, the capillary bed is also lost, causing a low diffusing capacity for carbon monoxide.

RESTRICTION

A restrictive pattern is defined by a reduced TLC, although presence of a restrictive pattern may be inferred by a reduced VC in the presence of a normal or elevated FEV_1/FVC. A restrictive defect may be caused by parenchymal lung disease, by chest wall disorders, or by neuromuscular disorders. Patients whose vital capacity and residual volume are reduced by about the same amount have symmetrical reduction in lung volumes. These patients often have fibrotic diseases that shrink the lungs, such as pulmonary fibrosis and sarcoidosis. Destruction of the lung parenchyma results in reduced volume of the capillary bed and a lower $D_{L_{CO}}$. Patients with neuromuscular diseases often show a pattern in which the vital capacity is reduced but the residual volume is normal or even slightly increased. Weak inspiratory muscles limit the size of a maximal breath. Weak expiratory muscles, especially a weak rectus abdominis muscle, prevent complete emptying. Muscle weakness does not affect the lung parenchyma, and the $D_{L_{CO}}$ is normal when corrected for the lung volume. Patients with chest wall abnormalities caused by kyphosis and scoliosis have a similar pattern.

ISOLATED LOW DIFFUSION

If the patient is not anemic or does not have an elevated carboxyhemoglobin level, isolated low diffusion suggests loss of the pulmonary capillary bed, either because of pulmonary vascular disease (e.g., pulmonary emboli or pulmonary hypertension) or because of an early interstitial lung disorder that has not yet reduced the lung volumes or of early emphysema that has not yet produced airflow obstruction.

MUSCLE FUNCTION

Assessment of respiratory function is frequently performed in the intensive care unit (ICU), often to help determine reasons for prolonged ventilator dependence. The diaphragm is the chief inspiratory muscle. Abnormalities in diaphragmatic function generally manifest as exertional dyspnea accompanying reductions in VC and TLC with a preserved $D_{L_{CO}}$. Patients with weakness or paralysis of one or both hemidiaphragms typically experience more dyspnea when they are supine than when they are seated or in an upright position (orthopnea). On physical examination, the abdominal wall may retract with inspiration (abdominal paradox), which indicates that the diaphragm is neither performing active inspiratory work nor providing a pressure barrier between the abdominal and pleural compartments.

In any severe obstructive or restrictive ventilatory disorder, the muscles of breathing carry a heavy workload in providing

ventilation, even at rest. High respiratory muscle loads elicit reflex responses that manifest as a rapid, shallow breathing pattern. High inspiratory workloads may also cause the appearance of abdominal paradox. These two findings, once thought to reflect respiratory muscle fatigue, are nonspecific indicators of respiratory muscle load and may be seen before diaphragmatic fatigue is actually present. Continuous loads can lead to the development of inspiratory muscle fatigue, defined as the loss of contractile strength despite maximal stimulation. Diaphragmatic function can be assessed in the laboratory by several techniques. In patients with significant diaphragmatic weakness or paralysis, vital capacity may decline by 20% or more in the supine position compared with the upright position. Muscle strength can be assessed by having the patient perform maximal inspiratory and expiratory efforts against a closed system and measuring the static pressures that are generated.[10]

A young, healthy adult can generate negative pressures in excess of –100 cm H_2O on inspiration and positive pressures in excess of +120 cm H_2O on maximal expiratory effort. In certain clinical settings, serial measurements of muscle strength may be of particular value. Examples include progressive neuromuscular disorders, such as Guillain-Barré syndrome, in which it is important to identify respiratory muscle weakness or paralysis early in its course (before the development of overt respiratory failure) and disorders that are characterized by fluctuating periods of muscle weakness, such as myasthenia gravis.

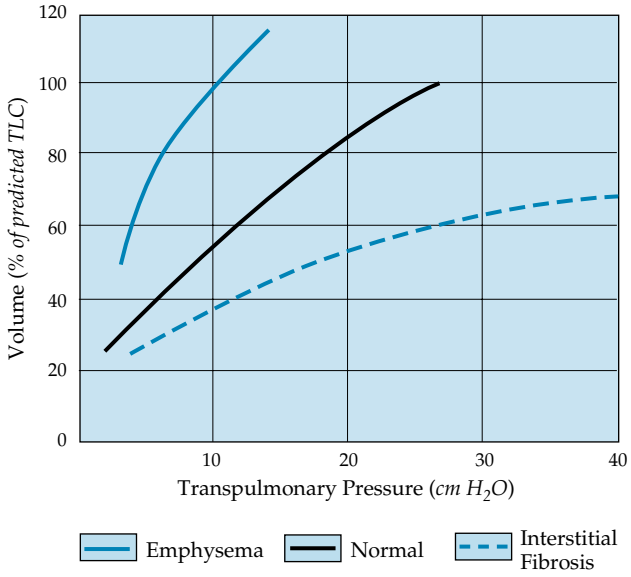


Figure 6 **Compliance is defined as the change in lung volume per change in unit pressure. The changes in pulmonary compliance that occur with various respiratory conditions are reflected in differences in the shape and position of the pressure-volume curve of the lung. Pressure-volume curves are obtained by measuring transpulmonary pressure at various lung volumes during a stepwise exhalation (or during in vitro deflation of the lung). The compliance of the lungs is abnormally high in emphysema and abnormally low in diffuse interstitial fibrosis.**

Nonroutine Pulmonary Function Tests

LUNG ELASTICITY

The elastic properties of the lung can be assessed by measuring the static transpulmonary pressure as a function of lung volume. Transpulmonary pressure is defined as the pleural pressure minus the oral pressure, and static denotes that the measurements are made at a time when no airflow is occurring. When the mouth is open, oral pressure is equal to atmospheric pressure, and pleural pressure can be approximated from intraesophageal pressure. Measurements of intraesophageal pressure require that the patient swallow a balloon-tipped catheter; thus, it is not a routine procedure in most clinical pulmonary function laboratories.

The static transpulmonary pressure, measured at various volume intervals during deflation of the lung, can be plotted against the percentage of the predicted total lung capacity [*see Figure 6*]. The change in lung volume divided by the change in pressure, measured over the volume interval extending from 500 ml above FRC to FRC, is known as the static compliance. Normal values for the static compliance range from 0.1 to 0.4 L/cm H_2O. Pressure-volume curves can also be plotted for obstructive and restrictive disorders and can be modeled mathematically for further analysis.[11] The static compliance is abnormally high in emphysema and abnormally low in interstitial fibrosis.

BRONCHIAL PROVOCATION

One of the defining characteristics of asthma is an increase in the responsiveness of the airways to a number of stimuli. If lung function is normal but the history suggests the presence of asthma, the demonstration of bronchial hyperresponsiveness to one of several constrictor challenges can be useful in establishing a diagnosis.[12]

Methacholine, carbacholine, and histamine aerosol challenges are common means of bronchial provocation. Because droplet size and the amount of drug delivered vary among aerosol-generating devices and because great care must be taken to monitor responses to avoid provoking severe obstructive episodes, such aerosol challenges should be performed only in laboratories with extensive experience in utilizing these techniques. In addition to the issue of safety, such laboratories have a large database on the distribution of airway responsiveness in a broad population, which allows a more reasonable assessment of the response measured.

An alternative constrictor challenge is 4 or 5 minutes of either exercise or voluntary hyperventilation with cold, dry air. Such approaches appear to be as valuable as drug challenges, and they avoid the problems of variations in aerosol delivery and the dangers of inadvertent overdose with a potent constrictor agent. Results of bronchoprovocation challenges are normally expressed as the provocative dose (PD) of a stimulus that produces a defined level of response in a pulmonary function parameter, most often the FEV_1 (e.g., $PD_{20}FEV_1$ is the dose that produces a 20% decrease in FEV_1) [*see Figure 7*]. The lower the PD_{20}, the greater the degree of responsiveness.

EXERCISE TESTING

Measurements of cardiovascular and respiratory performance at several levels of exercise are useful for the objective evaluation of patients for disability. Cardiopulmonary exercise tests can also be used to evaluate patients with unexplained dyspnea. The tests quantify the severity of the abnormality and may give clues to the etiology of the dyspnea.

Another use of exercise testing is to document the occurrence and severity of hypoxemia during exertion. Assessment of oxy-

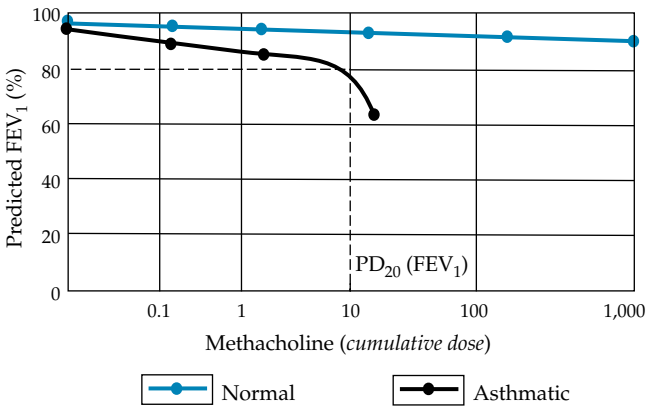


Figure 7 **Bronchoconstrictor challenge results are plotted as change in FEV_1 against cumulative dose of methacholine. Significant reaction is defined as a 20% drop in FEV_1. The dose at which this occurs is termed the PD_{20}. The lower the PD_{20}, the more reactive the airways.**

genation during exercise is particularly useful in two clinical situations. First, patients with restrictive defects caused by diffuse interstitial lung disease who have normal or near-normal $A\text{-}aDO_2$ values at rest often develop hypoxemia and widened $A\text{-}aDO_2$ values during exercise. This response has been attributed to a diffusion barrier between blood and alveolar gas that becomes apparent only during exercise, when contact time in the pulmonary capillaries is diminished by an increase in cardiac output. Second, patients with chronic airway obstruction who have only moderate hypoxemia at rest may experience more severe hypoxemia during exercise. This worsening of hypoxemia in the obstructive syndromes has been attributed to increasing ventilation-perfusion imbalance during the hyperpnea that accompanies exercise.

Hypoxemia can be documented by measuring arterial blood gases before and during exercise or by continuously monitoring capillary hemoglobin saturation by pulse oximetry. In both the restrictive and the obstructive syndromes, development or worsening of hypoxemia during exercise can explain, at least in part, why symptoms become more severe on exertion. This finding indicates the need for oxygen therapy during exercise, an intervention that may significantly improve exercise tolerance.

Oximeter readings during exercise may be falsely low because of vasoconstriction of arterial beds in nonexercising areas, such as the fingers and toes, ears, and skin. Low oxygen saturation recorded by oximetry during exercise should be confirmed by direct arterial blood measurements, especially if the oximetry readings are inconsistent with the clinical situation or if a prescription for portable oxygen during exercise is being considered. A formal exercise testing protocol is performed at graded levels of work output on a treadmill or cycle ergometer. Minute ventilation and tidal volume, heart rate, blood pressure, exhaled O_2 and CO_2 concentrations, and oxygen saturation are continuously monitored and correlated with the patient's perceived level of dyspnea on a visual analogue scale. Maximum oxygen uptake ($\dot{V}O_{2max}$) is reached when there is no further increase in $\dot{V}O_2$ with increasing work output. Anaerobic metabolism, reflecting a disproportionate increase in CO_2 output and ventilation with increasing $\dot{V}O_2$, occurs as a result of buffering the increase in lactic acid and signals the onset of anaerobic glycolysis; it is termed the anaerobic threshold. Common patterns of findings of exercise testing have been correlated with heart disease, obstructive and restrictive pulmonary diseases, pulmonary vasculopathies, and deconditioning.[13]

For severely disabled patients who are candidates for lung transplantation or lung volume reduction surgery, exercise capacity can be determined with the 6-minute-walk distance, a very simple test that requires only oximetry, a timepiece, and a measured distance. The 6-minute-walk distance test has been standardized and shown to be repeatable.[14]

POLYSOMNOGRAPHY

The term polysomnography is used to describe a collection of measurements made during sleep to assess sleep-associated respiratory and neurologic disturbances. These are discussed in detail elsewhere [*see 14:VI Ventilatory Control during Wakefulness and Sleep and 11:XIII Disorders of Sleep*].

Imaging Studies

STANDARD CHEST RADIOGRAPH

In patients with a pulmonary disorder, the relative lack of specificity of respiratory tract symptoms and the relatively poor reliability of the pulmonary physical examination make radiologic imaging techniques essential to the diagnostic process [*see Table 3*]. For this reason, the chest roentgenogram (posteroanterior and lateral) is the single most commonly utilized radiologic technique, accounting for some 40% of all radiologic procedures.

Standard radiographic imaging techniques depend on large differences in the densities of the various body components to discern structure and detect abnormalities. Four distinct densities—gas, water (solid tissues), fat, and metal (bones)—can be identified. Radiographic imaging brings the thoracic cage and the mediastinal structures into sharp contrast with the adjacent gas-containing lung, and any lesion within the lung is made apparent by this same contrast in densities. Several technically sophisticated diagnostic imaging techniques are also employed in the assessment of diseases of the thorax. These techniques include perfusion scans, ventilation scans, pulmonary angiography, computed tomography, magnetic resonance imaging, and ultrasonography.

RADIONUCLIDE SCANS

Ventilation Scans

Ventilation scans are used either to assess the distribution of pulmonary ventilation or to improve the diagnostic reliability of radionuclide perfusion scans performed for the diagnosis of pulmonary embolic disease. Ventilation scans can be performed with radioactive gases (usually xenon) or with a fine aerosol of particles labeled with technetium-99m (^{99m}Tc). When a gas is used, images are made after a wash-in phase (at equilibrium) and during a timed wash-out phase. Regions of lung that are poorly ventilated retain radioactivity. If a radioactive aerosol is used, multiple views can be obtained for comparison with the perfusion scan.

Perfusion Scans

Perfusion scanning is employed to assess the distribution of blood flow to the lung, either as a diagnostic tool for the evaluation of pulmonary embolic disease [*see 1:XVIII Venous Thromboembolism*] or to quantitate regional pulmonary blood flow for the assessment of physiologic operability for surgical pul-

Table 3 Summary of Imaging Techniques

Technique	*Indications*	*Contraindications*	*Comments*
Posteroanterior and lateral roentgenogram*	Wide variety of pulmonary disorders	—	Technique dependent; specificity and sensitivity are observer dependent and vary with disease process and clinical setting
Contrast angiography* Digital subtraction angiography*	Pulmonary thromboembolic disease	Contrast sensitivity, renal insufficiency, left bundle branch block	Gold standard for diagnosis of pulmonary embolism; used for evaluation of pulmonary vasculature; indicated when other studies are inconclusive and when treatment is invasive, hazardous, or contraindicated
Radionuclide scanning* Ventilation Perfusion	Pulmonary thromboembolic disease	Reports of complications, anecdotal only	Used to assess distribution of pulmonary ventilation and blood flow and in assessment of physiologic suitability for pulmonary resection
Computed tomography*	Emphysema, pulmonary nodules, infectious processes, parenchymal pulmonary disorders	Obesity (> 300 lb), inability to hold breath, contrast sensitivity	CT reconstructions provide sagittal sections of 10 mm thickness; increased sensitivity compared with plain film but with 100× radiation exposure, especially suited for evaluation of mediastinal anatomy and the pleura and for staging malignancies
Spiral CT*	Pulmonary thromboembolism		Spiral CT shortens breath-hold time; can be used for three-dimensional reconstructions of airway anatomy and lung masses
High-resolution CT*	Parenchymal pulmonary disorders, bronchiectasis		1–2 mm slices allow resolution of secondary pulmonary lobules by high-resolution CT
Magnetic resonance imaging MR angiography	Disorders of the chest wall, mediastinum, and soft tissues Pulmonary thromboembolic disorders	Presence of any ferrometallic objects, claustrophobia	No radiation exposure; provides coronal as well as sagittal reconstructions; resolves soft tissue anatomy
Positron emission tomography*	Moderate to large focal pulmonary lesions	—	Accuracy uncertain for lesions smaller than 1 cm
Ultrasonography	Pleural effusions	—	May help localize subpleural pulmonary nodules

*Radiation exposure requires caution for women of childbearing age.

monary resection. Isotopes that emit gamma rays, most often ^{99m}Tc, are attached to macroaggregated human albumin or to albumin microspheres having particle diameters of 10 to 50 μm. The radiolabeled particles are then injected into a peripheral vein and subsequently lodge in pulmonary capillaries. The distribution of the resulting emboli is proportional to the regional blood flow. Scanning is capable of resolving 3% of lung volume, with higher reliability for the left than for the right lung.

Because any pathophysiologic process that impedes circulation in that region may cause the abnormalities, a defect seen on perfusion scan is never in and of itself diagnostic of a specific disease or pathophysiologic process. The scan must be interpreted in the context of integrated information derived from clinical evaluation and a concurrent chest radiograph. Further, identification and assessment of the abnormalities are highly subjective. Only extreme scan results, such as totally normal perfusion or the absence of perfusion to a segment or lobe, are relatively free from observer variability.

Ventilation scans are almost always performed in conjunction with perfusion scans and are most useful in assessing the probability of pulmonary emboli. If a region with no perfusion ventilates well (ventilation-perfusion mismatch), there is an increased likelihood of pulmonary embolic disease. However, if ventilation and perfusion to a given region are both diminished (ventilation-perfusion matched defect), embolic disease is less likely. Perfusion scanning alone in combination with clinical history has been shown to have a positive predictive value of 0.99 and a negative predictive value of 0.97, suggesting that accurate diagnosis of pulmonary embolism is possible without benefit of a ventilation scan.[15] The degree of scintigraphic obstruction correlates well with pulmonary hemodynamics in acute pulmonary embolism, but less so in chronic pulmonary hypertension because of the presence of vascular remodeling.

According to the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study (which used pulmonary angiography as the standard), two or more large segmental mismatches, two or more moderate segmental mismatches and one large defect, or at least four moderate mismatches define a high probability scan that indicates an 80% or greater likelihood of pulmonary embolism. A nonsegmental perfusion defect, a perfusion defect much smaller than corresponding opacity on an accompanying chest roentgenogram, or three or fewer small matched defects in the presence of a normal chest film define a low-probability scan. A single moderate-sized mismatch or a moderate and a large defect, more than three matched defects in a lung zone or four in a whole lung, or a pattern that is neither high nor low probability defines an indeterminate scan. The positive predictive value of the ventilation-perfusion scan increases as the number and volume of mismatched defects increases. Two segmental mismatches carry an 80% positive predictive value for embolism.[16] As opposed to a completely normal study, 11% (8/75) of patients with a nearly normal ventilation-perfusion scan were subsequently found to have positive angiograms for pulmonary embolism. The PIOPED II study, which is currently under way, is assessing helical (spiral) CT in relation to pulmonary angiography, ventilation-perfusion scans, digital subtraction angiography, and lower extremity venous compression ultrasonography and contrast venography. PIOPED II is likely to generate new sensitivity and specificity data for these techniques. This study will be completed in 2004.[17]

Unfortunately, in more than 60% of cases, results of ventilation-perfusion scanning can neither confirm nor exclude pulmonary embolism. In a patient with normal D-dimer levels, an equivocal scan can reliably exclude pulmonary embolism.[18] In patients with an equivocal scan and elevated D-dimer levels, further clinical evidence is required [*see 1:XVIII Venous Thromboembolism*]. Positive noninvasive leg vein studies (plethysmography or venous Doppler ultrasonography) carry the same therapeutic implications as a definitive diagnosis of pulmonary embolism.

Pulmonary Angiography

Pulmonary angiography is currently the gold standard for the diagnosis of pulmonary thromboembolism [*see 1:XVIII Venous Thromboembolism*]. There are four indications for the use of pulmonary angiography in patients who may have pulmonary thromboembolic disease: (1) to confirm the diagnosis when anticoagulation carries significant risks to the patient or when a younger patient may face prolonged anticoagulation; (2) when aggressive therapy such as thrombolysis, vena caval interruption, or surgical clot extraction is contemplated; (3) to provide a definitive diagnosis when the clinical suspicion of thromboemboli is high, but less invasive studies are inconclusive; and (4) in the evaluation of pulmonary hypertension. If leg vein studies are negative and the clinical probability of pulmonary embolism is great, a pulmonary angiogram should be done because it is currently the only way to positively exclude pulmonary embolism. Pulmonary angiography as currently practiced is safe. Mortality is less than 1%. Major complications, such as hypotension, myocardial infarction, and renal failure, occur in 1.5%, and minor complications, particularly contrast nephrotoxicity, occur in 4.8%.[19] Most cases of renal failure occur in critically ill patients. Other risks include allergy to contrast media (0.1%), transient right bundle branch block, and nonsustained ventricular tachyarrhythmias. Pulmonary arterial pressure rises by an average of 4 mm Hg, linearly dependent on the size of the injectate. The rise in pulmonary arterial pressure is independent of the presence of pulmonary embolism. Most angiographic defects in pulmonary embolism are found in segmental or larger vessels. Reproducibility of interpretation is reduced in the 17% of patients with defects in subsegmental vessels.[20] When angiography is unavailable, excessively risky, or contraindicated, alternative imaging techniques may be considered.

Digital Subtraction Angiography

In digital subtraction angiography (DSA), images are taken before and after the injection of intravenous contrast medium, and a computer then subtracts the plain image from the contrast image. This enhances the resolution of the contrast image. DSA has been shown to be sufficiently sensitive to confidently exclude pulmonary embolism. Twelve-month follow-up of 54 patients with negative DSA examinations showed no indications of a missed diagnosis of pulmonary embolism, whether in the form of subsequent clinical events, positive angiograms, or positive scans.[21] The PIOPED II study (see above) will provide data on the utility of DSA in relation to other techniques.

COMPUTED TOMOGRAPHY

CT scanning utilizes standard radiographic signals that are processed by computer to provide detailed cross-sectional images of desired contrast. Although CT scanning offers significantly better visualization of anatomic structures than does standard radiography, a thoracic CT examination involves a radiation exposure equivalent to 100 standard chest radiographs. Risks and benefits must always be taken into consideration when ordering CT examinations.

CT scanning distinguishes gradations of density ranging from air (–1,000 Hounsfield units [HU]) to bone (+1,000 HU). Normal tissue has the density of water, 0 HU. Intravenous contrast media in modest doses allow CT scanning to be used to assess the intrapulmonary-versus-extrapulmonary location of a pulmonary lesion, the vascularity of lesions, and tracheobronchial pathology; to diagnose bronchiectasis; and to identify major vascular structures in regions of abnormality. The technique is employed to locate and characterize mediastinal masses and lymph nodes, intrapulmonary lesions, and pleural processes. CT scanning is now considered integral to the staging of most cases of lung cancer and to the evaluation of most mediastinal masses. CT scanning before bronchoscopy for suspected lung cancer increases the bronchoscopic yield from 71% to 90%.[22]

High-Resolution Computed Tomography

In high-resolution CT (HRCT), the thickness of the tomographic section is reduced from about 1 cm to 1 to 3 mm. Modified software improves resolution and shows fine morphologic detail. The technique is most useful for evaluating diffuse lung disease, and usually, only three to six scans are performed. Resolution of structures down to 200 μm can be achieved, so that secondary pulmonary lobules and even thickened intralobar septa can be visualized.

HRCT is a good screening test for interstitial lung disease, especially in patients with dyspnea whose chest roentgenogram is normal and whose lung volumes and diffusing capacity are in the low normal range or are mildly reduced. Negative results on HRCT are evidence against significant interstitial lung disease, although open lung biopsy specimens may still show evidence of abnormality in some cases. HRCT may be useful in assessing drug-induced lung damage; air-space abnormalities can be seen even when the chest roentgenogram is normal.

Certain patterns have been recognized and closely correlated with specific pathologic processes[23] [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]:

1. A ground-glass pattern correlates with active alveolitis. A higher proportion of ground-glass density in relation to septal thickening or honeycombing patterns suggests a more favorable prognosis in usual interstitial pneumonia.
2. Active tuberculosis is characterized on HRCT by ground-glass densities and centrilobular and poorly marginated nodules, whereas nontuberculous *Mycobacterium* infections frequently present as bronchiectasis associated with parenchymal nodules.
3. Idiopathic pulmonary fibrosis is characterized by crescentic subpleural reticular densities that are most dense in the posterior basal segments of the lower lobes. Scans taken with the patient in the prone position can ensure that such densities are not a result of increased perfusion of dependent zones. In advanced disease, reticular interstitial densities, regions of fibrosis, and cystic spaces are found throughout the lung. These changes are not specific to idiopathic pulmonary fibrosis but are seen in end-stage fibrotic lung disease from any cause.
4. Lymphangitic cancer has a rather characteristic appearance, with thickening of bronchovascular bundles and intralobar

septa. The thickened septa are often very prominent subpleurally.

5. Hypersensitivity pneumonitis may show an extensive interstitial and air-space infiltrate, giving a ground-glass appearance to large areas of the lung.
6. Parenchymal sarcoidosis has a wide range of appearances. The most characteristic abnormalities are nodular opacities along the bronchovascular bundles and, to a lesser degree, in the interlobular septa and pleura.
7. In lymphangioleiomyomatosis, numerous small cystic air spaces that are distributed widely throughout the lung can be detected, even when the chest roentgenogram is unremarkable.
8. Histiocytosis X is characterized by a combination of pulmonary nodules and cysts, most prominent in the upper lobes.

HRCT helps localize the distribution and severity of emphysema. This may be of importance in the evaluation of patients for lung volume reduction surgery. HRCT can clarify the extent and pattern of intrathoracic disease; if patterns are typical, they can be used in lieu of an open biopsy technique to establish a specific diagnosis of interstitial disease.[23]

Spiral Computed Tomography

Helical, or spiral, CT differs from conventional CT imaging by continuously acquiring imaging data while the patient moves through the CT gantry at a constant rate. Spiral CT has become the standard approach for thoracic imaging in many institutions because imaging time is substantially reduced, often to less than 1 minute. Contrast requirements are also minimized. Reconstructions can be constituted at any level of the thorax, decreasing the likelihood that a small lesion in between tomographic cuts could be missed. The rapidity of image acquisition allows a correlation of lung anatomy and physiology between inspiration and expiration.

Spiral CT is capable of resolving second- to fourth-generation pulmonary arterial branches. It has been shown to be a highly sensitive and specific tool in the diagnosis of central pulmonary embolism that, in most studies, offers performance superior to that of radionuclide imaging [*see Figure 8*]. A major advantage of spiral CT is its ability to demonstrate alternative diagnoses and ancillary findings of pulmonary embolism, such as right ventricular dilatation, pleural-based wedge-shaped infiltrates, linear banding, and increase in the size of affected arteries. Because of these advantages, the number of indeterminate studies is significantly reduced compared with ventilation-perfusion scanning. Spiral CT may be sufficiently sensitive to be used as a basis for withholding anticoagulation for suspected pulmonary embolism if combined with other clinical and laboratory data. However, this point needs confirmation from PIOPED II (see above). Unsuspected pulmonary embolism has been reported as an incidental finding in 5% of inpatient spiral CT examinations; most of these patients had cancer. Whether spiral CT should replace ventilation-perfusion scanning in the standard workup of pulmonary embolism has not been definitively resolved[24] and is currently being assessed in PIOPED II.

Three-dimensional reconstruction is possible from spiral CT data and has been used to create accurate virtual bronchoscopic examinations of the airways. Although these reconstructions do not identify subtle bronchial wall abnormalities, they may prove useful in identification of sites suitable for transbronchial needle sampling of lymphadenopathy and in the assessment of bronchial anatomy distal to high-grade obstructing lesions, which are inaccessible to fiberoptic endoscopes.[25]

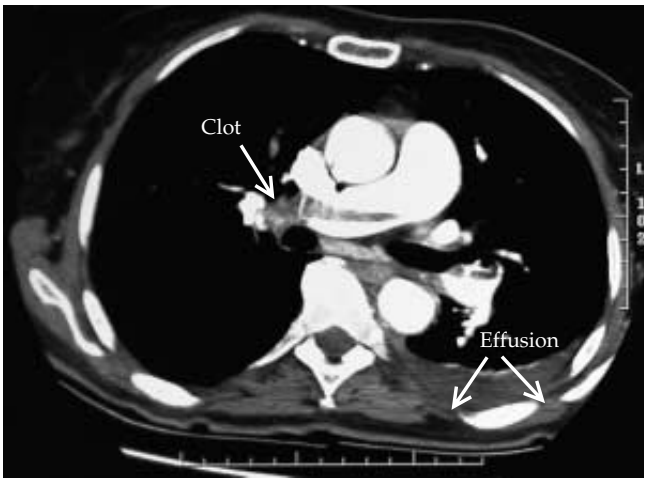


Figure 8 **A large filling defect is seen in the right pulmonary artery in this contrast-enhanced study. Note accompanying pleural effusion.**

MAGNETIC RESONANCE IMAGING

MRI reveals tissue properties by using physical principles different from those used and revealed in standard chest roentgenograms and CT. Because MRIs of the chest can be degraded by motion, thoracic MRIs are usually gated to the heart rate by ECG. MRI produces sagittal and coronal images that are 3 to 10 mm in thickness. Advantages of MRI are its increased soft tissue contrast, multiplanar capability, sensitivity to blood flow, and lack of ionizing radiation. Because of its ability to differentiate between tissues of varying compositions, MRI is particularly suited to evaluation of soft tissue processes. Disadvantages of MRI include inferior spatial resolution, degradation of image with respiratory motion, and the necessity of eliminating metallic objects from the imaging environment, making the method unsuitable for ventilated patients as well as those with metallic implants of any sort. The ability of MRI to differentiate contiguous structures makes it a useful technique for evaluating intrathoracic malignancies in relation to vascular and mediastinal structures. CT and MRI are comparable in accuracy for the evaluation of nodal spread of bronchogenic carcinoma.

MAGNETIC RESONANCE ANGIOGRAPHY

The advent of rapid imaging techniques has made magnetic resonance angiography (MRA) an attractive technology to supplement or supplant pulmonary angiography in the evaluation of pulmonary vascular disorders. MRA has been compared with pulmonary angiography in the diagnosis of pulmonary embolism. For segmental and larger pulmonary embolism, MRA is quite specific and sensitive. In limited studies, magnetic resonance venography has shown high sensitivity and specificity for diagnosis of deep vein thrombosis and may prove an important diagnostic modality.[24]

POSITRON EMISSION TOMOGRAPHY

Positron emission tomography (PET) utilizing 18-fluorodeoxyglucose as the positron emitter (to indicate active glucose metabolism) appears to be an accurate noninvasive test for the diagnosis of moderate sized and larger pulmonary focal le-

sions. Fewer data exist for lesions smaller than 1 cm. A sensitivity of 96.8% and a specificity of 77.8% have been reported from a meta-analysis.[26]

ULTRASONOGRAPHY

The normal air-containing lung is poorly visualized by two-dimensional ultrasonography, but in the presence of consolidation, ultrasonography may prove useful in the evaluation of a variety of pleuropulmonary disorders. Pleural effusions present as hypoechoic or nonechoic regions. Pleural lesions in suspected malignant effusion were visualized and biopsies performed with 100% specificity and 88% sensitivity in one study.[27] Consolidation is seen as hypoechoic regions that move with the patient's breathing. Air bronchograms are hyperechoic, whereas liquid-filled airways associated with atelectasis or endobronchial obstruction are hypoechoic. Abscesses are seen as irregularities in density in an otherwise homogeneous region, and tumors may be distinguished within a consolidated area by their relatively homogeneous consistency and by their regular borders. The specificity and sensitivity of the method are insufficient to permit ultrasonography to replace CT or other radiologic methods, but it may be viewed as a complementary technique.

Sampling Techniques

Several techniques are available for obtaining material for microscopic examination or, in the case of infectious agents, growing in culture media. Some sampling techniques are clearly superior in terms of yield, specificity, ease of performance, and safety. In any clinical setting, however, the local experience with, and availability of, specific sampling techniques will dictate the best choice in a given situation. In general, it is advisable to start with the least invasive approaches.

FIBEROPTIC BRONCHOSCOPY

Fiberoptic bronchoscopy has become the standard procedure for exploring the tracheobronchial tree. The fiberoptic bronchoscope possesses excellent optics, is flexible and easily manipulated, and has a small diameter. Topical anesthesia provides adequate comfort, and endotracheal intubation is generally not needed. These factors contribute to ease of performance and the excellent level of patient acceptance associated with this technique. From a diagnostic standpoint, the fiberoptic bronchoscope has virtually replaced the rigid bronchoscope, except in cases involving examination of the tracheobronchial tree during active massive hemoptysis or for extraction of aspirated foreign bodies [*see Table 4*].

In addition to allowing excellent visualization of the bronchi, the fiberoptic bronchoscope provides samples that can be used for smear, culture, cytologic study, and histologic examination.

Direct Vision

Contemporary fiberoptic bronchoscopes allow viewing of the first six generations of the airways from the posterior oropharynx and nasopharynx and the larynx to sub-subsegmental bronchi. Although this corresponds only to the inner third of the lung fields on a chest roentgenogram, a significant proportion of pathologic processes, including many lung tumors, aspirated foreign bodies, inhalation and aspiration injuries, and Kaposi sarcoma, occur in the proximal airways and are amenable to direct inspection. Localization of the source of hemoptysis is made by direct visualization, although it should be noted that in massive hemoptysis, a rigid bronchoscope may be preferable for its superior visualization and ability to maintain a clear airway.

Washings

Washings from the tracheobronchial tree using isotonic saline are useful for cytologic and microbiologic examination. Bronchial washing is particularly helpful in the diagnosis of sputum smear–negative tuberculosis, in which 40% of smears and up to 95% of cultures may be positive.

Bronchoalveolar lavage If the endoscopist wedges the bronchoscope in a segmental or subsegmental orifice before instilling and aspirating saline, the distal airways and air spaces may be sampled. Bronchoalveolar lavage (BAL) has proved extremely useful in the assessment of opportunistic infections in AIDS patients. The sensitivity of BAL for *Pneumocystis carinii* approaches 95% when compared with open lung biopsy and may exceed that of transbronchial biopsy. The yield for other opportunistic infections may be somewhat lower. Culture of BAL specimens has a sensitivity for tuberculosis of up to 88%. BAL has also been used in the evaluation of interstitial lung diseases and in the identification and assessment of rejection of lung transplants. The total number of cells recovered and the differential cell count help categorize alveolitis as neutrophilic or lymphocytic in nature and to gauge its intensity. BAL may be

Table 4 Common Indications and Contraindications for Fiberoptic Bronchoscopy

COMMON INDICATIONS
Inspection of Tracheobronchial Tree
Search for source of bleeding in patient with mild to moderate hemoptysis
Suspected endobronchial lesion (e.g., in patient with lobar collapse)
Search for occult cancer in patient with malignant cells in sputum but normal chest radiograph
Assessment of intractable cough
Staging of known lung cancer
Sampling from Lower Respiratory Tract
Microbiologic
Pulmonary infiltrate in immunocompromised host
Complicated pneumonia in immunocompetent host
Cytologic and histologic
Suspected pulmonary neoplasia
Diagnosis of unexplained pulmonary infiltrates
Aspiration of paratracheal or parabronchial lymph nodes
Cell count and differential
Assessing activity of chronic interstitial pneumonia (investigational)
Therapeutic Interventions
Suctioning mucous plugs
Guided intubation
Removal of foreign bodies (infrequent)
Placement of endobronchial stents
Endobronchial brachytherapy
CONTRAINDICATIONS
Uncooperative patient
Uncorrectable hypoxemia or hypercapnia
Uncorrectable bleeding diathesis
Severe asthma

diagnostic for eosinophilic granuloma when Langerhans cells are recovered in the washings. Milky BAL fluid indicates alveolar proteinosis; such fluid stains positively with the periodic acid–Schiff (PAS) reaction.

Brushings

Specimens obtained from the stiff bristles of a wire brush that can be passed through the bronchoscope may contribute to the cytologic diagnosis of pulmonary malignancies. Bronchial brushings have a much higher diagnostic yield for endobronchially visible proximal lesions but may be positive in 20% to 40% of peripheral lesions when the brushing is fluoroscopically guided. Lesions larger than 2 cm in diameter and lesions located more centrally tend to have higher cytologic yields, as do lesions that directly involve the bronchial tree. Bronchial brushing may contribute material for microbiologic smears and cultures but is unlikely to provide the sole positive result of a diagnostic evaluation.

Biopsy

Bronchoscopic biopsy forceps produce 2 to 3 mm specimens of lung tissue by tearing tissue away from the lung or bronchus. The biopsy is not painful for the patient. It is, however, associated with bleeding, which is usually minor but is occasionally significant or even life threatening.

Endobronchial Endobronchial biopsy of visible lesions through the bronchoscope was shown to be quite sensitive when compared with the aggregate yield of all other available diagnostic modalities, including bronchial brushings and washings, surgical biopsies, bronchoscopic and nonbronchoscopic needle biopsies, pleural fluid cytology, and biopsy of metastases. False negative results are sometimes encountered because a lesion has an overlying necrotic surface or because the process is more than 2 to 3 mm below the mucosal surface. Negative results in a patient who is strongly suspected of having a bronchogenic carcinoma should lead to a repeat procedure and biopsy attempt. CT scanning before the procedure improves the yield (see above).

Transbronchial Transbronchial lung biopsies provide small specimens of lung parenchyma that can be examined histologically and cultured. This technique is commonly used to sample focal and diffuse lesions beyond the range of direct vision. Under fluoroscopic guidance, the biopsy forceps are advanced to a region of interest, and samples are obtained. Usually, four to six individual biopsies provide acceptable diagnostic sensitivity. Transbronchial biopsy for diffuse parenchymal pulmonary disorders is helpful in the diagnosis of granulomatous disorders, such as sarcoidosis, and in infections or metastatic malignancy. However, its value in the evaluation of the nonlymphocytic alveolitides, such as idiopathic interstitial pulmonary fibrosis, is a subject of considerable debate. Chest radiographs are routinely obtained after transbronchial forceps biopsy procedures because pneumothorax complicates 5% to 10% of these cases. Approximately half of patients who experience pneumothorax after transbronchial lung biopsy require chest tube drainage.

Microbiologic Sampling

Quantitative culture methods using bronchoalveolar lavage or a protected catheter brush are useful for documenting bacterial pneumonia and providing guidance for antimicrobial therapy. Careful technique must be observed. Recovery of 10^3 organisms/ml by protected catheter brush, recovery of 10^4 organisms/ml from the fluid obtained from bronchoalveolar lavage, or the presence of intracellular organisms in more than 2% of alveolar cells usually indicates pneumonia, except in patients with chronic bronchitis or bronchiectasis. In an ICU patient, quantitative cultures are reliable if there has been no change in antibiotics for 48 to 72 hours, and they can be used to prove or exclude significant lung infection when infiltrates and fever of uncertain origin are present.[28] The use of invasive techniques for the diagnosis of bacterial pneumonia remains highly controversial, however; there is no evidence that these procedures improve outcome in severe hospital-acquired pneumonia.[29]

Because appropriate management of community-acquired bacterial pneumonia depends on the rapid (and usually empirical) administration of appropriate antibiotics, these techniques have only limited utility for the diagnosis of pneumonia in this setting. In practice, an invasive and expensive procedure, such as bronchoscopy with quantitative culture, is seldom done before first initiating a trial of antibacterial therapy.

Transbronchial Needle Aspiration

Lesions in the pulmonary parenchyma and in the mediastinum may be accessible to bronchoscopic sampling by transbronchial needle aspiration. A 1.3 cm, 22-gauge cytology needle or a 19-gauge cutting needle may be advanced through the sampling channel of a bronchoscope and passed through the bronchial wall to obtain cytologic or histologic material from tumors or adjacent lymph nodes. Transbronchial needle aspiration is particularly useful for the nodal evaluation of intrathoracic malignancy and for pulmonary nodules that are neither central nor peripheral in location. Correlation of endobronchial anatomy with CT findings is advisable before transbronchial needle aspiration is attempted. Percutaneous transthoracic needle aspiration is still better suited for small peripheral nodules, especially when there is a substantial likelihood of a benign diagnosis.

Complications of Fiberoptic Bronchoscopy

When reasonable precautions are taken, the complication rate of fiberoptic bronchoscopy is quite low, and those complications that do occur tend to be minor. Deaths resulting from bronchoscopy are extremely rare. Major complications—chiefly pneumothorax or severe hemorrhage—occur in 0.5% of routine bronchoscopies and about 7% of transbronchial biopsies. Minor complications, including bronchospasm, laryngospasm, epistaxis, and vasovagal syncope, occur in 0.8%.[30] Fever may result when liquid instilled into airways activates alveolar macrophages to release interleukin-1β (IL-1β), IL-6, and tumor necrosis factor–α.[31] Significant hypoxemia, occasionally leading to respiratory failure, may be seen in immunocompromised patients, especially those undergoing BAL for AIDS-related opportunistic infections. Myocardial ischemia has been objectively demonstrated to occur in elderly patients who undergo prolonged bronchoscopies. Careful monitoring and attention to oxygenation are needed in such patients. Because of its favorable safety profile, fiberoptic bronchoscopy is routinely performed as an outpatient procedure.

PERCUTANEOUS TRANSTHORACIC NEEDLE ASPIRATION

Cutting or cytologic needles may be passed through the chest wall under local anesthesia and fluoroscopic, CT, or ultrasound guidance to obtain tissue samples from the lung, mediastinum,

pleura, chest wall, and hila [*see Figure 9*]. Percutaneous transthoracic needle aspiration (PTNA) should be considered when fiberoptic bronchoscopy has been unsuccessful in obtaining diagnostic tissue or when a patient is inoperable or refuses thoracotomy for diagnosis. The major risks of both needle procedures are pneumothorax, severe hemoptysis, and air embolism. The risk of pneumothorax is considerably higher in patients with chronic obstructive pulmonary disease and varies directly with the amount of pulmonary parenchyma that must be traversed to reach the lesion in question. Pneumothorax occurs in 25% to 45% of cases, but only 4% to 15% require chest tube placement. Hemoptysis is seen in 2% to 16% of cases. Because bleeding originates from noncompressible sites, coagulopathy remains an absolute contraindication to PTNA or transbronchial needle aspiration. PTNA should not be attempted in patients with pulmonary hypertension, suspected arteriovenous malformations, or echinococcal cysts.

Depending on their design, percutaneous needles can be used either to aspirate a lesion—thereby providing samples for cytologic examination, smear, and culture—or to obtain a core of tissue that can provide material for histologic examination; their use carries a risk of pneumothorax. PTNA has been reported to have an 88% sensitivity and a 99% specificity for malignant lung lesions and to have a 95% sensitivity and an 81% specificity for benign lesions.[32]

THORACENTESIS

Thoracentesis, the aspiration of liquid from the pleural space, is a standard diagnostic technique. A pleural effusion without known cause should be sampled by thoracentesis for diagnostic purposes. The aspirated liquid can be analyzed for the number of red and white blood cells and the white blood cell differential; concentrations of glucose, lactate dehydrogenase, and total protein; appropriate cultures and smears; and, if indicated, cytology. Additional studies may be indicated in particular clinical circumstances. For example, measurement of pleural fluid amylase levels is helpful when pancreatitis or esophageal rupture is suspected as the etiology of the pleural effusion. Pleural liquid pH is useful in evaluating parapneumonic effusions. A pH of less than 7.29 (or less than 7.22 in the setting of low probability and higher risk) is associated with a complicated effusion that may require drainage.[33] pH paper should not be used; pH should be determined from an aliquot of pleural liquid placed in a heparinized syringe.

The most common complications of thoracentesis are bleeding and pneumothorax. The use of catheter-based thoracentesis kits has led to an increase in the incidence of retained catheters. The risk of complication from thoracentesis is a function of the experience of the operator and may be increased in certain settings, such as in patients with chronic obstructive lung disease or those who have undergone previous radiation therapy. If a single pass is required for the tap, if no air leak is encountered, and if the operator feels the procedure is uncomplicated, the rate of pneumothorax is less than 1%, and a postthoracentesis chest roentgenogram is not necessary.

PLEURAL BIOPSY

In contrast to thoracentesis, pleural biopsy should be performed by a specialist. The technique is employed in the evaluation of patients with unexplained exudative pleural effusions. The technique is performed using a large-bore needle with a cutting edge; small (2 to 3 mm) tissue samples are obtained blindly from the parietal pleura. Some pleural liquid must be present to ensure proper placement of the needle in the pleural space and to minimize the risk of lung puncture. Only 40% of submitted closed pleural biopsy specimens actually contain pleural tissue. It is advisable to submit tissue from at least six separate passes to obtain optimal diagnostic sensitivity. Its value is greatest in establishing the diagnosis of postprimary tuberculous involvement of the pleura: with an adequate number of biopsies submitted for microbiologic and histologic examination, closed pleural biopsy has a sensitivity of 87% for the diagnosis of tuberculous pleurisy (60% culture, 80% histology).[34] The diagnostic yield of biopsy in confirming malignant invasion of the pleura is only slightly greater than that obtained through cytologic examination of the aspirated liquid.

The major contraindication for these techniques is a bleeding disorder, because it is impossible to apply pressure to a pleural bleeding site without performing an open surgical procedure.

THORACOSCOPY

Thoracoscopy is currently employed in the diagnosis of pleural and mediastinal disorders, for the introduction of talc pleurodesis in the management of malignant pleural effusion or recurrent spontaneous pneumothorax, and for the evacuation of partially loculated empyema. Thoracoscopy has been successfully applied as an alternative to open lung biopsy in the evaluation of diffuse interstitial lung disease, for solitary pulmonary nodules, and for lung volume reduction surgery for advanced emphysema.[35] The technique requires access to the pleura through one or more small intercostal incisions, induction of artificial pneumothorax, and examination through a rigid thoracoscope (medical thoracoscopy) or via telescope attached to a video camera (video-assisted thoracoscopic surgery, or VATS). Medical thoracoscopies can be performed in an endoscopy suite under local anesthesia and conscious sedation, whereas VATS usually requires general anesthesia and double-lumen intubation in the operating room. Complication rates are higher than for mediastinoscopy, and hospital stays are longer.

Thoracoscopy should be undertaken for mediastinal disorders only when lesions are inaccessible to the mediastinoscope.

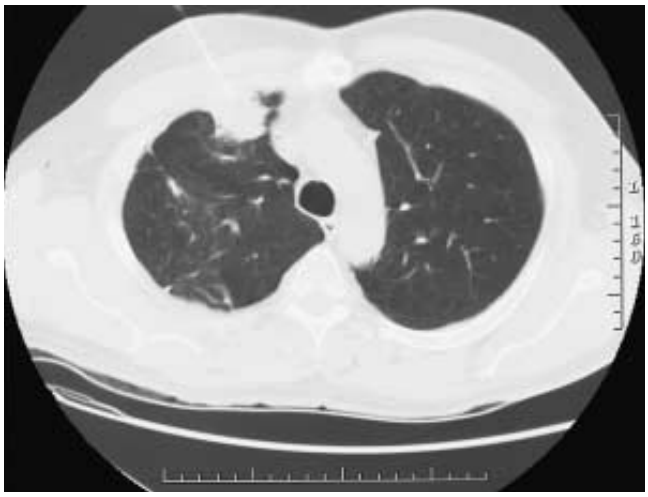


Figure 9 **Under CT guidance, a needle is passed through the chest wall into pleura-based right middle lobe lesions. Peripheral lesions are particularly suitable for this approach, which has a high diagnostic sensitivity.**

Mortality from medical thoracoscopy is reported to be 1% to 5%. Thoracoscopy has a 2% to 3% rate of complications, including prolonged air leak and pleural effusion. Empyema, significant bleeding, and tumor seeding of incisions are also reported. Approximately 10% of thoracoscopies require conversion to open thoracotomy to manage unexpected findings. For this reason, most experts recommend that all thoracoscopic procedures be performed by a thoracic surgeon.

MEDIASTINOSCOPY

Mediastinoscopy and anterior mediastinal exploration through a limited superior parasternal incision (mediastinotomy) are surgical procedures used for diagnostic biopsy of mediastinal masses and the staging of carcinoma of the lung. Mediastinoscopy is performed under general anesthesia and requires a small transverse incision just above the suprasternal notch. Blunt dissection along the pretracheal fascial plane is performed, and paratracheal lymph nodes can be sampled. Mediastinoscopy is especially suited for evaluation of the superior and anterior mediastinum. Mediastinoscopic access to posterior, subcarinal, and some para-aortic nodes is difficult and often necessitates an open exploration through a left-sided second intercostal space incision, the so-called Chamberlain procedure.

Mediastinoscopy is indicated in the nodal staging of lung cancer for superior sulcus tumors; for small cell cancer being considered for resection; for patients with poor ventilatory reserve in whom CT results are discordant with the expected likelihood of nodal metastasis; in evaluation of the superior vena cava syndrome (previously thought to be an absolute contraindication); and in settings of clinical uncertainty.

OPEN LUNG BIOPSY

Noninvasive diagnostic studies are subject to sampling error because of the size of the sample and the frequent nonhomogeneity of the pathologic process of interest. Samples may be crushed or distorted, leading to difficulty in pathologic interpretation. Biopsy by open thoracotomy provides the best obtainable specificity and sensitivity but at the cost of the risk of the open procedure and a painful intercostal incision. Histologic diagnoses are possible in 85% to 95% of cases. New or unexpected diagnoses may be made in almost half of these cases. Surgical mortality is reported in the range of 0% to 13% in immunocompetent patients but as high as 25% to 65% in immunocompromised and mechanically ventilated patients. Mortality is higher when open lung biopsy is performed on an emergent rather than an elective basis.[36] Significant morbidity may occur in another 25% to 50% of patients.[37]

Open biopsy should be considered when the overall prognosis is good, when less invasive modalities have failed to provide a useful diagnosis, and when an empirical therapeutic trial has not resulted in clinical improvement. Decisions for open lung biopsies, however, should give considerable weight as to whether a definitive diagnosis will change therapy and whether that therapy can be expected to lead to improvement in survival. An Israeli study of open lung biopsy in diffuse lung disease reported an 18% benefit, as defined above, at the cost of 13% mortality in immunocompetent patients. For immunocompromised patients, the benefit was 46% but at the cost of 39% operative mortality.[36]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Ernst A, Zibrak JD: Carbon monoxide poisoning. N Engl J Med 339:1603, 1998

2. Wahr JA, Tremper KK, Diab M: Pulse oximetry. Respir Care Clin North Am 1:77, 1995

3. Smetana GW: Preoperative pulmonary evaluation. N Engl J Med 340:937, 1999

4. Bancalari E, Clausen J: Pathophysiology of changes in absolute lung volumes. Eur Respir J 12:248, 1998

5. Coates AL, Peslin R, Rodenstein D, et al: Measurement of lung volumes by plethysmography. Eur Respir J 10:1415, 1997

6. Brown R, Leith DE, Enright PL: Multiple breath helium dilution measurement of lung volumes in adults. Eur Respir J 11:246, 1998

7. American Thoracic Society: Standardization of spirometry. 1994 update. Am J Respir Crit Care Med 152:1107, 1995

8. Pedersen OF, Brackel HJ, Bogaard JM, et al: Wave-speed–determined flow limitation at peak flow in normal and asthmatic subjects. J Appl Physiol 83:1721, 1997

9. Lunn WW, Sheller JR: Flow volume loops in the evaluation of upper airway obstruction. Otolaryngol Clin North Am 28:721, 1995

10. ATS/ERS Statement on respiratory muscle testing. Am J Respir Crit Care Med 166:518, 2002

11. Venegas JG, Harris RS, Simon BA: A comprehensive equation for the pulmonary pressure-volume curve. J Appl Physiol 84:389, 1998

12. American Thoracic Society: Guidelines for methacholine and exercise testing. Am J Respir Crit Care Med 161:309, 2000

13. Jones NL, Kieran JK: Exercise limitation in health and disease. N Engl J Med 343:632, 2000

14. ATS statement: guidelines for the six minute walk test. Am J Respir Crit Care Med 166:111, 2002

15. Miniati M, Pistolesi M, Marini C, et al: Value of perfusion lung scan in the diagnosis of pulmonary embolism: results of the Prospective Investigative Study of Acute Pulmonary Embolism Diagnosis (PISA-PED). Am J Respir Crit Care Med 154:1387, 1996

16. Ralph DD: Pulmonary embolism: the implications of prospective investigation of pulmonary embolism diagnosis. Radiol Clin North Am 32:679, 1994

17. Gottschalk A, Stein PD, Goodman LR, et al: Overview of prospective investigation of pulmonary embolism diagnosis II (PIOPED II). Semin Nucl Med 32:183, 2002

18. Kruip MJ, Slob MJ, SchijenJH, et al: Use of a clinical decision rule in combination with D-dimer in diagnostic work up of patients with suspected pulmonary embolism. Arch Intern Med 162:1631, 2002

19. Hudson ER, Smith TP, McDermott VG, et al: Pulmonary angiography performed with iopamidol: complications in 1,434 patients. Radiology 198:61, 1996

20. Diffin DC, Leyendecker JR, Johnson SP, et al: Effect of anatomic distribution of pulmonary emboli on interobserver agreement in the interpretation of pulmonary angiography. AJR Am J Roentgenol 171:1085, 1998

21. Forauer AR, McLean GK, Wallace LP: Clinical follow-up of patients after a negative digital subtraction pulmonary arteriogram in the evaluation of pulmonary embolism. J Vasc Interv Radiol 9:903, 1998

22. Laroche C, Fairbairn I, Moss H, et al: Role of computed tomographic scanning of the thorax prior to bronchoscopy in the investigation of suspected lung cancer. Thorax 55:359, 2000

23. American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. Am J Respir Crit Care Med 165:277, 2002

24. ACCP Consensus Committee on Pulmonary Embolism: Opinions regarding the diagnosis and management of venous thromboembolic disease. Chest 113:499, 1998

25. Fleiter T, Merkle EM, Aschoff AJ, et al: Comparison of real-time virtual and fiberoptic bronchoscopy in patients with bronchial carcinoma: opportunities and limitations. AJR Am J Roentgenol 169:1591, 1997

26. Gould MK, Maclean CC, Kuschner WG, et al: Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. JAMA 285:914, 2001

27. Adams RF, Gleeson FV: Percutaneous image-guided cutting-needle biopsy of the pleura in the presence of suspected malignant effusion. Radiology 219:510, 2001

28. Torres A, el-Ebiary M: Invasive diagnostic techniques for pneumonia: protected specimen brush, bronchoalveolar lavage, and lung biopsy methods. Infect Dis Clin North Am 12:701, 1998

29. Sanchez-Nieto JM, Torres A, Garcia-Cordoba F, et al: Impact of invasive and noninvasive quantitative culture sampling on outcome of ventilator-associated pneumonia: a pilot study. Am J Respir Crit Care Med 157:371, 1998

30. Pue CA, Pacht ER: Complications of fiberoptic bronchoscopy at a university hospital. Chest 107:430, 1995

31. Krause A, Hohberg B, Heine F, et al: Cytokines derived from alveolar macrophages induce fever after bronchoscopy and bronchoalveolar lavage. Am J Respir Crit Care Med 155:1793, 1997

32. Lacasse Y, Wong E, Guyatt GH, et al: Transthoracic needle aspiration biopsy for the diagnosis of localised pulmonary lesions: a meta-analysis. Thorax 54:884, 1999

33. Light RW: Pleural effusion. N Engl J Med 346, 2002

34. Kirsch CM, Kroe DM, Azzi RL, et al: The optimal number of pleural biopsy specimens for a diagnosis of tuberculous pleurisy. Chest 112:702, 1997

35. Hansen M, Faurschou P, Clementsen P: Medical thoracoscopy, results and complications in 146 patients: a retrospective study. Respir Med 92:228, 1998

36. Kramer MR, Berkman N, Mintz B, et al: The role of open lung biopsy in the management and outcome of patients with diffuse lung disease. Ann Thorac Surg 65:198, 1998

37. Neuhaus SJ, Matar KS: The efficacy of open lung biopsy. Aust N Z J Surg 67:181, 1997

II ASTHMA

Mitchell H. Grayson, m.d.
Michael J. Holtzman, m.d.

Asthma is a major public health issue. It affects nearly 5% of the United States population and is the reason for approximately two million emergency department visits, 470,000 hospitalizations, and 4,500 deaths annually.[1] Mortality has declined over the past 10 years, but rates remain 2.5 to 3 times higher in blacks than in whites.[2] Some evidence indicates that the quality of life of patients with asthma has improved with the development of new medications; however, inappropriate use of medications in the treatment of this disease is widespread. To address these issues, the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health and the World Health Organization (WHO) generated several expert panel reports on the diagnosis and treatment of asthma, including the National Asthma Education Prevention Program (NAEPP).[1] This chapter reviews the current understanding of asthma pathogenesis, diagnosis, and treatment.

Pathogenesis

Asthma has been defined as a chronic inflammatory disorder of the airways that is associated with recruitment of inflammatory cells and the clinical development of wheezing, shortness of breath, chest tightness, and cough. There is widespread but variable airflow obstruction that is often reversible, either spontaneously or as a result of treatment. In addition, asthmatic patients exhibit airway hyperreactivity that can be assessed by measuring the increased bronchoconstrictor response to inhaled agents such as histamine or methacholine. It has been proposed that airway hyperreactivity, as well as the clinical signs and symptoms of asthma, is driven by persistent airway inflammation.

Allergic response

The precise basis for the development of airway inflammation in patients with asthma is not fully defined. One traditional scheme postulates the development of an allergic response in the airways. In this case, the inflammatory cascade may begin when an antigen is inhaled into the airway and is taken up and processed by an antigen-presenting cell, usually a dendritic cell. The dendritic cell then migrates from the airway mucosa to the regional lymph nodes. Upon arrival at the lymph node, the dendritic cell, which has matured along the journey, presents the processed antigenic peptides to T cells. Rare antigen-specific T cells sample the peptides being presented in the lymph node; when an antigen-specific T cell finds a dendritic cell that presents the peptides for which the T cell is specific, it begins to proliferate [*see Figure 1*]. Simultaneously, the dendritic cell, through various signals, attempts to skew the cytokine profile of the developing T cells to cause them to become T helper type 1 (Th1) cells (which primarily produce interferon gamma [IFN-γ]) and interleukin 12 [IL-12]) or Th2 cells (which produce IL-4, IL-5, and IL-13). These cytokine profiles may be found in both $CD4^+$ and $CD8^+$ T cells; both types of T cells have been implicated in causing asthma.

The allergic response in general and the asthmatic allergic response in particular depend on a Th2 response to an antigen.[3] Th2 cells produce IL-4, which causes B cells to switch from usual production of IgM or IgG antibody to production of IgE antibody. Once produced, IgE antibodies bind to the surface of mast cells and basophils, where subsequent antigen cross-linking causes the generation and release of mediators that drive the asthma phenotype.

In an experimental model of the allergic response to inhaled antigen, Th2 cells alone were not sufficient to cause the asthma phenotype, because Th2 cell recruitment to the allergic site depends on help from Th1 cells.[4] Th1 cells are more readily recruited to the airway tissue and are responsible for tumor necrosis factor (TNF)–dependent expression of vascular cell adhesion molecules; the presence of Th1 cells in the airway tissue in turn allows Th2 cells to enter the tissue. The Th1 cells do not need to be of the same antigenic specificity as the Th2 cells.[5] This finding may help explain why inflammatory stimuli, such as respiratory viral infections, may facilitate the allergic response and contribute to flares of allergic asthma.

Allergen itself may also be sufficient to produce airway inflammation. When a sensitized individual is reexposed to allergen, the subsequent cross-linking of IgE on the surface of the mast cells and basophils can lead to the release of immediate-phase reactants such as histamine and TNF. Within 4 to 6 hours after exposure, the activated immune cells produce chemokines and cytokines, proteinases, enzymes, and lipid mediators such as the cysteinyl leukotrienes. Chemokines recruit and help activate additional Th2 cells and eosinophils in the airway. Proteinases and other enzymes may lead to airway damage, which promotes collagen deposition in the subepithelial basement membrane region. Cysteinyl leukotrienes can cause airway smooth muscle constriction and mucous cell secretion that further obstruct the airway lumen.

Leukotrienes in concert with chemokines (e.g., eotaxin) and cytokines (e.g., IL-5) may also recruit and activate eosinophils in the airway. Eosinophils are granulocytes that have a very short life span in the periphery and do not appear in large amounts in the circulation. These cells are produced in the bone marrow under the influence of Th2 cell production of IL-5. Newly produced eosinophils are released into the circulation and home in on the lung tissue, guided by a similar set of cell adhesion molecules that direct Th2 cell traffic. Upon entry into airway tissue, eosinophils and Th2 cells secrete their products, which, in addition to the mast cell/basophil products, are thought to lead to mucous cell metaplasia and smooth muscle hyperplasia. These changes in cellular behavior are the basis for the hypersecretion and hyperresponsiveness that is characteristic of the asthma phenotype. The Th2 cytokines IL-9 and IL-13 are especially effective in driving differentiation of mucous cells. Eosinophil granule constituents such as myelin basic protein, eosinophilic cationic protein, and eosinophil-derived neurotoxin may also contribute to airway inflammation, damage, and remodeling, although recent findings in mouse models and human subjects suggest that the contribution of the eosinophil may not be necessary for the asthma phenotype, at least under some conditions.[6-9] Recently, an additional type of T cell, the regulatory T cell (Tr), was identified. Tr cells produce IL-10 and transforming growth factor–β (TGF-β), which may downregulate the immune response. Studies in experimental models suggest that an inhibition of the Tr function may also drive the asthma phenotype,[10] but whether this phenomenon occurs in humans still needs to be determined.

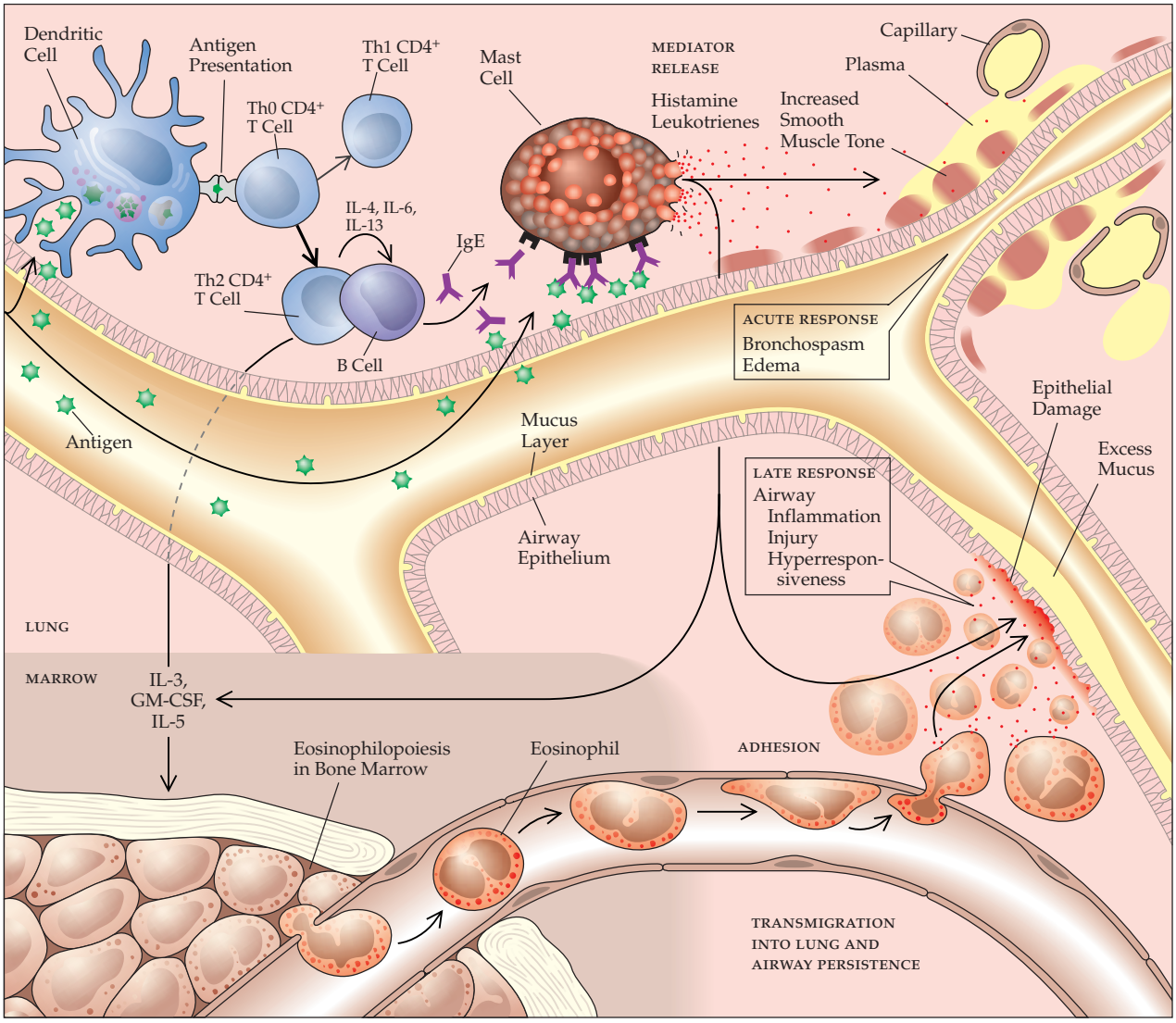


Figure 1 **Pathogenesis of allergic asthma. Inhaled antigen is processed by dendritic cells and presented to Th2 CD4$^+$ T cells. B cells are stimulated to produce IgE, which binds to mast cells. Inhaled antigen binds to IgE, stimulating the mast cell to degranulate, which in turn leads to the release of mediators of the immediate response and the late response. Histamine and the leukotrienes produce bronchospasm and airway edema. Released chemotactic factors, along with factors from the Th2 CD4$^+$ T cells, facilitate eosinophil traffic from the bone marrow to the airway walls. These late responses are proposed to lead to excessive mucus production, airway wall inflammation, injury, and hyperresponsiveness. (GM-CSF—granulocyte-macrophage colony-stimulating factor; IFN-γ—interferon gamma; IL—interleukin)**

GENETIC INFLUENCE

Family and twin studies have clearly shown that asthma is a heritable disease; however, it has proved difficult to identify specific linkage to individual genes. This difficulty likely rests on the fact that asthma is a complex genetic disorder in which multiple genes contribute to the development of the disease. In addition, asthma is a multifactorial disease in which a variety of environmental factors can influence the development of the disease phenotype. Nonetheless, several candidate genes have been linked to the development of asthma, at least in some populations. Genome-wide screens have shown associations between asthma and regions on chromosomes 2p, 4q, 5q23-31, 6p24-21, 11q13-21, 12q21-24, 13q12-14, 16q21-23, and 19q.[11] These chromosomal regions encode genes involved in antigen presentation and activation of T cells (i.e., *CTLA-4 HLA*) and cytokines such as IFN-α, IL-4, IL-5, IL-9, and IL-13, as well as components of the IgE receptor (i.e., FcεRIβ) and the beta agonist receptor. Through use of a candidate-gene approach, IL-13 and associated proteins (i.e., the α chains of the IL-13 and IL-4 receptors and STAT-6) were also found to exhibit linkage to asthma.[12-14] Other investigators have identified novel genes that are associated with asthma. *ADAM33* encodes a metalloproteinase found in airway smooth muscle and fibroblasts and is presumably involved in airway remodeling.[15] The B isoform of *GPRA*, a G protein–coupled receptor, is found in airway smooth muscle cells of asthmatic patients but not in those of nonasthmatic persons.[16] Two additional genes associated with asthma are the dipeptidyl peptidase family member *DPP10* and *PHF11*, which is a member of the family of proteins containing zinc fingers.[17,18] The function of these gene products is unknown. Future studies are likely to provide additional insight into the genetic basis of asthma.

VIRAL INFLUENCE

The cascade of events that surround the generation and acti-

vation of allergen-specific Th2 cells is best demonstrated in persons with allergy. However, many patients with allergy do not develop asthma, and some asthma patients have no demonstrable allergic component. Other than the generation of Th1 cells early in the allergic response, the Th2 hypothesis for asthma does not offer a useful explanation for how certain viral infections may not only exacerbate but initiate the development of asthma. Severe viral infections of the epithelium appear to generate a cascade of events that ultimately leads to the development of both an acute and a chronic asthma phenotype [*see Figure 2*]. Moreover, nearly 25% of children who develop a severe respiratory disease from infection with respiratory syncytial virus (RSV) before 3 years of age will subsequently develop asthma, whereas fewer than 5% of those who develop milder disease from RSV infection will manifest any long-term effects.[19] The basis of the response to RSV infection and the relationship of this response to the development of asthma still need to be defined.

One line of research has focused on the behavior of viral host cells (especially airway epithelial cells).[20] The findings have led to the proposal of an alternative paradigm for asthma pathogenesis that incorporates observed abnormalities in innate immune behavior of airway epithelial cells, viral capacity to initiate and sustain the asthma phenotype, and allergic predisposition of many asthma patients (the so-called Epi-Vir-All paradigm).[21] The contribution of each of these abnormalities to asthma pathogenesis likely varies in different types of asthmatic patients and at different times in the same patients. Moreover, some of the same abnormalities in innate and adaptive airway immunity may also underlie the pathogenesis of other chronic inflammatory airway diseases (e.g., chronic bronchitis) that manifest similar disease traits (e.g., airway hyperreactivity and mucous cell metaplasia). A challenge of current asthma research is to more precisely define these abnormalities and to develop biomarkers that can be used clinically for more accurate diagnosis and treatment of this condition.

Diagnosis

CLINICAL MANIFESTATIONS

The classic symptoms of asthma are wheezing, cough, and shortness of breath. Because of the intermittent nature of asthma symptoms, patients may be entirely asymptomatic between attacks. However, studies suggest that even with essentially normal lung function and an absence of symptoms, patients may still have persistent airway inflammation that requires ongoing therapy.

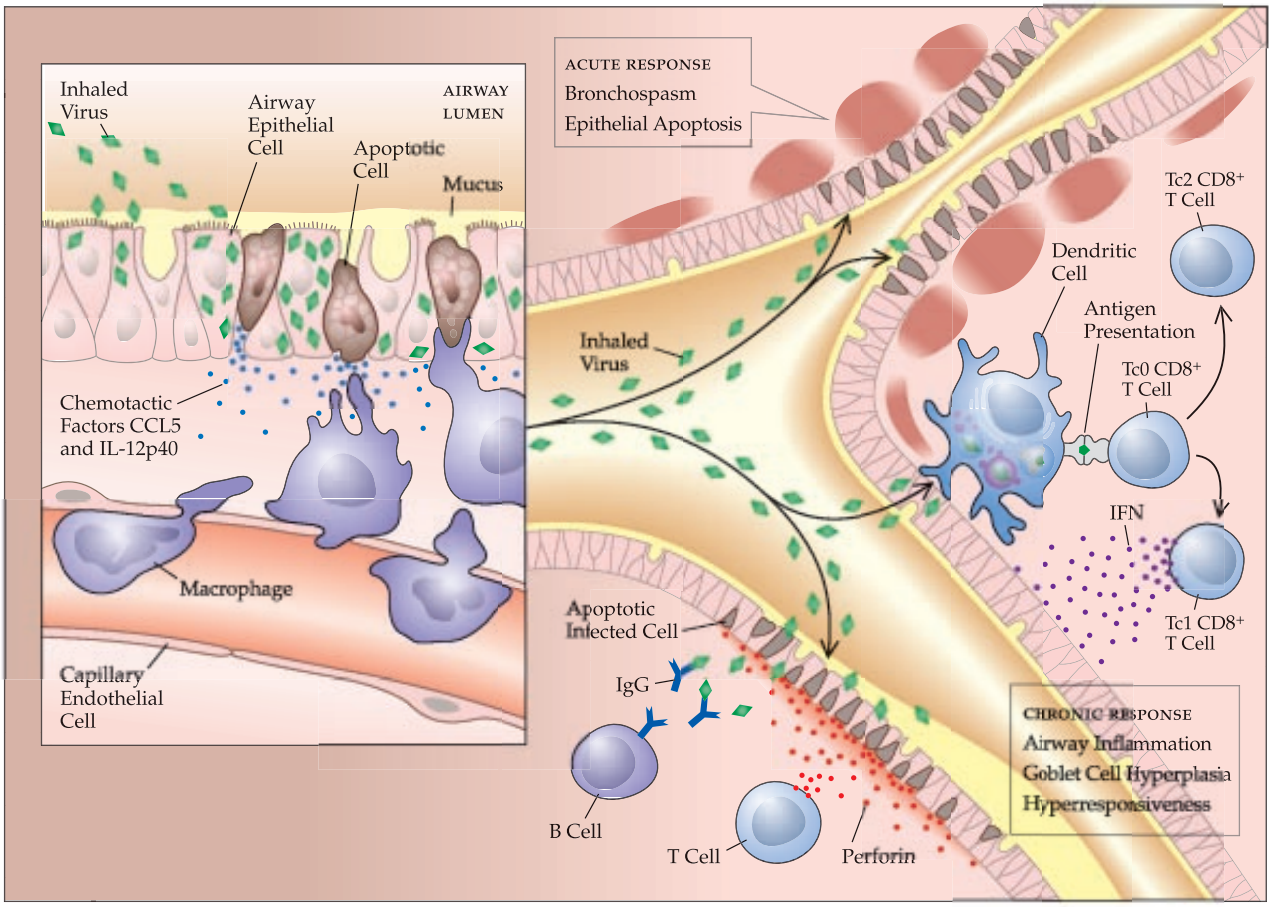


Figure 2 **Pathogenesis of viral-induced asthma. Inhaled virus infects epithelial cells and leads to apoptosis of some of them. The release of chemotactic factors promotes the recruitment of macrophages into the lung parenchyma, where they ingest the dead epithelium. An acute response consisting of bronchospasm occurs at this time. Similar to allergic asthma, the inhaled virus is processed by dendritic cells and presented to Th2 $CD8^+$ T cells. These cells produce copious amounts of IFN-γ. Perforin released from the T cells leads to apoptosis of infected cells. B cells produce IgG, which is capable of neutralizing the virus. These events are thought to be related to the chronic response, which consists of airway inflammation, goblet cell hyperplasia, and airway hyperresponsiveness. (IFN-γ—interferon gamma; IL—interleukin; CCL—chemokine ligand)**

Wheezing

Wheezing is the most common finding during acute airway obstruction, and the chest may be hyperresonant on percussion. As airflow obstruction becomes severe, a number of physical signs may become manifest and may offer clues to the severity of the attack. Tachypnea and tachycardia are common. A fall in systolic blood pressure of more than 10 to 12 mm Hg during inspiration (paradoxical pulse) is found in approximately one half of patients whose forced expiratory volume in one second (FEV_1) is 1 L or less during acute exacerbations. Accessory muscle usage and paradoxical pulse with decreasing intensity of breath sounds also signify severe airway obstruction. The cessation of wheezing in the absence of therapy is an especially ominous sign that may reflect a marked decrease in airflow. Bronchoconstriction can be triggered by a variety of stimuli that have little or no impact on the airways of nonasthmatic persons; these responses can be helpful diagnostically. The stimulus need not be a specific allergen or chemical in the workplace; a nonspecific (i.e., nonantigenic) stimulus, such as strenuous exercise, especially while breathing dry air, may trigger the response.

Cough

The cough of patients with asthma can be nonproductive or can raise copious amounts of sputum. In the absence of infection, sputum is typically mucoid and often tenacious. Eosinophils and their constituents may cause a yellow discoloration of sputum, even when infection is absent. Cough is occasionally the only manifestation of asthma; the term cough-varient asthma has sometimes been used to designate such cases.

Dyspnea

The level of dyspnea tends to vary greatly in individual patients over time, reflecting wide variations in the severity of airflow obstruction. At times, airflow obstruction prevents any significant physical exertion; at other times, strenuous exercise is possible but may trigger wheezing and shortness of breath [*see* Exercise-Induced Asthma, *below*]. During a severe attack, a desperate hunger for air is the overwhelming symptom. Chest tightness commonly occurs with dyspnea and may be confused with angina pectoris. Most patients associate their chest tightness with the sensation of being unable to take in a full and satisfying breath. Older patients are often less aware of airflow obstruction and may therefore require closer monitoring.[22]

STIMULI THAT TRIGGER ATTACKS

The stimuli that trigger asthmatic attacks vary among individual patients. For some patients, attacks of asthma are triggered by allergens such as ragweed or animal dander, house dust containing antigens from dust mites and cockroaches, strong odors or fumes, or ingested substances such as certain foods, sulfite agents, aspirin, and tartrazine [*see* Specific Forms and Complications of Asthma, *below*]. Emotional stress may trigger symptoms in some patients, but the precise role of the central nervous system in regulating airway function is difficult to quantitate. Reflux of gastric acid into the lower esophagus may exacerbate asthmatic symptoms, presumably through vagally mediated parasympathetic nervous reflexes,[23] but the role of gastroesophageal reflux in asthma remains controversial.[24] Persistent posterior drainage of nasal mucus may also be an aggravating factor. Indirect evidence indicates that nasal and sinus disease increases airway responsiveness and thereby exacerbates asthma.[25] Other stimuli are virtually universal precipitants

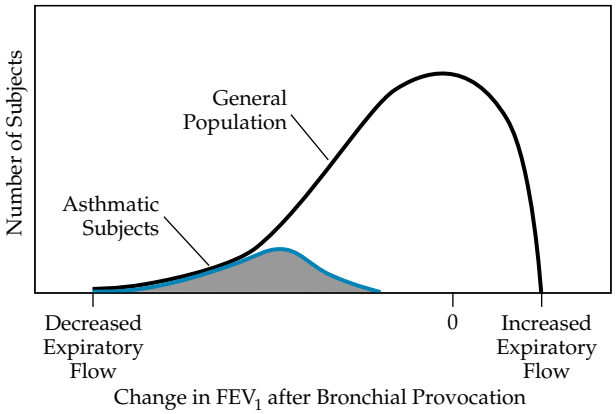


Figure 3 **The change in expiratory flow, as measured by forced expiratory volume in 1 second (FEV_1), after administration of a given concentration of bronchoprovocative agonist is shown for a general population. A large decrease in expiratory flow implies a greater bronchial reactivity. Bronchial reactivity is normally distributed, with a skew toward increased reactivity. The figure indicates that asthmatic persons exhibit varying degrees of bronchial hyperreactivity and that some persons with increased bronchial responsiveness do not manifest clinical signs of asthma.**

of asthma symptoms. Such stimuli include strenuous exercise, particularly if it is performed in cold, dry air; respiratory infections, usually caused by community-acquired viruses; inhaled pollutants and irritants such as ozone, sulfur dioxide, and cigarette smoke; beta blockers, angiotensin-converting enzyme inhibitors, and, in some Asian patients, ingestion of ethanol.[26] Specific antigen responsiveness may be more severe after exposure to environmental pollutants. Some women with asthma have been noted to have a significant increase in exacerbations during the preovulatory and perimenstrual periods.[27] The exact hormonal contributions to asthma flares in these patients are not known; however, increases in progesterone levels have been implicated on the basis of temporal association.

Asthma that is triggered by identifiable inhaled antigens (aeroallergens) is often associated with manifestations of atopic disease (i.e., seasonal allergic rhinitis or eczema). Exceptions to this association include cases of asthma caused by certain sensitizing antigens encountered in the workplace that may not elicit an IgE antibody response [*see* Occupational Asthma, *below*]. Laboratory tests can be used to support a diagnosis of atopic disease. These include increased levels of peripheral blood eosinophils, total serum IgE levels, and specific IgE levels as determined by radioallergosorbent testing (RAST) directed at a particular antigen or by positive wheal-and-flare reactions to antigens pricked or injected into the skin.

Some experts make a distinction between asthma in which there are known allergic precipitants of bronchoconstriction (i.e., extrinsic asthma) and asthma in which there are no known precipants (i.e., intrinsic asthma). In asthma cases in which there is evidence of atopy, lung function may decline at a greater rate than in nonatopic cases.[28] The distinction between extrinsic and intrinsic asthma has probably been overemphasized and may be misleading, because it implies a differentiation of etiology or pathogenesis that is not supported by current data. Categorization into extrinsic or intrinsic asthma groups is difficult: allergic precipitants may not be recognized, symptoms of other atopic diseases may create ambiguity, and laboratory test results may be inconclusive or falsely positive. As a result, it is probably best to avoid the use of the

terms intrinsic and extrinsic asthma. Fortunately, the management of asthmatic patients is not dependent on this distinction.

LABORATORY TESTS

No single laboratory test can establish a diagnosis of asthma, but a test for bronchodilator responsiveness can provide supportive evidence when asthma is suspected on clinical grounds. In patients with baseline airflow obstruction, a significant increase in expiratory airflow (e.g., ≥ 12% increase in FEV_1) after inhalation of a bronchodilator suggests asthma.[29] Unfortunately, as a diagnostic test, bronchodilator responsiveness lacks both sensitivity and specificity. Negative results may be found in asthmatic patients who have near-normal baseline lung function, are tested shortly after self-administration of a bronchodilator, or have obstruction caused by increased mucus secretions. Significant bronchodilator responses may also be observed in other types of chronic obstructive airway diseases, such as chronic bronchitis.

Because asthma is episodic, a diagnosis of asthma may be suspected on the basis of recurrent symptoms, even if there is normal pulmonary function. Measurement of airway reactivity may be a useful diagnostic test in this setting. In clinical practice, airway reactivity is defined by increased airway resistance or decreased maximal expiratory airflow (e.g., FEV_1) after a challenge with an inhaled nonspecific stimulus, such as methacholine or histamine. In asthmatic patients, airflow obstruction occurs with nonspecific stimuli that do not alter the airway mechanics of patients without asthma. The degree of nonspecific reactivity is often expressed as the dose required to reduce the FEV_1 by 20% from the measured baseline. This variable, the PD_{20} FEV_1, does not absolutely distinguish asthmatic patients from patients with other respiratory disorders (e.g., allergic rhinitis); however, as the degree of reactivity increases (with a decrease in the PD_{20} FEV_1), the confidence that asthma is present increases.

Other, less frequently used stimuli for bronchoprovocation include cold air, exercise, eucapnic hyperventilation, and adenosine 5′-monophosphate. Although bronchoprovocation tests are very sensitive, they are not specific for asthma [*see Figure 3*]. An alternative approach for the diagnosis of asthma is to demonstrate lability of maximal airflow. This can be demonstrated by having the patient periodically monitor his or her condition with a self-operated peak flowmeter and by keeping a diary of the results or by using a device that electronically collects and stores the information. A diurnal variation in peak expiratory airflow (PEF) of 15% or more is highly suggestive of a diagnosis of asthma.[30]

In an adult with asthmatic symptoms of new onset, a chest radiograph is also warranted to exclude alternative diagnoses (e.g., pneumonia, congestive heart failure, and pulmonary fibrosis). In patients with asthma, the chest radiograph is usually normal; occasionally, subtle changes indicative of bronchial wall thickening are detected, and during an episode of severe airflow obstruction, radiographic signs of hyperinflation may be present. Chest radiographs of patients treated for asthma exacerbations reveal unsuspected pulmonary infiltrates, atelectasis, pneumothorax, or pneumomediastinum only 2% of the time; therefore, radiographs are not routinely indicated in patients with documented asthma who are experiencing a flare of disease.

Differential Diagnosis

Acute respiratory illness can mimic the signs and symptoms of asthma, and it can exacerbate existing asthma. Commonly, adults and children with tracheobronchitis or bronchiolitis that is associated with respiratory viral infection present with signs of cough, wheezing, dyspnea, and increased secretions that are all driven by underlying airway inflammation and consequent airway hyperreactivity. Common viral pathogens include adenovirus, rhinovirus, influenza virus, metapneumovirus, and parainfluenza virus. In immunocompromised hosts, other less common pathogens (e.g., herpes simplex virus, respiratory syncytial virus, and possibly metapneumovirus) may be responsible [*see 7:XXV Respiratory Viral Infections*]. Much less commonly, asthmatic symptoms are caused by the Churg-Strauss syndrome [*see 14:IV Focal and Multifocal Lung Disease*], which is a form of vasculitis characterized by asthmalike airflow obstruction and wheezing.

In addition to acute conditions, flares of chronic obstructive airway diseases (especially chronic bronchitis but sometimes bronchiectasis or cystic fibrosis) [*see 14:III Chronic Obstructive Diseases of the Lung*] can cause transient increases in obstruction and lead to episodic wheezing similar to that seen in asthma. Other chronic lung diseases may also affect the airway in a manner that can be mistaken for asthma. For example, sarcoidosis may cause endobronchial granuloma formation with resultant airflow limitation, as well as cough, wheezing, and dyspnea that are refractory to the usual bronchodilator medications. In addition, rheumatoid arthritis may be associated with bronchiolitis, producing findings that mimic refractory asthma.

Congestive heart failure and pulmonary embolism can also cause dyspnea and wheezing. Wheezing in association with interstitial pulmonary edema is common enough to have been designated cardiogenic asthma. Improvement after administration of an inhaled bronchodilator does not exclude cardiogenic asthma as the cause of wheezing. Wheezing is a rare manifestation of pulmonary embolism. Although in the acute setting either pulmonary edema or pulmonary embolism may be mistaken for asthma, the discovery of rales, peripheral edema, and chest radiographic findings of pulmonary edema should allow for a quick distinction. Pulmonary emboli rarely cause wheezing or airflow obstruction sufficient enough to be confused with asthma exacerbations, and the sudden onset of dyspnea in a cancer patient or in a patient with an immobilized extremity should prompt additional diagnostic testing, such as an extremity ultrasound, perfusion lung scanning, computed tomography, or pulmonary angiography.

The conditions most likely to be confused with asthma over a more protracted period are those that cause partial upper airway obstruction [*see Table 1*]. In this context, the term upper airway refers to the single lumen airway from the carina upward. Dyspnea and wheezing associated with upper airway obstruction may be continuous and may fail to respond to bronchodilators—a pattern that suggests focal anatomic obstruction. However, in other cases, signs and symptoms may be intermittent and may be brought on by exercise (because of the increased airflow across the narrowed orifice) or by certain postures. Epinephrine or glucocorticoids occasionally relieve symptoms of upper airway obstruction by decreasing edema—a result that may lead one to mistake the cause of the symptoms as being asthma. Findings that suggest the correct diagnosis are the patient's perception that the problem is in the throat, wheezes that are heard loudest over the neck and transmitted with less intensity to the lung periphery, hoarseness, a cough that sounds unusual, and a normal arterial to alveolar oxygen difference ($A\text{-}aD_{O_2}$ gradient). A helpful diagnostic clue is a history of trauma, surgery, or prolonged intubation of the upper airway. Upper airway obstruction produces characteristic abnormalities in

flow-volume curves [*see 14:I Lung Function Assessment and Thoracic Diagnostic Techniques*]. An imaging test that may occasionally prove helpful in the diagnosis of upper airway obstruction is tracheal tomography, especially when performed with the latest generation of CT scanners. Laryngoscopy or fiberoptic bronchoscopy may also be useful in selected cases.

One particularly common type of upper airway obstruction that is frequently misdiagnosed as asthma is vocal cord dysfunction. Patients with vocal cord dysfunction can present with a change in vocal quality, wheezing, stridor, shortness of breath, or dyspnea on exertion. In patients with vocal cord dysfunction who do not have asthma, results of pulse oximetry, blood gas measurements, and chest radiography will be normal, and these patients will not exhibit altered airway reactivity. Flow-volume curves obtained during attacks typically show decreased inspiratory airflow consistent with an extrathoracic site of airway obstruction. Observing the vocal cords by endoscopy during an episode of vocal cord dysfunction can establish the diagnosis. The classic pattern consists of adduction of the anterior two thirds of the vocal cords with a posterior diamond-shaped chink. This occurs during inspiration but can be present during the entire respiratory cycle. It should be stressed that vocal cord dysfunction can exist alone or in combination with asthma; therefore, a diagnosis of vocal cord dysfunction does not rule out the existence of underlying asthma.

Table 1 Causes of Upper Airway Obstruction

EXTRINSIC COMPRESSION	
	Mediastinal neoplasm
	Retrosternal goiter
	Retropharyngeal abscess
	Fibrosing mediastinitis
	Thoracic aortic aneurysm
INTRALUMINAL OBSTRUCTION	
	Foreign-body aspiration
INTRINSIC STRUCTURAL ABNORMALITY	
	Infectious Disorders
	Epiglottitis
	Croup
	Leprosy
	Syphilis
	Diphtheria
	Neoplastic Disorders
	Oropharyngeal, laryngeal, or tracheal tumors
	Inflammatory and Degenerative Disorders
	Enlarged tonsils and adenoids
	Laryngeal or tracheal granulation tissue
	Cricoarytenoid arthritis
	Tracheobronchial amyloidosis
	Sarcoidosis
	Laryngomalacia
	Tracheomalacia
	Tracheal or laryngeal stenosis
	Neurologic Disorders
	Bilateral vocal cord paralysis
	Functional laryngospasm

Specific Forms and Complications of Asthma

EXERCISE-INDUCED ASTHMA

Exercise-induced asthma is distinct from exercise limitation caused by other forms of cardiopulmonary disease in which breathlessness develops after a certain level of exercise and gradually resolves with rest.[31] In contrast, an asymptomatic person with asthma and normal or near-normal lung function may exercise for several minutes without experiencing any symptoms. Five to 10 minutes after exercise, pulmonary obstruction develops (often accompanied by wheezing, shortness of breath, and chest tightness); the magnitude of obstruction is a function of drying and cooling of the airway mucosa and is thus a function of minute ventilation and of the temperature and humidity of the inspired air. The obstruction spontaneously resolves 30 to 60 minutes after onset.

The crucial determinants of exercise-induced asthma are the level of ventilation during exercise and the temperature and relative humidity of the inspired air. The higher the minute ventilation during exercise and the colder and drier the inspired air, the greater the airflow obstruction that develops after exercise. The intrathoracic airways become cooled and lose water as they condition large volumes of inspired air. It has been postulated that the fall in airway temperature and evaporative drying of the mucosa are the initiating stimuli for bronchoconstriction. The second step in this sequence is unknown; evidence is available regarding the possible activation of mast cells and the release of mediators. Another hypothesis attributes the airflow obstruction to a greater degree of reactive hyperemia of asthmatic airways in the postexercise period.[32]

NOCTURNAL ASTHMA

During periods of heightened disease activity, it is common for patients with asthma to experience symptoms at night.[33] They may awaken at 2:00 to 4:00 A.M. with cough, wheezing, and dyspnea. Measurements of pulmonary function before and after sleep usually document a significant worsening of obstruction in the morning, a phenomenon referred to as morning dipping. This phenomenon may contribute to the observed clustering of asthmatic deaths in the hours between midnight and 8:00 A.M. The precise cause of nocturnal asthma is uncertain, but possible mechanisms include sleep-related changes in airway tone, lung volumes, and airway inflammation; circadian variations in circulating histamine, cortisol, and epinephrine levels; prolonged exposure to allergens or irritants in the bedroom; late asthmatic reactions to daytime allergens or other inciting stimuli; gastroesophageal reflux related to the supine posture; retained airway secretions resulting from depressed cough reflex; and an increase in the intervals between antiasthmatic medication use.

NEAR-FATAL AND HYPERACUTE ASTHMA

There are data to suggest that patients at risk for near-fatal or fatal asthma differ histologically and pathophysiologically from other patients with asthma.[34] Advanced age, greater airway reactivity, previous use of mechanical ventilation, and long-term steroid therapy are important risk factors for mortality. Additional risk factors are previous hospitalizations for asthma, problems with compliance, major psychiatric diagnoses, use of major tranquilizers, reduced chemosensitivity to hypoxemia, and decreased perception of dyspnea. Patients with sudden attacks that worsen rapidly (referred to as hyperacute attacks) are at the greatest risk for mortality.

Histologically, patients who die less than 1 hour after the onset of symptoms have a larger proportion of neutrophils and fewer eosinophils than patients who die more than 2.5 hours af-

ter the onset of an attack, suggesting that pathogenesis may be different in those who have hyperacute attacks; such hyperacute attacks may possibly represent infection. The central airways of asthmatic patients who die during an attack have greater amounts of smooth muscle and submucosal glands than the central airways of patients in nonfatal cases and may be more responsive to bronchoconstrictive stimuli and less responsive to bronchodilators.

In some studies, regular use of beta agonists was associated with an unfavorable outcome in asthma patients. However, observations of patients with near-fatal asthma have shown that severe asphyxia and not cardiac arrhythmias may be the cause of near-fatal episodes, suggesting that undertreatment with anti-inflammatory medications, such as steroids, rather than overtreatment with beta agonists is the problem.

ASTHMA CAUSED BY ASPIRIN, SULFITES, OR TARTRAZINE

Current estimates are that 10% to 20% of patients with asthma exhibit an idiosyncratic reaction to ingested acetylsalicylic acid (aspirin).[35] Within 15 minutes to 4 hours after ingestion of as little as 10 mg of aspirin, susceptible patients may experience significant worsening of airflow obstruction and nasal or ocular symptoms (e.g., nasal congestion, rhinorrhea, and conjunctival injection). Nasal polyps are common in aspirin-sensitive asthmatic patients; the term aspirin triad has been used to describe the combination of asthma, nasal polyps, and idiosyncratic reactions to aspirin.

The most likely mechanism by which aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs) trigger bronchoconstriction in susceptible asthmatic patients is the blockade of the cyclooxygenase-mediated conversion of arachidonic acid to prostaglandins (particularly prostaglandin E_2 [PGE_2], a potent anti-inflammatory prostaglandin). This causes shunting of arachidonic acid toward the lipoxygenase pathway, where, possibly on the basis of genetic polymorphisms, there is increased expression of LTC4 synthase.[36] This in turn facilitates the formation of this bronchoconstrictor and proinflammatory leukotrienes.[37] The hypothesis does not account for the idiosyncratic nature of the reaction in only a minority of asthmatic persons or for the peculiar finding that, in rare cases, a patient with asthma actually exhibits improvement in lung function after administration of aspirin. Further, there have been conflicting data regarding leukotriene levels in these patients; however, at least some studies indicate an elevation in cysteinyl leukotrienes. This finding would help explain why leukotriene antagonists have shown good clinical efficacy in these patients. Another proposed mechanism for aspirin sensitivity is that endogenous prostaglandins normally inhibit inflammatory mediator release, especially from mast cells, and that prostaglandin blockade allows the disinhibition of this proinflammatory mechanism.

A history of asthmatic worsening or of nasal or conjunctival symptoms after aspirin ingestion is usually sufficient to identify aspirin-sensitive patients. Such patients should avoid any product containing aspirin or other NSAIDs that inhibit cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2). Acetaminophen and salsalate are weak COX-1 inhibitors, but with high doses, reactions can occur in a minority of aspirin-sensitive patients. Sodium salicylate, salicylamide, choline magnesium trisalicylate, benzydamine, chloroquine, azapropazone, and dextropropoxyphene do not inhibit COX enzymes and can be used safely. Selective COX-2 inhibitors have been shown to be safe in aspirin-sensitive patients.[38]

Sulfiting agents, including sodium and potassium bisulfite and metabisulfite, are used in the food-processing industry as sanitizing agents, preservatives, and antioxidants. These agents are also found in a great number of medications, including some bronchodilator solutions. Restaurant food, almost all wines and certain beers, fresh and dried fruits, peeled potatoes (including French fries), avocado dip, and shrimp and other shellfish are common sources of sulfites. The asthmatic reactions may be severe and life threatening, but they usually respond rapidly to bronchodilator therapy. An estimated 5% of persons with asthma are thought to have sulfite sensitivity, which currently can be properly diagnosed only by a controlled sulfite challenge.

Tartrazine, a food dye approved by the Food and Drug Administration, is a coal tar derivative widely used in prepared foods and drinks, medications, and mouth care products. Worsening of airflow obstruction in asthmatic patients has been reported in double-blind, placebo-controlled oral challenges with tartrazine. In most instances, positive responses occurred only in aspirin-sensitive asthmatic patients. Other reports have called into question the existence of tartrazine sensitivity. Tartrazine-free diets are so restrictive that careful documentation of sensitivity in a controlled fashion should always precede the imposition of such a diet on a patient.

ASTHMA IN PREGNANCY

Asthma is a potentially serious medical problem during pregnancy.[39] Pregnant women with asthma have higher rates of complications, such as hyperemesis gravidarum, uterine hemorrhage, preeclampsia, placenta previa, and the need for cesarian section; these complications are especially prevalent in patients with severe asthma. There is little evidence of an increase in maternal mortality associated with asthma.[39] Prematurity and intrauterine growth retardation of the fetus are associated with poor asthma control in the mother.[39] Improved control of asthma is associated with improvement in fetal outcome.[39] Of pregnant patients with asthma, one third will experience improvement in their asthma, one third will remain stable, and one third will experience a worsening of their asthma. Unfortunately, at this time it is not possible to accurately predict what will happen to a given patient. In fact, there is no evidence to suggest that the level of severity seen in one pregnancy will be seen in the next pregnancy of the same patient.

OCCUPATIONAL ASTHMA

Persons with asthma who are exposed to irritant dusts or fumes in the work environment are susceptible to exacerbations.[40] Occupational asthma, however, refers to asthma of new onset that is caused by prolonged exposure to a specific inhaled substance in the workplace. Occupational exposures may be involved in the development or worsening of as much as 10% of asthma cases.[40] The particular gas, dust, or vapor seems to sensitize the airways; continued exposure causes reversible airway narrowing and the development of nonspecific bronchial hyperreactivity. A typical history is that of a worker who after a few months (but sometimes up to several years) at a job notices cough, wheezing, and chest tightness shortly after arriving at the workplace, especially after a brief absence over the weekend (so-called Monday-morning asthma). Symptoms persist while the patient is at work but often abate after the patient returns home. Continued occupational exposure may lead to more persistent symptoms, more continuous airflow obstruction, and susceptibility to the typical spectrum of precipitants of asthma

in addition to the offending agents encountered in the work environment. Some persons with occupational asthma report a delayed onset of asthmatic symptoms: symptoms begin hours after the patient leaves the workplace, making recognition of an association with the offending agent more difficult. Occasionally, workers experience recurrent nocturnal asthmatic symptoms for several nights after a work-related exposure.

More than 200 substances have been linked to occupational asthma, and the list of agents continues to grow [*see Table 2*]. Major categories of offending agents include laboratory animals, birds, insects, and various animal products; plants and wood dust; biologic enzymes; isocyanates; anhydrides; metals; fluxes; latex; and drugs and other chemicals. For some of these agents, an immunologic mechanism involving the immediate hypersensitivity reaction has been demonstrated; for other substances, atopy does not predispose persons to the development of disease, and evidence of an IgE-mediated pathogenesis is lacking. Genetic factors also probably play a role in susceptibility among workers. The prevalence of disease varies with the sensitizing agent. For example, asthma develops in approximately 5% to 10% of workers exposed to toluene diisocyanate, in an estimated 20% of bakers exposed to wheat flour or rye flour, and in as many as 50% of people who work with platinum salts or proteolytic enzymes. Reducing concentrations of offending agents by better control of dust lowers the incidence of asthma exacerbations and may decrease sensitization.

Exposure to irritant gases can range from contact with pure chlorine gas in industrial or environmental spills to inhalation of smoke (which contains several toxic gases, as well as particulate matter and carbon monoxide). Such exposure can result in a wide spectrum of syndromes, ranging from acute tracheobronchitis to diffuse alveolar damage that results in acute respiratory distress syndrome. An intermediate response, termed reactive airway dysfunction syndrome, is a severe and persistent asthmatic state that can be difficult to treat and may result in long-term impairment of function.[41] Preexisting lung disease and long-term cigarette smoking are risk factors for more severe reactions.

Variations in lung function related to occupational exposures can be confirmed by the use of simple portable devices for recording PEF. Definitive diagnosis entails isolating the offending agent and then having the patient inhale carefully controlled concentrations of it while pulmonary function is sequentially recorded for several hours. Skin testing with the appropriate soluble extracts and RAST tests for specific IgE antibody assess only the presence of sensitization to the agent. Many workers exhibit positive skin-test or RAST results but have no evidence of asthma.

LATE ASTHMATIC REACTIONS

The airflow obstruction that develops after inhalation of certain allergens or industrial agents may follow one of several patterns [*see Figure 4*]. An immediate response typically occurs within 5 to 10 minutes after inhalation and usually resolves spontaneously over a period of 30 to 60 minutes, much like the airflow obstruction that develops after bronchoprovocation with histamine, methacholine, or exercise. This early asthmatic reaction often responds rapidly to inhaled bronchodilators. A delayed response to inhaled allergens and industrial agents is sometimes observed. Typically, this late asthmatic reaction occurs 6 to 8 hours after the inhalational challenge; it is often more severe than the immediate response, and it may resolve slowly

Table 2 Causes of Occupational Asthma

Potential Hazard	*Persons at Risk*	*Sensitizing Agent*
Laboratory animals, birds, insects, other animal products	Laboratory workers, animal handlers, veterinarians	Rats, mice, rabbits, guinea pigs
	Pigeon breeders, poultry workers, bird fanciers	Pigeons, chickens, budgerigars (shell parakeets)
	Grain workers	Grain mites
	Entomologists	Moths and butterflies
	Crab and prawn processors	Crabs and prawns
Plants, wood dust	Bakers	Wheat flour, rye flour
	Food processors	Coffee beans, castor beans
	Tea workers	Tea leaves
	Tobacco manufacturers	Tobacco leaves
	Carpenters, sawmill operators, cabinetmakers	Wood dust, including western red cedar dust
Biologic enzymes	Detergent industry workers	*Bacillus subtilis*
	Pharmaceutical industry workers, biomedical researchers	Pepsin, trypsin, bromelain
Isocyanates	Workers with polyurethane, plastics, and varnish	Toluene diisocyanate
	Automobile spray painters	Hexamethylene diisocyanate
Anhydrides	Workers with epoxy resins and plastics	Phthalic, trimellitic, and other anhydrides
Metals	Tanners	Chromium
	Platinum refiners	Platinum
	Metal platers	Nickel
Fluxes	Aluminum solderers	Aminoethylethanolamine
	Electronics workers	Colophony
Drugs, other chemicals	Pharmaceutical workers	Penicillins, cephalosporins, methyldopa, spiramycin, tetracycline
	Workers with plastics and rubbers	Azodicarbonamide
	Insulators	Urea, formaldehyde
	Refrigeration workers	Freon
	Hairdressers	Persulfate salts, henna

over a period of 12 to 24 hours. Treatment with glucocorticoids may prevent the occurrence of late asthmatic reactions and may be especially useful if the late response is relatively refractory to bronchodilators. Persons exhibiting both early and late responses are said to have dual reactions; occasionally, recurrent nocturnal airflow obstruction continues for several days after a single exposure.

Late reactions are important to our basic understanding of asthma, because often the clinical manifestations of asthma, particularly during subacute exacerbations, more closely mimic late asthmatic reactions than early reactions in terms of both timing and responsiveness to medications. Evidence from bronchoalveolar lavage specimens indicates that the late response correlates temporally with an influx of eosinophils and neutrophils into the airways.[42] Release of previously made factors by mast cells and basophils is thought to account for the immediate response and for the subsequent cellular infiltration. The release of a secondary wave of de novo synthesized mediators accounts for the delayed development of inflammation and obstruction. In animal models of the late asthmatic reaction, the development of bronchial hyperreactivity is coincident with the late response and is prevented by its inhibition.

ALLERGIC BRONCHOPULMONARY ASPERGILLOSIS

Allergic bronchopulmonary aspergillosis (ABPA), a hypersensitivity reaction to the colonization of the airways by *Aspergillus* species, rarely occurs except in patients with asthma. This disorder typically develops in patients with atopy and long-term asthma and is marked by fever, flulike symptoms, myalgias, and lassitude. The chest radiograph reveals pulmonary infiltrates, often with associated atelectasis of the involved segments or lobes caused by tenacious mucous plugs that occlude the proximal airways. Although this entity is frequently mistaken for bacterial pneumonia, sputum cultures are negative for pathogenic bacteria. Clues to the proper diagnosis include the presence of sputum and blood eosinophilia, the volume loss that accompanies the pulmonary infiltrates, and failure of the infiltrates to respond to antibiotics.

Abundant *Aspergillus* organisms are found in the bronchial lumina of patients with allergic bronchopulmonary aspergillosis, but significant tissue invasion does not occur. An exuberant inflammatory response involves both the bronchial wall and the surrounding lung. Acute and chronic inflammation in the walls of central bronchi, accompanied by necrosis, leads to bronchiectasis and large mucous plugs (mucoid impaction). At times, a granulomatous response results in the replacement of bronchial and peribronchial tissue (bronchocentric granulomatosis). An eosinophilic pneumonia commonly affects the surrounding lung tissue.

The pulmonary infiltrates may resolve spontaneously but commonly recur, leading to the appearance of migratory pulmonary infiltrates on chest x-ray [*see Figure 5*]. Chronic disease commonly involves the upper lobes; typical features include bronchiectasis (usually involving central airways)[43] and fibrosis with retraction. Chronic ABPA may be readily mistaken for tuberculosis, especially because hemoptysis occurs in one third to one half of cases. ABPA can occur in patients with cystic fibrosis, and a minority of patients with ABPA have mutations in the cystic fibrosis transmembrane regulator (*CFTR*) gene.[44]

The diagnosis of allergic bronchopulmonary aspergillosis can be confirmed by the following test results: (1) positive immediate skin-test reaction to *Aspergillus* antigen, (2) elevated total serum IgE level (usually > 1,000 ng/ml), (3) elevated levels of specific IgE and IgG antibodies against *Aspergillus* (levels are usually twice those of allergic asthmatic patients who do not have allergic bronchopulmonary aspergillosis), (4) the presence of precipitating antibodies against *Aspergillus*, and (5) peripheral blood eosinophilia together with the recurrent appearance of pulmonary infiltrates on chest x-ray.[45]

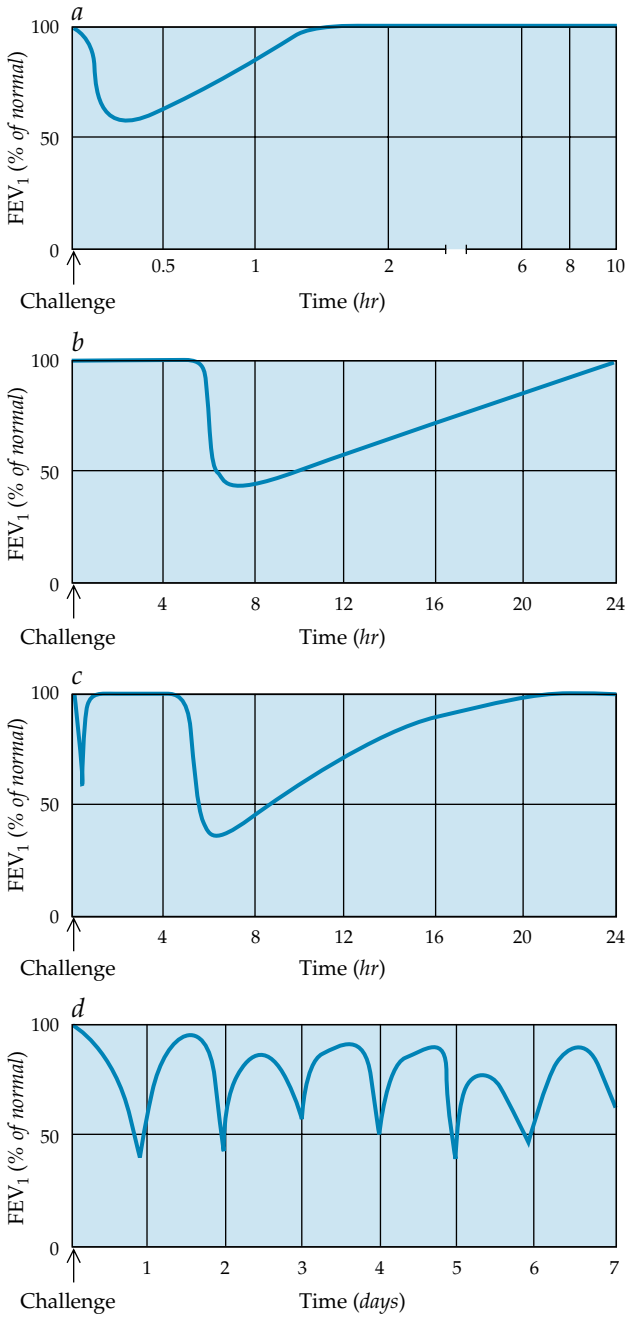


Figure 4 **Four patterns of bronchial response to an inhaled antigen are depicted. (*a*) An early response occurs within a few minutes after administration of the stimulus and resolves without treatment over the next 30 to 60 minutes. (*b*) A late response develops several hours after the bronchoprovocative challenge and resolves gradually over the next 12 to 24 hours. (*c*) A dual response combines features of both the early and the late reaction. (*d*) Occasionally, decreases in expiratory flow recur on several nights after a single challenge test. Note that different time scales are used.**

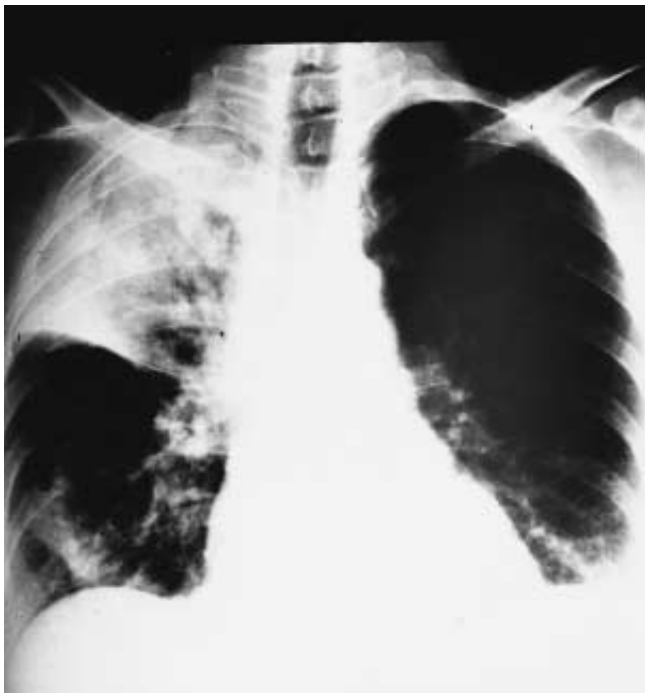


Figure 5 **This posteroanterior chest radiograph shows an asthmatic patient with allergic bronchopulmonary aspergillosis. Characteristic radiographic findings include the migratory nature of the infiltrates, the predominant upper lobe involvement, and the associated atelectasis. The minor fissure forms the lower border of the upper right lobe infiltrate; the minor fissure is displaced cephalad, indicating loss of volume in the right upper lobe.**

Management

In the United States, the total cost incurred because of asthma, including the direct costs of care and the indirect costs from lost productivity, is estimated to be $10.7 billion a year.[46] Almost half the costs arise from hospital care, and 80% of the costs are incurred by 20% of the patients.[46,47] Studies of patients hospitalized for exacerbations of asthma have documented deficiencies in the care that patients had been receiving before admission, which suggests that published guidelines[1,48] are not understood[49] and are not being used by primary care physicians.[50] In addition, many patients with asthma are not satisfied with their treatment regimen.[51] Education of the patient is a critical and often neglected aspect of care that has the potential to improve outcomes.[52] Self-monitoring with the use of inexpensive peak flowmeters may improve a patient's ability to judge disease severity and may reduce morbidity.[52] A specialist can provide more therapeutically effective[53] and more cost-effective treatment[54] and should be involved in the care of all patients with moderate to severe asthma.

ASTHMA THERAPIES

Medications used to treat asthma can be classified into two groups in accordance with their principal actions. The first group comprises bronchodilators that act primarily to relax airway smooth muscle; these agents include beta agonists and anticholinergic agents, although theophylline and its derivatives also have mild bronchodilator activity. The second group comprises anti-inflammatory drugs that act mainly as inhibitors and modifiers of airway inflammation; these agents include glucocorticoids, leukotriene modifiers, theophylline, and anti-IgE antibody [*see Table 3*].

The mainstay of therapy for asthma is the long-term use of medications to downregulate the airway inflammatory process and thereby control abnormal airway behavior. [*For treatment of specific presentations, see* Management Strategies, *below*.] Currently available drugs allow for adequate control of asthma in most patients, but there is significant need for—and ongoing effort to develop—agents that more fully and permanently correct the asthmatic condition.[55]

Bronchodilators

Beta agonists The most effective bronchodilators function by selectively and potently stimulating the $beta_2$ receptors on airway smooth muscle. Receptor activation causes adenylate cyclase to increase the levels of intracellular cyclic adenosine monophosphate (cAMP). Epinephrine is short-acting (30 to 60 minutes) when inhaled and activates both alpha receptors and beta receptors. Beta-adrenergic stimulation predominates in the airways, resulting in bronchodilatation, but dominant peripheral alpha-adrenergic stimulation results in vasoconstriction and may cause an increase in blood pressure. In clinical practice, epinephrine is given primarily as a subcutaneous or intramuscular injection for acute relief of severe airflow obstruction. It is also sold without prescription as an inhalant in metered-dose canisters. In general, epinephrine should not be used for the management of asthma; it is only useful in acute bronchospastic situations—although even then, an inhaled selective $beta_2$ agonist is preferred. Similarly, isoproterenol, a powerful nonselective beta agonist, has a short duration of action and is available as an inhalant for the treatment of asthma, but it has largely been replaced by newer, more selective $beta_2$ agonists.

Selective $beta_2$ agonists have structural modifications that make them effective when taken by inhalation. The duration of action of these compounds varies from 4 to 6 hours. Examples are metaproterenol, terbutaline, and albuterol [*see Table 4*]. Bitolterol has a slightly extended duration of action of 6 to 8 hours; it is currently available only in a metered-dose delivery system. A preparation that contains only the R isomer of albuterol (levalbuterol) has been reported to have reduced side effects and complications, but these claims are still being studied.[56] Salmeterol, an even longer-acting $beta_2$ agonist with a slow onset of action, can be given twice a day as maintenance therapy (not rescue therapy); it has resulted in better asthma control and quality of life than is seen with albuterol. Salmeterol does not compromise the bronchodilator response to the short-acting $beta_2$ agonists[57] and may have mild anti-inflammatory effects.[58] Formoterol, also a long-acting selective $beta_2$ agonist, has a duration of action similar to that of salmeterol, but it has a more rapid onset of action; formoterol thus has potential for use as a long-acting maintenance medication and a long-acting rescue medication in place of the short-acting $beta_2$ agonists.[59] Because of their potential to mask underlying inflammation, long-acting $beta_2$ agonists (i.e., salmeterol and formoterol) should not be used as monotherapy. In the United States, a combination dry-powder inhaler is available with salmeterol and the inhaled glucocorticoid fluticasone dipropionate. In most countries other than the United States, a combination of formoterol and the glucocorticoid budesonide is available for inhaled use. It is anticipated that this combination will soon be available in the United States as well.

The frequency and severity of side effects from the $beta_2$ agonists depend not only on the dose but also on the route of administration. Side effects are most prominent with parenteral admin-

Table 3 Summary of Drugs for Asthma

Drug	*Initial Dose*	*Maintenance Dose*	*Relative Efficacy*	*Cost*	*Comments*
Bronchodilators					
$Beta_2$ agonists	*See Table 4*		First choice for rescue		Oral or parenteral therapy associated with significant side effects; use short-acting drugs/inhaler p.r.n. rather than on regular schedule; use long-acting drugs regularly
Anticholinergic agents	*See Table 5*		Only for exacerbations		Can worsen prostatism, glaucoma; can be combined with beta agonists
Anti-inflammatory agents					
Systemic glucocorticoids					
Prednisone	0.5–1.0 mg/kg/day	None or minimal	Usual oral agent	$5.00/mo	q.d. initially, then wean off, if possible, or switch to q.o.d.
Methylprednisolone, oral	24–48 mg/day	None or minimal	Less commonly used oral agent	$9.00–43.00/mo	q.d. initially, then wean off, if possible, or switch to q.o.d.
Methylprednisolone, I.V.	60–125 mg q. 6–8 hr	None	Usual I.V. agent	$24.00–26.99/day	Oral therapy as effective
Inhaled glucocorticoids	*See Table 4*		First choice of anti-inflammatory agents		Various inhaled corticosteroid agents differ in potency [*see Table 5*]
Cromolyn sodium	2 puffs q.i.d.	Same	Much less potent than inhaled steroids	$30.00–39.99/canister	Used more in children; no steroid side effects; no longer preferred treatment
Nedocromil	2 puffs q.i.d.	Same	Much less potent than inhaled steroids	$50.00–59.99/mo	No longer preferred treatment
Leukotriene modifiers					As a group, less effective than inhaled corticosteroids; help with associated allergic rhinitis; should be used in aspirin-sensitive patients
Montelukast	10 mg q.h.s.	Same	First choice of leukotriene modifiers	$60.00–69.99/mo	No lab monitoring or restrictions related to meals
Zafirlukast	20 mg b.i.d.	Same		$50.00–59.00/mo	Should be taken at least 1 hr before or 2 hr after meals
Zileuton	600 mg q.i.d.	Same		$79.00–89.99/mo	Must monitor LFTs
Theophylline	100–200 mg b.i.d.	Adjust to serum level 10–20 μg/ml	Third choice	$10.00–19.99	Relatively weak bronchodilator; only used when all other agents optimized; significant toxicity, must monitor levels
Other anti-inflammatory agents					
Omalizumab	150–375 mg q. 2–4 wk	150–375 mg q. 2–4 wk	Only as add-on therapy in severe persistent asthma	$1,000/mo	Given S.C.; dosage based on pretreatment IgE levels and body mass
Methotrexate	7.5 mg/wk	Adjust to effect; maximum, 25 mg/wk	Efficacy controversial	$10.00–19.99/mo	Toxic drug, must monitor blood counts and LFTs closely; should only be given by asthma expert

LFT — liver function test

istration, are of intermediate prominence with oral administration, and are least prominent with inhaled administration. Common side effects are those of sympathetic stimulation, including nervousness and agitation, muscle tremors, and palpitations. Cardiac stimulation may lead to tachycardia and cardiac arrhythmias. Intracellular shifts of potassium can cause hypokalemia, although the clinical significance of this finding is uncertain. Long-term administration of beta agonists may be associated with the development of tolerance[60]—possibly caused by the decrease in the number and responsiveness of membrane beta receptors related to polymorphisms of the beta receptor genes[61]—and with increased airway responsiveness to allergen and nonspecific stimuli.

Over the years, $beta_2$ agonists given by inhalation have been

Table 4 Estimated Comparative Doses of Inhaled Bronchodilators for Asthma[38]

Drug	*Maintenance Dosage*	*Exacerbation Dosage*	*Relative Efficacy*	*Cost*	*Comments*
Inhaled Short-Acting Beta$_2$ Agonists			Inhaled short-acting beta$_2$ agonists are first-line p.r.n. therapy; no significant differences in efficacy among inhaled short-acting beta$_2$ agonists	$30.00–39.99/mo	Dilute aerosols to minimum of 4 ml at gas flow of 6–8 L/min
Albuterol					
Nebulizer, 5 mg/ml	1.25–5.0 mg q. 4–8 hr p.r.n.	5.0 mg q. 2 hr			
MDI, 84 μg/puff	2–4 puffs q. 6 hr p.r.n.	3–8 puffs q. 2 hr			MDI as effective as nebulizer when used with spacer
DPI, 200 μg/capsule	1–2 capsules q. 6 hr p.r.n.	Not studied in exacerbations			—
Bitolterol					
Nebulizer, 2 mg/ml	0.5–3.5 q. 4–8 hr p.r.n.	Not studied in exacerbations	—	N/A	—
MDI, 370 μg/puff	2–4 puffs q. 6 hr p.r.n.	—	—		—
Levalbuterol Nebulizer, 0.63 mg/3 ml	0.31–0.63 mg q. 6–8 hr p.r.n.	0.63–1.25 mg q. 6–8 hr p.r.n.	Similar to albuterol, slightly fewer side effects	$110.00–130.00/mo	—
Inhaled Long-Acting Beta$_2$ Agonists			Inhaled long-acting beta$_2$ agonists first-line scheduled bronchodilator therapy	$50.00–69.99/mo	Salmeterol has slower onset of action; not used as a rescue bronchodilator; should only be used with an inhaled glucocorticoid
Salmeterol					
MDI, 21 μg/puff	2 puffs q. 12 hr	Not recommended			
DPI, 50 μg/blister	1 blister q. 12 hr	—			
Formoterol DPI, 12 μg/capsule	1 capsule q. 12 hr	Not recommended	—	N/A	Faster onset of action than salmeterol; may be used as a rescue bronchodilator
Inhaled Anticholinergics			Indicated in combination with inhaled short-acting beta$_2$ agonists for exacerbations	$30.00–34.99/mo	
Ipratropium bromide					
Nebulizer, 0.25 mg/ml	Not recommended	0.5 mg q. 2–8 hr			May mix with albuterol in same nebulizer
MDI, 18 μg/puff	Not recommended	3–8 puffs q. 3–4 hr			MDI as effective as nebulizer when used with spacer
Combined Short-Acting Beta$_2$ Agonist and Anticholinergic			When both are indicated	$30.00–39.99/mo	When both are indicated
Albuterol + ipratropium bromide					
MDI albuterol (90 μg/puff) + ipratropium bromide (18 μg/puff)	Not recommended	3–8 puffs q. 2 hr			

DPI—dry-powder inhaler MDI—metered-dose inhaler

the mainstay of treatment for asthma because of their rapid onset of action, their effectiveness, and their convenience. In the mid-1980s, an increase in asthma deaths occurred in New Zealand after the introduction of the beta$_2$ agonist fenoterol. A study based in Saskatchewan, Canada, however, suggested that all beta$_2$ agonists, especially if used heavily, are associated with an increased risk of fatal and near-fatal asthmatic episodes, but a meta-analysis suggested that the risk is very small and may be confined to beta$_2$ agents given by nebulizer.[62] A recent case-control study found increased mortality in patients treated with short-acting beta$_2$ agonists for 1 to 5 years, but shorter courses were not associated with death.[63] In other studies, nebulized and oral beta$_2$ agonists were associated with increased risk of cardiovascular death, ischemic heart disease, and cardiac failure.[64,65] The most obvious question concerning these studies is whether the increased mortality occurred because of the administration of excessive amounts of the therapeutic agent or because of the severity of illness in those patients given the drug. Further studies have suggested that the regular use of beta$_2$ agonists leads to a mild increase in the degree of airway responsiveness to allergens.

Investigators have applied genetic analysis of the beta$_2$ receptor to try to determine whether certain patients are more susceptible to the detrimental effects of bronchodilator treatment. Initial retrospective studies indicated that a subgroup of patients with a specific polymorphism (homozygosity for arginine [Arg/Arg]) in the beta$_2$-adrenergic receptor gene (*beta$_2$ AR*) exhibited less improvement with long-term albuterol treatment than with intermittent therapy. Further, these patients experienced a significant decline in respiratory function when treatment with long-term beta-agonist therapy was discontinued.[66] A more recent prospective trial found a similar trend; in that study, patients with a homozygosity for glycine (Gly/Gly) experienced improvement of symptoms from albuterol therapy, whereas patients with a homozygosity for arginine (Arg/Arg) experienced improvement when albuterol therapy was discontinued.[67] Ongoing studies aim to define the risk associated with the use of long-acting beta$_2$

agonists; these studies also seek to determine the specific haplotypes that appear most susceptible to detrimental actions of $beta_2$ agonists. Early studies have found the greatest risk associated with Arg/Arg at 16th amino acid in the $beta_2$ receptor.[66] This haplotype is found at highest frequency in the African-American population,[68] but studies are under way to further define the association of additional mutations.

At present, these observations must be added to the list of potentially negative effects associated with the regular use of these beta agents. Currently, it is recommended that short-acting $beta_2$ agonists be used only on an as-needed basis.[1] This strategy not only avoids the potential negative side effects of the drug but also allows the clinician to detect an increase in asthmatic symptoms—as indicated by the patient's increased need for the $beta_2$ agonist—and a worsening of the underlying inflammation. When patients begin to feel the need to use inhaled $beta_2$ agonists more frequently and at increased doses, they should be seen by a physician and evaluated for factors that may be aggravating their asthma; the addition of other treatments should be considered to better control the asthma. Excessive reliance on inhaled bronchodilators, without sufficient attention to the underlying inflammatory component of the disease, continues to be a problem in the way asthma is managed. Inhaled glucocorticoids, rather than scheduled doses of $beta_2$ agonists, should be considered as first-line therapy for asthmatic patients with daily symptoms.[1]

Anticholinergic agents For hundreds of years, the anticholinergic agents atropine and stramonium, which are alkaloid derivatives of plants of the *Datura* species, have been used as bronchodilators. Because atropine is well absorbed into the blood across the respiratory tract mucosa, it causes significant systemic side effects, even when administered by inhalation. The atropinic congeners ipratropium bromide and methylatropine nitrate are poorly absorbed into the circulation when inhaled and therefore have limited side effects. Atropine and its derivatives compete with acetylcholine at muscarinic receptors, which are adjacent to the parasympathetic, postganglionic nerve endings.

Ipratropium bromide requires 60 minutes to reach peak effect—a slightly longer time than seen with inhaled $beta_2$ agonists—but it lasts 4 to 6 hours. It has a somewhat smaller bronchodilatory effect in patients with asthma. Potential advantages are (1) its minimal cardiac stimulatory effects, which make it particularly desirable for use in patients with coronary artery disease or cardiac arrhythmias, and (2) its additive bronchodilatory effect when used in combination with submaximal doses of $beta_2$ agonists. No benefit is gained by adding ipratropium bromide to a regimen of $beta_2$ agonists given at maximal doses in long-term therapy. There is a modest benefit for patients with acute asthma exacerbations.[69] The primary indication for anticholinergic agents is chronic obstructive pulmonary disease.

Anti-inflammatory Drugs

Systemic glucocorticoids The precise mechanism or mechanisms by which glucocorticoids downregulate asthmatic inflammation (or other types of inflammation) are yet to be defined. Nonetheless, considerable evidence indicates that glucocorticoids act at least in part by inhibiting the production of proinflammatory cytokines. Cellular targets of glucocorticoid action include T cells and eosinophils. The eosinophil, in particular, is especially sensitive to glucocorticoid treatment; it rapidly undergoes apoptotic cell death upon exposure to glucocorticoids. Glucocorticoids also exhibit a slight bronchodilatory effect, which is perhaps related to alterations in airway smooth muscle receptors; however, this action is a relatively minor component of glucocorticoid action in asthma.

The optimal dose, dosing schedule, preparation, and duration of therapy for the treatment of asthma with glucocorticoids are uncertain. As a consequence, physicians have widely varying preferences for glucocorticoid treatment regimens. Available studies suggest that lower dosages of intravenously delivered hydrocortisone (200 mg/day) are as effective as higher dosages (2,000 mg/day) in hospitalized patients[70] and that a prolonged taper of oral prednisone may not be necessary to prevent late exacerbations of asthma. Systemic glucocorticoids are commonly employed for short courses (e.g., 1 to 3 weeks) to treat exacerbations of asthma.

Therapy for less than 3 weeks is not associated with adrenal axis suppression; longer courses require tapering doses and measurement of A.M. cortisol levels to determine when it is safe to stop therapy. Prednisolone and prednisone, which is metabolized to prednisolone in the liver, are the most widely used oral preparations in the treatment of asthma; these two agents are equally effective in patients who do not have severe liver disease. Long-term maintenance therapy should be avoided because of the well-known side effects of extended use of systemic glucocorticoids. In those few cases in which relief of symptomatic airflow obstruction cannot be achieved by other means, orally administered glucocorticoids are given at the lowest effective dose. Alternate-day administration minimizes side effects and hypothalamic-pituitary-adrenal axis inhibition and may provide sufficient control of chronic symptoms.

A subset of asthmatic patients either are unresponsive to glucocorticoids or require persistently high doses of glucocorticoids for adequate control.[71] This phenomenon has led to the possibility that some patients develop "steroid-resistant" asthma, but the precise basis for defining this group remains uncertain. Some reports have indicated that these patients might have abnormalities of the intracellular mechanisms of glucocorticoid action.[72] Other studies have indicated that resistance may be reversed by high-dose glucocorticoid therapy. In either case, a proportion of these patients may simply have more severe disease; however, evidence of steroid-resistant asthma may suggest that use of other anti-inflammatory agents is indicated (see below).

Inhaled glucocorticoids Inhaled glucocorticoids are the primary long-term maintenance therapy aimed at preventing recurrent exacerbations of asthma. Regular use of inhaled glucocorticoids, even in low doses, has been shown to effectively prevent a major portion of asthma hospitalizations and deaths.[73] These benefits are realized, however, only when the proper dose is prescribed and the patient takes the medication on a regular basis and with proper technique of inhalation.[74] The particular advantages of inhaled glucocorticoids are that the agent is deposited locally in the airway and that it is poorly absorbed into the circulation; this results in a decrease in systemic side effects.

Five preparations of inhaled glucocorticoids are available in the United States: beclomethasone, triamcinolone, flunisolide, fluticasone, and budesonide[75] [*see Table 5*]. Formulations using the propellant hydrofluoroalkane instead of the chlorofluorocarbons, which harm the ozone layer, are available. Because these inhalers generate aerosols with smaller-size particles, more of the inhaled glucocorticoid is delivered into the small airways, resulting in improved therapeutic response.[76] A particularly popular combination of salmeterol and fluticasone is available

in this form; its increased efficacy may be related to better deposition of the combination or improved patient compliance.

Inhaled glucocorticoids effectively inhibit the late-phase response even when given after antigen challenge; if taken long term, they reduce airway inflammation and remodeling.[77] After a period of days to weeks of regular use of inhaled glucocorticoids, many patients manifest a decrease in airway hyperreactivity.[78] Conversely, a subset of patients who are withdrawn from therapy exhibit flares of asthma that appear linked to airway infiltration of $CD4^+$ and $CD8^+$ T cells.[9] These findings appear to fit with the capacity of glucocorticoids to suppress at least partially allergic or viral-induced airway hyperreactivity in experimental models of asthma.

High doses of inhaled glucocorticoids may be as effective as moderate doses of prednisone in the treatment of asthma exacerbations.[79] Long-term oral glucocorticoid therapy can be discontinued in some patients, and the frequency of bronchodilator administration can often be reduced. Inhaled glucocorticoids are often particularly effective during periods of increased exposure to aeroallergens. Local side effects include glossitis, sore throat, hoarseness, dysphonia, and oral candidiasis. Oral candidiasis occurs in 5% to 15% of patients who use inhaled glucocorticoids; rinsing the mouth with water after each use, utilizing a spacer device (if using a metered-dose inhaler), or both can decrease the incidence of this side effect. Despite extended use of inhaled glucocorticoid preparations, histologic changes in the bronchial mucosa and alterations in oropharyngeal flora do not occur. In adults, systemic side effects are usually seen only with high doses of inhaled glucocorticoids; these side effects include skin atrophy, reduction in bone density, cataracts, and suppression of the hypothalamic-pituitary-adrenal axis.[80]

Cromolyn and nedocromil Cromolyn (sodium chromoglycate) exhibits properties that are distinct from other antiasthmatic medications. It does not act as a bronchodilator and does not directly antagonize the action of inflammatory mediators or reverse established inflammation. Instead, this agent appears to inhibit inflammation by preventing the release of chemical mediators from mast cells, even in the presence of cell-bound IgE antibody and appropriate antigen. The mechanism by which cromolyn capacity stabilizes mast cells remains uncertain. Two effective delivery systems are available: a solution for nebulization and an aerosol delivered by a metered-dose pressurized canister. Controlled trials have shown that cromolyn is minimally beneficial for patients with asthma.[81] As a result of this modest effect and the development of safer inhaled glucocorticoids, the use of cromolyn and nedocromil is limited. Nonetheless, because these agents cause few if any side effects and exhibit efficacy in mast cell–dependent events, they may be useful for the pretreatment of patients at risk for allergen- or exercise-induced asthma.

Leukotriene modifiers Leukotriene modifiers have been developed to directly antagonize the capacity of leukotrienes to mediate airway smooth muscle contraction, mucus secretion, and inflammatory cell recruitment; they have been especially useful for the treatment of asthma induced by exercise, aspirin, and allergens.[82] The class of leukotrienes that contribute to asthma pathogenesis is derived from the action of the 5-lipoxygenase pathway that metabolizes arachidonic acid to biologically active lipid mediators, especially leukotriene D_4 (LTD_4). LTD_4 produces bronchospasm by binding to a specific receptor, cysteinyl leukotriene type 1 receptor (CysLT1R), on airway smooth muscle. Another leukotriene, LTB_4, is a potent chemotaxin, especially for neutrophils, but its role—as well as that of other lipoxygenase products—for producing the asthma phenotype remains uncertain.

Two types of leukotriene modifiers are available for the treatment of asthma. Zileuton inhibits 5-lipoxygenase itself and therefore blocks all leukotriene formation. The other type includes montelukast, zafirlukast, and pranlukast (available in Japan); these agents inhibit the binding of cysteinyl leukotriene to CysLT1R. The specific indications for these drugs in the treatment of asthma are still being established; however, patients with aspirin-induced and exercise-induced asthma respond well to either a 5-lipoxygenase or CysLT1R inhibitor. Leukotriene antagonists have been shown to be effective in patients

Table 5 Estimated Comparative Daily Dosages of Inhaled Glucocorticoids

Drug	*Low Dosage*	*Medium Dosage*	*High Dosage*	*Relative Efficacy*	*Cost*	*Comments*
Beclomethasone 42 μg/puff 84 μg/puff	168–504 μg 4–12 puffs/day 2–6 puffs/day	504–840 μg 12–20 puffs/day 6–10 puffs/day	> 840 μg > 20 puffs/day > 10 puffs/day	Third highest potency	$30.00–39.99/mo	Now available in HFA MDI, possibly improving effectiveness
Budesonide 200 μg/puff	200–400 μg 1–2 inhalations/day	400–600 μg 2–3 inhalations/day	> 600 μg > 3 inhalations/day	Second highest potency	$30.00–39.99/mo	DPI inhaler
Flunisolide 250 μg/puff	500–1,000 μg 2–4 puffs/day	1,000–2,000 μg 4–8 puffs/day	> 2,000 μg > 8 puffs/day	Lowest potency	$50.00–59.00/mo	
Fluticasone 44 μg/puff 110 μg/puff 220 μg/puff	83–264 μg 2–4 puffs/day 2 puffs/day —	264–660 μg — 2–6 puffs/day 1–2 puffs/day	> 660 μg — > 6 puffs/day > 3 puffs/day	Highest potency	$40.00–49.99/mo	Also formulated in combination with salmeterol in DPI
Triamcinolone 100 μg/puff	400–1,000 μg 4–10 puffs/day	1,000–2,000 μg 10–20 puffs/day	> 2,000 μg > 20 puffs/day	Lowest potency	$30.00–39.99/mo	Provided with spacer

DPI—dry-powder inhaler HFA—hydrofluoroalkane MDI—metered-dose inhaler

with allergic rhinitis.[83] Therefore, they may be considered useful in patients who have symptoms of both allergic rhinitis and asthma. These medications are effective in approximately 50% of patients with mild, persistent asthma or moderate asthma; they may be added to or substituted for inhaled or oral anti-inflammatory drugs,[84,85] but they are not as effective as inhaled glucocorticoids with or without long-acting beta$_2$ agonists.[86,87]

Because of associated liver dysfunction, patients taking zileuton must undergo frequent liver function testing. Doubling of liver transaminases necessitates the discontinuance of the drug. Zafirlukast has been reported to induce elevations in transaminase levels in rare situations; however, most authorities do not recommend regular monitoring of liver function in patients taking this drug. These side effects have not been found to occur with montelukast therapy. Leukotriene antagonists allow for the reduction of systemic glucocorticoids; in rare instances, patients have been noted to develop a vasculitis—the Churg-Strauss syndrome—once treatment with systemic glucocorticoids was reduced.[88] This phenomenon has also been reported with the use of potent inhaled glucocorticoids. In both cases, the development of Churg-Strauss syndrome was attributed to preexisting disease that was masked by glucocorticoid treatment and that became exacerbated when this treatment was reduced [*see 14:4 Focal and Multifocal Lung Disease*].

Theophylline Theophylline is a methylxanthine that is closely related in structure to caffeine.[89] It was thought that theophylline acted by increasing cAMP levels by inhibiting the activity of phosphodiesterase, an enzyme that facilitates the conversion of cAMP to the noncyclic adenosine 5′-monophosphate. Other modes of action for theophylline have been explored, including antagonism of adenosine at its receptor. Theophylline may also have anti-inflammatory effects that include reduction in the number of inflammatory cells, expression of cytokines, and acceleration of neutrophil apoptosis.[90,91] Recent studies suggest that theophylline may even influence transcription of genes in the airway tissue and thereby downregulate inflammation.[92]

When given orally, theophylline is almost completely absorbed from the gastrointestinal tract. The rate of absorption, however, can vary greatly, depending on the formulation. A rapid onset of action can be achieved with the elixir of theophylline (peak effect is seen at approximately 60 minutes), whereas a long duration of action results from slow-release preparations (peak effect occurs at 6 to 8 hours), making once- or twice-daily dosing schedules possible. Aminophylline (the ethylenediamine salt of theophylline) and oxtriphylline (the choline salt of theophylline) are more water soluble than theophylline. Aminophylline is the preparation used for I.V. administration; its bronchodilator activity is solely attributable to theophylline, which by weight constitutes 85% of aminophylline.

Oxidation and demethylation of theophylline take place in the liver; hepatic metabolites, along with a small amount of unaltered theophylline, are excreted in the urine. The average serum half-life of theophylline in nonsmoking adults is approximately 7 to 9 hours, but differences in metabolic rates among people and in the same person over time lead to considerable variability in theophylline serum concentrations after a given dose. Clearance of theophylline is accelerated in children, in cigarette smokers, and in persons receiving phenytoin; its clearance is delayed by numerous influences, including primary liver disease, right ventricular failure, and the use of certain drugs (e.g., cimetidine, erythromycin, ciprofloxacin, and oral contraceptives).

In mild to moderate asthma, the bronchodilator response varies directly with the theophylline serum concentration. Significant bronchodilation occurs at theophylline concentrations as low as 5 to 8 mg/L; at concentrations exceeding 20 mg/L, side effects become increasingly common. Traditionally, a therapeutic range of 10 to 20 mg/L has been recommended; however, lower levels may be effective against inflammation, and many experts recommend a dose range of 5 to 15 mg/L.[93] At levels greater than 40 mg/L, serious toxic reactions can be observed, including seizures and ventricular arrhythmias. Occasionally, life-threatening toxic reactions occur at lower serum levels (i.e., 25 to 40 mg/L). This relatively narrow therapeutic index, along with highly variable clearance rates, has necessitated measurement of serum theophylline concentrations in many clinical circumstances—especially in patients with severe airflow obstruction, in whom maximal bronchodilator effect is desired.

Side effects are common with theophylline and may occur at serum concentrations at or below the therapeutic target range. GI complaints include nausea and vomiting, abdominal pain, and diarrhea. Headache, nervousness, insomnia, and tremors are the most common neurologic side effects. Cardiac stimulation may lead to sinus tachycardia, extrasystoles, and atrial arrhythmias. In cases of severe theophylline toxicity caused by overdosage (i.e., theophylline concentration > 40 mg/L, especially > 60 mg/L), clearance of theophylline can be accelerated by the administration of activated charcoal, either orally or by nasogastric tube; this will remove any theophylline remaining in the stomach or intestine. On rare occasions, charcoal hemoperfusion is necessary to reduce theophylline concentrations rapidly.

The risk of life-threatening toxicity depends not only on the drug level but also on whether intoxication is acute or chronic. A theophylline level between 40 and 50 mg/L in an acutely intoxicated patient carries a very low risk of seizures; however, the same level in a patient receiving long-term theophylline therapy whose ability to clear the drug is impaired carries a significant risk of seizures and represents a potentially life-threatening intoxication. Moreover, when a toxic level is attributable to reduced clearance rather than to acute intoxication, the duration of risk is greater because of the long half-life of theophylline. Because of the multiple side effects, the need to monitor blood levels of the drug, the potential severity of toxicity, and the availability of newer classes of more effective drugs, theophylline is no longer considered a primary choice of therapy for patients with asthma.

Other Anti-inflammatory Agents

Additional anti-inflammatory agents are under development, but at least one has been approved by the Food and Drug Administration and is available in the United States for asthma therapy: a humanized monoclonal antibody (mAb) against IgE (omalizumab). This agent removes circulating IgE but is unable to bind to cell-bound antibody; this prevents the possibility of therapy-induced anaphylaxis. Omalizumab is administered subcutaneously on a biweekly or monthly schedule on the basis of pretreatment serum IgE levels and body mass index. Once therapy is begun, IgE levels should not be monitored (they will actually increase because of the formation of an IgE–anti-IgE complex). No significant side effects have been noted with omalizumab; however, the high cost of therapy is a significant obstacle to its use. This therapy, which has been shown to reduce asthma exacerbations and glucocorticoid dosage, is indicated for severe, persistent asthma; studies suggest that patients with markedly decreased lung function, those who have recently vis-

ited the emergency department, and those who require high-dose inhaled glucocorticoids are most likely to respond to treatment with omalizumab.[94,95]

Other agents have shown little effectiveness in the treatment of asthma. Antihistamines have not been recommended as part of the usual treatment of asthma. Newer-generation drugs, however, are more potent inhibitors of in vivo histamine effects and show some promise, possibly in combination with other mediator blockers, such as the leukotriene modifiers.[96] Methotrexate may have a glucocorticoid-sparing effect in glucocorticoid-dependent asthmatic patients.[97] However, most patients referred to specialists for methotrexate therapy can be managed with higher doses of inhaled glucocorticoids, long-acting $beta_2$ agonists, theophylline administered at therapeutic levels, or inhaled anticholinergic agents; in addition, precipitating problems, such as rhinosinusitis and gastroesophageal reflux, need to be eliminated. One study found that auranofin, an oral gold preparation, had glucocorticoid-sparing effects in asthmatic patients.[98] Cyclosporine, an inhibitor of T cell activation, has been shown to have efficacy in some patients with severe asthma.[99] Other treatments that have been found to be of possible benefit to these patients include hydroxychloroquine,[100] keliximab (a chimeric monoclonal antibody to $CD4^+$),[101] and high-dose intravenous immunoglobulin.[102] None of these agents have been accepted as being useful for asthma treatment.

Immunotherapy

Immunotherapy for asthma remains controversial.[103,104] The strategy involves repeated administration of an allergen in gradually increasing amounts to reduce sensitivity to the allergen. Immunotherapy has clearly been shown to be effective in allergic rhinitis and, thus, may be effective in some patients with specific types of allergic asthma. However, the role of immunotherapy in the majority of patients with asthma has not been determined and remains of limited use.

MANAGEMENT STRATEGIES

Outpatient Care

In newly diagnosed asthma patients, the identification of avoidable precipitants of airway obstruction should be sought by history, and whenever possible, precipitants should be removed from the environment.[105] Common sources of allergens and irritants in the home include pets, insects, dust mites, cigarette smoke, dust and mold, feather pillows, down comforters, and shag rugs. Because patients may be exposed to multiple allergens in their home, skin testing is useful in demonstrating hypersensitivity to a suspected allergen; this in turn allows for targeted and appropriate avoidance measures. Exposure to allergens may not always cause immediate symptoms; rather, it may induce a state of enhanced bronchial hyperreactivity to many nonspecific stimuli. The relation between allergen exposure and worsening of asthma may not be immediately apparent.

An air filter or air conditioner may be helpful during periods of heavy outside air contamination. Beta-blocking drugs, including ophthalmic solutions, should be avoided; if such a drug is absolutely necessary, a selective $beta_1$ blocking agent should be given in the lowest possible dose. Treatment of postnasal drip or esophageal reflux of gastric acid may ease asthmatic symptoms in some patients. Weight loss in obese patients with asthma may also improve symptoms and lung function.[106]

Several new types of devices have been developed to directly deliver medication to the airways.[107] Breath-actuated metered-dose inhalers, dry-powder inhalers, and metered-dose inhalers with nonchlorofluorocarbon (hydrofluoroalkane) propellants are now on the market. The dry-powder inhalers seem to be easier for patients to use than the other types of delivery devices, because no coordination of hand action and breath taking are required. However, most patients who use the new types of delivery devices still require a metered-dose inhaler (usually for their short-acting $beta_2$ agonist) as part of their regimen.

Some patients have difficulty using metered-dose aerosol systems, either because of a lack of coordination or because of a physical handicap. Spacer devices or reservoir bags (into which the medication is sprayed from the metered-dose canister and from which the patient then inhales) may greatly assist patients who have poor coordination; an adapter is available for patients with arthritic hands. Use of a spacer results in greater delivery of the drug to the airway[108] and a reduction in complications. The principal mechanisms underlying these effects are a reduction of larger particles that would otherwise adhere to the posterior pharynx and a slowing of the rate of inhalation (most devices have warning signals when inspiration is too fast), allowing the intermediate-sized particles to make the turn from the mouth into the airway.

On the basis of symptom frequency and measurements of airway function, the NAEPP guidelines divide asthma into four categories: mild intermittent, mild persistent, moderate persistent, and severe persistent. Although this section discusses discrete therapeutic options, it should be noted that asthma is an ever-changing disease. Therefore, a patient may have moderate persistent asthma at one visit, but with appropriate treatment, the asthma may become only mildly intermittent. The practitioner should be aware of this fact and continually adjust therapy to match the current level of disease severity. In this manner, patients will be exposed to the lowest possible dose of medication to adequately treat their disease; this will in turn lead to a lower incidence of unwarranted side effects and to better quality of life.

Mild intermittent asthma Patients with mild intermittent asthma have daily symptoms no more than twice a week or nocturnal symptoms no more than twice a month. Because airway function is not significantly compromised, these patients achieve an FEV_1 (or PEF) of at least 80% of predicted value/personal best value with less than 20% variability. These individuals have no symptoms between attacks. Although not proved, it is assumed that patients with mild intermittent asthma have no ongoing lung inflammation and therefore can be treated with only intermittent bronchodilator therapy for occasional symptoms.[1] However, there are scant data supporting or refuting the presence of inflammation in the airways of mild asthmatics, and some experts have begun to suggest the use of anti-inflammatory therapy even in patients with mild intermittent asthma. Short-acting $beta_2$ agonists delivered by metered-dose inhalers are preferred on the basis of rapid onset, potent bronchodilator effect, and few side effects. In addition, patients should attempt to avoid exposure to any known triggers in their environment.

Mild persistent asthma When symptoms are more frequent than twice a week but not yet daily or when nocturnal symptoms occur more than twice a month but not weekly, the patient is regarded as having mild persistent asthma. Although some slight increase in airway variability (20% to 30%) may be

seen, these patients do not have much loss of airway function at baseline (i.e., FEV_1 and PEF of at least 80% of predicted value/personal best value). These patients are regarded as having continuous airway inflammation, and long-term drug therapy should be employed to control this process. The NAEPP guidelines recommend use of a low-dose inhaled glucocorticoid as the preferred treatment.[1] However, alternative therapies may also be used in these patients, including leukotriene antagonists and theophylline. Appropriate environmental-control measures should be taken to limit the patients' exposure to known triggers of their symptoms.

Moderate to severe persistent asthma Because airway inflammation is the primary problem in moderate to severe asthma, the fundamental therapy is glucocorticoids, administered primarily by inhalation.[1] Some studies of patients with asthma have reported a slowing of the decline in lung function after the addition of inhaled glucocorticoids to a regimen of inhaled bronchodilators. Use of bronchial hyperreactivity as an additional indicator of the need for anti-inflammatory therapy has been reported to result in more aggressive treatment and improved outcomes.[108]

Patients with daily symptoms and persistent airflow obstruction should use inhaled glucocorticoids regularly. Inhaled $beta_2$ agonists should continue to be used on an as-needed basis. In patients receiving low-dose inhaled glucocorticoids, the addition of a long-acting selective $beta_2$ agonist at its regular dose may provide improved control.[109] Although long-acting selective $beta_2$ agonists should not be used in place of low-dose inhaled steroids,[110] their use can often lead to a reduction in the dose of inhaled steroids.[111] When inhaled glucocorticoids do not adequately control the patient's symptoms, a long-acting $beta_2$ agonist should be added to the regimen. This form of combination therapy (e.g., salmeterol and fluticasone) has been extremely useful in patients with moderate to severe persistent asthma, as well as in patients with less severe disease. Alternatively, in patients with moderate persistent asthma (i.e., those with FEV_1 values between 60% and 80%), the NAEPP guidelines allow for increasing the inhaled corticosteroid dose or for the addition of a leukotriene antagonist or theophylline in place of a long-acting $beta_2$ agonist.[1]

Severe persistent asthma (characterized by constant symptoms and an FEV_1 value of 60% or less than the predicted value) should be treated with a combination of a high-dose inhaled glucocorticoid and a long-acting $beta_2$ agonist. If this approach fails to fully control the patient's symptoms, the addition of a leukotriene antagonist, theophylline, or both can be considered. However, there are scant data pertaining to the use of these medications in the treatment of severe persistent asthma; as such, these agents are not included in the NAEPP recommendations. More severe and acute symptoms often warrant a short course of oral glucocorticoids. Often, patients show dramatic improvement with oral glucocorticoids but experience a relapse of symptoms shortly after the medication is discontinued. Relapses can be managed by reinstituting the oral glucocorticoids and by tapering the regimen to alternate-day therapy at the minimal dose necessary to maintain the patient's symptoms at an acceptable level.

An asthma specialist should evaluate all patients with moderate to severe persistent asthma. A specialist can verify the diagnosis, identify the specific allergens that trigger symptoms, and establish strategies by which the patient can avoid the triggers that cause the asthma to flare.

Exercise-induced asthma Exercise-induced asthma can be a significant source of morbidity in asthmatic patients, particularly in those who are physically active. This form of asthma is of particular concern in the treatment of athletes and children. Inhaled $beta_2$ agonists are active as prophylactic agents; they are also used therapeutically after an attack has occurred. One successful approach to controlling exercise-induced asthma is through the daily use of a leukotriene modifier.[112] The symptoms of exercise-induced asthma often can be reduced through the use of a warm-up routine before exercise and a cool-down routine after exercise. As with nocturnal asthma, overall treatment of the asthma will often relieve the exercise-induced component.

Aspirin-induced asthma Patients with aspirin-induced asthma should avoid aspirin and other NSAIDs. Otherwise, these patients should be treated in accordance with the guidelines outlined (see above), with leukotriene modifiers being a useful agent for disease control.[113]

Asthma in pregnancy Treatment of asthma during pregnancy is similar to the treatment of asthma in nonpregnant patients and should occur in consultation with an asthma specialist.[114] Education of the patient about asthma and the appropriate use of medications, regular peak flow monitoring in patients with moderate to severe asthma, and a self-management plan are important.

The short-acting $beta_2$ agonists metaproterenol, terbutaline, and albuterol have been found to be safe during pregnancy and should be used on an as-needed basis. The long-acting $beta_2$ agonists have not been studied adequately to determine their safety and should only be used after other medications have been optimized.

The inhaled glucocorticoids beclomethasone and budesonide have proved to be safe in pregnancy. There are few studies of safety of the other inhaled glucocorticoids in pregnancy. Use of inhaled glucocorticoids during pregnancy is a very useful therapy for reducing exacerbations of asthma and should be used in all patients who require as-needed short-acting $beta_2$ agonists more than once or twice weekly. In patients whose asthma is not well controlled with low doses of inhaled glucocorticoids, an increase in the dose of inhaled glucocorticoids or the addition of theophylline, which has been found to be safe in pregnancy, should be considered. Use of a long-acting $beta_2$ agonist in this situation should only occur after consideration of the risk. If these measures fail to control the asthma or if a severe exacerbation occurs, use of systemic glucocorticoids may be necessary. Prednisone and prednisolone have been found to be safe for the fetus in the vast majority of studies. Systemic glucocorticoid use may, however, be associated with an increase in the risk of the maternal complications of pregnancy. Once control is established, the doses of glucocorticoids should be gradually lowered at a rate that is well tolerated by the patient.

Occupational asthma Treatment of the milder airway responses may include the use of inhaled glucocorticoids, bronchodilators, and, if there are signs of airway infection, antibiotics. In most cases, removal of the offending agent or transfer of the patient from the site of the offending agent cures occupational asthma. Transfer of the patient to a job that merely reduces rather than eliminates exposure does not effectively relieve symptoms. In a few cases, asthma continues for years after the patient has left the workplace.

Allergic bronchopulmonary aspergillosis Systemic administration of glucocorticoids is a mainstay of treatment for asthmatic patients who have ABPA. Although antifungal therapy and inhaled glucocorticoids have not been thought to be useful, the combination of itraconazole—an antifungal with enhanced activity against *Aspergillus* species[115]—and high-dose inhaled beclomethasone[116] may be helpful. Besides symptomatic and radiographic evidence of a response to therapy, the total serum IgE concentration has been used to follow the course of the disease; in some instances, an increase in IgE concentration heralds an exacerbation.

Emergency Therapy

On arrival of the patient in the emergency department, oxygenation should immediately be assessed by pulse oximetry, and controlled doses of oxygen should be provided to increase the O_2 saturation to greater than 90%.[117] An arterial blood gas measurement should be obtained as soon as practical. Patients with severe hypoxemia should be rapidly evaluated for hypercapnia, pneumothorax, atelectasis, or pneumonia, because many patients with uncomplicated asthma do not have hypoxemia.

The most effective emergency treatment is repeated administration of a shorter-acting $beta_2$ agonist by inhalation. There is no advantage to giving $beta_2$ agonists systemically, and the inhaled route is preferred for initial therapy, even in severe exacerbations. Administration of a $beta_2$ agonist by metered-dose inhaler and a spacer device is as effective as administration by nebulization and requires less time with a respiratory therapist, making this approach potentially more cost-effective. In patients younger than 45 years, $beta_2$ agonists can be administered every 15 to 20 minutes for up to three or four doses without adverse hemodynamic consequences, although unpleasant side effects from adrenergic overstimulation are universal. Intravenous aminophylline has been found to be only a weak bronchodilator in this setting and adds little benefit. Systemic glucocorticoids should be given early for severe attacks because the glucocorticoids will not take effect for as long as 12 hours.[118] High doses are no more effective than routine doses of hydrocortisone (i.e., 200 mg I.V. bolus every 6 hours), methylprednisolone (40 to 60 mg I.V. bolus every 6 to 8 hours), or prednisone (40 to 60 mg orally every 6 to 8 hours).[118]

Evaluation of lung function through spirometry or measurement of peak expiratory airflow is useful in assessing the severity of the asthmatic attack and the response to treatment. Patients whose FEV_1 or PEF is decreased to 25% of normal or less are at risk for hypercapnia, and their arterial blood gas values should be obtained. Patients who have hypercapnia despite initial bronchodilator therapy should receive further treatment in an intensive care unit. Hypercapnia signals very severe and potentially life-threatening airway obstruction; if not reversed, intubation and mechanical ventilation will be required [*see 14:VIII Respiratory Failure*]. Some patients respond rapidly to an intensive bronchodilator regimen and can be discharged from the emergency department promptly. Because some degree of persistent airflow obstruction is likely to remain in such patients despite the resolution of their dyspnea and wheezing, most patients should be placed on an intensified home treatment program at the time of release. For most patients, this program should include a brief course of glucocorticoids. The routine use of oral glucocorticoids for 1 week after discharge from the emergency department reduces the rate of relapse in patients treated for an acute asthma exacerbation.[119] Other patients experience only slight, gradual improvement despite repeated administration of potent bronchodilators. In these cases, the airway obstruction may be accompanied by inflammation and edema of the airways, as well as by excessive mucus secretions.

All patients discharged from the emergency department should have a follow-up examination with their primary care provider or asthma specialist within 1 week of discharge. Although the patient may still be on a systemic glucocorticoid, it is important to remember that the asthma exacerbation occurred because of ineffective control of the underlying airway inflammation. Thus, an increase in anti-inflammatory medication would be warranted, as well as intensive asthma education, which is not often given during the emergency department visit.

In-Hospital Therapy

The most common indication for hospitalization because of asthma is the failure of signs and symptoms of obstruction to resolve after 3 or 4 hours of emergency therapy. Respiratory failure with hypercapnia severe enough to require endotracheal intubation or significant barotrauma requires immediate hospitalization. In-hospital therapy for asthma to a large extent is simply a continuation of the measures that were begun in the emergency department. Development of protocols for hospital care of asthma patients has resulted in improved outcomes.[120]

Inhaled bronchodilators continue to be a mainstay of asthma therapy. In severe, life-threatening attacks of asthma, short-acting inhaled $beta_2$ agonists can be given as often as every 1 to 2 hours if necessary. As the obstruction resolves, administration can be less frequent. A long-acting $beta_2$ agonist may be useful in reducing the need for frequent use of inhaled short-acting $beta_2$ agonists.[121] Glucocorticoid treatment is indicated in virtually all patients admitted for severe asthma. Two regimens commonly employed for the treatment of such patients are (1) hydrocortisone, 2 mg/kg by I.V. bolus, followed by 0.5 mg/kg/hr by continuous infusion, and (2) methylprednisolone, 60 to 125 mg by I.V. bolus every 6 hours. Some experts recommend giving a total daily dose of prednisone of 0.5 to 1 mg/kg orally. The optimal dosage and route of administration for glucocorticoids remain to be defined.

Most patients will improve with treatment, though progress may require several days. Serial determinations of pulmonary function are useful for monitoring the response to therapy. When the FEV_1 or PEF reaches 50% to 60% of normal, patients usually have minimal or no symptoms, and the wide fluctuations in airflow that often characterize the initial course are no longer present. The frequency and intensity of the bronchodilator regimen can then be decreased, I.V. medications can be switched to oral preparations, and the patient can be started on a program that can be continued after hospital discharge.

Occasionally, a patient deteriorates despite appropriate therapy. In this situation, it is necessary to monitor arterial blood gases closely. Intubation and mechanically assisted ventilation are needed in cases involving progressive hypercapnia and significant respiratory acidosis. For all intubated patients, lung function should be assessed. Lung compliance should be assessed by recording the airway pressure required to deliver a tidal volume (calculated by dividing tidal volume by plateau pressure). Flow resistive pressure should be determined (peak pressure minus plateau pressure). The presence and degree of intrinsic positive end-expiratory pressure (auto-PEEP) should be assessed. Baseline measurements can be compared to subsequent values as a means of assessing response to therapy or the need for accelerated treatment. If initial values reflect only modest obstruction in a

hypercapnic asthmatic patient, a substantial contribution of respiratory muscle fatigue or weakness to the respiratory failure is likely [*see 14:VIII Respiratory Failure*]. Extreme measures, such as I.V. administration of isoproterenol, bronchial lavage to remove mucous plugs, and general anesthesia, are potentially dangerous and are rarely indicated. Ancillary measures include use of controlled supplemental oxygen for patients with arterial oxygen desaturation (O_2 saturation < 90%) and empirical broad-spectrum antibiotic therapy (e.g., ampicillin, tetracycline, erythromycin, or trimethoprim-sulfamethoxazole) for patients with fever and sputum purulence that is not caused by eosinophils. Rapid diagnosis of viral infections is becoming more widely available through the use of real-time polymerase chain reaction and immunofluorescence staining, but treatment options are limited. Similarly, there is at best only marginal benefit from vigorous hydration, use of inhaled saline mists, mucolytic therapy (e.g., acetylcysteine), and chest physiotherapy.

Mitchell H. Grayson, M.D., receives grants for clinical research from Genentech, Inc., and Novartis AG, and he serves on the speakers' bureau of Merck & Co, Inc.

Michael J. Holtzman, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Guidelines for the diagnosis and management of asthma: update on selected topics—2002. National Asthma Education and Prevention Program. J Allergy Clin Immunol 110 (5 suppl):S141, 2002

2. Mannino DM, Homa DM, Akinbami LJ, et al: Surveillance for asthma—United States, 1980–1999. MMWR Surveill Summ 51:1, 2002

3. Cohn L, Elias JA, Chupp GL: Asthma: mechanisms of disease persistence and progression. Annu Rev Immunol 22:789, 2004

4. Randolph DA, Stephens R, Carruthers CJ, et al: Cooperation between Th1 and Th2 cells in a murine model of eosinophilic airway inflammation. J Clin Invest 104:1021, 1999

5. Stephens R, Randolph DA, Huang G, et al: Antigen-nonspecific recruitment of Th2 cells to the lung as a mechanism for viral infection–induced allergic asthma. J Immunol 169:5458, 2002

6. Leckie MJ, ten Brinke A, Khan J, et al: Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. Lancet 356:2144, 2000

7. Lee JJ, Dimina D, Macias MP, et al: Defining a link with asthma in mice congenitally deficient in eosinophils. Science 305:1773, 2004

8. Humbles AA, Lloyd CM, McMillan SJ, et al: A critical role for eosinophils in allergic airways remodeling. Science 305:1776, 2004

9. Castro M, Bloch SR, Jenkerson MV, et al: Asthma exacerbations after glucocorticoid withdrawal reflects T cell recruitment to the airway. Am J Respir Crit Care Med 169:842, 2004

10. Stock P, Akbari O, Berry G, et al: Induction of T helper type 1–like regulatory cells that express Foxp3 and protect against airway hyper-reactivity. Nat Immunol 5:1149, 2004

11. Wills-Karp M, Ewart SL: Time to draw breath: asthma-susceptibility genes are identified. Nat Rev Genet 5:376, 2004

12. Arima K, Umeshita-Suyama R, Sakata Y, et al: Upregulation of IL-13 concentration in vivo by the IL13 variant associated with bronchial asthma. J Allergy Clin Immunol 109:980, 2002

13. Hershey GK, Friedrich MF, Esswein LA, et al: The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. N Engl J Med 337:1720, 1997

14. Tamura K, Arakawa H, Suzuki M, et al: Novel dinucleotide repeat polymorphism in the first exon of the STAT-6 gene is associated with allergic diseases. Clin Exp Allergy 31:1509, 2001

15. Van Eerdewegh P, Little RD, Dupuis J, et al: Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. Nature 418:426, 2002

16. Laitinen T, Polvi A, Rydman P, et al: Characterization of a common susceptibility locus for asthma-related traits. Science 304:300, 2004

17. Allen M, Heinzmann A, Noguchi E, et al: Positional cloning of a novel gene influencing asthma from chromosome 2q14. Nat Genet 35:258, 2003

18. Zhang Y, Leaves NI, Anderson GG, et al: Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet 34:181, 2003

19. Sigurs N: A cohort of children hospitalised with acute RSV bronchiolitis: impact on later respiratory disease. Paediatr Respir Rev 3:177, 2002

20. Sampath D, Castro M, Look DC, et al: Constitutive activation of an epithelial signal transducer and activator of transcription (STAT) pathway in asthma. J Clin Invest 103:1353, 1999

21. Holtzman MJ, Morton JD, Shornick LP, et al: Immunity, inflammation, and remodeling in the airway epithelial barrier: epithelial-viral-allergic paradigm. Physiol Rev 82:19, 2002

22. Weiner P, Magadle R, Waizman J, et al: Characteristics of asthma in the elderly. Eur Respir J 12:564, 1998

23. Richter JE: Gastroesophageal reflux disease and asthma: the two are directly related. Am J Med 108:153S, 2000

24. Coughlan JL, Gibson PG, Henry RL: Medical treatment for reflux oesophagitis does not consistently improve asthma control: a systematic review. Thorax 56:198, 2001

25. Simons FE: Allergic rhinobronchitis: the asthma–allergic rhinitis link. J Allergy Clin Immunol 104:534, 1999

26. Takao A, Shimoda T, Kohno S, et al: Correlation between alcohol-induced asthma and acetaldehyde dehydrogenase-2 genotype. J Allergy Clin Immunol 101:576, 1998

27. Zimmerman JL, Woodruff PG, Clark S, et al: Relation between phase of menstrual cycle and emergency department visits for acute asthma. Am J Respir Crit Care Med 162:512, 2000

28. Ulrik CS, Backer V, Dirksen A: A 10 year follow up of 180 adults with bronchial asthma: factors important for the decline in lung function. Thorax 47:14, 1992

29. Standardization of spirometry, 1994 update. American Thoracic Society. Am J Respir Crit Care Med 152:1107, 1995

30. Jain P, Kavuru MS, Emerman CL, et al: Utility of peak expiratory flow monitoring. Chest 114:861, 1998

31. McFadden ER Jr, Gilbert IA: Exercise-induced asthma. N Engl J Med 330:1362, 1994

32. Cummiskey J: Exercise-induced asthma: an overview. Am J Med Sci 322:200, 2001

33. Martin RJ, Banks-Schlegel S: Chronobiology of asthma. Am J Respir Crit Care Med 158:1002, 1998

34. Barr RG, Woodruff PG, Clark S, et al: Sudden-onset asthma exacerbations: clinical features, response to therapy, and 2-week follow-up. Multicenter Airway Research Collaboration (MARC) investigators. Eur Respir J 15:266, 2000

35. Babu KS, Salvi SS: Aspirin and asthma. Chest 118:1470, 2000

36. Cowburn AS, Sladek K, Soja J, et al: Overexpression of leukotriene C4 synthase in bronchial biopsies from patients with aspirin-intolerant asthma. J Clin Invest 101:834, 1998

37. Szczeklik A, Sanak M: Genetic mechanisms in aspirin-induced asthma. Am J Respir Crit Care Med 161:S142, 2000

38. Stevenson DD, Simon RA: Lack of cross-reactivity between rofecoxib and aspirin in aspirin-sensitive patients with asthma. J Allergy Clin Immunol 108:47, 2001

39. Tan KS, Thomson NC: Asthma in pregnancy. Am J Med 109:727, 2000

40. Malo JL, Chan-Yeung M: Occupational asthma. J Allergy Clin Immunol 108:317, 2001

41. Lemiere C, Malo JL, Gautrin D: Nonsensitizing causes of occupational asthma. Med Clin North Am 80:749, 1996

42. O'Byrne P: Asthma pathogenesis and allergen-induced late responses. J Allergy Clin Immunol 102:S85, 1998

43. Ward S, Heyneman L, Lee MJ, et al: Accuracy of CT in the diagnosis of allergic bronchopulmonary aspergillosis in asthmatic patients. AJR Am J Roentgenol 173:937, 1999

44. Marchand E, Verellen-Dumoulin C, Mairesse M, et al: Frequency of cystic fibrosis transmembrane conductance regulator gene mutations and 5T allele in patients with allergic bronchopulmonary aspergillosis. Chest 119:762, 2001

45. Patterson R, Greenberger PA, Harris KE: Allergic bronchopulmonary aspergillosis. Chest 118:7, 2000

46. Weiss KB, Sullivan SD, Lyttle CS: Trends in the cost of illness for asthma in the United States, 1985–1994. J Allergy Clin Immunol 106:493, 2000

47. Smith DH, Malone DC, Lawson KA, et al: A national estimate of the economic costs of asthma. Am J Respir Crit Care Med 156:787, 1997

48. Guidelines for the diagnosis and management of asthma. National Heart, Lung, and Blood Institute, National Asthma Education and Prevention Program. Expert panel report 2, publication no 97-4051. National Institutes of Health, Bethesda, Maryland, 1997

49. Crim C: Clinical practice guidelines vs actual clinical practice: the asthma paradigm. Chest 118(2 suppl):62S, 2000

50. Taylor DM, Auble TE, Calhoun WJ, et al: Current outpatient management of asthma shows poor compliance with International Consensus Guidelines. Chest 116:1638, 1999

51. Markson LE, Vollmer WM, Fitterman L, et al: Insight into patient dissatisfaction with asthma treatment. Arch Intern Med 161:379, 2001

52. Partridge MR, Hill SR: Enhancing care for people with asthma: the role of communication, education, training and self-management. 1998 World Asthma Meeting Education and Delivery of Care Working Group. Eur Respir J 16:333, 2000

53. Vollmer WM, O'Hollaren M, Ettinger KM, et al: Specialty differences in the management of asthma: a cross-sectional assessment of allergists' patients and generalists' patients in a large HMO. Arch Intern Med 157:1201, 1997

54. Bartter T, Pratter MR: Asthma: better outcome at lower cost? The role of the expert in the care system. Chest 110:1589, 1996

55. Holtzman MJ: Drug development for asthma. Am J Respir Cell Mol Biol 29:163, 2003

56. Berger WE: Levalbuterol: pharmacologic properties and use in the treatment of pediatric and adult asthma. Ann Allergy Asthma Immunol 90:583, 2003

57. Korosec M, Novak RD, Myers E, et al: Salmeterol does not compromise the bronchodilator response to albuterol during acute episodes of asthma. Am J Med 107:209, 1999

58. Li X, Ward C, Thien F, et al: An antiinflammatory effect of salmeterol, a long-acting beta(2) agonist, assessed in airway biopsies and bronchoalveolar lavage in asthma. Am J Respir Crit Care Med 160:1493, 1999

59. Tattersfield AE, Lofdahl CG, Postma DS, et al: Comparison of formoterol and terbutaline for as-needed treatment of asthma: a randomised trial. Lancet 57:257, 2001

60. Jackson CM, Lipworth B: Benefit-risk assessment of long-acting beta$_2$-agonists in asthma. Drug Saf 27:243, 2004

61. Liggett SB: The pharmacogenetics of beta$_2$-adrenergic receptors: relevance to asthma. J Allergy Clin Immunol 105(2 pt 2):S487, 2000

62. Mullen M, Mullen B, Carey M: The association between beta-agonist use and death from asthma: a meta-analytic integration of case-control studies. JAMA 270:1842, 1993

63. Anderson HR, Ayres JG, Sturdy PM, et al: Bronchodilator treatment and deaths from asthma: case-control study. BMJ 330:117, 2005

64. Abramson MJ, Walters J, Walters EH: Adverse effects of beta-agonists: are they clinically significant? Am J Respir Med 2:287, 2003

65. Salpeter SR, Ormiston TM, Salpeter EE: Cardiovascular effects of beta-agonists in patients with asthma and COPD: a meta-analysis. Chest 125:2309, 2004

66. Israel E, Drazen JM, Liggett SB, et al: The effect of polymorphisms of the beta(2)-adrenergic receptor on the response to regular use of albuterol in asthma. Am J Respir Crit Care Med 162:75, 2000

67. Israel E, Chinchilli VM, Ford JG, et al: Use of regularly scheduled albuterol treatment in asthma: genotype-stratified, randomised, placebo-controlled cross-over trial. Lancet 364:1505, 2004

68. Weir TD, Mallek N, Sandford AJ, et al: Beta$_2$-adrenergic receptor haplotypes in mild, moderate and fatal/near fatal asthma. Am J Respir Crit Care Med 158:787, 1998

69. Rodrigo G, Rodrigo C, Burschtin O: A meta-analysis of the effects of ipratropium bromide in adults with acute asthma. Am J Med 107:363, 1999

70. Bowler SD, Mitchell CA, Armstrong JG: Corticosteroids in acute severe asthma: effectiveness of low doses. Thorax 47:584, 1992

71. Leung DY, Szefler SJ: Diagnosis and management of steroid-resistant asthma. Clin Chest Med 18:611, 1997

72. Lane SJ, Adcock IM, Richards D, et al: Corticosteroid-resistant bronchial asthma is associated with increased c-fos expression in monocytes and T lymphocytes. J Clin Invest 102:2156, 1998

73. Suissa S, Ernst P: Inhaled corticosteroids: impact on asthma morbidity and mortality. J Allergy Clin Immunol 107:937, 2001

74. Cochrane MG, Bala MV, Downs KE, et al: Inhaled corticosteroids for asthma therapy: patient compliance, devices, and inhalation technique. Chest 117:542, 2000

75. Barnes N: Relative safety and efficacy of inhaled corticosteroids. J Allergy Clin Immunol 101:S460, 1998

76. Vanden Burgt JA, Busse WW, Martin RJ, et al: Efficacy and safety overview of a new inhaled corticosteroid, QVAR (hydrofluoroalkane-beclomethasone extrafine inhalation aerosol), in asthma. J Allergy Clin Immunol 106:1209, 2000

77. Olivieri D, Chetta A, Del Donno M, et al: Effect of short-term treatment with low-dose inhaled fluticasone propionate on airway inflammation and remodeling in mild asthma: a placebo-controlled study. Am J Respir Crit Care Med 155:1864, 1997

78. van Grunsven PM, van Schayck CP, Molema J, et al: Effect of inhaled corticosteroids on bronchial responsiveness in patients with "corticosteroid naive" mild asthma: a meta-analysis. Thorax 54:316, 1999

79. Meijer RJ, Kerstjens HA, Arends LR, et al: Effects of inhaled fluticasone and oral prednisolone on clinical and inflammatory parameters in patients with asthma. Thorax 54:894, 1999

80. Lipworth BJ: Systemic adverse effects of inhaled corticosteroid therapy: a systematic review and meta-analysis. Arch Intern Med 159:941, 1999

81. Long-term effects of budesonide or nedocromil in children with asthma. The Childhood Asthma Management Program Research Group. N Engl J Med 343:1054, 2000

82. Salvi SS, Krishna MT, Sampson AP, et al: The anti-inflammatory effects of leukotriene-modifying drugs and their use in asthma. Chest 119:1533, 2001

83. Grayson M, Korenblat P: The emerging role of leukotriene modifiers in allergic rhinitis. Am J Respir Med 2:441, 2003

84. Virchow JC, Prasse A, Naya I, et al: Zafirlukast improves asthma control in patients receiving high-dose inhaled corticosteroids. Am J Respir Crit Care Med 162:578, 2000

85. Malmstrom K, Rodriguez-Gomez G, Guerra J, et al: Oral montelukast, inhaled beclomethasone, and placebo for chronic asthma: a randomized, controlled trial. Montelukast/Beclomethasone Study Group. Ann Intern Med 130:487, 1999

86. Busse W, Raphael GD, Galant S, et al: Low-dose fluticasone propionate compared with montelukast for first-line treatment of persistent asthma: a randomized clinical trial. J Allergy Clin Immunol 107:461, 2001

87. Busse W, Nelson H, Wolfe J, et al: Comparison of inhaled salmeterol and oral zafirlukast in patients with asthma. J Allergy Clin Immunol 103:1075, 1999

88. Weller PF, Plaut M, Taggart V, et al: The relationship of asthma therapy and Churg-Strauss syndrome: NIH workshop summary report. J Allergy Clin Immunol 108:175, 2001

89. Weinberger M, Hendeles L: Theophylline in asthma. N Engl J Med 334:1380, 1996

90. Lim S, Tomita K, Caramori G, et al: Low-dose theophylline reduces eosinophilic inflammation but not exhaled nitric oxide in mild asthma. Am J Respir Crit Care Med 164:273, 2001

91. Yasui K, Hu B, Nakazawa T, et al: Theophylline accelerates human granulocyte apoptosis not via phosphodiesterase inhibition. J Clin Invest 100:1677, 1997

92. Cosio BG, Mann B, Ito K, et al: Histone acetylase and deacetylase activity in alveolar macrophages and blood mononocytes in asthma. Am J Respir Crit Care Med 170:141, 2004

93. Evans DJ, Taylor DA, Zetterstrom O, et al: A comparison of low-dose inhaled budesonide plus theophylline and high-dose inhaled budesonide for moderate asthma. N Engl J Med 337:1412, 1997

94. Busse WW: Anti-immunoglobulin e (omalizumab) therapy in allergic asthma. Am J Respir Crit Care Med 164:S12, 2001

95. Bousquet J, Wenzel S, Holgate S, et al: Predicting response to omalizumab, an anti-IgE antibody, in patients with allergic asthma. Chest 125:1378, 2004

96. Wilson AM, Orr LC, Sims EJ, et al: Antiasthmatic effects of mediator blockade versus topical corticosteroids in allergic rhinitis and asthma. Am J Respir Crit Care Med 162:1297, 2000

97. Marin MG: Low-dose methotrexate spares steroid usage in steroid-dependent asthmatic patients: a meta-analysis. Chest 112:29, 1997

98. Bernstein IL, Bernstein DI, Dubb JW, et al: A placebo-controlled multicenter study of auranofin in the treatment of patients with corticosteroid-dependent asthma. Auranofin Multicenter Drug Trial. J Allergy Clin Immunol 98:317, 1996

99. Lock SH, Kay AB, Barnes NC: Double-blind, placebo-controlled study of cyclosporin A as a corticosteroid-sparing agent in corticosteroid-dependent asthma. Am J Respir Crit Care Med 153:509, 1996

100. Charous BL, Halpern EF, Steven GC: Hydroxychloroquine improves airflow and lowers circulating IgE levels in subjects with moderate symptomatic asthma. J Allergy Clin Immunol 102:198, 1998

101. Kon OM, Sihra BS, Compton CH, et al: Randomised, dose-ranging, placebo-controlled study of chimeric antibody to CD4 (keliximab) in chronic severe asthma. Lancet 352:1109, 1998

102. Salmun LM, Barlan I, Wolf HM, et al: Effect of intravenous immunoglobulin on steroid consumption in patients with severe asthma: a double-blind, placebo-controlled, randomized trial. J Allergy Clin Immunol 103:810, 1999

103. Adkinson NF Jr: Con: immunotherapy is not clinically indicated in the management of allergic asthma. Am J Respir Crit Care Med 164:2140, 2001

104. Bousquet J: Pro: immunotherapy is clinically indicated in the management of allergic asthma. Am J Respir Crit Care Med 164:2139, 2001

105. Platts-Mills TA, Vaughan JW, Carter MC, et al: The role of intervention in established allergy: avoidance of indoor allergens in the treatment of chronic allergic disease. J Allergy Clin Immunol 106:787, 2000

106. Stenius-Aarniala B, Poussa T, Kvarnstrom J, et al: Immediate and long term effects of weight reduction in obese people with asthma: randomised controlled study. BMJ 320:827, 2000

107. Anderson PJ: Delivery options and devices for aerosolized therapeutics. Chest 120:89S, 2001

108. Sont JK, Willems LN, Bel EH, et al: Clinical control and histopathologic outcome of asthma when using airway hyperresponsiveness as an additional guide to long-term treatment. The AMPUL Study Group. Am J Respir Crit Care Med 159:1043, 1999

109. Shrewsbury S, Pyke S, Britton M: Meta-analysis of increased dose of inhaled steroid or addition of salmeterol in symptomatic asthma (MIASMA). BMJ 320:1368, 2000

110. Lazarus SC, Boushey HA, Fahy JV, et al: Long-acting beta2-agonist monotherapy vs continued therapy with inhaled corticosteroids in patients with persistent asthma: a randomized controlled trial. JAMA 285:2583, 2001

111. Lemanske RF Jr, Sorkness CA, Mauger EA, et al: Inhaled corticosteroid reduction and elimination in patients with persistent asthma receiving salmeterol: a randomized controlled trial. JAMA 285:2594, 2001

112. Edelman JM, Turpin JA, Bronsky EA, et al: Oral montelukast compared with inhaled salmeterol to prevent exercise-induced bronchoconstriction: a randomized, double-blind trial. Exercise Study Group. Ann Intern Med 132:97, 2000

113. Dahlen B: Treatment of aspirin-intolerant asthma with antileukotrienes. Am J Respir Crit Care Med 161:S137, 2000

114. Namazy JA, Schatz M: Update in the treatment of asthma during pregnancy. Clin Rev Allergy Immunol 26:139, 2004

115. Stevens DA, Schwartz HJ, Lee JY, et al: A randomized trial of itraconazole in allergic bronchopulmonary aspergillosis. N Engl J Med 342:756, 2000

116. Imbeault B, Cormier Y: Usefulness of inhaled high-dose corticosteroids in allergic bronchopulmonary aspergillosis. Chest 103:1614, 1993

117. Chien JW, Ciufo R, Novak R, et al: Uncontrolled oxygen administration and respiratory failure in acute asthma. Chest 117:728, 2000

118. Rodrigo G, Rodrigo C: Corticosteroids in the emergency department therapy of acute adult asthma: an evidence-based evaluation. Chest 116:285, 1999

119. Rowe BH, Spooner CH, Ducharme FM, et al: Corticosteroids for preventing relapse following acute exacerbations of asthma. Cochrane Database Syst Rev (2):CD000195, 2000

120. Evans R 3rd, LeBailly S, Gordon KK, et al: Restructuring asthma care in a hospital setting to improve outcomes. Chest 116(4 suppl 1):210S, 1999

121. Peters JI, Shelledy DC, Jones AP Jr, et al: A randomized, placebo-controlled study to evaluate the role of salmeterol in the in-hospital management of asthma. Chest 118:313, 2000

Acknowledgment

Figures 1 and 2 Seward Hung.

III CHRONIC OBSTRUCTIVE DISEASES OF THE LUNG

Gerald W. Staton, Jr., m.d., f.a.c.p.

Chronic Obstructive Pulmonary Disease

definitions

Chronic obstructive pulmonary disease (COPD) is defined as a disease state characterized by airflow limitation that is not fully reversible. The airflow obstruction is usually both progressive and associated with an abnormal inflammatory response of the lungs to noxious particles and gases.[1] COPD, chronic obstructive lung disease, and chronic airflow obstruction (CAO) are synonyms used to describe the minimally reversible airflow obstruction caused by chronic bronchitis, emphysema, or both. These conditions are distinguished from asthma in that the abnormalities limiting airflow are for the most part irreversible, although irreversible airflow obstruction can develop in some patients with asthma[2] [*see 14:II Asthma*]. The inclusiveness of these terms stems from the fact that chronic bronchitis and emphysema usually occur together as the consequences of a common etiology (e.g., cigarette smoking). Chronic bronchitis and emphysema must be distinguished not only from asthma but also from bronchiectasis and obstructive bronchiolar disorders, which are covered in this chapter, and from upper airway obstruction and interstitial lung diseases that have an obstructive component (e.g., sarcoidosis and pulmonary lymphangioleiomyomatosis [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]).[3]

Chronic Bronchitis

Chronic bronchitis follows prolonged exposure of the tracheobronchial tree to nonspecific irritants and is characterized by hypersecretion of mucus and structural changes in the bronchi, including inflammation, metaplasia of the epithelium, and enlargement of the mucous glands. The inflammation and hypersecretion of mucus usually cause daily cough and sputum production, which help to identify the disease clinically. For epidemiologic purposes, chronic bronchitis is defined by the presence of cough and sputum on most days for at least 3 months of the year for a minimum of 2 years in succession.

Some cigarette smokers with chronic productive cough (often called smoker's cough) do not have dyspnea, and in fact, pulmonary function testing reveals no detectable obstruction of airflow in these patients. In other smokers, however, clinically significant narrowing of the airways develops. Because of these differing features, many physicians use the term chronic mucus hypersecretion or chronic bronchitis without airway obstruction to identify the former presentation and the term chronic obstructive bronchitis or chronic bronchitis with airway obstruction to identify the latter. The validity of such a distinction is supported by the fact that the two forms of chronic bronchitis have different natural histories.

In contrast to chronic bronchitis, acute bronchitis is a self-limited condition of the bronchi that is most often caused by viral infections in association with an upper respiratory tract illness or acute exposure to a nonspecific irritant (e.g., ammonia). Acute bronchitis, however, may also occur as a transient complication of chronic bronchitis.

Emphysema

Emphysema is a destructive process involving the lung parenchyma. It is defined in pathologic terms on the basis of findings of abnormal permanent enlargement of air spaces distal to the terminal bronchiole, accompanied by the destruction of the alveolar walls without obvious fibrosis. At autopsy, some areas of emphysema are found in the lungs of most persons older than 60 years.

epidemiology

Chronic bronchitis and emphysema are by far the most common causes of CAO; other causes include bronchiectasis and bronchiolitis. In the United States, an estimated 10 million adults report physician-diagnosed COPD; data suggest that 24 million adults have abnormal lung function, which points to significant underdiagnosis of COPD.[4] In 2000, COPD caused 119,054 deaths, representing a 128% increase since 1980.[4] COPD is the fourth leading cause of death in the United States and the fifth leading cause of death worldwide. From 80% to 90% of cases of COPD can be attributed to cigarette smoking. The risk of death from chronic bronchitis or emphysema is 30 times greater for heavy smokers (i.e., those who smoke 25 or more cigarettes a day) than for nonsmokers.

In cigarette smokers, the prevalence of chronic bronchitis increases with age (and, thus, with the increase in the total cumulative amount of cigarette smoke inhaled) and the number of cigarettes smoked. More than 50% of middle-aged men who smoke at least one pack of cigarettes a day report having chronic productive cough. Cigar and pipe smokers, as well as former cigarette smokers, have a rate of chronic bronchitis higher than the general population but lower than current cigarette smokers. Although occupational dust exposure is also associated with the development of chronic bronchitis, the effect of cigarette smoking overwhelms the occupational effect.

Some degree of emphysematous changes can be found in the autopsied lungs of 65% of men and 15% of women. Severe involvement, leading to a clinical diagnosis of emphysema, occurs in slightly less than 1% of the population. However, prevalence statistics are more relevant for cigarette smokers, who constitute the major population at risk. In a series involving more than 1,800 autopsies, the prevalence of emphysema judged to be advanced to far advanced was 0% in nonsmokers, 12% in smokers of less than one pack of cigarettes a day, and 19% in those who consumed one or more packs a day.[5]

Forced expiratory volume in 1 second (FEV_1) and the ratio of FEV_1 to forced vital capacity (FEV_1/FVC) can be used to assess the prevalence of CAO. If an FEV_1 value of less than 80% of predicted and an FEV_1/FVC of less than 70% is defined as abnormal, 7.4% of men and 5.8% of women in the general adult population would be considered to have obstructive lung disease.[3] There is a normal, or bell-shaped, distribution of FEV_1 values in the nonsmoking population. In cigarette smokers, this distribution curve is shifted toward lower FEV_1 values [*see Figure 1*]. The prevalence of abnormal values increases with increasing years of cigarette smoking. For instance, after 40 to 60 pack-years of cigarette smoking (number of pack-years = packs of ciga-

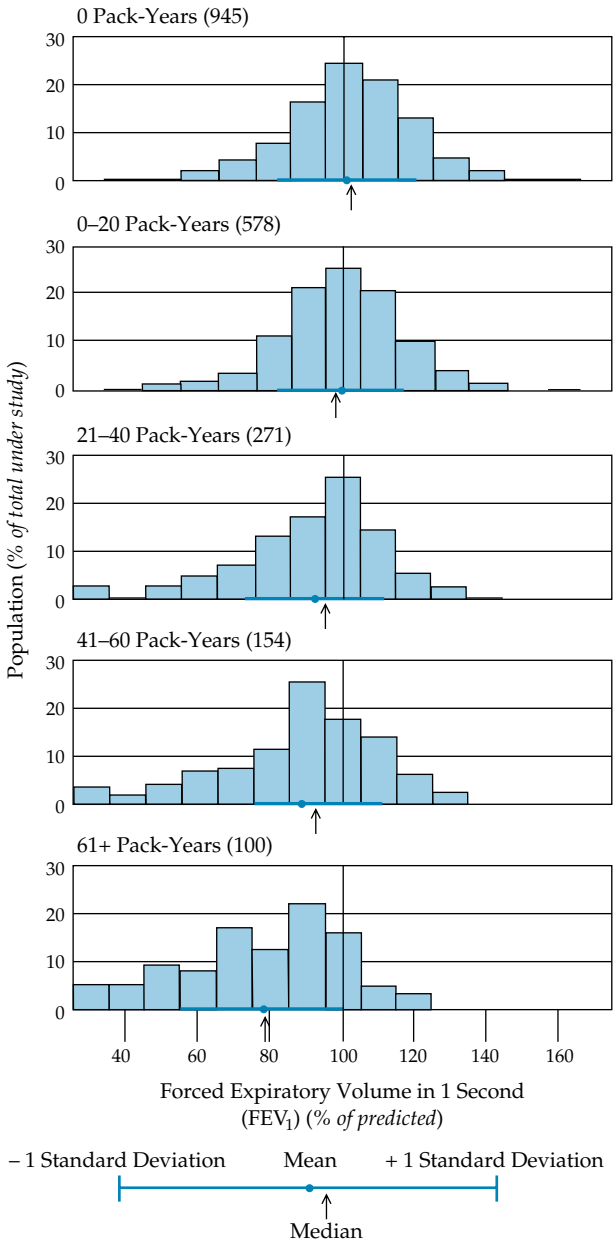


Figure 1 **There is a normal distribution of expiratory flow values, as measured by percent of the predicted forced expiratory volume in 1 second (FEV_1), among nonsmokers. Among cigarette smokers, the distribution is shifted toward lower FEV_1 values, with a larger skew at the lower range, as the number of pack-years of smoking increases.[182]**

rettes/day × number of years the person has smoked), most smokers have slightly reduced expiratory flow, and approximately 10% manifest CAO.

ETIOLOGY

Chronic bronchitis and emphysema are usually caused by cigarette smoking of at least 10 years' duration. The specific pathogenic substances in cigarette smoke have not been identified with certainty. It is likely that particle size and gas solubility determine which of the hundreds of noxious substances found in cigarette smoke are deposited in the trachea and the large bronchi, causing bronchitis, and which are deposited in the bronchioles and the gas exchange units, causing narrowing of the small airways and emphysema.

Cigarette smoking is such a dominant factor in the pathogenesis of COPD that it has been difficult to identify any causative or contributing factors other than α_1-antitrypsin deficiency [see α_1-Antitrypsin Deficiency, *below*]. Prolonged exposure to respirable dusts in the work environment has long been recognized as a cause of so-called industrial or occupational bronchitis in nonsmoking workers engaged in occupations such as coal or gold mining, textile manufacturing, and cement and steel making. An analysis of the accumulated data over the past 25 years suggests that these exposures are associated with the development of significant COPD.[6] Chronic HIV or hepatitis C infection,[7,8] low body mass,[9] and low intake of dietary antioxidants may be cofactors in some cigarette smokers.[10]

In otherwise healthy adults, passive cigarette smoking (i.e., indoor exposure of nonsmokers to cigarette smoke) and outdoor air pollution may cause nonspecific respiratory symptoms such as cough and may induce subtle abnormalities in pulmonary function, but they have not been shown to lead to the development of COPD.[11] A major unknown in the etiology of COPD is why clinically significant airway obstruction develops in only 10% to 15% of smokers. Hypotheses for the unique susceptibility of some cigarette smokers to the development of severe airflow obstruction suggest the presence of host factors and exposure factors. Host factors include preexisting airway hyperresponsiveness[12]; impaired lung growth; smaller size of terminal bronchioles; and genetic determinants favoring increased elastase activity, impaired antiprotease function, or other factors.[13,14] Exposure factors include deeper and more frequent inhalation of tobacco smoke, occupational dust and chemical exposure, indoor (including possible effects of passive smoking) and outdoor air pollution, and chronic bacterial or viral infection.[15]

NATURAL HISTORY

The dominant feature of the natural history of chronic bronchitis and emphysema is the development and progression of airflow obstruction. In healthy nonsmokers, FEV_1 begins declining at about 20 years of age at an average rate of about 0.02 to 0.04 L/yr. In smokers with obstructive lung disease, FEV_1 decreases, on average, two to three times faster than normal [*see Figure 2*]. A person with an FEV_1 of 2.0 L (50% of normal) at 45 years of age who continues to smoke cigarettes heavily is apt to be breathless on light exertion at 55 years age; for such a smoker, the FEV_1 would be expected to have fallen to about 1.2 L, or 30% of normal, at 55 years of age. At 60 years of age, the expected FEV_1 would be about 0.8 L (20% of normal). The person would probably be breathless at rest or on minimal exertion and would be at risk for recurrent episodes of acute respiratory failure. At the same estimated rate of lung function loss, this person would not be expected to live past 65 years of age.

Other factors may modify the development and progression of COPD. In addition to advanced age, variables that appear to confer a worse prognosis include frequent exacerbations,[16] persistent airway bacterial colonization,[17] hypoxemia, hypercapnia, cor pulmonale, and, possibly, increased airway responsiveness and eosinophilia (asthmatic bronchitis).[12] When smokers with mild to moderate airflow obstruction stop smoking, the rate of decline in expiratory flow reverts to that observed in nonsmokers, and there may be a slight improvement in FEV_1 during the first year.[18] Cough and phlegm production usually decrease, as well.[19] Unfortunately, when obstruction is already severe, smok-

ing cessation may not be as effective in altering future deterioration. There remains considerable variability among individuals, so that no single factor or group of factors can be used to predict the outcome in a given patient.

PATHOLOGY

Early Pathologic Changes

Pathologic changes are apparent in the lungs of smokers even before the appearance of COPD.[20] Breaks in the alveolar septal walls are seen, possibly representing a precursor to more generalized emphysema and inflammation of the bronchioles. The lungs of cigarette smokers, even in those as young as 25 years, can be distinguished from those of nonsmokers by the abnormal accumulation of pigmented macrophages in respiratory bronchioles, often accompanied by edema, epithelial hyperplasia, and fibrosis in adjacent bronchiolar and alveolar walls. The significance of these early findings is threefold. First, peripheral airways (i.e., bronchioles and small bronchi less than 2 mm in diameter) are an important site of airflow obstruction, both in asymptomatic cigarette smokers and in patients with advanced COPD. Second, the pathologic changes that occur in patients with the centriacinar form of emphysema [*see* Established Emphysema, *below*] initially develop in the respiratory bronchioles. Third, evidence suggests that the early abnormalities observed in cigarette smokers are for the most part reversible with cessation of smoking.

Airflow through healthy lungs is limited primarily by the resistance of the central airways (i.e., the trachea and the major bronchi) and the upper airways (i.e., the larynx and the pharynx) but also, to some degree, by the resistance of the small airways of the lung periphery. As COPD worsens, there is a marked increase in airflow resistance; almost all of this increase is in the peripheral airways. The degree of pathologic narrowing of the terminal bronchioles is strongly correlated with all physiologic indices of airflow obstruction, including the reduced FEV_1. The small airways either are directly narrowed by inflammatory changes or, as in emphysema, lose the tethering effect provided by the surrounding lung parenchyma, which normally limits narrowing during expiration.

Lungs of asymptomatic young smokers show clusters of macrophages in and around respiratory bronchioles. Other early pathologic lesions include squamous cell metaplasia of the epithelium, chronic inflammation of airway walls, small foci of peribronchiolar fibrosis, and muscular hypertrophy of the bronchiolar wall.[21] As COPD progresses, mural connective tissue, pigment, and muscle become more prominent. In addition, goblet cell hyperplasia develops,[22] and mucosal ulcerations appear.

Mild bronchiolitis is probably a universal finding in young smokers. The degree to which these terminal airways are narrowed by the inflammation has a strong correlation with obstruction, as measured by spirometry.[23]

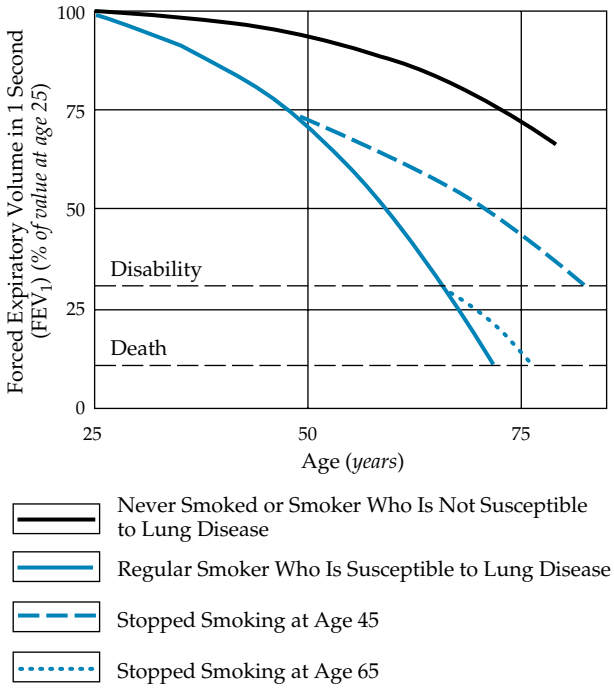


Figure 2 **Expiratory flow, as measured by FEV_1, declines up to two to three times faster in cigarette smokers who are susceptible to obstructive lung disease than in nonsmokers or in those who smoke but are not susceptible to lung disease. Smoking cessation can decrease the rate of fall in FEV_1 in susceptible cigarette smokers, delaying the onset of disability or death.[183]**

Late Pathologic Changes

Established chronic bronchitis The pathologic hallmark of chronic bronchitis is enlargement of the mucous glands in the major bronchi. The diameter of the mucous glands relative to the thickness of the bronchial wall is greater in patients with chronic bronchitis.

Although excess mucus in the airways is common in patients with chronic bronchitis, it has physiologic significance only when it obstructs the lumina of small airways. The small number and irregular distribution of goblet cells among the epithelial cells lining the airways make quantification difficult. It is possible that the mucus found plugging small airways is not produced locally but originates in the central airways. Functional disorders that may contribute to mucous plugging include secretion of mucus with altered viscoelastic properties and impaired ciliary clearance mechanisms.

It has not been established whether the volume of bronchial smooth muscle is increased in patients with chronic bronchitis, as it is in those with asthma. Smooth muscle hyperplasia has been found in the main and lobar bronchi in those patients with chronic bronchitis who have intermittent attacks of wheezing.

Bronchoscopic findings include generalized erythema, edema, and friability of the tracheobronchial mucosa. Mucous gland pits, which are depressions in the mucosa approximately 1 to 2 mm in diameter that represent the orifice of one or more mucous glands, are also present. Often, the most striking finding is dynamic expiratory collapse of large central airways; however, this is the result of obstruction of small peripheral airways rather than the cause of the expiratory airflow limitation.

Established emphysema The terminology commonly used by pathologists to describe pulmonary emphysema makes reference to two distinct units of lung structure: the acinus and the lobule. The acinus is an anatomic subdivision that consists of a respiratory bronchiole together with all the alveolar ducts and alveoli extending from it [*see Figure 3*]. A lobule is defined as the smallest discrete portion of the lung surrounded by connective tissue septa. Emphysema may involve the acinus and the lobule uniformly in a pattern called panacinar or panlobular emphyse-

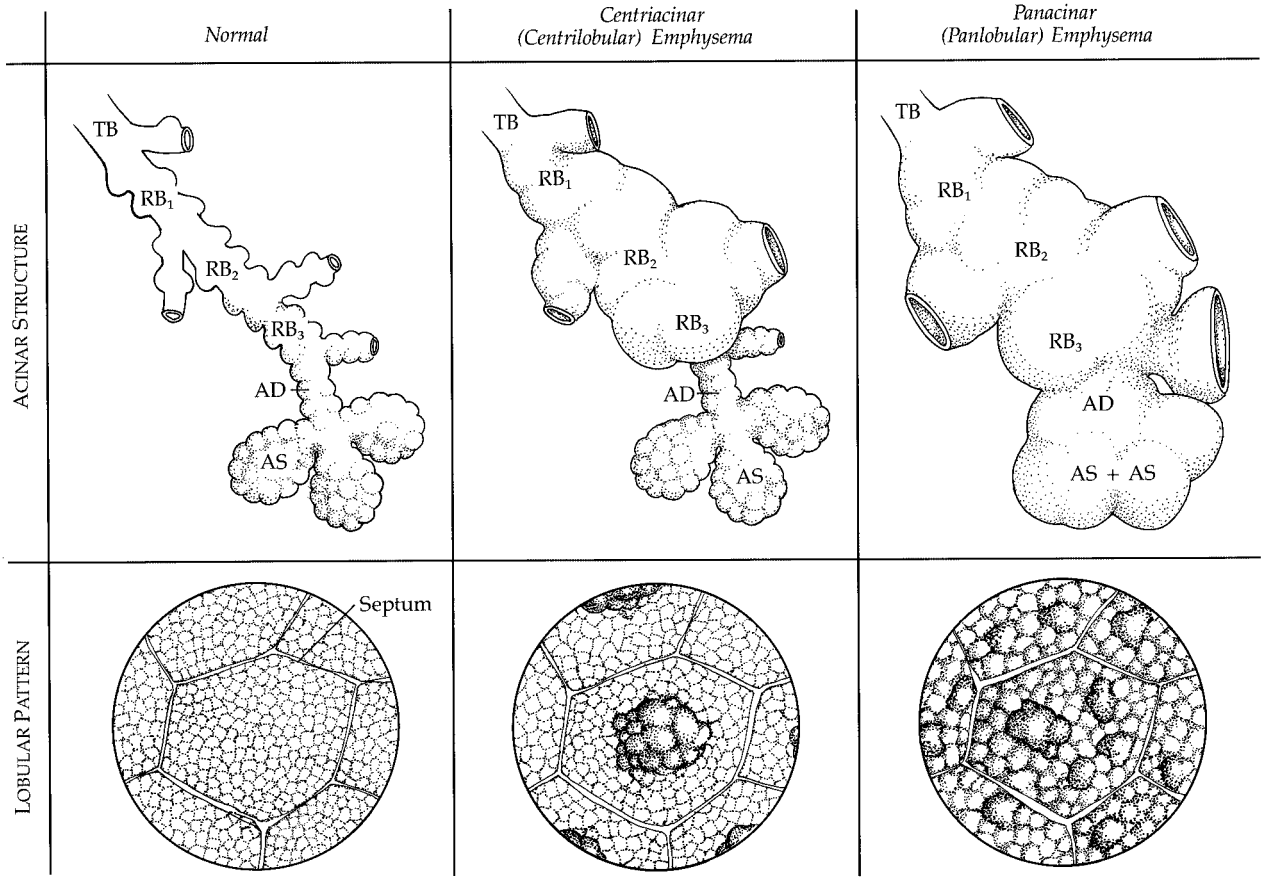


Figure 3 **In the top row, the acinar structure of normal lungs is compared with that of lungs in patients with centriacinar (centrilobular) emphysema or panacinar (panlobular) emphysema. The normal acinus has a clearly defined structure consisting of the terminal bronchiole (TB); first-, second-, and third-order respiratory bronchioles (RB_1, RB_2, and RB_3, respectively); alveolar ducts (AD); and alveolar sacs (AS). In centriacinar emphysema, there is selective enlargement and destruction predominantly of the respiratory bronchioles. In contrast, panacinar emphysema is defined by universally enlarged and destroyed air spaces throughout the acinus. Patterns of lobular destruction are depicted in the bottom row. The normal pulmonary lobule is a macroscopic structure, the borders of which can be identified by the presence of connective tissue septa. In centriacinar emphysema, the predominant site of overdistention is in the center of the lobule, with relative sparing toward the periphery—hence the name centrilobular emphysema. In contrast, there are uniform destructive changes throughout the lobule in panlobular emphysema.**

ma. Alternatively, it may involve primarily the respiratory bronchioles; this form of the disease is termed centriacinar, centrilobular, or, perhaps most correctly, proximal acinar emphysema.[24]

Panacinar emphysema is common in patients with α_1-antitrypsin deficiency [*see* α_1-Antitrypsin Deficiency, *below*]. Pathologic examination reveals enlargement of the alveolar ducts and sacs with loss of distinguishing features. Holes, or fenestrae, develop in the alveolar walls and progressively enlarge. The process of destruction ultimately leaves only strands of tissues containing supportive structures such as vessels, bronchi, and septa. The emphysematous spaces are spread uniformly across the lobule. Typically, the lower lobes show more involvement than the upper lobes.

Centriacinar emphysema is commonly found in cigarette smokers and is rare in nonsmokers. This form of emphysema primarily affects the respiratory bronchioles. The destroyed respiratory bronchioles enlarge and coalesce, forming emphysematous spaces surrounded by relatively normal alveolar ducts and alveoli. The emphysematous spaces are situated in the midportion of the lobule, separated from the septa by normal alveolar tissue. The small airways supplying these emphysematous spaces typically show signs of inflammation and are often narrowed. Centriacinar emphysema is usually more extensive and severe in the upper lobes.

Pure centriacinar emphysema is not a common finding at autopsy. In most cigarette smokers, a mixture of centriacinar and panacinar emphysema develops. This mixture may take the form of predominant centriacinar emphysema in the upper lung zones and panacinar emphysema in the lower zones or centriacinar emphysema that extends to the alveolar ducts and alveoli to varying degrees. Distinction between these two forms of emphysema is often difficult in advanced cases and is of no clinical significance. However, there may be subtle differences in the altered lung function between the two forms, possibly related to greater abnormalities of small airways and more fibrosis in centriacinar emphysema.

Other abnormalities found at autopsy in patients with COPD include atrophy of cartilage in the walls of segmental and subsegmental bronchi, obliteration of small airways (i.e., loss of bronchioles 0.4 to 0.6 mm in diameter), increased longitudinal smooth muscle in the intima of pulmonary arterioles and small arteries, and dilatation of bronchial arteries and veins.

Inflammatory cells are an important host factor in the development of COPD.[24-26] Activated macrophages, lymphocytes with a predominance of $CD8^+$ T cells, and neutrophils are found in increased numbers in the respiratory bronchioles of smokers. These activated macrophages release the neutrophil chemotactic factors leukotriene B_4 and interleukin-8 (IL-8), which are amplified by tumor necrosis factor–α (TNF-α). Complex interactions between these cells and mediators lead to increased amounts of proteases and oxidants in the lung parenchyma, airways, and possibly adjacent lung vessels.[27] Two interrelated causes—an imbalance of proteases and antiproteases and an imbalance of oxidants and antioxidants—are hypothesized to produce the chronic bronchitis and emphysema seen in cigarette smokers.

Protease-Antiprotease Imbalance

One current theory of the pathogenesis of emphysema holds that the disease may develop because of an imbalance between protease and antiprotease activities in the lungs.[24-26,28] The origins of this theory date to 1963, when the Scandinavian researchers Laurell and Eriksson identified a group of patients with a deficiency in α_1-antitrypsin (also called α_1-proteinase inhibitor), a major serum antiprotease [*see* α_1-Antitrypsin Deficiency, *below*]. Emphysema developed in a remarkably high percentage of these patients, typically at an unusually young age.

Most cigarette smokers have normal serum α_1-antitrypsin concentrations. In fact, cigarette smoking may cause a small (approximately 20%) rise in serum α_1-antitrypsin concentrations. (Increased amounts of α_1-antitrypsin often can be retrieved from the lungs of smokers by means of lavage.) Nonetheless, long-term inhalation of cigarette smoke has a number of effects that can lead to an excess of protease activity in the lung parenchyma [*see Figure 4*]. First, cigarette smoke activates alveolar macrophages to secrete proteases (multiple cathepsins and matrix metalloproteinases) and mediators that attract neutrophils; these or additional mediators also stimulate neutrophils to release more than the usual amount of cathepsin G and neutrophil elastase. Second, several oxidants that are present in cigarette smoke or are generated from products of cigarette smoke, such as the oxides of nitrogen, interact with hydrogen peroxide released from activated alveolar macrophages and neutrophils to oxidize and inactivate α_1-antitrypsin and other antiproteases.[29] Third, cigarette smoke may inhibit the synthesis of elastin, thereby retarding repair of damaged elastin fibers.

Oxidant-Antioxidant Imbalance

A number of studies have found evidence of an increased oxidative burden in cigarette smokers.[29] Sources of oxidants include cigarette smoke and the inflammatory cells (macrophages and neutrophils). In addition, the antioxidant system appears to be inadequate for dealing with the increased oxidants, producing an imbalance. Adverse effects of the oxidant stress include inactivation of antiproteases, membrane lipid peroxidation, DNA and matrix damage, epithelial injury, and stimulation of transcription of inflammatory cytokines, amplifying the inflammation.

Obviously, these two hypotheses are intertwined, suggesting that both mechanisms have a role in the production of COPD. As a consequence of the excesses of proteases and the inhibition of antiproteases, the walls of the respiratory bronchioles and the alveoli are damaged and the altered repair mechanisms prevent remodeling and fibrosis, resulting in emphysema. The oxidative imbalance damages the walls of the airway

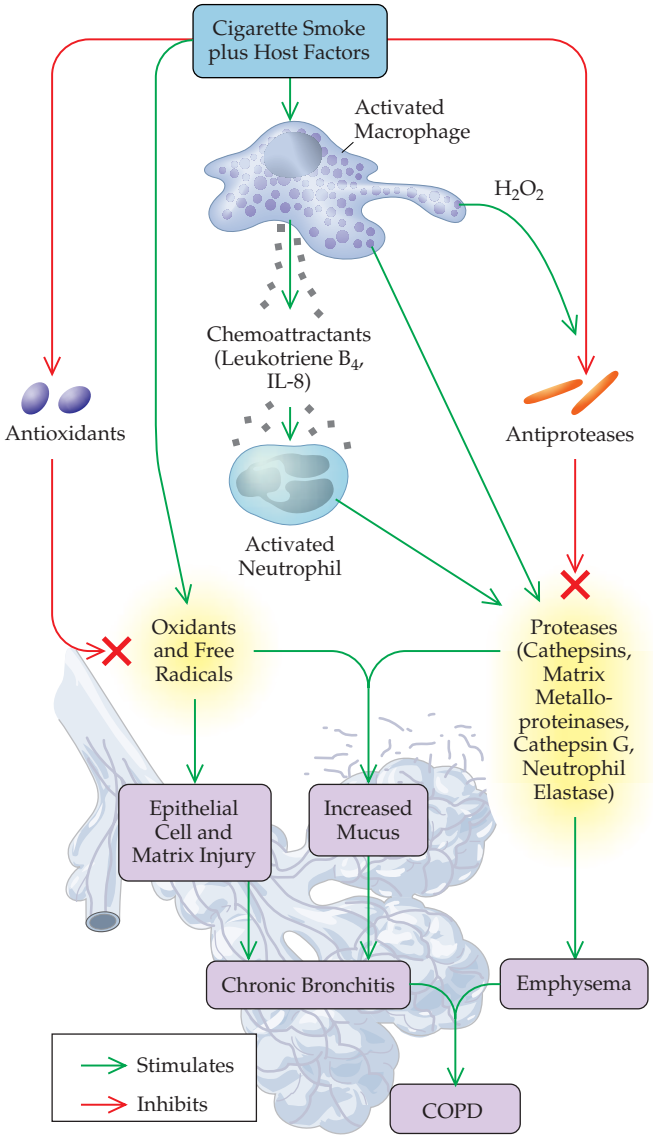


Figure 4 **Host factors play a role in the development of chronic obstructive pulmonary disease (COPD) in smokers. Cigarette smoke activates alveolar macrophages to secrete proteases (i.e., multiple cathepsins and matrix metalloproteinases). The activated macrophages release the neutrophil chemoattractants leukotriene B_4 and interleukin-8 (amplified by tumor necrosis factor–α [TNF-α]), and the chemoattractants also stimulate neutrophils to release more than the usual amount of cathepsin G and neutrophil elastase. Several oxidants that are present in cigarette smoke or are generated from products of cigarette smoke interact with hydrogen peroxide released from activated alveolar macrophages and neutrophils to oxidize and inactivate α_1-antitrypsin and other antiproteases. Cigarette smoke may inhibit the synthesis of elastin, thereby retarding repair of damaged elastin fibers. There is an increased oxidative burden from cigarette smoke and the inflammatory cells (macrophages and neutrophils), and the antioxidant system appears to be inadequate for dealing with the increased oxidants, producing an imbalance. Adverse effects of the oxidant stress include inactivation of antiproteases, membrane lipid peroxidation, DNA and matrix damage, and epithelial injury. As a consequence of the excesses of proteases and the inhibition of antiproteases, the walls of the respiratory bronchioles and the alveoli are damaged, and the altered repair mechanisms prevent remodeling and fibrosis, resulting in emphysema. The oxidative imbalance damages the walls of the airway and, along with the excess proteases, stimulates mucus hypersecretion, producing chronic bronchitis.**

Table 1 Differentiating Features in Advanced Chronic Airflow Obstruction

Features	*Type A—Pink Puffer (Predominant Emphysema)*	*Type B—Blue Bloater (Predominant Chronic Bronchitis)*
Symptoms and Signs	Dyspnea (first and predominant symptom); patients are usually thin, weight loss is common; minimal or no cough; hyperinflated lung fields; no signs of cor pulmonale	Cough and sputum production with frequent chest infections; stocky build; recurrent or persistent signs of right heart failure
Routine Laboratory Studies		
Chest radiograph	Hyperinflation; decreased vascular markings, bullae	Normal or increased markings at lung bases (so-called dirty-chest appearance)
Arterial blood gases	Mildly reduced P_aO_2; normal or decreased P_aCO_2	Marked reduction in P_aO_2; increased P_aCO_2
Total lung capacity	Increased	Normal or slightly increased
D_{LCO}	Decreased	Normal
Hematocrit	Normal	Increased
Specialized Laboratory Studies		
Inspiratory resistance	Normal	Increased
Pulmonary compliance	Increased	Normal
Ventilation-perfusion distribution	Increased V_D/V_T	Increased regions of low $\dot{V}_A/\dot{Q}$
Hemodynamics	Normal or decreased cardiac output Mild pulmonary hypertension	Normal cardiac output Marked pulmonary hypertension
Ventilatory Performance and Gas Exchange during Exercise	More wasted ventilation D_{LCO} fails to increase normally Decreased P_aO_2; small rise in P_aCO_2	Less wasted ventilation D_{LCO} increases normally P_aO_2 may increase; moderate rise in P_aCO_2
Gas Exchange during Sleep	Moderate degree of oxygen desaturation	Frequent periods of profound oxygen desaturation

D_{LCO}—diffusing capacity of lung for carbon monoxide P_aCO_2—arterial carbon dioxide tension P_aO_2—arterial oxygen tension V_D/V_T—ratio of dead space to tidal volume $\dot{V}_A/\dot{Q}$—ventilation/perfusion distribution

and, along with the excess proteases, stimulates mucus hypersecretion, producing chronic bronchitis.

DIAGNOSIS

Clinical Manifestations

The hallmarks of COPD are chronic productive cough and persistent, progressive exercise limitation because of breathlessness. Although these two symptoms are nonspecific, they usually point to the correct diagnosis of COPD in the context of prolonged cigarette smoking. Patients with predominant chronic bronchitis present with chronic productive cough, whereas patients with predominant emphysema complain of dyspnea [*see Table 1*]. Cough with expectoration develops 10 to 12 years after smoking begins. The symptom is often dismissed as a simple smoker's cough. However, persons with established chronic bronchitis commonly produce 2 oz or more of mucoid sputum a day. Transient periods of sputum discoloration caused by purulence commonly occur, often in association with respiratory tract infections. A small amount of hemoptysis may accompany a superimposed acute tracheobronchitis.

Wheezing is also common. It can occur transiently as mucus accumulates in airways and resolve suddenly with the expectoration of phlegm; however, some patients with chronic bronchitis experience prolonged and severe attacks of wheezing that mimic those of asthma. This combination of chronic bronchitis and reversible bronchospasm is commonly referred to as asthmatic bronchitis. These patients tend to respond more strongly to anti-inflammatory and bronchodilator drugs.

Shortness of breath, rather than cough, is more often the reason that patients in whom emphysema predominates seek medical attention. Although exertional dyspnea correlates in general with the degree of airflow obstruction, wide variability among individuals makes it impossible to predict the extent of respiratory impairment on the basis of any single value or set of values for expiratory flow. Nevertheless, as a broad guideline, only minimal limitation is imposed until the value for FEV_1 falls below 65% of normal. As airflow obstruction progresses, dyspnea develops with more moderate levels of exertion. When the FEV_1 value drops below 35% to 40% of normal, the patient may become breathless during activities of daily living such as making the bed or bathing.

Orthopnea is often present in patients with advanced airway disease, especially in cases in which increased airway secretions accompany significant airflow limitation. The orthopnea of COPD must be differentiated from the paroxysmal nocturnal dyspnea of chronic heart failure. This differentiation is based on the rapidity of onset of dyspnea with the patient lying supine: dyspnea is said to develop almost immediately in those with COPD but to be delayed for as long as a few hours in those with congestive heart failure. However, the distinction is not clear-cut, because patients with florid heart failure often refuse to lie supine even for brief periods, and patients with COPD may awaken after several hours of sleep with cough, shortness of breath, and chest congestion—symptoms that mimic paroxysmal nocturnal dyspnea. Relief is obtained by coughing up secretions. To further complicate matters, the two processes most often occur together.

COPD may have systemic effects. Possibly as a consequence of systemic inflammation,[30] nutritional abnormalities, and weight loss, COPD may be accompanied by skeletal muscle dysfunction (contributing to exercise limitations), coronary artery disease,[31] low serum testosterone levels (in males),[32] and osteoporosis.[33] Patients with COPD are also at increased risk for developing lung cancer.[34]

Physical Examination

Physical findings vary with the severity of disease and the relative contributions of chronic bronchitis and emphysema [*see*

Table 1]. In the early stages, physical examination may yield entirely normal results. With more advanced disease, tachypnea and a prolonged expiratory phase of the respiratory cycle are usually present. Emphysematous hyperinflation of the lungs may cause a hyperresonant percussion note and an unusually low position of the diaphragm. Breath sounds may be reduced by decreased airflow, and wheezes are heard in 40% or more of patients, especially if patients are in the supine position when examined. Rales are heard in some patients, particularly at the posterior lung bases, in the absence of heart failure. These are usually heard during the entire inspiratory phase, rather than solely at end-inspiration, which is the usual finding in interstitial lung disease (e.g., asbestosis or idiopathic pulmonary fibrosis). Clubbing of the digits is not a manifestation of COPD.

Laboratory Tests

In most patients who present with the symptoms and signs of COPD [*see* Clinical Manifestations, *above*], the order of testing should be as follows: (1) pulmonary function tests (including measurement of arterial blood gases), (2) chest radiography, and, in rare instances, (3) computed tomography of the chest.

Pulmonary function tests In the dyspneic patient, routine pulmonary function tests depict the characteristic pattern of volume-dependent airway obstruction [*see 14:I Lung Function Assessment and Thoracic Diagnostic Techniques*].[35] Spirometry, which should be performed in all smokers with any respiratory symptoms and all smokers older than 45 years,[36] reveals a reduction in FEV_1/FVC and an even greater relative decline in the forced expiratory flow between 25% and 75% of vital capacity (FEF_{25-75}). As the airflow obstruction worsens, a normal volume of gas can no longer be exhaled in the time available, and vital capacity also declines.

Measurement of lung volumes uniformly reveals an increased residual volume (RV) and a normal to increased functional residual capacity (FRC). RV may be two to four times higher than normal because of slowing of expiratory flow and trapping of gas behind prematurely closed airways. FRC increases by two mechanisms: dynamic hyperinflation and activation of inspiratory muscles during exhalation. As a result, tidal breathing may take place at lung volumes as high as 1 to 2 L above normal levels. The advantage of an increased FRC to the patient with significant airflow obstruction is enlarged airway diameter with greater radial support (which means less airway resistance) and increased driving pressure (i.e., elastic recoil) for exhalation. The cost to the patient is the greater work of breathing incurred at higher lung volume. Total lung capacity (TLC) is normal or increased in COPD. As would be predicted, there is a correlation between the extent of emphysematous lung destruction found at postmortem examination and the TLC. However, the correlation is not close, so that only a markedly increased TLC would be used to gauge the severity of emphysema.

Pulmonary function tests are useful in patients with COPD to confirm the obstructive abnormality, to quantify the severity of the defect, to assess the reversibility of the airflow obstruction in response to therapy,[37] and to monitor the course of the disease. A low FEV_1/FVC and a decrease in expiratory flow rates prove obstruction, but the best measurement for assessing the severity of the obstruction is the FEV_1. A scheme for determining severity of COPD has been proposed [*see Figure 5*]. Construction of flow-volume loops may be useful if upper airway obstruction is a diagnostic consideration.

Abnormalities in gas exchange It has long been recognized that the pattern of gas-exchange abnormalities in COPD may differ greatly among patients with airflow obstruction of identical severity. Early in the course of disease, when expiratory flow is only slightly reduced, mild hypoxemia may be the only blood gas abnormality. However, in advanced stages of COPD, two distinct patterns emerge [*see Table 1 and Figure 6*]. Patients in one group (type A) exhibit dyspnea with only mild to moderate hypoxemia (arterial oxygen tension [P_aO_2] levels are usually > 65 mm Hg) and maintain normal or even slightly reduced arterial carbon dioxide tension (P_aCO_2) levels. These patients are some-

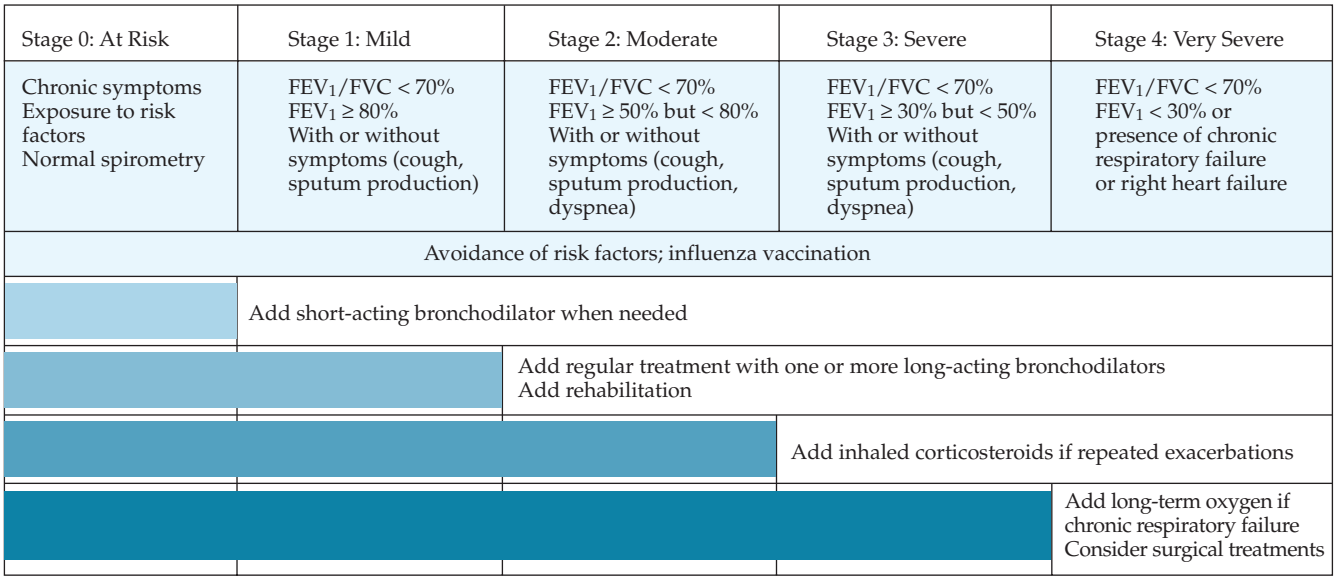
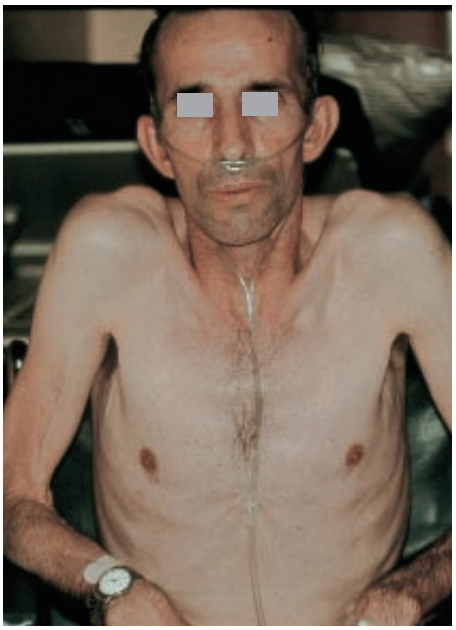


Figure 5 **Staging of COPD based on data derived from spirometry and recommended therapy for each stage. Therapy escalates as disease becomes more severe.[43]**



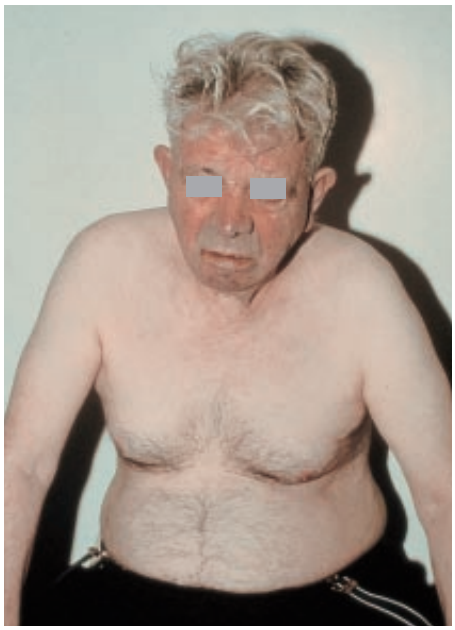


Figure 6 **In the advanced stages of chronic airflow obstruction, two distinct clinical types may emerge. Emphysema patients (so-called pink puffers, left) exhibit dyspnea without significant hypoxemia and tend to be thin, to have hyperinflated lung fields at total lung capacity, and to be free of signs of right heart failure. Chronic bronchitis patients (so-called blue bloaters, right) are characterized by marked hypoxemia with cyanosis and peripheral edema resulting from right heart failure.**

times referred to as pink puffers; they tend to be thin, to experience hyperinflation at total lung capacity, and to be free of signs of right heart failure. The pink puffer usually has severe emphysema. The other clinical group (type B) is characterized by marked hypoxemia and peripheral edema resulting from right heart failure. These patients, sometimes called blue bloaters, typically exhibit cough and sputum production. They have frequent respiratory tract infections, experience chronic carbon dioxide retention (P_aCO_2 > 45 mm Hg), and have recurrent episodes of cor pulmonale. The blue bloater may also have pathologic evidence of severe emphysema but, in addition, suffers from inflammation of large and small airways and possible defects in ventilatory control. These patients usually meet the criteria for chronic bronchitis. Many patients will demonstrate features of both types, giving rise to either mixed or intermediate clinical presentations.

Several explanations have been proposed for this disparate response to severe airflow obstruction, including differences in pulmonary mechanics, in the central control of ventilation, and in respiratory disturbances during sleep. Type A patients appear to have advanced emphysema with relatively little evidence of airway disease. These patients have abnormally compliant lungs, and airflow obstruction is related primarily to a loss of lung elastic recoil. In contrast, type B patients typically have intrinsic airway disease that manifests clinically as cough and sputum production (i.e., chronic bronchitis) and physiologically as increased resistance to airflow during both expiration and inspiration (airflow resistance during inspiration is not as apparent in patients with emphysema and loss of lung elastic recoil). Type B patients may compensate for the increased work of breathing imposed by this inspiratory resistive load by limiting their total amount of ventilation, causing P_aCO_2 levels to rise. Patients with both reduced elastic recoil and increased inspiratory resistance (i.e., those with a mixed type of clinical presentation) would be predicted to have chronic hypercapnia, and this is indeed the case.

It is also possible that patients with severe COPD in whom hypercapnia ultimately develops have intrinsic defects that cause relatively depressed ventilatory responses to acute rises in P_aCO_2 levels, falls in P_aO_2 levels, or both. This hypothesis is supported by the finding that the healthy first-degree relatives of patients with COPD and chronic carbon dioxide retention have blunted ventilatory responses to hypercapnia and hypoxia, as compared with the first-degree relatives of patients with COPD but no chronic hypercapnia.

Alternatively, arousal responses during sleep may be abnormally depressed in type B patients. Failure to arouse and resume normal ventilation may explain why, during sleep, type B patients experience more severe and prolonged decreases in P_aO_2 levels and increases in P_aCO_2 levels than normal persons or type A patients. Over time, repeated periods of hypercapnia during sleep may diminish the potency of hypercapnia as a stimulus to increase ventilation, eventually leading to fixed carbon dioxide retention.[38]

The more profound hypoxemia observed in type B patients, as compared with type A patients, is in part a result of overall alveolar hypoventilation. In addition, ventilation-perfusion ($\dot{V}/\dot{Q}$) mismatching in type B patients involves a greater number of alveoli receiving low levels of ventilation for the amount of blood flow delivered (i.e., a low ratio of alveolar ventilation to perfusion [$\dot{V}_A/\dot{Q}$]).

These two different patterns of abnormal gas exchange have very different consequences for the cardiovascular system in patients with advanced COPD. In the type B patient, both alveolar hypoxia and acidosis (secondary to chronic hypercapnia) stimulate pulmonary arterial vasoconstriction, and hypoxemia stimulates erythrocytosis. Increased pulmonary vascular resistance, increased pulmonary blood volume, and, possibly, increased blood viscosity from secondary erythrocytosis all contribute to pulmonary arterial hypertension.[39] In response to long-term pulmonary hypertension, cor pulmonale generally develops. In cor pulmonale, the right ventricle becomes hypertrophic; to increase cardiac output, abnormally high right ventricular filling pressures then occur. Additional hemodynamic loads may cause the right ventricle to fail, with the consequent development of systemic venous hypertension, manifested by jugular venous dis-

tention, peripheral edema, passive hepatic congestion, and, sometimes, ascites. (Pleural effusion is not a manifestation of cor pulmonale in the absence of left heart failure.) Echocardiographic evaluation of right heart function and estimation of pulmonary arterial systolic pressure are useful in quantifying the degree of pulmonary hypertension.[39]

The emphysematous lung destruction characteristically seen in type A patients leads to a restricted vascular bed because of the loss of pulmonary capillaries from the destroyed alveolar walls. This condition is reflected in the reduced diffusing capacity of the lung for carbon monoxide ($D_{L_{CO}}$) observed in type A (but not in type B) patients.[40] However, because P_aO_2 levels are only mildly depressed in type A patients, pulmonary vasoconstriction is minimal, and secondary erythrocytosis does not develop. Cardiac output may be slightly reduced. As a result, pulmonary hypertension in type A patients is milder than in type B patients, and cor pulmonale develops infrequently, usually only in the terminal phase of the illness.

Differences in gas exchange during exercise also distinguish the two clinical types. Type A patients have abnormally high levels of ventilation for a given workload (expressed in terms of oxygen consumption, $\dot{V}O_2$), but nevertheless, the P_aO_2 falls with exercise. The $D_{L_{CO}}$ in type A patients rises minimally and remains low during exercise, reflecting inadequacies in the alveolar-capillary surface area available for gas exchange. Patients with COPD who have $D_{L_{CO}}$ levels less than 55% of normal can be expected to have oxygen desaturation with exercise. In contrast, type B patients have subnormal increases in ventilation for a given level of work (decreased ratio of minute ventilation to oxygen consumption [$\dot{V}_E/\dot{V}O_2$]) and a greater rise in P_aCO_2 levels with exercise than type A patients, and yet, P_aO_2 levels may increase during exercise. This increase in P_aO_2 levels must reflect improved matching of ventilation to perfusion during exercise.

Chest Imaging

Radiographic abnormalities may be minimal, even in cases of advanced COPD.[41] When correlations are made between radiographic and pathologic findings in advanced COPD, the results of chest radiography suggest a diagnosis of emphysema in fewer than half of the cases, even among those with the highest emphysema scores pathologically.

Three types of radiographic abnormalities, when paired with the appropriate clinical history, suggest the diagnosis of emphysema [*see Table 1*].[41] The first abnormality is arterial deficiency in the lung periphery; on chest radiograph, narrowed or absent vessels in the lung periphery are associated with hyperlucency, usually in a symmetrical, bilateral distribution. The second abnormality relates to hyperinflation and is evident in the standard posteroanterior and lateral chest radiographs, which are obtained at TLC; radiographic signs of hyperinflation in these films include a low position of the diaphragm (i.e., at or below the seventh rib anteriorly), increased depth of the retrosternal air space, and a narrow, vertically oriented cardiac silhouette. Perhaps the most useful sign of hyperinflation is a flattening of the diaphragmatic contour and loss of the diaphragm's normal domed appearance, especially as visualized on the lateral film. The third abnormality is bullous disease. The presence of a bulla together with either of the other two radiographic findings is virtually diagnostic of emphysema, although only a small percentage of patients afflicted with emphysema have bullae [*see* Bullous Lung Disease, *below*].

Chronic bronchitis is rarely recognized on a chest radiograph [*see Table 1*].[41] On occasion, the diagnosis may be suspected because of visualization of thickened bronchial walls, particularly in a parahilar bronchus viewed end-on. In other patients, bronchovascular markings at the lung bases may be accentuated—a pattern that has been dubbed the dirty chest of chronic bronchitis. A similar radiographic appearance has been referred to as the increased-markings pattern of emphysema, especially when increased vascular markings are observed in the presence of pulmonary hypertension and cor pulmonale. The precise pathologic correlate of this radiographic image is unknown.

CT of the chest can be useful in the differential diagnosis of COPD.[41] Patients without abnormalities in pulmonary function may have extensive upper lobe emphysema, as demonstrated by CT. Occasional smokers who have normal airflow on spirometry but diminished $D_{L_{CO}}$ may have emphysema that can be detected only by high-resolution CT scanning.

TREATMENT

Reviews and guidelines on the treatment of COPD have been published,[25,35,42-44] but the various guidelines do not agree on recommendations.[45] Suboptimal prescription of and adherence to appropriate therapies further complicate management of COPD.[46] Of the therapeutic measures available for patients with chronic bronchitis and emphysema, only smoking cessation and long-term administration of supplemental oxygen to the chronically hypoxemic patient have been definitively shown to alter the natural history of the disease favorably. Data suggest, however, that the combination of a long-acting $beta_2$ agonist and inhaled corticosteroids may also improve survival of patients with COPD.[47]

Helping a patient to quit smoking is probably the single most important intervention; effective methods include counseling by physicians and nurses, use of nicotine replacement therapy, behavior intervention (e.g., individual or group therapy), and several pharmacologic interventions (e.g., bupropion and nortriptyline)[48] [*see CE:III Reducing Risk of Injury and Disease*]. Spirometric evidence of airway obstruction in patients who smoke, when combined with advice to stop smoking, may contribute to increased smoking-cessation rates.[49] Smoking cessation generally causes the symptoms of chronic bronchitis to diminish or entirely remit,[19] and it stops the accelerated loss of lung function observed with continued cigarette smoking.[18,50]

Reduction in morbidity and mortality in patients with COPD can be achieved with yearly vaccinations against influenza.[51] Administration of the pneumococcal vaccine is also advisable, although its efficacy is less well documented in this patient population.[52] In the future, immunization with *Haemophilus influenzae* vaccine may be recommended.[53]

A variety of other therapies offer potential relief of symptoms in patients with COPD. These include the use of bronchodilators; anti-inflammatory therapy; administration of antibiotics during acute purulent exacerbations; pulmonary rehabilitation programs, including physical exercise and respiratory muscle training; and, for patients with cor pulmonale, the use of diuretics. Based on the GOLD (*G*lobal Initiative for the Diagnosis, Management, and Prevention of Chronic *O*bstructive *L*ung *D*isease) staging system, a stepwise plan for management of patients with COPD has been proposed [*see Figure 6*].[43]

Management of Chronic Disease

Bronchodilators COPD has been considered a disorder of irreversible airflow obstruction, in contradistinction to asthma.

Table 2 Estimated Comparative Doses of Inhaled Bronchodilators for COPD

Category (Relative Efficacy)	*Drug*	*How Available*	*Maintenance Dosage*	*Exacerbation Dosage*	*Comment*
Inhaled long-acting anticholinergics (first-line therapy)	Tiotropium	DPI, 18 µg/puff	1 capsule q. day	Not recommended	Anticholinergic of choice for maintenance therapy
Inhaled short-acting anticholinergics	Ipratropium bromide	Nebulizer, 0.25 mg/ml	0.5 mg q. 6 hr	0.5 mg q. 2–8 hr	May mix with albuterol in same nebulizer
		MDI, 18 µg/puff	2–6 puffs q. 6 hr	3–8 puffs q. 3–4 hr	MDI as effective as nebulizer when used with spacer
Inhaled long-acting beta$_2$ agonists (may replace anticholinergic)	Salmeterol	MDI, 21 µg/puff	2 puffs q. 12 hr	Not recommended	Slower onset of action than short-acting beta$_2$ agonists
		DPI, 50 µg/blister	1 blister q. 12 hr	Not recommended	
	Formoterol	DPI, 12 µg/capsule	1 capsule q. 12 hr	Not recommended	Faster onset of action than salmeterol
Inhaled short-acting beta$_2$ agonists (second-line therapy)	Albuterol	Nebulizer, 5 mg/ml	1.25–5.0 mg q. 4–8 hr p.r.n.	5.0 mg q. 2 hr	Dilute aerosols to minimum of 4 ml at gas flow of 6–8 L/min
		MDI, 84 µg/puff	2–4 puffs q. 6 hr p.r.n.	3–8 puffs q. 2 hr	MDI as effective as nebulizer when used with spacer
		DPI, 200 µg/capsule	1–2 capsules q. 6 hr p.r.n.	Not studied in exacerbations	—
	Bitolterol	Nebulizer, 2 mg/ml	0.5–3.5 q. 4–8 hr p.r.n.	Not studied in exacerbations	—
		MDI, 370 µg/puff	2–4 puffs q. 6 hr p.r.n.	Not studied in exacerbations	—
	Pirbuterol	MDI, 200 µg/puff	2–4 puffs q. 6 hr p.r.n.	Not studied in exacerbations	—
Combined short-acting beta$_2$ agonist and anticholinergic (when both are indicated)	Albuterol + ipratropium bromide	MDI; albuterol, 90 µg/puff + ipratropium, 18 µg/puff	2–4 puffs q. 6 hr p.r.n.	3–8 puffs q. 2 hr	—

DPI—dry powder inhaler MDI—metered-dose inhaler

Nevertheless, most patients with COPD who are given a sufficiently strong bronchodilating medication will exhibit at least a 10% increase in maximal expiratory airflow.[37] Dyspneic patients should be given a trial of bronchodilators even if pulmonary function testing shows that they do not manifest significant bronchodilation, because bronchodilator responsiveness may vary over time.[37] Additionally, some of the symptomatic benefit from bronchodilators in COPD patients may come from reduction in exercise-induced dynamic hyperinflation, a change that may not be detected by spirometry.[25,54]

Bronchodilator therapy should be administered on the basis of the patient's symptoms and in a stepwise fashion on the basis of disease stage. For COPD patients in stages 1 through 4, a short-acting inhaled bronchodilator such as albuterol should be used as needed for treatment of symptoms [*see Figure 5 and Table 2*]. For symptomatic patients in stages 2 through 4, long-acting bronchodilators should be prescribed and used on a regular basis[55] [*see Figure 5 and Table 2*]. Current recommendations suggest the long-acting bronchodilator of choice is tiotropium, an approved anticholinergic drug that is inhaled once daily [*see Table 2*]. Alternatively, a long-acting beta$_2$ agonist (i.e., salmeterol or formoterol) may be used[55] [*see Table 2*]. An additional benefit of either type of long-acting bronchodilator therapy may be a reduction in the frequency of exacerbations.[55]

If the patient's symptoms are incompletely controlled with one long-acting inhaled bronchodilator, a second long-acting inhaled bronchodilator should be added. Whether the combination of anticholinergic and beta$_2$-agonist therapies will result in additive or synergistic bronchodilation and a reduction in exacerbations is unknown. When a combination of a long-acting beta$_2$ agonist and inhaled corticosteroid is required, a single an inhaler containing salmeterol and fluticasone can be used, potentially improving compliance.[56]

If the patient remains symptomatic on optimized inhaled medication, a trial of theophylline is indicated.[57] Because of theophylline's narrow therapeutic window, dosing must be individualized; serum levels from 8 to 12 mg/dl are optimal for COPD patients.

Bronchodilators may be delivered by three types of devices: nebulizers, metered-dose inhalers (MDIs), and dry-powder inhalers (DPIs). Nebulizers are useful in patients who cannot operate MDIs or DPIs, but they are bulky and messy to use. MDIs are familiar to many patients but require excellent coordination or the use of a spacer. DPIs come in several forms, some easier to use than others. Patients must be instructed in the use of each apparatus, and the technique of their use must be reviewed with the patient periodically. All three types of devices can be used to deliver medication to spontaneously breathing patients. Only nebulizer systems and MDIs can be used in intubated patients. There is no evidence that nebulized bronchodilators are of

greater benefit than properly administered dry-powder or metered-dose inhaled medications used with a spacer.[58] Use of bronchodilators in the setting of acute exacerbations is discussed elsewhere [*see* Management of Acute Exacerbations, *below*].

Corticosteroids The role of corticosteroids in the treatment of COPD is controversial; however, the current understanding that all forms of COPD are associated with a tissue inflammatory response [*see Figure 4*] suggests that anti-inflammatory therapy may alter disease progression and symptoms.

Patients who have recurrent attacks of wheezing, a significant acute response to inhaled bronchodilators (> 20% increase in FEV_1), and sputum eosinophilia may have COPD with an asthmatic component (asthmatic bronchitis); these patients may respond well to corticosteroids.[59]

Oral corticosteroid therapy should be avoided in COPD except in the setting of acute exacerbations [*see* Management of Acute Exacerbations, *below*]. Oral steroids required on a regular basis should be given in the lowest possible dose, preferably on alternate days.

Inhaled corticosteroids may have significant beneficial effects on airway inflammation [*see Table 3*].[60] In randomized trials, inhaled steroids did not arrest decline in lung function in patients with COPD[61]; however, their use was associated with reductions in respiratory symptoms, in acute exacerbations, and in use of health care services.[62] Medium to high doses of inhaled corticosteroids may be of greater benefit than lower doses [*see Table 3*].[47] Analyses of large databases have suggested that inhaled corticosteroids, especially when used with long-acting $beta_2$ agonists, may also increase survival.[47,63] The most recent guidelines recommend inhaled corticosteroids for COPD patients with stage 3 or 4 disease who have frequent exacerbations [*see Figure 5*].[43,64]

The main issue in corticosteroid therapy for COPD, as in any therapy, is how to achieve maximal benefit with minimal toxicity. If prolonged oral corticosteroid therapy is given to patients who do not show objective improvement in pulmonary function, the net effect will be negative. Complications associated with prolonged oral corticosteroid therapy include weight gain, osteoporosis,[33] hypertension, diabetes, cataracts, and myopathy. Serious pulmonary infections, including locally invasive aspergillosis, can ensue.

Side effects from inhaled corticosteroids in COPD patients are much less than those associated with oral corticosteroid use: osteoporosis[65] and adrenal suppression[66] have not been detected, and weight gain, hypertension, diabetes, and myopathy are not seen. Inhaled corticosteroids may carry a small risk for the development of cataracts.

Antibiotics In the 1950s and 1960s, clinical trials of daily antibiotic use in patients with mild COPD demonstrated that this intervention did not significantly alter either the degree of disability or the rate of disease progression. Only patients with very frequent purulent exacerbations (so-called chronic purulent bronchitis) derived some symptomatic benefit from regular antibiotic use.

Management of Acute Exacerbations

Acute exacerbations are short-term deteriorations of clinical and functional status of COPD patients.[67,68] Acute exacerbations of COPD are associated with worsening dyspnea (often at rest), increased cough (often with purulent sputum), and the development of wheezing, fatigue, fever, tachypnea, cyanosis, edema, and alteration of consciousness.

Symptoms that suggest an exacerbation of COPD can be caused by other diseases or complications of COPD. Examples of the diseases that can mimic a COPD exacerbation include pneumothorax, pneumonia, congestive heart failure, cardiac arrhythmia, pulmonary embolism, drug toxicity, metabolic disorders, stroke, and myopathy.

The immediate decision in the management of an acute worsening of COPD is whether the exacerbation is severe enough to require emergent evaluation and possible hospitalization. Criteria for immediate evaluation include a requirement of maximum therapy to control symptoms, alteration of consciousness, dyspnea at rest, a respiratory rate greater than 25 breaths/min, a pulse rate greater than 110 beats/min, increased cyanosis, and use of accessory muscles. As COPD becomes more advanced, patients are more likely to require emergent care when exacerbations occur. Other factors that should be considered include comorbidities; the medical sophistication, judgment, and reliability of the patient and caregivers; and the distance to the medical center or clinic.

Table 3 Estimated Comparative Daily Dosages of Inhaled Corticosteroids[184]

Drug	*Low Dosage*	*Medium Dosage*	*High Dosage*	*Relative Efficacy*
Beclomethasone 42 µg/puff 84 µg/puff	168–504 µg 4–12 puffs/day 2–6 puffs/day	504–840 µg 12–20 puffs/day 6–10 puffs/day	> 840 µg > 20 puffs/day > 10 puffs/day	Third-highest potency
Budesonide 200 µg/puff	200–400 µg 1–2 inhalations/day	400–600 µg 2–3 inhalations/day	> 600 µg > 3 inhalations/day	Second-highest potency
Flunisolide 250 µg/puff	500–1,000 µg 2–4 puffs/day	1,000–2,000 µg 4–8 puffs/day	> 2,000 µg > 8 puffs/day	Lowest potency
Fluticasone 44 µg/puff 110 µg/puff 220 µg/puff	 88–264 µg 2–6 puffs/day 1–2 puffs/day	 264–660 µg 6–15 puffs/day 2–6 puffs/day	 > 660 µg > 15 puffs/day > 6 puffs/day	Highest potency
Triamcinolone 100 µg/puff	400–1,000 µg 4–10 puffs/day	1,000–2,000 µg 10–20 puffs/day	> 2,000 µg > 20 puffs/day	Lowest potency

Antibiotics Exacerbations are often precipitated by respiratory infection, which may be caused by viral[69] or bacterial pathogens or a combination of the two. In patients with symptoms of infection, the duration of symptoms and the risk of serious deterioration in lung function can be reduced by a 7- to 10-day course of broad-spectrum antibiotics. Recent guidelines suggest that with advancing age, severity of disease, and the presence of comorbidity, the pathogens encountered in exacerbations tend to be more resistant, necessitating use of broader-spectrum antibiotics; sputum culture is required in the most severe illnesses.[70] For patients with simple chronic bronchitis (i.e., those with an $FEV_1 > 50\%$), in the absence of pneumonia, empirical treatment with ampicillin, tetracycline, or trimethoprim-sulfamethoxazole is satisfactory. In patients whose FEV_1 is less than 50%, many of the isolates of *H. influenzae, Moraxella* (*Branhamella*) *catarrhalis,* and *Streptococcus pneumoniae* are β-lactam resistant; and other organisms, such as members of the Enterobacteriaceae family and *Pseudomonas* species, become more common.[69] In these patients, a quinolone, a β-lactam–β-lactamase inhibitor combination, a second-generation oral cephalosporin, or a second-generation macrolide may be needed.

Bronchodilators The bronchodilator of choice in exacerbations of COPD is a short-acting $beta_2$ agonist (e.g., albuterol), mainly because of its rapid onset of action. The initial dose is four puffs every 4 hours. In patients who do not respond, doses can be increased and the frequency of dosing decreased to three to eight puffs every 2 hours [*see Table 2*].

Once short-acting $beta_2$-agonist therapy is started, inhaled anticholinergic therapy (e.g., ipratropium) should be initiated or increased. The dosage is usually three or four puffs but can be increased to five to eight puffs every 3 to 4 hours [*see Table 2*]; the onset of action is slower than that of beta agonists, but the duration of action is longer.

A nebulizer has no advantage over an MDI used with a large-volume spacer. The only instances in which a nebulizer is preferred to an MDI is in patients who have extreme dyspnea or altered consciousness. There are no data suggesting that the addition of theophylline to bronchodilator therapy is beneficial for exacerbations of COPD.[71] In patients already on theophylline, measurement of theophylline level is indicated because acute illness and some of the medications used to treat exacerbations can precipitate theophylline toxicity.

Corticosteroids Corticosteroids are beneficial in exacerbations of COPD,[72] but they are associated with side effects of varying severity. Patients should be considered for prednisone therapy (1) if they are currently on oral or inhaled steroids, (2) if they have recently finished a regimen of oral steroids, (3) if they previously responded to oral or inhaled steroids, or (4) if they are not responding to initial bronchodilator therapy. The dosage of prednisone is 0.6 to 1.0 mg/kg/day. In addition, patients whose oxygen saturation is 90% or less or whose peak expiratory flow rate is 100 L/min or less should be considered for treatment with corticosteroids regardless of their current therapy.

Exacerbations are often associated with hypercapnia and worsening of hypoxemia. Oxygen supplementation should be adjusted to keep oxygen saturation around 90% to 92% to maintain tissue oxygenation while minimizing the risk of worsening hypercapnia. In patients whose hypercapnia increases and whose blood pH values are less than 7.35 despite initial bronchodilator therapy, ventilatory support should be considered [*see 14:VII Respiratory Failure*].[73]

Management of Advanced Disease

In patients with far-advanced COPD who have cor pulmonale and right heart failure, therapies are often directed specifically at the hemodynamic consequences of the disease [*see 14:XI Pulmonary Hypertension, Cor Pulmonale, and Primary Pulmonary Vascular Diseases*].[74] Diuretics are administered for symptomatic relief of peripheral edema; they may also reduce pulmonary arterial pressure by decreasing intrapulmonary blood volume. Care must be taken to avoid chloride depletion from long-term diuretic use because the resultant hypochloremic metabolic alkalosis depresses respiratory drive and may aggravate CO_2 retention. Digoxin and aminophylline are only weak inotropic stimulants of the right ventricle. The increases in right ventricular output produced by these medications are not clinically significant.

A variety of arterial vasodilators, including hydralazine and nifedipine, have been used in an attempt to reduce pulmonary arterial hypertension secondary to severe COPD. A limited number of studies have demonstrated that long-term (i.e., 6 to 12 month) vasodilator administration produces persistent reductions in pulmonary arterial pressures, but there is no evidence that such therapy produces sustained symptomatic improvement or prolongs survival.[74] A potential risk of vasodilator therapy, particularly in the patient with fixed (i.e., anatomic) pulmonary hypertension, is the development of systemic hypotension and renal hypoperfusion. In general, relief of hypoxemia with long-term supplemental oxygen administration has proved to be more effective than drug therapy in reducing pulmonary arterial resistance.

Pulmonary Rehabilitation

Physical-training programs, such as treadmill walking, significantly increase the exercise capacity of patients with even far-advanced chronic bronchitis and emphysema.[75] These results have been achieved despite the fact that lung function, as reflected in such measurements as vital capacity and FEV_1, is not affected and maximal heart rate is generally not reached during the training sessions.

Much interest has focused on the possibility that resistive or endurance exercises aimed at strengthening the inspiratory muscles might increase respiratory muscle strength and endurance and thereby improve exercise capacity, at least in cases in which respiratory muscle fatigue contributes to exercise limitation. For instance, it was shown that patients with severe COPD who breathed through an inspiratory resistor (with the amount of resistive load carefully tailored to the particular patient) for 15 minutes twice daily for 1 month derived a number of benefits. Such a regimen not only increased the endurance of the inspiratory muscles but also enabled patients to walk farther during a fixed period, and it increased their endurance during submaximal exercise on a bicycle. However, it is not clear whether the benefit derived from respiratory muscle training exceeds that derived from general physical conditioning. As with other forms of physical training, deconditioning occurs within a few weeks of cessation of the exercise program.

Long-term Oxygen Administration

Probably the single most important advance in recent years in the treatment of advanced COPD has been the widespread

home use of supplemental oxygen.[76] Chronic hypoxemia causes secondary erythrocytosis and contributes to exercise limitation, pulmonary hypertension, right heart failure, and impaired neuropsychiatric function. Each of these consequences of chronic hypoxemia can be ameliorated or even corrected by the administration of oxygen at levels that are sufficient to maintain the arterial oxygen saturation at values exceeding approximately 90%. This target value, which corresponds to a P_aO_2 of 60 to 80 mm Hg, can usually be achieved with oxygen supplementation delivered at a rate of 2 L/min via nasal cannulas, but the rate of oxygen administration should be titrated to the individual patient's needs according to arterial blood gas measurements taken when the patient is breathing supplemental oxygen.

Supplemental oxygen administration for a minimum of 15 hr/day may suffice to ameliorate some of the adverse consequences of chronic tissue hypoxia, particularly secondary erythrocytosis and pulmonary hypertension. However, survival is significantly improved when supplemental oxygen is given for 24 hr/day.

In general, patients should receive long-term home oxygen supplementation if the P_aO_2 is 55 mm Hg or less or if the P_aO_2 is 59 mm Hg or less and one or more of the following conditions are present: peripheral edema (a sign of cor pulmonale), a hematocrit of 55% or greater, and P pulmonale on the electrocardiogram. The resting P_aO_2 should be assessed when the patient is free of an exacerbation and has received intensive bronchodilator therapy. With adequate oxygen therapy, phlebotomy for polycythemia (i.e., hematocrit > 50%) is rarely necessary.

Oxygen for home use can be stored in cylinders as compressed gas or as liquid oxygen, or it can be generated from ambient air by machines called oxygen concentrators that are the size of a bedside table. Portable tanks of oxygen, filled from a liquid oxygen reservoir tank, can provide supplemental oxygen for several hours of use outside the home. Newer methods to conserve oxygen consumption and prolong the time spent outside the home while using portable oxygen tanks include devices that release oxygen only during the inspiratory phase of the respiratory cycle and direct transtracheal administration of oxygen via an indwelling tracheal catheter.[77]

Long-term Ventilatory Support

Intermittent application of negative pressure or noninvasive positive pressure ventilation has been used in the treatment of patients with chronic hypercapnia caused by advanced chronic bronchitis and emphysema.[78] During periods of mechanical ventilation, the respiratory muscles are rested and gas exchange improves. Trial results in a limited number of patients suggested that daytime symptoms, arterial blood gas levels, and exercise performance may be improved when only a few hours of ventilation are applied nightly or even less frequently.

Surgery

Interest in surgical therapy for diffuse emphysema has been revived because of advances in the operative technique. Emphysematous lung tissue is removed by one of several approaches and techniques.[79] The rationale is to remove overdistended, poorly functioning emphysematous regions, thereby allowing more-normal or less-affected regions to expand with a decrease in the FRC and an improvement in inspiratory muscle function.[80] The surgery produces a 13% to 96% improvement in FEV_1 and a reduction in RV; many patients who previously required supplemental oxygen are able to discontinue it or reduce their use after surgical therapy. However, results vary enormously among patients.[81] Operative mortality is 5% to 7%. Patients usually stay in the hospital 7 to 10 days; the main morbidity is prolonged air leaks.[82] Decreases in TLC, increases in elastic recoil pressures, and lengthening of inspiratory muscles have been shown to be directly related to functional improvement.[83] The surgery can be performed on many patients who are not candidates for lung transplantation, or it may be performed instead of lung transplantation or as a bridge to transplantation.[84] By 5 years after surgery, only 8% of the patients continue to have a significant improvement in FEV_1.[85] In two small randomized trials, patients treated with pulmonary rehabilitation plus surgery, compared with patients treated with pulmonary rehabilitation alone, were shown to have improvement in lung function, exercise capacity, and quality of life.[86,87]

A randomized, multicenter clinical trial comparing lung volume reduction surgery with continued medical treatment in 1,218 patients with severe emphysema found that the surgery increased the chance of improved exercise capacity but did not confer a survival advantage, except in patients who had both predominantly upper lobe emphysema and low exercise capacity after rehabilitation.[88] Patients with an FEV_1 20% or less than predicted and either homogeneous emphysema on high-resolution CT scanning or a diffusion capacity 20% or less than predicted had a higher 30-day postoperative mortality than the medically treated group.[89] Patients with non–upper lobe emphysema and high baseline exercise capacity proved to be poor candidates because of operative mortality and negligible functional gains.[88]

In younger patients with far-advanced COPD, lung transplantation is a therapeutic option. Criteria for referral for lung transplantation include an FEV_1 less than 20% of predicted value, hypercapnia, homogeneous emphysema on high-resolution CT scan, and pulmonary hypertension.[90] In patients with pure emphysema, single-lung transplantation has been performed successfully despite the concern that the remaining emphysematous lung would become further overexpanded because of the normal recoil of the transplanted lung.[91] In patients with chronic bronchitis who have any evidence of concomitant bronchiectasis, bilateral lung transplantation is performed at many centers to eliminate the potential risk of persistent infection in the remaining lung. Heart-lung transplantation is not needed even by patients with cor pulmonale, because the right ventricle rapidly recovers when pulmonary vascular resistance is reduced. The 5-year survival rate after lung transplantation is 57%.[92]

Other Conditions Associated with CAO

α_1-ANTITRYPSIN DEFICIENCY

Degradation of interstitial elastin fibers by elastolysis is a central pathologic process in the development of emphysema [*see* Protease-Antiprotease Imbalance, *above*].[93] The main defense of normal lung tissue against this enzymatic destruction is alveolar α_1-antitrypsin, which inactivates elastase by forming a stable complex with it. α_1-Antitrypsin is synthesized predominantly in the liver and is present in the serum of normal persons at concentrations of 83 to 200 mg/dl if measured by nephelometry and 150 to 350 mg/dl (20 to 50 μM) if measured by radial immunodiffusion.

Genetic Variants Affecting α_1-Antitrypsin Levels

The genetic model describing inheritance of α_1-antitrypsin

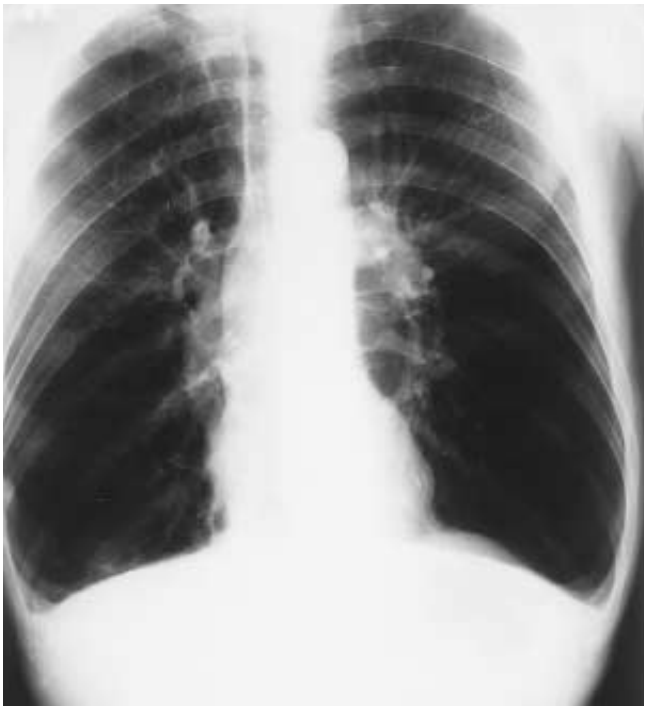


Figure 7 **This chest radiograph of a patient with homozygous α_1-antitrypsin deficiency reveals hyperinflation and hyperlucent lung fields with bilateral basilar bullae. Other findings include prominent central pulmonary arteries and rapid pruning of the pulmonary arterial branches, consistent with pulmonary arterial hypertension.**

genes invokes multiple autosomal codominant alleles at a single locus.[94] The locus on chromosome 14 has been designated Pi, which stands for protease inhibitor system. The most prevalent α_1-antitrypsin allele, found with a frequency of 0.95 in whites in the United States, is labeled Pi^M. Variant α_1-antitrypsin molecules are denoted with letters on the basis of their electrophoretic mobility. The variant protein molecule with the lowest rate of electrophoretic mobility is Pi^Z.

Most of the various Pi phenotypes are associated with normal amounts of serum α_1-antitrypsin.[93] However, in persons homozygous for the Pi^Z allele, serum concentrations of α_1-antitrypsin are only 10% to 15% of normal. In the United States, the prevalence of α_1-antitrypsin deficiency caused by a homozygous Pi^Z genotype is one in 3,000 persons[62]; it is less common in African Americans.[95] The substitution of a single amino acid in the normal M α_1-antitrypsin protein leads to the formation of polymers of the protein in the endoplasmic reticulum and impairs the release of the Z protein from hepatocytes into the circulation.[94] The protein appears as periodic acid–Schiff–positive and diastase-resistant granular cytoplasmic inclusions in the livers of patients heterozygous or homozygous for Pi^Z. The secreted Pi^Z α_1-anti-trypsin is also a less potent inhibitor of neutrophil elastase than Pi^M.[94]

Because of the deficient and less potent α_1-antitrypsin, endogenous proteases are not effectively inhibited, resulting in a breakdown of elastin and the development of emphysema.[93,94] This hypothesis is supported by studies that found evidence for uninhibited elastase activity in patients with α_1-antitrypsin deficiency.[96] Uninhibited elastase may stimulate macrophages to release IL-8 and leukotriene B_4, which is a chemoattractant for neutrophils that perpetuates the local inflammation and increased amounts of neutrophil elastase.[97]

Emphysema Caused by α_1-Antitrypsin Deficiency

Emphysema develops in at least 80% of patients with homozygous Pi^Z α_1-antitrypsin deficiency. The mean age at onset of dyspnea is 45 to 50 years in nonsmokers and approximately 10 years earlier in those who smoke. Remarkably, some nonsmoking patients with homozygous Pi^Z α_1-antitrypsin deficiency never manifest symptomatic emphysema.[98] There is no escape for smokers with this deficiency, however: all such patients have symptomatic airflow obstruction at a young age.

Diagnosis The typical pathologic picture is panacinar emphysema, but as many as 25% to 30% of nonsmoking patients and 60% of cigarette smokers report symptoms of chronic bronchitis as well. Many of the patients have evidence of enhanced airway reactivity.[99] High-resolution CT scanning has also detected a significant incidence of bronchiectasis.[100] Involvement of the lower lobes often predominates, perhaps because of increased neutrophil traffic and release of neutrophil elastase in the lower lung fields. The radiographic manifestation of this phenomenon is most commonly attenuation of the pulmonary vasculature to the lower lobes; in more advanced cases, basilar bullae may be seen [*see Figure 7*]. Features that would suggest α_1-antitrypsin deficiency as the cause of a particular patient's emphysema would thus include a family history of emphysema (especially among nonsmokers), the onset of symptoms at 30 to 50 years of age, the development of significant emphysema in a nonsmoker, and basilar predominance of the radiographic abnormalities.[98]

Serious liver disease, usually in the form of cirrhosis, occurs in 5% to 10% of adults with α_1-antitrypsin deficiency and may provide a clue to the underlying enzyme deficiency in some patients. Symptomatic liver disease is more common in children, who may present with neonatal hepatitis. Many infants and children recover, but 10% to 20% of those with hepatitis progress to cirrhosis and liver failure during childhood or early adulthood. It is not certain why some Pi^Z patients have symptomatic liver disease in childhood and others do not and acquire emphysema later in life. Other possible complications of α_1-antitrypsin deficiency include vasculitis, aneurysms, panniculitis, and pancreatitis.[93]

In persons who are heterozygous for the Z variant (Pi^{MZ}), serum α_1-antitrypsin concentrations are approximately 70% of normal.[86] Whether increased respiratory symptoms and airflow obstruction are greater in persons with the Pi^{MZ} phenotype than in matched Pi^M control subjects is not certain. Two long-term studies found greater deterioration in lung function in Pi^{MZ} patients than in Pi^{MM} patients.[101,102] The uncertainty might be explained by different degrees of deficiency in antiprotease levels. It is thought that α_1-antitrypsin levels higher than 40% of normal afford protection against the development of emphysema.

Treatment For patients with homozygous Pi^Z deficiency, smoking cessation and prevention of passive-cigarette-smoke exposure are essential. Consideration should be given to administration of purified α_1-antitrypsin, which is commercially available for replacement therapy.[93] Weekly infusions (60 mg/kg) can elevate serum levels above the hypothetical minimum protective level (50 mg/dl, 11 μM) for the entire week. Alternatively, a higher dose (250 mg/kg) can be given every 4 weeks. Clinical data support this longer interval, but unlike the weekly infusion, this regimen has not been approved for the product label. Studies of both dosing regimens using bronchoalveolar lavage have shown that adequate levels of α_1-antitrypsin are achieved

not only in the serum but also in the lung, where they are crucial. Preliminary data suggest reduction in the breakdown of elastin in patients on replacement therapy.[103]

There are numerous unresolved questions about antiprotease replacement therapy. The most important question is whether it will actually retard the progression of emphysema. Preliminary data from uncontrolled trials suggest a reduction in progression of disease with weekly replacement therapy.[104,105] For the present, individual treatment decisions have to be made without full information, with cost[106] and potential risk weighed against potential benefit. Replacement therapy probably is not indicated for patients who have normal pulmonary function; careful follow-up should be sufficient in such cases. Patients with mild to moderate airflow obstruction might benefit from replacement therapy. Smokers with CAO who have similar degrees of obstruction do benefit at this stage of the illness if the protease burden is reduced by the cessation of smoking. Patients with very severe airflow obstruction ($FEV_1 < 1$ L) may be beyond the point at which the natural course of the disease can be favorably altered by replacement therapy, much like smokers with very severe CAO, whose clinical course may not be altered by smoking cessation if there is little lung function left to preserve. In the future, it may be feasible to use antiprotease therapy[107] or gene therapy to restore normal production of α_1-antitrypsin.[108]

For homozygous Pi^Z patients with very severe airflow obstruction, lung volume reduction surgery[109] or single-lung transplantation offers hope for improved function and prolonged survival. Survival after transplantation is similar to that in patients with COPD.[110]

BULLOUS LUNG DISEASE

A bulla is defined pathologically as an emphysematous space greater than 1 cm in diameter; it is recognized radiographically as a localized hyperlucent area of lung demarcated by a curved hairline rim. Formation of bullae may complicate either panacinar or centriacinar emphysema. As many as one third of patients with radiographic evidence of emphysema have bullae on chest radiographs, although large, clinically significant bullae develop in only a small percentage of such patients.

Pathophysiology

Bullae may also form in the presence of normal surrounding lung tissue. The pathologic picture in this circumstance is most often described as paraseptal, or periacinar, emphysema. This condition involves destructive enlargement of predominantly the acinar ducts and the alveoli and primarily affects the periphery of the lobule adjacent to the surrounding septa. These subpleural bullae may produce a so-called soap-bubble appearance along the margin of the lung or may coalesce into a single bulla or multiple large bullae, typically in the lung apices. The perimeter of a bulla is formed by a thin layer of fibrous tissue and collapsed adjacent lung parenchyma. By contrast, the wall of congenital lung cysts (i.e., bronchogenic cysts) consists of bronchial epithelium.

For the most part, bullae behave as space-occupying lesions in the lung. Although bullae are inflated with gas, they participate minimally in overall ventilation. The virtual absence of ventilation of bullae in vivo can be demonstrated by (1) the absence of a change in the size of a bulla on an expiratory chest radiograph, (2) the large discrepancy between total lung capacity as measured by body plethysmography and as measured by helium dilution techniques, and (3) diminished or absent ventilation on radionuclide lung scanning with xenon-133. Perfusion scans reveal that bullae are also underperfused or nonperfused, and angiography demonstrates that large pulmonary vessels are not present within bullae. Thus, if a bulla contributes to hypoxemia in a patient, it is usually as a result of ventilation-perfusion imbalance (low $\dot{V}/\dot{Q}$) in the surrounding atelectatic lung tissue.

Diagnosis

Bullae that are surrounded by normal lung tissue are often asymptomatic. Until more than one third of the radiographic volume of a lung is occupied by a bulla, no clinically significant change in measured vital capacity occurs. Dyspnea may develop if the bulla is massively enlarged, or the patient may come to medical attention because of such complications as pneumothorax or infection. A bulla may be present with emphysema. Tests that may be useful for assessing the severity of associated emphysema in patients with bullae include CT of the chest, radionuclide lung scans, measurement of the $D_{L_{CO}}$, expiratory chest radiographs, and static pressure-volume curves of the lungs.

Treatment

Although a bulla may remain stable in size for many years, there is usually growth over time. Surgical resection of a large, symptomatic bulla may markedly increase vital capacity and P_aO_2 by allowing the surrounding normal lung to expand fully.[111]

A more difficult decision regarding the management of bullous lung disease is encountered in patients who have concomitant generalized emphysema throughout their lungs. In these patients, bullae simply represent a local exaggeration of a widespread process, and lung function does not normalize after resection of a bulla. Nevertheless, some patients with mild to moderate emphysema benefit from bullectomy. Improvement may be disappointingly brief, however, because bullae tend to recur.

However, there are no established guidelines for the management of bullous lung disease in patients with emphysema, and the decision to operate on a giant bulla is ultimately based on various subjective factors.

BRONCHIECTASIS

Bronchiectasis is a chronic suppurative disease of the airways that if sufficiently widespread may cause CAO.[112] In the preantibiotic era, it was a relatively frequent sequela of pulmonary infections in childhood, typically leading to chronic respiratory disability and death by 40 years of age. In modern practice, bronchiectasis is far less common and carries a much less dire prognosis. Nevertheless, it continues to be an important cause of chronic productive cough with sputum purulence, and it accounts for a significant percentage of cases of massive hemoptysis.

Etiology

Bronchiectasis is a localized, irreversible bronchial dilatation caused by a destructive inflammatory process involving the bronchial walls. Necrotizing bacterial or mycobacterial infection is thought to be responsible for most cases of bronchiectasis. A typical history is that of a childhood respiratory infection, such as whooping cough or bacterial superinfection complicating a viral pneumonia, followed by recurrent or persistent so-called chest colds.

Adult-onset bronchiectasis may result from an untreated or

inadequately treated bronchopneumonia that is caused by virulent organisms such as staphylococci or gram-negative bacilli. Mycobacterial infection frequently causes bronchiectasis, but because reactivation tuberculosis usually involves the upper lobes of the lungs, the clinical consequences differ considerably from the clinical consequences of other forms of bronchiectasis, which usually involve the right middle lobe, the lingula, or the lower lobes.

Generalized bronchiectasis may develop when systemic or pulmonary defense mechanisms are impaired in such a way as to predispose the patient to recurrent or chronic bacterial infections involving the airways. Rare congenital abnormalities of lung structure such as bronchial cartilage deficiency (Williams-Campbell syndrome) and tracheobronchomegaly (Mounier-Kuhn syndrome) may lead to generalized bronchiectasis, as may inherited deficiencies of immunoglobulins, impaired phagocytosis, complement deficiency, and α_1-antitrypsin deficiency.[113] In adults, widespread bronchiectasis occurs in association with hypogammaglobulinemia,[114] including isolated IgG subclass deficiencies[115]; cystic fibrosis [*see* Clinical Variants of Bronchiectasis, Cystic Fibrosis, *below*]; primary ciliary dyskinesia [*see* Clinical Variants of Bronchiectasis, Primary Ciliary Dyskinesia, *below*]; and several systemic diseases,[116] including rheumatoid arthritis,[117] other connective tissue diseases, inflammatory bowel disease, sarcoidosis, yellow nail syndrome, and HIV infection.[118]

Pathogenesis

Bronchiectasis results from the destruction of airways. The continued progression of disease over time likely results from continued infection, often with destructive organisms such as *P. aeruginosa,*[119] and the continued presence of tissue inflammation[120] with release of cytokines that are chemotactic for neutrophils.[121] These neutrophils locally release proteolytic enzymes such as elastase and other neutrophil enzymes[122] that may be involved in further airway and parenchymal damage and also may have an enhancing effect on mucus hypersecretion.[123]

The most important consequence of bronchiectatic damage to a portion of the airways is an increased susceptibility to recurrent or persistent bacterial infections. Normal defense mechanisms against infection are breached: ciliary function is disrupted by squamous metaplasia or by ulceration of the epithelial lining; excess mucus is secreted and pools in dilated spaces; and cough becomes less effective in clearing mucus from dilated bronchial segments. Once bacterial superinfection is established, it is virtually impossible to eradicate, and daily expectoration of purulent sputum is the end result in advanced cases. In bronchiectasis involving the upper lobes, stasis of mucus is prevented by gravitational drainage, resulting in the dry bronchiectasis of tuberculosis, in which cough and sputum production are usually absent but hemoptysis may occur.

Diagnosis

Three types of bronchiectasis have been described on the basis of bronchographic-pathologic findings: cylindrical, varicose, and cystic. Distinction between these three types of bronchiectasis is not useful clinically, however, because the manifestations and course of bronchiectasis are not correlated with the bronchographic pattern.

Clinical manifestations In most cases, the clinical presentation and a plain chest radiograph suffice for a presumptive diagnosis of bronchiectasis. Factors in the history, such as chronic cough and sputum purulence originating from a serious respiratory tract infection, often in childhood, strongly suggest the diagnosis. In addition, chronic sinusitis frequently accompanies bronchiectasis, and its presence should raise the suspicion of concomitant chronic lower respiratory tract infection.

In other patients, the clinical picture is one of frequent lower respiratory tract infections limited to the same area or areas of the lungs. Symptoms, physical findings, and abnormalities on the chest radiograph may not clear completely between each episode of pneumonia. Some patients have a mucoid sputum that becomes intermittently infected; disease course mimics that of chronic bronchitis with episodic infectious exacerbations. In this case, the local findings on chest examination and on the chest radiograph usually point to the diagnosis of bronchiectasis.

Clubbing of the digits occurs in the majority of patients with significant bronchiectasis and is a valuable diagnostic clue, especially because clubbing of the digits is not a manifestation of COPD [*see* Chronic Bronchitis and Emphysema, Diagnosis, *above*]. Auscultation of the chest usually reveals localized findings. Typically, paninspiratory, coarse crackles are heard over the involved region, and there may also be variable low-pitched wheezes if secretions are present in the airways.

Chest imaging Bronchiectasis may be seen on the plain chest radiograph in a number of different patterns. Cystic bronchiectasis is most readily recognized because of the distinctive appearance of a collection of thin-walled cystic spaces, sometimes accompanied by air-fluid levels, arranged in a segmental distribution. A localized increase in interstitial markings that follow the general orientation of the bronchovascular bundles may indicate milder degrees of ectasia. Occasionally, the thickened walls of a dilated bronchus can be visualized as the bronchus courses with its longitudinal axis perpendicular to the x-ray beam. These parallel lines are approximately 1 mm thick and are referred to as tramlines. Atelectasis often accompanies extensive bronchiectasis, in which case the radiographic appearance may mimic a postobstructive pneumonia. In approximately 7% of patients with bronchiectasis, the plain chest radiograph is entirely normal.

The current generation of CT machines provide excellent magnified images of bronchiectatic airways. CT scanning can be used to confirm a clinical suspicion of bronchiectasis, to suggest the specific cause,[124] and to evaluate the extent of disease[125] [*see Figure 8*].

Sputum examination Examination of purulent sputum produced by a patient with bronchiectasis may suggest the underlying diagnosis in two ways. First, if sputum is collected in a container for several hours until a sufficient volume is obtained, the sputum may settle into a characteristic three-layered pattern: foamy on top, purulent in the middle, and liquid at the bottom. Occasionally, the same pattern is observed in sputum from patients with chronic bronchitis or suppurative lung abscess. Second, routine bacterial culture of the sputum may grow *Pseudomonas* species. In an immunocompetent host, *Pseudomonas* species are almost never isolated from the sputum unless the host has bronchiectasis, is receiving broad-spectrum antibiotics, has a long-term tracheostomy, or is in the intensive care unit. *Staphylococcus aureus,* gram-negative bacilli other than *Pseudomonas* (especially *H. influenzae*), and *Mycobacterium avium* complex[126] may also infect the airways of patients with bronchiectasis. Bronchiec-

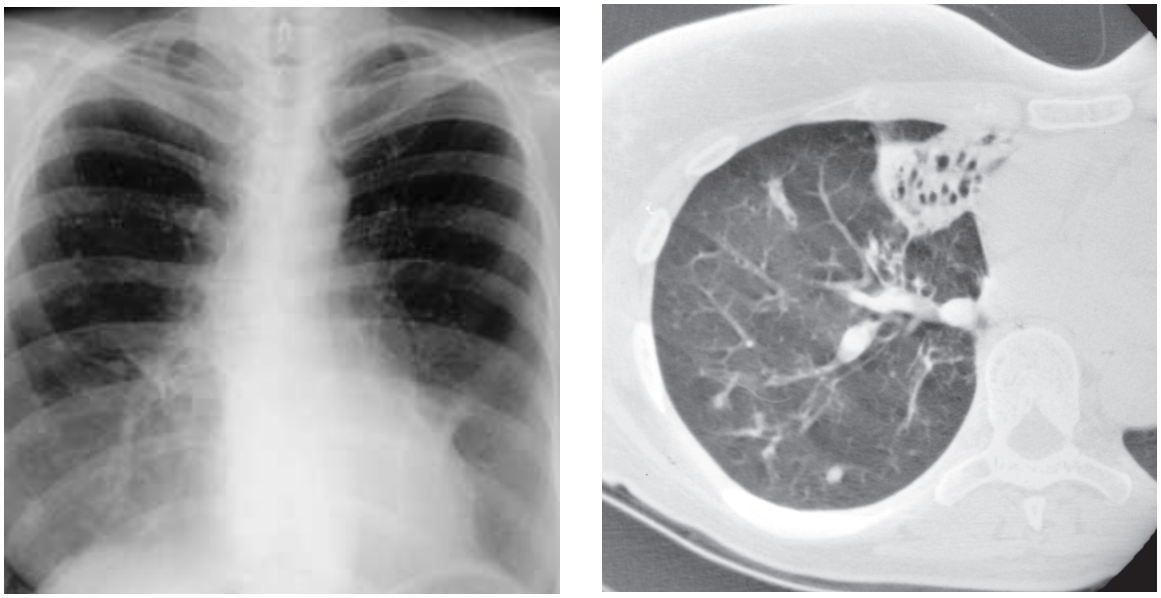


Figure 8 **In a comparison of a frontal chest radiograph with a CT scan of a patient with hemoptysis, the radiograph (left) suggests only an infiltrate in the right middle lobe; the CT scan (right) shows bronchiectasis in the right middle lobe that was the source of the bleeding.**

tasis is also a feature of allergic bronchopulmonary aspergillosis [*see 14:II Asthma*].

Pulmonary function tests Pulmonary function tests may remain normal if only a small portion of the tracheobronchial tree is affected. Widespread bronchiectasis causes chronic obstruction of expiratory airflow and may also cause a restrictive deficit if there is sufficient associated atelectasis or involvement of lung parenchyma by the infectious process. However, airflow obstruction is generally the main abnormality.

Treatment

The mainstays of therapy for bronchiectasis (including cystic fibrosis and primary ciliary dyskinesia), as for any chronic suppurative disease, are administration of antibiotics and drainage.

Antibiotics The use of antibiotics in the treatment of bronchiectasis has not been subjected to careful scientific investigation, and no one method of administration has proved to be superior in clinical experience. It is reasonable to culture the sputum periodically, because in patients from whom *S. aureus* or *H. influenzae* has repeatedly been isolated, antibiotics with appropriate spectra of activity can be selected. Oral antibiotics with effective antipseudomonal activity, such as the quinolones, show promise; however, because of the potential emergence of resistant strains of *Pseudomonas,* these drugs probably should not be used as single agents for long-term suppressive therapy. Nevertheless, many patients seem to benefit from broad-spectrum oral antibiotics, even when *Pseudomonas* is the only pathogen in the sputum and in vitro sensitivity testing shows that the antibiotics lack activity against *Pseudomonas*. In a patient who has daily purulent sputum production and is not allergic to sulfonamides, trimethoprim-sulfamethoxazole (one double-strength tablet twice daily) can be given continuously. Alternatively, antibiotics may be given intermittently or on a schedule in which different antibiotics are rotated.[127] Nebulized antibiotics may also be effective.[128]

If oral antibiotic therapy has failed, serious infectious complications (such as persistent fever with new areas of infiltration detected on the chest radiograph or the development of pleuritic chest pain) are generally best treated with a 10- to 14-day course of intravenous antibiotics. Two synergistic antibiotics with appropriate in vitro activity should be used for *Pseudomonas* infection.

Drainage Drainage of secretions is partially achieved by coughing and expectoration of sputum. However, because the diseased airways collapse during coughing and forced exhalation and because there is pooling of secretions distal to the areas of collapse, it is useful to include postural drainage as part of the management of bronchiectasis. Chest physiotherapy (e.g., the use of chest percussion and vibration) is often used to aid bronchopulmonary drainage, although it is difficult to demonstrate that physiotherapy produces benefits beyond those that are produced by postural drainage alone. There is no role for aerosolized recombinant human deoxyribonuclease (DNase) in bronchiectasis not associated with cystic fibrosis.[129]

Bronchodilator therapy Many patients with bronchiectasis will have significant airflow obstruction, manifested clinically by wheezing and detected by pulmonary function testing. Theophylline, beta$_2$ agonists, and anticholinergic bronchodilators can be used in this setting, although the evidence for their effectiveness is limited.[130-133] Bronchodilator therapy may promote the clearance of airway secretions if the bronchodilating agents are administered before each postural drainage session.

Anti-inflammatory therapy During episodes of exacerbation, oral corticosteroids may help improve the patient's condition, although the benefit of this approach has not been established in randomized trials.[134] High-dose inhaled corticosteroids have been shown to reduce markers of airway inflammation[135] and improve lung function,[136] but the potential role of long-term therapy is uncertain. Leukotriene modifiers have no role in the

treatment of bronchiectasis, except in the setting of allergic bronchopulmonary aspergillosis.[137]

Therapy for hemoptysis Significant hemoptysis can also usually be controlled with appropriate antibiotic therapy. Massive hemoptysis (> 200 ml of blood over a 24-hour period) that occurs as a complication of bronchiectasis was traditionally managed with surgical resection of the involved lung. Now, however, massive hemoptysis caused by bronchiectasis is often effectively treated with bronchial arterial embolization, an invasive radiologic procedure involving catheterization of the bronchial arteries. The dilated bronchial arteries that perfuse the airways in bronchiectasis are particularly suited for the application of this technique. The procedure requires a skilled angiographer.

Surgery In the modern antibiotic era, the role of surgery in the management of bronchiectasis has been declining.[138] In patients with widespread bilateral disease, diseased lung tissue is better able to support gas exchange than no lung tissue at all. In patients with only limited localized disease, symptoms can usually be well controlled with the measures described above. In addition, experience indicates that after lobar resections for localized bronchiectasis have been performed, clinically evident recurrences of the disease are common in parts of the lung that had previously been thought to be uninvolved. In the rare instance in which severe symptoms or recurrent complications in a young patient lead to consideration of resection, the localized nature of the bronchiectasis must first be demonstrated radiographically. If there is bronchiectasis distal to an obstructing bronchial lesion, surgery is indicated to remove the obstruction along with the diseased lung tissue.

CLINICAL VARIANTS OF BRONCHIECTASIS

Cystic Fibrosis

Although cystic fibrosis is an inherited disease that usually manifests itself in early childhood, a discussion of the condition in the context of general adult medicine is worthwhile for two reasons. First, increasing numbers of children with cystic fibrosis are now surviving into young adulthood: the median survival in the United States is 32 years.[139] Second, some patients have a variant form of the disease in which symptoms first appear during adolescence or adulthood.

Pathogenesis The genetic defects responsible for cystic fibrosis have been identified.[139] The cystic fibrosis locus is on the long arm of chromosome 7; it codes a polypeptide comprising 1,480 amino acids that has been named the cystic fibrosis transmembrane regulator (CFTR).[139] In 70% of patients with cystic fibrosis, the 508th amino acid of this sequence is missing (ΔF508 mutation). The abnormal protein derived from the altered sequence is not glycosylated; it is retained in the endoplasmic reticulum and is not transferred to the cell membrane. The result is a defective membrane with decreased apical chloride conductance and increased sodium absorption.[140] Excessive dehydration of respiratory secretions may alter the character of the sol phase, in which the cilia normally beat, making it thicker and more viscous. Patients who are homozygous for the ΔF508 mutation have a more severe form of the disease than those who are heterozygous.[141] A number of other defects of the cystic fibrosis gene have also been identified, and the resultant defects in the production of CFTR can be grouped into five classes [*see Figure 9*]. Because these defects cause the CFTR to function differently, phenotypic severity differs among the classes.[142] Several of these defects are associated with mild lung disease and even normal sweat chloride concentrations.[143] As a result, patients may present at a later age, and diagnosis may be difficult. In addition, polymorphisms of other genes involved in the immune response may alter the phenotypic severity.[144]

It is likely that impaired tracheobronchial clearance of the abnormal secretions leads to widespread mucous plugging of airways, resulting in secondary bacterial infection, persistent inflammation, and consequent generalized bronchiectasis.[145] The bacterial flora in the airways are highly stereotyped: early in the course of the disease, *S. aureus* is found in the sputum; subsequently, mucoid strains of *P. aeruginosa* are isolated (mucoid in this context refers to a slimy substance secreted by the colony of organisms growing on a culture plate). Despite the presence of these highly virulent pathogens in the lower respiratory tract, infection remains confined to the airways. Although lung abscess and empyema are common complications of *Staphylococcus* or *Pseudomonas* pneumonia, they very rarely develop in patients with cystic fibrosis.

Diagnosis CAO is present in virtually all adult patients with cystic fibrosis and follows a relentlessly progressive course. Thus, cough, chronic purulent sputum production, and exertional dyspnea are cardinal symptoms of cystic fibrosis. Increased airway reactivity is found in approximately 20% to 25% of patients with this disease; in this subgroup, episodic wheezing may be a prominent manifestation, leading to a misdiagnosis of asthma.[146] Nasal polyposis and chronic sinusitis are common upper respiratory tract findings in patients with cystic fibrosis and may be mistaken for signs of allergic disease. Even findings consistent with allergic bronchopulmonary aspergillosis are observed in as many as 9% of patients.[145] A small number of patients become colonized, although they are rarely infected, with atypical mycobacteria.[145,147] Two important complications of lower respiratory tract disease in patients with cystic fibrosis are hemoptysis and pneumothorax. Minor hemoptysis occurs intermittently in a majority of patients. In approximately 7% of adult patients, rupture of dilated bronchial arteries leads to massive, potentially fatal hemoptysis. Pneumothoraces may complicate the course of advanced obstructive lung disease in approximately one sixth of adult patients and are frequently recurrent.[148] Hypoxemia and hypercapnia, occurring initially during exercise and sleep,[149] are prominent complications and can be associated with the development of pulmonary hypertension.[150] Osteoporosis producing significant kyphosis is frequent in adult patients with cystic fibrosis.[151]

The chest radiograph may strongly suggest the diagnosis of cystic fibrosis. The generalized bronchiectasis manifests itself as a diffuse increase in interstitial markings, and discrete bronchiectatic cysts are often visible; typically, involvement of the upper lobes predominates. The obstructive aspect of the disease is reflected in the typical finding of pulmonary hyperinflation. This combination of diffusely increased markings with cystic spaces, upper lobe predominance, and hyperinflation is highly characteristic of cystic fibrosis; rare alternative radiographic diagnoses include eosinophilic granuloma and lymphangiomyomatosis [*see 14:V Chronic Diffuse Infiltrative Lung Diseases*]. In the late stages of the disease, cardiomegaly and signs of pulmonary arterial hypertension appear on the chest radiograph as cor pulmonale develops.

Extrapulmonary manifestations may also suggest the diagnosis of cystic fibrosis. Prominent among these findings are pancreatic insufficiency with consequent steatorrhea, recurrent partial intestinal obstruction caused by abnormal fecal accumulation (so-called meconium ileus equivalent), heat prostration, hepatic cirrhosis, and aspermia in males.

The diagnosis of cystic fibrosis should therefore be suspected in the adolescent or young adult who has widespread bronchiectasis and *Staphylococcus* or *Pseudomonas* infection of the airways. The diagnosis should also be considered in the young patient with so-called refractory asthma, especially if the asthma symptoms are accompanied by clubbing of the digits, chronic sputum purulence, a persistently abnormal chest radiograph, or symptoms of pancreatic insufficiency. The diagnosis can be established by a finding of abnormal results on a sweat test performed in a qualified laboratory using pilocarpine iontophoresis.[152] In persons younger than 20 years, a sweat chloride level exceeding 60 mEq/L confirms the diagnosis; a value exceeding 80 mEq/L is required for diagnosis in persons 20 years of age or older. The diagnosis can be confirmed by genotyping for the most common CFTR mutations. Genotyping should also be used when the sweat chloride tests results are equivocal or normal and cystic fibrosis is strongly suspected.

With the identification of the gene for cystic fibrosis, genetic screening has become available. A National Institutes of Health panel suggested that genetic testing for cystic fibrosis be offered to adults with a positive family history of the disease, to partners of people with the disease, to couples currently planning a pregnancy, and to couples seeking prenatal care, but not to the general population or newborns.[153]

Treatment Treatment of the pulmonary aspects of cystic fibrosis is similar to that of bronchiectasis [*see* Bronchiectasis, Treatment, *above*] and includes management of infections and the use of respiratory therapy modalities that are designed to mobilize secretions, including regular percussion and postural drainage, and to reduce airway obstruction.[139,145,154] A pneumatic bronchial drainage vest or a flutter device makes it easier to vibrate the chest so as to enhance the removal of thick secretions.[155,156] No randomized, controlled trials have established the efficacy of any airway clearance regimen in cystic fibrosis, however.[157] Aerosolized antibiotics such as tobramycin may have a role in reducing the burden of infection in those who have become chronically infected with *P. aeruginosa*.[145,158] Treatment with intravenous antibiotics is usually required for episodes of symptomatic infection with *P. aeruginosa*[145,159] or other organisms. Chronic treatment with macrolide antibiotics may have beneficial effects related to mechanisms other than antibacterial activity—namely, reduction in biofilm formation and anti-inflammatory properties.[160] Because the viscosity of the mucus in cystic fibrosis is partially caused by DNA released from cells, recombinant human DNase administered by inhalation is effective.[161] In patients with reversible airflow obstruction, treatment with bronchodilators (e.g., beta agonists), anticholinergics, and theophylline and with low-dose alternate-day oral or daily inhaled corticosteroids may be of benefit.[162-164] Strategies to reduce airway inflammation and to reduce the burden of neutrophil proteases are also being evaluated.[165,166] Attention to nutrition, physical conditioning, and emotional health must be part of an effective care plan. Bronchial artery embolization is useful in patients with significant hemoptysis.[167] Mechanical ventilatory support at night using noninvasive techniques may be useful in patients with chronic respiratory failure.[168] Lung transplantation is now being performed with good results in cystic fibrosis patients whose FEV_1 is less than 30% of predicted value.[169,170]

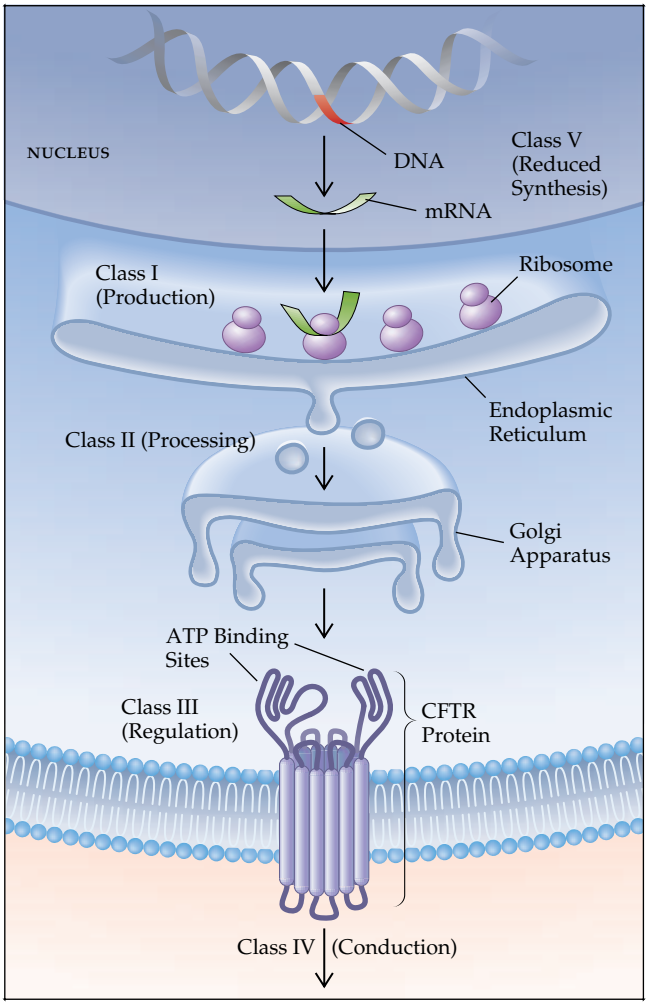


Figure 9 **Categories of cystic fibrosis transmembrane conductance regulator (CFTR) mutations. Cystic fibrosis can be produced by abnormalities at several points in the pathway from gene to functional protein on the cell surface. Class I mutants are associated with decreased transcription of the DNA or translation of the RNA. Class II mutants are attributable to abnormalities in processing of the protein in the endoplasmic reticulum, resulting in the degradation of the protein. Class III mutants are associated with abnormal regulation of the protein. Class IV mutants are associated with abnormal function of the CFTR protein on the cell surface. Class V mutants are associated with decreased synthesis of the CFTR protein. Some mutants (e.g., ΔF508, class II and class III) can be associated with more than one defect.**

The discovery of a specific genetic defect raises the possibility of more specific and perhaps more effective therapy. Therapy could be either pharmacologic (aimed at altering transport through involved or uninvolved ion channels) or genetic (aimed at replenishing the CFTR). In pilot studies, inhalation of amiloride, an epithelial sodium channel blocker, led to objective improvement in sputum character. Other sodium channel blockers that are more potent and longer-acting may be available in the future. Triphosphate nucleotides (adenosine triphosphate and uridine triphosphate) have been found to be effective chloride secretagogues in vivo but have not been tested in long-

term therapy.[171] Another approach is to rescue the mutant CFTR protein from the endoplasmic reticulum and improve the function of the abnormal channel when it is delivered to the plasma membrane.[172] Strategies aimed at correcting the genetic defect are advancing, albeit slowly.[108] An altered adenovirus (incapable of replication) has been used to introduce the cystic fibrosis gene into patients, and phase II trials of this therapy are under way.[108] Several other approaches are being taken and may be successful in the near future.[173]

Primary Ciliary Dyskinesia

In 1933, Dr. Manes Kartagener identified a group of patients with bronchiectasis who also suffered from chronic sinusitis and situs inversus. This triad of findings came to be known as Kartagener syndrome. Approximately 40 years later, it was recognized that male infertility was associated with this syndrome: men with Kartagener syndrome were found to have live sperm with absent or ineffective motility. Sperm tails and the cilia of respiratory tract epithelial cells share an ultrastructure, and in 1975, it was recognized that an inherited abnormality in that ultrastructure (i.e., an absence of the adenosine triphosphatase [ATPase]–containing dynein arms of the outer microtubular doublets) led to nonfunctioning respiratory tract cilia and immotile spermatozoa[174] [*see Figure 10*].

The consequences of congenital nonfunctioning cilia of the upper and lower respiratory tracts are chronic sinusitis, secretory otitis media, and daily productive cough dating from birth or early childhood; bronchiectasis develops during childhood in the majority of patients. With respect to situs inversus, it is speculated that the normal asymmetrical positioning of body organs is dependent on normal functioning of cilia on embryonic epithelium. In the absence of normal ciliary function, placement of organs to either the left or the right is random, and as expected, about one half of patients with congenitally nonfunctioning cilia manifest situs inversus. Thus, the term immotile cilia syndrome was coined to include all patients with chronic sinusitis and bronchiectasis resulting from ultrastructural abnormalities of cilia. Fertility is reduced not only in men but also in women with this syndrome, because of deficient cilia in the oviducts and fimbriae.

Primary ciliary dyskinesia highlights the importance of normal ciliary function in clearing airway mucus, as well as the importance of other mechanisms, especially cough, in defending against disease. Ciliary dysfunction in the lower respiratory tract leads to the retention of secretions, bacterial superinfection, and bronchiectasis. Other protective mechanisms appear to prevent more serious sequelae, including acute pneumonias and progressive airflow obstruction.

Pathogenesis A variety of abnormalities in addition to absent or deficient dynein arms may impair the structure and function of cilia and sperm tails.[175] Many of these abnormalities involve derangement of the normal configuration of microtubules. In each case, in vitro microscopic studies of ciliary motility have demonstrated a pattern of decreased, uncoordinated, or ineffective beating. Because at least some movement of the cilia is observed, the term primary ciliary dyskinesia has been proposed as a more accurate description of this condition than the term immotile cilia syndrome. Inheritance is thought to be autosomal recessive. Any given ultrastructural abnormality is found consistently in the cilia of the upper and lower respiratory tract (as well as in sperm tails in men), and all affected members of the same family have the same defect. The clinical syndrome that results is a common expression of impaired ciliary motility independent of the specific ultrastructural defect.

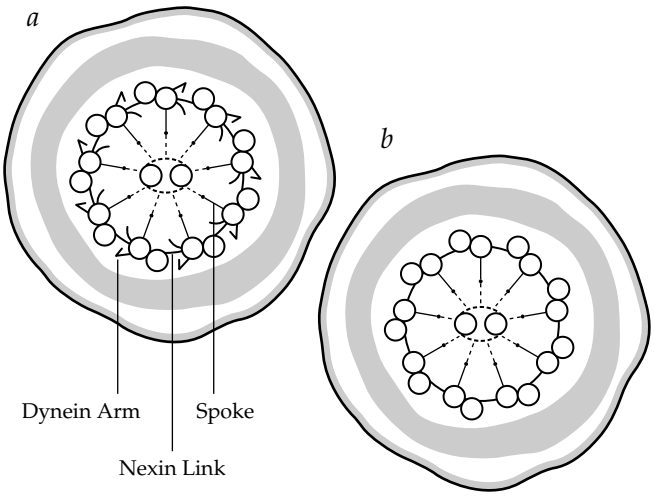


Figure 10 **A cross section (*a*) of the tail of a normal sperm reveals dynein arms. In a cross section of the tail of a sperm from a patient with Kartagener syndrome (*b*), dynein arms are missing; such a cell would be immotile. Similar changes have been noted in the cilia in the respiratory tract and paranasal sinuses in these patients.**

Diagnosis The diagnosis of primary ciliary dyskinesia can be made clinically in all patients with Kartagener syndrome.[176] A clinical diagnosis can also be made on the basis of the following criteria: a history of chronic sinusitis; a productive cough since childhood; and either the presence of live but immotile spermatozoa or a family history in which the patient has a sibling with Kartagener syndrome. In patients who have had chronic sinusitis and productive cough since childhood but who meet neither of the two additional criteria, three types of specialized laboratory studies can be employed to support the diagnosis of primary ciliary dyskinesia: (1) electron microscopic examination of sperm tails or of cilia from bronchial or nasal biopsy specimens, (2) in vitro light microscopic examination of the motility of cilia, and (3) measurement of mucociliary clearance in the nose or the tracheobronchial tree. An example of the last technique listed is inhalation of an aerosol of radiolabeled particles followed by external scanning over the thorax for at least 2 hours. Normal values for the ciliary structural and functional studies have been established.[177] These studies should be performed during periods of clinical stability, because acute inflammation can reversibly alter ciliary function.[178]

Differential diagnosis The respiratory disease associated with primary ciliary dyskinesia may be contrasted with that associated with cystic fibrosis. In cystic fibrosis, in the absence of acute infectious exacerbations, ciliary function is normal and mucociliary transport only modestly decreased. In primary ciliary dyskinesia, however, mucociliary transport along nasal or tracheobronchial mucosa is virtually absent. Nevertheless, disease of the lower respiratory tract is usually far milder in patients with primary ciliary dyskinesia than in those with cystic fibrosis. Bronchiectasis in primary ciliary dyskinesia usually involves the lower or the middle lung zones and is less widespread than in cystic fibrosis. Also, in primary ciliary dyskinesia,

bacterial infection of the airways is more commonly caused by *Haemophilus, Neisseria,* or *Streptococcus* organisms than by *Staphylococcus* or *Pseudomonas* organisms, and acute pneumonias are relatively infrequent. Finally, airflow obstruction is usually mild in primary ciliary dyskinesia, and progression to cor pulmonale is uncommon. Patients with primary ciliary dyskinesia usually can remain fully active and may survive to old age.

Treatment Treatment with postural drainage and antibiotics for infections aids in maintaining stable lung function [*see* Bronchiectasis, Treatment, *above*].[179]

BRONCHIOLITIS

Disorders of the bronchioles can be divided into three categories: (1) primary bronchiolar disorders, which are covered in this section; (2) parenchymal disorders with prominent bronchiolar involvement, and (3) bronchiolar involvement in large airway diseases.[180] Parenchymal disorders with prominent bronchiolar involvement such as hypersensitivy pneumonitis, respiratory bronchiolitis/interstitial lung disease, and bronchiolitis obliterans organizing pneumonia are covered in another chapter [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. The larger airway diseases associated with bronchiolar involvement are COPD, bronchiectasis [*see* Bronchiectasis, *above*], and asthma [*see 14:II Asthma*].

The primary bronchiolar disorders can be further divided into five types: acute bronchiolitis, constrictive bronchiolitis (obliterative bronchiolitis or bronchiolitis obliterans), diffuse panbronchiolitis, mineral dust airway disease, and follicular bronchiolitis.[180]

Etiology

Acute bronchiolitis is mostly a disease of childhood.[180] One form of the disease, an acute bronchiolitis that occurs in infants, is often the result of infection with respiratory syncytial virus. Adenovirus infection may cause a more serious necrotizing form of bronchiolitis in children; in some of these children, the healing process is characterized by exuberant inflammation and fibrosis that obliterate the bronchiolar lumen. If only one lung is affected, the obstructive disease may appear on radiography as a unilateral hyperlucency of the lung field because of distal alveolar overdistention (i.e., air trapping) and decreased vascularity in the affected lung. This syndrome of a unilateral hyperlucent lung in a patient with bronchiolitis bears two eponyms: Swyer-James syndrome, named for the two physicians who first described the disease in children, and Macleod syndrome, named after the physician who reported the first adult case.

Constrictive bronchiolitis is a rare cause of CAO in adults, occurring after viral pneumonia (e.g., pneumonia caused by measles, influenza, or adenovirus infection), after inhalation of toxic gases (e.g., chlorine or nitrogen dioxide), and as an idiopathic phenomenon.[180] Constrictive bronchiolitis has been documented as a complication of collagen vascular diseases, particularly rheumatoid arthritis, and as a sequela of stem cell transplantation in the setting of graft versus host disease. A possible association with the drug penicillamine has also been suggested.[180]

Diffuse panbronchiolitis has been seen primarily in Asia, particularly in Japanese adults. The cause is unknown, although a genetic predisposition is suspected.[180]

Mineral dust airway disease is associated with exposure to a number of inorganic dusts, including asbestos, iron and aluminum oxide, talc, mica, silica, silicate, and coal.[180]

Follicular bronchiolitis can be idiopathic; it can also be associated with collagen vascular diseases (particularly rheumatoid arthritis) or immunodeficiency syndromes such as HIV infection.[180]

Diagnosis

Clinical manifestations In patients with an acute infection or who have had a toxic exposure, fever, nonproductive cough, and dyspnea may develop 2 to 4 weeks after the initial event, often after an asymptomatic interval. In other patients, the gradual onset of dyspnea and dry cough are the presenting symptoms. Chest examination typically reveals high-pitched inspiratory crackles, wheezing, and a highly characteristic midinspiratory squeak.

Pulmonary function testing shows marked airflow obstruction and little or no reversibility of obstruction in response to bronchodilators.

Chest radiograph The characteristic chest radiographic finding is a pattern of diffuse, nodular densities, sometimes with a fine nodularity mimicking miliary tuberculosis. Although these findings are typical of bronchiolitis, a spectrum of presentations is possible. Radiographically, hyperinflation and vascular attenuation mimicking emphysema may be the only findings, or there may be scattered nonhomogeneous patchy infiltrates. High-resolution CT scanning of the chest, particularly if performed on inspiration and expiration, may show nodular lesions, regions of ground-glass attenuation, bronchocentric infiltrates, and small regions of lucency, denoting obstruction and air trapping at a small airway level.[181]

Treatment

Treatment of constrictive bronchiolitis is usually ineffective, although corticosteroids are often tried. Initially, high-dose corticosteroids (e.g., prednisone given in an oral dosage of 1 mg/kg/day) are used in an attempt to suppress the inflammatory reaction within and around the bronchioles. Although corticosteroids often benefit patients with idiopathic bronchiolitis obliterans organizing pneumonia, they are rarely effective against other forms of bronchiolitis.

Prognosis

In some patients with bronchiolitis, the disease progresses to severe airflow obstruction, respiratory failure, and death. In others, the inflammation remits, and chest x-ray and pulmonary function test results return to normal.

Gerald W. Staton, Jr., M.D., F.A.C.P., has received grants for educational activities from GlaxoSmithKline and Boeringer Ingleheim and serves as an advisor to GlaxoSmithKline.

Inhaled steroids have not been approved by the FDA for use in COPD.

References

1. Snider GL: Nosology for our day: its application to chronic obstructive pulmonary disease. Am J Respir Crit Care Med 167:678, 2003

2. Ulrik CS, Backer V: Nonreversible airflow obstruction in life-long nonsmokers with moderate to severe asthma. Eur Respir J 14:896, 1999

3. Ryu JH, Scanlon PD: Obstructive lung diseases: COPD, asthma, and many imitators. Mayo Clin Proc 76:1144, 2001

4. Mannino DM, Homa DM, Akinbami LJ, et al: Chronic obstructive pulmonary disease surveillance—United States, 1971–2000. MMWR Morb Mortal Wkly Rep 51:1, 2002

5. Sethi JM, Rochester CL: Smoking and chronic obstructive pulmonary disease. Clin Chest Med 21:67, 2000

6. Balmes J, Becklake M, Blanc P, et al: American Thoracic Society Statement: occupational contribution to the burden of airway disease. Am J Respir Crit Care Med 167:787, 2003

7. Diaz PT, King MA, Pacht ER, et al: Increased susceptibility to pulmonary emphysema among HIV-seropositive smokers. Ann Intern Med 132:369, 2000

8. Kanazawa H, Hirata K, Yoshikawa J: Accelerated decline of lung function in COPD patients with chronic hepatitis C virus infection: a preliminary study based on small numbers of patients. Chest 123:596, 2003

9. Harik-Khan RI, Fleg JL, Wise RA: Body mass index and the risk of COPD. Chest 121:370, 2002

10. Hu G, Cassano PA: Antioxidant nutrients and pulmonary function: the Third National Health and Nutrition Examination Survey (NHANES III). Am J Epidemiol 151:975, 2000

11. Sunyer J: Urban air pollution and chronic obstructive pulmonary disease: a review. Eur Respir J 17:1024, 2001

12. Postma DS, Rijcken B: The role of atopy and hyperresponsiveness in the development of COPD. Eur Respir Rev 7:159, 1997

13. Silverman EK, Chapman HA, Drazen JM, et al: Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease: risk to relatives for airflow obstruction and chronic bronchitis (pt 1). Am J Respir Crit Care Med 157:1770, 1998

14. Sandford AJ, Joos L, Pare PD: Genetic risk factors for chronic obstructive pulmonary disease. Curr Opin Pulm Med 8:87, 2002

15. Sethi S: Bacterial infection and the pathogenesis of COPD. Chest 117(suppl 1):286S, 2000

16. Donaldson GC, Seemungal TA, Bhowmik A, et al: Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 57:847, 2002

17. Wilkinson TM, Patel IS, Wilks M, et al: Airway bacterial load and FEV_1 decline in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 167:1090, 2003

18. Scanlon PD, Connett JE, Waller LA, et al: Smoking cessation and lung function in mild-to-moderate chronic obstructive pulmonary disease (pt 1). The Lung Health Study. Am J Respir Crit Care Med 161:381, 2000

19. Kanner RE, Connett JE, Williams DE, et al: Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the Lung Health Study. Am J Med 106:410, 1999

20. Nagai A: Pathology and pathophysiology of chronic obstructive pulmonary disease. Intern Med 41:265, 2002

21. Jeffery PK: Remodeling in asthma and chronic obstructive lung disease. Am J Respir Crit Care Med 164:S28, 2001

22. Saetta M, Turato G, Baraldo S, et al: Goblet cell hyperplasia and epithelial inflammation in peripheral airways of smokers with both symptoms of chronic bronchitis and chronic airflow limitation (pt 1). Am J Respir Crit Care Med 161:1016, 2000

23. Di Stefano A, Capelli A, Lusuardi M, et al: Severity of airflow limitation is associated with severity of airway inflammation in smokers. Am J Respir Crit Care Med 158:1277, 1998

24. Hogg JC, Senior RM: Chronic obstructive pulmonary disease, part 2: pathology and biochemistry of emphysema. Thorax 57:830, 2002

25. Calverley PM, Walker P: Chronic obstructive pulmonary disease. Lancet 362:1053, 2003

26. Barnes PJ, Shapiro SD, Pauwels RA: Chronic obstructive pulmonary disease: molecular and cellular mechanisms. Eur Respir J 22:672, 2003

27. Voelkel NF, Cool CD: Pulmonary vascular involvement in chronic obstructive pulmonary disease. Eur Respir J Suppl 46:28S, 2003

28. Snider GL: Clinical relevance summary: collagen vs elastin in pathogenesis of emphysema; cellular origin of elastases; bronchiolitis vs emphysema as a cause of airflow obstruction. Chest 117(suppl 1):244S, 2000

29. Boots AW, Haenen GR, Bast A: Oxidant metabolism in chronic obstructive pulmonary disease. Eur Respir J Suppl 46:14S, 2003

30. Agusti AG, Noguera A, Sauleda J, et al: Systemic effects of chronic obstructive pulmonary disease. Eur Respir J 21:347, 2003

31. Sin DD, Man SF: Why are patients with chronic obstructive pulmonary disease at increased risk of cardiovascular diseases? The potential role of systemic inflammation in chronic obstructive pulmonary disease. Circulation 107:1514, 2003

32. Creutzberg EC, Casaburi R: Endocrinological disturbances in chronic obstructive pulmonary disease. Eur Respir J 46:76s, 2003

33. Ionescu AA, Schoon E: Osteoporosis in chronic obstructive pulmonary disease. Eur Respir J Suppl 46:64S, 2003

34. Kishi K, Gurney JW, Schroeder DR, et al: The correlation of emphysema or airway obstruction with the risk of lung cancer: a matched case-controlled study. Eur Respir J 19:1093, 2002

35. Chitkara RK, Sarinas PS: Recent advances in diagnosis and management of chronic bronchitis and emphysema. Curr Opin Pulm Med 8:126, 2002

36. Van Schayck CP, Loozen JM, Wagena E, et al: Detecting patients at a high risk of developing chronic obstructive pulmonary disease in general practice: cross sectional case finding study. BMJ 324:1370, 2002

37. Calverley PM, Burge PS, Spencer S, et al: Bronchodilator reversibility testing in chronic obstructive pulmonary disease. Thorax 58:659, 2003

38. O'Donoghue FJ, Catcheside PG, Ellis EE, et al: Sleep hypoventilation in hypercapnic chronic obstructive pulmonary disease: prevalence and associated factors. Eur Respir J 21:977, 2003

39. Barbera JA, Peinado VI, Santos S: Pulmonary hypertension in chronic obstructive pulmonary disease. Eur Respir J 21:892, 2003

40. Cotton DJ, Soparkar GR, Grahan BL: Diffusing capacity in the clinical assessment of chronic airflow limitation. Med Clin North Am 80:549, 1996

41. Muller NL, Coxson H: Chronic obstructive pulmonary disease. 4: Imaging the lungs in patients with chronic obstructive pulmonary disease. Thorax 57:982, 2002

42. Sin DD, McAlister FA, Man SF, et al: Contemporary management of chronic obstructive pulmonary disease: scientific review. JAMA 290:2301, 2003

43. Fabbri LM, Hurd SS, et al: Global strategy for the diagnosis, management and prevention of COPD: 2003 update. GOLD Scientific Committee. Eur Respir J 22:1, 2003

44. Celli BR, MacNee W: Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J 23:932, 2004

45. Lacasse Y, Ferreira I, Brooks D, et al: Critical appraisal of clinical practice guidelines targeting chronic obstructive pulmonary disease. Arch Intern Med 161:69, 2001

46. Roche N, Lepage T, Bourcereau J, et al: Guidelines versus clinical practice in the treatment of chronic obstructive pulmonary disease. Eur Respir J 18:903, 2001

47. Sin DD, Man SF: Inhaled corticosteroids and survival in chronic obstructive pulmonary disease: does the dose matter? Eur Respir J 21:260, 2003

48. Morgan MD, Britton JR: Chronic obstructive pulmonary disease. 8: Non-pharmacological management of COPD. Thorax 58:453, 2003

49. Gorecka D, Bednarek M, Nowinski A, et al: Diagnosis of airflow limitation combined with smoking cessation advice increases stop-smoking rate. Chest 123:1916, 2003

50. Connett JE, Murray RP, Buist AS, et al: Changes in smoking status affect women more than men: results of the Lung Health Study. Am J Epidemiol 157:973, 2003

51. Poole P, Chacko E, Wood-Baker R, et al: Influenza vaccine for patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev (4):CD002733, 2000

52. Dear K, Holden J, Andrews R, et al: Vaccines for preventing pneumococcal infection in adults. Cochrane Database Syst Rev (4):CD000422, 2003

53. Foxwell AR, Cripps AW, Dear KB: *Haemophilus influenzae* oral whole cell vaccination for preventing acute exacerbations of chronic bronchitis. Cochrane Database Syst Rev (3):CD001958, 2003

54. Di Marco F, Milic-Emili J, Boveri B, et al: Effect of inhaled bronchodilators on inspiratory capacity and dyspnoea at rest in COPD. Eur Resp J 21:86, 2003

55. Tashkin DP, Cooper CB: The role of long-acting bronchodilators in the management of stable COPD. Chest 125:249, 2004

56. Calverley P, Pauwels R, Vestbo J, et al: Combined salmeterol and fluticasone in the treatment of chronic obstructive pulmonary disease: a randomised controlled trial. Lancet 361:449, 2003

57. Barnes PJ: Theophylline: new perspectives for an old drug. Am J Respir Crit Care Med 167:813, 2003

58. Turner MO, Patel A, Ginsburg S, et al: Bronchodilator delivery in acute airflow obstruction: a meta-analysis. Arch Intern Med 157:1736, 1997

59. Hargreave FE, Leigh R: Induced sputum, eosinophilic bronchitis, and chronic obstructive pulmonary disease (pt 2). Am J Respir Crit Care Med 160:S53, 1999

60. Hattotuwa KL, Gizycki MJ, Ansari TW, et al: The effects of inhaled fluticasone on airway inflammation in chronic obstructive pulmonary disease: a double-blind, placebo-controlled biopsy study. Am J Respir Crit Care Med 165:1592, 2002

61. Highland KB, Strange C, Heffner JE: Long-term effects of inhaled corticosteroids on FEV_1 in patients with chronic obstructive pulmonary disease: a meta-analysis. Ann Intern Med 138:969, 2003

62. Alsaeedi A, Sin DD, McAlister FA: The effects of inhaled corticosteroids in chronic obstructive pulmonary disease: a systematic review of randomized placebo-controlled trials. Am J Med 113:59, 2002

63. Sin DD, Tu JV: Inhaled corticosteroids and the risk of mortality and readmission in elderly patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 164:580, 2001

64. Jones PW, Willits LR, Burge PS, et al: Disease severity and the effect of fluticasone propionate on chronic obstructive pulmonary disease exacerbations. Eur Respir J 21:68, 2003

65. Jones A, Fay JK, Burr M, et al: Inhaled corticosteroid effects on bone metabolism in asthma and mild chronic obstructive pulmonary disease. Cochrane Database Syst Rev (1):CD003537, 2002

66. Eichenhorn MS, Wise RA, Madhok TC, et al: Lack of long-term adverse adrenal effects from inhaled triamcinolone: Lung Health Study II. Chest 124:57, 2003

67. Stoller JK: Clinical practice: acute exacerbations of chronic obstructive pulmonary disease. N Engl J Med 346:988, 2002

68. Snow V, Lascher S, Morttur-Pilson C: Evidence base for management of acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med 134:595, 2001

69. White AJ, Gompertz S, Stockley RA: Chronic obstructive pulmonary disease. 6: The aetiology of exacerbations of chronic obstructive pulmonary disease. Thorax 58:73, 2003

70. Niederman MS: Antibiotic therapy of exacerbations of chronic bronchitis. Semin Respir Infect 15:59, 2000

71. Barr RG, Rowe BH, Camargo CA Jr: Methylxanthines for exacerbations of chronic obstructive pulmonary disease. Cochrane Database Syst Rev (2):CD002168, 2003

72. Singh JM, Palda VA, Stanbrook MB, et al: Corticosteroid therapy for patients with acute exacerbations of chronic obstructive pulmonary disease: a systematic review. Arch Intern Med 162:2527, 2002

73. Calverley PM: Respiratory failure in chronic obstructive pulmonary disease. Eur Respir J Suppl 47:26S, 2003

74. Maloney JP: Advances in the treatment of secondary pulmonary hypertension. Curr Opin Pulm Med 9:139, 2003

75. Lacasse Y, Brosseau L, Milne S, et al: Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev (3):CD003793, 2002

76. Crockett AJ, Cranston JM, Moss JR, et al: Domiciliary oxygen for chronic obstructive pulmonary disease. Cochrane Database Syst Rev (3):CD002883, 2001

77. Yaeger ES, Goodman S, Hoddes E, et al: Oxygen therapy using pulse and continuous flow with a transtracheal catheter and a nasal cannula. Chest 106:854, 1994

78. Wijkstra PJ, Lacasse Y, Guyatt GH, et al: A meta-analysis of nocturnal noninvasive positive pressure ventilation in patients with stable COPD. Chest 124:337, 2003

79. Geddes DM: Lung volume reduction surgery. Thorax 54(suppl 2):S14, 1999

80. Marchand E, Gayan-Ramirez G, De Leyn P, et al: Physiological basis of improvement after lung volume reduction surgery for severe emphysema: where are we? Eur Respir J 13:686, 1999

81. Teschler H, Thompson AB, Stamatis G: Short- and long-term functional results after lung volume reduction surgery for severe emphysema. Eur Respir J 13:1170, 1999

82. Naunheim KS, Kaiser LR, Bavaria JE, et al: Long-term survival after thoracoscopic lung volume reduction: a multiinstitutional review. Ann Thorac Surg 68:2026, 1999

83. Ingenito EP, Loring SH, Moy ML, et al: Interpreting improvement in expiratory flows after lung volume reduction surgery in terms of flow limitation theory. Am J Respir Crit Care Med 163:1074, 2001

84. Meyers BF, Yusen RD, Guthrie TJ, et al: Outcome of bilateral lung volume reduction in patients with emphysema potentially eligible for lung transplantation. J Thorac Cardiovasc Surg 122:10, 2001

85. Gelb AF, McKenna RJ Jr, Brenner M, et al: Lung function 5 yr after lung volume reduction surgery for emphysema. Am J Respir Crit Care Med 163:1562, 2001

86. Criner GJ, Cordova FC, Furukawa S, et al: Prospective randomized trial comparing bilateral lung volume reduction surgery to pulmonary rehabilitation in severe chronic obstructive pulmonary disease. Am J Respir Crit Care Med 160:2018, 1999

87. Geddes D, Davies M, Koyama H, et al: Effect of lung-volume-reduction surgery in patients with severe emphysema. N Engl J Med 343:239, 2000

88. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. National Emphysema Treatment Trial Research Group. N Engl J Med 348:2059, 2003

89. Patients at high risk of death after lung-volume-reduction surgery. National Emphysema Treatment Trial Research Group. N Engl J Med 345:1075, 2001

90. Glanville AR, Estenne M: Indications, patient selection and timing of referral for lung transplantation. Eur Respir J 22:845, 2003

91. Low DE, Trulock EP, Kaiser LR, et al: Morbidity, mortality, and early results of single versus bilateral lung transplantation for emphysema. J Thorac Cardiovasc Surg 103:1119, 1992

92. Cassivi SD, Meyers BF, Battafarano RJ, et al: Thirteen-year experience in lung transplantation for emphysema. Ann Thorac Surg 74:1663, 2002

93. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. Am J Respir Crit Care Med 168:818, 2003

94. Lomas DA, Mahadeva R: Alpha$_1$-antitrypsin polymerization and the serpinopathies: pathobiology and prospects for therapy. J Clin Invest 110:1585, 2002

95. de Serres FJ: Worldwide racial and ethnic distribution of alpha$_1$-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. Chest 122:1818, 2002

96. Weitz JI, Silverman EK, Thong B, et al: Plasma levels of elastase-specific fibrinopeptides correlate with proteinase inhibitor phenotype: evidence for increased elastase activity in subjects with homozygous and heterozygous deficiency of alpha 1-proteinase inhibitor. J Clin Invest 89:766, 1992

97. Woolhouse IS, Bayley DL, Stockley RA: Sputum chemotactic activity in chronic obstructive pulmonary disease: effect of alpha(1)-antitrypsin deficiency and the role of leukotriene B(4) and interleukin 8. Thorax 57:709, 2002

98. Seersholm N, Kok-Jensen A: Clinical features and prognosis of lifetime non-smokers with severe alpha 1-antitrypsin deficiency. Thorax 53:265, 1998

99. Eden E, Hammel J, Rouhani FN, et al: Asthma features in severe alpha$_1$-antitrypsin deficiency: experience of the National Heart, Lung, and Blood Institute Registry. Chest 123:765, 2003

100. King MA, Stone JA, Diaz PT, et al: Alpha 1-antitrypsin deficiency: evaluation of bronchiectasis with CT. Radiology 199:137, 1996

101. Dahl M, Tybjaerg-Hansen A, Lange P, et al: Change in lung function and morbidity from chronic obstructive pulmonary disease in alpha$_1$-antitrypsin MZ heterozygotes: a longitudinal study of the general population. Ann Intern Med 136:270, 2002

102. Tarjan E, Magyar P, Vaczi Z, et al: Longitudinal lung function study in heterozygous PiMZ phenotype subjects. Eur Respir J 7:2199, 1994

103. Stone PJ, Morris TA III, Franzblau C, et al: Preliminary evidence that augmentation therapy diminishes degradation of cross-linked elastin in alpha-1-antitrypsin-deficient humans. Respiration 62:76, 1995

104. Dirksen A, Dijkman JH, Madsen F, et al: A randomized clinical trial of alpha(1)-antitrypsin augmentation therapy (pt 1). Am J Respir Crit Care Med 160:1468, 1999

105. Wencker M, Fuhrmann B, Banik N, et al: Longitudinal follow-up of patients with alpha(1)-protease inhibitor deficiency before and during therapy with IV alpha(1)-protease inhibitor. Chest 119:737, 2001

106. Gildea TR, Shermock KM, Singer ME, et al: Cost-effectiveness analysis of augmentation therapy for severe alpha$_1$-antitrypsin deficiency. Am J Respir Crit Care Med 167:1387, 2003

107. Snider GL, Lucey EC, Stone PJ: Pitfalls in antiprotease therapy of emphysema. Am J Respir Crit Care Med 150:S131, 1994

108. Flotte TR: Recombinant adeno-associated virus gene therapy for cystic fibrosis and alpha(1)-antitrypsin deficiency. Chest 121:98S, 2002

109. Gelb AF, McKenna RJ, Brenner M, et al: Lung function after bilateral lower lobe lung volume reduction surgery for alpha$_1$-antitrypsin emphysema. Eur Respir J 14:928, 1999

110. Trulock EP: Lung transplantation for alpha 1-antitrypsin deficiency emphysema. Chest 110:284S, 1996

111. Snider GL: Reduction pneumoplasty for giant bullous emphysema: implications for surgical treatment of nonbullous emphysema. Chest 109:540, 1996

112. Barker AF: Bronchiectasis. N Engl J Med 346:1383, 2002

113. Cuvelier A, Muir JF, Hellot MF, et al: Distribution of alpha (1)-antitrypsin alleles in patients with bronchiectasis. Chest 117:415, 2000

114. Kainulainen L, Liipo K, Svedareom E, et al: Pulmonary abnormalities in patients with primary hypogammaglobulinemia. J Allergy Clin Immunol 104:1031, 1999

115. Hill SL, Mitchell JL, Burnett D, et al: IgG subclasses in the serum and sputum from patients with bronchiectasis. Thorax 53:463, 1998

116. Cohen M, Sahn SA: Bronchiectasis in systemic diseases. Chest 116:1063, 1999

117. Shadick NA, Fanta CH, Weinblatt ME, et al: Bronchiectasis: a late feature of severe rheumatoid arthritis. Medicine (Baltimore) 73:161, 1994

118. Bard M, Coudere LJ, Saimot AG, et al: Accelerated obstructive pulmonary disease in HIV infected patients with bronchiectasis. Eur Respir J 11:771, 1998

119. Evans SA, Turner SM, Bosch BJ, et al: Lung function in bronchiectasis: the influence of *Pseudomonas aeruginosa*. Eur Respir J 9:1601, 1996

120. Gaga M, Bently AM, Humbert M, et al: Increases in CD4$^+$ T lymphocytes, macrophages, neutrophils and interleukin 8 positive cells in the airways of patients with bronchiectasis. Thorax 53:685, 1998

121. Mikami M, Llewellyn-Jones CG, Bayley D, et al: The chemotactic activity of sputum from patients with bronchiectasis. Am J Respir Crit Care Med 157:723, 1998

122. Schaaf B, Wieghorst A, Aries SP, et al: Neutrophil inflammation and activation in bronchiectasis: comparison with pneumonia and idiopathic pulmonary fibrosis. Respiration 67:52, 2000

123. Fahy JV, Schuster A, Ueki I, et al: Mucus hypersecretion in bronchiectasis: the role of neutrophil proteases. Am Rev Respir Dis 146:1430, 1992

124. Cartier Y, Kavanagh PV, Johkoh T, et al: Bronchiectasis: accuracy of high-resolution CT in the differentiation of specific diseases. AJR Am J Roentgenol 173:47, 1999

125. Lynch DA, Newell J, Hale V, et al: Correlation of CT findings with clinical evaluations in 261 patients with symptomatic bronchiectasis. AJR Am J Roentgenol 173:53, 1999

126. Wallace RJ Jr, Zhang Y, Brown BA, et al: Polyclonal *Mycobacterium avium* complex infections in patients with nodular bronchiectasis. Am J Respir Crit Care Med 158:1235, 1998

127. Evans DJ, Bara AI, Greenstone M: Prolonged antibiotics for purulent bronchiectasis. Cochrane Database Syst Rev (4):CD001392, 2003

128. Barker AF, Couch L, Fiel SB, et al: Tobramycin solution for inhalation reduces sputum *Pseudomonas aeruginosa* density in bronchiectasis (pt 1). Am J Respir Crit Care Med 162:481, 2000

129. Crockett AJ, Cranston JM, Latimer KM, et al: Mucolytics for bronchiectasis. Cochrane Database Syst Rev (1):CD001289, 2001

130. Steele K, Greenstone M, Lasserson J: Oral methyl-xanthines for bronchiectasis. Cochrane Database Syst Rev (1):CD002734, 2001

131. Sheikh A, Nolan D, Greenstone M: Long-acting beta-2-agonists for bronchiectasis. Cochrane Database Syst Rev (4):CD002155, 2001

132. Lasserson T, Holt K, Evans D, et al: Anticholinergic therapy for bronchiectasis. Cochrane Database Syst Rev (4):CD002163, 2002

133. Franco F, Sheikh A, Greenstone M: Short acting beta-2 agonists for bronchiectasis. Cochrane Database Syst Rev (3):CD003572, 2003

134. Lasserson T, Holt K, Greenstone M: Oral steroids for bronchiectasis (stable and acute exacerbations). Cochrane Database Syst Rev (4):CD002162, 2001

135. Tsang KW, Ho PL, Lam WK, et al: Inhaled fluticasone reduces sputum inflammatory indices in severe bronchiectasis. Am J Respir Crit Care Med 158:723, 1998

136. Kolbe J, Wells A, Ram FS: Inhaled steroids for bronchiectasis. Cochrane Database Syst Rev (2):CD000996, 2000

137. Corless JA, Warburton CJ: Leukotriene receptor antagonists for non-cystic fibrosis bronchiectasis. Cochrane Database Syst Rev (4):CD002174, 2000

138. Corless J, Warburton C: Surgery versus non-surgical treatment for bronchiectasis. Cochrane Database Syst Rev (4):CD002180, 2000

139. Ratjen F, Doring G: Cystic fibrosis. Lancet 361:681, 2003

140. Mall M, Bleich M, Greger R, et al: The amiloride-inhibitable Na^+ conductance is reduced by the cystic fibrosis transmembrane conductance regulator in normal but not in cystic fibrosis airways. J Clin Invest 102:15, 1998

141. Johansen HK, Nir M, Hoiby N, et al: Severity of cystic fibrosis in patients homozygous and heterozygous for DF508 mutation. Lancet 337:631, 1991

142. McKone EF, Emerson SS, Edwards KL, et al: Effect of genotype on phenotype and mortality in cystic fibrosis: a retrospective cohort study. Lancet 361:1671, 2003

143. Highsmith WE, Burch LH, Zhou Z, et al: A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. N Engl J Med 331:974, 1994

144. Merlo CA, Boyle MP: Modifier genes in cystic fibrosis lung disease. J Lab Clin Med 141:237, 2003

145. Gibson RL, Burns JL, Ramsey BW: Pathophysiology and management of pulmonary infections in cystic fibrosis. Am J Respir Crit Care Med 168:918, 2003

146. Balfour-Lynn IM, Elborn JS: "CF asthma": what is it and what do we do about it? Thorax 57:742, 2002

147. Griffith DE: Emergence of nontuberculous mycobacteria as pathogens in cystic fibrosis. Am J Respir Crit Care Med 167:810, 2003

148. Flume PA: Pneumothorax in cystic fibrosis. Chest 123:217, 2003

149. Bradley S, Solin P, Wilson J, et al: Hypoxemia and hypercapnia during exercise and sleep in patients with cystic fibrosis. Chest 116:647, 1999

150. Fraser KL, Tullis DE, Sasson Z, et al: Pulmonary hypertension and cardiac function in adult cystic fibrosis: role of hypoxemia. Chest 115:1321, 1999

151. Aris RM, Renner JB, Winders AD, et al: Increased rate of fractures and severe kyphosis: sequelae of living into adulthood with cystic fibrosis. Ann Intern Med 128:186, 1998

152. Rosenstein BJ: What is a cystic fibrosis diagnosis? Clin Chest Med 19:423, 1998

153. Genetic testing for cystic fibrosis. National Institutes of Health Consensus Development Conference Statement on genetic testing for cystic fibrosis. Arch Intern Med 159:1529, 1999

154. Yankaskas JR, Marshall BC, Sufian B, et al: Cystic fibrosis adult care: consensus conference report. Chest 125(1 suppl):1S, 2004

155. Arens R, Gozal D, Omlin KJ, et al: Comparison of high frequency chest compression and conventional chest physiotherapy in hospitalized patients with cystic fibrosis. Am J Respir Crit Care Med 150:1154, 1994

156. App EM, Kieselmann R, Reinhardt D, et al: Sputum rheology changes in cystic fibrosis lung disease following two different types of physiotherapy: flutter vs autogenic drainage. Chest 114:171, 1998

157. van der Schans C, Prasad A, Main E: Chest physiotherapy compared to no chest physiotherapy for cystic fibrosis. Cochrane Database Syst Rev (2):CD001401, 2000

158. Hodson ME, Gallagher CG, Govan JR: A randomised clinical trial of nebulised tobramycin or colistin in cystic fibrosis. Eur Resp J 20:658, 2002

159. Elborn JS, Prescott RJ, Stack BH, et al: Elective versus symptomatic antibiotic treatment in cystic fibrosis patients with chronic *Pseudomonas* infection of the lungs. Thorax 55:355, 2000

160. Nguyen T, Louie SG, Beringer PM, et al: Potential role of macrolide antibiotics in the management of cystic fibrosis lung disease. Curr Opin Pulm Med 8:521, 2002

161. Jones AP, Wallis CE: Recombinant human deoxyribonuclease for cystic fibrosis. Cochrane Database Syst Rev (3):CD001127, 2003

162. Cropp GJ: Effectiveness of bronchodilators in cystic fibrosis. Am J Med 100:19S, 1996

163. Dezateux C, Walters S, Balfour-Lynn I: Inhaled corticosteroids for cystic fibrosis. Cochrane Database Syst Rev (2):CD001915, 2000

164. Cheng K, Ashby D, Smyth R: Oral steroids for cystic fibrosis. Cochrane Database Syst Rev (2):CD000407, 2000

165. Allen ED: Opportunities for the use of aerosolized alpha 1-antitrypsin for the treatment of cystic fibrosis. Chest 110:256S, 1996

166. Konstan MW: Therapies aimed at airway inflammation in cystic fibrosis. Clin Chest Med 19:505, 1998

167. Barben J, Robertson D, Olinsky A, et al: Bronchial artery embolization for hemoptysis in young patients with cystic fibrosis. Radiology 224:124, 2002

168. Moran F, Bradley J: Non-invasive ventilation for cystic fibrosis. Cochrane Database Syst Rev (2):CD002769, 2003

169. Liou TG, Cahill BC, Adler FR, et al: Selection of patients with cystic fibrosis for lung transplantation. Curr Opin Pulm Med 8:535, 2002

170. Vricella LA, Karamichalis JM, Ahmad S, et al: Lung and heart-lung transplantation in patients with end-stage cystic fibrosis: the Stanford experience. Ann Thorac Surg 74:13, 2002

171. Zeitlin PL: Future pharmacological treatment of cystic fibrosis. Respiration 67:351, 2000

172. Gelman MS, Kopito RR: Rescuing protein conformation: prospects for pharmacological therapy in cystic fibrosis. J Clin Invest 110:1591, 2002

173. Welsh MJ: Gene transfer for cystic fibrosis. J Clin Invest 104:1165, 1999

174. Afzelius BA: Genetics and pulmonary medicine: 6. Immotile cilia syndrome: past, present, and prospects for the future. Thorax 53:894, 1998

175. Tsang KW, Tipoe G, Sun J, et al: Severe bronchiectasis in patients with "cystlike" structures within the ciliary shafts (pt 1). Am J Respir Crit Care Med 161:1300, 2000

176. Noone PG, Leigh MW, Sannuti A, et al: Primary ciliary dyskinesia: diagnostic and phenotypic features. Am J Respir Crit Care Med 169:459, 2004

177. Bush A, Cole P, Hariri M, et al: Primary ciliary dyskinesia: diagnosis and standards of care. Eur Respir J 12:982, 1998

178. Rayner CF, Rutman A, Dewar A, et al: Ciliary disorientation in patients with chronic upper respiratory tract inflammation. Am J Respir Crit Care Med 151:800, 1995

179. Ellerman A, Bisgaard H: Longitudinal study of lung function in a cohort of primary ciliary dyskinesia. Eur Respir J 10:2376, 1997

180. Ryu JH, Myers JL, Swensen SJ: Bronchiolar disorders. Am J Respir Crit Care Med 168:1277, 2003

181. Hansell DM: Small airways diseases: detection and insights with computed tomography. Eur Respir J 17:1294, 2001

182. Burrows B, Knudson RJ, Cline MG, et al: Quantitative relationships between cigarette smoking and ventilatory function. Am Rev Respir Dis 115:195, 1977

183. Fletcher C, Peto R: The natural history of chronic airflow obstruction. Br Med J 1:1645, 1977

184. National Asthma Education and Prevention Program Expert Panel Report 2: Guidelines for the diagnosis and management of asthma. National Institutes of Health publication no. 97-4051, Washington, DC, 1997

http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm

Acknowledgments

Figures 1 and 2 Al Miller.

Figure 4 Dana Burns-Pizer.

Figures 9 and 10 Seward Hung.

The previous version of this chapter was coauthored by Roland H. Ingram, Jr., M.D., F.A.C.P.

IV FOCAL AND MULTIFOCAL LUNG DISEASE

Gerald W. Staton, Jr., m.d., f.a.c.p.

Chest physicians, when consulted, most often begin their consultation by examining the results of imaging studies—standard posteroanterior and lateral chest radiographs, computed tomography scans of the chest, or both. In doing so, they assess the differential diagnostic probabilities and possibilities on the basis of the radiographic patterns. This approach serves to focus their efforts on history taking and the physical examination, and it facilitates their defining of proper diagnostic measures, which in turn guide therapeutic and management advice. In this chapter, the approach taken by chest physicians is used to discuss and categorize a wide range of lung diseases.

Assessment of Chest Radiographs

Focal and multifocal lung diseases are classified into seven categories on the basis of chest radiography and CT: (1) focal pulmonary infiltrates, (2) multifocal pulmonary infiltrates, (3) true segmental infiltrates, (4) cavitary infiltrates, (5) single small nodules, (6) large masses, and (7) multiple nodules. These radiographic patterns may be caused by infectious diseases; neoplastic diseases; or noninfectious, nonneoplastic disorders. The differential diagnoses for the seven categories are dissimilar but by no means mutually exclusive. Many diseases that usually cause focal infiltrates can produce multifocal infiltrates [*see* Focal Pulmonary Infiltrates, *below*]. Other disorders nearly always present as multiple infiltrates, and the pathology only rarely localizes to one area [*see* Multifocal Pulmonary Infiltrates, *below*]. Infiltrates that conform perfectly to the segmental anatomy of the lung usually result from an abnormality of the bronchus or pulmonary artery leading to the infiltrate [*see* True Segmental Infiltrates, *below*]. Cavitary infiltrates, single small nodules, large masses, and multiple nodules have distinct differential diagnoses.

Diagnosis of a focal or multifocal lung disorder starts with the abnormal chest radiograph or with abnormal findings on chest CT. In many instances, the chest CT scan may be abnormal when the chest radiograph is normal or shows very indistinct changes. In each category of radiographic pattern, the clinical features of the illness, the presence or absence of associated pleural or mediastinal abnormalities, and ancillary laboratory tests all serve to narrow the differential diagnosis. In some disorders, the combined radiographic, clinical, and laboratory presentation is virtually specific. In other disorders, cytologic, histopathologic, or microbiologic information is necessary to make a specific diagnosis.

In each of the seven categories, infectious diseases and neoplastic diseases are more common than noninfectious, nonneoplastic disorders. Many of the infectious and neoplastic entities, however, are discussed in great detail in other chapters; this chapter focuses on the noninfectious, nonneoplastic disorders that are relatively common.

Focal Pulmonary Infiltrates

When a focal infiltrate is dense, it is likely that pus, blood, water, or tissue is filling alveolar spaces. A focal infiltrate that is patchy and less dense suggests a less advanced stage of disease process. Many conditions can cause a focal infiltrate that is visible on the chest radiograph [*see Table 1*].

Infectious Diseases

Bacterial Pneumonia

The most common cause of a focal infiltrate is bacterial pneumonia.[1] Five clinical features in combination strongly suggest the diagnosis: (1) acute onset, (2) a new or increasing infiltrate on the chest radiograph [*see Figure 1a*], (3) fever, (4) purulent sputum, and (5) a white blood cell count that is either high, low, or shifted to the left. Absence of one or more of these features does not eliminate the possibility of bacterial pneumonia but does increase the probability of an alternative diagnosis.

Because many different types of bacteria can cause pneumonia, a precise etiologic diagnosis cannot be made either on clinical grounds or by a chest radiograph. Positive blood cultures have near-perfect specificity but low sensitivity, whereas positive sputum cultures can only suggest a specific etiology. *Streptococcus pneumoniae* is the most common cause of bacterial pneumonia, accounting for perhaps 85% of all cases in otherwise healthy young adults. However, many other bacteria also cause pneumonia, with a higher incidence in patients with chronic medical conditions and advanced age. These include *Haemophilus influenzae, Klebsiella pneumoniae, Staphylococcus aureus,* and *Legionella pneumophila* [*see 7:XX Pneumonia and Other Pulmonary Infections*]. An even wider range of pathogens cause pneumonia in patients in intensive care units. These include virulent gram-negative bacilli with high potential for antibiotic resistance, such as *Pseudomonas aeruginosa* and *Enterobacter, Serratia,* and *Proteus* species. Other infectious causes of focal infiltrates include *Mycoplasma pneumoniae, Chlamydophila pneumoniae,* tuberculosis (especially in patients with advanced-stage HIV infection and other immunosuppressed patients), viruses, rickettsiae, fungi, and parasites.

Neoplastic Diseases

Alveolar cell carcinoma and lymphoma commonly present as focal pulmonary infiltrates, though they also cause multifocal in-

Table 1 Major Causes of Focal Pulmonary Infiltrates

Cause	*Examples*
Infectious	Bacterial pneumonia Tuberculosis
Neoplastic	Primary lung cancer Alveolar cell carcinoma Lymphoma
Noninfectious, nonneoplastic	Radiation pneumonitis Lipoid pneumonia Lung contusion Pulmonary embolism Lobe torsion

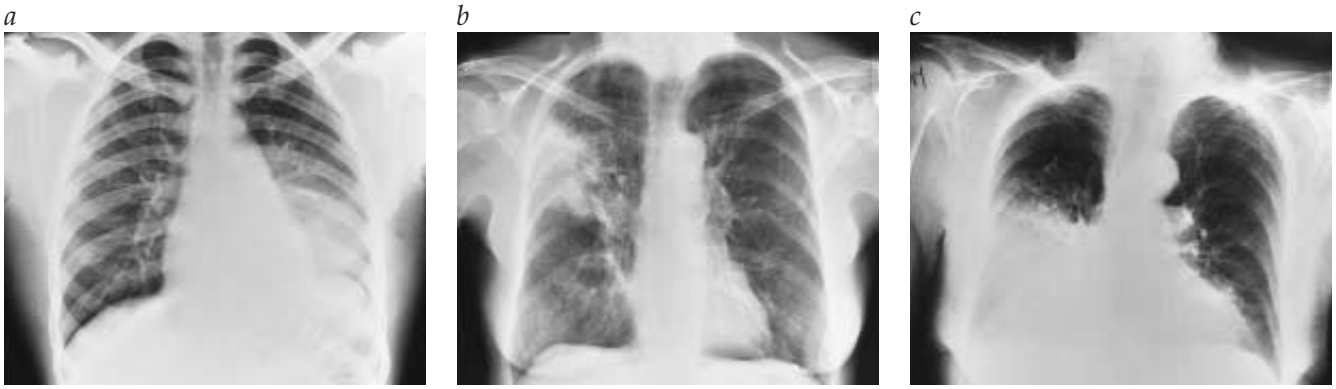


Figure 1 **(*a*) This chest radiograph demonstrates a focal left lower lobe infiltrate caused by bacteremic pneumococcal pneumonia in a 22-year-old man. (*b*) A focal right upper lobe infiltrate caused by alveolar cell carcinoma is revealed in this chest radiograph of a 71-year-old woman. (*c*) The focal right lower lobe infiltrate in this chest radiograph is the result of lipoid pneumonia in a 68-year-old woman.**

filtrates. Bronchogenic lung cancer usually produces a nodule or a mass but can cause a focal dense infiltrate. The absence of fever and purulent secretions (unless there is a postobstructive pneumonia) and the longer duration of symptoms suggest a noninfectious cause.

Alveolar Cell Carcinoma

Alveolar cell carcinoma may result in one or more areas of airspace opacity, presenting as an area of focal indistinct infiltrate or, sometimes, mimicking lobar pneumonia [*see Figure 1b*]. Chest CT often shows areas of "ground-glass opacity"; less commonly, the infiltrates are consolidative.[2] Patients cough and produce mucoid sputum; a few patients produce large volumes of sputum (bronchorrhea) that in rare instances has a salty taste. Weight loss and malaise are common. Fever and chills are absent. Metastases are less common than with other primary lung neoplasms, and the course of the illness is longer. Alveolar cell carcinoma is not related to smoking. Diagnostic tests should begin with sputum cytology, followed by fiberoptic bronchoscopy with bronchoalveolar lavage and transbronchial biopsy or, if needed, open lung biopsy by a traditional or a video-assisted thoracoscopic approach. Localized disease can sometimes be resected; radiation therapy and chemotherapy have no important role.

Lymphoma

Lymphoma can also produce a focal, dense consolidation. With Hodgkin disease, there may be multiple contiguous areas of tumor involvement that grow and merge, resulting in a dense infiltrate with irregular borders. Hilar and mediastinal adenopathy are nearly always present. Pleural effusions occur in as many as 30% of patients. Diagnostic strategies include fiberoptic bronchoscopy, aspiration biopsy of the infiltrate or of the mediastinal nodes (using needles that obtain a tissue core), mediastinoscopy, and pleural biopsy. Biopsy of associated cervical nodes or of the bone marrow is helpful in some cases. In non-Hodgkin lymphoma, the chest radiograph may show a dense infiltrate with regular margins, similar to the infiltrate seen in bacterial pneumonia or alveolar cell carcinoma. Mediastinal and hilar node involvement may be absent. Pleural effusion occurs in one third of patients [*see 12:XI Lymphomas*].

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Pulmonary embolism can cause a focal infiltrate or focal atelectasis; in other cases, the chest radiograph is normal or shows multifocal infiltrates or multiple nodules.[3] Pulmonary infarcts are always based at a visceral pleural surface, including fissures [*see 1:XVIII Venous Thromboembolism*]. Several other noninfectious, nonneoplastic disorders typically present as a focal pulmonary infiltrate. In making a diagnosis, the clinical history is very important because these disorders occur in specific clinical settings.

Radiation Pneumonitis

Radiation pneumonitis follows pulmonary irradiation after a lag time that is somewhat dose dependent.[4] Symptoms include nonproductive cough, dyspnea, and fever. These symptoms develop approximately 8 weeks after completion of a course of radiation therapy consisting of 4,000 cGy; they develop 1 week earlier for each additional 1,000 cGy administered.[1] Chemotherapeutic agents such as dactinomycin, doxorubicin, bleomycin, and busulfan can potentiate the effects of radiation. Onset is usually subacute. There may be rales and signs of consolidation. Hyperpigmentation of the skin overlying the irradiated lung is common but does not correlate with the severity of lung injury. Laboratory findings include leukocytosis and hypoxemia. Bronchoalveolar lavage may rule out tumor and infection—a determination that is most important in making a differential diagno-

Table 2 Major Causes of Multifocal Pulmonary Infiltrates

Cause	*Examples*
Infectious	Bacterial pneumonia *Pneumocystis jiroveci* pneumonia Influenza Tuberculosis Endemic fungal pneumonias Invasive aspergillosis
Neoplastic	Alveolar cell carcinoma Lymphoma
Noninfectious, nonneoplastic	Drug reactions Simple eosinophilic pneumonia Chronic eosinophilic pneumonia Bronchiolitis obliterans organizing pneumonia Allergic granulomatosis and angiitis Sarcoidosis

sis—and may demonstrate dysplastic type II cells, the presence of which suggests radiation injury. Lavage also reveals excess lymphocytes in both involved and uninvolved areas. Drug-induced lung injury causes similar cytopathic changes in type II cells, but usually, neutrophilic inflammation is also present.

The chest radiograph shows an infiltrate of variable density. A highly characteristic infiltrate has sharp edges and conforms exactly to the radiation port. Occasionally, high-resolution CT shows regions of air-space consolidation that are not visible on routine chest radiography. Many patients gradually improve over a few weeks. When the disease is severe, glucocorticoid therapy is often used, with uncertain benefit. Because of fever and fear of superinfection, antibacterial therapy is often given, also with uncertain benefit. Typically, the involved area of the lung scars and contracts with time, and the chest radiograph shows progressive volume loss.

Lipoid Pneumonia

Lipoid pneumonia is a noninfectious, inflammatory lung disorder caused by the aspiration of mineral oil or other oily substances. It is most common in elderly patients and others with impaired swallowing. Whether or not swallowing is impaired, lipoid pneumonia can result from the use of petroleum jelly or other oily substances applied to the lips or nose to relieve chronic dryness or mineral oil taken by mouth for relief of constipation. The most common symptom is a chronic cough, which may be caused by coexisting lung disease rather than lipoid pneumonia; fever is uncommon. Often, the disease is discovered on a routine chest radiograph that shows a focal, dense infiltrate, usually in a lower lobe or in the right middle lobe [*see Figure 1c*].[5] The radiographic appearance of such an infiltrate in a relatively asymptomatic patient suggests chronic pneumonia or lung cancer. CT scanning may show an extremely low density infiltrate produced by accumulated lipid; the density typically ranges from –60 to –150 Hounsfield units (water is 0). In contrast, the density of lung cancers usually ranges from +60 to +150 Hounsfield units. Lipid-laden macrophages can be demonstrated with the oil red O stain, which colors lipid droplets bright red. The stain can be applied to bronchoalveolar lavage specimens or transthoracic aspirates. A positive test supports the diagnosis, but some caution is necessary. Endogenous lipoid pneumonia may occur distal to an obstructed bronchus, and in such cases, the lipid is derived from the breakdown of cell membranes. Thus, bronchoscopy is still needed to rule out an obstructed bronchus even after demonstration of lipid-laden macrophages by needle aspiration. A variety of other lung disorders, such as pulmonary hemorrhage and primary and metastatic cancers, can also be associated with lipid-laden macrophages. To establish the diagnosis of exogenous lipoid pneumonia, it is necessary to analyze carefully all the clinical, cytopathologic, and radiographic findings, including the results of CT scanning. Many cases are diagnosed only after thoracotomy for resection of a presumed malignancy. The only specific therapy is avoidance of exposure to mineral oil and other lipid-containing agents. Lipoid pneumonia usually improves slowly after exposure to the agent is eliminated, though complete clearing of the infiltrate does not always occur.

Lung Contusion

Lung contusion is an important cause of focal, usually dense infiltrate. It results from blunt chest trauma, most often from falls or motor vehicle accidents. Usually within hours after trauma, an infiltrate develops deep beneath the impact point, representing blood and edema in the lung. Associated injuries, such as rib fractures and traumatic pneumothorax, may be present. Focal shunting through the area can cause refractory hypoxemia. If the injury is severe enough, the entire area may become necrotic and form a large cavity with irregular inner margins.

Lobe Torsion

Torsion of a lobe of the lung is rare and usually occurs postoperatively, particularly after resection of the left upper lobe. The vascular pedicle of the remaining left lower lobe twists and is compromised, and the lobe increases in density as it fills with blood and edema fluid. The diagnosis is often made by radiography but can be difficult. The differential diagnosis includes unilateral lung infection, edema, and hemorrhage. Treatment involves surgical relief of the torsion in early cases or resection of the lobe if it is no longer viable.

Multifocal Pulmonary Infiltrates

Most disorders that cause single infiltrates can also cause multiple infiltrates [*see Table 2*]. Pneumococcal pneumonia and other bacterial pneumonias are occasionally multifocal [*see Figure 2a*]; viral pneumonias are commonly multifocal or diffuse.[1] Clinical features of pneumonia with multiple infiltrates are similar to clin-

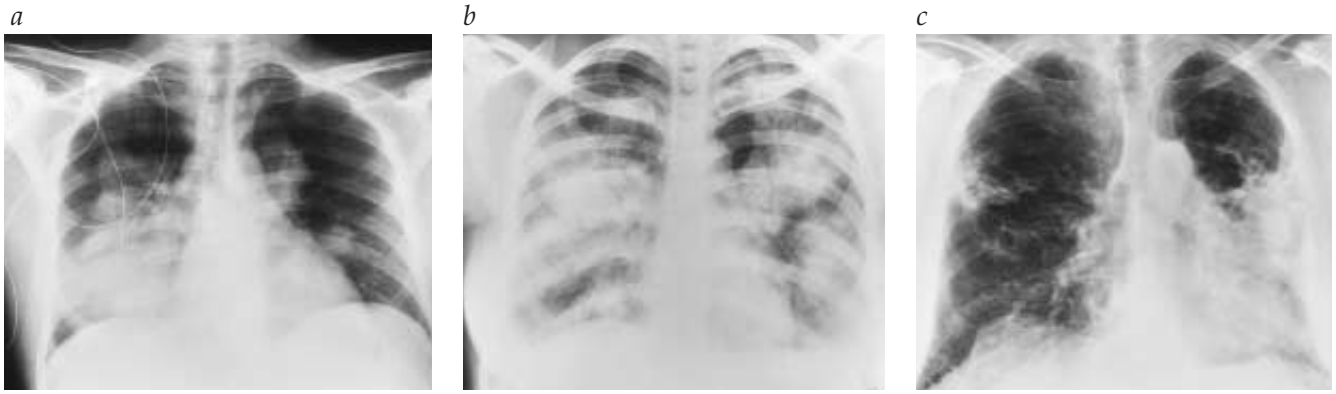


Figure 2 **(*a*) Bacteremic pneumococcal pneumonia caused the extensive bilateral multifocal infiltrates revealed in this chest radiograph of a 27-year-old man. (*b*) Alveolar sarcoidosis is the cause of the extensive bilateral multifocal infiltrates shown in this chest radiograph. The patient is a 22-year-old woman. (*c*) Alveolar cell carcinoma often presents as multifocal infiltrates, as seen in this chest radiograph of a 65-year-old man.**

ical features of pneumonia with one infiltrate, except that severity increases with extent of disease. Pulmonary thromboemboli can also produce multifocal infiltrates; a normal chest radiograph, unilateral or bilateral pleural effusions, and focal infiltrate or atelectasis are other possible radiographic patterns for pulmonary thromboemboli.[3] Septic pulmonary emboli often cause multiple infiltrates [*see* Multiple Nodules and Masses, *below*]. Finally, sarcoidosis is perhaps the most protean of all the noninfectious and nonmalignant lung disorders [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. A diffuse infiltrate with or without hilar adenopathy is the usual presentation [*see Figure 2b*].

INFECTIOUS DISEASES

Pneumocystis jiroveci *Pneumonia*

Pneumocystis jiroveci (formerly *P. carinii*) is an organism that was previously thought to be a protozoan but, on the basis of genetic studies, is currently considered to be a fungus. *Pneumocystis* pneumonia (PCP) typically occurs in patients who are immunocompromised, most commonly those with HIV/AIDS. Radiographically, PCP presents as multifocal infiltrates, usually hazy opacities in the perihilar regions.[6] These infiltrates can become diffuse and associated with severe respiratory compromise. CT shows ground-glass attenuation that may be diffuse or in a mosaic pattern. Less commonly, PCP presents as solitary or multiple, solid or cavitary opacities. Cysts, often associated with pneumothorax, can also be seen.

Tuberculosis

Primary tuberculosis is often focal and often associated with ipsilateral hilar adenopathy, especially in children. Most diagnosed adult cases of tuberculosis, however, result from reactivation of latent infection. Reactivation tuberculosis is often multifocal [*see 7:II Tuberculosis*].[7] Bilateral infiltrates in the upper lung zones are most characteristic. The upper lung zones are favored sites, because a higher ratio of ventilation to perfusion results in higher local oxygen tension, which enhances growth of *Mycobacterium tuberculosis.* The apical and posterior segments of the upper lobes are most commonly involved, followed by the apical-posterior segments of the lower lobes. The lingula, the middle lobe, and the basal segments of the lower lobes can all be involved by bronchogenic spread. Concomitant pleural disease occurs in a minority of patients.

Cavitation is frequent, but even in the absence of cavitation, the diagnosis of tuberculosis should be considered when multifocal infiltrates are present. Atypical radiographic patterns, including isolated lower lobe infiltrates, are particularly common in elderly, debilitated patients. These patterns can complicate and delay diagnosis. Atypical radiographic features, including midlung or lower lung field infiltrates, hilar adenopathy, and absence of cavitation, can be seen in immunocompromised patients, especially in patients with AIDS [*see 7:XXXIII HIV and AIDS*].

Endemic Fungal Pneumonias

Endemic fungal pneumonias are acquired by inhalation of aerosolized particles (usually small respirable spores) rather than by microaspiration of pharyngeal organisms—the mechanism of infection for most pyogenic bacterial pneumonias. Severe cases of endemic fungal pneumonia usually involve multiple areas; blastomycosis and coccidioidomycosis often cause multiple dense alveolar infiltrates [*see 7:XXXVII Mycotic Infections*].

Melioidosis

Melioidosis is an indolent bacterial pneumonia caused by the gram-negative bacillus *Burkholderia pseudomallei*. It is most common in Southeast Asia and may appear in patients months or even years after they have migrated to nonendemic areas. Multifocal infiltrates are characteristic. The clinical illness resembles tuberculosis except that upper lobe predominance is not as striking in melioidosis.

Invasive Aspergillosis

An opportunistic fungal infection that is also acquired by inhalation, invasive aspergillosis occurs in patients who experience a decrease in the number or function of phagocytes.[8] Prolonged neutropenia from cytotoxic chemotherapy for acute leukemia is the most common predisposing factor. High-dose glucocorticoid therapy also predisposes to this infection, as do several immunosuppressive regimens used for solid-organ transplantation and, especially, for bone marrow transplantation. Clinical features include fever, dyspnea, pleuritic chest pain, hemoptysis, and hypoxemia. The chest radiograph shows characteristic multiple dense infiltrates, which may be nondescript, wedge shaped and peripheral, or even nodular. Individual lesions may cavitate, particularly as the number of neutrophils begins to rebound after chemotherapy. During this recovery time, the patient is at the highest risk for major hemoptysis [*see 7:XXXVIII Mycotic Infections in the Compromised Host*]. CT shows a characteristic pattern of nodules surrounded by ground-glass attenuation (halo sign).

Diagnosis is difficult because actual tissue invasion must be determined. However, any positive culture for *Aspergillus* from sputum or bronchoscopic specimens has high predictive value in the patient with profound neutropenia. Such cultures are valuable but less specific in transplant recipients or other moderately immunosuppressed patients and are often not helpful in patients with chronic bronchitis, who are frequently colonized with this organism. Negative cultures never rule out the diagnosis of invasive aspergillosis.

NEOPLASTIC DISEASES

Two neoplasms that present as focal infiltrates [*see* Focal Pulmonary Infiltrates, *above*] also commonly present as multifocal infiltrates: alveolar cell carcinoma[2] [*see Figure 2c*] and Hodgkin disease. Hodgkin disease, however, almost always has associated hilar and mediastinal adenopathy, whereas alveolar cell carcinoma almost never does. Lymphomatoid granulomatosis, an angiocentric T cell lymphoma of variable grade, usually presents as multiple pulmonary nodules [*see* Multiple Nodules and Masses, *below*] but can also present as multifocal pulmonary infiltrates.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Several noninfectious, nonneoplastic pulmonary diseases may produce multiple areas of lung infiltration.

Drug Reactions

Multifocal infiltrates can occur as a manifestation of drug toxicity, especially acute methotrexate, nitrofurantoin-related, and amiodarone-related pulmonary toxicity [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. With amiodarone-related lung disease, the infiltrates may be bilateral and more prominent in upper lung zones. The drug profile must be reviewed carefully in any patient with unexplained multifocal pulmonary infiltrates.

Simple Eosinophilic Pneumonia

Patients with simple eosinophilic pneumonia, often called Löffler syndrome, have nonproductive cough and fever [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. The chest radiograph shows patchy infiltrates, typically in the lower lung fields; the infiltrates resolve within 1 to 2 weeks. The main clue to the diagnosis is an increase in the number of peripheral eosinophils, often to 60% or more of the total white blood cell count. Bronchoalveolar lavage is seldom indicated, because patients are only mildly ill; when lavage is done, it reveals pulmonary eosinophilia. The main causes of this pneumonia are drugs and worms—usually nematodes such as *Ascaris* and *Strongyloides,* which migrate through the lung during one phase of infection [*see 7:XXXV Helminthic Infections*]. Drugs that cause the syndrome include but are not limited to carbamazepine, chlorpromazine, cocaine, imipramine, isoniazid, naproxen, nitrofurantoin, penicillins, sulfonamides, and tetracycline.

Tropical Pulmonary Eosinophilia

Tropical pulmonary eosinophilia, which is caused by immunologic hyperreactivity to microfilariae, is characterized by dyspnea, wheezing, and coughing, all of which are worse at night, as well as fever, weight loss, and fatigue [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. The chest radiograph reveals interstitial micronodular lesions that may be diffuse or multifocal. Peripheral blood eosinophilia is in excess of 3,000/mm^3; the serum IgE levels are extremely high, and antibodies to microfilariae are present. Two filarial parasites, *Wuchereria bancrofti* and *Brugia malayi,* are causative agents [*see 7:XXXV Helminthic Infections*]. Blood eosinophilia is less prominent in patients with chronic forms of the disease. Patients who have lived in areas where tropical pulmonary eosinophilia is endemic can present with restrictive or obstructive lung disorders that are the residua of episodes of tropical eosinophilia that occurred earlier in life.

Chronic Eosinophilic Pneumonia

Chronic eosinophilic pneumonia is more serious than simple eosinophilic pneumonia[9] [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. It primarily affects women between 20 and 40 years of age; however, it has been reported in patients of both sexes and of all ages. Most patients have moderate to severe illness lasting from 1 month to several months, with fever, night sweats, nonproductive cough, shortness of breath, and weight loss. About one third of patients have a history of asthma, and up to one half have some atopic history, including allergic rhinitis and nasal polyps. About one half of patients have wheezing as part of their clinical presentation, some for the first time. The chest radiograph may show dense peripheral infiltrates that are often referred to as reverse pulmonary edema or the photographic negative of pulmonary edema. When the clinical presentation and chest radiograph are highly typical, the diagnosis can be made clinically. About two thirds of patients have peripheral eosinophilia; a minority have an elevated serum IgE level. Bronchoalveolar lavage shows increased eosinophils and increased lymphocytes. Open lung biopsy, which is seldom required, shows lymphocytes and eosinophils in alveolar walls and spaces, as well as bronchiolitis obliterans in up to one third of specimens.

The etiology of chronic eosinophilic pneumonia is unknown. The disease may represent a hypersensitivity response to unknown antigens or may be an idiopathic immunologic activation.

Acute Eosinophilic Pneumonia

Idiopathic acute eosinophilic pneumonia (AEP) is another syndrome that presents as multifocal infiltrates[10] [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. Blood eosinophilia is usually absent, and the onset is acute, generally within several days. Patients may have a fever and severe hypoxemia and may progress to respiratory failure. There is usually no history of atopy. There are rales on physical examination, but there is rarely any wheezing. The prime diagnostic consideration is severe community-acquired pneumonia. The chest radiograph shows bilateral infiltrates that may progress to diffuse infiltrates. Unlike the case with chronic eosinophilic pneumonia, there is no tendency for peripheral subpleural localization. CT shows multifocal air-space opacities, often with an interstitial component. Increased septal markings (Kerley B lines) are seen in about one third of cases. Diagnosis of AEP is usually made by bronchoalveolar lavage, which shows over 25% eosinophils (mean, 37 ± 2.5). Culture for respiratory pathogens is negative.

The etiology of this immune disorder is unknown. In addition to the increase in pulmonary eosinophils, pathobiologic features include an increase in the number of helper T cells and neutrophils in the lung and elevated concentrations of interleukin-5 (IL-5) in lavage fluid.

Bronchiolitis Obliterans Organizing Pneumonia

Bronchiolitis obliterans alone is a distinct pathologic process in which distal airways are filled with plugs of loose connective tissue containing fibroblasts and inflammatory cells [*see 14:III Chronic Obstructive Diseases of the Lung*]. Clinical manifestations in patients with this pathology include severe airflow obstruction and poor response to bronchodilator therapy; chest radiographs show no infiltrates. Some cases are idiopathic, but others are associated with specific immunologic abnormalities, such as rheumatoid arthritis, graft versus host disease (GVHD) in bone marrow transplant recipients, and chronic rejection in lung transplant recipients.

Bronchiolitis obliterans organizing pneumonia (BOOP) is a pathologic entity that is also (and perhaps more descriptively) called cryptogenic organizing pneumonia [*see 14:V Chronic Diffuse Infiltrative Lung Disease*].[11] The lung parenchyma is involved, and alveolar spaces are filled with loose connective tissue that contains fibroblasts and mixed inflammatory cells. The process also involves contiguous distal airways, so the characteristic plugs of bronchiolitis obliterans are also seen. The chest radiograph often shows multiple dense peripheral infiltrates.[9] A less common pattern of idiopathic BOOP is a dense lobar infiltrate, which resembles bacterial pneumonia in radiographic appearance. Because of the long course of the illness, clinical suspicion often focuses on chronic pulmonary infections, such as blastomycosis or even alveolar cell carcinoma. The most common features on CT scanning, which shows the extent of disease better than a chest radiograph, include patchy bilateral air-space consolidation, small nodular opacities (sometimes with a reversed halo sign), and bronchial wall thickening and dilatation. In half of cases, these abnormalities are mainly subpleural; in the other half of cases, they are distributed throughout the affected lobes.[12]

The diagnosis of BOOP can often be made on transbronchial lung biopsy. In cases in which an open lung biopsy is needed, the video-assisted thoracoscopic technique has become the established procedure.

Allergic Granulomatosis and Angiitis

Also known as Churg-Strauss syndrome, allergic granulomatosis and angiitis is rare [*see 15:VIII Systemic Vasculitis Syndromes*].[13] Almost all patients with allergic granulomatosis and angiitis have a strong history of atopic allergy with preexisting bronchial asthma for an average of 5 to 10 years before diagnosis of the vasculitis. Patients present with fever, shortness of breath, and a variety of complaints related to skin and nerve involvement (e.g., purpura, painful skin nodules, skin infarction, footdrop, wristdrop, or painful neuropathy). Myocardial involvement occurs in a minority of patients but on occasion can dominate the clinical presentation. Wheezing is common. Clinically significant renal involvement is uncommon. In 25% of cases, the chest radiograph shows peripheral infiltrates or nodules that seldom cavitate. Pleural effusion, pericardial effusion, or both may occur with or without infiltrates. Churg-Strauss syndrome, though rare, most often occurs in patients with bronchial asthma. There is an association between Churg-Stauss syndrome and treatment of asthma with leukotriene inhibitors.[14] Onset of the vasculitis usually occurs as glucocorticoids are tapered during leukotriene inhibitor therapy, and it is possible that some of these cases represent occult Churg-Strauss disease that is initially suppressed with glucocorticoids and then flares with glucocorticoid withdrawal. However, at least one case has been reported in a patient with no recent glucocorticoid use.[15]

Other Noninfectious, Nonneoplastic Multifocal Disorders

Other immunologic lung diseases can produce multifocal infiltrates. Wegener granulomatosis is the most common lung vasculitis. The radiographic spectrum of illness includes infiltrates and nodules that are usually multiple and often cavitate [*see* Cystic and Cavitary Infiltrates, *below*].

Collagen vascular disease can be associated with multifocal lung involvement.[16] Pneumonitis associated with systemic lupus erythematosus (SLE) may be diffuse but may also appear as dense lower lobe infiltrates. Other clinical and serodiagnostic features of SLE are present. Systemic sclerosis and mixed connective tissue disease usually present as diffuse interstitial infiltrates, often with basilar predominance [*see 14:V Chronic Diffuse Infiltrative Lung Disease* and *15:IV Systemic Lupus Erythematosus*].

Lung disease associated with rheumatoid arthritis takes many forms, including pulmonary fibrosis, pleural effusions, and necrobiotic pulmonary nodules, the pathologic appearance of which is similar to that of subcutaneous rheumatoid nodules [*see 14:V Chronic Diffuse Infiltrative Lung Disease* and *15:II Rheumatoid Arthritis*]. Peripheral dense, masslike infiltrates may also occur. Some lesions may cavitate; such lesions are most common in men with high-titer rheumatoid factor and subcutaneous rheumatoid nodules.

Other noninfectious, nonmalignant disorders that can cause multifocal infiltrates include silicosis and other diseases caused by inorganic dust inhalation.[17] With extensive disease, progressive massive fibrosis may develop, producing large opacities. These opacities begin in peripheral areas of the lung and migrate centrally as lung volume is lost. The chest radiograph almost always shows a background of small nodules that are most prominent in the upper lobe. Patients have productive cough and dyspnea on exertion. The physiologic findings are typical of restrictive disease, but many patients also have airflow obstruction resulting from cigarette smoking or from airway damage caused by massive dust overload.

Table 3 Major Causes of True Segmental Infiltrates

Cause	*Examples*
Infectious	Bacterial pneumonia Tuberculosis Invasive aspergillosis, *Mucor* pneumonia, *Pseudomonas* pneumonia Bronchiectasis
Neoplastic	Primary lung cancer Carcinoid tumors Cylindromas, mucoepidermoid carcinomas
Noninfectious, nonneoplastic	Foreign body Amyloidosis and sarcoidosis Asthma with mucous plugging Allergic bronchopulmonary aspergillosis Pulmonary infarction Bronchial fracture Postoperative atelectasis

Caplan syndrome occurs in patients with rheumatoid arthritis who have been occupationally exposed to coal dust or other particulates, such as silica, asbestos, aluminum dust, and iron dust. The tendency to form large, masslike infiltrates is more pronounced in these patients than it is in other patients with rheumatoid arthritis. The infiltrates are multiple and may be more than 5 cm in diameter. They are often somewhat nodular and may cavitate and even calcify. Histologically, they resemble other rheumatoid necrobiotic nodules except that the foreign particulate matter can be easily demonstrated. The occupational history and the nodules and fibrosis evident on the chest radiograph are diagnostic clues.

In ankylosing spondylitis, a spondyloarthritis that leads to a stiff spine, calcification of spinal ligaments results in a characteristic radiographic appearance, the so-called bamboo spine [*see 15:III Seronegative Spondyloarthritis*]. Chest wall expansion is impaired, and lung volumes are somewhat reduced. Patients may develop characteristic dense fibrous or fibrobullous infiltrates that are often limited to the upper lung zones. These chest radiographic findings mimic those of tuberculosis, and cavitation may occur. Once cavities have developed, abnormal air spaces may become colonized with *Aspergillus* species, causing fungus balls and even a locally invasive disease termed chronic necrotizing aspergillosis. CT studies of groups of patients with ankylosing spondylitis have found interstitial lung disease that is more diffuse and not confined to the apexes of the lung.[18] The lung disease associated with ankylosing spondylitis does not progress to respiratory insufficiency. Dyspnea on exertion and nonproductive cough are the usual symptoms. Once *Aspergillus* colonization or infection has occurred, symptoms from the local fungal infection may predominate; such symptoms include hemoptysis, productive cough, and mild to moderate constitutional symptoms. Locally invasive *Aspergillus* infection may extend directly to the pleura but almost never spreads hematogenously to distant sites.

Exogenous lipoid pneumonia may also cause multifocal infiltrates. These infiltrates are dense and are usually seen in both lower lobes.[5]

True Segmental Infiltrates

A chest radiograph that shows nearly complete involvement of a single lung segment, especially if associated with volume

loss, should raise suspicion that there is disease in the bronchus or in the pulmonary artery supplying that segment [*see Table 3*].[19]

INFECTIOUS DISEASES

Common bacterial pneumonias caused by *S. pneumoniae* and *H. influenzae* are usually not truly segmental unless an entire lobe is involved. In contrast, involvement of one or more discrete segments, often at the lung bases, is common in *Mycoplasma* pneumonia [*see Figure 3a*].[1]

Three types of pneumonia exhibit a strong tendency for angioinvasion and cause dense, wedge-shaped peripheral infiltrates resulting from combined infection and infarction of a lung segment. Patients may have fever and purulent sputum from the infection, and they may have hemoptysis and pleural pain from lung infarction. *Aspergillus* and *Mucor* species are important causes of this syndrome in immunosuppressed patients, particularly patients with neutropenia or those receiving high-dose glucocorticoid therapy [*see 7:XXXVIII Mycotic Infections in the Compromised Host*], and *Pseudomonas* pneumonia is an important cause of this syndrome in debilitated patients [*see 7:X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas*].

Bronchiectasis is a disease of the airways that can present as a peripheral segmental infiltrate, from dilated bronchi filled with mucus and infiltrate, and as an associated volume loss. Air-fluid levels may be seen in dilated saccular airways. Because CT scanning is more sensitive and specific than chest radiography in documenting bronchiectasis, it has replaced bronchography in the diagnosis of this process. The usual clinical presentation is a chronic cough that produces purulent sputum [*see 14:III Chronic Obstructive Diseases of the Lung*].

Tuberculosis is a parenchymal lung disease but may also involve the airways heavily. Bronchiectasis is a common late sequela. Because it occurs most often in bronchi of the upper lobes, it is well drained; hence, little or no sputum is produced (so-called dry bronchiectasis). Hemoptysis, often massive, can occur in patients who have no history of chronic cough and sputum. True segmental infiltrates may be seen early or late in the course of this chronic infection.

Pulmonary aspergillosis is an uncommon and, usually, very late complication of AIDS. Endobronchial aspergillosis is one manifestation of AIDS that often presents as true segmental infiltrates of the involved segment or segments. Pulmonary aspergillosis in AIDS may also present as multiple nodules and single or multiple infiltrates.

NEOPLASTIC DISEASES

Lung tumors that arise in the lung parenchyma present as nodules or masses. Tumors that arise in central airways block a segment of the lung, leading to distal infection, atelectasis, or both; the infection or atelectasis conforms perfectly to the obstructed segment, lobe, or lung. Symptoms include cough and hemoptysis or purulent sputum, as well as fever from the postobstructive pneumonia.

Carcinoid Lung Tumors

Carcinoid tumors of the lung are low-grade adenocarcinomas [*see 12:VIII Lung Cancer*].[20] Most carcinoid tumors arise in central airways, so they present as segmental infiltrate or atelectasis more often than do primary lung cancers [*see Figure 3b*]. Carcinoid tumors grow more slowly and metastasize less often than primary lung cancers, and they are not related to cigarette smoking. Although the mean age of incidence is 55 years, carcinoid tumors are almost evenly distributed across adulthood; in contrast, the incidence of primary lung cancer increases markedly with age. Despite their relative infrequency, carcinoid tumors account for a high percentage of lung tumors in patients in their third and fourth decades but only a tiny percentage of tumors in patients of advanced age.

Other Lung Tumors

Several other lung tumors typically grow in central airways and also present as segmental infiltrate, atelectasis, or both. Adenoid cystic carcinomas, known as cylindromas, and mucoepidermoid carcinomas are low-grade cancers of the mucous glands that usually occur in the trachea or central bronchi. The clinical presentation is similar to that of carcinoid tumors, but mucous gland tumors occur less than one tenth as often as carcinoid tumors. The histopathologies of cylindromas, mucoepidermoid carcinomas, and carcinoid tumors differ.

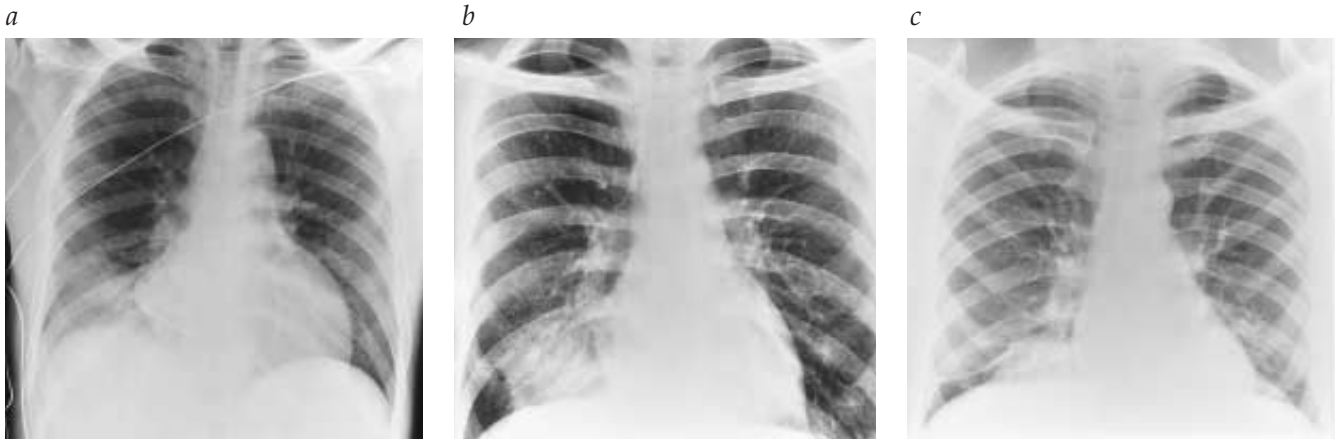


Figure 3 **(*a*) Unlike common bacterial pneumonias, *Mycoplasma* pneumonia frequently involves one or more specific segments. In this chest radiograph of a 42-year-old man, *Mycoplasma* pneumonia involves the anterior basal segment of the right lower lobe. (*b*) This chest radiograph of a 24-year-old man demonstrates a carcinoid tumor obstructing the medial basal segment of the right lower lobe. (*c*) Allergic bronchopulmonary aspergillosis may cause a segmental infiltrate or segmental atelectasis. In this chest radiograph of a 40-year-old man, the infiltrate revealed in the medial basal segment of the right lower lobe is the result of allergic bronchopulmonary aspergillosis.**

Benign lung tumors are uncommon. Most are hamartomas, which present peripherally as nodules [*see* Single Small Nodules, *below*]. Benign tumors presenting in the central airways are extremely rare; most are fibromas or lipomas.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Foreign Bodies

Aspiration of a foreign body is an important cause of a segmental infiltrate. However, only 10% of foreign bodies are radiodense. When they are, they can often be identified immediately on the radiograph. When the foreign body is invisible on the radiograph, the diagnosis requires a bronchoscopic examination. Bronchoscopy should be performed if the clinician can combine the clinical history (e.g., altered consciousness, choking spell, irritative cough, repeated infections, or hemoptysis) with the finding of a true segmental infiltrate on the chest radiograph.

Amyloidosis and Sarcoidosis

Pulmonary amyloidosis exhibits many different radiographic patterns, including diffuse infiltrates and single or multiple nodules or masses. Amyloid deposition in a central airway may cause obstruction, leading to segmental infiltrates. Sarcoidosis also displays many patterns [*see 14:V Chronic Diffuse Infiltrative Lung Disease*]. It commonly presents as diffuse interstitial disease with or without enlarged hilar and mediastinal lymph nodes, but it also has many uncommon presentations, including multiple nodules and endobronchial sarcoidosis with a segmental infiltrate.

Asthma

Simple mucous plugging during an acute exacerbation of asthma can lead to a segmental or lobar infiltrate and atelectasis. This pattern frequently occurs as the patient is improving, perhaps because hyperinflation is resolving and the airways are no longer so widely dilated. Bronchoscopy is not indicated and may be dangerous during an acute exacerbation [*see 14:II Asthma*].

Allergic Bronchopulmonary Aspergillosis

Allergic bronchopulmonary aspergillosis, which is also associated with asthma, is a hypersensitivity disease that primarily affects the central bronchi [*see 14:II Asthma* and *7:XXXVIII Mycotic Infections in the Compromised Host*]. Immediate hypersensitivity and delayed hypersensitivity are involved in pathogenesis. Onset of disease occurs most often in the fourth and fifth decades, and virtually all patients have long-standing atopic asthma. Even those few patients who do not have a history of documented asthma exhibit airflow obstruction when they present with this disorder. The typical patient has a long history of intermittent wheezing, after which the illness evolves into a more chronic and more highly symptomatic disorder with fever, chills, pulmonary infiltrates, and productive cough.

The chest radiograph may show a segmental infiltrate or segmental atelectasis, most commonly in the upper lobes [*see Figure 3c*]. One infiltrate may clear, only to be followed by another infiltrate in a different location. When the patient presents with typical symptoms, the branching, fingerlike shadows from mucoid impaction of dilated central bronchi are pathognomonic of allergic bronchopulmonary aspergillosis. CT scans showing typical central bronchiectasis are highly supportive of the diagnosis of allergic bronchopulmonary aspergillosis.[21]

Pulmonary Infarction

Obstruction of a segmental artery by thromboembolism can cause pulmonary infarction [*see 1:XVIII Venous Thromboembolism*]. The infiltrate is often peripheral and wedge shaped and occasionally has a clearly defined central limit that forms a pleurally based truncated cone, the so-called Hampton hump. Dyspnea, pleuritic chest pain, and, occasionally, hemoptysis are the clinical features of pulmonary infarction. An associated pleural effusion is commonly seen.

Bronchial Fracture

Segmental, lobar, or lung atelectasis can occur after trauma as a result of displacement of the fractured ends of bronchi. Pneumothorax, pneumomediastinum, and subcutaneous emphysema are often seen in this clinical setting.

Postoperative Atelectasis

Segmental or lobar infiltrates are common after surgery. In some instances, mucous plugging produces obstruction with volume loss. More commonly, particularly after cardiac surgery, atelectasis of the left lower lobe without bronchial obstruction occurs, in association with pleural effusion and diaphragmatic elevation.

Table 4 Major Causes of Cystic and Cavitary Pulmonary Infiltrates

Cause	*Examples*
Infectious	Necrotizing bacterial pneumonia Tuberculosis Blebs and bullae with secondary infection Other chronic suppurative infections Parasitic infections
Neoplastic	Primary lung cancer Metastatic lung cancer
Noninfectious, nonneoplastic	Wegener granulomatosis Sarcoidosis Rheumatoid nodules Pulmonary laceration

Cystic and Cavitary Infiltrates

Cavitation of the lung results from a suppurative infection, a rapidly growing neoplasm, or a destructive immune process such as vasculitis [*see Table 4*]. As with other focal and multifocal infiltrates, infections and neoplasms are most common and must be considered first.

INFECTIOUS DISEASES

Pyogenic bacteria can cause necrosis and thus cavitation of the lung. *S. pneumoniae* pneumonia rarely cavitates; when cavitation occurs, it usually indicates a mixed infection with anaerobic bacteria. Gram-negative aerobic pneumonias, including those caused by *K. pneumoniae* or *P. aeruginosa*, cavitate more frequently. A gram-negative pneumonia may produce a large, dense infiltrate, with multiple cavities appearing as the illness progresses. Staphylococcal pneumonia acquired from microaspiration may complicate influenza or may occur de novo [*see 7:I Infections Due to Gram-Positive Cocci*]. Multiple small cavities are frequent. Pyemic staphylococcal infection, often caused by right-sided endocarditis, typically causes scattered, round infiltrates, which often enlarge and then cavitate. Parapneumonic effusions and

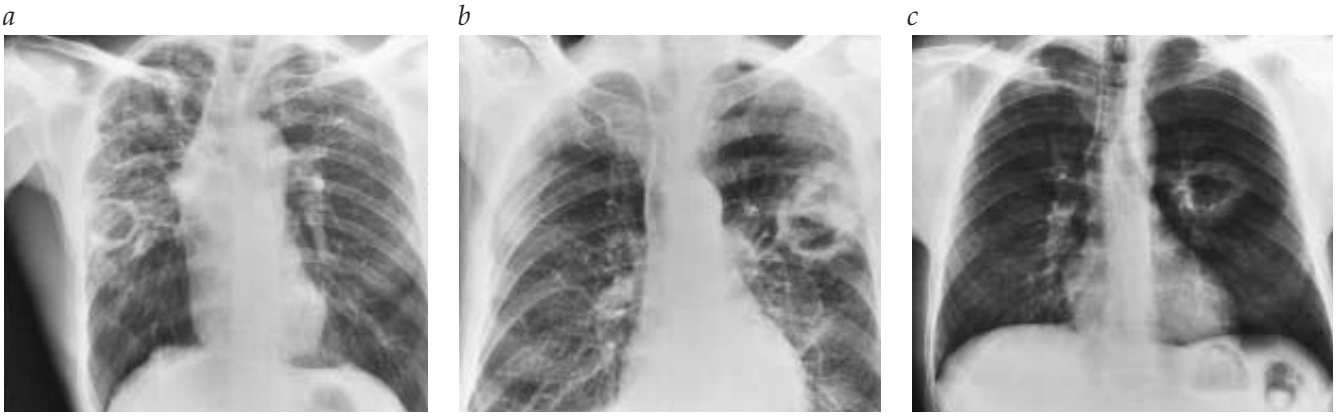


Figure 4 **(*a*) This chest radiograph of a 40-year-old man shows a large cavity and extensive infiltrate in the right upper lobe, caused by tuberculosis. (*b*) Central necrosis of a large cell carcinoma in a male smoker who is 65 years of age has left a large, thick-walled cavity in the upper lobe, as this chest radiograph reveals. (*c*) In 70% of cases of Wegener granulomatosis, the chest radiograph reveals that at least one of the lesions is cavitary. A large, thick-walled cavity caused by Wegener granulomatosis can be seen in the left upper lobe of this 37-year-old man.**

empyema are fairly common with either form of staphylococcal pulmonary infection.

A mixed anaerobic infection often produces a single, rounded, putrid lung abscess, which can be as large as 10 cm in diameter and frequently has an air-fluid level. Mixed infections may also cause a focal infiltrate with multiple small cavities. In lung gangrene (a rare entity), vascular compromise, focal vasculitis, and thrombosis of the involved lobe occur. The entire lobe cavitates, leaving chunks of necrotic lung floating in liquid pus. Anaerobic bacteria are probably involved in the etiology of most cases of lung gangrene, either alone or as part of a mixed infection with aerobic organisms. Diagnosis of this entity is based on the radiographic findings. Other infections complicated by cavitary infiltrates include *Pneumocystis* infections, septic thromboembolism, and bronchiectasis.

Blebs and bullae are thin-walled structures that may be congenital but usually result from emphysema. When bacterial superinfection is present, these cystic spaces may develop air-fluid levels and surrounding infiltrates. In the absence of old chest radiographs, it may be difficult to differentiate true cavitation from superinfection of bullae, or so-called bullitis.

Suppurative Infections

Chronic suppurative infections typically cause necrosis. Tuberculosis usually involves the upper lobes [*see 7:II Tuberculosis*]. Cavities are generally small to medium in size, have thick walls, and do not exhibit air-fluid levels [*see Figure 4a*]. There are often associated nodular infiltrates and fibrosis. Atypical mycobacterial infections, especially those caused by *M. avium* complex, often present as larger and more dramatic cavities, partly because such infections tend to occur in patients with underlying bullous lung disease. Infected, preexisting, abnormal air spaces caused by the underlying disease are difficult to distinguish from new necrosis of the lung caused by *M. avium* complex.

Histoplasmosis is a granulomatous infection that rarely cavitates in normal hosts. Blastomycosis and coccidioidomycosis have a mixed granulomatous and pyogenic histopathology, and lung necrosis is more common [*see 7:XXXVII Mycotic Infections*]. *Aspergillus* and *Mucor* infections lead to lung necrosis and infarction and frequently cavitate [*see 7:XXXVIII Mycotic Infections in the Compromised Host*]. *Nocardia* infections are subacute but also suppurative and may present as single or multiple cavitary lesions. Individual lesions tend to be round and well circumscribed; air-fluid levels are common [*see 7:IV Infections Due to Gram-Positive Bacilli*].

Parasitic Infections

Parasitic infections with cavities include echinococcal cysts, which are located in the lower lobes; the cysts are thin walled and have an irregular liquid level caused by collapse of the cyst wall [*see 7:XXXV Helminthic Infections*].[22] Although uncommon in the United States, echinococcosis is very common in parts of Africa and Asia. Paragonimiasis is most often acquired in East and Southeast Asia; infiltrates, pleural effusions, and impressive eosinophilia are sometimes seen [*see 7:XXXV Helminthic Infections*]. A more chronic form of the disease may present as fibrous nodules. Small, thin-walled cysts and cavities can also be created when small nodules soften and then cavitate as the contents are expectorated. Cysts and cavities average 2 to 4 cm in diameter.

Amebic lung abscesses, which are also uncommon in the United States, usually occur in the right lower lobe and generally have thick, irregular walls; they are frequently associated with hepatic abscesses.

NEOPLASTIC DISEASES

Any rapidly growing primary lung cancer can cavitate [*see 12:VIII Lung Cancer*]. Squamous cell carcinoma cavitates most frequently; adenocarcinoma and large cell cancers cavitate less frequently [*see Figure 4b*]; and despite having rapid growth rates, small cell cancers cavitate least often. Cavities are thick walled, as would be expected, because they result from central necrosis of a growing mass.

Hematogenous metastases can also cavitate. Cavitation has been reported in many different types of metastatic disease but occurs more frequently in squamous cell cancers that originate in the head and neck, cervix, or skin than in adenocarcinomas.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Cavitation in noninfectious, nonneoplastic disease is usually caused by a destructive immunologic process, such as vasculitis or rheumatoid necrobiosis.

Wegener Granulomatosis

Wegener granulomatosis is the most common lung vasculitis and is characterized by necrotizing granulomas of the upper and lower respiratory tracts, necrotizing glomerulonephritis, and

other features of systemic vasculitis [*see 15:VIII Systemic Vasculitis Syndromes*]. The mean age at onset is 40 years, but Wegener granulomatosis may occur at any age; it is slightly more prevalent in men. Clinical features vary greatly and relate to the sites of involvement. Systemic symptoms include fever, malaise, and weight loss. Upper airway findings include rhinorrhea, sinus pain and drainage, nasal and nasopharyngeal ulcers, and otitis media. Respiratory symptoms include cough, dyspnea, and hemoptysis. Skin, eye, or joint involvement is common, and renal involvement is extremely common. The chest radiograph or CT scan shows nodules or infiltrates, which are multiple about 90% of the time. About 70% of the time, at least one lesion is cavitary [*see Figure 4c*].[23]

Other Noninfectious, Nonneoplastic Multifocal Disorders

Other immunologically mediated processes such as sarcoid and rheumatoid nodules can cavitate [*see* Multiple Nodules and Masses, *below*]. Uncommonly, closed chest trauma results in one or more cysts in the lung that can be air filled or can be completely or partly filled with blood.

Single Small Nodules

Single small nodules, often called coin lesions, are usually primary lung cancers or granulomas [*see Table 5*].[24] It is important to detect and thoroughly evaluate such lesions because over 90% of cancers discovered at this stage can be cured. In general, given the high frequency of lung cancer, nodules—even small nodules detected only by CT—must be assumed to be malignant and must be removed unless they can be proved to be benign; unless they can be assumed to be benign on the basis of their having remained unchanged or on the basis of the young age of the patient; or unless the patient cannot tolerate a surgical procedure.

A benign etiology can be assumed if a chest radiograph taken 2 or more years earlier shows the lesion to have been the same size as or larger than it is currently. Such a situation could arise if the lesion went unrecognized on the initial film. (Unless the patient is younger than 35 years, it is inappropriate simply to follow a coin lesion.) There are also classic benign patterns of calcification that obviate further assessment of single small nodules. For granulomas, such patterns include dense, perfectly central targets of calcium; ring calcification; and solid, dense calcification of the whole nodule. For pulmonary hamartomas, patterns include the so-called popcorn calcification pattern. (It should be noted that small specks of calcium and eccentric clumps of dystrophic calcium are often seen in malignant tumors). Thin-section CT with contrast enhancement is another way to prove that a nodule is benign. Lesions of about –20 Hounsfield units that do not enhance with contrast are always benign; lesions that do enhance are either tumors or active granulomas.[25] A fine-needle biopsy may provide a specific benign diagnosis by histopathologic examination, cytologic examination, or culture of a microorganism.

If a patient with a pulmonary nodule is younger than 35 years, the chance of malignancy is low enough to justify serial follow-up every 6 months for 2 years. For patients of any age with a proven malignancy or a growing nodule and for all patients older than 35 years in whom the nodule cannot be proved benign, the lesion should be resected unless the patient cannot tolerate the required surgery. The patient's tolerance is largely determined by cardiopulmonary reserve and associated illnesses. Video-assisted thoracoscopic surgery decreases the morbidity of nodule resection.

Table 5 Major Causes of a Single Small Nodule

Cause	*Examples*
Infectious	Granuloma Parasites
Neoplastic	Primary lung cancer Hamartoma Carcinoid Metastasis
Noninfectious, nonneoplastic	Arteriovenous malformation Bronchogenic cysts Intrapulmonary lymph node Rounded atelectasis

INFECTIOUS DISEASES

Most infectious lung nodules are granulomas. Histoplasmomas are by far the most common, especially in the vast areas of the central United States drained by the Ohio and Mississippi rivers. Histoplasmomas rarely cause harm; the problem they pose, however, is in proving that they are not cancerous [*see above and Figure 5a*]. In the deserts of the southwestern United States, coccidioidomycosis is a common cause of a peripheral nodule [*see 7:XXXVII Mycotic Infections*]. Tuberculomas are uncommon in the United States but are frequent in countries with a high incidence of tuberculosis [*see 7:II Infections Due to Mycobacteria*]. Cryptococcomas can be small nodules or large masses [*see* Large Masses, *below, and 7:XXXVIII Mycotic Infections in the Compromised Host*].

A rare cause of a lung nodule that is almost always discovered after resection for presumed lung cancer is infection with *Dirofilaria immitis,* the dog heartworm [*see 7:XXXV Helminthic Infections*].[26] In areas of the world endemic for *D. immitis*, echinococcus is a common cause of a solitary pulmonary nodule.[22]

A lesion cannot be presumed to be benign just because there is immunologic evidence of infection by a particular pathogen remote to the lung, such as a positive skin test or positive serodiagnostic titers. To exclude the diagnosis of cancer, the organism must be identified directly or by culture from the sputum or from material obtained by needle aspiration.

NEOPLASTIC DISEASES

Lung cancers of each cell type can produce solitary nodules. This presentation is most characteristic of adenocarcinoma [*see Figure 5b*], but squamous cell, large cell, small cell, and even alveolar cell cancers can present as nodules. A single nodule is seldom a metastatic lesion. About 25% of carcinoid tumors present as solitary nodules.

Hamartomas are the most frequent benign tumors presenting as a nodule; they constitute 5% of all nodules. A nodule with scattered clumps of calcium throughout the whole lesion, termed popcorn calcification, is very likely a hamartoma. Unless such calcification is evident, hamartomas are often resected when cancer cannot otherwise be ruled out.[27] Other benign pulmonary tumors are very rare and are much more commonly present either intratracheally or intrabronchially.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Nodules are rarely noninfectious and nonneoplastic. Bronchogenic cysts[28] and arteriovenous malformations can present as a single nodule [*see Figure 5c*]. Pulmonary sequestrations are usually larger [*see* Large Masses, *below*]. Rheumatoid nodules, We-

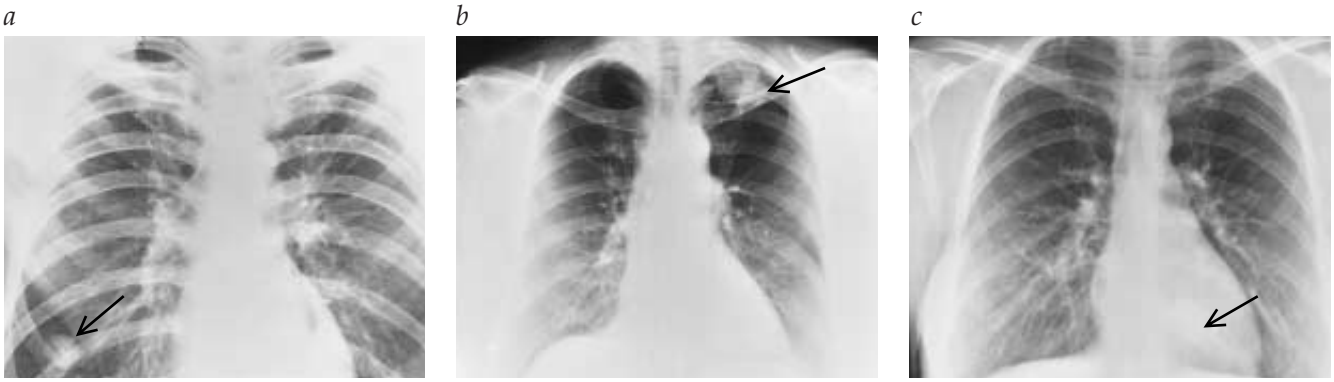


Figure 5 **(*a*) A histoplasmoma presents as a solitary nodule in the upper right lower lobe (arrow) in this chest radiograph of a 50-year-old man. (*b*) This solitary nodule in the apex of the left lung (arrow) proved to be an adenocarcinoma. The patient is a 63-year-old woman who has a history of heavy smoking. (*c*) In this chest radiograph of a woman who is 24 years old, the solitary nodule of a bronchogenic cyst can be seen in the left lower lobe, behind the cardiac shadow (arrow).**

gener granulomatosis, and pulmonary amyloidosis are rare disorders that seldom present as a single nodule.

Intrapulmonary lymph nodes can also present as small, solitary lung nodules. In one series, 17 of 96 patients who underwent excision of well-circumscribed peripheral pulmonary nodules had this pathology.[29] All the intrapulmonary lymph nodes were located within 20 mm of a visceral pleural surface and were in the lower lobes or middle lobe. They could not be distinguished radiographically from neoplasm or granuloma; however, improved radiologic image quality and use of CT will likely increase the number of intrapulmonary lymph nodes detected as small nodules.

A relatively uncommon cause of a solitary lesion is rounded atelectasis, which can produce a lesion of up to 5 cm in diameter. Rounded atelectasis is always adjacent to the pleura and may have a comet-tail appearance, with the "tail" pointing toward the hilum. The nodular density represents the curling up of atelectatic lung tissue adjacent to a chronic pleural process. A more common pseudotumor is found with pleural effusions caused by congestive heart failure. Collections of liquid appear as rounded lesions on the posteroanterior radiograph, most often in the major fissure. Lateral films often show fissural loculation of liquid.

Large Masses

Neoplasms are the most common cause of lesions larger than 6 cm in diameter. Other conditions only rarely cause such large masses [*see Table 6*].

Table 6 Major Causes of Large Pulmonary Masses (> 6 cm)

Cause	*Examples*
Infectious	Bacterial pneumonia (round pneumonia) Blastomycosis Cryptococcosis Invasive aspergillosis, mucormycosis
Neoplastic	Primary lung cancer Alveolar cell carcinoma Lymphoma
Noninfectious, nonneoplastic	Sequestration Bronchogenic cyst Progressive massive fibrosis

INFECTIOUS DISEASES

Bacterial pneumonia occasionally presents as a round infiltrate. An anaerobic lung abscess can also produce a homogeneous round infiltrate before cavitation. This can occur when the bronchi associated with an anaerobic lung abscess are occluded, thereby preventing pus from escaping and air from entering.

Cryptococcal infection can produce a solitary mass of up to 10 cm in diameter, often with few inflammatory symptoms. Blastomycosis, however, is the fungal infection that most frequently causes a rounded mass. It is often found in the superior segment of the lower lobe, projecting over the hilum on a standard posteroanterior chest radiograph. Blastomycosis is one of the more common benign conditions that are discovered at thoracotomy for presumed lung cancer [*see 7:XXXVII Mycotic Infections*].

Invasive pulmonary aspergillosis[8] and pulmonary mucormycosis occur in patients with neutropenia or in patients with depressed neutrophil function, usually caused by high-dose glucocorticoid therapy. Although each of these diseases most often presents as multifocal infiltrates [*see* Multifocal Pulmonary Infiltrates, *above*], these entities can also produce a focal infiltrate or a large, rounded mass [*see Figure 6a*].

NEOPLASTIC DISEASES

Large masses are most often large cell carcinoma [*see Figure 6b*] but are not uncommonly adenocarcinoma or squamous cell carcinoma. Alveolar cell carcinoma commonly presents as a slowly growing mass, often occupying most of a lobe or lung. Hodgkin and non-Hodgkin lymphoma may present as masses [*see 12:XI Lymphomas*]. The former, but not the latter, almost always displays hilar adenopathy, mediastinal adenopathy, or both.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Bronchopulmonary sequestration may be intralobar or extralobar and typically presents as a large mass with well-defined edges [*see Figure 6c*]. Intralobar sequestrations are located in the posterior basal segment adjacent to the diaphragm, and two of three sequestrations are located on the left. Extralobar sequestrations are contiguous with the diaphragm, and 90% are located on the left; they are not as limited as intralobar sequestrations to the posterior segment.

Lipoid pneumonia and pulmonary contusion usually present as infiltrates [*see* Focal Pulmonary Infiltrates, *above*] but can sometimes mimic round masses. Bronchogenic cysts can also present as large masses.[28] Progressive massive fibrosis in silicosis

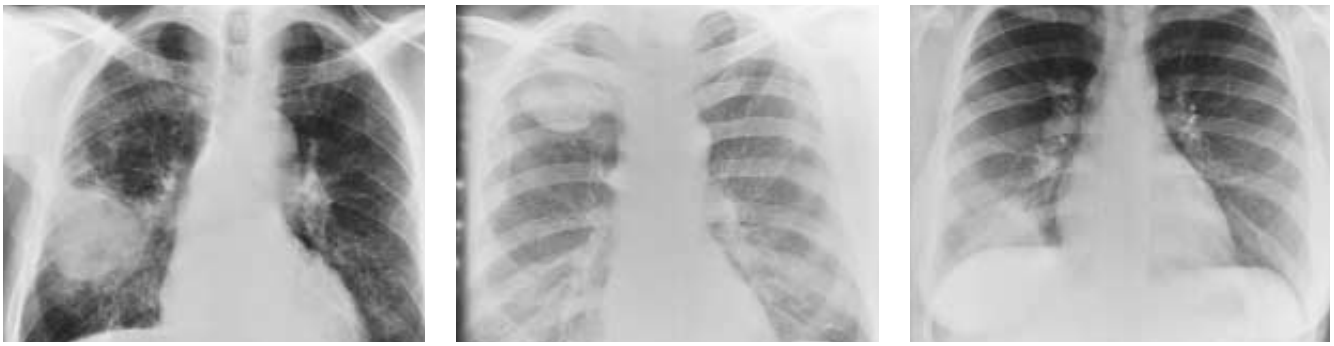


Figure 6 **(*a*) Mucor infection caused the large dense mass revealed in the right lung on this chest radiograph. The patient is a 74-year-old man who has had prolonged neutropenia, caused by myelofibrosis. (*b*) A large mass, the most common presentation of large cell carcinoma, can be seen in the right upper lobe in the chest radiograph of a male smoker who is 66 years of age. (*c*) This chest radiograph of a 31-year-old woman shows a large mass in the right lower lobe adjacent to the diaphragm. The mass is an intralobar sequestration.**

and coal workers' pneumoconiosis can have the appearance of a large mass.[17]

Multiple Nodules and Masses

Hematogenous metastases are the most likely cause of multiple nodules, especially if the patient is not febrile and the nodules vary widely in size. Because hematogenous spread to the lung usually occurs over time, different nodules have shorter or longer growth intervals before discovery. Other causes of multiple nodules are uncommon [*see Table 7*]. The differential diagnosis for multiple nodules is different from that for single small nodules or large masses.

INFECTIOUS DISEASES

Endocarditis and Endovascular Infections

Pyemic abscesses[30] are most commonly caused by right-sided endocarditis [*see 7:XVIII Infective Endocarditis*] or by other endovascular infections; *S. aureus* is the most common pathogen. Most patients have high fever and shaking chills. The white blood cell count is usually elevated and shifted toward neutrophils and neutrophil precursors, and blood cultures are usually positive. If the cause is endocarditis, a tricuspid murmur may be heard, and brisk venous pulsations in the neck (large V waves with rapid Y descent) may provide evidence of a leaky valve. Transthoracic echocardiography reveals vegetations in about half of cases; transesophageal echocardiography reveals vegetations in more than 80% of cases. The chest radiograph demonstrates multiple small nodules, which are usually 1 to 3 cm in diameter [*see Figure 7a*]. With time, the lesions may grow and cavitate. Subpleural nodules are frequently associated with pleural effusions, which may be sterile but often are highly inflammatory parapneumonic effusions or even frank empyema.

Other Infections

Multiple pulmonary nodules may also be caused by subacute or chronic infections. Melioidosis, for example, can present as multiple nodules [*see* Multifocal Pulmonary Infiltrates, *above*]. Patients often have chronic symptoms that are similar to those of tuberculosis. Nodules may be grouped in one area of the lung and may grow, coalesce, and cavitate as the disease progresses. *Nocardia* infections can also produce multiple nodules, which frequently cavitate [*see 7:IV Infections Due to Gram-Positive Bacilli*]. Over half of the patients who have *Nocardia* infections are immunosuppressed; sometimes, low-grade fever is the only symptom. Other patients have productive or nonproductive cough. Patients may also have associated subcutaneous nodules or brain abscesses, both of which are the result of hematogenous spread from the lung.

Cryptococcosis,[31] coccidioidomycosis, and paragonimiasis can produce scattered nodules. Often, the disease is relatively inactive and the only clinical finding is an abnormal chest radiograph. Individual lesions may soften and then be evacuated by expectoration of central necrotic areas, leaving thin-walled cavities. Histoplasmosis is a very frequent cause of multiple asymptomatic lung nodules. There are usually fewer than five nodules, and the lesions rarely if ever cavitate. Calcification is common but takes many years—at least 10 years in adults and somewhat less time in children. Dense, nearly total calcification is generally a strong indicator of a benign lesion. Small amounts of eccentric calcification, however, are frequently seen in rapidly growing primary and metastatic tumors, perhaps as a result of necrosis and subsequent calcification.

Extensive echinococcal infection can produce multiple large masses in the lung.[22] One or more of the lesions may be cavitary. The patient may be afebrile, and the initial diagnosis is often metastatic cancer. However, a CT scan will show that the density of echinococcal lesions in Hounsfield units is close to zero. In this way, metastatic cancer can be eliminated as a diagnosis. Other echinococcal lesions may be found in the liver, kidneys, or

Table 7 Major Causes of Multiple Pulmonary Nodules and Masses

Cause	*Examples*
Infectious	Septic emboli Multiple granulomas Fungal and parasitic infections
Neoplastic	Metastatic carcinoma Lymphoma
Noninfectious, nonneoplastic	Arteriovenous malformations Wegener granulomatosis Sarcoidosis and amyloidosis Rheumatoid nodules

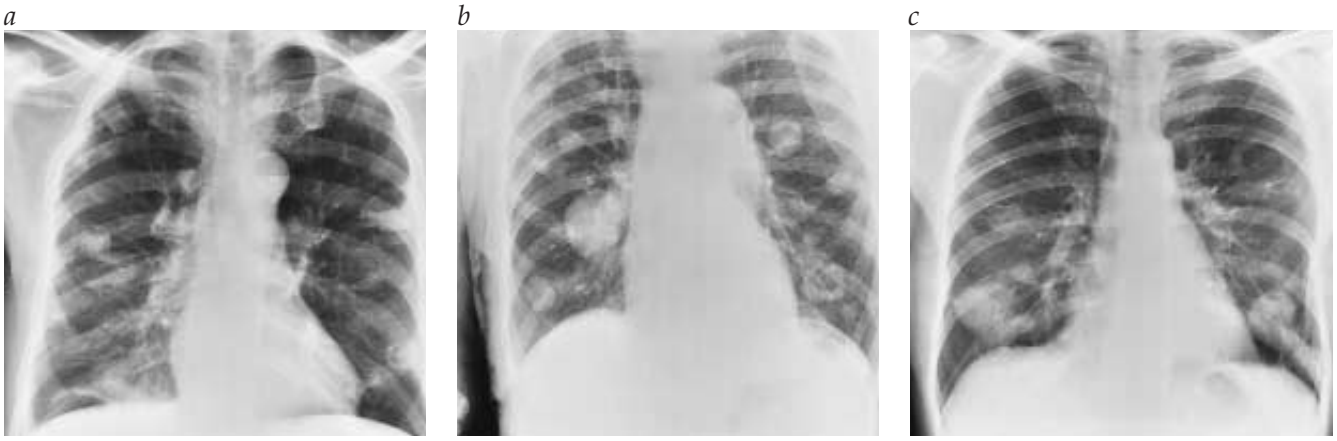


Figure 7 **(*a*) Multiple nodules, some with cavitation and some with air-fluid interfaces, are visible in this chest radiograph of a male I.V. drug abuser who is 34 years of age. The nodules are the result of tricuspid valve endocarditis caused by *Staphylococcus aureus*. (*b*) This chest radiograph of a 37-year-old woman with primary laryngeal cancer reveals metastatic squamous cell carcinoma. As is characteristic, the nodules vary widely in size, and some demonstrate cavitation. (*c*) This radiograph shows a 49-year-old man with Wegener granulomatosis, characterized by multiple nodules that vary widely in size and that have cavitated in some cases.**

other viscera. Surgery is impossible because of the extent of the disease (e.g., bilateral and extrapulmonary), but prolonged courses of antiparasitic therapy can be surprisingly effective.

NEOPLASTIC DISEASES

Metastatic carcinoma is the single most important cause of multiple pulmonary nodules [*see Figure 7b*]. Diagnosis is made by presumption if there are extensive metastases and an obvious primary cancer or by transthoracic needle aspiration biopsy if there is clinical uncertainty.

Lymphomatoid granulomatosis is a rare pulmonary disorder that usually presents as multiple pulmonary nodules or as multiple or diffuse infiltrates.[32] The radiographic presentation mimics metastatic carcinoma or pulmonary vasculitis—particularly Wegener granulomatosis. Immunohistochemical methods have established that lymphomatoid granulomatosis is an angiocentric T cell lymphoma of variable grade. The disease is extranodal, involving the lung and, in a substantial minority of cases, the skin and the central nervous system. It is most common in men in their sixth and seventh decades. In untreated cases, survival is less than 1 year, but survival may be prolonged by the use of combination therapy with prednisone and cyclophosphamide or, if that fails, by more aggressive combination chemotherapy as employed against T cell lymphoma.

Posttransplant lymphoproliferative disorders (PTLDs) often present as multiple well-circumscribed pulmonary nodules.[33] These disorders are related to infection with Epstein-Barr virus and may have polymorphic or monomorphic populations of lymphocytes, the latter being more aggressive and frankly malignant. AIDS-related lymphoma is a non-Hodgkin B cell lymphoma, also related to Epstein-Barr virus infection; it is similar to the monoclonal type of PTLD and often presents as multiple pulmonary nodules. Patients with other types of intensive immunosuppression also may occasionally develop multinodular pulmonary lymphoma. A feature that is shared by all of these disorders is the absence of associated mediastinal adenopathy. Pulmonary Kaposi sarcoma in patients with AIDS and other immunosuppressive conditions can also present as multiple pulmonary nodules.[34]

Two other related neoplastic entities, found exclusively in women, produce multinodular lung lesions that are most often asymptomatic: multiple pulmonary fibroleiomyomatous hamartomas, which are thought to arise within the lung, and so-called benign metastasizing uterine leiomyomas, which are thought to arise in the uterus. The two neoplasms are identical histopathologically and have often been shown to have estrogen receptors. These hamartomas and leiomyomas are most often benign and may regress spontaneously. Hysterectomy, oophorectomy, and antihormone therapy have been tried in rare cases of progression, without much effect.

NONINFECTIOUS, NONNEOPLASTIC DISORDERS

Only a few noninfectious, nonneoplastic conditions present as multiple pulmonary nodules. Pulmonary arteriovenous malformations are multiple in as many as one third of cases; half of these are associated with hereditary hemorrhagic telangiectasia (Osler-Weber-Rendu disease) [*see 5:XIII Hemorrhagic Disorders*]. Sometimes, the chest radiograph is so suggestive of this disorder that it is virtually diagnostic, especially when a feeding artery and a draining vein are visualized. To prove the diagnosis and to identify all the lesions, however, pulmonary angiography is always necessary. Patients may present with a number of clinical syndromes, including hemoptysis, hypoxemia, congestive heart failure, and systemic embolization. Invasive radiologic procedures have been developed to embolize the fistulas. These procedures are usually preferable to resection. Even if a large symptomatic lesion is resected, smaller lesions may grow and become symptomatic, leading to a series of resections and extensive loss of lung.

Wegener granulomatosis often causes multiple lung nodules [*see* Cystic and Cavitary Infiltrates, *above, and Figure 7c*]. Rheumatoid nodules [*see* Multifocal Pulmonary Infiltrates *and* Cystic and Cavitary Infiltrates, *above*] should also be included in the category of multiple pulmonary nodules. The infiltrates are multiple, are frequently discrete and well circumscribed (i.e., nodular), and sometimes cavitate. Temporal arteritis can produce multiple nodular lesions with fever, cough, and weight loss and should be considered in elderly patients. Pulmonary hyalinizing granuloma is an obscure cause of multiple bilateral pulmonary nodules.[35]

Sarcoidosis[36] and amyloidosis may produce multiple well-defined nodules, mimicking metastatic carcinoma. The nodules may be the only manifestations of these diseases. Histopatholog-

ic diagnosis is required, either by transthoracic needle aspiration biopsy or by open lung biopsy. There is no specific treatment for nodular amyloidosis, which has a better prognosis than other forms of pulmonary amyloidosis (e.g., diffuse interstitial infiltrates or endobronchial amyloidosis).

Additional Information

Additional information on focal and multifocal lung disease may be obtained from the National Heart, Lung, and Blood Institute (http://www.nhlbi.nih.gov) and the American Thoracic Society (http://www.thoracic.org).

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Gharib AM, Stern EJ: Radiology of pneumonia. Med Clin North Am 85:1461, 2001

2. Sabloff BS, Truong MT, Wistuba II, et al: Bronchioalveolar cell carcinoma: radiologic appearance and dilemmas in the assessment of response. Clin Lung Cancer 6:108, 2004

3. Coche E, Verschuren F, Hainaut P, et al: Pulmonary embolism findings on chest radiographs and multislice spiral CT. Eur Radiol 14:1241, 2004

4. Choi YW, Munden RF, Erasmus JJ, et al: Effects of radiation therapy on the lung: radiologic appearances and differential diagnosis. Radiographics 24:985, 2004

5. Baron SE, Haramati LB, Rivera VT: Radiological and clinical findings in acute and chronic exogenous lipoid pneumonia. J Thorac Imaging 18:217, 2003

6. Reittner P, Ward S, Heyneman L, et al: Pneumonia: high-resolution CT findings in 114 patients. Eur Radiol 13:515, 2003

7. Koh DM, Bell JR, Burkill GJ, et al: Mycobacterial infections: still a millennium bug—the imaging features of mycobacterial infections. Clin Radiol 56:535, 2001

8. Buckingham SJ, Hansell DM: *Aspergillus* in the lung: diverse and coincident forms. Eur Radiol 13:1786, 2003

9. Arakawa H, Kurihara Y, Niimi H, et al: Bronchiolitis obliterans with organizing pneumonia versus chronic eosinophilic pneumonia: high-resolution CT findings in 81 patients. AJR Am J Roentgenol 176:1053, 2001

10. Philit F, Etienne-Mastroianni B, Parrot A, et al: Idiopathic acute eosinophilic pneumonia: a study of 22 patients. Am J Respir Crit Care Med 166:1235, 2002

11. Epler GR: Bronchiolitis obliterans organizing pneumonia. Arch Intern Med 161:158, 2000

12. Kim SJ, Lee KS, Ryu YH, et al: Reversed halo sign on high-resolution CT of cryptogenic organizing pneumonia: diagnostic implications. AJR Am J Roentgenol 180:1251, 2003

13. Conron M, Beynon HL: Churg-Strauss syndrome. Thorax 55:870, 2000

14. Wechsler ME, Garpestad E, Flier SR, et al: Pulmonary infiltrates, eosinophilia, and cardiomyopathy following corticosteroid withdrawal in patients with asthma receiving zafirlukast. JAMA 279:455, 1998

15. Tuggey JM, Hosker HSR: Churg-Strauss syndrome associated with monteleukast therapy. Thorax 55:805, 2000

16. Tanaka N, Newell JD, Brown KK, et al: Collagen vascular disease-related lung disease: high-resolution computed tomography findings based on the pathologic classification. J Comput Assist Tomogr 28:351, 2004

17. Kim JS, Lynch DA: Imaging of nonmalignant occupational lung disease. J Thorac Imaging 17:238, 2002

18. El Maghraoui A, Chaouir S, Abid A, et al: Lung findings on thoracic high-resolution computed tomography in patients with ankylosing spondylitis: correlations with disease duration, clinical findings and pulmonary function testing. Clin Rheumatol 23:123, 2004

19. Ashizawa K, Hayashi K, Aso N, et al: Lobar atelectasis: diagnostic pitfalls on chest radiography. Br J Radiol 74:89, 2001

20. Jeung MY, Gasser B, Gangi A, et al: Bronchial carcinoid tumors of the thorax: spectrum of radiologic findings. Radiographics 22:351, 2002

21. Wark PA, Gibson PG: Allergic bronchopulmonary aspergillosis: new concepts of pathogenesis and treatment. Respirology 6:1, 2001

22. Tor M, Ozvaran K, Ersoy Y, et al: Pitfalls in the diagnosis of complicated pulmonary hydatid disease. Respir Med 95:237, 2001

23. Lee KS, Kim TS, Fujimoto K, et al: Thoracic manifestation of Wegener's granulomatosis: CT findings in 30 patients. Eur Radiol 13:43, 2003

24. Hartman TE: Radiologic evaluation of the solitary pulmonary nodule. Semin Thorac Cardiovasc Surg 14:261, 2002

25. Chouri N, Langin T, Lantuejoul S, et al: Pulmonary nodules with the CT halo sign. Respiration 69:103, 2002

26. Oshiro Y, Murayama S, Sunagawa U, et al: Pulmonary dirofilariasis: computed tomography findings and correlation with pathologic features. J Comput Assist Tomogr 28:796, 2004

27. Gjevre JA, Myers JL, Prakash UBS: Pulmonary hamartomas. Mayo Clin Proc 71:14, 1996

28. Yoon YC, Lee KS, Kim TS, et al: Intrapulmonary bronchogenic cyst: CT and pathologic findings in five adult patients. AJR Am J Roentgenol 179:167, 2002

29. Bankoff MS, McEniff NJ, Bhadelia RA, et al: Prevalence of pathologically proven intrapulmonary lymph nodes and their appearance on CT. AJR Am J Roentgenol 167:629, 1996

30. Iwasaki Y, Nagata K, Nakanishi M, et al: Spiral CT findings in septic pulmonary emboli. Eur J Radiol 37:190, 2001

31. Murayama S, Sakai S, Soeda H, et al: Pulmonary cryptococcosis in immunocompetent patients: HRCT characteristics. Clin Imaging 28:191, 2004

32. Lee JS, Tuder R, Lynch DA: Lymphomatoid granulomatosis: radiologic features and pathologic correlations. AJR Am J Roentgenol 175:1335, 2000

33. Pickhardt PJ, Siegel MJ, Anderson DC, et al: Chest radiography as a predictor of outcome in posttransplantation lymphoproliferative disorder in lung allograft recipients. AJR Am J Roentgenol 171:375, 1998

34. Fonseca R, Qwitzig TE, Olson LJ, et al: Disseminated Kaposi's sarcoma after heart transplantation: association with Kaposi's sarcoma-associated herpesvirus. J Heart Lung Transplant 17:732, 1998

35. Ramirez J, Mehta JB, Taylor RA, et al: Symptomatic pulmonary hyalinizing granuloma. South Med J 91:867, 1998

36. Nakatsu M, Hatabu H, Morikawa K, et al: Large coalescent parenchymal nodules in pulmonary sarcoidosis: "sarcoid galaxy" sign. AJR Am J Roentgenol 178:1389, 2002

Acknowledgment

The previous version of this chapter was authored by Roland H. Ingram, Jr., M.D., who with Scott F. Davies, M.D., originated its format.

V CHRONIC DIFFUSE INFILTRATIVE LUNG DISEASE

Gerald W. Staton, Jr., m.d.
Roland H. Ingram, Jr., m.d.

Chronic diffuse infiltrative lung disease encompasses a wide variety of disorders that primarily affect the alveoli and lung interstitium; involvement of the airways and the pulmonary vasculature is secondary to the interstitial process.

Chronic diffuse infiltrative lung disease affects approximately 70 per 100,000 population.[1] Postinflammatory pulmonary fibrosis and idiopathic pulmonary fibrosis (IPF) are the most common diagnoses.[1] Other common diseases are sarcoidosis, intersititial lung disease associated with collagen vascular disorders, and infiltrative lung disease secondary to occupational or environmental exposures or to drugs and radiation. In the United States, some of these disorders are more common in African Americans, though data are incomplete.[2]

A combination of environmental exposures and genetic susceptibility may play a significant role in the development of diffuse infiltrative lung disease.[3] Despite the diversity of contributing factors, the different forms of the disease appear to share a common pathogenetic pathway. The steps involved include initiation of tissue injury, maintenance of inflammation, and either resolution or development of remodeling and fibrosis.[4] In animal studies, several causative agents (e.g., bleomycin, paraquat, radiation, and cyclophosphamide) have been shown to produce a pattern of injury that resembles common idiopathic varieties of the disorder. In these models, there is initial damage to type I epithelial cells (which form the air-blood barrier) and capillary endothelial cells. After a phase of edema and mild hemorrhage, sometimes with extravascular fibrin accumulation, there is an influx of neutrophils. Newly recruited lymphocytes and macrophages appear later. Under the influence of various growth factors, type II epithelial cells (which normally proliferate in response to injury) replicate and spread out, replacing the damaged type I cells. Within 2 weeks, collagen, elastin, and other extracellular matrix components are present, and eventually, extensive lung fibrosis becomes evident. Variations in this common pathway lead to the specific diseases discussed in this subsection.

Approach to the Patient with Suspected Chronic Diffuse Infiltrative Lung Disease

Although more than 100 causes of diffuse infiltrative lung disease have been described, a smaller group of causes accounts for the vast majority of cases. Given the nonspecificity of the clinical, radiographic, and sometimes even pathologic features, the differential diagnosis should be approached carefully.[5] In all cases, a history and physical examination should be performed, a chest radiograph should be taken, and pulmonary function should be evaluated. Although such measures often do not lead to a definitive diagnosis, they help determine the most likely causes in a given case and thus serve to markedly limit the diagnostic possibilities, thereby reducing costs and effort of further studies.

presenting symptoms and patient history

The most common presenting complaints in diffuse infiltrative lung disease are dyspnea of gradual onset—initially with extreme exertion—and nonproductive cough. Chest pain and hemoptysis are uncommon except in specific disorders.

The history should include careful questioning as to the duration of symptoms—usually cough or dyspnea. Often, symptoms have been present for many months or even years, which would render an infectious cause very unlikely. In contrast, a more rapidly evolving course, over days to weeks, increases concern about infection. Exclusion of infection as the cause of diffuse infiltrative lung disease is of paramount importance because of the need for specific antimicrobial therapy and because many noninfectious causes of this disorder are treated with glucocorticoids. Fever may be observed in several noninfectious types of diffuse infiltrative lung disease (e.g., sarcoidosis, hypersensitivity pneumonitis, collagen vascular disease, and drug-induced disease), but its presence should always alert the clinician to the possibility of infection. Infectious diffuse infiltrative lung disease that evolves subacutely over weeks to months is rarely caused by common viral or bacterial organisms. On the other hand, mycobacterial and pathogenic fungal infections (e.g., histoplasmosis, coccidioidomycosis, and blastomycosis) can present subacutely and are usually associated with fever. Perhaps the most common infectious cause is *Pneumocystis carinii* pneumonia (PCP) in a patient with AIDS. AIDS-related PCP may present as dyspnea that evolves slowly over many weeks, occasionally without fever or other evidence of systemic toxicity; this pattern may resemble that of common noninfectious causes of diffuse infiltrative lung disease. A careful history regarding risk factors for HIV infection is a critical part of the evaluation of all patients with recent-onset diffuse infiltrative lung disease.

In addition to establishing the duration of illness and the presence or absence of fever and other systemic complaints, the history should focus on the patient's occupation and medications. Many different occupational exposures can lead to diffuse infiltrative lung disease, but most are rarely encountered in clinical practice. Occupation-related varieties are caused by exposure to the inorganic dusts, such as asbestos and silica, or to organic dusts that cause hypersensitivity pneumonitis. Asbestosis typically becomes evident many years after exposure, so in taking the history, it is important to note the patient's prior occupations. The most common form of occupation-related hypersensitivity pneumonitis is farmer's lung, which primarily affects dairy farmers. A farmer should be questioned carefully about any relation between respiratory or systemic symptoms and exposure to organic material (moldy hay or grain) that may contain thermophilic actinomycetes. The history should also include questions about possible exposure to pigeons or pet birds and to humidifiers, because such exposure can result in hypersensitivity pneumonitis [*see Table 1*].

Many drugs have been reported to cause diffuse infiltrative lung disease. The history should focus on both prescribed and over-the-counter medications and should include questions about medications taken during the preceding weeks or months [*see Table 2*].

Table 1 Common Causes of Hypersensitivity Pneumonitis

Disease	*Antigen Source*	*Probable Antigen(s)*
Farmer's lung	Moldy hay	Thermophilic actinomycetes, *Micropolyspora faeni, Thermoactinomyces vulgaris*
Bagassosis	Moldy pressed sugarcane (bagasse)	Thermophilic actinomycetes, *T. sacchari, T. vulgaris*
Multiple bird handler's diseases	Bird droppings, products, and feathers	Bird proteins
Laboratory worker's lung	Rat fur	Rat urine proteins
Diisocyanates and trimellitic anhydrides	Chemical exposures	Altered proteins
Ventilator lung	Contaminated humidifiers, dehumidifiers, air conditioners, and heating systems	Thermophilic actinomycetes, *T. candidus, T. vulgaris, Penicillium* species, *Amoeba* species, *Klebsiella* species, and *Candida* species

Finally, the clinician should ask about symptoms consistent with an underlying collagen vascular disease or vasculitis. A past history of Raynaud phenomenon, photosensitivity, skin rashes, or arthritis carries great diagnostic relevance.

physical examination

The physical examination has less diagnostic utility than the history but is important nonetheless. Collagen vascular disorders may be suggested by detection of synovitis, telangiectasia, sclerodactyly, or a malar rash. Sarcoidosis also involves extrapulmonary organs and should be suspected in patients with uveitis, erythema nodosum, or plaquelike skin lesions consistent with cutaneous sarcoidosis. Malignant diffuse infiltrative lung disease (i.e., lymphangitic spread) may be associated with findings relevant to the primary tumor, such as an abdominal or breast mass, hepatomegaly, or guaiac-positive stools. Finally, cardiovascular examination is very important because on occasion, chronic pulmonary venous congestion resulting from occult mitral stenosis or left ventricular failure may present as diffuse infiltrative lung disease.

laboratory examination

The laboratory examination comprises both routine studies (e.g., complete blood count, biochemical screening) and those that are more specialized. Sarcoidosis may be suggested by cytopenias if the bone marrow is involved or by hypercalcemia or elevated serum liver enzyme levels. Collagen vascular disease and malignancy may also result in cytopenias. Peripheral eosinophilia would suggest chronic eosinophilic pneumonia. Uncommonly, eosinophilic granuloma presents as a diffuse pulmonary disorder; these patients may have laboratory evidence of diabetes insipidus.

Additional laboratory tests, such as autoimmune serology and determination of angiotensin-converting enzyme (ACE) level, may prove useful, depending on the clinical and radiographic features in a given case. Antinuclear antibodies or rheumatoid factor is usually present in the serum of patients with collagen vascular disorders, but both are also found in low titer in up to 50% of patients with IPF. Increased serum levels of ACE are suggestive of sarcoidosis, but ACE can also be elevated in miliary tuberculosis, berylliosis, asbestosis, and silicosis. Hypersensitivity pneumonitis is nearly always associated with the presence of serum antibody against the offending antigen. However, the presence of antibody against one of the causative agents of hypersensitivity pneumonitis in serum does not prove that hypersensitivity pneumonitis is the cause of diffuse infiltrative lung disease but only that there has been sufficient exposure to the antigen to elicit an immunologic response. Resolu-

Table 2 Drugs That Commonly Induce Chronic Parenchymal Lung Disease

Category	*Drugs*	*Notes*
Chemotherapeutic agents	Bleomycin	Dose related, worse with O_2
	Busulfan	Induces pulmonary fibrosis, with high mortality and occasionally alveolar proteinosis
	Cyclophosphamide	Variable onset and course
	Vinblastine	Synergistic with mitomycin-C
	Nitrosoureas	Dose related, delayed onset
Cardiovascular drugs	Amiodarone	Long half-life, induces ARDS after procedures
	Hydralazine, others	Induce systemic lupus erythematosus (SLE)
Anti-inflammatory drugs	Methotrexate	Produces granulomas, as found on biopsy
	Penicillamine	Induces Goodpasture-like syndrome, SLE, bronchiolitis obliterans
	Gold	Injectable only; induces BAL lymphocytosis
Antibiotics	Nitrofurantoin	Induces acute and chronic parenchymal lung disease
	Sulfasalazine	BOOP, PIE, or pulmonary fibrosis
Illicit drugs	Methylphenidate	I.V. injection produces talc granulomas
	Cocaine	Induces BOOP, alveolar hemorrhage

ARDS—acute respiratory distress syndrome BAL—bronchoalveolar lavage BOOP—bronchiolitis obliterans organizing pneumonia PIE—peripheral eosinophilia

tion of pneumonitis with disappearance of antibody after stopping the exposure provides a more conclusive association.

chest radiography

Radiographic features are for the most part nonspecific. Diffuse infiltrative lung disease may be characterized by bilaterally symmetrical interstitial, alveolar, or mixed alveolar-interstitial radiographic patterns. In some cases, the lung fields appear completely normal on the chest radiograph despite the presence of significant clinical and physiologic abnormalities. There are certain ancillary radiographic clues that, if present, may help in narrowing the list of possible causes [*see Table 3*].

high-resolution computed tomography

The use of high-resolution computed tomography (HRCT) represents a significant advance in the evaluation of diffuse parenchymal lung disease. With the use of HRCT, the extent, location, and pattern of lung involvement can be determined with great accuracy [*see Figure 1*]. HRCT can often detect abnormalities in patients who have symptoms of interstitial lung disease but whose chest radiographs are normal. When combined with clinical data and chest radiography, HRCT of the chest can lead to a specific diagnosis in 60% to 80% of cases.[6] When HRCT indicates a specific diagnosis (e.g., eosinophilic granuloma), the need for a lung biopsy is eliminated.[5]

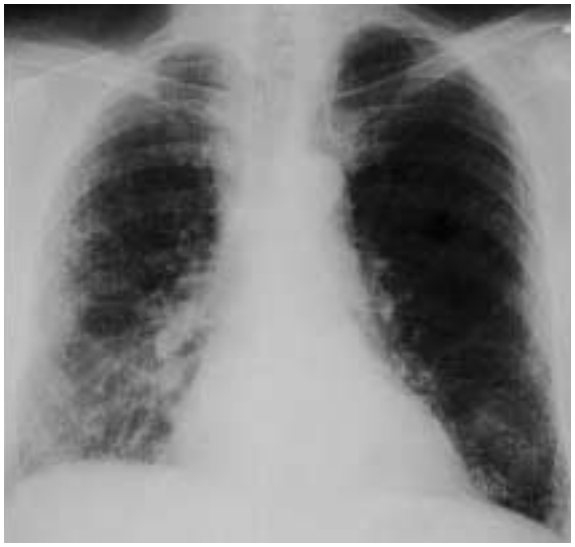
pulmonary function testing

Pulmonary function testing provides diagnostic clues to diffuse infiltrative lung disease, helps in establishing disease severity and defining prognosis, and is useful for monitoring response to therapy and disease progression.[7] The hallmarks of the disorder include a restrictive ventilatory pattern (reduced lung volume), a normal or increased ratio of forced expiratory volume in 1 second to forced vital capacity (FEV_1/FVC), a reduction in the diffusing capacity of the lung for carbon monoxide ($D_{L_{CO}}$), and a reduction in arterial oxygen tension (P_aO_2) associated with normal or reduced arterial carbon dioxide tension (P_aCO_2). In addition, there is usually significant exercise limitation resulting from a fall in P_aO_2, abnormalities in respiratory mechanics, associated pulmonary vascular disease, or a combination of these factors.

bronchoalveolar lavage

In most cases, the cause of lung disease remains uncertain despite careful clinical, radiographic, laboratory, and physiologic evaluation. The next step is usually to perform bronchoscopy with bronchoalveolar lavage (BAL) and transbronchial lung biopsy. Certain causes can be diagnosed solely by BAL, others only by biopsy, and some by either technique. BAL is most useful for diagnosing infectious causes, especially PCP. The diagnostic sensitivity of this procedure for AIDS-related PCP is approximately 90% to 95%. Thus, it is the procedure of choice for diagnosing the 20% to 50% of cases of AIDS-related PCP that cannot be diagnosed by induced sputum examination. Other opportunistic infections, such as cytomegalovirus pneumonia and disseminated fungal or tuberculous infection, can also be diagnosed by BAL. Noninfectious causes that can be diagnosed by this technique include alveolar proteinosis, lymphangitic carcinomatosis, and alveolar cell carcinoma. In addition, BAL may provide helpful information by revealing one of the following changes: (1) increased numbers of eosinophils in chronic eosinophilic pneumonia, (2) asbestos bodies in asbestosis, (3) so-

a



b

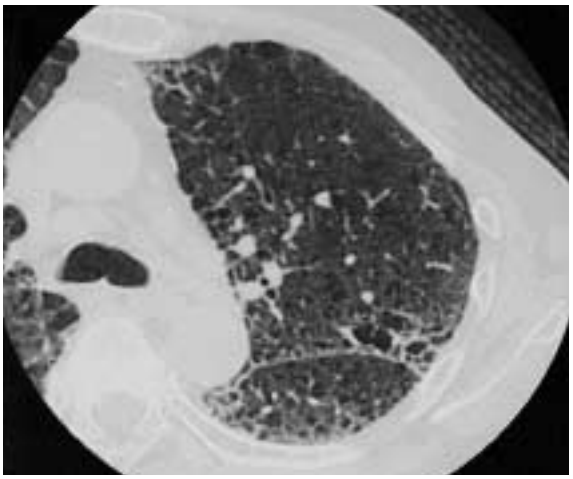


Figure 1 **High-resolution computed tomography (HRCT) in patients with chronic diffuse interstitial lung disease can occasionally suggest a specific diagnosis. (*a*) Posteroanterior radiography demonstrates a diffuse interstitial process, but differential diagnosis is lengthy. (*b*) HRCT shows a combined interstitial and cystic process that is virtually diagnostic of eosinophilic granuloma of the lung.**

Table 3 Radiographic Clues to Diagnosis of Diffuse Infiltrative Lung Disease

Associated Radiographic Finding	*Primary Diagnostic Considerations*
Hilar adenopathy	Sarcoidosis, lymphoma, carcinoma, granulomatous infection
Pleural effusion	Collagen vascular disease, asbestosis, lymphangiomyomatosis (chylous), tuberculosis
Pneumothorax	Eosinophilic granuloma, *Pneumocystis carinii* pneumonia, lymphangiomyomatosis
Upper lung zone predominance	Silicosis, eosinophilic granuloma, sarcoidosis
Peripheral predominance	Eosinophilic pneumonia, bronchiolitis obliterans organizing pneumonia, drug-induced injury

Table 4 Diagnostic Efficacy of Transbronchial Lung Biopsy

Diseases Frequently Diagnosed by TLB Findings
Infectious diffuse infiltrative lung disease
Sarcoidosis
Lymphangitic carcinoma
Alveolar cell carcinoma
Talc-induced diffuse infiltrative lung disease
Diseases Suggested by TLB Findings
Cytotoxic-induced disease
Bronchiolitis obliterans organizing pneumonia (occasionally)
Eosinophilic pneumonia
Hypersensitivity pneumonitis
Not Diagnosable by TLB Findings
Idiopathic pulmonary fibrosis
Diseases commonly misdiagnosed as idiopathic pulmonary fibrosis
Eosinophilic granuloma
Lymphangiomyomatosis
Bronchiolitis obliterans organizing pneumonia
Hypersensitivity pneumonitis
Sarcoidosis (occasional cases)
Lymphangitic carcinoma (occasional cases)

TLB—transbronchial lung biopsy

called foamy cells with lamellar inclusions in amiodarone-induced disease, (4) hyperplastic and atypical type II pneumocytes in cytotoxic drug-induced lung injury, (5) Langerhans cells in eosinophilic granuloma, and (6) a bloody effluent with abundant hemosiderin in alveolar macrophages in diffuse alveolar hemorrhage. Quantitating the number and distribution of inflammatory cells (e.g., macrophages, lymphocytes, and neutrophils) may suggest a specific diagnosis.[8] The alveolar lavage liquid of healthy nonsmokers typically contains 84% to 99% macrophages, 1% to 14% lymphocytes, and 0% to 1% neutrophils. The common causes of nongranulomatous diffuse infiltrative lung disease (e.g., IPF, some of the collagen vascular disorders, and asbestosis) are often characterized by a neutrophilic alveolitis, whereas sarcoidosis and hypersensitivity pneumonitis are associated with increases in lymphocytes. However, overlap exists (e.g., IPF with an increased number of lymphocytes or sarcoidosis with a normal number of lymphocytes).

lung biopsy

In many cases, the diagnosis of chronic diffuse infiltrative lung disease remains unknown until a lung biopsy is obtained. Pathologically, these disorders may be characterized by variable degrees of involvement of alveolar septa or alveoli by inflammatory cells, mesenchymal cells, fibrosis, granuloma, or neoplastic cells. In rare instances, the abnormal substance that accumulates in the lung parenchyma is blood, proteinaceous material (alveolar proteinosis), amyloid, smooth muscle (lymphangiomyomatosis or tuberous sclerosis), or an abnormal material that is deposited as a result of an inherited storage disorder (Gaucher disease, Niemann-Pick disease [*see 5:VII Nonmalignant Disorders of Leukocytes*]).

Lung tissue may be obtained either by transbronchial lung biopsy during bronchoscopy [*see Table 4*] or by one of two open biopsy techniques: traditional open lung biopsy or video-assisted thoracoscopy.[9] Transbronchial lung biopsy is most useful in diagnosing infectious diffuse infiltrative lung disease and sarcoidosis. Noncaseating granulomas can be demonstrated on transbronchial lung biopsy specimens in 70% to 90% of patients with sarcoidosis. The diagnostic yield is highest (approximately 90%) in patients with radiographically apparent lung disease (stages II and III) and lowest (approximately 70%) in patients with hilar adenopathy alone (stage I). Whether one proceeds to open lung biopsy after obtaining a nonspecific transbronchial lung biopsy depends on the particular clinical and radiographic features in an individual case, the impact that a more precise diagnosis would have on therapy, and an assessment of the risk of open lung biopsy in a given case[5] [*see Figure 2*]. The frequency with which open lung biopsy is performed to better define the histopathologic features of noninfectious forms of the disease varies greatly among different centers. Use of video-assisted thoracoscopy rather than open lung biopsy has not resulted in the anticipated decrease in morbidity and cost.[10]

treatment

The treatment recommendations for the lung diseases addressed in this subsection are based on small, nonrandomized series (often without control groups), case reports, and clinical experience. There are very few randomized, controlled trials of treatment for these diseases.

Even before a specific diagnosis is known, preventive and nonspecific measures can be initiated. Immunization against pneumococcal antigens (every 5 years) and yearly immunizations against influenza are indicated. There may be an increased risk of gastroesophageal reflux in patients with certain interstitial lung diseases (e.g., IPF and scleroderma), which should be treated appropriately if present.[11] Patients who have a reversible obstructive defect may benefit from bronchodilators. Many patients will be treated with corticosteroids, so the usual side effects and complications of such agents should be anticipated and treated.[12] Supplemental oxygen given at rest and with exercise will often improve patients' tolerance of activities of daily living.[13] Pulmonary hypertension and cor pulmonale occurring in patients with far advanced disease can be ameliorated with appropriate treatment [*see 14:XI Pulmonary Hypertension, Cor Pulmonale, and Primary Pulmonary Vascular Diseases*]. In patients with the most severe disease, referral for consideration of lung transplantation is appropriate.[14]

Chronic Diffuse Infiltrative Lung Disease of Known Etiology

drug-induced disease

Many different drugs have been reported to cause diffuse infiltrative lung disease [*see Table 2*].[15] Updated information on drug-induced lung disease is available on the Internet at http://www.pneumotox.com. Estimates are that drug-induced lung disease affects several hundred thousand patients each year and accounts for approximately 3% of cases of infiltrative lung disease.[1] To minimize the morbidity and mortality of drug-induced disease, early recognition is critical. Discontinuance of the offending agent is often followed by spontaneous improvement, whereas failure to appreciate the causal relation between the drug and the pulmonary disease can lead to irreversible lung injury. Unfortunately, certain aspects of drug-induced disease can hinder the recognition of this cause-and-ef-

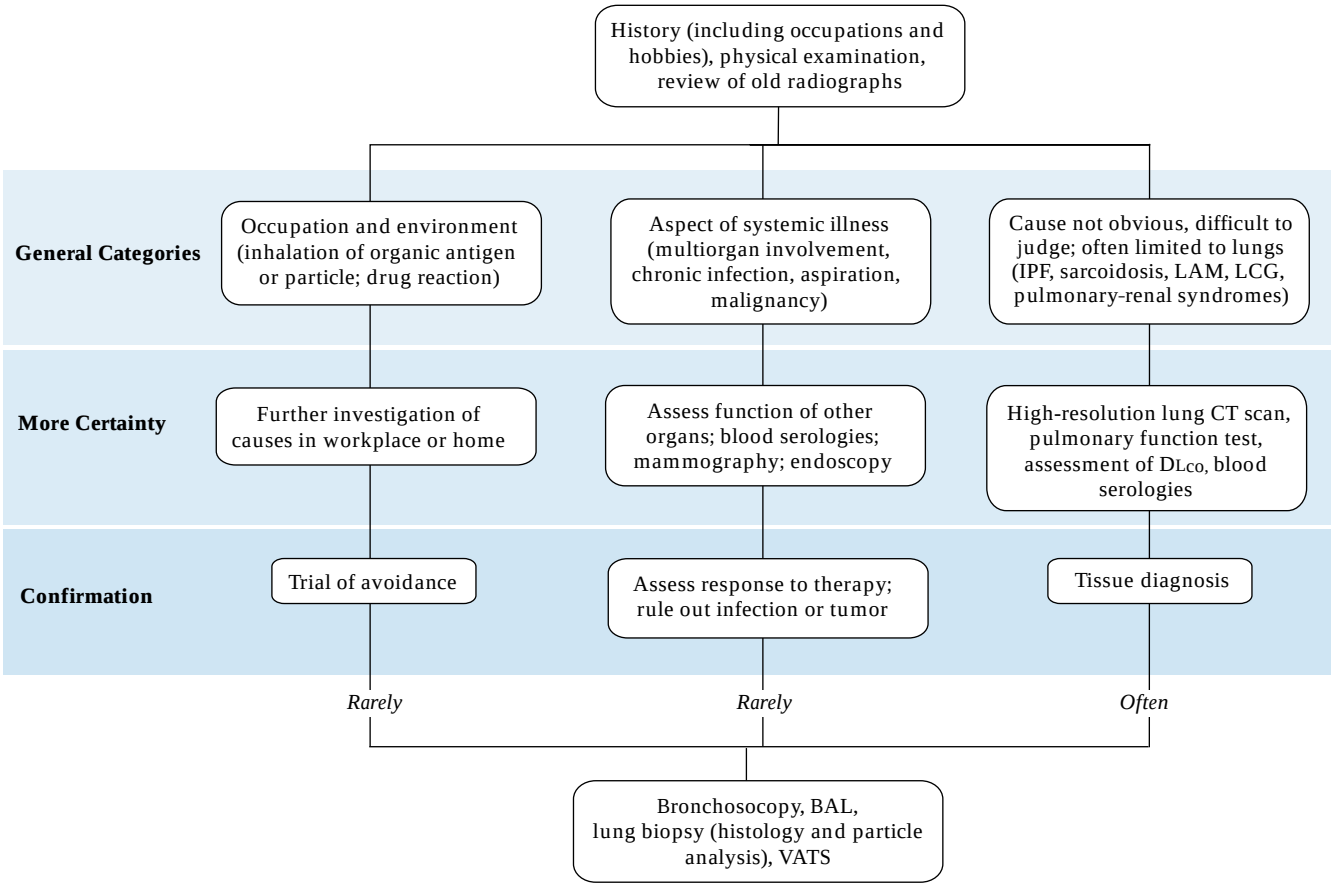


Figure 2 **Algorithm for the evaluation of diffuse interstitial lung disease.[5] (BAL—bronchoalveolar lavage; IPF—idiopathic pulmonary fibrosis; LAM—lymphangioleiomyomatosis; LCG—Langerhans cell granulomatosis; $D_{L_{CO}}$—diffusing capacity of the lung for carbon monoxide; VATS—video-assisted thoracoscopic surgery)**

fect relation. First, although many drugs can cause diffuse infiltrative lung disease, very few of the patients who receive such drugs experience this disease. Second, the onset of the pulmonary disease may occur weeks to months after the drug is begun. In the case of cytotoxic drug–induced disease, the onset of respiratory symptoms can occur many weeks after the last exposure to the offending agent. Finally, the drugs that cause diffuse infiltrative lung disease are often prescribed for conditions that are themselves associated with an increased risk for the disease. For example, a patient who receives an anticancer chemotherapeutic agent is at risk for infection and tumor infiltration of the lung. Similarly, the anti-inflammatory drugs that cause parenchymal lung disease are often prescribed for one of the collagen vascular disorders that can have diffuse infiltrative lung disease as one of its features. For all these reasons, the physician must have a heightened awareness of the possibility of a drug-induced etiology in patients with diffuse infiltrative lung disease.

Disease Induced by Cytotoxic Drugs

Diffuse infiltrative lung disease is a major cause of morbidity and mortality in patients undergoing treatment for an underlying malignancy. In some cases, the drug that is used to treat the malignancy is the direct cause of the pulmonary lesions. Some patients die of cytotoxic drug–induced pulmonary toxicity, and others experience permanent impairment of lung function after being cured of their malignancy. Most of the drugs used in the treatment of malignancy have the potential to cause diffuse infiltrative lung disease. The major offending agents are bleomycin, cyclophosphamide, methotrexate, and the nitrosoureas. Cyclophosphamide and methotrexate are also being used increasingly for nonmalignant conditions, though they are administered at much lower doses in the treatment of these disorders than in the treatment of malignancy. Drug-induced pulmonary disease is less common at these lower doses, but it can occur.

Clinical manifestations The clinical features of cytotoxic drug–induced disease are nonspecific. Cough and dyspnea are often prominent, and fever is not uncommon. The chest radiograph usually reveals bilateral and symmetrical interstitial infiltrates, though asymmetry may be seen early in the course of disease. The pace of the disease, both clinically and radiologically, is variable. Most often the onset is subacute, with cough and dyspnea occurring over several weeks. A more explosive onset, with features of acute respiratory distress syndrome (ARDS) and an urgent need for mechanical ventilatory support, can also be seen.[16] Lung fibrosis resulting from therapy with bleomycin and other agents can occur insidiously over many months.

Pathologic features The pathologic features of cytotoxic drug–induced disease are distinctive but not pathognomonic. Interstitial inflammatory cell infiltration and fibrosis may be

observed. However, the most characteristic finding of cytotoxic drug–induced disease is an increase in type II pneumocytes that show marked atypia. This pathologic feature is highly suggestive of cytotoxic drug–induced disease, but similar changes in type II cells can be seen with severe viral infections and during the reparative phase of ARDS.

A somewhat different pathologic picture is seen with disease induced by cytarabine or methotrexate. Cytarabine has been associated with an often fatal form of noncardiac pulmonary edema. Methotrexate-induced disease has been associated with granulomas, lymphocytic alveolitis, and a significant increase in helper T cells or with a diffuse alveolar damage pattern.[17]

Diagnosis Diagnosis of cytotoxic drug–induced lung disease is established by identifying a previous exposure to the offending agent, excluding infection as the cause of the lung damage, and demonstrating pathologic features that are consistent with drug-induced injury. The temporal relation between drug exposure and the onset of lung disease is variable. An interval of a few weeks between the last exposure to the drug and the onset of symptoms is not unusual; in rare cases, this interval can be as long as several months. The clinical and radiographic features of lung disease induced by a cytotoxic drug are indistinguishable from those caused by opportunistic infection. Thus, an aggressive approach to diagnosis is required. Failure to detect an infectious etiology by bronchoscopy increases the likelihood of drug-induced disease, especially if the characteristic type II cell changes are observed in a transbronchial lung biopsy specimen. However, open lung biopsy may be required to confidently exclude infection and to better define the histopathologic features of cytotoxic drug–induced injury. Even with open lung biopsy, the diagnosis of cytotoxic drug–induced disease remains inferential because there are no pathognomonic features.

Treatment Elimination of further drug exposure is essential in managing cytotoxic drug–induced lung disease. Anecdotal reports indicate that glucocorticoid therapy has been associated with rapid improvement in gas exchange and reversal of chest radiograph abnormalities. If the cytotoxic drug–induced disease is very severe or appears to progress despite elimination of further drug exposure, an empirical course of glucocorticoids is advisable.

Prevention Cytotoxic drug–induced lung disease is not easily prevented, because cytotoxic drugs are often necessary for optimal treatment of a potentially fatal malignancy. For patients with testicular carcinoma who receive bleomycin, the risk of pulmonary lesions appears to be dose related. Monitoring of the $D_{L_{CO}}$ and spirometry during therapy may in some cases permit detection of lung injury early enough to allow discontinuance of the drug and lessen the likelihood of permanent and severe lung impairment. However, therapy for non-Hodgkin lymphoma with bleomycin in combination with doxorubicin, cyclophosphamide, vincristine, and prednisone causes diffuse infiltrative lung disease in a non–dose-dependent fashion, thereby impeding efforts to prevent it. Perhaps the most effective strategy to reduce morbidity and mortality from cytotoxic drug–induced disease is not to prevent it altogether but, rather, to diagnose it as early as possible; further drug exposure can then be avoided by switching to alternative chemotherapy regimens whenever feasible.

Disease Induced by Noncytotoxic Drugs

Diffuse infiltrative lung disease may result from exposure to a variety of noncytotoxic agents, including antibiotics, anti-inflammatory agents, antiarrhythmics, and illicit drugs. In addition, this disorder may be one of the manifestations of drug-induced systemic lupus erythematosus (SLE).

Nitrofurantoin-induced disease Nitrofurantoin, an antibacterial agent used primarily for the treatment of urinary tract infection, is one of the most common causes of drug-induced lung disease. Both acute and chronic pulmonary toxicity can occur, but the acute syndrome is much more common. The acute pleuropulmonary reaction begins 2 to 10 days after initial drug exposure and is manifested by dyspnea, cough, and often fever. Pleurisy occurs in one third of patients. The chest radiograph shows a pattern of alveolar or interstitial infiltrates, sometimes accompanied by a pleural effusion. Peripheral blood eosinophilia may be observed. The disorder is diagnosed on the basis of a history of recent exposure to nitrofurantoin and spontaneous resolution of the clinical and radiographic changes 1 to 4 days after discontinuance of the drug. This rate of resolution is much faster than would occur with infection and serves to establish the diagnosis with a high degree of certainty. Chronic toxicity is not associated with systemic symptoms and has a clinical and radiographic picture that is indistinguishable from that of IPF. If there is no improvement within 2 to 3 months after withdrawal of the drug, corticosteroid therapy is indicated.

Amiodarone-induced disease Amiodarone is an antiarrhythmic agent that is used in the treatment of many types of tachyarrhythmia. This drug has a variety of dose-dependent toxic effects on different organs, but the major limitation to its use is pulmonary toxicity, which occurs in 5% to 7% of cases. Amiodarone-induced pulmonary toxicity is usually heralded by the onset of cough and dyspnea that may initially be attributed to heart failure. Systemic symptoms, including low-grade fever, are not uncommon. The chest radiograph may show a bilateral, symmetrical interstitial process similar to that observed in other types of diffuse infiltrative lung disease. However, more unusual radiographic patterns occur, including unilateral disease and isolated upper lobe disease; the latter finding may suggest tuberculosis. Pleural effusion occurs occasionally.

A diagnosis of amiodarone-induced lung disease is established primarily by excluding other likely causes of the pulmonary disorder, especially infection and heart failure. Bronchoscopy with BAL and biopsy helps exclude infection and reveals the presence of so-called foamy macrophages with lamellar inclusions (visualized by electron microscopy). These changes within macrophages are indicative of exposure to amiodarone but do not prove that the drug is the cause of the pulmonary process, because similar changes are seen in asymptomatic persons who are receiving the drug.

Interestingly, there have been several reports of acute amiodarone-induced lung disease in patients who have undergone relatively minor surgical procedures, such as placement of an automatic defibrillator, or in those who have undergone pulmonary angiography. Often, the disease follows a fulminant course, and sometimes the outcome is fatal. Because the half-life of amiodarone is several weeks, patients who undergo these procedures are still at risk for acute lung injury if the drug was discontinued 1 to 2 months before the procedure.

Withdrawal of the drug is the cornerstone of treatment for

amiodarone-induced lung disease. Glucocorticoids seem to be useful in more severe or persistent cases.

Disease induced by gold and penicillamine Gold and penicillamine, used primarily for the management of rheumatoid arthritis and other collagen vascular disorders, can lead to diffuse infiltrative lung disease. Even with an open lung biopsy, it may be difficult to determine whether the pulmonary damage is caused by the therapeutic agent or by the underlying disorder. Discontinuance of the drug, along with empirical corticosteroid therapy if symptoms are severe, is advisable. Penicillamine has also been implicated in cases of diffuse alveolar hemorrhage with glomerulonephritis, drug-induced SLE, and panbronchiolitis with severe airflow obstruction.

Disease induced by illicit drugs Illicit drugs can also cause diffuse infiltrative lung disease. Talc-induced granulomatosis and pulmonary fibrosis can result from injection of crushed and dissolved amphetamine or narcotic pills that have talc as the filler. The disease can progress years after the last exposure, because the talc persists in the lungs and continues to elicit an inflammatory response. Open lung or transbronchial lung biopsy reveals a granulomatous inflammation with abundant talc particles, which can be visualized with polarizing microscopy. Failure to evaluate granulomas with a polarizing microscope may lead to an erroneous diagnosis of sarcoidosis. The diagnosis can also be made by detecting talc in the retina. Heroin use can lead to acute pulmonary edema but does not in itself cause a chronic diffuse infiltrative lung disease. Cocaine has been associated with acute pulmonary edema and has been reported to cause diffuse alveolar hemorrhage and a syndrome of pulmonary infiltrates with eosinophilia. Cocaine-related pulmonary reactions do not lead to chronic lung lesions.

pneumoconioses

Environmental or occupational exposure to particulate matter can cause several pneumoconioses, including asbestosis, silicosis, coal worker's pneumoconiosis (CWP), and berylliosis. Whether a person develops disease after exposure to such agents is determined by (1) differences in delivery and the persistence of the agent in the lung, (2) genetic and acquired differences in enzymatic and nonenzymatic defense systems, and (3) genetically determined differences in the propensity to develop inflammation in response to the agent.[18]

Pneumoconioses manifest as asthma, chronic bronchitis, or diffuse parenchymal (mostly interstitial) disease. Complications that may be common to a wide variety of pneumoconioses include carcinoma of the lung, especially in patients with diffuse interstitial fibrosis,[19] and chronic necrotizing pulmonary aspergillosis.[20]

Asbestosis

Exposure to asbestos, a fibrous silicate used in insulation, in friction-bearing surfaces, and to strengthen materials, is associated with the development of pleural changes (plaques and effusions), an increased incidence of malignancy (bronchogenic carcinoma and mesothelioma), and diffuse interstitial fibrosis. Only diffuse interstitial fibrosis is called asbestosis. The specific form of asbestos is important in the subsequent development of fibrosis: short-fiber chrysotile asbestos disappears from the lung in a short time and is associated with a lower frequency of disease, whereas amphibole asbestos (amosite or crocidolite) contains longer fibers that persist in the lung and is associated with a higher frequency of disease.[21]

Asbestosis is characterized by the gradual onset of dyspnea 20 to 30 years after exposure to asbestos. Cough is usually present but is nonproductive unless the patient has been a smoker and has complicating chronic bronchitis. Fine end-inspiratory rales can be heard before the chest radiograph becomes abnormal, and digital clubbing is common. In the late stages of asbestosis, signs of cor pulmonale may develop.

Early in the course of disease, the chest radiograph may be normal, but small, irregular linear shadows gradually develop in the lower lung zones. Pulmonary function tests may show restriction, decreased $D_{L_{CO}}$, and exercise-induced hypoxemia.[22] Diagnosis of asbestosis is made when a patient has a history of exposure to asbestos, pleural plaques (an objective indicator of exposure), and, in uncertain cases, excess asbestos in specimens obtained from the lung (by BAL, transbronchial lung biopsy, or open lung biopsy).

There is no specific treatment for patients with asbestosis.

Silicosis

Silicosis is a chronic fibrotic lung disease caused by exposure to crystalline free-silica.[21] The occupational settings in which significant free-silica exposure may occur include certain types of mining; the cutting, polishing, and carving of stone; foundry work; and abrasive cleaning (sandblasting). Exposure of approximately 5 years' duration is usually required for the development of silicosis unless the exposure is very heavy. There are two forms of silicosis: simple nodular silicosis and progressive massive fibrosis (PMF). Many patients with simple nodular silicosis are asymptomatic or suffer only from chronic bronchitis secondary to tobacco use, whereas patients with PMF may develop disabling dyspnea.

Physical findings are often absent. Radiographically, simple nodular silicosis is characterized by diffuse, small, rounded opacities that tend to be more prominent in the upper lobes. Hilar node enlargement may be seen, sometimes with concentric (eggshell) calcifications. In patients with PMF, the opacities coalesce into large, irregularly shaped masses. In patients with simple nodular silicosis, pulmonary function may be normal or may exhibit a mixed pattern of obstruction and restriction.[22] Severe restriction and hypoxemia, as well as pulmonary hypertension, develop in patients who have PMF.

Diagnosis of silicosis can usually be made if the patient has a history of exposure and the chest radiograph is consistent with silicosis (see above). In atypical cases, the presence of silica in BAL liquid or in lung biopsy specimens (often guided by CT scan of the chest) can establish the diagnosis and exclude tuberculosis and carcinoma.

Tuberculosis and infection with atypical mycobacteria occur with increased frequency in patients with silicosis and may be confused with the progression of the silicosis, making conclusive diagnosis difficult. Because cell-mediated immunity appears to be normal in patients with silicosis, the tuberculin skin test can be helpful in identifying patients in whom diagnosis of tuberculosis should be explored.

Silicosis is generally considered an untreatable disease, though one report suggests that oral glucocorticoids reduce lung inflammation and improve lung function.[23] Prevention of further exposure to silica and prevention and treatment of active tuberculosis are important aspects of care.

Coal Worker's Pneumoconiosis

CWP is an uncommon cause of pulmonary fibrosis in workers who are exposed to coal dust and graphite.[21] Most of these patients are miners.

Many patients with CWP have a chronic cough that is sometimes productive of gray or black sputum, which may be caused by chronic bronchitis that is related to tobacco use or exposure to coal dust. Much like silicosis, CWP comprises simple CWP and PMF. PMF is characterized by severe dyspnea, whereas simple CWP is associated with less dyspnea.[22] The radiographic changes characteristic of simple CWP are small, rounded opacities; opacities that are larger than 1 cm are arbitrarily classified as PMF. Simple CWP does not progress if dust exposure is eliminated, whereas PMF can progress after exposure has stopped.

There is no specific treatment for either form of disease.

Berylliosis

Beryllium is a metal used in high-technology industries. High-intensity exposure to beryllium can cause an acute chemical bronchitis and pneumonitis, which is very rare currently because of strict environmental controls. However, chronic berylliosis is still frequently seen. Chronic berylliosis is characterized by multisystemic granulomatous disease that has many similarities to sarcoidosis and that occurs months to years after exposure.[24] The number of chronic cases appears to be decreasing, probably because of increased awareness of the disease and the adoption of environmental controls. Beryllium is thought to bind to host proteins that are carried throughout the body. These beryllium-protein complexes stimulate a delayed hypersensitivity response, which produces granulomatous inflammation at the sites of disease activity.[24]

The symptoms of berylliosis are dyspnea, cough, chest pain, fatigue, weight loss, and arthralgia. Rales are not an early feature. The radiographic findings, which may precede clinical symptoms, include ill-defined nodular and irregular opacities that are sometimes associated with hilar adenopathy. As the disease progresses, the lungs become smaller and areas of honeycombing develop. Pulmonary function tests may be normal; more often, however, they are characterized by a low $D_{L_{CO}}$ with or without restriction, or they demonstrate mild obstruction. Other abnormalities include hyperuricemia, hypercalcemia, hypercalciuria, and elevated serum levels of ACE.

The criteria for diagnosis include a well-documented history of exposure or a demonstration of excess beryllium in patient specimens; objective evidence, on radiography and pulmonary function testing, of lower respiratory tract disease consistent with berylliosis; and demonstration of granulomatous inflammation in the lung, lymph nodes, or other sites of disease activity.

Differentiating berylliosis from sarcoidosis may be difficult. Sarcoidosis is suggested by the presence of uveitis and erythema nodosum, the presence of hilar adenopathy without parenchymal infiltrates, or improvement after short-term steroid treatment. Berylliosis is suggested by lymphocyte transformation in response to beryllium compounds; this testing is performed on lymphocytes obtained from blood or BAL.

The most important step in the management of berylliosis is prevention of further exposure. Glucocorticoids may improve the course of the disease, but these agents may have to be given for the rest of the patient's life.[24]

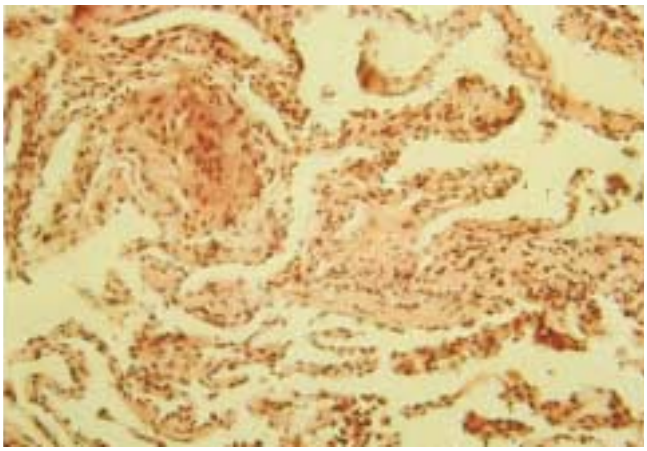
hypersensitivity pneumonitis

The term hypersensitivity pneumonitis (allergic alveolitis) refers to interstitial lung disease that results from inhalation of organic antigens.[25] Several organic antigens cause this syndrome [*see Table 1*], but the most common form of hypersensitivity pneumonitis, called farmer's lung, is caused by inhalation of proteins from thermophilic *Actinomyces* organisms present in moldy hay and grain. Other relatively common forms of hypersensitivity pneumonitis include pigeon breeder's disease and bird fancier's disease, in which inhaled proteins from pigeons or pet birds induce the syndrome. Humidifier lung disease is a type of hypersensitivity pneumonitis that results from exposure to contaminated forced-air systems with humidifiers or dehumidifiers. The responsible antigens in such cases are from *Actinomyces* organisms, as in farmer's lung, but also can include *Aureobasidium* organisms and various amebas. Forms of hypersensitivity pneumonitis other than farmer's lung, pigeon breeder's disease, bird fancier's disease, and humidifier lung disease are rare and generally result from exposure to antigens in rather arcane occupations (e.g., paprika slicing and mummy wrapping).

Pathogenesis and Pathology

The pathogenesis of hypersensitivity pneumonitis is not fully understood. The current hypothesis is that inhaled material causes direct activation of the alternative complement pathway, leading to the formation of chemotactic factors for neutrophils

a



b

Figure 3 **Open lung biopsy specimen from a patient with hypersensitivity pneumonitis (farmer's lung disease) shows characteristic features of (*a*) scattered granulomas and (*b*) bronchiolitis obliterans.**

and in turn to acute neutrophilic alveolitis. The products of complement activation also stimulate the alveolar macrophages to produce monokines, causing a sequence of events that results in granuloma formation.[25]

The pathologic changes in hypersensitivity pneumonitis have been well defined by examination of open lung biopsy specimens of a large number of patients with farmer's lung disease. All patients have an interstitial pneumonitis, and two thirds have scattered granulomas. Fibrosis may occur in more chronic forms of the disease, and as many as 50% of patients may have bronchiolitis obliterans. The pathologic picture of lymphocytic interstitial pneumonitis with granuloma formation and bronchiolitis obliterans is characteristic of hypersensitivity pneumonitis [*see Figure 3*]. BAL studies have examined the nature of the lymphocyte population in the lung and have demonstrated that the lymphocytes are predominantly of the suppressor-cytotoxic T cell subclass. This inversion of the ratio of helper T cells to suppressor-cytotoxic T cells in hypersensitivity pneumonitis contrasts with the predominance of helper T cells in patients with sarcoidosis. The finding of an intense lymphocytic alveolitis with inversion of the ratio of helper T cells to suppressor-cytotoxic T cells suggests hypersensitivity pneumonitis, but this pattern cannot be considered specific, because a similar lavage profile may be seen in other disorders (e.g., IPF associated with a lymphocytic alveolitis).

Clinical Manifestations

Hypersensitivity pneumonitis has two basic clinical forms: acute and chronic.

Acute hypersensitivity pneumonitis Acute hypersensitivity pneumonitis is characterized by fever, chills, cough, dyspnea, and malaise that typically occur 4 to 8 hours after antigenic exposure. Most patients with acute hypersensitivity pneumonitis improve spontaneously after exposure to the antigen ceases. Occasionally, patients experience a more severe episode characterized by hypoxemia and by considerable respiratory distress and require hospitalization. Because of the prominence of systemic symptoms during the acute episode, the patient may be erroneously diagnosed as having a respiratory infection. With continuing exposure to the inhaled antigen—as may occur in people living on a farm, for example—the patient may have daily episodes of acute hypersensitivity pneumonitis that blend into each other. If this pattern occurs, the patient may complain of having a flulike illness for many weeks. Respiratory symptoms, such as cough and wheezing, that occur immediately after exposure to an organic material should not be attributed to hypersensitivity pneumonitis but rather to an irritant response within the airways.

Chronic hypersensitivity pneumonitis In chronic hypersensitivity pneumonitis, there are usually no systemic symptoms; rather, the symptom complex is marked by chronic dyspnea and cough on exertion. The clinical presentation is that of interstitial lung disease of uncertain etiology.

Diagnosis

There is no single diagnostic test that can be used to establish an unequivocal diagnosis of hypersensitivity pneumonitis. Therefore, one must rely on a combination of history, radiography, serologic study, and exclusion of other possible causes.[25] A history of previous episodes of acute hypersensitivity pneumonitis is of great diagnostic value. In taking the history, the clinician should focus on establishing whether a relevant antigenic exposure has occurred and, if so, try to determine whether typical acute symptoms occurred several hours after the exposure. Such a history is more likely to be obtained in cases of heavy exposure (e.g., to moldy hay or a pigeon coop) than in cases of low-level exposure (e.g., to a humidifier or a single bird). In the latter circumstance, the exposure is more or less continuous, and it is much more difficult to establish a causal relation between antigenic exposure and symptoms. Patients should be questioned carefully as to their occupations and hobbies and whether symptoms resolve when they are on vacation or at other times when they are away from the putative source of illness. At least 50% of patients who present with the chronic form of the disease deny having a history of an acute episode. Such acute episodes may well have occurred, but the patient may have attributed the symptoms to respiratory tract infections. The patient may spontaneously improve on hospitalization for evaluation of a respiratory illness, only to experience a recurrence of symptoms soon after returning home. Such a history should always raise suspicion about hypersensitivity pneumonitis.

Chest radiography Chest radiographs of patients with acute hypersensitivity pneumonitis may range from normal to demonstration of alveolar, nonlobar infiltrates. As chronic disease develops, a reticular, nodular, or combined infiltrate evolves, and fibrosis may become evident during continued antigenic exposure. HRCT can be very useful in suggesting the diagnosis and differentiating hypersensitivity pneumonitis from IPF and other infiltrative lung diseases.[6]

Serologic study Serologic study is useful primarily in establishing exposure to a specific antigen. Nearly all patients with farmer's lung and pigeon breeder's disease have antibodies to thermophilic *Actinomyces* organisms and pigeon serum, respectively. However, these antibodies may also be present in the serum of a significant percentage of asymptomatic persons. Hence, the presence of antibody in serum does not establish a diagnosis of hypersensitivity pneumonitis but demonstrates that there has been significant exposure to the antigen.

Bronchoscopy Patients who present with subacute to chronic hypersensitivity pneumonitis should undergo diagnostic bronchoscopy with BAL and transbronchial biopsy. The lavage liquid in hypersensitivity pneumonitis typically shows an intense lymphocytosis, often with more than 50% lymphocytes. As noted (see above), these lymphocytes are primarily suppressor-cytotoxic T cells. As time passes after the last exposure to antigen, BAL results tend to return to normal.[26]

Lung biopsy Transbronchial lung biopsy shows a nonspecific interstitial pneumonitis in most patients and may show evidence of granuloma or bronchiolitis obliterans. Even in the two thirds of cases in which granulomas are present, these lesions are more scattered than sarcoid granulomas and are therefore not as likely to be demonstrated by transbronchial biopsy. Open lung biopsy is usually reserved for especially difficult and confusing cases in which a diagnosis cannot be established by a combination of history, radiography, serologic study, and bronchoscopy. Few patients who have hypersensitivity pneumonitis require open lung biopsy for diagnosis.

Table 5 Chronic Diffuse Parenchymal Lung Diseases of Unknown Cause*

Entity	*Clinical and Laboratory Clues*	*Chest Radiographic Clues*	*High-Resolution CT (HRCT) Findings*	*Pulmonary Function*
Sarcoidosis	Young African-American females; minimal symptoms; skin lesions; increased serum ACE activity	BHA; upper lung field infiltrates	Nodules with a perilymphatic distribution along bronchovascular bundles, interlobular septa, and subpleurally; ground-glass attenuation; reticular shadowing, distortion most commonly in the upper zones; air trapping; associated hilar and mediastinal lymphadenopathy	Restriction with decreased Dl_{CO}; some have obstruction
Idiopathic pulmonary fibrosis (IPF)	Older men; slowly progressive dyspnea; dry end-inspiratory rales; clubbing of digits	Sometimes normal initially; bibasilar reticular infiltrates	Lower-zone and subpleural predominance, occurring posteriorly at the bases and becoming increasingly anterior as disease progresses superiorly; reticular pattern with associated honeycombing in up to 95% of cases	Diffusion decreased more than lung volumes
Desquamative interstitial pneumonia / respiratory bronchiolitis interstitial lung disease (DIP / RBILD)	Middle-aged men more than women; slowly progressive dyspnea; clubbing of digits	May be normal; bibasilar hazy opacities	Ground-glass attenuation, basally and peripherally centered but diffusely spread throughout all zones; approximately half also show features of fibrosis, such as anatomic distortion or traction bronchiectasis	Restrictive defects less severe than in IPF
Acute interstitial pneumonia (AIP)	Rapid onset of dyspnea, cough, fever evolving to respiratory failure	Diffuse bilateral air-space opacities	Widespread disease, which less frequently demonstrates lower-zone predominance; ground-glass attenuation admixed with features of fibrosis, usually with air-space consolidation, possibly emphysema	Rapidly progressive restrictive defect with hypoxemia
Lymphocytic interstitial pneumonia (LIP)	Older women more than men; associated disorder may dominate clinical picture	Reticulonodular infiltrates	Centrilobular nodules, ground-glass attenuation, septal and bronchovascular thickening, thin-walled cysts	Restrictive defect with decreased Dl_{CO}
Nonspecific interstitial pneumonia and fibrosis (NSIP / F)	Middle-aged; women slightly greater than men; dyspnea and cough; often with fever	Bilateral hazy opacities	Bilateral patchy and subpleural areas of ground-glass opacity with lower-zone distribution, variably accompanied by areas of consolidation and irregular linear opacity; features suggestive of fibrosis, such as fissural distortion and bronchial dilatation in the absence of gross honeycombing	Restrictive defect with decreased Dl_{CO}
Bronchiolitis obliterans organizing pneumonia (BOOP)	Idiopathic BOOP seen in older patients; abrupt onset of cough, dyspnea, fever; rales and wheezes	Some have solitary focal infiltrates; bilateral patchy alveolar or ground-glass infiltrates	Bilateral areas of consolidation with air on bronchograms of the lower lung zones in a predominantly subpleural or peribronchial distribution, with associated nodules up to 4 cm in diameter; ground-glass attenuation in some cases	Mixed restriction and obstruction
Collagen vascular disease	Often asymptomatic; progressive dyspnea, cough, sometimes with chest pain, hemoptysis; positive serologies	Usually, bilateral diffuse or patchy infiltrates	Ground-glass, fibrotic, or combination	Restrictive defect with decreased Dl_{CO}

*After ruling out infections, drugs, and inhalants (organic and inorganic).
ACE—angiotensin-converting enzyme ARDS—acute respiratory distress syndrome BHA—bilateral hilar adenopathy DAD—diffuse alveolar damage Dl_{CO}—diffusing capacity of the lung for carbon monoxide ILD—interstitial lung disease LDH—lactate dehydrogenase MCTD—mixed connective tissue disease PM / DM—polymyositis / dermatomyositis PSS—progressive systemic sclerosis SLE—systemic lupus erythematosus UIP—usual interstitial pneumonia

Treatment

Treatment of hypersensitivity pneumonitis primarily involves stopping the exposure to the causative antigen. This step may be accomplished rather early in the case of humidifier lung disease or bird fancier's disease but is somewhat more difficult to accomplish in patients with farmer's lung. Sometimes, a dairy farmer is able to have another person do the work that involves exposure to the moldy hay. High-efficiency masks may help reduce the number of spores the patient inhales. Improvement in farming practices, including better ventilation and other steps to reduce the moisture content of hay, may significantly lessen the antigenic load. Referral of the patient to a center that deals with agricultural lung disease is appropriate.

Glucocorticoids are used primarily for acute hypersensitivity pneumonitis characterized by severe systemic symptoms and gas exchange abnormalities. Patients with subacute to chronic hypersensitivity pneumonitis may also improve more rapidly if they are given a course of glucocorticoids for several weeks. Elimination of further antigen exposure is essential when glucocorticoids are administered. Otherwise, the symptoms may be masked by use of the glucocorticoids, and continued exposure to the antigen could result in further lung damage. When antigenic exposure ceases, most patients undergo spontaneous improvement. Patients who have entered a fibrotic phase of the disease may not show substantial improvement, and there is a small subset of patients who seem to deteriorate even though antigenic exposure has apparently been eliminated. In the latter group, it may be beneficial to follow the serum antibody level; with lack of further exposure, the titer will gradually fall. An unchanging or rising titer would suggest continuing antigenic exposure.

malignant diffuse infiltrative lung disease

Diffuse involvement of the lung by malignancy occurs with

Table 5 (*continued*)

Associated Diseases, Exposures, Risk Factors	*Extrathoracic Manifestations*	*Histopathology*	*Therapy*	*Prognosis*
Familial incidence sometimes	Neurologic, eye, skin, heart, liver, hypercalcemia, joints	Noncaseous granulomas; must exclude granulomatous infection, berylliosis	For many—none; when indicated, steroids, antimalarials, methotrexate	Excellent
Familial IPF, cigarette smoking, chronic aspiration, ? viral infections or exposures	Arthralgias; cor pulmonale late in disease	UIP; fibrosis with widening of the alveolar septa; fibroblastic foci of varying age; minimal inflammation	Steroids, colchicine, and cytotoxic agents probably of no benefit; ? interferon gamma	Complete recovery not possible; mortality 68%; median survival 2–3 yr
Highly associated with cigarette smoking	None	Homogeneous filling of air spaces, especially in the peribronchiolar areas	Good response to discontinuance of smoking; good response to steroids	Complete recovery possible; mortality 27%; mean survival 12 yr
Viral prodrome	None	Diffuse alveolar damage; cannot be differentiated from ARDS pathologically	Supportive care with oxygen, mechanical ventilation; not clear whether steroids of any benefit	Mortality 33%–78%; survival often less than 6 mo; recurrences and progressive ILD seen
Autoimmune disorders (Sjögren syndrome), dysproteinemias, viral infections (HIV), drugs, bone marrow transplantation	Depends on associated disorder	Lymphocytic infiltrate of alveolar walls and alveoli that must be distinguished from lymphoma	Treatment of underlying disorder; some respond to steroids, others require immunosuppressive agents	Depends on underlying disorder; can be progressive
Collagen vascular disorders, exposures to agents known to cause hypersensitivity pneumonitis, other environmental exposures, drugs, HIV infection	None	Often confused with UIP; temporally uniform mixture of alveolar wall inflammation (mostly lymphocytes and plasma cells and fibrosis)	Good response to steroids	Complete recovery possible; median survival 13.5 yr; survival worse with increased fibrosis
Many idiopathic but some associated with infections, collagen vascular disorders, aspiration, transplantation, inhaled toxins, drugs, inflammatory bowel disease	Depends on associated disorder	Intraluminal plugs of connective tissue in the bronchioles, alveolar ducts, and alveolar spaces	Focal infiltrates do not require therapy; for others, steroids are successful but with high incidence of relapse	Patients with focal BOOP have no increased mortality; those with diffuse disease have a good prognosis with steroids
Rheumatoid arthritis, SLE, PSS, Sjögren syndrome, MCTD, PM / DM	Arthritis, skin rash or thickening, dysphagia, serositis, dry eyes / mouth, myositis, pulmonary hypertension	Some have UIP pattern, but NSIP / F, LIP, BOOP, DAD also seen	In some, steroids useful (PM / DM, Sjögren syndrome); in others, cytotoxic therapy indicated	Depends on the associated process, response to therapy

(*continued*)

lymphangitic carcinomatosis, alveolar cell carcinoma, lymphoma, and leukemia.

Disease Caused by Pulmonary Lymphangitic Carcinomatosis

Pulmonary lymphangitic carcinomatosis most often results from adenocarcinomas. The adenocarcinomas usually originate in the breast, gastrointestinal tract, or lung. The symptoms referable to lymphangitic tumor—that is, cough and dyspnea—are the same as those of nonmalignant diffuse infiltrative lung disease. However, lymphangitic carcinomatosis generally progresses much faster than other interstitial lung diseases; survival beyond 3 to 6 months is unusual. The chest radiograph shows bilateral reticular or reticulonodular infiltrates and, often, Kerley B lines. The latter finding should, in the absence of heart failure, raise suspicion that malignancy is responsible for the pulmonary disorder. Pleural effusions and hilar adenopathy may also be observed. Less often, lymphangitic carcinomatosis presents as unilateral involvement, or the patient has an entirely normal chest radiograph despite severe dyspnea and hypoxemia. In patients with lymphangitic carcinomatosis and a normal chest radiograph, tumor cells are often found in abundance in small pulmonary vessels, occasionally resulting in severe pulmonary hypertension (tumor embolism syndrome). Diagnosis is generally made by bronchoscopy, either by BAL alone or by transbronchial biopsy. An alternative means of establishing the diagnosis is by microvascular cytologic examination of blood aspirated through a wedged pulmonary arterial catheter. This technique may be considered in the very breathless patient, in whom right heart catheterization may be better tolerated than bronchoscopy. Open lung biopsy is only occasionally required to diagnose lymphangitic carcinomatosis.

Disease Caused by Alveolar Cell Carcinoma

Alveolar cell carcinoma may present as a simple nodule or mass; as multiple nodules; or with air-space disease that can be focal, multifocal, or quite diffuse.[27] When the tumor presents as diffuse disease, the differential diagnosis includes nonmalignant causes of diffuse alveolar filling, such as pulmonary edema,

Table 5 (*continued*)

Entity	*Clinical and Laboratory Clues*	*Chest Radiographic Clues*	*High-Resolution CT (HRCT) Findings*	*Pulmonary Function*
Eosinophilic granuloma	Age 20–50 years; sometimes asymptomatic; cough and dyspnea	Pneumothorax common, sometimes at presentation; nodules and cysts	Bizarre cyst shapes and associated nodules spread throughout the lung fields but sparing the costophrenic angles and tips of the lingual and middle lobes	Mixed restriction and obstruction with decreased Dl_{CO}
Alveolar proteinosis	Abnormal radiograph, acute complicating infection, or progressive cough and dyspnea; increased LDH	Bilateral alveolar filling that looks like pulmonary edema without signs of congestion	Perihilar areas of patchy or geographic ground-glass opacities with some thickening of the interlobular septa, resulting in so-called crazy-paving pattern	Restrictive defect with decreased Dl_{CO}; increased shunt fraction
Chronic eosinophilic pneumonia (CEP)	Middle-aged women; productive cough, dyspnea, fever, wheezing; eosinophilia	Bilateral peripheral; often upper lung field infiltrates	Patchy unilateral or bilateral peripheral middle- and upper-lobe consolidation, with other areas of ground-glass attenuation	Restrictive and obstructive defects
Anti–glomerular basement membrane (anti-GBM) antibody disease	Young male cigarette smokers; most have hemoptysis; drop in hemoglobin, hematuria, iron deficiency; positive anti-GBM antibody	Fluffy perihilar acinar shadows	Perihilar alveolar filling	Restrictive pattern that improves with clinical resolution; Dl_{CO} increased with bleeding
Lymphangiomyomatosis	Women in childbearing years; dyspnea, hemoptysis, pneumothorax, chylothorax	—	Homogeneously distributed thin-walled parenchymal cysts, varying from a few millimeters to several centimeters in diameter; associated with retrocrural adenopathy, pleural effusion, thoracic duct dilatation, pericardial effusion, and pneumothorax	Obstruction with increased lung volumes; Dl_{CO} decreased

*After ruling out infections, drugs, and inhalants (organic and inorganic).
ACE—angiotensin-converting enzyme ARDS—acute respiratory distress syndrome BHA—bilateral hilar adenopathy DAD—diffuse alveolar damage Dl_{CO}—diffusing capacity of the lung for carbon monoxide ILD—interstitial lung disease LDH—lactate dehydrogenase MCTD—mixed connective tissue disease PM/DM—polymyositis/dermatomyositis PSS—progressive systemic sclerosis SLE—systemic lupus erythematosus UIP—usual interstitial pneumonia

alveolar proteinosis, alveolar sarcoidosis, and alveolar hemorrhage. Diagnosis of alveolar cell carcinoma is usually established fairly easily by bronchoscopy with BAL and transbronchial biopsy.

Disease Caused by Lymphoma or Leukemia

Lymphoma and leukemia may cause diffuse infiltrative lung disease. Lymphomatous involvement of the lung parenchyma is usually seen in association with mediastinal lymphadenopathy and evidence of extrathoracic lymphoma.[28] On occasion, isolated pulmonary disease is the presenting manifestation of lymphoma. When the underlying leukemia is poorly controlled, leukemic infiltration of the lung is usually seen and may contribute to death by causing respiratory failure. Diagnosis of leukemic or lymphomatous infiltration of the lung may require open lung biopsy to adequately characterize the nature of the infiltrating process and to completely exclude infection. However, in some cases, bronchoscopy may suffice for making the diagnosis.

Chronic Diffuse Infiltrative Lung Disease of Unknown Etiology

An overview of diagnosis and treatment for chronic diffuse infiltrative lung disease of unknown etiology is presented [*see Table 5*].

sarcoidosis

Sarcoidosis is a multisystem disease of unknown etiology that is characterized pathologically by the presence of noncaseating granulomas in various tissues.[29] Any organ system may be affected. Most patients are asymptomatic and seek medical attention because of unrelated respiratory symptoms; most of those patients who are asymptomatic are first recognized through a chest radiograph abnormality. Fewer than 5% of patients with sarcoidosis have a normal chest radiograph. Further, the morbidity and mortality that are associated with sarcoidosis result most often from pulmonary involvement. Thus, it is appropriate to think of sarcoidosis as primarily a respiratory tract disorder that may also have features of extrapulmonary disease.

Many infectious and noninfectious conditions are associated with granuloma formation and a histopathologic picture indistinguishable from that of sarcoidosis.[30] Thus, diagnosis of sarcoidosis requires the exclusion of other causes of granuloma formation. Berylliosis and histoplasmosis, for example, can have clinicopathologic features identical to those of sarcoidosis [*see* Approach to the Patient with Suspected Chronic Diffuse Infiltrative Lung Disease, *above*].

Epidemiology

Sarcoidosis occurs worldwide, in all ethnic groups, but there are striking regional and ethnic differences in its incidence and prevalence.[1] For example, in the United States, sarcoidosis is 10 times more common in blacks than in whites and tends to be more severe in black patients. However, this increased prevalence is not observed in blacks in Europe. It has been estimated that the prevalence of sarcoidosis in the United States is between one in 10,000 and one in 2,500 population. Persons 20 to 40 years of age are most often affected, but the disease can present in children and the elderly. Males and females are affected almost equally, with the prevalence of sarcoidosis being only slightly higher in females. Familial sarcoidosis has been reported in a number of kindreds[31]; less commonly, sarcoidosis has

Table 5 (*continued*)

Associated Diseases, Exposures, Risk Factors	*Extrathoracic Manifestations*	*Histopathology*	*Therapy*	*Prognosis*
Cigarette smoking	Bone lesions, diabetes insipidus	Nodular lesions containing Langerhans cells with Birbeck granules	Smoking cessation; steroids not clearly of benefit	Median survival 6 yr; poor prognosis with older age, airflow obstruction, hyperinflation
Most cases idiopathic; associated with acute silica exposure, several hematologic malignancies, busulfan	None	Granular, eosinophilic material within alveolar spaces that stains positively with the periodic acid–Schiff (PAS) staining	Some patients require no therapy; whole lung lavage is treatment of choice; ? GM-CSF	Some patients undergo spontaneous remission; excellent response to whole lung lavage
None	None	Massive, mixed interstitial and alveolar inflammatory infiltrates that have a high eosinophil content	Steroids are effective, producing rapid response; relapse common	Spontaneous remissions in 10%; many require long-term steroids; with steroids, death uncommon from CEP
Cigarette smoking, hydrocarbon exposure	Glomerulonephritis	Alveolar hemorrhage, sometimes with linear immunofluorescent staining; kidney biopsy shows crescentic glomerulonephritis and linear immunofluorescent staining	Immunosuppressive agents, glucocorticoids, plasmapheresis	50% mortality; dialysis needed after 2 yr
Tuberous sclerosis produces identical radiographic picture	Renal angiomyolipomas	Diffuse proliferation of smooth muscle cells in the lung and visceral pleura, with cyst formation	Oophorectomy, progesterone treatment, or a combination of the two; lung transplantation for end-stage disease	25% mortality at 8 yr

been diagnosed in both a husband and his wife, which would suggest a common environmental factor.

Pathogenesis

A specific sequence of events leads to granuloma formation in sarcoidosis[32] [*see Figure 4*]. Expansion of the helper T cell population in the lungs likely occurs through the activation of T cells by macrophages via interleuken-1 (IL-1) and the subsequent release of IL-2 (also termed T cell growth factor) by the helper T cell. IL-2 causes self-replication of the existing T cell population. This population of helper T cells (dominantly of the Th1-type), which are greatly increased in number and in degree of activity, recruits monocytes from peripheral blood into the lung by releasing monocyte chemotactic factor and colony-stimulating factors and thereby participates in the formation of granulomas. The monocytes become tissue macrophages and, ultimately, the epithelioid cells and multinucleated giant cells that form the core of the granuloma. Monocyte- and macrophage-derived fibroblast growth factors may, in some patients, lead to the development of fibrosis. Another effect of the expanded and activated helper T cell population at sites of disease activity is local stimulation of B cells to produce immunoglobulin, which accounts for the hypergammaglobulinemia that is often seen in this disorder.

Clinical Manifestations of Intrathoracic Sarcoidosis

The most common manifestations of intrathoracic sarcoidosis are bilateral hilar lymphadenopathy and diffuse infiltrative lung disease.[33] These cases are categorized according to a five-part staging system, as follows: stage 0, normal chest radiograph (8% of cases at presentation); stage I, bilateral hilar lymphadenopathy alone (40% of cases); stage II, bilateral hilar lymphadenopathy and diffuse infiltrative lung disease (37% of cases); stage III, diffuse infiltrative lung disease alone (10% of cases); and stage IV, lung fibrosis, often with upper lobe cystic disease (5% of cases) [*see Figure 5*].[33]

Symptoms resulting from parenchymal lung involvement include dyspnea and cough. Pulmonary function abnormalities are found in nearly all symptomatic patients and in some patients who are asymptomatic. Physiologic abnormalities in patients with symptomatic sarcoidosis typically consist of a reduction in $D_{L_{CO}}$ and vital capacity without airflow obstruction. The $D_{L_{CO}}$ typically becomes abnormal before the vital capacity becomes abnormal, but both are usually affected in persons with moderate to severe symptoms. Airflow obstruction is relatively uncommon except in patients with advanced disease or with endobronchial involvement of larger airways. In rare instances, diffuse endobronchial granulomas in small airways lead to a predominantly obstructive abnormality in patients who present with stage I disease.

Most patients with stage I sarcoidosis are asymptomatic, and the abnormality is detected on a routine chest radiograph. When stage I disease is symptomatic, the symptoms are nonpulmonary in nature and consist of systemic complaints of fever, malaise, arthralgia, or erythema nodosum. The combination of fever, bilateral hilar lymphadenopathy, arthralgia or arthritis, and erythema nodosum is known as Löfgren syndrome.[34] Approximately 10% of patients who have stage I disease display evidence of extrapulmonary organ involvement (e.g., involvement of the eyes, nervous system, or lacrimal glands).

The frequency of symptoms is higher in patients with stage II sarcoidosis, but asymptomatic cases are not unusual. Symptoms may be primarily systemic, as in stage I disease, or may arise from pulmonary involvement. A presentation of extensive radiographic abnormalities associated with only minimal respiratory symptoms is not uncommon.

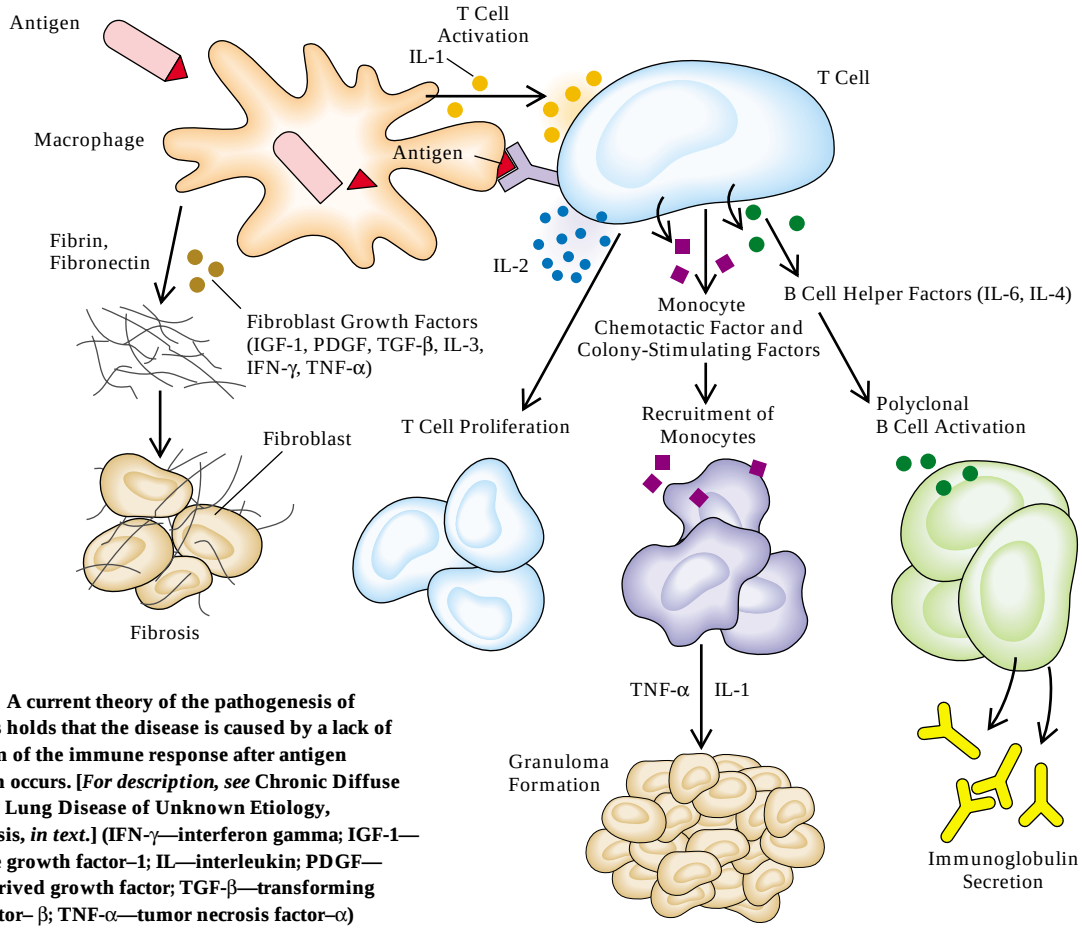


Figure 4 **A current theory of the pathogenesis of sarcoidosis holds that the disease is caused by a lack of modulation of the immune response after antigen recognition occurs. [*For description, see* Chronic Diffuse Infiltrative Lung Disease of Unknown Etiology, Pathogenesis, *in text.*] (IFN-γ—interferon gamma; IGF-1—insulinlike growth factor–1; IL—interleukin; PDGF—platelet-derived growth factor; TGF-β—transforming growth factor– β; TNF-α—tumor necrosis factor–α)**

Stage III disease is found at presentation in 5% to 15% of patients. Respiratory symptoms are common, but as in stage II disease, the chest radiograph may make the patient appear much worse than is actually the case. Certain extrapulmonary manifestations of chronic sarcoidosis, such as infiltrative skin disease, occur much more often in patients with stage III disease than in patients with stage I or stage II disease. This chronic course of stage III disease is not surprising, because patients who are diagnosed at this point in the disease process may represent those persons who earlier had bilateral hilar lymphadenopathy (stages I and II) and failed to undergo spontaneous remission.

Patients with stage IV disease have shrinking of the upper lobes, with upward retraction of the hili, as well as formation of bullae, cysts, and bronchiectasis. Many of these patients will have problems with recurrent bacterial infections and hemoptysis, which is often associated with the development of one or more aspergillomas in the upper lobe cysts.

Unusual manifestations of intrathoracic sarcoidosis include pleural effusion, alveolar infiltrates, large nodular opacities, cavitation, atelectasis, and calcification. Pleural involvement occurs in 1% to 4% of patients and consists of pleural effusion or thickening of the visceral and parietal pleurae. Rarely, chylothorax, hemothorax, or pneumothorax occurs.[35] Effusions are rarely large, and they typically contain a high percentage of lymphocytes. Pleural biopsy reveals granulomas, and the major alternative diagnosis to consider in patients with pleural sarcoidosis is tuberculosis. Alveolar sarcoidosis has variable presentations, ranging from patchy infiltrates in asymptomatic patients to rather extensive air-space consolidation in patients with respiratory failure. Nodules 2 to 10 cm in diameter occur and rarely cavitate. The differential diagnosis of nodular sarcoidosis includes fungal infection, tuberculous infection, and, especially, metastatic tumor. Atelectasis can be caused by endobronchial sarcoidosis. Calcification in lymph nodes is a late manifestation of disease and is seen in 5% of patients.

Necrotizing sarcoid granulomatosis, characterized by masses of confluent granulomas with some degree of vasculitis, is probably a variant of sarcoidosis. The two disorders have many features in common, including hilar lymphadenopathy and extrapulmonary granulomatous inflammation.

Clinical Manifestations of Extrathoracic Sarcoidosis

Although extrathoracic sarcoidosis is a much less common cause of morbidity than intrathoracic disease, it can dominate the clinical picture in some patients.

Erythema nodosum, a nongranulomatous panniculitis, is the most common cutaneous manifestation of sarcoidosis. More than 90% of patients with sarcoidosis who develop erythema nodosum have a stage I radiograph; the other 10% have stage II radiographs. Granulomatous skin lesions occur in 10% to 30% of patients and are especially common in African Americans. Lupus pernio is a potentially disfiguring bluish-purple swollen lesion on the nose, cheeks, earlobes, fingers and toes, lips, or knees. Skin plaques tend to be violaceous, angular, and flat with a raised edge. Psoriasislike plaques may also be seen.

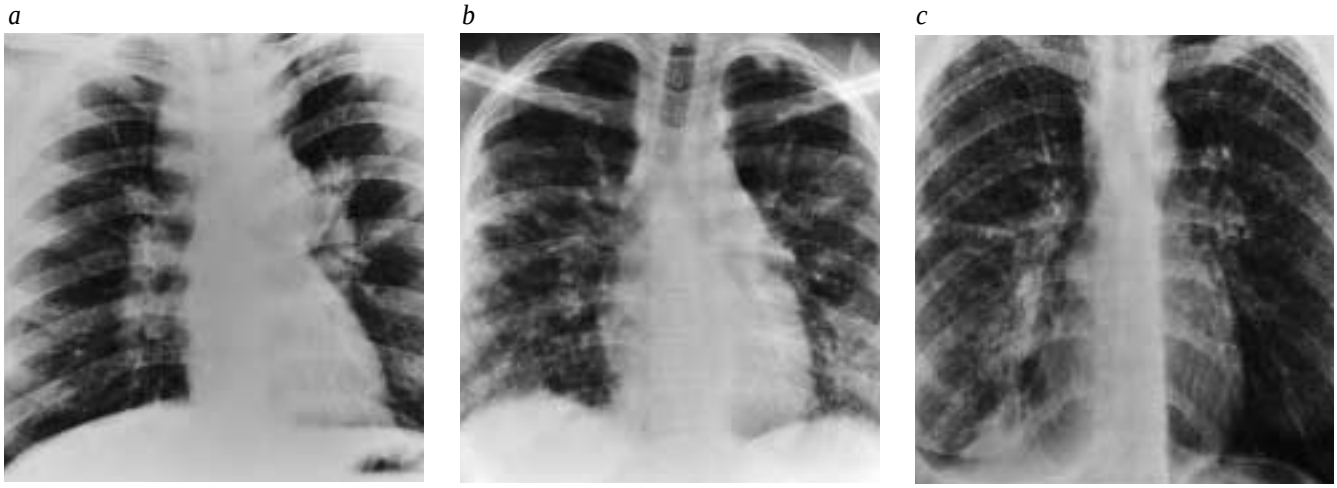


Figure 5 **(*a*) In stage I sarcoidosis, the chest radiograph reveals bilateral hilar and right paratracheal adenopathy with normal lung parenchyma. (*b*) In stage II sarcoidosis, adenopathy and parenchymal infiltration are apparent on the radiograph. (*c*) Stage III sarcoidosis is characterized by parenchymal infiltration, without definite adenopathy.**

Nodules may be elevated or may arise in the subcutaneous tissue. As a rule, these granulomatous skin manifestations of sarcoidosis are seen with chronic and persistent stage III pulmonary disease.

Ocular involvement occurs in about 25% of patients with sarcoidosis. The uveal tract is most often affected; sarcoidosis is responsible for 2% to 4% of all cases of uveitis. In some instances of chronic uveitis, sarcoidosis is suspected but cannot be confirmed.[36] Anterior uveal tract disease with iridocyclitis usually presents as tearing and photophobia with little or no pain. Heerfordt syndrome, or uveoparotid fever, is an uncommon manifestation of sarcoidosis that consists of uveitis, parotid enlargement, cranial nerve palsies, subacute meningitis, and systemic symptoms. Posterior uveitis (chorioretinitis) causes blurring of vision, or patients may be asymptomatic. Chorioretinitis may be difficult to detect when anterior uveitis is present.

The nervous system is affected in fewer than 5% of patients with sarcoidosis. Central nervous system manifestations include chronic meningitis, encephalopathy, hypothalamic lesions, cranial nerve involvement, and seizures. Peripheral sensory and motor neuropathy may be seen. Chronic meningitis in sarcoidosis may be difficult to distinguish from tuberculous or fungal meningitis and is typically associated with cerebrospinal fluid lymphocytosis, increased CSF protein levels, and, sometimes, low CSF glucose levels. Sarcoidosis can give rise to space-occupying lesions in various locations, resulting in diverse neurologic abnormalities. Magnetic resonance imaging is particularly useful in the diagnosis and management of patients with brain parenchymal lesions.[37]

Evidence of cardiac sarcoidosis is seen much more often at autopsy than would be suspected on the basis of clinical manifestations. Rhythm disturbances are the most common manifestations of cardiac sarcoidosis. Both ventricular tachyarrhythmias and complete heart block can occur and can lead to sudden death. Other, less common features of cardiac sarcoidosis include heart failure, pericardial disease, papillary muscle dysfunction, and ventricular aneurysm in the absence of coronary artery disease. Echocardiography, Holter monitoring, and thallium-201 scanning may show abnormalities but do not necessarily indicate that granulomas are in the myocardium. Age, comorbidities, and the types of abnormalities must be taken into account to assess the probability of myocardial sarcoidosis.

Bone or joint involvement is observed in 1% to 10% of patients.[38] Bone disease is characterized by cystic lesions and tends to be restricted to the fingers and toes. Swollen digits may be seen. Synovial sarcoidosis can present as monoarticular or polyarticular chronic arthritis.

Patients with sarcoidosis often have liver and spleen granulomas, but liver dysfunction or massive splenomegaly is unusual. In rare cases, sarcoidosis leads to cirrhosis and portal hypertension. Upper respiratory tract involvement is usually manifested by nasal or laryngeal symptoms. Epistaxis and nasal congestion, which may be mistaken for allergic or vasomotor rhinitis, are common symptoms. Sarcoidosis should be considered when nasal obstruction is refractory to conventional therapy. Laryngeal sarcoidosis involves the supraglottic structures and, on occasion, the vocal cords. Hoarseness is the most common symptom, but upper airway obstruction can occur.

Endocrine abnormalities in sarcoidosis include hypercalcemia[39] and pituitary dysfunction, which may present as diabetes insipidus. Hypercalcemia is seen much less often than hypercalciuria. Hypercalcemia and hypercalciuria in sarcoidosis are caused by the unregulated, increased synthesis of 1,25-dihydroxyvitamin D by activated macrophages in the granulomas, resulting in an increase in calcium absorption in the gut.

Diagnosis

The diagnosis of sarcoidosis is established by demonstration of noncaseating granulomas in tissue [*see Figure 6*] and exclusion of other causes of granulomas. Intrathoracic sarcoidosis is diagnosed most easily by bronchoscopy with transbronchial lung biopsy. The yield of transbronchial lung biopsy depends on the radiographic stage. In patients with pulmonary infiltrates (stages II and III), transbronchial lung biopsy demonstrates noncaseating granulomas in approximately 90% of cases. In stage I disease, the yield is 60% to 70%. If a diagnosis is not established by transbronchial lung biopsy, mediastinoscopy is indicated and will provide a diagnosis in over 95% of cases in which mediastinal adenopathy is present. Biopsy of extrapulmonary tissues can be used for diagnosis when clinically indicated (e.g., in patients with peripheral lymph node enlargement or skin lesions).

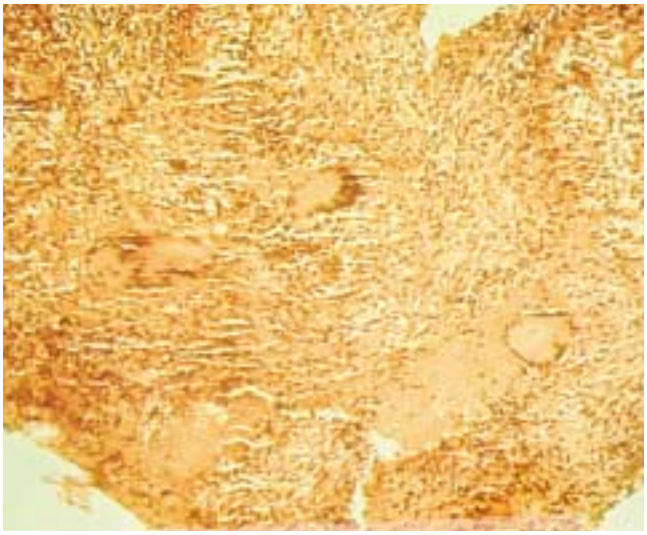


Figure 6 **Noncaseating granulomas can be seen in this tissue biopsy specimen from a patient with sarcoidosis.**

When noncaseating granulomas are found in tissue, possible causes of granuloma formation other than sarcoidosis must always be excluded. Occasionally, lymphoma or carcinoma is associated with granulomatous inflammation in local lymph nodes. Infectious causes of granulomas in the lungs and mediastinal lymph nodes, particularly tuberculosis and histoplasmosis, should be ruled out. A tuberculin skin test, fungal serologic studies, and the use of special stains on tissue biopsy specimens are necessary. For example, such findings as hilar lymphadenopathy, pulmonary infiltrates, and noncaseating granulomas in tissue may prompt the physician to make a presumptive diagnosis of sarcoidosis while these special studies are pending. If a high complement fixation titer for histoplasmosis were detected, the diagnosis would be changed to histoplasmosis.

At the time of bronchoscopy, BAL is usually performed in addition to biopsy. In active sarcoidosis, the lavage liquid usually shows increased numbers of lymphocytes, and the ratio of helper T cells to suppressor-cytotoxic T cells is elevated. Because other pulmonary disorders, both infectious and noninfectious, may be associated with lymphocytosis, this finding lacks diagnostic specificity. Hypersensitivity pneumonitis may be confused with sarcoidosis because both disorders are associated with lavage-liquid lymphocytosis and noncaseating granulomas in tissue. However, an inverted ratio of helper T cells to suppressor-cytotoxic T cells is typically seen in hypersensitivity pneumonitis. Levels of ACE are usually elevated in active sarcoidosis. Therefore, the measurement of ACE levels may aid in the diagnosis of sarcoidosis. Diagnosis of sarcoidosis on the basis of an elevated ACE level alone, without confirmation of the diagnosis by demonstration of noncaseating granulomas in tissue, is not recommended. Elevated ACE levels have also been detected in miliary tuberculosis, leprosy, biliary cirrhosis, silicosis, and asbestosis. With the exception of miliary tuberculosis, it is not likely that these diseases would be clinically confused with sarcoidosis. When the ACE level is elevated, repeated ACE measurements during treatment of sarcoidosis may be helpful because the ACE level often correlates with disease activity.

Differential Diagnosis

The differential diagnosis of stage I sarcoidosis primarily includes lymphoma and granulomatous infections such as tuberculosis and histoplasmosis. More infrequent causes of bilateral hilar lymphadenopathy include metastatic carcinoma (especially renal cell carcinoma) and amyloidosis. Most patients who have lymphoma that presents as bilateral hilar lymphadenopathy have peripheral lymphadenopathy, splenomegaly, or systemic symptoms such as fever, weight loss, and night sweats. Young adults who present with asymptomatic bilateral hilar adenopathy but neither peripheral lymphadenopathy nor splenomegaly are very unlikely to have lymphoma. It is therefore reasonable to make a presumptive diagnosis of stage I sarcoidosis in such patients without confirming the diagnosis by examination of a tissue specimen. However, a tuberculin skin test and fungal serologic studies should be performed. Clinical and radiographic follow-up is required to ensure that there is no progression of the disease. If disease progression is observed, it is imperative to perform a tissue biopsy to establish the diagnosis.

Treatment

Glucocorticoids are standard therapy for symptomatic sarcoidosis, though there is no definitive proof that they influence long-term outcome. Most patients respond very well to therapy, and often, low dosages of prednisone (e.g., 15 mg every other day) are effective in suppressing disease activity. A higher dosage (e.g., 40 to 60 mg / day) is usually given for a few weeks to induce regression of disease activity. The primary indication for glucocorticoids is symptomatic pulmonary involvement. Other indications include significant systemic symptoms (e.g., fever or weight loss), hypercalcemia, and involvement of extrapulmonary tissues that leads to established functional impairment or a risk of organ dysfunction. The required duration of therapy is uncertain. It is common practice to treat pulmonary sarcoidosis with prednisone for 1 year, then to withdraw the drug to see whether the disease is in remission. Relapses of the disease are treated in a similar fashion, with higher initial dosages of prednisone that are tapered to a low dosage for maintenance. Some patients experience relapses repeatedly as prednisone is withdrawn and, as a result, require low-dose suppressive therapy indefinitely.

If the use of steroids is strictly contraindicated or if the patient does not respond to steroid therapy, methotrexate[40] or chlorambucil may be tried. Chloroquine is often useful in patients with skin disease.

Lung transplantation in sarcoidosis has been associated with histologic evidence of recurrence in the transplanted lung, but the survival of these patients has been comparable to the survival of patients undergoing lung transplantation for other indications.[41]

Prognosis

The natural history of sarcoidosis is highly variable.[20] In 60% to 70% of all cases, spontaneous remission will occur; in 10% to 30%, the disease will have a progressive course. The mortality is 1% to 5%; death usually results from progressive lung disease, neurosarcoidosis, or cardiac disease. Other adverse prognostic features include lupus pernio, chronic uveitis, onset after 40 years of age, hypercalcemia or nephrocalcinosis, black race, nasal mucosal involvement, and cystic bone disease.

Despite many investigative efforts over the past 30 years, no test has been shown to be better than the chest radiographic staging system in predicting the clinical course of sarcoidosis. Approximately 75% of patients who have stage I sarcoidosis will undergo spontaneous remission within 2 years after presentation. Erythema nodosum at initial presentation increases the likelihood of spontaneous remission. Patients in whom bilateral hilar lymphadenopathy fails to remit within 2 years may remain stable, may undergo spontaneous remission at a later time, or may develop progressive pulmonary disease. No more than 10% to 15% of patients develop progressive pulmonary disease. Roughly 50% of patients with stage II sarcoidosis will be in remission 2 years after presentation. Stage II disease is more likely than stage I disease to be progressive and to follow a chronic symptomatic course than stage I disease. Only one third of patients with stage III sarcoidosis will be in remission 2 years after presentation. The disease commonly takes a chronic course that eventually leads to lung fibrosis (stage IV). Patients with stage IV disease have irreversible disease, though many such patients have clinically inactive disease and therefore do not require immunosuppressive treatment.

idiopathic pulmonary fibrosis

IPF, also known as cryptogenic fibrosing alveolitis, is one of the most common and most serious causes of chronic diffuse infiltrative lung disease.[1,42] Although this disorder has characteristic clinical, radiographic, physiologic, and histopathologic features, none of these features is pathognomonic. Therefore, a diagnosis of IPF requires not only a compatible clinicopathologic picture but also the exclusion of all other causes of diffuse disease. For example, the pulmonary manifestations of IPF may be indistinguishable from diffuse disease associated with collagen vascular disorders; asbestosis; chemotherapy-induced lung fibrosis and several closely associated disorders; bronchiolitis obliterans organizing pneumonia (BOOP); lymphocytic interstitial pneumonia (LIP); desquamative interstitial pneumonia / respiratory bronchiolitis interstitial lung disease (DIP / RBILD); acute interstitial pneumonia (AIP); and nonspecific interstitial pneumonia / fibrosis (NSIP / F).[43]

Pathogenesis

The pathogenesis of IPF is not well understood and is the topic of continued debate.[44,45] A viral trigger has been suspected of playing a role in certain instances because a substantial minority of patients date the onset of their respiratory symptoms to a flulike illness. Cigarette smoking; chronic aspiration; exposure to various environmental factors (metal and wood dusts); latent viral infections, especially from the herpesvirus family; and use of antidepressant drugs have been suggested as risk factors.[42] A genetic susceptibility may also be important, as evidenced by familial clustering,[46] the recent identification of a candidate gene,[47] and the occurrence of pulmonary fibrosis in the inherited disorder Hermansky-Pudlak syndrome.[48] It may be that IPF results when a genetically susceptible person is exposed to a viral or environmental agent that triggers a set of events leading to a cascade of inflammatory, possibly immune, events that result in fibrosis.

Whatever the stimulus, the pathology of IPF suggests that the lung is subjected to repeated episodes of injury that lead to activation of pathways resulting in pulmonary fibrosis [*see Figure 7*]. Neutrophilic inflammation, previously thought to play an important role in initiating the process, is probably an epiphenomenon.[49] There may be a role for T helper type 2 (Th2) lymphocytes and their cytokine products in the process. The current hypothesis[42,49] is that repeated episodes of injury to the epithelium results in formation of a microscopic lesion—in essence, a wound—in the wall of the alveolus, with the development of a fibrin clot in the alveolar space and disruption of the basement membrane. The alveolar cells secrete growth factors (platelet-derived growth factor [PDGF], transforming growth factor–β [TGF-β], and tumor necrosis factor–α [TNF-α]) that induce migration and differentiation of fibroblasts. These cells produce gelatinases that further disrupt the normal extracellular matrix and the basement membrane. Angiogenic factors (vascular endothelial growth factor [VEGF] and fibroblast growth factor–2 [FGF-2]) induce new-vessel formation. An imbalance of increased synthesis of extracellular matrix proteins, mainly collagen, and decreased collagenolytic activity results in the progressive deposition of matrix. Deficient apoptosis of fibroblasts and increased epithelial cell death impair reepithelialization. This aberrant wound healing, as it may be termed, leads to the development of pulmonary fibrosis.

Clinical Manifestations

As with other types of diffuse infiltrative lung disease, dyspnea on exertion and cough are the most prominent symptoms. Physical examination typically reveals fine, dry inspiratory rales. Cyanosis may be noted when hypoxemia is severe, and features of cor pulmonale (e.g., elevated jugular venous pressure, edema, and a prominent pulmonic second heart sound) indicate advanced disease. Clubbing of the digits is noted in 25% to 50% of cases but is often absent early in the course of the disease. There are no other extrathoracic manifestations of disease. Clinical evidence of extrapulmonary disease (e.g., arthritis, skin disease, or serositis) suggests a systemic disorder, such as sarcoidosis or one of the collagen vascular diseases, rather than IPF.

Diagnosis

A diagnosis of IPF is established by correlating histopathologic and clinical findings. Clinically, IPF is suspected when a diffuse infiltrative lung disease occurs with no involvement of other organ systems and there is no apparent relation to infection, environmental exposure, or drugs. The chest radiograph shows no evidence of hilar adenopathy or pleural effusion. Laboratory tests in IPF are generally unrevealing, except for the presence of a positive antinuclear antibody or positive rheumatoid factor in up to 50% of cases. Thus, serum antinuclear antibody and rheumatoid factor cannot be relied on to differentiate IPF from diffuse infiltrative lung disease associated with collagen vascular disorders if the titers are low.

The initial approach to diagnosing IPF is to exclude as many causes of diffuse infiltrative lung disease as possible by a combination of history and clinical examination. However, a number of causes remain that cannot be reliably identified without a lung biopsy. Such causes include stage III sarcoidosis, lymphangitic carcinomatosis, lymphangioleiomyomatosis, eosinophilic granuloma, DIP / RBILD, NSIP / F, LIP, AIP, and the more diffuse forms of BOOP.

Imaging studies The radiographic features are nonspecific and most often consist of a bilateral reticular or reticulonodular pattern that typically appears in the lower lung fields; if cor pulmonale develops, enlargement of the pulmonary arteries and hypertrophy of the right ventricle occurs. In as many as

10% of patients with symptomatic IPF, the chest radiograph may be entirely normal. Radiographic findings in patients with IPF are limited to the lung fields. Hilar lymphadenopathy or pleural effusion suggests a different cause of diffuse infiltrative lung disease.

HRCT may be helpful at this stage of the evaluation because the pattern of abnormalities seen in IPF is specific, and many of these other causes have characteristic and very different patterns.[6] Also, HRCT is likely to reveal abnormalities in cases in which the chest radiograph is normal.

The HRCT pattern in IPF includes patchy, peripheral, subpleural, bibasal reticular abnormalities with minimal ground-glass opacity. With advanced disease, honeycomb lung and traction bronchiectasis / bronchiolectasis, indicating end-stage lung fibrosis, are seen. Extensive areas of ground-glass haziness, indicating an acinar filling process, are not characteristic of IPF but may be seen in AIP, LIP, DIP / RBILD, and NSIP / F.

Pulmonary function testing The physiologic abnormalities of IPF are basically the same as those described for diffuse infiltrative lung disease. The classic composite physiologic picture of IPF is reduced $D_{L_{CO}}$, restriction of lung volume, exercise-induced oxygen desaturation, and absence of airflow obstruction. Some of the alterations correlate with the pathologic findings.[50]

Histopathologic features The histopathologic features of IPF are nonspecific. Similar histopathologic findings are observed in diffuse infiltrative lung disease associated with collagen vascular disease, various types of drug-induced diffuse disease, chronic hypersensitivity pneumonitis, asbestosis, and other disorders. Chronic fibrosis and a scant inflammatory cell infiltrate, the essential pathologic features of IPF, are common reactions to a variety of agents.

The histopathologic picture in IPF is termed usual interstitial pneumonitis (UIP) and is characterized by minimal interstitial inflammatory round cell infiltrate, widening of alveolar septa, and fibrosis with fibroblastic foci[43] [*see Figure 8*]. The distribution of the lesion is irregular: areas of intense fibrosis can coexist with areas of near-normal lung in the same open lung biopsy specimen. In some cases, more than one histologic pattern is seen; in such cases, if any areas of UIP are seen, the clinical diagnosis should be IPF.[51]

UIP must be differentiated from DIP / RBILD, AIP, LIP, NSIP / F, and BOOP, because of different prognoses [43] and treatment approaches (see below).

Lung biopsy Transbronchial lung biopsy is a sensitive tool for diagnosing sarcoidosis and lymphangitic cancer and may reveal pathologic features suggestive of IPF (i.e., fibrosis and widened alveolar septa with scant inflammatory cell infiltrate). Open lung biopsy is needed to exclude other causes of diffuse disease. It is up to the clinician to decide whether to proceed to open lung biopsy when clinical evaluation and transbronchial lung biopsy are highly suggestive of IPF. Open lung biopsy is advisable when there is any doubt about whether there is an infectious cause. Open lung biopsy should usually be performed in younger patients to establish the diagnosis of the underlying disorder with a reasonable degree of certainty.[5] Older patients with typical clinical features and a compatible transbronchial lung biopsy result may reasonably be spared the morbidity associated with open lung biopsy. In certain cases, it may be reasonable to diagnose IPF without any tissue biopsy. For example, a tissue biopsy may not be needed to diagnose IPF in an elderly patient who has had progressive dyspnea on exertion for many months to years and who has fine, end-inspiratory crackles and clubbing of the digits on examination, bilateral coarse interstitial infiltrates, low-titer serum antinuclear antibody, and no evidence of an environmental or a drug-related disorder.[52]

Prognosis and Treatment

The median survival for patients with newly diagnosed IPF is 2 to 3 years. Most deaths result from progressive pulmonary impairment. Adverse prognostic factors include cigarette smoking, severe dyspnea, low lung compliance, and high numbers of fibroblastic foci in the lung biopsy specimen.[53] Patients with IPF are also at increased risk for carcinoma of the lung.[54]

Corticosteroids have previously been used in the initial management of IPF. However, reviews of the lung biopsies of the patients in the studies supporting steroid use have suggested that the 20% to 30% of patients responding to steroids probably did not have UIP by the current pathologic criteria.[55]

Cyclophosphamide and azathioprine have been used in IPF, often in conjunction with prednisone or in patients in whom prednisone failed. The rationale for the use of these agents is similar to that for the use of steroids. A recent prospective trial found that cyclophosphamide had limited efficacy in a group of IPF patients in whom steroid therapy had failed or had produced unacceptable side effects.[56] Only one study provides evidence that azathioprine has any significant efficacy.[57] Current opinion holds that these drugs are more toxic than beneficial in IPF.[42] Colchicine, a drug thought to have antifibrotic properties, was not found to affect survival in one study.[58] Single-lung transplantation has been used successfully for end-stage lung fibrosis.[59] Patients who are potential transplantation candidates and whose $D_{L_{CO}}$ has dropped below 40% should be referred to a transplant center.[60]

Alternative therapies for IPF have been proposed.[61] A trial of interferon beta was unsuccessful. In a randomized controlled trial, a combination of interferon gamma-1b and low-dose prednisolone given to IPF patients whose condition was unresponsive to steroids or other immunosuppressive agents led to greater improvement in lung function over 12 months than was seen with prednisolone alone.[62] A larger randomized trial of interferon gamma-1b is under way.

desquamative interstitial pneumonia / respiratory bronchiolitis interstitial lung disease

Patients with DIP / RBILD present with cough and dyspnea of insidious onset.[43,63] Clubbing of digits is present in about 50% of cases. The average age at onset of symptoms is the early to middle 40s, which is significantly younger than the age at onset of IPF symptoms. Men are affected twice as often as women. Almost all of these patients are cigarette smokers, suggesting that the disease may be caused or initiated by tobacco use.

The chest radiograph typically shows vague, bibasilar opacities that correlate with ground-glass densities on HRCT. Some studies have found more reticulonodular and linear changes.[43,63]

The lung biopsy specimen in DIP / RBILD is characterized by a fairly homogeneous pattern in which alveolar spaces are filled with pigmented alveolar macrophages. These accumulations of macrophages are often accentuated in peribronchiolar air spaces, sparing the more distal air spaces [*see Figure 9*].[43,63]

Discontinuance of smoking has been associated with im-

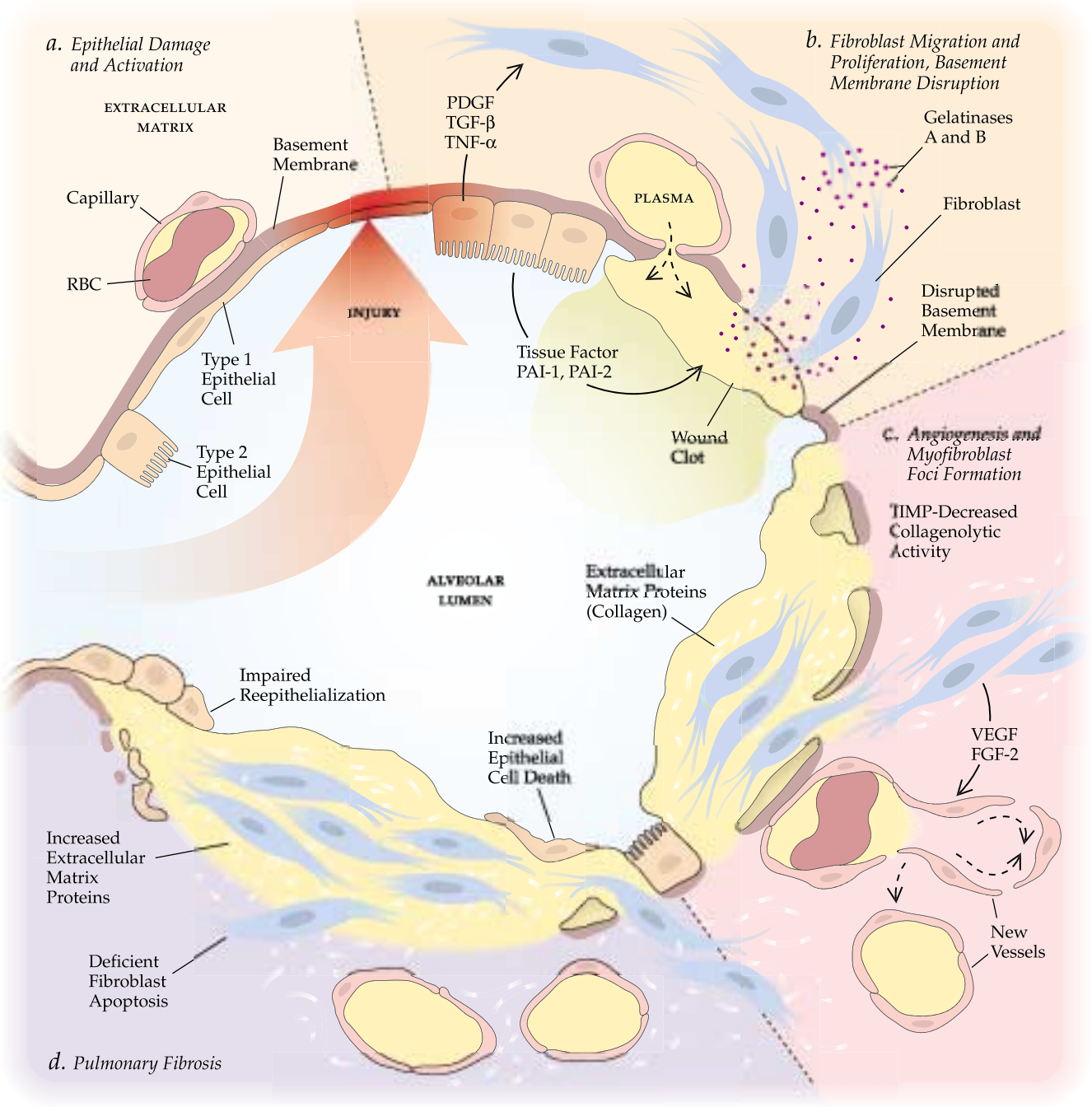


Figure 7 **The current hypothesis is that idiopathic pulmonary fibrosis (IPF) originates with repeated episodes of injury to the epithelium that result in formation of a microscopic lesion—in essence, a wound—in the wall of the alveolus (*a*). A fibrin clot then develops in the alveolar space, and the alveolar cells secrete growth factors (PDGF, TGF-β, TNF-α) that induce migration and differentiation of fibroblasts; in turn, the fibroblasts produce gelatinases that further disrupt the normal extracellular matrix and the basement membrane (*b*). Angiogenic factors (VEGF, FGF-2) induce new vessel formation. Small foci of myofibroblasts appear (*c*). An imbalance between increased synthesis of extracellular matrix proteins, mainly collagen, and decreased collagenolytic activity results in the progressive deposition of matrix, and deficient apoptosis of fibroblasts and increased epithelial cell death impair reepithelialization. The result of this aberrant healing process is pulmonary fibrosis (*d*). (FGF—fibroblast growth factor; MMP—matrix metalloproteinase; PAI—plasminogen activator inhibitor; PDGF—platelet-derived growth factor; TGF—transforming growth factor; TIMP—tissue inhibitor of metalloproteinase; TNF—tumor necrosis factor; VEGF—vascular endothelial growth factor)**

provement. Many patients improve without therapy other than smoking cessation. Steroids have been associated with a beneficial response in about 60%. The mortality is 20% to 30%, and the mean survival is 12 years.

acute interstitial pneumonia

Acute interstitial pneumonia is a distinct idiopathic condition that produces respiratory failure of rapid onset.[43,64] AIP is also known as Hamman-Rich syndrome.

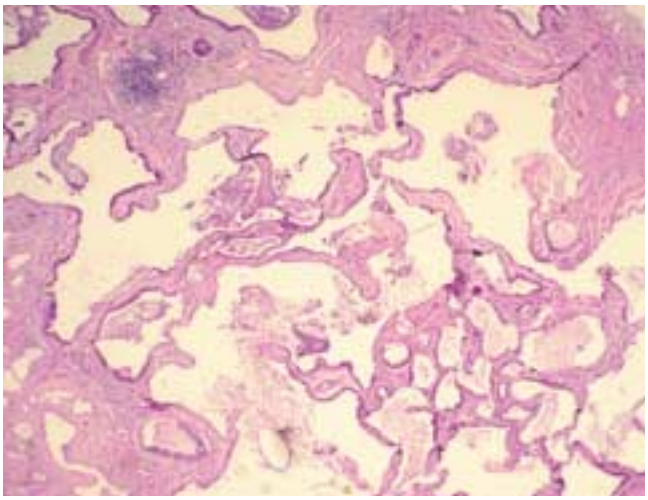


Figure 8 **Usual interstitial pneumonia is characterized by alveolar septal thickening, caused by fibrosis; minimal inflammatory cell infiltrate; honeycombing; and irregular involvement from one region to the next.**

The clinical presentation of AIP is similar to that of ARDS except that, in patients with AIP, no predisposing factor can be identified. A viral prodrome often precedes onset. Over a period of less than 30 days, a symptom complex of dyspnea, cough with mucoid sputum, and fever evolves into respiratory failure. Imaging studies by routine chest radiograph or HRCT show diffuse pulmonary infiltrates with ground-glass changes and consolidation.

The diagnosis is made by the recognition of a clinical illness compatible with ARDS in the absence of a predisposing factor after the exclusion of other alveolar-filling diseases, such as infectious pneumonia (especially PCP), alveolar hemorrhage, and eosinophilic pneumonia. The diagnosis is usually made through bronchoscopy with bronchoalveolar lavage.

AIP is characterized pathologically by the findings of diffuse alveolar damage. It cannot be differentiated from ARDS caused by sepsis, toxins, or shock [*see Figure 10*]. AIP evolves through the same sequence of pathologic patterns as ARDS [*see 14:X Pulmonary Edema*].

Treatment is largely supportive, through use of oxygen with either noninvasive or invasive mechanical ventilation. Steroids are advocated, but their use is not supported by controlled trials.

The mean 6-month mortality is 33% to 78%. Survivors may have recurrences or develop chronic progressive interstitial lung disease.

lymphocytic interstitial pneumonia

LIP is an interstitial lung disease characterized by diffuse or localized lymphocytic infiltration of the alveolar and interstitial areas of the lung.[43,65] This disorder can occur in association with a number of autoimmune processes (especially Sjögren syndrome), dysproteinemias, immunodeficiency (AIDS and common variable immunodeficiency), drug reactions (e.g., to phenytoin), or bone marrow transplantation; or it can occur as an idiopathic process.

LIP occurs more often in women than in men. The mean age at onset is 56 years. Patients present with dyspnea and cough, but the associated disorder may dominate the clinical picture. Rales and lymphadenopathy are common physical findings.

Imaging studies typically show reticulonodular infiltrates. The presence of hilar or mediastinal adenopathy or pleural effusion suggests lymphoma. HRCT will usually show a mixture of interstitial and alveolar (ground-glass) changes. Thin-walled cysts are often present.

Bronchoscopic specimens can be useful, showing a striking lymphocytosis (mostly B cells) on BAL and revealing lymphocytic infiltration on transbronchial lung biopsy [*see Figure 11*]. The pathologist will often need to use special studies to differentiate LIP from neoplastic forms of lymphocytic infiltration of the lung.

Patients with LIP should be treated for any underlying disorder. Some patients respond to steroids, though many require immunosuppressive agents.

nonspecific interstitial pneumonia / fibrosis

NSIP / F has often been confused with IPF. In retrospective analyses of patients who had been diagnosed with IPF, many were found to have NSIP / F and were noted to have a better prognosis than patients with IPF.[43]

NSIP / F occurs in middle-aged adults, with a mean age at onset of 49 years, but it can also affect children and older adults. There is a slight female predominance. Although some cases are idiopathic, many patients have collagen vascular diseases or immunodeficiency (including HIV infection), and some have a history of environmental or therapeutic drug exposures. A few patients have a history of an acute lung injury, such as that caused by pneumonia, ARDS, or surgery.

Dyspnea, cough, and sometimes fever are the dominant clinical features, with imaging studies showing bilateral interstitial infiltrates. HRCT scanning demonstrates bilateral, patchy areas of ground-glass abnormality, but this pattern is also seen with other disorders and therefore does not permit the confident diagnosis of NSIP / F.

NSIP / F can be diagnosed only by lung biopsy. As with IPF, transbronchial lung biopsy is not likely to produce an adequate specimen, so open biopsy is required. Pathologically, NSIP / F is characterized by a temporally uniform mixture of alveolar wall inflammation (mostly lymphocytes and plasma cells) and fibrosis [*see Figure 12*].

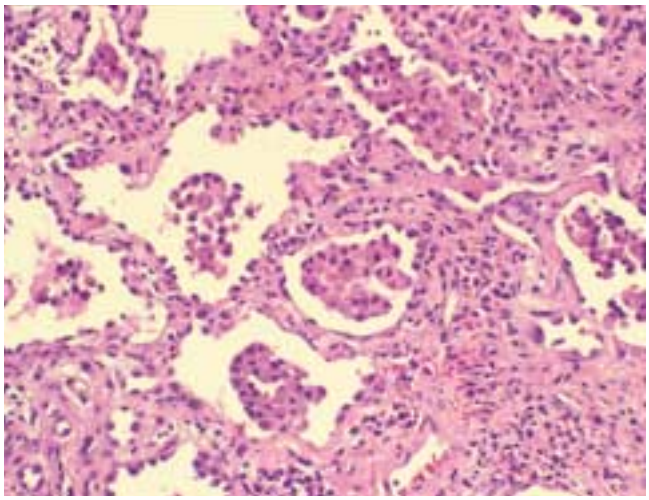


Figure 9 **Desquamative interstitial pneumonia/respiratory bronchiolitis interstitial lung disease. The alveoli and septal walls are filled with macrophages, and no fibrosis is apparent.**

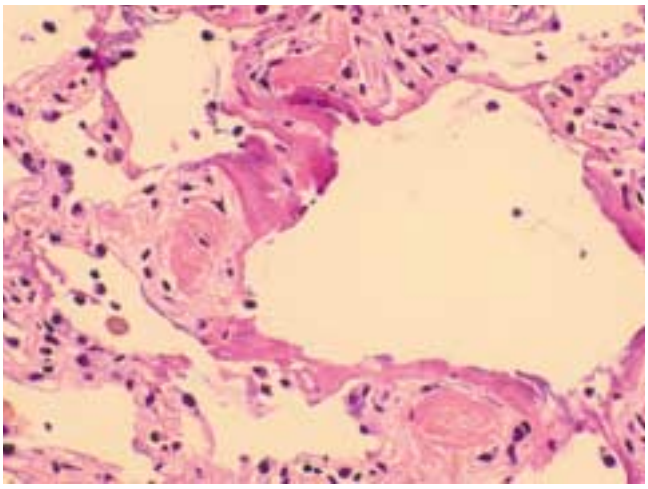


Figure 10 **Acute interstitial pneumonia. The biopsy specimen shows diffuse alveolar damage.**

Steroids appear to provide improvement in more than half of the patients with NSIP/F, and the prognosis for patients with NSIP/F is significantly better than that for patients with IPF.[66] Some patients experience relapses, and a minority have a progressive illness despite treatment.

bronchiolitis obliterans organizing pneumonia

BOOP [*see 14:IV Focal and Multifocal Lung Disease*], also called cryptogenic organizing pneumonitis,[43,67] occurs in the interstitium, within the alveoli, and in the small airways. BOOP is another disorder sometimes incorrectly diagnosed as IPF. The disease may result from any number of secondary causes, including proximal airway obstruction, reaction to drugs or inhalation of toxic fumes (e.g., oxides of nitrogen), viral infections, aspiration, lung transplantation, inflammatory bowel disease, collagen vascular disorders, and other lung diseases, such as eosinophilic lung disease, IPF, and hypersensitivity pneumonitis.

Clinical Manifestations

BOOP occurs in three clinical patterns: symptomatic idiopathic, secondary, and focal asymptomatic.[68] Although a thorough history sometimes identifies one of the secondary causes mentioned above, in some patients with idiopathic BOOP, an open lung biopsy is required to distinguish the disease from other diffuse parenchymal processes.

Clinical onset of BOOP usually occurs at 50 to 60 years of age. The illness frequently begins abruptly and is characterized by dry cough and dyspnea; fever, malaise, and fatigue are more common in BOOP than in IPF. Physical examination sometimes discloses rales that are often associated with wheezes. Clubbing of the digits, which is common in IPF, is absent.

Diagnosis

The diagnosis of BOOP is made by the correlation of the clinical findings with the results of imaging studies, particularly HRCT. Lung biopsy is often required in patients with idiopathic BOOP.

The chest radiograph typically shows bilateral patchy alveolar or ground-glass infiltrates, in contrast to the interstitial infiltrates that occur in IPF. HRCT often shows bilateral areas of consolidation on air bronchograms—involving mainly the lower-lobe subpleural region, the peribronchovascular region, or both—that are quite suggestive of the diagnosis.[6] Pulmonary function tests demonstrate obstruction, restriction, or a mixed pattern with decreased diffusing capacity and hypoxemia.

The lung biopsy in BOOP is characterized by the presence of intraluminal plugs of connective tissue in the bronchioles, alveolar ducts, and alveolar spaces [*see Figure 13*].

Treatment

Treatment with glucocorticoids has been quite successful, and many patients respond dramatically within days after starting therapy. When glucocorticoids are withdrawn, however, many patients experience a relapse, with recurrent systemic symptoms, infiltrates, and hypoxemia. Response to re-treatment is often good in such cases. Patients with focal asymptomatic BOOP do not require therapy.

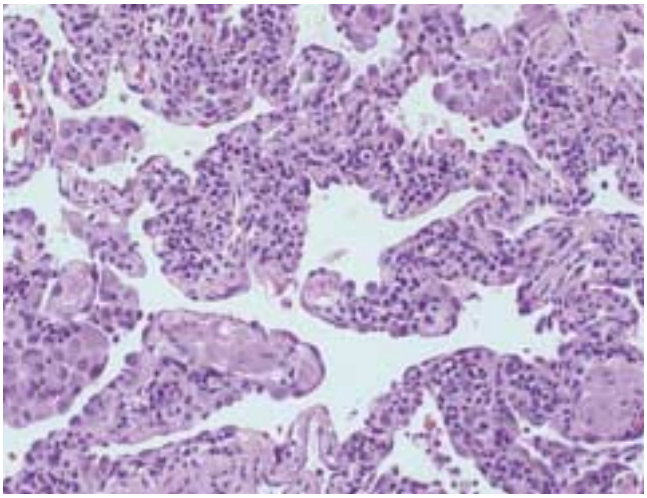


Figure 11 **Lymphocytic interstitial pneumonia. The biopsy specimen shows infiltration of the alveolar walls and alveoli with mature lymphocytes.**

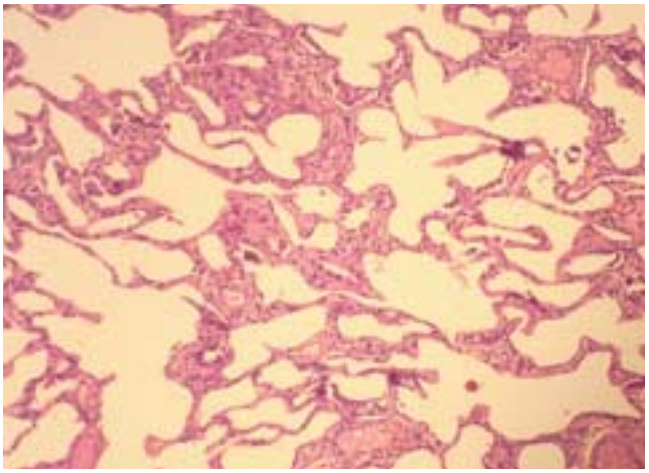


Figure 12 **Nonspecific interstitial pneumonia/fibrosis. The biopsy specimen shows a temporally uniform mixture of alveolar wall inflammation (mostly lymphocytes and plasma cells and fibrosis).**

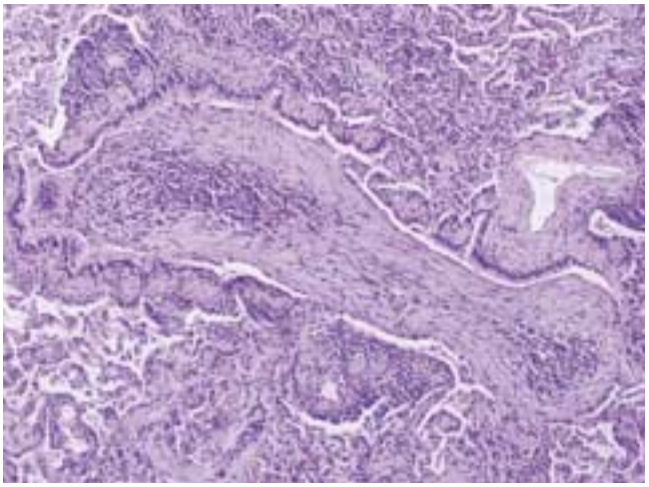


Figure 13 **Bronchiolitis obliterans organizing pneumonia. The biopsy specimen shows intraluminal plugs of connective tissue in the bronchioles, alveolar ducts, and alveolar spaces.**

collagen vascular diseases with associated pulmonary involvement

Collagen vascular diseases commonly affect the lung.[69] As many as two thirds of patients have clinical evidence of pleuropulmonary involvement, and nearly all have abnormal findings at autopsy. The immune mechanism is thought to be immune complex deposition, but the evidence that this mechanism is at work is more convincing in some collagen vascular disorders than in others. Pulmonary function tests in patients with interstitial infiltrates and fibrosis show restriction, small airway obstruction, and reduced $D_{L_{CO}}$ with hypoxemia at rest and worsening hypoxemia with exercise. HRCT of the chest often detects subtle interstitial disease that cannot be seen by routine chest radiography. Although the chronic interstitial disease seen in some of these disorders is pathologically indistinguishable from IPF, patients with a collagen vascular disease seem to have a better prognosis than those with IPF[70]; this is true even for those patients in whom collagen vascular disease develops after IPF.[71] Other patients with collagen vascular disease will have a pathologic pattern consistent with one of the other IPF-like disorders, such as LIP, NSIP / F, or BOOP.

Systemic Lupus Erythematosus

SLE is one of the most prevalent of the collagen vascular diseases, especially in African-American women of childbearing age; however, as many as 18% of cases occur after the fifth decade of life. Drug-induced SLE is a common clinical problem, especially in elderly patients with heart disease. Drugs that commonly cause SLE include hydralazine, procainamide, isoniazid, phenytoin, quinidine, methyldopa, and several of the beta-blocking agents. A few other drugs have been implicated, but the evidence is not as strong.

Pleuropulmonary complications of SLE occur frequently (in 38% to 89% of cases) and often must be differentiated from infections.[69,72] Pleural disease, atelectasis, acute lupus pneumonitis, and chronic intersititial pneumonitis are the most common clinical problems. Less commonly, alveolar hemorrhage or pulmonary vascular disease (acute reversible hypoxemia syndrome, vasculitis, or thromboembolism) may occur. Pleuropulmonary disease is frequently the presentation of patients with drug-induced SLE.

Acute lupus pneumonitis initially presents as sudden onset of cough and dyspnea. Radiographically, it appears as patchy basilar opacities. Occasionally, radiography shows only horizontal linear shadows that have been attributed to small areas of atelectasis; patients in whom this occurs usually respond well to glucocorticoids.

Chronic interstitial pneumonitis occurs as a complication in about 10% of patients with SLE. Many of the patients with chronic interstitial pneumonitis have mild, often asymptomatic disease detectable only by lung function testing and HRCT. Some cases are the result of prior episodes of acute lupus pneumonitis. Pathologically, most of these cases are similar to cases of NSIP / F. Rarely, chronic interstitial pneumonitis evolves into severe pulmonary fibrosis. Steroids and other immunosuppressive agents may be beneficial, but no trials have been performed to investigate this.

Patients with SLE may present with overwhelming alveolar hemorrhage or have occult hemorrhage during the course of their disease. Severe alveolar hemorrhage in SLE is associated with a 70% mortality. The symptoms are cough and dyspnea with or without hemoptysis; chest radiographs show fluffy alveolar infiltrates that may be patchy or confluent. Treatment is similar to that of Goodpasture syndrome and includes corticosteroids, plasmapheresis, and immunosuppressive therapy [*see 10:V Glomerular Diseases*].

Rheumatoid Arthritis

Although the incidence of rheumatoid arthritis is higher in women than in men, pleuropulmonary complications occur more frequently in men than in women.[69,73] Smokers are also at increased risk. Pleuropulmonary complications are clinically apparent in up to 14% of patients within 2 years of presentation[74] and in approximately 50% of patients over the course of the illness; a much higher percentage have pathologic involvement. In occasional cases, thoracic involvement becomes apparent before the articular disease, making diagnosis difficult.

The most frequent pleuropulmonary problems are interstitial lung disease, bronchiolitis, pleural disease, and parenchymal rheumatoid nodules. Less common problems include chronic airflow obstruction, bronchiectasis, BOOP, bronchiolitis obliterans, and pulmonary vasculopathy. These problems are usually associated with active arthritis, high titers of rheumatoid factor, circulating immune complexes, and cryoglobulinemia.

Diffuse interstitial pneumonitis with fibrosis is the most common and serious pulmonary problem in patients with rheumatoid arthritis, seen in up to 58% of patients if sensitive methods of detection are used. Other types of infiltrative lung disease—especially related to drugs commonly used in the treatment of rheumatoid arthritis, such as methotrexate and gold—must be considered. Dyspnea, sometimes with cough, is the most common symptom. Clubbing of the fingers may be present. The chest radiograph evolves from fine nodularity to coarse reticulation and finally to a honeycomb pattern. HRCT shows ground-glass attenuation and basal honeycombing, sometimes associated with pleural disease, bronchiectasis, or emphysema. Results of BAL may be abnormal, demonstrating increased numbers of lymphocytes, neutrophils, eosinophils, or all three; this variability has not been shown to have a bear-

ing on management. The pathologic patterns on open lung biopsy specimens include UIP, DIP, NSIP/F, LIP (in cases complicated by Sjögren syndrome), and BOOP. In patients with interstitial disease that is characterized by increased cellularity in lung biopsy specimens or the presence of progressive symptoms, glucocorticoids should be administered. Other drugs that have been anecdotally reported to benefit these patients include azathioprine, cyclophosphamide, methotrexate, and D-penicillamine.[73]

Rheumatoid (necrobiotic) nodules may be found in many tissues, including the lung parenchyma, endobronchial mucosa, and pleurae; in any of these locations, the nodules may cavitate. These lesions have a characteristic histopathologic appearance on microscopic examination; unless percutaneous transthoracic needle biopsy or open lung biopsy is performed, such lesions can be difficult to differentiate from those of malignancy or tuberculosis. Most rheumatoid nodules do not cause symptoms, and treatment is rarely necessary once a diagnosis is established.

Progressive Systemic Sclerosis

Progressive systemic sclerosis (PSS), also called scleroderma, is a disorder that predominantly affects women in the fourth to sixth decades of life. The prognosis is unfavorable, especially in African Americans, men, and patients with pulmonary disease.

Pulmonary complications, particularly interstitial lung disease, occur in as many as 90% of confirmed cases of PSS.[69,75] The presence of the serum autoantibody antitopoisomerase (Scl-70) is associated with the development of infiltrative lung disease. Interstitial lung disease in PSS is characterized by basilar reticular or reticulonodular fine infiltrates that become coarser as the disease progresses. In addition, radiography may show a loss of lung volume over time, and pneumothorax may occur. Evidence of pulmonary hypertension may be greater than would be expected from the degree of radiographic abnormality or pulmonary function disturbance, especially when Raynaud phenomenon is part of the clinical syndrome.

Glucocorticoids are not useful as therapy for interstitial lung disease associated with PSS; cyclophosphamide may be effective in patients who are shown by BAL to have active alveolitis.[69,75] As a result of esophageal motility disturbance, aspiration of gastric and esophageal contents often occurs; aspiration occurs more frequently when patients are in the recumbent position. Aspiration can be prevented by elevating the upper body to the near-upright position.

CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia) is a variant of PSS. Patients with CREST syndrome have prominent pulmonary vascular disease. This pulmonary hypertension may be associated with interstitial fibrosis, as is seen in PSS, or it may occur as an isolated finding [*see 15:V Scleroderma and Related Diseases*].

Polymyositis and Dermatomyositis

Polymyositis and dermatomyositis are autoimmune disorders that are characterized by weakness and occasionally by pain in proximal limb and neck muscles, skin rash (dermatomyositis), and neoplasm. The most frequent pulmonary complications of polymyositis and dermatomyositis are respiratory failure caused by respiratory muscle weakness, aspiration secondary to posterior pharyngeal and proximal esophageal weakness, and interstitial lung disease.[69,76]

Interstitial lung disease is seen in a minority of patients and may be associated with anti–Jo-1 antibody, an autoantibody that is specific to the cellular enzyme histidyl-tRNA synthetase. These patients may have BOOP, NSIP/F, UIP, or diffuse alveolar damage. In contrast to the interstitial disease associated with many of the other connective tissue disorders, that associated with polymyositis and dermatomyositis responds well to glucocorticoid therapy. Other immunosuppressive agents, such as azathioprine or methotrexate, are often used concurrently for treatment of the polymyositis/dermatomyositis.

Sjögren Syndrome

Sjögren syndrome comprises the triad of keratoconjunctivitis sicca, xerostomia, and recurrent swelling of the parotid glands and is often (in 60% of cases) associated with other collagen vascular diseases. Sjögren syndrome frequently includes pleuropulmonary problems, but in many cases these problems are secondary to the underlying connective tissue disease. Thoracic complications that may be directly related to Sjögren syndrome include interstitial infiltrates, often lymphocytic in nature (LIP); pleurisy with or without effusion; follicular bronchiolitis; and desiccation of the tracheobronchial tree, leading to impaired tracheobronchial clearance and consequent bronchiectasis and recurrent bronchitis or pneumonitis.[69,77]

Dyspnea is the primary symptom of the interstitial disease associated with Sjögren syndrome, whereas hoarseness and cough that is productive of thick, tenacious sputum are symptoms of laryngeal and tracheobronchial involvement. Chest radiography may show interstitial infiltrates with a prominent nodular component. These infiltrates have a lymphoplasmacytic histology that can be difficult to differentiate from that seen in patients with lymphoma. Results of pulmonary function tests demonstrate restrictive, obstructive, or mixed patterns. The lymphocytic interstitial pneumonitis and bronchiolitis associated with Sjögren syndrome respond to glucocorticoid treatment or immunosuppressive therapy.

Mixed Connective Tissue Disease

Several reports have described patients who have features of SLE, PSS, and polymyositis, often with associated pulmonary disease.[69,78] A hallmark of such mixed connective tissue disease is the presence of high titers of antibodies to extractable nuclear ribonucleoprotein (anti-nRNP). This syndrome is usually a relatively benign disease that responds well to glucocorticoids, although some cases display a less benign course that is associated with fatal diffuse interstitial lung disease or pulmonary hypertension.

eosinophilic granuloma of the lung

Eosinophilic granuloma of the lung, or pulmonary histiocytosis X, is an uncommon cause of diffuse infiltrative lung disease that primarily affects persons who are 20 to 50 years of age.[79] The disease is uncommon in African Americans and extraordinarily rare in Asians. The etiology of eosinophilic granuloma is unknown, but there is a strong association with cigarette smoking. In children and young adults, eosinophilic granuloma is a systemic disease with a similar histologic appearance, and either diabetes insipidus resulting from pituitary involvement or eosinophilic granuloma of bone may be seen in 15% to 20% of patients.

Clinical Manifestations

Most patients with pulmonary eosinophilic granuloma have cough or dyspnea, but as many as 25% of affected persons have only radiographic abnormalities and no symptoms. Between 10% and 20% of patients develop pneumothoraces, and a spontaneous pneumothorax may be the initial manifestation of the disease. The association of spontaneous pneumothorax with radiographic evidence of diffuse interstitial infiltrates suggests a diagnosis of eosinophilic granuloma or AIDS-related PCP.

The patterns seen on chest radiographs and HRCT in eosinophilic granuloma are variable.[6] Nodular and ground-glass opacities with thick-walled cysts are typically present early in the course of the disease. Later, thin-walled cysts, linear opacities, and emphysematous lesions are seen.

Eosinophilic granuloma often causes a reduction in $D_{L_{CO}}$ and vital capacity. However, unlike most other types of diffuse infiltrative lung disease, airflow obstruction is a characteristic feature, particularly in the later stages of the disease. The high incidence of airflow obstruction in eosinophilic granuloma may result from bronchiolar involvement in earlier stages of the disease and from bullae in late-stage disease. Smoking may also be a contributing factor to the high frequency of airflow obstruction in this disorder. Significant impairment of exercise performance is frequent and is greater than that which might be expected from the degree of volume restriction. Increasing wasted ventilation ratio (V_D / V_T) and falling P_aO_2 with exercise result in poor exercise performance in patients with pulmonary vascular abnormalities, which appear to be very prominent in eosinophilic granuloma.[80]

Histopathologic Features

The histopathologic features of eosinophilic granuloma are unique and include the presence of cells that resemble Langerhans cells in the skin. These cells stain positively with anti-CD1a monoclonal antibody and are shown by electron microscopy to contain characteristic inclusions termed X bodies, or Birbeck granules.[79] These cells can be seen in other types of diffuse infiltrative lung disease, but their numbers are much greater in eosinophilic granuloma. In patients with eosinophilic granuloma, $CD4^+$ T cells are in apposition to Langerhans cells, and there are increased numbers of neuroendocrine cells.[81] Granulocyte-macrophage colony-stimulating factor (GM-CSF) and TGF-β are found in the lesions, suggesting that these factors are potential mediators of the stimulation of Langerhans cells and fibroblasts, respectively. These findings suggest that the pathogenesis of eosinophilic granuloma involves interactions among these cells that result in fibroblast stimulation and fibrosis.

Diagnosis

The diagnosis of eosinophilic granuloma can be made most confidently by open lung biopsy. However, it has been suggested that the finding of more than 5% $CD1a^+$ cells in BAL fluid, along with an appropriate clinical and radiographic picture, may suffice for diagnosis.[79] HRCT may prove to be a useful noninvasive technique for diagnosing eosinophilic granuloma.

Treatment and Prognosis

Patients with eosinophilic granuloma have shown improvement after smoking cessation, and treatment with glucocorticoids has been reported to be effective, though no therapy has been sufficiently studied or has produced sufficiently convincing evidence of clinical benefit to justify a definitive recommendation.

Patients with eosinophilic granuloma have a median survival of 12.5 years after diagnosis. Respiratory failure causes a substantial proportion of the deaths. Increased age, airflow obstruction, reduced $D_{L_{CO}}$, and hyperinflation are predictors of a poor prognosis.[82]

alveolar proteinosis

Alveolar proteinosis is a rare disease characterized by the intra-alveolar accumulation of a cellular lipoproteinaceous material that resembles surfactant.[83] The disease can be acquired, congenital, or associated with other diseases.

Pathogenesis

Recent studies appear to have elucidated the pathogenesis of alveolar proteinosis.[83] About 90% of cases are acquired and are caused by autoantibodies to GM-CSF, a protein that is important in controlling the alveolar macrophage clearance of surfactant from the alveolus. In congenital cases, either mutations of surfactant proteins or defects in GM-CSF receptors are at fault. In secondary cases, a disorder that is clinically similar to alveolar proteinosis occurs in persons with lysinuric protein intolerance, acute silica exposure (usually from sandblasting) or other inhalational syndromes, several immunodeficiency syndromes, or any of several hematologic disorders. The potential role of GM-CSF in the secondary cases has not been clearly established.

The lipoproteinaceous material also may inhibit the macrophage's defense against infection, thereby producing the increased incidence of infection with intracellular pathogens that has been observed in patients with alveolar proteinosis.

Clinical Manifestations

Patients with alveolar proteinosis present in one of three ways: (1) without symptoms but with an abnormal chest radiograph; (2) with abrupt onset of cough, fever, and chest pain caused by the disease and complicated by opportunistic infection (with *Nocardia* species, fungi, or *Mycobacterium* species); and (3) with dyspnea and cough (sometimes productive) of gradual onset. Physical findings are often minimal. The chest radiograph or HRCT usually shows patchy or geographic ground-glass opacities with some thickening of the interlobular septa, resulting in a pattern that has been called crazy-paving. In later stages of the disease, coarse interstitial markings resulting from superimposed fibrosis can be seen. An elevated serum lactate dehydrogenase level without elevation of other serum enzyme levels may help make the diagnosis.

Histopathologic Features

The histopathologic changes in alveolar proteinosis include the presence of a granular, eosinophilic material within alveolar spaces that stains positively with the periodic acid–Schiff (PAS) reagent [*see Figure 14*]. Inflammation and fibrosis have been detected in some cases but are usually not prominent features. Electron microscopy reveals that the intra-alveolar substance contains numerous lamellar bodies, similar to those found in type II epithelial cells.

Diagnosis

Diagnosis of pulmonary alveolar proteinosis is made most conveniently by BAL. The effluent appears grossly turbid, and staining reveals large amounts of the PAS-positive material [*see Figure 13*]. PCP causes accumulation of a similar intra-alveolar

substance, and this infection must be excluded by special stains. The lavage liquid should be cultured for *Nocardia* species, fungi, and *Mycobacterium* species.

Treatment

Treatment of alveolar proteinosis is indicated if significant dyspnea is present at the time of diagnosis or if lesser symptoms fail to remit spontaneously after an observation period of several months. Approximately 10% of patients undergo spontaneous remission, so it is preferable to wait several months before instituting treatment. In some cases, however, the degree of hypoxemia and the severity of symptoms necessitate immediate therapy. The current best therapy for pulmonary alveolar proteinosis consists of whole lung lavage with sterile saline. The nonlavaged lung is protected by a double-lumen endotracheal tube. Most patients improve substantially after removal of the alveolar material by lavage, and the contralateral lung can be lavaged at a later date.

Now that the pathogenesis of alveolar proteinosis is known, specific therapy may be on the horizon.[83] In initial phase II trials of subcutaneous GM-CSF in patients with acquired disease, response rates of 36% to 75% have been seen. Further studies are necessary to determine the role of this therapy in relation to whole lung lavage.

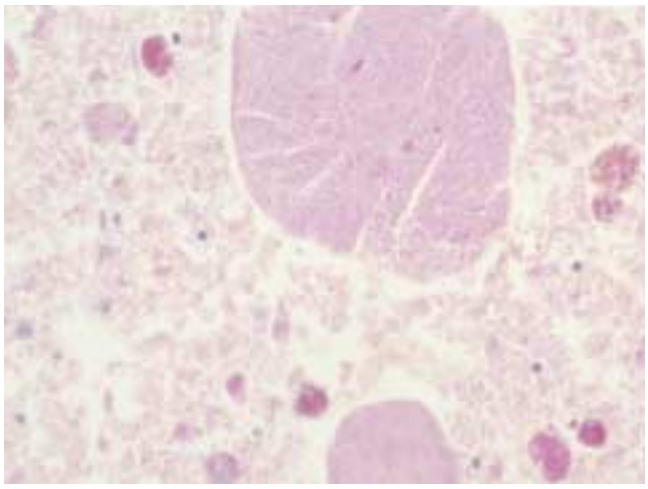


Figure 14 **Bronchoalveolar lavage effluent reveals distinctive periodic acid–Schiff–positive material that is diagnostic of alveolar proteinosis.**

eosinophilic pneumonia

There are seven forms of pulmonary infiltrates with peripheral eosinophilia (PIE): (1) simple pulmonary eosinophilia (Löffler syndrome); (2) prolonged pulmonary eosinophilia without asthma; (3) pulmonary eosinophilia with asthma [*see 14:II Asthma*]; (4) tropical pulmonary eosinophilia; (5) pulmonary vasculitis (allergic granulomatosis and angiitis) [*see 14:IV Focal and Multifocal Lung Disease*]; (6) hypereosinophilic syndrome [*see 5:VII Nonmalignant Disorders of Leukocytes*]; and (7) acute eosinophilic pneumonia.[84] In many of these disorders, substances secreted by eosinophils may help cause injury to the lung parenchyma, airways, or both. Miscellaneous disorders that may also cause pulmonary infiltrates with blood eosinophilia include infections (e.g., tuberculosis, brucellosis, and fungal diseases), neoplastic diseases (e.g., bronchogenic carcinoma, Hodgkin disease, and immunoblastic lymphadenopathy), and immune processes (e.g., rheumatoid lung disease and sarcoidosis).

Simple Pulmonary Eosinophilia (Löffler Syndrome)

Löffler syndrome is characterized by transient and sometimes migratory infiltrates and by mild symptoms lasting less than 1 month. The infiltrates are usually homogeneous in density; either a single infiltrate or multiple infiltrates may occur. Symptoms may be absent, or dyspnea and dry cough may develop. Pathologic examination reveals interstitial and intra-alveolar accumulation of eosinophils, macrophages, and fluid. Löffler syndrome may be idiopathic or may be caused by parasitic infestation (e.g., with *Ascaris* or *Strongyloides*) or a drug (e.g., nitrofurantoin or penicillin). Therapy consists of stopping the offending drug or treating the parasitic infestation; in most cases, no other therapy is required. Occasionally, patients with very symptomatic idiopathic disease require glucocorticoid therapy.

Prolonged Pulmonary Eosinophilia without Asthma

Prolonged pulmonary eosinophilia (chronic eosinophilic pneumonia) is an idiopathic disease that predominantly affects middle-aged women. It is characterized by productive cough, dyspnea, malaise, weight loss, night sweats, and fever associated with progressive peripheral pulmonary infiltrates. HRCT may help in differentiating this disorder from the other PIE syndromes.[85] Hemoptysis and wheezing may be present, which may lead to confusion with other causes of PIE. Pulmonary function test results may be normal or may show either restriction or obstruction.[86] Open lung biopsy specimens show massive, mixed inflammatory infiltrates that have a high eosinophil content. Some pulmonary vessels may contain a few inflammatory cells, but true vasculitis is not present.

Diagnosis is based on the clinical syndrome, chest radiography, and whether or not blood eosinophilia is present. If chest radiography shows typical dense peripheral infiltrates with central sparing, the diagnosis can be made without tissue examination. Usually, confirmatory information can be easily obtained by BAL and transbronchial lung biopsy.

Spontaneous remissions occur in as many as 10% of cases, but respiratory failure can also occur. Treatment with glucocorticoids is rapidly effective. Because relapses are frequent, therapy often must be continued for as long as 5 years.

Pulmonary Eosinophilia with Asthma

Allergic bronchopulmonary aspergillosis can cause pulmonary infiltrates in a patient with asthma [*see 14:II Asthma*].

Tropical Pulmonary Eosinophilia

The diagnosis of tropical pulmonary eosinophilia is suggested by the onset of asthma, fever, and marked blood eosinophilia and the presence of basilar mixed reticulonodular and alveolar infiltrates in a person who has recently traveled to the Far East.[87] This disease probably represents a form of filariasis (caused by *Wuchereria bancrofti* or other organisms). Despite therapy with diethylcarbamazine, many patients with this disorder have persistent inflammation and develop chronic interstitial disease.

Pulmonary Vasculitis (Allergic Granulomatosis and Angiitis)

Allergic granulomatosis and angiitis is a multisystem disorder characterized by vasculitis and necrotizing granulomatous

inflammation that involve the lungs, nervous system, and skin.[88] Although the kidneys, heart, spleen, and GI tract are much less commonly affected, any organ or tissue may be involved. The prominence of eosinophilia and elevated IgE levels suggest that immediate hypersensitivity mechanisms may play an important role in pathogenesis [*see 14:IV Focal and Multifocal Lung Disease*].

Hypereosinophilic Syndrome

Hypereosinophilic syndrome is characterized by a wide range of clinical manifestations that occur when mature eosinophils infiltrate organs [*see 5:VII Nonmalignant Disorders of Leukocytes*].

Acute Eosinophilic Pneumonia

Acute eosinophilic pneumonia is characterized by the acute onset of cough, dyspnea, fever, tachypnea, and rales; patients frequently require mechanical ventilation.[89] There is no clear association with cigarette smoking, environmental exposures, or drug intake, and patients tend to be young. Blood eosinophilia is not seen in most cases and is therefore not useful in diagnosing this disorder. Chest radiography and CT scanning show nonspecific combinations of alveolar (ground-glass) infiltrates and mixed alveolar-interstitial infiltrates that are sometimes associated with pleural effusions. The diagnosis can be made by bronchoalveolar lavage, which shows dramatic increases in the percentage of eosinophils, often to greater than 20% (the normal count is less than 1%). The administration of corticosteroids leads to rapid improvement, and relapses after steroid withdrawal are rare.

diffuse alveolar hemorrhage

Alveolar hemorrhage syndromes are characterized by diffuse parenchymal bleeding in the absence of blood aspiration, coagulopathy, elevation of pulmonary venous pressure, or an identifiable local cause (e.g., pulmonary emboli, cancer, or bronchitis). Frequently, diffuse alveolar hemorrhage is immunologically mediated and is often associated with glomerulonephritis.

The diseases that cause this syndrome are anti–glomerular basement membrane (anti-GBM) antibody disease (Goodpasture syndrome), idiopathic and rapidly progressive glomerulonephritis with or without immune complexes, vasculitis (microscopic polyangiitis, Wegener granulomatosis, and other forms of antineutrophil cytoplasmic antibody [ANCA]–associated capillaritis), SLE and other collagen vascular disorders, antiphospholipid antibody syndrome, certain drugs, and idiopathic pulmonary hemosiderosis.[90] Although each of these disorders has suggestive clinical characteristics, in many cases the diagnosis cannot be made on clinical grounds. In addition, many patients with diffuse alveolar hemorrhage are acutely ill, making a rapid and specific evaluation necessary. A suggested diagnostic approach to such a patient consists of the following steps:

1. Obtain serum to test for anti-GBM antibody, antinuclear and anti-DNA antibody, ANCA, complement levels, and immune complexes.
2. Perform bronchoalveolar lavage with an iron stain of the obtained specimen. If hemosiderin-laden macrophages are present in large numbers, the diagnosis of alveolar hemorrhage is confirmed.
3. If step 1 fails to confirm a diagnosis or the results are not rapidly available, a biopsy of the kidney, lung, or another involved site should be performed. Immunofluorescence and electron microscopy with routine tissue examination should be included.

Almost all cases of alveolar hemorrhage can be rapidly diagnosed if the physician follows this approach. Treatment with high doses of intravenous glucocorticoids should be started while the evaluation is in progress, because alveolar hemorrhage can rapidly become life threatening.

Anti-GBM Antibody Disease

Anti-GBM antibody disease occurs in young men who smoke cigarettes. It is characterized by glomerulonephritis and diffuse alveolar hemorrhage[91] and is caused by cytotoxic antibody against the α3 chain of type IV collagen in glomerular and alveolar basement membranes.

Many patients present with hemoptysis, which can be massive; however, some episodes of hemorrhage occur without hemoptysis. Most patients complain of dyspnea, and gross hematuria is common. Iron deficiency anemia is present in nearly all patients, and renal failure is present initially in 50% of cases.

Chest radiography initially shows fluffy perihilar acinar shadows; later, interstitial changes are seen. Pulmonary function tests demonstrate a restrictive pattern that improves with clinical resolution. $D_{L_{CO}}$ may be increased because of carbon monoxide uptake by extravascular hemoglobin. As the disease becomes chronic, interstitial radiographic changes and restrictive pulmonary function abnormalities may persist.

The diagnosis is confirmed when circulating anti-GBM antibody is detected (95% of patients have anti-GBM antibody) or linear deposits of IgG (occasionally IgA) are demonstrated on the glomerular or alveolar basement membrane. Anti-GBM antibody can also be eluted from the involved tissue.

Current recommendations for treatment include immunosuppressive agents, glucocorticoids, and plasmapheresis. Despite this aggressive therapy, 50% of patients die or require long-term dialysis after 2 years. Death is often caused by massive alveolar hemorrhage that is frequently precipitated by infection.

Idiopathic Pulmonary Hemosiderosis

Idiopathic pulmonary hemosiderosis, a disorder characterized by recurrent alveolar hemorrhage, occurs primarily in children (usually younger than 10 years) and young adults. When it occurs in adults, men are more often affected than women. Presentation of the disease is similar to that of Goodpasture syndrome except for the absence of renal involvement. Recurrent episodes of clinical and subclinical hemorrhage occur over time, resulting in lung damage that may be obstructive or restrictive in nature. Lung tissue shows changes of acute and chronic hemorrhage, but no vasculitis, immune complexes, or linear-staining deposits are found.

Patients with idiopathic pulmonary hemosiderosis are treated with glucocorticoids. Anecdotal evidence supports the use of immunosuppressive agents and plasmapheresis. Mean survival is 3 to 5 years, but spontaneous remissions and long-term survival do occur.

lymphangioleiomyomatosis

The onset of lymphangioleiomyomatosis (LAM), a rare disease (approximately 100 new cases per year worldwide) that affects women of childbearing age almost exclusively, is often heralded by some combination of dyspnea, cough, hemoptysis, pneumothorax, and chylothorax.[92] At the onset of symptoms, there may be no diffuse radiographic abnormality, though a fine reticular or reticulonodular process, predominating at the lung bases, may be seen. Because of lymphatic obstruction, sep-

tal (Kerley B) lines can be seen. Occasionally, there is a diffuse interstitial pattern in association with hyperinflation. In contrast to most diffuse infiltrative lung diseases, this disorder is often characterized by obstructive physiology with increased lung volumes, which has been attributed to peribronchial smooth muscle hyperplasia. In addition, $D_{L_{CO}}$ is usually decreased.

The diagnosis may be strongly suspected in a young nonsmoking woman with recurrent pneumothorax, chylothorax, fixed airway obstruction, and a diffuse interstitial radiographic pattern.[92] Nuclear ventilation lung scans show a so-called speckled pattern caused by accumulation of radionuclide in the lung cysts. HRCT is diagnostic when it shows homogeneously distributed thin-walled parenchymal cysts, varying from a few millimeters to several centimeters in diameter; these cysts may be associated with retrocrural adenopathy, pleural effusion, thoracic duct dilatation, pericardial effusion, and pneumothorax. Approximately 50% of patients also have renal angiomyolipomas as an associated finding, often first identified on CT scanning. Recently, a high incidence of meningiomas has been noted in patients with LAM.[93] All of these imaging findings are identical to those of tuberous sclerosis (TS). Of patients with TS, about one third of females and none of the males have LAM.[94] The overlap between the two diseases has led to the discovery that patients with LAM without TS may have somatic mutations of the *TSC2* gene, which has been implicated in the pathogenesis of TS.[95]

Open lung biopsy in LAM reveals proliferation of smooth muscle cells along lymphatic, vascular, and bronchial structures; in the alveolar walls; and in the subpleural areas of the visceral pleura. Cyst formation may occur secondary to bronchiolar obstruction. The clinical differential diagnosis includes many of the other infiltrative lung diseases, such as IPF, eosinophilic granuloma (which can produce similar cystic changes on HRCT), and sarcoidosis. Pathologically, the differential diagnosis includes diffuse pulmonary lymphangiomatosis, lymphangiomas, pulmonary lymphangiectasis, metastatic uterine leiomyosarcoma, and benign metastasizing uterine leiomyomata.

More than 75% of patients with LAM survive longer than 8 years.[92] There is some evidence that disease progression may be hormone dependent, because some slowing of progression has been reported with oophorectomy, progesterone treatment, or a combination of the two.[92] Lung transplantation may be considered for end-stage disease. Although transplantation may be lifesaving, there have been cases of recurrence in the transplanted lung, emphasizing the systemic nature of the disorder.[92]

pulmonary alveolar microlithiasis

Pulmonary alveolar microlithiasis is an exceedingly rare, sometimes familial disorder that is most often diagnosed in an asymptomatic patient by a routine chest radiograph. The chest radiograph reveals a strikingly distinct and diffuse scattering of small calcified nodules that, on histopathologic examination, are seen to be calcified spheres (microliths or calcispherites) that occupy the alveolar space.[96] This condition may progress to respiratory failure. A possible response to treatment with disodium etidronate, a bisphosphonate, has been reported in a single patient.[96] Lung transplantation is an option in severe cases.

Gerald W. Staton, Jr., M.D., participates in the speakers' bureaus of GlaxoSmithKline and Intermune Pharmaceuticals, Inc.; Roland H. Ingram, Jr., M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Demedts M, Wells AU, Anto JM, et al: Interstitial lung diseases: an epidemiological overview. Eur Respir J Suppl 32:2s, 2001

2. Respiratory diseases disproportionately affecting minorities. The NHLBI Working Group. Chest 108:1380, 1995

3. du Bois RM: The genetic predisposition to interstitial lung disease: functional relevance. Chest 121:14S, 2002

4. Strieter RM: Mechanisms of pulmonary fibrosis: conference summary. Chest 120:77S, 2001

5. Reynolds HY: Diagnostic and management strategies for diffuse interstitial lung disease. Chest 113:192, 1998

6. Glaspole IN, du Bois RM, Wells AU: The application of high-resolution CT to diagnosis in diffuse parenchymal lung disease. Monaldi Arch Chest Dis 56:233, 2001

7. Alhamad EH, Lynch JP III, Martinez FJ: Pulmonary function tests in interstitial lung disease: what role do they have? Clin Chest Med 22:715, 2001

8. Baughman RP, Drent M: Role of bronchoalveolar lavage in interstitial lung disease. Clin Chest Med 22:331, 2001

9. Glaspole IN, Wells AU, du Bois RM: Lung biopsy in diffuse parenchymal lung disease. Monaldi Arch Chest Dis 56:225, 2001

10. Miller JD, Urschel JD, Cox G, et al: A randomized, controlled trial comparing thoracoscopy and limited thoracotomy for lung biopsy in interstitial lung disease. Ann Thorac Surg 70:1647, 2000

11. Ing AJ: Interstitial lung disease and gastroesophageal reflux. Am J Med 111:41S, 2001

12. Walsh LJ, Wong CA, Oborne J, et al: Adverse effects of oral corticosteroids in relation to dose in patients with lung disease. Thorax 56:279, 2001

13. Harris-Eze AO, Sridhar G, Clemens RE, et al: Oxygen improves maximal exercise performance in interstitial lung disease. Am J Respir Crit Care Med 150:1616, 1994

14. Sulica R, Teirstein A, Padilla ML: Lung transplantation in interstitial lung disease. Curr Opin Pulmon Med 7:314, 2001

15. Camus PH, Foucher P, Bonniaud PH, et al: Drug-induced infiltrative lung disease. Eur Respir J Suppl 32:93s, 2001

16. Shanholtz C: Acute life-threatening toxicity of cancer treatment. Crit Care Clin 17:483, 2001

17. Imokawa S, Colby TV, Leslie KO, et al: Methotrexate pneumonitis: review of the literature and histopathological findings in nine patients. Eur Respir J 15:373, 2000

18. Nemery B, Bast A, Behr J, et al: Interstitial lung disease induced by exogenous agents: factors governing susceptibility. Eur Respir J Suppl 32:30s, 2001

19. Katabami M, Dosaka-Akita H, Honma K, et al: Pneumoconiosis-related lung cancers: preferential occurrence from diffuse interstitial fibrosis–type pneumoconiosis. Am J Respir Crit Care Med 162:295, 2000

20. Kato T, Usami I, Morita H, et al: Chronic necrotizing pulmonary aspergillosis in pneumoconiosis: clinical and radiologic findings in 10 patients. Chest 121:118, 2002

21. Fujimura N: Pathology and pathophysiology of pneumoconiosis. Curr Opin Pulmon Med 6:140, 2000

22. Wang XR, Christiani DC: Respiratory symptoms and functional status in workers exposed to silica, asbestos, and coal mine dusts. J Occup Environ Med 42:1076, 2000

23. Sharma SK, Pande JN, Verma K: Effect of prednisolone treatment in chronic silicosis. Am Rev Respir Dis 143:814, 1991

24. Saltini C, Amicosante M: Beryllium disease. Am J Med Sci 321:89, 2001

25. Patel AM, Ryu JH, Reed CE: Hypersensitivity pneumonitis: current concepts and future questions. J Allergy Clin Immunol 108:661, 2001

26. Drent M, van Velzen-Blad H, Diamant M, et al: Bronchoalveolar lavage in extrinsic allergic alveolitis: effect of time elapsed since antigen exposure. Eur Respir J 6:1276, 1993

27. Akira M, Atagi S, Kawahara M, et al: High-resolution CT findings of diffuse bronchioloalveolar carcinoma in 38 patients. AJR Am J Roentgenol 173:1623, 1999

28. Bragg DG, Chor PJ, Murray KA, et al: Lymphoproliferative disorders of the lung: histopathology, clinical manifestations, and imaging features. AJR Am J Roentgenol 163:273, 1994

29. Statement on sarcoidosis. Joint statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders (WASOG). Am J Respir Crit Care Med 160:736, 1999

30. Gal A, Koss MN: The pathology of sarcoidosis. Curr Opin Pulmon Med 8:445, 2002

31. Rybicki BA, Iannuzzi MC, Frederick MM, et al: Familial aggregation of sarcoidosis: a case-control etiologic study of sarcoidosis (ACCESS). Am J Respir Crit Care Med 164:2085, 2001

32. Costabel U: Sarcoidosis: clinical update. Eur Respir J Suppl 32:56s, 2001

33. Baughman RP, Teirstein AS, Judson MA, et al: Clinical characteristics of patients in a case control study of sarcoidosis. Am J Respir Crit Care Med 164:1885, 2001

34. Mañá J, Gómez-Vaquero C, Montero A, et al: Löfgren's syndrome revisited: a study of 186 patients. Am J Med 107:1999, 1999

35. Soskel NT, Sharma OP: Pleural involvement in sarcoidosis. Curr Opin Pulmon Med 6:455, 2000

36. Takahashi T, Azuma A, Abe S, et al: Significance of lymphocytosis in bronchoalveolar lavage in suspected ocular sarcoidosis. Eur Respir J 18:515, 2001

37. Dumas JL, Valeyre D, Chapelon-Abric C, et al: Central nervous system sarcoidosis: follow-up at MR imaging during steroid therapy. Radiology 214:411, 2000

38. Barnard J, Newman LS: Sarcoidosis: immunology, rheumatic involvement, and therapeutics. Curr Opin Rheumatol 13:84, 2001

39. Sharma OP: Hypercalcemia in granulomatous disorders: a clinical review. Curr Opin Pulmon Med 6:442, 2000

40. Baughman RP, Lower EE: A clinical approach to the use of methotrexate for sarcoidosis. Thorax 54:742, 1999

41. Walker S, Mikhail G, Banner N, et al: Medium term results of lung transplantation for end stage pulmonary sarcoidosis. Thorax 53:281, 1998

42. Gross TJ, Hunninghake GW: Idiopathic pulmonary fibrosis. N Engl J Med 345:517, 2001

43. American Thoracic Society / European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. Am J Respir Crit Care Med 165:277, 2002

44. Gauldie J: Pro: Inflammatory mechanisms are a minor component of the pathogenesis of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 165:1205, 2002

45. Strieter RM: Con: inflammatory mechanisms are not a minor component of the pathogenesis of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 165:1206, 2002

46. Hodgson U, Laitinen T, Tukiainen P: Nationwide prevalence of sporadic and familial idiopathic pulmonary fibrosis: evidence of founder effect among multiplex families in Finland. Thorax 57:338, 2002

47. Thomas AQ, Lane K, Phillips J III, et al: Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. Am J Respir Crit Care Med 165:1322, 2002

48. Brantly M, Avila NA, Shotelersuk V, et al: Pulmonary function and high-resolution CT findings in patients with an inherited form of pulmonary fibrosis, Hermansky-Pudlak syndrome, due to mutations in HPS-1. Chest 117:129, 2000

49. Selman M, King TE, Pardo A: Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med 134:136, 2001

50. Cherniack RM, Colby TV, Flint A, et al: Correlation of structure and function in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 151:1180, 1995

51. Flaherty KR, Travis WD, Colby TV, et al: Histopathologic variability in usual and nonspecific interstitial pneumonias. Am J Respir Crit Care Med 164:1722, 2001

52. Hunninghake GW, Zimmerman MB, Schwartz DA, et al: Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 164:193, 2001

53. King TE Jr, Schwarz MI, Brown K, et al: Idiopathic pulmonary fibrosis: relationship between histopathologic features and mortality. Am J Respir Crit Care Med 164:1025, 2001

54. Ma Y, Seneviratne CK, Koss M: Idiopathic pulmonary fibrosis and malignancy. Curr Opin Pulmon Med 7:278, 2001

55. Lynch JP III, White E, Flaherty K: Corticosteroids in idiopathic pulmonary fibrosis. Curr Opin Pulmon Med 7:298, 2001

56. Zisman DA, Lynch JP III, Toews GB, et al: Cyclophosphamide in the treatment of idiopathic pulmonary fibrosis: a prospective study in patients who failed to respond to corticosteroids. Chest 117:1619, 2000

57. Raghu G, DePaso WJ, Cain K, et al: Azathioprine combined with prednisone in the treatment of idiopathic pulmonary fibrosis: a prospective double-blind randomized, placebo controlled clinical trial. Am Rev Respir Dis 144:291, 1991

58. Douglas WW, Ryu JH, Schroeder DR: Idiopathic pulmonary fibrosis: impact of oxygen and colchicine, prednisone, or no therapy on survival. Am J Respir Crit Care Med 161:1172, 2000

59. Meyers BF, Lynch JP, Trulock EP, et al: Single versus bilateral lung transplantation for idiopathic pulmonary fibrosis: a ten-year institutional experience. J Thorac Cardiovasc Surg 120:99, 2000

60. Mogulkoc N, Brutsche MH, Bishop PW, et al: Pulmonary function in idiopathic pulmonary fibrosis and referral for lung transplantation. Am J Respir Crit Care Med 164:103, 2001

61. Lasky JA, Ortiz LA: Antifibrotic therapy for the treatment of pulmonary fibrosis. Am J Med Sci 322:213, 2001

62. Ziesche R, Hofbauer E, Wittmann K, et al: A preliminary study of long-term treatment with interferon gamma-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. N Engl J Med 341:1264, 1999

63. Ryu JH, Colby TV, Hartman TE, et al: Smoking-related interstitial lung diseases: a concise review. Eur Respir J 17:122, 2001

64. Vourlekis JS, Brown KK, Cool CD, et al: Acute interstitial pneumonitis: case series and review of the literature. Medicine (Baltimore) 79:369, 2000

65. Travis WD, Galvin JR: Non-neoplastic pulmonary lymphoid lesions. Thorax 56:964, 2001

66. Nagai S, Kitaichi M, Itoh H, et al: Idiopathic nonspecific interstitial pneumonia / fibrosis: comparison with idiopathic pulmonary fibrosis and BOOP. Eur Respir J 12:1010, 1998

67. Epler GR: Bronchiolitis obliterans organizing pneumonia. Arch Intern Med 161:158, 2001

68. Lohr RH, Boland BJ, Douglas WW, et al: Organizing pneumonia: features and prognosis of cryptogenic, secondary, and focal variants. Arch Intern Med 157:1323, 1997

69. Lamblin C, Bergoin C, Saelens T, et al: Interstitial lung diseases in collagen vascular diseases. Eur Respir J Suppl 32:69s, 2001

70. Agustí C, Xaubet A, Roca J, et al: Interstitial pulmonary fibrosis with and without associated collagen vascular disease: results of a two year follow up. Thorax 47:1035, 1992

71. Homma Y, Ohtsuka Y, Tanimura K, et al: Can interstitial pneumonia as the sole presentation of collagen vascular diseases be differentiated from idiopathic interstitial pneumonia? Respiration 62:248, 1995

72. Keane MP, Lynch JP III: Pleuropulmonary manifestations of systemic lupus erythematosus. Thorax 55:159, 2000

73. Gochuico BR: Potential pathogenesis and clinical aspects of pulmonary fibrosis associated with rheumatoid arthritis. Am J Med Sci 321:83, 2001

74. Gabbay E, Tarala R, Will R, et al: Interstitial lung disease in recent onset rheumatoid arthritis. Am J Respir Crit Care Med 156:528, 1997

75. Cheema GS, Quismorio FP Jr: Interstitial lung disease in systemic sclerosis. Curr Opin Pulmon Med 7:283, 2001

76. Douglas WW, Tazelaar HD, Hartman TE, et al: Polymyositis-dermatomyositis-associated interstitial lung disease. Am J Respir Crit Care Med 164:1182, 2001

77. Cain HC, Noble PW, Matthay RA: Pulmonary manifestations of Sjögren's syndrome. Clin Chest Med 19:687, 1998

78. Prakash UB: Respiratory complications in mixed connective tissue disease. Clin Chest Med 19:733, 1998

79. Vassallo R, Ryu JH, Colby TV, et al: Pulmonary Langerhans'-cell histiocytosis. N Engl J Med 342:1969, 2000

80. Fartoukh M, Humbert M, Capron F, et al: Severe pulmonary hypertension in histiocytosis X. Am J Respir Crit Care Med 161:216, 2000

81. Aguayo SM, King TE Jr, Waldron JA Jr, et al: Increased pulmonary neuroendocrine cells with bombesin-like immunoreactivity in adult patients with eosinophilic granuloma. J Clin Invest 86:838, 1990

82. Vassallo R, Ryu JH, Schroeder DR, et al: Clinical outcomes of pulmonary Langerhans'-cell histiocytosis in adults. N Engl J Med 346:484, 2002

83. Seymour JF, Presneill JJ: Pulmonary alveolar proteinosis: progress in the first 44 years. Am J Respir Crit Care Med 166:215, 2002

84. Allen JN, Davis WB: Eosinophilic lung diseases. Am J Respir Crit Care Med 150:1423, 1994

85. Johkoh T, Muller NL, Akira M, et al: Eosinophilic lung diseases: diagnostic accuracy of thin-section CT in 111 patients. Radiology 216:773, 2000

86. Marchand E, Reynaud-Gaubert M, Lauque D, et al: Idiopathic chronic eosinophilic pneumonia: a clinical and follow-up study of 62 cases. Groupe d'Etudes et de Recherche sur les Maladies "Orphelines" Pulmonaires (GERM"O"P). Medicine (Baltimore) 77:299, 1998

87. Ong RK, Doyle RL: Tropical pulmonary eosinophilia. Chest 113:1673, 1998

88. Conron M, Beynon HL: Churg-Strauss syndrome. Thorax 55:870, 2000

89. Pope-Harman A, Davis W, Allen E, et al: Acute eosinophilic pneumonia: a summary of 15 cases and review of the literature. Medicine (Baltimore) 75:334, 1996

90. Specks U: Diffuse alveolar hemorrhage syndromes. Curr Opin Rheumatol 13:12, 2001

91. Ball JA, Young KR Jr: Pulmonary manifestations of Goodpasture's syndrome: antiglomerular basement membrane disease and related disorders. Clin Chest Med 19:777, 1998

92. Kelly J, Moss J: Lymphangioleiomyomatosis. Am J Med Sci 321:17, 2001

93. Moss J, DeCastro R, Patronas NJ, et al: Meningiomas in lymphangioleiomyomatosis. JAMA 286:1879, 2001

94. Moss J, Avila NA, Barnes PM, et al: Prevalence and clinical characteristics of lymphangioleiomyomatosis (LAM) in patients with tuberous sclerosis complex. Am J Respir Crit Care Med 164:669, 2001

95. Pacheco-Rodriguez G, Kristof AS, Stevens LA, et al: Giles F. Filley Lecture: Genetics and gene expression in lymphangioleiomyomatosis. Chest 121:56S, 2002

96. Chan ED, Morales DV, Welsh CH, et al: Calcium deposition with or without bone formation in the lung. Am J Respir Crit Care Med 165:1654, 2002

Acknowledgments

Figure 1 Courtesy of Dr. Sam Aguayo, Veterans Affairs Medical Center, Decatur, Georgia.

Figure 4 Andy Christie.

Figure 7 Seward Hung.

Figures 7 through 12 Courtesy of Dr. Anthony Gal, Emory University Hospital, Decatur, Georgia.

VI VENTILATORY CONTROL DURING WAKEFULNESS AND SLEEP

KINGMAN P. STROHL, M.D.

Overview

Ventilation is a critical function for eliminating carbon dioxide and acquiring oxygen. At or near sea level, ventilation maintains arterial carbon dioxide tension (P_aCO_2) values in the range of 38 to 42 mm Hg and arterial oxygen tension (P_aO_2) values in the range of 85 to 100 mm Hg. What is remarkable is that P_aCO_2 values vary relatively little over the human life span despite substantial alterations in the mechanical properties of the chest wall and lungs that accompany birth, maturation, and aging. The control system for ventilation not only optimizes gas exchange but also serves a role in acid-base balance, speech, deglutition, defecation, and posture.

The components of ventilatory behavior (i.e., breathing rate and depth) are the result of a feedback control system in which the brain (controller) organizes neuromuscular output to the respiratory muscles of the upper airway, chest wall, and lungs (controlled system). The controlled system alters arterial pH, CO_2, and O_2 in response to impulses from the brain. Specialized sensors located in the respiratory system monitor the rate of gas exchange and send impulses to sensors located in the brain to prompt adjustments in system regulation. A feedback control model of the respiratory system provides insights into the effects of sleep on ventilation and gas exchange [*see Figure 1*].

Genetic factors influence disorders of ventilatory control (e.g., sleep apnea); however, there are adaptive components. One example is the periodic breathing during sleep that occurs with adaptation to high altitude. Such functional flexibility, or plasticity, is an essential feature of development, maintenance, and expression of effective ventilation and is not merely the result of mechanical operation of the lungs and chest wall. Consequently, genes, maturation, and experience all influence the adult phenotype for breathing and sleep and the clinical disorders resulting from this physiology.

This chapter focuses on the respiratory control system and how its elements contribute to sleep apnea and other state-related disorders of ventilation.

Physiology of Ventilatory Control

Inhalation begins with the discharge of inspiratory neural impulses from respiratory centers located in the medulla.[1] This neural network is embedded in a system of adjacent medullary neurons, pontine neurons, and regions such as the nucleus tractus solitarius (NTS) that receive neural impulses resulting from lung inflation, blood pressure, and other afferent systems. Inhalation continues until the respiratory centers receive negative feedback from the adjacent medullary neurons and from peripheral receptors, some of which are activated by lung inflation. The intensity of the activity of medullary neurons is affected by input from chemoreceptors [*see Figure 1*]. Influences from higher centers also adjust inspiratory and expiratory activity for speech and swallowing. With inhibition, inspiration ceases and expiration continues until inhibitory influences wane sufficiently to allow initiation of the next inspiration. Simply put, the rate of inspiration is determined by the intensity of medullary discharge; duration and depth of inspiraton are determined by the timing and intensity of inhibitory influences.

Chemoreceptors are specialized cells that sense O_2 and CO_2 through changes in pH. The peripheral chemoreceptors (i.e., the carotid and aortic bodies) are highly vascular collections of specialized sensory cells.[2] The carotid bodies are located bilaterally at the bifurcations of the common carotid arteries; the aortic bodies are situated anterior and posterior to the arch of the aorta and the left main pulmonary artery. The peripheral chemoreceptors are stimulated primarily by a low P_aO_2, although hypercapnia, acidemia, and possibly hyperthermia may influence an increased response to hypoxemia. Impulses travel from the carotid and aortic bodies to the NTS in the brain stem via sensory ganglia and the afferent nerves that follow along the ninth and 10th cranial nerves, respectively. Increases in P_aCO_2 stimulate cells on the ventral medullary surface (VMS), primarily by lowering the pH of the medullary extracellular fluid.[3] In the steady state, the pH of cerebrospinal fluid reflects the pH of the medullary microenvironment and may differ significantly from blood pH. This discordance is thought to result in transient stimulation of ventilation, even in the presence of a respiratory alkalosis (e.g., in persons who return to sea level after spending several weeks at high altitude).

Specialized sensory cells (i.e., mechanoreceptors) located in the upper airway, chest wall, and lung detect mechanical deformation and temperature changes resulting from inhalation and exhalation.[4] Afferent nerve signals from mechanoreceptors are directed to the medulla (NTS), where information is integrated with chemoreceptor information to influence the medullary timing and volume of ventilation; integrated information is relayed through thalamic connections to the cortex. In the presence of parenchymal lung disease, the information received by the cortex may contribute to perceptions of breathlessness (dyspnea). Thus, lung inflammation and bronchoconstriction will activate unmyelinated pulmonary C fibers, thereby resulting in hyperventilation, tachypnea, and dyspnea. Myelinated fibers from stretch receptors carry impulses that influence the duration of inspiration and expiration. Impulses from stretch receptors and C fibers travel through the pulmonary branch of the 10th cranial nerve. Segmental intercostal nerves carry impulses to the brain stem from the chest wall, muscle spindles, and joint proprioceptors.[5] Mechanoreceptors in these areas are influenced by the position of the rib cage and by the muscular tension required to inflate the lungs, and the rate of change in the afferent neurogram resembles flow rate. This receptor system acts in concert with chemoreceptors to control the timing of exhalation and control ventilation.

Other brain centers (e.g., the hypothalamus and cortex) provide input to pontomedullary respiratory centers.[6] These pathways coordinate neuromuscular outputs with voluntary respiratory acts, such as talking or expulsive maneuvers, and coordinate ventilation with metabolism, posture, and swallowing. Hypothalamic influences are in part responsible for the so-called wakefulness stimulus—the increased activity of medullary neurons in the cortex that occurs during wakefulness, as opposed to the activity that occurs during sleep. Some cortical cerebral path-

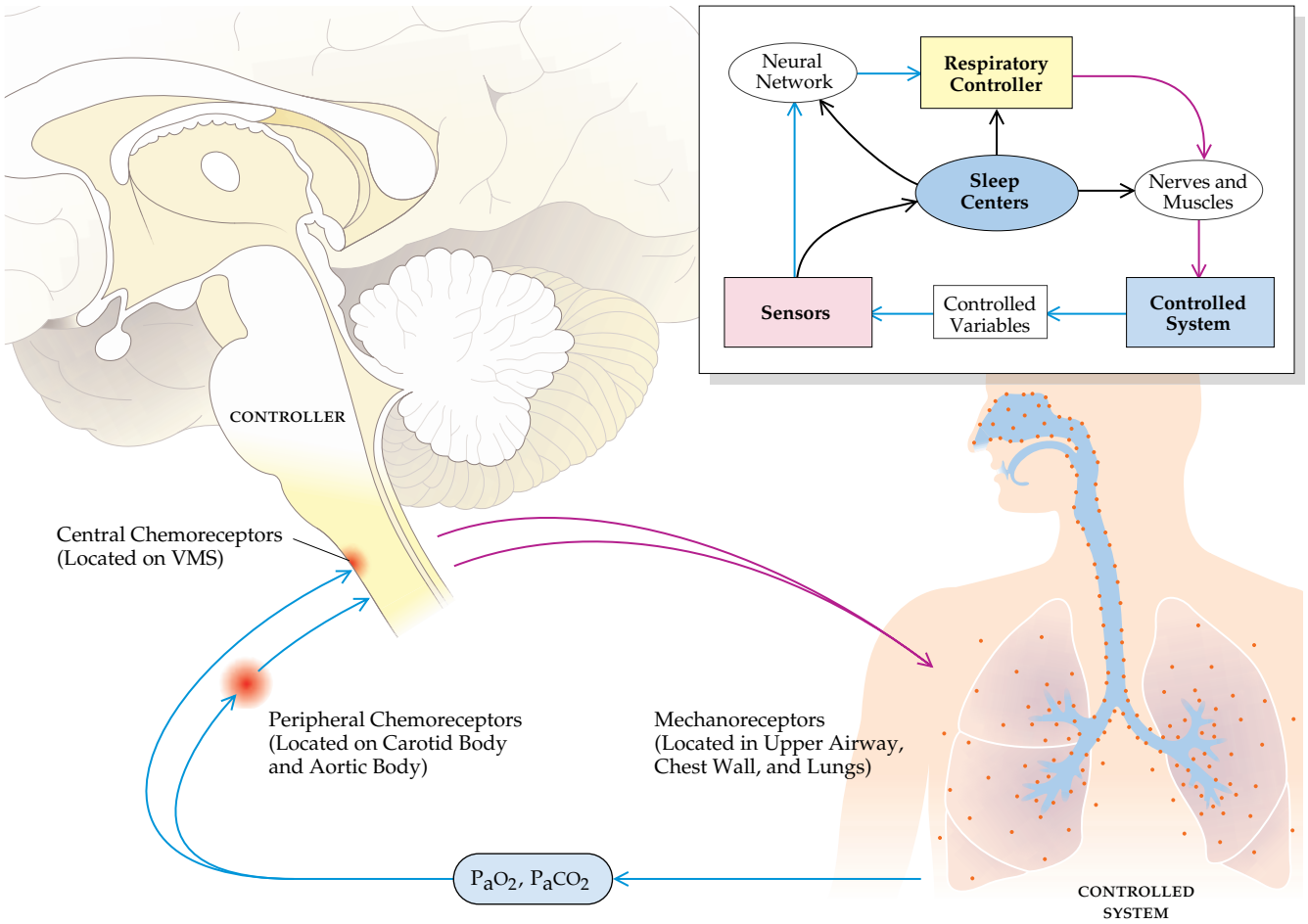


Figure 1 **This schematic representation illustrates the key elements of the respiratory control system—namely, the respiratory controller (the brain regions generating neuromuscular drive), the controlled system (upper airway, lungs, and chest wall), and the specialized sensors that send signals between the two. The controlled system alters arterial pH, CO_2, and O_2 according to signals received from specialized sensors. Peripheral chemoreceptors located on the carotid and aortic bodies are stimulated primarily by decreases in oxygen (P_aO_2) and to a lesser extent by carbon dioxide tension (P_aCO_2) (blue arrows); neural signals received by these chemoreceptors are sent to the brain stem. Central chemoreceptors—for example, those located in the medulla near the ventral medullary surface (VMS)—are stimulated by increases in PCO_2. Information received from the chemoreceptors is integrated in the medulla, and neural impulses from the medullary system (red arrows) travel to the muscles of the upper airway and chest wall to influence timing and volume of ventilation. Peripheral mechanoreceptors in the upper airway, chest wall, and lung detect mechanical deformation and temperature changes resulting from inhalation and exhalation; neural signals from these mechanoreceptors are sent to the central and peripheral chemoreceptors. This scheme of respiratory feedback control is a basic concept for understanding how sleep affects ventilation and gas exchange.**

ways circumvent the medulla and pass directly to respiratory muscles via pyramidal tracts. The cerebellum plays a role in both coordinating and adjusting respiratory neural output to the upper airway and chest wall muscles.

Putative set points for the ventilatory control system help ensure homeostatic control of acid-base balance (pH) and O_2 delivery. One example of a set point is the apneic threshold, which is defined as that level of arterial (or central) CO_2 below which there is little or no inspiratory activation. This set point is higher in sleep.[7] Certainly, brain centers other than the medulla and pons contribute to breathing rate and depth and, to a certain extent, can override brain stem mechanisms for breathing. During sleep, brain centers have less influence and may actively inhibit respiratory activity, unless they are engaged in an arousal response from quiet sleep or triggered by changes associated with rapid eye movement (REM) sleep [*see Figure 2*].

Integration and coordination of neuromuscular activity are most apparent during inhalation. In health, exhalation occurs as a result of the passive recoil of the lungs and chest wall; however, the duration of expiration and the start of a new inspiration are actively controlled events. In a healthy person, breathing is a sequence of inhalations and exhalations that serve to maintain alveolar ventilation at a level that is appropriate for meeting metabolic demands during wakefulness, exercise, and sleep. The increased metabolic requirements of exercise are met by increases in respiratory frequency and tidal volume (and therefore increases in minute and alveolar ventilations). During exercise, activation of abdominal and intercostal muscles during expiration allows more rapid emptying of the lungs at higher tidal volumes.[4,8] Environmental stresses, metabolic disturbances, hormonal changes, drugs, sleep-wake activity, and exercise may influence the output of a normal control system. Excessive respira-

tory suppression results in alveolar hypoventilation (hypercapnia), whereas overstimulation results in alveolar hyperventilation (hypocapnia).

Disorders with Increased Ventilatory Drive

Interstitial lung diseases (e.g., pulmonary fibrosis) increase resting ventilation and lower P_aCO_2 as a result of increased activity of lung receptors (probably C fibers).[9] The hyperventilation that accompanies pulmonary edema, pneumonia, interstitial disease, and the acute respiratory distress syndrome is a rapid, shallow breathing pattern that results from activation of these lung receptors. Hypocapnia with dyspnea may occur in the absence of hypoxemia in this setting. Unilateral vagal interruption in patients with parenchymal lung disease has been shown to reduce ventilation, as well as dyspnea, and may contribute to the improvement in breathlessness after unilateral lung transplantation.[9]

Hyperventilation is regularly produced by exposure to high altitude or other hypoxic environments, metabolic acidosis, pregnancy and other conditions associated with elevated progestational hormones, anxiety states, and mildly toxic doses of salicylates, amphetamines, or other CNS-stimulating drugs. Unlike the hyperventilation associated with parenchymal lung disease, the hyperventilation that occurs during progesterone stimulation (e.g., that which occurs during pregnancy) or metabolic acidosis is associated with an increased tidal volume and little increase in respiratory rate. The hyperventilation characterized by high tidal volume and relatively low frequency that accompanies diabetic ketoacidosis (Kussmaul respiration) is pH mediated and may not be as apparent to an observer as the breathlessness that occurs in interstitial lung disease.

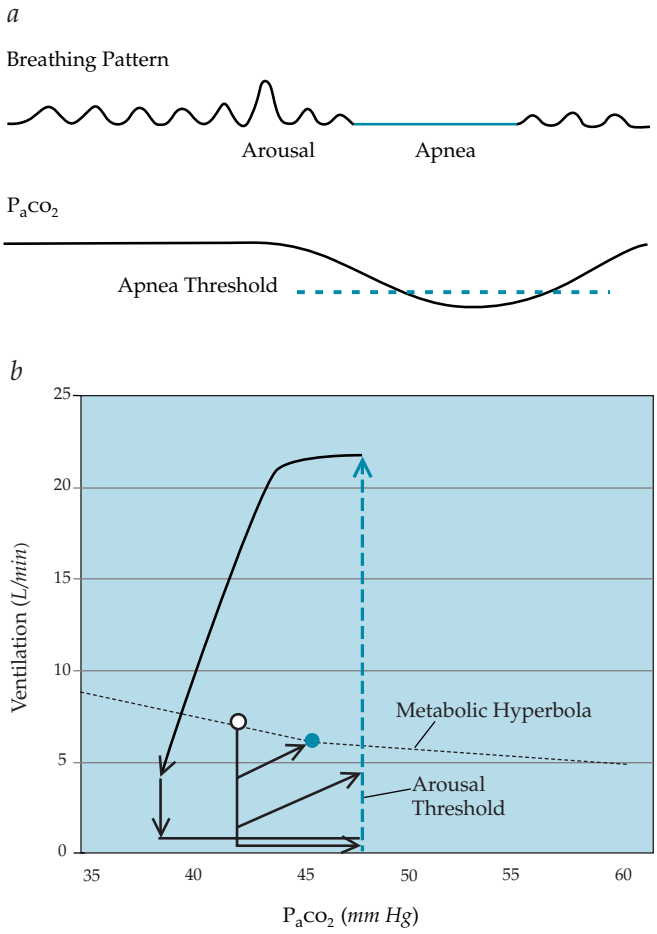


Figure 2 **(*a*) The existence of an apnea threshold (broken line) for arterial carbon dioxide tension (P_aCO_2) during sleep provides an explanation for the changes in breathing that occur at the onset of sleep. A transient increase in ventilation from brief arousal results in a lowering of P_aCO_2 below the apnea threshold. Breathing effort ceases until P_aCO_2 rises above the CO_2 threshold. The effect of lowering P_aCO_2 on ventilation is trivial in awake individuals. Sleep is associated with expression of this apnea threshold and may even raise the threshold, causing hypoventilation and apnea to occur more readily. Conditions that produce frequent arousals and large breaths may result in apneas or hypopneas as sleep resumes after the arousal. (*b*) The graph shows the cycles that occur when the set-point (white circle) is moved from wakefulness to sleep. With snoring, a new set point (blue circle) is established. A small increase in ventilation may lower the P_aCO_2, resulting in reductions in or cessation of breathing. P_aCO_2 will then rise and increase ventilation abruptly as the arousal threshold is reached. The length of an apneic episode depends in large part on the arousal threshold, the recovery mechanisms, and the tendency for sleep to persist without an arousal.**

Disorders with Decreased Ventilatory Drive

Hypoventilation occurs when alveolar ventilation is insufficient to eliminate metabolically produced CO_2. Hypoventilation may be caused by metabolic or mechanical factors. Metabolic causes of hypoventilation may include metabolic alkalosis, deficiency of thyroid hormone, and excess doses of sedative and narcotic agents. In each of these conditions, there is a relatively steady breathing pattern accompanied by a lowered respiratory rate, a lowered tidal volume, or both. Dyspnea is often absent despite an elevation of resting P_aCO_2.

In diseases that mechanically restrain ventilation (e.g., ankylosing spondylitis and chronic obstructive pulmonary disease [COPD]), hypoventilation may occur despite preserved activity of the medullary inspiratory neurons.[10] A perception of dyspnea occurs because of the increased work of breathing and the incongruity between central inspiratory activity and activation patterns detected by mechanoreceptors of the chest wall and lungs. In some persons, ventilation may be reduced to a degree that is out of proportion to the mechanical properties of the lungs or chest wall (possibly because of a genetic predisposition), and dyspnea may be a less prominent feature. As hypoventilation becomes chronic, adaptation of receptors, of central inspiratory neurons, of metabolic alkalosis, or of all three may occur. Adaptation to chronic hypoventilation in sleep apnea, COPD, neuromuscular disease, and chest wall disease may eventually depress responsiveness to CO_2 and depress ventilation during rest. Both resting CO_2 and ventilatory responsiveness to CO_2 may be increased by treatment of sleep apnea or by lowering CO_2 with ventilatory support. Hypoventilation can also be caused by hemodialysis, during which CO_2 removal lowers the P_aCO_2 sufficiently to depress the rhythmic activity of medullary respiratory neurons and produce apneas.

An uncommon condition called primary alveolar hypoventilation can be present at birth (congenital hypoventilation syndrome)[11] or can be acquired as a result of morbid obesity, cerebrovascular accidents, meningitis, encephalitis, bulbar poliomyelitis, or damage to afferent pathways in the cervical spinal cord. In all these conditions, however, no structural abnormality is found at autopsy. Presenting symptoms, which

are a result of blood gas derangement, often include lethargy, somnolence, headaches, and dependent edema. Such patients may not complain of shortness of breath. Physical findings may include cyanosis and evidence of right-sided heart failure. Secondary erythrocytosis is common. In congenital hypoventilation syndrome, alveolar ventilation is improved by exercise, indicating that the disturbance in ventilatory control is functionally determined.

The diagnosis of hypoventilation syndrome is one of exclusion and is considered when hypercapnia cannot be accounted for by disorders of the lungs, chest wall, respiratory muscles, or breathing during sleep. All hypoventilation syndromes worsen during sleep. In extreme cases, breathing occurs only with voluntary efforts and ceases entirely with inattention or during sleep. Management of hypoventilation related to CNS defects includes ventilatory assistance at night with or without respiratory stimulants such as medroxyprogesterone.

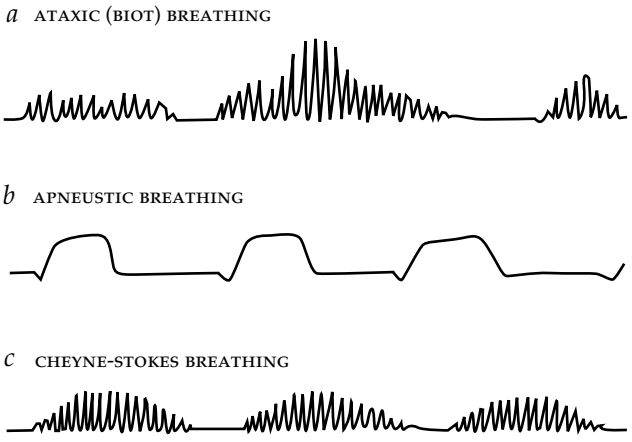


Figure 3 **Irregular breathing patterns may reflect central nervous system disease or an inherent alteration in the apneic threshold. Three examples of irregular breathing are illustrated: (*a*) Ataxic breathing is characterized by an unpredictable sequence of breaths varying in rate and depth and is associated with medullary disease. (*b*) Apneustic breathing involves repetitive gasps, with pauses at full inspiration lasting a few seconds, and is associated with pontine disease. (*c*) Cheyne-Stokes respiration is cyclic, with a crescendo-decrescendo pattern interrupted by apneas.**

Abnormal Breathing Patterns and Sleep Reports

Brain injury and certain drugs and toxins affect breathing patterns during both wakefulness and sleep; however, there is a growing awareness of how abnormal breathing only during sleep may affect health. Periods of cessation of airflow into and out of the lungs (apnea) regularly occur at sleep onset, and episodes of partial upper airway obstruction during inspiration (snoring) are very common. Some irregularity of breathing is considered normal during sleep, including mild CO_2 retention and a reduction in P_aO_2, as well as irregular breathing at sleep onset or with dreaming. As with many biologic phenomena, breathing irregularities that occur during sleep are designated as abnormalities only if they are sufficient in magnitude and frequency to disrupt sleep continuity or impair oxygenation enough to affect a person during wakefulness.

ataxic and apneustic breathing

Ataxic (Biot) breathing is a random pattern of shallow and deep breaths interspersed with irregular pauses [*see Figure 3*]. Ataxic breathing results from disruption of medullary neural pathways by trauma, hemorrhage, or extrinsic compression caused by cerebellar or pontine hemorrhage; it can be seen in terminally ill patients because respiratory control systems are affected by multisystem failure.[12] Complete apnea may ensue, especially in patients given sedative or narcotic drugs. Another disturbance, apneustic breathing, is characterized by an end-inspiratory pause of 2 to 3 seconds before exhalation is begun [*see Figure 3*]. Apneustic breathing is associated with caudal pontine lesions and is sometimes intermixed with ataxic breathing patterns.

Three patterns of apnea, or cessation of breathing, can be observed during sleep. These apneas are defined as episodes of a reduction in airflow of more than 80% occurring for more than 10 seconds.[13] Apneas may be classified as central (or nonobstructive), obstructive, or mixed [*see Figure 4*]. In central apnea, which implies a cessation of respiratory activity at a brain stem level, both airflow and respiratory efforts are absent. During obstructive apnea, respiratory efforts persist, although airflow is absent at the nose and mouth. Obstructive and central apneas are related clinically and pathophysiologically. Many adult patients exhibit mixed apneas in which both central and obstructive patterns occur. In a single apneic episode, there may be a period in which no efforts occur, followed by the appearance of respiratory efforts, also without airflow. In addition, in the same night, patients may have all three types of apneas in varying proportion. If more than 80% of apneas are of a central type, the patient is classified as having central sleep apnea. If apneas are predominantly mixed and obstructive apneas, the patient is classified as having obstructive apnea.[13]

Hypopneas or hypoventilation during sleep may arise by mechanisms similar to those producing apnea. Hypopneas are defined as episodes of a reduction in airflow of 30% to 80% occurring for more that 10 seconds.[13] Hypoventilation (hypopnea) leads to increased CO_2 and decreased O_2 levels in arterial blood and causes arousals from sleep; as with apneas, hypopnea may result from reduction in respiratory efforts or partial upper airway obstruction. Snoring is a form of partial airway obstruction and is called obstructive hypopnea. Snoring is common, but some patients who snore have symptoms similar to those of sleep apnea syndrome even if complete cessation of airflow (apnea) never occurs during sleep. Moreover, such patients may exhibit abnormal sleep and cardiorespiratory changes.

ventilatory behavior in sleep

The transition from wakefulness to non–rapid eye movement (NREM) sleep is accompanied by a reduction in metabolic rate and therefore a reduced need to breathe. Consequences of sleep onset include reduced tidal volume, changes in lung mechanics, reduced activity and upper airway dilators, reduced upper airway caliber, and loss of load compensation [*see* Load Compensation, *below*].[14]

Sleep is accompanied by reduced postural muscle tone. In NREM sleep, the ratio of rib cage displacement to abdominal displacement is greater than it is during wakefulness, whereas in REM sleep it is less.[15] These changes in displacement may affect the distribution of ventilation in the lungs, increasing ventilation-perfusion mismatching and contributing to hypoxia; the development of hypoxia, in turn, may necessitate changes in respiratory output, which may initiate an unstable breathing pattern.

Upper Airway Function

Upper airway caliber is reduced during sleep, and air passage is further impaired by decreased activity of upper airway muscles,[16,17] especially the muscles involved with tonic activity (independent of the phase of respiration), such as the tensor veli palatini muscle.[18] The mechanical consequence of reduced airway caliber is increased upper airway resistance.[19] Because pharyngeal compliance increases during NREM sleep, negative intrathoracic pressures normally produced in the upper airway during inspiration will result in airway collapse. Even in healthy persons, negative intrathoracic pressure during NREM sleep limits inspiratory flow, resulting in an inspiratory plateau that persists in the presence of increasing negative pressure.[19]

Curiously, the retropalatal airway is less compliant during REM sleep, when muscle activity is much reduced, than during NREM sleep.[20] This finding points to the significance of nonneuromuscular factors (e.g., bony and cartilaginous support) in the maintenance of upper airway patency during sleep.

Load Compensation

When the ratio of load to inhalation is increased (whether because of resistive factors or obstructive factors), a concomitant increase in breathing effort is required to restore tidal volume (i.e., load compensation). During sleep, however, immediate and subsequent load compensation is compromised and results in decreased tidal volume and minute ventilation, which thereby results in alveolar hypoventilation. The ensuing elevation of arterial P_aCO_2 restores CO_2 elimination toward normal levels.[7] The inability to perceive and immediately respond to increased loads allows for sleep to continue undisturbed. Thus, the main consequence of sleep is an increase in P_aCO_2 of 4 to 5 mm Hg. Such elevations in P_aCO_2 result in mild acidosis in both healthy persons and in persons with cardiopulmonary disorders but without sleep-disordered breathing (SDB).

Heavy snorers may not arouse from sleep despite continuous generation of subatmospheric intraluminal pressure that is several times higher than that which occurs during wakefulness [*see Figure 5*]. If increased resistance and inspiratory flow limitation are prolonged, the increased work of breathing or hypoventilation, or both, leads to respiratory-related arousals (RERA) from sleep. Partial obstruction of the upper airway (with RERAs) and daytime sleepiness are the features associated with upper airway resistance syndrome (UARS).[21]

The Hypocapnic-Apneic Threshold

In NREM sleep, a highly reproducible hypocapnic-apneic threshold is unmasked, and a central apnea will occur if the P_aCO_2 is lowered, even by a small amount.[22] As a result, hypocapnia is the most important inhibitory factor to breathing during NREM sleep. This threshold level of P_aCO_2 is decreased by hypoxia, possibly by excitation caused by miscellaneous nonchemical stimuli. One major cause of SDB is breathing instability produced by this threshold effect and by arousals, hypoxia, and other factors that alter this threshold over time.

Sleep Effects on Cardiovascular Physiology

The cardiovascular system adjusts to the changes in gas exchange that accompany sleep and to the apneas and hypopneas that may interrupt sleep. Normally, during NREM sleep there is a withdrawal of sympathetic tone, both neural and humoral, and an increase in parasympathetic tone—changes that result in a reduction in heart rate, blood pressure, and cardiac output.[23] The decreased cardiac workload and O_2 demand are accompanied by a diminished ability to provoke an arrhythmia.

Gradual awakening is accompanied by a modest increase in sympathetic outflow without much evidence of parasympathetic withdrawal. In contrast, with an abrupt arousal caused by noise or sleep apnea, there occurs an abrupt increase in sympathetic drive manifested by increases in blood pressure and heart rate and by marked parasympathetic withdrawal.[24]

In REM sleep, cardiovascular and breathing systems are relatively independent of metabolic drive and inhibition of muscle activity. Sympathetic activation increases to levels seen during wakefulness but is often episodic, leading to transient changes in heart rate, blood pressure, and breathing. Such surges in blood pressure may play a part in triggering ischemic events in patients with heart disease or diabetes.[23]

In general, however, sleep is cardioprotective in healthy persons. Sleep apnea disrupts cardiovascular regulation during sleep because of repetitive arousals, hypoxemia, and increased

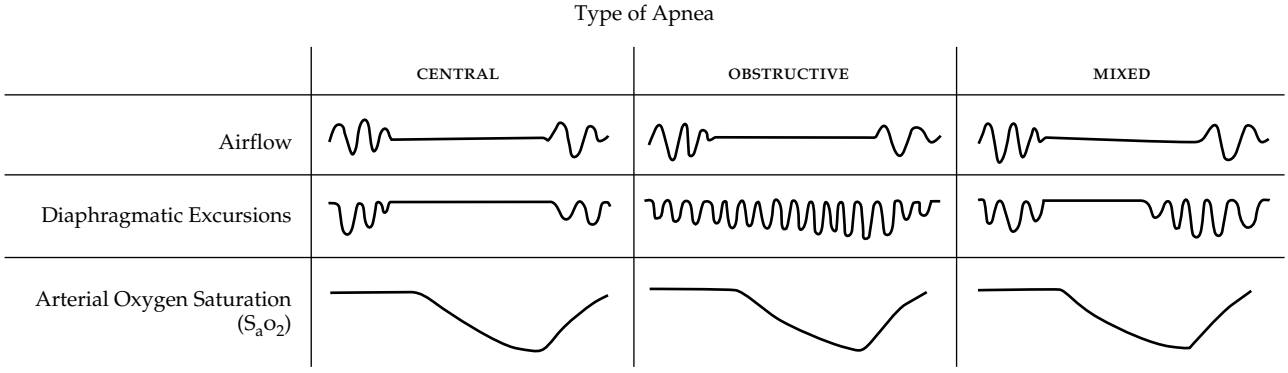


Figure 4 **This schematic representation of the ventilatory signals recorded during a sleep study (polysomnogram) illustrates the different patterns found in central, obstructive, and mixed apneas. In each example, the presence of apnea is confirmed by the cessation of airflow at the nose and mouth (top), and the consequence of apnea—namely, hypoxemia—is demonstrated by the development of oxygen desaturation on the continuous record of arterial oxygen saturation (S_aO_2) (bottom). The three types of apnea are distinguished by the respiratory efforts made during the episode (middle). In central apnea, no respiratory efforts are made; in obstructive apnea, diaphragmatic contractions continue and often intensify during the episode; and in mixed apnea, a period of absent respiratory efforts is followed by active inspiratory muscle contractions against an occluded upper airway.**

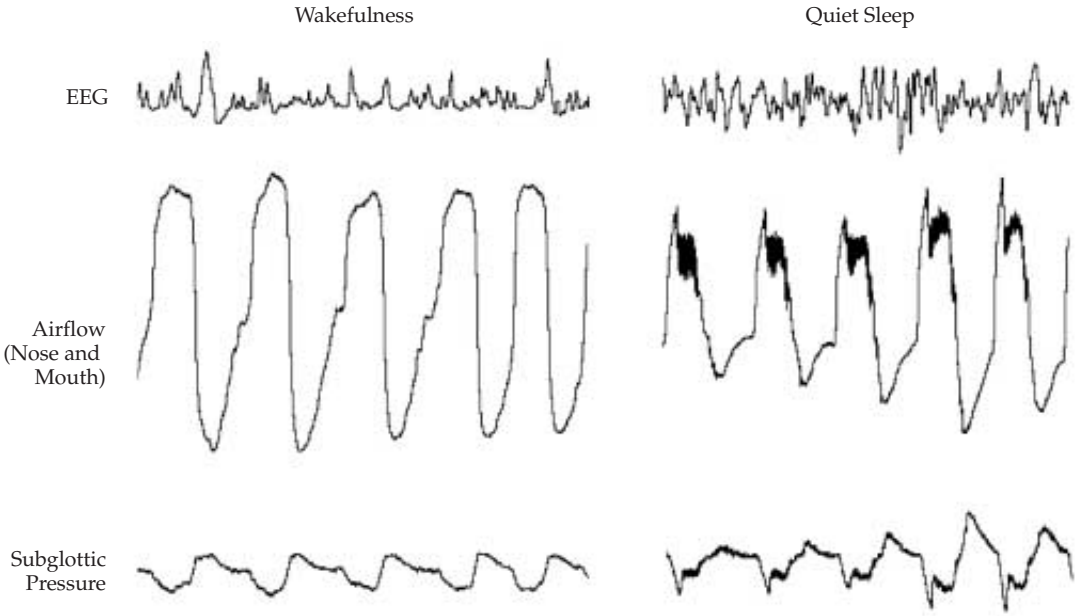


Figure 5 **Several cycles of breathing during EEG-defined wakefulness and sleep are shown for wakefulness (left) and quiet sleep (right) in a healthy person. Flow limitation is present during inspiration, with a chopped-off flow pattern and rapid fluttering of flow indicative of snoring. This pattern can represent a steady-state condition, balancing load compensation and chemical drive.**

intrathoracic pressure changes, which result in preload and afterload effects on the heart.[25]

EVALUATING SLEEP DISTURBANCES

Monitoring a person with electrodes during sleep results in a classification of sleep into two states: non–rapid eye movement, or NREM, sleep and rapid eye movement, or REM, sleep. NREM sleep can be further subdivided into stages I and II (light or transitional sleep) and stages III and IV (deep sleep), depending on the frequency and amplitude of brain waves. States are distinguished by recording electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) measurements. The combination of these measures and the cardiopulmonary monitoring of airflow, respiratory effort, oxygen saturation, and heart rate—along with identification of body position—constitute polysomnography, which is a common test used to diagnose sleep apnea.

Respiratory Disturbance

Various measurements are used to quantify respiratory disturbances during sleep.[15] The apnea-hypopnea index (AHI) is the total number of apneas and hypopneas occurring during sleep divided by the hours of sleep time. Values of AHI can be computed for the different stages of sleep. Another term for AHI is the respiratory disturbance index (RDI). The term desaturation index, also called oxygen desaturation index, refers to the number of times per hour that O_2 saturation falls by more than 3% to 4%, and it may be reported as an independent measure of cardiorespiratory instability. The snoring index (SI) is the percentage of time spent snoring during sleep.

The arousal index (AI) is computed as the number of transient awakenings per hour, and it is defined by a change in state from sleep to waking that is longer than 2 seconds but less than 3 minutes.[13] This number is used to estimate individual exposure to transient arousals from sleep, and it is distinguished from nocturnal awakenings by the length of the bout of wakefulness. Included in this index are spontaneous brief awakenings caused by external and internal stimuli (e.g., noises and leg jerks, respectively). The AI may differ from the AHI or RDI because many (approximately 20%) apneas or hypopneas are not accompanied by arousals, and because the AI count includes arousals that are not apnea-induced.

Oxygen Saturation

Various measurements of O_2 saturation, as plotted over time, indicate the extent of exposure to hypoxemia during sleep. One measure is the O_2 saturation profile, in which values of O_2 saturation are presented in the frequency domain, which plots the pattern and extent of O_2 deficiency during sleep.[26] Values reported include estimations of the lowest O_2 saturation and the length of time spent below a specific O_2 saturation (e.g., 90%, 85%, 80% O_2 saturation of hemoglobin). In addition, recordings can be examined in the time domain to estimate the degree to which O_2 saturation exhibits periodic behavior.

Hypoventilation

Hypoventilation is not directly measured during routine sleep studies, because tests for arterial blood gases are uncomfortable and incur an unfavorable risk-to-benefit ratio. Markers for $P_{a}CO_2$ include end-tidal values of CO_2 or transcutaneous estimates of CO_2. Both are qualitative. The former makes the assumption of adequate sampling of alveolar gas, and the latter provides trends rather than precise numbers. Neither is used routinely during sleep studies in adults.

Sleep-Disordered Breathing

Healthy individuals may exhibit obstructive or central apneas at sleep onset or during periods of REM sleep.[27] Episodes are usually less than 15 seconds in duration and are not repetitive.

Occasionally, longer periods of apnea (lasting 30 seconds or more) are seen during REM sleep. These episodes may not be accompanied by arousal or sleep-state changes.

Healthy young men have more frequent apneas during sleep than young women, but after the sixth decade of life, respiratory disturbances during sleep increase in number and occur with equal frequency in men and women.[27] Patients with a clinically important sleep apnea may be distinguished from patients with normal respiratory disturbances by the presence of repetitive apneas longer than 10 to 15 seconds that occur during stages I and II of NREM sleep and during REM sleep and that are frequently accompanied by daytime sleepiness. If treated, patients with significant apnea show improvement in daytime symptoms and general performance.

DEFINITION OF SLEEP-DISORDERED BREATHING

In the United States, 9% to 12% of women and 27% to 35% of men may have an AHI greater than 5, a number often quoted as a threshold value for normality; however, many people with an AHI greater than 5 have no clinically apparent illness.[27] If the definition of illness is the presence of daytime sleepiness or cardiovascular complications such as hypertension, it is estimated that about 2% of women and about 4% of men have symptomatic SDB. Studies suggest that patients with symptomatic SDB who drive are at increased risk for vehicular accidents in which they may incur substantial disability. Medical practitioners often fail to recognize the presence of sleep apnea syndrome.

ETIOLOGY OF SLEEP-DISORDERED BREATHING

Predisposing Factors

Snoring is generally considered a predisposing feature in the development of SDB and symptoms of sleep apnea.[28] Snoring increases with age; approximately 45% of men and 30% of women 65 years of age or older are said to snore. Persons who snore are two to three times more likely to have hypertension[29] and 1.5 times more likely to have diabetes than people who do not snore, even after age and obesity are taken into account as risk factors for these diseases.[30]

Genetic Factors

Sleep apnea has a genetic component. Symptoms relating to apnea occur two to four times more often in family members of affected patients than in a control population. Sleep apnea events occur more often in first-degree relatives of sleep apnea patients than in control subjects matched for age, sex, and socioeconomic status. Such studies reveal that the symptomatic sequelae of multiple apneas are quite variable, probably because of an interaction between both genetics and the environment.[31]

Patients with sleep apnea exhibit a twofold increase in a polymorphism for apolipoprotein E associated with cardiovascular disease and Alzheimer disease.[32] Using a statistical approach to estimate inheritance, the Cleveland Family Study found that 27% of the variation in AHI in the community could be accounted for by perhaps only a few genetic factors.[33] Transmission patterns in both the white and the African-American patients were consistent with mendelian inheritance. Adjustment for body mass index (BMI) significantly reduced the significance of a genetic effect in whites but not in African Americans. Thus, an underlying genetic basis for sleep apnea could be independent of the contribution of BMI to the disease in African Americans.

Specific craniofacial morphology (e.g., a short mandible and round head) are predisposing factors for the development of snoring, apneas, or both.[34] It is also known that there are familial traits in hypercapnic and hypoxic sensitivity; these could relate to the tendency to breathe periodically during sleep.[35] In addition, obesity and alcoholism (factors associated with SDB) can be family traits and, to the extent that these factors are causally related to apneas, may account for the familial clustering of sleep apnea. It is not known whether there is a familial trait involving the respiratory coordination of muscles of the chest wall and upper airway. A role for genetic transmission of ventilatory behavior (respiratory frequency, tidal volume, and minute ventilation) is directly supported by reports of nearly absent respiratory depression in several gene knockout models and by studies of inbred rat and mouse strains.[36] Given the current evidence, sleep apnea does not appear to be the result of a single mutation or protein action.

Central Sleep Apnea

PATHOPHYSIOLOGY

The instability of breathing that occurs with central apnea or hypopnea reflects brain stem interplay between the sensitivity to CO_2 and the hypocapnic-apneic threshold. The inhibitory effects of sleep on ventilatory responsiveness are offset by neural mechanisms that stabilize ventilation. One such mechanism is short-term potentiation (STP), or a transient increase in ventilation occurring after a large breath (a sigh) or an apnea. STP may be abolished by prolonged hypoxia, which may explain the development of periodic breathing in patients at high altitudes or in patients with cardiopulmonary disease who have modest hypoxemia. Finally, upper airway obstruction may reflexly inhibit central neural ventilatory output and provoke central apnea in some patients.[37]

The occurrence of a central apnea or, for that matter, any apnea appears to set in motion events that conspire to promote further breathing instability [*see Figure 6*].[38] First, time delays in the ventilatory control system prevent resumption of rhythmic breathing after apnea until arterial CO_2 levels increase by 4 to 6 mm Hg above the set point. Second, central apnea is associated with narrowing or occlusion of the pharyngeal airway.[39] Resumption of ventilation thus requires opening of a narrowed or occluded airway, which involves overcoming mucosal adhesion and gravitational forces. (This narrowing of the upper airway may explain the overlap between central and obstructive apnea, or mixed apnea, and the successful use of nasal continuous positive airway pressure [CPAP] in some patients with central sleep apnea.) Third, a combination of hypoxia, hypercapnia, and transient arousal results in ventilatory overshoot, subsequent hypocapnia, and further apnea or hypopnea.

Mathematical models and studies in humans have focused on statistical correlations between the incidence of periodic breathing and the incidence of hypoxic sensitivity.[40,41] Such periodic breathing during sleep occurs more frequently in individuals with higher peripheral chemosensitivity.[42,43] Alternative explanations for repetitive apneas during sleep are that (1) patients with repetitive apneas have the same oscillations as normal individuals, but excitatory stimuli contribute to a larger amplitude of these oscillations in sleep apnea,[41] or (2) recurrent apneas result from an intrinsic property in the feedback control of breathing in regard to either stability or instability in ventilation over time.[44]

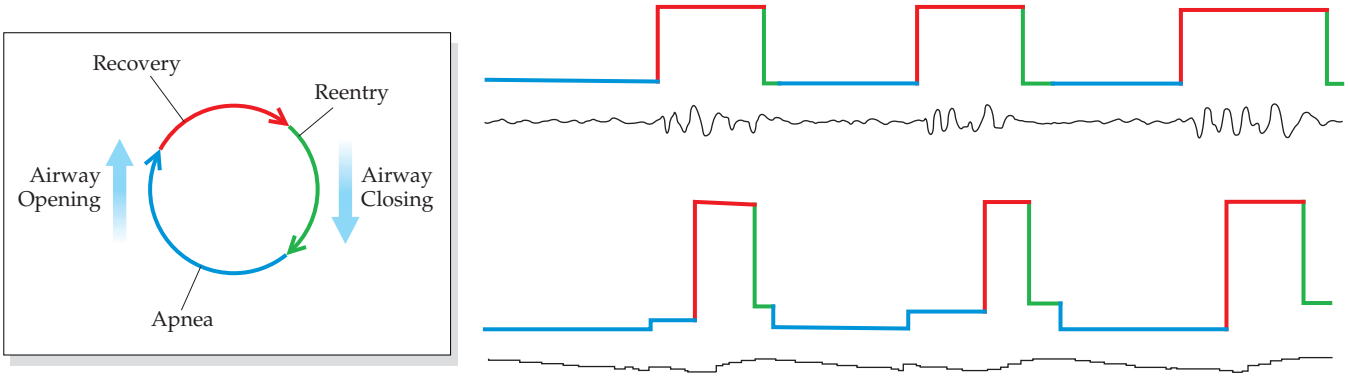


Figure 6 **Apneas can be thought of as a recurrent or "reentry" arrhythmia of ventilation, characterized in this example by a relatively long apnea followed by phases of recovery and reentry. Such periodicity is evident in both airflow and oxygen saturation measurements.**

RISK FACTORS

Several factors predispose persons to sleep apnea, and some of these factors are more selective for central apnea than for obstructive events [*see Table 1*].

Central sleep apnea is also more common in stroke patients than in control subjects matched for age, weight, and BMI.[45] Interestingly, there is no difference between the prevalence of sleep apnea in patients with hemispheric involvement and that in patients with brain stem involvement. Although central sleep apnea is associated with stroke, the natural history and consequences of central sleep apnea in stroke patients remain uncertain. Available data on the effects of sleep apnea on blood pressure and sympathetic motor output suggest that it is prudent to identify and treat sleep apnea in the poststroke period.[46]

Patients with hypothyroidism and renal failure have an unexpectedly high prevalence of both central and obstructive sleep apnea (OSA).[47,48] Similarly, patients with acromegaly have a high rate of central and obstructive apnea, which correlates with higher biochemical markers of disease activity and higher chemoresponsiveness.[49-51]

DIAGNOSIS

Clinical Presentation

Cheyne-Stokes respiration, or CSR, is the signature feature of central apnea. CSR is marked by a crescendo-decrescendo pattern of breaths, followed by central apneas or central hypopneas, some lasting as long as 30 seconds [*see Figure 7*]. Patients are often hypocapnic during wakefulness and sleep. CSR commonly seen in healthy persons at high altitude results from hypoxic stimulation of breathing and resultant hypocapnia. Conditions that promote CSR at normal altitude include hypoxia, decreased lung volume, decreased metabolic rate, renal failure, and cerebrovascular disease.

CSR occurs in 25% of congestive heart failure (CHF) patients. A seemingly minor ventilatory disturbance may initiate an appropriate change in ventilation and overcompensation. Then, a change in ventilation in the opposite direction occurs, with overshoot, and an oscillating breathing pattern is established.

Diagnostic Testing

The severity of central apneas during sleep is determined by a polysomnographic study of the frequency of respiratory events per hour of sleep (i.e., the apnea-hypopnea index or respiratory disturbance index), the severity and frequency of O_2 desaturation or hypercapnia, changes in sleep-stage distribution, and clinical symptoms produced by the disorder.

TREATMENT

Central apneas in otherwise healthy persons are usually not treated. Approximately 25% of patients with CHF, even those who are well compensated, have clinically significant central sleep apnea.[52] CHF patients with CSR have a higher mortality and a greater need for cardiac transplantation than CHF patients without CSR events.[53,54] The clinical management of CSR in the setting of CHF is well documented, and data indicate CPAP as the recommended therapy.

In a randomized study, CHF patients with CSR and central sleep apnea who received treatment with CPAP for up to 6 hours a night had a significantly greater rate of transplant-free survival than control subjects, who did not receive CPAP.[55] In patients with stable chronic CHF who have CSR, CPAP has been shown to reduce CSR, left ventricular afterload, plasma catecholamine

Table 1 Risk Factors for Sleep Apnea with Associations to Apnea Type

Risk Factors	*Apnea Type*
Increasing age	CSA, OSA
Male gender predominance	CSA, OSA
Family history of apnea or snoring	OSA
Head form (craniofacial morphology)	OSA
Poor physical fitness	OSA
BMI	CSA (lower BMI); OSA (higher BMI)
Alcohol ingestion	OSA
Smoking exposure	OSA
Sleep restriction	OSA
Cardiovascular disease	CSA, OSA

BMI—body mass index CSA—central sleep apnea and Cheyne-Stokes respiration OSA—obstructive sleep apnea

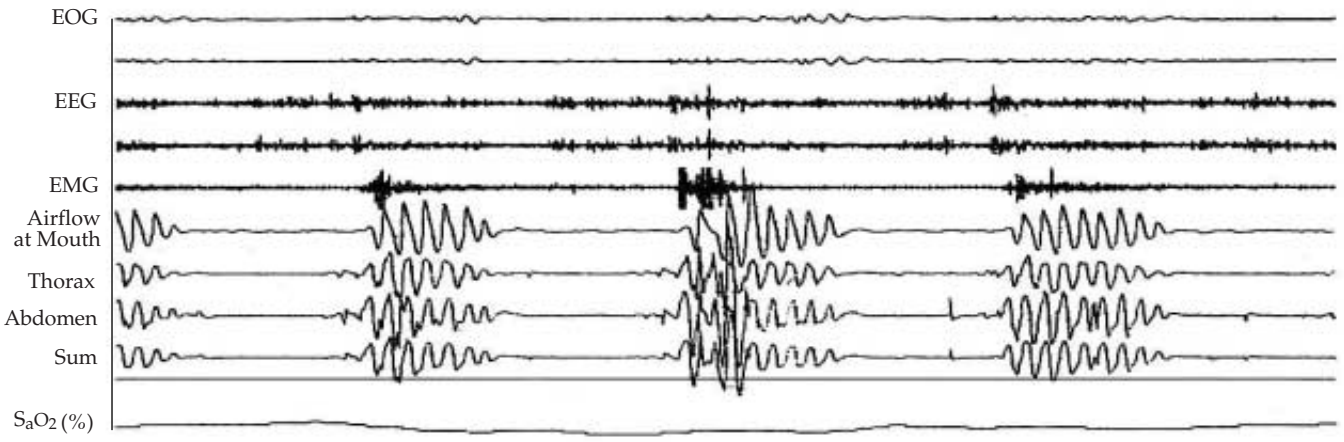


Figure 7 **A fragment of a polysomnogram shows a series of central apnea episodes during slow wave sleep. Each episode of absent flow is accompanied by a cessation of thoracoabdominal efforts. Central sleep apnea has this pattern as its signature.**

levels, and sympathetic nervous system activity and to improve left ventricular ejection fraction and quality of life.[56] Application of CPAP therapy is discussed elsewhere [*see* Obstructive Sleep Apnea, Continuous Positive Airway Pressure, *below*].

Supplemental O_2 has also been shown to reduce the severity of CSR, decrease urinary catecholamine levels, and improve exercise capacity.[57] Theophylline reduces the severity of central sleep apnea[43]; however, long-term use of theophylline is associated with proarrhythmic effects,[58] and sustained benefits are unproven. Thus, further study is necessary before theophylline or O_2 therapy can be recommended over CPAP.

Nonhypercapnic idiopathic central sleep apnea without CSR may improve with supplemental O_2 therapy or treatment with acetazolamide. Treatment for hypercapnic central sleep apnea, a disorder of the ventilatory control system, should be aimed at improving alveolar ventilation. Noninvasive ventilation or tracheostomy and mechanical ventilation should improve P_aCO_2, cor pulmonale, and symptoms of daytime hypersomnolence.

Obstructive Sleep Apnea

The fundamental feature of obstructive apneas and hypopneas, including snoring, is a functional narrowing or collapse of the upper airway. Illness occurs in the context of repetitive events.

PATHOPHYSIOLOGY

The presence of an anatomic abnormality is not sufficient or necessary to produce overt disease. Nevertheless, there is evidence that persons with OSA have a smaller-caliber pharyngeal airway during wakefulness than the airway in persons without OSA. In addition, the pharyngeal airway in patients with OSA syndrome has an elliptical anterior-posterior configuration, in contrast to a more circular configuration in normal persons.[34] This asymmetry seems to predispose patients to anterior-posterior collapse.

A collapsing transmural pressure is generated either by a negative intraluminal pressure or a collapsing surrounding pressure; however, pharyngeal obstruction does not require negative pressure, because complete upper airway collapse occurs during central apnea.[35] The occurrence of upper airway obstruction in the absence of negative intraluminal pressure is consistent with the hypothesis that the intrinsic properties of the upper airway (e.g., stiffness of the pharyngeal wall) will or will not permit collapse to occur.

Another factor in airway collapse is the mechanical interdependence of the upper airway and the thoracic cage and mediastinum. An increased lung volume is associated with increased upper airway caliber in awake humans, independent of muscle dilatation in the upper airway.[59] Caudal traction may stiffen the pharyngeal airway and permit greater dilating force or, at the very least, prevent inspiratory collapse, both of which effects have been demonstrated in model systems. Therefore, a reduction in functional residual capacity could also contribute to a reduction in upper airway patency during sleep.

Proposed mechanisms for the pathophysiology of OSA involve alterations in the neuromuscular control of upper airway muscle, the resting size of the upper airway, and the degree of stiffness of the upper airway wall. The underlying defect is a pharynx that is susceptible to narrowing and collapse. The change in respiratory drive that occurs with sleep onset leads to reduced ventilatory motor output to upper airway muscles, which triggers the cascade of events leading to pharyngeal obstruction during sleep. Upper airway obstruction often occurs during experimentally induced periodic breathing at the lowest point of respiratory drive. Central ventilatory control instability is a key mechanism for repetitive obstructive apnea.[60]

RISK FACTORS

Obesity is the strongest risk factor for OSA [*see Table 1*], and the prevalence of OSA substantially increases with age. OSA has a strong familial component that likely involves multiple genetic influences. Several anatomic abnormalities may predispose a patient to OSA; in men, a large neck size (>17 in) may be a risk factor for OSA.[61] Additional risk factors include ethanol and nicotine use, because these agents promote airway collapse.

OSA-ASSOCIATED MORBIDITY AND MORTALITY

The mortality associated with sleep apnea has not been satisfactorily explained. Early reports suggested mortality from cardiorespiratory failure, pulmonary embolus, and renal failure. Death has been reported to result from sedative drug use, particularly preoperative medications, which suppress breathing and the arousal response to an obstructive apnea. Automobile acci-

dents related to excessive daytime sleepiness may have a substantial impact on morbidity and mortality. Taking a broader view, sleep quantity and quality are associated epidemiologically with hypertension, all-cause cardiovascular risk, and early mortality.[62-64]

The occurrence of sleep apnea with the features of syndrome X—hypertension, obesity, diabetes, and hyperlipidemia—has prompted a call to rename this disorder syndrome Z.[65] This cluster of diseases and disorders may occur through a common set of neuroendocrine factors, genetic predispositions, or both. Both the physiologic plausibility that sleep problems relate to insulin resistance[66] and the epidemiologic association between snoring and cardiovascular risk factors offer a rationale to explore the pathogenesis of OSA as it relates to obesity, race, and cardiovascular disease and the pathogenesis of central apnea as it relates to aging, cardiovascular disease, and stroke. It may be that prevention of disease progression in sleep apnea may be nested within the factors that increase cardiovascular risk—namely, obesity, hypertension, lack of exercise, and alcohol use.

DIAGNOSIS

Clinical Presentation

Self-reported snoring and excessive daytime sleepiness are the major presenting symptoms of OSA. One might suspect that the bed partner rather than the patient would report snoring more reliably; however, reports by the patient and bed partner are similar.[67] Other complaints include apneas, choking during sleep, restless sleep, and, in some patients, insomnia-like symptoms along with excessive daytime sleepiness.

Diagnostic Testing and Imaging

Overnight polysomnography is the current gold standard for the diagnosis of OSA and should be considered in any patient suspected of having this disorder [*see Figure 8*]. Split night polysomnograms provide adequate time for both diagnosis and proper CPAP prescription. Home-based studies provide data comparable to data from laboratory studies; however, devices vary substantially. Oximetry alone cannot be recommended as a screening tool for OSA. Nasal-pressure changes detected during inspiration and expiration reflect changes in airflow with greater accuracy than thermistors.

TREATMENT

OSA patients should be counseled about the risk of sleepiness and offered therapy when it is clinically indicated.[68]

Therapy for sleep apnea is diverse and includes correction of associated medical conditions (e.g., cardiopulmonary disease) and treatment of SDB by specific interventions (e.g., surgery or use of mechanical aids). The principles of management are directed toward improvement in gas exchange, sleep continuity,[15] and chronic cardiopulmonary symptoms during wakefulness. Management approaches start with consideration of mechanical devices; when mechanical devices prove inadequate, surgery to remove nasal or pharyngeal obstruction is considered.

Treatment is customized to the individual patient and may require avoidance of agents that provoke apneic episodes (e.g., alcohol, sedatives, and androgens). Obese patients with OSA who are treated for obesity may show improvement in the severity of sleep apnea, but long-term weight reduction is often difficult to achieve. Dietary management alone is effective in producing long-term weight reduction in a small minority of morbidly obese patients. However, even modest weight reduction achieved by dieting can improve SDB. Gastric stapling can achieve weight reduction in approximately 60% of patients who undergo the procedure, and appears to be effective in improving or reversing sleep apnea in most morbidly obese patients.

Continuous Positive Airway Pressure

Indication CPAP is considered the first choice of treatment for OSA.[69] Treatment with CPAP improves vigilance and cognitive function in persons who report hypersomnolence; however, patients who do not have daytime hypersomnolence may not experience substantial improvement of cognitive function with CPAP.[70] Treatment with CPAP will reduce blood pressure in normotensive and hypertensive patients; in the latter patients, control of blood pressure with drugs may become easier, or

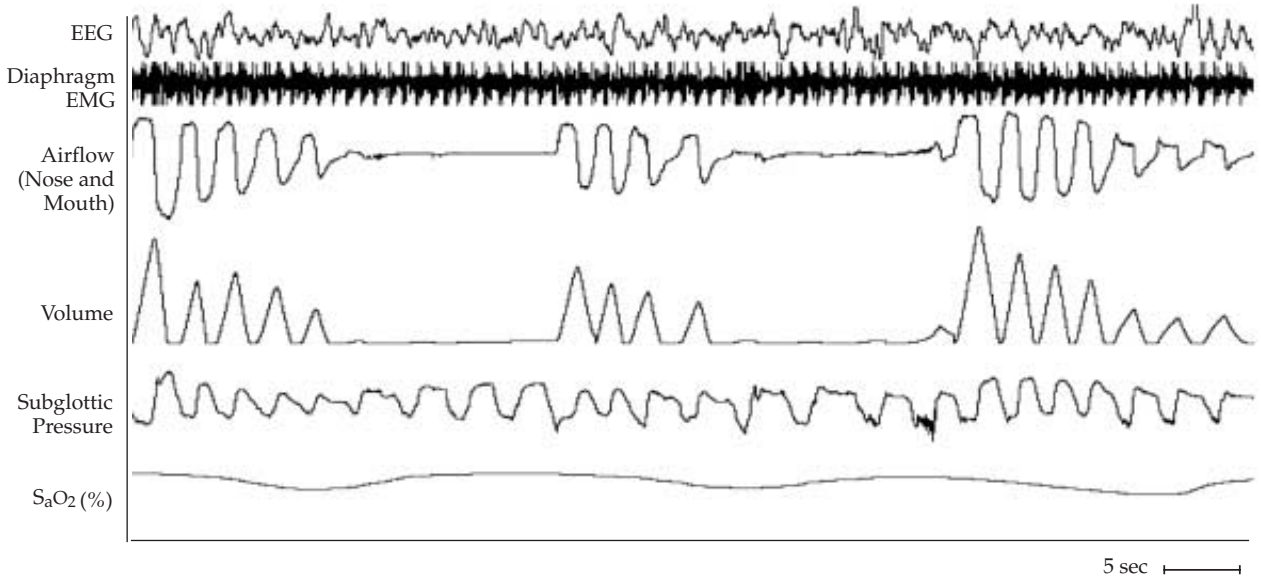


Figure 8 **A fragment of a polysomnogram shows a series of obstructive apneas during sleep. Each episode of absent flow is accompanied by continued thoracoabdominal efforts. Apneas are terminated by arousals. Obstructive sleep apnea has this pattern as its signature.**

drugs may no longer be needed.[71] Contraindications include bronchopleural fistula, acute pneumothorax, and sinus-communicating pneumocephalus. A relative contraindication is nasal or pharyngeal obstruction, which must be managed before CPAP treatment is attempted.

Application Nasal CPAP is immediately effective in reversing airway obstruction in most patients with OSAS. The patient is outfitted with a nasal mask, which is attached to a blower that applies an adjustable positive pressure. The amount of pressure required to relieve upper airway obstruction (usually 5 to 20 cm H_2O) is most often determined empirically during a sleep study. Effective CPAP is defined as the pressure required to prevent inspiratory collapse of the upper airway when the patient is in the supine position and during all stages of sleep. Positive pressure applied to the nose presses the soft palate to the back of the tongue and thereby prevents leakage of air out of the mouth. If the apparatus fails, the patient is free to breathe through the oral cavity. Full-face masks or chinstraps may be needed for patients in whom air escapes through the mouth. Monitoring of pressure can be used to document the hours of use of CPAP and the effectiveness of CPAP in maintaining airway patency. Heated-humidification devices may increase patient compliance and reduce symptoms of nasal obstruction. For patients in whom standard CPAP is unsuccessful or intolerable, bilevel or self-adjusting modes may be attempted.

The routine use of bilevel therapy (i.e., devices that sense inspiration efforts and apply a higher positive pressure during inhalation) does not improve patient adherence; however, bilevel pressure is able to eliminate apneas with lower expiratory pressure and mean airway pressure in patients who are unable to tolerate high pressures. Devices that sense and then automatically triturate pressure to eliminate apneas operate in a way similar to personal titration; the routine use of unattended devices in the home is currently under study.

Oral Appliances

Intraoral devices to reposition the mandible are increasingly being used in the management of sleep apnea, both as primary therapy for heavy snoring and as an alternative therapy for patients who cannot tolerate CPAP.[72] An oral appliance will reduce the apnea-hypopnea index and the arousal index and will improve O_2 saturation even in patients with mild to moderate OSA. Oral appliances are modestly preferred over CPAP; however, CPAP is more effective than oral appliances in reducing the apnea-hypopnea index.[73]

Surgery

Surgical management of OSA encompasses several strategies[74-76]: (1) bypass of the anatomic obstruction by way of tracheostomy, (2) alteration of the bony structural support of the upper airway, and (3) alteration of the soft tissue attachments or deposits to improve airway patency. Of the surgical options, only tracheostomy is routinely effective in eliminating OSA. Central apneas may persist after tracheostomy. Despite the development of a new flap-type tracheostomy and low-profile tracheostomy tubes, local stoma problems are relatively common. Infections, local discomfort, formation of granulation tissue, and the distortions of self-image produced by tracheostomy have limited the use of this procedure to relatively few patients. Adaptation to a tracheostomy often takes a year or more.

Pharmacologic Therapy

Pharmacologic therapy plays a very limited role in the management of upper airway obstructions during sleep. Modafinil is the only drug approved by the Food and Drug Administration for treatment of OSA,[77] and it is effective for the management of residual sleepiness after treatment with CPAP. Use of modafinil may be considered when CPAP compliance is acceptable and other behavioral issues, such as sleep restriction, have been addressed.

Medroxyprogesterone is not effective therapy for OSA in patients with normal levels of P_aCO_2 , but it may be used in the management of obesity hypoventilation syndrome. Acetazolamide therapy is not effective treatment for OSA, but it will reduce the CO_2-apneic threshold and thus may improve central sleep apnea. Protriptyline (a nonsedating antidepressant) increases alertness and may modestly reduce the apnea-hypopnea index, but its anticholinergic side effects limit its routine use. Nicotine may modestly reduce the apnea-hypopnea index in patients with OSA; however, current delivery methods and available doses do not have clinical utility. Serotonin agonists do not generally affect the apnea-hypopnea index or the neurocognitive defects associated with sleep apnea. Thus, these agents are not currently recommended for management of OSA, but they can be used to treat comorbid depression or mood disorder.

Other Conditions Associated with Sleep-Disordered Breathing

ASTHMA

Cough and cough-induced arousals from sleep may be the presenting complaint of patients with reactive airway disease. Cough may be caused by changes in airway smooth muscle tone during NREM sleep and by bronchoconstriction during REM sleep. Gastroesophageal reflux may contribute to awakenings and bronchoconstriction. Nighttime cough occurring in a patient with uncomplicated asthma may indicate inadequate therapeutic effect of asthma medication or exacerbation of airway disease by exposure to allergens [*see 14:II Asthma*]. Morning dipping refers to the fall in lung function that occurs in the early morning hours and represents an extreme form of diurnal variation in lung function present in most patients with airway reactivity. Reports describing morning dipping emphasize that lung function measured at midday may be normal, whereas nighttime values may show moderately severe airway obstruction.

UPPER AIRWAY DISEASE

Patients with disease of the nose, larynx, and pharynx may be disturbed during sleep by apnea or aspiration of secretions caused by excessive production of mucus (e.g., chronic allergic rhinitis) or by impaired swallowing (e.g., bilateral recurrent laryngeal nerve paralysis). In both instances, frequent arousals from sleep are associated with cough or a choking sensation. During sleep, particularly REM sleep, the cough response is less than it is during wakefulness. As a result, greater amounts of secretions are tolerated before a cough ensues. After awakening, this larger amount of material may precipitate paroxysmal cough. Patients who are being treated with hypnotic medications may tolerate greater amounts of secretions before cough-induced arousal from sleep, which increases the likelihood of aspiration injury to the lungs. Hypnotic medications used to manage insomnia should be used with caution in these patients.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Patients with COPD may present with a variety of sleep problems. Nocturnal cough can be related to bronchitis. Insomnia may be the consequence of therapy with methylxanthines (e.g., aminophylline). Hypoxemia during sleep may occur as a consequence of a mechanical airway impairment present during wakefulness that is exacerbated during sleep by the normal fluctuations in gas exchange. Hypoxemia that occurs only during sleep in patients with moderately severe COPD is associated with the development of cor pulmonale. In these persons, complications attributed to hypoxemia and hypercapnia are associated neither with a severe mechanical defect (forced 1-second expiratory volume of less than 1 L) nor with symptomatic apnea. Other diagnostic entities, such as recurrent pulmonary emboli or chest wall muscle weakness, should certainly be considered.

The combination of COPD and sleep apnea is called the overlap syndrome in the European literature, and it is estimated to be present in 20% to 25% of patients presenting with moderate and moderately severe COPD. There is usually historical evidence for snoring, pauses during sleep, and arousals from sleep, and the patient should be treated for OSA [*see* Obstructive Sleep Apnea, Treatment, *above*]. Patients with SDB that is not caused by OSA may experience a good sleep response to supplemental O_2. Symptoms persisting after nocturnal O_2 may warrant a sleep study.

INTERSTITIAL LUNG DISEASE

Respiratory disturbances during sleep for patients with interstitital lung disease include cough and hypoxemia. Patients also may have a concomitant sleep apnea. Sleep hypoxemia may be a factor in the development of pulmonary hypertension. A restrictive defect on pulmonary function testing and interstitial fibrosis on chest roentgenogram may reflect a history of chronic aspiration. In patients with interstitial lung disease, during sleep, the tone of the gastroesophageal junction relaxes, allowing stomach contents to regurgitate to the level of the pharynx. In such patients, it may be useful to measure pH levels in the pharynx and esophagus during sleep.

NEUROMUSCULAR DISORDERS

Respiratory disturbances caused by OSA during sleep may occur because the underlying disease process affects upper airway muscles. Sleep disturbances associated with cough, choking, or aspiration and with SDB may be the first indications of ventilatory problems in patients with neuromuscular disease. Occasionally, sleep fragmentation and the effects of sleep deprivation dominate the clinical presentation of the patient with neuromuscular disease. After treatment for SDB, the clinical manifestations of the primary neuromuscular disorder may not appear so severe.

KYPHOSCOLIOSIS

Treatment of hypocapnic respiratory failure by tracheostomy with or without positive-pressure ventilator support during sleep can reverse cor pulmonale and improve the appearance of the chest roentgenogram.

The author has received grants for clinical research from Respironics, Inc., and the National Institutes of Health.

References

1. Richter DW, Spyer KM: Studying rhythmogenesis of breathing: comparison of in vivo and in vitro models. Trends Neurosci 24:464, 2001

2. Forster HV, Pan LG, Lowry TF, et al: Important role of carotid chemoreceptor afferents in control of breathing of adult and neonatal mammals. Respir Physiol 119:199, 2000

3. Joseph V, Pequignot JM, Van Reeth O: Neurochemical perspectives on the control of breathing during sleep. Respir Physiol Neurobiol 130:253, 2002

4. Page C, Lee LY: Summary: peripheral pharmacology of cough. Pulm Pharmacol Ther 15:217, 2002

5. Gandevia SC, Allen GM, Butler JE, et al: Human respiratory muscles: sensations, reflexes and fatiguability. Clin Exp Pharmacol Physiol 25:757, 1998

6. Horn EM, Waldrop TG: Oxygen-sensing neurons in the caudal hypothalamus and their role in cardiorespiratory control. Respir Physiol 110:219, 1997

7. Dempsey JA, Smith CA, Harms CA, et al: Sleep-induced breathing instability. Sleep 19:236, 1996

8. Iscoe S: Control of abdominal muscles. Prog Neurobiol 56:433, 1998

9. Manning HL, Mahler DA: Pathophysiology of dyspnea. Monaldi Arch Chest Dis 56:325, 2001

10. Gaultier C: Abnormalities of the chemical control of breathing: clinical correlates in infants and children. Pediatr Pulmonol Suppl 23:114, 2001

11. Spengler CM, Gozal D, Shea A: Chemoreceptive mechanisms elucidated by studies of congenital central hypoventilation syndrome. Respir Physiol 129:247, 2001

12. Webber CL Jr, Speck DF: Experimental Biot periodic breathing in cats: effects of changes in PiO_2 and $PiCO_2$. Respir Physiol 46:327, 1981

13. Sleep-related breathing disorders in adults: recommendations for symptom definition and measurement techniques in clinical trials: the report of the American Academy of Sleep Medicine. Sleep 22:667, 1999

14. Dempsey JA, Skatrud JB: Apnea following mechanical ventilation may be caused by nonchemical neuromechanical influences. Am J Respir Crit Care Med 163:1297, 2001

15. Strohl KP, Cherniack NS, Gothe B: Physiologic basis of therapy for sleep apnea. Am Rev Respir Dis 134:791, 1986

16. Rowley JA, Zahn BR, Babcock MA, et al: The effect of rapid eye movement (REM) sleep on upper airway mechanics in normal human subjects. J Physiol 510:963, 1998

17. Rowley JA, Zhou X, Vergine I, et al: Influence of gender on upper airway mechanics: upper airway resistance and Pcrit. J Appl Physiol 91:2248, 2001

18. Mezzanotte WS, Tangel DJ, White DP: Influence of sleep onset on upper-airway muscle activity in apnea patients versus normal controls. Am J Respir Crit Care Med 153:1880, 1996

19. Henke KG, Dempsey JA, Badr MS, et al: Effect of sleep-induced increases in upper airway resistance on respiratory muscle activity. J Appl Physiol 70:158, 1991

20. Rowley JA, Sanders CS, Zahn BR, et al: Effect of REM sleep on retroglossal cross-sectional area and compliance in normal subjects. J Appl Physiol 91:239, 2001

21. Guilleminault C, Quo SD: Sleep-disordered breathing. A view at the beginning of the new millennium. Dent Clin North Am 45:643, 2001

22. Dempsey JS, Skatrud JB: A sleep-induced apneic threshold and its consequences. Am Rev Respir Dis 133:1163, 1986

23. Somers VK, et al: Sympathetic-nerve activity during sleep in normal subjects. N Engl J Med 328:303, 1993

24. Launois SH, Averill N, Abraham JH, et al: Cardiovascular responses to nonrespiratory and respiratory arousals in a porcine model. J Appl Physiol 90:114, 2001

25. Phillips BG, Somers VK: Neural and humoral mechanisms mediating cardiovascular responses to obstructive sleep apnea. Respir Physiol 119:181, 2000

26. Slutsky AS, Strohl KP: Quantification of oxygen saturation during episodic hypoxemia. Am Rev Respir Dis 121:893, 1980

27. Boehlecke BA: Epidemiology and pathogenesis of sleep-disordered breathing. Curr Opin Pulm Med 6:471, 2000

28. Dalmasso F, Prota R: Snoring: analysis, measurement, clinical implications and applications. Eur Respir J 9:146, 1996

29. Peppard PE, Young T, Palta M, et al: Prospective study of the association between sleep-disordered breathing and hypertension. N Engl J Med 342:1378, 2000

30. Elmasry A, Lindberg E, Berne C, et al: Sleep-disordered breathing and glucose metabolism in hypertensive men: a population-based study. J Intern Med 249:153, 2001

31. Tishler PV, Larkin EK, Schluchter MD, et al: Incidence of sleep-disordered breathing in an urban adult population: the relative importance of risk factors in the development of sleep-disordered breathing. JAMA 289:2230, 2003

32. Kadotani H, Kadotani T, Young T, et al: Association between apolipoprotein E epsilon4 and sleep-disordered breathing in adults. JAMA 285:2888, 2001

33. Buxbaum SG, Elston RC, Tishler PV, et al: Genetics of the apnea hypopnea index in Caucasians and African Americans: I. Segregation analysis. Genet Epidemiol 22:243, 2002

34. Schwab RJ, et al: Upper airway and soft tissue anatomy in normal subjects and patients with sleep-disordered breathing: significance of the lateral pharyngeal walls. Am J Respir Crit Care Med 152:1673, 1995

35. Vizek M, Pickett CK, Weil JV: Interindividual variation in hypoxic ventilatory response: potential role of carotid body. J Appl Physiol 63:1884,1987

36. Tankersley CG: A genomic model for differential hypoxic ventilatory responses. Adv Exp Med Biol 475:75, 2000

37. Malhotra A, Fogel RB, Edwards JK, et al: Local mechanisms drive genioglossus activation in obstructive sleep apnea. Am J Respir Crit Care Med 161:1746, 2000

38. Cherniack NS: Apnea and periodic breathing during sleep. N Engl J Med 341:985, 1999

39. Badr MS, Toiber F, Skatrud JB, et al: Pharyngeal narrowing/occlusion during central sleep apnea. J Appl Physiol 78:1806, 1995

40. Ahmed M, Serrette C, Kryger MH, et al: Ventilatory instability in patients with congestive heart failure and nocturnal Cheyne-Stokes breathing. Sleep 17:527,1994

41. Khoo MC, Anholm JD, Ko SW, et al: Dynamics of periodic breathing and arousal during sleep at extreme altitude. Respir Physiol 103:33, 1996

42. Khoo MC: Determinants of ventilatory instability and variability. Respir Physiol 122:167, 2000

43. Javaheri S: Treatment of central sleep apnea in heart failure. Sleep Suppl 4:S224, 2000

44. Gottschalk A, Khoo MC, Pack AI: Multiple modes of periodic breathing during sleep. Adv Exp Med Biol 393:105, 1995

45. Bassetti C, Aldrich MS: Sleep apnea in acute cerebrovascular diseases: final report on 128 patients. Sleep 22:217, 1999

46. Parra O, Arboix A, Bechich S, et al: Time course of sleep-related breathing disorders in first-ever stroke or transient ischemic attack. Am J Respir Crit Care Med 161:375, 2000

47. Zoccali C, Mallamaci F, Tripep G: Sleep apnea in renal patients. J Am Soc Nephrol 12: 2854, 2001

48. Zoccali C: Sleep apnoea and nocturnal hypoxaemia in dialysis patients: mere risk-indicators or causal factors for cardiovascular disease? Nephrol Dial Transplant 15:1919, 2000

49. Grunstein RR, Ho KY, Berthon-Jones M, et al: Central sleep apnea is associated with increased ventilatory response to carbon dioxide and hypersecretion of growth hormone in patients with acromegaly. Am J Respir Crit Care Med 150:496, 1994

50. Grunstein RR, Ho KY, Sullivan CE: Sleep apnea in acromegaly. Ann Intern Med 115:527, 1991

51. Hochban W, Ehlenz K, Conradt R, et al: Obstructive sleep apnoea in acromegaly: the role of craniofacial changes. Eur Respir J. 14:196, 1999

52. Javaheri S, Parker TJ, Wexler L, et al: Occult sleep-disordered breathing in stable congestive heart failure. Ann Intern Med 122:487, 1995

53. Hanly PJ, Zuberi-Khokhar NS: Increased mortality associated with Cheyne-Stokes respiration in patients with congestive heart failure. Am J Respir Crit Care Med 153:272,1996

54. Lanfranchi PA, Braghiroli A, Bosimini E, et al: Prognostic value of nocturnal Cheyne-Stokes respiration in chronic heart failure. Circulation 99:1435, 1999

55. Sin DD, Logan AG, Fitzgerald FS, et al: Effects of continuous positive airway pressure on cardiovascular outcomes in heart failure patients with and without Cheyne-Stokes respiration. Circulation 102:61, 2000

56. Leung RS, Bradley TD: Sleep apnea and cardiovascular disease. Am J Respir Crit Care Med 164:2147, 2001

57. Andreas S, Schulz R, Werner GS, et al: Prevalence of obstructive sleep apnoea in patients with coronary artery disease. Coron Artery Dis 7:541, 1996

58. Sessler CN, Cohen MD: Cardiac arrhythmias during theophylline toxicity: a prospective continuous electrocardiographic study. Chest 98:672, 1990

59. Van de Graaf WB: Thoracic influence on the upper airway. J Appl Physiol 65: 2124, 1988

60. Onal E, Lopata M: Periodic breathing and the pathogenesis of occlusive sleep apneas. Am Rev Respir Dis 126:676, 1982

61. Dancey DR, Hanly PJ, Soong C, et al: Gender differences in sleep apnea: the role of neck circumference. Chest 123:1544, 2003

62. Bliwise DL, King AC, Harris RB: Habitual sleep durations and health in a 50-65 year old population. J Clin Epidemiol 47:35, 1994

63. Nieto FJ, Young TB, Lind BK, et al: Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. JAMA 283:1829, 2000

64. Vgontzas AN, Kales A: Sleep and its disorders. Annu Rev Med 50:387, 1999

65. Wilcox I, McNamara SG, Wessendorf T, et al: "Syndrome Z": the interaction of sleep apnoea, vascular risk factors and heart disease. Thorax 53:S25, 1998

66. Vgontzas AN, Bixler EO, Chrousos GP: Metabolic disturbances in obesity versus sleep apnoea: the importance of visceral obesity and insulin resistance. J Intern Med 254:32, 2003

67. Kump K, Whalen C, Tishler PV, et al: Assessment of the validity and utility of a sleep-symptom questionnaire. Am J Respir Crit Care Med 150:735, 1994

68. Flemons WW: Clinical practice: obstructive sleep apnea. N Engl J Med 347:498, 2002

69. White J, Cates C, Wright J: Continuous positive airways pressure for obstructive sleep apnoea. Cochrane Database Syst Rev (2):CD001106, 2002

70. Patel SR, White DP, Malhotra A, et al: Continuous positive airway pressure therapy for treating sleepiness in a diverse population with obstructive sleep apnea: results of a meta-analysis. Arch Intern Med 163:519, 2003

71. Kaneko Y, Floras JS, Usui K, et al: Cardiovascular effects of continuous positive airway pressure in patients with heart failure and obstructive sleep apnea. N Engl J Med 348:1233, 2003

72. Ferguson K: Oral appliance therapy for obstructive sleep apnea: finally evidence you can sink your teeth into. Am J Respir Crit Care Med 163:1294, 2001

73. Engleman HM, McDonald JP, Graham D, et al: Randomized crossover trial of two treatments for sleep apnea/hypopnea syndrome: continuous positive airway pressure and mandibular repositioning splint. Am J Respir Crit Care Med 166:855, 2002

74. Cohen SR, Holmes RE, Machado L, et al: Surgical strategies in the treatment of complex obstructive sleep apnoea in children. Paediatr Respir Rev 3:25, 2002

75. Li KK, Powell NB, Riley RW, et al: Long-term results of maxillomandibular advancement surgery. Sleep Breath 4:137, 2000

76. Kerschner JE, Lynch JB, Kleiner H, et al: Uvulopalatopharyngoplasty with tonsillectomy and adenoidectomy as a treatment for obstructive sleep apnea in neurologically impaired children. Int J Pediatr Otorhinolaryngol 62:229, 2002

77. Smith I, Lasserson T, Wright J: Drug treatments for obstructive sleep apnoea. Cochrane Database Syst Rev (2):CD003002, 2002

Acknowledgment

Figure 1 Seward Hung.

Figures 5, 7, and 8 Courtesy of M. Safwan Badr, M.D.

VII DISORDERS OF THE CHEST WALL

Gerald W. Staton, Jr., m.d.
Roland H. Ingram, Jr., m.d.

The chest wall consists of the parietal pleura, rib cage, and muscles. The abdominal contents and abdominal wall function as part of the rib cage in that they influence the resting position and movement of the diaphragm.

The respiratory pump apparatus is composed of the rib cage and its musculature, including the intercostal muscles, diaphragm, and accessory muscles of respiration.[1] Optimal pumping action requires structural integrity and the synchronized contraction of the intercostal muscles and diaphragm. Respiratory pump function may be impaired by mass loading from obesity, skeletal abnormalities, neuromuscular disorders, or restriction of lung movement from pleural disease. Cervical strap or abdominal muscles may be recruited for assistance when respiratory pump function is impaired. Because the diaphragm is the major power generator for the respiratory pump, loss of rib cage function alone may be insufficient to cause ventilatory failure; ankylosis and paralysis of the rib cage are often associated with a normal resting carbon dioxide tension (P_{CO_2}).

Respiratory Pump Dysfunction

Respiratory pump dysfunction may vary from trivial to severe. Severe dysfunction restricts lung expansion and may cause dyspnea and hypercapnic ventilatory failure because of the small tidal volumes that increase the proportion of wasted ventilation per breath, despite a compensatory increase in the rate of respiration [*see Table 1*]. The result is a decrease in alveolar ventilation with hypercapnia and hypoxemia. Hypercapnia is often most severe during sleep because of a decrease in ventilatory drive and a sleep-associated increase in upper airway resistance. Hypercapnia and the associated hypoxia may in turn cause vasoconstrictive pulmonary hypertension and cor pulmonale [*see 14:XI Pulmonary Hypertension, Cor Pulmonale, and Primary Pulmonary Vascular Diseases*]. Weakness of expiratory muscles from neuromuscular disease may produce ineffective cough and result in recurring atelectasis or infections. Severe respiratory pump dysfunction differs from alveolar and interstitial lung disease, in which ventilatory abnormalities result from alterations in the lung parenchyma [*see Table 2*].

When the respiratory pump is impaired, ventilation will be determined by (1) the efficiency of the inspiratory muscles, (2) the strength of the inspiratory muscles, and (3) the impedance to the pumping action of these muscles. The efficiency of the respiratory muscles is determined by their length and by the resulting mechanical action on the pump apparatus. Shortening of the inspiratory muscles from hyperinflation or chest wall deformity reduces their pumping efficiency. Paralysis of either the chest wall or the diaphragm produces an observable paradoxical movement of the paralyzed component during inspiration; this movement results in inefficiency of the pump apparatus. The strength of the inspiratory muscles may be reduced by neuromuscular disease or metabolic disturbances, such as hypokalemia or hypophosphatemia. Ventilatory ability is proportional to the remaining respiratory muscle strength but may be disproportionately reduced if there are concomitant mechanical problems of the respiratory system. Finally, either increased airway resistance or decreased respiratory system compliance may impede the pumping action of the inspiratory muscles and reduce ventilation. Respiratory system compliance may be reduced by morbid obesity, chest wall deformity, circumferential pleural disease, or parenchymal disease. Patients with a poorly compliant respiratory system must exert more effort than healthy patients to achieve equivalent tidal volumes [*see Figure 1*], so they take smaller breaths to minimize respiratory muscle fatigue but must compensate by increasing their breathing rate.

Obesity and Its Impact on Respiratory Function

Obesity imposes a restrictive load on the thoracic cage, both directly because weight has been added to the rib cage and indirectly because of the large abdominal panniculus, which impedes the motion of the diaphragm when the person is supine.[2,3] In addition, obese patients, particularly males, may experience increased respiratory resistance and resultant airflow limitation that may be related to breathing at lower lung volumes, increases in pulmonary blood volume, or both.[4,5]

Obesity characteristically causes a decrease in functional residual capacity that becomes significant only in the presence of coexisting conditions such as obstructive lung disease, in which airway closure occurs at lower lung volume, leading to hypoxemia.

Obesity in otherwise healthy patients causes little interference with lung function at rest. Generally, vital capacity and total

Table 1 Relations among Tidal Volume, Respiratory Frequency, and Arterial and Alveolar Oxygen Tension

	Volume (L)			V_D/V_T	f	$\dot{V}_A$ (L/min)	P_{aCO_2} (mm Hg)	P_{aO_2} (mm Hg)	$A\text{-}aD_{O_2}$ (mm Hg)
	V_T	V_A	V_D						
Normal respiratory function	0.50	0.35	0.15	30%	12/min	4.20	40	90	10
Restrictive disorder	0.25	0.10	0.15	60%	30/min	3.00	56	70	10*

Note: this table demonstrates how restrictive disorders of the respiratory system resulting from neuromuscular or chest wall disease may produce hypercapnia and hypoxemia by decreasing the amount of ventilation per breath.

*The elevated P_aCO_2 has resulted in arterial and alveolar hypoxia. However, in the absence of atelectasis or another concomitant disease that would increase the ventilation-perfusion mismatch, the $A\text{-}aDo_2$ remains normal.

$A\text{-}aDo_2$—alveolar-arterial difference in oxygen f—respiratory frequency P_aCO_2—arterial carbon dioxide tension P_aO_2—arterial oxygen tension V_A—alveolar portion of the tidal volume $\dot{V}_A$—alveolar ventilation ($f \times V_A$) V_D—dead space portion of the tidal volume V_D/V_T—ratio of functional dead space volume to tidal volume, or the wasted ventilation ratio V_T—tidal volume

Table 2 Respiratory Parameters in Kyphoscoliosis, Neuromuscular Syndrome, and Diffuse Parenchymal Lung Disease

	Kyphoscoliosis	*Neuromuscular Syndrome*	*Diffuse Parenchymal Lung Disease*
Example	—	Amyotrophic lateral sclerosis	Idiopathic pulmonary fibrosis
Chest radiograph	Vertebral deformity	Elevated diaphragms; possible basilar atelectasis	Diffuse reticulonodular infiltrates
Lung volumes:			
Total lung capacity	↓	↓	↓
Functional residual capacity	↓	Normal	↓
Residual volume	↓	↑	↓
Diffusing capacity of the lung for carbon monoxide corrected for alveolar volume (DL/VA)	Normal (except in severe disease, scoliotic angle > 100°)	Normal	↓
Gas exchange:			
Alveolar-arterial difference in oxygen ($A\text{-}aDo_2$)	Normal or slightly ↑	↑ With atelectasis	↑
Arterial carbon dioxide tension (P_aco_2)	↑ As spinal deformity worsens	↑ As inspiratory muscle strength decreases	↓ (↑ Only in late stage as V_D/V_T increases)
Cough	Normal	↓	Normal
Maximal inspiratory pressure	Normal or ↓	↓	Normal

lung capacity remain normal except in the most severe instances of morbid obesity.

In patients with impaired ventilatory drive, the mechanical work load imposed by obesity may not be countered by increased respiratory effort. Under such circumstances, chronic daytime hypercapnia may develop, most commonly in the setting of obstructive sleep apnea but also in the absence of sleep-disordered breathing. The pathogenetic role of obesity in obstructive sleep apnea [*see 14:VI Ventilatory Control during Wakefulness and Sleep and 11:XIII Disorders of Sleep*] may relate in part to fatty encroachment on the upper airways.

Obese patients may experience significant dyspnea during exercise because of the increased work required to move the heavy chest and abdomen and because of overall poor conditioning. The tachypneic shallow breathing pattern during exercise in morbidly obese patients reflects the combined effects of this mass loading and diminished compliance of the respiratory system [*see Figure 1*].

Weight loss is the most important therapy for patients with respiratory problems related to obesity. For patients with associated sleep disorders of breathing, appropriate treatment of obesity may alleviate hypercapnia [*see 14:VI Ventilatory Control during Wakefulness and Sleep and 11:XIII Disorders of Sleep*]. Noninvasive ventilation may help decrease the symptoms of daytime hypercapnia.[6]

Skeletal Abnormalities that Affect Respiratory Function

Deformities of the costovertebral skeletal structures may affect compliance of the thoracic cage, its shape and volume, and, ultimately, pulmonary compliance. Effects vary from undetectable to severe.

KYPHOSCOLIOSIS

The two basic types of costovertebral skeletal deformity—scoliosis, a lateral curvature with rotation of the vertebral column, and kyphosis, an anterior flexion of the spine—are usually found in combination. Approximately 80% of cases of kyphoscoliosis are idiopathic. Idiopathic kyphoscoliosis commonly begins in late childhood or early adolescence and may progress in severity during these years of rapid skeletal growth. The incidence of kyphoscoliosis in females is four times higher than that in males. The remaining 20% of cases of kyphoscoliosis are found in association with neuromuscular disorders (e.g., syringomyelia, neurofibromatosis, or poliomyelitis), congenital vertebral defects (e.g., hemivertebrae), acquired vertebral abnormalities (e.g., tuberculous spondylitis or osteomalacia), or deforming chest wall processes (e.g., sequelae of empyema or as a result of thoracoplasty).

Respiratory Compromise

Of the various chest deformities that produce ventilatory failure, kyphoscoliosis is the most common. The degree to which ventilation is reduced is determined by the severity of deformity

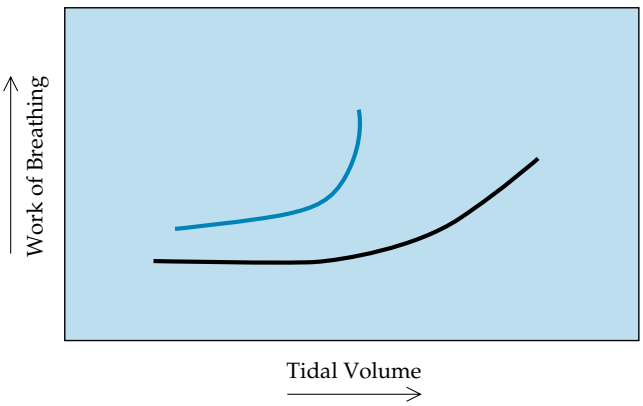


Figure 1 **When the respiratory pump is encumbered by structural changes such as obesity, kyphoscoliosis, or ankylosing spondylitis (blue line), the work of breathing at rest is higher than it is for a healthy person (black line). The work of breathing required to maintain the tidal volumes needed during exercise may be prohibitive for patients with these disorders, forcing them to adopt a shallow tachypneic breathing pattern in response to an increased ventilatory requirement.**

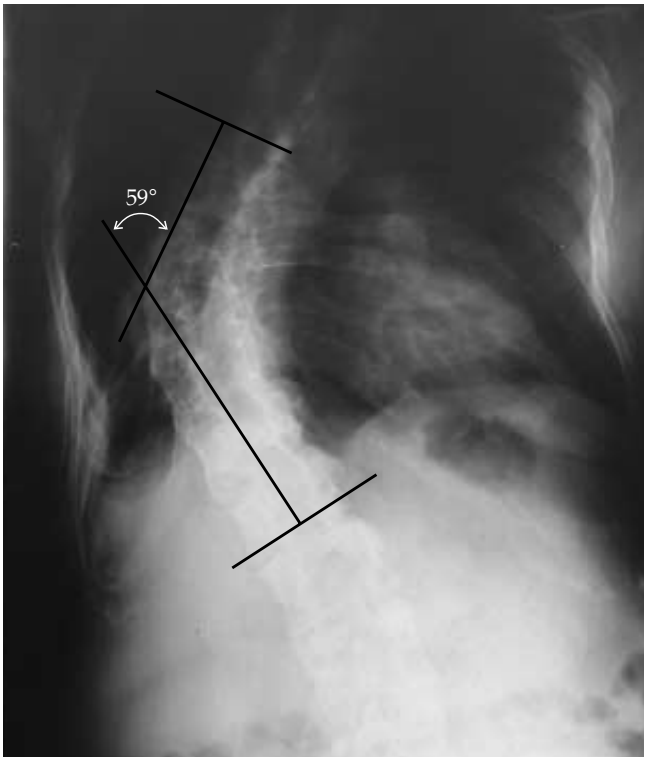


Figure 2 **In this radiograph of the spine in a patient with kyphoscoliosis, straight lines are passed through the upper and lower limbs of the curvature. The angle inscribed by these two lines defines the scoliotic angle.**

and the degree of neuromuscular weakness. A standard technique for measuring the degree of spinal curvature gives an indication of the potential risk for respiratory compromise [*see Figure 2*]. Mild to moderate deformities (scoliotic angle < 60°) are associated with minimal to mild restrictive ventilatory defects.

Dyspnea may occur during exercise and is most often caused by deconditioning and lack of regular aerobic exercise rather than by any alteration of lung function.[7] As the scoliotic curvature worsens, vital capacity and total lung capacity decline significantly, and dyspnea on mild or moderate exertion becomes a common complaint. Kyphoscoliosis may distort the respiratory pump, so that inspiratory power becomes limited even in the absence of a neuromuscular disease, such as poliomyelitis. The severity of hypercapnia is therefore related to both the severity of deformity and the degree of inspiratory muscle weakness.

Severe deformities (scoliotic angle > 100°) can be associated with prominently restricted lung volumes; typically, total lung capacity is reduced to 50% or less of the predicted value. Such restriction may lead to chronic alveolar hypoventilation, hypoxemia, pulmonary hypertension, and, ultimately, cor pulmonale.[8] Long-term follow-up of patients with kyphoscoliosis suggests that those with a vital capacity less than 45% of the predicted value and a scoliotic angle greater than 110° are at the greatest risk for respiratory failure.[9] Kyphoscoliosis of such severity may cause compression of underlying lung tissue, thereby elevating the alveolar-arterial difference in oxygen ($A\text{-}aDO_2$). In most cases of kyphoscoliosis, however, the $A\text{-}aDO_2$ remains at normal or near-normal levels, and significant hypoxemia is present only when hypercapnia develops. These findings contrast with restrictive ventilatory disorders caused by diffuse parenchymal lung disease, in which the $A\text{-}aDO_2$ is characteristically elevated and hypoxemia is often associated with hypocapnia until late in the course of disease.

Treatment

Acute complications Patients with severe kyphoscoliosis may live for many years without succumbing to respiratory insufficiency. Such patients, however, are vulnerable to any respiratory tract infection or central nervous system depressant. Because breathing is chronically restricted, increased central neural output and physical work are required to maintain ventilation; relatively minor insults, such as bacterial or viral bronchitis or pneumonia, may represent an increment in load sufficient to produce frank respiratory failure. In addition, standard doses of narcotics or sedatives may suppress chronically hyperactive control mechanisms to a level sufficient to precipitate acute respiratory failure.

Thus, immunization with influenza and pneumococcal vaccines, early treatment of respiratory tract infections, and strict avoidance of CNS depressants are important in the management of kyphoscoliosis. Episodes of hypercapnic respiratory failure precipitated by reversible conditions respond well to short-term supportive measures, including bronchopulmonary drainage, mechanical ventilatory support, and oxygen supplementation.

Chronic complications Chronic respiratory insufficiency may ensue after several years. Older patients with kyphoscoliosis are at risk for respiratory failure because the angle of curvature typically worsens with age. Although most cases of idiopathic scoliosis stabilize just after puberty, further spinal deformity may result from the osteoporosis, vertebral body weakening, and loss of muscle tone that accompany older age. Surgical procedures to straighten and stabilize the vertebral column usually fail to restore ventilatory capacity. Such procedures are useful early in the course of kyphoscoliosis, when they may prevent progression of the deformity before respiratory compromise develops.[10]

Supportive measures may sustain meaningful life for many years, even in patients with chronic respiratory failure. Many can adapt very well to a state of chronically disordered gas exchange. Chronic hypoxemia accompanied by secondary erythrocytosis, worsening of pulmonary hypertension, and cor pulmonale should be treated with supplemental oxygen administration. Such therapy can be augmented by nocturnal ventilatory support.[11]

Although kyphoscoliosis is not a primary sleep-related breathing disorder, the degree of oxygen desaturation that occurs during sleep is greater than that observed in other types of lung disease, possibly because the baseline hypoxemia and hypercapnia are more severe and lung volumes are smaller.[12] Thus, nocturnal oxygen therapy and mechanical ventilatory support during sleep often improve functional status and symptoms. In fact, many patients who achieve normal levels of arterial carbon dioxide tension (P_aCO_2) during sleep by means of mechanical ventilators can sustain normal or near-normal arterial blood gas levels throughout the day. Devices used for mechanical ventilation in such patients include the iron lung, specially fitted thoracoabdominal negative pressure ventilators, and positive pressure ventilators applied via tracheostomy or nasal mask.[13]

ANKYLOSING SPONDYLITIS

Ankylosing spondylitis may affect the thoracic cage because of arthritic involvement of the costovertebral articulations[14] [*see*

15:III Seronegative Spondyloarthropathies]. The chest may become relatively fixed in a hyperexpanded position, leading to an elevated midposition lung volume. The reduced compliance of the chest wall causes moderate restriction of vital capacity and total lung capacity. A typical physical finding is limited expansion of the chest wall on inspiration, despite normal findings on auscultation and percussion of the chest and normal muscle strength. Nonetheless, the alteration in function is almost never severe enough to produce symptoms, and this deformity does not produce respiratory failure.

DEFORMITIES OF THE STERNUM

Deformities of the sternum and costochondral articulations are potentially dramatic in radiographic and physical appearance and may induce psychological problems, but functional consequences are rare. There are two main varieties of deformity: pectus excavatum, an inward concavity of the lower sternum, and pectus carinatum, an outward protuberance of the upper, middle, or lower sternum.

Respiratory Compromise

Pectus deformity is present in fewer than 0.5% of the general population and appears to be more common in patients with other evidence of structural or connective tissue disease, such as scoliosis, Marfan syndrome, Poland syndrome, or Pierre Robin syndrome. In these circumstances, ventilatory function impairments may result from the underlying disease rather than the pectus deformity. For example, defects in bronchial cartilage development may lead to repeated pneumonia and result in bronchiectasis.

Treatment

In most cases of pectus deformity, no significant functional limitations are caused by the deformity. Lung volumes are preserved, and cardiovascular function is normal. Surgical correction is therefore generally restricted to patients who have severe deformity accompanied by evidence of lung restriction or cardiovascular dysfunction. Although the severity of the pectus deformity may be assessed by determining the ratio of the transverse diameter of the chest to the anteroposterior diameter as measured by computed tomography, it is not clear whether this index predicts improvement in lung function with surgery.[15] A few patients with cardiac compression or with lung restriction from pectus excavatum experience functional improvement after surgical repair. Right and left ventricular end-diastolic volumes also may improve as cardiac compression is relieved.[16]

Surgical correction may result in modest improvements in lung volumes, ventilatory capacity, and exercise capacity in patients with severe pectus deformities but may worsen the condition of patients with good preoperative lung function.[17,18] Surgical correction for only cosmetic reasons may be associated with worsening of lung function. New surgical techniques are being developed that may improve outcome.[19]

FLAIL CHEST

Flail chest is an acute process that may lead to life-threatening abnormalities of gas exchange and mechanical function. Stability of the thoracic cage is necessary for the muscles of inspiration to inflate the lung. In flail chest, a locally compliant portion of the chest wall moves inward as the remainder of the thoracic cage expands during inhalation; the same portion then moves outward during exhalation. Consequently, tidal volume is diminished because the region of lung associated with the chest wall abnormality paradoxically increases its volume during exhalation and deflates during inhalation. The result is progressive hypoxemia and hypercapnia. Multiple rib fractures, particularly when they occur in a parallel vertical orientation, can produce a flail chest. The degree of dysfunction is directly proportional to the volume of lung involved in paradoxical motion. Patient management may be complicated by other manifestations of trauma to the chest, such as splinting of ventilation because of pain, contusion of underlying lung, or hemothorax or pneumothorax. Positive pressure inflation of the lung or negative pressure applied to the chest wall corrects the abnormality until more definitive stabilization procedures can be undertaken.

The pathophysiologic disturbances of flail chest may also result from nonclosure of the wound after median sternotomy is performed. Any dehiscence of the sternal wound will lead to separation, loss of stability, and prominent inward motion during inspiration. The magnitude of the inward motion is directly related to the extent of the sternal separation and to the degree of negativity of inspiratory intrathoracic pressures. This condition is often the cause of difficulty in weaning a patient from mechanical ventilatory support after major cardiac surgery.

Other, rare causes of localized chest wall instability include destruction of the ribs from malignant disease (e.g., multiple myeloma) or from metabolic disorders (e.g., osteitis fibrosa cystica).

Neuromuscular Disorders That Affect Respiratory Function

Processes that interfere with the transfer of central neural output to the muscles that expand the rib cage, such as abnormalities of the spinal cord, peripheral nerves, neuromuscular junctions, or muscles, can lead to ventilatory impairment. Whereas central control problems allow creation of adequate inspiratory pressures by voluntary efforts [*see 14:VI Ventilatory Control during Wakefulness and Sleep*], central neural output abnormalities are characterized by an inability to generate normal respiratory pressures, either automatically or intentionally. Some diseases, such as poliomyelitis, can involve both the central controller and the peripheral neuromuscular apparatus.

OVERVIEW OF NEUROMUSCULAR DISORDERS AFFECTING RESPIRATION

Pathophysiology

Several factors are common to the neuromuscular disorders of the thoracic cage. The respiratory midposition volume is maintained at near-normal levels, whereas total lung capacity decreases (because of inspiratory muscle weakness) and residual volume increases (because of expiratory muscle weakness). Vital capacity is diminished along with maximal static inspiratory pressure.

Because muscle strength and vital capacity can be substantially diminished without causing respiratory failure, the presence of respiratory failure with hypoxemia and hypercapnia indicates either extreme progression of the primary process or the effects of complications such as atelectasis caused by retained secretions from ineffective cough, pneumonia, or pulmonary thromboembolism. Onset of hypoxemia, hypercapnia, or both in the presence of reasonable inspiratory muscle function suggests the presence of a complication rather than progression of the primary process. In the acute setting, the need

to distinguish between a complication and progression makes monitoring of maximal static inspiratory pressures (which assesses muscle strength) superior to measuring serial vital capacity, because vital capacity may be diminished by either a complication or progression.

Neuromuscular disorders that persist for months are associated with chronic decreases in compliance of both the chest wall and the lungs. It is unknown whether the decreases in lung compliance are the result of microatelectasis, altered surfactant, or mild fibrotic changes resulting from recurrent infections. Ventilation-perfusion mismatch occurs in the lungs of patients with these disorders and may lead to hypoxemia that is disproportionate to the degree of hypoventilation. Decreases in chest cage compliance have been attributed to gradual stiffening of the costochondral and costovertebral articulations and to fibrotic changes or spasticity of the muscles of the rib cage.

The diminution in lung volume that occurs in chronic neuromuscular disorders is caused by the combined effects of muscle weakness and secondary alterations in the mechanical properties of the lung and chest wall. Hence, for patients with chronic disease, measurement of vital capacity is a more accurate indicator of the total impact of the disorder than is maximal inspiratory pressure. Attempts to improve lung compliance by periodic hyperinflations with intermittent positive pressure breathing have usually not proved successful.

In contrast to the mechanical disorders of the thoracic cage, which preserve an effective cough, expiratory muscle weakness in the neuromuscular disorders prevents generation of sufficient expiratory velocities for a forceful cough. The extreme example is cervical spinal cord injuries in which paralysis of the abdominal and intercostal muscles severely reduces but does not eliminate spontaneous cough. Ineffective or absent cough eliminates a first-line defense against respiratory tract infection and is particularly troublesome when combined with airway mucus hypersecretion, as occurs in asthma or chronic bronchitis. Pneumonia followed by respiratory failure is a common cause of death in patients with neuromuscular syndromes.

Respiratory Compromise

Patients with neuromuscular disorders must be awake to maintain ventilation. During sleep, hypoxemia and hypercapnia develop or worsen and may contribute to complications such as cor pulmonale. The degree of hypoxemia that develops with sleep is related to the severity of the abnormalities in lung mechanics and to the degree of derangement in gas exchange that is present while the patient is awake.[20]

In the absence of major complications, the patient with neuromuscular involvement is often disproportionately tachypneic in relation to the decrease in tidal volume. The resulting increase in minute ventilation more than offsets the increase in dead space ventilation. Thus, early in the course of the illness, P_aCO_2 is often low. The basis for the tachypnea may be microatelectasis, which also accounts for mild arterial hypoxemia. Microatelectasis probably develops because of the patient's inability to take intermittent deep breaths or sighs, which results in changes in alveolar surface forces. As weakness progresses, tidal volume decreases, dead space ventilation increases, and alveolar hypoventilation with worsening hypoxemia ensues [*see Table 1*]. The decision whether to treat with mechanical ventilatory support must then be made. Long-term results depend on the nature and prognosis of the neuromuscular process and on the potential success of a specific therapy.

Treatment

Whether the primary disorder is acute (e.g., Guillain-Barré syndrome), intermittent (e.g., myasthenia gravis), progressive (e.g., amyotrophic lateral sclerosis), or chronic (e.g., quadriplegia), onset of a pulmonary complication and the accompanying increase in mechanical load and decrease in gas-exchanging ability may precipitate overt and life-threatening respiratory failure [*see Table 3*]. Because the patient is unable to produce an effective cough, even minor causes of increased airway secretions, such as a viral tracheobronchitis, may lead to major respiratory compromise. Maintenance of bronchopulmonary drainage and the early treatment of infections are essential for avoidance of complications. Acute episodes precipitated by such complications usually respond well to specific treatment plus supportive measures, including bronchopulmonary drainage and mechanical ventilation.

DIAPHRAGMATIC PARALYSIS

In the absence of respiratory complications, neuromuscular syndromes rarely progress to the point of hypercapnic respiratory failure unless diaphragmatic weakness or paralysis is present. Thus, quadriplegic patients who have a preserved phrenic nerve and diaphragmatic function (e.g., C7 spinal cord transection) almost never progress to hypercapnic respiratory failure unless a major pulmonary complication supervenes or CNS-depressant drugs are administered. Because diaphragmatic paralysis or paresis uniformly accompanies hypercapnic respiratory failure caused by any of the neuromuscular syndromes, it is not usually considered apart from these disorders. However, because certain forms of diaphragmatic paralysis have distinguishing clinical features, they are best considered as discrete entities.

Bilateral Diaphragmatic Paralysis

Respiratory compromise Bilateral phrenic nerve interruption or injury may result in an isolated partial or complete diaphragmatic paralysis. Causes include cervical and thoracic surgery, cold cardioplegia for cardiac surgery,[21] trauma, multiple sclerosis, and neuralgic amyotrophy.[22] Orthopnea may be a prominent symptom. With the patient supine, the hydrostatic force of the abdominal contents pushes the patient's diaphragm into the thorax. Negative intrapleural pressures generated by the accessory muscles cause the diaphragm to be sucked further into

Table 3 Neuromuscular Syndromes Associated with Respiratory Failure

Site of Lesion	*Disorder*
Spinal cord	Quadriplegia Amyotrophic lateral sclerosis Poliomyelitis Spinal muscular atrophies
Peripheral nerves	Guillain-Barré syndrome Diphtheritic neuropathy
Neuromuscular junctions	Myasthenia gravis Eaton-Lambert syndrome Botulism Drug-induced weakness
Muscles	Muscular dystrophies (e.g., Duchenne dystrophy, myotonic dystrophy)

the thorax during inspiration, producing a paradoxical inward motion of the upper abdomen as the thorax expands [*see Figure 3*]. As a result, mechanical and gas exchange abnormalities similar to those seen in flail chest develop. In the upright position, patients often experience a dramatic increase in vital capacity, improvement in gas exchange, and alleviation of symptoms because the weight of the abdominal contents offsets the negative intrapleural pressures and, therefore, the diaphragm no longer ascends with inspiration. Despite improvements in lung function when in the upright position, patients with bilateral diaphragm paralysis experience a significant reduction in exercise capacity.[23]

Treatment Sleeping in an upright position or nocturnal use of a thoracoabdominal cuirass negative-pressure device can facilitate breathing during sleep. In a limited study, bilateral plication of the diaphragm resulted in improved lung function and allowed patients to sleep in a supine position.[24] The bilateral diaphragmatic pacemaker, although expensive and time consuming to place and stabilize, is an alternative for selected patients.

Unilateral Diaphragmatic Paralysis

Unilateral diaphragmatic paralysis is most often detected as an asymptomatic radiographic finding. The paralyzed hemidiaphragm maintains its normal contour but is displaced cephalad. On fluoroscopy, the paralyzed hemidiaphragm may descend slightly on normal inspiration, mimicking normal contraction. However, with a sudden forceful inspiration (so-called sniff test), the paralyzed portion of the diaphragm ascends further into the thorax, opposite to the direction of the normally functioning side. It is driven cephalad by the sudden increase in intra-abdominal pressure and the sudden fall in intrathoracic pressure occasioned by the sniff. Thus, a fluoroscopic sniff test helps confirm the diagnosis of unilateral diaphragmatic paralysis.

Most cases of unilateral diaphragmatic paralysis are the result of neoplastic invasion of the phrenic nerve. Compression or destruction of the phrenic nerve by surgery, trauma, or enlarging lymph nodes or aneurysmal vessels may also cause the condition. Idiopathic cases, which may stem from an isolated phrenic neuropathy or acute infectious neuritis, appear to be evenly divided between the right and left sides and are usually permanent. Reversible paralysis is a rare complication of acute pneumonia and more commonly follows cardiac surgery in which the phrenic nerve is transiently injured by the ice slurry used to achieve cardioplegia. A thorough history and a CT scan of the thorax usually suffice for the workup of patients with unilateral diaphragmatic paralysis.

Respiratory compromise Vital capacity and total lung capacity are reduced 15% to 20% below normal levels with the patient upright and are reduced significantly more with the patient supine. A modest degree of reduction in exercise capacity is also seen.[23] Nonetheless, in the absence of associated pleuropulmonary disease, most adult patients with unilateral diaphragmatic paralysis but without a coexisting pulmonary disease remain asymptomatic.

Treatment Infants are more dependent than adults on bilateral diaphragm function for adequate respiratory pump function because their more deformable chest wall moves inward with inspiration along with the paralyzed hemidiaphragm. Plication of the diaphragm may be necessary to prevent flail motion in infants with ventilatory failure associated with unilateral diaphragm paralysis. Plication may also improve the lung function of symptomatic adults with unilateral diaphragmatic paralysis.[25]

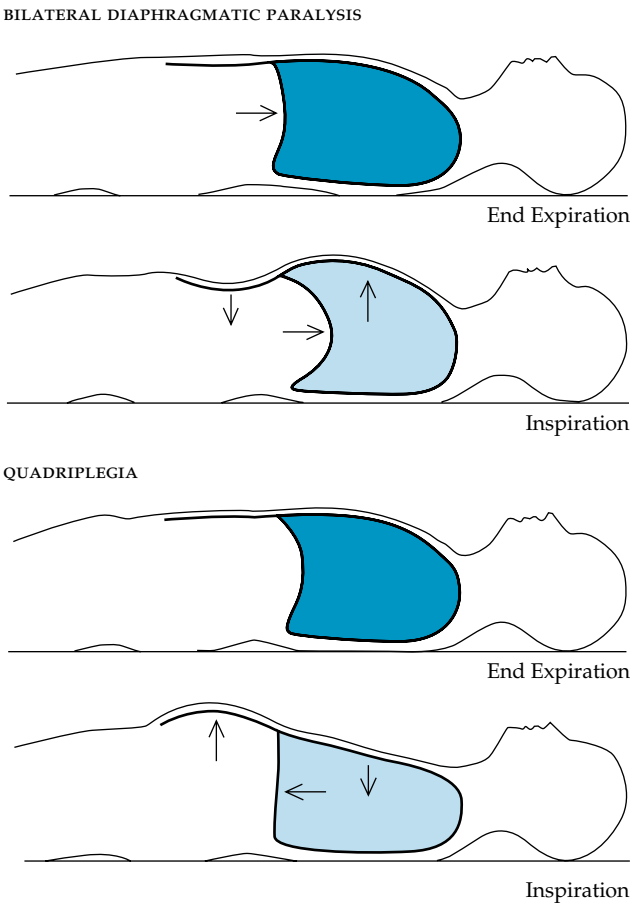


Figure 3 **In a supine patient with bilateral diaphragmatic paralysis (top), at end expiration the weight of the abdominal contents forces the paralyzed diaphragm cephalad. With inspiration, only the accessory muscles contract, elevating the anterior chest wall, creating a negative intrathoracic pressure, and forcing the diaphragm further cephalad. The result is a paradoxical inward motion of the abdominal wall with inspiration. The diaphragm of a supine patient with quadriplegia caused by a spinal cord injury at or below the level of C4 (bottom) contracts and descends normally on inspiration, causing the anterior abdominal wall to protrude. However, because the paralyzed intercostal muscles fail to stabilize the rib cage, the anterior chest wall is pulled paradoxically inward during inspiration by the negative intrathoracic pressure generated by the diaphragm.**

Postoperative Diaphragmatic Dysfunction

A transient form of diaphragmatic dysfunction occurs after upper abdominal surgery. In the hours after surgery, lung volumes and maximal inspiratory pressures fall and A-aD_{O_2} increases. At the same time, tidal volume decreases and respiratory frequency increases. These changes may be caused by irritation of the diaphragm that produces reflex inhibition of phrenic nerve function. As a result of postoperative diaphragmatic dysfunction, atelectasis and hypoxemia occur. Deep breathing that focuses on inspiratory effort, possibly facilitated by incentive spirometry or by direct encouragement, may alleviate these abnormalities.

SPINAL CORD SYNDROMES

Functional Transection

Functional transection, most often caused by trauma from motor vehicle or diving accidents, must be at or below the level of the C4 cervical nerve segment if the patient is to survive. If transection occurs above this level, the diaphragm ceases to function and breathing stops. For the quadriplegic patient, breathing is maintained solely or predominantly by the diaphragm.

Respiratory compromise The degree of compromise of lung function is directly related to the level of injury.[26] Patients with mid- to low-cervical injury are completely dependent on diaphragmatic function, producing problems with several aspects of breathing. Because the diaphragm is active only during inspiration, cough—which requires activity by expiratory muscles, including those of the abdominal wall—is almost totally absent. Intercostal muscles are required to stabilize the upper rib cage against inward collapse when negative intrathoracic pressures are produced by descent of the diaphragm. Thus, with diaphragmatic breathing alone, there is a paradoxical inward motion of the upper thorax during inspiration [*see Figure 3*]. The result is a diminished tidal volume for any level of diaphragmatic activation. Any sparing of the lower cervical segments allows some activity by accessory muscles and diminishes paradoxical motion. Paradoxical motion also diminishes as the condition becomes chronic because of stiffening of the thorax. In patients with injuries of the thoracic spine, inspiratory function is preserved, but expiratory function proves difficult because of abdominal muscle weakness.

Treatment When the quadriplegic patient is upright, the weight of the abdominal contents pulls on the diaphragm, and because abdominal muscle tone has been lost and cannot provide restraining action, diaphragmatic shortening ensues. Thus, the diaphragm is less effective with the patient in this position, and platypnea may result. Abdominal binders serve to replace lost abdominal muscle tone and should be used whenever tidal volume falls with the patient upright. An inflatable anterior air bladder in the binder may be used to assist ventilation in patients with marginal respiratory function. External abdominal compression may be used to help these patients cough. Quadriplegic patients have a mild degree of bronchial hyperresponsiveness caused by parasympathetic tone from the uninjured vagus nerve that is unopposed by sympathetic tone from the spinal cord.[27] Use of an anticholinergic bronchodilating agent or a $beta_2$-adrenergic agonist will reverse this abnormality.[28] Inspiratory muscle training with a resistive device can improve lung function and reduce dyspnea in these patients.[29]

Respiratory failure almost never occurs in quadriplegic patients in the absence of a complication such as pneumonia, atelectasis, septicemia, or pulmonary thromboemboli; therefore, mechanical ventilation is required only during such complications.

Anterior Horn Cell Disease

Amyotrophic lateral sclerosis Anterior horn cell disease is most commonly seen in patients with amyotrophic lateral sclerosis[30] [*see 11:IV Cerebrovascular Disorders*]. Although the disease causes weakness of the muscles of breathing and an associated restrictive abnormality in most patients, only a minority of patients have respiratory symptoms at presentation. Respiratory failure may not be seen until an episode of bronchitis or aspiration pneumonia produces an acute event.

Supportive measures administered during the acute event often return the patient to a stable condition. In most patients, however, respiratory failure caused by diaphragmatic involvement proves fatal within 3 to 4 years of the onset of symptoms.

Poliomyelitis Sporadic cases of poliomyelitis still occur in the United States. When anterior horn cells involving innervation of the diaphragm are affected, the patient may require ventilatory support, either continuously or during acute pulmonary events.

Spinal muscular atrophy Spinal muscular atrophy represents a heterogeneous collection of heredofamilial disorders that primarily involve spinal motor neurons. Patients with the slowly progressive infantile form, Werdnig-Hoffmann disease, may survive until late childhood. With the onset of a rapid growth phase during puberty, kyphoscoliosis develops and complicates the muscle weakness associated with these disorders.

DISORDERS OF THE PERIPHERAL NERVES

Guillain-Barré Syndrome

Respiratory insufficiency that requires artificial ventilation develops in 20% to 25% of patients with Guillain-Barré syndrome[25] [*see 11:II Diseases of the Peripheral Nervous System*]. Duration of ventilatory support averages 2 months, but periods of up to 30 months have been reported.[31,32] Approximately 30% of patients, however, can be extubated within 2 weeks. Mortality is less than 5%; the majority of survivors recover completely. A small minority may have persistent weakness and may be susceptible to recurring episodes of respiratory failure in association with respiratory infections. The chronic form of demyelinating polyneuropathy may account for some of these persistent abnormalities [*see 11:II Diseases of the Peripheral Nervous System*].

Toxin-Induced Neural Dysfunction

Peripheral nerve function may also be impaired by toxins from algae or plankton. Certain fish (grouper and snapper) or shellfish consume and concentrate toxins that may in turn be consumed by humans. Although many of the shellfish-related toxins have been associated with respiratory failure, mortality has been highest in connection with the ingestion of puffer fish, a delicacy in Japan.

Diphtheria

Cardiac and neural toxicity occurs in approximately 20% of cases of diphtheria [*see 7:IV Infections Due to Gram-Positive Bacilli*]. Clinical features and management of the neurologic syndrome are similar to those of Guillain-Barré syndrome.

DISORDERS OF NEUROMUSCULAR TRANSMISSION

Myasthenia Gravis

Myasthenia gravis is the most common disorder affecting neuromuscular transmission. Respiratory failure may occur during myasthenic crisis (acute episode of the basic disease process) or cholinergic crisis (increased weakness caused by an excess of anticholinergic medication), or it may occasionally follow initiation of glucocorticoid therapy. In patients with long-standing myasthenia, myopathy with severe diaphragmatic paresis may

develop and lead to chronic respiratory failure. The principles of respiratory monitoring and ventilatory support for myasthenia are similar to those for the other neuromuscular syndromes.[33]

Myasthenic Syndrome

The myasthenic (Eaton-Lambert) syndrome may be confused with myasthenia gravis. As with other paraneoplastic syndromes, there may be coexistent cerebellar ataxia or carcinomatous neuropathy. The prognosis is not good, and respiratory failure may occur as a terminal event.

Drug-related Muscle Paralysis

Paralysis lasting several hours can occur after a single dose of succinylcholine in patients with reduced functional pseudocholinesterase, as seen in those with severe liver disease, myxedema, or malnutrition; pregnant patients; and patients with a genetic deficiency. A syndrome of prolonged muscle paralysis or weakness can develop after the extended administration of postsynaptic paralytic agents to facilitate mechanical ventilation, especially in patients with renal failure.[34]

A number of other commonly used drugs can interfere with neuromuscular transmission. The mechanism can involve anesthetic-like action at the presynaptic level (e.g., with clindamycin or propranolol), postsynaptic curarelike action (e.g., with lincomycin, polymyxin B sulfate, chloroquine, or procainamide), or stabilization of postsynaptic membranes (e.g., with gentamicin, streptomycin, or neomycin).[35] Delayed recovery from anesthesia and difficulty in withdrawing a patient from ventilatory support should suggest a possible drug effect. Patients affected by these drugs often have mild or latent myasthenia; occasionally, concurrent electrolyte disturbances, such as hypokalemia, hypocalcemia, or hypomagnesemia, combine with drug effects to produce contributory muscle weakness in myasthenic patients. In severe cases, especially those involving antibiotic-induced postoperative muscle weakness, respiratory failure can ensue; ventilatory support, correction of associated electrolyte disturbances, and withdrawal of the drug usually lead to recovery.

Botulism is an uncommon condition caused by ingestion of a polypeptide toxin produced by *Clostridium botulinum* [*see 7:V Anaerobic Infections*]. In addition to specific therapy, careful monitoring and provision of ventilatory support in the event of respiratory failure are required.

DISORDERS OF THE MUSCLES

A number of muscle disorders can produce respiratory abnormalities. Patients with mitochondrial myopathy can present with unexplained dyspnea.[36] A specific abnormality of the diaphragm thought to be myopathic can occur in patients with systemic lupus erythematosus, producing the so-called shrinking-lungs syndrome.[37] Patients with pseudohypertrophic (Duchenne) dystrophy, myotonic dystrophy, and other forms of muscular dystrophy have a more severe disorder and are predisposed to pulmonary complications, and respiratory failure is a frequent cause of death.[38,39] Chronic alveolar hypoventilation caused by inspiratory muscle weakness may develop late in the course of a disease. Expiratory muscle weakness impairs cough in some patients, and accompanying weakness of the muscles of deglutition often leads to aspiration of food, which may precipitate acute deterioration. Chronic alveolar hypoventilation may also develop in patients with adequate muscle strength, which suggests that their disease may involve a defect in central control mechanisms. As with all neuromuscular syndromes, CNS-depressant drugs should be avoided whenever possible or given in minimal doses when necessary. Inspiratory muscle training may delay the onset of chronic respiratory failure.[40] Nocturnal ventilation with noninvasive techniques, such as nasal intermittent positive pressure ventilation or external negative pressure ventilation, may be useful in the later stages of these diseases.[41]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. ATS/ERS statement on respiratory muscle testing. Am J Respir Crit Care Med 166:518, 2002

2. Koenig SM: Pulmonary complications of obesity. Am J Med Sci 321:249, 2001

3. Rubinstein I, Zamel N, DuBarry L, et al: Airflow limitation in morbidly obese, nonsmoking men. Ann Intern Med 112:828, 1990

4. Sahebjami H: Dyspnea in obese healthy men. Chest 114:1373, 1998

5. Unterborn J: Pulmonary function testing in obesity, pregnancy, and extremes of body habitus. Clin Chest Med 22:759, 2001

6. Masa JF, Celli BR, Riesco JA, et al: The obesity hypoventilation syndrome can be treated with noninvasive mechanical ventilation. Chest 119:1102, 2001

7. Kearon C, Viviani GR, Killian KJ: Factors influencing work capacity in adolescent idiopathic thoracic scoliosis. Am Rev Respir Dis 148:295, 1993

8. Bergofsky EH: Respiratory failure in disorders of the thoracic cage. Am Rev Respir Dis 119:643, 1979

9. Pehrsson K, Bake B, Larsson S, et al: Lung function in adult idiopathic scoliosis: a 20-year follow-up. Thorax 46:474, 1991

10. Pehrsson K, Danielsson A, Nachemson A: Pulmonary function in adolescent scoliosis: a 25 year follow up after surgery or start of brace treatment. Thorax 56:388, 2001

11. Schonhofer B, Wallstein S, Wiese C, et al: Noninvasive mechanical ventilation improves endurance performance in patients with chronic respiratory failure due to thoracic restriction. Chest 119:1371, 2001

12. Midgren B: Oxygen desaturation during sleep as a function of the underlying respiratory disease. Am Rev Respir Dis 141:43, 1990

13. Masa JF, Celli BR, Riesco JA, et al: Noninvasive positive pressure ventilation and not oxygen may prevent overt ventilatory failure in patients with chest wall diseases. Chest 112:207, 1997

14. Lee-Chiong TL Jr: Pulmonary manifestations of ankylosing spondylitis and relapsing polychondritis. Clin Chest Med 19:747, 1998

15. Haller JA Jr, Kramer SS, Lietman SA: Use of CT scans in selection of patients for pectus excavatum surgery: a preliminary report. J Pediatr Surg 22:904, 1987

16. Peterson RJ, Young WG Jr, Godwin JD, et al: Noninvasive assessment of exercise cardiac function before and after pectus excavatum repair. J Thorac Cardiovasc Surg 90:251, 1985

17. Morshuis W, Folgering H, Barentsz J, et al: Pulmonary function before surgery for pectus excavatum and at long-term follow-up. Chest 105:1646, 1994

18. Lacquet LK, Morshuis WJ, Folgering HT: Long-term results after correction of anterior chest wall deformities. J Cardiovasc Surg 39:683, 1998

19. Sigalet DL, Montgomery M, Harder J: Cardiopulmonary effects of closed repair of pectus excavatum. J Pediatr Surg 38:380, 2003

20. Gay PC, Westbrook PR, Daube JR, et al: Effects of alterations in pulmonary function and sleep variables on survival in patients with amyotrophic lateral sclerosis. Mayo Clin Proc 66:686, 1991

21. Tripp HF, Bolton JW: Phrenic nerve injury following cardiac surgery: a review. J Card Surg 13:218, 1998

22. Mulvey DA, Aquilina RJ, Elliott MW, et al: Diaphragmatic dysfunction in neuralgic amyotrophy: an electrophysiologic evaluation of 16 patients presenting with dyspnea. Am Rev Respir Dis 147:66, 1993

23. Hart N, Nickol AH, Cramer D, et al: Effect of severe isolated unilateral and bilateral diaphragm weakness on exercise performance. Am J Respir Crit Care Med 165:1265, 2002

24. Stolk J, Versteegh MI: Long-term effect of bilateral plication of the diaphragm. Chest 117:786, 2000

25. Simansky DA, Paley M, Refaely Y, et al: Diaphragm plication following phrenic nerve injury: a comparison of paediatric and adult patients. Thorax 57:613, 2002

26. Baydur A, Adkins RH, Milic-Emili J: Lung mechanics in individuals with spinal cord injury: effects of injury level and posture. J Appl Physiol 90:405, 2001

27. Grimm DR, Arias E, Lesser M, et al: Airway hyperresponsiveness to ultrasonically nebulized distilled water in subjects with tetraplegia. J Appl Physiol 86:1165, 1999

28. DeLuca RV, Grimm DR, Lesser M, et al: Effects of a beta$_2$-agonist on airway hyperreactivity in subjects with cervical spinal cord injury. Chest 115:1533, 1999

29. Rutchik A, Weissman AR, Almenoff PL, et al: Resistive inspiratory muscle training in subjects with chronic cervical spinal cord injury. Arch Phys Med Rehabil 79:293, 1998

30. Kaplan LM, Hollander D: Respiratory dysfunction in amyotrophic lateral sclerosis.

Clin Chest Med 15:675, 1994

31. Teitelbaum JS, Borel CO: Respiratory dysfunction in Guillain-Barré syndrome. Clin Chest Med 15:705, 1994

32. Sharshar T, Chevret S, Bourdain F, et al: Early predictors of mechanical ventilation in Guillain-Barre syndrome. Crit Care Med 31:278, 2003

33. Varelas PN, Chua HC, Natterman J, et al: Ventilatory care in myasthenia gravis crisis: assessing the baseline adverse event rate. Crit Care Med 30:2663, 2002

34. Segredo V, Caldwell JE, Matthay MA, et al: Persistent paralysis in critically ill patients after long-term administration of vecuronium. N Engl J Med 327:524, 1992

35. Swift TR: Disorders of neuromuscular transmission other than myasthenia gravis. Muscle Nerve 4:334, 1981

36. Flaherty KR, Wald J, Weisman IM, et al: Unexplained exertional limitation: characterization of patients with a mitochondrial myopathy. Am J Respir Crit Care Med 164:425, 2001

37. Warrington KJ, Moder KG, Brutinel WM: The shrinking lungs syndrome in systemic lupus erythematosus. Mayo Clin Proc 75:467, 2000

38. Lynn DJ, Woda RP, Mendell JR: Respiratory dysfunction in muscular dystrophy and other myopathies. Clin Chest Med 15:661, 1994

39. Phillips MF, Quinlivan RC, Edwards RH, et al: Changes in spirometry over time as a prognostic marker in patients with Duchenne muscular dystrophy. Am J Respir Crit Care Med 164:2191, 2001

40. Koessler W, Wanke T, Winkler G, et al: 2 years' experience with inspiratory muscle training in patients with neuromuscular disorders. Chest 120:765, 2001

41. Hill NS, Eveloff SE, Carlisle CC, et al: Efficacy of nocturnal nasal ventilation in patients with restrictive thoracic disease. Am Rev Respir Dis 145:365, 1992

Acknowledgment

Figure 1 Janet Betries.

VIII RESPIRATORY FAILURE

MARIN KOLLEF, M.D.

The major physiologic function of the lungs is gas exchange between the pulmonary capillaries and the alveolar space. Oxygen is taken up from the alveolar space onto hemoglobin in red blood cells for transport to peripheral tissues; carbon dioxide is eliminated after it is carried to the lungs via the venous circulation. Acute respiratory failure is the inability of the respiratory system to provide adequate gas exchange. This can take the form of inadequate oxygenation of the arterial blood, insufficient removal of carbon dioxide from venous blood, or both. Acute respiratory failure can be caused by lesions affecting several parts of the respiratory system, including the airways, lung parenchyma, chest wall and respiratory muscles, and neuromuscular processes involved in breathing [*see Table 1*].

Definitions and Pathogenesis

ACUTE HYPOXEMIC RESPIRATORY FAILURE

Definition

Acute hypoxemic respiratory failure is defined as a decrease in the delivery of oxygen from the atmosphere to the blood and, more specifically, as an arterial oxygen tension (P_aO_2) of less than 60 mm Hg. Although somewhat subjective, this value was selected on the basis of the beginning of the descent of the oxygen-hemoglobin dissociation curve at a P_O_2 of 60 mm Hg.

Pathogenesis

The causes of acute hypoxemic respiratory failure are a low inspired concentration of oxygen, impairment of oxygen diffusion, alveolar hypoventilation, ventilation-perfusion ($\dot{V}/\dot{Q}$) mismatch, intrapulmonary shunting, and a low mixed venous oxygen content.

Low inspired concentration of oxygen is an uncommon cause of acute hypoxemic respiratory failure. This can occur at high altitudes or when toxic gases are inhaled (e.g., smoke inhalation). In patients with other cardiopulmonary disease processes, an inappropriately low fraction of inspired oxygen (F_IO_2) can contribute to hypoxic respiratory failure. Acute hypoxemic respiratory failure resulting from diffusion impairment is also relatively uncommon. This usually occurs in the setting of acute or chronic interstitial lung disease. Patients with acute respiratory distress syndrome (ARDS) can also have diffusion impairments contributing to hypoxemia, but shunting is the more important physiologic derangement in this disorder (see below). Both low inspired concentrations of oxygen and diffusion impairments usually respond to supplemental oxygen therapy.

Pure alveolar hypoventilation is a relatively rare form of acute hypoxemic respiratory failure that is caused by neuromuscular or central nervous system dysfunction (e.g., opiate overdose). The lung parenchyma is essentially normal. This condition is characterized by an acute reduction in effective alveolar ventilation and a subsequent decrease in the amount of CO_2 that is eliminated by the lungs. Therefore, the arterial carbon dioxide tension (P_aCO_2) and the P_aO_2 are always increased in patients with pure alveolar hypoventilation. Hypoxemia occurs as a result of the displacement of oxygen in the alveolar space resulting from the failure to eliminate CO_2. Calculation of the alveolar-arterial oxygen gradient or difference [A-aD_O_2] is helpful in determining whether acute hypoxemic respiratory failure is from pure alveolar hypoventilation. With pure alveolar hypoventilation, the A-aD_O_2 is in the normal range of 5 to 25 mm Hg when the patient is breathing room air. However, when acute hypoxemic respiratory failure is caused by $\dot{V}/\dot{Q}$ mismatching or an intrapulmonary shunt, the A-aD_O_2 is always increased.

$\dot{V}/\dot{Q}$ mismatching is the most common pathophysiologic cause of acute hypoxemia. It develops when there is a decrease in ventilation to normally perfused regions of the lung, a decrease in perfusion to normally ventilated regions of the lung, or some combination of a decrease in both ventilation and perfusion. Regions of the lung with low $\dot{V}/\dot{Q}$ ratios caused by inadequate ventilation primarily result in arterial hypoxemia. By contrast, regions of the lung with high $\dot{V}/\dot{Q}$ ratios caused by inadequate perfusion result in wasted ventilation, or dead-space ventilation, and are typically associated with hypercapnia when severe. The degree of hypoxemia in patients with pure $\dot{V}/\dot{Q}$ mismatching improves in response to an increase in the F_IO_2. This correction occurs because airways to poorly ventilated alveoli remain patent, and the increased inspired oxygen will eventually reach pulmonary capillary blood. Another clinical use for the calculation of the A-aD_O_2 is to identify $\dot{V}/\dot{Q}$ mismatching, when the measured P_aO_2 is normalized by hyperventilation. For example, a P_aO_2 of 90 mm Hg and a P_aCO_2 of 20 mm Hg when breathing room air would represent significant $\dot{V}/\dot{Q}$ mismatching, as evidenced by the calculated A-aD_O_2 of 35 mm Hg.

Physiologic shunting occurs when venous blood bypasses ventilated alveoli and enters the arterial system. In the normal lung, a 2% to 3% shunt normally occurs because of the bronchial artery circulation and drainage of some coronary venous blood directly into the left ventricle via the thebesian veins. In patients with severe pneumonia, atelectasis, or pulmonary edema, intrapulmonary shunting occurs when pulmonary capillary blood passes next to alveoli that are completely collapsed or filled with edema fluid or inflammatory cells. Shunting can be differentiated from $\dot{V}/\dot{Q}$ mismatching on the basis of the differences in the response to inhalation of 100% oxygen. P_aO_2 levels in patients with shunting who receive 100% oxygen will not improve to normal. In fact, the change in the P_aO_2 level that occurs in response to F_IO_2 values allows the shunt to be estimated as a percentage of cardiac output, assuming normal values for the difference in oxygen content between arterial and mixed venous blood and P_aCO_2 [*see Figure 1*]. With a shunt of 30% or greater, the P_aO_2 rises little, if at all, with increasing F_IO_2. By contrast, the rise in P_aO_2 with oxygen therapy is appreciable even with a severe $\dot{V}_A/\dot{Q}$ mismatch.

Low mixed venous oxygenation can also contribute to hypoxemia. However, it is uncommon for this to be the only factor contributing to acute hypoxemic respiratory failure. Normally, the lungs fully oxygenate pulmonary arterial blood, and mixed venous oxygen tension (P–vO_2) does not affect P_aO_2 significantly. However, a decreased P–vO_2 can lower the P_aO_2 significantly when either intrapulmonary shunting or $\dot{V}/\dot{Q}$ mismatch is present. Factors that can contribute to low mixed venous oxygenation include anemia, hypoxemia, inadequate cardiac output

Table 1 Causes of Acute Respiratory Failure

Acute Hypoxemic Respiratory Failure	*Acute Hypercapnic Respiratory Failure*	*Mixed Respiratory Failure*
Lung parenchymal disease process	Lung parenchymal disease process	Includes many disorders that can be associated with both hypoxemia and hypercapnia
Pneumonia	Emphysema	
Aspiration	Interstitial lung disease (usually end stage)	
Acute lung injury/acute respiratory distress syndrome	Airway disease process (upper and lower airways)	
Asthma	Asthma	
Emphysema/chronic bronchitis	Obstructive sleep apnea	
Smoke inhalation	Vocal cord dysfunction	
Interstitial lung disease	Airway obstruction	
Atelectasis	Chest wall/pleural space disease process	
Radiation injury	Obesity hypoventilation syndrome	
Cardiogenic pulmonary edema	Kyphoscoliosis	
Pulmonary vascular disease process	Pleural effusion	
Pulmonary embolism	Pleural fibrosis	
Fat embolism	Traumatic flail chest	
Air embolism	Pneumothorax	
Airway disease process	Neuromuscular disease process	
Airway obstruction	Brain	
Asthma	Narcotic overdose	
Airway edema (heat injury, allergic reaction)	Cerebrovascular accident	
	Encephalitis/meningitis	
	Spinal cord	
	Amyotrophic lateral sclerosis	
	Poliomyelitis	
	Tetanus	
	Peripheral nerve	
	Phrenic nerve injury	
	Botulism	
	Guillain-Barré syndrome	
	Muscle	
	Electrolyte disorders (hypophosphatemia)	
	Muscular dystrophies	
	Diaphragmatic atrophy (severe emphysema)	

(CO) as occurs in cardiogenic shock, and increased oxygen consumption. Improving oxygen delivery to tissues by increasing hemoglobin or CO usually decreases oxygen extraction and improves P–vO_2 and, subsequently, P_aO_2.

ACUTE HYPERCAPNIC RESPIRATORY FAILURE

Definition

Acute hypercapnic respiratory failure is defined as a P_aCO_2 greater than 45 to 50 mm Hg along with respiratory acidosis. Chronic failure is also marked by elevated P_aCO_2 levels, but in patients with chronic respiratory failure, renal compensation tends to normalize the pH. The distinction of acute from chronic hypercapnic respiratory failure is important, because the two have clearly different prognostic significance and therapeutic implications.

Pathogenesis

Three processes, alone or in combination, can produce acute hypercapnic respiratory failure: a reduction in minute ventilation, an increase in wasted ventilation, and an increase in CO_2 production.

Reduced minute ventilation can occur as a consequence of central nervous system disorders (e.g., brain injury or spinal cord lesions), peripheral nerve diseases (e.g., Guillain-Barré syndrome, botulism, myasthenia gravis or amyotrophic lateral sclerosis), muscle disorders (e.g., polymyositis or muscular dystrophy), chest wall abnormalities (e.g., thoracoplasty or scoliosis), drug overdoses, metabolic abnormalities (e.g., myxedema, hypokalemia), and upper airway obstruction. These disorders normally are associated with a normal A-aDO_2 unless lung disease is also present.

Wasted ventilation is defined as the ratio of the dead space volume (V_D) to the tidal volume (V_T). An increase in wasted ventilation (i.e., an increased V_D/V_T ratio) is caused by overventilation of regions of the lung relative to their perfusion. This may occur in intrinsic lung diseases (e.g., emphysema, asthma, cystic fibrosis or pulmonary fibrosis) and in chest wall disorders associated with parenchymal abnormalities (e.g., scoliosis). Usually, these disorders are also associated with widened A-aDO_2 gradients.

Increased CO_2 production ($\dot{V}CO_2$) in hospitalized patients is usually a result of infection, trauma, burns, or other major stresses that lead to hypermetabolism. Agitation, myoclonus, or other causes of muscle activity can increase $\dot{V}CO_2$ and contribute to the development of hypercapnic respiratory failure. During refeeding, lipogenesis from the oxidation of carbohydrates can increase the metabolic respiratory quotient significantly. The respiratory quotient (i.e., the ratio of the volume of carbon dioxide released to the volume of oxygen consumed), which is normally 0.8, may rise as high as 2.0 in such cases; this basically doubles $\dot{V}CO_2$. In patients who have severe lung disease or are on fixed mechanical ventilation, acute hypercapnia may occur.

Diagnosis

Acute Hypoxemic Respiratory Failure

When the P_aO_2 rapidly falls below 40 to 50 mm Hg, harmful effects may be observed in various organ systems. Patients may experience headache, somnolence, confusion, and seizures. With more severe hypoxemia, permanent encephalopathy may occur. Cardiovascular sequelae from mild hypoxemia, including tachycardia and hypertension, may also develop. With severe hypoxemia, opposite effects may occur, such as bradycardia and hypotension. Patients who have disorders associated with hypoxia (i.e., decreased delivery of oxygen to the peripheral tissues) without concurrent hypoxemia can present with similar clinical signs and symptoms. Causes of hypoxia without hypoxemia include anemia, decreased cardiac output, and carbon monoxide or cyanide poisoning.

Acute Hypercapnic Respiratory Failure

Signs and symptoms of hypercapnia depend not only on the absolute level of P_aCO_2 but also on the rate at which the level increases. A P_aCO_2 above 100 mm Hg may be well tolerated if the hypercapnia develops slowly and acidemia is minimized by renal compensatory changes. However, acute increases in P_aCO_2 levels are associated with several neurologic sequelae, including increased cerebral blood flow and elevation in intracranial pressure.

Acute elevation in P_aCO_2 to 80 to 90 mm Hg may produce many neurologic signs and symptoms, including confusion, headaches, seizures, and coma. A careful neurologic examination of a patient with acute hypercapnia may reveal agitation, coarse tremor, slurred speech, asterixis, and occasionally papilledema. These effects of hypercapnia on the central nervous system are fully reversible, as opposed to the potentially permanent neurologic sequelae that are associated with acute hypoxemia. An elevated P_aCO_2 is also associated with myocardial depression, arrhythmias, hyperkalemia, and gastrointestinal bleeding.

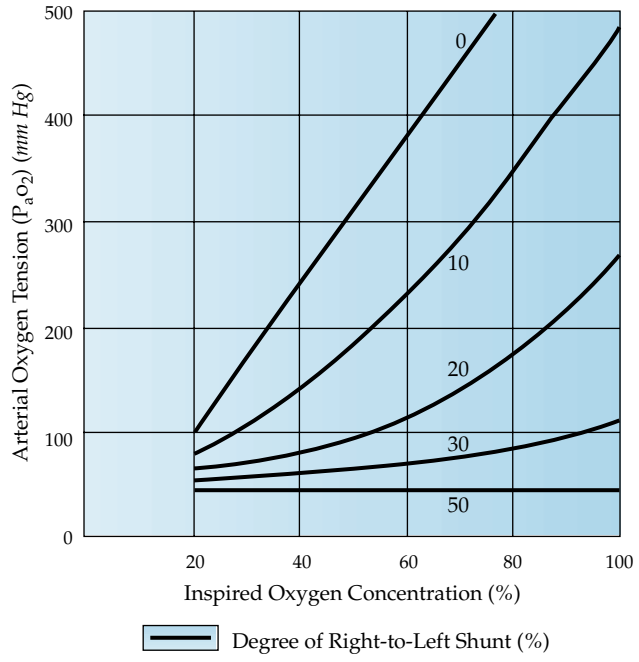


Figure 1 **The oxygen concentration of inspired gas is plotted against arterial oxygen tension (P_aO_2) for right-to-left shunts ranging from 0% to 50% of cardiac output. The arteriovenous oxygen content difference is assumed to be normal. Note that with right-to-left shunts in excess of 30%, there is virtually no increase in P_aO_2 with oxygen enrichment.**

Table 2 Radiographic Appearance of Disorders That Cause Acute Respiratory Failure

Diffuse or Patchy Infiltrates Present	*Relatively Clear Lungs*
Acute respiratory distress syndrome	Acute hypoxemic respiratory failure
Pneumonia	Pulmonary embolism
Cardiogenic pulmonary edema	Air embolism
Atelectasis	Fat embolism
Aspiration	Acute hypercapnic respiratory failure
Interstitial lung disease	Asthma
Pulmonary contusion	Exacerbation of COPD
Alveolar hemorrhage syndrome	Narcotic overdose
Metastatic or lymphatic spread of tumor	Neuromuscular disease
	Obesity-hypoventilation syndrome

COPD—chronic obstructive pulmonary disease

Differential Diagnosis

The many disorders that cause acute respiratory failure can be classified into two categories on the basis of the presence or absence of infiltrates on chest radiographs [*see Table 2*].[1] Most patients with acute respiratory failure have patchy or diffuse infiltrates and primarily develop hypoxemic respiratory failure. In patients whose chest radiographs show infiltrates, the most common causes of acute hypoxemic respiratory failure are pneumonia, cardiogenic pulmonary edema, noncardiogenic pulmonary edema (ARDS), atelectasis, aspiration, progressive interstitial lung disease, pulmonary contusion, and alveolar hemorrhage syndromes (e.g., Goodpasture syndrome, Wegener granulomatosis, and systemic lupus erythematosus).

Acute respiratory failure in a patient with a relatively clear chest radiograph can be either hypoxemic or hypercapnic. Acute hypoxemic respiratory failure in such patients is most often the result of pulmonary embolism or some other form of pulmonary vascular disease. Acute hypercapnic respiratory failure has five principal causes: (1) acute exacerbation of chronic obstructive pulmonary disease (COPD) or asthma, (2) neuromuscular dysfunction, (3) obesity-hypoventilation syndrome, (4) obstructive sleep apnea, and (5) drug overdose associated with a depressed respiratory drive.

Management

The three major steps in the management of acute respiratory failure are to (1) ensure an open airway, (2) restore oxygenation, and (3) maintain or restore ventilation to eliminate CO_2.

ENSURING AIRWAY PATENCY

Airway obstruction may develop in patients with depressed consciousness, dysfunction of the upper airway muscles caused by neuromuscular disease, or an inability to cough and clear secretions. Simple measures such as pushing the mandible for-

Table 3 Procedure for Direct Orotracheal Intubation

1. Administer oxygen by face mask. Ensure arterial oxygen saturation > 95% before attempting intubation, if possible.
2. Gather basic equipment: oxygen source, bag-valve device, suctioning device, endotracheal (ET) tube, blunt stylet, laryngoscope, 20 ml syringe. Have all equipment easily accessible.
3. Place patient on nonmobile rigid surface.
4. If patient is in hospital bed, remove backboard and adjust bed height.
5. Depress patient's tongue with tongue depressor and administer topical anesthesia to patient's pharynx.
6. Position patient's head in sniffing position by placing a small towel under the occiput and extending the head and neck.
7. Administer intravenous sedation and neuromuscular blocker if necessary.*
8. Have assistant apply Sellick maneuver (compressing cricothyroid cartilage against vertebral bodies) to occlude esophagus and prevent regurgitation and aspiration of stomach contents.
9. Grasp laryngoscope handle in left hand while opening patient's mouth with gloved right hand.
10. Insert laryngoscope blade on right side of patient's mouth and advance to base of tongue, displacing tongue to the left.
11. Lift laryngoscope away from patient at a 45° angle using arm and shoulder strength. Do not use patient's teeth as a fulcrum.
12. Suction oropharynx and hypopharynx if necessary.
13. Grasp ET tube with inserted stylet in right hand and insert it into right corner of patient's mouth, avoiding obscuration of epiglottis and vocal cords.
14. Advance ET tube through vocal cords until cuff is no longer visible, then remove stylet.
15. Inflate cuff with enough air to prevent significant air leakage.
16. Verify correct ET tube positioning by auscultation of both lungs and the abdomen.
17. Obtain a chest radiograph to verify correct position of the ET tube.

*Neuromuscular blockade can result in complete airway collapse and airway obstruction; personnel who are skilled in establishment of an emergency surgical airway should be available if paralysis is used.

ward, laying the patient in a lateral decubitus position, or placing a nasal trumpet or oral airway may alleviate upper airway obstruction caused by relaxed upper airway muscles. However, endotracheal intubation should be performed when bronchial secretions are excessive or the patient is at high risk for the aspiration of gastric contents. Emergency cricothyrotomy or tracheostomy is required only if there is glottic or infraglottic anatomic obstruction or if endotracheal intubation cannot be performed for other reasons, such as trauma to the oropharynx.

An endotracheal tube may be passed orally or nasally. The oral route is easier and faster and allows the use of a larger-diameter tube. Proper patient positioning and preparation are required to optimize success with endotracheal intubation [*see Table 3*]. The larger tube facilitates suctioning of secretions and fiberoptic bronchoscopy and decreases the risk of sinusitis.[2] Nasotracheal intubation may be preferable in awake and unanesthetized patients and in patients with limited mobility of the cervical vertebrae, such as that caused by ankylosing spondylitis or rheumatoid arthritis.

Complications of endotracheal intubation attempts occur immediately if the tube enters the esophagus. Other early complications of endotracheal intubation include cervical spine injury in patients with trauma or arthritis and injury to the teeth, nose, pharynx, larynx, and tracheobronchial tree. In addition, both prolonged tracheal intubation and tracheostomy can lead to the serious sequelae of tracheal stenosis or tracheomalacia, which can result from ischemic injury of the trachea by the inflatable cuff of the tube. Such difficulties have been reduced since the advent of high-volume and low-pressure cuffs and can be further reduced by not overinflating the cuff. Tracheostomies and endotracheal tubes are associated with the same long-term complications. Tracheostomy may provide greater comfort and more effective secretion removal and should be considered if mechanical ventilation is needed beyond 2 to 3 weeks.[3]

RESTORING ARTERIAL OXYGENATION

An increase in the F_IO_2 may suffice to restore normal oxygenation. On the basis of the shape of the oxygen-hemoglobin dissociation curve, achieving a P_aO_2 of 60 mm Hg or higher (> 90% oxygen saturation of hemoglobin) is an optimal goal for most acutely hypoxemic patients. In patients with chronic hypoxemia and hypercapnia, a P_aO_2 of 50 to 55 mm Hg may be needed to prevent respiratory depression and worsening hypercapnia.

Supplemental oxygen can be administered by a number of techniques to the spontaneously breathing patient. Nasal cannulas allow patients to eat, drink, and speak during oxygen administration. Their disadvantage is that the exact F_IO_2 delivered is not known, because it is influenced by the patient's peak inspiratory flow demand. As an approximation, the following guide can be used: 1 L/min of nasal-prong oxygen flow is approximately equivalent to an F_IO_2 of 24%, with each additional 1 L of flow increasing the F_IO_2 by approximately 4%. Flow rates should be limited to less than 5 L/min. Venturi masks allow more precise administration of oxygen to be delivered, with calibrated mask F_IO_2 values between 24% and 50%. Often, Venturi masks are useful in patients with COPD and hypercapnia because one can titrate the P_aO_2 to minimize carbon dioxide retention. Nonrebreathing masks achieve higher oxygen concentrations (approximately 80% to 90%) than partial rebreathing systems. A one-way valve prevents exhaled gases from entering the reservoir bag in a nonrebreathing system, thereby maximizing the F_IO_2.

A continuous positive airway pressure (CPAP) mask can be used if the P_aO_2 is less than 60 to 65 mm Hg during use of a nonrebreathing mask and the patient is conscious and cooperative, able to protect the lower airway, and hemodynamically stable. CPAP is delivered by a tight-fitting mask equipped with pressure-limiting valves. Many patients cannot tolerate a CPAP mask because of persistent hypoxemia, hemodynamic instability, or feelings of claustrophobia or aerophagia. In such patients, endotracheal intubation should be performed. Initially, 3 to 5 cm H_2O of CPAP should be applied while monitoring the P_aO_2 or arterial oxygen saturation (S_aO_2). If the P_aO_2 is still less than 60 mm Hg (S_aO_2 < 90%), the level of CPAP should be increased in increments of 3 to 5 cm H_2O up to a level of 10 to 15 cm H_2O. CPAP may improve oxygenation by opening previously closed alveoli and decreasing intrapulmonary shunting. Although the use of CPAP may prevent the need for intubation and mechanical ventilation, it may result in such complications as gastric distention, drying of the eyes from air leaks, and skin breakdown, especially on the bridge of the nose.[4] Bilevel positive airway pressure (BiPAP) is a method of noninvasive ventilation in which inspiratory and expiratory pressure can be applied through a mask during the patient's respiratory cycle. The in-

spiratory support decreases the patient's work of breathing. The expiratory support (CPAP) improves gas exchange by preventing alveolar collapse. Noninvasive ventilation using face or nasal masks may obviate endotracheal intubation and mechanical ventilation in patients with neuromuscular disease, COPD, and postoperative respiratory insufficiency.

In using BiPAP, a pressure-support ventilation (PSV) level of 5 to 10 cm H_2O and a CPAP level of 3 to 5 cm H_2O are reasonable starting points. The PSV level can be increased in increments of 3 to 5 cm H_2O, using the patient's respiratory rate as a guide of effectiveness.

MECHANICAL VENTILATORY SUPPORT

When adequate oxygenation cannot be maintained by noninvasive means or if progressive hypoventilation and hypercapnia with respiratory acidosis occurs, endotracheal intubation and mechanical ventilatory support should be initiated. Mechanical ventilation can produce positive pressure at the airway opening or create negative pressure around the chest wall. Use of negative pressure ventilators (e.g., an iron lung) is generally restricted to patients with chronic neuromuscular weakness or chest wall deformity.

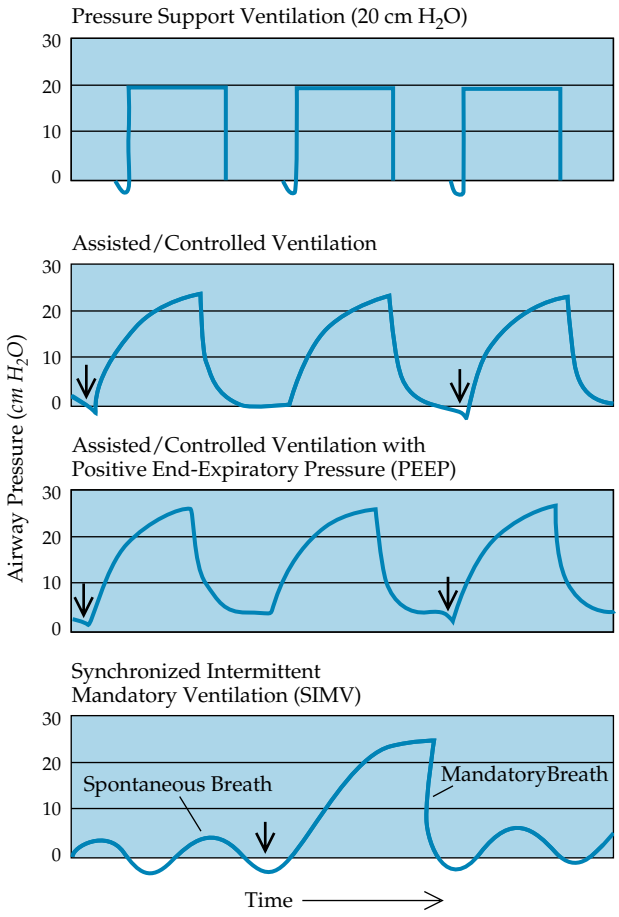


Figure 2 **Pressures measured at the endotracheal tube are shown as a function of time to illustrate the effects of various modes of pressure-preset (top panel) or constant tidal volume (bottom three panels) mechanical ventilation. Supra-atmospheric pressures have positive values, whereas subatmospheric pressures have negative values. Arrows indicate initiation of inspiration by the patient, triggering the ventilator to deliver an assisted breath.**

Volume-Cycled and Pressure-Cycled Ventilation

There are two basic types of positive pressure ventilation: volume cycled and pressure cycled.[5,6] Volume-cycled ventilation, the more commonly used mode in adults, supplies a fixed tidal volume, making inflation pressure the dependent variable. When a volume-cycled ventilator is used, changes occurring in pulmonary impedance are associated with alterations in the airway pressures during inflation. A pressure limit can be set using a pop-off valve that prevents further inflation, avoiding excessive overdistention, and functions as an alarm.

Pressure-cycled ventilation provides gas flow until a preset pressure is reached, so that the tidal volume becomes the dependent variable. With this mode, changes in pulmonary impedance are associated with alterations in tidal volume and therefore minute ventilation.

With volume-cycled ventilation, two primary modes are used: assisted/controlled (A/C) ventilation and synchronized intermittent mandatory ventilation (SIMV) [*see Figure 2*]. In the A/C mode, the ventilator guarantees a preset number of breaths (backup rate) supplied at a preset tidal volume. If the patient desires to breathe at a respiratory rate higher than the preset rate, the ventilator will deliver the entire preset tidal volume every time the patient generates a small negative airway pressure (–1 to –2 cm H_2O). In the SIMV mode, the ventilator also guarantees a preset number of breaths (backup rate) supplied at a preset tidal volume. However, the tidal volume for any additional breath above the preset rate is determined by the effort of the patient. Therefore, the SIMV mode combines a preset number of ventilator-delivered mandatory breaths with the ability to assist intermittent patient-generated spontaneous breaths. Most patients can be effectively ventilated with either A/C or SIMV, and both modes can be used with pressure-controlled ventilators. The A/C mode has certain advantages over SIMV in that the former usually requires less respiratory effort, reduces oxygen consumption, and is more likely to rest respiratory muscles. However, the respiratory muscles may continue to perform considerable work during A/C breathing, especially when respiratory drive and minute ventilation are increased. Potential advantages of SIMV include the exercising of respiratory muscles, the prevention of respiratory alkalemia, and, possibly, improved patient-ventilator coordination. However, a change from A/C ventilation to SIMV does not usually correct respiratory alkalemia caused by an increased respiratory drive.

Pressure-Support Ventilation

PSV is a pressure-targeted, flow-cycled mode that requires the patient to initiate every breath [*see Figure 2*]. During the inspiratory phase, pressures rise rapidly to a preset plateau level. The pressure terminates when the inspiratory airflow created by the patient falls below a certain level. Therefore, the total work of breathing for each breath is usually generated by a combination of patient effort and mechanical support. Intubated patients can be ventilated with pressure support alone, as long as they have an adequate ventilatory drive and are strong enough to initiate a sufficient number of breaths to maintain adequate minute ventilation. When used as the primary mode of ventilation, pressure support can be used with certain amounts of CPAP and generally is well tolerated by patients who are being weaned from mechanical ventilation. The pressure support mode of ventilation can also be used in conjunction with SIMV, in which case the level of pressure support is applied only to patient-generated breaths.

Selecting Appropriate Mechanical Ventilatory Settings

After the patient is intubated, the respiratory rate and tidal volume (in volume-controlled ventilation) should be set to maintain an adequate minute ventilation that will result in an appropriate pH level. The tidal volume should be large enough to prevent microatelectasis and progressive hypoxemia but without causing barotrauma. For most patients, an initial tidal volume between 6 and 10 ml/kg body weight is appropriate [*see* Management of Respiratory Failure in Specific Clinical Settings, Acute Respiratory Distress Syndrome, *below*].

Positive end-expiratory pressure (PEEP) is used to improve oxygenation and reduce the possibility of oxygen toxicity for patients requiring high levels of inspired oxygen. PEEP works by opening or recruiting previously closed alveoli and redistributing lung water from the alveoli to the interstitial spaces. CPAP works in a similar manner for patients who are breathing spontaneously. The benefits of PEEP must be weighed against its deleterious effects, which include decreased cardiac output, increased intracranial pressure, and hyperinflation.

Selecting the appropriate mode, respiratory rate, tidal volume (or pressure, in pressure-controlled ventilation), and amount of PEEP is a dynamic process. No single setting is appropriate for all patients or for a given patient at all stages of acute illness. Optimal ventilatory settings require constant monitoring and numerous readjustments based on acute changes in gas exchange, airway pressures, breathing patterns, and hemodynamics in conjunction with the resolution or progression of the underlying disease process.

During volume-cycled ventilation, peak and plateau airway pressures can be easily seen from most ventilators. Peak airway pressure can be subdivided into three components: (1) flow resistive, (2) elastic distending pressure, and (3) PEEP (either set or intrinsic) [*see Figure 3*]. The plateau pressure is a close approximation of the alveolar pressure and can be subdivided into only two components: (1) elastic distending pressure and (2) PEEP (either set or intrinsic). Because positive pressure is required to overcome the flow-resistive properties of the airway and external apparatus, the peak pressure will always be equal to or greater than the plateau pressure. The plateau pressure is measured by interrupting flow or pausing the ventilator at full inspiration. The static compliance of the respiratory system is computed by dividing the tidal volume delivered to the patient by the change in pressure (plateau pressure minus PEEP). A reduction in the normal static compliance is indicative of a stiffer lung or chest wall.

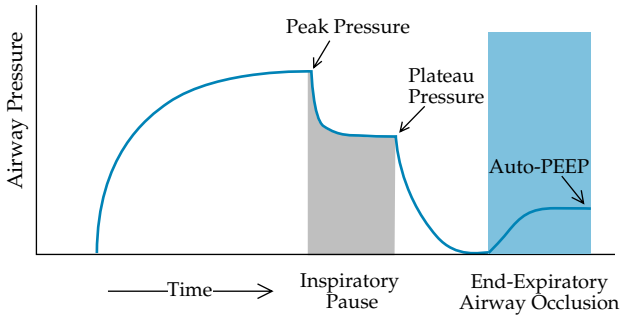


Figure 3 **In volume-cycled pressure ventilation, the total pressure required to overcome flow-resistive and elastic forces and deliver the preset tidal volume is termed peak airway pressure. The plateau pressure is the pressure required to keep the lung and chest wall inflated with the preset tidal volume during a brief period in which there is no flow (inspiratory pause). The difference between the peak and the plateau pressures is a reflection of the flow-resistive component of ventilation and is increased when airway resistance or inspiratory flow rate is increased. The plateau pressure is a function of lung and chest wall compliance and is unaffected by airway resistance and inspiratory flow rate unless auto-PEEP (positive end-expiratory pressure) occurs. Incomplete emptying is associated with auto-PEEP, the magnitude of which can be assessed with end-expiratory airway occlusion.**

During volume-cycled ventilation, changes in airway pressures can be helpful in determining the cause of an acute deterioration in mechanically ventilated patients. A large increase in peak airway pressure (out of proportion to the increase in plateau pressure) signals a change in the flow-resistive properties. This increase is usually observed in patients with airway problems (e.g., kinking of the endotracheal tube, mucous plug, or bronchospasm). Treatment should be directed at improving airway function, including repositioning the endotracheal tube, suctioning the airways, and administering bronchodilator therapy. When an increase is primarily in plateau pressure, the problem is localized to the lung parenchyma or chest wall rather than the airways. This condition can be seen with worsening of pulmonary edema, tension pneumothorax, a large region of atelectasis, or intubation of the right mainstem bronchus.

With both positive and negative pressure ventilators, exhalation is passive and is provided by the elastic recoil of the inflated lungs and chest wall. Exhalation rate depends on the resistance and compliance of the respiratory system. A high resistance is observed in patients who require a prolonged exhalation time, such as those with chronic airway obstruction or asthma. If the subsequent tidal volume is delivered before exhalation is complete, positive airway pressure will be maintained throughout the respiratory cycle, resulting in increases in end-expiratory lung volume and end-expiratory alveolar pressure. PEEP that results from this process of dynamic hyperinflation has been termed auto-PEEP (or intrinsic PEEP). Auto-PEEP increases both peak and plateau pressures and can be measured by briefly occluding the airway at the end of exhalation.[7] When the airflow is interrupted, pressures equilibrate quickly, and the airway-opening pressure rises to the level of the previous alveolar pressure, which is the level of auto-PEEP. The appropriate compliance of the respiratory system can be calculated by dividing the tidal volume by the difference between the plateau pressure and the auto-PEEP.

Management of Respiratory Failure in Specific Clinical Settings

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Patients with severe COPD are often hypoxemic and hypercapnic on a long-term basis and adapt, albeit precariously, to their abnormal state. Acute deterioration is most often triggered by infection, but it may also result from factors such as pneumothorax, congestive heart failure, and increased CO_2 production associated with febrile states. The worsening hypoxemia and hypercapnia that accompany acute deterioration lead to increasing dyspnea, sleep disruption, and, occasionally, alterations in consciousness. Depressed consciousness leads to retention of secretions and further worsening of gas exchange. This cycle can be broken by identifying and rectifying the processes that have precipitated the acute deterioration and by providing

support to improve gas exchange while the underlying disorders are being corrected.

Arterial Blood Gas Analysis

Arterial blood gas analysis is crucial to the proper assessment and management of acute exacerbations. The first priority is to achieve a P_aO_2 level of 50 to 60 mm Hg but no higher. An elevation in P_aCO_2 of 10 to 15 mm Hg is common when oxygen is given to patients with chronic airway obstruction and does not represent a failure of controlled oxygen therapy, provided that there is no critical reduction in blood pH (i.e., ≤ 7.2). Therapy should include promotion of bronchopulmonary drainage by encouragement of cough, administration of inhaled bronchodilators and systemic corticosteroids, and treatment of any underlying infection.[8]

Ventilatory Support

Many patients with severe exacerbations of COPD experience persistent respiratory acidosis and excessive work in breathing even after their initial treatment with bronchodilators. The level for P_aCO_2 at which ventilatory assistance becomes necessary cannot be specified, but ventilatory supports should be considered if hypercapnia is severe enough to cause profound acidemia (pH < 7.2) or if the patient shows signs of altered mental status or respiratory muscle fatigue. Intubation should be performed if hemodynamic instability or somnolence occurs or if secretions cannot be cleared. However, if the hypercapnic patient remains alert and cooperative, delivery of noninvasive positive pressure ventilation through a facial or nasal mask may reverse or prevent fatigue and thereby eliminate the need for conventional mechanical ventilation through an endotracheal tube. In selected patients with severe exacerbations of COPD, noninvasive ventilation decreases the rate of complications, reduces the need for intubation, shortens hospital stay, and may lower mortality.[9,10] Patients who do not tolerate the mask or whose acute exacerbation fails to improve should be intubated. The delay in intubation incurred by an unsuccessful trial of noninvasive ventilation should not pose a significant risk to the patient, provided that personnel skilled in airway management and intubation are readily available and the patient does not have hemodynamic instability or significant underlying cardiac disease. The success rate for noninvasive ventilation is highly dependent on the skill and commitment of the respiratory therapists and nurses who work with the patient.

Noninvasive ventilation The most common method of noninvasive ventilation for patients with airway obstruction is BiPAP. This type of ventilation delivers a specified amount of inspiratory positive airway pressure (IPAP), usually in the form of PSV, that supports each spontaneous breath (usually set at 15 to 20 cm H_2O) in conjunction with a low level of expiratory positive airway pressure (EPAP) (usually set at 3 to 5 cm H_2O). Some patients with acute hypercapnic respiratory failure tolerate a full face mask better than a nose mask because of large leaks that occur with the nose mask. However, securing the mask so tightly that all leaks are prevented may increase discomfort and decrease the ultimate likelihood of success.

Invasive ventilation For exacerbations that require intubation, a reasonable initial ventilator setting is a tidal volume of 8 to 10 ml/kg at 11 to 14 breaths a minute.[11] It is important to remember that P_aCO_2 levels in patients with chronic hypercapnia should not be lowered to the normal range, because this could result in alkalemia, which increases the risk of cardiac dysrhythmias and seizures. In addition, overventilation for more than 2 to 3 days may result in renal restoration of the pH to normal. As a consequence, during subsequent trials of spontaneous ventilation, as the P_aCO_2 rises to the baseline hypercapnic level, the patient becomes acidemic or the patient's respiratory muscles become fatigued because of the greater minute ventilation required for the reset baseline pH and P_aCO_2. To prevent this condition, adjustments in respiratory frequency and tidal volume should be aimed at maintaining the patient's baseline pH and P_aCO_2 or allowing some further degree of tolerated respiratory acidosis.

Dynamic hyperinflation with auto-PEEP occurs frequently during mechanical ventilation of patients with airflow obstruction [*see* Mechanical Ventilatory Support, *above*]. For airflow obstruction, the most critical determinant of the severity of dynamic hyperinflation is the minute ventilation delivered by the mechanical ventilator. Therefore, in addition to the effects on P_aCO_2 and pH levels, excessive ventilation of patients with airflow obstruction increases the risk of pulmonary hyperinflation. The major consequences of excessive auto-PEEP include barotrauma, misinterpretation of central venous and pulmonary arterial wedge pressures, decreased cardiac output secondary to reduced venous return, and increased work in breathing. The effort required to breathe increases because the patient's inspiratory muscles must generate a negative pressure equal to the auto-PEEP before the proximal airway pressure can be lowered to the –1 to –2 cm H_2O required to trigger the ventilator. Auto-PEEP should be suspected in a patient with airflow obstruction when the ventilator does not deliver an assisted breath even though the patient is making obvious inspiratory efforts. Inspiratory effort can be aided by the application of external PEEP at a level equal to or slightly less than the auto-PEEP level. Applying external PEEP causes the ventilator to deliver an assisted breath when the proximal airway pressure is lowered to 1 to 2 cm H_2O below the level of applied PEEP rather than to –1 to –2 cm H_2O. In effect, the application of external PEEP reduces the inspiratory effort required to trigger the ventilator. In patients with chronic airflow obstruction, applied PEEP does not increase lung volume or airway pressures as long as the level of applied PEEP does not exceed the level of auto-PEEP.[12]

ASTHMA

Respiratory failure in children and adults with asthma is usually precipitated by the underlying disease itself, viral or bacterial infection, irritation of the airway from aspiration or inhalation of a toxic gas, or medical noncompliance. Asthmatics usually have a mixed form of respiratory failure with both hypoxemia and hypercapnia.

Ventilatory Support

The goals of ventilatory support in a patient with asthma are to reverse the hypoxemia by administering supplemental oxygen and to reverse the hypercapnia and respiratory acidosis (assuming the patient is not a chronic retainer of CO_2). The approach to ventilatory support is similar to that outlined above for patients with acute exacerbations of COPD. However, patients with asthma may have a greater degree of airway obstruction, which may necessitate additional modes of therapy. The use of inhaled mixtures of helium and oxygen (usually composed of at least 70% helium) can reduce the density of the inspired gases, improving overall ventilation and reducing the pa-

tient's work associated with ventilation. However, in patients requiring increasing supplemental oxygen because of concomitant hypoxemia, the beneficial effects of the helium-oxygen mixture (heliox) may be lost.

ACUTE RESPIRATORY DISTRESS SYNDROME

Hypoxemic respiratory failure in patients with ARDS results in an unacceptably high mortality (30% to 60%). It may occur in previously healthy patients without preexisting lung disease. The pathophysiology, clinical features, and nonventilatory therapies for ARDS are outlined elsewhere[13] [*see 14:X Pulmonary Edema*].

Ventilatory Support

The initiation of mechanical ventilation for ARDS patients clearly has beneficial effects, yet a growing body of evidence indicates that mechanical ventilation can produce pathologic changes in normal lung tissue that are similar to ARDS. Experimentally, large inflation pressures produce capillary leak and, over time, permeability pulmonary edema. Harmful effects of large inflations and subsequent high alveolar pressures have been termed volutrauma.[14] Mechanical ventilation strategies for ARDS patients have been supported by studies that examined computed tomography scans and pressure-volume curves. Although chest radiographs in ARDS patients suggest homogeneous involvement of the lungs, CT scans reveal a markedly heterogeneous pattern.[15] Dependent regions of the lung are consolidated, but more superior regions of the lung appear normal. The inflation portion of the pressure-volume curve of ARDS patients exhibits a lower inflection point, which corresponds to the onset of reopening closed alveoli, and an upper inflection point, which signals the beginning of the overdistention of patent alveoli[16] [*see Figure 4*].

A specific ventilatory strategy that focuses on ventilating patients between the upper and lower inflection points has been proposed for ARDS patients. To limit transpulmonary pressures (plateau pressure of < 30 cm H_2O), low tidal volumes should be used. The subsequent decrease in minute ventilation, which is caused by the use of low tidal volumes and may result in hypercapnia and respiratory acidosis, is termed permissive hypercapnia. This strategy of mechanical ventilation with lower tidal volumes (6 ml/kg) versus traditional ventilation with larger tidal volumes (12 ml/kg) results in decreased mortality and shortens the amount of time that mechanical ventilation is required.[17]

Use of PEEP

PEEP is an effective way to improve oxygenation. However, the optimal amount of PEEP that should be used is controversial. The titration of PEEP to a level above the lower inflection point may decrease the damage to alveoli that would be caused by repetitive reopening and closing of lung units. The application of this level of PEEP, called the open lung approach, was associated with improved mortality when compared with conventional ventilatory strategies.[18] The beneficial effects of PEEP must always be weighed against a possible decrease in cardiac output and increased risk of barotrauma.

Inverse-Ratio Ventilation

One strategy for improving oxygenation without increasing F_IO_2 or PEEP is to prolong inspiratory time by adding an end-inspiratory pause, keeping alveolar pressure briefly at the plateau level.[19] When inspiratory time is prolonged, exceeding expiratory time, the ventilation is termed inverse-ratio ventilation. Inverse-ratio ventilation may improve oxygenation in some ARDS patients; however, prospective studies have found that most patients do not benefit from it.[20] Caution is required in applying inverse-ratio ventilation because barotrauma and hypotension can result from the development of excessive auto-PEEP as expiratory time is shortened.

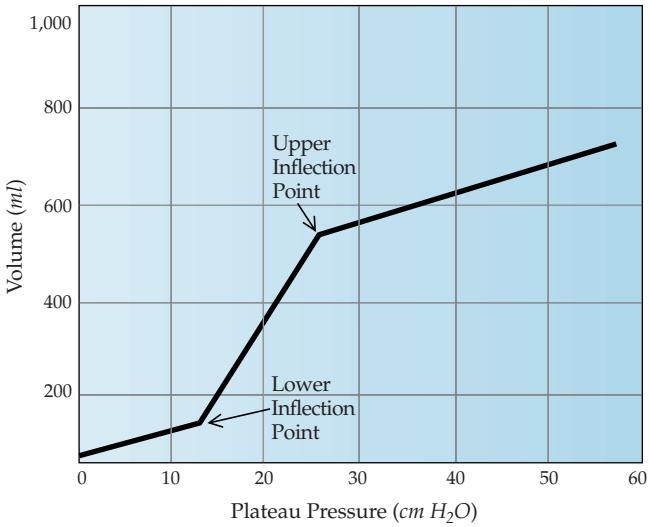


Figure 4 **A static pressure-volume curve can be constructed by plotting the tidal volume versus the elastic recoil pressure of the respiratory system for several different tidal volumes. In patients with acute respiratory distress syndrome, the inflation curve demonstrates a lower inflection point, which corresponds to the onset of reopening closed alveoli, and an upper inflection point, which signals the beginning of the overdistention of patent alveoli.**

Prone Positioning

An effective method for improving oxygenation is to place the patient in the prone position.[21,22] The mechanism by which the prone position improves oxygenation is related to the reduction of shunting and correction of $\dot{V}/\dot{Q}$ mismatching. The recruitment of dorsal atelectatic lung is thought to be produced by a more even distribution of pleural pressure. The transmural distending pressure is greater in the dorsal regions but is not significantly reduced in the dependent ventral region. Studies have found that in 60% to 80% of ARDS patients who are placed in the prone position, the P_aO_2-F_IO_2 ratio improves (from 80 to 200).[22] However, it is not clear whether prone positioning alters mortality for ARDS patients.

Complications of Mechanical Ventilation

PULMONARY COMPLICATIONS

Serious pulmonary complications of intubation and mechanical ventilation can be divided into three categories: (1) infection related to the presence of an endotracheal tube, (2) alveolar overdistention, and (3) atelectasis.

Infection

Mechanically ventilated patients are at high risk (13% to 38%) for nosocomial pneumonia.[23] Early-onset pneumonia, occurring 48 to 72 hours after intubation, is usually the result of aspiration during the intubation process.[24] These infections are most often

caused by antibiotic-sensitive organisms, including oxacillin-sensitive *Staphylococcus aureus*, *Haemophilus influenzae*, and *Streptococcus pneumoniae*. Ventilator-associated pneumonia that occurs more than 72 hours after intubation is frequently caused by antibiotic-resistant pathogens, including *Pseudomonas aeruginosa*, oxacillin-resistant *S. aureus*, *Acinetobacter* species, and *Enterobacter* species. The pathogenesis of ventilator-associated pneumonia usually requires two important steps: (1) bacterial colonization of the aerodigestive tract and (2) aspiration of contaminated secretions into the lower airways. The risk of a ventilator-associated pneumonia appears to be higher in trauma or burn patients, and in all patients, the risk rises as the duration of ventilation increases.[23,25] Diagnosis of ventilator-associated pneumonia may be difficult because many other processes may cause pulmonary infiltrates and fever. Also, cultures obtained by suctioning secretions through the endotracheal tube do not reliably differentiate between pneumonia and bacterial colonization of the trachea. The use of the fiberoptic bronchoscope to obtain specimens with a protected brush or quantitative cultures of bronchial lavage fluid may be helpful in excluding a pulmonary source of infection in intubated patients who have new clinical signs that may be caused by a noninfectious process.[26]

Nonpharmacologic strategies may decrease the incidence of ventilator-associated pneumonia, including adequate hand washing after contact with each patient. Whenever possible, intubated patients should be kept in a semirecumbent position (45° from horizontal), and gastric distention should be avoided to prevent aspiration.[27] In addition, the incidence of ventilator-associated pneumonia can be significantly reduced with the continuous aspiration of subglottic secretions.[28] Several pharmacologic strategies may prevent the development of ventilator-associated pneumonia, such as avoiding the use of unnecessary antibiotics, rotating the class of antibiotics used in the empirical treatment of a suspected bacterial infection, and administering chlorhexidine oral rinse.[29]

Nosocomial sinusitis is strongly related to the nasotracheal route of intubation. In one study, CT demonstrated fluid in the maxillary sinus in 95.5% of patients who underwent both nasotracheal and nasogastric intubation for 1 week, compared with only 22.5% of patients in whom endotracheal and feeding tubes were placed via the oral route.[30] Aspiration of the sinus may reveal nonpurulent mucoid material, but in a significant number of cases, sinusitis will be evidenced by positive stains and cultures of aspirated pus. Treatment of nosocomial sinusitis includes administration of antibiotics, replacement of nasal tubes with oral tubes, and use of decongestants to facilitate drainage.[31]

Alveolar Overdistention

Alveolar overdistention results in two potentially life-threatening problems: hypotension and barotrauma. Ventilator-associated hypotension occurs most often in patients with obstructive airway disease, because the markedly increased lung volume with auto-PEEP impedes venous return. The risk of life-threatening hypotension is greatest at the time of intubation, when preexisting volume depletion and the administration of sedative agents limit the patient's ability to maintain blood pressure. In addition to receiving rapid infusion of intravenous fluids, hypotensive patients who have underlying airway disease should be allowed to reduce the overdistention themselves by interrupting mechanical inflation, before ventilation is resumed at a reduced frequency [*see Figure 5*]. Although less common, ventilator-associated hypotension also occurs in ARDS patients who have high levels of PEEP and when intrinsic PEEP develops with airway obstruction.

Extra-alveolar air caused by positive pressure ventilation is termed barotrauma. Examples include subcutaneous emphysema, pneumomediastinum, pulmonary interstitial emphysema, pneumoperitoneum, arterial gas embolism, and pneumothorax [*see Figure 6*]. Extra-alveolar air usually originates from overdistended alveoli that rupture into the surrounding interstitial space. High alveolar (plateau) pressure and infections that produce lung necrosis increase the risk of barotrauma. Tension pneumothorax is the most common life-threatening manifestation of barotrauma. Tension pneumothorax leads to worsening hypoxemia and decreased venous return with hypotension. Hyperresonance and a reduction of breath sounds on the side of the pneumothorax are common, and inflation pressures are increased. In patients with severe airflow obstruction, differentiat-

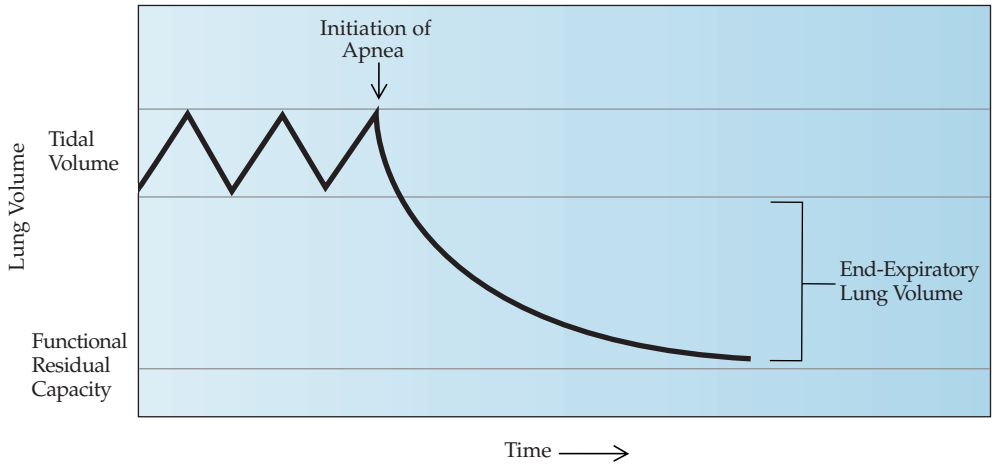


Figure 5 **The degree of pulmonary hyperinflation above the functional residual capacity can be determined by measuring the total exhaled volume during a period of apnea of 20 to 40 seconds. This allows the patient to reach the passive relaxation volume of the respiratory system. The difference between the total exhaled volume and the tidal volume represents the amount of pulmonary hyperinflation or the end-expiratory lung volume.**

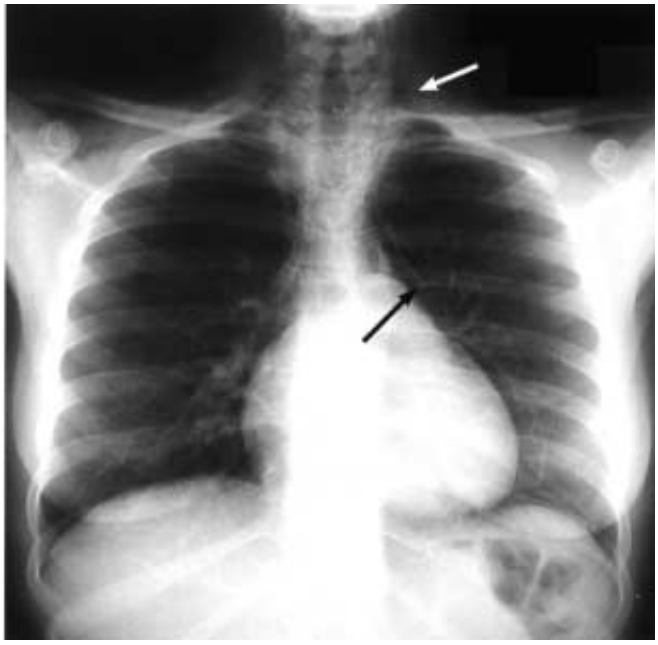


Figure 6 **Chest radiograph displays several signs of barotrauma, including subcutaneous emphysema (white arrow) and pneumomediastinum (black arrow).**

ing a tension pneumothorax from hyperinflation as the cause of hypotension may be difficult at bedside. Therefore, it is advisable to allow a period for deflation to determine whether a reduction of hyperinflation improves the blood and inflation pressure [*see Figure 5*].

Atelectasis

Atelectasis is a common cause of severe hypoxemia that develops during mechanical ventilation. Left lung atelectasis may result from intubation of the right mainstem bronchus, a problem that also may lead to overdistention of the right lung signaled by increased inflation pressures [*see Figure 7*]. A common cause of atelectasis is mucoid impaction of the bronchi. The right lung is more easily suctioned because the right mainstem bronchus follows a more direct course. The left lung is more likely to be affected by retained secretions because of the more horizontal course of the left mainstem bronchus. With atelectasis of an entire lung, breath sounds are diminished or absent on the affected side, and the trachea is shifted toward that side. A chest radiograph will reveal increased opacity in the affected hemithorax, together with ipsilateral tracheal shift and elevation of the hemidiaphragm. These findings are crucial for radiographic differentiation of whole lung atelectasis from a massive pleural effusion. Massive pleural effusion should cause the trachea to deviate away from the involved lung. A large region of atelectasis may produce a significant intrapulmonary shunt, giving rise to profound hypoxemia caused by shunting that is refractory to an increase in F_IO_2.

Atelectasis should be suspected when a sudden onset of severe, refractory hypoxemia occurs in the absence of hemodynamic instability. Other causes of profound hypoxemia, such as massive pulmonary embolism or tension pneumothorax, produce concomitant hypotension. Placing the patient in the lateral decubitus position with the atelectatic lung superior may significantly improve oxygenation, because gravity will redistribute blood flow to the dependent lung. Bronchoscopy should be performed to remove excess mucus if it cannot be easily extracted with endotracheal suctioning after chest percussion.[32]

Approach to Pulmonary Complications

Worsening respiratory distress or arterial oxygen desaturation may develop suddenly as a result of changes in the patient's cardiopulmonary status or secondary to a mechanical malfunction. The first priority is to ensure patency and correct positioning of the patient's airway so that adequate oxygenation and ventilation can be administered during the ensuing evaluation. Briefly, note ventilator alarms, airway pressures, and tidal volume. Low-pressure alarms with decreased exhaled tidal volumes may suggest a leak in the ventilator circuit. Disconnect the patient from the ventilator and manually ventilate with an anesthesia bag, using 100% oxygen. For patients receiving PEEP, manual ventilation with a PEEP valve should be used to prevent atelectasis and hypoxemia. If manual ventilation is difficult, check airway patency by passing a suction catheter through the endotracheal tube or tracheostomy tube. Additionally, listen for prolonged expiration continuing up to the point of the next manual breath. This suggests the presence of gas trapping and auto-PEEP. Check vital signs and perform a rapid physical examination, with attention paid to the patient's cardiopulmonary status. Be attentive to asymmetry in breath sounds or tracheal deviation suggesting tension pneumothorax. Note other parameters, including cardiac rhythm and hemodynamics**.** Treat appropriately on the basis of the foregoing evaluation.

Treatment should be specific to the identified problems. If the presence of gas trapping and auto-PEEP is suspected, a reduction in the minute ventilation is appropriate. In some circumstances, periods of hypoventilation (4 to 6 breaths/min) or even apnea for 30 to 60 seconds may be necessary to reverse the hemodynamic sequelae of auto-PEEP (e.g., shock and electromechanical dissociation). Return the patient to the ventilator only after checking its function. Increase the level of support provided by the ventilator to the patient after an episode of respiratory distress or arterial oxygen desaturation. Usually, this adjustment means increasing the F_IO_2 and the delivered minute ventilation unless significant auto-PEEP is present. An acute increase in the peak airway pressure usually implies either a decrease in lung compliance or an increase in airway resistance. At a minimum, considerations that should be entertained as causes of increased airway pressure include (1) pneumothorax, hemothorax, or hydropneumothorax; (2) occlusion of the patient's airway; (3) bronchospasm; (4) increased accumulation of condensate in the ventilator circuit tubing; (5) mainstem intubation; (6) worsening pulmonary edema; or (7) the development of gas trapping with auto-PEEP. Loss of tidal volume, as evidenced by a difference between the tidal volume setting and the delivered tidal volume, implies a leak in either the ventilator or the inspiratory limb of the circuit tubing. A difference between the delivered tidal volume and the expired tidal volume implies the presence of a leak at the patient's airway, from cuff malfunction, malpositioning of the airway (e.g., positioning of the cuff at or above the level of the glottis), or a leak within the patient (e.g., a bronchopleural fistula in a patient with a chest tube).

Asynchronous breathing (so-called fighting or bucking the ventilator) occurs when a patient's breathing coordinates poorly with the ventilator. This difficulty may indicate unmet respiratory demands. A careful evaluation is mandated, with attention focused at the identification of leaks in the ventilator system or

airway, inadequate F_IO_2, or inadequate ventilatory support. The problem can be alleviated by adjustments in the mode of mechanical ventilation, rate, tidal volume, inspiratory flow rate, and level of PEEP. The identification of gas trapping with auto-PEEP may require changing multiple settings to allow adequate time for exhalation (e.g., decreasing rate and tidal volume, increasing inspiratory flow rate, or switching from assist-control to SIMV in selected cases). Measures aimed at reducing the work of breathing with mechanical ventilation also may resolve the problem (addition of flow-by triggering or low levels of PSV to patients taking spontaneous breaths). If these adjustments are unsuccessful, sedation should be attempted. Muscle paralysis should be reserved for patients in whom effective gas exchange and ventilation cannot be achieved with other measures.

NONPULMONARY COMPLICATIONS

Nonpulmonary complications occur in critically ill patients either because of the natural course of the underlying disease or because of iatrogenesis. A major role of the physician is to limit the occurrences of these nonpulmonary complications by anticipating common problems and initiating specific prophylactic measures.

Venous Thromboembolism

Venous thromboembolic disease is a significant cause of morbidity and mortality in critically ill patients. On routine clinical screening, 33% of patients in medical intensive care units have deep vein thrombosis (DVT) and 18% of trauma patients have proximal DVT.[33,34] Specific independent risk factors associated with the development of venous thromboembolism include trauma, underlying malignancy, immobilization, heart failure, and obesity. Prophylaxis is recommended for all high-risk patients and has been reported to decrease the incidence of DVT by 68%.[35] The most commonly used regimen is low-dose unfractionated heparin, 5,000 units administered subcutaneously two or three times daily. Other prophylactic therapies, including intermittent pneumatic compression stockings and graded elastic stockings, can be used for patients who cannot tolerate anticoagulation. In patients with major trauma, low-molecular-weight heparin is more effective in preventing the development of proximal DVT than is prophylaxis with standard unfractionated heparin.[36] In medical patients, however, the use of low-molecular-weight heparin as prophylaxis has not been shown to provide greater benefit than standard unfractionated heparin.[37]

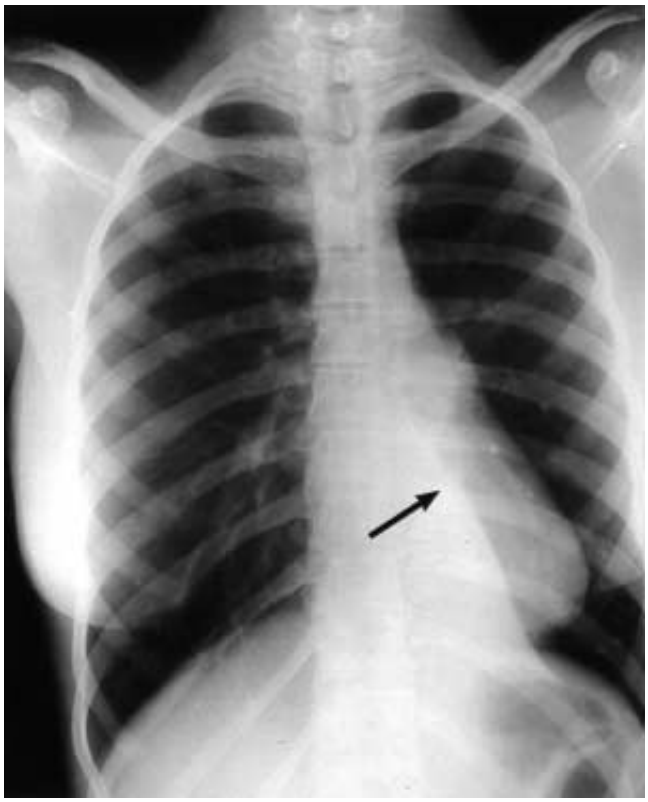


Figure 7 **Chest radiograph shows a collapsed left lower lobe of the lung (arrow).**

Gastrointestinal Bleeding

Gastrointestinal bleeding caused by stress ulceration is another important nonpulmonary complication and occurs in 1% to 10% of all critically ill patients.[38] Patients with coagulopathy, burns, head injury, or respiratory failure requiring mechanical ventilation are at increased risk for clinically significant bleeding. H_2 receptor antagonists should be administered to high-risk patients to reduce the possibility of the development of clinically significant bleeding.[39] Other therapeutic options include the cytoprotective agent sucralfate, which does not alter the gastric pH level and may be associated with a lower risk of late-onset pneumonia.[40]

Pressure Ulcers

Critically ill patients are also at an increased risk for pressure ulcers or localized areas of tissue necrosis that develop when soft tissue is compressed between a bony prominence and an external surface.[41] Pressure ulcers occur in 33% to 56% of patients in the ICU and are a source of infection that can result in bacteremia and osteomyelitis. Prevention programs that include regular repositioning of patients, which reduces the accumulation of moisture on skin, and adequate nutritional supplementation decrease the incidence of pressure sores. Air-suspension beds that redistribute body weight away from bony prominences also reduce the risk of pressure sores in certain critically ill patients.[42]

Neuromuscular Weakness

Patients who undergo mechanical ventilation may be at risk for neuromuscular weakness that persists long after the cause of respiratory failure has been resolved. A common cause of diffuse weakness is critical illness polyneuropathy, an axonal disorder that occurs with sepsis and multiorgan failure.[43] When present, critical illness polyneuropathy may be an important cause of delayed weaning from mechanical ventilation. Use of neuromuscular paralysis may be associated with weakness that persists after the neuromuscular blocking agents have been discontinued. Prolonged neuromuscular blockade can be diminished by appropriate dosing, adequate monitoring of the degree of neuromuscular blockade, and avoidance of medications that potentiate the action of specific neuromuscular blocking agents.[44] Neuromuscular blocking agents may also contribute to the development of acute myopathy, particularly in patients who receive concomitant corticosteroids. The risk of myopathy is not influenced by the chemical structure of the agent that is used to induce paralysis but, rather, is strongly correlated with the duration of paralysis.

Acute Renal Failure

Acute renal failure is a common nonpulmonary complication of patients in the ICU. Despite considerable advances in the management of critically ill patients and renal replacement therapy, the mortality associated with acute renal failure remains greater than 50%. Multiple system organ failure and other comorbidities contribute to the high mortality associated with acute renal failure, but the acute renal failure is also independently associated with an increase in morbidity and mortality.[45,46] In addition, long-term dialysis support is sometimes required for survivors of acute renal failure. Renal dysfunction in ICU patients is usually caused by intrinsic renal disease.[47] Hypotension, sepsis, the use of aminoglycosides, and volume depletion are all risk factors for the development of acute renal failure.

Withdrawal of Mechanical Ventilatory Support

To decrease complications and improve patient comfort, mechanical ventilatory support should be removed as soon as possible. However, premature withdrawal of mechanical ventilation is also associated with adverse events that may further delay appropriate extubation, such as aspiration or severe cardiopulmonary decompensation. Determination of the proper time to extubate a patient is based on several factors, including reversal or improvement of the underlying acute illness, adequate respiratory function, lack of excessive secretions, and ability to protect the airway.

Simple screening criteria have been developed that identify patients who may be ready to be removed from mechanical ventilation. These criteria include a minimal requirement for supplemental oxygenation ($F_IO_2 \leq 0.40$ or $P_aO_2/F_IO_2 < 200$), a PEEP level not exceeding 5 cm H_2O, adequate cough during suctioning, and no infusion of vasoactive agents. Measures of adequate respiratory function that are compatible with extubation have also been proposed, including a measure of rapid and shallow breathing, calculated as the respiratory rate divided by the tidal volume during spontaneous breaths. A threshold of 105 breaths/min/L, measured after 1 minute of spontaneous breathing, provides an excellent means to predict a successful extubation, with a positive and a negative predictive value of 0.78 and 0.95, respectively.[48]

Most patients can be removed rapidly from mechanical ventilation once the acute illness requiring mechanical ventilation has been reversed, and over 75% of patients who meet these initial screening criteria can be extubated successfully.[49] These patients can be quickly identified by a 30-minute to 2-hour trial of spontaneous breathing while they remain connected to the ventilator. Patients who appear comfortable after a trial of spontaneous breathing (respiratory rate < 35 breaths/min; heart rate < 140 beats/min; adequate arterial oxygen saturation; and no evidence of anxiety, diaphoresis, or extreme hypotension or hypertension) will likely be successfully removed from mechanical ventilation. Therefore, the concept of weaning, or slowly removing the patient from mechanical ventilation, does not apply to the majority of patients who require mechanical ventilation.

The remaining 25% of patients who require mechanical ventilation for acute respiratory failure cannot be removed rapidly from the ventilator.[49] There are several options for this group of patients. The three most commonly used modes of weaning are the following:

Table 4 Factors to Consider in Patients Having Difficulty Being Liberated from Mechanical Ventilation

Weaning parameters
$P_aO_2 \geq 60$ mm Hg with an $F_IO_2 \leq 50\%$
PEEP ≤ 5 cm H_2O
P_aCO_2 and pH in the acceptable range for patient's lung function
Spontaneous tidal volume ≥ 5 ml/kg
Vital capacity ≥ 10 ml/kg
Minute ventilation (MV) < 10 L/min
Maximum voluntary ventilation double of MV
Maximum inspiratory pressure (MIP) ≥ 25 cm H_2O
Respiratory rate < 30 breaths/min
Rapid shallow breathing index* < 105
Endotracheal tube
Use largest tube possible
Consider use of supplemental pressure-support ventilation
Suction secretions
Arterial blood gases
Avoid or treat metabolic alkalosis
Maintain P_aO_2 at 60 to 65 mm Hg to avoid blunting of respiratory drive
For patients with CO_2 retention, keep P_aCO_2 at or above the baseline level
Nutrition
Ensure adequate nutritional support
Avoid electrolyte deficiencies
Avoid excessive calories
Secretions
Clear regularly
Avoid excessive dehydration
Neuromuscular factors
Avoid neuromuscular depressant drugs
Avoid unnecessary corticosteroids
Airway patency
Use bronchodilators when appropriate
Exclude foreign bodies within the airway
Wakefulness
Avoid oversedation
Wean in morning or when patient is most awake

*Respiratory rate divided by tidal volume during spontaneous breaths; measured after 1 min of spontaneous breathing and calculated in breaths/min/L
F_IO_2—fraction of inspired oxygen PEEP—positive end-expiratory pressure P_aO_2—partial pressure of oxygen in arterial blood P_aCO_2—partial pressure of carbon dioxide in arterial blood

1. SIMV, which allows spontaneous breathing and diminishing numbers of mandatory breaths per minute until the patient is breathing unassisted.
2. The use of a T-piece circuit, which allows intermittent trials with total removal of mechanical support.
3. Use of decreasing levels of pressure-support ventilation.

Two large studies compared these three methods of weaning. The first one found that a gradual decrease in the level of pressure support was the most effective.[50] The second study found that a daily T-piece trial was associated with the shortest duration of mechanical ventilation.[49] The disparate conclusions of these two studies may in part be the result of differing criteria for extubation. Currently, it appears that both the gradual reduction in the level of pressure support and the use of intermittent T-piece trials are effective methods for weaning, but neither

offers a clear advantage over the other. Gradual reduction in the number of machine-supported breaths by use of SIMV appears to be the least effective method of weaning.

Several other measures can promote successful removal from mechanical ventilation. Daily screening of the respiratory function of patients receiving mechanical ventilation, followed by trials of spontaneous breathing initiated by a respiratory therapist, can reduce the duration of mechanical ventilation.[51,52] Daily interruption of sedative infusions has also been associated with decreased duration of mechanical ventilation and length of stay in the ICU.[53] In addition, it has been shown that fewer diagnostic tests to assess change in mental status are performed in patients assigned to daily interruption of sedative agents. Finally, noninvasive ventilation has been used as a technique to expedite weaning for patients with COPD.[54] Extubation and the application of noninvasive ventilation by face mask attempted after 48 hours has been associated with a significantly shorter mean duration of mechanical ventilation. For patients having difficulty being weaned from mechanical ventilation, a systematic approach to the evaluation of factors contributing to potential weaning failure is required [*see Table 4*].

Extubation

Usually, extubation should be performed early in the day, when full ancillary staff are available. The patient should be clearly educated about the procedure, the need to cough, and the possible need for reintubation. Elevation of the head and trunk to more than 30° to 45° improves diaphragmatic function. Equipment for reintubation should be available, and a high-humidity, oxygen-enriched gas source with a higher-than-current F_IO_2 setting should be available at the bedside. The patient's airway and the oropharynx above the cuff should be suctioned. The cuff of the endotracheal tube should be deflated partially, and airflow around the outside of the tube—indicating the absence of airway obstruction—should be detected. After the cuff is deflated completely, the patient should be extubated, and high-humidity oxygen should be administered via face mask. Coughing and deep breathing should be encouraged while the examiner monitors the patient's vital signs and upper airway for stridor. Inspiratory stridor may result from glottic and subglottic edema. If clinical status permits, treatment with nebulized 2.5% racemic epinephrine (0.5 ml in 3 ml normal saline) should be administered. If upper airway obstruction persists or worsens, reintubation should be performed. Extubation should not be reattempted for 24 to 72 hours after reintubation for upper airway obstruction. Otolaryngology consultation may be beneficial to exclude other causes of upper airway obstruction and to perform tracheostomy if upper airway obstruction persists.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Greene KE, Peters JI: Pathophysiology of acute respiratory failure. Clin Chest Med 15:1, 1994

2. Einarsson O, Rochester CL, Rosenbaum S: Airway management in respiratory emergencies. Clin Chest Med 15:13, 1994

3. Heffner JE: Timing of tracheotomy in mechanically ventilated patients. Am Rev Respir Dis 147:768, 1993

4. Hillberg RE, Johnson DC: Noninvasive ventilation. N Engl J Med 337:1746, 1997

5. Tobin MJ: Advances in mechanical ventilation. N Engl J Med 344:1986, 2001

6. Brochard L: Mechanical ventilation: invasive versus noninvasive. Eur Respir J Suppl 47:31s, 2003

7. Ranieri VM, Grasso S, Fiore T, et al: Auto-positive end-expiratory pressure and dynamic hyperinflation. Clin Chest Med 17:379, 1996

8. Niewoehner DE, Erbland ML, Deupree RH, et al: Effect of systemic glucocorticoids on exacerbations of chronic obstructive pulmonary disease. N Engl J Med 340:1941, 1999

9. Ram F, Picot J, Lightowler J, et al: Non-invasive positive pressure ventilation for treatment of respiratory failure due to exacerbations of chronic obstructive pulmonary disease. Cochrane Database Syst Rev (1):CD004104, 2004

10. Keenan SP, Sinuff T, Cook DJ, et al: Which patients with acute exacerbation of chronic obstructive pulmonary disease benefit from noninvasive positive-pressure ventilation? A systematic review of the literature. Ann Intern Med 138:861, 2003

11. Corbridge TC, Hall JB: The assessment and management of adults with status asthmaticus. Am J Respir Crit Care Med 151:1296, 1995

12. Passam F, Hoing S, Prinianakis G, et al: Effect of different levels of pressure support and proportional assist ventilation on breathing pattern, work of breathing and gas exchange in mechanically ventilated hypercapnic COPD patients with acute respiratory failure. Respiration 70:355, 2003

13. Ware LB, Matthay MA: The acute respiratory distress syndrome. N Engl J Med 342:1334, 2000

14. Dreyfuss D, Saumon G: Ventilator-induced lung injury: lessons from experimental studies. Am J Respir Crit Care Med 157:294, 1998

15. Gattinoni L, Caironi P, Pelosi P, et al: What has computed tomography taught us about the acute respiratory distress syndrome? Am J Respir Crit Care Med 164:1701, 2001

16. Amato MBP, Barbas CSV, Medeiros DM, et al: Beneficial effects of the open lung approach with low distending pressures in acute respiratory distress syndrome: a prospective randomized study on mechanical ventilation. Am J Respir Crit Care Med 152:1835, 1995

17. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. The Acute Respiratory Distress Syndrome Network. N Engl J Med 342:1301, 2000

18. Amato MBP, Barbas CSV, Medeiros DM, et al: Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. N Engl J Med 338:347, 1998

19. Shanholtz C, Brower R: Should inverse ratio ventilation be used in adult respiratory distress syndrome? Am J Respir Crit Care Med 149:1354, 1994

20. Lessard MR, Guerot E, Lorino H, et al: Effects of pressure-controlled with different I:E ratios versus volume-controlled ventilation on respiratory mechanics, gas exchange, and hemodynamics in patients with adult respiratory distress syndrome. Anesthesiology 80:983, 1994

21. Lamm WJE, Graham MM, Albert RK: Mechanism by which the prone position improves oxygenation in acute lung injury. Am J Respir Crit Care Med 150:184, 1994

22. Nakos G, Tsangaris I, Kostanti E, et al: Effect of the prone position on patients with hydrostatic pulmonary edema compared with patients with acute respiratory distress syndrome and pulmonary fibrosis. Am J Respir Crit Care Med 161:360, 2000

23. Cook DJ, Kollef MH: Risk factors for ICU-acquired pneumonia. JAMA 279:1605, 1998

24. Kollef MH: The prevention of ventilator-associated pneumonia. N Engl J Med 340:627, 1999

25. Cook DJ, Walter SD, Cook RJ, et al: Incidence of and risk factors for ventilator-associated pneumonia in critically ill patients. Ann Intern Med 129:433, 1998

26. Fagon JY, Chastre J, Wolff M, et al: Invasive and noninvasive strategies for management of suspected ventilator-associated pneumonia. Ann Intern Med 132:621, 2000

27. Drakulovic MB, Torres A, Bauer TT, et al: Supine body position as a risk factor for nosocomial pneumonia in mechanically ventilated patients: a randomised trial. Lancet 354:1851, 1999

28. Valles J, Artigas A, Rello J, et al: Continuous aspiration of subglottic secretions in preventing ventilator-associated pneumonia. Ann Intern Med 122:179, 1995

29. Rumbak MJ, Cancio MR: Significant reduction in methicillin-resistant *Staphylococcus aureus* ventilator-associated pneumonia associated with the institution of a prevention protocol. Crit Care Med 23:1200, 1995

30. Rouby JJ, Laurent P, Gosnach M, et al: Risk factors and clinical relevance of nosocomial maxillary sinusitis in the critically ill. Am J Respir Crit Care Med 150:776, 1994

31. Talmor M, Li P, Barie PS: Acute paranasal sinusitis in critically ill patients: guidelines for prevention, diagnosis, and treatment. Clin Infect Dis 25:1441, 1997

32. Marini JJ, Pierson DJ, Hudson LD: Acute lobar atelectasis: a prospective comparison of fiberoptic bronchoscopy and respiratory therapy. Am Rev Respir Dis 119:971, 1979

33. Hirsch DR, Ingenito EP, Goldhaber SZ: Prevalence of deep venous thrombosis among patients in medical intensive care. JAMA 274:335, 1995

34. Geerts WH, Code KI, Jay RM, et al: A prospective study of venous thromboembolism after major trauma. N Engl J Med 331:1601, 1994

35. Geerts WH, Heit JA, Clagett P, et al: Prevention of venous thromboembolism. Chest 119(suppl):1325, 2001

36. Geerts WH, Jay RM, Code KI, et al: A comparison of low-dose heparin with low-molecular-weight heparin as prophylaxis against venous thromboembolism after major trauma. N Engl J Med 335:710, 1996

37. Weitz JI: Low-molecular-weight heparins. N Engl J Med 337:688, 1997

38. Cook DJ, Fuller HD, Guyatt GH, et al: Risk factors for gastrointestinal bleeding in critically ill patients. N Engl J Med 330:377, 1994

39. Cook D, Guyatt G, Marshall J, et al: A comparison of sucralfate and ranitidine for the prevention of upper gastrointestinal bleeding in patients requiring mechanical ventilation. N Engl J Med 338:791, 1998

40. Prod'hom G, Leuenberger P, Koerfer J, et al: Nosocomial pneumonia in mechanically ventilated patients receiving antacid, ranitidine, or sucralfate as prophylaxis for stress

ulcer: a randomized controlled trial. Ann Intern Med 120:653, 1994

41. Peerless JR, Davies A, Klein D, et al: Skin complications in the intensive care unit. Clin Chest Med 20:453, 1999

42. Inman KJ, Sibbald WJ, Rutledge FS, et al: Clinical utility and cost-effectiveness of an air suspension bed in the prevention of pressure ulcers. JAMA 269:1139, 1993

43. Paresis acquired in the intensive care unit: a prospective multicenter study. Groupe de Reflexion et d'Etude des Neuromyopathies en Reanimation. JAMA 288:2859, 2002

44. Shapiro BA, Warren J, Egol AB, et al: Practice parameters for sustained neuromuscular blockade in the adult critically ill patient: an executive summary. Crit Care Med 23:1601, 1995

45. Levy EM, Viscoli CM, Horwitz RI: The effects of acute renal failure on mortality: a cohort analysis. JAMA 275:1489, 1996

46. Murray P, Hall J: Renal replacement therapy for acute renal failure. Am J Respir Crit Care Med 162:777, 2000

47. Brivet FG, Kleinknecht DJ, Loirat P, et al: Acute renal failure in intensive care units: causes, outcome, and prognostic factors of hospital mortality: a prospective, multicenter study. Crit Care Med 24:192, 1996

48. Yang KL, Tobin MJ: A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. N Engl J Med 324:1445, 1991

49. Esteban A, Frutos F, Tobin MJ, et al: A comparison of four methods of weaning patients from mechanical ventilation. N Engl J Med 332:345, 1995

50. Brochard L, Rauss A, Benito S, et al: Comparison of three methods of gradual withdrawing from ventilatory support during weaning from mechanical ventilation. Am J Respir Crit Care Med 150:896, 1994

51. Ely EW, Baker AM, Dunagan DP, et al: Effect on the duration of mechanical ventilation of identifying patients capable of breathing spontaneously. N Engl J Med 335:1864, 1996

52. Kollef MH, Shapiro SD, Silver P, et al: A randomized, controlled trial of protocol-directed versus physician-directed weaning from mechanical ventilation. Crit Care Med 25:567, 1997

53. Kress JP, Pohlman AP, O'Connor MF, et al: Daily interruption of sedative infusions in critically ill patients undergoing mechanical ventilation. N Engl J Med 342:1471, 2000

54. Nava S, Ambrosino N, Clini E, et al: Noninvasive mechanical ventilation in the weaning of patients with respiratory failure due to chronic obstructive pulmonary disease. Ann Intern Med 128:721, 1998

IX DISORDERS OF THE PLEURA, HILA, AND MEDIASTINUM

Gerald W. Staton, Jr., m.d.
Roland H. Ingram, Jr., m.d.

Pleurisy

Pleurisy (also known as pleuritis) is characterized by chest pain that results from inflammation of the pleural surfaces from any cause. The pain originates from the parietal pleura, which derives most of its innervation from the intercostal nerves. The central portion of the diaphragmatic parietal pleura receives phrenic innervation, so that inflammation of the diaphragmatic surfaces causes referred pain to the ipsilateral shoulder. Inspiratory chest pain, however, is not uniquely pleural and may occur with pericarditis or chest wall disease.

Pleuritic chest pain may result from primary involvement of the pleura by neoplasia, infection, trauma, or inflammation or from secondary spread of one of these processes from subjacent lung tissue. Thus, a common cause of pleuritic chest pain is pneumococcal pneumonia, which typically begins at the lung periphery and spreads to the adjacent pleural surfaces. Many middle-aged and elderly persons who report having experienced pleurisy in the past probably suffered from pneumonia.

diagnosis

Pleuritic chest pain is typified by intensification on deep inspiration. It is usually sharp in quality, may be present continuously, and is characteristically made worse by movements of the thorax, as well as by coughing, sneezing, or other sudden respiratory movements. If the pain is severe, inability to take a deep breath without aggravating the pain may lead to the sensation of shortness of breath. Often, pleuritic chest pain must be differentiated from chest wall pain of musculoskeletal origin (e.g., localized muscle strain, costochondritis, or rib fracture) and from the pain associated with pericarditis. Superficial tenderness of the chest wall on light palpation favors chest wall pain, but tenderness on deep palpation does not reliably exclude pleuritis. A pleural friction rub heard on auscultation of the chest establishes the presence of a pleural disorder.

A chest radiograph can be useful in suggesting the underlying cause of pleuritic pain. If the chest radiograph shows the presence of effusions, the causes of pleuritic pain are limited to the processes that produce pleural effusion [*see* Pleural Effusion, *below*]. If the chest radiograph is normal, pleuritic chest pain has a relatively limited differential diagnosis. Major etiologic possibilities include pulmonary embolism, viral pleurisy, and serositis in association with collagen vascular disease, especially systemic lupus erythematosus (SLE). Less common causes of pleuritic chest pain include uremia, sickle cell crisis, and pleuropericarditis that occurs after myocardial infarction or pericardiotomy (Dressler syndrome). Some patients present with an acute illness characterized by low-grade fever, headache, and myalgia; in many such episodes, the cause cannot be determined but is presumed to be viral. These acute illnesses are usually self-limited, resolving within a few days to 1 to 2 weeks. Coxsackievirus B and other enteroviruses cause an epidemic form of viral pleuritis called epidemic pleurodynia, or Bornholm disease, which often affects multiple family members. Relapses may occur after the patient is asymptomatic for several days.

treatment

Idiopathic or viral pleuritis can be treated effectively with nonsteroidal anti-inflammatory drugs and, if necessary, narcotic analgesics. However, the diagnosis is one of exclusion, and pulmonary embolism, a potentially lethal condition, is the most important cause to exclude.

Pleural Effusion

Pleural effusion, the abnormal accumulation of liquid in the pleural space, may affect as many as 800,000 persons in the United States each year. The most common causes are congestive heart failure, malignancy, pneumonia, and pulmonary emboli.

pathophysiology

A number of factors favor pleural effusion, including (1) altered permeability of the pleural membranes; (2) decreased intravascular oncotic pressure and, once a pleural effusion has formed, increased pleural liquid oncotic pressure; (3) increased hydrostatic pressure in the pleural capillaries as a result of heart failure; (4) greater negativity of pressure in the pleural space (e.g., if the lung is unable to expand normally); (5) lymphatic obstruction; (6) migration of ascitic liquid across the diaphragm; and (7) migration of pulmonary edema liquid across the visceral pleura.[1] Pleural effusion produces a restrictive defect that is correlated with the size of the effusion. Because both the air spaces and the pulmonary circulation are compressed and because of pulmonary hypoxic vasoconstriction, there is little shunting and only mild hypoxemia. Removal of a large effusion can result in modest improvement in lung function, but often, the underlying cause of the effusion (e.g., heart failure or lymphangitic carcinoma) causes persistent functional abnormalities. Various mediators are involved in the production of altered permeability and the evolution of pleural effusions.[2,3]

classification and etiology

Various types of liquid may accumulate in the pleural space. Accumulation of serous liquid is referred to as a hydrothorax. If blood accumulates, the condition is referred to as a hemothorax. An effusion composed of lipids is known as a chylothorax. Accumulation of pus is known as a pyothorax, or empyema. Although imaging studies can be helpful, the actual condition can be distinguished only by analysis of the liquid itself.

Hydrothoraces fall into two major categories on the basis of mechanisms of pleural liquid accumulation: transudation and exudation. Transudation of liquid into the pleural space occurs when there is an imbalance between the hydrostatic and the oncotic pressures governing the normal rates of pleural liquid formation and resorption. The most common cause of a transudative pleural effusion is congestive heart failure. Other causes are constrictive pericarditis, superior vena cava obstruction, and the hypoalbuminemic states associated with cirrhosis and the nephrotic syndrome [*see Table 1*]. Pulmonary arterial hypertension and right heart failure usually do not by themselves cause pleural effusions.

Table 1 Causes of Hydrothorax

Transudative
- Congestive heart failure
- Constrictive pericardial disease
- Cirrhosis
- Nephrotic syndrome
- Superior vena cava obstruction
- Ascites (transudative)
- Peritoneal dialysis

Exudative or Transudative
- Hypothyroidism
- Pulmonary embolism
- Trapped lung

Exudative
- Infections (i.e., parapneumonic effusion, tuberculosis)
- Malignant disorders
  - Primary lung cancer
  - Cancer metastases to the lungs or pleura
  - Lymphoma
  - Mesothelioma
- Collagen vascular diseases and vasculitides
- Gastrointestinal diseases
  - Pancreatitis and pancreatic pseudocyst
  - Esophageal rupture
  - Abdominal or retroperitoneal abscess
  - Postabdominal surgery
  - Postendoscopic variceal sclerotherapy

Miscellaneous
- Benign asbestos effusion
- Meigs syndrome
- Dressler syndrome (after myocardial infarction or pericardiotomy)
- Post–coronary artery bypass[97]
- Uremia
- Sarcoidosis and necrotizing sarcoid granulomatosis
- Radiation therapy
- Drugs (e.g., nitrofurantoin, dantrolene, methysergide, all-*trans*-retinoic acid[98])
- Yellow nail syndrome

Exudation of liquid into the pleural space results from any process that disrupts the integrity of the endothelial membrane that lines the pleural capillaries and venules. Obstruction of lymphatic drainage from the pleural space is another mechanism that can cause a protein-rich effusion. Exudative effusions are associated with a broad range of disorders, including a variety of infectious, neoplastic, inflammatory, embolic, and vasculitic diseases. In addition, exudative effusions may be caused by the effects of certain drugs and physical agents [*see Table 1*].

DIAGNOSIS

Clinical Manifestations

A pleural effusion may be suspected on the basis of physical examination. Physical findings include dullness on percussion, diminished or absent breath sounds, decreased fremitus, and egophony at the level of the pleural liquid meniscus.

Imaging Studies

The chest radiograph usually establishes the diagnosis; patients are often further evaluated by other imaging techniques (e.g., ultrasonography and computed tomography).[4]

Radiography When the patient is in the upright position, liquid collects first in the posterior sulcus, the most inferiorly located recess of the pleural space. Blunting of the normally sharp posterior costophrenic angle on a lateral chest radiograph indicates the presence of at least 25 to 50 ml of pleural liquid. As additional liquid accumulates (approximately 150 ml total), the lateral costophrenic angle on a posteroanterior radiograph becomes obliterated. Greater amounts of pleural liquid displace the lung centrally and produce a characteristic homogeneous opacity that forms a concave meniscus with the chest wall [*see Figure 1*].

A massive pleural effusion may opacify an entire hemithorax and displace mediastinal structures to the opposite side of the chest [*see Figure 2*]. Large effusions may reduce venous return and thus cardiac output, creating hemodynamic compromise. The displacement force that is exerted is proportional to the height of the effusion. A contralateral shift of the mediastinum in a patient with massive pleural effusion may go undetected if there is ipsilateral atelectasis of the lung or if the mediastinum is fixed by an invasive tumor or fibrosis. A mediastinal shift toward the side of the effusion indicates almost complete atelectasis of the underlying lung, most often resulting from an obstructing tumor of the mainstem bronchus.

When intense pleural inflammation leads to the formation of adhesions between the visceral and parietal surfaces, localized effusions may collect in the resultant pockets. These loculated effusions may form along any part of the pleural surface and at times may be mistaken for infiltrates or masses within the lung parenchyma [*see Figure 3*]. A radiographic sign that favors the diagnosis of pleural loculation is the oblique angle formed by the chest wall and the margin of the pleural density. A subpleural lung nodule, in contrast, usually forms an acute angle with the chest wall.

An atypical presentation of pleural effusion is the subpulmonary collection of liquid [*see Figure 3*]. For unknown reasons, sizable amounts of pleural liquid sometimes collect between the diaphragm and the base of the lung without significantly distorting the contour of the inferior lung margin. In a patient with a

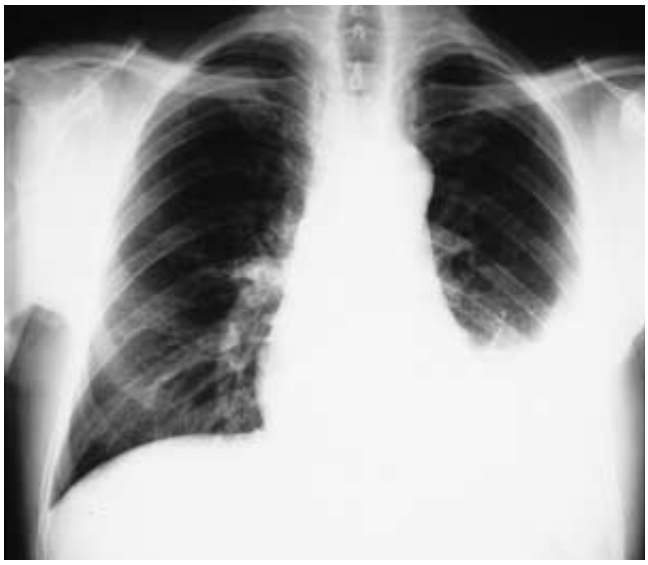


Figure 1 **This chest radiograph demonstrates the typical configuration of a pleural effusion as seen on posteroanterior projection. The pleural liquid forms a meniscus with the left lateral chest wall.**

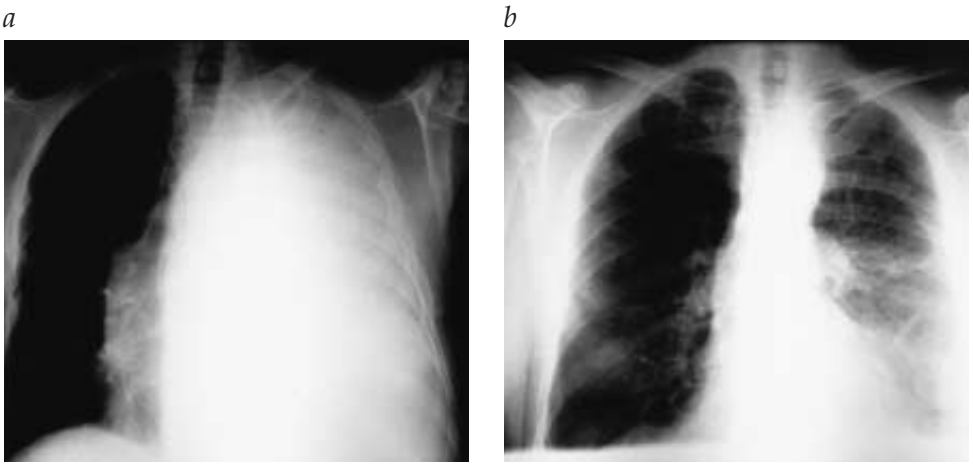


Figure 2 **(*a*) A massive hydrothorax resulting from a subpleural adenocarcinoma of the lung with a malignant pleural effusion is present in this patient. The heart and mediastinum are deviated to the right. Compression of the heart and increased intrathoracic pressure cause a reduction in venous return and resulting tachycardia (heart rate, 120 beats/min). (*b*) The heart and mediastinum return to the midline after evacuation of liquid; heart rate falls to 80 beats/min.**

subpulmonary effusion, it is possible to mistake the superior border of the effusion for the diaphragmatic silhouette, prompting a needless search for causes of an elevated hemidiaphragm. Several radiographic findings suggest the correct diagnosis. First, the contour of the base of the lung is slightly altered by a subpulmonary effusion; the normal domelike curve formed by the diaphragm is replaced by a hockey stick–like shape, with lateralization of the apex of the dome. Second, on the left side, the distance between the base of the lung and the gastric gas bubble is increased. The diagnosis of a subpulmonary effusion is confirmed by a lateral decubitus chest radiograph, which will show layering of the effusion along the lateral chest wall.

In patients with congestive heart failure, pleural liquid may collect between the two visceral pleural surfaces that line the interlobar fissures [*see Figure 4*]. Interlobar effusions, which may be mistaken for tumors of the lung parenchyma, disappear with effective treatment of the heart failure and thus have been called vanishing tumors, or pseudotumors. Interlobar effusions have a characteristic lenticular shape, with the tapered ends oriented in the plane of the fissure.

In some cases of pleural effusion, the only radiograph available for interpretation is an anteroposterior chest radiograph obtained with the patient in the supine position. In this position, a free-flowing pleural effusion spreads along the posterior costal surfaces. The radiograph shows a uniform increase in the opacity of the involved hemithorax, but normal lung markings are visible through the opacity [*see Figure 5*].

On lateral decubitus radiographs, small amounts of pleural liquid can be identified if the patient is positioned carefully. Free-flowing effusions can be distinguished from loculated effusions or thickened pleural tissue by comparing the pleural density with its appearance on a radiograph taken with the patient in the upright position.

Ultrasonography Ultrasonography is also a very sensitive test for detecting pleural effusions. It is particularly effective in cases of loculated pleural effusions because it can identify the precise site on the chest wall where a needle can be introduced to aspirate a sample.

Computed tomography CT scanning is helpful in distinguishing a pleural mass from a loculated effusion and distinguishing a hydropneumothorax from a lung abscess. CT scanning performed with contrast usually shows parietal pleural thickening in patients with exudates and enhancement of the parietal pleura in patients with empyema.[5]

Laboratory Studies

Examination of pleural liquid is useful for diagnosis. Thoracentesis should be performed in virtually all patients with pleural effusions of a significant size and of uncertain etiology. Pneumothorax is relatively common but rarely serious, so routine chest radiography is not necessary after uncomplicated thoracentesis.[6] Occasionally, pulmonary edema develops acutely in the reexpanded lung after rapid removal of large pleural effusions, especially if they have been present for a long time [*see 14:X Pulmonary Edema*]. This complication can be avoided by re-

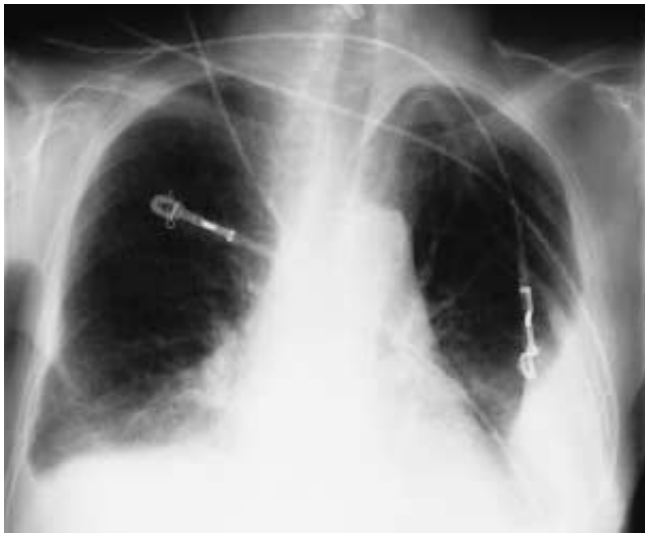


Figure 3 **This chest radiograph demonstrates two patterns of atypical pleural liquid accumulation. In the left pleural space, a loculated effusion has formed along the lateral chest wall; it is presumably trapped in this position by fibrous adhesions between adjacent areas of visceral and parietal pleurae, preventing free flow of the liquid to the most dependent portions of the pleural space. In the right pleural space, liquid has accumulated in a subpulmonic position, between the base of the lung and the superior surface of the diaphragm. The gas-liquid interface has a shape slightly different from the normal contour of the diaphragmatic silhouette: it is more flattened and curves inferiorly at a point more laterally displaced. A small meniscus sign at the lateral chest wall hints at the presence of this subpulmonic effusion.**

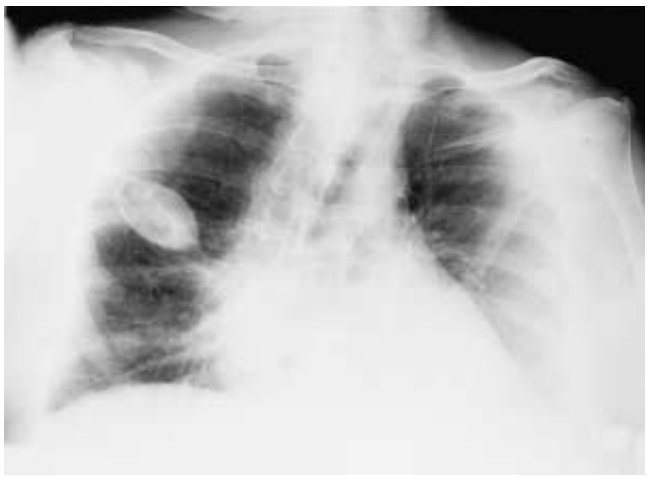


Figure 4 **This anteroposterior radiograph was made in a patient with congestive heart failure. The ovoid or lenticular opacity in the right upper lung zone is an interlobar effusion collected in the minor fissure; such effusions are sometimes mistaken for tumors of the lung parenchyma. Interlobar effusions resolve with treatment of the heart failure; hence, they are sometimes called vanishing tumors, or pseudotumors. The radiograph also demonstrates the presence of osteopenia and scoliosis in this patient.**

moving a limited amount of liquid (1,000 to 1,500 ml) during each drainage procedure or by terminating the procedure if highly negative pressures are required to remove additional liquid.[7,8] The gross appearance of pleural liquid may occasionally provide useful diagnostic information. Gross pus or putrid-smelling liquid is diagnostic of infection. A highly viscous clear liquid reflects the presence of hyaluronic acid elaborated by a mesothelioma, whereas a chocolate-sauce or anchovy-paste appearance suggests a hepatopleural fistula caused by amebiasis.

Three general categories of laboratory tests have been distinguished: (1) those that provide specific information, such as cytologic studies and microbiologic cultures and stains; (2) certain tests that, if the results are abnormal, indicate a limited number of diagnostic possibilities, such as measurements of amylase, glucose, pH (measured by a pH meter, not by dipstick), triglycerides, and antinuclear antibody (ANA) titers and differential cell counts; and (3) those that are useful in distinguishing between transudates and exudates, such as measurements of lactate dehydrogenase (LDH), protein, and cholesterol. Routine laboratory studies pertinent to most cases include total and differential white cell counts, microbiologic stains and cultures, cytologic analysis, pH measurement, and determinations of protein, LDH, amylase, and glucose levels. Measurement of pleural liquid triglycerides or ANA titers may be appropriate in certain patients with pleural effusions. When pleural fluid appears sanguineous, red cell counts may be useful.

Clinical and radiographic findings may indicate whether a pleural effusion is transudative or exudative. Patients who have painful effusions, pleural rubs, loculation of effusion on chest radiograph, or conditions known to produce pleural inflammation (e.g., pneumonia or active lupus) are more likely to have pleural exudates. Patients with cirrhosis and ascites, heart failure, or hypoproteinemia with bilateral free effusions are likely to have transudates. Clinical information may be helpful in designing strategies for pleural liquid analysis [*see Figure 6*].

If a pleural effusion is not pus, chyle, or blood, it is the clinical context that allows categorization as an exudate or transudate. In fact, the gold standards that are used to assess the sensitivities and specificities of the various pleural liquid measurements for differentiating exudates from transudates are based solely on clinical data. Pleural effusions resulting from exudation or obstruction of lymphatic drainage typically have a high protein concentration, usually 3 g/dl or greater. However, four criteria differentiate exudative from transudative pleural effusions more effectively than the absolute protein concentration[9]: (1) a ratio of pleural protein to serum protein greater than 0.5, (2) a ratio of pleural LDH to serum LDH greater than 0.6, (3) a pleural LDH concentration greater than two thirds the upper limit of normal for serum LDH, or (4) a pleural cholesterol level of 60 mg/dl or higher.[10] If one or more of these criteria are met, the effusion is usually exudative or caused by lymphatic obstruction. If none of these characteristics are present, a transudative mechanism would be expected. Although these criteria are extremely useful, they are not absolute, and the results must be considered in the clinical context.

No additional information about the etiology of a transudative pleural effusion can be obtained by further testing of the pleural liquid. However, results of certain laboratory studies can be used to narrow the differential diagnosis of exudative pleural effusions.

Amylase level Determination of the pleural liquid amylase level is warranted in patients with unexplained left-sided pleural effusions, particularly in the presence of coexistent abdominal disease. In patients who have pleural effusions associated with acute pancreatitis or pancreatic pseudocysts, the pleural liquid amylase level typically exceeds serum levels and remains elevated long after the concentration of amylase in the serum has returned to normal. Elevated amylase levels are also commonly seen in patients with malignancy.[11]

Red cell count Most bloody pleural effusions in which the red cell count is greater than 100,000/mm^3 are caused by malig-

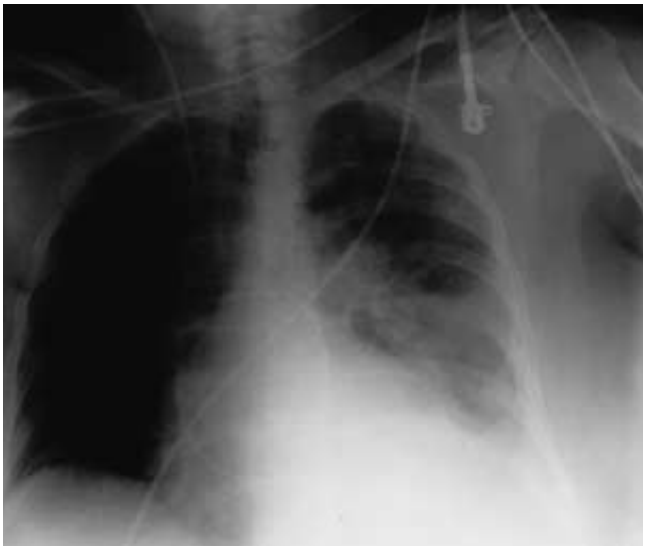


Figure 5 **In a chest radiograph obtained with the patient in the supine position, a pleural effusion may be apparent as only a diffusely increased opacity or haze over the affected hemithorax, as seen in the left pleural space of this patient. In this position, the effusion uniformly layers along the posterior chest wall; the x-ray beam penetrates perpendicular to the thin pleural liquid layer, producing a diffuse haze without a discrete meniscus or gas-liquid interface.**

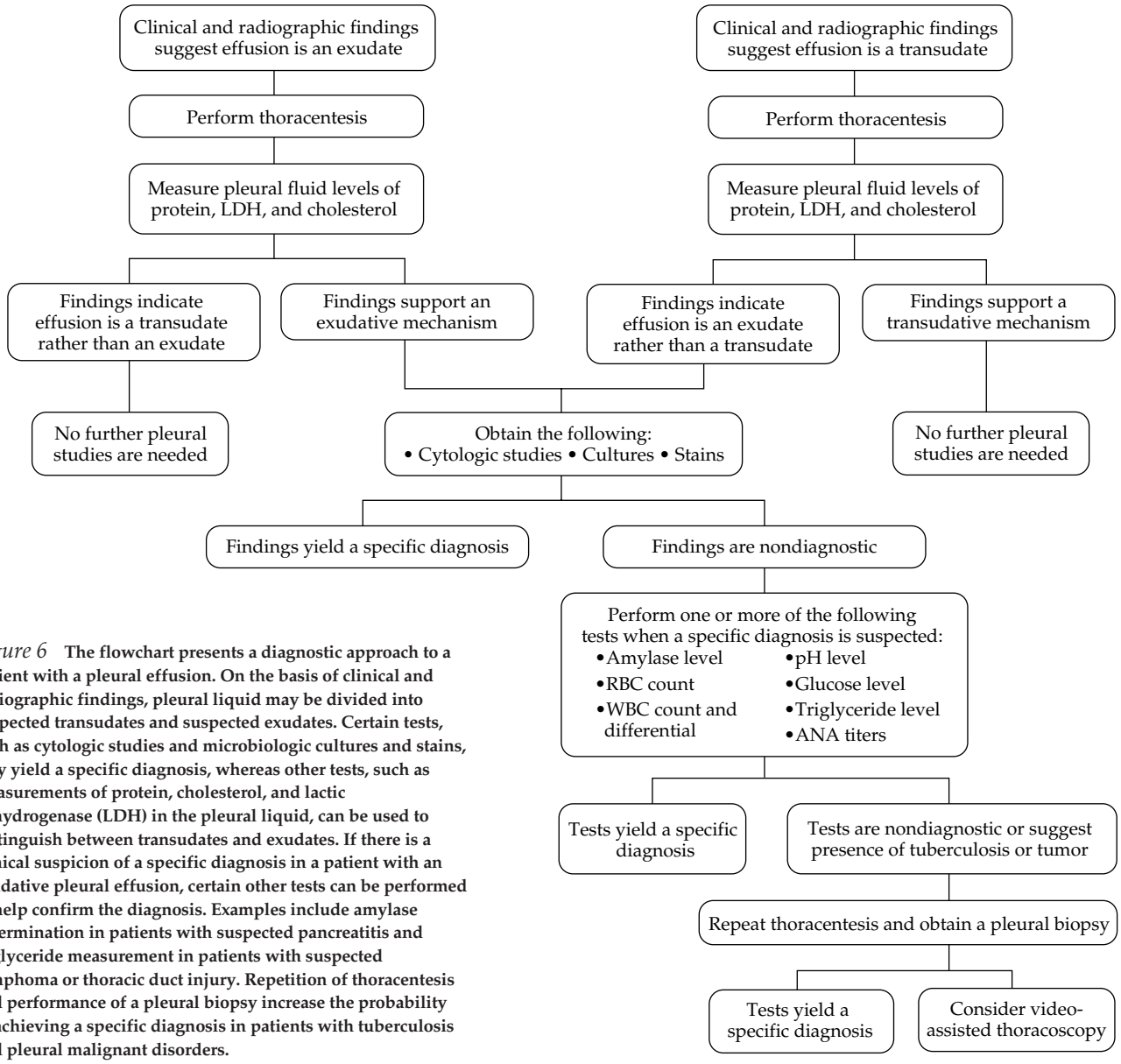


Figure 6 **The flowchart presents a diagnostic approach to a patient with a pleural effusion. On the basis of clinical and radiographic findings, pleural liquid may be divided into suspected transudates and suspected exudates. Certain tests, such as cytologic studies and microbiologic cultures and stains, may yield a specific diagnosis, whereas other tests, such as measurements of protein, cholesterol, and lactic dehydrogenase (LDH) in the pleural liquid, can be used to distinguish between transudates and exudates. If there is a clinical suspicion of a specific diagnosis in a patient with an exudative pleural effusion, certain other tests can be performed to help confirm the diagnosis. Examples include amylase determination in patients with suspected pancreatitis and triglyceride measurement in patients with suspected lymphoma or thoracic duct injury. Repetition of thoracentesis and performance of a pleural biopsy increase the probability of achieving a specific diagnosis in patients with tuberculosis and pleural malignant disorders.**

nant disorders, injuries to the chest, or pulmonary embolism. Other possible causes include tuberculous pleurisy, esophageal rupture, pancreatitis, and benign asbestos effusions. A pleural liquid hematocrit that exceeds half the simultaneous peripheral blood hematocrit indicates frank bleeding into the pleural space and is diagnostic of a hemothorax. On the other hand, the finding that pleural liquid is blood tinged is not diagnostically useful, because as little as 1 or 2 ml of blood added to 1 L of fluid results in a red cell count of 5,000 to 100,000/mm^3 and a serosanguineous appearance. Serosanguineous effusions may be found in transudative and exudative effusions of many different causes.

White cell count Total and differential white cell counts may further narrow the differential diagnosis. Pleural liquid white cell counts exceeding 50,000 to 100,000/mm^3 are usually associated with grossly visible pus; such effusions are by definition pyothoraces or empyemas. Absolute white cell counts below this range cannot be used to differentiate infected pleural effusions from other inflammatory processes. Likewise, a predominance of polymorphonuclear leukocytes in the differential white cell count simply indicates an acute inflammatory process; the etiology may be empyema, parapneumonic effusion, pulmonary embolism, pancreatitis, viral pleuritis, benign asbestos effusion, malignant disease, early tuberculous effusion, or other, less common disorders.

More than 90% of patients with exudates containing a predominance of small lymphocytes have a malignant or tuberculous effusion. Pleural liquid eosinophilia is rarely the result of a fungal or parasitic infection. Much more commonly, the eosinophilia is a nonspecific finding; in some cases, it is thought to result from prior introduction of air or blood into the pleural space. The finding of more than 10% eosinophils in the pleural liquid is relatively uncommon in effusions associated with tuberculosis or malignant disease but does not exclude these diagnoses.[12] Finally, the presence of mesothelial cells in an exudative pleural effusion may be of diagnostic utility because tuberculous pleural effusions, except perhaps at their very onset, almost never contain more than 5% mesothelial cells.

pH level In congestive heart failure, the pH of the pleural liquid is very close to that of the serum (7.35 to 7.45), and in patients who also have acidemia, the pleural liquid pH falls below 7.35 as the serum pH falls. Certain exudative pleural effusions are acidotic (pH < 7.3), presumably because of metabolism of glucose to carbon dioxide and fixed acids (e.g., lactic acid). A pleural liquid pH of less than 7.3 in the absence of systemic acidosis suggests one of the following diagnoses: empyema, malignant disease (usually associated with a pleural effusion of several months' duration), collagen vascular disease, tuberculosis, esophageal rupture, or hemothorax. In patients who have a pleural effusion associated with bacterial pneumonia (parapneumonic effusion), a pleural liquid pH of less than 7.0 is suggestive of an infected pleural space (empyema).[13] A pH of 6.0 or less suggests esophageal rupture.

Glucose level Low pleural liquid glucose values (60 mg/dl) are found in tuberculous effusions, parapneumonic effusions and empyemas, and large or highly cellular effusions associated with malignant disorders. In these conditions, consumption of glucose by microorganisms or by inflammatory or malignant cells is thought to be the cause of the low glucose concentrations. Glucose values are extremely low in effusions associated with rheumatoid arthritis caused by impaired transport of glucose into the pleural space.

ANA and rheumatoid factor titer Identification of lupus erythematosus cells or an ANA titer equal to or greater than that found in the serum is diagnostic of SLE as the cause of the pleural effusion. In contrast, latex fixation titers of rheumatoid factor as high as 1:160 may be found in the pleural liquid of patients who have a pleural effusion related to a variety of conditions other than rheumatoid arthritis.[14]

Repetition of pleural liquid analysis and pleural biopsies increases the probability of achieving a diagnosis in patients with tuberculosis or pleural malignant disorders. When tuberculosis or malignancy is still suspected, thoracoscopy can be used to visualize the pleura and to obtain biopsies of visible lesions.[15] Follow-up in the clinic is indicated in the minority of patients in whom exhaustive evaluation has failed to yield a specific diagnosis. In most of these patients, the disease will have a benign course, with resolution of the effusion occurring over several months.[16]

HYDROTHORAX

Hydrothorax Caused by Congestive Heart Failure

Pleural liquid accumulation may be a relatively early sign of congestive heart failure. A portion of the liquid comes directly from the lung into the pleural space.[1] Radiographic evidence of pleural effusion (i.e., blunting of the costophrenic angles) will appear during the stage of interstitial pulmonary edema formation. Pulmonary venous hypertension, especially in combination with elevated systemic venous pressures, favors the development of transudative pleural effusions. However, at any given level of venous pressure elevation, the likelihood of pleural liquid formation varies widely among individuals; the reason for this variability remains unexplained.

Most often, pleural effusions caused by congestive heart failure form either bilaterally or on the right side only. Successful treatment of congestive heart failure usually leads to clearing of the associated effusions within hours to a few days. The protein concentration of these effusions increases slightly after diuresis, and the resultant protein concentration is occasionally in a range that would be confused with a pleural exudate.[17] In most patients with heart failure, frankly exudative effusions are caused by a process other than congestive heart failure.[18]

Ascites

Liquid that collects in the abdomen from cirrhosis,[19] ovarian tumors, or peritoneal dialysis may migrate into the pleural space. In all these conditions, ascitic liquid is usually evident clinically, and there is a preponderance (60% to 70%) of right-sided effusions. The pressure gradient between the peritoneal cavity (supra-atmospheric pressure) and the pleural cavity (subatmospheric pressure) favors this direction of flow. The transdiaphragmatic migration of liquid appears to occur via lymphatic vessels or physical openings in the diaphragm. Thoracentesis is recommended to establish the diagnosis of hepatic hydrothorax, because a significant proportion of the patients with pleural effusion will have alternative diagnoses (e.g., infection, malignancy).[20] Symptomatic liquid collection may be controlled by diuresis in cirrhosis and by the resection of benign ovarian tumors when these are associated with accumulations in the chest. Repetition of therapeutic thoracentesis is discouraged because this procedure results in protein and volume depletion. A preferable approach to the management of hepatic hydrothorax is to use video-assisted thoracoscopy to close the defects in the diaphragm, which may result in the resolution of the effusions.[21]

Hydrothorax Caused by Malignant Disease

Neoplastic invasion of the pleura, usually involving both the visceral and the parietal surfaces, can cause a pleural effusion.[22] The effusion most often has the characteristics of an exudate. Mediastinal obstruction of lymphatic channels by a tumor frequently contributes to the pathogenesis of pleural liquid formation.

The lung and breast are the most common primary sites from which pleural metastases arise; stomach and ovarian carcinomas are next in frequency. Breast, stomach, and ovarian cancers usually spread to the pleura indirectly from hepatic metastases, although contiguous spread through the chest wall occasionally occurs with breast cancer and spread across the diaphragm occasionally occurs with ovarian or stomach cancer.

Virtually any carcinoma can metastasize to the pleura, and in some cases, pleural carcinomatosis is found with no identifiable primary site.[23] Lymphomas can also cause pleural effusions; the mechanism usually involves infiltration of mediastinal lymph nodes and consequent obstruction of lymphatic drainage from the pleural cavity.

Pleural effusions are a common manifestation of primary pleural neoplasms, such as mesothelioma,[24] a rare neoplasm that can be either localized and benign or diffuse and malignant. Risk factors for the malignant form include exposure to asbestos, zeolite (erionite), and therapeutic radiation. Survival of patients with malignant mesothelioma is poor with all forms of therapy.

Diagnosis Pleural effusions associated with malignant disease typically produce the following pleural liquid findings: (1) the pleural liquid is serosanguineous; (2) protein and LDH concentrations are elevated, characteristic of an exudative effusion; (3) the differential white cell count reveals more than 50% lymphocytes; (4) the pleural liquid glucose level is normal (> 60 mg/dl); and (5) the pleural liquid pH is greater than 7.3. However, none of these findings is consistently present.

Two clinical circumstances in which malignant disease is the likely cause of pleural effusion are (1) massive pleural effusions, in which the pleural liquid opacifies an entire hemithorax,[25] and (2) effusions in which the etiology eludes thoracentesis and closed pleural biopsy.

In a majority of cases of malignant pleural effusion, the diagnosis can be made by cytologic examination. Routine studies include Papanicolaou stain of the centrifuged material and hematoxylin-eosin stain of the paraffin-embedded cell pellet (so-called cell block). When a lymphoma is suspected as the cause of a lymphocytic pleural effusion, special studies of cell surface markers can be of use in differentiating a monoclonal lymphocytic process from a polyclonal inflammatory reaction. Other immunohistochemical studies can assist in the diagnosis of carcinoma or mesothelioma.[26] The greater the amount of material provided for analysis, the higher the diagnostic yield of malignant cells; thus, the number of positive diagnoses increases with the second and third diagnostic thoracenteses. Closed pleural biopsy, which provides samples of parietal pleura 1 to 2 mm in diameter, will identify only an additional 7% of cases of pleural malignancy when repeated thoracenteses have been negative. Image-guided percutaneous biopsy may provide a better yield than traditional closed pleural biopsy.[27] Fiberoptic bronchoscopy is useful in cases that present with clinical features suggestive of bronchogenic carcinoma, such as hemoptysis, a mass, an infiltrate, or atelectasis seen on radiographs.[28] Thoracoscopy is needed for diagnosis in a minority of cases.

Treatment Management of symptomatic pleural effusions should be directed at draining the pleural liquid and, when necessary, obliterating the pleural space. In terminally ill patients with slowly reaccumulating pleural effusions, repetitive thoracenteses may be sufficient to control dyspnea for a few weeks to months.[29] Often, however, a longer-term solution to this problem is needed, in which case chemical pleurodesis may be indicated. This technique requires that the pleural space be drained by a chest tube until a minimal amount of liquid remains, allowing apposition of the visceral and parietal surfaces. Small (14 French) tubes can be used to drain the pleural space, and the procedure can be performed on selected ambulatory patients.[30] In patients with multiloculated malignant effusions, use of an intrapleural fibrinolytic agent may help achieve adequate drainage.[31] Once sufficient liquid is removed, a sclerosing agent (e.g., doxycycline, talc slurry, bleomycin, or iodopovidone[32]) is introduced into the pleural space through the chest tube; during healing of the resultant inflammatory process, a fibrous bond forms between the two pleural surfaces.[33]

Talc is the least costly sclerosing agent available for pleurodesis, and the results with talc are no different from those with doxycycline and bleomycin.[34] When pleurodesis is successful, pleural sclerosis prevents further accumulation of pleural liquid or limits accumulation of liquid to small pockets between pleural adhesions, and the patient experiences remarkably little compromise of ventilatory function. However, pleurodesis will be ineffective if the pleural liquid pH is less than 7.3 or if the lung does not expand to fill the space when pleural liquid is removed. Talc poudrage performed during thoracoscopy is an alternative and possibly more effective technique than pleurodesis.[35] In patients who do not respond to pleurodesis or talc poudrage, implantable catheters or pleuroperitoneal shunts may be useful in controlling pleural effusions.[36,37]

Hydrothorax Caused by Tuberculosis

Pleural effusion is more often a manifestation of primary tuberculosis than of reactivation tuberculosis. Tuberculous effusions occur more commonly in young adults than in children. Typically, the effusion develops 3 to 6 months after initial exposure to *Mycobacterium tuberculosis*. A tuberculin skin test is positive at the time of clinical presentation in 70% of patients; in the remaining 30% of patients, the skin test becomes positive within 6 weeks after an initial negative result. In many cases, pleural effusion is the only abnormality detected on chest radiograph. Tuberculous pleural effusions are thought to be caused by the rupture of a small subpleural focus of infection into the pleural space, with the discharge of tuberculoprotein and viable tubercle bacilli. Widespread granulomatous infection of the parietal and visceral pleurae ensues, as indicated by the high yield of closed pleural biopsy in this setting. In patients with primary tuberculosis, untreated pleural effusions resolve spontaneously in approximately 2 to 4 months. However, active tuberculosis develops in two thirds of such patients during the ensuing 5 years.

Diagnosis Tuberculous pleuritis in patients with primary tuberculosis may present acutely as a febrile illness accompanied by pleuritic chest pain or subacutely as anorexia, weight loss, and dyspnea on exertion. The pleural liquid is usually serous or serosanguineous. In most cases, the differential white cell count reveals a lymphocytosis. A finding of more than 10% eosinophils or more than 5% mesothelial cells suggests a diagnosis other than pleural tuberculosis. In approximately 20% of cases, the pleural liquid glucose level drops below 60 mg/dl and the pleural liquid pH below 7.3. Increased levels of adenosine deaminase and interferon gamma distinguish tuberculous effusions from those of other causes.[38] Use of polymerase chain reaction techniques to analyze pleural liquid for detection of mycobacterial DNA may become the method of choice for identifying tuberculous effusions.[38,39]

Acid-fast bacilli are rarely seen in pleural liquid, and cultures are positive in only 20% to 40% of patients. However, closed-needle biopsy of the pleura reveals caseating or noncaseating granulomas in approximately 70% of cases and provides material that is culture positive in approximately 75% of cases. Thus, the total diagnostic yield, based on histopathology and culture, is 90% to 95%.

Treatment Patients who have pleural effusions from primary tuberculosis should be treated with at least two antituberculous drugs.[40] Treatment schedules should be identical to those used for parenchymal pulmonary tuberculosis. Corticosteroids have not been found to be of benefit[41] [*see 7:VIII Infections Due to Mycobacteria*]. Chest tube drainage is not necessary. On the other hand, in cases of reactivation tuberculosis, infection may spill directly into the pleural space, causing a thick, purulent effusion known as tuberculous empyema. In these cases, chest tube drainage is required, as it is for other forms of empyema.

Hydrothorax Caused by Pulmonary Embolism

Pleural effusion may develop in patients with pulmonary embolism, even though pulmonary infarction (as manifested by a parenchymal infiltrate on chest radiograph) may not be present. In a large series of patients with emboli documented by angiography, 28% of the patients had radiographic evidence of pleural effusion at presentation.[42] Spiral CT scans used for the diagnosis of pulmonary embolism detect pleural effusions in 57% of cas-

es.[43] The precise pathogenetic mechanism for pleural liquid formation in the absence of lung infarction is not known, although increased capillary permeability in the lung and visceral pleura and systemic venous hypertension have been suggested as contributing factors.

Diagnosis In many cases of pleural effusion caused by pulmonary embolism, pleural liquid analysis reveals an exudative effusion that is serosanguineous in appearance (> 10,000 red cells/mm^3). A differential white cell count typically reveals a predominance of polymorphonuclear leukocytes. However, the characteristics of such an effusion are highly variable. The pleural liquid may be serous, may have a low protein concentration and meet the diagnostic criteria for a transudate, and may contain more than 50% lymphocytes. In addition, the total white cell count may vary from less than 100/mm^3 to more than 50,000/mm^3. Thus, none of the findings on routine pleural liquid analysis can be used to exclude a diagnosis of pulmonary embolism.[44]

Treatment Management of pulmonary embolism need not be modified because of the presence of a pleural effusion, even if the effusion is bloody. In this setting, pleural effusions are maximal at their onset and gradually resorb with time, although resolution often requires more than 7 days if the effusion is large and associated with pulmonary consolidation. An enlarging effusion or the subsequent development of a contralateral effusion should raise suspicion of recurrent embolization or another complication, such as empyema.

HEMOTHORAX

Direct hemorrhage into the pleural space most commonly results from trauma to the thorax. The trauma may be either blunt (e.g., sustained in a motor vehicle accident), in which case rib fractures are usually present, or penetrating (e.g., a knife or bullet wound). In a majority of cases, air or alveolar gas enters the pleural space along with blood, causing a hemopneumothorax. Hemothorax may also complicate invasive diagnostic or therapeutic procedures that lacerate pleural or mediastinal blood vessels; such procedures include thoracentesis, pleural biopsy, and cannulation of the subclavian or internal jugular veins. In rare instances, abnormal vascular structures in the mediastinum or lung periphery rupture into the pleural space; examples include hemothorax associated with arteriovenous malformation, pleural endometriosis,[45] and thoracic aortic aneurysm. Rare causes of nontraumatic hemothorax include pleural metastasis and iatrogenic or disease-related coagulopathy.[46]

Diagnosis

When hemothorax is suspected, the hematocrit of the pleural liquid should be measured. It is important to remember that the pleural liquid may appear bloody even when its hematocrit is less than 1%. However, in true hemothorax, the pleural liquid hematocrit exceeds 50% of the peripheral blood hematocrit. Hemothorax, even if it is sterile, may be a cause of transient fever.

Treatment

In a large or rapidly accumulating hemothorax, blood should be promptly drained from the pleural space with a wide-bore chest tube. If rapid bleeding (100 to 200 ml/hr) continues, a thoracoscopy[47] or thoracotomy will be necessary. After bleeding has stopped, intrapleural fibrinolytic therapy to lyse clots not removed by the chest tube may obviate thoracotomy.[48] Potential adverse consequences of undrained pleural blood include empyema and fibrothorax. Empyema may develop because blood provides a rich culture medium for growth of bacteria. Fibrothorax is a late sequela of moderate or large hemothoraces and results from organization of clotted blood into a dense fibrous peel surrounding the lung. A small, self-limited hemothorax associated with minor trauma such as a rib fracture will resolve without drainage or surgery.

CHYLOTHORAX AND PSEUDOCHYLOTHORAX

Chyle is the lipid-rich liquid transported from small intestinal villi—the so called lacteals—to systemic veins in the thorax via the thoracic duct. Disruption or compression of the thoracic duct may lead to leakage of chyle first into the posterior mediastinum and then into the pleural space. Alternatively, an atretic or obstructed thoracic duct may cause reversal of lymph flow through dilated collateral channels. Rupture of these lymphatic channels may also produce chylothorax.

The thoracic duct follows a variable course through the mediastinum. Commonly, it crosses the diaphragm to the right of the vertebral column, entering the chest cavity through the aortic hiatus. It then crosses to the left of the vertebral column between the seventh and fifth thoracic vertebrae, arches above the level of the clavicle, and enters the systemic venous circulation in the region of the left jugular and subclavian veins. Depending on the particular anatomy of a given patient and the level at which the thoracic duct is disrupted, chylous effusions may be left sided, right sided, or, occasionally, bilateral.

Chylothorax

Etiology Various conditions can cause chylothorax. Mediastinal tumors are the most common cause, with lymphomas exceeding metastatic carcinomas in frequency.[49] Trauma is the other major cause of chylothorax in adults. In some cases, major chest trauma has been sustained. In other cases, seemingly minor actions, such as hyperextension of the back, are the only identifiable antecedent events. Thoracic and cardiovascular surgery occasionally results in transection of the thoracic duct and the subsequent development of chylothorax. In addition, chylothorax frequently occurs as a complication of the rare disease lymphangiomyomatosis.[50] Other conditions producing mediastinal lymphatic disruption or obstruction may cause chylothorax. Such conditions include congenital lymphangiectasis, mediastinal irradiation, fibrosing mediastinitis, granulomatous mediastinitis, left subclavian vein thrombosis, and esophageal sclerotherapy.

Diagnosis Chylothorax is usually not suspected until thoracentesis reveals a milky-white pleural liquid. Identification of chylomicrons by lipoprotein analysis establishes the diagnosis of chylothorax and distinguishes it from other causes of an opalescent pleural liquid. However, several less expensive screening tests also may be useful. For example, a chylous effusion remains opaque after centrifugation, whereas the supernatant in empyema is clear. Furthermore, if a chylous effusion sample is allowed to remain undisturbed for 12 to 24 hours, a creamy layer of chylomicrons floats to the top, and the addition of a few drops of ethyl ether rapidly causes the liquid to clear. In most cases of chylothorax, the triglyceride concentration exceeds 110 mg/dl; exceptions usually are limited to patients in whom feedings have been withheld, such as postoperative patients.[49] A CT scan of the chest is indicated in most patients with chylothorax to de-

Table 2 Classification and Treatment of Parapneumonic Effusions and Empyema

Type of Effusion	*Radiographic and Laboratory Findings*	*Treatment*
Typical parapneumonic pleural effusion	> 10 mm maximum thickness on lateral decubitus chest radiograph; glucose level and pH are normal; Gram stain and culture are negative	Antibiotics alone
Complicated parapneumonic pleural effusion	May be nonloculated or multiloculated; no obvious pus; glucose level and pH are low and/or Gram stain and culture are positive	Tube thoracostomy plus antibiotics; if loculated, use intrapleural fibrinolytics; surgery rarely required
Empyema	Free-flowing single loculus or multiloculated; obvious pus present	Tube thoracostomy plus antibiotics; if loculated, use intrapleural fibrinolytics; may require thoracoscopy or decortication

termine if the etiology is carcinoma, and pleural liquid cytology in these patients occasionally yields malignant cells.

Chylous effusions elicit little pleural inflammation and are only very rarely complicated by empyema because of the bacteriostatic properties of chyle. The major consequence of chylous effusions is the rapid and recurrent accumulation of liquid in the pleural space. Normally, the thoracic duct transports chyle at a rate of 1.5 to 2.5 L/day. In patients with chylothorax, much or all of this liquid may enter the pleural space.

Treatment Repeated thoracenteses or chest tube drainage can avert lung compression caused by pleural liquid buildup. However, these procedures may result in large losses of protein, fat, and circulating lymphocytes, rapidly leading to malnutrition and possible immunosuppression.

Definitive treatment of chylothorax varies with the specific etiology. Radiation therapy, with or without chemotherapy, is frequently effective for patients with mediastinal malignant disease, especially lymphomas. Thoracotomy, with oversewing of the thoracic duct leak or ligation of the duct below the leak, is curative in cases of accidental or intraoperative trauma. Pleurodesis [*see* Hydrothorax Caused by Malignant Disease, *above*] may prevent reaccumulation of chyle in patients with unresponsive malignant disease or in other poor operative candidates.

Pseudochylothorax

Occasionally, in patients with long-standing pleural effusions, cholesterol crystals collect in the pleural liquid, causing a milky-white appearance that is indistinguishable from chylothorax on gross inspection.[51] Such pseudochylous, or chyliform, effusions can usually be readily differentiated from true chylothorax on the basis of the clinical setting. Pseudochylous effusions typically occur as a complication of rheumatoid or tuberculous effusions that have been present for several years and are associated with extensive pleural thickening. Chyliform effusions have been reported with paragonimiasis. These effusions should be drained and the underlying process treated.

Parapneumonic Effusions and Empyema

Thoracic empyema most often results from contiguous spread of infection from an underlying region of pneumonia and occasionally from a lung abscess or bronchiectasia. Bacteria are the most common pathogens, although any microorganism capable of causing pneumonia may also cause empyema. In a report from an inner-city municipal hospital, as many as 7% of patients admitted with acute pneumonia had empyema on presentation.[52]

The pleural space may also become infected as a result of seeding of pathogens after thoracic surgery, penetrating trauma, thoracentesis, or tube thoracostomy. Direct spread of infection from a subdiaphragmatic site, hematogenous spread of infection during septicemia, and embolic spread during septic thrombophlebitis are other mechanisms by which the pleural space may become infected. Finally, empyema may occur as a complication of spontaneous pneumothorax, mediastinitis, or esophageal rupture.

Etiology Microorganisms causing empyema have changed considerably during the past 50 years, largely because of the introduction and increasingly widespread use of potent broad-spectrum antibiotics. In the 1930s and 1940s, *Streptococcus pneumoniae* (pneumococcus) and hemolytic streptococci were the pathogens most frequently isolated from empyemas. In the 1950s and 1960s, *Staphylococcus aureus* and gram-negative bacilli (e.g., *Escherichia coli, Haemophilus influenzae, Klebsiella pneumoniae,* and *Pseudomonas aeruginosa*) became the predominant pathogens found in empyemas. Some series, employing modern techniques for culturing anaerobic bacteria, have found a preponderance of anaerobes, either alone or in combination with aerobic bacteria.[53] Fungi have been isolated from empyemas more frequently in recent years, particularly in hospitalized patients with significant comorbidities who are immunocompromised.[54]

Diagnosis The most common presenting symptoms of empyema are fever, chest pain, cough, and dyspnea. These symptoms, however, are not specific enough to distinguish patients who have pneumonia and empyema from those who have pneumonia alone. Patients with anaerobic empyema are more likely to have an indolent presentation, characterized by low-grade fever, anorexia, weight loss, or anemia. Between 10% and 15% of patients with empyema do not have fever or an elevated white cell count.

In patients with pneumonia and pleural effusion, it is important to determine whether the effusion is a typical parapneumonic effusion (formerly known as a sympathetic effusion), a complicated parapneumonic effusion, or an empyema. These three types of effusion can be distinguished on the basis of the radiographic appearance and characteristics of the pleural liquid obtained at thoracentesis [*see Table 2*].

Treatment In general, very small pleural effusions (maximum thickness < 10 mm as measured on lateral decubitus chest radiograph) in the setting of acute pneumonia will resolve with

appropriate antibiotic treatment of the pneumonia and do not require further investigation by thoracentesis. However, most larger effusions require thoracentesis. Loculations on radiographic studies, low pleural liquid pH and glucose concentrations, high LDH, or positive Gram stain and cultures indicate the presence of a complicated parapneumonic effusion (in cases in which there is no obvious presence of pus) or empyema (in cases in which there is obvious presence of pus). A pleural liquid pH of 7.2 or higher favors the diagnosis of typical parapneumonic effusion. Pleural liquid pH values between 7.0 and 7.2 are considered to be indeterminate (borderline complicated pleural effusion) and necessitate repeated thoracenteses, often from different sites, to determine whether a parapneumonic effusion is typical or complicated.

Systemic antibiotic therapy and prompt drainage of the pleural space are key.[55] In general, antibiotics penetrate the pleural space well and should be given systemically rather than intrapleurally. The choice of antibiotics should be guided by the results of a Gram stain and culture of pleural samples, blood samples, sputum samples, or a combination of these. For anaerobic empyema, clindamycin is the drug of choice because as many as 10% to 15% of the anaerobic isolates are penicillin-resistant *Bacteroides fragilis*.

Pleural drainage usually requires prompt placement of a chest tube because complicated parapneumonic effusions and empyema liquid reaccumulate rapidly after thoracentesis. In addition, continuous drainage is required to sterilize the pleural space, and drainage prevents the formation of loculi. In some patients with complicated parapneumonic effusions and empyema that do not resolve radiographically after placement of a chest tube, intrapleural streptokinase or urokinase[56] may allow adequate drainage and obviate the use of multiple chest tubes or surgical drainage, although the evidence for this approach is not strong.[57] Urokinase may be preferred over streptokinase because urokinase is associated with less frequent febrile reactions.[58] Use of intrapleural fibrinolytic agents may also allow small-bore pigtail catheters to be placed into loculi with CT or fluoroscopic guidance.[59] This approach may be unsuccessful with thick pleura, multiple loculi, or both.[60] Early thoracoscopic therapy may be more cost-effective than attempts at drainage and the use of fibrinolytic agents in such cases.[61]

With appropriate treatment, most patients will become afebrile within 4 days, and peripheral white cell counts will return to normal within a week. In a minority of cases, adequate pleural drainage cannot be achieved, and fever and leukocytosis persist; in such cases, a surgical procedure, either video-assisted thoracoscopy (early stage) or thoracotomy (late stage), is usually required.[62] Stripping the adherent inflammatory peel from the pleural surfaces and removing the purulent exudate from the pleural cavity (i.e., decortication and empyemectomy) is a demanding surgical procedure. In elderly or severely debilitated patients, a more conservative surgical procedure may be used in which a partial rib resection is performed to create an open pleurocutaneous tract.[63] The chronic empyema cavity that is formed can then be managed in the same way as any other chronic open wound and eventually may close spontaneously or can be closed by reconstructive surgery.

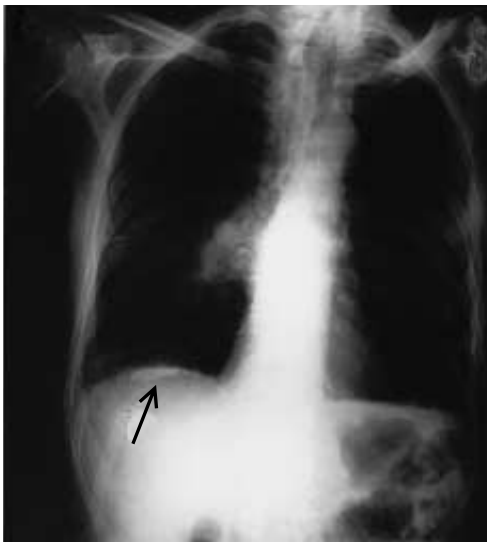
If left untreated, empyema can lead to septic shock. Other complications of untreated empyema include localized dissection of infection and consequent rupture through the skin (empyema necessitatis) or into the bronchial tree (bronchopleural fistula).

Pleural Plaques, Diffuse Pleural Thickening, and Pleural Calcification

PLEURAL PLAQUES

Pleural plaques are asymptomatic collections of collagenous connective tissue that form on the parietal pleura. They usually occur as a consequence of heavy exposure to asbestos fibers. In rare cases, plaques may form in response to inhalation of inorganic fibers other than asbestos, or they may arise in association with hyaloserositis or other conditions. Pleural plaques are discrete lesions that are covered by a layer of normal mesothelium and are clearly demarcated from normal surrounding pleura. They vary in diameter from a few millimeters to several centimeters and have an average thickness of approximately 0.5 cm. They typically form on the lateral and posterior pleural surfaces and the central (aponeurotic) portion of the diaphragm, sparing both the costophrenic sulci and the upper third of the thorax. They are often difficult to identify on chest radiograph unless their orientation is tangential to the x-ray beam, and only a small percentage of noncalcified pleural plaques identified at surgery

a



b

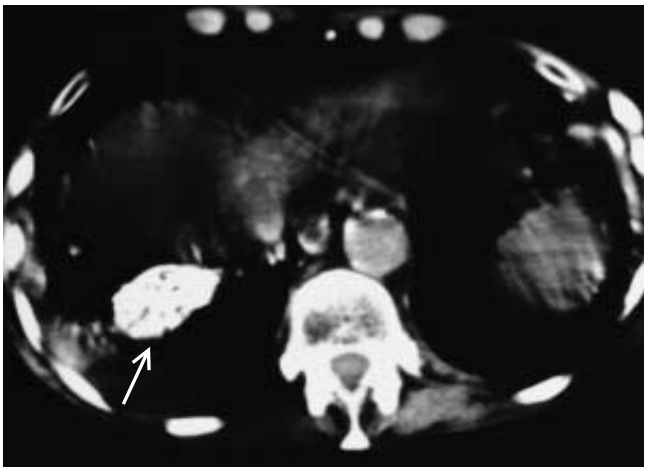


Figure 7 **(*a*) Dense, white linear opacity along the left diaphragm (arrow) represents calcification of the parietal pleura in a patient with long-standing asbestos exposure. (*b*) CT scan shows area of dense calcification along the central dome of the diaphragm (arrow).**

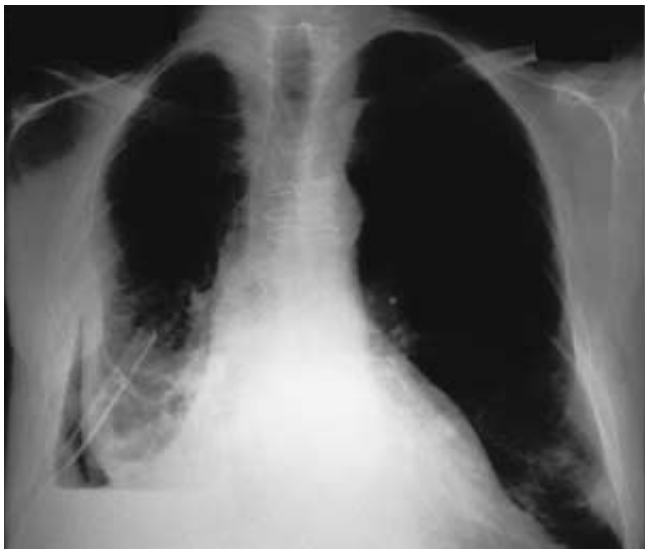


Figure 8 **Radiograph reveals a so-called trapped lung, which was caused by marked thickening of the visceral pleura from a chronic empyema. Despite the application of a highly negative pressure to the pleural space, the right lung does not expand. A pyopneumothorax is present in the right thorax because evacuated pus has been replaced by air. There is no bronchopleural fistula (air leak from the lung). Obstruction of the mainstem bronchus may also cause a trapped lung.**

or postmortem examination are recognized radiographically. CT scanning is more sensitive than plain films of the chest for detection of plaques. After several years, pleural plaques calcify, causing a radiographically distinct image [*see Figure 7*].

Pleural plaques develop slowly over a period of many years, and they usually do not become apparent until 20 years after initial asbestos exposure. After 40 years, as many as 50% of persons who had occupational exposure to asbestos have pleural plaques. These hyaline plaques do not degenerate into neoplastic disease, but minimal changes in lung function may be associated with their presence.[64] Their main significance is as a marker of probable asbestos exposure.

Other processes that resemble plaques include metastases, callus formation surrounding rib fractures, and normal areas of muscle insertion or fat accumulation. The distinction can usually be made easily by overpenetrated oblique views on radiograph or, if necessary, by CT scanning.

PLEURAL THICKENING

Pleural thickening develops when pleural inflammation of any cause heals with the formation of fibrous tissue involving the visceral or parietal pleural surfaces. The costophrenic angle is most frequently involved, causing blunting of the normal recess. Localized pleural thickening over the apex of the lung—the so-called apical pleural cap—usually involves fibrous scarring of the apical lung merging into the adjacent visceral pleura. Although apical pleural thickening was once thought to be indicative of granulomatous disease from tuberculosis, more recent studies have documented the absence of any correlation with tuberculosis. The etiology of this asymptomatic finding is unknown.

Pleural thickening may be bilateral or unilateral. Extensive bilateral pleural thickening, extending along the lateral margins of the chest and even to the apex of the lung, is frequently the result of exposure to inorganic fibers—most often asbestos and, occasionally, talc. Other causes of bilateral pleural thickening include uremia and collagen vascular diseases, especially SLE and rheumatoid arthritis. Diffuse unilateral pleural thickening usually results from one of three causes: hemothorax, bacterial empyema, or tuberculous pleurisy.

Diffuse pleural thickening may be sufficiently dense to entrap the underlying lung, preventing its full expansion and impairing gas exchange [*see Figure 8*].[65] This condition is referred to as fibrothorax. During the months immediately after hemothorax or empyema, inflammatory pleural thickening may gradually resorb. However, once an organized fibrous peel has formed, surgical decortication is the only available therapy. If the underlying lung tissue is relatively normal, resection of a thick pleural peel may result in significant improvement in ventilatory function.

In rare cases, localized or diffuse pleural thickening can entrap a small region of lung, producing a masslike lesion termed rounded atelectasis. It can be difficult to differentiate such lesions from other, more serious mass lesions (e.g., neoplasm).[66]

PLEURAL CALCIFICATION

Diffuse, sheetlike calcification may develop in patients with hemothorax, bacterial empyema, or tuberculous pleurisy. Unlike the pattern of calcification that is observed in asbestos-related pleural plaques, calcium deposition after these conditions always takes place along the visceral pleura, thereby outlining the inner margin of the pleural thickening. In addition, calcification is generally found at the level of the midthorax, sparing the diaphragmatic pleural surface.

Pneumothorax

ETIOLOGY

Pneumothorax, the presence of gas within the pleural space, indicates that disruption of the visceral or parietal pleura has occurred.[67] The gas that enters the pleural space may come from various sources. For example, in a penetrating injury involving the chest wall and parietal pleura, the gas enters from the outside environment. Alternatively, the gas may come from a gas-filled gastrointestinal structure; this situation might arise as a result of a ruptured esophagus, or it could arise from a ruptured intra-abdominal viscus and subsequent escape of gas across the diaphragm from a pneumoperitoneum. Most commonly, the source of the gas is the lung—after alveolar or tracheobronchial injury, after blunt or penetrating trauma, or as a complication of invasive diagnostic and therapeutic procedures, such as thoracentesis, attempted percutaneous cannulation of a central vein, acupuncture, or intercostal nerve block.

Pneumothorax caused by disruption of the visceral pleura may also result from focal pulmonary processes. Focal destructive processes may involve the visceral pleura as a primary site or may extend to the pleura from adjacent lung tissue. Examples of focal pulmonary processes that cause pneumothorax include bronchogenic carcinoma, rheumatoid lung nodule, thoracic endometriosis,[45] necrotizing pneumonia, and pulmonary infarct. In the past, tuberculosis was frequently implicated as a cause of pneumothorax. *Pneumocystis carinii* pneumonia, a complication of AIDS, has become a common cause of pneumothorax, particularly in patients receiving aerosolized pentamidine, those who smoke cigarettes, and those with pneumatoceles seen on chest radiograph or CT scan.[68]

Diffuse diseases of the lung parenchyma can also cause pneumothorax. Such diseases can greatly distort the lung architecture, resulting in an uneven distribution of ventilation within the lung. In this setting, localized alveolar overdistention, along with weakened alveolar walls, leads to an increased incidence of associated pneumothorax. The most common of these diffuse processes are the obstructive lung diseases, specifically emphysema, asthma, and cystic fibrosis.[69] Rarer conditions that also carry an increased risk of pneumothorax include pulmonary lymphangiomyomatosis,[50] tuberous sclerosis, eosinophilic granuloma, scleroderma, and congenital disorders of connective tissue. In some cases, the visceral pleura remains intact but alveolar gas gains entry to the pleural space via the mediastinum. When perivascular alveoli rupture, alveolar gas can dissect centripetally along the bronchovascular interstitium to the mediastinum, entering the pleural space through the mediastinal parietal pleura.

CLASSIFICATION OF PNEUMOTHORAX

A commonly used classification of pneumothorax recognizes traumatic pneumothorax and iatrogenic pneumothorax as distinct clinical entities and lumps all other causes of pneumothorax under the somewhat misleading label of spontaneous pneumothorax. When spontaneous pneumothorax occurs in patients with underlying pleural or parenchymal disease, it is called secondary spontaneous pneumothorax. When no underlying lung disease is evident, it is called idiopathic spontaneous pneumothorax. Although patients with idiopathic spontaneous pneumothorax are otherwise healthy, most have subpleural apical blebs, frequently associated with more diffuse centrilobular emphysema detectable by CT scan[70]; during surgery, the blebs are often found to have ruptured into the pleural space. These abnormalities may have a genetic etiology.[71] Secondary spontaneous pneumothorax is a more serious problem than idiopathic spontaneous pneumothorax because patients with secondary spontaneous pneumothorax typically have impaired lung function.[67]

EPIDEMIOLOGY

Secondary Spontaneous Pneumothorax

The incidence of secondary spontaneous pneumothorax depends on the underlying disease process. The incidence of pneumothorax in patients with chronic obstructive lung disease is approximately 26 per 100,000 per year, and the incidence is directly related to the severity of obstruction.[67] Pneumothorax will develop in 5% to 8% of cystic fibrosis patients at some point in their lifetime; however, it occurs in 16% to 20% of cystic fibrosis patients older than 18 years. Pneumothorax occurs in 2% to 6% of HIV patients and is almost always associated with *P. carinii* pneumonia.[67] Twenty-five percent of patients with eosinophilic granuloma and 80% of patients with pulmonary lymphangiomyomatosis have pneumothorax at some point in their disease course, and pneumothorax can be a presenting manifestation in both diseases.[67]

For most underlying lung diseases, the rate of recurrent pneumothorax is similar to that of idiopathic spontaneous pneumothorax (39% to 47%),[67] although patients with cystic fibrosis have a much higher recurrence rate (68% to 90%).[69]

Idiopathic Spontaneous Pneumothorax

Idiopathic spontaneous pneumothorax has an incidence of approximately 4.3 cases per 100,000 patient-years. The peak incidence is in persons between 20 and 30 years of age, and the male-to-female ratio is approximately 5:1. Patients often have a tall, thin stature[72] and very frequently are cigarette smokers. The precise mechanism whereby male sex, asthenic habitus, and cigarette smoking predispose to apical pleural bleb formation or rupture is unknown. One study detected anomalies of the bronchial tree in the majority of patients with spontaneous pneumothorax, suggesting associated congenital abnormalities of lung structure.[73] How these anomalies may relate to the occurrence of pneumothorax is unclear. Common misconceptions are that strenuous physical activity is frequently a trigger for the development of pneumothorax and that patients are at increased risk during airplane travel. In fact, most studies have found that the onset of symptoms of pneumothorax usually occurs at rest or during light activity, and a study of pneumothoraces among pilots in the United States Air Force found that very few episodes occurred during flight.[74]

DIAGNOSIS

The most common symptoms of pneumothorax are chest pain and dyspnea. The pain may be a dramatic, severe, stabbing unilateral chest pain with a sudden, explosive onset, sometimes radiating to the ipsilateral shoulder or scapular area. In other cases, the discomfort may be more modest and more easily tolerated. Often, patients will recall previous transient episodes of pain that were similar to, although milder in degree or shorter in duration than, the one that finally caused them to seek medical attention. Dyspnea develops in most patients. The dyspnea is more severe when the pneumothorax is large and when there is significant underlying lung disease (i.e., secondary spontaneous pneumothorax).

Secondary Spontaneous Pneumothorax

In a patient with secondary spontaneous pneumothorax associated with underlying emphysema, the diagnosis is particularly difficult to make on the basis of physical findings. Decreased lung elastic recoil and residual hyperinflation keep the lung from fully collapsing and limit changes in the size of the thoracic cage. In addition, in patients with emphysema, physical findings such as hyperresonance and diminished breath sounds may be found over the contralateral lung. In a patient with chronic airflow obstruction, the sudden onset of chest pain and worsened dyspnea should raise the suspicion of pneumothorax. Confirmation of the diagnosis of pneumothorax usually requires a chest radiograph.

Idiopathic Spontaneous Pneumothorax

Characteristic physical findings on the involved side include expansion of the hemithorax (caused by release of the ipsilateral chest wall from the recoil forces of the lung), hyperresonance, diminished fremitus, diminished transmission of voice sounds, and distant or absent breath sounds.

As gas collects in the pleural space, the lung recoils from the chest wall toward the hilum. The presence of a pneumothorax can be identified on a chest radiograph by visualization of a thin (=1 mm) linear shadow made by the visceral pleura as it passes along a plane tangential to the x-ray beam. This linear shadow, marking the outer rim of the lung, follows the contour of the inner aspect of the chest wall, and no lung markings (i.e., bronchovascular shadows) can be seen peripheral to it. In cases in which the pneumothorax is small and gas collects over the apex of the lung, it may be difficult to distinguish the visceral pleural line from superimposed rib margins. When the chest radiograph

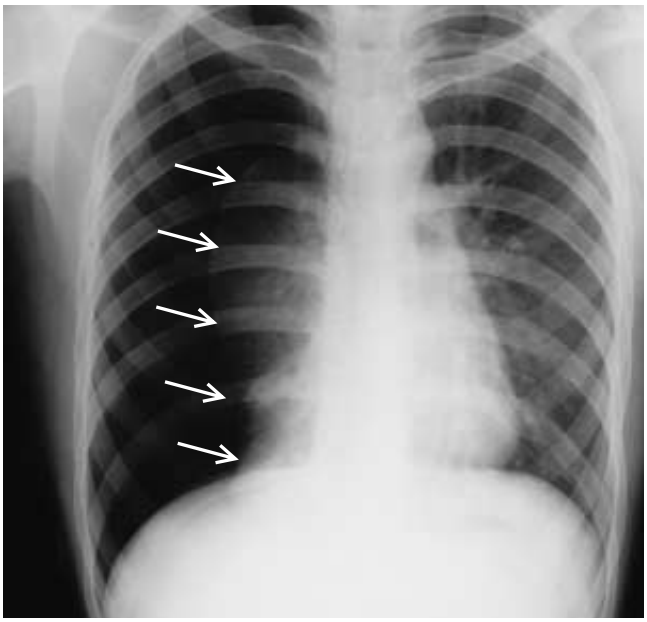


Figure 9 **This chest radiograph of a patient with pneumothorax demonstrates virtually complete collapse of the right lung. There has been a slight shift of the mediastinum toward the contralateral side. The visceral pleura (arrows) can be clearly identified because gas is present on both sides.**

is taken with the patient in the supine position, a pneumothorax will collect along the costophrenic sulcus rather than along the apex of the lung. This so-called deep sulcus sign may be useful in identifying occult pneumothorax in hypotensive patients in whom upright chest radiographs are inadvisable. Chest radiographs taken during full expiration have not been shown to enhance the detection of pneumothorax.[75] Pitfalls in the radiographic diagnosis of pneumothorax include confusion regarding two other causes of curvilinear shadows in the chest: (1) extrathoracic skin folds and (2) intrapulmonic cysts or bullae.

It is difficult to accurately estimate the size of a pneumothorax relative to the size of the hemithorax by casual inspection of the chest radiograph. As a rough indicator, a collection of gas around the lung that has an average thickness of 1 in. represents a 30% pneumothorax. When collapse of the lung is complete (100% pneumothorax), the lung forms a fist-sized opacity near the hilum, and the mediastinum may shift slightly toward the contralateral lung; the diagnosis is obvious on x-ray [*see Figure 9*].

A small pleural effusion is commonly present in pneumothorax. It is detected on radiograph by blunting of the costophrenic sulcus. The pleural liquid is usually bloody. The effusion is probably formed as a result of the rupture of small blood vessels within pleural adhesions.

TREATMENT

Secondary Spontaneous Pneumothorax

Because initial or recurrent episodes of secondary spontaneous pneumothorax can be life threatening, aggressive treatment is required.[67,76] Patients should be hospitalized and a chest tube placed for drainage; observation and simple aspiration are not adequate treatment. The methods for preventing recurrence of secondary spontaneous pneumothorax are the same as those of idiopathic spontaneous pneumothorax (e.g., thoracotomy, video-assisted thoracotomy, pleurodesis) (see below), although many experts feel that application of such methods should be utilized on the first occurrence of secondary spontaneous pneumothorax.[76]

Idiopathic Spontaneous Pneumothorax

Treatment of idiopathic spontaneous pneumothorax is directed in part at allowing the collapsed lung to expand fully again and in part at preventing recurrences.[67] In mildly symptomatic patients with moderate pneumothorax, simple aspiration of the pneumothorax may be successful in 60% to 70% of cases.[77] A patient with a small (15% to 20%) pneumothorax who is asymptomatic or minimally symptomatic can be safely observed, and the pneumothorax can be allowed to resorb spontaneously. Complete resolution of a pneumothorax usually requires approximately 10 days, provided there is no further air leak. Resolution can be accelerated by the administration of supplemental oxygen. Supplemental oxygen lowers the nitrogen content of blood, thereby increasing the pressure gradient for nitrogen that favors transfer of gas from the pleural space into the venous end of the pleural capillaries. Strict bed rest does not hasten resolution and is not necessary. A large pneumothorax in a symptomatic patient should be evacuated promptly with an intercostal small-bore (14 French) or large-bore (28 French or greater) chest tube passed cephalad into the apex of the chest. The air leak from the lung may seal immediately or may persist for 3 to 5 days until the tear in an apical bleb heals. An air leak from the lung that persists for more than 7 days is considered by many physicians to be an indication for surgical intervention; at thoracoscopy or thoracotomy, the blebs can be oversewn or excised by wedge resection and the pleura abraded or pleurodesis performed.

Incidence of recurrence Idiopathic spontaneous pneumothorax often recurs. At least 20% to 30% of patients with idiopathic spontaneous pneumothorax will experience an ipsilateral recurrent pneumothorax within the ensuing 5 years; most recurrences occur within a year after the initial event. According to some reports, the rate of initial recurrence may be as high as 50%. Recurrences are more common in women and taller men and are reduced by smoking cessation.[78] Ninety percent or more of recurrences are ipsilateral, despite the fact that the underlying abnormality (i.e., apical subpleural blebs) is bilateral in more than half the cases.[79] Simultaneous bilateral idiopathic spontaneous pneumothoraces are fortunately infrequent, occurring in approximately 1% of cases; surprisingly, when they do occur, they are rarely fatal. After the first ipsilateral recurrence of a pneumothorax, subsequent recurrences become increasingly likely.

Management of recurrence There is debate regarding the best method of preventing recurrences and the optimal timing for such an intervention.[76] Because half or more of patients with idiopathic spontaneous pneumothorax will never suffer a recurrence, it seems reasonable to withhold preventive treatment until after a recurrence. At that point, the probability of further recurrences is quite high, and the discomfort and inconvenience associated with recurrent pneumothoraces would have become increasingly apparent to the patient. Traditional and definitive treatment for prevention of recurrences involves thoracotomy: apical blebs are oversewn or excised by wedge resection, and adhesions are induced between the lung and chest wall by abrasion of the pleural surface with dry gauze or by partial parietal

pleurectomy. Recurrent pneumothorax occurs after this procedure in 0% to 2% of patients. Video-assisted thoracoscopy has allowed these procedures to be performed without thoracotomy, resulting in less pain and shorter hospital stays.[80] Attempts have also been made to achieve pleurodesis without thoracotomy via a chest tube. This technique is the same as that used for the management of pleural effusions in patients with malignant disease [*see* Hydrothorax Caused by Malignant Disease, *above*]. Recurrence of spontaneous pneumothorax is reduced after doxycycline administration or talc pleurodesis.

TENSION PNEUMOTHORAX

Occasionally, gas enters the pleural space during the inspiratory phase and is prevented from escaping during expiration, presumably because an airway or tissue flap acts as a one-way valve. Under these circumstances, there is a progressive increase in the amount of pleural gas, and the pleural gas is under increased pressure (i.e., tension). This situation is referred to as tension pneumothorax. Tension pneumothorax occurs in only 1% to 2% of the cases of idiopathic spontaneous pneumothorax. However, it is a more common manifestation of the barotrauma that may occur as a complication of positive pressure mechanical ventilation.

Diagnosis

Tension pneumothorax is a medical emergency. Patients are often dyspneic at rest and gasping for breath. Cyanosis and hypotension are common. The diagnosis cannot be established by a plain chest radiograph, although a marked contralateral shift of the mediastinum and depression or inversion of the ipsilateral hemidiaphragm are suggestive.

Treatment

If there is acute distress, immediate action must be taken to remove gas from the pleural space. Introduction of a small-bore plastic catheter together with a needle through an intercostal space may suffice for emergency relief if delay in placing a full-sized chest tube is anticipated. In patients with tension pneumothorax, the gas that is under pressure will rush out of the chest through the open catheter. If a Heimlich flutter valve, which allows one-way passage of gas, is attached to the catheter, a series of coughs or Valsalva maneuvers will allow almost complete evacuation of the remainder of the pneumothorax that is not under tension. Because of the life-threatening nature of tension pneumothorax, a procedure to prevent recurrence should be undertaken after the first such event in cases involving idiopathic spontaneous pneumothorax.

Hilar and Mediastinal Disorders

NORMAL ANATOMY OF THE HILA AND MEDIASTINUM

The mediastinum is the intrathoracic compartment situated between the two lungs. It is bordered anteriorly by the sternum, posteriorly by the vertebral column, and laterally by parietal pleura. It extends from the thoracic inlet to the superior surface of the diaphragm. Several important structures are contained within or pass through the mediastinum, including the heart and great vessels, the esophagus, the trachea and mainstem bronchi, lymphatic vessels and nodes, and nerves.

It is convenient to divide the mediastinum into three parts on the basis of imaginary coronal sections [*see Figure 10*]. The anterior mediastinum (i.e., the portion anterior and superior to the anterior surface of the pericardium) contains the thymus, lymph nodes, and mesenchymal tissue. The middle mediastinum consists of the heart and pericardium, the major vessels as they enter and leave the heart, the trachea and main bronchi, lymph nodes, and portions of the phrenic and vagus nerves. Between the posterior aspect of the pericardium and the vertebral column lies the posterior mediastinum, which contains the descending aorta, the esophagus, the thoracic duct, lymph nodes, and a portion of the vagus nerve. The sympathetic nerve chain, which runs in the costovertebral gutter, is conventionally included in the posterior mediastinum.

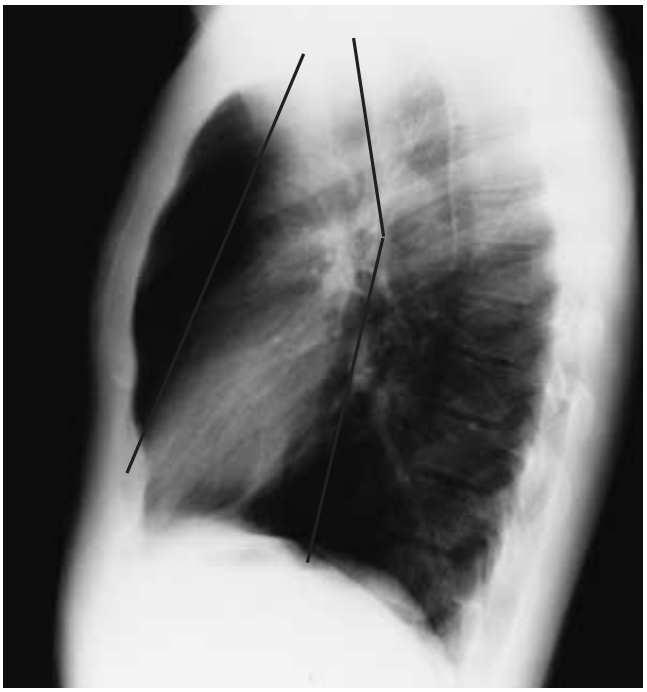


Figure 10 **Lines drawn on this normal lateral chest x-ray mark the theoretical division of the mediastinum into anterior, middle, and posterior compartments. Focal abnormalities of the mediastinum can be identified as occurring predominantly in one compartment or another, facilitating differential diagnosis.**

As blood vessels, airways, and nerves leave the mediastinum and pass into the lungs, they form the pulmonary hila. On chest radiographs, the hilar shadows are composed primarily of branches of the pulmonary arteries. The right pulmonary artery bifurcates just before entering the hilum as the truncus anterior artery (superior branch) and the right interlobar artery (inferior branch). The left pulmonary artery divides within the hilum into superior branches and the larger left interlobar artery. In 97% of normal persons, the left hilum is positioned superior and posterior to the right hilum; displacement from this normal position may be caused by mass lesions or lobar atelectasis. Central bronchopulmonary lymph nodes that are important for drainage of the lung parenchyma are situated within the pulmonary hila along the bronchial tree, especially within bifurcations.

HILAR ENLARGEMENT

The most common causes of enlargement of the hilar shadows on chest radiographs are vascular engorgement and adenopathy.

Enlargement Caused by Vascular Engorgement

Vascular engorgement may result from increased blood flow in the pulmonary circulation, as in atrial septal defect with left-to-right intracardiac shunt, or from pulmonary arterial hypertension of any cause. The caliber of the pulmonary arteries can be assessed by measurement of the transverse diameter of the right interlobar artery on a posteroanterior radiograph; in normal persons, this value is usually less than 16 mm. In rare cases, localized aneurysmal dilatation of the pulmonary arteries occurs; this condition is referred to as the Hughes-Stovin syndrome.

On plain chest radiographs, it is sometimes difficult to distinguish vascular from nodal enlargement of the hila. Pulmonary angiography was formerly needed to make this distinction. Currently, however, contrast-enhanced CT scanning of the chest can effectively identify vascular structures in the hila without the need for invasive procedures. Occasionally, there is a strong contraindication to the use of intravenous contrast agents in some patients; in such cases, magnetic resonance imaging, which does not require the use of contrast agents, can be used.

Enlargement Caused by Adenopathy

Hilar lymph node enlargement may be unilateral or bilateral. Unilateral lymphadenopathy may accompany virtually any pneumonia, although it is most characteristic of granulomatous infections (e.g., tuberculosis and atypical mycobacteriosis, histoplasmosis, and coccidioidomycosis) and certain atypical pneumonias (e.g., *Mycoplasma* infections, tularemia, pertussis, and psittacosis). Neoplastic enlargement of hilar lymph nodes usually results from spread of bronchogenic carcinoma; extrathoracic cancers that metastasize to the hilar and mediastinal lymph nodes include cancers of renal cell origin, which are implicated especially often, and cancers of the breast and GI tract. Hodgkin disease and other lymphomas may also cause unilateral or asymmetrical hilar adenopathy. Finally, about 1% to 3% of patients with sarcoidosis have unilateral hilar adenopathy.

A common diagnostic challenge is the evaluation of a patient with bilateral hilar adenopathy. This condition is often found in association with mediastinal adenopathy and sometimes in association with parenchymal infiltrates. Nonspecific symptoms of cough, chest pain, dyspnea, or malaise may have prompted the chest radiograph, or bilateral hilar adenopathy may have been detected incidentally on a chest radiograph obtained for unrelated reasons. The most common etiology, especially in patients between 20 and 40 years of age, is sarcoidosis. The differential diagnosis of bilateral hilar adenopathy includes the following: (1) lymphoma, which is usually accompanied by extrathoracic manifestations, such as systemic symptoms, peripheral adenopathy, and anemia; (2) metastatic cancer, in which the primary malignant disease is most often known; (3) chronic granulomatous infections, such as tuberculosis or histoplasmosis, in which the adenopathy is more commonly unilateral; and (4) berylliosis, which can precisely mimic sarcoidosis but which can be readily diagnosed with a careful occupational history.

In a retrospective analysis of 100 cases of bilateral hilar adenopathy not caused by infection, it was found that patients who were asymptomatic and had a normal physical examination or those who had erythema nodosum or uveitis as the only manifestation of their disease invariably had sarcoidosis. On the other hand, it was found that patients who were symptomatic or patients with other abnormal physical findings (e.g., peripheral adenopathy, hepatomegaly, or splenomegaly) in some cases had sarcoidosis and in other cases had neoplastic node involvement.

The approach to diagnosis in patients with bilateral hilar adenopathy depends on the relative likelihood of the various diagnostic probabilities, as determined by clinical assessment. Patients strongly suspected of having sarcoidosis may be observed with serial chest radiographs or may undergo transbronchial lung biopsy via the fiberoptic bronchoscope for tissue confirmation (i.e., identification of noncaseating granulomas). In patients with granulomatous infection, CT scanning can indicate the presence of active infection when areas of low attenuation (necrosis), peripheral rim enhancement, and larger nodes are seen, whereas inactive infection is suggested by smaller nodes, homogeneous density, and calcification.[81] When lymphoma is suspected, lymph node biopsy by mediastinoscopy should provide diagnostic material. In patients suspected of having carcinoma of the lung, an increase in the uptake of fluorodeoxyglucose by the thoracic lymph nodes, as shown through use of positron emission tomography, indicates high risk of malignancy.[82] The definitive diagnosis of cancerous lymph nodes can be determined with a high yield by needle aspiration of the hilar mass under fluoroscopic, CT, or transesophageal endosonographic guidance.[83] It should be noted, however, that direct histologic analysis of hilar lymph nodes, which might be necessary in cases of lymphoma in which there is no mediastinal involvement, requires at least a limited anterior thoracotomy to obtain a tissue sample of adequate size.

Enlargement Caused by Calcified Lymph Nodes

Calcified hilar lymph nodes most often indicate previous granulomatous infection, especially tuberculosis or histoplasmosis. Hilar and mediastinal lymph node calcification in patients with silicosis may appear in a highly characteristic pattern referred to as eggshell calcification, in which calcium outlines only the perimeter of the lymph nodes. Sarcoidal lymph nodes rarely become calcified.

ACUTE MEDIASTINITIS

Acute bacterial mediastinitis usually occurs as the consequence of esophageal perforation and the subsequent release of acidic gastric juices, often along with anaerobic bacteria, into the mediastinum. Rupture of the lower esophagus from violent vomiting is referred to as Boerhaave syndrome; free gas in the mediastinum (i.e., pneumomediastinum) and pleural effusion or pneumothorax (which is usually left sided) are common manifestations of this syndrome. Esophageal rupture may also occur as a complication of foreign-body ingestion, esophageal carcinoma, penetrating or blunt chest trauma, and medical procedures that dilate the esophagus. Occasionally, infection spreads to the mediastinum from adjacent sites; empyema, lung abscess, pericarditis, and retropharyngeal abscess[84] may all be complicated by acute mediastinitis. Mediastinitis can also be a complication of sternotomy for cardiac or thoracic surgery.[85] Finally, a mediastinal bronchogenic cyst may become infected and discharge its contents into the mediastinum.

Patients who have acute mediastinitis are usually acutely and severely ill. Symptoms of the disorder include fever, dysphagia, and a lancinating chest pain. The chest radiograph may reveal, in addition to mediastinal widening, such abnormalities as the presence of pneumomediastinum and pleural effusions.

Treatment consists of the administration of broad-spectrum antibiotics, including antibiotics that are effective against anaerobic bacteria, because the infections are usually polymicrobial.[86] In addition, open surgical drainage is required in most cases.[87] Per-

Table 3 Causes of Mediastinal Widening[88]

- *Diffuse Mediastinal Widening*
  - Acute mediastinitis
  - Hemorrhage
  - Lipomatosis
  - Fibrosing mediastinitis
- *Anterior Mediastinal Masses*
  - Thymus disorders
    - Thymoma
    - Thymic cyst
    - Thymolipoma
    - Thymic hyperplasia
    - Thymic carcinoma
    - Thymic carcinoid
  - Teratoma and dermoid cyst
  - Intrathoracic goiter or thyroid carcinoma
  - Lymphoma
  - Parathyroid masses
  - Malignant germ cell neoplasms
    - Endodermal sinus tumor
    - Seminoma
    - Primary choriocarcinoma
    - Mixed germ cell tumor
  - Mesenchymal neoplasms
    - Lipoma
    - Fibroma
    - Hemangioma
    - Lymphangioma (cystic hygroma)
- *Middle Mediastinal Masses*
  - Lymphoma
  - Carcinoma of the trachea
  - Carcinoma metastases to the lymph nodes
  - Granulomatous mediastinitis
  - Bronchogenic cyst
  - Pleuropericardial cyst
  - Diaphragmatic hernia (through foramen of Morgagni)
  - Benign lymph node hyperplasia (Castleman disease)
  - Vascular dilatation
- *Posterior Mediastinal Masses*
  - Neurogenic tumors
  - Cysts
    - Neurenteric
    - Gastroenteric
    - Thoracic duct
  - Esophageal neoplasms and diverticula
  - Diaphragmatic hernia (through foramen of Bochdalek)
  - Diseases of the thoracic spine
  - Extramedullary hematopoiesis

cutaneous catheter drainage may be successful in some cases of limited esophageal leakage.

MEDIASTINAL WIDENING AND MASS LESIONS

A great diversity of benign and malignant lesions may cause the mediastinum to have an abnormal appearance on the chest radiograph. Occasionally, a specific set of findings may prompt a search for a mediastinal lesion. For example, the muscle weakness of myasthenia gravis may prompt a search for associated thymomas. Alternatively, the constellation of abnormalities suggesting the superior vena cava syndrome (e.g., headaches, jugular venous distention, engorgement of the head and neck, and a prominent collateral venous pattern across the upper thorax) may prompt a search for an obstructing lesion [*see 12:VIII Lung Cancer and 12:XII Oncologic Emergencies*]. In some patients, the mediastinal disorder induces nonspecific symptoms referable to the thorax, such as cough, dyspnea, chest pain or pressure, hemoptysis, dysphagia, hoarseness, or wheezing. Most often, however, diagnostic evaluation is prompted by an asymptomatic radiographic finding.

The differential diagnosis of abnormal radiographic opacities within the mediastinum includes benign and malignant neoplasms, cysts, vascular abnormalities, ectopic thyroid tissue, granulomatous diseases, and several other disorders.[88] It is useful to distinguish abnormalities that cause diffuse widening of the mediastinum from those in which the abnormality is confined primarily to the anterior, middle, or posterior mediastinum [*see Table 3*]. Selected examples of each category will be presented.

With the use of intravenous contrast agents, CT scanning can distinguish vascular structures, fat density, cysts, and calcifications from soft tissue density. Often, benign mediastinal lesions can be identified with reasonable certainty by the CT image alone; nonneoplastic lesions include those associated with mediastinal lipomatosis, congenital cysts, vascular aneurysms, diaphragmatic hernias, and intrathoracic goiters. In other cases, CT scanning can be used with a high degree of sensitivity and specificity to guide the biopsy of suspected malignant lesions.[89] MRI may provide advantages over CT in cases suspected to be vascular.[90]

Diffuse Mediastinal Widening

Diffuse mediastinal widening may occur acutely or chronically. Acute widening of the mediastinal shadows may result from an acute mediastinitis or from mediastinal hemorrhage. Important causes of chronic mediastinal widening include mediastinal lipomatosis and mediastinal granuloma and fibrosis.

Mediastinal Lipomatosis

Mediastinal lipomatosis is a consistently asymptomatic condition in which excess fat tissue is deposited between and around mediastinal structures; it is found in some patients with Cushing disease, iatrogenic Cushing syndrome, or obesity. The only significance of mediastinal lipomatosis is that it must be distinguished from other, more serious causes of mediastinal widening. CT scanning of the thorax readily establishes the diagnosis by identifying material of fat density deposited diffusely through the mediastinum.

Mediastinal Granuloma and Fibrosis

Mediastinal lymph node involvement is common in patients with chronic granulomatous infections, such as histoplasmosis and tuberculosis. Infrequently, an excessive fibrotic response develops around a caseous focus within lymph nodes. In some cases, fibrosis causes a capsule 2 cm or greater in thickness to form around a caseous focus, giving rise to a mediastinal mass that is usually situated in a subcarinal or right paratracheal location. In other patients, the fibrosis invades or compresses adjacent structures and at times extends diffusely through the mediastinum. The latter condition, called fibrosing mediastinitis, is the most common nonmalignant cause of superior vena cava syndrome.[91] Depending on which structures the fibrosis impinges, patients may also exhibit bronchial obstruction, pulmonary arterial or venous obstruction, or esophageal obstruction.[91]

In some cases of fibrosing mediastinitis, histologic examination of a surgical or postmortem specimen reveals only dense fi-

brosis. In other cases, a small caseous focus of infection can be identified, almost always caused by infection with *Histoplasma capsulatum*.[91] The presumed mechanism in all cases is rupture of caseous material into the mediastinum, leading to an intense inflammatory reaction that heals with fibrosis.

Even when viable fungal organisms can be identified, the therapeutic response to amphotericin B is poor, and corticosteroids are generally ineffective as well. Some studies have suggested that ketoconazole may be helpful.[92] In some patients, surgical extirpation of the fibrotic mass is possible. Endovascular balloon angioplasty and placement of stents have been successful in the management of superior vena cava and pulmonary artery obstructions. In general, progression of the fibrosis is slow, and long-term survival is possible.

Anterior Mediastinal Masses

In addition to lymphomas, common causes of masses found in the anterior mediastinum are thymomas, teratomas, dermoid cysts, and retrosternal goiters.[88]

Thymomas Thymomas may be found in adults of any age. They may arise from epithelial or lymphocytic cell lines. Most thymomas are benign, but some behave as malignant lesions and exhibit local invasion into adjacent structures. Reports of distant metastases are rare. Calcification may be present at the perimeter of or throughout the thymoma, but its presence does not necessarily signify that the lesion is benign. As many as one quarter to one third of patients with thymomas have myasthenia gravis, and some of these patients will experience a remission of symptoms after removal of the tumor. Other rare paraneoplastic syndromes that are associated with thymomas include red cell aplasia, Cushing syndrome, Graves disease, carcinoid syndrome, and hypogammaglobulinemia. Because of their malignant potential, suspected thymomas should be surgically excised in most cases. Other thymic lesions include thymic carcinoma, thymic carcinoid, thymolipoma, and thymic cysts.

Dermoid cysts and teratomas Germ cell tumors are most common in young adults 18 to 25 years of age. They originate from germ cell rests that were deposited in the mediastinum during embryogenesis. Dermoid cysts are usually benign cystic tumors composed of epidermal and dermal tissue; on occasion, hair, bone, or teeth form within these lesions. Teratomas consist of cells from all three embryonic origins (i.e., ectodermal, mesodermal, and endodermal) and may be cystic (usually benign) or solid (usually malignant). As with thymomas, calcification may be present along the periphery of the tumor. A potential complication of dermoid cysts and cystic teratomas is rupture and subsequent discharge of the cyst contents into the mediastinum or tracheobronchial tree. Surgical excision is the treatment of choice. Other germ cell tumors of the mediastinum include seminoma and nonseminomatous malignant tumors.

Goiters Cervical goiters may extend inferiorly into the thorax.[93] On occasion, masses within the mediastinum, usually located anteriorly, that have no palpable cervical connection prove to be intrathoracic goiters. These lesions are usually nodular colloid goiters that arise from the lower pole or isthmus of the thyroid gland and extend down into the chest; occasionally, such lesions prove to be adenomas or malignant tumors.

Most patients with nodular colloid goiters are asymptomatic, but some lesions grow sufficiently large to compress the trachea, causing dyspnea and stridor. Dysphagia, vascular compression, and vocal cord paresis or paralysis are other potential presenting manifestations. Uptake of radiolabeled iodine by a mediastinal mass on thyroid scan is diagnostic, but this is a relatively infrequent finding because most of these lesions are nonfunctioning goiters. Certain characteristic features of the CT image, including anatomic continuity with the cervical thyroid gland and particular patterns of calcification, have proved useful in identifying intrathoracic goiters. Patients with small, asymptomatic intrathoracic goiters can be observed or possibly given suppressive thyroid therapy; however, symptomatic patients require thyroidectomy, which can usually be achieved with a suprasternal incision.

MIDDLE MEDIASTINAL MASSES

Mediastinal bronchogenic cysts usually originate near the main carina of the tracheobronchial tree, but they may extend into any of the three mediastinal compartments.[88] They are lined with respiratory epithelium and contain a milky-white or brown mucoid material. Direct communication with the tracheobronchial tree is rare, although the potential exists for infection within the cyst and subsequent rupture into the airways or mediastinum. Bronchogenic cysts are usually discovered during childhood or early adulthood; in the latter age group, related symptoms are uncommon. Surgical excision is warranted for infected bronchogenic cysts or for cysts communicating with the tracheobronchial tree.

POSTERIOR MEDIASTINAL MASSES

Neurogenic tumors are the most common primary neoplasms of the posterior mediastinum.[88] They are round or oval masses in the paravertebral sulcus. Most patients are asymptomatic, although in some patients, chest or back pain or symptoms of bronchial compression develop. Tumors that arise from the nerve sheath (schwannomas) are usually benign; tumors derived from nerve cells may be benign (ganglioneuromas) or malignant (neuroblastomas or ganglioneuroblastomas). Neurofibromas derive from all nerve elements, including axons, sheath cells, and connective tissue; most are benign. Mediastinal neurofibromas may represent an isolated finding or may be part of generalized neurofibromatosis (von Recklinghausen disease). A rare tumor arising from paraganglionic cells is a mediastinal pheochromocytoma, which may be hormonally active.

Even benign neurogenic tumors may erode into an adjacent rib or cause pleural effusion. Some neurogenic tumors, most commonly neurofibromas, may extend into the spinal canal, causing widening of the intervertebral foramen; as a result, such tumors may assume a dumbbell shape. CT and MRI have for the most part supplanted myelography in the assessment of spinal extension of these so-called dumbbell tumors, which often require a coordinated thoracic and neurosurgical approach for resection.

Pneumomediastinum

Gas outside of normal GI structures may enter the mediastinum by several routes. First, it may collect in the mediastinum after esophageal or tracheobronchial rupture. Second, it may dissect into the mediastinum along fascial planes from the neck or oropharynx above (e.g., after dental extraction or tracheotomy) or from the retroperitoneum below (e.g., after colonic rupture or duodenal perforation). Third, after alveolar rupture,

gas may track along the perivascular interstitium in the lung and enter the mediastinum at the pulmonary hilum. This third pathway is the likely route of gas entry in patients subjected to barotrauma from mechanical positive pressure ventilation or in those who suffer pneumomediastinum as a complication of bronchoscopy. Pneumomediastinum occurring in the setting of mechanical ventilation may be associated with the more serious complications of barotrauma, which include pneumothorax, gas embolization, and pneumopericardium. Most often, however, pneumomediastinum occurs independently of traumatic or iatrogenic causes, in which case it is referred to as spontaneous pneumomediastinum.

Etiology

Spontaneous pneumomediastinum occasionally occurs as a complication of pneumonia or asthma, but most often, it is found in young and otherwise healthy individuals.[94] Symptoms are typically abrupt in onset and usually follow exaggerated respiratory efforts. Presumably, distended marginal alveoli rupture into the interstitium when the transpulmonary pressure is high; the elevated transpulmonary pressure would occur, for example, when a person inhales deeply or when a scuba diver holds his or her breath during ascent from an underwater depth. Thus, events that commonly precede spontaneous pneumomediastinum include parturition, cough, emesis, and straining at stool. In addition, numerous reports have appeared describing spontaneous pneumomediastinum after cocaine or marijuana use.

Diagnosis

Typical symptoms are retrosternal chest pain and dyspnea. The chest pain mimics that of pericarditis in that it is improved by sitting up and leaning forward; coughing, swallowing, and deep inspiration generally aggravate the pain.[94] Because the free mediastinal gas often escapes cephalad into the subcutaneous tissues of the neck and supraclavicular area, patients may also complain of neck pain and sore throat, dysphagia, or a peculiar swelling and crepitation in the upper chest and neck. Physical findings may include a crunchlike sound heard in synchrony with the heartbeat (Hamman mediastinal crunch), diminished dullness on percussion of the heart, and subcutaneous emphysema. The last finding, which is detected by palpation of crepitations resulting from the formation of gas bubbles just below the surface of the skin, may in extreme cases of pneumomediastinum extend as far down as the arms, abdominal wall, and genitals. Gas in the mediastinum may also decompress into the pleural space, leading to a concomitant pneumothorax, usually on the left side.

The diagnosis is established with chest radiography.[95] On posteroanterior view, a lucent zone of up to a few centimeters in width separates the cardiac silhouette from the medial border of the parietal and visceral pleurae. Gas may be seen outlining the aortic knob and extending in linear streaks up into the neck. On lateral view, gas collects between the sternum and the anterior border of the heart and outlines the aorta and other mediastinal soft tissue structures. In more subtle cases, lateral views of the neck reveal gas in the pretracheal fascia**.** In mild cases, CT may be necessary to detect the air in the mediastinum.[96]

Treatment

Spontaneous pneumomediastinum in spontaneously breathing adults resolves without specific therapy within a few days, and recurrences are uncommon. In contrast to spontaneous pneumothoraces, spontaneous pneumomediastinum is not recurrent. Thus, other than treating any underlying lung disease, management consists of reassurance, avoidance of strenuous activities, and use of analgesics if needed. Breathing-enhanced concentrations of oxygen may accelerate the rate of gas resorption. In adults, it is exceedingly rare that gas within the mediastinum is of sufficient pressure to compress vascular structures; therefore, surgical decompression of the mediastinal space by techniques such as tracheotomy is not indicated.

Gerald W. Staton, Jr., M.D., participates in the speakers' bureaus of GlaxoSmithKline and InterMune Pharmaceuticals, Inc.

Roland H. Ingram, Jr., M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Miserocchi G: Physiology and pathophysiology of pleural fluid turnover. Eur Respir J 10:219, 1997
2. Grove CS, Lee YC: Vascular endothelial growth factor: the key mediator in pleural effusion formation. Curr Opin Pulm Med 8:294, 2002
3. Lee YC, Lane KE: The many faces of transforming growth factor-beta in pleural diseases. Curr Opin Pulm Med 7:173, 2001
4. Muller NL: Imaging of the pleura. Radiology 186:297, 1993
5. Aquino SL, Webb WR, Gushiken BJ: Pleural exudates and transudates: diagnosis with contrast-enhanced CT. Radiology 192:803, 1994
6. Aleman C, Alegre J, Armadans L, et al: The value of chest roentgenography in the diagnosis of pneumothorax after thoracentesis. Am J Med 107:340, 1999
7. Tarver RD, Broderick LS, Conces DJ Jr: Reexpansion pulmonary edema. J Thorac Imaging 11:198, 1996
8. Villena V, Lopez-Encuentra A, Pozo F, et al: Measurement of pleural pressure during therapeutic thoracentesis. Am J Respir Crit Care Med 162:1534, 2000
9. Light RW: Diagnostic principles in pleural disease. Eur Respir J 10:476, 1997
10. Vaz MA, Marchi E, Vargas FS: Cholesterol in the separation of transudates and exudates. Curr Opin Pulm Med 7:183, 2001
11. Villena V, Perez V, Pozo F, et al: Amylase levels in pleural effusions: a consecutive unselected series of 841 patients. Chest 121:470, 2002
12. Martinez-Garcia MA, Cases-Viedma E, Cordero-Rodriguez PJ, et al: Diagnostic utility of eosinophils in the pleural fluid. Eur Respir J 15:166, 2000
13. Heffner JE, Brown LK, Barbieri C, et al: Pleural fluid chemical analysis in parapneumonic effusions: a meta-analysis. Am J Respir Crit Care Med 151:1700, 1995
14. Joseph J, Sahn SA: Connective tissue diseases and the pleura. Chest 104:262, 1993
15. Blanc FX, Atassi K, Bignon J, et al: Diagnostic value of medical thoracoscopy in pleural disease: a 6-year study. Chest 121:1677, 2003
16. Ferrer JS, Munoz XG, Orriols RM, et al: Evolution of idiopathic pleural effusion: a prospective, long-term follow-up study. Chest 109:1508, 1996
17. Romero-Candeira S, Fernandez C, Martin C, et al: Influence of diuretics on the concentration of proteins and other components of pleural transudates in patients with heart failure. Am J Med 110:681, 2001
18. Gotsman I, Fridlender Z, Meirovitz A, et al: The evaluation of pleural effusions in patients with heart failure. Am J Med 111:375, 2001
19. Lazaridis KN, Frank JW, Krowka MJ, et al: Hepatic hydrothorax: pathogenesis, diagnosis, and management. Am J Med 107:262, 1999
20. Xiol X, Castellote J, Cortes-Beut R, et al: Usefulness and complications of thoracentesis in cirrhotic patients. Am J Med 111:67, 2001
21. Milanez de Campos JR, Filho LO, de Campos Werebe E, et al: Thoracoscopy and talc poudrage in the management of hepatic hydrothorax. Chest 118:13, 2000
22. Sahn SA: Management of malignant pleural effusions. Monaldi Arch Chest Dis 56:394, 2001
23. Ang P, Tan EH, Leong SS, et al: Primary intrathoracic malignant effusion: a descriptive study. Chest 120:50, 2001
24. Sterman DH, Kaiser LR, Albelda SM: Advances in the treatment of malignant pleural mesothelioma. Chest 116:504, 1999
25. Porcel JM, Vives M: Etiology and pleural fluid characteristics of large and massive effusions. Chest 124:978, 2003
26. Wick MR, Moran CA, Mills SE, et al: Immunohistochemical differential diagnosis of pleural effusions, with emphasis on malignant mesothelioma. Curr Opin Pulm Med 7:187, 2001
27. Adams RF, Glesson FV: Percutaneous image-guided cutting-needle biopsy of the pleura in the presence of a suspected malignant effusion. Radiology 219:510, 2001
28. Poe RH, Levy PC, Israel RH, et al: Use of fiberoptic bronchoscopy in the diagnosis of bronchogenic carcinoma: a study in patients with idiopathic pleural effusions. Chest 105:1663, 1994
29. Burrows CM, Mathews WC, Colt HG: Predicting survival in patients with recurrent symptomatic malignant pleural effusions: an assessment of the prognostic values of

physiologic, morphologic, and quality of life measures of extent of disease. Chest 117:73, 2000

30. Parulekar W, Di Primio G, Matzinger F, et al: Use of small-bore vs large-bore chest tubes for treatment of malignant pleural effusions. Chest 120:19, 2001
31. Davies CW, Traill ZC, Gleeson FV, et al: Intrapleural streptokinase in the management of malignant multiloculated pleural effusions. Chest 115:729, 1999
32. Olivares-Torres CA, Laniado-Laborin R, Chavez-Garcia C, et al: Iodopovidone pleurodesis for recurrent pleural effusions. Chest 122:581, 2002
33. Walker-Renard PB, Vaughan LM, Sahn SA: Chemical pleurodesis for malignant pleural effusions. Ann Intern Med 120:56, 1994
34. Belani CP, Pajeau TS, Bennett CL: Treating malignant pleural effusions cost consciously. Chest 113 (suppl):78S, 1998
35. Diacon AH, Wyser C, Bolliger CT, et al: Prospective randomized comparison of thoracoscopic talc poudrage under local anesthesia versus bleomycin instillation for pleurodesis in malignant pleural effusions. Am J Respir Crit Care Med 162:1445, 2000
36. Pollak JS: Malignant pleural effusions: treatment with tunneled long-term drainage catheters. Curr Opin Pulm Med 8:302, 2002
37. Schulze M, Boehle AS, Kurdow R, et al: Effective treatment of malignant pleural effusion by minimal invasive thoracic surgery: thoracoscopic talc pleurodesis and pleuroperitoneal shunts in 101 patients. Ann Thorac Surg 71:1809, 2001
38. Villegas MV, Labrada LA, Saravia NG: Evaluation of polymerase chain reaction, adenosine deaminase, and interferon-gamma in pleural fluid for the differential diagnosis of pleural tuberculosis. Chest 118:1355, 2000
39. Nagesh BS, Shegal S, Jindal SK, et al: Evaluation of polymerase chain reaction for detection of *Mycobacterium tuberculosis* in pleural fluid. Chest 119:1737, 2001
40. Canete C, Galarza I, Granados A, et al: Tuberculous pleural effusion: experience with six months of treatment with isoniazid and rifampicin. Thorax 49:1160, 1994
41. Matchaba PT, Volmink J: Steroids for treating tuberculous pleurisy. Cochrane Database Syst Rev (2):CD001876, 2000
42. Worsley D, Alavi A, Aronchick J, et al: Chest radiographic findings in patients with acute pulmonary embolism: observations from the PIOPED study. Radiology 189:133, 1993
43. Shah AA, Davis SD, Gamsu G, et al: Parenchymal and pleural findings in patients with and patients without acute pulmonary embolism detected at spiral CT. Radiology 211:147, 1999
44. Light RW: Pleural effusion due to pulmonary emboli. Curr Opin Pulm Med 7:198:2001
45. Joseph J, Sahn SA: Thoracic endometriosis syndrome: new observations from an analysis of 110 cases. Am J Med 100:164, 1996
46. Martinez FJ, Villanueva AG, Pickering R, et al: Spontaneous hemothorax: report of 6 cases and review of the literature. Medicine (Baltimore) 71:354, 1992
47. Landreneau RJ, Keenan RJ, Hazelrigg SR, et al: Thoracoscopy for empyema and hemothorax. Chest 109:18, 1996
48. Jerjes-Sanchez C, Ramirez-Rivera A, Elizalde JJ, et al: Intrapleural fibrinolysis with streptokinase as an adjunctive treatment in hemothorax and empyema: a multicenter trial. Chest 109:1514, 1996
49. Hillerdal G: Chylothorax and pseudochylothorax. Eur Respir J 10:1157, 1997
50. Taylor JR, Ryu J, Colby TV, et al: Lymphangioleiomyomatosis: clinical course in 32 patients. N Engl J Med 323:1254, 1990
51. Garcia-Zamalloa A, Ruiz-Irastorza G, Aguayo FJ, et al: Pseudochylothorax: report of 2 cases and review of the literature. Medicine (Baltimore) 78:200, 1999
52. Varkey B, Rose HD, Kutty CP, et al: Empyema thoracis during a ten-year period: analysis of 72 cases and comparison to a previous study (1952 to 1967). Arch Intern Med 141:1771, 1981
53. Chen KY, Hsueh PR, Liaw YS, et al: A 10-year experience with bacteriology of acute thoracic empyema: emphasis on *Klebsiella pneumoniae* in patients with diabetes mellitus. Chest 117:1685, 2000
54. Ko SC, Chen KY, Hsueh PR, et al: Fungal empyema thoracis: an emerging clinical entity. Chest 117:1672, 2000
55. Lim TK: Management of parapneumonic pleural effusion. Curr Opin Pulm Med 7:193, 2001
56. Bouros D, Schiza S, Tzanakis N, et al: Intrapleural urokinase versus normal saline in the treatment of complicated parapneumonic effusions and empyema: a randomized, double-blind study. Am J Respir Crit Care Med 159:37, 1999
57. Cameron R, Davies H: Intra-pleural fibrinolytic therapy vs. conservative management in the treatment of parapneumonic effusions and empyema. Cochrane Database Syst Rev (3):CD002312, 2000
58. Bouros D, Schiza S, Patsourakis G, et al: Intrapleural streptokinase versus urokinase in the treatment of complicated parapneumonic effusions: a prospective, double-blind study. Am J Respir Crit Care Med 155:291, 1997
59. Moulton JS, Benkert RE, Weisiger KH, et al: Treatment of complicated pleural fluid collections with image-guided drainage and intracavitary urokinase. Chest 108:1252, 1995
60. Park CS, Chung WM, Lim MK, et al: Transcatheter instillation of urokinase into loculated pleural effusion: analysis of treatment effect. AJR Am J Roentgenol 167:649, 1996
61. Waller DA: Thoracoscopy in management of postpneumonic pleural infections. Curr Opin Pulm Med 8:323, 2002
62. Coote N: Surgical versus non-surgical management of pleural empyema. Cochrane Database Syst Rev (2):CD001956, 2002
63. Maruyama R, Ondo K, Midami K, et al: Clinical course and management of patients undergoing open window thoracostomy for thoracic empyema. Respiration. 68:606, 2001
64. Van Cleemput J, De Raeve H, Verschakelen JA, et al: Surface of localized pleural plaques quantitated by computed tomography: no relation with cumulative asbestos exposure and no effect on lung function. Am J Respir Crit Care Med 163:705, 2001
65. Yates DH, Browne K, Stidolph PN, et al: Asbestos-related bilateral diffuse pleural thickening: natural history of radiographic and lung function abnormalities. Am J Respir Crit Care Med 153:301, 1996
66. Batra P, Brown K, Hayashi K, et al: Rounded atelectasis. J Thorac Imaging 11:187, 1996
67. Sahn SA, Heffner JE: Spontaneous pneumothorax. N Engl J Med 342:868, 2000
68. Afessa B: Pleural effusions and pneumothoraces in AIDS. Curr Opin Pulm Med 7:202, 2001
69. Flume PA: Pneumothorax in cystic fibrosis. Chest 123:217, 2003
70. Bense L, Lewander R, Eklund G, et al: Nonsmoking, non-alpha 1-antitrypsin deficiency-induced emphysema in nonsmokers with healed spontaneous pneumothorax, identified by computed tomography of the lungs. Chest 103:433, 1993
71. Bense L, Eklund G, Lewander R: Hereditary pulmonary emphysema. Chest 121:297, 2002
72. Fujino S, Inoue S, Tezuka N, et al: Physical development of surgically treated patients with primary spontaneous pneumothorax. Chest 116:899, 1999
73. Bense L, Eklund G, Wiman LG: Bilateral bronchial anomaly: a pathogenetic factor in spontaneous pneumothorax. Am Rev Respir Dis 146:513, 1992
74. Voge VM, Anthracite R: Spontaneous pneumothorax in the USAF aircrew population: a retrospective study. Aviat Space Environ Med 57:939, 1986
75. Seow A, Kazerooni EA, Pernicano PG, et al: Comparison of upright inspiratory and expiratory chest radiographs for detecting pneumothoraces. AJR Am J Roentgenol 166:313, 1996
76. Baumann MH, Strange C, Heffner JE, et al: Management of spontaneous pneumothorax: an American College of Chest Physicians Delphi consensus statement. Chest 119:590, 2001
77. Noppen M, Alexander P, Driesen P, et al: Manual aspiration versus chest tube drainage in first episodes of primary spontaneous pneumothorax: a multicenter, prospective, randomized pilot study. Am J Resp Crit Care Med 165:1240, 2002
78. Sadikot RT, Greene T, Meadows K, et al: Recurrence of primary spontaneous pneumothorax. Thorax 52:805, 1997
79. Sihoe AD, Yim AP, Lee TW, et al: Can CT scanning be used to select patients with unilateral primary spontaneous pneumothorax for bilateral surgery? Chest 118:380, 2000
80. Yim AP, Ng CS: Thoracoscopy in the management of pneumothorax. Curr Opin Pulm Med 7:210, 2001
81. Moon WK, Im JG, Yeon KM, et al: Mediastinal tuberculous lymphadenitis: CT findings of active and inactive disease. AJR Am J Roentgenol 170:715, 1998
82. Gupta NC, Graeber GM, Bishop HA: Comparative efficacy of positron emission tomography with fluorodeoxyglucose in evaluation of small (< 1 cm), intermediate (1 to 3 cm), and large (> 3 cm) lymph node lesions. Chest 117:773, 2000
83. Larson SS, Krasnik M, Vilmann P, et al: Endoscopic ultrasound guided biopsy of mediastinal lesions has a major impact on patient management. Thorax 57:98, 2002
84. Freeman RK, Vallieres E, Verrier ED, et al: Descending necrotizing mediastinitis: an analysis of the effects of serial surgical debridement on patient mortality. J Thorac Cardiovasc Surg 119:260, 2000
85. Combes A, Trouillet JL, Baudot J, et al: Is it possible to cure mediastinitis in patients with major postcardiac surgery complications? Ann Thorac Surg 72:1592, 2001
86. Brook I, Frazier EH: Microbiology of mediastinitis. Arch Intern Med 156:333, 1996
87. El Oakley RM, Wright JE: Postoperative mediastinitis: classification and management. Ann Thorac Surg 61:1030, 1996
88. Yoneda KY, Louie S, Shelton DK: Mediastinal tumors. Curr Opin Pulm Med 7:226, 2001
89. Morrissey B, Adams H, Gibbs AR, et al: Percutaneous needle biopsy of the mediastinum: review of 94 procedures. Thorax 48:632, 1993
90. Link KM, Samuels LJ, Reed JC, et al: Magnetic resonance imaging of the mediastinum. J Thorac Imaging 8:34, 1993
91. Mole TM, Glover J, Sheppard MN: Sclerosing mediastinitis: a report on 18 cases. Thorax 50:280, 1995
92. Urschel HC Jr, Razzuk MA, Netto GJ, et al: Sclerosing mediastinitis: improved management with histoplasmosis titer and ketoconazole. Ann Thorac Surg 50:215, 1990
93. Buckley JA, Stark P: Intrathoracic mediastinal thyroid goiter: imaging manifestations. AJR Am J Roentgenol 173:471, 1999
94. Abolnik I, Lossos IS, Breuer R: Spontaneous pneumomediastinum: a report of 25 cases. Chest 100:93, 1991
95. Bejvan SM, Godwin JD: Pneumomediastinum: old signs and new signs. AJR Am J Roentgenol 166:1041, 1996
96. Kaneki T, Kubo K, Kawashima A, et al: Spontaneous pneumomediastinum in 33 patients: yield of chest computed tomography for the diagnosis of the mild type. Respiration 67:408, 2000
97. Light RW: Pleural effusions after coronary artery bypass graft surgery. Curr Opin Pulm Med 8:308, 2002
98. Jung JI, Choi JE, Hahn ST, et al: Radiologic features of all-trans-retinoic acid syndrome. AJR Am J Roentgenol 178:475, 2002

X PULMONARY EDEMA

Gerald W. Staton, Jr., m.d., f.a.c.p.

Acute pulmonary edema can be divided into two categories: edema caused by increased capillary pressure (hydrostatic, or cardiogenic, edema) and edema caused by increased capillary permeability (noncardiogenic pulmonary edema, or acute respiratory distress syndrome [ARDS]). In some cases of pulmonary edema, both pressure and permeability are increased.

Pathogenesis

Pulmonary edema is an abnormal, diffuse, extravascular accumulation of liquid in the pulmonary tissues and air spaces. It is the most common noninfectious, diffuse parenchymal lung disease. The pathogenesis of pulmonary edema can be best understood by examination of the Starling equation, which defines the determinants of liquid flux across the pulmonary capillary membrane:

$$Q = K[(P_{cap} - P_{int}) - \sigma(\pi_{cap} - \pi_{int})]$$

where Q is liquid flux; K, the filtration coefficient, which is directly proportional to the endothelial surface area and inversely proportional to alveolar capillary wall thickness; P_{cap}, the intravascular (capillary) hydrostatic pressure; P_{int}, the interstitial hydrostatic pressure; σ, the reflection coefficient for protein (i.e., the degree of permeability to macromolecules); π_{cap}, the plasma oncotic pressure; and π_{int}, the interstitial oncotic pressure. In this equation, $P_{cap} - P_{int}$ represents the hydrostatic force, and $\pi_{cap} - \pi_{int}$ represents the colloid osmotic force.

An alteration in any of the factors in the Starling equation could conceivably lead to an increase in transvascular liquid flux. However, in clinical practice, only two of these factors commonly lead to pulmonary edema: (1) an increase in capillary pressure, which leads to cardiogenic, or hydrostatic, pulmonary edema, and (2) a decrease in the reflection coefficient, which leads to noncardiogenic, or permeability, pulmonary edema.

The primary defense against pulmonary edema is provided by the lymphatic system. Normally, liquid that is filtered across the capillary membrane is removed by the lymphatic vessels. The lymphatic reserve is such that even a fourfold to sixfold increase in transcapillary liquid flux can be tolerated without an increase in lung water. A secondary defense mechanism and pathway for the removal of edema liquid is the active transport of Na^+ with passive osmotic water reabsorption in type II alveolar lining cells.[1]

Once the liquid removal reserve is overwhelmed, lung water increases, first in the interstitium around the airways, then in the interstitium around the alveoli, and finally within the alveoli. Clinical or radiographic evidence of pulmonary edema therefore implies a large increase in transcapillary liquid flux.

Approach to the Patient with Suspected Pulmonary Edema

Because of the different therapeutic approaches in cardiogenic and noncardiogenic pulmonary edema, it is important to differentiate between these conditions, though this is sometimes difficult, and some patients have components of both. The patient's history provides some important clues [*see Table 1*]. Patients with cardiogenic pulmonary edema often have a history of cardiac disease or hypertension, whereas patients with noncardiogenic edema may have a history of alcoholism or other problems that increase the risk of infection. The patient with cardiogenic pulmonary edema may have symptoms of a new cardiac event (ischemic chest pain) or a hypertensive emergency. The patient with noncardiogenic pulmonary edema will almost always have a clear precipitating event, such as sepsis, trauma, aspiration of gastric contents, multiple transfusions, or pneumonia.

The symptoms and signs are similar in the two forms of pulmonary edema, with some important exceptions [*see Table 1*]. Although both groups of patients experience dyspnea, patients with cardiogenic pulmonary edema have cool, diaphoretic skin

Table 1 Cardiogenic versus Noncardiogenic Pulmonary Edema

Clinical Features	*Cardiogenic Pulmonary Edema*	*Noncardiogenic Pulmonary Edema*
Clinical setting	Prior MI, hypertension, cardiomyopathy, MI, hypertensive emergency, dietary indiscretion, others	Alcoholism, immunosuppression, sepsis, trauma, aspiration, pneumonia, others
Symptoms	Dyspnea, cough productive of pink sputum	Dyspnea, cough
Signs	Tachypnea, tachycardia, weak pulse, hypertension, cool skin, diaphoresis, central or peripheral cyanosis, wheezing, rales, LV heave, gallop	Tachypnea, fever, hypothermia, bounding pulse, warm skin, central cyanosis, hyperdynamic precordium
Findings on chest radiography	Venous cephalization, wide vascular pedicle, increased cardiothoracic ratio, perihilar infiltrates, Kerley B lines	Diffuse infiltrates, air bronchograms
Findings on echocardiography	Decreased LV ejection fraction, diastolic dysfunction, valvular disease, segmental wall motion abnormalities	Hyperdynamic LV, RV dilatation, pulmonary hypertension
Findings on pulmonary artery catheterization	Increased PCWP, normal or decreased CO, increased pulmonary arterial pressure with normal PVR, increased SVR	Normal or low PCWP; when sepsis is present: increased CO, increased pulmonary arterial pressure with increased PVR, decreased SVR

CO—cardiac output LV—left ventricular MI—myocardial infarction PCWP—pulmonary capillary wedge pressure PVR—pulmonary vascular resistance RV—right ventricular SVR—systemic vascular resistance

with evidence of left ventricular (LV) dysfunction (LV heave, gallop); by contrast, most patients with noncardiogenic pulmonary edema have warm skin and show evidence of a hyperdynamic circulation.

Pulmonary edema of either type is typically manifested by bilateral, symmetrical alveolar opacities that involve all four quadrants, as seen on a standard anteroposterior chest radiograph [*see Table 1*].[2] A predominantly perihilar distribution is common, and occasionally, there is a very sharp demarcation between the central area of pulmonary edema and the lung periphery, leading to a so-called bat's-wing or butterfly pattern. This pattern is more typical of cardiogenic pulmonary edema than noncardiogenic pulmonary edema [*see Figure 1*]. This sharp line of demarcation does not correspond to any anatomic boundaries but may be caused by physiologic gradients of ventilation, perfusion, and lymph flow from the central to the peripheral portions of the lung. Less often, the edema is markedly asymmetrical or entirely unilateral. Asymmetrical pulmonary edema can occur as a result of the patient's lying on the involved side as the edema develops, or it may result from vascular obstruction (thromboembolism) or attenuation (emphysema) in the more radiolucent regions of the lung. However, the reasons why some cases are asymmetrical and others are unilateral are often impossible to determine.

Ancillary features that can be routinely visualized on an anteroposterior chest radiograph made with a portable x-ray machine may help differentiate cardiogenic from noncardiogenic pulmonary edema. A widened vascular pedicle and an increase in the cardiothoracic ratio suggest increased pulmonary capillary pressure[3]; distinct air bronchograms are more common with noncardiogenic pulmonary edema. The presence or absence of pleural effusions have less value in making the differential diagnosis. Unfortunately, cardiogenic and noncardiogenic pulmonary edema often cannot be confidently differentiated using bedside radiography, owing to anteroposterior exposure and lack of full inspiration, both of which magnify the cardiac silhouette.

Bedside echocardiography can be very useful in differentiating cardiogenic from noncardiogenic pulmonary edema [*see Table 1*]. In patients with cardiogenic pulmonary edema, echocardiography can detect and quantitate the abnormal left ventricular function and can be used to diagnose many of the causes (e.g., valvular dysfunction, diastolic dysfunction, cardiomyopathy, and focal wall motion abnormalities). In patients with noncardiogenic pulmonary edema, a normal or hyperdynamic left ventricle is seen.

If differentiation of cardiogenic from noncardiogenic pulmonary edema is not possible with the noninvasive evaluation outlined above, placement of a pulmonary arterial catheter may be considered. Pulmonary capillary wedge pressure (PCWP) is the most helpful measurement, but other measurements can support the diagnosis and may help in treating the patient [*see Table 1*].[4]

Despite the logical appeal of the use of pulmonary arterial catheters, no beneficial effect on outcome has been attributed to their use. A study of a large number of patients in intensive care units has suggested that patients who had pulmonary arterial catheters had a higher mortality at a higher financial cost than patients who did not undergo catheterization.[5] The report has been criticized but is currently the only published analysis of the effect of the use of these catheters on the outcome of patients in the ICU.

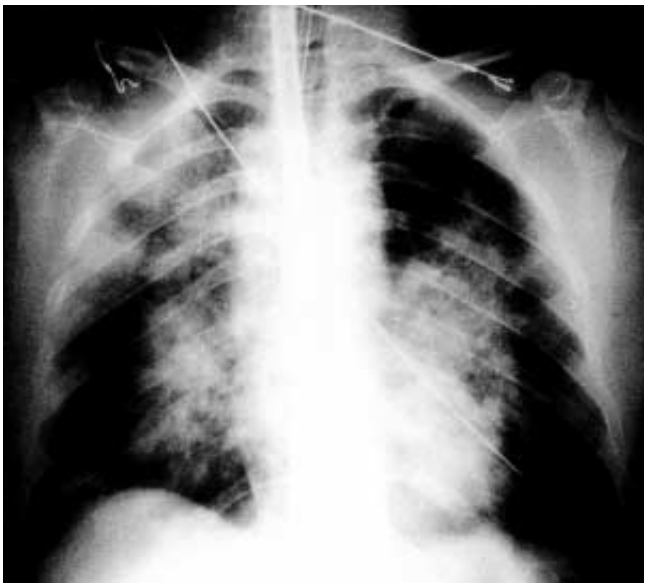


Figure 1 **Bat's-wing pattern characteristic of acute pulmonary edema is evident in this chest radiograph.**

Cardiogenic Pulmonary Edema

PATHOGENESIS

Pulmonary edema caused by increased capillary pressure can occur as a result of systolic or diastolic dysfunction of the left ventricle, mitral valvular disease, hypervolemia associated with normal left heart function (as might occur in a patient with renal failure), or pulmonary venous obstruction. The most common cause of cardiogenic pulmonary edema is left ventricular dysfunction. In congestive cardiomyopathy, the systolic performance of the left ventricle is impaired, the ventricle is dilated, and left ventricular end-diastolic pressure (LVEDP) is increased. The rise in LVEDP leads to an increase in pulmonary capillary pressure. Other types of heart disease can also increase LVEDP, despite normal systolic function and euvolemia, by reducing left ventricular compliance and producing diastolic dysfunction.[6] Reduced compliance may be persistent (as seen with left ventricular hypertrophy or restrictive cardiomyopathy from infiltrative heart disease) or transient (as from myocardial ischemia). Increased capillary pressure despite normal LVEDP is uncommon but does occur with mitral stenosis or as a result of obstructed flow in the pulmonary veins (pulmonary venoocclusive disease).

DIAGNOSIS

The salient clinical features of cardiogenic pulmonary edema are extreme breathlessness, tachypnea, and signs of increased sympathetic activity, such as tachycardia, hypertension, and diaphoresis [*see Table 1 and* Approach to the Patient with Suspected Pulmonary Edema, *above*]. Hypotension is uncommon but may occur if pulmonary edema results from a large myocardial infarction. Breathlessness is rarely alleviated by correction of hypoxemia, suggesting that the cause may be activation of intrapulmonary stretch receptors rather than hypoxemia. In cardiogenic pulmonary edema, central cyanosis may be observed if there is profound arterial hypoxemia; more often, cyanosis is peripheral and results from intense cutaneous vasoconstriction and a decreased cardiac output. Use of accessory muscles of res-

piration is common because of the marked increase in the work of breathing. The effort required to breathe is often so great that endotracheal intubation and mechanical ventilation are required to correct or prevent the development of frank hypercapnic respiratory failure. Early in the acute course, there may be wheezing caused by airway edema and intraluminal liquid; later, diffuse coarse rales are heard.

TREATMENT

Supplemental oxygen should be given by mask or nasal cannula. If hypoxemia cannot be corrected by establishing maximum oxygen flow rates and by use of reservoir bags, mechanical ventilation with a mask[7] or endotracheal tube will be required [*see 14:VIII Respiratory Failure*]. The positive intrathoracic pressures created by the ventilator open collapsed alveoli and impede venous return. While hypoxemia is being treated, one or more of several therapeutic options should simultaneously be exercised, depending on the underlying pathophysiologic processes [*see Table 2*]. The proper combination of therapeutic measures depends on the pathophysiology. For example, when pulmonary edema occurs as a complication of malignant hypertension, vasodilators and diuretics may suffice. In the case of a causative or strongly contributing tachyarrhythmia, antiarrhythmic therapy may be the key intervention. Specific interventions for causative cardiac conditions are described elsewhere [*see Section 1 Cardiovascular Medicine*].

OUTCOME

Patients with acute cardiogenic pulmonary edema often are elderly and have multiple medical problems, including ischemic heart disease, diabetes, and valvular heart disease. As a consequence, the mortality for these patients ranges from 6% to 30%.[8] Of those patients who survive to discharge, approximately 70% will survive 1 year, and 50% will have relatively good functional status for longer periods.[9]

Noncardiogenic Pulmonary Edema: Acute Respiratory Distress Syndrome

ARDS is characterized by diffuse pulmonary endothelial injury, which leads to pulmonary edema as a result of an increase in capillary permeability to water, solutes, and macromolecules.[10,11] The pulmonary edema seen in patients with ARDS is characterized by a higher concentration of protein in the edema liquid than is seen in patients with cardiogenic pulmonary edema. In patients with ARDS, this concentration is often as high as 80% to 90% of the plasma protein. Furthermore, in ARDS patients, the underlying inflammatory response causes high levels of neutrophils and their secretory products in bronchoalveolar lavage liquid; this characteristic distinguishes noncardiogenic edema from cardiogenic edema.

Table 2 Treatment of Cardiogenic Pulmonary Edema

Effect Sought	*Therapeutic Approach*
Decrease venous return	Venodilator
Decrease impedance to ventricular systole	Arteriolar dilator
Decrease intravascular volume	Diuretic agent
Stimulate the myocardium	Inotropic agent
Correct arrhythmias	Antiarrhythmic agent, pacemaker
In the Presence of Coronary Arterial Occlusion	
Alleviate ischemia	Angioplasty, thrombolytic agent, coronary bypass
Prevent clot propagation	Aspirin, other antiplatelet agent, anticoagulant

Table 3 Causes of ARDS[11]

Condition	*Examples*
Direct Lung Injury	
Diffuse pulmonary infection	Bacterial, viral (including SARS[53]) or fungal pneumonia; *Pneumocystis carinii* pneumonia; tuberculosis
Chemical pneumonitis caused by aspiration	Aspiration of gastric contents or of water in near-drowning
Inhalation injury	Inhalation of smoke, chlorine, or nitrous oxide
Direct pulmonary trauma	Automobile accident
Indirect Lung Injury	
Systemic reaction to nonpulmonary infection	Bacteremias, nonbacteremic sepsis, toxic-shock syndrome
Systemic reaction to nonpulmonary tissue inflammation or injury	Pancreatitis, trauma, fat embolism, amniotic fluid embolism
Transfusion reaction	Red cell transfusion
Drug toxicity	Salicylates, cytotoxic agents
Reperfusion injury	Cardiopulmonary bypass, post–lung transplantation
Other	Marathon running

SARS—severe acute respiratory syndrome

Many clinical disorders are associated with the development of ARDS [*see Table 3*], which may arise from direct injury to the lung or from extrapulmonary processes that injure the lung indirectly.[10,12] A patient's risk of developing ARDS varies with the predisposing disorder, and the risk increases as the number of predisposing disorders increases.[11] The risk of ARDS may be further increased in patients with a history of alcohol abuse[13] or cigarette smoking,[14] or it may be increased by the presence of low serum pH or hypoproteinemia[15] at the time of the insult. Genetic factors may also play a role in predisposition.[16]

PATHOGENESIS

Diffuse alveolar damage is a descriptive term for the nonspecific but predictable sequence of changes that characteristically occur in patients with ARDS.[17] The causative agent or process usually cannot be determined from the histopathologic pattern. In addition, the abnormalities may resolve at any point in the clinical course.

The histologic appearance of diffuse alveolar damage varies during the period between the precipitating event and the biopsy or autopsy, progressing through the following three phases: an acute exudative phase (days 0 through 7); a subacute proliferative, or organizing, phase (days 7 through 14); and a chronic phase (after day 14).[17]

The earliest part of the acute exudative phase is characterized by interstitial and intra-alveolar edema, neutrophil infiltration, hemorrhage, and fibrin deposition. A mixture of fibrin and cellular debris is deposited in the alveolar space to form the so-called hyaline membranes that are prominent 3 to 7 days after injury. Sloughing of the cells of the alveolar lining leaves a de-

nuded basement membrane, which plays an important role in subsequent repair or fibrosis. An interstitial infiltrate of inflammatory cells becomes more pronounced around day 7 and persists throughout the proliferative phase.[17]

Denudation of the basement membrane causes type II pneumocytes to proliferate (days 3 through 7), producing a pattern of hyperplasia in the cells of the alveolar lining. In patients in whom the syndrome resolves, these proliferating cells ultimately differentiate into type I pneumocytes, restoring the epithelial side of the alveolocapillary wall and returning gas exchange to normal.[17]

The proliferative phase of ARDS is characterized by inflammation and fibroblast proliferation, initially in the interstitium. The fibroblasts invade the alveolar spaces through basement membrane defects, a process that produces regions of intra-alveolar fibrosis. During this phase, the hyaline membranes disappear as a result of phagocytosis or of organization involving the incorporation of the exudate into intra-alveolar plugs of proliferating fibroblasts.[17]

The chronic phase of ARDS is characterized by regions of intense fibrosis, focal regions of overexpansion, and pulmonary vascular obliteration. Histologically, this phase of the disease can be similar to idiopathic pulmonary fibrosis; in contrast to idiopathic pulmonary fibrosis, however, the chronic phase of ARDS may improve with time.[17]

Extensive investigations have led to a better understanding of the mechanisms that lead to ARDS[10,17,18] [*see Figure 2*].

DIAGNOSIS

Clinical Manifestations

The major clinical signs of noncardiogenic pulmonary edema overlap those of cardiogenic pulmonary edema [*see Table 1 and* Approach to the Patient with Suspected Pulmonary Edema, *above*]. In noncardiogenic pulmonary edema, however, there is often less sympathetic activity[19]; cyanosis, if present, is often caused by arterial hypoxemia. The skin may be warm (rather than cool, moist, and pallid), and the pulse may be racing.

Imaging Studies

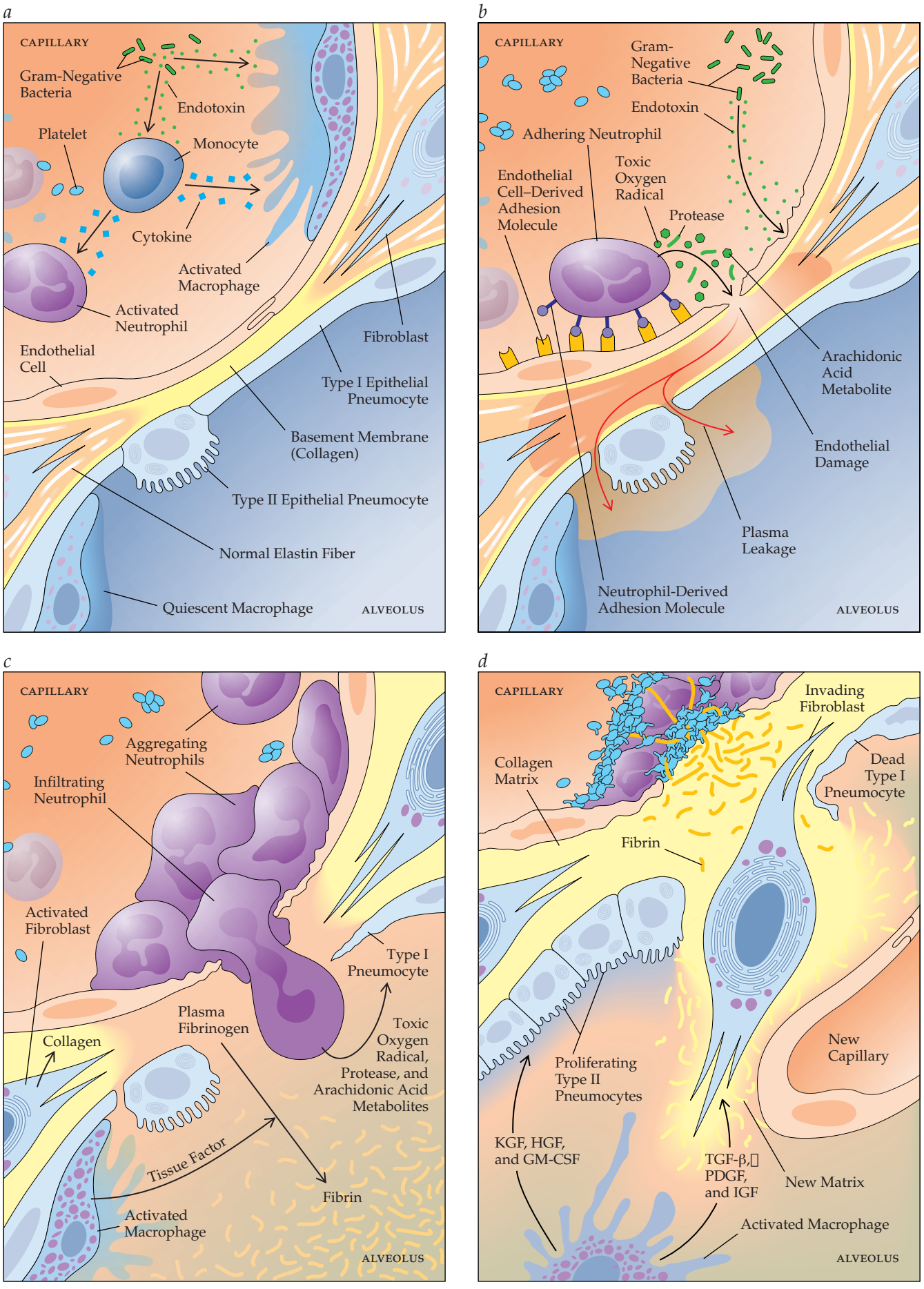
Typically, portable anteroposterior chest radiography reveals a diffuse and homogeneous alveolar filling process.[20] When examined by computed tomography, however, the air-space filling pattern frequently appears less homogeneous. Radiographs with the patient in the supine position typically show a greater degree of consolidation in posterior lung zones than in anterior lung zones; this distribution, however, may be reversed by placing the patient in the prone position for a few hours. With the patient in the prone position, the anterior regions become more consolidated, because of the influence of gravity, and the posterior portions of the lung show improved aeration. This finding demonstrates the contribution of Starling forces (i.e., capillary pressure) to the severity of edema in the setting of increased permeability.

Physiologic Testing

Gas exchange in ARDS is characterized initially by hypoxemia that is refractory to increasing concentrations of inspired oxygen, implying the presence of increased intrapulmonary shunting. Intrapulmonary shunting is primarily a consequence of diffuse alveolar filling, collapse, or both, with microatelectasis. Diffuse alveolar collapse is believed to be caused by abnormalities in surfactant, a substance that normally helps maintain alveolar distention by lowering the surface tension in the air-liquid interface of the lung.

Initially, arterial carbon dioxide tension (P_aCO_2) is either low or in the normal range with only a modest increase in minute ventilation. However, the ratio of dead space to tidal volume (V_D/V_T) tends to increase over time, so that increasing amounts of minute ventilation are required to achieve a normal P_aCO_2.[21] In some cases, the increase in V_D/V_T is so extreme that hypercapnia cannot be avoided even when minute ventilation is increased to the maximal achievable level. This increase in physiologic dead space results from damage to the pulmonary capillary bed, which creates regions of high ventilation relative to perfusion. A secondary cause of the increased minute-ventilation requirement in ARDS is the presence of increased carbon dioxide production from hypermetabolism. The high V_D/V_T in ARDS is slow to resolve; this, together with the reduction of lung compliance (which imposes a burden on spontaneous breathing), accounts in part for the frequent need for mechanical ventilatory support for prolonged periods. Such support often must be continued even though hypoxemia has improved sufficiently to make only a modest increase in inspired oxygen concentration necessary.

Figure 2 **(*a*) The first site of injury in ARDS is the endothelium of the pulmonary capillaries. In ARDS caused by sepsis, endotoxin stimulates the production or release of several proinflammatory mediators such as activated complement fragments, coagulation factors, platelet activating factor, eicosanoids, and cytokines/chemokines (e.g., tumor necrosis factor-α [TNF-α], interleukin-1 [IL-1], IL-6, IL-8, and other cytokines), as well as anti-inflammatory substances such as IL-10, IL-11, cytokine receptor antagonists, anticytokine autoantibodies, antioxidants, and antiproteases.[51] Release of some of these mediators may be induced by the mechanical ventilation that is often required in such cases, and these mediators may play a role in inducing the systemic inflammatory response syndrome and multiorgan failure.[52] If there is an imbalance in favor of the inflammatory mediators, the mediators act as primers and secretagogues for neutrophils and other monocytes/macrophages, resulting in aggregation and embolization of these cells in the pulmonary vasculature. (*b*) The neutrophils adhere to the endothelium by way of specific receptors on the neutrophils (integrins and L-selectin) and receptors on the endothelium (intercellular adhesion molecule–1 [ICAM-1] and E- and P-selectin) and release injurious oxidants, proteolytic enzymes, and arachidonic acid metabolites, resulting in endothelial cell dysfunction and destruction and in denudation of the endothelial side of the basement membrane. Endotoxin may also injure endothelial cells directly. Increased permeability of the alveolocapillary membrane allows plasma to enter the interstitial spaces of the lung and, ultimately, the alveoli. (*c*) The second site of injury is the alveolar epithelium. Neutrophils penetrate the alveolocapillary membrane and, along with activated macrophages, release oxidants and proteolytic enzymes. As a result, type I pneumocytes die, denuding the alveolar side of the basement membrane. Macrophages produce procoagulant substances, such as tissue factor and factor VII, that produce fibrin, and the alveolus becomes filled with proteinaceous exudate and cellular debris. (*d*) In response to the death of the type I pneumocytes, type II pneumocytes, under the control of growth factors (keratinocyte growth factor [KGF], hepatocyte growth factor [HGF], and granulocyte-monocyte colony-stimulating factor [GM-CSF] from macrophages), begin to proliferate and attempt to cover the denuded basement membrane. If the injury to the epithelial cells and the basement membrane is severe, progressive interstitial and intra-alveolar fibrosis ensues. Fibroblasts, which are controlled by growth factors (transforming growth factor–β [TGF-β], platelet-derived growth factor [PDGF], and insulinlike growth factor [IGF]) derived from macrophages, invade the exudate, where they proliferate and synthesize collagen (type I), elastin, and other new matrix components. The normal lung architecture is progressively replaced at this point (approximately 2 weeks from onset) by fibrous tissue that severely impairs gas exchange.**



Lung mechanics in ARDS is characterized primarily by a reduction in lung compliance (increased lung elastance); thus, high transpulmonary pressures are required to achieve normal tidal ventilation. Early in ARDS, when edema predominates, much of the distending pressure needed to inflate the lung is expended in opening collapsed alveoli. Indeed, lung compliance (i.e., the slope of the pressure-volume curve) may be in the normal range if it is measured after these collapsed alveoli have been opened. However, there is a significant reduction in compliance as the disease process evolves or when alveolar fibrosis becomes predominant. This apparent reduction may be caused by a diffuse thickening of the alveolocapillary membrane as a result of fibrosis or by the loss of a large percentage of the alveolocapillary units that are available for ventilation. It has been suggested that in patients with ARDS, lung mechanics is best conceptualized by regarding the lung as being small rather than stiff.[22] Increased airway resistance that responds to inhaled bronchodilators is also seen in ARDS.[23]

In addition to their effect on V_D/V_T, the pulmonary vascular changes in ARDS result in increased pulmonary vascular resistance and pulmonary hypertension.[24] Indeed, the pulmonary arterial pressure is elevated in nearly all patients who have moderate to severe ARDS. The etiology of pulmonary hypertension in ARDS is likely to be multifactorial; however, a major underlying cause seems to be the presence of small pulmonary arterial thrombi. The pathogenesis of the thrombi in ARDS is unknown, but it is likely that they are formed in situ.

Many patients with ARDS have certain markers of accelerated intravascular coagulation.[25] However, despite the apparent importance of thrombosis in producing pulmonary hypertension, it has not been determined whether traditional anticoagulation benefits patients with ARDS.

TREATMENT

The treatment of ARDS includes addressing the precipitating cause of ARDS, general ICU support (e.g., nutrition), prevention of complications in the ICU (e.g., stress ulcer prophylaxis), and management of edema[26] [*see Table 4*]. Ventilatory and cardiorespiratory issues in the management of respiratory failure in patients with ARDS are discussed in detail elsewhere [*see 14:VIII Respiratory Failure*].

Treatment of the cause of ARDS, when feasible, should be instituted as soon as possible. For example, in patients with ARDS that is associated with sepsis, appropriate antibiotics should be started immediately, and the source of the infection should be identified and treated (e.g., abscesses should be drained).

The initial management in the ICU should include several general measures. Prevention of stress ulceration and associated gastrointestinal bleeding, as well as prevention of deep vein thrombosis and pulmonary embolism, is indicated. Activated protein C (drotrecogin alfa [activated]), an endogenous protein with antithrombotic, profibrinolytic, and anti-inflammatory properties, should be administered to patients with sepsis who are not at risk for bleeding.[27] In septic patients with adrenal insufficiency, replacement doses of corticosteroids are indicated.[28] Attempts to reduce nosocomial infection, particularly ventilator-associated pneumonia, through the use of topical and systemic antibiotics have had favorable results,[29] but concerns about the induction of resistant organisms have made this type of therapy controversial. Enteral feeding is indicated if the GI tract is functioning; otherwise, parenteral nutrition should be provided. Special enteral formulations enriched with "immunonutrients" (e.g., arginine, glutamine, omega-3 fatty acids, and nucleotides) have been found to reduce length of hospital stay, infection rates, and duration of mechanical ventilation, but not mortality.[30]

The treatment of ARDS is focused on reducing existing pulmonary edema, preventing further edema, and modifying the evolution of the disease. Once the patient is hemodynamically stable, fluid restriction, diuresis (possibly with albumin infusion), or both are indicated.[31] Numerous attempts have been made to reduce the inflammatory response in ARDS. Except in patients with relative adrenal insufficiency, use of corticosteroids early in the course of disease has no beneficial effect; however, steroids given 7 to 14 days later may modify the fibroproliferative phase.[32] Numerous other agents have not been found to modify the disease process or improve outcome [*see Table 4*].[33,34]

OUTCOME

Although the prognosis of patients with ARDS is related to the degree of lung injury, other parameters more accurately predict outcome; this is not surprising, because ARDS is frequently part of a systemic inflammatory response syndrome. The cause of the syndrome is often unclear, but the possibilities include bronchopneumonia, translocation of bacterial products across the intestine, and persistent release of endogenous mediators in the absence of ongoing infection. Sepsis, which is often associated with vasodilatation that is unresponsive to vasoconstrictors, is the most common cause of death during the course of illness. As a result of state-of-the-art ventilatory-support techniques, respiratory failure is the cause of death in less than 20% of cases—a fact that highlights the importance of dysfunction of other organ systems (e.g., hemodynamic failure with refractory shock or progressive renal failure) in causing morbidity and mortality. Data suggest that the outcome of patients with ARDS may be improving, possibly because of improvements in therapy.[10,19]

One way to predict mortality is by the number of organ systems that fail. Mortality increases with the number of failing organs and the number of days of failure [*see Figure 3*].[35] There is a further increase in mortality in patients who are older than 65 years. For example, in patients younger than 65 years who have single-organ malfunction for 5 days, mortality is 27%, whereas in patients with single-organ malfunction who are older than 65 years, mortality is 48%. Likewise, the failure of two organs for 2 days results in a 47% mortality in patients younger than 65 years and a 73% mortality in patients older than 65 years. The failure of three or more organs for 5 days in patients of any age is associated with a 97% mortality.[35] These data provide a useful guide for decision making for patients with catastrophic illness and their families.

If patients survive the acute illness that causes ARDS, the prognosis for return of lung function is good. Factors associated with poor pulmonary functional outcome are severity of ARDS, lowest measured compliance, and duration of positive pressure ventilation.[10] Lung function improves rapidly over the first several weeks, then more slowly over a period of as long as 2 years. Common symptoms and signs include exertional dyspnea, cough, wheezing, and persistent rales. Pulmonary function tests may demonstrate the presence of restrictive disease; obstructive disease, often with increased airway reactivity; or decreased diffusing capacity. CT may show a persistent reticular pattern in 85% of patients.[36] One year or more after the onset of ARDS, more than 75% of patients either have normal respiratory function or suffer only mild impairment. However, many will suffer

Table 4 Treatments of ARDS That Do Not Involve Ventilation[26]

Treatment	*Purpose*	*Evidence Grade*	*Is This Treatment Recommended?*
Stress ulcer prophylaxis	Prevent stress ulcers and GI bleeding	A (studied in ICU patients)	Yes, gastric acid neutralization allowable
Prophylaxis of deep vein thrombosis and pulmonary embolism	Prevent deep vein thrombosis and pulmonary embolism	A (studied in ICU patients)	Yes
Activated protein C	Antithrombotic, profibrinolytic, anti-inflammatory	B	Yes, for patients with sepsis without risk of hemorrhage
Corticosteroids, replacement doses	Relative adrenal insufficiency	B	Yes, for patients with sepsis with relative adrenal insufficiency
Topical plus systemic prophylactic antibiotics	Prevent nosocomial infections	A (studied in ICU patients)	Controversial; evidence favors
Enteral nutrition	Prevent malnutrition	Ungraded	Yes, if GI function satisfactory
Parenteral nutrition, with low carbohydrate and high fat	Prevent malnutrition	Ungraded	Yes, when GI function inadequate
Immune-enhancing nutritional formulations (e.g., omega-3 fatty acids)	Alter immune system and prostenoids	B	Controversial; evidence favors
Early fluid restriction and/or diuresis, possibly with albumin infusion	Prevent or reduce edema	B	Yes, if hemodynamically stable, trial under way
Corticosteroids (high dose) early in course of disease	Reduce inflammation	A	No
Corticosteroids (high dose) late in course of disease	Prevent or reduce intra-alveolar fibroproliferation	B	Controversial; one randomized clinical trial reports benefit when given at 7–14 days; large randomized clinical trial is in progress
Exogenous surfactant	Prevent alveolar collapse	B	No; no trials under way
Acetylcysteine, procysteine	Antioxidant	B	No
Ketoconazole	Inhibit thromboxane and leukotriene synthesis	A	No
Ibuprofen, other NSAIDs	Reduce inflammation	B	No
Alprostadil (prostaglandin E_1)	Inhibit neutrophil activity	B	No
Pentoxifylline, lisofylline	Inhibit cytokine secretion	A	No
Antiendotoxin and anticytokines	Prevent ARDS	A	No
Inhaled nitric oxide	Improve $\dot{V}/\dot{Q}$ balance; reduce hypoxemia	A	No
Nebulized prostacyclin (epoprostenol)	Improve $\dot{V}/\dot{Q}$ balance; reduce hypoxemia	C	No
Inhaled beta-adrenergic agonists	Reduce airway resistance, speed edema resorption	D	Yes

ARDS—acute respiratory distress syndrome GI—gastrointestinal NSAIDs—nonsteroidal anti-inflammatory drugs $\dot{Q}$—perfusion $\dot{V}$—ventilation

neuropsychological sequelae such as impaired memory, attention, and concentration; decreased mental-processing speed; or all four effects.[37] Patients also experience reduced quality of life because of impaired physical functioning that results from muscle wasting and weakness.[19,38,39]

Miscellaneous Causes of Pulmonary Edema

NEUROLOGIC INSULTS

Pulmonary edema can occur as a result of various insults to the central nervous system, including grand mal seizures, head trauma, subarachnoid hemorrhage, intracerebral hemorrhage, and subdural hematoma. The common denominator for these CNS insults is that they are severe and occur acutely. Most often, pulmonary edema is acute, occurring minutes to hours after the CNS event. Occasionally, there may be a more delayed onset and gradual progression over several days. The pathogenesis of neurogenic pulmonary edema is not well understood. The acute form, which is more common, may result in large part from intense sympathetic activity associated with systemic hypertension and central pooling of blood volume. This combination of factors can lead to extraordinary, albeit transient, increases in pulmonary capillary pressure and a predomi-

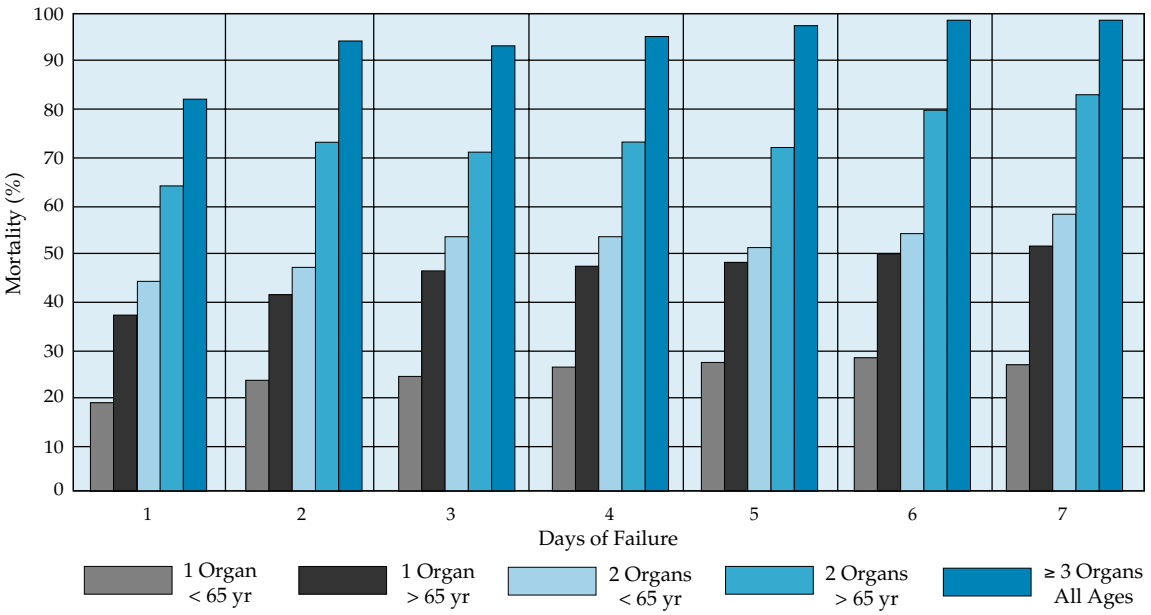


Figure 3 **In patients with ARDS, the number of failing organs and the patient's age have a greater influence on mortality than does the number of days of failure.**

nantly cardiogenic pattern of pulmonary edema. Increased permeability caused by pressure-induced mechanical injury to pulmonary capillaries or possibly by CNS control over pulmonary capillary permeability is also believed to play a role in neurogenic pulmonary edema.[40]

Clinically, neurogenic pulmonary edema is usually diagnosed through its association with a rather dramatic preceding CNS insult. The major differential diagnosis is aspiration injury to the lungs. Unlike neurogenic pulmonary edema, chemical pneumonitis resulting from aspiration frequently persists for more than a few days and is often complicated by secondary bacterial infection. If the pulmonary process clears rapidly (i.e., over a few days), the likely diagnosis is neurogenic pulmonary edema. The management of neurogenic pulmonary edema is basically supportive. Diuretics should not be used in the absence of hypervolemia because of the risk of hypovolemic hypotension, which could aggravate the CNS injury in the setting of increased intracranial pressure.

EXPOSURE TO HIGH ALTITUDE

High-altitude pulmonary edema (HAPE) has been reported to occur at heights of 2,500 to 5,000 m (8,202 to 16,404 ft), with an incidence of 0.5% to 15%; greater risk is associated with young age, male sex, more rapid ascent, heavy exertion, and cold environment.[41] HAPE can also occur in persons residing at high altitude who return from a few days' stay at a lower altitude. Persons with previous episodes of HAPE have a 60% chance of recurrence.

Symptoms are cough, which is sometimes productive of pink or bloody sputum, dyspnea on exertion, and fever; onset of symptoms is often gradual but typically occurs within 48 to 96 hours at high altitude.[41] Fulminant pulmonary edema may be preceded by the less severe symptoms of acute mountain sickness. The edema may be diffuse, yet patchy, or quite asymmetrical.

The primary pathophysiologic abnormality underlying high-altitude pulmonary edema is increased capillary permeability, yet the mechanism of the increased permeability is uncertain. It has been suggested that pronounced hypoxia-induced pulmonary vasoconstriction may lead to overperfusion of the less obstructed portions of the vascular bed and subsequent endothelial injury that results in fluid leakage.[41] Evidence for this mechanism comes in part from the observation that persons who have experienced HAPE have more exaggerated hypoxic pulmonary vasoconstriction than those who have not. Several mechanisms have been examined as contributing factors in this process. Exaggerated release of vasoconstrictors (e.g., endothelin-1) or impaired production of vasodilators (e.g., nitric oxide) may play a role.[41,42] Edematogenic mediators released from endothelial or inflammatory cells may also be involved.[41] In addition, impaired activity of alveolar NA+,K+-ATPase may contribute to slowed clearance of alveolar fluid.[43]

The risk of HAPE can be decreased by ascending slowly and steadily to high altitudes. In addition, nifedipine has been found to prevent HAPE in susceptible persons.[32] Inhaled β-adrenergic agonists may also prevent HAPE.[43] Descent to a lower altitude when symptoms of acute mountain sickness develop should also reduce the risk of pulmonary edema. Once full-blown pulmonary edema has occurred, administration of oxygen, continuous positive airway pressure, nifedipine, and prompt descent are useful treatments.[41] Descending even only a few hundred meters may be beneficial. Inhalation of nitric oxide has been shown to improve arterial oxygenation and may be useful in patients who cannot be evacuated to a lower altitude.[44] Persons who have experienced HAPE are at increased risk for its recurrence and should be advised to avoid high altitudes.

REEXPANSION OF COLLAPSED LUNG

Rapid reexpansion of a collapsed lung may lead to ipsilateral or, occasionally, bilateral pulmonary edema.[45] The risk of reexpansion pulmonary edema after evacuation of a pneumothorax or pleural effusion is related to the amount of air or liquid in the pleural space, the duration of collapse, the rapidity of reexpansion, and the suctional pressures required to reexpand the lung. The development of highly negative pleural pressure during removal of pleural air or liquid, with resultant marked reduction in interstitial hydrostatic pressure, may be important in the

pathogenesis of reexpansion pulmonary edema. The high protein concentration in the edema fluid suggests enhanced membrane permeability. This increased permeability could be caused in part by mechanical stretching and deformation of endothelial pores or by generation of toxic oxygen radicals during reperfusion of the rapidly expanded lung. Depletion of surfactant in the collapsed lung may play a role in the genesis of the reduction in interstitial hydrostatic pressure during reexpansion. The risk of reexpansion pulmonary edema is very low during evacuation of a pneumothorax that has been present for a day or less. For a pneumothorax that is thought to have been present for longer than a day, evacuation under water seal, rather than by application of negative pressure, may reduce the risk of edema. Evacuation of pleural liquid does not usually lead to pulmonary edema unless more than 1.0 to 1.5 L of liquid is removed rapidly. It has been suggested that any amount of pleural liquid can be removed safely as long as pleural pressure is maintained at a level higher than –20 cm H_2O. It is not certain, however, whether this approach will always prevent reexpansion pulmonary edema; therefore, it is advisable to remove very large effusions gradually over several hours whenever possible. Treatment of reexpansion pulmonary edema is supportive. There is no evidence that treatment with diuretics is beneficial.

UPPER AIRWAY OBSTRUCTION

Pulmonary edema has been reported to occur after episodes of upper airway obstruction caused by postextubation laryngospasm, tumors, strangulation, or obstructive sleep apnea.[46] The pathogenesis is thought to be related to the development of highly negative intrapleural pressure (–50 to –100 cm H_2O) caused by vigorous inspiratory efforts against an obstructed airway (Müller maneuver). The highly negative intrapleural pressure decreases the interstitial hydrostatic pressure, increases venous return, and imposes an afterload on the left ventricle. In addition, such pressure may lead to intense sympathetic activation, systemic hypertension, and central pooling of blood volume. These factors together can lead to acute pulmonary edema by increasing the transcapillary pressure gradient (i.e., the difference between the capillary pressure and the interstitial hydrostatic pressure). The condition resolves rapidly after the obstruction is removed.

DRUGS

Acute noncardiogenic pulmonary edema can occur after administration of a number of drugs [*see Table 5*]. Acute pulmonary edema can occur after intravenous injection of heroin or other narcotics.[47] Because the edema fluid has a high protein concentration, it has been suggested that a permeability defect could be a pathogenetic factor, but this finding could result from a transient, extreme increase in capillary pressure produced by a so-called neurogenic mechanism. Onset usually occurs within a few hours after narcotic use, but occasionally, it may be delayed for as long as 24 hours. In addition to the clinical and radiographic features of pulmonary edema, typical signs of narcotic intoxication are present, such as pupillary constriction, decreased respiration, and altered mentation. Fever and leukocytosis do not necessarily indicate the presence of infection. As with neurogenic pulmonary edema, the primary differential diagnostic consideration is aspiration, because of the altered level of consciousness.

Management is supportive and should generally include intubation with mechanical ventilation, both to guarantee adequate ventilatory support and to provide airway protection against aspiration. The role of naloxone is uncertain. Certainly, a patient who has overdosed on narcotics and is experiencing life-threatening hypotension or bradycardia should be given naloxone. Likewise, if naloxone is given to an unresponsive and hypopneic patient who does not necessarily require mechanical ventilation for pulmonary edema, the patient may be spared intubation. In contrast, for a patient who is intubated on an emergency basis because of acute pulmonary edema and who becomes clinically stable without hemodynamic compromise, better management may be to allow the narcotic intoxication to reverse gradually rather than precipitously. There is no evidence that naloxone helps speed resolution of narcotic-induced pulmonary edema. In fact, naloxone has been reported to cause pulmonary edema.[32] Furthermore, acute reversal of narcotic intoxication in a long-term addict could result in agitation, with marked sympathetic activation and a less stable clinical course.

Cocaine causes acute pulmonary edema, usually when used as free-base cocaine.[48] The pathophysiology is uncertain. Like heroin, cocaine leads to a high-protein pulmonary edema that suggests endothelial cell injury and increased capillary permeability. However, as has been suggested with heroin, cocaine could lead to extreme sympathetic activation with a steep, extreme increase in capillary pressure that could produce a transient increase in protein leakage across the capillary membrane. Cocaine also causes coronary vasoconstriction, with acute myocardial ischemia or infarction, resulting in pulmonary edema.

LUNG RESECTION

Pulmonary edema can occur as a complication after lung resection, especially pneumonectomy. The clinical picture is consistent with acute lung injury or ARDS and occurs in approximately 6% of pneumonectomies, 3.7% of lobectomies, and 1% of minor resections. Mortality in these cases is 64.5%.[49]

The pathogenesis of post–lung-resection pulmonary edema is multifactorial. Perioperative fluid overload, impaired lymphatic drainage from lymph node dissection, damage from high concentrations of oxygen, and ischemia and reperfusion injury are probably involved.[50]

Management of post–lung-resection pulmonary edema includes mechanical ventilation, with particular attention paid to airway pressures, and fluid restriction.

Table 5 Drugs Associated with Acute Pulmonary Edema

Narcotics	Intrathecal methotrexate
Cocaine	Bleomycin
Salicylates	Dextran
Nonsteroidal anti-inflammatory drugs	Contrast media
Naloxone	Ethchlorvynol
Amiodarone (after general anesthesia)	Phenothiazines
Thiazide diuretics	Colchicine
Tocolytic agents	Interleukin-2
Protamine	Insulin
Cytarabine	All-*trans*-retinoic acid[54]

The author has received grants for educational activities from and participates in the speakers' bureaus of Glaxo-SmithKline and Merck & Co., Inc.

References

1. Sartori C, Matthay MA: Alveolar epithelial fluid transport in acute lung injury: new insights. Eur Respir J 20:1299, 2002

2. Gluecker T, Capasso P, Schnyder P, et al: Clinical and radiologic features of pulmonary edema. Radiographics 19:1507, 1999

3. Ely EW, Haponik EF: Using the chest radiograph to determine intravascular volume status: the role of vascular pedicle width. Chest 121:942, 2002

4. Duane PG, Colice GL: Impact of noninvasive studies to distinguish volume overload from ARDS in acutely ill patients with pulmonary edema: analysis of the medical literature from 1966 to 1998. Chest 118:1709, 2000

5. Connors AF Jr, Speroff T, Dawson NV, et al: The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. JAMA 276:889, 1996

6. Andrew P: Diastolic heart failure demystified. Chest 124:744, 2003

7. Brochard L: Noninvasive ventilation for acute respiratory failure. JAMA 288:932, 2002

8. Edoute Y, Roguin A, Behar D, et al: Prospective evaluation of pulmonary edema. Crit Care Med 28:330, 2000

9. Adnet F, Le Toumelin P, Leberre A, et al: In-hospital and long-term prognosis of elderly patients requiring endotracheal intubation for life-threatening presentation of cardiogenic pulmonary edema. Crit Care Med 29:891, 2001

10. Ware LB, Matthay MA: The acute respiratory distress syndrome. N Engl J Med 342:1334, 2000

11. Atabai K, Matthay MA: The pulmonary physician in critical care. 5: Acute lung injury and the acute respiratory distress syndrome: definitions and epidemiology. Thorax 57:45, 2002

12. Callister ME, Evans TW: Pulmonary versus extrapulmonary acute respiratory distress syndrome: different diseases or just a useful concept? Curr Opin Crit Care 8:21, 2002

13. Moss M, Burnham EL: Chronic alcohol abuse, acute respiratory distress syndrome, and multiple organ dysfunction. Crit Care Med 31:S207, 2003

14. Iribarren C, Jacobs DR Jr, Sidney S, et al: Cigarette smoking, alcohol consumption, and risk of ARDS: a 15-year cohort study in a managed care setting. Chest 117:163, 2000

15. Mangialardi RJ, Martin GS, Bernard GR, et al: Hypoproteinemia predicts acute respiratory distress syndrome development, weight gain, and death in patients with sepsis. Ibuprofen in Sepsis Study Group. Crit Care Med 28:3137, 2000

16. Marshall RP, Webb S, Bellingan GJ, et al: Angiotensin converting enzyme insertion/deletion polymorphism is associated with susceptibility and outcome in acute respiratory distress syndrome. Am J Respir Crit Care Med 166:646, 2002

17. Bellingan GJ: The pulmonary physician in critical care * 6: the pathogenesis of ALI/ARDS. Thorax 57:540, 2002

18. Shimabukuro DW, Sawa T, Gropper MA: Injury and repair in lung and airways. Crit Care Med 31:S524, 2003

19. Steinberg KP, Hudson LD: Acute lung injury and acute respiratory distress syndrome: the clinical syndrome. Clin Chest Med 21:401, 2000

20. Goodman PC: Radiographic findings in patients with acute respiratory distress syndrome. Clin Chest Med 21:419, 2000

21. Nuckton TJ, Alonso JA, Kallet RH, et al: Pulmonary dead-space fraction as a risk factor for death in the acute respiratory distress syndrome. N Engl J Med 346:1281, 2002

22. Pelosi P, Crotti S, Brazzi L, et al: Computed tomography in adult respiratory distress syndrome: what has it taught us? Eur Respir J 9:1055, 1996

23. Wright PE, Carmichael LC, Bernard GR: Effect of bronchodilators on lung mechanics in the acute respiratory distress syndrome (ARDS). Chest 106:1517, 1994

24. Moloney ED, Evans TW: Pathophysiology and pharmacological treatment of pulmonary hypertension in acute respiratory distress syndrome. Eur Respir J 21:720, 2003

25. Idell S: Anticoagulants for acute respiratory distress syndrome: can they work? Am J Respir Crit Care Med 164:517, 2001

26. Cranshaw J, Griffiths MJ, Evans TW: The pulmonary physician in critical care–part 9: non-ventilatory strategies in ARDS. Thorax 57:823, 2002

27. Bernard GR, Vincent JL, Laterre PF, et al: Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 344:699, 2001

28. Annane D, Sébille V, Charpentier C, et al: Effect of treatment with low doses of hydrocortisone and fludrocortisone on mortality in patients with septic shock. JAMA 288:862, 2002

29. Liberati A, D'Amico R, Pifferi S, et al: Antibiotics for preventing respiratory tract infections in adults receiving intensive care. Cochrane Database Syst Rev (2):CD000022, 2000

30. Beale R, Bryg D, Bihari D: Immunonutrition in the critically ill: a systematic review of clinical outcome. Crit Care Med 27:2799, 1999

31. Martin GS, Mangialardi RJ, Wheeler AP, et al: Albumin and furosemide therapy in hypoproteinemic patients with acute lung injury [comment]. Crit Care Med 30:2175, 2002

32. Thompson BT: Glucocorticoids and acute lung injury. Crit Care Med 31:S253, 2003

33. McIntyre RC Jr, Pulido EJ, Bensard DD, et al: Thirty years of clinical trials in acute respiratory distress syndrome. Crit Care Med 28:3314, 2000

34. Taylor RW, Zimmerman JL, Dellinger RP, et al: Low-dose inhaled nitric oxide in patients with acute lung injury: a randomized controlled trial. JAMA 291:1629, 2004

35. Knaus W, Wagner DP: Multiple systems organ failure: epidemiology and prognosis. Crit Care Clin 5:221, 1989

36. Desai SR, Wells AU, Rubens MB, et al: Acute respiratory distress syndrome: CT abnormalities at long-term follow-up. Radiology 210:29, 1999

37. Hopkins RO, Weaver LK, Pope D, et al: Neuropsychological sequelae and impaired health status in survivors of severe acute respiratory distress syndrome. Am J Respir Crit Care Med 160:50, 1999

38. Herridge MS, Cheung AM, Tansey CM, et al: One-year outcomes in survivors of the acute respiratory distress syndrome. N Engl J Med 348:683, 2003

39. Orme J Jr, Romney JS, Hopkins RO, et al: Pulmonary function and health-related quality of life in survivors of acute respiratory distress syndrome. Am J Respir Crit Care Med 167:690, 2003

40. Smith WS, Matthay MA: Evidence for a hydrostatic mechanism in human neurogenic pulmonary edema. Chest 111:1326, 1997

41. Basnyat B, Murdoch DR: High-altitude illness. Lancet 361:1967, 2003

42. Busch T, Bartsch P, Pappert D, et al: Hypoxia decreases exhaled nitric oxide in mountaineers susceptible to high-altitude pulmonary edema. Am J Respir Crit Care Med 163:368, 2001

43. Sartori C, Allemann Y, Duplain H, et al: Salmeterol for the prevention of high-altitude pulmonary edema. N Engl J Med 346:1631, 2002

44. Anand IS, Prasad BA, Chugh SS, et al: Effects of inhaled nitric oxide and oxygen in high-altitude pulmonary edema. Circulation 98:2441, 1998

45. Tarver RD, Broderick LS, Conces DJ Jr: Reexpansion pulmonary edema. J Thorac Imaging 11:198, 1996

46. Deepika K, Kenaan CA, Barrocas AM, et al: Negative pressure pulmonary edema after acute upper airway obstruction. J Clin Anesth 9:403, 1997

47. Sporer KA, Dorn E: Heroin-related noncardiogenic pulmonary edema: a case series. Chest 120:1628, 2001

48. Albertson TE, Walby WF, Derlet RW: Stimulant-induced pulmonary toxicity. Chest 108:1140, 1995

49. Kutlu CA, Williams EA, Evans TW, et al: Acute lung injury and acute respiratory distress syndrome after pulmonary resection. Ann Thorac Surg 69:376, 2000

50. Jordan S, Mitchell JA, Quinlan GJ, et al: The pathogenesis of lung injury following pulmonary resection. Eur Respir J 15:790, 2000

51. Park WY, Goodman RB, Steinberg KP, et al: Cytokine balance in the lungs of patients with acute respiratory distress syndrome. Am J Respir Crit Care Med 164:1896, 2001

52. Pinhu L, Whitehead T, Evans T, et al: Ventilator-associated lung injury. Lancet 361:332, 2003

53. Lew TW, Kwek TK, Tai D, et al: Acute respiratory distress syndrome in critically ill patients with severe acute respiratory syndrome. JAMA 290:374, 2003

54. Jung JI, Choi JE, Hahn ST, et al: Radiologic features of all-trans-retinoic acid syndrome. AJR Am J Roentgenol 178:475, 2002

Acknowledgments

Figure 2 Seward Hung.

Figure 3 Talar Agasyan.

The previous version of this chapter was authored with Roland H. Ingram, Jr., M.D.

XI PULMONARY HYPERTENSION, COR PULMONALE, AND MISCELLANEOUS VASCULAR CONDITIONS

LEWIS J. RUBIN, M.D.

Pulmonary hypertension is a hemodynamic abnormality that is caused by a variety of disorders. It can be classified into four broad categories: (1) pulmonary arterial hypertension; (2) pulmonary venous hypertension (e.g., left ventricular dysfunction and mitral valve disease); (3) pulmonary hypertension associated with disorders of the respiratory system, usually with hypoxemia (e.g., chronic obstructive and restrictive lung disease); and (4) pulmonary hypertension caused by chronic thrombotic or embolic disease [*see Table 1*].[1] This chapter covers the processes that chronically affect the pulmonary vasculature directly as well as the common features of acute and chronic cor pulmonale, irrespective of cause.

Pulmonary Hypertension

Pulmonary hypertension is defined as a mean pulmonary arterial pressure greater than 25 mm Hg at rest or 30 mm Hg with exercise. Pulmonary hypertension may be suggested when estimated pulmonary arterial systolic pressure exceeds approximately 40 to 50 mm Hg (tricuspid regurgitation velocity of 3.0 to 3.4 m/sec), as assessed by echocardiography. The prevalence of pulmonary hypertension is difficult to measure precisely, but it is very common; most patients with heart failure have some degree of pulmonary hypertension.[2]

PHYSIOLOGY

The normal pulmonary circulation is capable of accommodating the entire cardiac output at perfusion pressures that are one fifth of those in the systemic circulation, even when cardiac output increases severalfold during exercise. The pulmonary circulation accomplishes this by dilation of the vasculature already receiving the cardiac output and recruitment of unused vasculature (i.e., arterioles and capillaries); by these mechanisms, the pulmonary circulation minimizes increases in perfusion pressure and maximizes gas exchange surface area. There are two unique characteristics of the pulmonary vasculature. First, the effect of gravity on blood flow through the lungs is greater than its effect on ventilation, which results in diminishing zones of perfusion from base to apex in the upright position. Upstream pressure, such as the filling pressure from the left side of the heart, also increases pulmonary arterial pressures by distending and recruiting vasculature. Second, there are two portions of the vasculature that are influenced in opposite directions by changes in lung gas volume. Alveolar vessels are lengthened and narrowed monotonically with increases in lung volume; this in turn produces an increase in their resistance to blood flow. In series with the alveolar vessels are the extra-alveolar vessels that are tethered by the lung parenchyma; their size increases with increases in lung gas volume. The result of this interplay of these two characteristics is a rise in pulmonary vascular resistance both when lung volume decreases below usual levels and when it increases above usual levels [*see Figure 1*].

Accordingly, the major responses of the normal pulmonary circulation to upstream and downstream vascular pressures and to changes in lung volume are generally passive responses. Neural and humoral vasomotor responses are normally modest, in keeping with the paucity of vascular smooth muscle.

PATHOGENESIS AND PATHOPHYSIOLOGY

Pulmonary hypertension can be caused by narrowing of the precapillary vessels (arteries and arterioles), loss of vascular surface area, or passive back pressure from the postcapillary vessels [*see Table 2*].

Precapillary pulmonary hypertension can be produced by several mechanisms. Embolic material, such as venous thrombi, can lodge in the pulmonary artery, producing acute obstruction

Table 1 Revised Nomenclature and Classification of Pulmonary Hypertension (2003)[1]

- Pulmonary arterial hypertension (PAH)
  - Sporadic (IPAH)
  - Familial (FPAH)
  - Conditions associated with PAH
    - Collagen vascular disease
    - Congenital systemic to pulmonary shunts (large, small, repaired or nonrepaired)
    - Portal hypertension
    - HIV infection
    - Drugs and toxins
    - Other (glycogen storage disease, Gaucher disease, hereditary hemorrhagic telangiectasia, hemoglobinopathies, myeloproliferative disorders, splenectomy)
    - Significant venous or capillary involvement
      - Pulmonary veno-occlusive disease
      - Pulmonary capillary hemangiomatosis
- Pulmonary venous hypertension
  - Left-sided atrial or ventricular heart disease
  - Left-sided valvular heart disease
- Pulmonary hypertension associated with hypoxemia
  - Chronic obstructive pulmonary disease
  - Interstitial lung disease
  - Sleep-disordered breathing
  - Alveolar hypoventilation disorders
  - Chronic exposure to high altitude
- Pulmonary hypertension due to chronic thrombotic and/or embolic disease
  - Thromboembolic obstruction of proximal pulmonary arteries
  - Thromboembolic obstruction of distal pulmonary arteries
  - Pulmonary embolism (tumor, parasites, foreign material)
- Miscellaneous
  - Sarcoidosis
  - Histiocytosis X
  - Lymphangiomatosis
  - Compression of pulmonary vessels (adenopathy, tumor, fibrosing mediastinitis)

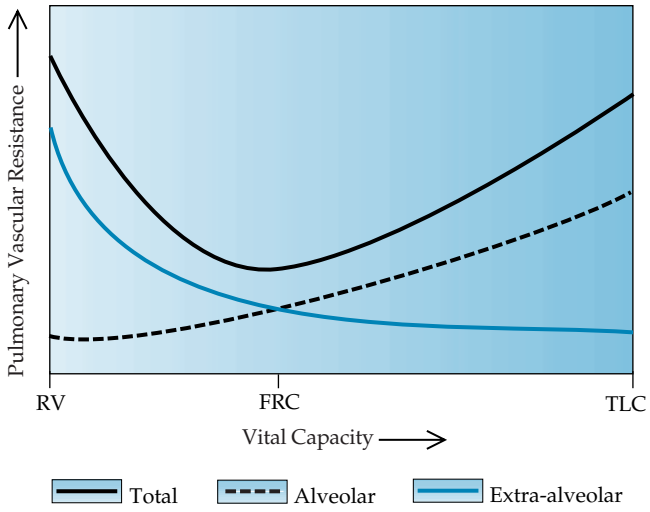


Figure 1 **Effects of lung volume on vascular resistance in alveolar, extra-alveolar, and total pulmonary circulation. The total vascular resistance is lowest at functional residual capacity (FRC) and is higher at low lung volume (residual volume [RV]) and high lung volume (total lung capacity [TLC]).**

or, if unresolved and organized into the vessel wall, chronic obstruction. In situ thrombosis can also occur. Chronically increased blood flow, as seen in large left-to-right intracardiac shunts, is associated with remodeling of the pulmonary arterial walls to vessels that resemble systemic arteries and arterioles; this results in an increase in pulmonary vascular resistance and, ultimately, reversal of the shunt (a condition known as Eisenmenger syndrome). Remodeling of the pulmonary arterial and arteriolar walls as a result of inflammation or endothelial dysfunction can also occur, for example in idiopathic pulmonary arterial hypertension (IPAH) or pulmonary arterial hypertension (PAH) caused by connective tissue diseases.

Loss of the pulmonary vascular bed as a result of destructive processes such as emphysema or interstitial fibrotic disease will increase resistance to blood flow and produce pulmonary hypertension. Hypoxia-induced pulmonary vasoconstriction and vascular remodeling further augment the degree of pulmonary hypertension in this setting.

On rare occasions, the pulmonary veins can be obstructed by a primary process (pulmonary veno-occlusive disease) or during passage of the pulmonary veins through the mediastinum (neoplasm or mediastinal fibrosis). Any process that increases left atrial pressure (mitral stenosis or regurgitation) or increases left ventricular end-diastolic pressure (LVEDP) will also increase pulmonary arterial pressure, with less dramatic increases in intrinsic pulmonary arteriolar resistance.

Regardless of the etiology, when pulmonary hypertension occurs, the vasculature responds by undergoing changes that further increase its resistance [*see Figure 2*].[3] The patterns of histopathologic change seen in pulmonary hypertension are medial hypertrophy, intimal thickening, plexogenic pulmonary arteriopathy, thrombotic pulmonary arteriopathy, and veno-occlusive disease.[4] Historically, these patterns were felt to be specific for the different causes of pulmonary hypertension. However, more recent studies indicate that these changes likely represent a final common pathway of response to pulmonary vascular injury and persistent pulmonary hypertension.

DIAGNOSIS

The diagnostic evaluation of pulmonary hypertension [*see Figure 3*] begins with a careful history and physical examination.

Many of the clinical features of pulmonary hypertension are similar regardless of the underlying cause.[5] Early in the process, the symptoms of pulmonary hypertension may be minimal and nonspecific. Dyspnea, weakness, and fatigue are common; these symptoms are sometimes associated with chest pain that can mimic angina pectoris. Syncope, which is often exertional, occurs late in pulmonary hypertension; it is a sign of poor prognosis because it implies an inability to augment cardiac output during exertion. Hoarseness caused by compression of the recurrent laryngeal nerve (Ortner syndrome) and hemoptysis related to rupture of hypertensive, atherosclerotic small pulmonary vessels could occur. Signs and symptoms of right ventricular failure, such as edema and ascites, occur relatively late in the disease; as with syncope, these symptoms indicate a poor prognosis.

Physical Examination

During physical examination, the jugular veins may be distended, and there may be prominent A waves, signifying decreased right ventricular (RV) compliance. Also, increased V waves may indicate tricuspid regurgitation. Palpation of the chest may detect an RV heave in the parasternal area or, in patients with COPD, in the subxiphoid area. On auscultation of the heart, there may be an increased P_2, an RV S_4, an RV S_3, a pulmonic ejection click, the murmur of tricuspid regurgitation at the lower right sternal border that is increased with inspiration, and, occasionally, pulmonic regurgitation (Graham Steell murmur). Hepatomegaly, ascites, and lower extremity edema are each extrathoracic indicators of right ventricular failure.

Table 2 Relation between Site, Pathogenesis, and Disorders of the Pulmonary Circulation

Site	*Pathogenesis*	*Disorders*
Precapillary	Intravascular obstruction, increased blood flow, vascular remodeling or inflammation, vasospasm, increased blood viscosity	Pulmonary emboli or in situ thrombosis, left-to-right shunt, alveolar hypoxia, vasculopathy caused by collagen vascular disorder or primary pulmonary hypertension, polycythemia
Capillary	Destruction of capillary bed	Emphysema, interstitial lung disease, surgical removal
Postcapillary	Passive back pressure from pulmonary venous obstruction, high left atrial pressure, high LVEDP	Pulmonary veno-occlusive disease, pulmonary venous obstruction in the mediastinum, mitral regurgitation or stenosis, left ventricular failure

LVEDP—left ventricular end-diastolic pressure

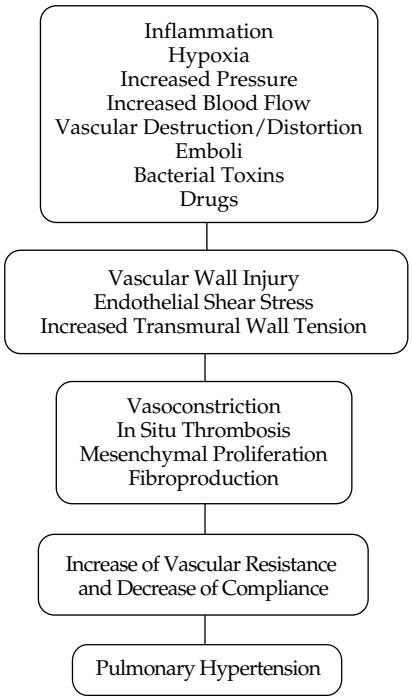


Figure 2 **Mechanisms of pulmonary hypertension. Irrespective of the initiating event or events, pulmonary hypertension worsens over time because of vascular wall injury, endothelial shear stress, and an increase in transmural wall tension. These processes result in vasoconstriction, vascular wall remodeling, and in situ thrombosis.**

Initial Imaging and Physiologic Testing

Chest radiography The chest radiograph can provide clues to the diagnosis and etiology of pulmonary hypertension. Symmetrical enlargement of the pulmonary arteries, with rapid tapering of the distal vessels (pruning) and enlargement of the right ventricle, can be found but usually is not seen until later stages of the disorder. Asymmetrical enlargement of the central pulmonary arteries may be seen in patients with chronic thromboembolic pulmonary hypertension (CTEPH). Radiographic findings that suggest the underlying cause of pulmonary hypertension include pulmonary venous congestion (e.g., LV failure), hyperinflation (COPD), and interstitial lung disease (e.g., interstitial pulmonary fibrosis).

Electrocardiography In mild cases of pulmonary hypertension, the electrocardiogram may be normal. In more severe cases, the ECG will show changes of right ventricular hypertrophy and right atrial enlargement. These changes are to be contrasted with the typical ECG findings seen in COPD, which largely reflect the hyperinflation of the lungs and low diaphragms.

Transthoracic echocardiography Transthoracic echocardiography with Doppler estimation of pulmonary arterial pressure is the noninvasive method of choice in screening populations of patients with a high incidence of pulmonary hypertension; such patients include those with the scleroderma spectrum of disease, those who have a family history of pulmonary artery hypertension, patients with HIV, and patients with cirrhosis who are being evaluated for liver transplantation. Transthoracic echocardiography is also useful in evaluating patients with symptoms suggestive of pulmonary hypertension. Findings indicative of pulmonary hypertension include hypertrophy, enlargement of the right ventricle and atrium, and abnormal motion of the septum [*see Figure 4*]. In some patients, transesophageal echocardiography may be necessary to detect congenital defects. If hypoxemia suggestive of a right-to-left intracardiac shunt is present, injection of saline filled with air bubbles during the echocardiogram will allow detection and localization of the shunt.

Tests to Diagnose Underlying Cause of Pulmonary Hypertension

It is often necessary to perform additional testing to determine the etiology of pulmonary hypertension. The sequence for this testing generally begins with noninvasive studies, followed by more invasive studies as needed.

Pulmonary function testing Measurements of pulmonary function can be useful in evaluating patients with pulmonary hypertension. Detection of severe obstructive or restrictive defects may indicate that all or a portion of the pulmonary hypertension is caused by intrinsic lung disease. By contrast, isolated reduction of the diffusion capacity or minimal reduction in the lung volumes can be seen in any of the causes of pulmonary vasculopathy. Arterial blood gases at rest and pulse oximetry with exercise detect complicating resting or exercise hypoxemia that should be remedied therapeutically with supplemental oxygen. The finding of hypercapnia is most compatible with severe chronic airflow obstruction, sleep apnea, or restrictive chest wall disease.

Polysomnography Patients suspected of having a sleep disorder of breathing that may be causing or contributing to pulmonary hypertension should undergo nocturnal polysomnography.

Ventilation-perfusion lung scanning Ventilation-perfusion lung scanning is a critical test in evaluating patients with pulmonary hypertension, especially when chronic thromboembolic pulmonary hypertension is suspected. Patients with chronic thromboembolic pulmonary hypertension will have multiple, bilateral perfusion defects of different sizes (usually interpreted as indicating a high probability of pulmonary embolism), whereas patients with other causes of pulmonary hypertension will have either homogeneous or mildly mottled perfusion [*see Figure 5*]. The presence of radioactivity in the head or kidney suggests a right-to-left intracardiac or intrapulmonary shunt.

Computed tomography Computed tomography of the chest using spiral or helical or electron-beam techniques can visualize central pulmonary thromboemboli. High-resolution CT of the chest can detect emphysema or interstitial lung disease not seen on routine chest radiography and may also provide clues to the presence of pulmonary veno-occlusive disease.

Cardiac catheterization and pulmonary arteriography Right heart catheterization, with pulmonary arteriography and/or left heart catheterization as clinically indicated, should be performed to confirm the presence of pulmonary hypertension and determine its severity and to exclude congenital heart disease, proximal or peripheral pulmonary arterial stenosis, and valvular or ventricular left sided heart disease. Because of the increased risk of complications, experienced angiographers should perform pulmonary arteriography in this situation. In those patients who are thought to have PAH, vasodilator testing with such agents as ni-

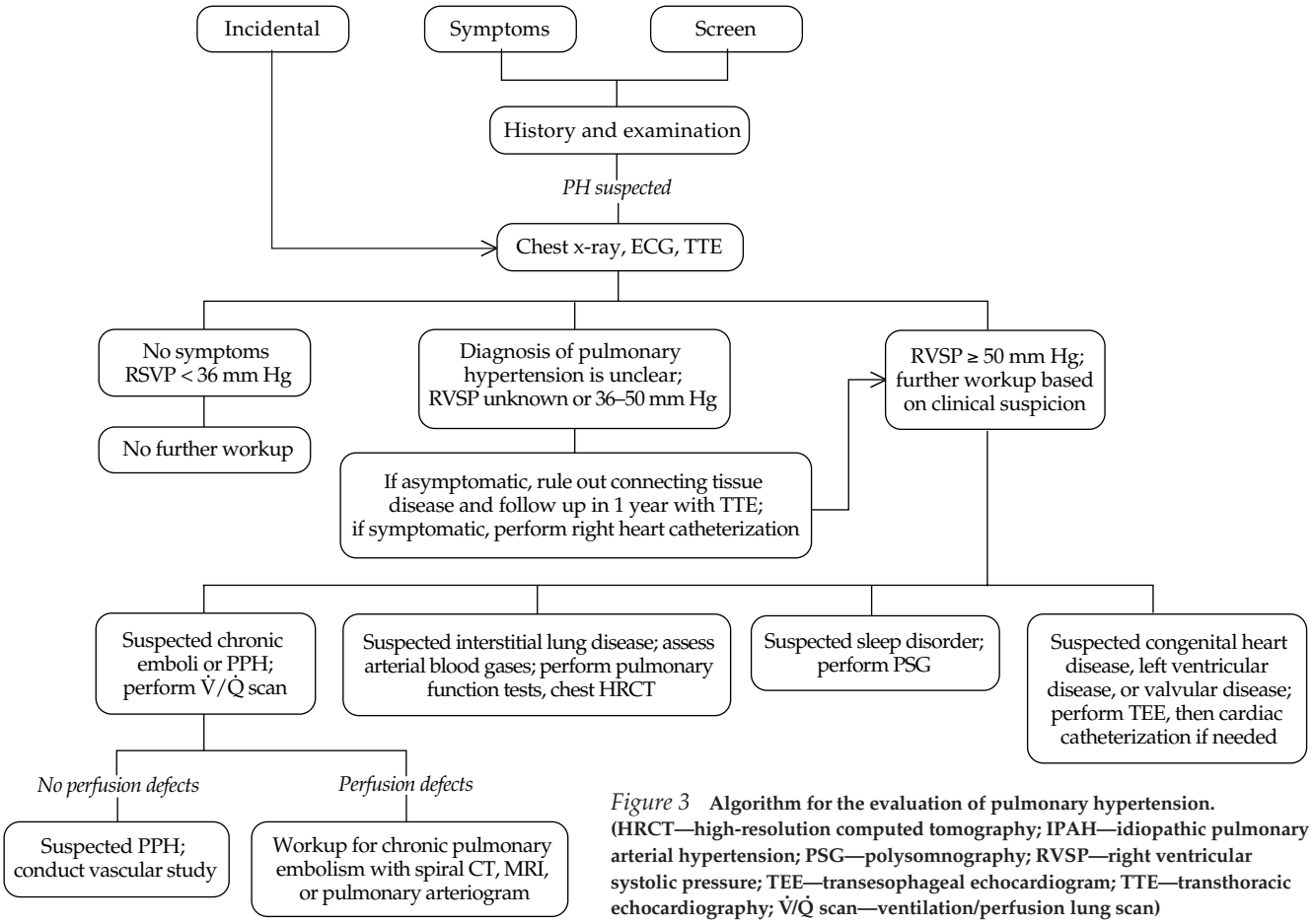


Figure 3 **Algorithm for the evaluation of pulmonary hypertension. (HRCT—high-resolution computed tomography; IPAH—idiopathic pulmonary arterial hypertension; PSG—polysomnography; RVSP—right ventricular systolic pressure; TEE—transesophageal echocardiogram; TTE—transthoracic echocardiography; V̇/Q̇ scan—ventilation/perfusion lung scan)**

tric oxide, prostacyclin, and adenosine, with monitoring of gas exchange and pulmonary hemodynamics, should be performed.[5]

Lung biopsy It is very uncommon that a lung biopsy is required to establish the cause of pulmonary hypertension. The only exceptions would be in patients in whom one of the interstitial lung diseases or pulmonary arteritis is suspected as a cause of pulmonary hypertension. Bronchoscopic lung biopsy is contraindicated in patients with severe pulmonary hypertension; in such cases, open or video-assisted biopsy is the technique of choice. Such a biopsy poses a greater risk for patients with pulmonary hypertension than for patients without pulmonary hypertension.

TREATMENT

Treatment of pulmonary hypertension depends on the underlying cause. Disorders that affect the pulmonary circulation acutely (e.g., pulmonary embolism, pulmonary edema, and the acute respiratory distress syndrome) are covered elsewhere [*see 1:XVIII Venous Thromboembolism and 14:X Pulmonary Edema*], as are valvular heart defects [*see 1:XI Valvular Heart Disease*], congenital heart defects that can cause pulmonary hypertension [*see 1:XV Adult Congenital Heart Disease*], and diseases that affect the lung airways and parenchyma, such as COPD and the interstitial lung diseases [*see 14:III Chronic Obstructive Diseases of the Lung and 14:V Chronic Diffuse Infiltrative Lung Disease*].

Cor Pulmonale

DEFINITION

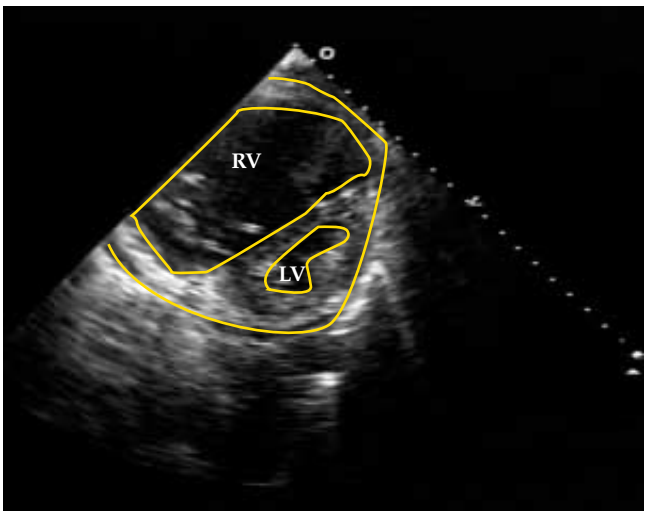
Cor pulmonale is the term used for pulmonary hypertension resulting from disorders of the pulmonary parenchyma, the thoracic cage, or the neuromuscular system, excluding congenital heart disease and disorders of the left side of the heart. Cor pulmonale can occur acutely in settings of rapid-onset right ventricular overload, or chronically with the slow onset of pulmonary hypertension.

ACUTE COR PULMONALE

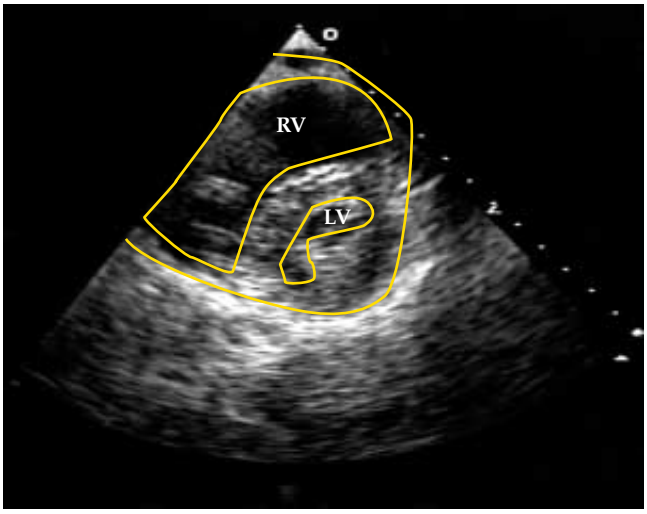
Pathogenesis

The right ventricle normally pumps against a low afterload, even when cardiac output is dramatically increased by exercise. In response to an acute increase in pulmonary vascular resistance, the right ventricle distends, producing an increase in right ventricular systolic and diastolic volume, but is unable to generate high pressures (the maximal pressure generated is usually approximately 40 mm Hg). If the right ventricle cannot adequately compensate, increases in right ventricular end-diastolic pressure (RVEDP) and right atrial pressure occur, producing acute right heart failure. Additionally, the reduced cardiac output from the right ventricle to the left ventricle and the shift of the interventricular septum toward the left ventricle cause a re-

a



b



c

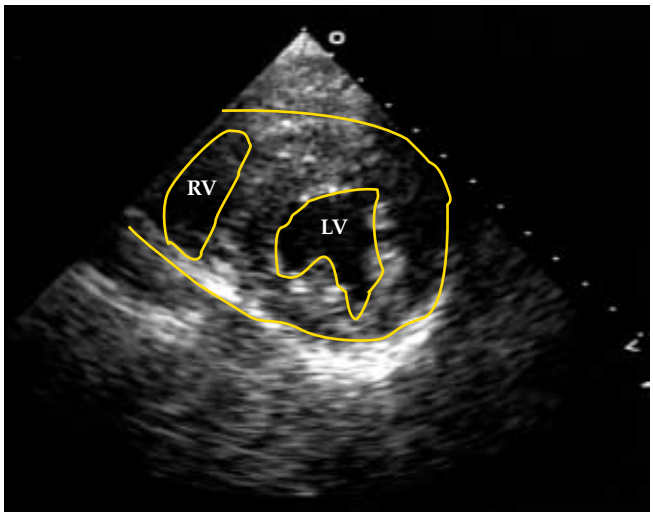


Figure 4 **(*a*) Echocardiogram of a patient with chronic pulmonary hypertension. (*b*) Echocardiogram of a patient with acute pulmonary hypertension. (*c*) Echocardiogram of the patient shown in Figure B after clot lysis. (LV—left ventricle; RV—right ventricle)**

duction in left ventricular filling, resulting in systemic hypoperfusion. Decreased coronary perfusion caused by low systemic diastolic pressure further reduces the ability of the right ventricle to overcome the added resistance, producing a rapid decline to death.

The disorders that cause acute cor pulmonale are diseases that produce sudden obstruction of the pulmonary vasculature, such as massive pulmonary thromboembolism[6]; acute embolism caused by other materials, such as air, bone marrow, fat, amniotic fluid, or tumor; or obstruction of the microvasculature caused by high airway pressure or destruction, as is seen in acute respiratory distress syndrome.[7]

Diagnosis

In the setting of acute respiratory failure or shock with evidence of acute right heart failure, immediate bedside echocardiography can demonstrate acute cor pulmonale. Features that are seen include pulmonary hypertension (usually mild), right ventricular dilatation without hypertrophy, tricuspid regurgitation, septal flattening or paradoxical septal motion, and left ventricular diastolic dysfunction [*see Figure 5*].[8]

Treatment

Rapid recognition of acute cor pulmonale and relief of the pulmonary vascular obstruction, if possible, are key to survival for patients with acute cor pulmonale. For example, thrombolytic therapy for patients with acute massive pulmonary embolism may result in complete resolution of the acute cor pulmonale [*see Figure 6*].[9]

CHRONIC COR PULMONALE

Pathogenesis

In response to a chronic increase in pulmonary vascular resistance, the right ventricle will distend and undergo hypertrophy. When the ability of the right ventricle to compensate is overwhelmed, increases in RVEDP and right atrial pressure occur, resulting in right heart failure.

The pulmonary hypertension that characterizes chronic cor pulmonale is produced by increased pulmonary vascular resistance caused by varying combinations of pulmonary vascular destruction or obstruction, dynamic vasoconstriction caused by hypoxia or acidosis, and remodeling of the pulmonary vasculature. Some of these changes (e.g., hypoxic vasoconstriction and some remodeling) are reversible with therapy, whereas others (vascular bed destruction) are not.

Diagnosis

Early detection of chronic cor pulmonale may be difficult because the manifestations of the underlying lung disease dominate the clinical picture. The clinical features and evaluation are the same as those described for pulmonary hypertension [*see* Pulmonary Hypertension, *above*] Echocardiography has become the most useful means of detecting right heart changes caused by pulmonary hypertension.

Treatment

The nature and severity of the underlying lung disease determine the outcome of patients with chronic cor pulmonale. For example, cor pulmonale secondary to sleep apnea may be entirely reversible with treatment directed at the underlying ventilatory disorder, whereas cor pulmonale caused by idiopathic pul-

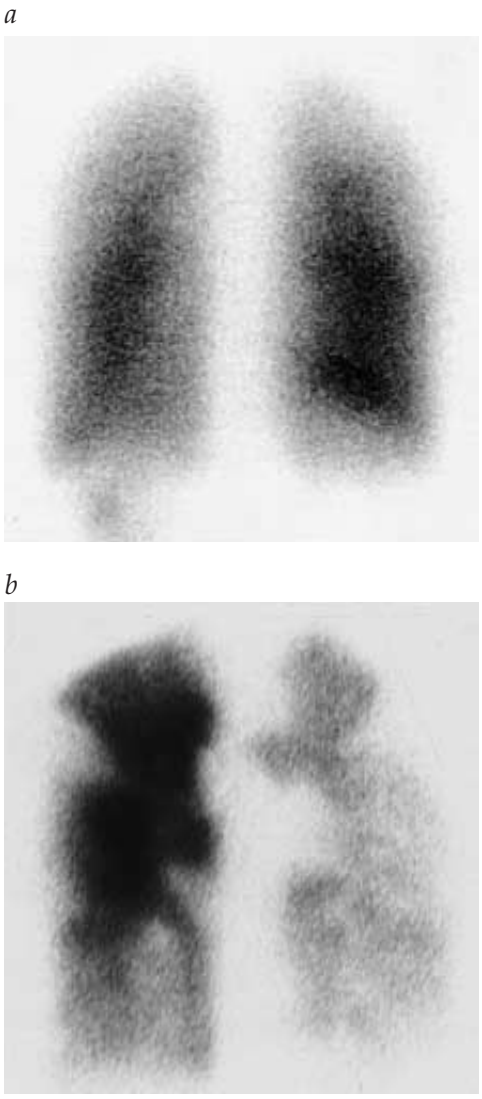


Figure 5 **Perfusion lung scans of a patient with primary pulmonary hypertension showing homogeneous perfusion (*a*) and of a patient with chronic thromboembolic pulmonary hypertension, showing large bilateral perfusion defects (*b*)**

monary fibrosis is often irreversible. In patients with COPD, an increased incidence of right ventricular involvement may correlate with increasing severity of lung dysfunction.

The most important treatment of chronic cor pulmonale is the treatment of the underlying lung disease [*see 14:III Chronic Obstructive Diseases of the Lung, 14:V Chronic Diffuse Infiltrative Lung Disease, 14:VI Ventilatory Control during Wakefulness and Sleep, and 14:VII Disorders of the Chest Wall*]. In patients with hypoxemia, controlled-flow supplemental O_2 should be given at rest, during exercise, and during sleep to maintain O_2 saturation above 90%. Patients should be monitored for acute increases in arterial carbon dioxide pressure (P_aCO_2).

Diuretics are useful in reducing the edema, ascites, and liver congestion associated with cor pulmonale, but these agents must be used carefully to avoid reducing right ventricular filling pressures, which may lead to decreasing cardiac output. Phlebotomy in patients with secondary polycythemia[10] and noninvasive mechanical ventilation of patients with chronic respiratory failure[11] may also improve hemodynamics.

Although vasodilator agents such as calcium channel blockers block hypoxic pulmonary vasoconstriction and cause pulmonary vasodilatation, the impairment in gas exchange in patients with chronic cor pulmonale negates any potential positive effect of these drugs. Epoprostenol (prostacyclin) and three of its analogues have been studied in several types of chronic pulmonary hypertension and are approved for use in PAH, but they have not been studied extensively in cor pulmonale.[12] Except in cases of coexistent left ventricular failure, clinical studies do not support the use of digitalis in patients with cor pulmonale; in addition, use of digitalis may be associated with greater toxicity.

Intrinsic Pulmonary Vascular Diseases

IDIOPATHIC PULMONARY ARTERIAL HYPERTENSION

Idiopathic pulmonary arterial hypertension (IPAH), formerly known as primary pulmonary hypertension (PPH), is a condition in which the pulmonary vasculature is the exclusive site of the pulmonary hypertension.[1,13] By definition, other causes of pulmonary hypertension are excluded.

Epidemiology

IPAH is rare, having an estimated incidence of 1 to 2 per million people a year. Although the disease can occur at any age, the mean age at diagnosis is 36 years.[13] The disease is more frequent in females than in males (ratio, 1.7:1 to 3.5:1) and is equally represented in all races. Approximately 6% to 12% of cases are familial, now termed familial pulmonary artery hypertension (FPAH).[1,14]

In the United States, the mortality from IPAH rose substantially from 1979 to 1996, possibly because of the introduction of anorexigens (see below).[15]

Etiology and Pathogenesis

The cause of IPAH is unknown, although there are several clinical conditions associated with its development. For example, the ingestion of certain drugs or other materials has been associated with IPAH.[16] In the 1960s and again in the 1990s, ingestion

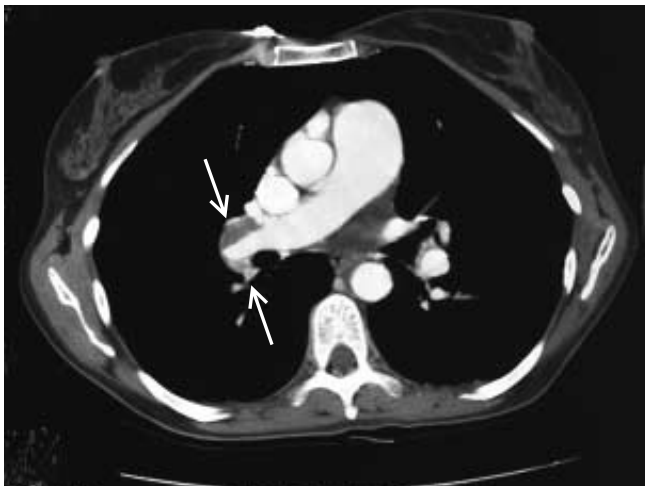


Figure 6 **Spiral CT scan of the chest of a patient with chronic thromboembolic pulmonary hypertension showing large defects in the right main pulmonary artery (arrows).**

of appetite suppressants such as aminorex, fenfluramine, and dexfenfluramine were associated with an increased incidence of pulmonary hypertension.[17] Additionally, ingestion of contaminated rapeseed oil, L-tryptophan, amphetamines, and cocaine are associated with IPAH.[16] Other conditions associated with IPAH are splenectomy/asplenia, portal hypertension, and HIV infection.[18,19]

As is true of many diseases, it is thought that PAH in general, and IPAH in particular, results from an inciting factor in a genetically susceptible individual. Studies of inheritance have suggested an autosomal dominant pattern with markedly reduced penetrance in FPAH. The gene for familial pulmonary hypertension, *PPH1,* has now been identified as heterogeneous germline mutations of the gene coding for the bone morphogenic protein receptor 2 (BMPR2).[20] The exact mechanisms accounting for how abnormalities of this receptor produce pulmonary hypertension are being investigated.[21]

Studies of the pulmonary vasculature in IPAH suggest endothelial injury and dysfunction occur early in the process.[22] In IPAH, expression of endothelial nitric oxide synthetase is reduced (nitric oxide is a pulmonary vasodilator) and expression of endothelin 1, a potent pulmonary vasoconstrictor and a mitogen, is increased. There is an excess of thromboxane (a vasoconstrictor and potent stimulus for platelet aggregation) relative to prostacyclin, a pulmonary vasodilator. Other vasoactive mediators, such as serotonin, another pulmonary vasoconstrictor, may also play a role, especially in pulmonary hypertension associated with appetite suppressants, which inhibit serotonin reuptake and increase expression of its transporter. Abnormal α_1-adrenoceptor affinity and responsiveness may produce downstream signaling events that cause defects in ion channel activity and control of intracellular calcium and could contribute to vasoconstriction and to smooth muscle proliferation and growth.[23] As a consequence of the endothelial dysfunction and resultant pulmonary vasoconstriction, intimal proliferation, smooth muscle hyperplasia and hypertrophy, and other remodeling phenomena occur, further increasing pulmonary vascular resistance.[19] In situ thrombosis may also play a role in endothelial injury and vascular obstruction.

Diagnosis

The diagnosis of IPAH can be made when clinical findings (e.g., raised pulmonary artery pressure, dyspnea, evidence of right heart failure) [*see* Pulmonary Hypertension, *above, and Figure 3*] are present and other causes of pulmonary hypertension have been excluded.[13]

Treatment

Medical therapy Current treatment options slow the progression of the disease but do not halt it. Patients with IPAH should receive long-term warfarin anticoagulation therapy to achieve an INR (international normalized ratio) of 1.5 to 2.5, and they should avoid medications that interfere with warfarin metabolism.[13] Medications that worsen pulmonary hypertension or right heart function (e.g., decongestants, beta blockers) should also be avoided. Symptom-limited physical activity should be encouraged. Pregnancy is poorly tolerated and should be prevented (oral contraceptives should not be used, because they may increase the risk of thrombosis). Invasive medical procedures should be avoided whenever possible. Oxygen should be used for resting or exercise-induced hypoxemia or if the patient will be exposed to high altitudes, such as on an airplane flight. Diuretics are useful to help control edema and ascites. Use of digoxin is controversial.[24] Vasodilators may be helpful for only a select portion of patients with pulmonary hypertension who have evidence of significant vasoreactivity, generally considered to be 10% of those with IPAH and very few with other forms of PAH. For patients who have an acute vasodilator response to the agents used (i.e., inhaled nitric oxide, inhaled iloprost, intravenous epoprostenol, or intravenous adenosine), a trial of a calcium channel blocker, such as nifedipine or diltiazem, should be initiated. The patient should be monitored closely during this time because some patients will deteriorate despite this therapy. If the patient responds, the therapy can be continued and titrated to maximum benefit. Patients should be warned that abrupt withdrawal of therapy can lead to rebound pulmonary hypertension that can be fatal.[24]

Patients who do not show a vasodilator response during right heart catheterization should be started on one of the other available agents, including the oral endothelin receptor antagonists (e.g., bosentan), subcutaneous or intravenous treprostinil, inhaled iloprost, or intravenous epoprostenol.[13,25-27] Epoprostenol must be given intravenously and, because of its 3- to 5-minute half-life, delivered by continuous infusion. Abrupt discontinuance of therapy may be required because of drug-delivery system malfunction or I.V. access infections. Minor side effects include headache, jaw pain, rash, diarrhea, and joint pain. The dose must be gradually increased over time to maintain maximum benefit. Epoprostenol appears to have effects that go beyond vasodilation and may include the decreased production of endogenous vasoconstrictor substances and antiplatelet and antiproliferative properties that seem to ameliorate what previously appeared to be irreversible vascular changes.[28] Bosentan is administered orally twice daily. The major toxicity is hepatic, which necessitates monthly monitoring of liver function. Treprostinil administered subcutaneously produces site pain that often limits its usefulness. Inhaled iloprost obviates the need for intravenous delivery of prostanoid medication; however, it must be administered by nebulization six to nine times daily.

Surgical therapy Creation of a small interatrial communication by percutaneous balloon atrial septostomy may result in decompression of the right ventricle and improvement in symptoms related to right heart failure.[29] Indications include recurrent syncope and right heart failure despite maximum medical therapy, deterioration despite maximum medical therapy while the patient is awaiting transplantation, and exhaustion of all other options. Because right-to-left shunting at the atrial level occurs, hypoxemia worsens but is usually well tolerated in patients without severe preprocedure hypoxemia.

Lung transplantation Lung transplantation is indicated for patients in whom IPAH has progressed despite optimal medical and surgical therapy. In most centers, heart-lung transplantation is no longer performed because it was found that the right ventricle recovered both form and function when pressure fell after even single-lung transplantation. Indications for referral for transplant evaluation include New York Heart Association (NYHA) functional class III or IV despite medical therapy[29]; failure of epoprostenol therapy and the occurrence of severe side effects from epoprostenol are additional indications for referral for transplant evaluation. These guidelines take into consideration the course of the disease and the waiting time for transplantation.

The surgical mortality for patients with IPAH is higher than that for patients receiving lung transplantation for other forms of pulmonary disease,[30] partly because of the greater complexity of the transplant surgery. The 1-year survival for patients receiving lung transplantation for IPAH is 65%; the 3-year survival is 55%; and the 5-year survival is 44%. No randomized study of medical therapy versus transplantation has been performed. Comparison of survival data from different studies suggests that survival may be higher with medical therapy than with transplantation, although it is likely that only the more severely ill patients received lung transplantation.[31] Recurrences of IPAH after lung transplantation has not been reported.

Prognosis

A National Institutes of Health (NIH) registry of patients with IPAH in the 1980s defined the natural history of the disease.[32] The mean life expectancy from diagnosis was 2.8 years, and the 5-year survival was 22% to 38%. Patients younger than 14 years and older than 65 years had lower survival, as did patients with more severe symptoms. Patients with acute responses to vasodilators had a better prognosis. This study also developed a formula, utilizing data from right heart catheterization (right atrial pressure, cardiac index, and mean pulmonary arterial pressure), that predicted survival of patients before the current era of efficacious medical therapy and lung transplantation. With these new therapies, the natural history of the disease has changed. A recent evaluation study identified African-American and Asian descent as factors associated with an increased risk of death in patients with IPAH; cardiac function and acute reactivity of the pulmonary vascular bed remained strong independent predictors of outcome.[33]

CHRONIC THROMBOEMBOLIC PULMONARY HYPERTENSION

Epidemiology

The epidemiology of chronic thromboembolic pulmonary hypertension (CTEPH) is unknown. An estimated 0.1% to 4% of patients with acute pulmonary embolism will experience chronic pulmonary hypertension, suggesting that several thousand cases occur each year in the United States.[34] There are conflicting data on the male-to-female ratio. The disease has been seen in adults of all ages, but more than 50% of patients are younger than 45 years. The reason for the development and lack of resolution of pulmonary emboli in these patients is unknown. A minority of the patients is found to have hypercoagulability states, such as deficiencies of protein C or S or of antithrombin III. About 10% will have circulating lupus anticoagulant.[35] The incidence of factor V Leiden and other hypercoagulable syndromes has not been adequately studied. Abnormalities of the fibrinolytic system have been sought, but no consistent patterns have been detected.[35] Embolization of so-called aged clot, lack of therapy for the initial episode of embolism, and recurrent emboli have all been hypothesized. Medical conditions that may be associated with an increased risk for the development of CTEPH are splenectomy, ventriculo-atrial shunt for the treatment of hydrocephalus, and chronic inflammatory disorders such as osteomyelitis and inflammatory bowel disease.[36]

Pathogenesis

CTEPH results from the organization (rather than lysis) of the clots from a single massive episode of pulmonary embolism or multiple episodes of pulmonary embolism. As a consequence of obstruction and distortion of the proximal pulmonary vasculature, the pulmonary vascular resistance is increased, and over time, pulmonary hypertension develops and right ventricular function worsens.[35]

Diagnosis

Clinical manifestations Patients with CTEPH usually present months or years after the initial (and often asymptomatic) embolic event(s) with symptoms of pulmonary hypertension.[35] The initial event may have been diagnosed as an episode of embolism (in only 50% of cases) or may have been undiagnosed or misdiagnosed as pneumonia or another clinical entity. The reason for the delay in onset of symptoms is unknown, although it is suspected that organization of the clot and the increased pulmonary arterial pressure result in progressive remodeling of the obstructed and unobstructed pulmonary vasculature, gradually worsening the pulmonary hypertension and finally resulting in right ventricular failure.

Physical examination A pulmonary arterial flow murmur is the only finding on physical examination that is characteristic of chronic thromboembolic pulmonary hypertension. It is best heard over the lung fields while the patient holds his or her breath.

Imaging and physiologic testing The chest radiograph may be normal or may show abnormalities suggestive of pulmonary hypertension. Findings that suggest chronic thromboembolic pulmonary hypertension include asymmetrical enlargement of the pulmonary arteries, regions of hyperperfusion and hypoperfusion, and focal fibrotic areas of old infarction that may be associated with local pleural thickening or cavitation.

The electrocardiogram may be normal or have changes of right ventricular hypertrophy or strain, right atrial enlargement, and right bundle branch block.

Pulmonary function tests will most commonly show no abnormality of lung volumes or of spirometry measurements, although about 20% of patients will have restriction, probably related to previous infarction.[35] Frequently, carbon monoxide diffusion in the lung is mildly reduced; this reduction is thought not to be proportional to the degree of obstruction. Hypoxemia is common and is often worsened by exercise. The hypoxemia is the result of a combination of ventilation-perfusion inequality, low cardiac output, and, sometimes, a patent foramen ovale with a right-to-left shunt.[35]

The ventilation-perfusion lung scan is the best method of screening patients with pulmonary hypertension to identify those in whom the disorder may be caused by chronic thromboembolism.[35,37] Patients with chronic thromboembolism will have at least one segmental or larger perfusion defect [*see Figure 5*]. This pattern can also be seen with tumor or fibrosing mediastinitis. In chronic thromboembolism, the degree of obstruction is underestimated by the defects seen on the perfusion scan.[35]

When large perfusion defects are seen on the ventilation-perfusion scan, additional testing is needed to confirm that the defects are caused by chronic thromboembolism. Spiral CT is a noninvasive imaging technique that can allow identification of proximal chronic clots and detect other causes of pulmonary vascular obstruction [*see Figure 6*].[35] However, pulmonary angiography is needed to accurately assess the full extent of pulmonary vascular obstruction and determine candidacy for pulmonary thromboendarterectomy.

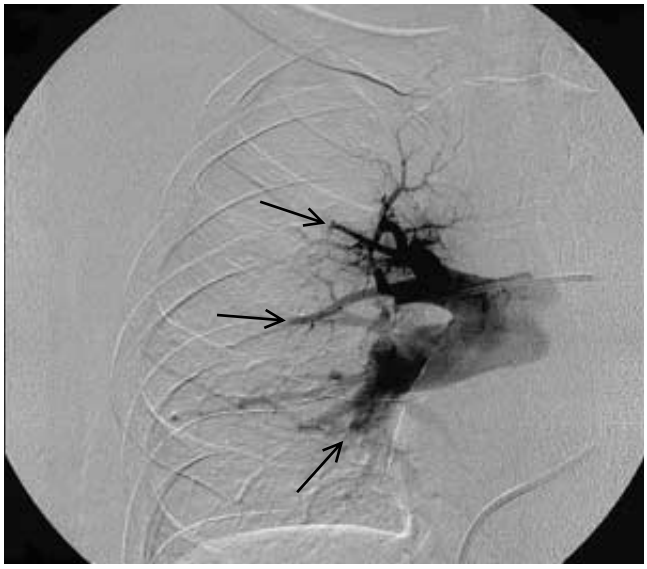


Figure 7 **Pulmonary arteriogram of a patient with chronic thromboembolic pulmonary hypertension showing multiple abrupt terminations (arrows) of the pulmonary arterial branches, with no distal flow characteristic of chronic pulmonary embolism.**

The use of pulmonary angiography in patients with pulmonary hypertension has been thought to be associated with a significant risk of death or adverse outcome. Use of oxygen, small amounts of nonionic contrast dye, and a limited number of injections has reduced the risk dramatically.[35] Results of pulmonary angiography must be interpreted with the understanding that the patterns of pulmonary hypertension are different from those of acute embolism [*see Figure 7*].[35] In some cases, direct visualization of the pulmonary vasculature by pulmonary angioscopy is useful to confirm the presence and extent of obstruction.[35]

Treatment

There is no effective medical therapy for CTEPH. Vasodilators, angioplasty of the obstructed areas, and thrombolytic therapy have been investigated in small studies and have limited utility. Preoperative anticoagulation is needed to prevent further embolism and in situ thrombosis. Placement of an inferior vena cava filter is recommended.[35]

Surgery for CTEPH is pulmonary endarterectomy, a procedure that is associated with significant risk and therefore should not be considered unless the patient meets criteria for surgery. The criteria used in the centers with expertise are: an elevated pulmonary vascular resistance (> 300 dynes·sec·cm^{-5}), severe disability (at least NYHA class III), and surgically accessible disease.[35] Contraindications or factors that increase risk include significant comorbid disease, especially renal disease, inoperable coronary artery disease, other lung disease, and massive obesity. At highly experienced centers, survival has continued to improve over time: before 1990, mortality was 15.8%; from 1998 to 2002, mortality was reduced to 4.4%.[38]

Pulmonary thromboendarterectomy is performed through a median sternotomy, with cardiopulmonary bypass, deep hypothermia, and periods of complete arrest (because of the brisk backbleeding from the increased bronchial collateral circulation), allowing better visualization in a bloodless field.[35] Dissection of the pulmonary artery starts in the proximal vessels and extends into the subsegmental branches, including the intima and some media. If an atrial septal defect or a patent foramen ovale is found, it is corrected.

Besides experiencing the usual complications associated with cardiac surgery, these patients have some specific problems, including reperfusion pulmonary edema in the previously obstructed areas, resulting in hypoxemia caused by shunting of blood from the previously unobstructed but remodeled pulmonary arterial bed into these newly opened areas.[35] Altered mental status is very common and seems to relate to the total circulatory arrest time.

Patients who are not candidates for pulmonary thromboendarterectomy may be considered for medical therapy with drugs used to treat other forms of pulmonary hypertension, although experience to date is limited.[39-41]

Prognosis

Without surgery, the survival rate for patients with CTEPH is poor. In patients whose mean pulmonary arterial pressure is 40 to 50 mm Hg, 5-year survival is 30%; in those whose pulmonary arterial pressure is greater than 50 mm Hg, 5-year survival is 10%.[35]

Mortality associated with pulmonary thromboendarterectomy ranges from 5% to 24% and is lowest in the institutions with the greatest experience. The patients who survive have improvement of functional status from NYHA class III or IV to NYHA class I or II.[35] The right heart recovers quickly, and the pulmonary arterial pressure may continue to fall for up to 1 year after surgery. Many patients are able to return to work.

CONDITIONS ASSOCIATED WITH PULMONARY ARTERIAL HYPERTENSION

Collagen Vascular Diseases

Pulmonary hypertension with little or no parenchymal lung disease is a life-threatening manifestation of the collagen vascular diseases.[42] The clinical course and pathologic changes are similar to those of IPAH. The pathophysiology is unknown, although autoantibodies, immunoglobulin, and complement have been found in the vessel walls, suggesting involvement of immune complexes.[43] Patients with scleroderma, particularly those with CREST syndrome (calcinosis, Raynaud phenomenon, esophageal involvement, sclerodactyly, and telangiectasias) have a high incidence of pulmonary hypertension; the incidence ranges from 2% to 35% in patients with scleroderma to 50% in those with CREST syndrome. The incidence in patients with other disorders varies from 23% to 53% in those with mixed connective tissue disease to 0.5% to 14% in those with systemic lupus erythematosus; it is rare in those with rheumatoid arthritis, Sjögren syndrome, and dermatomyositis. Treatment of pulmonary hypertension in patients with collagen vascular diseases should be patterned after the protocol for IPAH detailed above.[25,44,45]

The role of immunosuppressive therapy in patients with pulmonary hypertension secondary to collagen vascular diseases is unclear. Sanchez and coworkers[42] reviewed the literature and found improvement in pulmonary hemodynamics with various immunosuppressive regimens in seven of 11 published cases. There have been no randomized, controlled trials with these agents, and this approach is not recommended as the primary treatment strategy.

Although patients with pulmonary hypertension secondary to collagen vascular disorders are sometimes excluded from con-

sideration of lung transplantation, transplants have been performed in such patients, with prolonged survival.[46]

Several disorders are associated with vasculitis involving the pulmonary vasculature; however, pulmonary hypertension associated with vasculitis is rare. The exceptions to this rule are Takayasu disease [*see 15:VIII Systemic Vasculitis Syndromes*], in which pulmonary arteritis and hypertension are common findings,[47] and rheumatoid arthritis[48] and systemic lupus erythematosus, in which pulmonary hypertension occurs much less commonly.[49]

Pulmonary Veno-occlusive Disease

Pulmonary veno-occlusive disease is a rare but distinct form of pulmonary hypertension characterized by obstruction of the small intrapulmonary veins.[50] About one third of patients are children, and there are some cases that seem to be related to HIV infection, use of chemotherapy drugs, or bone marrow transplantation.

Patients with this disorder experience increasing dyspnea, sometimes with hemoptysis. Findings on chest radiography suggest left ventricular failure; such findings include enlarged pulmonary arteries, Kerley B lines, pulmonary edema, and pleural effusions. The diagnosis should be considered when these radiographic findings are associated with no echocardiographic evidence of left ventricular dysfunction, mitral valvular disease, or obstruction to flow in the left atrium. CT scan of the chest can provide support for the diagnosis[51] and can help exclude obstruction of the pulmonary veins in the mediastinum (fibrosing mediastinitis or tumor). The diagnosis can be confirmed by open lung biopsy, which will show the combination of chronic congestion, pulmonary hypertensive changes in the pulmonary arteries, and narrowing or occlusion of the pulmonary veins by eccentric or concentric intimal fibrosis.

Therapies that are used in IPAH are inconsistently successful in veno-occlusive disease, and epoprostenol in particular has been associated with the development of pulmonary edema. Accordingly, lung transplantation should be considered once the diagnosis is established.[50,52]

Hematologic Disorders

Pulmonary hypertension is common in adults with sickle cell disease and most likely is a complication of hemolysis resulting from interference with the nitric oxide regulatory pathway [*see 5:IV Hemoglobinopathies and Hemolytic Anemias*].[53]

Patients with chronic myeloproliferative disorders,[54] amyloidosis,[55] and the POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes) syndrome[56] have been noted to have complicating pulmonary hypertension.

Pulmonary Arteriovenous Malformations

Abnormal, direct communications between branches of the pulmonary arteries and veins, producing a right-to-left shunt, are called pulmonary arteriovenous malformations (AVMs). Such lesions may be acquired (e.g., as a result of trauma or as a complication of cirrhosis) [*see 4:IX Cirrhosis of the Liver*] or congenital.[57]

Congenital pulmonary AVMs can be either single or multiple. They may be isolated to the lung or, much more commonly, can occur as part of an autosomal dominant syndrome of widespread AVMs in skin, mucous membranes, lungs, and other internal organs. The autosomal dominant syndrome is known as Osler-Weber-Rendu disease or hereditary hemorrhagic telangiectasia (HHT)[57] [*see 5:XIII Hemorrhagic Disorders*].

Diagnosis Pulmonary AVMs may produce severe symptoms or complications, but most patients are asymptomatic. The diagnosis is suggested by solitary or multiple pulmonary nodules found on routine chest radiography.[58] Dyspnea is the most common symptom, occurring at first with exercise and later at rest. Dyspnea that worsens in the upright position and improves on reclining (platypnea) has been seen in these patients. It may be caused by worsening of the right-to-left shunt in the lung bases with the patient in the upright position, producing hypoxemia (orthodeoxia). Hemoptysis, usually mild and related to bronchial mucosal telangiectasia, is the second most common symptom. Neurologic complaints are also seen.[58,59] Headache, tinnitus, seizures, symptoms that mimic transient ischemic attack, and completed stroke may relate to complicating polycythemia, paradoxical embolus, or brain abscess. Patients with HHT may have bleeding from AVMs outside the lungs, which is often sufficient to produce anemia. Physical findings include mucous membrane telangiectasia in the patients with HHT, cyanosis, digital clubbing, and intrapulmonary bruits that increase with inspiration. Patients with pulmonary AVMs have hypoxemia, with the degree of hypoxemia being proportional to the magnitude of the shunt. As a consequence, chronic respiratory alkalosis and polycythemia may occur. Pulmonary function tests are normal except for a decreased diffusing capacity.

The diagnosis is most often suggested by the finding on chest radiograph of a well-defined solitary pulmonary nodule or multiple nodules with a feeding artery, draining vein, or both. Echocardiography with contrast can also suggest the diagnosis.[58] Spiral CT with contrast will confirm the vascular nature of lesions and may detect others that were not obvious from standard radiography. The definitive diagnostic procedure, however, is pulmonary angiography, which allows demonstration of all significant lesions; this is important in planning therapy.[59,60]

Treatment The natural history of pulmonary AVMs is poorly understood. In asymptomatic patients, the treatment decisions must be made with the recognition of the risk of hemoptysis, paradoxical embolization, and brain abscess.

The treatment options include thoracotomy with selective resection or percutaneous embolization. In patients with solitary or unilateral lesions, resection can be performed with minimal morbidity and mortality.[61] The only consideration is the possibility of the development or persistence of pulmonary hypertension after the low-resistance AVMs are removed in patients whose pulmonary vasculature has become remodeled in response to the high flow through the AVMs, or who have the ALK-1 genetic mutation associated with HHT and pulmonary hypertension.[62] In such patients, percutaneous balloon obstruction of the AVMs allows preoperative measurement of pulmonary hemodynamics.

The techniques and safety of percutaneous embolization have been improved. With the use of detachable balloons and metal coils, single or multiple AVMs can be treated.[61] The main risk is systemic embolization of the obstructing material. This technique is rapidly replacing surgery as the treatment of choice.

Pulmonary Arterial Aneurysms

Aneurysms of the pulmonary arteries are very rare.[63] They are seen as congenital anomalies, often along with abnormalities of the other great vessels and heart; as part of connective tissue disorders (e.g., Marfan syndrome); as a consequence of trauma (pseudoaneurysms caused by Swan-Ganz catheters may be confused for pulmonary aneurysms), infection (syphilis, tuberculo-

sis, pyogenic bacteria, or fungi), or immunologic disorders (Behçet disease[64] or Hughes-Stovin syndrome); and in association with pulmonary diseases such as pulmonary hypertension or bronchiectasis.

In many cases, pulmonary arterial aneurysms are asymptomatic. However, cough, dyspnea, and particularly hemoptysis can be seen. Rupture of the aneurysm into an airway can be associated with sudden, massive, and usually fatal hemoptysis. Dissection of the pulmonary artery can also occur. The diagnosis can be made from a contrast-enhanced CT scan or by pulmonary angiography.

Aneurysms of the main and proximal right and left pulmonary arteries can be surgically repaired.[65] Aneurysms in smaller pulmonary arteries can be removed by surgical resection of the involved area of lung.

OTHER CAUSES OF PULMONARY HYPERTENSION

Residence at high altitude may be associated with the development of pulmonary hypertension.[66] This response is maladaptive and occurs in only a small portion of the population living at high altitude.[67] In these patients, moving to a lower altitude best reduces pulmonary arterial pressure. Patients with obstructive sleep apnea without other forms of lung disease often have complicating pulmonary hypertension.[68] Development of pulmonary hypertension is associated with greater body mass index and degree of daytime hypoxemia, small airway closure during tidal breathing, and heightened pulmonary pressor responses to hypoxia and increased pulmonary blood flow.[69,70] Treatment of sleep apnea is associated with improvement in pulmonary hemodynamics[70] [*see 14:VI Ventilatory Control during Wakefulness and Sleep*]. Because these patients are often sedentary, obese, and have risk factors for left-sided heart disease, the clinical evaluation should assess for recurrent pulmonary thromboembolic disease and left ventricular dysfunction (systolic and/or diastolic) as underlying causes for the pulmonary hypertension.

Compression or stenosis of the pulmonary arteries or veins in the mediastinum by tumor or fibrosing mediastinitis [*see 14:IX Disorders of the Pleura, Hila, and Mediastinum*] or stenosis of multiple peripheral segments of the pulmonary artery can produce precapillary or postcapillary pulmonary hypertension. These disorders can usually be diagnosed by spiral CT or pulmonary angiography.

A rare cause of pulmonary hypertension is pulmonary capillary hemangiomatosis, a disorder that clinically mimics or can be a component of pulmonary veno-occlusive disease.[1,71] The lung biopsy specimens of these patients have patchy regions of severe congestion that contain capillary-sized blood vessels that appear to invade the walls of the pulmonary veins and, to a lesser degree, the pulmonary arteries. A CT scan can assist in differentiating this disorder from IPAH.[72] The only effective treatment is lung transplantation.

The author has received grants for clinical research and served as an advisor or consultant to Actelion, Pfizer, Inc., Cotherix, Inc., Myogen, Inc., and Schering-Plough Corp.

Sildenafil has not been approved by the FDA for uses described in this chapter.

References

1. Simonneau G, Galie N, Rubin LJ, et al: Clinical classification of pulmonary hypertension. J Am Coll Cardiol 43:5S, 2004

2. Rubin LJ: Introduction. Diagnosis and Management of Pulmonary Arterial Hypertension: ACCP Evidence-Based Clinical Practice Guidelines. Chest 176:7S, 2004

3. Humbert M, Morrell NW, Archer SL, et al: Cellular and molecular pathobiology of pulmonary arterial hypertension. J Am Coll Cardiol 43:13S, 2004

4. Pietra GG, Capron F, Stewart S, et al: Pathologic assessment of vasculopathies in pulmonary hypertension. J Am Coll Cardiol 43:25S, 2004

5. McGoon M, Gutterman D, Steen V, et al: Screening, early detection and diagnosis of pulmonary arterial hypertension. Chest 126:14S, 2004

6. Vieillard-Baron A, Page B, Augarde R, et al: Acute cor pulmonale in massive pulmonary embolism: incidence, echocardiographic pattern, clinical implications and recovery rate. Intensive Care Med 27:1481, 2001

7. Vieillard-Baron A, Schmitt JM, Augarde R, et al: Acute cor pulmonale in acute respiratory distress syndrome submitted to protective ventilation: incidence, clinical implications, and prognosis. Crit Care Med 29:1551, 2001

8. Goldhaber SZ: Echocardiography in the management of pulmonary embolism. Ann Intern Med 136:691, 2002

9. Arcasoy SM, Kreit JW: Thrombolytic therapy of pulmonary embolism: a comprehensive review of current evidence. Chest 115:1695, 1999

10. Borst MM, Leschke M, Konig U, et al: Repetitive hemodilution in chronic obstructive pulmonary disease and pulmonary hypertension: effects on pulmonary hemodynamics, gas exchange, and exercise capacity. Respiration 66:225, 1999

11. Schonhofer B, Barchfeld T, Wenzel M, et al: Long term effects of non-invasive mechanical ventilation on pulmonary haemodynamics in patients with chronic respiratory failure. Thorax 56:524, 2001

12. Galie N, Manes A, Branzi: Prostanoids for pulmonary arterial hypertension. Am J Respir Med 2:123, 2003

13. Runo JR, Loyd JE: Primary pulmonary hypertension. Lancet 361:1533, 2003

14. Thomas AQ, Gaddipati R, Newman JH, et al: Genetics of primary pulmonary hypertension. Clin Chest Med 22:477, 2001

15. Lilienfeld DE, Rubin LJ: Mortality from primary pulmonary hypertension in the United States, 1979–1996. Chest 117:796, 2000

16. Humbert M, Nunes H, Sitbon O, et al: Risk factors for pulmonary arterial hypertension. Clin Chest Med 22:459, 2001

17. Michelakis ED, Weir EK: Anorectic drugs and pulmonary hypertension from the bedside to the bench. Am J Med Sci 321:292, 2001

18. Seoane L, Shellito J, Welsh D, et al: Pulmonary hypertension associated with HIV infection. South Med J 94:635, 2001

19. Farber HW, Loscalzo J: Pulmonary artery hypertension. N Engl J Med 351:1655, 2004

20. Tuder RM, Yeager ME, Geraci M, et al: Severe pulmonary hypertension after the discovery of the familial primary pulmonary hypertension gene. Eur Respir J 17:1065, 2001

21. Archer S, Rich S: Primary pulmonary hypertension: a vascular biology and translational research "work in progress." Circulation 102:2781, 2000

22. Salvi SS: Alpha$_1$-adrenergic hypothesis for pulmonary hypertension. Chest 115:1708, 1999

23. Botney MD: Role of hemodynamics in pulmonary vascular remodeling: implications for primary pulmonary hypertension. Am J Respir Crit Care Med 159:361, 1999

24. Naeije R, Vachiery JL: Medical therapy of pulmonary hypertension: conventional therapies. Clin Chest Med 22:517, 2001

25. Liu C, Cheng J: Endothelin receptor antagonists for pulmonary arterial hypertension. Cochrane Database Syst Rev (1)CD004434, 2005

26. Simonneau G, Barst RJ, Galie N, et al: Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. Am J Respir Crit Care Med 165:800, 2002

27. Ghofrani HA, Wiedemann R, Rose F, et al: Combination therapy with oral sildenafil and inhaled iloprost for severe pulmonary hypertension. Ann Intern Med 136:515, 2002

28. Clapp LH, Finney P, Turcato S, et al: Differential effects of stable prostacyclin analogues on smooth muscle proliferation and cyclic AMP generation in human pulmonary artery. Am J Respir Cell Mol Biol 26:194, 2002

29. Doyle RL, McCrory D, Channick RN, et al: Surgical treatments/interventions for pulmonary arterial hypertension: ACCP evidence-based clinical practical guidelines. Chest 126:63S, 2004

30. Trulock EP: Lung transplantation for primary pulmonary hypertension. Clin Chest Med 22:583, 2001

31. Gaine SP, Rubin LJ: Medical and surgical treatment options for pulmonary hypertension. Am J Med Sci 315:179, 1998

32. D'Alonzo GE, Barst RJ, Ayres SM, et al: Survival in patients with primary pulmonary hypertension: results from a national prospective registry. Ann Intern Med 115:343, 1991

33. Kawut SM, Horn EM, Gerekashvilil KK, et al: New predictors of outcome in idiopathic pulmonary arterial hypertension. Am J Cardiol 95:199, 2005

34. Pengo V, Lensing AW, Prins, MH, et al: Incidence of chronic thromboembolic pulmonary hypertension after pulmonary embolism. N Engl J Med 350:2257, 2004

35. Fedullo PF, Auger WR, Kerr KM, et al: Chronic thromboembolic pulmonary hypertension. N Engl J Med 345:1465, 2001

36. Bonderman D, Jakowitsch J, Adlbrecht C, et al: Medical conditions increasing the risk of chronic thromboembolic pulmonary hypertension. Thromb Haemost 93:512, 2005

37. Dartevelle P, Fadel E, Mussot S, et al: Chronic thromboembolic pulmonary hypertension. Eur Respir J 23:637, 2004

38. Jamieson SE, Kapelanski DP, Sakakibara N, et al: Pulmonary endartectomy: experience and lessons learned in 1,500 cases. Ann Thorac Surg 76:1457, 2003

39. Ono F, Nagay N, Okumura H, et al: Effect of orally active prostacyclin analogue on survival in patients with chronic thromboembolic pulmonary hypertension without major vessel obstruction. Chest 123:1583, 2003

40. Ghofrani HA, Schermuly RT, Rose F, et al: Sildenafil for long-term treatment of nonoperable chronic thromboembolic pulmonary hypertension. Am J Respir Crit Care Med 167:1139, 2003

41. Bresser P, Fedullo PF, Auger WG, et al: Continuous intravenous epoprosterol for chronic thromboembolic pulmonary hypertension. Eur Respir J 23:595, 2004

42. Sanchez O, Humbert M, Sitbon O, et al: Treatment of pulmonary hypertension secondary to connective tissue diseases. Thorax 54:273, 1999

43. Quismorio FP Jr, Sharma O, Koss M, et al: Immunopathologic and clinical studies in pulmonary hypertension associated with systemic lupus erythematosus. Semin Arthritis Rheum 13:349, 1984

44. Barst RJ, McGoon M, McLaughlin V, et al: Beraprost therapy for pulmonary arterial hypertension. J Am Coll Cardiol 41:2219, 2003

45. Badesch DB, Tapson VF, McGoon MD, et al: Continuous intravenous epoprostenol for pulmonary hypertension due to the scleroderma spectrum of disease: a randomized, controlled trial. Ann Intern Med 132:425, 2000

46. Huerd SS, Hodges TN, Grover FL, et al: Secondary pulmonary hypertension does not adversely affect outcome after single lung transplantation. J Thorac Cardiovascular Surg 119:458, 2000

47. Robles M, Reyes PA: Takayasu's arteritis in Mexico: a clinical review of 44 consecutive cases. Clin Exp Rheumatol 12:381, 1994

48. Young ID, Ford SE, Ford PM: The association of pulmonary hypertension with rheumatoid arthritis. J Rheumatol 16:1266, 1989

49. Rubin LA, Geran A, Rose TH, et al: A fatal pulmonary complication of lupus in pregnancy. Arthritis Rheum 38:710, 1995

50. Mandel J, Mark EJ, Hales CA: Pulmonary veno-occlusive disease. Am J Respir Crit Car Med 162:1964, 2000

51. Swensen SJ, Tashjian JH, Myers JL, et al: Pulmonary venoocclusive disease: CT findings in eight patients. AJR Am J Roentgenol 167:937, 1996

52. Holcomb BW Jr, Loyd JE, Ely EW, et al: Pulmonary veno-occlusive disease: a case series and new observations. Chest 118:1671, 2000

53. Gladwin MT, Sachdev V, Jison ML, et al: Pulmonary hypertension as a risk factor for death in sickle cell disease. N Eng J Med 350: 886, 2004

54. Dingli D, Utz JP, Krowka MJ, et al: Unexplained pulmonary hypertension in chronic myeloproliferative disorders. Chest 120:801, 2001

55. Dingli D, Utz JP, Gertz MA: Pulmonary hypertension in patients with amyloidosis. Chest 120:1735, 2001

56. Paciocco G, Bossone E, Erba H, et al: Reversible pulmonary hypertension in POEMS syndrome: another etiology of triggered pulmonary vasculopathy? Can J Cardiol 16:1007, 2000

57. Gossage JR, Kanj G: Pulmonary arteriovenous malformations: a state of the art review. Am J Respir Crit Care Med 158:643, 1998

58. Swanson KL, Prakash UB, Stanson AW: Pulmonary arteriovenous fistulas: Mayo Clinic experience, 1982–1997. Mayo Clinic Proc 74:671, 1999

59. Maher CO, Piepgras DG, Brown RD Jr, et al: Cerebrovascular manifestations in 321 cases of hereditary hemorrhagic telangiectasia. Stroke 32:877, 2001

60. Kjeldsen AD, Oxhoj H, Andersen PE, et al: Pulmonary arteriovenous malformations: screening procedures and pulmonary angiography in patients with hereditary hemorrhagic telangiectasia. Chest 116:432, 1999

61. Gupta P, Mordin C, Curtis J, et al: Pulmonary ateriovenous malformations: effect of emoblization on right-to-left shunt, hypoxemia, and exercise tolerance in 66 patients. AJR Am J Roentgenol 179: 347, 2002

62. Trembath RC, Thomson JR, Machado RD, et al: Clinical and molecular genetic features of pulmonary hypertension in patients with hereditary hemorrhagic telengiectasia. N Engl J Med 345:325, 2001

63. Bartter T, Irwin RS, Nash G: Aneurysms of the pulmonary arteries. Chest 94:1065, 1988

64. Erkan F, Gul A, Tasali E: Pulmonary manifestations of Behçet's disease. Thorax 56:572, 2001

65. Kuwaki K, Morishita K, Sato H, et al: Surgical repair of the pulmonary trunk aneurysm. Eur J Cardiothorac Surg 18:535, 2000

66. Naeije R: Pulmonary circulation at high altitude. Respiration 64:429, 1997

67. Groves BM, Droma T, Sutton JR, et al: Minimal hypoxic pulmonary hypertension in normal Tibetans at 3,658 m. J Appl Physiol 74:312, 1993

68. Bady E, Achkar A, Pascal S, et al: Pulmonary arterial hypertension in patients with sleep apnoea syndrome. Thorax 55:934, 2000

69. Sajkov D, Wang T, Saunders NA, et al: Daytime pulmonary hemodynamics in patients with obstructive sleep apnea without lung disease. Am J Respir Crit Care Med 159:1518, 1999

70. Sajkov D, Wang T, Saunders NA, et al: Continuous positive airway pressure treatment improves pulmonary hemodynamics in patients with obstructive sleep apnea. Am J Respir Crit Care Med 165:152, 2002

71. Eltorky MA, Headley AS, Winer-Muram H, et al: Pulmonary capillary hemangiomatosis: a clinicopathologic review. Ann Thorac Surg 57:772, 1994

72. Dufour B, Maitre S, Humbert M, et al: High-resolution CT of the chest in four patients with pulmonary capillary hemangiomatosis or pulmonary venoocclusive disease. AJR Am J Roentgenol 171:1321, 1998

I INTRODUCTION TO THE RHEUMATIC DISEASES

SHAUN RUDDY, M.D., F.A.C.P.

In 2002, arthritis and chronic joint disease affected one in three adults in the United States, or 70 million persons. Arthritis and musculoskeletal disorders are the leading causes of disability in persons 18 to 65 years of age and are common causes of disability related to employment. Risk factors for these disorders include lower education level, sedentary lifestyle, high body-mass index, and female sex. Age is a major risk factor: the prevalence of these disorders increases with age, with more than 60% of those age 65 and older being affected[1] [*see Figure 1*]. By 2030, when the percentage of the population that is 65 and older will have risen to 20%, the number of people in this age group with arthritis will nearly double, from 21.4 million to 41.1 million.[2]

One in seven patients who visit a physician's office has a complaint regarding the musculoskeletal system.[3] Although many of these patients have benign, self-limited conditions that respond to simple remedies, some patients have serious, complex problems for which timely intervention may be crucial for a successful outcome.

Approach to Diagnosis

Diagnosis of the rheumatic diseases primarily relies on the history and physical examination.[4,5] Expensive laboratory and imaging studies are usually of little use; most have low sensitivity and specificity, and their findings are seldom definitively diagnostic. For example, the serum level of uric acid, a substance intimately involved in the pathogenesis of gout, is of little use in diagnosing gouty arthritis. Uric acid levels are normal in 20% of patients with gout (i.e., false negative results), and most persons with elevated levels will never have gouty arthritis (false positive results) [*see 15:IX Crystal-Induced Joint Disease*]. Wide-ranging and expensive investigations often lead to the wrong conclusions. The dearth of useful laboratory aids to diagnose rheumatic disease makes the clinical skills of the physician very important and the diagnosis of such disease an exciting undertaking.

History

A careful and detailed history is the most important part of the evaluation of a patient with arthritis. It focuses the subsequent physical examination and laboratory studies. Relevant factors in the history include the patient's age, sex, and race. In addition, the physician should elicit the location of the pain, which includes both the distribution of affected joints and the point of origin of the pain (i.e., whether it arises from a joint or surrounding structures) [*see Table 1*]. Symmetrical involvement of the small joints of the hands and feet but not the distal interphalangeal joints suggests rheumatoid arthritis. In contrast, distal interphalangeal involvement often occurs in psoriatic arthritis. Bony overgrowth with distal interphalangeal involvement suggests osteoarthritis. Asymmetrical involvement of the large joints accompanied by back pain in a young man is characteristic of spondyloarthropathy. The sudden onset of severe pain in the great toe is a feature of gout (podagra).

Knowing the nature of the pain helps the physician determine whether the disease is inflammatory or noninflammatory [*see Table 1*]. The date of onset and the temporal course of the pain should also be elicited [*see Figure 2*]. The new involvement of joints that had not previously been affected in a patient with joint involvement is common in rheumatoid arthritis and systemic lupus erythematosus (SLE). Migratory polyarthritis, in which one affected joint becomes asymptomatic as another becomes painful, occurs in gonococcal arthritis, Reiter syndrome, and acute rheumatic fever. Intermittent arthritis, in which asymptomatic intervals are punctuated by acute flares, is common in crystal-induced arthritis.

RHEUMATOLOGIC EMERGENCIES

Although rheumatologic emergencies seldom occur, failure to recognize any of a few important symptoms or signs [*see Table 2*] may result in permanent disability or death; thus, the initial contact with the patient requires being alert for these conditions.[5] If a rheumatologic emergency is suspected, prompt initiation of appropriate diagnostic testing and treatment is essential.

After emergencies have been excluded, the approach to diagnosis need not be rushed. Although patients often come to the office with the preconceived notion that they have arthritis and expect quick diagnostic confirmation and treatment, they often must be observed for some time before the diagnosis can be made. Considerable patience and tact may be required in communicating this need to the patient. He or she should be in-

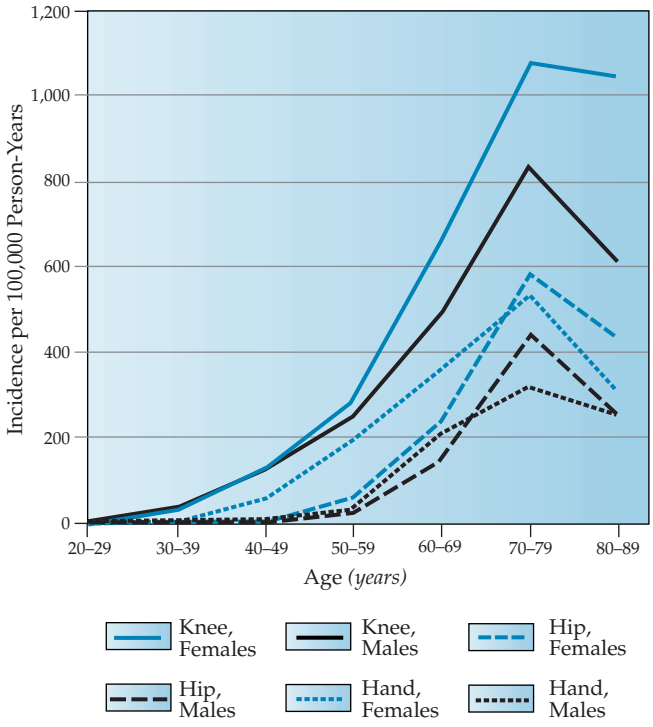


Figure 1 **Incidence of osteoarthritis of the hand, hip, and knee, in members of the Fallon Community Health Plan, 1991–1992, by age and sex.[20]**

Table 1 Common Rheumatic Diseases

	Intra-articular	*Extra-articular*
Inflammatory	Rheumatoid arthritis Systemic lupus erythematosus Septic arthritis Gout Pseudogout Spondyloarthropathy	Tendinitis Bursitis Polymyositis Vasculitis
Noninflammatory	Osteoarthritis	Fibromyalgia

formed that an uncertain diagnosis at presentation is usually a good prognostic sign; the prognosis for patients with clear-cut, easily diagnosed disease is often poor. Temporizing with an indefinite diagnosis—for example, knee pain of uncertain etiology—is better than prematurely classifying a patient's musculoskeletal complaints into a particular diagnostic category. In at least 10% of patients, a diagnosis cannot be made with certainty.

LOCATION OF PAIN

Patients are often not precise in identifying the location of musculoskeletal pain; it helps to ask them to put their hands on the place that hurts. A patient who complains of pain in the hip may point to one of three areas: the inguinal ligament and anterior thigh, suggesting involvement of the true hip joint; the lateral hip girdle, characteristic of trochanteric bursitis; or the buttock, consistent with sacroiliac joint disease or radiation of back pain along the sciatic nerve. During the physical examination, detailed attention to the location of pain helps identify the structures affected. If a specific structure cannot be identified, the pain may be referred from elsewhere. Pain described as diffuse or poorly localized may suggest a serious systemic disease, such as polymyositis, but more commonly is a manifestation of fibromyalgia or a related pain syndrome.

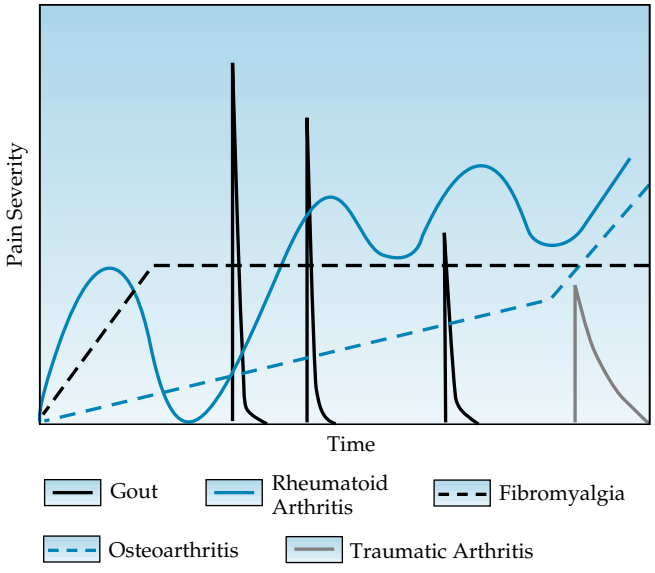


Figure 2 **Variations in the temporal course and severity of pain in the rheumatic diseases. The pain of rheumatoid arthritis may vary in intensity over time. The pain of osteoarthritis tends to be slowly progressive until late in the course of disease, when it may become very severe. The pain of fibromyalgia remains constant. Gout has an intermittent course of high-intensity pain flares. Traumatic arthritis improves more slowly than gout.[4]**

CHARACTER OF PAIN

Pain that arises from joints and that is dull, alleviated by rest, and worsened by weight bearing or movement of the joint suggests joint damage such as that which occurs in osteoarthritis. Pain that is more intense, accompanied by swelling, and present when the patient is at rest or pain that awakens the patient at night suggests an inflammatory process such as rheumatoid arthritis. Inflammation also causes stiffness after prolonged immobility—the so-called gelling phenomenon, which may last longer than 30 minutes. Patients who have inflammatory joint pain caused by rheumatoid arthritis or SLE typically have significant morning stiffness that lasts as long as several hours and that improves as the day goes on—or at least until they are overwhelmed by the deep fatigue that also accompanies these diseases. The duration of morning stiffness is a rough measure of the activity of the inflammatory disease. Patients who have the noninflammatory joint pain that characterizes osteoarthritis, tendinitis, or bursitis describe focal morning pain that lasts for a few minutes; they are most uncomfortable at the end of the day, after prolonged activity. Shooting, burning, or so-called pins-and-needles discomfort is usually neurogenic.

Symptoms of giving way or locking in a weight-bearing joint generally suggest a mechanical process, such as a cartilage or ligament tear or a loose body within the joint. Locking may also occur when soft tissue inflammation impairs mobility, as when triggering of a finger is caused by a nodule in a flexor tendon as a result of tenosynovitis.

WEAKNESS

A complaint of weakness may reflect one of several disease processes. In patients with focal loss of muscle power in a specific region, indicated by complaints such as, "I can't raise my arm," the obvious possibilities include a neurologic lesion; local

Table 2 Symptoms and Signs of Rheumatologic Emergencies[5]

Symptom or Sign	*Condition*
History of significant trauma	Fracture, compartmental syndrome, rhabdomyolysis
Systemic symptoms (fever, weight loss)	Infection (septic arthritis, osteomyelitis, endocarditis, necrotizing fasciitis)
Weakness	
Focal	Radiculopathy, entrapment neuropathy, compartmental syndrome, motor neuron disease, mononeuritis multiplex
Global or progressive	Spinal cord compression, myelopathy, transverse myelitis, myositis, response to toxin
Neurogenic pain (burning, numbness, paresthesias)	Radiculopathy, entrapment neuropathy
Point tenderness	Fracture
Red, hot, swollen joint	Septic arthritis, gout, pseudogout
Asymmetrical painful, swollen leg	Deep vein thrombosis
Diffuse shoulder and hip pain, headache in elderly patient	Temporal arteritis
Bowel or bladder dysfunction	Cauda equina syndrome

muscle atrophy; failure of a musculoskeletal unit, such as a ruptured tendon or torn muscle; or weakness secondary to joint pain. A patient who says, "I feel weak all over," should be asked to distinguish true muscle weakness (i.e., loss of power or endurance) from a more generalized sense of asthenia, malaise, or fatigue. The patient with true muscle weakness has difficulty starting or maintaining an activity that requires muscle strength, such as arising from a chair, getting out of bed, or walking up stairs. Physical examination usually confirms the muscular nature of such problems, with the proximal musculature being most affected in myopathies and the distal muscles involved in neuropathies. If the examination shows no objective weakness or other neurologic abnormalities, such as loss of deep tendon reflexes, the patient's report of weakness probably corresponds to asthenia, fatigue, and loss of sense of well-being. If this occurs together with other constitutional symptoms, such as anorexia, weight loss, and low-grade fever, it may indicate an active systemic rheumatic disease such as rheumatoid arthritis or SLE. If the asthenia and fatigue are not accompanied by objective constitutional symptoms, fibromyalgia is more likely.

FUNCTIONAL IMPAIRMENT AND PSYCHOSOCIAL FACTORS

After pain, loss of function is of the greatest concern to patients, making the functional assessment an important part of the history. Exploring this area often affords the physician insights into the way the patient views the disease and what the patient expects. Simple questions asked by the physician may include the following: How has your disease affected you? What can you not do now that you could do before? What can you not do that you would like to do? Patients with significant functional impairment should be asked more-detailed questions about routine activities of daily living. Explicit questions about sexual function may uncover problems that the patient might otherwise hesitate to describe. The sexual history also identifies risk factors for sexually transmitted joint diseases, such as gonococcal arthritis, Reiter syndrome, and HIV-associated arthropathy. A vocational history may identify specific tasks that exacerbate the disease and that need modification; it may also indicate whether the patient is likely to be exposed to ticks carrying Lyme disease. The history of use of devices such as a cane or crutches and information as to when the patient began using them is helpful in assessing the temporal course of the disease. Claims for workers' compensation or disability or other pending litigation should also be noted.

FAMILY HISTORY

Many rheumatic diseases have strong familial predispositions. Ankylosing spondylitis is the best known, but other autoimmune diseases, such as rheumatoid arthritis and SLE, also have a genetic basis, as do gout and, to a lesser extent, pseudogout.

REVIEW OF SYSTEMS

Once a specific disease is suspected, the patient should be questioned about the presence of other systemic features of the disease. A partial list of clues to systemic disease that can be gleaned from the review of systems is shown [*see Table 3*].

Table 3 Diagnostic Clues from the Review of Systems

Feature	*Condition*
Facial rash, photosensitivity	Systemic lupus erythematosus
Hair loss	Systemic lupus erythematosus
Uveitis	Spondyloarthropathy, sarcoidosis, Behçet syndrome
Conjunctivitis	Reactive arthritis
Dry eyes and/or mouth	Sjögren syndrome
Oral ulcers	Systemic lupus erythematosus
Nasal discharge, ulcers	Wegener granulomatosis
Diabetes	Diabetic stiff-hand syndrome, Dupuytren contractures, reflex sympathetic dystrophy, carpal tunnel syndrome
Thyroid problems	Osteoporosis, myopathy
Pleurisy	Systemic lupus erythematosus
Hemoptysis	Systemic lupus erythematosus, Wegener granulomatosis
Abdominal pain, bowel dysfunction	Crohn disease, fibromyalgia (with irritable bowel syndrome)
Urethral burning or discharge	Reactive arthritis, septic arthritis
Kidney stones	Gout
Numbness or paresthesias	Carpal tunnel syndrome, vasculitis
Vaginal burning or discharge	Behçet syndrome, septic arthritis

Physical Examination

GENERAL EXAMINATION

By the time the history is finished, the clinician should have formulated some diagnostic hypotheses that can be used to guide the physical exam. If a patient's complaints are localized and the history elicits no suggestion of a more generalized process, the examination may be limited to the region involved. When systemic symptoms are present, a complete and detailed examination is required. Asymmetrical joint disease and inflammatory back pain raise the possibility of psoriatic arthritis and spondylitis, in which case the physician should very carefully inspect the skin for a patch of psoriasis of which the patient may be unaware. In the sexually active patient with asymmetrical involvement of large joints and a history of conjunctivitis, detailed examination of the genitalia for the lesions of Reiter syndrome should be performed. The patient with complaints of sinus trouble, arthritis, and hemoptysis should undergo thorough scrutiny of the nasopharynx and sinuses for the lesions of Wegener granulomatosis. Other important findings are iritis or conjunctivitis, nodules, pericardial or pleural rubs, hepatic or splenic enlargement, lymphadenopathy, and neurologic abnormalities.

JOINT EXAMINATION

By looking at and palpating the joints, the physician can identify the precise anatomic structures that are the source of the patient's pain and decide whether the pain is caused by inflammation. A goal of the examination is to reproduce the patient's pain, either by motion of the joint or by palpation. Frank redness of the skin overlying a joint is unusual; however, increased temperature, best detected by palpation with the backs of the fingers, is not unusual and, when present, indicates inflammation.

Apparent swelling may be caused by periarticular edema, an effusion in the joint, synovial proliferation, or bony overgrowth; in all cases, it indicates organic disease. Palpation for tenderness may reveal whether the problem lies within the joint or is discretely localized to an overlying bursa or tendon sheath. The finding of fine crepitus with motion of the structure corresponds to the grinding of subchondral bone, denuded of articular cartilage, against opposing bone. Coarser (so-called creaking) crepitus is associated with the fibrinous tendinitis that occurs in scleroderma or traumatic tendinitis. When the examiner is able to move the joint through a passive range of motion that exceeds the active range of motion accomplished voluntarily by the patient, failure of a musculoskeletal unit (e.g., rupture of the rotator cuff in the shoulder) or a neurologic deficit should be suspected. Examination of the spine is the most neglected part of the musculoskeletal examination, probably because it entails moving the patient from the sitting position through the supine and prone positions and then to the standing position. Patients with findings suggestive of a lumbar radiculopathy should have detailed testing of the motor, sensory, and reflex systems in the legs. Reviews of such testing are available on the Internet, at http://www.neuropat.dote.hu/neurology.htm and at http://www.neuroexam.com. An experienced physician can perform a complete musculoskeletal examination in less than 10 minutes.

Laboratory Studies

THE ACUTE-PHASE RESPONSE

The cellular response to inflammation or tissue injury elaborates cytokines such as interleukin-1 (IL-1), IL-6, and tumor necrosis factor (TNF), which have profound effects on the hepatic synthesis of plasma proteins.[6] Concentrations of C-reactive protein (CRP) increase as much as 100-fold within 1 or 2 days after tissue damage; parallel increases in serum amyloid A protein occur. Slower and less marked increases occur in coagulation proteins such as fibrinogen and prothrombin, most of the complement components, normal plasma protease inhibitors, and transport proteins such as ferritin and haptoglobin. Corresponding decreases occur in serum transferrin and albumin, accounting for the low serum iron and albumin levels that accompany inflammatory diseases.

The erythrocyte sedimentation rate (ESR) is the time-honored test used to detect the acute-phase response.[7] The increased rate of fall of the column of erythrocytes is caused by stacking of the cells into rouleaux, induced mainly by increases in the highly asymmetrical fibrinogen molecule; the increased levels of immunoglobulins seen in chronic inflammatory conditions also favor rouleaux formation. The ESR is influenced by many extraneous factors [*see Table 4*], the most important of which is age. The upper limit of normal for men is obtained by dividing the age in years by 2; for women, it is obtained by adding 10 to the age in years and then dividing that number by 2.

The CRP level is measured by immunoassay and is not influenced by most of the extraneous factors that affect the ESR. It also increases more rapidly than the ESR, which may take several days to increase.

The ESR and CRP tests are nonspecific: ESR and CRP levels may be elevated in a number of inflammatory conditions, such as malignancy, chronic infection, pneumonia, and acute myocardial infarction. Even increases in very low levels of CRP, as detected on a so-called ultrasensitive test, serve as a biomarker for atherosclerosis.[8] These tests are most useful in excluding significant inflammatory disease. For example, in a patient with diffuse pain and tenderness at trigger points, suggesting fibromyalgia, a normal ESR value supports this diagnosis; an ESR of 90 mm/hr dictates close scrutiny for other diseases. If the patient with these symptoms is older than 60 years, polymyalgia rheumatica and temporal arteritis are the primary possibilities. The acute-phase tests are also moderately useful in distinguishing inflammatory from noninflammatory arthritis and in monitoring the course of an inflammatory disease such as rheumatoid arthritis or polymyalgia rheumatica.

Table 4 Factors That Influence ESR[7]

Factors That Increase ESR	*Factors That Decrease ESR*
Advancing age	Congestive heart failure
Female sex	Sickle cell disease
Pregnancy	Altered erythrocyte shape (e.g., anisocytosis, spherocytosis, acanthosis, microcytosis)
Hypercholesterolemia	
B cell neoplasm (e.g., myeloma, macroglobulinemia, cryoglobulinemia)	Polycythemia
	Extreme leukocytosis
	Cachexia
Renal failure	Hypofibrinogenemia
	Cryoglobulinemia

ESR—erythrocyte sedimentation rate

IMMUNOLOGIC TESTS

As a class, immunologic tests have low specificity and only moderate sensitivity. They are also more expensive and less reproducible than most other clinical laboratory tests. They should never be used as screening tests; their greatest utility occurs when the pretest probability of disease is high. The misuse of immunologic tests frequently confounds the diagnosis. Two common examples of unnecessary rheumatology referrals are the octogenarian with arthritis of the knees or shoulders who tests positive for rheumatoid factor (unnecessary because positivity in healthy persons increases with age, and osteoarthritis of these joints is common in octogenarians) and the young woman with fatigue; diffuse pains; and a positive, usually low-titer, antinuclear antibody test (unnecessary because low titers of antinuclear antibody are found in as many as 32% of young women[9]; the patient probably has depression or fibromyalgia). The use of so-called arthritis panels, in which many serologic tests are bundled together, increases the likelihood of an abnormal result in a patient without rheumatic disease and should be avoided. Highly specific (and expensive) tests, such as those for antineutrophil cytoplasmic antibody, Lyme disease serology, HLA-B27, antiphospholipid antibody, and antibodies against individual nuclear constituents, should be performed only when the pretest probability of a particular disease is high. In one study of patients with swelling of at least one joint, the rheumatoid factor test had a sensitivity of 65% and a specificity of 87% for the diagnosis of rheumatoid arthritis.[10] At disease onset, a positive test result for rheumatoid factor is predictive of increased severity, as assessed by radiographic evidence of erosions.[11] Antibodies to cyclic citrullinated peptides (anti-CCP) formed by posttranslational modification of arginine residues in proteins such as fillagrin are found in the serum of many patients with rheumatoid arthritis. The 70% sensitivity and 98% specificity of the anti-CCP assays make them very useful in the diagnosis of rheumatoid arthritis, especially during the early

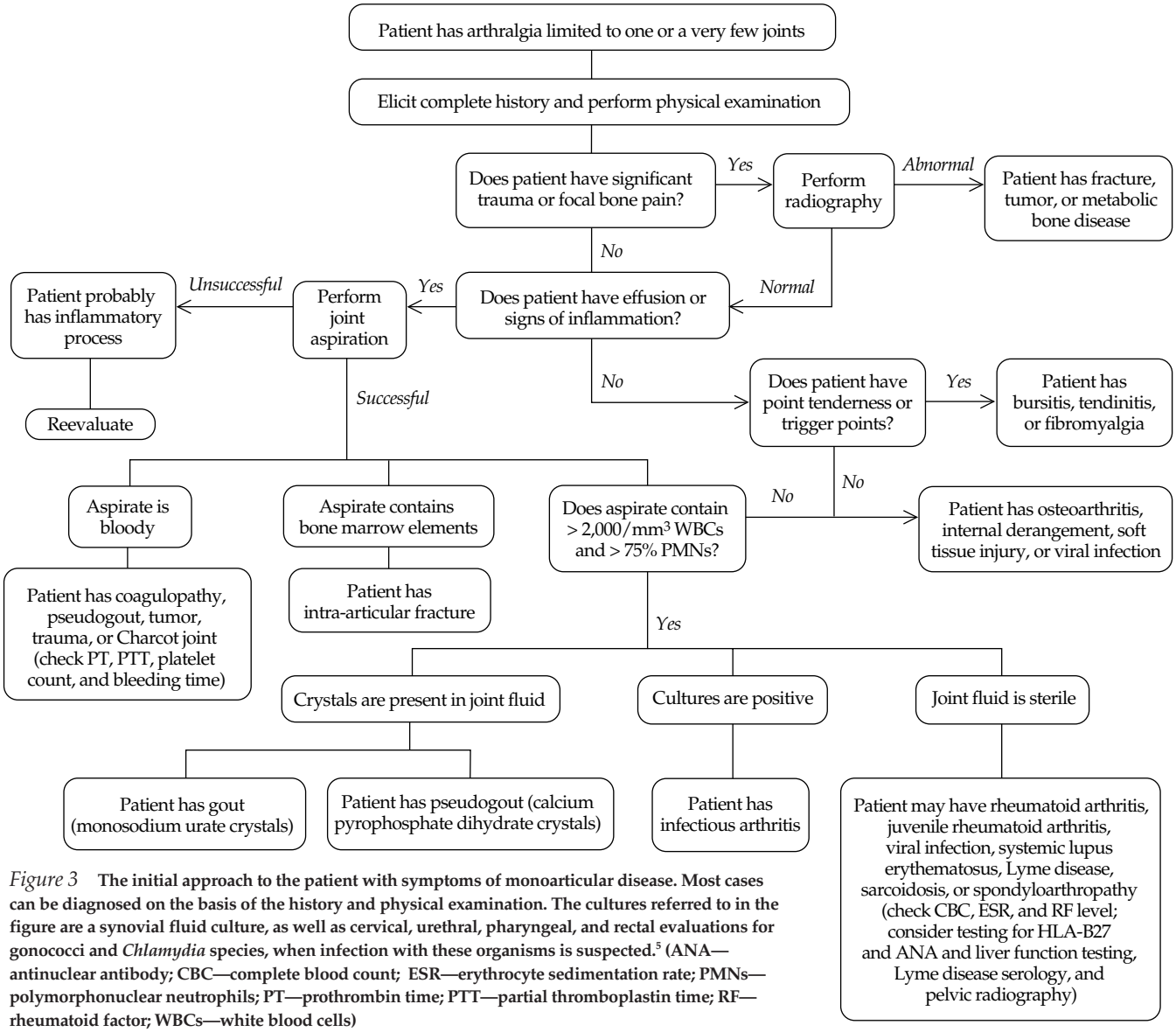


Figure 3 **The initial approach to the patient with symptoms of monoarticular disease. Most cases can be diagnosed on the basis of the history and physical examination. The cultures referred to in the figure are a synovial fluid culture, as well as cervical, urethral, pharyngeal, and rectal evaluations for gonococci and *Chlamydia* species, when infection with these organisms is suspected.[5] (ANA—antinuclear antibody; CBC—complete blood count; ESR—erythrocyte sedimentation rate; PMNs—polymorphonuclear neutrophils; PT—prothrombin time; PTT—partial thromboplastin time; RF—rheumatoid factor; WBCs—white blood cells)**

phases of the disease, when the anti-CCP assay may become positive before the rheumatoid factor test.[12]

SYNOVIAL FLUID ANALYSIS

Examination of synovial fluid is perhaps the most important diagnostic test in rheumatology. It gives the physician one of the few opportunities available for the precise diagnosis of rheumatic disease and permits immediate initiation of specific and effective therapy. In addition, aspiration of synovial fluid is a low-risk procedure: the frequency of iatrogenic infection is less than one in 10,000. Joint aspiration should be performed with aseptic technique as part of the evaluation of every case of acute monoarthritis.

Analysis of the synovial fluid includes a white blood cell count and differential, appropriate cultures and stains for microorganisms, and polarized-light microscopy. The white blood cell count in the synovial fluid is useful in distinguishing inflammatory from noninflammatory arthritis: levels greater than 2,000 cells/mm^3 are consistent with inflammation. Patients with crystal-induced arthritis usually have counts in excess of 30,000 cells/mm^3. The reliability of the examination of fluid for crystals by polarized-light microscopy depends very much on the laboratory doing the test. The finding of monosodium urate or calcium pyrophosphate dihydrate crystals on polarized-light microscopy is pathognomonic for gout and pseudogout, respectively. The absence of crystals does not exclude these diagnoses, nor does their presence exclude the possibility of coexistent infection. Gram stain and culture may be diagnostic of infection. If patients with established arthritis have fever and an apparent flare, joint infection should be excluded by joint aspiration because septic arthritis occurs more frequently in such patients.

IMAGING

The findings on radiography are unlikely to be abnormal in most patients with acute arthritis, mechanical back pain, tendinitis, or bursitis; radiography should not be used in these cases. In patients with acute back pain of less than 6 weeks' duration, imaging is not indicated unless there is a high suspicion of systemic disease or progressive neurologic deficit[13] [*see 15:XII Back Pain and Common Musculoskeletal Problems*]. For patients

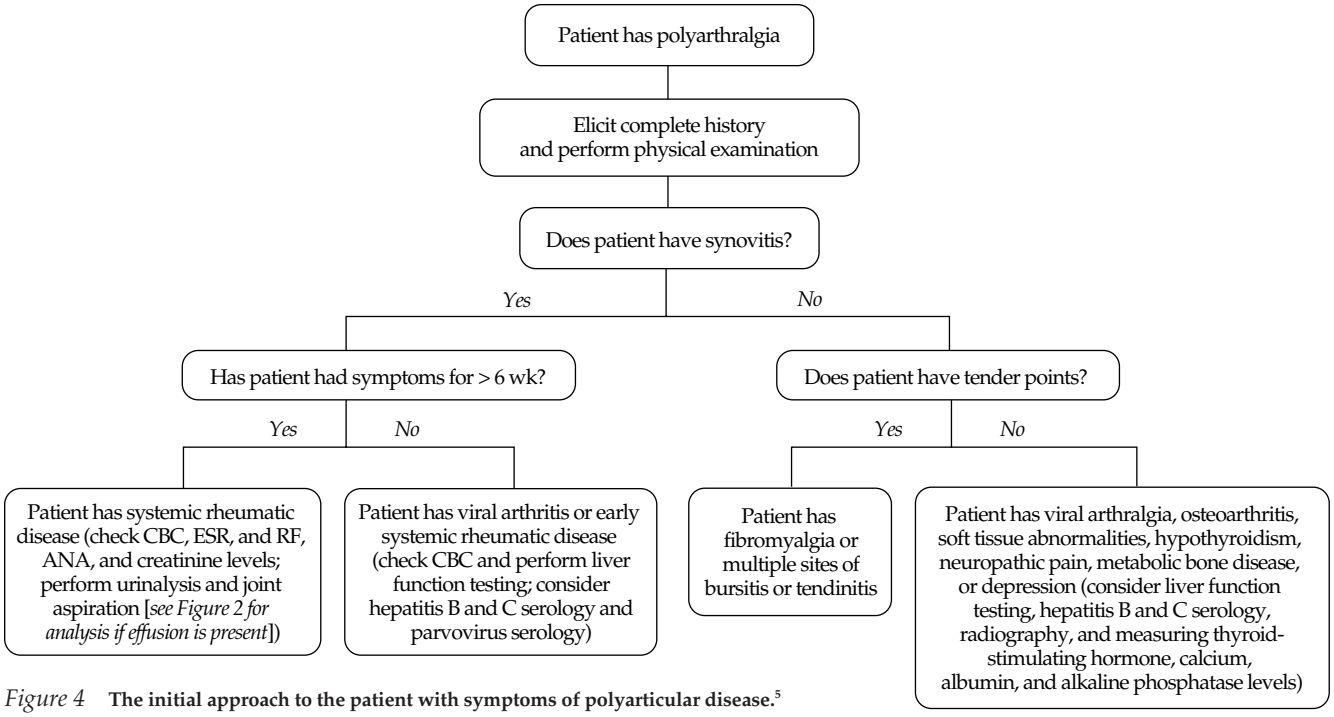


Figure 4 **The initial approach to the patient with symptoms of polyarticular disease.**[5]

with established arthritis, such as rheumatoid arthritis, it takes longer than 6 months for radiographic abnormalities such as joint space narrowing and marginal erosions to appear. In patients with osteoarthritis, radiography may be useful in assessing the extent of joint damage, but the correlation between its findings and patients' symptoms is surprisingly poor.

Plain radiography is most useful in patients with significant trauma that suggests the possibility of fracture, in those who experience a sudden loss of function (e.g., an inability to bear weight), in those with symptoms that do not improve despite appropriate treatment, and in those with a suspected infection or neoplastic disease. Computed radiography and digital imaging improve the quality of plain radiographs, make long-distance transmission easy, and eliminate the problem of the lost radiograph. When it is likely that a fracture is present, special views may be required; repeat imaging 7 to 10 days later may detect callus formation at a previously unrecognized fracture site. In patients with chronic disease, repeat radiography is useful in assessing the progress of disease or the necessity of surgical intervention.

Although computed tomography, magnetic resonance imaging, and radionuclide bone scanning are powerful techniques for obtaining information about bones and joints, they should be used only when their results will be used to make important diagnostic or therapeutic decisions. As is true for plain radiography, myelography, and even autopsy, MRI of the spine often reveals abnormalities in people who do not have back pain. In one study, only 36% of patients free of back pain had normal findings on MRI of the lumbar spine.[14] Bone scans are useful in detecting osteomyelitis, stress fractures, and metastases to bone.

Clinical Presentation and Initial Approach

MONOARTICULAR DISEASE

Patients with symptoms that involve one or, at most, a few joints may have posttraumatic syndromes, bursitis, tendinitis, septic arthritis, crystal-induced inflammation (gout or pseudogout), or an atypical presentation of a systemic arthritis such as rheumatoid arthritis [*see Figure 3*]. Arthralgia, in which pain arises from structures surrounding the joint, must be distinguished from arthritis, in which there is evidence of frank involvement of the joint itself. Monoarticular arthritis should be considered to have an infectious etiology until proved otherwise. Prompt aspiration and examination of synovial fluid are usually indicated.

POLYARTICULAR DISEASE

In contrast to monoarticular disease, which often requires an aggressive and invasive diagnostic approach, arthritis involving numerous joints usually requires a more gradual and expectant strategy. The one exception is septic arthritis, which may have a polyarticular onset in up to 20% of patients. In other diseases, there is less urgency in arriving at the diagnosis, and observation for 6 weeks or more may be required. During this period, it is unwise to alarm the patient unnecessarily by musing about the possibility that he or she has a particular systemic rheumatic disease; a wait-and-see attitude is more appropriate. Eventually, most patients with polyarthralgia will prove not to have a systemic rheumatic disease but to have a more benign condition [*see Figure 4*].

DIFFUSE ACHES AND PAINS

Most patients with diffuse aches and pains have a benign, self-limited illness of unknown cause that improves after 1 or 2 weeks of observation. The history often identifies antecedent viral infection or exertional stress as a precipitating factor. Another substantial number of patients prove to have fibromyalgia, which is suggested by an inability to precisely locate the anatomic origins of the pain on physical examination and by the finding of tender trigger points [*see 15:XIII Fibromyalgia*]. A hemogram and measurements of the ESR or CRP, thyroid-stimulating hormone, creatine kinase (if weakness is an issue), calcium, and phosphate levels are usually sufficient to exclude other diseases [*see Table 5*].

Table 5 Differential Diagnosis of Diffuse Aches and Pains

Benign postviral syndromes
Postexertional syndromes
Fibromyalgia
Polymyalgia rheumatica
Temporal arteritis
Hypothyroidism
Metabolic bone disease
Hypophosphatemia
Atypical onset of systemic rheumatic disease (e.g., rheumatoid arthritis, systemic lupus erythematosus)

Prognosis

Patients who present with complaints of joint pain often express the opinion, "Nothing can be done for my arthritis—I'll just have to learn to live with it." Correcting this misconception is one of the most important things the physician can do during the initial contact. Most kinds of arthritis can be managed quite effectively, and the patient must understand that treatment greatly improves the condition of most patients with arthritis. Even for the most severe diseases, such as rheumatoid arthritis, very effective treatments have been available for 20 years and have brought distinct improvements in long-term outcome.[15,16] New treatments with biologic agents that block the inflammatory cytokines TNF-α and IL-1 are very effective, have a rapid onset of action, and prevent radiographic progression of the disease.[17-19] Educating the patient about the effectiveness of treatment is the first step toward a successful outcome.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Public health and aging: projected prevalence of self-reported arthritis or chronic joint symptoms among persons aged > 65 years—United States, 2005–2030. MMWR Morb Mortal Wkly Rep 52:489, 2003

2. Prevalence of self-reported arthritis or chronic joint symptoms among adults—United States, 2001. MMWR Morb Mortal Wkly Rep 51:948, 2002

3. Schappert SM: Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 1997. Vital Health Stat 13 143:1, 1999

4. Liang MH, Roberts WN, Robb-Nicholson C, et al: Sorting out musculoskeletal complaints. Rheumatology: Problems in Primary Care. Medical Economics Books, Oradell, New Jersey, 1990

5. Guidelines for the initial evaluation of the adult patient with acute musculoskeletal symptoms. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. Arthritis Rheum 39:1, 1996

6. Gabay C, Kushner I: Acute-phase proteins and other systemic responses to inflammation. N Engl J Med 340:448, 1999

7. Sox HC, Liang MH: The erythrocyte sedimentation rate: guidelines for rational use. Ann Intern Med 104:515, 1986

8. Labarrere CA, Zaloga GP: C-reactive protein: from innocent bystander to pivotal mediator of atherosclerosis. Am J Med 117:499, 2004

9. Tan EM, Feltkamp TE, Smolen JS, et al: Range of antinuclear antibodies in "healthy" individuals. Arthritis Rheum 40:1601, 1997

10. Saraux A, Berthelot JM, Chales G, et al: Value of laboratory tests in early prediction of rheumatoid arthritis. Arthritis Rheum 47:155, 2002

11. Bukhari M, Lunt M, Harrison BJ, et al: Rheumatoid factor is the major predictor of increasing severity of radiographic erosions in rheumatoid arthritis: results from the Norfolk Arthritis Register Study, a large inception cohort. Arthritis Rheum 46:906, 2002

12. Morrow DA, Ridker PM: C-reactive protein, inflammation, and coronary risk. Med Clin North Am 84:149, 2000

13. Jarvik JG, Deyo RA: Diagnostic evaluation of low back pain with emphasis on imaging. Ann Intern Med 137:586, 2002

14. Jensen MC, Brant-Zawadzki MN, Obuchowski N, et al: Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 331:69, 1994

15. Fries JF, Williams CA, Morfeld D, et al: Reduction in long-term disability in patients with rheumatoid arthritis by disease-modifying antirheumatic drug-based treatment strategies. Arthritis Rheum 39:616, 1996

16. O'Dell JR, Haire CE, Erikson N, et al: Treatment of rheumatoid arthritis with methotrexate alone, sulfasalazine and hydroxychloroquine, or a combination of all three medications. N Engl J Med 334:1287, 1996

17. Kremer JM: Rational use of new and existing disease-modifying agents in rheumatoid arthritis. Ann Intern Med 134:695, 2001

18. Weinblatt ME, Keystone EC, Furst DE, et al: Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum 48:35, 2003

19. Cohen S, Hurd E, Cush J, et al: Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 46:614, 2002

20. Oliveria SA, Felson DT, Reed JI, et al: Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis Rheum 38:1134, 1995

Acknowledgment

Figures 2 through 4 Marcia Kammerer.

II RHEUMATOID ARTHRITIS

GARY S. FIRESTEIN, M.D.

Rheumatoid arthritis (RA) is the most common chronic inflammatory arthritis and affects about 1% of adults; it is two to three times more prevalent in women than in men. RA may begin as early as infancy, but onset usually occurs in the fifth or sixth decade. There are no specific laboratory tests for RA; diagnosis depends on a constellation of signs and symptoms that can be supported by serology and radiographs. Involvement of the small joints of the hands and feet is often the key to the diagnosis. Specific clinical criteria have evolved [*see Table 1*], but in practice, diagnosis is established by careful observation of the pattern of disease activity over time.

Immunogenetics

Genetic makeup plays a critical role in susceptibility to RA. Identical twins show 30% to 50% concordance for the disease; first-degree relatives of patients with RA have about a twofold to threefold increased incidence. Study of the major histocompatibility complex (MHC) has identified a shared epitope on the β chains of certain HLA-DR haplotypes in RA patients. This susceptibility epitope is associated with the third hypervariable region of DR β chains, which contains amino acids 70 through 74 (glutamine-leucine-arginine-alanine-alanine, also known as QKRAA) found in *DRB1*0401*, *DRB1*0404*, and other immunologically distinct alleles.[1] This sequence is common to most RA patients, although the disease develops in only a small fraction of those with the epitope.

Susceptibility to RA is likely polygenic; for instance, certain immunoglobulin genotypes and, perhaps, genetic differences in the galactosylation of immunoglobulin may be predisposing factors. Further studies have identified associations with microsatellite alleles of cytokines. Tumor necrosis factor (TNF) alleles are in linkage disequilibrium with the *DR β* gene and may be independent risk factors for RA.[2] In addition, a polymorphism of the interleukin-1α (IL-1α) gene is associated with juvenile RA.[3] Studies of IL-10 promoter polymorphisms have been variable,

Table 1 American Rheumatism Association Criteria for the Classification of Rheumatoid Arthritis[96]

Criteria

Morning stiffness: morning stiffness in and around the joints lasting at least 1 hr before maximal improvement

Arthritis of three or more joint areas: at least three joint areas have simultaneously had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician; the 14 possible joint areas (right and left) are proximal interphalangeal (PIP), metacarpophalangeal (MCP), wrist, elbow, knee, ankle, and metatarsophalangeal (MTP) joints

Arthritis of hand joints: at least one joint area swollen as above in wrist, MCP, or PIP joint

Symmetrical arthritis: simultaneous involvement of the same joint areas (as in arthritis of three or more joint areas, above) on both sides of the body (bilateral involvement of PIP, MCP, or MTP joints is acceptable without absolute symmetry)

Rheumatoid nodules: subcutaneous nodules over bony prominences or extensor surfaces or in juxta-articular regions that are observed by a physician

Serum rheumatoid factor: demonstration of abnormal amounts of serum rheumatoid factor by any method that has been positive in fewer than 5% of normal control subjects

Radiographic changes: radiographic changes typical of rheumatoid arthritis on posteroanterior hand and wrist x-rays, which must include erosions or unequivocal bony decalcification localized to or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify)

Exclusions

The presence of any of the following excludes the diagnosis of rheumatoid arthritis:

Typical rash of systemic lupus erythematosus (SLE)

High concentration of lupus erythematosus cells (four or more in two smears); because of the frequent finding of LE cells in patients with clinically typical rheumatoid arthritis, however, it is suggested that such patients be listed separately

Histologic evidence of polyarteritis nodosa with segmented necrosis of arteries associated with nodular leukocytic infiltration extending perivascularly, including many eosinophils

Persistent muscle swelling of dermatomyositis or weakness of neck, trunk, and pharyngeal muscles

Definite scleroderma (not limited to the fingers)

A clinical picture characteristic of rheumatic fever with migratory joint involvement and evidence of endocarditis, especially if accompanied by subcutaneous nodules, erythema marginatum, or chorea (an elevated antistreptolysin titer will not rule out the diagnosis of rheumatoid arthritis)

A clinical picture characteristic of gouty arthritis with acute attacks of swelling, redness, and pain in one or more joints, especially if responsive to colchicine

Tophi

A clinical picture characteristic of acute infectious arthritis of bacterial or viral origin with an acute focus of infection or a close association with a disease of known infectious origin; chills; fever; acute joint involvement, usually initially migratory (especially if organisms are present in the joint fluid or there is a response to antibiotic therapy)

Tubercle bacilli in joints or histologic evidence of joint tuberculosis

A clinical picture characteristic of reactive arthritis with urethritis and conjunctivitis associated with acute joint involvement, usually initially migratory

A clinical picture characteristic of the shoulder-hand syndrome with unilateral involvement of shoulder and hand and diffuse swelling of the hand, followed by atrophy and contractures

A clinical picture characteristic of hypertrophic osteoarthropathy with clubbing of fingers or hypertrophic periostitis, or both, along the shafts of the long bones, especially if an intrapulmonary lesion is present

A clinical picture characteristic of neuroarthropathy with condensation and destruction of bones of involved joints and associated neurologic findings

Homogentisic acid in the urine grossly detectable by alkalinization

Histologic evidence of sarcoid or a positive Kveim test

Multiple myeloma evidenced by marked increase in plasma cells in the bone marrow or by Bence Jones protein in the urine

Characteristic skin lesions of erythema nodosum

Leukemia or lymphoma with characteristic cells in peripheral blood, bone marrow, or tissues

A clinical picture characteristic of ankylosing spondylitis, psoriasis, ulcerative colitis, or regional enteritis

Note: for classification purposes, a patient is said to have rheumatoid arthritis if he or she has satisfied at least four of the above seven criteria. The first four must be present for at least 6 wk. Patients with two clinical diagnoses are not excluded. Designation as classic, definite, or probable rheumatoid arthritis is *not* to be made.

although on balance, there does not appear to be an association between a specific polymorphism and susceptibility to RA.

Etiology

It is unlikely that a single etiologic factor accounts for all cases of adult RA. A pathogenic organism is often assumed to be responsible, but despite some suggestive data, no conclusive evidence implicates bacteria or mycoplasmas. Viruslike particles have been isolated from synovial effusions in RA,[4] and some RA patients exhibit evidence of a recent parvovirus B19 infection.[5] The potential role of parvovirus B19 is controversial. Whereas some studies have not shown any correlation between RA and serologic evidence of previous infection or presence of *B19* genes in synovial tissue, one study demonstrated parvovirus B19 proteins by immunohistochemistry and potentially infectious virus particles in RA synovium.[6] Some data suggest that B19 can infect cultured synovial fibroblasts and increase invasion into cartilage matrix.[7]

Other viruses that have been isolated from synovial fluid include rubella and Epstein-Barr virus (EBV). Sera from most RA patients contain greater amounts of antibodies to various EBV-derived antigens than normal sera. Suppression of EBV infection by lymphocytes from RA patients is impaired, possibly because T cells mount an insufficient response with low levels of interferon gamma. Although EBV infection is probably not the initial event in RA, it may contribute to persistent immunologic stimulation by acting as a polyclonal activator of B cells, thereby augmenting the production of autoantibodies.

Lymphocytes from some RA patients respond to a region of EBV glycoprotein gp110 that contains the same QKRAA sequence as the susceptibility epitope on DR β chains.[8] Thus, molecular mimicry may lead to autoimmunity in certain EBV-infected individuals. Other xenoproteins, most notably *Escherichia coli* DNA J protein, also contain QKRAA and may contribute to a response against self-MHC.[9] Although many patients undergo an autoimmune response to type II collagen, this response is nonspecific and may be of secondary etiologic significance.

Retroviruses could also serve as infectious causes of RA-like diseases. Synovial human T cell lymphotropic virus type I (HTLV-I) infection is associated with chronic arthritis, and in vitro transduction of synoviocytes with the HTLV-I *tax* gene leads to increased growth.[10] Retrovirus-like particles have been observed in some synovial samples, and expression of zinc-finger proteins associated with retroviral infections offers some support.[11]

Some RA patients show evidence of autoimmunity long before the appearance of clinical arthritis. For instance, rheumatoid factors and other autoantibodies (e.g., anticyclic citrullinated peptides) can be detected in the blood of patients many years before the onset of disease.[12] Although autoimmunity occurs in patients with RA, it may not be responsible for the initiation of the disease. An alternative hypothesis ascribes the initiation of disease to the activation of innate immunity in the synovium of susceptible persons.[13] This process, which involves primitive pattern-recognition receptors on macrophages, dendritic cells, mast cells, and neutrophils, leads to nonspecific articular inflammation. A local immune response then occurs as the synovium permits the influx of lymphocytes, which, in the appropriate cytokine milieu, recognize a variety of xenoantigens and autoantigens. In this scenario, no single etiologic agent is required. Instead, nonspecific inflammation in a patient with a particular gene set can lead to local responses directed at many articular antigens.

Pathogenesis

SYNOVIAL HISTOPATHOLOGY AND INVASION

The synovial tissue in RA becomes markedly hyperplastic, with redundant folds, frondlike villi, and edema. In the earliest stages, blood vessel proliferation and endothelial damage are prominent. Hyperplasia of the synovial intimal lining (the region in direct contact with synovial fluid) can occur, although the sublining inflammatory infiltrate can be mild. As the chronic phase begins, intimal lining hyperplasia becomes more prominent, increasing up to fivefold from the normal thickness of one or two cell layers [*see Figure 1*]. Synovial lining hyperplasia is caused, in part, by local proliferation of the fibroblast-like type B synoviocytes and migration of new macrophage-like type A synoviocytes from bone marrow and blood into the joint. The rate of cell death also determines tissue cellularity. Many cells of the intimal lining contain damaged DNA that normally leads to apoptosis (programmed cell death), but relatively few cells complete this process.[14] RA synovial cells possibly have defective apoptosis that contributes to hyperplasia.

In chronic RA, inflammatory cells (including T cells, B cells, macrophages, and plasma cells) accumulate in the sublining region. Lymphocytes can organize into discrete aggregates, although diffuse mononuclear cell infiltration or relatively acellular fibrous tissue can also be present. The majority of T cells are $CD4^+$ memory cells with small nuclei and scant cytoplasm. Although the cells are functionally quiescent, many express surface antigens that suggest previous activation. An increased number of blood vessels remains a prominent finding in the chronic phase. Capillary morphometry studies suggest that the capillary network is more disorganized than normal, and the tissue bulk outstrips the proliferation of blood vessels.

Rheumatoid synovitis is usually accompanied by increased synovial effusions. The white blood cell (WBC) count in synovial fluid in active RA is about 10,000/mm^3 (about 70% neutrophils). In contrast to the synovium, there are more $CD8^+$ T cells than $CD4^+$ T cells in synovial effusions. Total WBC counts sometimes

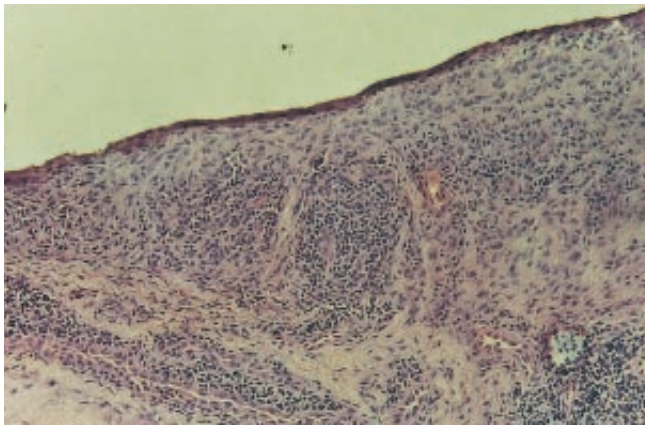


Figure 1 **Section of a proliferative synovium from a patient with classic rheumatoid arthritis reveals synovial lining hyperplasia and a sublining lymphocyte infiltration and aggregation.**

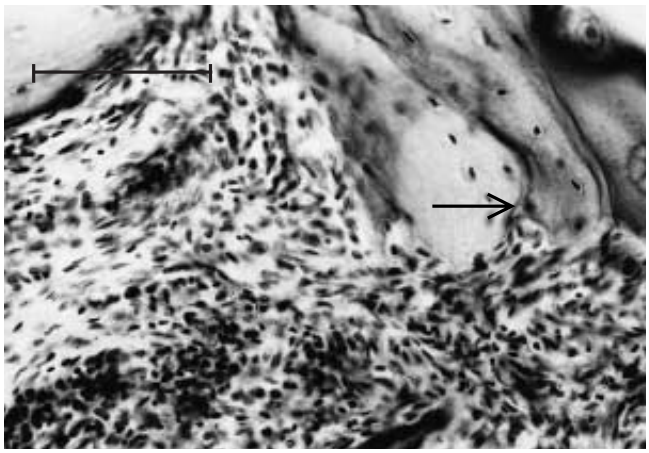


Figure 2 **At the junction between a proliferative inflamed rheumatoid synovium and the bone, scalloped regions of erosion can be seen (arrow). Section is stained with hematoxylin and eosin (bar scale = 100 μm).**

exceed 50,000/mm^3 and include 90% to 95% polymorphonuclear leukocytes. The polymorphonuclear leukocytes are drawn into the joint fluid along a gradient formed by chemotactic substances that include leukotriene B4, platelet-activating factor, the C5a fragment of complement, and chemokines such as IL-8. Lymphocytes, macrophages, and shed lining cells are also seen in synovial fluid. Surprisingly, very few neutrophils are present in RA synovium, even though they are abundant in the effusions.

Pannus, the invasive region of synovium that erodes into cartilage and bone [*see Figure 2*], contains macrophages and primitive mesenchymal cells but very few lymphocytes. It is not clear whether these mesenchymal cells are related to type B synoviocytes, but morphologic and functional studies suggest that pannus-derived fibroblasts (pannocytes) have distinctive characteristics (e.g., very high expression of vascular cell adhesion molecule–1).[15] Mesenchymal stem cells have also been described in RA synovial tissue; these cells express distinct surface proteins (e.g., bone morphogenic protein receptors and endoglin) and can migrate into the synovium directly through pores in the bone or through the circulating blood.[16]

Damage to bone and cartilage by synovial tissue and pannus is mediated by several families of enzymes, including serine proteases and cathepsins. The most damaging enzymes are the metalloproteinases (e.g., collagenase, stromelysin, and gelatinase) and cathepsins (especially cathepsin K), which can degrade the major structural proteins in the joint. Cytokines such as IL-1 and TNF-α are potent inducers of metalloproteinase gene expression. Although protease inhibitors, like tissue inhibitors of metalloproteinases, are expressed by the rheumatoid synovial lining, the balance between proteases and inhibitors appears to favor the former in RA.[17] Chondrocytes in the cartilage, synoviocytes in pannus, and osteoclasts in the bone are the primary sources of proteases. The receptor activator of nuclear factor κB (RANK) and the RANK ligand (RANKL) together play a critical role in local osteoclast activation and bone destruction; the RANKL/RANK system is counterbalanced by the natural inhibitor osteoprotegerin (OPG). In animal models of arthritis, administration of OPG markedly decreases bone destruction, even though inflammation is unaffected.[18]

Destruction of extracellular matrix by rheumatoid synovium mesenchymal cells may occur either as a result of a normal response to the inflammatory cytokine milieu or as a result of abnormal synoviocyte function.[19] Evidence of partial transformation of RA synoviocytes, including adhesion-independent growth and loss of contact inhibition in vitro, suggests that immunosuppression may slow but not necessarily halt joint destruction. Cultured RA synoviocytes that have been coimplanted with cartilage explants into mice with severe combined immunodeficiency disease invade the cartilage matrix, whereas osteoarthritis synoviocytes and normal dermal fibroblasts do not.[20] Somatic mutations in the genes encoding key regulatory proteins, such as the *p53* tumor-suppressor gene, may contribute to the transformed phenotype of synoviocytes.[21] Such mutations are likely caused by the high local concentration of oxidants in the rheumatoid joint. Hence, the invasive component of rheumatoid synovitis potentially functions as an autonomous cytokine-independent tissue that erodes into cartilage.

CELLULAR IMMUNITY

Attempts to identify an etiologic agent by determining the proliferative response of synovial T cells to specific antigens have been relatively unrewarding. Articular T cells are often less responsive than peripheral blood cells. For instance, the proliferation of lymphocytes in synovial fluid in response to mitogens or recall antigens (e.g., tetanus toxoid) is significantly lower than the proliferation of blood T cells. Production of cytokines (e.g., interferon gamma and IL-2) by synovial fluid T cells in vitro is also low after stimulation by nonspecific mitogens. The mechanism of defective T cell responses in RA appears to be related to abnormal intracellular redox balance, which interferes with transduction of the T cell receptor signal.[22] Mycobacterial antigens and the 60 kd heat shock protein appear to be exceptions in that lymphocyte proliferation in response to these antigens is greater in cells from rheumatoid effusions than in blood cells. However, this response is not specific to RA and is even more prominent in reactive arthritis.

Immune dysregulation has been observed in peripheral blood T cells in patients with RA, especially with EBV infection. The deficient T cell response can be correlated with disease activity, but it also occurs in patients with other forms of arthritis. A more specific defect is observed in the autologous mixed lymphocyte reaction, in which T cells proliferate and produce cytokines in response to MHC class II antigens expressed on autologous antigen-presenting cells. Autoimmune responses directed toward joint-specific antigens can contribute to synovitis. In addition to type II collagen, which is localized to hyaline cartilage, other articular antigens have been implicated. For instance, T cell immunity directed against heat shock proteins, cartilage protein gp39, cartilage link protein, and proteoglycans have been variably implicated in RA. Many of these antigens can induce arthritis in mice or rats when the animals are immunized with the antigen in combination with complete Freund adjuvant. An unusual T cell phenotype ($CD4^+$, $CD28^-$) has been noted in the synovial tissue of patients with RA that might possess functions of both innate and adaptive immunity.[23]

HUMORAL IMMUNITY

Rheumatoid Factors and Other Autoantibodies

Rheumatoid factors are immunoglobulins with antibody specificity for the Fc region of IgG. The tests usually employed in clinical diagnosis (latex fixation, sensitized sheep red blood cell agglutination, nephelometry, and enzyme-linked immunosor-

bent assay) detect only IgM rheumatoid factors. The tests are positive in up to 90% of patients with classic RA, depending on the method used. Although patients with classic RA may have negative test results, a high-titer positive result indicates a poor prognosis—an unremitting course and a greater degree of joint damage.

Rheumatoid factor is not a specific finding for RA. Significant titers are found in patients with related diseases (e.g., systemic lupus erythematosus [SLE], progressive systemic sclerosis, and dermatomyositis) and in patients with nonrheumatic chronic inflammatory disorders and infections. Healthy elderly persons, particularly women, often have positive test results. Rheumatoid factor may be a feature of the early immune response to many proteins, facilitating antigen clearance by macrophages.

IgM rheumatoid factor is most commonly detected; IgG and, less frequently, IgA rheumatoid factors are also sometimes found. The presence of IgG rheumatoid factor is associated with a higher rate of systemic complications (e.g., necrotizing vasculitis).[24] Rheumatoid factors in RA may result from somatic mutation in response to an antigen-driven immune response.[25] Rheumatoid factors can be synthesized by B cells and plasma cells that infiltrate the synovium in RA patients, including some seronegative patients.

Another autoantibody system has been described that could have considerable clinical utility. Antibodies directed against keratin or to the filament-aggregating protein filaggrin were defined in patients with RA. Their specificity is determined by their ability to bind to a modified arginine residue known as citrulline. Many proteins are citrullinated in synovial tissue, and citrullination occurs in animal models of arthritis.[26] The enzyme responsible for altering this amino acid is peptidylarginine deiminase (PAD); and certain isoforms of PAD, such as PAD4, are overexpressed in inflamed joint tissues. Certain haplotypes of the enzyme PAD4 appear to confer an increased risk of developing RA. Although citrullinated peptides are commonly produced at sites of inflammation, antibodies directed against PAD epitopes appear to be specific for RA. B cells that produce these IgG and IgM anticitrulline antibodies are present in rheumatoid synovial tissue. Anticitrulline antibodies (often assayed using cyclic citrullinated peptide [CCP]) are as prevalent as rheumatoid factor and may be more specific for RA. High titers of anti-CCP antibodies also correlate with more aggressive and destructive disease.

Other autoantibodies also have a role in RA, including antibodies directed at joint-specific antigens such as gp39, RA33, and p205.[27] Antibodies to glucose-6-phosphate isomerase (GPI), a ubiquitous antigen, can cause arthritis in mice and have also been detected in patients with RA and other inflammatory arthropathies.[28] Anti-GPI antibodies appear to localize in the joints and activate complement, perhaps because GPI can adhere to articular cartilage. The relative contribution of autoantibodies to RA as either a primary or a secondary phenomenon is still uncertain. Anti-GPI antibodies are present in only a small percentage of RA patients; they are also detected in the blood of patients with other forms of inflammatory arthritis.

Complement Activation

Interaction of rheumatoid factors with normal IgG activates complement and thereby starts a chain of events that includes production of anaphylatoxins and chemotactic factors. Polymorphonuclear leukocytes then engulf the rheumatoid factor–IgG-complement complexes and release lysosomal enzymes and other products. Complexes of IgG rheumatoid factor with IgG and complement components are readily detected in the synovium, synovial fluid, and extra-articular lesions. Although the synovium is a rich source of complement production, the levels in rheumatoid synovial fluid are low because of local consumption. Deposits of immunoglobulin and complement have been identified in avascular cartilage and other collagenous tissues of rheumatoid joints and may play a role in the formation of the destructive lesion of RA. These deposits, which are highly specific for RA, may be an attractant for the invasive pannus.

Cytokines

Early studies suggested an unrestricted abundance of cytokines in the rheumatoid joint. However, later experiments demonstrated a relative paucity of many T cell–derived cytokines, including IL-2, IL-4, and TNF-β.[29] One exception is IL-17, which can regulate cartilage metabolism and may be produced by $CD4^+$ T cells in the joint.[30] T cells can also potentially contribute to macrophage and synoviocyte activation by inducing metalloproteinase gene expression via direct cell-cell contact.

T helper cells can be divided into subsets that mediate distinct functions of the immune system. T helper type 1 (Th1) cells produce interferon gamma and IL-2 but not IL-4, IL-5, or IL-10; T helper type 2 (Th2) cells produce the opposite cytokine profile. Th1 overactivity predominates in most animal models of autoimmunity, whereas Th2 cytokines mediate disease suppression.[31] The small amounts of T cell cytokines that can be detected in RA are biased toward the Th1 phenotype, including IL-17. In contrast, Th2 cytokines (especially IL-4) are virtually absent from the joint. Some IL-10 is present but is derived mainly from macrophages, and the amount is not sufficient to suppress Th1 cytokine production.[32] The relative lack of suppressive Th2 cytokines may contribute to the pathogenesis of rheumatoid synovitis. Levels of other suppressive cytokines, such as the natural IL-1 receptor antagonist (IL-1ra), are also low in RA joint tissues.[33]

Macrophage- and fibroblast-derived cytokines (e.g., IL-1, IL-6, TNF-α, and granulocyte-macrophage colony-stimulating factor [GM-CSF]) are abundantly expressed in the rheumatoid joint.[34] Although many of these cytokines are involved in the pathogenesis of RA, TNF-α and IL-1 are major pathogenic factors: both can induce synoviocyte proliferation, collagenase production, and prostaglandin release; overexpression of either TNF-α or IL-1 can induce arthritis in animal models. IL-15 is produced by macrophages but shares many activities of the T cell–derived cytokine IL-2. It increases the ability of T cells to induce TNF-α production by macrophages through an antigen-independent mechanism that involves cell-cell contact.[35] IL-18 is also present in the RA joint and can bias T cell responses toward Th1 or directly activate macrophages to produce proinflammatory mediators.[36] Cytokine networks can potentially establish paracrine or autocrine networks that can perpetuate arthritis long after the etiologic agent has been cleared. Recent studies suggest that anti-cytokine therapy (including therapy with IL-1, TNF-α, and IL-6) is effective in severe RA and demonstrates the importance of fibroblast and macrophage products in chronic synovitis.

Diagnosis

CLINICAL FEATURES AND DISEASE COURSE

The onset of RA in adults may be either acute or insidious. In the latter case, systemic manifestations may precede overt symp-

toms of arthritis by months. In some patients, external events (e.g., major infections, surgical procedures, trauma, or childbirth) precede the clinical onset. How these events relate to pathogenesis is unknown. Small joints of the hands and feet are usually involved at the outset, although large joints (e.g., knees and ankles) are sometimes affected first. In about 10% of cases, monoarthritis of a large joint can presage progression to polyarticular RA.

An insidious onset followed by progression to polyarticular involvement is the most common course. Most patients experience some degree of joint stiffness, especially in the morning after awakening, which may accompany or precede joint swelling or pain. These symptoms are hallmarks of disease activity and help distinguish RA from noninflammatory diseases such as osteoarthritis. However, joint stiffness and swelling are not specific for RA and can occur with other types of inflammatory arthritis. RA patients frequently complain of morning stiffness that lasts more than 30 minutes (often up to several hours).

Examination of the joints reveals varying degrees of swelling, warmth over the involved joint, tenderness to palpation, and limitation of active and passive range of motion. Swelling may be caused by thickening, edema, and increased vascularity of the synovium; by synovial effusions; or by combinations of these factors. In small joints, such as metacarpophalangeal joints, effusions may be difficult to detect: the presence of synovial thickening causes loss of the anatomic landmarks and can obscure the peaks and valleys formed by the joints. In large joints, especially the knees, effusions are usually easy to demonstrate. Unlike acute inflammatory arthritides (e.g., gout or septic arthritis), RA tends not to cause marked erythema, and swelling usually does not extend far beyond the articulation. In elderly patients, the most prominent manifestation may be diffuse swelling of the hands accompanied by aching and marked stiffness in the absence of erythema. This can be difficult to distinguish from polymyalgia rheumatica, especially in patients lacking rheumatoid factor.

Classically, RA is symmetrical. When RA is progressive and unremitting, nearly every peripheral joint may eventually be affected, although the thoracic, lumbar, and sacral spine are usually spared. This clinical presentation is observed in perhaps 10% of patients. In about 75%, the disease waxes and wanes over a period of years. In the remaining patients, complete remissions may be achieved with no evidence of inflammation. Remissions may be only partial, with mild clinical disease persisting despite clear improvement. When the course is progressive, the periods of remission may become shorter, and less impressive decreases in symptoms and findings may occur.

A relatively favorable course with long remissions tends to be associated with age less than 40 years, acute onset restricted to a few large joints, disease duration less than 1 year, and negative test results for rheumatoid factor. Conversely, an unfavorable prognosis is often associated with insidious onset, constitutional symptoms (e.g., weight loss, low-grade fever, and profound fatigue), rapid appearance of rheumatoid nodules, and high titers of rheumatoid factor. Homozygosity for the QKRAA sequence in the HLA-DR locus is also associated with more severe disease with extra-articular manifestations. The duration and intensity of inflammation correlates with long-term disability, and there is a significant relationship between persistent elevations in the level of C-reactive protein and poor outcome. The appearance of bone erosions early in the course of disease also portends a worse prognosis.

Pregnancy often relieves the symptoms of RA in the second or third trimester through a poorly clarified mechanism. One possible explanation is that the placenta produces large amounts of the suppressive cytokine IL-10. The risk of developing RA appears to be lower in women who have been pregnant. The effect of oral contraceptives on disease susceptibility is controversial; the effect, if any, is probably small.[37] In long-term studies, multiple pregnancies or the use of oral contraceptives did not significantly alter the course of RA.[38]

Mortality is higher in RA patients than in the normal population. For the most part, RA patients die of the same causes as the general public, albeit earlier. In severe RA, mortality can approach that of severe congestive heart failure or Hodgkin disease, thereby justifying aggressive early management. Cardiovascular disease accounts for about 40% to 45% of deaths in RA patients; cancer, about 15%; and infection, about 10%. The inflammatory response, especially when associated with an increase in C-reactive protein and the use of proatherogenic treatments such as corticosteroids, correlates with an increased incidence of coronary artery disease.[39] The incidence of lymphoproliferative diseases is increased in patients with RA; non-Hodgkin lymphoma, leukemia, multiple myeloma, and Hodgkin disease account for most excess malignancies.

Specific Joint Disease

Hands and wrists Involvement of the hands and wrists is the most characteristic finding. Swelling and tenderness are usually noted first at the metacarpophalangeal and proximal interphalangeal joints [*see Figure 3*]. Fusiform swelling at the proximal interphalangeal joints is typical. Distal interphalangeal joints are usually spared. Grip strength is decreased because of pain and mechanical derangement. Flexor tenosynovitis is common; progressive flexion limitation prevents the making of a fist.

Depending on the site and severity of the rheumatoid lesions, varying degrees of ulnar deviation and subluxation at the metacarpophalangeal joints result. These deformities are, in large part, caused by inflammation and radial deviation at the wrist. As the wrist abnormalities progress, the extensor tendons apply torque across the metacarpophalangeal joints and tend to pull the digits into the classic ulnar deviation position. Other changes in the phalanges include (1) hyperextension at the prox-

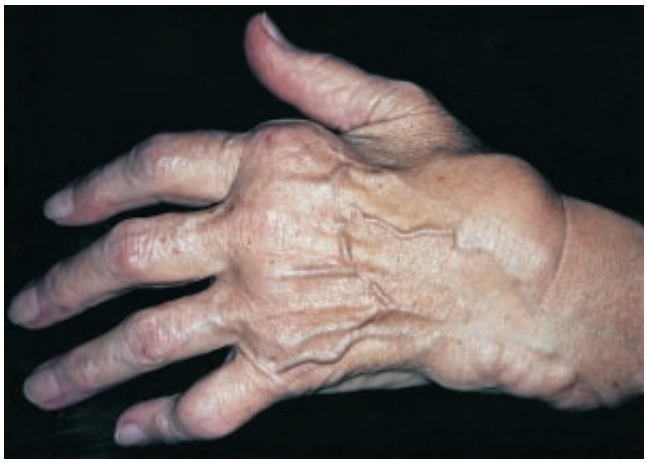


Figure 3 **The hand and wrist are common sites of synovitis in rheumatoid arthritis. Marked swelling in the wrist and metacarpophalangeal joints is caused by synovial proliferation. Modest ulnar deviation of the fingers is also present.**

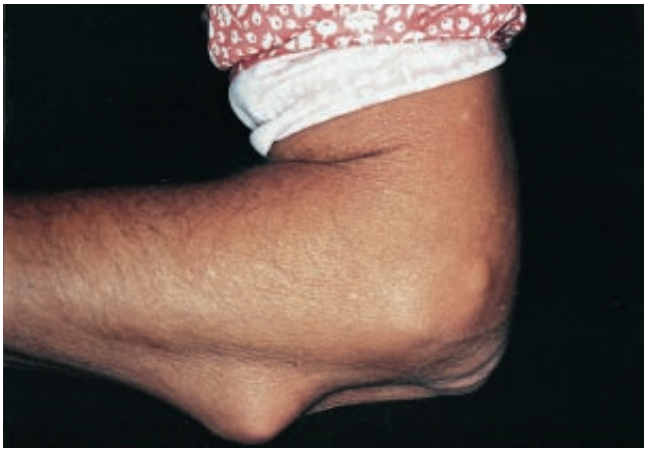


Figure 4 **Rheumatoid nodules commonly form near the extensor surface of the elbow. They can be fixed to the underlying periosteum or can be freely mobile.**

imal interphalangeal joint and flexion at the distal interphalangeal joint (so-called swan-neck deformity) and (2) flexion at the proximal interphalangeal joint and extension at the distal interphalangeal joint (boutonnière deformity). Several deformities also affect the thumb and interfere with grasp and pinch. In extreme instances, the fingers are markedly deformed and flail as a result of destruction of cartilage and bone.

In early RA, relatively painless swelling of the dorsum of the wrist may be noted. Most often, the wrist is painful and is the source of functional limitations (e.g., inability to remove the lid from a jar). At the volar aspect, median nerve compression caused by synovial expansion can produce carpal tunnel syndrome. On the dorsal surface, synovial proliferation may erode and rupture the extensor tendons of the fingers, rendering the patient unable to extend the fingers actively at the metacarpophalangeal joints. Decreased dorsiflexion and plantar flexion of the wrist caused by fusion of carpal bones is common in severe disease. Volar subluxation and radial deviation are also common deformities; the ulnar styloid is often one of the first sites of bone erosion.

Elbows and shoulders Synovitis of the elbow joint and inflammation and nodules in the olecranon bursa are frequent in established RA [*see Figure 4*]. Mild flexion contractures occur early; late in the disease, more severe flexion contractures cause functional disability, especially when associated with decreased shoulder abduction and rotation. Pain with decreased range of motion is commonly caused by synovitis of the glenohumeral joint; occasionally, large anterior effusions are evident. Shoulder pain commonly causes difficulty sleeping at night and functional disability. In chronic RA, the joint space becomes contracted, and rupture of the rotator cuff is very common. On physical examination, true glenohumeral joint arthritis can usually be distinguished from acromioclavicular pain, rotator cuff tendinitis, and subdeltoid bursitis.

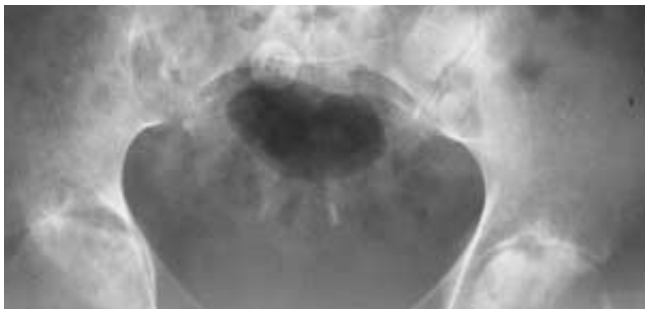
Hips The hip is affected later than most other joints. In osteoarthritis, the femoral head tends to migrate superiorly in the acetabulum, but in RA, symmetrical destruction of cartilage leads to axial migration. End-stage rheumatoid disease with typical cartilage loss produces acetabular protrusion of the femoral head [*see Figure 5*].

Knees Knee arthritis is common and is occasionally a primary manifestation in early RA. Swelling and thickening of the synovium and effusions are usually simple to detect; arthrocentesis readily provides synovial fluid for analysis. Occasionally, large effusions expand into the suprapatellar pouch. Atrophy of muscles around the knee, especially the quadriceps, and resultant weakness can be detected early. Persistent synovitis eventually limits walking because of cartilage destruction, ligament laxity, joint instability, and contractures.

Baker cysts of the popliteal space are lined with synovial membrane and usually communicate with the cavity of the knee joint. The high pressure generated during knee flexion may be propagated posteriorly and cause rupture or dissection of these cysts. Calf swelling, pain, and erythema result, mimicking thrombophlebitis. Rupture of cysts is not specific to RA, occurring in other forms of inflammatory synovitis as well. Diagnosis of popliteal cysts can be confirmed by ultrasonography or arthrography. Generally, treatment of the cyst is directed toward the underlying knee synovitis. Corticosteroid injections are usually directed into the knee rather than into the cyst.

Ankles and feet Inflammation of the ankle joints and of the small joints of the feet is common. Pain on flexion and extension is a result of tibiotalar arthritis, whereas pain on inversion and eversion is caused by subtalar disease. The metatarsophalangeal joints are sites of early synovitis, which causes pain in the ball of

a



b

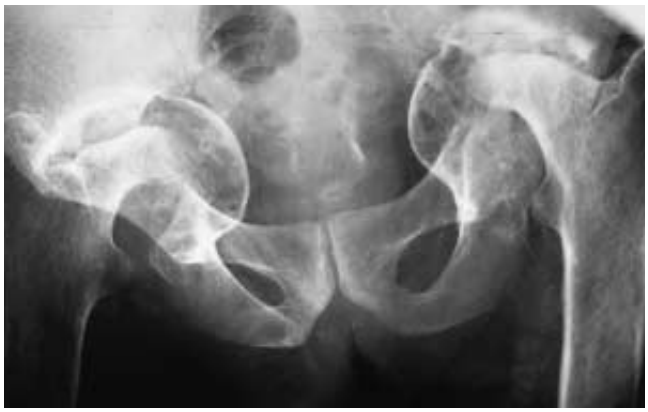


Figure 5 **(*a*) A pelvic roentgenogram of a patient with classic seropositive rheumatoid arthritis was taken early in the course of the disease. (*b*) Another roentgenogram taken 4 years later demonstrates marked acetabular protrusion and resorption of the femoral heads, both of which are characteristic of the disease.**

the foot on weight bearing [*see Figure 6*]. Later in the disease, there is subluxation with protrusion of the metatarsal heads, hallux valgus, and collapse of the arch.

Cervical spine Joints of the thoracic, lumbar, and sacral spine are relatively unaffected in adult RA, but cervical spine disease is frequent and may result in severe pain or neurologic complications.[40] The lesion that has received the most attention is atlantoaxial subluxation and consequent separation at the atlantoodontoid articulation [*see Figure 7*]. This deformity is best seen on lateral roentgenograms obtained with the neck flexed, so that the separation of the anterior margin of the odontoid process from the posterior margin of the anterior arch of the atlas can exceed 3 mm. When the separation is severe, the odontoid process may protrude into the foramen magnum and exert pressure on the spinal cord, causing paresthesia or even muscle weakness in the arms and hands. Often, the odontoid process itself is eroded, which minimizes pressure complications but produces instability. Prophylactic surgery to correct subluxation is usually not recommended because of the high morbidity and mortality associated with the procedure. Surgical fixation is indicated in the presence of neurologic signs and symptoms related to spinal cord compression. If the patient requires other surgical procedures, the anesthesiologist should be alerted to the presence of atlantoaxial subluxation to minimize complications of intubation.

Other cervical spine lesions are also seen, including subluxation at multiple levels, erosions at end plates or apophyseal joints, or fusion at these joints. Management of cervical spine pain in the absence of significant subluxation can be frustrating. Traction can be gently applied, but one must always be cognizant of instability. Soft collars can provide some temporary relief, but if used excessively, they can exacerbate the problem by weakening the cervical muscles.

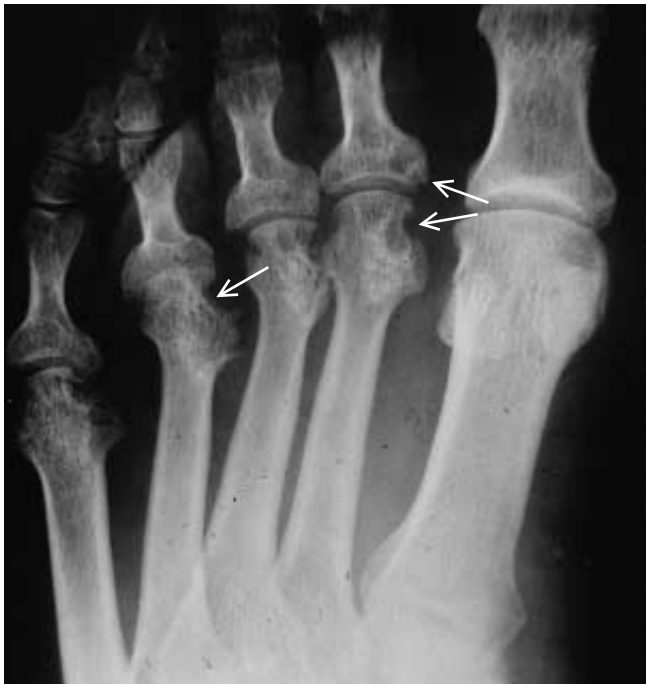


Figure 6 **Erosions (arrows) are visible in the metatarsal heads and in some of the phalanges in this roentgenogram of the foot of a patient with classic seropositive rheumatoid arthritis.**

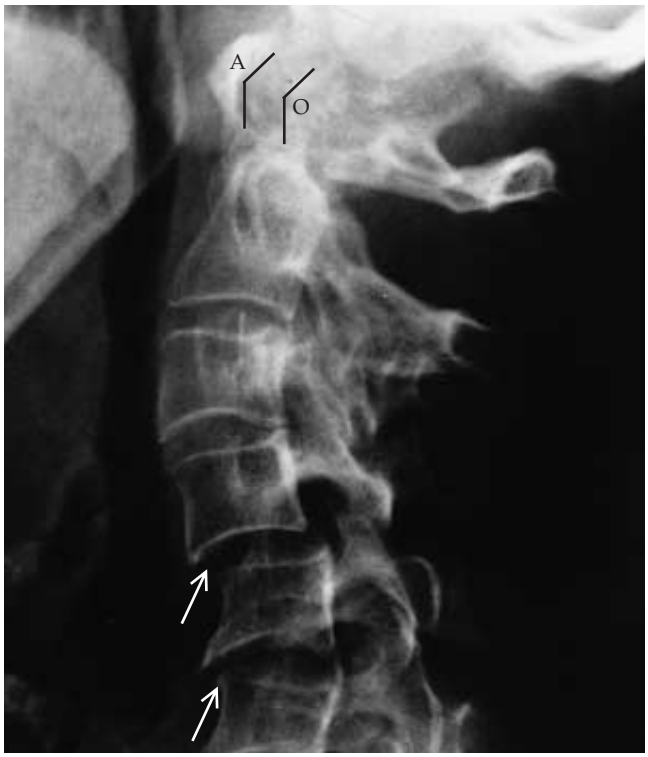


Figure 7 **The anterior edge of the odontoid process (O) is abnormally separated from the posterior margin of the arch of the atlas (A) in this lateral roentgenogram of the cervical spine of a patient with rheumatoid arthritis. Subluxations of the lower cervical vertebral bodies (arrows) are also visible.**

Other joints Synovitis of the temporomandibular joints may produce pain on chewing and limit jaw motion. If the joint is sufficiently destroyed, posterior subluxation of the jaw may cause a receding chin. Sternoclavicular arthritis is uncommon but occurs in patients with widespread arthritis. In acute cricoarytenoid arthritis, hoarseness and pain on swallowing may accompany tenderness over the larynx.

Extra-articular Manifestations

RA is a systemic disease, even though it characteristically affects structures in and around the joints. Its systemic manifestations include mild fever, anorexia, weight loss, fatigue, and muscular weakness. Specific organ involvement usually occurs in the context of severe RA, with high titers of rheumatoid factor and nodule formation.

Rheumatoid nodules Rheumatoid nodules are the most common extra-articular manifestation, occurring in about 15% of patients. Almost all patients in whom nodules develop are seropositive for rheumatoid factor and have erosive disease [*see Figure 8*]. Nodules are usually subcutaneous and often are found in areas exposed to pressure—for example, over the extensor surfaces of the forearm, the olecranon bursa, the knuckles, the ischial regions, the Achilles tendon, and the bridge of the nose (if glasses are worn). They also occur in viscera. Rheumatoid nodules are firm and are either freely movable or attached to connective tissue (e.g., periosteum or tendons). They range from a few millimeters to more than 2 cm in diameter and often occur in clusters. Nodules typically have a rubbery or gritty feel and can

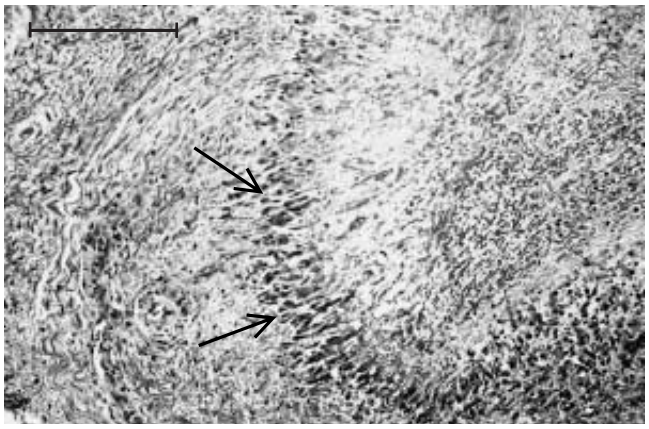


Figure 8 **A typical rheumatoid nodule contains an area of fibrinoid necrosis (center) surrounded by palisading histiocytes (arrows). At the periphery are round cells (predominantly lymphocytes). Stain is hematoxylin and eosin (bar scale = 300 μm).**

be indistinguishable from gouty tophi on physical examination. The lesion contains a center of fibrinoid necrosis (a mixture of fibrin and other proteins, such as degraded collagen) surrounded by a zone of histiocytes, which tend to be arranged radially. Lymphocytes and plasma cells form an outer layer. The pathogenesis of nodules is likely similar to that of synovitis, with early vascular involvement and local cytokine production.

There is no specific therapy for nodules other than treatment of the underlying arthritis. Surgical removal is often ineffective because nodules can return; it is generally reserved for severe functional impairment or obvious cosmetic problems. The appearance of fresh crops of nodules can indicate active disease. In some cases, exuberant nodule production (rheumatoid nodulosis) is a complication of methotrexate therapy. Rheumatoid nodules are not specific to RA, occurring in other connective tissue diseases (e.g., SLE) or in isolation (e.g., granuloma annulare).

Eyes The sicca syndrome, which is part of Sjögren syndrome, is the most frequent ocular manifestation of RA. Symptoms include sensations of grittiness, accumulation of dried mucoid material (especially in the morning immediately after waking up), and decreased tear production. The relative paucity of tears is demonstrated by decreased wetting of a filter paper strip in a Schirmer test. The dryness is not limited to the eyes and involves other exocrine glands, including those in the nose, the mouth, the rectum, and the vagina. Marked enlargement of lacrimal and salivary glands can occur in severe cases, although this is more common in primary Sjögren syndrome than in RA.

The genetic basis of RA with dry eyes is different from that of primary Sjögren syndrome, which is associated with HLA-DR3 antigen rather than HLA-DR4 antigen.[41] In patients who have the primary syndrome without joint inflammation, there is a greater frequency of recurrent parotitis, Raynaud phenomenon, purpura, lymphadenopathy, myositis, and renal involvement. In all patients with sicca syndrome, lacrimal and salivary glands are characterized histologically by lymphocyte infiltration and distortion of ductal structures. Patients with Sjögren syndrome often have high titers of anti-Ro antibody (also called anti–SS-A). Biopsy of minor salivary glands in the lip can help establish the diagnosis.

Scleritis is painful and may lead to perforation of the sclera and blindness. Episcleritis is common and can often be managed with topical corticosteroids. Matrix loss around the limbus and corneal melting may also lead to perforation. Uveitis and iritis occur no more often in adults with RA than in control populations.

Lungs The most common form of lung involvement in RA is pleurisy with effusions.[42] Evidence of pleuritis is often found at postmortem examination, but symptomatic pleurisy occurs in fewer than 10% of patients. Clinical features include gradual onset and variable degrees of pain and dyspnea. The effusions generally have protein concentrations greater than 3 to 4 g/dl, as well as glucose concentrations lower than 30 mg/dl; the latter finding has been ascribed to a primary defect in glucose transport. The leukocyte count is rarely higher than 5,000/mm^3 and is dominated by lymphocytes. The lactate dehydrogenase level is often markedly elevated; occasionally, the lipid content is also high. Complement levels are usually low, and rheumatoid factors are present. Pleural biopsy usually reveals nonspecific fibrosis or granulomas. The pleural effusions usually resolve spontaneously within months. Occasionally, repeated aspirations are required to relieve dyspnea; if effusions are troublesome, instillations of glucocorticoids are useful.

Rheumatoid nodules occur in the pulmonary parenchyma and on the pleural surface. They range in size from just detectable to several centimeters in diameter. They may be single or multiple, and at times, they cavitate. Such nodules can be difficult to distinguish radiologically from tuberculous or malignant lesions and often require further evaluation, including biopsy.

Progressive, symptomatic interstitial pulmonary fibrosis that produces coughing and dyspnea in conjunction with radiographic changes of a diffuse reticular pattern (i.e., honeycomb lung) is usually associated with high titers of rheumatoid factor. The lesion is histologically indistinguishable from idiopathic pulmonary fibrosis. Chest radiographs show pleural thickening, nodules, diffuse or patchy infiltrates, and a restrictive ventilatory defect that is characterized by a decreased CO diffusion rate. These abnormalities are often associated with cigarette smoking, other extra-articular manifestations, and active disease. Bronchiolitis obliterans, an unusual form of airway obstruction that usually has a viral or toxic etiology, may also develop.

Heart Cardiac involvement in RA is common but rarely symptomatic. Echocardiographic evidence of pericardial effusion or thickening has been found in about one third of patients studied.[43] Autopsy findings include rheumatoid nodules, healed or active pericarditis, myocarditis, endocarditis, and valvular fibrosis.

Symptomatic pericarditis is most frequent in patients with severe seropositive disease. Overt manifestations include chest pain, friction rub, and associated pleural effusions. The pericardial effusions resemble the pleural effusions in RA. Cardiac tamponade is rare, as is constrictive pericarditis.

Rheumatoid nodules and inflammation in the valves and the conduction system may cause conduction disturbances, including complete heart block. Aortic regurgitation secondary to aortitis and dilation of the aortic root may lead to congestive heart failure.

Blood Mild anemia of chronic disease is characteristic of active RA, although the hemoglobin level is usually greater than 10 mg/dl. Nonsteroidal anti-inflammatory drugs (NSAIDs) often cause GI blood loss, leading to iron deficiency. Although its lev-

els are not reduced in RA, administration of erythropoietin alleviates anemia.[44]

The constellation of RA with splenomegaly and leukopenia is known as Felty syndrome.[45] The mean serum leukocyte count in such patients is usually 1,500 to 2,000/mm^3, and the mean granulocyte count is 500 to 1,000/mm^3. Severe thrombocytopenia is uncommon. Infections, particularly of the skin, the perianal region, and the lungs, are frequent and are usually caused by common organisms. Other findings in Felty syndrome include hepatomegaly, lymphadenopathy, and chronic cutaneous ulcerations.

The neutropenia is the result of excessive vascular margination of leukocytes, increased peripheral destruction of leukocytes caused by IgG and IgM antigranulocyte antibodies,[46] and the inhibitory effects of T cells on granulopoiesis. Some cases of Felty syndrome are associated with oligoclonal or monoclonal expansion of large granular lymphocytes in the blood and represent a form of chronic leukemia.[47] Splenectomy usually produces an increase in the leukocyte count, but this increase is sustained in only 30% of patients. Lithium chloride may alleviate the neutropenia, as may treatment of active arthritis with disease-modifying drugs such as methotrexate.[48] Treatment with recombinant colony-stimulating factors, such as granulocyte colony-stimulating factor (G-CSF), can increase peripheral granulocyte counts in patients with Felty syndrome. The drug must be given for extended periods because discontinuance leads to relapse.

Neuromuscular involvement Weakness of muscles adjacent to joints with active synovitis is common. The most common neuropathy is median nerve compression caused by synovitis of the wrist. Entrapment of the ulnar nerve at the elbow or branches of the sural nerve in the tarsal tunnel also occurs.

Mononeuritis multiplex is seen in patients who have severe disease with necrotizing vasculitis and, frequently, deposits of immune complexes in the walls of the blood vessels supplying the involved nerves.[49] In milder cases, only segmental demyelination without vascular abnormalities may be found. Aseptic meningitis resulting from a hypersensitivity reaction to NSAIDs has been documented.[50]

Blood vessels Vasculitis in small synovial vessels is a hallmark of early RA, but more widespread vascular inflammation of medium-sized muscular arteries also occurs in older men with advanced disease, rheumatoid nodules, and high titers of rheumatoid factor. The involvement of larger vessels is distinct from small vessel disease; such involvement includes leukocytoclastic vasculitis or nail-fold infarcts. The course and prognosis of systemic rheumatoid vasculitis are similar to those of polyarteritis nodosa. Clinically, patients with rheumatoid vasculitis demonstrate polyneuropathy (mononeuritis multiplex), skin ulcerations, purpura, and cutaneous infarctions (sometimes progressing to gangrene). Manifestations of visceral ischemia, including bowel perforations, myocardial infarctions, and cerebral infarctions, are also common. Treatment usually requires high-dose corticosteroids or cyclophosphamides, or both, and still may not be effective. This feared complication has become quite rare, perhaps because of improved therapy for the underlying disease.

Other systems Apart from gastric and duodenal lesions caused by NSAIDs, GI complications are rare in RA. Rheumatoid nodules may involve the pharynx and esophagus. Mild elevations of liver enzymes are common and are usually drug related. Other hepatic abnormalities, particularly elevations in serum alkaline phosphatase and 5′-nucleotidase levels, occur in Sjögren syndrome. These hepatobiliary lesions are ascribed to immune responses to cross-reacting salivary and biliary antigens. Hepatitis C infection is also associated with the development of the sicca syndrome.[51] RA rarely causes specific renal lesions; NSAIDs and amyloid are more often responsible.

IMAGING FEATURES

Because early joint pathology in RA is confined to the synovium, standard radiographs are often not useful. Periarticular osteopenia of the metacarpophalangeal joints and proximal interphalangeal joints in the hand can be evident within months of onset. Joint space narrowing, caused by the loss of articular cartilage, indicates irreversible damage to such cartilage; RA must be active for at least 6 months for such damage to occur. Arthroscopic visualization of articular cartilage (e.g., in the knee) identifies damage to cartilage considerably earlier, but such findings have no value in the management of RA. Subchondral sclerosis is a feature of osteoarthritis but not of RA. Prominent periostitis with new bone formation is much more common in psoriatic arthritis or reactive arthritis syndrome.

Radiographically visualized bone erosions are best seen at the margins of the joint, where the synovium is reflected near the attachment of the capsule. The bone in this region (the so-called "bare area") is not protected by a layer of cartilage and is directly attacked by the invading synovium and osteoclasts. The erosions associated with RA may be difficult to distinguish from those of gout: the latter tend to have sharper borders and overhanging edges of bone, whereas the former are usually small and irregularly shaped. Cystlike radiolucencies may be seen in larger joints. Entire portions of bone adjacent to the joints, such as the metacarpal ends and the ulnar styloid process, may be resorbed. Cartilage destruction caused by RA tends to be evenly distributed within a joint. For instance, both the medial and the lateral compartments of the knee joint are narrowed in RA, whereas the medial compartment is more often affected in osteoarthritis. Progression of erosions takes time in RA; it is rarely necessary to repeat radiographs more often than every 12 months. Damage that is radiographically evident often occurs during the first 2 to 5 years and can progress inexorably in the absence of treatment.

MRI can distinguish synovial pannus from cartilage and synovial fluid and thus can detect pannus as it invades joint structures. The use of intravenous contrast materials, such as gadolinium, permits accurate assessment of synovial invasion and volume. MRI has replaced arthrography for the investigation of large joints, such as the knees.[52] The use of MRI to monitor response to therapy is still experimental because of a lack of uniform standards for judging damage. Although most erosions persist or progress, up to one quarter of them heal spontaneously.[53] Because of the lack of standardization, plain radiographs remain the gold standard for following disease progression.

LABORATORY EVALUATION

A mild normochromic, normocytic anemia and an elevated platelet count are usually present in patients with RA. The leukocyte count is generally normal, although neutropenia occurs in association with splenomegaly in Felty syndrome. The erythrocyte sedimentation rate (ESR) and the C-reactive protein level are usually elevated in active RA and are useful in monitoring disease activity and response to therapy. Results of serum

chemistry studies are normal, although the use of either NSAIDs or methotrexate can lead to elevations in liver enzyme levels. Urinalysis is generally normal.

About 80% to 85% of patients with RA are seropositive for rheumatoid factor. If seropositivity develops, it usually does so before the end of the first year of disease. From 1% to 5% of healthy persons test positive for rheumatoid factor, with the higher percentage noted in the elderly. Many chronic inflammatory conditions besides RA are associated with positive rheumatoid factor test results, although the titers are usually lower. Compared with rheumatoid factor, testing for antibodies to CCP appears to be more specific for RA (85% to 90%) but may be less sensitive (50% to 60%). Testing for anti-CCP antibodies could be useful as a diagnostic test in selected cases.

Other serologic tests commonly used to diagnose rheumatic diseases are of limited value in RA. Antinuclear antibodies (ANA) are often present in low titer. If anti-DNA antibodies are detected, they are almost always directed against single-stranded DNA rather than native double-stranded DNA. Antibodies to the antigens associated with Sjögren syndrome (SS-A and SS-B) may be positive. Serum complement levels are normal in uncomplicated RA; hypocomplementemia suggests systemic rheumatoid vasculitis. Serologic tests for viruses may help identify patients with postrubella arthritis or parvovirus B19 infection. Hepatitis B and C serologies can also provide useful information, because these infections can cause a self-limited symmetrical polyarthritis that mimics RA.

Analysis of synovial fluid provides supportive data but is rarely diagnostic. Synovial fluid in RA usually appears straw colored and mildly turbid. Bits of fibrin and, occasionally, small fronds of synovium may be aspirated. Leukocyte counts range from 2,000 to 20,000/mm^3. On differential counts, most of the cells (50% to 80%) are neutrophils; the remainder are lymphocytes (mainly T cells) and monocytes. The synovial fluid glucose level is usually normal, which distinguishes RA from acute infection. Synovial fluid complement levels are usually low in inflamed rheumatoid joints despite abundant production of complement proteins by synovium. Tests for rheumatoid factor, antinuclear antibodies, total protein, or lactate dehydrogenase in synovial fluid are not clinically useful. Synovial biopsies, either blind or arthroscopically directed, can be used in clinical trials to assess response to therapy. However, their utility for differential diagnosis or in predicting response to therapy remains limited.

Differential Diagnosis

The onset and course of RA can be highly variable, and the lack of a specific biologic marker makes diagnosis difficult. Prolonged observation and the integration of clinical and laboratory data that meet established criteria are often required.

Diagnostic criteria have been formulated by the American College of Rheumatology [*see Table 1*]. Conditions mimicking RA should be ruled out, if possible. Patients with other rheumatic diseases often have a symmetrical polyarticular arthritis resembling RA. The presence of high-titer ANA and anti–double-stranded DNA, a low serum complement level, and major organ system involvement (especially nephritis) are clues to the diagnosis of SLE. Careful physical examination often helps the clinician distinguish other rheumatic diseases. Elderly persons with polymyalgia rheumatica can present with peripheral synovitis, although prominent proximal muscle stiffness and a very high ESR (often higher than 80 to 100 mm/hr) are useful differential findings. Viral arthritis mediated by immune complex deposi-

Table 2 Comparison of Various Antirheumatic Treatments

Drug	*Response Rate; Onset of Action*	*Magnitude of Efficacy (0 to ++++)*	*Major Toxicities*	*Dosage*
Adalumimab	50%–70%; 2–4 wk	+++	Injection-site reactions, infections	40 mg S.C. every 2 wk
Anakinra	30%; 1–3 mo	+ to ++	Injection site reactions, infection	100 mg/day S.C.
Azathioprine	30%–50%; 2–3 mo	++	Hematologic, immunosuppression, cholestasis	100–150 mg/day
Cyclosporine	30%; 2–3 mo	++	Renal (irreversible), hypertension, hypertrichosis, immunosuppression	2.5–5.0 mg/kg/day
Etanercept	50%–70%; 2–4 wk	+++	Injection-site reaction, ?immune surveillance	25 mg S.C. twice a week
Gold	30%; 3–6 mo	++	Skin rash, hematologic, renal	5.0 mg/wk I.M. × 6 mo
Hydroxychloroquine	30%–50%; 2–6 mo	++	Retinopathy, myopathy, hyperpigmentation	200 mg b.i.d.
Infliximab	50%–70%; 2–4 wk	+++	Increased infection	3 mg/kg I.V. q. 8 wk
Leflunomide	50%; 2–3 mo	++	Liver, teratogen, gastrointestinal, skin rash	100 mg/day × 3, then 20 mg/day
Methotrexate	> 70%; 6–8 wk	+++	Liver (fibrosis, elevated enzymes), hematologic, oral ulcers	7.5–15 mg/wk
NSAIDs	> 75%; < 2 wk	+	Gastric erosion, renal	Varied
Prednisone	> 90%; < 1 wk	+++	Skin atrophy, cataracts, osteoporosis, avascular necrosis	5.0–7.5 mg/day
Sulfasalazine	> 30%; 2–3 mo	++	Dyspepsia, hemolysis in glucose-6-phosphate dehydrogenase deficiency	1 g b.i.d. or t.i.d.

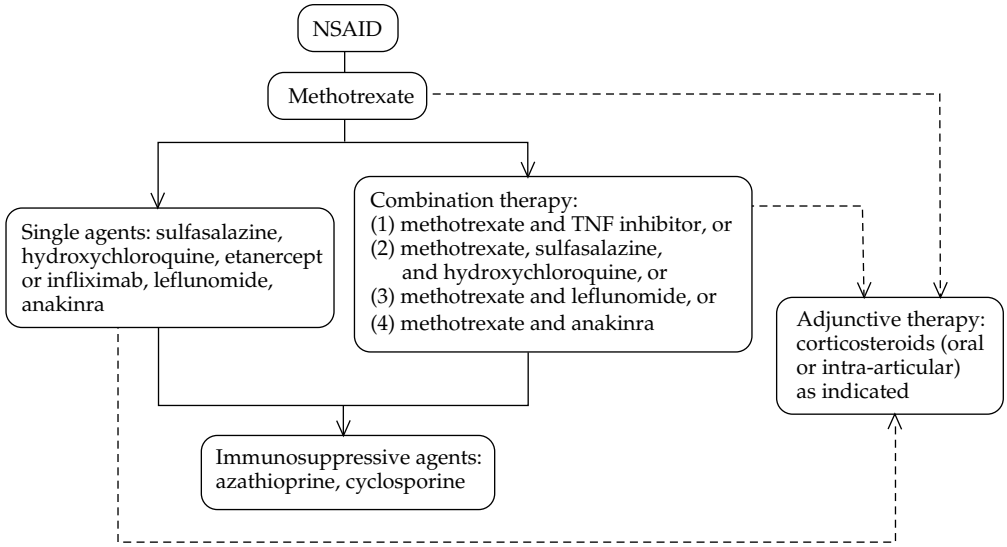


Figure 9 **Proposed algorithm for pharmacologic management of rheumatoid arthritis. The solid lines indicate the standard management options, and the broken lines indicate stages at which adjunctive therapy may be introduced. Most patients require rapid advancement from nonsteroidal anti-inflammatory drugs to a second-line agent, most often methotrexate.**

tion, as in hepatitis B or rubella, often has the same distribution as RA but is transient.

Metabolic disorders such as gout and calcium pyrophosphate deposition arthropathies can mimic RA. Radiographs may indicate characteristic gouty erosions or chondrocalcinosis. The finding of crystals in synovial fluid distinguishes these disorders from RA. Septic arthritis is also relatively easy to identify through clinical examination and evaluation of synovial fluid.

Seronegative spondyloarthropathies can present a diagnostic challenge when they exhibit peripheral polyarticular disease. They can generally be distinguished by the lack of symmetry, the distribution of affected joints (usually, lower extremities are affected more than upper, and large joints more than small), the absence of rheumatoid factor, the presence of proliferative bone changes on radiographs, and characteristic skin lesions. However, in some cases, psoriatic arthritis has a clinical picture almost identical to that of RA.

Morning stiffness and involvement of the wrist and metacarpophalangeal joints are uncommon in osteoarthritis, which typically affects the weight-bearing joints and the distal interphalangeal joints. Patients with osteoarthritis usually are seronegative for rheumatoid factor and lack marginal erosions. Synovial effusions are noninflammatory, with a WBC lower than 2,000/mm^3 and a predominance of mononuclear cells.

Rheumatoid joints are more susceptible to bacterial infection than normal joints, and superimposed sepsis may not be readily apparent. The usual signs (e.g., localized erythema, increased pain, and limitation of motion) may be difficult to distinguish from the underlying rheumatoid synovitis or may be suppressed by antirheumatic therapy. Multiple joints may be infected simultaneously; the diagnosis requires arthrocentesis and culture [*see 7:XV Septic Arthritis*].

Management

Optimal management requires an awareness of the variable course of the disorder. Statistical predictions about outcome can be made from clinical features and laboratory abnormalities, but remissions and exacerbations are common, and the risks associated with drugs and surgery must be viewed in the light of this uncertainty. The patient should be aware of the risks both of taking a drug and of not taking it. Active synovitis that persists for a year or more after the onset of RA results in irreversible cartilage damage, joint destruction, and increased mortality. Thus, every effort should be made to suppress the synovitis by pharmacologic methods during the early months.

No specific climate or diet alters the course of RA. Alternative therapies, including cartilage extracts, rarely have more than a placebo effect. However, the power of placebo effects should not be underestimated; most arthritis and pain studies demonstrate a "therapeutic" response in 20% to 30% of patients who receive placebo. Glucosamine and chondroitin sulfate have been studied in osteoarthritis, but few data support their use in RA.

In addition to conveying to the patient an understanding of the disease, management involves efforts to relieve pain and discomfort, preserve strength and joint function, prevent deformities, and attend to systemic complications. Surgical intervention is important not only for replacing destroyed joints but also, at times, for restoring function or preventing further damage.

DRUG THERAPY

Drugs are used for analgesia, to control inflammation, and to alter the natural history of disease. Only empirical data support the use of some currently available agents; the mechanism of action of these agents is unknown. In formal studies of individual drug therapies, systematic measurements are taken of the number of inflamed joints, the extent of swelling, and the range of joint motion. Laboratory measurements (e.g., ESR and hematocrit value) and assessments of subjective features (e.g., pain and morning stiffness) are also made. The information is then assessed globally. Some objective and subjective improvement is usually observed with any agent; placebo effects may be striking [*see Table 2*].

General Recommendations

The appropriate management of RA is rapidly evolving; previous treatment algorithms based on a gradual escalation of treatment (i.e., the traditional pyramid approach) have been replaced by more aggressive treatment approaches.[54] The change has been fomented by a variety of factors, including the follow-

ing: (1) active RA significantly decreases the life span of affected individuals; (2) active inflammation is associated with increased morbidity and mortality; (3) more effective therapy is available; and (4) the advent of combination therapy and new drugs has had a clear effect on the natural history of disease. No single algorithm can capture the complexity of RA management today because of the extensive pharmacopoeia, although broad guidelines can be given [*see Figure 9*].

Most patients require rapid advancement from NSAIDs to a second-line agent, most often methotrexate. Because symptoms of RA will not be adequately controlled in 70% of patients by methotrexate alone, the clinician is usually faced with the choice of either "add-on" therapy or testing a series of single agents (e.g., sulfasalazine, antimalarials, leflunomide, a TNF inhibitor, or gold). In the United States, most rheumatologists prefer to increase the methotrexate dosage rapidly to 20 to 25 mg/wk and then add another agent within 2 to 3 months if necessary. Morning stiffness lasting longer than 30 minutes, continued pain, or evidence of active synovitis on physical examination is an indication for advancing therapy even if the patient has experienced significant improvement on methotrexate. Typically, one would add either a TNF inhibitor, leflunomide, or sulfasalazine (with or without hydroxychloroquine). Care must be exercised, especially with combinations of methotrexate and leflunomide, because of hepatotoxicity.

Few data demonstrate the superiority of one combination of drugs over another, although there is an increasing bias toward use of TNF inhibitors early in management.[55] Prednisone is generally reserved for patients requiring adjunctive "bridge" therapy to improve their ability to perform activities of daily living; the prednisone therapy is maintained until a single-drug or combination therapy permits its use to be tapered or discontinued. The most recent addition to the armamentarium, the IL-1 inhibitor anakinra, is modestly effective; it is usually reserved for patients who do not respond to combinations of methotrexate and TNF inhibitors.[56]

Alternative algorithms have also been proposed for the management of RA. Some data suggest that early management with triple therapy (e.g., sulfasalazine, hydroxychloroquine, and methotrexate) is very effective.[57] Other options include early high-dose corticosteroid treatment, with the dose tapered over several months, in combination with methotrexate and sulfasalazine given in stable doses. In such protocols, patients tend to improve quite rapidly because of the steroid. This form of early aggressive treatment appears to have a persistent effect and is associated with less radiographically evident joint damage, even after corticosteroids or one of the other agents has been discontinued.[58] However, it is very difficult to assess the efficacy of second-line drugs if prednisone is added early in the course of treatment.

Although the number of cases of treatment-resistant disease is decreasing, a fraction are recalcitrant to all of the aforementioned drugs. Immunosuppressive agents such as azathioprine or cyclosporine can be used, although the therapeutic ratio is narrow. Experimental approaches can also be considered with appropriate oversight from institutional review boards.

Nonsteroidal Anti-inflammatory Drugs

The use of aspirin for RA has decreased substantially because of the availability of newer NSAIDs. Nonacetylated salicylate compounds, such as choline salicylate and choline magnesium trisalicylate, produce less gastric irritation than aspirin but are more expensive and exert less anti-inflammatory effect.

Other NSAIDs are probably no more effective than aspirin but may have certain advantages (e.g., fewer GI effects, better pharmacokinetics, and, usually, fewer pills to be taken daily, which may enhance compliance). The prostaglandin analogue misoprostol, proton pump inhibitors, and, to a lesser extent, H_2 receptor blockers can suppress the GI toxicity of NSAIDs[59] and should be considered for patients who take a nonselective cyclooxygenase inhibitor but have a history of duodenal ulcer or gastritis.[60] *Helicobacter pylori* infection, use of salicylates, and use of NSAIDs are likely independent risk factors for peptic ulceration.[61]

Pharmacokinetics aside, there are a few pharmacologic differences between the various NSAIDs. Sulindac may have less renal toxicity than other NSAIDs, but the clinical relevance is not certain. Most currently available NSAIDs nonselectively block both the constitutively expressed cyclooxygenase COX-1 and the inducible cyclooxygenase COX-2, although the relative selectivity for COX-2 can be favorable with some compounds (e.g., meloxicam and diclofenac). Selective COX-2 inhibitors (e.g., celecoxib and valdecoxib) are also effective anti-inflammatory and analgesic agents.[62] They have fewer GI side effects than nonselective COX inhibitors. In addition, they do not block platelet function and may be used in some situations in which traditional NSAIDs may be contraindicated. Both selective and nonselective agents, however, can alter renal blood flow and glomerular filtration rate. Celecoxib is administered in a dosage of 100 to 200 mg orally once or twice daily, whereas valdecoxib is usually given in a once-daily oral dosage of 10 mg. The association of cardiovascular events with prolonged use of the COX-2 inhibitor rofecoxib (Vioxx) prompted the withdrawal of the drug from the global market in 2004.[63,64] There is scant evidence to assess the long-term risk of other COX-2 inhibitors. When prescribing COX-2 inhibitors, clinicians should weigh the potential cardiovascular risks against anticipated benefits and consider issues such as dose and comorbid conditions.

Advancement from NSAIDs to second-line agents is recommended if (1) symptoms have not improved sufficiently after a short trial of NSAIDs, (2) the patient has aggressive seropositive disease, or (3) there is radiographic evidence of erosions or joint destruction. The trend today is for more aggressive treatment, and the majority of patients require additional pharmacotherapy.

Methotrexate

Methotrexate is one of the most effective second-line drugs. Methotrexate not only alleviates the signs and symptoms of RA but also decreases the ESR and raises the hematocrit value.[65] It probably also slows the rate of bone erosion in RA, perhaps by decreasing expression of destructive enzymes such as collagenase in synovium.[66]

Methotrexate is usually given in weekly oral doses, beginning at 7.5 mg and, if necessary, increasing to 15 mg over 2 to 3 months. Response to therapy is usually relatively rapid, with improvement occurring in 6 to 8 weeks. If no response is observed, the dose can be increased to 20 to 25 mg/wk [*see Figure 9*]. Further increases up to 45 mg/wk administered parenterally had minimal additional benefit.[67] Over 70% of patients have a response to methotrexate, and half remain on the drug for at least 5 years. Efficacy remains excellent even after years of use.[68] Complete remissions are uncommon, although patients' sense of well-being is often dramatically improved. Inflammation usually reappears within weeks after discontinuance. Monitoring of hematologic and liver parameters every 4 to 8 weeks is required. The primary action of methotrexate is anti-inflammatory at the

doses used in RA, although immunosuppression resulting in *Pneumocystis* pneumonia has been reported. Improved survival as a result of a reduction in cardiovascular mortality has been reported in patients treated with methotrexate.[69]

Risk factors for toxicity include alcoholism, diabetes, obesity, advanced age, and renal disease. The major concern is hepatic fibrosis, although marrow toxicity and sterility are also important side effects. Methotrexate is a potent teratogen and should not be used in women of childbearing age unless they are using a reliable form of contraception. Idiosyncratic interstitial lung disease can develop even at low doses. Other adverse reactions to methotrexate include nausea, stomatitis, leukopenia, diarrhea, and elevations of serum aminotransferase levels. Some of these toxicities (especially oral ulcers) can be minimized by prophylactic administration of folic acid, 1 mg/day.[70] Hepatotoxicity leading to clinically significant fibrosis or cirrhosis is exceedingly uncommon; accordingly, most centers have abandoned routine monitoring with liver biopsies.[71] A biopsy should be performed if a patient develops persistent abnormalities in blood tests for liver enzymes that do not promptly resolve with discontinuance. Other medications that interfere with folate metabolism (e.g., trimethoprim) should be used with caution in patients taking methotrexate.

Leflunomide

Leflunomide is an effective antirheumatic agent that blocks the pyrimidine synthesis required for stimulated lymphocytes to proliferate. In vitro, it inhibits mitogen-stimulated proliferation of both B cells and T cells. Randomized, controlled trials have demonstrated that leflunomide is approximately as effective as methotrexate in the treatment of active RA.[72] In addition to being clinically effective, leflunomide slows radiographic progression of RA.[73]

Because the half-life of leflunomide is 2 weeks, loading doses are necessary to achieve therapeutic blood levels promptly. A loading dose of 100 mg/day for 3 days is sometimes given, followed by daily dosing with 20 mg/day. Leflunomide can be used in combination with methotrexate, although in this setting the initial dose of leflunomide is usually 10 mg/day; patients must be followed very carefully for hepatotoxicity.[74]

Adverse effects include diarrhea, liver toxicity, rash, oral ulcers, and reversible hair loss.[75] Periodic monitoring of liver enzymes is required. In animal studies, leflunomide has been associated with birth defects; thus, this agent should not be used by pregnant women or women of childbearing age who are not using a reliable form of contraception.

Antimalarial Drugs and Sulfasalazine as Single and Combination Agents

The antimalarial drug hydroxychloroquine is useful as early second-line therapy for RA.[76] Its response rate is lower than that of methotrexate, and less improvement is seen; however, its relative safety makes it an ideal choice for patients with mild early disease or as an additive agent in combination therapy. Adverse reactions to antimalarials occur, particularly retinopathy that may lead to an irreversible decrease in vision; this reaction is rare and can be minimized with regular ophthalmologic examinations.

Sulfasalazine is also useful.[77] It is effective in at least 30% of patients at a dosage of 2 to 3 g/day in divided doses. Only moderate side effects have been reported, especially GI upset. Sulfasalazine is well tolerated; over 30% of patients continue to take it for at least 5 years.[78]

A significant percentage of RA patients do not experience satisfactory symptomatic relief with a single disease-modifying antirheumatic drug. This lack of response has led clinical investigators to study combination regimens,[79] but most of the studies performed have been poorly controlled. A prospective study examining azathioprine and methotrexate alone and in combination demonstrated safety but no additive benefits.[80] A 2-year study showed that a combination of methotrexate, sulfasalazine, and hydroxychloroquine is more effective than any of the agents alone.[81]

Gold Salts

Injectable gold salts decrease inflammation and increase the likelihood of remission. Gold was once the primary second-line agent, but its relatively modest efficacy and high side-effect profile, coupled with the increased use of methotrexate and anticytokine therapy, have led to a dramatic decrease in its use.

Treatment usually consists of intramuscular administration of a compound such as gold sodium thiomalate in a test dose of 10 mg, followed by a single 25 mg dose 1 week later and then by 50 mg once weekly for 20 weeks. If the response is satisfactory at this point, the compound can be given every other week for another several months and monthly thereafter for an indefinite period. The response rate to gold is about 30%, although toxicity and lack of efficacy result in discontinuance in almost 90% of patients within 5 years.

Major adverse side effects involve hematologic, renal, and dermatologic reactions. Oral gold preparations appear to be less toxic than injectable ones but are less effective.

Anticytokine Therapy

TNF inhibitors Interference with the function of TNF-α, either by soluble receptor blockade or by giving monoclonal antibody, is effective in treating RA.[82] By itself, the soluble 75 kd TNF receptor has a very short biologic half-life. Recombinant technology was used to engineer the fusion of the receptor to the Fc portion of the immunoglobulin molecule (etanercept), greatly prolonging the half-life and making twice-weekly subcutaneous injections feasible. The fusion protein is entirely human in origin, and antibody formation against it is minimal.

Etanercept has been approved for use in the United States (25 mg S.C. twice a week) and has been shown to be beneficial in patients who have only a partial response to methotrexate used either in combination therapy or as monotherapy.[83] A single injection of 50 mg/wk appears to have similar efficacy. The fusion protein is also effective in children with RA.

Two monoclonal antibodies have been approved for use in RA: the human-mouse chimeric monoclonal antibody infliximab and the human monoclonal antibody adalumimab. There are subtle differences in the pharmacology of etanercept and the monoclonal antibodies. Etanercept binds both TNF-α and TNF-β, whereas the monoclonal antibodies have a greater affinity for the TNF-α ligand and bind only to TNF-α. However, the clinical impact of this difference is not clear, and the response rates are similar.

Infliximab is used in combination with methotrexate for treatment of RA; this appears to permit long-term use of infliximab with less formation of neutralizing antibodies.[84] Infliximab is administered by intravenous infusion; the recommended dose is 3 to 10 mg/kg every 8 weeks. As with etanercept and all TNF inhibitors, the drug must be used with care in the presence of infections.

Adalimumab has demonstrated efficacy similar to that of infliximab and etanercept in RA. It is generally administered at a dose of 40 mg S.C. every 2 weeks. In perhaps 10% of patients, the frequency of administration must be increased to once a week.

The clinical response to TNF inhibitors can be dramatic, and an improved sense of well-being can occur within days. Typically, patients who respond to TNF inhibitors will notice decreased pain and stiffness after a few weeks of therapy. About two thirds of patients have at least 20% improvement in the number of swollen and tender joints, and half of these patients will have more than 50% improvement. Concomitant with decreased symptoms, serum and synovial cytokine levels decrease, and peripheral blood markers of inflammation (i.e., ESR and the C-reactive protein level) decline. Although the clinical effect occurs sooner with the TNF inhibitors than with methotrexate, a significant benefit over methotrexate has not been demonstrated after 3 months of therapy.[83]

In addition to providing symptomatic improvement, TNF inhibitors prevent or slow bone and cartilage destruction in RA. Perhaps most intriguing is the observation that radiographic progression is prevented even in patients who do not experience clinical improvement after treatment with infliximab. This observation could have important implications for the widespread clinical use of these agents if long-term studies confirm these findings. For now, however, these agents are generally used only after an adequate trial of methotrexate.

Because TNF is important in host defense and possibly in tumor surveillance, impairment of these functions in treated patients could increase the risk of infection and even cancer. The most common adverse reaction to etanercept is local inflammation at the injection site, although serious bacterial and mycobacterial infections have also been observed with TNF inhibitors.[85] Hence, these agents should be used with caution in patients with active infections. Before therapy with an anti-TNF agent is initiated, a skin test should be performed and a chest x-ray obtained to rule out tuberculosis; prophylactic treatment with isoniazid should be initiated if indicated. More intriguing is the induction of antinuclear and anti-DNA antibodies and, in a few cases, frank SLE in some RA patients treated with anti–TNF-α therapy. Long-term studies of larger numbers of patients are required to assess these problems, although very few serious side effects have been noted in patients treated for over 5 years with etanercept. However, demyelinating syndromes and aplastic anemia have been noted in a small number of patients receiving TNF inhibitors. These agents appear to exacerbate multiple sclerosis; this finding is possibly related to the protective role TNF-α may play in the central nervous system.[86]

Although no overall increase in malignancies has been observed, some concerns have been raised regarding an increased risk of lymphoma in patients receiving TNF inhibitors. However, this question is confounded by the fact that the prevalence of lymphoma is higher in patients with RA, compared with that in normal individuals, and it is unclear whether the use of TNF inhibitors further increases the risk of lymphoma. Another important consideration is whether the improved control of disease activity provided by TNF inhibitors diminishes mortality in RA overall, even if their use may be associated with a slight risk of certain malignancies.

IL-1 inhibitor The natural inhibitor to IL-1, IL-1ra (anakinra), demonstrates modest anti-inflammatory activity in RA.[87] Anakinra also appears to have disease-modifying activity, evidenced by a decrease in radiographic progression. Anakinra is administered subcutaneously at a dose of 100 mg by daily injection, either alone or in combination with methotrexate.[88] The most common side effect is injection-site reaction, although an increased incidence of infections has also been reported. Use of anakinra in combination with TNF inhibitors may markedly increase infectious complications; therefore, such combinations should be avoided until further studies clarify the risks and benefits. Anakinra has been approved for use in the United States.

Other anticytokine therapies Antibodies to the IL-6 receptor also may be effective in RA. Similarly, anti–B cell therapy with rituximab and blockade of T cell costimulation with CTLA4-Ig looked promising in phase II clinical trials. A preliminary study using IL-10 to treat RA did not show significant alterations in synovial histology, synovial cytokine production, or clinical disease activity.[89] Many other approaches are also under investigation, including the use of anti–IL-12 antibodies, IL-18 binding proteins, and inhibitors of specialized chemoattractant cytokines called chemokines.

Glucocorticoids

Glucocorticoids are some of the most potent anti-inflammatory agents available. When used systemically, they decrease joint swelling, pain, and morning stiffness and improve ability to function. Unfortunately, the dosages necessary to maintain such improvement are usually high enough to be associated with long-term side effects (e.g., osteoporosis, osteonecrosis, increased susceptibility to infection, cataracts, myopathy, and poor wound healing). Alternate-day therapy usually cannot be used to reduce these side effects, because RA patients are usually symptomatic on the off days. The conventional wisdom is that glucocorticoids neither alter the course of the disease nor affect the ultimate degree of damage to joints or other structures; however, some evidence indicates that low-dose prednisolone given early in RA can slow the rate of radiographic progression.[90] Although there is considerable difference of opinion regarding the most appropriate use of steroids in RA, they are typically employed as "bridge" therapy for severe disease while awaiting a therapeutic response from other second-line agents. Systemic glucocorticoids, however, even in these low dosages, are associated with accelerated bone loss. Supplemental calcium (1,000 to 1,500 mg daily) and vitamin D (400 to 800 I.U. daily) can be given to postmenopausal women who require daily prednisone therapy, even in low doses.[91] Dual-energy x-ray absorptiometry (DEXA) scans should be used to monitor bone mineral density, and bisphosphonates should be used to treat or prevent osteoporosis in susceptible patients.

Intra-articular glucocorticoids are also useful in limited flares. Administration of such agents requires careful aseptic technique. The procedure involves injection of a local anesthetic agent (e.g., lidocaine) and an insoluble glucocorticoid preparation (e.g., triamcinolone hexacetonide, 20 to 40 mg) into a large joint; smaller amounts (5 to 10 mg) may be injected into small joints, bursae, and tendon sheaths. These local injections usually result in relief of pain and inflammation within days and do not often produce serious side effects. The risk of infection is probably about one in 10,000 procedures. The beneficial effects may last for weeks or months (on average, about 3 to 6 months), but repeated injections may result in increased cartilage destruction, osteonecrosis, and tendon rupture. Thus, intra-articular glucocorticoids are only occasionally useful in relieving inflammation in one or two joints that are particularly symptomatic.

Immunosuppressive Agents

Alkylating agents, particularly cyclophosphamide, and antimetabolites such as azathioprine have been used in patients with severe progressive disease who have not responded to the measures already described. These drugs decrease inflammation and possibly reduce the frequency of new joint erosions. Cyclophosphamide is also effective in controlling rheumatoid vasculitis in some patients. Hematologic toxicity and GI toxicity can be severe; hemorrhagic cystitis is a disturbing complication of oral cyclophosphamide therapy. Azathioprine is safer but has only modest efficacy. The additional potential hazard of inducing neoplasias and chromosomal abnormalities also restricts the usefulness of these agents.

Cyclosporine is a more focused immunosuppressive drug than azathioprine or cyclophosphamide because it targets T cells; it has been used extensively in allograft rejection. Hypertension and decreased creatinine clearance are common side effects, generally related to the cumulative dose. In RA, cyclosporine at a dosage of 2.5 to 5.0 mg/kg/day is an alternative to cyclophosphamide or azathioprine in patients who need immunosuppressive therapy, provided patients are closely monitored for renal toxicity.[92] Despite its potent immunosuppressive effects, a minority of patients have a meaningful clinical response to cyclosporine.

Other Chemotherapeutic Agents

In some studies, minocycline appeared to provide benefit for RA patients.[93] Its mechanism of action is unclear and may be related to the ability of tetracycline analogues to inhibit metalloproteinases. Current research into therapies for RA includes trials of recombinant products and humanized monoclonal antibodies. Some of these preparations showed promising results in early open studies, only to fail in controlled trials. This is especially true of antibodies directed against T cell markers (e.g., CD4).[94] Other novel therapies have had mixed success. Attempts to induce tolerance by oral administration of type II collagen or cartilage protein gp39 have not provided significant benefit. A diet supplemented with fish oils, which contain omega-3 fatty acids, reduces synthesis of inflammatory arachidonate metabolites and may be a useful adjuvant therapy for selected patients.[95]

PHYSICAL THERAPY

Hospitalization is occasionally helpful in the management of patients with RA, although this approach in an era of managed care and cost constraints is not feasible. Removing the patient from a stressful home environment and instituting a program of rest combined with physical therapy are of great value. Splinting inflamed joints may decrease synovitis.

Physical therapy has a role in the management of RA, although data supporting its ability to change outcome are lacking. Passive range-of-motion exercises help prevent contractures. Isometric exercises build up muscle strength without subjecting inflamed joints to excessive wear, and isotonic exercises further increase muscle strength and help preserve function. Most physical measures, such as whirlpools, heated wax, ultrasonography, and diathermy, make patients feel better during the procedure and perhaps for a short time afterward but offer no significant long-term functional, anti-inflammatory, or disease-modifying benefit; consequently, many patients eventually become disillusioned with them. It is important for patients to maintain an active life, and guidance from physical therapists for range-of-motion exercises and aerobic training is useful. Swimming or other water exercises are especially useful aerobic stresses that minimize the load on the lower extremities.

SURGERY

Indications for surgical intervention include intractable pain and impaired function. Eroded cartilage, ruptured ligaments, and progressive destruction of bone can lead to severe functional derangement that is amenable only to surgical correction. Besides helping restore function to weight-bearing joints, surgery may also restore function in severely deformed hands. In a joint such as the wrist, dorsal synovectomy may prevent extensor tendon ruptures. Although proliferative synovitis often recurs after synovectomy, it may take 1 or 2 years to return and may be less intense than it was initially. Surgery is also useful for removing frayed menisci and other loose bodies that interfere with joint function. In the hands and wrists, operations on periarticular structures (e.g., repair of capsules and replacement of tendons) may restore appearance and function; release of carpal tunnel compression usually relieves pressure on the median nerve. Arthroscopic surgery to remove cartilaginous fragments and to perform a partial synovectomy may be useful when a large, accessible joint (e.g., the knee) is involved with proliferative synovitis.

If gross deformity and joint destruction have occurred, more definitive procedures may be required. In some joints, such as the wrist and ankle, function may be improved by stabilizing the joint through fusion, albeit at the cost of loss of motion. For destroyed joints, total replacement may be necessary. Hip prostheses provide a stable, pain-free joint with a good range of motion in more than 90% of patients. Metal-to-plastic prostheses are also useful in reconstruction of knee, elbow, and shoulder joints. Joint replacement procedures involve a relatively high risk of thromboembolism, but serious infections are rare. Loosening of the components has been observed within several years in as many as 20% of patients.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Nepom GT, Byers P, Seyfried C, et al: HLA genes associated with rheumatoid arthritis: identification of susceptibility alleles using specific oligonucleotide probes. Arthritis Rheum 32:15, 1989

2. Cvetkovic JT, Wallberg-Jonsson S, Stegmayr B, et al: Susceptibility for and clinical manifestations of rheumatoid arthritis are associated with polymorphisms of the TNF-alpha, IL-1beta, and IL-1Ra genes. J Rheumatol 29:212, 2002

3. McDowell TL, Symons JA, Ploski R, et al: A genetic association between juvenile rheumatoid arthritis and a novel interleukin-1 alpha polymorphism. Arthritis Rheum 38:221, 1995

4. Stransky G, Vernon J, Aicher WK, et al: Virus-like particles in synovial fluids from patients with rheumatoid arthritis. Br J Rheumatol 32:1044, 1993

5. Stahl HD, Pfeiffer R, Von Salis-Soglio G, et al: Parvovirus B19–associated mono- and oligoarticular arthritis may evolve into a chronic inflammatory arthropathy fulfilling criteria for rheumatoid arthritis or spondylarthropathy. Clin Rheumatol 19:510, 2000

6. Takahashi Y, Murai C, Shibata S, et al: Human parvovirus B19 as a causative agent for rheumatoid arthritis. Proc Natl Acad Sci USA 95:8227, 1998

7. Ray NB, Nieva DR, Seftor EA, et al: Induction of an invasive phenotype by human parvovirus B19 in normal human synovial fibroblasts. Arthritis Rheum 44:1582, 2001

8. Albani S, Carson DA: A multistep molecular mimicry hypothesis for the pathogenesis of rheumatoid arthritis. Immunol Today 17:466, 1996

9. Auger I, Roudier J: A function for the QKRAA amino acid motif: mediating binding of DnaJ to DnaK: implications for the association of rheumatoid arthritis with HLA-DR4. J Clin Invest 99:1818, 1997

10. Nakajima T, Aono H, Hasunuma T, et al: Overgrowth of human synovial cells driven by the human T cell leukemia virus type I *tax* gene. J Clin Invest 92:186, 1993

11. Aicher WK, Heer AH, Trabandt A, et al: Overexpression of zinc-finger transcription factor Z-225/Egr-1 in synoviocytes from rheumatoid arthritis patients. J Immunol 152: 5940, 1994

12. Rantapaa-Dahlqvist S, de Jong BA, Berglin E, et al: Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum 48:2741, 2003

13. Firestein GS, Zvaifler NJ: How important are T cells in chronic rheumatoid synovitis? II. T cell-independent mechanisms from beginning to end. Arthritis Rheum 46:298, 2002

14. Firestein GS, Yeo M, Zvaifler NJ: Apoptosis in rheumatoid arthritis synovium. J Clin Invest 96:1631, 1995

15. Zvaifler NJ, Firestein GS: Pannus and pannocytes: alternative models of joint destruction in rheumatoid arthritis. Arthritis Rheum 37:783, 1994

16. Corr M, Zvaifler NJ: Mesenchymal precursor cells. Ann Rheum Dis 61:3, 2002

17. Mengshol JA, Mix KS, Brinckerhoff CE: Matrix metalloproteinases as therapeutic targets in arthritic diseases: bull's-eye or missing the mark? Arthritis Rheum 46:13, 2002

18. Gravallese EM, Goldring SR: Cellular mechanisms and the role of cytokines in bone erosions in rheumatoid arthritis. Arthritis Rheum 43:2143, 2000

19. Firestein GS: Invasive fibroblast-like synoviocytes in rheumatoid arthritis: passive responders or transformed aggressors? Arthritis Rheum 39:1781, 1996

20. Muller-Ladner U, Kriegsmann J, Franklin BN, et al: Synovial fibroblasts of patients with rheumatoid arthritis attach to and invade normal human cartilage when engrafted into SCID mice. Am J Pathol 149:1607, 1996

21. Tak PP, Zvaifler NJ, Green DR, et al: Rheumatoid arthritis and p53: how oxidative stress might alter the course of inflammatory diseases. Immunol Today 21:78, 2000

22. Gringhuis SI, Leow A, Papendrecht-Van Der Voort EA, et al: Displacement of linker for activation of T cells from the plasma membrane due to redox balance alterations results in hyporesponsiveness of synovial fluid T lymphocytes in rheumatoid arthritis. J Immunol 164:2170, 2000

23. Warrington KJ, Takemura S, Goronzy JJ, et al: $CD4^+$,$CD28^-$ T cells in rheumatoid arthritis patients combine features of the innate and adaptive immune systems. Arthritis Rheum 44:13, 2001

24. Wollheim FA: Predictors of joint damage in rheumatoid arthritis. APMIS 104:81, 1996

25. Ermel RW, Kenny TP, Chen PP, et al: Molecular analysis of rheumatoid factors derived from rheumatoid synovium suggests an antigen-driven response in inflamed joints. Arthritis Rheum 36:380, 1993

26. Vossenaar ER, Nijenhuis S, Helsen MM, et al: Citrullination of synovial proteins in murine models of rheumatoid arthritis. Arthritis Rheum 48:2489, 2003

27. Goldbach-Mansky R, Lee J, McCoy A, et al: Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset. Arthritis Res 2:236, 2000

28. Benoist C, Mathis D: Autoimmunity provoked by infection: how good is the case for T cell epitope mimicry? Nat Immunol 2:797, 2001

29. Firestein GS, Zvaifler NJ: How important are T cells in chronic rheumatoid synovitis? Arthritis Rheum 33:768, 1990

30. Chabaud M, Lubberts E, Joosten L, et al: IL-17 derived from juxta-articular bone and synovium contributes to joint degradation in rheumatoid arthritis. Arthritis Res 3:168, 2001

31. Liblau RS, Singer SM, McDevitt HO: Th1 and Th2 $CD4^+$ T cells in the pathogenesis of organ-specific autoimmune diseases. Immunol Today 16:34, 1995

32. Katsikis KD, Chu CQ, Brennan FM, et al: Immunoregulatory role of interleukin 10 in rheumatoid arthritis. J Exp Med 179:1517, 1994

33. Firestein GS, Boyle D, Yu C, et al: Synovial interleukin-1 receptor antagonist and interleukin-1 balance in rheumatoid arthritis. Arthritis Rheum 37:644, 1994

34. Firestein GS, Alvaro-Gracia JM, Maki R: Quantitative analysis of cytokine gene expression in rheumatoid arthritis. J Immunol 144:3347, 1990

35. McInnes IB, Leung BP, Sturrock RD, et al: Interleukin-15 mediates T cell–dependent regulation of tumor necrosis factor–alpha production in rheumatoid arthritis. Nat Med 3:189, 1997

36. Gracie JA, Forsey RJ, Chan WL, et al: A proinflammatory role for IL-18 in rheumatoid arthritis. J Clin Invest 104:1393, 1999

37. Brennan P, Bankhead C, Silman A, et al: Oral contraceptives and rheumatoid arthritis: results from a primary care-based incident case-control study. Semin Arthritis Rheum 26:817, 1997

38. Drossaers-Bakker KW, Zwinderman AH, van Zeben D, et al: Pregnancy and oral contraceptive use do not significantly influence outcome in long term rheumatoid arthritis. Ann Rheum Dis 61:405, 2002

39. Weyand CM, Goronzy JJ, Liuzzo G, et al: T-cell immunity in acute coronary syndromes. Mayo Clin Proc 76:1011, 2001

40. Kauppi M, Sakaguchi M, Konttinen YT, et al: Pathogenetic mechanism and prevalence of the stable atlantoaxial subluxation in rheumatoid arthritis. J Rheumatol 23:831, 1996

41. Foster H, Stephenson A, Walker D, et al: Linkage studies of HLA and primary Sjögren's syndrome in multicase families. Arthritis Rheum 36:473, 1993

42. Helmers R, Galvin J, Hunninghake GW: Pulmonary manifestations associated with rheumatoid arthritis. Chest 100:235, 1991

43. MacDonald WJ, Crawford MH, Klippel JH, et al: Echocardiographic assessment of cardiac structure and function in patients with rheumatoid arthritis. Am J Med 63:890, 1977

44. Peeters HR, Jongen-Lavrencic M, Vreugdenhil G, et al: Effect of recombinant human erythropoietin on anaemia and disease activity in patients with rheumatoid arthritis and anaemia of chronic disease: a randomised placebo controlled double blind 52 weeks clinical trial. Ann Rheum Dis 55:739, 1996

45. Sienknecht CW, Urowitz MB, Pruzanski W, et al: Felty's syndrome: clinical and serological analysis of 34 cases. Ann Rheum Dis 36:500, 1977

46. Hartman KR: Anti-neutrophil antibodies of the immunoglobulin M class in autoimmune neutropenia. Am J Med Sci 308:102, 1994

47. Starkebaum G, Loughran TP Jr, Gaur LK, et al: Immunogenetic similarities between patients with Felty's syndrome and those with clonal expansions of large granular lymphocytes in rheumatoid arthritis. Arthritis Rheum 40:624, 1997

48. Rashba EJ, Rowe JM, Packman CH: Treatment of the neutropenia of Felty syndrome. Blood Rev 10:177, 1996

49. Puechal X, Said G, Hilliquin P, et al: Peripheral neuropathy with necrotizing vasculitis in rheumatoid arthritis: a clinicopathologic and prognostic study of thirty-two patients. Arthritis Rheum 38:1618, 1995

50. Moris G, Garcia-Monco JC: The challenge of drug-induced aseptic meningitis. Arch Intern Med 159:1185, 1999

51. Jorgensen C, Legouffe MC, Perney P, et al: Sicca syndrome associated with hepatitis C virus infection. Arthritis Rheum 39:1166, 1996

52. Poleksic L, Musikic P, Zdravkovic D, et al: MRI evaluation of the knee in rheumatoid arthritis. Br J Rheumatol 35(suppl 3):36, 1996

53. McQueen FM, Benton N, Crabbe J, et al: What is the fate of erosions in early rheumatoid arthritis? Tracking individual lesions using x rays and magnetic resonance imaging over the first two years of disease. Ann Rheum Dis 60:859, 2001

54. Guidelines for the management of rheumatoid arthritis: 2002 update. Arthritis Rheum 46:328, 2002

55. Furst DE, Breedveld FC, Burmester GR, et al: Updated consensus statement on tumour necrosis factor blocking agents for the treatment of rheumatoid arthritis (May 2000). Ann Rheum Dis 59(suppl 1):i1, 2000

56. Jiang Y, Genant HK, Watt I, et al: A multicenter, double-blind, dose-ranging, randomized, placebo-controlled study of recombinant human interleukin-1 receptor antagonist in patients with rheumatoid arthritis: radiologic progression and correlation of Genant and Larsen scores. Arthritis Rheum 43:1001, 2000

57. Neva MH, Kauppi MJ, Kautiainen H, et al: Combination drug therapy retards the development of rheumatoid atlantoaxial subluxations. Arthritis Rheum 43:2397, 2000

58. Landewe RB, Boers M, Verhoeven AC, et al: COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 46:347, 2002

59. Raskin JB, White RH, Jackson JE, et al: Misoprostol dosage in the prevention of nonsteroidal anti-inflammatory drug-induced gastric and duodenal ulcers: a comparison of three regimens. Ann Intern Med 123:344, 1995

60. Gabriel SE, Jaakkimainen RL, Bombardier C: The cost-effectiveness of misoprostol for nonsteroidal antiinflammatory drug–associated adverse gastrointestinal events. Arthritis Rheum 36:447, 1993

61. Furst DE: Are there differences among nonsteroidal antiinflammatory drugs? Comparing acetylated salicylates, nonacetylated salicylates, and nonacetylated nonsteroidal antiinflammatory drugs. Arthritis Rheum 37:1, 1994

62. Hinz B, Brune K: Cyclooxygenase-2: 10 years later. J Pharmacol Exp Ther 300:367, 2002

63. Singh D: Merck withdraws arthritis drug worldwide. BMJ 329:816, 2004

64. Bombardier C, Laine L, Reicin A, et al: Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 343:1520, 2000

65. Alarcon GS: Methotrexate use in rheumatoid arthritis: a clinician's perspective. Immunopharmacology 47:259, 2000

66. Lopez-Mendez A, Daniel WW, Reading JC, et al: Radiographic assessment of disease progression in rheumatoid arthritis patients enrolled in the cooperative systematic studies of the rheumatic diseases program randomized clinical trial of methotrexate, auranofin, or a combination of the two. Arthritis Rheum 36:1364, 1993

67. Lambert CM, Sandhu S, Lochhead A, et al: Dose escalation of parenteral methotrexate in active rheumatoid arthritis that has been unresponsive to conventional doses of methotrexate: a randomized, controlled trial. Arthritis Rheum 50:364, 2004

68. Weinblatt ME, Maier AL, Fraser PA, et al: Longterm prospective study of methotrexate in rheumatoid arthritis: conclusion after 132 months of therapy. J Rheumatol 25:238, 1998

69. Choi HK, Hernan MA, Seeger JD, et al: Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. Lancet 359:1173, 2002

70. Morgan SL, Baggott JE, Vaughn WH, et al: Supplementation with folic acid during methotrexate therapy for rheumatoid arthritis: a double-blind, placebo-controlled trial. Ann Intern Med 121:833, 1994

71. Erickson AR, Reddy V, Vogelgesang SA, et al: Usefulness of the American College of Rheumatology recommendations for liver biopsy in methotrexate-treated rheumatoid arthritis patients. Arthritis Rheum 38:1115, 1995

72. Mladenovic V, Domljan Z, Rozman B, et al: Safety and effectiveness of leflunomide in the treatment of patients with active rheumatoid arthritis: results of a randomized, placebo-controlled, phase II study. Arthritis Rheum 38:1595, 1995

73. Sharp JT, Strand V, Leung H, et al: Treatment with leflunomide slows radiographic progression of rheumatoid arthritis: results from three randomized controlled trials of leflunomide in patients with active rheumatoid arthritis. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 43:495, 2000

74. Weinblatt ME, Kremer JM, Coblyn JS, et al: Pharmacokinetics, safety, and efficacy of combination treatment with methotrexate and leflunomide in patients with active rheumatoid arthritis. Arthritis Rheum 42:1322, 1999

75. Siva C, Eisen SA, Shepherd R, et al: Leflunomide use during the first 33 months after food and drug administration approval: experience with a national cohort of 3,325 patients. Arthritis Rheum 49:745, 2003

76. Conaghan PG, Brooks P: Disease-modifying antirheumatic drugs, including methotrexate, sulfasalazine, gold, antimalarials, and D-penicillamine. Curr Opin Rheumatol 5:276, 1993

77. Rains CP, Noble S, Faulds D: Sulfasalazine: a review of its pharmacological properties and therapeutic efficacy in the treatment of rheumatoid arthritis. Drugs 50:137, 1995

78. McEntegart A, Porter D, Capell HA, et al: Sulfasalazine has a better efficacy/toxicity profile than auranofin—evidence from a 5 year prospective, randomized trial. J Rheumatol 23:1887, 1996

79. Paulus HE: History of combination therapy of rheumatoid arthritis. J Rheumatol Suppl 44:38, 1996

80. Willkens RF, Sharp JT, Stablein D, et al: Comparison of azathioprine, methotrexate, and the combination of the two in the treatment of rheumatoid arthritis: a forty-eight-week controlled clinical trial with radiologic outcome assessment. Arthritis Rheum 38:1799, 1995

81. O'Dell JR, Haire CE, Erikson N, et al: Treatment of rheumatoid arthritis with methotrexate alone, sulfasalazine and hydroxychloroquine, or a combination of all three medications. N Engl J Med 334:1287, 1996

82. Moreland LW, Baumgartner SW, Schiff MH: Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75) fusion protein. N Engl J Med 337:141, 1997

83. Weinblatt ME, Kremer JM, Bankhurst AD, et al: A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. N Engl J Med 340:253, 1999

84. Maini R, St Clair EW, Breedveld F, et al: Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. Lancet 354:1932, 1999

85. Keane J, Gershon S, Wise RP, et al: Tuberculosis associated with infliximab, a tumor necrosis factor alpha–neutralizing agent. N Engl J Med 345:1098, 2001

86. Mohan N, Edwards ET, Cupps TR, et al: Demyelination occurring during anti–tumor necrosis factor alpha therapy for inflammatory arthritides. Arthritis Rheum 44:2862, 2001

87. Bresnihan B, Alvaro-Gracia JM, Cobby M, et al: Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. Arthritis Rheum 41:2196, 1998

88. Cohen S, Hurd E, Cush J, et al: Treatment of rheumatoid arthritis with anakinra, a recombinant human interleukin-1 receptor antagonist, in combination with methotrexate: results of a twenty-four-week, multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 46:614, 2002

89. Smeets TJ, Kraan MC, Versendaal J, et al: Analysis of serial synovial biopsies in patients with rheumatoid arthritis: description of a control group without clinical improvement after treatment with interleukin 10 or placebo. J Rheumatol 26:2089, 1999

90. Kirwan JR: The effect of glucocorticoids on joint destruction in rheumatoid arthritis. The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. N Engl J Med 333:142, 1995

91. American College of Rheumatology Ad Hoc Committee on Glucocorticoid-Induced Osteoporosis: Recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis: 2001 update. Arthritis Rheum 44:1496, 2001

92. Stein CM, Pincus T, Yocum D, et al: Combination treatment of severe rheumatoid arthritis with cyclosporine and methotrexate for forty-eight weeks: an open-label extension study. The Methotrexate-Cyclosporine Combination Study Group. Arthritis Rheum 40:1843, 1997

93. Stone M, Fortin PR, Pacheco-Tena C, et al: Should tetracycline treatment be used more extensively for rheumatoid arthritis? Metaanalysis demonstrates clinical benefit with reduction in disease activity. J Rheumatol 30:2112, 2003

94. Moreland LW, Pratt PW, Mayes MD, et al: Double-blind, placebo-controlled multicenter trial using chimeric monoclonal anti-CD4 antibody, cM-T412, in rheumatoid arthritis patients receiving concomitant methotrexate. Arthritis Rheum 38:1581, 1995

95. Kremer JM, Lawrence DA, Petrillo GF, et al: Effects of high-dose fish oil on rheumatoid arthritis after stopping nonsteroidal antiinflammatory drugs: clinical and immune correlates. Arthritis Rheum 38:1107, 1995

96. Arnett FC, Edworthy SM, Bloch DA, et al: The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31:315, 1988

III SERONEGATIVE SPONDYLOARTHRITIS

Frank C. Arnett, m.d., f.a.c.p.

Definition

The term spondyloarthritis encompasses a family of clinically, epidemiologically, and genetically related inflammatory diseases that primarily affect spinal and peripheral joints. Once considered a variant of rheumatoid arthritis, spondyloarthritis has been shown to differ in such fundamental clinical and pathogenetic ways from rheumatoid disease that the two are now considered distinctly separate entities [*see Table 1*]. The term seronegative refers to the uniform absence of serum IgM autoantibodies to IgG (rheumatoid factor) in patients with spondyloarthritis. Other distinguishing characteristics are the following:

1. The sacroiliac joints are affected (sacroiliitis); ascending spinal inflammation and bony fusion (spondylitis) often develop after sacroiliitis.
2. Peripheral joints are affected, typically in an oligoarticular and asymmetrical pattern.
3. There is inflammation of sites of ligamentous insertions into bone (entheses), referred to as enthesitis or enthesopathy, as well as inflammation of joint synovium. Inflammation occurs both along the spine and near peripheral joints.
4. There may be inflammation of extra-articular sites, including the eye, the aortic valve, the gastrointestinal tract, the genitourinary system, and the skin.
5. Disease onset typically occurs in young adulthood.
6. There is a strong familial tendency and a striking genetic association with the histocompatibility antigen HLA-B27.
7. Certain bacteria play important pathogenetic roles.

Classification

Spondyloarthritis includes the prototypical spinal arthritis, ankylosing spondylitis; reactive arthritis (formerly known as Reiter syndrome); psoriatic arthritis; enteropathic arthritis (accompanying ulcerative colitis and Crohn disease); juvenile spondyloarthritis (or juvenile ankylosing spondylitis); and such rare disorders as acne-associated arthritis, or SAPHO (synovitis, acne, pustulosis, hyperostosis, osteitis) syndrome, and Whipple disease. The various forms of spondyloarthritis can usually be distinguished from one another by the pattern of joint involvement and associated extra-articular features [*see Table 2*]. However, some patients have overlapping clinical manifestations that defy categorization; these patients are usually designated as having undifferentiated spondyloarthritis. Because of such patients, the European Spondyloarthropathy Study Group (ESSG) proposed classification criteria for spondyloarthritis that may be useful in clinical diagnosis and epidemiologic studies.[1]

Epidemiology

Estimations of prevalence rates for spondyloarthritis using the ESSG criteria are few.[1-3] Among Germans in Berlin, the prevalence of spondyloarthritis has been reported to be 1.9%; in Inuits in Alaska and Siberia, rates of 2% to 3.4% have been reported. Spondyloarthritis appears to be rare in African and Japanese populations. These differences among ethnic groups are explainable in large part by differences in the frequency of HLA-B27.[1]

Pathogenesis

The various forms of spondyloarthritis appear to be complex disorders resulting from the interplay of several genetic and environmental factors, only a few of which have been identified.

genetic factors

Heredity plays a major role in predisposition.[4,5] Family studies have shown that 15% to 20% of patients with ankylosing spondylitis have one or more first-degree relatives with the same disease. In the families of some patients with ankylosing spondylitis, there are relatives with other types of spondyloarthritis or other associated disorders, such as uveitis, psoriasis, and inflammatory bowel disease. Concordance for ankylosing spondylitis in monozygotic twins approaches 63% to 75%, compared with 13% to 23% in dizygotic twins. Genetic modeling in twins and families indicates that ankylosing spondylitis is associated with a multiplicative, polygenic pattern of inheritance, with 97% of the susceptibility to the disease attributed to genetics. These studies suggest that the environmental factors that contribute to development of the disease are probably ubiquitous.

The HLA-B27 allele encoded by the class I HLA-B locus within the major histocompatibility complex (MHC) is the one genetic factor identified thus far that is strongly associated with spondyloarthritis. This allele is present in 90% of patients with ankylosing spondylitis and confers a relative risk for the disease of over 100, but it is found less often in patients with other forms of spondyloarthritis [*see Table 2*].[5] HLA-B27 shows linkage to ankylosing spondylitis in families and appears to contribute 30% to 50% of the genetic risk; in most cases, it appears essential for disease expression.[4] Other HLA alleles, including HLA-B60, HLA-DR1, and HLA-DR8, also appear to increase the risk of ankylosing spondylitis. In addition, different HLA alleles predispose to psoriasis and psoriatic arthritis, including HLA-B13, HLA-B17, HLA-Cw6, HLA-B38, and HLA-B39.[6] The HLA region shows genetic linkage to inflammatory bowel diseases, but specific HLA alleles show only weak associations.[7] Ongoing human genome searches have revealed additional non-HLA loci linked to ankylosing spondylitis,[4,5] some of which also may be common to Crohn disease[8,9] and psoriasis.[5,10]

Laboratory evidence strongly suggests that the *HLA-B27* gene itself, rather than a linked locus, directly participates in the pathogenesis of ankylosing spondylitis and reactive arthritis. Transgenic rats expressing the human *HLA-B27* and β_2-microglobulin genes spontaneously develop colitis, peripheral and spinal arthritis, enthesitis, skin and nail lesions resembling psoriasis, and genitourinary inflammation.[11] Littermates raised in a germ-free environment do not develop most of these manifestations, however. That finding emphasizes the importance of both the *HLA-B27* gene and gut bacteria, possibly *Bacteroides* species, in pathogenesis and suggests that antibiotics (e.g., sulfasalazine) may be useful in the treatment of reactive arthritis and ankylosing spondylitis in humans. The mechanism by which HLA-B27 promotes disease is unknown, but the following are the prevailing hypotheses[12]: (1) in its function as an MHC class I molecule,

HLA-B27 presents a so-called arthritogenic self-peptide or bacterial peptide to cytotoxic $CD8^+$ T cells, which causes an autoimmune attack on various self-structures; (2) HLA-B27 contains stretches of amino acid sequences that also occur in bacterial proteins, and as a result of this molecular mimicry, a cytotoxic or humoral immune response to these bacterial sequences also involves HLA-B27; (3) HLA-B27, either intracellularly or extracellularly, promotes bacterial persistence or dissemination to joints and other structures; and (4) HLA-B27 is unique among HLA class I molecules in forming so-called homodimers, which may cause misfolding of the protein, resulting in an inflammatory response.[5,13,14]

ENVIRONMENTAL FACTORS

Reactive arthritis provides the strongest evidence of bacterial pathogenesis in spondyloarthritis. Enteric infections by *Shigella flexneri, Salmonella* (many species), *Yersinia enterocolitica, Y. pseudotuberculosis*, and *Campylobacter jejuni* have all been implicated as triggers of the disease in various epidemics and in sporadic cases, especially in HLA-B27–positive persons.[15,16] Similarly, sexually acquired infections with *Chlamydia trachomatis*[15,16] and perhaps *Ureaplasma urealyticum* may cause reactive arthritis.[17] Pulmonary infection with *Chlamydophila pneumoniae* (formerly known as *Chlamydia pneumoniae*) has also been implicated.[18] Patients with chronic reactive arthritis have been found to have IgA antibodies to the initiating microbe, suggesting a persistent mucosal infection.[19] Moreover, synovial fluid T cells were found to proliferate when challenged with the bacterium that triggered the arthritis.[12] There is no evidence, however, that these microorganisms cause ankylosing spondylitis. Normal gut flora seem more likely to be implicated in ankylosing spondylitis, as suggested by studies of the HLA-B27 transgenic rat[11] and by a high frequency of asymptomatic foci of gut inflammation in patients with ankylosing spondylitis or reactive arthritis.[20]

Pathology

Chronic inflammation with infiltrating mononuclear cells (macrophages, T cells, and B cells) occurs in both peripheral and axial joint structures of patients with spondyloarthritis.[12,21,22] $CD4^+$ helper T cells and $CD8^+$ suppressor-cytotoxic T cells appear to be equally represented. A high concentration of the inflammatory cytokine tumor necrosis factor–α (TNF-α) has been found in the dense cellular infiltrates in synovial portions of sacroiliac joints.[21] When cytokines from the joints and blood of patients with spondyloarthritis were compared with those of patients with rheumatoid arthritis, the cytokines from patients with spondyloarthritis showed a higher ratio of immunosuppressive cytokines, such as interleukin-4 (IL-4) and IL-10, to inflammatory cytokines, such as TNF-α and interferon gamma. This leads to a blunted T helper type 1 (Th1) response in patients with spondyloarthritis.[22] Inherent levels of cytokines, such as TNF-α and IL-10, are determined by genetic polymorphisms in their respective genes.[22] In ankylosing spondylitis, the observed tendency for ligamentous ossification, enthesopathy, and widespread new bone formation is associated with the finding of transforming growth

Table 1 Comparison between Spondyloarthritis and Rheumatoid Arthritis

	Spondyloarthropathies	*Rheumatoid Arthritis*
Distribution	Racial (more prevalent in whites)	Worldwide
Prevalence	0.2%–1.9%	1%–2%
Etiology	Genetic and bacterial	Genetic and unknown
Positive family history	Frequent	Rare
Sex distribution	More frequently diagnosed in males	More common in females
Age at onset	Peak incidence at 20–30 yr	All ages affected; peak incidence 30–50 yr
Joint involvement	Oligoarthritis; asymmetrical; large joints; lower limbs more than upper limbs	Polyarthritis; symmetrical; small and large joints; upper and lower limbs
Sacroiliac involvement	Yes	No
Spinal involvement	Ascending; all segments with fusion	Cervical only; erosions and instability
Subcutaneous nodules	No	Yes
Aortic regurgitation	Yes	No
Ocular involvement	Uveitis, conjunctivitis	Sicca syndrome; scleritis; scleromalacia perforans
Lung involvement	Upper lobe pulmonary fibrosis	Pleural effusions; lower lobe pulmonary fibrosis; nodules; Caplan syndrome
Rheumatoid factor and/or anti-CCP	No	Yes
HLA-B27	Yes	No (normal frequency)
HLA-DR4	25% (normal frequency)	60%–70%
Pathology	Synovitis and enthesopathy	Synovitis
Radiographic findings	Asymmetrical erosive arthritis and periostitis; new bone formation and ankylosis; sacroiliitis, spondylitis	Symmetrical erosive arthritis with bony destruction

anti-CCP—antibodies to cyclic citrullinated peptides

Table 2 Features of Seronegative Spondyloarthritis

	Ankylosing Spondylitis	*Reactive Arthritis (Reiter Syndrome)*	*Psoriatic Arthritis*	*Enteropathic Arthritis*
Sex distribution	Male > female	Male > female	Female > male	Female = male
Age at onset (years)	≥ 20	≥ 20	Any age	Any age
Mode of onset	Gradual	Sudden	Gradual	Peripheral sudden Spinal gradual
Peripheral joints	Often lower limbs Asymmetrical	Usually lower limbs Asymmetrical	Upper > lower limbs Asymmetrical	Lower > upper limbs Symmetrical
Enthesopathy	+	+	+	– Peripheral + Spinal
Heel pain	Occasional	Frequent	Occasional	Infrequent
Spinal involvement	+++ (always)	+ (20%)	+ (20%)	+ (10%)
Symmetry (sacroiliitis and syndesmophytes)	+	+/–	+/–	+
Familial aggregation	++	+	++	++
HLA-B27 positive	90%	63%–75%	20% (50% with sacroiliitis)	10% (50% with sacroiliitis)
Risk for B27-positive person	2% (20% when a relative)	20% (when infected)	?	?
Urethritis	–	+	–	–
Skin involvement	–	+	+++	+
Nail involvement	–	+	+++	–
Mucous membrane involvement	–	++	–	+
Cardiac involvement	+	+	Rare	Rare
Self-limiting	–	+	–	++ Peripheral – Spinal

factor–β (TGF-β) near these sites. TGF-β is a reparative cytokine that stimulates connective tissue matrix formation.

Reactive arthritis was once considered a sterile joint disease triggered in some unknown manner by a distant infection, but more recent studies of synovial fluids and tissues affected by reactive arthritis have consistently revealed the presence of intracellular bacterial antigens from each of the known offending microorganisms.[14,23,24] Moreover, with electron microscopy and polymerase chain reaction, living but dormant *C. trachomatis* has been detected in synovial macrophages and fibroblasts even after many years of disease.[23] It is still unclear whether the enteric pathogens causing reactive arthritis are viable.[14,24] Spinal joint tissue from patients with ankylosing spondylitis is difficult to obtain. Limited studies of sacroiliac joint biopsies have not revealed bacterial antigens.[25]

Ankylosing Spondylitis

EPIDEMIOLOGY

The prevalence of ankylosing spondylitis parallels the frequency of HLA-B27 in different ethnic populations.[1,26] HLA-B27 occurs in 7% to 9% of the white population, and the disease has a prevalence of approximately 0.2% to 0.9%.[2,3,26] One study from Norway, where the frequency of HLA-B27 is twice that seen in the white populations of the United States and the United Kingdom, found that ankylosing spondylitis occurred in 1.9% to 2.2% of men and 0.3% to 0.6% of women.[3] The disease is distinctly rare in African and Japanese populations, in which HLA-B27 is found in low frequency; however, ankylosing spondylitis is common in certain Native-American groups, such as the Haida and Pima, in which the frequency of HLA-B27 is high.[26]

In randomly chosen cohorts of whites possessing HLA-B27, ankylosing spondylitis developed in approximately 2% to 6%.[3] In HLA-B27–positive relatives of patients with ankylosing spondylitis, however, the risk of disease is 20% to 30%. Similar estimates are not available for other ethnic groups, but rates may differ because multiple molecular subtypes of HLA-B27 have been discovered, each with different distributions among various ethnic groups.[5,25-27] HLA-B*2705, followed in frequency by HLA-B*2702, is predominantly found in whites; HLA-B*2704 is found in Chinese; and HLA-B*2703 is found in Africans. Most HLA-B27 subtypes appear to predispose to ankylosing spondylitis, with the possible exceptions of HLA-B*2706, found in Indonesians and Thais, and HLA-B*2709, found in Sardinians.

Ankylosing spondylitis was once considered to be almost exclusively a disease of males, but recent studies suggest a more uniform distribution by sex (the ratio of males to females is 3:1).[1] In females, the disease may be diagnosed less frequently and later in the course of disease because physicians still consider it primarily a disorder of males. Some studies suggest that females have milder disease, with less progressive spinal involvement and more peripheral arthritis. Other studies suggest that the overall pattern of disease is similar in males and females.

Onset typically occurs between 16 and 30 years of age, peaking at around 24 years; ankylosing spondylitis seldom begins in patients older than 40 years. Childhood onset before 16 years

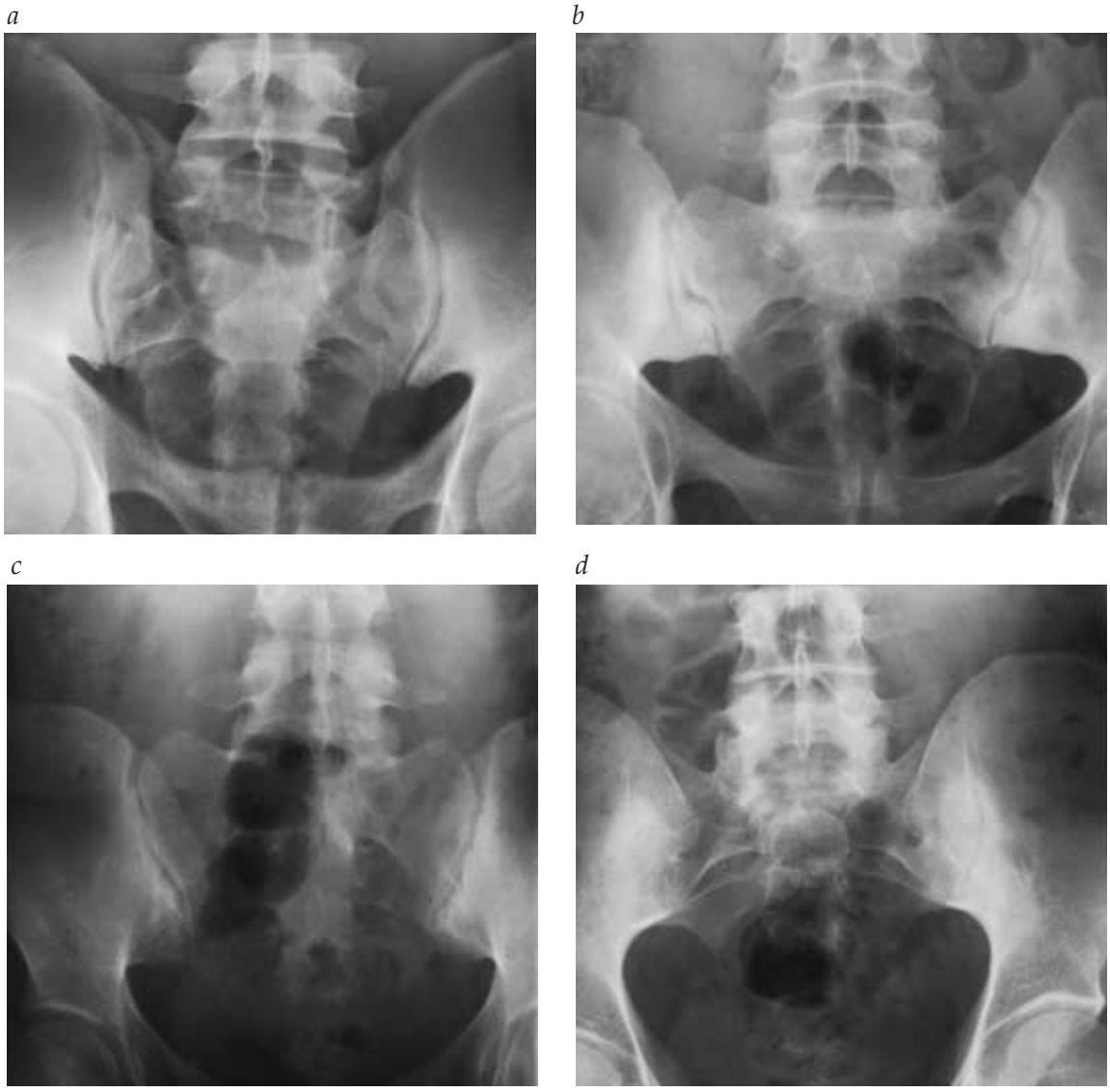


Figure 1 **(*a*) Radiograph of normal sacroiliac joints showing clearly defined joint margins and no sclerosis (grade 0). (*b*) Sclerosis on both margins of each sacroiliac joint but no joint erosions (grade II). (*c*) Sclerosis and erosions of both sacroiliac joints (grade III). (*d*) Complete bony fusion of both sacroiliac joints (grade IV).**

of age occurs in approximately 10% to 20% of cases in the United States and Europe but is more common (54%) in developing countries, suggesting earlier exposure to the environmental triggers.[28]

DIAGNOSIS

The modified New York criteria[1] are currently used to diagnose ankylosing spondylitis. A patient should have one or more of the following clinical criteria:

1. Low back pain of at least 3 months' duration that is alleviated by exercise and is not relieved by rest.
2. Restricted lumbar spinal motion.
3. Decreased chest expansion relative to normal values for age and sex.

In addition, the patient must have definitive radiographic evidence of sacroiliitis (i.e., bilateral sacroiliitis of grade II to IV or unilateral sacroiliitis of grade III or IV) [*see Figure 1*].

A simpler approach in diagnosis is to accept symptomatic sacroiliitis as an adequate definition. Sacroiliitis, as defined radiographically, should be definitive (i.e., grade III or IV changes should be evident) and should be present bilaterally [*see Figure 1*]. In addition, the patient should have no other diseases that could cause sacroiliitis (i.e., reactive arthritis, psoriasis, or inflammatory bowel disease).

Clinical Presentation

Low back pain and stiffness are the usual presenting symptoms of ankylosing spondylitis. Because back pain is such a common complaint in the general population and its causes are myriad, certain characteristics that specifically suggest inflammatory back pain have been formulated:

1. Onset in a person younger than 40 years.
2. Insidious rather than abrupt onset.
3. Persistence of back symptoms for 3 months or longer.
4. Worsening of back pain or stiffness with inactivity.
5. Subsiding of back pain or stiffness with exercise.

Some patients describe buttock pain that often alternates from one side to the other and sometimes radiates down the posterior leg, which is indicative of sacroiliac joint disease. Other patients present with a peripheral arthritis, typically monoarticular or oligoarticular, that affects joints of the lower extremity, often the knee. Careful questioning about subtle musculoskeletal symptoms in such patients is often fruitful. Fatigue can be a major symptom in patients with ankylosing spondylitis and has been found to correlate with level of disease activity, functional ability, global well-being, and mental health status.[29] Elicitation of a history of uveitis or the presence of spondyloarthritic features in family members also strongly suggests the disease. Radiologic evidence of sacroiliitis in any of these clinical presentations, however, is essential in confirming a diagnosis of ankylosing spondylitis [*see* Radiographic Features, *below*]. In patients whose sacroiliac radiographs are normal, the presence of HLA-B27 is highly suggestive but not definitive evidence of the disease. Follow-up studies of patients in whom the diagnosis was strongly suspected on the basis of the clinical picture and HLA-B27 positivity showed that sacroiliac joint abnormalities eventually appear on plain x-rays, but the evolution may occur over as many as 10 years. MRI of the sacroiliac joints is a very sensitive method for detecting early sacroiliitis, as well as inflammation elsewhere in the spine.

Patients with juvenile-onset ankylosing spondylitis typically present with peripheral oligoarthritis, often with enthesopathy and infrequently with spinal symptoms.[28] Such patients may be misdiagnosed as having juvenile rheumatoid arthritis [*see* Juvenile Spondyloarthritis, *below*]. Spinal involvement usually appears later, in young adulthood.

Physical Examination

Examination of the back may be relatively normal early in the course of the disease. Sacroiliac joints are usually painful when palpated or stressed. When the disease advances into the lumbar spine, the normal lordotic curvature may be lost, and paravertebral muscle spasm is prominent. Forward bending, or flexion, may be restricted, as measured by the Schober test. In this measurement, two points are drawn with the patient standing erect, one at the L5–S1 region and the other 10 cm above this region. With normal flexion, the distance between these two points increases by 4 to 6 cm, but when the lumbar spine becomes fused, there may be little or no increase in distance between the two points. Lateral lumbar bending and extension are also typically restricted.

Thoracic spine involvement causes an exaggerated dorsal kyphosis; in patients with costovertebral joint fusion, chest expansion (as measured circumferentially at the fourth intercostal space from full expiration to inspiration) is reduced to 2.5 cm or less. When the disease ascends into the neck and causes fusion, cervical lordosis is lost and a fixed flexion deformity may occur. Spinal fusion often results in the patient's being severely stooped forward with neck immobile and flexed; the patient has difficulty looking straight ahead.

Peripheral arthritis, especially of the hips, shoulders, and knees, occurs in approximately 30% of patients with ankylosing spondylitis and further increases disability. Peripheral enthesopathic features may include Achilles tendinitis, plantar fasciitis, or costochondritis.[30]

Laboratory Findings

The HLA-B27 histocompatibility antigen is present in more than 90% of ankylosing spondylitis patients. HLA-B27 testing of individual patients, however, is indicated only in atypical cases, when the clinical suspicion is high but the most definitive finding—radiographic evidence of sacroiliitis—is not present. In HLA-B27–positive patients and HLA-B27–negative patients, the patterns and severity of arthritis are similar. HLA-B27–negative patients differ from HLA-B27–positive patients in that in HLA-B27–negative patients, disease onset occurs at an older age, there is no family history of spondylitis, and uveitis or cardiac complications occur infrequently.[1,26] HLA-B27 is found less commonly (50%) in patients with ankylosing spondylitis who are of African ancestry.

Elevation of the erythrocyte sedimentation rate (ESR) occurs in many patients, but it may be normal despite severe disease. C-reactive protein (CRP) levels may be elevated. Serum IgA levels are often elevated.[19] Some patients have a mild normocytic normochromic anemia because of chronic inflammation; in these patients, the platelet count may be high.

Radiographic Features

Bilateral sacroiliitis is the most specific finding that supports a diagnosis of ankylosing spondylitis, and meticulous interpretation of the radiographs is imperative. A grading system that assesses each sacroiliac joint for juxta-articular bony sclerosis, blurring or erosion of joint margins, and bony fusion has been formulated and tested [*see Figure 1*]. Grade 0 findings are normal. Grade I findings are suspicious but not definitive. Grade II findings show sclerosis on both sides of a joint; such findings are even more suspicious when they occur bilaterally, but they should be interpreted with great caution. Findings of grades III and IV are definitive. Another radiographically defined entity, osteitis condensans ilii, may be misinterpreted as sacroiliitis, and vice versa. Patients with osteitis condensans ilii have sclerosis on the iliac side of both sacroiliac joints; the condition occurs in women who have borne children. Although quantitative radionuclide scans, computed tomography, and MRI have been suggested as superior diagnostic methods, well-performed plain radiographs of the sacroiliac joints (Ferguson view or oblique view) are usually adequate.

An early spinal change seen on radiographs is squaring of the normally concave anterior side of vertebral bodies [*see Figure 2*]. This phenomenon is caused by inflammation and bony erosion at the site of insertion (enthesitis) of the outer fibers of the annulus fibrosus. Later changes are ossification of ligaments, which are seen on radiographs as syndesmophytes that bridge adjacent vertebral bodies [*see Figure 3*], producing the classic bamboo-spine appearance [*see Figure 4*]. Zygapophyseal joints become fused into solid bone. Finally, diffuse osteoporosis may occur, making the spine susceptible to fracture. Bony fusion across joint spaces of affected peripheral joints in ankylosing spondylitis may be the most distinctive change seen on radiographs.

Similar spinal changes are seen in primary ankylosing spondylitis and in the spondylitis associated with inflammatory bowel disease. In spondylitis associated with reactive arthritis and psoriatic arthritis, the sacroiliitis and syndesmophytes tend to be asymmetrical.[31] Another disease that may mimic ankylosing spondylitis is diffuse idiopathic skeletal hyperostosis (DISH).[32] DISH occurs in middle-aged and older persons, especially men; it is characterized by large, flowing syndesmophytes that restrict spinal motion; sacroiliitis is not found, however, and there is no association with HLA-B27.

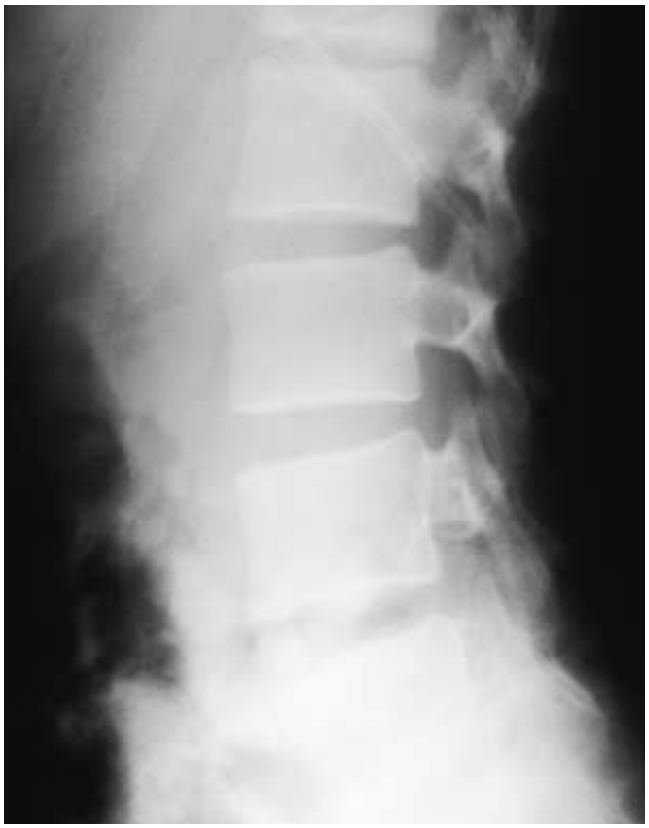


Figure 2 **Loss of the normal anterior concavity of vertebral bodies, resulting in so-called squaring in early ankylosing spondylitis.**

Extra-articular Manifestations

A number of extraskeletal features may complicate the course of ankylosing spondylitis and contribute to morbidity and mortality.

Ocular involvement Acute anterior uveitis, usually occurring episodically and affecting one eye at a time, occurs in 25% of patients. Acute pain, redness, and photophobia are the usual symptoms. Prompt referral to an ophthalmologist for treatment is essential. Uveitis does not correlate with arthritis activity or severity and shows a strong association with HLA-B27, even in patients without spondyloarthritis.[33]

Cardiovascular disease A fibrosing cardiovascular lesion occurs in 2% to 10% of patients with ankylosing spondylitis. The lesion causes the aortic valve and proximal root to thicken, and it often extends into the conducting system, causing aortic regurgitation, atrioventricular block, or both. The lesion probably occurs with a similar frequency in patients with reactive arthritis.[34] In rare instances, mitral regurgitation may also occur. One study emphasized a high prevalence of underlying spondyloarthritis, often undiagnosed, in men requiring cardiac pacemakers for bradyarrhythmias.[34] In addition, this study revealed the strong association of the clinical combination of lone aortic regurgitation and heart block with HLA-B27, with or without apparent arthritis. Fulminant cardiac disease typically appears only after the patient has had spondyloarthritis for many years, but such disease has been described even in very early spondyloarthritis. Echocardiography may detect cardiac abnormalities in some patients without clinical signs.[35] No treatment is known to prevent progression of spondylitic heart disease; most patients require permanent cardiac pacemakers, aortic valve replacement, or both.

Pulmonary disease Despite restriction of chest wall motion by joint fusion, respiratory function is preserved in most patients with ankylosing spondylitis, owing to good diaphragmatic function. Severe kyphotic deformity, however, may compromise breathing. Approximately 1% of patients with ankylosing spondylitis, usually those with severe disease, also have fibrosis in the upper lung fields that mimics tuberculosis.[36] Cavitation may occur and may be complicated by *Aspergillus* infection. Cough, dyspnea, and even hemoptysis are typical symptoms. Currently available treatment is unsatisfactory.

Renal disease Kidney function is usually normal. The appearance of proteinuria, with or without a nephrotic syndrome, usually indicates complicating amyloidosis or IgA nephropathy. Secondary amyloidosis occurs in approximately 4% of patients and can be diagnosed with abdominal fat-pad or rectal biopsy.[37]

IgA nephropathy is being increasingly recognized. It correlates with high serum IgA levels. Renal function may become impaired, but episodes are usually self-limited.[19]

Neurologic disease Spinal fracture is a major cause of morbidity and mortality in patients with ankylosing spondylitis; cord compression occurs even with seemingly minor trauma.[1] A rigid and osteoporotic cervical spine is most susceptible to fracture, usually at the C6 or C7 level. A high degree of suspicion for fracture is always warranted in patients with localized spinal pain, even when plain x-rays fail to reveal an acute abnormality;

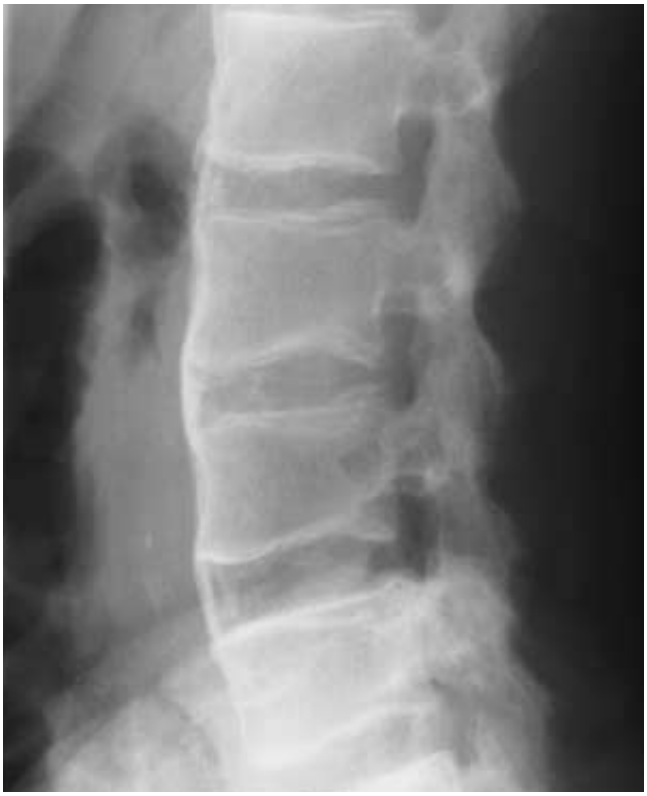


Figure 3 **Progression of ankylosing spondylitis is demonstrated by ossification of the anterior fibers of the annulus fibrosus (syndesmophytes).**

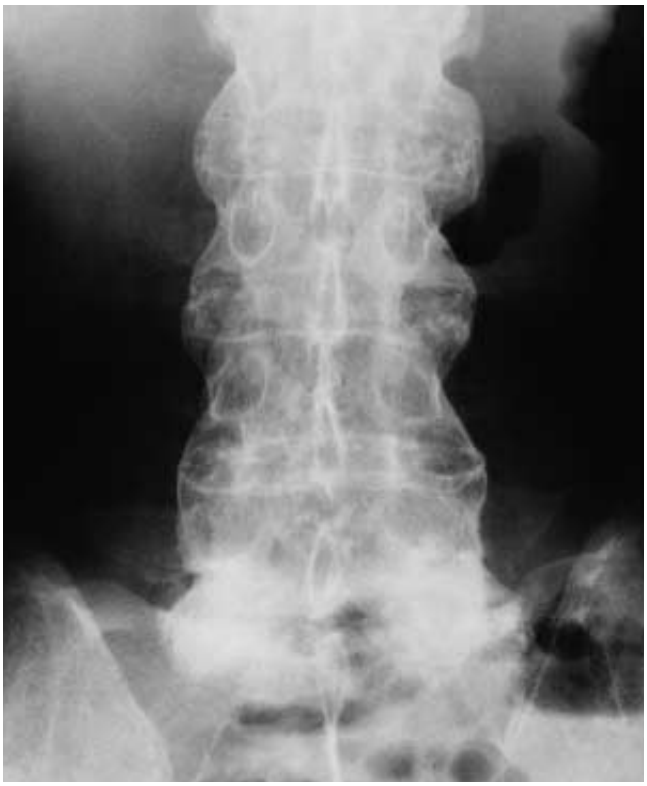


Figure 4 **Patients with severe ankylosing spondylitis may develop the classic bamboo spine, as shown in this radiograph.**

additional imaging with CT is often necessary.

A cauda equina syndrome occurs in rare instances, usually because of arachnoiditis around sacral nerves that leads to progressive leg weakness, paresthesias, and sphincter dysfunction.[38]

Retroperitoneal fibrosis Fibrosis in the retroperitoneum may be another extra-articular feature of ankylosing spondylitis.[39]

TREATMENT AND PROGNOSIS

Early diagnosis and treatment of ankylosing spondylitis appear to improve functional outcome, but it is not clear whether any drug modifies the disease pathology. Objectives of treatment are pain relief, reduction of inflammation, and maintenance of good posture and spinal function.[40] Patient education is very important. Excellent sources for patient education are available at www.spondylitis.org and www.arthritis.org.

Nonsteroidal anti-inflammatory drugs (NSAIDs) relieve inflammatory symptoms of pain and stiffness and allow patients to engage in an appropriate exercise program. In clinical practice, certain NSAIDs appear to be more often effective than others as treatment for spondyloarthritis [*see Table 3*]. However, the efficacy of individual agents varies greatly from patient to patient; for that reason, some patients may need to try several NSAIDs before finding one that provides relief. There is no strong evidence that any NSAIDs alter disease progression.

There are now incontrovertible data that selective cyclooxygenase-2 (COX-2) inhibitors, and possibly other NSAIDs, increase a person's risk for cardiovascular disease. Therefore, careful assessment of other risk factors and of the potential benefit compared with risk is essential in all patients. Gastrointestinal intolerance of any of the NSAIDs may present as nausea, gastric discomfort, diarrhea, or, more seriously, gut hemorrhage or perforation. Concomitant use of a gastroprotective agent, such as misoprostol or a proton pump inhibitor, may significantly reduce GI toxicity in patients treated with NSAIDs. All of these drugs may decrease renal tubular capacity to secrete potassium and can cause an abrupt reduction in renal function when used in patients with renal disease or with renal hypoperfusion resulting from ineffective circulatory volume. Because patients with ankylosing spondylitis will probably take NSAIDs for many years, physicians must diligently monitor for renal and GI tract damage.

In a 6-month randomized, controlled clinical trial, the bisphosphonate pamidronate, given monthly by intravenous infusion, was shown to be effective in improving symptoms and function in patients with ankylosing spondylitis whose disease was refractory to treatment with NSAIDs.[41]

Low-dose corticosteroids (e.g., prednisone, 5 to 10 mg daily) may be necessary to quell inflammation in some patients with highly active disease, but these agents should be used sparingly because they promote osteoporosis and do not improve spinal disease. Injection of repository corticosteroids into affected peripheral joints also may be useful. Injection into the sacroiliac joint, guided by either CT or MRI, may offer relief.

Sulfasalazine, 2 to 3 g daily in two divided doses, has been shown in several placebo-controlled trials to be an effective long-term treatment of ankylosing spondylitis, as well as of other types of spondyloarthritis. Sulfasalazine is very effective for peripheral joint symptoms but not especially effective for axial joint symptoms.[1,40] The drug moiety responsible for the efficacy of sulfasalazine has been proved to be sulfapyridine rather than salicylate; however, it is not clear whether the efficacy results from antimicrobial or other properties of the drug.[42] Because sulfasalazine has been shown to lower acute-phase reactants, such as the ESR and the CRP level, it may modify disease progression; however, this desirable effect has yet to be proved.

Other long-acting agents used to treat rheumatoid arthritis, including gold salts, penicillamine, and hydroxychloroquine, are not effective in ankylosing spondylitis.[40] Methotrexate therapy, which is highly effective for rheumatoid arthritis, is clearly effective in psoriatic arthritis but not in other forms of spondyloarthritis.[6] Administration of radiation therapy to the spine was once used successfully but is no longer recommended because of the risk of subsequent malignancy.

The TNF antagonists etanercept and infliximab have been approved for the treatment of ankylosing spondylitis, as well as psoriatic arthritis. An increasing number of controlled and open-label studies of the use of these agents in each of the forms of spondyloarthritis have shown dramatic and rapid improvement in symptoms; significantly reduced inflammatory changes in the spine and peripheral joints, as evidenced on MRI; and lowered acute-phase reactants such as ESR and CRP. Long-term efficacy and modification of disease progression and outcome have yet to be determined.[5,43-45] Treatment with TNF-α antagonists also has been shown to halt progression of secondary amyloidosis.[46] Because of the high cost of these agents and still-unanswered questions about their long-term safety, guidelines have been developed by international consensus to facilitate the judicious use of TNF antagonists.[5] Many patients with mild disease may never require TNF antagonists.

All patients with ankylosing spondylitis should be informed of potential spinal deformities and how to prevent them. Good posture should be emphasized. A firm mattress and minimal pillow support are recommended. An exercise program of spinal extension and peripheral joint range-of-motion exercises, along

Table 3 Treatment for Spondyloarthritis

Drug	*Dose*	*Efficacy Rating*	*Comments*
Indomethacin*	50 mg t.i.d. or 75 mg SR, q. 12 hr	Effective for symptoms	Side effects: headaches, changes in mentation, peptic ulcers, GI toxicity, intolerance, renal insufficiency
Tolmetin*	600 mg t.i.d.	Effective for symptoms	Side effects: peptic ulcers, GI toxicity, renal insufficiency
Piroxicam*	20 mg q.d.	Effective for symptoms	Side effects: peptic ulcers, GI toxicity, renal insufficiency
Diclofenac*	75 mg b.i.d.	Effective for symptoms	Side effects: peptic ulcers, GI toxicity, renal insufficiency
Sulfasalazine†	1–3 g daily in two divided doses	Long-term efficacy; lowers acute-phase reactants	Side effects: headache, GI intolerance
Methotrexate†	7.5–20 mg weekly	Effective for skin and arthritis in psoriatic arthritis; effectiveness in other diseases unproved	Side effects: GI intolerance, hepatotoxicity, marrow suppression, pulmonary disease
Doxycycline†	100 mg bi.d.	Effective in preventing relapse and in long-term treatment of reactive arthritis only	Side effects: GI intolerance, photosensitivity
Infliximab†	5 mg/kg I.V. every 6–8 wk after loading	Highly effective and immediate response; improved inflammation in joints by MRI; long-term effects unknown	Side effects: allergic reactions; increased susceptibility to infection, especially tuberculosis
Etanercept†	25 mg subcutaneous injections twice a week	Same as infliximab	Injection-site reactions, increased risk of infections

*Concomitant treatment with a proton pump inhibitor or misoprostol recommended for gastric protection.
†Potentially disease-modifying agents.
GI—gastrointestinal MRI—magnetic resonance imaging NSAIDs—nonsteroidal anti-inflammatory drugs SR—slow release

with hydrotherapy, should be prescribed. Swimming is a very effective means of achieving exercise goals.[1,5]

Some patients who experience hip involvement—a major cause of disability—greatly benefit from total hip replacement. Wedge osteotomy for severe spinal kyphosis is available only at a few medical centers. Treatment of spinal fractures is controversial. Pregnancy does not appear to be significantly affected by ankylosing spondylitis.

The prognosis for individual patients is often difficult to ascertain.[47] Worse outcomes have been associated primarily with hip joint involvement and, to a lesser extent, early age at onset. The course of the disease in its first 10 years appears to predict its future course and the functional outcome. Despite long-standing and severe disease, ankylosing spondylitis often does not affect a patient's ability to work. Mortality from the disease is infrequent but may result from cardiac or neurologic complications or amyloidosis.

Reactive Arthritis

Reactive arthritis was originally defined as the triad of nongonococcal urethritis, conjunctivitis, and arthritis. It is now recognized that most patients present with arthritis alone and have no clinical evidence of urethritis or conjunctivitis.[15,16,23,24] The concept of reactive arthritis arose from observations that the disease followed certain enteric infections (such cases are termed epidemic or postenteric) and sexually acquired infections (such cases are termed endemic or postvenereal) [*see* Pathogenesis, Environmental Factors, *above*]. Despite this association with previous infection, affected sites were seemingly sterile when cultured for bacteria. It has been found that bacterial antigens, if not viable microorganisms, are present in the joints of affected patients[15,16,26] [*see* Pathology, *above*]. Like ankylosing spondylitis, reactive arthritis may be complicated by sacroiliitis, spondylitis, uveitis, and cardiac lesions. It is also strongly associated with HLA-B27 [*see Table 2*].

EPIDEMIOLOGY

Reactive arthritis probably has a worldwide distribution, but most epidemiologic and clinical studies have come from Europe and the United States.[15-17] The prevalence of the disease is difficult to ascertain because it changes over time, depending on sexual behavior and the prevalence of enteric pathogens in different populations.[48] It was estimated that from 1950 to 1980 in Rochester, Minnesota, the incidence of reactive arthritis in men younger than 50 years was 0.035%; however, 10- to 20-fold higher rates were reported in homosexual men and in certain Native Americans in whom the frequency of HLA-B27 was high (30% to 40%) and who had endemic exposure to enteric or venereal pathogens.[15]

Reactive arthritis, probably the postvenereal form, is the most common cause of inflammatory arthritis in young men. The disease is recognized in women far less frequently; the reasons for this are unclear, because the ratio of affected men to affected women after epidemics of gastroenteritis is typically 1:1 and, overall, the incidence of reactive arthritis approaches 1% to 2% of persons infected with any of the triggering pathogens.[15,16,26] The incidence appears to have fallen significantly since the HIV epidemic and the adoption of safer sexual practices.[48]

HLA-B27 is found in 63% to 75% of patients with both forms of reactive arthritis and confers a relative risk of approximately 37. Of persons with HLA-B27 who are infected with one of the causative bacteria, reactive arthritis develops in approximately 20%.

DIAGNOSIS

Clinical Presentation

Reactive arthritis typically develops 10 to 30 days after an

episode of gastroenteritis or sexual exposure to a venereal pathogen; however, many patients deny any such antecedent events.[15] Episodes of urethritis or conjunctivitis may have been mild and transient or not perceived at all. Thus, recognition of the pattern of musculoskeletal involvement, as well as several other mucocutaneous manifestations, is important in establishing the correct diagnosis.

The arthritis usually is oligoarticular and asymmetrical and predominantly affects lower-extremity joints, most often the knees, ankles, and feet. Diffuse, painful swelling of entire digits (sausaging or dactylitis) occurs frequently. Pain in the heels from Achilles tendinitis or plantar fasciitis, or both, reflects the most common sites of enthesitis; however, enthesopathic pain at other sites is also frequent.[30] Low back pain is a complaint of 60% of patients, and 20% ultimately experience radiographically detectable sacroiliitis. An ascending spondylitis ensues in approximately 10% to 12% of patients.

One or more of the mucocutaneous features can be found on examination in more than 50% of patients, usually early in the disease. Keratoderma blennorrhagica is a papulosquamous skin rash that usually begins on the soles or palms as painless and nonpruritic excrescences resembling mollusk shells [*see Figure 5*]. With time, these lesions evolve into scaling plaques that may coalesce into a more generalized exfoliative dermatitis. Keratoderma blennorrhagica is clinically and histopathologically the same as the disorder pustular psoriasis [*see 2:III Psoriasis*]. A similar scaling rash on the glans penis in circumcised men is termed circinate balanitis. Moist, shallow ulcers characterize balanitis in uncircumcised men, who may be unaware of the lesions unless the foreskin is retracted [*see Figure 6*]. Similar painless oral ulcers may be found on the tongue or palate. Nails may become hyperkeratotic, thickened, and deformed, but the characteristic nail pitting of psoriasis is usually absent [*see Figure 6*]. It is important to search for all of these lesions; they are frequently asymptomatic but are definitive and can establish a diagnosis.

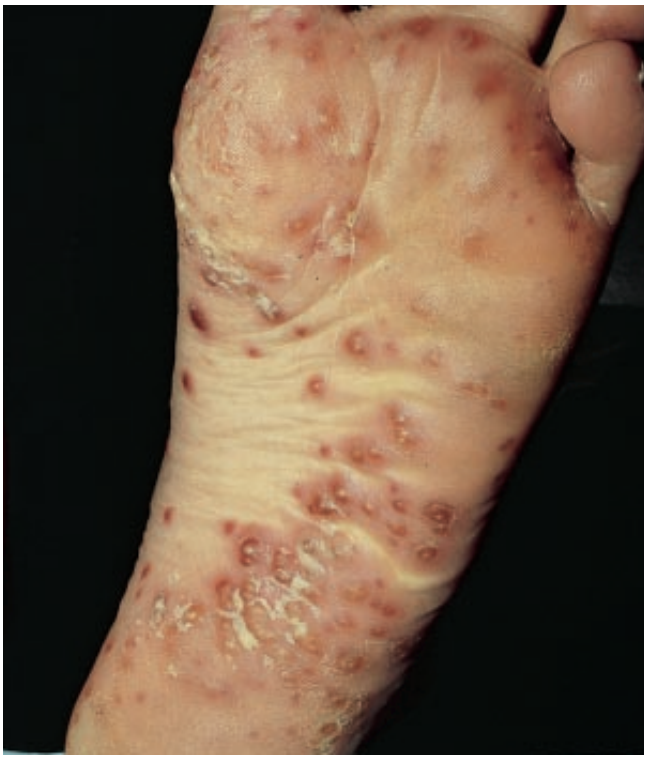


Figure 5 **Typical keratoderma blennorrhagica rash of reactive arthritis on the sole of the foot.**

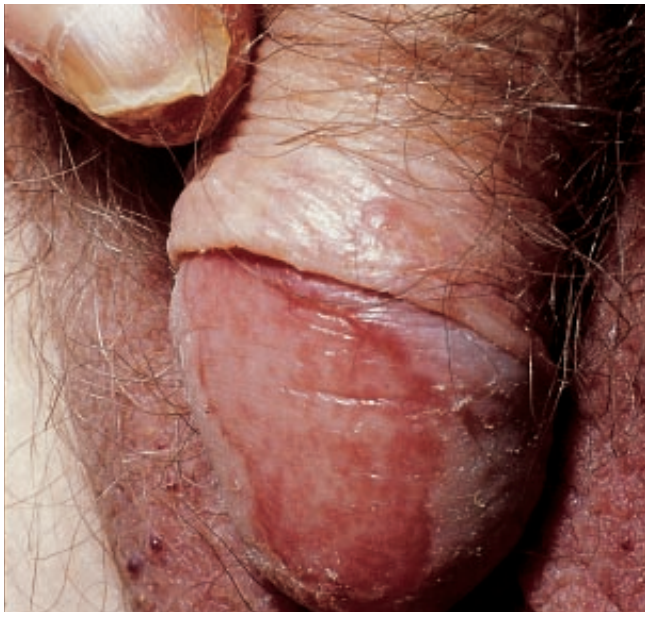


Figure 6 **Superficial penile ulceration of circinate balanitis in an uncircumcised patient with reactive arthritis. Also note dystrophic fingernail.**

Some patients experience low-grade or high fever at disease onset; malaise—or even prostration—and significant weight loss may ensue. Acute anterior uveitis occurs in approximately 20% of patients with reactive arthritis. Cardiac bradyarrhythmia, aortic regurgitation, or both may also occur during the acute disease phase or may appear later in patients whose illness follows a chronic course.[34] Patients with reactive arthritis who are HLA-B27 positive are more likely to experience sacroiliitis and spondylitis, as well as uveitis, cardiac lesions, or both, and to experience a prolonged disease course.

Reactive arthritis has been frequently described in patients with HIV infection; the joint and skin disease may be more severe than usual in such persons.[15,16,26] This association is now believed to result from sexually acquired enteric and venereal pathogens common to both diseases.

Laboratory Evaluation

Tests of patients with reactive arthritis usually show a modest leukocytosis, thrombocytosis, and anemia, along with elevation of the ESR, reflecting systemic inflammation. Examination of the synovial fluid reveals inflammatory changes of poor mucin clot and leukocytosis; but in contrast to septic arthritis, the glucose level is not low, and bacterial cultures are negative. Polymerase chain reaction (PCR) analysis of synovial fluid or tissue biopsies has been used successfully to detect specific bacterial DNA or RNA in research laboratories; PCR kits should become clinically available soon.[49] Cultures or molecular probes for *C. trachomatis* should be obtained in patients with venereal exposure, genitourinary symptoms, or both.[50] At the same time, tests for concomitant gonorrhea, syphilis, and HIV infection should be performed. In patients with preceding GI symptoms, stool cultures for the triggering organisms are usually negative by the time joint symptoms appear. Serologic tests for *Salmonella* and other enteric pathogens are usually unreliable but may be useful in some cases.[51]

Radiographic Features

X-rays are of no diagnostic value early in the disease; however, MRI may show inflammatory changes of enthesitis and arthritis. After several months of persistent joint symptoms, enthesopathic symptoms, or both, radiographs may show the distinctive changes of periostitis and bony ankylosis. Patients with chronic heel pain may show a fluffy periosteal reaction or spur formation at the Achilles or plantar tendon insertions.[30] Similar radiographic changes may be seen along metatarsal or phalangeal bones of the feet; bony fusion across joints may be visible. Sacroiliitis, when present, is more often unilateral than bilateral, and large asymmetrical syndesmophytes may be seen in the lumbar spine.[31]

TREATMENT AND PROGNOSIS

Reactive arthritis runs a self-limited course in most patients, lasting 4 to 12 months, although annoying residual musculoskeletal symptoms may persist for years.[51] From 15% to 30% of patients suffer permanent disability.[40] Relapses are not uncommon; it is unclear whether they result from repeat infection or other endogenous mechanisms. The same NSAIDs used to treat ankylosing spondylitis [*see* Ankylosing Spondylitis, Treatment and Prognosis, *above*] [*see Table 3*] are usually effective in quieting inflammatory joint symptoms. Some patients with highly active disease, however, may require short courses of low-dose systemic corticosteroids or repository corticosteroid injections into joints.

Early treatment of genitourinary infections with appropriate antibiotics (e.g., tetracycline or erythromycin) has been shown to reduce the likelihood of subsequent reactive arthritis; however, even early antibiotic use in patients with gastroenteritis does not appear to prevent reactive arthritis.[15,40] A blinded, placebo-controlled trial of the use of tetracycline for the treatment of reactive arthritis demonstrated that the duration of disease was shortened only in patients who had *Chlamydia*-induced disease.[40] Ciprofloxacin has not been shown to shorten the course of chronic reactive arthritis.[40] Controlled studies have shown that sulfasalazine, in dosages similar to those used in the treatment of ankylosing spondylitis, is effective in all forms of spondyloarthritis.[40] Whether any of these antibiotic approaches change the natural history of the disease remains to be proved. Patients with spondyloarthritis that persists despite treatment with NSAIDs and antibiotics may benefit from the use of anti-TNF agents. An increasing number of studies are documenting immediate and dramatic benefit from the use of TNF-α antagonists (e.g., infliximab and etanercept) in such patients.[43] Physical therapy is important in maintaining joint motion and preventing disability.

Psoriatic Arthritis

EPIDEMIOLOGY

The prevalence of cutaneous psoriasis is estimated to be 2% in most white populations; it appears to be lower in populations who are of African or Asian ancestry.[52] An inflammatory arthropathy attributable to psoriasis appears in 5% to 7% of patients with the skin disease, especially in those whose nails are affected.[6] Psoriasis is highly familial, and there is strong evidence that it is a complex genetic disease associated with several HLA alleles and other non–HLA-linked loci[5,53] [*see* Pathogenesis, *above*]. Genomic studies now strongly suggest major but yet unidentified loci for psoriasis susceptibility near HLA-C in the MHC region and on chromosome 17.[5,10] HLA-B27 is only weakly associated with psoriasis and peripheral psoriatic arthritis, but it occurs in 50% of persons who have psoriatic spondylitis. Potential environmental triggers are streptococcal infection and physical trauma. Psoriatic arthritis is slightly more common in females than in males. Psoriasis frequently first appears in childhood; psoriatic arthritis typically appears in early or middle adulthood, although there are many exceptions. The arthritis may appear before the psoriasis in as many as 40% of children and 15% of adults. Although the incidence of psoriasis and psoriatic arthritis in HIV-positive persons is similar to that in uninfected persons, severe exacerbations of both skin disease and joint disease have been observed in patients with HIV infection, especially as the number of $CD4^+$ T cells declines.

DIAGNOSIS

Clinical Presentation

In general, there is little relation between joint and skin severity. In fact, psoriatic skin lesions may be found only after careful scrutiny of the scalp, the umbilicus, or the gluteal region, and nail pitting or other changes may be the only clues supporting a diagnosis of psoriatic arthritis. Several clinical patterns of joint involvement, often overlapping, have been described:

1. Asymmetrical oligoarthritis of both small and large joints is the most common form of psoriatic arthritis. Involvement of distal interphalangeal joints and sausage-shaped toes or fingers are highly suggestive signs. A disparity is often noted between clinical appearance and subjective symptoms; overtly involved joints may be largely asymptomatic, unlike the concordance usually found in rheumatoid arthritis.
2. Symmetrical polyarthritis may resemble rheumatoid arthritis, although tests for rheumatoid factor and antibodies to cyclic citrullinated peptides (anti-CCP) should be negative. Anti-CCP is a newly discovered autoantibody marker for rheumatoid arthritis that is 65% sensitive and 96% specific. Uncertainty about classification is reasonable because psoriasis and rheumatoid arthritis are both relatively common diseases and are expected to occur together by chance.
3. Arthritis mutilans is the most destructive form of psoriatic arthritis; it occurs in approximately 5% of patients with psoriatic arthritis. Striking bone resorption and telescoping of fingers (opera-glass hand) are characteristic. Affected patients often have concomitant spinal involvement.
4. Psoriatic spondylitis occurs in approximately 20% of patients with psoriatic arthritis, often with unilateral sacroiliitis and large asymmetrical syndesmophytes, similar to the pattern seen in patients with reactive arthritis.
5. Dominant or exclusively distal interphalangeal joint involvement with psoriatic nail changes may occur.

Laboratory Findings

An elevated ESR or CRP level, anemia, and hyperuricemia may be found. Rheumatoid factor, anti-CCP, and antinuclear antibody tests are negative. Synovial fluid shows nonspecific inflammatory changes.

Radiographic Features

A characteristic change is whittling of the distal ends of phalanges, giving the joints a so-called pencil-in-cup appearance, which is radiographically distinctive for psoriatic arthritis. Periostitis—which results in whiskering around joints—bony erosions, and joint fusion in the absence of osteopenia also are common and diagnostically useful findings.

TREATMENT AND PROGNOSIS

NSAIDs similar to those used for ankylosing spondylitis [*see* Ankylosing Spondylitis, Treatment and Prognosis, *above*] [*see Table 3*] are the mainstay of arthritis therapy in most patients but have no effect on the skin disease, which may require separate dermatologic approaches [*see 2:III Psoriasis*]. Sulfasalazine, methotrexate, or cyclosporine may be beneficial for both skin and joint disease in NSAID-resistant or severe, progressive disease.[6] Well-controlled studies have demonstrated that TNF-α antagonists (etanercept and infliximab) are highly effective for symptoms and probably modify outcomes for both the arthritis and the skin disease.[43] Gold, penicillamine, and hydroxychloroquine are not useful agents.

Psoriatic arthritis usually runs a more benign course than rheumatoid arthritis does, although clearly there are many patients with severe disease. Many patients with psoriatic arthritis maintain reasonable function, often despite extensive deformities.

Enteropathic Arthritis

Two major clinical patterns of arthritis associated with inflammatory bowel diseases are peripheral arthritis and spondylitis.

PERIPHERAL ARTHRITIS

Approximately 20% of patients with Crohn disease or ulcerative colitis experience an acute peripheral arthritis.[20] Symmetrical swelling of the knees, ankles, or wrists is the most common articular pattern; large effusions may occur. The pathogenesis of the arthritis is unknown, but the disease occurs during periods of active inflammation of the gut and may be the first sign of a bowel flare-up. HLA-B27 is not increased in frequency among inflammatory bowel disease patients with peripheral arthritis, as compared with the normal population. Extraskeletal and extraintestinal manifestations may occur simultaneously and include fever, acute anterior uveitis, painful oral ulcers, erythema nodosum (in Crohn disease), and pyoderma gangrenosum (in ulcerative colitis). Treatment of the arthritis should be aimed at controlling the inflammatory bowel disease. The arthritis seldom results in deformities.

SPONDYLITIS

Sacroiliitis develops in about 10% of patients with inflammatory bowel disease. Clinically, the spondylitis may progress to total spinal ankylosis; radiographically, it is indistinguishable from ankylosing spondylitis.[20,31] There is no correlation of the spondylitis with activity of the bowel disease. HLA-B27 is found in approximately 50% of such patients. Therapy is largely the same as for ankylosing spondylitis [*see* Ankylosing Spondylitis, Treatment and Prognosis, *above*]. Despite the bowel disease, NSAIDs are usually well tolerated.

UNDIFFERENTIATED SPONDYLOARTHRITIS

Inevitably, the presentations of many patients do not conform to the typical presentations described above, and the symptoms and signs defy specific disease classification.[1,54] Examples are a patient with unilateral sacroiliitis, a sausage digit, and uveitis; a patient with typical reactive arthritis who experiences psoriatic arthritis; and a patient with typical ankylosing spondylitis who years later experiences Crohn disease. Such patients are often designated as having undifferentiated spondyloarthritis. The ESSG criteria[1] now make the classification of patients with spondyloarthritis more definitive. There remains, however, a large number of patients with formes frustes that do not fulfill the new criteria but probably fall within the spectrum of spondyloarthritis. Such entities, which are strongly associated with HLA-B27, are chronic inflammatory back and chest pain syndromes (in which radiographs are normal), chronic dactylitis, chronic plantar fasciitis or Achilles tendinitis, pustular psoriasis (keratoderma blennorrhagica), circinate balanitis, acute anterior uveitis, and spondylitic heart disease without evidence of arthritis. In patients suspected of having a limited form of spondyloarthritis, typing for HLA-B27 may prove clinically useful in supporting such a diagnosis.[1,54]

Juvenile Spondyloarthritis

Until recently, the term juvenile rheumatoid arthritis was used, inappropriately, to describe all forms of chronic childhood arthritis. Careful clinical evaluation, autoantibody testing, and HLA typing have revealed a heterogeneous group of diseases in which only a small proportion of affected children truly have rheumatoid arthritis.

Juvenile spondyloarthritis occurs most often in boys; it typically begins in late childhood or adolescence with lower extremity oligoarthropathy and enthesopathy.[28] Spinal symptoms are rare initially but often appear years later. Bony ankylosis of the tarsal bones has been described in some of these patients. Acute anterior uveitis is not uncommon. Such patients are seronegative for rheumatoid factor, anti-CCP, and antinuclear antibodies but are positive for HLA-B27. Less often, a patient may present with chronic polyarthritis with prominent cervical spine fusion rather than lower spine involvement.

Subsets of juvenile arthritis include the following:

1. Oligoarthritis appearing in early childhood, more often in girls; it is associated with antinuclear antibodies, a high risk of chronic iridocyclitis and blindness, and HLA-DR5 (DR11), HLA-DR8, or HLA-DR6, as well as HLA-DP2, but not HLA-B27.
2. Polyarthritis appearing in early childhood, more often in girls who are seronegative for rheumatoid factor and antinuclear antibodies; it is associated with HLA-DR8 and HLA-DP3 but not HLA-B27.
3. Polyarthritis associated with rheumatoid factor, anti-CCP, and HLA-DR4 (but not HLA-B27), which probably represents true juvenile rheumatoid arthritis.
4. Still disease, characterized by high, spiking fever, evanescent rash, hepatosplenomegaly, lymphadenopathy, and polyarthritis in patients who are seronegative and HLA-B27 negative.

Miscellaneous Arthropathies

ACNE-ASSOCIATED ARTHRITIS

A rare inflammatory oligoarthritis may occur in patients with severe forms of acne, including acne conglobata, acne fulminans, hidradenitis suppurativa, and dissecting cellulitis of the scalp.[55] Such patients experience fever and inflamed joints; symptoms resemble those of septic arthritis, but the joints are sterile by culture. Sacroiliitis has been described in some patients.

SAPHO is an acronym for a syndrome that consists of synovitis, severe acne, palmoplantar pustulosis, hyperostosis, and osteitis and that may be a form of spondyloarthritis.[55-57] These arthritides may represent forms of reactive arthritis, but patients are usually HLA-B27 negative. Antibiotic therapy is usually of little or no benefit, but some patients respond to NSAIDs or low-

dose corticosteroids. Surgical excision of the affected skin, when possible, has been reported to resolve the arthritis.

WHIPPLE DISEASE

Whipple disease is a rare multisystem disorder that usually affects men (the ratio of affected men to women is 9:1). Patients may present with arthralgias or transient episodes of additive, symmetrical polyarthritis that is nondeforming. Sacroiliitis has been reported in rare instances, and the frequency of HLA-B27 may be increased in patients with Whipple disease. Patients usually have GI symptoms, including diarrhea, steatorrhea, and profound weight loss. Other clues to diagnosis are skin hyperpigmentation, serositis (pleural effusions), lymphadenopathy, uveitis, nervous system disease (ocular palsies or encephalopathy), leukocytosis, and thrombocytosis. The diagnosis traditionally has been based on small-bowel biopsies showing deposits on periodic acid–Schiff staining or electron microscopic demonstration of rodlike bacillary organisms in intestinal macrophages. The causative organism has been identified by RNA sequence analysis and cultured as a gram-positive actinomycete named *Tropheryma whippelii*.[58] Diagnosis can be made on the basis of results from PCR analysis of DNA from affected tissues or blood samples. Long-term treatment with tetracycline usually results in complete remission.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Khan MA: Update on spondyloarthropathies. Ann Intern Med 136:896, 2002

2. Lawrence RC, Helmick CG, Arnett FC, et al: Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 41:778, 1998

3. Braun J, Bollow M, Remlinger G, et al: Prevalence of spondyloarthropathies in HLA-B27 positive and negative blood donors. Arthritis Rheum 41:58, 1998

4. Wordsworth P: Genes in the spondyloarthropathies. Rheum Dis Clin North Am 24:845, 1998

5. Reveille JD, Arnett FC: Spondyloarthritis: update on pathogenesis and management. Am J Med 118:592, 2005

6. Gladman DD: Psoriatic arthritis. Rheum Dis Clin North Am 24:829, 1998

7. Hampe J, Shaw SH, Saiz R, et al: Linkage of inflammatory bowel disease to human chromosome 6p. Am J Hum Genet 65:1647, 1999

8. Hugot JP, Chamaillard M, Zouali H, et al: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411:599, 2001

9. Rioux JD, Daly MJ, Silverberg MS, et al: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn's disease. Nat Genet 29:223, 2001

10. Höhler T, Märker-Hermann E: Psoriatic arthritis: clinical aspects, genetics and the role of T cells. Curr Opin Rheumatol 13:273, 2001

11. Taurog JD, Maika SD, Satumtira N, et al: Inflammatory disease in HLA-B27 transgenic rats. Immunol Rev 169:209, 1999

12. Märker-Hermann E, Höhler T: Pathogenesis of human leukocyte antigen B27–positive arthritis: information from clinical materials. Rheum Dis Clin North Am 24:865, 1998

13. Elkman P, Saarinen M, He Q, et al: HLA-B27–transfected (*Salmonella* permissive) and HLA-A2–transfected (*Salmonella* permissive) human monocyte U937 cells differ in their production of cytokines. Infect Immun 70:1609, 2002

14. Lerisalo-Repo M, Hannu T, Mattila L, et al: Microbial factors in spondyloarthropathies: insights from population studies. Curr Opin Rheumatol 15: 408, 2003

15. Amor B: Reiter's syndrome: diagnosis and clinical features. Rheum Dis Clin North Am 24:677, 1998

16. Seiper J: Disease mechanisms in reactive arthritis. Curr Opin Rheumatol 16:110, 2004

17. Vittecoqoq O, Schaeverbeke T, Favre S, et al: Molecular diagnosis of *Ureaplasma urealyticum* in an immunocompetent patient with destructive reactive polyarthritis. Arthritis Rheum 40:2084, 1997

18. Hannu T, Puolakkainen M, Leirisalo-Repo M: *Chlamydia pneumoniae* as a triggering infection in reactive arthritis. Rheumatology 38:411, 1999

19. Montenegro V, Monteiro RC: Elevation of serum IgA in spondyloarthropathies and IgA nephropathy and its pathogenic role. Curr Opin Rheumatol 11:265, 1999

20. De Keyser F, Elewaut D, DeVos M, et al: Bowel inflammation and the spondyloarthropathies. Rheum Dis Clin North Am 24:785, 1998

21. Braun J, Bollow M, Neure L, et al: Use of immunohistologic and in situ hybridization techniques in the examination of sacroiliac joint biopsy specimens from patients with ankylosing spondylitis. Arthritis Rheum 38:499, 1995

22. Rudwaleit M, Höhler T: Cytokine gene polymorphisms relevant for the spondyloarthropathies. Curr Opin Rheumatol 13:250, 2001

23. Gerard HC, Branigan PJ, Schumacher HR, et al: Synovial *Chlamydia trachomatis* in patients with reactive arthritis/Reiter's syndrome are viable but show aberrant gene expression. J Rheumatol 25:734, 1998

24. Nikkari S, Rantakokko K, Ekman P, et al: *Salmonella*-triggered reactive arthritis. Arthritis Rheum 42:84, 1999

25. Braun J, Tuszewski M, Ehlers S, et al: Nested polymerase chain reaction strategy simultaneously targeting DNA sequences of multiple bacterial species in inflammatory joint diseases: examination of sacroiliac and knee joint biopsies of patients with spondyloarthropathies and other arthritides. J Rheumatol 24:1101, 1997

26. Lau CS, Burgos-Vargas R, Louthrenoo W, et al: Features of spondyloarthritis around the world. Rheum Dis Clin North Am 24:753, 1998

27. Blanco-Gelaz MA, Lopez-Vazquez A, Garcia-Fernandez S, et al: Genetic variability, molecular evolution, and geographic diversity of HLA-B27. Hum Immunol 62:1042, 2001

28. Burgos-Vargas R, Pacheco-Tena C, Vazquez-Mellado J: Juvenile-onset spondyloarthropathies. Rheum Dis Clin North Am 23:569, 1997

29. van Tubergen A, Coenen J, Landewe R, et al: Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. Arthritis Rheum 47:8, 2002

30. Francois RJ, Braun J, Khan MA: Entheses and enthesitis: a histopathologic review and relevance to spondyloarthropathies. Curr Opin Rheumatol 13:255, 2001

31. Helliwell PS, Hickling P, Wright V: Do the radiological changes of classic ankylosing spondylitis differ from the changes found in the spondylitis associated with inflammatory bowel disease, psoriasis, and reactive arthritis? Ann Rheum Dis 57:135, 1998

32. Resnick D, Shapiro RF, Wiesner KB, et al: Diffuse idiopathic skeletal hyperostosis (ankylosing hyperostosis of Forestier and Rotes-Querol). Semin Arthritis Rheum 7:153, 1978

33. Banares A, Hernandez-Garcia C, Fernandez-Gutierrez B, et al: Eye involvement in the spondyloarthropathies. Rheum Dis Clin North Am 24:771, 1998

34. Bergfeldt L: HLA-B27–associated cardiac disease. Ann Intern Med 127:621, 1997

35. Roldan CA, Chavez J, Wiest PW, et al: Aortic root disease and valve disease associated with ankylosing spondylitis. J Am Coll Cardiol 32:1397, 1998

36. Casserly IP, Fenlon HM, Breatnach E, et al: Lung findings on high-resolution computed tomography in idiopathic ankylosing spondylitis: correlation with clinical findings, pulmonary function testing and plain radiography. Br J Rheumatol 36:677, 1997

37. Gratacos J, Orellana C, Sanmarti R, et al: Secondary amyloidosis in ankylosing spondylitis: a systematic survey of 137 patients using abdominal fat aspiration. J Rheumatol 24:912, 1997

38. Charlesworth CH, Savy LE, Stevens J, et al: MRI demonstration of arachnoiditis in cauda equina syndrome of ankylosing spondylitis. Neuroradiology 38:462, 1996

39. LeBlanc CM, Inman RD, Dent P, et al: Retroperitoneal fibrosis: an extraarticular manifestation of ankylosing spondylitis. Arthritis Rheum 47:210, 2002

40. Leirisalo-Repo M: Prognosis, course of disease, and treatment of the spondyloarthropathies. Rheum Dis Clin North Am 24:737, 1998

41. Maksymowych WP, Jhangri GS, Fitzgerald AA, et al: A six-month randomized, controlled, double-blind, dose-response comparison of intravenous pamidronate (60 mg versus 10 mg) in the treatment of nonsteroidal antiinflammatory drug–refractory ankylosing spondylitis. Arthritis Rheum 46:766, 2002

42. Taggart A, Gardiner P, McEvoy F, et al: Which is the active moiety of sulfasalazine in ankylosing spondylitis? A randomized, controlled study. Arthritis Rheum 39:1400, 1999

43. Braun J, de Keyser F, Brandt J, et al: New treatment options in spondyloarthropathies: increasing evidence for significant efficacy of anti–tumor necrosis factor therapy. Curr Opin Rheumatol 13:245, 2001

44. Gorman JD, Sack KE, Davis JC Jr: Treatment of ankylosing spondylitis by inhibition of tumor necrosis factor alpha. N Engl J Med 346:1349, 2002

45. Braun J, Brandt J, Listing J, et al: Treatment of active ankylosing spondylitis with infliximab: a randomised controlled multicentre trial. Lancet 359:1187, 2002

46. Fernandez-Nebro A, Tomero E, Ottiz-Santamaria V, et al: Treatment of rheumatic inflammatory disease in 25 patients with secondary amyloidosis using tumor necrosis factor alpha antagonists. Am J Med 118:552, 2005

47. Kerr HE, Sturrock RD: Clinical aspects, outcome assessment, disease course, and extra-articular features of spondyloarthropathies. Curr Opin Rheumatol 11:235, 1999

48. Iliopoulos A, Karras D, Ioakimidis D, et al: Change in the epidemiology of Reiter's syndrome (reactive arthritis) in the post-AIDS era? An analysis of cases appearing in the Greek army. J Rheumatol 22:252, 1995

49. Li F, Schumacher HR, Kieber-Emmons T, et al: Molecular detection of bacterial DNA in venereal-associated arthritis. Arthritis Rheum 39:950, 1996

50. Sieper J, Rudwaleit M, Braun J, et al: Diagnosing reactive arthritis: role of clinical setting in the value of serologic and microbiologic assays. Arthritis Rheum 46:319, 2002

51. Thomson GD, DeRubeis DA, Hodge MA, et al: Post-*Salmonella* reactive arthritis: late clinical sequelae in a point source cohort. Am J Med 98:13, 1995

52. Schon MP, Boehncke WH: Psoriasis. N Engl J Med 352:1899, 2005

53. Gladman DD, Farewell VT: The role of HLA antigens as indicators of disease progression in psoriatic arthritis. Arthritis Rheum 38:845, 1995

54. Olivier I, Salvarani C, Cantini F, et al: Ankylosing spondylitis and undifferentiated spondyloarthropathies: a clinical review and description of a disease subset with older age at onset. Curr Opin Rheumatol 13:280, 2001

55. Olafsson S, Khan MA: Musculoskeletal features of acne, hidradenitis suppurativa, and dissecting cellulitis of the scalp. Rheum Dis Clin North Am 18:215, 1992

56. Winchester R: Psoriatic arthritis and the spectrum of syndromes related to the SAPHO (synovitis, acne, pustulosis, hyperostosis, and osteitis) syndrome. Curr Opin Rheumatol 11:251, 1999

57. Hayem G, Bouchaud-Chabot A, Banali K, et al: SAPHO syndrome: a long-term follow up study of 120 cases. Semin Arthritis Rheum 29:159, 1999

58. Raoult D, Birg ML, La Scola B, et al: Cultivation of the bacillus of Whipple's disease. N Engl J Med 342:620, 2000

IV SYSTEMIC LUPUS ERYTHEMATOSUS

Michael D. Lockshin, m.d.

Disease Definition and Subclassification

Lupus is a chronic autoimmune illness characterized by autoantibodies directed at nuclear antigens and causing a variety of clinical and laboratory abnormalities, including rash, arthritis, leukopenia and thrombocytopenia, alopecia, fever, nephritis, and neurologic disease. Most or all of the symptoms of acute lupus are attributable to immunologic attack on the affected organs. Many complications of long-term disease are attributable both to the disease and to its treatment.[1]

The term lupus applies to several variants of the illness [*see Table 1*], of which systemic lupus erythematosus (SLE) is the most serious and most common. SLE is the prototype of a systemic autoimmune illness, involving multiple organ systems in pathogenically similar ways. Characteristically, patients with SLE progress through periods of active inflammation (flare) and periods of quiescence (remission), both of which may occur spontaneously; periods of flare and quiescence may also be induced. The reasons for the varying course are unknown. Intense sun exposure, drug reactions, and infections are circumstances that are known to induce flare; the aim of treatment is to induce remission.

SLE may occur as an overlap syndrome that shares features with other autoimmune illnesses, such as mixed or undifferentiated connective tissue disease, dermatomyositis, Sjögren syndrome, rheumatoid arthritis, and scleroderma. Organ-specific autoimmune diseases, such as thyroiditis, autoimmune hemolytic anemia, and idiopathic thrombocytopenia, frequently accompany and may be part of SLE.

Lupus may also appear as a skin disease only. Discoid lupus occurs as a destructive, scarring rash, unaccompanied by systemic symptoms or autoantibodies.[2] Subacute cutaneous lupus comprises a characteristic persistent, polycyclic rash; relatively minor visceral symptoms; and strongly positive blood tests.

Drugs such as procainamide and some anticonvulsants induce a lupuslike syndrome, which is called drug-induced lupus.[3] Uncommonly, persons (often relatives of lupus patients) have positive blood tests for lupus but are clinically well. In the absence of symptoms, such persons are not considered to have lupus.

Approximately one third of lupus patients have antiphospholipid antibody, which induces blood clots and fetal death. The presence of this antibody, in the absence of clinical lupus, is referred to as primary antiphospholipid antibody syndrome.[4]

Neonatal lupus is a syndrome consisting of rash, thrombocytopenia, and congenital heart block occurring in infants born of mothers who carry antibody to the SS-A (Ro) and SS-B (La) antigens.[5] It does not evolve into SLE.

Epidemiology and Genetics

SLE is primarily a disease of young women, but the female predominance of SLE has not been explained. Women between 15 and 45 years of age are the most commonly affected; the female-to-male ratio in this age group is between 6:1 and 9:1. African Americans are four times as likely to develop lupus as are whites.[6] The disease incidence in Asians, Hispanics, and Native Americans falls between that of blacks and whites [*see Figure 1*]. No cogent explanation offered to date suggests why African Americans are more frequently affected; racial differences in SLE incidence persist when socioeconomic differences have been controlled. SLE severity in men is similar to that in women. Overall survival is lower in African Americans.[7]

The familial aspects of lupus are striking: approximately 10% of persons with lupus have family members with lupus or other autoimmune disease. Susceptibility to lupus is higher in persons with specific genetic deficiencies [*see* Other Genetic Susceptibilities, *below*].

Pathophysiology and Pathogenesis

AUTOANTIBODIES

Circulating antibodies to a broad list of autoantigens characterize SLE. Antinuclear antibodies, usually defined by immunofluorescence, are present in almost all lupus patients; tests for antinuclear antibody constitute a screening test (sensitive, but not specific) for the illness. Autoantibodies to nuclear constituents—primarily to double-stranded (native) DNA but also to single-stranded (denatured) DNA, histones, ribonuclear proteins, and other nuclear antigens, such as the Smith (Sm) antigen—confirm the diagnosis and are likely pathogenic. For example, these autoantibodies cause glomerulonephritis by inciting inflammation when deposited as complement-fixing immune complexes on glomerular basement membranes (GBMs) or by binding directly to the GBM.[8] Both animal models and clinical observations suggest that autoantibodies in the presence of complement mediate lupus-associated glomerulonephritis, hemolysis, and thrombocytopenia. Antibodies to phospholipid-binding proteins (beta$_2$-glycoprotein I, prothrombin, and others) mediate thrombosis and fetal loss. Lupus rash and arthritis are less clearly linked to autoantibodies, but immune reactants and inflammation are demonstrable in relevant biopsy specimens, primarily at the GBM.[9] Among lupus manifestations, neurologic lupus is least clearly caused by anti-DNA or other autoantibody; however, anti–ribosomal P antibody may define mood disorders in neurologic lupus.[10] A hypothesis that antibody to a glutamate receptor has diagnostic and pathogenic importance in SLE-associated cognitive dysfunction has proved to be untrue.[11]

ABNORMAL INNATE AND ADAPTIVE IMMUNITY

Genetic defects of immune complex processing are unusually frequent in lupus patients, suggesting that SLE arises because of incomplete or improper disposal of exogenous material.[12] Such defects include abnormalities in complement (deficiencies of C1q, C2, or C4), Fc receptor, apoptotic pathways, and phagocytic cells.[13] Defective clearance of immune complexes may result in their persistence in large quantities,[14] and autoantibodies may be a protective mechanism to neutralize them. Other theories of pathogenesis argue that genetic predispositions that promote T helper type 2 cell (Th2) responses or cytokine dysregulation are the underlying defects leading to the development of SLE.

In animal models, several immune defects that may cause SLE have been identified. These include abnormalities in genes affecting overall immunoreactivity (apoptosis [*Fas*, *Fas* ligand,

Table 1 Characteristic Features That Distinguish Lupus from Lupuslike Diseases

Organ System	*Lupus*				*RA*	*Sjögren Syndrome*
	SLE	*Discoid*	*Drug-Induced*	*Neonatal*		
Skin	Specific rashes, alopecia, mucosal ulcers, periungual telangiectasia	Specific rash, alopecia, mucosal ulcers	Rash	Rash	Subcutaneous nodules	Dry eyes, dry mouth
Joints	Symmetrical nondestructive arthritis	—	Symmetrical nondestructive arthritis	—	Symmetrical destructive arthritis	Symmetrical destructive arthritis
Renal	Glomerulonephritis, renal failure	—	—	—	Amyloidosis (late)	Tubular dysfunction
CNS	Seizures, psychosis, cognitive dysfunction, stroke, myelopathy, neuropathy	—	—	—	Peripheral neuropathy	Peripheral and cranial neuropathy
Cardiac	—	—	—	Heart block	—	—
Blood						
ANA	Strong positive, any pattern	May be positive	Strong positive	Positive	May be positive	Positive
Complement	Low with renal disease or hemolytic anemia	Normal	Normal	Normal or low	Commonly high	Low or high
Diagnostic autoantibodies	Anti-dsDNA, anti-Sm	—	Antihistone	Anti– SS-A, anti–SS-B	Anti-IgG (rheumatoid factor)	Anti–SS-A, anti–SS-B
Other autoantibodies	Anti–SS-A (Ro), anti–SS-B (La), anti-RNP, anti-IgG, anti-ssDNA	Anti-ssDNA	Antihistone	Anti-RNP	—	Anti-IgG
Other abnormalities	Leukopenia, thrombocytopenia, hemolysis	—	Leukopenia	Thrombocytopenia, hemolysis	Leukocytosis, thrombocytosis	Hyperglobulinemia

ANA—antinuclear antibody CNS—central nervous system CPK—creatinine phosphokinase dsDNA—double-stranded DNA LLD—lupuslike disease MCTD—mixed connective tissue disorder PAPS— primary antiphospholipid syndrome RA—rheumatoid arthritis RNP—ribonucleoprotein ssDNA—single-stranded DNA SLE—systemic lupus erythematosus UCTD—undifferentiated connective tissue disease

Bcl-2]), B cell activation (*FcγRIIB, SHP-1, CD22, CD19, PD-1, Lyn, Blys-1*), T cell activation (*TGF-β, TGF- βR, PD-1*), cell proliferation (*p21, Fli-1*), and cytokines (*IFN-γ, IL-4, IL-10, TNF-α*); and genes affecting autoantigen clearance (*C1q, C4, SAP, DNAse-1*).[15,16] Gene array data strongly indicate that interferon genes are markedly upregulated in patients with active SLE. Whether this is causative or a result of disease is not known.[17]

OTHER GENETIC SUSCEPTIBILITIES

Twin and family studies of SLE make it abundantly clear that the illness is highly heritable. HLA types DR3 and DR4 predominate in SLE patients.[18] Specific susceptibility loci on chromosomes 1, 4, and 7, among others, have been identified.[19] Persons with genetic deficiencies of complement appear to be more susceptible to the development of lupus.[20] Specific FcIII gamma receptor alleles increase the severity of lupus nephritis, particularly in whites.[21] The genetics of lupus are extremely complex, however, and no single genetic trait is unequivocally linked to susceptibility to the illness. Several national registries are currently attempting to definitively describe the genetics of lupus.

INFECTIONS

Although an infectious trigger of SLE has long been suspected, no single infection has been found. Universal exposure of children with SLE to Epstein-Barr virus has been noted (at an age when 50% exposure is the norm), suggesting a possible link of this virus to disease.[22] Autoantibodies can be identified in serum specimens up to a decade before the earliest symptoms of SLE. Autoantibodies first appear as one specific antibody, then generalize just before clinical onset. Whether this progression reflects response to infection or autoimmunity is unknown.[23]

ESTROGEN

Some investigators attribute the female predominance of SLE and its occurrence in childbearing years to the upregulating effect of estrogen on the immune system, a phenomenon demonstrable largely in vitro. However, this argument applies to autoimmunity in general, not specifically to lupus, and it fails to explain why other autoimmune diseases have much less striking female-to-male ratios. Furthermore, postmenopausal estrogen replacement and oral contraceptive use do not significantly alter SLE incidence or severity, nor does pregnancy; minor differences in incidence or susceptibility have on occasion been reported.[24-26] Alternative explanations for a high female-to-male ratio include an estrogen-sensitive threshold mechanism[27] or sex differences of exposure to exogenous agents (although none has been convincingly suggested.[28] Reports of patients with Klinefelter syndrome and SLE have prompted investigators to consider the possible involvement of male hypogonadism in the pathogenesis of lupus; patients with Klinefelter syndrome may be unusually susceptible to lupus.

COMPLICATIONS OF CHRONIC ILLNESS

Most current information on SLE pathogenesis focuses on upregulation or downregulation of components of the immune response, genetic controls of immunity, and potential etiologic agents. However, the long-term damage of chronic disease, from tissue injury or treatment, is as important to patients as acute inflammatory disease. Some elements of damage are clearly attributable to therapy: osteoporosis, osteonecrosis, cataracts, and tendon ruptures are all associated with long-term corticosteroid therapy. Other elements of damage result directly from the inflammatory and immunologic aspects of the illness, which lead to tissue necrosis and scarring: progressive renal failure, destruc-

Table 1 (*continued*)

MCTD	*UCTD*	*PAPS*	*Scleroderma*	*LLD*	*Dermatomyositis*
Sclerodactyly	—	Livedo reticularis	Scleroderma, periungual telangiectasia	—	Specific rash, periungual telangiectasia
Symmetrical nondestructive arthritis	Symmetrical nondestructive arthritis	—	Transient, symmetrical, early arthritis	Symmetrical nondestructive arthritis	—
—	—	Thrombotic microangiopathy	Angiotensin-driven renal crisis	—	—
—	—	Stroke, myelopathy	Hypertensive crisis	—	Myopathy
—	—	—	—	—	—
Positive, speckled Normal	May be positive Normal	May be positive Normal	Positive, speckled, nucleolar, centromere Normal	May be positive Normal	Positive Normal
Anti-RNP	—	Anticardiolipin, lupus anticoagulant	Anti–Scl-70, anticentromere (topoisomerase I)	—	Anti–Jo-1
—	—	Anti-β_2-glycoprotein I	—	—	—
—	—	Thrombocytopenia	High renin during crisis	—	High CPK and aldolase

tive joint disease, and brain infarcts. Many of these complications are more severe in patients of minority races or lower socioeconomic classes.[29] Still other elements of chronicity bear an uncertain relationship to disease and treatment: accelerated atherosclerosis,[30] valvular heart disease, cognitive dysfunction, and psychosocial dysfunction.

Diagnosis

The American College of Rheumatology (ACR) has defined criteria for the classification of SLE [*see Table 2*].[31] Although the ACR criteria are useful for ensuring uniformity of patients reported in medical journals, the criteria are often mistakenly used as diagnostic criteria. For individual patients, the criteria have high false negative and false positive rates.[32] For instance, a patient with biopsy-proven lupus nephritis, positive antinuclear antibody, and anti-Sm antibody as the only manifestations of disease does not fulfill ACR criteria, whereas a patient with rheumatoid arthritis who has a positive antinuclear antibody, low positive anti-DNA antibody, and leukopenia (e.g., from Felty syndrome) does fulfill ACR criteria. As a rule, characteristic disease of one organ system (kidney, joints, skin) plus a high-titer anti-dsDNA or Sm antibody suffices to make the clinical diagnosis.

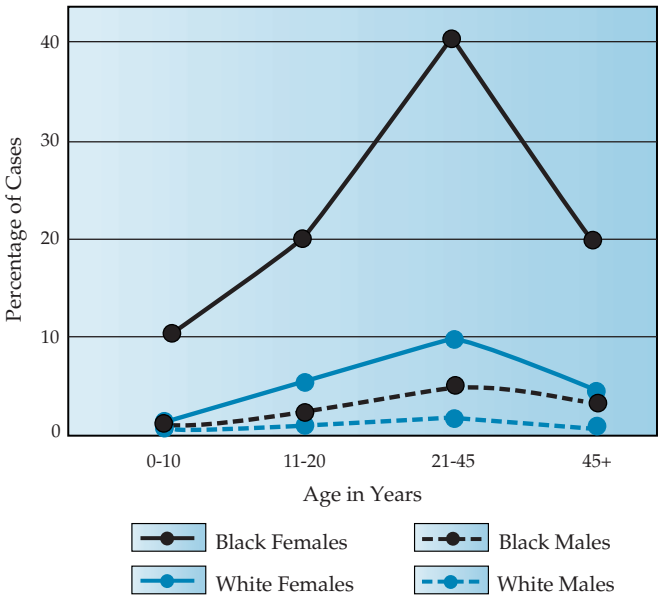


Figure 1 **Age, sex, and race distribution of the incidence of SLE.[6]**

In clinical practice, diagnosis of SLE is based on a combination of autoantibody assays, clinical manifestations, and laboratory studies of affected organ systems. The clinical manifestations of lupus are protean. Patients with lupus activity or damage may be asymptomatic or may present with findings that reflect the specific organ systems involved [*see Table 3*].

Symptoms and signs accumulate over time in patients with SLE [*see Table 4*].[33] At any given time, especially at the onset of illness, most often only a few manifestations are present. Arthritis, malaise, cytopenias, and rashes are the most prominent early findings. Nephritis (with renal failure), arthritis, osteoporosis and osteonecrosis (corticosteroid complications), neurologic disease, accelerated atherosclerosis, and cardiac valvular disease dominate the late course. With disease activity and with its treatment, the risk of opportunistic infection is high. For conceptual purposes, it is easiest to consider disease activity and manifestations separately for each affected organ system.

SYSTEMIC SIGNS AND SYMPTOMS

Malaise, arthralgia, myalgias, fever (usually low grade), and weight loss are common manifestations of active SLE. Some patients will have high temperatures (> 40° C [104° F]). Even with very high fever, shaking chills are unusual; when present, they suggest infection. Like the organ-specific manifestations of lupus, the systemic symptoms vary considerably during the day and

Table 2 American College of Rheumatology Criteria for the Classification of SLE*

Malar rash
Discoid rash
Photosensitivity
Oral ulcers
Arthritis
Serositis (pleuritis or pericarditis)
Renal disorder (proteinuria > 0.5 g/day or cellular casts)
Neurologic disorder (seizures or psychosis)
Hematologic disorder (hemolytic anemia, leukopenia, lymphopenia, or thrombocytopenia)
Immunologic disorder (anti-DNA, anti-Sm, or antiphospholipid antibodies [anticardiolipin, lupus anticoagulant, or biologic false positive test for syphilis])
Antinuclear antibody

Note: these are not diagnostic criteria.
*Four criteria are required to include a patient in an SLE cohort of a research study.

over weeks. In approximately one third of lupus patients, sun exposure, usually intense, will induce systemic flare. Sun exposure that is mild or of short duration does not harm most patients.

SKIN AND MUCOSAL INVOLVEMENT

Up to half of lupus patients manifest some degree of alopecia. Typically, this takes the form of broken frontal hairs and diffuse thinning, which recovers when health is regained. Severe alopecia may occur. Discoid rashes cause focal patches of hair loss.

Most patients develop a rash at some point during their illness. The well-known butterfly rash, on both cheeks and across the bridge of the nose [*see Figure 2*], occurs in only a minority of patients, but most rashes involve the face in some manner. Commonly the tip of the chin, the upper lip, the eyebrows, and the hairline are also involved. The rash may consist of erythema only or may be papular and scaly or deeply pigmented (discoid rash). Discoid rashes, which may scar, are hyperpigmented at the circumference and often depigmented centrally [*see Figure 3*]. Patients with discoid rashes may have either discoid lupus or SLE. Other types of rashes occur only in SLE.

Table 3 Physical Examination Abnormalities in Acute and Chronic SLE *

Organ	*Acute Disease*		*Chronic Disease*	
	Common	*Uncommon*	*Common*	*Uncommon*
General	Fever, weight loss	Asthenia	Cachexia	
Skin	Malar rash, rash elsewhere, alopecia	Periungual telangiectasia, vasculitis	Malar rash, rash elsewhere, alopecia, striae, atrophy, pigment change	Periungual telangiectasia, skin ulcers
Nodes	Lymphadenopathy	—	—	—
Breasts	—	—	—	—
Eyes	—	Retinal hemorrhages, exudates	Hypertensive changes	Retinal hemorrhages, exudates
Ears	—	Rash in ear canal, decreased hearing	—	Scarring in ear canal, decreased hearing
Nose	—	Septal ulceration	—	Septal perforation
Throat	—	Mucosal ulcer (hard palate)	—	Mucosal scarring (hard palate)
Chest	—	Rales, pleural rub, effusion	—	Rales, effusion
Heart and vessels	Raynaud phenomenon	Pericardial rub, enlargement	Raynaud phenomenon, valve disease	Enlargement, valve insufficiency, arrhythmia
Abdomen	—	Hepatomegaly, splenomegaly	—	Hepatomegaly, splenomegaly, ascites
Muscles	Weakness, tenderness	—	Weakness, atrophy, tendon rupture	—
Bones	—	—	Fracture (vertebrae, hip), osteonecrosis	—
Joints	Synovitis, restricted motion	—	Synovitis, deformity, restricted motion	Jaccoud deformities
Neuromotor	—	Stroke, mononeuritis multiplex, seizure	—	Stroke, mononeuritis multiplex, seizure
Neurosensory	—	Peripheral neuropathy, mononeuritis multiplex	—	Peripheral neuropathy, mononeuritis multiplex
Cognitive	Depression	Psychosis, dementia	Depression	Psychosis, dementia

*This table is not comprehensive; it does not include rare abnormalities.

Table 4 Frequencies of Various Manifestations of SLE by Disease Stage[33]

Manifestation	*Early Disease (%)*	*Late Disease (%)*
Arthritis	46–53	83–95
Rash	9–11	81–88
Fever	3–5	77
Mucosal ulcers	—	7–23
Alopecia	—	37–45
Serositis	5	63
Pulmonary inflammation	—	9
Liver function test abnormalities	1	—
Vasculitis	—	21–27
Myositis	—	5
Osteoporosis	—	High
Osteonecrosis	—	7–24
Leukopenia	41–66	41–66
Thrombocytopenia	2	19–45
Anemia	2	57–73
CNS abnormalities	3	55–59
Nephritis	6	31–53
Renal failure	< 1	20

Diagnostic inflammatory rashes are somewhat raised, scaly, and relatively uniform in appearance across the lesion; they may ulcerate; and they have sharp borders. Less diagnostic rashes occur on the extensor surfaces of the upper arms, the blush area of the neck and shoulders, and the extensor surfaces of the elbows and fingers. These are usually erythematous macular rashes. The erythematous rashes evolve over days to weeks and often leave hyperpigmentation as they recede; they may appear more prominent with fever or pregnancy. Vasculitic rashes (usually small, ulcerating papules) occur on the extensor tips of the elbows; painful, erythematous vascular lesions occur at the distal fingers and palms (lupus pernio). A polycyclic, persistent rash, primarily on the trunk, and specifically associated with anti–SS-A antibody, is known as subacute cutaneous lupus [*see Figure 4*]. A painful subcutaneous lesion that is deeply indurated and tender and may ulcerate is called lupus profundus.

Some patients have only erythema in distributions typical of lupus rashes. Unlike the other rashes, erythematous rashes are not by themselves diagnostic of lupus, but they add to the overall diagnostic information. Lupus rashes are often confused with rosacea (although rosacea is more oily and more papular), polymorphous light eruptions, and allergic reactions.

Chronic nasal ulcers and recurring painless mouth ulcers [*see Figure 5*] (particularly on the hard palate, but also on the gums and buccal mucosa) are characteristic of more severe disease. Mucosal ulcers are irritating rather than severely painful. Periungual telangiectasias and small, ulcerating, vasculitic ulcers on the elbows also occur in more severe disease. A particular type of palmar and digital pulp erythema, known as lupus pernio, is a form of vasculitis. In rare cases, subcutaneous inflammation leads to lupus profundus, consisting of local fat necrosis and painful nodules.

LYMPH NODE INVOLVEMENT

Lymphadenopathy is common in active disease. It is modest in extent and generalized (often noted on computed tomography scans of the abdomen or chest). It resolves rapidly in patients started on corticosteroid therapy for other manifestations of SLE.

CARDIOPULMONARY INVOLVEMENT

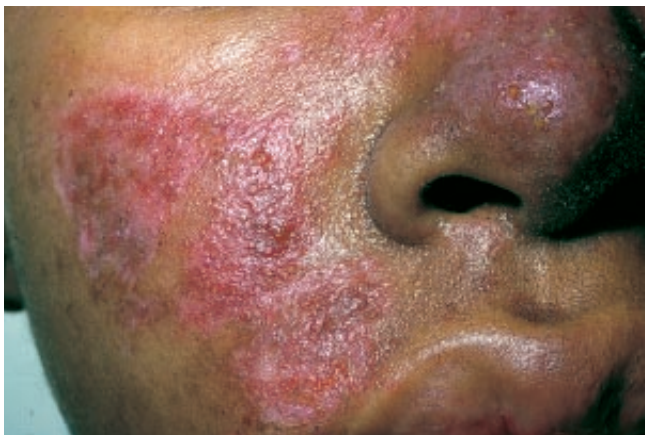
SLE is associated with a range of cardiopulmonary disorders [*see Table 5*]. These usually cause symptoms or abnormal physical findings; it is unnecessary to test asymptomatic patients for cardiovascular disease.

Pleuropericarditis is frequently symptomatic but is not usually life-threatening. It causes pain on breathing, as well as elevation of the diaphragm that is evident on physical examination or x-ray. Atelectasis at the bases of the lungs is audible as fine crackling rales or is visible on x-ray as horizontal lines (plate atelectasis). Pulmonary function tests may reflect reduced diffusion capacity, reduced lung volumes, and reduced lung elasticity.

Pleural effusions also occur, but they are not usually large. On thoracentesis, LE cells (polymorphonuclear leukocytes with ingested nuclear debris appearing as a homogeneous round inclusion) may be found on a Wright stain of the fluid buffy coat.

A minority of patients develop respiratory insufficiency from pulmonary fibrosis. A rare manifestation is so-called lupus lung.

a



b

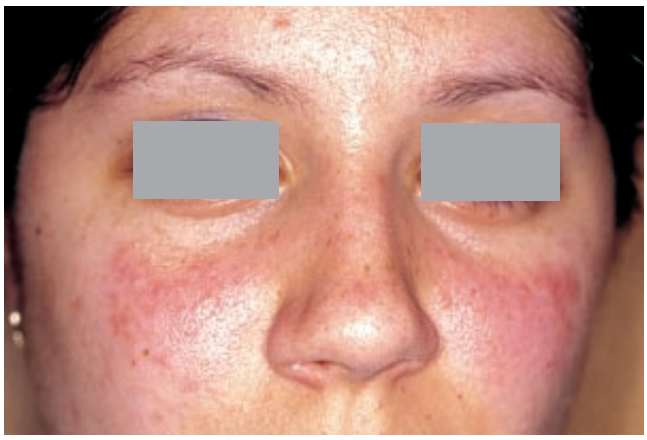


Figure 2 **Most lupus rashes involve the face (a). The butterfly rash of lupus, on both cheeks and across the bridge of the nose (b), occurs in only a minority of patients.**

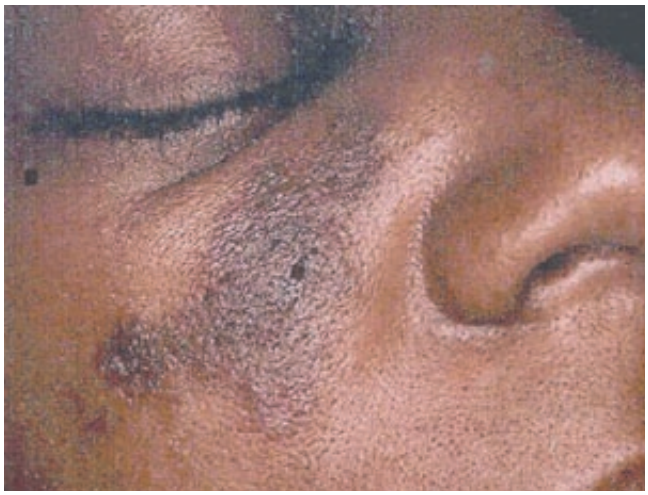


Figure 3 **Discoid lupus rashes are hyperpigmented at the circumference and often depigmented centrally.**

This consists of inflammatory lung disease or pulmonary hemorrhage, either of which is life threatening.[34] Pulmonary hypertension is uncommon but serious when it occurs, which is most often in patients with intense Raynaud phenomenon, recurrent pulmonary emboli, or pulmonary fibrosis.

Transthoracic echocardiography demonstrates valvular heart disease in up to 30% of patients with long-standing SLE; with transesophageal echocardiography, the frequency is much higher.[35] Libman-Sacks lesions occur primarily on the mitral valve but also occur on the aortic valve and, rarely, on the pulmonic or tricuspid valves. Symptomatic valve disease may be more common in patients with antiphospholipid antibody. Pericardial effusions or thickening occurs during active disease, but otherwise, they are uncommon. Small pericardial effusions, which may be symptomatic or asymptomatic, occur often in patients with active SLE. Life-threatening large effusions are uncommon.

Accelerated atherosclerosis is a risk of long-standing lupus,[36,37] leading to myocardial infarction and other vascular occlusive manifestations before the age of 40. Myocardial infarction may also be caused by coronary vasculitis, but this is less common. It is important to consider atherosclerosis, together with vasculitis and thrombosis (from antiphospholipid antibody), in patients with long-standing lupus who present with complaints of vascular insufficiency.

Diffuse nonischemic myocarditis may also occur. Newborns suffering the neonatal SLE syndrome may have complete congenital heart block and may die of congestive heart failure.

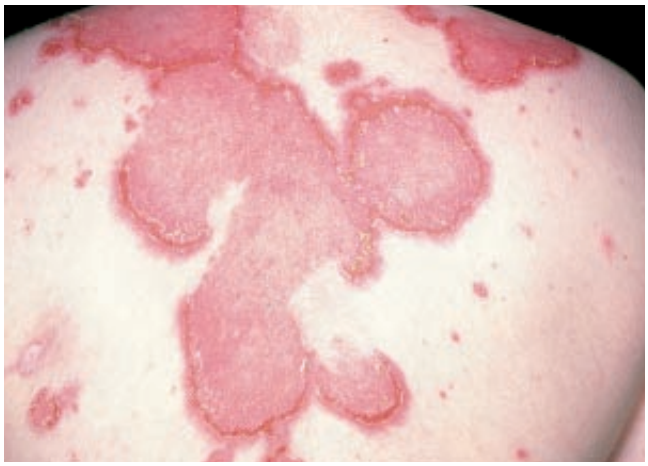


Figure 4 **The rash of subacute cutaneous lupus is polycyclic and persistent.**

MUSCULOSKELETAL INVOLVEMENT

The arthritis of SLE is typically painful, transient, and symmetrical, involving the wrists, small joints of the hands, elbows, knees, and ankles. Swelling and redness are modest. Less often, SLE arthritis will present as asymmetrical oligoarthritis or intensely inflamed, sustained polyarthritis resembling that of rheumatoid arthritis. Although deformity may occur as a result of ligamentous laxity (reversible subluxations, Jaccoud arthropathy),[38] rheumatoid-like joint destruction is uncommon.

Inflammatory myositis occurs primarily in patients with overlap features with scleroderma or dermatomyositis. It presents as proximal myopathy; serum levels of muscle enzymes are modestly elevated; and results of electromyography, magnetic resonance imaging, and muscle biopsy, if done, are similar to those seen in dermatomyositis. However, abnormal enzyme levels associated with proximal muscle tenderness or weakness are sufficient for diagnosis in a patient with established SLE.

Lupus does not involve bone directly. However, bone involvement can occur secondary to organ system failure (e.g., renal failure), severe illness (e.g., osteoporosis from inactivity or catabolic state), or treatment (e.g., corticosteroid-induced osteoporosis or avascular necrosis).

Osteoporosis presents as atraumatic fractures of vertebrae or long bones. Its occurrence is a severe threat to SLE patients, even premenopausal women, because of the frequent use of corticosteroids for treatment and because of inactivity attendant upon polyarthritis and systemic illness.

Avascular necrosis (osteonecrosis) most often occurs in patients who have had a severe flare treated with high-intensity corticosteroid therapy, but this complication can develop in patients who have never received corticosteroid treatment. Marked cushingoid features during steroid treatment and Raynaud phenomenon may be predictors of its occurrence.[39] The femoral head is the most commonly involved site, but shoulders, ankles, wrists, metacarpals, and shafts of long bones are also vulnerable.[40] Typically, affected areas become painful at the initial occurrence of infarction and again years later when the necrotic bone collapses. The most typical presentation of osteonecrosis is sudden hip pain 2 or 3 years after a major flare of lupus. Some patients receiving infusions of high-dose intravenous methylprednisolone complain of intense pain at preexisting osteonecrotic sites during and shortly after the infusion. Reducing the corticosteroid dose at the time of occurrence of pain has no effect on the course of the complication.

GASTROINTESTINAL AND HEPATIC INVOLVEMENT

Esophageal dysfunction is rare in SLE; it occurs primarily in patients with severe Raynaud phenomenon or in patients with scleroderma overlap disease. Gastroduodenal ulcer may occur as a result of treatment but is not directly linked to SLE. Ischemia of the small and large intestines may result from systemic vasculitis; it presents as abdominal angina, pneumatosis intestinalis, infarction or perforation, or pseudo-obstruction. Intestinal ischemia is a rare complication, occurring only in the most severely ill patients.[41]

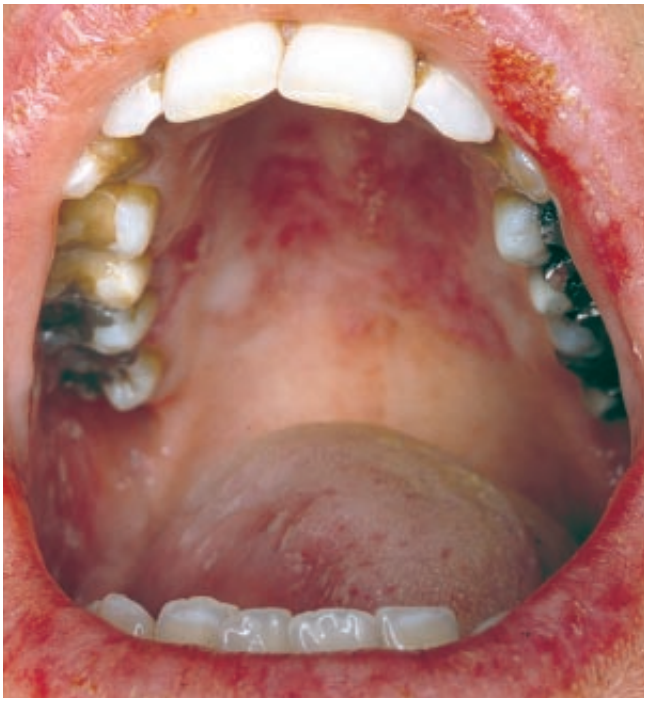


Figure 5 **Painless mouth ulcers, most often found on the hard palate but also found on the gums and buccal mucosa, are characteristic of more severe lupus.**

Diverticulitis often develops in patients with long-standing SLE, especially after prolonged treatment with corticosteroids. The symptoms of diverticulitis are easily masked by corticosteroid therapy. Consequently, diverticular perforation or abscess is frequently misdiagnosed, especially in young SLE patients.

Chemical hepatitis may follow use of nonsteroidal anti-inflammatory drugs (NSAIDs) (lupus patients appear to be unusually susceptible to this side effect) or other drugs, such as azathioprine.[42] Occasionally, patients suffer concomitant autoimmune hepatitis or primary biliary cirrhosis. In the absence of other causes, abnormalities of liver enzyme levels because of SLE are uncommon.

HEMATOLOGIC INVOLVEMENT

Leukopenia is such a regular feature of SLE that its absence, in untreated disease, should raise suspicion that the diagnosis is incorrect or that infection or tissue necrosis is present. Usually, lymphocyte counts show greater reductions than do granulocyte counts: a leukocyte count of about 3.5/mm^3, with 10% lymphocytes, is usual. Leukopenia of this degree seldom places patients at serious risk of infection. There is usually no need to administer granulocyte-macrophage colony-stimulating factor (GM-CSF); there are anecdotal reports that administration of this agent may induce lupus flare.

Thrombocytopenia in SLE is usually low grade, with platelet counts greater than 50,000/mm^3. Severe thrombocytopenia may occur, however; idiopathic thrombocytopenic purpura (ITP) may be an initial presentation of SLE.

SLE may result in anemia of chronic disease and anemia from autoimmune hemolysis. The anemia of chronic disease in SLE patients responds to administration of recombinant erythropoietin.

NEUROLOGIC INVOLVEMENT

Neurologic signs and symptoms represent one of the most serious and least understood aspects of SLE. The primary neurologic manifestations of SLE consist of generalized and focal (usually vascular) brain disease, myelopathy, peripheral neuropathy, mononeuritis multiplex, and cognitive dysfunction. Secondary neurologic events can also occur; these include seizures from hypertension or hemorrhage, delirium from drugs or uremia, brain or spinal cord abscess, and stroke from atheroma or embolus. Attribution of a specific neurologic symptom to active lupus (which is treatable with immunosuppression), as opposed to a complication of lupus or its treatment (which is treatable by ameliorating the offending problem) requires deep investigation and good clinical judgment. Confusion about diagnostic criteria for neurologic lupus led the ACR to publish nomenclature and case-definition criteria for these syndromes.[43]

General Brain Disease

Patients with SLE frequently complain of progressive cognitive dysfunction, such as confusion, forgetfulness, and so-called foggy thinking.[44,45] Retrospective and cross-sectional studies document a high frequency of poor performance on tests of cognitive function, particularly in the executive, short-term memory, and verbal-processing spheres.[46,47] It is not known whether this deficit results from immunologic attack on the brain (by antineuronal or other autoantibodies) or diffuse vascular disease. Cognitive dysfunction may respond to corticosteroid therapy. It seldom progresses to advanced dementia.

Headaches are common in SLE. A special form of migraine called lupus headache has been described, but whether it exists as a definable entity remains a matter of debate.

Focal Brain Disease

Seizures, strokes, cranial neuropathies (including blindness), and cerebellar dysfunction may occur in SLE. These events are assumed to result from vascular occlusion, but they may occur in patients with no known thrombotic diathesis, embolization, atherosclerosis, or vasculitis. Stroke is one of the most common presentations of the antiphospholipid antibody syndrome. Seizures are most common in severely active, febrile, multisystem dis-

Table 5 Cardiopulmonary Manifestations of SLE

Pleuropericarditis
Libman-Sacks endocarditis
Valve insufficiency
Valve stenosis
Ischemic cardiomyopathy
Accelerated atherosclerosis
Antiphospholipid antibody syndrome
Hypertensive heart disease
Vasculitis
Hypertensive heart disease
Pulmonary hypertension
Peripheral arterial insufficiency
Vasculitis
Atherosclerosis
Antiphospholipid antibody syndrome
Peripheral venous thrombosis
Raynaud phenomenon
Complete congenital heart block in newborns with neonatal lupus erythematosus

ease. In this circumstance, they generally do not persist after the disease is brought under control.[48]

Myelopathy

Transverse myelitis occurs in two patterns: (1) abrupt onset, with progression in hours from the first symptom, often heralded by a burning, dysesthetic pain in the legs; and (2) slower progression, in a stuttering fashion, worsening over days. Unless treated immediately and aggressively, both forms may progress to advanced paraparesis or paraplegia. Although few direct data exist to support these hypotheses, it is likely that the first form represents vascular occlusion with spinal cord ischemia and the second form represents inflammatory disease. It is mandatory to exclude a space-occupying mass in all such patients.

A slowly progressive and intermittent myelopathy, very much resembling multiple sclerosis (so-called MS-like or lupoid sclerosis), develops in some lupus patients. There is no definitive way to exclude concomitant MS in these patients except by the association of the myelopathy with SLE and by its failure to progress in the way MS usually does. In this form of lupus myelopathy, cerebrospinal fluid examinations may reveal oligoclonal bands, but MRI studies are atypical for MS.

Peripheral Neuropathy

Stocking-and-glove neuropathy is a slowly progressive lesion that tends to occur in patients with continuing, active disease. Its pathogenesis is unclear; it may result from direct immune attack on peripheral nerves or from vasculitic occlusion of the vasa nervora. Abrupt loss of motor and sensory function, such as sudden occurrence of footdrop or wristdrop, is diagnosed as mononeuritis multiplex. This is a very serious manifestation indicating vasculitis of the vasa nervora; it implies systemic vasculitis, as well.

RENAL INVOLVEMENT

Approximately half of lupus patients develop lupus nephritis, and approximately 10% overall will progress to dialysis or transplantation. Lupus nephritis presents as proteinuria (or an otherwise abnormal urinalysis), hypertension, or a rising serum creatinine level, all of variable degree. In its early stages, lupus nephritis is painless and asymptomatic. In more advanced stages, edema, anemia, symptomatic hypertension, and symptomatic uremia occur. Patients with inflammatory forms of nephritis are usually hypocomplementemic; most have high levels of anti-DNA or anti-Sm antibody. Signs or symptoms of disease active in other organ systems need not accompany lupus nephritis.

The World Health Organization (WHO) pathologic classification of lupus nephritis has been revised. The revised criteria differ from previous classifications by taking into account normal biopsies, scarring, and tubulointerstitial changes; in addition, they incorporate information from immunofluorescence and electron microscopy studies [*see Table 6*].[49] This classification includes indices of disease activity and chronicity, which delineate acute necrosis, inflammatory infiltrate, crescent formation, scarring, and tubular atrophy to provide further prognostic information.[9] Electron microscopy demonstrates immune complex deposits in subepithelial spaces in membranous lupus nephritis and in subendothelial spaces in proliferative lupus nephritis, as well as in mesangial locations. Characteristic tubuloreticular structures, thought to be RNA degradation products, also appear. Immunofluorescence studies demonstrate IgG, IgM, and C3 deposits in the same distributions. Vascular inflammation or endothelial proliferation is also seen.

Although lupus nephritis may present as anuria, acute hypertension, or fluid retention, most often it is first noted by an abnormal urinalysis. If left untreated, lupus nephritis progresses to renal insufficiency over months to years. Biopsy is necessary primarily when the result will change treatment. Urinalysis and blood chemistry results correlate only roughly with biopsy findings [*see Table 7*].

SPECIAL PRESENTATIONS

Neonatal Lupus Syndrome

Approximately 25% of infants born to mothers with anti–SS-

Table 6 International Society of Nephrology/Renal Pathology Society (ISN/RPS) 2003 Classification of Lupus Nephritis

Class	*Name*	*Description*	*Clinical Presentation*
I	Minimal mesangial	Normal light microscopy, mesangial immune deposits by immunofluorescence	Normal urinalysis, normal function
II	Mesangial proliferative	Infiltrating cells and proliferation of mesangium	Mild proteinuria, celluria, normal function
III A III A/C III C	Focal active Focal active and chronic Focal chronic	Infiltrating cells and immune complex deposits in portions of the glomerulus and in < 50% of glomeruli (active) or scarring (chronic)	Variable proteinuria, celluria, normal function
IV A IV G IV-S A/C IV-G A/C IV-S C IV-G C	Diffuse and active (A), global (G), segmental (S), chronic (C)	Infiltrating cells and moderate immune complex deposits in entire glomeruli and in ≥ 50% of glomeruli with segmental or global lesions, with or without scarring	Variable proteinuria, celluria; often severe, decreasing function
V	Membranous	Global or segmental subendothelial immune deposits by light, immunofluorescence, or electron microscopy, with or without mesangial lesions	Marked proteinuria, slowly decreasing function
VI	Advanced sclerotic	≥ 90% of glomeruli globally sclerosed without residual activity	Variable proteinuria, decreased function

Table 7 Likely Renal Biopsy Findings According to Urinalysis and Serum Creatinine Results

Urinalysis		*Creatinine Level*	*Most Likely Pathology*
Protein	*Cells*		
None	None	Normal	Normal or mesangial
Little	WBCs	Normal	Mesangial, focal proliferative
Moderate	WBCs, RBCs	Normal	Mesangial, focal or diffuse proliferative
Moderate	WBCs, RBCs, casts	Normal or elevated	Focal or diffuse proliferative
Severe	WBCs, RBCs, casts	Normal or elevated	Diffuse proliferative, membranoproliferative
Severe	Few	Normal or elevated	Membranous
Moderate	WBCs	Normal or elevated	Interstitial (tubular) disease (in patients with acidosis or electrolyte abnormalities)

A or anti–SS-B antibody will develop a photosensitive rash or thrombocytopenia, both of which are transient. A very small number of these infants will develop complete congenital heart block in utero. Both the cardiac and the skin manifestations constitute the neonatal lupus syndrome. Either can be present independently. The syndrome appears to result from transplacental passage of maternal antibody, and it subsides when the antibody disappears. However, the heart block persists and may be lethal. The antibody likely targets transiently expressed antigens in the fetal conducting system; signals for apoptosis and fibrosis are upregulated.[50]

Antiphospholipid Antibody Syndrome

Between one third and one half of lupus patients have anticardiolipin antibody, lupus anticoagulant, or both. When either of these antibodies is present in high titer, patients are susceptible to recurrent thromboembolic disease, thrombocytopenia, livedo reticularis, and cardiac valvular disease. Women are susceptible to recurrent pregnancy loss. These symptoms, combined with positive blood tests, constitute the antiphospholipid syndrome (APS).[51] In the absence of lupus, the disorder is termed primary APS (PAPS); and when lupus or another rheumatic disease is present, the syndrome is designated secondary APS (SAPS). Current research suggests that the true antigen for the syndrome is the phospholipid binding protein, beta$_2$-glycoprotein I, rather than negatively charged phospholipids themselves. In some patients, antibody to an alternative phospholipid binding protein, such as prothrombin, induces the same syndrome. It is not known what induces clotting events in individual patients, but evidence of endothelial activation or injury, such as circulating endothelial cells, appears to be associated with thromboembolic episodes. The sites of thrombosis are not inflammatory and are best treated by anticoagulation rather than by immunosuppression. However, studies of pregnancy loss in animal models indicate that complement activation is a critical component of fetal injury, suggesting anew the possible involvement of the innate immune system in the development of APS.

LABORATORY TESTS

Tests of a variety of body fluids may be abnormal in patients with SLE [*see Table* 8]. Not all tests are abnormal in all patients. If lupus is suspected, an antinuclear antibody test, a complete blood count, and a urinalysis should be performed; if the results of these tests are all normal, SLE is excluded. However, because lupuslike illnesses are also usually suspected, it is often efficient also to obtain at first visit the following tests: erythrocyte sedimentation rate (ESR) or C-reactive protein level; assays for antibodies against dsDNA, Sm, RNP, SS-A, and SS-B; partial thromboplastin time (or other screening test for lupus anticoagulant) and cardiolipin antibodies; and a chemistry profile that includes liver function tests and serum creatinine level.

The antinuclear antibody (ANA) assay is a screening test for lupus. The ANA assay is almost always positive in high titer (> 1:80) in untreated patients with active disease, but a positive result does not by itself confirm a diagnosis of lupus. Only the anti-dsDNA antibody and anti-Sm antibodies, when present in high titer, are diagnostic of lupus. Anti-dsDNA antibody and complement levels are rough guides to disease activity, but many patients remain well for long periods of time with severely abnormal tests. Hypocomplementemia reflects proliferative lupus nephritis but not other aspects of SLE, including membranous lupus nephritis. The ESR remains elevated in many otherwise well SLE patients, as does the C-reactive protein level.

Brain Imaging Studies

Evaluation of neurologic involvement in SLE is complex. MRI scans, usually with contrast, are indicated for any clinical suspicion of central nervous system disease, such as seizures, cognitive dysfunction, new severe headache, chorea, or stroke symptoms. CT scans are far less definitive, except in stroke. Cerebral angiography or magnetic resonance angiography (MRA) is rarely helpful.

CT and MRI scans of the brain frequently demonstrate atrophy and infarcts (the latter including hyperintense areas in the white matter). These lesions correlate poorly with neurologic disease other than stroke syndromes.[13] Findings on fluorodeoxyglucose positron emission tomography (PET), magnetic resonance spectrography (MRS), and single-photon emission computed tomography (SPECT) are frequently abnormal even in asymptomatic patients and correlate poorly with all but the most severe neuropsychiatric disease.[52] Interpretation of abnormal findings in asymptomatic patients is uncertain.[53,54]

Vascular Evaluation

Vascular evaluation is indicated when there is clinical suspicion of medium-size vascular occlusion. Ultrasound, Doppler studies, angiography, and MRA can demonstrate thromboembolic disease from antiphospholipid antibody or atherosclerosis. The small vessel vasculitis that occurs in SLE is usually beyond the resolution of these technologies.

Table 8 Commonly Abnormal Tests on Body Fluids in SLE

Test	*Abnormality*	*Interpretation*
CBC	Normochromic anemia, leukopenia (WBC ~3,000, thrombocytopenia)	Active SLE
ESR and CRP	Elevated	Active SLE
Urinalysis	Proteinuria, hematuria, leukocyturia, cylindruria	Active lupus nephritis
Coombs and reticulocyte count	Positive, high	Hemolytic anemia
APTT, dRVVT	High	If confirmed with mixing test, lupus anticoagulant
Antinuclear antibody	Strongly positive	Positive in almost all patients during active disease; not specific for lupus
Anti-dsDNA antibody	Strongly positive	Positive in two thirds to three quarters of patients during active disease; diagnostic of lupus
Anti-Sm antibody	Positive	Positive in one quarter to one third of patients; diagnostic of lupus
Anti–SS-A, anti–SS-B, and anti-RNP antibodies	Positive	Positive in one third of patients; nonspecific
Anticardiolipin antibody	Positive	Antiphospholipid antibody syndrome
Complement C3, C4, and CH50	Low	Lupus nephritis likely; also hemolytic anemia and cryoglobulinemia
Cryoglobulin	Present	Active SLE
BUN and serum creatinine	Elevated	Severe lupus nephritis, drug toxicity
Liver function tests	Elevated	Drug toxicity (rarely, active SLE)
CSF protein and cells	Elevated	Present in a minority of patients with CNS SLE
Synovial fluid	WBC 5,000–10,000, normal glucose level	Lupus arthritis
Pleural fluid, pericardial fluid	WBC 5,000–10,000, normal glucose level, low complement, LE cells present	Lupus serositis

APTT—activated partial thromboplastin time BUN—blood urea nitrogen CBC—complete blood count CNS—central nervous system CRP—C-reactive protein CSF—cerebrospinal fluid dRVVT—dilute Russell viper venom time ESR—erythrocyte sedimentation rate WBC—white blood cells

Renal Evaluation

All lupus patients should have urinalyses performed, preferably at each clinic visit, because renal disease may appear de novo at any time. All patients with any abnormality on urinalysis or with an abnormal blood urea nitrogen (BUN) or serum creatinine level should have monitoring of 24-hour urine protein and creatinine clearance no less often than every 6 months. It is important to consider the results of renal testing in context: a serum creatinine level of 1.2 mg/dl may be within the laboratory range of normal, but in a 110 lb young woman, a level that high is very likely abnormal. Falling creatinine clearance always demands evaluation, even when the patient is clinically well.

Kidney biopsy The primary indication to perform a kidney biopsy is to help the physician make a treatment decision. Although abnormal biopsy results may be found in asymptomatic patients with normal urinalyses, it is not clear whether treatment of such patients improves outcome. The well patient with normal urinalysis results and normal renal function generally does not need a kidney biopsy, even if the anti-DNA antibody level is high and the complement level is low. The very ill patient with multisystem disease, including abnormal urinalysis results and abnormal renal function, likely will be treated aggressively anyway and does not need a kidney biopsy. The patient with mild systemic disease, mild urinary abnormalities, or both generally should undergo biopsy, because the decision for conservative or aggressive treatment may depend on the result. Occasionally, a biopsy is done to document end-stage, untreatable disease and thereby permit withdrawal of therapy. However, renal ultrasonography can usually provide the same information.

Cardiac Evaluation

Cardiac monitoring with echocardiography or stress tests is unnecessary on a routine basis. Because of the high frequency of accelerated atherosclerosis, however, any occurrence of dyspnea, dyspepsia, or shoulder or arm pain merits consideration of ischemic cardiac disease.

Differential Diagnosis

The differential diagnosis of lupus involves two linked questions: does the patient have a rheumatic disease, and if so, which one?

NONRHEUMATIC ILLNESSES THAT MIMIC SLE

The syndrome of fever, cytopenia, rash, and adenopathy suggests many infections, including HIV, cytomegalovirus, mononucleosis, and bacterial endocarditis. Acute polyarthritis, rash,

and cytopenias can result from many viral infections, such as hepatitis, parvovirus, and rubella. These syndromes resolve spontaneously within several weeks. This syndrome also suggests hematologic malignancies, primarily the lymphomas, leukemias, and myelodysplastic syndromes. Although the presence of antinuclear antibody is common in many of these illnesses, the presence of anti-DNA or anti-Sm antibodies is not; also uncommon are the specific rashes of lupus, nephritis, and vascular manifestations such as periungual telangiectasia and vasculitic papules. Photosensitivity and frontal alopecia are also characteristics of lupus that do not occur in these other illnesses.

Non-SLE causes of nephritis include poststreptococcal glomerulonephritis, Goodpasture disease, genetic nephropathies, and toxemia. The rash of rosacea is commonly mistaken for that of lupus. ITP or autoimmune hemolytic anemia may occur as isolated illnesses or as part of the multisystemic involvement of lupus. In these circumstances, full clinical and serologic evaluation for lupus will place the findings in proper context.

RHEUMATIC ILLNESSES THAT RESEMBLE SLE

Lupus may resemble a variety of other rheumatic diseases [*see Table 1*]. These include rheumatoid arthritis and Sjögren syndrome [*see 15:II Rheumatoid Arthritis*], as well as scleroderma [*see 15:V Scleroderma and Related Diseases*]. The polyarthritic presentation of lupus is very similar to that of rheumatoid arthritis, Lyme disease, and other rheumatic illnesses. Early scleroderma often presents with bilateral hand edema that is mistaken for polyarthritis.

Dermatomyositis [*see 15:VI Idiopathic Inflammatory Myopathies*] peaks in three age groups: 5 to 10 years of age, late teens and early 20s, and older than 45 years. The rash of dermatomyositis is similar to that of lupus, but they tend to involve the eyes differently: in dermatomyositis, telangiectasia causes the so-called heliotrope appearance of the eyelids; lupus rashes involve the eyebrows, but the malar rash stops abruptly at the orbits. Also, the rash in dermatomyositis commonly spares the ear canals, whereas the ear canals are commonly involved in lupus. Periungual telangiectasia is more dramatic in dermatomyositis than in lupus; it also occurs in scleroderma. Rash over the small joints of the hands suggests dermatomyositis; rash between the joints suggests lupus.

Compared with SLE, rheumatoid arthritis occurs in older persons (40 to 60 years of age) and has less of a female predominance (2:1 to 3:1). Although morning stiffness, fatigue, and weight loss are common in patients with rheumatoid arthritis, specific visceral multisystem disease is not. Leukocytosis rather than leukopenia is characteristic of rheumatoid arthritis. Renal disease is very rare, and when it does occur, it is attributable to tubular disease or amyloidosis rather than to glomerulonephritis. High fever does not occur in rheumatoid arthritis. From 10% to 20% of patients with rheumatoid arthritis have antinuclear antibodies; a small percentage have low-titer anti-DNA antibodies, and a minority have anti–SS-A and anti–SS-B antibodies. Complement levels are usually elevated. Rheumatoid factor is present in 80% of patients with rheumatoid arthritis, compared with its presence in 25% of lupus patients.

OVERLAP DISEASE

Some patients have symptoms suggestive of lupus (most commonly, arthritis, pleuritic pain, and cytopenia) but lack the specific diagnostic criteria for lupus (e.g., butterfly rash, glomerulonephritis, and high-titer anti-DNA or anti-Sm antibody). Other patients have lupuslike symptoms together with findings suggestive of rheumatoid arthritis, dermatomyositis, or scleroderma. Patients with no definable serology and a nondescript clinical picture are defined as having undifferentiated connective tissue disease (UCTD). Still other patients have inflammatory myositis, Raynaud phenomenon, and sclerodactyly together with very high titer antibodies to the ribonucleoprotein antigen (U1 RNP) and no anti-DNA or anti-Sm antibody. This set of findings is defined as mixed connective tissue disease (MCTD).

The differentiation of SLE from UCTD, MCTD, and Sjögren syndrome depends on the extent and pattern of different organ involvement (glomerulonephritis is rare in all these disorders except lupus) and on the accompanying serologic abnormalities. High-titer anti-DNA antibody or anti-Sm antibody generally indicates lupus; high-titer anti-RNP antibody with no other positive antibodies indicates MCTD; and anti–SS-A and anti–SS-B antibodies are consistent with Sjögren syndrome but occur in lupus and rheumatoid arthritis, as well. Occasionally, patients have characteristic rheumatoid destructive arthritis, subcutaneous nodules, and high-titer rheumatoid factor and anti-DNA antibody. These patients, as well as those with other overlap features, should be treated as if they have both diseases.

The prognosis in patients with UCTD tends to be more benign than that in patients with SLE. Patients with MCTD do not develop glomerulonephritis, but the long-term prognosis for patients with this disorder is worsened by the eventual development of pulmonary hypertension.

Treatment

ACUTE DISEASE

Management recommendations for the acute symptoms of lupus depend on the severity and organ systems involved. Non–life-threatening manifestations, such as minor arthritis, arthralgia, malaise, myalgias, serositis, and low-grade fever can often be controlled with full doses of NSAIDs. There is no specific preference among the NSAIDs, but lupus patients are unusually susceptible to hepatic and renal toxicities, which must be monitored. Also, in rare cases, lupus patients have developed abrupt high fever and meningitis after taking ibuprofen and similar drugs. Patients who do not respond to NSAIDs usually do respond to low doses (5 to 10 mg/day) of prednisone. As a rule, patients with inflammatory rashes, as well as patients anticipated to be on treatment for months or longer, should receive antimalarial therapy with hydroxychloroquine, 200 mg twice daily. Over a course of 3 months or more, hydroxychloroquine reduces arthralgia, myalgia, rash, fatigue, malaise, and similar symptoms.[55] Patients expected to take corticosteroids for more than a few weeks should strongly consider bone-protective measures (*see* Osteoporosis, *below*). Such patients should also consider the addition of a lipid-lowering agent, such as a statin, to the regimen. Facial rashes, especially erythematous lesions with edema or telangiectasia, may respond to topical therapy with corticosteroid creams.

Alopecia may recover spontaneously, but it is not otherwise easily amenable to therapy. Wigs and falls or hair extenders are useful. Skillful use of makeup can cover most pigment changes caused by discoid lupus.

Low-dose corticosteroid therapy may be appropriate for modest thrombocytopenia or anemia. Leukopenia does not usually require treatment. High-dose corticosteroid therapy (60 mg

of prednisone daily) is used for patients with severe systemic symptoms, renal disease, or other visceral disease that is potentially life threatening. Treatment should be initiated in split doses during the day, maintained for 4 to 6 weeks, and then tapered; too early reduction of dose usually results in recurrence of disease activity. If longer-term use of corticosteroids is anticipated, if vasculitis or life-threatening disease is present, or if corticosteroid toxicity is unacceptable, it is generally advisable to add immunosuppressive therapy. Immunosuppressive agents used for lupus include cyclophosphamide administered orally or intravenously, azathioprine, and mycophenolate mofetil. A standard regimen for active lupus nephritis includes a high-dose corticosteroid and intravenous cyclophosphamide. The cyclophosphamide is given at a dosage of 1 g/m^2 monthly for 6 months and then every 3 months for 2 years.

Acute cerebral symptoms (other than stroke) are usually treated with high-dose corticosteroids. Hallucinations and other psychotic symptoms respond to antipsychotic medications such as haloperidol, which is often administered in conjunction with corticosteroids. Because psychosis may also result from the use of a high-dose of a corticosteroid alone, withdrawal of corticosteroids may be necessary in some cases. The distinction between so-called steroid psychosis and lupus psychosis is quite difficult. No single set of criteria distinguishes between the two; evidence of ongoing active lupus in other organ systems is an indication to treat the patient for lupus psychosis. Acute lupus episodes are often treated with bolus doses of a corticosteroid (usually, 1,000 mg of methylprednisolone administered by rapid I.V. infusion [1 hr] once daily for 3 days). Very few formal studies support this practice, but clinical experience suggests its efficacy and relative safety. Bolus corticosteroid treatment may cause abrupt increases in blood pressure, acute vasospasm leading to stroke, cardiac infarct, or intestinal infarct. Transient oliguria and increased serum creatinine levels may also occur. To prevent these complications, bolus therapy should be monitored closely and withheld if hypertension is not controlled. Access to renal replacement therapy must be available in the event that the patient experiences diminished renal function.

Many new therapies are being investigated. These are largely biologic therapies and include the use of drugs directed against receptors of immune-activating cells or recognition cells and the use of modulators of immune response, such as CD154, CTLA-4, and anti-C5b. Removal of antibody by passing patient plasma over an absorptive column is also being studied. Tumor necrosis factor–α (TNF-α) inhibition, which was once thought to be dangerous for SLE patients, is under reconsideration. Attempts at hormone manipulation, as with dehydroepiandrosterone (DHEA), have had only modest success.

No single test informs the physician whether treatment for acute SLE is successful or not. Instead, it is necessary to monitor the entire clinical picture, including symptoms and results of physical examination, routine laboratory tests, and immune function studies.

INDICES OF DISEASE SEVERITY

Flare

Increase of inflammation in any SLE-affected organ system is known as flare. In a subpopulation of SLE patients, flare is a continuous, not a dichotomous, variable. In a given patient, it may occur in different organ systems at different rates and intensities; for instance, rash may become severe while nephritis remains stable. As a result, several different schemas for measuring flare exist. They differ in giving different weights to individual measures of disease activity (for instance, does new nephritis count more or less than new rash?) and whether serologic measures (antinuclear antibody titer, anti-DNA antibody, complement) do or do not count in the determination. The available indices—SLE Disease Activity Index (SLEDAI), Systemic Lupus Activity Measure (SLAM), and British Isles Lupus Assessment Group (BILAG)[56]—generally agree in identifying flare in populations of patients but often disagree in specifics. There is poor consensus about distinguishing between day-to-day variation of disease activity and a definite flare. Several components of the indices (e.g., quantitation of rash or of arthritis) are sufficiently subjective that investigators in clinical trials must undergo standardization training before they can validate their scores on individual patients.

Flare in pregnancy Proteinuria, thrombocytopenia, and other pregnancy events that occur in the absence of lupus invalidate most scoring systems for pregnant patients. A specific instrument, the SLE Pregnancy Disease Activity Index (SLEPDAI), has been devised for use in pregnant patients.[57]

Damage

Recurring inflammation and vascular occlusion induce irreversible scarring and such permanent deficits as stroke, cataract, skin thinning, osteoporosis, osteonecrosis, and renal failure. It is common practice, therefore, to score SLE patients according to their activity (flare) indices and their damage indices. The most widely used damage index is the Systemic Lupus International Collaborating Clinics (SLICC).[58]

SLE DURING PREGNANCY

Pregnancy in patients with SLE, once thought to be contraindicated, is now a routine event. The complications of pregnancy in these patients are related to three major issues: abnormal renal function, the presence of antiphospholipid antibody, and the presence of anti–SS-A and anti–SS-B antibody. It remains debatable whether lupus is exacerbated by pregnancy, but consensus now exists that pregnant SLE patients do not need prophylactic increases of corticosteroid therapy; rather, they should be treated in the same manner as patients who are not pregnant, except that drugs with fetal toxicity should not be given.

Renal disease, particularly renal insufficiency, strongly predisposes to toxemia of pregnancy. Hypertension, reduced creatinine clearance, and active SLE all threaten the viability of the fetus. Conversely, women who enter pregnancy with no renal disease and no hypertension usually do well. Recurrent pregnancy loss, particularly in the second trimester, is one of the prime clinical manifestations of the antiphospholipid antibody syndrome; for antiphospholipid antibody–positive patients, the peripartum period is one of high risk for thromboembolic disease. Women who have antibody to the SS-A or SS-B antigen are at risk of delivering children with the neonatal lupus syndrome.

CHRONIC DISEASE AND COMPLICATIONS

The major treatment issue for long-term lupus patients is the prevention or management of damage to the arteries, kidneys, bones, and brain rather than the control of immune response and inflammation. The physician must anticipate chronic effects of both the disease and its therapy.

During treatment with high-dose corticosteroids, with or without immunosuppressive agents, avoidance of infections is a

primary concern. Herpes zoster, tuberculosis, and a variety of bacterial infections are the primary threats. *Pneumocystis jiroveci* (formerly *Pneumocystis carinii*) infection is seen relatively infrequently; most rheumatologists do not routinely suggest prophylaxis against this organism. Complications of long-term corticosteroid therapy include osteoporosis (see below) and cataracts, cutaneous striae, cutaneous hemorrhage, diabetes, and oral and vaginal candidiasis. In patients with long-standing disease, these complications produce as much morbidity as the disease itself.

Antiphospholipid Antibody Syndrome

Treatment of antiphospholipid antibody syndrome is anticoagulation to an international normalized ratio (INR) of 2.0 to 3.0. Warfarin and low-dose aspirin are used to prevent thrombotic manifestations of the syndrome; heparin or low-molecular-weight heparin is used in pregnant patients.[59] Recent data suggest that the addition of statin drugs, to downregulate endothelial activation, may also be of benefit.[60]

Atherosclerosis

Early-onset, severe atherosclerosis is a common problem in patients with long-standing SLE. Atherosclerosis most commonly presents as coronary and cerebral artery occlusion; peripheral vascular occlusion also occurs. The cause is unknown, but chronic inflammation, corticosteroid therapy, uncontrolled hypertension, diabetes, smoking, and other factors have been implicated. Most specialists in this area recommend early and vigorous treatment of known risk factors in all lupus patients. The atherosclerosis of lupus is managed in the same manner as atherosclerosis in other situations, except that vascular interventions in patients with antiphospholipid antibody syndrome are hazardous.

Osteonecrosis

Although the mechanism of osteonecrosis is not clearly known, many authorities believe that a steroid-induced increase in the volume of lipocytes increases pressure in the bone marrow, cutting off blood flow to the vulnerable areas. Consequently, if osteonecrosis is recognized before the joint has collapsed (usually by bone scan or MRI), trephining the bone to reduce intraosseous pressure (so-called core decompression) has been recommended. However, the validity of this theory and the efficacy of the treatment remain unproved. Usually, joint replacement is eventually required.[61]

Osteoporosis

Osteoporosis follows long-term corticosteroid therapy with sufficient frequency that all patients receiving such therapy should receive prophylaxis for this complication. High-dose oral calcium (i.e., 1,500 mg daily), vitamin D, a bisphosphonate drug, and parathyroid hormone are the primary preventive measures; estrogen replacement may be considered in postmenopausal women who do not have antiphospholipid antibody. Weight-bearing exercise should be encouraged. Other prophylactic measures, including calcitonin and parathyroid hormone, may be appropriate. Because lupus patients are photosensitive, increased sun exposure to prevent osteoporosis is unwise. Bisphosphonates should not be used in women anticipating pregnancy.

Cardiac Disease

Inflammatory cardiomyopathy responds to corticosteroids; ischemic cardiomyopathy does not. Valvular insufficiencies and thromboemboli are late complications of lupus cardiac disease. Bacterial endocarditis rarely complicates this abnormality. Valvulitis generally does not respond to treatment, although it has been reported that acute valvulitis will respond to corticosteroid therapy.[62] Small numbers of patients require valve replacement, usually of the aortic or mitral valve. The mechanism of valvulopathy in SLE is unknown, as are methods of prevention.

PALLIATIVE CARE

A common mistake in the treatment of SLE is to assume that a given complaint reflects ongoing inflammatory disease, rather than irreversible damage, and that it can be controlled with anti-inflammatory and immunosuppressive therapy rather than with palliation. Examples of such symptoms include seizures, dementia, and other neurologic syndromes associated with brain infarcts; cutaneous ulcers caused by long-standing vascular insufficiency; embolic phenomena from atherosclerosis; respiratory insufficiency from pulmonary fibrosis or pulmonary hypertension; arthritis from osteonecrosis, erosive rheumatoid-like arthritis, or tendinosis; and progressive renal insufficiency from arteriolonephrosclerosis, interstitial fibrosis, or glomerulosclerosis.

Dementia

Chronic neurologic disease, often in the form of dementia, is a long-term sequela of SLE. Some causes are stroke (atherosclerotic, hypertensive, or thrombotic associated with antiphospholipid antibody), autoantibody attack on specific brain targets, small vessel occlusive disease, and drugs. Occasionally, patients are severely disabled. No effective prophylaxis is known.

Renal Failure

Patients with renal disease often need angiotensin-converting enzyme inhibitors for proteinuria, antihypertensives for hypertension, erythropoietin for anemia, and diuretics for edema. Renal failure in lupus occurs in three modes. In the first, acute inflammatory nephritis is characterized by distinctly abnormal urinalyses and clinically and serologically evident disease activity. Patients with this type of renal picture have rapidly rising serum creatinine levels and enter renal failure early after diagnosis. If treated aggressively, with high-dose corticosteroids and immunosuppressive drugs, renal failure will be reversible in approximately one third of these patients.[63] In the second mode of renal failure, which occurs only occasionally, patients will have renal failure from drug toxicity—usually NSAIDs—or acute tubular necrosis. Other manifestations of SLE are modified in uremia: rash is less prominent; and fever, cachexia, mucosal ulcers, and cytopenias are more prominent.

In the third mode, which is the most common, lupus patients enter renal failure slowly after many years of disease. Characteristically, at the time renal failure first appears, the patient has little systemic illness and has had months to years of modest renal insufficiency (with creatinine clearance at 10 to 30 ml/min and serum creatinine below 3.5 mg/dl), slowly rising serum creatinine levels, relatively noninflammatory urinary sediments, and progressive anemia. Then, over a few months, the patient develops hypertension and fluid retention with or without cardiac failure. Abdominal pain is frequently present. Ultrasound shows small, fibrotic kidneys or thin, scarred renal cortices. Renal failure is only transiently reversible in such patients. Aggressive immunosuppressive therapy is not helpful at this stage; on the contrary, it may hasten renal deterioration and will complicate initiation of dialysis.

Less commonly, renal failure is caused by antiphospholipid antibody-associated thrombotic microangiopathy. In these cases, the presentation comprises modest proteinuria with bland urine sediment and slowly rising serum creatinine levels, often with moderate hypertension. Lupus vasculitis involving the kidneys tends to be abrupt in onset, with severely abnormal urinalyses, severe hypertension, and rapidly progressive renal failure. This complication is treated with high-dose corticosteroids and immunosuppression. However, full recovery is uncommon.

Lupus patients, particularly those who enter renal failure slowly, tolerate dialysis and renal transplantation well. However, preexisting cardiac, cerebral, and osteoarticular damage may be limiting factors. Patients who enter renal failure acutely often have other active systemic disease, with seizures and cytopenias being the most common. The common belief that lupus becomes inactive in renal failure is likely not true. Rather, in the majority of patients who enter dialysis, the lupus was already inactive systemically and remains so; probably, the renal failure results not from continuing disease but from progressive scarring. In the minority of patients who enter renal failure during an acute systemic flare, usually early in the course of the illness, active systemic disease tends to continue, and it represents a relative contraindication to renal transplantation. Patients on dialysis who have active SLE usually have high anti-DNA antibody levels and low complement levels. They respond to corticosteroid therapy, usually at lower doses than patients who are not on dialysis.

Prognosis

Prognosis in SLE has four elements: immediate prognosis for life, immediate prognosis for individual organ systems, and long-term prognosis for organ systems and for life.

During the early phases of lupus, complete reversal of almost all manifestations (i.e., rash, arthritis, fever, cytopenias, and nephritis) with aggressive therapy is expected. Exceptions include the scarring discoid rash, brain or spinal cord infarcts, and severe nephritis or nephrosis. In rare cases (< 5%), patients have such severe disease that despite treatment, they rapidly progress to death within 2 years. Because most newly diagnosed patients respond to therapy, 5-year survival is 80% to 90%, 10-year survival is 70% to 90%, and 20-year survival is nearly 70%.[6] Determinants of survival are age, renal disease, and race, with African Americans having lower overall survival than whites.[64]

One organ system (e.g., kidneys or platelets) may fail to respond completely to treatment but not directly threaten the patient's life. Such patients may be monitored without intervention. A patient may develop chronic renal insufficiency (e.g., serum creatinine, 3.0 mg/dl) or have persistent thrombocytopenia (platelet count, 30,000/mm^3) and be considered to be in remission and in no need of treatment.

Long-term prognosis is a function of organ damage from either SLE or its treatment. Pulmonary fibrosis and pulmonary hypertension respond poorly to therapy, although new protocols, such as bosentan or infusion of prostacyclin, may be useful for pulmonary hypertension. Patients with cardiopulmonary failure who have no other system disease limitations are candidates for heart and lung transplantation. Lupus patients with renal failure are candidates for dialysis and for renal transplantation.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Zonana-Nacach A, Barr SG, Magder LS, et al: Damage in systemic lupus erythematosus and its association with corticosteroids. Arthritis Rheum 43:1801, 2000

2. Mimouni D, Nousari CH: Systemic lupus erythematosus and the skin. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 855

3. Mongey A-B, Hess EV: Drug and environmental lupus: clinical manifestations and differences. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 1121

4. Lockshin MD: Antiphospholipid syndrome. Kelley's Textbook of Rheumatology. Ruddy S, Harris ED Jr, Sledge C, Eds.WB Saunders, Philadelphia, 2001, p 1145

5. Buyon JP: Neonatal lupus syndromes. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 449

6. Gladman DD: Epidemiology of systemic lupus erythematosus. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 697

7. Karlson EW, Daltroy LH, Lew RA, et al: The relationship of socioeconomic status, race, and modifiable risk factors to outcomes in patients with systemic lupus erythematosus. Arthritis Rheum 40:47, 1997

8. Hahn BH: Antibodies to DNA. N Engl J Med 338:1359, 1998

9. Balow JE, Boumpas DT, Austin HA III: Systemic lupus erythematosus and the kidney. Systemic Lupus Erythematosus, 3rd ed. Lahita RG, Ed. Academic Press, San Diego, 1999, p 657

10. Bonfa E, Golombek SJ, Kaufman LD, et al: Association between lupus psychosis and anti–ribosomal P protein antibodies. N Engl J Med 317:265, 1987

11. Harrison MJ, Ravdin LD, Volpe B, et al: Anti-NR2 antibody does not identify cognitive dysfunction in a general SLE population. Arthritis Rheum 50:s596, 2004

12. Walport MJ: Lupus, DNAse, and defective disposal of cellular debris. Nat Genet 25:135, 2000

13. Blanco P, Palucka AK, Gill M, et al: Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. Science 294:1540, 2001

14. Davies KA, Robson MG, Peters AM, et al: Defective Fc-dependent processing of immune complexes in patients with systemic lupus erythematosus. Arthritis Rheum 46:1028, 2002

15. Marrack P, Kappler J, Kotzin BL: Autoimmune disease: why and where it occurs. Nat Med 7:899, 2001

16. Davidson A, Diamond B: Autoimmune diseases. N Engl J Med 345:340, 2001

17. Crow MK, Kirou KA: Interferon-alpha in systemic lupus erythematosus. Curr Opin Rheumatol 16:541, 2004

18. Reveille JD: Major histocompatibility complex class II and non–major histocompatibility complex genes in the pathogenesis of systemic erythematosus. Systemic Lupus Erythematosus, 3rd ed. Lahita RG, Ed. Academic Press, San Diego, 1999, p 67

19. Harley JB, Moser KL, Gaffney PM, et al: The genetics of human systemic lupus erythematosus. Curr Opin Immunol 10:690, 1998

20. Abel G, Agnello V: Complement deficiency and systemic lupus erythematosus. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 173

21. Zuniga R, Ng J, Peterson MG, et al: Low binding alleles of Fc gamma receptor types IIa and IIIa are inherited independently and are associated with systemic lupus erythematosus in Hispanic patients. Arthritis Rheum 44:361, 2001

22. James JA, Kaufman KM, Farris AD, et al: An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. J Clin Invest 100:3019, 1997

23. Arbuckle MR, McClain MT, Rubertone MV, et al: Development of autoantibodies before the clinical onset of systemic lupus erythematosus. N Engl J Med 349:1526, 2003

24. Petri M, Robinson C: Review: oral contraceptives and systemic lupus erythematosus. Arthritis Rheum 40:797, 1997

25. Lockshin MD, Sammaritano LR, Schwartzman S: Lupus pregnancy. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 659

26. Ruiz-Irastorza G, Lima F, Alves J, et al: Increased rate of lupus flare during pregnancy and the puerperium: a prospective study of 78 pregnancies. Br J Rheumatol 35:133, 1996

27. Bynoe MS, Grimaldi CM, Diamond B: Estrogen up-regulates Bcl-2 and blocks tolerance induction of naive B cells. Proc Natl Acad Sci USA 97:2703, 2000

28. Cooper GS, Dooley MA, Treadwell EL, et al: Hormonal, environmental, and infectious risk factors for developing systemic lupus erythematosus. Arthritis Rheum 41:1714, 1998

29. Alarcon GS, McGwin G Jr, Bartolucci AA, et al: Systemic lupus erythematosus in three ethnic groups: IX. Differences in damage accrual. LUMINA Study Group. Arthritis Rheum 44:2797, 2001

30. Salmon JE, Roman MJ: Accelerated atherosclerosis in systemic lupus erythematosus: implications for patient management. Curr Opin Rheumatol 13:341, 2001

31. Hochberg MC: Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus (letter). Arthritis Rheum 40:1725, 1997

32. Lockshin MD, Sammaritano LR, Schwartzman S: Brief report: validation of the Sapporo Criteria for antiphospholipid antibody syndrome. Arthritis Rheum 43:440, 2000

33. Lahita RG: The clinical presentation of systemic lupus erythematosus in adults. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 435

34. Lawrence EC: Systemic lupus erythematosus and the lung. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 961

35. Roldan CA, Shively BK, Crawford MH: Echocardiographic study of valvular heart disease associated with lupus erythematosus. N Engl J Med 335:1424, 1996

36. Lockshin MD, Salmon JE, Roman MJ: Atherosclerosis and lupus: a work in progress (editorial). Arthritis Rheum 44:2215, 2001

37. Roman MJ, Shanker BA, Davis A, et al: Prevalence and correlates of accelerated atherosclerosis in systemic lupus erythematosus. N Engl J Med 349:2399, 2003

38. Bywaters EGL: Jaccoud's syndrome: a sequel to the joint involvement of systemic lupus erythematosus. Clin Rheum Dis 1:125, 1975

39. Mont MA, Glueck CJ, Pacheco IH, et al: Risk factors for osteonecrosis in systemic lupus erythematosus. J Rheumatol 24:645, 1997

40. Nagasawa K, Tsukamoto H, Tada Y, et al: Imaging study on the mode of development and changes in avascular necrosis of the femoral head in systemic lupus erythematosus: long-term observations. Br J Rheumatol 33:343, 1994

41. Dotan I, Mayer L: Nonhepatic gastrointestinal manifestations of systemic lupus erythematosus. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 975

42. Mackay IR: Hepatic disease and systemic lupus erythematosus: coincidence or convergence. Systemic Lupus Erythematosus, 4th ed. Lahita RG, Ed. Academic Press, San Diego, 2004, p 993

43. The American College of Rheumatology Nomenclature and Case Definitions for Neuropsychiatric Lupus Syndromes. ACR Ad Hoc Committee on Neuropsychiatric Lupus Nomenclature. Arthritis Rheum 42:599, 1999

44. Hanly JG: Evaluation of patients with CNS involvement in SLE. Baillière's Clin Rheumatol 12:415, 1998

45. Harrison MJ, Gershengorn J: Despite low rate of global impairment in SLE patients, cognitive performance is less than expected (abstr). Arthritis Rheum 44:s197, 2001

46. Kozora E, Thompson LL, West SG, et al: Analysis of cognitive and psychological deficits in systemic lupus erythematosus patients without overt central nervous system disease. Arthritis Rheum 39:2035, 1996

47. Sanna G, Piga M, Terryberry JW, et al: Central nervous system involvement in systemic lupus erythematosus: cerebral imaging and serological profile in patients with and without overt neuropsychiatric manifestations. Lupus 9:573, 2000

48. Gladman DD, Urowitz MB, Slonim D, et al: Evaluation of predictive factors for neurocognitive dysfunction in patients with inactive systemic lupus erythematosus. J Rheumatol 27:2367, 2000

49. Weening JJ, D'Agati VD, Schwartz MM, et al: The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int 65:521, 2004

50. Clancy RM, Kapur RP, Molad Y, et al: Immunohistologic evidence supports apoptosis, IgG deposition, and novel macrophage/fibroblast crosstalk in the pathologic cascade leading to congenital heart block. Arthritis Rheum 50:173, 2004

51. Wilson WA, Gharavi AE, Koike T, et al: International consensus statement on preliminary classification criteria for definite antiphospholipid syndrome: report of an international workshop. Arthritis Rheum 42:1309, 1999

52. Kozora E, West SG, Kotzin BL, et al: Magnetic resonance imaging abnormalities and cognitive deficits in systemic lupus erythematosus patients without overt central nervous system disease. Arthritis Rheum 41:41, 1998

53. Sibbitt WL Jr, Sibbitt RR, Brooks W: Neuroimaging in neuropsychiatric systemic lupus erythematosus. Arthritis Rheum 42:2026, 1999

54. West SG, Emlen W, Wener MH, et al: Neuropsychiatric lupus erythematosus: a ten-year prospective study on the value of diagnostic tests. Am J Med 99:153, 1995

55. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. Canadian Hydroxychloroquine Study Group. N Engl J Med 324:150, 1991

56. Liang MH, Socher SA, Roberts WN, et al: Measurement of systemic lupus erythematosus activity in clinical research. Arthritis Rheum 31:817, 1988

57. Buyon JP, Kalunian KC, Ramsey-Goldman R, et al: Assessing disease activity in SLE patients during pregnancy. Lupus 8:677, 1999

58. Gladman DD, Goldsmith CH, Urowitz MB, et al: The Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index for systemic lupus erythematosus international comparison. J Rheumatol 27:373, 2000

59. Khamashta MA, Cuadrado MJ, Mujic F, et al: The management of thrombosis in the antiphospholipid-antibody syndrome. N Engl J Med 332:993, 1995

60. Meroni PL, Raschi E, Testoni C, et al: Statins prevent endothelial activation induced by antiphospholipid (anti-beta$_2$-glycoprotein I) antibodies: effect on the proadhesive and proinflammatory phenotype. Arthritis Rheum 44:2862, 2001

61. Mont MA, Fairbank AC, Petri M, et al: Core decompression for osteonecrosis of the femoral head in systemic lupus erythematosus. Clin Orthop 334:91, 1997

62. Nesher G, Ilany J, Rosenmann D, et al: Valvular dysfunction in antiphospholipid syndrome: prevalence, clinical features and treatment. Semin Arthritis Rheum 27:27, 1997

63. Kimberly RP, Lockshin MD, Sherman RL, et al: Reversible "end-stage" lupus nephritis: analysis of patients able to discontinue dialysis. Am J Med 74:361, 1983

64. Trends in deaths from systemic lupus erythematosus—United States, 1979–1998. MMWR Morb Mortal Wkly Rep 51(17):371, 2002

Acknowledgment

Figure 2b © 2002 Hospital for Special Surgery. http://www.Rheumatology.HSS.edu All rights reserved. Used by permission.

V SCLERODERMA AND RELATED DISEASES

George Moxley, m.d.

Scleroderma

definition and classification

Scleroderma, or systemic sclerosis, is a rare, slowly progressive rheumatic disease characterized by deposition of fibrous connective tissue in the skin and other tissues. It is accompanied by vascular lesions, especially in the skin, lungs, and kidneys. No cure is known.

Scleroderma may be either systemic or localized, with the systemic illness being either diffuse or limited. The limited form of systemic scleroderma, or CREST syndrome (calcinosis, Raynaud phenomenon, esophageal involvement, sclerodactyly, and telangiectasias), involves internal organs less often than diffuse scleroderma. Except when pulmonary hypertension is present, patients with the limited form have a better prognosis than those with the diffuse form [*see Table 1*]. Systemic scleroderma can be fatal.

Localized scleroderma is confined to the skin, subcutaneous tissue, and muscle and is not accompanied by the Raynaud phenomenon, acrosclerosis, or visceral involvement. There are two forms of localized scleroderma: morphea, which presents as variable-sized plaques of skin induration, and linear, which presents as bands of skin induration on the face or a single extremity. Linear scleroderma may be associated with muscle atrophy and involvement of the underlying bone. It usually afflicts children or young adults and may lead to significant growth impairment of the involved part. In the morphea version, the lesions may persist for months or years, after which improvement may occur [*see Table 1*]. Although there is a possibility of disfigurement, localized scleroderma is not a severe illness, and patients with the disease generally have a normal life span.

epidemiology

The Raynaud phenomenon is associated with scleroderma [*see* Diagnosis, Clinical Manifestations, *below*]. Primary Raynaud phenomenon (Raynaud phenomenon without underlying illness) is quite common, with up to 30% of young women having episodes. A meta-analysis showed that the transition rate to a defined inflammatory rheumatic disease (e.g., scleroderma or lupus) is 3.2 per 100 patient-years of observation; the eventual development of an inflammatory rheumatic disease occurred in 12.6% of individuals. The best predictor of development of inflammatory disease is an abnormal nailfold capillary pattern, which has a predictive value of 47%; a positive antinuclear antibody test has a predictive value of 30%.[1] Racial, genetic, and environmental factors have been proposed as influences of scleroderma risk and disease pattern.[2]

The overall global incidence of scleroderma is approximately 17 to 19 per one million population per year, with higher rates for women than for men. African Americans experience diffuse scleroderma more often than other ethnic groups; whites are more often diagnosed with the CREST variant.[3] The prevalence in the United States is approximately 24 per 100,000 population[3]—fourfold to ninefold the prevalence in other countries. Mortality factors include diffuse disease, older age at onset, and internal organ involvement, particularly pulmonary and renal involvement. Some investigators have suggested that scleroderma is associated with various environmental exposures.[4] For example, workers exposed to polyvinylchloride may experience the Raynaud phenomenon and scleroderma-like skin thickening.[5] However, no substance has yet been convincingly linked with scleroderma.[5] Silicone breast implants have not been found to be associated with scleroderma.[6]

etiology

The etiology of scleroderma is largely unknown. The disease shows familial aggregation consistent with a genetic component, and family history is the strongest risk factor.[3] In the United States, about 1.6% of relatives are affected, an over 60-fold higher prevalence than in the general population (0.026%).[3] One etiologic factor may be a fibrillin defect. Fibrillin is a macromolecule that is a component of elastic fibers, and it is defective in Marfan syndrome. In an animal model (the tight-skin mouse, *tsk1*), a scleroderma-like condition is caused by an insertion into the fibrillin gene that apparently encodes a latent binding region for the transforming growth factor–β (TGF-β) cytokine. One proposed explanation is that the abnormal fibrillin binds an increased number of fibroblast growth factors that influence nearby fibroblasts. The current knowledge extends to Japanese and Oklahoma Choctaw scleroderma patients, who likely have a scleroderma-related genetic defect (e.g., secreted protein acidic and rich in cystein [SPARC]) in the fibrillin chromosomal region either in or near the gene.[7] However, other genes (e.g., *IL-1A, TGF-β,* and *IL-4*) or environmental exposures may also be involved.

pathophysiology and pathogenesis

The common pathologic features of tissues with scleroderma involvement are progressive fibrosis, vascular abnormalities, and inflammation. Fibrosis involves an accumulation of excessive collagen and other extracellular matrix constituents, such as glycosaminoglycans and fibronectin. The vascular abnormalities are intimal hyperplasia with collagen deposition and adventitial fibrosis, capillary dropout, dilatation, tortuosity, and fibrotic atherosclerosis. The inflammatory changes may include cellular infiltration. These pathologic characteristics are believed to be the result of three or more interacting components: autoimmunity, an endothelial abnormality, and a skin fibroblast lesion. An alternative explanation is that these characteristics are the result of a disease process akin to graft versus host disease (GVHD).

Table 1 Classification of Scleroderma

Form	*Syndrome*
Scleroderma (systemic sclerosis)	Diffuse skin involvement Limited skin involvement (CREST syndrome) Overlapping features of mixed connective tissue disease
Localized scleroderma	Morphea: single or multiple plaques or generalized lesions Linear scleroderma

CREST—calcinosis, Raynaud phenomenon, esophageal involvement, sclerodactyly, and telangiectasias

Immune Cell and Cytokine Abnormalities

Activated thymus-derived lymphocytes predominate among the cells that infiltrate involved tissues; activated inflammatory cells are also present. Such immune and inflammatory cells release a plethora of cytokines and soluble mediators. Of particular note are cytokines influencing fibroblast function: interleukin-1 (IL-1), IL-4, IL-6, IL-8, RANTES (regulated upon activation, normal T cell expressed and secreted), tumor necrosis factor (TNF), and TGF-β. TNF and interferon gamma are antifibrotic,[8] but the action of TGF-β results in a pattern of tissue damage similar to the pattern seen in scleroderma. TGF-β stimulates fibroblasts and vascular smooth muscle cells to make collagen, and it stimulates endothelial cells to make endothelin-1, which, in turn, causes vasoconstriction and collagen production. B cells are also activated in patients with scleroderma, and some autoantibodies are relatively specific for scleroderma. Whether such humoral responses result in scleroderma tissue damage is under investigation.

Fibroblast Abnormalities

As with lymphocytes and macrophages showing activation in scleroderma, fibroblasts are also metabolically activated. Such activated cells overproduce collagen, other extracellular matrix molecules, and cellular adhesion molecules.[9] The reasons why collagen gene expression is increased and sustained have not been fully explained. One possible explanation is that the presence of cytokines such as IL-4 and TGF-β strongly foster collagen production.[8] TGF-β acts in large part through specific receptors that lead to activation of Smads, a family of second messenger/transcription factor proteins.[10] Some genetic studies have supported the proposition that scleroderma is associated with TGF-β markers.[11]

Vascular Abnormalities

Vessels in involved tissues show disrupted pattern and function, characterized by altered endothelial permeability, adhesion of platelets and leukocytes to endothelium, and the presence of inflammatory cells. One current pathogenetic theory is that endothelial homeostasis is disrupted, leading to increased endothelin-1 production, reduced prostacyclin release, and enhanced coagulation.[12] Endothelial cells express increased numbers of adhesion molecules necessary for inflammatory cell adhesion and extravasation. Capillaries become obliterated but are not replaced; involved tissues may become ischemic and then reperfused. The Raynaud phenomenon [*see* Diagnosis, Clinical Manifestations, *below*] is a hallmark vascular lesion that likely represents an exaggerated response of a stiff vessel wall to a typical environmental exposure. Microscopic analysis of the nailfolds of patients with scleroderma and Raynaud phenomenon reveals capillary disappearance and dilatation. In scleroderma, small arteries develop concentric intimal fibrosis and thus become narrow; this narrowing in turn greatly increases the vascular reactivity to alpha$_2$-adrenergic agents.

Chimerism and Microchimerism

Chimerism denotes a state in which a person has cells derived from two or more other people. Male DNA consistent with microchimerism has been found more frequently and in higher concentrations in the circulation of women with scleroderma who have had a previous male delivery than in women without scleroderma who have also had a male delivery. In addition, microchimerism has been found in tissues of mothers with scleroderma but not in the tissues of healthy mothers.[13] However, cause and effect are not yet established,[14] and parous women are reported to have a reduced scleroderma risk (odds ratio, 0.3).[15]

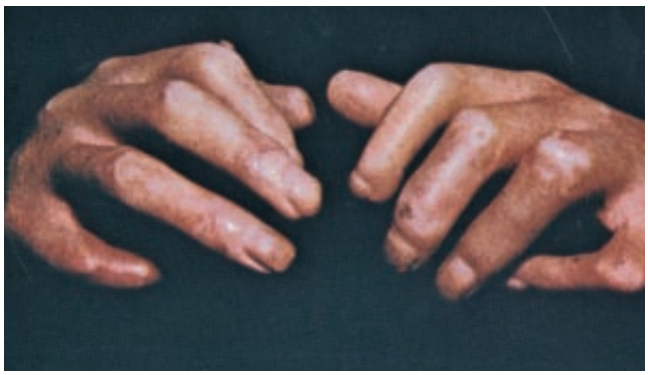


Figure 1 **Severe involvement of the hands in a patient with long-standing scleroderma includes flexion contractures of the fingers related to fibrosis of the skin and of subcutaneous tissues. Increased pigmentation has occurred, and melanin loss (vitiligo) is evident in some areas. The distal aspects of the terminal phalanges in some fingers have undergone resorption or shortening. This process, termed autoamputation, usually occurs without ulceration of the terminal digit; the mechanism is unknown.**

DIAGNOSIS

Clinical Manifestations

Skin The first signs of scleroderma in the skin are swelling and thickening of the fingers and hands, with or without involvement of the face; later in the illness, other areas of skin may become thickened. Involvement of the trunk and arms proximal to the elbows is associated with visceral involvement and a poorer prognosis. The skin continues to thicken during the first 2 to 3 years after the onset of disease; the thickening then ceases and may recede, giving the impression that the skin is softening. In subsequent years, skin atrophy occurs, with con-

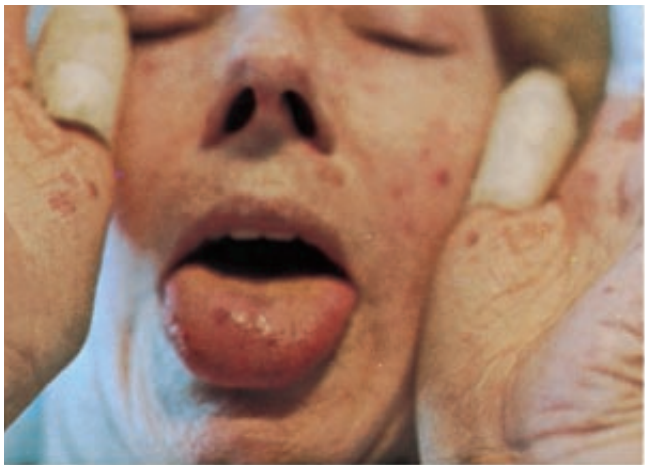


Figure 2 **Telangiectasias appear on the hands, face, and tongue in a patient with the CREST (calcinosis, Raynaud phenomenon, esophageal involvement, sclerodactyly, and telangiectasias) variant of scleroderma. Thumbs are bandaged because of chronic ulcerations associated with the Raynaud phenomenon.**

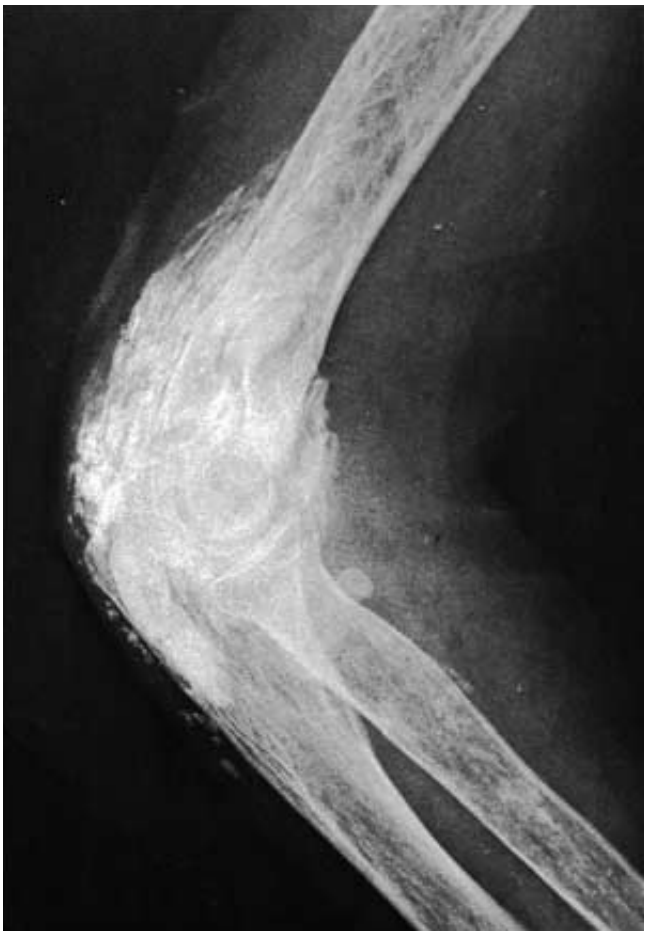


Figure 3 **In scleroderma, extensive calcinosis (hydroxyapatite crystal deposition) may be found in connective tissues and around joints. If extensive, calcinosis is usually associated with at least partial loss of joint motion.**

comitant loss of hair, sebaceous glands, and sweat glands, as well as a loss of pliability. In addition, the skin becomes hidebound—tightly drawn and bound to underlying structures.

Skin involvement is often most prominent in the hands and fingers (sclerodactyly); frequently, the face is also affected. A tightening of facial skin results in decreased skin lines, a pursed appearance, and a diminution in the oral aperture. The skin tightness may limit mobility, especially in the fingers. Flexion contractures may also develop in the fingers [*see Figure 1*]. Several other skin abnormalities may accompany these changes. Telangiectasias occur frequently and may be numerous [*see Figure 2*]. They are often most prominent on the face, hands, and oral mucosa. Calcinosis—the deposition of hydroxyapatite crystals in subcutaneous areas—may be limited or widespread and is usually located around joint capsules [*see Figure 3*]. Skin ulceration over calcific deposits may lead to drainage of a white material with a consistency resembling toothpaste. A diffuse increase in melanotic pigmentation may extend over the entire skin surface; areas of hypopigmentation are also commonly seen.

The Raynaud phenomenon is an episodic manifestation of numbness or pain accompanied by a two- or three-phase color change in the digits; these changes are triggered by cold temperatures or emotional stress and are relieved by warming the involved part. In severe cases, however, the relation to ambient temperature is sometimes less obvious. The episode typically begins with pallor, followed by cyanosis and, finally, by redness caused by reactive hyperemia [*see Figure 4*]. Prolonged ischemia may lead to painful digits, ulceration, and even gangrene. Almost all patients (95%) with diffuse or limited scleroderma experience the Raynaud phenomenon. The Raynaud phenomenon is also seen in patients with other disorders, including other autoimmune diseases such as systemic lupus erythematosus (SLE), polymyositis, and several forms of vasculitis; in patients who are receiving certain drugs, such as bleomycin, ergot derivatives, beta blockers, and methysergide; and after occupational exposure to vinyl chloride, cold temperatures, and vibrating tools.

In scleroderma, the Raynaud phenomenon is a manifestation of vasculopathy involving small arteries and capillaries; it occurs not only in the extremities but also in some involved viscera, such as the lungs and kidneys. Patients with scleroderma and the Raynaud phenomenon have characteristic capillary changes on nailfold microscopy. Nailfold microscopy consists of observation of the capillary structure of the periungual tissues with a handheld magnifying lens, such as that used in a standard ophthalmoscope. Patients with underlying scleroderma may exhibit loss of some capillaries and dilatation of capillaries in other nailfold areas. Such changes often occur in asso-

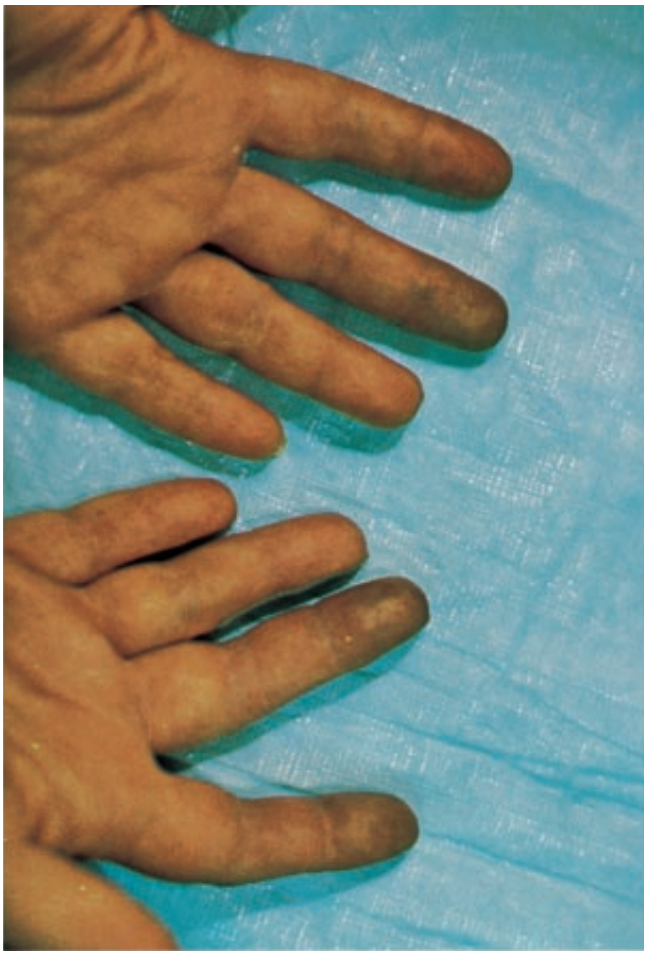


Figure 4 **Vascular pathology usually manifests itself as the Raynaud phenomenon in patients with scleroderma. The cyanotic phase of the Raynaud phenomenon often involves the distal two thirds of the second and third fingers of both hands.**

Table 2 Antinuclear Antibodies in Scleroderma

Immunofluorescent Pattern of Antinuclear Antibody Staining	*Antigens*	*Clinical Pattern*	*Approximate Frequency (%)*	*Specificity*
Nucleolar	Nuclear ribonucleoproteins (nRNPs)	Diffuse or limited scleroderma	50	Moderate
	RNA polymerases I, II, and III	Diffuse scleroderma	23	High
	Nuclear proteins PM-1 (PM-Scl) and Ku	Scleroderma-polymyositis overlap	< 5	High, for scleroderma and polymyositis
Centromeric (large speckles)*	Centromere proteins (CENP-A, CENP-B, and CENP-C)	Usually in limited scleroderma (CREST syndrome)	50 (of patients with CREST syndrome)	High
Diffuse (fine speckles)	Topoisomerase I (Scl-70)	Diffuse scleroderma	20–33	High
Homogeneous	Histones (mainly H1 and H3)	Localized scleroderma	50	Moderate

*Requires a human epithelial carcinoma cell line (HEp-2).

ciation with internal organ involvement and may be used to predict the development of diffuse scleroderma when visceral involvement is not clinically apparent.

Proposed criteria for early diagnosis of scleroderma include either objective observation or measurement of cold-induced vasospasm, plus either abnormal nailfold microscopy, presence of antibodies directed toward characteristic autoantigens [*see Table 2*],[16] or patient reports of Raynaud phenomenon and the presence of nailfold microscopic changes and autoantibodies.

Musculoskeletal system A mild, usually symmetrical, inflammatory arthritis can occur in scleroderma. Juxta-articular bone erosions occur frequently, especially in the distal interphalangeal joints, but the degree of destruction is usually less than that seen in rheumatoid arthritis. Flexion contractures of the fingers often develop and are most likely related to fibrosis

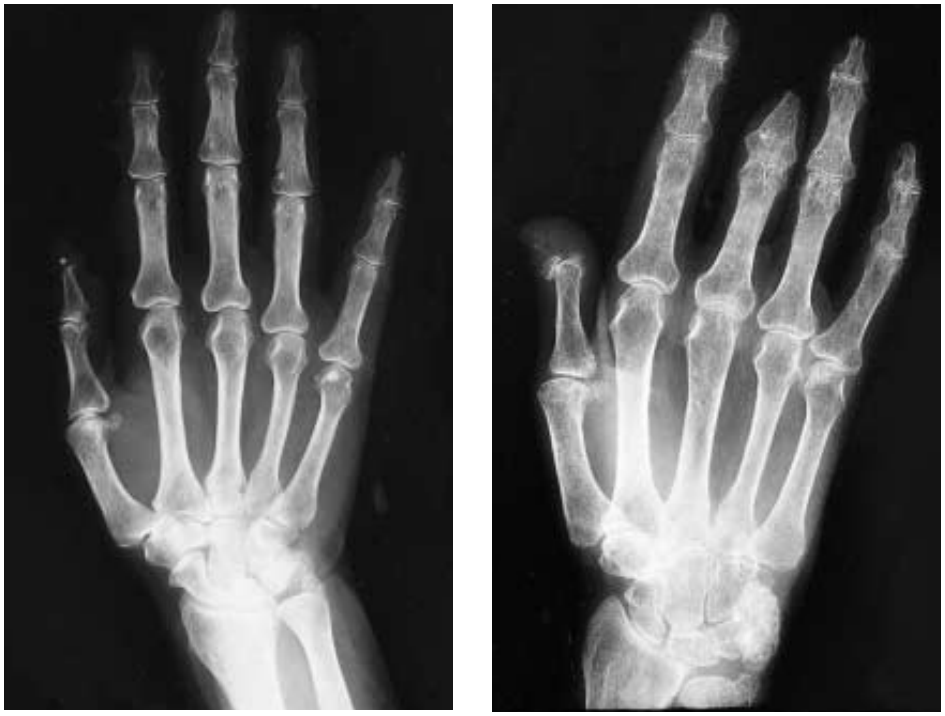


Figure 5 **Over an 8-year period, radiographs taken of the hand of a patient with scleroderma demonstrate a progressive, terminal resorption of the digits. The earlier film (left) shows a loss of the spherical terminal portion of the distal phalanx of the thumb and a small, dense calcific deposit at the terminal aspect of the thumb. After an 8-year period, dramatic changes can be seen (right). An almost complete loss of the terminal phalanx of the thumb and a partial loss of the distal phalanges of the remaining fingers are observed. In addition, the entire distal phalanx of the third finger is lost along with part of the middle phalanx. Loss of the middle phalanx is less common than loss of the distal phalanges. A generalized osteopenia is present in the later stage, which is probably related to osteoporosis of disuse, and a calcific deposit has formed in the ulnar aspect of the wrist. The apparent narrowing of interphalangeal joint spaces may be associated with flexion contractures of the fingers, which are caused by the fibrous thickening of the connective tissues of the hand.**

of the tendons and joint capsules. Crepitus and friction rubs, detected by palpation or auscultation over tendons and bursas, are characteristic findings related to fibrotic changes of underlying tissues. Another skeletal complication is acral osteolysis, which is the resorption of the terminal phalanges and surrounding soft tissue with consequent shortening of the digits [*see Figure 5*]. It may occur without infection or ulceration. Patients with scleroderma may have one or more of a variety of muscle disorders,[17] such as fatigue without objective evidence of muscle damage, a simple myopathy, or clear-cut inflammatory myositis.

GI tract Almost every part of the GI tract may be involved in scleroderma.[18] Sjögren syndrome, which causes dry eyes and dry mouth, occurs in about one third of patients with scleroderma. Esophageal hypomotility, which is demonstrated by cinefluoroscopic examination and manometric studies, occurs in more than 90% of patients with scleroderma, many of whom are asymptomatic [*see Figure 6*]. The absence of normal peristaltic waves in the lower two thirds of the esophagus may cause dysphagia; incompetence of the gastroesophageal sphincter leads to reflux esophagitis and sometimes may result in esophageal stricture. Barrett esophagus, esophageal carcinoma, and candidal esophagitis may ensue. Similarly, hypomotility of the stomach, small intestine, colon, and anorectal area may occur, possibly causing gastroparesis, pseudo-obstruction, colonic impaction, or impaired anorectal function. Telangiectasias may be present in the gastric mucosa and the mucosa of the small intestine and colon. Gas may dissect into the intestinal wall (pneumatosis intestinalis) and leak into the peritoneal cavity, simulating a perforated viscus. Characteristic wide-mouthed diverticula of the colon may develop; these are pathognomonic of scleroderma. The early lesions of the GI tract may be caused by autonomic nerve dysfunction; autonomic nerve dysfunction may in time lead to smooth muscle atrophy and irreversible muscle fibrosis of the gut. Primary biliary cirrhosis and drug-induced hepatitis may also be associated with scleroderma.

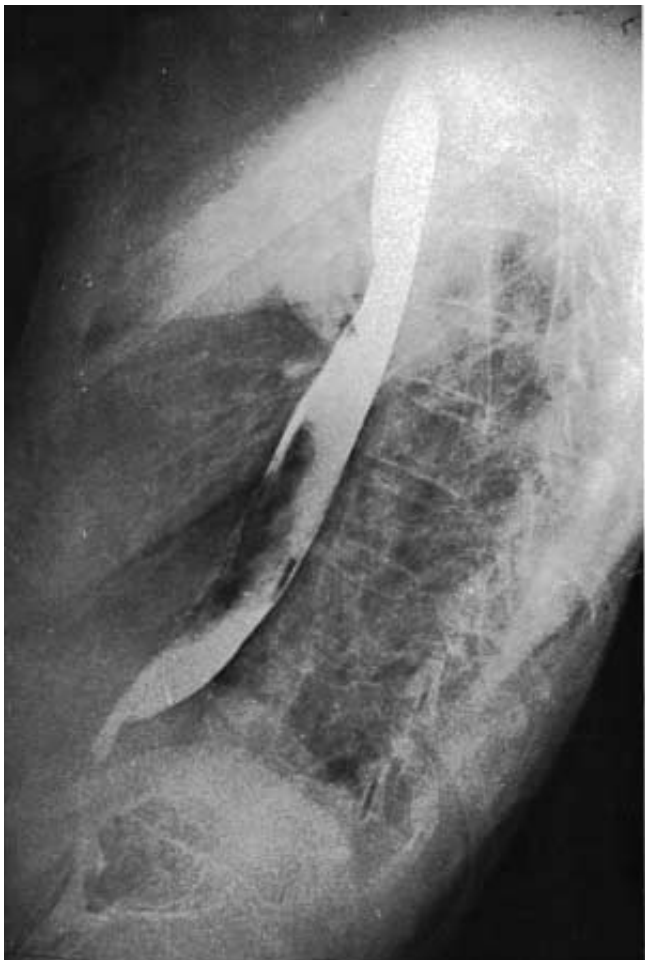


Figure 6 **Hypomotility of the esophagus in scleroderma, which is demonstrated by the lack of peristaltic waves in the barium column, is frequently an asymptomatic finding, but it may be associated with dysphagia and incompetence of the gastric sphincter, which results in reflux esophagitis.**

Lungs Pulmonary involvement represents an important cause of scleroderma-related morbidity and mortality[19]; lung disease is the most frequent cause of scleroderma-related death. Although scleroderma-related findings may range from associated malignancy and silicosis to calcinosis and hemorrhage, the most common findings are pulmonary vascular disease and interstitial inflammation and fibrosis. Isolated pulmonary hypertension is typically found in the CREST syndrome (prevalence is 50% to 65% in the limited cutaneous subset and up to 35% in the diffuse subset)[20]; interstitial pulmonary fibrosis may be found in both limited and diffuse scleroderma (at postmortem examination, the frequency is about 75%).

Isolated pulmonary hypertension usually results in cough, dyspnea, and syncope; it has a severe prognostic outlook (5-year survival < 10%). Pulmonary hypertension is defined by a resting mean pulmonary arterial pressure greater than 25 mm Hg or an exercise-induced mean pulmonary arterial pressure greater than 30 mm Hg; in addition, pulmonary hypertension is often associated with abnormal diffusing capacity for carbon monoxide. In persons with limited cutaneous involvement, decreasing diffusing capacity of the lung for carbon monoxide (D_{LCO}) is an excellent predictor of subsequent pulmonary hypertension; one study of 106 patients reported that at baseline, patients who subsequently developed pulmonary hypertension had a mean pulmonary arterial pressure only slightly higher than patients who did not, but the D_{LCO} in patients with pulmonary hypertension declined from a mean of 80% to a mean of 35%. The early clinical features associated with subsequent pulmonary hypertension included more severe Raynaud phenomenon and digital tip ulcers, as well as positivity for serum autoantibodies directed toward nucleoli.[21]

Interstitial fibrosis is typically accompanied by dyspnea and cough, and the 5-year survival is about 45%. Although chest roentgenograms may show linear and reticular abnormalities, high-resolution CT scanning is favored for the detection of early disease; on CT scans, alveolitis appears as patchy areas with a ground-glass appearance. Interstitial fibrosis appears to result from inflammatory alveolitis: the release of various cytokines and chemokines in the course of the inflammatory process results in fibroblast activation and extracellular matrix remodeling. Association between severe esophageal involvement and interstitial fibrosis suggests that gastroesophageal reflux may contribute to fibrotic changes.[22]

Heart Patients with clinically evident scleroderma-related heart disease have a poor prognosis.[23] The myocardium is

involved in approximately 20% to 25% of clinical cases of systemic scleroderma. Scleroderma-associated myocardial disease is typically characterized by patchy areas of myocardial fibrosis replacing normal muscle; this may cause hypertrophy and diminished cardiac output, particularly with exercise. Diastolic dysfunction[24] and atherosclerotic coronary artery disease are common. Myocardial infarctions may develop in patients with scleroderma who have normal findings on coronary artery catheterization. Myocarditis may also occur in patients with scleroderma-associated inflammatory myositis. The patchy myocardial fibrosis and conduction system involvement may result in various arrhythmias and sudden death. Pericardial disease is common at autopsy, but clinically evident pericardial manifestations occur in only 5% to 16% of cases. Such pericardial disease may cause acute pericarditis, arrhythmias, pericardial effusions, or sudden death. In limited cutaneous scleroderma, apart from a similar frequency of cardiac arrhythmias and conduction defects, heart involvement is generally less frequent and less severe than in diffuse scleroderma.

Kidneys Chronic mild proteinuria and mild hypertension are common effects of scleroderma but typically do not result in significant renal dysfunction. The most significant disease process associated with scleroderma is renal crisis,[25] which consists of the rapid development of malignant hypertension, hyperreninemia, microangiopathic hemolytic anemia, and oliguric renal failure. Renal crisis typically occurs in the scleroderma subset characterized by rapidly progressive diffuse skin disease. Renal crisis was formerly the most common cause of death in patients with scleroderma, but aggressive treatment with angiotensin-converting enzyme (ACE) inhibitors early in the course of disease has greatly improved outcome.

Other organ systems Although scleroderma rarely involves the central nervous system, peripheral neuropathy may occur. The most frequent form is unilateral or bilateral trigeminal neuropathy of one or more of the three trigeminal branches; this neuropathy presents as progressive numbness and pain. Widespread autonomic nervous system dysfunction underlies the propensity for the Raynaud phenomenon and intestinal involvement. Hematopoietic consequences of scleroderma are uncommon.

Laboratory Findings

More than 85% of patients with scleroderma have positive test results for antinuclear antibodies[26] [*see Table 2*]. Antibodies to certain nuclear antigens are specific for scleroderma,[27] and each type of antibody is associated with a particular clinical pattern of disease. Antibodies to Scl-70 and to RNA polymerases (usually RNA polymerase III) are seen in patients with systemic scleroderma. In a meta-analysis, anti–Scl-70 antibodies had a positive predictive value of 70% for diffuse cutaneous scleroderma.[28] Anticentromere antibodies are associated with limited scleroderma (the CREST syndrome); the meta-analysis indicated a positive predictive value of 88% for limited cutaneous scleroderma.[28] Antibodies to nuclear ribonucleoprotein (nRNP) are associated with diffuse and limited scleroderma. Antibodies to PM-1 and Ku are infrequent. The presence of these antibodies is usually associated with overlapping clinical features of polymyositis and scleroderma. About 50% of patients with localized scleroderma have antibodies to histones [*see Table 2*].

MANAGEMENT

Management of scleroderma is often a severe challenge,[29] but it includes several potentially lifesaving interventions. Close attention to detecting renal and pulmonary involvement will typically improve eventual outcomes.

Renal Crisis

Renal crisis is a serious scleroderma-related manifestation that is manageable with appropriate therapy. Onset typically is within the first 2 years of the disease course and is more common in diffuse scleroderma than in the limited form of the disease[30]; renal crisis occurs in about 15% of patients with diffuse scleroderma. A clinician should instruct patients with early diffuse scleroderma about daily blood pressure monitoring and initiate antihypertensive therapy if the patient's blood pressure exceeds 130/80 mm Hg. Renal crisis is recognized by diminished renal function with new-onset hypertension (even of modest degree), microscopic hematuria, and proteinuria. Treatment of this medical emergency with ACE inhibitors and other potent antihypertensive drugs appears to arrest the deterioration in renal function. Even in patients who initially require dialysis, ACE inhibitors may restore renal function enough to make dialysis unnecessary.[31] However, this therapy has been shown to be more beneficial when initiated before serum creatinine levels have exceeded 3 mg/dl. In some patients, renal failure progresses despite good control of blood pressure, necessitating dialysis and possibly renal transplantation. Despite reduced frequency with ACE inhibitor therapy and availability of dialysis, renal crisis is associated with poor survival rates.[32]

Pulmonary Involvement

A clinician should be able to detect pulmonary involvement through yearly or twice-yearly pulmonary arterial pressure estimates (from echocardiograms) and measurements of forced vital capacity, carbon monoxide diffusing capacity, and exercise arterial blood gas levels. Patients with pulmonary function abnormalities may then be evaluated with bronchoalveolar lavage, where available, and high-resolution CT to detect the extent of lower respiratory inflammation. Progression is more likely to occur in patients with more extensive lung disease and patients who are positive for anti-Scl-70 antibodies. In diffuse scleroderma, meticulous assessment is warranted, particularly in the first 5 years after onset.[33] In patients with early fibrosing alveolitis, progression of pulmonary fibrosis may be halted by long-term therapy with cyclophosphamide and low-dose prednisone (< 10 mg/day), but this measure is not yet supported by a randomized, controlled trial.[19,33] Cyclophosphamide therapy is a powerful immunosuppressive medication; its use is sometimes associated with adverse events such as bacterial infections, herpes zoster, varicella-zoster virus infection, interstitial cystitis, and malignancies such as bladder transitional cell carcinoma. Often, intravenous cyclophosphamide is effective and somewhat less toxic than oral regimens.[33]

Pulmonary hypertension may be a particularly serious manifestation of scleroderma. In patients with suspected pulmonary arterial hypertension, on the basis of physiologic findings and serial echocardiographic examinations, a right heart catheter measurement of pulmonary arterial pressure should be obtained. Individuals with pulmonary arterial hypertension should receive influenza and pneumococcal immunizations and be counseled regarding cigarette smoking and other im-

portant exposures. When hypoxemia is present, supplemental oxygen is appropriate.

If serious pulmonary hypertension occurs, the patient should be referred to a specialist for this condition. Referral is important because current therapy is rapidly evolving, the medications used are not part of the usual formulary, and the specialist's experience will lead to reduced risk of complication and toxicity. Internet resources may prove useful in identifying local specialists and support programs available to scleroderma patients [*see Sidebar* Internet Resources for Information on Scleroderma].

The therapeutic measures employed for pulmonary arterial hypertension include the use of calcium channel blockers, anticoagulation, prostaglandin medications, phosphodiesterase V inhibitors such as sildenafil, and endothelin receptor antagonists such as bosentan. In a small fraction of patients, high-dose calcium channel blockers will reduce pulmonary arterial pressure and vascular resistance without causing serious adverse effects such as hypotension[34]; in this subset of patients, it is necessary to administer a diuretic to reduce risk of right-sided heart failure. Warfarin therapy (achieving an international normalized ratio of 2.0 to 2.5) increases survival.[34]

Various prostaglandins are helpful for therapy of pulmonary artery hypertension. The currently available prostaglandins include continuous intravenous epoprostanol (2 ng/kg/min) and continuous subcutaneous treprostinil (1.25 ng/kg/min); other oral and inhaled prostaglandins are under study. Epoprostanol has been shown to prolong life and decrease symptoms and improve hemodynamics[35-37]; however, it is exacting in terms of discomfort and disability. Subcutaneous treprostinil therapy is more convenient and improves symptoms and exercise capacity but is sometimes associated with infusion-site pain.[38] Oral sildenafil appears to improve gas exchange[39] and may act synergistically when administered with an inhaled prostaglandin (e.g., iloprost).[40] The most dramatic therapeutic development has been bosentan, an orally administered endothelin-1 antagonist (125 mg twice daily). Bosentan therapy leads to improved exercise capacity, decreased symptoms, and improved hemodynamic effects, but hepatotoxicity is a concern.[41] However, the most beneficial therapies for severe pulmonary arterial hypertension are costly (annual cost of bosentan, $36,000; epoprostanol, $72,000; and treprostinil, $93,000).

Internet Resources for Information on Scleroderma*

National Institute of Arthritis and Musculoskeletal and Skin Diseases Information Clearinghouse
http://www.niams.nih.gov
This clearinghouse provides information about various forms of arthritis and rheumatic diseases, distributes patient and professional education materials, and refers people to other sources of information.

American College of Rheumatology
http://www.rheumatology.org
This association provides referrals to doctors and health professionals who work on arthritis, rheumatic diseases, and related conditions. The association also provides educational materials and guidelines.

Scleroderma Foundation
http://www.scleroderma.org
The foundation publishes information on scleroderma and offers patient education seminars, support groups, physician referrals, and information hotlines.

Scleroderma Research Foundation
http://www.srfcure.org
The foundation's goal is to find a cure for scleroderma by funding and facilitating the most promising, highest-quality research and by placing the disease and its need for a cure in the public eye. The foundation distributes patient handbooks and a twice-yearly, research-related newsletter.

Arthritis Foundation
http://www.arthritis.org
The foundation is a voluntary organization devoted to supporting research on arthritis and other rheumatic diseases, including scleroderma. It also provides up-to-date information on treatments, nutrition, alternative therapies, and self-management strategies. Chapters nationwide offer exercise programs, classes, support groups, physician referral services, and free literature.

*Descriptions of Web sites are derived from www.niams.nih.gov.

Cutaneous Involvement

Most patients with the Raynaud phenomenon may be managed with such measures as avoidance of cold exposure. Instructions to wear warm clothing, gloves, and a hat when exposed to a cold environment may be sufficient. Cigarette smokers have a fourfold risk of digital vascular complications in surgical debridement, amputation, or intravenous vasodilator therapy[42]; smoking cessation is thus a key instruction. In addition, patients should avoid vasoconstrictive agents (e.g., decongestants, caffeine, amphetamines, beta blockers, and ergot alkaloids). For persons with attacks of the Raynaud phenomenon so frequent or so severe as to interfere with daily activities or to put them at risk for skin necrosis, pharmacologic therapy may be employed. Low-dose aspirin (81 mg daily) is recommended.[30] The calcium channel blocking agents promote vasodilatation and generally reduce the frequency and severity of Raynaud phenomenon.[43] Although nifedipine in daily doses of 30 to 60 mg has been effective for most patients, approximately one third do not respond, and some patients experience adverse effects. In double-blind, placebo-controlled trials, other calcium channel blockers have also shown efficacy for the Raynaud phenomenon; these agents include amlodipine, 5 to 10 mg daily, and felodipine, 5 to 10 mg daily. Ischemic digital ulcerations may be difficult to treat and may progress to gangrene of the fingertips. Sympathectomy should be reserved for severe ischemic crises.

GI Tract Involvement

Because gastroesophageal reflux is nearly universal in scleroderma, clinicians typically will instruct scleroderma patients to elevate the head of the bed and will administer proton pump inhibitors. Esophageal motility may be enhanced by use of metoclopramide. Esophageal dilatation may be required if strictures are present. Patients with diminished gastric emptying may be instructed regarding the frequent taking of small meals and may be given prokinetic medications such as metoclopramide. Small bowel motility may present as pain, distention, and vomiting; most episodes can be managed by increasing dietary fiber and avoiding medications that affect motility (e.g., opiates). Octreotide is used for severe small bowel dysmotility. For small bowel bacterial overgrowth with malabsorption, empirical antibiotic therapy with ciprofloxacin, metronidazole, doxycycline, or erythromycin is recommended.

Potential Disease-Modifying Therapies

Potential disease-modifying therapies for scleroderma have also been studied. Penicillamine is no longer recommended because of associated toxicity and minimal efficacy,[44] and methotrexate therapy is of uncertain benefit.[45,46] Stem cell transplants are associated with a higher procedure-related mortality but typically lead to decreased skin manifestations and stabilization of lung function.[47,48] A truly effective disease-modifying therapy is not yet available.

CLINICAL COURSE

In the CREST syndrome, skin involvement is relatively limited, usually affecting only the hands and face; the prognosis for patients with this syndrome is generally favorable unless viscera are involved. However, even the limited CREST variant tends to be unremitting and slowly progressive. Many patients experience an indolent course with little change over several years, although the progression may be more rapid. Diffuse scleroderma is highly variable in its course and manifestations; therefore, its rate of progression is difficult to predict. With the exception of some of the sclerodermatous changes in mixed connective tissue disease, diffuse scleroderma rarely remits completely. Involvement of the viscera, such as the heart, lungs, or, particularly, the kidneys, indicates a poor prognosis. Two other features of scleroderma also indicate a poor outcome: (1) active inflammation, as manifested by an elevated erythrocyte sedimentation rate, and (2) evidence of cardiopulmonary disease, renal disease, or both within 1 year after diagnosis.[49] The life-threatening complications of scleroderma—severe skin involvement, pulmonary fibrosis, and renal crisis—usually occur within the first 2 to 5 years after the onset of disease. After this interval, the disease tends to run an indolent course. The 5-year survival for patients with diffuse scleroderma is approximately 50%**,** but pediatric patients with the disease have a much better prognosis.[50]

Eosinophilic Fasciitis

Eosinophilic fasciitis can superficially resemble scleroderma. It is characterized by pain, swelling, and tenderness of the extremities, after which induration of the skin and subcutaneous tissues occurs. Joint motion may be limited, but the Raynaud phenomenon, sclerodactyly, and other manifestations of scleroderma are not seen. Laboratory test abnormalities include peripheral blood eosinophilia, which may be marked; elevation of the erythrocyte sedimentation rate; and hyperglobulinemia. Antinuclear antibody and rheumatoid factor test results are negative. Biopsy specimens of involved areas have shown inflammation and thickening of the fascia deep to the subcutaneous tissues. The skin appears normal, but the underlying deep fascia is infiltrated with lymphocytes, plasma cells, histiocytes, and sometimes eosinophils. Eosinophilic fasciitis seems to be either self-limited or responsive to low doses of glucocorticoids. Its etiology remains unknown, but several cases have been reported after strenuous muscle exertion.

Eosinophilia-Myalgia Syndrome

In 1989, a previously unrecognized syndrome associated with ingestion of contaminated L-tryptophan appeared.[51] The eosinophilia-myalgia syndrome is characterized by peripheral eosinophilia, severe and incapacitating myalgias, and fatigue of several weeks' duration. Dyspnea and cough may also be present. Skin involvement consists of variable rashes, edema, and scleroderma-like changes, usually without the visceral manifestations of scleroderma. Interstitial pulmonary infiltrates, hypoxia, pulmonary hypertension, and hypersensitivity pneumonitis may occur. Polyneuropathy has been described in a pattern of mononeuritis multiplex. Neurocognitive disorders, such as memory disturbances and difficulty in concentration, have been reported. Most patients continue to manifest symptoms from 2 to 4 years after onset but have no new signs of inflammation.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Spencer-Green G: Outcomes in primary Raynaud phenomenon: a meta-analysis of the frequency, rates, and predictors of transition to secondary diseases. Arch Intern Med 158:595, 1998

2. Michet C: Update in the epidemiology of the rheumatic diseases. Curr Opin Rheumatol 10:129, 1998

3. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, et al: Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis Rheum 48:2246, 2003

4. Nietert PJ, Silver RM: Systemic sclerosis: environmental and occupational risk factors. Curr Opin Rheumatol 12:520, 2000

5. Steen VD: Occupational scleroderma. Curr Opin Rheumatol 11:490, 1999

6. Janowsky EC, Kupper LL, Hulka BS: Meta-analyses of the relation between silicone breast implants and the risk of connective-tissue diseases. N Engl J Med 342:781, 2000

7. Ahmed SS, Tan FK: Identification of novel targets in scleroderma: update on population studies, cDNA arrays, SNP analysis, and mutations. Curr Opin Rheumatol 15:766, 2003

8. Widom RL: Regulation of matrix biosynthesis and degradation in systemic sclerosis. Curr Opin Rheumatol 12:534, 2000

9. Sato S: Abnormalities of adhesion molecules and chemokines in scleroderma. Curr Opin Rheumatol 11:503, 1999

10. Varga J: Scleroderma and Smads: dysfunctional Smad family dynamics culminating in fibrosis. Arthritis Rheum 46:1703, 2002

11. Denton CP, Abraham DJ: Transforming growth factor-beta and connective tissue growth factor: key cytokines in scleroderma pathogenesis. Curr Opin Rheumatol 13:505, 2001

12. Schachna L, Wigley FM: Targeting mediators of vascular injury in scleroderma. Curr Opin Rheumatol 14:686, 2002

13. Johnson KL, Nelson JL, Furst DE, et al: Fetal cell microchimerism in tissue from multiple sites in women with systemic sclerosis. Arthritis Rheum 44:1848, 2001

14. Tan FK, Arnett F: Genetic factors in the etiology of systemic sclerosis and Raynaud phenomenon. Curr Opin Rheumatol 12:511, 2000

15. Pisa FE, Bovenzi M, Romeo L, et al: Reproductive factors and the risk of scleroderma: an Italian case-control study. Arthritis Rheum 46:451, 2002

16. LeRoy EC, Medsger TA Jr: Criteria for the classification of early systemic sclerosis. J Rheumatol 28:1573, 2001

17. Olsen NJ, King LEJ, Park JH: Muscle abnormalities in scleroderma. Rheum Dis Clin North Am 22:783, 1996

18. Sjogren RW: Gastrointestinal features of scleroderma. Curr Opin Rheumatol 8:569, 1996

19. Bolster MB, Silver RM: Assessment and management of scleroderma lung disease. Curr Opin Rheumatol 11:508, 1999

20. Magliano M, Isenberg DA, Hillson J: Pulmonary hypertension in autoimmune rheumatic diseases: where are we now? Arthritis Rheum 46:1997, 2002

21. Steen V, Medsger TA Jr: Predictors of isolated pulmonary hypertension in patients with systemic sclerosis and limited cutaneous involvement. Arthritis Rheum 48:516, 2003

22. Marie I, Dominique S, Levesque H, et al: Esophageal involvement and pulmonary manifestations in systemic sclerosis. Arthritis Rheum 45:346, 2001

23. Deswal A, Follansbee WP: Cardiac involvement in scleroderma. Rheum Dis Clin North Am 22:841, 1996

24. Coghlan JG, Mukerjee D: The heart and pulmonary vasculature in scleroderma: clinical features and pathobiology. Curr Opin Rheumatol 13:495, 2001

25. Steen VD: Scleroderma renal crisis. Rheum Dis Clin North Am 22:861, 1996

26. Okano Y: Antinuclear antibody in systemic sclerosis (scleroderma). Rheum Dis Clin North Am 22:709, 1996

27. Harvey GR, McHugh NJ: Serologic abnormalities in systemic sclerosis. Curr Opin Rheumatol 11:495, 1999

28. Spencer-Green G, Alter D, Welch HG: Test performance in systemic sclerosis: anti-

centromere and anti-Scl-70 antibodies. Am J Med 103:242, 1997

29. Sule SD, Wigley FM: Update on management of scleroderma. Bull Rheum Dis 49:1, 2000

30. Walker JG, Ahern MJ, Smith MD, et al: Scleroderma renal crisis: poor outcome despite aggressive antihypertensive treatment. Intern Med J 33:216, 2003

31. Steen VD, Medsger TA Jr: Long-term outcomes of scleroderma renal crisis. Ann Intern Med 133:600, 2000

32. DeMarco PJ, Weisman MH, Seibold JR, et al: Predictors and outcomes of scleroderma renal crisis: the high-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis trial. Arthritis Rheum 46:2983, 2002

33. Latsi PI, Wells AU: Evaluation and management of alveolitis and interstitial lung disease in scleroderma. Curr Opin Rheumatol 15:748, 2003

34. Rich S, Kaufmann E, Levy PS: The effect of high doses of calcium-channel blockers on survival in primary pulmonary hypertension. N Engl J Med 327:76, 1992

35. Badesch DB, Tapson VF, McGoon MD, et al: Continuous intravenous epoprostenol for pulmonary hypertension due to the scleroderma spectrum of disease: a randomized, controlled trial. Ann Intern Med 132:425, 2000

36. Barst RJ, Rubin LJ, Long WA, et al: A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension. The Primary Pulmonary Hypertension Study Group. N Engl J Med 334:296, 1996

37. Klings ES, Hill NS, Ieong MH, et al: Systemic sclerosis–associated pulmonary hypertension: short- and long-term effects of epoprostenol (prostacyclin). Arthritis Rheum 42:2638, 1999

38. Simonneau G, Barst RJ, Galie N, et al: Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. Am J Respir Crit Care Med 165:800, 2002

39. Ghofrani HA, Wiedemann R, Rose F: Sildenafil for treatment of lung fibrosis and pulmonary hypertension: a randomised controlled trial. Lancet 360:895, 2002

40. Ghofrani HA, Wiedemann R, Rose F, et al: Combination therapy with oral sildenafil and inhaled iloprost for severe pulmonary hypertension. Ann Intern Med 136:515, 2002

41. Rubin LJ, Badesch DB, Barst RJ, et al: Bosentan therapy for pulmonary arterial hypertension. N Engl J Med 346:896, 2002

42. Harrison BJ, Silman AJ, Hider SL, et al: Cigarette smoking as a significant risk factor for digital vascular disease in patients with systemic sclerosis. Arthritis Rheum 46:3312, 2002

43. Thompson AE, Shea B, Welch V, et al: Calcium-channel blockers for Raynaud's phenomenon in systemic sclerosis. Arthritis Rheum 44:1841, 2001

44. Clements PJ, Furst DE, Wong WK, et al: High-dose versus low-dose D-penicillamine in early diffuse systemic sclerosis: analysis of a two-year, double-blind, randomized, controlled clinical trial. Arthritis Rheum 42:1194, 1999

45. van den Hoogen FH, Boerbooms AM, Swaak AJ, et al: Comparison of methotrexate with placebo in the treatment of systemic sclerosis: a 24 week randomized double-blind trial, followed by a 24 week observational trial. Br J Rheumatol 35:364, 1996

46. Pope JE, Bellamy N, Seibold JR, et al: A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum 44:1351, 2001

47. Binks M, Passweg JR, Furst D, et al: Phase I/II trial of autologous stem cell transplantation in systemic sclerosis: procedure related mortality and impact on skin disease. Ann Rheum Dis 60:577, 2001

48. Furst DE: Stem cell transplantation for autoimmune disease: progress and problems. Curr Opin Rheumatol 14:220, 2002

49. Bulpitt KJ, Clements PJ, Lachenbruch PA, et al: Early undifferentiated connective tissue disease: III. Outcome and prognostic indicators in early scleroderma (systemic sclerosis). Ann Intern Med 118:602, 1993

50. Foeldvari I: Diffuse and limited cutaneous systemic scleroderma. Curr Opin Rheumatol 12:435, 2000

51. Varga J, Kahari VM: Eosinophilia-myalgia syndrome, eosinophilic fasciitis, and related fibrosing disorders. Curr Opin Rheumatol 9:562, 1997

VI IDIOPATHIC INFLAMMATORY MYOPATHIES

Nancy J. Olsen, m.d.
Beth L. Brogan, m.d.

Idiopathic inflammatory myopathies, which include polymyositis and dermatomyositis, primarily affect skeletal muscle. The common features of these diseases are weakness of and inflammatory changes in skeletal muscle. In general, the idiopathic inflammatory myopathies are serious disorders that respond variably to therapy. Polymyositis and dermatomyositis may be linked with other rheumatic diseases, notably scleroderma, and with malignancies. Prognosis varies according to the specific syndrome that is expressed.

Classification

Classification of these heterogeneous muscle disorders into subtypes is useful for determining diagnostic and therapeutic approaches.[1,2] Categories are defined on the basis of clinical and histologic features rather than on laboratory or radiologic tests [*see Table 1*].

dermatomyositis

Patients with dermatomyositis usually show symmetrical proximal muscle weakness in all extremities, accompanied by a characteristic skin rash. Neck and back muscles may also be weak. Areas of skin most commonly affected by the rash include extensor surfaces of the hands and knees. Subtypes of dermatomyositis include the juvenile form [*see* Juvenile Myositis, *below*]. Another recognized subtype is amyopathic dermatomyositis.[3-5] Patients with this disorder have the characteristic rash but do not have demonstrable muscle abnormalities. One study of such patients has shown that sensitive magnetic resonance imaging techniques can reveal changes after exercise that indicate a metabolic abnormality.[6]

polymyositis

Patients with polymyositis have symmetrical proximal muscle weakness similar to that experienced in dermatomyositis, but the rash is absent. Onset of polymyositis may be more difficult to determine, in part because no rash is available as an indicator of possible inflammation. Muscle weakness and atrophy may be more profound than that usually seen in patients with dermatomyositis. However, no formal studies of long-term outcome have been carried out that prove this assertion.

juvenile myositis

Children ranging in age from younger than 5 years through the teen years may be affected by juvenile myositis. Most children have a skin rash, and vasculitis and soft tissue calcifications are much more common in children than in adults.[7] Although residual dermatologic changes, muscle fibrosis, and calcification may occur, the long-term outlook is generally favorable.[8]

myositis with malignancy

Most, but not all, cases of malignancy-associated myositis are accompanied by the typical rash of dermatomyositis. A recent study showed that patients with dermatomyositis had a relative risk of malignancy of 6.2. For patients with polymyositis or inclusion body myositis, the risk was lower but still significant.[9-11] The incidence of underlying malignancy increases with age[12] and decreases with increasing time from diagnosis.[9] Onset of the myositis may precede or follow discovery of the malignancy. Adults with dermatomyositis should be screened for occult malignancies in the first 2 years after onset of disease.

myositis with other rheumatic diseases

Inflammatory myositis may occur with another established rheumatic disease, most commonly scleroderma. Other conditions that can occur with myositis include rheumatoid arthritis, systemic lupus erythematosus (SLE), and Sjögren syndrome. Recent reports have linked scleromyxedema to dermatomyositis[13] and have linked inclusion body myositis to subacute cutaneous lupus.[14,15] Many patients with these overlap conditions have a relatively mild form of muscle inflammation that responds well to treatment. However, a small subset of patients, especially those with coexistent scleroderma, may have a severe and very debilitating muscle weakness that is resistant to therapy.[16]

inclusion body myositis

The pattern and severity of muscle weakness in inclusion body myositis (IBM) differs from the pattern of severity seen in the other idiopathic inflammatory myopathies. In addition to the presence of proximal weakness, distal muscles may be involved; and in some cases, muscle abnormalities are asymmetrical. Unlike most of the other muscle disorders discussed in this subsection, IBM afflicts more men than women, with approximately two thirds of affected persons being men. Response to treatment is generally poor.

Epidemiology

prevalence and incidence

The estimated prevalence of idiopathic inflammatory myopathies is approximately one case per 100,000 individuals.

Table 1 Classification of Myositis Syndromes

Clinicopathologic Category	*Characteristics*
Dermatomyositis	Proximal weakness, skin rash; amyopathic variant with rash only
Polymyositis	Proximal weakness without rash
Juvenile myositis	Myositis in childhood, usually with a rash
Myositis with malignancy	Myositis with associated underlying neoplastic disease
Myositis with another connective tissue disease	Coexistent syndrome, usually scleroderma, rheumatoid arthritis, or systemic lupus erythematosus
Inclusion body myositis	Severe weakness with characteristic inclusions on muscle biopsy

This prevalence makes these disorders about 1,000 times less common than rheumatoid arthritis. The rarer syndromes, such as IBM, may constitute 20% or less of all cases. In one study, the annual incidence of idiopathic inflammatory myopathies was 5.5 cases per million population.[17] Incidence rates, however, may be increasing, possibly because of improved methods of detection.

ETHNIC, RACIAL, AND GENDER GROUP DIFFERENCES

No ethnic clustering of the idiopathic inflammatory myopathies has been reported. It has been suggested that incidence rates in North America are increasing faster in African Americans than in whites.[17] In adults, polymyositis is more common than dermatomyositis, whereas in children and young adults, dermatomyositis is the predominant form. It has been suggested that incidence rates are higher in regions that have greater amounts of sun exposure.[18] Polymyositis and dermatomyositis show a female-to-male ratio of approximately 2:1. Risk of underlying malignancy increases significantly after 40 years of age.[12] Malignancies in children are rare but have been reported. The diagnosis of inclusion body myositis is rarely made in persons younger than 50 years.

Etiology and Pathogenesis

The etiology of inflammatory muscle disease remains unknown. The most widely accepted hypotheses suggest multiple factors. One possible scenario is that an initial insult—for example, a virus or another infectious agent or an environmental toxin—leads to muscle damage in a genetically susceptible host. This process in turn triggers an immune response, subsequently causing chronic muscle inflammation.[19]

INFECTIOUS AGENTS

A role for viruses in the etiology of idiopathic inflammatory myopathies has been suggested by seasonal and geographic clustering of new cases. Furthermore, infection with HIV or hepatitis C virus has been associated with the development of myopathy.[20] Most studies looking for evidence of viral genomic material in muscle tissue have failed to find such evidence.[21] Immunoreactivity for hepatitis C in involved muscle tissues has been reported in a single case.[20] Viruses may mediate tissue damage, which may in turn lead to immunologic responses that target or damage muscle tissues.[22] The relative rarity of the myositis syndromes would suggest that if a common infectious agent were involved, coexistent factors would also be required. These factors could include host-specific genetic loci that control the immune response or other noninfectious factors such as drugs or environmental toxins.[18]

NONINFECTIOUS FACTORS

Lipid-lowering agents such as clofibrate and the statin group of drugs have been associated with elevated levels of serum muscle enzymes and with muscle weakness in a small number of patients. However, most patients are asymptomatic. The list of drugs reported to be associated with development of myopathy is very long. For this reason, concomitant medications should be examined closely in any patient with unexplained muscle weakness.[23] Both HIV infection and drugs used in its treatment, such as zidovudine (AZT), have been implicated in the development of myopathy. It is possible that as yet undefined environmental toxins play a role.

GENETIC FACTORS

Familial clustering of inflammatory myositis syndromes occurs, but the great majority of cases are sporadic. Sporadic cases have been linked to HLA-DRB1*0301, whereas familial cases have shown increased prevalence of HLA-DQA1 (DQA1*0501). A form of hereditary IBM has been described in several ethnic groups. Chromosomal links with this disorder have been identified, but candidate genes are as yet undefined.[24] Many of the reported studies of genetic links in inflammatory myopathies have grouped several types of syndromes together. It is probable that future studies that perform separate analyses of the various distinct clinical syndromes will show stronger associations with genetic markers.

AUTOIMMUNE FACTORS

The presence of cellular infiltrates in muscle tissues is a defining feature of inflammatory muscle diseases [*see Figure 1*]. Light microscopic examination of these infiltrates reveals different patterns of infiltration. In tissues from patients with dermatomyositis, the lymphocytes are generally located around blood vessels and at the periphery of the muscle bundles. Invasion of muscle fibers by mononuclear cells is rarely observed, and there is a relative paucity of necrotic muscle fibers. Complement-mediated capillary damage is also more commonly observed in biopsy samples from dermatomyositis patients, especially those patients with an underlying malignancy. Some studies suggest that dermatomyositis patients with capillary damage who do not have a malignancy have a more acute syndrome that responds better to immunosuppressive treatment. In polymyositis, muscle fibers may be invaded by the mononuclear infiltrates, and focal areas of muscle destruction are seen. Tissues from patients with IBM usually show some degree of inflammation accompanied by intracellular rimmed vacuoles.[25]

Two groups have identified chimeric cells of maternal origin in the peripheral blood and inflammatory lesions of children with myositis. These findings support the hypothesis that

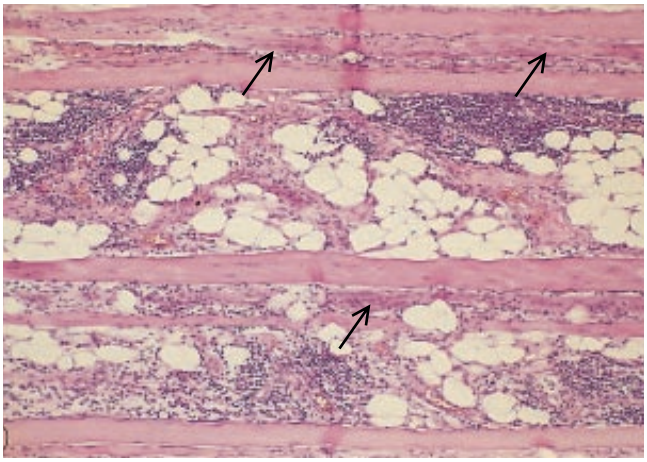


Figure 1 **Extensive pathologic changes can be seen in involved muscle in myositis. These changes include a decrease in the number of striated muscle fibers and a loss of cross-striations in the remaining fibers. Some fibers demonstrate increased numbers of rounded nuclei and basophilic staining (arrows), which suggests attempted regeneration. There is also intense infiltration of the muscle by mononuclear inflammatory cells, predominantly lymphocytes and plasma cells.**

childhood myositis is a manifestation of a graft-versus-host reaction.[26,27]

Differences between polymyositis and dermatomyositis are revealed by immunophenotyping of the cellular infiltrates.[28] Mononuclear cell infiltrates in polymyositis and probably in IBM tissues are predominantly of the $CD8^+$ cytotoxic T cell phenotype. The $CD8^+$ T cells in polymyositis show evidence of clonal expansion, which is most likely driven by muscle-specific antigens.[29] Activated $CD8^+$ T cells probably mediate cytotoxic, immune-mediated, and antigen-specific muscle cell destruction. In dermatomyositis, T cells, predominantly of the $CD4^+$ helper-inducer phenotype, are present along with B cells; restricted clonality is not seen.[29] These differences in histology support the hypothesis that polymyositis and dermatomyositis are distinct disorders with different etiologies.

Diagnosis

Major diagnostic criteria that were proposed by Bohan and Peter in 1975[1,2] remain useful for defining most of the myositis syndromes. However, IBM, which was not recognized at the time these criteria were written, differs somewhat from polymyositis and dermatomyositis. Although sophisticated diagnostic tests, including autoantibody profiles and imaging techniques, are now available, findings obtained through a careful history and physical examination remain indispensable for both making the initial diagnosis and evaluating responses to treatment.

CLINICAL FEATURES

Muscle Weakness

In polymyositis and dermatomyositis, the weakness is predominantly proximal. Distal strength is usually preserved. In IBM, the weakness may be asymmetrical, and diminished distal strength is commonly seen. Muscle strength can be tested in the office or at the bedside and estimated on a semiquantitative scale from 1 to 5. Devices for quantitative assessment of muscle strength, some of which can be used at the bedside, are also available.

Characteristic Skin Rash

The rash of dermatomyositis is a deep-red erythematous eruption, with or without mild scaling and atrophy. It occurs on the face, neck, upper chest, and extensor surfaces of joints such as elbows and those of the hands. Periorbital edema may appear, as may heliotrope erythema, which is characterized by a violet or lilac color, especially of the eyelids. Occasionally, the rash is more widespread or takes different forms [*see Figure 2*]. Erythema and telangiectasia also occur in periungual areas. In adults, vasculitis is usually confined to the skin and takes the form of urticaria, subcutaneous nodules, periungual infarcts, or digital ulcerations. Cutaneous vasculitis has been associated with underlying malignant disease.

Pulmonary Involvement

Pulmonary involvement occurs in nearly 50% of patients who have myositis, with pneumonia being the most common pulmonary abnormality. Aspiration pneumonia, which is often recurrent, is prevalent in patients who have pharyngeal muscle weakness. Ineffective coughing caused by ventilatory muscle weakness also occurs but is far less common than swallowing

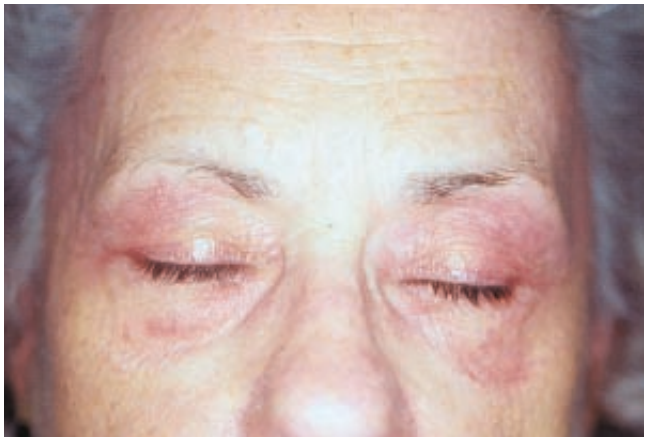


Figure 2 **Skin eruption in a patient with dermatomyositis consists of a deep-red, erythematous, papular rash over the nasal and forehead areas and a lilac-colored, or heliotrope, erythema of the upper eyelid and orbital area.**

problems. In general, the patient with recurrent aspiration pneumonia has a poor prognosis; it indicates marked dysfunction of many muscle groups. Bacterial pneumonia caused by aspiration is a major cause of death in elderly patients.[12] Opportunistic infections may occur in patients undergoing immunosuppressive drug therapy. In addition, some of these drugs, most notably methotrexate, can be associated with the development of pneumonitis, which is usually reversible but is potentially fatal.

Interstitial lung disease (ILD) occurs in up to 30% of myositis patients and in approximately 60% of patients who have antibodies directed against aminoacyl-transfer RNA (tRNA) synthetases. The advent and application of newer, sensitive diagnostic techniques such as high-resolution computed tomography may lead to an increase in the detection of pulmonary abnormalities. The most common presentation is with progressive shortness of breath, which may be accompanied by a nonproductive cough. On physical examination, basilar crepitant rales are usually detected. Progression may be slow, and symptoms may occur in patients with established disease; or onset may be rapid and may occur at the same time as the muscle weakness.[30] Hypoxemia and respiratory alkalosis may be present. In some patients, these abnormalities are detected only after exercise. High-resolution CT scanning is useful for detection of interstitial fibrosis that might not be appreciated on routine chest radiography. Pulmonary function tests may reveal reduced lung volume and diminished diffusion capacity. One of three forms of histology of ILD is usually found: interstitial pneumonia, diffuse alveolar damage, or bronchiolitis obliterans with or without organizing pneumonia. ILD occurs with or without skin involvement. There is no correlation between the development of ILD and the severity of muscle involvement, and ILD may precede or follow the onset of muscle weakness. ILD is associated with a high mortality. Treatment with cyclophosphamide or azathioprine has been reported to be beneficial in some patients.[30] A small number of patients with acute pneumonitis may respond to corticosteroid treatment alone.

Cardiac Abnormalities

Clinically significant involvement of heart muscle is unusual and is probably associated with a poor prognosis. Cardiac abnormalities may take many forms, ranging from rhythm or

conduction disturbances to myocardial inflammation or fibrosis. Cardiac muscle abnormalities may be detected by radionuclide scanning studies. However, many histologic and electrical abnormalities are not clinically significant.[31] Therefore, evaluation beyond routine diagnostic studies is rarely indicated.

Calcinosis

Soft tissue calcification is seen most commonly in children. Deposits may be deep along fascial planes or in superficial dermal areas, sometimes with ulceration through the skin. Treatments have been based on largely anecdotal reports; no systematic studies have been carried out.[32] Agents that have been found to be of use in some cases include probenecid, diltiazem, and warfarin. Some patients show spontaneous regression of calcinosis without specific treatment.

Vascular Abnormalities

Raynaud phenomenon is most commonly observed in patients whose myositis is associated with another rheumatic disease (e.g., scleroderma). Clinically significant vasculitis is unusual in adults, although dermatomyositis patients may show vascular changes on histologic examination.

LABORATORY TESTS

Muscle Biopsy

Histologic confirmation of muscle inflammation is required in many, but not all, cases of inflammatory muscle disease. Patients with the characteristic skin rash of dermatomyositis and with elevated serum muscle enzyme levels may be treated without a muscle biopsy, because these two indicators can be used to follow the course of disease. In the absence of a skin rash or elevations in muscle enzyme levels, diagnosis is more difficult; in most patients, a biopsy is needed to confirm the presence of muscle inflammation. Two types of biopsy approaches are used: open surgical and closed needle. The closed-needle approach offers the advantages of decreased morbidity and of lower cost because an operating room is not required. Tissue samples obtained with the closed-needle approach can provide sufficient diagnostic information for interpretation by the muscle pathologist. However, the quality of the specimen obtained is dependent on the skill and experience of the operator. In the absence of such a resource or in special cases in which it is desirable to obtain extra tissue, the open surgical approach is preferable. Imaging studies such as MRI or CT can be used to determine the optimal site for biopsy.

All biopsy specimens require immediate handling by an experienced surgical team working closely with the pathology laboratory to ensure optimal results. Light microscopic analysis is sufficiently informative for most purposes. Because the treatment of polymyositis and that of dermatomyositis are the same, immunophenotyping of cellular infiltrates to distinguish between these two disorders is not indicated for routine diagnostic specimens. Electron microscopy may be required to demonstrate the inclusion bodies that define IBM. Examination of the biopsy specimen by a specialist in neuromuscular pathology may be helpful, because many pathologists do not see these diseases on a regular basis.

Muscle Enzymes

Most patients with inflammatory myopathy have increased muscle enzyme levels at some point during the course of active myositis.[33] The presence of intracellular muscle enzymes in the serum most likely reflects damage to muscle cell membranes. The most commonly used muscle enzyme measurement is the creatine kinase (CK) level. The CK level may rise to many times normal. The MB isozyme of CK may be elevated because of the presence of this isoform in regenerating skeletal muscle. Measurement of CK may be confounded by the presence of naturally occurring inhibitors of this enzyme. Furthermore, racial and gender variations exist for normal levels of CK, with black males generally showing the highest values.[34] Aldolase is another muscle enzyme that may be measured in the serum and may have less variability. However, aldolase is present in tissues other than muscle, and therefore, it is not specific for muscle damage. MRI studies have shown that active muscle inflammation may exist in patients with persistently normal CK serum levels.[35] Reasons for this discordance are not known, but the findings suggest that treatment strategies should be focused on the clinical status of the patient rather than on the muscle enzyme levels.[36]

Autoantibodies

Autoantibodies to nuclear and cytoplasmic antigens are found in as many as 90% of patients with an inflammatory myopathy. These antibodies are often useful in differentiating inflammatory myopathies from diseases that are not autoimmune disorders. Some of these autoantibodies are nonspecific and are seen in several autoimmune disorders. Other autoantibodies are relatively specific for the inflammatory myositis syndromes in general or for specific diagnostic categories. About 25% of patients with inflammatory myositis test positive for antinuclear antibody; in patients with overlapping rheumatic disease syndromes, the percentage is higher. The antinuclear antibody test is generally not helpful in establishing a diagnosis of myositis or one of its subsets. Autoantibodies that are in large part directed against cytoplasmic ribonucleoproteins have been designated as myositis-specific autoantibodies (MSA). Approximately 30% of patients with myositis have one or more of these autoantibodies. They are thus relatively specific but not sensitive to the presence of myositis, and as such, these autoantibodies cannot be used to screen for the presence of disease.

Three groups of patients can be defined by the MSA specificities. These subgroups differ in clinical presentation and prognosis.[37] The first group is defined by the presence of antibodies directed against aminoacyl-tRNA synthetases. The presentation in the first group is generally characterized by an acute onset of muscle disease, with a high incidence of associated interstitial lung disease. Patients in this group may also have arthritis and a hyperkeratotic rash on the hands, known as mechanic's hands. A majority of patients in this group test positive for HLA-DR3. Responses to treatment are variable, and mortality is significant. The second group includes patients with antibodies to the signal recognition particle (SRP). This protein complex facilitates translocation of newly synthesized polypeptides across the endoplasmic reticulum. Patients with anti-SRP have an abrupt onset of muscle weakness and may have associated involvement of cardiac muscle. The majority of patients in this group are African-American women. Responses to treatment are not good, and the prognosis is poor. In one series, the 5-year mortality for anti-SRP patients was 75%. A third group is identified by antibodies to Mi-2, which is a nuclear protein with unknown function. The majority of these patients have the dermatomyositis clinical syndrome with the so-

called shawl-sign pattern of rash on the trunk and with cuticular overgrowth. Responses to treatment are generally good, and mortality is lower than that in the other groups. Most of these clinical associations, which were originally described in North American patients, have been confirmed in a large group of European patients.[38] Preliminary reports suggest that antibodies to a novel 155 kd protein may also be useful in identifying patients with amyopathic dermatomyositis.[39]

Electromyography

In most patients, electromyography reveals low-amplitude, polyphasic motor unit potentials, indicating a lack of synchronous contracture in muscle fibers within motor units. This finding correlates with the usually inhomogeneous distribution of muscle degeneration shown by histopathologic examination. Fibrillations and insertional irritability are evidence of membrane abnormalities. These findings are characteristic of, but not specific for, myositis.

IMAGING AND SPECTROSCOPY TECHNIQUES

Conventional radiographs have little value in evaluating skeletal muscle. However, other techniques, including ultrasonography, CT, and MRI, can enhance diagnostic approaches to many myopathies.[40] Of these modalities, MRI has been the most useful in the evaluation and longitudinal management of inflammatory muscle syndromes. However, it may not be the method of choice in all circumstances, and in some of these cases, the alternative modalities of ultrasonography and CT can provide helpful information. Advantages of these three techniques are that they are noninvasive and offer the possibility of examining a volume of muscle larger than that which can be obtained by biopsy. In patients for whom biopsy may be a difficult or traumatic experience, such as young children, imaging may provide sufficient information to proceed with treatment.

Magnetic Resonance Imaging

MRI is a very accurate method for muscle imaging that has been very useful in the diagnosis and management of patients with inflammatory muscle diseases of many kinds. Full assessment requires both T_1- and T_2-weighted images. The T_1 image is most useful for outlining muscle anatomy because it detects changes in muscle mass caused by atrophy or fat infiltration. Inflammation is readily detected on the T_2-weighted image, where the abnormal areas appear as brightness against the usually dark background of normal muscle [*see Figure 3*]. Studies using MRI have clearly demonstrated the patchy nature of the muscle inflammation, perhaps explaining why some patients with significant weakness have normal biopsy results. In dermatomyositis, inflammation in the thigh muscles is seen in predominantly anterior muscle compartments, and muscle mass is generally preserved. In patients with polymyositis and inclusion body myositis, extensive fat infiltration and muscle atrophy, which can include all muscle groups, are more likely to be seen. Longitudinal MRI studies can be used to document the effectiveness of immunosuppressive therapy. Patients may be studied in the usual body coil, which allows for visualization of both legs.[41] Other studies have utilized a knee coil positioned over the anterior quadriceps, which provides a greater level of detail.[35] As in all MRI studies, patients must be very carefully questioned for the presence of any indwelling metals before being placed in the magnet.

Ultrasonography

Ultrasonography is a readily available and relatively inexpensive technique that has been used to examine a wide variety of muscle disorders.[42,43] Inflammation within muscle tissues appears on ultrasonography as areas of decreased echogenicity. In addition, blood-flow changes can be measured with related techniques such as color Doppler imaging. Ultrasonography may be useful in guiding the choice of site for needle or open muscle biopsy.

Computed Tomography

CT is not useful for the detection of inflammatory muscle changes. However, areas of atrophy or fat infiltration cause decreased muscle density, which is easily detected by CT. Soft tissue calcifications such as those seen in juvenile dermatomyositis are best visualized with CT. Sometimes, these calcifications are in deep areas that cannot be readily appreciated on physical examination.

a *b*

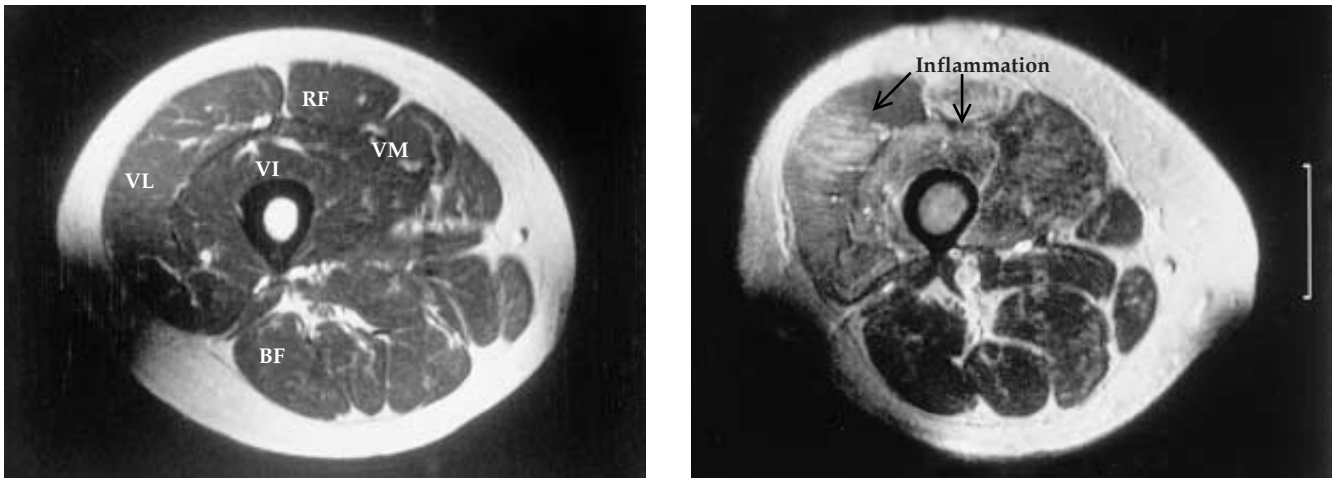


Figure 3 **Magnetic resonance images of thigh muscles in a patient with dermatomyositis. The T_1-weighted image (*a*) shows uniform density in all muscle groups, identified as VL, vastus lateralis; VI, vastus intermedius; VM, vastus medialis; RF, rectus femoris; and BF, biceps femoris. The T_2-weighted image (*b*) illustrates inflammation in muscles of the quadriceps group, shown as areas of brightness or increased signal intensity.[35]**

Magnetic Resonance Spectroscopy

Spectroscopy is primarily a research tool. However, studies have shown the utility of this noninvasive approach for evaluating muscle function, and applications in the clinic may become available in the near future.[35,44] In patients with dermatomyositis, loss of high-energy phosphate compounds needed for efficient muscle contraction has been documented with P-31 magnetic resonance spectroscopy.[44] Longitudinal studies have documented that correction of these metabolic abnormalities may lag behind improvement in muscle inflammation.

EVALUATION FOR UNDERLYING MALIGNANCY

Patients with dermatomyositis and polymyositis are at increased risk for an underlying malignancy. The magnitude of this risk is difficult to determine and varies greatly between reports. One study in a population-based cohort estimated the relative risk of cancer as 1.8 in males and 1.7 in females.[45] A more recent study from Scotland has indicated that the relative risk may be as high as 7.7 in patients with dermatomyositis, with a greater risk in females than in males.[46] In general, the risk is greater in patients with dermatomyositis than in patients with polymyositis and in all patients who are older than 40 years. There is general agreement that routine screening for malignancies should include chest radiography, mammography (in women), examination of stool for occult blood, complete gynecologic examination, and assessment for prostate-specific antigen (in men). Abnormalities seen on these screening tests may suggest the need for additional studies such as endoscopy, colonoscopy, and tissue biopsy. The most difficult malignancies to detect are those arising in the ovary. Uterine transvaginal ultrasonography or CT of the pelvis should be done in women older than 40 years, but some occult ovarian malignancies escape detection even with these tests. Some investigators advocate lower gastrointestinal studies to detect colon cancer in patients older than 65 years.[12] Other, more extensive screening tests for occult malignancies are generally not recommended.

Differential Diagnosis

Diagnosis of dermatomyositis is aided by the presence of the characteristic rash. However, because the rash has features of SLE, this diagnosis may be confused with SLE, especially when the antinuclear antibody test is positive. Patients with polymyositis may be difficult to distinguish from patients with other myopathic disorders [*see Table 2*]. These other disorders include metabolic myopathies, endocrine dysfunction, drug-induced disorders, infections, and miscellaneous syndromes such as sarcoidosis. Some types of dystrophies should also be considered in patients who have muscle weakness and elevated muscle enzyme levels. Myalgia syndromes, such as polymyalgia rheumatica, in which stiffness is a predominant complaint, may confuse the diagnosis in some patients. Fibromyalgia, which is associated with a primary symptom of fatigue rather than muscle weakness, is characterized by the presence of discrete tender points that are not usually present in myositis patients.

Table 2 Differential Diagnosis of Inflammatory Myositis

Cause	*Effect*
Metabolic myopathies	Myophosphorylase deficiency (McArdle disease) Myoadenylate deaminase deficiency Carnitine palmitoyltransferase deficiency Glycogen storage disease Periodic paralysis Hypokalemia, hypomagnesemia
Endocrine disorders	Cushing syndrome Thyroid dysfunction
Drug-induced disorders	Ethanol toxicity Penicillamine toxicity Lipid-lowering drug (statin) toxicity Zidovudine toxicity
Infections	Viral: HIV, coxsackievirus, adenovirus, influenza virus, echovirus
Other rheumatic disorders	Systemic lupus erythematosus (rash) Polymyalgia rheumatica
Miscellaneous disorders	Sarcoidosis, eosinophilia

Treatment

Guidelines for treatment of the idiopathic inflammatory myopathies are not well established for several reasons. The diseases are uncommon, making it difficult to accumulate sufficient numbers of patients to carry out randomized, controlled trials. In addition, some forms of these diseases have a slow, prolonged course, requiring long periods of observation. Finally, there is as yet no uniformly accepted classification scheme for these disorders; thus, comparisons of therapies administered to different groups of patients at different times and places may not be valid. As examples, in the past, polymyositis and dermatomyositis have been included in the same category, and IBM may not have been recognized. It is now clear that different forms of these diseases vary in prognosis and in response to therapy.

DRUG THERAPY

A table showing drugs for the treatment of inflammatory muscle diseases is provided [*see Table 3*].

Glucocorticoids

Corticosteroids are the mainstay of initial therapy. Most patients with documented muscle inflammation should be started on these drugs at relatively high levels (1 mg/kg/day), given in divided doses. A standard approach has been to maintain this dosage for up to 3 months or until clinical improvement occurs. After this initial period of high-dose therapy, the dose can be consolidated into a single morning dose and then tapered, with the total daily dose being reduced by 20% to 25% each month and a maintenance dose of 5 to 10 mg daily being achieved in about 6 to 8 months. The addition of second-line drugs to the prednisone regimen is now recommended within 3 months after initiation of treatment. Older patients with comorbid conditions such as diabetes and osteoporosis are especially at risk from side effects of steroids. Side effects may include a cushingoid appearance, compression fractures, avascular necrosis, cataracts, and infections. One study has suggested that the side effects of corticosteroid therapy contribute significantly to the morbidity of polymyositis and dermatomyositis.[47] For these reasons, any patient with severe muscle weakness,

Table 3 Drugs for the Treatment of Inflammatory Muscle Diseases

Drug	*Dose*	*Efficacy Rating*	*Comments*
Prednisone	5–60 mg/day	Highly effective for initial treatment	Side effect: cushingoid syndrome Avoid prolonged use at high doses; taper to 10 mg/day or less
Methotrexate	15–25 mg/wk	Effective, steroid-sparing	Side effects: liver abnormalities, pneumonitis Supplement with folate
Azathioprine	100–150 mg/day	Effective, steroid-sparing	Side effect: bone marrow suppression Can be combined with methotrexate
Hydroxychloroquine	200 mg b.i.d.	Effective for skin manifestations	Side effect: retinal toxicity Can be combined with other agents
Cyclophosphamide	100–150 mg daily or as intravenous pulses every 6 wk	Possibly effective for lung involvement	Side effects: bone marrow suppression, hemorrhagic cystitis
Cyclosporine	3 mg/kg/day	Use after other immunosuppressants	Side effects: hypertension, renal dysfunction
Intravenous immunoglobulin	1 g/kg/day, 2 consecutive days monthly	Use in patients in whom other regimens have failed	High cost and limited supply

limited functional status, or underlying conditions that make steroids a high risk (e.g., diabetes mellitus or osteoporosis) should be started on second-line immunosuppressive drugs at the outset.

Methotrexate and Azathioprine

The most commonly used second-line agents for the treatment of inflammatory myopathy are methotrexate and azathioprine. Methotrexate may be given orally or subcutaneously at an initial dosage of 7.5 to 10 mg weekly and then increased gradually to 25 mg weekly. As the dosage of methotrexate is increased, the dosage of prednisone is usually tapered. In general, methotrexate is well tolerated by patients with inflammatory myopathy, but there have been reports of toxicities similar to those seen in patients with rheumatoid arthritis who have taken methotrexate. Regular monitoring of liver function is necessary. Measurement of enzymes other than aminotransferases is required to prevent interference by the ongoing muscle inflammation. γ-Glutamyltranspeptidase is a liver-specific alternative. Methotrexate may be useful in the treatment of interstitial lung disease associated with myositis, but because this drug may in rare cases cause pulmonary toxicity, it is relatively contraindicated in patients with significant lung problems.

Azathioprine has been shown to be effective in patients with myositis in a prospective, controlled, double-blind trial, but treatment for at least 6 months may be required for improvement to occur. Azathioprine therapy should be initiated at a dosage of 50 to 100 mg/day, and the dosage should be increased gradually to a maximum of 150 to 200 mg/day. Side effects include bone marrow suppression and development of infections and, possibly, malignancies. Azathioprine and methotrexate have similar efficacy in these disorders, and the choice of which to use may depend on tolerability or comorbid conditions. Patients with myositis in whom therapy with glucocorticoids and either methotrexate or azathioprine has failed may respond to a combination of methotrexate and azathioprine.[48,49]

Other Immunosuppressive Agents

Cyclophosphamide has been given both as intravenous pulse therapy and by daily oral administration. Some data suggest that it may be useful in adults with the antisynthetase syndrome and in children with vasculitis-related complications of dermatomyositis. Cyclophosphamide may be useful in the treatment of the complication of interstitial lung disease.[30] Other drugs that may be of value in patients in whom other therapies have failed include cyclosporine, FK506 (tacrolimus), chlorambucil, and mycophenolate mofetil.[50-52] The adenine analogue fludarabine has also shown some benefit in one study of refractory patients.[53]

Intravenous Immune Globulin

Intravenous immune globulin appears to benefit some patients with either polymyositis or dermatomyositis. A controlled trial of immune globulin in dermatomyositis patients demonstrated efficacy when given at a dosage of 1 g/kg/day for 2 days, repeated monthly for 3 months.[54] The combination of intravenous immune globulin and cyclosporine may be of value.[55] One controlled study suggested that intravenous immune globulin may be of benefit in IBM,[56] but studies by another group failed to show clinical improvements.[57] Treatment with intravenous immune globulin is limited by the restricted supply and very high cost and should be reserved for severe cases not responding to other therapies.

SKIN PROTECTION

The rash of dermatomyositis is usually photosensitive. Therefore, attention to protection from the sun is very important, and patients should be advised to avoid sun exposure as much as possible. Sunscreen preparations, sun-protective clothing, and tinting of windows are often effective. Some dermatologists recommend use of β-carotene, 25 to 30 mg, taken twice daily initially and then increasing to no more than five times a day. Antimalarials may be of benefit, and one report suggests the use of topical tacrolimus.[58]

PHYSICAL THERAPY

Physical therapy plays an important role in the rehabilitation of patients with myositis. During the phase of active inflammatory disease, passive range-of-motion exercises are necessary to prevent contractures. Once the inflammatory compo-

nent of the disease is controlled, active resistive exercises are useful in regaining muscle strength.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this subsection.

References

1. Bohan A, Peter JB: Polymyositis and dermatomyositis (pt I). N Engl J Med 292:344, 1975

2. Bohan A, Peter JB: Polymyositis and dermatomyositis (pt II). N Engl J Med 292:403, 1975

3. Euwer RL, Sontheimer RD: Amyopathic dermatomyositis (dermatomyositis sine myositis). J Am Acad Dermatol 24:959, 1991

4. el-Azhara RA, Pakza SY: Amyopathic dermatomyositis: retrospective review of 37 cases. J Am Acad Dermatol 46:560, 2002

5. Caproni M, Cardinali C, Parodi A, et al: Amyopathic dermatomyositis: a review by the Italian Group of Immunodermatology. Arch Dermatol 138:114, 2002

6. Park JH, Olsen NJ, King LE, et al: MRI and P-31 magnetic resonance spectroscopy detect and quantify muscle dysfunction in the amyopathic and myopathic variants of dermatomyositis. Arthritis Rheum 38:68, 1995

7. Pachman LM, Hayford JR, Chung A, et al: Juvenile dermatomyositis at diagnosis: clinical characteristics of 79 children. J Rheumatol 25:1198, 1998

8. Collison CH, Sinal SH, Jorizzo JL, et al: Juvenile dermatomyositis and polymyositis: a follow-up study of long-term sequelae. South Med J 91:17, 1998

9. Buchbinder R, Forbes A, Hall S, et al: Incidence of malignant disease in biopsy-proven inflammatory myopathy: a population-based cohort study. Ann Intern Med 134:1087, 2001

10. Sparsa A, Liozon E, Herrmann F, et al: Routine vs extensive malignancy search for adult dermatomyositis and polymyositis: a study of 40 patients. Arch Dermatol 138:969, 2002

11. Hill CL, Zhang Y, Sigurgeirsson B, et al: Frequency of specific cancer types in dermatomyositis and polymyositis: a population-based study. Lancet 357:96, 2002

12. Marie I, Hatron PY, Levesque H, et al: Influence of age on characteristics of polymyositis and dermatomyositis in adults. Medicine (Baltimore) 78:139, 1999

13. Launay D, Hatron PY, Delaporte E, et al: Scleromyxedema (lichen myxedematosus) associated with dermatomyositis. Br J Dermatol 144:359, 2001

14. Wenzel J, Uerlich M, Gerdsen R, et al: Association of inclusion body myositis with subacute cutaneous lupus erythematosus. Rheumatol Int 21:75, 2001

15. Lindvall B, Bengtsson A, Ernerudh J, et al: Subclinical myositis is common in primary Sjögren's syndrome and is not related to muscle pain. J Rheumatol 29:717, 2002

16. Olsen NJ, King LE Jr, Park JH: Muscle abnormalities in scleroderma. Rheum Dis Clin North Am 22:783, 1996

17. Oddis CV, Conte CG, Steen VD, et al: Incidence of polymyositis-dermatomyositis: a 20-year study of hospital diagnosed cases in Allegheny County, PA 1963-1982. J Rheumatol 17:1329, 1990

18. Reed AM: Myositis in children. Curr Opin Rheumatol 13:428, 2001

19. Englund P, Nennesmo I, Klareskog L, et al: Interleukin-1alpha expression in capillaries and major histocompatibility complex class I expression in type II muscle fibers from polymyositis and dermatomyositis patients: important pathogenic features independent of inflammatory cell clusters in muscle tissue. Arthritis Rheum 46:1044, 2002

20. Kase S, Shiota G, Fujii Y, et al: Inclusion body myositis associated with hepatitits C virus infection. Liver 21:357, 2001

21. Pachman LM, Litt DL, Rowley AH, et al: Lack of detection of enteroviral RNA or bacterial DNA in magnetic resonance imaging–directed muscle biopsies from twenty children with active untreated juvenile dermatomyositis. Arthritis Rheum 38:1513, 1995

22. Ytterberg SR: Infectious agents associated with myopathies. Curr Opin Rheumatol 8:507, 1996

23. Pascuzzi RM: Drugs and toxins associated with myopathies. Curr Opin Rheumatol 10:511, 1998

24. Dalakas MC: Molecular immunology and genetics of inflammatory muscle diseases. Arch Neurol 55:1509, 1998

25. Vogel H: Inclusion body myositis: a review. Adv Anat Pathol 5:164, 1998

26. Reed AM, Picornell YJ, Harwood A, et al: Chimerism in children with juvenile dermatomyositis. 357:887, 2002

27. Artlett CM, Ramos R, Jiminez SA, et al: Chimeric cells of maternal origin in juvenile idiopathic inflammatory myopathies. Childhood Myositis Heterogeneity Collaborative Group. Lancet 356:2155, 2000

28. Engel AG, Arahata K: Mononuclear cells in myopathies: quantitation of functionally distinct subsets, recognition of antigen-specific cell-mediated cytotoxicity in some diseases, and implications for the pathogenesis of the different inflammatory myopathies. Hum Pathol 17:704, 1986

29. Mantegazza R, Andreetta F, Bernasconi P, et al: Analysis of T cell receptor repertoire of muscle-infiltrating T lymphocytes in polymyositis: restricted Vα/β rearrangements may indicate antigen-driven selection. J Clin Invest 91:2880, 1993

30. Schwarz MI: The lung in polymyositis. Clin Chest Med 19:701, 1998

31. Gonzalez-Lopez L, Gamez-Nava JI, Sanchez L, et al: Cardiac manifestations in dermato-polymyositis. Clin Exp Rheumatol 14:373, 1996

32. Spiera R, Kagen L: Extramuscular manifestations in idiopathic inflammatory myopathies. Curr Opin Rheumatol 10:556, 1998

33. Hochberg MC, Feldman D, Stevens MB: Adult onset polymyositis/dermatomyositis: an analysis of clinical and laboratory features and survival in 76 patients with a review of the literature. Semin Arthritis Rheum 15:168, 1986

34. Worrall JG, Phongsathorn V, Hooper RL, et al: Racial variation in serum creatinine kinase unrelated to lean body mass. Br J Rheumatol 29:371, 1990

35. Park JH, Vital T, Ryder N, et al: MR imaging and P-31 MR spectroscopy provide unique quantitative data for longitudinal management of patients with dermatomyositis. Arthritis Rheum 37:736, 1994

36. Dalakas MC: Polymyositis, dermatomyositis, and inclusion-body myositis. N Engl J Med 325:1487, 1991

37. Love LA, Leff RL, Fraser DD, et al: A new approach to the classification of idiopathic inflammatory myopathy: myositis-specific autoantibodies define useful homogeneous patient groups. Medicine (Baltimore) 70:360, 1991

38. Brouwer R, Hengstman GJ, Vree Egberts W, et al: Autoantibody profiles in the sera of European patients with myositis. Ann Rheum Dis 60:116, 2001

39. Sontheimer RD: Would a new name hasten the acceptance of amyopathic dermatomyositis (dermatomyositis sine myositis) as a distinctive subset within the idiopathic inflammatory dermatomyopathies spectrum of clinical illness? J Am Acad Dermatol 46:626, 2002

40. Olsen NJ, Park J: Skeletal muscle imaging for the evaluation of myopathies. Diseases of Skeletal Muscle. Wortmann R, Ed. Lippincott, Williams & Wilkins, New York, 1999, p 293

41. Fraser DD, Frank JA, Dalakas M, et al: Magnetic resonance imaging in the idiopathic inflammatory myopathies. J Rheumatol 18:1693, 1991

42. Reimers CD, Fleckenstein JL, Witt TN, et al: Muscular ultrasound in idiopathic inflammatory myopathies of adults. J Neurol Sci 116:82, 1993

43. Fleckenstein JL, Reimers CD: Inflammatory myopathies: radiologic evaluation. Radiol Clin North Am 34:427, 1996

44. Newman ED, Kurland RJ: P31 magnetic resonance spectroscopy in polymyositis and dermatomyositis: altered energy utilization during exercise. Arthritis Rheum 35:199, 1992

45. Sigurgeirsson B, Lindelof B, Edhag O, et al: Risk of cancer in patients with dermatomyositis or polymyositis. N Engl J Med 326:363, 1992

46. Stockton D, Doherty VR, Brewster DH: Risk of cancer in patients with dermatomyositis or polymositis, and follow-up implications: a Scottish population-based cohort study. Br J Cancer 85:41, 2001

47. Clarke AE, Bloch DA, Medsger TA, et al: A longitudinal study of functional disability in a national cohort of patients with polymyositis/dermatomyositis. Arthritis Rheum 38:1218, 1995

48. Joffe MM, Love, LA, Leff RL, et al: Drug therapy of the idiopathic inflammatory myopathies: predictors of response to prednisone, azathioprine, and methotrexate and a comparison of their efficacy. Am J Med 94:379, 1993

49. Villalba ML, Hicks JE, Thornton B, et al: A combination of oral methotrexate and azathioprine is more effective than high dose intravenous MTX with leucovorin rescue in treatment-resistant myositis. Arthritis Rheum 38:S307, 1995

50. Vencovsky J, Jarosova K, Machacek S: Cyclosporine A versus methotrexate in the treatment of polymyositis and dermatomyositis. Scand J Rheumatol 29:95, 2000

51. Chaudhry V, Cornblath DR, Griffin JW, et al: Mycophenolate mofetil: a safe and promising immunosuppressant in neuromuscular diseases. Neurology 56:94, 2001

52. Mowzoon N, Sussman A, Bradley WG: Mycophenolate (CellCept) treatment of myasthenia gravis, chronic inflammatory polyneuropathy and inclusion body myositis. J Neurol Sci 185:119, 2001

53. Adams EM, Pucino F, Yarboro C, et al: A pilot study: use of fludarabine for refractory dermatomyositis and polymyositis, and examination of endpoint measures. J Rheumatol 26:352, 1999

54. Dalakas MC, Illa I, Dambrosia JM, et al: A controlled trial of high-dose intravenous immune globulin infusions as treatment for dermatomyositis. N Engl J Med 329:1993, 1993

55. Danieli MG, Malcangi G, Palmieri C, et al: Cyclosporin A and intravenous immunoglobulin treatment in polymyositis/dermatomyositis. Ann Rheum Dis 61:37, 2002

56. Walter MC, Lochmuller H, Toepfer M, et al: High-dose immunoglobulin therapy in sporadic inclusion body myositis: a double-blind, placebo-controlled study. J Neurol 247:22, 2000

57. Dalakas MC, Koffman B, Fujii M, et al: A controlled study of intravenous immunoglobulin combined with prednisone in the treatment of IBM. Neurology 56:323, 2001

58. Jorizzo JL: Dermatomyositis: practical aspects. Arch Dermatol 138:114, 2002

VIII SYSTEMIC VASCULITIS SYNDROMES

Brian F. Mandell, m.d., ph.d.

The diagnosis of a primary vasculitic syndrome is dependent on documentation of vasculitis and the exclusion of diseases that can cause secondary vasculitis. The diagnosis of a specific primary vasculitic disorder depends on the pattern of organ involvement, the histopathology, and the size of affected blood vessels.

The major determinants of prognosis and therapy include the type of vasculitis, the severity and extent of critical organ involvement, the rate of disease progression, and the etiology, if identifiable. The inflammatory process is often associated with nonspecific symptoms and laboratory abnormalities (e.g., elevated erythrocyte sedimentation rate, anemia, and fevers) that do not distinguish vasculitic diseases from other inflammatory, infectious, or neoplastic diseases. The toxic nature of the therapies for systemic vasculitis dictates the need for an accurate diagnosis.

Approach to the Patient with Suspected Vasculitis

EVALUATION

The physician should not be reluctant to pursue invasive testing in the diagnostic evaluation of patients with a multisystem illness, but biopsy of clinically uninvolved tissue and the use of less specific tests should be eschewed. An approach directed toward "ruling in" a specific form of vasculitis and ruling out reasonable specific alternatives should be pursued.

The first step in the diagnosis of vasculitis is to perform a detailed patient history and physical examination to document specific organ involvement. Special attention should be paid to the skin, eyes, ears, upper airway, joints, urinalysis, lymph nodes, peripheral nerves, and large vessels. A few laboratory tests [*see Table 1*] should be selectively included in the initial evaluation. Specialized studies, including serologies, should be obtained only after a differential diagnosis is formulated. If the urine dipstick test indicates blood, leukocytes, or protein, the

Table 1 Selected Laboratory Tests for Patients with Multisystem Disease and Possible Vasculitis

Test	*Comments*
Platelet count	Thrombocytosis may parallel the acute-phase response Thrombocytopenia is not expected in primary vasculitic syndromes; consider SLE, marrow infiltration, hairy-cell leukemia, TTP, DIC, hypersplenism, APLS, HIV, scleroderma renal crisis, and heparin-induced thrombocytopenia
White blood cell count	Leukopenia is not expected in primary vasculitis; consider SLE, leukemia, hypersplenism, sepsis, myelodysplasia, and HIV Eosinophilia is common in Churg-Strauss syndrome; it may occur in WG, rheumatoid arthritis, or normotensive scleroderma renal crisis
Erythrocyte sedimentation rate	Relatively low ESR is seen in DIC, liver failure, and hyperviscosity; ESR is frequently normal in HSP, may be low in Takayasu arteritis, and is normal in ≤ 20% of giant cell arteritis
Transaminases	ALT or AST is elevated in liver disease, myositis, rhabdomyolysis, hemolysis, or myocardial necrosis
Anti–glomerular basement membrane	Useful for evaluation of alveolar hemorrhage, with or without glomerulonephritis; also useful for evaluation of normocomplementemic glomerulonephritis
Antinuclear antibody	Order when there is clinical suspicion of SLE, not as a general screening test for sick patients; negative test makes SLE very unlikely
Antineutrophil cytoplasmic antibody	Order when there is clinical suspicion of WG or MPA; order specific anti-PR3 and antimyeloperoxidase
Drug screen	Order for unexplained CNS symptoms, myocardial ischemia, vascular spasm, panic attacks with systemic features, or tachycardia; urine screen should be done
Blood cultures	Useful for any patient with febrile, multisystem, or wasting illness; pulmonary infiltrates; or focal ischemia/infarction. Cultures are easy to obtain
APLA/PTT/RVVT	Order for unexplained venous or arterial thrombosis or thrombocytopenia
Purified protein derivative (± anergy)	Use in any patient who may require steroid therapy or who has unexplained sterile pyuria or hematuria, granulomatous inflammation, chronic meningitis, or possible exposure to tuberculosis
Examination of fresh urinary sediment	Perform in all patients with an unexplained febrile or multisystem illness
Hepatitis serologies	Order for abnormal transaminases or elevated hepatic alkaline phosphatase; portal hypertension; PAN or MPA syndrome; or unexplained cryoglobulinemia, polyarthritis, or cutaneous vasculitis
Complement C3, C4	Not a screening test for vasculitis; useful in the differential diagnosis of glomerulonephritis; low in cryoglobulinemia; may be low in endocarditis; usually normal in PAN, MPA, HSP, WG; may be low in viral hepatitis–related glomerulonephritis or vasculitis
Aldolase	Aldolase has no organ specificity; it has similar organ distribution as lactic dehydrogenase

ALT—alanine aminotransferase APLA—antiphospholipid antibody APLS—antiphospholipid antibody syndrome AST—aspartate aminotransferase DIC—disseminated intravascular coagulation ESR—erythrocyte sedimentation rate HSP—Henoch-Schönlein purpura MPA—microscopic polyangiitis PAN—polyarteritis nodosa PR3—proteinase 3 PTT—partial thromboplastin time RVVT—Russell viper venom test SLE—systemic lupus erythematosus TTP—thrombotic thombocytopenic purpura WG—Wegener granulomatosis

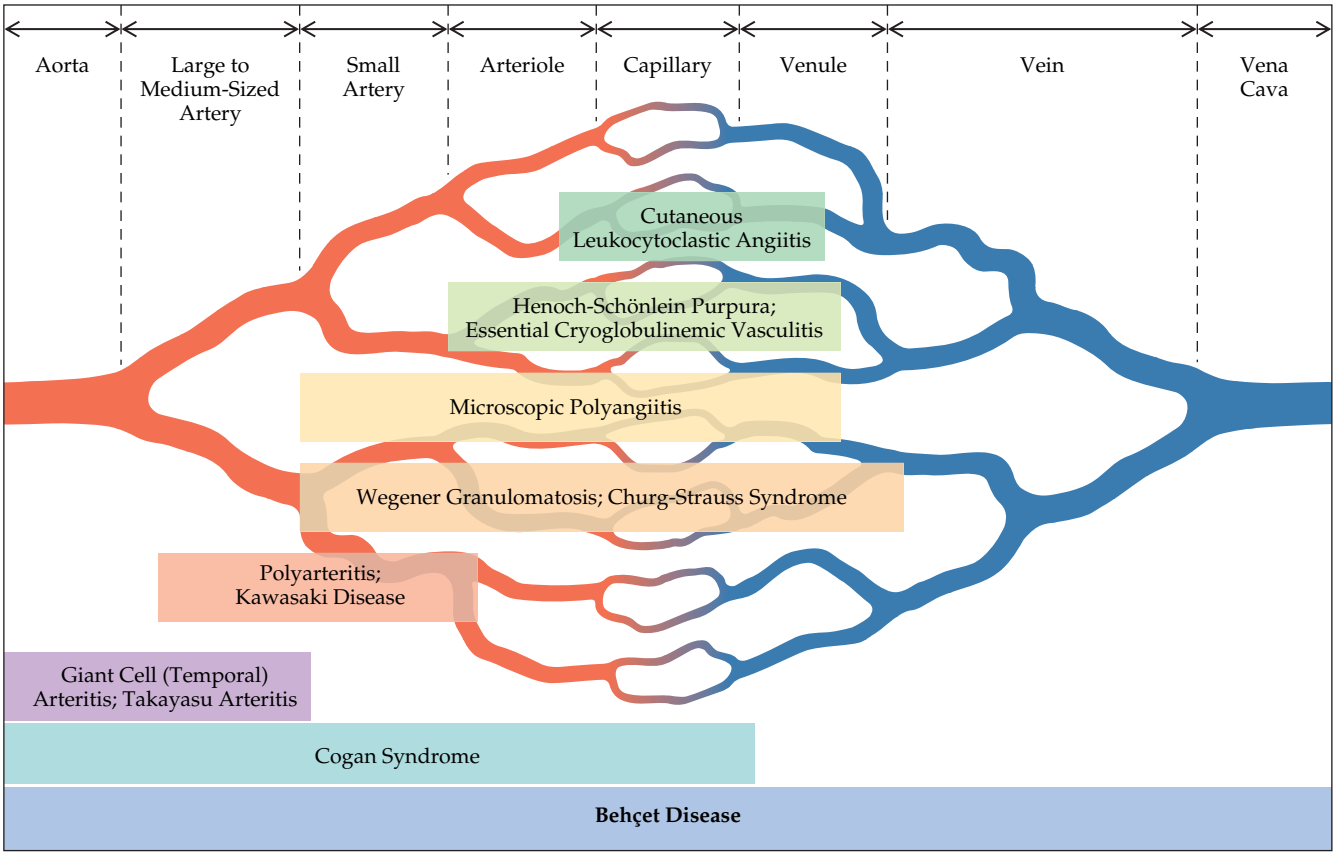


Figure 1 **Classification of the systemic vasculitis syndromes.**[1]

physician must promptly examine several fresh urine sediments. Urine that has been sitting for several hours before analysis is not as useful for identification of cellular casts, which rapidly degenerate ex vivo. The presence of red blood cell casts is highly suggestive of glomerulonephritis, but white cell casts may also be seen. Glomerulonephritis is usually asymptomatic. On the basis of the pattern of organ involvement, a differential diagnosis that includes specific types of systemic vasculitis and other disorders can then be generated.

CLASSIFICATION

Several classification schemes have been proposed for organizing the systemic vasculitic disorders into a consistent paradigm. These classifications are useful in distinguishing the clinical disorders that have distinct differences in prognosis and response to treatment.[1] No scheme is universally accepted. They all reiterate the characteristics of fulminant or classic disease, placing an emphasis on specificity of diagnosis. If a classification scheme is strictly adhered to, the newly ill patient without fully expressed disease is frequently left without a definitive diagnosis. The physician must recognize that until specific etiologies are defined, diagnostic entities remain conceptual, and overlap between diseases is not unusual. This must not be a deterrent to instituting therapy in the patient at risk for rapidly progressive organ damage. Nonetheless, classification systems provide useful constructs for communication and the design of research protocols [*see Figure 1*]. The most widely used classification schemes are based on the caliber of affected blood vessels, the pattern of organ involvement, and the presence or absence of granulomas, significant immune complex deposition, and eosinophilic infiltrates. Some authors have proposed a diagnostic role for the presence or absence of specific serum antineutrophil cytoplasmic antibodies (ANCAs), particularly antibodies to proteinase 3 and myeloperoxidase. At present, the appropriate role of these tests is to support a rationally developed clinical diagnosis, not to define one. In patients who do not fit neatly into a well-defined diagnostic category, these serologic tests should not supplant an attempt to obtain a tissue diagnosis. The presence of ANCA is not sufficient to make a diagnosis of a primary vasculitic syndrome; ANCA is not a screening test.

When the dominant symptoms and findings (i.e., neuropathy and cutaneous vasculitis) do not suggest a single specific vasculitic disorder, targeted physical examination and serologic testing may be helpful. Most valuable is biopsy confirmation of the specific disorder. The value of indiscriminate testing for antinuclear antibodies, ANCA, rheumatoid factor, and angiotensin-converting enzyme is arguable. Alternatively, infection with hepatitis B or C can be associated with a broad range of vasculitic syndromes, and these infections must be routinely excluded.[2]

OVERVIEW OF TREATMENT

The systemic vasculitides are potentially life threatening and may require potent anti-inflammatory and immunosuppressive therapy. Diagnoses should be made with as much certainty as possible. However, questions regarding alternative diagnoses or coexistent diseases may linger. Hence, even after therapy is initiated, physicians should maintain a high degree of vigilance to detect unrelated medical problems, complications of therapy,

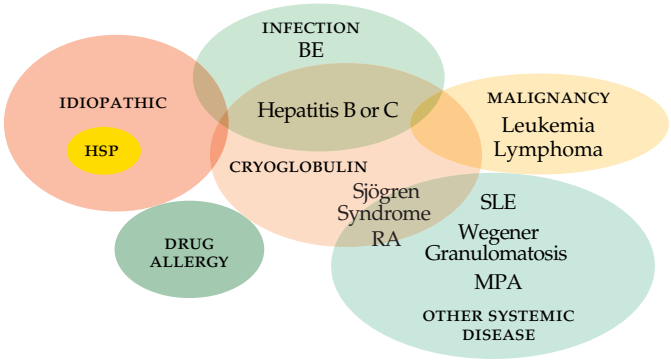


Figure 2 **A Venn diagram illustrates the relations between the causes of small vessel ("hypersensitivity") vasculitis. (BE—bacterial endocarditis; HSP—Henoch-Schönlein purpura; MPA—microscopic polyangiitis; RA—rheumatoid arthritis; SLE—systemic lupus erythematosus)**

or both. The signs and symptoms of unrecognized infection may transiently resolve with steroid therapy.[3] With the initiation of potent immunosuppressive therapy, there is a prolonged window of increased susceptibility to opportunistic infection. The greatest risks occur in patients with marked neutropenia or those receiving high doses of corticosteroids. Physicians must be particularly wary about attributing new problems to "flares" in the underlying disease without first excluding a new or recrudescent infection. Patients with varicella-zoster virus may present with fever and pain before appearance of the vesicles. *Pneumocystis carinii,* cytomegalovirus, and systemic fungal infections and reactivation of mycobacterial disease are observed more frequently in patients with systemic vasculitides than in the general population. Immunosuppression from steroids and other medications is frequently associated with mucosal candidiasis, less commonly associated with molluscum contagiosum, and rarely associated with Kaposi sarcoma.

Methotrexate, azathioprine, and cyclophosphamide may cause leukopenia and, less often, other cytopenias. In patients with decreased renal function, methotrexate must be used with caution, if at all; the dose of cyclophosphamide should be decreased and carefully monitored because the pro-drug (cyclophosphamide) is renally excreted. Bladder-emptying dysfunction is a relative contraindication to the long-term use of cyclophosphamide because increased exposure to toxic metabolites of the drug may predispose to bladder cancer or cystitis. The recent trend in the treatment of patients with certain potentially life-threatening systemic vasculitic syndromes has been to introduce therapy with a short course of corticosteroids (often with a second immunosuppressive agent), tapered from a high dose to a low dose, to induce remission and then, depending on the disease, to continue immunosuppressive therapy with an alternative regimen of corticosteroids to maintain remission. The second regimen may initially consist of cyclophosphamide, which is felt to be the most potent of these agents, but cyclophosphamide is then replaced with an agent that has a better safety profile (e.g., methotrexate or azathioprine). Therapy with that agent is then continued for many months.

Small Vessel Vasculitis

Vasculitis that affects capillaries and venules is the most common form of vasculitis and almost invariably involves the skin. It can occur at any age and affects men and women with equal frequency.

ETIOLOGY

Small vessel vasculitis can occur as an idiopathic disorder or secondary to drug allergy, bacterial endocarditis, viral infections such as those caused by hepatitis B or C, disseminated *Neisseria,* and rickettsiae; it can be part of a defined systemic autoimmune disorder such as Sjögren syndrome, systemic lupus erythematosus (SLE), or rheumatoid arthritis; or it can occur in association with hematologic, lymphoid, and solid-organ malignancies [*see Figure 2*]. Small vessel vasculitis can accompany diseases commonly associated with the involvement of larger vessels (e.g., Wegener granulomatosis [WG]).

DIAGNOSIS

Clinical Manifestations

Cutaneous involvement can occur in many of the primary or secondary vasculitic syndromes. Large, medium-sized, or small vessel occlusion can cause livedo, Raynaud phenomenon, or necrosis. Purpura is the most common manifestation of small vessel vasculitis. Small vessel vasculitis, particularly when associated with infections, is frequently associated with immune complex deposition. Vasculitis primarily involving the postcapillary venules has been termed hypersensitivity vasculitis in older literature.[4] Primary small vessel vasculitis may be limited to the skin or may be associated with visceral involvement, including alveolar hemorrhage, intestinal ischemia or hemorrhage, and glomerulonephritis.

Purpura tends to occur in recurrent crops of lesions of similar age and is more pronounced in gravity-dependent areas [*see Figure 3*]. When purpura is not primarily in gravity-dependent areas, cold agglutinin disease, cryoglobulinemia (which may be associated with an infection such as hepatitis C or with lymphoma), embolism, and infiltrative diseases should be excluded. Cutaneous vasculitis of any etiology may be associated with striking dependent edema.

In a case series of cutaneous small vessel vasculitis,[4] almost 100% of patients younger than 20 years had disease limited to the skin, whereas approximately 40% of the 172 patients older

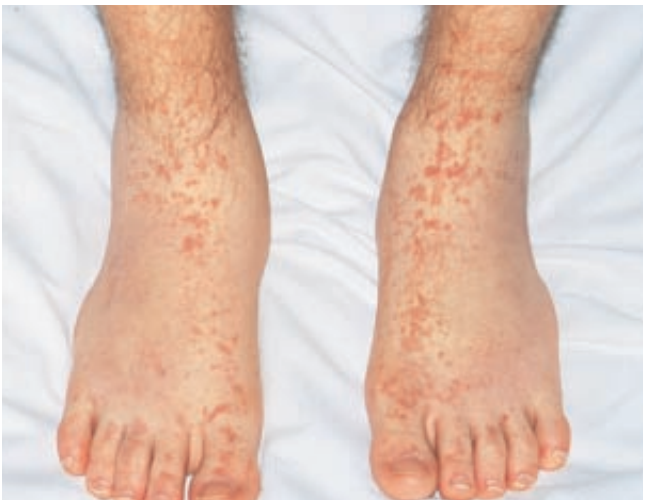


Figure 3 **Palpable purpura of the distal extremities is the most common presentation of small vessel vasculitis.**

Table 2 Immunosuppressive Therapies for Vasculitis

Drug	*Dose*	*Efficacy Rating*	*Comments*
Prednisone	Often used at 1 mg/kg daily (split doses in severe disease) initially; tapered, with goal of discontinuance by 6 months or sooner if possible; utilize other drugs to enable this if possible	Primary therapy in all forms of life- or organ-threatening forms of vasculitis; probably most rapid-acting therapy	Ideally, check baseline PPD status; consider prophylaxis against *Pneumocystis* (when using high doses) and osteoporosis; monitor for development of glaucoma in elderly patients
Cyclophosphamide	1–3 mg/kg p.o. daily; avoid neutropenia; nadir is usually 9–14 days after initiation of therapy or change in dose; decrease dose in setting of renal insufficiency; monthly "pulse" dosing has been used (0.5–1 g/m^2), but there may be greater likelihood of relapse; give pulse dose after dialysis	Most potent nonsteroidal immunosuppressive therapy; unclear onset of action but should be given when severe disease recognized, particularly rapidly progressive glomerulonephritis	Major side effects limit long-term use of this drug: leukopenia, myeloproliferative disease, bladder damage, and malignancy; current trend is to induce remission in WG and other severe forms of vasculitis with prednisone and cyclophosphamide, with tapering of prednisone and change of cyclophosphamide to a less toxic (but likely less effective) medication (e.g., azathioprine or methotrexate)
Azathioprine	2–3 mg/kg daily p.o.	Less potent than cyclophosphamide; useful to maintain remission while trying to spare corticosteroid dosing	Not usually given as primary induction therapy; avoid leukopenia; can cause a confusing hypersensitivity reaction that includes high fever, with or without rash and eosinophilia
Methotrexate	Given once weekly (up to approximately 0.3 mg/kg/dose) along with daily folic acid (1 mg)	Less potent than cyclophosphamide; useful to maintain remission while trying to spare corticosteroid dosing; decrease dose for mild renal insufficiency; *avoid in patients with creatinine > 2.5 mg/dl*	Useful in maintaining remission; has been used as primary induction therapy with prednisone in patients with mild WG; significant frequency of relapse in WG patients maintained on this drug alone; monitor WBC, creatinine and transaminase levels (causes hepatitis and can cause cirrhosis; avoid any ethanol ingestion); can be given orally or by weekly injection; folic acid reduces "nuisance" side effects

PPD—purified protein derivative WBC—white blood cell count WG—Wegener granulomatosis

than 20 years had an associated or underlying systemic disorder. Seventeen adults had a systemic necrotizing vasculitis, four had malignancy, four had a bacterial infection causing the vasculitis, 11 had cryoglobulinemia, and 59 had Henoch-Schönlein purpura. The prevalence of infection with hepatitis C virus, likely the most common cause of mixed cryoglobulinemia,[2] was not reported in this series.

Laboratory Tests

Biopsy is most useful in excluding causes of nonvasculitic purpura such as amyloidosis, leukemia cutis, Kaposi sarcoma, T cell lymphomas, and cholesterol or myxomatous emboli. Tissue immunofluorescent staining is useful to support the diagnosis of Henoch-Schönlein purpura (specifically, IgA staining), SLE, or infection **(**the percentage of cases with positive results on immunofluorescent staining is not known**)**. The cells infiltrating and perhaps destroying the vessel wall may be neutrophils or lymphocytes, depending on the etiology. The pathology in most cases of small vessel vasculitis is leukocytoclastic angiitis (LCA). Hepatitis C infection should be excluded routinely in patients who present with unexplained purpura—an important example of the fact that even the demonstration of LCA does not indicate that a patient's illness is the result of a primary vasculitic syndrome.

CLINICAL SUBSETS

Henoch-Schönlein Purpura

Henoch-Schönlein purpura is a clinically defined small vessel vasculitic syndrome in which cutaneous features are usually striking and in which significant visceral involvement is less common. Henoch-Schönlein purpura, which occurs less frequently in adults than in children,[5] is usually associated with vascular and renal deposition of IgA-containing immune complexes. Common manifestations of Henoch-Schönlein purpura include purpura; urticaria; abdominal pain; gastrointestinal bleeding or intussusception (mostly in children); arthralgias or arthritis; and glomerulonephritis. Visceral symptoms may precede the skin lesions. Henoch-Schönlein purpura may be precipitated by medications or streptococcal or viral infections. It is usually a self-limited disorder, but the associated glomerulonephritis may, in rare instances (most often in adults), progress to renal failure. In the absence of renal dysfunction, Henoch-Schönlein purpura is often a self-limited but frequently recurrent syndrome that may require only symptomatic therapy.

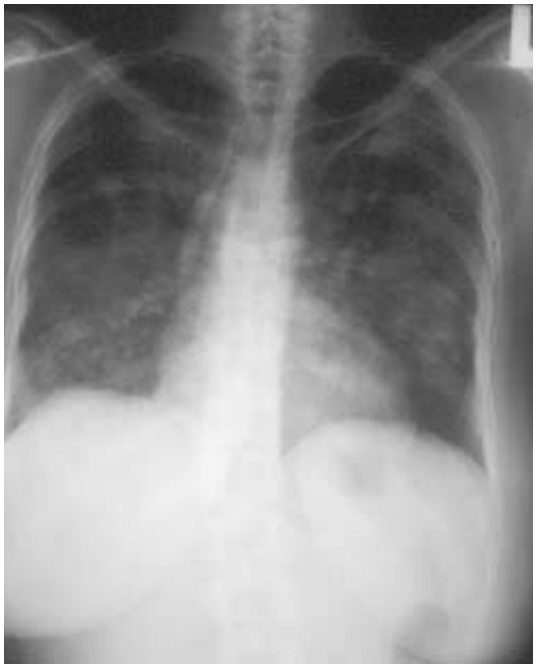
Urticarial Vasculitis

Urticarial vasculitis represents a peculiar subset of small vessel vasculitis.[6] The clinical presentation is that of wheals or serpentine papules, sometimes with surrounding or geographically separate angioedema. Individual lesions are slow to resolve, often lasting for several days; the disease follows a more prolonged course than typical urticaria. There is frequently a burning, dysesthetic discomfort from the lesions. Like purpura, the lesions of urticarial vasculitis are frequently located in gravity-dependent areas and often heal with skin hyperpigmentation or an ecchymotic area. Most cases are idiopathic, although an association with an underlying systemic autoimmune disorder such as SLE, IgM paraproteinemia, or a viral infection has been described. In rare cases, urticarial vasculitis has been associated with a syndrome that includes hypocomplementemia and interstitial pulmonary disease. This syndrome is distinct from C1 esterase deficiency associated angioedema, which does not cause urticaria.

TREATMENT

Therapy for cutaneous vasculitis is first directed at eliminat-

a



b

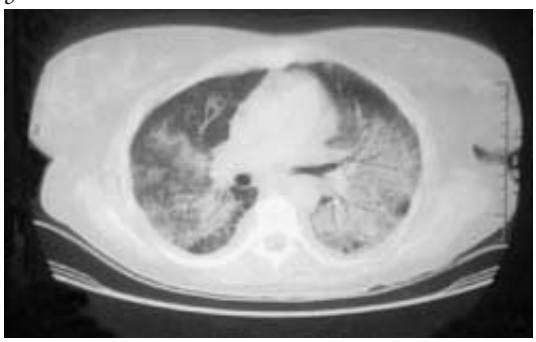


Figure 4 **The nodular infiltrates of the lung in Wegener granulomatosis are shown less extensively in a standard radiograph (*a*) than in a computed tomographic scan (*b*).**

ing any underlying precipitant. Infectious etiologies should be sought out and treated. Potential offending drugs should be withdrawn. Association with myelodysplasia and myeloproliferative disease should be considered, especially if there are any hematologic abnormalities. If no precipitants are apparent, low-risk therapy can be attempted with nonsteroidal anti-inflammatory drugs, colchicine, pentoxifylline, dapsone, or short-term low-dose corticosteroids. Long-term corticosteroid therapy should be eschewed if at all possible. Compressive support stockings or panty hose may be useful in limiting the significant edema that often accompanies cutaneous vasculitis of the legs.

Visceral involvement with organ dysfunction may necessitate a more aggressive approach than that used in limited cutaneous vasculitis. Moderate-dose corticosteroids are generally effective. In the setting of potential complications from chronic corticosteroid use or the setting of severe visceral involvement, methotrexate, azathioprine, cyclophosphamide, or other immunosuppressive agents may occasionally be required [*see Table 2*]. When treating chronic, refractory small vessel disease that is not organ or life threatening, one must pay close attention to the risk-to-benefit ratio of selected therapies.

Wegener Granulomatosis

WG is a relatively uncommon, potentially lethal disease characterized by necrotizing granulomatous inflammation and vasculitis of small and medium-sized vessels.[7,8] Males and females of all ages can be affected.

DIAGNOSIS

Clinical Manifestations

WG is characterized by parenchymal necrosis with a variable contributory component of vasculitis. Multiple organs are often involved; there is a predilection for the upper and lower respiratory tracts, eyes, and kidneys.

Upper respiratory tract involvement Upper airway disease may be striking but is often attributed for months or even years to routine sinus disease until other manifestations of WG are recognized. Even after the diagnosis is made and immunosuppressive treatment is provided, sinus disease may be recalcitrant to therapy. This chronicity may be caused in part by superinfection of damaged tissue by *Staphylococcus aureus.* Anatomic damage can include septal perforations and saddle-nose deformities. Laryngotracheal involvement may result in subglottic stenosis, which is best treated by local corticosteroid injection therapy. Ear involvement is common, particularly otitis media, which may produce conductive hearing loss. Orbital pseudotumors may cause proptosis with intractable pain and loss of vision; these inflammatory and fibrous masses may be refractory to anti-inflammatory therapy, immunosuppressive therapy, and even radiation therapy. Conjunctivitis, uveitis, and scleritis alone or in combination commonly occur.

Lower respiratory tract involvement Lung involvement may be absent at the onset of disease or present dramatically as diffuse alveolar hemorrhage. One third of pulmonary lesions noted on imaging studies [*see Figure 4*] are asymptomatic (CT scanning is more sensitive than radiography). Nodules often undergo necrosis leading to cavity formation. Bronchospasm is not characteristic of WG. If airway obstruction is suspected, bronchoscopy should be considered to exclude endobronchial or subglottic stenoses. It is frequently necessary to rule out infectious causes of the pulmonary infiltrates, and bronchoscopy is useful in this regard. However, tissue obtained from transbronchial biopsy is usually of insufficient quantity to confirm the pathologic diagnosis of WG.

Open lung biopsy is often the optimal method for demonstrating the typical pathologic findings of WG and for excluding malignancies and atypical infections. Typical open lung biopsies[9] may contain areas of necrosis, frequently in a broad pattern; giant cells in the parenchymal tissue; and vasculitis. Not all histopathologic features may be present in the same biopsy section. Because these findings may also occur in chronic mycobacterial or fungal infections, special stains and cultures for these agents are essential.

Glomerulonephritis Glomerulonephritis is a common cause of morbidity and mortality in WG. Its presence or absence defines the generalized or limited forms of the disease.

Table 3 Clinical Features of Vasculitis

Disorder	*Common Target Organs*	*Special Pathologic Features*	*Special Laboratory Studies*	*Comments*
Microscopic polyangiitis	Nerve, glomerulus, lung (small vessels), GI tract	No giant cells, vasculitis, proliferative GN (no or rare immune deposits*)	p-ANCA (antimyeloperoxidase)	Rule out hepatitis B and C
Polyarteritis nodosa	Nerve, GI tract	Arteritis of medium muscular arteries, no giant cells, no GN	No ANCA	No small vessel involvement; rule out hepatitis B and C
Wegener granulomatosis	Upper airway, eye, lung (small vessels), glomerulus, nerve, musculoskeletal system	Giant cells, geographic necrosis, mild eosinophilia, vasculitis, proliferative GN (no or rare immune deposits)	c-ANCA (anti-PR3)	Chronic sinus or ear disease
Churg-Strauss syndrome	Nerve, lung infiltrates, heart, skin	Giant cells, eosinophilia, vasculitis, proliferative GN (no or rare immune deposits)	Eosinophilia ± ANCAs	Positive atopic history

*Presence of immune deposits suggests possible hepatitis B or C infection.
ANCA—antineutrophil cytoplasmic antibody c-ANCA—cytoplasmic ANCA GN—glomerulonephritis p-ANCA—perinuclear ANCA PR3—proteinase 3

Glomerulonephritis is often aggressive, or it may be relatively indolent. It may be clinically and pathologically indistinguishable from idiopathic rapidly progressive crescentic glomerulonephritis, and it is usually clinically silent. The evolution from subclinical to dialysis-dependent renal disease may occur over several weeks. Glomerulonephritis may be present at the outset of the disease, or it may develop only after the patient has been ill with an apparently limited form of the disease. The importance of frequent microscopic urinalyses in the initial and follow-up evaluation of patients with WG cannot be overemphasized. Especially in elderly or debilitated patients, valuable information may be obtained by occasional 24-hour urine collections, which can establish a more accurate estimate of the glomerular filtration rate (GFR) than that provided by the serum creatinine measurement. Renal biopsy may reveal focal and segmental glomerulonephritis with variable glomerular proliferative changes, crescent formation, and necrosis, in the absence of significant immune complex deposition. Although supportive of the diagnosis of WG, these findings are not diagnostic of the disease, and renal biopsy is not the preferred study to confirm the specific diagnosis of WG.

Additional clinical manifestations Musculoskeletal involvement occurs in over half of patients with WG. Symptoms may include arthralgias or arthritis; these symptoms may be migratory, additive, or of fixed distribution. Rheumatoid factor is frequently present in patients with WG, and it may cause diagnostic confusion with rheumatoid arthritis when joint symptoms are significant. The joint disease of WG only rarely produces bone erosions. Neurologic signs and symptoms occur in fewer than 50% of patients, peripheral neuropathy in fewer than 20%, and involvement of the central nervous system in fewer than 10%. Oculomotor defects may occur because of impingement by a retro-orbital mass or sinus disease. Gastrointestinal ischemia and ulcerations are infrequent but may be confused with inflammatory bowel disease, especially because the latter can be associated with ANCA (usually perinuclear ANCA, or p-ANCA). Up to 50% of WG patients exhibit cutaneous involvement with purpura, panniculitis, or ulcerations. The activity of the skin disease generally parallels systemic disease activity.

Laboratory Tests

Unexplained chronic inflammation of the respiratory tract or eye or the presence of glomerulonephritis is consistent with the diagnosis of WG. The probability of WG is increased when multiple organ involvement is present, upper airway disease is destructive, and pulmonary nodules (especially with cavities) are demonstrated by radiography. Any combination of organ involvement is possible, but most patients exhibit upper airway involvement at the time of diagnosis.

If the entire clinical picture is compatible with WG and if alternative diagnoses have been appropriately ruled out, the finding of circulating cytoplasmic ANCA (c-ANCA) with anti–proteinase 3 specificity is sufficient to make the provisional diagnosis and initiate therapy without a tissue diagnosis. Approximately 20% of patients with WG may have p-ANCA with antimyeloperoxidase specificity. If there are any atypical features or special concerns regarding the initiation of immunosuppressive therapy or if the patient does not respond appropriately to therapy, histopathologic confirmation of the diagnosis is mandatory. The presence of ANCA is not equivalent to the presence of vasculitis; ANCAs can be found in other diseases. The ANCA level is not a reliable means to follow disease activity.[10-12] Because WG generally requires therapy with a glucocorticoid plus a cytotoxic agent, it should be distinguished from other inflammatory disorders, including other vasculitic syndromes [*see Table 3*], which may be effectively treated with a less toxic regimen.

TREATMENT

Initial treatment of generalized WG virtually always requires dual-drug immunosuppressive therapy. Corticosteroids may produce symptomatic improvement in the upper airway, lungs, skin, and musculoskeletal system, but tapering usually results in a flare in the disease. Acutely serious disease, particularly renal disease that is progressing, is treated initially with corticosteroids and daily cyclophosphamide with subsequent tapering of the corticosteroids over several months. Many authors recommend that once remission is achieved, cyclophosphamide therapy should be replaced by methotrexate or azathioprine therapy for an additional 12 months of therapy [*see Table 2*]. There are some strong relative contraindications to the long-term use of cyclophosphamide, including bladder dysfunction (increased risk of drug metabolite–induced cystitis and bladder cancer) and leukopenia. In milder or limited WG, weekly doses of methotrexate (0.20 to 0.30 mg/kg, adjusted for renal function) with folic acid or leucovorin may be substituted for cyclophosphamide. Patients undergoing treatment with immunosuppressives must be continuously monitored for flares

in disease, opportunistic infections, and side effects. Flares may be more frequent in patients treated with methotrexate than in those receiving longer courses of cyclophosphamide.[12] Side effects include cytopenias and drug-induced pneumonitis. Methotrexate may cause hepatitis, marrow suppression, and, on rare occasions, cirrhosis. It should be avoided in the setting of renal insufficiency or alcohol use. Some authors have suggested using trimethoprim-sulfamethoxazole as adjunctive therapy for the treatment of WG and for prevention of bacterial infections that may promote flares of upper airway disease. This approach remains highly controversial. Administration of trimethoprim-sulfamethoxazole three times weekly is useful in protecting patients against *P. carinii* pneumonia while they are receiving intensive immunosuppressive therapy. Local nasal and sinus toilet and otolaryngoscopic evaluations are a routine part of the care of patients with upper airway disease. Prophylactic measures to prevent osteoporosis should always be considered when corticosteroids are used on a long-term basis.

Churg-Strauss Syndrome

Churg-Strauss syndrome (CSS), or allergic granulomatosis angiitis, is a rare syndrome that affects small to medium-sized arteries and veins in association with bronchial asthma.

DIAGNOSIS

Clinical Manifestations

CSS displays clinical similarities to WG in terms of organ involvement and pathology, especially in patients with upper or lower airway disease or glomerulonephritis. It can follow a rapidly progressive course. CSS differs most strikingly from WG in that the former occurs in patients with a history of atopy, asthma, or allergic rhinitis, which is often ongoing. In the prevasculitic atopy phase, as well as during the systemic phase of the illness, eosinophilia is characteristic and often of striking degree (≥ 1,000 eosinophils/mm^3). When eosinophilia is present in WG, it is usually more modest (~500 eosinophils/mm^3).

Systemic features of CSS include some combination of pulmonary infiltrates, cardiomyopathy, coronary arteritis, pericarditis, polyneuropathy (symmetrical or mononeuritis multiplex), ischemic bowel disease, eosinophilic gastroenteritis, ocular inflammation, nasal perforations, glomerulonephritis, cutaneous nodules, and purpura.[13,14]

The patchy pulmonary infiltrates of CSS are often transient and may be associated with alveolar hemorrhage. Pulmonary nodules are uncommon and rarely cavitate. Pleural effusions are common and often contain abundant eosinophils. Clinical distinction from hypersensitivity pneumonitis, allergic aspergillosis, and pulmonary lymphoma is at times difficult. Several cases of CSS have been reported to have occurred after the introduction of inhibitors of 5-lipoxygenase and while patients with chronic bronchial asthma are being weaned off corticosteroids.

Cardiac disease can be severe and is a leading cause of mortality. Valvular heart disease is not as striking or as common as it is in the idiopathic hypereosinophilic syndrome. Neurologic involvement occurs in more than 60% of patients. Such involvement may be severe; it is generally attributable to arteritis. Cutaneous purpura, urticaria, polymorphous erythematous eruptions, and nodules occur. Gastrointestinal involvement resulting from ischemic vasculitis, eosinophilic gastroenteritis, or both may cause pain, cramping, and diarrhea.

Laboratory Tests

Histopathology typically exhibits extravascular granulomatous inflammation, with a prominent eosinophilic infiltrate and vasculitis. Vasculitis in a given tissue section may be granulomatous or nongranulomatous. Granulomas can be found in tissue at areas separate from the demonstrable vasculitis. Eosinophilic infiltrates are more striking than in WG. Neither abundant eosinophils, granulomas, nor giant cells are found in classic polyarteritis nodosa (PAN) or microscopic polyangiitis (MPA). The pathology of the nodules is not by itself sufficient to make a diagnosis of CSS, because similar pathology can be seen in lymphoma and sarcoidosis. Glomerulonephritis is frequently not as severe as in WG, but when present, it is usually focal and segmental and indistinguishable from other forms of so-called pauci-immune glomerulonephritis (i.e., glomerulonephritis that is without significant tissue deposition of immune complexes).

TREATMENT

CSS is generally responsive to corticosteroid therapy. Most patients are able to be withdrawn from steroids. However, bronchial asthma and sinus disease may require ongoing therapy, even if the vasculitic component of the disease has remitted. Patients with severe or refractory visceral organ involvement are empirically treated with additional agents such as cyclophosphamide, methotrexate, or azathioprine; the corticosteroids are tapered after remission is achieved [*see Table 2*].

Polyarteritis Nodosa and Microscopic Polyangiitis

CLASSIFICATION

Attempts to separate PAN and MPA, two forms of necrotizing small to medium-sized vessel arteritis, have not been universally accepted. A recent international conference proposed that the diagnosis of these disorders be based on the absence of granulomatous inflammation in both and by involvement of arterioles, capillaries, venules, and glomerular capillaries in MPA but *not* in PAN. Older studies of patients with PAN did not uniformly make this distinction. Even more important, patients with viral hepatitis B or C were not excluded from older studies. The recognition of viral hepatitis is crucially important because chronic hepatitis B or C[2,15] can elicit a secondary vasculitic syndrome indistinguishable from PAN or MPA in presentation but distinct in response to therapy.[16] MPA involves vessels ranging in size from capillaries and venules to medium-sized arteries [*see Figure 1*].[17] Clinically, MPA can mimic WG, although some authors have arbitrarily defined MPA as excluding involvement of the upper airway.

DIAGNOSIS

Clinical Manifestations

Glomerulonephritis, particularly rapidly progressive glomerulonephritis, and alveolar hemorrhage are common in MPA and absent, by definition, in classic PAN.

PAN affects the medium-sized muscular arteries and, like MPA, is associated with peripheral neuropathy and bowel ischemia.[18-20] Azotemia and hypertension in PAN may occur because of arteritis of the renal arteries but not because of glomerulonephritis. Microaneurysm formation in medium-sized visceral arteries may be striking, and they may rupture.

Constitutional symptoms such as fever, asthenia, and myalgias are common in both PAN and MPA. Elevated acute-phase reactants, thrombocytosis, leukocytosis, and the anemia of inflammatory disease are common, although they are not uniformly present.

When the clinical syndrome of PAN or MPA is suspected, bacterial infection (e.g., endocarditis) and viral infection (e.g., hepatitis B or C) must be excluded. The association with hepatitis B or C infection may not dramatically alter the presentation of the PAN or MPA syndrome, except that membranous glomerulonephritis, cryoglobulinemia, immune complex–associated glomerulonephritis, hepatic failure, and thrombocytopenia are more likely to occur with viral hepatitis–associated vasculitis.

Antiphospholipid antibody syndrome (APLS) can mimic PAN by presenting as mesenteric ischemia or renal insufficiency caused by thrombotic occlusion of mesenteric and renal vessels.[21] Features of APLS and arteritis affecting muscular arteries include livedo reticularis [*see Figure 5*]. Glomerulonephritis cryoglobulinemia, immune complex–associated glomerulonephritis, and peripheral neuropathy are not expected in APLS unless the patient also has SLE. Thrombocytopenia can occur with APLS but is not expected in PAN. Cholesterol embolization should also be considered as a cause of livedo, renal insufficiency, eosinophilia, and constitutional symptoms[22]; the clinical history of a recent vascular procedure and the performance of a biopsy will help confirm the diagnosis.

Laboratory Tests

The diagnosis of MPA and PAN should ideally be based on histopathologic demonstration of arteritis and the clinical pattern of disease. A biopsy specimen of clinically involved, nonnecrotic tissue that demonstrates the presence of arteritis of muscular arteries is the ideal supportive finding for the diagnosis of arteritis of a medium-sized vessel, but such a biopsy is not always possible. The presence of serum p-ANCA with antimyeloperoxidase specificity (in 60% of MPA patients) supports the clinical diagnosis of MPA, but p-ANCA is not specific for this disease. ANCAs are not characteristic of PAN. MPA is a form of pauci-immune glomerulonephritis; that is, the renal biopsy tissue in MPA, as in WG and CSS, does not contain extensive immune complexes on immunofluorescent staining and electron microscopy. Lung biopsy in the setting of pulmonary infiltrates or hemorrhage reveals capillaritis, a histopathologic pattern that can also be seen in WG, SLE, and anti–glomerular basement membrane disease. Biopsy is most useful in ruling out alternative pulmonary diagnoses; open lung and thoracoscopic techniques have a higher yield for demonstrating vasculitis than transbronchial biopsy. Classic PAN does not cause glomerulonephritis or pulmonary parenchymal disease.

The demonstration of arteritis in PAN may be difficult, especially in the setting of dominant constitutional symptoms and the absence of easily accessible, disease-affected tissue. Biopsy efforts should be directed toward tissue that is abnormal as demonstrated by symptoms or objective testing. Sural nerve biopsy has become a popular option when attempting to diagnose an arteritis that is affecting medium-sized muscular vessels. The sural nerve is an accessible pure sensory nerve, and its vasa nervorum contains small as well as medium-sized muscular arteries. Nerve conduction studies can identify a diseased ischemic sural nerve before the appearance of clinical symptoms.[23] Multiple reports have emphasized the low diagnostic yield from the biopsy of asymptomatic and electrically normal nerve. Even nerves exhibiting abnormal conduction have reportedly showed no diagnostic pathology 46% of the time.[24] There is notable morbidity associated with sural nerve biopsy; 13 of 60 patients experienced wound infections or delayed healing, and three patients suffered from new pain in the distribution of the sural nerve that underwent biopsy.[24] Biopsy of clinically uninvolved tissue (i.e., asymptomatic muscle) has a diagnostic yield of less than 30%.

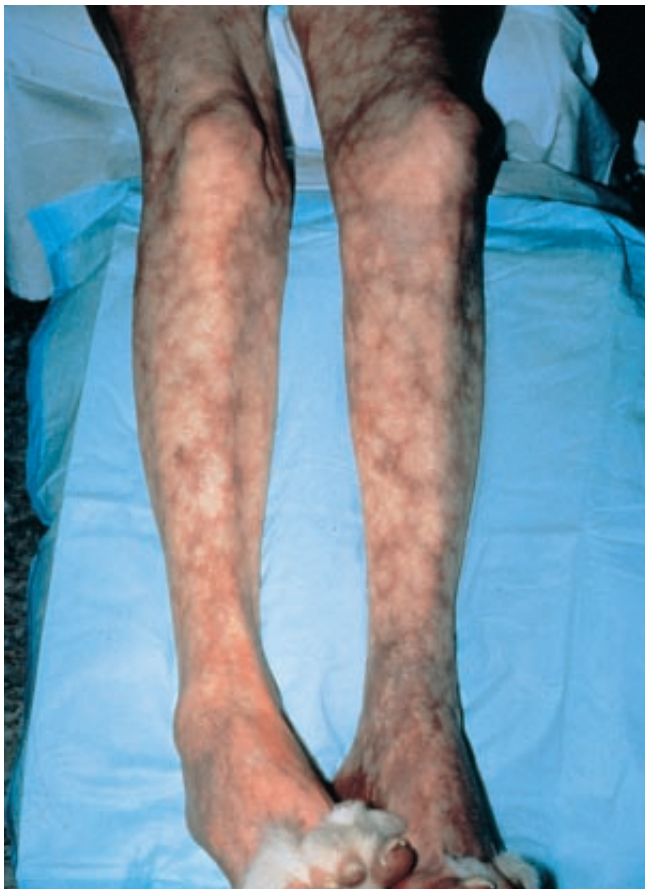


Figure 5 **Livedo reticularis is characterized by reddish-blue mottling of the extremities caused by occlusion of the deep dermal arterioles.**

Abdominal angiography is frequently utilized in the evaluation of patients who may have medium-sized vessel arteritis when biopsy has been unrewarding or is not an option. Arteries affected by polyarteritis nodosa and other disorders of medium-sized muscular arteries may develop microaneurysms or stenoses that can be visualized by angiography. When angiography is used in an effort to diagnose systemic necrotizing vasculitis in the absence of pathologic evidence of the disease, several caveats must be noted. Angiography has limited spatial resolution; smaller vessels are not well seen. In patients with primarily smaller vessel disease, the angiogram will not likely be diagnostic. In one study, angiograms were diagnostic in only four of 30 patients with MPA, a disease that affects both small and medium-sized arteries.[17] Different investigators have reported aneurysms in 60% to 90% of patients with PAN. Aneurysms take time to develop and may not be present early in the course of the illness. In addition to being associated with aneurysms, arteritis may be associated with stenoses, which may be longer and smoother than typical atherosclerotic lesions or occlusion. To maximize the yield from the procedure, angiography should in-

clude the celiac, renal, and mesenteric vessels. Lack of clinical involvement of an organ (i.e., no intestinal ischemia) does not exclude the possibility of finding abnormal vessels on angiography. It has been suggested that the visualization of aneurysms in PAN denotes more severe disease; it is unclear whether their presence may alternatively relate to the actual duration of the untreated illness. Aneurysms may resolve with successful treatment of primary or viral hepatitis–associated disease. The presence of visceral microaneurysms is not diagnostic of PAN. They have also been anecdotally described in patients with WG and MPA, likely representing medium-sized muscular artery involvement in these diseases. Microaneurysms also occur in nonvasculitic disorders. Isolated case reports have described aneurysms in patients with atrial myxoma, bacterial endocarditis, peritoneal carcinomatosis, or severe arterial hypertension and after methamphetamine abuse. Inadequate data are available to assess the sensitivity and specificity or the predictive value of abdominal angiography in the diagnosis of necrotizing arteritis. As is the case when interpreting a biopsy result of suspected vasculitis, imaging studies must be considered in the light of the entire clinical profile. Angiography is generally avoided in the setting of progressive or significant renal insufficiency.

TREATMENT

Treatment of both PAN and MPA is empirical[25] [*see Table 2*]. Corticosteroids in high doses (1 mg/kg daily of prednisone or its equivalent) remain the initial mainstay of therapy for both disorders in the acutely ill patient. Use of corticosteroids alone may be sufficient in patients who do not have critical organ involvement, defined as renal insufficiency, gastrointestinal ischemia, cardiomyopathy, or dense peripheral neuropathy. Therapy with corticosteroids alone may fail more frequently in MPA than in PAN, given the tendency for frequent relapses in MPA.[17] Patients who require long-term corticosteroid therapy for disease control or patients who have clinical markers of severe disease are usually treated with glucocorticoids and an additional immunosuppressive agent such as cyclophosphamide. The indications for initial combination therapy have not been adequately studied.

When active hepatitis B or C infection is present, a relatively short course of steroids should be considered on the basis of disease severity and the organs at acute risk for failure, in conjunction with aggressive antiviral therapy.

Kawasaki Disease

Kawasaki disease (KD) was first described in 1967 as mucocutaneous lymph node syndrome.[26] It typically affects infants and young children, causing dominant cutaneous manifestations, fever, and coronary arteritis. It can on rare occasions affect adults.

DIAGNOSIS

The presence of characteristic clinical features has permitted the establishment of diagnostic criteria for KD [*see Table 4*]. Vasculitis may involve vessels ranging in size from venules to the aorta. Prominent inflammation is noted in the larger coronary arteries, which results in aneurysm formation in approximately 25% of untreated patients. The immediate and delayed life-threatening cardiac complications of the disease, coupled with its unique therapy (aspirin and intravenous γ-globulin), mandates prompt clinical diagnosis. Biopsy is generally not necessary, nor is it likely to yield a specific diagnosis.

High, spiking fevers may persist for 1 to 2 weeks if left untreated. Rapid defervescence is usually observed with initiation of appropriate therapy. Nonexudative conjunctivitis often appears with the fever. Aseptic (lymphocytic) meningitis is common. Oral involvement includes erythema, dryness and fissuring of the lips, nonexudative pharyngitis, and tongue erythema with very prominent papillae. Mucosal ulcerations are not characteristic of this illness. Distal limb swelling may appear days after the fever, with erythema and tenderness that are not limited to the joints. Desquamation, often in sheets, may begin days to a few weeks after the onset of fever. When desquamation occurs early in KD, it may appear concurrently with a truncal rash and eye and lip changes; it may mimic a drug reaction or Stevens-Johnson syndrome. The rash is usually diffuse and polymorphous, with urticarial, morbilliform, annular, or plaque components, but it is not vesicular. Adenopathy, which is present in 75% of patients, is most apparent in the cervical region.

The morbidity and mortality (< 3%) of KD is overwhelmingly associated with the development of inflammatory coronary artery aneurysms, most of which are asymptomatic at the time of formation. Aneurysms may be detected by echocardiography. Thrombosis can occur in the aneurysms, resulting in direct or embolic coronary artery occlusion. Coronary events may occur weeks or even many years after the febrile illness. A baseline echocardiogram should be obtained at the time of the acute illness and repeated 2 and 6 weeks later. Early recognition of the disease and treatment with intravenous immunoglobulin and aspirin have significantly decreased the frequency of aneurysm formation and thrombotic coronary events.

TREATMENT

Treatment of KD should be initiated with intravenous immunoglobulin (2 g/kg as a single dose) and aspirin (80 to 100 mg/kg/day every 6 hours) as soon as the disease is seriously suspected.[27] Aspirin is more effective than corticosteroids in preventing aneurysms. Corticosteroid therapy is usually unnecessary, and some authors feel that it is relatively contraindicated. Symptoms tend to respond within several days after the institution of aspirin and intravenous immunoglobulin. In resistant cases, however, corticosteroids are frequently added to the above therapies.

Table 4 Diagnostic Criteria for Kawasaki Disease

- Persistent fever (> 5 days)
  - *plus*
- Four of the following five conditions:
  - Nonpurulent bilateral conjunctivitis
  - Oral mucosal involvement
    - Erythematous pharynx
    - Red or fissured lips
    - Strawberry tongue
  - Soft tissue abnormalities of hands and feet
    - Edema/erythema
    - Desquamation
  - Polymorphous, nonvesicular rash
  - Cervical adenopathy

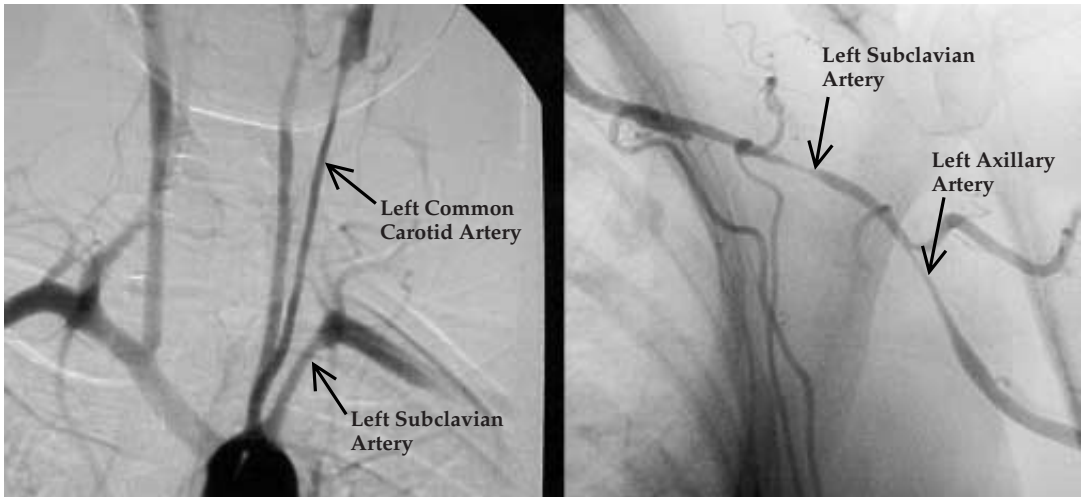


Figure 6 **Angiograms of a patient with Takayasu aortitis demonstrating long, smooth stenotic lesions of the left subclavian artery and involvement of other branches of the aortic arch vessels.**

Large Vessel Arteritis

Temporal, or giant cell, arteritis (GCA) of the elderly and Takayasu arteritis (TA) are the most common inflammatory diseases of the aorta and its major branches. Similar vascular targeting may occur in Behçet disease, Cogan syndrome, and sarcoidosis. The last two conditions are recognized by the pattern of extra-aortic organ involvement. It is uncertain whether TA and GCA are distinct disorders or are the same disorder with modified expression in different age groups.

TEMPORAL OR GIANT CELL ARTERITIS

GCA generally affects individuals older than 50 years.[28,29] In many patients, it is associated with the syndrome of polymyalgia rheumatica (PMR). PMR is characterized by proximal muscle pain, with nocturnal and early morning worsening. There may be a subjective sense of weakness, without true weakness on examination and without elevation of serum muscle enzyme levels.

GCA is variably associated with fever, scalp tenderness, headache, masticatory muscle claudication, peripheral vascular disease, inflammatory aortic aneurysms, and retinal ischemic syndromes. Oligoarticular arthritis, often in the upper extremity, and acute carpal tunnel syndrome can occur. The ischemic symptoms and signs may be clinically indistinguishable from those occurring in arteriosclerotic obliterative disease.

Examination for disparate four-extremity blood pressure readings, abdominal aneurysms, and bruits must be part of the routine follow-up visits of patients with GCA or PMR. Pathologic findings of GCA can occur in superficial temporal arteries of patients with PMR, even without any symptoms of GCA. However, routine biopsy of the superficial temporal arteries in patients with PMR, without any other symptoms of GCA, is not warranted.

Levels of acute-phase reactants are elevated in more than 80% of patients. Definitive diagnosis of GCA is generally made by biopsy of the superficial temporal artery. Pathology in GCA usually reveals chronic mononuclear cell infiltrates; destruction of the internal elastic lamina; and giant cells. The presence of giant cells is not requisite to make the diagnosis. The presence of characteristic clinical features such as new headache and jaw claudication, especially with concurrent PMR, may allow for a presumptive diagnosis in the absence of a biopsy or even when the superficial temporal artery biopsy is negative. However, because other conditions can mimic GCA, including atherosclerosis, an attempt to diagnose GCA by biopsy is warranted in most patients.[30] Corticosteroid therapy will not rapidly affect the biopsy results and should not be withheld from a patient strongly suspected of having GCA who is awaiting biopsy. Bilateral superficial temporal artery biopsy increases the diagnostic yield.

TAKAYASU ARTERITIS

Takayasu arteritis (pulseless disease) is a chronic inflammatory disease affecting the aorta and its major branches.[31] Usually diagnosed in younger, predominantly female patients of reproductive age, TA can also occur in young children and older patients of either sex. TA is more commonly associated with stenoses and aneurysms of the aorta and aortic branch vessel than is GCA.

The presenting clinical syndrome may include a prolonged flulike illness, including a polymyalgia rheumatica pattern of muscle pain. Many patients initially present with symptoms of limb, cerebral, or cardiac ischemia in the absence of any constitutional features. The characteristic features of the disease reflect the ischemia produced by the inflammatory stenoses of the aorta and its major branches. Renal ischemia can elicit high renin hypertension. Predominant sites of stenosis are the aortic arch vessels, particularly the subclavian arteries [*see Figure 6*]. Arm claudication with bruits is common. Superficial artery pain and tenderness (e.g., carotidynia) may be found on examination but are not diagnostic of TA. Severe central hypertension caused by renal artery stenosis may not be recognized because of coexistent arm artery stenosis; thus, four-extremity blood pressure readings must be evaluated initially and monitored on a frequent basis. Occasionally, stenoses exist in all major vessels of the extremities, and cuff monitoring may be an unreliable measure of central aortic pressures. Stroke is not uncommon and is often related to undetected central hypertension. It is extremely difficult to assess the activity of TA; the presence or absence of constitutional features or elevated acute-phase reactants are poor measures of disease activity. This impression is supported by vessel histopathology ob-

tained during reconstructive surgery. Over 40% of vascular specimens from patients thought to be in remission revealed active inflammation.

Diagnosis of TA is usually made by arteriographic demonstration of stenotic lesions; aneurysms are less commonly observed. The entire arch, as well as the abdominal aorta and renal vessels, should be evaluated. It is of paramount importance that central arterial pressure be routinely obtained at the time of angiography and compared with simultaneously obtained arm and leg cuff pressures. The role of sequential vascular magnetic resonance imaging in the evaluation and follow-up of these patients is currently under investigation.[32] This technique may reveal therapy-related changes in vessel wall thickness and edema as well as changes in lumen size. Pathologic documentation is difficult to obtain in TA, but the histopathology, usually obtained at the time of bypass surgery, is similar to that for GCA. Preoperative discussion with the vascular surgeon is mandatory to ensure that appropriate tissue samples are obtained if possible.

TREATMENT OF GCA AND TA

Corticosteroids are the initial treatment for both TA and GCA. GCA is generally very responsive to steroid therapy, although the most appropriate initial dose remains controversial. Initial daily doses of between 20 mg and 1 mg/kg have been advocated, with tapering over 8 to 12 months. It is generally recommended (without the support of data from controlled trials) that patients with any symptoms of ocular ischemia be initially treated with high-dose corticosteroids (at least 1 mg/kg of prednisone or its equivalent, with some authors suggesting I.V. methylprednisolone in doses of up to 1 g daily for several days). A significant proportion of patients with GCA require several years of therapy. Measurement of acute-phase reactants provides an imperfect index of disease activity and should not be the sole guide for adjustment of steroid dosing. If significant steroid side effects occur or if patients experience relapses during tapering, a second-line agent such as methotrexate is often added on an empirical basis to the corticosteroid therapy. However, the value of adjunctive steroid-sparing agents in GCA is currently unproved. A recent large prospective, randomized trial was unable to demonstrate a positive effect from methotrexate therapy.[33] Vascular reconstructive surgery, angioplasty, and stent placement are adjunctive therapeutic options in some patients. Although very preliminary experience suggests a high degree of stent failure, the frequent involvement of the subclavian vessels in TA must be taken into consideration when choosing the graft implantation site for coronary or carotid bypass procedures. High-dose corticosteroid therapy, especially in the elderly, has potentially dangerous side effects. Special attention must be paid to the prevention of opportunistic infections, osteoporosis, glaucoma, hyperglycemia, and hyperlipidemia.

The author has received grants from Abbott Laboratories and Wyeth Pharmaceuticals, and he has received grants from and served as advisor or consultant for Pfizer/Pharmacia and Centocor, Inc.

The drugs cytoxan, corticosteroids, methotrexate, pentoxifylline, colchicine, and dapsone have not been approved by the FDA for uses described in this chapter.

References

1. Jennette C, Falk RJ, Andrassy K, et al: Nomenclature of systemic vasculitides: proposal of an international consensus conference. Arthritis Rheum 37:187, 1994

2. Vassilopoulos D, Calabrese L: Hepatitis C virus infection and vasculitis. Arth Rheum 46:585, 2002

3. Lawrence EC, Mills J: Bacterial endocarditis mimicking vasculitis with steroid-induced remission. West J Med 124:333, 1976

4. Blanco R, Martinez-Taboada VM, Rodriguez-Valverde V, et al: Cutaneous vasculitis in children and adults: associated disease and etiologic factors in 303 patients. Medicine (Baltimore) 77:403, 1998

5. Szer IS: Henoch-Schönlein purpura: when and how to treat. J Rheumatol 23:1661, 1996

6. O'Donnell B, Black AK: Urticarial vasculitis. Int Angiol 14:166, 1995

7. Hoffman GS, Kerr GS, Leavitt RY, et al: Wegener's granulomatosis: an analysis of 158 patients. Ann Intern Med 116:488, 1992

8. Rheinhold-Keller E, Beuge N, Latza U, et al: An interdisciplinary approach to the care of patients with Wegener's granulomatosis: long term outcome in 155 patients. Arth Rheum 43:1021, 2000

9. Travis WD, Hoffman GS, Leavitt RY, et al: Surgical pathology of the lung in Wegener's granulomatosis. Am J Surg 15:315, 1991

10. Hoffman GS: Classification of the systemic vasculitides: antineutrophil cytoplasmic antibodies, consensus and controversy. Clin Exp Rheumatol 16:111, 1998

11. Hoffman GS, Specks U: Antineutrophil cytoplasmic antibodies: diagnostic value in systemic vasculitis. Arthritis Rheum 41:1521, 1998

12. Reinhold-Keller E, Fink CO, Herlyn K, et al: High rate of renal relapse in 71 patients with Wegener's granulomatosis under maintenance of remission with low-dose methotrexate. Arth Rheum 47:326, 2002

13. Guillevin L, Cohen P, Gayraud M, et al: Churg-Strauss syndrome: clinical study and long-term follow up of 96 patients. Medicine (Baltimore) 78:26, 1999

14. Reid AC, Harrison BW, Watts RA, et al: Churg-Strauss syndrome in a district hospital. Q J Med 91:219, 1998

15. Hadziyannis SJ: The spectrum of extrahepatic manifestations in hepatitis C virus infection. J Viral Hepat 4:9, 1997

16. Guillevin L, Lhote F, Cohen P, et al: Polyarteritis nodosa related to hepatitis B virus: a prospective study with long-term observation of 41 patients. Medicine (Baltimore) 74:238, 1995

17. Guillevin L, Durand-Gasselin B, Cevallos R, et al: Microscopic polyangiitis: clinical and laboratory findings in 85 patients. Arthritis Rheum 42:421, 1999

18. Travers RL, Allison DJ, Brettle RP, et al: Polyarteritis nodosa: a clinical and angiographic analysis of 17 cases. Semin Arthritis Rheum 8:184, 1979

19. Guillevin L, Lhote F, Gayraud M, et al: Prognostic factors in polyarteritis nodosa and Churg-Strauss syndrome. Medicine (Baltimore) 75:17, 1996

20. Mandell BF, Hoffman GS: Differentiating the vasculitides. Rheum Dis Clin North Am 20:409, 1994

21. Triplett DA: Protean clinical presentation of antiphospholipid-protein antibodies. Thromb Haemost 74:329, 1995

22. Om A, Ellahham S, DiScascio G: Cholesterol embolism: an underdiagnosed clinical entity. Am Heart J 124:1321, 1992

23. Wees SJ, Sunwoo IN, Oh SJ: Sural nerve biopsy in systemic necrotizing vasculitis. Am J Med 71:525, 1981

24. Rappaport WD, Valente J, Hunter GC, et al: Clinical utilization and complications of sural nerve biopsy. Am J Surg 166:252, 1993

25. Guillevin L, Lhote F: Treatment of polyarteritis nodosa and microscopic polyangiitis. Arthritis Rheum 41:2100, 1998

26. Schulman ST, Inocencio JD, Hirsch R, et al: Kawasaki disease. Pediatr Clin North Am 21:1013, 1995

27. Leung DY, Schlievert PM, Meissner HC: The immunopathogenesis and management of Kawasaki syndrome. Arthritis Rheum 41:1538, 1998

28. Evans JM, Hunder GG: Polyangiitis rheumatica and giant cell arteritis. Clin Geriatr Med 14:455, 1998

29. Hoffman GS: Treatment of giant cell arteritis: where we have been and why we must move on (review). Cleve Clin J Med 69(suppl II):SII117, 2002

30. Ponge T, Barrier JH, Grolleau JY, et al: The efficacy of selective unilateral temporal artery biopsy versus bilateral biopsies for diagnosis of giant cell arteritis. J Rheumatol 15:997, 1988

31. Kerr GS, Hallahan CW, Giordano J, et al: Takayasu's arteritis. Ann Intern Med 120:919, 1994

32. Tso E, Flamm SD, White RD, et al: Takayasu arteritis: utility and limitations of magnetic resonance imaging in diagnosis and treatment. Arth Rheum 46:1634, 2002

33. Hoffman GS, Cid MC, Hellmann DB, et al: A multicenter, randomized, double-blind, placebo-controlled trial of adjuvant methotrexate treatment for giant cell arteritis. Arthritis Rheum 46:1309, 2002

Acknowledgments

Figures 1 and 2 Seward Hung.

Figure 6 Gary S. Hoffman.

IX CRYSTAL-INDUCED JOINT DISEASE

Christopher Wise, m.d.

The presence of precipitated crystals in the synovium or synovial fluid can be associated with an inflammatory response that usually manifests itself as an acute arthritis associated with synovial fluid leukocytosis. The identification of monosodium urate (MSU) crystals in the synovial fluid of patients with acute gout in 1961 by McCarty and Hollander represented the initial recognition of arthritis associated with articular crystal deposition.[1] This development was followed in 1962 by the recognition of so-called pseudogout, which is associated with calcium pyrophosphate dihydrate (CPPD) crystals.[2] Since then, a great deal has been learned about these two common types of arthritis. In addition, the role of crystals in the pathogenesis of osteoarthritis and other arthropathies has been further explored. The diagnosis of crystal-induced arthritis requires the identification of crystals in synovial fluid or tissue; in most cases, this acute arthritis is self-limited.

Gout

DEFINITION AND CLASSIFICATION

Gout is defined as an arthritis associated with the presence of MSU crystals in synovial fluid or tissue. Gout is often classified as primary or secondary [*see Table 1*], and both forms are associated with hyperuricemia. Gout associated with an inborn error in metabolism or decreased renal excretion without other renal disease is referred to as primary gout, whereas gout associated with an acquired disease or the use of a drug is called secondary gout. In both primary and secondary gout, chronic hyperuricemia may be the result of overproduction of uric acid caused by increased purine intake, synthesis, or breakdown, or it may be the result of decreased renal excretion of urate.

EPIDEMIOLOGY

Gout is predominantly a disease of middle-aged men, but there is a gradually increasing prevalence in both men and women in older age groups. The annual incidence of gout in men in most studies is in the range of one to three per 1,000, but the incidence is much lower in women. In the Framingham Study, for example, the 2-year incidence of gout was 3.2 per 1,000 men versus 0.5 per 1,000 women.[3] A study of gout in Rochester, Minnesota, over the past 2 decades has suggested that the incidence of gout may be increasing.[4]

The overall prevalence of self-reported gout in the general population is 0.7% to 1.4% in men and 0.5% to 0.6% in women. However, in people older than 65 years, prevalence increases to 4.4% to 5.2% in men and 1.8% to 2.0% in women.[5] In male populations, the prevalence of gout reaches impressive levels by the fifth decade. In a study of male medical students, the prevalence of gout reached 5.8% in whites and 10.9% in African Americans surveyed for a mean of 28 years after graduation.[6] In patients who experience the onset of gout after 60 years of age, the prevalence in men and the prevalence in women are almost equal; in those who experience onset after 80 years of age, the prevalence is greater in women.[7]

The incidence and prevalence of gout are parallel to the incidence and prevalence of hyperuricemia in the general population. Serum urate levels increase by 1 to 2 mg/dl in males at the time of puberty, but females exhibit little change in urate levels until after menopause, when concentrations approach those seen in males.[8] Most patients with elevated serum uric acid levels do not have gout, but hyperuricemia is clearly associated with an increased risk of gout.[9] For example, in persons with serum urate levels greater than 10 mg/dl, the annual incidence of gout is 70 per 1,000 and the 5-year prevalence is 30%, whereas in persons with levels less than 7 mg/dl, the annual incidence is only 0.9 per 1,000 and the 5-year prevalence is 0.6%. Additional factors that correlate strongly with serum urate levels and the prevalence of gout in the general population include serum creatinine levels, body weight, height, blood pressure, and alcohol intake.

PATHOGENESIS AND ETIOLOGY

The development of gout tends to be associated with chronically increased serum levels of uric acid. However, a substantial minority of patients with acute gout have normal uric acid levels, and hyperuricemia does not always lead to the development of gout. Humans are one of the few species with an inactive uricase gene, which results in elevated levels of uric acid. It has been postulated that humans have a propensity to develop hyperuricemia because uric acid confers protection against degenerative diseases by acting as an antioxidant.[10,11] Urate at high levels and under certain conditions will precipitate into MSU crystals, and the deposition of these crystals within the synovium or synovial fluid may lead to the development of gout (see below).

Table 1 Classification of Hyperuricemia and Gout

Primary Hyperuricemia and Gout with No Associated Condition	*Secondary Hyperuricemia and Gout with Identifiable Associated Condition*
Uric acid undersecretion (80%–90%)	Uric acid undersecretion
Idiopathic	Renal insufficiency (any cause)
Urate overproduction (10%–20%)	Polycystic kidney disease
Idiopathic	Lead nephropathy
HGPRT deficiency	Drugs
PRPP synthetase overactivity	Diuretics
	Salicylates (low dose)
	Pyrazinamide
	Ethambutol
	Niacin
	Cyclosporine
	Didanosine
	Urate overproduction
	Myeloproliferative diseases
	Lymphoproliferative diseases
	Hemolytic anemias
	Polycythemia vera
	Other malignancies
	Psoriasis
	Glycogen storage disease
	Dual mechanism
	Obesity
	Ethanol consumption
	Hypoxemia and tissue hypoperfusion

HGPRT—hypoxanthine-guanine phosphoribosyltransferase
PRPP—phosphoribosylpyrophosphate

Gout has been recognized as a familial disorder since the time of Sir Alfred Garrod. About 40% of patients in most series report a family history of gout, and the hereditary component for serum uric acid levels in the general population has been estimated to be approximately 40%.[8,12] The mechanisms for this association are still not understood, but most available data suggest that serum uric acid levels are controlled by multiple genes involving both production and excretion of uric acid.

Hyperuricemia

The plasma concentration of uric acid is maintained at a relatively constant level in humans because of a balance between production and excretion. Uric acid derives from exogenous and endogenous sources: it is the end product of the metabolism of dietary purines and occurs as a result of the breakdown of purines from nucleic acids during cell turnover. A very small amount of uric acid is passively eliminated through the gastrointestinal tract. Almost all plasma uric acid is filtered at the glomerulus, and 80% is reabsorbed in the proximal tubule. Some of this plasma uric acid is subsequently secreted back into the lumen; a small amount undergoes distal reabsorption.[8,11]

Hyperuricemia can result from decreased renal excretion or increased production of uric acid. In 80% to 90% of patients with primary gout, hyperuricemia is caused by renal underexcretion of uric acid, even though renal function is otherwise normal. The defect in renal excretion of uric acid in patients with primary gout may be attributed to reduced filtration, enhanced reabsorption, or decreased secretion, but it is unclear which of these mechanisms is most important. Even patients with high levels of urate excretion (overproducers) demonstrate a relative decrease in urate clearance compared to patients with normal levels of uric acid production.[13]

Hyperuricemia may develop secondary to numerous conditions (e.g., renal insufficiency, myeloproliferative diseases, obesity, alcohol consumption, and drug intake) [*see Table 1*]. Patients with secondary gout related to renal disease are hyperuricemic because of a decreased filtered load of uric acid, although decreased tubular secretion may play a role in some patients. Patients with lead nephropathy seem to be particularly prone to the development of gout, and recent studies have suggested that subclinical exposure to environmental lead may contribute to some of the hyperuricemia and gout seen in the general population.[14,15] The hyperuricemia associated with diuretic therapy results from volume depletion, which leads to a decreased filtered load, and from enhanced tubular reabsorption.[16] A renal mechanism is the cause of most other cases of drug-associated hyperuricemia. Low-dose aspirin can cause significant changes in renal handling of urate within a week after therapy is started, particularly in elderly patients.[17] Hyperuricemia and gout may be associated with cyclosporine therapy in renal and cardiac transplantation patients, and it appears to be the result of a combined effect of cyclosporine on renal blood flow and tubular function.[18-20]

Overproduction of uric acid, caused by increased purine synthesis, is seen in about 10% to 20% of patients with primary gout. In addition, four specific heritable defects of purine synthesis have been identified: phosphoribosylpyrophosphate synthetase overactivity, glucose-6-phosphatase deficiency, fructose-1-phosphate aldolase deficiency, and hypoxanthine-guanine phosphoribosyltransferase (HGPRT) deficiency. Of these heritable defects, the best-known is HGPRT deficiency. Complete deficiency of this enzyme is associated with the Lesch-Nyhan syndrome in children, and a partial deficiency has been associated with early-onset gout and nephrolithiasis.

Most diseases that cause secondary hyperuricemia characterized by overproduction of uric acid are associated with increased nucleic acid turnover. These diseases include multiple myeloma, polycythemia, pernicious anemia, hemoglobinopathies, thalassemia, other hemolytic anemias, other myeloproliferative and lymphoproliferative disorders, and other neoplasms. In addition, some critically ill patients may experience hyperuricemia resulting from accelerated breakdown of adenosine triphosphate (ATP).

Uric Acid Precipitation and Crystal-Induced Inflammation

Uric acid dissociates almost completely to the urate anion form at a pH of 7.4. At concentrations greater than 6.5 to 7.0 mg/dl, urate precipitates in the form of MSU crystals. Local conditions in tissues responsible for crystal precipitation and deposition include lower temperature (as is found in peripheral joints), lower pH level in extracellular fluid, and reduced urate binding to plasma proteins. Other local factors that contribute to precipitation and deposition of crystals are trauma and rapid increases in local urate concentration as a result of mobilization of water from peripheral tissues (as occurs when edematous feet are elevated during sleep).

The factors responsible for the inflammatory response to crystals are not completely understood.[8,21] The phlogistic properties of crystals seem to be linked to their ability to bind immunoglobulins and other proteins, particularly complement and lipoproteins. These complexes bind to surface receptors on macrophages and mast cells, leading to activation and release of proinflammatory cytokines, chemotactic factors, and other mediators. An influx of phagocytic cells—particularly neutrophils—follows. Crystals are engulfed, and subsequent disruption of lysosomes releases arachidonate metabolites, collagenases, and oxygen radicals. Several factors have been postulated as contributing to the self-termination of attacks. These factors include digestion of crystals by myeloperoxidase, increased heat and blood flow leading to dissolution and removal of crystals from the joint, alteration of the crystal properties by the inflammatory process itself, and phagocytosis of crystals by more mature macrophages later in the attack.[22,23]

CLINICAL STAGES

Acute Gouty Arthritis

Acute gouty arthritis is usually characterized by a sudden and dramatic onset of pain and swelling, usually in a single joint. This condition occurs most often in lower extremity joints and evolves within hours to marked swelling, warmth, and tenderness. The process often extends beyond the confines of the joint and may mimic cellulitis. The pain of gout is often severe enough to make even the light pressure of bedclothes intolerable, and weight bearing is usually very difficult. Even without treatment, attacks of gout usually subside within a few days, although some attacks may last a few weeks. Early in the course of gout, affected joints usually return to normal after attacks.

The initial attack of gout is monoarticular in 85% to 90% of patients. At least half of initial attacks occur in the first metatarsophalangeal joints (a condition known as podagra), but other joints of the foot may be involved simultaneously or in subsequent attacks. Other lower extremity joints, including the ankles

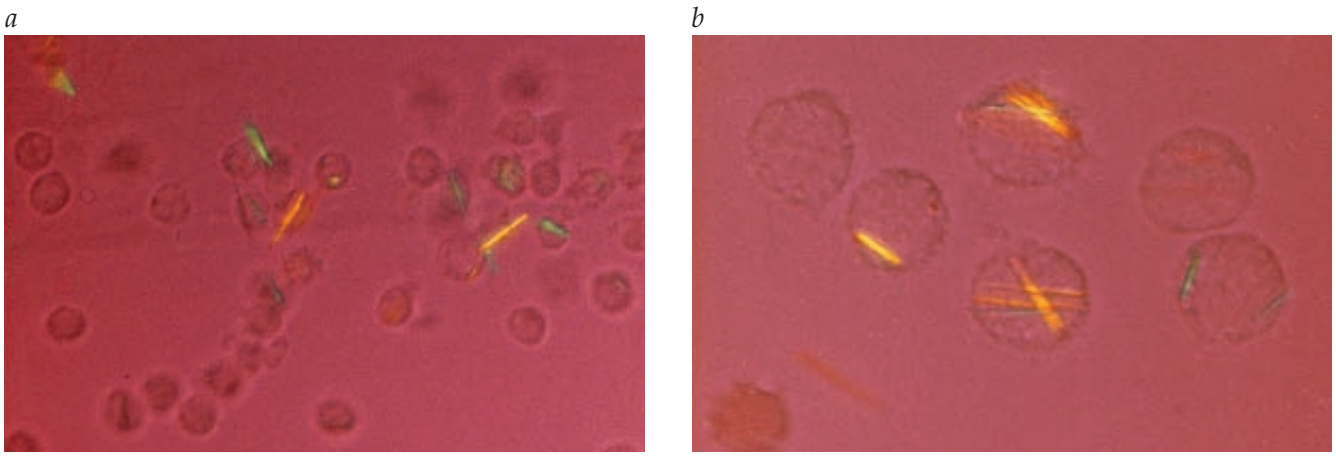


Figure 1 **Gout can be diagnosed by demonstration of negatively birefringent monosodium urate crystals in synovial fluid examined by polarized-light microscopy, either free (*a*) or within polymorphonuclear leukocytes (*b*).**

and knees, are often affected; in more advanced gout, attacks may occur in upper extremity joints, such as the elbow, wrist, and small joints of the fingers. In older women in particular, involvement of the small joints of the fingers (previously affected by osteoarthritis) is more commonly seen earlier in the course of the disease.[24,25] Acute episodes may also involve the bursae, particularly in the olecranon or prepatellar areas. Polyarticular gout occurs as the initial manifestation in about 10% to 15% of patients and may be associated with fever.[26]

Almost all synovial fluid aspirated early in an acute attack contains typical needlelike crystals, which are negatively birefringent and may be extracellular or may occur within polymorphonuclear leukocytes [*see Figure 1*]. The leukocyte count in most gouty synovial fluid rises to a range of 10,000 to 60,000/mm^3, but it may be much higher in some patients.

Intercritical Gout

After the initial attack of gout subsides, the clinical course of gout may follow one of several patterns. A minority of patients never have another attack of gout, and some may not have another attack for several years. Most patients, however, have recurrent attacks over a period of years. In a study done before the use of hypouricemic agents, 78% of patients had a second attack within 2 years and 93% had a second attack within 10 years.[27] In many patients, symptom-free intervals between attacks become progressively shorter as episodes of acute arthritis increase in frequency. In chronic disease, soft tissue swelling and joint effusions persist for longer periods after each attack. Finally, after 10 to 20 years of recurrent gouty attacks, patients typically develop chronic tophaceous gout.

Chronic Tophaceous Gout

Persistent hyperuricemia with increasingly frequent attacks of gout eventually leads to joint involvement of wider distribution and chronic joint destruction resulting from deposition of massive amounts of urate in and around joints [*see Figure 2a*]. Without therapy to lower serum uric acid levels, the average interval from the first gouty attack to the development of chronic arthritis or tophi is about 12 years.[28] After 20 years, 75% of patients have tophi; patients with the highest urate levels are at highest risk. In

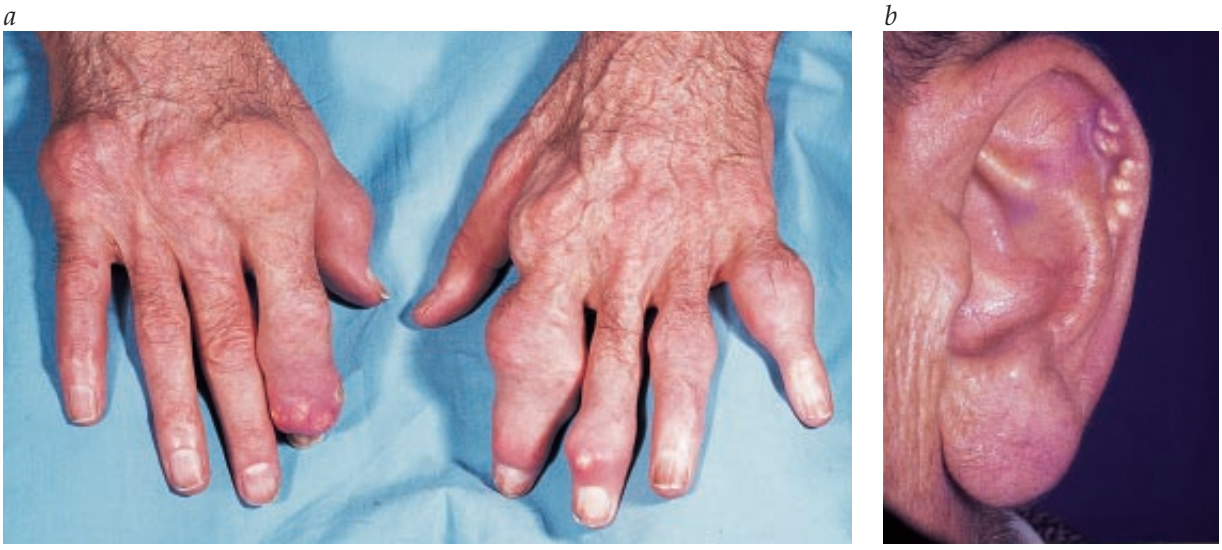


Figure 2 **Tophaceous gout, demonstrating chronic swelling in and around the joints of the hand caused by bone destruction and tophaceous deposits in the hands (*a*). Tophi may also be found in extra-articular areas, such as the pinna of the ear (*b*).**

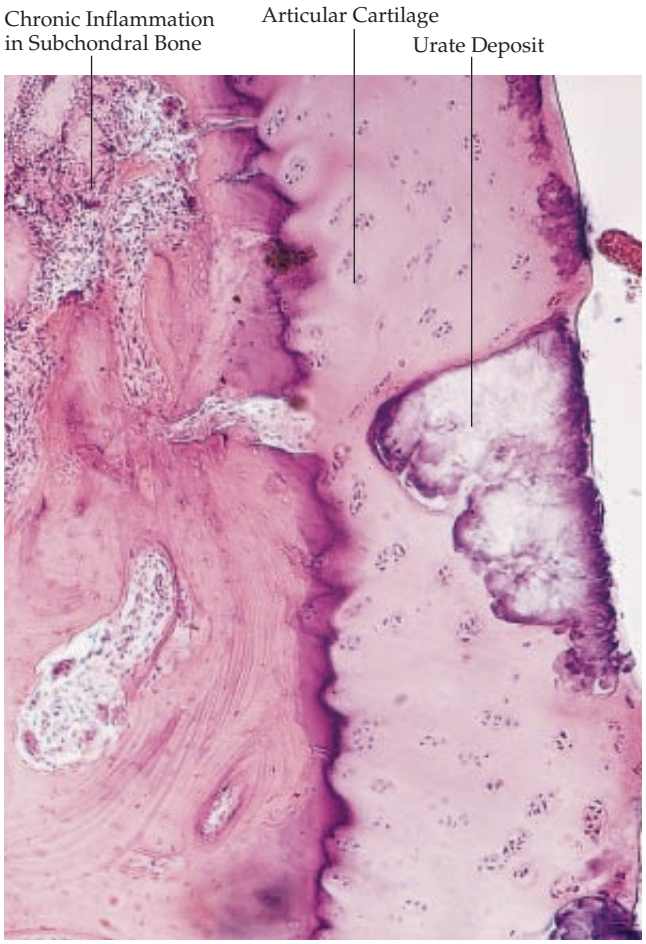


Figure 3 **Microscopic appearance of sodium urate deposits causing a defect in articular cartilage and chronic inflammation in the subchondral bone.**

elderly patients, particularly women, tophi may appear earlier in the course of the disease, sometimes in patients without a history of gouty attacks.[25]

Subcutaneous tophi begin to appear in periarticular and bursal tissues, especially around the knees and elbows, along tendons of the hands and feet, and around the interphalangeal and metacarpophalangeal joints of the hands. Tophaceous deposits are usually firm and movable, and the overlying skin may be normal or thin and reddened. When close to the surface, deposits exhibit a characteristic chalky appearance and may be cream-colored or yellowish. Tophi have also been described in areas not associated with joints, such as the pinna of the ear [*see Figure 2b*], and in unusual visceral locations, such as the myocardium, pericardium, aortic valves, and extradural spinal regions.

Destruction of the articular cartilage and subchondral bone eventually occurs in patients with chronic articular involvement [*see Figure 3*]. Erosive bony lesions may be seen on x-rays as well-defined punched-out lesions in periarticular bone, often associated with overhanging edges of bone.[29] These erosions are usually 5 mm or more in diameter and are larger than those seen in rheumatoid arthritis. Bone mineralization appears to be generally normal in chronic tophaceous gout, and periarticular osteopenia, which is seen in rheumatoid arthritis, is usually not present. The distribution of destructive joint disease in gout is often asymmetrical and patchy.

Associated Conditions

A number of chronic illnesses may be associated with gout and hyperuricemia, either in primary or secondary form. The best-known association is with renal disease, which frequently occurs in patients with gout.[30,31]

Most cases of renal disease in patients with primary forms of gout are believed to be the result of nephrosclerosis related to hypertension. However, experimental evidence suggests that hyperuricemia may play a direct pathogenetic role in the development of renal disease and hypertension.[32] In addition, the presence of intrarenal urate deposits associated with an inflammatory reaction in some patients and the improvement of renal function associated with control of uric acid levels suggest that urate has a causal role in the renal disease seen in patients with chronic tophaceous gout.[33,34]

Renal stones occur in 10% to 25% of patients with gout.[35] Most stones in patients with gout are composed of uric acid. However, some are composed of calcium oxalate and other constituents, and hyperuricemia is believed to contribute to the formation of these stones as well.[36]

An acute urate nephropathy associated with the tumor lysis syndrome has been described in patients with leukemia or lymphoma who are undergoing chemotherapy.[37] This condition is associated with acute oliguria and an elevated urinary urate-to-creatinine ratio (> 1.0) and is usually treated prophylactically with allopurinol and vigorous hydration. An association of gout with renal disease and chronic lead intoxication has been noted in some populations (saturnine gout).[38] In the United States, this association has most often been attributed to the drinking of illicit whiskey produced in lead-lined stills, but it has also been attributed to occupational lead exposure.

Gout has long been associated with obesity, diabetes mellitus, hyperlipidemia, and atherosclerotic cardiovascular disease. The association with diabetes has been variously reported; in lipid disorders, the association is primarily with hypertriglyceridemia, which may in turn be linked to alcohol intake.[39] In addition, a correlation has been found between hyperuricemia and the insulin-resistance syndrome (also referred to as the metabolic syndrome), possibly associated with body-fat distribution and triglyceride levels.[40,41] The association of gout and hyperuricemia with cardiovascular disease appears to be related to the link between these metabolic disorders and hypertension.[42]

Alcohol consumption has long been associated with the precipitation of gouty attacks in susceptible patients. In addition, long-term heavy alcohol consumption promotes hyperuricemia by interfering with renal excretion and increasing production of urate; some alcoholic beverages, particularly beer, serve as a source of dietary purine.[9] Patients with gout have an increased prevalence of hypothyroidism, and urate levels have been shown to decrease with the institution of thyroid replacement therapy, probably through a renal mechanism.[43]

DIAGNOSIS

A diagnosis of gout can be made with certainty only by confirmation of the presence of monosodium urate crystals in synovial fluid or tissue. Elements of the patient's history, physical examination, and laboratory studies can be very helpful in diagnosing gout [*see Table 2*].[44] A typical presentation of podagra in a middle-aged man with known hyperuricemia may be sufficient for an initial tentative diagnosis of gout, particularly if the condition responds well to colchicine. Nodular deposits on the olecra-

Table 2 Diagnosis of Gout

Test	*Sensitivity*	*Specificity*	*Comments*
Microscopic examination of synovial fluid by polarized microscopy	High during acute attacks; some potential between attacks	100% specific	Presence of needle-shaped, negatively birefringent urate crystals is diagnostic
Serum uric acid levels	Unreliable, even during attacks (elevated levels are a risk factor for, but not diagnostic of, gout)	Low in unselected patients (only 30% of people with uric acid > 10 mg/dl will have gout over 5 yr)	Serum levels are lower during attacks in some studies; serial serum levels over months and years may suggest risk of gout and severity
X-rays	Moderate, only in patients with chronic disease; even typical erosions may be difficult to differentiate from other erosions (e.g., as seen in rheumatoid arthritis)	Moderate, depending on nature of changes; otherwise may be difficult to differentiate from other forms	May be useful in patients with chronic disease if interpreted carefully (see text)

non processes, dorsal aspects of the fingers, or finger pads should be sought, particularly in patients with a history of joint problems. Patients with gout may have a normal serum urate level at the time of an attack. With most patients, however, a review of old records reveals a history of chronic hyperuricemia. Radiographs are seldom useful during an acute attack, unless previous attacks have occurred in the area examined and unless, after years of disease, well-defined erosions in or around joints, with characteristic overhanging edges, can be seen.

The detection of needle-shaped, negatively birefringent urate crystals in synovial fluid examined under polarized light microscopy is the definitive diagnostic finding for gout. Although this test is best done on fluid obtained during an acute attack, aspiration of synovial fluid from previously affected joints or aspiration of a subcutaneous nodule suspected of being a tophus may be helpful.[45,46] The synovial fluid should be examined by someone experienced in crystal identification, because an inexperienced person may not recognize the presence of crystals.[47]

Alternative diagnoses should be considered in all patients suspected of having gout. Acute arthritis can be caused by infection, other crystal-induced arthropathies, or other diseases. A Gram stain and culture of the synovial fluid and radiographs may be needed in some patients to rule out these disorders. Gout can be accompanied by fever, particularly during polyarticular attacks, and should be considered in patients suspected of having acute bacterial arthritis whose cultures are negative.[26] In addition, gout and infection can coexist in the same joints, making therapeutic decisions difficult in individual cases. Thus, synovial fluid cultures are essential in any patient who is suspected of having gout and who has fever or purulent-appearing synovial fluid. Pseudogout [*see* Pseudogout (Calcium Pyrophosphate Dihydrate Deposition Disease), *below*] may cause acute monoarthritis or oligoarthritis that is similar to gout. Radiographs in such patients may show chondrocalcinosis, and CPPD crystals in the synovial fluid are usually easily distinguishable from urate crystals. However, some patients may have both gout and pseudogout in the same joint.

TREATMENT

The goals of therapy for patients with gout include termination of the acute attack, prevention of further attacks during the intercritical period, assessment for associated and contributing factors, and consideration of long-term hypouricemic therapy.[48,49] Each aspect of therapy should be considered separately, and there should be no confusion between efforts to suppress inflammation in acute attacks and efforts to lower serum urate levels, decrease the frequency of attacks, and prevent complications in the future.

Acute Gout

Treatment of acute gout should be initiated as early in the attack as possible. Agents available for terminating the acute attack include colchicine, nonsteroidal anti-inflammatory drugs (NSAIDs), adrenocorticotropic hormone (ACTH), and corticosteroids [*see Table 3*]. Each agent has a toxicity profile, with advantages and disadvantages applicable to individual circumstances. The patient's overall health and coexistent medical problems, particularly renal disease and gastrointestinal disease, often dictate the choice among these approaches. Corticosteroids and ACTH have been used more often in recent years in patients with multiple comorbid conditions because of the relatively low toxicity profile of these agents.

Colchicine has been used for centuries to treat acute attacks of gout. Given in oral dosages of 0.6 to 1.2 mg initially, followed by 0.6 mg every 2 hours, colchicine begins relieving most attacks of gout within 12 to 24 hours. However, most patients experience nausea, vomiting, abdominal cramps, and diarrhea with these dosages. Colchicine should be given more cautiously in elderly patients and should be avoided in patients with renal or hepatic insufficiency and patients already on long-term colchicine therapy.[50] Intravenous colchicine has been used for acute attacks, but recognition of the potential for bone marrow suppression and other systemic toxicities has resulted in guidelines for restricting dosage and even in a lack of availability in some countries.[51,52]

NSAIDs are useful in most patients with acute gout and remain the agents of choice for young, healthy patients without comorbid diseases. Indomethacin has been the most widely used agent over the years; it usually begins to provide relief within hours after the initial oral dose. Most NSAIDs are comparable in efficacy, although studies comparing NSAIDs in acute gout are few. The use of all NSAIDs is limited by the risks of gastric ulceration and gastritis, acute renal failure, fluid retention, interference with antihypertensive therapy, and, in older patients, problems with mentation. Aspirin is usually avoided because of its dose-related and variable effect on urate excretion. Newly developed NSAIDs with a high specificity for cyclooxygenase-2 are now available for the treatment of rheumatoid arthritis and osteoarthritis. These agents are generally much less toxic and ap-

Table 3 Drug Treatment of Acute Gout

Drug	*Dosage*	*Relative Efficacy*	*Comments*
Colchicine	0.6–1.2 mg p.o. initially, then 0.6 mg p.o., q. 2 hr	Moderate efficacy in high dose, but low threshold for toxicity	Common side effects are nausea, vomiting, cramps, and diarrhea; use with caution in elderly; avoid in patients with renal or hepatic insufficiency and those already on long-term colchicine Reasonably effective in low dose (q.d.–b.i.d.) as prophylaxis for future attacks
NSAIDs		Moderately effective; most are comparably effective	NSAIDs carry risk of gastric ulceration, gastritis, acute or chronic renal failure, fluid retention, congestive heart failure interference with antihypertensive therapy; risk of mentation problems in elderly COX-2–specific agents reduce risk of GI complications but not other complications
Indomethacin	50–75 mg b.i.d., t.i.d.		
Ibuprofen	600–800 mg t.i.d., q.i.d.		
Naproxen	500 mg b.i.d., t.i.d.		
Diclofenac	75 mg b.i.d		
Piroxicam	20 mg q.d.		
Meloxicam	15 mg q.d.		
Nabumetone	1,500 mg q.d.		
COX-2 inhibitors			
Rofecoxib	25–50 mg q.d.		
Celecoxib	200 mg q.d., b.i.d.		
Valdecoxib	20 mg q.d., b.i.d.		
Corticosteroids		Moderately to extremely effective	Caution in patients with brittle diabetes; some caution in patients with fluid retention or hypertension; overall, safer than NSAIDs or colchicine in these patients
Triamcinolone acetonide (and others)			
Intra-articular	5–10 mg for small joints; 40–60 mg for large joints		
Intramuscular	40–60 mg		
Prednisone (oral)	40–60 mg p.o. per day to start, taper over 5–14 days		
ACTH intramuscular	40–60 mg single dose		

ACTH—adrenocorticotropic hormone COX-2—cyclooxygenase-2 NSAIDs—nonsteroidal anti-inflammatory drugs

pear to have very low potential for gastrointestinal toxicity and inhibition of platelet function in clinical studies. These cyclooxygenase-2–specific NSAIDs should be useful in treating acute gout, and possibly in long-term prophylaxis, in patients at risk for gastrointestinal toxicity from the currently available NSAIDs.[53]

Corticosteroids have become more widely used in the treatment of acute gout in recent years.[54] Intra-articular steroids after arthrocentesis are extremely useful in providing relief, particularly in large effusions, in which the initial aspiration of fluid results in rapid relief of pain and tightness in the affected joint. The dosage of the steroid triamcinolone depends on the size of the joint, ranging from 5 to 10 mg for small joints of the hands or feet to 40 to 60 mg for larger joints, such as the knee. Systemic corticosteroids may also be useful in patients for whom colchicine or NSAIDs are inadvisable and in patients with polyarticular attacks. Oral prednisone, administered in tapered doses starting at 40 to 60 mg daily, and single intramuscular injections of ACTH (40 units) or triamcinolone (40 to 60 mg) have all been shown to be as effective as NSAIDs in treating acute gout. In most studies of systemic steroids for acute gout, only a small proportion of patients have required repeated therapy for rebound attacks in the first several days after therapy.

Interval Follow-up and Evaluation

Patients remain at increased risk for another attack of gout for several weeks after resolution of the initial attack; prophylaxis with small doses of colchicine or NSAIDs should be used for most patients. Colchicine (0.6 mg one or two times a day) prevents attacks in over 80% of patients. Prophylaxis should be continued for 1 to 2 months after an acute attack and for several months in patients with a history of frequent attacks; it should also be employed when urate-lowering drugs are initiated.[55] The dose of colchicine should be reduced or the duration of therapy shortened in patients with reduced renal function because bone marrow suppression and myoneuropathy have been reported in patients on long-term low-dose colchicine therapy with a creatinine clearance of less than 50 ml/min.[56]

After an acute attack of gout, a patient can be monitored for recurrent attacks and assessed for potential underlying causes of hyperuricemia. A spot midmorning urine sample to determine the ratio of urinary uric acid per deciliter of glomerular filtrate (urinary urate × plasma creatinine/urinary creatinine) or a 24-hour urine collection helps to classify the patient as an overproducer or underexcretor of urate; this is useful in identifying the optimal drug treatment for lowering serum urate levels, if indicated [*see Sidebar* Determining Overproducers and Underexcretors of Uric Acid].[1,57] If urinary urate excretion exceeds 0.6 mg/dl of glomerular filtrate or 600 to 700 mg a day, allopurinol is the most appropriate agent for lowering urate levels, by decreasing

Determining Overproducers and Underexcretors of Uric Acid

Urinary urate × plasma creatinine/urinary creatinine
- > 0.6 highly suggestive of overproducer
- < 0.6 suggestive of underexcretor

Urate present in 24-hour urine collection
- > 600–700 mg confirms overproducer status
- < 600 mg evidence of underexcretor if serum urate level is elevated

Table 4 Drug Treatment of Chronic Gout (Urate-Lowering Agents)

Drug	*Dosage*	*Relative Efficacy*	*Comments*
Probenecid	1–2 g/day in divided doses	Modest	Limited benefit to patients with decreased renal function; risk of renal stones
Sulfinpyrazone	Up to 400–800 mg/day in divided doses	Modest	Limited benefit to patients with decreased renal function; risk of renal stones
Allopurinol	300–600 mg/day; 200 mg for patients with GFR < 60 ml/min; 100 mg for patients with GFR < 30 ml/min	Extremely effective in lowering serum urate levels	Inhibits metabolism of azathioprine and other drugs; may cause hypersensitivity reaction (discontinue immediately if rash develops)

GFR—glomerular filtration rate

urate production; with lower excretion levels, a uricosuric drug may be useful.

Management after an acute attack should include a review of the patient's overall health, which may reveal important coexistent diseases, medications, and habits that could contribute to hyperuricemia. In particular, alcohol consumption should be discussed as an important factor in hyperuricemia and the precipitation of attacks. A review of the patient's diet may reveal heavy consumption of purine-rich foods, such as organ meats, seafood, or various legumes or other vegetables. In addition, the intercritical period is an excellent time to assess for obesity, hyperlipidemia, and hypertension, which often accompany gout and are correctable risk factors for premature cardiovascular mortality. Although the purine content of the diet usually contributes only about 1.0 mg/dl to the serum urate concentration, a diet that emphasizes calorie reduction, complex carbohydrates, and unsaturated fats will reduce urate levels and improve lipoprotein profiles in patients with gout.[48,58,59] Weight loss and physical activity are important strategies for lowering cardiovascular risk in obese patients; however, evidence suggests that increased physical activity is the perferred treatment[48,58,59]; sedentary, obese patients who have had an acute attack of gout should be strongly encouraged to increase their level of physical activity [*see CE:IV Diet and Exercise*].

Chronic Hyperuricemia

In general, patients with asymptomatic hyperuricemia should not be treated with hypouricemic agents. However, patients with persistent marked hyperuricemia (levels > 10 mg/dl) or hyperuricosuria (levels > 1,000 mg/24 hr) should be followed carefully for manifestations of gout or renal stones. Drug therapy to lower urate levels should be considered for patients who have had crystal-proven gout with recurrent attacks and persistent hyperuricemia despite efforts to identify and correct contributing factors [*see Table 4*]. Patients who have had more than two or three attacks, who have tophi or radiographic evidence of joint damage, or who have chronic renal insufficiency and recurrent gout should be treated with hypouricemic therapy if they are willing to comply with a long-term regimen. Reduction of serum urate levels to well into the normal range (i.e., < 6.0 mg/dl) eventually leads to prevention of further attacks and resorption of tophi.[42] Low-dose colchicine or NSAIDs should be used to prevent attacks that can occur for several months after hypouricemic therapy is started.[60]

Agents that increase renal excretion of urate (i.e., uricosuric drugs) can be used in patients with normal renal function who have no history of nephrolithiasis and whose 24-hour excretion of urate is less than 700 mg/day. Probenecid (1 to 2 g/day) is the most commonly used agent in this class, although sulfinpyrazone (up to 400 to 800 mg/day) can be used as well. Both agents are of limited use in patients with decreased renal function and carry a risk of precipitating renal stones. High urine volume and alkalinization with bicarbonate intake decrease this risk.

In up to 25% of patients, urate levels are not well controlled with uricosuric drug therapy. Benzbromarone is a uricosuric agent that has been available in Europe for over 20 years but is not available in the United States. Studies have shown that this agent may be useful in lowering uric acid levels in some patients with renal disease.[61]

Allopurinol, the only available inhibitor of xanthine oxidase, reduces serum urate levels in almost all compliant patients and may be used in overproducers or underexcretors. A daily dose of 300 mg is standard in patients with normal renal function, although some patients may require as much as 600 mg to achieve optimal serum urate levels. The dose should be reduced to 200 mg in patients whose glomerular filtration rate (GFR) is less than 60 ml/min and to 100 mg in those whose GFR is less than 30 ml/min. The dose of some other drugs, particularly azathioprine, must be reduced in patients receiving allopurinol, because allopurinol inhibits metabolism. Approximately 2% of patients taking allopurinol develop a hypersensitivity rash that progresses to a severe exfoliative dermatitis.[62] This disorder is more likely to occur in patients taking ampicillin or in those with renal insufficiency. Severe rashes may be accompanied by a syndrome of vasculitis, hepatitis, and interstitial renal disease, with a mortality risk of 20% reported in some series. Because of this risk, allopurinol should be discontinued in any patient who experiences a rash. Allopurinol may be reinstituted in such patients if the rash is mild. Reinitiation of allopurinol therapy should be gradual, starting with oral doses of 50 µg daily and increasing to 100 mg daily over a 4-week period; this strategy effectively desensitizes the majority of patients with prior hypersensitivity reactions to allopurinol therapy.[63]

Pseudogout (Calcium Pyrophosphate Dihydrate Deposition Disease)

DEFINITION AND CLASSIFICATION

CPPD crystals may be found in deposits in and around joints and are characterized by calcification of articular cartilage, menisci, synovium, and other periarticular tissues. McCarty and colleagues first described CPPD crystals in synovial fluids from patients with goutlike attacks in 1962.[2] They used the term pseudogout for this new arthropathy, which is characterized by intra-articular calcifications (chondrocalcinosis), crystals in the synovial fluid, and acute arthropathy. Since then, other clinical presentations and a variety of disease processes have been asso-

ciated with CPPD crystals. Thus, the term CPPD deposition disease has come to include the various clinical presentations as part of the same general clinical syndrome.

EPIDEMIOLOGY

CPPD deposition disease is generally a disease of the elderly; the average age of patients is approximately 70 years.[64] The prevalence of articular chondrocalcinosis is very low in people younger than 40 years but increases with age and is quite common in older populations. When multiple radiologic studies are obtained, the documented prevalence in the general population is 10% to 15% in those 65 to 75 years of age and over 40% in people older than 80 years.[65,66] CPPD deposition occurs in males and females in differing distribution in different studies, but there does not seem to be a major gender predominance. An increased prevalence of CPPD deposition in certain diseases and familial groups has been reported.

PATHOGENESIS AND ETIOLOGY

The metabolic basis for CPPD formation and deposition is less well understood than that for urate crystals. CPPD crystal formation occurs almost exclusively in the articular and periarticular tissue, most often near the surface of chondrocytes.[64,67] Crystal formation is enhanced by elevated levels of either calcium or inorganic pyrophosphate (PP_i) within local tissues or local factors in the cartilage matrix that promote crystal formation. An abnormal substrate of matrix collagen and proteoglycan, as well as variations in mineral content, may promote crystal deposition. Local elevations of PP_i levels appear to be related to two factors: the overactivity of a cell surface enzyme (ectoenzyme) known as nucleoside triphosphate pyrophosphohydrolase (NTPPH), which catalyzes the extracellular hydrolysis of ATP, and the extracellular transport of PP_i by the transmembrane protein ANK.[68] In addition, some of the excess PP_i production may take place intracellularly through NTPPH or as a by-product of cellular proteoglycan and protein synthesis. Other factors that may contribute to excess PP_i and crystal formation include decreased activity of pyrophosphatase, degenerating cellular debris, abnormal matrix collagen, and even the local influence of growth factors (i.e., transforming growth factor and insulinlike growth factor). The mechanisms by which CPPD crystals induce inflammation are believed to be similar to those observed in gout.

CLINICAL VARIANTS

Most joints with radiographically observed chondrocalcinosis are asymptomatic, although subtle articular symptoms are more common in asymptomatic patients with chondrocalcinosis than in patients without these findings.[69] Clinically symptomatic CPPD deposition disease may take any of several forms that tend to present in acute or chronic fashion, mimicking other arthropathies.

Acute Pseudogout

Acute pseudogout is slightly more common in males than in females. Attacks of this form of CPPD deposition disease are usually acute, increase in intensity over 12 to 36 hours, and last for a few days to a few weeks. Most acute attacks of pseudogout are less intense than attacks of gout. The most commonly involved joint is the knee (seen in over half of patients), followed by the wrist and ankle. In rare cases, attacks in the first metatarsophalangeal joint may be seen.[64] Affected joints previously involved are more likely to be involved in subsequent attacks. Attacks may occur in clusters over short periods, and polyarticular attacks occur in a few patients. A moderate synovial fluid leukocytosis is common, and marked elevations in the white blood cell count, mimicking a septic joint, may be seen in some patients. Mild fever and leukocytosis have been described, but not as frequently as in gout.[70] Between attacks, the joint is usually asymptomatic unless there is coexistent osteoarthritis. As in gout, attacks of pseudogout seem to be precipitated in some patients by stressful events, such as surgery, trauma, and acute medical illness. The intra-articular injection of hyaluronate for the symptomatic treatment of osteoarthritis has also been reported to trigger attacks of acute pseudogout.[71]

Table 5 Conditions Associated with Calcium Pyrophosphate Dihydrate Deposition Disease*

Definite association	Possible or doubtful association
Hemochromatosis	Gout
Hyperparathyroidism	Familial hypocalciuric hypocalcemia
Hypophosphatasia	Acromegaly
Hypomagnesemia	X-linked hypophosphatemic rickets
Probable association	Neuropathic joints
Hypothyroidism	Amyloidosis
	Trauma

See reference 62.
*As pseudogout or radiographic chondrocalcinosis.

Chronic Rheumatoid-like Arthritis

About 5% to 10% of patients with CPPD deposition disease experience a polyarticular process resembling rheumatoid arthritis. The disease is indolent, symmetrical in distribution, and characterized by inflammation. Chronic swelling, morning stiffness, and predominant wrist and knee involvement are seen in this group of patients. Because of the relatively high frequency of incidental chondrocalcinosis and positive rheumatoid factors in elderly persons, differentiating CPPD deposition from rheumatoid arthritis can be difficult. A history of acute exacerbations in these patients may help suggest CPPD deposition, whereas the presence of subcutaneous nodules and very high titers of rheumatoid factor favor a diagnosis of rheumatoid arthritis.

Osteoarthritis and CPPD Deposition

About half of patients with CPPD deposition have a chronic degenerative arthritis involving multiple joints, usually in a symmetrical pattern. Women predominate in this group of patients.[64] The knees are most commonly involved, followed by the wrist, metacarpophalangeal joints, hips, spine, and shoulders. CPPD-associated osteoarthritis may be differentiated from typical osteoarthritis by the presence of changes in atypical joints, such as the wrists, elbows, and shoulders. Some patients in this group may not have chondrocalcinosis, but CPPD crystals may be found in the synovial fluid of most of those patients without radiographic findings. In addition, radiographic features of predominantly patellofemoral involvement and femoral cortical erosions in the knee suggest CPPD deposition.[72]

Conditions Associated with CPPD Deposition Disease

Most cases of CPPD deposition disease are sporadic. However, a number of kindreds with familial forms of disease and asso-

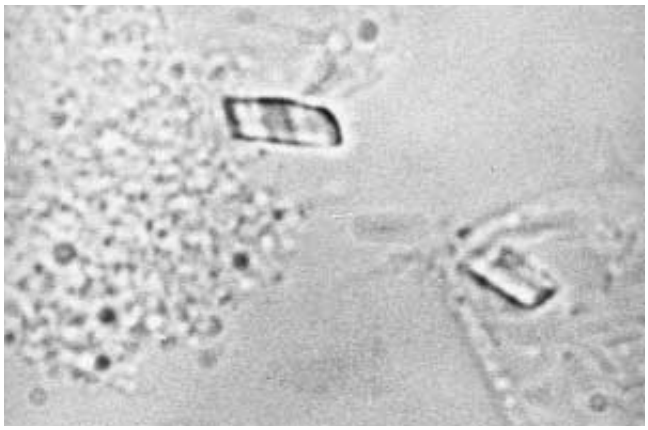


Figure 4 **Calcium pyrophosphate dihydrate crystals typical of pseudogout are rhomboid and demonstrate a weakly positive birefringence under polarized light.**

ciations with metabolic diseases have been reported. Most of the familial forms have shown an autosomal dominant transmission but have displayed a variety of clinical presentations, and most appear to involve single gene mutations.[73,74] Associations with several endocrine and metabolic conditions have been reported, many of which probably represent no more than a chance occurrence of common age-related conditions [*see Table 5*].

Definite associations exist between CPPD deposition disease and hemochromatosis, hyperparathyroidism, hypophosphatasia, and hypomagnesemia. A distinct form of arthritis associated with hemochromatosis was first reported in 1964.[75] This arthropathy is similar to osteoarthritis and rheumatoid arthritis and may be the initial presenting feature in some patients. The most frequently involved joints are the metacarpophalangeal joints (primarily the second and third), wrists, and hips; radiologic changes consisting of hooklike osteophytes at the metacarpal heads are a characteristic finding. CPPD deposition has been described in 20% to 30% of patients with primary hyperparathyroidism, more often in older patients.

Attacks of acute pseudogout after parathyroidectomy have been described and are often the first manifestation of CPPD deposition in these patients.[76] Reports have described pseudogout in patients with hypomagnesemia after liver transplantation, possibly related to tacrolimus therapy.[77]

DIAGNOSIS

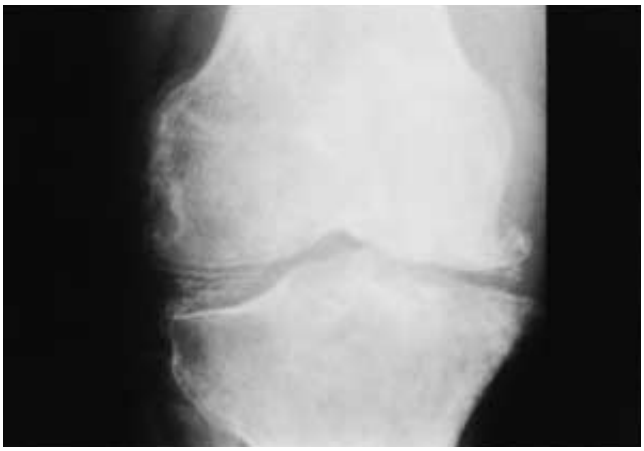
Synovial fluid aspiration and examination for crystals are essential to the diagnosis. The synovial fluid in pseudogout is usually inflammatory but may be hemorrhagic. A leukocyte count of about 10,000 to 20,000 cells/mm^3 is the rule, but in the small joints, such as the wrist, very high counts may be seen. Synovial fluid should be examined first under regular microscopy, because CPPD crystals are weakly birefringent under polarized microscopy. The crystals are rhomboid or rod-shaped and may be intracellular or extracellular [*see Figure 4*]. Because of their weak birefringence, CPPD crystals may be missed on initial examination, so it is essential that someone experienced in crystal identification examine the fluid.

Radiographic studies of affected joints often reveal chondrocalcinosis of the articular cartilage. The fibrocartilage of the menisci in the knees [*see Figure 5a*] or of the triangular ligament at the radioulnar joint at the wrist [*see Figure 5b*] may have punctate or linear calcifications; similar changes may be seen in the symphysis pubis, shoulder, hip, and intervertebral disks. Linear calcification of the hyaline cartilage in these joints may be seen as well. Other features may include narrowing and sclerosis of the radiocarpal and patellofemoral joints, femoral cortical erosions above the knee, and extra-articular calcifications involving tendons or ligaments.

TREATMENT

Management of the patient with pseudogout is similar to management of the patient with acute gout, with the main goal of therapy being control of the acute inflammatory reaction.[49] Rest of the inflamed joint (or joints) and administration of NSAIDs or intra-articular corticosteroid preparations are the mainstay of therapy. Aspiration of the joint is sufficient to significantly relieve pain and discomfort in some patients. Colchicine is effective in patients with acute pseudogout but should be used cautiously in older patients. At lower doses of 0.6 mg one or two times a day, colchicine can be helpful in preventing further attacks.[78] In some patients, intramuscular or subcutaneous ACTH (40 units) or intramuscular triamcinolone (60 mg) can control the acute inflammatory reaction.[79] For those with chronic pain and inflammation,

a



b

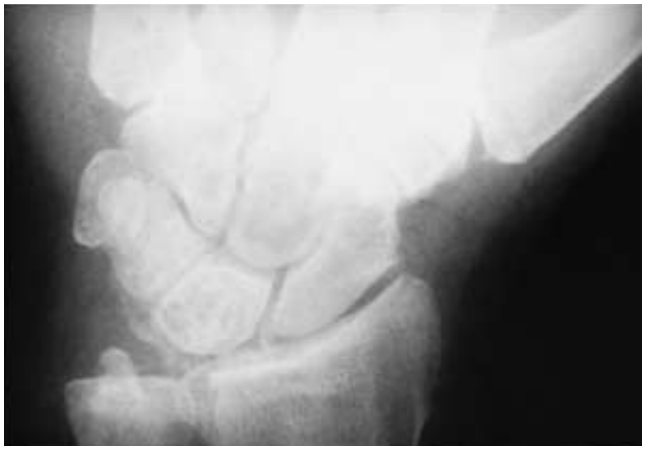


Figure 5 **Typical radiographs of chondrocalcinosis seen in CPPD deposition disease, with evidence of intra-articular calcification in the meniscus and hyaline cartilage of the knee (*a*) and the triangular cartilage of the wrist (*b*).**

alternatives for management are physiotherapy, analgesics, and NSAIDs. An evaluation of the patient for underlying metabolic abnormalities, particularly hemochromatosis and hyperparathyroidism, should be considered. However, successful treatment of these associated conditions has not been shown to alter the radiographic or clinical course of CPPD deposition disease.

Other Forms of Crystal-Associated Arthritis

BASIC CALCIUM PHOSPHATE DEPOSITION

A group of apatitelike (basic calcium phosphate) crystals has been identified in pathologic synovial fluids and articular and periarticular tissues in a variety of musculoskeletal disorders. These crystals may be found in 30% to 60% of synovial fluids from patients with osteoarthritis and may contribute to the low-grade inflammatory process and cartilage destruction seen in typical osteoarthritis.[80,81] A severe destructive arthropathy of the shoulder and knee, known as the Milwaukee shoulder-knee syndrome, predominantly affects older women.[82] This process is associated with rotator cuff degeneration and rupture, joint instability, and periarticular calcification. The synovial fluid may be serosanguineous and contains few cells; hydroxyapatite crystals may appear as clumps or may look like intracellular shiny coins, but they are not birefringent under polarized-light microscopy. The treatment of this condition is difficult, but joint aspiration and intra-articular corticosteroid injections have been helpful in some patients.

Basic calcium phosphate crystals are also associated with acute calcific periarthritis that may affect the shoulder or other joint areas in periarticular structures. In patients with this condition, periarticular calcific deposits may be found in the shoulder; near the lateral trochanter of the hip; around the wrists, fingers, or knees; or in the ankle and foot. These deposits may be well defined radiographically at the beginning of attacks but often disappear over several weeks. NSAIDs and local corticosteroid injections are usually useful in the treatment of this condition.

OTHER CRYSTALS FOUND IN SYNOVIAL FLUID

A variety of other crystals in synovial fluid have been described. Cholesterol crystals may be seen in some chronic effusions and are most often associated with chronic rheumatoid bursal effusions.[83] Other lipid crystals have been seen after joint trauma, and another type of lipid crystal, which resembles a Maltese cross, has been described and may be responsible for an acute inflammatory reaction in rare cases. In addition, calcium oxalate crystals have been found in the synovial fluid of patients with end-stage renal disease and in soft tissues of patients with primary hyperoxaluria and may cause articular symptoms.[84] The pathogenetic significance of each of these types of crystals is uncertain; they most probably represent incidental secondary phenomena.

The author has received grants for clinical research from Amgen Inc., Abbott Laboratories Inc., Isis Pharmaceuticals, Inc., and Bristol-Myers Squibb Company and has served as a consultant for TAP Pharmaceutical Products Inc.

References

1. McCarty DJ, Hollander JL: Identification of urate crystals in gouty synovial fluids. Ann Intern Med 54:452, 1961

2. McCarty DJ, Kohn NN, Faires JS: The significance of calcium phosphate crystals in the synovial fluid of arthritis patients: the "pseudogout syndrome": I. Clinical aspects. Ann Intern Med 56:711, 1962

3. Abbott RD, Brand FN, Kannel WB, et al: Gout and coronary artery disease: the Framingham Study. J Clin Epidemiol 41:237, 1988

4. Arromdee E, Michet CJ, Crowson CS: Epidemiology of gout: is the incidence rising? J Rheumatol 29:2403, 2002

5. Lawrence RC, Helmick CG, Arnett FC, et al: Estimates of the prevalence of arthritis and selected musculoskeletal diseases in the United States. Arthritis Rheum 41:778, 1998

6. Hochberg MC, Thomas J, Thomas DJ, et al: Racial differences in the incidence of gout. Arthritis Rheum 38:628, 1995

7. Ter Borg EJ, Rasker JJ: Gout in the elderly: a separate entity? Ann Rheum Dis 46:72, 1987

8. Wortmann RL, Kelley WN: Gout and hyperuricemia. Textbook of Rheumatology, 6th ed. Ruddy S, Harris ED, Sledge CB, Eds. WB Saunders Co, Philadelphia, 2001, p 1339

9. Campion EW, Glynn RJ, DeLabry LO: Asymptomatic hyperuricemia: risks and consequences in the Normative Aging Study. Am J Med 82:421, 1987

10. Agudelo CA, Wise CM: Gout: diagnosis, pathogenesis, and clinical manifestations. Curr Opin Rheumatol 13:234, 2001

11. Wortmann RL: Gout and hyperuricemia. Curr Opin Rheumatol 14:281, 2002

12. Wilk JB, Djousse L, Borecki I, et al: Segregation analysis of serum uric acid in the NHLBI Family Heart Study. Hum Genet 106:355, 2000

13. Perez-Ruiz F, Calabozo M, Erauskin GG, et al: Renal underexcretion of uric acid is present in patients with apparent high urinary uric acid output. Arthritis Rheum 47:610, 2002

14. Shadick NA, Kim R, Weiss S, et al: Effect of low level lead exposure on hyperuricemia and gout among middle aged and elderly men: the normative aging study. J Rheumatol 27:1708, 2000

15. Lin JL, Tan DT, Ho HH, et al: Environmental lead exposure and urate excretion in the general population. Am J Med 113:563, 2002

16. Scott JT, Higgens CS: Diuretic induced gout: a multifactorial condition. Ann Rheum Dis 51:259, 1992

17. Caspi D, Lubart E, Graff E, et al: The effect of mini-dose aspirin on renal function and uric acid handling in elderly patients. Arthritis Rheum 43:103, 2000

18. Lin HY, Rocher LL, McQuillan MA, et al: Cyclosporine-induced hyperuricemia and gout. N Engl J Med 321:287, 1989

19. Burack DA, Griffith BP, Thompson ME, et al: Hyperuricemia and gout among heart transplant recipients receiving cyclosporine. Am J Med 92:141, 1992

20. Clive DM: Renal transplant-associated hyperuricemia and gout. J Am Soc Nephrol 11:974, 2000

21. Schiltz C, Liote F, Prudhommeaux F, et al: Monosodium urate monohydrate crystal-induced inflammation in vivo: quantitative histomorphometric analysis of cellular events. Arthritis Rheum 46:1643, 2002

22. Yagnik DR, Hillyer P, Marshall D, et al: Noninflammatory phagocytosis of monosodium urate monohydrate crystals by mouse macrophages: implications for the control of joint inflammation in gout. Arthritis Rheum 43:1779, 2000

23. Landis RC, Yagnik DR, Florey O, et al: Safe disposal of inflammatory monosodium urate monohydrate crystals by differentiated macrophages. Arthritis Rheum 46:3026, 2002

24. Fam AG, Stein J, Rubenstein J: Gouty arthritis in nodal osteoarthritis. J Rheumatol 23:684, 1996

25. Agudelo CA, Wise CM: Crystal-associated arthritis. Clin Geriatr Med 14:495, 1998

26. Hadler NM, Franck WA, Bress NM, et al: Acute polyarticular gout. Am J Med 56:715, 1974

27. Gutman AB: Gout. Textbook of Medicine. Beeson PB, McDermott W, Eds. WB Saunders Co, Philadelphia, 1963, p 1255

28. Gutman AB: The past four decades of progress in the knowledge of gout with an assessment of the present status. Arthritis Rheum 16:431, 1973

29. Barthelemy CR, Nakayama DA, Carrera GF, et al: Gouty arthritis: a prospective radiographic evaluation of sixty patients. Skeletal Radiol 11:1, 1984

30. Yü T-F, Berger L: Renal disease in primary gout: a study of 253 gout patients with proteinuria. Semin Arthritis Rheum 4:293, 1975

31. Yü T-F, Berger L: Impaired renal function in gout: its association with hypertensive vascular disease and intrinsic renal disease. Am J Med 72:95, 1982

32. Johnson RJ, Kang DH, Feig D, et al: Is there a pathogenetic role for uric acid in hypertension and cardiovascular disease and renal disease? Hypertension 41:1183, 2003

33. Tarng D-C, Lin H-Y, Shyong M-L, et al: Renal function in gout patients. Am J Nephrol 15:31, 1995

34. Perez-Ruiz F, Calabozo M, Herrero-Beites AM, et al: Improvement of renal function in patients with chronic gout after proper control of hyperuricemia and gouty bouts. Nephron 86:287, 2000

35. Kramer HM, Curhan G: The association between gout and nephrolithiasis: the National Health and Nutrition Examination Survery III, 1988–1994. Am J Kidney Dis 40:37, 2002

36. Pak CYC, Barilla DE, Holt K, et al: Effect of oral purine load and allopurinol on the crystallization of calcium salts in urine of patients with hyperuricosuric calcium urolithiasis. Am J Med 65:593, 1978

37. Cohen LF, Balow JE, Magrath IT, et al: Acute tumor lysis syndrome: a review of 37 patients with Burkitt's lymphoma. Am J Med 68:486, 1980

38. Reynolds PP, Knapp MJ, Baraf HSB, et al: Moonshine and lead: relationship to the pathogenesis of hyperuricemia in gout. Arthritis Rheum 26:1057, 1983

39. Takahashi S, Yamamoto T, Moriwaki Y, et al: Impaired lipoprotein metabolism in

patients with primary gout: influence of alcohol intake and body weight. Br J Rheumatol 33:731, 1994

40. Bo S, Cavallo-Perin P, Gentile L, et al: Hypouricemia and hyperuricemia in type 2 diabetes: two different phenotypes. Eur J Clin Invest 31:318, 2001

41. Kahn HS, Valdez R: Metabolic risks identified by the combination of enlarged waist and elevated triacylglycerol concentration. Am J Clin Nutr 78:928, 2003

42. Culleton BF, Larson MG, Kannel WB, et al: Serum uric acid and risk for cardiovascular disease and death: the Framingham Heart Study. Ann Intern Med 131:7, 1999

43. Erickson AR, Enzenauer RJ, Nordstrom DM, et al: The prevalence of hypothyroidism in gout. Am J Med 97:231, 1994

44. Wallace SL, Robinson H, Masi AT, et al: Preliminary criteria for the classification of acute arthritis of primary gout. Arthritis Rheum 20:895, 1977

45. Agudelo CA, Weinberger A, Schumacher HR, et al: Definitive diagnosis of gout by identification of urate crystals in asymptomatic metatarsophalangeal joints. Arthritis Rheum 22:559, 1979

46. Pascual E, Batlle E, Martinez A, et al: Synovial fluid analysis for diagnosis of intercritical gout. Ann Intern Med 131:756, 1999

47. Pascual E: Gout update: from lab to the clinic and back. Curr Opin Rheumatol 12:213, 2000

48. Terkeltaub RA: Clinical practice: gout. N Engl J Med 349:1647, 2003

49. Agudelo CA, Wise CM: Crystal deposition diseases. Treatment of the Rheumatic Diseases, 2nd ed. Weisman MH, Weinblatt ME, Louie J, Eds. WB Saunders Co, Philadelphia, 2001, 447

50. Roberts WN, Liang MH, Stern SH: Colchicine in acute gout: reassessment of risks and benefits. JAMA 257:1920, 1987

51. Wallace SL, Singer JZ: Systemic toxicity associated with the intravenous administration of colchicine: guidelines for use (review). J Rheumatol 15:495, 1988

52. Bonnel RA, Villalba ML, Karwoski CB, et al: Deaths associated with inappropriate intravenous colchicine administration. J Emerg Med 22:385, 2002

53. Fam AG: Treating acute gouty arthritis with selective COX 2 inhibitors. BMJ 325:980, 2002

54. Fam AG: Current therapy of acute microcrystalline arthritis and the role of corticosteroids. J Clin Rheumatol 3:35, 1997

55. Yü TF: The efficacy of colchicine prophylaxis in articular gout: a reappraisal after 20 years. Semin Arthritis Rheum 12:256, 1982

56. Wallace SL, Singer JZ, Duncan GJ, et al: Renal function predicts colchicine toxicity: guidelines for the prophylactic use of colchicine in gout. J Rheumatol 18:264, 1991

57. Simkin PA: When, why, and how should we quantify the excretion rate of urinary uric acid? J Rheumatol 28:1207, 2001

58. Dessein PH, Shipton EA, Stanwix AE, et al: Beneficial effects of weight loss associated with moderate calorie/carbohydrate restriction, and increased proportional intake of protein and unsaturated fat on serum urate and lipoprotein levels in gout: a pilot study. Ann Rheum Dis 59:539, 2000

59. Snaith ML: Gout: diet and uric acid revisited. Lancet 358:525, 2001

60. Bull PW, Scott JT: Intermittent control of hyperuricemia in the treatment of gout. J Rheumatol 16:1246, 1989

61. Perez-Ruiz F, Calabozo M, Fernandez-Lopez MJ, et al: Treatment of chronic gout in patients with renal function impairment: an open, randomized, actively controlled study. J Clin Rheumatol 5:49, 1999

62. Rosenthal AK, Ryan LM: Calcium pyrophosphate crystal deposition disease, pseudogout, and articular chondrocalcinosis. Arthritis and Allied Conditions, 14th ed. Koopman WJ, Ed. Williams & Wilkins, Baltimore, 2001, p 2348

63. Fam AG, Dunne SM, Iazzetta J, et al: Efficacy and safety of desensitization to allopurinol following cutaneous reactions. Arthritis Rheum 44:231, 2001

64. Hande KR, Noone RM, Stone WJ: Severe allopurinol toxicity: description and guidelines for prevention in patients with renal insufficiency. Am J Med 76:47, 1984

65. Wilkins E, Dieppe P, Maddison P, et al: Osteoarthritis and articular chondrocalcinosis in the elderly. Ann Rheum Dis 42:280, 1983

66. Doherty M, Dieppe P: Crystal deposition disease in the elderly. Clin Rheum Dis 12:97, 1986

67. Reginato AJ, Reginato AM: Diseases associated with deposition of calcium pyrophosphate or hydroxyapatite. Textbook of Rheumatology, 6th ed. Ruddy S, Harris ED, Sledge CB, Eds. WB Saunders Co, Philadelphia, 2001, p 1377

68. Hirose J, Ryan LM, Masuda I: Up-regulated expression of cartilage intermediate-layer protein and ANK in articular hyaline cartilage from patients with calcium pyrophosphate dihydrate crystal deposition disease. Arthritis Rheum 46:3218, 2002

69. Ellman MH, Levin B: Chondrocalcinosis in elderly persons. Arthritis Rheum 18:43, 1975

70. Bong D, Bennett R: Pseudogout mimicking systemic disease. JAMA 246:1438, 1981

71. Fam AG: What is new about crystals other than monosodium urate? Curr Opin Rheumatol 12:228, 2000

72. Resnick D, Williams G, Weisman MH, et al: Rheumatoid arthritis and pseudorheumatoid arthritis in calcium pyrophosphate dihydrate crystal deposition disease. Radiology 140:615, 1981

73. Jones AC, Chuck AJ, Arie EA, et al: Diseases associated with calcium pyrophosphate deposition disease. Semin Arthritis Rheum 22:188, 1992

74. Timms AE, Zhang Y, Russell RG, et al: Genetic studies of disorders of calcium crystal deposition. Rheumatology (Oxford) 41:725, 2002

75. Schumacher HR Jr: Hemochromatosis and arthritis. Arthritis Rheum 7:41, 1964

76. Rynes RI, Merzig EG: Calcium pyrophosphate crystal deposition disease and hyperparathyroidism: a controlled, prospective study. J Rheumatol 5:460, 1978

77. Perez-Ruiz F, Testillano M, Gastaca MA, et al: "Pseudoseptic" pseudogout associated with hypomagnesemia in liver transplant patients. Transplantation 71:696, 2001

78. Alvarellos A, Spilberg I: Colchicine prophylaxis in pseudogout. J Rheumatol 13:804, 1986

79. Roane DW, Harris MD, Carpenter MT, et al: Prospective use of intramuscular triamcinolone acetonide in pseudogout. J Rheumatol 24:1168, 1997

80. Jaovisidha K, Rosenthal AK: Calcium crystals in osteoarthritis. Curr Opin Rheumatol 14:298, 2002

81. Morgan MP, McCarthy GM: Signaling mechanisms involved in crystal-induced tissue damage. Curr Opin Rheumatol 14:292, 2002

82. Halverson PB, Carrera GF, McCarty DJ: Milwaukee shoulder syndrome: fifteen additional cases and a description of contributing factors. Arch Intern Med 150:677, 1990

83. Wise CM, White RE, Agudelo CA: Synovial fluid lipid abnormalities in various disease states: review and classification. Semin Arthritis Rheum 16:222, 1987

84. Reginato AJ: Calcium oxalate and other crystals of particles associated with arthritis. Arthritis and Allied Conditions. Koopman WJ, Ed. Williams and Wilkins, Baltimore, 2000, p 2393

X OSTEOARTHRITIS

Christopher Wise, m.d.

Definition

Osteoarthritis is a common form of arthritis characterized by degeneration of articular cartilage and reactive changes in surrounding bone and periarticular tissue. The disease process results in pain and dysfunction of affected joints and is a major cause of disability in the general population. Osteoarthritis is also frequently referred to as degenerative joint disease; other terms that have been used include osteoarthrosis, hypertrophic arthritis, and atrophic arthritis.

Classification

primary osteoarthritis

Patients without a specific inflammatory or metabolic condition known to be associated with arthritis and without a history of specific injury or trauma are considered to have primary osteoarthritis. However, a number of underlying processes are considered to be important in patients with primary osteoarthritis [*see* Etiologic Factors, Risk Factors, *below*]. In most patients, involvement is limited to one or a small number of joints or joint areas. In some patients, however, multiple joint areas are involved, and these patients are considered to have a separate variant called primary generalized osteoarthritis. Another variant, termed erosive osteoarthritis, is characterized by polyarticular involvement of the small joints of the hand and tends to occur more often in middle-aged and elderly women.

secondary osteoarthritis

Secondary osteoarthritis has been associated with several conditions that cause damage to articular cartilage through a variety of mechanisms, including mechanical, inflammatory, and metabolic processes [*see Table 1*]. Acute trauma, particularly intra-articular fractures and meniscal tears, can result in articular instability or incongruity and can lead to osteoarthritis years after an injury.

The role of chronic trauma from certain occupational or avocational activities is not as well established as the role of acute trauma in the development of secondary osteoarthritis. Neurologic disorders that result in the loss of sensory nerve function may be associated with a particularly destructive type of degenerative arthritis (i.e., neuropathic arthritis and Charcot joint) in which cartilage and bone fragmentation are seen with relatively little pain.

Many types of inflammatory arthritis can cause destruction of articular cartilage. The best example of cartilage damage is seen in chronic rheumatoid arthritis, but similar cartilage damage can be seen in postinfectious arthritis, psoriatic arthritis, reactive arthritis, and ankylosing spondylitis.

Congenital and developmental diseases that cause joint incongruity may result in osteoarthritis. This condition is best recognized in epiphyseal dysplasia, Perthes disease, and other processes affecting the femoral head and hip and has also been associated with generalized joint hypermobility, as seen in Ehlers-Danlos syndrome.

Primary bone disorders that affect the mechanics and articular surfaces of nearby joints may also lead to degenerative cartilage changes, particularly around major joints such as the shoulder, hip, and knee. Several metabolic and endocrine disorders have been associated directly or indirectly with the development of osteoarthritis, often with atypical patterns or in unusual locations. In most of these conditions, cartilage damage is associated with the accumulation, in articular cartilage, of a particular substance associated with the metabolic condition (e.g., uric acid or iron). In hemochromatosis, the mechanism of joint damage may also be related to an association with calcium pyrophosphate crystal deposition. In acromegaly, overgrowth of articular cartilage and subsequent mechanical problems appear to be important in the pathogenesis of the disease.

Epidemiology

Osteoarthritis is the most common type of arthritis, and it is one of the most common causes of disability and dependence in the United States.[1,2] Estimating the prevalence of osteoarthritis in the general population is difficult because of the high prevalence of asymptomatic radiographic changes of osteoarthritis and differences in case definition. The prevalence of radiographic changes of osteoarthritis in the population in general, regardless of symptoms, is roughly 30% for the hands, 21% for the feet, and

Table 1 Causes of Secondary Osteoarthritis

Trauma
Acute injury
Chronic occupational overuse
Sports overuse
Neuropathic arthropathy (Charcot joint)
Inflammatory arthritis
Rheumatoid arthritis
Infectious arthritis
Psoriatic arthritis
Reactive arthritis
Ankylosing spondylitis
Dysplastic and hereditary conditions
Congenital hip dysplasia
Epiphyseal dysplasia
Chondrodysplasias
Perthes disease
Kashin-Bek disease
Joint hypermobility
Bone disorders
Osteonecrosis (avascular necrosis)
Osteochondritis
Paget disease of bone
Metabolic and endocrine disorders
Crystal deposition disease (gout, calcium pyrophosphate deposition, basic calcium phosphate)
Hemochromatosis
Ochronosis
Wilson disease
Bleeding disorders
Acromegaly

3% for the knees and hips. In persons older than 65 years, changes are seen in the knee in 33% and in the hands in almost 100%. Fortunately, most patients with radiographic changes found in population-based surveys have few symptoms or functional limitations. Men and women 30 to 60 years of age have equal overall prevalence of symptomatic osteoarthritis (approximately 6% have affected knees and 4% have affected hips). For adults older than 60 years, however, the prevalence of symptomatic osteoarthritis (all joints) increases to 17% in men and 30% in women.[1,2]

Men and women tend be affected equally by osteoarthritis in middle age, but after 50 years of age, women are affected more often, particularly in the interphalangeal joints of the fingers.[3] Osteoarthritis is seen in all population groups, although prevalence can vary with certain geographic areas and ethnic groups. For example, osteoarthritis of the hip is least common in Japanese, Saudi Arabian, Chinese, and African populations; and knee involvement is most common in African-American women. Comparisons of osteoarthritis prevalence have shown that hip involvement is less common, but knee involvement is more common, in Chinese men and women than in white men and women in the United States.[4-6]

Etiologic Factors

RISK FACTORS

A number of risk factors are believed to contribute to the development of primary osteoarthritis, including age, obesity, joint malalignment, bone density, hormonal status, nutritional factors, joint dysplasia, trauma, occupational factors, and hereditary factors.[2,7]

Age is the factor most strongly associated with radiographic and clinically significant osteoarthritis, with an exponential increase seen in more severely involved joints. The cellular or biomechanical changes in articular cartilage that occur with aging are not necessarily those seen in osteoarthritis. However, it has been speculated that these changes may facilitate the development of disease.

Obesity is clearly associated with osteoarthritis of the knee. The increased load carried by obese persons and the alterations in gait and posture that redistribute the load contribute to cartilage damage. A study in young men suggested that each increase in weight of 8 kg results in a 70% increase in the risk of symptomatic arthritis of the knee in later years.[8] This association is particularly high in patients with varus malalignment of the knee, and obese patients with malalignment are at risk for more rapid progression of established osteoarthritis in the knee.[9,10] Most of the association of obesity with osteoarthritis of the knee appears to be related to environmental, rather than genetic, factors.[11] The relation of obesity to osteoarthritis in other weight-bearing joints is not as clear-cut and may not be much of a factor at all for hip involvement.

An association between increased bone density and osteoarthritis has been noted in several studies.[12,13] Women with osteoporosis and hip fractures have a decreased risk of osteoarthritis, and those affected by osteoarthritis have significantly increased bone density. This negative association suggests that soft subchondral bone absorbs impact and protects articular cartilage better than dense bone. Paradoxically, however, estrogen deficiency may contribute to the increased prevalence of osteoarthritis in women who have recently entered menopause.[2] In addition, a study showing that patients with low dietary vitamin D intake have more rapid progression of disease suggests that strong subchondral bone may be particularly important in preventing progression of osteoarthritis once it is established.[14]

Chronic repetitive impact loading is known to cause rapid degenerative changes in articular cartilage in laboratory animals. This mechanism probably accounts for the high frequency of osteoarthritis in certain occupational and athletic settings. In particular, occupational activities that require frequent knee bending increase the risk of knee involvement, and frequent lifting appears to be a risk factor for hip involvement.[15,16] Long-term weight-bearing sports activity is associated with an increased risk of developing radiographic evidence of osteoarthritis. In patients without a history of injury, clinical symptoms do not always correlate well with radiographic changes, and radiographic changes do not often progress significantly, even in older long-distance runners.[17,18] However, a history of specific joint injury, usually related to sports and recreational activities, is an important risk factor for knee and hip disease.[19-21]

Decreased strength and proprioception have been demonstrated in patients with osteoarthritis and likely play a role in pathogenesis of the disease. Patients with radiographic changes of osteoarthritis and no pain have decreased muscle strength in the affected leg, and decreased proprioception has been demonstrated in unaffected knees of patients with unilateral disease.[22,23] In addition, local injection to relieve pain will only partially improve muscle activity and proprioceptive and gait defects.[24] The importance of muscle strength and proprioception to the health of normal cartilage has been further suggested in studies demonstrating cartilage thinning after spinal cord injury.[25,26]

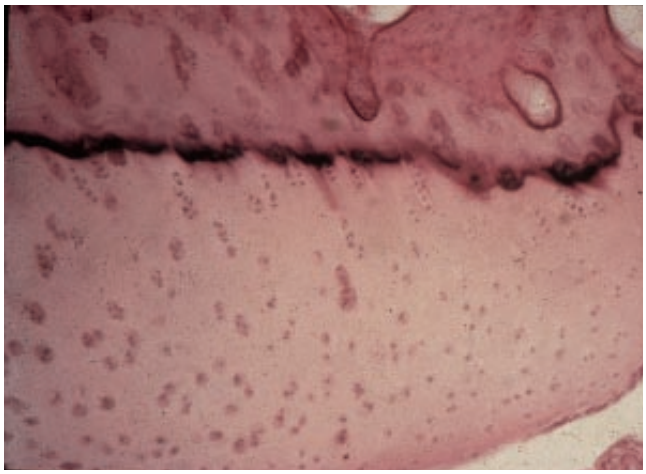
Many patients with osteoarthritis have a family history of the disorder, and multiple genetic factors may be responsible in various forms of osteoarthritis.[27] Osteoarthritis with finger joint involvement in women is probably the best recognized form of arthritis with familial associations,[28] but hereditary factors are also important in osteoarthritis of the hip.[29,30] Metabolic abnormalities related to the hereditary component of osteoarthritis have been found in a number of studies. These abnormalities include associations between variations of collagen genes in familial osteoarthritis and lumbar disk disease; between estrogen and vitamin D receptor genes and osteoarthritis of the knee; and between a gene linked with hemochromatosis in older patients and hand involvement.[31-35] The significance of these findings in relation to osteoarthritis in the general population are uncertain.[36] In addition to the known heritable and acquired joint dysplasias that cause secondary osteoarthritis, subclinical degrees of dysplasia may be a factor in patients with primary osteoarthritis, particularly of the hip.[37] Many different chromosomal markers have been associated with various patterns of osteoarthritis, suggesting that the genetic component in osteoarthritis most likely involves multiple genes.[38-40]

NORMAL ARTICULAR CARTILAGE

Articular cartilage is specialized connective tissue that covers the weight-bearing surfaces of diarthrodial joints. It is composed of sparsely scattered cells (chondrocytes) within an extracellular matrix composed of collagen, proteoglycans, and water, with a very small component of calcium salt.[41]

Most of the collagen in cartilage is type II collagen, which is arranged in thick bundles and is parallel to the surface of the cartilage in outer portions and more perpendicular to the surface in deeper layers. This arrangement of collagen serves as a

a



b

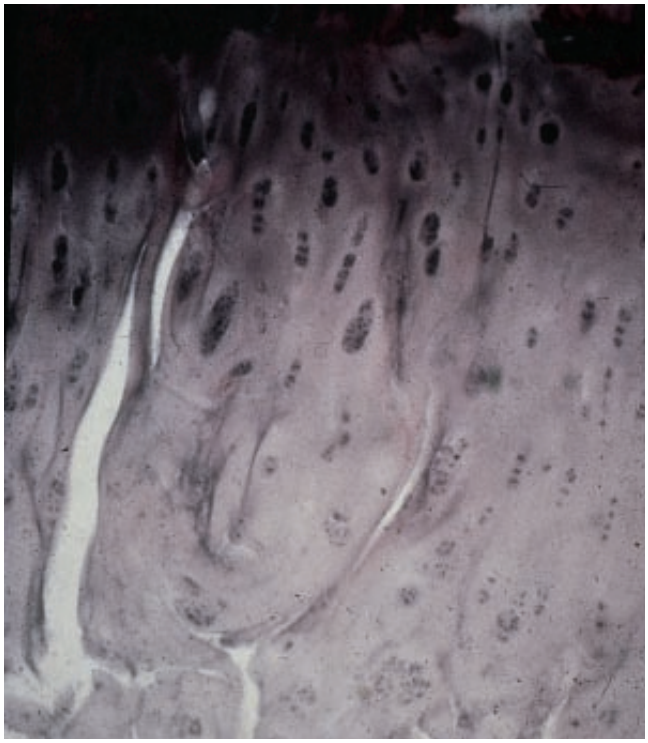


Figure 1 **Microscopic appearance of normal articular cartilage (*a*) and osteoarthritic (*b*) articular cartilage. In normal cartilage, the cartilage surface is smooth and chondrocytes are regularly arranged, mostly as single cells; the background proteoglycan staining is homogeneous; and the subchondral bony plate is intact. In osteoarthritis, there is splitting fissuring of the surface, proliferation and clustering of the chondrocytes, and decreased and irregular staining of the background proteoglycan.**

limiting membrane, distributes compressive forces, and tethers the uncalcified cartilage to the more basilar calcified cartilage and subchondral bone.

The proteoglycan component of the matrix of articular cartilage is composed predominantly of a large molecule called aggrecan, which consists of a large core protein with covalently attached side chains of glycosaminoglycans, most of which are chondroitin sulfate and keratan sulfate. A link protein connects aggrecan to hyaluronic acid, a long, unbranched polysaccharide molecule that can bind several hundred aggrecan molecules. This aggregate of aggrecan molecules forms a very large molecule with a molecular weight of 100 million daltons or more. The molecule has a high fixed negative charge, which allows the retention of large amounts of water.

The collagen matrix and hydrophilic proteoglycan component form a resilient tissue that holds water under pressure and is capable of dissipating much of the force of weight bearing, protecting soft tissues and subchondral bone.

In normal cartilage, the turnover rate of collagen is relatively slow, whereas proteoglycan turnover is rapid. The normal turnover of these matrix components is mediated by the chondrocytes, which synthesize the components and the proteolytic enzymes responsible for their breakdown. Chondrocytes are, in turn, influenced by a number of factors, including polypeptide growth factors and cytokines, structural and physical stimuli, and even the components of the matrix itself.

CHANGES IN OSTEOARTHRITIC CARTILAGE

Pathologic findings suggest that articular cartilage is the site of the primary abnormality in osteoarthritis. There is a loss of homogeneity, and disruption and fragmentation of the surface occur. Uneven staining for proteoglycans is seen in the matrix, and the deeper layers of cartilage are invaded by capillaries from the calcified cartilage. Chondrocytes, which exist as isolated cells in normal cartilage, begin to proliferate and are found in large clusters and clones, and osteophytes are formed, which are covered by irregular hyaline and fibrocartilage [*see Figure 1*].

In early osteoarthritis, the water content of diseased cartilage increases and the cartilage swells, and the collagen fibers are usually smaller and not as tightly organized. The proteoglycan content of cartilage decreases markedly as disease progresses, with shortening of the glycosaminoglycan chains and impaired molecular aggregation.

Osteoarthritic cartilage is characterized by an increase in anabolic and catabolic activity. In the early stages, the synthesis of collagen, proteoglycans, and hyaluronate is increased and chondrocytes tend to replicate. At the same time, synthesis of degradative enzymes such as collagenase, stromelysin, gelatinase, and hyaluronidase is increased, whereas some of the substances that inhibit cartilage destruction are themselves destroyed or inhibited. In the later stages, the anabolic activities of the chondrocytes become insufficient to keep up with the degradative process. The final result is a matrix that is less structurally sound and less well organized on a macromolecular basis, decreasing its ability to withstand the forces required of articular cartilage.

The biochemical and metabolic changes in cartilage that are considered to be potential etiologic factors in osteoarthritis include abnormalities in collagen structure, crystal deposition, inflammatory mediators, and chondrocyte metabolism. The discovery of a familial form of osteoarthritis associated with a specific genetic defect in collagen has led to speculation that similar abnormalities in collagen or other structural components of cartilage may have etiologic importance. In addition, the association of deposition of hemosiderin, copper, or various crystals with secondary forms of osteoarthritis suggests that substances that alter matrix composition can be responsible for degenerative changes.

The relation of calcium-containing crystals to osteoarthritis is complex.[42] Both calcium pyrophosphate dihydrate and basic calcium phosphate crystals have been associated with osteoarthrit-

ic cartilage. In vitro measurement of the by-products of cartilage breakdown suggests that these crystals magnify the degenerative process by stimulation of mitogenesis in fibroblasts and secretion of proteases by cells that ingest the calcium-containing crystals.

The reasons for the increased anabolic and catabolic activities of chondrocytes in osteoarthritis are not well understood.[41] Chondrocytes are influenced by a number of humoral, mechanical, synovial, and cartilage matrix mediators. In particular, prostaglandins, nitric oxide, interleukin-1 (IL-1), transforming growth factor–β, estrogen, and insulinlike growth factor–1 have a variety of stimulatory and inhibitory effects that may be pathogenetically important. In addition, the finding of increased leptin levels in cartilage and synovial fluid of osteoarthritic joints has suggested a role for this substance in the development of osteoarthritis and a role in the relationship between obesity and osteoarthritis.[43,44] Whether the observed abnormalities in these factors are etiologic or merely represent the response of the chondrocyte to other injury is not yet known.[45] The role of inflammation and the potential for damaged cartilage to invoke a more intense inflammatory response than normal cartilage are also areas of ongoing research.[46,47]

The pain of osteoarthritis appears to be derived from inflammation of soft tissue structures surrounding bone, as well as from edema of subchondral bone.[48,49] In addition, edema in subchondral bone is associated with further progression of cartilage damage over time.[50]

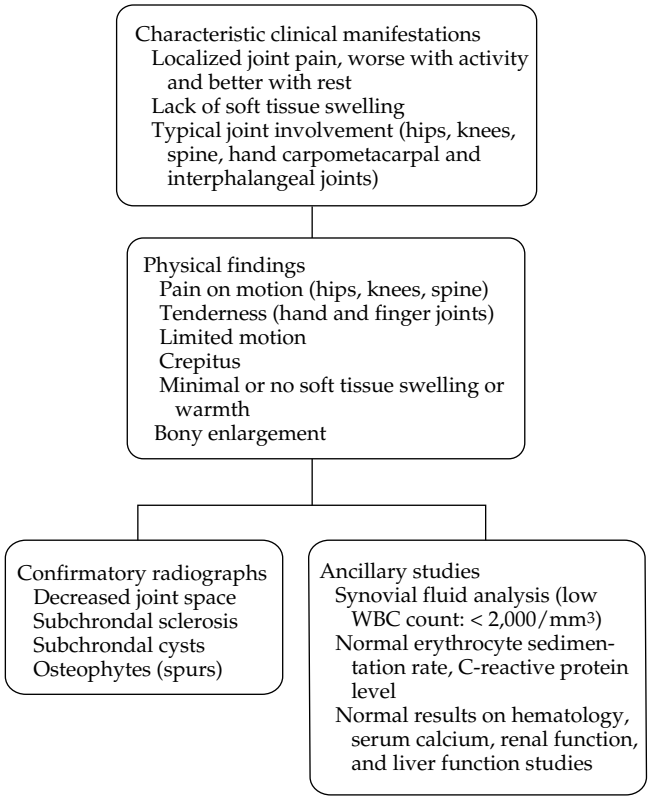


Figure 2 **Diagnosis of osteoarthritis.**

Diagnosis

Characteristic radiographic features are usually considered essential for diagnosis but should be corroborated by the presence of compatible symptoms [*see Figure 2*]. Laboratory studies are useful in the evaluation of patients with osteoarthritis only in that they help to exclude other diagnoses. Thus, the erythrocyte sedimentation rate (ESR), rheumatoid factor, and routine hematologic and biochemical parameters should be normal in patients with osteoarthritis unless the osteoarthritis is attributable to comorbid conditions. Synovial fluid from involved joints is noninflammatory, with leukocyte counts of less than 2,000 cells/mm^3 in most patients. The presence of birefringent calcium pyrophosphate dihydrate crystals is diagnostic of a separate process that frequently is concurrent with typical osteoarthritis. Basic calcium phosphate crystals, which are not birefringent, may be seen frequently in typical osteoarthritis if special stains are used.

Even though some patients have multiple joint involvement, specific joints should be considered individually so that no important problem-causing nonarticular or superimposed process is overlooked.

CLINICAL MANIFESTATIONS

General

Typical symptoms of osteoarthritis include pain, stiffness, swelling, deformity, and loss of function. Pain is usually chronic and localized to the involved joint or joints or referred to nearby areas. Pain may be mild or moderate early in the disease but tends to worsen gradually over many years. Most of the pain is made worse with activity and improves with rest. Morning stiffness is not as prolonged as in patients with inflammatory diseases; morning stiffness in patients with osteoarthritis usually lasts less than an hour. Many patients complain of stiffening, or so-called gelling, during the day, particularly after sitting for extended periods of time. Swelling tends to be mild or moderate and is often related to bony enlargement rather than soft tissue edema. Deformity and loss of function are later manifestations, occurring after many years of disease.

Physical findings in osteoarthritis include crepitus, pain on motion, bony enlargement, and periarticular tenderness. Synovial effusions may be present, particularly in the knee. Erythema and warmth are unusual and should suggest the presence of coexistent crystal-induced inflammation or other conditions. In more advanced disease, limited range of motion, deformity, and instability may become more prominent findings.

Specific Joint Involvement and Complications

Osteoarthritis has a characteristic pattern of involvement in most patients. Joints frequently involved include the distal and proximal interphalangeal joints, as well as the first carpometacarpal joints in the hands, the cervical and lumbar spines, the hips, the knees, and, less commonly, the small joints of the feet or the acromioclavicular joint. The wrists, metacarpophalangeal joints, elbows, shoulders, and ankles are usually not affected unless there is a history of injury to the specific joint, occupational overuse, or underlying condition that might be a cause of secondary osteoarthritis.

Hands The most commonly affected joints in the hands in patients with osteoarthritis are the distal and proximal interphalangeal joints, in which bony enlargement occurs (i.e., Heberden and Bouchard nodes, respectively) [*see Figures 3 and 4*]. The progressive enlargement of these joints occurs slowly over many

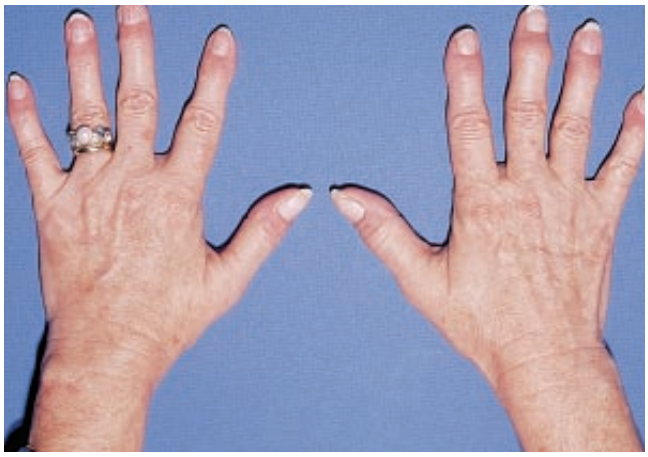


Figure 3 **The hands of a patient with typical primary osteoarthritis show bony enlargement of multiple distal interphalangeal joints (Heberden nodes), mostly on the right, and early bony enlargement of the proximal interphalangeal joints (Bouchard nodes) on the left.**

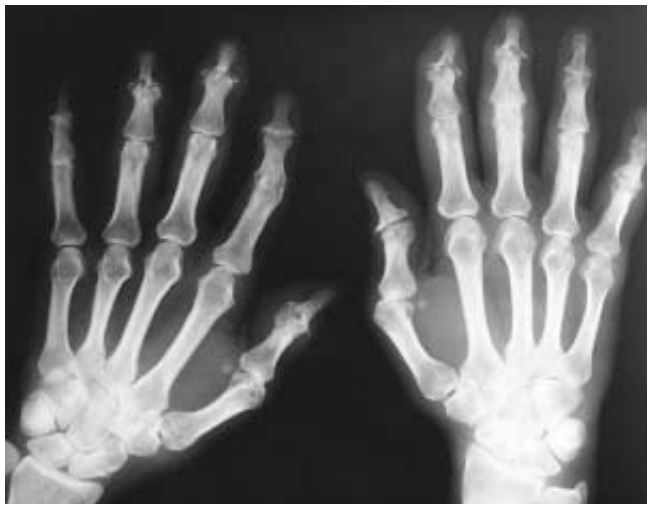


Figure 4 **Severe destructive changes involving all of the distal interphalangeal joints and some of the proximal interphalangeal joints are characteristic of erosive osteoarthritis. Osteophytes are present at the margins of involved joints. Sharply demarcated bone erosions are seen in several distal interphalangeal joints, and bony fusion is seen in the proximal interphalangeal joint of the left second digit.**

years, is frequently familial, and occurs most often in middle-aged or elderly women. Individual joints may go through inflammatory phases with redness and increased swelling and pain, most of which eventually subsides to a bony enlargement. Small gelatinous cysts may develop over the dorsal aspect of the distal interphalangeal joints and either persist or resolve spontaneously. Many patients with Heberden and Bouchard nodes have very little pain most of the time and therefore may not seek medical attention. The carpometacarpal joint of the thumb is another frequently involved joint, either by itself or along with the more distal joints; in such cases, patients experience pain, bony enlargement, and limited motion of the thumb.

Knees Osteoarthritis frequently affects the knees and may be a cause of significant disability. Most patients present with pain that is worse with activity and improves with rest; they report difficulty getting out of chairs or going up steps. Osteoarthritic knees will almost always have crepitus, limited motion, and pain on motion; effusions may or may not be present. In more advanced disease, bony enlargement, instability, and varus angulation may be present. Many patients have involvement of the patellofemoral compartment, but isolated disease in this area should suggest the presence of calcium pyrophosphate deposition disease.

Hips Osteoarthritis of the hips is another common cause of significant pain and disability [*see Figure 5*]. Most patients experience a progressive disabling pain, usually in the upper thigh or inguinal region, sometimes radiating to the knee. Pain is worse with ambulation and may cause the patient to limp. Patients may also complain of difficulty with activities such as tying shoes, and limited hip motion is found on physical examination.

Spine Osteoarthritis of the cervical and lumbar spine is referred to as spondylosis. Involvement of the intervertebral disk spaces or the posterior spinal facet joints may cause chronic back or neck pain that worsens with activity and improves with rest. Disk degeneration may be complicated by protrusion of the nucleus pulposus, causing nerve root compression with radicular pain or muscle weakness. In patients with extensive degenerative changes with fibrosis and osteophytes, stenosis of the spinal canal can occur, resulting in chronic cord compression in the cervical spine or compression of the cauda equina in the lumbar region. Lumbar spinal stenosis, causing chronic radicular leg pain that is worse with activity and better with rest (neurogenic claudication), is a common complication in elderly patients. A variant of spinal osteoarthritis occurring in the thoracic spine,

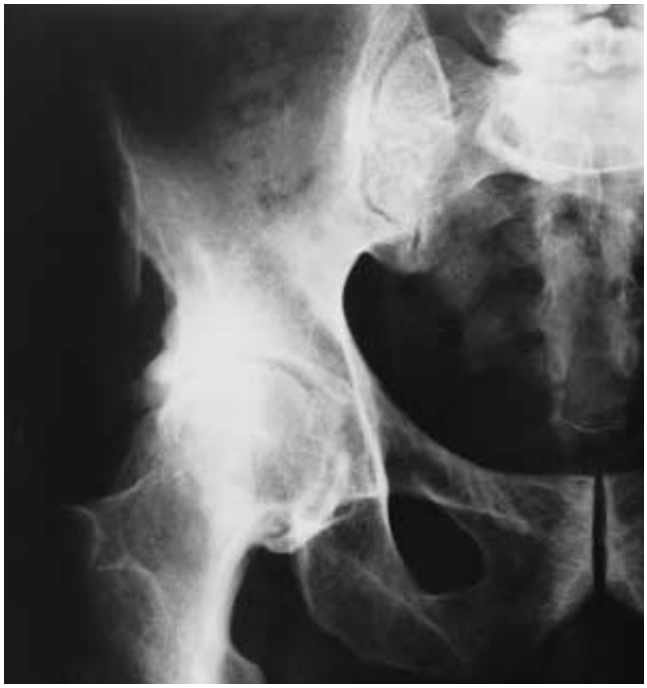


Figure 5 **The hip is a common site of involvement in osteoarthritis. Joint space narrowing is most prominent at the superior and lateral aspects of the joint. Increased bony density (sclerosis) is seen in the subchondral bone on both sides of the joint, along with early subchondral cysts and osteophytes (bony spurs) over the superior and inferior aspects of the acetabulum.**

known as diffuse idiopathic skeletal hyperostosis (DISH), is characterized by extensive bridging osteophytes and may cause loss of motion but little pain.

Radiologic Features

Typical radiographic findings in osteoarthritis include joint space narrowing, subchondral bone sclerosis, subchondral cysts, and osteophytes (bony spurs) [*see Figure 5*]. Joint space narrowing, resulting from loss of cartilage, is often asymmetrical and may be the only finding early in the disease process. In weight-bearing joints such as the knees, narrowing may be seen only in a standing view and may be missed in a radiograph obtained with the patient in the recumbent position. In more chronic disease, the hypertrophic features of subchondral sclerosis and osteophyte formation become more prominent, and subluxations or fusion of the joint may become apparent in more severely affected joints. In the small interphalangeal joints of the fingers, central erosions may be seen within the joint space; these erosions should be easily distinguishable from the periarticular erosions of rheumatoid arthritis.

Differential Diagnosis

Because of the high frequency of incidental radiographic changes in the general population, it is important not to attribute all musculoskeletal pain to osteoarthritis, even in patients with radiographic abnormalities. Alternative diagnoses should be made or coexistent conditions suspected in patients who are considered to be at low risk for osteoarthritis (e.g., younger patients) or in those who present with atypical pain patterns or atypical joint involvement. Patients with a relatively sudden onset of pain or with severe pain early in their presentation most often have something other than osteoarthritis. Problems in the wrists, elbows, shoulders, or ankles should raise concerns about other types of arthritis or secondary types of osteoarthritis.

Crystal-induced arthritis should always be considered in patients with acute pain, particularly if swelling and erythema are prominent. Calcium pyrophosphate deposition disease is common in the knees and hips and often coexists with osteoarthritis. Other joints frequently involved are the wrists and shoulders. Detection of chondrocalcinosis on x-ray or of crystals in synovial fluid confirms the diagnosis. Gout usually affects foot and ankle joints in early disease and is not often confused with osteoarthritis, but involvement of the knees is common in later disease. In addition, elderly women with Heberden and Bouchard nodes in the hands may have superimposed attacks in these joints as an initial manifestation of gout. Thus, examination of fluid from these joints for urate crystals may be essential in differentiating gout from an inflammatory flare of erosive osteoarthritis.

Rheumatoid arthritis can usually be distinguished from osteoarthritis on the basis of a different pattern of joint disease, more prominent morning stiffness, and soft tissue swelling and warmth on physical examination. In some patients, the patterns of joint disease may overlap, particularly in the proximal interphalangeal joints, hips, and knees. Thus, in some patients, the presence of an elevated ESR, a high-titer rheumatoid factor, or periarticular erosive changes may be the only way to distinguish these two common conditions.

Polymyalgia rheumatica is a disease of the elderly and is often seen in patients with underlying osteoarthritis. Patients typically have a change in the pattern of pain, more localized to the shoulder and hip girdles, with few peripheral joint symptoms. Morning stiffness is a prominent feature, and the diagnosis is usually more likely if the ESR is markedly elevated. However, because modest elevations of ESR are seen in many normal elderly individuals, the differentiation of this condition from osteoarthritis is often difficult. In some patients, a rapid response of symptoms to a low dose of corticosteroid is helpful in making a diagnosis.

Ankylosing spondylitis is usually a disease that first manifests in young adulthood and should not be confused with spinal osteoarthritis. However, some patients have only mild levels of pain and may not seek medical attention until later years. In such patients, the radiographic changes in the cervical and lumbar spine in the two conditions should make differentiation between them relatively easy.

Psoriatic arthritis, when present in a classic distribution in the distal interphalangeal joints of the fingers, may mimic Heberden nodes. Psoriatic arthritis usually occurs in younger persons and is more common in males, but differentiation between psoriatic arthritis and Heberden nodes may still be difficult. In young patients with arthritis of the distal interphalangeal joints, a careful search for psoriatic skin lesions and nail changes is essential. On physical examination, the swelling of involved joints is usually greater in the soft tissues, with less bony enlargement. Radiography will usually show more erosive changes and fewer osteophytic changes than in typical osteoarthritis.

Disorders of bone near joints can be confused with osteoarthritis. Osteonecrosis of the hip, knee, or shoulder may cause pain and restricted motion without significant signs of inflammation. Radiographs may be normal initially, and follow-up radiographs or magnetic resonance imaging may be necessary to differentiate this condition from osteoarthritis. Paget disease or osteoporotic fractures may cause pain in the back and hip girdle that is similar to that of osteoarthritis, although the pain is often more severe and acute in patients with fractures.

Nonarticular pain syndromes involving tendons, bursae, peripheral nerves, and internal joint structures may cause pain similar to that of osteoarthritis. Examples include de Quervain tenosynovitis or carpal tunnel syndrome in the hand, trochanteric bursitis or meralgia paresthetica in the hip, anserine bursitis or meniscal tears in the knee, and plantar fasciitis and interdigital neuromas in the feet. Knowledge of nonarticular pain syndromes and the characteristic patterns of symptoms and physical findings in each is essential to diagnosing and differentiating these syndromes from osteoarthritis in the same area.

Management

There is no cure for osteoarthritis and no therapy known to prevent or retard the degenerative biologic process in articular cartilage. Thus, the treatment of osteoarthritis is focused primarily on relieving symptoms and improving function.[51-53] Treatment decisions should be based on the severity and distribution of joint involvement, considered in the light of the patient's other medical problems that might affect the safety and effectiveness of any chosen therapy [*see Table 2*].

NONPHARMACOLOGIC MEASURES

Nonpharmacologic measures that have the potential to improve outcomes in osteoarthritis include patient education, physical and occupational therapy assessment and interventions, exercise, weight loss, and dietary measures.[7] Exercise, in particular, should be a part of the therapeutic regimen in every

Table 2 Treatment of Osteoarthritis

Treatment Type	*Useful for What or Whom?*	*Measure*	*Comments*
Nonpharmacologic	All patients	Exercise	Range of motion and strengthening of muscles around affected joints
		Weight loss	Particularly valuable in patients with involvement of weight-bearing joints
		Dietary measures	Adequate intake of calcium, vitamin C, and vitamin D
Pharmacologic	Most or all patients	Simple analgesics	Acetaminophen, tramadol, narcotics in selected cases
		NSAIDs	Nonselective NSAIDs for patients at low risk for GI complications; otherwise, consider addition of misoprostol, a proton pump inhibitor, or an H_2 antagonist or use of a cyclooxygenase-2–specific NSAID (coxib)
Ancillary medical and surgical	Selected joints or patients	Splints	Specific for each joint
		Canes or other orthotics	—
		Corticosteroid injections	Knees, fingers; other joints in selected cases
		Hyaluronic acid injections	Knees in some patients
		Arthroscopic surgery	For patients with mechanical symptoms or findings
		Osteotomy	Knees in selected patients
		Total joint replacement	—

NSAIDs—nonsteroidal anti-inflammatory drugs

patient. Quadriceps weakness contributes significantly to disability in patients with osteoarthritis of the knee, and exercises designed to strengthen quadriceps have potential to lessen pain and disability.[54] In addition, aerobic exercise, such as a walking program, can improve function and reduce pain. Most studies of patients with osteoarthritis have found that regular activity is associated with a better outcome.[55-57] However, compliance with exercise programs is often low, and regular supervised follow-up may be helpful.

The role of obesity as an etiologic factor in osteoarthritis of the knee is well established, and some data suggest that weight loss may reduce the risk of development of symptoms in patients predisposed to osteoarthritis. Even though few prospective studies have been done, patients with osteoarthritis of the knees and hips should be encouraged to lose weight if they are above ideal body weight. In addition, epidemiologic studies have suggested a role for adequate dietary vitamin C and D intake in reducing the risk of progression of established osteoarthritis.[14,58] In some patients, measures designed to alter the biomechanical forces on diseased joints should be considered, including patellar taping, wedged insoles, bracing, canes, and crutches.

PHARMACOLOGIC THERAPY

The primary goal of drug therapy in osteoarthritis is to relieve pain. Acetaminophen may be as effective as nonsteroidal anti-inflammatory drugs (NSAIDs),[59] so it should be prescribed initially in most patients. In most cases, up to 3,000 to 4,000 mg of acetaminophen a day can be given. Doses should be limited in patients with exposure to other potentially hepatotoxic substances. In particular, patients who take acetaminophen regularly should be advised to limit alcohol ingestion and be warned about the increased risk of acetaminophen hepatotoxicity in heavy drinkers. Opioids are generally avoided in osteoarthritis but may be useful in selected patients. However, these agents should be used with caution in elderly patients.[60] Tramadol, a centrally acting analgesic with dual mechanisms, may give relief comparable to that achieved with acetaminophen and codeine. Topical capsaicin may be useful in some patients, especially those with involvement of the knees and hands.

A number of other nutritional supplements and topical therapies have been investigated for the treatment of osteoarthritis. These studies have been poorly controlled or limited, however. Little evidence exists to support widespread use of any of these alternative or complementary therapies.[61,62]

NSAIDs are useful in osteoarthritis mostly for their analgesic effects and, in most patients, are more effective than acetaminophen.[63-65] Unfortunately, NSAIDs are associated with an increased risk of gastric ulcers and bleeding, particularly in patients with a history of gastrointestinal disease, those on concomitant steroids or anticoagulants, and those older than 65 years. Strategies to reduce this toxicity include the use of lower doses of NSAIDs or concomitant use of misoprostol, histamine$_2$ receptor antagonists, and proton pump inhibitors. The cyclooxygenase-2 (COX-2)–specific NSAIDs (celecoxib and valdecoxib) have been shown to reduce the risk of serious gastrointestinal complications, compared with the nonselective COX inhibitors. The association of cardiovascular events with the prolonged use of the COX-2 inhibitor rofecoxib (Vioxx) prompted the withdrawal of the drug from the global market in 2004.[66] The Adenomatous Polyp Prevention on Vioxx (APPROVe) study documented that patients taking the drug for 18 months were twice as likely to experience cardiovascular events as patients taking placebo (after 3 years, the rates of cardiovascular events were 1.5% for rofecoxib and 0.7% for placebo).[66] Data from several sources are consistent in finding an increase of hypertension and congestive heart failure in patients taking rofecoxib.[67-69] There is scant evidence to assess the long-term risk of other COX-2 inhibitors. When prescribing COX-2 inhibitors, clinicians should weigh the potential cardiovascular risks against anticipated benefits and consider issues such as dose and comorbid conditions.

Intra-articular corticosteroid injections may be useful in treating selected joints, particularly during exacerbations characterized by increased pain and effusion, and injections in symptomatic knees every 3 months may be a safe and effective means of reducing pain and improving function over longer periods of time.[70] Some animal and in vitro studies have suggested that steroids have a detrimental effect on articular cartilage, but there are few clinical data to support this concern in patients with osteoarthritis. Intra-articular hyaluronic acid derivatives (Hyalgan and Synvisc), given in a series of three to five weekly injections, have been shown to be superior to placebo in most studies and may be useful in relieving pain in selected patients with less advanced disease.[71,72]

SURGERY

In patients with badly damaged knees and hips, total joint replacement is an effective option. Almost all patients experience significant pain relief, and some have improved range of motion. Joint loosening and infection are potential late complications in prosthetic joints but are uncommon. Arthroscopic debridement of affected joints is no better than placebo in unselected patients with osteoarthritis of the knee.[73] Thus, this procedure should be reserved for patients with mechanical symptoms suggesting internal derangement. Realignment of a degenerative knee to allow redistribution of forces is sometimes attempted by a high tibial wedge osteotomy, particularly in younger patients with valgus deformities.

Biologic approaches to the surgical treatment of osteoarthritis have been explored. These include local enhancement of bone marrow progenitor cells and various forms of cartilage transplantation.[51] In addition, stem cell transplantation has been investigated in animal models of osteoarthritis.[74] Most of these approaches are in early stages of development, and none is likely to enter clinical use in the foreseeable future.

PROTECTING CARTILAGE

Therapies with potential to prevent or retard the progression of articular cartilage breakdown have received a great deal of attention in recent years.[75] Agents considered to have so-called chondroprotective potential include tetracyclines, protease inhibitors, glycosaminoglycan compounds, growth factors, and cytokine inhibitors. Oral glucosamine and chondroitin sulfate have been promoted as health food supplements to improve cartilage, but most of the clinical studies with these agents have demonstrated only modest pain relief, compared with placebo, and studies to assess the effect on cartilage are ongoing.[76-79] Inconsistent dosages in studies and lack of standardization of available preparations have complicated assessment of these agents' value.

Tetracyclines have been shown to reduce the severity of osteoarthritis in animals, probably by inhibiting metalloprotease activity, and are being studied in early human trials. Other approaches to disease modification being investigated in animal models include other agents that inhibit metalloproteases or nitric oxide synthase inhibitors.[80]

Biologic therapies designed to augment growth factors or inhibit cytokines have also been investigated in animal models of osteoarthritis. These have included attempts to introduce growth factors or IL-1 receptor antagonist through the use of gene therapy.[81]

Prognosis

Osteoarthritis is a slowly progressive condition with a variable prognosis.[2] Radiographically, most joints will either remain stable or gradually worsen over a 5- to 15-year period. In most patients, symptoms evolve over many years and may spontaneously remit for long periods of time without explanation. Progression of osteoarthritis of the hand is particularly hard to measure because pain levels frequently improve after involved joints become fused. Disease may progress more rapidly in the hips and knees of older women with osteopenic bone. However, in general, predicting the prognosis in patients with osteoarthritis is difficult.

The author has received grants for clinical research from Abbott Laboratories, Amgen, Inc, Bristol-Myers Squibb Co., and Isis Pharmaceuticals, Inc., and has served as advisor or consultant to TAP Pharmaceutical Products, Inc.

References

1. Lawrence RC, Helmick CG, Arnett FC, et al: Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 41:778, 1998

2. Felson DT, Lawrence RC, Dieppe PA, et al: Osteoarthritis: new insights. I: The disease and its risk factors. Ann Intern Med 133:635, 2000

3. Poole J, Sayer AA, Hardy R, et al: Patterns of interphalangeal hand joint involvement of osteoarthritis among men and women: a British cohort study. Arthritis Rheum 48:3371, 2003

4. Felson DT, Nevitt MC, Zhang Y, et al: High prevalence of lateral knee osteoarthritis in Beijing Chinese compared with Framingham Caucasian subjects. Arthritis Rheum 46:1217, 2002

5. Nevitt MC, Xu L, Zhang Y, et al: Very low prevalence of hip osteoarthritis among Chinese elderly in Beijing, China, compared with whites in the United States: the Beijing osteoarthritis study. Arthritis Rheum 46:1773, 2002

6. Zhang Y, Xu L, Nevitt MC, et al**:** Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: the Beijing Osteoarthritis Study. Arthritis Rheum 44:2065, 2001

7. Felson DT, Zhang Y: An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum 41:1343, 1998

8. Gelber AC, Hochberg MC, Mead LA, et al: Body mass index in young men and the risk of subsequent knee and hip osteoarthritis. Am J Med 107:542, 1999

9. Sharma L, Lou C, Cahue S, et al: The mechanism of the effect of obesity in knee osteoarthritis: the mediating role of malalignment. Arthritis Rheum 43:568, 2000

10. Sharma L, Song J, Felson DT, et al: The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA 286:188, 2001

11. Manek NJ, Hart D, Spector TD, et al: The association of body mass index and osteoarthritis of the knee joint: an examination of genetic and environmental influences. Arthritis Rheum 48:1024, 2003

12. Hart DJ, Cronin C, Daniels M, et al: The relationship of bone density and fracture to incident and progressive radiographic osteoarthritis of the knee: the Chingford Study. Arthritis Rheum 46:92, 2002

13. Lane NE, Nevitt MC**:** Osteoarthritis, bone mass, and fractures: how are they related? Arthritis Rheum 46:1, 2002

14. McAlindon TE, Felson DT, Zhang Y, et al: Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. Ann Intern Med 125:353, 1996

15. McAlindon TE, Wilson PFW, Aliabadi P, et al: Level of physical activity and the risk of radiographic and symptomatic knee osteoarthritis in the elderly: the Framingham study. Am J Med 106:151, 1999

16. Coggon D, Croft P, Kellingray S, et al: Occupational physical activities and osteoarthritis of the knee. Arthritis Rheum 43:1443, 2000

17. Spector TD, Harris PA, Hart DJ, et al: Risk of osteoarthritis associated with long-term weight bearing sports: a radiologic survey of the hips and knees in female ex-athletes and population controls. Arthritis Rheum 39:988, 1996

18. Lane NE, Oehlert JW, Bloch DA, et al: The relationship of running to osteoarthritis of the knee and hip and bone mineral density of the lumbar spine: a 9 year longitudinal study. J Rheumatol 25:334, 1998

19. Gelber A, Hochberg M, Mead L, et al: Joint injury in young adults and risk for subsequent knee and hip osteoarthritis. Ann Intern Med 133:321, 2000

20. Biswal S, Hastie T, Andriacchi TP, et al: Risk factors for progressive cartilage loss in the knee. Arthritis Rheum 46:2884, 2002

21. Lievense AM, Bierma-Zeinstra SM, Verhagen AP, et al: Influence of sporting activities on the development of osteoarthritis of the hip: a systematic review. Arthritis Rheum 49:228, 2003

22. Slemenda C, Brandt KD, Heilman DK, et al: Quadriceps weakness and osteoarthritis of the knee. Ann Intern Med 127:97, 1997

23. Sharma L, Pai YC, Holtkamp K, et al: Is knee joint proprioception worse in the arthritic knee versus the unaffected knee in unilateral knee osteoarthritis? Arthritis Rheum 40:1518, 1997

24. Hassan BS, Doherty SA, Mockett S, et al: Effect of pain reduction on postural sway, proprioception, and quadriceps strength in subjects with knee osteoarthritis. Ann Rheum Dis 61:422, 2002

25. Vanwanseele B, Eckstein F, Knecht H, et al: Knee cartilage of spinal cord-injured patients displays progressive thinning in the absence of normal joint loading and movement. Arthritis Rheum 46:2073, 2002

26. Vanwanseele B, Eckstein F, Knecht H, et al: Longitudinal analysis of cartilage atrophy in the knees of patients with spinal cord injury. Arthritis Rheum 48:3377, 2003

27. Holderbaum D, Haqqi TM, Moskowitz RW: Genetics and osteoarthritis: exposing the iceberg. Arthritis Rheum 42:397, 1999

28. Jonsson H, Manolescu I, Stefansson SE, et al: The inheritance of hand osteoarthritis in Iceland. Arthritis Rheum 48:391, 2003

29. Lanyon P, Muir K, Doherty S, et al: Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. BMJ 321:1179, 2000

30. MacGregor AJ, Antoniades L, Matson M, et al: The genetic contribution to radiographic hip osteoarthritis in women: results of a classic twin study. Arthritis Rheum 43:2410, 2000

31. Ross JM, Kowalchuk RM, Shaulinsky J, et al: Association of heterozygous hemochromatosis C282Y gene mutation with hand osteoarthritis. J Rheumatol 30:121, 2003

32. Bergink AP, van Meurs JB, Loughlin J, et al: Estrogen receptor alpha gene haplotype is associated with radiographic osteoarthritis of the knee in elderly men and women. Arthritis Rheum 48:1913, 2003

33. Knowlton RG, Katzenstein PL, Moskowitz RW, et al: Genetic linkage of polymorphism to type II procollagen gene (COL 2A1) to primary osteoarthritis associated with mild chondrodysplasia. N Engl J Med 322:526, 1990

34. Keen RW, Hart DJ, Lanchbury JS, et al: Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. Arthritis Rheum 40:1444, 1997

35. Paassilta P, Lohiniva J, Goring H, et al: Identification of a novel common genetic risk factor for lumbar disc disease. JAMA 285:1843, 2001

36. Baldwin CT, Cupples LA, Joost O, et al: Absence of linkage or association for osteoarthritis with the vitamin D receptor/type II collagen locus: the Framingham Osteoarthritis Study. J Rheumatol 29:161, 2002

37. Lane NE, Lin P, Christiansen L, et al: Association of mild acetabular dysplasia with an increased risk of incident osteoarthritis in elderly white women: the study of osteoporotic fractures. Arthritis Rheum 43:400, 2000

38. Gillaspy E, Spreckley K, Wallis G, et al: Investigation of linkage on chromosome 2q and hand and knee osteoarthritis. Arthritis Rheum 46:3386, 2002

39. Chapman K, Mustafa Z, Dowling B, et al: Finer linkage mapping of primary hip osteoarthritis susceptibility on chromosome 11q in a cohort of affected female sibling pairs. Arthritis Rheum 46:1780, 2002

40. Demissie S, Cupples LA, Myers R, et al: Genome scan for quantity of hand osteoarthritis: the Framingham Study. Arthritis Rheum 46:946, 2002

41. Goldring MB: The role of the chondrocyte in osteoarthritis. Arthritis Rheum 43:1916, 2000

42. Jaovisidha K, Rosenthal AK**:** Calcium crystals in osteoarthritis. Curr Opin Rheumatol 14:298, 2002

43. Dumond H, Presle N, Terlain B, et al**:** Evidence for a key role of leptin in osteoarthritis. Arthritis Rheum 48:3118, 2003

44. Loeser RF**:** Systemic and local regulation of articular cartilage metabolism: where does leptin fit in the puzzle? Arthritis Rheum 48:3009, 2003

45. Clements KM, Price JS, Chambers MG, et al: Gene deletion of either interleukin-1beta, interleukin-1beta-converting enzyme, inducible nitric oxide synthase, or stromelysin 1 accelerates the development of knee osteoarthritis in mice after surgical transection of the medial collateral ligament and partial medial meniscectomy. Arthritis Rheum 48:3452, 2003

46. Pelletier JP, Martel-Pelletier J, Abramson SB: Osteoarthritis, an inflammatory disease: potential implication for the selection of new therapeutic targets. Arthritis Rheum 44:1237, 2001

47. Yuan GH, Masuko-Hongo K, Kato T, et al: Immunologic intervention in the pathogenesis of osteoarthritis. Arthritis Rheum 48:602, 2003

48. Felson DT, Chaisson CE, Hill CL, et al: The association of bone marrow lesions with pain in knee osteoarthritis. Ann Intern Med 134:594, 2001

49. Hill CL, Gale DR, Chaisson CE, et al: Periarticular lesions detected on magnetic resonance imaging: prevalence in knees with and without symptoms. Arthritis Rheum 48:2836, 2003

50. Felson DT, McLaughlin S, Goggins J, et al: Bone marrow edema and its relation to progression of knee osteoarthritis. Ann Intern Med 139:330, 2003

51. Felson DT, Lawrence RC, Hochberg MC, et al: Osteoarthritis: new insights: II. Treatment approaches. Ann Intern Med 133:726, 2000

52. Recommendations for the medical management of osteoarthritis of the hip and knee. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Arthritis Rheum 43:1905, 2000

53. Pendelton A, Arden N, Dougados M, et al: EULAR recommendations for the management of knee osteoarthritis: report of a task force of the Standing Committee for the International Clinical Studies Including Therapeutic Trials (ESCIST). Ann Rheum Dis 59:936, 2000

54. Deyle GD, Henderson NE, Matekel RL, et al: Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee: a randomized, controlled study. Ann Intern Med 132:173, 2000

55. Van Baar ME, Assendelft WJJ, Dekker J, et al: Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 42:1361, 1999

56. Penninx BW, Messier SP, Rejeski WJ, et al: Physical exercise and the prevention of disability in activities of daily living in older persons with osteoarthritis. Arch Intern Med 161:2309, 2001

57. Sharma L, Cahue S, Song J, et al: Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. Arthritis Rheum 48:3359, 2003

58. McAlindon T, Jacques P, Zhang Y, et al: Do antioxidant micronutrients protect against the development and progression of knee osteoarthritis? Arthritis Rheum 39:648, 1996

59. Bradley JD, Brandt KD, Katz BP, et al: Comparison of an anti-inflammatory dose of ibuprofen, an analgesic dose of ibuprofen, and acetaminophen in the treatment of patients with osteoarthritis of the knee. N Engl J Med 325:87, 1991

60. Peleso PM: Opioid therapy for osteoarthritis of the hip and knee: use it or lose it (editorial)? J Rheumatol 28:6, 2001

61. Hochberg MC: Multidisciplinary integrative approach to treating knee pain in patients with osteoarthritis. Ann Intern Med 139:781, 2003

62. Long L, Soeken K, Ernst E: Herbal medicines for the treatment of osteoarthritis: a systematic review. Rheumatology (Oxford) 40:779, 2001

63. Felson DT: The verdict favors nonsteroidal antiinflammatory drugs for treatment of osteoarthritis and a plea for more evidence on other treatments. Arthritis Rheum 44:1477, 2001

64. Pincus T, Koch GG, Sokka T, et al: A randomized, double-blind, crossover clinical trial of diclofenac plus misoprostol versus acetaminophen in patients with osteoarthritis of the hip or knee. Arthritis Rheum 44:1587, 2001

65. Case JP, Baliunas AJ, Block JA: Lack of efficacy of acetaminophen in treating symptomatic knee osteoarthritis: a randomized, double-blind, placebo-controlled comparison trial with diclofenac sodium. Arch Intern Med 163:169, 2003

66. Singh D: Merck withdraws arthritis drug worldwide. BMJ 329:816, 2004

67. Bombardier C, Laine L, Reicin A, et al: Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 343:1520, 2000

68. Solomon DH, Schneeweiss S, Glynn RJ, et al: Relationship between selective cyclooxygnease-2 inhibitors and acute myocardial infarction in older adults. Circulation 209:2068, 2004

69. Solomon DH, Schneeweiss S, Levin R, et al: Relationship between COX-2 specific inhibitors and hypertension. Hypertension 44:140, 2004

70. Raynauld JP, Buckland-Wright C, Ward R, et al: Safety and efficacy of long-term intraarticular steroid injections in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 48:370, 2003

71. Brandt K, Smith G, Simon L: Intraarticular injection of hyaluronan as treatment for knee osteoarthritis: what is the evidence (review)? Arthritis Rheum 43:1192, 2000

72. Felson DT, Anderson JJ: Hyaluronate sodium injections for osteoarthritis: hope, hype, and hard truths. Arch Intern Med 162:245, 2002

73. Moseley JB, O'Malley K, Petersen NJ, et al: A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 347:81, 2002

74. Murphy JM, Fink DJ, Hunziker EB, et al: Stem cell therapy in a caprine model of osteoarthritis. Arthritis Rheum 48:3464, 2003

75. Dieppe PA: The management of osteoarthritis in the third millennium. Scand J Rheumatol 29:279, 2000

76. McAlindon TE, LaValley MP, Bulin JP, et al: Glucosamine and chondroitin for treatment of osteoarthritis: a systematic quality assessment and meta-analysis. JAMA 283:1469, 2000

77. Reginster JY, Deroisy R, Rovati CL, et al: Long-term effects of glucosamine sulphate on osteoarthritis progression: a randomised, placebo-controlled clinical trial. Lancet 357:251, 2001

78. Richy F, Bruyere O, Ethgen O, et al: Structural and symptomatic efficacy of glucosamine and chondroitin in knee osteoarthritis: a comprehensive meta-analysis. Arch Intern Med 163:1514, 2003

79. Mengshol JA, Mix KS, Brinckerhoff CE: Matrix metalloproteinases as therapeutic targets in arthritic diseases: bull's-eye or missing the mark? Arthritis Rheum 46:13, 2002

80. Towheed TE: Current status of glucosamine therapy in osteoarthritis. Arthritis Rheum 49:601, 2003

81. Gelse K, Von Der Mark K, Aigner T, et al: Articular cartilage repair by gene therapy using growth factor-producing mesenchymal cells. Arthritis Rheum 48:430, 2003

Acknowledgment

Figure 1 Courtesy of Richard Hard, M.D.

XII BACK PAIN AND COMMON MUSCULOSKELETAL PROBLEMS

CHRISTOPHER WISE, M.D., F.A.C.P.

A large proportion of the musculoskeletal problems for which patients seek medical attention are related to periarticular structures and do not represent a true articular process or a more generalized systemic illness.[1] Knowledge of the common nonarticular regional rheumatic disorders is important because of their high prevalence in primary care practice, the dependence on clinical findings for diagnosis, and the high cost that can result from unnecessary laboratory evaluations. The ability to recognize important patterns of pain and associated physical signs is essential to making a correct diagnosis; in most cases, radiographic and laboratory studies are not needed. Diagnostic studies should be utilized judiciously and must be interpreted in the light of existing clinical findings and prestudy suspicion for specific diagnoses.

Most regional rheumatic disorders respond to local measures, such as application of heat or cold, splinting, and injection of glucocorticoids. Nonsteroidal anti-inflammatory drugs (NSAIDs) or mild analgesic medications are often helpful therapeutic adjuncts. Referral for surgical intervention may be indicated for patients with certain conditions. For example, in cases of cervical or lumbar disk disease or spinal stenosis with definite nerve entrapment or spinal cord compression, well-timed decompression may be necessary to restore function or prevent further functional impairment. Arthroscopic intervention is sometimes useful to better define and treat refractory knee and shoulder pain syndromes. Surgical release is indicated for entrapment neuropathies when there is evidence of motor dysfunction. Surgical consultation may be useful for a variety of other syndromes when the response to conservative measures proves to be less than optimal. Physical therapy and occupational therapy are useful for many patients—particularly those patients who have persistent back and shoulder pain—though these therapies may constitute an important part of the treatment of almost any refractory regional pain syndrome.

Definitions

Common regional rheumatic disorders include various types of bursitis, tendinitis, tenosynovitis, myofascial pain, and entrapment neuropathies. Bursitis results from mechanical or inflammatory changes of one of the many bursae in the body. Bursae are synovia-lined sacs around the joints that serve to minimize friction between tendons, ligaments, and bony structures. Tendinitis usually results from trauma or overuse of tissues near sites where tendons attach to bone or at the musculotendinous junction. Myofascial pain originates at sites within muscle groups and surrounding fascial tissues that become tender and painful as a result of localized injury or overuse. Entrapment neuropathies occur at sites where peripheral nerves are compressed as they traverse periarticular areas that allow relatively little room for free movement of the affected nerves.

Neck Pain

Neck pain may result from degenerative changes in the cervical disks and zygapophyseal (facet) joints or from a variety of muscular, ligamentous, and tendinous conditions.[2] In whiplash injuries occurring after rapid acceleration or deceleration and hyperextension of the head in motor vehicle accidents, a number of structures may be injured.[3] Recovery from whiplash injuries is often incomplete, and a combination of physical and psychosocial factors may contribute to prolongation of pain.[4] Judicious use of analgesics, muscle relaxants, and physiotherapy proves helpful in some patients. The zygapophyseal joints may be the source of pain, and local nerve block with an anesthetic or ablation may be helpful in selected patients.

The term cervical sprain denotes transient neck pain associated with muscle tenderness and spasm. Cervical sprain usually responds to heat, rest, and, occasionally, immobilization and traction. Manual therapy, range-of-motion or strength-training exercises, or acupuncture may provide relief in some patients.[5-7] In cervical disk herniation, nerve root impingement results in pain, paresthesia, and sometimes muscle weakness in the distribution of the affected nerve (usually at the C5 to C7 level). In such patients, radiographic documentation and surgical decompression are sometimes needed if symptoms do not improve with rest or traction or if significant neurologic deficit is present. In some patients with long-standing cervical spondylosis, cervical stenosis may cause chronic compression of the spinal cord (most often at the C3 to C5 level). Surgical decompression is indicated in patients with evolving myelopathy.

Back Pain

Low back pain is the most common musculoskeletal complaint requiring medical attention; it is the fifth most common reason for all physician visits.[8-10] Over half of the general population will seek medical attention for back pain at some point in their lives. An increased risk of back pain is associated with male sex, smoking, frequent lifting of heavy objects, poor general health and conditioning, and certain occupational and sports activities. In addition, preexisting psychological distress, compensation issues, other chronic pain, and job dissatisfaction may play a role in persistent back pain.[11,12] In most patients, the cause of pain cannot be determined with any degree of certainty and is usually attributed to muscular or ligamentous strain, facet joint arthritis, or disk pressure on the annulus fibrosus, vertebral end plate, or nerve roots.[13]

ACUTE BACK PAIN

Diagnosis

For patients with acute back pain, the initial history should be used to identify those who are at increased risk for serious underlying conditions, such as fracture, infection, tumor, or major neurologic deficit[8] [*see Table 1*]. The presence of such factors in patients with acute back pain may indicate the need for radiographic and laboratory studies earlier than in patients without such factors. The initial physical examination should include

Table 1 Indications That Acute Back Pain May Involve Underlying Conditions

Patient demographics	Age > 70 yr History of cancer Glucocorticoid or immunosuppressive drug therapy Alcohol or I.V. drug abuse
Historical features	Weight loss Fever Pain increased by rest
Neurologic symptoms	Bowel or bladder dysfunction Saddle block anesthesia Progressive motor weakness

evaluation for areas of localized bony tenderness and assessment of flexion and straight leg raising. Because acute low back pain will improve within a month in over 90% of patients, further evaluation is usually unnecessary. Plain radiographs should be reserved for patients at high risk for more serious underlying conditions [*see Table 1*], because abnormal findings on plain films are common and do not correlate with back pain. Early use of magnetic resonance imaging or computed tomography has not been shown to change treatment choices or outcomes; consequently, these studies should be reserved for patients with persistent pain and sciatica.[14,15]

Treatment

A number of therapeutic interventions are available for acute back pain, but data supporting efficacy are minimal for most therapies. Strict bed rest should be kept to a minimum (no more than 2 to 4 days), and the continuation of normal activities within the limits permitted by pain should be encouraged. Mild analgesics (including NSAIDs) and muscle relaxants may be useful for early symptom control; opiates should be used sparingly.[16] Spinal manipulation, massage or other physical therapy, graded activity, self-management, or other specific exercise programs may be effective in acute back pain, but most controlled studies suggest little to no advantage of any particular regimen.[17-21] Patient education about the natural history of back pain may result in fewer demands for further diagnostic tests and physician visits and should improve patient satisfaction.

CHRONIC BACK PAIN

Diagnosis

Patients whose pain persists after 4 to 6 weeks of conservative treatment measures should be reassessed. Plain radiography and basic laboratory studies (e.g., complete blood count, erythrocyte sedimentation rate or C-reactive protein level, chemistry profile, and urinalysis) should be considered to screen for systemic illnesses. A herniated lumbar disk should be considered in patients with symptoms of radiculopathy, as suggested by pain radiating down the leg with symptoms reproduced by straight leg raising. MRI may be necessary to confirm a herniated disk, but findings should be interpreted with caution because many asymptomatic persons have disk abnormalities.[22] Electromyography may be useful in differentiating lumbar radiculopathy from other causes of radicular leg pain. Most lumbar disk herniations producing sciatica occur at the L4-L5 and L5-S1 levels. Surgical intervention is indicated in patients with persistent sciatica and clear-cut evidence of a herniated disk on MRI or myelography–CT scanning.

Treatment

A number of nonpharmacologic approaches to therapy have been studied in patients with chronic low back pain, but no specific regimen appears to be superior to others. Most analyses of prior studies suggest that individualized programs promoting strength and range of motion have modest efficacy.[23-25] Behavioral therapies and using a mattress of medium firmness may provide short-term improvement in some patients, as well.[26,27] Judicious use of analgesics, NSAIDs, and tricyclic antidepressants may help the patient function more fully and may improve outcome.[28] Local therapies in the form of local injections of intradiscal or periarticular steroids, hypertonic irritant solutions (prolotherapy), or acupuncture are no more effective than placebo in most studies.[29-31]

Lumbar Stenosis

Lumbar spinal stenosis, usually a result of extensive degenerative disk disease and osteophytes, should be suspected in elderly patients with chronic back pain associated with sciatica.[32] Patients typically complain of pain, numbness, and weakness in the buttocks that extends to one or both legs. Symptoms are usually brought on by standing or walking and improve when the patient stoops forward, sits, or lies down (i.e., neurogenic claudication or pseudoclaudication). The diagnosis may be confirmed by MRI or myelogram–CT scanning.[33] Conservative measures are helpful in many patients, allowing stable or improved function. Surgical decompression by multilevel laminectomy and fusion should be considered in patients with poor response to conservative measures or progressive functional deterioration.[34-36]

Shoulder Pain

Shoulder pain is one of the most common musculoskeletal problems seen in the outpatient setting.[37] Most shoulder pain results from conditions of the periarticular structures of the joint; true arthritis of the glenohumeral joint is uncommon [*see Figure 1*].

The initial evaluation of shoulder pain should include consideration of pain that may be referred from the neck, thorax, or abdomen. The examination should assess active and passive range of flexion, abduction, and internal and external rotation of the shoulder, along with forward elevation. In addition, areas of localized tenderness may help differentiate the various potential causes of shoulder pain. Plain radiographs are seldom diagnostic but are indicated in patients with a history of trauma or refractory pain or when true glenohumeral joint disease is suspected. For patients who respond poorly to conservative therapy, a variety of specialized tests (e.g., arthrography, arthroscopy, and MRI) are available for further definition of lesions that may require surgery.

ROTATOR CUFF TENDINITIS (IMPINGEMENT SYNDROME)

Rotator cuff tendinitis, or impingement syndrome, is often associated with bursitis of the overlying subacromial bursa and is the cause of most nontraumatic cases of shoulder pain. Rotator cuff tendinitis results from inflammation, degeneration, and attrition of the rotator cuff by mechanical impingement on the acromion, coracoacromial ligament, and sometimes the acromioclavicular joint.

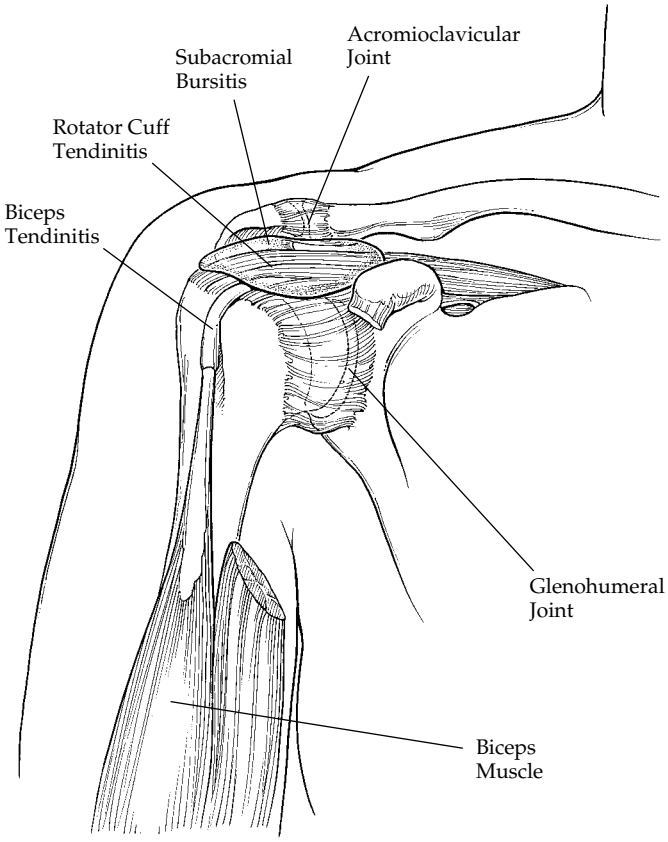


Figure 1 **Tendinitis of the rotator cuff and subacromial bursitis cause pain that is felt over the lateral aspect of the shoulder, whereas bicipital tendinitis, acromioclavicular joint disease, and glenohumeral joint disease cause anterior shoulder pain.**

Rotator cuff tendinitis is seen most commonly in patients 35 to 60 years of age and appears to be partially related to work or activities performed with the arms in highly elevated positions.[38] Younger patients may be affected as a result of athletic activities involving overhand throwing. Patients report an insidious pain that may be diffuse over the lateral deltoid or more localized to the anterior acromial region. Pain worsens with reaching and may be accompanied by a catch as the patient brings the arm into an overhead position. Rotator cuff pain is often particularly bothersome at night and interferes with sleep. On examination, pain may limit movement and may be reproduced by resistance of active movement. The so-called impingement sign is elicited by forced forward elevation of the arm with the scapula stabilized from behind. A coexistent rotator cuff tear may be suspected if the patient cannot hold the arm in a horizontal position against gravity.

The goal of therapy for rotator cuff tendinitis is to relieve pain and maintain or restore range of motion. Treatment should begin with rest and a progressive program of stretching and strengthening exercises, facilitated by an NSAID.[39] Injection of glucocorticoids and local anesthetic into the subacromial space or glenohumeral joint may result in dramatic relief of symptoms and may allow a more rapid, full recovery in some patients.[40,41] Avoidance of repetitive overhead activities of the arms is necessary during recovery, and job modification may be needed to prevent recurrence. In refractory cases, surgical division of the coracoacromial ligament or acromioplasty may be indicated.

CALCIFIC TENDINITIS

Calcific tendinitis is the cause of pain in a subset of patients with apparent rotator cuff disease. In most cases, a more chronic tendinitis is implicated, with associated deposition of calcium in the rotator cuff; calcification in the subacromial space is apparent radiographically. Patients usually have a more acutely painful condition, similar to that seen in crystal-induced arthritis. NSAIDs and local glucocorticoid injections are usually helpful, and surgery is indicated in selected cases. Ultrasound therapy has been shown to provide short-term improvement in symptoms and radiographic signs of calcification, as compared with placebo.[42]

BICIPITAL TENDINITIS

Bicipital tendinitis occurs in the region of the anterior shoulder, where the long head of the biceps tendon passes through the bicipital groove of the humerus and through the joint to insert over the glenoid cavity. Diagnosis is based on the localization of tenderness anteriorly, though this condition may coexist with rotator cuff tendinitis. Rupture of the tendon may occur occasionally, particularly in older patients, and often presents as a bulge in the biceps muscle. Treatment with local measures and range-of-motion exercises is effective, as in rotator cuff disease. Surgical repair of a ruptured tendon is indicated only in younger patients with acute rupture.

FROZEN SHOULDER (ADHESIVE CAPSULITIS)

Frozen shoulder, or adhesive capsulitis, is characterized by progressive pain and global loss of motion in the shoulder. This condition is usually seen in patients with an underlying rotator cuff tendinitis or bicipital tendinitis but has also been associated with stroke, myocardial infarction, cervical radiculopathy, and pulmonary disease. The pathophysiology of frozen shoulder is unclear, and controversy exists as to how significantly capsular inflammation or fibrosis really contributes to the loss of motion that is characteristic of the condition. Treatment is directed toward pain relief and restoration of function, often with a combination of exercises, local heat, ultrasonography, and NSAIDs or mild analgesic medications. Maximal rehabilitation of a frozen shoulder often requires 1 to 2 years. Suprascapular nerve blockade, short courses of systemic steroids, and joint distention with intra-articular steroids have been shown to provide short-term benefit in some studies.[43-45] Manipulative therapy or intra-articular steroid injections combined with physical therapy may provide improvement over longer periods.[46,47] Surgical procedures and capsular distention with saline injection have reportedly proved useful in individual cases.

MYOFASCIAL SHOULDER PAIN SYNDROME

Myofascial shoulder pain syndromes are characterized by pain over the trapezius or medial or lateral scapular borders posteriorly, with the finding of reproducible trigger points. These poorly characterized syndromes usually respond to local injection with glucocorticoids and an anesthetic, though local modalities may be needed in more chronic cases.

Chest Wall Pain

Musculoskeletal chest wall pain syndromes account for about 10% to 15% of cases in which adults are seen for chest pain in the emergency department setting, and they account for about 15% to 20% of patients who have had chest pain but whose coronary angiograms are negative.[48] The diagnosis of musculoskeletal

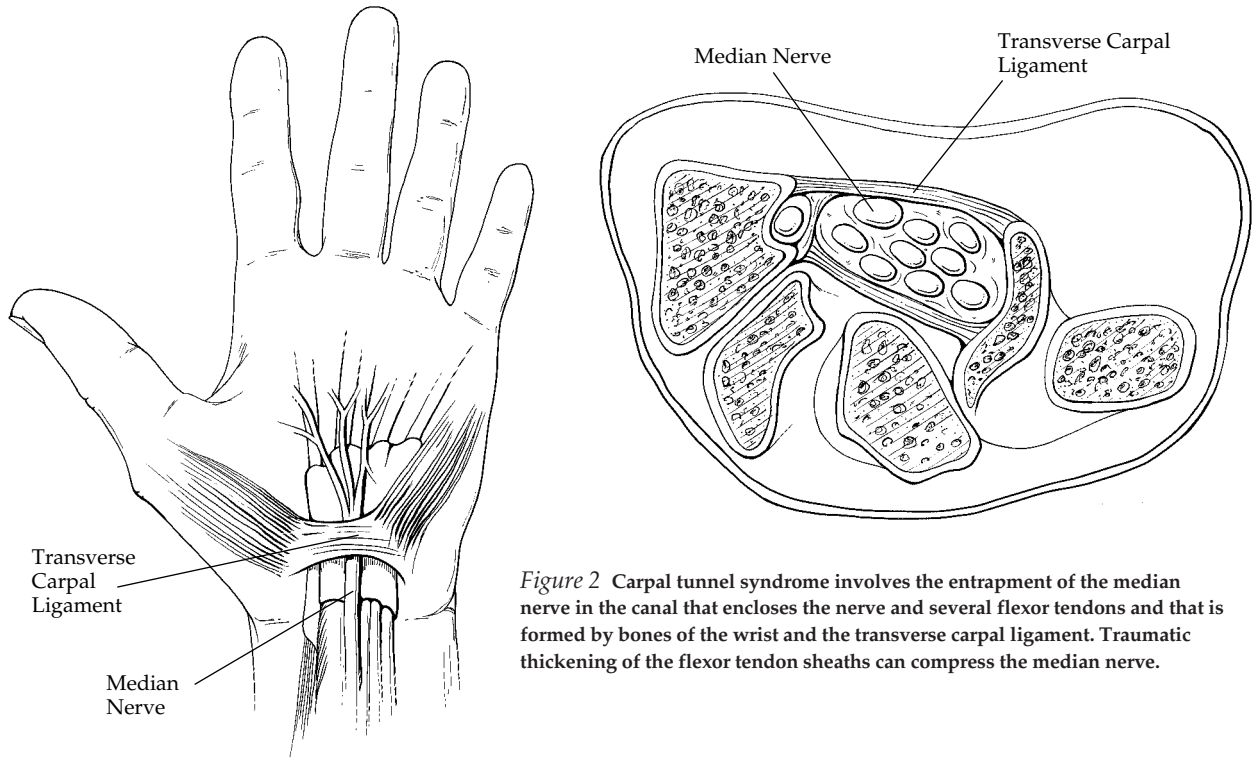


Figure 2 **Carpal tunnel syndrome involves the entrapment of the median nerve in the canal that encloses the nerve and several flexor tendons and that is formed by bones of the wrist and the transverse carpal ligament. Traumatic thickening of the flexor tendon sheaths can compress the median nerve.**

chest wall pain requires the finding of consistent areas of tenderness that reproduce the patient's pain. In rare cases, chest pain may result from Tietze syndrome—a benign, painful, nonsuppurative localized swelling of the costosternal, sternoclavicular, or costochondral joints, most often involving the area of the second and third ribs. In most cases, only one area is involved. Young adults are more commonly affected.

More often, patients with musculoskeletal chest wall syndromes have a more diffuse pain syndrome, termed costochondritis or costosternal syndrome, the specific etiology of which is not well understood. Areas of tenderness are not accompanied by warmth, erythema, or swelling; multiple areas of tenderness are found, usually in the upper costochondral or costosternal junctions. A number of less common chest wall syndromes have been described, each defined by the area of tenderness (e.g., xiphoidalgia, sternalis syndrome, and slipping rib syndrome). Musculoskeletal chest wall syndromes are usually self-limited and respond to analgesics, local heat, stretching exercises, and local glucocorticoid injection.

Elbow Pain

The most common nonarticular syndromes of the elbow include epicondylitis, olecranon bursitis, and ulnar nerve entrapment.

EPICONDYLITIS

Epicondylitis is caused by an inflammation at the origin of the tendons and muscles serving the forearm; it is usually caused by overuse or by repetitive activity. Patients typically complain of elbow and forearm pain with activity. When the extensor muscles are involved (i.e., tennis elbow), tenderness is maximal over the lateral epicondyle and aggravated by extension of the wrist against resistance. A similar, less common process may affect the flexor muscles originating at the medial epicondyle (i.e., golfer's elbow).

Epicondylitis usually responds to rest, local heat or ice, NSAIDs, and forearm support to reduce tension at the epicondyle. Local infiltration of glucocorticoids and lidocaine often results in more rapid improvement than other measures in the first month or two but does not appear to affect the outcome over 6 to 12 months.[49,50]

OLECRANON BURSITIS

Olecranon bursitis presents as a discrete swelling with palpable fluid over the tip of the elbow. Traumatic bursitis is characterized by minimal heat or surrounding erythema. The fluid aspirated is noninflammatory and often contains multiple red cells. Infectious bursitis—usually caused by gram-positive skin organisms—is accompanied by heat, erythema, and induration. When infection is suspected, prompt aspiration and culture of the fluid are mandatory. Antibiotics should be started empirically, and the bursa should be reaspirated frequently until the fluid no longer reaccumulates and cultures are negative.[51] Olecranon bursitis may also be part of rheumatoid arthritis or gout, usually in a patient in whom a diagnosis has already been made. On occasion, an initial diagnosis of gout is made by examination of bursal fluid for urate crystals.

ULNAR NERVE ENTRAPMENT

Ulnar nerve entrapment is caused by compression of the ulnar nerve as it passes through the ulnar groove at the elbow[52] [*see 11:II Diseases of the Peripheral Nervous System*]. Patients typically complain of pain and numbness that radiates from the elbow to the little finger and the medial side of the hand. An increase in paresthesia with elbow flexion is helpful in making the diagnosis, but nerve conduction studies are often needed to confirm the diagnosis. Conservative therapy with a loose cast may help limit elbow flexion and improve symptoms in some patients; surgical decompression is indicated in patients with disabling pain or weakness.

Hand and Wrist Pain

Painful conditions of the tendons and tendon sheaths of the hand and wrist are often related to repetitive or unaccustomed activities. The resultant edema, inflammation, and fibrosis of the structures interfere with the normal function of the tendon as it moves within the sheath.

DE QUERVAIN TENOSYNOVITIS AND FLEXOR TENOSYNOVITIS

De Quervain tenosynovitis affects the abductor pollicis longus and extensor pollicis brevis. Typical symptoms are pain over the radial aspect of the wrist during activities and tenderness that is usually found over the affected tendons proximal to the level of the carpometacarpal joint of the thumb. Pain is reproduced by stretching the tendons with the thumb inside a closed fist (i.e., the Finkelstein maneuver). Flexor tenosynovitis, or trigger finger, is caused by involvement of the flexor tendons of the digits, usually at the level of the metacarpophalangeal joint. Patients complain of locking of the affected digit in a flexed position, often with a sudden painful release on extension. Treatment of de Quervain tenosynovitis and flexor tenosynovitis may require rest, local heat, immobilization with a splint, or local infiltration with glucocorticoids. Surgical release is rarely required.

CARPAL TUNNEL SYNDROME

Carpal tunnel syndrome is caused by compression of the median nerve at the wrist as it courses with the flexor tendons[53] [*see Figure 2 and 11:II Diseases of the Peripheral Nervous System*]. Entrapment is usually associated with flexor tenosynovitis related to overuse, vibratory tool use, or trauma.[54] In addition, an association has been observed with cigarette smoking, body weight, and such medical conditions as diabetes mellitus, rheumatoid arthritis, pregnancy, and hypothyroidism, as well as with rare conditions such as amyloidosis, acromegaly, and localized infection.

Carpal tunnel syndrome is relatively common in the general population. A 1999 study found that 14% of the general population has symptoms suggestive of carpal tunnel syndrome; such symptoms were confirmed by clinical examination and electrophysiologic studies in 2% to 3% of the patients studied.[55] In addition, 18% of asymptomatic persons were found to have electrophysiologic evidence of median nerve entrapment. Carpal tunnel syndrome is more common in persons with occupations that require repetitive wrist movements, awkward wrist positions, or the use of vibrating tools or great force. Patients report numbness, tingling, and pain over the palmar radial aspect of the hand; these symptoms are often worse at night or after use. Reproduction of paresthesia with maximal wrist flexion (i.e., the Phalen test) or tapping over the volar aspect of the wrist (i.e., the Tinel sign) are often considered to be helpful clinical findings. However, a review of published studies suggests that the pattern of pain and findings of decreased sensation and weakness of thumb abduction are the most reliable diagnostic findings.[56] Because of the uncertainties in the reliability of diagnostic findings, electrodiagnostic testing is usually necessary to confirm a diagnosis, particularly when surgical intervention is considered.

Conservative treatment measures include use of NSAIDs and placement of a wrist splint in a neutral position. Local injection of glucocorticoids affords short-term relief in most patients, but long-term improvement is less predictable.[57,58] Surgical decompression by sectioning of the volar carpal ligament results in excellent outcome in 67% to 80% of patients; it is indicated in patients whose conditions respond poorly to conservative therapy, patients with chronic or recurrent symptoms, or patients with weakness or atrophy of the thenar muscles.[59] Predictors of less favorable surgical outcome include poor upper extremity function; alcohol use; worse mental health status; and, in workers, involvement of an attorney.[60]

DUPUYTREN CONTRACTURE

Dupuytren contracture is a fibrosing condition of the palmar and digital fascia that results in thickening and puckering of the palmar skin with subcutaneous nodules and often in flexion contracture of the underlying digit. Dupuytren contracture may be associated with other fibrosing syndromes; with an autosomal dominant inheritance pattern; and possibly with liver disease, epilepsy, and alcoholism. Although spontaneous improvement may be seen, surgical intervention to improve function may be useful in individual cases.

STIFF-HAND SYNDROME

The stiff-hand syndrome, resembling scleroderma, is characterized by thickening of the skin and subcutaneous tissues and generalized limitation of hand and wrist motion. This condition is seen almost exclusively in young patients with long-standing type 1 diabetes mellitus.[61]

Hip Girdle Pain

Pain around the hip girdle is a common complaint in clinical practice. Patients with pain resulting from diseases of the hip joint usually describe pain in the anterior thigh or inguinal region that worsens with weight bearing. More commonly, patients with a chief complaint of hip pain have a problem in one of the nonarticular structures of the hip girdle, usually located posteriorly or laterally [*see Table 2*]. A multitude of bursae have been described in the hip girdle region. Pain in the upper buttock in and around the gluteal muscles is often referred to as myofascial hip pain or gluteal bursitis. Pain in this area is often difficult to differentiate from referred lumbar pain. Local therapy with heat,

Table 2 Differential Diagnosis of Hip Girdle Pain

Clinical Syndrome	*Location of Pain*	*Diagnostic Features and Comments*
Acetabular joint pain	Anterior hip (inguinal)	Worse with weight bearing Radiographic confirmation
Ileopectineal bursitis	Anterior hip (inguinal)	Pain with extension Normal radiograph ? Ultrasound or CT scanning
Meralgia paresthetica	Anterior hip (midthigh)	Numbness and tingling Normal hip movement
Trochanteric bursitis	Lateral hip, posterior hip, or both	Normal hip movement Point tenderness Relief with glucocorticoid injection
Myofascial pain	Posterior hip	Localized tenderness Relief with glucocorticoid injection ? Mimics lumbar disease
Gluteal bursitis	Posterior hip	Localized tenderness Relief with glucocorticoid injection ? Mimics lumbar disease
Ischiogluteal bursitis	Posterior hip	Normal hip movement Point tenderness

stretching, or glucocorticoid injection is usually helpful, but many patients require long-term therapy.

TROCHANTERIC BURSITIS (GREATER TROCHANTERIC PAIN SYNDROME)

Trochanteric bursitis is probably the most common cause of hip girdle pain, although a study using MRI suggests that most patients with this pain syndrome may have tendinitis or a partial tear of the gluteus medius tendon.[62] Patients typically complain of pain over the lateral aspect of the hip girdle, sometimes radiating down the thigh, that is worse at night when they lie on the affected side. Pain is sometimes present when the patient arises from a chair, but it tends to improve with ambulation. Point tenderness over the lateral or posterior aspect of the greater trochanter is usually diagnostic, though some patients with referred lumbar facet or disk disease may have a similar presentation. Patients with more severe pain may have a positive Trendelenburg sign on physical examination. Local heat and NSAIDs may be helpful, and a local glucocorticoid injection is curative in most patients. In refractory cases, repeated injections, physical therapy, and, in rare instances, surgical excision of the bursa may be indicated.

ISCHIOGLUTEAL BURSITIS

Ischiogluteal bursitis results from an irritation of the bursa in the area of the attachments of the hamstring and gluteal muscles at the ischial tuberosity. The condition may be brought on by prolonged sitting or by pressure in the area and usually responds to local heat, stretching, or glucocorticoid injection.

ILIOPECTINEAL BURSITIS

Iliopectineal bursitis, which is caused by irritation of the bursa between the iliopsoas muscle and the inguinal ligament, is an uncommon cause of inguinal pain and may mimic true hip joint disease. The diagnosis is suggested by the presence of inguinal pain that is aggravated by extension of the hip (in a patient whose hip x-ray is normal). Confirmation by ultrasonography or CT scanning may be required. Treatment is usually with local measures or, in rare cases, by means of surgical excision.

MERALGIA PARESTHETICA

Meralgia paresthetica is characterized by intermittent paresthesia, hypoesthesia, or hyperesthesia over the upper anterolateral thigh. The syndrome is caused by an entrapment of the lateral femoral cutaneous nerve at the level of the anterosuperior iliac spine where the nerve passes through the lateral end of the inguinal ligament. Causes include local trauma, rapid weight gain, and the wearing of constrictive garments around the hips. Useful therapies include avoidance of pressure in the area, weight loss, and local infiltration of glucocorticoids at the level of the nerve exit.

Knee and Lower Leg Pain

Clinically, the differentiation of articular knee pain from nonarticular pain can be difficult. Most patients with articular knee pain have a relatively diffuse pain that is not well localized to one area of the knee. Physical examination shows loss of motion, crepitus (in osteoarthritis), warmth (in inflammatory arthritis), or the presence of effusion. If knee pain is localized or if the knee has full range of motion without warmth, crepitus, or effusion, one of the following nonarticular syndromes should be considered: infrapatellar tendinitis, Osgood-Schlatter disease, prepatellar bursitis, anserine bursitis, anterior knee pain syndromes, and restless legs syndrome.

INFRAPATELLAR TENDINITIS

Infrapatellar tendinitis, or jumper's knee, causes anterior knee pain below the patella and is often related to athletic activities. Tenderness is localized to the infrapatellar tendon, with no associated swelling, and conservative measures almost always result in resolution of symptoms.

OSGOOD-SCHLATTER DISEASE

Osgood-Schlatter disease is characterized by pain and swelling over the tibial tubercle at the tendon insertion point. This condition is seen predominantly in adolescent males and is thought to represent a traumatic avulsion injury. Symptoms usually resolve with temporary immobilization and slow resumption of activities.

PREPATELLAR BURSITIS

Prepatellar bursitis, or housemaid's knee, causes pain and swelling in the anterior knee superficial to the patella and infrapatellar tendon. An area of localized fluid collection is usually detectable; aspiration is often needed for diagnosis. As in olecranon bursitis of the elbow, prepatellar bursitis may be associated with trauma, localized bacterial infection, and, less commonly, gout, rheumatoid arthritis, and atypical infections. The differentiation between trauma and infection is particularly important for initiation of appropriate therapy.

ANSERINE BURSITIS

Anserine bursitis, which is caused by irritation of the bursa near the attachment of the sartorius and hamstring muscles at the medial tibial condyle, is a common cause of medial knee pain. Patients with this condition complain of pain at night or when climbing stairs, and an area of localized tenderness can be found on examination. Coexistent osteoarthritis of the knee joint is present in many patients, and relief with local heat or injection of glucocorticoids and anesthetic may be helpful both diagnostically and therapeutically.

ANTERIOR KNEE PAIN SYNDROMES

Anterior knee (patellofemoral) pain syndromes usually manifest themselves as pain and crepitus associated with activities that require knee flexion under load conditions (e.g., stair climbing). Physical findings that help with diagnosis include (1) reproduction of pain with pressure over the patella during knee motion and (2) tenderness over the medial surface of the patella. The cause of most anterior knee pain syndromes is uncertain, but the pain may be related to misalignment of the quadriceps with lateral patellar subluxation, patella alta, hypermobility, or findings of chondromalacia of the patella on arthroscopic evaluation. Local measures and an exercise program that emphasizes isometric quadriceps strengthening is helpful in most patients. Some patients require arthroscopic intervention to diagnose and correct articular irregularities or patellar misalignment.

Ankle and Foot Pain

Nonarticular foot and ankle pain is best approached with a consideration of the region affected: the forefoot, midfoot, or hindfoot [*see Figure 3*].

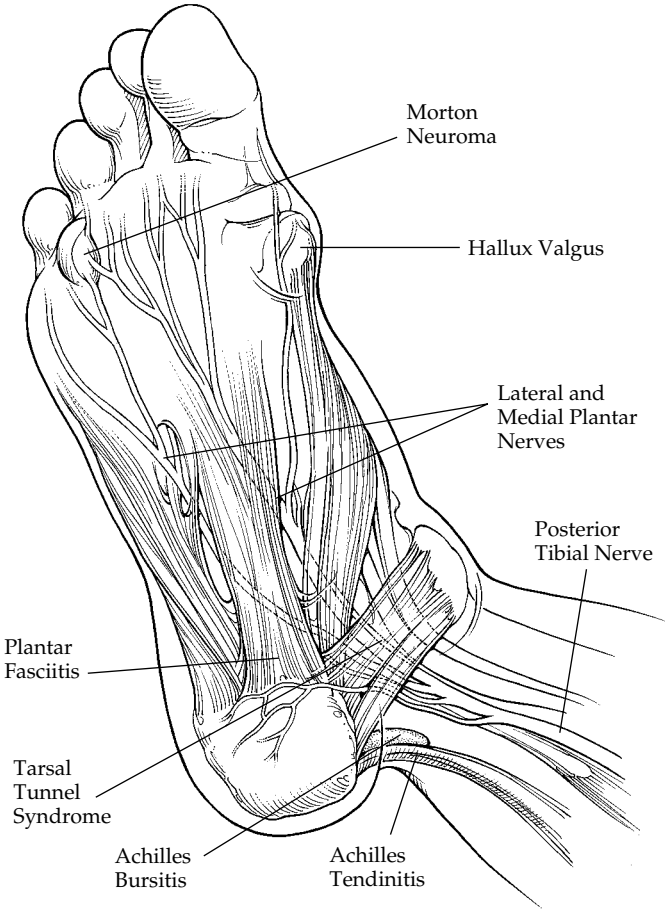


Figure 3 **In the anterior foot, hallux valgus may cause diffuse pain, whereas Morton neuroma is usually localized. Tarsal tunnel syndrome causes paresthesias over the medial and plantar aspects. Plantar fasciitis and Achilles tendinitis are common causes of posterior foot pain.**

FOREFOOT PAIN

Hallux valgus is the leading cause of forefoot pain. It is a common deformity that causes pain because of direct pressure over the first metatarsophalangeal joint resulting from footwear or because of pressure over the lateral toe joints caused by crowding of the toes. In the lateral toes, hammer toe (i.e., plantar flexion of the proximal interphalangeal joint), claw toe (i.e., plantar flexion of the proximal and distal interphalangeal joints), or mallet toe (i.e., isolated flexion contracture of the distal interphalangeal joint) may be associated with a dorsiflexion contracture of the metatarsophalangeal joint. Initial treatment of these problems should begin with adequate footwear that allows ample width for the metatarsal heads, individualized orthoses, and surgical correction (reserved for patients with persistent pain). Morton neuroma is an entrapment neuropathy of the interdigital nerve, with or without an associated plantar neuroma, that is most commonly seen between the third and fourth metatarsal heads. Patients report pain and paresthesia radiating into the affected toes; tenderness between the metatarsal heads that reproduces the described symptoms will also be found. Orthoses to decrease pressure in the area, local glucocorticoid injection, or surgical excision of the neuroma may be needed to relieve symptoms.

MIDFOOT PAIN

Midfoot pain is usually the result of deformities of the arch of the foot or arthritic changes of the midfoot joints. Patients with a cavus foot deformity, peripheral neuropathies, or previous ligamentous injuries from sprains may be predisposed to excessive stresses on the midfoot and early osteoarthritic changes. Tarsal tunnel syndrome is caused by entrapment of the posterior tibial nerve under the flexor retinaculum on the medial side of the ankle. Symptoms of pain and paresthesia over the plantar and distal foot and toes are usually present, and the Tinel sign may be positive. Tarsal tunnel syndrome is much less common and more difficult to diagnose than carpal tunnel syndrome in the wrist. Treatment consists of splinting and NSAIDs. Local glucocorticoid injection and surgical decompression are not as predictably successful as in carpal tunnel syndrome.

HINDFOOT PAIN

Plantar fasciitis is one of the most common causes of hindfoot pain, accounting for up to 15% of all foot symptoms requiring medical attention.[63] The condition is more common in runners and military personnel; other risk factors include obesity, prolonged standing, pes planus, and reduced ankle motion. Patients report pain over the plantar aspect of the heel and midfoot that worsens with walking. Although radiographic spurs in the affected area are common, they may also be seen in asymptomatic persons and are therefore not diagnostic. A majority of patients (approximately 80%) improve regardless of therapy within 12 months of onset. Orthoses, night splints, plantar and heel cord stretching exercises, NSAIDs, and local glucocorticoid injection may be helpful, but few studies have shown conclusive results with any single therapy. Surgical release may be useful for carefully selected patients with poor response to conservative measures. Posterior heel pain is usually caused by Achilles tendinitis or by bursitis of the bursae that lie superficial or deep to the insertion of the Achilles tendon at the calcaneus. Although usually associated with running and other sports activities, Achilles tendinitis may also be part of ankylosing spondylitis and Reiter syndrome and has been reported in association with fluoroquinolone therapy and treatment with systemic or local steroids.[64,65] NSAIDs and orthoses designed to reduce stress on the tendon (e.g., heel lifts) are usually helpful. In most cases, glucocorticoid injections in the Achilles tendon area should be avoided because of the risk of tendon rupture.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Sheon RP, Moskowitz RW, Goldberg VM: Soft Tissue Rheumatic Pain: Recognition, Management, Prevention, 3rd ed. Lea & Febiger, Philadelphia, 1996

2. Posner J, Glew C: Neck pain. Ann Intern Med 136:758, 2002

3. Eck JC, Hodges SD, Humphreys SC: Whiplash: a review of a commonly misunderstood injury. Am J Med 110:651, 2001

4. Cassidy JD, Carroll LJ, Cote P, et al: Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash injury. N Engl J Med 342:1179, 2000

5. Hoving JL, Koes BW, de Vet HC, et al: Manual therapy, physical therapy, or continued care by a general practitioner for patients with neck pain: a randomized, controlled trial. Ann Intern Med 136:713, 2002

6. Ylinen J, Takala EP, Nykanene M, et al: Active neck training in the treatment of chronic neck pain in women: a randomized controlled trial. JAMA 289: 2509, 2003

7. White P, Lewith G, Prescott P, et al: Acupuncture versus placebo for the treatment of chronic mechanical neck pain: a randomized, controlled trial. Ann Intern Med 141:911, 2004

8. Deyo RA, Weinstein JN: Low back pain. N Engl J Med 344:363, 2001

9. Carragee EJ: Clinical practice: persistent low back pain. N Engl J Med 352:1891, 2005

10. Manek N: Epidemiology of back disorders: prevalence, risk factors, and prognosis. Curr Opin Rheumatol 17:134, 2005

11. Coste J, Lefrancois G, Guillemin F, et al: Prognosis and quality of life in patients with acute low back pain: insights from a comprehensive inception cohort study. Arthritis Rheum 51:168, 2004

12. Carragee EJ, Alamin TF, Miller JL, et al: Discographic, MRI, and psychosocial determinants of low back pain disability and remission: a prospective study in subjects with benign persistent back pain. Spine J 5:24, 2005

13. Jarvik JG, Deyo RA: Diagnostic evaluation of low back pain with emphasis on imaging. Ann Intern Med 137:586, 2002

14. Jarvik JG, Hollingsworth W, Martin B, et al: Rapid magnetic resonance imaging vs radiographs for patients with low back pain: a randomized controlled trial. JAMA 289: 2810, 2003

15. Gilbert FJ, Grant AM, Gillan MG, et al: Low back pain: influence of early MRI imaging or CT on treatment and outcome—a multicenter randomized trial. Radiology 231:343, 2004

16. van Tulder MW, Touray T, Furlan AD, et al: Muscle relaxants for nonspecific low back pain: a systematic review within the framework of the Cochrane collaboration. Spine 28:1978, 2003

17. Assendelft WJ, Morton SC, Yu EI, et al: Spinal manipulative therapy for low back pain: a meta-analysis of effectiveness relative to other therapies. Ann Intern Med 138: 871, 2003

18. Cherkin DC, Sherman KJ, Deyo RA, et al: A review of the evidence for the effectiveness, safety, and cost of acupuncture, massage therapy, and spinal manipulation for back pain. Ann Intern Med 138:898, 2003

19. Hayden JA, van Tulder MW, Malmivaara AV, et al: Meta-analysis: exercise therapy for nonspecific low back pain. Ann Intern Med 142:765, 2005

20. Staal JB, Hlobil H, Twisk JW, et al: Graded activity for low back pain in occupational health care: a randomized, controlled trial. Ann Intern Med 140:77, 2004

21. Damush TM, Weinberger M, Perkins SM, et al: The long-term effects of a self-management program for inner-city primary care patients with acute low back pain. Arch Intern Med 163:2632, 2003

22. Kjaer P, Leboeuf-Yde C, Korsholm L, et al: Magnetic resonance imaging and low back pain in adults: a diagnostic imaging study of 40-year-old men and women. Spine 30:1173, 2005

23. Hayden JA, van Tulder MW, Tomlinson G: Systematic review: strategies for using exercise therapy to improve outcomes in low back pain. Ann Intern Med 142:776, 2005

24. Niemisto L, Lahtinen-Suopanki T, Rissanen P, et al: A randomized trial of combined manipulation, stabilizing exercises, and physician consultation compared to physician consultation alone for chronic low back pain. Spine 28:2185, 2003

25. Schonstein E, Kenny DT, Keating J, et al: Work conditioning, work hardening, and functional restoration for workers with back and neck pain. Cochrane Database Syst Rev (1):CD001822, 2003

26. Ostelo R, Tulder M, Vlaeyen J, et al: Behavioural treatment for chronic low back pain. Cochrane Database Syst Rev (1):CD002014, 2005

27. Kovacs FM, Abraira V, Pena A, et al: Effect of firmness of mattress on chronic nonspecific low-back pain: randomized, double-blind, controlled, multicentre trial. Lancet 362:1599, 2003

28. Salerno SM, Browning R, Jackson JL: The effect of antidepressant treatment on chronic back pain: a meta-analysis. Arch Intern Med 162:19, 2002

29. Khot A, Bowditch M, Powell J, et al: The use of intradiscal steroid therapy for lumbar spinal discogenic pain: a randomized controlled trial. Spine 29:833, 2004

30. Yelland MJ, Del Mar C, Pirozzo S, et al: Prolotherapy injections for chronic low back pain: a systematic review. Spine 29:2126, 2004

31. Manheimer E, White A, Berman B, et al: Meta-analysis: acupuncture for low back pain. Ann Intern Med 142:651, 2005

32. Arbit E, Pannullo S: Lumbar stenosis: a clinical review. Clin Orthop 384:137, 2001

33. Saint-Louis LA: Lumbar spinal stenosis assessment with computed tomography, magnetic resonance imaging, and myelography. Clin Orthop 384:122, 2001

34. Amundsen T, Weber H, Nordal HJ, et al: Lumbar spinal stenosis: conservative or surgical management? A prospective 10-year study. Spine 25:1424, 2000

35. Atlas SJ, Keller RB, Wu YA, et al: Long-term outcomes of surgical and nonsurgical management of lumbar spinal stenosis: 8 to 10 year results from the Maine lumbar spine study. Spine 30:936, 2005

36. Deyo R, Nachemson A, Mirza SK: Spinal-fusion surgery—the case for restraint. N Engl J Med 350:722, 2004

37. Gomoli AH, Katz JN, Warner JJ, et al: Rotator cuff disorders: recognition and management among patients with shoulder pain. Arthritis Rheum 50:3751, 2004

38. Svendsen SW, Gelineck J, Mathiassen SE, et al: Work above shoulder level and degenerative alterations of the rotator cuff tendons: a magnetic resonance imaging study. Arthritis Rheum 50:3314, 2004

39. Hay EM, Thomas E, Paterson SM, et al: A pragmatic randomized controlled trial of local corticosteroid injection and physiotherapy for the treatment of new episodes of unilateral shoulder pain in primary care. Ann Rheum Dis 62:394, 2003

40. Green S, Buchbinder R, Glazier R, et al: Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. BMJ 316:354, 1998

41. van der Windt DA, Koes BW, Deville W, et al: Effectiveness of corticosteroid injections versus physiotherapy for treatment of painful stiff shoulder in primary care: randomised trial. BMJ 317:1292, 1998

42. Ebenbichler GR, Erdogmus CB, Resch KL, et al: Ultrasound therapy for calcific tendinitis of the shoulder. N Engl J Med 340:1533, 1999

43. Dahan TH, Fortin L, Pelletier M, et al: Double blind randomized clinical trial examining the efficacy of bupivacaine suprascapular nerve blocks in frozen shoulder. J Rheumatol 27:1464, 2000

44. Buchbinder R, Hoving JL, Green S, et al: Short course prednisolone for adhesive capsulitis (frozen shoulder or stiff painful shoulder): a randomized, double blind, placebo controlled trial. Ann Rheum Dis 63:1460, 2004

45. Buchbinder R, Green S, Forbes A, et al: Arthrographic joint distension with saline and steroid improves function and reduces pain in patients with painful stiff shoulder: results of a randomized, double blind, placebo controlled trial. Ann Rheum Dis 63:302, 2004

46. Bergman GJ, Winters JC, Groenier KH, et al: Manipulative therapy in addition to usual medical care for patients with shoulder dysfunction and pain: a randomized, controlled trial. Ann Intern Med 141:432, 2004

47. Carette S, Moffet H, Tardif J, et al: Intraarticular corticosteroids, supervised physiotherapy, or a combination of the two in the treatment of adhesive capsulitis of the shoulder: a placebo-controlled trial. Arthritis Rheum 48:829, 2003

48. Wise CM: Chest wall syndromes. Curr Opin Rheumatol 6:197, 1994

49. Hay EM, Paterson SM, Lewis M, et al: Pragmatic randomised controlled trial of local corticosteroid injection and naproxen for treatment of lateral epicondylitis of elbow in primary care. BMF 319:964, 1999

50. Smidt N, van der Windt DA, Assendelft WJ, et al: Corticosteroid injections, physiotherapy, or a wait-and-see policy for lateral epicondylitis: a randomised controlled trial. Lancet 359:657, 2002

51. Laupland KB, Davies HD: Olecranon septic bursitis managed in an ambulatory setting. The Calgary Home Parenteral Therapy Program Study Group. Clin Invest Med 24:171, 2001

52. Dawson D: Entrapment neuropathies of the upper extremities. N Engl J Med 329:2013, 1993

53. Katz JN, Simmons BP: Clinical practice. Carpal tunnel syndrome. N Engl J Med 346:1807, 2002

54. Nathan PA, Meadows KD, Istvan JA: Predictors of carpal tunnel syndrome: an 11-year study of industrial workers. J Hand Surg (Am) 27:644, 2002

55. Atroshi I, Gummesson C, Johnsson R, et al: Prevalence of carpal tunnel syndrome in a general population. JAMA 282:153, 1999

56. D'Arcy CA, McGee S: The rational clinical examination: does this patient have carpal tunnel syndrome? JAMA 283:3110, 2000

57. O'Gardaigh D, Merry P: Corticosteroid injection for the treatment of carpal tunnel syndrome. Ann Rheum Dis 59:918, 2000

58. Gerritsen AA, de Vet HC, Scholten RJ, et al: Splinting vs surgery in the treatment of carpal tunnel syndrome: a randomized controlled trial. JAMA 288:1245, 2002

59. Ly-Pen D, Andreu JL, deBlas G, et al: Surgical decompression versus local steroid injection in the carpal tunnel syndrome: a one-year, prospective, randomized, open, controlled clinical trial. Arthritis Rheum 52:612, 2005

60. Katz JN, Losina E, Amick BC 3rd, et al: Predictors of outcomes of carpal tunnel release. Arthritis Rheum 44:1184, 2001

61. Jacobs-Kosmin D, DeHoratius RJ: Musculoskeletal manifestations of endocrine disorders. Curr Opin Rheumatol 17:64, 2005

62. Bird PA, Oakley SP, Shnier R, et al: Prospective evaluation of magnetic resonance imaging and physical examination findings in patients with greater trochanteric pain syndrome. Arthritis Rheum 44:2138, 2001

63. Buchbinder R: Clinical practice. Plantar fasciitis. N Engl J Med 350:2159, 2004

64. Paavola M, Kannus P, Jarvinen TA, et al: Achilles tendinopathy. J Bone Joint Surg Am 84:2062, 2002

65. van der Linden PD, Sturkenboom CJM, Herings RMC, et al: Increased risk of Achilles tendon rupture with quinolone antibacterial use, especially in elderly patients taking oral corticosteroids. Arch Intern Med 163:1801, 2003

Acknowledgments

Figures 1 and 3 Susan E. Brust, C.M.I.

Figure 2 Lynn O'Kelley.

XIII FIBROMYALGIA

JOHN BUCKNER WINFIELD, M.D.

Definition

Fibromyalgia is a chronic syndrome that occurs predominantly in women and is marked by generalized pain, multiple defined tender points [*see Figure 1*],[1] fatigue, disturbed or nonrestorative sleep, and numerous other somatic complaints. It is not a discrete disease; rather, it lies at the far end of a continuum of psychological distress and chronic pain in the population. Currently, fibromyalgia is best classified as one of a series of disorders that are variously termed symptom-based conditions,[2] functional somatic syndromes,[3] or affective spectrum disorders.[4] Advances in our understanding of the psychophysiologic dysregulation[5] in these illnesses undoubtedly will lead to revision of such classification in the future. There appear to be discrete subgroups of patients with respect to pain sensitivity and psychological factors,[6,7] and these subgroups vary in response to therapy and in prognosis.[8]

Fibromyalgia largely overlaps with other syndromes with unexplained symptoms, such as chronic fatigue syndrome and irritable bowel syndrome, all of which are related to, but not fully dependent on, depression and anxiety.[9] Fibromyalgia frequently coexists with organically defined disease, such as systemic lupus erythematosus (SLE) or rheumatoid arthritis.

Epidemiology

Otherwise-unexplained widespread pain occurs in about 10% of the general adult population in Western countries, with approximately half of those, mostly women, meeting American College of Rheumatology classification criteria for fibromyalgia.[10]

Fibromyalgia becomes more common after 60 years of age but occurs not infrequently in children. On a typical day, primary care physicians should expect to interact with several patients with fibromyalgia, many of whom will be seeking care for illness other than fibromyalgia. For example, more than 25% of patients with SLE exhibit painful tender points and other clinical and psychological features of fibromyalgia.

Etiology

The cause of fibromyalgia is unknown. Despite extensive research, no definitive organic pathology has been identified. Psychological factors associated with chronic distress appear to be very important.[11] Most researchers believe that pain in fibromyalgia reflects an abnormality of central pain processing, but there is no consensus regarding the mechanism. Nevertheless, current research-based biomedical and sociobehavioral data are now sufficient for a clinically useful understanding of some of the variables that contribute to chronic pain and fatigue in this disorder.

BIOPSYCHOSOCIAL MODEL

At one level, the pain and associated symptoms of fibromyalgia can be viewed according to Engel's biopsychosocial model[12] of chronic illness: health status and outcomes are influenced by the interaction of biologic, psychological, and sociologic variables. Important biologic variables are genetics, gender, sleep, physical condition, stress/neuroendocrine and autonomic dysregulation, and central sensitization to pain. Psychological (cognitive-behavioral) variables contributing to the chronic pain and fatigue of fibromyalgia include pain beliefs/attributions, mood, depression, anxiety, personality traits/disorders, pain behaviors, hypervigilance, coping strategies, and perceived self-efficacy for pain control. Sociologic (environmental and sociocultural) variables consist of experiences influenced by life and culture that have impact on the course of chronic pain and fatigue, such as psychosocial experiences during childhood, family support, work environment, job satisfaction, and ethnologic factors.

Biologic Variables

Genetics Genetic analysis of community-based populations[13] and twins[14] suggests that somatic symptoms, such as pain and fatigue, are etiologically distinct from symptoms of anxiety and depression and reflect, at least in part, the action of genes. The short (S) allele of the promoter region of the serotonin transporter gene *5-HTT* has been associated with depres-

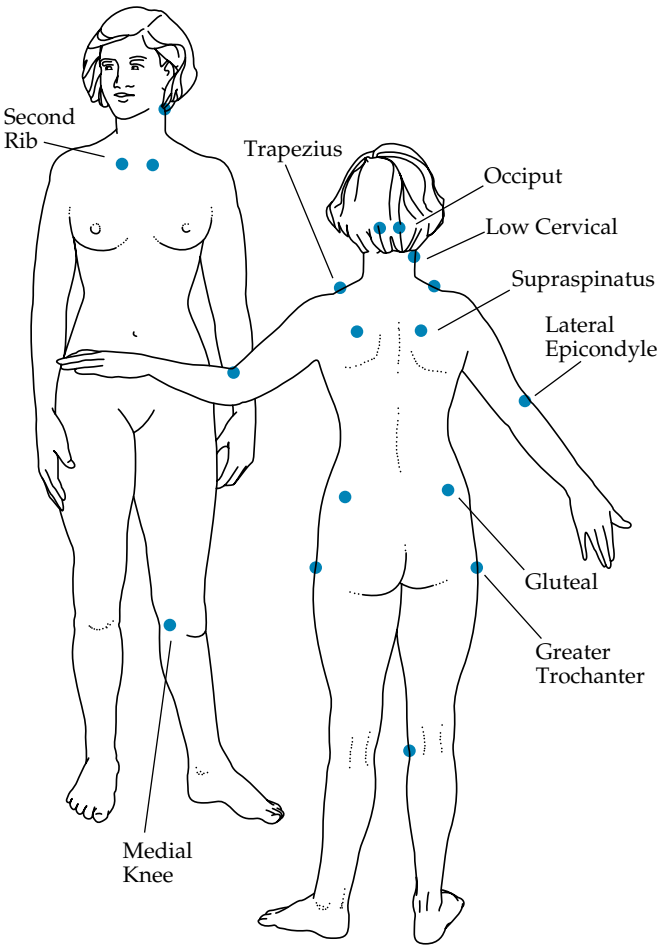


Figure 1 **Tender points in fibromyalgia.**

sion and suicidality in response to stressful life events,[15] and a higher frequency of the S/S genotype has been reported in patients with fibromyalgia than in healthy control subjects.[16] An association with *5-HTT* or its alleles is not evident in fibromyalgia without accompanying depression and psychological distress, however. Other susceptibility genes undoubtedly will be discovered in the near future.

Female gender In general, females exhibit higher pain sensitivity than males and are at greater risk for many pain conditions.[17] Central pain-modulatory systems and inflammatory cytokine levels in females are influenced by phasic alterations in reproductive hormone levels; and females appear to lack diffuse noxious inhibitory control (DNIC), a pain-inhibitory mechanism that is found in normal males.[18] Stressful or aversive stimuli evoke greater sympathetic nervous system, neuroendocrine, and psychological responses in females than in males. Reporting of unexplained symptoms, independent of psychiatric morbidity, also is more common in females.[19] Certain data suggest that the neural circuits activated by emotional experiences and those involved in encoding of emotional experiences into memory show much more overlap in females than in males.[20] Together with the higher prevalence of abuse in women, these biologic variables may explain, at least in part, the increased susceptibility to fibromyalgia in women.

Sleep disturbance Poor sleep is almost universal in fibromyalgia, but fibromyalgia is not primarily a sleep disorder, as had been thought. Pain interferes with sleep, and disturbances in sleep contribute to the experience of pain.[21] In turn, nonrestful sleep and pain underlie the experience of fatigue.

Psychological distress and the stress response system The number of painful tender points in patients with fibromyalgia correlates strongly with their levels of psychological distress.[22] Much evidence suggests that there is an association between chronic, unrelieved psychological stress/distress and functional alterations of the stress-response system (the autonomic nervous system and hypothalamic-pituitary-adrenocortical [HPA] axis) that, in turn, contributes to symptoms of anxiety, pain, and fatigue.[5,23,24] It is unclear, however, whether abnormalities in the stress-response system reflect preexisting vulnerability to fibromyalgia or are a consequence of chronic pain and fatigue.

External triggers Fibromyalgia patients often have fixed beliefs that minor traumatic events, pathogens, chemicals, or other physical agents caused their illness. The available evidence does not support any of those factors as causes, and furthermore, such beliefs can be a barrier to recovery.

Psychological Variables

Cognitive-behavioral variables play a central role in the development and maintenance of persistent pain and functional disability. For example, negative emotions (depression and anxiety), other negative psychological factors (loss of control, unpredictability in one's environment), and certain cognitive aspects (negative beliefs and attributions, catastrophic interpretation of events) can lower pain thresholds and tolerances.[25] Pain behaviors are the actions or expressions by which an individual communicates feelings of pain to the outside world. The response of the outside world (e.g., spouse, physician, employer) then positively or negatively reinforces the pain experience.

Environmental and Sociocultural Variables

Multiple experiences related to forces in the environment, life, and culture can influence the course of chronic pain and fatigue. These influences can be either positive (e.g., good job satisfaction in a patient with work-related back strain) or negative (e.g., a physician's suggestion that a minor traffic accident may have left the patient with long-term damage). In the United States, negative sociocultural elements include the promotion of fear and suggestibility by the media and by society in general, as well as focus on definable causes by patients, physicians, and attorneys. Adverse experiences during childhood, such as poor family environment or childhood sexual abuse, increase susceptibility to the development of chronic pain in adulthood.[26]

Pathophysiology

There are four principal categories of pain: nociceptive, neuropathic, psychogenic, and chronic pain of complex etiology. Nociceptive pain involves stimulation of peripheral pain receptors during inflammation, injury, or destruction of tissues, and it is characterized by a pain experience that corresponds with the noxious stimulus. Neuropathic pain results from direct injury to nerves, such as the radiculopathic pain of degenerative spondylosis. Psychogenic pain occurs in more strictly psychiatric illness, such as somatization disorder or hysteria. Chronic pain of complex etiology is the type of pain characteristic of fibromyalgia, as well as of various regional pain syndromes.

Although chronic pain of complex etiology is as yet incompletely understood, many of the biologic elements operant in the abnormal pain modulation and the associated symptoms seen in fibromyalgia have been identified. Such elements include increased levels of inflammatory cytokines, decreased serotonin levels, increased levels of substance P in the cerebrospinal fluid, deficiency of biogenic amines that normally regulate the release of substance P, glial activation,[27] decreased somatomedin C levels, intrusion of alpha waves into the brain's electrical field during non–random eye movement sleep, and dysregulation of the stress response system. Such dysregulation may manifest as neurally mediated hypoten-

Table 1 Symptoms of Fibromyalgia Syndrome[48]

Musculoskeletal	Pain at multiple sites Stiffness "Hurt all over" Swollen feeling in soft tissues
Nonmusculoskeletal	Fatigue (most times of the day) Morning fatigue (symptom of nonrestorative sleep) Poor sleep Paresthesias
Associated symptoms	Self-assessed anxiety Headaches Dysmenorrhea Irritable bowel syndrome Self-assessed depression Restless legs syndrome Sicca symptoms Raynaud phenomenon Female urethral syndrome

sion; a hypofunctional sympathetic reflex response to stressors; or abnormalities of the HPA axis and other neuroendocrine axes, such as growth hormone secretion in response to exercise.

In fibromyalgia, negative psychological elements constituting stress and distress are a major contributor to the development of increased pain sensitivity and myriad other symptoms. Functional magnetic resonance imaging of the brain during the application of painful pressure in fibromyalgia patients has demonstrated augmented central pain processing.[28] There is no consensus, however, regarding the relative contribution to pain in fibromyalgia of so-called central sensitization (enhanced excitability of dorsal neurons of the spinal cord), dysfunction of descending inhibitory pain control systems (e.g., DNIC), windup (temporal summation of second pain, described as "dull," "aching," or "burning"), thalamic activity, neuroglial activation,[29] and genes.

Diagnosis

CLINICAL MANIFESTATIONS

Pain is the hallmark of fibromyalgia. The pain radiates diffusely from the axial skeleton and is localized to muscles and muscle-tendon junctions of the neck, shoulders, hips, and extremities. Fibromyalgia patients describe the pain with such terms as exhausting, miserable, or unbearable. Generalized hyperalgesia is a cardinal feature. Patients frequently complain that even gentle touch is unpleasant, a condition known as allodynia—pain from normally nonpainful stimuli.

In addition to their persistent widespread pain, fibromyalgia patients experience severe fatigue, insomnia, and low mood or depression [*see Table 1*]. Fatigue occurring most times of the day on most days, together with subjective weakness and nonrestorative sleep, is almost universal. Cognitive complaints, such as difficulties with concentration and memory, may be prominent. Depression, anxiety disorders, and personality disorders contribute to ongoing psychological distress. Other complaints result from somatization, which can be defined as translating psychological distress into somatic symptoms (which are considered more socially acceptable) and seeking care for those symptoms.

Patients with fibromyalgia have a strong tendency toward external attribution of symptoms, with fixed beliefs that minor trauma, viruses (e.g., Epstein-Barr), *Candida,* mold (e.g., so-called black mold), toxic chemicals, or other physical agents (e.g., silicone breast implants) caused their illness. This can be a barrier to recovery, as can ongoing litigation regarding causation of fibromyalgia, disability determination proceedings, or workers' compensation claims.

Functional impairment is usually present, at least in patients with fibromyalgia who seek care. Patients report difficulty doing usual activities of daily living and lack of exercise—indeed, fear and avoidance of exercise.

Regional pain syndromes, such as headache, temporomandibular joint syndrome, or irritable bowel syndrome, are extremely common in fibromyalgia. It is essential that the physician not automatically attribute all such symptoms to fibromyalgia, however, because fibromyalgia frequently coexists with other organically defined disorders. Optimum therapy requires recognition of both fibromyalgia and comorbid disease.

In taking the history, the physician should inquire about sleep quality, ongoing and past stressors, and feelings of anxiety or depression. Recognition of difficulties in these areas is essential in the overall management of chronic pain and fatigue. Onset of fibromyalgia usually antedates clinical diagnosis by years. Exploration of life events surrounding the onset often reveals major stressors, such as breakup of marriage, loss of job, or bankruptcy. The open-ended question, "During your childhood, did you have any bad experiences, such as physical, emotional, or sexual abuse?" not infrequently leads to a catharsis of a significant psychological burden. Current research on the biology of emotion suggests that traumatic experiences during childhood, a period when the brain is still developing, profoundly shape emotions and the subsequent development of functional somatic pain in adulthood. The consequences of abuse are not limited to children; adult domestic violence is an important antecedent of fibromyalgia.[30]

It is valuable to ask whether the patient has previously sought treatment for manifestations of fibromyalgia. Discussion of what treatments have been prescribed and how the patient responded can guide the physician in developing a therapeutic plan. Narcotic use and unsuccessful prior referrals to multidisciplinary pain centers suggest a poor prognosis. Fibromyalgia patients are especially likely to use complementary and alternative medicine (CAM), in part because of the limited efficacy of conventional medical care. Because some CAM agents are not safe and many have the potential to interact with conventional pharmacologic agents, questions about CAM use are an important part of the history.

PHYSICAL EXAMINATION FINDINGS

A patient with uncomplicated fibromyalgia will have normal results on general physical examination. This reassures both the physician and the patient that a significant alternative cause for the symptoms is unlikely. Evidence of synovitis (joint effusion, warmth over joint, pain on joint motion), objective muscle weakness, or other definite physical or neurologic signs suggests the presence of either comorbid disease or an alternative diagnosis. It is essential to identify concomitant painful diseases such as osteoarthritis of individual joints, degenerative spondylosis, bursitis, or other inflammatory soft tissue conditions, because the nociceptive pain from such common problems is amplified in fibromyalgia. Many patients exhibit signs of regional pain syndromes that are often associated with depression or anxiety, such as tenderness in the jaw area (temporomandibular disorder), subtle lower abdominal tenderness to palpation (irritable bowel syndrome), psychomotor slowing (depression), and irritability or hostility (anxiety, panic disorder, or personality disorder).

Tender Points

Eighteen specific tender points have been identified in fibromyalgia [*see Figure 1*].[1] A patient with fibromyalgia will have pain, not just tenderness, on palpation at many of these tender points. Palpation is performed with the thumb, using approximately 4 kg of pressure—about the pressure necessary to blanch the examiner's thumbnail. Attempting to confirm pain at all 18 tender points is not necessary for diagnosis and is often uncomfortable for patients, many of whom find tender-point palpation quite distressing. If an algometer is available, it can provide semiquantitative information regarding pressure pain threshold at tender points. Some fibromyalgia patients complain of pain when pressure is applied anywhere on the body, even relatively insensitive areas such as the forehead or

thumbnail. Unfortunately, in situations of potential secondary gain (e.g., injury litigation, pending disability determination), the physician must be alert for malingering; some of these patients are well informed regarding the location of tender points. Firm pressure over the trapezii or posterior thorax with a stethoscope, rather than a thumb, can provide insight in such cases.

LABORATORY TESTS

There are no specific laboratory test abnormalities in fibromyalgia. Nevertheless, it is appropriate to conduct limited screening for commonly associated disorders and for other diseases that can cause pain and fatigue.

Blood Tests

Useful tests in fibromyalgia include the following: antinuclear antibody (ANA), complete blood count, erythrocyte sedimentation rate (ESR) or C-reactive protein, thyroid-stimulating hormone (TSH), creatine kinase (CK), aspartate aminotransferase, and alanine aminotransferase. Tests for Lyme disease, Epstein-Barr virus infection, and endocrinologic status (e.g., measurement of estrogen, testosterone, growth hormone, and dehydroepiandrosterone) are usually unnecessary.

Other Tests

Urinalysis may be useful. Tests for autonomic dysfunction (tilt-table test), studies of nerve conduction velocity and electromyography, and imaging studies should not be done unless there is a specific indication.

Differential Diagnosis

The differential diagnosis of fibromyalgia is very broad [*see Table 2*]. Nevertheless, extensive laboratory testing or imaging studies should not be done unless the patient has objective indications of other disease on physical or neurologic examination. Major causes of diffuse musculoskeletal pain and key clinical and laboratory features are as follows:

- Fibromyalgia (diffuse pain, tenderness, fatigue, stiffness, tender points, normal laboratory findings)
- Rheumatoid arthritis (symmetrical synovitis, presence of rheumatoid factor, elevated ESR and C-reactive protein)
- SLE (constitutional symptoms, rash, arthralgias, presence of ANA and other autoantibodies)
- Polymyalgia rheumatica (age greater than 50 years, shoulder and hip girdle pain, very high ESR)
- Spondyloarthropathy (morning back pain and stiffness, asymmetrical oligoarthritis, sacroiliitis on pelvic x-rays)
- Inflammatory myopathy (muscle tenderness, objective proximal muscle weakness, elevated serum CK)
- Hypothyroidism (myalgias, weight gain, dry skin, fatigue, cold sensitivity, hyporeflexia, elevated TSH level)
- Osteomalacia (diffuse bone pain and tenderness, proximal myopathy with weakness, low serum phosphate and 25-hydroxyvitamin D levels)

It is usually not helpful clinically to distinguish fibromyalgia from chronic fatigue syndrome or the many regional pain syndromes. Giving patients a name for their illness, however, often enables them to concentrate on getting better, rather than continuously searching for a cause and cure.

Treatment

Four principles govern the treatment of fibromyalgia: (1) validation of distress, (2) diagnostic and therapeutic conservatism, (3) an individualized combination of pharmacologic and nonpharmacologic measures, and (4) care rather than cure. Validation of the patient's symptoms and distress begins with the initial history and physical examination, when the physician explores adverse developmental, social, and behavioral variables and past and current stressors. Failure to provide validation, initially or later in the therapeutic relationship, serves only to perpetuate pain and fatigue and may constitute an insurmountable barrier to treatment.

Strong evidence for therapeutic efficacy in fibromyalgia on the basis of randomized, controlled trials has been difficult to obtain. Confounding factors include the current poor understanding of the syndrome's origin and pathophysiology, the complexity of its symptoms, lack of consensus regarding nosology and clinically meaningful outcome measures, small sample size, short trial duration, and a strong placebo effect because of the close attention patients receive in trials. Much current treatment is empirical and based on proposed, rather than established, models of pathophysiology. Nevertheless, a wealth of published information on treatment of fibromyalgia is now available, including two monographs[31,32] and a series of systematic reviews and meta-analyses of controlled trials.[33-35] Clinical practice guidelines are pending from the American Pain Society[36] The available data suggest that the pain, fatigue, non-

Table 2 Differential Diagnosis of Fibromyalgia

Major Disorder Group	*Selected Specific Disorders*
Rheumatologic	Systemic lupus erythematosus* Rheumatoid arthritis, Sjögren syndrome* Polyarticular osteoarthritis, degenerative spondylosis Polymyalgia rheumatica* Polymyositis, statin myopathy Regional pain syndromes* Osteomalacia Hypermobility syndromes
Neurologic	Carpal tunnel syndrome* Cervical radiculopathy* Metabolic myopathies Multiple sclerosis* Cervical cord compression
Chronic infection	Subacute bacterial endocarditis Brucellosis Hepatitis HIV
Endocrine	Hypothyroidism* Diabetes mellitus type 2 Hyperparathyroidism
Neoplastic	Metastatic (e.g., breast, lung, prostate) Myeloma
Psychiatric	Pain disorder associated with psychological factors* (formerly, somatoform pain disorder) Somatization disorder (hysteria, Briquette syndrome)

*Common alternative diagnosis.

restorative sleep, depression, and anxiety respond to a multifaceted therapeutic approach that combines drug therapy with physical, psychological, and behavioral treatments. An overarching goal of therapy is the promotion of self-efficacy—the patient's firmly held belief that he or she can control symptoms of pain and fatigue.

TREATMENT IN GERIATRIC AND PEDIATRIC CASES

Treatment of diffuse pain in older persons[37] and in children[38] requires special approaches. Pain in older persons is often neglected or ignored. Compared with younger persons, older persons exhibit more physical abnormalities, are more sensitive to opioids, have lower self-efficacy, and use fewer cognitive coping methods. There are many barriers to accurate pain assessment in older persons, including fear of diagnostic tests or medications and reluctance to report pain because pain is an expected part of aging.

The general principles of treatment in the elderly are as follows: discuss goals, hopes, and trade-offs openly; start medications at a low dose and raise the dose slowly; pay attention to timing of medications; be aware of economic barriers; include nonpharmacologic treatment as an integral part of management; and educate the patient and caregiver.

In the pediatric population, unexplained diffuse or localized pain is most common in preadolescent to adolescent girls, often in association with incongruent affect; disproportionate impairment of performance in school; and psychological distress in the child, the family, or both. Key aspects of therapy in this population are discontinuance of all medications, psychological evaluation and psychotherapy if necessary, and a program of intense exercise. Prognosis is good for most children with persistent unexplained pain.

PHARMACOLOGIC MANAGEMENT

Diffuse Pain

A well-established conventional approach to initial therapy is the use of a tricyclic antidepressant (TCA) at bedtime. Amitriptyline, starting at 10 mg and escalating slowly to 50 mg, is a common choice. In approximately one third of fibromyalgia patients, low doses of amitriptyline produce moderate short-term improvements in pain, disturbed sleep, patient and physician global assessments, physical status, psychological status, and capacity for activities of daily living. Improvements in fatigue, tenderness, and stiffness are more modest. Long-term improvement with TCAs superior to that with placebo has not been established, however, and patient acceptance of these agents is poor because of their anticholinergic and sedative effects and tendency to cause weight gain. The selective serotonin reuptake inhibitors (SSRIs) fluoxetine[39] and citalopram have proved effective in randomized, controlled trials, and the combination of a TCA with an SSRI typically produces greater improvement in pain, sleep, and overall well-being than either drug used alone. Dual-action (serotonin/noradrenaline) reuptake inhibitors (SNRIs), such as venlafaxine and the investigational agents duloxetine and milnacipran, also show promise. Centrally acting skeletal muscle relaxants (e.g., cyclobenzaprine, baclofen, tizanidine) generally are not effective as single agents for the diffuse pain of fibromyalgia, although such usage is common practice and is supported by a few clinical trials. However, muscle relaxants given at low doses and in combination with a TCA or SSRI may provide some benefit, at least over the short term. Topical capsaicin is useful when gently massaged into painful areas twice a day. The patient should be informed that the initial discomfort often encountered with capsaicin will subside with time and that beneficial effects may not be apparent until after 3 to 4 weeks of therapy.

Table 3 Selected Antiepileptic Drugs for Treatment of Fibromyalgia Pain

Agent (Trade Name)	*Dosage*
Gabapentin (Neurontin)	300 mg h.s. to 600 mg t.i.d.; escalate over several weeks
Topiramate (Topamax)	25–50 mg h.s. to 200 mg b.i.d.; escalate by 25–50 mg at weekly intervals
Tiagabine (Gabitril)	2–4 mg h.s. to 56 mg q.d. in two to four divided doses; escalate by 4–8 mg at weekly intervals

Addition of an antiepileptic drug (AED) is indicated in patients with marked allodynia and hyperalgesia. AEDs have efficacy for pain sensitivity and as adjunctive medications for disturbed sleep and depression. Many choices are available, including gabapentin, topiramate, and tiagabine [*see Table 3*]; the experimental AED pregabalin has also shown promise for this purpose. Adjunctive therapy with anxiolytic drugs is often beneficial. Many choices are available, including benzodiazepines, buspirone, or SSRIs [*see 13:VIII Anxiety Disorders*]. AEDs may have anxiolytic effects, as well. Other drugs that may improve pain through effects on the abnormal central nociceptive processing in fibromyalgia include mexiletine, clonidine, and competitive antagonists of 5-hydroxytryptamine-3 [5-HT_3], such as tropisetron.

Opioids should be avoided if at all possible. A few patients can achieve a reasonable quality of life and daily functioning in no other way, however, so blanket interdiction of all opioid medications in fibromyalgia is inappropriate. When opioids prove necessary, they must of course be used with extreme caution. Patients must be closely monitored, agree to seek psychotherapy, and complete a signed promise to use only one prescribing physician and one dispensing pharmacy. So-called doctor shopping is not uncommon, especially among patients taking fentanyl or oxycodone-containing preparations, which are associated with increased psychological dependency. Tramadol, methadone, and hydrocodone are preferable.

Comorbid Nociceptive or Neuropathic Pain

When patients have discrete sources of pain, such as osteoarthritis, inflammatory arthritis, or degenerative disk disease, treatment should include nonsteroidal anti-inflammatory drugs (NSAIDs), with or without opioid or nonopioid analgesics. Analgesic drugs should be given in a stepwise approach on the basis of pain intensity and response. Clonidine (0.1 mg three times a day) and various AEDs [*see* Diffuse Pain, *above*] are helpful in neuropathic pain. Clonidine also is useful in decreasing withdrawal symptoms when opioids are being tapered. Different combinations of corticosteroid injections, activity modification, splints, local heat or cold, and other physical-therapy modalities may be indicated, depending on the specific musculoskeletal disorder.

Fatigue

Fatigue generally improves with effective treatment of pain, depression, and sleep disturbances in combination with a grad-

ed aerobic exercise program. Modafinil,[40] SSRIs, or 5-HT_3 receptor antagonists may benefit patients in whom overwhelming fatigue is a persistent complaint.

Poor Sleep

Sleep disturbances should be managed aggressively, beginning with instruction in the elements of good sleep hygiene, such as avoidance of daytime naps and caffeine. Most patients require medication. A single bedtime dose of a TCA can be given; other drugs that can be given in addition to or instead of a TCA include alprazolam, cyclobenzaprine, temazepam, trazodone, triazolam, and zolpidem.

Clonazepam or levodopa-carbidopa at bedtime may be effective for restless legs syndrome, which is very common in patients with fibromyalgia. A formal sleep study to identify sleep apnea and restless legs syndrome is indicated when the above simple measures are ineffective [*see 11:XIII Disorders of Sleep*].

Depression

Depression requires aggressive pharmacologic treatment. Many well-tolerated drugs are available [*see 13:II Depression and Bipolar Disorder*]. Both SNRIs (e.g., venlafaxine) and certain SSRIs (fluoxetine, citalopram, fluvoxamine) have been shown to improve pain and overall well-being in fibromyalgia independently of their effects on depression. Antidepressant drug treatment should be combined with psychotherapeutic management (formal or informal counseling) [*see 13:II Depression and Bipolar Disorder*].

NONPHARMACOLOGIC MANAGEMENT

Nonpharmacologic treatments generally have been of greater benefit in fibromyalgia than pharmacologic treatments alone.[34,41] Nondrug treatments may include exercise, meditation, yoga, biofeedback and stress management, counseling, and support groups. Patient education is important [*see Sidebar* Recommended Internet Sites on Fibromyalgia]. Optimum management of pain and fatigue in many patients is best achieved by interdisciplinary interaction of the primary care physician with allied health professionals and psychologists. Of particular efficacy are such approaches as biofeedback training and water aerobics, which promote active and independent participation by the patient, thereby enhancing self-management and self-efficacy. Conversely, referral of patients to so-called pain clinics that focus on various types of injection therapies and other passive procedures contributes little to the outcome of diffuse pain syndromes.

Biofeedback and Exercise

Significant improvement in physical status and self-reported fibromyalgia symptoms has been achieved by biofeedback training, alone or in combination with relaxation training and aerobic exercise. A systematic review of published trials indicates that supervised aerobic exercise training has beneficial effects on physical capacity, pain threshold, and global well-being.[42] Unfortunately, prescription of progressive home-based aerobic exercise is of limited efficacy unless the patient is highly motivated and has effective coping skills. Exercise should be of graded intensity; high-intensity fitness programs should be avoided.

Psychological Intervention

Psychotherapeutic counseling is helpful for many patients. Although formal cognitive-behavioral therapy is theoretically attractive, its cost-effectiveness and additional benefit over other interventions have not been established. Other psychological interventions, such as guided imagery, may improve self-reported pain over the short term.

Sidebar Recommended Internet Sites on Fibromyalgia

Fibromyalgia Information from The Oregon Fibromyalgia Foundation
http://www.myalgia.com
Fibromyalgia Information
http://www.ncf.carleton.ca/fibromyalgia
National Fibromyalgia Association
http://fmaware.org
Fibromyalgia—Mayo Clinic.com
http://www.mayoclinic.com/invoke.cfm?id=DS00079
National Fibromyalgia Research Association
http://www.nfra.net

TREATMENTS OF UNCERTAIN BENEFIT

Trigger-point injections, botulinum toxin injections, ultrasound, laser therapy, sphenopalatine blocks, and numerous other interventions have been advocated for fibromyalgia. These modalities either have no place in the treatment of fibromyalgia or are of questionable benefit.

COMPLEMENTARY AND ALTERNATIVE MEDICINE

CAM is used almost universally by patients with fibromyalgia, at least in part because of distrust of physicians and frustration with the limited efficacy of much of traditional care. Acupuncture, hypnotherapy, relaxation techniques (e.g., yoga, tai chi, meditation), certain herbal and nutritional supplements (magnesium, S-adenosylmethionine [SAMe]), and massage may be of benefit [*see CE:XII Complementary and Alternative Medicine*]. Vegetarian diets, magnet therapies, and chiropractic manipulation are not recommended.[43-45]

WHEN TO REFER

Consultative referral to a rheumatologist familiar with fibromyalgia is indicated when the diagnosis is unclear, when fibromyalgia is complicated by comorbid autoimmune or musculoskeletal disease, and when response to therapy is poor. Psychiatric referral is essential for severe depression with suicidal ideation and for comorbid psychosis.

Complications

Marital and other personal relationships, family health, and capacity to work productively are all threatened by severe fibromyalgia. Addiction to such drugs as opioids, benzodiazepines, and muscle relaxants is a rare but real risk in fibromyalgia patients. Nevertheless, the physician should be cognizant that drug-seeking behavior is often a sign of inadequate symptom control (pseudoaddiction) rather than drug dependency. Abrupt cessation of such medications may be associated with withdrawal symptoms. Exacerbation of symptoms often is iatrogenic—an unsympathetic, uninformed physician who is dismissive of fibromyalgia as a diagnostic entity and who fails to validate suffering can be a major perpetuating factor in this illness.

Prevention

Physicians may unwittingly contribute to the development of chronic pain in patients with acute pain from neck strain—typically, incurred in a minor traffic accident—by diagnostic indecision that permits patients to misconstrue the severity of their condition. In these cases, it is important for physicians to avoid such measures as open-ended referral for physical therapy, prolonged release from work, prescription of a neck brace, and failure to reassure the patient that recovery will be rapid and full. The term posttraumatic fibromyalgia should not be used.

Prognosis

Although improved treatment provides optimism for better outcomes in fibromyalgia, prospective long-term longitudinal studies in academic medical centers have shown little improvement in health status, disease severity, health service utilization, and costs, with approximately 25% of patients with fibromyalgia receiving disability or other compensation payments.[46,47] Persons with fibromyalgia suffer much more than those with other chronic rheumatologic diseases, such as rheumatoid arthritis.

Even though hyperalgesia and allodynia cannot be reversed entirely, most patients can expect substantial improvement in symptoms and in overall quality of life. Resolution of ongoing stress and promotion of the patient's self-efficacy for control of pain are of pivotal importance. Prognosis varies among three fairly distinct subsets of patients, who have been termed adaptive copers, interpersonally distressed, and dysfunctional.[6] Adaptive copers, many of whom do not seek care for fibromyalgia, do well with respect to self-reported pain, disturbed sleep, and fatigue. Interpersonally distressed patients also respond to a comprehensive interdisciplinary therapeutic approach. Dysfunctional patients with high levels of pain, anxiety, and opioid dependence do poorly, as do patients with pending litigation. The treatment goal that responds least to therapy is improvement in daily functioning.

The author has been a member of the speakers' bureau of Wyeth and a member of the advisory committee of Sanofi-Synthelabo, Inc., during the past 12 months.

References

1. Wolfe F, Smythe HA, Yunus MB, et al: The American College of Rheumatology 1990 Criteria for the Classification of Fibromyalgia. Report of the Multicenter Criteria Committee. Arthritis Rheum 33:160, 1990

2. Hyams KC: Developing case definitions for symptom-based conditions: the problem of specificity. Epidemiol Rev 20:148, 1998

3. Barsky AJ, Borus JF: Functional somatic syndromes. Ann Intern Med 130:910, 1999

4. Hudson JI, Mangweth B, Pope HG Jr, et al: Family study of affective spectrum disorder. Arch Gen Psychiatry 60:170, 2003

5. Okifuji A, Turk DC: Stress and psychophysiological dysregulation in patients with fibromyalgia syndrome. Appl Psychophysiol Biofeedback 27:129, 2002

6. Turk DC, Okifuji A, Sinclair JD, et al: Pain, disability, and physical functioning in subgroups of patients with fibromyalgia. J Rheumatol 23:1255, 1996

7. Giesecke T, Williams DA, Harris RE, et al: Subgrouping of fibromyalgia patients on the basis of pressure-pain thresholds and psychological factors. Arthritis Rheum 48:2916, 2003

8. Turk DC, Okifuji A, Sinclair JD, et al: Differential responses by psychosocial subgroups of fibromyalgia syndrome patients to an interdisciplinary treatment. Arthritis Care Res 11:397, 1998

9. Henningsen P, Zimmermann T, Sattel H: Medically unexplained physical symptoms, anxiety, and depression: a meta-analytic review. Psychosom Med 65:528, 2003

10. Neumann L, Buskila D: Epidemiology of fibromyalgia. Curr Pain Headache Rep 7:362, 2003

11. Winfield JB: Psychological determinants of fibromyalgia and related syndromes. Curr Rev Pain 4:276, 2000

12. The Biopsychosocial Approach: Past, Present, Future. Frankel R, Quill T, McDaniel S, Eds. University of Rochester Press, Rochester, New York, 2003

13. Gillespie NA, Zhu G, Heath AC, et al: The genetic aetiology of somatic distress. Psychol Med 30:1051, 2000

14. Buchwald D, Herrell R, Ashton S, et al: A twin study of chronic fatigue. Psychosom Med 63:936, 2001

15. Caspi A, Sugden K, Moffitt TE, et al: Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science 301:386, 2003

16. Offenbaecher M, Bondy B, de Jonge S, et al: Possible association of fibromyalgia with a polymorphism in the serotonin transporter gene regulatory region. Arthritis Rheum 42:2482, 1999

17. Fillingim RB: Sex-related influences on pain: a review of mechanisms and clinical implications. Rehabil Psychol 48:165, 2003

18. Staud R, Robinson ME, Vierck CJ Jr, et al: Diffuse noxious inhibitory controls (DNIC) attenuate temporal summation of second pain in normal males but not in normal females or fibromyalgia patients. Pain 101:167, 2003

19. Kroenke K, Spitzer RL: Gender differences in the reporting of physical and somatoform symptoms. Psychosom Med 60:150, 1998

20. Canli T, Desmond JE, Zhao Z, et al: Sex differences in the neural basis of emotional memories. Proc Natl Acad Sci USA 99:10789, 2002

21. Moldofsky H: Sleep and pain. Sleep Med Rev 5:385, 2001

22. Wolfe F: The relation between tender points and fibromyalgia symptom variables: evidence that fibromyalgia is not a discrete disorder in the clinic. Ann Rheum Dis 56:268, 1997

23. Elenkov IJ, Wilder RL, Chrousos GP, et al: The sympathetic nerve—an integrative interface between two supersystems: the brain and the immune system. Pharmacol Rev 52:595, 2000

24. Crofford LJ: The hypothalamic-pituitary-adrenal axis in the pathogenesis of rheumatic diseases. Endocrinol Metab Clin North Am 31:1, 2002

25. Clauw DJ, Crofford LJ: Chronic widespread pain and fibromyalgia: what we know, and what we need to know. Best Pract Res Clin Rheumatol 17:685, 2003

26. Fillingim RB, Wilkinson CS, Powell T: Self-reported abuse history and pain complaints among healthy young adults. Psychosom Med 61:111, 1999

27. Watkins LR, Milligan ED, Maier SF: Glial activation: a driving force for pathological pain. Trends Neurosci 24:450, 2001

28. Gracely RH, Petzke F, Wolf JM, et al: Functional magnetic resonance imaging evidence of augmented pain processing in fibromyalgia. Arthritis Rheum 46:1333, 2002

29. Staud R: Evidence of involvement of central neural mechanisms in generating fibromyalgia pain. Curr Rheumatol Rep 4:299, 2002

30. Walker EA, Keegan D, Gardner G, et al: Psychosocial factors in fibromyalgia compared with rheumatoid arthritis: II. Sexual, physical, and emotional abuse and neglect. Psychosom Med 59:572, 1997

31. Pain management in the rheumatic diseases. Rheum Dis Clin North Am 25:1, 1999

32. Rational management of fibromyalgia. Rheum Dis Clin North Am 28:1, 2002

33. Arnold LM, Keck PE Jr, Welge JA: Antidepressant treatment of fibromyalgia: a meta-analysis and review. Psychosomatics 41:104, 2000

34. Rossy LA, Buckelew SP, Dorr N, et al: A meta-analysis of fibromyalgia treatment interventions. Ann Behav Med 21:180, 1999

35. Sim J, Adams N: Systematic review of randomized controlled trials of nonpharmacological interventions for fibromyalgia. Clin J Pain 18:324, 2002

36. Guideline for the Management of Pain in Fibromyalgia. American Pain Society (in preparation) http://www.ampainsoc.org

37. The management of chronic pain in older persons. American Geriatrics Society Panel on Chronic Pain in Older Persons. J Am Geriatrics Soc 46:635, 1998

38. Schanberg LE: Widespread pain in children: when is it pathologic? Arthritis Rheum 48:2402, 2003

39. Arnold LM, Hess EV, Hudson JI, et al: A randomized, placebo-controlled, double-blind, flexible-dose study of fluoxetine in the treatment of women with fibromyalgia. Am J Med 112:191, 2002

40. Schaller JL, Behar D: Modafinil in fibromyalgia treatment. J Neuropsychiatry Clin Neurosci 13:530, 2001

41. Williams DA: Psychological and behavioural therapies in fibromyalgia and related syndromes. Best Pract Res Clin Rheumatol 17:649, 2003

42. Busch A, Schachter CL, Peloso PM, et al: Exercise for treating fibromyalgia syndrome. Cochrane Database Syst Rev (3):CD003786, 2002

43. Crofford LJ, Appleton BE: Complementary and alternative therapies for fibromyalgia. Curr Rheumatol Rep 3:147, 2001

44. Ebell MH, Beck E: How effective are complementary/alternative medicine (CAM) therapies for fibromyalgia? J Fam Pract 50:400, 2001

45. Holdcraft LC, Assefi N, Buchwald D: Complementary and alternative medicine in fibromyalgia and related syndromes. Best Pract Res Clin Rheumatol 17:667, 2003

46. Wolfe F, Anderson J, Harkness D, et al: Health status and disease severity in fibromyalgia: results of a six-center longitudinal study. Arthritis Rheum 40:1571, 1997

47. Wolfe F, Anderson J, Harkness D, et al: A prospective, longitudinal, multicenter study of service utilization and costs in fibromyalgia. Arthritis Rheum 40:1560, 1997

48. Yunus MB, Masi AT: Fibromyalgia, restless legs syndrome, periodic limb movement disorder and psychogenic pain. Arthritis and Allied Conditions: A Textbook of Rheumatology. McCarty DJ Jr, Koopman WJ, Eds. Lea & Febiger, Philadelphia, 1993, p 1383

INTRODUCTION TO WOMEN'S HEALTH: THE PRIMARY CARE OF WOMEN

JANET B. HENRICH, M.D.

Definition of Women's Health

Women's health can be defined as diseases or conditions that are unique to women or that involve gender differences that are particularly important to women.[1] This definition acknowledges the increasing scientific evidence supporting a focus on sex and gender and expands the concept of women's health beyond the traditional focus on reproductive organs and their function.[2,3] Over time, the definition has come to include an appreciation of wellness and prevention, the interdisciplinary and holistic nature of women's health, the diversity of women and their health needs over the lifespan, and the central role of women as patients and as active participants in their health care.[4]

This broader interdisciplinary perspective has important implications for clinicians providing care to women. In addition to understanding basic female physiology and reproductive biology, clinicians need to appreciate the complex interaction between the environment and the biology and psychosocial development of women. When dealing with conditions that are not specific to women, clinicians need to be aware of those aspects of disease that are different in women or have important gender implications. The ability to apply this information requires that clinicians adopt attitudes and behavior that are culturally and gender sensitive. Clinicians also must understand the changes that are occurring in women's patterns of health-seeking and forms of communication and interaction with the health care system.

Goals of the Women's Health Section

The women's health section in *ACP Medicine* is intended to furnish clinicians with the knowledge base and conceptual framework needed to provide comprehensive care to female patients of all ages. Chapters on reproductive function are written for generalists in practice and discuss normal and abnormal menstruation and common problems in younger reproductive-age women, including the aspects of fertility and pregnancy that are under the purview of generalist clinicians. Other chapters discuss the conditions that cause the greatest morbidity and mortality in women as they age; many of these conditions are linked to the menopause and the decline in estrogen levels that accompany waning ovarian function. This chapter contains additional information on some of these topics, as well as links to other chapters in *ACP Medicine* covering important women's health topics.

Factors That Influence Women's Health

HORMONAL INFLUENCES

A dominant theme in women's health is the effect of marked changes in endogenous hormone levels that occur throughout a woman's lifetime. A health-issue category unique to women is the influence, during the reproductive years, of sex hormones on sexual development and reproductive function. As women grow older and sex hormone levels decrease with menopause, women's risk factors for certain diseases increase dramatically and become more similar to those of men. Although women develop the same diseases that affect men, biologic mechanisms and psychosocial factors influence the course of disease differently in women.

SOCIAL FACTORS

One of the important social trends over the past 50 years is the increasing participation of women in the workforce. Currently, it is estimated that 59% of women 16 years of age and older are in the workforce—including 71% of women with children younger than 18 years—and 60% of working women are employed full time throughout the year.[5] The cumulative effects of playing multiple roles, work stress, and new environmental exposures on women's health and reproductive status are largely unknown, but they are certain to have important health and social ramifications.

Paralleling the growing numbers of women in the workforce is the increasing number of single-parent families headed by women, especially minority women. Many of these families live in poverty. Increasing evidence indicates that socioeconomic factors are major indicators of health and that, for some health outcomes, poverty and lack of education are more important determinants of health than ethnicity.[6] However, socioeconomic status cannot wholly explain important ethnic and racial differences that remain in women's susceptibility and response to certain diseases. For example, the mortalities associated with coronary heart disease (CHD), stroke, and breast cancer are higher in black women than in white women, whereas the mortality associated with lung cancer is higher in white women.[6] Regardless of their racial or ethnic designation, minority-group women have a lower life expectancy than white women and experience greater health problems. The differences are most pronounced in areas related to reproductive issues and childbearing, the occurrence and course of chronic disease, the incidence and outcome of cancer, and the risk of interpersonal violence.[6]

MORBIDITY AND MORTALITY IN WOMEN

At the beginning of the 20th century, the average lifespan of women in the United States was 48 years, compared with 46 years for men. Since then, the life expectancy for women has increased more than 30 years and is currently close to 80 years, compared with 74 years for men.[6] The dramatic increase in overall life expectancy is thought to be related to lower infant mortality, control of infectious diseases, and progress in the treatment of chronic diseases, such as diabetes and cardiovascular disease.

Despite a dramatic decline in mortality from heart disease that has occurred in both sexes over the past 2 decades, heart disease remains the leading cause of death for women, accounting for 30% of all deaths in women [*see Table 1*]. Heart disease occurs in women 10 to 15 years later than it does in men. This delayed onset, which is attributed primarily to the protective effect of estrogen, accounts for the fact that 90% of heart disease–related mortality in women occurs after the menopause. Evidence suggests that heart disease in women is more serious than that in

men, resulting in higher mortality. In addition to being affected by biologic factors, the poorer survival of women may stem from the older age and increased prevalence of comorbid conditions in women at the time of diagnosis, as well as to less well defined social factors that influence the diagnosis and treatment of heart disease in women.

Cancer is the second leading cause of death in women and is the most common cause of premature death. Whereas death rates for all cancers combined changed little for women during the last half of the 20th century, major advances in the diagnosis and treatment of cervical and uterine cancers were offset by an increase in mortality from lung and breast cancer. Although breast cancer is the most common cancer diagnosed in women, lung cancer is the overall leading cause of cancer deaths.

Death from breast cancer is the second overall leading cancer death in women and the most common in women younger than 55 years. Although the incidence of breast cancer continues to rise, mortality has begun to decline for the first time since cancer registries began tracking this disease in the 1940s. This divergence is thought to be related partly to the widespread use of screening mammography and the detection of cancers in earlier stages that have a more favorable prognosis.

There are significant racial differences in breast cancer incidence. White women continue to have the highest rates of breast cancer, followed closely by black women. Hispanic and Asian/Pacific Islander women are at intermediate risk, and Native American women are at lowest risk.[6]

There are also significant racial differences in breast cancer mortality. Although the incidence of breast cancer is 22% higher in white women than in black women, mortality is 36% higher in black women.[6] The reasons for the racial differences with breast cancer incidence and mortality are unclear, but the differences may be related to socioeconomic and biologic factors, as well as to certain health behaviors, such as participation in screening mammography.

Although stroke-related deaths have declined by more than 60% in the United States over the past 25 years, stroke still causes 8% of all deaths in women and ranks third as the cause of death for all persons in the United States. Striking racial differences exist in stroke mortality, which is almost twice as high in black women as in white women throughout most of the lifespan.[6] Most of the fatal strokes in women result from thromboembolic disease and occur in older women.

Death rates from chronic lower respiratory diseases have increased steadily for both women and men during the past 25 years; however, the increase has been greater in women. Because this increase has been linked to patterns in cigarette smoking—during the past century, smoking rates rose later in women than in men—the increase in death rates in women for pulmonary disease, as well as for lung cancer, are expected to continue to rise.

The reported mortality of diabetes is most likely underestimated because of its strong association with other life-threatening medical conditions, such as cardiovascular disease, stroke, and kidney failure. Eight percent of all women 20 years of age and older have diabetes; however, prevalence rates are higher in black, Hispanic, and Native American women.[6] In addition to its mortality, diabetes is a significant cause of morbidity and, in women of childbearing age, has important adverse effects on pregnancy and pregnancy outcome, resulting in an increased risk of toxemia, macrosomia, hydramnios, congenital malformations, cesarean section, and fetal and perinatal mortality.

Table 1 Leading Causes of Death and Number of Deaths in Women in the United States in 2002[6]

Cause of Death	*Number of Deaths*	*Total Deaths (%)*
All causes	1,244,123	100.0
Heart disease	356,014	28.6
Cancer	268,503	21.6
Stroke	100,050	8.0
Chronic lung diseases	64,103	5.2
Alzheimer disease	41,877	3.4
Diabetes	38,948	3.1
Unintentional injuries	37,485	3.0
Pneumonia/influenza	36,763	3.0
Kidney disease	21,279	1.7

Although the mortality associated with HIV infection and AIDS began to decline in the mid-1990s as a result of highly effective combination treatment for HIV infection and better prevention of opportunistic infections, HIV remains a leading cause of death in younger women. Women overall account for an increasing proportion of AIDS cases, and ethnic-minority women are disproportionately affected.[6] As the features of this epidemic change, with heterosexual transmission accounting for the majority of new HIV cases in women, these rates are expected to continue to rise.

Mortality data alone do not provide a complete picture of women's health status. Although women live longer than men, overall measures of health status are worse in women. Estimates from the 2002 National Health Interview Survey suggest that more women than men report symptoms of or seek care for acute medical conditions and that women are more disabled by these self-limited illnesses, as measured by the number of days spent in bed or lost from work.[6] In addition, several chronic conditions occur more frequently in women and cause significant disability, such as arthritis, chronic pain syndromes (including migraine and neck, back, and facial pain), and chronic respiratory disorders (including chronic bronchitis, sinusitis, and asthma).[6] Women are also more likely than men to experience severe psychological stress.[6] Data from the National Institute of Mental Health show that affective disorders, especially major depressive episodes, and the anxiety disorders are significantly more prevalent in women.[7]

WOMEN'S HEALTH ACROSS THE LIFESPAN

Many of the important health issues in women have their onset or greatest impact at certain ages and are intricately linked with women's psychosocial and sexual development. To develop a more integrated concept of women's health, it is instructive to look at the important health issues in women within the major lifespan groups.[8]

Birth to Young Adulthood

As girls reach puberty, the health issues that emerge are related primarily to developmental changes involving physical and sexual growth and changing relationships within and outside the family. Central to the psychosocial development of young

women is the process of gender identification and orientation and the development of self-esteem. Intentional and unintentional injuries, including an increasing frequency of acts of physical and sexual violence, are the primary cause of death and disability in young women. Chronic disease or disability develops in a small proportion of girls. Most of these conditions (e.g., lupus erythematosus, juvenile rheumatoid arthritis, and thyroid disease) have an autoimmune component. Because of hormonal influences, many of these conditions first occur or are exacerbated during puberty.

15 to 44 Years

The second lifespan group stretches from ages 15 to 44. As women progress through this age group, cancers of the breast and reproductive tract emerge as the leading cause of death, followed by unintentional injury and heart disease. Of the unintentional and intentional injuries, motor vehicle accidents, homicide, and suicide account for three fourths of all injury-related deaths.[6] Death rates from homicide and suicide have shown a downward trend in young women. Nevertheless, black women, like black men, are more likely than members of other races to be homicide victims, and firearms are used in more than half of those deaths. Intimate partners play a substantial role in violence against women; one third of murders of women are perpetrated by an intimate partner.[9]

The most dramatic trend in this age group, beginning in the 1980s and peaking in the mid-1990s, was the emergence and rapid rise of HIV infection as a major cause of death. Although overall AIDS incidence and death rates have decreased yearly since 1996, the rate of decline has been smaller in women. The consequences of this disease for gynecologic care and reproductive counseling for women are unique. Because HIV can be transmitted during pregnancy and more than 40% of pregnancies are unintended, routine medical care should include discussions about effective contraceptive methods, the effects of pregnancy on HIV infection and treatment, and the potential for perinatal transmission of HIV. Treatment strategies for women who may become or who are pregnant should take into consideration regimens that maximally suppress maternal viral load and reduce transmission to the fetus while minimizing toxicity. A three-part regimen of zidovudine (AZT) reduces the risk of perinatal transmission by 70% and is effective even in women with advanced disease.[10] United States Public Health Service recommendations for antiretroviral chemoprophylaxis to reduce perinatal HIV transmission are evolving rapidly and take into consideration the now standard use of more aggressive combination drug therapies to treat HIV, as well as the clinical status and antiviral drug history of the woman. Zidovudine should be part of the antepartum drug regimen in all pregnant HIV-infected women, if feasible. Treatment guidelines are updated periodically and are available on the Internet (http://AIDSinfo.nih.gov).

An important role of clinicians in the care of young women is to recognize and reduce risk-taking and other unhealthy behaviors. For example, early or unprotected sexual activity increases women's risk for sexually transmitted diseases (STDs). Not only are these diseases transmitted more easily from men to women, but women are disproportionately affected because of infectious complications that can lead to disorders of reproductive function, such as pelvic inflammatory disease and subsequent ectopic pregnancy or infertility. Unfortunately, efforts at risk reduction, particularly in the use of harmful substances, are hampered by industry and market forces and other social factors that influence women's lives. For example, the adverse effects of cigarette smoking on lung cancer and other respiratory diseases, heart disease, osteoporosis, and pregnancy are well documented, yet 20% of women continue to smoke and 18% of teenage mothers smoke during pregnancy.[6]

Social and cultural factors have also contributed to the increasing prevalence of dieting and eating disorders. It is estimated that up to 4% of young women suffer from anorexia nervosa or bulimia, and an additional 5% have less specific eating disorders that are characterized by aberrant eating patterns and weight-management habits.[11] These statistics most likely underestimate the prevalence of eating disorders in young women. According to findings from the 2003 Youth Risk Behavior Surveillance System developed by the Centers for Disease Control and Prevention, during the month before the survey, 56% of adolescent women had attempted dieting, 18% had gone more than 24 hours without eating, 11% had taken diet aids without professional advice, and 8% had induced vomiting or taken laxatives for weight control.[12]

The female lifespan group of 15 to 44 years of age delineates the reproductive years. In addition to having the traditional childbearing and family responsibilities, women are increasingly assuming new, additional roles. The effect of multiple and often conflicting roles on women's mental and physical health remains to be determined, but it is almost certainly closely linked to reproductive freedom and health. Thus, clinicians need to understand the safety and effectiveness of current methods of contraception, including extended oral contraception and emergency contraception, as well as the acceptability of these methods to women of various cultures.[13-15]

Many common disorders of reproductive function are not exclusively gynecologic problems. For example, polycystic ovary syndrome is associated with insulin resistance and an increased risk of diabetes and cardiovascular disease[16] [*see 16:V Polycystic Ovary Syndrome*].

Autoimmunity is a common theme in many of the medical disorders that have the highest prevalence in women 15 to 44 years of age. Most of the autoimmune diseases occur more often, as well as cause greater morbidity, in women than in men. Many of these diseases are influenced by changes in estrogen levels, particularly during pregnancy. The prevalence rates of collagen vascular diseases such as rheumatoid arthritis, systemic lupus erythematosus, and scleroderma are three to nine times higher in women than in men. Many autoimmune-related endocrinopathies, such as Hashimoto thyroiditis and Graves disease, have a female-to-male ratio as high as 10:1. Less well recognized is the role of autoimmunity in recurrent pregnancy loss and infertility in women.

Among the mental disorders, depressive illnesses are twice as common in women as in men. Each year, an estimated 6.5% of women experience a major depressive episode, and twice that many have chronic low-grade depression. The increased risk of depression in women begins at puberty and declines after the menopause. In addition, many women experience mood, cognitive, or behavioral changes associated with cyclic changes in hormone levels during the menstrual cycle or with the marked changes in hormone levels that occur during the postpartum period and the menopausal transition. The genetic, biologic, and environmental contributions to women's susceptibility to depression are not fully understood; however, hormonal factors are thought to play a major role. Women are also twice as likely as men to be diagnosed with an anxiety disorder, including pan-

ic disorder, posttraumatic stress disorder, generalized anxiety disorder, agoraphobia, and simple phobia.[7]

A major cause of psychosocial morbidity in women is physical and sexual abuse. According to the National Violence Against Women Survey conducted by the National Institute of Justice and the Centers for Disease Control and Prevention in 1998, 52% of women have been physically assaulted at some time during their life, and 18% have experienced a rape.[17] Young women are at particular risk for rape; of those women who have been raped, more than half were younger than 18 years when rape first occurred.[17]

Physical and sexual assault of women is primarily a problem of partner violence. Three fourths of women who experience physical or sexual abuse after 18 years of age are assaulted by a current or former spouse or a male intimate. Unfortunately, clinicians often fail to recognize or address symptoms of abuse, whether because of inadequate knowledge of physical and sexual violence, misconceptions about it, or inadequate training in its management. Adequate screening tools are especially crucial in the emergency department, where up to one third of women who have been assaulted seek care.

45 to 64 Years

Death rates for women 45 to 64 years of age have declined by 30% in the past 25 years. Previously, the leading cause of death in this age group was heart disease; however, cancer is now ranked number one. This shift in mortality reflects primarily a decline in death rates for heart disease—a decline that has been observed in both sexes and that is attributed to changes in lifestyle, better control of hypertension, and lower blood cholesterol levels.

Many of the important chronic conditions in women first appear in this age group; some of them, such as heart disease, osteoporosis, and cancer, are inextricably linked to the menopause and the associated marked decline in estrogen levels. Hormone replacement is the most effective therapy for vasomotor and vaginal symptoms associated with the menopause; in addition, hormone therapy (HT) decreases bone loss and the risk of osteoporotic fractures and colon cancer. However, because of the adverse effects of HT shown in the Women's Health Initiative Trial (i.e., an increased incidence of breast cancer, heart disease, stroke, dementia, and urinary problems),[18-24] HT should not be used as preventive therapy in postmenopausal women [*see 16:XI Menopause*].

Although the menopause encompasses many of the physiologic changes that define this period, women also experience major transitions in social roles and life circumstances that profoundly affect their physical and mental health: children leave home; many women become widowed or divorced; parenting roles change as women are called upon to care for aging parents; and disabilities increase, making it difficult for some women to function within and outside the home. An understanding of these life events is essential to the comprehensive care of mature women.

65 Years and Older

Heart disease is the leading cause of death in women 65 years of age and older, followed by cancer and stroke. Mortality for all three disorders rises steeply after 65 years of age and begins to approach the rates for men.

As the longevity of women increases, they bear the burden of illnesses that are seen primarily in the very old. Of these, the neurologic degenerative diseases, such as dementia, sleep disorders, and neurosensory and movement disorders, are particularly common in women. Unfortunately, the added years of life in women are often spent in a frail or dependent state and often result in institutionalization.

The social and psychological changes that women experience as they age add to the burden of illness. Social isolation increases because of the death of loved ones, loss of financial stability, and increasing physical disabilities. In addition to an increasing incidence of dementia with age, mental health problems become more prevalent or serious. The role of primary care clinicians is to recognize and help reduce the impact of these accumulated conditions on women's ability to function and on their quality of life.

Clinical Evaluation of Women

HISTORY

Most elements of the medical history are similar for women and men. Those that are unique to women or are different in women are discussed here.

The goal of the dietary, exercise, and weight history in women is to identify patterns that may indicate an eating disorder or weight-management problem [*see 13:IX The Eating Disorders*]. An assessment of calcium and vitamin D intake is important for bone health at all ages—including the teenage years and 20s, when peak bone mass is built—as is questioning about folic acid supplementation in women of childbearing age to help prevent development of fetal neural tube defects in early pregnancy.

The menstrual and reproductive history, including past and current forms of contraception and hormone use, is important in evaluating current problems and in assessing risk for future conditions. The focus of the sexual history varies depending on age, with an emphasis on STD and pregnancy risk assessment in young women, the identification of medical conditions that may adversely affect fertility and pregnancy outcome in women who are considering pregnancy [*see 16:IX Medical Complications in Pregnancy*], and conditions that may interfere with healthy sexual function in older women. Symptoms of urinary incontinence and pelvic floor dysfunction should be sought specifically [*see 16:XII Urinary Incontinence and the Overactive Bladder*]. The need for Papanicolaou (Pap) smear screening—and, at an appropriate age, screening mammography—should be ascertained.

Screening questions for depression and interpersonal violence are appropriate in women of all ages and socioeconomic class because of the high prevalence of these conditions in women.

Primary clinicians are in a unique position to identify patients at risk for inherited cancer syndromes. Careful questioning about a family history of breast, ovarian, and related cancers can identify women at high risk for the breast cancer susceptibility genes *BRCA1* and *BRCA2* who may benefit from genetic testing and subsequent preventive measures if they test positive for a mutation [*see 3:VIII Genetic Diagnosis and Counseling*].

When obtaining the medication history, clinicians need to be aware of important sex differences in the effects of certain classes of medications that may put women at increased risk for adverse drug reactions.[25,26] For example, because women have longer baseline QT intervals than men, the use of drugs that further prolong the QT interval, such as certain antibiotics, antihistamines, antiarrhythmics, and antipsychotics, may put women at increased risk for torsade de pointes, a potentially lethal ventric-

ular arrhythmia.[27,28] The University of Arizona maintains an online registry of drugs that prolong the QT interval or induce torsade de pointes (http://www.arizonacert.org/medical-pros/drug-lists/drug-lists.htm). Also, women with depressed left ventricular function who use digoxin for heart failure are at greater risk of death than men with a similar diagnosis.[29] This risk is attributed partly to higher serum digoxin levels in women. Because of this potential harm, clinicians are advised to weigh carefully the risks versus benefits of digoxin therapy in women with heart failure.[30]

The hepatic cytochrome P-450 family of enzymes is involved in the metabolism of endogenous and exogenous steroids and many other drugs. Anticonvulsants can lead to oral contraceptive (OC) failure by inducing the cytochrome P-450 enzyme system to metabolize estrogen more rapidly. A similar mechanism has been attributed to antibiotics, leading to drug warnings about the risk of pregnancy when antibiotics are prescribed to women on OCs. Because the evidence supporting drug interactions between antibiotics and OCs is weak,[31] the Council on Scientific Affairs of the American Medical Association convened a panel to review the data and provide recommendations. The panel concluded that rifampin, acting primarily as an inducer of the cytochrome P-450 3A4 isoenzyme, is the only antibiotic tested that significantly reduces plasma concentrations of OC-derived estrogens and increases OC failure rates.[32] However, women using several other commonly prescribed antibiotics have wide variations in OC-derived hormone levels; some women have elevated follicle-stimulating hormone levels or breakthrough bleeding, suggesting that ovulation may occur. On the basis of these findings, the panel concluded that even though the risk of OC failure with these antibiotics is very small, some women may be at risk, particularly those using low-dose or very low dose OCs. The panel recommended nonhormonal contraceptive methods for women who are concomitantly using OCs and rifampin, as well as for women who are taking other antibiotics and are concerned about a small risk of pregnancy, have had previous contraceptive failure, or develop breakthrough bleeding.[32]

Because women are generally smaller and have more body fat than men, there is an increased risk of bleeding in women who are on anticoagulants if doses are not carefully adjusted for body size. Women also have an enhanced response to certain fat-soluble drugs, such as antipsychotics.[25] High body weight by itself can influence drug efficacy in women. For example, women who weigh 70.5 kg or more are at increased risk for contraceptive failure on OCs—particularly with low-dose and very low dose OCs.[33] This effect is attributed to lower circulating drug levels as a result of higher metabolic rates in women with a large body habitus, increased absorption of OCs by fat cells in women with high fat levels, or both.

PHYSICAL EXAMINATION

Most primary clinicians include the breast and pelvic examinations as part of the routine physical examination in women. Skill in these examinations requires an appreciation of the spectrum of normal findings in women of different age groups and the ability to recognize pathology. Clinicians who wish to increase their skills in performing the breast and pelvic examination can attend workshops or training sessions in one of several teaching associate programs affiliated with academic medical centers or can study teaching modules developed by the American College of Physicians and other professional organizations.

LABORATORY TESTS

Routine Blood Tests

Body structure and hormones, particularly estrogen, affect routine blood test results in women. Serum creatinine and creatine kinase (CK) levels are generally lower in women than in men because women have less muscle mass. The MB fraction of CK (CK-MB) is specific for cardiac tissue and is used as a marker for cardiac injury. CK-MB reference limits for diagnosing acute coronary syndromes are lower in women than in men because women generally have smaller hearts. There is also a small contribution to CK-MB from skeletal muscle, leading to higher levels in men.[34] Premenopausal women have lower serum uric acid levels than men because estrogen increases renal urate excretion.

Alkaline phosphatase levels increase gradually with aging in both women and men and can reach levels that are 50% higher than those reported in young adults. The source of the increase in postmenopausal women is bone; the liver is the source in older men.

Estrogen and the serum estrogen receptor modulators tamoxifen and raloxifene increase thyroid-binding globulin levels in women, resulting in an increase in total thyroxine (T_4) and triiodothyronine (T_3) concentrations but not in the free concentrations of thyroid hormones; thus, thyroid function remains normal.

Women have lower ferritin levels than men as a result of iron losses during menses, pregnancy, and lactation, and a high proportion of women have absent iron stores, resulting in lower red blood cell (RBC) measures (e.g., hemoglobin concentration, hematocrit, and RBC count). Because anemia is defined as RBC values more than two standard deviations below the mean, normal ranges for women are lower than they are for men; however, these differences diminish with age, because RBC measures decline in both women and men as part of normal aging.

Breast Cancer Screening

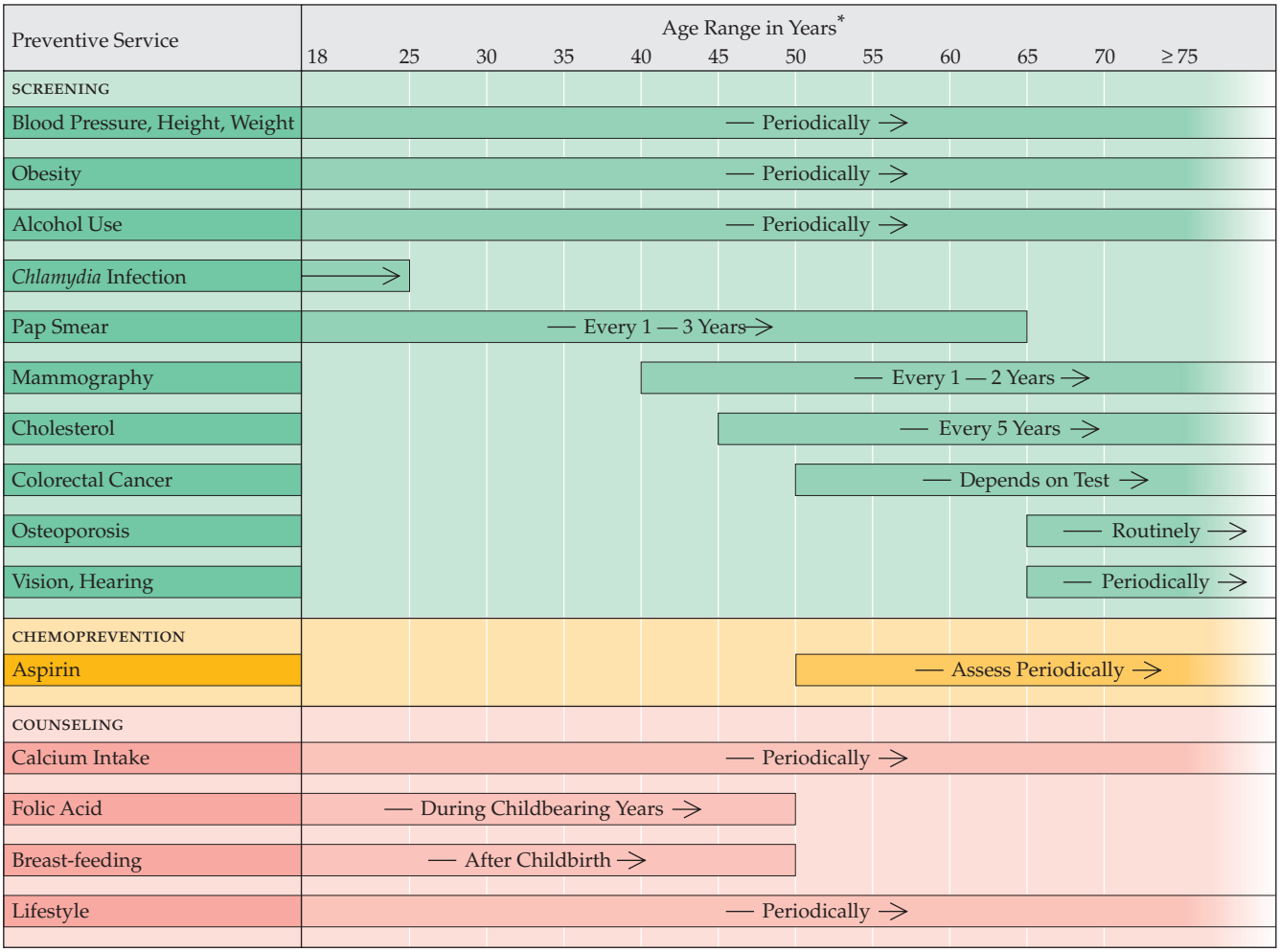
Primary care clinicians need to be familiar with the risks and benefits of screening mammography and of newer breast cancer screening technologies offered to high-risk women, including breast magenetic resonance imaging[35-38] [*see 12:VII Breast Cancer*]. Many clinicians do a clinical breast examination (CBE) before ordering a screening mammogram to help prevent the possibility of a biased exam if the mammography report is normal. A mass found on CBE should be investigated further even if it is not identified by mammography[39] [*see 16:XIV Approach to the Patient with a Breast Mass*].

Osteoporosis Screening

Clinicians should be knowledgeable about the indications for bone densitometry and should be able to interpret a woman's fracture risk on the basis of her bone mass and other factors that influence the risk of hip and other fractures [*see 3:VI Diseases of Calcium Metabolism and Metabolic Bone Disease*].

Cardiovascular Disease Testing

Noninvasive tests for CHD generally have lower diagnostic accuracy in women than in men. Consequently, clinicians should be familiar with the sensitivity and specificity of exercise testing and of newer imaging techniques whose accuracy is lower in women [*see 16:XIX Cardiovascular Disease in Women*].



* Upper age limits should be individualized for each patient.
For cardiovascular disease, when indicated by risk assessment.
Tobacco cessation, drug and alcohol use, sexually transmitted diseases and HIV, nutrition, physical activity, sun exposure, oral health, injury prevention, and polypharmacy.

Figure 1 **United States Preventive Services Task Force guidelines for preventive primary care in women.[42] These recommendations are conservative and represent the minimum level of preventive services that should be offered.**

Cervical Cancer Screening

Cervical cancer screening requires proficiency in obtaining the traditional Pap smear or in using liquid-based cytology or other new cervical cancer screening technologies. Knowledge of the Bethesda system for reporting results of cervical cytology is essential to interpreting Pap test results and in managing patients with atypical cells, including DNA testing for high-risk types of human papillomavirus (HPV).[40,41]

Vaginitis and Sexually Transmitted Disease Testing

Because STDs are often asymptomatic in women, screening is vital for preventing their spread, as well as their progression to pelvic inflammatory disease. Nonculture technologies have become the preferred method for diagnosing *Chlamydia* infections and gonorrhea in women and in identifying the cause of genital ulcers [*see 7:XXII Vaginitis and Sexually Transmitted Diseases*]. Clinicians need to be familiar with the use of these tests and the proper way to collect specimens. None of these newer technologies, however, replace the simple saline wet mount as a valuable tool for establishing the cause of a vaginal discharge.

Primary Prevention Guidelines in Nonpregnant Women

Many of the conditions that cause the greatest morbidity and mortality in women can be prevented or delayed by the early recognition and treatment of risk factors and the identification of individuals who might benefit the most from early intervention. The United States Preventive Services Task Force (USPSTF) publishes primary preventive guidelines that can assist clinicians in providing care to women [*see Figure 1*].[42] These guidelines are available online (http://www.ahrq.gov/clinic/uspstfix.htm). The recommendations are conservative and represent the minimum level of preventive services that should be offered.

The USPSTF recommends that all women have periodic measurement of blood pressure, height, and weight, as well as undergo screening for obesity and depression. Cervical cancer screening with Pap smears is recommended in sexually active women. Although other policy-making organizations offer different guidelines about when to start or stop Pap smear screening and the optimal screening interval in low-risk women, the USPSTF recommends initiating screening 3 years after sexual activity begins or at age 21, whichever occurs first, and screening

low-risk women at least every 3 years after two or three annual normal Pap smears. Routine screening is not recommended for low-risk women after age 65 if they have had adequate recent screening or for women who have had a total hysterectomy for benign conditions.

Newer screening methods using liquid-based cytology offer advantages over conventional Pap smears, such as the option of reflex HPV testing, but are more expensive. Until more information is available from trials assessing the role of HPV testing in cervical cancer screening, the USPSTF holds that the evidence is insufficient to recommend the use of this new technology over the routine Pap smear. Sexually active women younger than 25 years should be screened routinely for *Chlamydia*; screening for *Chlamydia*, gonorrhea, syphilis, and HIV should be done in any woman with risk factors for these diseases.

The USPSTF recommends that breast cancer screening with mammography, with or without clinical breast exam, be done every 1 to 2 years starting at age 40, although women 50 to 65 years of age and those at increased risk for breast cancer benefit the most [*see CE:V Adult Preventive Health Care*]. Clinicians should discuss the risks and benefits of breast cancer chemoprophylaxis with women who are at high risk for breast cancer. Routine screening for ovarian cancer is not recommended.

According to the USPSTF, women should receive cholesterol screening every 5 years starting at age 45; women with risk factors for cardiovascular disease, such as hypertension or hyperlipidemia, should be screened further for diabetes. As with men, colorectal screening is recommended beginning at age 50. Routine screening for osteoporosis in women is recommended starting at age 65. Many younger women, however, have conditions that put them at increased risk for osteoporotic fractures; in these women, screening is recommended to be started at age 60. Periodic vision and hearing screening is recommended for both women and men after age 65.

On the basis of findings from studies that were conducted primarily in men, the USPSTF currently recommends that clinicians discuss the use of aspirin for the primary prevention of CHD with postmenopausal women and with younger women with risk factors for CHD. However, results from the first randomized trial that assessed the risks and benefits of aspirin as chemoprevention for CHD in women found a beneficial effect only in women 65 years of age and older.[43] In younger women, aspirin reduced the risk of thromboembolic stroke but had little effect on CHD, findings opposite to those described in men. As in men, the use of aspirin in older women was associated with an increased risk of hemorrhagic stroke and major gastrointestinal hemorrhage. Until more information is available about aspirin's effects in women, discussions about its use should be individualized for each patient.

The USPSTF recommends counseling women of childbearing age about folic acid supplementation to reduce the risk of neural tube defects and recommends counseling women of all ages about adequate calcium and vitamin D intake. All women should be counseled about tobacco and other substance use, strategies to decrease the transmission of STDs, effective methods of contraception, a healthy diet and increased physical activity, and injury prevention. In older women, counseling should focus on prevention of falls and the potential dangers of multiple medication use.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Women's Health: report of the Public Health Service Task Force on Women's Health Issues. US Department of Health and Human Services, Public Health Service, Washington, DC, 1985

2. Agenda for Research on Women's Health for the 21st Century. A report of the Task Force on the NIH Women's Health Research Agenda for the 21st Century. National Institutes of Health, Office of the Director, Office of Research on Women's Health, Bethesda, Maryland, 1999

3. Exploring the Biological Contributions to Human Health. Does Sex Matter? Wizemann TM, Pardue ML, Eds. National Academy Press, Washington, DC, 2001

4. Women's health in the curriculum: a resource guide for faculty. Donoghue G, Ed. National Academy on Women's Health Education, Philadelphia, 1996

5. Women in the Labor Force: a Databook. Bureau of Labor Statistics, United States Department of Labor, Washington, DC, 2005

6. Health, United States, 2004 with Chartbook on Trends in the Health of Americans. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2004
http://www.cdc.gov/nchs/data/hus/hus04.pdf

7. The Numbers Count: Mental Disorders in America. National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, 2001

8. Report of the National Institutes of Health: Opportunities for Research on Women's Health. Office of Research on Women's Health, National Institutes of Health, Bethesda, Maryland, 1992

9. Rennison CM, Welchans S: Bureau of Justice Statistics Special Report: Intimate Partner Violence, 1993–2001, US Department of Justice, 2000
http://www.ojp.usdoj.gov/bjs/pub/pdf/ipv.pdf

10. Connor EM, Sperling RS, Gerber R, et al: Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N Engl J Med 331:1173, 1994

11. Practice guideline for the treatment of patients with eating disorders (revision). American Psychiatric Association Work Group on Eating Disorders. Am J Psychiatry 157(1 suppl):1, 2000

12. Grunbaum JA, Kann L, Kinchen SA, et al: Youth risk behavior surveillance—United States, 2003. MMWR Surveill Summ 53:1, 2004

13. Edelman A, Gallo M, Jensen J, et al: Continuous or extended cycle vs. cyclic use of combined oral contraceptives for contraception. Cochrane Database Syst Rev (3): CD004695, 2005

14. Kaunitz AM: Beyond the pill: new data and options in hormonal and intrauterine contraception. Am J Obstet Gynecol 192:998, 2005

15. Raine TR, Harper CC, Rocca CH, et al: Direct access to emergency contraception through pharmacies and effect on unintended pregnancy and STIs: a randomized controlled trial. JAMA 293:54, 2005

16. Ehrmann DA: Polycystic ovary syndrome. N Engl J Med 352:1223, 2005

17. Tjaden P, Thoennes N: Prevalence, Incidence and Consequences of Violence Against Women: Findings From the National Violence Against Women Survey. National Institute of Justice, Centers for Disease Control and Prevention, 1998

18. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

19. Manson JE, Hsia J, Johnson KC, et al: Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med 349:523, 2003

20. Wassertheil-Smoller S, Hendrix SL, Limacher M, et al: Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. JAMA 289:2673, 2003

21. Shumaker SA, Legault C, Rapp SR, et al: Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 289:2651, 2003

22. Rapp SR, Espeland MA, Shumaker SA, et al: Effect of estrogen plus progestin on global cognitive function in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 289:2663, 2003

23. Hendrix SL, Cochrane BB, Nygaard IE, et al: Effects of estrogen with and without progestin on urinary incontinence. JAMA 293:935, 2005

24. Anderson GL, Limacher M, Assaf AR, et al: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 291:1701, 2004

25. Anderson GD: Sex and racial differences in pharmacological response: where is the evidence? Pharmacogenetics, pharmacokinetics, and pharmacodynamics. J Womens Health (Larchmt) 14:19, 2005

26. Meibohm B, Beierle I, Derendorf H: How important are gender differences in pharmacokinetics? Clin Pharmacokinet 41:329, 2002

27. Anthony M: Male/female differences in pharmacology: safety issues with QT-prolonging drugs. J Womens Health (Larchmt) 14:47, 2005

28. Rodriguez I, Kilborn MJ, Liu XK, et al: Drug-induced QT prolongation in women during the menstrual cycle. JAMA 285:1322, 2001

29. Rathore SS, Wang Y, Krumholz HM: Sex-based differences in the effect of digoxin for the treatment of heart failure. N Engl J Med 347:1403, 200

30. Rathore SS, Krumholz HM: Digoxin therapy for heart failure: safe for women? Ital Heart J 4:148, 2003

31. Archer JS, Archer DF: Oral contraceptive efficacy and antibiotic interaction: a myth debunked. J Am Acad Dermatol 46:917, 2002

32. Dickinson BD, Altman RD, Nielsen NH, et al: Drug interactions between oral con-

traceptives and antibiotics. Obstet Gynecol 98:853, 2001

33. Holt VL, Cushing-Haugen KL, Daling JR: Body weight and risk of oral contraceptive failure. Obstet Gynecol 99:820, 2002

34. Apple FS, Quist HE, Doyle PJ, et al: Plasma 99th percentile reference limits for cardiac troponin and creatine kinase MB mass for use with European Society of Cardiology/American College of Cardiology consensus recommendations. Clin Chem 49:1331, 2003

35. Mandelblatt JS, Schechter CB, Yabroff KR, et al: Toward optimal screening strategies for older women. J Gen Intern Med 20:487, 2005

36. Elmore JG, Armstrong K, Lehman CD, et al: Screening for breast cancer. JAMA 293:1245, 2005

37. Kriege M, Brekelmans CT, Boetes C, et al: Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. Magnetic Resonance Imaging Screening Study Group. N Engl J Med 351:427, 2004

38. Warner E, Plewes DB, Hill KA, et al: Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. JAMA 292:1317, 2004

39. Haas JS, Kaplan CP, Brawarsky P, et al: Evaluation and outcomes of women with a breast lump and a normal mammogram result. J Gen Intern Med 20:692, 2005

40. Solomon D, Davey D, Kurman R, et al: The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 287:2114, 2002

41. Wright TC Jr, Cox JT, Massad LS, et al: 2001 Consensus Guidelines for the management of women with cervical cytological abnormalities. JAMA 287:2120, 2002

42. The Guide to Clinical Preventive Services, 2005: Recommendations of the U.S. Preventive Services Task Force. Agency for Healthcare Research and Quality, US Department of Health and Human Services, 2005

43. Ridker PM, Cook NR, Lee IM, et al: A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med 352:1293, 2005

I AMENORRHEA

Robert L. Barbieri, m.d.

The menstrual cycle is orchestrated through the interaction of the hypothalamus, pituitary gland, ovaries, and uterus. The hypothalamus secretes gonadotropin-releasing hormone (GnRH), which stimulates the pituitary to secrete luteinizing hormone (LH) and follicle-stimulating hormone (FSH). LH and FSH stimulate follicular growth, ovulation, corpus luteum formation, and the secretion of estradiol and progesterone by the ovaries, which leads to the cyclic growth and shedding of the uterine endometrium. At the endometrial level, the menstrual cycle has three key phases: estradiol stimulates endometrial growth, progesterone induces differentiation of the endometrium, and withdrawal of estradiol and progesterone results in sloughing of the endometrium and menstrual bleeding.

Failure of any part of this process can result in the absence of menstruation. A girl may not start to menstruate when she reaches puberty (primary amenorrhea); or as is far more common, a woman who has been menstruating may have her cycles cease (secondary amenorrhea).

Pathophysiology

The hypothalamus contains approximately 10,000 GnRH-secreting neurons that drive the menstrual cycle by secreting pulses of GnRH. The embryonic precursors of the GnRH neurons develop in the olfactory bulb and migrate to the arcuate and preoptic nuclei. Improper development of the olfactory bulb in early embryogenesis can result in both anosmia and amenorrhea because of the absence of the GnRH neurons (Kallman syndrome).

The main function of the GnRH neurons is to receive neural signals from the brain and transform them into an endocrine output, the pulsatile release of GnRH. This conversion of electrical signal into endocrine output takes place in the arcuate nucleus. To determine the appropriate pulse frequency and amplitude of GnRH secretion, the hypothalamus monitors numerous environmental cues, including body composition, stress, nutritional status, and emotion. From a teleologic perspective, it is inefficient to ovulate and reproduce if the environment is hostile to the nurturing of a newborn.

The hypothalamus is the conductor that sets the tempo for the menstrual cycle. When the hypothalamus secretes GnRH at a low pulse frequency and amplitude, the pituitary gland is not driven to secrete LH and FSH, so the ovary and endometrium become quiescent. This causes amenorrhea. When the hypothalamus secretes GnRH at an abnormally elevated pulse frequency and amplitude, there is an exaggerated secretion of LH, causing the ovary to become androgenic and secrete testosterone. Follicular growth is blocked and no ovulation occurs. This results in the polycystic ovary syndrome (PCOS), which can be associated with oligomenorrhea or amenorrhea.

The pituitary gland is the main link between the brain and ovarian function. Secretion of the gonadotropins LH and FSH by the pituitary gland is not only stimulated by GnRH from the hypothalamus but also modulated by the negative feedback of steroid and protein hormones from the ovaries, especially estradiol, progesterone, and inhibin A and B [*see Figure 1*]. Estradiol and inhibin A are secreted by growing follicles and the corpus luteum. Inhibin B is secreted by the small antral follicles and growing follicles. Progestersone is secreted by the corpus luteum.

The follicle is the functional unit of the ovary. At puberty, the ovary contains approximately 300,000 follicles, of which only a few hundred will be ovulated in the woman's lifetime. The follicle has three components: an outer shell of thecal cells that respond to LH and secrete the androgen precursor androstenedione; an inner cell mass of granulosa cells that respond to FSH by converting androstenedione to estradiol; and, at the center of the follicle, the oocyte [*see Figure 2*]. Resting fol-

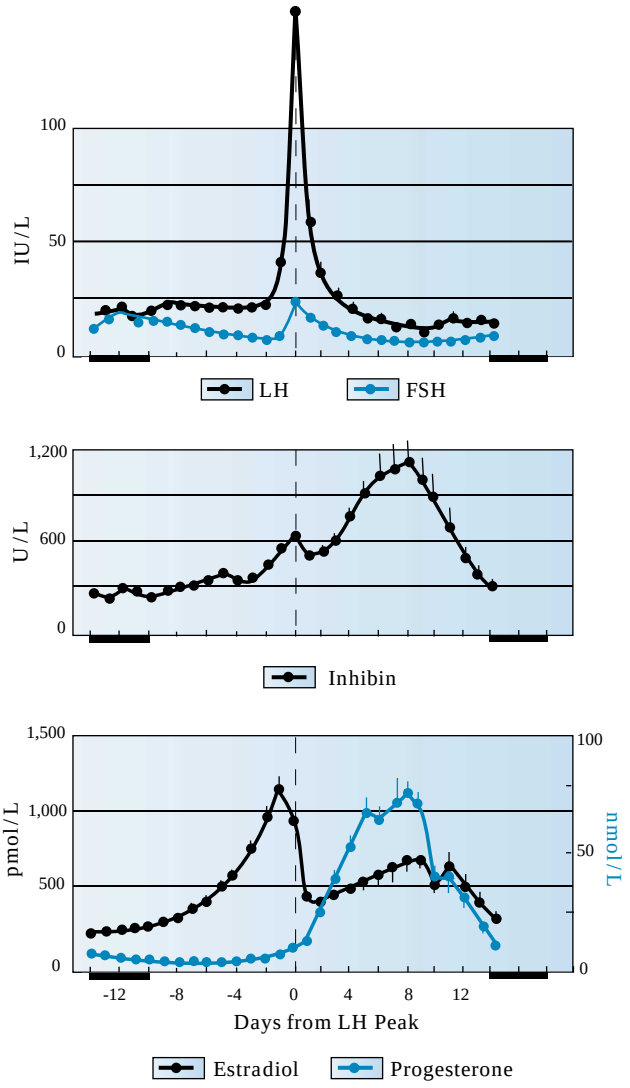


Figure 1 **Interaction between pituitary and ovarian hormones. The surge of luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which prompts ovulation, is followed by an increase in inhibin and progesterone secretion and a decrease in estradiol.[28]**

licles are recruited into a cohort of active follicles, only one of which will be destined to ovulate each cycle; the remainder undergo atresia. Under FSH stimulation, granulosa cell numbers increase dramatically, from approximately 10 cells in the primordial follicle to approximately 50 million cells in the preovulatory dominant follicle. The dominant follicle, which is the one destined to ovulate, can be identified early in its development by three characteristics: it has captured large amounts of FSH in its follicular fluid, it has the optimal number of granulosa cells for its size, and it produces much more estradiol than testosterone. Excess stimulation by LH results in an androgen-dominant follicle, which is characterized by thecal cell overactivity and a preference to produce testosterone over estradiol. In PCOS, the ovary contains many androgen-dominant follicles that do not have the potential to ovulate.

The purpose of the menstrual cycle is to generate a single oocyte for fertilization and to prepare the endometrium for implantation. Estradiol stimulates endometrial proliferation, gland formation, and vascular growth in the endometrium.

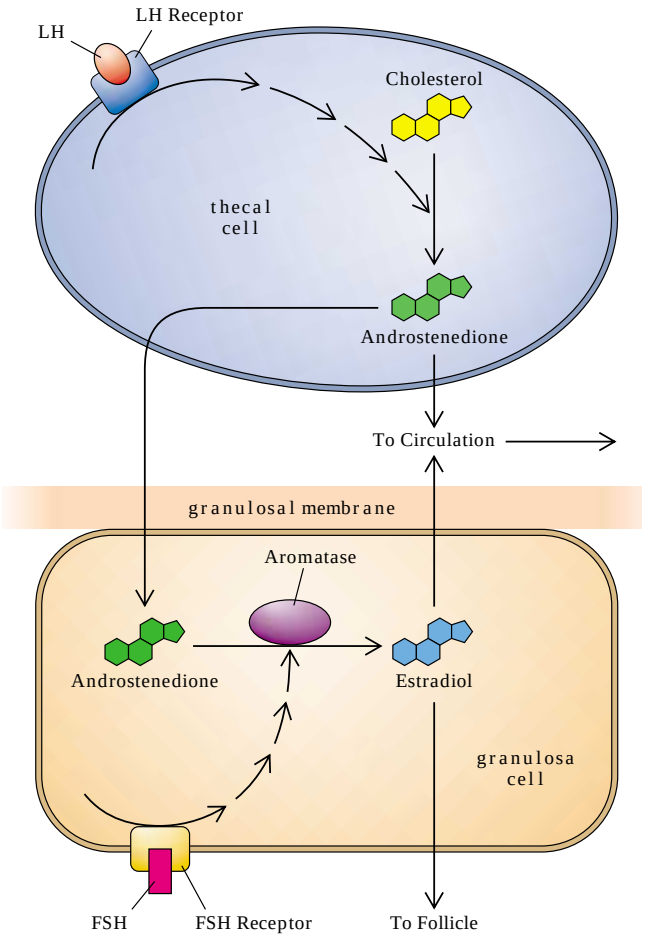


Figure 2 **Relationship between granulosa cells and thecal cells of the ovarian follicle. Luteinizing hormone (LH) stimulates thecal production of androstenedione, which diffuses to the granulosa cells. In the granulosa cells, follicle-stimulating hormone (FSH) stimulates the conversion of androstenedione to estradiol.[29]**

When stimulated by estradiol, the endometrium increases production of its own intracellular estrogen receptors, which augment its response to estradiol; it also produces more progesterone receptors. After ovulation, progesterone causes gland development and differentiation of the endometrium; in addition, glycogen is stored in preparation for embryo implantation. If pregnancy does not occur, the decline in ovarian production of estradiol and progesterone causes vasospasm of the endometrial blood vessels and sloughing of the endometrium, resulting in menstrual bleeding.

Abnormalities in GnRH secretion, LH or FSH secretion, ovarian follicular function, or endometrial function can cause amenorrhea or oligomenorrhea. In a given patient, the differential diagnosis—and hence the evaluation—depends on whether the amenorrhea is secondary or primary.

Secondary Amenorrhea

Secondary, or adult-onset, amenorrhea is present when a woman who had been menstruating has no menses for longer than three of her previous cycles, or 6 months. Determining the cause of secondary amenorrhea starts with measuring serum levels of human chorionic gonadotropin (hCG), prolactin, FSH, and testosterone and calculating the body mass index (BMI) [*see Figure 3*].

The most common cause of secondary amenorrhea is pregnancy. Pregnancy is best diagnosed by measuring the serum or urine hCG level. The hCG pregnancy test is one of the most accurate in medicine, with a sensitivity and specificity exceeding 99%.

In women who are not pregnant, the most common causes of secondary amenorrhea are as follows: hypothalamic dysfunction (low GnRH pulse frequency, amplitude, or both), pituitary dysfunction (low LH and FSH production), loss of all ovarian follicles (ovarian failure), PCOS, Asherman syndrome (intrauterine adhesions), and thyroid disease [*see Table 1*].[1]

amenorrhea caused by hypothalamic dysfunction

Low GnRH Secretion

Low BMI, vigorous exercise, psychosocial stress, and nutritional abnormalities decrease GnRH production. This reduces LH and FSH secretion and can cause amenorrhea.

Diagnosis The patient typically has a history of regular vigorous exercise, psychosocial stress, or reduced caloric intake. On physical examination, the BMI is often less than 20 kg/m^2 [*see Figure 4*]. Serum FSH, prolactin, and testosterone levels are usually reported as normal in women with secondary amenorrhea caused by low GnRH secretion. Women with secondary amenorrhea from hypothalamic hypofunction should be screened for eating disorders; the prevalence of eating disorders in this population is 5% to 10%. Rarely, hypothalamic dysfunction can be caused by structural abnormalities of the hypothalamus, including lymphoma, histiocytosis X, sarcoidosis, and hypothalamic cysts.

The absence of menses in association with hypothalamic hypofunction suggests severe estrogen deficiency, although varying degrees of hypoestrogenism may be present. The severity of the hypoestrogenism can be tested by performing a progestin withdrawal test [*see Figure 5*].[2] Serum estradiol assays are not sufficiently accurate for this purpose. Women with amenorrhea

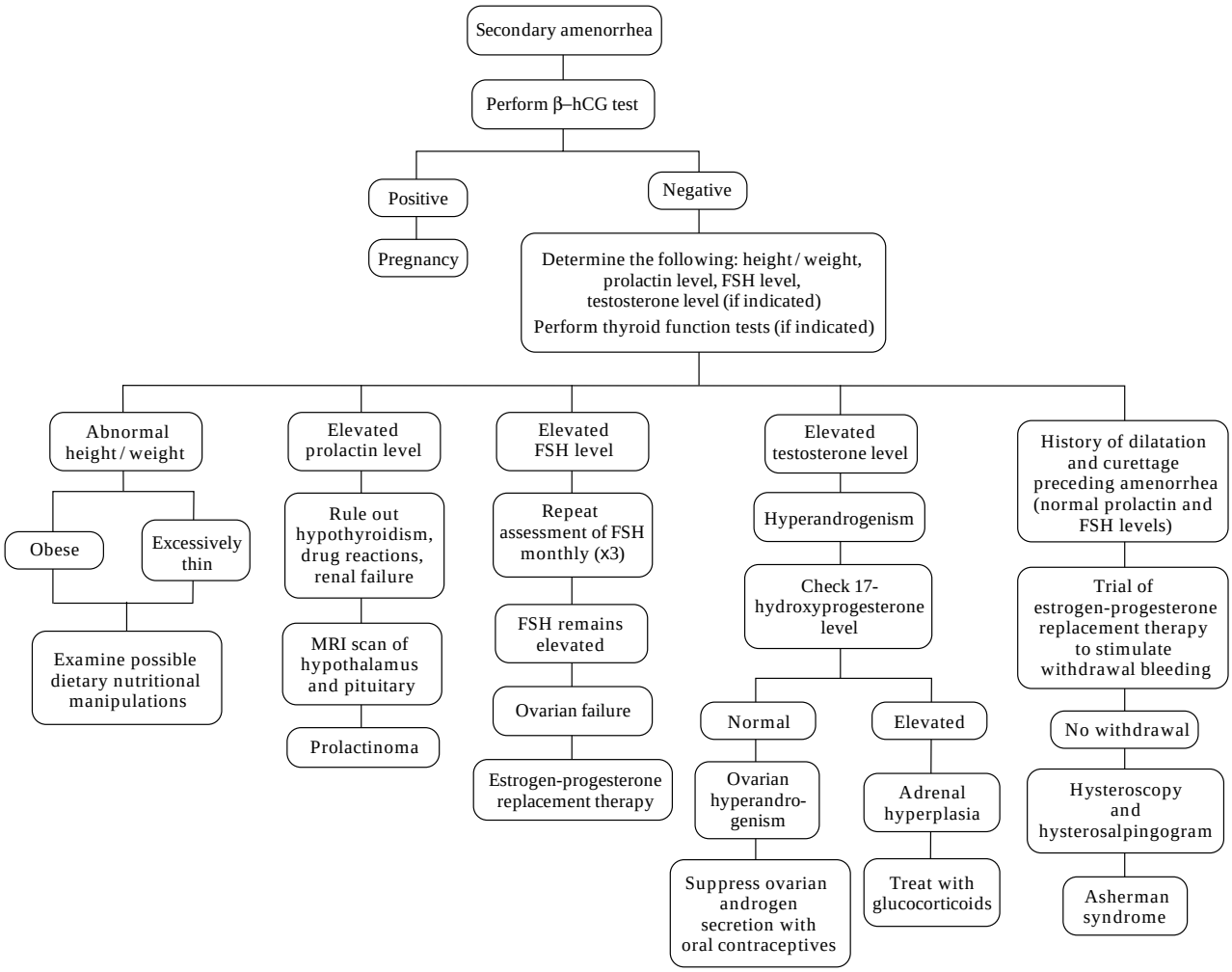


Figure 3 **Diagnosis and treatment of secondary amenorrhea. (β-hCG—β-human chorionic gonadotropin; FSH—follicle-stimulating hormone)**

resulting from hypothalamic hypofunction often have triiodothyronine (T_3) levels less than 70 ng / dl, reverse T_3 levels greater than 40 ng / dl (similar to those of nutritionally deprived individuals), and elevated levels of cortisol secretion (as seen in depressed women or women under significant stress).

Treatment Reversing the underlying cause of hypothalamic hypofunction (reducing psychosocial stress, gaining weight, lowering exercise intensity) often results in resumption of ovulatory menses. However, many women with amenorrhea from low GnRH secretion prefer to maintain the exercise and nutritional regimens that cause the amenorrhea, and those with an eating disorder may not respond to treatment of it. These women are best treated with hormone replacement. Therapeutic choices include oral contraceptives and low-dose hormone replacement, such as a standard combined continuous regimen of conjugated estrogens, 0.625 mg, plus medroxyprogesterone acetate, 2.5 mg, daily. Cyclic hormone replacement may also be used. The use of sustained-release vaginal gel to supply the progesterone component of hormone replacement therapy has been reported.[3] Vitamin D (400 IU / day) and calcium (1,200 to 1,500 mg / day) supplements should be administered to slow the rate of decline in bone mineral density associated with low estrogen levels.

amenorrhea caused by pituitary dysfunction

The most common pituitary diseases that cause secondary amenorrhea are prolactin-secreting pituitary tumors (prolactinomas), the empty sella syndrome, Sheehan syndrome, and other pituitary tumors, such as those that secrete adrenocorticotropic hormone (ACTH) or growth hormone.

Prolactin-Secreting Pituitary Tumors

Most pituitary tumors are monoclonal, which indicates that they arise from a somatic mutation in a single progenitor cell. In general, pituitary tumors are benign and slow growing.

Diagnosis The most common causes of an elevation in the serum prolactin level are, in order of frequency, advanced pregnancy; the use of psychotropic medications that are

Table 1 The Most Common Causes of Secondary Amenorrhea in Women Who Are Not Pregnant

Organ	*Cause*	*Relative Frequency (%)*
Hypothalamus	Abnormalities of height / weight and nutrition	15
	Exercise	10
	Psychosocial stress	10
	Infiltrative disease or tumors of the hypothalamus (sarcoidosis, histiocytosis, craniopharyngioma)	< 0.1
Pituitary	Prolactin-secreting pituitary tumor	17
	Empty sella syndrome	1
	Sheehan syndrome	< 1
	ACTH-secreting tumor (Cushing disease)	< 1
	GH-secreting tumor	< 1
Ovary	Premature ovarian failure	10
	Polycystic ovary syndrome	30
Uterus	Asherman syndrome (intrauterine synechiae)	5
Other	Nonclassical adrenal hyperplasia	< 1
	Thyroid disease	1
	Ovarian tumors	< 1

ACTH—adrenocorticotropic hormone GH—growth hormone

dopamine antagonists, such as haloperidol; prolactin-secreting pituitary tumors; hypothyroidism; and renal failure. After excluding those causes, the workup should focus on the pituitary.

A magnetic resonance imaging scan of the hypothalamus and pituitary can confirm the diagnosis of prolactinoma. The MRI is also used to determine whether the diameter of the tumor is less than 10 mm (microprolactinoma) or greater than 10 mm (macroprolactinoma), because this measurement has clinical implications. Finally, the MRI can assess for possible involvement of the sella turcica and the optic chiasm.

If the MRI shows a pituitary tumor, it is important to also measure serum insulinlike growth factor–1 (IGF-1). IGF-1 levels will be elevated in patients whose pituitary tumor secretes growth hormone; this is a more reliable test than measurement of growth hormone itself.

Treatment In general, microprolactinomas have a benign course and can be managed by the patient's primary care physician. Observational studies indicate that over a period of 4 to 6 years, 95% of microprolactinomas do not increase in size.[4,5] Macroprolactinomas, however, can be associated with significant complications, such as pituitary apoplexy and compression of the optic chiasm, and should be managed by an endocrinologist. The initial treatment of both microprolactinomas and macroprolactinomas should be medical therapy, not surgery.

The two best approaches to management of microprolactinomas in women with amenorrhea are low-dose oral contraceptives and a dopamine agonist (bromocriptine, pergolide, or cabergoline). Both contraceptives and dopamine agonists can initiate regular withdrawal bleeding and prevent osteoporosis. In women with microprolactinomas, treatment with an estrogen-progestin oral contraceptive is safe and is not associated with clinically significant tumor growth.[6]

Women with amenorrhea caused by a prolactinoma who wish to become pregnant should receive treatment with a dopamine agonist to induce ovulation. Dopamine agonists directly suppress prolactin production by the tumor and cause an increase in endogenous GnRH secretion, which stimulates pituitary secretion of LH and FSH and consequently induces follicle development and ovulation. In addition, dopamine agonists decrease the size of prolactin-secreting pituitary tumors.[7,8]

Bromocriptine has been used to induce ovulation in women with hyperprolactinemia for more than 25 years.[9] In one study of 280 women with hyperprolactinemia, bromocriptine normalized the circulating prolactin level in 82% of the women.[10] The main side effects associated with the use of bromocriptine are nausea, vomiting, and orthostatic hypotension. To minimize these potential side effects, it is recommended that bromocriptine be initiated at a dose of 1.25 mg at bedtime. After 1 week, the dosage can be increased to 1.25 mg twice daily. The dosage can then be increased to 2.5 mg twice daily, a standard dosage that successfully reduces serum prolactin in most women with hyperprolactinemia.[10] Long-acting oral and injectable forms of bromocriptine have been developed,[11,12] but those are not yet available in the United States.

Pergolide, an ergot dopamine agonist, has been approved by the Food and Drug Administration for the treatment of Parkinson disease but not for the treatment of hyperprolactinemia. Unlike bromocriptine, pergolide can be given once a day. Pergolide is the least expensive of the dopamine agonists; its cost is about one sixth that of cabergoline.

Cabergoline is a non-ergot dopamine agonist that is administered once or twice a week and causes less nausea than bromocriptine or pergolide.[13] The FDA-approved dosage of cabergoline is 0.25 mg twice a week. Many clinicians start cabergoline at a dosage of 0.5 mg a week, then increase the dosage to 1 mg once or twice a week, depending on the response of the serum prolactin level (see below). In about 25% of women, the serum prolactin level returns to normal through therapy with cabergoline at a dosage of 1 mg a week; in these patients, the dosage can be reduced to 0.5 mg a week and the serum prolactin level will remain normal. About one half of women who do not respond to bromocriptine treatment will respond to treatment with cabergoline.[14] Many authorities believe that cabergoline is more effective than bromocriptine in treating hyperprolactinemia.[15] In a series of 459 women with hyperprolactinemia and amenorrhea, 83% of the women treated with cabergoline experienced normalization of their prolactin levels, compared with 52% of those treated with bromocriptine.[16] Cabergoline is significantly more expensive than bromocriptine, however.

With dopamine agonist therapy, near-maximal decreases in serum prolactin levels are typically achieved after 4 weeks of treatment. Serum prolactin levels should be measured approximately 1 month after initiating therapy and about 1 month after a change in dose or drug. If the serum prolactin concentration is normal and no side effects have occurred, the initial dose should be continued. If serum prolactin has not decreased to normal and no side effects are present, the dose should be gradually increased. Maximal dosages of the dopamine agonists are as follows: bromocriptine, 5 mg twice daily; pergolide, 0.25 mg once daily; and cabergoline, 1.5 mg two or three times weekly. If the serum prolactin level does not decrease to normal, switching to a different dopamine agonist may be effec-

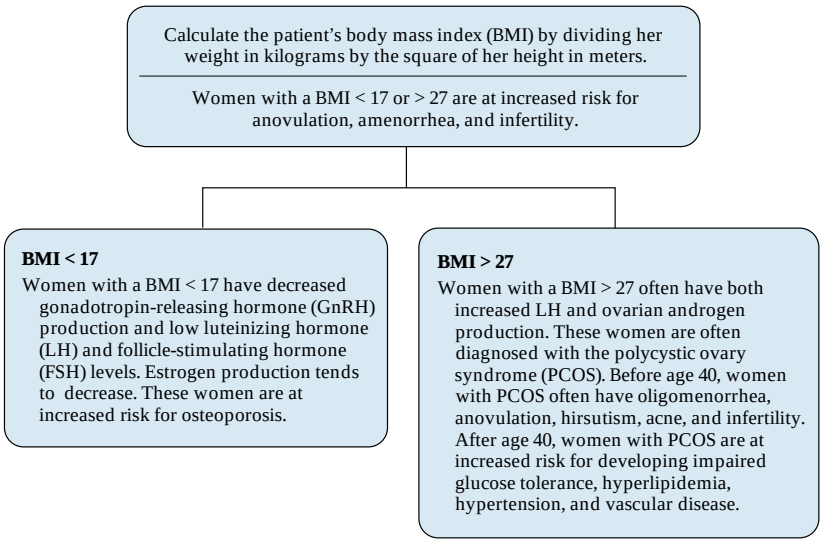


Figure 4 **Relation of body mass index to anovulation.**

tive. If the patient cannot tolerate the side effects of the dopamine agonist initially prescribed, a different dopamine agonist may be tried. If the patient experiences side effects with all the dopamine agonists, then vaginal administration of bromocriptine can be tried.[17] If all attempts at medical therapy fail, transsphenoidal surgery is indicated. Successful removal of the tumor will result in the normalization of the prolactin secretion level and resumption of ovulatory menses.

After correction of hyperprolactinemia, about 80% of women will ovulate; cumulative pregnancy rates of 80% are commonly observed.[18] Treatment is usually discontinued once a pregnancy is diagnosed. However, in women with a macroprolactinoma, therapy should be continued throughout pregnancy to reduce the risk that the tumor will grow and cause neurosurgical complications, such as compression of the optic nerve.

Empty Sella Syndrome

The roof of the pituitary gland (the diaphragm of the sella) is perforated by the pituitary stalk, which connects the hypothalamic median eminence to the pituitary. If the perforation in the diaphragm of the sella is excessively large, the pia mater and accompanying cerebrospinal fluid can herniate into the pituitary fossa. Herniation of this fluid, which is under reasonably high pressure, can produce compression atrophy of the pituitary gland, resulting in hypopituitarism and amenorrhea.

The empty sella syndrome can be diagnosed on the basis of high-resolution MRI or computed tomography of the pituitary. Therapy is directed to the specific replacement of documented hormonal abnormalities.

Sheehan Syndrome

Sheehan syndrome is the onset of hypothalamic and pituitary dysfunction after severe obstetric hemorrhage and maternal hypotension at delivery. During pregnancy, the pituitary volume increases by approximately 100%. The increase in pituitary size and the low-flow, low-pressure nature of the portal circulation may make the pituitary and parts of the hypothalamus susceptible to ischemia brought on by obstetric hemorrhage and hypotension. Worldwide, Sheehan syndrome is the most common cause of hypopituitarism.

Diagnosis Every possible pattern of pituitary hormone deficiency has been reported in Sheehan syndrome, but growth hormone and prolactin deficiencies are the most common presentations. In a study of 10 African women with Sheehan syndrome, all had both prolactin and growth hormone deficiency, nine had cortisol deficiency, eight had TSH deficiency, seven had LH deficiency, and four had FSH deficiency.[19] Clinical studies have demonstrated that many women with Sheehan syndrome also have mild defects in vasopressin secretion.[20] The best test to diagnose Sheehan syndrome is to administer thyrotropin-releasing hormone (TRH), 100 µg intravenously, and measure prolactin at 0 and 30 minutes. The ratio of prolactin at 30 minutes to prolactin at 0 minutes should be greater than 3.[21] If the ratio is less than 3, the patient should undergo a complete evaluation for panhypopituitarism.

Treatment Treatment of Sheehan syndrome is with hormone replacement. The particular regimen is based on the patient's pattern of hormone deficiency.

Premature Ovarian Failure

Ovarian failure is the loss of all functional ovarian follicles. Ovarian failure in patients younger than 40 years is termed premature ovarian failure. The causes of premature ovarian failure include genetic abnormalities (e.g., microdeletions of the X chromosome), autoimmune processes (e.g., polyglandular autoimmune disease and myasthenia gravis), chemotherapy (especially with alkylating agents), and pelvic radiotherapy (> 500 cGy to the ovaries).

Diagnosis Loss of all follicles results in a decrease in estradiol and inhibin B production. In the absence of the negative feedback effect of these two hormones, FSH secretion increases markedly. Therefore, ovarian failure is most accurately diagnosed by measurement of serum FSH. Complete ovarian failure is associated with serum FSH levels greater than 25 U/L. In

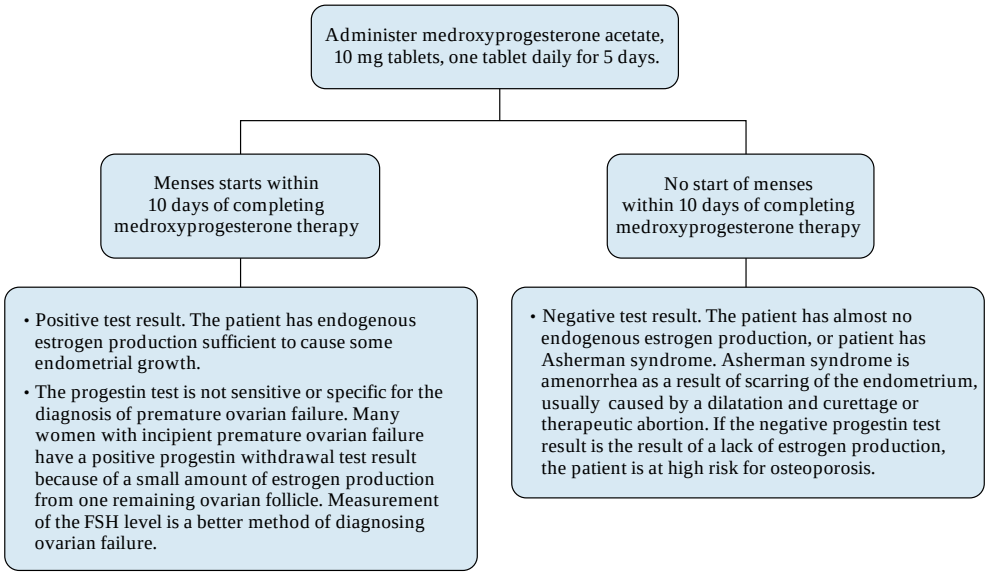


Figure 5 **The progestin withdrawal test. (FSH—follicle-stimulating hormone)**

women with incipient ovarian failure, FSH levels are often between 15 and 25 U/L and can fluctuate.[22] Clinically, women with premature ovarian failure often experience vasomotor symptoms (hot flashes) or vaginal dryness.

Treatment There are no proven therapies specifically for ovarian failure. Because women with premature ovarian failure (like all women with estrogen deficiency) are at high risk for osteoporosis, they should be treated with estrogen-progestin replacement. Although women with ovarian failure are by definition infertile, they may be able to bear a child through oocyte donation. In this process, a donor egg and sperm from the woman's partner are incubated in vitro. The embryo is then transferred to the woman's uterus, after she undergoes treatment with estrogen and progesterone to mature the endometrium.

Hyperandrogenism and the Polycystic Ovary Syndrome

Hyperandrogenism is marked by clinical evidence of excess androgens, including severe hirsutism, and elevated serum androgen levels. PCOS causes 80% of hyperandrogenism; idiopathic hirsutism causes 15%, and nonclassical adrenal hyperplasia (NCAH) causes 4%. Rare causes, such as androgen-secreting ovarian or adrenal tumors, account for 1% of cases. Idiopathic hirsutism is hirsutism associated with regular menses and therefore is not a cause of amenorrhea. PCOS, NCAH, and androgen-secreting tumors can all cause amenorrhea. Women who have amenorrhea with hyperandrogenism are at increased risk for endometrial carcinoma.

Diagnosis Androgen-secreting tumors are usually associated with circulating testosterone levels greater than 200 ng/dl. In menopausal women, small ovarian hilar cell androgen-secreting tumors can be associated with elevations in the serum testosterone level in the range of 150 to 200 ng/dl. In most women with PCOS, the serum testosterone level ranges from 50 and 200 ng/dl. The 17-hydroxyprogesterone level, when measured at 8:00 A.M., is greater than 4 ng/ml in women with NCAH and is less than 4 ng/ml in women with PCOS. A sonogram that shows multiple small ovarian follicles can support the diagnosis of PCOS.[22]

Treatment Ovarian hyperandrogenism, such as PCOS, is treated by suppression of ovarian androgen secretion with oral contraceptives. Adrenal hyperplasia is treated with glucocorticoids.

Table 2 The Most Common Causes of Primary Amenorrhea

Cause	*Rate of Occurrence (%)*
Gonadal dysgenesis, including 45X (Turner syndrome)	45
Physiologic delay of puberty	20
Müllerian agenesis	15
Obstructed outflow tract: transverse vaginal septum or imperforate hymen	5
Absence of hypothalamic gonadotropin-releasing hormone production (Kallmann syndrome)	5
Anorexia nervosa	2
Hypopituitarism	2
Androgen insensitivity	1
Hyperprolactinemia	1
Adrenal hyperplasia	1
Hypothyroidism	1
Pituitary tumors	1
Craniopharyngioma	1

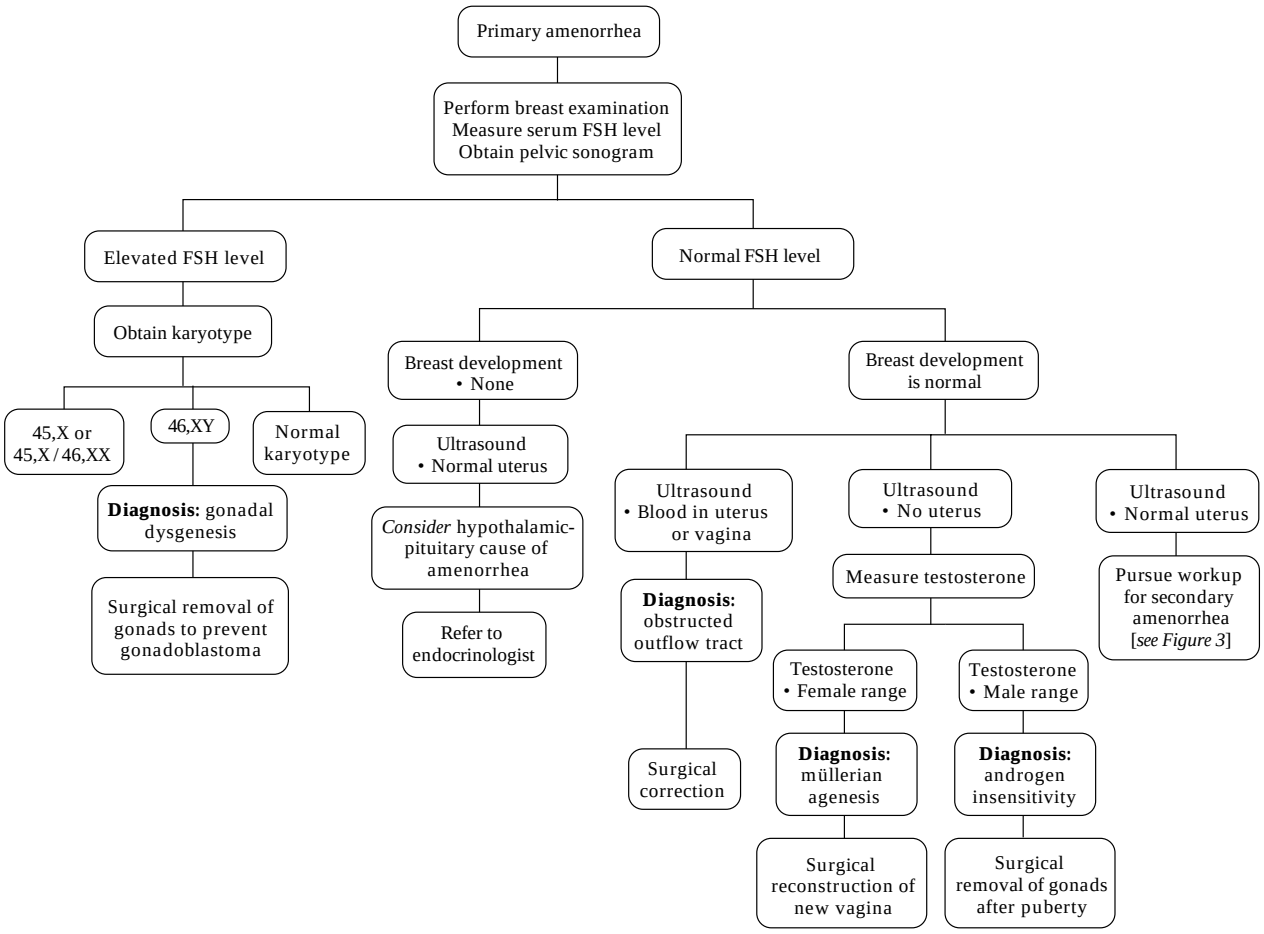


Figure 6 **Diagnosis and treatment of primary amenorrhea. (FSH—follicle-stimulating hormone)**

Asherman Syndrome (Intrauterine Adhesions)

Asherman syndrome is the presence of intrauterine scar tissue that interferes with normal endometrial growth and shedding. In women with this condition, intrauterine scar tissue usually develops after vigorous curettage of infected endometrium early in pregnancy.

Diagnosis The history is an important clue to Asherman syndrome. For patients whose history is suggestive, a commonly used way to assess endocrine function is to prescribe conjugated estrogens, 2.5 mg for 35 days, plus medroxyprogesterone acetate, 10 mg daily, on days 26 to 35. The estrogen and progestin are then discontinued. Absence of withdrawal bleeding after such a challenge strongly suggests Asherman syndrome. The diagnosis can be confirmed radiologically by hysterosalpingogram or by direct visualization of the scar tissue with hysteroscopy.

Treatment The treatment of Asherman syndrome involves surgical lysis of the intrauterine adhesions by operative hysteroscopy, followed by stimulation of endometrial growth with estrogen. Some women who become pregnant after treatment experience placental defects such as placenta accreta. This is probably the result of disruption of function of the endometrial stroma by the disease and treatment.

Primary Amenorrhea

Primary amenorrhea is present when the first menses has not occurred by the time a girl reaches 16 years of age. Girls who reach 14 years of age and have no breast development should also undergo a workup similar to that for primary amenorrhea.[23] Primary amenorrhea is usually caused by genetic or congenital disorders and is often associated with developmental problems during puberty.[24] These developmental problems include delayed height and weight gain, delayed breast development, and delayed development of pubic and axillary hair.

In a large case series, the most common causes of primary amenorrhea were found to be gonadal dysgenesis resulting from chromosomal abnormalities such as 45,X, 45,X/46,XX, and 46,XY (45% of cases); physiologic delay of puberty (20% of cases); müllerian agenesis (15% of cases); transverse vaginal septum or imperforate hymen (5% of cases); absence of hypothalamic GnRH production, such as seen in Kallman syndrome (5% of cases); anorexia nervosa (2% of cases); and hypopituitarism (2% of cases).[25] Less common causes of primary amenorrhea are hyperprolactinemia, hypothyroidism, pituitary tumors, Cushing disease, and craniopharyngiomas [*see Table 2*].

Primary amenorrhea is evaluated by focusing on the serum FSH level; breast development; and the presence of the cervix

and uterus, as determined by an imaging study (pelvic sonogram or MRI) [*see Figure 6*].

If the FSH level is elevated, a diagnosis of gonadal dysgenesis can be made in most cases. A karyotype should be obtained in such cases to determine whether a Y chromosome is present. Girls with gonadal dysgenesis and a Y chromosome on the karyotype have a high rate of malignant transformation of the gonad to a dysgerminoma or a gonadoblastoma. In these cases, the ovaries need to be removed surgically.

If the FSH level is normal, then breast development should be assessed, and it should be determined whether a uterus and a cervix are present. If breast development is absent and a uterus is present, the differential diagnosis includes hypothalamic and pituitary causes of amenorrhea; the patient should then be referred to an endocrinologist. If breast development is normal and a uterus is present, the differential diagnosis is the same as that for secondary amenorrhea (see above). However, if the pelvic ultrasound shows not only that the patient has a uterus but that the uterus or vagina is dilated by blood, the diagnosis is a blocked reproductive outflow tract (transverse vaginal septum or imperforate hymen). Surgical relief of the outflow tract is necessary to prevent the development of hematocolpos, hematometria, and endometriosis.[26]

If breast development is present and no uterus is present on ultrasound, then the serum testosterone level should be assessed. If the testosterone level is in the normal range for a girl, the likely diagnosis is müllerian agenesis.[27] Treatment of müllerian agenesis is surgical construction of a new vagina, which should be scheduled for the time when the patient is expected to begin having sexual relations.

In a patient with breast development, no uterus, and a testosterone level that is elevated to the male range, the likely diagnosis is androgen insensitivity. Females with androgen insensitivity usually do not have pubic or axillary hair. Breast development is usually normal because in the absence of a tissue androgen receptor, low levels of estrogen have a markedly stimulating effect. Although the karyotype in these patients is 46,XY, their likelihood of developing a gonadoblastoma is low until after puberty. Removal of the gonads is typically delayed until puberty is complete.

References

1. Reindollar RH, Novak M, Tho SP, et al: Adult-onset amenorrhea: a study of 262 patients. Am J Obstet Gynecol 155:531, 1986

2. Hull MG, Knuth UA, Murray MA, et al: The practical value of the progestogen challenge test, serum oestradiol estimation or clinical examination in assessment of the oestrogen state and response to clomiphene in amenorrhoea. Br J Obstet Gynaecol 86:799, 1979

3. Warren MP, Biller BM, Shangold MM: A new clinical option for hormone replacement therapy in women with secondary amenorrhea: effects of cyclic administration of progesterone from the sustained-release vaginal gel Crinone (4% and 8%) on endometrial morphologic features and withdrawal bleeding. Am J Obstet Gynecol 180:42, 1999

4. Sisam DA, Sheehan JP, Sheeler LR: The natural history of untreated microprolactinoma. Fertil Steril 48:67, 1987

5. Schlechte J, Dolan K, Sherman B, et al: The natural history of untreated hyperprolactinemia. J Clin Endocrinol Metab 68:412, 1989

6. Corenblum B, Donovan L: The safety of physiological estrogen plus progestin replacement therapy and with oral contraceptive therapy in women with pathological hyperprolactinemia. Fertil Steril 59:671, 1993

7. Shimon I, Melmed S: Management of pituitary tumors. Ann Intern Med 129:472, 1998

8. Touraine P, Plu-Bureau G, Beji C, et al: Long-term follow-up of 246 hyperprolactinemic patients. Acta Obstet Gynecol Scand 80:162, 2001

9. Molitch ME: Disorders of prolactin secretion. Endocrinol Metab Clin North Am 30:585, 2001

10. Vance ML, Evans WS, Thorner MO: Drugs five years later: bromocriptine. Ann Intern Med 100:78, 1984

11. Zgliczynski W, Zgliczynski S, Makowska A, et al: New long-acting bromocriptine (Parlodel MR and Parlodel LAR) in the treatment of pituitary tumours with hyperprolactinemia. Endokrynol Pol 43:234, 1992

12. Jamrozik SI, Bennet AP, James-Deidier A, et al: Treatment with long acting repeatable bromocriptine (Parlodel-LAR) in patients with macroprolactinomas: long-term study in 29 patients. J Endocrinol Invest 19:472, 1996

13. Biller BM, Molitch ME, Vance ML, et al: Treatment of prolactin-secreting macroadenomas with the once-weekly dopamine agonist cabergoline. J Clin Endocrinol Metab 81:2338, 1996

14. Verhelst J, Abs R, Maiter D, et al: Cabergoline in the treatment of hyperprolactinemia. J Clin Endocrinol Metab 84:2518, 1999

15. Di Sarno A, Landi ML, Cappabianca P, et al: Resistance to cabergoline as compared with bromocriptine in hyperprolactinemia: prevalence, clinical definition, and therapeutic strategy. J Clin Endocrinol Metab 186:5256, 2001

16. Webster J, Piscitelli MD, Polli A, et al: A comparison of cabergoline and bromocriptine in the treatment of hyperprolactinemic amenorrhea. N Engl J Med 331:904, 1994

17. Ricci G, Giolo E, Nucera G, et al: Pregnancy in hyperprolactinemic infertile women treated with vaginal bromocriptine: report of two cases and review of the literature. Gynecol Obstet Invest 51:266, 2001

18. Crosignani PG: Management of hyperprolactinemia in infertility. J Reprod Med 44(suppl):1116, 1999

19. Jialal I, Maidoo C, Norman RJ, et al: Pituitary function in Sheehan's syndrome. Obstet Gynecol 63:15, 1984

20. Jialal I, Desai RK, Rajput MC: An assessment of posterior pituitary function in patients with Sheehan's syndrome. Clin Endocrinol (Oxf) 27:91, 1987

21. Barbieri RL, Cooper DS, Daniels GH, et al: Prolactin response to thyrotropin-releasing hormone (TRH) in patients with hypothalamic-pituitary disease. Fertil Steril 43:66, 1985

22. Lewis V: Polycystic ovary syndrome: a diagnostic challenge. Obstet Gynecol Clin North Am 28:1, 2001

23. Lee PA, Guo SS, Kulin HE: Age of puberty: data from the United States of America. APMIS 109:81, 2001

24. Pletcher JR, Slap GB: Menstrual disorders: amenorrhea. Pediatr Clin North Am 46:505, 1999

25. Reindollar RH, Byrd JR, McDonough PG: Delayed sexual development: a study of 252 patients. Am J Obstet Gynecol 140:371, 1981

26. Valdes C, Malini S, Malinak LR: Ultrasound evaluation of female genital tract anomalies: a review of 64 cases. Am J Obstet Gynecol 149:285, 1984

27. Buttram VC Jr, Gibbons WE: Müllerian anomalies: a proposed classification. (An analysis of 144 cases.) Fertil Steril 32:40, 1979

28. Roseff SJ, Bangah ML, Kettel LM, et al: Dynamic changes in circulating inhibin levels during the luteal-follicular transition of the human menstrual cycle. J Clin Endocrinol Metab 69:1033, 1989

29. Ryan KJ: The endocrine pattern and control of the ovulatory cycle. Infertility: Male and Female. Insler V, Lunenfeld B, Eds. Churchill Livingstone, New York, 1986

II ABNORMAL VAGINAL BLEEDING

Alan H. DeCherney, m.d.

Definition

Abnormal vaginal bleeding is classified on the basis of the duration, amount, and frequency of bleeding. The duration of a menstrual cycle is defined as the length of time between the first day of bleeding of one menstrual cycle and the first day of bleeding of the next cycle. The normal menstrual cycle lasts about 28 days (plus or minus 7 days). Any deviation in cycle duration is considered abnormal. Polymenorrhea refers to an abnormally shortened menstrual cycle, with bleeding occurring every 21 days or sooner. A cycle duration of 35 days or longer is also considered abnormal and is termed oligomenorrhea.

Women with normal menses bleed for a maximum of 7 days, losing a total of 60 ml of blood. Menstruation that occurs at regular intervals but lasts for more than 7 days and involves the loss of more than 80 ml of blood with each menstrual cycle is called menorrhagia. In contrast, women with a decrease in monthly blood loss are considered to have hypomenorrhea.

Intermenstrual bleeding, metrorrhagia, and menometrorrhagia are other abnormal bleeding patterns. Intermenstrual bleeding is bleeding that occurs between regular menstrual periods. Metrorrhagia refers to bleeding that is frequent and irregular. When this pattern of bleeding becomes prolonged, it is called menometrorrhagia.

Etiology

Abnormal vaginal bleeding occurs in women of all ages and can result from a number of causes, including anatomic abnormalities, pregnancy complications, malignancies, infections, systemic diseases, and endocrinologic disturbances [*see Table 1*]. The physician must be able to recognize all causes of this condition and treat the specific underlying cause. The bleeding pattern experienced by a woman and the age at which it occurs may be indicative of the cause. It is thus useful to distinguish between premenopausal vaginal bleeding [*see Figure 1*] and postmenopausal vaginal bleeding [*see Figure 2*]. Once the cause of bleeding is identified, treatment can be started and tailored to meet the needs of the patient.[1,2]

Premenopausal Vaginal Bleeding

PREGNANCY

An abnormal pregnancy is the first possibility to consider when a patient in her reproductive years presents with abnormal vaginal bleeding. The complications of pregnancy associated with abnormal vaginal bleeding include threatened, incomplete, and undetected abortions; ectopic pregnancies; trophoblastic disease (a benign or malignant tumor arising in placental tissue); and other abnormalities of placental location. These conditions can be fatal if not treated in an expeditious and judicious manner. Thus, early diagnosis and treatment are critical [*see 16:VIII Ectopic Pregnancy and Spontaneous Abortion*].

NONPREGNANT PATIENTS

A detailed history and physical examination are essential for making an accurate diagnosis. Inspection of the vagina and external genitalia can easily establish the diagnosis in patients whose bleeding is caused by either trauma or the presence of a foreign body, as well as in some patients with anatomic abnormalities. Treatment in such cases is specific to the lesion.

Adjunctive blood tests (e.g., for anemia or clotting abnormalities) and imaging studies are also essential in most patients. Pelvic ultrasonography and hysteroscopy may be useful in differentiating anatomic from nonanatomic causes of bleeding.[3,4]

In a premenopausal woman older than 35 years who experiences changes in her menstrual bleeding pattern, a Papanicolaou (Pap) smear and endometrial biopsy should also be obtained. These tests will reveal whether premalignant or malignant cells are present. Although these tumors are more common in postmenopausal women, premenopausal women should always be evaluated.

Anatomic abnormalities of the uterus (e.g., polyps, uterine myomas, adenomyosis, and endometriosis) should be considered when abnormal bleeding occurs in an ovulatory woman. Pelvic ultrasonography, hysterosalpingography, or diagnostic laparoscopy or hysteroscopy may aid in the diagnosis of anatomic uterine abnormalities. Surgical correction of the anatomic problem is usually necessary to alleviate and ultimately resolve the bleeding disorder.

Reproductive Tract Diseases

Malignancy Vaginal bleeding may be an initial symptom of malignancies involving the cervix, endometrium, vagina, vulva, or ovaries. The incidence of these malignant tumors increases with advancing age. Identification of a malignancy warrants intervention with radiotherapy, chemotherapy, surgery, or a combination of these modalities.

Infection Lower or upper genital tract infection may result in abnormal bleeding. Urethritis, cervicitis, erosions of the cervix

Table 1 Causes of Abnormal Uterine Bleeding

- Anatomic abnormalities of the uterus
  - Polyps
  - Uterine myomas
  - Adenomyosis
  - Endometriosis
- Pregnancy complications
- Malignancy
  - Cervical
  - Endometrial
  - Vaginal
  - Vulvar
  - Ovarian
- Infection
  - Urethritis
  - Cervicitis
  - Cervical or vaginal erosions
  - Pelvic inflammatory disease
- Systemic disease
  - Coagulation defects
  - Liver failure
  - Thyroid dysfunction
- Trauma
- Foreign bodies
- Medications
  - Long-acting contraceptives
  - Digitalis
  - Phenytoin
  - Anticoagulants
  - Corticosteroids
  - Psychotropic agents

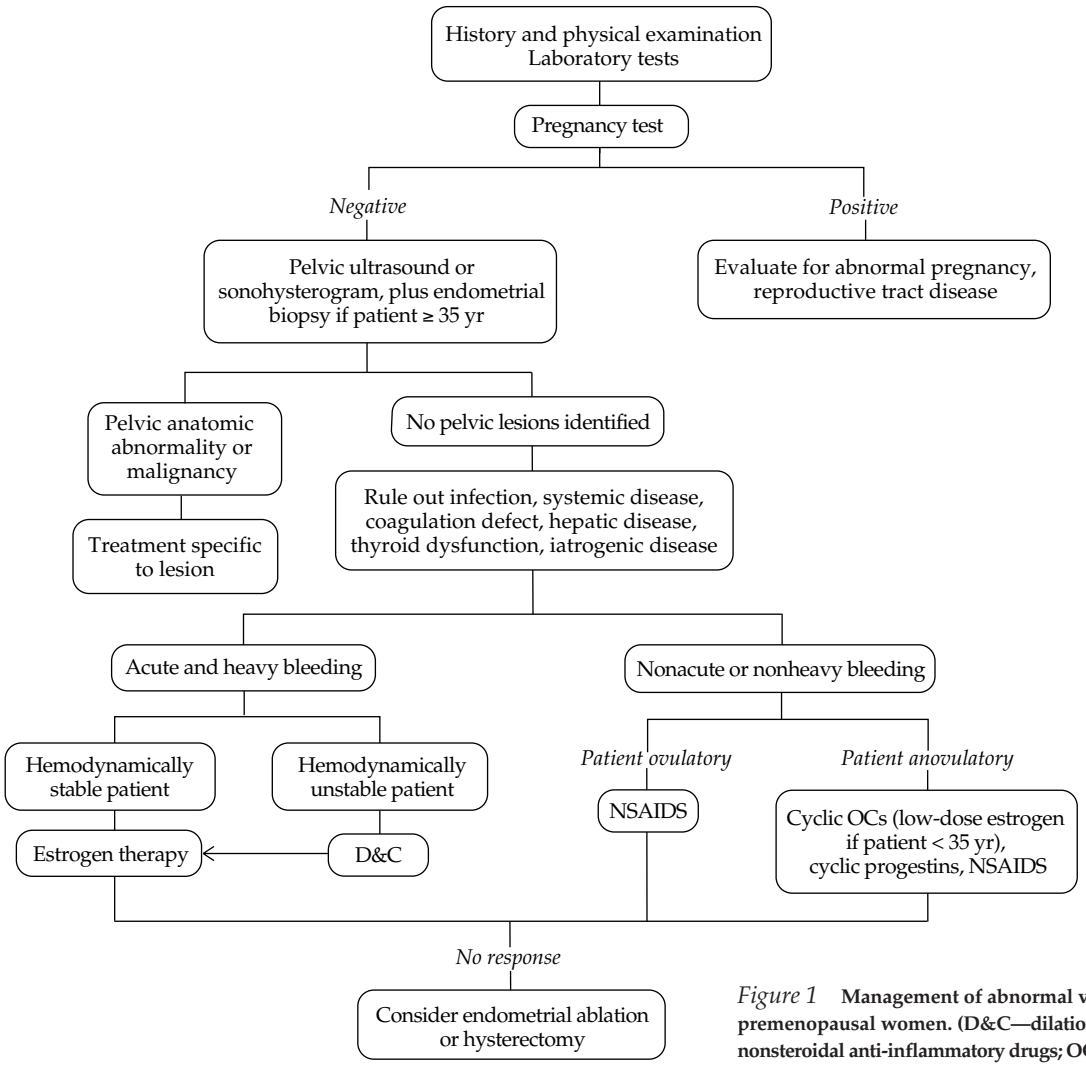


Figure 1 **Management of abnormal vaginal bleeding in premenopausal women. (D&C—dilation and curettage; NSAIDs—nonsteroidal anti-inflammatory drugs; OCs—oral contraceptives)**

and vagina, and emerging pelvic inflammatory disease can cause menstrual irregularities. Affected patients often experience pelvic and lower abdominal pain, fever, chills, and bloody vaginal discharge.

Diagnosis is usually made with vaginal and cervical cultures and an endometrial biopsy. Ultimately, treatment is targeted at the causative organism [*see 7:XXII Sexually Transmitted Diseases*].

Systemic Disease

Systemic diseases, including coagulation defects, liver failure, and thyroid dysfunction, can manifest as abnormal vaginal bleeding. In such cases, other manifestations typically point to the diagnosis.

Coagulation defects Patients with coagulopathies usually bruise easily and bleed from numerous sites, including the nose, gingiva, uterus, urinary tract, and gastrointestinal tract. The cessation of menstruation relies partly on a normal coagulation profile and the body's ability to form clots. Thrombocytopenia, von Willebrand disease, and other disorders of coagulation disrupt this natural process. Because coagulation disorders generally become apparent at the time of menarche, adolescents who display uncontrolled menorrhagia should be evaluated for these conditions[5] [*see 5:VI Disorders of Hemostasis and Coagulation*].

A coagulation defect is detected in 5% to 20% of adolescents who are hospitalized for abnormal vaginal bleeding. Von Willebrand disease is the most common inherited bleeding disorder in women and is believed to occur in 0.1% to 0.8 % of the population.[6,7] Acquired and congenital deficiencies of coagulation factors are less frequent but should not be overlooked. It is imperative that an accurate diagnosis be made so that appropriate treatment can be administered.

Liver failure Cirrhosis of the liver may affect vaginal bleeding in two ways. First, estrogen metabolism may be compromised in advanced liver disease, leading to elevated levels of free estrogen. Second, the liver produces components of the intrinsic pathway of the coagulation cascade. When production of these factors becomes compromised, bleeding irregularities occur.

Thyroid dysfunction Hypothyroidism can cause menstrual disturbances. The prevalence of hypothyroidism in women with menorrhagia is about 2.5%. Patients who complain of weight gain, fatigue, cold intolerance, and constipation may have a thyroid disorder and warrant further evaluation. The thyroid-stim-

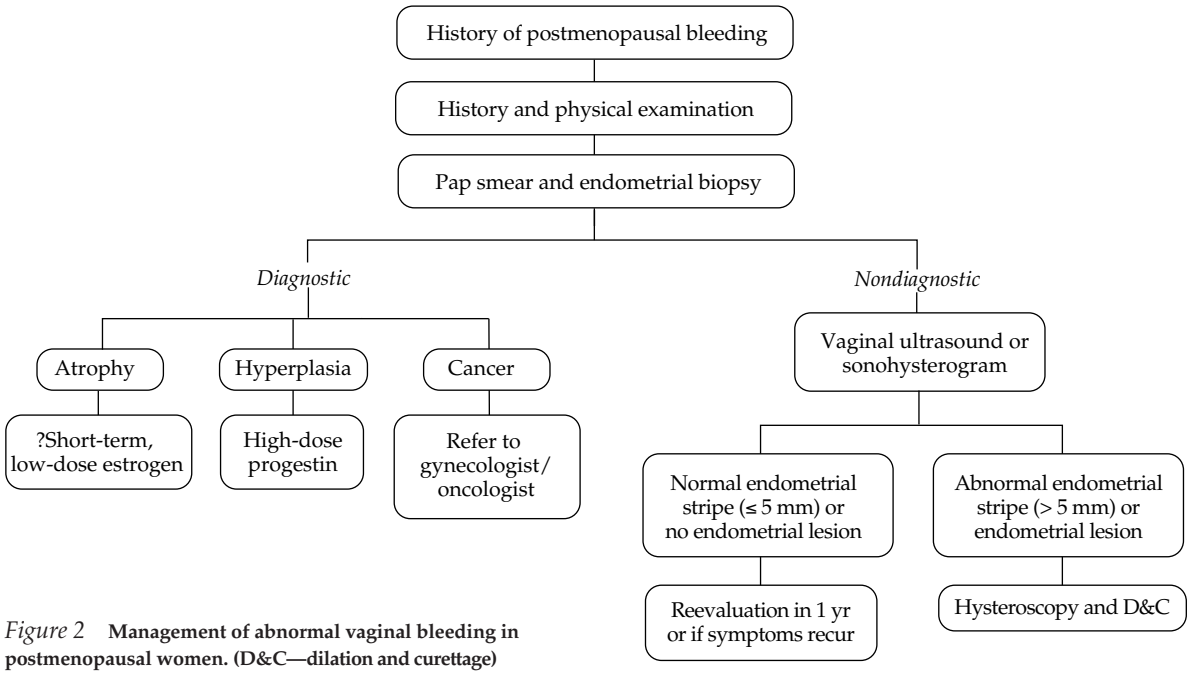


Figure 2 **Management of abnormal vaginal bleeding in postmenopausal women. (D&C—dilation and curettage)**

ulating hormone (TSH) level should be determined during the initial workup. Treatment entails oral administration of exogenous thyroid hormone. The dosage should be titrated until symptoms resolve and the TSH level becomes normal.

Hyperthyroidism can also cause abnormalities in menstrual bleeding. Uncommonly, women with hyperthyroidism experience menorrhagia. Patients should be treated with propylthiouracil until a normal TSH value is obtained. Beta blockers should be considered for patients with severe tachycardia. Treatment for hyperthyroidism is discussed in detail elsewhere [*see 3:I Thyroid*].

Iatrogenic Causes of Vaginal Bleeding

Women who use medications that contain estrogen and progesterone may experience changes in their bleeding pattern. Long-acting contraceptives such as depot medroxyprogesterone acetate and levonorgestrel-releasing implants commonly cause irregular menses and eventually amenorrhea in patients who use these medications for longer than 6 months. Other pharmacologic agents that cause menstrual disturbances are digitalis, phenytoin, anticoagulants, corticosteroids, and psychotropic agents. A detailed account of the patient's current medication usage will reveal risk for iatrogenically induced uterine bleeding.

Abnormal Uterine Bleeding

Abnormal uterine bleeding (AUB) is the term used to describe uterine bleeding that is not caused by an anatomic lesion. Two variations of AUB exist, ovulatory and anovulatory. Prostaglandins have been implicated as a cause of AUB. As the menstrual cycle progresses to the luteal phase, the production of prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) from arachidonic acid increases. $PGF_{2\alpha}$ causes vasoconstriction of the spiral arteries, whereas prostaglandin E_2 (PGE_2) induces vasodilation. In normal menstruating women, the level of $PGF_{2\alpha}$ relative to PGE_2 increases steadily during the menstrual cycle.

Another hypothesis regarding the regulation of menstrual bleeding considers the role of matrix metalloproteinase–2 (MMP-2). This enzyme initiates menstruation when progesterone is withdrawn from decidualized endometrial stroma. MMP-2 acts on the vascular and epithelial basement membrane of the endometrium, causing breakdown and shedding of the endometrium during a normal menstrual cycle.[8]

Ovulatory AUB Ovulatory AUB is caused by an abnormality in the ratio of $PGF_{2\alpha}$ to PGE_2. The production of $PGF_{2\alpha}$ in ovulatory AUB is reduced, while the synthesis of PGE_2 and prostacyclin is increased. This distortion in prostaglandin synthesis leads to abnormal uterine bleeding, despite the presence of a luteinized endometrium.

Anovulatory AUB In most cases, AUB arises from anovulation. Anovulatory AUB is especially likely to affect women shortly after menarche and before menopause. When ovulation does not take place, progesterone is not produced, and the effects of estrogen predominate. Continuous unopposed estrogen production causes the endometrial lining to proliferate without undergoing cyclical changes. Eventually, the endometrial lining outgrows its blood supply, which results in focal areas of necrosis and bleeding. Because a uniform shedding of the endometrial lining does not occur, a prolonged and abnormal bleeding pattern may develop. In addition, the $PGF_{2\alpha}$ production decreases and PGE_2 production remains the same, stimulating heavier bleeding.

Treatment of AUB Nonsteroidal anti-inflammatory drugs (NSAIDs) can reduce the amount of uterine bleeding in ovulatory and anovulatory women. These drugs inhibit the activity of cyclooxygenase and therefore block the production of prostaglandins [*see 16:IV Dysmenorrhea*]. Thromboxane is a prostaglandin that promotes platelet aggregation, and prostacyclin is a prostaglandin that inhibits platelet aggregation. The benefit of NSAIDs may derive from an alteration in the ratio of thromboxane to prostacyclin. Women with AUB who are treated with NSAIDs may experience as much as a 50% reduction in blood loss during their menstrual period.[9] There is no reason to

Table 2 Pharmacologic Treatment Options for Dysfunctional Uterine Bleeding

Ovulatory AUB
Nonsteroidal anti-inflammatory drugs
Anovulatory AUB
Combination oral contraceptives
or
Cyclic progestin
Medroxyprogesterone acetate, 10 mg/day for 7–14 days
Medroxyprogesterone acetate in oil, 50 mg I.M. q. 28 days
Depot medroxyprogesterone acetate, 150 mg I.M. q. 3 mo
and
Nonsteroidal anti-inflammatory drugs

AUB—abnormal uterine bleeding

choose one NSAID over another,[10] although some reports have suggested that naproxen may be superior. Cyclooxygenase-2 (COX-2) inhibitors are still being evaluated for this indication.

Management of anovulatory AUB is accomplished with hormone therapy, not by curettage. In treating anovulatory AUB, NSAIDs should be combined with either oral contraceptives or progestin therapy to optimize the control of bleeding [*see Table 2*]. Combination oral contraceptive therapy is preferred to other treatments because of its superior efficacy in controlling menstrual bleeding. Abnormal bleeding will usually cease within 24 hours after the start of this regimen.

Cyclic progestin can also be used to control anovulatory bleeding. Progesterone prevents endometrial proliferation, and its withdrawal causes an organized breakdown and uniform slough. Progesterone also increases the formation of arachidonic acid and the production of $PGF_{2\alpha}$, inducing vasoconstriction of the spiral arteries. Together, these two effects of progesterone cause an organized slough of the endometrium and a swift termination of bleeding.

Progesterone can be administered orally as medroxyprogesterone acetate, 10 mg a day for 7 to 14 days, or intramuscularly as a single dose of 50 mg progesterone in oil every 28 days. Alternatively, a single dose of a long-acting progesterone (e.g., depot medroxyprogesterone acetate, 150 mg) can be given once every 3 months.

Some patients have acute, heavy bleeding as a result of decreasing estrogen levels. Provided the patient is hemodynamically stable, she can be treated with estrogen alone, either intravenously or orally.[11] Therapy with progestins should not be used, because these agents are ineffective in controlling acute and heavy bleeding. When bleeding is pronounced and leads to hypovolemia, dilation and curettage (D&C) should be performed before the induction of medical therapy. However, a D&C should not be performed routinely for anovulatory bleeding.

Conjugated estrogen is usually administered intravenously in doses of 20 mg every 4 hours over 12 to 24 hours. After completion of intravenous estrogen administration, oral supplements with 2.5 to 3.75 mg of conjugated estrogen should follow for 3 weeks before the addition of progesterone supplements to obtain a withdrawal bleed. The use of estrogen without progesterone is not the preferred treatment for anovulatory AUB.

Women who fail to respond to medical management should be considered for surgical therapy. Conservative surgery with endometrial ablation can be used to treat women who wish to retain their uterus.[12] Heavy bleeding recurs in nearly half of these women within 1 year after the procedure, however, and it may need to be performed several times before the bleeding disorder is corrected. Ablation of the endometrium should not be performed on women who desire future pregnancies or on those who have not tried other medical modalities.

Hysterectomy is the ultimate and most definitive treatment for AUB. The removal of the uterus should be reserved for women with ovulatory AUB in whom all other therapies have failed. If an alternative indication for performing this procedure coexists with AUB, a hysterectomy should be performed.

Postmenopausal Bleeding

Bleeding in postmenopausal women is most often caused by atrophic changes of the endometrium. Endometrial atrophy accounts for 45% of such cases. Polyps, submucous myomas, and proliferative changes of the endometrium each account for about 10%, and endometrial hyperplasia accounts for another 15%. Thus, all postmenopausal women who experience an episode of bleeding should be evaluated promptly [*see Figure 2*].

DIAGNOSTIC EVALUATION

The evaluation should begin with a thorough history and physical examination. A detailed account of current medications should be obtained.

Women who use hormone replacement therapy (HRT) may experience episodes of uterine bleeding.[13] In most cases, bleeding that occurs with HRT is the result of an anatomic structural abnormality and not cancer. Although HRT is thought to reduce the risk of endometrial cancer, this possibility should still be considered in the differential diagnosis.

A Pap smear and endometrial biopsy should be obtained. These tests are diagnostic if they show atrophy, hyperplasia, or endometrial cancer. An endometrial biopsy cannot detect an intracavitary mass, however, so pelvic sonography or sonohysterography should be performed in patients with nondiagnostic results, to evaluate the endometrial cavity and search for an intracavitary mass.[14,15] If an endometrial biopsy cannot be performed in the office, such as in the case of a patient with a stenotic cervical os, an ultrasound should be obtained to evaluate the thickness of the endometrial lining. An endometrium that is less than 5 mm thick is considered normal and unlikely to harbor a malignancy. However, 4% of endometrial cancers will be missed by this criterion, with a false positive rate as high as 50%, so endometrial thickness measurement in symptomatic women does not reduce the need for invasive diagnostic testing.[16]

MANAGEMENT

Postmenopausal bleeding should be treated on the basis of the predisposing etiology. If a submucous myoma or polyp is responsible, surgical removal of the lesion should be considered. Endometrial atrophy is not amenable to treatment; some gynecologists use short-term, low-dose estrogen for this purpose, but there are no data to support this practice. High-dose progestin therapy should be instituted for women with endometrial hyperplasia. Women with endometrial cancer should be treated according to the stage and grade of the cancer detected.

If an identifiable cause of bleeding is found, such as a polyp or myoma, hysteroscopy should be performed. Hysteroscopy should also be done, followed by a D&C, when an ultrasound obtained during a screening reveals an abnormally thickened endometrial lining.

References

1. Brenner PF: Differential diagnosis of abnormal uterine bleeding. Am J Obstet Gynecol 175:766, 1996

2. Rosenfeld JA: Treatment of menorrhagia due to dysfunctional uterine bleeding. Am Fam Physician 53:165, 1996

3. Long CA: Evaluation of patients with abnormal uterine bleeding. Am J Obstet Gynecol 175:784, 1996

4. Clark TJ, Voit D, Gupta JK, et al: Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review. JAMA 288:1610, 2002

5. Kadir RA, Economides DL, Sabin CA, et al: Frequency of inherited bleeding disorders in women with menorrhagia. Lancet 351:485, 1998

6. Ewenstein BM: The pathophysiology of bleeding disorders presenting as abnormal uterine bleeding. Am J Obstet Gynecol 175:770, 1996

7. Federici AB, Mannucci PM: Advances in the genetics and treatment of von Willebrand disease. Curr Opin Pediatr 14:23, 2002

8. Irwin JC, Kirk D, Gwatkin RB, et al: Human endometrial matrix metalloproteinase-2, a putative menstrual proteinase: hormonal regulation in cultured stromal cells and messenger RNA expression during the menstrual cycle. J Clin Invest 97:438, 1996

9. Davies AJ, Anderson AB, Turnbull AC: Reduction by naproxen of excessive menstrual bleeding in women using intrauterine devices. Obstet Gynecol 57:74, 1981

10. Lethaby A, Augood C, Duckitt K: Nonsteroidal anti-inflammatory drugs for heavy menstrual bleeding. Cochrane Database Syst Rev (1):CD000400, 2002

11. Chuong CJ, Brenner PF: Management of abnormal uterine bleeding. Am J Obstet Gynecol 175:787, 1996

12. DeCherney AH, Diamond MP, Lavy G, et al: Endometrial ablation for intractable uterine bleeding: hysteroscopic resection. Obstet Gynecol 70:668, 1987

13. Spencer CP, Cooper AJ, Whitehead MI: Management of abnormal bleeding in women receiving hormone replacement therapy. BMJ 315:37, 1997

14. Weber AM, Belinson JL, Bradley LD, et al: Vaginal ultrasonography versus endometrial biopsy in women with postmenopausal bleeding. Am J Obstet Gynecol 177:924, 1997

15. O'Connell LP, Fries MH, Zeringue E, et al: Triage of abnormal postmenopausal bleeding: a comparison of endometrial biopsy and transvaginal sonohysterography versus fractional curettage with hysteroscopy. Am J Obstet Gynecol 178:956, 1998

16. Tabor A, Watt HC, Wald NJ: Endometrial thickness as a test for endometrial cancer in women with postmenopausal vaginal bleeding. Obstet Gynecol 99:663, 2002

III PREMENSTRUAL SYNDROME

Sarah L. Berga, m.d.

Premenstrual syndrome (PMS) is a recurrent constellation of affective and physical symptoms that begin during the luteal phase of the menstrual cycle and resolve completely or almost completely during the follicular phase. The number, severity, and duration of symptoms occur along a spectrum. It is estimated that 20% to 40% of women report premenstrual symptoms during the luteal phase, but only 5% to 10% of women report symptoms severe enough to significantly interfere with their lifestyle. The long-term natural history of PMS is unclear.

The fourth edition of the *Diagnostic and Statistical Manual* of the American Psychiatric Association (DSM-IV) has defined a related syndrome, premenstrual dysphoric disorder (PMDD), to facilitate an accurate psychiatric diagnosis [*see Table 1*].[1] The criteria for PMDD are also used to define research populations so that therapeutic responses can be quantitated and generalized. PMDD is often considered a variant of depression, an impression buttressed by the treatment efficacy of antidepressants, particularly selective serotonin reuptake inhibitors (SSRIs), in women who meet the criteria for PMDD.[2-4]

Pathogenesis

Most evidence suggests that PMS/PMDD is caused by aberrant responses of target tissues, particularly the brain, to normal fluctuations in serum levels of ovarian gonadal steroids.[5-7] The causes of the untoward central nervous system responses are not firmly established[8] [*see Table 2*], but several pathogenic mechanisms are currently being explored. Altered metabolism of progesterone by the CNS may lead to altered CNS reactivity and neurotransmission. Specifically, one of the principal metabolites of progesterone, allopregnanolone, decreases anxiety. Women with PMS show lower levels of allopregnanolone during their luteal phase[9]; they may instead produce a predominance of anxiogenic progesterone metabolites.

PMS/PMDD may involve alterations in CNS neurotransmission that cause heightened reactivity to normal excursions in gonadal steroid levels. This exaggerated reactivity may be inherited or acquired. An acquired cause of altered CNS neurotransmission is chronic stress. Chronobiologic disturbances documented in women with PMDD have been interpreted as evidence of an underlying aberration in CNS function.[10] High levels of estrogen or of estrogen and progesterone may elicit or aggravate this underlying brain dysfunction, which would explain why symptoms are greatest during the luteal phase or at ovulation. In women with PMS/PMDD, but not in women without it, the symptom complex can be replicated by exposure to, followed by withdrawal from, exogenous estrogen and progesterone.[11] There is also an association between PMS and dysmenorrhea, further suggesting that PMS is an exaggerated tissue response to normal hormonal changes.

Diagnosis

To warrant medical attention, evaluation, and intervention, premenstrual symptoms must be recurrent and sufficiently severe to interfere with daily work and social activities. To establish the diagnosis of PMS/PMDD, the clinician must confirm that the patient has the characteristic manifestations of the disorder at the appropriate time in her menstrual cycle.

It is also important to identify any concurrent conditions likely to complicate treatment [*see Table 3*]. Many women with PMS/PMDD have a personal or family history of alcoholism. A history of sexual abuse, particularly in childhood or adolescence, may be common in women with severe PMS.[12] This population also has an increased personal and family history of posttraumatic stress disorder, mood disorders, schizophrenia, eating disorders, postpartum depression or psychosis, personality disorders, and anxiety disorders. A positive family history does not necessarily imply an inherited biologic vulnerability, because persistent exposure to dysfunctional family interactions is a chronic stress that can alter underlying CNS function.

Conclusive diagnosis of PMS or PMDD requires the documentation of concordance between symptoms and the luteal or periovulatory phase. The diagnosis of PMS/PMDD cannot be made in an anovulatory patient. Ideally, symptoms and menstrual dates should be followed prospectively to establish synchrony between the luteal phase and increase in symptoms. Two or more cycles and at least five symptoms [*see Table 1*] should be charted before the diagnosis is made. However, with a patient who is suffering severe psychological distress, there may not be time for prospective evaluation. Immediate referral to a psychiatrist may be indicated to prevent suicide or homicide.

If the menstrual cycle is irregular (menstrual cycle is ordinarily between 26 and 30 days) and there is no clear pattern of symptoms, the progesterone level should be measured weekly

Table 1 DSM-IV Criteria for Premenstrual Dysphoric Disorder

Symptoms occur in the luteal phase, with prospective confirmation of a 30% increase in symptoms during the luteal phase above the level in the follicular phase

Not an exacerbation of major depression, panic dysthymia, or personality disorder

Marked disturbance in functioning

At least five of the following

- Marked lability
- Marked irritability
- Marked anxiety
- Markedly depressed mood
- Decreased interest
- Lethargy
- Difficulty concentrating
- Food craving
- Hypersomnia or insomnia
- Physical symptoms (breast tenderness or swelling, headache, joint or muscle pain, bloating, weight gain)
- Avoidance of social activities
- Decreased productivity
- Increased sensitivity to rejection
- Feeling overwhelmed
- Feeling out of control
- Increased interpersonal conflict

Table 2 Potential Causes of PMS/PMDD

Aberrant responses of target tissues, especially the brain, to normal gonadal steroid exposures mediated by the following:
- Opioid withdrawal
- Serotonergic imbalance
- Entrainment to endogenous cycles
- Chronobiologic disturbance
- Membrane effects of steroids or steroid metabolites
- Genomic effects of steroids
- Variation in steroid metabolism

throughout a cycle to determine whether there is a luteal phase. A progesterone concentration greater than 5 ng/ml, or 15 nmol/L, is generally considered evidence that a woman is in the luteal phase and ovulation is impending. No rise in progesterone indicates anovulation, which may be stress induced. These patients may require further evaluation [*see 16:I Amenorrhea*]. Symptoms should be charted concurrently with progesterone levels. To meet the criteria for PMS/PMDD, the patient's symptoms should become at least 30% more severe during the luteal phase (or when the progesterone concentration is greater than 5 ng/ml) than they were in the follicular phase.

DIFFERENTIAL DIAGNOSIS

If a patient has regular menses and severe dysmenorrhea but no behavioral symptoms, she should be evaluated for possible endometriosis [*see 16:X Endometriosis*]. As with all psychiatric diagnoses, it is important to exclude organic causes. In PMS/PMDD, it is especially important to exclude thyroid dysfunction (hyperthyroidism or hypothyroidism) and drug abuse or dependence as contributing factors.

Premenstrual changes in hormone levels can exacerbate underlying medical conditions, including migraine, epilepsy, asthma, irritable bowel syndrome, and diabetes mellitus.[13] This is not PMS but may resemble it.

If behavioral symptoms are severe and are present throughout the menstrual cycle and if there is no clear pattern of increase in symptom severity during the luteal phase, another psychiatric diagnosis must be considered. The following psychiatric disorders must be excluded: major depression, panic and anxiety disorders, dysthymia, and personality disorder. Such patients should be referred to a psychiatrist for definitive diagnosis and treatment. The most important condition to exclude is depression.

Treatment

Available therapies for PMS/PMDD range from lifestyle modification to surgery. Sustained improvement in a woman with PMS/PMDD generally requires a combination of modalities. The severity of a patient's symptoms and her response to particular modalities should guide the choice of therapies and the pace of their introduction. Mild cases can be treated with lifestyle modification and nonpharmacologic options; severe cases deserve immediate and aggressive intervention.

NONPHARMACOLOGIC THERAPY

Lifestyle interventions for PMS include institution of good sleep patterns and regular exercise. The patient should reduce or eliminate the use of tobacco, alcohol, and other drugs.

Dietary treatment helps some patients. Calcium supplementation may be beneficial.[14] It has been suggested that diets high in carbohydrates and protein buttress the serotonergic axis.[15] A diet high in tryptophan, a precursor of serotonin, may also be of benefit for mild PMS.[16]

Full-spectrum bright-light therapy given in the evening has been shown to markedly reduce symptoms of PMS/PMDD. Its use can be limited to the luteal phase.[17,18]

Stress management is integral to lifestyle treatment. Biofeedback,[19] massage,[20] and other relaxation methods may be helpful. Education, emotional support, and attention from the physician or therapist are instrumental. However, almost any intervention can be temporarily helpful, as the placebo response is quite high in this disorder.

Some women may wish to treat their PMS with herbal remedies, such as oil of primrose, chaste tree berries, or St. John's wort [*see Clinical Essentials:XII Complementary and Alternative Medicine*]. The use of herbal medicine and other complementary and alternative measures for PMS/PMDD has not been strongly validated in randomized, controlled trials, however.[21]

Behavioral Therapy

Patients with PMS may benefit from cognitive-behavioral therapy or interpersonal therapy. These are formal, structured psychotherapies designed to help patients institute behavioral changes and address cognitive patterns that sustain maladaptative behavior. Response to treatment may take as long as 6 months, but the effects persist indefinitely. If a patient is having difficulty coping with her symptoms during the early months of psychotherapy, there is no reason not to add pharmacologic treatment. The effects of medication are more rapid in onset than those of psychotherapy, but the effects persist only as long as the patient takes the medication. The model of combined psychotherapy and pharmacotherapy is considered the most effective approach to major depression; it has not been formally tested in PMS/PMDD, but there is no reason to believe it would not work well in this disorder, given the similarities between depression and PMS/PMDD.

Disposition

Benefit from nonpharmacologic interventions should be evident within two menstrual cycles. If the patient has shown no improvement at all during that time, the clinician should move on to pharmacotherapy. Complete recovery is not to be expected, however; that seldom or never occurs, regardless of the therapy chosen. Nevertheless, it is unwise for the clinician to lower the patient's expectations about these modalities, because enthusiasm inspires patients to participate in therapy and enhances the placebo effect. To set the stage for follow-up without lowering expectations, the clinician can tell the patient,

Table 3 Pertinent History in PMS/PMDD

- Reproductive events
- Dysmenorrhea
- Psychosocial adjustments and stressors
- Diet
- Exercise
- Sleep
- Drug, alcohol, and medication use
- Endocrine disorders
- Family and personal history of psychiatric disorders

"I want to see how you are doing in 2 months. If you are feeling perfectly fine and don't want to come back in, just give me a call. In case you aren't perfectly fine, though, I would rather see you back."

It is reasonable to wait more than 2 months for a response in a patient who has initiated behavioral therapy in a formal psychiatric setting. On the other hand, with a patient who is in severe psychological distress, 2 months may be too long to wait. The worst-case scenario is that such a patient will interpret a prescription of lifestyle modification and a distant follow-up appointment as a dismissal by her physician and, in the meantime, commit suicide or a homicide or ruin her life in some way. Clinicians who have limited psychiatric expertise or who practice in a stringent managed-care setting that severely restricts follow-up should refer severely distressed patients to a psychiatrist.

Pharmacologic Therapies for Somatic Symptoms

Bromocriptine (2.5 mg/day orally) has been promoted as treatment for breast tenderness. Spironolactone (25 to 50 mg/day orally) has been given to alleviate bloating. Nonsteroidal anti-inflammatory drugs can be effective treatment for dysmenorrhea [*see 16:IV Dysmenorrhea*].

Progesterone treatment has been shown to be ineffective for PMS.[22] Oral contraceptives are likely to aggravate rather than attenuate PMS/PMDD symptoms.

Pharmacologic Therapies for Affective Symptoms

Because the pathogenesis of PMS/PMDD probably involves an aberrant CNS response to normal ovarian function, the first-line treatment is to buttress CNS function with antidepressants. The SSRIs fluoxetine and sertraline have been shown to be effective[2-4]; other agents in this class presumably would work but are not as well studied. Although use of SSRIs can be limited to the luteal phase (10 to 14 days), that approach is impractical for many patients, who may have difficulty determining when to start the drug. It is simpler for patients simply to take the medication every day.

Sertraline has a shorter half-life than fluoxetine. The advantage of a shorter half-life is that if the patient experiences unacceptable side effects, the side effects will fade more rapidly once the medication is discontinued. The most prominent side effect of SSRIs in patients with PMS/PMDD—and a common reason for poor compliance with SSRI therapy for the disorder—is sexual dysfunction.[23] Some patients are willing to accept impaired libido as a trade-off for the relief of their symptoms. Buspirone is mildly effective for PMS/PMDD and may be a useful alternative in patients who find the sexual side effects unacceptable.[24] Classified as an atypical antidepressant, buspirone tends to be used for the anxious variety of depression.

Benzodiazepine therapy with alprazolam, taken during the luteal phase, is appropriate for patients whose main symptom is anxiety.[25] However, alprazolam has many more side effects than do SSRIs, even though the dose can be titrated to minimize side effects.

Pharmacologic Interventions That Alter Ovarian Steroid Exposure

Patients with PMS/PMDD who fail to respond adequately to lifestyle modification and SSRI therapy or who refuse or are unable to follow such measures can be treated with gonadotropin-releasing hormone (GnRH) agonist therapy (e.g., leuprolide, nafarelin, or goserelin).[26-29] GnRH agonists effect a medical oophorectomy. They cannot be continued indefinitely without hormone replacement therapy because of concerns about the long-term deleterious effects of sustained hypoestrogenism. Add-back hormone regimens generally involve continuous exposure to small amounts of both estrogen and progestin,[29] thereby obviating the hormonal changes associated with a menstrual cycle.

There is a variety of hormonal preparations available for add-back regimens. It is usually a good idea to administer the estrogen and progestin separately at first, so that the response to each can be monitored. The progestin dose must be large enough to prevent endometrial hyperplasia but below the threshold for triggering PMS symptoms. Ongoing exposure to even small amounts of progestin (e.g., oral medroxyprogesterone acetate, 2.5 mg daily) may provoke symptoms, however.

The synthetic androgen danazol can be used to temporarily suppress endogenous ovarian function and provide an androgenic environment. However, its side effects may be as problematic as the PMS/PMDD symptoms. The androgenic side effects of danazol include voice changes, hirsutism, and breast regression, all of which may be permanent.

Surgical Therapy

GnRH agonist therapy is expensive, so if a patient is responding very well to a GnRH agonist and has completed her childbearing, oophorectomy and hysterectomy may be a reasonable step. Surgery may also be the therapy of choice for patients who have sustained improvement with GnRH-agonist therapy but experience recurrent symptoms with add-back hormone regimens. Postoperatively, these patients are given hormone replacement therapy with continuous estrogen alone. Continuing SSRI therapy may be indicated.

References

1. American Psychiatric Association: Diagnostic and Statistical Manual of Psychiatric Disorders, 4th ed. United States Department of Health and Human Services, Washington, DC, 1996, p 714

2. Steiner M, Steinberg S, Stewart D, et al: Fluoxetine in the treatment of premenstrual dysphoria. Canadian Fluoxetine/Premenstrual Dysphoria Collaborative Study Group. N Engl J Med 332:1529, 1995

3. Steiner M, Korzekwa M, Lamont J, et al: Intermittent fluoxetine dosing in the treatment of women with premenstrual dysphoria. Psychopharmacol Bull 33:771, 1997

4. Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment: a randomized controlled trial. Sertraline Premenstrual Dysphoric Collaborative Study Group. JAMA 278:983, 1997

5. Sundstrom I, Andersson A, Nyberg S, et al: Patients with premenstrual syndrome have a different sensitivity to a neuroactive steroid during the menstrual cycle compared to control subjects. Neuroendocrinology 67:126, 1998

6. Rabin DS, Schmidt PJ, Campbell G, et al: Hypothalamic-pituitary-adrenal function in patients with the premenstrual syndrome. J Clin Endocrinol Metab 71:1158, 1990

7. Korzewka MI, Lamont JA, Steiner M: Late luteal phase dysphoric disorder and the thyroid axis revisited. J Clin Endocrinol Metab 81:2280, 1996

8. Berga SL: Understanding premenstrual syndrome. Lancet 351:465, 1998

9. Rapkin AJ, Morgan M, Goldman L, et al: Progesterone metabolite allopregnanolone in women with premenstrual syndrome. Obstet Gynecol 90:709, 1997

10. Parry BL, Berga SL, Kripke DF, et al: Altered waveform of plasma nocturnal melatonin secretion in premenstrual depression. Arch Gen Psychiatry 47:1139, 1990

11. Schmidt PJ, Nieman LK, Danaceau MA, et al: Differential behavioral effects of gonadal steroids in women with and in those without premenstrual syndrome. N Engl J Med 338:209, 1998

12. Golding JM, Taylor DL, Menard L, et al: Prevalence of sexual abuse history in a sample of women seeking treatment for premenstrual syndrome. J Psychosom Obstet Gynaecol 21:69, 2000

13. Case AM, Reid RL: Menstrual cycle effects on common medical conditions. Compr Ther 27:65, 2001

14. Calcium carbonate and the premenstrual syndrome: effects on premenstrual and menstrual symptoms. Premenstrual Syndrome Study Group. Am J Obstet Gynecol 179:444, 1998

15. Wurtman JJ, Brzezinski A, Wurtman RJ, et al: Effect of nutrient intake on premenstrual depression. Am J Obstet Gynecol 161:1228, 1989

16. Sayegh R, Schiff I, Wurtman J, et al: The effect of a carbohydrate-rich beverage on mood, appetite, and cognitive function in women with premenstrual syndrome. Obstet Gynecol 86:520, 1995

17. Parry BL, Berga SL, Mostofi N, et al: Morning versus evening bright light treatment of late luteal phase dysphoric disorder. Am J Psychiatry 146:1215, 1989

18. Parry BL, Mahan AM, Mostofi N, et al: Light therapy of late luteal phase dysphoric disorder: an extended study. Am J Psychiatry 150:1417, 1993

19. Van Zak DB: Biofeedback treatments for premenstrual and premenstrual affective syndromes. Int J Psychosom 41:53, 1994

20. Hernandez-Reif M, Martinez A, Field T, et al: Premenstrual symptoms are relieved by massage therapy. J Psychosom Obstet Gynaecol 21:9, 2000

21. Stevinson C, Ernst E: Complementary/alternative therapies for premenstrual syndrome: a systematic review of randomized controlled trials. Am J Obstet Gynecol 185:227, 2001

22. Wyatt K, Dimmock P, Jones P, et al: Efficacy of progesterone and progestogens in management of premenstrual syndrome: systematic review. BMJ 323:776, 2001

23. Sundstrom-Poromaa I, Bixo M, Bjorn I, et al: Compliance to antidepressant drug therapy for treatment of premenstrual syndrome. J Psychosom Obstet Gynaecol 21:205, 2000

24. Landen M, Eriksson O, Sundblad C, et al: Compounds with affinity for serotonergic receptors in the treatment of premenstrual dysphoria: a comparison of buspirone, nefazodone and placebo. Psychopharmacology (Berl) 155:292, 2001

25. Freeman EW, Rickels K, Sondheimer SJ, et al: A double-blind trial of oral progesterone, alprazolam, and placebo in treatment of severe premenstrual syndrome. JAMA 274:51, 1995

26. Freeman EW, Sondheimer SJ, Rickels K: Gonadotropin-releasing hormone agonist in the treatment of premenstrual syndrome with and without ongoing dysphoria: a controlled study. Psychopharmacol Bull 33:303, 1997

27. Brown CS, Ling FW, Andersen RN, et al: Efficacy of depot leuprolide in premenstrual syndrome: effect of symptom severity and type in a controlled trial. Obstet Gynecol 84:779, 1994

28. Mezrow G, Shoupe D, Spicer D, et al: Depot leuprolide acetate with estrogen and progestin add-back for long-term treatment of premenstrual syndrome. Fertil Steril 62: 932, 1994

29. Mortola JF, Girton L, Fischer U: Successful treatment of severe premenstrual syndrome by combined use of gonadotropin-releasing hormone agonist and estrogen/progestin. J Clin Endocrinol Metab 72:252A, 1991

IV DYSMENORRHEA

Alan H. DeCherney, m.d.

Dysmenorrhea—colicky pain in the lower abdomen and pelvis around the time of menses—is a common condition that disturbs the lives and families of the women who suffer from it. Severe pain and discomfort often lead to absenteeism from work or school, resulting in an overall reduction in productivity and enormous economic loss.

Dysmenorrhea can be classified as primary or secondary, depending on its cause. Secondary dysmenorrhea results from a known pathologic process occurring within the pelvis. Primary dysmenorrhea does not have an anatomic cause.

Because many women seek help for this disorder, the physician should be sensitive to the concerns of each patient and provide a treatment that is specific to each patient's needs. The goal of therapy is to enable the patient to resume her daily life without the fear of being disabled by pain.

Primary Dysmenorrhea

The prevalence of primary dysmenorrhea is between 40% and 90%, with an average of 75%. It occurs predominantly in women younger than 25 years.

pathogenesis

Primary dysmenorrhea results from tissue hypoxia and ischemia. An elevation in the basal tone of the uterus, combined with an increase in contraction strength and frequency, leads to vasospasm and a reduction in uterine blood flow.[1] Increased levels of prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$), leukotrienes, and vasopressin are responsible for these alterations [*see Figure 1*].[2] It is important to note that this abnormality in prostaglandin production occurs only in endometrial tissue that has been exposed to both estrogen and progesterone.

diagnosis

Diagnosis of primary dysmenorrhea is based on patient history; results of physical examination are usually normal. Symptoms characteristically begin within 6 to 12 months after menarche, as ovulatory cycles become established. The pain occurs only during ovulatory cycles and lasts about 48 to 72 hours each month. In most patients, the pain starts a few hours before menstruation or at the onset of menstruation.

The degree of pain suffered is variable, but fewer than 15% of women with primary dysmenorrhea have severe pain. Patients with severe pain may also experience other symptoms, such as nausea, vomiting, dizziness, and diarrhea.

treatment

Treatment of primary dysmenorrhea includes nonsteroidal anti-inflammatory drugs (NSAIDs), oral contraceptives, and transcutaneous electrical nerve stimulation (TENS).[3] Other nonpharmacologic approaches that may be effective are the use of a lower abdominal heating pad,[4] supplemental vitamin B_1 (100 mg daily) or magnesium (400 mg daily),[5] and a low-fat vegetarian diet.[6]

NSAIDs inhibit the action of cyclooxygenase and prevent the conversion of arachidonic acid into prostaglandins (PGs). NSAIDs significantly alleviate pain in approximately 75% of patients. The fenamates (e.g., mefenamic acid) are considered the

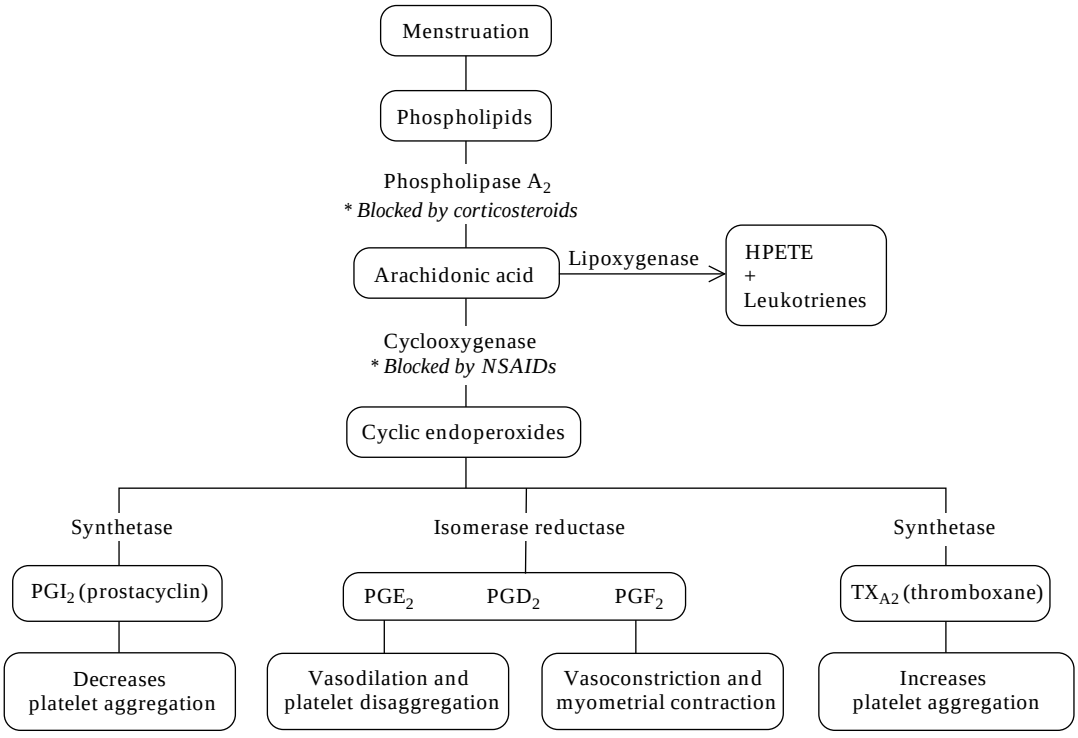


Figure 1 **Prostaglandin synthesis. (HPETE—hydroperoxyeicosatetraenoic acid)**

Table 1 Common Prostaglandin Synthesis Inhibitors

Chemical Group	*Derivative*	*Usual Dosage*
Benzoic acid	Acetylsalicylic acid (aspirin)	650 mg p.o. every 4 hr
Fenamates	Mefenamic acid	250 mg p.o. every 6 hr
	Meclofenamate sodium	100 mg p.o., t.i.d., for up to 6 days
Indoleacetic acid	Indomethacin	25–50 mg p.o. every 8 hr
Arylpropionic acid	Ibuprofen	400–800 mg p.o. every 4 hr
	Naproxen	250–500 mg p.o. every 12 hr
	Naproxen sodium	275–550 mg p.o. every 12 hr
Cyclooxygenase-2 (COX-2) inhibitors	Celecoxib	200 mg p.o., b.i.d.
	Rofecoxib	50 mg p.o., q.d.
	Valdecoxib	20 mg p.o., b.i.d.

best choice of NSAIDs because they act as antiprostaglandins, preventing both the production of PGs and the binding of the PG to its receptor [*see Table 1*].[7,8] Therapy with NSAIDs should be discontinued if adverse side effects occur.[9] Definitive research on the use of cyclooxygenase-2 (COX-2) inhibitors (e.g., celecoxib [Celebrex] or rofecoxib [Vioxx]) in dysmenorrhea has not been completed, but these agents seem promising for this purpose.

If NSAIDs are ineffective or poorly tolerated, the next step is the use of combined estrogen-progestin oral contraceptives. In a normal menstrual cycle, the progesterone level increases after ovulation and steadily decreases during the luteal phase. As the level of progesterone decreases, lysosomal enzymes within the endometrial cells are released, causing an increase in the production of PGs. Oral contraceptives prevent fluctuations of endogenous progesterone levels, reducing the amount of pain and symptoms associated with primary dysmenorrhea.[10] Regular-cycle oral contraceptives are not usually effective for treatment of primary dysmenorrhea; continuous oral contraceptives are preferable. Many patients consider continuous oral contraceptives unnatural but are willing to compromise by taking long-cycle contraceptives, so that they have three or four menses a year [*see 16:VI Contraception*].

If neither NSAIDs nor oral contraceptives alone alleviate primary dysmenorrhea, the two can be used together. If combination therapy also fails, the patient should be reevaluated and a diagnostic workup initiated for secondary dysmenorrhea.

Secondary Dysmenorrhea

The pain associated with secondary dysmenorrhea is the direct result of a pathologic process. Unlike primary dysmenorrhea, secondary dysmenorrhea varies with regard to the patient's age at onset and the causative condition. Some of the conditions that can cause secondary dysmenorrhea include endometriosis, adenomyosis, pelvic adhesions and infection, pelvic congestion, cervical stenosis, psychological stress, and psychological disturbances.

endometriosis

Endometriosis is the presence of endometrial glands and stroma outside the uterus [*see 16:X Endometriosis*]. Approximately 7% of women in the United States suffer from this disorder. Endometriosis causes intra-abdominal hemorrhage, fibrosis, and adhesion formation. Consequently, dyspareunia, infertility, and pelvic pain occur.[11]

The pain usually begins 2 to 3 days before menses and worsens during menstruation. Tender nodules along the uterosacral ligament, a posteriorly fixed uterus, and enlarged cystic ovaries are characteristic findings; however, results of physical examination are often normal. Definitive diagnosis requires direct visualization during laparoscopy, with or without a tissue biopsy.

Treatment may entail either medical intervention or surgery. Oral contraceptives, intramuscular injection of leuprolide acetate depot, oral danazol, or high-dose progestins (oral or intramuscular) are all beneficial in suppressing the endometrial implants and relieving the symptoms of pain.

adenomyosis

Adenomyosis is the presence of ectopic endometrial glands and stroma in the myometrium of the uterus. Unlike the ectopic glands in endometriosis, the ectopic glands in adenomyosis do not undergo monthly cyclical changes.

Symptoms of adenomyosis classically include dysmenorrhea and menorrhagia (heavy menstrual bleeding). As the disease progresses, so does the dysmenorrhea. On physical examination, the uterus is soft, globular, and uniformly enlarged. Typically, the uterus is tender just before and during menstruation.

Diagnostic aids include pelvic sonography, magnetic resonance imaging, and hysterosalpingography. Unfortunately, most cases go undiagnosed until histologic evaluation is made at the time of a hysterectomy.

Treatment starts with medical suppression of ovarian function and culminates in hysterectomy if symptoms do not abate. Thermal balloon ablation of the endometrium, which is an effective treatment for some patients with dysmenorrhea from other causes,[12] does not work in adenomyosis.

cervical stenosis

When menstrual flow is impeded at the level of the internal cervical os, intrauterine pressure increases and pelvic pain occurs. A narrow or stenotic os may be a congenital abnormality or the result of trauma, infection, or surgery.

The diagnosis of cervical stenosis should be considered in women who have a history of hypomenorrhea and severe pelvic pain during menses or if the diameter of the external cervical os is less than 5 mm.

During the physical examination, the physician should attempt to pass a uterine sound into the endometrial cavity. Inability to document a clear passage through the cervical canal warrants further investigation. Diagnostic workup with hysterosalpingography may reveal a narrow cervical canal.

Treatment consists of dilating the cervical canal with laminaria tents or performing a formal dilatation and curettage (D and C) under anesthesia. These procedures have limited therapeutic benefit and need to be repeated frequently. Complete resolution of symptoms typically occurs with pregnancy and vaginal delivery, which therefore is considered the ultimate therapy.

pelvic inflammatory disease

Most pelvic infections are caused by *Chlamydia*, *Neisseria gonorrhoeae*, and mixed microbial organisms. Pelvic anatomy is often distorted as a consequence of dense adhesion formation. During menstruation, adhesion edema and venous congestion

result in severe pelvic pain and discomfort. This pain may eventually become chronic.

Patients at risk for pelvic inflammatory disease (PID) include current or past users of intrauterine devices (IUDs) and women with more than one sexual partner. The workup includes cervical cultures, endometrial biopsy, and pelvic sonography.

Treatment of the dysmenorrhea associated with PID includes NSAIDs for pain management and antibiotics for acute infection. Surgery can be offered to patients with chronic pain and to those with a known tubo-ovarian abscess or hydrosalpinx. Although lysis of adhesions can be performed, results are usually poor because recurrence is high.

pelvic congestion syndrome

Engorgement and thrombosis of the pelvic veins are another cause of dysmenorrhea.[13] The pooling of blood in the pelvic vasculature results in a burning and throbbing pain. The pain is characteristically worse at night and after prolonged periods of standing. Bimanual examination often reveals a uterus that is mildly enlarged and tender to the touch. The diagnosis of pelvic congestion syndrome is made almost exclusively during laparoscopic evaluation.

Although the underlying cause of pain is not well understood, treatment often entails NSAIDs and psychological therapy. New treatment approaches that utilize uterine artery embolization show promising results. Hysterectomy should be reserved for patients who do not respond to other therapeutic modalities.

other causes of chronic pelvic pain

Secondary dysmenorrhea can also be caused by psychological problems, including stress, tension, and abnormal conditioned behavior. For these patients, resolution of symptoms is best achieved through lifestyle and behavior modification. Chronic pelvic pain, rather than acute pain, is more common among women with psychological disorders.

Patients who have pain for more than 6 months are considered to have chronic pelvic pain. In addition to a gynecologic cause of the pelvic pain, the physician should always consider other causes. The basic workup should include gastrointestinal, urologic, musculoskeletal, and psychological evaluations. Once a diagnosis has been established, treatment should focus on correcting the underlying disorder. Treatment should be initiated with medical therapy. If this fails, more aggressive treatment can be attempted, including presacral neurectomy or a laparoscopic uterine nerve ablation (LUNA) procedure.[14-16]

References

1. Dawood MY: Dysmenorrhea. Clin Obstet Gynecol 26:719, 1983

2. Smith RP: Cyclic pelvic pain and dysmenorrhea. Obstet Gynecol Clin North Am 20:753, 1993

3. Kaplan B, Rabinerson D, Pardo J, et al: Transcutaneous electrical nerve stimulation (TENS) as a pain-relief device in obstetrics and gynecology. Clin Exp Obstet Gynecol 24:123, 1997

4. Akin MD, Weingand KW, Hengehold DA, et al: Continuous low-level topical heat in the treatment of dysmenorrhea. Obstet Gynecol 97:343, 2001

5. Wilson ML, Murphy PA: Herbal and dietary therapies for primary and secondary dysmenorrhoea. Cochrane Database Syst Rev (3):CD002124, 2001

6. Barnard ND, Scialli AR, Hurlock D, et al: Diet and sex-hormone binding globulin, dysmenorrhea, and premenstrual symptoms. Obstet Gynecol 95:245, 2000

7. Owen PR: Prostaglandin synthetase inhibitors in the treatment of primary dysmenorrhea: outcome trials reviewed. Am J Obstet Gynecol 148:96, 1984

8. Rosenwaks Z, Seegar-Jones G: Menstrual pain: its origin and pathogenesis. J Reprod Med 25:207, 1980

9. Brooks P: Use and benefits of nonsteroidal anti-inflammatory drugs. Am J Med 104 (suppl 3A):9S, 1998

10. Crosignani PG, Vegetti W, Biachedi D: Hormonal contraception and ovarian pathology. Eur J Contracept Reprod Health Care 2:207, 1997

11. Fedele L, Bianchi S, Bocciolone L, et al: Pain symptoms associated with endometriosis. Obstet Gynecol 79:767, 1992

12. Ulmsten U, Carstensen H, Falconer C, et al: The safety and efficacy of Meno Treat, a new balloon device for thermal endometrial ablation. Acta Obstet Gynecol Scand 80:52, 2001

13. Cordts PR, Eclavea A, Buckley PJ, et al: Pelvic congestion syndrome: early clinical results after transcatheter ovarian vein embolization. J Vasc Surg 28:862, 1998

14. Polan ML, DeCherney A: Presacral neurectomy for pelvic pain in infertility. Fertil Steril 34:557, 1980

15. Kwok A, Lam A, Ford R: Laparoscopic presacral neurectomy: a review. Obstet Gynecol Surv 56:99, 2001

16. Carter JE: Surgical treatment for chronic pelvic pain. JSLS 2:129, 1998

V POLYCYSTIC OVARY SYNDROME

Robert L. Barbieri, m.d.

Polycystic ovary syndrome (PCOS) is defined as hyperandrogenism and reduced frequency of ovulation in the absence of other hyperandrogenic disorders. The clinical manifestations of PCOS in women with hyperandrogenism are cosmetic and reproductive. These patients present with hirsutism, acne, and irregular menstrual periods; ovulation may be infrequent or absent, and infertility can occur.

When assessing a patient with possible PCOS, the physician must rule out other conditions that can produce clinical hyperandrogenism, such as androgen-secreting tumors and nonclassic congenital adrenal hyperplasia. It is also necessary to identify patients whose hirsutism or acne results from increased sensitivity to normal androgen levels.

Epidemiology

PCOS is one of the most common endocrine disorders of women. In three population-based studies, the average prevalence of PCOS in women of reproductive age was reported to be about 6%.[1-3] Among anovulatory women, the prevalence of PCOS is approximately 30%.[4]

Pathophysiology

In women, luteinizing hormone (LH) and adrenocorticotropic hormone (ACTH) normally drive the secretion of androgens by the ovaries and adrenal glands, respectively. PCOS is caused by abnormally increased secretion of LH, insulin, or both. Increased levels of those hormones stimulate the ovarian theca and stroma to produce excess quantities of androgen, including testosterone and androstenedione. Elevated ovarian androgen secretion tends to block the growth of a dominant ovarian follicle. Instead, many small follicles accumulate; these follicles range in size from 4 to 8 mm in diameter. In the absence of a dominant follicle, an LH surge is not triggered, and ovulation does not occur regularly.

About 95% of women with PCOS have elevated levels of LH secretion, and 50% have hyperinsulinemia.[5] Those rates are influenced by body mass and genetics. In a population with a high prevalence of obesity, up to 100% of the women with PCOS will have hyperinsulinemia. Indeed, there may be two major phenotypes of PCOS: (1) lean women with markedly elevated levels of LH secretion and minimal or no insulin resistance and (2) obese women with slightly elevated or normal levels of LH secretion and insulin resistance and with markedly elevated levels of insulin secretion.

Women with PCOS show an abnormal increase in both the amplitude and the frequency of LH pulses in the early follicular phase of the menstrual cycle.[6] The elevated LH pulse frequency suggests an underlying increase in the pulse frequency of gonadotropin-releasing hormone (GnRH) secretion by the hypothalamus; this in turn suggests that PCOS results from a neuroendocrine disorder. The neuroendocrine mechanisms that raise GnRH pulse frequency are poorly characterized but may include alterations in hypothalamic opioid and catecholamine tone.

The insulin resistance in women with PCOS, as in other patients, has many possible causes, including genetic mutations in the insulin receptor and autoantibodies to the insulin receptor. The most common cause, however, is obesity. When pancreatic function is normal, resistance to the action of insulin in the liver, adipose tissue, muscle, and other insulin-sensitive tissues results in a compensatory and chronic hypersecretion of insulin. Laboratory studies suggest that insulin, especially in high concentrations, can stimulate ovarian androgen secretion.[7] Why insulin resistance develops in muscle and fat but not the ovary remains unclear. It is possible that insulin stimulates ovarian androgen secretion indirectly by binding to the insulin-like growth factor–1 receptor in the theca and the stroma.

Hirsutism

At birth, all areas of the body except the scalp and eyebrows are covered with vellus hair, which is light colored; individual vellus hairs have a very narrow diameter. Androgens can transform vellus hair into terminal hair, which is dark; individual terminal hairs have a thick diameter. The amount of androgen necessary to stimulate this transformation depends on many factors, including the sensitivity of hair follicles to androgens and the site of the hair follicle on the body. When girls reach puberty, small quantities of adrenal and ovarian androgens stimulate the transformation of vellus hair in the pubic region and axilla into terminal hair. Substantially larger amounts of androgen are required to stimulate the growth of terminal hair in a male-pattern distribution—that is, on the face, chest, and abdomen.

The term hirsutism denotes an increase in terminal hair in a male-pattern distribution. Hirsutism varies considerably by race: it is rare in Asian women, for example, because their hair follicles are relatively insensitive to androgens. In some families, women may inherit heightened sensitivity of the follicles to androgens and therefore may experience a degree of hirsutism at normal levels of circulating androgens.

Acne

The hair follicle is part of the pilosebaceous unit, which also contains a sweat or sebaceous gland. Androgenic stimulation of the pilosebaceous unit promotes not only hair growth but also the production of sebum. Blockage of the follicle by excessive sebum (in concert with inflammation from free fatty acids produced by bacteria and yeast in the follicle) produces acne. The pilosebaceous unit is both a target organ responsive to androgen stimulation and a site of androgen production and metabolism.[8] Consequently, acne, like hirsutism, reflects both circulating androgen concentrations and genetic makeup.

Diagnosis

In PCOS, hyperandrogenism and oligo-ovulation or anovulation are present, but other hyperandrogenic disorders, such as androgen-secreting tumors and nonclassic adrenal hyperplasia, are absent.

No single feature is pathognomonic for PCOS. The history, physical examination, and laboratory evaluation can all provide evidence that establishes the diagnosis of PCOS and that excludes other conditions associated with those features. Clinical evidence of hyperandrogenism includes hirsutism, acne, and menstrual irregularity. Laboratory evidence of hyperandrogenism includes an elevated total or free serum testosterone concentration [*see Table 1*].

HISTORY

The history can provide key information in the differential diagnosis of androgen excess (and heightened androgen sensitivity) in women.

Age of Onset

In PCOS, oligomenorrhea, hirsutism, and acne typically begin in the perimenarchal or teenage years. The onset of severe hirsutism in menopause suggests an ovarian neoplasm.

Menstrual History

Patients with PCOS typically experience irregular menstrual cycles starting at menarche. Regular cycles are more consistent with familial or idiopathic hirsutism. A history of initially regular periods followed by onset of oligomenorrhea or amenorrhea and hirsutism with virilization in adult life suggests an androgen-secreting tumor.

Pace of Progression of Hirsutism

In PCOS, hirsutism tends to progress slowly, over many years. Rapid progression to severe hirsutism suggests a virilizing disorder from an androgen-secreting tumor. Patients with androgen-secreting tumors typically report other manifestations of virilization, such as deepening of the voice and secondary amenorrhea. Virilization will be evident on the physical examination (see below).

Family History

Approximately 50% of women with PCOS have a family history of PCOS, type 2 diabetes mellitus, or both. Alternatively, the history may disclose familial hirsutism, which begins at puberty and is accompanied by regular menstrual cycles and normal concentrations of circulating androgens. In the absence of a positive family history, hirsutism that is disproportionate to the patient's racial background and is accompanied by regular periods is considered idiopathic, if laboratory test results are normal (see below).

Medication Use

Some medications appear to cause increased LH secretion and thus promote the development of PCOS. In particular, long-term use of the anticonvulsant valproate is strongly associated with the onset of PCOS.[9]

Cigarette Smoking

Women who smoke have higher concentrations of androstenedione and testosterone than do nonsmoking women.[10] Hence, smoking may contribute to hyperandrogenism.

PHYSICAL EXAMINATION

Hirsutism can be assessed objectively with the Ferriman-Gallwey scoring system [*see Figure 1*].[11] Along with providing a baseline measurement of hirsutism, this system can also be used to follow the efficacy of treatment.

Various physical findings point to insulin resistance [*see Table 2*]. Excess weight is a major determinant of insulin resistance and hyperinsulinemia.[12] Relative weight is best assessed by means of the body mass index (BMI), which is calculated by dividing the patient's weight in kilograms by the square of the patient's height in meters. Women with a BMI of greater than 27 (i.e., those who are overweight) are often insulin resistant and usually demonstrate hyperinsulinemia in response to a glucose stimulus. Women with a BMI of greater than 30 (i.e., those who are obese) are almost always insulin resistant. Women with a BMI of less than 22 are unlikely to be insulin resistant unless they have one of a relatively rare group of acquired or inherited lipodystrophic disorders.

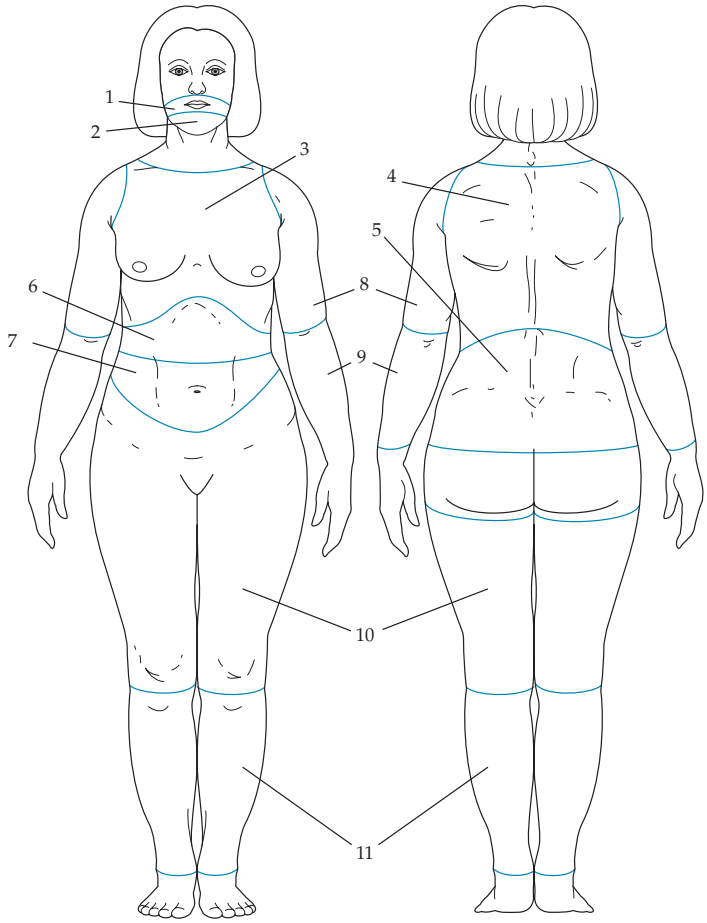
Other physical findings that suggest insulin resistance are a waist-to-hip ratio greater than 0.85 and a waist circumference greater than 90 cm (35.5 in). The presence of acanthosis nigricans or achrochordons (skin tags) suggests hyperinsulinemia. The syndrome of hyperandrogenism, insulin resistance, and acanthosis nigricans (HAIR-AN syndrome) is the most severe form of the insulin-resistant phenotype of PCOS.

Unfortunately, the physical findings that are associated with insulin resistance tend to be specific but not sensitive. For example, patients with acanthosis nigricans are almost always in-

Table 1 Characteristics of Common Causes of Androgen Excess

Diagnosis	*Cause*	*Ovulation Status*	*Testosterone Level*	*8 A.M. Follicular-Phase 17-Hydroxyprogesterone Level*
PCOS	Elevated LH, serum insulin, or both	Anovulation or oligo-ovulation	Elevated (0.75–2 ng/ml)	Normal (< 4 ng/ml)
Idiopathic hirsutism	Elevated production of androgen in the pilosebaceous unit, or a mild form of PCOS	Regular ovulation	Normal (< 0.75 ng/ml)	Normal (< 4 ng/ml)
Adrenal hyperplasia	Decrease in 21-hydroxylase activity because of a gene mutation	Anovulation or oligo-ovulation	Elevated (0.75-2 ng/ml)	Elevated (> 4 ng/ml)
Ovarian or adrenal tumor	Disorder of cell growth	Anovulation	Markedly elevated (> 2 ng/ml)	May be elevated

LH— luteinizing hormone PCOS—polycystic ovary syndrome



Site	Grade	Definition
1. Upper Lip	1	Few hairs at outer margin
	2	Small mustache at outer margin
	3	Mustache extending halfway from outer margin
	4	Mustache extending to midline
2. Chin	1	Few scattered hairs
	2	Scattered hairs with small concentrations
	3 and 4	Complete cover, light and heavy
3. Chest	1	Circumareolar hairs
	2	With midline hair in addition
	3	Fusion of these areas, with three-quarters cover
	4	Complete cover
4. Upper Back	1	Few scattered hairs
	2	Rather more, still scattered
	3 and 4	Complete cover, light and heavy
5. Lower Back	1	Sacral tuft of hair
	2	With some lateral extension
	3	Three-quarters cover
	4	Complete cover
6. Upper Abdomen	1	Few midline hairs
	2	Rather more, still midline
	3 and 4	Half and full cover
7. Lower Abdomen	1	Few midline hairs
	2	Midline streak of hair
	3	Midline band of hair
	4	Inverted V-shaped growth
8. Arm	1	Sparse growth affecting not more than one quarter of limb surface
	2	More than this; cover still incomplete
	3 and 4	Complete cover, light and heavy
9. Forearm	1, 2, 3, and 4	Complete cover of dorsal surface; 2 grades of light and 2 of heavy growth
10. Thigh	1, 2, 3, and 4	As for arm
11. Leg	1, 2, 3, and 4	As for arm

Figure 1 **Ferriman-Gallwey scoring system for quantitating hirsutism. The 11 sites are graded from 0 (no terminal hair) to 4 (severe hirsutism). Women with a total score greater than 8 are considered hirsute.[11]**

sulin resistant, but many women with insulin resistance do not have acanthosis nigricans.[13] Although the identification of severe insulin resistance on the basis of clinical manifestations may be relatively simple, the detection of mild insulin resistance may be difficult.

A key aspect of the physical examination is a search for signs of virilization, such as clitoromegaly, increased upper body muscle mass, and male pattern baldness. These may indicate the presence of an androgen-secreting tumor.

Women with PCOS have enlarged ovaries, although the ovaries typically are not palpable on pelvic examination. If pelvic examination discloses a large, complex mass, the patient may have an adrenal or ovarian tumor.

LABORATORY TESTS

The goals of the laboratory evaluation of hyperandrogenism are to rule out an adrenal and ovarian tumor, assess the severity of the androgen excess, and determine whether the source of the hyperandrogenism is adrenal or ovarian. Laboratory tests that are the most useful in the evaluation of hyperandrogenism include determination of the total serum testosterone level; determination of the 8 A.M. follicular phase 17-hydroxyprogesterone level; determination of the serum dehydroepiandrosterone sulfate (DHEAS) level (if fertility is an issue); and determination of the serum prolactin level (if the patient has amenorrhea).

Testosterone

The serum testosterone concentration provides the best laboratory estimate of the severity of androgen overproduction. Either total or free testosterone can be measured. Total testosterone measurement is performed by all clinical laboratories; these tests are well standardized and are less expensive than free testosterone measurement. However, because the level of sex hormone–binding globulin decreases as testosterone production increases, the total testosterone level does not fully reflect the degree of hyperandrogenism, especially if the overproduction of testosterone is minimal. Many women with mild PCOS have a total testosterone level in the upper end of the normal range. If the total testosterone level is greater than 2 ng/ml (200 ng/dl), the patient probably has ovarian stromal hyperthecosis or an adrenal or ovarian tumor and needs a detailed evaluation, which should include imaging studies of the ovary and adrenal glands.

The free testosterone measurement is more sensitive in detecting mild androgen overproduction. Nevertheless, because the total testosterone adequately identifies women with marked

Table 2 Physical Findings Associated with Insulin Resistance

Body mass index* > 27	Acanthosis nigricans
Waist-to-hip ratio > 0.85	Numerous achrochordons (skin tags)
Waist > 90 cm	

*Calculated by dividing the patient's weight in kilograms by the square of her height in meters.

androgen overproduction who need additional evaluation and because free testosterone is usually more expensive to assay, measurement of free testosterone is not routinely indicated.

17-Hydroxyprogesterone

Approximately 2% of women who present with hyperandrogenism and oligo-ovulation or anovulation have nonclassic adrenal hyperplasia resulting from a 21-hydroxylase deficiency. The prevalence of this congenital disorder varies markedly among different ethnic groups, from below 1% in Hispanic populations to as high as 5% to 8% in Ashkenazi Jewish populations. The decision to screen for the disorder depends on the cost-benefit assessment of detection and the baseline prevalence of the disorder in the patient's ethnic group.

If the 17-hydroxyprogesterone level at 8 A.M. (measured in the follicular phase of the menstrual cycle) is greater than 4 ng/ml, the patient probably has nonclassic adrenal hyperplasia resulting from a 21-hydroxylase deficiency. This diagnosis can be confirmed by a 60-minute ACTH stimulation test. The test utilizes a form of synthetic ACTH (cosyntropin) that contains the first 24 of the 39 amino acids of natural ACTH: 0.25 mg is given intravenously or intramuscularly, and the 17-hydroxyprogesterone level is measured 60 minutes later. A post-ACTH 17-hydroxyprogesterone level greater than 10 ng/ml confirms the diagnosis of nonclassic adrenal hyperplasia resulting from a 21-hydroxylase deficiency.

DHEAS

DHEAS, an androgen prohormone that can be converted to testosterone in the periphery, is secreted almost exclusively by the adrenal glands. The normal DHEAS level in premenopausal women is 0.12 to 5.35 µg/dl. A DHEAS level above 10.70 µg/dl—that is, more than twice the upper limit of normal—should raise concern over a possible adrenal tumor. Many women with PCOS have a DHEAS level in the upper range of normal, for reasons that have not been clearly identified. In infertile women with PCOS whose DHEAS level is greater than 2 µg/ml, the combination of clomiphene and a glucocorticoid may result in higher pregnancy rates than clomiphene alone (see below).

Serum Prolactin

If the patient has amenorrhea, the laboratory workup should include an assessment of the serum prolactin level to rule out a prolactin-secreting pituitary tumor [*see 16:I Amenorrhea*]. Many clinicians also routinely measure serum follicle-stimulating hormone (FSH) and thyroid-stimulating hormone (TSH) levels in amenorrheic patients.

Serum Luteinizing Hormone

The measurement of serum LH presents a special problem in the laboratory evaluation of PCOS. In the research setting—using multiple serum LH measurements (every 10 minutes for at least 8 hours) and a precise and reliable LH assay—elevated LH levels can be documented in about 95% of women with PCOS. However, because LH secretion is pulsatile and the standard commercial assays are not as precise as research assays, measurement of LH in clinical practice is of only modest utility. An elevated LH level is reasonably specific for PCOS, provided the sample was not taken during a preovulatory LH surge. A normal LH value does not necessarily exclude PCOS, however, because the test sample may have been drawn when the patient was at the nadir of an LH pulse. Another important point is that as BMI increases, the normal range for LH decreases [*see Figure 2*].[14] Nomograms that control serum LH for BMI are not widely available.

Pelvic Imaging

Demonstration of polycystic ovaries on pelvic ultrasonography is not essential for the diagnosis of PCOS. Pelvic imaging is indicated only if the ovaries are palpable on physical examination or the total testosterone concentration is greater than 200 mg/dl.

Tests for Detection of Insulin Resistance and Hyperinsulinemia

About 50% of women with PCOS have insulin resistance and hyperinsulinemia. There is no clear consensus on how to detect those conditions. A major problem is that the least resource-intensive laboratory techniques for diagnosing insulin resistance and hyperinsulinemia are specific but not sensitive. Elevation of the fasting serum insulin level or a fasting serum insulin-to-glucose ratio of less than 4.5 is almost always associ-

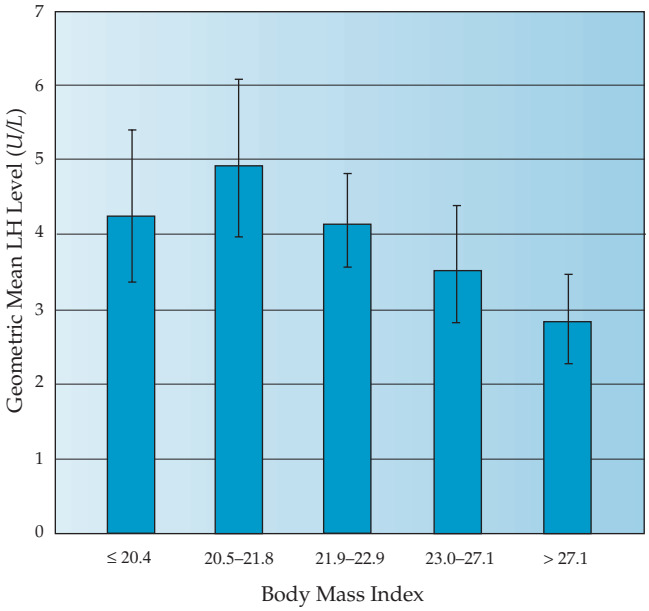


Figure 2 **Relationship between body mass index (BMI) and basal luteinizing hormone (LH) levels in women in the follicular phase of the menstrual cycle.[14]**

ated with insulin resistance, but many insulin-resistant women do not have fasting hyperinsulinemia. Laboratory techniques that are both specific and sensitive for detecting insulin resistance, such as euglycemic hyperinsulinemic clamp studies, are too complex and expensive for application in general clinical practice. Until both specific and sensitive laboratory tests that can be widely applied in practice become available, clinicians should use clinical findings and, if necessary, simple laboratory tests—such as assessment of fasting insulin levels or assessment of insulin response to an oral glucose challenge—to identify women with insulin resistance. Even nonobese women with PCOS have marked increases in circulating insulin after a glucose challenge [*see Figure 3*].[15]

Differential Diagnosis

IDIOPATHIC HIRSUTISM

Idiopathic hirsutism is defined as hirsutism in a woman with regular, ovulatory menstrual cycles. Women with idiopathic hirsutism have circulating testosterone and androstenedione concentrations at the upper limit of the normal range, but in these patients, such levels are lower than the levels observed in women with PCOS. Women with idiopathic hirsutism often have sisters with PCOS, and they tend to have a lower BMI than the sisters with PCOS.[16,17] Many authorities believe that idiopathic hirsutism is a mild form of PCOS in which hyperandrogenism is present but the disease has not progressed to the point where ovulatory menses have become disrupted. Other authorities believe that idiopathic hirsutism is the result of overactive skin conversion of weak precursor androgens (e.g., as androstenedione) to potent androgens (e.g., dihydrotestosterone) directly in the pilosebaceous unit. Regardless of the etiology, idiopathic hirsutism is best treated with the same approach as that for hirsutism in women with PCOS (see below).

VIRILIZATION SYNDROMES

Women with the rapid onset of virilization or a serum testosterone level greater than 2 ng/ml (200 ng/dl) should be evaluated for an adrenal or ovarian tumor. Use of magnetic resonance imaging to screen for an adrenal tumor and use of pelvic sonography to screen for an ovarian tumor are helpful.

Adrenal carcinoma presents with the rapid onset of virilization; it is often associated with systemic symptoms such as fatigue, weakness, and weight loss. DHEAS concentration is often greater than 8 µg/ml, and 24-hour urinary 17-ketosteroid excretion is markedly increased, to about 30 mg/dl.

OVARIAN HYPERTHECOSIS

Careful histologic examination of ovaries from women with PCOS often reveals islands of luteinized, steroid-secreting stromal cells (stromal hyperthecosis) in the medullary portion of the ovary that are not associated with follicular structures.[18] Severely hyperthecotic ovaries may contain only a small number of follicles, each 4 to 8 mm in diameter. PCOS patients with more severe hyperinsulinemia seem to be at highest risk for hyperthecosis.

Only a small subset of women with hyperandrogenism have stromal hyperthecosis. The diagnosis should be considered in a

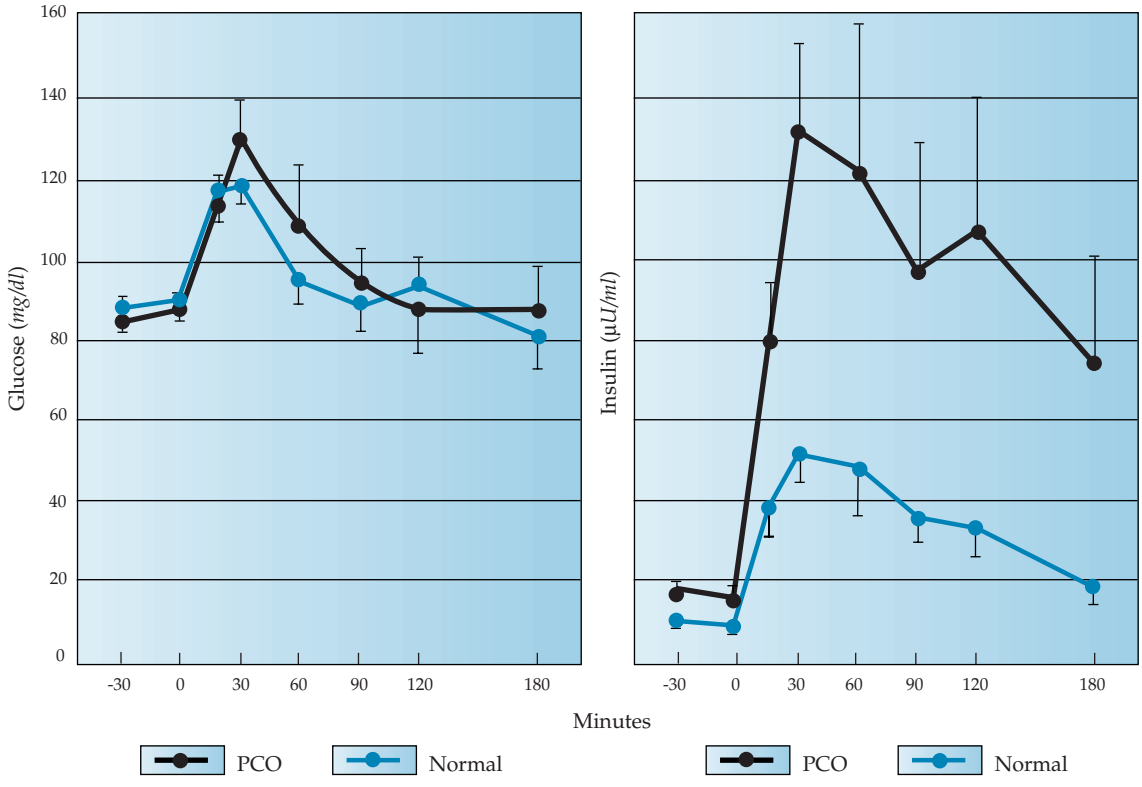


Figure 3 **In response to an oral glucose challenge, nonobese women with PCOS experience an exaggerated increase in circulating insulin compared with weight-matched control subjects; the increase persists for over 3 hours.[15]**

patient who presents with virilization, a total serum testosterone concentration of greater than 2 ng/ml, a normal LH level, and marked insulin resistance and hyperinsulinemia. Pathologic confirmation of the diagnosis, which requires removal of the ovaries, is not necessary.

Differentiation of ovarian hyperthecosis from PCOS is important because women with ovarian hyperthecosis often do not have significant suppression of circulating testosterone when treated with estrogen-progestin contraceptive alone. Instead, treatment with a GnRH analogue (e.g., leuprolide acetate depot, 3.75 mg intramuscularly every 4 weeks) plus an estrogen-progestin contraceptive often results in the normalization of circulating androgens.[19] A possible interpretation of this observation is that LH must be profoundly suppressed to decrease ovarian androgen production in women with hyperthecotic ovaries. Low-dose estrogen-progestin contraceptives alone do not suppress pituitary LH secretion as completely as they do in combination with a GnRH agonist analogue.

Treatment

Treatment for PCOS may be general—directed at the underlying hormonal imbalance—or specific to a particular manifestation (e.g., hirsutism or infertility) [*see Table 3*]. PCOS should be treated, because it poses long-term risks of endometrial cancer, diabetes mellitus, and possibly cardiovascular disease.

Regardless of the patient's presenting complaint, treatment of PCOS in a woman who smokes cigarettes includes smoking cessation. Discontinuance of smoking may result in a reduction in circulating androgens; also, smoking is a contraindication to the use of oral contraceptives, which are often prescribed for patients with PCOS.

TREATMENT OF HIRSUTISM

The mainstay of the treatment of hirsutism resulting from androgen excess is the combination of an estrogen-progestin oral contraceptive (used in regular or long cycles) and an antiandrogen. Patients should be advised that response to therapy tends to be slow; for more immediate results, patients may prefer the use of techniques that directly destroy the hair follicle.

Oral Contraceptives

Cyclic estrogen-progestin oral contraceptives produce multiple beneficial effects in patients with PCOS. These effects include the following: (1) decreased LH secretion, which suppresses ovarian androgen production; (2) increased liver production of sex hormone–binding globulin, which decreases free testosterone concentration; (3) decreased adrenal androgen production, through an unidentified pathway; (4) prevention of endometrial hyperplasia; and (5) regular uterine withdrawal bleeding. Of course, oral contraceptives also prevent pregnancy. The choice of agent does not seem important; it appears that any oral contraceptive, regardless of the estrogen dose or the progestin employed, can be effective in the treatment of hirsutism.

By suppressing both ovarian and adrenal androgen production, oral contraceptives decrease the stimulus to terminal hair growth. In patients with hirsutism, therapy with an oral contraceptive first brings a decrease in the diameter of the hair shaft and in the intensity of pigment in the hair—the unwanted hair becomes thinner and lighter in color. With prolonged treatment, the linear rate of hair growth diminishes. Ultimately, the hair follicle becomes senescent. If no new hair follicle starts to grow, the hirsutism then diminishes.

Long-cycle oral contraceptives When oral contraceptives are used in a standard cyclic manner (21 days of active pills followed by 7 days of inactive pills), LH secretion tends to increase during the 7 days in which inactive pills are taken. This stimulates a small rise in ovarian testosterone secretion, which, in turn, can stimulate hair growth. As an alternative, oral contraceptives can be taken in a continuous manner. In the initial treatment of hirsutism caused by PCOS, use of continuous oral contraceptives may be more effective than use of cyclic oral contraceptives for the suppression of LH and androgen production.[20]

Table 3 Treatment of Hirsutism, Anovulatory Infertility, and Endometrial Hyperplasia in Women with PCOS

Presenting Problem	*Standard Treatment*	*Alternative Treatment*
Hirsutism	Oral contraceptive plus spironolactone Weight loss	Oral contraceptive used in a long-cycle regimen, plus spironolactone Oral contraceptive plus GnRH analogue Insulin sensitizer, preferably metformin Finasteride
Acne	Oral contraceptive	Topical agents, oral antibiotics
Infertility	Weight loss Clomiphene Metformin plus clomiphene Clomiphene plus glucocorticoids Low-dose FSH injections Low-dose FSH injections plus metformin	Ovarian surgery IVF-ET
Endometrial hyperplasia	Oral contraceptives High-dose progestins	Weight loss

FSH—follicle-stimulating hormone GnRH—gonadotropin-releasing hormone IVF-ET—in vitro fertilization with embryo transfer

Most women with PCOS would like to restore a regular pattern of menses. To accommodate this desire while maximally suppressing LH levels, the clinician can employ the following regimen: the patient takes an active pill daily for 63 days (using the active pills from three packs), then takes inactive pills for 7 days. This long-cycle oral contraceptive regimen may induce regular withdrawal bleeding every other month and produce better suppression of LH and testosterone than oral contraceptives used in monthly cycles.

Antiandrogens

Antiandrogens are a cornerstone of therapy for hirsutism. In a 1-year trial, antiandrogen therapy using cyproterone acetate, 50 mg daily, plus ethinyl estradiol on cycle days 16 through 25, was compared with glucocorticoid suppression with hydrocortisone, 20 mg daily. Although hydrocortisone was more effective at normalizing plasma androgen levels, cyproterone provided superior treatment of hirsutism.[21] This study demonstrated that in established hirsutism, suppressing circulating androgens to normal levels is less important than blocking peripheral androgen action in the pilosebaceous unit.

Cyproterone acetate is not available in the United States. However, spironolactone has proved to be an effective antiandrogen, and its effects on hirsutism are similar to those of cyproterone acetate.[22] Because spironolactone has been used for many decades in the treatment of hirsutism and has an excellent safety profile, it should be considered the first-line agent for this indication in the United States. The usual dosage of spironolactone as an antiandrogen is 100 mg once daily.

The 5α-reductase inhibitor finasteride, which the Food and Drug Administration has approved only for treatment of men with benign prostatic hyperplasia, also reduces hirsutism in women.[23] The dosage of finasteride is 5 mg daily.

The androgen receptor antagonist flutamide (Eulexin) is as effective as spironolactone and finasteride in the treatment of hirsutism, but flutamide has liver toxicity and may cause fulminant liver failure.[24] Although the risk of liver failure with flutamide is small, hirsutism is a benign condition, so treatment regimens must have a high degree of safety.

Spironolactone, finasteride, and flutamide are believed to be human teratogens and may induce abnormal genital tract development in male fetuses. Women taking these agents need to use effective birth control. In addition, spironolactone, finasteride, and flutamide do not protect against endometrial hyperplasia and are not likely to induce regular menstrual cycles. These considerations support the use of an oral contraceptive in combination with an antiandrogen in the treatment of hirsutism in women of reproductive age.

Weight Loss

Numerous studies have demonstrated the benefits of weight loss in hyperandrogenic, insulin-resistant women.[25-27] In these studies, mean weight loss ranging from about 10 to 20 kg has been associated with a decrease in insulin levels and testosterone concentration and with ovulation and subsequent pregnancy in many women.[28-30]

Weight loss is difficult to achieve. A structured program that includes consultation with a nutritionist, encouragement by the physician, a low-calorie diet, and initiation of an exercise program may be the most effective nonsurgical approach in these patients. Surgical methods of weight reduction can be very effective, especially in women whose BMI is greater than 40.

Gonadotropin-Releasing Hormone Agonists

Women with hirsutism who do not respond to treatment with an oral contraceptive plus an antiandrogen can often be successfully treated with a combination of a GnRH agonist (e.g., leuprolide acetate) and an oral contraceptive,[31] with or without an antiandrogen. It is likely that the combination of leuprolide and an oral contraceptive produces greater suppression of LH and androgens than the oral contraceptive alone. Leuprolide is not ordinarily recommended as a first-line therapy for hirsutism because it is expensive (costing approximately $400 a month). Also, when used as a single agent, a GnRH agonist produces profound hypoestrogenism and accelerated bone loss. The addition of an oral contraceptive prevents GnRH agonist–induced hypoestrogenism and vasomotor symptoms and preserves bone density.

Glucocorticoids

Treatment with glucocorticoids may be appropriate in women with ACTH-dependent adrenal androgen overproduction, such as those with nonclassic adrenal hyperplasia resulting from 21-hydroxylase deficiency,[32,33] but it is not appropriate for the treatment of hirsutism in PCOS. A major problem with glucocorticoid therapy is that the complete suppression of ACTH production often requires giving more glucocorticoid than would normally be produced by the adrenal glands. As a result, patients receiving long-term glucocorticoid treatment are at increased risk for iatrogenic Cushing syndrome, osteoporosis, and diabetes mellitus. In addition, the corticotropin-releasing hormone–ACTH–cortisol axis may become so suppressed that the adrenal response to stress is blunted. For those reasons, I avoid using glucocorticoids to treat hirsutism. If the clinician does decide to use glucocorticoid therapy, use of low-dose glucocorticoids (5 or 7.5 mg of prednisone daily) or an alternate-day regimen of glucocorticoids may minimize those risks. Also, because almost all women treated with glucocorticoids gain weight and many develop osteoporosis, the clinician should carefully monitor weight and bone density in these patients.

Ovarian Surgery

Ovarian surgery can be used to decrease the mass of androgen-secreting thecal and stromal tissue. No randomized, controlled trials have been published concerning the benefits and risks of ovarian surgery for hyperandrogenism. In my opinion, the risks of ovarian surgery are greater than the potential benefits for women with hirsutism. I recommend ovarian surgery only for women with PCOS and infertility in whom conservative therapy has failed (see below).

Laser, Electrolysis, and Depilatory Treatment

Techniques that directly destroy the hair follicle are helpful in the treatment of hirsutism. Mechanical techniques include plucking, waxing, and shaving. These methods do not lead to a worsening of hirsutism, as some people believe, but may cause skin irritation. Electrolysis can destroy both the hair and the cells responsible for the growth of the hair, thereby producing a more prolonged beneficial effect than mechanical techniques.

Laser therapy is an evolving technique that is highly effective in the treatment of hirsutism. Multiple case series have

been reported using the ruby laser to treat hirsutism. The melanin in the hair is used as a natural chromophore. The energy delivered by the ruby laser causes photothermal damage to the hair and key cells surrounding the hair, which may prevent hair regrowth. In one study, two treatments with the ruby laser resulted in a 50% reduction in hair follicle density.[34] Unwanted side effects of laser treatment of hirsutism include skin hyperpigmentation and hypopigmentation and pitting of the skin surface. Most of these changes resolve spontaneously over 6 months.[35]

A new chemical inhibitor of hair growth is eflornithine 13.9% cream (Vaniqa). Eflornithine irreversibly inhibits the activity of skin ornithine decarboxylase, an enzyme that is necessary for the synthesis of polyamines and hence for hair growth. In clinical studies, which excluded pregnant and nursing women, eflornithine reduced facial hirsutism.[36]

TREATMENT OF ACNE

As with treatment of hirsutism, treatment of acne in PCOS patients begins with an oral contraceptive. Randomized, placebo-controlled clinical trials have demonstrated the benefits of combination estrogen-progestin oral contraceptives for acne after 6 months of treatment.[37,38] Because norgestrel is more androgenic than gestodene and desogestrel, some authorities recommend, on theoretical grounds, that oral contraceptives containing one of the latter progestins be used to treat acne.

Topical agents and oral antibiotics may also be indicated for acne in women with PCOS [*see 2:XII Acne Vulgaris and Related Disorders*].

TREATMENT OF ANOVULATION

Ovulation induction in PCOS follows a stepwise approach [*see Table 4*]. If the BMI is greater than 27, weight loss is an important goal. If normalization of the BMI cannot be achieved, clomiphene citrate is often prescribed because it is relatively inexpensive and has an excellent safety profile. Hyperandrogenic, insulin-resistant women are more likely to fail to ovulate and become pregnant with clomiphene than are women who are not insulin resistant. If both weight loss and clomiphene do not induce ovulation and result in pregnancy, the currently available choices for ovulation induction include insulin-sensitizing agents (for patients with insulin resistance), FSH injections, ovarian surgery, and in vitro fertilization with embryo transfer (IVF-ET).

Table 4 A Stepwise Approach to the Induction of Ovulation in Infertile Women with PCOS*

Step 1: if BMI is > 27, weight loss of at least 10%
Step 2: clomiphene
Step 3: if DHEAS > 2 µg/ml (200 µg/dl), clomiphene plus glucocorticoid therapy
Step 4: metformin for 8 to 12 wk
Step 5: metformin plus clomiphene
Step 6: low-dose FSH therapy
Step 7: metformin plus low-dose FSH therapy
Step 8: in vitro fertilization
Step 9: laparoscopic ovarian surgery to reduce ovarian androgen production

*Steps proceed in order of increasing resource intensity.

Weight Loss

Many women with PCOS are overweight or obese. In such women, weight loss (see above) is associated with a decrease in insulin secretion, a decrease in LH secretion, and a decrease in androgen production. The result is often a resumption of regular ovulation and, in some women, pregnancy.

Clomiphene Citrate

Clomiphene citrate is the most widely used agent for ovulation induction in women with PCOS. The FDA-approved doses for clomiphene are 50 or 100 mg daily for a maximum of 5 days per cycle. After a spontaneous menses or the induction of menses with a progestin withdrawal maneuver (medroxyprogesterone acetate, 10 mg p.o. daily for 5 days), clomiphene (50 mg daily for 5 days) is started on cycle day 3, 4, or 5. In properly selected women, 50% will ovulate through the use of this clomiphene regimen. Another 25% will ovulate if the dose of clomiphene is increased to 100 mg daily. During each cycle, determination of ovulation should be attempted by use of basal body temperature charts, ultrasound monitoring of follicle growth and rupture, or luteal-phase progesterone measurements. Some clinicians use endometrial biopsies to document ovulation in cycles where conception is not attempted. In most women, ovulation occurs approximately 5 to 12 days after the last dose of clomiphene. Measurement of the urinary LH surge is recommended to assist the patient in prospectively determining the periovulatory interval.

Although the FDA has approved maximal clomiphene doses of 100 mg daily, many clinicians have used clomiphene at doses of up to 250 mg daily. Women who fail to ovulate after taking clomiphene in doses of 100 mg daily for 5 days may ovulate if they are treated with clomiphene at doses of 150 mg daily for 5 days. Some authorities advocate use of clomiphene at doses up to 250 mg daily for up to 14 days. As many as 70% of the women who fail to ovulate with doses of 100 mg daily will ovulate with higher doses, but fewer than 30% of those become pregnant. In my opinion, there are few data to support the use of clomiphene at doses greater than 150 mg daily. Women who do not become pregnant at that dose should consider other approaches to ovulation induction (see below).

Clomiphene treatment can be associated with adverse changes in the reproductive tract, including induction of a luteal-phase defect (delay of endometrial maturation) and the creation of a hostile cervical environment from low quantity and quality of cervical mucus. Some clinicians recommend endometrial biopsy in a test cycle of clomiphene treatment to assess whether clomiphene induces luteal-phase deficiency. Many clinicians recommend that a postcoital test be performed during the first clomiphene cycle of treatment to screen for poor cervical mucus properties.

Multiple pregnancy is a well known outcome of clomiphene use. The absolute risk of high-order multiple gestation with clomiphene treatment is low: in a manufacturer's study of 2,369 clomiphene-induced pregnancies, 7% resulted in twins, 0.5% triplets, 0.3% quadruplets, and 0.13% quintuplets. However, because clomiphene is a heavily prescribed medication, the number of triplets resulting from clomiphene is substantial. The rate of spontaneous abortion after clomiphene-induced ovulation

and pregnancy is approximately 15%. The most common side effects of clomiphene include vasomotor symptoms (20%), adnexal tenderness (5%), nausea (3%), headache (1%), and, rarely, blurring of vision or scotomata. Most clinicians permanently discontinue clomiphene in women who experience visual changes from the drug.

Clomiphene plus Glucocorticoid

Anovulatory women with PCOS whose serum DHEAS concentration is above the midnormal range (2 μg/ml) appear to have reduced ovulation and pregnancy rates when they are treated with clomiphene. Some studies suggest that adding a glucocorticoid (e.g., dexamethasone) to clomiphene improves pregnancy rates in these women.[39]

Insulin Sensitizers

A major advance in reproductive endocrinology is the discovery that insulin sensitizers can induce ovulation in infertile women with oligo-ovulation, hyperandrogenism, and insulin resistance. Insulin sensitizers that have been approved for the treatment of diabetes include metformin, rosiglitazone, and pioglitazone. The insulin sensitizer D-chiro-inositol has been demonstrated to induce ovulation in hyperandrogenic insulin-resistant women, but it is currently available for use in research trials only.[40]

Metformin An oral biguanide antihyperglycemic agent approved for the treatment of type 2 diabetes mellitus, metformin has been evaluated in multiple studies for its ability to induce ovulation in women with hyperandrogenism and insulin resistance.

Metformin decreases blood glucose by inhibiting hepatic glucose production and enhancing peripheral glucose uptake. It increases insulin sensitivity at the postreceptor level and stimulates insulin-mediated glucose disposal. Unlike the sulfonylureas, metformin's mechanism of action does not involve increased insulin secretion.

Metformin increases the number of ovulatory cycles in infertile women with hyperandrogenism and insulin resistance. When used together with clomiphene, metformin significantly increases the rate of ovulation and of pregnancy resulting in live-born singleton births.[41-44] Metformin has also been shown to enhance response to the induction of ovulation with FSH injections in oligo-ovulatory, hyperandrogenic, insulin-resistant women.[45]

Metformin is commonly used at a dosage of 500 mg three times daily. To minimize gastrointestinal side effects, such as nausea, many clinicians start metformin at 500 mg daily for the first week, then increase the dosage to 500 mg twice daily for the second week, and then increase the dosage again to 500 mg three times daily. After the full dosage is reached, some clinicians switch to a regimen of 850 mg twice daily to enhance compliance. Although metformin is not approved for ovulation induction by the FDA, it may be significantly less expensive than FSH injections, ovarian surgery, or IVF-ET and may have fewer serious side effects than these treatments.

The most common side effects associated with metformin are GI disturbances, including diarrhea, nausea, vomiting, and abdominal bloating. In rare cases, metformin treatment has caused fatal lactic acidosis, but most of these patients had some degree of renal insufficiency or were severely hypoxic. Before starting treatment with metformin, it is advisable to confirm that the patient's serum creatinine level is less than 1.4 mg/dl.

If a patient has not ovulated after 5 to 10 weeks of metformin treatment, clomiphene can be added (see above). If the patient becomes pregnant, metformin can be discontinued, although it is a category B drug for pregnant women and has been used by some clinicians to treat diabetes in pregnant women.[46]

Thiazolidinediones The thiazolidinediones increase cellular sensitivity to the effects of insulin. Agents in this category include pioglitazone, rosiglitazone, and troglitazone. Several studies of troglitazone reported a decrease not only in fasting insulin but also in LH and testosterone levels, along with an increase in ovulatory cycles.[47-49] Troglitazone was reported to induce ovulation in obese, infertile women with hyperandrogenism and severe insulin resistance who had previously failed to ovulate when treated with clomiphene.[50] In this study, some women ovulated and became pregnant with troglitazone alone, whereas others responded to troglitazone plus clomiphene.

Troglitazone was removed from the market because of its association with the risk of death from liver failure.[51] The risk was extremely small but measurable, affecting approximately 1 in 50,000 patients treated with the drug. Pioglitazone and rosiglitazone, although now widely used in the treatment of diabetes mellitus, have not been extensively studied for their impact on ovulation. One might reasonably assume that they would offer benefits similar to troglitazone. However, until their efficacy and safety in PCOS is well established, it is probably best to use metformin as the main insulin sensitizer in women with PCOS.

Clomiphene plus Gonadotropin Induction

In women who fail to ovulate after therapy with clomiphene alone, gonadotropin injections can be added to clomiphene treatment to induce ovulation.[52] Typically, the injections are started after clomiphene, 100 to 200 mg daily, has been given for 5 days. The main benefit of this combination is that it tends to reduce the quantity of gonadotropins (an expensive medication) needed to induce ovulation during each cycle, because the rise in endogenous LH and FSH levels induced by clomiphene increases the sensitivity of the follicles to the injected gonadotropins. This regimen has been associated with a 50% decrease in the dosage of gonadotropin required to induce ovulation.[53]

Gonadotropins

The gonadotropins currently available for ovulation induction include (1) FSH produced by recombinant DNA technology and immunopurification and (2) LH plus FSH derived from menopausal urine. The recombinant FSH preparations can be given as subcutaneous injections and are available in ampules of 37.5 or 75 IU. FSH is the primary hormone responsible for follicular recruitment and growth in humans; it can be used as a single agent to induce ovulation in most anovulatory women. Women with PCOS generally do not require exogenous LH to induce follicular development, because their levels of LH secretion are already increased.

In women with PCOS, induction of ovulation with long-term, low-dose FSH treatment appears to result in a high pregnancy rate with a low rate of complications such as high-order multiple gestation and ovarian hyperstimulation.[54] In this approach, 75 units of FSH are given daily for the first 14 days; the

dose is then raised by 37.5 units every 7 days until follicular ripening is complete. If FSH treatment fails to result in pregnancy, consideration should be given to the combination of metformin and FSH, ovarian surgery, or IVF-ET.

During gonadotropin induction of ovulation, as many as 20% of patients experience mild to moderate enlargement of the ovaries. Some women treated with gonadotropins develop increased vascular permeability and accumulation of fluid in the peritoneal cavity and pleural space, a condition termed the ovarian hyperstimulation syndrome (OHSS). Clinical manifestations of OHSS include abdominal pain, abdominal distention, nausea, vomiting, diarrhea, and dyspnea. Other physical and laboratory findings of OHSS include weight gain, ovarian enlargement, ascites, pleural effusion, hemoconcentration, electrolyte imbalances, renal dysfunction, and thrombosis.[55] Treatment includes bed rest, maintenance of intravascular volume, prophylaxis against thrombosis, and surgical correction of ovarian torsion.

Before the utilization of repetitive estradiol measurements and sonographic evaluation of the follicular development, OHSS occurred in as many as 5% of women receiving gonadotropin treatment. In recent series that employed intense monitoring with those techniques, the rate of OHSS was approximately 0.5%. OHSS may be more severe and have a longer course if a successful pregnancy occurs. Multiple births occur in approximately 15% of pregnancies that take place after ovulation induction with gonadotropins.

Ovarian Surgery

Many gynecologists recommend that in infertile women with PCOS, ovarian surgery be attempted before FSH injections or IVF-ET. Laparoscopic drilling of the ovary is the most widely studied surgical treatment for ovulation in PCOS; approximately 1,000 cases have been reported, although no controlled studies have been undertaken.[56] These reports demonstrate that surgery to induce ovulation causes a decrease in circulating LH (50% decline) and testosterone (30% decline) and an increase in FSH (30% increase). The pregnancy rate is in the range of 50% at 12 months and 70% at 24 months.

The surgical techniques used for ovarian drilling vary between centers. However, all use a laser or electrosurgery to make multiple millimeter-size punctures in each ovary.[57-59]

In Vitro Fertilization with Embryo Transfer

IVF-ET has recently been demonstrated to be effective in the treatment of infertile women with PCOS who fail to become pregnant with gonadotropin injections.[60] In preliminary reports, IVF-ET treatment of infertile women with PCOS has been associated with a per-cycle pregnancy rate of 0.24 to 0.27.[61,62]

Metabolic Abnormalities Associated with PCOS

DIABETES MELLITUS

Obese women with PCOS may be at high risk for diabetes. In one study, testing in 254 PCOS patients found new-onset diabetes in 7%, most of whom had a BMI greater than 30.[63] Another study showed that by 40 years of age, 10% of women with PCOS will have been diagnosed with diabetes.[64]

It is important to assess for diabetes in an obese infertile woman with PCOS before using an ovulation-inducing agent. Poorly controlled diabetes is associated with a significantly increased risk of fetal malformations. Sacral agenesis and caudal dysplasia are 400 times more common in the offspring of women with diabetes than in those of women with normal glucose metabolism. Other anomalies associated with diabetes include anencephaly, open spina bifida, renal agenesis, ventricular septal defects, and transposition of the great vessels. Control of diabetes before conception decreases the risk of fetal malformations.

HYPERLIPIDEMIA

Women with PCOS have hyperandrogenism, and obesity and insulin resistance are common. These conditions are often associated with hyperlipidemia. In one study, levels of total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglycerides were higher in obese women with PCOS than in weight-matched control subjects. Elevated LDL levels were the predominant lipid abnormality observed in women with PCOS, independent of obesity.[65] High-density lipoprotein (HDL) cholesterol levels were also higher, however, which could provide some degree of protection against cardiovascular disease. One epidemiologic study that included a 30-year follow-up of women with PCOS found no increase in mortality from cardiovascular disease.[66]

References

1. Knochenhauer ES, Key TJ, Kahsar-Miller M, et al: Prevalence of PCOS in unselected black and white women of the southeastern United States: a prospective study. J Clin Endocrinol Metab 83:3078, 1998

2. Diamanti-Kandarakis E, Kouli CR, Bergiele AT, et al: A survey of the polycystic ovary syndrome in the Greek island of Lesbos: hormonal and metabolic profile. J Clin Endocrinol Metab 84:4006, 1999

3. Asuncion M, Calvo RM, San Millan JL, et al: A prospective study of the prevalence of the polycystic ovary syndrome in unselected Caucasian women from Spain. J Clin Endocrinol Metab 85:4182, 2000

4. Reindollar RH, Novak M, Tho SP, et al: Adult-onset amenorrhea: a study of 262 patients. Am J Obstet Gynecol 155:531, 1986

5. Smith S, Ravnikar VA, Barbieri RL: Androgen and insulin response to an oral glucose challenge in hyperandrogenic women. Fertil Steril 48:72, 1987

6. Waldstreicher J, Santoro NF, Hall JE, et al: Hyperfunction of the hypothalamic-pituitary axis in women with polycystic ovary disease: indirect evidence for partial gonadotroph desensitization. J Clin Endocrinol Metab 66:165, 1988

7. Barbieri RL, Makris A, Randall RW, et al: Insulin stimulates androgen accumulation in incubations of ovarian stroma obtained from women with hyperandrogenism. J Clin Endocrinol Metab 62:904, 1986

8. Pochi PE, Strauss JS: Endocrinologic control of the development and activity of the human sebaceous gland. J Invest Dermatol 62:191, 1974

9. Joffe H, Taylor AE, Hall JE: Polycystic ovarian syndrome: relationship to epilepsy and antiepileptic drug therapy (editorial). J Clin Endocrinol Metab 86:2946, 2001

10. Longcope C, Johnstone CC: Androgen and estrogen dynamics in pre- and postmenopausal women: a comparison between smokers and non-smokers. J Clin Endocrinol Metab 67:379, 1988

11. Ferriman D, Gallwey JD: Clinical assessment of body hair growth in women. J Clin Endocrinol Metab 21:1440, 1961

12. Weyer C, Bogardus C, Mott DM, et al: The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. J Clin Invest 104:787, 1999

13. Dunaif A: Insulin resistance and the polycystic ovary syndrome: mechanism and implications for pathogenesis. Endocrin Rev 18:774, 1997

14. Bohlke K, Cramer D, Barbieri RL: Relation of luteinizing hormone levels to body mass index in premenopausal women. Fertil Steril 69:500, 1998

15. Chang JR, Nakamura RM, Judd HL, et al: Insulin resistance in nonobese patients with polycystic ovarian disease. J Clin Endocrinol Metab 57:356, 1983

16. Glickman SO, Rosenfield L: Androgen metabolism by isolated hairs from women with idiopathic hirsutism is usually normal. J Inv Derm 82:62, 1984

17. Escobar-Morreale HF, Serrano-Gotarredona J, Garcia-Robles R, et al: Mild adrenal and ovarian steroidogenic abnormalities in hirsute women without hyperandrogenism: does idiopathic hirsutism exist? Metabolism 46:902, 1997

18. Hughesdon PE: Morphology and morphogenesis of the Stein Leventhal ovary and of so called hyperthecosis. Obstet Gynecol Surv 37:58, 1982

19. Adashi EY: Potential utility of gonadotropin releasing hormone agonists in the management of ovarian hyperandrogenism. Fertil Steril 53:765, 1990

20. Ruchhoft EA, Elkind-Hirsch KE, Malinak R: Pituitary function is altered during the same cycle in women with polycystic ovary syndrome treated with continuous or cyclic oral contraceptives or a gonadotropin releasing hormone agonist. Fertil Steril 66:54, 1996

21. Spritzer P, Billaud L, Thalabard JC, et al: Cyproterone acetate versus hydrocortisone treatment in late onset adrenal hyperplasia. J Clin Endocrinol Metab 70:642, 1990

22. O'Brien RC, Cooper ME, Murray RML, et al: Comparison of sequential cyproterone acetate/estrogen versus spironolactone plus oral contraceptive in the treatment of hirsutism. J Clin Endocrinol Metab 72:1008, 1991

23. Wong IL, Morris RS, Chang L, et al: A prospective randomized trial comparing finasteride to spironolactone in the treatment of hirsute women. J Clin Endocrinol Metab 80:233, 1995

24. Andrade RJ, Lucena MI, Fernandez MC, et al: Fulminant liver failure associated with flutamide therapy for hirsutism (letter). Lancet 353:983, 1999

25. Bates GW, Whitworth NS: Effect of body weight reduction on plasma androgens in obese infertile women. Fertil Steril 38:406, 1982

26. Pasquali R, Antenucci D, Casimirri F, et al: Clinical and hormonal characteristics of obese and amenorrheic women before and after weight loss. J Clin Endocrinol Metab 8:173, 1989

27. Clark AM, Thornley B, Tomlinson L, et al: Weight loss in obese infertile women results in improvements in reproductive outcome for all forms of fertility treatment. Human Reprod 13:1502, 1998

28. Hollman M, Runnebaum B, Gerhard I: Effects of weight loss on the hormonal profile in obese infertile women. Human Reprod 11:1884, 1996

29. Clark AM, Ledger W, Galletly C, et al: Weight loss results in significant improvement in pregnancy and ovulation rates in anovulatory obese women. Human Reprod 10:2705, 1995

30. Guzick DS, Wing R, Smith D, et al: Endocrine consequences of weight loss in obese, hyperandrogenic, anovulatory women. Fertil Steril 1:598, 1994

31. Elkind-Hirsch KE, Anania C, Mack M, et al: Combination gonadotropin releasing hormone agonist and oral contraceptive therapy improves treatment of hirsute women with ovarian hyperandrogenism. Fertil Steril 63:970, 1995

32. Rittmaster RS, Loriaux DL, Cutler GB: Sensitivity of cortisol and adrenal androgens to dexamethasone suppression in hirsute women. J Clin Endocrinol Metab 81:462, 1985

33. Carmina E, Lobo RA: Ovarian suppression reduces clinical and endocrine expression of late-onset congenital adrenal hyperplasia due to 21-hydroxylase deficiency. Fertil Steril 62:738, 1994

34. Gault DT, Grobbelaar AO, Grover R, et al: The removal of unwanted hair using a ruby laser. Br J Plastic Surg 52:173, 1999

35. Sommer S, Render C, Sheehan-Dare R: Facial hirsutism treated with the normal-mode ruby laser: results of a 12-month follow-up study. J Am Acad Dermatol 41:974, 1999

36. Balfour JA, McClellan K: Topical eflornithine. Am J Clin Dermatol 2:197, 2001

37. Lucky AW, Henderson TA, Olson WH, et al: Effectiveness of norgestimate and ethinyl estradiol in treating moderate acne vulgaris. J Am Acad Dermatol 37:746, 1997

38. Lemay A, DeWailly SD, Grenier R, et al: Attenuation of mild hyperandrogenic activity in postpubertal acne by a triphasic oral contraceptive containing low doses of ethinyl estradiol and *d,l*-norgestrel. J Clin Endocrinol Metab 71:8, 1990

39. Daly DC, Walters CA, Soto-Albors CE, et al: A randomized study of dexamethasone in ovulation induction with clomiphene citrate. Fertil Steril 41:844, 1984

40. Nestler JE, Jakubowicz DJ, Reamer P, et al: Ovulatory and metabolic effects of D-chiro-inositol in the polycystic ovary syndrome. N Engl J Med 340:1314, 1999

41. Nestler JE, Jakubowicz DJ, Evans WS, et al: Effects of metformin on spontaneous and clomiphene-induced ovulation in the polycystic ovary syndrome. N Engl J Med 338:1876, 1998

42. Vandermolen DT, Ratts VS, Evans WS, et al: Metformin increases the ovulatory rate and pregnancy rate from clomiphene citrate in patients with polycystic ovary syndrome who are resistant to clomiphene citrate alone. Fertil Steril 75:310, 2001

43. Sarlis NJ, Weil SJ, Nelson LM: Administration of metformin to a diabetic woman with extreme hyperandrogenemia of non-tumoral origin: management of infertility and prevention of inadvertent masculinization of a female fetus. J Clin Endocrinol Metab 84:1510, 1999

44. Morin-Papunen LC, Koivunen RM, Ruokonen A, et al: Metformin therapy improves the menstrual pattern with minimal endocrine and metabolic effects in women with polycystic ovary syndrome. Fertil Steril 69:691, 1998

45. DeLeo V, LaMarca A, Ditto A, et al: Effects of metformin on gonadotropin induced ovulation in women with polycystic ovary syndrome. Fertil Steril 72:282, 1999

46. Coetzee EJ, Jackson WU: Metformin in management of pregnant insulin-dependent diabetics. Diabetologia 16:241, 1979

47. Hasegawa I, Murakawa H, Suzuki M, et al: Effect of troglitazone on endocrine and ovulatory performance in women with insulin resistance–related polycystic ovary syndrome. Fertil Steril 71:323, 1999

48. Dunaif A, Scott D, Finegood D, et al: The insulin sensitizing agent troglitazone improves metabolic and reproductive abnormalities in the polycystic ovary syndrome. J Clin Endocrinol Metab 81:3299, 1996

49. Elkind-Hirsch KE, McWilliams RB: Pregnancy after treatment with the insulin sensitizing agent troglitazone in an obese woman with the hyperandrogenic insulin-resistant acanthosis nigricans syndrome. Fertil Steril 71:943, 1999

50. Mitwally MFM, Kuscu N, Yalcinkaya TM: High ovulatory rates with use of troglitazone in clomiphene resistant women with polycystic ovary syndrome. Human Reprod 14:2700, 1999

51. Neuschwander-Tetri BA, Isley WL, Oki JC, et al: Troglitazone induced hepatic failure leading to liver transplantation. Ann Intern Med 130:163, 1999

52. Kistner RW: Sequential use of clomiphene citrate and human menopausal gonadotropin in ovulation induction. Fertil Steril 27:72, 1976

53. Jarrell J, McInnes R, Crooke R: Observations on the combination of clomiphene citrate-hMG-hCG in the management of anovulation. Fertil Steril 35:634, 1981

54. Homburg R, Levy R, Ben Rafael Z: A comparative study of conventional regimens with low dose FSH for ovulation in PCOS. Fertil Steril 63:729, 1995

55. Schenker JG, Weinstein D: Ovarian hyperstimulation syndrome: a current survey. Fertil Steril 30:255, 1978

56. Donesky BW, Adashi EY: Surgically induced ovulation in the polycystic ovary syndrome: wedge resection revisited in the age of laparoscopy. Fertil Steril 63:439, 1995

57. Armar NA, McGarrigle HG, Honour J, et al: Laparoscopic ovarian diathermy in the management of anovulatory infertility in women with polycystic ovaries: endocrine changes and clinical outcome. Fertil Steril 53:45, 1990

58. Gjonnaess H: Polycystic ovarian syndrome treated by ovarian electrocautery through the laparoscope. Fertil Steril 41:20, 1984

59. Sanfillipo JS, Rock JA: Surgery for benign disease of the ovary. TeLinde's Operative Gynecology. Rock JA, Thompson JD, Eds. Lippincott, Williams & Wilkins, Philadelphia, 1997

60. Buyalos RP, Lee CT: Polycystic ovary syndrome: pathophysiology and outcome with in vitro fertilization. Fertil Steril 65:1, 1996

61. Wada I, Matson PL, Troup SA, et al: Assisted conception using buserelin and human menopausal gonadotropins in women with polycystic ovary syndrome. Br J Obstet Gynecol 100:3665, 1993

62. Urman B, Fluker MR, Yuen BH, et al: The outcome of in vitro fertilization and embryo transfer in women with polycystic ovary syndrome failing to conceive after ovulation induction with exogenous gonadotropins. Fertil Steril 57:1269, 1992

63. Legro RS, Kunselman AR, Dodson WC, et al: Prevalence and predictors of risk for type 2 diabetes mellitus and impaired glucose tolerance in polycystic ovary syndrome: a prospective, controlled study in 254 affected women. J Clin Endocrinol Metab 84:165, 1999

64. Ehrmann DA, Cavaghan MK, Barnes RB, et al: Prevalence of impaired glucose tolerance and diabetes in women with PCOS. Diabetes Care 22:141, 1999

65. Legro RS, Kunselman AR, Dunaif A: Prevalence and predictors of dyslipidemia in women with PCOS. Am J Med 111:665, 2001

66. Pierpoint T, McKeigue PM, Isaacs AJ, et al: Mortality of women with polycystic ovary syndrome at long term follow up. J Clin Epidemiol 51:581, 1998

VI CONTRACEPTION

Sarah L. Berga, m.d.

Contraception means to prevent conception, but in common medical usage, it also refers to methods that prevent implantation. The goal of contraception is to make every child a wanted child. Most methods of contraception (e.g., barrier methods and hormone preparations) also reduce the risk of sexually transmitted diseases (STDs), but intrauterine devices (IUDs) may increase the risk of STDs or their consequences.

Contraceptive methods are generally categorized as reversible [*see Table 1*] or irreversible. Irreversible methods are often referred to as sterilization procedures. Pregnancy termination is not typically regarded as contraception, but the availability of medical methods of abortion has blurred this distinction. Emergency postcoital contraceptive methods are also available.

Most methods of contraception are designed for use by women. Hormonal methods of male contraception are under development, but they are not yet commercially available.[1] In any case, unless such products prove to be almost free of side effects, their acceptance by men is likely to be limited.

Historically, methods to prevent or terminate pregnancy have been subject to intense legal regulation. In the United States, legal regulations vary widely from state to state regarding the provision of services to minors, waiting periods for termination and sterilization, husband and parental consent, reporting of complications and deaths, and restrictions on advance directives by pregnant or potentially pregnant women. Some states also have regulations regarding the type of practitioner or the type of facility in which contraceptive and fertility management procedures can be provided, but no state currently bans the use of reversible contraceptives. Medical insurance coverage for contraception varies widely.

The percentage of women using a contraceptive method rose from 56% in 1982 to 64% in 1995.[2] The most widely used methods in 1995 were female sterilization, combined estrogen and progestin oral contraceptives (COCs), and the male condom. Male condom use is common among unmarried couples; this popularity is due in part to the protection its use affords against certain STDs, particularly HIV infection.

Combined Estrogen-Progestin Contraceptives

COCs are formulated with estrogen (in the form of ethinyl estradiol), in doses ranging from 20 to 35 μg, and a variety of progestins derived from 19-nortestosterone. Individual formulations may have a fixed dose of progestin (monophasic) or may have doses that vary by cycle phase (triphasic). Two preparations contain varying estrogen doses.

efficacy and mechanism of action

Combined estrogen and progestin oral contraceptives are a highly effective method of birth control. Theoretical efficacy is about 99.9%, but the typical efficacy is around 97%.[3] The contraceptive effect of COCs derives principally from the suppression of the hypothalamic release of gonadotropin-releasing hormone (GnRH) and the concomitant suppression of the pituitary release of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). The decrease in LH reduces ovarian androgen secretion. The suppression of GnRH is primarily caused by the progestin component. Estrogen also independently suppresses FSH at the pituitary level, thereby retarding folliculogenesis.

COCs are progestin dominant. Progestin exposure causes endometrial decidualization and atrophy, rendering the endometrium unfavorable for implantation and thickening of the cervical mucus, thereby blocking the entry of sperm and bacteria into the upper genital tract. There are three generations of progestins. Third-generation progestins, which are theoretically less androgenic, include desogestrel, gestodene, and norgestimate. The clinical superiority of one progestin over any other has not been demonstrated.

reduction of androgen exposure

All COC preparations reduce androgen exposure by two mechanisms: (1) suppression of ovarian androgen production, as a result of the reduction in LH stimulation of the ovarian theca compartment and (2) elevation of the level of sex hormone–binding globulin protein, which binds androgens and thereby lowers the unbound fraction of circulating androgens. COCs reduce facial or androgen-dependent hair growth and acne by reducing the circulating concentrations of androgens available to occupy androgen receptors on the pilosebaceous unit. Although women with polycystic ovary syndrome have elevated GnRH and LH levels, COCs containing 30 to 35 μg of estrogen adequately suppress their androgen secretion.[4]

continuous versus cyclic coc regimens

Traditional COCs follow a 28-day cycle, with 21 days of active pills and 7 days of placebo pills. The 7-day placebo window permits significant follicular development, and higher estrogen and progestin doses are then needed to inhibit further folliculogenesis and ovulation. Further, if there is a delay in starting the next pill pack, a so-called escape ovulation may result.

In patients taking COCs to effect ovarian suppression for medical purposes, this 28-day cycle may be ineffective. These patients may benefit from a long-cycle regimen—42 to 105 days of active pills, and then 1 week off—or a continuous regimen.

Table 1 Categories of Commercially Available Reversible Contraceptives

- Barrier methods
  - Condoms
  - Diaphragms
  - Cervical caps
- Hormonal
  - Combined estrogen-progestin preparations
    - Oral
    - Vaginal ring
    - Transdermal
    - Injectable
  - Progestin preparations
    - Oral
    - Injectable
    - Implantable
- Intrauterine devices

For long-cycle regimens, patients use the active pills from several packs; however, a commercial long-cycle preparation that includes 63 days' worth of active pills is under development. The use of continuous COC regimens has not been well studied, but it is common to omit the placebo week in women undergoing hormonal treatment for disorders such as polycystic ovary syndrome or endometriosis, as well as in those who desire amenorrhea or have headaches or other symptoms provoked by estrogen withdrawal during the placebo week. There are several advantages to a continuous approach, including better suppression of ovarian function. Persistent ovarian cysts may be less likely with a continuous regimen than with a cyclic regimen.

Increased breakthrough bleeding is a potential side effect of a continuous regimen as compared with a cyclic regimen. The increase in breakthrough bleeding is attributable partly to the development of fragile endometrial vessels coursing along the surface of the endometrium and partly to impaired local hemostasis.[5] The only near-continuous COC preparation that avoids the increase in breakthrough bleeding while increasing follicular suppression is Mircette, which has 2 placebo days, 5 days of 10 µg estrogen, and 21 days of 20 µg estrogen plus 150 µg desogestrel.[6] One objective of these modifications is to lower overall sex steroid exposure and minimize the unwanted consequences of pill use while minimizing breakthrough bleeding.

nonoral combined contraceptives

Several combined estrogen-progestin contraceptives that do not use the oral route have recently been developed. All offer the convenience of less-frequent dosing.

The hormonal vaginal contraceptive ring (NuvaRing) was approved by the Food and Drug Administration in October 2001. This product is a flexible polymer ring, about 2 in. in diameter, which the patient inserts in her vagina. The ring releases a continuous low dose of estrogen and etonogestrel. The ring is left in place for 3 weeks, then removed for the week during which the patient will have her menstrual period. Neither patients nor their partners can tell that the ring is in place. Comparison of the vaginal ring with a COC has shown a lower incidence of irregular bleeding and a higher incidence of a normal intended bleeding pattern with the ring.[7]

The transdermal contraceptive patch (Ortho Evra) was approved by the FDA in November 2001. The patch delivers estrogen and norelgestromin over the course of a week. Patches are applied once a week for 3 consecutive weeks to the skin of the buttocks, the abdomen, the upper torso, or the upper outer arm.[8] In general, the efficacy and cycle control provided by the patch are comparable to those of COCs, but the efficacy of the patch may be lower in women who weigh 198 lb (90 kg) or more.[9] The overall rate of patch detachment is about 4%; about 2% of users experience skin irritation at the site of application.

An injectable estrogen-progestin contraceptive (Lunelle) is available in the United States. The preparation, which is given intramuscularly once a month, contains medroxyprogesterone acetate and estradiol cypionate in a timed-release form. In clinical trials, efficacy and patient satisfaction have been comparable to that seen with COCs.[10]

side effects

A myriad of serious and nuisance side effects are associated with COCs [*see Table 2*]. Smoking markedly increases the risk of venous thromboembolism. Smoking is a relative contraindication to COC use, particularly in women older than 35 years.[11]

Table 2 Potential Side Effects of Oral Contraceptives

Serious side effects	Metabolic changes Decreased insulin action Increase in clotting factors Elevation of triglyceride levels Increase in the metabolic work load of the liver Increase in renin substrate Clinical manifestations* Venous thromboembolic events Fatty liver or hepatoma Cholestasis or cholecystitis Diabetes mellitus Hypertension Cardiovascular events Exacerbation of depression Drug interactions
Nuisance side effects	Mastodynia Reduced libido Reduced vaginal lubrication Increased appetite Weight gain Fatigue Bloating

*These events are rare in healthy women younger than 50 years.

The crucial clinical issue is to convince the woman who smokes to stop smoking rather than deny her access to an acceptable form of contraception. For nonsmokers who take COCs containing 35 µg or less of estrogen, the risk of nonfatal venous thromboembolism is approximately one half that of pregnancy (60 per 100,000 women) but greater than that observed in healthy women who do not take oral contraceptives (5 per 100,000 women). The alleged excess mortality from venous thromboembolism attributable to third-generation progestins as compared with other progestins is less than two per million women per year.[12] However, women with familial thrombophilia caused by factor V Leiden or prothrombin mutation 20210A have a greatly increased risk of venous thromboembolism when using any oral contraceptive[13] [*see 5:XIV Thrombotic Disorders*]. In white women, the carrier rate of factor V Leiden is approximately 3% and that of prothrombin mutation 20210A is less than 2%, so screening has not been routinely advised.[14] Also, there are other known causes of thrombophilia, but not all the thrombophilias can be detected.

COCs may decrease insulin action, an effect that has been attributed to the progestin component. The use of COCs does not increase the risk of diabetes mellitus in women who do not have other risk factors for the disease. However, in a nonrandomized clinical trial that followed Latin-American women with gestational diabetes for 7.5 years post partum, the rate of development of diabetes mellitus was 8.7% for those given nonhormonal contraception, 10.4% for those who used COCs, and 26.5% for those who used progestin-only pills. Life-table analysis showed an increase in diabetes mellitus within 2 years in the progestin-only group.[15] Given these considerations, it is prudent to avoid prescribing any progestin-only form of contraception in women predisposed to diabetes mellitus or in frankly diabetic women. Fortunately, COC use by insulin-dependent diabetic women does not increase the risk of diabetic retinopathy or nephropathy.[16]

COCs have a negligible effect on the overall risk of cancer.[17] Among women who take oral contraceptives for 8 years, the estimated increase in the number of cases of cancer is 125 per 100,000 for cervical cancer and 41 per 100,000 for liver cancer; those increases are offset, however, by decreases in endometrial cancer and ovarian cancer [*see* Benefits, *below*].[17] Oral contraceptive use does not appear to increase the risk of breast cancer significantly—regardless of the dose, duration of use, or age at use—even in women with a family history of breast cancer.[18] However, the effect of oral contraceptive use on risk of breast cancer in carriers of *BRCA1* and *BRCA2* has not been defined.

Exposure to high doses of estrogen or progestin may provoke depression and mood disturbances, but this effect is limited to women with an underlying diathesis.[19] It is not known whether oral contraceptive use increases the lifetime risk of depression or hastens its onset in women so predisposed.

Progestins have mineralocorticoid activity that results in the retention of up to 2 lb of water in sensitive women. The progestins also increase plasma renin activity; in predisposed women, hypertension may result.

Oral contraceptive use can cause drug interactions by increasing liver production of proteins that bind other drugs, by inhibiting oxidative metabolism in the liver by the P-450 and P-448 cytochrome systems, and by competing for or accelerating conjugation. Conversely, drugs that stimulate the hepatic microsomal system, such as oral antifungal agents or rifampin, may decrease plasma levels of contraceptive steroid and lead to unintended pregnancy. Antibiotics such as ampicillin, tetracyclines, and metronidazole do not interfere with COC efficacy.

contraindications

There are specific contraindications to oral contraceptives [*see Table 3*]. Lactating women should probably not take COCs. Women with hypertension, epilepsy, depression, hepatitis, gallbladder disease, migraine, or premenstrual syndrome (PMS) need to be carefully monitored. Fibroids are not a contraindication to COC use.

benefits

Women use oral contraceptives primarily for birth control. If side effects are tolerable, they are pleased to gain the other benefits [*see Table 4*]. Women generally appreciate the lighter and predictable menses. The option of long cycles or continuous use to schedule or skip bleeding episodes is a major advantage for any busy woman, particularly one who travels or spends time outdoors. Other women use oral contraceptives to treat an underlying disorder, such as polycystic ovary syndrome, dysmenorrhea, endometriosis, or idiopathic hirsutism. In general, the same benefits accrue.

Table 3 Contraindications to Oral Contraceptives

- Active liver or gallbladder disease
- Medically significant hypertriglyceridemia
- Active or past venous thromboembolic events
- Atherosclerotic heart disease
- Undiagnosed vaginal bleeding
- Estrogen-dependent neoplasia
  - Breast cancer
  - Endometrial cancer
- Symptomatic mitral valve prolapse
- Smoking after 35 years of age

Table 4 Potential Benefits of Oral Contraceptives

- Reduced risk of the following disorders:
  - Ectopic pregnancy
  - Benign breast disease
  - Anemia
  - Ovarian cysts and cancer
  - Endometrial cancer
- Lighter and predictable menses
- Reduction or elimination of dysmenorrhea
- Bone accretion

One of the major benefits of COC use is bone accretion. One study showed that as little as 10 µg of estrogen was bone sparing in women older than 40 years, and 5 µg of estrogen plus 1 mg of norethindrone caused bone accretion.[20] Women who are hyperandrogenic but eumenorrheic have greater bone mass than do hyperandrogenic women who are oligomenorrheic. The latter have slightly higher bone mass than eumenorrheic, nonhirsute women. Women with polycystic ovary syndrome or idiopathic hirsutism who take oral contraceptives will have a decrement in endogenous androgen exposure, but COC use in this setting is expected to be bone sparing. Women with hypothalamic hypogonadism, particularly those with an eating disorder, have underlying metabolic disturbances that render their bones less responsive to exogenous steroid exposure. These women may continue to lose or not accrue bone even if they use COCs.

Long-term use of COCs reduces the incidence of endometrial cancer and ovarian cancer. Among women who take oral contraceptives for 8 years, it is estimated that there will be 197 fewer cases of endometrial cancer per 100,000 users and 193 fewer cases of ovarian cancer per 100,000 users.[17] Newer COCs, which contain 20 µg of estrogen, appear to provide identical risk reduction for ovarian cancer as did older formulations, which contained 50 µg or more of estrogen.[21] A recent study showed that oral contraceptives also markedly reduce the risk of ovarian cancer in carriers of *BRCA1* and *BRCA2* mutations (who are at increased risk for ovarian cancer and premenopausal breast cancer). The longer the duration of use, the greater is the protection from ovarian cancer.[22]

Progestin-Only Contraceptives

efficacy and mechanism of action

Two progestin-only contraceptives are commercially available. Norplant is composed of Silastic rods impregnated with the progestin levonorgestrel. The new Norplant system contains only two rods. A single-rod system will soon be available. The rods are inserted subdermally, generally in the upper arm. Diffusion of levonorgestrel through the wall of each capsule provides a continuous low dose of progestin for at least 5 years. The progestin modestly inhibits the hypothalamic-pituitary-ovarian axis to block ovulation; it also induces endometrial shedding, making implantation unlikely, and thickens cervical mucus,

thereby retarding the entry of sperm and bacteria to the upper genital tract. The birth-control efficacy is greater than 99.9%.

Depot medroxyprogesterone acetate (DMPA) is an aqueous suspension of 150 mg designed to be given intramuscularly every 3 months. The birth-control efficacy is greater than 99%.

side effects

Subdermal and injectable progestins have side effects [*see Table 5*]. The principal side effect is breakthrough bleeding caused by the development of fragile endometrial vessels and local derangement of hemostatic mechanisms as a result of excess progestin exposure relative to estrogen exposure.[5] The breakthrough bleeding may respond to the administration of an estrogen such as transdermal estradiol. Progestin implants and injections are relatively contraindicated in women with past or active depression or other psychiatric disorders [*see Table 6*]. There is some suggestion that progestin-only contraceptives are more mood destabilizing than COCs. The long-term effect on bone accretion depends on the extent of ovarian suppression and its attendant decline in estradiol secretion and on the age of the patient. Younger patients who have not attained peak bone mass may be more adversely affected. Levonorgestrel may be more bone sparing than DMPA. Progestin-only contraceptives have been found to increase the risk of diabetes mellitus in Latin-American women who have had gestational diabetes (see above).[15] This may be partly caused by the lack of estrogen, which is an insulin sensitizer. The long-term cardiovascular risks are largely unknown, but in some experimental settings, synthetic progestins provoke vasoconstriction, an effect not seen with progesterone. A recent epidemiologic analysis from the World Health Organization (WHO) found no excess risk of cardiovascular disease with either combined or progestin-only methods other than an increased risk of stroke in hypertensive women who were given the progestin-only contraceptives.[23]

Another common side effect of progestin-only contraceptives is delay in return of menses. DMPA is given at 90-day intervals, but patients who discontinue this method may not experience immediate return of menses because of variability in the metabolism of the depot form and variable sensitivity of the hypothalamic GnRH pulse generator to low levels of progestin.

Intrauterine Devices

IUDs interfere with sperm migration, fertilization, ovum transport, and implantation, presumably by causing a sterile salpingitis, endometritis, or both. The birth-control efficacy is greater than 97%.

One of the principal benefits of IUDs is that they provide a nonhormonal method of birth control. They are ideal for women who have completed childbearing and who desire a low-maintenance, reversible method of contraception. IUDs are also relatively economical.

Table 5 Potential Side Effects of Injectable and Implantable Progestin Contraceptives

Breakthrough bleeding	Mastodynia
Headaches	Acne
Mood changes	Bone loss
Weight gain	

Table 6 Contraindications to Injectable and Implantable Progestin Contraceptives

Active liver disease	Active thromboembolic disease
Diabetes mellitus	Active cardiovascular disease
Unexplained vaginal bleeding	Depression
Breast cancer	Other psychiatric disorders

Two IUDs are currently available: the Copper T 380A, which lasts 10 years, and the 5-year, levonorgestrel-releasing Mirena. With the copper IUD, both the inert plastic device and the copper contribute to the spermicidal effect and prevention of implantation. With the progestin-containing IUD, part of the efficacy is attributed to the effects of the progestin on the endometrium that retard implantation.[24]

The main side effect associated with IUD use is pelvic inflammatory disease (PID). Most of the increased risk of PID occurs in the first 3 weeks after insertion. Women with more than one sexual partner who are at risk for contracting gonorrhea and chlamydial infection also are at increased risk for PID. Patient selection, rigorous aseptic insertion technique, and screening for STDs may minimize this risk. Routine antibiotic prophylaxis during insertion may not be necessary.[25] Uterine perforation is a rare insertion risk. Dysmenorrhea and menorrhagia have been reported with the copper IUD, whereas decreased menstrual flow, dysmenorrhea, and increased risk of ectopic pregnancy have been reported with the progesterone-releasing IUD.

The primary contraindication to IUD use is a history of PID. Nulligravidity is a relative contraindication. Sexual monogamy should be emphasized as a means of minimizing the risk of STDs and PID.

Barrier Methods

Male and female barrier contraceptives are available. When used correctly, the male condom protects against pregnancy and STDs. The theoretical efficacy of barrier methods for birth control is 98%, but the actual efficacy is about 88%. The efficacy gap results from inconsistent use and condom breakage. The female condom is more difficult to use and has not gained popularity. The diaphragm and cervical cap do not protect against STDs as effectively as condoms. When they are used with spermicides, the birth-control efficacy of diaphragms and cervical caps is theoretically 94%; in practice, however, the efficacy is about 82%. Both cervical caps and diaphragms require fitting and a prescription. They also require user training and diligence. Instructions on their use are provided in the products' package inserts. Spermicides may independently decrease the risk of STDs. When used alone, spermicides have a birth-control efficacy of about 79%.[3] Spermicides that also have antimicrobial activity are in development.

Allergic reactions to latex and hypersensitivity to spermicides occur. Diaphragm use may increase the risk of urinary tract infections because the rim presses against the symphysis pubis and urethra, which may cause incomplete emptying of the bladder.

The main contraindication to barrier methods is lack of user motivation and hypersensitivity to spermicides or allergy to latex. The primary benefits of condoms are that they are avail-

able without prescription, inexpensive, relatively easy to use, nonhormonal, and protective against STDs. Other barrier methods are only slightly more difficult to use but require fitting and a prescription, so the need for birth control must be anticipated.

Periodic Abstinence

Periodic abstinence, or natural family planning, depends on recognition of the periovulatory window and avoidance of sexual intercourse during that window. As such, it requires that a woman have highly regular menstrual cycles and that both partners be motivated to avoid intercourse when the woman is fertile. There are several methods of detecting the fertile window, including avoiding intercourse on days 9 to 14 of a 28-day cycle, monitoring cervical mucus and body temperature, and monitoring salivary estradiol levels.

Successful use of fertility-awareness methods for birth control requires not only dedication but education. Family health centers, family planning centers, and church-affiliated centers may offer courses on this subject. Information is available on the Internet at sites such as http://my.webmd.com/encyclopedia/article/1819.51010.

Mastering the concepts of menstrual-cycle physiology can be empowering, and couples can use this information to plan a pregnancy as well as to avoid it. There are no known contraindications. There are no religious prohibitions against periodic abstinence, so it is theoretically available to all women who have predictable cycles.

Periodic abstinence can be frustrating, however, and it is less reliable than other forms of contraception, with an estimated efficacy of 80%. Several factors can interfere with fertility awareness. Even women who usually have very regular cycles may occasionally have a cycle that deviates from normal. Vaginitis may obscure the recognition of midcycle mucus. Fever may mimic the progesterone-induced rise in body temperature that normally indicates the onset of the luteal phase, thereby falsely signaling that the fertile period has passed.

Sterilization

Sterilization procedures generally entail occlusion or ligation of the fallopian tubes in women or the vas deferens in men. The birth-control efficacy of sterilization procedures is greater than 99%; they are meant to be permanent. Reversal procedures are available, but the reversibility of tubal ligation or vasectomy is not guaranteed. Sterilization procedures may fail if the fallopian tube is not properly identified or if it recannulates. Vasectomy failures primarily result from not waiting a sufficient length of time after the procedure before having unprotected sexual intercourse. In women, sterilization procedures can be done post partum, but interval procedures are safer and more effective. Interval procedures employ laparoscopy, with or without general anesthesia. The fallopian tubes are either fulgurated or banded.

Patients may experience feelings of regret after a tubal ligation or vasectomy. Appropriate counseling can minimize this emotional side effect. There is no concrete evidence that vasectomy causes heart disease or prostate cancer. A recent review of tubal ligation found no evidence of increased rates of premenstrual distress, menorrhagia, dysmenorrhea, or menstrual irregularities in women 30 years of age or older who had undergone interval tubal ligation.[26]

The main contraindication to sterilization is ambivalence. In addition, women who undergo laparoscopic procedures must be suitable surgical candidates. The main benefit of sterilization is its permanence. Because sterilization is a one-time procedure with high efficacy, it is highly cost-effective in appropriately selected candidates. Tubal ligation may decrease the risk of PID and ovarian cancer.

Emergency Contraception

Postcoital contraception aims to desynchronize endometrial development and prevent implantation. Various methods have been proposed.[27] They include high doses of COCs taken within 72 hours after intercourse; levonorgestrel taken within 72 hours after intercourse; high doses of estrogen; danazol; mifepristone, as a single 600 mg dose; and insertion of a copper IUD up to 5 days after ovulation. The contraceptive efficacy of mifepristone or IUDs is at least 99%. One study compared a treatment consisting of 100 µg of estrogen plus 0.5 mg of levonorgestrel taken twice, 12 hours apart, with a treatment consisting of levonorgestrel, 0.75 mg, taken twice, 12 hours apart.[28] Both treatments were taken within 72 hours after intercourse. Levonorgestrel alone had an efficacy of 85% and was associated with fewer side effects than the combined therapy, which had an efficacy of 57%. The efficacy of levonorgestrel taken within 24 hours after intercourse was greater than 99%. This would appear to be the treatment of choice because it is inexpensive, widely available, well tolerated, and highly effective. A recent Scottish study also suggested that women given a single emergency contraceptive kit used it correctly without experiencing significant side effects, and they had a lower unintended pregnancy rate.[29] Given the safety and efficacy of emergency contraception, many physicians strongly advocate that it be made available over the counter.[30] Patient information on emergency contraception is available on the Internet at http://ec.princeton.edu.

Choosing a Contraceptive Method

There is no perfect contraceptive; all may fail, and all have drawbacks and side effects. Age, motivation, marital status, partner attitude, perceived risk of pregnancy, frequency of intercourse, medical conditions, costs, cultural considerations, and religious beliefs affect the choice of contraceptive methods. The patient's or couple's medical history and preferences must guide the selection of a contraceptive [*see Table 7*]. Patients should be encouraged to revise their choice on the basis of side effects and changing circumstances.

Patients should be advised to inform the physician of new symptoms before discontinuing a contraceptive method. The physician must remain sensitive to the patient's concerns. Even if a symptom sounds trivial from a medical perspective, it may alarm the patient and cause her to discontinue the method.

The role of condoms and other contraceptives in the reduction of STD transmission must be emphasized so that patients can choose properly from among the available options. Emergency contraception should be discussed and offered to those not seeking long-term contraception.

reversible method desired

The first decision point in the choice of a reversible contraceptive method hinges on whether the patient has more than one sexual partner or is in a long-term monogamous relationship. If

Table 7 Contraceptive Characteristics Affecting Choice[29]

Characteristic	*Method*
High efficacy	Combined oral contraceptives Intrauterine devices Depot medroxyprogesterone acetate Subdermal progestin implants
Limited or no systemic side effects	Barriers Spermicides Periodic abstinence
Minimal effort	Intrauterine devices Subdermal progestin implants Depot medroxyprogesterone acetate
Low cost	Male condom Spermicides Combined oral contraceptives
Nonprescription	Male condom Spermicides Periodic abstinence
No religious prohibitions	Periodic abstinence
Protection against sexually transmitted diseases	
Cervical gonorrhea and chlamydial infection	Barriers
Salpingitis	Barriers, hormone contraceptives
HIV infection	Male and female latex condoms
Other health benefits	Hormone contraceptives
Minimal risk to future fertility	Hormone contraceptives Barriers Periodic abstinence

the patient has more than one sexual partner, condoms with or without hormonal contraception should be recommended. User reluctance and lack of familiarity are the main limitations to condom use. Condoms are ideal for unplanned intercourse.

For healthy women, the ancillary health benefits of combined estrogen-progestin contraceptives should be emphasized. These include a reduced risk of ovarian and endometrial cancer and preservation or accretion of bone mass. The option of using combined hormonal contraceptives to regulate menstrual timing should be discussed as a means of aiding compliance.

COCs, particularly the generic brands, are relatively inexpensive—in the range of $10 to $20 a month. COCs work best if taken daily, and some women find it difficult to remember to do so; they may prefer a vaginal ring, transdermal patch, or injectable contraceptive. Women who do not take the pills reliably will have increased rates of pregnancy and side effects, such as breakthrough bleeding. They should be counseled to use a barrier method or spermicide if they miss two consecutive pills or if they start the next package of pills after a hiatus of 8 or more days. In healthy nonsmokers without predisposing medical conditions, the pill is a safe and highly efficacious method of contraception that can be used in women up to 50 years of age. In women who are approaching menopause, COC use not only provides effective contraception but also can regularize menstrual cycles, relieve vasomotor symptoms, and stabilize bone mass.[31]

Women with epilepsy may need to have their antiseizure medications adjusted when they start oral contraceptive therapy. Not all the newer antiseizure drugs interact with oral contraceptives, however,[32] so patients need to be evaluated on a case-by-case basis. Consultation with a pharmacist may be useful.

Follow-up is important in women who choose hormonal contraceptives. Blood pressure should be measured around 3 months after the start of COCs and annually thereafter. If hypertension results, it is prudent to discontinue COCs. Women using hormonal contraception who develop a severe, unremitting headache should be evaluated for possible stroke and cerebral thrombosis.

With prolonged use of COCs—even on a cyclic regimen—some women develop endometrial atrophy and amenorrhea. Once pregnancy has been excluded, it is prudent to recommend a long-cycle or a continuous regimen or one with a shortened placebo window.

It is prudent for women who develop serious mood disturbances to discontinue COCs. Women with active or past PMS and depression, including postpartum depression, should be advised about the potential for negative mood effects associated with COCs.

Switching to a lower-dosage regimen may reduce nuisance side effects. Physically smaller women or women who metabolize synthetic sex steroids slowly (such as women of Asian descent) should start with a 20 μg pill. Although most women do well on 20 μg pills, there may be a slight increase in breakthrough bleeding with some formulations. The benefit is fewer estrogen-dependent side effects, such as breast tenderness or nausea. Women with a history of headaches may do better on the lowest dose given in a continuous or nearly continuous (Mircette) regimen.

If a patient has a low risk of depression, PMS, and osteoporosis, Norplant or DMPA may be an appropriate method of contraception. The patient must desire extended protection and be willing to undergo the insertion procedure or an injection. Subdermal implants and DMPA are relatively expensive.

The DMPA cannot be removed if a patient experiences adverse effects, and continuance rates are low with DMPA. Insertion of subdermal implants requires training and skill. If adverse effects occur with subdermal implants, a surgical procedure is required for their removal. Removal is often more difficult than introduction because of scarring around the capsules. In appropriate patients, however, subdermal implants provide a long-term, low-maintenance birth-control method.

IUDs, barrier methods, or periodic abstinence may be considered if the patient is not a candidate for hormone contraception and is in a long-term monogamous relationship.

Prospective users of IUDs must be made aware of the attendant risks and benefits. To make an informed decision, users need to understand the risks and potential consequences of PID.

Diaphragms and cervical caps are ideal for highly motivated users who desire a nonhormonal method of birth control. Spermicides increase the efficacy of all barrier methods of contraception but also increase cost and bother. Use of barrier methods is increased by public health education and by suggestion that the barrier method be incorporated into foreplay.

Women who report vaginal pruritus, irritation, inflammation, pain, or discharge associated with the use of a barrier method may have a latex allergy or hypersensitivity to spermicide. Latex allergies are particularly common in health care workers, many of whom are women. Formal allergy testing

can be done to detect latex allergy, but current tests are not highly reliable. Latex allergy can provoke life-threatening anaphylaxis. Although most spermicides contain nonoxynol-9 as the active contraceptive ingredient, the other constituents may vary. Therefore, it may be possible to minimize irritation by switching to a different preparation. Men may also report allergies to latex or hypersensitivity to spermicides. Latex is preferred for condoms and barrier methods because it is impermeable to HIV.

nonreversible method desired

All potential reversible and permanent options must be reviewed before a sterilization procedure is chosen. In general, if a couple seeks sterilization, a vasectomy is recommended because it is safer and less expensive than a laparoscopic tubal ligation. Postpartum tubal ligation is the most risky, least effective, and most likely to cause regret.

Laparoscopic tubal ligation is expensive and must be performed by a skilled surgeon in an appropriate setting, so availability may be limited by cost or access to an appropriate physician. Many states require mandatory waiting periods. Some require spousal consent. Counseling is the cornerstone of success.

References

1. Anawalt BD, Amory JK: Male hormonal contraceptives. Expert Opin Pharmacother 2:1389, 2001

2. Piccinino LJ, Mosher WD: Trends in contraceptive use in the United States: 1982–1995. Fam Plann Perspect 30:4, 1998

3. Trussell J, Hatcher RA, Cates W Jr, et al: A guide to interpreting contraceptive efficacy studies. Obstet Gynecol 76:558, 1990

4. Daniels TL, Berga SL: Resistance of gonadotropin releasing hormone drive to sex steroid-induced suppression in hyperandrogenic anovulation. J Clin Endocrinol Metab 82:4179, 1997

5. Runic R, Schatz F, Krey L, et al: Alterations in endometrial stromal cell tissue factor protein and messenger ribonucleic acid expression in patients experiencing abnormal uterine bleeding while using Norplant-2 contraception. J Clin Endocrinol Metab 82:1983, 1997

6. Killick SR, Fitzgerald C, Davis A: Ovarian activity in women taking an oral contraceptive containing 20 microgram ethinyl estradiol and 150 microgram desogestrel: effects of low estrogen doses during the hormone-free interval. Am J Obstet Gynecol 179:S18, 1998

7. Bjarnadottir RI, Tuppurainen M, Killick SR: Comparison of cycle control with a combined contraceptive vaginal ring and oral levonorgestrel / ethinyl estradiol. Am J Obstet Gynecol 186:389, 2002

8. Burkman RT: The transdermal contraceptive patch: a new approach to hormonal contraception. Int J Fertil Womens Med 47:69, 2002

9. Zieman M, Guillebaud J, Weisberg E, et al: Contraceptive efficacy and cycle control with the Ortho Evra / Evra transdermal system: the analysis of pooled data. Fertil Steril 77(2 suppl 2):S13, 2002

10. Kaunitz AM: Lunelle monthly injectable contraceptive: an effective, safe, and convenient new birth control option. Arch Gynecol Obstet 265:119, 2001

11. Farley TM, Collins J, Schlesselman JJ: Hormonal contraception and risk of cardiovascular disease: an international perspective. Contraception 57:211, 1998

12. Hampton N, Kubba A: Contraception: a slow train gathers speed. Lancet 352 (suppl 4):SIV3, 1998

13. Martinelli I, Sacchi E, Landi G, et al: High risk of cerebral-vein thrombosis in carriers of a prothrombin-gene mutation and in users of oral contraceptives. N Engl J Med 338:1793, 1998

14. Bertina RM, Rosendaal FR: Venous thrombosis: the interaction of genes and environment. N Engl J Med 338:1840, 1998

15. Kjos SL, Peters RK, Xiang A, et al: Contraception and the risk of type 2 diabetes mellitus in Latina women with prior gestational diabetes mellitus. JAMA 280:533, 1998

16. Garg SK, Chase HP, Marshall G, et al: Oral contraceptives and renal and retinal complications in young women with insulin-dependent diabetes mellitus. JAMA 271:1099, 1994

17. Schlesselman JJ: Net effect of oral contraceptive use on the risk of cancer in women in the United States. Obstet Gynecol 85:793, 1995

18. Marchbanks PA, McDonald JA, Wilson HG, et al: Oral contraceptives and the risk of breast cancer. N Engl J Med 346:2025, 2002

19. Schmidt PJ, Nieman LK, Danaceau MA, et al: Differential behavioral effects of gonadal steroids in women with and in those without premenstrual syndrome. N Engl J Med 338:209, 1998

20. Speroff L, Rowan J, Symons J, et al: The comparative effect on bone density, endometrium, and lipids of continuous hormones as replacement therapy (CHART study): a randomized controlled trial. JAMA 276:1430, 1996

21. Risk of ovarian cancer in relation to estrogen and progestin dose and use characteristics of oral contraceptives. SHARE Study Group. Steroid Hormones and Reproductions. Am J Epidemiol 152:233, 2000

22. Oral contraceptives and risk of hereditary ovarian cancer. Hereditary Ovarian Cancer Clinical Study Group. N Engl J Med 339:424, 1998

23. Cardiovascular disease and use of oral and injectable progestogen-only contraceptives and combined injectable contraceptives: results of an international, multicenter, case-control study. World Health Organization Collaborative Study of Cardiovascular Disease and Steroid Hormone Contraception. Contraception 57:315, 1998

24. Kaunitz AM: Reappearance of the intrauterine device: a "user-friendly" contraceptive. Int J Fertil Womens Med 42:120, 1997

25. Walsh T, Grimes D, Frezieres R, et al: Randomised controlled trial of prophylactic antibiotics before insertion of intrauterine devices. IUD Study Group. Lancet 351:1005, 1998

26. Gentile GP, Kaufinan SC, Helbig DW: Is there any evidence for a post-tubal sterilization syndrome? Fertil Steril 69:179, 1998

27. Glasier A: Emergency postcoital contraception. N Engl J Med 337:1058, 1997

28. Randomised controlled trial of levonorgestrel versus the Yuzpe regimen of combined oral contraceptives for emergency contraception. Task Force on Postovulatory Methods of Fertility Regulation. Lancet 352:428, 1998

29. Glasier A, Baird D: The effects of self-administering emergency contraception. N Engl J Med 339:1, 1998

30. Grimes DA, Raymond EG, Jones BS: Emergency contraception over-the-counter: the medical and legal imperatives. Obstet Gynecol 98:151, 2001

31. Kaunitz AM: Oral contraceptive use in perimenopause. Am J Obstet Gynecol 185 (2 suppl):S32, 2001

32. Hachad H, Ragueneau-Majlessi I, Levy RH: New antiepileptic drugs: review on drug interactions. Ther Drug Monit 24:91, 2002

VII INFERTILITY

Alan H. DeCherney, m.d.

Infertility is defined as the inability to conceive after 1 or more years of regular coital activity without contraception. Fecundity is the statistical probability of achieving a pregnancy (resulting in a live birth) within 1 month (one menstrual cycle) of unprotected sexual intercourse. Monthly fecundity for a fertile couple ranges from 20% to 35% [*see Table 1*].[1] Within 1 year, more than 85% of all couples trying to conceive will achieve a pregnancy.

The incidence of infertility has been increasing over the past 3 decades. It is estimated that nearly 10% of all couples in the United States have disorders associated with infertility. As a result, about 2.5 million couples seek advice for treatment each year in the United States.[2,3]

The practitioner needs to approach the infertile couple in a rational and organized manner. After the initial evaluation, all the reproductive factors should be assessed and a treatment plan proposed that addresses the risk,[4] benefit, and cost[5] to the couple [*see Figures 1 and 2*]. Couples who go through this process and have little or no success may experience severe emotional and psychological distress. For this reason, patients should be counseled about the probability of their achieving a pregnancy before embarking on this potentially expensive treatment course, and psychological counseling during treatment may be valuable.[6]

Evaluation and Treatment of the Infertile Couple

During the first encounter, both partners should be interviewed. The physician should take note of each partner's age, duration of infertility, past pregnancies, past surgeries, frequency of coital activity, and problems encountered during intercourse (e.g., dyspareunia, impotence, anorgasmia, and lack of libido). All potential problems that are revealed during the initial examination should be addressed and treated accordingly. The initial evaluation should also cover other conditions that can affect fertility rates in women. Obese and sedentary women are at higher risk for infertility,[7] because obesity promotes anovulation. Cigarette smoking[8] and long-term use of nonsteroidal anti-inflammatory drugs (NSAIDs)[9] can contribute to infertility. These effects are most noticeable at the extremes (i.e., marked obesity, heavy smoking, or high NSAID doses).

age

A woman's reproductive capability decreases with advancing age. After women reach 30 years of age, pregnancy rates decrease.[10] By 40 years of age, a woman's monthly fecundity is less than 5%.[11] Poor oocyte quality is the predominant factor responsible for this decline in pregnancy rate.

Male fertility also diminishes with age. Sperm quality declines, and the frequency of ejaculation decreases as men grow older. Nearly 50% of the infertility experienced by couples older than 40 years is caused by problems associated with advancing age.[12] Therefore, age is an independent risk factor that is instrumental in determining a couple's chance of achieving a successful pregnancy.

The duration of infertility may be just as important a risk factor as the age of the couple. The longer the duration of infertility, the lower is the probability of achieving a successful pregnancy.

male infertility

Infertility in the male partner is the primary cause of infertility in approximately 35% of couples who seek help.[10] Any previous testicular injury, infection, surgery, radiation, or chemotherapy should be documented during the initial history. The physical examination should focus on penile and testicular anomalies (e.g., hypospadia, cryptorchidism, and varicoceles). If any such anomaly is present, the patient should be referred to a urologist for evaluation and treatment.

After the initial physical examination, a semen analysis should be performed. The specimen is collected after 48 hours of abstinence from coital activity and evaluated no more than 1 hour after collection.[13] If any parameters are abnormal [*see Table 2*], two additional semen analyses should be performed 2 weeks apart. Persistent abnormalities may warrant urologic evaluation and workup for diabetes mellitus, prolactin elevation, and chromosomal abnormalities.

Other tests that can be performed are the sperm penetration assay and the immunobead-binding assay. These tests can detect abnormalities in sperm penetration and motility, respectively, but may not indicate the true nature of a patient's problem.

If the results of the semen tests are normal and the female partner's evaluation appears normal, a diagnosis of unexplained infertility is appropriate [*see* Unexplained Infertility, *below*].

Idiopathic oligospermia is the most common cause of infertility in men. Although there is no cure for this problem, treatment may include in vitro fertilization (IVF) with microinjection of sperm into egg (intracytoplasmic sperm injection [ICSI]) [*see Table 3*]. In severe cases of infertility, this technique can achieve fertilization rates as high as 65%.[14]

female infertility

Tubal and Pelvic Factors

Nearly 35% of the infertility experienced by couples and 40% of the infertility in women is of pelvic origin. Uterine, tubal, and other pelvic abnormalities are responsible for this type of infertility. The practitioner should elicit information regarding any history of sexually transmitted diseases, pelvic inflammatory disease, appendicitis with rupture, pelvic tuberculosis, or adnexal surgery. Many patients with tubal or pelvic damage have a history that includes a previous diagnosis of endometriosis, ectopic pregnancy, or submucous myomas. Al-

Table 1 Fecundity of Normal Couples over Time

Time (Months)	*Couples Achieving Pregnancy (%)*
1	20–36
3	57
6	72
12	85
24	93

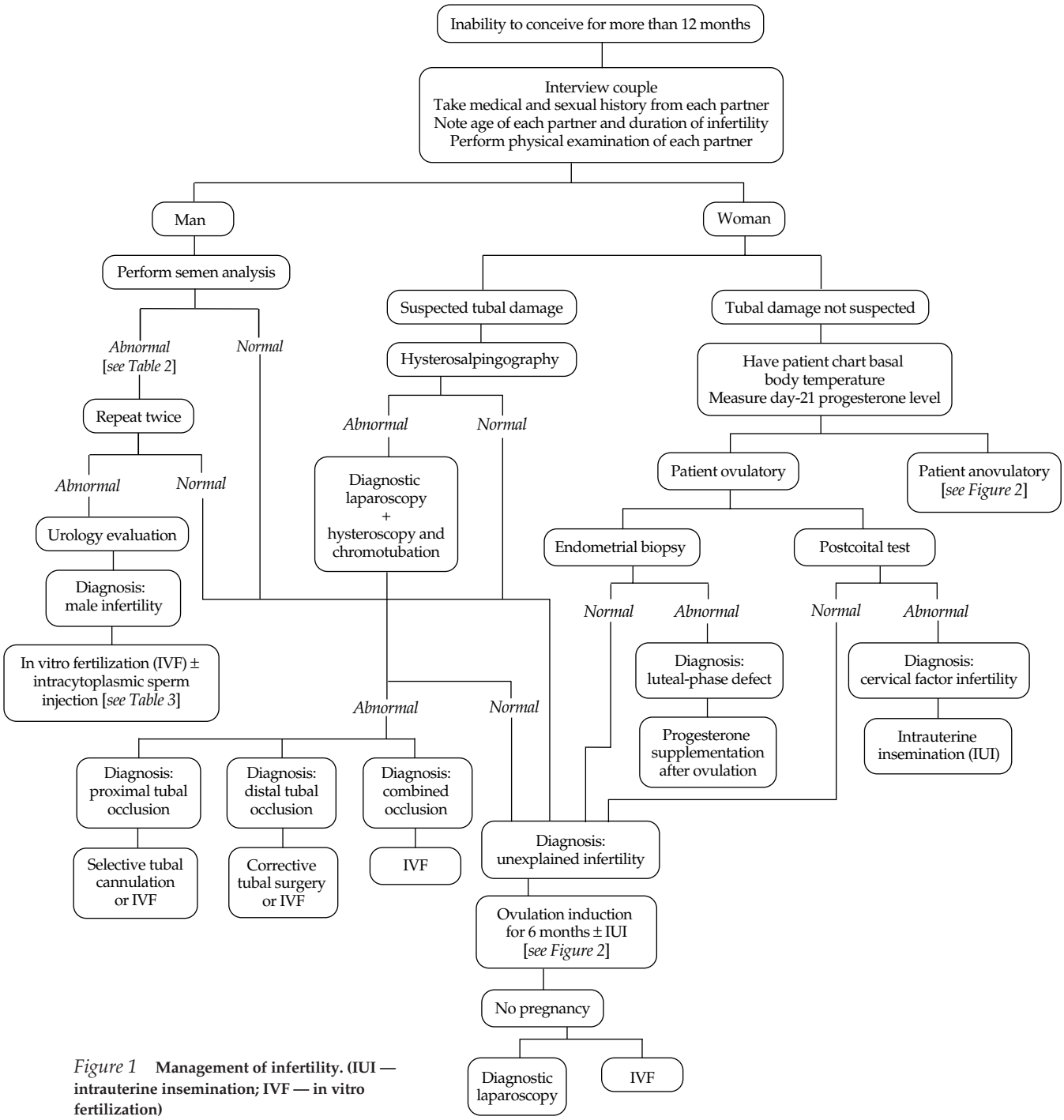


Figure 1 **Management of infertility. (IUI — intrauterine insemination; IVF — in vitro fertilization)**

though uterine myomas seldom cause infertility, they may cause recurrent early pregnancy loss and preterm labor.

Hysterosalpingography (HSG) is one of the initial diagnostic tests used to evaluate uterine, tubal, and pelvic abnormalities. This test is performed during the early proliferative phase, after the cessation of menstrual flow (cycle day 5 to 10). HSG can help identify abnormalities of uterine filling caused by submucous myomas, polyps, uterine synechiae (adhesions), and congenital malformations.

Tubal patency should be evaluated at the time of HSG. A delayed set of radiographs can detect pelvic adhesions and other pelvic abnormalities that prevent the release of contrast material into the pelvis. Once the site of blockage (which may be proximal or distal) has been identified, it can be dealt with accordingly.[15] Patients who are known to be anovulatory may forgo an initial HSG. If ovulation induction is successfully attempted for at least four consecutive cycles and conception does not occur, however, HSG should be performed.

HSG should never be performed on a woman with acute salpingitis, a tender pelvic mass, or allergy to iodine. Women with a known contraindication are better evaluated directly by laparoscopy. Patients who undergo HSG should receive prophylactic treatment for chlamydial infection; oral doxycycline, 100 mg twice daily for 7 days, is effective.

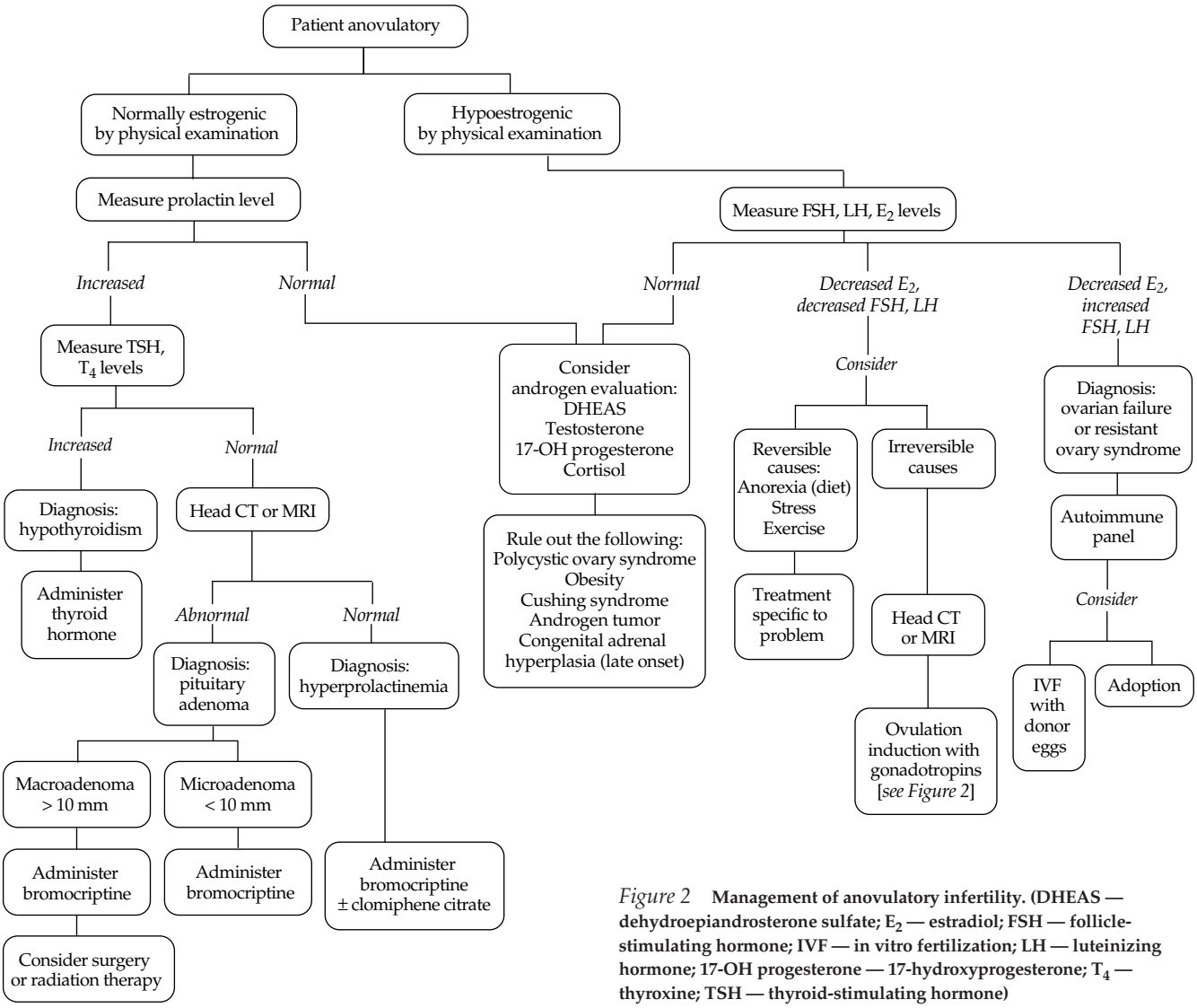


Figure 2 **Management of anovulatory infertility. (DHEAS — dehydroepiandrosterone sulfate; E_2 — estradiol; FSH — follicle-stimulating hormone; IVF — in vitro fertilization; LH — luteinizing hormone; 17-OH progesterone — 17-hydroxyprogesterone; T_4 — thyroxine; TSH — thyroid-stimulating hormone)**

Laparoscopy with chromotubation (intrauterine injection of colored liquid [indigo carmine] to confirm tubal patency) is indicated for patients with abnormal findings on HSG and for patients with unexplained infertility. This procedure may be omitted from the diagnostic workup if IVF is the main focus of treatment. However, surgical resection of the diseased tube should be considered for patients who are diagnosed with a hydrosalpinx on the basis of HSG or laparoscopy. Studies indicate that the mere presence of a hydrosalpinx can adversely affect embryo implantation and the success of IVF.[16]

If the laparoscopy and chromotubation reveal tubal occlusion, surgery may be indicated. Isolated proximal or distal tubal occlusions may be treated by various surgical techniques. However, combined proximal and distal occlusions are not well corrected with surgery, and IVF should be recommended as the treatment of choice to achieve pregnancy in these patients. Age is also important when deciding between tubal surgery and IVF. Older couples should be encouraged to have IVF rather than tubal surgery because the probability of their achieving a pregnancy is higher with IVF.

Cervical Factors

Abnormal cervical mucus is the recognized cause of infertility in 5% to 10% of couples trying to conceive. The postcoital test provides information regarding both the quality of the cervical mucus and its interaction with sperm. This test is performed on cycle day 11 to 13 (24 to 48 hours before ovulation); the male partner must abstain from ejaculation for 48 hours before testing. The cervical mucus is examined 2 to 8 hours after intercourse. The consistency of the cervical mucus is examined and the number of motile sperm per high-power field (hpf) is

Table 2 Semen Analysis Parameters

Parameter	*Normal Value*
Volume of semen	≥ 2.0 ml
pH	7.2–8.0
Sperm concentration	$\geq 20 \times 10^6$ spermatozoa/ml
Total sperm count	$\geq 40 \times 10^6$ spermatozoa/ejaculate
Motility	≥ 50% with forward progression
	≥ 25% with rapid progression
Morphology	≥ 30% with normal forms
Vitality	≥ 75% or more living
White blood cell count	$\leq 1 \times 10^6$/ml

Table 3 In Vitro Fertilization

Step 1	Stimulation of ovulation with injectable gonadotropins.
	Monitoring of follicular development with vaginal ultrasound.
	When mean follicle diameter is ≥ 15 mm, ovulation is triggered with hCG administered intramuscularly.
Step 2	Collection of eggs 34 to 36 hours after hCG injection.
Step 3	Collection of sperm on day of ovum capture or obtain frozen sample.
Step 4	Laboratory (in vitro) incubation of egg(s) with sperm for fertilization and embryo growth.
	If sperm are of poor quality, fertilization is facilitated by microinjection (ICSI) of sperm into egg.
Step 5	Transfer of embryo to uterus 3 to 5 days after oocyte aspiration.
	Administration of progesterone in oil, 50 to 100 mg/day I.M., or vaginal progesterone suppositories.
Step 6	14-day wait for pregnancy or menstruation.
	Measure β-hCG.

hCG—human chorionic gonadotropin ICSI—intracytoplasmic sperm injection

determined. Normal mucus is acellular, clear, thin, and elastic; the mucus should stretch approximately 8 to 10 cm when placed on a glass slide and pulled. This elasticity of the cervical mucus is known as spinnbarkeit. The mucus should also contain at least 5 to 10 progressively motile sperm/hpf.

Cervical mucus that is of poor quality (i.e., thick and nonelastic) will demonstrate a globular rather than a fernlike pattern after drying on a microscope slide. Absent or poor-quality cervical mucus may reflect either inaccurate timing of the test or an abnormality in mucus production. Cervical trauma and infection have been implicated as antecedents to abnormal mucus production.

Sperm that are both shaky and immotile on microscopic inspection are found in the cervical mucus of women who produce antisperm antibodies. When all the sperm from a postcoital test are found to be immotile, the patient should be asked whether lubricants or spermicides were used during coitus.

Intrauterine insemination (IUI) is the treatment of choice for those patients with cervical factor infertility.[17] This procedure bypasses the cervix and allows the physician to place washed sperm directly into the endometrial cavity.[18]

Ovulatory Factors

Ovulatory dysfunction is the cause of 15% of the infertility detected in couples and 40% of the infertility found in women. Anovulation and oligo-ovulation account for most menstrual abnormalities. Shortened menstrual cycles and luteal phase defects are less common causes of ovulatory dysfunction.

A patient's ovulatory status can be determined by several techniques. The cheapest and least invasive technique is to have the patient chart her basal body temperature (BBT). When done correctly, charting can aid the clinician by providing indirect evidence of ovulation. A biphasic temperature curve (i.e., an elevated temperature for at least 11 to 16 days) is an indication that ovulation probably occurred. The patient's own assessment of premenstrual molimina further strengthens the indirect evidence of ovulation.

Measurement of the progesterone level on day 21 of the menstrual cycle also provides an indirect assessment of ovulatory status. This method is less time-consuming than BBT charting. Progesterone values of more than 15 ng/ml are consistent with ovulation. A value of less than 5 ng/ml may indicate that ovulation has not occurred. Because progesterone is secreted in a pulsatile manner, only elevated values of progesterone are diagnostically useful. Levels between 5 and 15 ng/ml probably indicate ovulation but give insufficient information regarding the adequacy of the luteal phase.

Of all menstrual cycles in normally menstruating women, 5% to 30% involve a luteinized unruptured follicle. Although ovulatory symptoms and elevated progesterone levels occur during these cycles, an oocyte is not released and fertilization is impossible. Thus, the predictive value of indirect measures of ovulatory status is limited.

An endometrial biopsy can be performed on cycle days 23 to 26. This test can assess both the ovulatory status of the patient and the adequacy of the luteal phase. A luteal phase defect is defined as a lag in the histologic development of the endometrium by 2 or more days when compared with the cycle day of sampling. This defect in the luteal phase is presumably caused by inadequate progesterone secretion from the corpus luteum.

Treatment has entailed prolonging the luteal phase by administering progesterone either intramuscularly or intravaginally. The benefit of this approach, however, has not been substantiated.

Anovulation

Measurement of the prolactin level should be done during the initial evaluation of patients who are believed to be anovulatory [*see Figure 2*].[19] Elevated prolactin levels have a negative feedback effect on the hypothalamus, preventing the pulsatile release of gonadotropin-releasing hormone (GnRH). This, in turn, prevents secretion of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) from the anterior pituitary. Consequently, follicular development and ovulation do not occur.

Hyperprolactinemia is responsible for 15% of all ovulatory disturbances.[20] If the patient has an elevated prolactin level, with or without galactorrhea, the thyroid-stimulating hormone (TSH) level should be measured to rule out primary or secondary hypothyroidism. If the TSH level is normal, a computed tomography scan or magnetic resonance image of the head should be obtained to determine whether the patient has a prolactinoma. Of import is that prolactin levels may also be elevated as a result of the use of specific medications. Pharmacologic agents that deplete dopamine reserves (i.e., antidepressants, antipsychotics, and other psychotropic agents) may also result in hyperprolactinemia and anovulation.

If the CT or MRI findings are abnormal or reveal a pituitary adenoma, the patient should be treated with oral bromocriptine, starting at a dosage of 2.5 to 5 mg daily, or oral cabergoline, 0.25 mg twice weekly [*see 16:I Amenorrhea*]. These medications should be titrated until prolactin levels return to normal. When prolactin levels are normalized, restoration of ovulatory function should occur.[21] Patients with symptomatic macroadenomas may require ablative therapy with either surgery or radiation if medical therapy does not reduce the size of the tumor or if symptoms associated with the tumor persist or worsen.

Patients with hyperprolactinemia and oligomenorrhea (except those with primary and secondary hypothyroidism, who require thyroid hormone replacement) should be treated with bromocriptine only if they are bothered by symptoms (i.e., galactorrhea) or desire fertility. If a patient remains anovulatory despite treatment with bromocriptine, oral clomiphene cit-

rate, starting at a dosage of 50 mg daily for 5 days, can be added as an adjunctive therapy to stimulate ovulation.

If, at the time of initial examination, the patient is determined to be hypoestrogenic (i.e., she has an atrophic vagina and perineum and reports hot flashes and lack of lubrication during coital activity), the clinician should obtain serum levels of FSH, LH, and estradiol (E_2). These values will identify patients with hypogonadotropic hypogonadism and those with ovarian failure. Patients with hypogonadotropic hypogonadism should be evaluated with a GnRH-stimulation test to determine whether the problem is reversible.

Special attention should be given to anovulatory women who have normal levels of estrogen and prolactin and who have signs of hyperandrogenism and virilization. In these patients, measurements should be made of dehydroepiandrosterone sulfate (DHEAS), total testosterone, 17-hydroxyprogesterone, and 8 A.M. free urine cortisol levels. These tests will help to identify patients with polycystic ovary syndrome (PCOS), ovarian and adrenal neoplasms, congenital adrenal hyperplasia, or Cushing syndrome [*see 16:V Polycystic Ovary Syndrome*].

Patients with PCOS that is associated with elevated insulin or glucose levels who wish to conceive may benefit from a combined regimen of oral metformin, 850 mg twice daily, and clomiphene citrate. Studies have shown that women treated with this combination have a higher rate of ovulation than those treated with clomiphene citrate alone.[22] Hirsutism and acne should not be treated medically during ovulatory induction cycles.

Elevated levels of both FSH and E_2 on cycle day 3 signify a decrease in ovarian reserve (i.e., a decrease in the total number of follicles present for maturation and ovulation). The diagnosis of premature ovarian failure is reserved for women who are younger than 40 years and have gonadotropin (FSH and LH) levels in the menopausal range.

Depending on the incipient age of ovarian failure, a complete autoimmune profile and possibly a genetic karyotype should be considered to establish a diagnosis.[23] Women with an autoimmune disorder are at increased risk for developing multiple organ failure and should be screened annually. These women should be counseled to consider IVF with donor eggs or adoption.

UNEXPLAINED INFERTILITY

The incidence of unexplained infertility is estimated to be between 15% and 20%. Couples who do not receive treatment have a monthly fecundity of 3% and a cumulative 3-year pregnancy rate of 60%. However, when a couple has experienced long-standing infertility (> 3 years) and the female partner is older than 35 years, the probability of achieving a pregnancy is markedly reduced.[24]

The treatment for couples with unexplained infertility includes inducement of superovulation with either clomiphene citrate or gonadotropins[25,26] and IUI or one of the assisted reproductive technologies (e.g., IVF, gamete intrafallopian transfer, and zygote intrafallopian transfer).[27] IUI with ovulation induction using gonadotropins produces higher pregnancy rates for couples with male-factor or unexplained infertility than does either procedure performed alone.[17]

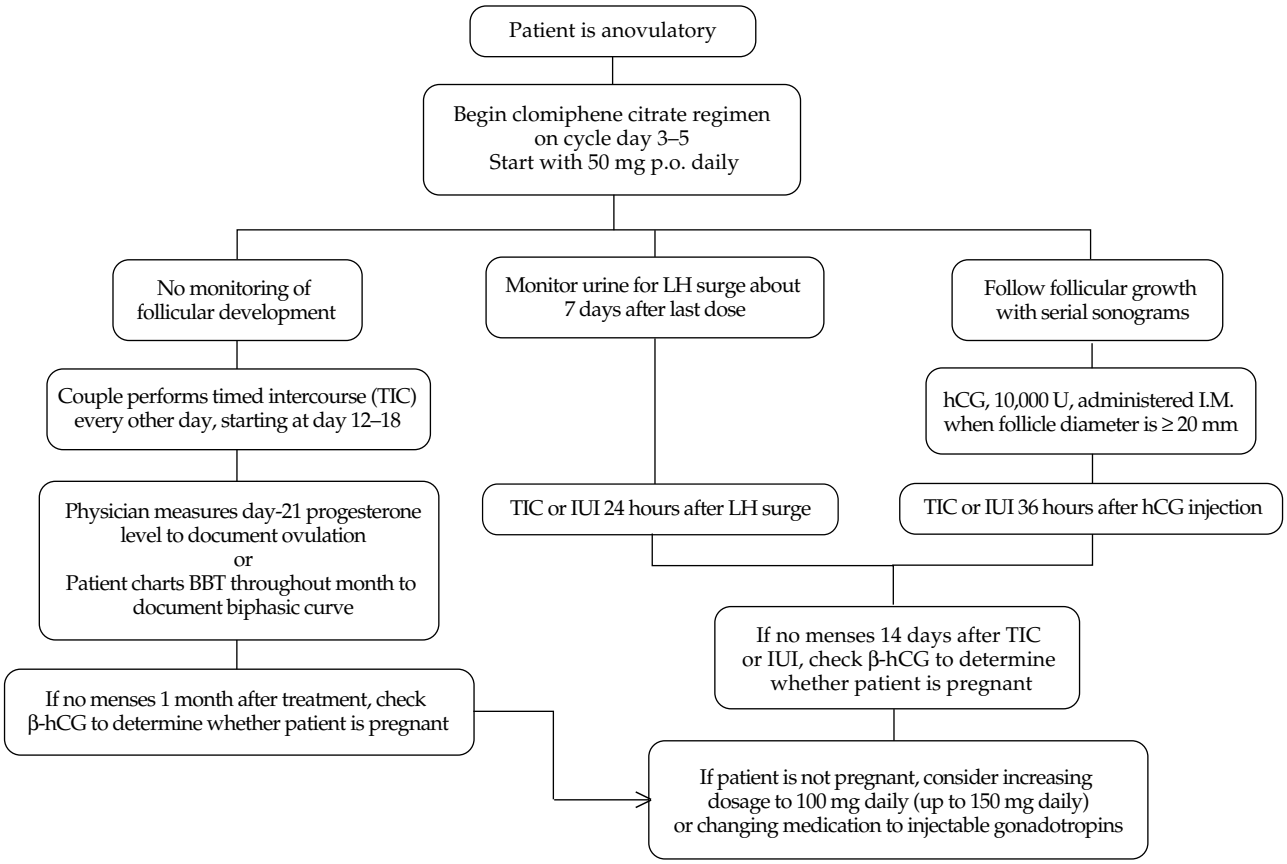


Figure 3 **Ovulation induction with clomiphene citrate. (BBT — basal body temperature; hCG — human chorionic gonadotropin; IUI — intrauterine insemination; TIC — timed intercourse)**

Ovulation Induction with Clomiphene Citrate

To induce ovulation in a woman who has been determined to be anovulatory, clomiphene citrate, 50 mg orally daily, is begun on cycle day 3 to 5 and is continued for a total of 5 days [*see Figure 3*]. The couple and the physician must decide whether to add IUI to the ovulation induction regimen or have the couple perform timed sexual intercourse.

The couple and physician must also decide whether to monitor follicular development and timing for intercourse or IUI and, if so, whether to use a low-, moderate-, or high-technology method for monitoring. Low-technology monitoring entails charting the BBT. A moderate level of monitoring by the patient can be achieved with an LH kit. The kit enables urinary detection of the LH surge, which usually occurs about 7 days after the last dose of clomiphene citrate. High-technology monitoring entails serial vaginal sonography, with administration of human chorionic gonadotropin (hCG) to trigger ovulation when appropriate; hCG is given in a dose of 10,000 units intramuscularly when the follicle diameter is at least 20 mm.

If the clinician is not monitoring follicular development, the couple is directed to perform timed intercourse every other day, starting on cycle days 12 to 18. Timed intercourse or IUI should begin 24 hours after urinary detection of the LH surge or 36 hours after the administration of hCG. One study has shown that in couples with anovulatory, male, or unexplained infertility, clinical pregnancy rates with IUI and clomiphene citrate did not depend on the method used to establish the timing for IUI.[28] Thus, if cost is a consideration for the couple, urinary LH testing may lower the expense by reducing the number of patient visits and eliminating the midcycle ultrasound.

If menses does not begin 14 days after timed intercourse or IUI, the hCG level should be checked to determine whether the patient is pregnant. If the patient is not pregnant but did ovulate, as evidenced by an elevation in the progesterone level on day 21 or a biphasic BBT chart, she should undergo another stimulation cycle with the same dosage of clomiphene citrate. This method can be repeated for as long as 6 months. If pregnancy is not achieved by that time, the use of injectable gonadotropins to stimulate ovulation should be considered [*see 16:V Polycystic Ovary Syndrome*].

It is important for the clinician to realize that the incidence of multiple gestations (e.g., twins) is nearly 8% in patients taking clomiphene citrate and as high as 35% in patients who use injectable gonadotropins.[17] Therefore, extreme caution and judgment should be exercised with these medications.

For patients who do not ovulate with the initial 50 mg/day dosage of clomiphene citrate (as determined on the basis of a low day-21 progesterone level or a monophasic BBT chart), the dose can be increased to 100 or 150 mg daily. If ovulation still does not occur with the higher dosage, ovulation induction with gonadotropins should be employed.[29]

References

1. Guttmacher AF: Factors affecting normal expectancy of conception. JAMA 161:855, 1956

2. Chandra A, Mosher WD: The demography of infertility and the use of medical care for infertility. Study Design and Statistics for Infertility Research 64:781, 1994

3. Stephen EH, Chandra A: Use of infertility services in the United States: 1995. Fam Plann Perspect 32:132, 2000

4. Schover LR, Thomas AJ, Falcone T: Attitudes about genetic risk of couples undergoing in-vitro fertilization. Hum Reprod 13:862, 1998

5. Van Voohis BJ, Stovall DW, Allen BD, et al: Cost-effective treatment of the infertile couple. Fertil Steril 70:995, 1998

6. Kainz K: The role of the psychologist in the evaluation and treatment of infertility. Womens Health Issues 11:481, 2001

7. Rich-Edwards JW, Spiegelman D, Garland M, et al: Physical activity, body mass index, and ovulatory disorder infertility. Epidemiology 13:184, 2002

8. Delayed conception and active and passive smoking. The Avon Longitudinal Study of Pregnancy and Childhood Study Team. Fertil Steril 74:725, 2000

9. Mendonca LL, Khamashta MA, Nelson-Piercy C, et al: Non-steroidal anti-inflammatory drugs as a possible cause for reversible infertility. Rheumatology (Oxford) 39:880, 2000

10. Collins JA, Crosignani PG: Unexplained infertility: a review of diagnosis, prognosis, treatment efficacy and management. Int J Gynaecol Obstet 39:267, 1992

11. Menken J, Trussel J, Larsen U: Age and infertility. Science 233:1389, 1986

12. van Kooij RJ, Looman CW, Habbema JD, et al: Age-dependent decrease in embryo implantation rate after in vitro fertilization. Fertil Steril 66:769, 1996

13. World Health Organization: WHO Laboratory Manual for the Examination of Human Semen and Semen-Cervical Mucus Interaction, 3rd ed. Cambridge University Press, New York, 1992

14. Hlinka D, Herman M, Vesela J, et al: A modified method of intracytoplasmic sperm injection without the use of polyvinylpyrrolidone. Hum Reprod 13:1922, 1998

15. Penzias AS, DeCherney AH: Is there ever a role for tubal surgery? Am J Obstet Gynecol 174:1218, 1996

16. Murray DL, Sagoskin AW, Widra EA, et al: The adverse effect of hydrosalpinges on in vitro fertilization pregnancy rates and the benefit of surgical correction. Fertil Steril 69:41, 1998

17. Guzick DS, Carson SA, Coutifaris C, et al: Efficacy of superovulation and intrauterine insemination in the treatment of infertility. National Cooperative Reproductive Medicine Network. N Engl J Med 340:177, 1999

18. Keck C, Gerber-Schafer C, Wilhelm C, et al: Intrauterine insemination for treatment of male infertility. Int J Androl 20(suppl 3):55, 1997

19. Cunnah D, Besser M: Management of prolactinomas. Clin Endocrinol 34:231, 1991

20. Colao A, Annunziato L, Lombardi G: Treatment of prolactinomas. Ann Med 30:452, 1998

21. Shimon I, Melmed S: Management of pituitary tumors. Ann Intern Med 129:472, 1998

22. Nestler JE, Jakubowicz DJ, Evans WS, et al: Effects of metformin on spontaneous and clomiphene-induced ovulation in the polycystic ovary syndrome. N Engl J Med 338:1876, 1998

23. Conway GS, Kaltsas G, Patel A, et al: Characterization of idiopathic premature ovarian failure. Fertil Steril 65:337, 1996

24. Guzick DS, Sullivan MW, Adamson GD, et al: Efficacy of treatment for unexplained infertility. Fertil Steril 70:207, 1998

25. Kousta E, White DM, Franks S: Modern use of clomiphene citrate in induction of ovulation. Hum Reprod Update 3:359, 1997

26. Rust LA, Isreal R, Mishell DR Jr: An individualized graduated therapeutic regimen of clomiphene citrate. Am J Obstet Gynecol 120:785, 1974

27. Dodson WC, Haney AF: Controlled ovarian hyperstimulation and intrauterine insemination for treatment of infertility. Fertil Steril 55:457, 1991

28. Deaton JL, Clark RR, Pittaway DE: Clomiphene citrate ovulation induction in combination with a timed intrauterine insemination: the value of urinary luteinizing hormone versus human chorionic gonadotropin timing. Fertil Steril 68:43, 1997

29. ACOG Practice Bulletin. Clinical management guidelines for obstetrician-gynecologists number 34, February 2002. Management of infertility caused by ovulatory dysfunction. American College of Obstetricians and Gynecologists. ACOG Committee on Practice Bulletins–Gynecology. Obstet Gynecol 99:347, 2002

VIII ECTOPIC PREGNANCY AND SPONTANEOUS ABORTION

Alan H. DeCherney, m.d.

Ectopic Pregnancy

Ectopic pregnancy is the implantation of an embryo outside the endometrial cavity (i.e., lining of the uterus). The embryo may be implanted in the fallopian tubes, ovaries, abdomen, or cervix. More than 95% of ectopic pregnancies occur in the fallopian tubes, with nearly 80% of ectopic pregnancies occurring in the ampullary portion of the fallopian tube.

Ectopic pregnancies comprise approximately 2% of all reported pregnancies in the United States. If left untreated, ectopic pregnancy can result in rupture of the fallopian tube, which can lead to hemorrhagic shock and death. For that reason, ectopic pregnancy is a major cause of morbidity and mortality in women of reproductive age. It is the third leading cause of maternal mortality and is responsible for nearly 9% of all pregnancy-related deaths.[1] Between 1970 and 1992, the incidence of ectopic pregnancy increased sixfold, which is consistent with the increased prevalence of important risk factors for ectopic pregnancy (see below).[2]

Improvements in diagnostic skills have enabled physicians to treat ectopic pregnancies at an earlier gestational age and by more conservative approaches. Therefore, it is imperative that the primary care physician be able to diagnose and intervene as early as possible to reduce the incidence of irreversible tubal damage and the risk of future infertility [*see Figure 1*].[3]

DIAGNOSIS

Clinical Manifestations

The principal diagnostic task in ectopic pregnancy is distinguishing it from intrauterine pregnancy or threatened abortion. The symptoms of an ectopic pregnancy [*see Table 1*] vary with its location and the rate of its growth. Nearly all women with ectopic pregnancies complain of a colicky abdominal pain that is vague in location. As the pregnancy grows, capillaries are broken and blood spills into the abdominal cavity. The blood that fills the intraperitoneal cavity causes irritation of the left hemidiaphragm, resulting in left shoulder pain; the liver occupies the space directly under the right hemidiaphragm and prevents the blood from reaching the right hemidiaphragm. Almost one quarter of the patients with ruptured ectopic pregnancies have left shoulder pain.

Amenorrhea, the second most common symptom, occurs in more than 75% of patients. The duration of amenorrhea depends on the site of implantation and usually lasts 6 to 8 weeks before the onset of other symptoms.

Vaginal bleeding is also common and may occur a few days to weeks before the patient's initial visit. The pattern of bleeding is most often described as spotting and usually is preceded by the onset and worsening of abdominal pain.

Other symptoms, which occur less frequently, are dizziness, fainting, nausea, vomiting, other signs of pregnancy, the urge to defecate, and passage of tissue through the vagina.

History and Physical Examination

At the initial visit, the patient's medical history should be taken. Patients in whom ectopic pregnancy is suspected often have a history of infertility, pelvic inflammatory disease, endometriosis, or tubal damage [*see Table 2*]. In rare cases, ectopic pregnancy can occur in patients who have undergone tubal sterilization, even many years after the procedure.[4] If the diagnosis of ectopic pregnancy is suspected, a pregnancy test should be performed (e.g., measurement of the human chorionic gonadotropin [hCG] level in urine). The urine hCG test is extremely sensitive; false negative results are very rare.

Patients who have a negative urine hCG test result should be evaluated for other gynecologic problems, including ovarian torsion, a ruptured ovarian cyst, and pelvic inflammatory disease. Gastrointestinal disorders and possible surgical conditions (e.g., appendicitis) should also be investigated.

If the urine hCG test result is positive, a physical examination should be performed and a pelvic sonogram obtained.[5] The physician should assess the degree of abdominal and pelvic pain experienced by the patient and try to elicit signs of peritonitis, which could indicate rupture of an ectopic pregnancy.

Inspection of the patient's cervix with a speculum can help distinguish ectopic pregnancy from spontaneous or threatened abortion. If the cervical os is open, fetal tissue should be observed at the internal cervical os; the diagnosis of an inevitable abortion or incomplete abortion should be made, and a dilatation and curettage (D&C) should be performed. If the cervical os is closed, the examiner should determine both the amount of blood present in the vagina and the amount emanating from the external cervical os. Most women with ectopic pregnancies do not experience heavy vaginal bleeding and have only a light bloody vaginal discharge. In the case of threatened or complete abortion, the patient should be instructed to have pelvic rest and bed rest until the symptoms resolve.

Laboratory Testing

A pelvic sonogram should be obtained to document whether an intrauterine pregnancy (IUP) or extrauterine pregnancy (EUP) exists. The sonographer should focus attention on the uterus, the adnexa, and the cul-de-sac. The endometrium should be inspected, and the presence of a gestational sac and fetal pole should be verified. Attempts should be made to observe the fetal heartbeat to help distinguish normal from abnormal pregnancies. If nothing is present in the uterus, the adnexa should be inspected. If a gestational sac and fetal pole are identified outside the uterus, the diagnosis of an EUP should be made. The ectopic pregnancy can then be treated by medical or surgical means [*see* Treatment, *below*].

If no IUP or EUP can be documented on sonography, further laboratory evaluation should be pursued. The evaluation should include measurement of the serum β-hCG level, a complete blood count (CBC), and measurement of the prothrombin time (PT) or partial thromboplastin time (PTT). A progesterone assay is optional (see below). A β-hCG value of 1,500 mIU/ml occurs

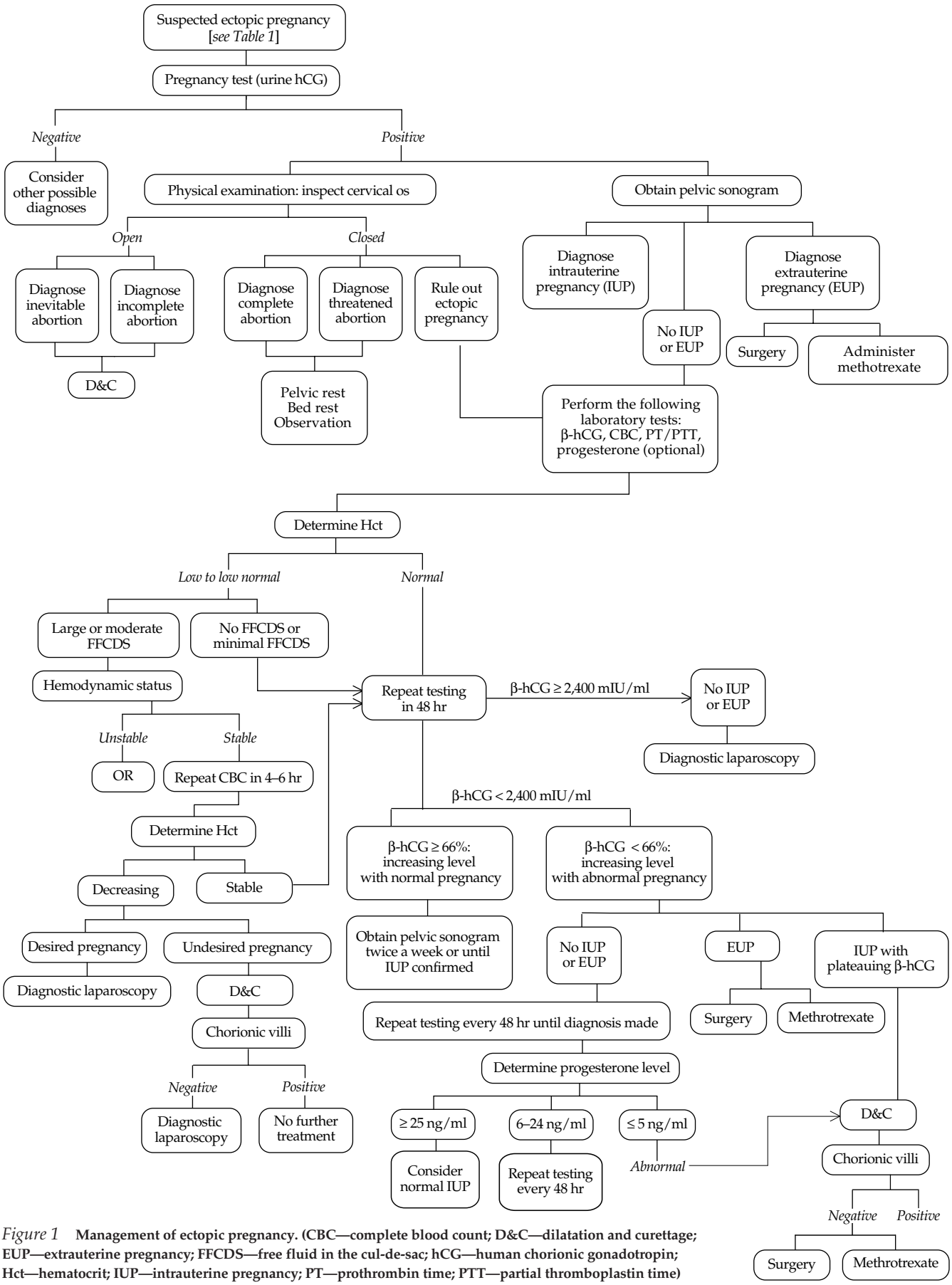


Figure 1 **Management of ectopic pregnancy. (CBC—complete blood count; D&C—dilatation and curettage; EUP—extrauterine pregnancy; FFCDS—free fluid in the cul-de-sac; hCG—human chorionic gonadotropin; Hct—hematocrit; IUP—intrauterine pregnancy; PT—prothrombin time; PTT—partial thromboplastin time)**

Table 1 Signs and Symptoms of Ectopic Pregnancy

Abdominal pain
Amenorrhea
Vaginal bleeding
Dizziness and fainting
Other symptoms of pregnancy

Table 2 Patient History Consistent with Suspected Ectopic Pregnancy

Prior ectopic pregnancy
History of infertility
Past infection with *Chlamydia* or *Neisseria gonorrhoeae* or history of pelvic inflammatory disease
Past or present use of an intrauterine device (IUD)
Current pregnancy conceived with in vitro fertilization (IVF)
History of endometriosis
Current pregnancy conceived while taking oral contraceptives

around the time a normal IUP first becomes visible on pelvic sonography. If no IUP is observed and the β-hCG value is 2,400 mIU/ml or higher, an ectopic pregnancy should be suspected.

A baseline hematocrit, along with sonographic evaluation of the patient's cul-de-sac, will provide insight for prognosis and possible treatment options. Free fluid in the cul-de-sac may be blood; this finding increases the likelihood of ectopic pregnancy.[6] Patients with aborting ectopic and early nonruptured ectopic pregnancies may present with blood in the cul-de-sac. If the patient has a normal hematocrit and minimal to no free fluid in the cul-de-sac, repeating pelvic sonography and measurements of the β-hCG level and a CBC in 48 hours is recommended. Patients with low to low-normal hematocrits in conjunction with mild to moderate amounts of free fluid in the cul de sac need further evaluation, including a repeat CBC.

If a moderate to large amount of free fluid is found in the cul-de-sac, blood pressure needs to be checked immediately. If the blood pressure is unstable, the patient should be taken to the operating room for an exploratory laparotomy and transfusion with packed red blood cells and crystalloids to replenish intravascular losses.

Some patients require hospitalization for confirmation of hemodynamic stability. The CBC should be repeated within 4 to 6 hours in patients who are hemodynamically stable. If the hematocrit remains stable, the laboratory tests and sonographic evaluation should be repeated within 48 hours (see below).

Patients with decreasing hematocrits should be evaluated with either a D&C or diagnostic laparoscopy, depending on their desire to maintain the pregnancy. When a D&C is performed for removal of fetal tissue, histology of the removed tissue should show chorionic villi. If no villi are obtained, an ectopic pregnancy should be suspected.

Repeat hCG levels The doubling time of the β-hCG level in early pregnancies ranges from 48 to 72 hours. A rise in the β-hCG level of at least 66% in 2 days is generally indicative of a normal IUP. Patients who have normal doubling values of their β-hCG level on repeat evaluation should be followed up in 1 week with a repeat sonogram to confirm a pregnancy in utero.[7]

An abnormally rising β-hCG level (< 66% higher than original values) should be further investigated. Correlation with a repeat sonogram and hematocrit will help guide the clinician to the correct diagnosis and treatment. Although the presence of an ectopic pregnancy should be suspected, an abnormally developing IUP cannot be ruled out. Treatment should be decided not on the basis of only two β-hCG values but, rather, on the entire clinical picture. An abnormally rising β-hCG level that is not substantiated by other laboratory or radiographic evidence should not be treated as an ectopic pregnancy. Surgical or medical treatment of these patients should be considered only when the diagnosis is confirmed and an EUP is documented.

Serum progesterone level Progesterone values can help the clinician determine the viability of a pregnancy, but only in rare instances can they reveal an ectopic pregnancy.[8] A progesterone level of greater than 25 ng/ml is associated with a normal IUP in nearly 97% of cases. Values of less than 5 ng/ml are associated with abnormal pregnancies, and values between 5 and 25 ng/ml are indeterminate. The usefulness of the progesterone assay is limited because more than 85% of the values obtained are between 5 and 25 ng/ml. Furthermore, most centers are unable to process this test in a timely fashion, with results being unavailable for review and interpretation on the same day that the sample is drawn.

TREATMENT

Ectopic pregnancy can be treated medically or surgically. The choice of treatment should be tailored to the patient's clinical circumstances and preferences.[9,10]

Medical Therapy

Methotrexate should be considered as the primary modality of treatment in all patients who meet the criteria for medical therapy [*see Table 3*].[11,12] Methotrexate is an antimetabolite that interferes with the conversion of dihydrofolic acid to tetrahydrofolic acid, inhibiting DNA synthesis and cell division and thereby terminating the pregnancy. Because methotrexate may, in rare cases, produce hepatotoxicity or bone marrow suppression, baseline liver function tests (LFTs) and a CBC, along with a β-hCG level, must

Table 3 Criteria and Contraindications for Methotrexate Treatment

Criteria
Patient is diagnosed with ectopic pregnancy
Patient is reliable and expected to comply with regimen
Patient is hemodynamically stable
The ectopic pregnancy is no greater than 3.5 to 4 cm in diameter
Contraindications
Absolute
Hepatic dysfunction
Moderate to severe anemia
Relative
Human chorionic gonadotropin concentration ≥ 10,000 mIU/ml
Fetal cardiac activity detected with sonography

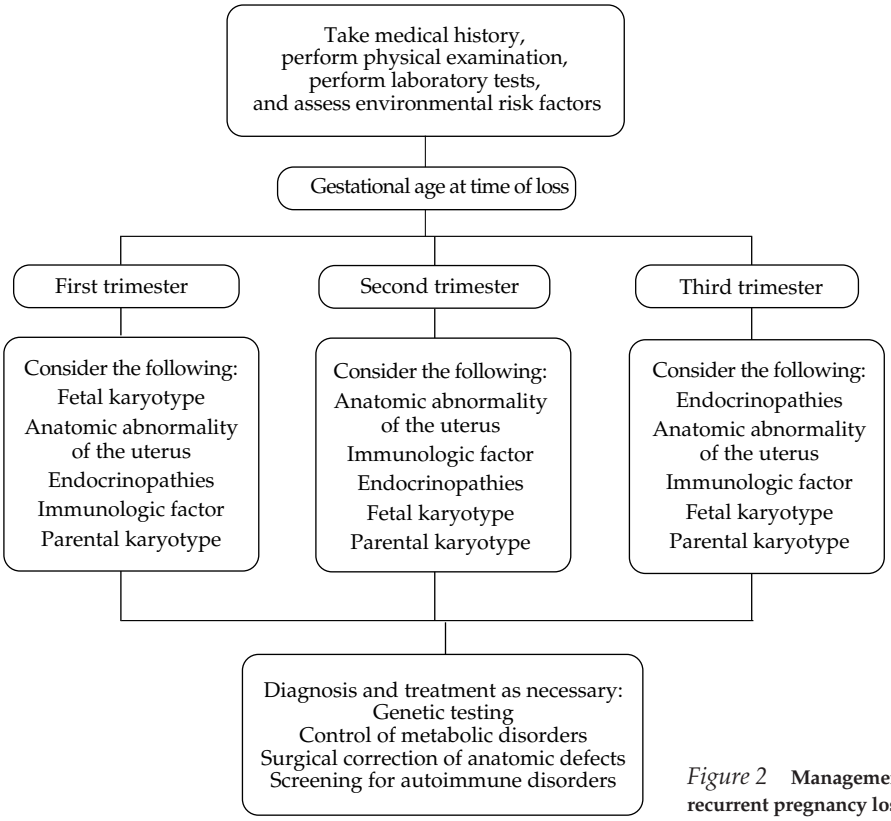


Figure 2 **Management of spontaneous abortion and recurrent pregnancy loss.**

be done in all patients being considered for methotrexate therapy. LFTs and CBCs must be monitored during treatment with repeated doses of methotrexate, and methotrexate should be discontinued if test results become abnormal.

Methotrexate therapy can be given by local injection into the area of the ectopic pregnancy, which requires ultrasound or laparoscopic guidance, or by intramuscular injection as a single-dose or multiple-dose regimen. The multiple-dose regimen is rarely used today, unless the patient fails to respond to the single-dose regimen. The multiple-dose regimen includes citrovorum rescue to protect maternal cells. In the single-dose regimen, 50 mg/m^2 of methotrexate is given and baseline studies are repeated on day 3 or 4. If the β-hCG level decreases by less than 15%, the treatment can be repeated in 1 week. Of the patients who receive methotrexate, 64% are cured with a single dose.[13] An additional 14% require two or three doses. This gives an overall cure rate of 78%. Patients who are cured with methotrexate have fertility rates equivalent to those of patients who are treated surgically. Unfortunately, 20% of the patients whose ectopic pregnancy is treated pharmacologically will ultimately require surgical intervention.

Surgical Therapy

Surgical modalities for ectopic pregnancy include laparoscopy and laparotomy [*see Figure 1*]. Laparoscopy with conservative tubal therapy (salpingostomy or salpingotomy) is the preferred surgical method in hemodynamically stable patients, for the following reasons: (1) Less intraoperative blood loss occurs, (2) less postoperative analgesia is required, (3) the hospital stay is shorter, and (4) the cost savings per patient is greater. After laparoscopy, β-hCG levels should be measured serially until they decrease to nonpregnant levels. If the β-hCG level plateaus, increases, or does not decrease more than 15% in 48 hours, treatment with methotrexate or salpingectomy is required.

In the past, laparotomy and radical tubal surgery (salpingectomy) was recommended for hemodynamically unstable patients. Because of improvements in anesthesia and cardiovascular monitoring, together with advances in laparoscopic surgical skills and experience, operative laparoscopy can now be justified for surgical treatment of ectopic pregnancy even in women with hemodynamic instability.[14] Serial measurement of β-hCG levels is not required after salpingectomy.

Fertility after surgery is not a function of the surgical method employed. Rather, future fertility in these cases depends on three patient factors. The first is a history of infertility; patients with prior infertility have a fourfold lower pregnancy rate than patients without. The second is the status of the contralateral fallopian tube. The third is the extent of adhesions involving the ipsilateral tube.

Spontaneous Abortions

A spontaneous abortion is defined as the spontaneous termination of a pregnancy before 20 weeks' gestation (from the onset of the last menstrual period) or the loss of a pregnancy with a fetal weight of less than 500 g. All pregnancy losses that occur later than 20 weeks' gestational age are termed miscarriages.

Spontaneous abortions occur in about 15% to 20% of all known pregnancies.[15] It has been estimated that more than 50% of all conceptions end in spontaneous abortion. This rate is higher than previous estimates, because spontaneous abortions often

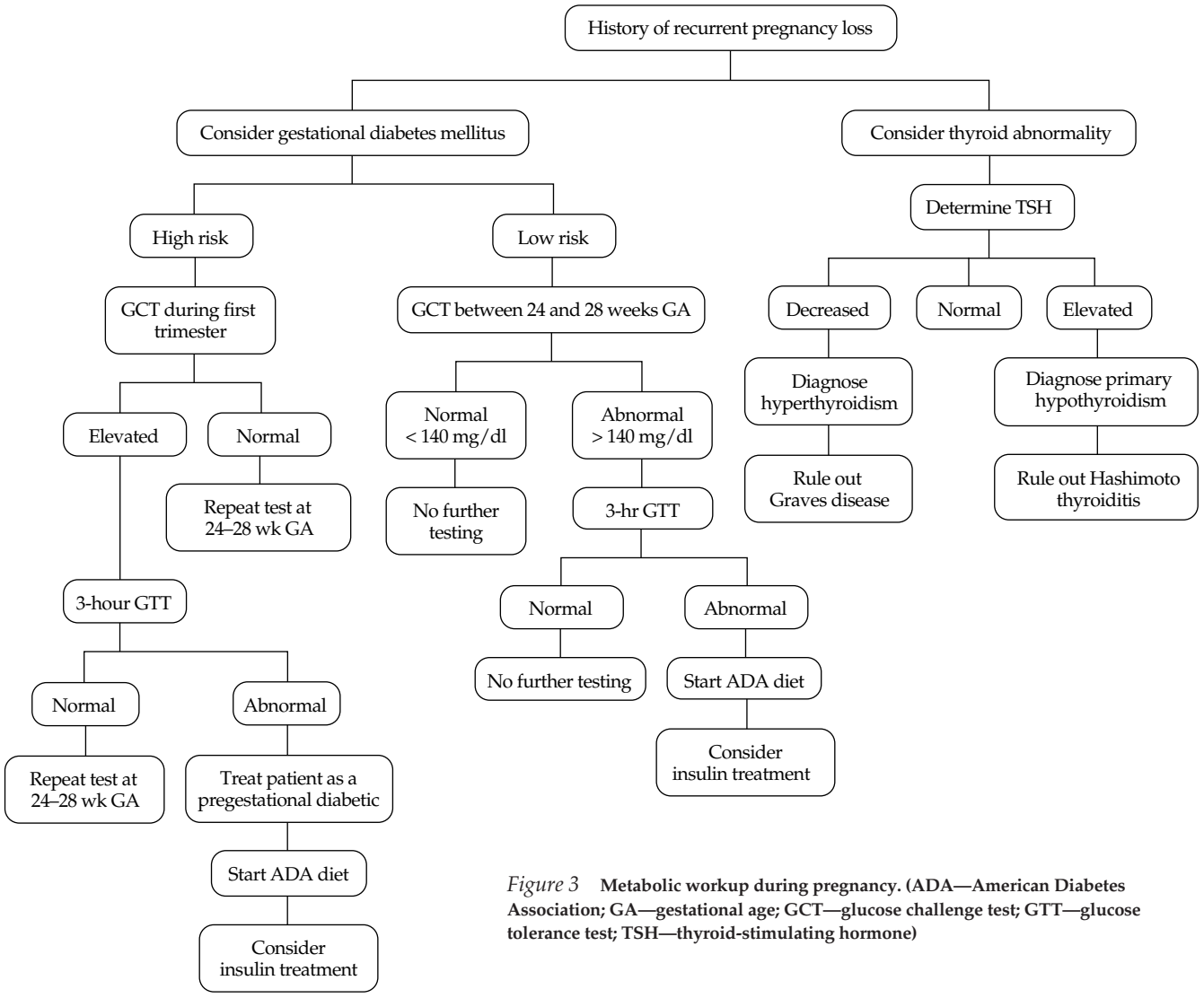


Figure 3 **Metabolic workup during pregnancy. (ADA—American Diabetes Association; GA—gestational age; GCT—glucose challenge test; GTT—glucose tolerance test; TSH—thyroid-stimulating hormone)**

occur around the time of expected menses and are not recognized as abortions.

Traditionally, women who had three or more consecutive pregnancy losses were designated habitual aborters, and it was considered likely that they would continue to have frequent spontaneous abortions. It is now known, however, that in women who have two or more consecutive losses and no history of a live birth, the maximum likelihood that a subsequent pregnancy will result in spontaneous abortion is only about 45%. Women who have at least one live birth before three or more losses have an abortion frequency of less than 30%.[16]

Factors responsible for pregnancy termination vary from one trimester to the next. A strong correlation exists between the gestational age at which the loss occurs and its cause. Therefore, it is important to recognize these potential risks and begin treatment when necessary [*see Figure 2*].

DETERMINING CAUSE OF SPONTANEOUS ABORTION

When a spontaneous abortion has occurred, the physician must investigate the possible causes. A medical history should be taken, possible environmental risk factors should be assessed, and laboratory tests for common infections should be performed.

Depending on whether the abortion occurred during the first, second, or third trimester, the likelihood of genetic, anatomic, or endocrinologic causes varies (see below).

Infections

Organisms that can cause spontaneous abortions include *Ureaplasma, Chlamydia, Listeria,* mycoplasmas, and the TORCH (toxoplasmosis, other [e.g., hepatitis B] rubella, cytomegalovirus, herpes) organisms. Diagnosis is made by cultures of blood or amniotic fluid. Treatments may include antibiotics, γ-globulins, or vaccinations.

Environmental Factors

Cigarette smoking, alcohol consumption, and caffeine consumption can cause spontaneous abortion. The greater the number of cigarettes smoked in a day, the greater the risk of fetal loss. Cleaning solvents and anesthetic gases have been linked to fetal wastage. It is important that pregnant women take precautions and limit their exposure to these agents.

Genetic Factors

Chromosomal abnormalities are responsible for more than 70% of all first-trimester abortions.[17] Aneuploidy problems pre-

dominate, with trisomies as a whole constituting the largest group (nearly 50%). Monosomy X (Turner syndrome) is the single most common chromosomal abnormality, accounting for nearly 25% of spontaneous abortions. Maternal nondisjunction during metaphase I is the most common cause of trisomies. It is estimated that only 30% of second-trimester losses and 3% of third-trimester losses are the result of chromosomal defects.

Balanced translocations are the most commonly transmitted parental chromosomal abnormalities responsible for recurrent losses. Thus, genetic testing should be done in couples who have experienced recurrent spontaneous abortion, to determine whether either partner is a carrier. Genetic testing is often done on spontaneously aborted fetuses.

Endocrine Factors

Patients with uncontrolled diabetes mellitus or thyroid disease have an increased risk of fetal demise and spontaneous abortion. Workup for an endocrine disorder should be considered for all patients with recurrent pregnancy losses. Any history or physical examination that is suspect for a metabolic problem should prompt the physician to investigate immediately.

Second- and third-trimester losses are more often the result of a maternal endocrine abnormality. Recognition and control of the metabolic disturbance are the goals of treatment [*see Figure 3*]. Initial screening begins with a 1-hour glucose challenge test (GCT). This test should be performed in all pregnant women between 24 and 28 weeks' gestational age. A 3-hour glucose tolerance test (GTT) is performed only in women with abnormal GCT values. Evaluation for a thyroid abnormality should be performed at the first prenatal visit. Determining the thyroid-stimulating hormone (TSH) level will distinguish women who have a suspected thyroid dysfunction from those who are euthyroid.[18,19]

Luteal-phase defect is another factor thought to be responsible for early pregnancy loss. Women with this problem characteristically have abnormally shortened secretory phases (< 10 days) or an endometrial lining that is not in phase with the presumed day of the menstrual cycle. An endometrial biopsy that is 2 or more days out of phase with the menstrual cycle is considered diagnostic. Treatment entails the use of progesterone suppositories (25 mg twice daily) to support the luteal phase. However, the benefit of this treatment regimen has not been substantiated.

Anatomic Abnormalities

Uterine cavity malformations are responsible for nearly 15% of recurrent pregnancy losses. Most of these losses occur in the second and third trimesters. Uterine septa are most commonly associated with recurrent abortions. Other causes are submucous myomas, intrauterine adhesions, congenital abnormalities, and cervical incompetence.

A pelvic sonogram and hysterosalpingogram are useful for establishing an initial diagnosis. A magnetic resonance image of the pelvis or a combined laparoscopy and hysteroscopy may be needed for further examination of the uterus. Treatment entails surgical correction of the anatomic defect.

Immunologic Factors

Autoimmune disorders have also been associated with pregnancy losses.[20] These losses commonly occur in the second and third trimesters. Approximately 30% of patients with recurrent pregnancy losses test positive for antinuclear antibodies. Antiphospholipid antibodies, lupus anticoagulants, anticardiolipin antibodies, anti–SS-A (Ro), and anti–SS-B (La) have also been implicated as risk factors. Diagnosis is made by screening the patient for common autoimmune disorders [*see 15:IV Systemic Lupus Erythematosus*].

References

1. Atrash HK, Friede A, Hogue CJ: Ectopic pregnancy mortality in the United States, 1970-1983. Obstet Gynecol 70:817, 1987

2. Ectopic pregnancy—United States, 1990–1992. MMWR Morb Mortal Wkly Rep 44:46, 1995

3. Stovall TG, Ling FW, Carson SA, et al: Nonsurgical diagnosis and treatment of tubal pregnancy. Fertil Steril 54:537, 1990

4. The risk of ectopic pregnancy after tubal sterilization. U.S. Collaborative Review of Sterilization Working Group. N Engl J Med 336:762, 1997

5. Gracia CR, Barnhart KT: Diagnosing ectopic pregnancy: decision analysis comparing six strategies. Obstet Gynecol 97:464, 2001

6. Dart R, McLean SA, Dart L: Isolated fluid in the cul-de-sac: how well does it predict ectopic pregnancy? Am J Emerg Med 20:1, 2002

7. Kador N, Romero R: Serial human chorionic gonadotropin measurements in ectopic pregnancy. Am J Obstet Gynecol 158:1239, 1988

8. Stovall TG, Ling FW, Cope BJ, et al: Preventing ruptured ectopic pregnancy with a single serum progesterone. Am J Obstet Gynecol 160:1425, 1989

9. Yao M, Tulandi T: Current status of surgical and nonsurgical management of ectopic pregnancy. Fertil Steril 67:421, 1997

10. Stovall T, Ling FW, Buster JE: Outpatient chemotherapy of unruptured ectopic pregnancy. Fertil Steril 51:435, 1989

11. Stovall TG, Ling FW, Gray LA, et al: Methotrexate treatment of unruptured ectopic pregnancy: a report of 100 cases. Obstet Gynecol 77:749, 1991

12. Carson SA, Buster JE: Ectopic pregnancy. N Engl J Med 329:1174, 1993

13. Stovall TG, Ling FW, Cope BJ, et al: Single-dose methotrexate for treatment of ectopic pregnancy. Obstet Gynecol 77:754, 1991

14. Sagiv R, Debby A, Sadan O, et al: Laparoscopic surgery for extrauterine pregnancy in hemodynamically unstable patients. J Am Assoc Gynecol Laparosc 8:529, 2001

15. Wilcox AJ, Weinberg CR, O'Connor JF, et al: Incidence of early loss of pregnancy. N Engl J Med 319:189, 1988

16. Harger JH, Archer DF, Marchese SG, et al: Etiology of recurrent pregnancy losses and outcome of subsequent pregnancies. Obstet Gynecol 62:574, 1983

17. Byrne J, Warburton D, Kline J, et al: Morphology of early fetal deaths and their chromosomal characteristics. Teratology 32:297, 1985

18. Kjos SL, Buchanan TA: Gestational diabetes mellitus. N Engl J Med 342:896, 2000

19. ACOG Practice Bulletin. Thyroid disease in pregnancy. Obstet Gynecol 98(5 pt 1):879, 2001

20. Coulam CB: Immunologic tests in the evaluation of reproductive disorders: a critical review. Am J Obstet Gynecol 167:1844, 1992

IX MEDICAL COMPLICATIONS IN PREGNANCY

Ellen W. Seely, m.d.
Jeffrey Ecker, m.d.

Medical complications and intercurrent disease have long presented challenges to obstetricians and other medical providers caring for pregnant women. Contemporary medical practice and treatments have only added to these challenges. Advances in disease management mean that patients with some conditions (e.g., cystic fibrosis) whose life expectancies in the past would have precluded pregnancy are now living to reproductive age. Cures and treatments for other conditions can restore fertility to patients who previously had limited fecundity (e.g., renal transplants in women with renal insufficiency). In vitro fertilization and other assisted reproductive technologies allow the barrier of age, as well as anatomic and genetic barriers, to be surmounted.

All these advances emphasize the need for careful and considered collaboration between clinicians caring for women of reproductive age who are not pregnant and those who care for them during pregnancy. Contraceptive planning and management is vital for patients with significant medical problems, especially those undergoing treatment that is likely to restore fertility (e.g., prescription of metformin for insulin resistance in women with polycystic ovarian syndrome). Primary care clinicians and obstetricians must together evaluate the risk associated with particular diseases in pregnancy and plan for their management, ideally through prepregnancy consultation.

In this chapter, we first review the basic structure and issues for prepregnancy or early pregnancy consultation. We then outline the principles of teratogenesis necessary to evaluate the safety of using medications and other treatments during pregnancy. Next, we discuss the normal physiologic changes of pregnancy. Finally, we discuss specific diseases and conditions that may predate pregnancy or arise as complications of pregnancy and that often require comanagement by obstetricians and medical specialists.

Pregnancy Planning and Counseling

The prognosis for pregnancy complicated by medical conditions is usually improved by pregnancy planning and counseling. In addition, there are a few conditions in which pregnancy may so significantly complicate underlying medical conditions that pregnancy itself is either ill-advised or undertaken only with the greatest of caution. Such conditions include primary pulmonary hypertension and Eisenmenger syndrome (pulmonary hypertension in association with a left-to-right cardiac shunt), each of which carries an approximately 25% to 50% risk of maternal mortality.[1,2] Marfan syndrome with involvement of the aortic root also carries a significant risk of mortality during pregnancy and the puerperium—25% by some estimations.[3] In women with such conditions who have an unplanned pregnancy or who become pregnant in spite of careful contraceptive management, early pregnancy termination should be considered. However, because maternal hormones and physiology change so early in gestation, even early pregnancy interruption does not eliminate all the maternal morbidity and mortality associated with these conditions.

Consultation should also include an evaluation of the risk that a child will inherit the parent's condition. For many disorders, inheritance is defined by mendelian genetics (e.g., autosomal dominant, autosomal recessive, or X-linked disorders). For other conditions that are felt to result from a mix of genetic and environmental exposures (e.g., congenital heart disease), empirical data on inheritance are available. An increasing number of gene defects have been linked to specific disorders, and for many of these, prenatal diagnosis is available. All such diagnoses are facilitated by screening a couple before pregnancy or early in gestation to determine whether genetic testing is informative (i.e., if there is an identifiable mutation or abnormality that can be used in screening DNA from fetal cells). For some couples, family history or ethnic background may indicate the need for screening for specific disorders. The American College of Obstetricians and Gynecologists recommends that white couples with Northern European background be offered screening for cystic fibrosis.[4] Ashkenazi Jews are offered screening for Tay-Sachs disease, Canavan disease, and cystic fibrosis; in addition, Ashkenazi couples may choose screening for an expanded panel of recessive conditions that are more common in this population. Black couples of African background are offered screening to determine whether they are sickle cell carriers. Any fetal testing is facilitated by completing parental screening in advance of pregnancy so that the prospective parents, practitioners, and genetic counselors can anticipate the need for prenatal diagnosis. Such advance planning is essential if couples are considering preimplantation diagnosis, in which embryos created using in vitro fertilization undergo biopsy, their genetic material is studied, and appropriate embryos are transferred for implantation.

Finally, women or couples seeking advice in advance of pregnancy should be counseled about specific dietary or lifestyle changes that may optimize outcome of future pregnancy. Folic acid supplementation of at least 400 μg daily, starting at least 3 months in advance of conception, is recommended to decrease the likelihood of neural tube defects.[5] Planning for a pregnancy may be an impetus to address smoking and the use of alcohol and drugs of abuse. In women with diabetes mellitus (DM), improved glucose control before conception can reduce the risk of birth defects.

Principles of Teratogenesis

Patients and physicians often worry that exposure to medications or environmental agents may increase the risk of birth defects or pregnancy complications. There are, however, few well-designed, prospective studies that address these questions. Such a deficit, while unfortunate, is understandable, because few pregnant women would choose to enroll in prospective, placebo-controlled trials designed simply to examine drug safety. As a result, most of the studies evaluating teratogenesis are retrospective and observational and, therefore, compromised by all the attendant limitations of this approach. In particular, it can be difficult to

separate the effect of medications from that of the condition being treated (e.g., determining whether growth restriction in fetuses of women taking beta blockers for hypertension is caused by the hypertension or the beta blocker). Recognizing the inherent difficulty in human studies, some investigators have turned to animal models; with regard to teratogenesis, however, it is clear that mice are not men (or women). Thalidomide, for example, was not recognized as a teratogen in laboratory animals but proved to cause dramatic phocomelia in humans. Conversely, exposure to steroids in early pregnancy has a much stronger association with cleft lips in mice pups than in humans.

There are several other important challenges in determining whether a particular drug or exposure is teratogenic. Potential teratogenic effects may be small (e.g., lithium is associated with Ebstein anomaly in only one in 1,000 exposed pregnancies) or distant from the incident exposure (e.g., clear cell carcinoma in daughters whose mothers used diethylstilbestrol [DES] occurs decades after the mother's use). In addition, many diseases require treatment with multiple medications. As limited as the data are on individual agents, however, information on outcomes associated with combinations of medications and their potential synergistic effects is even more limited.

Given these recognized limitations, several principles can guide counseling regarding medication use or environmental exposure in pregnancy. Because organogenesis begins in early pregnancy, the first trimester is a critical period for teratogenesis. Early exposure is most directly linked with birth defects; however, later exposure may have adverse effects as well. Use of warfarin or ethanol during the second and third trimester, for example, is associated with clear fetal consequences, because brain development continues well into the neonatal period. Good practice, therefore, recommends that throughout pregnancy, physicians should use the fewest number of medications and the lowest doses appropriate to treat the underlying symptoms or conditions.

There are several resources that providers can use to evaluate exposures in anticipation of or during pregnancy. Databases on reproductive risk are available as online references at many hospitals and health care facilities or by subscription (e.g., the Reprorisk system, which is made up of four such databases, is available online [http://www.micromedex.com/products/reprorisk]. The Food and Drug Administration's classification of drugs according to teratogenic potential [*see Table 1*] and compendia such as the *Physicians' Desk Reference* (PDR) are often used as resources for information on drugs in pregnancy and as guides to patients and providers. Each has important limitations, however. Ratings in the FDA system are often more a reflection of available studies than of evidence of teratogenesis, and the PDR reports all associations between drugs and outcome without rigorously evaluating causality.

The baseline risk of congenital anomalies in the United States is approximately 2% to 3%.[6] To be considered teratogenic, a drug or exposure must be associated with a risk that is higher than threshold, ideally with anomalies clustered in a specific pattern. In fact, only a few drugs are identified teratogens [*see Table 2*].[7] Family history may also increase the risk that a couple will have a child with birth defects. For example, the risk for congenital heart defects is 0.4% to 0.8% in the general population but rises to 3% to 10% if the mother or father had a heart defect at birth.[8] Other maternal characteristics or behaviors may increase the risk of anomalies; for example, maternal obesity is associated with an increased risk of cardiac and neural tube defects.[9] Finally, environmental exposures, such as exposure to heat or ionizing radiation, especially during the critical first trimester, can increase the risk of miscarriage or birth defects such as spina bifida.[10]

Table 1 The Food and Drug Administration Drug Classification System for Pregnancy

Class A	Controlled studies in women fail to demonstrate a risk to the fetus in the first trimester, and there is no evidence of a risk in later trimesters; the possibility of fetal harm appears remote.
Class B	Either animal-reproduction studies have not demonstrated a fetal risk (and there are no controlled studies in pregnant women) or animal-reproduction studies have shown an adverse effect (other than a decrease in fertility) that was not confirmed in controlled studies in women in the first trimester, and there is no evidence of a risk in later trimesters.
Class C	Either studies in animals have revealed adverse effects on the fetus (teratogenic or embryocidal or other) and there are no controlled studies in women, or studies in women and animals are not available. Drugs should be given only if the potential benefit justifies the potential risk to the fetus.
Class D	There is positive evidence of human fetal risk, but the benefits from use in pregnant women may be acceptable despite the risk (e.g., if the drug is needed in a life-threatening situation or for a serious disease for which safer drugs cannot be used or are ineffective).
Class X	Studies in animals or human beings have demonstrated fetal abnormalities, there is evidence of fetal risk based on human experience, or both, and the risk of the use of the drug in pregnant women clearly outweighs any possible benefit. The drug is contraindicated in women who are or may become pregnant.

Physiologic Changes in Pregnancy

A variety of physiologic adaptations take place over the course of pregnancy. For example, blood volume increases by as much as 50%; red blood cell mass also increases, but to a lesser extent, resulting in a mild dilutional anemia. Cardiac output rises in compensation. The total blood leukocyte count increases to as much as 15,000 cells/ml in the third trimester. The albumin concentration decreases, which tends to increase free levels of protein-bound drugs. The glomerular filtration rate and the renal plasma flow rate increase until midpregnancy, typically by 40%.

Respiratory changes include an increase in tidal volume. The majority of pregnant women experience dyspnea in the first trimester as a paradoxical result of lower carbon dioxide levels.

Placental secretion of hormones, such as placental lactogen, promotes maternal insulin resistance with subsequent postprandial hyperglycemia; glucose is shunted to the fetus, and the mother uses ketones and triglycerides to meet her metabolic needs.

Liver enzyme levels change during pregnancy. Alkaline phosphatase levels double, whereas there is a slight decrease in levels of aspartate aminotransferase, alanine aminotransferase, γ-glutamyl transpeptidase, and bilirubin.

Changes in the coagulation system include marked increases in levels of fibrinogen and factor VIII, with lesser increases in factors VII, IX, X, and XII; protein S levels (both free and bound) decrease. Together with the increase in venous stasis in the lower extremities, these changes can promote thrombosis.

Cardiovascular Disease

HYPERTENSION

Epidemiology

Up to 5% of pregnant women have chronic hypertension.[11] Moreover, the prevalence of chronic hypertension in pregnancy is increasing, for two reasons: (1) women are having children at older ages, when chronic hypertension is more common, and (2) with the increase in obesity in the general population, chronic hypertension is developing at younger ages.

Diagnosis

Chronic hypertension is defined by the Working Group on High Blood Pressure in Pregnancy as a blood pressure greater than 140/90 mm Hg before pregnancy or in the first half of pregnancy.[11]

Management

Blood pressure should be followed closely throughout pregnancy, to guide decision making about initiation or adjustment of medication. In normal pregnancy, blood pressure falls in the late first trimester and returns to prepregnancy ranges near term. Blood pressure in most women with chronic hypertension follows this pattern.

During pregnancy, blood pressure must be kept high enough to maintain placental perfusion. Consequently, blood pressure goals for pregnant women are higher than those for nonpregnant women. The Working Group on High Blood Pressure in Pregnancy recommends antihypertensive treatment for pregnant women with a diastolic blood pressure of 105 mm Hg or higher to decrease risk of maternal stroke and intracerebral hemorrhage.[11] A meta-analysis of existing studies indicates that antihypertensive therapy in women with mild to moderate chronic hypertension does not influence the risk of developing superimposed preeclampsia.[12]

No antihypertensive medications are designated category A in pregnancy, but methyldopa is recommended as first-line therapy, given its long history of use without adverse effects.[11] However, because methyldopa often causes fatigue and has limited potency, labetalol, beta blockers, or calcium channel blockers are often used. Trials of antihypertensive therapies and trials comparing the effectiveness of different agents in pregnancy are very limited; consequently, few data exist to guide choices among these agents.

Angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers are considered category C in the second and third trimester, because of associated neonatal renal failure.[13,14] Some studies have shown that exposure to ACE inhibitors during the first trimester does not result in ill effects.[13] This finding is important for counseling women who conceive inadvertently while on this class of medication. However, given the number of unplanned pregnancies and the fact that many women do not present for prenatal care in the first trimester, it is prudent that women on these agents be switched to agents of another class when they are planning pregnancy.

Because chronic hypertension in pregnancy is associated with an increased risk of fetal growth restriction and other manifestations of uteroplacental insufficiency, including stillbirth, and because intensive antihypertensive treatment in pregnancy has been associated with intrauterine fetal growth restriction,pregnant women who have chronic hyertension should be carefully monitored. This is usually accomplished by regular evaluation of fetal growth with ultrasonography and, as the pregnancy approaches term, weekly or biweekly evaluation of fetal well-being with nonstress tests (20- to 30-minute evaluations of the pattern of fetal heart rate tracing) or biophysical profiles (ultrasonographic evaluation of fetal movement and amniotic fluid volume). Recognizing the increased risk attendant with these pregnancies, most practitioners plan delivery before the 40th week of gestation even if test results up to that point are reassuring.

Prognosis

Most women with chronic hypertension do well during pregnancy. However, pregnant women with poorly controlled hypertension are at risk for stroke and cerebral hemorrhage, and they are at increased risk for placental abruption.[15] In addition, preeclampsia (exacerbation of hypertension with onset or worsening of proteinuria [*see* Pregnancy-Specific Conditions, Preeclampsia, *below*] develops in approximately 25% of women with chronic hypertension, compared with an overall incidence of 5% in the general population.[15] Preeclampsia has the highest risk of maternal and fetal problems associated with essential hypertension. Despite much research, efforts to decrease the incidence of preeclampsia in women at increased risk for it have been unsuccessful. For these reasons, it is important that women with chronic hypertension be followed more closely during pregnancy. Measuring the 24-hour urinary protein level before pregnancy can facilitate the diagnosis of preeclampsia during pregnancy in women with chronic hypertension.

VALVULAR DISEASE

In general, valvular heart diseases such as pulmonic and tricuspid lesions or mitral valve prolapse or regurgitation are well tolerated during pregnancy. Previously undiagnosed and asymptomatic mitral stenosis, however, may become symptomatic during pregnancy as a result of the increased plasma volume. Women with moderate or severe mitral stenosis are at particular risk in the postpartum period, when the fluid previously sequestered in the placenta or lower extremities returns to the systemic circulation, augmenting the existing increased volume. During labor, careful attention to fluid management and judicious diuresis is required for women with mitral stenosis. Epidural anesthesia may help both by providing vascular relaxation,

Table 2 Selected Drugs with Suspected or Known Teratogenic Potential[7]

Alcohol	Misoprostol
Androgens	Nonsteroidal anti-inflammatory drugs
Danazol	Phenytoin
Angiotensin-converting enzyme inhibitors	Quinolones
Anticholinergic drugs	Retinoids and derivatives
Antithyroid drugs	Acitretin
Methimazole	Etretinate
Propylthiouracil	Isotretinoin
Carbamazepine	Tetracycline
Cyclophosphamide	Thalidomide
Cocaine	Valproic acid
Lithium	Warfarin
Methotrexate	

which slows increases in plasma volume after delivery, and by providing pain relief, which can help slow the patient's heart rate. Lower heart rates in these patients improve the heart's efficiency by promoting adequate atrial filling and the "kick" needed to push blood past the stenotic mitral valve.[16] Symptomatic mitral stenosis has become less common with the fall in the incidence of rheumatic heart disease, but patients who experience critical stenosis can be treated safely during pregnancy with balloon valvuloplasty.[17]

ARRHYTHMIAS

Increased plasma volume may contribute to the onset or worsening of maternal dysrhythmias during pregnancy. Many antiarrhythmic agents are appropriate for use during pregnancy, however, and cardioversion, implantable defibrillators, and pacemakers have all been safely used to control symptomatic dysrhythmias in pregnant women.[18]

PERIPARTUM CARDIOMYOPATHY

Peripartum cardiomyopathy affects one in 1,500 to 15,000 pregnancies and is marked by heart failure in the months before or after delivery. The cause is unknown. Diagnosis is confirmed by echocardiographic demonstration of decreased ejection fraction in the absence of another recognized etiology.[19] Treatment is supportive and focuses on supplemental oxygen, inotropic medication, and diuresis. With such support, 50% of patients improve over a period of weeks. In those who do not improve, cardiac transplantation may be required. In subsequent pregnancies, women who have previously had pregnancies complicated by peripartum cardiomyopathy demonstrate further decline in left ventricular ejection fraction. Death occurs principally in women who enter their next pregnancy with persistent left ventricular dysfunction; mortality in this group can be as high as 20%.[20]

Diabetes Mellitus

EPIDEMIOLOGY

Approximately 1% to 2% of pregnant women have DM that predates their pregnancy (i.e., pregestational DM). In another 2% to 3%, DM develops during the pregnancy (i.e., gestational DM) [*see* Pregnancy-Specific Conditions, Gestational Diabetes Mellitus, *below*]. As with chronic hypertension, changes in the demographic profile of women who become pregnant have led to an increase in the frequency of type 2 DM in pregnancy.

MANAGEMENT

Women with pregestational DM should receive preconception care that includes counseling about the importance of preconception glucose control. Preconception care improves glucose control and decreases maternal hospitalizations, the need for neonatal intensive care, birth defects, and fetal and neonatal deaths.[21]

Diet therapy is a mainstay of the management of DM outside of pregnancy and is key in pregnancy. Insulin is the preferred agent for the treatment of DM in pregnancy. Women with type 1 DM will already be on insulin; women with type 2 DM who are on oral agents (e.g., hypoglycemics and insulin sensitizers) should be switched to insulin before conception. Insulin therapy should be intensified before conception to achieve a glycosylated hemoglobin level in the normal range, when possible. Neutral protamine Hagedorn (NPH) and regular insulins have a long track record of use in pregnancy. The short-acting insulins, in particular insulin lispro, have been used increasingly in pregnancy and appear to be safe, but long-term follow-up of infants exposed to these agents in utero is not yet available. The long-acting insulins, such as insulin glargine, have not been extensively tested in pregnancy. Goals for whole-blood glucose levels recommended by the American Diabetes Association (ADA) during pregnancy are fasting levels below 95 mg/dl and 1-hour postprandial levels below 140 mg/dl.

Insulin resistance is a hallmark of normal pregnancy, and it occurs in women with DM as well. As a result, insulin requirements typically increase in pregnancy.

Both type 1 and type 2 DM are associated with an increased risk of fetal growth abnormalities. Women with type 1 DM, particularly those with vascular complications, are at risk for fetal growth restriction. Conversely, women with type 2 DM are at risk for large (macrosomic) newborns. In either case, fetal growth in women with diabetes can be monitored with serial ultrasound scans during pregnancy. DM, particularly when poorly controlled, has also been linked to intrauterine fetal demise; accordingly, for pregnant women with diabetes, most practitioners employ a careful program of fetal surveillance that includes nonstress tests, biophysical profiles, and fetal movement counting. As with pregnancies that are at increased risk from other medical complications, delivery is often planned before the 40th week of gestation.

COMPLICATIONS AND PROGNOSIS

DM is associated with increased risks for both the mother and the fetus. For the mother with DM, there may be acceleration of retinopathy and nephropathy during pregnancy. In addition, women with DM are at increased risk for preeclampsia. For the fetus, poor maternal glycemic control at the time of conception may be associated with a birth defect rate as high as 20%. This rate can be reduced to a rate similar to that of the general population if glycemic control is adequate at conception.

Unless insulin doses are increased to compensate for the insulin resistance of pregnancy, women with type 1 DM may experience diabetic ketoacidosis. In the setting of uncontrolled maternal hyperglycemia, the fetus responds by increasing insulin secretion. Insulin acts as a growth factor for the fetus and can cause macrosomia, which increases the chance of injury during vaginal delivery, as well as the need for cesarean section. This elevation in insulin secretion does not resolve immediately with delivery and can lead to neonatal hypoglycemia. The neonates of women who had DM during pregnancy are also at increased risk for respiratory distress syndrome.

Thyroid Disease

EPIDEMIOLOGY

In general, thyroid disease is more common in women than in men. Both hypothyroidism and hyperthyroidism are common in pregnant women.

HYPERTHYROIDISM

Hyperthyroidism complicates about 0.2% of pregnancies.[22] Graves disease is the most common cause of hyperthyroidism in pregnancy.

Diagnosis

The diagnosis of hyperthyroidism during pregnancy is often complicated by the changes in thyroid function that take place in

normal pregnancy. In early pregnancy, the rise in levels of human chorionic gonadotropin, which has thyroid stimulatory activity, results in a compensatory fall in levels of thyroid-stimulating hormone (TSH). Although the fall in TSH concentration is usually within the physiologic range, in some women the TSH falls into the suppressed range. Therefore, the diagnosis of Graves disease should be based not solely on a low TSH level but also on the presence of symptoms of hyperthyroidism and of elevated thyroxine (T_4) levels. Symptoms of hyperthyroidism in pregnancy include weight loss or absence of weight gain, anxiety, and palpitations. Although it is possible that a small degree of thyroid enlargement occurs in normal pregnancy, secondary to increase in plasma volume and thyroid blood flow, a palpable goiter should not be attributed to pregnancy.

Management

Antithyroid medications (thionamides) do cross the placenta, and the fetal thyroid appears to be more sensitive to these agents than the maternal thyroid. Therefore, maternal treatment of euthyroidism is associated with fetal hypothyroidism in about 25% of cases.[23] In addition to fetal and neonatal hypothyroidism, neonatal goiter can be present and can cause respiratory compromise from tracheal compression.

Treatment of maternal hyperthyroidism is recommended only for mothers who are symptomatic. The treatment goal is a T_4 level in the upper range of normal. Hyperthyroidism in pregnancy is one of the few clinical situations in which therapy is guided by the T_4 level rather than the TSH level; normalization of the TSH level is not the goal in these patients.

Propylthiouracil (PTU) and methimazole are the most common drugs used for hyperthyroidism in the United States. There has been some suggestion that PTU crosses the placenta to a lesser degree than methimazole, although the data are limited. Furthermore, methimazole, but not PTU, has been associated with the condition aplasia cutis, a skin defect most commonly involving the scalp. As a result, PTU is the preferred drug for pregnancy. If antithyroid medication does not control the hyperthyroidism and the woman has severe symptoms, thyroidectomy during pregnancy may be required. Radioactive iodine (RAI) ablation is contraindicated during pregnancy, because it can ablate the fetal thyroid and cause congenital hypothyroidism. Concern over uncontrollable hyperthyroidism in pregnancy has led some endocrinologists to recommend RAI thyroid ablation or surgery in women with Graves disease who are planning a pregnancy and in whom control of Graves disease has been difficult, as a result of disease activity or problems with medication. Women with Graves disease who are going to be managed on antithyroid medications should be followed by a clinician with experience in thyroid disease during pregnancy, because the goals of treatment differ from those in nonpregnant patients.

Complications and Prognosis

Severe hyperthyroidism may be associated with stillbirth, preterm delivery, intrauterine growth restriction, preeclampsia, or congestive heart failure.[24] Graves disease usually follows a characteristic course during pregnancy, with exacerbation in the first trimester and improvement or even remission in the third trimester, with exacerbation again about 2 months post partum.

HYPOTHYROIDISM

Maternal hypothyroidism complicates 2% to 5% of pregnancies. For unknown reasons, women with primary hypothyroidism have an increased requirement for thyroid hormone during pregnancy.[25]

Diagnosis

Hypothyroidism is best diagnosed by determination of the TSH level. The upper limit of normal in pregnancy is similar to that outside of pregnancy.

Management

Women with hypothyroidism should have their TSH level measured at the time of a pregnancy test to guide adjustment of their thyroid hormone replacement dose. The TSH level should then be checked at a minimum of each trimester, as well as 4 to 6 weeks after each dose adjustment. Because of the increase in thyroid hormone requirement in pregnancy, some advocate regulating dosing of thyroid hormone before conception to achieve a TSH level of around 1 IU/ml. The increased requirement resolves with delivery, and therefore, the thyroid hormone dose can be reduced to the prepregnancy dose at that time.

Complications and Prognosis

Maternal hypothyroidism has been reported to increase the risk for preeclampsia, placental abruption, and stillbirths. A study of over 400 pregnant women with subclinical hypothyroidism (defined as a TSH level of ≥ 97.5% of the upper range of normal and a T_4 level within normal range) found that rates of hypertension were no different from those in euthyroid pregnant women; however, subclinical hypothyroidism was associated with an increased risk of placental abruption and preterm delivery.[26] The offspring of mothers who were hypothyroid during pregnancy have been reported to have intelligent quotients (IQs) 7 points lower than those of the offspring of euthyroid mothers, although IQs remain in the normal range.[27]

POSTPARTUM THYROIDITIS

The immunologic flare that follows the relative immune suppression of pregnancy may result in thyroiditis. Postpartum thyroiditis is typically transient; it is marked by a period of hyperthyroidism followed by hypothyroidism.

Epidemiology

Postpartum thyroiditis occurs after approximately 8% of pregnancies.[28] Women with type 1 DM are at threefold higher risk; the risk is also higher in women who have had Hashimoto thyroiditis and in those who had postpartum thyroiditis after a previous pregnancy.

Diagnosis

About 2 months after delivery, women with postpartum thyroiditis may experience hyperthyroidism as a result of leakage of thyroid hormone from the inflamed thyroid gland. This phase is often missed because of its short duration (usually 4 weeks or less) and because the patient and her caregivers tend to ascribe its symptoms of anxiety and palpitations to the stress of being a new mother. A TSH level measured at this time will be low. About 6 months after delivery, the hypothyroid phase develops. At this time, patients usually present with fatigue and weight gain or an inability to lose pregnancy-associated weight. Depression can also be a presenting symptom; for that reason, it is important to check the TSH level before making a diagnosis of postpartum depression.

Management

Whether to initiate thyroid hormone therapy in a woman with postpartum thyroiditis depends on the severity of her symptoms. If thyroid hormone replacement therapy is chosen, thyroid hormone can be withdrawn at 1 year post partum to determine whether long-term therapy is indicated. The TSH level should be checked 4 to 6 weeks after stopping therapy, or it should be checked sooner if symptoms of hypothyroidism develop.

Prognosis

In the majority of women with postpartum thyroiditis, thyroid function returns to normal in the ensuing postpartum year. However, some women remain permanently hypothyroid. Overall, approximately 25% of women who have postpartum thyroiditis progress to permanent hypothyroidism in the ensuing 10 years.[28]

Thrombophilia

The risk for thromboembolism during pregnancy is increased, largely as a result of increased venous stasis in the lower extremities and pelvis. In addition, changes in some serum coagulation factors (i.e., elevation of factor VII, factor X, factor VIII, fibrinogen, and von Willebrand factor and decrease of protein S) may contribute to the pregnancy-associated risk for thrombosis. This risk may be particularly prominent in women who have another risk factor for thrombosis, such as inherited or acquired thrombophilia or a history of thrombosis unassociated with trauma—in particular, thrombosis associated with the use of oral contraceptives. Prophylaxis with aspirin or heparin during pregnancy may be indicated for women with any of these risk factors.[29] Although the safety of low-molecular-weight heparin (LMWH) during pregnancy is not definitively established, expert panels have concluded that LMWH may be safely used during pregnancy.[30] Because the timing and course of labor and delivery are not easily predicted and because LMWH has a long duration of action, conversion to unfractionated heparin is generally planned late in the third trimester to minimize bleeding complications from delivery or regional anesthesia. The use of LMWH or unfractionated heparin during pregnancy requires careful monitoring, because changes in plasma volume may affect the usual dosing.[31]

Some studies have suggested that thrombophilias may also be associated with an increased risk of miscarriage and other pregnancy complications, such as stillbirth, growth restriction, and placental abruption.[32] A detailed review of these investigations is beyond the scope of this chapter. Data are, at best, inconclusive, and there is little evidence that intervention (e.g., treatment with heparin, aspirin, or steroids) improves outcome, except possibly in women with antiphospholipid antibody syndrome.[33]

Asthma

As many as 6% of pregnant women have asthma. This disease has a variable course during pregnancy: it is equally likely to worsen, improve, or remain unchanged.[34] In contrast to past reports, current studies indicate that asthma is not associated with an increased risk of preterm delivery and growth restriction.[35] The clinical course in an individual patient may be linked, at least in part, to the patient's and her clinician's willingness to use needed medications during pregnancy. Guidelines from the National Institute of Child Health and Human Development emphasize that continued use of beta agonists and steroids during pregnancy is safe and appropriate.[36] As in nonpregnant women, clinical evaluations such as peak flow measurements can be used to judge disease activity and guide therapy.

Infectious Diseases

Vertical transmission of maternal infection is rare during pregnancy, but some pathogens carry particular pregnancy-associated risks. In addition, infections that are of little or no clinical significance to a woman when she is not pregnant may have serious implications during pregnancy (e.g., parvovirus or cytomegalovirus infection). It is important to note that pregnancy is not a contraindication to most prophylactic measures against infection—including, as appropriate, malaria prophylaxis or vaccination with inactive agents. Seasonal vaccination with influenza is recommended for pregnant women.

HERPES SIMPLEX VIRUS INFECTION

Neonatal infection with herpes simplex virus (HSV) acquired during delivery as a consequence of viral shedding from the cervix and vagina can have devastating consequences for the newborn. Cesarean delivery, especially if performed before active labor when the amniotic membranes are still intact, largely eliminates the possibility of vertical transmission.[37] The risk of delivering an HSV-infected infant is greatest in women with a primary infection with HSV type 1 or 2 at the time of delivery. Although secondary infection carries lower neonatal risk, as a result of the protection afforded by circulating HSV antibodies shared across the placenta with the fetus, cesarean delivery to further limit vertical transmission is still generally recommended when women present in labor with signs or symptoms of secondary infection. To avoid cesarean delivery and reduce neonatal risk, prophylactic treatment with antiviral agents such as acyclovir may be recommended late in the third trimester in an effort to minimize the chances of an outbreak at the time of labor and delivery.[38]

PARVOVIRUS INFECTION

Infection with parvovirus B19, which produces a viral exanthem known as fifth disease, may at times be epidemic in schools or child care settings. Parvovirus infection is usually not of clinical consequence to immunocompetent adults or children. Rarely, fetuses of newly infected pregnant women can develop significant anemia or other complications as a result of transplacental infection. Estimating the precise fetal risk has proved to be challenging, but some series suggest that the risk of the loss of pregnancy when a new maternal infection occurs before the 20th week of gestation may range from 2.5% to 10%.[39] Both early and later infections can result in fetal bone marrow suppression and subsequent anemia; in turn, the anemia leads to pericardial effusion, pleural effusion, and generalized body edema—a condition known as hydrops fetalis. It should be emphasized that such complications are rare, however.[40] Although the anemia is often self-limited, either delivery or in utero fetal red blood cell transfusion may be considered if hydrops fetalis develops.

HIV INFECTION

Although transplacental transmission of HIV infection from mother to fetus has been described, vertical transmission of HIV occurs largely at the time of delivery and is directly linked to maternal viral load. To reduce transmission and optimize maternal

health, antiretroviral regimens that were initiated before pregnancy are continued during pregnancy.[41] Because of mitochondrial toxicity, the combination of didanosine and stavudine should be avoided during pregnancy. Additional concern has been raised about teratogenic effects of efavirenz in animals. For women who have not previously been treated with antiviral agents and whose viral loads and $CD4^+$ T cell counts do not themselves dictate treatment, zidovudine is prescribed during the antepartum and intrapartum period, with postpartum treatment for the newborn, because such treatment has been linked to a reduction of vertical transmission.[42] Cesarean delivery has also been shown to protect against vertical transmission in women with HIV infection and high viral loads.[43] Consequently, cesarean delivery should be offered to patients with more than 1,000 copies/ml.[41] Breast-feeding is not recommended for women with HIV infection if appropriate alternatives (e.g., clean water and formula) exist.[44]

HEPATITIS

Women with active hepatitis B or C are at risk for vertical transmission of such infections to their fetus, particularly if either the presence of e-antigen (in the case of hepatitis B) or copy counts indicate high viral loads. Administration of hepatitis B vaccine and immunoglobulin reduces neonatal infection with hepatitis B by 90%. Cesarean delivery is not routinely recommended in the setting of either of these maternal infections. Breast-feeding is not contraindicated in mothers with hepatitis C, nor is it contraindicated in mothers with hepatitis B whose newborns have received appropriate prophylaxis.

GROUP B *STREPTOCOCCUS* INFECTION

In 20% of women, group B *Streptococcus* (GBS) can be cultured from the rectovaginal area. Such carriage, of no consequence to adults, can result in early- or late-onset GBS sepsis and mortality in three of 1,000 infants delivered to women who are carriers. Screening for GBS colonization is now recommended during pregnancy; when linked with intrapartum treatment, such screening has reduced neonatal infection to less than one in 1,000 deliveries.[45]

Renal Disease

A 1996 study of the outcomes of 87 pregnancies in 67 women with moderate or severe preexisting renal disease described preterm delivery in 59% and fetal growth restriction in 37%—rates considerably higher than those seen in the general population (11% and 10%, respectively).[46] In addition, pregnancy-related loss of renal function was noted in 43% of these women, although distinguishing natural disease progression from progression caused by the pregnancy is extremely challenging at an individual level and can be difficult even at a population level. Accordingly, careful monitoring of both mother and baby are required during such pregnancies. Some conditions (e.g., autosomal dominant polycystic kidney disease and structural anomalies such as single kidney) increase the risk of fetal renal disease.

Renal transplantation can restore fertility in women with end-stage renal disease. However, it is generally recommended that transplant recipients delay conception until graft function has been stable for 1 to 2 years, blood pressure is well controlled, and the immunosuppressive regimen is optimized. In one review of 2,300 pregnancies in 1,600 women who had had a renal transplant, 13% of patients had a spontaneous abortion (a rate similar to that in the general population), and 27% elected pregnancy termination; of pregnancies continuing beyond the first trimester, 92% were termed successful, although many of these were complicated by issues of prematurity and growth restriction.[47] Worsening graft function has been noted during posttransplant pregnancies, but whether the deterioration is any worse than that expected outside of pregnancy remains a matter of debate.

Autoimmune Diseases

Whether pregnancy worsens maternal autoimmune disease, such as systemic lupus erythematosus (SLE), remains controversial. Women with such conditions are at risk for preeclampsia and fetal growth restriction. Outcomes for mother and baby are, in general, felt to be best when maternal disease is quiescent.[48] In some cases, continued treatment with steroids may be needed to maintain quiescence and optimize outcomes. Fetuses of women with SLE or other connective tissue diseases and anti-Ro/SS-A or anti-La/SS-B antibodies are at special risk for neonatal lupus, a condition that may be marked by fetal or neonatal heart block, skin disease, and hepatic inflammation.[49] Testing for these serologic markers is important for guiding both fetal and neonatal care. For example, congenital heart block can be detected in utero, and some have suggested that early detection, when linked with steroid treatment of the mother, can improve outcomes.[50]

Recurrent miscarriage, stillbirth, growth restriction, and preeclampsia more frequently complicate pregnancies of women with antiphospholipid antibodies. In small studies, treatment with heparin, aspirin, or both has been shown to improve pregnancy outcome in some women with symptomatic antiphospholipid antibody syndrome.[51] Such data are limited, however, and treatment should carefully be considered on a case-by-case basis.

Cancer

There is no evidence that pregnancy increases the development, hastens the progression, or promotes the recurrence of any cancer. Cancers, particularly those more prevalent in populations of young women (e.g., breast and cervical cancer), will be found during pregnancy, however. Indeed, given that pregnancy may be one of the few times that many young women seek medical care, the examination and screening conducted during pregnancy may improve detection of incidental malignancies. Appropriate diagnostic studies, including mammography, Papanicolaou smear and colposcopy, and endoscopy should not be deferred because of pregnancy.

Management of cancer that is coincident with pregnancy presents many challenges. Because some forms of cancer and its treatment carry risks for both the woman and her fetus, some women may choose pregnancy interruption. In particular, management of invasive cervical cancer requiring surgical excision presents particular challenges. For other cancers in women who choose to continue their pregnancies, surgical and medical management need not be delayed by pregnancy. Several series describe successful chemotherapeutic treatment of malignancies diagnosed early in gestation, in concert with careful fetal monitoring.[52] For malignancies diagnosed later in gestation, early delivery may be considered so as to facilitate treatment.

Neurologic Diseases

EPILEPSY

The management of epilepsy before and during pregnancy demonstrates the careful balance of risks and benefits attendant with treating women of reproductive age, particularly those who require long-term medication. Most antiseizure drugs are associated with an increased risk of congenital anomalies—6% to 9%, compared with the 3% risk for birth defects in the general United States population. These defects cluster in a characteristic pattern that includes facial defects and microcephaly.[53] The risk is further increased by the need for multiple medications or the use of a few particularly teratogenic medications; for example, valproic acid is associated with a 1% to 3% risk of open neural tube defects. Balanced against these risks are the morbidities associated with seizures if medications are discontinued or are inappropriately dosed during pregnancy. Although pregnancy itself does not appear to either lower seizure thresholds or otherwise increase seizure frequency, changes in plasma volume may require increases in dosing to maintain effective drug levels. With careful monitoring, most women with epilepsy can anticipate healthy outcomes for themselves and their babies.[54]

MULTIPLE SCLEROSIS

Multiple sclerosis generally improves during pregnancy—a time of relative immunosuppression. The disease may flare post partum, however.[55] Although steroids might be used during pregnancy if a woman were to have worsening symptoms, many other contemporary treatments are either poorly studied in pregnancy (i.e., glatiramer acetate) or generally discouraged during pregnancy (i.e., interferons).

MYASTHENIA GRAVIS

Infants of women with myasthenia gravis are at risk for a self-limited neonatal myasthenia syndrome that results when antibodies to the acetylcholine receptor cross the placental circulation. Once delivery and cord clamping interrupt the placental circulation, the newborn clears these antibodies and, over a period of weeks, the syndrome improves. In the interval, affected newborns may require ventilatory and nutritional support until muscles involved in breathing and feeding again function normally.

Substance Use

The use of substances, legal and illegal, can have important consequences for pregnancy. Cigarette smoking is associated with an increased risk of intrauterine growth restriction, preterm delivery, and sudden infant death. Although nicotine is labeled class D (i.e., there is positive evidence of its association with human fetal risk) by the FDA, many clinicians use nicotine replacement to help pregnant women quit smoking.[56] In fact, there is reason to believe that nicotine might carry fewer risks when it is absorbed via transdermal patch or chewing gum than when it is delivered together with the carbon monoxide and other toxins contained in inhaled tobacco smoke.

Although the effects of cocaine on the behavior and neural development of newborns have in the past been exaggerated,[57] the use of cocaine during pregnancy may cause placental abruption. Heroin use may cause fetal growth abnormalities, whereas methadone therapy can be initiated and safely maintained during pregnancy.[58] Infants of mothers who use methadone or other prescription and nonprescription narcotics during pregnancy may be born with drug dependence and require careful monitoring after birth to prevent acute withdrawal. In such newborns, tincture of opium is sometimes given and then slowly titrated to minimize the consequences of acute withdrawal.

Fetal alcohol syndrome—marked by abnormal facial features, growth retardation, and central nervous system disturbances—is a consequence of long-term alcohol use. Although higher frequency and volumes of drinking are associated with an increased risk of fetal alcohol syndrome, no safe level of alcohol use during pregnancy can be defined or recommended.[59]

Data examining the effects of caffeine on pregnancy are indeterminate, but some small studies suggest that high exposures (i.e., more than three to five cups of regular coffee a day) may increase the risk of miscarriage. Prudence therefore argues for moderation in women planning to become pregnant.[60]

Both because of the recognized fetal risks and because prenatal care offers a series of regular practitioner contacts, pregnancy may offer an ideal opportunity to address problems of substance use and abuse. Intervention, however, requires a careful balance that encourages treatment without limiting maternal liberties. Mandated screening and reporting may drive women from prenatal care and otherwise limit the therapeutic opportunity that pregnancy presents.

Intrahepatic Cholestasis

Intrahepatic cholestasis appears to result from impaired processing of bile salts and acids in the liver. The etiology remains uncertain, but evidence suggests a genetic predisposition. Intrahepatic cholestasis occurs in 0.1% to 0.01% of pregnant women. The accumulation of these salts in the skin can cause an intense and characteristic pruritus (worse on palms and soles) without associated skin lesions. Jaundice develops in only about 10% of patients.

Liver function tests in women with intrahepatic cholestasis of pregnancy show elevations in levels of γ-glutamyl transpeptidase, alanine aminotransferase, and aspartate aminotransferase. The most specific test is an elevation in the fasting serum total bile acid concentration.

Whether deposition of bile salts in the placenta results in fetal morbidity and mortality is much debated, but some series indicate an increased risk of stillbirth and postpartum hemorrhage in women with cholestasis. Consequently, many practitioners institute a program of fetal monitoring and maternal prothrombin measurement in cholestatic pregnancies and may recommend induced delivery between 36 and 39 weeks of gestation.[61] Ursodeoxycholic acid can relieve pruritus and help normalize liver enzyme levels and may improve fetal outcome.[62]

Pregnancy-Specific Conditions

There are several medical complications of pregnancy that occur only during pregnancy and that resolve once the pregnancy is completed. Three of these pregnancy-specific conditions are hyperemesis gravidarum, preeclampsia, and gestational diabetes.

HYPEREMESIS GRAVIDARUM

Nausea and vomiting complicate as many as 70% of pregnancies. The cause of these symptoms remains obscure but may be related to levels of human chorionic gonadotropin and other hormones that are affected by pregnancy. Older theories suggesting that nausea and vomiting are a psychosomatic reflection of a woman's rejection of her pregnancy have been discredited. Never-

theless, different cultures and individuals may manage symptoms differently, and in any woman, the symptoms may cause stress.

Unless other findings suggest specific causes for nausea and vomiting (e.g., hyperthyroidism, hepatitis or other viral syndromes, or hydatidiform mole), little testing is required; care is supportive. Women may be reassured that the dietary limitations their symptoms require are unlikely to compromise their pregnancy, and in fact, pregnancies accompanied by nausea and vomiting are more likely to be healthful than those without such symptoms.[63] Several randomized, controlled trials have demonstrated that supplementation with vitamin B_6 (10 to 50 mg two or three times a day) may improve symptoms.[64] Some women may require other medications, including antihistamines and dopamine antagonists, to control symptoms.[65]

Fortunately, in most women, nausea and vomiting abate between the 12th and 16th weeks of gestation. Only rarely is hospitalization or enteral nutrition required.

PREECLAMPSIA

Preeclampsia is defined as the onset of hypertension (i.e., blood pressure greater than 140/90 mm Hg) and proteinuria in the second half of pregnancy.[11] Proteinuria in pregnancy is defined as a finding of more than 300 mg of protein on a 24-hour urine collection. Preeclampsia may also be superimposed on chronic hypertension; in these cases, there is an exacerbation of hypertension accompanied by new-onset proteinuria or worsening of preexisting proteinuria. Preeclampsia complicates 3% to 5% of pregnancies. It occurs most commonly in first pregnancies.

Pathogenesis

The cause of preeclampsia is unknown. At the root of the disorder is endothelial dysfunction and placental insufficiency. Several potential factors in preeclampsia include immune dysregulation, genes, insulin resistance, and placental factors such as soluble fms-like tyrosine kinase–1.[66]

Management

Because the cause of preeclampsia is not known, effective interventions have been elusive. The only definitive treatment is delivery of the fetus. Antihypertensive treatment can decrease the maternal risk of stroke but has not been shown to improve pregnancy outcome. Prophylactic treatment with intravenous magnesium sulfate decreases the risk of eclampsia. Preventive efforts to lower the risk for preeclampsia with aspirin or calcium supplementation have been unsuccessful. Close surveillance of both the mother and the fetus is essential and includes more frequent visits or inpatient observation with periodic evaluation of liver and renal function, as well as other blood markers of disease progression (e.g., platelet count), antenatal testing (e.g. nonstress testing, biophysical profile, or evaluation of amniotic fluid alone), and evaluation of fetal growth.

Complications

Complications of preeclampsia affect both the mother and the fetus. Maternal complications include stroke and intracerebral hemorrhage. The HELLP syndrome (hemolysis, elevated liver function, and low platelets) is a life-threatening maternal complication of preeclampsia.

Prognosis

Preeclampsia resolves with delivery or shortly thereafter. Women with a history of preeclampsia are at higher risk for the disorder in future pregnancies. In addition, these women may be at increased risk for the later development of essential hypertension and other cardiovascular disease.

GESTATIONAL DIABETES MELLITUS

Epidemiology

Gestational diabetes mellitus (GDM) affects about 7% of all pregnancies. The rate of GDM is higher in obese women and in women belonging to certain ethnic groups (e.g., Hispanics).

Pathogenesis

Insulin resistance increases in normal pregnancy. Women who cannot overcome this insulin resistance develop GDM.

Diagnosis

GDM is defined as carbohydrate intolerance diagnosed during pregnancy. However, no universal criteria exist for screening or diagnosis; these criteria differ within and between countries. The ADA recommends screening for GDM in all pregnant women with at least one of the following risk factors: age 25 years or older; ethnicity other than white; family history of DM in a first-degree relative; being overweight before pregnancy or experiencing greater than usual weight gain during pregnancy; or a personal history of either abnormal blood glucose levels or a previous obstetric complication. In the United States, screening is typically performed with a 1-hour glucose tolerance test (GTT) using a 50 g oral glucose load. For women with blood glucose values greater than 140 mg/dl, the next step is a fasting 3-hour 100 g GTT. The diagnosis of GDM is made when patients have two or more GTT values higher than those of either the National Diabetes Data Group or Carpenter-Coustan criteria [*see Table 3*].[67]

Management

The mainstay of treatment for GDM is diet. The therapeutic goals are fasting whole-blood glucose levels below 95 mg/dl and 1-hour postprandial glucose levels below 140 mg/dl.[67] Women with GDM are taught to perform home glucose monitoring.

If glucose levels remain elevated despite dietary intervention, insulin therapy should be initiated. Although this recommendation has been made routinely for some time, evidence supporting the benefit of treatment for pregnancy outcome was lacking until 2005, when a study demonstrated that perinatal complications were less frequent in newborns of women with GDM who were randomized to management with home glucose monitor-

Table 3 Venous Plasma Glucose Criteria for the Diagnosis of Gestational Diabetes after a 100 g Oral Glucose Tolerance Test*

Time	*NDDG Criteria (mg/dl)*	*Carpenter-Coustan Criteria (mg/dl)*
0 (fasting)	105	95
1 hour	190	180
2 hour	165	155
3 hour	145	140

*Gestational diabetes mellitus is diagnosed when two or more results exceed these values.[67]
NDDG—National Diabetes Data Group

ing, dietary therapy, and insulin, compared with the newborns of women who received routine care.[68] A 2004 study suggested that GDM may be effectively managed with oral glyburide rather than insulin,[69] but further studies are needed before this approach is advocated.

Women with GDM who have poor glycemic control or who require insulin or oral agents to regulate blood sugar may benefit from antenatal fetal testing. Clinical or sonographic evaluation for fetal macrosomia may lead physicians to recommend cesarean section rather than vaginal delivery.

Complications

Elevated fasting glucose levels have been associated with fetal demise. Macrosomia risk is also increased. As with preexisting DM, macrosomia increases the chance of shoulder dystocia during vaginal delivery and resultant cesarean section. Like pregestational DM, GDM may result in neonatal hypoglycemia.

Prognosis

GDM is an important risk factor for the future development of type 2 DM. As many as 50% of women who have had GDM will develop type 2 DM over the ensuing 10 years.[70] Obesity further increases this risk. Consequently, questions about GDM should be a standard part of the primary care evaluation in parous women. The ADA recommends yearly screening for DM in women with a history of GDM. Whether screening should entail measurement of a fasting glucose level or a postprandial glucose level has not been established. It seems prudent that women with a history of GDM be advised to reduce weight and to exercise, although studies supporting this recommendation are lacking. Some studies suggest that the use of insulin sensitizers may reduce the rate of conversion of GDM to type 2 DM[71]; these data need to be confirmed in large-scale trials, and cost-effectiveness must be taken into consideration before the use of these agents can be recommended.

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Gleicher N, Midwall J, Hochberger D, et al: Eisenmenger's syndrome and pregnancy. Obstet Gynecol Surv 34:721, 1979

2. Weiss BM, Hess OM: Pulmonary vascular disease and pregnancy: current controversies, management strategies, and perspectives. Eur Heart J 21:104, 2000

3. Shabetai R: Cardiac Diseases in Maternal Fetal Medicine, 4th ed. Creasy RK, Resnik R, Eds. WB Saunders Co, Philadelphia, 1999, p 793

4. Mennuti MT, Thomson E, Press N: Screening for cystic fibrosis carrier state. Obstet Gynecol 93:456, 1999

5. ACOG practice bulletin. Clinical management guidelines for obstetrician-gynecologists. Number 44, July 2003. Neural tube defects. Obstet Gynecol 102:203, 2003

6. Nelson K, Holmes LB: Malformations due to presumed spontaneous mutations in newborn infants. N Engl J Med 320:19, 1989

7. Koren G, Pastuszak A, Ito S: Drugs in pregnancy. N Engl J Med 338:1128, 1998

8. Whittemore R, Wells JA, Castellsague X: A second-generation study of 427 probands with congenital heart defects and their 837 children. J Am Coll Cardiol 23:1459, 1994

9. Werler MM, Louik C, Shapiro S, et al: Prepregnant weight in relation to risk of neural tube defects. JAMA 275:1089, 1996

10. Milunsky A, Ulcickas M, Rothman KJ, et al: Maternal heat exposure and neural tube defects. JAMA 268:882, 1992

11. Report of the National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy. Am J Obstet Gynecol 183:S1, 2000

12. Abalos E, Duley L, Steyn DW: Antihypertensive drug therapy for mild to moderate hypertension in pregnancy. Cochrane Database Syst Rev (2):CD002252, 2001

13. Hanssens M, Keirse MJ, Vankelecom F, et al: Fetal and neonatal effects of treatment with angiotensin-converting enzyme inhibitors in pregnancy. Obstet Gynecol 78:128, 1991

14. Alwan S, Polifka JE, Friedman JM: Angiotensin II receptor antagonist treatment during pregnancy. Birth Defects Res A Clin Mol Teratol 73:123, 2005

15. Sibai BM, Lindheimer M, Hauth J, et al: Risk factors for preeclampsia, abruptio placentae, and adverse neonatal outcomes among women with chronic hypertension. National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. N Engl J Med 339:667, 1998

16. Reimold SC, Rutherford JD: Clinical practice: valvular heart disease in pregnancy. N Engl J Med 349:52, 2003

17. Fawzy ME, Kinsara AJ, Stefadouros M, et al: Long-term outcome of mitral balloon valvotomy in pregnant women. J Heart Valve Dis 10:153, 2001

18. Page RL: Treatment of arrhythmias during pregnancy. Am Heart J 130:871, 1995

19. Pearson GD, Veille JC, Rahimtoola S, et al: Peripartum cardiomyopathy: National Heart, Lung, and Blood Institute and Office of Rare Diseases (National Institutes of Health) workshop recommendations and review. JAMA 283:1183, 2000

20. Elkayam U, Tummala PP, Rao K, et al: Maternal and fetal outcomes of subsequent pregnancies in women with peripartum cardiomyopathy. N Engl J Med 344:1567, 2001

21. Korenbrot CC, Steinberg A, Bender C, et al: Preconception care: a systematic review. Matern Child Health J 6:75, 2002

22. Neale D, Burrow G: Thyroid disease in pregnancy. Obstet Gynecol Clin North Am 31:893, 2004

23. Momotani N, Noh J, Oyanagi H, et al: Antithyroid drug therapy for Graves' disease during pregnancy: optimal regimen for fetal thyroid status. N Engl J Med 315:24, 1986

24. Davis Le, Lucas MJ, Hawkins GD, et al: Thyrotoxicosis complicating pregnancy. Am J Obstet Gynecol 160:63, 1989

25. Mandel SJ, Larsen PR, Seely EW, et al: Increased need for thyroxine during pregnancy in women with primary hypothyroidism. N Engl J Med 323:91, 1990

26. Casey BM, Dashe JS, Wells E, et al: Subclinical hypothyroidism and pregnancy outcomes. Obstet Gynecol 105:239, 2005

27. Haddow JE, Palomaki GE, Allen WC, et al: Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. N Engl J Med 341:549, 1999

28. Stagnaro-Green A: Postpartum thyroiditis. Best Pract Res Clin Endocrinol Metab 18:303, 2004

29. Bates SM: Treatment and prophylaxis of venous thromboembolism during pregnancy. Thromb Res 108:97, 2003

30. American College of Obstetricians and Gynecologists. ACOG Committee Opinion: safety of Lovenox in pregnancy. Obstet Gynecol 100:845, 2002

31. Sephton V, Farquharson RG, Topping J, et al: A longitudinal study of maternal dose response to low molecular weight heparin in pregnancy. Obstet Gynecol 101:1307, 2003

32. Kupferminc MJ, Eldor A, Steinman N, et al: Increased frequency of genetic thrombophilia in women with complications of pregnancy. N Engl J Med 340:9, 1999

33. Infante-Rivard C, Rivard GE, Yotov WV, et al: Absence of association of thrombophilia polymorphisms with intrauterine growth restriction. N Engl J Med 347:19, 2002

34. Clark SL: Asthma in pregnancy. National Asthma Education Program Working Group on Asthma and Pregnancy. National Institutes of Health, National Heart, Lung and Blood Institute. Obstet Gynecol 82:1036, 1993

35. Dombrowski MP, Schatz M, Wise R, et al: Asthma during pregnancy. Obstet Gynecol 103:5, 2004

36. Managing Asthma During Pregnancy: Recommendations for Pharmacologic Treatment—Update 2004. NIH Publication Number 05-3279. National Institutes of Health, Washington, DC, 2005

37. Brown ZA, Wald A, Morrow RA, et al: Effect of serologic status and cesarean delivery on transmission rates of herpes simplex virus from mother to infant. JAMA 289:203, 2003

38. Corey L, Wald A, Patel R, et al: Once-daily valacyclovir to reduce the risk of transmission of genital herpes. Valacyclovir HSV Transmission Study Group. N Engl J Med 350:11, 2004

39. Young NS, Brown KE: Parvovirus B19. N Engl J Med 350:586, 2004

40. Harger JH, Adler SP, Koch WC, et al: Prospective evaluation of 618 pregnant women exposed to parvovirus B19: risks and symptoms. Obstet Gynecol 91:413, 1998

41. Watts DH: Management of human immunodeficiency virus infection in pregnancy. N Engl J Med 346:1879, 2002

42. Connor EM, Sperling RS, Gelber R, et al: Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. N Engl J Med 331:1173, 1994

43. The mode of delivery and the risk of vertical transmission of human immunodeficiency virus type 1. The International Perinatal HIV Group. N Engl J Med 340:977, 1999

44. Van de Perre P: Postnatal transmission of human immunodeficiency virus type 1: the breast-feeding dilemma. Am J Obstet Gynecol 173:483, 1995

45. American College of Obstetricians and Gynecologists, ACOG Committee Opinion: number 279, December 2002. Prevention of early-onset group B streptococcal disease in newborns. Obstet Gynecol 100:1405, 2002

46. Jones DC, Hayslett JP: Outcome of pregnancy in women with moderate or severe renal insufficiency. N Engl J Med 335:226, 1996

47. Davison JM: Dialysis, transplantation and pregnancy. Am J Kidney Dis 17:127, 1991

48. Cervera R, Font J, Carmona F, et al: Pregnancy outcome in systemic lupus erythematosus: good news for the new millennium. Autoimmun Rev 1:354, 2002

49. Brucato A, Frassi M, Franceschini F, et al: Risk of congenital complete heart block in newborns of mothers with anti-Ro/SSA antibodies detected by counterimmunoelectrophoresis: a prospective study of 100 women. Arthritis Rheum 44:1832, 2001

50. Shinohara K, Miyagawa S, Fujita T, et al: Neonatal lupus erythematosus: results of

maternal corticosteroid therapy. Obstet Gynecol 93:952, 1999

51. Cowchock FS, Reece EA, Balaban D, et al: Repeated fetal losses associated with antiphospholipid antibodies: a collaborative randomized trial comparing prednisone with low-dose heparin treatment. Am J Obstet Gynecol 166:1318, 1992

52. Cardonick E, Iacobucci A: Use of chemotherapy during human pregnancy. Lancet Oncol 5:283, 2004

53. Holmes LB, Harvey EA, Coull BA, et al: The teratogenicity of anticonvulsant drugs. N Engl J Med 344:1132, 2001

54. Richmond JR, Krishnamoorthy P, Andermann E, et al: Epilepsy and pregnancy: an obstetric perspective. Am J Obstet Gynecol 190:371, 2004

55. Confavreux C, Hutchinson M, Hours MM, et al: Rate of pregnancy-related relapse in multiple sclerosis. Pregnancy in Multiple Sclerosis Group. N Engl J Med 339:285, 1998

56. National Institute for Clinical Excellence: Guidance on the use of nicotine replacement therapy (NRT) and bupropion for smoking cessation. Technology Appraisal Guidance No 39. National Institute for Clinical Excellence, London, 2002

57. Singer LT, Arendt R, Minnes S, et al: Cognitive and motor outcomes of cocaine-exposed infants. JAMA 287:1952, 2002

58. Effective medical treatment of opiate addiction. National Consensus Development Panel on Effective Medical Treatment of Opiate Addiction. JAMA 280:1936, 1998

59. Fetal alcohol syndrome and alcohol-related neurodevelopmental disorders. American Academy of Pediatrics, Committee on Substance Abuse and Committee on Children With Disabilities. Pediatrics 106:358, 2000

60. Eskenazi B: Caffeine: filtering the facts. N Engl J Med 341:1688, 1999

61. Mullally BA, Hansen WF: Intrahepatic cholestasis of pregnancy: review of the literature. Obstet Gynecol Surv 57:47, 2002

62. Palma J, Reyes H, Ribalta J, Hernandez I, et al: Ursodeoxycholic acid in the treatment of cholestasis of pregnancy: a randomized, double-blind study controlled with placebo. J Hepatol 27:1022, 1997

63. Weigel RM, Weigel MM: Nausea and vomiting of early pregnancy and pregnancy outcome: a meta-analytical review. Br J Obstet Gynaecol 96:1312, 1989

64. Vutyavanich T, Wongtra-ngan S, Ruangsri R, et al: Pyridoxine for nausea and vomiting of pregnancy: a randomized, double-blind, placebo-controlled trial. Am J Obstet Gynecol 173:881, 1995

65. American College of Obstetricians and Gynecologists: ACOG (American College of Obstetrics and Gynecology) Practice Bulletin: nausea and vomiting of pregnancy. Obstet Gynecol 103:803, 2004

66. Maynard SE, Min Y, Merchan J, et al: Excess placental soluble fms-like tyrosine kinase (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. J Clin Invest 111:649, 2003

67. American Diabetes Association Position Statement: Gestational diabetes. Diabetes Care 26:S103, 2003

68. Crowther C, Hiller JE, Moss JR, et al: Effect of treatment of gestational diabetes on pregnancy outcomes. N Engl J Med 352:2477, 2005

69. Langer O, Conway DL, Berkus MD, et al: A comparison of glyburide and insulin in women with gestational diabetes mellitus. N Engl J Med 343:1134, 2004

70. Kim C, Newton KM, Knopp RH: Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 25:1862, 2002

71. Buchanan TA, Xiang AH, Peters RK, et al: Preservation of pancreatic beta-cell function and prevention of type 2 diabetes by pharmacological treatment of insulin resistance in high-risk Hispanic women. Diabetes 51:2796, 2002

X ENDOMETRIOSIS

Robert L. Barbieri, m.d

Definition and Pathophysiology

Endometriosis is a condition in which tissue resembling endometrial glands or stroma occurs outside the uterus. Endometriosis lesions are most often found in the pelvis. Common sites are the peritoneal surfaces posterior to the uterus; the ovary; the peritoneal surfaces anterior to the uterus; the bowel; the bladder; and the appendix. Rarely, endometriosis lesions occur at sites outside the pelvis, such as the respiratory diaphragm.

Endometriosis lesions are heterogeneous, ranging from 1 mm superficial peritoneal lesions to 4 cm deeply invasive lesions in the rectovaginal septum. Ovarian lesions of endometriosis can grow to 4 to 10 cm in size, necessitating surgical resection. Endometriosis lesions undergo cycles of growth and bleeding in tandem with the menstrual cycle. Intraperitoneal bleeding from the lesions elicits an inflammatory response in the pelvis that is associated with pain and infertility.

Epidemiology

Many authorities believe that approximately 5% of women between 15 and 45 years of age have endometriosis.[1] The precise incidence is difficult to determine because there is no inexpensive, highly reliable method for diagnosing endometriosis. The current gold standard for the diagnosis of endometriosis is surgical visualization of endometriosis lesions, usually by laparoscopy, and so (as with any disease that requires expensive and invasive procedures for diagnosis) a significant number of cases may be missed.[2] Collection of definitive data would require selecting a random sample of women and performing laparoscopy on them to determine whether they have endometriosis; understandably, no such study has been done.

Endometriosis is rare before menarche and after menopause, when estrogen production is low. Most cases of endometriosis are diagnosed in women in their 20s who have never had a child. Full-term pregnancy and delivery appear to markedly reduce the risk of developing endometriosis. Multiple full-term pregnancies further reduce the risk. Long periods of amenorrhea (for example, the amenorrhea of athletes) is associated with a reduced risk of endometriosis, as is aerobic exercise for more than 7 hours a week.

Pathogenesis

The pathogenesis of endometriosis lesions involves mechanical, hormonal, immunologic, and genetic factors. The prominence of particular factors may vary from case to case; indeed, it is possible that endometriosis comprises several different diseases with a common clinical outcome.

mechanical factors

In women with a normal uterus, 99.9% of menstrual blood flow occurs in an antegrade direction—that is, from the endometrium through the cervix and into the vagina. Numerous clinical observations as well as experiments in laboratory animals indicate that anatomic changes, such as cervical stenosis, that hinder antegrade flow are associated with an increased risk of endometriosis. In women with cervical stenosis, the relative obstruction at the level of the cervix causes blood to flow from the uterus back through the fallopian tubes and into the peritoneal cavity. This retrograde menstrual flow contains blood, growth factors, and viable bits of endometrial tissue. The greater the amount of retrograde blood flow, the higher the risk of endometriosis. For example, about 80% of women with congenital cervical stenosis and a functioning endometrium will develop endometriosis. Epidemiologic studies suggest that more prolonged menstrual flow (> 8 days) and more frequent menses (cycle length < 27 days) are also associated with an increased risk of endometriosis.[3]

hormonal factors

Steroid hormones control the growth and function of endometriosis lesions. Estradiol stimulates growth, and androgens cause atrophy of endometriosis lesions [*see Table 1*]. High doses of progestins induce terminal differentiation in endometriosis lesions, a process called pseudodecidualization. Once endometriosis tissue undergoes pseudodecidualization, it can no longer grow. The reason pregnancy reduces the risk of endometriosis is probably that the extremely high progesterone levels that occur in pregnancy cause pseudodecidualization of endometriosis lesions.

Organochlorine chemicals (e.g., dioxin) can disrupt steroid metabolism; exposure to these pollutants has been proposed as a factor in the development of endometriosis. In animal models, dioxin has been found to increase the incidence and severity of endometriosis,[4] possibly by interfering with the action of progesterone,[5] but the effect in humans has yet to be confirmed.

Table 1 Effects of Different Steroids on Endometrium and Endometriosis Lesions

Steroid	*Effect on Endometrium*	*Effect on Endometriosis Lesions*
Estrogen	Growth	Growth
Androgen	Atrophy	Atrophy
Progesterone at physiologic concentrations	Differentiation and secretory changes	No effect on lesions that have no progesterone receptors; differentiation and secretory changes in lesions with progesterone receptors
Progesterone at high concentrations	Decidualization	Pseudodecidualization (a terminal differentiation step)

immunologic factors

Numerous studies indicate that in women with endometriosis, the pelvic peritoneal environment is immunologically abnormal, with increased concentrations of white blood cells, cytokines, and growth factors. Indeed, elevated levels of cytokines—specifically, tumor necrosis factor–α in peritoneal fluid and interleukin-6 in serum—have been proposed as a potential diagnostic marker for endometriosis.[6] One group of researchers has found an increased incidence of autoimmune disease in women with endometriosis—a finding that supports the concept that immunologic abnormalities play a role in the development of endometriosis.[7]

Some authorities believe that in women with endometriosis, a primary immunologic abnormality prevents the clearance, from the peritoneal environment, of the endometrial tissue fragments deposited by retrograde menstruation.[8] This postulated primary alteration in the immune response allegedly contributes to the development of endometriosis. Other authorities believe that the observed peritoneal immunologic changes are not a cause of endometriosis but a consequence of it: the endometriosis lesions produce a chronic pelvic inflammation, which leads to an increase of immune cells in the peritoneal fluid. Interestingly, factors secreted by these immune cells appear to promote angiogenesis and cause endometriosis lesions to grow. It is likely that there is cross-talk between the immune system and endometriosis lesions: endometriosis lesions cause inflammation, inducing immune cells to enter the peritoneal environment; in turn, immune cells secrete factors that can stimulate the growth of endometriosis lesions.

genetic abnormalities

The risk of endometriosis is approximately doubled in first-degree relatives of women with endometriosis.[9] The heritable aspects of endometriosis may involve alterations in the immune response that predispose women to ectopic transplantation and survival of endometrial tissue.

Ovarian endometriosis cysts (endometriomas) are monoclonal and appear to arise from a somatic mutation in a precursor cell, although those mutations have not been characterized.[10] This finding suggests that a small number of genes play a central role in the pathogenesis of endometriosis.

endometriosis and infertility

An association between endometriosis and infertility in women has long been noted,[11] and many possible mechanisms for the infertility have been identified. Nevertheless, the hypothesis that endometriosis decreases fertility has not been definitively proved by consistent data from rigorous studies.

In advanced endometriosis, infertility can have an anatomic cause: adhesions interfere with the release of the ovum from the ovary and its uptake into the fallopian tube. Although women with early-stage endometriosis often have reduced fertility, a causal link between the endometriosis and the infertility is not clear.

Abnormalities in peritoneal, tubal, and endometrial function caused by endometriosis may inhibit fertility, especially in women with early-stage disease.[11] Numerous investigators have reported peritoneal abnormalities in women with endometriosis, including an increased volume of peritoneal fluid[12] and increased concentrations of activated macrophages,[13] prostaglandin, interleukin-1, tumor necrosis factor, and proteases.[14] Peritoneal fluid from women with advanced endometriosis appears to inhibit sperm function, thereby possibly reducing fertility.[15]

A few investigators have reported that women with endometriosis may have increased levels of antiendometrial antibodies, which may impair endometrial function.[16,17] Some women with early-stage endometriosis have luteal phase dysfunction,[18] abnormal follicle growth,[19] multiple premature luteinizing hormone surges,[20] and luteinized unruptured follicle syndrome.

Intrauterine endometrium may be abnormal in women with endometriosis, which suggests the possibility of a so-called field defect in the müllerian tract. Significant suppression of β_3 integrin has been reported in the endometrium of women with early-stage endometriosis.[21] This decrease in β_3 integrin expression may be associated with an impaired interaction of the embryo with the endometrium. In addition, elevated levels of the müllerian antigen CA-125 have been found on endometrial biopsies taken during the luteal phase of the menstrual cycle from women with advanced endometriosis[22] and in the menstrual discharge of women with endometriosis.[23]

Diagnosis

Although endometriosis is a common disorder, it remains remarkably difficult to diagnose. In one cohort study, women with endometriosis reported that, on average, 4 years elapsed between their first presentation with symptoms caused by endometriosis and their diagnosis.

clinical presentation

Women with endometriosis typically present because of chronic pelvic pain or infertility. Other possible symptoms include secondary dysmenorrhea, dyspareunia, pain with bowel movements (dyschezia), and pelvic pain not associated with menses. The rare cases of diaphragmatic endometriosis have been associated with chest pain at the onset of menstruation.[24]

physical examination

In most women with endometriosis, the physical examination is normal. However, certain findings on physical examination suggest the presence of endometriosis. These include tender, thickened, or nodular uterosacral ligaments and fixed adnexal masses. A retroverted, fixed uterus suggests involvement of the cul-de-sac with endometriosis.

The uterosacral ligaments connect the base of the uterus to the sacrum. Nodularity of the ligaments is evident on bimanual pelvic examination as pea-sized nodules palpable at 4 o'clock and 8 o'clock at the base of the cervix. These nodules most often are implants of endometriosis.

Two less common physical findings in endometriosis are cervical stenosis[25] and lateral displacement of the cervix. Lateral displacement of the cervix occurs when one uterosacral ligament becomes severely involved with endometriosis, shortens as a result of scarring, and pulls the cervix to the side.[26]

noninvasive laboratory tests

A complete blood count, urinalysis, and endocervical cultures for gonococci and *Chlamydia* should be performed to rule out infectious causes of pelvic pain in women. Results of all these tests will be normal in women with endometriosis. In most women with endometriosis, the pelvic sonogram is normal, but other conditions, such as uterine leiomyomas, will be evident on sonography. Although many conditions can cause

adnexal masses, including dermoids (mature teratomas), serous and mucinous cysts, and hemorrhagic corpora lutea, endometriomas have classic characteristics on ultrasound, which aids in their diagnosis.

surgical staging

The current gold standard for the diagnosis of endometriosis is the surgical visualization of lesions, usually by laparoscopy. The normal peritoneal surface is smooth and glistening, like the inner surface of the oral mucosa. Classic endometriosis lesions are often black, purple, or red and measure 1 to 5 mm in diameter; they stud the surface of the peritoneum. Atypical endometriosis lesions are often translucent or yellow, and they may take the form of either flat plaques or vesicles.

Unfortunately, surgeons vary considerably in their ability to detect endometriosis lesions reliably. One study reported pathologic confirmation rates of visually diagnosed endometriosis at 42%, 65%, and 76% for three different surgeons.[27]

Endometriosis is staged surgically using the American Society of Reproductive Medicine staging system. This system divides the disease into four stages: stage I, minimal; stage II, mild; stage III, moderate; and stage IV, severe. As with detection, however, staging is not always performed consistently. Studies of intersurgeon and intrasurgeon variability in the staging of endometriosis report low reproducibility and a kappa coefficient in the range of 0.28.[28]

histologic diagnosis

Biopsy and histologic analysis of lesions found on laparoscopy may enable more reliable diagnosis of endometriosis than does visual inspection alone. The criteria for histologic diagnosis of endometriosis include the presence of one of the following components: (1) both endometrial glands and stroma; (2) glandular epithelium with hemosiderin; or (3) endometrial stroma–like tissue with hemosiderin. One weakness of histologic diagnosis for endometriosis is that diagnostic criteria vary among pathologists.[29,30] Furthermore, no study has demonstrated high interobserver reproducibility in the histologic diagnosis of endometriosis.

clinical diagnosis

An innovative approach to the diagnosis of endometriosis is to use a combination of history, physical examination, and noninvasive laboratory testing.[31] This approach is called clinical diagnosis.

differential diagnosis

Pelvic Pain

Chronic pelvic pain, defined as the presence of pain below the umbilicus for more than 6 months, is a common gynecologic problem. In one study of primary care practices that included 284,162 women 12 to 70 years of age, the reported prevalence of chronic pelvic pain was 3.8%.[32] Along with endometriosis, other common gynecologic causes of chronic pelvic pain include chronic pelvic inflammatory disease, adenomyosis, and uterine leiomyomata. Nongynecologic diseases such as irritable bowel syndrome and fibromyalgia, as well as psychiatric diseases such as somatization, may also contribute to chronic pelvic pain. In populations in which the prevalence of sexually transmitted diseases is low, endometriosis is the most common cause of chronic pelvic pain. In three large studies, 70% to 80% of women with chronic pelvic pain had endometriosis as the cause.[31,33,34]

Infertility

Endometriosis is considered to be responsible for 8% of all cases of infertility. The most common causes of infertility, accounting for about 75% of cases, are ovulatory disorders, tubal disease, and semen abnormalities. Miscellaneous factors, such as cervical or immunologic abnormalities and uterine synechiae, cause 2% of cases; 15% are unexplained.[35-37]

Treatment of Pelvic Pain

Interventions that reduce estradiol production are the most reliable way to cause atrophy of endometriosis lesions and are the most effective in treating pain symptoms. A variety of hormonal and surgical interventions are available for this purpose. Most authorities recommend a stepwise approach to the use of these interventions [*see Table 2*].

Table 2 Stepwise Treatment of Pelvic Pain

Step	*Description*	*Recommendation*
1	Thorough history and physical examination	Detailed history and physical examination forms for evaluating pelvic pain are available on the Internet at www.pelvicpain.org
2	Noninvasive laboratory testing	Pelvic ultrasound, complete blood count, urinalysis, endocervical cultures for gonococci and *Chlamydia*
3	Empirical therapy	Oral contraceptive plus nonsteroidal anti-inflammatory medication
4	Surgical diagnostic procedure	Laparoscopy to determine the cause of pain if empirical therapy does not result in sufficient relief of pain
5	GnRH agonist therapy	For regimens, see Table 3
6	GnRH agonist therapy plus steroid add-back	Consider for reduction of GnRH agonist side effects; for regimens, see Table 4
7	Progestin-only treatment	If GnRH agonists cannot be tolerated because of side effects; for regimens, see Table 5

GnRH—gonadotropin-releasing hormone

hormonal treatment for relief of pain

Randomized clinical trials have demonstrated that combination estrogen-progestin oral contraceptives, gonadotropin-releasing hormone (GnRH) agonist analogues, danazol, and progestins are all effective in relieving pelvic pain caused by endometriosis. GnRH agonist analogues are the most effective; combination estrogen-progestin oral contraceptives are the least expensive.

Combination Estrogen-Progestin Oral Contraceptives

Oral contraceptives are sometimes effective in the treatment of pelvic pain caused by endometriosis because progestins can block the growth of endometrium and endometriosis lesions. Although estrogen stimulates the growth of endometriosis lesions, modern oral contraceptives are progestin dominant and contain low doses of estrogen.

In the United States, almost all women with chronic pelvic pain are initially treated empirically with a combination of cyclic oral contraceptives and nonsteroidal anti-inflammatory drugs (NSAIDS), such as ibuprofen. In contrast, in some European countries, the standard practice is to perform laparoscopy on women with chronic pelvic pain to determine the cause of the pain before starting hormonal treatment. In one randomized study, women with endometriosis who had not previously undergone hormonal treatment were randomized to receive treatment with either low-dose cyclic oral contraceptives or a GnRH agonist analogue. Both groups experienced significant improvement in pelvic pain and dysmenorrhea. However, the group treated with GnRH agonists had better relief of dyspareunia.[38]

Oral contraceptives can be used in monthly cycles or long-cycle regimens. If a regimen of oral contraceptives taken in monthly cycles does not relieve the pain, many physicians will try a regimen of long-cycle oral contraceptives. In long-cycle regimens, the active pills are taken for 42 to 105 days in a row; no pills are taken for a period of 1 week between cycles.

If oral contraceptives and NSAIDs fail to relieve chronic pelvic pain, most physicians recommend laparoscopy to definitively determine whether endometriosis is present.

GnRH Agonist Analogues

Several GnRH agonists have been approved for use in endometriosis [*see Table 3*]. These agents are analogues of the native decapeptide GnRH, with substitutions in amino acids 6 and 10. The introduction of D-amino acids at position 6 of native GnRH produces GnRH analogues that are resistant to degradation by endopeptidases and have long half-lives, high affinity for the GnRH receptor, and long receptor occupancy.

Paradoxically, initial treatment with a GnRH agonist analogue stimulates the secretion of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). Prolonged treatment, however, suppresses gonadotropin secretion through the cellular processes of downregulation and desensitization. The suppression of secretion is greater for LH than for FSH. The suppression of pituitary gonadotropin secretion results in suppression of ovarian follicle growth and a 95% decrease in estrogen production. In women treated with many GnRH analogues, the circulating estradiol concentration is suppressed to about 15 pg/ml, which is in the range observed in menopausal women. In essence, this therapy constitutes a reversible medical oophorectomy.

Numerous clinical trials have demonstrated that approximately 85% of women with endometriosis and pelvic pain who are treated with GnRH agonist analogues experience relief of their pain. In one placebo-controlled trial, treatment with a GnRH agonist resulted in better relief of pelvic pain than the administration of placebo (85% and 30%, respectively).[39]

Table 3 GnRH Agonists Approved for the Treatment of Endometriosis*

GnRH Agonist	*Dose*
Leuprolide acetate depot	3.75 mg I.M. every 4 wk
Goserelin acetate	3.6 mg subcutaneous implant every 4 wk
Nafarelin acetate	200 µg twice daily as a nasal spray

*Note: In the United States, GnRH agonist therapy is approved for 6 mo as single-agent therapy and for 1 yr when used in combination with a steroid add-back.

GnRH Agonist Analogues plus Steroid Add-Back

GnRH agonist treatment is associated with hypoestrogenic side effects such as vasomotor symptoms (hot flashes), decreased libido, dry vagina, and decreased bone density. Recent trials have demonstrated that use of a steroid (either high-dose progestin or very low dose estrogen) in so-called add-back therapy can minimize these side effects. GnRH agonist treatment combined with low-dose steroid add-back causes atrophy of endometriosis lesions and improves pelvic pain while minimizing hypoestrogenic vasomotor symptoms and bone loss. In one clinical trial, women with endometriosis and chronic pelvic pain were randomized to four different hormone treatment groups: GnRH agonist alone, GnRH agonist plus progestin only (norethindrone, 5 mg daily), GnRH agonist plus low-dose estrogen-progestin (conjugated equine estrogen, 0.625 mg daily, plus norethindrone acetate, 5 mg daily), or GnRH agonist plus high-dose estrogen plus progestin (conjugated equine estrogen, 1.25 mg daily, plus norethindrone acetate, 5 mg daily). All women were treated with the GnRH agonist leuprolide acetate, given in a depot injection of 3.75 mg I.M. every 4 weeks for 1 year. The rate of treatment discontinuance because of continuing pain was significantly higher in the group that received the combination of GnRH agonist and high-dose estrogen than in any of the other treatment groups.[40] The high-dose estrogen probably stimulated continuing function of the endometriosis implants. Consequently, treatment with a combination of GnRH agonist and high-dose estrogen is not recommended for most women with endometriosis and pelvic pain. The women in the three other groups experienced similar decreases in their pelvic pain, suggesting that all three regimens are effective. Bone density de-

Table 4 Steroid Hormone Regimens for Pelvic Pain from Endometriosis

Regimen	*Comments*
Transdermal estradiol patch, 25 µg daily, plus medroxyprogesterone acetate, 2.5 mg daily[27]	Does not completely prevent bone loss; achieves estradiol concentration in the range of 30 pg/ml
Norethindrone acetate, 5 mg daily[26]	A high dose of progestin; may be associated with symptoms such as bloating and mood changes
Conjugated equine estrogen, 0.625 mg daily, plus norethindrone, 5 mg daily[26]	Preserves bone density and markedly reduces vasomotor symptoms

creased significantly in the women who received the GnRH agonist alone. Bone density was preserved in the groups that were treated with a combination of a GnRH agonist and steroid add-back therapy, and vasomotor symptoms were significantly reduced. This study and others suggest that an optimal treatment of pelvic pain from endometriosis may involve the use of GnRH agonists to suppress ovarian estrogen production, followed by add-back therapy with low doses of estrogen-progestin or progestin alone [*see Table 4*].

Endometriosis lesions grow when serum estradiol concentration is in the premenopausal range (30 to 300 pg / ml), and they regress when estradiol levels are in the menopausal range (< 20 pg / ml). An important clinical question is, What concentration of estradiol will minimize the growth of endometriosis implants but not cause severe hypoestrogenic side effects? Treatments that achieve estradiol levels in the range of 20 to 30 pg / ml are associated with amenorrhea and regression of endometriosis lesions. In addition, these treatments are associated with fewer side effects than treatments that target estradiol levels to less than 20 pg / ml.[41]

Danazol

The first hormonal treatment of endometriosis was the intramuscular administration of testosterone. High-dose parenteral testosterone therapy was demonstrated to cause regression in endometriosis lesions. Unfortunately, many women became virilized by this treatment. Androgen treatment of endometriosis was resurrected after the development of synthetic oral androgens, such as danazol, which had attenuated androgen properties.[42]

Randomized clinical trials that have directly compared danazol and the GnRH agonists have demonstrated that both treatments improve pelvic pain in approximately 85% of treated women.[43] The side effects of these two treatments are very different. The main side effects of the GnRH agonists are those associated with hypoestrogenism (see above). The main side effects of danazol are weight gain (on average, approximately 4 kg at doses of 800 mg / day), muscle cramps, decrease in breast size, oily skin, and hirsutism.[44] In the United States, these side effects have limited the use of danazol for the treatment of endometriosis.[45] Many of the side effects of danazol are dose dependent. Doses of 50, 100, and 200 mg daily can be effective in relieving pelvic pain caused by endometriosis and are associated with less severe side effects than daily doses of 400 or 800 mg. Doses of danazol of less than 400 mg / day do not reliably suppress ovulation. Danazol crosses the placenta and is a known teratogen, so patients who are taking low doses of danazol must use a reliable method of contraception.

Progestins

High-dose synthetic progestins have been demonstrated to be effective in the treatment of pelvic pain in women with endometriosis [*see Table 5*]. These agents have multiple mechanisms of action: (1) suppression of LH and FSH secretion, which suppresses estradiol production; (2) direct antiestrogenic effects on endometriosis lesions; and (3) induction of pseudodecidualization. A problem with progestin treatment is that many women gain weight or experience symptoms typical of the premenstrual period, such as mood changes and bloating.

Table 5 Progestins Effective for Single-Agent Treatment of Endometriosis

Progestin	*Dose*
Norethindrone acetate	5 mg p.o. daily
Medroxyprogesterone acetate	50 mg p.o. daily; 150 mg I.M. every 90 days
Norgestrel	0.075 mg p.o. daily

surgical treatment

Surgical treatment of endometriosis is termed either conservative or definitive. In conservative surgery, all the pelvic organs are preserved; in definitive surgery, both ovaries are removed.

Conservative Surgery

Conservative endometriosis surgery is best accomplished by laparoscopy because postoperative recovery is very rapid, with discharge usually occurring within 1 day. Most surgeons utilize sharp excision to remove endometriosis lesions, electrosurgery to ablate endometriosis lesions, or a combination of the two methods. In one clinical trial, women with pelvic pain caused by endometriosis were randomized to undergo diagnostic laparoscopy and aspiration of peritoneal fluid or to undergo conservative surgery with laparoscopy and resection or ablation of endometriosis lesions. Six months after surgery, 63% of the women treated with surgical resection of endometriosis lesions reported relief of pain, whereas 23% of those treated with diagnostic laparoscopy without surgical resection reported pain relief.[46]

Conservative surgery typically fails to provide permanent relief of endometriosis, however. Within 2 years after surgical treatment, pain recurs in most women with endometriosis.[47] Also, surgical treatment may result in pelvic adhesions, which can become a primary cause of continuing pelvic pain.

Definitive Surgery

Definitive surgery for endometriosis involves removal of both ovaries. Typically, the uterus is removed as well; indeed, in the United States, endometriosis is second only to uterine fibroids as a reason for performing hysterectomy. Many large cohort studies report that about 90% of women with endometriosis and pelvic pain experience long-term relief of their pain through bilateral oophorectomy.[48,49]

After bilateral oophorectomy for endometriosis, patients are typically started on low-dose estrogen replacement. Low-dose estrogen therapy prevents vasomotor symptoms and osteoporosis; pelvic pain usually does not recur.

surgical excision of ovarian masses

Endometriomas require surgical excision if they are causing pain or are enlarging. Large ovarian cysts may be the result of ovarian cancer. Surgical removal of the cyst allows a definitive diagnosis of the cause of the cyst to be made. A randomized clinical trial demonstrated that surgical removal of endometriomas resulted in better long-term results than simple aspiration and fenestration of the cyst.[50]

Treatment of Infertility

early-stage endometriosis and infertility

Women with minimal or mild (stage I or II) endometriosis and infertility have a baseline fecundity of approximately 0.03

Table 6 Stepwise Treatment of Infertility in Early-Stage Endometriosis

Step	*Description*	*Recommendation*
1	Identify and treat all reversible causes of infertility	Proper timing of coitus in relation to ovulation Optimal coital frequency Cessation of cigarette smoking Optimal body mass index Reduce consumption of alcohol and caffeine
2	Laparosocopic surgery to resect endometriosis and remove adhesions	Attempt to restore pelvic anatomy to normal
3	Ovarian stimulation with clomiphene plus intrauterine insemination	Insemination timed to the day before and day of ovulation
4	Ovarian stimulation with gonadotropin injections plus intrauterine insemination	Insemination timed to the day before and day of ovulation; because of increased risk of twin, triplet, and quadruplet pregnancy, some clinicians prefer to skip step 4 and move directly to step 5
5	In vitro fertilization and embryo transfer	—

(3% per cycle, compared with 20% to 36% in normal couples). Numerous randomized studies have demonstrated that a stepwise approach to treatment can increase pregnancy rates in women with early-stage endometriosis [*see Table 6*].

Treatment of Infertility from Other Causes

The first step in the management of early-stage endometriosis and infertility is to identify and treat all reversible causes of infertility in the couple. Many couples have multiple causes of their infertility (e.g., endometriosis in the female partner and a low sperm count in the male partner).

Laparoscopic Surgery

If other causes of infertility have been addressed but the woman is still unable to conceive, the next step is to consider a laparoscopic surgical procedure to ablate or excise endometriosis implants and adhesions and to attempt to restore the pelvis to normal. In one randomized, prospective trial, diagnostic laparoscopy alone was compared with diagnostic laparoscopy combined with surgical resection or ablation of endometriosis in 341 women with early-stage endometriosis. During 36 weeks of postoperative follow-up, fecundity was 0.024 in the diagnosis-only group and 0.047 in the surgically treated group; cumulative pregnancy rates during follow-up were 18% and 31%, respectively.[51]

Intrauterine Insemination

Women who fail to become pregnant after laparoscopic surgery can be treated with intrauterine insemination (IUI) in combination with either clomiphene or gonadotropin injections. Clomiphene is far less expensive than gonadotropins; therefore it is generally used first. These methods are designed to cause multifollicle development and multiple ovulation. In addition, IUI places a large number of motile sperm high in the reproductive tract. Thus, the spermatazoa do not have to travel through the vagina, cervix, and lower portion of the uterus. Both of these methods have been demonstrated to improve pregnancy rates in women with early-stage endometriosis. In one randomized study in 40 women with early-stage endometriosis, fecundity was 0.045 (4.5% per cycle pregnancy rate) in the group that received no treatment and 0.15 in the group treated with three cycles of gonadotropin injections in combination with IUI.[52] Similar findings have been reported by other groups.[53-56]

Table 7 Stepwise Treatment of Infertility in Advanced Endometriosis

Step	*Description*	*Recommendation*
1	Identify and treat all reversible causes of infertility	Proper timing of coitus in relation to ovulation Optimal coital frequency Cessation of cigarette smoking Optimal body mass index Reduce consumption of alcohol and caffeine
2	Surgery to resect endometriosis and remove adhesions	Attempt to restore pelvic anatomy to normal
3	Ovarian stimulation with clomiphene plus intrauterine insemination	Insemination timed to the day before and day of ovulation; limited to patients with patent fallopian tubes and no dense ovarian adhesions
4	Ovarian stimulation with gonadotropin injections plus intrauterine insemination	Insemination timed to the day before and day of ovulation; limited to patients with patent fallopian tubes and no dense ovarian adhesions; because of increased risk of twin, triplet, or quadruplet pregnancy, some clinicians prefer to skip step 4 and move directly to step 5
5	In vitro fertilization and embryo transfer	—

Many authorities believe that the per-cycle pregnancy rate drops significantly after three or four cycles of clomiphene or gonadotropin injections in combination with IUI. Consequently, after three cycles of such treatment, the clinician should review with the couple the advantages of proceeding to the next step, which is in vitro fertilization (IVF) with embryo transfer.[57]

In Vitro Fertilization

There are no prospective, large-scale, randomized trials that demonstrate the efficacy of IVF in the treatment of infertility caused by endometriosis. However, the use of IVF in women with endometriosis and infertility routinely results in treatment-cycle pregnancy rates of approximately 0.30, a 10-fold increase over the baseline fecundity seen in such women.[58-60] It should be noted, however, that the outcome of IVF is highly influenced by the woman's age: women younger than 37 years have much better success with IVF than do women older than 37 years.

advanced endometriosis and infertility

In women with moderate or severe endometriosis and infertility, a stepwise approach to treatment is warranted [*see Table 7*].

Treatment of Infertility from Other Causes

The first step is to identify and correct all other reversible causes of infertility (see above).

Surgical Treatment

The second step is to perform surgical resection for ovarian endometriosis, peritoneal endometriosis, and pelvic adhesions to restore pelvic anatomy and function. There are no randomized, prospective studies that demonstrate the efficacy of surgery in the treatment of advanced endometriosis. However, most authorities believe that surgery improves fertility in these women. One retrospective analysis reviewed the outcome in 130 infertile women with endometriosis who were treated with expectant management, conservative surgery, or expectant management followed by surgery. Although no significant difference was noted between expectant management and surgery in women with mild or moderate endometriosis, women with severe endometriosis appeared to benefit from surgery. Of the 32 women with advanced endometriosis who were observed over 231 months of cumulative follow-up, none became pregnant. Of the 34 women with advanced endometriosis who underwent conservative surgery, 10 became pregnant during 702 cumulative months of follow-up.[61] Similar results have been reported in a meta-analysis of the impact of surgery on fertility in women with endometriosis.[62] These studies suggest that expectant management is not warranted in the treatment of infertility associated with advanced endometriosis and that surgical treatment may improve fecundity.

Pregnancy rates are highest in the 6 to 18 months after the surgical procedure. Additional surgical procedures have not been shown to be effective in increasing fecundity[63]; therefore, if pregnancy does not occur after the first surgery, the clinician should usually move on to intrauterine insemination. Physicians should carefully weigh the limited benefits of second and third operative procedures to enhance fertility against the potential risks of major surgery.

Intrauterine Insemination

Clomiphene or gonadotropin injections in combination with IUI are used empirically in patients with advanced endometriosis and infertility. Most of the clinical trials that have tested these modalities have focused on women with early-stage endometriosis (see above). However, many authorities believe that the benefits of these measures probably extend to women with advanced disease. In patients with severe pelvic adhesions, clinicians may choose to move directly from surgery to in vitro fertilization. Clomiphene or gonadotropins in combination with IUI should not be recommended for women with tubal blockage or dense ovarian adhesions.

In Vitro Fertilization

There are no large, randomized, controlled clinical trials that definitively demonstrate that IVF increases pregnancy rates in women with advanced endometriosis. In one small study involving 21 women with endometriosis and infertility, none of the six women randomized to undergo expectant management became pregnant, whereas five of the 15 women who were treated with IVF became pregnant.[64] Because of the small sample size, however, this study did not have sufficient statistical power to detect true differences between the two groups. One analysis of various infertility treatments demonstrated that for infertile women with advanced endometriosis, rapid progression through the steps to IVF is the most cost-effective treatment approach.[65] In the United States, the median projected cost per IVF cycle in 2001 was $9,226.[66] The cost of having a child with IVF is within the range of the cost of adopting a child. Furthermore, over the past decade, IVF success rates have increased.[67]

IVF is less successful in women with advanced endometriosis who have previously undergone bilateral ovarian surgery; after unilateral oophorectomy and a contralateral ovarian cystectomy, ovarian stimulation is often ineffective, and the pregnancy rate is low. Reduced pregnancy rates for women with advanced endometriosis (compared with women who have early-stage endometriosis or tubal factor infertility) may be the result of a premature depletion of the ovarian follicle pool,[68] abnormal folliculogenesis,[69] or reduced fertilization potential of oocytes.[70]

References

1. Houston DE, Noller KL, Melton LJ, et al: Incidence of pelvic endometriosis in Rochester, Minnesota, 1970–1979. Am J Epidemiol 125:959, 1987

2. Holt VL, Weiss NS: Recommendations for the design of epidemiologic studies of endometriosis. Epidemiology 11:654, 2000

3. Cramer DW, Wilson E, Stillman RJ, et al: The relation of endometriosis to menstrual characteristics, smoking and exercise. JAMA 225:1904, 1986

4. Birnbaum LS, Cummings AM: Dioxins and endometriosis: a plausible hypothesis. Environ Health Perspect 110:15, 2002

5. Bruner-Tran KL, Rier SE, Eisenberg E, et al: The potential role of environmental toxins in the pathophysiology of endometriosis. Gynecol Obstet Invest 48(suppl 1):45, 1999

6. Bedaiwy MA, Falcone T, Sharma RK, et al: Prediction of endometriosis with serum and peritoneal fluid markers: a prospective controlled trial. Hum Reprod 17:426, 2002

7. Sinaii N, Cleary SD, Ballweg ML, et al: Autoimmune and related disease among women with endometriosis: a survey analysis. Fertil Steril 77(suppl 1):S7, 2002

8. Gazvani R, Templeton A: New considerations for the pathogenesis of endometriosis. Int J Gynaecol Obstet 76:117, 2002

9. Simpson JL, Elias S, Malinak LR, et al: Heritable aspects of endometriosis. Am J Obstet Gynecol 137:327, 1980

10. Jimbo J, Hitami Y, Yoshikawa H, et al: Evidence for monoclonal expansion of epithelial cells in ovarian endometrial cysts. Am J Pathol 150:1173, 1997

11. Barbieri RL, Missmer S: Endometriosis and infertility: a cause-effect relationship? Ann N Y Acad Sci 955:23, 2002

12. Haney AF, Muscato JJ, Weinberg JB: Peritoneal fluid cell populations in infertility patients. Fertil Steril 41:122, 1984

13. Halme J, Becker S, Haskill S: Altered maturation and function of peritoneal macrophages: possible role in the pathogenesis of endometriosis. Am J Obstet Gynecol 156:783, 1987

14. Fakih H, Baggett B, Holtz G: Interleukin-1: a possible role in the infertility associated with endometriosis. Fertil Steril 47:213, 1987

15. Oral E, Arici A, Olive DL, et al: Peritoneal fluid from women with moderate or se-

vere endometriosis inhibits sperm motility: the role of seminal fluid components. Fertil Steril 66:787, 1996

16. Badawy SZ, Cuenca V, Stitzel A: Autoimmune phenomena in infertile patients with endometriosis. Obstet Gynecol 63:271, 1984

17. Weed JC, Aquembourg PC: Endometriosis: can it produce an autoimmune response resulting in infertility? Clin Obstet Gynecol 23:885, 1980

18. Cheesman KL, Cheesman SD, Chatterton RT: Alterations in progesterone metabolism and luteal function in infertile women with endometriosis. Fertil Steril 29:270, 1978

19. Wardle PG, McLaughlin EA, McDermott A: Endometriosis and ovulatory disorder: reduced fertilization in vitro compared with tubal and unexplained infertility. Lancet 2:236, 1985

20. Polan ML, Totora M, Caldwell BV, et al: Abnormal ovarian cycles as diagnosed by ultrasound and serum estradiol levels. Fertil Steril 37:342, 1982

21. Lessey BA, Castlebaum AJ, Sawin SW, et al: Aberrant integrin expression in the endometrium of women with endometriosis. J Clin Endocrinol Metab 79:643, 1994

22. McBean JH, Brumsted JR: In vitro CA-125 secretion from women with advanced endometriosis. Fertil Steril 59:89, 1993

23. Takahashi K, Nagata H, Musa AA, et al: Clinical usefulness of CA-125 levels in the menstrual discharge of women with endometriosis. Fertil Steril 54:360, 1990

24. Redwine DB: Diaphragmatic endometriosis: diagnosis, surgical management, and long-term results of treatment. Fertil Steril 77:288, 2002

25. Barbieri RL: Stenosis of the external cervical os: an association with endometriosis in women with chronic pelvic pain. Fertil Steril 70:571, 1998

26. Propst AM, Storti K, Barbieri RL: Lateral cervical displacement: an association with endometriosis. Fertil Steril 70:568, 1998

27. Pardanani S, Barbieri RL: The gold standard for the surgical diagnosis of endometriosis: visual findings or biopsy results. J Gynecol Tech 4:121, 1998

28. Hornstein MD, Friedman AJ, Gleason RE, et al: The reproducibility of the revised American Fertility Society classification of endometriosis. Fertil Steril 59:1015, 1993

29. Jansen RS, Russell P: Non-pigmented endometriosis: clinical, laparoscopic and pathologic definitions. Am J Obstet Gynecol 155:1154, 1986

30. Blaustein A: Pelvic endometriosis. Blaustein's Pathology of the Genital Tract, 2nd ed. Blaustein A, Ed. Springer-Verlag, New York, 1982, p 464

31. Randomized controlled trial of depot leuprolide in patients with chronic pelvic pain and clinically suspected endometriosis. Pelvic Pain Study Group. Obstet Gynecol 93:51, 1999

32. Zondervan KT, Yudkin PL, Vessey MP, et al: Prevalence and incidence of chronic pelvic pain in primary care: evidence from a national general practice database. Br J Obstet Gynaecol 106:1149, 1999

33. Carter JE: Combined hysteroscopy and laparoscopic findings in patients with chronic pelvic pain. J Am Assoc Gynecol Laparoscopy 2:43, 1994

34. Koninckx PR, Lessafre E, Meuleman C, et al: Suggestive evidence that pelvic endometriosis is a progressive disease, whereas deeply infiltrating endometriosis is associated with pelvic pain. Fertil Steril 55:759, 1991

35. Collins JA, Crosignani PG: Unexplained infertility: a review of diagnosis, prognosis, treatment efficacy and management. Int J Gynaecol Obstet 39:267, 1992

36. Templeton AA, Penney GC: The incidence, characteristics, and prognosis of patients whose infertility is unexplained. Fertil Steril 37:175, 1982

37. Guzick DS, Grefenstette I, Baffone K, et al: Infertility evaluation in fertile women: a model for assessing the efficacy of infertility testing. Hum Reprod 9:2306, 1994

38. Vercellini P, Trespidi L, Colombo A, et al: A gonadotropin-releasing hormone agonist versus a low-dose oral contraceptive for pelvic pain associated with endometriosis. Fertil Steril 60:75, 1993

39. Lupron depot (leuprolide acetate for depot suspension) in the treatment of endometriosis: a randomized, placebo-controlled, double-blind study. Lupron Study Group. Fertil Steril 54:419, 1990

40. Leuprolide acetate depot and hormonal add-back in endometriosis: a 12-month study. Lupron Add-Back Study Group. Obstet Gynecol 91:16, 1998

41. Howell R, Edmonds D, Dowsett M, et al: Gonadotropin releasing hormone analogue plus hormone replacement therapy for the treatment of endometriosis: a randomized clinical trial. Fertil Steril 64:474, 1995

42. Barbieri RL, Ryan KJ: Danazol: endocrine pharmacology and therapeutic applications. Am J Obstet Gynecol 141:453, 1981

43. Henzl MR, Corson SL, Moghissi K, et al: Administration of nasal nafarelin as compared with oral danazol for endometriosis. N Engl J Med 318:485, 1988

44. Barbieri RL, Evans S, Kistner RW: Danazol in the treatment of endometriosis: analysis of 100 cases with a 4-year follow-up. Fertil Steril 37:737, 1982

45. Selak V, Farquhar C, Prentice A, et al: Danazol for pelvic pain associated with endometriosis. Cochrane Database Syst Rev (2):CD000068, 2002

46. Sutton CJ, Ewen SP, Jacobs SA, et al: Laser laparoscopic surgery in the treatment of ovarian endometriosis. J Am Assoc Gynecol Laparoscopy 4:319, 1997

47. Hornstein MD, Hemmings R, Yuzpe AA, et al: Use of nafarelin versus placebo after reductive laparoscopic surgery for endometriosis. Fertil Steril 68:860, 1997

48. Ranney B: Endometriosis. 3. Complete operations. Reasons, sequelae, treatment. Am J Obstet Gynecol 109:1137, 1971

49. Hickman TN, Namnoum AB, Hinton EL, et al: Timing of estrogen replacement following hysterectomy with oophorectomy for endometriosis. Obstet Gynecol 91:673, 1998

50. Beretta P, Franci M, Ghezzi F, et al: Randomized clinical trial of two laparoscopic treatments of endometriomas: cystectomy versus drainage and coagulation. Fertil Steril 70:1176, 1998

51. Laparoscopic surgery in infertile women with minimal or mild endometriosis. The Canadian Collaborative Group on Endometriosis. N Engl J Med 337:217, 1997

52. Fedele L, Bianchi S, Marchini M, et al: Superovulation with human menopausal gonadotropins in the treatment of infertility associated with minimal or mild endometriosis. Fertil Steril 58:28, 1992

53. Tummon IS, Asher LJ, Martin JS, et al: Randomized controlled trial of superovulation and insemination for infertility associated with minimal or mild endometriosis. Fertil Steril 68:8, 1997

54. Nulsen JC, Walsh S, Dumez S, et al: A randomized and longitudinal study of human menopausal gonadotropin with intrauterine insemination in the treatment of infertility. Obstet Gynecol 82:780, 1993

55. Chaffkin LM, Nulsen JC, Luciano AA, et al: A comparative analysis of the cycle fecundity rates associated with combined human menopausal gonadotropin and intrauterine insemination versus either hMG or IUI alone. Fertil Steril 55:252, 1991

56. Guzick DS, Carson SA, Coutifaris C, et al: Efficacy of superovulation and intrauterine insemination in the treatment of infertility. N Engl J Med 340:177, 1999

57. Isaksson R, Tiitinen A: Superovulation with combined insemination or timed intercourse in the treatment of couples with unexplained infertility and minimal endometriosis. Acta Obstet Gynecol Scand 76:550, 1997

58. Olivenenes F, Feldberg D, Liu HC, et al: Endometriosis: a stage by stage analysis: role for in vitro fertilization. Fertil Steril 64:392, 1995

59. Oehninger S, Acosta AA, Kreiner D, et al: In vitro fertilization and embryo transfer: an established and successful therapy for endometriosis. J In Vitro Fertilization and Embryo Transfer 5:249, 1988

60. Chillik CF, Acosta AA, Garcia JE, et al: The role of in vitro fertilization in infertile patients with endometriosis. Fertil Steril 44:56, 1985

61. Olive DL, Lee KL: Analysis of sequential treatment protocols for endometriosis-associated infertility. Am J Obstet Gynecol 154:613, 1986

62. Adamson GD, Pasta DJ: Surgical treatment of endometriosis-associated infertility: meta-analysis compared with survival analysis. Am J Obstet Gynecol 171:1404, 1994

63. Pagidas K, Falcone T, Hemmings R, et al: Comparison of reoperation for moderate and severe endometriosis-related infertility with in vitro fertilization-embryo transfer. Fertil Steril 65:791, 1996

64. Soliman S, Daya S, Collins J: A randomized trial of in vitro fertilization versus conventional treatment for infertility. Fertil Steril 59:1239, 1993

65. Phillips Z, Barraza-Llorens M, Posnett J: Evaluation of the relative cost-effectiveness of treatments for infertility in the UK. Human Reproduction 15:95, 2000

66. Collins J: Cost-effectiveness of in vitro fertilization. Semin Reprod Med 19:279, 2001

67. Cramer DW, Liberman RF, Powers DR, et al: Recent trends in assisted reproductive techniques and associated outcomes. Obstet Gynecol 95:61, 2000

68. Hornstein MD, Barbieri RL, McShane PM: The effects of previous ovarian surgery on the follicular response to ovulation induction in an in vitro fertilization program. J Reprod Med 34:277, 1989

69. Toya M, Saito H, Ohta N, et al: Moderate and severe endometriosis is associated with alterations in the cell cycle of granulose cells in patients undergoing in vitro fertilization and embryo transfer. Fertil Steril 73:344, 2000

70. Pal L, Shifren JL, Isaacson KB, et al: Impact of varying stages of endometriosis on the outcome of in vitro fertilization–embryo transfer. J Assisted Reproduction and Genetics 15:27, 1998

XI MENOPAUSE

Susan D. Reed, m.d., m.p.h.
Eliza L. Sutton, m.d.

Definitions

The female reproductive system matures in a continuous, natural process from menarche to menopause, as the finite numbers of oocytes produced during fetal development are gradually lost to ovulation and senescence. Menopause is defined as the permanent cessation of menses[1]; by convention, the diagnosis of menopause is not made until the individual has had 12 months of amenorrhea. Menopause is thus characterized by the menstrual changes that reflect oocyte depletion and subsequent reduction in ovarian hormone production. However, the manifestations that occur around the time of menopause are caused by the underlying ovarian changes, rather than by the cessation of menstruation itself. Therefore, a woman who has undergone a hysterectomy but who retains her ovaries will experience normal menopausal symptoms as oocyte depletion leads to hypoestrogenism, even though cessation of menstruation occurred with surgery.

Natural menopause occurs at or after 40 years of age and has no underlying pathologic cause [*see* Natural Menopause, *below*]. Induced menopause may occur after chemotherapy, pelvic radiation, or, most commonly, bilateral oophorectomy. Menopause is considered premature when it occurs before 40 years of age but is otherwise natural [*see* Premature Ovarian Failure, *below*].

The climacteric, a term now used infrequently, refers to the time of waning ovarian function associated with menstrual irregularity and vasomotor symptoms. Perimenopause is the time between the onset of the climacteric and the year after the last menses. Menopausal transition is replacing perimenopause and climacteric as the preferred term to describe the time of physiologic change around the cessation of ovarian function [*see Figure 1*].[2] Premenopause is the entire reproductive span before onset of the menopausal transition, and postmenopause is the span of life after menopause.

In the past, natural menopause was considered to be an endocrinopathy, with the ovary depicted as a failing organ and estrogen considered the optimal therapy. Given that menopause is a normal transition in the lives of most women and that significant risks have been associated with postmenopausal hormone "replacement," the viewpoint of menopause as an endocrinopathy is no longer espoused.

Natural Menopause

EPIDEMIOLOGY

The menopausal transition, which precedes menopause, has an average duration of 4 years, with a range of 0 to 10 years.[3-5] The mean age at which menopause occurs in developed countries is 51 years[4,6,7] and may be increasing.[8] The standard deviation around this mean is about 2 years.[4,9] Approximately 95% of women experience menopause by 55 years of age.[4] Several factors appear to influence the age at which women experience menopausal symptoms and the final menstrual period; for example, menopause occurs approximately 1 year earlier in smokers[6,7,10] and nulliparous women.[6,7] Menopause may also occur earlier in women who have had ovarian cystectomies or unilateral oophorectomies.[11]

PHYSIOLOGY AND GENETICS OF REPRODUCTIVE AGING

Ovarian follicular depletion, by means of atresia, is the final common pathway in female reproductive aging. At 5 months of fetal age, the ovaries contain their peak number of primordial follicles, totaling approximately two million. At birth, girls have one million primordial follicles, approximately 25% of which remain at puberty. During the reproductive years, many follicles will begin to develop during each ovulatory cycle; all but one, the dominant follicle, become atretic. An estimated 1,000 follicles remain in the ovaries of a woman 51 years of

Stages of reproductive aging	Reproductive Years			Menopausal Transition			Postmenopause	
	Early	Peak	Late	Early	Late*		Early*	Late
				Perimenopause				
Duration of stage	Variable			Variable		1 yr	4 yr	Until demise
Menstrual cycles	Variable to regular	Regular		Variable cycle length (> *7 days different from normal*)	≥ 2 skipped cycles and an interval of amenorrhea (≥ *60 days*)	0	None	
Endocrine function (FSH levels)	Normal		Elevated or normal	Elevated or normal		↑	Elevated	

*Stages most likely to be characterized by vasomotor symptoms.

Final Menstrual Period

Figure 1 **The Stages of Reproductive Aging Workshop (STRAW) reproductive staging system showing the relationship of the final menstrual period with menstrual cycle changes and FSH serum concentrations.[2] (FSH—follicle-stimulating hormone, ↑—elevated)**

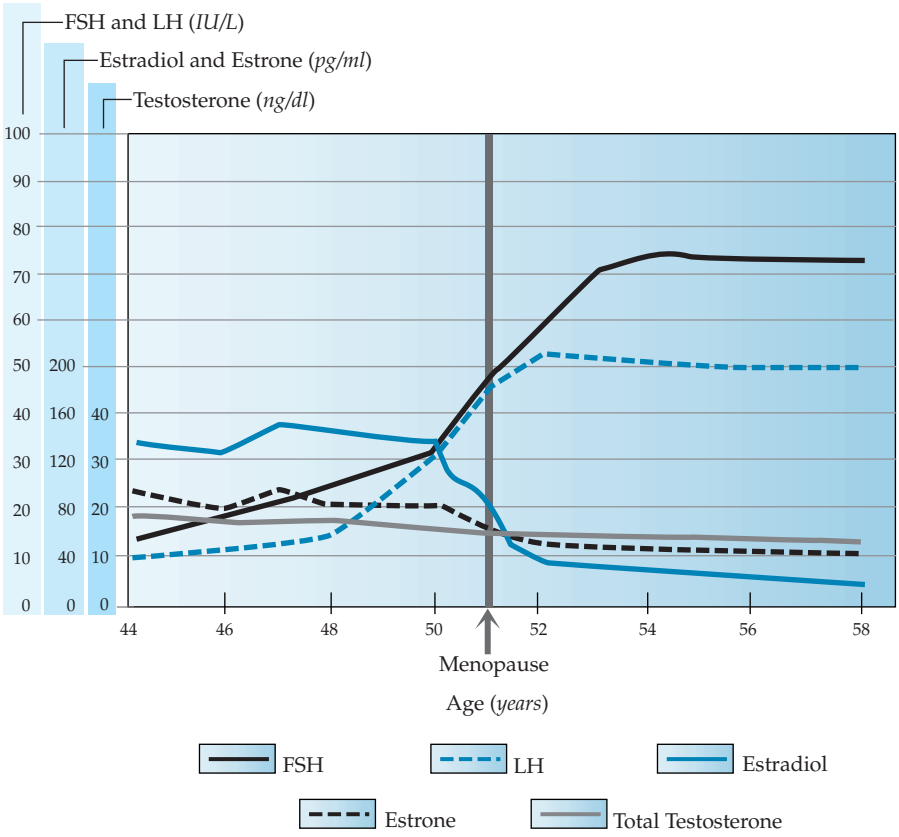


Figure 2 **Approximate average serum concentrations of estradiol, estrone, FSH, LH, and total testosterone during the menopausal transition and postmenopause. A subtle rise in FSH occurs first, followed by a rise in LH and a decline in estradiol and estrone. There are no abrupt changes in testosterone, but a gradual continuous decline occurs that begins before the menopausal transition.[130]**

age.[12] Some poorly responsive follicles persist for a few years after the menopause.[13] This progressive loss of follicles that accompanies aging is characteristic of all mammals studied to date; however, the controlling factors for this process have not been well defined.

Beginning as early as 10 to 15 years before menopause, the length of the menstrual cycle progressively decreases, owing to a shortening of the follicular phase of the cycle. The observed decrease in cycle length continues until the onset of the menopausal transition, when both the average cycle length and the standard deviation of cycle length begin to increase as follicles are depleted and ovulation occurs less frequently.[4,14] Insufficient follicular development results in inadequate estrogen production. With little estrogen available to stimulate the endometrium, amenorrhea results.

There is good evidence that the timing of natural menopause is genetically programmed,[15-17] but the specific genes involved are yet to be well defined. Common allelic variants of the estrogen receptor gene (estrogen receptor–α [*ER*-α] and *ER*-β) contribute to the variability in the timing of menopause.[18] In addition, all of the steroid receptors, as well as the proteins and enzymes involved in steroid biosynthesis and metabolism, are known to be coded by polymorphic sites (genetic changes found in at least 1% of the population). This genetic variability adds to the complexity of the actions and interactions of the reproductive steroids and to the timing and extent of menopausal symptoms.

PHYSIOLOGIC CHANGES IN MENOPAUSE

Hormonal Changes

A subtle rise in the concentration of follicle-stimulating hormone (FSH) is the earliest and most consistent clinically measurable hormonal change noted in studies of reproductive aging.[19,20] An FSH level measured during the early follicular stage of the menstrual cycle that is greater than two standard deviations above the mean level in women of reproductive age is a marker of impending menopausal transition.[2] Luteinizing hormone (LH) levels remain normal initially, but they eventually become elevated as ovarian steroid secretion falls and gonadotropin-releasing hormone (GnRH) increases [*see Figures 2 and 3*]. The early selective increase in FSH appears to be caused by decreased secretion of the hormone inhibin B by the ovarian granulosa cells and is a marker of follicular atresia. Inhibin A and B, hormones that are involved in directing follicular development and were first characterized in the 1990s, suppress pituitary FSH production.[21,22] As anovulation predominates, FSH and LH remain chronically elevated (i.e., there is a 10-fold to 20-fold increase in the FSH level and a threefold to fivefold increase in the LH level),[19,21,23] and estradiol levels fall below 50 pg/ml [*see Figure 2*].

The physiologic changes that are associated with menopause are predominantly reflected by changes in circulating levels of estrogens, androgens, and progesterone [*see Figures 2 and 3*]. The hormonal system is made more complex by fluctuations in steroid hormones that alternate between free and bound states. Sex hormone–binding globulin affects serum levels of all steroid

hormones, binding preferentially to testosterone, estrogen, and progesterone, in that order.

During the reproductive years, estradiol (E_2) is the principal estrogen, both in quantity and in potency; estrone (E_1) is present in a significant amount but is less potent than estradiol. Estriol (E_3), a weak estrogen, is a metabolite of estrone and estradiol. Despite diminished fertility and ongoing follicular atresia, the ovulatory cycles of women in the menopausal transition have normal to high concentrations of circulating estradiol and estrone. In fact, as women approach the menopausal transition, preovulatory estradiol levels can be higher than those seen in younger women.[24,25]

After menopause, estradiol production drops by 90%,[19,21] owing to follicular atresia [*see Figures 2 and 3a*]. What little estradiol is produced after menopause comes primarily from peripheral conversion of estrone. Estrone, the dominant estrogen after menopause, is produced through peripheral conversion of adrenal androstenedione by aromatase, primarily in adipose tissues [*see Figure 3d*]. Fatty breast tissue is a principal site of aromatase activity, but activity is also present in the brain, muscle, liver, and, minimally, the ovary of a postmenopausal woman.

As reproductive aging progresses, serum levels of androgens decrease but not to the extent that estrogen levels diminish. Androstenedione levels drop by approximately 50%,[26-28] ovarian

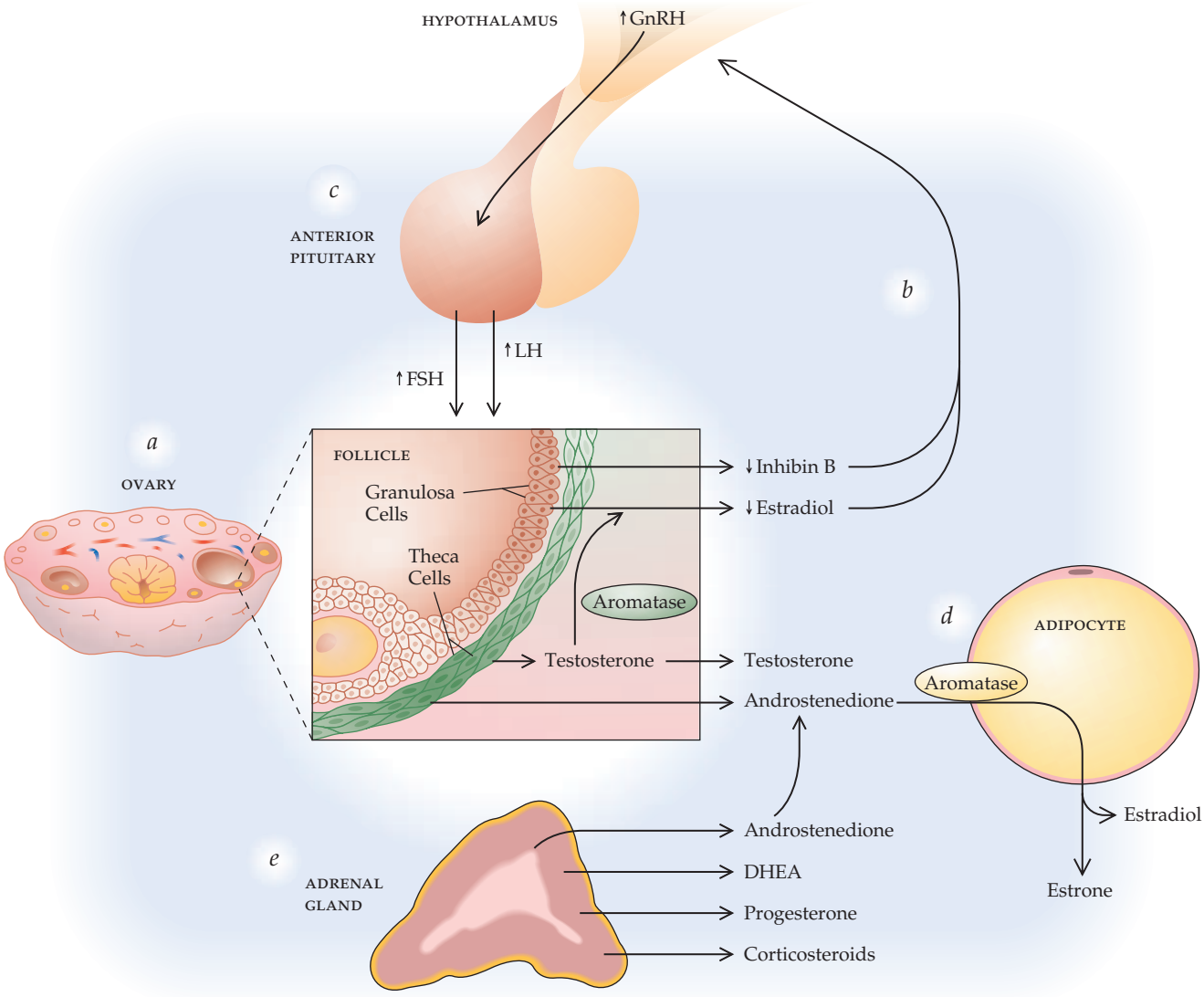


Figure 3 **Multiple hormonal changes are associated with reproductive aging. (*a*) Within the ovary, secretion of inhibin B by granulosa cells decreases when a woman is in her mid-30s, and follicular depletion results in increasing rates of anovulation and diminished ovulatory surges of estradiol and estrone by her early 40s. Ovarian testosterone secretion continues; some ovarian testosterone is converted to estradiol by the enzyme aromatase, and the remainder is secreted as testosterone or the androgen precursor androstenedione. (*b*) In the menopausal transition, decreased circulating levels of inhibin and, subsequently, decreasing estradiol concentrations result in stimulation of the hypothalamus to increase secretion of GnRH. (*c*) Elevated circulating GnRH levels stimulate the anterior pituitary to increase secretion of FSH, followed by an increase in LH. Eventually, attempts by the brain to drive the ovary to produce estrogen fail, but production of androstenedione and testosterone by the ovarian theca cells continues in early menopause. (*d*) With diminished serum estrogen levels, adipocytes are stimulated to convert androstenedione to estrone via the enzyme aromatase. (*e*) Hormonal synthesis by the adrenal gland remains fairly constant, undergoing changes associated with aging, not menopause per se.**

production declines [*see Figure 3a*], and adrenal output remains relatively constant [*see Figure 3e*]. Testosterone decreases by approximately 30% and continues to be secreted by the ovarian stroma, under the influence of LH [*see Figure 3a*].[26-28] Serum concentrations of the adrenal androgen precursor dehydroepiandrosterone (DHEA) decrease with biologic aging, beginning before the final menstrual period [*see Figure 3e*].[26]

During the reproductive years, the principal source of progesterone is the corpus luteum; small concentrations of progesterone continue to be produced by the adrenal gland after the menopause [*see Figure 3e*].

The overall changes in reproductive steroid hormones observed following menopause include the following:

- Negligible estradiol production by the ovary
- A shift from the ovary to the adrenal gland as the primary source of estrogen precursors
- Emergence of estrone as the dominant estrogen
- Continued testosterone production by the ovarian stroma
- An overall increase in the ratio of androgens to estrogens
- A decrease in progesterone levels resulting from anovulation

Target Tissues

During the past decade, remarkable advances in the understanding of steroid biosynthesis, metabolism, and receptor tissue specificity have occurred.[29] At least two estrogen receptors, ER-α and ER-β, and two progesterone receptors, PRA and PRB, have been identified. Estrogen receptors are found in the genitourinary, cardiovascular, and gastrointestinal tracts and in the brain, bone, and integument. Different tissues have a predominance of specific receptors, depending on the individual's endogenous hormonal profile. The complexity of the system leads to variations in the clinical manifestations of reproductive aging. Recent advances in the understanding of these complex physiologic processes have important implications for designing specific targeted therapies and have led to new classifications of pharmaceuticals: the selective estrogen receptor modulators (SERMs). Selective progesterone receptor modulators and selective androgen receptor modulators are also in development [*see* Preventive Health Care, *below*].

Endometrium During normal ovulatory cycles, progesterone, which is produced by the corpus luteum, causes the endometrium to mature to a secretory state. During the menopausal transition, endometrial shedding occurs less frequently because of anovulation, and oligomenorrhea results.[4] Bleeding may be quite heavy in anovulatory cycles because estrogen is still produced, although at diminished levels, and stimulates the endometrium unopposed by progesterone. Furthermore, with increasing anovulation, longer cycles predominate and result in a thicker endometrium. Eventually, as anovulation predominates and estradiol production by the ovary becomes negligible, amenorrhea results.

Genitourinary epithelium The vagina is a principal target tissue for estrogen. Estrogen matures the vaginal epithelium, making it thicker and rugated. The estrogen-stimulated epithelial cells produce more glycogen, which in turn changes the bacterial flora and increases vaginal acidity.[30] Hypoestrogenism results in thinning of the vaginal and vulvar epithelium. The base of the bladder is also derived from müllerian tissue and likewise is estrogen sensitive. Epithelial changes in the bladder are similar to those occurring in the vagina and vulva and result in thin, pale, friable tissues.

Central and sympathetic nervous systems Fluctuations in estrogen levels are associated with hot flushes.[31] Hot flushes are caused by thermoregulatory dysfunction that is most likely initiated by the hypothalamus in response to estrogen withdrawal. Small elevations in core body temperature are followed by peripheral vasodilation. This results in a sensation of warmth and perspiration, both of which occur at a core body temperature that is lower than normal.[32] To be susceptible to hot flushes, a woman needs to have been exposed to reproductive levels of estrogen and then experience estrogen withdrawal. For example, women with Turner syndrome, who never attain reproductive levels of estrogen, do not experience hot flushes.

Alterations in dopamine, norepinephrine, and serotonin pathways[33-35] associated with systemic estrogen fluctuations may contribute to the vasomotor symptoms experienced during the menopausal transition and postmenopause. In addition to hot flushes and diaphoresis, symptoms may include a sense of prickling of the skin, heart palpitations, and anxiety.

Sleep disruption from vasomotor instability can result in insomnia[36] and daytime fatigue, and it may also contribute to mood and other neuropsychiatric changes.[37] The prevalence of sleep-disordered breathing, including snoring and obstructive sleep apnea, increases after menopause,[38,39] most likely because of the estrogen responsiveness of the upper airway musculature.[40] Other pathophysiologic alterations resulting in changes in cognition, mood, and sleep are not as well studied.

Women may experience a decline in libido during the menopausal transition or after the menopause.[41] The contributing factors are complex and may include fatigue or stress (e.g., multiple responsibilities, including caretaking and employment), urogenital atrophy leading to dyspareunia, decreased testosterone levels [*see Figure 2*] (particularly in women who undergo surgical removal of both ovaries), sexual inactivity or dysfunction in a partner, and physical or emotional separation from a partner.

Bone Estrogen suppresses bone resorption. At the menopausal transition, bone resorption exceeds formation, and an accelerated loss in bone mass may occur.[42] Bone mass may be lost at an annual rate of 3% to 5% in the first few years after the final menstrual period, but eventually, this rate of loss slows and continues at 1% to 2% a year.[43] Trabecular bone, the predominant type of bone in the spine, hip, and distal radius, is affected first and to a greater degree than cortical bone [*see 3:VI Diseases of Calcium Metabolism and Metabolic Bone Disease*].

Cardiovascular system Cardiovascular risks[44] and events[45] increase after menopause. Among the factors that contribute to an increased risk of cardiovascular events are the levels of low-density lipoprotein (LDL) cholesterol and apolipoprotein B, which are higher after menopause.[46] Estrogen receptors have been found in the muscularis of arteries in cardiovascular tissue.[29] Estrogens appear to have a direct vasodilatory effect on the coronary artery, mediated by the formation and release of endothelium-derived relaxing factor, reduction of endothelin levels, and the promotion of prostacyclin production.[47]

Coagulation factors Menopause has been associated with increases in factor VII, factor VIII, plasminogen activator inhibitor–1 (PAI-1), and fibrinogen; all of these changes can lead to hypercoagulable states. Conversely, menopause has been associated with increased levels of antithrombin III and activated protein C, which may be beneficial in that these factors diminish

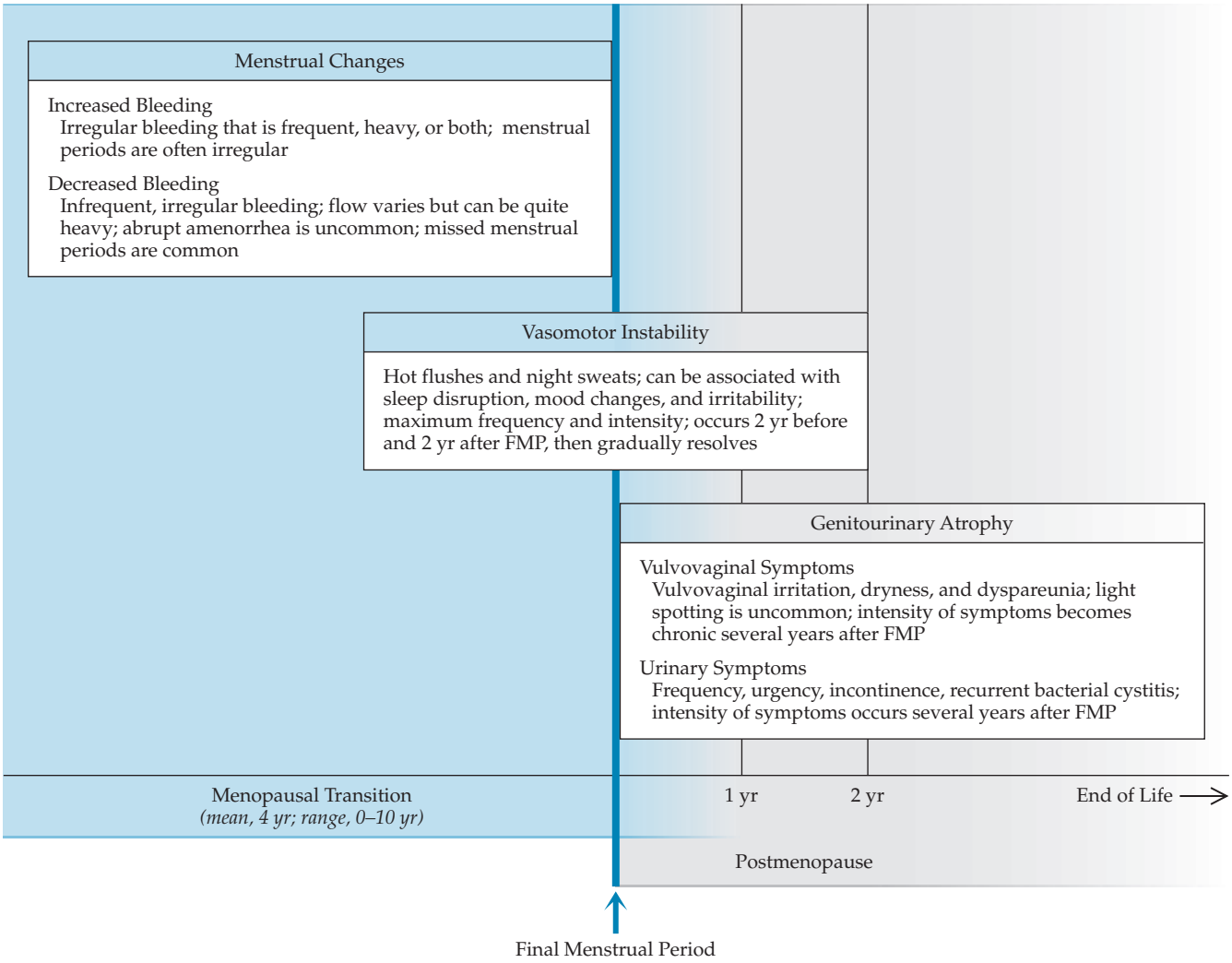


Figure 4 **Characteristic symptoms of the menopausal transition and menopause. Peak vasomotor symptoms occur around the time of the final menstrual period. Menstrual changes are common before the menopause; abrupt amenorrhea preceded by normal cycles is unusual. Genitourinary atrophy is most common in postmenopause.**

coagulation.[48] It is unknown whether these changes are caused by hypoestrogenism alone or by a combination of hormonal changes observed at the time of menopause.

Integument A decrease in the production of dermal collagen and a subsequent reduction in dermal thickness[49] result in significant changes in women's skin, including wrinkles and dryness. In addition, a reduced rate of cutaneous wound healing has been associated with diminished secretion of transforming growth factor–β1 (TGF-β1) by dermal fibroblasts.[50]

Target-sensitive tissues and neoplastic growth Changes in the balance of reproductive hormones at the time of menopause have been associated with increased neoplastic growth in specific tissues. Hormonally sensitive neoplasms of the breast, colon, ovary, endometrium, and myometrium (leiomyoma) are widely recognized. Leiomyomas commonly increase in size during the menopausal transition but diminish in the postmenopausal period, presumably as a result of low levels of estradiol and progesterone. Other less common neoplasms occur in the gastrointestinal tract (esophageal and gastric)[51]; blood vessels; adipose and angiolymphatic tissues (angiomyolipoma,[52] lymphangiomyomatosis)[53]; and the central nervous system (meningioma).[54] All of these neoplastic tissues have been found to have reproductive hormone receptors and appear to be sensitive to steroid hormones. Contrary to previous evidence, recent observations suggest that melanoma is not progesterone sensitive.[55]

DIAGNOSIS

The diagnosis of menopausal transition may be suspected on the basis of symptoms (e.g., menstrual irregularity and vasomotor symptoms) in a woman older than 40 years [*see Figure 4*] well before it can be proven by FSH testing [*see* Laboratory Tests, *below*]. The clinical diagnosis of natural menopause is made if a woman is of an appropriate age and has had 12 months of amenorrhea accompanied by symptoms suggestive of ovarian failure, at which point the FSH serum concentration is so certainly elevated that testing is usually not useful.

Clinical Manifestations

Reproductive system changes The most common changes in the bleeding pattern in the menopausal transition are shortened cycle length, heavier flow (menorrhagia), and irregular cycle length (metrorrhagia). Intermenstrual bleeding may also oc-

cur, but it warrants specific attention because of its association with endometrial neoplasia in women older than 40 years [*see* Laboratory Tests, Biopsy, *below*]. A woman with vasomotor symptoms who has completely missed a menses is likely to experience her final menstrual period within the next 1 to 2 years.[14] Menopause usually occurs several years after the onset of menstrual changes; however, about 10% of women experience abrupt onset of amenorrhea.[3,4] Infertility and the cessation of menses are the only universal manifestations of menopause.

Genitourinary atrophy Genitourinary atrophy is typically mild and asymptomatic during the menopausal transition, but it is progressive and can become quite severe in the postmenopausal years.[3,31] Atrophic vulvovaginitis can present as vaginal dryness, vulvovaginal pruritus, vaginal dyspareunia, or postcoital spotting. Atrophic urethritis and recurrent cystitis can manifest as dysuria, frequency, and incontinence.

Vasomotor symptoms Vasomotor symptoms (i.e., hot flushes and night sweats) are common manifestations of the menopausal transition; for example, 80% of white women experience vasomotor symptoms.[56] Women typically describe hot flushes as a strong sensation of warmth accompanied by flushing, a prickling sensation of the skin, and perspiration that seems to move from the trunk toward the head before it dissipates. These flushes are spontaneous, uncomfortable, and unpredictable, and they can occur any time of the day or night. Each episode is self-limited and typically lasts several minutes. A number of women describe feeling excessively warm in a more continuous pattern. The frequency of vasomotor symptoms may be represented by a bell-shaped curve that peaks around the time of the final menses.[3,37] The occurrence of vasomotor symptoms usually ceases within 4 to 5 years from first onset, although 10% of women may suffer symptoms for much longer (up to 10 years).[14,37]

Changes in libido, sleep, mood, and cognition Changes in mood, libido, and sleep may also occur but are variable in severity[37] and have a wider differential diagnosis. Snoring and daytime sleepiness suggest the possibility of obstructive sleep apnea. Women may complain of mildly diminished cognitive capacity, particularly during the menopausal transition; however, this has not been well studied.

Physical Examination

There are no pathognomonic physical findings in the menopausal transition. However, the physical examination may provide information that suggests the presenting symptoms are the result of an underlying pathologic condition and are not related to normal menopausal transition. Palpation of the thyroid gland and examination for physical signs of hypothyroidism or hyperthyroidism are warranted, particularly if menstrual irregularity, excessive diaphoresis, or neurocognitive changes are present. When intermenstrual bleeding is reported, speculum examination should be performed to rule out cervical or vaginal lesions, such as endocervical polyps. Bimanual pelvic examination is indicated when bleeding is heavy or frequent, to rule out the presence of adnexal masses and evaluate the uterus for fibroids; it is also indicated when pregnancy is possible. When the clinical presentation of oligomenorrhea or amenorrhea is not classic for the menopausal transition, prolactinoma may be suspected, in which case visual-field testing for bitemporal hemianopsia and breast examination for galactorrhea are appropriate. In addition, inspection of the skin for needle tracks from possible injection use of heroin and evaluation for low body weight or significant weight loss may be useful, because these findings suggest a hypothalamic cause for oligomenorrhea or amenorrhea.

After menopause, vulvovaginal atrophy typically occurs. The vulvovaginal skin may appear pale, thin, and friable and may exhibit a loss of rugae and possible fissuring and erythema. The uterus is smaller, measuring 5 to 6 cm in length, and the ovaries are usually nonpalpable. The cervix may become stenotic and flush with the vagina.

Laboratory Tests

Urine or serum β–human chorionic gonadotropin (β-hCG) testing is crucial in the evaluation of any woman suspected to be in menopausal transition but who has the potential for pregnancy and who presents with a missed period, oligomenorrhea, or irregular vaginal bleeding with or without pain. In addition, testing for high-sensitivity thyroid-stimulating hormone (TSH) should be considered when menorrhagia, excessive diaphoresis, or neurocognitive changes—all potentially associated with the menopausal transition—suggest thyroid dysfunction. If the clinical picture suggests hemorrhagic diathesis, it may be helpful to obtain a platelet count, prothrombin time, and partial thromboplastin time. Additional evaluation for coagulopathies, such as von Willebrand factor, should follow, if appropriate.

With the onset of menstrual irregularity, there are wide variations in the production of FSH, estradiol, and LH.[23] Because of these wide variations, measurement of serum concentrations of these hormones is generally not useful during the menopausal transition and is not indicated unless the clinical situation is atypical and suggests an underlying condition. An elevated follicular-stage FSH level demonstrates that ovarian function is declining, but the FSH level cannot predict when the final menstrual period will occur.[19,21,23] Once a woman has had 12 months without a menses, the FSH is reliably elevated at 25 IU/L, and the estradiol level is less than 50 pg/ml.

Oral contraceptive use during the menopausal transition will treat menopausal symptoms and mask menopause. Oral contraceptives suppress FSH; therefore, the FSH should be drawn on the seventh day of placebo pills or the seventh day of the pill-free week. If menopause has occurred, the serum FSH level will be greater than 25 IU/L when drawn on two separate occasions.

Measurement of FSH serum concentration can assist in the diagnosis of menopause in a woman with vasomotor symptoms who has had a hysterectomy without oophorectomy. FSH testing may also be appropriate in the evaluation of atypical clinical situations; for example, in a case of abrupt-onset amenorrhea in a 40-year-old woman with negative β-hCG testing, measurement of FSH, prolactin, and TSH concentrations should be performed to evaluate for premature ovarian failure, prolactinoma, or thyroid dysfunction [*see* Differential Diagnosis, *below*].

Tests of other body fluids Vaginal fluid pH is elevated after menopause, and vaginal cytology shows a decreased maturation index (increase in parabasal cells).[30] These tests are not typically performed, nor are they necessary, to establish the diagnosis of menopause. In evaluation of vaginal dyspareunia or vulvovaginal pruritus, vaginal fluid pH testing and microscopic examination of vaginal fluid (saline and 10% potassium hydroxide preparations) should be performed to rule out common vaginal infections such as candidiasis, trichomoniasis, or bacterial vaginosis. It should be noted that with genitourinary atro-

phy, the shift toward a more basic pH can precipitate bacterial overgrowth and concomitant infection.

Imaging Studies

No imaging study is useful in establishing the diagnosis of the menopausal transition or menopause, although pelvic ultrasound may be indicated in the diagnostic evaluation of women with abnormal vaginal bleeding before or after menopause. In women who present with metrorrhagia or menorrhagia during the menopausal transition, pelvic ultrasound can confirm a diagnosis of leiomyomas or endometrial polyps and can suggest a diagnosis of adenomyosis; however, ultrasound cannot rule out endometrial neoplasia in premenopausal women [*see* Biopsy, *below*]. In contrast, ultrasound can serve as a screening test for endometrial neoplasia in postmenopausal women who experience bleeding or spotting spontaneously or in conjunction with hormone therapy (HT). In a postmenopausal woman, a homogeneous endometrial thickness of 4 mm or less confers assurance that endometrial hyperplasia or cancer is not present in more than 96% of cases.[57-59]

Biopsy In the menopausal transition, endometrial biopsy should be performed in women who experience intermenstrual bleeding (i.e., bleeding at intervals of fewer than 21 days) or in obese women who present with menometrorrhagia. If the biopsy results are normal or if examination suggests leiomyoma or adenomyosis, ultrasonography should follow. Endometrial biopsy is also indicated in postmenopausal women at heightened risk for endometrial neoplasia (e.g., women with diabetes or obesity) who experience any bleeding after 12 months of amenorrhea or who have an ultrasound result that demonstrates an endometrium at least 4 mm in thickness.

DIFFERENTIAL DIAGNOSIS

Menstrual Changes

For women 45 to 55 years of age who are experiencing progressive oligomennorhea, the most likely diagnosis is the menopausal transition, especially if there are associated vasomotor symptoms. In this setting, a wider differential diagnosis rarely needs to be considered. In younger women who have no vasomotor symptoms or whose menstrual changes are abrupt, a wider differential should be considered [*see* Premature Ovarian Failure, *below*]. The differential diagnosis for oligomenorrhea and secondary amenorrhea should always include pregnancy, prolactinoma, thyroid dysfunction, and medication or supplement use.

For women with excessive or intermenstrual bleeding, the differential diagnosis includes hypothyroidism, hyperthyroidism, blood dyscrasias, leiomyoma, adenomyosis, endometrial polyps, endometriosis, endometrial or cervical neoplasia, and hormone-secreting neoplasms such as granulosa cell ovarian cancer. Increased menstrual bleeding induced by medication or supplement use should also be considered.

Genitourinary Atrophy

Multiple conditions can cause genitourinary symptoms similar to those associated with hypoestrogenism occurring in the menopausal transition and menopause. Vulvovaginal symptoms (e.g., vaginal dryness, pruritus, dyspareunia, and postcoital spotting) may be caused by trichomonas vaginitis, yeast vulvovaginitis, bacterial vaginosis, desquamative inflammatory vaginitis, vestibulitis, allergic vulvovaginitis, and vulvar dysplasia or cancer. Urinary symptoms (e.g., dysuria, urinary frequency, and incontinence) may be caused by dietary bladder irritants, detrusor instability, urinary tract infection, and interstitial cystitis. The presence of isolated microscopic hematuria on urinalysis should prompt evaluation for neoplasia of the urinary tract.

Hot Flushes and Night Sweats

Hot flushes and night sweats may be symptoms of a number of disease processes, including hyperthyroidism, pheochromocytoma, carcinoid, and occult infection or neoplasm (e.g., tuberculosis, HIV), and lymphoma with B symptoms. Nonvolitional weight loss or documented fevers suggest a possible underlying disease. On the other hand, weight gain or existing obesity, which provides insulation against loss of body heat, may explain easy perspiration and a sensation of excess warmth in some women.

Changes in Libido, Sleep, Mood, and Cognition

The changes in libido, sleep patterns, mood, and cognition associated with the menopausal transition and menopause may also be induced by mood or anxiety disorders, thyroid dysfunction, and stress. Medications or other substances may cause insomnia, anxiety, mood abnormalities, cognitive changes, and sexual dysfunction. Other symptoms, such as fatigue, may be the result of an unrecognized sleep disorder (e.g., obstructive sleep apnea and restless legs syndrome), an inflammatory or neoplastic process, or multiple sclerosis. New cognitive dysfunction may be the first manifestation of dementia.

MENOPAUSAL TRANSITION AND POSTMENOPAUSAL SYMPTOM MANAGEMENT

Management of women experiencing menopausal symptoms is best approached by (1) defining the reproductive phase[2] of the patient [*see Figure 1*]; (2) identifying the menopausal symptoms for which treatment is desired [*see Figure 4*]; and (3) identifying the medical conditions that might influence management options [*see Sidebar* Internet Resources for Information on Menopause].

The menopausal transition and menopause do not warrant management in and of themselves. However, women who experience bothersome symptoms may want to consider treatment. HT is effective in controlling symptoms of the menopausal transition and menopause, but it carries risks [*see* Hormone Therapy Risks and Benefits, *below*].

The Food and Drug Administration has recommended that HT be used only for women with symptoms severe enough to warrant its use and at the lowest dose and for the shortest duration required to ease the menopausal transition. The FDA further recommends that tissue-targeted therapies be used whenever possible.[60]

Because of the potential risks associated with HT, it is recommended that all women taking HT be evaluated on an annual basis. The woman who is taking HT should be instructed to refrain from taking HT 1 week before the annual assessment to allow the physician and patient to evaluate the current severity of symptoms. Women who choose to stop HT may require a slow taper, ranging from 3 to 6 months, for successful cessation. Women on HT are encouraged to attempt cessation after 5 years of use.

Uterine Bleeding

Vaginal bleeding during the menopausal transition is best managed (after appropriate evaluation) with low-dose, combi-

nation oral contraceptives (containing 20 µg ethinyl estradiol) or a progestin intrauterine device; both protect against pregnancy and reduce menstrual blood loss. The overall effect of ethinyl estradiol at 20 µg/day is estimated to be three to four times that of 0.625 mg/day of conjugated estrogen; head-to-head trials comparing the clinical effects of the two estrogens do not exist. Although the chance of pregnancy is low (< 1% after age 50),[61] pregnancy may occur during the menopausal transition. Women 40 to 49 years of age have a rate of unintended pregnancy that is now higher than that of any other age group, even teenagers[62]; thus, it is important to address the issue of contraception with every potentially fertile woman until she has experienced 12 months of amenorrhea. Oral contraceptives can be discontinued and symptoms reassessed at approximately age 50 [*see* Laboratory Testing, *above*]. Those with persistent and severe vasomotor symptoms and amenorrhea may be transitioned to postmenopausal HT.

Genitourinary Atrophy

Symptoms of genitourinary atrophy may be present during the menopausal transition but typically become more prominent after menopause. Symptoms of genitourinary atrophy usually improve within 2 weeks after initiation of estrogen therapy and should be controlled after 1 to 3 months of use.[63] Estrogen can be administered topically with excellent local effect. Vaginal estrogen creams result in little or no systemic absorption when used at extremely low doses of less than one-eighth applicator (< 0.15 mg conjugated estrogen cream) and one-sixteenth applicator (< 0.025 mg estradiol cream); a full applicator of vaginal estrogen cream can deliver a dose equivalent to that of an oral formulation, although the rate of absorption varies considerably. Initial therapy constitutes nightly application for 2 to 6 weeks; thereafter, maintenance doses can be applied one to three times a week, depending on the severity of symptoms. Use of low-dose vaginal creams does not necessitate the use of a progestin; however, cessation of therapy results in the return of genitourinary atrophy. Low-dose vaginal estrogen rings effectively treat genitourinary atrophy with little or no systemic absorption and no significant endometrial stimulation.

Nonhormonal alternatives include lubricants for use during intercourse and vaginal moisturizers.

Vasomotor Symptoms

Estrogen is highly effective for the treatment of vasomotor symptoms. For women in the menopausal transition who are at risk for pregnancy and who have heavy or frequent menses, treatment with low-dose oral contraceptives provides amelioration of vasomotor symptoms, control of bleeding, and contraception. Cyclical HT is preferable for women in the menopausal transition who are predominantly anovulatory and who do not need contraception, because it provides lower doses of hormones than oral contraceptives. Continuous HT, which is commonly used in women who are approximately 12 months from the final menstrual period, is often associated with bothersome bleeding patterns in women in earlier stages of the menopausal transition.

Low-dose oral contraceptives and HT result in prompt resolution of symptoms within 1 to 2 weeks in 80% of women; they should be titrated to the lowest dose possible to achieve acceptable symptom relief. Systemic administration of estrogen can be achieved orally, transdermally (in the form of a gel or patch), or transmucosally (at a higher dose via the use of a vaginal ring).

Internet Resources for Information on Menopause

General Information about Menopause for Clinicians and Patients
- North American Menopause Society
  http://www.menopause.org
- Women's Health Initiative
  http://www.whi.org

Alternative Therapies for Management of Menopausal Symptoms
- University of Washington School of Medicine, Department of Family Medicine, on Complementary and Alternative Medicine
  http://www.fammed.washington.edu/predoctoral/CAM
- ConsumerLab.com
  http://www.consumerlab.com
- The Longwood Herbal Task Force
  http://www.mcp.edu/herbal
- Natural Medicines Comprehensive Database
  http://www.naturaldatabase.com

Some formulations include progestins. Implants and intramuscular injections are less preferable forms of delivery because they release extremely high levels of HT at the time of administration or placement. The lowest doses of estrogens found to be effective for vasomotor symptoms include oral conjugated estrogen 0.03 mg, oral estradiol 0.5 mg, and transdermal estradiol 0.025 mg.

Several nonhormonal alternatives for treatment of vasomotor symptoms may have some efficacy. These include venlafaxine,[64,65] selective serotonin reuptake inhibitors (SSRIs) (e.g., fluoxetine and paroxetine),[66-68] and gabapentin.[69] Less evidence supports the use of clonidine[70,71] and vitamin E[72] as being effective in the control of vasomotor symptoms. Results from controlled clinical trials evaluating the effectiveness of phytoestrogens (including dietary soy) for vasomotor symptoms vary, but most studies indicate that the use of phytoestrogens offers no significant improvement over placebo in reducing the frequency of hot flushes.[73] Black cohosh, a possible phytoestrogen, may be effective, but no large controlled trials have been conducted.[74] Progestin alone is effective[75,76] but is not recommended because of a potential increased risk of breast cancer.[77] Behavioral modification[78,79] and increased exercise[78,80] may diminish the severity of hot flushes. Red clover extract, dong quai, evening primrose oil, and Siberian ginseng have not been found to be effective in small randomized, controlled trials. Other botanicals purported to be effective, including valerian, motherwort, and chasteberry, have not been studied in clinical trials.[81]

Libido, Sleep, Mood, and Cognition

Treatment of sexual dysfunction depends on the underlying etiology. If the cause of decreased libido is not predominantly psychosocial, testosterone therapies have been shown, in some circumstances, to improve sexual function, interest and frequency of desire, and psychological well-being.[82,83] There are no FDA-approved products for diminished libido in women; however, esterified estrogen combined with methyltestosterone is commonly used. Vaginal estrogen therapy, if indicated, can play an important role in the treatment of diminished sexual function resulting from urogenital atrophy.[63]

Vasomotor instability may contribute to disruption of sleep;

thus, estrogen is effective for some women who begin to experience insomnia during the menopausal transition. Alternative therapies include short-term zolpidem and low-dose trazodone.

Estrogen may be beneficial in the treatment of depression in the menopausal transition.[84] Estrogen alone, without an antidepressant, does not appear to be sufficient to treat significant clinical depression in postmenopausal women.[85] However, some investigators support the use of estrogen as an adjunct to other therapies, such as SSRIs, particularly in older women.[86]

MANAGEMENT CONSIDERATIONS

Risks and Benefits of Hormone Therapy

Although the risks and benefits of using HT for the relief or prevention of symptoms in women in the menopausal transition have not been evaluated in clinical trials, information has been established on the risk-to-benefit profile of short-term and long-term use of HT for postmenopausal women 50 to 79 years of age. Historically, it was believed that the estrogen deprivation that accompanies menopause increases the risk of some chronic diseases—specifically, heart disease, osteoporosis, and dementia. On the basis of observational data, long-term postmenopausal HT was recommended during the 1980s and 1990s not only for symptom relief but also to reduce the risk of chronic disease and to prolong life. However, two large randomized, controlled trials (i.e., Heart and Estrogen/Progestin Replacement Study [HERS][87] and the Women's Health Initiative [WHI])[88,89] called this practice into question. HT is no longer recommended for primary or secondary prevention of these conditions in women older than 50 years.

HERS demonstrated no evidence to support the use of HT for the secondary prevention of heart disease[87]; more important, WHI found that HT use conferred an increased risk of cardiovascular disease,[90] stroke,[91] dementia,[92] thromboembolism,[88] and breast cancer[93] in women 50 to 79 years of age. Striking discrepancies in the findings of the randomized trials and the earlier nonrandomized (observational) studies can be explained by selection biases in participants in the observational studies. In the observational studies, women opting for HT therapy tended to be healthier and of higher socioeconomic status than non-HT users. In addition, these women were more likely to be carefully screened for chronic disease before starting HT and, therefore, had a lower risk of developing chronic disease than nonusers of HT.[94] The selection biases inherent in the observational studies were virtually eliminated in the randomized trials.

Counseling about the risks of postmenopausal HT use should now be based on the evidence provided by WHI.[88,89,91-100] The WHI postmenopausal estrogen and progestin therapy (EPT) and estrogen therapy (ET) trials are discussed in greater detail below.

Estrogen and progestin therapy The WHI prematurely halted its clinical trial of EPT in 2002; participants had been followed for an average of 5.2 years. The trial randomized over 16,000 postmenopausal women who were 50 to 79 years of age to take either conjugated equine estrogen (0.625 mg/day) plus medroxyprogesterone acetate (2.5 mg/day) or placebo. The study was halted because the rates of adverse events (i.e., cardiovascular events, stroke, thromboembolism, and breast cancer) were 1% higher in the intervention group and overshadowed the reduced risk of osteoporotic fractures and colon cancer.[88] There was no difference in overall or disease-specific mortality between the HT and placebo groups. The study reported the following risks: (1) thromboembolic events were highest in the first year and remained elevated over 5 years (absolute risk difference, 21/10,000/yr); (2) ischemic cardiac events were highest in the first year and remained elevated and statistically unchanged thereafter (absolute risk difference, 7/10,000/yr); (3) stroke risk was not elevated in the first year, rose slightly in the second year, and remained elevated through year 5 (absolute risk difference, 8/10,000/yr); and (4) breast cancer risk was not appreciably higher in years 1 to 3 but became elevated in year 4 (absolute risk difference, 8/10,000/yr), with the increased breast cancer risk being strongest in the approximately 25% of women who had taken HT before enrolling in the study [*see Table 1*]. The EPT portion of the WHI study showed a

Table 1 WHI Findings: Outcomes Associated with Use of Combined Estrogen and Progestin and Estrogen Alone in Healthy Postmenopausal Women

	*Combined Estrogen and Progestin**		*Estrogen Alone†*	
Outcomes	*Relative Risk (95% CI)*	*Absolute Risk Difference‡*	*Relative Risk 95% (CI)*	*Absolute Risk Difference‡*
Adverse/neutral				
Deep vein thrombosis[88,89]	2.07 (1.49–2.87)	13	1.47 (1.04–2.08)	6
Pulmonary embolism[88,89]	2.13 (1.39–3.25)	8	1.34 (0.87–2.06)	11
Coronary artery disease[89,90]	1.24 (1.00–1.54)	7	0.91 (0.75–1.12)	5
Ischemic stroke[89,91]	1.44 (1.09–1.90)	8	1.39 (1.10–1.77)	12
Breast cancer[89,93]	1.24 (1.01–1.54)	8	0.77 (0.59–1.01)	7
Probable dementia[§95,98]	2.05 (1.21–3.48)	23	1.49 (0.83–2.66)	12
Beneficial/neutral				
Colorectal cancer[88,89,100]	0.56 (0.38–0.81)	6	1.08 (0.75–1.55)	1
All fractures[88,89,96]	0.76 (0.69–0.85)	44	0.70 (0.63–0.79)	56
Mortality[88,89]	0.98 (0.82–1.18)	1	1.04 (0.88–1.22)	3

*Patients received 0.625 mg/day of conjugated estrogen and 2.5 mg/day of medroxyprogesterone acetate.
†Hysterectomized patients received 0.625 mg/day of conjugated estrogen.
‡Annual per 10,000 women.
§Ages: 65–79 yr.
CI—confidence interval

reduction in the risk of hip fractures (absolute risk difference, 6/10,000/yr)[96] and colorectal cancer (absolute risk difference, 6/10,000/yr).[100]

The Women's Health Initiative Memory Study (WHIMS), a substudy of the WHI continuous combined HT intervention trial, observed the effect of HT on memory and cognition in women 65 to 79 years of age (average age, 73 years).[92,95] A reduction in memory and thinking abilities (as measured by the Modified Mini-Mental State Examination)[95] and an increase in dementia of all types (absolute risk increase, 2/1,000/yr) were observed in women who took HT.[92]

Estrogen therapy The estrogen-only arm of WHI was halted prematurely in early 2004 because of increased risk of stroke; participants had been followed for an average of 6.8 years.[89] Over 10,000 women 50 to 79 years of age were randomized to receive 0.625 mg/day of conjugated equine estrogen or placebo. As in the EPT portion of WHI and HERS, the ET portion of the WHI trial found that estrogen use conveyed an increased risk of deep vein thrombosis (1.47 relative risk; 95% confidence interval [CI], 1.04 to 2.08) and stroke (1.39 relative risk; 95% CI, 1.10 to 1.77) [*see Table 1*]. Women taking ET had 12 more strokes and six more events of deep vein thrombosis a year than the women taking placebo. The study showed a reduction in the incidence of hip fractures (0.61 relative risk; 95% CI, 0.41 to 0.91) and an unanticipated, though not statistically significant, reduction in breast cancer incidence, a finding that requires further investigation. Observational studies support an increased risk of thromboembolism,[101-103] cholecystitis,[104] and breast cancer[77] in women taking ET.

In contrast to the EPT portion of the WHIMS study, women taking estrogen alone did not have a statistically increased risk of dementia.[98] For women taking ET, as compared with placebo, the risk of having a 10-unit decrease in the Modified Mini-Mental State examination scores (greater than two standard deviations) was 1.47 (95% CI, 1.04 to 2.07). The risk was greater in women with lower cognitive function at initiation of ET.[99]

Type, route of administration, and dose of HT Two observational studies[77,105] and a population-based study from Southern California[106] have increased current understanding of the type, route of administration, and dose of HT with associated breast cancer risk. The findings are as follows: (1) use of estrogen therapy confers greater risk than nonuse[77,107]; (2) use of estrogen plus progestin confers greater risk than use of estrogen alone[77,105,106]; (3) risk increases with duration of estrogen use[77,105,106]; (4) risk with estrogen use is increased in women with low or normal body mass index but not in overweight and obese women[105]; (5) increased risk of breast cancer is associated with any dose and type of commonly used estrogen (i.e., conjugated estrogen and estradiol) and progestin (i.e., medroxyprogesterone acetate, norethisterone, and levonorgestrel/norgestrel); and (6) transdermal formulations of estrogen also confer increased risk.[77] Surprisingly, the use of tibolone, a synthetic steroid with estrogenic, progestogenic, and androgenic properties that is marketed in Europe for its favorable effect on breast symptoms (e.g., tenderness and mastalgia), was also associated with a greater risk of breast cancer than nonuse of steroid hormones.[77] Current evidence indicates that all forms of HT are associated with an increased risk of breast cancer.

Transdermal delivery of estrogen may not be safer if administration is long term; however, transdermal patches and transmucosal delivery systems (including estrogen vaginal rings that provide systemic levels of estrogen for treatment of vasomotor symptoms) avoid the first-pass effect through the liver and may carry a lower risk of thromboembolism,[48,102,108,109] elevation of bile acids,[110] and hypertriglyceridemia[111] than oral estrogen. Studies of moderate- to high-dose regimens suggest that transdermal systems may have a more favorable effect on the coagulation pathway[112] and C-reactive protein[113] than oral products. Transdermal and transmucosal products have been shown to be effective as treatment for vasomotor symptoms[114] and maintenance of bone mineral density.[115]

Given the newly appreciated risks of oral progestins,[77,88] alternative approaches to progestin therapy are gaining popularity. There is a widely held belief that natural progesterone is better than synthetic progestins, but this hypothesis has never been studied. Lower-dose oral micronized progesterone formulations have been widely used in Canada and Europe and were evaluated in the Postmenopausal Estrogen/Progestin Intervention (PEPI) trial.[116] The PEPI study of 596 postmenopausal women found that micronized progesterone combined with estrogen sufficiently diminished the hyperplastic endometrial changes associated with estrogen-only therapy.[116] No other randomized clinical trials exist to better inform us about risks and safety of micronized progesterone, particularly with respect to breast cancer. Likewise, over-the-counter transdermal progesterone creams have not been studied in this regard.

In addition, attention has been directed at nonsystemic therapies. Off-label use of a progestin intrauterine system (IUS) (20 µg/day of levonorgestrel) in postmenopausal women taking estrogen provides low systemic levels of progestin and attenuation of endometrial stimulation by estrogen.[117] Intrauterine levonorgestrel at doses of 10 µg and 14 µg/day have been studied in Europe.[117,118] No increased risk of endometrial hyperplasia or cancer has been observed in women taking estrogen with a progestin IUS in place.[117] Vaginal application of progesterone creams result in local uterine and systemic effects.[119]

Preventive Health Care

The menopausal transition offers women and their health care providers the opportunity to review and focus on preventive health care measures, including basic health habits, such as regular exercise, good nutrition with calcium and vitamin D supplementation, and avoidance of smoking [*see Table 2*].

WHI demonstrated that HT is effective for the prevention of osteoporotic fractures and colorectal cancer in postmenopausal women, but the risks outweigh the benefits[88] [*see Table 1*]. The prevalence of certain medical conditions (e.g., dementia, coronary artery disease, breast cancer, colon cancer, and diabetes mellitus) increase with age, rising more steeply after loss of ovarian function. Recommended management for these conditions is almost always nonhormonal. HT is not indicated for primary prevention of disease, unless a woman at high risk for osteoporosis chooses HT over other options after consideration of the risks and benefits.[60] Two chronic disease processes associated with hypoestrogenism and aging greatly impact women's health in the postmenopausal years, namely cardiovascular disease and osteoporosis. Cardiovascular disease is the leading cause of death in women in the United States, and osteoporosis is a major cause of morbidity. The management of these diseases is discussed more fully elsewhere [*see 16:IX Cardiovascular Disease in Women and 3:VI Diseases of Calcium Metabolism and Metabolic Bone Disease*].

Table 2 Preventive and Screening Measures for Common Conditions in Postmenopausal Women

Condition	*Prevention*	*Early Detection*
Dementia	Participation in cognitive leisure activities Regular exercise Treatment of hypertension Statins and possibly other lipid-lowering agents Long-term NSAID use* Avoidance of HT initiation in postmenopause†	
CAD	Smoking cessation Regular exercise Diet high in nuts, whole-grains, and total fiber (especially water-soluble fiber), folate, and marine n-3 fatty acids Diet low in saturated fat, *trans* fatty acids, and glycemic load Daily, low-dose alcohol Prevention and treatment of hypertension, diabetes mellitus, and hypercholesterolemia Consideration of low-dose daily aspirin if risk of CAD events is ≥ 0.7%/yr Avoidance of HT initiation in postmenopause Statins, aspirin, and beta blockers for secondary prevention (underutilized)	
Breast cancer	Minimal exposure to HT (estrogen and/or progestin) Regular exercise Avoidance of increase in weight and waist circumference Weight loss if overweight/obese Reduction of alcohol intake to 0–20 g/day Raloxifene* if at average risk, or tamoxifen if risk ≥ 1.67%/yr	Screening mammography every 1–2 yr, with or without clinical breast exam regardless of age until clinically significant comorbid conditions
Osteoporosis	Adequate calcium and vitamin D intake Weight-bearing exercise Thiazide* diuretics Antiresorptive treatment before first osteoporotic fracture Antiresorptive treatment after osteoporotic fracture HT† if intolerant of or unresponsive to first-line agents	Screening DEXA at age 65 (earlier if risk factors)
Colon cancer	High-fiber diet for primary, but not secondary, prevention of polyps Aspirin* if personal history of adenoma or colon cancer Removal of adenomatous polyps HT* effective but not advised for this indication†	Periodic screening‡ by fecal occult blood testing or sigmoidoscopy
Diabetes mellitus	Regular exercise Weight loss if overweight/obese Metformin, acarbose, and possibly thiazolidinediones for those at high risk HT† effective but not advised for this indication	

*Off-label indication.

†Increased risk of adverse outcomes has been demonstrated for HT initiation in the postmenopausal years. HT is not advised for the primary prevention of disease, because associated risks outweigh benefits for most women; in limited cases, HT may be used for prevention of osteoporosis, after consideration of all other options. HT should be used only for severe and debilitating symptoms, in the lowest dose and most directed therapy possible, and for the shortest time necessary to accomplish symptom control.

‡The United States Preventive Services Task Force (USPSTF) recommends screening for colon cancer starting at age 50 using either annual fecal occult blood testing, sigmoidoscopy (periodicity unspecified), or both.

HT—hormone therapy CAD—coronary artery disease DEXA—dual x-ray absorptiometry NSAID—nonsteroidal anti-inflammatory drug

Premature Ovarian Failure

Premature ovarian failure (POF) is defined as menopause that occurs before 40 years of age that is not iatrogenically induced. The prevalence of POF is approximately 1%.[120] The Study of Women Across the Nation (SWAN) investigated risk factors associated with POF[121] and found that ethnicity influences risk: POF occurs in 1.1% of white women and 1.4% of African-American and Hispanic women, but it occurs in only 0.5% of Chinese-American women and 0.1% of Japanese-American women. Higher body mass index is associated with increased likelihood of POF, especially in African-American women. Disability and current smoking are associated with greater risk in white women.

ETIOLOGY

There is good evidence to suggest that the timing of the age of menopause is genetically programmed[15,16] and that genes play a significant role in the etiology of premature ovarian failure.[120] Rare genetic and chromosomal causes of premature ovarian failure include familial predisposition, FSH receptor mutations, galactosemia, 17α-hydroxylase deficiency, alterations in gonadotropin structure or function, and structural alterations of the X

chromosome (e.g., Turner syndrome mosaicism). A common genetic cause of POF is the fragile X premutation. Up to 3% to 5% of women with POF are carriers of the fragile X premutation, the most common cause of mental retardation in males.[122] Approximately 16% of women who are fragile X premutation heterozygotes have POF.[123]

Premature menopause may be immune-mediated in 30% to 50% of women with POF.[123] Family history is often positive for autoimmune conditions,[124] and other autoimmune diseases may be present in the patient herself,[124] including autoimmune thyroiditis, type 1 diabetes mellitus, autoimmune hemolytic anemia, Addison disease, hypoparathyroidism, idiopathic thrombocytopenic purpura, Crohn disease, myasthenia gravis, rheumatoid arthritis, systemic lupus erythematosus, vitiligo, and polyendocrine failure.

DIAGNOSIS

Clinical Manifestations

The presentation of premature ovarian failure is identical to that observed in natural menopause, with the exception that POF occurs before 40 years of age. It is more common, however, for women with POF to experience waxing and waning of symptoms over longer periods than it is for women who have natural menopause, and some women will ovulate several years after a diagnosis of POF is made.

Physical Examination

A targeted examination for women with oligomenorrhea or secondary amenorrhea is described elsewhere [*see* Natural Menopause, Physical Examination, *above*]. Less common etiologies of POF, such as Turner Syndrome (i.e., short stature, webbed neck, shield chest, small fourth metacarpal, and minimal breast development with normal hair distribution), can be detected with a specifically targeted physical examination. Findings of other autoimmune conditions often associated with POF [*see* Etiology, *above*] may be present in some women, including signs of thyroid disease (i.e., enlarged, asymmetrical, or nodular thyroid gland; dry skin; lateral eyebrow thinning; delayed relaxation phase on deep tendon reflexes; and myxedema), adrenal dysfunction (i.e., hyperpigmentation and orthostatic hypotension), and systemic lupus erythematosus (i.e., synovitis or malar rash). Galactorrhea suggests an elevated prolactin (PRL) prolactinoma [*see* Differential Diagnosis, *below*].

Laboratory Tests

Ovarian failure is most accurately confirmed by measurement of serum FSH. In women with incipient ovarian failure, FSH levels are often between 15 and 25 IU/L and can fluctuate. Complete ovarian failure is associated with repeated serum FSH levels greater than 25 IU/L. Therefore, FSH levels persistently greater than 25 IU/L (drawn on at least two separate occasions) can be useful in making the diagnosis of POF. Testing of urine or serum β-hCG, TSH, and prolactin concentrations should not be deferred if indicated in the evaluation of oligomenorrhea or secondary amenorrhea [*see* Natural Menopause, Laboratory Tests, *above*]. If a diagnosis of POF is made, consideration of genetic testing for a premutation allele of fragile X may be advisable, providing that this information would benefit family members and that the patient agrees to testing. If a woman younger than 30 years is diagnosed with POF, a karyotype test should be considered to rule out Turner syndrome mosaicism. When POF may be caused by autoimmunity, the complete blood count (CBC), erythrocyte sedimentation rate (ESR), rheumatoid factor (RF), antinuclear antibody (ANA), glucose, calcium, and phosphorus levels can point to associated autoimmune conditions that may not otherwise be clinically apparent.

DIFFERENTIAL DIAGNOSIS

Hypergonadotropic amenorrhea can be caused by thyroid dysfunction, hyperprolactinemia, heroin addiction, and the use of some antidepressant and antipsychotic medications.

MANAGEMENT

All women with POF should be treated with exogenous estrogen, either in the form of a low-dose estrogen-progestin combination contraceptive or a postmenopausal HT formulation to manage symptoms and decrease the risk of osteoporosis and osteopenia. Bone mineral density should be obtained at baseline and followed at intervals of 3 to 5 years. It is recommended that women continue estrogen replacement until at least age 50 (approximately the time of natural menopause). Progestin therapy is recommended for women who have a uterus. Women who are at risk for unintended pregnancy should receive exogenous estrogen and progestin in the form of a contraceptive. For those desiring pregnancy, artificial reproductive technology is available. In vitro fertilization utilizing donor eggs and hormonal manipulation to mature the endometrium result in successful pregnancy in women with POF as often as in women with infertility from other causes.[125]

Treatment with oral contraceptives in the general population is associated with an increased risk of thromboembolic disease, cardiovascular disease in smokers, and stroke in women with migraine headaches or hypertension.[126] The risks of using oral contraceptives and postmenopausal HT in women with POF has not been specifically studied.

COMPLICATIONS AND PROGNOSIS

The chance of spontaneous pregnancy in POF is estimated to be less than 10%.[127] Women with POF may be at higher risk for younger onset of cardiovascular disease.[128] It is estimated that women with POF who do not take estrogen have a lower background risk of breast cancer and thromboembolism than the general population.[129] New onset of autoimmune disorders is not uncommon after the diagnosis of POF has been made.

Early-age mortality may occur in women with POF because of autoimmune phenomena, cardiovascular disease, and osteoporosis. A few epidemiologic studies suggest that an earlier age at menopause is associated with substantially increased mortality[128,129]; thus, careful screening for and management of chronic disease processes associated with hypoestrogenism [*see Table 2*] may be important for sustaining long-term health and quality of life. In addition, careful management of any coexisting autoimmune disorder and reduction, when possible, of potential risks posed by medications used to treat such disorders (e.g., corticosteroids) may be crucial for the long-term health of affected women.

Susan D. Reed, M.D., is a research consultant for Pfizer Inc.

Eliza L. Sutton, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

The progestin intrauterine system and esterified estrogen combined with methyltestosterone have not been approved by the FDA for uses described in this chapter.

References

1. Research on the Menopause in the 1990's. Proceedings of a meeting. Geneva, Switzerland, 14–17 June 1994. Maturitas 23:109, 1996

2. Soules MR, Sherman S, Parrott E, et al: Executive summary: Stages of Reproductive Aging Workshop (STRAW). Climacteric 4:267, 2001

3. McKinlay SM, Brambilla DJ, Posner JG: The normal menopause transition. Maturitas 14:103, 1992

4. Treloar AE: Menstrual cyclicity and the pre-menopause. Maturitas 3:249, 1981

5. Kronenberg F: Hot flashes: epidemiology and physiology. Ann N Y Acad Sci 592:52, 1990

6. Kato I, Toniolo P, Akhmedkhanov A, et al: Prospective study of factors influencing the onset of natural menopause. J Clin Epidemiol 51:1271, 1998

7. Gold EB, Bromberger J, Crawford S: Factors associated with age at natural menopause in a multiethnic sample of midlife women. Am J Epidemiol 153:865, 2001

8. Rodstrom K, Bengtsson C, Milsom I, et al: Evidence for a secular trend in menopausal age: a population study of women in Gothenburg. Menopause 10:538, 2003

9. McKinlay SM: The normal menopause transition: an overview. Maturitas 23:137, 1996

10. van Noord PA, Dubas JS, Dorland M, et al: Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. Fertil Steril 68:95, 1997

11. Cooper GS, Thorp JM Jr: FSH levels in relation to hysterectomy and to unilateral oophorectomy. Obstet Gynecol 94:969, 1999

12. Baker TG: Radiosensitivity of mammalian oocytes with particular reference to the human female. Am J Obstet Gynecol 110:746, 1971

13. Soules MR, Bremmer WJ: The menopause and climacteric: endocrinologic basis and associated symptomatology. J Am Geriatr Soc 30:547, 1982

14. Mitchell ES, Woods NF, Mariella A: Three stages of the menopausal transition from the Seattle Midlife Women's Health Study: toward a more precise definition. Menopause 7:334, 2000

15. Snieder H, MacGregor AJ, Spector TD: Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. J Clin Endocrinol Metab 83:1875, 1998

16. Treloar SA, Do KA, Martin NG: Genetic influences on the age at menopause. Lancet 352:1084, 1998

17. Torgerson DJ, Thomas RE, Reid DM: Mothers and daughters menopausal ages: is there a link? Eur J Obstet Gynecol Reprod Biol 74:63, 1997

18. Weel AE, Uitterlinden AG, Westendorp IC, et al: Estrogen receptor polymorphism predicts the onset of natural and surgical menopause. J Clin Endocrinol Metab 84:3146, 1999

19. Rannevik G, Jeppsson S, Johnell O, et al: A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG and bone mineral density. Maturitas 21:103, 1995

20. Cooper GS, Baird DD, Hulka BS, et al: Follicle-stimulating hormone concentrations in relation to active and passive smoking. Obstet Gynecol 85:407, 1995

21. Burger HG, Dudley EC, Hopper JL, et al: Prospectively measured levels of serum follicle-stimulating hormone, estradiol, and the dimeric inhibins during the menopausal transition in a population-based cohort of women. J Clin Endocrinol Metab 84:4025, 1999

22. Klein NA, Battaglia DE, Woodruff TK, et al: Ovarian follicular concentrations of activin, follistatin, inhibin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-2 (IGFBP-2), IGFBP-3, and vascular endothelial growth factor in spontaneous menstrual cycles of normal women of advanced reproductive age. J Clin Endocrinol Metab 85:4520, 2000

23. Landgren BM, Colllins A, Csemiczky G: Menopause transition: annual changes in serum hormonal patterns over the menstrual cycle in women during a nine-year period prior to menopause. J Clin Endocrinol Metab 89:2763, 2004

24. Dennerstein L, Smith AM, Morse C, et al: Menopausal symptoms in Australian women. Med J Aust 159:232, 1993

25. Santoro N, Brown JR, Adel T, et al: Characterization of reproductive hormonal dynamics in the perimenopause. J Clin Endocrinol Metab 81:1495, 1996

26. Zumoff B, Strain GW, Miller LK, et al: Twenty-four-hour mean plasma testosterone concentration declines with age in normal premenopausal women. J Clin Endocrinol Metab 80:1429, 1995

27. Burger HG, Dudley EC, Cui J, et al: A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone-binding globulin levels through the menopause transition. J Clin Endocrinol Metab 85:2382, 2000

28. Labrie F, Belanger A, Cusan L, et al: Marked decline in serum concentrations of adrenal C19 sex steroid precursors and conjugated androgen metabolites during aging. J Clin Endocrinol Metab 82:2396, 1997

29. Scobie GA, Macpherson S, Millar MR, et al: Human oestrogen receptors: differential expression of ER alpha and beta and the identification of ER beta variants. Steroids 67:985, 2002

30. Garcia-Closas M, Herrero R, Bratti C, et al: Epidemiologic determinants of vaginal pH. Am J Obstet Gynecol 180:1060, 1999

31. Dennerstein L, Dudley EC, Hopper JL, et al: A prospective population-based study of menopausal symptoms. Obstet Gynecol 96:351, 2000

32. Freedman RR, Krell W: Reduced thermoregulatory null zone in postmenopausal women with hot flashes. Am J Obstet Gynecol 181:66, 1999

33. Berendsen HH: The role of serotonin in hot flushes. Maturitas 36:155, 2000

34. Slopien R, Maczekalski B, Warenik-Szymankiewicz A: Relationship between climacteric symptoms and serum serotonin levels in postmenopausal women. Climacteric 6:53, 2003

35. Warenik-Szymankiewicz A, Meczekalski B: Neuroendocrine aspects of menopause. Ginekol Pol 68:620, 1997

36. Jansson C, Johansson S, Lindh-Astrand L, et al: The prevalence of symptoms possibly related to the climacteric in pre- and postmenopausal women in Linkoping, Sweden. Maturitas 45:129, 2003

37. Moe KE: Reproductive hormones, aging, and sleep. Semin Reprod Endocrinol 17:339, 1999

38. Bixler EO, Vgontzas AN, Lin HM, et al: Prevalence of sleep-disordered breathing in women: effects of gender. Am J Respir Crit Care Med 163:608, 2001

39. Young T, Rabago D, Zgierska A, et al: Objective and subjective sleep quality in premenopausal, perimenopausal, and postmenopausal women in the Wisconsin Sleep Cohort Study. Sleep 26:667, 2003

40. Popovic RM, White DP: Upper airway muscle activity in normal women: influence of hormonal status. J Appl Physiol 84:1055, 1998

41. Dennerstein L, Randolph J, Taffe J, et al: Hormones, mode, sexuality, and the menopausal transition. Fertil Steril 77(suppl 4):S42, 2002

42. Seifert-Klauss V, Mueller JE, Luppa P, et al: Bone metabolism during the perimenopausal transition: a prospective study. Maturitas 41:23, 2002

43. Recker R, Lappe J, Davies K, et al: Characterization of perimenopausal bone loss: a prospective study. J Bone Miner Res 15:1965, 2000

44. Do KA, Green A, Guthrie JR, et al: Longitudinal study of risk factors for coronary heart disease across the menopausal transition. Am J Epidemiol 151:584, 2000

45. Gordon T, Kannel WB, Hjortland MC, et al: Menopause and coronary heart disease. The Framingham Study. Ann Intern Med 89:157, 1978

46. Schaefer EJ, Lamon-Fava S, Cohn SD, et al: Effects of age, gender, and menopause status on plasma low-density lipoprotein cholesterol and apolipoprotein B levels in the Framingham Offspring Study. J Lipid Res 35:779, 1994

47. Wagner JD: Rationale for hormone replacement therapy in atherosclerosis prevention. J Reprod Med 45(3 suppl):245, 2000

48. Peverill RE: Hormone therapy and venous thromboembolism. Best Pract Res Clin Endocrinol Metab 17:149, 2003

49. Bolognia JL: Aging skin. Am J Med 98:99S, 1995

50. Ashcroft GS, Dodsworth J, van Boxtel E, et al: Estrogen accelerates cutaneous wound healing associated with an increase in TGF-beta 1 levels. Nat Med 3:1209, 1997

51. Di Leo A, Messa C, Cavallini A, et al: Estrogens and colorectal cancer. Curr Drug Targets Immune Endocr Metabol Disord 1:1, 2001

52. L'Hostis H, Deminiere C, Ferriere J, et al: Renal angiomyolipoma: a clincopathologic, immunohistochemical, and follow-up study of 46 cases. Am J Surg Pathol 23:1011, 1999

53. Ferrans VJ, Yu ZX, Nelson WK, et al: Lymphangioleiomyomatosis (LAM): a review of clinical and morphological features. J Nippon Med Sch 67:311, 2000

54. Blankenstein MA, Verheijen FM, Jacobs JM, et al: Occurrence, regulation, and significance of progesterone receptors in human meningioma. Steroids 65:795, 2000

55. Pfahlberg A, Hassan K, Wille L, et al: Systematic review of case-control studies: oral contraceptives show no effect on melanoma risk. Public Health Rev 25:309, 1997

56. Keenan NL, Mark S, Fugh-Berman A, et al: Severity of menopausal symptoms and use of both conventional and complementary/alternative therapies. Menopause 10:507, 2003

57. Gull B, Karlsson B, Milsom I, et al: Can ultrasound replace dilation and curettage? A longitudinal evaluation of postmenopausal bleeding and transvaginal sonographic measurement of the endometrium as predictors of endometrial cancer. Am J Obstet Gynecol 188:401, 2003

58. Ferrazzi E, Torri V, Trio D, et al: Sonographic endometrial thickness: a useful test to predict atrophy in patients with postmenopausal bleeding: an Italian multicenter study. Ultrasound Obstet Gynecol 7:315, 1996

59. Tabor A, Watt HC, Wald NJ: Endometrial thickness as a test for endometrial cancer in women with postmenopausal vaginal bleeding. Obstet Gynecol 99:663, 2002

60. Bren L: The estrogen and progestin dilemma: new advice, labeling guidelines. FDA Consum 37:10, 2003

61. Narayan H, Buckett W, McDougall W, et al: Pregnancy after fifty: profile and pregnancy outcome in a series of elderly multigravidae. Eur J Obstet Gynecol Reprod Biol 47:47, 1992

62. Ventura SJ, Abma JC, Mosher WD, et al: Revised pregnancy rates, 1990–97, and new rates for 1998–99: United States. Natl Vital Stat Rep 52:1, 2003

63. Pisani G, Facioni L, Fiorani F, et al: Psychosexual problems in menopause. Minerva Ginecol 50:77, 1998

64. Loprinzi CL, Kugler JW, Sloan JA, et al: Venlafaxine in management of hot flashes in survivors of breast cancer: a randomised controlled trial. Lancet 365:2059, 2000

65. Barton D, La VB, Loprinzi C, et al: Venlafaxine for the control of hot flushes: results of a longitudinal continuation study. Oncol Nurs Forum 29:33, 2002

66. Stearns V, Beebe KL, Iyengar M, et al: Paroxetine controlled release in the treatment of menopausal hot flashes: a randomized controlled trial. JAMA 289:2827, 2003

67. Loprinzi CL, Sloan JA, Perez EA, et al: Phase III evaluation of fluoxetine for treatment of hot flashes. J Clin Oncol 20:1578, 2002

68. Weitzner MA, Moncello J, Jacobsen PB, et al: A pilot trial of paroxetine for the treatment of hot flashes and associated symptoms in women with breast cancer. J Pain Symptom Manage 23:337, 2002

69. Guttuso T Jr, Kurlan R, McDermott MP, et al: Gabapentin's effects on hot flashes in postmenopausal women: a randomized controlled trial. Obstet Gynecol 101:337, 2003

70. Wren BG, Brown LB: A double-blind trial with clonidine and a placebo to treat hot flushes. Med J Aust 144:369, 1986

71. Sonnendecker WY, Polakow ES, Gerdes L: Psycho-endocrine differences and correlations in symptomatic and asymptomatic climacteric women — the possible role of prolactin. S Afr Med J 60:661, 1981

72. Barton DL, Loprinzi CL, Quella SK, et al: Prospective evaluation of vitamin E for hot flashes in breast cancer survivors. J Clin Oncol 16:495, 1998

73. Kronenberg F, Fugh-Berman A: Complementary and alternative medicine for menopausal symptoms: a review of randomized, controlled trials. Ann Intern Med 137:805, 2002

74. Wuttke W, Seidlova-Wuttke D, Gorkow C: The *Cimicifuga* preparation BNO 1055 vs. conjugated estrogens in a double-blind placebo-controlled study: effects on menopause symptoms and bone markers. Maturitas 44:S67, 2003

75. Loprinzi FL, Michalak JC, Quella SK, et al: Megestrol acetate for the prevention of hot flashes. N Engl J Med 331:347, 1994

76. Leonetti HB, Longo S, Anasti JN: Transdermal progesterone cream for vasomotor symptoms and postmenopausal bone loss. Obstet Gynecol 94:225, 1999

77. Beral V, Banks E, Reeves G, et al: Breast cancer and hormone-replacement therapy: the Million Women Study. Lancet 362:1330, 2003

78. Irvin JH, Domar AD, Clark C, et al: The effects of relaxation response training on menopausal symptoms. J Pyschosom Obstet Gynaecol 17:202, 1996

79. Freedman RR, Woodward S: Behavioral treatment of menopausal hot flushes: evaluation by ambulatory monitoring. Am J Obstet Gynecol 167:436, 1992

80. Ivarsson T, Spetz AC, Hammar M: Physical exercise and vasomotor symptoms in postmenopausal women. Maturitas 29:139, 1998

81. Pinn G: Herbs used in obstetrics and gynaecology. Aust Fam Physician 30:351, 2001

82. Shifren J, Braunstein GD, Simon JA, et al: Transdermal testosterone treatment in women with impaired sexual function after oophorectomy. N Engl J Med 343:682, 2000

83. Lobo RA, Rosen RC, Yang HM, et al: Comparative effects of oral esterified estrogens with and without methyltestosterone on endocrine profiles and dimensions of sexual function in postmenopausal women with hypoactive sexual desire. Fertil Steril 79:1341, 2003

84. Schmidt PJ, Nieman LK, Danaceau MA, et al: Estrogen replacement in perimenopause-related depression: a prelimnary report. Am J Obstet Gynecol 183:414, 2000

85. Derman RJ, Dawood MY, Stone S: Quality of life during sequential hormone replacement therapy. Int J Fertil Menopaus Stud 40:73, 1995

86. Schneider LS, Small GW, Clary CM: Estrogen replacement therapy and antidepressant response to sertraline in older depressed women. Am J Geriatr Psychiatry 9:393, 2001

87. Hulley S, Grady D, Bush T, et al: Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA 280:605, 1998

88. Rossouw JE, Anderson FL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

89. Anderson GL, Limacher M, Assaf AR, et al: Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. JAMA 291:1701, 2004

90. Manson JE, Hsia J, Johnson KC, et al: Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med 349:523, 2003

91. Wassertheil-Smoller S, Hendrix SL, Limacher M, et al: Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. JAMA 289:2673, 2003

92. Shumaker SA, Legault C, Rapp SR, et al: Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 289:2651, 2003

93. Chlebowski RT, Hendrix SL, Langer RD, et al: Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. JAMA 289:3243, 2003

94. Col NF, Pauker SG: The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? Ann Intern Med 139:923, 2003

95. Rapp SR, Espeland MA, Shumaker SA, et al: Effect of estrogen plus progestin on global cognitive function in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. JAMA 289:2663, 2003

96. Cauley JA, Robbins J, Chen Z, et al: Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial. JAMA 290:1729, 2003

97. Anderson GL, Judd HL, Kaunitz AM, et al: Effects of estrogen plus progestin on gynecologic cancers and associated diagnostic procedures: the Women's Health Initiative Investigators. JAMA 290:1739, 2003

98. Shumaker SA, Legault C, Kuller L, et al: Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women. Women's Health Initiative Memory Study. JAMA 291:2941, 2004

99. Espeland MA, Rapp SR, Shumaker SA, et al: Conjugated equine estrogens and global cognitive function in postmenopausal women. Women's Health Initiative Memory Study. JAMA 291:2959, 2004

100. Chlebowski RT, Wactawski-Wende J, Ritenbaush C, et al: Estrogen plus progestin and colorectal cancer in postmenopausal women. N Engl J Med 350:991, 2004

101. Jick H, Derby LE, Myers MW, et al: Risk of hospital admission of idiopathic venous thromboembolism among users of postmenopausal oestrogens. Lancet 348:981, 1996

102. Daly E, Vessey MP, Hawkins MM, et al: Risk of venous thromboembolism in users of hormone replacement therapy. Lancet 348:977, 1996

103. Grodstein F, Stampfer MJ, Goldhaber SZ, et al: Prospective study of exogenous hormones and risk of pulmonary embolism in women. Lancet 348:938, 1996

104. Grady D, Gebretsadik T, Kerlikowske K, et al: Hormone replacement therapy and endometrial cancer risk: a meta-analysis. Obstet Gynecol 85:304, 1995

105. Schairer C, Lubin J, Troisi R, et al: Estrogen-progestin replacement and risk of breast cancer. JAMA 284:691, 2000

106. Ross RK, Paganini-Hill A, Wan PC, et al: Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. J Natl Cancer Inst 92:328, 2000

107. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 350:1047, 1997

108. Scarabin PY, Alhenc-Gelas M, Plu-Bureau G, et al: Effects of oral and transdermal estrogen/progesterone regimens on blood coagulation and fibrinolysis in postmenopausal women: a randomized controlled trial. Arterioscler Thromb Vasc Biol 17:3071, 1997

109. Perez Gutthann S, Garcia Rodriguez LA, Castellsague J, et al: Hormone replacement therapy and risk of venous thromboembolism: population based case-control study. BMJ 314:796, 1997

110. Van Erpecum KJ, Van Berge Henegouwen GP, Verschoor L, et al: Different hepatobiliary effects of oral and transdermal estradiol in postmenopausal women. Gastroenterology 100:482, 1991

111. Sendag F, Karadadas N, Ozsener S, et al: Effects of sequential combined transdermal and oral hormone replacement therapies on serum lipid and lipoproteins in postmenopausal women. Arch Gynecol Obstet 266:38, 2002

112. Chen FP, Lee N, Soong YK, et al: Comparison of transdermal and oral estrogen-progestin replacement therapy: effects on cardiovascular risk factors. Menopause 8:347, 2001

113. Decensi A, Omodei U, Robertson C, et al: Effect of transdermal estradiol and oral conjugated estrogen on C-reactive protein in retinoid-placebo trial in healthy women. Circulation 106:1224, 2002

114. Shulman LP, Yankov V, Uhl K: Safety and efficacy of a continuous once-a-week 17beta-estradiol/levonorgestrel transdermal system and its effects on vasomotor symptoms and endometrial safety in postmenopausal women: the results of two multicenter double-blind, randomized, controlled trials. Menopause 9:195, 2002

115. O'Connell D, Robertson J, Henry D, et al: A systematic review of the skeletal effects of estrogen therapy in postmenopausal women: II. An assessment of treatment effects. Climacteric 1:112, 1998

116. Effects of hormone replacement therapy on endometrial histology in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. The Writing Group for the PEPI Trial. JAMA 275:370, 1996

117. Raudaskoski T, Tapanainen J, Tomas E, et al: Intrauterine 10 µg and 20 µg levonorgestrel systems in postmenopausal women receiving oral oestrogen replacement therapy: clinical, endometrial and metabolic response. BJOG 109:136, 2002

118. Wildemeersch D, Schacht E, Wildemeersch P: Performance and acceptability of intrauterine release of levonorgestrel with a miniature delivery system for hormonal substitution therapy, contraception and treatment in peri and postmenopausal women. Maturitas 44:237, 2003

119. Franchin R, De Ziegler D, Bergeron C, et al: Transvaginal administration of progesterone. Obstet Gynecol 90:396, 1997

120. Sherman SL: Premature ovarian failure in the fragile X syndrome. Am J Med Genet 97:189, 2000

121. Luborsky JL, Meyer P, Sowers MF, et al: Premature menopause in a multi-ethnic population study of the menopause transition. Hum Reprod 18:199, 2003

122. Gersak K, Meden-Vrtovec H, Peterlin B: Fragile X permutation in women with sporadic premature ovarian failure in Slovenia. Hum Reprod 18:1637, 2003

123. Hundscheid RD, Sistermans EA, Thomas CM, et al: Imprinting effect in premature ovarian failure confined to paternally inherited fragile X permutations. Am J Hum Genet 66:413, 2000

124. Alper MM, Jolly EE, Garner PR: Pregnancies after premature ovarian failure. Obstet Gynecol 67:59S, 1986

125. Lydic ML, Liu JH, Rebar RW, et al: Success of donor oocyte in vitro fertilization-embryo transfer in recipients with and without premature ovarian failure. Fertil Steril 65:98, 1996

126. Schwingl PJ, Ory HW, Visness CM: Estimates of the risk of cardiovascular death attributable to low-dose oral contraceptives in the United States. Am J Obstet Gynecol 180:241, 1999

127. Rebar RW, Connolly HV: Clinical features of young women with hypergonadotropic amenorrhea. Fertil Steril 53:804, 1990

128. van der Schouw YT, van der Graaf Y, Steyerberg EW, et al: Age at menopause as a risk factor for cardiovascular mortality. Lancet 347:714, 1996

129. Snowdon DA, Kane RL, Beeson WL, et al: Is early natural menopause a biologic marker of health and aging? Am J Public Health 79:709, 1989

130. Rannevik G, Jeppson S, Johnell O, et al: A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG, and bone mineral density. Maturitas 21:103, 1995

Acknowledgments

Figure 1 Modified from "Executive Summary: Stages of Reproductive Aging Workshop (STRAW)," by M. R. Soules, S. Sherman, E. Parrott, et al., in *Fertility and Sterility* 76:874, 2001.

Figure 3 Seward Hung.

XII URINARY INCONTINENCE AND THE OVERACTIVE BLADDER

Robert L. Barbieri, m.d.

The involuntary loss of urine is an extremely common problem in women—more common than Alzheimer disease or osteoporosis. Urinary incontinence affects approximately 15% of women younger than 65 years, 25% of women older than 65 years, and 50% of women who are nursing home residents.[1,2] It is underdiagnosed and undertreated; patients may fail to report it, and physicians may fail to address it.[3]

Pathogenesis

The normal pattern of urinary voiding requires the interaction of the central nervous system, the sacral parasympathetic and thoracolumbar sympathetic systems, the bladder muscle (detrusor muscle), the urinary sphincter muscle, and the mechanical support of the pelvic fascia. In the resting state, urine collects in the bladder, the bladder muscle is quiescent, and the tone of the urinary sphincter muscle is high, ensuring that urine remains in the bladder. Urination is initiated when signals from the CNS induce the bladder muscle to contract and the urethral sphincter muscle to relax. At the completion of voiding, the bladder muscle relaxes and the urethral sphincter contracts. A key feature of the urinary system is that in the normal resting state, the urethral pressure is greater than the pressure in the bladder. Reversal of that differential—whether from a rise in bladder pressure, a decrease in urethral pressure, or both—will result in incontinence.

Urinary Incontinence in the Nongeriatric Population

In the nongeriatric population, the five main causes of urinary incontinence are (1) loss of fascial support of the urethra (stress incontinence), (2) an overactive bladder muscle (detrusor instability), (3) intrinsic sphincter deficiency, (4) neuropathies, and (5) urinary tract fistulas. Many women with incontinence have both stress incontinence and bladder muscle overactivity. In a study of 303 women with urinary incontinence, 43% were diagnosed as having stress incontinence, 21% as having bladder muscle overactivity, and 36% as having both disorders.[4] In some women, the stress incontinence is predominant, with bladder muscle overactivity playing less of a role; in other women, the bladder muscle overactivity is predominant, and there is less of a contribution from stress incontinence. Race may be a factor: African-American women have lower rates of stress incontinence than Hispanic and white women but have higher rates of detrusor instability.[5]

The initial approach to urinary incontinence should include a history; physical examination; and laboratory tests, including a urinalysis and urine culture. In the history, the patient should be asked to describe the episodes of incontinence. Women with stress incontinence typically report urinary leakage with events that increase intra-abdominal pressure—for example, coughing, laughing, sneezing, and the lifting of heavy weights. Women with detrusor instability often report a sudden and intense urge to void that occurs just before the loss of urine. Symptoms such as frequency, nocturia, hesitancy, and terminal dribbling are not specific to incontinence from any one cause.

A voiding diary is a reliable method to establish the severity of urinary incontinence. In the voiding diary, the patient reports the time of continent and incontinent voids, the volume of continent voids, and any events that precipitate incontinence. To facilitate measurement of continent voids, patients can be given a graduated plastic receptacle (a so-called hat) to place on their toilet seat at home.

In the history, the physician should assess for substances that may contribute to incontinence. Caffeine, which is both a diuretic and a bladder irritant, may worsen incontinence in some cases. Other possible dietary bladder irritants include alcohol, citrus or other highly acidic fruit, tomatoes, spicy foods, dairy products, and sugar. Alpha blockers and some muscle relaxants (e.g., dantrolene) can exacerbate incontinence by relaxing the urinary sphincter.

The assessment should include a complete review of the patient's obstetric history, including the number of deliveries, their route, and any complications. Details of any abdominal or pelvic surgery should also be obtained.

The physical examination should include inspection of the vagina for mucosal atrophy and herniation of the pelvic structures.

Urinalysis and culture are necessary to screen for urinary tract infection or glucosuria. If those conditions are found, they should be treated before one proceeds with further evaluation of urinary incontinence. Curing the urinary tract infection or controlling the diabetes may correct the urinary incontinence in such cases.

Additional initial tests for incontinence include a clinical stress test and measurement of postvoid residual urinary volume. One approach to these two tests is to ask the patient to void and then catheterize the bladder to determine the postvoid residual urinary volume. If the postvoid residual volume is more than 200 ml, the physician should assess for a neurogenic process or a bladder outflow obstruction. Using the same catheter, 300 ml of fluid (sterile water or sterile saline) at body temperature can be instilled into the bladder. The patient is then asked to cough while in the supine and upright positions. Immediate leakage of urine with the cough is diagnostic of stress incontinence.

urodynamic testing with multichannel equipment

Urodynamic testing with multichannel pressure recording devices is the gold standard for identifying both stress incontinence and bladder overactivity. In this procedure, the intravesical pressure is directly measured by means of a bladder catheter, and intra-abdominal pressure is measured by use of a vaginal or rectal probe. Subtracting intra-abdominal pressure from intravesical pressure yields the true detrusor pressure. A separate channel in the bladder catheter allows fluid to be instilled into the bladder [*see Figure 1*].

Urodynamic testing is not necessary for every incontinent patient. Although widely available, the testing is expensive and time consuming. Some clinicians prefer to forgo urody-

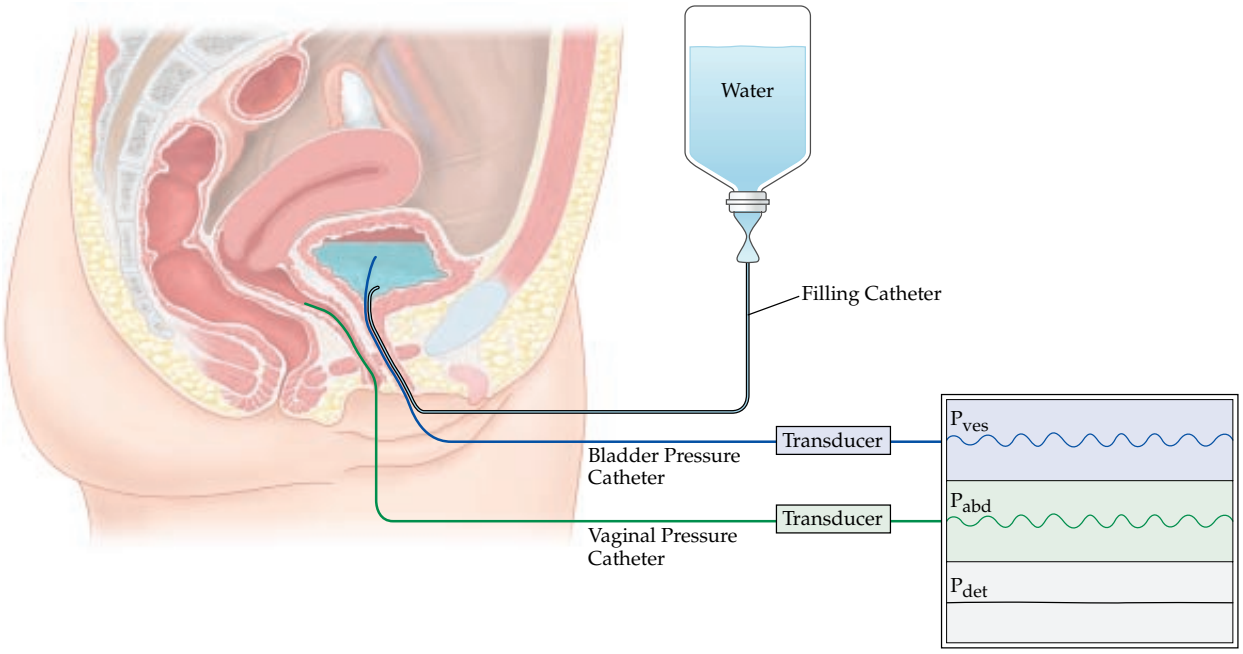


Figure 1 **Subtracted cystometry. Intra-abdominal pressure (Pabd) is measured with an intravaginal or intrarectal pressure catheter. Bladder pressure (Pves) is measured with an intravesical catheter. Subtraction of Pabd from Pves yields the true detrusor pressure (Pdet). Fluid can be instilled into the bladder through a separate channel in the intravesical channel. Bladder contractions that occur as the bladder is filling and that the patient is unable to completely suppress indicate detrusor instability.**

namic testing and treat on the basis of clinical diagnosis, whereas others (especially urogynecologists) use the testing routinely.

STRESS INCONTINENCE

Stress incontinence refers to a sudden involuntary loss of urine, usually secondary to a sudden increase in intra-abdominal pressure, in the absence of a bladder contraction. Intra-abdominal pressure can increase as a result of coughing, sneezing, running, climbing stairs, standing, or lifting heavy weights.

Normally, the pelvic fascia holds the urethral sphincter in an intra-abdominal position; consequently, any increase in intra-abdominal pressure is transmitted equally to the urethra and bladder, and the urethral sphincter remains closed. Damage to the pelvic fascia, laxity in the pelvic ligaments, and partial pelvic muscle tears and denervation—all common sequelae of childbirth—can cause the urethra to drop below the pelvic floor. If intra-abdominal pressure then increases suddenly, the pressure will be transmitted primarily to the bladder; bladder pressure will then exceed urethral pressure, the urethral sphincter will open involuntarily, and urine will spill.

Diagnosis

If the history suggests stress incontinence, the vaginal walls should be examined for laxity of the pelvic structures. Possible underlying abnormalities in stress incontinence include cystocele (a hernia of the bladder into the anterior vaginal wall), rectocele (a hernia of the rectum into the posterior vaginal wall), enterocele (a hernia of the small bowel into the posterior apex of the vaginal wall), and uterine prolapse (a hernia of the uterus into the lower portion of the vagina).

Treatment

If the physical examination reveals the cervix at the vaginal opening (uterine prolapse) or a cystocele, rectocele, or enterocele at the vaginal opening, the patient should be referred to a gynecologist or urogynecologist for surgical repair. If the physical examination reveals no significant herniations, a nonsurgical approach can be recommended.

Nonsurgical treatment The nonsurgical approach to the treatment of stress incontinence includes exercises to strengthen the pelvic floor musculature (Kegel exercises) and electrical stimulation of the pelvic muscles. The Kegel exercise is the voluntary contraction of the pubococcygeal muscles in sets of 10 contractions, five to 10 times daily. Contraction of the pubococcygeal muscle can be taught by asking the patient to stop her stream of urine while voiding. This requires contraction of pelvic muscles and relaxation of the abdominal muscles. Many patients have difficulty isolating the correct muscles and may benefit from instruction sheets on performing the exercises; one source for such instructional material is http://www.niddk.nih.gov/health/urolog/uibcw/exerc/exerc.htm.

Kegel exercises alleviate the symptoms in about 70% of women with stress incontinence.[6] Improvement in continence may not become apparent for 6 to 12 weeks, but once attained, the benefits may be maintained for years.[7]

Vaginal cones can also be used to teach women to contract pelvic floor muscles without contracting the abdominal muscles.[8] Vaginal cones come in sets of five cones, each of a different weight [*see Figure 2*]. Depending on the manufacturer, the cones may be all the same size or may decrease in diameter as they increase in weight. The patient inserts the lightest cone

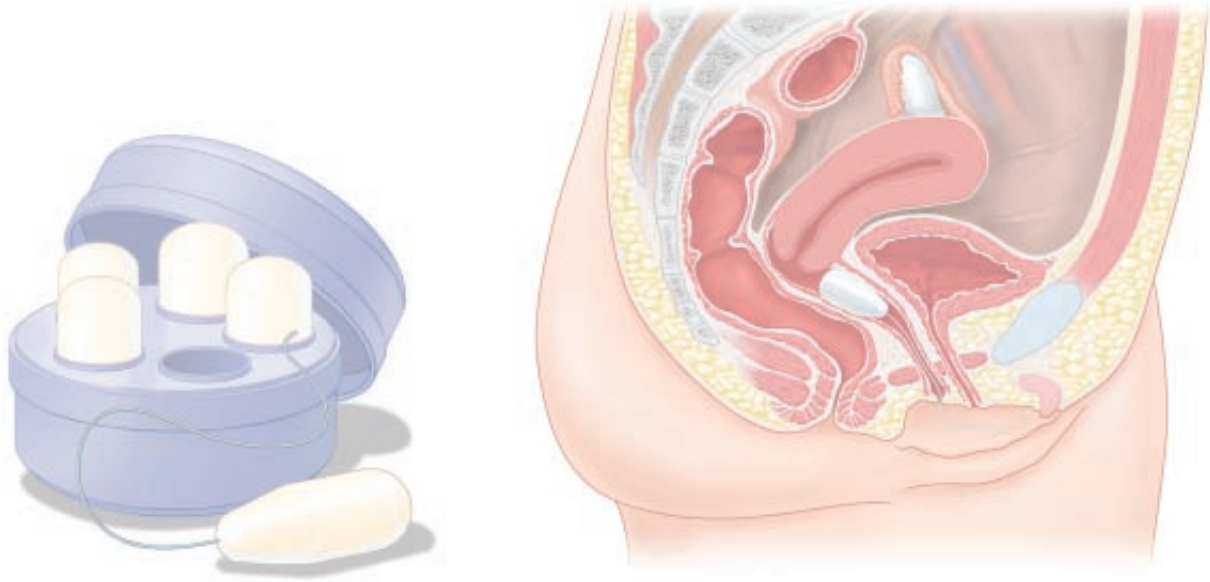


Figure 2 **Vaginal cones. Women with stress incontinence can use vaginal cones to learn how to contract their pelvic floor muscles. The cones come in sets of graduated weight and are shaped like a tampon, with a string to facilitate removal from the vagina (*a*). The patient inserts the cone (*b*) and keeps it in her vagina for 15 minutes. Successively heavier cones are used as the pelvic floor muscles increase in strength.**

into the vagina and uses the pelvic floor muscles to hold the cone in the vagina while walking. Once those muscles can hold the cone in place for at least 15 minutes, the patient progresses to the next heavier cone. For some women, the cones are superior to Kegel exercises for strengthening the pelvic floor musculature. As with the Kegel exercises, patients may not experience benefit for a number of weeks.

An alternative to Kegel exercises and vaginal cones is to electrically stimulate the paravaginal musculature. An approved device for strengthening the pelvic musculature consists of a probe that is placed in the vagina twice daily for 15 to 30 minutes. The probe stimulates the afferent fibers of the pudendal nerve, contracting the muscles of the pelvic floor and periurethral muscles. This stimulation strengthens the muscles that maintain the urethral pressure profile.

Estrogen therapy has been suggested for women with atrophic vaginitis and stress incontinence. Estrogen has not been documented to improve symptoms of stress incontinence,[9,10] but some researchers have speculated that topical estrogen may be effective even if oral estrogen is not[11]; and the combination of estrogen and pelvic floor exercises has been found to be more effective than exercises alone.[12] In one study, imipramine, 25 mg three times a day, produced significant improvement or cure in 60% of patients with stress incontinence.[13]

Surgical treatment If nonsurgical measures fail, the patient with stress incontinence should be referred to a urogynecologist for further evaluation. Incontinence surgery is very effective in the treatment of stress incontinence. In women older than 65 years, incontinence surgery has an overall mortality of less than 0.3%,[14] and surgery for stress incontinence has high cure rates.[15]

The current state-of-the art surgical procedure for stress incontinence is the tension-free vaginal tape procedure.[16] This procedure can be performed on an outpatient basis using local or regional anesthesia and conscious sedation. In this technique, a strip of polypropylene mesh (the tape) is used to create a urethral sling. The tape is inserted under the wall of the vagina, beneath the midurethra, and the ends of the tape are passed retropubically up to two small incisions on the lower abdomen near the superior border of the pubic hair. The tape is elevated just enough to eliminate urine leakage with coughing. It stays in place without sutures because of its Velcro-like surface, and it is further anchored by postoperative fibrosis. The tape allows increases in intra-abdominal pressure to be efficiently transmitted to the urethra, preventing loss of urine. In one study of women with stress incontinence, the tension-free vaginal tape procedure resulted in cure in about 85% of patients.[17] Other bladder neck suspension procedures for stress incontinence include the Burch colposuspension and the Raz needle vaginal suspension.

DETRUSOR INSTABILITY—OVERACTIVE BLADDER MUSCLE

Detrusor instability is defined as spontaneous or provoked bladder contractions during the filling phase of cystometry that cannot be completely suppressed by a neurologically normal patient.[18] A problem with this definition is that it requires an invasive test, cystometry, for the diagnosis. Some authorities prefer the term overactive bladder muscle to describe the condition. In the nonvoiding state, the bladder muscle should not contract spontaneously. Spontaneous contractions of the bladder muscle can raise the bladder pressure above that of the urethral sphincter, resulting in the involuntary loss of urine. Bladder muscle from women with detrusor instability is more sensitive to electrical and acetylcholine stimulation than is bladder muscle from healthy women.[19]

Detrusor hyperreflexia is defined as an overactive bladder muscle caused by a neurologic disorder, such as multiple scle-

rosis, Parkinson disease, cerebrovascular disease, or spinal cord injury. As many as 80% of women with multiple sclerosis have detrusor hyperreflexia.[20]

Diagnosis

Women with detrusor instability typically complain of urinary urgency (a desire to urinate immediately), urinary frequency (a need to urinate every half hour or hour), urge incontinence, and nocturia. A severe urge to void followed by incontinence at rest is common. Incontinent episodes typically involve the loss of large amounts of urine. Women who void 10 or more times a day often have detrusor instability. Physical examination discloses no specific findings in women with overactive bladder. However, because stress incontinence and overactive bladder can coexist, the physical findings associated with stress incontinence may be observed in women with overactive bladder muscle.

Treatment

Bladder retraining and acetylcholine blockade are the two most effective approaches to the treatment of overactive bladder. Surgery is not recommended for women with bladder overactivity.

Bladder retraining The goal of bladder retraining is to elicit behavioral changes that will lead to an increase in bladder capacity and will prolong the interval between episodes of voiding. At the beginning of bladder retraining, the patient is asked to urinate approximately once every hour. No nighttime schedule is recommended. As training progresses, the woman is asked to void once every 90 minutes, then once every 2 hours. Millard and Oldenburg reported that bladder retraining alleviated symptoms in as many as 75% of women with overactive bladder.[21] Biofeedback using a vaginal or anorectal probe may help some women better control urge incontinence.[22] Many physical therapy programs teach bladder retraining and biofeedback techniques.

Behavioral treatment may be as effective as drug treatment for many women with bladder overactivity. In one trial, behavioral therapy resulted in an 86% improvement in incontinent episodes, versus a 69% improvement with oxybutynin.[23]

Drug therapy Detrusor muscle stimulation is mediated by cholinergic fibers. Anticholinergic agents [*see Table 1*] can decrease detrusor muscle activity, reduce the urge to void, and result in improvement in women with urge incontinence. Traditional anticholinergic agents, such as oxybutynin, block detrusor muscle stimulation, but they also block cholinergic effects in the salivary glands, causing dry mouth. The typical dose of immediate-release oxybutynin is 2.5 or 5 mg three or four times a day. Contraindications to oxybutynin include narrow-angle glaucoma and cardiac arrhythmia. A new, extended-release formulation of oxybutynin (starting dose, 5 mg once daily, titrated to 20 to 30 mg once daily) appears to have fewer side effects, such as dry mouth, than the original rapid-acting formulation of the drug.[24] In a prospective, randomized, controlled trial, extended-release oxybutynin proved significantly more effective than tolterodine in improving incontinence and urinary frequency.[25]

Tolterodine, one of a new class of anticholinergic agents, has anticholinergic effects on the bladder but has little effect on the salivary glands. In one study, the mean number of incontinence episodes was reduced by 47% in women treated with tolterodine (2 mg twice daily), compared with 17% in women treated with placebo.[26] For young women, a dosage of 2 mg twice daily is usually effective in reducing the frequency of voiding. For older women, dosages of tolterodine as low as 1 mg daily may control the problem of frequent voiding.

Although alpha blockers are not often used for treating overactive bladder in women, these agents have a modulating effect on bladder smooth muscle. One prospective study found doxazosin [*see Table 1*], 2 mg at bedtime, to be effective in women with urinary frequency and urgency.[27]

INTRINSIC SPHINCTER DEFICIENCY

Intrinsic sphincter deficiency (ISD) is a form of urinary incontinence caused by the failure of the urethra to close completely. Urethral dysfunction in these patients results from mucosal and muscular atrophy and denervation. Postoperative scarring of the urethra can also cause ISD. Urinary leakage in ISD is often continual and can occur without an increase in intra-abdominal pressure. Surgical treatments for ISD include sling procedures, such as the tension-free vaginal tape procedure, or the periurethral injection of bulking agents, such as glutaraldehyde cross-linked bovine collagen (Contigen)[28] or carbon beads suspended in gel (Durasphere)[29] [*see Figure 3*]. Short-term success rates with injectable agents are very high, but long-term durability remains in question.[30]

Table 1 Drugs Used in the Treatment of Overactive Bladder

Class	*Agent (Brand Name)*	*Dosage*	*Cost*	*Comment*
Anticholinergic agents	Oxybutynin	2.5–5 mg, t.i.d.–q.i.d	5 mg t.i.d.: $10–19.99/mo	Dry mouth a common side effect
	Oxybutynin, extended release (Ditropan XL)	5–30 mg q.d.	5 mg: $100.99/mo 10 mg: $107.99/mo	Dry mouth a common side effect
	Tolterodine (Detrol)	1 mg q.d.–2 mg b.i.d.	2 mg b.i.d.: $70–79.99/mo	Lower dose often effective in older women
	Tolterodine, extended release (Detrol LA)	2–4 mg q.d.	2 mg: $100.99/mo	—
Alpha blocker	Doxazosin (Cardura)	2 mg h.s.	2 mg q.d.: $20—29.99/mo	—

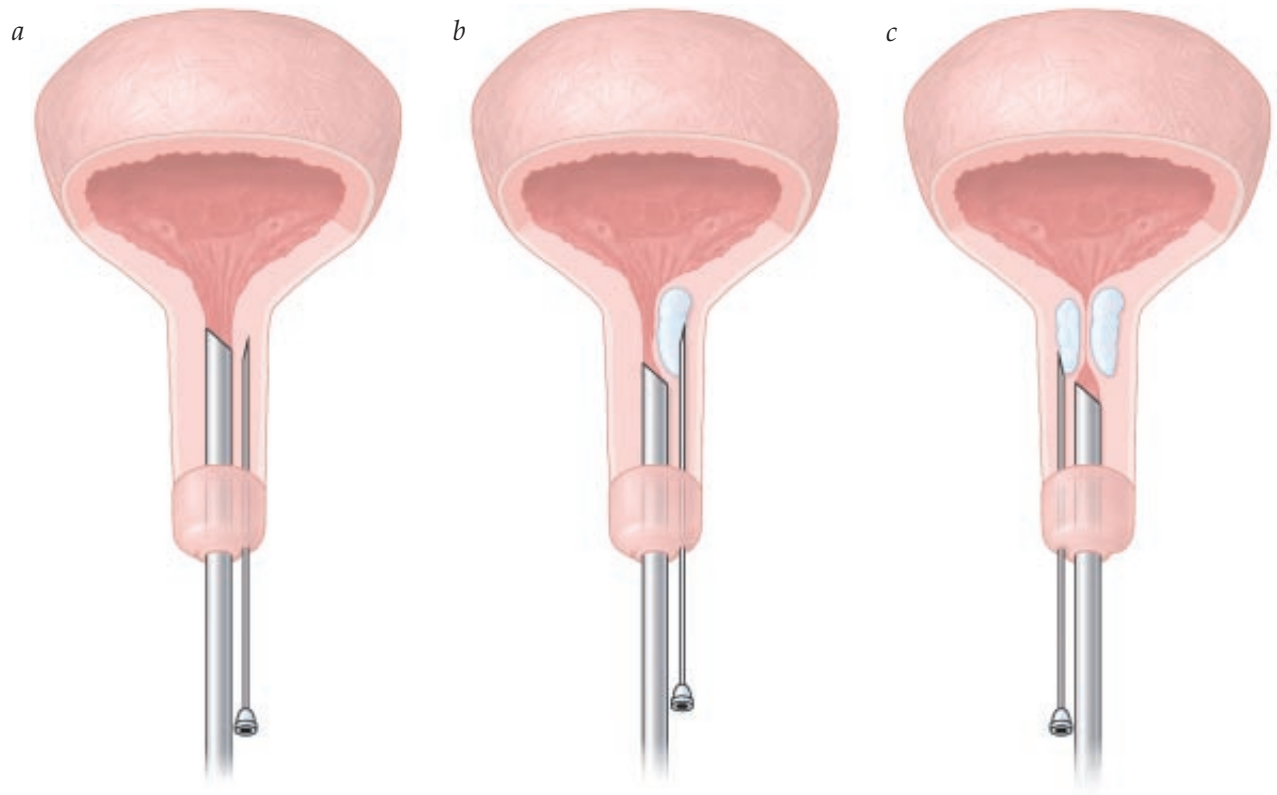


Figure 3 **Injection of bulking agents to treat intrinsic sphincter deficiency. The needle is advanced to the proximal urethra just below the bladder neck (*a*), and the bulking agent is then injected (*b*). Injection of the agent bilaterally closes off the proximal urethra (*c*).**

OVERFLOW INCONTINENCE

Rarely, neurologic problems that result in denervation of the detrusor muscle can lead to overfilling of the bladder and so-called overflow incontinence. Overdistention of the bladder can also be caused by outflow obstruction. Pharmacologic contributors to overflow incontinence include calcium channel blockers, which can relax the detrusor muscle, and alpha agonists, which can increase urethral resistance. Denervation of the detrusor muscle leading to overflow incontinence can result from cerebrovascular accidents, multiple sclerosis, spinal cord injury, cauda equina tumors, diabetes, and pelvic nerve damage during surgery.

A simple way to test the status of the sacral nerves is to look for the anal wink reflex. If the perianal region is lightly stroked with the wooden end of a cotton-tipped applicator, the anal sphincter should contract. Presence of this reflex suggests that the sacral dermatomes S2, S3, and S4 are intact. If an anal wink is not present, a complete neurologic evaluation is warranted.

URINARY TRACT FISTULAS

Fistulas between the bladder and vagina are a rare cause of incontinence. Most urinary tract fistulas occur after a pelvic surgical procedure, such as cesarean section, surgical vaginal delivery (e.g., involving repair of severe lacerations or forceps use), hysterectomy, or bladder surgery. Spontaneous fistulas are uncommon. Incontinence resulting from fistulas usually involves chronic, continuous leakage of urine.

In many cases, fistulas between the bladder and vagina can be observed on speculum examination of the vagina. Small fistulas can be detected by giving the patient a dose of oral pyridium and placing a tampon in the vagina. If the tampon turns orange after 2 hours, a fistula between the urinary tract and vagina is likely.

Urinary Incontinence in the Geriatric Population

As many as half of the women in nursing homes have urinary incontinence. Aging is associated with decreased urethral pressure and increased involuntary detrusor contractility, which increase the risk of symptomatic bladder overactivity. In addition, in the elderly, urine production increases during the night. Resnick has suggested the mnemonic DIAPPERS for the evaluation of incontinence in the geriatric population.[31] This mnemonic summarizes the most common causes of incontinence in the elderly: delirium, infection, atrophic urethritis, pharmaceuticals that interfere with bladder and urethral function, psychiatric causes (depression), excess urinary output, restricted mobility, and stool impaction. Treatment of these problems can often significantly alleviate the incontinence.

The incontinent woman with cognitive impairment often responds well to prompted voiding. One program of prompted voiding requires that the woman be asked every 2 hours if she would like to void. If she responds in the affirmative, she is escorted to the toilet.

References

1. Milsom I, Ekelund P, Molander U, et al: The influence of age, parity, oral contraception, hysterectomy and menopause on the prevalence of urinary incontinence in women. J Urol 149:1459, 1993

2. Ouslander JG, Schnelle JF: Incontinence in the nursing home. Ann Intern Med 122:438, 1995

3. Stoddart H, Donovan J, Whitley E, et al: Urinary incontinence in older people in the community: a neglected problem? Br J Gen Pract 51:548, 2001

4. Walter S, Olesen KP: Urinary incontinence and genital prolapse in the female: clinical, urodynamic and radiological examinations. Br J Obstet Gynaecol 89:393, 1982

5. Duong TH, Korn AP: A comparison of urinary incontinence among African American, Asian, Hispanic, and white women. Am J Obstet Gynecol 184:1083, 2001

6. Tchou DC, Adams C, Varner RE, et al: Pelvic-floor musculature exercises in treatment of anatomical urinary stress incontinence. Phys Ther 68:652, 1988

7. Bo K, Talseth T: Long-term effect of pelvic floor muscle exercise 5 years after cessation of organized training. Obstet Gynecol 87:261, 1996

8. Peattie AB, Plevnik S, Stanton SL: Vaginal cones: a conservative method of treating genuine stress incontinence. Br J Obstet Gynecol 95:1049, 1988

9. Fantl JA, Bump RC, Robinson D, et al: Efficacy of estrogen supplementation in the treatment of urinary incontinence. Obstet Gynecol 88:745, 1996

10. Grady D, Brown JS, Vittinghoff E, et al: Postmenopausal hormones and incontinence: the heart and estrogen/progestin replacement study. Obstet Gynecol 97:116, 2001

11. Ouslander JG, Greendale GA, Uman G, et al: Effects of oral estrogen and progestin on the lower urinary tract among female nursing home residents. J Am Geriatr 49:803, 2001

12. Ishiko O, Hirai K, Sumi T, et al: Hormone replacement therapy plus pelvic floor muscle exercise for postmenopausal stress incontinence: a randomized, controlled trial. J Reprod Med 46:213, 2001

13. Lin HH, Sheu BC, Lo MC, et al: Comparison of treatment outcomes for imipramine for female genuine stress incontinence. Br J Obstet Gynaecol 106:1089, 1999

14. Sultana CJ, Campbell JW, Pisanelli WS, et al: Morbidity and mortality of incontinence surgery in elderly women: an analysis of Medicare data. Am J Obstet Gynecol 176:344, 1997

15. Leach GE, Dmochowski RR, Appell RA, et al: Female Stress Urinary Incontinence Clinical Guidelines Panel summary report on surgical management of female stress urinary incontinence. J Urol 158:875, 1997

16. Bezerra CA, Bruschini H: Suburethral sling operations for urinary incontinence in women (Cochrane Review). Cochrane Database Syst Rev 3:CD001754, 2001

17. Ulmsten U, Johnson P, Rezapour M: A three year follow up of tension free vaginal tape for surgical treatment of female stress urinary incontinence. Br J Obstet Gynecol 106:345, 1999

18. Abrams P, Blaivas JG, Stanton SL, et al: The standardization of terminology of lower urinary tract function. Scan J Urol Nephrol Suppl 114:5, 1988

19. Kinder RB, Mundy AR: Pathophysiology of idiopathic detrusor instability and detrusor hyperreflexia: an in vivo study of human detrusor muscle. Br J Urol 60:509, 1987

20. Blaivas JG, Bhimani G, Labib KB: Vesicourethral dysfunction in multiple sclerosis. J Urol 122:342, 1979

21. Millard RJ, Oldenburg BF: The symptomatic, urodynamic and psychodynamic results of bladder re-education programs. J Urol 130:715, 1983

22. Burgio KL, Engel BT: Biofeedback-assisted behavioral training for elderly men and women. J Am Geriatr Soc 38:338, 1990

23. Burgio KL, Locher JL, Goode PS, et al: Behavioral versus drug treatment for urge urinary incontinence in older women: a randomized controlled trial JAMA 280:1995, 1998

24. Anderson RU, Mobley D, Blank B, et al: Once-daily controlled versus immediate release oxybutynin chloride for urge urinary incontinence. J Urol 161:1809, 1999

25. Prospective randomized controlled trial of extended-release oxybutynin chloride and tolterodine tartrate in the treatment of overactive bladder: results of the OBJECT Study. Overactive Bladder: Judging Effective Control and Treatment Study Group. Mayo Clin Proc 76:358, 2001

26. Abrams P, Freeman R, Anderstrom C, et al: Tolterodine, a new antimuscarinic agent: as effective but better tolerated than oxybutynin in patients with an overactive bladder. Br J Urology 81:801, 1998

27. Serels S, Stein M: Prospective study comparing hyoscyamine, doxazosin, and combination therapy for the treatment of urgency and frequency in women. Neurourol Urodyn 17:31, 1998

28. Nataluk EA, Assimos DG, Kroovand RL: Collagen injections for treatment of urinary incontinence secondary to intrinsic sphincter deficiency. J Endourol 9:403, 1995

29. Lightner D, Calvosa C, Andersen R, et al: A new injectable bulking agent for treatment of stress urinary incontinence: results of a multicenter, randomized, controlled, double-blind study of Durasphere. Urology 58:12, 2001

30. Herschorn S: Current status of injectable agents for female stress urinary incontinence. Can J Urol 8:1281, 2001

31. Resnick NM: Geriatric incontinence. Urol Clin North Am 23:55, 1976

XIV APPROACH TO THE PATIENT WITH A BREAST MASS

Valerie L. Staradub, m.d.
Monica Morrow, m.d.

More than half of the patients who present to a breast clinic have the chief complaint of a breast mass.[1] The identification of a breast mass causes a great deal of anxiety in women, although the majority of breast masses are benign. The most important task of the physician who is evaluating a breast mass is to exclude the presence of malignancy. Once malignancy is ruled out, the physician must provide an accurate diagnosis, suitable treatment, and reassurance to the patient.

Assessment of Normal Organ Function

The normal breast is a mixture of epithelial (glandular) elements, stromal tissue, and fat. This heterogeneity is responsible for the lumpiness that is characteristic of normal breasts, particularly in premenopausal women. The upper outer quadrant and the inframammary ridge are usually the most nodular areas of the normal breast. In women older than 40 years, small pealike nodules can often be felt beneath the areola. These nodules represent dilated ducts and are of no clinical concern. Most normal areas of nodularity can be readily identified by their presence in both breasts.

In premenopausal women, the normal hormonal fluctuations of the menstrual cycle often result in changes in breast nodularity that may be mistaken for disease processes. The progesterone surge at ovulation results in mammary duct differentiation and alveolar epithelial cell differentiation into secretory cells. Clinically, this translates to a greater degree of nodularity in the upper outer quadrants of the breasts and may also result in breast tenderness or discomfort. Cyclical nodularity generally decreases after the onset of menses, which is the rationale for the recommendation that a patient perform breast self-examination in the week after her period, when the breasts are the least nodular. These cyclical changes in breast nodularity are often erroneously termed fibrocystic disease, but they are in fact a component of normal physiology.[2] After menopause with its concomitant withdrawal of estradiol and progesterone, the epithelial elements of the breast atrophy, making the breasts softer, less nodular, and easier to examine.

History and Physical Examination of the Patient with a Breast Mass

The key to evaluating the patient who presents with a breast mass is to determine whether a dominant mass is present and to define the level of suspicion for malignancy associated with the mass, should one be detected. These determinations are initially made on the basis of a careful history and physical examination and will direct the approach to diagnosis and management for each patient.

HISTORY

The initial step in obtaining the pertinent history is to characterize the mass by determining the mass's duration, fluctuation with the menstrual cycle, associated tenderness, and whether it has changed in size since the patient first identified it. The patient should be asked whether she has a history of breast problems, including cyst aspirations and biopsies. A menstrual history is important, including the date of the last period, any recent menstrual irregularities, use of oral contraceptives or hormone replacement therapy, and recent changes in hormone preparation.

An assessment of the patient's level of risk for breast cancer is appropriate [*see Table 1*], although the characteristics of the breast mass rather than the patient's level of risk for cancer should be the primary determinant of the appropriate workup. The characteristics of the breast mass take precedence over the assessment of cancer risk because the majority of women with breast cancer lack identifiable risk factors.[3] When eliciting a family history, it is important to obtain information on both maternal and paternal relatives, because breast cancer on either side of the family is associated with an increased level of risk.

PHYSICAL EXAMINATION

Physical examination is important to confirm the presence of a mass. Often, a mass identified by a patient or a primary care physician is actually an area of normal glandular nodularity or normal breast tissue and underlying structures. In a study of 605 women younger than 40 years, Morrow and colleagues reported that referral by a primary care provider to a surgeon for the evaluation of a breast lump led to confirmation of the presence of a dominant mass in 29% of women, whereas patient-detected masses were confirmed by the surgeon in 36% (a difference that was not statistically significant).[4]

Examination should be carried out with the patient in both an upright sitting position and a supine position. The breasts should be evaluated for symmetry with the arms relaxed and with the arms raised over the head. The presence of skin or nipple retraction, edema, or erythema should be noted. In many women, the breasts are not precisely the same size, and in some women, there may be a significant difference in size. If a size discrepancy is noted, the patient should be questioned regarding its duration. Similarly, many women have bifid or chronically inverted nipples, the latter of which occur particularly after lactation; however, bifid nipples or chronically inverted nipples are of no concern, even if the chronically inverted nipples are present in a patient who has never lactated.

Table 1 Factors Used for Assessment of Breast Cancer Risk

- Patient age
- Patient race
- Age at menarche
- Age at first live birth
- Age at menopause
- History of postmenopausal hormone replacement therapy
- Number of relatives with breast and/or ovarian cancer
  - Relationship to patient
  - Age at diagnosis
- Number of previous breast biopsies
- Pathologic findings at biopsy

Palpation of both breasts, as well as the axilla, should follow. The axilla should be examined with the patient seated and the ipsilateral arm supported to relax the pectoral muscle. Small palpable nodes are not uncommon in slender women, and any palpable nodes must be assessed for worrisome characteristics such as fixation, large size, or hardness.

Breast palpation should be performed with the patient both in the upright position and in the supine position. In the supine position, the ipsilateral arm should be placed behind the head to spread the breast tissue across the chest wall. The pads, rather than the tips, of the first three fingers should be used for palpation, and pinching of the breast tissue between the fingers should be avoided. The goal of the examination is to determine if a dominant mass is present. Dominant masses are distinguished from nodular breast tissue by having three dimensions and a texture different from the adjacent normal breast. If a mass is identified, it should be measured with a ruler, the consistency should be noted (e.g., soft, rubbery, firm, hard), and the characteristics of its margins described (e.g., well circumscribed, poorly defined) [*see Table 2*]. Fixation of the mass within the breast or to the chest wall should also be noted. If there is uncertainty whether a finding represents a true dominant mass, comparison with the mirror-image location in the opposite breast is often helpful.

Once the examination is complete, regardless of whether a mass is identified, the patient should be asked to indicate the area that concerns her. This ensures that the area of concern to the patient is not overlooked by the physician. At the conclusion of the examination, the patient can be categorized according to four possible assessments: (1) no abnormal finding is appreciated; (2) a prominent nodularity is present, but it does not have the characteristics of a dominant mass; (3) a dominant mass with clinically benign characteristics is present; and (4) a dominant mass suspicious for cancer is present. The appropriate imaging and diagnostic workup is specific to the outcome of the physical examination.

Table 2 Physical Characteristics of Benign and Malignant Breast Masses

Characteristic	*Benign*	*Malignant*
Borders	Well circumscribed	Irregular
Texture	Firm or rubbery	Hard
Mobility	Mobile	Fixed to surrounding tissue
Skin changes	None	Dimpling, retraction
Nipple changes	None	Retraction, bloody discharge, scaling

Evaluation of a Breast Mass

NO ABNORMALITY DETECTED BY PHYSICIAN

If no abnormality is detected during a clinical breast examination, even after careful examination of the area of concern, the patient should be reassured of the absence of worrisome findings. Women 40 years of age and older who have not had a mammogram within the past year should receive a mammogram to screen for nonpalpable abnormalities. In younger patients, no imaging should be recommended unless their level of risk for malignancy indicates screening as a prudent measure. To ensure that no worrisome finding was overlooked, a follow-up examination 2 to 3 months after the patient's initial visit is appropriate for physicians who do not have extensive experience in evaluating breast masses. The follow-up visit is also a good time to review the woman's age and cancer risk to determine the appropriate type and frequency of screening tests.

NODULARITY

It can sometimes be difficult to confidently differentiate a nodularity from a dominant mass. In women between 35 and 40 years of age, mammography is usually not helpful in making this determination. Morrow and colleagues reported that in 197 women who were referred for evaluation of a lump but whose physical examinations were considered by the surgeon to be normal or characterized only by glandular nodularity, only three had a mass identified on mammography that resulted in a breast biopsy, and all three were benign.[4] More commonly, the imaging study led to a recommendation for a 6-month follow-up mammogram to monitor abnormalities that were classified as "probably benign" and were independent of the patient's reason for presentation. None of the abnormalities monitored by follow-up mammograms proved to be carcinoma. Similarly, Harris and Jackson reported that no malignant lesions were identified in their study of 625 women younger than 35 years when mammography was used to examine lumpy breasts and suspected fibrocystic disease.[5]

Directed ultrasound of the area in question is the initial study recommended in young women when there is uncertainty regarding the presence of a dominant mass. If no suspicious findings are revealed on ultrasound, a short-interval follow-up examination in 1 to 2 months is appropriate.

In women older than 40 years, a diagnostic imaging workup should be performed when a dominant mass is identified on physical examination. The mammogram should include placement of a skin marker over the area of interest and extra views of the indicated area, if these are determined to be appropriate by the radiologist. If no abnormality is seen on mammography, a directed ultrasound study should be performed to exclude the presence of mammographically occult carcinoma. If these imaging studies reveal no evidence of a breast mass, then a short-interval follow-up examination constitutes appropriate management.

Attempts at needle aspiration to reassure both the patient and her physician that no worrisome abnormalities are present are not usually helpful in the absence of a dominant mass. Normal breast tissue is significantly less cellular than dominant masses; thus, the rate of nondiagnostic aspirates is significantly higher in cases in which no discrete abnormality is detected than in cases in which a dominant mass is present.[4,6] An aspirate with insufficient material for diagnosis is generally considered an indication for surgical biopsy; therefore, the use of fine-needle aspiration (FNA) for vague areas of nodularity may lead to unnecessary biopsies.

In patients with nodularity on physical examination and a negative imaging workup, a follow-up examination should be performed in 1 to 3 months to ensure the stability and benign nature of the nodularity. If the finding persists, another examination after 6 months is appropriate to ensure that a discrete mass is not evolving [*see* Dominant Masses with Clinically Benign Features, *below*].

Imaging Evaluation

In the woman whose physical examination detected a dominant mass with benign clinical features, the initial step in evaluation is to determine whether the mass is cystic or solid. Cysts cannot be reliably diagnosed by physical examination alone, and thus, ultrasound or aspiration of fluid and subsequent resolution of the mass are required for diagnosis [*see Figure 1*].

Ultrasound In women younger than 35 years, ultrasound is often the only diagnostic study needed for the evaluation of a clinically benign breast mass. Mammography is recommended only if the mass is considered suspicious for malignancy, because cancer is rare in this group and mammograms are usually inconclusive because of breast density. Women older than 35 years who present with a breast mass should be evaluated by mammography and ultrasound.

Ultrasound, unlike mammography, has the capacity to differentiate solid masses from cystic masses. Simple cysts are seen on ultrasound as round or oval with sharply defined margins and posterior acoustic enhancement—that is, the tissue deep to the cyst appears brighter than other breast tissue found at the same level—and without any internal echoes. If a cyst is seen to have a solid component on ultrasound, further workup is warranted.

Aspiration Complex cysts are defined as those with septations or internal echoes and are traditionally managed by aspiration. However, in a study of 308 complex cysts, the incidence of malignancy was only 0.3%, suggesting that many complex cysts can be managed adequately by short-term follow-up.[7]

Biopsy Cystic lesions with thick and indistinct walls, thick septations, or any solid component should be biopsied, because some of these masses are malignant. In a study by Berg and colleagues, 18 out of 79 lesions with these characteristics proved to be malignant.[8]

It has been suggested that patients with a palpable breast mass exhibiting normal mammographic and ultrasound findings do not require tissue diagnosis[9]; however, evidence suggests that malignant masses may not be detected by mammography and ultrasonography. Edeiken reported that 22% of 499 women with a palpable breast cancer had a false negative mammogram,[10] and the majority of false negatives were observed in women younger than 50 years (a finding attributed primarily to the prevalence of breast density in young women). Ultrasound is an extremely operator-dependent technique. Although Dennis and colleagues[9] reported that ultrasound has a high sensitivity for the diagnosis of cancer, it is not clear that these results can be generalized to a variety of practice settings.

A patient with a palpable solid mass should be referred to a surgeon for a tissue diagnosis regardless of whether the mass is visualized by imaging studies. Solid masses may be diagnosed with FNA cytology, core-needle biopsy, or excisional biopsy [*see Figure 1*].

Triple Diagnosis Test

The triple diagnosis test uses a combination of physical examination, imaging studies, and FNA cytology as an alternative to surgical excision to establish that a breast mass is benign. The triple test is considered to identify the mass as benign if the physical examination, mammogram, and FNA all indicate a benign process. If the lesion cannot be visualized on mammogram or if the FNA contains insufficient cells for diagnosis, the triple test cannot be confirmatory for a benign lesion. In a study of 191 patients who had confirmatory surgical biopsy, Steinberg and

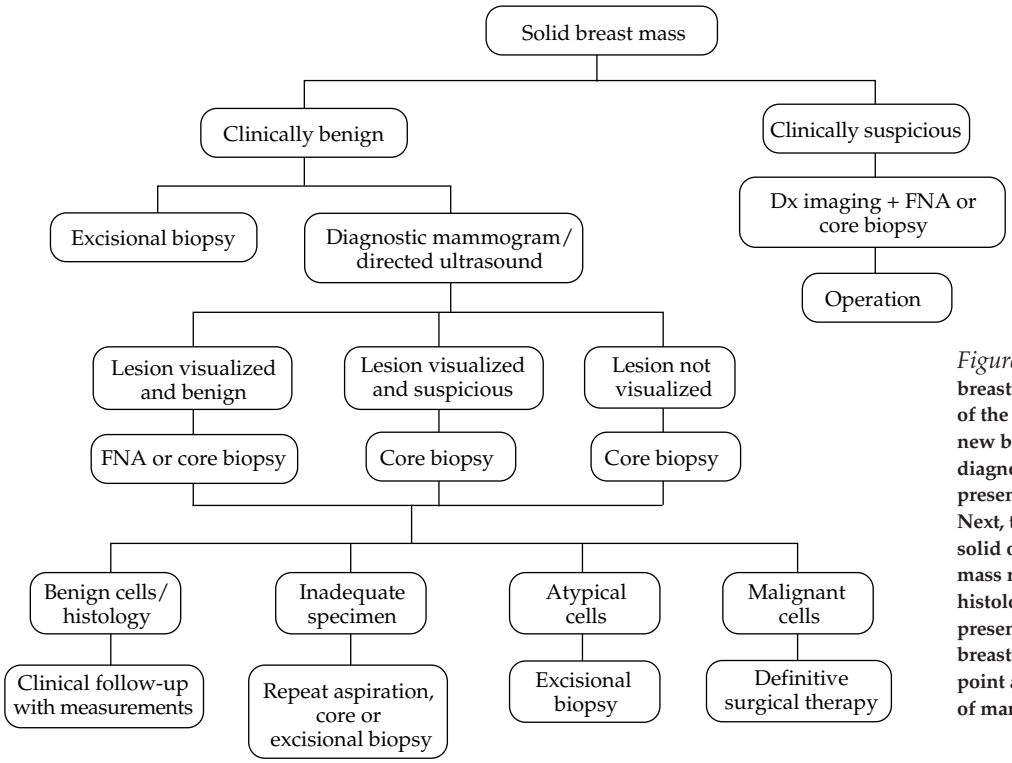


Figure 1 **Management of a solid breast mass. The most important facet of the evaluation and management of a new breast mass is the exclusion of a diagnosis of breast cancer. First, the presence of a mass must be confirmed. Next, the mass is classified as either solid or cystic [*see Figure 2*]. A solid mass requires either cytologic or histologic sampling to exclude the presence of malignancy. Referral to a breast specialist is indicated at any point at which the diagnosis or choice of management is in doubt.**

colleagues reported that the triple test had a sensitivity of 95.5% and a specificity of 100%.[11] Vetto and colleagues reported that the triple diagnosis test is accurate and results in a substantial reduction in the need for surgical excisional biopsies of benign lesions.[12] In their study of 46 breast lesions identified in 43 patients, the triple test produced concordant results in 21 lesions.[12] Twelve triple tests gave concordant benign results, and biopsy was confirmatory in all of these cases (negative predictive value of 100%). In nine cases, there were concordant malignant results, and final pathology on all of these confirmed malignancy (positive predictive value of 100%). There were 25 discordant triple test results (54%): in nine of these cases, the final pathology was benign, and the remaining 16 demonstrated malignancy (positive predictive value of 64%).

No single mode of evaluation used in the triple test is as accurate as the combination of the three.[12] Bicker and colleagues reported that only seven of 2,184 (0.32%) patients assessed as having benign disease by the triple test were subsequently found to have carcinoma, with five of the seven cancers diagnosed within the first year of observation.[13] Overall, FNA has been shown to be quite accurate in the evaluation of benign breast masses [*see Table 3*].[11,12,14-16] FNA cytology alone has a sensitivity ranging from 65% to 98% and a specificity of 34% to 100%. In a review of 29 studies, the likelihood of identifying malignant cytology in patients with breast cancer ranged from 35% to 92%.[17] Lower sensitivity is associated with smaller tumors and younger patient age; sensitivity is quite high when FNA is performed by trained personnel and interpreted by an experienced cytologist.[18]

Following a benign, concordant triple diagnosis test, an identified mass must be monitored for growth by serial examination and imaging studies, which are generally recommended to be performed every 6 months for 2 years, until stability is documented. Growth of the lesion should prompt surgical excision. Patients opting for observation should be counseled about the small possibility of a delay in the diagnosis of cancer. Particular caution should be used when taking a wait-and-see approach in women 50 years of age or older, because in this group benign breast masses are infrequent and carcinoma is more common.

Core-Needle Biopsy

An alternative approach to the diagnosis of a clinically and radiographically benign lesion is core-needle biopsy. This approach has two advantages: (1) it provides a histologic specimen that can be interpreted by a general pathologist, rather than requiring a specialized cytopathologist, and (2) it provides a specific histologic diagnosis rather than simply classifying the mass as benign. In one study, 286 breast lesions (232 of which were palpable) were evaluated both by FNA and core-needle biopsy, and the two tests were reported to have equal sensitivity, positive predictive value for malignancy, and equally low rates of samples inadequate for diagnosis.[15] Core biopsy is associated with slightly greater discomfort and higher costs than FNA. Because core biopsies may sample adjacent tissue, rather than the lesion itself, a follow-up evaluation is necessary, even if a benign result is obtained.

Before undertaking an extensive workup to establish with a high degree of certainty that a mass is benign, it is important to ascertain whether the patient will be comfortable with a palpable abnormality left in place in her breast. For the highly anxious patient, the patient in whom follow-up is difficult, or the unreliable patient, excisional biopsy may be the preferred diagnostic strategy, even for lesions felt to be clinically benign.

SUSPICIOUS DOMINANT BREAST MASSES

Imaging Evaluation

If a breast mass has characteristics suggestive of carcinoma, a diagnostic mammogram is the initial step in evaluation. The purpose of the mammogram is to define the absence or presence and extent of nonpalpable disease associated with the mass and to identify additional abnormalities in the ipsilateral or contralateral breast that may influence the choice of local therapy. The purpose of the mammogram is *not* to diagnose the palpable finding. Between 15% and 30% of palpable cancers are not seen on mammography[19]; therefore, a normal mammogram does not ensure the absence of cancer. However, the likelihood of a false negative result can be minimized through the use of a diagnostic mammogram, which may include multiple views in comparison with the two standard views used in screening mammography. In a diagnostic mammogram, a marker is placed on the palpable finding to ensure that the area of the breast containing the mass is included on the films. Failure to visualize lesions in the periphery of the breast is a well-documented cause of false negative mammograms. Extra views are obtained of the tumor site to define the extent of the primary tumor, and these views are critically important to the surgeon when assessing the local therapy options available to the patient. However, regardless of the findings of mammography, women with clinically suspicious breast masses require a histologic diagnosis to exclude the presence of cancer.

Histologic Diagnosis

Needle biopsy, either core biopsy or FNA, is the preferred technique for diagnosing clinically suspicious breast masses.

Table 3 Accuracy of Fine-Needle Aspiration

Study	*Number of Patients*	*Positive Predictive Value (%)*	*Negative Predictive Value (%)*	*Sensitivity (%)*	*Specificity (%)*
Morris[14]	261	100	95.5	96	100
Westenend[15]	286	100	—	92	82
Steinberg[11]	191	98	—	49	99.5
Vetto[12]	46	100	95.5	96	100
Patel[16]	731	99.4	85	91	56*

*Included specimens insufficient for diagnosis.

Needle biopsy is more cost-effective than primary surgical excision and does not involve incisions that may interfere with mastectomy incision placement or affect breast skin needed for breast reconstruction, should the patient choose such a treatment option.

False negative rates for core biopsy and FNA, as discussed previously, are similar. A small incidence of false positive findings with FNA has been reported. The likelihood of having a benign final diagnosis after a frankly malignant FNA result is between 0% and 0.2%. However, the likelihood of having a benign final diagnosis after an FNA result reported as suspicious for malignancy is as high as 6.2%; thus, a finding that is suspicious for malignancy on FNA should prompt a surgical excision for definitive diagnosis before proceeding to definitive cancer treatment.[17]

Core biopsy is preferred by many physicians because it provides information about tumor histology and reliably distinguishes invasive from intraductal carcinoma. If needle-biopsy techniques do not provide a definitive diagnosis of malignancy and the mass has suspicious characteristics on physical examination or mammogram, a surgical biopsy should be performed. When surgical excision is undertaken, the procedure should be performed as a lumpectomy that includes excision of some adjacent normal breast tissue and placement of orienting sutures for the pathologist. If the margins are free of cancer, the diagnostic procedure serves as the definitive breast procedure.

Differential Diagnosis and Management of Common Benign Breast Masses

CYSTS

Cysts are most frequent in the perimenopausal years, with 63% of cysts occurring in women between the ages of 40 and 49.[20] Cysts are less common in younger women, accounting for only 10% of breast masses in women younger than 40 years.[4] Cysts are also relatively uncommon in postmenopausal women, except in those receiving hormone replacement therapy. Multiple cysts in the same patient, whether synchronous or metachronous, are not uncommon.[21]

Cysts are usually round or oval and have smooth, well-demarcated borders. They may fluctuate with the menstrual cycle, and they may be tender, especially if they have filled rapidly. Cysts are indistinguishable from solid masses on physical examination and mammography, but they may be suspected because of a history of menstrual-cycle variation and the clinical finding of a well-circumscribed, firm mass. There are two methods to definitively characterize a mass as a cyst: aspiration and ultrasound [*see Figure 2*]. Aspiration is performed with a 20- to 22-gauge needle and is successful if fluid is obtained and the mass completely resolves. The fluid may be yellow, clear, or a murky greenish-brown. Bloody cyst fluid is an indication for excision, and if blood is identified in the aspirate, the procedure should be promptly halted in an attempt to preserve a part of the mass. Cyst aspiration fluid should not be sent for cytology, because dead epithelial cells are shed into a cyst and atypical findings are likely, despite the low likelihood of malignancy.[20] Additionally, once the cyst has been aspirated, there is no palpable target for the excision mandated by the finding of atypia. In a study of 6,747 cysts with nonbloody aspirates, no cancers were found.[20] If a mass is still present after aspiration, the mass should be evaluated as a dominant solid mass, and a histologic diagnosis should be obtained. If a cyst recurs rapidly (within 1 to 3 weeks) after aspiration, it should be reevaluated. Fewer than 20% of cysts will recur rapidly after aspiration,[22] and rapid refilling of the cyst raises the possibility of a mass resulting in ductal obstruction or a growth within the cyst wall. A second aspiration may be undertaken, but with multiple recurrences, surgical biopsy should be performed to ensure that a cancer with a cystic component is not overlooked.

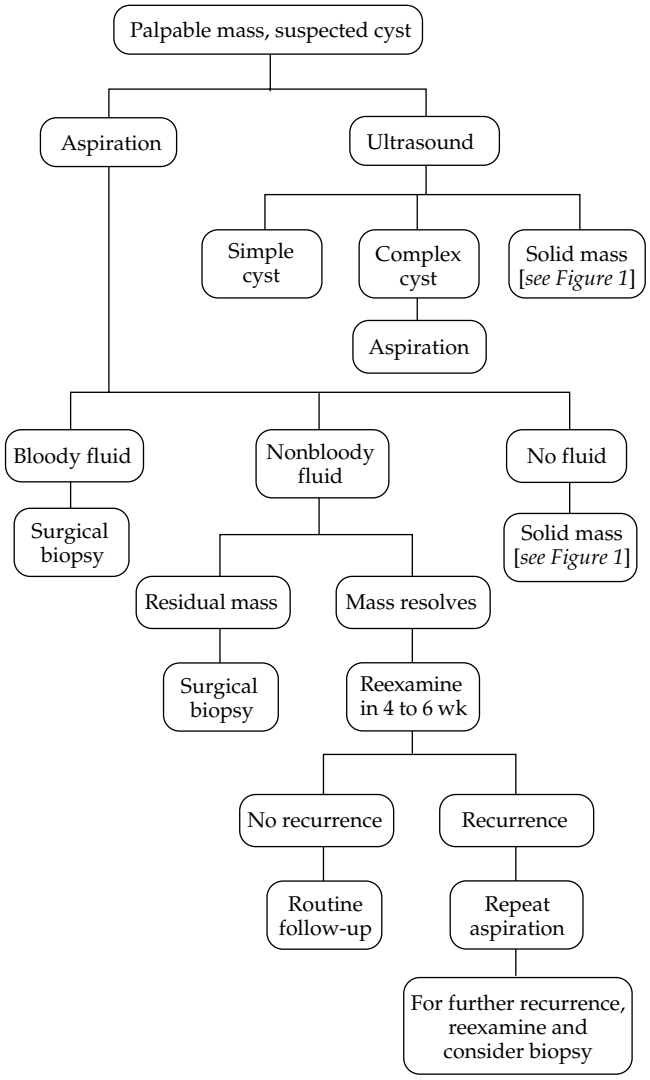


Figure 2 **Management of a breast cyst. Cysts can be identified either by aspiration of fluid or by ultrasound. Prompt diagnosis and management reduce the anxiety level of the patient.**

Patients who do not wish to undergo aspiration of the cyst may opt for targeted ultrasound of the mass. Ultrasound is 98% to 100% accurate in diagnosing simple cysts when careful criteria are used.[23] As previously stated, the criteria for identifying a simple cyst on ultrasound are its round or oval shape, sharply defined margins, posterior acoustic enhancement, and absence of internal echoes. Simple cysts positively identified on ultrasound do not require any treatment, although painful cysts may be aspirated for symptom relief.

Fibroadenomas

Fibroadenomas are a frequent cause of breast mass and occur most commonly in younger women. Fibroadenomas tend to occur at an earlier age in blacks than in whites.[24] A biphasic incidence is reported in white women, with peaks at 25 and 48 years of age[25]; in black women, the peak incidence is between 16 and 25 years of age.[25] Fibroadenoma cannot be definitively diagnosed by clinical examination alone.

Fibroadenomas in young women are usually firm; rubbery; well circumscribed; nontender; and very mobile, because they grow by displacing the surrounding breast tissue. Fibroadenomas can be diagnosed by core biopsy or excisional biopsy, or they may be observed after a benign and concordant triple diagnosis test, whichever is preferable to the patient. Excision is indicated if any aspect of the triple test is discordant or if the lesion enlarges after diagnosis.

Ultrasound is useful to document the stability of a lesion identified as a fibroadenoma that is being managed conservatively. Benign fibroadenomas can change in size. In a study of 1,070 cases diagnosed by FNA and followed for 3 years, a mean change in each of the three dimensions of 20% or less over a 6-month interval was not associated with malignancy.[26] A more significant change in size may be an indication of malignancy or phyllodes tumor.[27,28] Fibroadenomas may increase in size significantly during pregnancy, and younger women considering expectant management of a fibroadenoma should be educated regarding this possibility. Fibroadenomas can also regress entirely or become smaller over time. In a study of 92 fibroadenomas confirmed by FNA and followed for a mean of 47 weeks, 15 (16%) resolved and 30 (33%) enlarged throughout the study.[29] The remaining lesions either remained static or enlarged slightly and then remained unchanged. Tumor resolution is more likely to occur in fibroadenomas in very young women, in new-onset fibroadenomas, in small fibroadenomas, and in fibroadenomas occurring during pregnancy.

Although fibroadenomas themselves are benign, having a fibroadenoma may slightly increase the long-term risk for breast cancer development. For example, a retrospective cohort study showed a 2.17 relative risk for breast cancer in women with fibroadenomas, compared with the study's control subjects.[30] However, in the subset of patients without a family history of breast cancer and with a noncomplex fibroadenoma, no increase in risk was seen, whereas those with complex fibroadenomas had a relative risk between 2 and 4.

Complex fibroadenomas are characterized by epithelial calcification, apocrine metaplasia, and sclerosing adenosis and are larger than 3 mm.[30] Fibroadenomas rarely contain epithelial hyperplasia with atypia, with a frequency of 0.3%.[31] Furthermore, in one study, the presence of such atypia did not correlate with an additional increased risk of future breast carcinoma.[32] Fibroadenomas are regarded by the Cancer Committee of the College of American Pathologists as lesions that do not confer an increased risk for the development of breast cancer.[33]

Phyllodes Tumors

Phyllodes tumors are much less common than fibroadenomas. In one study of 515 benign masses, the ratio of fibroadenomas to phyllodes tumors was 29:1.[34] The majority of phyllodes tumors are benign, but they may grow to a very large size. A rapid change in the size of a mass thought to be a fibroadenoma is suggestive of a phyllodes tumor and is an indication for excision. Benign phyllodes tumors have a propensity to recur locally in about 20% of cases unless excised with a 1 to 2 cm margin of normal breast tissue.[35] For very large tumors, adequate resection may necessitate mastectomy.

Table 4 Benign Breast Lesions by Category

Nonproliferative lesions
Simple cysts
Mild hyperplasia of the usual type
Papillary apocrine change
Proliferative lesions without atypia
Fibroadenomas
Intraductal papillomas
Sclerosing adenosis
Moderate to florid hyperplasia
Proliferative lesions with atypia
Atypical ductal hyperplasia
Atypical lobular hyperplasia

Other Benign Masses

Hamartomas Hamartomas are uncommon benign breast masses that are usually well circumscribed and contain elements of fat, glandular tissue, and fibrous tissue.[36] Because they contain the same tissues as the normal breast, hamartomas are often difficult to feel and are more typically diagnosed by mammography. Hamartomas do not have specific diagnostic histologic features; thus, tissue diagnosis must be correlated to physical findings and mammographic imaging.[36] The cytologic findings overlap with findings in other benign masses, and they are unlikely to be confused with malignancy.[37] There are extremely rare reports of in situ or invasive carcinoma identified in a breast hamartoma, but for the most part, hamartomas are benign masses.

Fat necrosis Fat necrosis results from trauma to the breast, which can be secondary to operation, accident (e.g., seat-belt injury and falls), or radiation therapy. In up to 50% of cases, the patient cannot recall any antecedent trauma.[38]

Clinically, fat necrosis results in a firm irregular breast mass. Because fat necrosis often occurs superficially in the breast, it may result in skin retraction that mimicks carcinoma. There is a spectrum of imaging findings associated with fat necrosis, including radiolucent oil cysts (which are diagnostic of the condition), dystrophic calcifications, and spiculated masses mimicking carcinoma.[38] It can be difficult to differentiate fat necrosis from malignancy using clinical and radiographic features alone, and biopsy is usually required.

Benign Breast Lesions and Breast Cancer Risk

Much confusion has surrounded the relationship between benign breast disease and breast cancer risk. The seminal work in this area was by Dupont and Page, who reviewed over 10,500 women with benign breast biopsies and reported results based on 3,303 of these women, who were followed for a median of 17 years.[39] Benign disease was categorized into three groups: nonproliferative disease, proliferative disease without atypia, and proliferative disease with atypia [*see Table 4*].

Women with nonproliferative lesions were not found to be at increased risk for subsequent breast cancer, whereas women with proliferative lesions and no atypia had a relative risk for developing cancer of 1.9. Women with atypical ductal or lobular hyperplasia had a cancer risk 5.3 times that of women with nonproliferative lesions, and women with atypical hyperplasia and a family history of breast cancer had a relative risk for developing cancer of 11. These findings have been validated in a subsequent prospective, case-control study by London and colleagues, who evaluated 121 women with breast cancer who had a prior biopsy for benign breast disease and 448 control subjects (women without cancer matched for year of birth and year of benign biopsy).[40] This study reported a relative risk of 1.6 for patients with proliferative disease and no atypical hyperplasia and a relative risk of 3.7 for patients with proliferative disease and atypical hyperplasia. It is important to recognize that atypical hyperplasia is uncommon in clinically detected lesions, accounting for only 3.6% of patients in the study by Dupont and Page.[39] Although proliferative lesions increase the relative risk of breast cancer development by a factor of 1.5 to 2.0, the absolute risk of breast cancer remains low in the absence of other risk factors. Atypical hyperplasia is a significant risk factor, but the absolute risk of cancer will vary with the individual's age and other risk factors. The modified Gail model is a useful tool for evaluating a woman's overall level of risk after a diagnosis of atypical hyperplasia.[41]

The authors have no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Hindle WH: Breast mass evaluation. Clin Obstet Gynecol 45:750, 2002

2. Love SM, Gelman RS, Silen W: Sounding board. Fibrocystic "disease" of the breast—a nondisease? N Engl J Med 307:1010, 1982

3. Madigan MP, Ziegler RG, Benichou J, et al: Proportion of breast cancer cases in the United States explained by well-established risk factors. J Natl Cancer Inst 87:1681, 1995

4. Morrow M, Wong S, Venta L: The evaluation of breast masses in women younger than forty years of age. Surgery 124:634, 1998

5. Harris VJ, Jackson VP: Indications for breast imaging in women under age 35 years. Radiology 172:445, 1989

6. Rimsten A, Stenkvist B, Johanson H, et al: The diagnostic accuracy of palpation and fine-needle biopsy and an evaluation of their combined use in the diagnosis of breast lesions: report on a prospective study in 1244 women with symptoms. Ann Surg 182:1, 1975

7. Venta LA, Kim JP, Pelloski CE, et al: Management of complex breast cysts. AJR Am J Roentgenol 173:1331, 1999

8. Berg WA, Campassi CI, Ioffe OB: Cystic lesions of the breast: sonographic-pathologic correlation. Radiology 227:183, 2003

9. Dennis MA, Parker SH, Klaus AJ, et al: Breast biopsy avoidance: the value of normal mammograms and normal sonograms in the setting of a palpable lump. Radiology 219:186, 2001

10. Edeiken S: Mammography and palpable cancer of the breast. Cancer 61:263, 1988

11. Steinberg JL, Trudeau ME, Ryder DE, et al: Combined fine-needle aspiration, physical examination and mammography in the diagnosis of palpable breast masses: their relation to outcome for women with primary breast cancer. Can J Surg 39:302, 1996

12. Vetto J, Pommier R, Schmidt W, et al: Use of the "triple test" for palpable breast lesions yields high diagnostic accuracy and cost savings. Am J Surg 169:519, 1995

13. Bicker T, Schondorf H, Naujoks H: Long-term follow-up in patients with mammary gland changes found unsuspicious by aspiration cytology. Cancer Detect Prev 11:319, 1988

14. Morris A, Pommier RF, Schmidt WA, et al: Accurate evaluation of palpable breast masses by the triple test score. Arch Surg 133:930, 1998

15. Westenend PJ, Sever AR, Beekman-De Volder HJ, et al: A comparison of aspiration cytology and core needle biopsy in the evaluation of breast lesions. Cancer 93:146, 2001

16. Patel JJ, Gartell PC, Smallwood JA, et al: Fine needle aspiration cytology of breast masses: an evaluation of its accuracy and reasons for diagnostic failure. Ann R Coll Surg Engl 69:156, 1987

17. Giard RW, Hermans J: The value of aspiration cytologic examination of the breast: a statistical review of the medical literature. Cancer 69:2104, 1992

18. Cariaggi MP, Bulgaresi P, Confortini M, et al: Analysis of the causes of false negative cytology reports on breast cancer fine needle aspirates. Cytopathology 6:156, 1995

19. Barton MB, Harris R, Fletcher SW: The rational clinical examination. Does this patient have breast cancer? The screening clinical breast examination: should it be done? How? JAMA 282:1270, 1999

20. Ciatto S, Cariaggi P, Bulgaresi P: The value of routine cytologic examination of breast cyst fluids. Acta Cytol 31:301, 1987

21. Hughes LE, Bundred NJ: Breast macrocysts. World J Surg 13:711, 1989

22. Leis HP Jr: Gross breast cysts: significance and management. Contemp Surg 39:13, 1991

23. Sickles EA, Filly RA, Callen PW: Benign breast lesions: ultrasound detection and diagnosis. Radiology 151:467, 1984

24. Oluwole SF, Freeman HP: Analysis of benign breast lesions in blacks. Am J Surg 137:786, 1979

25. Nigro DM, Organ CH Jr: Fibroadenoma of the female breast: some epidemiologic surprises. Postgrad Med 59:113, 1976

26. Gordon PB, Gagnon FA, Lanzkowsky L: Solid breast masses diagnosed as fibroadenoma at fine-needle aspiration biopsy: acceptable rates of growth at long-term follow-up. Radiology 229:233, 2003

27. Hermann G, Keller RJ, Tartter P, et al: Interval changes in nonpalpable breast lesions as an indication of malignancy. Can Assoc Radiol J 46:105, 1995

28. Sickles EA: Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. Radiology 179:463, 1991

29. Wilkinson S, Anderson TJ, Rifkind E, et al: Fibroadenoma of the breast: a follow-up of conservative management. Br J Surg 76:390, 1989

30. Dupont WD, Page DL, Parl FF, et al: Long-term risk of breast cancer in women with fibroadenoma. N Engl J Med 331:10, 1994

31. Kuijper A, Mommers EC, van der Wall E, et al: Histopathology of fibroadenoma of the breast. Am J Clin Pathol 115:736, 2001

32. Carter BA, Page DL, Schuyler P, et al: No elevation in long-term breast carcinoma risk for women with fibroadenomas that contain atypical hyperplasia. Cancer 92:30, 2001

33. Fitzgibbons PL, Henson DE, Hutter RV: Benign breast changes and the risk for subsequent breast cancer: an update of the 1985 consensus statement. Cancer Committee of the College of American Pathologists. Arch Pathol Lab Med 122:1053, 1998

34. Hindle WH, Alonzo LJ: Conservative management of breast fibroadenomas. Am J Obstet Gynecol 164:1647, 1991

35. Mangi AA, Smith BL, Gadd MA, et al: Surgical management of phyllodes tumors. Arch Surg 134:487, 1999

36. Tse GM, Law BK, Ma TK, et al: Hamartoma of the breast: a clinicopathological review. J Clin Pathol 55:951, 2002

37. Herbert M, Schvimer M, Zehavi S, et al: Breast hamartoma: fine-needle aspiration cytologic finding. Cancer 99:255, 2003

38. Bilgen IG, Ustun EE, Memis A: Fat necrosis of the breast: clinical, mammographic and sonographic features. Eur J Radiol 39:92, 2001

39. Dupont WD, Page DL: Risk factors for breast cancer in women with proliferative breast disease. N Engl J Med 312:146, 1985

40. London SJ, Connolly JL, Schnitt SJ, et al: A prospective study of benign breast disease and the risk of breast cancer. JAMA 267:941, 1992

41. Spiegelman D, Colditz GA, Hunter D, et al: Validation of the Gail et al. model for predicting individual breast cancer risk. J Natl Cancer Inst 86:600, 1994

XV APPROACH TO THE PATIENT WITH A PELVIC MASS

Joseph T. Chambers, m.d., ph.d.
Carolyn D. Runowicz, m.d.

The finding of a pelvic mass may occur in a female patient of any age; fortunately, most masses are associated with a benign neoplasm or process. Nevertheless, each year, between 5% and 10% of women in the United States undergo surgery for a suspected ovarian neoplasm; only 13% to 21% of these women prove to have a malignant ovarian neoplasm.[1]

A pelvic mass may be found on an abdominal or pelvic examination in an asymptomatic patient at a scheduled health maintenance assessment or on examination of a symptomatic patient who presents with a complaint. A mass may also be noted as an incidental finding on imaging studies, usually computed tomographic scans of the abdomen or pelvis or pelvic ultrasonography, obtained either for medical indications unrelated to a suspected pelvic disease or as a screening study for gynecologic malignancies. Increasingly, a pelvic mass is found on imaging studies obtained in the emergency department for the evaluation of vague abdominal pain or trauma.[2]

The differential diagnosis is extensive [*see Table 1*], because a pelvic mass may be of gynecologic or nongynecologic origin and may be associated with congenital, functional, neoplastic (either benign or malignant), obstructive, or inflammatory processes; a mass may also be associated with pregnancy. When a mass is of gynecologic origin, it may arise from the ovary, fallopian tube, broad ligament, round ligament, uterus, cervix, or uterosacral ligament. Adnexal masses arise from an area comprising the ovary, the fallopian tube, and the ligaments of the uterus. When the mass is of nongynecologic origin, it may arise from the urinary tract system, the gastrointestinal system, or the pelvic vessels or nerves. Rarely, a mass may arise directly from the peritoneum, the retroperitoneum, or the omentum.

The possible etiologies differ markedly, depending on the patient's age and symptoms. Consequently, the complete documentation of symptoms; the past medical, family, and surgical histories; the physical examination; and the initial imaging studies and laboratory tests will narrow the differential diagnosis and direct appropriate evaluation and referral. The overview of the diagnostic evaluation is presented [*see* Evaluation of the Patient with a Pelvic Mass, *below*], followed by a review of the distinctions in diagnostic workup required for patients of each age group [*see* Age-Specific Considerations in Patient Evaluation, *below*].

Evaluation of the Patient with a Pelvic Mass

MEDICAL HISTORY

The patient's medical history is essential in the evaluation of a pelvic mass; it includes a complete menstrual history, which establishes menarche and cyclicity of menses. Because ovulation generally occurs midcycle, a new pelvic mass found midcycle could be consistent with a functional cyst, whereas a mass identified later in the cycle would be more consistent with a corpus luteum cyst. A patient with a long history of irregular menses or even secondary amenorrhea with nontender bilateral adnexal masses may have polycystic ovaries. Pelvic inflammatory disease usually presents about a week after menstruation, and missed menses suggest a pregnancy. Therefore, an accurate menstrual history is useful in directing the patient evaluation.

Menstrual Bleeding

Increased menstrual bleeding often occurs with submucous leiomyomas. Postmenopausal bleeding may be caused by endometrial cancer or, rarely, by hormonally functioning ovarian or fallopian tube neoplasms. Irregular bleeding, pain, and missed menses suggest an ectopic pregnancy.

Although rare, estrogen-producing neoplasms may cause abnormal uterine bleeding, breast tenderness, or hirsutism. In children, precocious puberty may develop.

In women of reproductive age, the medical history must include questions about contraceptive practices, because a patient on oral contraceptive pills is less likely to develop physiologic and functional cysts (e.g., follicular and corpus luteum cysts). In addition, a patient with multiple sexual partners who does not use barrier contraception may develop pelvic inflammatory disease and a chronic tubo-ovarian abscess.

Abdominal Pain

Pain, if present, must be characterized. The onset, pattern of occurrence, and extent or severity of pain may give important indications of the etiology of the suspected pelvic mass. For example, new onset of midcycle pain in premenopausal women suggests the presence of a physiologic cyst. Pain following intercourse may be related to a ruptured cyst, whereas chronic pain during intercourse is suggestive of endometriosis.

Acute pain Acute severe abdominal or pelvic pain accompanies torsion of an adnexal structure or pedunculated leiomyoma (in which case the pain is usually intermittent); hemorrhage or rupture of an ovarian cyst; dilation of the fallopian tube in association with an ectopic pregnancy; or rupture of a pelvic abscess. In addition, a degenerating leiomyoma may also cause acute pain. The rapid stretching of the ovarian capsule by an expanding functional cyst or germ cell neoplasm may produce acute pain. Acute pain associated with nausea, vomiting, and fever suggests an inflammatory process, such as pelvic inflammatory disease, appendicitis, peritonitis from a perforation or infarction of bowel, or diverticulitis.

Chronic, cyclic pain More chronic, cyclic pain, particularly pain associated with dysmenorrhea, dyspareunia, and menorrhagia, suggests endometriosis. In addition, chronic pelvic or abdominal discomfort is often reported with long-standing tubo-ovarian abscesses or chronic hydrosalpinx. Chronic abdominal pain or vague abdominal discomfort accompanied by bloating suggests an ovarian neoplasm.

Changes in Bowel Habits

Changes in bowel habits and constitution, as well as changes in appetite, weight, and energy, may be more consistent with a benign colonic disease or with a cancer arising in any pelvic or-

gan. The medical history of gynecologic cancers or precancerous disease, breast cancer, melanoma, or any cancer may direct the workup. The family history is important, because hereditary cancers, such as those associated with a *BRCA1* or *BRCA2* mutation (e.g., breast, ovary, endometrial cancer), or nonpolyposis colon cancer may present as a pelvic mass.

PHYSICAL EXAMINATION

Before the pelvic examination, the patient should empty her bladder. There must be adequate light for a visual inspection of the external genitalia for signs of androgen excess. The vagina and cervix must be inspected for signs of infection and hormones. A Papanicolaou smear should be obtained if the patient's age and findings on physical examination indicate the need for it. The abdominovaginal examination gives information on the size, shape, consistency, location, mobility, laterality, and tenderness of the mass. This information must be confirmed by the rectovaginal examination, which also evaluates the uterosacral ligaments, the cul-de-sac, and the anorectal area. Physical examination may suggest whether the mass is benign or malignant [*see Table 2*]. When a pelvic mass is identified, a complete physical examination must be performed. Particular attention should be paid to the examination of the breasts, the respiratory system, the nodal areas, and the abdomen.

LABORATORY TESTS

In the evaluation of a pelvic mass, especially in women of reproductive age, a urinary or serum β-human chorionic gonadotropin (β-hCG) test is required. To determine pregnancy, the urinary test is adequate at the time of missed menses. This test, however, is qualitative in nature, and for the management of an ectopic pregnancy or molar pregnancy, a quantitative determination of the β-hCG serum concentration is necessary. In addition to being an indication for pregnancy, elevated levels of β-hCG may be associated with theca-lutein cysts, especially in women with choriocarcinoma, diabetes mellitus, and Rh sensitization; elevated levels of β-hCG are also associated with clomiphene use, human menopausal gonadotropin or human chorionic gonadotropin ovulation induction, and use of gonadotropin-releasing hormone analogues.[3]

A complete blood count will help in assessing anemia; an elevated white cell count suggests an infectious etiology. The erythrocyte sedimentation rate is nonspecific and does not narrow the differential diagnosis.

IMAGING STUDIES

The most important initial diagnostic tool in evaluating a pelvic mass is the pelvic ultrasound, which indicates whether the mass is more likely to be uterine, adnexal, or gastrointestinal in origin. A pelvic ultrasound scan also provides information on the size and consistency of the mass—characteristics that suggest whether the mass is malignant or benign.

Unilocular ovarian cysts are overwhelmingly benign and resolve spontaneously in 3 to 6 months; therefore, observation is the recommended approach to management. If malignancy is suspected on pelvic ultrasound, additional imaging studies such as abdominal pelvic computed tomography or magnetic resonance imaging may assist in confirming the diagnosis. The CT scan may further characterize the malignant potential of the mass and give information regarding evidence of metastatic spread to lymph nodes, the retroperitoneal space, or adjacent structures. If the origin of the pelvic mass is uncertain, an MRI may help clarify whether the mass arises from the uterus, adnexa, or another structure (e.g., muscles or nerves).

Table 1 Differential Diagnosis of Pelvic Mass

- Ovary
  - Functional cyst
  - Endometrioma
  - Benign neoplasm
  - Malignant neoplasm
- Fallopian tube
  - Tubo-ovarian abscess
  - Hydrosalpinx
  - Paratubal cyst
  - Ectopic pregnancy
  - Benign neoplasm
  - Malignant neoplasm
- Uterus
  - Fibroid (pedunculated or interligamentous)
  - Intrauterine pregnancy
  - Sarcoma
- Gastrointestinal tract
  - Bowel loops with feces
  - Diverticular disease
  - Inflammatory bowel disease
  - Appendicitis
  - Benign small bowel neoplasm
  - Colon cancer
- Urinary tract
  - Distended bladder
  - Pelvic kidney
  - Urachal cyst
- Retroperitoneum
  - Abdominal wall hematoma or abscess
  - Sarcoma, lymphoma, or teratoma
  - Benign neoplasm

Ultrasound

Transabdominal and transvaginal ultrasound Two methods of ultrasound provide extensive information about the characteristics of a pelvic mass. Transabdominal ultrasound is better tolerated by patients than is transvaginal ultrasound, and it gives more information about abdominal processes. Transvaginal ultrasound, however, provides better resolution and more precise information of the mass within the pelvic organ. The unique imaging patterns offered by each of these ultrasound modalities frequently help narrow the differential diagnosis.

Physiologic ovarian cysts are usually oval and filled with clear fluid. They may or may not have septations, and they have no echogenic or solid components, with the exception of the corpus luteum. Cystic teratomas are partially cystic and solid with echogenic foci of abnormal tissue from foreign tissues and hemorrhage. Endometriomas are simple cysts with echogenic elements secondary to old and new hemorrhage. Leiomyomas are solid or semisolid, well-circumscribed masses; they are similar in appearance to the myometrium but have no vascular vessels within the mass.

Several morphologic scoring systems have been introduced to predict whether a pelvic mass evaluated by ultrasound is benign or malignant. Most systems agree that the following characteristics are suggestive of malignancy: irregularity in the wall of the mass, the presence of thick septations within the mass, any papillary projection within or emerging from the mass, and a mass containing solid components. Size itself is an important characteristic; a mass larger than 8 cm in diameter raises concern. In general, the more of these characteristics that are present, the greater the chance that the mass is malignant.

Color Doppler ultrasound The use of color flow Doppler imaging may help in determining whether the adnexal mass is malignant or benign, because malignancies have an increased neovascularity and, thus, lower resistance and pulsatile indices. Currently, there is no agreement concerning pulsatile indices that indicate malignancy; nevertheless, specific indices may be

less important in suggesting malignancy than overall morphologic pattern, blood flow, and tumor location.

Taking into consideration the findings of regular and color flow ultrasound, the overall sensitivity, using morphologic criteria for malignant disease, ranges from 82% to 100%; specificity ranges from 60% to 95%. It is unclear, however, whether the addition of the color flow Doppler significantly improves these percentages. Two large studies reported that the combined approach (using regular and color flow ultrasound) has a sensitivity of 88% to 97%, a specificity of 97% to 100%, and an accuracy of 83% to 99%.[4]

Computed tomography CT is less frequently used in the initial evaluation of a pelvic mass than other imaging studies. However, the sensitivity, specificity, and accuracy of CT for determining whether the mass is benign or malignant is reported to be comparable to other modalities (i.e., sensitivity, 89%; specificity, 96% to 99%; accuracy, 92% to 94%).[4] CT is better than ultrasound in assessing the retroperitoneal spaces (i.e., nodal systems, pancreas, and spleen) and the omentum. CT is useful in establishing the extent of intra-abdominal and retroperitoneal disease in patients in whom an ovarian malignancy is highly suspected.

Magnetic Resonance Imaging

MRI may be used in the evaluation of a pelvic mass when findings from ultrasound studies are unclear or indeterminate. MRI is particularly useful in clarifying the origin of the mass as either uterine or ovarian. In addition, its accuracy in assessing fatty and hemorrhagic components of a mass can help in the diagnosis of dermoid cysts, hemorrhagic corpus luteum cysts, and endometriosis. This modality may be especially useful in the pregnant patient, if more information is needed than is provided by the pelvic ultrasound.[5]

Positron Emission Tomography

Positron emission tomography (PET) has limited application in the initial evaluation of pelvic masses. It is used primarily to detect recurrent pelvic malignancies.

SERUM TUMOR MARKERS

Serum tumor markers are not used for screening; they are used in the evaluation of patients with suspected malignant pelvic neoplasms. In deciding which serum markers to use, consideration must be given to the patient's age and medical history. For instance, in prepubertal girls and young women, germ cell tumors are the most frequent ovarian malignancy; appropriate serum markers for suspected germ cell tumors include α-fetoprotein (AFP), lactate dehydrogenase (LDH), and β-hCG. In older women with suspected malignancy, it is important to measure epithelial tumor markers, such as CA125. In patients with confirmed malignancy, a significant elevation in the level of a tumor marker may indicate malignant recurrence, whereas decreasing levels reflect response to treatment. Functional genomics and proteomics—the study of human gene sequences and protein sequences, respectively—show promise in identifying novel cellular targets that may be exploited in the future as screening tests for cancer.

Table 2 Physical Findings Associated with Pelvic Mass

Characteristic of Mass	*Suggestive of Benign Process*	*Suggestive of Malignant Process*
Unilateral	Yes	Occasionally
Bilateral	Occasionally	Yes
Cystic	Yes	No
Solid	No	Yes
Mobile	Yes	Occasionally
Fixed	Occasionally	Yes
Irregular contour	Occasionally	Yes
Smooth contour	Yes	No
Presence of ascites	No	Yes
Cul-de-sac nodules	Usually no	Yes
Rapid growth rate	No	Yes
Pain	Yes	Usually no
Size	< 5 cm	≥ 10 cm

Measurement of AFP, LDH, and β-hCG

Dysgerminomas are the most common germ cell tumors and can occur at any age; these malignancies are associated with elevations in AFP, LDH, and β-hCG. Placental alkaline phosphatase and LDH levels are sometimes elevated in patients with dysgerminomas. Endodermal sinus tumors can be monitored by measuring AFP levels, and choriocarcinomas can be monitored by measuring β-hCG levels. Embryonal carcinoma may secrete both AFP and β-hCG. Immature teratomas usually do not secrete any markers. Mixed germ cell tumors may secrete combinations of the above markers. Granulosa cell tumors are associated with elevated levels of inhibin and estradiol in prepubertal and postmenopausal women.

Measurement of CA125

CA125 is a glycoprotein expressed by fetal amniotic and coelomic epithelium and müllerian epithelium. Elevated levels of serum CA125 may suggest an increased risk of malignancy in certain patients; however, elevations of CA125 are also associated with normal and benign conditions. Therefore, CA125 measurement is not useful as a screening test for ovarian cancer. A CA125 serum concentration greater than 35 U/ml is found in 83% of patients with epithelial ovarian cancer, but it is reported to be elevated in only 50% of patients with stage I disease that is limited to the ovary.[6] In addition, CA125 is increased in patients with other malignancies; in those with benign gynecologic conditions; and in patients with diverticulitis and cirrhosis. It is also increased in normal conditions of pregnancy and menstruation [*see Table 3*].[7] Approximately 1% of healthy women have elevated CA125 serum levels.[8] In women with benign gynecologic conditions, the levels are usually less than 200 U/ml.

In postmenopausal women, CA125 measurement may have some application as a screening test for malignancy but is not a routine screening test. In a prospective Swedish study of 4,290 volunteer women who were at least 50 years of age, the specificity of CA125 measurement for ovarian cancer was reported to be 97% and the positive predictive value was 4.6%, using a CA125 level greater than 30 U/ml.[9] However, an elevated CA125 level in postmenopausal women with a pelvic mass suggests a malignancy. The positive predictive value for elevations of CA125 in this age group has been reported to be 97%.[10] In women of reproductive age, elevations of serum CA125 may raise the suspicion of malignancy when imaging findings (i.e., tumor location, morphologic pattern, and vascularity) are also consistent with malignancy.

Table 3 Conditions Associated with an Elevated Serum CA125 Level

Gynecologic malignancies	Epithelial ovarian cancers Germ cell cancers Sex chord stromal tumors Fallopian tube cancers Endometrial cancers Adenocarcinoma of the cervix
Benign gynecologic conditions	Adenomyosis Benign ovarian neoplasms Endometriosis Functional ovarian cysts Leiomyomas Meigs syndrome Menstruation Pregnancy Ovarian hyperstimulation Pelvic inflammation
Nongynecologic conditions	Liver disease and cirrhosis Colitis Congestive heart failure Diabetes Diverticulitis Lupus Mesothelioma Pericarditis Polyarteritis nodosa Surgery Previous irradiation Renal disease Sarcoidosis Tuberculosis
Nongynecologic cancers	Breast Colon Lung Pancreas Lymphoma

Measurement of Other Tumor Markers

CA19-9 and carcinoembryonic antigen (CEA) are commonly used to follow mucinous tumors. However, the sensitivity and specificity of these markers are lower than those found with CA125 measurement.

FINE-NEEDLE ASPIRATION

Fine-needle aspiration (FNA) is not routinely used in the evaluation of pelvic masses. Although it can easily be performed with ultrasound or CT guidance, FNA has limited diagnostic accuracy, especially in the evaluation of cystic structures. In a study of the use of FNA in 235 patients with cystic ovarian masses, 56% of the aspirates were devoid of diagnostic cells. The sensitivity for specific lesions ranged from 35% to 83%, and the specificity approached 100%.[11] Thus, FNA is not accurate, and should rupture of the cyst contents occur, dissemination of malignant cells may result.

BIOPSY

Rarely, a directed biopsy may be indicated in an individual whose imaging studies are consistent with advanced ovarian intra-abdominal disease and in whom surgery is contraindicated because of significant medical problems. A directed biopsy of a peritoneal tumor implant may be used to identify the histology of the tumor and thus assist in selecting appropriate nonsurgical management (e.g., initial or neoadjuvant chemotherapy) for ovarian cancer. Immunohistochemical staining profiles may provide additional information consistent with ovarian or peritoneal cancer.

Age-Specific Considerations in Patient Evaluation

PELVIC MASS IN PREPUBERTAL GIRLS

Ovarian cysts occur in 2% to 5% of prepubertal girls.[12] During the first months of life, ovarian cysts are generally functional cysts caused by maternal gonadotropin stimulation of the newborn ovary. Persistence of cysts or the finding of a solid or complex component (i.e., a component in which both cystic and solid elements are present) of the adnexal mass suggests other disorders, including Wilms tumors, neuroblastomas, or gastrointestinal tract abnormalities.

The older literature suggests a very high rate of malignancy for ovarian neoplasms in children, with germ cell tumors being the most frequent malignancy and dysgerminomas being the most common germ cell tumor.[13] More recently, it was reported that in girls younger than 10 years who undergo surgery for an adnexal mass, 60% of the masses were not neoplasms, and two thirds of the neoplasms were benign.[14] If there are signs of early sexual development, the child should be evaluated for precocious puberty or a hormonally functioning ovarian neoplasm.

If torsion of the adnexal mass is suspected, pelvic ultrasound may confirm this diagnosis. If a solid component of the adnexal mass is detected, the risk of a germ cell tumor must be considered and serum levels of appropriate tumor markers (i.e., AFP, LDH, and β-hCG) must be obtained.

Because of the rarity of gynecologic diseases in this age group, more common diagnoses, including acute appendicitis, intussusception of the bowel, gastroenteritis, genitourinary disorders, and chronic constipation, must be considered.

PELVIC MASS IN ADOLESCENTS

The differential diagnosis of a pelvic mass in an adolescent girl is broader than that for younger patients because adolescence is accompanied by functioning ovaries and the beginning of sexual activity. A pelvic mass may be caused by a benign neoplasm, a malignant neoplasm, an anatomic abnormality, or an ectopic pregnancy. Sexually transmitted disease (STD) and pelvic inflammatory disease (PID) must be included in the differential diagnosis.

The patient history should include questions about sexual activity, previous history of STD and PID, and use of contraception. Depending on the adolescent's sexual activity, a vaginal examination may also be appropriate. Adolescents deny sexual activity for multiple reasons, but studies show that 50% of adolescent girls have had sexual intercourse by 17 years of age.[15] In adolescents, a pelvic mass that is not associated with pregnancy may require further evaluation by abdominal CT or MRI. When anatomic abnormalities are suspected, the evaluation should include MRI. Genetic studies may also be indicated if a separate adnexal mass is found, because in 25% of the patients with a Y chromosome, dysgenetic gonads are malignant.[16]

Cystic Adnexal Masses

The majority of cystic adnexal masses are related to the normal physiologic ovary; such masses include follicular cysts, cor-

pus luteum cysts, and theca-lutein cysts. Usually unilocular and less than 8 to 10 cm in diameter, cystic adnexal masses commonly resolve within 6 to 8 weeks. Combination monophasic oral contraceptive pills with progestin, as well as estrogen at a dose higher than 50 μg, are reported to reduce the risk of further ovarian cysts. The use of lower-dose oral contraceptives may be less effective.[17,18]

Theca-lutein cysts that occur in pregnancy are usually bilateral, large, and multicystic. They are associated with high β-hCG levels.[3] Spontaneous resolution usually occurs post partum.

Cystic Teratoma

Overall, mature cystic teratomas (dermoid cysts) account for more than half of the ovarian neoplasms in children and adolescents younger than 20 years.[19] These neoplasms usually range from 5 to 10 cm in diameter; 15% are bilateral. Because they arise from pluripotential germ cell lines, they may contain hair, teeth, sebaceous material, neural elements, and other tissues not usually found in the ovary. Some of these elements show unique imaging patterns (e.g., calcified materials on plain radiograph and fat density on pelvic ultrasound).

Malignant Neoplasms

The risk of ovarian malignant neoplasms is lower in adolescents than in younger children.[14] In reports from referral centers, the rate of malignancy in ovarian neoplasms was 35% in prepubertal girls and adolescents; however, in community-hospital centers, the rate of malignancy was 10% in these patients.[20]

Germ cell tumors account for approximately 70% of the malignant ovarian tumors in girls younger than 15 years.[14] Dysgerminomas are the most common such tumors, followed by immature or malignant teratomas, endodermal sinus tumors, embryonal carcinomas, and choriocarcinomas. Stromal tumors and epithelial carcinomas each make up 15% of the ovarian tumors.[14]

Miscellaneous Masses

Pregnancy luteomas, sclerotic ovaries, and endometriotic cysts occur in adolescent girls. These may be incidental findings on physical examination, or they may be associated with symptoms of pain or irregular menses. Torsion, rupture, or leakage of the content of these cysts and subsequent peritoneal irritation may cause the pain. Laparoscopy may be required for full evaluation of the condition that causes the pain.

Anatomic Abnormalities

Around the time of expected menarche, anatomic abnormalities in the development of the müllerian system can cause obstruction of the uterovaginal outflow tract, resulting in a pelvic mass. The anomalies include imperforated hymen, transverse vaginal septa, vaginal agenesis with normal functional endometrium, vaginal duplications with obstructing longitudinal septa, and obstructed uterine horns. If these anomalies cause a blockage of the vagina or the uterus, a mass may develop; such masses may result in cyclic pain, prompting these women to seek treatment.

Ectopic Pregnancy

In women younger than 17 years, approximately 82% of pregnancies are unintended; 75% of pregnancies are unintended in women 18 and 19 years of age.[21] Ectopic pregnancy is associated with pelvic pain, an adnexal mass, and missed or irregular menses. The risk of ectopic pregnancy is increased in women who have a history of STD or PID, as well as in women who fail to use contraception; oral contraceptives lower the risk.

Inflammatory Processes

The differential diagnosis of a pelvic mass in adolescents includes several infectious processes. When a patient presents with fever, an elevated white cell count, and a lower abdominal, pelvic, or adnexal tender mass that is associated with cervical motion tenderness and mucopurulent cervical discharge, a tuboovarian abscess or pyosalpinx must be considered. Often, a history of recent unprotected intercourse with a new partner will help establish the diagnosis. Appendicitis must also be considered in this setting. In the less acute setting, a patient with a pelvic mass and a history of STD should be evaluated for PID. Laparoscopy may be useful in confirming the diagnosis of PID; the clinical diagnosis of PID has been reported to be incorrect in up to one third of patients.[22]

PELVIC MASS IN WOMEN OF REPRODUCTIVE AGE

The detection of an asymptomatic pelvic mass is frequent during the reproductive years because women undergo annual examinations for family planning and gynecologic cancer screening. The differential diagnosis includes all the conditions that may cause a pelvic mass in adolescents (i.e., cystic adnexal masses, cystic teratomas, malignant neoplasms, and ectopic pregnancy), as well as leiomyomas, endometriomas, and metastatic neoplasms involving the ovary.

In a series of 100 women undergoing laparotomy for a pelvic mass, the most common diagnoses by age group were cancer, reported in 56% in women 50 years of age or older; endometriosis, reported in 27% of women 31 to 49 years of age; and cystic teratomas, reported in 33% of women younger than 30 years. In women younger than 30 years, only 10% had an ovarian malignancy, and most of these were tumors of low malignant potential. Thus, most pelvic masses that occur during reproductive years will be benign uterine neoplasms or benign ovarian neoplasms.[23]

A pregnancy test is required in all women of reproductive age who present with a suspected pelvic mass. In pregnant women, the use of abdominal and pelvic CT scans must be avoided. If ultrasound studies are inconclusive, an MRI may help identify the source of the mass.[5] If patient age, physical examination, and findings on ultrasound suggest malignancy, it is appropriate to measure epithelial serum tumor markers, such as CA125. However, elevated levels of CA125 may be associated with normal gynecologic conditions, as well as benign uterine and ovarian neoplasms. Risk of malignancy increases with patient age, positive family history, severity of symptoms, and number of imaging findings consistent with malignancy.

Uterine Neoplasms

Epidemiology Leiomyomas (fibroids) are the most common benign uterine neoplasm. They can also rarely arise from the ovary, the cervix, the pelvic ligaments, or other pelvic structures.

Leiomyomas are clinically apparent on examination in approximately 25% of women, but there is a marked difference in racial groups. In the United States, in women 25 to 44 years of age, the incidence rates of leiomyomas that were confirmed by ultrasound or hysterectomy were 8.9 for white women and 30.6 for black women per 1,000 women-years.[24] When uteri are surgically removed for treatment of noncancerous presentations, pathologic examination reveals leiomyomas in 89% of black

women and 59% of white women.[25] Similar results are obtained when screening women with ultrasonography: by 50 years of age, more than 80% of black women and 70% of white women will show fibroids.[26]

Leiomyomas are hormonally dependent; thus, these benign neoplasms usually shrink after menopause. They also frequently increase in size during pregnancy, as well as with the use of high-dose exogenous estrogens and, occasionally, with tamoxifen.

Sarcomatous degeneration of leiomyomas is rare. The incidence is reported to range from 0.4% to 1.4%. However, rapid increase in the size of a leiomyoma raises concern, although the definition of rapid growth has not been quantified. In fact, in a retrospective review of 371 patients operated on for rapidly growing leiomyomas, the incidence of leiomyosarcoma was 0.23%. When rapidly growing leiomyoma was defined as an increase of 6 weeks' gestational size over 1 year, none of 198 patients who satisfied this criterion were found to have a sarcoma.[27]

Clinical manifestations Most women with leiomyomas are asymptomatic, but symptoms may occur during the third and fourth decades. Leiomyomas, which are usually nontender, are most frequently found on clinical pelvic examination; but increasingly, they are identified by pelvic ultrasound during evaluation of nonspecific abdominal or pelvic symptoms.

Symptomatic patients may complain of pelvic discomfort, pressure, pain, menorrhagia, and dysmenorrhea. With degeneration or infarction, severe lower abdominal or pelvic pain develops. This may be associated with fever and an elevated white cell count. If the leiomyoma is pedunculated, torsion may cause severe pain, which may be intermittent. The pain may become part of a chronic pelvic pain pattern. Urinary symptoms include urinary frequency from extrinsic pressure on the bladder or, rarely, urinary retention secondary to urethral obstruction from a cervical or lower uterine leiomyoma. Depending on the location and the size of the leiomyoma, rectosigmoid compression and constipation may develop.

Leiomyomas coming through the cervical os can cause severe cramping; if necrotic, a foul vaginal discharge may develop. Abnormal uterine bleeding may be associated with a leiomyoma that disrupts the endometrial lining; the bleeding associated with a leiomyoma is cyclic, occurring in response to ovarian hormones.

Leiomyomas are usually discrete, firm, rounded, rubbery masses; they can vary in size from several millimeters to masses large enough to fill the abdominal pelvic cavity. They can be hard (if calcified) or soft (if cystic). Usually they cause an asymmetrical enlargement of the uterus, but multiple small leiomyomas cause a symmetrically enlarged uterus. Within the uterus, the leiomyoma may be located within the myometrium, beneath the endometrial lining, or on the surface of the uterus. When pedunculated or located posterior in the cul-de-sac, leiomyomas can give the clinical impression of a solid adnexal mass.

Ovarian Neoplasms

About two thirds of all ovarian neoplasms are discovered during the reproductive years; however, in women younger than 45 years, the chance that such neoplasms are malignant is 5% to 18%.[23,28,29] The most common ovarian neoplasms are endometriomas, cystic teratomas, and epithelial ovarian neoplasms. Most ovarian neoplasms produce few specific symptoms, the most common being vague abdominal pelvic pain or discomfort, abdominal distention, pelvic pressure, and urinary or gastrointestinal symptoms. Occasionally, in hormonally active neoplasms, irregular vaginal bleeding may occur.

Endometriomas Endometriomas are benign ovarian masses arising from ectopic endometrial tissue. Their incidence has not been determined. Frequently, endometriomas partially or almost completely replace normal ovarian tissue. Bilateral involvement of the ovaries has been reported in one third to one half of cases.[30] Endometriomas, which are usually less than 15 cm in diameter, may spontaneously rupture or resolve.

Patients who have an endometrioma usually complain of pelvic pain, dysmenorrhea, and dyspareunia; often, patients have an established history of endometriosis and infertility. On imaging evaluation, a mass 6 to 8 cm in diameter may be found. Endometriomas may by characterized by septations, debris, or solid components. These masses may not resolve over time. Endometriomas may be accompanied by an elevation in the CA125 serum level, which may cause concern regarding a malignancy; generally, however, CA125 serum levels associated with endometriomas are less than 200 U/ml.

Cystic teratomas Cystic teratomas, or dermoid cysts, are benign ovarian germ cell tumors. More than 80% of cystic teratomas are diagnosed during the reproductive years.[31] In a 10-year retrospective review, cystic teratomas constituted 62% of all ovarian neoplasms in women younger than 40 years.[29] The malignant transformation of these tumors is less than 2% and mostly occurs in women older than 40 years. There is a 15% risk of torsion and a 10% chance of bilateral presentation.

The risk of epithelial ovarian neoplasms increases with age. Bilateral ovarian neoplasms carry a 2.6-fold increased risk of malignancy, as compared with unilateral neoplasms.[29] Other causes of ovarian enlargement in this age group include metastatic cancer, especially from the breast or the gastrointestinal tract.

PELVIC MASS IN POSTMENOPAUSAL WOMEN

A pelvic mass in postmenopausal women may arise from the gynecologic organs, but increasingly in this age group, a mass may arise from nongynecologic organs. With decreasing ovarian hormone production, leiomyomas should undergo regression, and functional ovarian cysts are less likely. Endometriotic tumors are also not usually found in this age group. Thus, a newly found pelvic mass raises the suspicion of a malignancy.

Epidemiology The incidence of ovarian cancer increases with age, and 30% to 60% of ovarian masses in women older than 50 years are malignant.[23] The average age of a woman when diagnosed with ovarian cancer is 56 to 60 years. The majority of these tumors are epithelial malignancies. Fallopian tube cancer is rare. The differential diagnosis for a pelvic mass in postmenopausal women includes colon cancer, which is the third most common cancer in women.

Ovarian cysts have been reported in 3% to 17% of asymptomatic postmenopausal women undergoing pelvic ultrasound. In a study of 83 patients with thin-walled ovarian cysts less than 5 cm in diameter, 43 underwent surgery; no ovarian cancers were found in this group. In the remaining patients, 32 underwent serial ultrasound studies. In this group, 12 cysts resolved, seven decreased in size, four remained unchanged, and one increased slightly in size. The remaining eight patients underwent cyst aspirations; all the aspirated cysts were benign.[32]

Clinical manifestations The presentation of ovarian cancer is not specific. Patients may complain of vague gastrointestinal symptoms, including dyspepsia, early satiety, anorexia, bloating, and, occasionally, constipation. In a retrospective survey of 1,725 patients with ovarian cancer, 95% reported symptoms that were categorized as abdominal (77%), gastrointestinal (70%), pain (58%), constitutional (50%), urinary (34%), and pelvic (26%).[33] Fallopian tube cancer may present as uterine bleeding, pelvic pain, and an adnexal mass. Classically, profuse watery vaginal discharge is seen, although this finding is rare.

The findings on examination consistent with advanced disease include abdominal distention with ascites, an abdominal/pelvic mass, and nodularity in the cul-de-sac on rectovaginal examination. An ultrasound of the abdomen and pelvis may show ascites, bilateral complex adnexal masses, and omental implants. In addition, there may be a pleural effusion. Evaluation of these patients should include abdominal and pelvic CT scans and chest x-ray. Again, if imaging studies raise the suspicion of malignancy, it is important to measure serum levels of CA125. An elevated CA125 level in postmenopausal women with a pelvic mass suggests a malignancy, because the positive predictive value for elevated CA125 (i.e., > 65 U/ml) in this age group has been reported to be 97%.[10] Early referral to a gynecologic oncologist is appropriate for patients with a mass that raises suspicion of malignancy.

A pelvic ultrasound may also identify nonmalignant cysts. Studies have suggested that women with simple cysts that are less than 10 cm in diameter and that are without any excrescences, septations, or ascites should undergo serial ultrasound studies.[34,35]

A pelvic mass in a postmenopausal woman with a history of bowel symptoms may suggest colon cancer; the evaluation of the stool may be positive for occult or frank blood. Diverticular disease must also be considered.

Management

The management of a pelvic mass depends on the patient's age, history, tumor characteristics, and likelihood of malignancy. For all age groups, surgery is required for masses that are greater than 10 cm in diameter and for those that are solid, fixed, or bilateral. When these findings are accompanied by significantly elevated levels of tumor markers, the presence of ascites, or a finding on imaging or physical examination that suggests malignancy, the patient should be referred to a gynecologic oncologist.

INFANTS AND PREPUBESCENT GIRLS

Infants and prepubescent girls with suspected physiologic or functional cysts should undergo serial ultrasound studies approximately every 6 weeks. Aspiration of unilocular cysts in prepubescent girls, either with ultrasound guidance or laparoscopy, is associated with a 50% recurrence rate[3] and is usually not recommended. If the mass increases in size, persists after 6 months, or becomes complex, surgery via either laparotomy or laparoscopy is necessary. Conservative management is indicated if the malignancy is confined to one ovary. Consultation with a gynecologic oncologist is necessary.

ADOLESCENT GIRLS

The management of adnexal masses in adolescent girls should have as its aim the preservation of ovarian function. An asymptomatic simple cyst that measures less than 10 cm in diameter may be observed and followed with serial ultrasound studies. To prevent new formation of physiologic cysts, ovarian suppression with oral contraception should begin. If the cyst increases in size, becomes complex, or causes symptoms, then surgery should be performed.

For benign neoplasms, cystectomy is recommended. Because most malignant tumors are unilateral, only the ovary or adnexa need be removed. The contralateral ovary should be inspected. If the ovary appears grossly normal, biopsy need not be performed, nor should the ovary be bivalved (i.e., the surface of the ovary divided and the cortex inspected), because these procedures could lead to peritubal or periovarian adhesions. However, if suspicious areas are identified, biopsies must be performed and a frozen section ordered for histiologic analysis during surgery. Histiologic analysis will help determine malignancy and indicate the need for consultation with a gynecologic oncologist to establish surgical staging. If the frozen section does not clearly establish malignancy, a second surgery is preferable to performing unnecessary initial surgery. However, the choice of surgery should be undertaken cautiously, because any adnexal surgery may result in tubal adhesions, which could interfere with future fertility.

PID in adolescents should be managed medically. Surgical management of nonmalignant presentations is rarely indicated in adolescents. Surgery, however, may be required to treat a ruptured tubo-ovarian abscess; it may also be required if the disease fails to respond to broad-spectrum antibiotics. Ectopic pregnancy can be managed medically, providing the pregnancy is small and the patient is hemodynamically stable; surgical management is required if these conditions are not met. If surgery is indicated, the procedure should be conservative and aimed at preserving fertility.

WOMEN OF REPRODUCTIVE AGE

The management of a pelvic mass in women of reproductive age will depend on the malignant potential of the mass. Most often, the mass will be a benign uterine leiomyoma. The initial approach will depend on whether the patient is symptomatic and has completed childbearing.

Leiomyomas Asymptomatic leiomyomas should be followed with periodic pelvic examinations to ensure that there is not a rapid growth in size. The clinical records should document the location; a pelvic ultrasound can more accurately estimate the size. Rapid growth in the postmenopausal years may indicate transformation into a sarcoma. The risk, however, is reported to be less than 2 to 3 per 1,000.

In patients who will soon enter menopause or who are planning to undergo surgery for mildly symptomatic leiomyomas, hormonal therapy using gonadotropin-releasing hormone (GnRH) analogues results in a 40% to 60% decrease in uterine volume. GnRH treatment causes hypoestrogenic states that result in bone loss and hot flashes. Regrowth of the leiomyoma occurs within a few months of treatment cessation in one half of patients. Use of GnRH may be considered (1) as neoadjuvant therapy to shrink the size of the leiomyoma before surgery to permit a vaginal approach, (2) as treatment for anemia secondary to hemorrhage associated with leiomyomas, and (3) as treatment in perimenopausal women in an effort to avoid surgery.

Symptomatic leiomyomas require surgery. The usual indications include abnormal uterine bleeding with anemia that is unresponsive to hormone therapy; chronic pelvic pain with dysmenorrhea and dyspareunia; acute pelvic pain associated with

torsion of pedunculated leiomyoma; prolapsing leiomyoma; urinary frequency with hydronephrosis; and symptoms of pelvic or rectal pressure caused by a significantly enlarged leiomyoma. Rarely, infertility caused by a leiomyoma obstructing the fallopian tubes or loss of a pregnancy secondary to a leiomyoma may be an indication for myomectomy. The finding of a mass during pregnancy demands the same management approach as for a nonpregnant patient. If surgery is necessary during pregnancy, the second trimester is the safest period.

Once childbearing is complete, hysterectomy is traditionally the definitive management for symptomatic leiomyomas. However, other treatment options have become available, including laparoscopic myomectomy and hysteroscopic resection of submucosal leiomyoma. In addition, endometrial ablation (e.g., laser, thermal, or chemical ablation, as well as selective arterial embolization) can decrease the bleeding caused by intramural leiomyomas.[36]

Endometriomas Endometriomas that do not spontaneously resolve are managed with surgical excision.

Cystic teratomas In women younger than 45 years, the treatment for a cystic teratoma is ovarian cystectomy, which often can be performed laparoscopically, especially if the mass is less than 10 cm in diameter.

Epithelial ovarian neoplasms Surgical management of epithelial ovarian neoplasms includes removal of the adnexa and surgical staging. Whether the surgery can be conservative (i.e., a unilateral salpingo-oophorectomy) will depend on the extent of the disease and the degree of malignancy (i.e., invasive tumor versus tumor of low malignant potential). These decisions are made in consultation with a gynecologic oncologist.

POSTMENOPAUSAL WOMEN

The risk of a malignancy increases with age, and thus, the threshold for conservative management decreases in postmenopausal women. It has been reported that a suspicious mass seen on ultrasound combined with a CA125 level greater than 65 U/ml has a specificity of 96.1%, a sensitivity of 91.7%, and an accuracy of 94.3% for detecting an ovarian neoplasm in postmenopausal women.[37] On the other hand, a postmenopausal woman with an ovarian simple cyst that is less than 5 cm in diameter and a normal CA125 serum concentration has a 0% risk of malignancy. Thus, the former patient should be referred to a gynecologic oncologist for appropriate management, whereas the latter may be followed with serial ultrasound studies every 4 to 6 months for a year and, provided the tumor remains stable and the patient asymptomatic, annually thereafter.

Joseph T. Chambers, M.D., Ph.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

Carolyn D. Runowicz, M.D., has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. The adnexal mass and early ovarian cancer. Clinical Gynecologic Oncology, 6th ed., DiSaia PJ, Creasman WT, Eds. Mosby, St. Louis, 2001, p 259

2. Slanetz PJ, Hahn PF, Hall DA, et al: The frequency and significance of adnexal lesions incidentally revealed by CT. AJR Am J Roentgenol 168:647, 1997

3. Adams Hillard PJ: Benign diseases of the female reproductive tract: symptoms and signs. Novak's Gynecology, 13th ed. Novak E, Berek JS, Eds. Lippincott Williams & Wilkins, Philadelphia, 2002, p 351

4. Ascher SM, Imaoka I, Hricak H: Diagnostic imaging techniques in gynecologic oncology. Principles and Practice of Gynecologic Oncology, 3rd ed. Hoskins WH, Perez CA, Young RC, et al, Eds. Lippincott Williams & Wilkins, Philadelphia, 2000, p 655

5. Kier R, McCarthy SM, Scoutt LM, et al: Pelvic masses in pregnancy: MR imaging. Radiology 176:709, 1990

6. Jacobs I, Bast RC Jr: The CA 125 tumor-associated antigen: a review of the literature. Hum Reprod 4:1, 1989

7. Rosenthal A, Jacobs I: Ovarian cancer screening. Semin Oncol 25:315, 1998

8. Bast RC Jr, Klug TL, St. John E, et al: A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 309:883, 1983

9. Einhorn N, Sjovall K, Knapp RC, et al: Prospective evaluation of serum CA 125 levels for early detection of ovarian cancer. Obstet Gynecol 80:14, 1992

10. Brooks SE: Preoperative evaluation of patients with suspected ovarian cancer. Gynecol Oncol 55:S80, 1994

11. Mulvany NJ: Aspiration cytology of ovarian cysts and cystic neoplasms: a study of 235 aspirates. Acta Cytol 40:911, 1996

12. Russell DJ: The female pelvic mass: diagnosis and management. Med Clin North Am 79:1481, 1995

13. Norris HJ, Jensen RD: Relative frequency of ovarian neoplasms in children and adolescents. Cancer 30:713, 1972

14. van Winter JT, Simmons PS, Podratz KC: Surgically treated adnexal masses in infancy, childhood and adolescence. Am J Obstet Gynecol 170:1780, 1994

15. Abma JC, Sonenstein FL: Sexual activity and contraceptive practices among teenagers in the United States, 1988 and 1995. Vital Health Stat 21:1, 2001

16. Schellhas HF: Malignant potential of the dysgenetic gonad. Part 1. Obstet Gynecol 41:74, 1974

17. Holt VL, Daling JR, McKnight B, et al: Functional ovarian cysts in relation to the use of monophasic and triphasic oral contraceptives. Obstet Gynecol 79:529, 1992

18. Lanes SF, Birmann B, Walker AM, et al: Oral contraceptive type and functional ovarian cysts. Am J Obstet Gynecol 166:956, 1992

19. Kozlowski JD: Ovarian masses. Adolesc Med 10:337, 1999

20. Diamond MP, Baxter JW, Peerman GCJ, et al: Occurrence of ovarian masses in childhood and adolescence: a community-wide evaluation. Obstet Gynecol 71:858, 1988

21. Henshaw SK: Unintended pregnancy in the United States. Fam Plann Perspect 30:24, 1998

22. Cibula D, Kuzel D, Fucikova Z, et al: Acute exacerbation of recurrent pelvic inflammatory disease: laparoscopic findings in 141 women with a clinical diagnosis. J Reprod Med 46:49, 2001

23. Hernandez E, Miyazawa K: The pelvic mass: patients' age and pathologic findings. J Reprod Med 33:361, 1988

24. Marshall LM, Spiegelman D, Barbieri RL, et al: Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. Obstet Gynecol 90:967, 1997

25. Kjerulff KH, Langenberg P, Seidman JD, et al: Uterine leiomyomas: racial differences in severity, symptoms and age at diagnosis. J Reprod Med 41:483, 1996

26. Day Baird D, Dunson DB, Hill MC, et al: High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol 188:100, 2003

27. Parker WH, Fu YS, Berek JS: Uterine sarcoma in patients operated on for presumed leiomyoma and rapidly growing leiomyoma. Obstet Gynecol 83:414, 1994

28. Killackey MA, Neuwirth RS: Evaluation and management of the pelvic mass: a review of 540 cases. Obstet Gynecol 71:319, 1988

29. Koonings PP, Campbell K, Mishell DR Jr, et al: Relative frequency of primary ovarian neoplasms: a 10-year review. Obstet Gynecol 74:921, 1989

30. Egger H, Weigmann P: Clinical and surgical aspects of ovarian endometriotic cysts. Arch Gynecol 233:37, 1982

31. Talerman A: Germ cell tumors of the ovary. Blaustein's Pathology of the Female Genital Tract, 5th ed. Blaustein A, Kurman RJ, Eds. Springer Verlag, New York, 2002, p 998

32. Kroon E, Andolf E: Diagnosis and follow-up of simple ovarian cysts detected by ultrasound in postmenopausal women. Obstet Gynecol 85:211, 1995

33. Goff BA, Mandel L, Muntz HG, et al: Ovarian carcinoma diagnosis. Cancer 89:2068, 2000

34. Aubert JM, Rombaut C, Argacha P, et al: Simple adnexal cysts in postmenopausal women: conservative management. Maturitas 30:51, 1998

35. Bailey CL, Ueland FR, Land GL, et al: The malignant potential of small cystic ovarian tumors in women over 50 years of age. Gynecol Oncol 69:3, 1998

36. ACOG committee opinion: uterine artery embolization. Obstet Gynecol 103:403, 2004

37. Maggino T, Gadducci A, D'Addario V, et al: Prospective multicenter study on CA 125 in postmenopausal pelvic masses. Gynecol Oncol 54:117, 1994

XVI APPROACH TO THE PATIENT WITH AN ABNORMAL PAP SMEAR

Carolyn D. Runowicz, m.d.

Cervical cancer is the third most common gynecologic cancer in the United States. An estimated 10,370 new cases of invasive cervical cancer and 3,710 deaths occur annually, representing 1.4% of cancer deaths in women.[1] In the United States, the incidence of cervical cancer decreased by more than 70% between 1950 and 2000, largely as a result of screening[2]; more than half of the incident cases of invasive cancer are diagnosed in women who have not been adequately screened. Screening cytology methods, such as the Papanicolaou (Pap) smear, are excellent means of identifying preinvasive disease; however, false positive rates are relatively high. Each year, approximately 3.5 million cervical cytologic tests are interpreted as indicating an abnormality requiring additional follow-up or evaluation.[3]

Epidemiology and Risk Factors for Cervical Cancer

The occurrence of invasive cervical cancer is related to age. Premalignant cervical lesions are usually diagnosed in women younger than 40 years, which is 10 to 15 years earlier than in women diagnosed with invasive cervical cancer. This age gap suggests a long latency period for malignant transformation. For example, the diagnosis of cervical intraepithelial neoplasia (CIN) is usually made in women in their twenties, whereas the diagnosis of carcinoma in situ (CIS) is made in women 25 to 35 years of age, and invasive cancer is diagnosed in women older than 40 years.

Infection with high-risk strains of human papillomavirus (HPV) is the most important risk factor for cervical cancer.[4,5] HPV infection is usually transient[6,7]; however, persistent infection by high-risk HPV virus—most commonly, subtypes 16, 18, 31, and 45—is a prerequisite for the development of grade 1 and 2 CIN and invasive cervical cancer.[8]

HPV is usually acquired sexually; high-risk sexual behavior (e.g., having multiple sexual partners and promiscuous sexual partners) increases the risk for exposure to the virus.[9] HPV is a necessary precursor of CIN, but it does not act alone; host factors such as age, immune function,[10] a history of sexually transmitted disease (e.g., *Chlamydia trachomatis*),[11] and smoking[12] are surrogate markers for oncogenesis.

Natural History of Cervical Cancer

Early invasive cervical cancer is frequently asymptomatic, a fact that underscores the importance of routine screening. An abnormal Pap smear is frequently the first indication of a precancerous condition. The abnormalities observed on a cytologic smear or tissue biopsy of the cervix represent alterations in the degree of differentiation of cervical epithelial cells. An understanding of the natural history of low-grade and high-grade CIN lesions is central to the clinical management of patients who have abnormal cervical cytology.

CIN is a preinvasive pathologic intermediate of cervical cancer; it is slow to progress and can be easily detected and treated. The severity of CIN is designated by the extent to which the lesion involves the epithelial thickness. CIN 1 refers to intraepithelial neoplasia in which cellular changes are confined to the basal third of the epithelium; CIN 2 refers to intraepithelial neoplasia in which cellular changes are confined to the basal two thirds of the epithelium; and CIN 3 refers to cellular dysplasia encompassing more than two thirds of the epithelial thickness, including full-thickness lesions. CIS demonstrates full-thickness evidence of neoplasia without invasion of the basement membrane; CIN 3 and CIS may persist unchanged for 10 to 15 years, but eventually the lesion progresses to invasive carcinoma.

Low-grade lesions do not necessarily progress to high-grade lesions. A large cohort study indicated that CIN 1 lesions regressed to normal within 2 years in 44% of patients; they regressed to normal within 5 years in 74% of patients.[13] This series noted that the rates of progression of CIN 1 at 2 and 5 years were 2% and 6%, respectively; rates of progression of CIN 2 were 16% and 25%, respectively.

A primary goal of cervical cytologic screening is to identify women at risk for high-grade lesions; however, cytologic results are often equivocal. To improve the accuracy of cytologic interpretation, a standard system of terminology was adopted to distinguish findings most likely to be precancerous [*see* Test Interpretation—the Bethesda Reporting System, *below*]. Essentially, this reporting system classified a broad range of atypical findings into two categories: those findings that were more likely to represent high-risk lesions and those whose significance was undetermined.

Screening for Cervical Cancer

who should be screened?

Abundant evidence indicates that regular gynecologic examinations and cervical cytology decrease cervical cancer incidence and mortality. However, among policy-making organizations, there is some variation in the recommendations concerning the age at which screening should start, the interval of screening, and the age at which routine screening should stop.

Commencement of Screening

The United States Preventive Services Task Force (USPSTF) recommends beginning cytologic screening within 3 years of onset of sexual activity or by age 21, whichever comes first.[14] The recommendations of the American College of Obstetricians and Gynecologists (ACOG)[15] and the American Cancer Society (ACS)[16] are consistent with these guidelines. Other North American organizations, such as the Canadian Task Force on Preventive Health Care (CTFPHC), recommend that screening begin at onset of sexual activity or at 18 years of age.[17]

There is little value in screening women who have never been sexually active; however, many North American organizations recommend routine screening by age 18 or 21 on the basis of the generally high prevalence of sexual activity by that age and concerns that clinicans may not always obtain accurate sexual histories.

Table 1 The 2001 Bethesda System (Abridged)[29]

- Specimen adequacy
  - Satisfactory for evaluation *(note presence or absence of endocervical/transformation zone component)*
  - Unsatisfactory for evaluation *(specify reason)*
    - Specimen rejected/not processed *(specify reason)*
    - Specimen processed and examined but unsatisfactory for evaluation of epithelial abnormality because of *(specify reason)*
- General categorization (optional)
  - Negative for intraepithelial lesion or malignancy
  - Epithelial cell abnormality
  - Other
- Interpretation/result
  - Negative for intraepithelial lesion or malignancy
    - Organisms
      - *Trichomonas vaginalis*
      - Fungal organisms morphologically consistent with *Candida* species
      - Shift in flora suggestive of bacterial vaginosis
      - Bacteria morphologically consistent with *Actinomyces* species
      - Cellular changes consistent with herpes simplex virus
    - Other non-neoplastic findings (optional to report; list not comprehensive)
      - Reactive cellular changes associated with the following:
        - Inflammation (includes typical repair)
        - Radiation
        - Intrauterine contraceptive device
      - Glandular cells status post hysterectomy
      - Atrophy
  - Epithelial cell abnormalities
    - Squamous cell
      - Atypical squamous cells (ASC)
        - Of undetermined significance (ASC-US)
        - Cannot exclude HSIL (ASC-H)
      - Low-grade squamous intraepithelial lesion (LSIL) (cellular changes consistent with HPV, mild dysplasia, CIN 1)
      - High-grade squamous intraepithelial lesion (HSIL) (moderate to severe dysplasia, CIN 2, CIN 3, CIS) *(indicate if there are features suspicious of invasion)*
      - Squamous cell carcinoma
    - Glandular cell
      - Atypical glandular cells *(specify endocervical, endometrial, or not otherwise specified)*
      - Atypical glandular cells, favor neoplastic *(specify endocervicalor not otherwise specified)*
      - Endocervical adenocarcinoma in situ (AIS)
      - Adenocarcinoma
  - Other (list not comprehensive)
    - Endometrial cells in a women ≥ 40 years

CIN—cervical intraepithelial neoplasia CIS—carcinoma in situ

Screening Interval

Because cervical cancer is slow growing, considerable uncertainty surrounds the issue of the screening interval. The USPSTF found no direct evidence that annual screening achieves better outcomes than screening every 3 years. The most direct evidence on which to base a recommendation of a screening interval comes from a prospective cohort analysis of a randomized controlled trial.[18] Among 2,561 women (mean age, 66.7 years) with normal Pap tests at baseline, 110 had an abnormal Pap test within the next 2 years. No woman was found to have CIN 2, CIN 3, or invasive cancer; only one woman had CIN 1 or CIN 2. Thus, the positive predictive value of screening 1 year after a negative Pap test was 0%; after 2 years, the positive predictive value was 0.9%. The authors concluded that Pap tests should not be repeated within 2 years after a negative test. A large study of women younger than 65 years, which included data from the National Breast and Cervical Cancer Early Detection Program and which used a model to estimate the rate at which intraepithelial neoplasia progresses to cancer, found that little further mortality reduction from cervical cancer was achieved by screening every year as compared with every 3 years.[19]

On the basis of the limited evidence, the USPSTF recommends screening at least every 3 years. The ACS guidelines recommend waiting until age 30 before lengthening the screening interval from 1 to 3 years. The ACOG recommends initiating screening with annual smears for 2 or 3 years; if these are negative, intervals of up to 3 years may be appropriate.[20] The ACOG identifies additional risk factors that might justify annual screening, including a history of cervical neoplasia, infection with HPV or other sexually transmitted diseases, or high-risk sexual behavior; however, data by which to determine the benefits of these strategies are limited.[20]

Cessation of Screening

The USPSTF recommendations state that screening can be discontinued in women who have had a total hysterectomy for benign disease (i.e., disease in which there is no evidence of cervical neoplasia or cancer), given the low yield of screening and the potential harms from false positive results in this population.[18] In women with a cervix, the optimal age to discontinue screening is not clear, but the risk of cervical cancer and the yield of screening decline steadily through middle age. However, screening is recommended in older women who have not been previously screened or when information about previous screening is unavailable.[14]

The USPSTF recommends discontinuing routine screening for women older than 65 years who have had adequate recent screening with normal Pap smears and are not otherwise at high risk for cervical cancer.[18] The ACS guidelines recommend that screening can be safely stopped in older women who have had three or more documented, consecutive, technically satisfactory, normal cervical cytologic tests and who have had no abnormal cytologic tests within the past 10 years; routine screening may be discontinued at age 70.[16]

AVAILABLE TESTS

There are several methods for cervical cancer screening: the conventional cytologic Pap smear, the liquid-based Pap smear, and HPV DNA testing in combination with cervical cytology. The purpose of these tests is to screen for cellular abnormalities that are associated with an increased risk of the development of cervical cancer.

Conventional Cytology (Pap Smear)

The Pap smear is the standard screening test for genital tract neoplasia. The reported sensitivity of a single Pap smear varies widely, ranging from 32% to 92%.[21] This low sensitivity prompted the recommendation found in early guidelines that cytologic screening be made annually.[22] False negative rates have been at-

tributed to poor sample preparation, in which precursor cells were obscured by blood, pus, air-drying artifacts, and other cells.[23] A conventional Pap smear costs $25 to $40.

Liquid-Based Cervical Cytology

Liquid-based cytology offers higher sensitivity and comparable specificity to that of the conventional Pap smear.[24] The liquid-based test costs $45 to $60. If used at 3-year intervals, the liquid-based test is cost-effective.[25] Evidence-based reports show that both liquid-based and conventional cytology are acceptable screening tests.[26] One advantage of liquid-based cytology is that HPV testing can be performed on the same preparation (see below).

Cytology and HPV DNA Testing

HPV DNA testing in combination with conventional or liquid-based cytology has been approved by the Food and Drug Administration for primary screening for cervical cancer in women older than 30 years.[27] In this age group, the combination of cytology and HPV DNA testing has been reported to have a sensitivity approximately 10% to 20% greater than that of a single conventional cytologic smear; however, specificity is lower.[28] Because of the high negative predictive value of these combined tests, women who test negative on both the cytologic and HPV DNA testing can increase their screening interval to 3 years.[27]

TEST INTERPRETATION—THE BETHESDA REPORTING SYSTEM

The most common abnormal cervical cytologic result is one of uncertainty. Interpretation of equivocal cytologic findings is complicated by confusion among laboratories and clinicians concerning the use of multiple classification systems and inconsistently defined numerical grading conventions. The Bethesda System, which was introduced in 1988 and is periodically updated (most recently in 2001), was devised as a uniform system of terminology to guide the interpretation of cytologic findings. A significant contribution of the Bethesda System was the standardized laboratory report that includes a description of specimen adequacy (to improve the consistency and quality of reporting) and that uses simplified terminology for the interpretation of equivocal cytologic findings [*see Table 1*].

Specimen Adequacy

The 2001 update of the Bethesda System qualifies specimens as being either satisfactory or unsatisfactory for evaluation.[29] Minimal squamous cellularity varies with the specimen type: an estimated 8,000 to 12,000 well-visualized squamous cells are acceptable for conventional smears, and 5,000 squamous cells are acceptable for liquid-based preparations. Epithelial cells may be obscured by blood or inflammation and still be considered satisfactory; however, if more than 75% of epithelial cells are obscured, the specimen is unsatisfactory. For specimens containing adequate squamous cellularity, the cytologic report notes the presence or absence of an endocervical/transformation zone component. Adequate endocervical cellularity consists of at least 10 well-preserved endocervical or squamous metaplastic cells.

Interpretaton of Specimen

The Bethesda System stipulates that cervical cytology is primarily a screening test, and the interpretation of morphologic findings described by the cytologic report must be integrated into a clinical context to establish a diagnosis.

Specimens are broadly defined as negative for intraepithelial lesion or malignancy or positive for epithelial cell abnormality. Epithelial abnormalities include atypical squamous cells, low- and high-grade squamous intraepithelial lesions (LSIL and HSIL), and atypical glandular cells (AGC). A finding of atypical squamous cells that cannot be determined as precancerous is the most common result, and its correct interpretation poses a clinical challenge.

Atypical squamous cells The 2001 Bethesda System qualified a finding of atypical squamous cells (ASC) in two ways: (1) ASC of undetermined significance (ASC-US) and (2) ASC for which HSIL cannot be excluded (ASC-H) [*see Table 1*]. A finding of undetermined significance emphasizes that some cases of ASC-US are associated with underlying CIN 2,3. ASC-H is used when there are cytologic features suggestive of HSIL but definite evidence is lacking. The ASC-H category constitutes approximately 5% to 10% of all ASC, but it includes women at greatest risk for CIN 2,3.[30,31] HSIL is more often associated with viral persistence and higher risk of progression, whereas LSIL is generally the result of a transient infection of HPV.[32,33]

Squamous epithelial lesions The Bethesda System classifies squamous intraepithelial lesions as low-grade (LSIL) or high-grade (HSIL). Cellular changes consistent with HPV, mild dysplasia, and CIN 1 are combined within the category of LSIL. Moderate to severe dysplasia, CIN 2, CIN 3, and CIS are combined within the category of HSIL. In the Bethesda System, CIN and dysplasia terminology can be used either as substitute terms for squamous intraepithelial lesions or as additional descriptors of intraepithelial lesions.

Atypical glandular cells The Bethesda System classifies glandular cell abnormalities into three types: atypical endocervical cells, endometrial cells, and glandular cells. In the majority of cases, morphologic features permit differentiation between atypical endometrial and endocervical cells. The management of patients with glandular abnormalities may vary significantly, depending on cell type, and distinguishing between these cell types is justified, when possible. The Bethesda System distinguishes AGC (either endocervical, endometrial, or AGC that are not otherwise specified [AGC-NOS]) from AGC (either endocervical or AGC-NOS) that favor neoplasia, because these two categories are associated with different degrees of risk of significant disease. Biopsy-confirmed high-grade lesions, including CIN 2,3, adenocarcinoma in situ, and invasive cancer, have been found in 9% to 41% of women who have AGC-NOS, as compared with 27% to 96% of those who have AGC that favors neoplasia.[34-36]

Management of Cytologic Abnormalities

ATYPICAL SQUAMOUS CELLS

Each year, an estimated two to three million women are diagnosed as having cervical cytology containing ASC.[37,38] The majority of women with ASC do not have a clinically significant lesion. However, 5% to 17% of patients with atypical ASC cytology have CIN 2,3 confirmed by biopsy.[30,39,40] A large, prospective study of routinely screened women reported that 39% of cases of high-grade squamous lesions were detected in women with ASC cytology.[41]

Increased age and a history of treatment of CIN have been reported to increase the risk of CIN in patients with ASC cytology.[42] High viral levels of HPV types known to be associated

with cervical cancer have been found to be strongly predictive of high-grade CIN in patients with ASC Pap smears.[43] ASC findings suggestive of neoplastic processes carry a greater risk of high-grade lesions and carcinoma than ASC findings suggestive of reactive processes.

Management Options

The evaluation and management of women with ASC-US cytology is a topic of considerable controversy. The best approach to the management of abnormal cervical cytology is to use the fewest number of tests to resolve the clinical question posed by the presence of ASC. Management options include (1) repeat cytology at a designated interval, (2) high-risk HPV DNA testing, (3) immediate colposcopy, (4) a combination of repeat cervical cytology and HPV DNA testing, (5) direct visual inspection in combination with conventional cytology, (6) referral for colposcopy, and (7) a combination of these strategies.

Repeat cervical cytology A repeat cervical cytology obtained at a later date appeals to many health care providers, because most histologic abnormalities found with ASC-US will be low-grade cervical neoplasia that is likely to regress without definitive therapy. The repeat Pap smear should be obtained 4 to 6 months after the index cytology [*see Figure 1*]. This is considered the optimal interval; a repeat cytology obtained after less than 4 months is thought to be associated with decreased sensitivity,[40] and after 6 months, 7% of patients with ASC-US smears will have CIN 2,3.[44]

High-risk HPV DNA testing HPV DNA is associated with virtually all cervical cancers and high-grade precursor lesions; the identification of these types of HPV that are associated with a high risk of oncogenesis is useful in identifying patients with ASC-US who are at increased risk for neoplasia.[45]

The probability of HPV expression is influenced by several factors, including age, menstrual cycling, use of exogenous hormones, and immunocompetence.[10,46] The addition of HPV DNA testing to cervical cytologic screening for women with ASC-US may be most effective in women 30 to 35 years of age, who are past the peak of incidence of acute infections [*see Table 2*].[46] In women whose cytology was classified as ASC-US, those who were 30 years of age or older had a lower prevalence of HPV positivity and a lower referral to colposcopy than younger women (30% versus 65%); this improvement in specificity was not accompanied by a decrement in sensitivity. Establishing age- and population-specific analytic cutoff points for the detection of HPV may further enhance the specificity of HPV DNA testing. Other strategies for improving specificity without compromising sensitivity may include the use of serial HPV DNA testing or lengthening the interval between the index ASC-US cytology and HPV DNA testing.

Of the commercial HPV DNA detection kits, Hybrid Capture 2 (HC2) (Digene Diagnostics, Silver Spring, Maryland) has a higher sensitivity (94.8% versus 84.4% for high-grade precursor lesions) and detects a broader range of high-risk HPV types than other methods.[47] HC2 is a nonradioactive, rapid assay that can detect 18 HPV DNA types,[48] 13 of which convey high risk of oncogenesis. A retrospective study of 398 women showed the HC2 assay to be as sensitive and specific as a single repeat cytologic smear for the detection of CIN.[40] However, it has been argued by some that its use identifies large numbers of low-grade lesions (CIN 1) and transient HPV infections and may lead to overevaluation and overtreatment because of its low positive predictive value.[47]

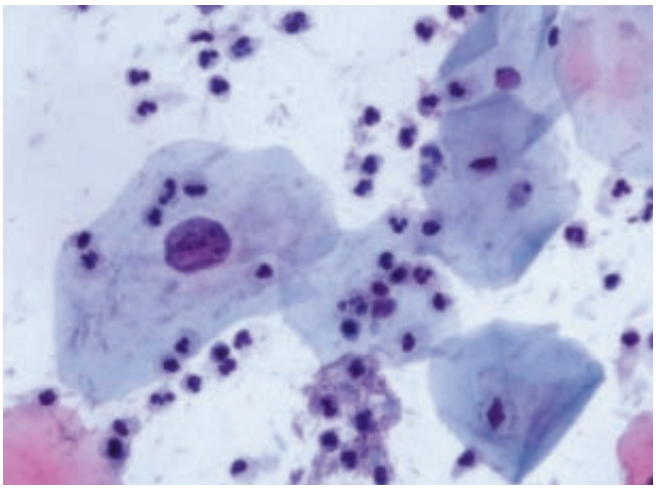


Figure 1 **A Pap smear showing atypical squamous cells of undetermined significance (ASC-US).**

Immediate colposcopy Colposcopy would be expected to detect almost all of the cases of high-grade CIN, but it has drawbacks, including expense and the risk of overevaluation and overtreatment in women who do not have CIN. Immediate colposcopy, however, may be indicated for women with a history of CIN, for poorly compliant women, and for women for whom waiting creates undue anxiety.

Studies of the natural history of HPV infection have demonstrated that most HPV infections produce only minor, transient infections.[10,11] A positive HPV DNA test in combination with an ASC-US cervical cytologic finding does not indicate high-grade disease or the presence of cancer, but it does indicate some increased risk of cancer now or in the future. Because of imperfect sensitivity, the initial colposcopy and directed biopsy will not detect about 10% of those women who will have histologically confirmed CIN 2,3 within 2 years of follow-up.[49]

Repeat cytology and direct visualization To compensate for the low sensitivity of a single repeat cervical cytology in women with atypical squamous cells, it has been suggested that repeat cytology be combined with a visual screening method. Direct visual screening methods include cervicoscopy (direct visual inspection of the cervix after an acetic acid wash), speculoscopy (direct visualization of the cervix under low magnification after application of an acetic acid wash), and cervicography (visual inspection of the cervix in which a static photographic image is used to document cervical abnormalities after an acetic acid wash).[50] Several studies that evaluated the combined use of cervicography and repeat cytology reported that this screening approach had a high sensitivity for the detection of CIN 2,3.[51,52] Because the data are limited, more studies are needed before recommending repeat cytology and direct visualization as a screening approach in patients with ASC-US.

Repeat cytology and HPV DNA testing Several studies have reported that the combination of HPV DNA testing and a repeat Pap smear has a sensitivity similar to that of colposcopy for the detection of high-grade CIN.[43,53-55] The negative predictive

value of DNA testing for high-risk types of HPV is generally reported to be 98% or greater.[40,56] However, other investigators question the cost-effectiveness of this combined screening strategy in women with ASC-US.[57,58] Use of liquid-based cytology permits the residual transport fluid to be used to test for HPV DNA (a technique referred to as reflex HPV DNA testing).[39] In a large study, Manos and colleagues evaluated HPV DNA testing of residual material from liquid-based cervical cytology using the HC2 assay. An overall sensitivity of 96.9% (95% confidence interval [CI], 88.3% to 99.5%) was achieved when colposcopy of HPV DNA–positive women was performed immediately after the reflex HPV DNA test. The authors concluded that reflex HPV DNA testing of cervical cytology aids in identifying those women at risk for high-grade lesions.[44]

The initial report of the ASC-US/LSIL Triage Study (ALTS) found that reflex HPV DNA testing was more sensitive than a single repeat cytologic smear in detecting CIN 3 in women with ASC-US cytology.[40] The sensitivity of HPV DNA testing for the detection of CIN of grade 3 or higher was 96.3% (95% CI, 91.6% to 98.8%); 56.1% (95% CI, 54.1% to 58.1%) of patients were referred for colposcopy. The sensitivity of a single repeat cytologic specimen for the detection of ASC-US or findings associated with higher risk was 85.3% (95% CI, 78.2% to 90.8%); 58.6% (95% CI, 56.5% to 60.6%) of patients were referred for colposcopy. However, the conventional clinical management strategy for cytologic follow-up is based on a series of repeat cytologic specimens, not on the sensitivity of a single cytologic sampling. A 2-year follow-up of the ALTS showed that a repeat cytologic specimen is as sensitive as HPV DNA testing at an ASC-US referral threshold, but this approach requires two follow-up visits and ultimately more colposcopic examinations than HPV triage (67.1% versus 53.1%).[59]

Reflex HPV DNA testing offers advantages over HPV DNA testing using conventional Pap smears: women do not need an additional clinical examination for specimen collection, and 40% to 60% of women with negative test results are spared a colposcopic examination. However, with this approach, the number of patients needing colposcopies remains high. The combination of HPV DNA testing and cervical cytology increases sensitivity at the expense of specificity, because even potentially oncogenic HPV-type infections are found in women without cervical neoplasia, particularly young, sexually active women. The positive predictive value of HPV DNA testing is similar to that of cytology (18% to 25%), but the negative predictive value is 99.8% to 100%.[60-62]

The drawbacks of reflex HPV DNA testing lie in its tendency to produce false negative and false positive results. A screening study of cervical cytology found a substantial degree of cross-reactivity (6.4%) between the reflex HPV DNA probe and the HPV types not included on the probe; this was detected by retesting all HPV DNA–positive samples with a polymerase chain reaction (PCR) assay. The PCR retest found a significant false positive rate of 3.6% and a false negative rate of 6.1% of all samples defined by the reflex HPV DNA test.[63] Thus, reflex HPV DNA testing has the potential for both overevaluation and underevaluation of high-risk HPV infections. However, the combination of a negative reflex HPV DNA test and a negative cytology indicated the absence of CIN 3 or cancer to a certainty of 100%, with a specificity, positive predictive value, and negative predictive value of 93.8%, 8.6%, and 100%, respectively.[63]

Table 2 Prevalence of High-Risk HPV Infections Stratified by Age of Women at the First Examination[61]

Age	*Women (%)*	*High-Risk HPV (%)*
< 20	418 (5.3)	84 (20.1)
21–30	1,843 (23.2)	435 (23.6)
31–40	2,076 (26.2)	289 (13.9)
41–50	1,925 (24.3)	235 (12.2)
51–60	1,014 (12.8)	110 (10.8)
> 60	656 (8.3)	61 (9.3)
Total	7,932	1,214 (15.3)

HPV—human papillomavirus

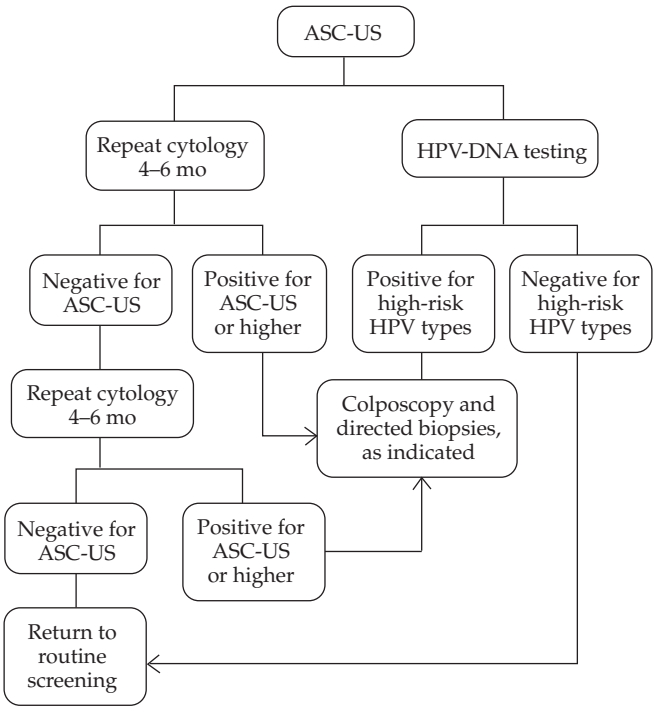


Figure 2 **Management scheme for atypical squamous cells of undetermined significance (ASC-US), based on the consensus guidelines developed by American Society for Colposcopy and Cervical Pathology.[64] (HPV—human papillomavirus)**

Recommended Management

Different policy-making organizations vary in their recommended approaches to the management of ASC. The American Society for Colposcopy and Cervical Pathology (ASCCP) established consensus guidelines for the management of ASC in 2001. The consensus guidelines classify ASC into two groups: ASC-US and ASC-H (see above).[64] For women with ASC-US cytology, a program of repeat cytologic testing, immediate colposcopy, or DNA testing for high-risk types of HPV are acceptable management options [*see Figure 2*].[64] When liquid-based cytologic is used or when cocollection for HPV DNA testing can be performed, reflex HPV DNA testing is the preferred approach. Women with ASC-US who test negative for high-risk HPV DNA should undergo repeat cytologic testing at 12 months. Women who are managed with immediate colposcopy and who are found not to have CIN should undergo repeat cytologic testing at 12 months.

When a program of repeat cervical cytologic testing is used, women with ASC-US should undergo repeat cytology (either conventional or liquid-based) at 4- to 6-month intervals until two consecutive results that are negative for intraepithelial lesion or malignancy are obtained. In most instances, women with ASC-H, LSIL, HSIL, and AGC should be referred for immediate colposcopic evaluation [*see* Squamous Epithelial Lesions, *below*].[64]

The National Comprehensive Cancer Network guidelines are consistent with the recommendation of the consensus guidelines.[65]

The Society of Obstetricians and Gynecologists of Canada (SOGC) guidelines recommend repeating the cytologic smear every 3 to 6 months until three consecutive negative smears are obtained, after which annual cytologic examinations can be resumed. If cytology continues to demonstrate ASC-US, colposcopy should be performed.[66]

Management of ASC-US in Special Circumstances

Immunosuppressed women with ASC-US are at increased risk for CIN 2,3; high-risk types of HPV are frequently detected in these women. Referral for colposcopy is recommended in all immunosuppressed women who have ASC-US.[64]

In postmenopausal women with ASC-US, the risk of CIN 2,3 is lower than in premenopausal women. Treatment with a course of intravaginal estrogen followed by a repeat cervical cytology 1 week after therapy is an acceptable option.[40] It is also acceptable to manage postmenopausal women who have ASC-US with immediate colposcopy or HPV DNA testing.

ASC-US AND ONCOGENIC HPV DNA

As demonstrated by ALTS and other studies, it is now possible to identify many women with ASC-US who do not need colposcopy. However, women who have oncogenic HPV DNA and ASC-US present a sizable management challenge. It is not known how to manage women with ASC-US who test positive for high-risk HPV DNA but who are not found to have CIN by colposcopy and biopsy. Expert opinion and review of the literature indicate that such women are at low risk for high-grade cervical neoplasia and that repeated colposcopy should not be performed in this setting. Instead, HPV DNA testing along with repeat cytology at 6 and 12 months is recommended.[27] The likelihood that these repeat tests will be negative and that patients will subsequently forgo further surveillance screening is not known.

SQUAMOUS EPITHELIAL LESIONS

For women with LSIL, colposcopy is the recommended management.[64] If the colposcopy results are negative for CIN and cancer, appropriate management entails either repeat cytology at 6 and 12 months or HPV DNA testing at 12 months; subsequent management entails (1) a repeat colposcopy if results are positive for ASC or HPV or (2) a return to routine screening if results are negative. If colposcopy reveals the presence of a lesion, the patient is managed in accordance with the guidelines recommended by the ASCCP.

For women with HSIL, colposcopy with endocervical assessment is the recommended management.[64] If colposcopy reveals no lesion or only biopsy-proven CIN 1, then cytology, colposcopy, and biopsy results should be reviewed; subsequent management depends on the final interpretation of tests. If colposcopy indicates the presence of a lesion, the patient is managed in accordance with the guidelines recommended by the ASCCP.

Management of LSIL and HSIL may vary if the patient is pregnant, postmenopausal, or an adolescent.

ATYPICAL GLANDULAR CELLS

The finding of AGC is significant because AGC is associated with a greater risk of high-grade lesions than the risk associated with ASC. On follow-up evaluation, high-grade lesions (either squamous or glandular) may be seen in 10% to 39% of patients with an AGC cytologic result[35,36]; in comparison, 5% to 17% of patients with atypical ASC cytology have CIN 2,3 confirmed by biopsy.[30,39,40]

Women with atypical endometrial cells should initially be evaluated with endometrial sampling.[64] If no neoplasia is identified, it is recommended that the patient undergo follow-up evaluation using a program of repeat cervical cytologic testing at 4- to 6-month intervals until four consecutive results that are negative for intraepithelial lesion or malignancy are obtained. If a result of ASC or LSIL is obtained on any of the follow-up smears, acceptable options include a repeat colposcopic examination.[64] Continued follow-up evaluation is needed.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Jemal A, Murray T, Ward E, et al: Cancer statistics, 2005. CA Cancer J Clin 55:10, 2005

2. United States Cancer Statistics: 2000 Incidence. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute, Atlanta, 2003

3. Jones BA, Davey DD: Quality management in gynecologic cytology using interlaboratory comparison. Arch Pathol Lab Med 124:672, 2000

4. Schiffman MH, Bauer HM, Hoover RN, et al: Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 85:958, 1993

5. Human papillomaviruses: Summaries and Evaluations, Vol 64. International Agency for Research on Cancer (IARC), Lyons, 1995
http://www.inchem.org/documents/iarc/vol64/hpv.html

6. Ho GY, Bierman R, Beardsley, et al: Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med 338:423, 1998

7. Moscicki AB, Shiboski S, Broering J, et al: The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. J Pediatr 132:277, 1998

8. Koutsky LA, Holmes KK, Critchlow CW, et al: A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. N Engl J Med 327:1272, 1992

9. Shepherd J, Weston R, Peersman G, et al: Interventions for encouraging sexual lifestyles and behaviours intended to prevent cervical cancer. Cochrane Database Syst Rev (2):CD001035, 2000

10. Frisch M, Biggar RJ, Engels EA, et al: Association of cancer with AIDS-related immunosuppression in adults. JAMA 285:3090, 2001

11. Smith JS, Bosetti C, Munoz N, et al: *Chlamydia trachomatis* and invasive cervical cancer: a pooled analysis of the IARC multicentric case-control study. Int J Cancer 111:431, 2004

12. Trimble CL, Genkinger JM, Burke AE, et al: Active and passive cigarette smoking and the risk of cervical neoplasia. Obstet Gynecol 105:174, 2005

13. Holowaty P, Miller AB, Rohan T, et al: Natural history of dysplasia of the cervix. J Natl Cancer Inst 91:252, 1999

14. U.S. Preventive Services Task Force: Screening for Cervical Cancer: Guide to Clinical Preventive Services. AHRQ Publication No. 03-515A, Agency for Healthcare Research and Quality, Rockville, MD, 2003
http://www.ahrq.gov/clinic/uspstf/uspscerv.htm

15. Cervical cancer screening in adolescents: ACOG committee opinion number 300. American College of Obstetricians and Gynecologists Committee on Adolescent Health Care. Obstet Gynecol 104:885, 2004

16. Cervical Cancer: ACS Cancer Detection Guidelines. American Cancer Society, Inc. Atlanta, 2005
http://www.cancer.org

17. Morrison BJ: Screening for cervical cancer. Canadian Task Force on the Periodic Health Examination. Canadian Guide to Clinical Preventive Health Care. Ottawa, 1994, p. 870
http://www.ctfphc.org

18. Sawaya GF, Grady D, Kerlikowske K, et al: The positive predictive value of cervical

smears in previously screened postmenopausal women. The Heart and Estrogen/Progestin Replacement Study (HERS). Ann Intern Med 133:942, 2000

19. Sawaya GF, McConnell KJ, Kulasingam SL, et al: Risk of cervical cancer associated with the interval between cervical-cancer screenings. N Engl J Med 349:1501, 2003

20. Routine cancer screening. ACOG committee opinion number 185. American College of Obstetricians and Gynecologists. Washington DC, 1997

21. Nanda K, McCrory DC, Myers ER, et al: Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 132:810, 2000

22. U.S. Preventive Services Task Force: Screening for Cervical Cancer: Guide to Clinical Preventive Services. Agency for Healthcare Research and Quality, Rockville, MD, 1996

23. McGoogan E, Reith A: Would monolayers provide more representative samples and improved preparations for cervical screening? Overview and evaluation of systems available. Acta Cytol 40:107, 1996

24. Abulafia O, Pezzullo JC, Sherer DM: Performance of Thin-Prep liquid-based cervical cytology in comparison with conventionally prepared Papanicolaou smears: a quantitative survey. Gynecol Oncol 90:137, 2003

25. Andy C, Turner LF, Neher JO: Clinical inquiries: is ThinPrep better than conventional Pap smear at detecting cervical cancer? J Fam Pract 53:313, 2004

26. ACOG practice bulletin: clinical management guidelines for obstetrician-gynecologists. Obstet Gynecol 102:417, 2003

27. Wright TC Jr, Schiffman M, Solomon D, et al: Interim guidance for the use of human papillomavirus DNA testing as an adjunct to cervical cytology for screening. Obstet Gynecol 103:304, 2004

28. Ratnam S, Franco EL, Ferenczy A: Human papillomavirus testing for primary screening of cervical cancer precursors. Cancer Epidemiol Biomarkers Prev 9:945, 2000

29. Solomon D, Davey D, Kurman R, et al: The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 287:2114, 2002

30. Malik SN, Wilkerson EJ, Drew PA, et al: Do qualifiers of ASCUS distinguish between low-and high-risk patients? Acta Cytol 43:376, 1999

31. Sherman ME, Tabbara SO, Scott DR, et al: "ASCUS, rule out HSIL": cytologic features, histologic correlates, and human papillomavirus detection. Mod Pathol 12:335, 1999

32. Park TJ, Richart RM, Sun XW, et al: Association between HPV type and clonal status of cervical squamous intraepithelial lesions (SIL). J Natl Cancer Inst 88:355, 1996

33. zur Hausen H: Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. J Natl Cancer Inst 92:690, 2001

34. Kennedy AW, Salmieri SS, Wirth SL, et al: Results of the clinical evaluation of atypical glandular cells of undetermined significance (AGCUS) detected on cervical cytology screening. Gynecol Oncol 63:14, 1996

35. Soofer S, Sidawy M: Atypical glandular cells of undetermined significance: clinically significant lesions and means of patient follow-up. Cancer Cytopathol 90:207, 2000

36. Eddy GL, Stumpf KB, Wojtowycz MA, et al: Biopsy findings in five hundred thirty one patients with atypical glandular cells of uncertain significances as defined by the Bethesda System. Am J Obstet Gynecol 177:1188, 1997

37. Kurman RJ, Henson De, Herbst AL, et al: Interim guidelines for management of abnormal cervical cytology. JAMA 271:1866, 1994

38. Davey DD, Woodhouse S, Styer P, et al: Atypical epithelial cells and specimen adequacy: current laboratory practices of participants in the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. Arch Pathol Lab Med 124:203, 2000

39. Wright TC, Lorincz AT, Ferris DG, et al: Reflex human papillomavirus deoxyribonucleic acid testing in women with abnormal Papanicolaou smears. Am J Obstet Gynecol 178:962, 1998

40. Solomon D, Schiffman M, Tarone R: Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. ALTS Study group. J Natl Cancer Inst 93:293, 2001

41. Kinney WK, Manos MM, Hurley LB, et al: Where's the high-grade cervical neoplasia? The importance of the minimally abnormal Papanicolaou diagnoses. Obstet Gynecol 91:973, 1998

42. Wright TC, Sun XW, Koulos J: Comparison of management algorithms for the evaluation of women with low-grade cytologic abnormalities. Obstet Gynecol 85:202, 1995

43. Cox JT, Lorincz AT, Schiffman MH, et al: Human papillomavirus testing by hybrid capture appears to be useful in triaging women with a cytologic diagnosis of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 172:946, 1995

44. Manos MM, Kinney WK, Hurley LB, et al: Identifying women with cervical neoplasia using human papillomavirus DNA testing for equivocal Papanicolaou results. JAMA 281:1605, 1999

45. Human papillomaviruses. IARC Monogr Eval Carcinog Risks Hum 64:277, 1995

46. Sherman M, Schiffman M, Cox JT: Effects of age and human papilloma viral load on colposcopy triage: data from the randomized Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesion Triage Study (ATLS). J Natl Cancer Inst 94:102, 2002

47. Arbyn M, Buntinx F, Van Ranst M, et al: Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst 96:250, 2004

48. Riethmuller D, Gay C, Bertrand X, et al: Genital human papillomavirus infection among women recruited for routine cervical cancer screening or for colposcopy determined by Hybrid Capture II and polymerase chain reaction. Diagn Mol Pathol 8:157, 1999

49. Cox JT, Schiffman M, Solomon D: Prospective follow-up suggests similar risk of subsequent cervical intraepithelial neoplasia grade 2 or 3 among women with cervical intraepithelial neoplasia grade 1 or negative colposcopy and directed biopsy. The ASCUS-LSIL Triage Study (ALTS) Group. Am J Obstet Gynecol 188:1406, 2003

50. Wright TC Jr: Cervical cancer screening using visualization techniques. J Natl Inst Monogr, No 31. National Cancer Institute, Atlanta, 2003, p 66

51. Spitzer M, Krumholz B, Seltzer VL, et al: Comparative utility of repeat Papanicolaou smear, cervicography, and colposcopy in the evaluation of atypical Papanicolaou smears. Obstet Gynecol 69:731, 1987

52. Eskridge C, Begneaud WP, Landweher C: Cervicography combined with repeat Papanicolaou test as triage for low-grade cytologic abnormalities. Obstet Gynecol 92:351, 1998

53. Cox JT, Shiffman MH, Winzelberg AJ, et al: An evaluation of human papillomavirus testing as part of referral to colposcopy clinics. Obstet Gynecol 80:389, 1992

54. Hatch KD, Schneider A, Abdel-Nour MW: An evaluation of human papillomavirus testing for intermediate- and high-risk types as triage before colposcopy. Am J Obstet Gynecol 172:1150, 1995

55. Bergeron C, Jeannel D, Poveda J, et al: Human papillomavirus testing in women with mild cytologic atypia. Obstet Gynecol 95:821, 2000

56. Shlay JC, Dunn T, Byers T, et al: Prediction of cervical intraepithelial neoplasia grade 2–3 using risk assessment and human papillomavirus testing in women with atypia on Papanicolaou smears. Obstet Gynecol 96:410, 2000

57. Kaufman RH, Adam E, Icenogle J, et al: Human papillomavirus testing as triage for atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesions: sensitivity, specificity, and cost-effectiveness. Am J Obstet Gynecol 177:930, 1997

58. Raab SS, Steiner AL, Hornberger J: The cost-effectiveness of treating women with a cervical vaginal smear diagnosis of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 179:411, 1998

59. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. The ASCUS-LSIL Triage Study (ALTS) Group. Am J Obstet Gynecol 188:1383, 2003

60. Cuzick J, Szarewski A, Cubie H, et al: Management of women who test positive for high-risk types of human papillomavirus: the HART study. Lancet 362:1871, 2003

61. Clavel C, Masure M, Bory JP, et al: Human papillomavirus testing in primary screening for the detection of high-grade cervical lesions: a study of 7,932 women. Br J Cancer 84:1616, 2001

62. Sherman ME, Lorincz AT, Scott DR, et al: Baseline cytology, human papillomavirus testing, and risk for cervical neoplasia: a 10-year cohort analysis. J Natl Cancer Inst 95:46, 2003

63. Petry KU, Menton S, Menton M, et al: Inclusion of HPV testing in routine cervical cancer screening for women above 29 years in Germany: results for 8466 patients. Br J Cancer 88:1570, 2003

64. Wright TC, Cox JT, Massad LS, et al: 2001 consensus guidelines for the management of women with cervical cytologic abnormalities. JAMA 287:2120, 2002

65. Cervical Cancer: Clinical Practice Guidelines in Oncology. National Comprehensive Cancer Network, 2004
http://www.nccn.org/professionals/physician_gls/PDF/cervical.pdf

66. Management of the abnormal Papanicolaou smear. SOGC Clinical Practice Guidelines, No. 70, 1998

Acknowledgment

Figure 1 Courtesy of Joan Jones, M.D.

XVIII HIRSUTISM

Robert L. Barbieri, m.d.

As an isolated clinical condition, excessive hair growth can significantly detract from a woman's quality of life. Hirsutism that occurs along with certain other clinical findings may also signal significant endocrine disease. For example, hirsutism in the presence of oligomenorrhea is often caused by polycystic ovary syndrome (PCOS) and may be associated with infertility, endometrial hyperplasia, and diabetes mellitus. Hirsutism in the presence of virilization may be caused by androgen secretion from an adrenal or ovarian tumor. Fortunately, almost all cases of hirsutism can be effectively treated with a combination of hormonal and nonhormonal therapies.

Definition

Hirsutism is the presence, in a woman, of coarse, dark terminal hair in a male pattern, often involving the upper lip, chin, sideburns, and chest. Terminal hairs are thick, stiff, and pigmented. In men, they are normally found on the face, chest, abdomen, and back. From a practice perspective, a woman who complains of hirsutism has hirsutism. Clinically, hirsutism is present if the patient's Ferriman-Gallwey score is greater than 8 [*see Figure 1*].[1,2]

Hirsutism must also be defined in the context of cultural, ethnic, and racial norms. Most Asian and Native American women have less body hair than white women; women from Mediterranean backgrounds have greater numbers of terminal hairs. In some southern European cultures, significant terminal hair on a woman's upper lip is considered normal. In other cultures, a similar amount of terminal hair might be viewed as abnormal.

Hirsutism must be distinguished from hypertrichosis, which is an increase in total body hair, including sites where terminal hair is not usually found, such as the forehead. Most cases of hypertrichosis are associated with the use of drugs such as phenytoin, penicillamine, diazoxide, minoxidil, and cyclosporine. Systemic diseases such as anorexia nervosa, malnutrition, porphyria, and hypothyroidism can also cause hypertrichosis. Treatment of hypertrichosis consists of discontinuance of the inciting drug or treatment of the systemic disease. If these approaches are not successful, treatment with an antiandrogen such as spironolactone (200 mg daily) has been reported to be moderately effective.[3]

Epidemiology

On the basis of a Ferriman-Gallwey score of more than 8, approximately 4% of women have hirsutism; only 1% of women have a Ferriman-Gallwey score above 10.[1,2] Approximately 10% of a population-based cohort of women in Finland self-reported having hirsutism.[4] In this study about 2.5 % of the women reported having both hirsutism and oligomenorrhea, indicating that in the general population, the prevalence of PCOS is lower than the prevalence of hirsutism. However, a greater proportion of women with PCOS seek medical care for their condition than women with hirsutism. Consequently, in a specialty endocrine practice, women with PCOS are overrepresented compared with women with self-reported hirsutism.

Pathogenesis

Hirsutism is caused by an excess of androgen production, androgen action, or both.[5,6] Androgen overproduction can take place in the ovary, the adrenal gland, or the skin itself (specifically, in the hair follicle). In many hirsute women, all three organs overproduce androgens.

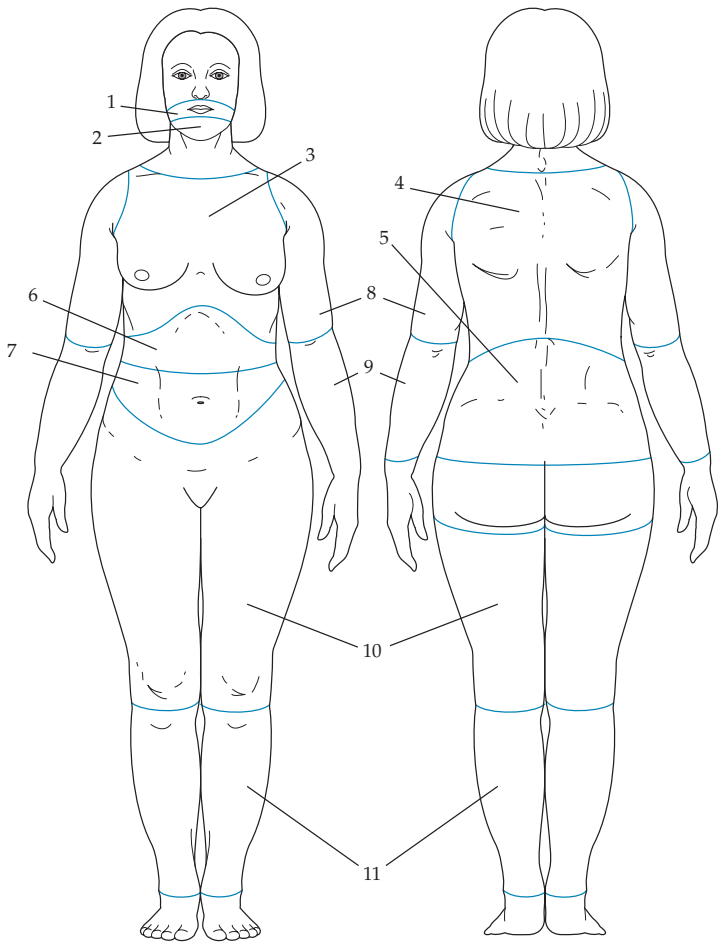
There are three types of hair: lanugo, vellus, and terminal. Lanugo is the soft, unmedullated hair seen in the fetus; it is shed in utero in the third trimester or shortly after birth. Vellus hairs are the thin, soft, unpigmented hairs that cover many areas of the body, such as the forehead. Terminal hairs are thick, coarse, and pigmented. Terminal hair is composed of an inner sheath of pigment with an outer sheath of keratin. Androgens stimulate the conversion of vellus hairs to terminal hairs in areas of the body sensitive to androgens, including the face and chest. Paradoxically, androgens cause the loss of terminal hairs in certain areas of the body, such as the frontal and parietal regions of the scalp.

The adrenal gland produces large quantities of the major androgens dehydroepiandrosterone sulfate (DHEAS), dehydroepiandrosterone (DHEA), and androstenedione. The ovary produces androstenedione and testosterone. In end organs, such as the pilosebaceous unit (hair follicle), androstenedione and testosterone can be converted to the potent androgen dihydrotestosterone (DHT) by the enzyme 5-α-reductase type 2. DHT and testosterone are the most potent androgens in humans; both bind with high affinity to the androgen receptor and initiate gene transcription that stimulates the growth of the pilosebaceous units on the face and chest and decreases the growth of pilosebaceous units on the frontal and parietal scalp. DHEA and DHEAS have no inherent androgen activity but can be metabolized to the active androgens testosterone and DHT. In research studies, nearly all hirsute women have been found to have increased production of testosterone, but standard clinical assays may not be sensitive enough to reliably detect the elevations in serum testosterone levels in these patients.[5,7]

Many women with hirsutism have several hormonal defects, including the following: (1) overproduction of adrenal androgens, (2) overproduction of ovarian androgens,[8] (3) increased conversion of androstenedione and testosterone to DHT in the hair follicle, and (4) increased sensitivity of the hair follicle to androgen action. Androgen overproduction in many hirsute women probably arises from multiple pathophysiologic defects, including (1) increased pituitary secretion of luteinizing hormone (LH), which stimulates ovarian production of testosterone and androstenedione; (2) elevated insulin levels, because of insulin resistance, which decrease the production of sex-hormone–binding globulin (SHBG) and thus increase free testosterone levels; (3) mild biochemical defects in the adrenal steroid enzymes that produce cortisol, which increase the ratio of androgen-to-cortisol secretion and result in adrenal androgen overproduction; and (4) increased activity of 5-α-reductase in the hair follicle, which increases the conversion of androgens to DHT in the follicle.

Etiology

The two most common causes of hirsutism are idiopathic hirsutism and PCOS. Idiopathic hirsutism is defined as hirsutism



Site	Grade	Definition
1. Upper Lip	1	Few hairs at outer margin
	2	Small mustache at outer margin
	3	Mustache extending halfway from outer margin
	4	Mustache extending to midline
2. Chin	1	Few scattered hairs
	2	Scattered hairs with small concentrations
	3 and 4	Complete cover, light and heavy
3. Chest	1	Circumareolar hairs
	2	With midline hair in addition
	3	Fusion of these areas, with three-quarters cover
	4	Complete cover
4. Upper Back	1	Few scattered hairs
	2	Rather more, still scattered
	3 and 4	Complete cover, light and heavy
5. Lower Back	1	Sacral tuft of hair
	2	With some lateral extension
	3	Three-quarters cover
	4	Complete cover
6. Upper Abdomen	1	Few midline hairs
	2	Rather more, still midline
	3 and 4	Half and full cover
7. Lower Abdomen	1	Few midline hairs
	2	Midline streak of hair
	3	Midline band of hair
	4	Inverted V-shaped growth
8. Arm	1	Sparse growth affecting not more than one quarter of limb surface
	2	More than this; cover still incomplete
	3 and 4	Complete cover, light and heavy
9. Forearm	1, 2, 3, and 4	Complete cover of dorsal surface; 2 grades of light and 2 of heavy growth
10. Thigh	1, 2, 3, and 4	As for arm
11. Leg	1, 2, 3, and 4	As for arm

Figure 1 **Ferriman-Gallwey system for clinical scoring of hirsutism.[1,58] Each of the 11 designated body areas is assigned a score of 0 (absence of coarse dark terminal hairs) to 4 (extensive terminal hair growth). A score higher than 8 indicates hirsutism. Hair scores over the forearms and lower leg do not contribute significantly to the distinction of hirsutism from nonhirsutism.**

(patient self-report or a Ferriman-Gallwey score above 8) in a woman with regular ovulatory menses. PCOS is the combination of oligomenorrhea (oligo-ovulation) and hyperandrogenism, as manifested by hirsutism or elevation in levels of a serum androgen such as androstenedione, testosterone, or DHEA. In many cases, idiopathic hirsutism may be a mild form of PCOS, in which androgen levels are elevated enough to cause hirsutism but not elevated enough to produce oligo-ovulation and oligomenorrhea.[9] In a very small number of women, hirsutism (usually severe hirsutism, the so-called bearded-lady phenomenon) is a manifestation of serious diseases such as adrenal or ovarian tumors. One study of 350 British women who presented with hirsutism or androgenic alopecia found that 68% had idiopathic hirsutism, 30% had PCOS, and 2% had serious underlying disease, including congenital adrenal hyperplasia, ovarian tumors, adrenal tumors, prolactinoma, and acromegaly.[10]

Controversy exists as to a possible relationship between hyperprolactinemia, excess production of adrenal androgens, and hirsutism. Several investigators have reported a relationship between hyperprolactinemia and excess production of DHEAS,[11] but others have not confirmed these observations.[12,13] In one of the most detailed studies, Schiebinger and colleagues[13] found that in women with hyperprolactinemia, successful treatment decreased the DHEAS production rate from 27 mg to 17 mg a day and increased the metabolic clearance rate from 16 L to 21 L a day. Along with these changes, serum DHEAS decreased from 2.5 μg/ml to 1.8 μg/ml. These results suggest that hyperprolactinemia may play a modest role in stimulating adrenal androgen production and may occasionally contribute to the development of hirsutism.

Diagnosis

The main goal in the evaluation of women with hirsutism is to determine whether idiopathic hirsutism or PCOS is present and to exclude rare, medically serious causes of hyperandrogenism, such as ovarian and adrenal tumors.[14] This can best be done on the basis of the history and physical examination, along with limited laboratory testing [*see Table 1*].

CLINICAL FEATURES

The differentiation between the two most common causes of hirsutism can be made by menstrual history. Women with idiopathic hirsutism have normal menstrual cycles lasting 23 to 35

Table 1 Differential Diagnosis of Hirsutism in Women of Reproductive Age

Associated Finding	*Most Likely Diagnosis*
Oligomenorrhea	Polycystic ovary syndrome (PCOS)
Regular ovulatory cycles	Idiopathic hirsutism
Fasting 17-hydroxyprogesterone level > 4 ng/ml	Nonclassic adrenal hyperplasia
Virilization	Ovarian or adrenal androgen-secreting tumor, or ovarian stromal hyperthecosis
Acanthosis nigricans	Severe insulin resistance syndrome, often associated with PCOS or ovarian stromal hyperthecosis
Galactorrhea and elevated serum prolactin level	Hyperprolactinemia

days. Women with PCOS have oligomenorrhea, with some menstrual cycles lasting longer than 35 days. Many women with PCOS have fewer than six spontaneous menstrual cycles a year.

Serious causes of hyperandrogenism (e.g., ovarian or adrenal tumors) typically manifest themselves clinically as virilization, with frontal balding, acne, clitoromegaly, increased upper body muscle mass, and deepening of the voice. Almost all women with signs of virilization are amenorrheic; none have regular ovulatory menstrual cycles. In addition, most women who present with virilization from an ovarian or adrenal tumor report recent onset (within the past year) of their hirsutism and rapid progression of their condition. In contrast, most women with idiopathic hirsutism or PCOS notice the onset of hirsutism many years before presenting for medical care, often during puberty.

When taking the history, the clinician should ask whether the patient has been using androgenic medications such as danazol or testosterone and its derivatives, which is a rare cause of hirsutism. Some female athletes may use anabolic steroids.

LABORATORY TESTING

Serum Androgen Assays

Patients with clinical evidence of hirsutism should undergo measurement of serum testosterone (total or free) and serum 17-hydroxyprogesterone levels. Unfortunately, most of the androgen assays available in clinical practice are designed to differentiate between levels found in normal women and men, and these assays are neither sensitive enough nor specific enough for detecting differences between normal and hirsute women. For example, the mean circulating total testosterone level is approximately 25 ng/dl in normal women, 50 ng/dl in hirsute women, and 600 ng/dl in men. Clinical assays are excellent at differentiating between 25 ng/dl and 600 ng/dl but are poor at differentiating between 25 ng/dl and 50 ng/dl.

Another problem with clinical assays is that testosterone circulates in both free and SHBG-bound forms. Testosterone bound to SHBG is not able to stimulate end organs, such as the pilosebaceous unit. The amount of free testosterone depends on both testosterone production rates and the concentration of SHBG. To complicate the situation, an increase in testosterone production causes a decrease in SHBG production, which results in a decrease in total testosterone (the form of testosterone that is most often measured in practice) and an increase in free testosterone (a form that is not typically measured in clinical practice).

In most clinical assays, the upper limit of normal for serum testosterone is 80 ng/dl. Almost all women with virilization have serum testosterone levels greater than 150 ng/dl.[15-17] It is probably wise for primary care physicians to seek the help of an endocrinologist if they identify a virilized woman or a woman with a serum testosterone concentration above 150 ng/dl.

Other Tests

In a hirsute woman who has amenorrhea, the serum prolactin level should be measured. Women with PCOS, especially those who are obese, should be screened for diabetes mellitus. This is particularly true of patients in whom physical examination reveals acanthosis nigricans, which suggests significant insulin resistance.

Between 1% and 5% of women who present with hirsutism and oligomenorrhea have nonclassic congenital adrenal hyperplasia (NCAH) from 21-hydroxylase deficiency that was too mild to be detected at birth.[18] All of these women have elevated levels of 17-hydroxyprogesterone (above 4 ng/ml) in the early morning (8 A.M.) during the follicular phase of the menstrual cycle. Consequently, measurement of a follicular-phase 8 A.M. 17-hydroxyprogesterone level can help rule out NCAH. It is not clear whether all women with hirsutism and oligomenorrhea should be screened for NCAH, because the disorder has a low prevalence in many populations. In addition, many women with NCAH respond well to the standard treatments for PCOS. Screening might be warranted in women of ethnic groups that have an increased prevalence of NCAH (e.g., Ashkenazi Jews, Inuits). One middle-of-the-road approach is to screen women for NCAH only if their serum testosterone concentration is above a clearly elevated level, such as 100 ng/dl or 150 ng/dl. This approach will detect the most clinically significant cases of NCAH.

Treatment

The goals of treatment of hirsutism include the following: (1) rule out a serious underlying medical condition such as a virilizing tumor of the adrenal or ovary, (2) slow or stop new hair growth, (3) remove or hide existing hair, (4) evaluate and treat associated hormonal problems such as oligomenorrhea, and (5) anticipate long-term health conditions that can occur in women with hyperandrogenism, such as an increased risk of diabetes.

Hormonal treatment of hirsutism typically does not result in normalization of the Ferriman-Gallwey score (i.e., to below 8). In many studies, patients' Ferriman-Gallwey scores are in the range of 20 to 25 at the initiation of treatment. After 12 months of treatment, the Ferriman-Gallwey scores are in the range of 10 to 15. On the basis of these general observations, clinicians can assure women that treatment is effective but should warn them that treatment is unlikely to reduce facial hair to a level similar to that in nonhirsute women. In addition, clinicians should advise women that it will take at least 6 months for treatment to reach near maximal effects. Finally, most patients report that if they discontinue hormone treatment, the hirsutism recurs over the next 3 to 9 months.

Treatment of hirsutism should be multimodal [*see Table 2*]. It should include hormonal suppression of androgen production

or action, along with a nonhormonal method of controlling hair growth. A common strategy is to combine an estrogen-progestin contraceptive formulation with an antiandrogen and to use electrolysis or shaving as a nonhormonal adjuvant. An alternative hormonal approach is treatment with an antiandrogen alone. Alternative nonhormonal treatments include the use of a facial antihair cream or laser destruction of hair follicles [*see* Nonhormonal Approaches, *below*].

WEIGHT LOSS

Body mass index (BMI) is a major determinant of insulin resistance and hyperinsulinemia. Women with a BMI greater than 27 kg/m^2 are often insulin resistant.[19] Women with a BMI under 25 kg/m^2 are seldom insulin resistant, unless they have a relatively rare genetic cause of insulin resistance, such as lipodystrophy. Hyperinsulinemia causes hyperandrogenism through several mechanisms. In combination with LH, elevated insulin levels cause the ovarian theca and stroma to produce excess quantities of androstenedione and testosterone.[20] In addition, hyperinsulinemia suppresses hepatic production of SHBG, which causes an increase in free androgens. Excess ovarian production of androgens and decreased SHBG work synergistically to increase the risk of development of hirsutism.

Numerous studies have demonstrated the benefits of weight loss in hyperandrogenic, insulin-resistant women.[21-25] In these studies, mean weight loss ranging from about 10 to 20 kg has been associated with decreases in insulin levels and testosterone concentration and with ovulation and pregnancy in many women.

Weight loss is difficult to achieve. A structured program that includes consultation with a nutritionist, encouragement by the physician, a low-calorie diet, and initiation of an exercise program may be the most effective nonsurgical approach in these patients. Surgical methods of weight reduction can be very effective, especially in women whose BMI is greater than 40 [*see 3:X Obesity*].

ESTROGEN-PROGESTIN CONTRACEPTIVES

Combination estrogen-progestin formulations at doses used for contraception (in the form of pills, transdermal patches, or vaginal rings) suppress LH secretion, which in turn suppresses ovarian androgen production.[26,27] In addition, the estrogen in estrogen-progestin combinations increases liver production of SHBG, which binds testosterone and decreases circulating bioavailable testosterone, reducing androgen stimulation of the pilosebaceous units.[28] A further advantage for women with PCOS is that estrogen-progestin combinations, when given cyclically, induce regular withdrawal bleeding.

In a study that examined the effects of 5 years of estrogen-progestin contraceptive treatment for hirsutism, the investigators reported that hirsutism continued to improve through all 5 years of the treatment. Acne improved significantly by 2 years of treatment, but minimal further improvement occurred between 2 and 5 years. Within 6 months of discontinuance of treatment, hirsutism and acne reappeared.[29]

The effectiveness of estrogen-progestin contraceptives may be influenced by body mass. In heavier women, it may be best to select a medication with 35 μg of ethinyl estradiol.

Some estrogen-progestin contraceptive formulations may differ with regard to their effect on hepatic production of SHBG. Norgestrel is an androgenic progestin (derived from 19-nortestosterone) that partially blocks estrogen-induced SHBG production. Contraceptives that contain androgenic progestins, such as norgestrel, do not result in as great an increase in SHBG production as contraceptives that contain less androgenic progestins, such as drospirenone, desogestrel, ethynodiol diacetate, and norgestimate [*see Table 3*]. However, for the treatment of hirsutism, any estrogen-progestin contraceptive is superior to no estrogen-progestin contraceptive.

Table 2 Treatment of Hirsutism

Category	*Treatment*	*Potential Toxic Effects*
Hormonal	Combination estrogen-progestin contraceptive	Deep vein thrombosis
	Spironolactone, 50–200 mg daily; typical dosage, 100 mg daily	Hyperkalemia in patients with renal disease
	Combination estrogen-progestin contraceptive plus spironolactone, 100 mg daily	—
	Flutamide, 62.5–500 mg daily; typical dosage, 125 mg daily	Liver toxicity
	Finasteride, 2.5–5 mg daily; typical dosage, 5 mg daily	Teratogenicity in male fetuses
	Metformin, 1,500–2,250 mg daily; typical dosage, 2,000 mg daily	Nausea, vomiting; fatal lactic acidosis in patients with renal disease, heart failure, or liver disease
	Gonadotropin-releasing hormone analogue	Hypoestrogenism, including vasomotor symptoms and bone loss, when used as a single agent
Nonhormonal	Shaving; depilatories; bleaching; laser photothermolysis; nonlaser, rapid pulse, intense light therapy; eflornithine cream; electrolysis	—

Note: treatment is best accomplished with the combination of a hormonal treatment and a nonhormonal treatment.

ANTIANDROGENS

Antiandrogens are a cornerstone of therapy for hirsutism. Used alone, antiandrogens appear to be as effective as oral-contraceptive treatment.[30] As with estrogen-progestin treatment, cessation of antiandrogen therapy is followed by recurrence of hirsutism within 1 year in the majority of women.[31]

The antiandrogens with demonstrated efficacy against hirsutism include spironolactone, finasteride, flutamide, and ketoconazole [*see Table 2*].[32-38] In one high-quality clinical trial, 40 hirsute women were randomized to receive spironolactone (100 mg daily), flutamide (250 mg daily), finasteride (5 mg daily), or placebo. Placebo did not produce a decrease in hirsutism, but all three active treatments were associated with a significant reduction in hirsutism scores. The reductions in the Ferriman-Gallwey score observed after 6 months of treatment were 41% with spironolactone, 39% with flutamide, and 32% with finasteride.[39] Similarly, in a placebo-controlled trial of finasteride (5 mg daily), flutamide (500 mg daily), and estrogen-progestin, the three active treatments resulted in similar decreases in Ferriman-Gallwey scores.[40] A randomized trial that compared flu-

tamide (250 mg daily), finasteride (5 mg daily), ketoconazole (300 mg daily), and estrogen-progestin in 66 hirsute women found that after 1 year of treatment, all three agents significantly decreased the Ferriman-Gallwey score. The magnitude of the reduction in the Ferriman-Gallwey score was 55% with flutamide, 44% with finasteride, 53% with ketoconazole, and 60% with estrogen-progestin; the differences between the agents were not statistically significant.[41]

Individual Antiandrogens

Spironolactone The most widely used antiandrogen for the treatment of hirsutism, spironolactone inhibits the binding of testosterone and DHT to the androgen receptor. Because spironolactone and its active metabolites have very long half-lives, the entire dose can be given once daily. The usual dosage of spironolactone is between 50 and 200 mg daily. Many authorities recommend a dosage of 100 mg daily because it is near the top of the dose-response curve and is associated with fewer side effects than 200 mg daily. Approximately 70% of women treated with spironolactone have reported improvement. In one study of spironolactone (200 mg daily) for the treatment of moderate to severe hirsutism, 19 of 20 women reported improvement, with beneficial effects noticeable within 2 months, peaking at 6 months, and maintained through 12 months of therapy.[42] Because spironolactone may be a teratogen, sexually active patients who are not taking spironolactone along with an estrogen-progestin contraceptive should use barrier contraception.

Finasteride Finasteride inhibits type 1 (skin) and type 2 (prostate) 5-α-reductase, the enzyme that transforms testosterone to the most potent androgen, DHT. The Food and Drug Administration has placed a so-called black-box warning on finasteride, advising that no woman of childbearing age should take the medication or touch a broken pill because of the potential teratogenic effects of an antiandrogen on a male fetus. Consequently, a physician who prescribes finasteride to a hirsute woman may find that the pharmacist refuses to dispense the medication or that the patient's insurance company urges the patient to stop taking the medicine. These practical considerations argue against prescribing finasteride for the treatment of hirsutism.

Table 3 Androgenicity of Progestins Used in Oral Contraceptives

Level of Androgenicity	*Progestin*	*Brand Name of Products*
High	Norgestrel	LoOvral
	Levonorgestrel	Alesse, Levlite, Nordette, Trilevlen, Tri-Phasil
Moderate	Norethindrone	Genora 1/35, Norinyl 1/35, OrthoNovum 1/35, Ovcon 35
	Norethindrone acetate	Loestrin 1/20 Loestrin 1.5/35
Low	Ethynodiol diacetate	Demulen 1/35
	Norgestimate	Ortho-TriCyclen Ortho-TriCyclen Lo
	Desogestrel	Desogen Ortho-Cept
	Drosperinone	Yasmin

Flutamide An androgen receptor antagonist, flutamide is widely used in Europe for the treatment of hirsutism. It is effective at dosages ranging from 67.5 to 500 mg daily. Flutamide is probably a more potent antiandrogen than spironolactone. However, flutamide is much more expensive than spironolactone and can be associated with fatal hepatotoxicity.[43]

Ketoconazole Ketoconazole is an antifungal agent that inhibits cytochrome P-450 enzymes, which are involved in the synthesis of androgens. Treatment with ketoconazole (300 mg daily) inhibits androgen production and is as effective as finasteride, spironolactone, and flutamide in the treatment of hirsutism. However, ketoconazole has more side effects than the other antiandrogens that are commonly used to treat hirsutism.[41] The most commonly reported side effects are nausea, alopecia, headache, and elevated liver enzyme levels. For that reason, I do not recommend the use of ketoconazole to treat hirsutism.

ESTROGEN-PROGESTIN PLUS AN ANTIANDROGEN

The combination of an oral contraceptive and an antiandrogen may be somewhat more effective against hirsutism than either agent used alone. One trial studied the use of an oral contraceptive plus either finasteride (5 mg daily) or placebo for the treatment of hirsutism. After 1 year of treatment, the women who received combination therapy had greater reductions in their Ferriman-Gallwey score than women who received the oral contraceptive alone (48% versus 38%, respectively; $P < 0.05$).[44] On the other hand, a trial comparing an oral contraceptive plus flutamide (250 mg daily) with flutamide alone found the two treatments to be equally effective.[45]

One disadvantage of combined therapy is that it is more expensive than single-agent therapy. Also, the side effects may be additive.

GONADOTROPIN-RELEASING HORMONE ANALOGUES

Long-term administration of a gonadotropin-releasing hormone (GnRH) agonist analogue suppresses pituitary secretion of LH and ovarian secretion of testosterone. Treatment with a GnRH analogue has been demonstrated to reduce Ferriman-Gallwey scores. In head-to-head studies, however, antiandrogens have produced greater reductions in Ferriman-Gallwey scores than GnRH agonist treatment.[46] GnRH agonists are also generally more expensive than antiandrogens, limiting the clinical utility of GnRH agonists as first-line agents in the treatment of hirsutism.

METFORMIN

Many women with hirsutism have insulin resistance, and abnormally elevated concentrations of circulating insulin appear to stimulate ovarian androgen production. Metformin improves insulin dynamics, decreases circulating insulin levels, reduces serum testosterone levels, and reduces the rate of abnormal terminal hair growth. Clinical trials suggest that metformin at a dosage of 1,500 mg daily is effective in the treatment of hirsutism. For example, one randomized study that compared metformin with a cyclic oral contraceptive in 52 women with PCOS and hirsutism found that Ferriman-Gallwey scores after 12 months of treatment were significantly re-

duced in patients taking metformin (25% decrease; $P < 0.01$) but not in those taking oral contraceptives (5% decrease; P = not significant).[47] Mean hair diameter was significantly reduced with both treatments. Similar results have been reported by other investigators.[48]

THIAZOLIDINEDIONES

Like metformin, the thiazolidinediones increase insulin sensitivity. Neither pioglitazone nor rosiglitazone has been extensively studied for its efficacy against hirsutism, however. Before being withdrawn from the market because of hepatic toxicity, troglitazone was demonstrated to be effective in the treatment of hyperandrogenism and hirsutism.[49]

GLUCOCORTICOIDS

Suppression of adrenal gland androgen production with glucocorticoids is effective in the treatment of hirsutism, but long-term glucocorticoid treatment can lead to such serious side effects as weight gain, osteopenia, impaired glucose metabolism, and suppression of adrenal function. Consequently, most endocrinologists do not recommend the long-term use of glucocorticoids for the treatment of hirsutism.

NONHORMONAL APPROACHES

Eflornithine Cream

Topical eflornithine (α-difluoromethylornithine) retards hair growth through mechanisms that are not fully characterized. Eflornithine inhibits ornithine decarboxylase and is active as an anthelmintic agent. Eflornithine is not a depilatory agent; rather, it retards hair growth. The FDA has approved eflornithine for the treatment of hirsutism, and it is available as a 13.9% cream. The standard treatment regimen is application of a thin film to the upper lip and other affected areas of the face twice daily. The applications should be at least 8 hours apart. Patients should not wash the treated areas for 4 hours. To minimize inflammatory side effects, eflornithine should not be applied until at least 5 minutes after plucking, tweezing, cutting, or shaving of hair. In one study, after 24 weeks, approximately 32% of women treated with eflornithine reported marked improvement in facial hirsutism, compared with 8% of women treated with a placebo cream.[50] Improvement is seen within 4 to 8 weeks after starting treatment. Hirsutism may return to pretreatment levels 12 weeks after discontinuing the medication.

The major side effects of eflornithine are stinging and burning at the site of application and inflammation, with reddening of the irritated skin. Less than 1% of the applied dose is absorbed through the skin.

Laser Treatment

Laser therapy can induce selective photothermolysis of the pilosebaceous unit. A brief laser pulse at a wavelength that will be absorbed by the melanin in darkly pigmented hair, but not by lightly pigmented skin, is used to selectively heat the pilosebaceous unit. The use of brief pulses of laser energy minimizes the lateral diffusion of the thermal effect, resulting in maximal delivery of energy to the target and minimal effects in the adjacent skin. Three to six treatments performed every 8 weeks are necessary to control hair growth. Laser hair removal is most successful in patients with lighter skin color and dark hairs. Lasers with ruby and alexandrite generators appear to be most widely utilized.[51] Patients who have undergone laser treatment of hirsutism report a high degree of satisfaction with the therapy and its results, although many women have recurrence of hirsutism 6 to 12 months after the treatment.[52]

Intense Pulsed Light

An alternative to laser treatment for hirsutism is the use of intense pulsed light that is not coherent (nonlaser). In one form of this treatment, thermal energy is created in hair follicles through the use of broad-spectrum noncoherent radiation with wavelengths from 550 to 1,200 nm administered in macropulses divided into three minipulses. The thermal energy destroys the epithelial cells in the hair follicle and thereby reduces hair growth.[53] In another form of this treatment, nonlaser light with wavelengths of approximately 550 to 585 nm is delivered in energy pulses of 38.7 joules/cm.[54] In small series, many of which had no control group, intense pulsed light has been demonstrated to be effective in reducing facial hirsutism.[55]

Hair Removal and Lightening

Shaving is a safe and useful method for hair removal but may need to be done multiple times each week; depilatory agents and wax can remove hair; and bleaches can lighten hair. These treatments, however, can cause skin irritation and erythema, which is especially troublesome when it occurs on the face. Electrolysis, which destroys the pilosebaceous unit with electrical energy one hair follicle at a time, is effective but labor intensive and expensive. In contrast, laser therapy removes hundreds of hairs simultaneously.

EXPERIMENTAL TREATMENT

Finasteride cream, 0.25%, applied to facial areas has been found superior to placebo in hirsute women. In a small study, administration of finasteride cream for 6 months reduced hair counts by approximately 40%, whereas placebo had no beneficial effects.[56] The FDA has not approved finasteride cream; a compounding pharmacy would need to prepare the medication.

Prognosis

Women with hirsutism should be warned not to expect significant improvement for at least 3 to 6 months after starting hormone therapy. Hair follicles have a life cycle of up to 6 months, and the actively growing hair must reach the end of its life cycle before it will cease to grow. The first response that a patient may observe is that hair is thinner (has a smaller diameter) and lighter in color. This occurs because the pigmented core of the hair shrinks in diameter before the linear growth rate of the hair slows.

Therapy for hirsutism is usually continued indefinitely because increased androgen production or sensitivity may persist for decades. In one study, hormone therapy for 2 years led to a marked decrease in hirsutism, but 80% of patients noted recurrence of hirsutism within 6 months after discontinuing therapy.[57] In most women, excess androgen production begins to decline in the late portion of the fourth decade of life.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

None of the oral contraceptives or hormonal agents discussed in this chapter has been approved by the FDA for the specific treatment of hirsutism.

References

1. Ferriman D, Gallwey JD: Clinical assessment of body hair growth in women. J Clin Endocrinol Metab 21:1440, 1961

2. Knochenhauer ES, Key TJ, Kahsar-Miller M, et al: Prevalence of the polycystic ovary syndrome in unselected black and white women of the southeastern United States. J Clin Endocrinol Metab 37:615, 1998

3. Darendeliler F, Bas F, Balaban S, et al: Spironolactone therapy in hypertrichosis. Eur J Endocrinol 135:604, 1996

4. Taponen S, Martikainen H, Jarvelin MR, et al: Hormonal profile of women with self-reported symptoms of oligomenorrhea and/or hirsutism: Northern Finland Birth Cohort 1966 study. J Clin Endocrinol Metab 88:141, 2003

5. Kirschner MA, Samojlik E, Silber D: A comparison of androgen production and clearance in hirsute and obese women. J Steroid Biochem 19:607, 1983

6. Matteri RK, Stanczyk FZ, Gentzschein EE, et al: Androgen sulfate and glucuronide conjugates in nonhirsute and hirsute women with the polycystic ovary syndrome. Am J Obstet Gynecol 167:1807, 1992

7. Samojilk E, Kirschner MA, Silder D, et al: Elevated production and metabolic clearance rates of androgens in morbidly obese women. J Clin Endocrinol Metab 59:949, 1984

8. Escobar-Morreale HF, Serrano-Gotarredona J, Garcia-Robles R, et al: Mild adrenal and ovarian steroidogenic abnormalities in hirsute women without hyperandrogenemia: does idiopathic hirsutism exist? Metabolism 46:902, 1997

9. Carmina E, Lobo RA: Polycystic ovaries in hirsute women with normal menses. Am J Med 111:602, 2001

10. Higuchi K, et al: Prolactin has a direct effect on adrenal androgen secretion. J Clin Endocrinol Metab 59:714, 1984

11. Belisle S, Menard J: Adrenal androgen production in hyperprolactinemic states. Fertil Steril 33:396, 1980

12. Parker LN, Chang S, Odell WD: Adrenal androgens in patients with chronic marked elevation of prolactin. Clin Endocrinol 8:1, 1978

13. Schiebinger RJ, Chrousos GP, Cutler GB Jr, et al: The effect of serum prolactin on plasma adrenal androgens and the production and metabolic clearance rate of dehydroepiandrosterone sulfate in normal and hyperprolactinemic subjects. J Clin Endocrinol Metab 62:202, 1986

14. O'Driscoll JB, Mamtora H, Higginson J, et al: A prospective study of the prevalence of clear cut endocrine disorders and polycystic ovaries in 350 women presenting with hirsutism or androgenic alopecia. Clin Endocrinol 41:231, 1994

15. Friedman CI, Schmidt GE, Kim MH, et al: Serum testosterone concentrations in the evaluation of androgen-producing tumors. Am J Obstet Gynecol 153:44, 1985

16. Meldrum DR, Abraham GE: Peripheral and ovarian venous concentrations of various steroid hormones in virilizing ovarian tumors. Obstet Gynecol 53:36, 1979

17. Derksen J, Nagesser SK, Meidners AE, et al: Identification of virilizing adrenal tumors in hirsute women. N Engl J Med 331:968, 1994

18. Chetkowski RJ, DeFazio J, Shamonki I, et al: The incidence of late-onset congenital adrenal hyperplasia due to 21-hydroxylase deficiency among hirsute women. J Clin Endocrinol Metab 58:595, 1984

19. Weyer C, Bogardus C, Mott DM, et al: The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. J Clin Invest 104:787, 1999

20. Barbieri RL, Smith S, Ryan KJ: The role of hyperinsulinemia in the pathogenesis of ovarian hyperandrogenism. Fertil Steril 50:197, 1988

21. Bates GW, Whitworth NS: Effect of body weight reduction on plasma androgens in obese infertile women. Fertil Steril 38:406, 1982

22. Pasquali R, Antenucci D, Casimirri F, et al: Clinical and hormonal characteristics of obese and amenorrheic women before and after weight loss. J Clin Endocrinol Metab 8:173, 1989

23. Clark AM, Thornley B, Tomlinson L, et al: Weight loss in obese infertile women results in improvements in reproductive outcome for all forms of fertility treatment. Hum Reprod 13:1502, 1998

24. Hollman M, Runnebaum B, Gerhard I: Effects of weight loss on the hormonal profile in obese infertile women. Hum Reprod 11:1884, 1996

25. Clark AM, Ledger W, Galletly C, et al: Weight loss results in significant improvement in pregnancy and ovulation rates in anovulatory obese women. Hum Reprod 10:2705, 1995

26. Givens JR, Andersen RN, Wiser WL: The effectiveness of two oral contraceptives in suppressing plasma androstenedione, testosterone, LH and FSH and in stimulating plasma testosterone binding capacity in hirsute women. Am J Obstet Gynecol 124:333, 1976

27. Givens JR, Andersen RN, Wiser WL, et al: Dynamics of suppression and recovery of plasma FSH, LH, androstenedione and testosterone in polycystic ovary disease using an oral contraceptive. J Clin Endocrinol Metab 38:727, 1974

28. Kullberg G, Hamburger L, Mattson LA, et al: Effects of a low-dose desogestrel-ethinyl estuarial combination on hirsutism, androgens and sex hormone binding globulin in women with polycystic ovary syndrome. Act Obstet Gynecol Scand 64:195, 1985

29. Falsetti L, Gambera A, Tisi G: Efficacy of the combination ethinyl estradiol and cyproterone acetate on endocrine, clinical and ultrasonographic profile in the polycystic ovary syndrome. Hum Reprod 16:36, 2001

30. Spritzer PM, Lisboa KO, Mattiello S, et al: Spironolactone as a single agent for long-term therapy of hirsute patients. Clin Endocrinol 52:587, 2000

31. Yucelten D, Erenus M, Gurbuz O, et al: Recurrence rate of hirsutism after 3 different antiandrogen therapies. J Am Acad Dermatol 41:64, 1999

32. Bayram F, Muderris II, Guven M, et al: Comparison of high-dose flutamide versus low-dose flutamide in the treatment of hirsutism. Eur J Endocrinol 147:467, 2002

33. Falsetti L, Gambera A, Lengrenzi L, et al: Comparison of finasteride versus flutamide in the treatment of hirsutism. Eur J Endocrinol 141:361, 1999

34. Venturoli S, Paradisi R, Banoli A, et al: Low-dose flutamide (125 mg/day) as maintenance therapy in the treatment of hirsutism. Horm Res 56:25, 2001

35. Farquhar C, Lee O, Toomath R, et al: Spironolactone versus placebo or in combination with steroids for hirsutism and/or acne. Cochrane Database Syst Rev (4):CD000194, 2003

36. Muderris II, Bayram F, Guven M: Treatment of hirsutism with lowest dose flutamide (62.5 mg daily). Gynecol Endocrinol 14:38, 2000

37. Petrone A, Civitillo RM, Galante L, et al: Usefulness of a 12-month treatment with finasteride in idiopathic and polycystic ovary syndrome associated hirsutism. Clin Exp Obstet Gynecol 26:213, 1999

38. Barth JH, Cherry CA, Wojnarowska F, et al: Spironolactone is an effective and well-tolerated systematic antiandrogen therapy for hirsute women. J Clin Endocrinol Metab 68:966, 1989

39. Moghetti P, Tosi F, Tosti A, et al: Comparison of spironolactone, flutamide and finasteride efficacy in the treatment of hirsutism: a randomized, double blind, placebo controlled trial. J Clin Endocrinol Metab 85:89, 2000

40. Fruzzetti F, Bersi C, Parrini D, et al: Treatment of hirsutism: comparisons between different antiandrogens with central and peripheral effects. Fertil Steril 71:445, 1999

41. Venturoli S, Marescalchi O, Colombo FM, et al: A prospective randomized trial comparing low dose flutamide, finasteride, ketoconazole and cyproterone acetate-estrogen regimens in the treatment of hirsutism. J Clin Endocrinol Metab 84:1304, 1999

42. Cumming DC, Yang JC, Rebar RW, et al: Treatment of hirsutism with spironolactone. JAMA 247:1295, 1982

43. Wysowski DK, Freiman JP, Touretlot JB, et al: Fatal and nonfatal hepatotoxicity associated with flutamide. Ann Intern Med 119:1150, 1993

44. Sahin Y, Dilber S, Keletimur F: Comparison of Diane 35 and Diane 35 plus finasteride in the treatment of hirsutism. Fertil Steril 75:496, 2001

45. Taner C, Inal M, Basogul O, et al: Comparison of the clinical efficacy and safety of flutamide versus flutamide plus an oral contraceptive in the treatment of hirsutism. Gynecol Obstet Invest 54:105, 2002

46. Pazos F, Escobar-Morreale HF, Balsa J, et al: Prospective randomized study comparing the long-acting gonadotropin-releasing hormone agonist triptorelin, flutamide and cyproterone acetate used in combination with an oral contraceptive in the treatment of hirsutism. Fertil Steril 71:122, 1999

47. Harborne L, Fleming R, Lyall H, et al: Metformin or antiandrogen in the treatment of hirsutism in polycystic ovary syndrome. J Clin Endocrinol Metab 88:4116, 2003

48. Kelly CJ, Gordon D: The effect of metformin on hirsutism in polycystic ovary syndrome. Eur J Endocrinol 147:217, 2002

49. Azziz R, Ehrmann D, Legro RS, et al: Troglitazone improves ovulation and hirsutism in the polycystic ovary syndrome. J Clin Endocrinol Metab 86:1626, 2001

50. Balfour JA, McClellan K: Topical eflornithine. Am J Clin Dermatol 2:197, 2001

51. Sanchez LA, Perez M, Azziz R: Laser hair reduction in the hirsute patient: a critical assessment. Hum Reprod Update 8:169, 2002

52. Loo WJ, Lanigan SW: Laser treatment improves the quality of life of hirsute females. Clin Exp Dermatol 27:439, 2002

53. Sadick NS, Weiss RA, Shea CR, et al: Long-term photoepilation using a broad-spectrum intense pulsed light source. Arch Dermatol 136:1336, 2000

54. Schroeter CA, Raulin C, Thurllimann W, et al: Hair removal in 40 hirsute women with an intense laser-like light source. Eur J Dermatol 9:374, 1999

55. Moreno-Arias GA, Castelo-Branco C, Ferrando J: Side-effects after IPL photodepilation. Dermatol Surg 28:1131, 2002

56. Lucas KJ: Finasteride cream in hirsutism. Endocr Pract 7:5, 2001

57. Kokaly W, McKenna TJ: Relapse of hirsutism following long-term successful treatment with estrogen-progestogen combination. Clin Endocrinol 52:379, 2000

58. Hatch R, Rosenfield RL, Kim MH, et al: Hirsutism: implications, etiology and management. Am J Obstet Gynecol 140:815, 1981

Acknowledgment

Figure 1 Seward Hung.

XIX CARDIOVASCULAR DISEASE IN WOMEN

JoAnne Micale Foody, m.d.

Heart disease is the leading cause of death in women in the United States: in 2002, almost half a million women died of heart disease, which is more than 10 times as many as died of breast cancer.[1] Indeed, almost one out of every two women in the United States will die of some cardiovascular event—most likely, myocardial infarction (MI), hypertensive heart disease, or stroke.[1] In addition, heart failure is a growing epidemic in women, particularly older women.

Despite the importance of cardiovascular disease in women, it has been viewed as a disease of middle-aged men, and most knowledge of it comes from studies in middle-aged men.[2] Yet there is increasing evidence that significant gender differences do exist in cardiovascular risk factors and in their modification in women. The clinical presentation of heart disease,[2,3] the impact of individual cardiovascular risk factors,[1,4-7] and the results of risk-factor interventions differ dramatically by gender.[8-17] Moreover, although the mortality associated with coronary artery disease (CAD) in men has shown major improvements, mortality in women continues to increase [*see Figure 1*]. This chapter reviews those uniquely female attributes that are associated with differences in cardiovascular disease presentation and discusses ways to reduce risk and improve clinical outcomes.

Coronary Artery Disease

EPIDEMIOLOGY

Coronary artery disease is the leading cause of mortality in women in the United States. One third of women develop CAD, and one in three die of the disease—in 2002, a toll of almost 242,000 deaths.[1,18] CAD is not limited to older women; a significant number of younger women also develop CAD.[19,20]

RISK FACTORS

The increasing rates of CAD are not surprising, in view of the fact that coronary risk factors are highly prevalent and are increasing in women in the United States. Of women 20 to 74 years of age, a third have hypertension, more than 25% have hyperlipidemia, more than 25% smoke cigarettes, more than 25% are overweight, and more than 25% report having sedentary lifestyles.[1] These risk factors are more prevalent in men than in women, but as women grow older, their risk-factor profile approaches, and, in some instances, surpasses that of their male counterparts.

Although multiple clinical trials have identified important risk factors and effective therapies for CAD, few of these studies have included sufficient numbers of women to draw meaningful conclusions.[21] Thus, much of the evidence that supports contemporary recommendations for testing, prevention, and treatment of CAD in women is extrapolated from studies conducted predominantly in middle-aged men. Applying the findings of studies in men to management of CAD in women may not be appropriate, because the symptoms of CAD, its natural history, and the response to therapy may differ in men and women.

Dyslipidemia

It is well established that the higher the level of serum cholesterol, the higher the risk of CAD in both men and women.[22,23] In women, however, lipid and lipoprotein concentrations vary with ovarian function; consequently, CAD risk varies as ovarian function changes.

After 20 years of age, the plasma total cholesterol level increases progressively in both sexes; however, until the sixth decade, the rate of increase is slower in women. After the sixth decade, women's total-cholesterol levels generally increase faster, exceeding those of men of similar age. Although elevations in total cholesterol are significantly associated with CAD in women, women have lower absolute rates of disease than men for any given cholesterol level. For example, the Framingham investigators found that with total-cholesterol levels above 295 mg/dl, the rate of MI in women was 60% that in men.[5]

Similarly, high-density lipoprotein (HDL) cholesterol levels are different between men and women. Until puberty, HDL levels are slightly higher in boys than in girls, but as girls pass through puberty, HDL levels do not fall as they do in boys of the same age. After puberty, HDL levels in women increase slowly until before menopause, whereas they remain constant in men. Thus, average HDL values in women are approximately 20% higher than those in men.[24] HDL levels decline by approximately 3.5 mg/dl after menopause, but they generally remain higher in women than men throughout life.[25]

Elevated HDL levels protect both men and women against the development of CAD, but the protective effect appears to be more powerful in women. In the Framingham Study, every 10 mg/dl increase in HDL was associated with a 40% to 50% reduction in coronary risk; the Framingham investigators suggest that the protective effect of HDL in women is approximately twice the atherogenic effect of low-density lipoprotein (LDL).[26-29] Both the Framingham Study and the Donolo–Tel Aviv Study have documented increased CAD incidence and mortality in women whose HDL levels are greater than 35 mg/dl but less than 50 mg/dl.[30,31]

LDL levels are lower in women than in men throughout the

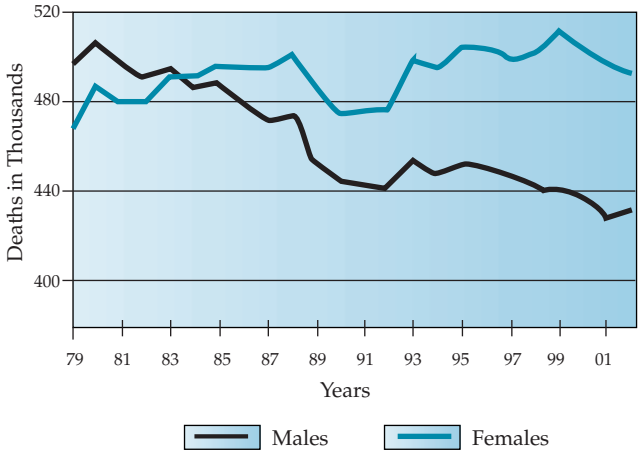


Figure 1 **Cardiovascular disease mortality trends for males and females, United States 1979–2002.[111]**

first two thirds of the life span. After the middle of the fifth decade, as a result of menopause, LDL levels in women rise abruptly, whereas LDL levels in men remain stable. In the last third of life, women generally have higher LDL levels.[5] Although an LDL level of 130 mg/dl or higher is considered a risk factor for CAD in both men and women, independent of the HDL level, the role of LDL as a risk factor in women is somewhat controversial. Few prospective studies have examined LDL and CAD risk in women. Very high LDL concentrations carry the same poor prognosis in both sexes, but mild elevations of LDL are not associated with the same risk burden in women as in men, perhaps because of women's generally higher levels of HDL.

Little difference in triglyceride concentrations is noted between sexes until puberty. Then, triglyceride concentrations increase in both sexes with increasing age, albeit at a much slower rate in women. Triglyceride levels in men actually decrease in middle age, whereas they continue to increase gradually in women. By the age of 70 years, mean levels in women equal those of men.

Triglyceride levels appear to be statistically independent predictors of CAD risk even after adjustment for multiple risk factors, including total-cholesterol levels. Elevated triglyceride levels are common in older postmenopausal women, in whom they may be markers for the presence of metabolic syndrome, insulin resistance, and associated atherogenic lipoproteins.

Diabetes Mellitus

Diabetes mellitus is one of the most important risk factors for CAD in women, regardless of age. The risk of MI is twice as high in women with diabetes as in nondiabetic women of the same age. The incidence of CAD in diabetic women is almost three times that of nondiabetic women, whereas the incidence is doubled in diabetic men. The presence of diabetes tends to attenuate any gender-related differences in cardiovascular morbidity and mortality. In fact, the risk of cardiovascular events in diabetic women is higher than that of both diabetic men and nondiabetic women, even after adjustment has been made for age and other cardiovascular risk factors.

The higher risk in diabetic women may be partially explained by the clustering of multiple risk factors, such as hypertension, smoking, and obesity. Aside from diabetes or high blood glucose levels, diabetic women have significantly greater degrees of four cardiovascular risk factors: low HDL levels, high triglyceride levels, high systolic blood pressure, and high body mass index (BMI). Differences in lipoprotein levels between diabetic and nondiabetic patients are much more dramatic in women than in men.[32] Whether the increased risk associated with diabetes in women results from the clustering of other risk factors or from diabetes itself remains to be fully explored.

Hypertension

Premenopausal women are less likely than men to have hypertension. Nevertheless, one in four white women 20 years of age or older have hypertension. The tendency of women to have lower blood pressure than men early in life is counterbalanced by the fact that blood pressure increases more steeply with aging in women than it does in men. Before menopause, the predominant risk factors for hypertension in women are gender specific: contraceptive use and pregnancy. With menopause, many women—African-American women in particular—develop hypertension. By age 65, more than half of all women are hypertensive, and about 80% are hypertensive by age 75.

Obesity

Obesity has been shown to be an independent risk factor for the development of CAD in women. In a 26-year follow-up of participants in the Framingham Study, relative weight in women was positively and independently associated with cardiovascular events. It is estimated that 30% to 40% of CAD cases in obese women can be attributed to the excess weight alone, and even mild to moderate overweight increases the risk of CAD in middle-aged women.[33] In the Nurses' Health Study, women with a BMI of 25 to 29 kg/m^2 had an age-adjusted relative risk for CAD of 1.8, compared with the leanest women, and obese women (i.e., those with a BMI greater than 29) had a relative risk for CAD of 3.3.[34,35] Approximately 25% of women in the United States between 35 and 64 years of age have BMIs of 29 or higher, and in this group, the relative risks of nonfatal MI and fatal CAD are 3.2 and 3.5, respectively.

The distribution of body fat may be more important than absolute weight in determining cardiovascular risk. Fat distribution, as assessed by the waist-to-hip ratio, is significantly related to coronary atherosclerosis in both women and men, and the waist-to-hip ratio is significantly greater in women with CAD. Truncal obesity (also known as android or male-pattern obesity), manifested as a high waist-to-hip ratio, correlates with higher LDL and lower HDL levels.

Smoking

Whereas diabetes is the most gender-differentiated biologic risk factor for CAD in women, cigarette smoking may be the most psychologically and sociologically distinctive behavioral risk factor. Although the prevalence of smoking has always been lower in women than in men, smoking prevalence has declined more rapidly in men than in women, and young women are now smoking at greater rates than young men. This narrowing of the gender gap in smoking prevalence is likely to contribute to a substantially greater burden of cardiovascular disease in women.

Physical Activity

Sedentary lifestyle is now recognized as a major risk factor for CAD.[36-38] Even moderately fit women demonstrate significantly lower blood glucose levels, blood pressures, and weight, in addition to better lipid profiles, than women in the lowest-fitness category. Consequently, exercise is a key component of any prevention program.[39-41]

PREVENTION

The American Heart Association (AHA) has released evidence-based guidelines for the prevention of cardiovascular disease in women, which are available on the Internet (http://circ.ahajournals.org/cgi/content/full/109/5/672).[42] To provide interventions tailored to individual patient risk, the guidelines provide a scoring sheet for calculating a woman's 10-year risk of having a coronary event, on the basis of the Framingham Risk Calculator [*see Table 1*]. The Framingham Risk Calculator can be used to assess a patient's 10-year absolute risk of CAD and, on that basis, to categorize her as being at high, intermediate, low, or optimal risk.

Once a patient's risk factors have been assessed and her risk level stratified, interventions can be tailored accordingly. Interventions can be directed toward lifestyle, major risk factors, cardiovascular preventive therapies, or measures to counter atrial fibrillation and prevent stroke, according to the patient's risk level [*see Tables 2 through 4*]

Hormone Replacement Therapy

Much controversy surrounds the role of hormone replacement therapy (HRT) in cardiovascular disease prevention. Although pathophysiologic and epidemiologic data suggest that estrogen replacement has beneficial cardiovascular effects, other data from large, randomized, controlled trials have demonstrated no benefit as well as potential harm from HRT when it is used for the prevention of CAD.[43-54] One arm of the Women's Health Initiative (WHI), a randomized, controlled trial of combined estrogen and progesterone replacement in 16,608 healthy postmenopausal women, was halted early because of an increased risk of breast cancer (hazard ratio [HR], 1.26); cardiovascular events were also increased (HR, 1.29), as was the risk of stroke (HR, 1.41) and pulmonary embolism (HR, 2.13).[43,44] In a meta-analysis of 22 small randomized trials of HRT encompassing 4,124 women, the calculated odds ratio for cardiovascular events in women assigned to hormones versus those not assigned to hormones was 1.39 (95% confidence interval [CI], 0.48–3.95).[45]

The first large-scale clinical trial to examine the cardioprotective effect of estrogen in postmenopausal women with CAD was the Heart and Estrogen/progestin Replacement Study (HERS), a multicenter, randomized, double-blind, placebo-controlled trial in 2,763 postmenopausal women.[46] Over the first year of the study, there was a statistically significant 52% excess risk of CAD events in the HRT group; an additional 6.8 years of follow-up also failed to show any evidence of long-term cardiovascular benefit.[47,48]

In the Estrogen in the Prevention of Reinfarction Trial (ESPRIT), a randomized , controlled trial in 1,017 postmenopausal women with prior MI, 2 years of estrogen treatment did not lower the frequency of reinfarction or cardiac death or reduce all-cause mortality.[49] Similarly, the Estrogen Replacement and Atherosclerosis (ERA) trial, a randomized, controlled trial in 309 postmenopausal women, found that after an average follow-up of 3.2 years, women assigned to HRT showed no significant difference in mean minimal coronary artery diameters, despite significant reductions in LDL levels and increases in HDL levels.[50,51] Finally, in the Women's Angiographic Vitamin and Estrogen (WAVE) trial, a randomized trial in 423 postmenopausal women that assessed estrogen and progesterone with or without vitamins E and C, there was greater progression of coronary atherosclerosis in the treatment groups.[52] Studies investigating the role of estrogen in the secondary prevention of cerebrovascular disease, including the Women's Estrogen for Stroke Trial (WEST), have failed to show benefit.[53]

Because of these trial results, long-term HRT is not recommended for either primary or secondary prevention of CAD. The AHA advises that there are insufficient data to support starting HRT in women who do not have CAD for the sole purpose of preventing CAD; the decision to initiate HRT—or to continue it, if the patient is already receiving it—should be made on the basis of non-CAD benefits and risks, possible CAD benefits and risks, and patient preference.[54,55] Women with CAD should not be started on HRT for the secondary prevention of CAD; if they are already receiving long-term HRT, the decision to continue or stop it should be made on the basis of established non-CAD benefits and risks and patient preference.

DIAGNOSIS OF CAD

The approaches to the diagnosis of CAD are similar in men and women [*see 1:I Approach to the Cardiovascular Patient and 1:IX Chronic Stable Angina*]. Regardless of the patient's gender, diagnosis encompasses a detailed history, a complete physical exami-

Table 1 Framingham Point Score for Estimation of 10-Year Cardiovascular Risk in Women

Age (yr)	*Points*
20–34	– 7
35–39	– 3
40–44	0
45–49	3
50–54	6
55–59	8
60–64	10
65–69	12
70–74	14
75–79	16

Total Cholesterol (mg/dl)	*Points*				
	Age 20–39 yr	*Age 40–49 yr*	*Age 50–59 yr*	*Age 60–69 yr*	*Age 70–79 yr*
< 160	0	0	0	0	0
160–199	4	3	2	1	1
200–239	8	6	4	2	1
240–279	11	8	5	3	2
≥ 280	13	10	7	4	2

	Points				
	Age 20–39 yr	*Age 40–49 yr*	*Age 50–59 yr*	*Age 60–69 yr*	*Age 70–79 yr*
Nonsmoker	0	0	0	0	0
Smoker	9	7	4	2	1

High-Density Lipoprotein Level (mg/dl)	*Points*
≥ 60	– 1
50–59	0
40–49	1
< 40	2

Systolic Blood Pressure (mm Hg)	*If Untreated*	*If Treated*
< 120	0	0
120–129	1	3
130–139	2	4
140–159	3	5
≥ 160	4	6

Total Points	*10-Year Risk (%)*	*Risk Category*
< 9	< 1	Optimal
9	1	Low Risk
10	1	
11	1	
12	1	
13	2	
14	2	
15	3	
16	4	
17	5	
18	6	
19	8	
20	11	Intermediate risk
21	14	
22	17	
23	22	High risk
24	27	
≥ 25	≥ 30	

Table 2 Lifestyle Interventions for Cardiovascular Disease in Women[42]

Intervention	*Comment*	*Evidence Rating*
Tobacco avoidance	Consistently encourage smoking cessation; encourage nonsmokers not to smoke and to avoid environmental tobacco smoke	Class I, level B
Physical activity	Cumulative ≥ 30 min moderate-intensity physical activity (e.g., brisk walking) on most, and preferably all, days of the week	Class I, level B
Cardiac rehabilitation	Women with a recent acute coronary syndrome or coronary intervention, new-onset or chronic angina should participate in a comprehensive physician-guided or community-based risk-reduction regimen	Class I, level B
Heart-healthy diet	Eat a variety of fruits, vegetables, grains, low-fat or nonfat dairy products, fish, legumes, and sources of protein low in saturated fat (e.g., poultry, lean meats, plant sources); limit saturated fat intake to 10% of calories; limit cholesterol intake to < 300 mg/day; limit intake of *trans*–fatty acids	Class I, level B
Weight maintenance/ reduction	Use appropriate balance of physical activity, caloric intake, and formal behavioral programs when indicated to maintain or achieve a BMI of 18.5–24.9 kg/m^2 and waist circumference < 35 in	Class I, level B
Psychosocial factors	Evaluate women with cardiovascular disease for depression; refer or treat when indicated	Class IIa, level B
Omega-3 fatty acids	Consider dietary supplementation with omega-3 fatty acids in high-risk women	Class IIb, level B
Folic acid	Consider dietary supplementation with folic acid in high-risk women (except after revascularization procedure) who have elevated homocysteine levels	Class IIb, level B

BMI—body mass index

nation, and an electrocardiogram. Once the initial evaluation is performed, laboratory blood tests, stress testing, and cardiac catheterization may be necessary to obtain further diagnostic insight. General recommendations have been published by the American College of Cardiology (ACC), the AHA, and the American College of Physicians (ACP); these are available on the Internet (http://www.annals.org/cgi/content/abstract/141/7/562).[56] An essential point is that because of a lower pretest probability of CAD at any age, tests have lower sensitivity in women.

Clinical Manifestations

Whereas MI and sudden cardiac death are the more common presentations in men with CAD, angina pectoris is the main initial and subsequent presenting symptom of CAD in women. Compared with men, women with angina are more likely to be older and to have hypertension, diabetes, and heart failure. They are less likely than men to have a history of either MI or percutaneous coronary intervention.[57-59]

When women present with symptoms of CAD, they are less likely to have typical or classic symptoms of heart disease and more likely to have atypical symptoms, particularly abdominal, neck, and shoulder pain. Older and diabetic women have a higher frequency of dyspnea and fatigue. Other ischemic symptoms in women are sleep disturbance, indigestion, and anxiety.

Laboratory Testing

Because women at all ages are less likely to have obstructive CAD, the pretest probability of CAD is lower in women with

Table 3 Major Risk Factor Interventions for Cardiovascular Disease in Women[42]

Intervention	*Comment (Evidence Rating)*
Blood pressure normalization	Lifestyle: encourage optimal BP of < 120/80 mm Hg (class I, level B) Drugs: treat when BP ≥ 140/90 mm Hg, or lower in patients with target-organ damage or diabetes; thiazide diuretics should be part of initial regimen in most patients unless contraindicated (class I, level A)
Lipid normalization—nonpharmacologic	Encourage lifestyle approaches to maintain optimal levels: LDL < 100 mg/dl, HDL > 50 mg/dl, triglycerides < 150 mg/dl, non–HDL cholesterol < 130 mg/dl (class I, level B) Diet: in woman at high risk or with elevated LDL, reduce saturated fat intake to < 7% of calories, cholesterol to < 200 mg/day; reduce *trans*-fatty acid intake (class I, level B)
Lipid normalization—pharmacologic	High-risk patients: LDL-lowering therapy—preferably a statin—plus lifestyle therapy in women with LDL ≥ 100 mg/dl (class I, level A); statin therapy in women with LDL < 100 mg/dl, unless contraindicated (class I, level B); niacin or fibrate therapy for women with low HDL or elevated non-HDL (class I, level B) Intermediate-risk patients : LDL-lowering therapy (preferably a statin) if LDL ≥ 130 mg/dl despite lifestyle therapy (class I, level A); niacin or fibrate therapy for women with low HDL or elevated non-HDL after LDL goal reached (class IIa, level B) Low-risk patients: consider LDL-lowering therapy in women with one or no risk factors when LDL ≥ 190 mg/dl or in women with multiple risk factors when LDL ≥ 160 mg/dl (class IIa, level B); consider niacin or fibrate therapy in women with low HDL or elevated non-HDL after LDL goal reached (class IIa, level B)
Diabetes treatment	Use lifestyle and pharmacotherapy to achieve glycosylated hemoglobin < 7% (class I, level B)

BP—blood pressure HDL—high-density lipoprotein cholesterol LDL—low-density lipoprotein cholesterol

Table 4 Preventive Drug Interventions for Cardiovascular Disease in Women[42]

Drug	*Comment (Evidence Rating)*
Aspirin	High-risk patients: 75–162 mg/day, unless contraindicated; clopidogrel in aspirin-intolerant patients (class I, level A) Intermediate-risk patients: consider 75–162 mg/day, if blood pressure is controlled and benefit is likely to outweigh risk of GI side effects (class IIa, level B) Low-risk patients: Routine aspirin use not recommended (class III, level B) Patients with chronic or paroxysmal atrial fibrillation: 325 mg/day, for stroke prevention when warfarin is contraindicated or stroke risk is low (< 1%/yr) (class I, level A)
Beta blockers	Unless contraindicated, should be used indefinitely in all women who have had an MI or who have ischemic syndromes (class I, level A)
ACE inhibitors	Unless contraindicated, should be used in women at high risk (class I, level A)
ARBs	In high-risk women with clinical evidence of heart failure or an LVEF < 40% who are intolerant to ACE inhibitors (class I, level B)
Warfarin	Treatment to INR of 2.0–3.0; for stroke prevention in women with chronic or paroxysmal atrial fibrillation, unless they are considered at low risk for stroke (< 1%/yr) or at high risk for bleeding (class I, level A)

ACE—angiotensin-converting enzyme ARBs—angiotensin receptor blockers GI—gastrointestinal INR—international normalized ratio LVEF—left ventricular ejection fraction MI—myocardial infarction

chest pain presenting for a diagnostic evaluation. Therefore, false positive results with exercise ECGs are more common in women than in men[60-62]; this also means that a negative result on exercise ECG in a woman has high predictive accuracy for the absence of clinically significant CAD.

The use of exercise echocardiography and single-photon emission computed tomography (SPECT) imaging has been promoted in women, but data on the utility of these tests in this setting are limited. In a cross-sectional study of the accuracy of exercise SPECT imaging with thallium-201, the sensitivity of exercise thallium SPECT imaging was lower in women than in men. Specificities were similar in men and women. A systematic review of SPECT versus exercise echocardiography showed that although the sensitivity of exercise SPECT was high (87%), its specificity was low (64%), resulting in a low likelihood ratio (1.9) for both men and women.[63]

Myocardial perfusion imaging improves the diagnostic accuracy of exercise testing; technetium-99m sestamibi SPECT imaging offers particular benefit. Stress echocardiography in women with adequate echocardiographic images displays sensitivity and specificity comparable to those in men. The most cost-effective diagnostic strategy is a sequential approach to testing tailored to patient risk and symptoms.

The most cost-effective diagnostic strategy is a sequential approach to exercise testing that is tailored to the patient's coronary risk factors and symptoms. According to the ACC/AHA practice guidelines for exercise testing, a woman's pretest likelihood for CAD may be crudely defined on the basis of age and the presence and characteristics of symptoms [*see Table 5*]; testing has the largest potential effect on diagnostic outcome in women at intermediate risk.[64] In addition, symptomatic women who have diabetes mellitus or multiple risk factors (i.e., the metabolic syndrome) are at increased risk for CAD and should be considered for testing.

Acute Coronary Syndromes

Although women with acute coronary syndromes (i.e., unstable angina, non–ST segment elevation MI, and ST elevation MI) are more likely to present with nausea and vomiting and indigestion than men, typical symptoms are the strongest predictors of acute coronary syndromes in women [*see I:X Unstable Angina and Non–ST Segment Elevation Myocardial Infarction*]. In the Myocardial Infarction Triage and Intervention Registry, the clinical presentation of MI was indistinguishable by sex, with comparable numbers of women and men having typical and atypical pain presentations.[65] However, in the Worcester Heart Study, more women than men had a history of angina pectoris before their initial MI.[66-70] In addition, women with acute coronary syndromes are more likely to have a higher Killip class, tachycardia, atrioventricular block, and pulmonary rales on presentation; complications such as shock, heart failure, recurrent chest pain, cardiac rupture, and stroke are more common in women.

Time to presentation may differ between men and women. In the first Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Arteries (GUSTO-I) trial, median time from onset of chest pain to admission was longer for women than for men.[71] Initial studies suggested that women were more likely to have unstable angina than to have documented acute MI.[72,73] Women with MI are less likely than men to have ST segment elevation MI. Compared with men with MI, women with MI tend to be older; are more likely to have diabetes and hypertension and to have had heart failure; and are less likely to have had an MI.

TREATMENT

There is currently no evidence that acute coronary syndromes should be treated differently in women than in men, and current therapeutic guidelines do not make recommendations based on gender. A number of randomized, controlled clinical trials of MI therapies—including antiplatelet drugs, beta blockers, calcium channel blockers, fibrinolytic drugs, and ACE inhibitors—confirm comparable reductions in mortality for women and men. Unfortunately, the use of these therapies after acute coronary syndromes is less common in women than in men. For example, in the GUSTO-I study, median time from onset of chest pain to thrombolysis was longer for women than for men. In addition, women are less likely than men to be treated aggressively. They are half as likely to be considered for acute catheterization, angioplasty, thrombolysis, or coronary artery bypass grafting.[71] Women are also less likely to be referred for cardiac rehabilitation after a cardiovascular event. The inequitable application of treatment guidelines may result in poorer outcomes for women.

Table 5 Pretest Probability of Coronary Artery Disease in Women[64]

Age	*Typical or Definite Angina Pectoris*	*Atypical or Probable Angina Pectoris*	*Nonanginal Chest Pain*	*Asymptomatic*
30–39	Intermediate	Very low	Very low	Very low
40–49	Intermediate	Low	Very low	Very low
50–59	Intermediate	Intermediate	Low	Very low
60–69	High	Intermediate	Intermediate	Low
≥ 70	High	Intermediate	Intermediate	Low

PROGNOSIS

Women with ST segment elevation MI have a higher mortality than men of the same age, whereas women with non–ST segment elevation MI have outcomes similar to those of men. Some of the differences in outcome may be explained by differences in disease burden or severity of MI on presentation.

In the second National Registry of Myocardial Infarction (NRMI-2), younger women with MI (< 70 years) had higher death rates during hospitalization than their male counterparts.[13] Only at older age was there gender parity, and in the very oldest age group (over 80 years), women had better outcomes. Whether these mortality differences reflect the contribution of gender per se or of residual confounders and treatment differences remains to be determined.[13,74-77] Certainly, increasing the use of proven standard therapies offers significant potential to improve clinical outcomes and increase survival in women with suspected acute MI.

Race also affects prognosis: African-American women have a higher mortality from CAD than white women. In the United States, the age-adjusted CAD mortality is 25% to 50% higher for African-American women than for white women; in particular, the mortality from MI is twice as high in African-American women.[78]

Heart Failure in Women

EPIDEMIOLOGY

Each year, more than one million women have heart failure in the United States and approximately 100,000 die of this disease. Men and women differ with respect to the risk, causes, and prognosis of heart failure.[79-81] The majority of deaths attributable to heart failure occur in women, even though women with heart failure have a lower risk of death than men.[82]

RISK FACTORS

Risk factors for development of heart failure differ by sex. Hypertension and diabetes mellitus have a greater role in women, with diabetes disproportionately increasing the risk of heart failure in women.[5,83-86] Women may also be more likely to develop heart failure after MI.[84] An excess number of cases of heart failure and pulmonary edema in women after revascularization was reported both in the Coronary Artery Surgery Study (CASS) and in the Bypass Angioplasty Revascularization Investigation (BARI), despite left ventricular ejection fractions (LVEFs) that were equivalent to those or better than those in men.[87,88]

PATHOPHYSIOLOGY

Although chronic heart failure from left ventricular systolic impairment is well characterized, less is known regarding heart failure with preserved left ventricular systolic function. From 30% to 50% of all patients with chronic heart failure have preserved left ventricular systolic function, and most of these patients are women, are elderly, and have hypertension. Patients with chronic heart failure and preserved left ventricular systolic function appear to have pathophysiologic derangements that are similar to, but less severe than, those with impaired systolic function.[89-92]

DIAGNOSIS

Irrespective of gender, heart failure has a wide spectrum of potential clinical presentations. The majority of patients have signs and symptoms of pulmonary congestion, including dyspnea, orthopnea, and paroxysmal nocturnal dyspnea. Others do not have congestive symptoms but instead have manifestations of low cardiac output, including fatigue, effort intolerance, cachexia, and renal hypoperfusion. Clinically, the New York Heart Association (NYHA) functional classification can be utilized to assess the severity of functional limitations and correlates well with prognosis in both men and women [*see Table 6*]. In addition, the ACC/AHA have issued guidelines to promote earlier identification of heart failure in patients at risk [*see I:II Heart Failure*].

Clinical Manifestations

Women with heart failure have more symptoms than men with similar LVEFs, are older, and are more likely to have hypertension, diabetes mellitus, and preserved systolic function.[79] Patients with heart failure and intact ventricular systolic function, predominantly women, experience dyspnea and fatigue, exercise intolerance, and resultant impaired life quality.

Laboratory Testing

In general, the evaluation of new-onset heart failure in women is similar to that in men [*see 1:II Heart Failure*].

TREATMENT

Pharmacologic Therapy

Although a significant body of evidence supports the use of angiotensin-converting enzyme (ACE) inhibitors, diuretics, and beta blockers in patients with reduced systolic function, few women were included in these studies, and data regarding the benefits of these therapies in women remain conflicting.

ACE inhibitors Afterload reduction and neurohormonal modulation with ACE inhibitors lower mortality; improve heart failure symptoms, exercise tolerance, and LVEF; and reduce emergency room visits and hospitalizations. In a meta-analysis of gender-stratified data for all seven major studies that assessed the impact of ACE inhibitors on mortality, however, the authors conclude that women with symptomatic left ventricular systolic dysfunction probably benefit from ACE inhibitors but that treatment with ACE

Table 6 New York Heart Association Classification of Heart Failure

NYHA Class	*Comment (Evidence Rating)*
I	No symptom limitation with ordinary physical activity
II	Ordinary physical activity somewhat limited by dyspnea (e.g., long-distance walking, climbing two flights of stairs)
III	Exercise limited by dyspnea at mild work loads (e.g., short-distance walking, climbing one flight of stairs)
IV	Dyspnea at rest or with very little exertion

inhibitors may not reduce mortality in women with asymptomatic left ventricular systolic dysfunction (pooled relative risk, 0.96; 95% confidence interval [CI], 0.75 to 1.22) [*see Figure 2*].[93]

Angiotensin receptor blockers Gender-specific analyses of angiotensin receptor blockers (ARBs) in heart failure have not been performed. In clinical trials, ARBs were superior to placebo but not better than ACE inhibitors in improving mortality. ARBs improve morbidity when given along with ACE inhibitors; they are recommended as second-line therapy in patients who cannot tolerant ACE inhibitors because of cough or angioedema.[94] They should not be substituted for ACE inhibitors in cases of hyperkalemia or renal dysfunction. ARBs may be useful for the treatment of diastolic heart failure.

Beta blockers Three beta blockers—carvedilol, metoprolol succinate, and bisoprolol—have been shown to increase survival in patients with heart failure.[95-99] Metoprolol tartrate has not been approved by the Food and Drug Administration for heart failure and was less effective than carvedilol in preventing sudden death in the Carvedilol Or Metoprolol European Trial.[100]

Although the addition of beta blockers to a treatment regimen of diuretics, ACE inhibitors, and digoxin has been recommended to improve clinical outcomes and decrease mortality and hospitalizations, questions have arisen regarding the benefit in women. However, a meta-analysis of the five major mortality studies of beta-blocker treatment in heart failure showed lower mortality in women (risk ratio in women, 0.63 [95% CI, 0.44 to 0.91]; risk ratio in men, 0.66 [95% CI, 0.59 to 0.75]; HR in women, 0.75 [95% CI, 0.51 to 1.09]; and HR in men, 0.68 [95% CI, 0.51 to 0.89].[93]

Despite gender-related differences in outcomes, current guidelines from the Heart Failure Society of America (HFSA) and the ACC/AHA recommend the use of beta blockers for NYHA classes I through III heart failure, irrespective of gender. Beta blockers are also indicated in NYHA class IV patients who are euvolemic—again, irrespective of gender.

Digoxin HFSA and ACC/AHA guidelines recommend digoxin for patients with left ventricular systolic dysfunction who remain symptomatic despite standard medical therapy, particularly if they are in atrial fibrillation. No distinctions are made for gender. Significant gender differences have been reported in response to digoxin for the management of heart failure, however. In the 6,800-patient Digitalis Investigation Group study, for example, women who were randomized to digoxin had a higher rate of death (33.1%) than women who were randomized to placebo (28.9%). In contrast, men who received digoxin or placebo had similar death rates. In the multivariable analysis, digoxin was associated with significantly higher risk of mortality in women (HR, 1.23; 95% CI, 1.02 to 1.47), but it had no significant effect on mortality in men (P = 0.014; HR, 0.93; 95% CI, 0.85 to 1.02).[101] This study suggests that digoxin therapy is associated with an increased risk of death from any cause in women with heart failure and depressed left ventricular systolic function.

Diuretics and aldosterone antagonists Diuretics and aldosterone antagonists (e.g., spironolactone) in the treatment of heart failure are used similarly in men and women [*see 1:II Heart Failure*].

Heart Failure with Preserved Ventricular Systolic Function

Most large clinical trials of heart failure management strategies have involved patients with a decreased LVEF. Older patients with heart failure, among whom women predominate, are more likely to have preserved ventricular systolic function, a problem not addressed in these major treatment trials.[11,102-104] The few heart failure trials in patients with preserved ventricular systolic function have failed to produce conclusive evidence for optimal pharmacologic management. Because no clinical trial evidence is consistent, the ACC/AHA guidelines recommend a pathophysiologic approach to treatment that consists of measures to control blood pressure and tachycardia, to decrease cen-

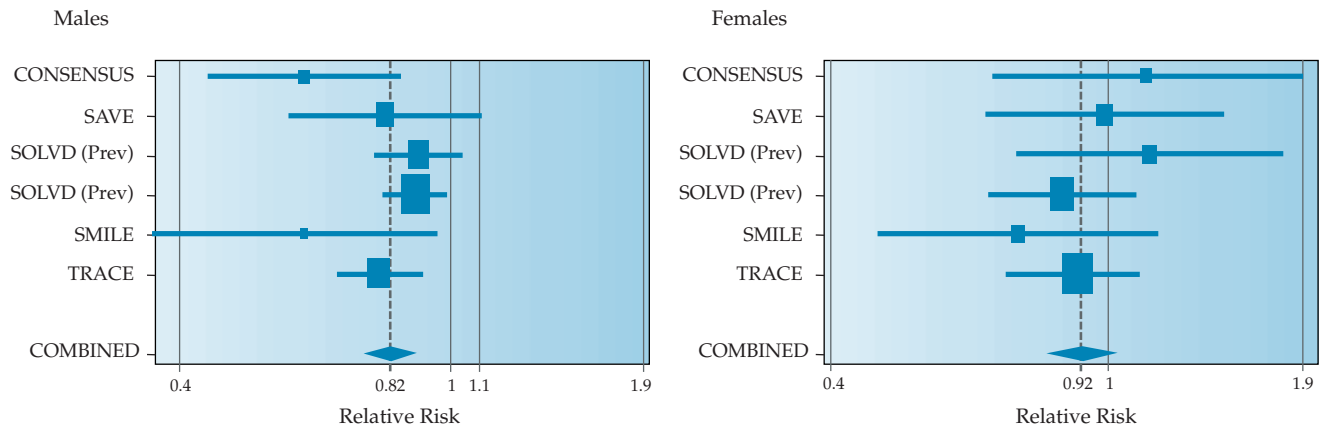


Figure 2 **Efficacy of angiotensin-converting enzyme inhibitors in the management of left ventricular systolic dysfunction according to sex.**[93]

tral blood volume, and to alleviate myocardial ischemia.

PROGNOSIS

Research on gender and heart failure prognosis has shown inconsistent results.[105-109] Epidemiologic data from the Framingham cohort suggest that the prognosis in women is better than that in men.[83] In contrast, Bourassa and colleagues demonstrated a poor prognosis in older women with heart failure.[82]

Differences in survival may reflect differences in etiology.[105,110] An Italian study found that women were more likely than men to have a nonischemic cause of their heart failure, and LVEF was higher in nonischemic women than nonischemic men; the survival advantage in women was in part attributed to better left ventricular function and in part to the primary etiology of heart failure.[111] In a United States study, survival was significantly better in women than in men when heart failure was from nonischemic causes; but when the principal etiology was ischemic heart disease, men and women had similar survival rates.[105]

Pathophysiology also affects prognosis in heart failure. Although mortality is significantly greater in patients with chronic heart failure who have impaired left ventricular systolic function than in those with preserved systolic function, even patients with preserved systolic function—most of whom are women—have a 25% 5-year mortality.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Major cardiovascular disease (CVD) during 1997–1999 and major CVD hospital discharge rates in 1997 among women with diabetes—United States. MMWR Morb Mortal Wkly Rep 50(RR-43):948, 2001

2. Lewis SJ: Cardiovascular disease in postmenopausal women: myths and reality. Am J Cardiol 89:5E, 2002

3. Kuster GM, Buser P, Osswald S, et al: Comparison of presentation, perception, and six-month outcome between women and men > or = 75 years of age with angina pectoris. Am J Cardiol 91:436, 2003

4. Knopp RH: Risk factors for coronary artery disease in women. Am J Cardiol 89:28E, 2002

5. Kannel WB: The Framingham Study: historical insight on the impact of cardiovascular risk factors in men versus women. J Gender-Specific Med 5:27, 2002

6. Bertsias G, Mammas I, Linardakis M, et al: Overweight and obesity in relation to cardiovascular disease risk factors among medical students in Crete, Greece. BMC Public Health 3:3, 2003

7. Prevalence of selected cardiovascular disease risk factors among American Indians and Alaska Natives—United States, 1997. MMWR Morb Mortal Wkly Rep 49(RR-21):461, 2000

8. Vaccarino V: Women and outcomes of coronary artery bypass surgery: do we have an answer? Am Heart J 146:935, 2003

9. Gorman Koch C, Mora Mangano C, Schwann N, et al: Is it gender, methodology, or something else? J Thorac Cardiovasc Surg 126:932, 2003

10. Abramson JL, Veledar E, Weintraub WS, et al: Association between gender and in-hospital mortality after percutaneous coronary intervention according to age. Am J Cardiol 91:968, 2003

11. Smith GL, Masoudi FA, Vaccarino V, et al: Outcomes in heart failure patients with preserved ejection fraction: mortality, readmission, and functional decline. J Am Coll Cardiol 41:1510, 2003

12. Vaccarino V, Abramson JL, Veledar E, et al: Sex differences in hospital mortality after coronary artery bypass surgery: evidence for a higher mortality in younger women. Circulation 105:1176, 2002

13. Vaccarino V, Parsons L, Every NR, et al: Sex-based differences in early mortality after myocardial infarction. National Registry of Myocardial Infarction 2 Participants. N Engl J Med 341:217, 1999

14. Vaccarino V, Chen YT, Wang Y, et al: Sex differences in the clinical care and outcomes of congestive heart failure in the elderly. Am Heart J 138:835, 1999

15. Krumholz HM, Philbin DM Jr, Wang Y, et al: Trends in the quality of care for Medicare beneficiaries admitted to the hospital with unstable angina. J Am Coll Cardiol 31:957, 1998

16. Nohria A, Vaccarino V, Krumholz HM: Gender differences in mortality after myocardial infarction. Why women fare worse than men. Cardiol Clin 16:45, 1998

17. Vaccarino V, Krumholz HM, Mendes de Leon CF, et al: Sex differences in survival after myocardial infarction in older adults: a community-based approach. J Am Geriatr Soc 44:1174, 1996

18. Wenger NK: Coronary risk reduction in the menopausal woman. Revista Portuguesa de Cardiologia 18(suppl 3):III39, 1999

19. Mukherjee D, Hsu A, Moliterno DJ, et al: Risk factors for premature coronary artery disease and determinants of adverse outcomes after revascularization in patients < or = 40 years old. Am J Cardiol 92:1465, 2003

20. Cole JH, Miller JI 3rd, Sperling LS, et al: Long-term follow-up of coronary artery disease presenting in young adults. J Am Coll Cardiol 41:521, 2003

21. Lee PY, Alexander KP, Hammill BG, et al: Representation of elderly persons and women in published randomized trials of acute coronary syndromes. JAMA 286:708, 2001

22. Hulley SB: The US National Cholesterol Education Program. Adult treatment guidelines. Drugs 36(suppl 3):100, 1998

23. Greenlund KJ, Giles WH, Keenan NL, et al: Prevalence of multiple cardiovascular disease risk factors among women in the United States, 1992 and 1995: the Behavioral Risk Factor Surveillance System. J Women's Health 7:1125, 1998

24. Rifkind BM, Tamir I, Heiss G, et al: Distribution of high density and other lipoproteins in selected LRC prevalence study populations: a brief survey. Lipids 14:105, 1979

25. Bittner V: Lipoprotein abnormalities related to women's health. Am J Cardiol 90:77i, 2002

26. Stensvold I, Urdal P, Thurmer H, et al: High-density lipoprotein cholesterol and coronary, cardiovascular and all cause mortality among middle-aged Norwegian men and women. Eur Heart J 13:1155, 1992

27. Kannel WB, McGee D, Gordon T: A general cardiovascular risk profile: the Framingham Study. Am J Cardiol 38:46, 1976

28. Wilson PW, Garrison RJ, Castelli WP, et al: Prevalence of coronary heart disease in the Framingham Offspring Study: role of lipoprotein cholesterols. Am J Cardiol 46:649, 1980

29. Wilson PW, D'Agostino RB, Levy D, et al: Prediction of coronary heart disease using risk factor categories. Circulation 97:1837, 1998

30. Livshits G, Weisbort J, Meshulam N, et al: Multivariate analysis of the twenty-year follow-up of the Donolo–Tel Aviv Prospective Coronary Artery Disease Study and the usefulness of high density lipoprotein cholesterol percentage. Am J Cardiol 63:676, 1989

31. Brunner D, Weisbort J, Meshulam N, et al: Relation of serum total cholesterol and high-density lipoprotein cholesterol percentage to the incidence of definite coronary events: twenty-year follow-up of the Donolo–Tel Aviv Prospective Coronary Artery Disease Study. Am J Cardiol 59:1271, 1987

32. Walden CE, Knopp RH, Wahl PW, et al: Sex differences in the effect of diabetes mellitus on lipoprotein triglyceride and cholesterol concentrations. N Engl J Med 311:953, 1984

33. Mosca L, Grundy SM, Judelson D, et al: Guide to preventive cardiology for women. AHA/ACC Scientific Statement Consensus panel statement. Circulation 99:2480, 1999

34. Kritz-Silverstein D, Barrett-Connor E: Long-term postmenopausal hormone use, obesity, and fat distribution in older women. JAMA 275:46, 1996

35. Barrett-Connor E: Heart disease in women. Fertil Steril 62(6 suppl 2):127S, 1994

36. Dubbert PM, Carithers T, Sumner AE, et al: Obesity, physical inactivity, and risk for cardiovascular disease. Am J Med Sci 324:116, 2002

37. Kohl HW 3rd: Physical activity and cardiovascular disease: evidence for a dose response. Med Sci Sports Exerc 33(6 suppl):S472, 2001

38. Lawler JM, Hu Z, Green JS, et al: Combination of estrogen replacement and exercise protects against HDL oxidation in post-menopausal women. Int J Sports Med 23:477, 2002

39. Miller TD, Fletcher GF: Exercise and coronary artery disease prevention. Cardiologia 43:43, 1998

40. Miller TD, Balady GJ, Fletcher GF: Exercise and its role in the prevention and rehabilitation of cardiovascular disease. Ann Behav Med 19:220, 1997

41. Grundy SM, Bazzarre T, Cleeman J, et al: Prevention Conference V: Beyond secondary prevention: identifying the high-risk patient for primary prevention: medical office assessment: Writing Group I. Circulation 101:E3, 2000

42. Mosca L, Appel L, Benjamin E, et al: Evidence-based guidelines for cardiovascular disease prevention in women. Circulation 109:672, 2004

43. Hays J, Ockene JK, Brunner RL, et al: Effects of estrogen plus progestin on health-related quality of life. N Engl J Med 348:1839, 2003

44. Rossouw JE, Anderson GL, Prentice RL, et al: Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321, 2002

45. Hemminki E, McPherson K: Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. BMJ 315:149, 1997

46. Grady D, Applegate W, Bush T, et al: Heart and Estrogen/progestin Replacement Study (HERS): design, methods, and baseline characteristics. Control Clin Trials 19:314, 1998

47. Hulley S, Grady D, Bush T, et al: Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA 280:605, 1998

48. Grady D, Herrington D, Bittner V, et al: Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II). JAMA 288:49, 2002 (Erratum, JAMA 288:1064, 2002)

49. Cherry N, Gilmour K, Hannaford P: Oestrogen therapy for prevention of reinfarction in postmenopausal women: a randomized placebo controlled trial. Lancet 360:2001, 2002 http://image.thelancet.com/extras/02art11268web.pdf

50. Herrington DM, Reboussin DM, Brosnihan KB, et al: Effects of estrogen replacement on the progression of coronary-artery atherosclerosis. N Engl J Med 343:522, 2000

51. Herrington DM, Howard TD, Hawkins GA, et al: Estrogen-receptor polymorphisms

and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease. N Engl J Med 346:967, 2002

52. Waters DD, Alderman EL, Hsia J, et al: Effects of hormone replacement therapy and antioxidant vitamin supplements on coronary atherosclerosis in postmenopausal women: a randomized controlled trial. JAMA 288:2432, 2002

53. Hurn PD, Brass LM: Estrogen and stroke: a balanced analysis. Stroke 34:338, 2003

54. Viscoli CM, Brass LM, Kernan WN, et al: A clinical trial of estrogen-replacement therapy after ischemic stroke. N Engl J Med 345:1243, 2001

55. Mosca L, Collins P, Herrington DM, et al: Hormone replacement therapy and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. Circulation 104:499, 2001

56. Snow V, Barry P, Fihn SD, et al: Primary care management of chronic stable angina and asymptomatic suspected or known coronary artery disease: a clinical practice guideline from the American College of Physicians. Ann Intern Med 141:562, 2004

57. Milner KA, Funk M, Arnold A, et al: Typical symptoms are predictive of acute coronary syndromes in women. Am Heart J 143:283, 2002

58. Milner KA, Funk M, Richards S, et al: Symptom predictors of acute coronary syndromes in younger and older patients. Nurs Res 50:233, 2001

59. Milner KA, Funk M, Richards S, et al: Gender differences in symptom presentation associated with coronary heart disease. Am J Cardiol 84:396, 1999

60. Miller TD, Roger VL, Hodge DO, et al: Gender differences and temporal trends in clinical characteristics, stress test results and use of invasive procedures in patients undergoing evaluation for coronary artery disease. J Am Coll Cardiol 38:690, 2001

61. Miller TD, Roger VL, Milavetz JJ, et al: Assessment of the exercise electrocardiogram in women versus men using tomographic myocardial perfusion imaging as the reference standard. Am J Cardiol 87:868, 2001

62. Daly SC, Roger VL, Leibson C, et al: Cardiology services after stress testing: are there sex differences? A population-based study. J Clin Epidemiol 53:661, 2000

63. Fleischmann KE, Hunink MG, Kuntz KM, et al: Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. J Nucl Cardiol 9:133, 2002

64. Mieres JH, Shaw LJ, Arai A, et al: Role of noninvasive testing in the clinical evaluation of women with suspected coronary artery disease: consensus statement from the Cardiac Imaging Committee, Council on Clinical Cardiology, and the Cardiovascular Imaging and Intervention Committee, Council on Cardiovascular Radiology and Intervention, American Heart Association. Circulation 111:682, 2005

65. Kudenchuk PJ, Maynard C, Martin JS, et al: Comparison of presentation, treatment, and outcome of acute myocardial infarction in men versus women (the Myocardial Infarction Triage and Intervention Registry). Am J Cardiol 78:9, 1996

66. Harrold LR, Esteban J, Lessard D, et al: Narrowing gender differences in procedure use for acute myocardial infarction: insights from the Worcester heart attack study. J Gen Intern Med 18:423, 2003

67. Goldberg RJ: A community-wide perspective of secular trends in the therapeutic management of patients with acute myocardial infarction. The Worcester Heart Attack Study. Cardiology 76:105, 1989

68. Goldberg RJ, Gore JM, Alpert JS, et al: Incidence and case fatality rates of acute myocardial infarction (1975–1984): the Worcester Heart Attack Study. Am Heart J 115:761, 1988

69. Goldberg RJ, Gore JM, Alpert JS, et al: Therapeutic trends in the management of patients with acute myocardial infarction (1975–1984): the Worcester Heart Attack Study. Clin Cardiol 10:3, 1987

70. Goldberg RJ, Gore JM, Alpert JS, et al: Recent changes in attack and survival rates of acute myocardial infarction (1975 through 1981). The Worcester Heart Attack Study. JAMA 255:2774, 1986

71. Newby LK, Rutsch WR, Califf RM, et al: Time from symptom onset to treatment and outcomes after thrombolytic therapy. GUSTO-1 Investigators. J Am Coll Cardiol 27:1646, 1996

72. Nabel EG, Selker HP, Califf RM, et al: Women's Ischemic Syndrome Evaluation: current status and future research directions: report of the National Heart, Lung and Blood Institute workshop: October 2–4, 2002: Section 3: diagnosis and treatment of acute cardiac ischemia: gender issues. Circulation 109:e50, 2004

73. Goldberg R, Goff D, Cooper L, et al: Age and sex differences in presentation of symptoms among patients with acute coronary disease: the REACT Trial. Rapid Early Action for Coronary Treatment. Coron Artery Dis 11:399, 2000

74. Vaccarino V, Horwitz RI, Meehan TP, et al: Sex differences in mortality after myocardial infarction: evidence for a sex-age interaction. Arch Intern Med 158:2054, 1998

75. Vaccarino V, Krumholz HM, Yarzebski J, et al: Sex differences in 2-year mortality after hospital discharge for myocardial infarction. Ann Intern Med 134:173, 2001

76. Vaccarino V, Lin ZQ, Kasl SV, et al: Sex differences in health status after coronary artery bypass surgery. Circulation 108:2642, 2003

77. Vaccarino V, Lin ZQ, Kasl SV, et al: Gender differences in recovery after coronary artery bypass surgery. J Am Coll Cardiol 41:307, 2003

78. Mensah GA, Mokdad AH, Ford ES, et al: State of disparities in cardiovascular health in the United States. Circulation 111:1233, 2005

79. Stromberg A, Martensson J: Gender differences in patients with heart failure. Eur J Cardiovasc Nurs 2:7, 2003

80. Ekman I, Ehrenberg A: Fatigue in chronic heart failure—does gender make a difference? Eur J Cardiovasc Nurs 1:77, 2002

81. Jaarsma T: Are women different than men? Aspects of heart failure in special populations: elderly women. Eur J Cardiovasc Nurs 1:29, 2002

82. Bourassa MG, Gurne O, Bangdiwala SI, et al: Natural history and patterns of current practice in heart failure. The Studies of Left Ventricular Dysfunction (SOLVD) Investigators. J Am Coll Cardiol 22(4 suppl A):14A, 1993

83. Ho KK, Anderson KM, Kannel WB, et al: Survival after the onset of congestive heart failure in Framingham Heart Study subjects. Circulation 88:107, 1993

84. Ho KK, Pinsky JL, Kannel WB, et al: The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol 22(4 suppl A):6A, 1993

85. Kannel WB, Belanger AJ: Epidemiology of heart failure. Am Heart J 121:951, 1991

86. Wilson PW, D'Agostino RB, Sullivan L, et al: Overweight and obesity as determinants of cardiovascular risk: the Framingham experience. Arch Intern Med 162:1867, 2002

87. Vasan RS, Benjamin EJ, Levy D: Congestive heart failure with normal left ventricular systolic function: clinical approaches to the diagnosis and treatment of diastolic heart failure. Arch Intern Med 156:146, 1996

88. McCormick JR, Schick EC Jr, McCabe CH, et al: Determinants of operative mortality and long-term survival in patients with unstable angina. The CASS experience. J Thorac Cardiovasc Surg 89:683, 1985

89. Mullany CJ, Mock MB, Brooks MM, et al: Effect of age in the Bypass Angioplasty Revascularization Investigation (BARI) randomized trial. Ann Thorac Surg 67:396, 1999

90. Vasan RS, Benjamin EJ, Levy D: Prevalence, clinical features and prognosis of diastolic heart failure: an epidemiologic perspective. J Am Coll Cardiol 26:1565, 1995

91. Kitzman DW, Little WC, Brubaker PH, et al: Pathophysiological characterization of isolated diastolic heart failure in comparison to systolic heart failure. JAMA 288:2144, 2002

92. Gottdiener JS, McClelland RL, Marshall R, et al: Outcome of congestive heart failure in elderly persons: influence of left ventricular systolic function. The Cardiovascular Health Study. Ann Intern Med 137:631, 2002

93. Shekelle PG, Rich MW, Morton SC, et al: Efficacy of angiotensin-converting enzyme inhibitors and beta-blockers in the management of left ventricular systolic dysfunction according to race, gender, and diabetic status: a meta-analysis of major clinical trials. J Am Coll Cardiol 41:1529, 2003

94. A randomized trial of the angiotensin-receptor blocker valsartan in chronic heart failure. Valsartan Heart Failure Trial Investigators. N Engl J Med 345:1667, 2001

95. Foody JM, Farrell MH, Krumholz HM: β-Blocker therapy in heart failure: scientific review. JAMA 287:883, 2002

96. Farrell MH, Foody JM, Krumholz HM: β-Blockers in heart failure: clinical applications. JAMA 287:890, 2002

97. Packer M, Bristow MR, Cohn JN, et al: The effect of carvedilol on morbidity and mortality in patients with chronic heart failure. N Engl J Med 334:1349, 1996

98. Hjalmarson A, Goldstein S, Fagerberg B, et al: Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF). JAMA 283:1295, 2000

99. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. Investigators and Committees C-I. Lancet 353:9, 1999

100. Poole-Wilson PA, Swedberg K, Cleland JG, et al: Comparison of carvedilol and metoprolol on clinical outcomes in patients with chronic heart failure in the Carvedilol Or Metoprolol European Trial (COMET): randomised controlled trial. Lancet 362:7, 2003

101. Rathore SS, Wang Y, Krumholz HM: Sex-based differences in the effect of digoxin for the treatment of heart failure. N Engl J Med 347:1403, 2002

102. Masoudi FA, Havranek EP, Smith G, et al: Gender, age, and heart failure with preserved left ventricular systolic function. J Am Coll Cardiol 41:217, 2003

103. Heiat A, Gross CP, Krumholz HM: Representation of the elderly, women, and minorities in heart failure clinical trials. Arch Intern Med 162:1682, 2002

104. Hall WD: Representation of blacks, women, and the very elderly (aged > or = 80) in 28 major randomized clinical trials. Ethn Dis 9:333, 1999

105. Adams KF Jr, Dunlap SH, Sueta CA, et al: Relation between gender, etiology and survival in patients with symptomatic heart failure. J Am Coll Cardiol 28:1781, 1996

106. Tandon S, Hankins SR, Le Jemtel TH: Clinical profile of chronic heart failure in elderly women. Am J Geriatr Cardiol 11:318, 2002

107. Muntwyler J, Abetel G, Gruner C, et al: One-year mortality among unselected outpatients with heart failure. Eur Heart J 23:1861, 2002

108. Kober L, Torp-Pedersen C, Pedersen OD, et al: Importance of congestive heart failure and interaction of congestive heart failure and left ventricular systolic function on prognosis in patients with acute myocardial infarction. Am J Cardiol 78:1124, 1996

109. Kimmelstiel CD, Konstam MA: Heart failure in women. Cardiology 86:304, 1995

110. Opasich C, Tavazzi L, Lucci D, et al: Comparison of one-year outcome in women versus men with chronic congestive heart failure. Am J Cardiol 86:353, 2000

111. Women and cardiovascular diseases—statistics. American Heart Association, 2004 http://www.americanheart.org/downloadable/heart/1109000876764FS10WM05REV.DOC

XXI MUSCULOSKELETAL PROBLEMS IN THE FEMALE ATHLETE

Carol C. Teitz, m.d.

Musculoskeletal injuries in the female athlete are for the most part similar to those in the male athlete. However, there are differences in the incidence of these injuries and in the sports in which they tend to occur.[1] Stress fractures are more common in the female athlete because of the higher prevalence of eating disorders and subsequent osteopenia in females. In addition, for reasons that are not completely elucidated, the female athlete has a higher rate of noncontact anterior cruciate ligament (ACL) injuries than the male athlete. Finally, as a function of greater participation by females in certain sports, such as dance and gymnastics, injuries specific to those sports are more common in females. This chapter addresses the injuries that are seen commonly in the female athlete.

Stress Fractures

Stress fractures result from an imbalance between bone resorption and formation secondary to repetitive submaximal loads. Extrinsic mechanical causes that can induce this imbalance include duration, frequency, and intensity of load; muscle fatigue; and the shock-absorbing capacity of sports equipment. Intrinsic causes for abnormalities in bone remodeling include the nutritional and hormonal milieu, as well as endocrine disorders that are beyond the scope of this chapter. Disordered eating, amenorrhea, and low bone mineral density—the so-called female athlete triad—can result in osteoporosis in young women.[1] Insufficient caloric intake for the expenditure required for sports is thought to lead to changes in the pulsatile secretion of gonadotropin-releasing hormone and abnormalities in luteinizing hormone, which in turn produce the menstrual irregularities found in some female athletes.[2]

Amenorrhea and oligomenorrhea may cause loss of 2% of bone mass a year, especially when these conditions occur during late adolescence and early adulthood—periods during which peak bone mass is acquired.[3] When amenorrhea is present for more than 3 years, treatment with calcium and estrogen will not reverse this bone loss; nor is there evidence that treatment with oral contraceptives prevents stress fractures.[4] Many studies report increased stress fracture rates in athletes with menstrual irregularities, even when dual-energy x-ray absorptiometry indicates that their bone density lies within the normal range.[4] Epidemiologic studies of stress fractures in athletes and military recruits have shown that these injuries are 1.5 to 4 times more common in females than in males.[5-7] This differential between male and female athletes is especially notable for stress fractures in the pelvis, femoral neck, and metatarsals.

General Treatment Guidelines

Recommending that the patient rest the affected body part—the standard treatment for stress fractures—is insufficient for female athletes. Instead, the physician needs to address any features of the athlete's training habits, nutritional and menstrual status, or equipment that may have contributed to the stress fracture. This will ensure that once the fracture heals, the patient will not be exposed to the same factors that led to the fracture.

Nonsteroidal anti-inflammatory drugs (NSAIDs) are contraindicated in the treatment of pain related to stress fractures, because there is evidence that NSAIDs may interfere with fracture healing.[8,9] When pharmacologic pain control is needed, acetaminophen usually suffices.

A stress fracture is healed when the involved bone is no longer tender to palpation and the patient is able to tolerate either impact on the bone or the use of muscles that place stress on the bone. Progressive return to the desired activity level ensues. For lower-limb fractures, the patient can begin a so-called amphibious rehabilitation program when she can walk without discomfort. In this program, patients begin by jogging in water that is chest high and gradually progress to waist-high water. They then move out of the water and onto land. The decrease in buoyancy provides a gradually increasing load to the bone, giving it a chance to adapt. The next step is sport-specific rehabilitation: the athlete resumes her former training regimen for her sport but starts with a time or distance that is about a third of that at the time of fracture. The mileage or time is then increased at a rate of 10% to 15% a week.

Pubic Ramus Stress Fractures

Pelvic stress fractures make up only 1% to 5% of all stress fractures in all runners, but they are significantly more common in female runners. The most common site is the inferior pubic ramus. In a military population, Hill and colleagues found that 11 of 12 pubic ramus fractures were in female recruits.[10]

Pathophysiology

A crossover running style and overstriding may contribute to the higher incidence of pubic ramus stress fractures in female athletes. The term crossover running refers to adduction of the lower extremity across the line of gait progression. This may occur more commonly in women because of their wider pelves. Overstriding refers to excessive flexion of the lower extremity; it is usually done either to gain speed or to keep up with a taller running partner. Both crossover running and overstriding create repetitive pull by the adductor musculature at its origin on the pubic ramus. Sometimes this pull results in adductor tendinitis; at other times, it results in inferior pubic ramus stress fractures.

Diagnosis

The patient complains of groin or medial thigh pain that increases with activity, especially impact, and is relieved by rest. The physician should ask whether the patient runs with a taller partner. The differential diagnoses of adductor tendinitis, osteitis pubis, and pain originating in the hip joint can be sorted out through careful physical examination and imaging studies.

Physical examination findings Painless range of motion of the hip essentially rules out the hip as the origin of the pain. The key physical findings in pelvic stress fracture include tenderness to palpation of the pubic ramus and difficulty standing on one

leg. Pain with resisted use of the adductor muscles is common but is not helpful for sorting out the other possible diagnoses. Pain during passive adductor stretch is more consistent with tendinitis.

Imaging studies Unless the fracture is very recent, it will appear on an anteroposterior radiograph of the pelvis as a lucent line in the pubic ramus, often surrounded by callus [*see Figure 1*]. When the radiograph is unremarkable but a stress fracture seems likely, the physician should obtain either a bone scan or a magnetic resonance imaging scan. Bone scans are very sensitive, although not specific, and usually suffice to make the diagnosis. MRI can reveal a stress fracture; alternatively, it may show fluid in the adductor tendons or edema on both sides of the symphysis pubis in patients with tendinitis or osteitis pubis, respectively.

Treatment

The patient should stop impact activity, as well as other activities that cause pain. If walking is painful, then crutches should be prescribed for a few weeks, or until the patient can walk without pain. Swimming is often well tolerated as long as the patient avoids the whip kick, which uses the adductor muscles. Bicycling may not be tolerated, because despite the lack of impact, the bicycle seat often presses right on the area of the fracture, creating discomfort.

Pubic ramus stress fractures may take from 2 to 5 months to heal. Surgical treatment is not indicated.

FEMORAL NECK STRESS FRACTURES

Femoral neck stress fracture is relatively uncommon, accounting for less than 10% of all stress fractures,[7] but it is two to four times more common in women than in men.[6,11] Moreover, this fracture typically does not occur in the recreational or novice athlete but rather in athletes who participate in high-level training six or more times a week and who have been involved in regular training for more than 2 years.[6] If the diagnosis is missed, a femoral neck stress fracture may progress to complete fracture, with potentially devastating consequences.

Pathophysiology

In epidemiologic studies, femoral neck stress fractures have been associated with overuse and with demineralization. Weak hip musculature and a discrepancy in leg length of greater than 0.5 in. may contribute as well.

Diagnosis

Patients with a femoral neck stress fracture complain of groin or thigh pain, often referred to the medial knee (via the obturator nerve). The pain increases with weight bearing or impact activity.

Physical examination findings Tenderness to palpation of the femoral neck is rare because of the thickness of the soft tissues in this area. Passive log-rolling of the lower limb may produce pain. If not, active straight-leg raising and passive movement of the hip to its maximal range of motion should be performed; these maneuvers will produce pain in patients with a femoral neck stress fracture. Standing on only the affected limb often produces pain. The Trendelenburg sign is sometimes positive. When none of these findings are present, the examiner should check patellar tendon reflexes, quadriceps muscle strength, and ankle dorsiflexion strength to assess L3 and L4 nerve root function. Strength testing of the hip flexor muscles can yield confusing results in this setting because weakness may be from inhibition by pain rather than a neurologic deficit.

Imaging studies In two thirds of the cases of femoral neck stress fractures, the fracture does not appear on radiographs.[12] Bone scans are so sensitive (93% to 100%) that increased uptake is commonly seen in military recruits whose fractures are asymptomatic. MRI is accurate (100% with MRI, as compared with 68% with bone scan), sensitive, and specific.[12] If an intrapelvic lesion seems possible, the radiologist can be asked to include the pelvis in the hip MRI, or vice versa.

Differential Diagnosis

The differential diagnosis for groin or anterior thigh pain includes adductor tendinitis, osteitis pubis, and pubic ramus stress fracture (see above). In addition, pressure on either the obturator nerve or the femoral nerve from an intrapelvic mass, as well as disk disease at the L2-L4 levels, can produce pain in the same distribution.

Treatment

The treatment depends on the location of the stress fracture in the femoral neck. The superolateral aspect of the neck is under tension; consequently, patients with fractures at this site are at increased risk for complete fracture and displacement. The inferomedial aspect of the neck is under compression; for that reason, complete fracture is less likely. Therefore, medial femoral neck stress fractures are usually treated by the avoidance of weight bearing on the involved limb until the patient is pain free. This may take months. Follow-up radiographs should be obtained monthly, or sooner if symptoms change. Swimming and cycling are not allowed because of the forces across the hip joint induced by the musculature. Patients with medial femoral neck stress fractures should be referred to an orthopedic surgeon if they show no improvement in symptoms after

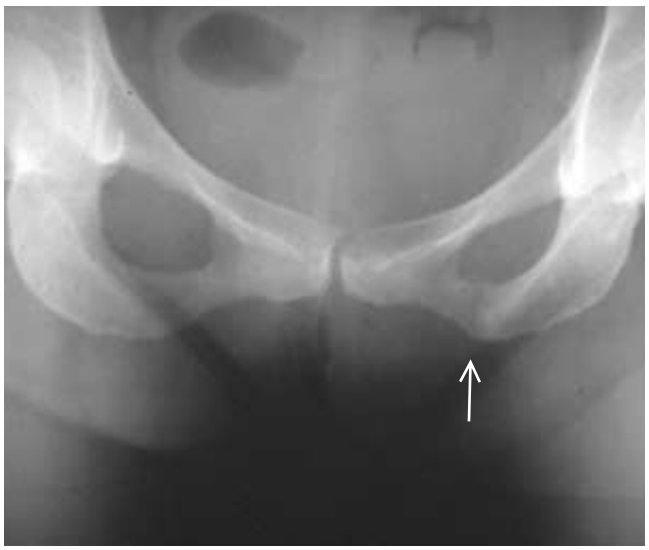


Figure 1 **In a patient with a stress fracture of the inferior pubic ramus, the fracture will appear on an anteroposterior radiograph of the pelvis as a lucent line in the pubic ramus (arrow), often surrounded by callus.**

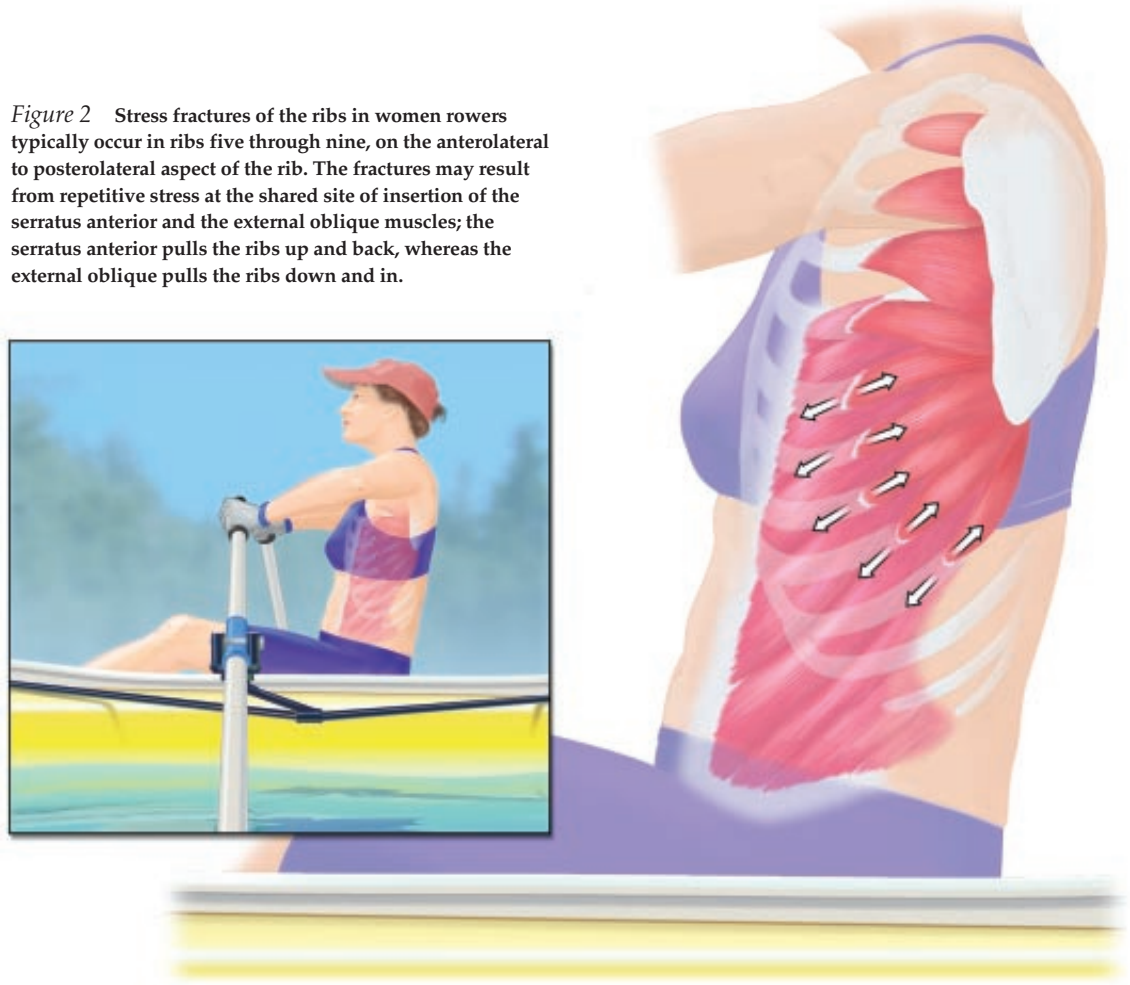


Figure 2 **Stress fractures of the ribs in women rowers typically occur in ribs five through nine, on the anterolateral to posterolateral aspect of the rib. The fractures may result from repetitive stress at the shared site of insertion of the serratus anterior and the external oblique muscles; the serratus anterior pulls the ribs up and back, whereas the external oblique pulls the ribs down and in.**

3 months or if follow-up radiographs show propagation of a fracture line.

Lateral femoral neck stress fractures are usually treated with internal fixation and thus require immediate referral to an orthopedic surgeon. Although internal fixation is debated, no prospective studies have compared operative and nonoperative treatment of the lateral femoral neck stress fracture because of the fear that nonoperative care will result in complete fracture, with its risk of subsequent avascular necrosis, malunion, or nonunion.[12]

RIB STRESS FRACTURES IN ROWERS

Rib stress fractures account for 7.4% of all injuries in the female rower. These fractures typically occur in ribs five through nine, on the anterolateral to posterolateral aspect of the rib.[13]

Pathophysiology

Ribs five through nine share the insertion of the serratus anterior and the external oblique muscles; the serratus anterior pulls the ribs up and back, whereas the external oblique pulls the ribs down and in [*see Figure 2*]. Holden and Jackson attribute these fractures to repetitive stress on the ribs from the pull of these muscles.[14]

Diagnosis

Clinical manifestations Rowers with rib stress fractures present with chest or thoracic back pain that increases during rowing or weight lifting. Coughing also produces pain.

Physical examination findings Patients have tenderness to palpation on the affected rib, usually between the anterior and posterior axillary lines. Compression of the rib cage in an anterior to posterior direction and from side to side increases pain in patients with rib fractures, because these maneuvers produce bending stress on the rib cage. This finding holds even when the compression is applied away from the tender rib. When the problem is muscular, rib-cage compression does not usually produce pain. Muscle and rib injures cannot be differentiated on the basis of resisted muscle use.

Imaging studies The diagnosis of rib stress fractures is confirmed by bone scan.

Differential Diagnosis

The differential diagnosis of rib stress fracture includes muscle strain involving the serratus anterior, pectoralis, or inferior trapezius. Pulmonary or cardiac problems must also be excluded.

Treatment

The athlete should stop rowing. If coughing produces significant pain, a rib belt can be prescribed. While waiting for the fracture to heal, the athlete may do core strengthening and scapular stabilization exercises, omitting the serratus anterior. Once the rib is nontender, serratus strengthening exercises should be

added, with gradual return to rowing. Surgical treatment is not indicated for this fracture.

Multidirectional Shoulder Instability

Shoulder problems are especially prevalent in athletes who engage in swimming, racquet sports, and softball. These sports involve frequent overhead use of the arm, which, in the presence of instability, can cause secondary impingement under the acromion. Shoulder problems are also common in gymnasts, who use the shoulder as a weight-bearing joint.[15] Acquired laxity has been noted in gymnasts, weight lifters, and swimmers who use the butterfly stroke or the backstroke. Laxity, however, is not synonymous with instability; the term instability refers to symptomatic laxity.[16]

Shoulder instability can result from acute trauma or repetitive microtrauma; it may also be atraumatic and multidirectional. Traumatic shoulder instability is more common in males, whereas multidirectional instability is believed to be more common in females, although the literature in this regard is controversial.[16] Generalized ligamentous laxity, which may play a role in shoulder instability, has been noted in numerous studies of multidirectional instability. McFarland and colleagues observed that in 178 asymptomatic high school and college athletes, the shoulder could be subluxated posteriorly in 65% of females compared with 51% of males; generalized laxity was also significantly more common in the female athletes.[17]

PATHOPHYSIOLOGY

Shoulders with multidirectional instability have large inferior capsular pouches that extend both anteriorly and posteriorly. Often, the tissue in the rotator interval between the supraspinatus and subscapularis tendons is attenuated.[15] Some authors have theorized that proprioception is abnormal in these patients, or that weak rotator cuff muscles fail to generate an adequate compressive force against an intact labrum. This type of muscular force is especially important for stabilizing a lax shoulder.[15]

DIAGNOSIS

Clinical Manifestations

Patients with multidirectional shoulder instability often present with pain rather than dislocations. The pain occurs during activities of daily living while the shoulder is in the middle of its range of motion, rather than in the extreme ranges normally associated with traumatic dislocations. The pain is aggravated by overhead use of the arm. Subluxations and dislocations may occur during low-demand activities such as carrying objects with the arm at the side. In extreme cases, the patient is able to voluntarily subluxate or dislocate the glenohumeral joint, as well as reduce it spontaneously.

When a patient presents with pain rather than instability, the differential diagnosis is broad and can include stress fractures, rotator cuff injuries, and tumors, in addition to referred pain from the cervical nerve roots or heart.

Physical Examination Findings

Diagnostic criteria for multidirectional shoulder instability include the demonstration of subluxation in multiple directions in a patient with a history of midrange symptoms.[15] Frequently, the contralateral shoulder is also lax but not symptomatic. The examination should start with an evaluation of the entire kinetic chain, looking for spinal postural abnormalities, hip weakness, and scapular dyskinesia. The one-legged–instability screening test can be informative.[18]

Figure 3 **In a patient with multidirectional shoulder instability, pulling down on the arm at the side may produce inferior translation of the humeral head, as demonstrated by an indentation or dimple beneath the acromion (sulcus sign).**

On examination of the shoulder girdle, the distance between the scapula and the thoracic spine should be symmetrical bilaterally. During range of motion of the arm, the scapula should rotate laterally as the arm elevates. The scapula should not wing away from the thorax. The rotator cuff muscles should be tested with and without scapular retraction. If muscle strength improves with scapular retraction, the functional weakness is more likely from poor scapular stabilization, rather than cuff weakness per se.

Pulling down on the arm at the side may produce inferior translation of the humeral head, as demonstrated by an indentation or dimple beneath the acromion (sulcus sign) [*see Figure 3*]. This test is positive when the inferior translation exceeds 1 to 2 cm; it reflects insufficiency of the superior glenohumeral ligament and capsular rotator interval. The same finding when the arm is tested while abducted 90° reflects attenuation in the inferior capsule.

The degree of instability can be further assessed with the so-called load-and-shift test. With the patient in the supine or lateral decubitus position, the examiner uses one hand to apply an axial load, pushing upward on the arm to force it into the socket, and uses the other hand to shift the patient's arm posteriorly or anteriorly.[15] The glenohumeral joint may dislocate or subluxate in more than one direction (e.g., inferiorly, posteriorly, or anteriorly), with the patient experiencing symptoms in one or more directions.

Imaging Studies

Shoulder radiographs are normal in patients with multidirectional instability. Nevertheless, radiographs are recommended because they may demonstrate Bankart or Hill-Sachs lesions and glenoid dysplasia, which can also cause shoulder instability.

TREATMENT

Operative treatment is not as successful for atraumatic shoulder instability (with or without impingement) as it is for trau-

matic instability. Consequently, surgery should not be the first choice of treatment. Burkhead and Rockwood found that 88% of patients with multidirectional shoulder instability had good or excellent results when treated with specific exercises to increase muscle strength and neuromotor coordination; these exercises are designed to stabilize the scapula and the humeral head in the glenoid fossa.[19] Overhead activities are restricted in the initial phases of rehabilitation. Rotator cuff strengthening alone is not sufficient to stabilize the humeral head. If the scapula fails to rotate normally during arm motion, then the glenohumeral relationship is disturbed and the humeral head is more likely to slide off the glenoid. The hip and trunk often begin the kinetic chain of many upper limb activities and also bear the proximal attachment of many shoulder muscles (e.g., the latissimus dorsi and pectoral muscles). Therefore, strengthening starts with the core muscles, followed by the scapular-stabilizing muscles, and finally the rotator cuff muscles. The scapular stabilizers include the serratus anterior, the latissimus dorsi, the lower trapezius, and the rhomboids. Kibler and Livingston recommend that these exercises first be done in closed kinetic chain fashion (i.e., with the distal limb fixed, or at least touching something; for example, closed kinetic chain exercises of the upper limb are typically done with the hands against a wall).[18] Ultimately, athletic activities require a combination of open and closed kinetic chain muscle function, and both should be addressed in the rehabilitation program. A patient who has failed to recover after 6 months of a good, supervised exercise program should be referred to an orthopedic surgeon for consideration of operative treatment.

Low Back Pain

When evaluating a female athlete with low back pain, the physician should not forget that pain in this area can be referred from the uterus. In addition, the female athlete, particularly in adolescence, has a higher prevalence of spondylolysis, a stress fracture through the pars interarticularis of the posterior elements of a vertebral body. This fracture occurs developmentally in 4% to 5% of the white population and has a familial association. Its prevalence, however, is significantly higher in athletes who hyperextend their spines, such as gymnasts (32%), dancers (12% to 15%), divers (63%), and football linemen.[20] Unlike stress fractures, spondylolysis has not been associated with bone mineral density abnormalities.

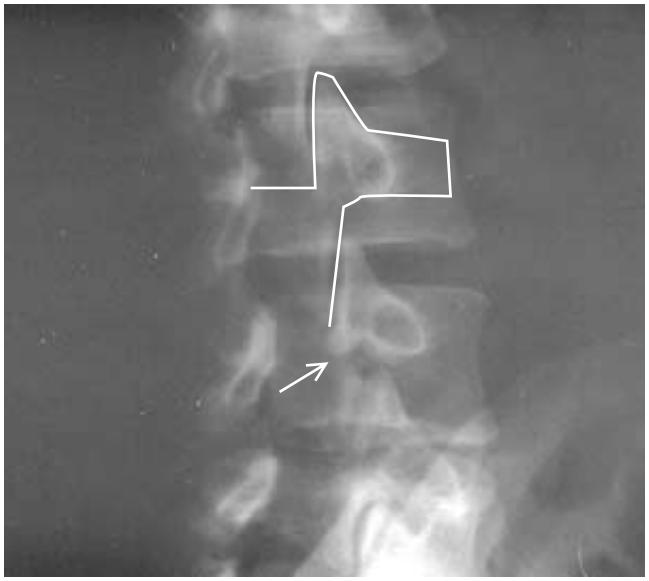


Figure 4 **Spondylolysis may be evident on radiographs, particularly oblique views of the lumbar spine, as the so-called Scottie dog sign. In this radiograph, the Scottie dog is outlined on the normal vertebra above the affected one. Spondylolysis produces the appearance of a collar on the dog (arrow).**

DIAGNOSIS

Clinical Manifestations

The patient presents with low back pain that is made worse by extension or hyperextension of the spine. Sciatica, weakness, and numbness are rare, although the athlete may note decreasing hamstring flexibility. The pain is not exacerbated by Valsalva maneuvers.

Physical Examination Findings

Excessive lumbar lordosis is common. The examiner may find spinal muscle spasm and tenderness to palpation just lateral to the midline of L4, L5, or S1. Pain increases with hyperextension of the spine or leg (i.e., the arabesque position in ballet). When the athlete is asked to stand on one leg and lean back, pain may be produced on the affected side. The straight-leg raising maneuver does not produce pain. Spondylolysis can be differentiated from a spinous process fracture by the fact that in the latter injury, the tenderness will be in the midline on the involved bone rather than lateral.

Imaging Studies

Radiographs, particularly oblique views of the lumbar spine, may show a spondylolysis as the so-called Scottie dog sign [*see Figure 4*]. However, radiographs are often unremarkable in patients with new stress fractures, and these views involve a relatively high dose of radiation. Therefore, bone scans, single-photon emission computed tomographic scans, or MRI scans are recommended.[20] CT can be used later to demonstrate healing. Associated findings in familial spondylolysis include scoliosis, spina bifida occulta, and transitional vertebrae.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis for low back pain that is exacerbated by extension includes injuries or arthritis of the facet joints, fracture of a spinous process, or erector spinae muscle strain.

TREATMENT

Although there is controversy regarding whether spondylolysis can heal, Micheli has demonstrated that antilordotic bracing is effective in the nonfamilial type seen in athletes—especially when the fracture is acute, is "hot" on bone scan, or shows edema and stress reaction on MRI.[20] All athletes with nonfamilial spondylolysis are advised to temporarily refrain from hyperextension activities and to participate in core strengthening exercises. When the spondylolysis is visible radiographically as a fracture line and bracing is used, the athlete probably should not return to activities that involve hyperextension until the fracture line has healed. When a scan shows that the fracture is inactive (i.e., chronic), the athlete may return to full activity when the pain has resolved. Surgery is recommended only when bilateral spondylolysis is present and has progressed to significant spondylolisthesis.[20] Surgical referral is indicated when symptoms have not improved after 6 to 8 weeks of treatment, when

any neurologic abnormalities are noted, or when spondylolisthesis is seen on radiographs.

Anterior Cruciate Ligament Tears

Noncontact ACL tears are two to eight times more common in females than in males. Since 1995, when this increased prevalence was first noted, research has focused on possible etiologic contributing factors and the effectiveness of prevention programs.[21]

PATHOPHYSIOLOGY

Investigators agree that there is no one identifiable cause of the increased incidence of ACL injuries in female athletes and that hormonal, anatomic, and proprioceptive factors may play a role. Studies on hormonal influences are inconclusive. Investigators have found estrogen receptors in the ACL and have reviewed ligamentous laxity, collagen synthesis, and ACL injuries in varying phases of the menstrual cycle. Human studies of the role of hormones in ACL injuries in female athletes have been limited by a failure to determine serum hormone levels in the athletes. Menstrual diaries, urinary hormone levels, and serum estrogen levels alone have been used as markers of the menstrual cycle phase. In addition, the variability in menstrual cycle characteristics makes correlation with injury difficult. A significant limitation of animal models is the difference between estrous and menstrual cycles. Finally, ligamentous laxity may not be directly related to injury risk. Moreover, if hormones do play a role in ACL injury, it may be primary or secondary: hormones could have a direct effect on collagenous tissue, or they could cause neuromuscular changes that in turn affect proprioception, balance, or agility.

Studies of anatomic factors that compared the size of the ACL and the size of the femoral intercondylar notch (home of the ACL) in males and females have yielded inconclusive results. However, a study at West Point found that a combination of increased body mass, generalized joint laxity, and decreased intercondylar notch width significantly increased the risk of ACL injury.[21]

Proprioceptive factors have garnered significant attention, in part because they are the most easily modified etiologic factors. A number of multicenter analyses of athletes videotaped while tearing their ACLs noted that the "position of no return" included apparent knee valgus, tibial rotation, and a relatively straight knee and hip.[22] Laboratory studies have shown that compared with males, females have less knee and hip flexion, increased knee valgus and internal rotation of the hip, and higher quadriceps muscle activity while cutting, pivoting, or landing after jumps.[23]

DIAGNOSIS

Clinical Manifestations

Athletes usually sustain a noncontact ACL injury while in the process of sudden deceleration or directional change. The athlete often feels or hears a pop, falls, and is unable to continue playing. Although athletes often sense that their knee was hyperextended at the time of injury, this has not been documented on video. The knee swells within the first 12 hours after injury, because of hemarthrosis; swelling that develops after that time is likely a synovial effusion. Two large studies have shown that when a knee hemarthrosis is present, even in the absence of demonstrable instability on physical examination, 72% to 80% of patients will have a torn ACL.[24,25]

Physical Examination Findings

In a patient with a torn ACL, the knee is swollen, the range of motion is limited, and motion is painful. Aspiration of the knee joint yields gross blood. An ACL tear that occurs without rupture of the surrounding synovial membrane can present without a hemarthrosis, but this injury is uncommon. The Lachmann test is usually, but not invariably, positive. Joint-line tenderness suggests meniscal tears or bone bruises, which are commonly found in association with ACL tears. Tenderness along the medial retinaculum and apprehension when the examiner pushes the patella laterally suggest a patellar dislocation rather than an ACL tear.

Imaging Studies

MRI is usually definitive in the diagnosis of ACL tears. Although tears most commonly occur "midsubstance," the ACL can avulse from the femoral origin; in children, the ACL can avulse off the tibial insertion. MRI will demonstrate any associated injuries. Finally, MRI will detect other conditions in the differential diagnosis of ACL tears, which includes patellar dislocation and intra-articular fractures.

Treatment

A patient with an ACL tear should be referred to an orthopedic surgeon to discuss therapeutic options. The treatment goal is to provide a stable knee and prevent episodes of instability that are likely to result in a tear of the meniscus or lead to osteoarthritis (OA). ACL reconstruction is recommended, especially in athletes younger than 50 years. Previous studies suggested that patients who experience instability only while engaging in pivoting sports could get by with a brace. However, more patients have opted for reconstruction since ACL reconstruction became an outpatient arthroscopic procedure. Subsequent studies have found that the incidence of OA does not seem to have changed despite reconstruction and early motion postoperatively.[26] Therefore, current avenues of inquiry are whether the OA in these patients results from articular damage at the time of ACL tears, the biochemical environment, or continued athletic activities.[26,27]

PREVENTION

Current prevention programs are resulting in a decreased incidence of ACL injuries.[22] However, these programs are still being studied to determine which components are the most important, the best time to institute such a program, how long the program should last, and which athletes are at risk and thus would benefit the most from such training.

Neuromuscular training is the mainstay of these prevention programs and includes one or more of the following: stretching and strengthening exercises; aerobic, agility, kinesthetic, and plyometric exercises; and risk-awareness education. Strength training and endurance training focus on hip abductor and external rotator muscles. Agility and skill drills emphasize rapid directional changes and therefore stimulate maximum cocontractions of agonist-antagonist muscle pairs to increase stiffness and reduce the number of unanticipated joint movements. Agility drills also improve reflex and cortical response time. Kinesthetic training emphasizes keeping the center of gravity forward and the athlete on her toes. Plyometric training decreases the time to peak torque and increases dynamic joint stability by decreasing landing forces and varus/valgus moments through increased muscle activation.[22]

Patellofemoral Pain

The term patellofemoral pain is used synonymously with anterior knee pain. It has replaced the term chondromalacia patellae because arthroscopic surgery in patients with anterior knee pain has shown no consistent correlation of articular cartilage damage with symptoms.[28]

Patellofemoral pain in the absence of trauma is quite common in women and occurs independently of pain from patellar instability, dislocation, or subluxation—all of which may well be associated with actual morphologic changes in the patellar articular cartilage. The differential diagnosis of anterior knee pain in the absence of trauma includes quadriceps or patellar tendinitis and, rarely, patellar stress fracture. Anterior knee pain can also be a manifestation of chronic regional pain syndrome, which is beyond the scope of this chapter.

PATHOPHYSIOLOGY

Although the pathophysiology of atraumatic patellofemoral pain remains incompletely defined, many factors have been shown to be associated with this type of knee pain. Modification of these factors leads to resolution or lessening of the symptom complex. Theories have centered around subtle abnormalities of patellar tracking related to misalignment of bone, the transmission of aberrant forces from the foot, or imbalance of the musculature controlling knee motion. Because bony alignment is not easily changed, this chapter focuses on soft-tissue factors that may contribute to the clinical problem and that are modifiable.

The lower limb kinetic chain includes the hip, the knee, and the ankle/foot. Aberrations in one area can affect normal mobility and function in another. The patella lies within the quadriceps muscle-tendon complex. Because of the way in which the patellar tendon attaches to the tibial tubercle, as well as the way in which the tibia rotates in response to pronation or supination of the foot, some theorize that excessive subtalar joint motion can produce abnormal patellar angular or rotational movement. At the other end of the kinetic chain, chronic excessive lumbar lordosis can result in a tightening of the hip flexor muscles, including the rectus femoris component of the quadriceps. This tightness, in turn, creates greater compressive patellofemoral forces. Hip stability may also influence the stability of the patella. When hip abductors and external rotator muscles are weak, the knee is more likely to go into a valgus position during weight bearing.[22] A valgus knee tends to cause the patella to move laterally in the trochlea. Finally, muscle balance also influences the movement of the patella in the trochlea. The retinacular attachments to the medial patella of the vastus medialis obliquus component of the quadriceps tendon can be thought of as resisting the pull of the iliotibial band fibers that attach to the lateral aspect of the patella.

DIAGNOSIS

Clinical Manifestations

The patient with patellofemoral pain complains of diffuse pain in the anterior knee. The pain is worse when the patient spends much of the day standing; goes up, and especially down, stairs; walks or hikes down hills; or has to sit with her knees flexed at an angle of 90° or greater (the so-called theater sign). Not uncommonly, when these patients go to the movies, they choose seats at the end of a row so that they can extend the affected leg into the aisle. There may or may not be a history of the knee swelling, clicking, catching, or popping, especially during the performance of deep knee bends. Tendinitis is not usually aggravated by standing or sitting but is made worse by jumping, kicking, or, in adolescents, rapid growth spurts.

Physical Examination Findings

Although the history of a patient with patellofemoral pain should raise the prospect of this diagnosis and therefore focus the physical examination, screening the patient for signs of meniscal and ligament problems of the knee is worthwhile. Any condition causing knee pain or effusion can lead to reflex quadriceps muscle inhibition and atrophy, with resulting changes in patellofemoral mechanics. Therefore, patellofemoral pain can be secondary, in which case the primary problem should be addressed first.

Assessment of the biomechanical factors associated with patellofemoral pain is as follows:

- With the patient standing, the examiner looks for signs of hip flexor tightness, such as increased lordosis or forward-leaning trunk. The orientation of the patellae is observed to determine whether they are facing straight ahead or appear "cross-eyed" (femoral anteversion) or "wall-eyed" (femoral retroversion). The structure of the foot is observed for high arch or flat foot. Finally, the patient attempts a minisquat (i.e., a squat to 20° of knee flexion) on one leg. Angular or rotational movement of the knee or shifting of the trunk during the single-legged minisquat implies weakness of the hip abductor and external rotator muscles.
- With the patient seated, the examiner observes patellar tracking during active knee extension. The patella usually moves laterally just as the knee "screws home" to a fully extended position. If the patella moves medially or the lateral motion begins earlier, abnormal kinetics are present. The knee is palpated for crepitus during active flexion and extension. Coarse crepitus does not correlate with anatomic findings unless the clunking or clicking sensation is persistent and is present at a specific point in the range of motion, in which case it may represent a horizontal plica in the knee or an articular surface irregularity on the back of the patella or on the trochlea. More often, however, the clicks will come and go. Fine crepitus (like sand in the joint) usually implies synovitis. Tenderness to palpation of the patellar or quadriceps tendon is more consistent with tendinitis. Pain produced when the patient extends her knee against the examiner's manual resistance is not useful as a differential finding. However, the absence of this pain is inconsistent with tendinitis.
- With the patient supine, the examiner looks for effusions about the knee. Effusions are uncommon, and their presence implies arthritis or meniscal damage. The patella should be passively mobile—laterally and medially as well as proximally and distally. In the absence of arthritis, restricted motion implies tightness of the retinacular tissues attached to the patella.
- With the patient prone, the examiner measures the range of motion of the hips in internal and external rotation. The range differs between the prone and supine positions because hip flexion and abduction are eliminated when the patient is prone. When internal rotation exceeds external rotation by more than 15°, the patient has femoral anteversion. The converse holds for femoral retroversion. The thigh-foot axis is then measured to assess tibial torsion. Tightness of the quadriceps muscle is checked by gently flexing the knee

to the point of resistance and measuring the distance between the heel and the buttock. This should be no greater than 2 inches. If the buttocks lift, a flexion contracture of the hip is present. If this maneuver produces pain, tendinitis is likely.

- With the patient lying on the unaffected side, the examiner performs an Ober test: The unaffected hip and knee are flexed sufficiently to eliminate lumbar lordosis, and the examiner holds the affected extremity with the knee flexed. The examiner's other hand stabilizes the patient's pelvis. The affected limb is then flexed, abducted, hyperextended, and finally adducted at the hip. If the iliotibial band is tight, the limb cannot be adducted back down to the examining table or opposite limb but will remain passively abducted; in such patients, typically, the distance from the underside of the knee to the surface of the examination table will exceed 2 inches.

Finally, examination of a runner's shoes may reveal abnormal wear, reflecting abnormal dynamic biomechanics that are not apparent during static examination.

Imaging Studies

Imaging studies are not necessary initially in a patient with atraumatic patellofemoral pain. However, if initial treatment is not successful, obtaining radiographic Merchant views (also known as sunrise, sunset, and axial views) of the patellofemoral joint may demonstrate gross misalignment of the patellae in the trochlear groove; narrowing of the patellofemoral joint space; or osteophytes, reflecting patellofemoral arthritis.

TREATMENT

Treatment of patellofemoral pain is tailored to the physical examination findings. A patient who is unstable in the single-legged squat position should undertake exercises to strengthen the hip abductors and external rotators. Patients with excessive lumbar lordosis are given pelvic tilt, core, and gluteal muscle strengthening exercises, as well as hip flexor stretching exercises. Patients with abnormal patellar tracking, a tight iliotibial band, or tight quadriceps muscles should be prescribed stretching exercises. Myofascial release by a physical therapist also may be of use. When excessive pronation or supination is present (either statically or dynamically), orthotic arch supports (over-the-counter or custom) may be beneficial. Bracing, straps, and patellar taping also change patellofemoral compression forces and may relieve pain.[29] Quadriceps muscle strengthening specifically directed at the vastus medialis obliquus is useful and is best done in a closed kinetic chain fashion.[30] Patients should avoid full arc extension exercises and deep squats, both of which markedly increase patellofemoral forces. Finally, if the patient is unable to do quadriceps muscle strengthening exercises because of pain, treatment can start with electrical muscle stimulation.

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Ireland ML, Ott SM: Special concerns of the female athlete. Clin Sports Med 23:281, 2004

2. Loucks AB, Verdun M, Heath EM: Low energy availability, not stress of exercise, alters LH pulsatility in exercising women. J Appl Physiol 84:37, 1998

3. Lane JM, Riley EH, Wirganowicz PZ: Osteoporosis: diagnosis and treatment. J Bone Joint Surg 78A:618,1996

4. Voss LA, Fadale PD, Hulstyn MJ: Exercise-induced loss of bone density in athletes. J Am Acad Orthop Surg 6:349, 1998

5. Jones BH, Harris JM, Vinh TN, et al: Exercise-induced stress fractures and stress reactions of bone: epidemiology, etiology, and classification. Exerc Sport Sci Rev 17:379, 1989

6. Hulkko A, Orava S: Stress fractures in athletes. Int J Sports Med 8:221, 1987

7. Matheson GO, Clement DB, McKenzie DC, et al: Stress fractures in athletes: a study of 320 cases. Am J Sports Med 15:46, 1987

8. Warden SJ: Cyclo-oxygenase–2 inhibitors: beneficial or detrimental for athletes with acute musculoskeletal injuries? Sports Med 35:271, 2005

9. Seidenberg AB, An YH: Is there an inhibitory effect of COX-2 inhibitors on bone healing? Pharmacol Res 50:151, 2004

10. Hill PF, Chatterji S, Chambers D, et al: Stress fracture of the pubic ramus in female recruits. J Bone Joint Surg 78B:383, 1996

11. Egol KA, Koval KJ, Kummer F, et al: Stress fractures of the femoral neck. Clin Orthop Rel Res 348:72, 1998

12. Shin AY, Gillingham BK: Fatigue fractures of the femoral neck in athletes. J Am Acad Orthop Surg 5:293, 1997

13. Hickey GJ, Fricker PA, McDonald WA: Injuries to elite rowers over a 10-yr period. Med Sci Sports Exerc 29:1567, 1997

14. Holden DL, Jackson DW: Stress fractures of the ribs in female rowers. Am J Sports Med 13:342,1985

15. Schenck TJ, Brems JJ: Multidirectional instability of the shoulder: pathophysiology, diagnosis, and management. J Am Acad Orthop Surg 6:65, 1998

16. Brown GA, Tan JL, Kirkley A: The lax shoulder in females: issues, answers, but many more questions. Clin Orthop 372:110, 2000

17. McFarland EG, Campbell G, McDowell J: Posterior shoulder laxity in asymptomatic athletes. Am J Sports Med 24:468, 1996

18. Kibler WB, Livingston B: Closed-chain rehabilitation for upper and lower extremities. J Am Acad Orthop Surg 9: 412, 2001

19. Burkhead WZ, Rockwood CA: Treatment of instability of the shoulder with an exercise program. J Bone Joint Surg 74A:890, 1992

20. Omey ML, Micheli LJ, Gerbino PG: Idiopathic scoliosis and spondylolysis in the female athlete: tips for treatment. Clin Orthop 372:74, 2000

21. Uhorchak JM, Scoville CR, Williams GN, et al: Risk factors associated with noncontact injury of the anterior cruciate ligament: a prospective four-year evaluation of 859 West Point cadets. Am J Sports Med 31:831, 2003

22. Griffin LY, Albohm MS, Arendt EA, et al: Update on ACL injury prevention: theoretical and practical considerations. A review of the Hunt Valley II meeting, February 2005. Am J Sports Med (in press)

23. Hewett TE, Myer GD, Ford KR, et al: Biomechanical measures of neuromuscular control and valgus loading of the knee predict ACL injury risk in female athletes: a prospective study. Am J Sports Med 33:492, 2005

24. Noyes FR, Bassett RW, Grood ES, et al: Arthroscopy in acute traumatic hemarthrosis of the knee: incidence of anterior cruciate tears and other injuries. J Bone Joint Surg Am 62:687, 1980

25. DeHaven KE: Diagnosis of acute knee injuries with hemarthrosis. Am J Sports Med 8:9, 1980

26. Asano H, Muneta T, Ikeda H, et al: Arthroscopic evaluation of the articular cartilage after anterior cruciate ligament reconstruction: a short-term prospective study of 105 patients. Arthroscopy 20:474, 2004

27. Cameron M, Buchgraber A, Passler H, et al: The natural history of the anterior cruciate ligament deficient knee: changes in synovial fluid cytokine and keratan sulfate concentrations. Am J Sports Med 25:751, 1997

28. Fulkerson JP: Patellofemoral pain disorders: evaluation and management. J Am Acad Orthop Surg 2:124, 1994

29. Powers, CM, Ward SR, Chan L, et al: The effect of bracing on patella alignment and patellofemoral joint contact area. Med Sci Sports Exerc 36:1226, 2004

30. Stensdotter AK, Hodges PW, Mellor R, et al: Quadriceps activation in closed and in open kinetic chain exercise. Med Sci Sports Exerc 35:2043, 2003

Acknowledgment

Figure 2 Tom Moore.

XXII FEMALE SEXUALITY: ASSESSING SATISFACTION AND ADDRESSING PROBLEMS

Jennifer Potter, m.d.

Historically, little attention has been devoted to studying female sexual function and dysfunction. A number of factors have contributed to this neglect. Gender bias remains strong.[1] Cultural proscriptions, such as those regarding premarital intercourse, tend to be much more stringent for women. Women's sexual pleasure is often valued less (exemplified in the common statement, "My husband's needs come first") or not at all (demonstrated in its most extreme form by female circumcision, in which sensitive vulvar structures are surgically removed).[2] Evaluation of sexual difficulties in women emphasized psychological origins, whereas biologic factors were relatively ignored.

Scientific investigation of female sexuality has been hampered by numerous methodological challenges. Arriving at a precise definition of female sexual dysfunction, for example, has proved difficult because sexual difficulties reported by women are not discrete, tending to occur together. Specific diagnostic criteria, such as duration of symptoms and degree of distress, continue to be debated. Sexual response is more difficult to evaluate objectively in women than in men, and until recently, few reliable animal models existed.

For all of these reasons, funding to support research on female sexuality has been suboptimal. Nevertheless, knowledge about female sexuality has expanded rapidly during the past decade. Although further study is needed to throroughly address remaining questions and controversies, there are now sufficient data to guide rational evaluation of women who present with sexual concerns.

As in men, sexual problems in women impact quality of life, general functioning, and adaptation to illness. Levels of distress associated with sexual dysfunction range from mild unhappiness, frustration, or a sense of sexual inadequacy to a more pervasive loss of self-esteem that can have profound effects on intimate relationships, as well as functioning in other social and occupational realms.[3] Sexual complaints can be important clues to the presence of an underlying illness, such as depression. Sexual side effects of medications are common and may result in lack of compliance if they are not adequately addressed. For these reasons, it is imperative for clinicians to become adept at addressing patients' sexual concerns. However, many practitioners lack formal training in this area and thus are unsure of what questions to ask and what to do when a sexual problem is identified. This chapter provides a practical approach to the evaluation and management of sexual dysfunction in women.

Epidemiology

In the Pfizer Global Study of Sexual Attitudes and Behaviors, a survey of over 27,500 persons in 29 countries, 63% of women (and 83% of men) 40 to 80 years of age described sex as extremely, very, or moderately important.[4] It is reasonable to assume that younger men and women find sex to be equally if not more important.

Large population-based studies in several countries have examined the prevalence of sexual difficulties in women and men. In all of the studies, a higher percentage of female respondents than male respondents reported experiencing sexual difficulties: 43% versus 31% in the United States,[5] 54% versus 35% in the United Kingdom,[6] 30% versus 27% in Sweden,[7] and 71% versus 46% in Australia.[8] The wide variation in these numbers may reflect differences in the specific questions that were asked, the time intervals that were studied, the demographics of the study populations, and whether or not unpartnered persons were included. The numbers may be inflated in that not all persons who report having sexual difficulties consider themselves to have a sexual dysfunction: only one third to one half of women who report having decreased sexual desire or response believe they have a problem or feel distress for which they would like help.[9] However, even when personal distress and the desire for intervention are factored in, the prevalence of sexual problems in women remains high (10% to 35% of women in these studies).

Some, but not all, studies of sexual function report an association between older age and lower desire; however, when the relationship between low interest and distress over sexual difficulties is taken into account, the age factor largely disappears.[9] Although menopause can be associated with decreased lubrication, vaginal dryness, and discomfort during sexual penetration, pain during sex seems to be significantly more common in younger women than in older ones.[5] In a large random telephone survey done in Australia, which included 8,282 women and 8,510 men 16 to 59 years of age, the most common problem cited by both women and men was lack of sexual interest.[8] Women and men felt equally anxious about their ability to perform. Women were more likely than men to report being unable to achieve orgasm, not finding sex pleasurable, experiencing physical pain during intercourse, and worrying during sex that their body looked unattractive. Few studies have examined the effect of factors such as ethnicity and sexual orientation on rates of sexual satisfaction and sexual difficulties in women.

Female Sexual Function

Female sexual function is far more than a simple biologic response to a stimulus. When a woman experiences sexual difficulty, there is almost never just one cause; a multifactorial etiology is the rule. Similarly, there is usually no quick fix. This may help explain why medications such as sildenafil and other phosphodiesterase (PDE5) inhibitors, which rapidly restore sexual function in many impotent men, are not particularly useful in most women.[10,11] To fully understand the nuances in each case, it is crucial to identify and address all of the factors that contribute to the patient's lack of satisfaction with her sex life. These factors can be readily enumerated if the clinician examines each case in two ways: first, from the perspective of a biopsychosocial model of sexual function and, second, by examining which functional domains (i.e., desire, arousal, orgasm, satisfaction, absence of pain) are affected.

Biopsychosocial model of female sexual function

Normal female sexual function involves successful integration of key factors in psychological, interpersonal, and biologic realms [*see Figure 1*]. For example, women raised in societies in

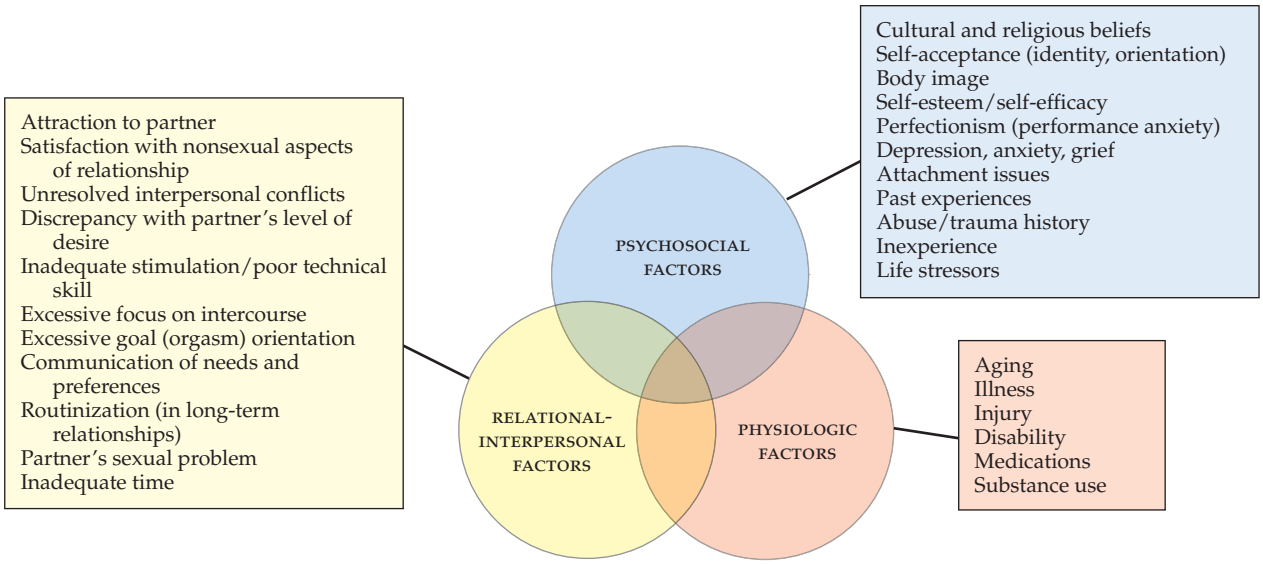


Figure 1 **The biopsychosocial model posits that overlapping physiologic, psychosocial, and relational factors influence sexuality.**

which discussion of sex is taboo are often ignorant about genital anatomy, sexual needs, and techniques. Young women whose families have strong expectations regarding sexual orientation and virginity may feel anxiety, confusion, or shame if their evolving sense of self turns out to be discordant with the values of their families. Finally, gender roles shape power dynamics in relationships and likely affect the degree to which different societies tolerate sexual coercion and violence. A pattern of sexual inhibition and avoidance often follows sexual trauma; disempowered women typically find it difficult to communicate sexual needs to their partner or to advocate for safer sex.

In addition to culture and contextual background, studies show considerable variation with respect to how women rate the importance of sex, the specific sexual practices they choose to participate in, the frequency of sexual activity they feel is optimal, and the intensity and duration of stimulation they need to achieve arousal and orgasm.[12] What would be a satisfying sex life for one woman might seem woefully inadequate to another. Clearly, the only person who can decide she has a sexual problem is the patient herself. Diagnosis of a sexual dysfunction should never be made unless a problem is persistent and causes distress and unless the patient articulates a desire for further evaluation.[13]

MODELS OF HUMAN SEXUAL RESPONSE

Traditional models of male and female sexual response were initially described by Masters and Johnson in the 1960s and embellished later by Helen Kaplan [*see Figure 2*].[14,15] In these linear models, events proceed in an orderly, stepwise manner; desire is a necessary first step; and orgasm is the explicit goal, although it may not always be attained. An important difference between the sexes is that orgasm in the male is followed by a longer refractory period; multiple sequential orgasms are more readily achieved in women.

Basson and colleagues developed a new model of sexual response in an attempt to convey more aptly the complexity of female sexual function.[16] This model is circular and contains multiple feedback loops through which sexual desire and arousal may be intensified or inhibited [*see Figure 3*]. Spontaneous desire is not a necessary factor to enter the cycle; simply feeling emotionally close and engaging in intimate touching can lead to arousal. This more realistic model acknowledges the fact that subjective response (emotional closeness to one's partner during

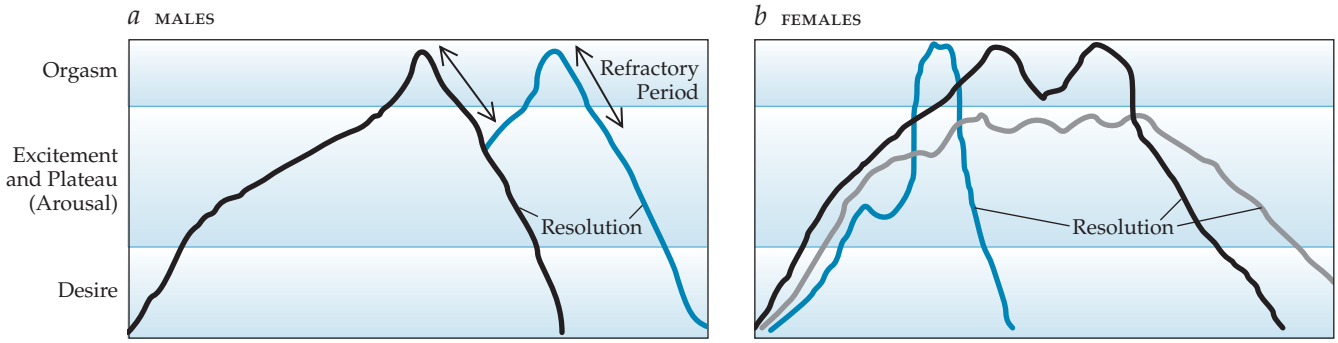


Figure 2 **Traditional models of human sexual response are linear.[14,15] In males (*a*), orgasm is followed by a short refractory period, after which further stimulation (blue line) can again lead to arousal and a second orgasm. Females (*b*) may experience a single-orgasm event (blue line), multiple orgasms (black line), or a plateau without orgasm (gray line).**

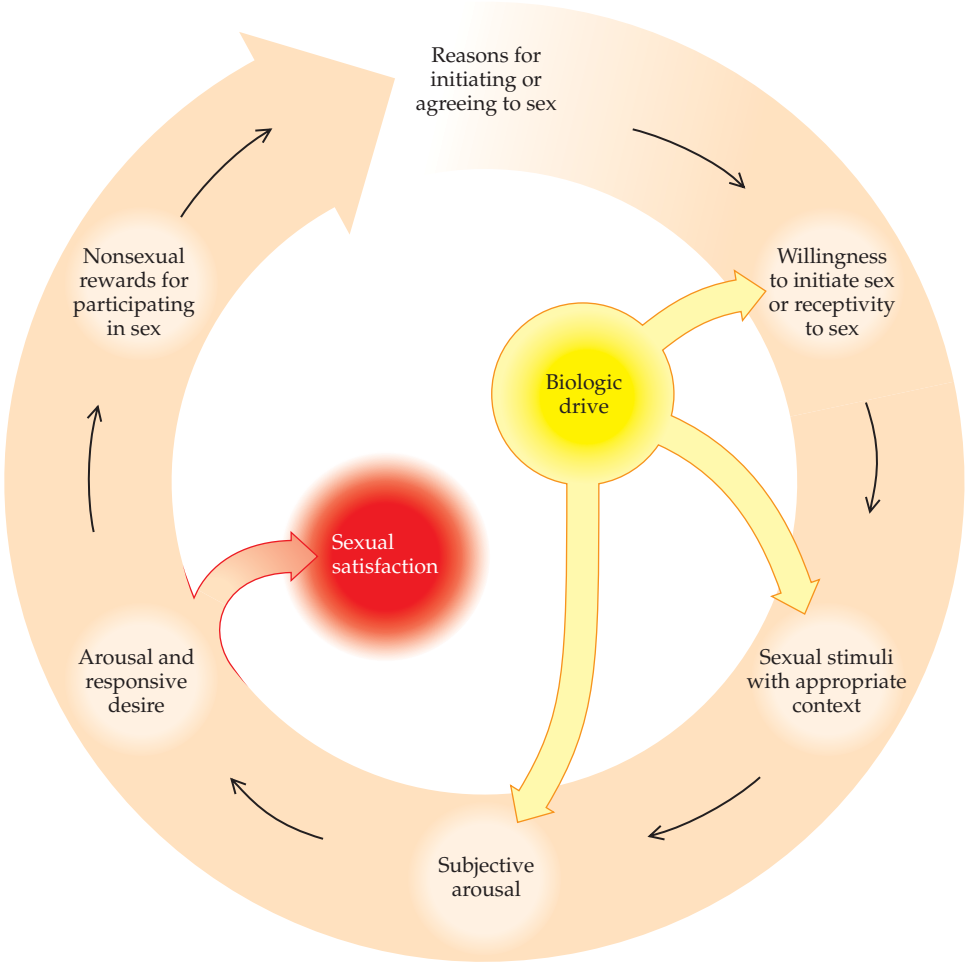


Figure 3 **Female sexual response is most accurately described by a circular model containing multiple feedback loops.[139] At several points in the cycle, sexual response can be either inhibited or stimulated by cognitive influences (e.g., negative thoughts can interrupt the response; erotica or fantasy can enhance the response). Sexual satisfaction can occur with or without orgasm. Nonsexual rewards may include emotional intimacy, feelings of well-being, or avoiding negative consequences for not participating in sex.**

sexual activity) can be as important for some women as physical response (attainment of orgasm).

STRUCTURE AND EROGENICITY OF THE FEMALE GENITALIA

Erogenous external genital structures in women include the clitoris, labia majora and minora, vaginal introitus, and anus. Of these, the densely innervated and highly sensitive clitoris is exceedingly important: arousal proceeds most readily and orgasms are strongest when the clitoris is stimulated.[17] Indeed, an excessive focus on penile-vaginal intercourse without clitoral stimulation is a chief cause of anorgasmia in women.

Internal genital structures involved in female sexual response include the vagina, cervix, and uterus. Of these, stimulation of the anterior vagina seems to be most pleasurable. Popular literature describes a so-called G-spot, or Grafenberg spot, an allegedly highly erogenous area located along the lower third of the anterior vaginal wall.[18] Although few studies provide objective evidence supporting the existence of the G-spot, stimulation of this area is purported to result in a kidney bean–size area of swelling and to be associated with high levels of arousal and powerful orgasms. Similarly, the existence of female ejaculation has been postulated because of the observation that fluid is released during orgasm in some women. Whether this fluid is simply urine or might represent a true ejaculate, such as emission of secretions from female paraurethral glands (analogous to the male prostate), is not clear from available data.[18]

Despite its proximity to the major pelvic or paracervical ganglion, the cervix itself is a relatively insensitive structure.[19] There are numerous theoretical reasons why removal of the uterus might be expected to interfere with sexual response, including anatomic changes (shortening of the vaginal vault and scar formation in the vaginal cuff), surgical damage to pelvic nerves and blood vessels, secondary hormonal changes,[20] and the possibility that uterine contraction itself contributes substantively to orgasmic pleasure. Although studies of sexual function after hysterectomy are fraught with methodological problems, the preponderance of evidence suggests that detrimental effects of either total or supracervical hysterectomy are rare and that preoperative sexual function is the most important predictor of postoperative sexual satisfaction.[21]

BIOCHEMISTRY AND METABOLISM OF SEX STEROIDS

Sex steroids and their receptors play a key role in the maturation, maintenance, and function of the tissues that are involved in female sexual response. The modulating effects of estrogen and androgen help explain the changes in sexual function that

are associated with changes in the levels of these hormones, whether the hormones are endogenous (as with changes that occur during puberty, within each menstrual cycle, during pregnancy and post partum, during lactation and after weaning, and with surgical or natural menopause) or exogenous (i.e., given as contraceptives or fertility drugs or for hormone replacement). Other hormones, including progesterone, oxytocin, and prolactin, probably also have important effects.

Synthesis of both estrogen and androgen takes place in the ovaries, adrenal glands, and peripheral tissues. Estradiol predominates before menopause. After menopause, estrogen is made extragonadally via aromatization of ovarian and adrenal androgens, and estrone predominates. The ovaries produce androstenedione, testosterone, and a small amount of dehydroepiandrosterone (DHEA). The adrenals produce androstenedione and DHEA sulfate (DHEAS). Both DHEA and DHEAS are converted in brain, bone, and adipose tissue to androstenedione or testosterone, which can then be either converted by 5-reductase to dihydrotestosterone (DHT) or aromatized to estrone or estradiol. Only free testosterone and DHT can bind to receptors and are therefore biologically active.

Before menopause, 25% of a woman's testosterone is produced in the ovaries, 25% in the adrenals, and the remaining 50% in peripheral tissues. Circulating androgen levels peak during the third decade of life, then decrease slowly with age because of reduced adrenal synthesis.[22] Diurnal and menstrual cycle–linked changes in testosterone and androstenedione also occur, with levels highest in the morning before 10 A.M. and during the middle third of the menstrual cycle.[22] Stromal cells of the ovary continue to make androgen precursors and testosterone after natural menopause; the amount of bioavailable testosterone may actually rise slightly during the first few years of the menopausal transition because of lower serum estrogen and, therefore, lower sex hormone–binding globulin (SHBG) levels.[23] Surgical menopause presents a dramatically different situation: bilateral salpingo-oophorectomy can lead to a sudden 50% decline in circulating levels of androstenedione and testosterone, with the effect being most pronounced in younger women.[24] Compromised ovarian function from premature ovarian failure, chemotherapy, pelvic radiotherapy, or administration of gonadotropin-releasing hormone agonists is also associated with low circulating androgen levels, as are adrenal insufficiency or adrenalectomy, glucocorticoid treatment (which suppresses pituitary adrenocorticotropic hormone secretion), and administration of exogenous estrogen (which increases SHBG levels and suppresses pituitary luteinizing hormone secretion).[25]

NEUROBIOLOGY OF FEMALE SEXUAL RESPONSE

Studies of the anatomy, physiology, and pathophysiology of female sexual function and dysfunction are limited. Much of what is known, or inferred, is extrapolated from animal studies, primarily in rodents, and drawn by analogy from studies in men.[19]

Normal female sexual function involves intact neural, vascular, and muscular circuitry; complex interactions between multiple neurotransmitter systems; and critical modulating influences from the endocrine system. Numerous neurotransmitters, bioactive substances, and sex steroids appear to be involved. These include, but are probably not limited to, dopamine, norepinephrine, serotonin, acetylcholine, nitric oxide, vasoactive intestinal peptide (VIP), prostaglandin E_1 (PGE_1), estrogen, testosterone, progesterone, oxytocin, and prolactin. Emerging evidence suggests that α-melanocortin–stimulating hormone (α-MSH) may also play an important role.[26] Some of these substances act centrally, in the brain and spinal cord, whereas others have peripheral sites of action (e.g., arteries, peripheral nerves, and the pubococcygeal muscles) [*see Figure 4*]. Some neurotransmitters have effects at both central and peripheral locations, but the nature of these effects (i.e., excitatory versus inhibitory) is not always the same at these two locations. For example, central norepinephrine activity appears to have stimulatory effects on female sexual response, whereas peripheral effects tend to be largely inhibitory. This has important implications for the design of pharmacologic therapies for female sexual problems. Because peripheral responses (e.g., vasocongestion) and observable behaviors (e.g., lordosis as a measure of proceptivity and receptivity in female rats) are easier to measure than central events, research to date has concentrated more on peripheral mechanisms than on central ones. The neurophysiology of each phase of the female sexual-response cycle—desire, arousal, and orgasm—can be considered separately.

Neurobiology of Desire (Libido)

Desire can be defined as a mental state created by external and internal stimuli that induce an urge to participate in sexual activity.[19] Manifestations of increased desire include sexual thoughts or fantasies and the motivation to initiate, or the willingness to be receptive to, sexual activity. It is useful to consider three non–mutually exclusive aspects of sexual desire: drive, motivation, and beliefs/values.

Sex drive has biologic roots and appears to be mediated by the excitatory action of dopamine and modulating effects of sex steroids in the mesolimbic system of the brain.[27] Higher-order cortical processes likely provide excitatory and inhibitory influences on lower cortical centers. Dopamine enhances sex drive and the wish to continue sexual activity once sexual stimulation has been initiated; these effects are inhibited by serotonin. Estrogen seems to have a small permissive effect, whereas testosterone and progesterone appear to influence initiation of sexual activity and receptivity to partner approach, respectively.[27] Serum prolactin concentrations increase after orgasm in men and women, and chronic hyperprolactinemia is associated with reduced libido in both sexes.[28] These observations suggest that prolactin may be a regulatory factor that signals central nervous system centers involved in the initiation or control of sexual behavior. Although sexual pathways in female humans need to be further elucidated, α-MSH agonists have been found to selectively stimulate solicitational behaviors in female rats and enhance desire and stimulate erection in men with psychogenic erectile dysfunction.[26]

Sexual motivation is fueled by anticipation of a risk or reward associated with initiation of or participation in sexual activity. Incentives are diverse and may include the following: feeling close to one's partner, giving or experiencing sexual pleasure, relieving tension, becoming pregnant, or exchanging sexual favors for tangible and intangible gifts. Conversely, the expectation, based on negative past experiences, of pain or injury resulting from sexual activity serves as a major disincentive.

Beliefs/values include unique personality and cultural factors that exert either excitatory or inhibitory influences on the wish to engage in sexual activity. Motivation and beliefs/values probably affect libido via excitatory and inhibitory cortical influences on lower cortical structures; further study is needed to elucidate specific pathways.

Knowledge of the basic neurophysiology of libido helps predict the effects of various exogenous substances. As expected, libido seems to be increased by dopamine enhancers, such as am-

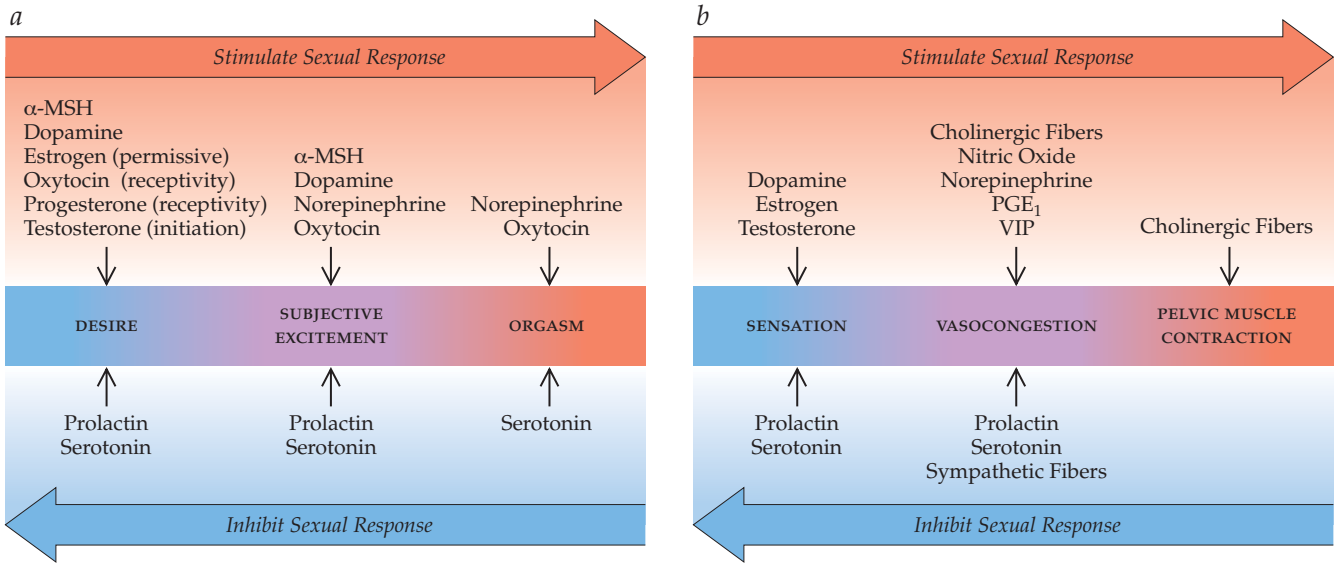


Figure 4 **The neurobiology of female sexual response includes both central and peripheral events.[27] Although these may be visualized as occurring sequentially, in reality, the physiologic and biochemical components of arousal may occur simultaneously and to some extent independently. Whether sexual desire, arousal, and orgasm are enhanced or inhibited depends on the coordinated activity of numerous bioactive substances, as well as the presence or absence of key environmental influences.**

(*a*) Central events are critical in the emergence of sexual desire and subjective excitement (arousal) and play an important role in orgasm. Increased levels of serotonin can result in inhibition of this response through reduction of dopamine and norepinephrine levels in the brain.

(*b*) Peripheral events include genital sensation, vasocongestion, and muscle contraction. The baseline state of the genital vasculature is tonic vasoconstriction, which is maintained by outflow from sympathetic fibers. In response to stimulation, release of vascular mediators (e.g., nitric oxide, prostaglandin E_1 [PGE_1], and vasoactive intestinal peptide [VIP]) promotes dilation of vessels and tissue vasocongestion. Both vasocongestion and pelvic muscle contractions are mediated by the release of acetylcholine from cholinergic fibers. Increasing feedback along afferent pathways to the spinal cord culminates in a spinal reflex arc that produces pelvic muscle contractions; transmission of these impulses cephalad to pleasure centers in the brain produces the sensation of orgasm.

phetamines and norepinephrine-dopamine reuptake inhibitors (NDRIs) (e.g., bupropion). Conversely, substances that reduce dopamine activity diminish libido: examples include D_2 dopamine receptor blockers and selective serotonin reuptake inhibitors (SSRIs). Hyperprolactinemia (caused by pituitary macroadenomas, lactation, or use of so-called typical antipsychotic drugs), intoxication with CNS depressants such as alcohol, and use of antiandrogens also reduce libido.

Fortunately, decreased libido does not have to relegate a woman to an unsatisfactory sex life. Many women whose spontaneous level of desire is reduced or absent retain the capacity for arousal and even orgasm by strong cognitive motivation to engage in sexual activity and skillful sexual-stimulation technique.

Neurobiology of Arousal

Sexual arousal includes both subjective excitement (i.e., awareness of, comfort with, and appreciation of erogenous stimulation) and objectively measurable signs (both nongenital and genital) of physiologic arousal.[27] It is important to distinguish between subjective and objective arousal, because studies show that women often experience measurable physiologic arousal in the absence of a subjective sense of excitement or pleasure.[13] Indeed, the feeling of sexual arousal in women seems to be heavily dependent on cognitive processing of stimulus meaning and content,[29] as opposed to being related purely to peripheral vasocongestive and neuromuscular events.

Subjective arousal is a central event that appears to be mediated by excitatory effects of both dopamine and norepinephrine; oxytocin, which is elevated throughout the follicular phase of the menstrual cycle, may also play a stimulatory role.[30,31] The effects of dopamine and norepinephrine are inhibited by serotonin; prolactin also has a negative effect on subjective excitement.

Genital sexual arousal appears to be mediated by spinal reflexes.[31] One such spinal reflex is the bulbocavernosus reflex, in which stimulation of light-touch pudendal sensory fibers activates pudendal motor neurons, resulting in contraction of striated perineal muscles. Activation of this reflex contributes to development of the orgasmic platform and strengthens urinary continence during arousal. The bulbocavernosus reflex can be evaluated during physical examination to check both sensory and motor function of the pudendal nerve. Another spinal reflex involves vaginal and clitoral cavernosal autonomic nerve stimulation, which results in clitoral, labial, and vaginal engorgement.[19] The efferent and afferent arms of these spinal reflexes have been relatively well defined in men; pathways in women are probably similar. Efferent facilitatory parasympathetic output arises in the sacral parasympathetic nucleus and is conveyed to the vagina and clitoris by the pelvic nerve. Efferent inhibitory sympathetic output arises in the dorsal gray commissure and the intermediolateral cell column at the thoracolumbar level and travels to the genitalia via the hypogastric nerve and the paravertebral sympathetic chain. The activity of the spinal nuclei is controlled both by descending projections from multiple supraspinal sites and by sensory afferents from the genitalia that are conveyed by the pudendal, hypogastric, pelvic, and vagus nerves. Serotoninergic projections from the brain to the spinal cord ap-

pear to inhibit the induction of genital arousal by afferent peripheral stimulation; noradrenergic projections also exist, but their functional significance is not fully understood.[31]

In the basal state, clitoral corporal and vaginal smooth muscles maintain contractile tone under noradrenergic mediation.[19] During sexual stimulation, neurogenic and endothelial release of nitric oxide, produced by the action of nitric oxide synthase on L-arginine, triggers a rise in cyclic guanosine monophosphate (cGMP) in postjunctional cells, causing calcium influx to vascular smooth muscle in the clitoris and vagina, vasodilatation, and tissue engorgement.[19] The result is extrusion of the glans and enhanced sensitivity. The vaginal epithelium continuously reabsorbs sodium from submucosal capillary plasma transudate. During sexual stimulation, release of nitric oxide and VIP induce increased capillary inflow; sodium reabsorption becomes overwhelmed; and 3 to 5 ml of vaginal transudate is produced, enhancing lubrication during sexual activity.[19] In addition to nitric oxide and VIP, a number of other vasoactive substances, including PGE_1, have been found in the clitoris and vagina; all deserve further study.[32] Cholinergic fibers innervate vascular smooth muscle in the vagina and may help mediate vaginal engorgement during arousal; however, the chief role of acetylcholine in the periphery is to increase pelvic striated muscle contractility.[33] Estrogen plays a critical role in maintaining vaginal tissue integrity and has local regulatory effects on nerve transmission (thereby increasing genital sensation) and mediators of vasocongestion (thereby enhancing genital blood flow and lubrication).[34] Preliminary evidence suggests that androgens may also affect genital sensation, genital hemodynamics, and mucin production (a minor component of sexual lubrication).[34] Peripherally, serotonin acts via 5-hydroxytryptamine-2A ($5\text{-}HT_{2A}$) receptors to reduce genital sensation and to inhibit nitric oxide synthase.[35] Exogenous substances that increase genital arousal include estrogens; pPDE5 inhibitors; PGE_1 agonists; and alpha blockers. PDE5 inhibitors potentiate the effects of the nitric oxide pathway by blocking the degradation of cGMP.[36] Substances that reduce genital arousal include antiestrogens, SSRIs, and anticholinergic medications.

Neurobiology of Orgasm

Arousal and orgasm can be viewed on a continuum. Arousal is necessary for buildup of the orgasmic platform; there can be no orgasm if there is no arousal. The experience of orgasm comprises both central events (experience of pleasure) and peripheral events (measurable pelvic muscle contraction). The central control of orgasm is poorly understood, but it is postulated that norepinephrine has stimulatory effects (via $alpha_1$- and $alpha_2$-adrenergic receptors) and that serotonin and prolactin have inhibitory effects, as they do on the production of subjective excitement.

Genital events that take place during orgasm are likely the result of a spinal reflex, as is the ejaculatory response in men. When afferent impulses from the genitals to the spinal centers reach threshold, efferent impulses result in stereotyped contractions (typically, five to 12 in number) of the vagina, uterus, and pelvic floor muscles at 0.8-second intervals; simultaneous messages are sent to cortical pleasure centers. Acetylcholine is the primary mediator of pelvic floor muscle contraction. Oxytocin may also play a role peripherally, in that plasma oxytocin levels rise substantially during orgasm and are correlated with the intensity of orgasmic contractions. Serotonin has an inhibitory effect on orgasm in the periphery, via stimulation of $5\text{-}HT_{2A}$ receptors. No definitive role for either estrogen or testosterone in orgasm has been established.

The neurobiology of orgasm helps explain the differential effects of several antidepressants [*see Table 1*]. Orgasm is inhibited by substances that reduce central levels of norepinephrine or increase levels of serotonin (e.g., SSRIs); anticholinergic medications also have negative effects. However, antidepressants that enhance the central effects of norepinephrine and dopamine, including NDRIs such as bupropion, as well as compounds that block $5\text{-}HT_{2A}$ receptors, seem to be less likely to affect sexual function adversely. Antidepressants that block $5\text{-}HT_{2A}$ receptors include mirtazapine, trazodone, and nefazodone. Mirtazapine is a noradrenergic and specific serotoninergic antidepressant with postsynaptic $alpha_2$-antagonist, $5\text{-}HT_{2A}$, $5\text{-}HT_{2C}$, and $5\text{-}HT_3$ receptor antagonism; trazodone and nefazodone are serotonin antagonist/reuptake inhibitors, which act by potent blockade of $5\text{-}HT_{2A}$ receptors combined with less potent inhibition of 5-HT reuptake. Some agents have different effects at different doses. For example, venlafaxine inhibits sexual function at low doses because of 5-HT inhibition; has less of an effect on sexual function at moderate to high doses because of norepinephrine reuptake inhibition; and has the least effect on sexual function at highest doses, at which it inhibits dopamine reuptake.

Sexual Behavior

Adolescence and young adulthood are periods of intense exploration of sexual identity and orientation and of experimentation with sexual behavior. According to the 2003 National Youth Behavior Risk Survey, 45% of young women have had penile-vaginal intercourse by the end of high school, and 4% have had intercourse before the age of 13.[37] The prevalence of other sexual behaviors, including mutual masturbation, oral sex, and anal sex, is also high. Of sexually active young women, 14% have had four or more sexual partners by age 19.

Sexual health is not a given during this time of self-discovery. Shame, depression, and increased rates of substance use are common in some groups of teens, particularly sexual-minority women (i.e., those who are lesbian, bisexual, or transgender), whose sense of self may conflict strongly with societal and family expectations. The majority of sexually active young women do not consistently engage in safe sexual practices. Only 63% of sexually active young women report using condoms, and only 17% report having been on birth control pills during their last sexual intercourse.[37] Moreover, 25% of teens report having been under the influence of alcohol or drugs at the time of intercourse.[37] Although the incidence of prostitution is difficult to determine, up to 2% of women report a history of engaging in oral, anal, or vaginal sex in exchange for money or drugs.[38] Moreover, young women are the most frequent targets for rape, whether by an acquaintance or a stranger. Not surprisingly, therefore, rates of unintended pregnancy and sexually transmitted infection (STI) remain high. The pregnancy rate in young women 15 to 19 years of age is 84 cases per 1,000[39]; each year, one in four teenagers in the United States contracts an STI,[40] and over their lifetime, half of all persons will contract an STI.[41] It is important to address these issues, because poor self-esteem, substance use, and depression have profound effects on sexual satisfaction and overall health, as do events such as unplanned pregnancy, abortion, development of STIs, and sexual trauma. Misconceptions and concerns must be addressed, or patients may be reluctant to use contraceptives and

Table 1 Medications with Sexual Side Effects[140, 141]

Drug Class	*Desire*	*Arousal*	*Orgasm*	*Vaginal Dryness*
Amphetamines	↑	?	?	—
Antiandrogens				
Cimetidine	↓	?	—	—
Spironolactone	↓	?		
Anticholinergics	—	↓	↓	↑
Anticonvulsants	↓	?	?	—
Antidepressants				
Selective serotonin reuptake inhibitors	↓	↓	↓	↑
Tricyclics	↓	↓	↓	↑
Antiestrogens	?	↓	↓	↑
Antihistamines	—	↓	—	↑
Antihypertensives				
Alpha blockers	↓	↑	?	—
Beta blockers	↓	↓		
Bupropion	↑	—	↑	—
Oral contraceptives (depends on androgenicity)	↑/↓	—	—	—
Sedatives (dose-dependent)				
Benzodiazepines	↑/↓	↓	↓	—
Alcohol	↑/↓	↓	↓	
Statins	↓	—	—	—
Typical antipsychotics	↓	—	—	—

safer-sex methods. Examples of common beliefs and misconceptions include the following: "I can't get an STI because I'm on the pill;" "If someone I am dating has an STI, I would know it;" and "If I ask my partner to use a condom, he will be suspicious or worried about my sexual history."

Misconceptions on the part of the clinician may interfere with the assessment of sexual risk. For example, it is unwise to assume that a particular patient is not sexually active (e.g., because she is too young, too old, or too ill); that the patient (or her partner) is exclusively heterosexual or participates solely in penile-vaginal intercourse; or that a person (or her partner) who is in a committed relationship is consistently monogamous. Same-sex sexual activity is quite common: in the 2002 National Survey of Family Growth (NSFG),[42] 7% of men and 11% of women 25 to 44 years of age reported same-sex activity at some point in their lives. Sexual activity does not always correlate with sexual orientation: in the Minnesota Adolescent Health Survey, self-identified bisexual or lesbian teens 12 to 19 years of age were as likely as their heterosexual peers to have had penile-vaginal intercourse and reported twice as high a prevalence of pregnancy.[43] Men and women engage frequently in sexual activities other than penile-vaginal intercourse: in the NSFG, 90% of male and 88% of female participants reported a history of oral sex with the opposite sex, and 40% of men and 35% of women reported a history of anal sex with the opposite sex.[42] Sexual activity is common outside the context of committed relationships. For example, 21% of men and 13% of women report having had sex with someone other than their spouse while married.[38] Clinicians who wish to provide pertinent risk-reduction counseling must consider all of these facts and remember that sexual behavior is fluid. It is crucial to ask all patients about their sexual behaviors and to ask repeatedly over time, because behaviors may change.

Many men and women are sexually active from extreme youth through old age, so both sexes stand to benefit from repeated risk-assessment and risk-reduction counseling throughout their lives. It may, in fact, be especially important to discuss sexual health with older people, who, because of the loss of a long-term partner, may be reentering the dating world after many years in a monogamous relationship. Many of these individuals have little knowledge of safer-sex methods and minimal or no experience having frank discussions about sexual safety with prospective partners. As a result of societal restrictions, women often feel particularly uncomfortable initiating these kinds of conversations.

Sexual pleasure tends to be enhanced during young to middle adulthood, as a result of experience, greater confidence, and, in many cases, greater intimacy in the context of committed relationships. However, women in this phase of life may develop sexual concerns related to pregnancy, lactation, parenting, infertility, and pregnancy loss.[44] Sexual function during pregnancy is affected by a woman's (and her partner's) emotional reactions to the pregnancy, body image/physical comfort, and hormonal changes. Women who feel excited about their pregnancy, close to their partner, and happy about being able to dispense with contraception often experience enhanced enjoyment. Conversely, women who feel anxious or who experience nausea and fatigue during the first trimester may be more inhibited. Desire and responsiveness tend to be enhanced during the second

trimester, though weight gain, body-image issues, or withdrawal of a woman's sexual partner can have detrimental effects. Some women find that desire and ease of orgasm increase during the last trimester, whereas others avoid sex because of physical discomfort or concerns about injuring the fetus or causing preterm labor.

In the absence of pregnancy complications, women and their partners should be reassured that sexual activity is safe throughout pregnancy. Creative positioning can be suggested to maximize comfort.

After the birth of a child, many couples experience changes in their sex life. A combination of factors, including initial perineal/pelvic discomfort, disrupted sleep and exhaustion, anxiety about being a new parent, hyperprolactinemia associated with breast-feeding, feelings of unattractiveness, fear of becoming pregnant again, or postpartum depression, can all have inhibiting effects. Over the short term, a period of reduced sexual activity post partum is beneficial, in that it allows healing of delivery trauma and adjustment to parenthood; after a few months, most couples establish a new sexual equilibrium. If sexual intimacy is not eventually restored and if the couple is distressed by the loss, it is appropriate to pursue an evaluation.

Infertility and pregnancy loss can also have profound effects on a couple's sex life, through the combination of hormonal influences and intense feelings (e.g., anger, grief, guilt) about the situation. Provision of emotional support, referral for counseling, and psychopharmacologic intervention, as appropriate, are critical.

Sexual function remains a critical ingredient of a satisfying life in older women. In the Modern Maturity Sexuality Study of 1,384 persons performed by the American Association of Retired Persons (AARP), a satisfying sexual relationship was reported as being important to the quality of life of 66% of women 45 to 59 years of age, 48% of those 60 to 74 years of age, and 44% of those 75 years of age or older.[45] In a study of 1,292 men and women 60 years of age or older that was performed by the National Council on Aging, 51% of women in their 60s, 30% of those in their 70s, and 18% of those 80 years of age or older had engaged in sexual activity at least once a month during the past year.[45] For heterosexual women, some of the reduction in sexual activity with age is related to the availability of a male partner; in the AARP study, only 21% of women 75 years of age or older had a current partner, compared with 78% of women 45 to 59 years of age and 53% of women 60 to 74 years of age.

With age, sexual desire normally shows a gradual decline. Contrary to popular belief, menopause, in and of itself, is not necessarily associated with a precipitous drop in libido. In the Harris Interactive Study, which surveyed 580 menopausal women, 45% of the women reported a decrease in sexual desire after menopause, whereas 37% reported no change and 10% actually noted an increase.[45] On the other hand, menopause results in clear changes in arousal: genital blood flow, vaginal lubrication, and genital sensation all decrease. Orgasmic capacity is maintained as women age, although typically, clitoral stimulation needs to be more direct, more intense, and of longer duration. Decreased muscle tension in the pelvic floor and decreased uterine contractions during climax may diminish the intensity of orgasm. Despite these changes, satisfaction tends to be preserved. In the Association of Reproductive Health Professionals Sexual Activity Survey of 1,000 persons in the United States, 52% of women 50 to 59 years of age, 42% of women 60 to 69 years of age, and 49% of women 70 years of age or older were as satisfied or more satisfied with their current sex life as compared with their sex life when they were younger.[45] There seems to be a difference between the so-called genital prime and the sexual prime. The former represents biologic responsiveness; the latter reflects positive effects of the intimacy that develops in many long-term relationships, the knowledge about one's body that accrues with experience, and the greater ability to ask for and accept pleasure that often accompanies maturity.

Importantly, some women do experience a distressing diminution in sexual satisfaction associated with aging, menopause, or both. Hormonal factors are only one aspect of this decline. Other contributing factors include poor body image and self-esteem; unrealistic expectations; relationship difficulties; lifestyle issues and stress; the onset of acute or chronic illnesses; use of medications with sexual side effects; having a partner with sexual dysfunction; or, often, the absence of a partner. Negative feelings about growing older, including lack of acceptance of physical changes, lead some women to worry that their partner will no longer find them attractive. Some older women may feel that sex is only for the young—that they shouldn't want to have sex any longer. Other women may expect sex to be just as it was during their 20s and 30s and may therefore have difficulty accepting and adapting to the changes associated with growing older. Although long-term relationships often result in increased trust, respect, and sexual satisfaction, some couples find that over time, sex becomes predictable and even boring.

Events that disrupt the equilibrium that has been reached in a long-term relationship, such as a man's increased interest in intercourse after beginning a PDE5 inhibitor, can be challenging for a female partner if she was happy with the status quo. Lifestyle changes that occur during midlife and beyond can upset decades-long patterns in a couple's life. Examples include children leaving home, retirement, financial stresses, aging parents, and concern about one's own or one's partner's health. Fortunately, sexual satisfaction can be maintained and even enhanced in older couples with attention to communication, making time for intimacy, and using a broad range of sexual techniques. In acute illness, a reduction in sexual function can be adaptive, allowing a woman to focus her energy on healing and survival. Chronic illness and disabilities are often associated with changes in sexual function. However, sexual activity continues to be very important to many affected women. A period of adjustment may be needed to allow her and her partner to adjust to a changed body that responds differently. Subsequently, a satisfying sex life can be achieved in many circumstances with adaptations such as timing sexual activity when the energy level is high, using creative positioning to maximize comfort, and learning creative ways of providing pleasurable stimulation.

Classification of Sexual Disorders

Classification of female sexual disorders is important for guiding research, for legitimizing clinical discussion of this topic, and for facilitating the development of diagnosis codes so that clinicians are more likely to be paid for managing these disorders. However, because female sexual function is nonlinear, multifaceted, and subject to individual interpretation, delineating a reliable classification of female sexual problems has proved difficult. Two classification schemes are in common clinical use: those found in the American Psychiatric Association's *Diagnostic and Statistical Manual, Fourth Edition, Text Revision* (DSM-IV-TR)[46]

and the International Statistical Classification of Disease and Related Health Problems, 10th Revision (ICD-10).[47] Both have been criticized because they are based on the sexual response cycle of men and are overly focused on penile-vaginal intercourse; they ignore the fact that changes in sexual function can be adaptive or normative at certain junctures in a woman's life; they overemphasize the medical aspects of the problem by downplaying psychological and relational factors; they categorize problems narrowly, whereas multifactorial etiologies are actually the rule; and they ignore the enormous variation that exists between individual women.[13]

The American Foundation for Urologic Disease (AFUD) convened a multidisciplinary group of experts to review and revise definitions of female sexual problems in 1998; initial recommendations were published in 2000,[48] and the definitions continue to be debated and updated. Overall, the AFUD classification contains more appropriate diagnostic descriptors for women because it attends to context and the presence of distress, is less focused on physical symptoms and intercourse, and distinguishes between subjective and objective events. The AFUD classification divides female sexual problems into the following major categories: sexual interest and desire disorder, sexual-aversion disorder, sexual-arousal disorders, orgasmic disorders, and sexual-pain disorders [*see Table 2*].[13] Unfortunately, the less appropriate DSM-IV and ICD-10 diagnostic codes are still needed to bill for clinical work with patients.

Diagnosis

SEXUAL AND RELATIONSHIP HISTORY

A general sexual assessment can be accomplished relatively quickly. Initial questions should be open-ended and nonjudgmental and should convey a willingness to discuss sexual issues [*see Table 3*]. Follow-up questions should address specific sexual activities and the risk of unintended pregnancy, STIs, and sexual abuse [*see Table 3*].

When a woman articulates concerns about her sex life, further questioning is appropriate. Useful questions include the following:

1. Do you notice a change in your interest in sex?
2. Do you have trouble becoming aroused or sufficiently lubricated?
3. Are you able to reach orgasm?
4. Do you experience pain or discomfort during sex?

If a problem is identified, further questions should be asked about its severity (global versus situational) and chronicity (primary versus secondary), as well as about the context in which it has developed (e.g., relating to psychiatric or medical illness, substance use, relationship problems, life stressors, or medication use) and the extent to which it causes the patient distress. When the woman has a partner, specific questions should be asked about communication, technical skill, and sexual repertoire. Both individuals in the relationship should be included in the evaluation whenever possible.

The History in Women with Sexual-Desire Problems

When decreased desire is primary and generalized, the problem often traces back to a highly sex-negative upbringing or an early traumatic sexual event; when the decreased desire is acquired or situational, it is more likely to be caused by relationship problems, stress, poor physical health, hormonal changes, and exogenous substances. Questions to ask women with low sexual desire include the following:

1. When did you notice a change in your level of desire?
2. What do you think is responsible for the change?
3. Do you ever feel motivated to have sex? If so, what motivates you?
4. Do you experience spontaneous sexual thoughts or fantasies?

Table 2 Classification of Female Sexual Dysfunction[13]

Desire Disorders

Sexual interest and desire disorder

Absent or diminished feelings of sexual interest or desire, absent sexual thoughts or fantasies, and a lack of responsive desire. Motivations (defined as reasons/incentives) for attempting to become sexually aroused are scarce or absent. The lack of interest is considered to be beyond a normative lessening with life cycle and relationship duration

Sexual-aversion disorder

Extreme anxiety, disgust, or both at the anticipation of having, or the attempt to have, any sexual activity

Arousal Disorders

Subjective sexual-arousal disorder

Absent or markedly diminished feelings of sexual arousal (sexual excitement and sexual pleasure) from any type of sexual stimulation, although vaginal lubrication or other signs of physical response still occur

Genital sexual-arousal disorder

Absent or impaired genital sexual arousal; self-report may include minimal vulvar swelling or vaginal lubrication from any type of sexual stimulation and reduced sexual sensations from caressing genitalia; subjective sexual excitement still occurs from nongenital sexual stimuli

Combined subjective and genital sexual-arousal disorder

Absence of, or markedly diminished feelings of, sexual arousal (sexual excitement and sexual pleasure) from any type of sexual stimulation, as well as complaints of absent or impaired genital sexual arousal (vulvar swelling, reduced lubrication)

Persistent sexual-arousal disorder

Spontaneous, intrusive, and unwanted genital arousal (e.g., tingling, throbbing, pulsating) in the absence of sexual interest and desire; any awareness of subjective arousal is typically but not invariably unpleasant; the arousal is unrelieved by one or more orgasms and the feelings of arousal persist for hours or days

Orgasmic Disorder

Self-report of high sexual arousal or excitement but lack of orgasm, markedly diminished intensity of orgasmic sensations, or marked delay of orgasm from any type of stimulation

Sexual-Pain Disorders

Dyspareunia

Persistent or recurrent pain with attempted or complete vaginal entry and/or penile vaginal intercourse

Vaginismus

Persistent difficulties in allowing vaginal entry of a penis, finger, or any object, despite the woman's expressed wish to do so; there is variable involuntary pelvic muscle contraction, (phobic) avoidance, and anticipation/fear/experience of pain; structural or other physical abnormalities must be ruled out or addressed

Noncoital sexual-pain disorder

Recurrent or persistent genital pain induced by noncoital sexual stimulation

Table 3 Helpful Questions to Ask for the Sexual History

Initial Questions

- Are you currently involved in a sexual relationship? Have you ever been?
- How old were you when you first had sex?
- How many partners have you had in the past month? Six months? Over your lifetime?
- Have your partners been men, women, or both?
- How are things going in your relationship(s)?
- How satisfied are you with the sexual aspect of your relationship(s)?
- Is there anything about your (or your partner's) sexual activity that you would like to change?

Follow-up Questions

- How often do you have sex?
- What types of sexual activity do you engage in (e.g., hand on genital, mouth on genital, penis in vagina, penis in anus, use of sex toys)?
- Do you think any of your sexual activities put you at risk? Do you take any precautions to reduce your risk?
- Do you feel comfortable discussing safer sex with your partner or partners?
- Has anyone ever coerced you or forced you to have sex against your will?
- Are you currently in a relationship where you feel physically, sexually, or emotionally threatened?

5. Are you "turned on" by erotic descriptions in books or sex scenes in movies?
6. How do you feel about your body?
7. Do you ever masturbate?
8. Are you attracted to your partner (or partners)?
9. Do you find other men or women attractive?

When the decrease in desire seems to stem from low self-esteem, interpersonal issues, or cultural prohibitions, it is appropriate to ask additional questions, such as the following:

1. What are your beliefs and values about sex in general?
2. (For women who are heterosexual) Were you raised with strong beliefs about premarital sex and virginity?
3. Were you raised with strong negative beliefs about specific sexual activities, such as oral sex, self- and partner masturbation, vaginal penetrative sex, and anal sex?
4. (For women who are not exclusively heterosexual or who may be questioning their sexual orientation) Were you raised with strong beliefs about same-sex attraction and same-sex relationships?
5. Do you feel comfortable with your gender identity and sexual orientation?
6. Has anything ever happened that made you feel guilty or bad about sex?

Questions should also be asked about life stressors, as well as about medical conditions, medications, and other substances that are known to be associated with diminished libido. A history of medical conditions (e.g., thromboembolism, active liver disease, hormone-responsive cancers, eating disorders, and seizures) that might contraindicate potential treatment options (e.g., estrogen, testosterone, and bupropion) should also be sought.

Sexual-aversion disorder is usually associated with low sexual desire and sometimes with vaginismus or dyspareunia. A history of sexual trauma is common, and patients with such a history often have extensive negative, unexpressed feelings about their relationships. Questions should therefore focus on psychological and relationship issues; ascertaining a history of abuse; and ruling out coexistent mental health conditions, including anxiety and obsessive-compulsive disorders. It can be helpful to find out whether a woman's aversion is generalized or specific. Generalized aversion refers to all partners or any type of sexual activity, whereas specific aversion refers to only certain partners or to anticipating or attempting a particular sexual act.

The History in Women with Sexual-Arousal Problems

Absent or impaired sexual arousal The etiology of decreased sexual arousal includes either a decrease in central excitatory activity or an increase in central inhibitory activity, as well as medical conditions that compromise genital blood flow and pelvic innervation. Diabetes mellitus is associated with decreased genital vasocongestion and arousal in women[49]; atherosclerosis caused by other risk factors should also be considered. Multiple sclerosis and other neurologic conditions can be associated with neurogenic impairment.[50] Obstetric trauma, including episiotomy, traumatic delivery, and cesarean section, is associated with sexual dysfunction, though painful sex is usually more common than reduced arousal. Decreased arousal and pain also occur frequently in women who have had pelvic surgery (e.g., hysterectomy, with or without oophorectomy, and procedures to correct urinary incontinence or pelvic organ prolapse) and women who have sustained pelvic trauma.[51] The incidence and severity of sexual problems after hysterectomy depend on preoperative sexual function and psychosocial state, the degree to which the surgery alleviates the symptoms for which it was performed, and the extent of the surgical procedure.[52] Surgical technique (i.e., whether blood vessels and nerves are spared) is probably a key factor. Clinical experience suggests that women who undergo pelvic evisceration for treatment of malignancy, whether gynecologic or nongynecologic, frequently develop arousal problems. Contrary to popular opinion, there is little sexual advantage to performing a supracervical rather than a total hysterectomy.[53] Persistent perineal pressure (e.g., from prolonged bicycle riding) may cause injury in women.

Questions to ask patients with reduced arousal include the following:

1. Do you feel aroused by sexual thoughts, fantasies, reading a sexy passage in a book, or seeing a sexy scene in a movie?
2. Do you feel aroused by having different parts of your body touched (by yourself or your partner)?
3. What parts of your body do you most like having touched?
4. What kinds of stimulation (hand-genital, mouth-genital, genital-genital, sex toy–genital) are most pleasurable?
5. Do you feel comfortable talking with your partner about what you like?
6. Do you lubricate enough to make sex comfortable? Have you tried using lubricants, and if so, which ones? Do you think they are effective?
7. Has there been a change in your level of arousal? If so, when did things change, and why do you think they changed?

The history should also include questions about menopausal status, medical illnesses, medications, obstetric injury, pelvic surgery, and pelvic trauma.

Persistent sexual-arousal disorder The etiology of persistent sexual-arousal disorder is unknown, because the entity is rarely seen and little studied. Anecdotal reports have described patients with both vascular (aneurysmal dilatation of genital blood vessels) and neurologic (persistent clitoral paresthesias from bicycle-seat trauma) abnormalities. Therefore, it is probably useful to ask questions about the onset and duration of the condition, as well as the presence of medical conditions, injuries, or surgeries that might have triggered a neuropathic response.

The History in Women with Orgasm Problems

The etiology of orgasm problems varies according to whether the lack of orgasm is lifelong and generalized or is situational and intermittent. Primary (lifelong, generalized) orgasmic disorder rarely has a physical cause; it is usually the result of a highly sex-negative upbringing, unpleasant or abusive sexual experiences, and ineffective sexual technique on the part of either the patient during self-masturbation or her partner during sex. One of the most common causes of failure to achieve orgasm is ignorance about the need for clitoral stimulation and an excessive focus on penile-vaginal intercourse as an optimal technique.

Secondary (situational/intermittent) orgasmic disorder can be caused by any inhibiting influence that dampens sexual arousal on a temporary basis. Common causes include psychological issues, relationship problems, an unskilled sexual partner, stress, illness, and medications. It is important to address the issue of expectations relating to orgasm, because women sometimes express satisfaction with a sexual encounter that does not include orgasm. On the other hand, it is quite common for a woman's partner to feel inadequate if she does not have an orgasm. Secondary orgasmic disorder can be permanent if severe illness or injury significantly impairs the neural circuitry necessary for normal arousal and orgasm.

Questions for women with primary orgasmic disorder should focus on beliefs and values; prior sexual experiences; and knowledge about female sexual anatomy, function, and technique. Women with secondary orgasmic disorder should be queried about self- and relational issues, partner skill, physical problems that interfere with sexual arousal, and medications that interfere with arousal and orgasm. In addition, all women presenting with sexual problems should be asked about urinary incontinence.

Sexual complaints are common in women with urinary incontinence and pelvic organ prolapse.[54,55] The prevalence of urinary incontinence that occurs during sexual activity (i.e., sexual incontinence) ranges from 2% in randomly selected population samples to as high as 56% in women in clinical samples.[56] Understandably, sexual incontinence can be associated with avoidance of sexual activity and sexual dissatisfaction; research using standard definitions and measures is needed to determine the frequency of sexual problems in this context. Limited data suggest that stress incontinence is most often associated with urinary leakage at the time of vaginal penetration, whereas women with urge incontinence are more likely to experience leakage during orgasm.[57]

The History in Women with Sexual-Pain Disorders

Traditional definitions of sexual-pain disorders (i.e., dyspareunia, vaginismus) focused on vaginal pain caused by insertion of a penis—that is, pain during penile-vaginal intercourse. Dyspareunia is a general term that does not specify etiology, whereas vaginismus refers to pain on insertion that is caused by involuntary spasm of the perineal and levator muscles. Revisions have broadened these descriptions to include vaginal pain on insertion of other objects; in addition, a new category, noncoital sexual-pain disorder, has been added [*see Table 2*]. This expanded view is appropriate, in that it embraces the existence of noncoital sexual activities and recognizes that sexual stimulation of external genital structures can also trigger a painful response rather than a pleasurable one.

The chief causes of noncoital sexual pain are vulvodynia and vestibulitis. The International Society for the Study of Vulvovaginal Disease has defined vulvodynia as vulvar discomfort, most often described as burning pain, occurring in the absence of relevant visible findings or a specific, clinically identifiable, neurologic disorder.[58] Vulvodynia is further characterized by whether the pain is generalized or localized and whether it occurs with or without provocation, or both. Vestibulitis is a term used to describe vulvodynia that is localized to the region of the vestibule. Vestibulitis may involve sexual pain associated with light touch around the vestibule, as well as pain on attempted penetration.

It is important to note that different pain disorders can coexist and that, at times, one may trigger another. For example, chronic yeast infection may trigger vestibulitis, which, in turn, may lead to avoidance of penetrative activity because of anticipated pain, with secondary vaginismus.

The history should include assessment of the onset, character, intensity (graded on a scale of 1 through 10), location, and duration of the pain. Clues to etiology can also be obtained by asking about apparent triggers of the pain (including timing during the menstrual cycle), as well as ameliorating factors (e.g., position changes, lubricants, analgesics, and pain-modulating agents). The clinician should ask whether the pain is present invariably (i.e., during each sexual encounter with every partner) or only intermittently. A primary, lifelong pain that is complete and generalized during penile-vaginal intercourse is usually the result of congenital anomalies, psychosocial issues, or vulvodynia/vestibulitis, whereas secondary, situational pain is more likely to have an acquired physical cause. It is important to ask about a history of sexual trauma, which is commonly associated with pain during sexual activity, especially vaginismus. Evaluation should always assess the impact of the problem on the patient and her relationship with her partner, as well as the willingness of both parties to participate in counseling, as appropriate.

The location of the pain (i.e., superficial, present on penetration, or deep within the pelvis) is the best clue to its etiology.[59] Superficial pain is associated with light touch around the clitoris, labia, and vestibule, and it tends to be caused by anatomic abnormalities, dermatitis, dermatosis, infection, and vulvodynia/vestibulitis. Pain on penetration occurs with entry of any object (finger, sex toy, penis, and, in premenopausal women, tampons) into the vagina, and it is usually the result of a structural abnormality (e.g., episiotomy scar, hymenal ring, vaginal stenosis, or vaginal septation/duplication), vaginitis (e.g., bacterial vaginosis, candidiasis, trichomoniasis, or chemical induced), vaginal atrophy, or vaginismus. Other causes of vaginal pain include inadequate lubrication associated with inhibited arousal caused by psychological or relationship factors, lack of estrogen, medications (e.g., antihistamines or anticholinergics), and diseases that cause dryness (e.g., Sjögren syndrome); urinary tract

problems (e.g., urethritis and interstitial cystitis); and vaginal shortening (congenital or postsurgical). Pain associated with deep penetration or thrusting is sometimes seen in women with retroverted uteri; position changes are typically helpful. Other causes of deep pain include pelvic adhesions; infections (e.g., chronic pelvic inflammatory disease); endometriosis; benign pelvic processes (e.g., ovarian cyst, uterine enlargement from adenomyosis or fibroids, interstitial cystitis, irritable bowel disease, and inflammatory bowel disease); pelvic organ prolapse; malignancy in the pelvis (e.g., cervical, ovarian, endometrial, urinary tract, or gastrointestinal tract); and a history of pelvic surgery for either benign or malignant conditions.

The timing of the pain is also informative. For example, pelvic pain during orgasm may be related to uterine contractions, whereas pain after sexual activity may be from pelvic congestion. Vulvovaginal itching, swelling, and erythema after intercourse may be caused by seminal plasma hypersensitivity; this diagnosis can be made by absence of symptoms with condom use and by positive skin testing with a pooled sample of seminal fluid.[59]

Sexual Questionnaires

Having patients complete a written questionnaire can sometimes help define key issues. Several multidimensional self-report instruments on female sexual dysfunction have been developed.[60] These include the 25-item Derogatis Interview for Sexual Functioning (DISF/DISF-SR),[61] the 22-item Brief Index of Sexual Functioning for Women (BSFI-W),[62] and the 19-item Female Sexual Function Index (FSFI).[63] The FSFI, which is the most commonly used of these questionnaires, is available free of charge on the Internet (http://www.FSFIquestionnaire.com). Each of these questionnaires is well standardized, relatively unobtrusive, and inexpensive and takes 15 to 20 minutes to administer and score; moreover, normative values for both clinical and nonclinical populations are available. There is reason to believe that use of questionnaires may result in a more accurate history, because women may be more willing to answer sensitive questions privately than during a face-to-face conversation.

PHYSICAL EXAMINATION AND LABORATORY TESTING

If a woman wishes to pursue an evaluation, the clinician should perform a comprehensive but pertinent physical examination, which should be adapted to her specific presenting complaints as well as to anticipated treatments. It is reasonable to palpate the thyroid in all women, because both hypothyroidism and hyperthyroidism can affect sexual function adversely. The breasts and liver should be examined when hormonal, especially systemic hormonal, treatments are a consideration, because of known toxicities. A detailed genital exam should be performed in all women, especially in those with decreased arousal, anorgasmia, or sexual pain [*see Table 4*]. The pelvic exam can be especially challenging in some women, especially those with a history of abuse or a negative experience during a past clinical encounter. Simple measures can help reduce anxiety and discomfort and maximize a woman's sense of control, including the presence of a support person or chaperone; a contract to stop the exam if requested by the patient; the use of a small speculum; and adequate lubrication.[59] In severe cases, administration of a short-acting benzodiazepine before the exam may also be useful.

The integrity of pelvic blood vessels and nerves should be assessed in women with arousal problems. Blood pressure and peripheral pulses provide an assessment of overall cardiovascular function. Several methods have been devised to measure pelvic blood flow objectively, including vaginal photoplethysmography,[64] Doppler ultrasound,[65] application of oxygen electrodes to measure oxygen tension,[66] and magnetic resonance imaging[67]; however, this equipment is costly and useful primarily in research. Sensation can be evaluated by light touch with a cotton-tipped applicator.[59] In this test, the patient is asked to report whether or not she is able to ascertain light touch of the swab to the labia majora and minora, clitoris, and clock-face locations around the vestibule. This test is also useful in determining the location and intensity of pain (graded from 1 through 10) in women with suspected vulvodynia/vestibulitis.

The biothesiometer, a device originally developed to test sensation in the feet of diabetic patients with peripheral neuropathy, can also be used to test sensation in the labia and around or on the clitoris. A vibrating probe is first applied to a neutral region of the body (e.g., the patient's finger), and the degree of vibration is carefully adjusted to determine the lowest level of stimulation that can be appreciated by the patient. Subsequent measurements of labial and clitoral sensation can then be obtained. Biothesiometry is a relatively inexpensive way to confirm the presence of reduced genital sensitivity, establish the degree of neurologic impairment, and provide objective measurements that can be tracked during treatment; devices are available that cost approximately $500.

Pelvic floor muscle tone can be assessed by grading the degree of vaginal and anal sphincter contraction that can be generated by a patient during digital vaginal and rectal examination (i.e., absent, mild, moderate, or strong) or by using a peritoneometer. The peritoneometer is a handheld device with a vaginal probe that measures vaginal muscle contraction. Rarely, pelvic ultrasonography and laparoscopy are necessary in the investigation of deep pelvic pain.

Laboratory testing should be tailored to individual circumstances, with careful consideration given to cost and evidence of clinical usefulness. Assessment of cardiovascular risk factors (e.g., elevated lipid or glucose levels) are useful when hemodynamic compromise is suspected. Measurement of pituitary hormone levels (i.e., follicle-stimulating hormone, thyroid-stimulating hormone, and prolactin) may be helpful when evaluating postpartum, lactating, and perimenopausal women and when manifestations of thyroid disease are present. Measurement of baseline lipid levels, liver function tests, and a complete blood count may be performed in women for whom androgen supplementation is anticipated, because excess androgen replacement can be associated with adverse changes in the lipid profile, abnormal liver function, and polycythemia. An appropriate evaluation should be performed when vulvar or vaginal lesions, vaginal discharge, or cervicitis is discovered, including a biopsy of suspicious lesions, vaginal pH testing, a wet mount, a potassium hydroxide preparation, and selected tests for STIs. Routine vaginal cultures are not useful.

Recommendations regarding serum androgen levels are more difficult to make, because there are no precise definitions of androgen deficiency, normal serum androgen levels for women of different ages are poorly characterized, the accuracy of available testosterone assays in women is questionable,[68] and not all studies show a correlation between low serum testosterone levels and low sexual desire.[69] One large study, however, has demonstrated a significant association between self-reported low sexual function and a low serum DHEAS level (below 10th percentile for age).[69] Further studies are needed to determine

Table 4 The Genital Examination in Women with Sexual Difficulties

Anatomic Feature or Examination Maneuver	*Possible Sources of Sexual Difficulty*
Clitoris	Phimosis, adhesions, female genital circumcision
Urethra	Erythema, carbuncle, prolapse
Labia majora and minora	Atrophy, lesions
Introitus	Erythema, scarring, stricture
Vagina	Atrophy, erythema, discharge
Valsalva maneuver	Uterine prolapse, urinary incontinence
Bimanual examination	Masses, pain
Vaginal and rectal sphincter contraction	Poor tone, prolapse
Bulbocavernosus reflex	Pudendal neuropathy

whether serum DHEAS measurement has any utility in identifying a subgroup of women with low desire who are most likely to benefit from treatment. Part of the difficulty is that circulating androgen levels are not good indicators of intracellular androgen activity, because extensive intracellular metabolism of androgens occurs without the release of active metabolites into the pericellular compartment.[70] Moreover, studies of the efficacy of androgen supplementation show inconsistent results. Some of this inconsistency may be explained by differences in the responsiveness of individual women to androgen supplementation as a result of variations in the amount or activity of 5-reductase and aromatase, as well as differences in the androgen receptor response.[71] At present, the major reasons to check serum androgen levels are to exclude women with generous levels from treatment and to monitor serum levels during treatment to prevent overreplacement. Routine screening with a clinical breast examination, mammography, and Papanicolaou (Pap) testing should follow usual guidelines.

Management

GENERAL TREATMENT RECOMMENDATIONS

Effective treatment for women with sexual problems must address each area of distress (i.e., psychological, relational, and physiologic) and must intervene in each affected domain (i.e., desire, arousal, orgasm, overall satisfaction, and comfort or absence of pain). The goal of treatment is to increase pleasure and satisfaction, rather than to achieve a perfect genital response. Certain interventions are useful in the treatment of all female sexual difficulties; these include basic education and information about lifestyle measures that enhance sexual satisfaction and, in many circumstances, referral for counseling.

Education and Counseling

As expected from the substantial placebo effects observed during trials of various interventions for female sexual problems, simply inviting women to talk about their sexual concerns can be extraordinarily therapeutic. Discussion should include basic education about normal female sexual anatomy and function, and permission should be given to explore self-pleasuring. Many women have unrealistic expectations about sex and feel that they should be immediately and reliably aroused and responsive, despite normal distractions or age-related changes. Education alone can be extremely beneficial. For example, a woman who has lost some of her sex drive but remains motivated to be close to her partner can be reassured that willingness to participate in intimate touching and skillful sexual technique can still lead to arousal and orgasm. Similarly, women with arousal problems often benefit from the knowledge that increased intensity and duration of stimulation, such as that which can be provided by a vibrator, can restore their ability to experience pleasurable genital sensations. Finally, because an emphasis on intercourse and ignorance regarding the importance of clitoral stimulation are common in women with primary anorgasmia, many of these patients are able to achieve their first climax after learning about anatomy and practicing appropriate sexual technique.

Simple lifestyle measures can go a long way toward enhancing interest and ensuring comfort. For example, in one study of the effects of free testosterone levels and other psychosocial variables on sexual function in women undergoing natural menopause, the only variable significantly associated with satisfaction was exercise.[72] Couples can be encouraged to maximize opportunities for closeness and intimacy by setting aside specific times for this purpose (i.e., making "dates"). Thought should be given toward creating environments conducive to intimacy; considerations include time of day, music, lighting, and temperature and children being away or asleep. If a woman and her partner are open-minded, the use of diversified sexual techniques, as well as experimentation with different sex positions, venues, sex toys, fantasy, and erotica, can be suggested as means to enhance excitement and eliminate routine. Strategies that maximize comfort can also be recommended, including relaxing in a warm bath before sex (for women with vaginismus), using effective lubricants (for women with atrophy or dryness), trying position changes (for women with pelvic discomfort from fibroids or organ prolapse), and strategic dosing of mild analgesics before engaging in sexual activity (for women with arthritis). Many different lubricants are available, including water-, oil-, and silicone-based products. Selection should be based on the need for safer sex protection, palatability, and glide. Oil-based and silicone-based lubricants last longest and produce the best glide during penetrative sex. Vegetable oils are inexpensive, reasonably physiologic, and taste good—an important consideration when oral sex is part of the sexual repertoire. However, oil-based products degrade latex; only water-based lubricants should be used with latex male condoms. The manufacturer's recommendations should be closely reviewed when choosing lubricants for use with sex toys. Many books and Internet resources now exist that provide high-quality information about sex [*see Sidebar,* Selected Resources for Female Sexuality].

Referral for individual, couples, or sex therapy can be beneficial when psychological problems or couples issues are present.[73] The purpose of sex therapy is to help women identify thoughts or behaviors that interfere with sexual enjoyment, become more aware of and comfortable with sexual feelings, learn what kind of stimulation is most pleasurable, and increase communication of sexual needs to their partners. Various techniques may be used, including sexual fantasy and skills training, masturbation exercises, the use of erotica and vibrators, and practicing communication using role playing. Cognitive restructuring techniques can be useful for women who experience distracting thoughts that inhibit arousal. Women who have experienced

Selected Resources for Female Sexuality

National Organizations

American Association of Sex Educators, Counselors, and Therapists (AASECT)
http://www.aasect.org

International Academy of Compounding Pharmacies
http://www.iacprx.org/index.html

International Society for the Study of Women's Sexual Health (ISSWSH)
http://www.isswsh.org

National Vulvodynia Organization
http://www.nva.org

North American Menopause Society (NAMS)
http://www.menopause.org

Web Sites

Sexualhealth.com
http://sexualhealth.com

The Society for Human Sexuality
http://www.sexuality.org

Sexuality Information and Education Council of the United States
http://www.siecus.org

San Francisco Sex Information
http://www.sfsi.org
Sex information via e-mail or telephone.

Scarleteen
http://www.scarleteen.com
Youth-oriented sex information.

Kinsey Institute for Research in Sex, Gender, and Reproduction
http://www.indiana.edu/~kinsey

Sex-Positive Sex Shops

Blowfish
www.blowfish.com
Online and catalog only.

Good Vibrations, San Francisco
www.goodvibes.com

Grand Opening, Brookline, Massachusetts
www.grandopening.com

Early2 Bed, Chicago
www.early2bed.com

Ruby's Pearl, Iowa City

The Smitten Kitten, Twin Cities, Minnesota
www.smittenkittenonline.com

Toys in Babeland, Seattle and New York
www.babeland.com

Vaginal Dilators

Soul Source Enterprises
www.soulsourceenterprises.com.
Colorful, slightly flexible silicone (set of 4, $150; 2 additional sizes available at extra cost).

Syracuse Medical Devices (315-449-0657)
Hard white plastic (set of four, $50).

Vaginal Weights

LadyCare Vaginal Exercise Weights
http://www.sportstek.net/qlc4.html

Books

GENERAL VULVOVAGINAL HEALTH

Stewart EG: The V Book: A Doctor's Guide to Complete Vulvovaginal Health. Bantam Books, 2002

"HOW TO" GUIDES

Winks C, Semans A: The Good Vibrations Guide to Sex: The Most Complete Sex Manual Ever Written, 3rd ed. Cleis Press, Inc., San Francisco, 2002

Inkeles G: The Art of Sensual Massage, 2nd ed. Arcata Arts, Arcata, California, 2000

GENERAL REFERENCES ABOUT SEXUAL PROBLEMS

Reichman J: I'm Not in the Mood: What Every Woman Should Know about Improving Her Libido. William Morrow, New York, 1998

Berman J, Berman L: For Women Only, Revised Edition: A Revolutionary Guide to Reclaiming Your Sex Life. Henry Holt & Co, New York, 2001

Glazer HI: The Vulvodynia Survival Guide: How to Overcome Painful Vaginal Symptoms and Enjoy an Active Lifestyle. New Harbinger Publications, Oakland, California, 2002

MASTURBATION/BECOMING ORGASMIC

Heiman J, LoPiccolo J, Palladini D: Becoming Orgasmic: A Sexual and Personal Growth Program for Women. Simon & Schuster, New York, 1987

Barbach L: For Yourself: The Fulfillment of Female Sexuality. New American Library, New York, 2000

Dodson D: Sex for One: The Joy of Self-Loving. Three Rivers Press, New York, 1996

Allison S: Tickle Your Fancy: A Woman's Guide to Sexual Self-Pleasure. Tickle Kitty Press, San Francisco, 2001

FOR SPECIAL POPULATIONS OF WOMEN

Blank H: Big Big Love: A Sourcebook on Sex for People of Size and Those Who Love Them. Greenery Press, Oakland, California, 2000

Paget L: Hot Mamas: The Ultimate Guide to Staying Sexy Throughout Your Pregnancy and the Months Beyond. Gotham Publishing, New York, 2004

Winks C, Semans A: Sexy Mamas: Keeping Your Sex Life Alive While Raising Kids. Inner Ocean Publishing Inc., Maui, Hawaii, 2004

Caster W, Bussel RK, May J: The Lesbian Sex Book, 2nd ed: A Guide for Women Who Love Women. Alyson Publications, New York, 2003

Barbach LG: The Pause: Positive Approaches to Perimenopause and Menopause. First Flume Printing, New York, 2000

Haines S: The Survivor's Guide to Sex: How to Have an Empowered Sex Life after Child Sexual Abuse. Cleis Press, San Francisco, 1999

Silverburg C, Kaufman M, Odette F: The Ultimate Guide to Sex and Disability: For All of Us Who Live with Disabilities, Chronic Pain and Illness. Cleis Press, San Francisco, 2003

Zilbergeld, B: Better Than Ever: Love and Sex at Midlife. Crown House Publishing, Norwalk, Connecticut, 2005

sexual trauma may be helped by a combination of cognitive-behavioral techniques, eye-movement desensitization and reprocessing (EMDR), and insight-oriented psychotherapy. For women whose instinct is to avoid sexual activities because of either prior abuse or pain, the use of nondemand, sensual touching exercises (so-called sensate focus), which focus on the pleasurable sensations associated with nongenital touch, can be extremely helpful. Typically, these exercises begin with the least threatening kinds of touch, such as back rubs or foot massage, and progress over time to nude, full-body caressing, including the genitals. Sexual-aversion disorder is best approached using gradual exposure and habituation exercises, which are also effective in the man-

agement of other phobias. A useful resource for sex therapists is the American Association of Sexuality Educators, Counselors, and Therapists (www.aasect.org). It should be noted that the use of surrogate partners, a practice that was utilized in the past, is fraught with ethical and legal issues and is no longer an accepted sex-therapy modality.

INDIVIDUALIZED EVALUATION AND MANAGEMENT

In addition to education, lifestyle changes, and referral for counseling, a number of individualized treatment options can be considered for selected women who present with specific symptoms and findings. Therapeutic measures in these patients are tailored to the domain of sexual response affected—desire, arousal, or orgasm, alone or in combination—and the particular cause.

Sexual-Desire Disorders

Postmenopausal women Reduced sexual desire after menopause—especially surgical menopause—may result from estrogen and androgen deficiency. Medical and psychiatric illnesses that could be contributing to the low libido should be ruled out. If any medications need to be prescribed for contributing illnesses, agents that do not themselves have a negative impact on desire should be selected, if possible. For some women with low sex drive, cognitive motivation to engage in sex, coupled with skillful technique, can compensate for their lowered libido. For women who are not satisfied with this strategy, it is reasonable to consider additional management options, including supplemental estrogen or androgen or both. Laboratory evaluation might include checking thyroid-stimulating hormone and baseline serum androgen levels, which, if androgen supplementation is elected, can subsequently be followed to avoid excessive dosing.

Most studies of systemic estrogen replacement in postmenopausal women demonstrate an improvement in sexual function in all phases of the sexual-response cycle.[74] It is not clear to what extent the effect on desire is mediated indirectly through amelioration of vaginal discomfort caused by atrophic changes or through alleviation of sleep deprivation caused by vasomotor symptoms. Lubricants and local application of topical estrogen[75] can help counteract atrophy. Systemic estrogen replacement may be indicated for vasomotor symptoms, in the absence of contraindications; however, it should be remembered that oral estrogen increases SHBG levels and, thus, may theoretically reduce rather than increase desire. Therefore, the optimal vehicle for systemic estrogen replacement may be a transdermal preparation, which avoids the hepatic first-pass effect. In patients with only minor vasomotor symptoms, topical estrogen applied to the genitalia might be a safer long-term strategy.

Studies of testosterone therapy in the form of oral preparations,[76,77] intramuscular injections,[78,79] and subcutaneous implants[80-82] have demonstrated significant improvement in sexual desire in postmenopausal women, and there is increasing use of testosterone for this purpose. Many of these studies, however, included small sample sizes, were of short duration, achieved supraphysiologic levels of serum testosterone (at which risks are more likely to outweigh benefits), or exclusively enrolled surgically menopausal women, in whom the dramatic decline in postoperative androgen levels is commonly associated with a significant loss of sexual desire. In general, it has been more difficult to demonstrate efficacy of androgen replacement in naturally menopausal women or to show a benefit in either surgically or naturally menopausal women with restoration of testosterone levels within the physiologic range—perhaps because of the large placebo effect seen in these trials, which could obscure the beneficial effects of testosterone. Two trials have demonstrated benefits of testosterone replacement to physiologic levels in surgically menopausal women using a 300 g transdermal patch.[83,84] A trial in naturally menopausal women has shown similar benefits.[85]

It has been suggested that the beneficial effect of testosterone results from an increase in bioavailable estrogen. Studies of hormone replacement therapy, however, show that the combination of estrogen and testosterone improves sexual function to a greater degree than estrogen alone,[86,87] and trials have not documented an increase in serum estrogen levels after the addition of testosterone.[83,84] Whether coadministration of estrogen is necessary for testosterone to achieve its beneficial effect is unknown, because all studies in postmenopausal women to date have included systemic estrogen replacement.

The steroidal agent tibolone appears to have beneficial effects on sexual function in postmenopausal women.[88-90] Tibolone is not available in the United States, however, and the efficacy and safety of this agent in the management of sexual problems in postmenopausal women need to be assessed in randomized, placebo-controlled trials.

It is important to identify women in whom low desire is the result of androgen deficiency and who are therefore most likely to benefit from androgen replacement. Consideration of androgen replacement is most appropriate for symptomatic, surgically menopausal women, especially those who underwent bilateral salpingo-oophorectomy at a young age. Clinicians should be circumspect about prescribing androgens, however, for a number of reasons. First, numerous factors impact sexual desire, and androgen deficiency probably plays a significant role in only a minority of women with low libido. This is consistent with the finding that most women with low self-reported sexual desire and sexual satisfaction do not have correspondingly low serum levels of total or free testosterone, androstenedione, or DHEAS.[91] Second, it is not clear that the benefits of androgen replacement outweigh the risks. The benefits found with replacement of androgen to physiologic levels appear to be quite modest. For example, Buster and colleagues found that surgically menopausal women receiving the testosterone patch engaged in one additional sexual episode per 2.5 weeks, compared with one additional episode per 5.5 weeks in those who received placebo.[84] Short-term adverse events associated with use of the 300 g patch include application-site reactions (usually mild) and a nonsignificant increase in androgenic side effects (e.g., acne, alopecia, facial hair growth, voice deepening).[84] Excessive androgen replacement can be associated with significant adverse effects, such as hyperlipidemia and hepatotoxicity. No meaningful changes in lipid profiles, carbohydrate metabolism, liver function tests, hematology parameters, or clotting measures were seen in recent studies, but the duration of these studies was only 24 weeks; concerns remain regarding the possibility of longer-term adverse effects, including possible increases in the incidence of atherogenesis and breast cancer.

Androgen therapy is contraindicated in women with a history of breast cancer, severe liver disease, or deep vein thrombosis. In general, androgen supplementation should be avoided in premenopausal women (see below), because there is little evidence of their efficacy, and adverse effects can be especially severe.

Because of the lack of long-term data regarding efficacy and safety, few countries have approved the use of androgen re-

placement for the treatment of sexual dysfunction in women. Consequently, the majority of available androgen preparations are formulated for use in men. A variety of androgen preparations are used for off-label treatment of low libido in women. Oral formulations include 1.25 mg of esterified estrogens plus 2.5 mg methyltestosterone (Estratest), 0.625 mg esterified estrogens plus 1.25 mg methyltestosterone (Estratest HS), and DHEA. The two Estratest preparations are approved by the Food and Drug Administration for the management of hot flashes. Nevertheless, these preparations, which were formulated before the publication of the Women's Health Initiative trials, contain estrogen doses that are higher than optimal; in addition, periodic progestin withdrawal in women who have not had hysterectomies is indicated to prevent endometrial hyperplasia.

DHEA is available as a nonregulated, over-the-counter dietary supplement. A double-blind, placebo-controlled, randomized crossover trial in 24 women with adrenal insufficiency showed that DHEA at a dosage of 50 mg/day for 4 months significantly increased sexual interest and the level of satisfaction with sex, as well as overall well-being and mood.[92] In women without adrenal insufficiency, however, the value of DHEA for treating low sexual desire remains questionable.[93,94] Clinical reports describe using a starting dose of 25 to 50 mg/day of DHEA and titrating the dose upward in 25 mg increments until a maximum dose of 100 mg is achieved.[95]

Topical androgen is commercially available as a 1% testosterone gel, which comes in a packet or a pump dispenser. Doses for men (one 5 g packet or four pumps) are excessive for women, who should start at approximately one tenth the male dose. Daily topical application of a pea-sized dollop to nongenital skin seems to be a reasonable starting point. Testosterone can also be formulated in a 1% or 2% ointment or cream by compounding pharmacies. Because of its central permissive effect on libido, systemic estrogen along with androgen may be needed to achieve an improvement in libido, although studies have not yet confirmed this need.

The optimal follow-up schedule, laboratory monitoring, and duration for androgen therapy have not been defined. Because of concerns about supraphysiologic dosing and long-term safety, however, it is advisable to monitor women at regular intervals. In trials of the 300 g testosterone patch, free, total, and bioavailable testosterone levels; free and total estradiol levels; estrone levels; and SHBG levels were measured at baseline, 12 weeks, and 24 weeks.[83,84] Androgen levels rose at 12 weeks and remained fairly stable over the remainder of the study period. Estrogen levels did not rise appreciably in either study. Maximal effects on sexual-function end points were seen at 8 to 12 weeks[84]; therefore, it is reasonable to see women at 8- to 12-week intervals initially. Monitoring should include an assessment of clinical efficacy (i.e., symptom reduction) and potential side effects (e.g., acne and hirsutism), as well as laboratory assessment of free and total testosterone levels. Androgen doses can be titrated up or down to achieve symptom reduction in the absence of side effects without allowing serum testosterone levels to exceed the top of the normal range. Periodic monitoring of serum lipid levels and liver function is also reasonable. Until data are available that demonstrate long-term safety, it is advisable to utilize systemic androgen supplements in short-term management only. Extrapolation from safety data regarding systemic estrogen use (which may or may not be germane to systemic androgen use) suggests that short-term management translates into 5 years or less of continuous treatment.

Premenopausal women In the postpartum period, common contributing factors to low sexual desire include fatigue and adjustment to life as a new parent. Hypothyroidism and hyperprolactinemia must be ruled out.

Use of oral contraceptives results in low libido in 5% to 13% of patients.[96] It is not helpful to check serum hormone levels in the setting of oral contraceptive administration. When decreased desire occurs in this context, four management options are available: (1) switching to a nonhormonal contraceptive (which may lead to a higher rate of unintended pregnancy), (2) reducing the estrogen dose (which will lower SHBG levels and result in increased free testosterone levels), (3) changing to a transdermal preparation (which eliminates the hepatic first pass and thereby also lowers SHBG levels), and (4) switching to a preparation with a more androgenic progesterone. Available progestins, listed in order of decreasing androgenicity, include levonorgestrel, desogestrel, norgestimate, norethindrone acetate, norethindrone, and ethynodiol diacetate.[97]

In general, when a woman develops sexual side effects related to a medication, management options include discontinuing the medication, substituting an alternative agent that is less likely to cause the problem, and adding an antidote. Antidotes are indicated when it is medically inadvisable to change the offending medication or when the addition of a second agent is likely to have benefits beyond those of reducing the side effect caused by the first drug. The best example of this is the addition of bupropion to SSRI treatment in the management of depression [*see* Orgasm Problems, *below*].

Whether bupropion is useful as a sexual stimulant for women with low libido who are not taking SSRIs remains to be determined. One study evaluated the use of sustained-release bupropion (300 mg a day for 112 days) in nondepressed, premenopausal women with hypoactive-sexual-desire disorder; significant increases were seen in sexual arousal, orgasm completion, and sexual satisfaction but not in sexual desire.[98] Larger trials are needed to elucidate these findings and to determine which specific sexual-function domains are impacted. It should be noted that the use of bupropion requires careful assessment for possible medication interactions, as well as precautions regarding seizures, particularly in young women who may have eating disorders. Adverse effects, particularly seizures, may be more frequent with immediate-release bupropion than with the sustained-release preparation.

There is little evidence regarding the efficacy of androgen supplementation in premenopausal women (see above).[99] Testosterone is a known teratogen, which can cause virilization of a female fetus. Therefore, most clinicians rarely prescribe androgens to premenopausal women, and in cases in which they are prescribed, they are used only in patients with excellent long-term contraception, such as tubal ligation or an intrauterine device.[100]

Arousal Problems

Subjective arousal problems Useful management strategies for women with subjective sexual-arousal difficulty include permission to use fantasy and erotica to enhance excitatory cognitive stimulation; referral for individual, couples, or sex therapy to address inhibitory cognitive influences; and cautious review and adjustment of any medications that might be interfering with central arousal. Such medications typically include agents that reduce central dopamine and norepinephrine concentrations, those that increase serotonin

levels [*see* Orgasm Problems, *below*], and those that increase prolactin levels.

Objective arousal problems General treatment strategies that are useful in women with genital-arousal problems include the use of lubricants during sexual activity, treatment with estrogen when genital-tissue atrophy is present, and the use of more direct and intense stimulation for longer periods. Additional interventions can be considered in certain circumstances, including PDE5 inhibitors or a clitoral pump (see below) to increase genital blood flow and strategies to combat sexual side effects caused by medications [*see* Orgasm Problems, *below*].

Prescription of an appropriate estrogen regimen requires careful consideration. Systemic estrogen has well-known adverse side effects. Oral estrogen increases levels of SHBG, resulting in lower serum testosterone levels, which may have negative effects on libido. Except in women who have undergone a hysterectomy, estrogen must be given with progesterone to prevent endometrial hyperplasia; however, progesterone downregulates estrogen receptors, which diminishes the benefits of estrogen. Local treatment with small doses of estrogen applied to the genitalia avoids systemic side effects and the need for progesterone. Vaginal suppositories and the estrogen-impregnated silicone ring are good options to relieve vaginal atrophy with minimal systemic absorption. The ring is inserted intravaginally every 12 weeks. Silicone lubricants should be avoided because they may degrade the matrix. The use of suppositories for restoration of vaginal mucosal integrity generally requires daily application for several weeks; frequency of use should then be reduced to the lowest exposure needed to maintain beneficial effects. Long-term intravaginal use of estrogen cream should generally be avoided because it results in significant systemic absorption, but a thin film of estrogen cream applied to the perineum helps restore external-tissue integrity and is probably safe. When systemic estrogen treatment is desired for the management of vasomotor symptoms, parenteral delivery with the transdermal patch is recommended, because this method avoids the hepatic first-pass effect. Direct application of androgen gels or creams (available at compounding pharmacies) to the genitalia is touted to increase sensation; this approach is supported by current data on the role of androgens in female genital response, but it has not yet been proved to be effective in clinical studies.

When decreased sexual arousal is postulated or shown to be vasculogenic, strategies that increase genital blood flow can be considered. Current options include PDE5 inhibitors and clitoral vacuum pumps. PDE5 inhibitors (i.e., sildenafil, tadalafil, and vardenafil) have been shown to be generally effective for men with erectile dysfunction, although the degree of efficacy varies according to the etiology and severity of the disorder.[101] Adverse effects, which include headache, flushing, rhinitis, nausea, and visual symptoms, are usually minor, although sudden cardiac death and stroke have been reported in high-risk patients. Two large studies have assessed the efficacy and safety of sildenafil in women receiving or not receiving estrogen therapy who have genital or subjective sexual-arousal problems.[10,102] In both studies, 50 mg of sildenafil was shown to increase genital blood flow, but subjective arousal did not improve significantly. There was a large placebo response (30% to 40%); no serious adverse events were reported. A smaller study of sildenafil in women receiving estrogen therapy who have clear-cut genital-arousal disorder stratified study patients according to degree of arousal deficit by using vaginal photoplethysmography.[11] In this study, sildenafil increased perception of genital arousal, improved subjective arousal, and reduced the latency to orgasm but only in patients with a low vaginal pulse amplitude response. This finding suggests that sildenafil may be useful in selected women with pure genital-arousal problems. Studies have examined the effects of PDE5 inhibitors in women with multiple sclerosis,[103] spinal cord injury,[104] and sexual dysfunction induced by SSRIs,[105] but numbers are too small to permit firm conclusions. PDE5 inhibitors should be used with caution in patients with recent myocardial infarction or stroke and in those with active coronary ischemia and episodes of heart failure. These agents are contraindicated in women who are taking nitrates.[106]

A large number of other pharmacologic agents have been demonstrated to increase genital blood flow in response to erotic stimuli. These include ephedrine,[107] phentolamine,[108,109] topical alprostadil (a PGE_1 that increases cyclic adenosine monophosphate),[110-112] and a combination of yohimbine (an $alpha_2$ antagonist derived from the bark of the African yohimbe tree) and L-arginine (a nitric oxide precursor).[113] To date, no large-scale, randomized, controlled trials of these agents have shown efficacy in the treatment of female sexual dysfunction; as with sildenafil, changes in genital blood flow have not correlated with changes in subjective arousal in most studies.

Numerous plant-derived remedies are purported to treat sexual problems.[114] Of these, both ginseng and ginkgo biloba have been shown to increase genital blood flow, but neither has yet been shown to benefit women with sexual-arousal disorders. However, a small study of ArginMax, a nutritional supplement containing ginseng, ginkgo, L-arginine, damiana, vitamins, and minerals, did demonstrate improvements in sexual desire, vaginal dryness, clitoral sensation, and frequency of sexual activity in normal female volunteers.[115]

Other options that can be used to treat women with genital-arousal problems are mechanical devices that work either through vibratory stimulation or by causing clitoral vascular engorgement using a vacuum system.[116] Vibrators supply stimulation that is high in intensity, can be delivered directly to the genital sites that have greatest sensitivity, and have the advantage that, unlike a human partner, they do not become fatigued with prolonged duration of use. These characteristics make vibrators invaluable tools for women who require high-intensity, direct clitoral stimulation for prolonged periods to overcome the effects of vascular or neurologic compromise.

The Eros–Clitoral Therapy Device is a small, handheld, battery-powered vacuum pump that can be used to increase blood flow to and around the clitoris during foreplay and self-stimulation. The device received FDA approval on the basis of a small study that showed that application for 3 to 5 minutes, three times a week, for a total of 6 weeks resulted in significant improvement in sensation, vaginal lubrication, ability to achieve orgasm, and greater overall sexual satisfaction in women with and without sexual-arousal disorder.[117] The device costs approximately $400; less expensive knockoffs, some of which also vibrate, can be obtained from sex boutiques.

Persistent sexual-arousal disorder Because persistent sexual-arousal disorder is rare, there are no studies to guide the management of women with the condition. A vigorous attempt should be made to determine the etiology in each case, utilizing specialized testing of genital sensation and blood flow, and treatment should be targeted accordingly. Anecdotal treatments that

have been tried include the following: cognitive reframing of the arousal as a healthy response; discontinuance of potentially offending medications such as trazodone (which causes a similar syndrome in men); administration of SSRIs (because of their dampening effects on arousal); genital application of topical anesthetic agents; administration of antiandrogens; and pelvic embolization (for a pelvic arteriovenous malformation with branches to the clitoris).

Orgasm Problems

Primary orgasmic disorder Classically, women with primary anorgasmia have been raised in a sexually repressive culture, feel guilt or shame about sexual urges, may have negative attitudes toward masturbation, and are relatively unassertive in their relationships. Both the woman and her partner typically are not very knowledgeable about female anatomy or sexual function.

Available data suggest that women with primary orgasm problems respond well to education (of both the patient and her partner) and directed masturbation.[118] Directed masturbation training involves self-stimulation in which the woman becomes aware of the type of stimulation she needs to increase her arousal and pleasure; this knowledge can subsequently be generalized to sexual encounters with a partner.[119-121] A wide variety of self-help books are available to help guide this activity. It may be necessary to address negative attitudes toward the use of erotica and vibrators before introducing these aids. The use of sensate focus exercises in combination with a temporary refrain from intercourse can be recommended as a nonthreatening way for both partners to explore each other's bodies. Counseling can be helpful to improve both sexual and nonsexual communication. Unfortunately, the few studies that have examined the effectiveness of these modalities suffer from numerous methodological deficiencies, including inadequate definitions of dysfunction and treatment success, failure to control for therapist or expectancy factors, failure to isolate specific effects of different treatments, small numbers of study patients, and an absence of control or comparison groups.

Secondary orgasmic disorder Acquired orgasmic dysfunction is a common side effect of medications, particularly antidepressants. In particular, sexual dysfunction is reported by as many as 30% to 60% of patients taking SSRIs.[122] Although these drugs can adversely affect any sexual-function domain, orgasm delay is the problem most frequently cited. It is imperative to address sexual side effects during SSRI treatment, because consequent nonadherence or premature treatment discontinuance can place patients at risk for recurrent depression, relapse, chronicity, and, in extreme cases, suicide.

When discontinuance of antidepressants is not an option, several other strategies can be considered. These include dose reduction or drug holiday, adding a second medicine as an antidote, or switching to an alternative agent with less sex-negative effects.[123] Sexual side effects appear to be dose related and may be reduced by lowering the daily dose. Similarly, drug holidays (stopping the drug for 48 hours before anticipated sexual activity, such as 2 days before a weekend) may reduce sexual side effects in women who are taking agents that have a short half-life. Such strategies, however, carry the risk of decreasing the antidepressant dose to subtherapeutic levels and should therefore be used with caution. Antidotes are especially appealing when the agent added not only helps to treat sexual side effects but also has additional benefits, such as augmenting antidepressant efficacy.

Two small, randomized, placebo-controlled trials have examined the effectiveness of adding sustained-release bupropion to SSRI treatment in euthymic men and women who had developed sexual side effects related to SSRI use. One trial did not demonstrate efficacy, but the dosage of bupropion was low (150 mg a day) and the duration of the trial was short (3 weeks).[124] In the second trial, patients receiving bupropion did demonstrate an increase in both the desire to engage in sexual activity and the frequency of sexual activity.[125] The bupropion dosage used in this trial was 300 mg a day, and the duration of the study was 4 weeks. Anecdotal reports suggest that a 75 mg dose of immediate-release bupropion taken several hours before anticipated sexual activity also helps improve libido, lubrication, orgasm delay, and orgasm quality in women taking SSRIs.[126] Larger trials are needed to further investigate these findings and to determine the optimal bupropion preparation and treatment dose.

Because the sexual side effects of SSRIs appear to be mediated in part through the nitric oxide pathway, PDE5 inhibitors are of theoretical benefit. Small studies suggest that sildenafil may be beneficial in women with SSRI-induced sexual dysfunction.[105] Case reports also suggest efficacy with the addition of methylphenidate and dextroamphetamine at dosages of 5 mg/day.[123]

Substitution therapy is sometimes a viable option. Both bupropion and nefazodone are less likely than SSRIs to cause sexual dysfunction; 10% or less of patients taking these drugs experience this side effect. Mirtazapine may also be associated with a lower rate of sexual side effects.

Anecdotal experience suggests that the use of pelvic floor exercises to increase pelvic floor muscle tension enhances sexual arousal and orgasmic pleasure in some women, including women who experience sexual incontinence. Unfortunately, studies have been small, have lacked appropriate controls, and have had inconsistent outcome measures. In one small, randomized trial, Bo and colleagues demonstrated a 50% improvement in sexual incontinence in women who performed pelvic-strengthening exercises, versus 10% in the control group.[127] Pelvic floor exercises are easily taught, pose little potential for harm, and have minimal cost; they are therefore a reasonable first-line treatment for women with stress incontinence. Effectiveness can be augmented by using biofeedback techniques, including self-use of a handheld peritoneometer, or by using vaginal weights.[128] Estrogen, either systemic or topical, has not been found to be effective for the management of stress incontinence.[129] Alpha-adrenergic agonists and tricyclic antidepressants may be helpful for the incontinence, but their use for this purpose is not FDA approved, and they may have sexual side effects.[130] Surgical treatment can be a double-edged sword: although it often restores continence, it sometimes reduces sensation and blood flow, resulting in impaired sexual arousal and orgasm.[131] In addition, some procedures performed to correct incontinence or prolapse are associated with an increase in dyspareunia.[132]

Women with acquired orgasmic dysfunction are often less satisfied with their overall relationship and are more distressed. Therefore, counseling is frequently a requisite aspect of treatment.[133]

Pain during Sexual Activity

Treatment for women who experience pain during sexual activity should be directed at the underlying etiology. A multidisciplinary approach, including involvement of a medical

provider, psychotherapist or sex therapist, and a specially trained physical therapist, is often very useful. It may not be possible to eliminate the pain completely. More realistic goals are to help the patient develop better coping strategies and gradually achieve or regain satisfactory sexual function. When the pain is the result of inadequate lubrication, causes to consider include inadequate foreplay before intercourse and vulvovaginal atrophy. Lubricants can be very useful in such patients; topical estrogen should also be considered in postmenopausal women, especially if atrophic changes are evident. Medications that cause vaginal dryness should be avoided or discontinued, if possible. Infections, dermatitis, dermatosis, correctable anatomic abnormalities, and conditions associated with deep pelvic pain must each be addressed appropriately. Women who have experienced pain during sexual activity often feel anxious that the pain will recur and frequently avoid subsequent sexual contact. Over time, this fear and withdrawal can become as problematic as the pain itself. By working with a sex therapist, affected women and their partners can learn to focus on activities that are pleasurable. The use of sensate focus exercises, which direct attention to activities and part of the body that do not provoke anxiety or cause pain, may be particularly helpful.

Vulvodynia and vestibulitis The causes of vulvodynia and vestibulitis are not known. Proposed etiologies include increased urinary oxalates, genetic or immune factors, hormonal factors, inflammation, infection, and neuropathic changes.[134] Treatments that have been tried include vulvar care measures; topical measures (e.g., local anesthetics, capsaicin, amitriptyline, nitroglycerin, and baclofen); oral medications (e.g., tricyclic antidepressants and anticonvulsants); injectable medications (e.g., interferon, steroids, local anesthetics, and botulinum toxin); pelvic floor muscle rehabilitation; low-oxalate diet and calcium citrate supplementation; acupuncture; hypnotherapy; and surgery (e.g., vestibulectomy, perineoplasty, and pudendal nerve decompression).[135] Of these, several are of little harm and fall within the purview of the general internist.

Vulvar care measures consist of wearing cotton underwear, avoiding exposure to irritants, gentle cleansing with water, patting dry (no rubbing or use of hair dryers), application of soothing emollients (e.g., vegetable oils and petrolatum) that do not contain preservatives, and use of ice or gel packs for discomfort. Referral to a physical therapist with expertise in vulvar pain management may be worthwhile, especially for patients with concomitant vaginismus. Topical products, such as 5% lidocaine ointment applied to painful areas 10 minutes before sexual activity, can be helpful. Topical preparations should be formulated as ointments, because creams contain more preservatives and can cause stinging or burning. Oral agents such as tricyclic antidepressants and gabapentin may also be tried; however, vaginal dryness can be a side effect. Finally, referral to a specialist with expertise in the management of vulvodynia and vestibulitis is warranted when simple measures do not provide relief.

Vaginismus Psychological issues, such as ambivalence about sex from a strict upbringing or avoidance because of past trauma, are common in women with vaginismus. Secondary vaginismus is a common conditioned response in women who have experienced genital pain (e.g., from urinary tract procedures performed during childhood). Vaginismus is characterized by pelvic floor hypertonicity or spasm and impaired ability to both contract and relax the pelvic floor musculature. The most effective treatment for women with vaginismus consists of a combination of psychotherapy to address personal issues, cognitive-behavioral therapy to break the pain-spasm-pain cycle, sphincter muscle exercises to achieve pelvic muscle relaxation,[135] and gradual desensitization using vaginal self-dilation with fingers or vaginal dilator insertion.[15] The use of functional electrical stimulation biofeedback to teach pelvic muscle relaxation improves the efficiency and success rate of treatment.[136] Injection of botulinum toxin may be useful in refractory cases.[137,138]

The author has no commercial relationships with manufacturers of products or providers of services discussed in this chapter.

References

1. Graziottin A: The challenge of sexual medicine for women: overcoming cultural and educational limits and gender biases. J Endocrinol Invest 26:139, 2003

2. Nour NM: Female circumcision and genital mutilation: a practical and sensitive approach. Contemp Ob Gyn 45:50, 2000

3. Stevenson RW: Sexual medicine: why psychiatrists must talk to their patients about sex. Can J Psychiatry 49:673, 2004

4. Nicolosi A, Laumann EO, Glasser DB, et al: Sexual behavior and sexual dysfunctions after age 40: the global study of sexual attitudes and behaviors. Urology 64:991, 2004

5. Laumann EO, Paik A, Rosen RC: Sexual dysfunction in the United States: prevalence and predictors. JAMA 281:537, 1999

6. Mercer CH, Fenton KA, Johnson AM, et al: Sexual function problems and help seeking behaviour in Britain: national probability sample survey. BMJ 327:426, 2003

7. Fugl-Meyer AR, Fugl-Meyer K: Sexual disabilities, problems and satisfaction in 18–74 year old Swedes. Scand J Sexology 2:79, 1999

8. Richters J, Grulich AE, de Visser RO, et al: Sex in Australia: sexual difficulties in a representative sample of adults. Aust N Z J Public Health 27:164, 2003

9. Bancroft J, Loftus J, Long JS: Distress about sex: a national survey of women in heterosexual relationships. Arch Sex Behav 32:193, 2003

10. Berman JR, Berman LA, Toler SM, et al: Safety and efficacy of sildenafil citrate for the treatment of female sexual arousal disorder: a double-blind, placebo-controlled study. J Urol 170:2333, 2003

11. Basson R, Brotto LA: Sexual psychophysiology and effects of sildenafil citrate in oestroegenized women with acquired genital arousal disorder and impaired orgasm: a randomized controlled trial. BJOG 110:1014, 2003

12. Cain VS, Johannes CB, Avis NE, et al: Sexual functioning and practices in a multi-ethnic study of midlife women: baseline results from SWAN. J Sex Res 40:266, 2003

13. Basson R, Leiblum S, Brotto L, et al: Definitions of women's sexual dysfunction reconsidered: advocating expansion and revision. J Psychosomat Obstet Gynecol 24:221, 2003

14. Masters WH, Johnson V: Human Sexual Response. Little, Browne & Co, Boston, 1966

15. Kaplan HS: Hypoactive sexual desire. J Sex Marital Ther 2:3, 1979

16. Basson R: Female sexual response revisited. J Soc Obstet Gynecol 378, 2000

17. Schober JM, Meyer-Bahlburg HF, Ransley PG: Self-assessment of genital anatomy, sexual sensitivity and function in women: implications for genitoplasty. BJU Int 94:589, 2004

18. Hines TM: The G-spot: a modern gynecologic myth. Am J Obstet Gynecol 185:359, 2001

19. Munarriz R, Kim NN, Goldstein I, et al: Biology of female sexual function. Urol Clin North Am 29:685, 2002

20. Maas CP, Weijenborg PTM, ter Kuile MM: The effect of hysterectomy on sexual functioning. Annu Rev Sex Res 14:83, 2003

21. Scott JR, Sharp HT, Dodson MK, et al: Subtotal hysterectomy in modern gynecology: a decision analysis. Am J Obstet Gynecol 176:1186, 1997

22. Dennerstein L, Randolph J, Taffe J, et al: Hormones, mood, sexuality, and the menopausal transition. Fertil Steril 77:S42, 2002

23. Burger HG, Dudley EC, Cui J: A prospective longitudinal study of serum testosterone, dehydroepiandrosterone sulfate, and sex hormone binding globulin levels through the menopause transition. J Clin Endocrinol Metab 85:2832, 2000

24. Laughlin GA, Barrett-Connor E, Kritz-Silverstein D: Hysterectomy, oophorectomy, and endogenous sex hormone levels in older women: the Rancho Bernardo Study. J Clin Endocrinol Metab 85:645, 2000

25. Davis SR: Androgens and female sexuality. J Gend Specif Med 3:36, 2000

26. Pfaus JG, Shadiack A, Van Soest T, et al: Selective facilitation of sexual solicitation in the female rat by a melanocortin receptor agonist. Proc Natl Assoc Sci USA 101:10201, 2004

27. Clayton AH: Sexual function and dysfunction in women. Psychiatr Clin North Am 26:673, 2003

28. Kruger TH, Haake P, Hartmann U, et al: Orgasm-induced prolactin secretion: feedback control of sexual drive? Neurosci Biobehav Rev 26:31, 2002

29. Graziottin A: The challenge of sexual medicine for women: overcoming cultural and educational limits and gender biases. J Endocrinol Invest 26:139, 2003

30. Salonia A, Nappi RE, Pontillo M, et al: Menstrual cycle-related changes in plasma oxy-

tocin are relevant to normal sexual function in healthy women. Horm Behav 47:164, 2005

31. McKenna KE: The neurophysiology of female sexual function. World J Urol 20:93, 2002

32. Giuliano F, Rampin O, Allard J: Neurophysiology and pharmacology of female sexual response. J Sex Marital Ther 28:101, 2002

33. Min K, O'Connell L, Munarriz R, et al: Experimental models for the investigation of female sexual function and dysfunction. Int J Impot Res 13:151, 2001

34. Traish AM, Kim N, Min K, et al: Role of androgens in female genital sexual arousal: receptor expression, structure, and function. Fertil Steril 77:S11, 2002

35. Clayton AH: Sexual function and dysfunction in women. Psychiatr Clin North Am 26:673, 2003

36. Modelska K, Cumings S: Female sexual dysfunction in postmenopausal women: systematic review of placebo-controlled trials. Am J Obstet Gynecol 188:286, 2003

37. Youth Risk Behavior Surveillance—United States, 2003. MMWR Morb Mortal Wkly Rep 53(SS-2):1, 2004

38. Davis JA, Smith TW: General Social Survey, 1972–1998: Cumulative Codebook. NORC, Chicago, 1998

39. Teen pregnancy, birth, and abortion. Sexuality Information and Education Council of the United States. SIECUS Rep 30:1, 2002
http://www.siecus.org/pubs/fact/fact0010.html

40. Sex and America's Teenagers. Alan Guttmacher Institute, New York, 1994

41. Koutsky L: Epidemiology of genital human papillomavirus infection. Am J Med 102:3, 1997

42. Mosher WD, Chandra A, Jones J: Sexual behavior and selected health measures: men and women 15–44 years of age, United States, 2002. Advance data from vital and health statistics; no. 362. National Center for Health Statistics, Hyattsville, Maryland, 2005

43. Saewyc EM, Bearinger LH, Heinz PA, et al: Sexual intercourse, abuse and pregnancy among adolescent women: does sexual orientation make a difference? Fam Plann Perspect 31:127, 1999

44. Roberts LW, Fromm LM, Bartlik BD: Sexuality of women through the life phases. Textbook of Women's Health. Wallis LA, Kasper AS, Reader GG, et al, Eds. Lippincott-Raven Publishers, Philadephia, 1998

45. Sexuality in Middle and Later Life. Sexuality Information and Education Council of the United States, Washington, DC, 2004
http://www.siecus.org/pubs/fact/fact0018.html

46. Diagnostic and Statistical Manual for Mental Disorders, 4th ed, Text Revision. American Psychiatric Association. American Psychiatric Press, Washington, DC, 2000

47. International Statistical Classification of Diseases and Related Health Problems, 10th ed. World Health Organization. World Health Organization, Geneva, 1992

48. Basson R, Berman J, Burnett A, et al: Report of the International Consensus Development Conference on Female Sexual Dysfunction: Definitions and Classifications. J Urol 163:888, 2000

49. Doruk H, Akbay E, Cayan S, et al: Effect of diabetes mellitus on female sexual function and risk factors. Arch Andrology 51:1, 2005

50. Sipski ML, Behnegar A: Neurogenic female sexual dysfunction: a review. Clin Autonom Res 11:279, 2001

51. Copeland CE, Bosse MJ, McCarthy ML, et al: Effect of trauma and pelvic fracture on female genitourinary, sexual, and reproductive function. J Orthopaed Trauma 11:73, 1997

52. Maas CP, Weijenborg PT, ter Kuile MM: The effect of hysterectomy on sexual functioning. Annu Rev Sex Res 14:83, 2003

53. Kim DH, Lee YS, Lee ES: Alteration of sexual function after classic intrafascial supracervical hysterectomy and total hysterectomy. J Am Assoc Gynecol Laparosc 10:60, 2003

54. Handa VL, Harvey L, Cundiff GW, et al: Sexual function among women with urinary incontinence and pelvic organ prolapse. Am J Obstet Gynecol 191:751, 2004

55. Barber MD, Visco AG, Wyman JF, et al: Sexual function in women with urinary incontinence and pelvic organ prolapse. Obstet Gynecol 99:281, 2002

56. Shaw C: A systematic review of the literature on the prevalence of sexual impairment in women with urinary incontinence and the prevalence of urinary leakage during sexual activity. Eur Urol 42:432, 2002

57. Barber MD, Mullen KJ, Dowsett SA, et al: The impact of stress urinary incontinence on sexual activity in women. Cleveland Clin J Med 72:225, 2005

58. Haefner HK, Collins ME, Davis GD, et al: The vulvodynia guideline. J Low Genit Tract Dis 9:40, 2005

59. Stewart EG: Treatment options for vulvar vestibulitis. Contemp Ob/Gyn 48:47, 2003

60. Jones le RA: The use of validated questionnaires to assess female sexual dysfunction. World J Urol 20:89, 2002

61. Derogatis LR: The Derogatis interview for sexual functioning (DISF/DISF-SR): an introductory report. J Sex Marital Ther 23:291, 1997

62. Mazer NA, Leiblum SR, Rosen RC: The Brief Index of Sexual Functioning for Women (BISF-W): a new scoring algorithm and comparison of normative and surgically menopausal populations. Menopause 7:350, 2000

63. Rosen RC, Brown C, Heiman J, et al: The Female Sexual Function Index (FSFI): a multidimensional self-report instrument for the assessment of female sexual function. J Sex Marital Ther 26:191, 2000

64. Geer JH: Development of the vaginal photoplethysmograph. Int J Impot Res 17:285, 2005

65. Khalife S, Binik YM, Cohen DR, et al: Evaluation of clitoral blood flow by color Doppler ultrasonography. J Sex Marital Ther 26:187, 2000

66. Sommer F, Caspers HP, Esders K, et al: Measurement of vaginal and labial oxygen tension for the evaluation of female sexual function. J Urol 165:1181, 2001

67. Suh DD, Yang CC, Cao Y, et al: MRI of female genital and pelvic organs during sexual arousal. J Psychosom Obstet Gynecol 25:153, 2004

68. Guay A, Davis SR: Testosterone insufficiency in women: fact or fiction? World J Urol 20:106, 2002

69. Davis SR, Davison SL, Donath A, et al: Circulating androgen levels and self-reported sexual function in women. JAMA 294:91, 2005

70. Labrie F: Intracrinology. Mol Cell Endocrinol 78:C113, 1991

71. Bancroft J: Sexual effects of androgens in women: some theoretical considerations. Fertil Steril 77:S55, 2002

72. Gerber JR, Johnson JV, Bunn JY, et al: A longitudinal study of the effects of free testosterone and other psychosocial variables on sexual function during the natural traverse of menopause. Fertil Steril 83:643, 2005

73. Leiblum SR, Wiegel M: Psychotherapeutic interventions for treating female sexual dysfunction. World J Urol 20:127, 2002

74. Alexander JL, Kotz K, Dennerstein L, et al: The effects of postmenopausal hormone therapies on female sexual functioning: a review of double-blind, randomized controlled trials. Menopause 11:749, 2004

75. Suckling J, Lethaby A, Kennedy R: Local oestrogen for vaginal atrophy in postmenopausal women. Cochrane Database Syst Rev 4:CD001500, 2003

76. Barrett-Connor E, Young R, Notelovitz M, et al: A two-year, double-blind comparison of estrogen-androgen and conjugated estrogens in surgically menopausal women: effects on bone mineral density, symptoms, and lipid profiles. J Reprod Med 44:1012, 1999

77. Lobo RA, Rosen RC, Yang HM, et al: Comparative effects of oral esterified estrogens with and without methyltestosterone on endocrine profiles and dimensions of sexual function in postmenopausal women with hypoactive desire disorder. Fertil Steril 79:1341, 2003

78. Sherwin BB, Gelfand MM: The role of androgen in the maintenance of sexual functioning in oophorectomized women. Psychosom Med 49:397, 1987

79. Sherwin BB, Gelfand MM, Brender W: Androgen enhances sexual motivation in females: a prospective, crossover study of sex steroid administration in the surgical menopause. Psychosom Med 47:339, 1985

80. Burger H, Hailes J, Nelson J, et al: Effect of combined implants of oestradiol and testosterone on libido in postmenopausal women. Br Med J 294:936, 1987

81. Burger HG, Hailes J, Menelaus M, et al: The management of persistent menopausal symptoms with oestradiol-testosterone implants: clinical, lipid and hormonal results. Maturitas 6:351, 1984

82. Davis SR, McCloud P, Strauss BJ, et al: Testosterone enhances estradiol's effects on postmenopausal bone density and sexuality. Maturitas 21:227, 1995

83. Braunstein GD, Sundwall DA, Katz M, et al: Safety and efficacy of a testosterone patch for the treatment of hypoactive sexual desire disorder in surgically menopausal women: a randomized, placebo-controlled trial. Arch Intern Med 165:1582, 2005

84. Buster JE, Kingsberg SA, Aguirre O, et al: Testosterone patch for low sexual desire in surgically menopausal women: a randomized trial. Obstet Gynecol 105:944, 2005

85. Kroll R, Davis S, Moreau M, et al: Testosterone transdermal patch (TTP) significantly improved sexual function in naturally menopausal women in a large phase III study. Fertil Steril 82(suppl 2):S77, 2004

86. Sarrel P, Dobay B, Wiita B: Estrogen and estrogen-androgen replacement in postmenopausal women dissatisfied with estrogen-only therapy: sexual behavior and neuroendocrine responses. J Reprod Med 43:847, 1998

87. Sherwin BB: Randomized clinical trials of combined estrogen-androgen preparations: effects on sexual functioning. Fertil Steril 77:S49, 2002

88. Egarter C, Topcuoglu A, Vogl S, et al: Hormone replacement therapy with tibolone: effects on sexual functioning in postmenopausal women. Acta Obstet Gynecol Scand 81:649, 2002

89. Kokcu A, Cetinkaya MB, Yanik F, et al: The comparison of effects of tibolone and conjugated estrogen-medroxyprogesterone acetate therapy on sexual performance in postmenopausal women. Maturitas 36:75, 2000

90. Nathorst-Boos J, Hammar M: Effect on sexual life—a comparison between tibolone and a continuous estradiol-norethistrone acetate regimen. Maturitas 26:15, 1997

91. Davis SR, Davison SL, Donath S, et al: Circulating androgen levels and self-reported sexual function in women. JAMA 294:91, 2005

92. Arlt W, Callies F, Allolio B: DHEA replacement in women with adrenal insufficiency—pharmocokinetics, bioconversion and clinical effects on well-being, sexuality and cognition. Endocr Res 26:505, 2000

93. Dayal M, Sammel MD, Zhao J, et al: Supplementation with DHEA: effect on muscle size, strength, quality of life, and lipids. J Womens Health (Larchmt) 14:391, 2005

94. Cameron DR, Braunstein GD: The use of dehydroepiandrosterone therapy in clinical practice. Treat Endocrinol 4:95, 2005

95. Guay A, Davis S: Testosterone insufficiency in women: fact or fiction? World J Urol 20:106, 2002

96. Slap GB: Oral contraceptives and depression: impact, prevalence and cause. J Adolesc Health Care 2:53, 1981

97. Hatcher RA, Zieman M, Creinin MD, et al: Contraceptive Technology, 18th ed. Bridging the Gap Foundation, Clayton, Georgia, 2005

98. Segraves RT, Clayton A, Croft H, et al: Bupropion sustained release for the treatment of hypoactive sexual desire disorder in premenopausal women. J Clin Pharmacol 24:339, 2004

99. Goldstat R, Briganti E, Tran J, et al: Transdermal testosterone improves mood, well being and sexual function in premenopausal women. Menopause 10:390, 2003

100. Robinson B, Feldman J, Striepe M, et al: Women's sexual health: an interdisciplinary approach to treating low sexual desire. Minn Med 86:34, 2003

101. Rosen RC, McKenna KE: PDE-5 inhibition and sexual response: pharmacological mechanisms and clinical outcomes. Annu Rev Sex Res 13:36, 2002

102. Basson RJ, McInnes R, Smith M, et al: Efficacy and safety of sildenafil citrate in women with sexual dysfunction associated with female sexual arousal disorder. J Womens Health Gend Based Med 11:339, 2002

103. DasGupta R, Wiseman OJ, Kanabar G, et al: Efficacy of sildenafil in the treatment of female sexual dysfunction due to multiple sclerosis. J Urol 171:1189, 2004

104. Sipski ML, Rosen RC, Alexander CJ, et al: Sildenafil effects of sexual and cardiovascular responses in women with spinal cord injury. Urology 55:812, 2000

105. Nurnberg HG, Hensley PL, Lauriello J, et al: Sildenafil for women patients with antidepressant-induced sexual dysfunction. Psychiatr Serv 50:1076, 1999

106. Alboni P, Bettiol K, Fuca G, et al: Sexual activity with and without the use of sildenafil: risk of cardiovascular events in patients with heart disease. Ital Heart J 5:343, 2004

107. Meston CM, Heiman JR: Ephedrine-activated physiological sexual arousal in women. Arch Gen Psychiatr 55:652, 1998

108. Rosen RC, Phillips NA, Gendrano N, et al: Oral phentolamine and female sexual arousal disorder: a pilot study. J Sex Marital Ther 25:137, 1999

109. Rubio-Aurioles E, Lopez M, Lipezker M: Phentolamine mesylate in postmenopausal women with FSAD. J Sex Marital Ther 28:S205, 2002

110. Islam A, Mitchell J, Rosen R, et al: Topical alprostadil in the treatment of FSAD. J Sex Marital Ther 27:531, 2001

111. Padma-Nathan H, Brown C, Fendl J, et al: Efficacy and safety of topical alprostadil cream for the treatment of female sexual arousal disorder (FSAD): a double-blind, multicenter, randomized, and placebo-controlled trial. J Sex Marital Ther 29:329, 2003

112. Bachara A, Bertolino MV, Casabe A, et al: Duplex Doppler ultrasound assessment of clitoral hemodynamics after topical administration of alprostadil in women with arousal and orgasmic disorders. J Sex Marital Ther 29:1, 2003

113. Meston CM, Worcel M: The effects of yohimbine plus L-arginine glutamate on sexual arousal in postmenopausal women with sexual arousal disorder. Arch Sex Behav 31:323, 2002

114. Rowland DL, Tai Wendi: A review of plant-derived and herbal approaches to the treatment of sexual dysfunctions. J Sex Marital Ther 29:185, 2003

115. Ito TY, Trant AS, Polan ML: A double-blind placebo-controlled study of ArginMax, a nutritional supplement for enhancement of female sexual function. J Sex Marital Ther 27:541, 2001

116. Billups KL: The role of mechanical devices in treating female sexual dysfunction and enhancing the female sexual response. World J Urol 20:137, 2002

117. Wilson SK, Delk JR, Billups KL: Treating symptoms of female sexual arousal disorder with the Eros-Clitoral Therapy Device. J Gend Specif Med 4:54, 2001

118. Heiman J, Meston CM: Empirically validated treatment for sexual dysfunction. Annu Rev Sex Res 8:148, 1997

119. Kohlenberg RJ: Directed masturbation and the treatment of primary orgasmic dysfunction. Arch Sex Behav 3:349, 1974

120. Nairne KD, Hemsley DR: The use of directed masturbation training in the treatment of primary anorgasmia. Br J Clin Psychol 22:283, 1983

121. Heiman J, Meston CM: Empirically validated treatment for sexual dysfunction. Annu Rev Sex Res 8:148, 1997

122. Gregorian RS, Golden KA, Bahce A, et al: Antidepressant-induced sexual dysfunction. Ann Pharmacother 36:1577, 2002

123. Zajecka J: Strategies for the treatment of antidepressant-related dysfunction. J Clin Psychiatr 62:S35, 2001

124. Masand PS, Ashton AK, Gupta S, et al: Sustained-release bupropion for selective serotonin reuptake inhibitor–induced sexual dysfunction: a randomized, double-blind, placebo-controlled, parallel-group study. Am J Psychiatr 158:805, 2001

125. Clayton AH, Warnock JK, Kornstein SG, et al: A placebo-controlled trial of bupropion SR as an antidote for selective serotonin reuptake inhibitor-induced sexual dysfunction. J Clin Psychiatr 65:62, 2004

126. Labbate LA, Grimes JB, Hines A, et al: Bupropion treatment of serotonin reuptake antidepressant-associated sexual dysfunction. Ann Clin Psychiatr 9:241, 1997

127. Bo K, Talseth T, Vinsnes A: Randomized controlled trial on the effect of pelvic floor muscle training on quality of life and sexual problems in genuine stress incontinent women. Acta Obstet Gynecol Scand 79:598, 2000

128. Perkins J: Vaginal weights for assessment and training of the pelvic floor. J Wound Ostomy Continence Nurs 25:206, 1998

129. Al Badr A, Ross S, Soroka D, et al: What is the available evidence for hormone replacement therapy in women with stress urinary incontinence? J Obstet Gynaecol Can 25:567, 2003

130. Viktrup L, Bump RC: Pharmacological agents used for the treatment of stress urinary incontinence in women. Curr Med Res Opin 19:485, 2003

131. Pauls RN, Berman JR: Impact of pelvic floor disorders and prolapse on female sexual function and response. Urol Clin North Am 29:677, 2002

132. Weber AM, Walters MD, Piedmonte MR: Sexual function and vaginal anatomy in women before and after surgery for pelvic organ prolapse and urinary incontinence. Am J Obstet Gynecol 182:1610, 2000

133. McCabe MP, Delaney SM: An evaluation of therapeutic programs for the treatment of secondary inorgasmia in women. Arch Sex Behav 21:69, 1992

134. Haefner HK, Collins ME, Davis GD, et al: The vulvodynia guideline. J Low Genit Tract Dis 9:40, 2005

135. Zukerman Z, Roslik Y, Orvieto R: Treatment of vaginismus with the Paula Garburg sphincter muscle exercises. Harefuah 144:246, 2005

136. Seo JT, Choe JH, Lee WS, et al: Efficacy of functional electrical stimulation-biofeedback with sexual cognitive-behavioral therapy as treatment for vaginismus. Urology 66:77, 2005

137. Ghazizadeh S, Nikzad M: Botulinum toxin in the treatment of refractory vaginismus. Obstet Gynecol 104:922, 2004

138. Shafik A, El-Sibai O: Vaginismus: results of treatment with botulin toxin. J Obstet Gynaecol 20:300, 2000

139. Basson R: Female sexual response: the role of drugs in the management of sexual dysfunction. Obstet Gynecol 98:350, 2001

140. Ragucci KR, Culhane NS: Treatment of female sexual dysfunction. Ann Pharmacother 37:546, 2003

141. Drugs that cause sexual dysfunction: an update. Med Lett Drugs Ther 34:73, 1992

NORMAL LABORATORY VALUES

Chemistry

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Adrenocorticotropin (ACTH)	P	6.0–76.0 pg/ml	1.3–16.7 pmol/L	Immunoassay	0.2202
Alanine aminotransferase (ALT, SGPT)	S			Kinetic method	0.01667
Female		7–30 U/L	0.12–0.50 μkat/L		
Male		10–55 U/L	0.17–0.92 μkat/L		
Albumin	S	3.1–4.3 g/dl	31–43 g/L	Colorimetry (bromocresol purple)	10
Aldolase	S	0–7 U/L	0–7 U/L	Kinetic method	1
Aldosterone (adult)				Immunoassay	
Supine, normal-sodium diet	S, P	2–9 ng/dl	55–250 pmol/L		27.74
Upright, normal-sodium diet	S, P	2–5 times supine value with normal-sodium diet			
Supine, low-sodium diet	S, P	2–5 times supine value with normal-sodium diet			
Urine, normal-sodium diet	U	2.3–21.0 μg/24 hr	6.38–58.25 nmol/24 hr		2.774
Alkaline phosphatase (adult)	S			Kinetic method	0.01667
Female		30–100 U/L	0.5–1.67 μkat/L		
Male		45–115 U/L	0.75–1.92 μkat/L		
Alkaline phosphatase, heat fractionated	S	20%–35%	0.20–0.35	Kinetic method	0.01
Alpha-fetoprotein (nonmaternal)	S	< 12.8 IU/ml	< 9.92 μg/L	Immunoassay	0.775
Ammonia	P	12–48 μmol/L	12–48 μmol/L	Enzymatic analysis	1
Amylase	S	53–123 U/L	0.88–2.05 nkat/L	Kinetic method	0.01667
	P	43–115 U/L	0.72–1.92 nkat/L		
	U	4–400 U/L	0.07–6.67 nkat/L		
Androstenedione (adult)	S	50–250 ng/dl	1.75–8.73 nmol/L	Immunoassay	0.0349
Angiotensin-converting enzyme	S			Kinetic method	1
Female		19–79 U/L	19–79 U/L		
Male		19–95 U/L	19–95 U/L		
Apolipoprotein	S			Nephelometry	
Apolipoprotein A-1		119–240 mg/dl	1.19–2.4 g/L		0.01
Apolipoprotein B		52–163 mg/dl	0.52–1.63 g/L		0.01
Apolipoprotein B:apolipoprotein A-1 ratio		0.35–0.98	0.35–0.98		1
Aspartate aminotransferase (AST, SGOT)	S			Kinetic method	0.01667
Female		9–25 U/L	0.15–0.42 μkat/L		
Male		10–40 U/L	0.17–0.67 μkat/L		
$Beta_2$-microglobulin	S, P	1.2–2.8 mg/L	1.2–2.8 mg/L	Immunoassay	1
	U	< 200 μg/L	< 200 μg/L		
Bicarbonate (HCO_3^-)	WB, S	22–26 mEq/L	22–26 mmol/L	Calculation	1
Bilirubin, direct	S	0.0–0.4 mg/dl	0–7 μmol/L	Colorimetry	17.1
Bilirubin, total	S	0.0–1.0 mg/dl	0–17 μmol/L	Colorimetry	17.1
C peptide (adult)	S, P	0.5–2.0 ng/ml	0.17–0.66 nmol/L	Immunoassay	0.33
C-reactive protein	S	0.0–12.0 mg/L	0–12 mg/L	Nephelometry	1
CA 15-3	S	0–30 U/ml	0–30 kU/L	Immunoassay	1
CA 19-9	S	0–37 U/ml	0–37 kU/L	Immunoassay	1
CA 27,29	S	0–32 U/ml	0–32 kU/L	Immunoassay	1
CA-125	S	0–35 U/ml	0–35 kU/L	Immunoassay	1
Calcitonin	S			Immunoassay	1
Female		2–17 pg/ml	2–17 ng/L		
Male		3–26 pg/ml	3–26 ng/L		
Calcium	S	8.5–10.5 mg/dl	2.1–2.6 mmol/L	Colorimetry	0.25
	U	0–300 mg/24 hr	0.0–7.5 mmol/24 hr	Colorimetry	0.025
Calcium, ionized	WB	1.14–1.30 mmol/L	1.14–1.30 mmol/L	Ion-selective electrode	1

Note: This table includes reference values, methods, and conversion factors for tests commonly ordered at the Massachusetts General Hospital (MGH) and recorded in the Case Records. The table revises the most recently published data (Normal Reference Laboratory Values. N Engl J Med 327:718–724, 1992). Laboratory values are expressed in the units used at MGH (conventional units) and the units of the Système International d'Unités (SI units), with a factor to convert conventional units to SI units by multiplication. The table is not intended to provide a comprehensive review of reference values, since this information is widely available in standard textbooks. Detailed information on specific methods and instruments is not provided, to avoid suggesting an endorsement of commercial products. Because reference values are affected by many variables, table reference ranges may differ from those of other laboratories as well as from the values given in subsections of *Scientific American Medicine*. Furthermore, the methods and units reported should not be interpreted as a judgment about the optimal way to perform or report a given laboratory test.

Specimen abbreviations: CSF—cerebrospinal fluid JF—joint fluid P—plasma S—serum U—urine WB—whole blood

Chemistry *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Carbon dioxide content, total	P	24–30 mmol/L	24–30 mmol/L	Carbon dioxide electrode	1
Carbon dioxide, partial pressure, arterial (P_aco_2)	WB	35–45 mm Hg	4.7–6.0 kPa	Carbon dioxide electrode	0.1333
Carboxyhemoglobin	WB	< 5% of total hemoglobin	< 0.05 fraction of total hemoglobin saturation	Co-oximetry	0.01
Carcinoembryonic antigen (CEA)	P, S	0.0–3.4 ng/ml	0.0–3.4 µg/L	Immunoassay	1
Catecholamines (adult)	U			High-pressure liquid chromatography	
Epinephrine		2–24 µg/24 hr	11–131 nmol/24 hr		5.458
Norepinephrine		15–100 µg/24 hr	89–591 nmol/24 hr		5.911
Dopamine		52–480 µg/24 hr	340–3,134 nmol/24 hr		6.53
Total (epinephrine + norepinephrine)		26–121 µg/24 hr	142–660 nmol/24 hr		5.458 (as normeta-nephrine)
Cerebrospinal fluid (adult)	CSF				
Albumin		11–48 mg/dl	0.11–0.48 g/L	Nephelometry	0.01
Cell count		0–5 mononuclear cells/µl	0–5 × 10^6 cells/L	Manual count	1 × 10^6
Chloride		120–130 mmol/L	120–130 mmol/L	Coulometry	1
Glucose		50–75 mg/dl	2.8–4.2 mmol/L	Enzymatic analysis	0.05551
IgG		8.0–8.6 mg/dl	0.08–0.086 g/L	Nephelometry	0.01
Pressure		70–180 mm of water	70–180 arbitrary units	Manual measurement	1
Protein					
Lumbar		15–45 mg/dl	0.15–0.45 g/L	Turbidometry	0.01
Cisternal		15–25 mg/dl	0.15–0.25 g/L		
Ventricular		5–15 mg/dl	0.05–0.15 g/L		
Ceruloplasmin	S	27–50 mg/dl	270–500 mg/L	Nephelometry	10
Chloride	P	100–108 mmol/L	100–108 mmol/L	Coulometry	1
	U	Depends on diet	Depends on diet		
Cholesterol	S			Colorimetry	0.02586
Desirable		< 200 mg/dl	< 5.17 mmol/L		
Borderline high		200–239 mg/dl	5.17–6.18 mmol/L		
High		> 239 mg/dl	> 6.18 mmol/L		
Cortisol	S			Immunoassay	27.59
Fasting, 8 A.M.–12 P.M.		5–25 µg/dl	138–690 nmol/L		
12 P.M.–8 P.M.		5–15 µg/dl	138–414 nmol/L		
8 P.M.–8 A.M.		0–10 µg/dl	0–276 nmol/L		
Cortisol, free	U	20–70 µg/24 hr	55–193 nmol/24 hr	Immunoassay	2.759
Creatine kinase (CK)	S			Kinetic method	0.01667
Female		40–150 U/L	0.67–2.50 µkat/L		
Male		60–400 U/L	1.00–6.67 µkat/L		
Creatine kinase isoenzyme index	S	0–2.5% relative index	None	$\frac{\text{ng/ml}}{\text{Total CK(U/L)} \times 100}$	None
Creatine kinase isoenzymes, MB fraction	S	0–5 ng/ml	0–5 µg/L	Immunoassay	1
Creatinine	P	0.6–1.5 mg/dl	53–133 µmol/L	Colorimetry	88.4
	U	15–25 mg/kg/day	0.13–0.22 mmol/kg/day		0.0884
Dehydroepiandrosterone (DHEA) (adult)	S			Immunoassay	0.03467
Female		130–980 ng/dl	4.5–34.0 nmol/L		
Male		180–1,250 ng/dl	6.24–43.3 nmol/L		
Dehydroepiandrosterone (DHEA) sulfate (adult)	S			Immunoassay	10
Female					
Premenopausal		12–535 µg/dl	120–5,350 µg/L		
Postmenopausal		30–260 µg/dl	300–2,600 µg/L		
Male		10–619 µg/dl	100–6,190 µg/L		
Deoxycorticosterone (DOC) (adult)	S	2–19 ng/dl	61–576 nmol/L	Immunoassay	30.3
11-Deoxycortisol (adult) (8 A.M. sample)	S	12–158 ng/dl	0.35–4.56 nmol/L	Immunoassay	0.02886
1,25-Dihydroxyvitamin D	S	18–62 pg/ml	43.2–148.8 pmol/L	Immunoassay	2.4
Estradiol	S, P			Immunoassay	3.671
Female					

Chemistry *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Menstruating					
Follicular phase		50–145 pg/ml	184–532 pmol/L		
Midcycle peak		112–443 pg/ml	411–1,626 pmol/L		
Luteal phase		50–241 pg/ml	184–885 pmol/L		
Postmenopausal		< 59 pg/ml	< 217 pmol/L		
Male		< 50 pg/ml	< 184 pmol/L		
Fatty acids, free (adult)	S	0.17–0.95 mmol/L	0.17–0.95 mmol/L	Spectrophotometry	1
Follicle-stimulating hormone (FSH)	S, P			Immunoassay	1
Female					
Menstruating					
Follicular phase		3.0–20.0 U/L	3.0–20.0 U/L		
Ovulatory phase		9.0–26.0 U/L	9.0–26.0 U/L		
Luteal phase		1.0–12.0 U/L	1.0–12.0 U/L		
Postmenopausal		18.0–153.0 U/L	18.0–153.0 U/L		
Male		1.0–12.0 U/L	1.0–12.0 U/L		
Gastrin	S	< 100 ng/L	< 100 ng/L	Immunoassay	1
Globulin	S	2.6–4.1 g/dl	26–41 g/L	Calculation: total protein – albumin	10
Glucagon	P	20–100 pg/ml	20–100 ng/L	Immunoassay	1
Glucose	U	< 0.05 g/dl	< 0.003 mmol/L	Enzymatic analysis	0.05551
Glucose, fasting	P	70–110 mg/dl	3.9–6.1 mmol/L	Enzymatic analysis	0.05551
γ-Glutamyltransferase (GGT)	S			Spectrophotometry	1
Female		1–70 U/L	1–70 U/L		
Male		1–94 U/L	1–94 U/L		
Growth hormone (resting)	S	2–5 ng/ml	2–5 μg/L	Immunoassay	1
Hemoglobin A_{1C}	WB	3.8%–6.4%	0.038–0.064	Liquid chromatography	0.01
High-density lipoprotein cholesterol, as major risk factor	S	< 35 mg/dl	< 0.91 mmol/L	Colorimetry	0.02586
Human chorionic gonadotropin (hCG) (nonpregnant women)	S	< 5 mIU/ml	< 5 IU/L	Immunoassay	1
5-Hydroxyindoleacetic acid (5-HIAA)	U	≤ 6 mg/24 hr	≤ 31.4 μmol/day	High-pressure liquid chromatography	5.23
17-Hydroxyprogesterone (adult)	S			Immunoassay	0.03
Female					
Follicular phase		20–100 ng/dl	0.6–3.0 nmol/L		
Midcycle peak		100–250 ng/dl	3–7.5 nmol/L		
Luteal phase		100–500 ng/dl	3–15 nmol/L		
Postmenopausal		≤ 70 ng/dl	≤ 2.1 nmol/L		
Male		5–250 ng/dl	0.15–7.5 nmol/L		
25-Hydroxyvitamin D	S	8–42 ng/ml	20–105 nmol/L	Immunoassay	2.496
Insulin	S, P	2–20 μU/ml	14.35–143.5 pmol/L	Immunoassay	7.175
Ketone (acetone)	S, U	Negative	Negative	Colorimetry (nitroprusside)	
17-Ketosteroids	U			Modified Zimmerman reaction	3.467
Female		5–15 mg/24 hr	17.3–52.0 μmol/24 hr		
Male		7–20 mg/24 hr	24.3–69.3 μmol/24 hr		
Lactic acid	P	0.5–2.2 mmol/L	0.5–2.2 mmol/L	Enzymatic analysis	1
Lactate dehydrogenase (LDH)	S	110–210 U/L	1.83–3.50 μkat/L	Kinetic method	0.01667
Lactate dehydrogenase isoenzymes	S				
LD_1		16%–29%	0.16–0.29	Electrophoresis	0.01
LD_2		30%–41%	0.30–0.41	Electrophoresis	0.01
LD_3		15%–24%	0.15–0.24	Electrophoresis	0.01
LD_4		6%–13%	0.06–0.13	Electrophoresis	0.01
LD_5		5%–29%	0.05–0.29	Electrophoresis	0.01
Total LDH (when isoenzymes determined)		90–250 U/L	1.5–4.17 μkat/L	Kinetic method	0.01667
Lipase	S	3–19 U/dl	0.5–3.17 μkat/L	Kinetic method	0.1667
Lipoprotein(a)	S	0–30 mg/dl	0–300 mg/L	Nephelometry	10
Low-density lipoprotein cholesterol	S			Calculation	0.02586
Desirable		< 130 mg/dl	< 3.36 mmol/L		
Borderline high risk		130–159 mg/dl	3.36–4.11 mmol/L		

Chemistry *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
High risk		≥ 160 mg/dl	≥ 4.13 mmol/L		
Luteinizing hormone (LH)	S, P			Immunoassay	1
Female					
Menstruating					
Follicular phase		2.0–15.0 U/L	2.0–15.0 U/L		
Ovulatory phase		22.0–105.0 U/L	22.0–105.0 U/L		
Luteal phase		0.6–19.0 U/L	0.6–19.0 U/L		
Postmenopausal		16.0–64.0 U/L	16.0–64.0 U/L		
Male		2.0–12.0 U/L	2.0–12.0 U/L		
Magnesium	S	1.4–2.0 mEq/L	0.7–1.0 mmol/L	Colorimetry	0.5
Metanephrines	U			Chromatography	
Metanephrine		45–290 μg/24 hr	245–1,583 nmol/24 hr		5.458
Normetanephrine		82–500 μg/24 hr	448–2,730 nmol/24 hr		5.46
Total		120–700 μg/24 hr	655–3,821 nmol/24 hr		5.458
Methemoglobin	P	0.4%–1.5% of total hemoglobin	0.004–0.015	Co-oximetry	0.01
Microalbumin, random urine	U	< 20 μg/ml	< 20 mg/L	Nephelometry	1
5′-Nucleotidase	S	0–11 U/L	0.02–0.18 μkat/L	Kinetic method	0.01667
Osmolality	S, P	280–296 mOsm/kg of water	280–296 mmol/kg of water	Freezing-point depression	1
Oxygen, partial pressure, arterial (P_aO_2) (room air, age dependent)	WB	80–100 mm Hg	10.7–13.3 kPa	Oxygen electrode	0.1333
Parathyroid hormone	S	10–60 pg/ml	10–60 ng/L	Immunoassay	1
Parathyroid hormone–related protein	P	< 1.3 pmol/L	< 1.3 pmol/L	Immunoassay	1
pH, arterial	WB	7.35–7.45 pH units	7.35–7.45 pH units	pH electrode	1
Phosphorus, inorganic (adult)	S	2.6–4.5 mg/dl	0.84–1.45 mmol/L	Spectrophotometry	0.3229
	U	average, 1 g/day	average, 32 mmol/day		32.29
Potassium	P	3.4–4.8 mmol/L	3.4–4.8 mmol/L	Ion-selective electrode	1
	S	3.5–5.0 mmol/L	3.5–5.0 mmol/L		
	U	Depends on diet	Depends on diet		
Prealbumin (adult)	S	19.5–35.8 mg/dl	195–358 mg/L	Nephelometry	10
Progesterone	S,P			Immunoassay	3.18
Female					
Follicular phase		< 1.0 ng/ml	< 3.18 nmol/L		
Midluteal phase		3–20 ng/ml	9.54–63.6 nmol/L		
Male		< 1.0 ng/ml	< 3.18 nmol/L		
Prolactin	S			Immunoassay	1
Female					
Premenopausal		0–20 ng/ml	0–20 μg/L		
Postmenopausal		0–15 ng/ml	0–15 μg/L		
Male		0–15 ng/ml	0–15 μg/L		
Prostate-specific antigen (PSA)	S			Immunoassay	1
Female		< 0.5 ng/ml	< 0.5 μg/L		
Male					
< 40 yr		0.0–2.0 ng/ml	0.0–2.0 μg/L		
≥ 40 yr		0.0–4.0 ng/ml	0.0–4.0 μg/L		
Prostate-specific antigen (PSA), free, in males 45–75 yr, with PSA values between 4 and 20 ng/ml	S	> 25% associated with benign prostatic hyperplasia	> 0.25 associated with benign prostatic hyperplasia	Immunoassay, calculation	0.01
Protein, total	S	6.0–8.0 g/dl	60–80 g/L	Colorimetry	10
	U	< 165 mg/day	< 0.165 g/day	Turbidometry	0.001
Renin (adult, normal-sodium diet)	P			Immunoassay	0.2778
Supine		0.3–3.0 ng/ml/hr	0.08–0.83 ng/(L · sec)		
Upright		1.0–9.0 mg/ml/hr	0.28–2.5 ng/(L · sec)		
Serotonin	WB	55–260 ng/ml	0.31–1.48 μmol/L	High-pressure liquid chromatography	0.00568
Sex hormone–binding globulin (adult)	S			Immunoassay	1
Female		8–85 mmol/L	8–85 mmol/L		
Male		6–44 mmol/L	6–44 mmol/L		
Sodium	P	135–145 mmol/L	135–145 mmol/L	Ion-selective electrode	1
	U	Depends on diet	Depends on diet		
Somatomedin C (insulinlike growth factor I)	S			Immunoassay	1

Chemistry *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
16–24 yr		182–780 ng/ml	182–780 μg/L		
25–39 yr		114–492 ng/ml	114–492 μg/L		
40–54 yr		90–360 ng/ml	90–360 μg/L		
> 54 yr		71–290 ng/ml	71–290 μg/L		
Testosterone, total (morning sample)	S			Immunoassay	0.03467
Female		6–86 ng/dl	0.21–2.98 nmol/L		
Male		270–1,070 ng/dl	9.36–37.10 nmol/L		
Testosterone, unbound (morning sample)	S			Immunoassay	34.67
Female					
20–40 yr		0.6–3.1 pg/ml	20.8–107.5 pmol/L		
41–60 yr		0.4–2.5 pg/ml	13.9–86.7 pmol/L		
61–80 yr		0.2–2.0 pg/ml	6.9–69.3 pmol/L		
Male					
20–40 yr		15.0–40.0 pg/ml	520–1,387 pmol/L		
41–60 yr		13.0–35.0 pg/ml	451–1,213 pmol/L		
61–80 yr		12.0–28.0 pg/ml	416–971 pmol/L		
Thyroglobulin	S	0–60 ng/ml	0–60 μg/L	Immunoassay	1
Thyroid hormone–binding index (THBI; T_3RU)	S	0.77–1.23	0.77–1.23	Immunoassay	1
Thyroid-stimulating hormone	S	0.5–5.0 μU/ml	0.5–5.0 μU/ml	Immunoassay	1
Thyroxine, total (T_4)	S	4.5–10.9 μg/dl	58–140 nmol/L	Immunoassay	12.87
Transferrin	S	191–365 mg/dl	1.91–3.65 g/L	Nephelometry	0.01
Triglycerides (fasting)	S	40–150 mg/dl	0.45–1.69 mmol/L	Spectrophotometry	0.01129
Triiodothyronine, total (T_3)	S	60–181 ng/dl	0.92–2.78 nmol/L	Immunoassay	0.01536
Troponin I	S	< 0.6 ng/ml >1.5 ng/ml consistent with acute myocardial infarct	< 0.6 μg/L > 1.5 μg/L	Immunoassay	1
Urea nitrogen (BUN) (adult)	P	8–25 mg/dl	2.9–8.9 mmol/L	Conductivity	0.357
Urea nitrogen, urine	U	6–17 g/day	6–17 g/day	Conductivity	1
Uric acid	S			Colorimetry	59.48
Female		2.3–6.6 mg/dl	137–393 μmol/L		
Male		3.6–8.5 mg/dl	214–506 μmol/L		
Urinalysis	U			Reflectance	
pH		5.0–9.0	5.0–9.0	Spectrophotometry	1
Specific gravity		1.001–1.035	1.001–1.035		1
Chemical screens		Negative	Negative		
Urine sediment	U			Manual method	
White cells		0–2/high-power field	0–2/high-power field		1
Red cells		0–2/high-power field	0–2/high-power field		1
Vasoactive intestinal polypeptide (VIP)	P	< 75 pg/ml	< 75 ng/L	Immunoassay	1
Xylose	U	4–9 g/5 hr	4–9 g/5 hr	Colorimetry	1
	S (fasting)	None detected	None detected		

Toxicology and Therapeutic Drug Monitoring

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Acetaminophen, toxicity	S	> 120 μg/ml at 2–4 hr	> 794 μmol/L at 2–4 hr	Liquid chromatography	6.62
Amikacin	S			Immunoassay	1.71
Trough		1–7 μg/ml	1.7–12 μmol/L		
Peak		15–25 μg/ml	26–43 μmol/L		
Amitriptyline	S			Liquid chromatography	3.61
Amitriptyline		120–250 μg/L	433–903 nmol/L		
Nortriptyline		50–150 μg/L	181–542 nmol/L		
Barbiturate screen	S	Negative	Negative	Liquid chromatography	
Basic drug screen	S	Negative	Negative	Liquid chromatography	
Benzodiazepine screen	S	Negative	Negative	Liquid chromatography	
Carbamazepine (adult)	S	4–12 μg/ml	17–51 μmol/L	Immunoassay	4.23
Chlordiazepoxide	S	1.0–3.0 mg/L	3.3–10 μmol/L	Liquid chromatography	3.336
Clomipramine	S	150–450 μg/L	476–1,427 nmol/L	Liquid chromatography	3.17
Clonazepam	S	10–70 μg/L	32–222 nmol/L	Liquid chromatography	3.17
Clozapine	S			Liquid chromatography	
Clozapine		150–500 μg/L	459–1,530 nmol/L		3.06
Norclozapine		100–450 μg/L	320–1,440 nmol/L		3.20
Desipramine	S	150–300 μg/L	563–1,125 nmol/L	Liquid chromatography	3.75
Diazepam	S	100–1,000 μg/L	0.35–3.50 μmol/L	Liquid chromatography	0.0035
Digoxin	S	0.9–2.0 ng/ml	1.2–2.6 nmol/L	Immunoassay	1.28
Doxepin	S	150–250 μg/L	537–895 nmol/L	Liquid chromatography	3.58
Ethanol	S	Clinical intoxication: > 1,000 mg/L	Clinical intoxication: > 1 g/L	Gas chromatography	0.001
Fluoxetine	S			Liquid chromatography	
Fluoxetine		50–450 μg/L	162–1,454 nmol/L		3.23
Norfluoxetine		50–450 μg/L	169–1,521 nmol/L		3.38
Fluvoxamine	S	12–240 μg/L	38–756 nmol/L	Liquid chromatography	3.15
Gentamicin	S			Immunoassay	
Trough		< 2.1 μg/ml	< 4.4 μmol/L		2.09
Peak		4–8 μg/ml	8.4–16.7 μmol/L		2.09
Imipramine	S			Liquid chromatography	
Imipramine		150–250 μg/L	536–893 nmol/L		3.57
Desipramine		150–300 μg/L	563–1,125 nmol/L		3.75
Lithium	S, P	0.5–1.5 mmol/L	0.5–1.5 mmol/L	Ion-selective electrode	1
Methadone	S	50–800 μg/L	0.16–2.58 μmol/L	Liquid chromatography	0.00323
Methotrexate	S			Immunoassay	1
24 hr after high-dose infusion		< 10 μmol/L	< 10 μmol/L		
48 hr after high-dose infusion		< 1 μmol/L	< 1 μmol/L		
72 hr after high-dose infusion		< 0.4 μmol/L	< 0.4 μmol/L		
Nordiazepam	S	100–800 μg/L	0.37–2.95 μmol/L	Liquid chromatography	0.00369
Paroxetine	S	20–190 μg/L	61–576 nmol/L	Liquid chromatography	3.03
Phenobarbital	S	15–50 μg/ml	65–216 μmol/L	Immunoassay	4.31
Phenytoin	S	5–20 μg/ml	20–79 μmol/L	Immunoassay	3.96
Quinidine	S	1.2–4.0 μg/ml	3.7–12.3 μmol/L	Liquid chromatography	3.08
Salicylate intoxication		> 500 mg/L	> 3.62 mmol/L		0.00724

Specimen abbreviations: P—plasma S—serum

Toxicology and Therapeutic Drug Monitoring *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Sertraline	S	40–160 µg/L	130–522 nmol/L	Liquid chromatography	3.26
Sulfamethoxazole	S	5–15 mg/dl	19.8–59.2 nmol/L	Liquid chromatography	3.95
Theophylline	S	10–20 µg/ml	56–111 µmol/L	Immunoassay	5.55
For apnea control		5–15 µg/ml	28–83 µmol/L		
Thiocyanate, toxic range	S	> 100 mg/L	> 1,720 µmol/L	Spectrophotometry	17.2
Tobramycin	S			Immunoassay	2.14
Trough		< 2.0 µg/ml	< 4.3 µmol/L		
Peak		4.0–8.0 µg/ml	8.6–17.1 µmol/L		
Trazodone	S	800–1,600 µg/L	2,152–4,304 nmol/L	Liquid chromatography	2.69
Valproic acid	S	50–100 µg/ml	347–693 µmol/L	Immunoassay	6.93
Vancomycin	S			Immunoassay	0.690
Trough		< 10.1 µg/ml	< 7.0 µmol/L		
Peak (2 hr after infusion)		18–26 µg/ml	12–18 µmol/L		

Immunology

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Alpha$_1$-antitrypsin (adult)	S	76–189 mg/dl	0.76–1.89 g/L	Nephelometry	0.01
Anti–glomerular basement membrane antibodies	S				
Qualitative		Negative	Negative	Western blot assay	
Quantitative		< 5 U/ml	< 5 kU/L	ELISA	1
Antineutrophil cytoplasmic autoantibodies, cytoplasmic (c-ANCA)	S				
Qualitative		Negative	Negative	Indirect immunofluorescence	
Quantitative (antibodies to proteinase 3)		< 2.8 U/ml	< 2.8 kU/L	ELISA	1
Antineutrophil cytoplasmic autoantibodies, perinuclear (p-ANCA)	S				
Qualitative		Negative	Negative	Indirect immunofluorescence	
Quantitative (antibodies to myeloperoxidase)		< 1.4 U/ml	< 1.4 kU/L	ELISA	1
Autoantibodies					
Antiadrenal antibody	S	Negative at 1:10 dilution	Not applicable	Indirect immunofluorescence	
Anti–double-stranded (native) DNA	S	Negative at 1:10 dilution	Not applicable	Indirect immunofluorescence	
Antigranulocyte antibody	S	Negative	Not applicable	ELISA	
Anti–Jo-1 antibody	S	Negative	Not applicable	ELISA	
Anti–La antibody	S	Negative	Not applicable	ELISA	
Antimitochondrial antibody	S	Negative	Not applicable	Indirect immunofluorescence	
Antinuclear antibody	S	Negative at 1:40 dilution	Not applicable	Indirect immunofluorescence	
Anti–parietal cell antibody	S	Negative at 1:20 dilution	Not applicable	Indirect immunofluorescence	
Anti-Ro antibody	S	Negative	Not applicable	ELISA	
Anti-RNP antibody	S	Negative	Not applicable	ELISA	
Anti–Scl-70 antibody	S	Negative	Not applicable	ELISA	
Anti-Smith antibody	S	Negative	Not applicable	ELISA	
Anti–smooth muscle antibody	S	Negative at 1:20 dilution	Not applicable	Indirect immunofluorescence	
Antithyroglobulin antibody	S	Negative	Not applicable	Immunoassay	
Antithyroid antibody	S	< 0.3 IU/ml	< 0.3 kIU/L	Immunoassay	1
Bence Jones protein	S	None detected	Not applicable	Immunoelectrophoresis	
Qualitative	U	None detected in a 50-fold concentration	Not applicable	Immunoelectrophoresis	
Quantitative	U			Nephelometry	
Kappa		< 2.5 mg/dl	< 0.03 g/L		0.01
Lambda		< 5.0 mg/dl	< 0.05 g/L		0.01
C1 esterase inhibitor protein	S				
Antigenic		12.4–24.5 mg/dl	0.12–0.25 g/L	Nephelometry	0.01
Functional		Present	Present	Immunodiffusion	
Complement					
C3 (adult)	S	86–184 mg/dl	0.86–1.84 g/L	Nephelometry	0.01
C4 (adult)	S	20–58 mg/dl	0.20–0.58 g/L	Nephelometry	0.01
Total complement (adult)	S	63–145 U/ml	63–145 kU/L	Immunoassay	1
Factor B	S	17–42 mg/dl	0.17–0.42 g/L	Nephelometry	0.01
Cryocrit	S	None detected	Not applicable	Manual method	
Cryoproteins	S	None detected	Not applicable	Manual method	
Cryoprotein identification	S	None detected	Not applicable	Immunodiffusion	
CSF	CSF				
Agarose electrophoresis		No banding seen in an 80-fold concentration	Not applicable	Agarose electrophoresis	
Quantitation of albumin (adult)		11.0–50.9 mg/dl	0.11–0.51 g/L	Nephelometry	0.01
Quantitation of IgG (adult)		0.0–8.0 mg/dl	0.0–0.08 g/L	Nephelometry	0.01
Haptoglobin	S	16–199 mg/dl	0.16–1.99 g/L	Nephelometry	0.01
Immunofixation	S	None detected	Not applicable	Agarose electrophoresis	
Immunoglobulin (adult)					
IgA	S	60–309 mg/dl	0.60–3.09 g/L	Nephelometry	0.01

Specimen abbreviations: CSF—cerebrospinal fluid ELISA—enzyme-linked immunosorbent assay JF—joint fluid S—serum

Immunology *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
IgE	S	10–179 IU/ml	24–430 µg/L	Immunoassay	2.4
IgG	S	614–1,295 mg/dl	6.14–12.95 g/L	Nephelometry	0.01
IgM	S	53–334 mg/dl	0.53–3.34 g/L	Nephelometry	0.01
Joint fluid crystal	JF	No crystals seen	Not applicable	Microscopy	
Joint fluid mucin	JF	Only type I mucin present	Not applicable	Manual method	
Rheumatoid factor	S, JF	< 30 IU/ml	< 30 kIU/L	Nephelometry	1
Serum protein electrophoresis	S	Normal pattern	Not applicable	Agarose electrophoresis	
Viscosity	S	1.4–1.8 relative viscosity units, as compared with water	1.4–1.8 relative viscosity units, as compared with water	Ostwald viscosimetry	1

Hematology and Coagulation

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Activated protein C resistance (factor V Leiden)	P	Ratio > 2.0	Not applicable	Automated clotting assay	
$Alpha_2$-antiplasmin	P	80%–130%	0.80–1.30	Chromogenic assay	0.01
Antiphospholipid-antibody panel					
Partial thromboplastin time–lupus anticoagulant screen	P	Negative	Negative	Dilute phospholipid clotting assay	
Platelet-neutralization procedure	P	Negative	Negative	Clotting assay	
Anticardiolipin antibody	S				
IgG		0–15 GPL units	0–15 arbitrary units	ELISA	1
IgM		0–15 MPL units	0–15 arbitrary units	ELISA	1
Antithrombin III	P				
Immunologic		22–39 mg/dl	220–390 mg/L	Immunoassay	10
Functional		80%–130%	0.8–1.30 U/L	Chromogenic assay	0.01
Anti–Xa assay (heparin assay)	P			Chromogenic assay	
Unfractionated heparin		0.3–0.7 IU/ml	0.3–0.7 kIU/L		1
Low-molecular-weight heparin		0.5–1.0 IU/ml	0.5–1.0 kIU/L		1
Danaparoid		0.5–0.8 IU/ml	0.5–0.8 kIU/L		1
Bleeding time (adult)		2–9.5 min	2–9.5 min	Surgicutt	1
Clot retraction	WB	50%–100%/2 hr	0.50–1.00/2 hr	Manual method	0.01
D-Dimer	P	< 0.5 μg/ml	< 0.5 mg/L	Latex agglutination	1
Differential blood count	WB			Automated cell counter or manual method	
Neutrophils		45%–75%	0.45–0.75		0.01
Bands		0%–5%	0.0–0.05		0.01
Lymphocytes		16%–46%	0.16–0.46		0.01
Monocytes		4%–11%	0.04–0.11		0.01
Eosinophils		0%–8%	0.0–0.8		0.01
Basophils		0%–3%	0.0–0.03		0.01
Erythrocyte count (adult)	WB			Automated cell counter	
Female		$4.10–5.10 \times 10^6/mm^3$	$4.10–5.10 \times 10^{12}/L$		1×10^6
Male		$4.50–5.30 \times 10^6/mm^3$	$4.50–5.30 \times 10^{12}/L$		1×10^6
Erythrocyte sedimentation rate	WB			Automated erythrocyte-sedimentation-rate instrument	1
Female		1–25 mm/hr	1–25 mm/hr		
Male		0–17 mm/hr	0–17 mm/hr		
Factor II, prothrombin	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor V	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor VII	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor VIII	P	50%–200%	0.50–2.00	Automated clotting assay	0.01
Factor IX	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor X	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor XI	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor XII	P	60%–140%	0.60–1.40	Automated clotting assay	0.01
Factor XIII screen	P	No deficiency detected	Not applicable	Urea clot dissolution	
Factor-inhibitor assay	P	< 0.5 Bethesda unit	< 0.5 Bethesda unit	Automated clotting assay	1
Ferritin	S			Immunoassay	1
Female		10–200 ng/ml	10–200 μg/L		
Male		30–300 ng/ml	30–300 μg/L		
Fibrin(ogen)-degradation products	P	< 2.5 μg/ml	< 2.5 mg/L	Latex agglutination	1
Fibrinogen	P	175–400 mg/dl	1.75–4.00 μmol/L	Automated clotting assay	0.01
Folate (folic acid)	S, P			Immunoassay	2.266
Normal		3.1–17.5 ng/ml	7.0–39.7 nmol/L		
Borderline deficient		2.2–3.0 ng/ml	5.0–6.8 nmol/L		
Deficient		< 2.2 ng/ml	< 5.0 nmol/L		
Excessive		> 17.5 ng/ml	> 39.7 nmol/L		
Glucose-6-phosphate dehydrogenase (erythrocyte)	WB	No gross deficiency	Not applicable	Visual fluorescence screening	
Hematocrit (adult)	WB			Automated cell counter	
Female		36.0–46.0	0.36–0.46		0.01
Male		37.0–49.0	0.37–0.49		0.01
Hemoglobin (adult)	WB			Automated cell counter	0.6206
Female		12.0–16.0 g/dl	7.4–9.9 mmol/L		

Specimen abbreviations: ELISA—enzyme-linked immunosorbent assay GPL—IgG phospholipid MPL—IgM phospholipid P—plasma PRP—platelet-rich plasma S—serum WB—whole blood

Hematology and Coagulation *(continued)*

Analyte	*Specimen*	*Conventional Units*	*SI Units*	*Method or Instrument*	*Factor for Conversion to SI Units*
Male		13.0–18.0 g/dl	8.1–11.2 mmol/L		
Hemoglobin A_2	WB	< 3.5%	< 0.04	Spectrophotometry	0.01
Hemoglobin F	WB	< 0.02%	< 0.0002	Spectrophotometry	0.01
Heparin-induced thrombocytopenia antibody	P	Negative	Negative	ELISA	
Iron	S	30–160 μg/dl	5.4–28.7 μmol/L	Colorimetry	0.1791
Iron-binding capacity	S	228–428 μg/dl	40.8–76.7 μmol/L	Colorimetry	0.1791
Leukocyte count (WBC)	WB	$4.5–11.0 \times 10^3/mm^3$	$4.5–11.0 \times 10^9/L$	Automated cell counter	1×10^6
Mean corpuscular hemoglobin (MCH)	WB	25.0–35.0 pg/cell	25.0–35.0 pg/cell	Automated cell counter	1
Mean corpuscular hemoglobin concentration (MCHC)	WB	31.0–37.0 g/dl	310–370 g/L	Automated cell counter	10
Mean corpuscular volume (MCV) (adult)	WB			Automated cell counter	1
Female		78–102 μm^3	78–102 fl		
Male		78–100 μm^3	78–100 fl		
Osmotic fragility of erythrocytes	WB	Increased hemolysis as compared with normal control	Not applicable	Spectrophotometry	
Partial thromboplastin time, activated	P	22.1–34.1 sec	22.1–34.1 sec	Automated clotting assay	1
Plasminogen	P				
Antigen		8.4–14.0 mg/dl	84–140 mg/L	Immunoassay	10
Functional		80%–130%	0.80–1.30	Chromogenic assay	0.01
Platelet aggregation	PRP	> 65% aggregation in response to adenosine diphosphate, epinephrine, collagen, ristocetin, and arachidonic acid	Not applicable	Platelet aggregometry	
Platelet count	WB	$150–350 \times 10^3/mm^3$	$150–350 \times 10^9/L$	Automated cell counter	1×10^6
Platelet, mean volume	WB	6.4–11.0 μm^3	6.4–11.0 fl	Automated cell counter	1
Prekallikrein assay	P	60%–140%	0.60–1.40	Chromogenic assay	0.01
Prekallikrein screen	P	No deficiency detected		Automated clotting assay	
Protein C	P				
Total antigen		70%–140%	0.70–1.40	Immunoassay	0.01
Functional		70%–140%	0.70–1.40	Chromogenic assay	0.01
Protein S	P				
Total antigen		70%–140%	0.70–1.40	Immunoassay	0.01
Functional		70%–140%	0.70–1.40	Automated clotting assay	0.01
Free antigen		70%–140%	0.70–1.40	Immunoassay	0.01
Prothrombin time	P	11.2–13.2 sec	11.2–13.2 sec	Automated clotting assay	1
Red cell distribution width	WB	11.5%–14.5%	0.115–0.145	Automated cell counter	0.01
Reptilase time	P	16–24 sec	16–24 sec	Automated clotting assay	1
Reticulocyte count	WB	0.5%–2.5% red cells	0.005–0.025 red cells	Flow cytometry	0.01
Ristocetin cofactor (functional von Willebrand factor)	P			Platelet aggregometry	0.01
Blood group O		75% mean of normal	0.75 mean of normal		
Blood group A		105% mean of normal	1.05 mean of normal		
Blood group B		115% mean of normal	1.15 mean of normal		
Blood group AB		125% mean of normal	1.25 mean of normal		
Sucrose hemolysis	WB	< 10%	< 0.1	Spectrophotometry	0.01
Thrombin time	P	16–24 sec	16–24 sec	Automated clotting assay	1
Total eosinophils	WB	$70–440/mm^3$	$70–440 \times 10^6/L$	Automated cell counter	1×10^3
Vitamin B_{12}	S, P			Immunoassay	0.7378
Normal		> 250 pg/ml	> 184 pmol/L		
Borderline		125–250 pg/ml	92–184 pmol/L		
Deficient		< 125 pg/ml	< 92 pmol/L		
von Willebrand factor (vWF) antigen (factor VIII:R antigen)	P			Immunoassay	
Blood group O		75% mean of normal	0.75 mean of normal		0.01
Blood group A		105% mean of normal	1.05 mean of normal		0.01
Blood group B		115% mean of normal	1.15 mean of normal		0.01
Blood group AB		125% mean of normal	1.25 mean of normal		0.01
von Willebrand factor multimers	P	Normal distribution	Normal distribution	Western blot assay	

Microbiology

Culture Specimen	*Routinely Cultured for*	*Also Reported*	*Normal Flora*
Throat	Group A β-hemolytic streptococci, pyogenic groups C and G β-hemolytic streptococci, *Arcanobacterium haemolyticum*	If a complete throat culture is requested, it will be examined for *Haemophilus influenzae, Staphylococcus aureus, Streptococcus pneumoniae, Neisseria meningitidis,* and yeast	α-Hemolytic streptococci, nonhemolytic streptococci, diphtheroids, coagulase-negative staphylococci, saprophytic *Neisseria*
Sputum	Pneumococci, *H. influenzae,* β-hemolytic streptococci, *S. aureus, Moraxella* (*Branhamella*) *catarrhalis, Pseudomonas,* Enterobacteriaceae, yeast	Presence or absence of normal throat flora	Carefully collected specimens should contain few or no normal throat flora
Urine	Aerobic bacteria and yeast: abundant, $> 10^5$ colony-forming units/ml; moderate, 10^4–10^5 colony-forming units/ml	Few, 10^3–10^4 colony-forming units/ml; rare, 10^2–10^3 colony-forming units/ml; no growth, $< 10^2$ colony-forming units/ml; these amounts may indicate clinicaly significant bacteriuria if accompanied by pyuria, clinical symptoms, or both	Carefully collected specimens should not contain mixed bacterial species (i.e., two or more of the following: lactobacilli, non–β-hemolytic streptococci, diphtheroids, coagulase-negative staphylococci, *Gardnerella vaginalis*)
Blood	Aerobic and anaerobic bacteria, yeasts	Growth in both bottles is usually more clinically significant than growth in a single bottle	None; aerobic and anaerobic diphtheroids and coagulase-negative staphylococci are common contaminants
Cerebrospinal and other fluids	Aerobic and anaerobic bacteria; yeasts, including *Cryptococcus*	Any organism isolated	None
Feces	Enteric pathogens: *Salmonella, Shigella, Campylobacter, Plesiomonas,* and *Aeromonas* when predominant	Moderate or abundant yeast or *S. aureus*; presence or absence of normal gram-negative enteric flora; special cultures can be specifically requested for *Yersinia, Vibrio cholerae, V. parahaemolyticus,* or hemorrhagic (O157) strains of *Escherichia coli*	Enterobacteriaceae, streptococci, *Pseudomonas,* small numbers of staphylococci, and yeast (and anaerobes that are not cultured routinely)
Wounds	Aerobic and anaerobic bacteria, yeasts		
Cervical or vaginal	Gonococci, β-hemolytic streptococci, *S. aureus,* and *G. vaginalis* when predominant	Enteric gram-negative rods and *Candida,* if present in large numbers	

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 29 No. 4 April 2006

What's New in ACP Medicine

DAVID C. DALE, MD, FACP, *Editor-in-Chief* DANIEL D. FEDERMAN, MD, MACP, *Founding Editor*
WENDY LEVINSON, MD, FACP, *Associate Medical Editor, What's New in ACP Medicine*

www.acpmedicine.com

PRACTICE OF MEDICINE

Update on Influenza: The Bird Flu Threat

JOHN T. BATTAILE, MD
MICHAEL J. HOLTZMAN, MD
Washington University School of Medicine

Influenza viruses are major causes of morbidity and mortality throughout the world. In the United States alone, between 1990 and 1999, an average of 36,000 deaths a year were attributed to influenza.[1] Recently, there has been renewed concern that influenza virus may cause pandemic disease. The capacity of influenza virus to cause such problems is based on variations in two viral envelope glycoproteins—hemagglutinin and neuraminidase. Point mutations in the RNA genes encoding these proteins allow the virus to evade immunity and cause disease outbreaks in susceptible populations on an almost yearly basis. However, more abrupt, major changes in these proteins, so-called antigenic shifts, can give rise to strains to which the majority of humans are immunologically naive. Pandemics can result if the new virus is also capable of efficient person-to-person transmission.

There are at least two major ways for influenza viruses to develop new pathogenic capabilities. One of these mechanisms is based on gene reassortment events, which was likely the basis for the 1957 Asian and 1968 Hong Kong influenza pandemics. Reassortment events occur when two different viral strains share genes while coinfecting the same cell. For example,

continued on page 2

THIS MONTH'S UPDATES

11 NEUROLOGY

VIII Headache

RANDOLPH W. EVANS, MD
University of Texas Medical School at Houston

Classifying Headache

The International Headache Society (IHS) criteria,[1] which were introduced in 1988 and updated in 2004, are the worldwide standard for headache classification. IHS criteria categorize headaches as primary or secondary. Primary headaches—those with no other underlying cause—account for 90% of headaches. This category includes migraine, tension, cluster, and miscellaneous headaches, such as primary exertional headaches. There are a large number of secondary headaches, which are classified according to their causes [*see Table, page 3*].

1. The International Classification of Headache Disorders: 2nd edition. Headache Classification Subcommittee of the International Headache Society. Cephalalgia 24(suppl 1):9, 2004 [PMID 14979299]

Triptans for Migraine

Over the past decade, seven triptans have become available in the United States: sumatriptan, zolmitriptan, naratriptan, rizatriptan, almotriptan, frovatriptan, and eletriptan [*see Table, page 4*].[1] In migraineurs who take oral triptans when their pain is moderate to severe in intensity, the 2-hour response rate (i.e., no pain or mild pain) is about 45% for naratriptan and frovatriptan and about 65% to 70% for the others. With all of the triptans, the 2-hour pain-free response rates are much higher if the drug is taken when the headache is mild; depending on the drug, the rate may exceed 70%.

The oral triptans may not be equally effective for all patients. If a patient has an unsatisfactory or inconsistent response, unpleasant side effects, or tachyphylaxis with one triptan, a different triptan may prove effective and tolerable. Patients who have prominent vomiting or nausea or who desire the quickest relief may benefit from subcutaneous sumatriptan (at 2 hours, 79% of patients show a response and 60% are pain free) or intranasal sumatriptan or zolmitriptan.

continued on page 3

In This Issue

PRACTICE OF MEDICINE

continued from page 1

avian influenza genes that confer pathogenicity can be transferred to a strain that is capable of human transmission. In addition to this mechanism, it has been proposed that the pandemic H1N1 strain responsible for the 1918 Spanish influenza pandemic was not the result of gene reassortment but rather emerged when an avian influenza virus adapted to spread to humans.[2] Sporadic human infections with avian influenza viruses, such as those being reported in Southeast Asia and elsewhere, serve to increase the likelihood of antigenic shift by creating the potential for gene reassortment or adaptation.

Recent attention has been focused on the pandemic potential for the avian H5N1 strain. Human cases have been reported in several Southeast Asian nations, including China, Thailand, and Vietnam, since 1997. More recent cases have occurred as far west as northern Iraq and Turkey. Migratory waterfowl have been implicated in the westward spread of the virus. In this case, migratory birds may have acquired the virus in Southeast Asia and then infected domestic birds along their migratory routes. The virus then spreads to humans via contact with infected domestic fowl. No sustained human transmission of the virus has been reported, but there have been isolated cases of human-to-human transmission. This phenomenon raises the concern that the virus could eventually acquire the ability to spread efficiently from person to person and spark a deadly pandemic.

The concern over avian flu also derives from the apparent severity of reported cases. Reported mortality for hospitalized patients with confirmed H5N1 infection has varied from 30% to over 80%.[3] In a series of 10 patients from Vietnam infected with the H5N1 virus, the majority required mechanical ventilation within 48 hours of hospital admission. [4] Perhaps most striking is the fact that many of the fatal H5N1 infections have occurred in otherwise healthy, young individuals. This finding has led to comparisons with the 1918 pandemic strain, which killed over 40 million people in a single year,[5] many of them young and previously healthy.

Controlling the next influenza pandemic will require immune and drug strategies. In that regard, work is under way for large-scale production of an effective vaccine. Vaccines with activity against other strains with pandemic potential, such as the H9N2 and H7N7 avian strains, are also being stockpiled in preparation for a pandemic. In individuals who are not effectively vaccinated, antiviral drug therapy may be useful for disease prevention and treatment. Two classes of drugs—adamantanes and neuraminidase inhibitors—have been used for human influenza outbreaks. The adamantanes amantadine and rimantadine are effective against influenza A but not influenza B and work to reduce viral replication by inhibiting uncoating of the virus. However, most avian influenza strains are resistant to the adamantanes. Even in susceptible strains, resistance can develop rapidly with treatment. In addition, resistance among circulating human influenza strains is increasingly common. For these and other reasons, neuraminidase inhibitors and not adamantanes should be used for human infections with avian or the novel flu strains. The neuraminidase inhibitors oseltamivir and zanamivir inhibit the release of virus from infected cells, thereby impeding viral spread. They are effective against both influenza A (including the H5N1 strain) and B and can be used for both disease prevention and treatment. Resistance is less common than with the adamantanes but has been reported in some strains, including recent reports of resistance in H5N1 isolates.[6] Higher doses of oseltamivir may be required to treat certain strains. To be effective, antiviral drug therapy for influenza should be initiated as soon as possible after symptom onset, ideally within the first 12 hours.[7] In addition to the pursuit of new antiviral agents with activity against influenza, other approaches to treatment are being investigated, such as strategies that would enable the infected host to more rapidly and efficiently clear viral infec-

ACP MEDICINE
A Publication of the American College of Physicians

Published by WebMD

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $349 for the first year ($324 for residents and students) and $269 for renewals ($249 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $449 for the first year and $369 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

PRACTICE OF MEDICINE

continued from page 2

tion by enhancing endogenous interferon efficacy.[8] Nonetheless, effective vaccination remains the cornerstone of current efforts to minimize the impact of influenza outbreaks, including future pandemics.

References

1. Thompson WW, Shay DK, Weintraub E, et al: Mortality associated with influenza and respiratory syncytial virus in the United States. JAMA 289:179, 2003 [PMID 12517228]

2. Taubenberger JK, Reid AH, Lourens RM, et al: Characterization of the 1918 influenza virus polymerase genes. Nature 437:889, 2005 [PMID 16208372]

3. Beigel JH, Farrar J, Han AM, et al: Avian influenza A (H5N1) infection in humans. N Engl J Med 353:1374, 2005 [PMID 16192482]

4. Tran TH, Nguyen TL, Nguyen TD, et al: Avian influenza A (H5N1) in 10 patients in Vietnam. N Engl J Med 350:1179, 2004 [PMID 14985470]

5. Palese P: Influenza: old and new threats. Nat Med 10(12 suppl):S82, 2004 [PMID 15577936]

6. Le QM, Kiso M, Someya K, et al: Avian flu: isolation of drug-resistant H5N1 virus. Nature 437:1108, 2005 [PMID 16228009]

7. Moscona A: Neuraminidase inhibitors for influenza. N Engl J Med 353:1363, 2005 [PMID 16192481]

8. Zhang Y, Takami K, Lo MS, et al: Modification of the Stat1 SH2 domain broadly improves interferon efficacy in proportion to p300/CREB-binding protein coactivator recruitment. J Biol Chem 280:34306, 2005 [PMID 16107341]

michaelholtzman@webmd.net

THIS MONTH'S UPDATES

continued from page 1

Patients may experience recurrence, which is defined as the return of headache (usually of moderate or severe intensity) within 24 hours after an initial response to acute treatment. When taken for moderate to severe pain, naratriptan, frovatriptan, almotriptan, and eletriptan have the lowest recurrence rates (about 20% to 25%). The recurrence rates for the other triptans are about 31% with zolmitriptan, 33% with sumatriptan, and 40% with rizatriptan. The time to recurrence is generally about 12 hours.[2]

1. Silberstein SD: Migraine. Lancet 363:381, 2004 [PMID 15070571]

2. Geraud G, Keywood C, Senard JM: Migraine headache recurrence: relationship to clinical, pharmacological, and pharmacokinetic properties of triptans. Headache 43:376, 2003 [PMID 12656709]

Best Medications for Migraine Prevention

Class I evidence indicates that the beta blocker propranolol, the tricyclic antidepressant amitriptyline, and the antiseizure medications divalproex sodium and topiramate are the most effective preventive medications, reducing the frequency of migraines by more than 50% in about 50% of patients. In general, preventive medications are more effective when patients are placed on a titration schedule with a minimum target dose. Titration schedules and minimum target doses are as follows: propranolol (either regular or long acting), 40 mg daily, increased weekly by 40 mg to a maximum daily dose of 120 to 160 mg; amitriptyline, 10 mg at bedtime, increased weekly by 10 mg to a maximum daily dose of 50 mg; divalproex sodium (either regular or extended release), 500 mg daily for 1 week and then 1,000 mg daily; and

Major Categories of Headache Disorders

Primary Headaches

Migraine

Tension-type headache

Cluster headache and chronic paroxysmal hemicrania

Miscellaneous headaches unassociated with structural lesion: idiopathic stabbing, external compression, cold stimulus, benign cough, benign exertional, associated with sexual activity

Secondary Headaches

Headache associated with head trauma

Headache associated with vascular disorder: acute ischemic cerebrovascular disorder, intracranial, hematoma, subarachnoid hemorrhage, unruptured vascular malformation, arteritis, carotid or vertebral artery pain, venous thrombosis, arterial hypertension, associated with other vascular disorder

Headache associated with nonvascular intracranial disorder: high and low cerebrospinal fluid pressure, intracranial infection, intracranial sarcoidosis and other noninfectious inflammatory disease, related to intrathecal injections, intracranial neoplasm, associated with other intracranial disorder

Headache associated with substances or their withdrawal: acute and long-term substance use or exposure, withdrawal after acute and long-term use, associated with substances with uncertain mechanism

Headache associated with noncephalic infection: viral infection, bacterial infection, other infection

Headache associated with metabolic disorder: hypoxia, hypercapnia, mixed hypoxia and hypercapnia, hypoglycemia, dialysis, other metabolic abnormality

Headache or facial pain associated with disorder of cranium, neck, eyes, ears, nose, sinuses, teeth, mouth, or other facial or cranial structures

Cranial neuralgias, nerve trunk pain, and deafferentation pain

Persistent pain of cranial nerve origin, trigeminal neuralgia, glossopharyngeal neuralgia, nervus intermedius neuralgia, superior laryngeal neuralgia, occipital neuralgia, central causes of head and facial pain other than tic douloureux

Triptans Available in the United States

Drug (Brand Name)	*Formulation*	*Strengths (mg)*
Almotriptan (Axert)	Tablets	12.5
Eletriptan (Relpax)	Tablets	20, 40
Frovatriptan (Frova)	Tablets	2.5
Naratriptan (Amerge)	Tablets	1, 2.5
Rizatriptan (Maxalt)	Tablets Orally disintegrating preparation* (Maxalt MLT)	5, 10
Sumatriptan (Imitrex)	Subcutaneous injection Tablets Nasal spray	6 25, 50, 100 5, 20
Zolmitriptan (Zomig)	Tablets Orally disintegrating preparation* (Zomig ZMT)	2.5, 5

*Dissolves on the tongue; can be taken without water (efficacy similar to that of tablet form).

topiramate, 25 mg daily for the first week, increased by 25 mg/wk in divided doses to a maximum daily dose of 100 mg administered at a dosage of 50 mg twice daily.[1]

Other beta blockers may also be effective [*see Table, page 5, top*]. Regarding the tricyclic antidepressants, the quality of evidence for nortriptyline is not as good as that for amitriptyline, but the clinical impression is one of similar efficacy with less sedation. Venlafaxine may be as effective as amitriptyline with fewer side effects.[2,3] Selective serotonin reuptake inhibitors are probably not effective for migraine prevention. Verapamil and gabapentin are only modestly effective for migraine prevention.

1. Evans RW, Bigal ME, Grosberg B, et al: Target doses and titration schedules for migraine preventive medications. Headache 46:160, 2006 [PMID 16412164]

2. Bulut S, Berilgen MS, Baran A, et al: Venlafaxine versus amitriptyline in the prophylactic treatment of migraine: randomized, double-blind, crossover study. Clin Neurol Neurosurg 107:44, 2004 [PMID 15567552]

3. Ozyalcin SN, Talu GK, Kiziltan E, et al: The efficacy and safety of venlafaxine in the prophylaxis of migraine. Headache 45:144, 2005 [PMID 15705120]

Natural Products for Migraine Prevention

There are natural products that may be beneficial for migraine prevention, including the herb feverfew (*Tanacetum parthenium*); extract from the butterbur plant, *Petasites hybridus* (Petadolex, 75 mg twice daily); riboflavin (400 mg a day); coenzyme Q10 (100 mg three times a day[1]); and oral magnesium supplements. Botulinum toxin injections may also be of benefit, especially in intractable cases. The relative benefit of these treatments may become clearer with additional studies, but for now, some migraineurs may prefer them because they have few if any side effects.

1. Sandor PS, Di Clemente L, Coppola G, et al: Efficacy of coenzyme Q10 in migraine prophylaxis: a randomized controlled trial. Neurology 64:713, 2005 [PMID 15728298]

11 NEUROLOGY

XIII Disorders of Sleep

SUDHANSU CHOKROVERTY, MD
New York Medical College

Idiopathic Hypersomnia

The International Classification of Sleep Disorders defines idiopathic hypersomnia as a disorder of excessive sleepiness that is associated with major sleep episodes, the cause of which is presumed to lie in the central nervous system. The disorder is further categorized as with or without long sleep time (over 10 hours or from 6 to 10 hours, respectively). The disease develops insidiously, generally between the ages of 15 and 30 years. It closely resembles narcolepsy. Affected patients generally sleep for hours, and the sleep is not refreshing. The patient does not give a history of cataplexy or snoring.

Idiopathic hypersomnia is a disabling and lifelong disorder. The Multiple Sleep Latency Test shows a mean sleep latency of less than 8 minutes without sleep-onset rapid eye movements. The treatment of idiopathic hypersomnia is similar to the stimulant treatment of narcolepsy; however, the therapeutic response is unsatisfactory.

Quieting Restless Legs

Four groups of drugs are available to treat restless legs syndrome (RLS) and periodic limb movement in sleep: dopaminergic drugs (e.g., carbidopa-levodopa and dopamine agonists such as pergolide, pramipexole, ropinirole, and cabergoline); benzodiazepines (e.g., clonazepam); opioids (e.g., codeine, propoxyphene, oxycodone, and hydrocodone); and anticonvulsants (e.g., gabapentin). The best drug for initial therapy in most cases is a dopamine agonist. The only drug currently approved by the Food and Drug Administration for use in RLS is ropinirole, but several clinical trials have proved that other agents can be used for this disorder.

For Insomnia, Briefly

For transient insomnia or insomnia of short duration, treatment with sedative-hypnotics (e.g., zolpidem, zaleplon, or eszopiclone) or short- or intermediate-acting benzodiazepines (e.g., temazepam) for a few nights to a few weeks is appropriate [*see Table, page 5, bottom*]. Hypnotic medications should not be used for chronic insomnia.

11 NEUROLOGY

III Diseases of Muscle and the Neuromuscular Junction

MARINOS C. DALAKAS, MD
National Institute of Neurological Disorders and Stroke

Unrealized Hopes for Gene Therapy

In a modified form of gene therapy, human myoblasts carrying normal dystrophin were injected into the mus-

Preventive Medications for Migraine

Drug Class	*Agent*	*Dosage*	*Typical Side Effects*
Beta blockers	Propranolol*	40–120 mg b.i.d.	Hypotension, tiredness, exacerbation of asthma
	Propranolol long acting*	60–160 mg/day	
	Metoprolol	50–200 mg/day	
	Nadolol	40–160 mg/day	
	Atenolol	50–100 mg/day	
	Timolol	10–30 mg/day	
Antidepressants	Amitriptyline*	25–150 mg h.s.	Drowsiness, dry mouth, weight gain, constipation
	Nortriptyline	25–150 mg h.s.	
	Venlafaxine	37.5–225 mg/day	Nausea and vomiting
Anticonvulsants	Divalproex sodium*	500–1,000 mg/day	Nausea, tremor, drowsiness, weight gain, alopecia, hematologic and liver abnormalities, fetal abnormalities
	Topiramate*	25–300 mg h.s.	Weight loss, paresthesias, cognitive disturbances, kidney stones
	Gabapentin	300–800 mg t.i.d.	Dizziness, fatigue, drowsiness

*Class I evidence indicates that these are the most effective medications for migraine prevention.

cles of patients with Duchenne muscular dystrophy (DMD), but the procedure failed to change the recipients' muscle function. Several prospective, randomized, placebo-controlled trials involving repeated myoblast injections to the same muscles failed to demonstrate any improvement in strength. The finding in dystrophic mice that aminoglycoside antibiotics can "read through" nonsense mutations and generate a full-length protein, thereby restoring dystrophin function, led to a trial examining the efficacy of gentamicin therapy in DMD patients who had a premature stop codon. The results were disappointing, although mild reduction in the serum CK level was noted. Future gene therapies may prove efficacious if proper vectors are found that can be used to effectively insert the gene into the muscle.[1]

1. Gregorevic P, Chamberlain JS: Gene therapy for muscular dystrophy: a review of promising progress. Expert Opin Biol Ther 3:803, 2003 [PMID 12880380]

Managing McArdle Disease

Muscle phosphorylase deficiency (also known as McArdle disease) is the prototypical glycogenosis: glycogen breakdown is inhibited, which leads to pyruvate shortage and impaired energy output. It is the second most common cause of recurrent myoglobinuria after carnitine palmitoyltransferase deficiency.

Diagnosis

This autosomal recessive disease presents as exercise intolerance and myoglobinuria in patients older than 15 years. If patients rest briefly after exercise-induced myalgia and stiffness, they can resume activity with better endurance (second-wind phenomenon), owing to increased mobilization and utilization of free fatty acids and glucose. Fixed muscle weakness may develop later in life. The resting serum creatine kinase level is often elevated. The inability to produce venous lactate after exercise has been traditionally examined with the ischemic forearm exercise test. This test, however, is falling out of favor as a diagnostic tool because it produces false positive results, is not specific, can be painful, and may result in focal muscle damage. A nonischemic exercise test is of equal diagnostic value and does not have the drawbacks of the ischemic test. Muscle biopsy shows an absence of phosphorylase, the presence of subsarcolemmal vacuoles, and increased glycogen accumulation. The diagnosis is confirmed by biochemical analysis of muscle and molecular analysis of blood cells. The defect is caused by mutations in the muscle isoform of phosphorylase on chromosome 11q13 and can be detected in the leukocytes in more than 90% of patients.

Treatment

No treatment is available for McArdle disease, but aerobic exercise training and a high-protein diet can be helpful. Vitamin B_6 supplementation has also been reported to be helpful. The results of a controlled trial indicate that creatine supplementation may

Short-Term Drug Therapy for Insomnia

Category	*Agent*	*Dose (mg)**
Sedative-hypnotics	Zaleplon	5–10
	Zolpidem	5–10
	Zolpidem extended release	6.25–12.5
	Eszopiclone	1–3
	Ramelteon	8
Short-acting or intermediate-acting benzodiazepines	Temazepam	7.5–30
	Triazolam	0.125–0.250
	Flurazepam	15–30
	Estazolam	1–2

*All agents are given at bedtime for a few nights to a few weeks.

FDA Approval Report

The following is selected from the FDA's list of recently approved products. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

New Treatment for Myelodysplastic Syndrome

Generic Name: Lenalidomide

Brand Name: Revlimid

Manufacturer: Celgene Corporation, Summit, New Jersey

The FDA has approved lenalidomide for the treatment of patients with a subtype of myelodysplastic syndrome (MDS). The subtype is that form of MDS associated with the deletion 5q cytogenetic abnormality. In clinical trials, patients treated with lenalidomide no longer needed transfusions; most patients became independent of transfusion after 3 months of therapy. The transfusion-free period lasted for an average of 44 weeks.

Lenalidomide is structurally similar to thalidomide, a drug known to cause severe birth defects. Additional studies are ongoing in animals to address whether there is a risk that lenalidomide will also cause birth defects when taken during pregnancy. While these studies are under way, lenalidomide is being marketed under a risk management plan called RevAssist, which is designed to prevent fetal exposure. Under RevAssist, only pharmacists and prescribers registered with the program will prescribe and dispense Revlimid. The program requires patients, including female patients undergoing mandatory pregnancy testing, to give informed consent before starting Revlimid. Physicians are to check pregnancy tests, limit prescriptions to a 1-month mail supply, and report any pregnancies to FDA. FDA and the manufacturer will reevaluate the risk-management plan when results of further animal testing for birth defects are completed. The labeling for lenalidomide will include a black-box warning and a medication guide regarding the prevention of fetal exposure. Additional black-box warnings include the potential need to lower the dose because of suppressed blood counts, as well as an increased risk of blood clots. Common side effects reported with lenalidomide include thrombocytopenia, neutropenia, diarrhea, pruritus, rash, and fatigue.

Source:
FDA Approves New Treatment for Myelodysplastic Syndrome (MDS). FDA News. U.S. Food and Drug Administration, December 28, 2005 (http://www.fda.gov/bbs/topics/news/2005/NEW01289.html)

improve muscle function.[1] Most encouraging was the result from a recent study showing that 75 g of sucrose before exercise markedly improved exercise tolerance and may protect against exercise-induced rhabdomyolysis.[2]

1. Quinlivan R, Beynon RJ: Pharmacological and nutritional treatment for McArdle's disease (glycogen storage disease type V). Cochrane Database Syst Rev (3):CD003458, 2004 [PMID 15266486]

2. Vissing J, Haller RG: The effect of oral sucrose on exercise tolerance in patients with McArdle's disease. N Engl J Med 349:2503, 2003 [PMID 14695410]

Easing the Course in ALS Patients

The course of amyotrophic lateral sclerosis (ALS) is invariably fatal. Therapy is primarily aimed at symptom management. For patients with dysphagia, percutaneous endoscopic gastrostomy (PEG) should be considered soon after the onset of symptoms. Initiation of noninvasive ventilation should be planned with care, with the risk of ventilator dependence balanced against that of sudden death. Algorithms outlining the indications for initiation of PEG and ventilatory assistance are available.[1] Physical therapy, rehabilitation techniques, and assistive devices such as canes, ankle-foot orthoses, and walking frames may prolong a patient's functional independence, but most patients will eventually require a wheelchair. Riluzole is the only drug available for the treatment of ALS that is approved by the Food and Drug Administration. It is believed to reduce glutamate-mediated excitotoxicity, although its exact mode of action is not well understood. Studies have shown that riluzole slows the decline in muscle strength and prolongs survival by up to 4 months[1]; it may be most useful in patients with early or mild disease. Other medications are helpful in controlling sialorrhea, pseudobulbar affect (emotional lability), dyspnea, and anxiety.[1] Hospice referral should be considered in the terminal phase.

1. Simmons Z: Management strategies for patients with amyotrophic lateral sclerosis from diagnosis to death. Neurologist 11:257, 2005 [PMID 16148733]

CLINICAL ESSENTIALS

V Adult Preventive Health Care

MARK HELFAND, MD
Oregon Health & Science University

Screening for Breast Cancer Genes

In women whose family history suggests an increased risk of deleterious *BRCA1* or *BRCA2* mutations, the United States Preventive Services Task Force (USPSTF) recommends referral for genetic counseling and evaluation for *BRCA* testing (grade B recommendation). However, the USPSTF recommends against routine testing for breast cancer susceptibility genes (*BRCA1* or *BRCA2*) or routine referral for genetic counseling in women whose family history does not suggest an increased risk of deleterious mutations in these genes

(grade D recommendation). Such screening and counseling have few or no benefits and could have important adverse ethical, legal, social, and medical consequences.

Noncancer Prevention Imperatives

The USPSTF assigns an overall grade of A, B, C, D, or I to each prevention service. The grades reflect the overall strength of evidence and the magnitude of benefit, defined as benefits minus harms. A grade of A indicates services that have solid supporting evidence and at least a moderate net benefit. Several preventive measures have earned an A grade on the strength of their good supportive evidence, substantially greater benefits than harms, and broadest applicability to primary care practice [*see Table, page 8*].

7 INFECTIOUS DISEASE

XXXI Viral Zoonoses

LYLE R. PETERSEN, MD, MPH
DUANE J. GUBLER, SC.D.
Centers for Disease Control and Prevention

Dealing with Dengue

Early diagnosis and prompt management with fluid replacement therapy can substantially reduce case-fatality rates in dengue fever.[1] Initial management and treatment decisions should not be delayed pending results of serologic tests. Clinical laboratory tests should be used to monitor vascular leakage.[1]

There is no licensed vaccine for dengue/dengue hemorrhagic fever, although significant progress is being made toward the development of live attenuated and recombinant candidate vaccines using infectious clone technology.[2,3] Currently, disease prevention depends exclusively on mosquito control and personal protective measures, such as the use of mosquito repellents.

1. Dengue Haemorrhagic Fever: Diagnosis, Treatment, Prevention and Control, 2nd ed. WHO Technical Report Series, World Health Organization, Geneva, Switzerland, 1997 http://www.who.int/csr/resources/publications/dengue/Denguepublication/en/

2. Chang GJ, Juno G, Purdy DE, et al: Recent advancement in flavivirus vaccine development. Expert Rev Vaccines 3:199, 2004 [PMID 15056045]

3. Blaney JE Jr, Matro JM, Murphy BR, et al: Recombinant, live-attenuated tetravalent dengue virus vaccine formulations induce a balanced, broad, and protective neutralizing antibody response against each of the four serotypes in rhesus monkeys. J Virol 79:5516, 2005 [PMID 15827166]

Preventing Yellow Fever

Yellow fever is an international reportable disease, and immunization is required for travelers to many countries of sub-Saharan Africa and tropical America. The live, attenuated 17D vaccine, delivered as a single 0.5 ml subcutaneous dose, is highly effective. Immunity is probably lifelong, but for travel certification, revaccination is recommended every 10 years. Information about indications for yellow fever vaccine and requirements for international travel are available at www.cdc.gov/travel/reference.htm.[1] For patients who require immunization, the locations of designated yellow fever vaccination centers can be obtained through local health departments in most areas.

Although the 17D vaccine is one of the safest vaccines, rare cases of severe and fatal infection from vaccination have been reported. The elderly may experience a higher incidence of serious adverse events.[2] Persons with documented egg allergy should not be immunized or should be skin-tested with the vaccine. The vaccine must not be given to children younger than 6 months, in whom there is a risk of postvaccinial encephalitis; it is best to delay vaccination until 9 months of age. On theoretical grounds, persons with immunosuppression, including those with clinical AIDS, should not be immunized. Immunization during pregnancy is generally contraindicated.

1. Centers for Disease Control and Prevention. Travelers Health. Atlanta: US Department of Health and Human Services, Public Health Service, 2005 http://www.cdc.gov/travel/vaccinat.htm

2. Khromava AY, Eidex RB, Weld LH, et al: Yellow fever vaccine: an updated assessment of advanced age as a risk factor for serious adverse events. Vaccine 23:3256, 2005 [PMID 15837230]

The Marburg Outbreak

In 2005, the World Health Organization confirmed an outbreak of Marburg virus infection in 124 persons in Angola.[1] Information for travelers to Angola is available from the CDC at http://www.dcd.gov/travel/other/marburg_vhf_angola_2005.htm. The ecology of Marburg virus remains virtually unknown. Marburg virus disease presents as an acute febrile illness; it can progress within 6 to 8 days to severe hemorrhagic manifestations. Clinical manifestations include fever, chills, headache, myalgia, maculopapular rash, nausea, vomiting, chest pain, and abdominal pain. Signs and symptoms can become increasingly more severe. Clinicians should consider the diagnosis of Marburg virus for febrile patients who, within 10 days before onset of fever, have traveled in northern Angola, had direct contact with blood or other body fluids of a person suspected of having hemorrhagic fever, or have worked in a laboratory that handles hemorrhagic fever viruses. No vaccine or curative treatment is available; treatment is supportive.

1. Outbreak of Marburg virus hemorrhagic fever: Angola, October 1, 2004–March 29, 2005. MMWR Morb Mortal Wkly Rep 54:308, 2005 [PMID 15800477]

SPECIAL – ONLINE ONLY!

16 WOMEN'S HEALTH

XXII Female Sexuality: Assessing Satisfaction and Addressing Problems

JENNIFER POTTER, MD
Harvard Medical School

Hysterectomy and Sexual Function

Despite its proximity to the major pelvic or paracervical ganglion, the cervix itself is a relatively insensitive structure. There are numerous theoretical reasons why removal of the uterus might be expected to interfere with sexual response, including anatomic changes (shortening of the vaginal vault, scar formation in the vaginal cuff), surgical damage to pelvic nerves and blood vessels, secondary hormonal changes, and the possibility that uter-

Strongly Recommended Noncancer Preventive Services in Adults*

Service	*Candidates*	*Established Benefits*
Aspirin for primary prevention of cardiovascular events	Adults at high cardiovascular risk	Reduces the risk of stroke
Blood pressure screening	All adults	Reduces the risk of stroke
Screening for lipid disorders	Men 35 yr of age and older; women 45 yr of age and older; and younger adults at increased risk for coronary artery disease	Reduces overall mortality, as well as mortality from cardiovascular disease
Chlamydial infection screening	Sexually active women 25 yr of age and younger; other asymptomatic women at increased risk for infection	Reduced the risk of pelvic inflammatory disease in one randomized trial
Hepatitis B virus (HBV) infection screening	Pregnant women	Reduces prenatal transmission of HBV
Syphilis screening	Persons at increased risk for infection; all pregnant women	Penicillin treatment during pregnancy reduces the risk to the fetus of acquiring congenital syphilis
HIV screening†	Pregnant women High-risk men and women	Reduces prenatal transmission of HIV Delays mortality from HIV disease and permits counseling to reduce transmission
Screening for asymptomatic bacteriuria	Pregnant women (urine culture at 12–16 weeks' gestation)	Prevents symptomatic urinary tract infections, low birth weight, and preterm delivery

*As per the United States Preventive Services Task Force. †As per the CDC.

ine contraction itself contributes substantively to orgasmic pleasure. Although studies of sexual function after hysterectomy are fraught with methodological problems, the preponderance of evidence suggests that detrimental effects of either total or supracervical hysterectomy are rare and that preoperative sexual function is the most important predictor of postoperative sexual satisfaction.

Using Sexual Questionnaires

Having patients complete a written questionnaire can sometimes help to define key issues. Several multidimensional self-report instruments on female sexual dysfunction have been developed. These include the 25-item Derogatis Interview for Sexual Functioning (DISF/DISF-SR), the 22-item Brief Index of Sexual Functioning for Women (BSFI-W), and the 19-item Female Sexual Function Index (FSFI). The FSFI, which is the most commonly used of these questionnaires, is available free of charge on the Internet (http://www.FSFIquestionnaire.com). Each of these questionnaires is well standardized, relatively unobtrusive, and inexpensive and takes 15 to 20 minutes to administer and score; moreover, normative values for both clinical and nonclinical populations are available. There is reason to believe that use of questionnaires may result in a more accurate history, because women may be more willing to answer sensitive questions privately than during a face-to-face conversation.

The Right Route for Estrogen Replacement

Most studies of systemic estrogen replacement in postmenopausal women demonstrate an improvement in sexual function in all phases of the sexual-response cycle.[1] It is not clear to what extent the effect on desire is mediated indirectly through amelioration of vaginal discomfort caused by atrophic changes or alleviation of sleep deprivation caused by vasomotor symptoms. Lubricants and local application of topical estrogen[2] can help counteract atrophy. Systemic estrogen replacement may be indicated for vasomotor symptoms, in the absence of contraindications; however, it should be remembered that oral estrogen increases sex hormone–binding globulin levels and thus may theoretically reduce rather than increase desire. Therefore, the optimal vehicle for systemic estrogen replacement may be a transdermal preparation, which avoids the hepatic first-pass effect. In patients with only minor vasomotor symptoms, topical estrogen applied to the genitalia might be a safer long-term strategy.

1. Alexander JL, Kotz K, Dennerstein L, et al: The effects of postmenopausal hormone therapies on female sexual functioning: a review of double-blind, randomized controlled trials. Menopause 11:749, 2004 [PMID 15543027]

2. Suckling J, Lethaby A, Kennedy R: Local oestrogen for vaginal atrophy in postmenopausal women. Cochrane Database Syst Rev (4):CD001500, 2003 [PMID 14583935]

Coming in May

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 29 No. 3 March 2006

What's New in ACP Medicine

DAVID C. DALE, MD, FACP, *Editor-in-Chief* DANIEL D. FEDERMAN, MD, MACP, *Founding Editor*
WENDY LEVINSON, MD, FACP, *Associate Medical Editor, What's New in ACP Medicine*

www.acpmedicine.com

PRACTICE OF MEDICINE

Vitamin D and Cancer: A Goldilocks Paradox?

KAREN ANTMAN, MD
Columbia University College of Physicians and Surgeons

Skin cancers are associated with excess solar exposure. On the other hand, inadequate solar exposure may result in vitamin D deficiency, which may be linked with an increased risk of prostate, colon, breast, and other cancers.

The relevant band of sunlight is ultraviolet B (UVB) light. In the United States, UVB levels are lowest in the Northeast and highest in the Southwest. Mortality for many cancers in the United States is highest in the Northeast and lowest in the Southwest, with the notable exceptions of bronchogenic cancer in the Southeast and skin cancers in the South. Regional UVB levels in the month of July show statistically significant inverse correlations with the incidence of bladder, breast, colon, esophageal, gastric, ovarian, prostate, rectal, renal, and uterine cancer and non-Hodgkin lymphoma; in addition, UVB may possibly be associated with an increase in the incidence of laryngeal, oral, cervical, gallbladder, and pancreatic cancers and Hodgkin lymphoma.

UVB levels depend on season, time of day, exposure time, latitude, altitude, the amount of skin surface exposed, cloud cover, smog, sunscreen use, and skin pigmentation. Clouds, pollution, smog, and ozone can block UVB. Window glass blocks about

continued on page 2

THIS MONTH'S UPDATES

7 INFECTIOUS DISEASE

X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas

SHAWN J. SKERRETT, MD
University of Washington School of Medicine

Treating *Haemophilus influenzae* Infection

The initial antibiotic of choice for meningitis and other invasive *H. influenzae* infections is a third-generation cephalosporin such as ceftriaxone or cefotaxime. These agents are highly active against all isolates of *H. influenzae*, penetrate the cerebrospinal fluid, and eradicate nasopharyngeal carriage of *H. influenzae* type b. The pediatric dose of ceftriaxone is 50 mg/kg every 12 hours and that for cefotaxime is 50 mg/kg every 6 hours. For adults, the maximum dose of ceftriaxone is 2 g every 12 hours and that for cefotaxime is 2 g every 6 hours. The preferred alternative treatment for *H. influenzae* meningitis is chloramphenicol (75 to 100 mg/kg/day in six divided doses). Children with *H. influenzae* meningitis also should receive dexamethasone (0.6 mg/kg/day in four divided doses). Other antibiotics that are highly active against nearly all strains of *H. influenzae* include carbapenems, aztreonam, β-lactam/β-lactamase inhibitor combinations, fluoroquinolones, azithromycin, and tetracycline[1] [*see Table, page 4, top*]. Ampicillin is highly effective against sensitive strains, but 35% to 40% of North American isolates of *H. influenzae* are resistant to ampicillin. Most ampicillin resistance is mediated by β-lactamase production, but conventional susceptibility testing is recommended for invasive isolates to detect all resistant strains. Intravenous antibiotic therapy is recommended for meningitis and endocarditis and for the initial treatment of other invasive infections. Less severe *H. influenzae* infections can be managed with oral antibiotics. Most *H. influenzae* infections can be cured with 7 to 10 days of therapy, but treatment should continue for 3 to 6

continued on page 3

In This Issue

PRACTICE OF MEDICINE

continued from page 1

95% of UVB light. Direct UV light on a sunny summer day allows skin to produce vitamin D; such production ceases after 10 to 20 minutes of exposure in light-skinned people. Because melanin blocks UVB, vitamin D production requires proportionately longer exposure in darker-skinned people; indeed, the need for vitamin D may provide the evolutionary explanation for lighter skin in northern latitudes. Between November and March in most of the United States, when the sun is at a lower angle and UVB is filtered through the atmosphere, even light-skinned people produce little vitamin D. Older or obese persons also produce vitamin D less efficiently.

Fortified foods are the major dietary sources of vitamin D (calciferol). Rickets in children was virtually eliminated in the 1930s by the fortification of milk with vitamin D. Vitamin D_3 (cholecalciferol) comes from fish oil, eggs, organ meats, and animal fat. Vitamin D_2 (ergocalciferol), which is less biologically active, is found in fortified foods and some supplements. Most vitamin D supplements provide 400 IU. A glass of fortified milk provides about 100 IU. For comparison, 10 to 20 minutes of exposure on a sunny beach can result in the production of 10,000 IU. Despite food fortification, a substantial fraction of the United States population is vitamin D deficient.[1] Vitamin D deficiency rates are particularly high in the elderly and dark-skinned people in northern latitudes.

In laboratory models, vitamin D reduces cell proliferation and increases cell differentiation, improves cell adhesion, and inhibits cancer progression and metastasis. Vitamin D ameliorates the growth-signaling effects of insulin-like growth factor–1 (IGF-1) and induces apoptosis and downregulation of telomerase.[2-5] Many tissues have vitamin D receptors (VDRs). Various alleles of the VDR genes may affect the risk of cancer.[4]

Clinical Studies

Skin cancer The development of basal cell cancers and melanoma correlates with intermittent weekend or vacation sun exposures and sunburns, whereas regular occupational exposure is protective. However, actinic keratosis and squamous cell carcinoma correlate with total lifetime UVB irradiation.

Colon cancer A number of studies have documented significant inverse associations of serum 25-hydroxyvitamin D_3 [25(OH)D_3] concentrations with the risk of colorectal cancer.[6-9] Analyses limited to dietary vitamin D tend to have mixed results because dietary sources provide only a portion of total vitamin D; supplements and solar UVB radiation provide the balance.[10] Thus, serum levels may be more reliable than dietary histories.

Prostate cancer Vitamin D inhibits prostate cancer growth via VDRs and increases IGF-binding protein expression. The IGF and vitamin D regulatory systems may interact to affect prostate cancer risk.[4] In a case-control study from Great Britain, prostate cancer risk correlated inversely with childhood sunburn, sunbathing, and foreign holidays; in addition, prostate cancer developed at an earlier age in men with low UV exposure than in those with higher exposure.[11]

Breast cancer In premenopausal women, consumption of vitamin D has been inversely associated with risk of breast cancer.[12] Breast cancer risk has also been correlated with lower mean 25(OH)D levels and 1,25$(OH)_2$D levels. For both metabolites, the trend was stronger in women older than 60 years.[13]

Conclusions

The interaction of sun exposure and cancer risk may be more complex than originally conceived—too much or too little vitamin D may increase cancer risk. Consensus is evolving on the "just right" level of solar exposure and whether increased oral vitamin D supplements can substitute for UVB exposure.

1. Thomas MK, Lloyd-Jones DM, Thadhani RI, et al: Hypovitaminosis D in medical inpatients. N Engl J Med 338:777, 1998 [PMID 9504937]

2. Lamprecht SA, Lipkin M: Chemoprevention of

ACP MEDICINE
A Publication of the American College of Physicians

Published by WebMD

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $349 for the first year ($324 for residents and students) and $269 for renewals ($249 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $449 for the first year and $369 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

PRACTICE OF MEDICINE

continued from page 2

colon cancer by calcium, vitamin D and folate: molecular mechanisms. Nat Rev Cancer 3:601, 2003 [PMID 12894248]

3. Ordonez-Moran P, Larriba MJ, Pendas-Franco N, et al: Vitamin D and cancer: an update of in vitro and in vivo data. Front Biosci 10:2723, 2005 [PMID 15970529]

4. Chokkalingam AP, McGlynn KA, Gao YT, et al: Vitamin D receptor gene polymorphisms, insulin-like growth factors, and prostate cancer risk: a population-based case-control study in China. Cancer Res 61:4333, 2001 [PMID 11389055]

5. Giovannucci E: The epidemiology of vitamin D and cancer incidence and mortality: a review (United States). Cancer Causes Control 16:83, 2005 [PMID 15868450]

6. Grant WB, Garland CF: A critical review of studies on vitamin D in relation to colorectal cancer. Nutr Cancer 48:115, 2004 [PMID 15231446]

7. Garland CF, Comstock GW, Garland FC, et al: Serum 25-hydroxyvitamin D and colon cancer: eight year prospective study. Lancet 2:1176, 1989 [PMID 2572900]

8. Feskanich D, Ma J, Fuchs CS, et al: Plasma vitamin D metabolites and risk of colorectal cancer in women. Cancer Epidemiol Biomarkers Prev 13:1502, 2004 [PMID 15342452]

9. Kampman E, Slattery ML, Caan B, et al: Calcium, vitamin D, sunshine exposure, dairy products and colon cancer risk (United States). Cancer Causes Control 11:459, 2000 [PMID 10877339]

10. Moan J, Porojnicu AC, Robsahm TE, et al: Solar radiation, vitamin D and survival rate of colon cancer in Norway. J Photochem Photobiol B 78:189, 2005 [PMID 15708515]

11. Luscombe CJ, Fryer AA, French ME, et al: Exposure to ultraviolet radiation: association with susceptibility and age at presentation with prostate cancer. Lancet 358:641, 2001 [PMID 11530156]

12. Shin MH, Holmes MD, Hankinson SE, et al: Intake of dairy products, calcium, and vitamin D and risk of breast cancer. J Natl Cancer Inst 94:1301, 2002 [PMID 12208895]

13. Bertone-Johnson ER, Chen WY, Holick MF, et al: Plasma 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 14:1991, 2005 [PMID 16103450]

karenantmanmd@webmd.net

THIS MONTH'S UPDATES

continued from page 1

weeks in cases of osteomyelitis or endocarditis.

1. Johnson DM, Sader HS, Fritsche TR, et al: Susceptibility trends of *Haemophilus influenzae* and *Moraxella catarrhalis* against orally administered antimicrobial agents: five-year report from the SENTRY Antimicrobial Surveillance Program. Diagn Microbiol Infect Dis 47:373, 2003 [PMID 12967753]

The New Front Runner

Nontypeable strains of *H. influenzae* are isolated from 15% to 30% of purulent middle-ear effusions in both children and adults. *H. influenzae* is now the leading cause of persistent otitis media in children immunized with the pneumococcal conjugate vaccine.[1]

As with all cases of otitis media, infants typically present with fever and irritability, and older children and adults complain primarily of ear pain. The clinical diagnosis is made by otoscopy. A bacteriologic diagnosis requires tympanocentesis, but this procedure is warranted only for evaluation of treatment failures.

1. Casey JR, Pichichero ME: Changes in frequency and pathogens causing acute otitis media in 1995–2003. Pediatr Infect Dis J 23:824, 2004 [PMID 15361720]

Managing *Moraxella*

Moraxella catarrhalis is the third most common cause of otitis media, sinusitis, and acute exacerbations of chronic bronchitis, after *H. influenzae* and *Streptococcus pneumoniae*.[1] The majority of lower respiratory tract infections attributed to *M. catarrhalis* in adults occur in patients with underlying chronic obstructive pulmonary disease. More than 90% of clinical isolates of *M. catarrhalis* produce β-lactamases that render ampicillin and amoxicillin ineffective.[2] Penicillin–β-lactamase–inhibitor combinations, second- and third-generation cephalosporins, fluoroquinolones, macrolides, and tetracyclines are active against nearly all strains of *M. catarrhalis* [*see Table, page 4, top*].

1. Sande MA, Gwaltney J: Acute community-acquired bacterial sinusitis: continuing challenges and current management. Clin Infect Dis 39:S151, 2004 [PMID 15546110]

2. Johnson DM, Sader HS, Fritsche TR, et al: Susceptibility trends of *Haemophilus influenzae* and *Moraxella catarrhalis* against orally administered antimicrobial agents: five-year report from the SENTRY Antimicrobial Surveillance Program. Diagn Microbiol Infect Dis 47:373, 2003 [PMID 12967753]

Best Tests for Legionnaires Disease

The diagnosis of legionellosis requires the isolation of the organism in culture, detection of microbial antigens or nucleic acids in body fluids, or demonstration of serologic evidence of infection [*see Table, page 4, bottom*]. However, an early clue to the diagnosis of Legionnaires disease can be obtained from nonspecific stains of sputum, bronchoalveolar lavage fluid, or lung tissue.

Rapid Diagnosis of *Legionella* Pneumonia

Urinary antigen detection is currently the most helpful test for the rapid diagnosis of legionellosis. Antigens appear in the urine within 3 days of the onset of illness and may persist for months, although most patients stop excreting antigen within 6 weeks. The sensitivity of this test is higher in cases of severe pneumonia than in milder disease.[1] The major limitation of urinary antigen testing is that currently available tests reliably detect only *L. pneumophila* serogroup 1, which accounts for more than 80% of legionellosis cases diagnosed in the United States.

1. Blazquez RM, Espinosa FJ, Martinez-Toldos CM, et al: Sensitivity of urinary antigen test in relation to clinical severity in a large outbreak of *Legionella* pneumonia in Spain. Eur J Clin Microbiol Infect Dis 24:488, 2005 [PMID 15997369]

Pseudomonas in the Diabetic Ear

Pseudomonas aeruginosa is the causative pathogen in nearly all cases of malignant (necrotizing) exter-

Antibiotics Effective against *Haemophilus influenzae* and *Moraxella catarrhalis*

Antibiotic	H. influenzae *Susceptible (%)*	M. catarrhalis *Susceptible (%)*
Ceftriaxone	100	100
Cefpodoxime	100	99
Cefdinir	98	100
Cefuroxime	98–99	99
Ampicillin	68–72	< 10
Amoxicillin-clavulanate	> 99	> 99
Fluoroquinolones	> 99	100
Azithromycin	> 99	100
Clarithromycin	60–90	100
Telithromycin	98	100
Tetracycline	> 98	100
Chloramphenicol	> 99	100
Trimethoprim-sulfamethoxazole	79–82	> 97

nal otitis, a chronic, invasive infection of the external auditory canal that predominantly afflicts elderly persons with diabetes mellitus.[1] Rarely, cases have been reported in AIDS patients and immunocompromised children. The infection spreads to the temporal bone and mastoid, often involving the adjacent cranial nerves as they exit the skull. Severe otalgia is the presenting symptom, and most patients have purulent otorrhea with an intact tympanic membrane. Facial nerve paresis is evident in 30% to 40% of cases; other cranial neuropathies are less common. Fever and other systemic signs of infection are usually absent. The erythrocyte sedimentation rate (ESR) is nearly always markedly elevated. Technetium-99m bone scanning is very sensitive but not specific for this condition. Computed tomography and magnetic resonance imaging are useful for defining the extent of disease, with MRI being more sensitive for delineating the soft tissue component.[1] The diagnosis can be made from the clinical presentation, imaging studies, and cultures of diseased bone or granulation tissue in the external canal.

1. Grandis JR, Bransteter BF IV, Yu VL: The changing face of malignant (necrotising) external otitis: clinical, radiological, and anatomic correlations. Lancet Infect Dis 4:34, 2004 [PMID 14720566]

7 INFECTIOUS DISEASE

XXIV Hyperthermia, Fever, and Fever of Undetermined Origin

HARVEY B. SIMON, MD, FACP
Harvard Medical School

Causes of Hyperthermia

Numerous clinical disorders can disrupt thermoregulatory homeostasis by causing increased heat production, decreased heat dissipation, or hypothalamic insult, thereby inducing hyperthermia [*see Table, page 5, top*]. Mild forms of hyperthermia are common. In dehydration, for example, cutaneous vasoconstriction and cessation of sweating occur in response to the decrease in intravascular volume as a means of conserving further loss of fluid and of minimizing the consequences of fluid loss. As a result, heat dissipation is impaired and body temperature may rise. The hyperthermia of dehydration is usually mild and is readily corrected by fluid replacement. In some cases, however, thermoregulatory disorders cause extreme pyrexia, in which body temperatures rise to 41.1° C (106.0° F) or higher; examples of these thermoregulatory disorders are heatstroke, malignant hyperthermia of anesthesia, thyroid storm, and hypothalamic insult caused by infection, tumor, or drugs.

A Half-century of Diagnostic Challenges

Although geographic factors are relevant, the leading causes of fever of undetermined origin (FUO) are reasonably uniform throughout the United States. The relative frequency of the etiologic categories responsible for FUO have been relatively stable over the past 5 decades [*see Table, page 5, bottom*].

The Serotonin Syndrome

The serotonin syndrome,[1] which results from excessive agonism of serotoninergic receptors in the central nervous system and periphery, has become more prevalent with the increasing use of selective serotonin reuptake inhibitors (SSRIs) and other serotoninergic agents. Because of its broad range of clinical manifestations and the large number of prescription, over-the-counter, and herbal products that have been implicated, the serotonin syndrome is often unrecognized. The actual prevalence is not known, but the syndrome occurs in 14% to 16% of patients who take overdoses of SSRIs.

Serotonin syndrome begins abruptly within 5 weeks of starting

Diagnostic Tests for Legionellosis

Test	*Specimen*	*Sensitivity (%)*	*Specificity (%)*	*Comment*
Culture	Sputum, BALF	11–80	100	All species; requires selective media
Direct fluorescent antibody	Sputum, BALF	22–75	> 90	*Legionella pneumophila* only
Urinary antigen	Urine	48–90	> 95	*L. pneumophila* serogroup 1 only
Serology				
Fourfold change*	Blood	60–80	>90	*L. pneumophila* only
Single titer		10–27	> 85	
PCR	Sputum, BALF, serum, urine	10–100	> 95	All species; not widely available

*Between acute and convalescent samples.
BALF—bronchoalveolar lavage fluid PCR—polymerase chain reaction

serotoninergic agents, either singly or in combination. In mild cases, patients exhibit hyperkinesia, intermittent tremors, hyperreflexia and clonus, hyperactive bowel sounds, diarrhea, mydriasis, and tachycardia. Patients with these milder cases may have shivering and diaphoresis, but body temperature is normal. In moderate to severe cases, however, hyperthermia is the rule, and it may be severe. Muscular rigidity, agitation, and ocular clonus or inducible clonus are characteristic features of the severe serotonin syndrome.

Management of the serotonin syndrome requires removal of the responsible drug or drugs and metabolic and hemodynamic support. In mild cases, agitation and tremors can often be controlled with benzodiazepines, but severely ill patients require sedation, neuromuscular paralysis, and ventilatory support. Antipyretics, beta blockers, bromocriptine, and dantrolene are not effective. Cyproheptadine, olanzapine, and chlorpromazine have been useful in isolated cases but have not been studied fully.[1]

1. Boyer EW, Shannon M: The serotonin syndrome. N Engl J Med 352:1112, 2005[PMID 15784664]

Causes of Hyperthermia

Cause	*Example*
Excessive heat production	Delirium tremens Drug abuse (amphetamines) Exertional hyperthermia Generalized tetanus Heatstroke (exertional)* Lethal catatonia Malignant hyperthermia of anesthesia Neuroleptic malignant syndrome* Pheochromocytoma Salicylate intoxication Serotonin syndrome Status epilepticus Thyrotoxicosis
Diminished heat dissipation	Anticholinergic drugs Autonomic dysfunction Dehydration Heatstroke (classic)* Neuroleptic malignant syndrome* Occlusive dressings Drug abuse (cocaine)
Hypothalamic dysfunction	Cerebrovascular accidents Encephalitis Idiopathic hypothalamic dysfunction Neuroleptic malignant syndrome* Sarcoidosis and granulomatous infections Trauma Tumors

*Pathogenesis of these disorders is mixed.

Chill Out

Patients with strokes are at risk for additional brain injury from fever; the result is a marked increase in morbidity and mortality, especially when the pyrexia occurs soon after the stroke. Because of this, even modest elevations in temperature should be suppressed in patients with acute strokes. In addition, moderate induced hypothermia may improve the outcome of strokes, even in patients who are afebrile on presentation. Hypothermia may also improve outcome after traumatic brain injury[1] in neonates with hypoxic-ischemic encephalopathy[2] and in patients receiving cardiopulmonary resuscitation.[3]

1. Shann F: Hypothermia for traumatic brain injury: how soon, how cold, and how long? Lancet 362:1950, 2003 [PMID 14683651]
2. Shankaran S, Laptook AR, Ehrenkranz RA, et al: Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. N Engl J Med 353:1574, 2005 [PMID 16221780]
3. Boddicker KA, Zang Y, Zimmerman B, et al: Hypothermia improves defibrillation success and resuscitation outcomes from ventricular fibrillation. Circulation 111:3195, 2005 [PMID 15956132]

11 NEUROLOGY

XI Alzheimer Disease and Other Major Dementing Illnesses

DAVID S. KNOPMAN, MD
Mayo Clinic School of Medicine

Imaging in Dementia

Brain imaging is necessary for the initial diagnostic evaluation of dement-

Causes of Fever of Undetermined Origin over 5 Decades

Cause	*Study*				
	New Haven, 1961 (% of 100 Patients)	*Boston, 1973 (% of 128 Patients)*	*Seattle, 1982 (% of 105 Patients)*	*Rhode Island, 1992 (% of 89 Patients)*	*Leuven, Belgium, 2003 (% of 290 Patients*)*
Infections	36	35	30	33	20
Neoplasms	19	23	31	24	20
Collagen vascular diseases	13	16	9	16	19
Other specific causes	25	18	18	18	18
Undiagnosed	7	8	12	9	34

*Percentages have been modified to conform to the diagnostic criteria used in the American series.

FDA Approval Report

The following is selected from the FDA's list of recently approved products. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

New Treatment for Advanced Kidney Cancer

Generic Name: Sorafenib tosylate

Brand Name: Nexavar

Manufacturer: Bayer Pharmaceuticals Corporation, Westhaven, Connecticut

The FDA has approved sorafenib tosylate for the treatment of adults with advanced renal cell carcinoma, the most common type of kidney cancer. In two studies in patients with advanced kidney cancer, the median time to tumor progression or death was 167 days in patients treated with sorafenib tosylate, compared with 84 days in patients not treated with the drug. In the larger of the two studies, most patients had previously received treatment with interleukin-2 or interferon. Common temporary side effects reported with sorafenib tosylate are rash; diarrhea; increased blood pressure; and redness, pain, swelling, or blisters on the palms of the hands or soles of the feet.

Source:
FDA Approves New Treatment for Advanced Kidney Cancer. FDA News. U.S. Food and Drug Administration, December 20, 2005 (http://www.fda.gov/bbs/topics/NEWS/2005/NEW01282.html)

ed patients. A brain computed tomography scan may be adequate, but a magnetic resonance imaging scan without contrast enhancement will yield more clinically useful information. The fundamental purpose of a brain-imaging study is to rule out space-occupying lesions such as tumors and subdural hematomas. MRI is becoming increasingly useful for assessing the burden of cerebral infarcts, as well. Elderly persons are at increased risk for small infarcts in deep brain structures that may occur covertly.

MRI can also detect hippocampal atrophy when imaging is performed in a plane perpendicular to the axis of the temporal lobe. Hippocampal atrophy is commonly seen in Alzheimer disease (AD), but this finding is suggestive rather than diagnostic of AD because it is also seen both in normal elderly patients and in those with other dementing illnesses.

Positron emission tomography (PET) using fluorodeoxyglucose as a tracer may show reduced metabolic activity in the temporoparietal regions in AD, but it is not clear what additive value this technique provides beyond the clinical diagnosis and at what cost. More recently, a PET tracer known as Pittsburgh compound B (PIB) has been shown to label brain amyloid in patients with AD. Studies are under way to determine whether PET imaging using PIB is of value in the diagnosis of AD.[1]

1. Klunk WE, Engler H, Nordberg A, et al: Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. Ann Neurol 55:306, 2004 [PMID 14991808]

Vitamin E Fails a Test

Vitamin E is often recommended for patients with AD because a 2-year study of AD patients with moderate dementia showed that vitamin E (α-tocopherol, 2,000 IU a day) delayed the progression to severe dementia by about 200 days. A subsequent study of the use of vitamin E in patients with mild cognitive impairment failed to show any benefit.[1] These results call into question the rationale for treating patients with mild AD with vitamin E.

1. Petersen RC, Thomas RG, Grundman M, et al: Vitamin E and donepezil for the treatment of mild cognitive impairment. N Engl J Med 352:23, 2005 [PMID 15829527]

A Pair for Alzheimer Treatment

Memantine is approved by the Food and Drug Administration for the treatment of AD. This agent, a glutamate modulator that is a noncompetitive receptor antagonist of *N*-methyl-D-aspartate, has been the subject of several clinical trials that reported positive results in moderate to severe dementia.[1,2] It is hypothesized that glutamatergic overstimulation is a part of the pathogenic cycle in AD. In the clinical trials, memantine delayed symptom progression. There is no evidence that memantine affects the biologic course of AD, however. Studies in patients with mild to moderate AD have been completed, but no results are yet available. On the basis of a study that paired memantine with donepezil, the current suggestion in moderate to severe AD is to use both agents.[2]

1. Reisberg B, Doody R, Stoffler A, et al: Memantine in moderate-to-severe Alzheimer's disease. N Engl J Med 348:1333, 2003 [PMID 12672860]

2. Tariot PN, Farlow MR, Grossberg GT, et al: Memantine treatment in patients with moderate to severe Alzheimer disease already receiving donepezil: a randomized controlled trial. JAMA 291:317, 2004 [PMID 14734594]

11 NEUROLOGY

XVI Acute Viral Central Nervous System Diseases

DONALD H. GILDEN, MD
University of Colorado Health Sciences Center

The Many Faces of West Nile Infection

The neurologic complications of West Nile virus infection are protean. Patients have presented not only with meningoencephalitis but also with encephalitis involving the cerebellum, pons, and medulla oblongata (rhombencephalitis), cerebellar ataxia affecting the muscles and eyes (opsoclonus-myoclonus),[1]

unilateral brachial plexopathy with meningoencephalitis,[2] and Guillain-Barré syndrome. As with other arbovirus infections, deep-seated lesions in the basal ganglia and thalamus are often seen on brain MRI.[3] Polymerase chain reaction testing of the cerebrospinal fluid often reveals West Nile virus RNA.

1. Khosla JS, Edelman MJ, Kennedy N, et al: West Nile virus presenting as opsoclonus-myoclonus cerebellar ataxia. Neurology 64:1095, 2005 [PMID 15781844]

2. Almhanna K, Palanichamy N, Sharma M, et al: Unilateral brachial plexopathy associated with West Nile virus meningoencephalitis. Clin Infect Dis 36:1629, 2003 [PMID 12802774]

3. Rosas H, Wippold FJ 2nd: West Nile virus: case report with MR imaging findings. AJNR Am J Neuroradiol 24:1376, 2003 [PMID 12917131]

Lab Tests for Viral CNS Infection

Laboratory methods that are used to identify the causative agent of viral meningitis and encephalitis include polymerase chain reaction detection of viral RNA or DNA in the cerebrospinal fluid, culturing virus from the CSF, antigen or antibody detection in the CSF, serologic tests (e.g., enzyme immunoassays and Western blot assay), and brain biopsy [*see Table, top, right*].

Viral Diagnosis of Aseptic Meningitis Syndrome and Encephalitis

Virus	*Tissue Culture*	*Antigen/Antibody Detection in CSF**	*PCR*
Adenovirus	×		
Cytomegalovirus	×	×	×
Enteroviruses	×	×	×
Epstein-Barr virus		×	×
Herpes simplex virus type 1	×	×	×
Herpes simplex virus type 2	×	×	×
HIV		×	
Influenza		×	
Measles		×	
Mumps		×	
Poliovirus	×	×	
Varicella-zoster virus	×	×	×

*Results are not usually available during acute disease.

Lymphocytic Choriomeningitis Infections

The arenaviruses are RNA viruses endemic in rodents. Human infection, which occurs mostly in South America and Africa, produces various hemorrhagic fevers that are often fatal. In the United States, the most common arenavirus infection is produced by lymphocytic choriomeningitis (LCM) virus. Infection is acquired from mice or pet hamsters or through occupational exposure in laboratory personnel working with the virus. Most cases present as aseptic meningitis, but fatal meningoencephalitis also occurs. LCM preceded by a lupuslike syndrome that included rash and the presence of a circulating anticoagulant has been described. In 2005, a cluster of cases of meningoencephalitis deaths in solid organ transplant recipients was associated with LCM infection.[1] As with mumps, orchitis and parotitis may develop concurrently with central nervous system disease; profound mononuclear pleocytosis and low glucose levels in the cerebrospinal fluid also occur. A specific diagnosis can be made by observing the development of choriomeningitis in weanling or adult mice inoculated intracerebrally with infected CSF or the development of LCM-specific antibody in the serum or CSF of infected humans.

1. Lymphocytic choriomeningitis virus infection in organ transplant recipients—Massachusetts, Rhode Island, 2005. MMWR Morb Mortal Wkly Rep 54:537, 2005 [PMID 15931158]

Epstein-Barr in Neurologic Clothing

Despite the prevalence of Epstein-Barr virus (EBV) infection, neurologic disease is rare. The most common presentation is meningoencephalitis, which is often associated with acute cerebellar ataxia. More serious meningoencephalopathy, presenting as athetosis and chorea or as stupor and coma, has been described. In a fascinating case of acute demyelinating encephalopathy after EBV infection, a patient presented with behavioral abnormalities, visual illusions, and a seizure. EBV mononucleosis may also be followed by recurrent aseptic meningitis. Chronic active EBV infection has been associated with calcification in the basal ganglia. EBV-associated cranial neuropathies, including acute autonomic neuropathy, have been described. In a comprehensive study of four patients with EBV myeloradiculitis and encephalomyeloradiculitis, none of the patients died or had brain swelling, but residual neurologic deficits were evident. A case of myeloradiculitis in an immunocompetent 72-year-old woman who was seronegative for EBV was proved to be caused by EBV by the detection of EBV DNA in both serum and CSF; quantitative polymerase chain reaction revealed an EBV DNA load not detected in healthy persons who have latent infection or are not infected.[1] Detection of EBV DNA in CSF and reductions in the ratio of EBV antibody in serum to EBV antibody in CSF have been used to diagnose infection in the peripheral nervous system and CNS.

1. Muhlau M, Bulow S, Stimmer H, et al: Seronegative Epstein-Barr virus myeloradiculitis in an immunocompetent 72-year-old woman. Neurology 65:1329, 2005 [PMID 16247075]

Coming in April

Clinical Essentials
V Adult Preventive Health Care

7 Infectious Disease
XXXI Viral Zoonoses

11 Neurology
III Diseases of Muscle and the Neuromuscular Junction
VIII Headache
XIII Disorders of Sleep

14 Respiratory Medicine
XIII Invasive Diagnostic and Therapeutic Techniques in Lung Disease

16 Women's Health
XXII Female Sexuality: Assessing Satisfaction and Addressing Problems

CLINICAL ESSENTIALS

VIII Quantitative Aspects of Clinical Decision Making

BRIAN HAYNES, MD, PHD, FACP
McMaster University Health Sciences Centre
HAROLD C. SOX, JR, MD, MACP
Annals of Internal Medicine

Tools for Staying Current

The current best evidence for clinical practice is becoming more and more accessible to clinicians. Resources include electronic access to MEDLINE and its clinical subsets, specialized compendia of studies such as Clinical Evidence and Physicians' Information and Educational Resource (PIER [http://pier.acponline.org/index.html?hp]); systematic reviews of studies, such as the Cochrane Library (http://www.update-software.com/cochrane); and alerting services for new, clinically relevant evidence, such as the British Medical Journal Updates (www.bmjupdates.com) and MEDSCAPE Best Evidence alerts (https://profreg.medscape.com/px/newsletter.do).

Test Results in Context

A physician should never interpret a test result in isolation but should always take into account the individual patient's pretest probability. The posttest probability after a positive test result will be greater if the pretest index of suspicion was high than if the pretest index of suspicion was low. The most important practical application of this reasoning is to be suspicious when a test result is negative in a patient whose clinical findings strongly point toward a disease or when a test is positive in a patient for whom the likelihood of disease is very low.

The evaluation of suspected pulmonary embolism (PE) using helical CT scanning is a good example of the practical use of these statistical terms and methods. To calculate the posttest odds of PE, the physician must combine the patient's pretest odds with the test's likelihood ratio by means of the odds ratio format of Bayes' theorem. The same logic can be applied to all screening and diagnostic tests for PE, such as D-dimer, a sensitive test with low specificity that is therefore more useful for ruling out pulmonary embolism when negative than ruling it in when positive.[1] It can also be used for calibrating clinical observations to enhance the quantitation of pretest probabilities.[2]

1. Stein PD, Hull RD, Patel KC, et al: D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism. Ann Intern Med 140:589, 2004 [PMID 15096330]

2. Chunilal SD, Eikelboom JW, Attia J, et al: Does this patient have pulmonary embolism? JAMA 290:2849, 2003 [PMID 14657070]

Judging Diagnostic Tests

In the past, articles usually described the performance of a diagnostic test only in terms of sensitivity and specificity. However, these two familiar terms do not directly describe the effect of a test result on the probability of disease. In addition to sensitivity and specificity, other clinically useful measures of diagnostic test performance include, for example, the likelihood ratio. Clinically useful measures of test interpretation include pretest odds, pretest probability, probability after a positive test result, and probability after a negative test result [*see Table, below*]. Physicians should memorize and internalize the definitions of these terms to avoid becoming muddled when they are attempting to use information that has been gathered from diagnostic tests in decision making.

Definitions of Clinically Useful Measures of Diagnostic Test Performance and Interpretation

The typical approach to evaluation of most diagnostic tests, particularly those with so-called binary outcomes (e.g., a positive or a negative test result, with no other categories), makes use of a 2 × 2 table, as follows:

Diagnostic Test Result	*Presence or Absence of Disease on a Reference Test (Gold Standard)*		*No. of Patients with Given Test Result*
	Present	*Absent*	
Positive	*a*	*b*	*a* + *b*
Negative	*c*	*d*	*c* + *d*
Total	*a* + *c*	*b* + *d*	

Measures of diagnostic test performance, defined below, are calculated from this table.

- Sensitivity: the proportion of people with a disease of interest who are detected by a diagnostic test; calculated as $a/(a+c)$.
- Specificity: the proportion of people who do not have a disease who are correctly identified by a negative result on a diagnostic test; calculated as $d/(b+d)$.
- Likelihood ratio: the amount by which the odds of having a disease change after a test result; calculated as $[a/(a+c)]/[b/(b+d)]$ for a positive test result and as $[c/(a+c)]/[d/(b+d)]$ for a negative test result.
- Pretest probability: the proportion of people with the disorder of interest in a group suspected of having the disorder; calculated as $(a+c)/(a+b+c+d)$.
- Odds: calculated as probability/(1 – probability).
- Probability: calculated as odds/(1+ odds).
- Posttest odds: calculated as pretest odds × likelihood ratio.
- Probability after a positive test: the proportion of people with a positive test result who have the disease of interest; calculated as $a/(a+b)$.
- Probability after a negative test: the proportion of people with a negative test result who have the disease of interest; calculated as $c/(c+d)$.

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 29 No. 2 February 2006

What's New in ACP Medicine

DAVID C. DALE, MD, FACP, *Editor-in-Chief* DANIEL D. FEDERMAN, MD, MACP, *Founding Editor*
WENDY LEVINSON, MD, FACP, *Associate Medical Editor, What's New in ACP Medicine*

www.acpmedicine.com

PRACTICE OF MEDICINE

Quality Improvement in Acute Myocardial Infarction

RAYMOND J. GIBBONS, MD
Mayo Clinic College of Medicine

There is a growing interest in improving the quality of care for patients with acute myocardial infarction. Over the past few years, the Joint Commission on Accreditation of Healthcare Organizations has collected data from hospitals on discharge performance measures in the care of patients with acute myocardial infarction. The public release of these sobering data has generated media attention, including this quote from an editorial in *USA Today*: "If you became seriously ill, you'd expect a hospital to provide what the medical profession considers essential care. At many hospitals, you'd be wrong. Thousands of patients needlessly die each year…."

The Center for Medicare and Medicaid Services has started several initiatives to improve the quality of care, including a new Web site (www.hospitalcompare.hhs.gov), which lists performance measures for four different conditions (i.e., heart attack, heart failure, pneumonia, and surgery). This Web site permits a consumer to examine individual performance measures for a local hospital and graphically compare the hospital's performance with that of other hospitals. The URL for this Web site has been

continued on page 2

THIS MONTH'S UPDATES

4 GASTROENTEROLOGY

V Diseases of the Pancreas

PETER DRAGANOV, MD
CHRIS E. FORSMARK, MD, FACP
University of Florida College of Medicine

Pancreatitis on the Upswing

Estimates of the incidence of acute pancreatitis range from about five to 25 cases per 100,000 population. In the United States, 166,000 to 224,000 patients are admitted each year with a primary diagnosis of acute pancreatitis.[1] The number of patients discharged from the hospital with a diagnosis of acute pancreatitis has shown a steady increase over the past 20 years.[2] Similar trends have been seen in other developed countries.[3] The explanation for this increase is unclear but may relate to increased risk of gallstones (one of the major causes of acute pancreatitis)[4]; the prevalence of gallstones is rising in most developed countries, as a result of both the aging of the population and the obesity epidemic.[5]

1. DeFrances CJ, Hall MJ, Podgornik MN: 2003 National Hospital Discharge Survey. U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, Advance data from vital and health statistics; no. 359. National Center for Health Statistics, July 8, 2005

2. Lowenfels AB, Sullivan T, Fiorianti J, et al: The epidemiology and impact of pancreatic diseases in the United States. Curr Gastroenterol Rep 7:90, 2005 [PMID 15802095]

3. Lindkvist B, Appelros S, Manjer J, et al: Trends in incidence of acute pancreatitis in a Swedish population: is there really an increase? Clin Gastroenterol Hepatol 2:831, 2004 [PMID 15354285]

4. Overweight and obesity: obesity trends: U.S. obesity trends 1985–2004. Centers for Disease Control and Prevention. http://www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/index.htm

5. Kang JY, Ellis C, Majeed A, et al: Gallstones—an increasing problem: a study of hospital admissions in England between 1989/1990 and 1999/2000. Aliment Pharmacol Ther 17:561, 2003 [PMID 12622765]

A Better Test for Acute Pancreatitis

Measurement of the serum lipase level is often used as an adjunct to or substitute for serum amylase testing as a confirmatory test for acute pancre-

continued on page 3

In This Issue

PRACTICE OF MEDICINE

continued from page 1

published in many newspapers and in *Newsweek* magazine.

Although these well-publicized, evidence-based performance measures have been carefully reviewed by multiple professional organizations, clinicians can reasonably question whether better performance at a given hospital with respect to these measures translates into better outcomes. Preliminary data from the National Registry of Myocardial Infarction (NRMI) suggest that this is the case. Using data from more than 250,000 patients in the NRMI registry, hospitals were divided into four quartiles on the basis of a composite of these discharge performance measures. There was a clear difference in patient mortality between hospitals in the poorest quartile (15.3%) and those in the best quartile (8.3%).

Some physicians have concluded that they can't correct the multiple systems problems that affect the quality of inpatient care. Nevertheless, several well-publicized efforts have suggested that rapid systems change is indeed possible with the assistance of highly motivated physician leaders. The American College of Cardiology's Guidelines Applied in Practice (GAP) project published results from 10 hospitals that showed significant improvement in nearly all performance indicators, particularly for hospitals that employed the available admission and discharge tools.[1] Similar improvement has been documented by the Get With The Guidelines project, sponsored by the American Heart Association, which has now been adopted by more than 1,000 hospitals across the United States.[2] The Intermountain Health Care System showed that a discharge medication program could improve compliance with medications and subsequent patient outcomes.[3]

The most dramatic evidence in support of quality-improvement efforts has recently been published by the ACC's GAP project.[4] During the project, hospital mortality after acute myocardial infarction declined from 13.6% to 10.4%, and 1-year mortality declined from 38.3% to 33.2%. The authors reported that use of the standard discharge document alone was able to reduce adjusted 1-year mortality by almost one half (odds ratio, 0.53; 95% confidence interval, 0.36 to 0.76; $P = 0.0006$).

In summary, the care of patients with acute myocardial infarction too often fails to conform to existing practice guidelines, and these shortcomings have drawn the attention not only of physicians but also of the general public. Several different quality-improvement efforts have proven ability to improve performance measures and patient mortality. All physicians and other health care professionals, individually and through their respective organizations, including the American College of Physicians, should take the lead in local and national efforts to improve the quality of care in acute myocardial infarction.

References

1. Mehta RH, Montoye CK, Gallogly M, et al: Improving quality of care for acute myocardial infarction: the Guidelines Applied in Practice (GAP) initiative. JAMA 287:1269, 2002 [PMID 11886318]

2. LaBresh KA, Ellrodt AG, Gliklich R, et al: Get with the guidelines for cardiovascular secondary prevention: pilot results. Arch Intern Med 164:203, 2004 [PMID 14744845]

3. Lappe JM, Muhlestein JB, Lappe DL, et al: Improvements in 1-year cardiovascular clinical outcomes associated with a hospital-based discharge medication program. Ann Intern Med 141:446, 2004 [PMID 15381518]

4. Eagle KA, Montoye CK, Riba AL, et al: Guideline-based standardized care is associated with substantially lower mortality in Medicare patients with acute myocardial infarction: the American College of Cardiology's Guidelines Applied in Practice (GAP) Projects in Michigan. J Am Coll Cardiol 46:1242, 2005 [PMID 16198838]

raymondgibbonsmd@webmd.net

Next Month's Column: "Vitamin D and Cancer: A Goldilocks Paradox?" by Karen Antman, MD

ACP MEDICINE
A Publication of the American College of Physicians

Published by WebMD

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $349 for the first year ($324 for residents and students) and $269 for renewals ($249 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $449 for the first year and $369 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

THIS MONTH'S UPDATES

continued from page 1

atitis. Accurate measurement of lipase was difficult in the past, but new methods provide high levels of precision. The lipase level is in fact slightly more sensitive and somewhat more specific for acute pancreatitis than is serum amylase.[1] In addition, it stays elevated longer and can confirm a diagnosis of acute pancreatitis up to 5 to 10 days after the onset of symptoms, by which time amylase levels have generally returned to normal. Like amylase, lipase may be elevated in other intra-abdominal conditions (e.g., intestinal ischemia and perforation, bowel obstruction, choledocholithiasis, cholelithiasis with cholecystitis, and acute appendicitis), as well as in renal failure. Elevations that are more than three times the upper limit of normal have the greatest diagnostic sensitivity and specificity, but this threshold may need to be increased to five times the upper limit of normal in patients with renal failure. Lipase measurement is probably preferable to amylase measurement as a confirmatory test because in addition to its greater specificity, it is no more costly and, in most hospitals, has equally rapid availability.

1. Al-Bahrani AZ, Ammori BJ: Clinical laboratory assessment of acute pancreatitis. Clin Chim Acta 362:26, 2005 [PMID 16024009]

The Enteral Alternative

Nutritional support is useful for patients with severe pancreatitis and for those with milder pancreatitis who nonetheless are unable to eat for more than 5 to 7 days. The preferred route of providing exogenous nutrients has changed. For years, total parenteral nutrition (TPN) has been the standard practice, but accumulating evidence suggests that enteral feeding is equivalent to or better than TPN.[1,2] Prospective, randomized trials have demonstrated that enteral feeding infused distal to the ligament of Treitz is associated with a decreased rate of complications (i.e., infection and hyperglycemia) and lower cost than TPN. Although the evidence is not yet definitive,[2] it has led to a shift in the preferred method of providing nutrition to patients with acute pancreatitis. The main practical challenge in using enteral jejunal feeding is placing and maintaining position of the nasojejunal tube.

1. Kaushik N, O'Keefe SJ: Nutritional support in acute pancreatitis. Curr Gastroenterol Rep 6:320, 2004 [PMID 15245702]

2. Al-Omran M, Groof A, Wilke D: Enteral versus parenteral nutrition for acute pancreatitis. Cochrane Database Syst Rev (1):CD002837, 2003 [PMID 12535441]

The Questionable Value of Antibiotic Prophylaxis

The role of antibiotics in the prevention of pancreatic infection is controversial. Early studies using prophylactic ampicillin demonstrated no reduction in pancreatic infections. It was then found, however, that ampicillin does not penetrate necrotic pancreas at adequate concentrations. Subsequent studies used regimens with adequate penetration into necrotic pancreas (e.g., imipenem, cefuroxime, and ofloxacin plus metronidazole); a meta-analysis of these studies concluded that intravenous prophylactic antibiotic therapy for 10 to 14 days reduces the risk of superinfection of pancreatic necrosis and mortality.[1] However, this meta-analysis did not include the most recently reported study—the only double-blind randomized trial to date—which did not demonstrate any benefit from prophylactic treatment with ciprofloxacin plus metronidazole.[2] This has led to differing opinions on the overall utility of prophylactic antibiotics, with both advocates and opponents of this practice.

1. Bassi C, Larvin M, Villatoro E: Antibiotic therapy for prophylaxis against infection of pancreatic necrosis in acute pancreatitis. Cochrane Upper Gastrointestinal and Pancreatic Diseases Group. Cochrane Database Syst Rev (4):CD002941, 2003 [PMID 14583957]

2. Isenmann R, Runzi M, Kron M, et al: Prophylactic antibiotic treatment in patients with predicted severe acute pancreatitis: a placebo-controlled, double-blind trial. Gastroenterology 126:997, 2004 [PMID 15057739]

Pancreatitis in the Genes

Hereditary pancreatitis is an autosomal dominant disease that typically presents in childhood or early adulthood and frequently is accompanied by steatorrhea, diabetes mellitus, and diffuse pancreatic calcifications. Pain and acute episodes of pancreatitis flares may also occur but are somewhat less common in hereditary pancreatitis than in alcoholic chronic pancreatitis. The initially identified genetic abnormality is a defect in the cationic trypsinogen (*PRSS1*) gene on chromosome 7.[1] Multiple mutations have been described, but two are more common. The described mutations appear to produce a trypsinogen that, once activated, is difficult or impossible to inactivate. The activated enzyme, trypsin, can in turn activate all the other pancreatic enzymes. Chronic pancreatitis appears to be caused in this situation by prolonged low-grade pancreatic injury from the activated proteases. Pancreatic adenocarcinoma frequently complicates the condition, with a 30% risk by age 70. The risk may be substantially higher in patients with paternal inheritance.

1. Howes N, Greenhalf W, Stocken DD, et al: Cationic trypsinogen mutations and pancreatitis. Gastroenterol Clin North Am 33:767, 2004 [PMID 15528017]

13 PSYCHIATRY

II Depression and Bipolar Disorder

MICHAEL T. COMPTON, MD, MPH
CHARLES B. NEMEROFF, MD, PHD
Emory University School of Medicine

Depression or Dementia?

In the elderly, depression with cognitive impairment that is reversed by antidepressant treatment may be an early predictor of the development of an irreversible dementia. Patients with late-onset depression are more likely to have cognitive impairment and to have deeper white-matter lesions on imaging studies. However, the so-called pseudodementia of depression is common and should not be mistaken for a primary dementia.

User-Friendly MAOIs?

In the United States, three monoamine oxidase inhibitors (MAOIs) are currently available: phenelzine, isocarboxazid, and tranylcypromine.

FDA Approval Report

The following is selected from the FDA's list of recently approved products. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

First Oral Drug for Chronic Iron Overload

Generic Name: Deferasirox

Brand Name: Exjade

Manufacturer: Novartis Pharmaceuticals Corp., Stein, Switzerland

The FDA has approved deferasirox, an oral iron chelator developed to treat chronic iron overload resulting from multiple blood transfusions. Deferasirox is the first orally administered medication to be approved for this use. Previously, treatment for iron overload required daily intravenous infusions lasting 8 to 12 hours.

Deferasirox was approved under a program that allows accelerated approval of products to treat serious or life-threatening diseases on the basis of early evidence of efficacy. The evidence for deferasirox was derived from clinical studies of 48 weeks' duration, which demonstrated reductions in liver iron concentrations in adult and pediatric patients who were receiving red blood cell transfusions on an ongoing basis. Deferasirox has also received Orphan Drug Designation, which is granted to products that treat diseases affecting fewer than 200,000 people in the United States.

In clinical studies of deferasirox, common side effects included nausea and abdominal pain. Elevations in measures of kidney and liver function were also noted. Less common side effects included hearing and visual disturbances and rashes. Monitoring of kidney and liver function and testing of hearing and vision before and during treatment are recommended.

Source:
FDA Approves First Oral Drug for Chronic Iron Overload. FDA News. U.S. Food and Drug Administration, November 9, 2005 (http://www.fda.gov/bbs/topics/NEWS/2005/NEW01258.html)

These agents are rarely used because their use is complicated by side effects (including hypotension), lethality in overdose, and lack of simplicity in dosing. Patients treated with MAOIs must follow a specific tyramine-free diet because of the potential for a pharmacodynamic interaction with tyramine that can result in a hypertensive crisis. Drugs that have been reported to interact with MAOIs include carbamazepine, cyclobenzaprine, dextromethorphan, fenfluramine, certain hypoglycemics, L-tryptophan, meperidine, selective serotonin reuptake inhibitors (SSRIs), stimulants, sympathomimetics, and tricyclic antidepressants. Nonetheless, the MAOIs appear to be more effective than other antidepressants in the treatment of atypical depression, though their use is usually limited to psychiatrists who have experience with these agents. MAOIs are also often effective in patients with depression that is refractory to treatment with other antidepressants.

Selective MAOIs (e.g., moclobemide), which are available outside of the United States, are easier to use because of the lack of dietary constraints and drug-drug interactions. A transdermal system for delivery of selegiline, an MAOI approved by the Food and Drug Administration for the treatment of Parkinson disease, may be effective in the treatment of depression.[1,2]

1. Bodkin JA, Amsterdam JD: Transdermal selegiline in major depression: a double-blind, placebo-controlled, parallel-group study in outpatients. Am J Psychiatry 159:1869, 2002 [PMID 12411221]

2. Amsterdam JD: A double-blind, placebo-controlled trial of the safety and efficacy of selegiline transdermal system without dietary restrictions in patients with major depressive disorder. J Clin Psychiatry 64:208, 2003 [PMID 12633131]

SSRIs and Side Effects

SSRIs have a relatively mild side-effect burden. Sexual side effects of these medications (i.e., delayed ejaculation, anorgasmia, and decreased libido) are being increasingly recognized and studied and are one of the leading causes of discontinuance or noncompliance. Although the correlation between SSRI treatment and increased suicide risk remains controversial, the potential for increased risk has prompted the Food and Drug Administration to require that manufacturers of all antidepressants include in their labeling a boxed warning (http://www.fda.gov/cder/drug/antidepressants/default.htm).[1,2]

1. Khan A, Khan S, Kolts R, et al: Suicide rates in clinical trials of SSRIs, other antidepressants, and placebo: analysis of FDA reports. Am J Psychiatry 160:790, 2003 [PMID 12668373]

2. Grunebaum MF, Ellis SP, Li S, et al: Antidepressants and suicide risk in the United States, 1985–1999. J Clin Psychiatry 65:1456, 2004 [PMID 15554756]

Atypical Antipsychotics in Bipolar Disorder

The atypical antipsychotic agents are effective for the treatment of acute mania. Olanzapine, risperidone, quetiapine, ziprasidone, and aripiprazole have FDA indications for bipolar mania. The risks of weight gain and metabolic disturbances are increasingly being recognized with some of these agents; therefore, clinical monitoring is necessary.[1,2]

1. Consensus development conference on antipsychotic drugs and obesity and diabetes. American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Diabetes Care 27:596, 2004 [PMID 14747245]

2. Casey DE, Haupt DW, Newcomer JW, et al: Antipsychotic-induced weight gain and metabolic abnormalities: implications for increased mortality in patients with schizophrenia. J Clin Psychiatry 65(suppl 7):4, 2004 [PMID 15151456]

7 INFECTIOUS DISEASE

XXVIII Enteric Viral Infections

NINO KHETSURIANI, MD, PhD
UMESH D. PARASHAR, MD, MPH
Centers for Disease Control and Prevention

Fever from Enteroviruses

Enteroviruses are a common cause of febrile illness, particularly during the summer and fall months. Fever may be the sole symptom, or infection may be accompanied by rashes or respiratory symptoms (the "summer cold"); in infants, symptoms include irritability, lethargy, anorexia, vomiting, and diarrhea. Because of the need to rule out potentially serious bacterial infections, a high proportion of infants with enteroviral febrile illness require evaluation for bacterial sepsis or meningitis.[1]

1. Rittichier KR, Bryan PA, Bassett KE, et al: Diagnosis and outcomes of enterovirus infections in young infants. Pediatr Infect Dis J 24:546, 2005 [PMID 15933567]

The End of Polio?

Through successful immunization programs, control of poliomyelitis has been achieved in numerous countries worldwide, including the entire Western Hemisphere, but wild polioviruses continue to circulate in some countries of sub-Saharan Africa and Southeast Asia. It is likely that the large-scale Polio Eradication Initiative, led by the World Health Organization since 1988, will result in global polio eradication[1]; until that occurs, there remains a risk of importation of poliomyelitis into the United States.

1. Progress toward interruption of wild poliovirus transmission—worldwide, January 2004–March 2005. MMWR Morb Mortal Wkly Rep 54:408, 2005 [PMID 15858461]

Rotavirus Vaccine, Round Two

A vaccine against severe rotavirus disease was licensed in the United States in 1998, but it was withdrawn in 1999 because its use was associated with intussusception. Efforts are ongoing to develop other rotavirus vaccines. Two large clinical trials that each involved more than 60,000 infants have been completed. The trials examined two leading vaccine candidates: a multivalent bovine-human reassortant rotavirus vaccine[1] and an attenuated single-human-strain rotavirus vaccine.[2] These trials demonstrated the safety of both vaccines with respect to intussusception and other potential adverse events,[3,4] and they indicated that the vaccines have an efficacy of more than 90% against severe rotavirus disease.[1,2] Data on the multivalent vaccine, which is administered orally to infants in three doses at 2, 4, and 6 months of age, were submitted for licensure to the FDA in April 2005; a decision is expected in 2006.

1. Salinas B, Schael I, Linhares AC, et al: Evaluation of safety, immunogenicity and efficacy of an attenuated rotavirus vaccine, RIX 4414: a randomized, placebo-controlled trial in Latin American infants. Pediatr Infect Dis J 24:807, 2005 [PMID 16148848]

2. Velasquez FR, Abate H, Clemens SA: The human monovalent G1P[8] rotavirus vaccine, Rotarix, is highly efficacious and provides cross protection against G1 and non-G1 serotypes. Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

3. Vesikari T, Matson D, Van Damme P, et al: Incidence of intussusception with the pentavalent (human-bovine) reassortant rotavirus vaccine (PRV) is similar to placebo. Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

4. Vesikari T, Matson D, Dennehy P, et al: Protection against rotavirus gastroenteritis of multiple serotypes by a pentavalent (human-bovine) reassortant rotavirus vaccine (PRV). Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID), Valencia, Spain, May 18–20, 2005

8 INTERDISCIPLINARY MEDICINE

I Management of Poisoning and Drug Overdose

KENT R. OLSON, MD, FACP
University of California, San Francisco, Schools of Medicine and Pharmacy
MANISH M. PATEL, MD
Emory University School of Medicine

Ipecac out

Ipecac-induced emesis has been almost completely abandoned in the clinical setting.[1] In 2003, the American Academy of Pediatrics advised against routine use of ipecac in the home and recommended disposing existing ipecac in the home.[2] One reason it has fallen out of favor is that treated patients run the risks of sudden, unexpected deterioration from the effects of overdose and subsequent pulmonary aspiration; more important, however, is the lack of evidence of the efficacy of ipecac-induced emesis, especially when emesis is induced more than 1 hour after the ingestion.

In addition to emesis, other techniques for gut decontamination include gastric lavage, administration of activated charcoal, and whole bowel irrigation [*see Table, page 6*]. However, nowhere in the field of toxicology is there more controversy than in the debate about gastrointestinal decontamination.

1. Position paper: ipecac syrup. J Toxicol Clin Toxicol 42:133, 2004 [PMID 15214617]

2. Poison treatment in the home. American Academy of Pediatrics Committee on Injury, Violence, and Poison Prevention. Pediatrics 112:1182, 2003 [PMID 14595067]

Parenteral Treatment for Acetaminophen Overdose

Acetylcysteine is highly effective in preventing liver damage from acetaminophen toxicity, especially if therapy is initiated within 8 to 10 hours after the ingestion of acetaminophen. Occasionally, however, patients cannot tolerate oral aetylcysteine because the drug has a disagreeable odor and they are already vomiting. In such cases, it is advisable to administer the drug by the intravenous route. In 2004, the FDA approved a 20-hour protocol for intravenous acetylcysteine treatment of acetaminophen overdose.[1]

The initial loading dose is 150 mg/kg in 200 ml of 5% dextrose in water (D5W) over 15 minutes. This is followed by 50 mg/kg in 500 ml D5W over 4 hours, then 100 mg/kg in 1 L D5W over the next 16 hours. Intravenous administration can cause an anaphylactoid reaction (i.e., skin flushing and hypotension); this can be minimized by slowing the rate of the initial loading dose, giving it over 45 to 60 minutes.

1. Safety labeling changes approved by FDA Center for Drug Evaluation and Research (CDER)—April 2004. Acetadote (acetylcysteine) injection. http://www.fda.gov/medwatch/SAFETY/2004/apr_PI/Acetadote_PI.pdf

Ephedra, Gone and Back Again

Ephedra (Ma Huang) is a common ingredient in herbal weight-loss products (herbal fen-phen), stimulants (herbal ecstasy), decongestants, and bronchodilators. The active moiety in ephedra is ephedrine and related alkaloids. Serious adverse reactions, in-

Methods of Gastrointestinal Decontamination

Method and Technique	*Useful Situations*	*Comments*
Emesis: give syrup of ipecac, 30 ml p.o. in adults (15 ml in children), along with one to two glasses of water; may repeat after 30 min if no emesis occurs; alternatively, give 1–2 tbsp of liquid handwashing or dishwashing soap	Possible benefit in rare circumstances after a potentially lethal ingestion when medical care is more than 60 min away, but only under the guidance of a poison control center	No longer used in emergency departments; American Academy of Pediatrics recommends against routine home use of ipecac; contraindicated in ingestions of corrosive agents and most hydrocarbons, when the patient is lethargic, or when the ingested substance is likely to cause abrupt onset of coma or seizures
Gastric lavage: insert large-bore nasogastric or orogastric tube, empty stomach contents, and lavage with 100–200 ml aliquots of water or saline until clear	Useful in obtunded or comatose patients, in recent ingestions (< 1 hr), or in ingestion of anticholinergic agents or salicylates (delayed gut emptying)	Obtunded patient should have prior endotracheal intubation to protect airway; best position is left lateral decubitus to reduce movement of poison into small intestine
Activated charcoal: give 50–60 g of charcoal slurry p.o. or by gastric tube; goal is approximately 10:1 ratio of charcoal to ingested poison; usually given with one dose of a cathartic agent	Often useful because it adsorbs most drugs and poisons; may be equally effective when given alone as when given after emesis or lavage	Not effective for ingestions of iron, lithium, potassium, sodium, or alcohols; may need to repeat two or three times or more for large ingestions; repeated dosing may also enhance elimination of some drugs
Whole bowel irrigation: give Colyte or GoLYTELY, 1–2 L/hr p.o. or by gastric tube, until rectal effluent is clear or x-ray is negative for radiopaque materials	Useful in ingestions of iron, lithium, sustained-release or enteric-coated pills, and drug packets or other foreign bodies	Generally well tolerated; no significant fluid or electrolyte gain or loss occurs; most useful in awake, ambulatory patients; may reduce effectiveness of activated charcoal

cluding hypertension, seizures, arrhythmias, heart attack, stroke, and death, have been reported. In 2004, the FDA declared dietary supplements containing ephedra to be unsafe and banned ephedra-containing supplements.[1] However, a federal judge reversed the ban in early 2005, and the future of this substance remains uncertain.

1. Final rule declaring dietary supplements containing ephedrine alkaloids adulterated because they present an unreasonable risk: final rule. Fed Regist 69:6787, 2004 [PMID 14968803]

11 NEUROLOGY

VII Anoxic, Metabolic, and Toxic Encephalopathies

MICHAEL J. AMINOFF, MD, DSC
University of California, San Francisco, School of Medicine

Will This Patient Awaken from Coma?

In the first several days after the onset of ischemic encephalopathy, diffusion-weighted magnetic resonance imaging may be useful in determining prognosis; hyperintense cortical lesions are associated with a poor outcome.[1,2]

Studies of evoked potentials can also have prognostic utility. In a systematic review of somatosensory evoked potentials performed early after onset of coma, rates of awakening in adults with hypoxic-ischemic encephalopathy ranged from 0% when somatosensory evoked potentials were absent to 52% when somatosensory evoked potentials were normal.[3] In a prospective study of 346 comatose patients, pupillary reflex was the strongest prognostic variable for awakening. The probability of awakening was higher when late auditory evoked potentials were also present, and it was higher still when middle-latency evoked potentials were present. No patient with cognitive potentials became permanently vegetative.[4]

1. McKinney AM, Teksam M, Felice R, et al: Diffusion-weighted imaging in the setting of diffuse cortical laminar necrosis and hypoxic-ischemic encephalopathy. AJNR Am J Neuroradiol 25:1659, 2004 [PMID 15569727]

2. Els T, Kassubek J, Kubalek R, et al: Diffusion-weighted MRI during early global cerebral hypoxia: a predictor for clinical outcome? Acta Neurol Scand 110:361, 2004 [PMID 15527448]

3. Robinson LR, Micklesen PJ, Tirschwell DL, et al: Predictive value of somatosensory evoked potentials for awakening from coma. Crit Care Med 31:960, 2003 [PMID 12627012]

4. Fischer C, Luaute J, Adeleine P, et al: Predictive value of sensory and cognitive evoked potentials for awakening from coma. Neurology 63:669, 2004 [PMID 15326240]

Encephalopathy at High Altitude

High-altitude sickness can lead to an encephalopathy characterized by headache, fatigue, anorexia, nausea, poor concentration, and sleep disturbances.[1] Symptoms of high-altitude sickness begin within hours or days of ascent to altitudes above 10,000 ft. In severe cases or at higher altitudes, consciousness is impaired and coma may occur—sometimes with a fatal outcome. Cerebral edema causes papilledema, retinal hemorrhages, cranial neuropathies, a variety of sensorimotor deficits, and behavioral disturbances. High-altitude cerebral edema arises because the low barometric pressure encountered at high altitudes causes a reduction in the partial pressure of oxygen (P_{O_2}); the condition often follows acute mountain sickness and may be an extreme form of that disorder. Acetazolamide (250 mg once or twice a day) or dexamethasone (2 mg every 6 to 8 hours) may prevent acute mountain sickness. High-altitude cerebral edema is treated with prompt descent to a lower altitude and administration of oxygen and dexamethasone (8 mg, then 4 mg every 6 hours).[1]

1. West JB: The physiologic basis of high-altitude diseases. Ann Intern Med 141:789, 2004 [PMID 15545679]

Encephalopathy with and without Anemia

Encephalopathy is a well-recognized complication of vitamin B_{12} deficien-

cy. It may be accompanied by myelopathy, optic neuropathy, peripheral neuropathy, or any combination of these conditions. The neurologic complications do not reflect the presence or the severity of any associated megaloblastic anemia. Folic acid masks the hematologic abnormality and fails to prevent the neurologic complications. It was feared that fortification of cereal grains with folic acid, which began in 1994 in the United States, might result in an increase in the incidence of vitamin B_{12} deficiencies in persons who are not anemic. However, a 2003 Veterans Affairs study found no evidence that such an increase has occurred since fortification began.[1]

1. Mills JL, Von Kohorn I, Conley MR, et al: Low vitamin B-12 concentrations in patients without anemia: the effect of folic acid fortification of grain. Am J Clin Nutr 77:1474, 2003 [PMID 12791626]

SPECIAL – ONLINE ONLY!

16 WOMEN'S HEALTH

Introduction to Women's Health: The Primary Care of Women

JANET B. HENRICH, MD
Yale University School of Medicine

Drugs and Women's Hearts

When obtaining the medication history, clinicians need to be aware of important sex differences in the effects of certain classes of medications that may put women at increased risk for adverse drug reactions.[1] For example, because women have longer baseline QT intervals than men, the use of drugs that further prolong the QT interval, such as certain antibiotics, antihistamines, antiarrhythmics, and antipsychotics, may put women at increased risk for torsade de pointes, a potentially lethal ventricular arrhythmia.[2] The University of Arizona Health Sciences Center maintains an online registry of drugs that prolong the QT interval or induce torsade de pointes (http://www.arizonacert.org/medical-pros/drug-lists/drug-lists.htm). Also, women with depressed left ventricular function who use digoxin for heart failure are at greater risk of death than men with a similar diagnosis. This risk is attributed partly to higher serum digoxin levels in women. Because of this potential harm, clinicians are advised to weigh carefully the risks versus benefits of digoxin therapy in women with heart failure.[3]

1. Anderson GD: Sex and racial differences in pharmacological response: where is the evidence? Pharmacogenetics, pharmacokinetics, and pharmacodynamics. J Womens Health (Larchmt) 14:19, 2005 [PMID 15692274]
2. Anthony M: Male/female differences in pharmacology: safety issues with QT-prolonging drugs. J Womens Health (Larchmt) 14:47, 2005 [PMID 15692277]
3. Rathore SS, Krumholz HM: Digoxin therapy for heart failure: safe for women? Ital Heart J 4:148, 2003 [PMID 12784740]

Oral Contraceptives versus Antibiotics

The hepatic cytochrome P-450 family of enzymes is involved in the metabolism of endogenous and exogenous steroids and many other drugs. Anticonvulsants can lead to oral contraceptive (OC) failure by inducing the cytochrome P-450 enzyme system to metabolize estrogen more rapidly. A similar mechanism has been attributed to antibiotics, leading to drug warnings about the risk of pregnancy when antibiotics are prescribed to women on OCs. Because the evidence supporting drug interactions between antibiotics and OCs is weak, the Council on Scientific Affairs of the American Medical Association convened a panel to review the data and provide recommendations. The panel concluded that rifampin, acting primarily as an inducer of the cytochrome P-450 3A4 isoenzyme, is the only antibiotic tested that significantly reduces plasma concentrations of OC-derived estrogens and increases OC failure rates. However, women using several other commonly prescribed antibiotics have wide variations in OC-derived hormone levels; some women have elevated follicle-stimulating hormone levels or breakthrough bleeding, suggesting that ovulation may occur.

On the basis of these findings, the panel concluded that even though the risk of OC failure with these antibiotics is very small, some women may be at risk, particularly those using low-dose or very low dose OCs. The panel recommended nonhormonal contraceptive methods for women who are concomitantly using OCs and rifampin, as well as for women who are taking other antibiotics and are concerned about a small risk of pregnancy, have had previous contraceptive failure, or develop breakthrough bleeding.

Answering the Aspirin Question

On the basis of findings from studies that were conducted primarily in men, the United States Preventive Services Task Force currently recommends that clinicians discuss the use of aspirin for the primary prevention of coronary heart disease (CHD) with postmenopausal women and with younger women with risk factors for CHD. However, results from the first randomized trial that assessed the risks and benefits of aspirin as chemoprevention for CHD in women found a beneficial effect only in women 65 years of age and older.[1] In younger women, aspirin reduced the risk of thromboembolic stroke but had little effect on CHD—findings opposite to those described in men. As in men, the use of aspirin in older women was associated with an increased risk of hemorrhagic stroke and major gastrointestinal hemorrhage. Until more information is available about aspirin's effects in women, discussions about its use should be individualized for each patient.

1. Ridker PM, Cook NR, Lee IM, et al: A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. N Engl J Med 352:1293, 2005 [PMID 15753114]

Coming in March

Clinical Essentials
VIII Quantitative Aspects of Clinical Decision Making

7 Infectious Disease
X Infections Due to Haemophilus, Moraxella, Legionella, Bordetella, and Pseudomonas
XXIV Hyperthermia, Fever, and Fever of Undetermined Origin

11 Neurology
XI Alzheimer Disease and Other Major Dementing Illnesses
XVI Acute Viral Central Nervous System Diseases

Best Dx/Best Rx by *ACP Medicine*

Best Dx/Best Rx is a free tool that provides fast access to expert guidance at the point of care. Each module of Best Dx/Best Rx is based on a chapter of ACP Medicine, *whose authors are recognized authorities in clinical medicine. Their recommendations in Best Dx/Best Rx are intended for use in clinical settings to help answer these essential questions:*

• How do I diagnose? • How do I treat? • What is the evidence?

Best Dx/Best Rx is available within chapters on *ACP Medicine Online*. The modules may also be accessed on any HTML browser, including wireless PDAs and computers.

Below is our new module on tachycardia. Other topics include rheumatoid arthritis, cirrhosis, hypertension, alzheimer disease, and many more. Go to www.acpmedicine.com/dxrxpromo1.htm to read other modules and for more information on how to get Best Dx/Best Rx.

Atrioventricular Nodal Reentry Tachycardia (AVNRT)

Melvin M. Scheinman, M.D., F.A.C.P.
University of California, San Francisco, School of Medicine

Definition/Key Clinical Features

- AVNRT is the most common form of supraventricular tachycardia; it results from conduction through a reentrant circuit comprising fast and slow atrioventricular nodal pathways
- Heart rate, 150–250 beats/min
- Neck pounding
- Palpitations, light-headedness, near-syncope
- Narrow QRS complexes on ECG
- Abrupt onset and termination of episodes
- More common in women than in men
- Frequently presents after 20 yr of age

Differential Diagnosis

- Atrial fibrillation
- Atrial flutter
- AV reentrant tachycardia

Best Tests

- 12-lead ECG
  - The P wave is either buried within the QRS complex or inscribed just after the QRS complex
- The P wave inscribed by retroconduction over the AV node is negative in the inferior leads and positive in lead V1; PSVT may manifest as small negative deflections in the inferior leads and a small positive deflection in V1 (pseudo r^1 pattern)

Best Therapy

Acute Therapy

Nonpharmacologic

- Carotid sinus massage

Pharmacologic

- Adenosine: for use when carotid massage fails to convert SVT
  - Possible adverse effects: headache, wheezing, flushing, which will disappear within 45–60 sec; atrial, ventricular, and junctional premature beats; atrial fibrillation in 3% to 5% of cases, which may result in serious problems for patients with accessory pathways
  - An external defibrillator should be readily available when adenosine is administered
  - Dose: initial dose: rapid bolus of 6 mg I.V., followed by a saline flush; if necessary, a 12 mg dose and finally an 18 mg dose can be given
- Metoprolol (5 mg I.V.) or verapamil (0.1 mg/kg I.V.) for patients who fail to respond to adenosine

Long-term Therapy

- Associated with frequent recurrences and adverse effects

Pharmacologic

- Beta blockers
- Sotalol
  - Common side effects: torsade de pointes, heart failure, bradycardia, exacerbation of chronic obstructive or bronchospastic lung disease
  - Adjust dose for renal function and QT-interval response during in-hospital initiation phase
  - Dose: 240–320 mg/day
  - Cost/mo: $138
- Calcium channel blockers (verapamil, diltiazem)
- Digoxin
  - Dose: 0.1 mg/day
  - Cost/mo: $9
- Antiarrhythmic agents: for patients without structural cardiac disease; more effective than beta blockers and calcium channel blockers, but with 25% to 35% recurrence rates
- Amiodarone
  - Common side effects: photosensitivity, pulmonary toxicity, polyneuropathy, GI upset, bradycardia, torsade de pointes (rare), hepatic toxicity, thyroid dysfunction
  - Dose: 100–400 mg/day
  - Cost/mo: $94
- Disopyramide
  - Common side effects: torsade de pointes, heart failure, glaucoma, urinary retention, dry mouth
  - Dose: 400–750 mg/day
  - Cost/mo: $84
- Dofetilide
  - Common side effect: torsade de pointes
  - Dose: 500–1,000 mg/day
  - Cost/mo: N/A
- Flecainide
  - Common side effects: ventricular tachycardia, heart failure, enhanced AV nodal conduction (conversion to atrial flutter)
  - Dose: 200–300 mg/day
  - Cost/mo: $115
- Procainamide
  - Common side effects: torsade de pointes, lupuslike syndrome, GI symptoms
  - Dose: 1,000–4,000 mg/day
  - Cost/mo: $59
- Propafenone
  - Common side effects: ventricular tachycardia, heart failure, enhanced AV nodal conduction (conversion to atrial flutter)
  - Dose: 490–900 mg/day
  - Cost/mo: $198
- Quinidine
  - Common side effects: torsade de pointes, GI upset, enhanced AV nodal conduction
  - Dose: 600–1,500 mg/day
  - Cost/mo: $67
- Single-dose agents for p.r.n. use: diltiazem, 120 mg, or propranolol, 80 mg; p.r.n. dosing regimens are used with caution and only after efficacy and safety have been established for the individual patient under ECG monitoring

Nonpharmacologic

- Catheter ablation
  - Procedure of choice for patients in whom drug therapy fails and those with milder symptoms who prefer to avoid long-term drug therapy
  - Success rate > 96%
  - AV block occurs in ~ 1% of patients

Best References

Kwaku KF, et al: Card Electophysiol Rev 6:414, 2002
Jackman WM, et al: N Engl J Med 327:313, 1992

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 29 No. 1 January 2006

What's New in ACP Medicine

www.acpmedicine.com

PRACTICE OF MEDICINE

Sensing a Role for Innate Immunity in Chronic Disease

JOHN P. ATKINSON, MD, FACP
Washington University School of Medicine

The immune system has a hard job. It must defend against infectious agents that can mutate at alarming rates. It must distinguish self from nonself, or the body will suffer the consequences of autoimmunity. At epithelial borders, it has the particularly difficult task of separating dangerous from benign microorganisms. This task, which is carried out by the innate immune system, has two primary components: sensing when a microorganism is present, and then triggering the appropriate response. Both components are critical. The timely recognition of a "bad guy" and the eventual destruction of same are necessary events. Otherwise, an acute infection may become lethal or lead to a chronic process. At the other extreme, mistaking the harmless for the perilous can result in an allergic response.

Progress in the understanding of the innate immune system maps the early history of the field of immunology. Phagocytosis and the lytic activity of the complement system were two innate immune processes recognized in the late 19th century. Investigation of antibody, complement, and phagocytosis dominated the field of immunology until the middle of the 20th century. The study of the host response to infectious

continued on page 2

THIS MONTH'S UPDATES

7 INFECTIOUS DISEASE

XXXVI Bacterial Infections of the Central Nervous System

JAN V. HIRSCHMANN, MD
University of Washington School of Medicine

Classic versus Likely Signs of Meningitis

The most common features of bacterial meningitis are headache, fever, nuchal rigidity, and neurologic findings. Less than half of patients have the classic triad of fever, neck stiffness, and a change in mental status; however, almost all have at least two of the four manifestations of headache, fever, neck stiffness, and altered mental status.[1] Fever is present in approximately 95% of patients and typically lasts 4 to 8 days after appropriate therapy has begun. Stiff neck is apparent in about 90% of patients. Mental changes ranging from lethargy to confusion, stupor, and coma occur in about 80% of patients, somewhat more frequently in the elderly.

1. van de Beek D, de Gans J, Spanjaard L, et al: Clinical features and prognostic factors in adults with bacterial meningitis. N Engl J Med 351:1849, 2004 (erratum in N Engl J Med 352:950, 2005) [PMID 15509818]

Tap First, Don't Scan?

In patients with suspected meningitis, clinicians commonly obtain a computed tomography scan before doing a lumbar puncture, on the basis of two beliefs: that brain herniation is a frequent risk in meningitis and that CT scans can accurately predict its development. Neither belief is correct. Herniation occurs in approximately 1% of patients with bacterial meningitis, sometimes without a preceding lumbar puncture. CT scans are often normal in those who later experience brain herniation, and most patients who develop this complication have focal neurologic findings before the lumbar puncture that suggest that herniation has already begun: dilated, fixed pupils; Cheyne-Stokes respiration; decerebrate posturing; hemiplegia; and coma.

In addition to focal neurologic deficits and an abnormal level of consciousness, findings that justify per-

continued on page 3

In This Issue

PRACTICE OF MEDICINE

continued from page 1

organisms and to vaccination, as well as the analysis of blood transfusion reactions, contributed much to our understanding of the workings of the immune system. A surprising discovery was that two components of blood that are required to opsonize and lyse microorganisms—namely, antibody and complement—also mediate mismatched transfusion reactions and autoimmunity. Thus, early on, the immune system was recognized as a double-edged sword, protecting us against infections on the one hand but mediating tissue damage through allergic, autoimmune, and transfusion reactions on the other.

With the discovery of B and T cells, the attention of immunologists turned to the adaptive immune system. Research into the generation of the adaptive response dominated immunology from the 1960s to the present and brought remarkable gains in both basic and clinical knowledge about adaptive immunity. For example, the discovery over 3 decades ago of the role of cytokines and lymphokines in immune responses led to highly effective therapeutic agents (i.e., interleukin-1 and tumor necrosis factor antagonists).

Interest in the innate immune system was rekindled about a decade ago by the discovery that Toll receptors act as sensors of microbes. Toll receptors are expressed on epithelial cells, where they interface with the environment. They recognize structural features of microorganisms that are distinct from those expressed by the host—for example, endotoxin, peptidoglycans, flagellin, and unique structural features of viral and bacterial DNA and RNA. These structures form the backbone of microorganisms and thus cannot be altered to any major degree. To do so would be like primates changing the basic elements of muscle or bone.

In addition to Toll receptors, the innate immune system has many other sensors for pathogens, including natural antibodies, lectins, natural killer cells, the complement system, and phagocytes themselves. These sensors are responsible for identifying pathogens in the skin, the gut, the upper airway, and the urogenital tract. This recognition event (which may occur either outside or inside the cell) calls forth an inflammatory response. In addition, instructions are sent to the adaptive immune system regarding the location and nature of the problem. For example, the innate immune sensors can distinguish gram-positive from gram-negative bacteria and between bacteria and viruses. The information from the innate immune system about who is trying to invade provides vital direction for the adaptive immune system.

An exciting discovery for clinical medicine is that many stubborn conditions whose cause has been obscure—for example, psoriasis, inflammatory bowel disease, sarcoidosis, and certain arthritides—may originate from defects in innate immune sensing. Thus, one could envision Crohn disease arising in two general ways: (1) from an overly active sensor on the epithelial cells that sparks an innate immune reaction to normal gut flora or (2) from a failure to properly regulate the tonic signals being sent by normal sensors from an epithelial surface (i.e., the gut) that is bathed in microorganisms. Recent genetic studies of familial forms of these perplexing syndromes have identified mutations in such sensors or in regulators of the response that is triggered by these sensors. In addition, defects in this system appear to be the source of rare, recurrent, inflammatory diseases such as familial Mediterranean fever and several related periodic syndromes. Similar types of defects may be at the heart of allergies and autoimmunity, as well.

johnatkinsonmd@webmd.net

Next Month's Column: Quality Improvement in Myocardial Infarction

by Raymond J. Gibbons, M.D., *Mayo Clinic College of Medicine*

ACP MEDICINE

A Publication of the American College of Physicians

Published by WebMD

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $349 for the first year ($324 for residents and students) and $269 for renewals ($249 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $449 for the first year and $369 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

THIS MONTH'S UPDATES

continued from page 1

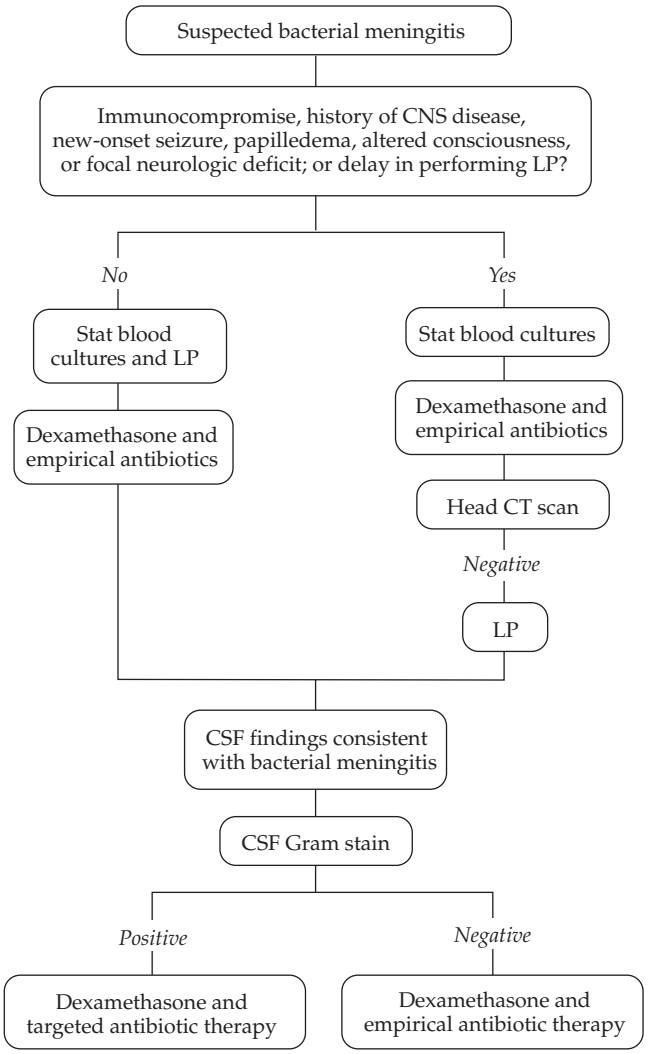
forming CT before lumbar puncture include the following: immune compromise; a history of central nervous system disease (e.g., mass lesion, stroke, or focal infection); papilledema (although the presence of venous pulsations suggests that the patient does not have increased intracranial pressure); and seizures within 1 week before presentation (some experts will not perform a lumbar puncture in patients with prolonged seizures, and they will delay lumbar puncture for 30 minutes in patients who have experienced short, convulsive seizures).[1] In the absence of such findings in patients with suspected bacterial meningitis, clinicians should not delay lumbar puncture to obtain a CT scan [*see Figure, right*].

1. Tunkel AR, Hartman BJ, Kaplan SL, et al: Practice guidelines for bacterial meningitis. Clin Infect Dis 39:1267, 2004 [PMID 15494903]

The Role of Steroids in Meningitis Treatment

In adults with acute pneumococcal meningitis, adjuvant treatment with dexamethasone has been shown to lower the risk of an unfavorable outcome and to lower mortality without increasing the likelihood of gastrointestinal bleeding.[1] Consequently, guidelines from the Infectious Diseases Society of America recommend that adults with suspected or proven pneumococcal meningitis receive adjunctive dexamethasone.[2] The dosage is 0.15 mg/kg every 6 hours for 2 to 4 days, with the first dose given 10 to 20 minutes before, or at least concomitant with, the first dose of antibiotics. Dexamethasone should be continued only if the cerebrospinal fluid Gram stain reveals gram-positive diplococci, or if blood or CSF cultures are positive for *Streptococcus pneumoniae*. Dexamethasone treatment is unlikely to improve the outcome in adult patients who have already received antimicrobial therapy and hence should not be used in this circumstance.

1. van de Beek D, de Gans J, McIntyre P, et al: Steroids in adults with acute bacterial meningitis: a systematic review. Lancet Infect Dis 4:139, 2004 [PMID 14998499]



Management of adult patients with suspected bacterial meningitis. (CSF—cerebrospinal fluid; LP—lumbar puncture)

2. Tunkel AR, Hartman BJ, Kaplan SL, et al: Practice guidelines for bacterial meningitis. Clin Infect Dis 39:1267, 2004 [PMID 15494903]

12 ONCOLOGY

V Colorectal Cancer

BERNARD LEVIN, MD, FACP
University of Texas M. D. Anderson Cancer Center

Testing for Colon Cancer Genes

Genetic testing is now the standard of care for familial adenomatous polyposis (FAP). Despite the detailed genetic knowledge of FAP that is now available, genetic testing is often poorly interpreted. Consequently, genetic counseling is an integral part of management and should precede genetic testing. Testing for FAP in a family is most informative when it begins with the affected family member, to identify the mutation responsible for FAP within that family. Once a causal mutation has been identified in an affected person, predictive testing can be done to identify other family members at risk.[1] DNA testing for *APC* gene mutations has a sensitivity of 70% to 90% and a specificity of 100%.

FDA Approval Report

The following is selected from the FDA's list of recently approved products. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

New Drug for T Cell Leukemia and Lymphoma

The FDA has approved nelarabine for the treatment of adults and children with T cell acute lymphoblastic leukemia (T-ALL) and T cell lymphoblastic lymphoma (T-LBL) whose disease has not responded to at least two chemotherapy regimens or has relapsed after such treatment. Nelarabine is the first drug to treat this limited population of patients.

Nelarabine was approved under a program that allows accelerated approval of products to treat cancer or other serious or life-threatening diseases on the basis of early evidence of efficacy. In this case, this evidence consisted of complete remission in 23% of 39 pediatric patients, with remission lasting 3.3 to 9.3 weeks, and complete remission in 21% of 28 adult patients, with remission lasting from 4 to more than 195 weeks. In those patients who responded to nelarabine, the disappearance of cancer cells was sometimes accompanied by the return of normal blood cell counts. Further studies are required to verify nelarabine's clinical benefit. Nelarabine also received Orphan Drug Designation, which is granted to products that treat rare diseases (i.e., those that affect fewer than 200,000 people in the United States). Approximately 1,600 patients are newly diagnosed with T-ALL/T-LBL each year in the United States. Of those, an estimated 500 patients have relapsed or refractory T-ALL/T-LBL; approximately 200 of these patients are children. Common side effects reported with nelabarine treatment are fatigue, nausea, vomiting, and diarrhea.

Generic Name: Nelarabine

Brand Name: Arranon

Manufacturer: GlaxoSmithKline (GSK), Research Triangle Park, North Carolina

Source:
FDA Approves Arranon for Rare Leukemia and Lymphoma: Drug Approved under Agency's Orphan Drug and Accelerated Approval Programs. FDA News. U.S. Food and Drug Administration, October 31, 2005
(http://www.fda.gov/bbs/topics/NEWS/2005/NEW01251.html)

MYH is a base excision repair gene that is involved in repairing oxidative damage to DNA. Loss of *MYH* is associated with G:C to T:A transversions in APC, leading to polyp formation and carcinoma. The two most common *MYH* mutations, G382D and Y165C, account for approximately 85% of MYH-associated polyposis. Appropriate candidates for mutational analysis testing are patients with multiple adenomas or FAP who have a family history compatible with a recessive pattern of inheritance (e.g., colon cancers or multiple adenomas in only one generation or in skipped generations) and in whom testing has failed to show a germline *APC* mutation.

1. Colorectal cancer screening. National Comprehensive Cancer Network. J Natl Compr Canc Netw 1:72, 2003

Postmenopausal Hormone Therapy and Cancer Risk

Many epidemiologic studies have examined the possible associations between exogenous estrogens and colorectal neoplasia risk. In a meta-analysis of 18 epidemiologic studies, postmenopausal hormone therapy (HT) was associated with a 33% reduction in the risk of colon cancer in recent users; the relative risk was 0.67, compared with a relative risk of 0.92 in women who had used HT more than 1 year ago. Similarly, HT may protect against adenoma formation. In the Women's Health Initiative trial, which included over 16,000 postmenopausal women, the combination of estrogen and progesterone was associated with a reduction in the risk of colorectal cancer (hazard ratio, 0.56; 95% confidence interval, 0.38–0.81; $P = 0.003$).[1] Invasive colorectal cancers in the hormone-treated group were similar in histologic features and grade to those in the group receiving placebo. However, in patients who received HT, the number of positive lymph nodes was slightly greater, and the cancer was more advanced regionally or metastatically.

1. Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, et al: Estrogen plus progestin and colorectal cancer in postmenopausal women. Women's Health Initiative Investigators. N Engl J Med 350:991, 2004 [PMID 14999111]

Molecular Detection of Colon Cancer

Detection of gene mutations in the stool has been possible for over a decade. It is technically feasible to detect *APC* and *p53* mutations, long DNA, and K-*ras* mutations. In addition, right-sided lesions can be detected by the identification of *BAT-26* mutations. In a 2004 study of over 2,500 persons, a fecal DNA mutation panel detected 29 of 71 (41%) invasive cancers and adenomas with high-grade dysplasia. The specificity was 94%. This is a rapidly advancing field, and improvements in technology are likely to enhance the sensitivity of detection.[1] The cost of such techniques is high, however, and it remains to be seen whether their use will be cost-effective relative to other techniques, including newer immunochemical tests for fecal occult blood.

1. Imperiale TE, Ransohoff DF, Itzkowitz SH, et al: Fecal DNA versus fecal occult blood for colorectal cancer screening in an average-risk population. N Engl J Med 351:2704, 2004 [PMID 15616205]

Monoclonal Antibodies for Treating Colorectal Cancer

After decades in which fluorouracil (5-FU) was the only chemotherapeutic agent available, the advent of new and more effective agents has completely changed the management of colorectal cancer. Although 5-FU remains the backbone of most regimens, new chemotherapeutic agents (i.e., irinotecan, oxaliplatin, and capecitabine) and new biologic agents (i.e., the monoclonal antibodies cetuximab and bevacizumab) offer additional benefit.

Cetuximab is a chimeric monoclonal antibody that bonds to the external growth factor receptor, which is commonly expressed on colorectal carcinoma cells. Dysregulation of cell signaling through this receptor is thought to be an important factor in the growth of many epithelial malignancies.

Patients treated with combinations of cetuximab and irinotecan have shown a higher response rate and a longer time to worsening of the metastatic disease than patients treated with either drug alone.[1] Side effects of cetuximab include an acneiform rash, malaise, and diarrhea.

Bevacizumab is a human chimeric antibody to vascular endothelial growth factor that stimulates angiogenesis. New blood vessel formation is important in tumor growth and invasion.

The addition of bevacizumab to combination therapy with irinotecan, leucovorin, and 5-FU has been shown to increase therapeutic efficacy: patients receiving the four-drug combination survived for approximately 20 months—about 5 months longer than those receiving the three-drug combination.[2] Combinations of bevacizumab and cetuximab are also being assessed because of preclinical and early clinical data indicating synergistic activity.

1. Cunningham D, Humblet Y, Siena S, et al: Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351:337, 2004 [PMID 15269313]

2. Hurwitz H, Fehrenbacher L, Novotny W, et al: Bevacizumab plus irinotecan, fluorouracil and leucovorin for metastatic colorectal cancer. N Engl J Med 350:2335, 2004 [PMID 15175435]

4 GASTROENTEROLOGY

III Diarrheal Diseases

LAWRENCE R. SCHILLER, MD, FACP
Baylor University Medical Center

Antibiotic Selection by Geography for Shigellosis

Because of growing resistance to the fluoroquinolones in the United States, trimethoprim-sulfamethoxazole is the recommended initial treatment for most patients with shigellosis. Shigellosis contracted overseas is initially treated with fluoroquinolones, because those strains are more likely to be resistant to trimethoprim-sulfamethoxazole.

Confirming and Curing Giardiasis

Enzyme-linked immunosorbent assay (ELISA) for *Giardia* antigen is superior to microscopic inspection of stool (so-called ova and parasites testing) for the detection of giardiasis. Therapy with tinidazole, metronidazole, or nitazoxanide is effective in most patients, but reinfection can occur.[1]

1. Ali SA, Hill DR: *Giardia intestinalis*. Curr Opin Infect Dis 16:453, 2003 [PMID 14501998]

Drugs in, Diarrhea out

Drug therapy is a key cause of secretory diarrhea.[1] Many drugs have diarrhea as a side effect. These include antibiotics; cardiovascular agents, such as beta-adrenergic antagonists, digitalis, and quinidine; cancer chemotherapy; nonsteroidal anti-inflammatory drugs; and colchicine. Thus, in taking the history of a patient with chronic diarrhea, it is critical to formulate a detailed drug list, including over-the-counter and alternative medications. A special category of drug-induced secretory diarrhea is surreptitious ingestion of stimulant laxatives.

1. Cappell M: Colonic toxicity of administered drugs and chemicals. Am J Gastroenterol 99:1175, 2004 [PMID 15180742]

Neoplastic Diarrhea

Endocrine causes of secretory diarrhea include a group of rare tumors of the endocrine cells of the gut, including gastrinomas, carcinoid tumors, vasoactive intestinal peptide tumors (VIPomas), somatostatinomas, and medullary carcinoma of the thyroid.[1,2] These tumors produce peptides and other mediators that affect intestinal mucosal and muscle function and thereby produce diarrhea. In most cases, rapid intestinal transit seems to be the major mechanism producing diarrhea in these disorders, although this remains controversial.

Other tumors that produce secretory diarrhea include colon cancer (mechanism uncertain), villous adenoma of the rectum, lymphoma, and mastocytosis. Mastocytosis (and probably some lymphomas) produce diarrhea by release of histamine or other mediators that affect gut function. Infiltration of the mucosa also may play a role in some cases.

1. Warner RR: Enteroendocrine tumors other than carcinoid: a review of clinically significant advances. Gastroenterology 128:1668, 2005 [PMID 15887158]

2. Modlin IM, Kidd M, Latich I, et al: Current status of gastrointestinal carcinoids. Gastroenterology 128:1717, 2005 [PMID 15887161]

Diagnosis at a Gulp?

The role of capsule endoscopy in the evaluation of patients with chronic diarrhea is under investigation; studies suggest that it may be helpful in detecting Crohn disease and, perhaps, celiac disease.[1-3] However, capsule endoscopy does not allow for biopsy of abnormalities that are visualized during the procedure, which limits its utility. Double-balloon enteroscopy offers the possibility of visualizing and obtaining biopsies from the entire small intestine; this technique may find a place in the evaluation of patients with watery secretory diarrhea.[4]

1. Petroniene R, Dubcenco E, Baker JP, et al: Given capsule endoscopy in celiac disease: evaluation of diagnostic accuracy and interobserver agreement. Am J Gastroenterol 100:685, 2005 [PMID 15743369]

2. Kalantzis N, Papanikolaou IS, Giannakoulopoulou E, et al: Capsule endoscopy: the cumulative experience from its use in 193 patients with suspected small bowel disease. Hepatogastroenterology 52:414, 2005 [PMID 15816447]

3. Sturniolo GC, Di Leo V, Vettorato MG: Clinical relevance of small-bowel findings detected by wireless capsule endoscopy. Scand J Gastroenterol 40:725, 2005 [PMID 16036534]

4. Matsumoto T, Moriyama T, Esaki M, et al: Performance of antegrade double-balloon enteroscopy: comparison with push enteroscopy. Gastrointest Endosc 62:392, 2005 [PMID 16111958]

2 DERMATOLOGY

II Papulosquamous Disorders

ELIZABETH A. ABEL, MD
Stanford University School of Medicine

Light and Dark on Pityriasis Rosea

Pityriasis rosea lesions resolve spontaneously after 6 to 8 weeks. The patient should be reassured that the disorder is benign and self-limited; such reassurance, together with educating the patient about the disease, is the most important aspect of treatment. Lesions are variably pruritic. Symptoms should be treated with bland emollients or systemic antipruritics. Sun exposure may accelerate clearing. Irradiation with ultraviolet B (UVB) sunlamps is beneficial in decreasing the severity of disease, especially when treatment is initiated within the first week of the eruption. One study found that 10 erythemogenic exposures of UVB substantially decreased the extent of pityriasis rosea, although it neither altered the duration of the disorder nor improved the itching. Other evidence suggests that UVB therapy may hasten resolution of the rash but may cause hyperpigmentation.[1]

1. Stulberg DL, Wolfrey J: Pityriasis rosea. Am Fam Physician 69:87, 2004 [PMID 14727822]

What Is the Link between Hepatitis and Lichen Planus?

The prevalence of viral hepatitis, especially hepatitis C, is increased in patients with lichen planus. In a multicenter study of 303 sequential patients with lichen planus, the prevalence of hepatitis C virus (HCV) was 19.1%, compared with 3.2% in control subjects.[1] The role of HCV in the pathogenesis of lichen planus is not clearly understood; some investigators suggest that the cause of lichen planus may relate to the pattern of immune dysregulation induced by HCV.[2] There are a number of reports of lichen planus occurring after administration of different types of hepatitis B vaccine.[3] This is a rare occurrence, considering the widespread use of this vaccine; several cases have been reported from France and Italy, and one case has been reported from the Middle East. An immunologic mechanism has been postulated as the cause. The latency period ranges from several days to 3 months after any one of the three usual injections of vaccine.

1. Lodi G, Giuliani M, Majorana A, et al: Lichen planus and hepatitis C virus: a multicentre study of patients with oral lesions and a systematic review. Br J Dermatol 151:1172, 2004 [PMID 15606512]
2. Harden D, Skelton H, Smith KJ: Lichen planus associated with hepatitis C virus: no viral transcripts are found in the lichen planus, and effective therapy for hepatitis C virus does not clear lichen planus. J Am Acad Dermatol 49:847, 2003 [PMID 14576663]
3. Callista D, Morri M: Lichen planus induced by hepatitis B vaccination: a new case and a review of the literature. Int J Dermatol 43:562, 2004 [PMID 15304176]

Treating Lichen Planus in the Mouth

For lichen planus that is localized to the oral mucosa, a high-potency corticosteroid such as clobetasol in a vehicle that is adherent to the mucosal surface (Orabase) is helpful. Intralesional injections of corticosteroids may be used to treat localized, recalcitrant lesions. Use of miconazole gel in combination with chlorhexidine mouth rinses is effective for prophylaxis against oral candidiasis. Topical isotretinoin gel is an effective alternative to corticosteroids, although relapses often occur after discontinuance of this medication. In a double-blind, placebo-controlled study of 22 patients with biopsy-proven oral lichen planus, an 8-week course of 0.1% isotretinoin gel was found to be effective. Cyclosporine mouth rinses have been helpful for some patients. A 6-month course of hydroxychloroquine, 200 to 400 mg daily, was successful in nine of 10 patients with oral lichen planus; ulcers healed and pain decreased after 1 to 2 months.

Topical tacrolimus, a macrolide that suppresses T cell activation, may be effective in treating erosive mucosal lichen planus that is resistant to conventional treatment; however, relapse after cessation of therapy is common. Topical pimecrolimus cream is being evaluated as a treatment for oral erosive lichen planus.[1] The role of these agents in the treatment of lichen planus must be further investigated, particularly in view of the 2005 Food and Drug Administration alert about a possible link between use of topical tacrolimus and pimecrolimus and cases of lymphoma and skin cancer.

1. Swift JC, Rees TD, Plemons JM, et al: The effectiveness of 1% pimecrolimus cream in the treatment of oral erosive lichen planus. J Periodontol 76:627, 2005 [PMID 15857105]

What Works for Seborrheic Dermatitis

Seborrheic dermatitis on the scalp usually responds well to frequent—as often as daily—shampooing with a preparation containing 3% to 5% sulfur and 2% to 3% salicylic acid. Good response has also been reported with use of ciclopirox 1% shampoo twice weekly.[1] For the face and nonhairy areas, a mild cream containing precipitated 3% sulfur and 3% salicylic acid is effective. Involved areas also respond well to low-potency topical glucocorticoids, such as 1% hydrocortisone cream or desonide cream. Caution, however, must be exercised in the use of high-potency fluorinated steroid preparations, especially on the face and in skin folds; prolonged application may lead to chronic skin changes, such as atrophy and telangiectasia. Wet dressings followed by a topical antibiotic preparation are helpful in treating intertriginous areas, in which maceration and superficial secondary infection may occur.

Topical antifungal agents have been used in the treatment of seborrheic dermatitis. In addition to their antifungal properties, certain azoles (e.g., bifonazole, itraconazole, and ketoconazole) have demonstrated anti-inflammatory activity, which may be beneficial in alleviating symptoms.[2]

1. Lebwohl M, Plott T: Safety and efficacy of ciclopirox 1% shampoo for the treatment of seborrheic dermatitis of the scalp in the US population: results of a double-blind, vehicle-controlled trial. Int J Dermatol 43(suppl 1):17, 2004 [PMID 15271196]
2. Gupta AK, Nicol K, Batra R: Role of antifungal agents in the treatment of seborrheic dermatitis. Am J Clin Dermatol 5:417, 2004 [PMID 15663338]

8 INTERDISCIPLINARY MEDICINE

II Bites and Stings

LAWRENCE M. LEWIS, MD
WILLIAM H. DRIBBEN, MD
MARK D. LEVINE, MD
Washington University School of Medicine, St. Louis

Are Human-Bite Wounds Not So Bad?

Traditionally, human-bite wounds have had a reputation for frequent

and severe complications. Current data, however, suggest an infection rate from human-bite wounds on the order of 10% to 50%, depending on the wound type and location. Occlusional/simple bite wounds to areas other than the hand probably are no more at risk for infection than any other type of bite wound and minimally more than for nonbite lacerations.[1] However, human-bite wounds to the hand are associated with infection rates of almost 50%. A clenched-fist injury is considered the most serious of all human-bite wounds.[1] These injuries may appear innocent at first but progress to serious infections that may include the joint, tendons, or various compartments of the hand. These injuries require meticulous wound care, appropriate antibiotic therapy, and consultation with a hand surgeon.

1. Broder J, Jerrard D, Olshaker J, et al: Low risk of infection in selected human bites treated without antibiotics. Am J Emerg Med 22:10, 2004 [PMID 14724871]

Prophylactic Antibiotics for Bite Wounds

The use of prophylactic antibiotics for any bite wound is debatable, but there is general consensus that certain wounds in all patients and most wounds in certain patients deserve prophylactic antibiotics [*see Table, upper right*]. The antibiotic of choice for prophylaxis of most mammalian bite wounds is amoxicillin-clavulanate. For the penicillin-allergic patient, a third-generation fluoroquinolone (e.g., moxifloxacin)[1] or a cephalosporin (e.g., cefotaxime)[2] serves as a good alternative. The timing of prophylactic antibiotics is important. Prophylactic antibiotics should be given as soon after the bite injury as possible. A systematic review suggests that prophylactic antibiotics may reduce the incidence of infection in all hand-bite wounds and human bites, regardless of location.[1]

1. Talan DA, Abrahamian FM, Moran GJ, et al: Clinical presentation and bacteriologic analysis of infected human bites in patients presenting to emergency departments. Clin Infect Dis 37:1481, 2003 [PMID 14614671]

2. Turner TW: Evidence-based emergency medicine/systematic review abstract. Do mammalian bites require antibiotic prophylaxis? Ann Emerg Med 44:274, 2004 [PMID 15332071]

Bite Wounds Requiring Prophylactic Antibiotics

Wound characteristics	Puncture wounds Full-thickness wounds Hand or foot wounds Wounds requiring surgical repair Treatment delay (> 24 hr) Human bites* Cat bites*
Patient characteristics	Age > 50 yr Immunosuppression (e.g., asplenia, alcoholism, corticosteroid use) Diabetes mellitus Peripheral vascular disease

*There is debate, but many authors recommend prophylactic antibiotic treatment for virtually all human and cat bites because of the high rate of infection.

Brown Recluse Bites That Aren't

The brown recluse spider is found in the south central United States, especially in Missouri, Kansas, Arkansas, Louisiana, eastern Texas, and Oklahoma. It is occasionally found in other states, but these cases likely represent spiders uprooted and transported from the endemic areas.[1] The diagnosis of true *Loxosceles* spider bites in the United States is considered to be far less than the number of patients who are treated for recluse spider envenomations. Strict inclusion criteria for recluse envenomation includes sighting of the biting spider and identification of that specific spider. In studies from areas endemic for *Loxosceles*, patients present with almost no true *Loxosceles* bites or envenomations despite exposure to dozens to hundreds of spiders; these results suggest that nonendemic areas should have minimal to no true *Loxosceles* bites.[2,3]

1. Swanson DL, Vetter RS: Bites of brown recluse spiders and suspected necrotic arachnidism. N Engl J Med 352:700, 2005 [PMID 15716564]

2. Isbister GK, White J: Clinical consequences of spider bites: recent advances in our understanding. Toxicon 43:477, 2004 [PMID 15066408]

3. Vetter RS, Cushing PE, Crawford RL, et al: Diagnoses of brown recluse spider bites (loxoscelism) greatly outnumber actual verifications of the spider in four western American states. Toxicon 42:413, 2003 [PMID 14505942]

CLINICAL ESSENTIALS

XI Management of Psychosocial Issues in Terminal Illness

JENNIFER RHODES-KROPF, MD
Mount Sinai School of Medicine
NED H. CASSEM, MD, FACP
Harvard Medical School

Cultural and Individual Attitudes toward Truth Telling

The desire for truth telling may vary among different ethnic groups.[1] In a study of elderly persons in the United States, Korean Americans and Mexican Americans were less likely than African Americans and European Americans to believe that a patient should be told the diagnosis of metastatic cancer. However, a population study in Hong Kong reported that the majority of persons canvassed would want to know if they had terminal cancer; this finding is at odds with the cultural preference of many Chinese, who usually prefer to withhold diagnostic information from terminally ill family members.[2] This study emphasizes that cultural preferences give only general indications of a patient's readiness to hear bad news; truth telling ultimately depends on the physician's assessment of what the patient wants to know and is prepared to know about the diagnosis.

1. Searight HR, Gafford J: Cultural diversity at the end of life: issues and guidelines for family physicians. Am Fam Physician 71:515, 2005 [PMID 15712625]

2. Tse CY, Chong A, Fok SY: Breaking bad news: a Chinese perspective. Palliat Med 17:339, 2003 [PMID 12822851]

Talk about Religion

Studies find that people who have a strong internalized faith possess a resource that helps significantly in coping with a fatal illness.[1,2] It is a well-documented finding that religious persons usually belong to a community that can be unusually thoughtful and generous in providing support.

Coming in February

4 Gastroenterology
V Diseases of the Pancreas

7 Infectious Disease
XXVIII Enteric Viral Infections

8 Interdisciplinary Medicine
I Management of Poisoning and Drug Overdose

11 Neurology
VII Anoxic, Metabolic, and Toxic Encephalopathies

13 Psychiatry
II Depression and Bipolar Disorder

16 Women's Health
Introduction to Women's Health: The Primary Care of Women

However, the community may not know of the patient's plight and may need to be contacted. Thus, the appreciation of a person's religion or spirituality is extremely important. Studies suggest that patients welcome inquiries about their spiritual well-being from the physicians, although interventions to address spiritual distress have not been well developed or well evaluated.[3]

1. Koenig HG, George LK, Titus P: Religion, spirituality, and health in medically ill hospitalized older patients. J Am Geriatr Soc 52:554, 2004 [PMID 15066070]

2. McClain CS, Rosenfeld B, Breibart W: Effect of spiritual well-being on end-of-life despair in terminally-ill patients. Lancet 361:1603, 2003 [PMID 12747880]

3. Morrison RS, Meier DE: Palliative care. N Engl J Med 350:2582, 2004 [PMID 15201415]

Truth and Silence

Even when the prognosis is grave, a calm statement of the treatments planned to counteract and contain the disorder is of value to the anxious patient. The more ominous the prognosis, the more important it is to encourage the patient to specify the fear, so that correspondingly true reassurances (e.g., "the medication can control pain") can be given. False comfort is not recommended. It robs the physician of credibility and, therefore, of the ability to reassure the patient as the illness progresses. An empathetic yet silent presence can sometimes be more helpful than well-meant counsel.[1]

1. Penson RT, Partridge RA, Shah MA, et al: Fear of death. Oncologist 10:160, 2005 [PMID 15709218]

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 28 No. 12 December 2005

What's New in ACP Medicine

www.acpmedicine.com

PRACTICE OF MEDICINE

The ACP and *ACP Medicine*

JOHN TOOKER, MD, MBA, FACP
Executive Vice President and Chief Executive Officer
American College of Physicians
STEVEN WEINBERGER, MD, FACP
Senior Vice President, Medical Knowledge and Education
American College of Physicians

More than at any time in the past, physicians are under great pressure to meet new educational and professional requirements. They are being asked to learn more and do more at the cost of time to spend on their patients, as well as on their personal lives and other interests. Yet, despite these demands, physicians everywhere are continuing to step forward to accept the additional burdens placed upon them.

The American College of Physicians (www.acponline.org) takes special pride in its role as an advocate for the practicing physician and as a contributor of increased medical education requirements that are fair and reasonable benchmarks of a physician's skills. One of the College's educational efforts is its sponsorship of *ACP Medicine*, which is published by WebMD in partnership with the College. The excellent 2006 bound edition of *ACP Medicine* is the latest manifestation of our joint efforts.

The partnership between the College and WebMD was the culmination of an extensive review of *ACP Medicine's* content and editorial processes by the College, as well as a

continued on page 2

THIS MONTH'S UPDATES

3 ENDOCRINOLOGY

X Obesity

JONATHAN Q. PURNELL, MD
Oregon Health and Science University

Laboratory Tests for Obese Patients

Screening of obese patients for macrovascular risk involves obtaining an electrocardiogram when appropriate. A fasting lipid profile should be obtained to complete the cardiovascular risk assessment, and, if necessary, treatment should be instituted according to guidelines from the National Cholesterol Education Program Expert Panel. This panel incorporated several nonlipid risk factors for cardiovascular disease into its recommendations for clinical care by defining criteria for a condition that has become known as the metabolic syndrome (also called syndrome X or the insulin-resistance syndrome). The metabolic syndrome includes the most common abnormalities of lipid and glucose metabolism that accompany abdominal obesity [*see Table, page 3, top*]. Identifying these abnormalities in a patient allows the practitioner to better assign that patient's risk for diabetes and coronary artery disease.[1] Laboratory screening tests for hepatosteatosis include a liver panel. In addition, for all overweight patients, normal thyroid function should be documented by measurement of the thyroid-stimulating hormone level.

1. Eckel RH, Grundy SM, Zimmet PZ: The metabolic syndrome. Lancet 365:1415, 2005 [PMID 15836891]

Diabetes Treatment and Weight Gain

Patients with type 1 or type 2 diabetes often gain weight after starting therapy. This weight gain is proportional to the degree of improved glycemic control; it results from a reduction in glucosuria and improvement in metabolic efficiency. Long-term studies have shown that intensive insulin treatment of type 1 diabetes can result in excessive weight gain and obesity in up to 25% of patients; in type 2 diabetes, intensive glycemic control with insulin, a sulfonylurea, or one of the thiazolidinediones may also result in greater weight gain than predicted by

continued on page 2

In This Issue

PRACTICE OF MEDICINE

continued from page 1

shared and long-standing commitment to helping physicians achieve excellence in clinical practice. David C. Dale, M.D., F.A.C.P., the Editor-in-Chief of *ACP Medicine*, reports to the College regarding content development and editorial standards related to the publication. A strong connection has long existed between the College and WebMD, with many of *ACP Medicine's* editors and authors having also been active in the College's educational programs and publications.

One of the many benefits of the College's sponsorship of *ACP Medicine* is that the text includes coverage of evolving programs and professional requirements, such as performance measurement and other quality-improvement programs. These requirements will obviously have great implications for how physicians practice medicine on a daily basis. Medicare already has started a series of pilot programs offering financial incentives on the basis of improved outcomes. Because of the importance of such performance measures, the front matter of the 2006 edition of *ACP Medicine* contains special coverage of this information by Kevin B. Weiss, M.D., F.A.C.P., and colleagues. Dr. Weiss is chair of the ACP Performance Measures Subcommittee, which has created a "starter set" of performance measures. In addressing these topics, Dr. Weiss provides an important overview of the changes to come in performance measures. Updated information on this topic and numerous others is available at the *ACP Medicine* Web site (www.acpmedicine.com).

Clearly, the need for medical education continues after completion of residency—it is a lifelong commitment. The fact that the educational requirements now include understanding and following new professional obligations, as well as remaining current on the latest diagnostic and therapeutic practices, makes the need for access to a textbook such as *ACP Medicine* all the more significant.

Physicians share a professional commitment to excellence in clinical practice through the use of evidence-based practice guidelines in the care of their patients. We hope that *ACP Medicine* provides an expedient and authoritative resource for physicians to achieve this goal of translating knowledge into care.

whatsnew@webmd.net

THIS MONTH'S UPDATES

continued from page 1

improved glycemic control alone. Therapy with metformin plus nighttime long-acting insulin may reduce or prevent this extra weight gain, and newer diabetes medications, such as pramlintide and exenatide, can improve glycemic control in both type 1 and type 2 diabetes with a modest weight loss.

Reasonable Expectations for Obesity Treatment

The weight loss goal for the treatment of obesity is a sustained loss of 5% or more of initial body weight. Although this goal does not result in attainment of a normal body weight (i.e., a body mass index [BMI] of 19 to 25) in the majority of patients, it still represents a weight loss that can be achieved with available intervention modalities and that has been associated with lower morbidity, including reductions in the risk of diabetes and heart disease.

For some patients who are experiencing a period of weight gain, weight stability may be their primary goal. This is especially common in patients

ACP MEDICINE

A Publication of the American College of Physicians

Published by WebMD Inc.

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $329 for the first year ($219 for residents and students) and $289 for renewals ($219 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $429 for the first year and $349 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, WebMD Inc., P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, WebMD Inc., P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

THIS MONTH'S UPDATES

continued from page 2

who have just completed a low-calorie weight loss program and are struggling to remain below their initial weight.

Obesity under the Knife

A National Institutes of Health expert panel has suggested that patients with a BMI of 40 or more or a BMI of 35 or more who also have obesity-related risk factors could be considered for surgical therapy. In general, these procedures can be classified into three types: restriction of food passage, malabsorption of nutrients, or a combination of the two [*see Table, below*]. As an example of a strictly restrictive procedure, vertical-banded gastroplasty involves the formation of a small stoma for the passage of food. The rationale behind this procedure is to increase a sense of fullness and reduce food intake.

Procedures resulting in malabsorption typically involve bypassing sections of small or large intestine. Early procedures in which large sections, or even the entire length, of the small intestine was bypassed often resulted in severe malabsorption and sometimes hepatic failure and death. Subsequent procedures involved bypass of shorter sections, which lowered morbidity. Patients must understand that bypass surgery is anatomically irreversible in most cases and has a potentially high postoperative complication rate.

A more recent technique, known as laparoscopic gastric banding, in which a restrictive band is placed around the upper stomach, has been shown to lead to sustained weight loss in studies of up to 4 years. However, this weight loss is generally felt to be similar to that of vertical-banded gastroplasty and is not as great as that obtained with the gastric bypass operation.[1]

1. Chapman AE, Kiroff G, Game P, et al: Laparoscopic adjustable gastric banding in the treatment of obesity: a systematic literature review. Surgery 135:326, 2004 [PMID 14976485]

Criteria for Metabolic Syndrome

Any Three of the Following:

- Increased waist circumference
  - Men: > 102 cm (40 in)
  - Women: > 88 cm (35 in)
- Fasting plasma glucose ≥ 100 mg/dl
- Elevated blood pressure
  - Systolic ≥ 130 mm Hg
  - Diastolic ≥ 85 mm Hg
- Serum triglyceride level ≥ 150 mg/dl
- Decreased high-density lipoprotein (HDL) cholesterol level
  - Men < 40 mg/dl
  - Women < 50 mg/dl

Most Commonly Used Bariatric Procedures

Type of Surgery	*Procedure*	*Average Weight Loss*	*Medical Complications*	*Management*
Restrictive	Vertical-banded gastroplasty	~17% (5 yr)	Nausea, vomiting Gastric distention	Reversal of procedure if complications unacceptable
	Laparoscopic adjustable gastric banding	17%–21% (3 yr)	Nausea, vomiting Slippage Erosion and leakage	Adjustment of band to minimize nausea; reversal of procedure if necessary
Malabsorptive	Biliopancreatic diversion	~27% (5 yr)	Diarrhea, fatty stools Protein-calorie malnutrition Anemia (low iron) Deficiency of vitamin D (and other fat-soluble vitamins) Hypocalcemia Hyperparathyroidism Metabolic bone disease Vitamin B_{12} deficiency, Wernicke encephalopathy (rare)	Monitoring of levels and replacement of nutritional deficiencies; patients often require high oral or parenteral doses
Combination	Gastric bypass	~27% (5 yr)	Dumping syndrome Reductions in levels of iron, vitamin B_{12}, folate, calcium, vitamin D Nesidioblastosis (rare)	Small, frequent meals; monitoring of nutrient levels and oral supplementation with iron tablets (325 mg), vitamin B_{12} (500 µg), folate (1 mg), calcium (500–1,000 mg), and vitamin D

FDA Approval Report

The following is selected from the FDA's list of announcements. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

Strengthened Risk Management Program for Isotretinoin

The FDA has approved a strengthened distribution program for isotretinoin. This program, called iPLEDGE, is intended to prevent use of isotretinoin during pregnancy. Isotretinoin (Accutane and its generics) is a highly effective drug for severe recalcitrant nodular acne, but it carries a significant risk of birth defects. Women who are pregnant or who might become pregnant should not take the drug.

Starting December 31, 2005, physicians and patients must register with the iPLEDGE program before receiving authorization to prescribe or use the drug. Compliance with iPLEDGE requires office visits, counseling, birth control, and other responsibilities. Key requirements for patients include completing an informed consent form, obtaining counseling about the risks and requirements for safe use of the drug, and, for women of childbearing age, complying with required pregnancy testing.

As of October 31, 2005, wholesalers and pharmacies have had to register with iPLEDGE to obtain isotretinoin from a manufacturer, to distribute the product, and to dispense it.

Physicians, patients, and pharmacies can obtain program information and can register with iPLEDGE via the Internet (www.ipledgeprogram.com) or by telephone (1-866-495-0654).

A reporting and collection system for serious adverse events associated with the use of isotretinoin has also been implemented. All pregnancy exposures to isotretinoin must be reported immediately to the FDA via MedWatch (1-800-FDA-1088); pregnancy exposures must also be reported to the iPLEDGE pregnancy registry, either by telephone (1-866-495-0654) or through the iPLEDGE Web site (www.ipledgeprogram.com).

In addition to approving the iPLEDGE program, the FDA has approved changes to the existing warnings, patient information, and informed consent document so that patients and prescribers can better identify and manage the risks of psychiatric symptoms and depression before and after prescribing isotretinoin.

Source:

FDA Announces Strengthened Risk Management Program to Enhance Safe Use of Isotretinoin (Accutane) for Treating Severe Acne. FDA News. U.S. Food and Drug Administration, August 12, 2004 (http://www.fda.gov/bbs/topics/NEWS/2005/NEW01218.html)

7 INFECTIOUS DISEASE

VIII Infections Due to Escherichia coli and Other Enteric Gram-Negative Bacilli

MICHAEL S. DONNENBERG, MD

University of Maryland School of Medicine

Treating Traveler's Diarrhea

As with all diarrheal disease, the initial management of enterotoxigenic *Escherichia coli* (ETEC) infection involves ensuring adequate fluid repletion. Several well-conducted clinical trials have demonstrated that traveler's diarrhea that is partly caused by ETEC responds rapidly to treatment with any of several regimens containing antibiotics, with or without antimotility agents. Ciprofloxacin, taken every 12 hours for 3 days in combination with loperamide, is a particularly effective regimen. Rifaximin, a nonabsorbable agent, is approved by the Food and Drug Administration for traveler's diarrhea caused by noninvasive strains of *E. coli*. Studies have found rifaximin to be comparable in efficacy to ciprofloxacin. Azithromycin is another effective alternative.[1] Rifaximin can be used as a single agent or in combination with loperamide. ETEC organisms are generally susceptible to fluoroquinolones; however, should resistance to fluoroquinolones increase, rifaximin may prove an important alternative to ciprofloxacin for the treatment of traveler's diarrhea.

1. Adachi JA, Ericsson CD, Jiang ZD, et al: Azithromycin found to be comparable to levofloxacin for the treatment of travelers with acute diarrhea acquired in Mexico. Clin Infect Dis 37:1165, 2003 [PMID 14557959]

The Wide Reach of *E. coli* O157:H7

Of the many Shiga toxin–producing *E. coli* strains that have been described in the literature, enterohemorrhagic *E. coli* (EHEC) of serotype O157:H7 is the most important, having caused both the largest number of outbreaks and the outbreaks involving the greatest number of patients. There are approximately 0.9 cases of EHEC O157:H7 infections per 100,000 persons annually in the United States.[1] The reservoir for EHEC is infected cattle, but the organism can be found in a variety of other ruminants. The disease often appears in outbreaks associated with the consumption of contaminated food. Undercooked ground beef has been associated with many outbreaks and is the leading risk factor for sporadic cases, but a variety of other foods and beverages, including other beef products, lettuce, sprouts, fruit, fruit juices, and milk, have been implicated. The disease has a low inoculum and can spread from person to person, particularly in day care centers or within the families of young children. An outbreak of EHEC infection resulting from airborne dispersal of bacteria after a country fair emphasizes the highly infectious nature of this pathogen.[2] It can also be spread through contamination of the water supply or through contamination of swimming pools, lakes, or

water parks, and it can be contracted directly from infected animals at farms and petting zoos.

1. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—10 sites, United States, 2004. MMWR Morb Mortal Wkly Rep 54:352, 2005 [PMID 15829864]

2. Varma JK, Greene KD, Reller ME, et al: An outbreak of *Escherichia coli* 0157 infection following exposure to a contaminated building. JAMA 290:2709, 2003 [PMID 14645313]

7 INFECTIOUS DISEASE

XIV Antimicrobial Therapy

TIMOTHY H. DELLIT, MD

University of Washington School of Medicine

HARVEY B. SIMON, MD, FACP

Harvard Medical School and Massachusetts Institute of Technology

THOMAS M. HOOTON, MD, FACP

University of Washington School of Medicine

Polymyxins Return

Polymyxins are cationic polypeptides that disrupt the bacterial cell membrane through a detergentlike mechanism. With the development of less toxic agents, such as extended-spectrum penicillins and cephalosporins, parenteral polymyxin use was largely abandoned, except for the treatment of multidrug-resistant pulmonary infections in patients with cystic fibrosis. More recently, however, the emergence of multidrug-resistant gram-negative bacteria, such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii*, and the lack of new antimicrobial agents have led to the revived use of the polymyxins, particularly colistin (polymyxin E) and polymyxin B.[1] Colistin is bactericidal against gram-negative bacilli, including strains of *Acinetobacter, P. aeruginosa, Klebsiella, Enterobacter, E. coli, Citrobacter, Morganella, Haemophilus influenzae*, and some strains of *Stenotrophomonas maltophilia*. It is not active against *P. mallei, Bacteroides cepacia, Proteus, Providencia*, or *Serratia*. Colistin sulfate is available as an oral formulation for the treatment of bowel decontamination; colistimethate sodium can be administered intravenously, intramuscularly, or by nebulization. The recommended intravenous dose is 2.5 to 5 mg/kg/day divided into two to four equal doses; however, dosing must be reduced if renal insufficiency is present.

Polymyxin B has been used extensively in topical otic and ophthalmic solutions, but there is more limited parenteral experience with it than with colistin. Originally, colistimethate sodium was thought to be less toxic than polymyxin B; however, if the drugs are administered at comparable doses, their toxicities may be similar.

Given the concern for potential toxicity, the use of colistin and polymyxin B should be viewed as a treatment of last resort for patients who have serious infections caused by multidrug-resistant gram-negative pathogens for which no other therapeutic options exist.

1. Falagas ME, Kasiakou SK: Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. Clin Infect Dis 40:1333, 2005 [PMID 15825037]

The First of a New Family

Tigecycline received approval by the Food and Drug Administration in 2005 as the first member of the glycylcycline family. Tigecycline is a semisynthetic derivative of minocycline. Its mechanism of action is similar to that of minocycline[1]; however, tigecycline avoids the two major forms of tetracycline resistance—namely, ribosomal protection and drug efflux. Tigecycline has a broad spectrum of bacteriostatic activity against gram-positive, gram-negative, atypical, and anaerobic bacteria. It is active against multidrug-resistant bacteria, such as penicillin-resistant *Streptococcus pneumoniae*, methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococcus* (VRE), *E. coli* containing extended-spectrum β-lactamases, and *Klebsiella pneumoniae*. It is also active against some strains of *Acinetobacter*; however, it is not active against *Pseudomonas* or *Proteus*. Tigecycline is currently indicated in the treatment of complicated intra-abdomial infections and complicated skin and soft-tissue infections.

Tigecycline is available only as an intravenous formulation. Excretion is primarily through the biliary system. Side effects are similar to those of the tetracyclines, with gastrointestinal disturbances such as nausea and emesis being the most common.

1. Rubinstein E, Vaughan D: Tigecycline: a novel glycylcycline. Drugs 65:1317, 2005 [PMID 15977966]

Using Linezolid Wisely

Linezolid has been used successfully in the treatment of multidrug-resistant gram-positive bacterial infections, including VRE and MRSA,[1] but clinical experience with deep-seated infections such as endocarditis and osteomyelitis is limited. Although resistance is uncommon, it can develop during therapy. As a result, it may be wise to reserve this unique antibiotic for serious infections caused by MRSA, VRE, or coagulase-negative staphylococci that do not respond to vancomycin. It also has a role in decreasing hospital stay for some patients with resistant gram-positive infections,[2] although these patients require close outpatient monitoring.

1. Eliopoulos GM: Quinupristin-dalfopristin and linezolid: evidence and opinion. Clin Infect Dis 36:473, 2003 [PMID 12567306]

2. Li JZ, Willke RJ, Rittenhouse BE, et al: Effect of linezolid versus vancomycin on length of hospital stay in patients with complicated skin and soft tissue infections caused by known or suspected methicillin-resistant staphylococci: results from a randomized clinical trial. Surg Infect 4:57, 2003 [PMID 12744768]

7 INFECTIOUS DISEASE

XXXVIII Mycotic Infections in the Compromised Host

JO-ANNE VAN BURIK, MD, FACP

University of Minnesota Medical School

New Antifungals

In the first years of the 21st century, several new antifungal agents have gained important roles in the treatment of candidiasis and other fungal infections. Voriconazole, a new triazole with wide-spectrum antifungal activity and high bioavailability, has proved effective as empirical antifungal therapy in patients with neutropenia and persistent fever. Caspofungin is the first in a new class of antifungal agents called echinocandins, which inhibit synthesis of an integral component of the fungal cell wall, β-(1,3)-D-glucan. The Food and Drug Administration has approved caspofungin for the first-line treatment of candidemia, esophageal candidiasis, and other *Candida* infections (e.g., intra-abdominal abscesses, peritonitis, and pleural space infections).[1] A comparison study found caspofungin to be at least as

SPECIAL – ONLINE ONLY!

New . . .

16 WOMEN'S HEALTH

IX Medical Problems in Pregnancy

ELLEN W. SEELY, MD
JEFFREY ECKER, MD
Harvard Medical School

Postpartum Thyroiditis or Postpartum Depression?

About 2 months after delivery, women with postpartum thyroiditis may experience hyperthyroidism as a result of leakage of thyroid hormone from the inflamed thyroid gland. This phase is often missed because of its short duration (usually 4 weeks or less) and because the patient and her caregivers tend to ascribe its symptoms of anxiety and palpitations to the stress of being a new mother. The thyroid-stimulating hormone (TSH) level is low at this time. About 6 months after delivery, the hypothyroid phase develops. At this time, patients usually present with fatigue and weight gain or an inability to lose pregnancy-associated weight. Depression can also be a presenting symptom; for that reason, it is important to check the TSH level before making a diagnosis of postpartum depression.

The Expectant Asthmatic

As many as 6% of pregnant women have asthma. This disease has a variable course during pregnancy: it is equally likely to worsen, improve, or remain unchanged. In contrast to past reports, current studies indicate that asthma is not associated with an increased risk of preterm delivery and growth restriction.[1] The clinical course in an individual patient may be linked, at least in part, to the patient's and her clinician's willingness to use needed medications during pregnancy. Guidelines from the National Institute of Child Health and Human Development emphasize that continued use of beta agonists and steroids during pregnancy is safe and appropriate.[2] As in nonpregnant women, clinical evaluations such as peak flow measurements can be used to judge disease activity and guide therapy.

1. Dombrowski MP, Schatz M, Wise R, et al: Asthma during pregnancy. Obstet Gynecol 103:5, 2004 [PMID 14704237]
2. Managing Asthma During Pregnancy: Recommendations for Pharmacologic Treatment—Update 2004. NIH Publication Number 05-3279. National Institutes of Health, Washington DC, 2005

Proven and Unproven Treatments of Gestational Diabetes

If a patient with gestational diabetes mellitus (GDM) continues to have elevated glucose levels despite dietary intervention, insulin therapy should be initiated. Although this recommendation has been made routinely for some time, evidence supporting the benefit of treatment for pregnancy outcome was lacking until 2005, when a study demonstrated that perinatal complications were less frequent in newborns of women with GDM who were randomized to management with home glucose monitoring, dietary therapy, and insulin, compared with the newborns of women who received routine care.[1] A 2004 study suggested that GDM may be effectively managed with oral glyburide rather than insulin,[2] but further studies are needed before this approach is advocated.

1. Crowther C, Hiller JE, Moss JR, et al: Effect of treatment of gestational diabetes on pregnancy outcomes. N Engl J Med 352:2477, 2005 [PMID 15951574]
2. Langer O, Conway DL, Berkus MD, et al: A comparison of glyburide and insulin in women with gestational diabetes mellitus. N Engl J Med 343:1134, 2004 [PMID 11036118]

effective as amphotericin B for the treatment of invasive candidiasis and, more specifically, candidemia. Caspofungin is also approved by the FDA for the empirical treatment of presumed fungal infection in febrile neutropenic patients; it appears to be as effective as liposomal amphotericin B when used for this purpose, and it is generally better tolerated than amphotericin.[2] Micafungin, the second echinocandin to become clinically available, is approved by the FDA for intravenous treatment of esophageal candidiasis and for *Candida* prophylaxis in hematopoietic stem cell transplant recipients.[3,4]

1. Kartsonis NA, Saah A, Lipka CJ, et al: Second-line therapy with caspofungin for mucosal or invasive candidiasis: results from the caspofungin compassionate-use study. J Antimicrob Chemother 53:878, 2004 [PMID 15044431]
2. Walsh TJ, Teppler H, Donowitz GR, et al: Caspofungin versus liposomal amphotericin B for empirical antifungal therapy in patients with persistent fever and neutropenia. N Engl J Med 351:1391, 2004 [PMID 15459300]
3. Micafungin (Mycamine) for fungal infections. Med Lett Drugs Ther 47:51, 2005 [PMID 15961968]
4. van Burik JA, Ratanatharathorn V, Stepan DE, et al: Micafungin versus fluconazole for prophylaxis against invasive fungal infections during neutropenia in patients undergoing hematopoietic stem cell transplantation. National Institute of Allergy and Infectious Diseases Mycoses Study Group. Clin Infect Dis 39:1407, 2004 [PMID 15546073]

Early Diagnosis of Aspergillosis

An enzyme-linked immunosorbent assay for the detection of *Aspergillus* galactomannan antigen in serum is commercially available and has become an important tool for the early diagnosis of invasive aspergillosis.[1] In one study, the test had a sensitivity of 75% and a specificity of 100%.[2] The assay is increasingly being used on specimens of body fluids other than serum, including urine, bronchoalveolar lavage fluid, and cerebrospinal fluid.[3]

1. Mennink-Kersten MA, Donnelly JP, Verweij PE: Detection of circulating galactomannan for the diagnosis and management of invasive aspergillosis. Lancet Infect Dis 4:349, 2004 [PMID 15172343]
2. Rovira M, Jimenez M, De La Bellacasa JP, et al: Detection of *Aspergillus* galactomannan by enzyme immunoabsorbent assay in recipients of allogeneic hematopoietic stem cell transplantation: a prospective study. Transplantation 77:1260, 2004 [PMID 15114095]
3. Klont RR, Mennink-Kersten MA, Verweij PE: Utility of *Aspergillus* antigen detection in specimens other than serum specimens. Clin Infect Dis 39:1467, 2004 [PMID 15546083]

Treating Zygomycosis

Zygomycosis is currently treated with maximal-dosage intravenous amphotericin, 1.0 to 1.5 mg/kg daily for 2 to 12 weeks or for the duration of clinical disease. In addition to intravenous infusions, intracavitary or topical amphotericin may be required for central nervous system disease; topical amphotericin can be applied to cutaneous disease. Posaconazole, an orally administered triazole, is effective against zygomycosis[1] but is not yet approved by the FDA.

Treatment should include adjunctive measures such as debridement of any adherent mycelial masses at accessible sites of infection. Debridement should be repeated as often as every other day until cultures of the debrided tissues are negative. Some patients may be candidates for granulocyte transfusions or hyperbaric oxygen.

1. Keating GM: Posaconazole. Drugs 65:1553, 2005 [PMID 16033292]

5 HEMATOLOGY

XIII Platelet and Vascular Disorders

LAWRENCE L. K. LEUNG, MD
Stanford University School of Medicine

Helicobacter pylori and ITP

The role of *Helicobacter pylori* eradication in the management of idiopathic thrombocytopenic purpura (ITP) is controversial. Eradication of *H. pylori* may have a limited value in improving the thrombocytopenia in young patients who have evidence of *H. pylori* infection and have relatively mild thrombocytopenia of short duration (i.e., platelet counts of 30,000 to 70,000/μl for less than 2 years). Eradication is not useful in patients with chronic, severe ITP.[1]

1. Stasi R, Rossi Z, Stipa E, et al: *Helicobacter pylori* eradication in the management of patients with idiopathic thrombocytopenic purpura. Am J Med 118:414, 2005 [PMID 15808140]

Monoclonal Antibodies for Refractory ITP

Rituximab, a chimeric anti-CD20 monoclonal antibody, administered intravenously at 375 mg/m^2 once weekly for 4 weeks, produces a lasting and substantial response in approximately one third of patients with chronic refractory ITP, although long-term follow-up is still limited.[1] The majority of the responses occur within 8 weeks of the first infusion. The therapy is generally well tolerated, with most of the side effects (i.e., fever, chills, mild hypotension, and bronchospasm) being infusion related and occurring with the first infusion. Rituximab produces a profound and prolonged peripheral B-cell depletion in all patients, which can last for more than a year, but serious infection is rare. Because of the concern of long-term bone marrow toxicity with older treatment regimens (i.e., azathioprine and cyclophosphamide), rituximab should be considered the first-line therapy in refractory ITP patients, if treatment is indicated.

1. Cooper N, Stasi R, Cunningham-Rundles S, et al: The efficacy and safety of B-cell depletion with anti-CD20 monoclonal antibody in adults with chronic immune thrombocytopenic purpura. Br J Haematol 125:232, 2004 [PMID 15059147]

Thrombocytopenia from Antiplatelet Drugs

Three parenteral platelet glycoprotein (GP) IIb-IIIa receptor antagonists—abciximab, eptifibatide, and tirofiban—have been approved for use in the treatment of acute coronary syndrome and as adjunctive therapy in coronary angioplasty. Treatment with GPIIb-IIIa antagonists can result in acute, often profound thrombocytopenia within a few hours of drug administration. This complication occurs in about 1% of patients with the first exposure to abciximab. After a second exposure, the incidence of thrombocytopenia rises to 4%. The incidence of drug-induced thrombocytopenia associated with eptifibatide and tirofiban is probably in a similar 1% range.[1] Consequently, a platelet count should be obtained in all patients before and within 2 to 4 hours after the initiation of an intravenous GPIIb-IIIa antagonist.

Because eptifibatide and tirofiban have very short half-lives and are cleared from the circulation within hours, thrombocytopenia resolves quickly once these drugs are discontinued. Abciximab, however, has a much longer half-life, with inhibition of platelet function reported up to 1 week after drug discontinuance; thrombocytopenia from this drug can persist for 5 to 7 days. It should be noted that in some patients, abciximab causes a delayed thrombocytopenia that develops 5 to 8 days after treatment. Limited published experience suggests that it is safe to administer eptifibatide or tirofiban to patients who are sensitive to abciximab, and vice versa.[1]

1. Aster RH: Immune thrombocytopenia caused by glycoprotein IIb/IIIa inhibitors. Chest 127:53S, 2005 [PMID 15706031]

5 HEMATOLOGY

XV Coagulation Disorders

LAWRENCE L. K. LEUNG, MD
Stanford University School of Medicine

Factor XI Deficiency, Surgery, and Childbirth

Patients with factor XI deficiency frequently come to medical attention when a prolonged activated partial thromboplastin time is detected in preoperative screening. It is most frequently observed in Ashkenazi Jews, although sporadic cases have been described in people of different ethnic origins. Factor XI deficiency is inherited as an autosomal recessive trait, and heterozygous deficiency is not associated with any clinical bleeding. Patients with homozygous or compound heterozygous deficiency generally have factor XI levels of less than 15%, and most bleeding manifestations in these patients are related to trauma or surgery, especially at sites of high fibrinolytic activity (e.g., the urinary tract, tonsils, and tooth sockets).[1] Factor XI plays a supportive role in the clotting cascade. It is activated by thrombin and then functions in a positive feedback manner to augment thrombin generation and clot stabilization. Thus, factor XI is primarily required in situations in which there is a significant hemostatic challenge; this explains the mild bleeding diathesis in factor XI deficiency. For patients with severe factor XI deficiency (< 15%) who require surgery, fresh frozen plasma should be used to replenish the plasma level to more than 50%. ε-Aminocaproic acid given orally at a dosage of 3 g three or four times daily is also effective for minor surgical or dental procedures. In a recent

Coming in January

Clinical Essentials
XI Management of Psychosocial Issues in Terminal Illness

2 Dermatology
II Papulosquamous Disorders

4 Gastroenterology
III Diarrheal Diseases

7 Infectious Disease
XXXVI Bacterial Infections of the Central Nervous System

8 Interdisciplinary Medicine
II Bites and Stings

12 Oncology
V Colorectal Cancer

retrospective study of 62 women with severe factor XI deficiency, about 70% of the women did not have any postpartum hemorrhage. Of the 30% who did have postpartum hemorrhage, some had a history of recurrent clinical bleeding. Postpartum hemorrhage had no relationship with the particular abnormal factor XI genotype or with the level of factor XI.[2]

1. Salomon O, Seligsohn U: New observation on factor XI deficiency. Haemophilia 10(suppl 4):184, 2004 [PMID 15479396]
2. Salomon O, Steinberg DM, Tamarin I, et al: Plasma replacement therapy during labor is not mandatory for women with severe factor XI deficiency. Blood Coagul Fibrinolysis 16:37, 2005 [PMID 15650544]

APC versus DIC

Recombinant human activated protein C (APC, or drotrecogin alfa [activated]) has been shown to significantly reduce mortality in patients with severe sepsis (mortality was 24.7% in patients given APC versus 30.8% in patients given placebo). Although it is associated with a slightly increased risk of bleeding, APC appears to be an effective agent in the treatment of severe disseminated intravascular coagulation (DIC) in patients with sepsis, even for patients with normal protein C levels.[1] In large randomized trials, neither recombinant tissue factor pathway inhibitor nor antithrombin concentrate reduced mortality in septic patients.[2] In cases of DIC associated with solitary or multiple hemangiomas, the hemangiomas can be excised when they are localized and occasionally show a good response to local irradiation. Attempts to control DIC with heparin, corticosteroids, aspirin, and estrogens have not been successful. The key to successful management of DIC associated with certain snakebites is identification of the type of snake and prompt administration of appropriate antivenin.

1. Dhainaut JF, Yan SB, Joyce DE, et al: Treatment effects of drotrecogin alfa (activated) in patients with severe sepsis with or without overt disseminated intravasular coagulation. J Thromb Haemost 2:1924, 2004 [PMID 15550023]
2. Abraham E, Reinhart K, Opal S, et al: Efficacy and safety of tifacogin (recombinant tissue factor pathway inhibitor) in severe sepsis: a randomized controlled trial. OPTIMIST Trial Study Group. JAMA 290:238, 2003 [PMID 12851279]

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (required by 39 U.S.C. 3685). 1. Title of publication: ACP Medicine. 2. Publication number: 482-310. 3. Date of filing: October 1, 2005. 4. Frequency of issue: monthly. 5. Number of issues published annually: 12. 6. Annual subscription price: $329.00. 7. Complete mailing address of known office of publication: 111 Eighth Ave., Suite 700, New York, NY 10011. 8. Complete mailing address of the headquarters or general business offices of the publisher: 111 Eighth Ave., Suite 700, New York, NY 10011. 9. Full names and complete mailing addresses of chairman, editor-in-chief, and editor: Publisher: Cynthia Chevins, 111 Eighth Ave., Suite 700, New York, NY 10011. Editor-in-Chief: David C. Dale, M.D., University of Washington Medical Center, Seattle, WA 98195. Editor: Daniel Federman, M.D., Harvard Medical School, Boston, MA 02115. 10. Owner/Stockholders: Emdeon, 669 River Drive, Center 2, Elmwood Park, NJ 07407; FMR Corporation, 82 Devonshire Street, Boston, MA 02109; Martin J. Wygod, c/o Emdeon, 669 River Drive, Center 2, Elmwood Park, NJ 07407. 11. Known bondholders, mortgagees, and other security holders owning 1 percent or more of total amount of bonds, mortgages, or other securities: none. 12. For completion by nonprofit organizations authorized to mail at special rates (Section 432.12 DMM). The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: not applicable to this organization. 13. Publication Title: ACP Medicine. 14. Issue Date for Circulation Data Below: October 2005. 15. Extent and nature of circulation: A. Total number of copies printed (net press run): average number of copies of each issue during preceding 12 months: 10,000; actual number of copies of single issue published nearest to filing date: 10,000. B. Paid circulation: 1. Paid/requested outside-county mail subscriptions stated on form 3541: average number of copies of each issue during preceding 12 months: 7,665; actual number of copies of single issue published nearest to filing date: 7,371. 2. Paid in-county subscriptions: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. 3. Sales through dealers and carriers, street vendors, and counter sales and other non-USPS paid distribution: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. 4. Other classes mailed through the USPS: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. C. Total paid circulation (sum of B1, B2, B3, and B4): average number of copies of each issue during preceding 12 months: 7,665; actual number of copies of single issue published nearest to filing date: 7,371. D. Free distribution by mail—samples, complimentary, and other free copies: 1. Outside county as stated on form 3541: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. 2. In county as stated on form 3541: average number of copies of each issue during preceding 12 months: 178; actual number of copies of single issue published nearest to filing date: 179. 3. Other classes mailed through USPS: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. E. Free distribution outside the mail—carriers or other means: average number of copies of each issue during preceding 12 months: none; actual number of copies of single issue published nearest to filing date: none. F. Total free distribution (sum of D and E): average number of copies of each issue during preceding 12 months: 178; actual number of copies of single issue published nearest to filing date: 179. G. Total distribution (sum of C and F): average number of copies of each issue during preceding 12 months: 7,843; actual number of copies of single issue published nearest to filing date: 7,550. H. Copies not distributed: average number of copies of each issue during preceding 12 months: 2,157; actual number of copies of single issue published nearest to filing date: 2,450. I. Total (sum of G and H): average number of copies of each issue during preceding 12 months: 10,000; actual number of copies of single issue published nearest to filing date: 10,000. J. Percent paid and/or requested circulation (C divided by G times 100): average number of copies of each issue during preceding 12 months: 97.7%; actual number of copies of single issue published nearest to filing date: 97.6%. 16. Publication of Statement of Ownership. Will be printed in the December 2005 issue of this publication. 17. I certify that the statements made by me above are correct and complete: (Signed) Cynthia Chevins, Publisher.

ACP MEDICINE
A Publication of the American College of Physicians
Vol. 28 No. 11 November 2005

What's New in ACP Medicine

www.acpmedicine.com

PRACTICE OF MEDICINE

Announcing *ACP Medicine* 2006

DAVID C. DALE, MD, FACP
Editor-in-Chief, ACP Medicine

We are proud to announce the publication of the 2006 edition of *ACP Medicine*, a medical textbook designed for the practicing physician.

The chapters in *ACP Medicine* are written with the goal of supporting the diagnostic and therapeutic decisions made daily in clinical practice. Our authors, who are among today's medical leaders, base their recommendations on the best current evidence, as well as on their substantial clinical experience. Because we realize that medicine is more than just searching for the latest study results, each chapter also explains the pathophysiology essential for making informed decisions and puts medical advances into context within existing knowledge.

The 2006 bound edition of *ACP Medicine* is larger than ever, with more than 3,000 pages, over 1,100 figures and algorithms, nearly 1,000 tables, and over 16,500 references. We start this edition off with a discussion of one of the key issues facing physicians in practice today: performance measurement. In the front matter of Volume 1, we have added a special review by Kevin B. Weiss, M.D., F.A.C.P., and colleagues of current work in performance measures. Dr. Weiss, who is Chair of the ACP Performance Measures Subcommittee, includes in his review a discussion of how these measures are struc-

continued on page 2

THIS MONTH'S UPDATES

4 GASTROENTEROLOGY

VI Gallstones and Biliary Tract Disease

KIMBERLY M. PERSLEY, MD
RAJEEV JAIN, MD, FACP
University of Texas Southwestern Medical Center at Dallas

The Murphy Sign Validated

Physical examination in patients with acute cholecystitis may reveal upper quadrant subcostal tenderness and pain on inspiration, often with inspiratory arrest (the Murphy sign). Of all physical examination findings, the Murphy sign has the highest positive likelihood ratio (LR) for acute cholecystitis (LR, 2.8; 95% confidence interval, 0.8 to 8.6).[1]

1. Trowbridge RL, Rutkowski NK, Shojania KG: Does this patient have acute cholecystitis? JAMA 289:80, 2003 [PMID 12503981]

Early Is Best for Laparoscopic Cholecystectomy

In acute cholecystitis, laparoscopic cholecystectomy should be performed within 96 hours of onset of symptoms because the increasing inflammatory changes that occur over time have been implicated in bile duct injury; these changes may necessitate converting the procedure to an open cholecystectomy.[1] Early laparoscopic cholecystectomy is recommended for acute cholecystitis, because a delay in surgery does not reduce morbidity, mortality, rate of conversion to open surgery, or mean hospital stay.[2,3] In skilled hands, the laparoscopic procedure carries approximately the same risk as that of open cholecystectomy, but it is associated with much less postoperative pain and a shorter convalescence. In patients with cirrhosis, laparoscopic cholecystectomy is performed for more emergent reasons and is associated with higher morbidity; however, the laparoscopic approach offers advantages of less blood loss, shorter operative time, and shorter length of hospitalization in patients with compensated cirrhosis.[4] In addition, laparoscopic cholecystectomy can be safely performed during pregnancy. Laparoscopic cholecystecto-

continued on page 2

In This Issue

PRACTICE OF MEDICINE

continued from page 1

tured and what they will mean for the individual physician. Other examples of expanded coverage include new chapters on breast mass, pelvic mass, and the Pap smear, which have been added to our Gynecology and Women's Health section, as well as a new rheumatology chapter on fibromyalgia. In addition, we have updated our cardiac resuscitation and adult preventive care chapters to include the latest guidelines. Overall, 45% of the material in the book either has been updated or is new.

ACP Medicine is available in four formats. In addition to the bound edition, we also publish a CD-ROM edition, our original loose-leaf edition, and an online edition, available at www.acpmedicine.com. Both the loose-leaf and the online editions are updated monthly. Our Web site provides more detailed information on *ACP Medicine*, including one development we are very excited about: the launch of a peer review process for new or updated chapters. While this process has only just started, new or revised chapters are now read before publication by two to four peer reviewers, who are invited to comment formally on content, clarity, the weighting of topic coverage, and the quality of evidence; they are also encouraged to suggest additional material for inclusion (especially information on therapies). We list our new peer review board in the front matter. Included free with purchase of the bound edition is 3 months of access to *ACP Medicine Online*.

ACP Medicine provides an economical online CME program, through which you can earn 120 Category 1 CME credits a year. The program is conveniently linked to the related chapters and has an easy-to-use logbook, which bookmarks your tests so that you can return to them later. It also tracks your credits, and you can print out an award certificate as needed.

We want our readers to play a role in the continuing evolution of *ACP Medicine*. For example, in response to numerous requests, we have included an index at the end of both Volume 1 and Volume 2 of the bound edition to make it easier for you to look up what you need right away; our recently redesigned Web site, with more flexible search and navigational features, reflects reader input as well. We welcome your ongoing comments and questions.

I'm sure you will find *ACP Medicine* to be a convenient and helpful tool for your daily practice of medicine.

daviddalemd@webmd.net

THIS MONTH'S UPDATES

continued from page 1

my is less expensive than minilaparotomy or open cholecystectomy in high-volume surgery.[5]

1. Bhattacharya D, Ammori BJ: Contemporary minimally invasive approaches to the management of acute cholecystitis: a review and appraisal. Surg Laparosc Endosc Percutan Tech 15:1, 2005 [PMID 15714147]

2. Shikata S, Noguchi Y, Fukui T: Early versus delayed cholecystectomy for acute cholecystitis: a meta-analysis of randomized controlled trials. Surg Today 35:553, 2005 [PMID 15976952]

3. Papi C, Catarci M, D'Ambrosio L, et al: Timing of cholecystectomy for acute calculous cholecystitis: a meta-analysis. Am J Gastroenterol 99:147, 2004 [PMID 14687156]

4. Puggioni A, Wong LL: A metaanalysis of laparoscopic cholecystectomy in patients with cirrhosis. J Am Coll Surg 197:921, 2003 [PMID 14644279]

5. Nilsson E, Ros A, Rahmqvist M, et al: Cholecystectomy: costs and health-related quality of life: a comparison of two techniques. Int J Qual Health Care 16:473, 2004 [PMID 15557357]

Which Imaging Study for Common Duct Stones?

Transabdominal ultrasonography (TUS) may detect only 50% of common bile duct stones; however, it can

ACP MEDICINE

A Publication of the American College of Physicians

Published by WebMD Inc.

ACP Medicine (ISSN 1548-9345) (USPS 482-310), formerly WebMD Scientific American® Medicine, is published monthly by WebMD Professional Publishing, 111 Eighth Avenue, Suite 700, New York, NY 10011. Periodical postage paid at New York, NY, and at additional mailing offices. Individual subscription rates–USA, its possessions, and Canada: $329 for the first year ($219 for residents and students) and $289 for renewals ($219 for residents and students). Institutional subscription rates–USA, its possessions, and Canada: $429 for the first year and $349 for renewals. Separate shipping and handling apply. POSTMASTER: Send address changes to ACP Medicine,WebMD Professional Publishing, WebMD Inc., P.O. Box 1819, Danbury, CT 06813-9663.

FOR ASSISTANCE WITH YOUR SUBSCRIPTION

Please address all inquiries to Fulfillment Department, WebMD Professional Publishing, WebMD Inc., P.O. Box 1819, Danbury, CT 06813-9663, or call 800-545-0554 or 203-790-2087, fax us at 203-790-2066, or e-mail us at acpmedicine@webmd.net. For change of address, please provide both your new and your old addresses (include your update mailing label if possible); be sure to notify us at least six weeks before you expect to move to avoid interruptions in your service.

YOUR FEEDBACK IS WELCOME

- E-mail: whatsnew@webmd.net
- Write: WebMD Professional Publishing
  111 Eighth Avenue, Suite 700
  New York, NY 10011

www.acpmedicine.com

THIS MONTH'S UPDATES

continued from page 2

often detect dilatation of the common bile duct and intrahepatic ducts. The sensitivity of TUS for detecting common duct stones increases to 76% when ductal dilatation of more than 6 mm is used as the primary end point for choledocholithiasis. CT is no more sensitive or specific than TUS. Cholescintigraphy may show common bile duct obstruction, particularly when symptoms are of recent onset, but not all common bile duct stones will cause complete bile duct obstruction. Magnetic resonance cholangiopancreatography (MRCP) and endoscopic ultrasonography have similar accuracies in detecting common bile duct stones. MRCP is noninvasive and may be preferred in cases where the suspicion of choledocholithiasis is mild to moderate.[1,2]

1. Kondo S, Isayama H, Akahane M, et al: Detection of common bile duct stones: comparison between endoscopic ultrasonography, magnetic resonance cholangiography, and helical-computed-tomographic cholangiography. Eur J Radiol 54:271, 2005 [PMID 15837409]

2. Aube C, Delorme B, Yzet T, et al: MR cholangiopancreatography versus endoscopic sonography in suspected common bile duct lithiasis: a prospective, comparative study. AJR Am J Roentgenol 184:55, 2005 [PMID 15615951]

2 DERMATOLOGY

X Malignant Cutaneous Tumors

ALLAN C. HALPERN, MD
PATRICIA L. MYSKOWSKI, MD
Joan and Sanford I. Weill Medical College of Cornell University

Where Skin Cancer Rates Are Climbing

Several lines of data suggest significant increases in the incidence of cutaneous squamous cell carcinoma (SCC). In Australia, for example, the incidence of SCC increased by 51% between the years 1985 and 1990. In the United States, some of the highest rates of nonmelanoma skin cancer have been detected in the southwest. A population-based survey in New Mexico found the incidence of SCC doubled in both males and females between 1978 and 1999.[1]

1. Athas WF, Hunt WC, Key CR: Changes in nonmelanoma skin cancer incidence between 1977–1978 and 1998–1999 in Northcentral New Mexico. Cancer Epidemiol Biomarkers Prev 12:1105, 2003 [PMID 14578151]

Staying Up to Date on Skin Cancer Treatment

Small SCCs evolving from an actinic keratosis can be adequately treated with simple curettage and electrodesiccation. Larger actinic lesions, as well as lesions arising in non–sun-exposed areas of skin, are best treated with definitive surgical excision with confirmation of negative margins. High-risk, ill-defined lesions, especially those occurring in the surgically sensitive areas of the face, genitalia, hands, and feet, are often best treated by Mohs micrographic surgery.

Fractionated radiation therapy is an alternative treatment of primary SCC in older patients who are poor surgical candidates. The benefits of adjuvant radiation therapy are less clear, as are the benefits of sentinel lymph node biopsy and elective lymph node dissection for patients with high-risk SCC of the head and neck.

Cytotoxic chemotherapy and biologic response modifiers have been used in patients who have advanced SCC; this therapeutic approach has been reported to have complete response rates up to 68%, but there are few long-term survivors. Actinic keratoses are treated with cryotherapy, curettage, topical therapies (e.g., fluorouracil, imiquimod, or diclofenac), photodynamic therapy, and laser resurfacing to prevent progression to SCC.[1] Regularly updated guidelines for the treatment of SCC and basal cell carcinoma are available through the National Comprehensive Cancer Network (NCCN).[2]

1. Jorizzo JL: Current and novel treatment options for actinic keratosis. J Cutan Med Surg 8(suppl 3):13, 2004 [PMID 15647860]

2. Basal Cell and Squamous Cell Skin Cancers. Clinical Practice Guidelines in Oncology, Version 2.2005. National Comprehensive Cancer Network, Jenkintown, Pennsylvania, 2005 http://www.nccn.org/professionals/physician_gls/PDF/nmsc.pdf

Diagnostic Aids for Melanoma

Several specialized aids to the diagnosis of melanoma in patients with dysplastic nevi are under development. Dermoscopy entails the use of a handheld otoscopelike device to magnify a pigmented lesion while applying pressure and oil to the surface. The technique allows the visualization of pigment patterns and features not apparent with simple visual inspection. With experience and training, dermoscopy can be a useful aid in distinguishing melanoma from benign pigmented lesions; however, when used inexpertly, dermoscopy may actually decrease diagnostic accuracy.[1,2]

Another aid to melanoma detection in high-risk individuals is photographically assisted follow-up. A baseline set of whole-body photographs of the skin is used during self-examination and professional follow-up examination to assess change in the lesions. This procedure helps to prevent unnecessary excision of stable lesions and improves the sensitivity of examinations in detecting change. New imaging technologies such as in vivo confocal scanning laser microscopy hold promise for future improvements in the noninvasive diagnosis of melanoma.[3]

1. Naeyaert JM, Brochez L: Clinical practice. Dysplastic nevi. N Engl J Med 349:2233, 2003 [PMID 14657431]

2. Braun RP, Rabinovitz HS, Oliviero M, et al: Dermoscopy of pigmented skin lesions. J Am Acad Dermatol 52:109, 2005 [PMID 15627088]

3. Marghoob AA, Swindle LD, Moricz CZ, et al: Instruments and new technologies for the in vivo diagnosis of melanoma. J Am Acad Dermatol 49:777, 2003 [PMID 14576657]

Kinder Cuts for Melanoma

Primary cutaneous melanoma is managed surgically with definitive reexcision. The wide excisions of the past have given way to resections with more modest margins. Multiple prospective, randomized trials have investigated the surgical resection of primary cutaneous melanoma utilizing different margins of resection; these studies have focused on varied and overlapping patient populations. On the basis of these data, the National Comprehensive Cancer Network recommends resection margins of 1 cm for melanomas less than 1 mm in thickness; margins of 1 to 2 cm for melanomas between 1 and 2 mm in thickness; and margins of 2 cm for melanomas greater than 2 mm in thickness.[1,2] Primary closure and

FDA Approval Report

The following is selected from the FDA's list of recently approved products. Complete, updated information on FDA approvals and notifications is available on the FDA Web site (http://www.fda.gov).

New Drug Combination for Heart Failure in Black Patients

The FDA has approved a combination of two older drugs, hydralazine and isosorbide dinitrate—neither of which had been approved for heart failure—for treatment of heart failure in black patients.

Generic Name: Hydralazine and isosorbide dinitrate

Brand Name: BiDil

Manufacturer: NitroMed, Inc., Lexington, Massachusetts

The approval of hydralazine and isosorbide dinitrate was based in part on the results of the African-American Heart Failure Trial (A-HeFT). A-HeFT, which included 1,050 self-identified black patients with severe heart failure who had already been treated with the best available therapy, was conducted because two previous trials of the use of these drugs for severe heart failure found no benefit in the general population of patients but suggested a benefit in black patients. In A-HeFT, patients on hydralazine and isosorbide dinitrate had 43% lower mortality and 39% lower rates of hospitalization for heart failure than patients receiving placebo; treated patients also experienced a decrease in symptoms of heart failure.

As an antihypertensive agent, hydralazine is an arterial dilator and decreases the work of the heart. Isosorbide dinitrate is an antianginal agent that causes dilatation of both veins and arteries. Isosorbide seems to work by releasing nitric oxide at the blood vessel wall, but its effect usually wears off after half a day. Hydralazine may prevent this loss of effect, but how the two drugs work together is not fully known. Common side effects with the use of hydralazine and isosorbide dinitrate include headache and dizziness.

Source:

FDA Approves BiDil Heart Failure Drug for Black Patients. FDA News. U.S. Food and Drug Administration, June 23, 2005 (http://www.fda.gov/bbs/topics/NEWS/2005/NEW01190.html)

reconstructive flaps are preferable, cosmetically and functionally, to skin grafts and should be used instead of grafts whenever possible.

1. Melanoma. Clinical Practice Guidelines in Oncology, Version 2.2005. National Comprehensive Cancer Network, Jenkintown, Pennsylvania, 2005 http://www.nccn.org/professionals/physician_gls/PDF/melanoma.pdf

2. Cook J: Surgical margins for resection of primary cutaneous melanoma. Clin Dermatol 22:228, 2004 [PMID 15262309]

13 PSYCHIATRY

IX The Eating Disorders

W. STEWART AGRAS, MD
Stanford University School of Medicine

Medical Complications of Refeeding in Anorexia

The medical treatment of anorexia nervosa includes rehydration and correction of electrolyte disturbances. The patient should be examined frequently, and the levels of electrolytes (sodium, potassium, chloride, bicarbonate, phosphorus, and magnesium) should be monitored frequently. It is important to warn the patient that temporary edema may occur because of rehydration and electrolyte correction; this information helps allay the patient's concerns about weight and shape. Congestive heart failure occasionally occurs early in the refeeding period; hence, careful evaluation of cardiovascular status during the early stages of refeeding is necessary. Hypophosphatemia may also develop at this time, when reserves are depleted.[1]

1. Ornstein RM, Golden NH, Jacobson MS, et al: Hypophosphatemia during nutritional rehabilitation in anorexia nervosa: implications for refeeding and monitoring. J Adolesc Health 32:83, 2003 [PMID 12507806]

Failure and Success in Anorexia Pharmacotherapy

Various psychopharmacologic approaches to the treatment of anorectic patients have been evaluated in controlled trials and have included the use of tricyclic antidepressants, fluoxetine, lithium, and antipsychotic compounds. There are various rationales for the use of these compounds—for instance, the observation that patients who take antipsychotic agents such as chlorpromazine tend to gain weight and the finding that anorectics often suffer from depressive symptoms. The results of these trials have been disappointing: no evidence of clinically useful effectiveness for any of these medications in the acute phase of treatment has been established.[1] Moreover, patients with anorexia nervosa are often reluctant to take medication, particularly if the medication might lead to weight gain. Symptoms such as depression often disappear as weight is gained.

Because obsessive-compulsive disorder is associated with anorexia nervosa and because the genetic findings implicate serotonin metabolism, interest has focused on the use of the selective serotonin reuptake inhibitors (SSRIs) that are effective in patients with obsessive-compulsive disorder. Small controlled and uncontrolled trials suggest that fluoxetine may be useful in helping patients maintain their weight gain. Open-label trials also suggest that the newer antipsychotics, such as olanzepine, may be effective in

reducing depression, anxiety, and core eating-disorder symptoms (e.g., food obsessions and distorted body image) and in enhancing weight gain.[2]

1. Powers PS, Santana C: Available pharmacological treatments for anorexia nervosa. Expert Opin Pharmacother 5:2287, 2004 [PMID 15500375]

2. Barbarich NC, McConaha CW, Gaskill J, et al: An open trial of olanzepine in anorexia nervosa. J Clin Psychiatry 65:1480, 2004 [PMID 15554759]

The Most Effective Treatment for Bulimia

It is generally agreed that the most effective treatment of bulimia nervosa is an outpatient trial of cognitive-behavioral therapy. A large number of controlled trials have demonstrated that such therapy is more effective than no treatment, other forms of psychotherapy, antidepressant medication, and placebo.[1] The usual course of treatment is 20 sessions over 6 months. Approximately 50% of patients who complete treatment with cognitive-behavioral therapy recover, and an additional 20% to 30% show significant improvement. Gains made during cognitive-behavioral therapy are well maintained over the long term.[1]

1. Hay PJ, Bacultchuk J, Stefano S: Psychotherapy for bulimia nervosa and binge eating. Cochrane Database Syst Rev (3):CD000562, 2004 [PMID 15266434]

14 RESPIRATORY MEDICINE

IV Focal and Multifocal Lung Disease

GERALD W. STATON, JR, MD, FACP
Emory University School of Medicine

Imaging and Managing Alveolar Cell Carcinoma

Alveolar cell carcinoma may result in one or more areas of air-space opacity, presenting as an area of focal indistinct infiltrate or, sometimes, mimicking lobar pneumonia. Chest CT often shows areas of "ground glass opacity"; less commonly, the infiltrates are consolidative.[1] Patients cough and produce mucoid sputum; a few patients produce large volumes of sputum (bronchorrhea) that in rare instances has a salty taste. Weight loss and malaise are common. Fever and chills are absent. Metastases are less common than with other primary lung neoplasms, and the course of the illness is longer. Alveolar cell carcinoma is not related to smoking. Diagnostic tests should begin with sputum cytology, followed by fiberoptic bronchoscopy with bronchoalveolar lavage and transbronchial biopsy or, if needed, open lung biopsy by a traditional or a video-assisted thoracoscopic approach.

1. Sabloff BS, Truong MT, Wistuba II, et al: Bronchioalveolar cell carcinoma: radiologic appearance and dilemmas in the assessment of response. Clin Lung Cancer 6:108, 2004 [PMID 15476596]

Pneumonia by Mouth

Lipoid pneumonia is a noninfectious, inflammatory lung disorder caused by the aspiration of mineral oil or other oily substances. It is most common in elderly patients and others with impaired swallowing. With or without impaired swallowing, lipoid pneumonia can result from the use of petroleum jelly or other oily substances applied to the lips or nose to relieve chronic dryness or mineral oil taken by mouth for relief of constipation. The most common symptom is a chronic cough, which may be caused by coexisting lung disease rather than lipoid pneumonia; fever is uncommon. Often, the disease is discovered on a routine chest radiograph that shows a focal, dense infiltrate, usually in a lower lobe or in the right middle lobe.[1]

The radiographic appearance of such an infiltrate in a relatively asymptomatic patient suggests chronic pneumonia or lung cancer. CT scanning may show an extremely low density infiltrate produced by accumulated lipid; the density typically ranges from –60 to –150 Hounsfield units (water is 0). In contrast, the density of lung cancers usually ranges from +60 to +150 Hounsfield units.

1. Baron SE, Haramati LB, Rivera VT: Radiological and clinical findings in acute and chronic exogenous lipoid pneumonia. J Thorac Imaging 18:217, 2003 [PMID 14561906]

Ankylosing Spondylitis and the Lungs

Chest wall expansion is impaired in ankylosing spondylitis, and lung volumes are somewhat reduced. Patients may develop characteristic dense fibrous or fibrobullous infiltrates that are often limited to the upper lung zones. These chest radiographic findings mimic those of tuberculosis, and cavitation may occur. Once cavities have developed, abnormal air spaces may become colonized with *Aspergillus* species, causing fungus balls and even a locally invasive disease termed chronic necrotizing aspergillosis. Studies of groups of patients with ankylosing spondylitis using CT have found interstitial lung disease that is more diffuse and not confined to the apices of the lung.[1] The lung disease associated with ankylosing spondylitis does not progress to respiratory insufficiency. Dyspnea on exertion and nonproductive cough are the usual symptoms.

Once *Aspergillus* colonization or infection has occurred, symptoms from the local fungal infection may predominate; such symptoms include hemoptysis, productive cough, and mild to moderate constitutional symptoms. Locally invasive *Aspergillus* infection may extend directly to the pleura but almost never spreads hematogenously to distant sites.

1. El Maghraoui A, Chaouir S, Abid A, et al: Lung findings on thoracic high-resolution computed tomography in patients with ankylosing spondylitis: correlations with disease duration, clinical findings and pulmonary function testing. Clin Rheumatol 23:123, 2004 [PMID 15045626]

CLINICAL ESSENTIALS

VI Occupational Medicine

LINDA ROSENSTOCK, MD, MPH, FACP
University of California, Los Angeles, School of Public Health

MARK R. CULLEN, MD
Yale University School of Medicine

Occupational Medicine on the Internet

Awareness of the impact of the work environment on health has increased dramatically in the past few decades. An increasing amount of information about occupational medicine is available on the Internet from the National Institute for Occupational Safety and Health (http://www.cdc.gov) and the Occupational Safety and Health Administration (http://www.osha.gov).

A Vicious Synergy

Many workplace hazards and toxins interact with one another and with nonoccupational factors to cause disease. Dose-response correlations for industrial hazards may be markedly shifted in the presence of other haz-

ards, habits, or medications. An important example is the likelihood of disease resulting from thermal stress (i.e., heat or cold) in the presence of hemodynamically active agents, such as calcium channel blockers, autonomic agents, and diuretics. Likewise, the effects of vibration trauma on wrists and digits may be amplified by nicotine. The effects of one hazard may be significantly altered in the presence of another; for example, the combined effect of noise and solvents on hearing loss and of asbestos and smoking on lung cancer are greater than the effect of exposure to each hazard alone.

Cancer from the Workplace

Only a small number of hazards found in the workplace have been clearly established as carrying substantive cancer risk for workers. An additional group of hazards are suspected, but additional studies are needed. The list of potential carcinogens is expanding; for example, evidence suggests that exposure to cadmium may play a role in the development of prostate cancer. Studies provide some indication that workers in print shops, service station employees, farm product vendors, horticulturists, farmers, and aircraft mechanics are at increased risk for renal cell carcinoma [*see Table, right*].[1]

1. Zhang Y, Cantor KP, Lynch CF, et al: A population-based case-control study of occupation and renal cell carcinoma risk in Iowa. J Occup Environ Med 46:235, 2004 [PMID 15091286]

Established Occupational Carcinogens

Cancer Site	*Hazard*	*Setting*
Lung	Asbestos	Insulation, textiles
	Ionizing radiation	Uranium mining
	Arsenic	Refining
	Polyaromatic hydrocarbons	Coke ovens
	Nickel	Nickel refining
	Chromium	Tanning, pigments
	Alkylating agents	Chemical industry
	Silica	Mining, stonecutting
	Ceramic fibers	Insulation
	Formaldehyde	Chemicals, plastics
	Beryllium	Nuclear weapons, aerospace industry
	Cadmium	Batteries
	Acrylonitrile	Plastics
	1,3-Butadiene	Rubber, plastics
Pleura and peritoneum	Asbestos	Construction materials
Upper respiratory tract	Wood dust	Carpentry
	Nickel	Refining
	Chromium	Plating
	Asbestos	Friction products
	Formaldehyde	Chemicals, plastics
Urinary bladder	Benzidine and related amines	Dyes, chemicals
	Polyaromatic hydrocarbons	Aluminum reduction
Liver	Vinyl chloride monomer	Plastics
	Arsenic	Pesticides
Upper GI tract	Asbestos	Shipbuilding
	Coal dust	Mining
	Acrylonitrile	Plastics
Hematologic system	Benzene	Chemicals, rubber
	Ionizing radiation	Defense industry
	Ethylene oxide	Chemicals, sterilizers
Soft tissue	Dioxin	Chemical industry
Brain	Vinyl chloride	Chemical industry
	Formaldehyde	Chemical industry

When Work Gets under Their Skin

Overwhelmingly, the major skin problem in the workplace remains dermatitis, either irritant induced or caused by allergy. Many agents may be responsible, including organic and inorganic chemicals, plastics and rubber, oils and lubricants, metals and construction materials, paints, and coatings.[1] The key to correct diagnosis is the history of skin contact and the temporal relation between contact and manifestations. Unfortunately, there is seldom a perfect or obvious correlation between the two, and some sleuthing is necessary, especially to discern the extent to which chemical contact may spread to places like the groin or areas where hand contact occurs. Airborne exposure may cause lesions in apparently untouched areas, such as the face; such occurrences are signs of likely hypersensitivity. Vexingly, symptoms do not always abate dramatically over weekends or during short periods in which exposure is avoided; removing the patient from the toxin for a week or two may be necessary to observe response. This, combined with observation of the patient during reexposure, is often the most valuable diagnostic test. Patch testing, performed by an experienced clinician aware of the exposures of concern, may be useful in difficult cases, though the clinician should keep in mind that irritants may yield false negative results and that even many healthy atopic persons will experience reactions to common contactants, such as nickel.[2]

1. Pratt MD, Belsito DV, DeLeo VA, et al: North American Contact Dermatitis Group patch-test results, 2001–2002 study period. Dermatitis 15:176, 2004 [PMID 15842061]

2. Krob HA, Fleischer AB Jr, D'Agostino R Jr, et al: Prevalence and relevance of contact dermatitis allergens: a meta-analysis of 15 years of published T.R.U.E. test data. J Am Acad Dermatol 51:349, 2004 [PMID 15337975]

7 INFECTIOUS DISEASE

XXXIX Infections Due to Mycobacterium leprae and Nontuberculous Mycobacteria

MICHAEL K. LEONARD, JR, MD
HENRY M. BLUMBERG, MD
Emory University School of Medicine

Diagnosing MAC Infection

Mycobacterium avium complex (MAC) comprises two closely related species, *M. avium* and *M. intra-*

cellulare. MAC is encountered most often as an opportunistic pathogen in immunosuppressed patients, especially those with advanced HIV infection or AIDS. However, MAC infection can occur in immunocompetent hosts; it is especially likely to cause pulmonary infections in persons with underlying chronic lung disease. The American Thoracic Society has developed criteria for the diagnosis of pulmonary infection by MAC, as well as infections from other nontuberculous mycobacteria [*see Table, below*]. Criteria for the diagnosis of these infections include repeated isolation of a potentially pathogenic species, the absence of other pathogens, and a compatible clinical, radiologic, or pathologic picture.

Adverse Reactions to Beneficial Effects

Immune reconstitution syndromes can be an adverse event in the treatment of patients with HIV infection who have disseminated infection from MAC. These syndromes sometimes occur in HIV-infected patients with unrecognized MAC who are started on antiretroviral agents. Pathogen-specific immune responses to MAC can occur in the first few months after the initiation of highly active antiretroviral therapy.[1] Immune reconstitution reactions to MAC usually manifest themselves as fever and lymphadenopathy (peripheral, intrathoracic, or intra-abdominal). Bacteremia does not occur. In severe cases, treatment with corticosteroids or nonsteroidal anti-inflammatory drugs may be necessary to ameliorate the symptoms associated with the immune response.

1. Lawn SD, Bekker L, Miller RF: Immune reconstitution associated with mycobacterial infections in HIV-infected individuals receiving antiretrovirals. Lancet Infect Dis 5:361, 2005 [PMID 15919622]

Mycobacteria in the Fast Lane

Rapidly growing mycobacteria produce mature growth on agar plates within 7 days. The skin and soft tissues are the most common sites of infection from these organisms; indeed, of all nontuberculous mycobacteria species, rapidly growing mycobacteria are the most common cause of skin and soft tissue infections. Furunculosis from *Mycobacterium fortuitum* and other organisms in the *M. fortuitum* group has been reported in association with whirlpool footbaths in customers of nail salons.[1] Shaving the legs with a razor before a pedicure was a risk factor for infection. In these cases, molecular typing techniques were used to match rapidly growing mycobacteria recovered from the footbaths with isolates recovered from patients.

Rapidly growing mycobacteria are also associated with skeletal infections (e.g., infections of bone, joint, or tendon) and pulmonary disease. Nosocomial or health care–associated infections caused by these organisms include infections from colonization of long-term venous access devices or peritoneal dialysis catheters, postinjection abscesses, and surgical wound infections (e.g., after cardiac bypass surgery, augmentation mammoplasty, facelifts,[2] and other plastic surgery; postoperative keratitis has been reported after ophthalmic surgery). Clusters of infections and outbreaks of true infection and pseudoinfections have occurred from contaminated fluids, irrigation with or exposure to tap water, injectable medicines, and topical skin solutions and markers.

1. Gira AK, Reisenauer AH, Hammock L, et al: Furunculosis due to *Mycobacterium mageritense* associated with footbaths at a nail salon. J Clin Microbiol 42:1813, 2004 [PMID 15071058]

2. *Mycobacterium chelonae* infections associated with face lifts—New Jersey, 2002–2003. MMWR Morb Mortal Wkly Rep 53:192, 2004 (Erratum in MMWR Morb Mortal Wkly Rep 53:246, 2004) [PMID 15017374]

Coming in December

3 Endocrinology
X Obesity

5 Hematology
XIII Thrombocytopenia and Other Primary Hemostatic Disorders
XV Disorders of Coagulation

7 Infectious Disease
VIII Infections Due to Escherichia coli and Other Enteric Gram-Negative Bacilli
XIV Antimicrobial Therapy
XXXVIII Mycotic Infections in the Compromised Host

16 Women's Health
IX Medical Complications in Pregnancy

American Thoracic Society Criteria for Diagnosis of Nontuberculous Mycobacterial Lung Disease*

Radiographic Criteria	*Laboratory Criteria*
Infiltrate, reticulonodular infiltrate, or cavitary disease on chest x-ray *or* Multifocal bronchiectasis or multiple small nodules on high-resolution chest CT	Two positive respiratory (sputum and/or bronchial wash) cultures within 12 mo, if one or both specimens are AFB smear positive *or* Three positive sputum/bronchial wash cultures within 12 mo, if none of the specimens are AFB smear positive *or* In patients unable to produce sputum, one positive bronchial wash culture with a 2+, 3+, or 4+ AFB smear and/or growth on solid media *or* A transbronchial biopsy yielding NTM *or* Biopsy showing mycobacterial histopathologic features (granulomatous infiltration and/or AFB) and one or more sputa/bronchial washings are positive for NTM

*For symptomatic patients who are HIV seropositive or HIV seronegative.
AFB—acid-fast bacilli CT—computed tomography NTM—nontuberculous mycobacteria

SPECIAL – ONLINE ONLY!

New . . .

16 WOMEN'S HEALTH

XXI Musculoskeletal Problems in the Female Athlete

CAROL C. TEITZ, MD
University of Washington School of Medicine

Not Just Rest for Stress Fractures

Recommending that the patient rest the affected body part—the standard treatment for stress fractures—is insufficient for female athletes. Instead, the physician also needs to address any features of the athlete's training habits, nutritional and menstrual status, or equipment that may have contributed to the stress fracture. This will ensure that once the fracture heals, the patient will not be exposed to the same factors that led to the fracture.

Nonsteroidal anti-inflammatory drugs (NSAIDs) are contraindicated in the treatment of pain related to stress fractures, because there is evidence that NSAIDs may interfere with fracture healing.[1,2] When pharmacologic pain control is needed, acetaminophen usually suffices.

1. Warden SJ: Cyclo-oxygenase–2 inhibitors: beneficial or detrimental for athletes with acute musculoskeletal injuries? Sports Med 35:271, 2005 [PMID 15831058]

2. Seidenberg AB, An YH: Is there an inhibitory effect of COX-2 inhibitors on bone healing? Pharmacol Res 50:151, 2004 [PMID 15177303]

Knees at Risk

Noncontact anterior cruciate ligament (ACL) tears are two to eight times more common in females than in males. Investigators agree that there is no one identifiable cause of the increased incidence of ACL injuries in female athletes and that hormonal, anatomic, and proprioceptive factors may play a role. Studies on hormonal influences are inconclusive. In contrast, a study at West Point found that a combination of increased body mass, generalized joint laxity, and decreased intercondylar notch width significantly increased the risk of ACL injury.[1]

1. Uhorchak JM, Scoville CR, Williams GN, et al: Risk factors associated with noncontact injury of the anterior cruciate ligament: a prospective four-year evaluation of 859 West Point cadets. Am J Sports Med 31:831, 2003 [PMID 14623646]

An Ounce for the ACL

Current prevention programs are resulting in a decreased incidence of ACL injuries.[1] Neuromuscular training is the mainstay of these prevention programs and includes one or more of the following: stretching and strengthening exercises; aerobic, agility, kinesthetic, and plyometric exercises; and risk-awareness education. Strength training and endurance training focus on hip abductor and external rotator muscles. Agility and skill drills emphasize rapid directional changes and therefore stimulate maximum cocontractions of agonist-antagonist muscle pairs to increase stiffness and reduce the number of unanticipated joint movements. Agility drills also improve reflex and cortical response time. Kinesthetic training emphasizes keeping the center of gravity forward and the athlete on her toes. Plyometric training decreases the time to peak torque and increases dynamic joint stability by decreasing landing forces and varus/valgus moments through increased muscle activation.

1. Griffin LY, Albohm MS, Arendt EA, et al: Update on ACL injury prevention: theoretical and practical considerations. A review of the Hunt Valley II meeting, February 2005. Am J Sports Med (in press)